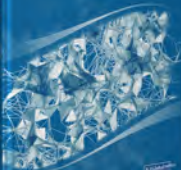


Encyclopedia of

Information Science and Technology

Fourth Edition



MAHMOUD KHALIL-POUR, D.Sc.

Encyclopedia of
Information Science
and Technology
Volume 1

Encyclopedia of Information Science and Technology, Fourth Edition

Mehdi Khosrow-Pour

Information Resources Management Association, USA

Published in the United States of America by

IGI Global
Information Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue
Hershey PA, USA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: <http://www.igi-global.com>

Copyright © 2018 by IGI Global. All rights reserved. No part of this publication may be reproduced, stored or distributed in any form or by any means, electronic or mechanical, including photocopying, without written permission from the publisher. Product or company names used in this set are for identification purposes only. Inclusion of the names of the products or companies does not indicate a claim of ownership by IGI Global of the trademark or registered trademark.

Library of Congress Cataloging-in-Publication Data

Names: Khosrow-Pour, Mehdi, 1951- editor.

Title: Encyclopedia of information science and technology / Mehdi

Khosrow-Pour, editor.

Description: Fourth edition. | Hershey, PA : Information Science Reference,

[2018] | Includes bibliographical references and index.

Identifiers: LCCN 2017000834 | ISBN 9781522522553 (set : hardcover) | ISBN

9781522522560 (ebook)

Subjects: LCSH: Information science--Encyclopedias. | Information
technology--Encyclopedias.

Classification: LCC Z1006 .E566 2018 | DDC 020.3--dc23 LC record available at <https://lccn.loc.gov/2017000834>

British Cataloguing in Publication Data

A Cataloguing in Publication record for this book is available from the British Library.

All work contributed to this book is new, previously-unpublished material. The views expressed in this book are those of the authors, but not necessarily of the publisher.

For electronic access to this publication, please contact: eresources@igi-global.com.

This book is dedicated to the memory of my late father for the love and care that he always displayed for his family, and for teaching me the importance of humanitarianism. Also, to my wife, Olga, and our son, Darius, for filling my life with so much love, joy, and happiness.

Editorial Advisory Board

Ari-Veikko Anttiroiko, *University of Tampere, Finland*
S. Annie Becker, *Florida Institute of Technology, USA*
Shirley M. Fedorovich, *Embry-Riddle Aeronautical University, USA*
Wen-Chen Hu, *University of North Dakota, USA*
Jerzy Kisielnicki, *Warsaw University, Poland*
In Lee, *Western Illinois University, USA*
James A. Rodger, *Indiana University of Pennsylvania, USA*
Ari Sigal, *Catawba Valley Community College, USA*
Lawrence A. Tomei, *Robert Morris University, USA*
John Wang, *Montclair State University, USA*
Liudong Xing, *University of Massachusetts, Dartmouth, USA*
Jordi Vallverdú, *Universitat Autònoma de Barcelona, Spain*

Editorial Review Board

Mifrah Ahmad, *Universiti Teknologi PETRONAS, Malaysia*
Bashar Shahir Ahmed, *University Abdelmalek Essaadi, Morocco*
Daniel Alemneh, *University of North Texas, USA*
Salam Omar Ali, *Brighton College, UAE*
Paul Ankomah, *North Carolina A&T University, USA*
Andrea Atzeni, *Politecnico di Torino, Italy*
Danilo Avola, *Sapienza University, Italy*
Ihuoma Babatope, *Delta State Univeristy Abraka/ Delta State College of Physical Education, Mosogar, Nigeria*
Mihaela Badea, *Petroleum-Gas University of Ploiesti, Romania*
Surajit Bag, *Tega Industries South Africa Pty Ltd, South Africa*
Kannan Balasubramanian, *Mepco Schlenk Engineering College, India*
Neeta Baporikar, *Namibia University of Science and Technology, Nigeria*
Nomathamsanqa (Thami) Batyashe, *Cape Peninsula University of Technology, South Africa*
Clementina Bruno, *University of Piemonte Orientale, Italy*
Claudia Cappelli, *Universidade Federal do Estado do Rio de Janeiro, Brazil*
Calin Ciufudean, *Stefan cel Mare University, Romania*
Ángela Coello, *University of Vigo, Spain*
Eduardo Contreras, *Autonomous University of Coahuila, Mexico*
Caroline Crawford, *University of Houston-Clear Lake, United States*
Phil Crosby, *Curtin University, Australia*
Nicoletta Dessì, *Università degli Studi di Cagliari, Italy*
Nic Dunham, *Unitec Institute of Technology, New Zealand*

Christophe Duret, *Université de Sherbrooke, Canada*
Alexander Fridman, *Russian Academy of Sciences, IIMM KSC RAS, Russia*
Dražena Gašpar, *University of Mostar, Bosnia and Herzegovina*
Ciara Heavin, *University College Cork, Ireland*
Masoud Hemmatpour, *Politecnico di Torino, Italy*
Adamkolo Mohammed Ibrahim, *University of Maiduguri, Nigeria*
Shigeru Ikuta, *Otsuma Women's University, Japan*
Giuseppe Iurato, *University of Palermo, Italy*
George Jamil, *Informações em Rede, Brazil*
Jurgen Janssens, *TETRADE Consulting, Belgium*
Michael Josefowicz, *Nemetics Institute Kolkata, USA*
Prateek Kalia, *I.K Gujral Punjab Technical University, India*
Yefim Katz, *Chestnut Hill College, USA*
Nadim Akhtar Khan, *University of Kashmir, India*
Maria Koleva, *Bulgarian Academy of Sciences, Bulgaria*
Raghuraman Koteeswaran, *SASTRA University, India*
Priya Krishnamoorthy, *SASTRA University, India*
Sefer Kurnaz, *Istanbul Esenyurt University, Turkey*
Trent Larson, *North Carolina A&T University, USA*
Elvira Locuratolo, *ISTI, Italy*
Ronald Lofaro, *Embry-Riddle Aeronautical University, USA*
Geetha Mariappan, *Kamaraj College of Engineering and Technology, India*
M.K. Marichelvam, *Mepco Schlenk Engineering College, India*
Gilberto Marzano, *Rezekne University of Applied Sciences, Latvia*
Lincy Mathews, *M S Ramaiah Institute of Technology, India*
Karim Mezghani, *Al Imam Mohammad Ibn Saud Islamic University (IMSIU), Tunisia*
Emilia Mikolajewska, *Ludwik Rydygier Collegium Medicum in Bydgoszcz, Nicolaus Copernicus University in Torun, Poland*
Dariusz Mikolajewski, *Kazimierz Wielki University, Poland*
Vardan Mkrttchian, *HHH University, Australia*
Jessina Muthee, *Kenyatta University, Kenya*
Antonio Osorio, *University of Minho, Portugal*
Leo Paschoal, *Federal University of Cruz Alta, Brazil*
Prantosh Paul, *Raiganj University, India*
Danilo Piaggese, *FRAmericas, International Knowledge Economy Program (IKEP), USA*
Diana Presada, *Petroleum-Gas University of Ploiesti, Romania*
Mohd Rahman, *Aligarh Muslim University, India*
N. Raghavendra Rao, *FINAIT Consultancy Services, India*
José Rascão, *Polytechnic Institute of Setúbal, Portugal*
Filipa Ribeiro, *University of Porto, Portugal*
Gerasimos Rigatos, *Industrial Systems Institute, Greece*
John Robinson, *Bishop Heber College, India*
Juan Rodriguez, *UNICEN CONICET, Argentina*
Isabel Romero, *University of Castilla-La Mancha, Spain*

Roya Rouzbehani, *University of Tehran, Iran*
John C. Sandvig, *Western Washington University, USA*
Francesca Sgobbi, *University of Brescia, Italy*
Huma Shah, *Coventry University, UK*
Dolly Sharma, *PEC University of Technology, India*
Teay Shawyun, *King Saud University, Saudi Arabia*
Zeljko Stojanov, *University of Novi Sad, Serbia*
Apostolos Syropoulos, *Greek Molecular Computing Group, Greece*
Safiye Turgay, *Sakarya University, Turkey*
Immanuel Umukoro, *Lagos Business School, Pan Atlantic University, Nigeria*
Tugba Ucma Uysal, *Mugla Sitki Koçman University, Turkey*
JT Velikovsky, *University of Newcastle, Australia*
Laura Vettraino, *Learning Community, Italy*
Revathi Viswanathan, *B.S. Abdur Rahman University, India*
Ye Wang, *Zhejiang Gongshang University, China*
Tingshao Zhu, *Institute of Psychology, Chinese Academy of Sciences, China*

List of Contributors

A, Vadivel / <i>National Institute of Technology Trichy, India</i>	212
Abad, Lalaine P. / <i>Department of Education, Philippines</i>	2466
Abdelhafez, Hoda Ahmed / <i>Suez Canal University, Egypt</i>	406
Abdou, Tamer / <i>Suez Canal University, Egypt</i>	7399
Abdrakhmanova, Gulnara / <i>National Research University Higher School of Economics, Russia</i>	4704
Aboud, Sattar J. / <i>University of Bedfordshire, UK</i>	4898
Adhikari, Arnab / <i>Indian Institute of Management Ranchi, India</i>	7234
Adibfar, Alireza / <i>University of Florida, USA</i>	539
Agrawal, Shaila / <i>Maulana Abul Kalam Azad University of Technology, India</i>	4872
Agrey, Renae / <i>The University of Queensland, Australia</i>	5433
Aguilar-Moreno, Estefanía / <i>Universitat Jaume I of Castellón, Spain</i>	3473
Agwae, Uruemu / <i>Carleton University, Canada</i>	1684
Ahmad, Mifrah / <i>Universiti Teknologi PETRONAS, Malaysia</i>	3337
Ahmed, Bashar Shahir / <i>University Abdelmalek Essaadi (LEROSA Laborator), Morocco</i>	1616
Ahmed, Rukhsana / <i>University of Ottawa, Canada</i>	6083
Akahane-Yamada, Reiko / <i>Advanced Telecommunications Research Institute International (ATR), Japan</i>	3850
Akalan, Nazif Ekin / <i>Istanbul University, Turkey</i>	298
Akhtar, Saliha / <i>Seton Hall University, USA</i>	7181
Akolkar, Rajshree Tushar / <i>Zeal College of Engineering and Research, India</i>	4099
Akyuz, Goknur Arzu / <i>University of Turkish Aeronautical Association, Turkey</i>	5285
Alba-Elías, Fernando / <i>Universidad de La Rioja, Spain</i>	1934
Alberg, Dima / <i>Shamoon College of Engineering (SCE), Israel</i>	364
Alemneh, Daniel G. / <i>University of North Texas, USA</i>	6748
Alharbi, Ali H. / <i>Qassim University, Saudi Arabia</i>	1249,3448
Ali, Salam Omar / <i>Brighton Collage Al Ain, UAE</i>	7647
Al-Khalifa, Hend S. / <i>King Saud University, Saudi Arabia</i>	4652
Al-Marzouqi, Mohamed / <i>UAE University, UAE</i>	3043
Almeida, Fernando / <i>Polytechnic Institute of Gaya, Portugal</i>	800
Almeida, Gustavo de Oliveira / <i>Federal University of the State of Rio de Janeiro, Brazil</i>	754
Almunawar, Hasan Jawwad / <i>P. T. Tegar Kupas Mediatama, Indonesia</i>	1101
Almunawar, Mohammad Nabil / <i>Universiti Brunei Darussalam, Brunei</i>	1101,5908,7369
Alnowibet, Khalid / <i>King Saud University, Saudi Arabia</i>	1570
Al-Rousan, Thamer / <i>Isra University, Pakistan</i>	7549

Al-Salman, Abdulmalik / King Saud University, Saudi Arabia.....	4652
Altun, Arif / Hacettepe University, Turkey	1441
Alves, Domingos / University of São Paulo, Brazil.....	3782
Amoros, Francisco / Miguel Hernandez University, Spain.....	6894
Amroush, Fadi / Universidad de Salamanca, Spain.....	1616
Amuda, Caleb Okoro / University of Ibadan, Nigeria.....	2303
Anderson, Billie / Ferris State University, USA	7140
Andrade, Maria Teresa / University of Porto, Portugal	6031
Andreev, Rumen D. / Bulgarian Academy of Sciences, Institute of Information and Communication Technologies, Bulgaria	2188
Ankomah, Paul / North Carolina A&T State University, USA.....	4113
Anshari, Muhammad / Universiti Brunei Darussalam, Brunei	5908,7369
Anton, Jose Luis Monroy / La Ribera University Hospital, Spain.....	5787
Antonino, Ardilio / Fraunhofer Institute for Industrial Engineering, Germany	4560
Aragaki, Madoka / Business Breakthrough University, Japan	4962
Arana, Loredana / University of Phoenix, USA.....	5650
Araujo, Andre / Texas A&M University, USA.....	720
Arias, Vladimir Santiago / Texas Tech University, USA.....	7069
Arner, Tracy / Kent State University, USA	6388
Arora, Richa / Regenesys Institute of Management, India	2882
Arshad, Noreen Izza / Universiti Teknologi PETRONAS, Malaysia.....	3337
Arslan, Yunus Ziya / Istanbul Univesity, Turkey.....	298,492,7470
Artizzu, Valentino / University of Cagliari, Italy.....	3273
Arumugam, Sivasankar / Sri Venkateswara College of Education, India	2599
Askarany, Davood / University of Auckland, New Zealand	2166
Atzeni, Andrea / Politecnico di Torino, Italy.....	5004
Atzori, Barbara / University of Florence, Italy	5955
Aubin, Verónica Inés / Universidad Nacional de La Matanza, Argentina.....	4643
Averweg, Udo Richard / eThekweni Municipality, South Africa.....	7106
Ávila, Liliana / University of Aveiro, Portugal.....	888
Avola, Danilo / University of Udine, Italy.....	6195
Aymerich-Franch, Laura / CNRS-AIST JRL (Joint Robotics Laboratory), AIST, Japan	4234
Azevedo, Ana / Polytechnic Institute of Porto, Portugal.....	1907
Aziz, Abdul / Usman Institute of Technology, Pakistan.....	166
Babatope, Ihuoma Sandra / Delta State College of Physical Education, Nigeria.....	5252
Badea, Mihaela / Petroleum-Gas University of Ploiesti, Romania.....	3945
Baeva, Liudmila Vladimirovna / Astrakhan State University, Russia	4189
Bag, Surajit / Tega Industries South Africa Pty Ltd., South Africa	3086
Bailetti, Tony / Carleton University, Canada	1684
Balamurugan B / VIT University, India	1075
Balasubramanian, Kannan / Mepco Schlenk Engineering College, India	4975
Balasubramanian, Senthilarasu / National Institute of Technology, India.....	49
Baldini, Gianmarco / European Commission – Joint Research Centre, Italy.....	6136
Ballesteros, Carlos / Universidad Pontificia Comillas, Spain	1638
Banaji, Shakuntala / LSE, UK	3667
Banihashemi, Saeed / University of Technology Sydney, Australia	539

Baporikar, Neeta / <i>Namibia University of Science and Technology, Namibia & University of Pune, India</i>	2989
Barlow, Michael / <i>University of New South Wales, Australia</i>	156
Barrett, Diane / <i>Bloomsburg University of Pennsylvania, USA</i>	1356
Bartens, Yannick / <i>University of Hamburg, Germany</i>	601
Basham, Randy / <i>University of Texas at Arlington, USA</i>	1407
Basu, Rounaq / <i>Indian Institute of Technology Bombay, India</i>	7843
Bates, Paul / <i>University of Southern Queensland, Australia</i>	7869
Battistella, Gaetano / <i>ISPRA, Italy</i>	3156
Batyashe, Thami / <i>Cape Peninsula University of Technology, South Africa</i>	810
Bayındır, Levent / <i>Ataturk University, Turkey</i>	6286
Bayousuf, Abeer / <i>King Saud University, Saudi Arabia</i>	4652
Bedesem, Pena L. / <i>Kent State University, USA</i>	6388
Belalem, Ghalem / <i>University of Es Senia OranI, Algeria</i>	1116
Belikov, Olga / <i>Brigham Young University, USA</i>	6779
Bellar, Wendi R. / <i>Texas A&M University, USA</i>	6161
Ben Hamida, Ahmed / <i>University of Sfax, Tunisia</i>	106
Ben Maati, Mohammed / <i>University Abdelmalek Essaadi, Morocco</i>	1616
Ben-Abdallah, Hanène / <i>King Abdulaziz University, Saudi Arabia</i>	765
Benazeer, Shahzada / <i>University of Antwerp, Belgium</i>	5317
Bennett, Sue / <i>University of Wollongong, Australia</i>	2512
Bergenti, Federico / <i>Università degli Studi di Parma, Italy</i>	6950
Bessette, Dustin / <i>National Graduate School of Quality Management, USA</i>	7898
Betts, Lucy R. / <i>Nottingham Trent University, UK</i>	4168,4245
Bhanot, Neeraj / <i>Dr. B. R. Ambedkar National Institute of Technology Jalandhar, India</i>	587
Bharwani, Sonia / <i>Indian School of Management and Entrepreneurship, India</i>	1529
Bhattacharya, Simanti / <i>University of Kalyani, India</i>	466
Bhuimali, A. / <i>Raiganj University, India</i>	4448,7201
Bhutkar, Ganesh D. / <i>Vishwakarma Institute of Technology, India</i>	4099
Biancofiore, Giovanni / <i>giovannibiancofiore.com, Italy</i>	7671
Bielenia-Grajewska, Magdalena / <i>University of Gdansk, Poland</i>	2580,5085
Bihl, Trevor J. / <i>Air Force Institute of Technology, USA</i>	6642
Bisht, Pankaja / <i>B. T. Kumaon Institute of Technology, India</i>	6253
Bisui, Sandip / <i>Indian Institute of Technology (IIT) Kanpur, India</i>	5901
Blanco, Beatriz / <i>University of Cantabria, Spain</i>	3870
Boccasini, Adele / <i>Lega del Filo d'Oro – Termini Imerese, Italy</i>	287
Bocci, Elena / <i>Sapienza University of Rome, Italy</i>	4038,4064,4404,7014,7044
Bodea, Constanța-Nicoleta / <i>Bucharest University of Economic Studies, Romania</i>	5158,6579
Boff, Felipe / <i>Lutheran University of Brazil (ULBRA), Brazil</i>	4797
Bonaiuti, Federica / <i>ISPRA, Italy</i>	3156
Bond, Emma / <i>University Campus Suffolk, UK</i>	7312
Bose, Ahana / <i>Indian Institute of Management Calcutta, India</i>	61
Bouaziz, Rafik / <i>University of Sfax, Tunisia</i>	2043,2054
Bouck, Emily C. / <i>Michigan State University, USA</i>	266
Boudriga, Noureddine / <i>University of Carthage, Tunisia</i>	4849
Boukadi, Khouloud / <i>University of Sfax, Tunisia</i>	765

Bowers, Clint / <i>University of Central Florida, USA</i>	995,3327
Braga, Adriana / <i>Pontifical Catholic University of Rio de Janeiro, Brazil</i>	1186
Brahmia, Zouhaier / <i>University of Sfax, Tunisia</i>	2043,2054
Bratsas, Charalampos P. / <i>Aristotle University of Thessaloniki, Greece</i>	1196
Brewer, James C. / <i>Texas Tech University, USA</i>	1499
Bringula, Rex Perez / <i>University of the East, Philippines</i>	2466
Bruno, Clementina / <i>University of Piemonte Orientale, Italy</i>	3144
Bruno, Giorgio / <i>Politecnico di Torino, Italy</i>	788
Bryan, Lisa Dotterweich / <i>Upper Iowa University, USA</i>	3612
Bryl, Lukasz / <i>Poznan University of Economics and Business, Poland</i>	5056
Bueno, José de Jesús Pérez / <i>Centro de Investigación y Desarrollo Tecnológico en Electroquímica, Mexico</i>	1277
Burton, Sharon / <i>Grand Canyon University, USA</i>	7898
Buss, Terry F. / <i>Carnegie Mellon University, Australia</i>	3526
Cabrera, Edison / <i>University of the East, Philippines</i>	2466
Cahlik, Tomas / <i>Charles University Prague, Czech Republic & University of Economics Prague, Czech Republic</i>	4920
Calabrese, Rita / <i>University of Salerno, Italy</i>	1206
Calmerin, Princess B. / <i>University of the East, Philippines</i>	2466
Campbell, Heidi A. / <i>Texas A&M University, USA</i>	6161
Cappelli, Claudia / <i>Federal University of the State of Rio de Janeiro, Brazil</i>	754
Carlos-Mancilla, Miriam A. / <i>CINVESTAV Unidad Guadalajara, Mexico</i>	6522
Carlucci, Carlo / <i>ISPRA, Italy</i>	3156
Carneiro de Sousa, Catarina / <i>Polytechnic Institute of Viseu, Portugal</i>	4146
Carnero, María Carmen / <i>University of Castilla-La Mancha, Spain & University of Lisbon, Portugal</i>	3131,3698
Caron, Franco / <i>Politecnico di Milano, Italy</i>	5679
Carrasco-Carrasco, Rocío / <i>University of Huelva, Spain</i>	3349
Carvalho, Henrique Cyrne / <i>Serviço de Cardiologia, Hospital de Santo António, Centro Hospitalar do Porto, Portugal</i>	1006
Castello, Valentina / <i>CIOFS FP, Italy</i>	1559
Castro, Ana / <i>Universidade do Porto, Portugal</i>	1006
Catacutan, Annaliza E. / <i>National University, Philippines</i>	2466
Celebi, Numan / <i>Sakarya University, Turkey</i>	5490
Cella, Francesca / <i>University of Cagliari, Italy</i>	3273
Chabukswar, Aniruddha R. / <i>MAEER's Maharashtra Institute of Pharmacy, India</i>	6767
Chakraborty, Aruna / <i>Maulana Abul Kalam Azad University of Technology, India</i>	4872,6961
Chamakiotis, Petros / <i>University of Sussex, UK</i>	1153
Chand, Trilok / <i>PEC University of Technology, India</i>	477
Chandan, Harish C. / <i>Argosy University, USA</i>	4358
Chang, Ni / <i>Indiana University – South Bend, USA</i>	2661
Chatterjee, D. / <i>Seacom Skills University, India</i>	4448,4723
Chaudhuri, Somnath / <i>Maldives National University, Maldives</i>	3403
Chedid, Marcello Fernandes / <i>University of Aveiro, Portugal</i>	3963
Chen, E. Jack / <i>BASF Corporation, USA</i>	1297,1856
Chen, Edward T. / <i>University of Massachusetts – Lowell, USA</i>	3077

Chen, Gaowei / <i>The University of Hong Kong, Hong Kong</i>	7969
Chen, Hsin-Liang (Oliver) / <i>University of Massachusetts Boston, USA</i>	7599
Chen, Jihui / <i>Illinois State University, USA</i>	2841
Chen, Shipping / <i>CSIRO Data61, Australia</i>	7563
Chen, Yangjun / <i>University of Winnipeg, Canada</i>	4502,7995
Chen, Ye-Sho / <i>Louisiana State University, USA</i>	930,2686,4822
Chen, Yibin / <i>University of Winnipeg, Canada</i>	4502
Chen, Yueguo / <i>Renmin University of China, China</i>	1947
Cheng, Po-Keng / <i>State University of New York, Stony Brook University, USA</i>	71
Cheung, Benny C. F. / <i>The Hong Kong Polytechnic University, Hong Kong</i>	4774
Cheung, H. H. / <i>The University of Hong Kong, Hong Kong</i>	505
Chickerur, Satyadhyam / <i>KLE Technological University, India</i>	847
Chiluwa, Innocent / <i>Covenant University OTA, Nigeria</i>	6275
Chiu, Ming M / <i>The Education University of Hong Kong, Hong Kong</i>	7969
Cho, Kyong James / <i>Texas A&M University, USA</i>	6161
Choi, S. H. / <i>The University of Hong Kong, Hong Kong</i>	505
Chong, Alberto / <i>Georgia State University, USA & Universidad del Pacifico, Peru</i>	7214
Chuang, Huey-Wen / <i>National Taichung University of Education, Taiwan</i>	3434
Chunpir, Hashim / <i>German Climate Computing Centre (DKRZ), Germany</i>	601
Cinque, Luigi / <i>Sapienza University, Italy</i>	6195
Ciufudean, Calin / <i>Stefan cel Mare University, Romania</i>	1814
Cockrell, Susan / <i>Austin Peay State University, USA</i>	5476,5516
Cohen, Yuval / <i>Afeka Tel Aviv College of Engineering, Israel</i>	1627
Collins, Geri / <i>Mercer University, USA</i>	1432
Contreras, Eduardo C. / <i>Autonomous University of Coahuila, Mexico</i>	2431
Contreras, Isis I. / <i>Saltillo Institute of Technology, Mexico</i>	2431
Copeland Jr., Robert L. / <i>Howard University, USA</i>	1794
Cordero-Gutiérrez, Rebeca / <i>University of Salamanca, Spain</i>	3359
Cornelius, Herbert / <i>Intel Corporation EMEA, Germany</i>	4004
Correia, Eduardo / <i>Christchurch Polytechnic Institute of Technology (CPIT), New Zealand</i>	1026
Corrin, Linda / <i>University of Melbourne, Australia</i>	2512
Cosby, Missy D. / <i>Michigan State University, USA</i>	266
Cotza, Alessandro / <i>University of Cagliari, Italy</i>	3273
Coulson, Neil S. / <i>University of Nottingham, UK</i>	3767
Cox, Sharon A. / <i>Birmingham City University, UK</i>	694,1384
Craigen, Dan / <i>Carleton University, Canada</i>	1684
Crawford, Caroline M. / <i>University of Houston – Clear Lake, USA</i>	1474,5149,7922
Crompton, Helen / <i>Old Dominion University, USA</i>	6347
Crosby, Phil / <i>Curtin University, Australia</i>	5690
Crossan, Adam / <i>Letterkenny Institute of Technology, Ireland</i>	6216
Cruz-Jimenez, Miriam Guadalupe / <i>Institute INAOE, Mexico</i>	6007
Cuff, Zianne / <i>City University of New York, USA</i>	7880
Cunningham, Tamrah D. / <i>New York University, USA</i>	7880
Curcio, Davide / <i>University of Cagliari, Italy</i>	3273
Curcio, Giuseppe / <i>University of L'Aquila, Italy</i>	3296,6124
Curran, Kevin / <i>Ulster University, UK</i>	6216

Czajkowski, Marcin / <i>Bialystok University of Technology, Poland</i>	2132
D'Abundo, Michelle Lee / <i>Seton Hall University, USA</i>	5810,5820
DaCosta, Boaventura / <i>Solers Research Group, USA</i>	6361
D'Agostino, Jerome V. / <i>The Ohio State University, USA</i>	5183
Daidj, Nabya / <i>Telecom Ecole de Management, France</i>	2345
Dalcher, Darren / <i>University of Hertfordshire, UK</i>	5660
Damron, Terry Stringer / <i>Austin Peay State University, USA</i>	5476,5516
Dantsuji, Masatake / <i>Kyoto University, Japan</i>	3850
Darmont, Jérôme / <i>Université de Lyon, Lyon 2, ERIC EA3083, France</i>	1772
Das, Amit / <i>University of Kalyani, India</i>	466
Dascalu, Maria-Iuliana / <i>University Politehnica of Bucharest, Romania</i>	5158
Dasso, Aristides / <i>Universidad Nacional de San Luis, Argentina</i>	1919,7609
Dattakumar, Ambica / <i>Nanyang Technological University, Singapore</i>	2280
David-West, Olayinka / <i>Pan-Atlantic University, Nigeria</i>	2724
Dawson, Maurice / <i>University of Missouri – St. Louis, USA</i>	7898
De Bruyn, Peter / <i>University of Antwerp, Belgium</i>	5317
de Matta, Renato / <i>University of Iowa, USA</i>	5503
de Mendoza, Cecilia / <i>Ministry of Production of Argentina, Argentina</i>	7214
de Oliveira, José Palazzo M. / <i>Federal University of Rio Grande do Sul (UFRGS), Brazil</i>	6424
DeCoito, Isha / <i>Western University, Canada</i>	1420
Delgado, José Carlos Martins / <i>Universidade de Lisboa, Portugal</i>	6566
Delgado-Pérez, Pedro / <i>University of Cádiz, Spain</i>	7459
DeLong, Ronald L. / <i>University of Dayton, USA</i>	1366
Deng, Shuyuan / <i>Dakota State University, USA</i>	3861
Dessi, Nicoletta / <i>Università degli Studi di Cagliari, Italy</i>	455
DeWitty Jr., Robert L. / <i>Providence Hospital, USA</i>	1794
Di Marco, Giuseppe / <i>ISPRA, Italy</i>	3156
Diana, Barbara / <i>Università di Milano-Bicocca, Italy</i>	6223
Dias de Castro, Teresa Sofia Pereira / <i>University of Minho, Portugal</i>	7312
Dickson-Deane, Camille / <i>University of Melbourne, Australia</i>	7599
Diković, Ljubica / <i>Business Technical College, Serbia</i>	3689
Dillen, Nicole Belinda / <i>St. Thomas' College of Engineering and Technology, India</i>	6961
Dimitrova, Dimitrina / <i>York University, Canada</i>	7057
Dimoulas, Charalampos A. / <i>Aristotle University of Thessaloniki, Greece</i>	2908,6476
Dinger, Michael / <i>University of South Carolina Upstate, USA</i>	8036
Dobrilovic, Dalibor / <i>University of Novi Sad, Serbia</i>	7514
Đogatović, Vesna Radonjić / <i>University of Belgrade, Serbia</i>	6546
Dolecek, Gordana Jovanovic / <i>Institute INAOE Puebla, Mexico</i>	4746,6007,6043,6171,6234
Domínguez-Jiménez, Juan José / <i>University of Cádiz, Spain</i>	7459
Doorn, Jorge Horacio / <i>Universidad Nacional del Oeste, Argentina & Universidad Nacional de La Matanza, Argentina</i>	4643,5127,7411,7422
Dos Santos, Leandro Rocha / <i>In3 Inteligência de Mercado, Brazil</i>	961
Dounias, Georgios / <i>University of the Aegean, Greece</i>	180
Dowell, Margaret-Mary Sulentic / <i>Louisiana State University, USA</i>	2326
Doyle, D. John / <i>Cleveland Clinic, USA</i>	5829
Drivet, Alessio / <i>GeoGebra Institute of Torino, Italy</i>	4629

Dryjanska, Laura / <i>Sapienza University of Rome, Italy</i>	4038,4064,4404,7044
Dunham, Nicola / <i>Massey University, New Zealand</i>	3839
Duong, Linh Nguyen Khanh / <i>Auckland University of Technology, New Zealand</i>	5335
Duret, Christophe / <i>Université de Sherbrooke, Canada</i>	4296
Durkin, Keith F. / <i>Ohio Northern University, USA</i>	1366
Duzgun, Sebnem / <i>Middle East Technical University, Turkey</i>	3503
Edgeman, Rick / <i>Utah State University, USA</i>	729
Edwards, John Steven / <i>Aston University, UK</i>	5046
Egedigwe, Eges / <i>Dallas County Community College, USA</i>	1129
Eichler, Marcelo Leandro / <i>Universidade Federal do Rio Grande do Sul, Brazil</i>	6376
Endo, Hiroko / <i>Rissho University, Japan</i>	7946
Erdt, Marius / <i>Fraunhofer IDM@NTU, Singapore</i>	381
Erener, Arzu / <i>Kocaeli University, Turkey</i>	3503,7831
Erickson, G. Scott / <i>Ithaca College, USA</i>	943
Eryilmaz, Mehmet Eymen / <i>Uludağ University, Turkey</i>	2998
Esfahanipour, Akbar / <i>Amirkabir University of Technology, Iran</i>	4570
Esparza, Juan Carlos Moctezuma / <i>Universidad Politécnica Metropolitana de Hidalgo, Mexico</i>	1277
Espy, Deborah D. / <i>Cleveland State University, USA</i>	5941
Essila, Jean C. / <i>Northern Michigan University, USA</i>	5345
Eteokleous, Nikleia / <i>Frederick University Cyprus, Cyprus</i>	2492,6859
Eustáquio, Luís / <i>Universidade do Porto, Portugal</i>	4146
Fadzil, Fadzlina Mohd / <i>Universiti Sains Malaysia, Malaysia</i>	5378
Fahy, Patrick J. / <i>Athabasca University, Canada</i>	5829
Faily, Shamal / <i>Bournemouth University, UK</i>	5004
Fakhfakh, Sonda Bouattour / <i>University of Tunis El-Manar, Tunisia</i>	7190
Fan, Qiuyan / <i>Western Sydney University, Australia</i>	3602
Fancott, Terrill / <i>Concordia University, Canada</i>	7577
Fang, Edna Ho Chu / <i>University of Malaya, Malaysia</i>	4382
Farahneh, Hasan / <i>Ryerson University, Canada</i>	6672
Farinosi, Manuela / <i>University of Udine, Italy</i>	2064
Farmer, Lesley S. J. / <i>California State University – Long Beach, USA</i>	2477
Favre, Liliana Maria / <i>Universidad Nacional Del Centro De La Provincia De Buenos Aires, Argentina</i>	7447
Fazlollahtabar, Hamed / <i>Sharif University of Technology, Iran & National Elites Foundation, Iran</i>	6825
Feng, Haoxian / <i>University of Ottawa, Canada</i>	1757
Feraco, Antonio / <i>Fraunhofer IDM@NTU, Singapore</i>	381
Fernando, Xavier / <i>Ryerson University, Canada</i>	6672
Fernando, Yudi / <i>Universiti Malaysia Pahang, Malaysia</i>	2802,5306,5327,5357,5378,5422,5446,5456,5465,5527,5550
Ferrari, Alberto / <i>University of Parma, Italy</i>	2392
Ferrer, Gemma García / <i>Rey Juan Carlos University, Spain</i>	5767
Ferrero, Renato / <i>Politecnico di Torino, Italy</i>	3989
Fessi, Boutheina / <i>University of Carthage, Tunisia</i>	4849
Figg, Candace / <i>Brock University, Canada</i>	3238

Figuerola, Paola / <i>Cinvestav-IPN, Mexico</i>	3794
Filténborg, Jacob / <i>Aarhus University, Denmark</i>	2920
Fino, Emanuele / <i>Psychologist, Psychometrician, Italy</i>	7014
Firdhous, Mohamed Fazil Mohamed / <i>University of Moratuwa, Sri Lanka</i>	1174,6556
Fisher, Joel / <i>Department of State, United States Government, USA</i>	1017
Fisher, Robert Leslie / <i>Independent Researcher, USA</i>	4136
Fister Jr., Iztok / <i>University of Maribor, Slovenia</i>	7348
Fister, Iztok / <i>University of Maribor, Slovenia</i>	7348
Fitch-Hauser, Margaret / <i>Auburn University, USA</i>	6985
Flanagan, Patrick / <i>St. John's University, USA</i>	4619
Flanagan, Sara M. / <i>University of Kentucky, USA</i>	266
Flor, Alexander G. / <i>University of the Philippines, Philippines</i>	5077
Fokides, Emmanuel / <i>University of the Aegean, Greece</i>	2616
Foresti, Gian Luca / <i>University of Udine, Italy</i>	2064,6195
Forge, John / <i>Independent Researcher, Australia</i>	3205
Forrest, Edward / <i>University of Alaska Anchorage, USA</i>	5748
Fowler, Angela / <i>Erikson Institute, USA</i>	7623
Fox, William P. / <i>Naval Postgraduate School, USA</i>	4594
Fragkogios, Antonios / <i>University of Thessaly, Greece</i>	5411
Franklin, Benjamin / <i>Northumbria University, UK</i>	1605
Freiman, Viktor / <i>Université de Moncton, Canada</i>	2314,7248
Fridman, Alexander Yakovlevich / <i>Institute for Informatics and Mathematical Modelling, Kola Science Centre of RAS, Russia</i>	1995
Frydman, Antonia / <i>The University of Texas at Austin, USA</i>	6685
Fuchs, Matthias / <i>Mid-Sweden University, Sweden</i>	349
Fuji, Kei / <i>University of Tsukuba, Japan</i>	7946
Funes, Ana / <i>Universidad Nacional de San Luis, Argentina</i>	1919,7609
Furlong, Shauneen / <i>University of Ottawa, Canada & John Moores Liverpool University, UK</i>	3621
Galés, Maria Nieves Lorenzo / <i>The Transformation Society, Spain</i>	6488
Gallo, Crescenzo / <i>University of Foggia, Italy</i>	440
Gallon, Ray / <i>The Transformation Society, France</i>	6488
Galloni, Ruggero / <i>Square Reply S.r.l., Italy</i>	5004
Ganaie, Shabir Ahmad / <i>University of Kashmir, India</i>	2264,4515
Gandino, Filippo / <i>Politecnico di Torino, Italy</i>	3989
Ganguly, Shromona / <i>Indian Institute of Management Calcutta, India & Reserve Bank of India, India</i>	6916,7234
Gani, Anisha Banu Dawood / <i>Universiti Sains Malaysia, Malaysia</i>	5306
García de Leaniz, Patricia Martínez / <i>University of Cantabria, Spain</i>	3183
García, Diego Marroquín / <i>Universidad Tecnológica de San Juan del Río, Mexico</i>	1277
García, Felix Buendia / <i>Polytechnical University of Valencia, Spain</i>	5787
García-Morales, Víctor / <i>University of Granada, Spain</i>	5775
García-Vaquero, Martín / <i>Nebrija University, Spain</i>	3020
Garg, Nishu / <i>VIT University, India</i>	4528
Garland, Virginia E. / <i>University of New Hampshire, USA</i>	2503
Garrido-Moreno, Aurora / <i>University of Malaga, Spain</i>	5775

Gaskins, Melvin / <i>Howard University Hospital, USA</i>	1794
Gasmelseed, Akram / <i>Qassim University, Saudi Arabia</i>	1249,3448
Gašpar, Dražena / <i>University of Mostar, Bosnia and Herzegovina</i>	2521
Gasparini, Isabela / <i>Santa Catarina State University (UDESC), Brazil</i>	6424
Gavurová, Beáta / <i>Technical University of Košice, Slovakia</i>	5841
Geetha, Mariappan / <i>Kamaraj College of Engineering and Technology, India</i>	4369
Ghasem, Nayef Mohamed / <i>UAE University, UAE</i>	3043
Giaquinto-Cilliers, Maria / <i>Kimberley Hospital Complex, South Africa</i>	6147
Gibbs, William J. / <i>Duquesne University, USA</i>	4210
Gierl, Mark / <i>University of Alberta, Canada</i>	2369
Giomelakis, Dimitrios / <i>Aristotle University of Thessaloniki, Greece</i>	8046
Giuliani, Raimondo / <i>European Commission – Joint Research Centre, Italy</i>	6136
Giuseffi, Frank G. / <i>Lindenwood University, USA</i>	2571
Gokhberg, Leonid / <i>National Research University Higher School of Economics, Russia</i>	4704
Goldkind, Lauri / <i>Fordham University, USA</i>	3569
Gomes, Jorge / <i>ISEG, Universidade de Lisboa, Portugal</i>	3756,5714
Gómez, Andrés / <i>University of Castilla-La Mancha, Spain</i>	3698
Gonsalves, Tad / <i>Sophia University, Japan</i>	144,229
González-Marcos, Ana / <i>Universidad de La Rioja, Spain</i>	1934
Gosney, Matthew W. / <i>University of Colorado – Health, USA</i>	4326
Goudos, Sotirios K. / <i>Aristotle University of Thessaloniki, Greece</i>	5967,6595
Goulette, Elizabeth / <i>Georgia State University, USA</i>	7682
Govindarajan, M. / <i>Annamalai University, India</i>	373
Graham, Charles R. / <i>Brigham Young University, USA</i>	1487
Grahn, Kaj J. / <i>Arcada University of Applied Sciences, Finland</i>	7715
Grandi, Fabio / <i>University of Bologna, Italy</i>	2043,2054
Granell-Canut, Carlos / <i>Universitat Jaume I of Castellón, Spain</i>	3473
Gretter, Sarah / <i>Michigan State University, USA</i>	2292
Grooms, Linda D. / <i>Regent University, USA</i>	2455,2588
Grotto, Rosapia Lauro / <i>University of Florence, Italy</i>	5955
Grover, Varun / <i>Clemson University, USA</i>	8036
Grunwald, Armin / <i>Karlsruhe Institute of Technology, Germany</i>	4267
Guan, Chong / <i>SIM University, Singapore</i>	2280
Guedes, Rui / <i>Faculdade de Medicina da Universidade do Porto, Portugal</i>	1006
Guerreiro, Sérgio Luís / <i>University of Lisbon, Portugal</i>	651,2154
Guglielman, Eleonora / <i>Learning Community, Italy</i>	1559
Guia, Angelita D. / <i>University of the East, Philippines</i>	2466
Gumussoy, Bilal / <i>Shell and Turcas Petrol Inc., Turkey</i>	530
Gumussoy, Cigdem Altin / <i>Istanbul Technical University, Turkey</i>	530
Gupta, Yash / <i>Maulana Abul Kalam Azad University of Technology, India</i>	4872
Gursoy, Guner / <i>Okan University, Turkey</i>	5285
Guspini, Marco / <i>educommunity – Educational Community, Italy</i>	1559
Gutiérrez-Artacho, Juncal / <i>University of Granada, Spain</i>	4471
Güzeldereli, Esra Ayça / <i>Afyon Kocatepe University, Turkey</i>	6789
Gwangwawa, Norman / <i>Botswana International University of Science and Technology, Botswana</i>	2626

Hadad, Graciela Dora Susana / <i>Universidad Nacional del Oeste, Argentina</i>	5127,7422
Haleem, P. P. Abdul / <i>Farook College, India</i>	7745
Halili, Munira / <i>Universiti Sains Malaysia, Malaysia</i>	5527
Hamam, Habib / <i>University of Moncton, Canada</i>	106
Hardin, J. Michael / <i>Samford University, USA</i>	7140
Hari, Seetha / <i>Vellore Institute of Technology, India</i>	1825
Hartsock, Ralph / <i>University of North Texas, USA</i>	6748
Hasan, Osman / <i>National University of Sciences and Technology, Pakistan</i>	3103,6872,6882
Hasan, Syed Rafay / <i>Tennessee Technological University, USA</i>	3103
Hassan, Azizul / <i>Cardiff Metropolitan University, UK</i>	4031
Hassan, Md. Salleh Hj. / <i>Universiti Putra Malaysia, Malaysia</i>	1704,2761
Hawkins, Donna Patterson / <i>University of Phoenix, USA</i>	1463,3805
Heavin, Ciara / <i>University College Cork, Ireland</i>	2558,5864
Hemmatzad, Nolan / <i>University of Nebraska at Omaha, USA</i>	7796
Hemmatpour, Masoud / <i>Politecnico di Torino, Italy</i>	3989
Henckell, Martha / <i>Southeast Missouri State University, USA</i>	1451
Hendayani, Ratih / <i>Telkom University, Indonesia</i>	5456
Hernando-Gómez, Ángel / <i>University of Huelva, Spain</i>	3976
Herpich, Fabrício / <i>Federal University of Rio Grande do Sul, Brazil</i>	7935
Herrera-Pérez, Gabriel / <i>Instituto Tecnológico Superior de Irapuato, Mexico</i>	2897
Herschel, Richard T. / <i>Saint Joseph's University, USA</i>	951
Highland, Sue Ann / <i>Grand Canyon University, USA</i>	5183
Hight, Craig / <i>University of Newcastle, Australia</i>	7539
Hirzalla, Fadi / <i>Erasmus University Rotterdam, The Netherlands</i>	3667
Hlahla, Catherine / <i>National University of Science and Technology, Zimbabwe</i>	2626
Hoanca, Bogdan / <i>University of Alaska Anchorage, USA</i>	5748
Hoffman, Hunter G. / <i>University of Washington, USA</i>	5955
Holland, Barbara Jane / <i>Brooklyn Public Library, USA</i>	899
Holmberg, Christopher / <i>University of Gothenburg, Sweden</i>	6940
Hope, John K. / <i>University of Auckland, New Zealand</i>	2421
Höpken, Wolfram / <i>University of Applied Sciences Ravensburg-Weingarten, Germany</i>	349
Horne, Jeremy / <i>The International Institute of Informatics and Systemics, USA</i>	1845
Hosseini, M. Reza / <i>Deakin University, Australia</i>	539
Hough, Michelle / <i>Pennsylvania State University, USA</i>	7026
Hsieh, Wan-Chiang / <i>National Taichung Girls' Senior High School, Taiwan</i>	3434
Hsu, Jeffrey / <i>Fairleigh Dickinson University, USA</i>	5388
Hsueh, Ya-Hui / <i>National Taichung University of Education, Taiwan</i>	3434
Hu, Nan / <i>University of Utah, USA</i>	2213
Hu, Wen-Chen / <i>University of North Dakota, USA</i>	6057
Hude, Rahul U. / <i>MAEER's Maharashtra Institute of Pharmacy, India</i>	6767
Hughes, Claretha / <i>University of Arkansas, USA</i>	4326
Hui, Hoo Yee / <i>Universiti Tunku Abdul Rahman, Malaysia</i>	2802
Hussain, Shabir / <i>University of Kashmir, India</i>	5262
Huysmans, Philip / <i>University of Antwerp, Belgium</i>	5317
Ibrahim, Adamkolo Mohammed Mohammed / <i>Universiti Putra Malaysia, Malaysia & University of Maiduguri, Nigeria</i>	1704,2761

Iengo, Giorgio Amedeo / <i>University of Cagliari, Italy</i>	3273
Igbinomwanhia, Osaro Rawlings / <i>University of Benin, Nigeria</i>	5036
Iguisi, Osarumwense / <i>University of Benin, Nigeria</i>	5036
Ikahihifo, Tarah B. / <i>Brigham Young University, USA</i>	1487
Ikolo, Violet E. / <i>Delta State University Library, Nigeria</i>	5726
Ikuta, Shigeru / <i>Otsuna Women's University, Japan</i>	6464
Ilavarasan, P. Vigneswara / <i>Indian Institute of Technology Delhi, India</i>	7126
Iqbal, Badar Alam / <i>Aligarh Muslim University, India</i>	6727
Iqbal, Mehree / <i>North South University, Bangladesh</i>	3579
Ishikawa, Yasushige / <i>Kyoto University of Foreign Studies, Japan</i>	3850
Iurato, Giuseppe / <i>University of Palermo, Italy</i>	7704
Ivashov, Arslan / <i>Kazakh Ablai Khan University of International Relations and World Languages, USA</i>	5626
Iyamu, Tiko / <i>Cape Peninsula University of Technology, South Africa</i>	810,2943,4480
Jacques, Laura / <i>Medical College of Wisconsin, USA</i>	5800
Jafari, Mostafa / <i>Regional Institute of Forest and Rangelands (RIFR), Iran</i>	3114,7815
Jagannathan, Veeraraghavan / <i>National Institute of Technology, India</i>	49
Jagdale, Swati C. / <i>MAEER's Maharashtra Institute of Pharmacy, India</i>	6767
Jaipal-Jamani, Kamini / <i>Brock University, Canada</i>	3238
Jakobs, Kai / <i>RWTH Aachen University, Germany</i>	4679
Jamil, Cecília C. / <i>Stockholm University, Sweden</i>	961
Jamil, George Leal / <i>Informações em Rede, Brazil</i>	961
Jamis, Marilou N. / <i>National University, Philippines</i>	2466
Jan, Rosy / <i>University of Kashmir, India</i>	7912
Jana, Arnab / <i>Indian Institute of Technology Bombay, India</i>	7843
Janakova, Milena / <i>Silesian University in Opava, Czech Republic</i>	6907
Janczyk-Strzała, Elżbieta / <i>Wroclaw School of Banking, Poland</i>	3910
Janeš, Aleksander / <i>University of Primorska, Slovenia</i>	5638
Janssens, Jurgen / <i>QSpin, Belgium</i>	682
Jasmi, Fairuz / <i>Universiti Sains Malaysia, Malaysia</i>	5465
Jaume, Sylvain / <i>Saint Peter's University, USA</i>	5388
Jayalakshmi, S. / <i>Kumaraguru College of Technology (KCT), India</i>	431
Jesus, Rui Alberto / <i>CESPU, Instituto de Investigação e Formação Avançada em Ciências e Tecnologias da Saúde, Portugal</i>	1548
Jhamb, Garvita / <i>University of Delhi, India</i>	4436
Jiang, Bo / <i>Zhejiang Gongshang University, China</i>	7588
Joe, Jennifer Ashley Wright / <i>Western Kentucky University, USA</i>	5204
Jones, James D. / <i>Liberty University, USA</i>	4736
Jones-Lillie, Michelle / <i>Lillie Pad Studios, USA</i>	4692
Josefowicz, Michael / <i>Nemetics Institute Kolkata, USA</i>	6488
Julien, Heidi / <i>State University of New York at Buffalo, USA</i>	2243
K, John Singh / <i>VIT University, India</i>	1667
Kaabouch, Naima / <i>University of North Dakota, USA</i>	6057
Kabeil, Magdy M. / <i>Al-Yamamah University, Saudi Arabia</i>	2006
Kabil, Ahmad M. / <i>University of Wisconsin – Whitewater, USA</i>	2006
Kajan, Ejub / <i>State University of Novi Pazar, Serbia</i>	2773

Kalia, Prateek / <i>I. K. Gujral Punjab Technical University, India</i>	2882
Kalita, Manashee / <i>NERIST, India</i>	4985
Kaljo, Kristina / <i>Medical College of Wisconsin, USA</i>	5800
Kalliris, George / <i>Aristotle University of Thessaloniki, Greece</i>	6476
Kalpić, Damir / <i>University of Zagreb, Croatia</i>	620,636,5607
Kamakula, Madhu Kishore Raghunath Raghunath / <i>GVP College of Engineering (Autonomous), India</i>	3009
Kamariotou, Maria / <i>University of Macedonia, Greece</i>	912,3032
Kamat, R. K. / <i>Shivaji University, India</i>	2411
Kamath, R. S. / <i>Chatrapati Shahu Institute of Business Education and Research, India</i>	2411
Kamel, Sherif H. / <i>The American University in Cairo, Egypt</i>	7223,7259
Kamthan, Pankaj / <i>Concordia University, Canada</i>	7399,7577
Kanaan, Yasmine M. / <i>Howard University, USA</i>	1794
Kanellopoulos, Dimitris N. / <i>University of Patras, Greece</i>	6435
Kang, Sunghyun Ryoo / <i>Iowa State University, USA</i>	8079
Karabulut, Ali Naci / <i>Mugla Sitki Kocman University, Turkey</i>	5737
Karabulut, Derya / <i>Halic University, Turkey</i>	492
Karagoz, Pinar / <i>Middle East Technical University, Turkey</i>	1974
Karahan, Çetin / <i>Directorate General of Civil Aviation, Turkey</i>	6789
Karahoca, Adem / <i>Bahcesehir University, Turkey</i>	7168
Karasavvidis, Ilias / <i>University of Thessaly, Greece</i>	3287
Karpinski, Aryn C. / <i>Kent State University, USA</i>	5183
Kasemsap, Kijpokin / <i>Suan Sunandha Rajabhat University, Thailand</i>	1584,1594,2253,2791,3591,3734,4199,4347,6412,7775
Katsaounidou, Anastasia N. / <i>Aristotle University of Thessaloniki, Greece</i>	2908
Kaur, Sarabjot / <i>IIT Kanpur, India</i>	7036,7161
Kavoura, Androniki / <i>Technological Educational Institute of Athens, Greece</i>	3827,4052,7002,7270
Kavurucu, Yusuf / <i>Turkish Naval Research Center Command, Turkey</i>	1974
Kawata, Shigeo / <i>Utsunomiya University, Japan</i>	4583
Kaya, Galip / <i>Havelsan Inc., Turkey</i>	1441
Kaye, Linda K. / <i>Edge Hill University, UK</i>	3317
Kayikci, Yasanur / <i>Turkish-German University, Turkey</i>	5367
Kefallonitis, Efstathios / <i>State University of New York at Oswego, USA</i>	3827,4052
Kerkhoff, Shea N. / <i>North Carolina State University, USA</i>	2235
Khalifeh, Ala' Fathi / <i>German Jordan University, Jordan</i>	6672
Khalil, Sadia / <i>NUST School of Electrical Engineering and Computer Science, Pakistan</i>	2869
Khan, Amadu Wurie / <i>University of Edinburgh, UK</i>	6695
Khan, Nadim Akhtar / <i>University of Kashmir, India</i>	4436
Khoo, Elaine / <i>University of Waikato, New Zealand</i>	7539
Kibret, Behailu / <i>Victoria University, Australia</i>	6319
Kikumori, Mai / <i>Ritsumeikan University, Japan</i>	6019
Kilani, Amal / <i>University of Gabes Tunisia, Tunisia</i>	106
Kilburn, Michelle / <i>Southeast Missouri State University, USA</i>	1451
Kilinç, Uğur / <i>Ondokuz Mayıs University, Turkey</i>	7338
Kille, Tarryn / <i>Griffith University, Australia</i>	7869

Kimmons, Royce / Brigham Young University, USA	6779
King, Samuel Olugbenga / Auburn University, USA	1880
Kinsell, Carolyn / Solers Research Group, USA	6361
Kiriakidis, Stavros / University of Crete, Greece	3827
Kitamura, Misato / Advanced Telecommunications Research Institute International (ATR), Japan	3850
Kitsios, Fotis / University of Macedonia, Greece	912,3032
Klock, Ana Carolina Tomé / Federal University of Rio Grande do Sul (UFRGS), Brazil	6424
Knepper, Richard / Indiana University, USA	1063
Koleva, Maria K. / Institute of Catalysis, Bulgarian Academy of Sciences, Bulgaria	220
Kolker, Alexander / GE Healthcare, USA	2095,3711
Kondo, Mutsumi / Kyoto University of Foreign Studies, Japan	3850
Konzack, Lars / University of Copenhagen, Denmark	8015
Kopeć, Katarzyna / Tischner European University, Poland	563
Kopec, Rafal / Pedagogical University of Krakow, Poland	7302
Köroğlu, Cemile Zehra / Uşak University, Turkey	4715
Köroğlu, Muhammet Ali / Uşak University, Turkey	4715
Korstanje, Maximiliano Emanuel / University of Palermo, Argentina	3637,4761
Kostic-Ljubisavljevic, Aleksandra / University of Belgrade, Serbia	1164
Koteeswaran, Raghuraman / SASTRA University, India	77
Kováč, Viliam / Technical University of Košice, Slovakia	5841
Krasniewicz, Jan / Birmingham City University, UK	1384
Kretowski, Marek / Bialystok University of Technology, Poland	2132
Krishnamoorthy, Priya / SASTRA University, India	77
Krogstie, John / Norwegian University of Science and Technology, Norway	4810
Kulik, Boris Alexandrovich / Institute of Problems in Mechanical Engineering RAS, Russia	1995
Kumar, Sameer / University of Malaya, Malaysia	2784,4382,6207,6993
Kumar, Sathish / VIT University, India	1075
Kumaradja, Richard / Renaissance Consulting, Indonesia	862
Kundur, Avinash Reddy / Griffith University, Australia	5853
Kurata, Noriko / Tokyo University of Science, Suwa, Japan	7327
Kurnaz, Sefer / Istanbul Esenyurt University, Turkey	7470
Kurt, Ganite / Gazi University, Turkey	5737
Kuznik, Lea / University of Maribor, Slovenia	4077
Kyriafinis, Georgios / AHEPA University Hospital, Greece	5886
Lagaris, Isaac E. / University of Ioannina, Greece	7765
Lahuerta-Otero, Eva / University of Salamanca, Spain	3359
Lai, Daniel T. H. / Victoria University, Australia	6319
Lai, Hollis / University of Alberta, Canada	2369
Lakkaraju, Santhosh Kumar / Dakota State University, USA	3861
Lakshika, Erandi / University of New South Wales, Australia	156
Lansiquot, Reneta D. / City University of New York, USA	7880
Lao, Eduardo A. / University of the East, Philippines	2466
Larson, Trent / North Carolina A&T University, USA	4113
Lau, Wilfred W. F. / The Chinese University of Hong Kong, China	3309
Law, Penny / Regenesys Business School, South Africa	2882

Lawless, Kimberly A. / <i>University of Illinois at Chicago, USA</i>	7888
Lay, Ah-Nam / <i>Institute of Teacher Education – Sultan Abdul Halim, Malaysia</i>	3248
Lazaro, Joan P. / <i>University of the East, Philippines</i>	2466
Leaning, Marcus / <i>University of Winchester, UK</i>	7106
Lecic, Dusanka Milorad / <i>Levi9 Global Sourcing Balkan, Serbia</i>	2932
Ledesma, Viviana Alejandra / <i>Universidad Nacional del Oeste, Argentina & Universidad Nacional de La Matanza. Argentina</i>	7422
Lee, D. Israel / <i>Southern Illinois University, USA & University of Phoenix, USA</i>	1463
Lee, Wanbil William / <i>The Computer Ethics Society, Hong Kong & Wanbil & Associates, Hong Kong</i>	4884,4909
Lekule, Chrispina S. / <i>St. Augustine University of Tanzania, Tanzania</i>	2314
Lenzi, Valentina Bartalesi / <i>ISTI-CNR, Italy</i>	5067
Leonard, Awie C. / <i>University of Pretoria, South Africa</i>	777
Leone, Sabrina / <i>Università Politecnica delle Marche, Italy</i>	1517,2545,7671
Leung, Carson K. / <i>University of Manitoba, Canada</i>	338
Lewis, John Kennedy / <i>Salve Regina University, USA</i>	5194
Lexhagen, Maria / <i>Mid-Sweden University, Sweden</i>	349
Li, Baobin / <i>University of Chinese Academy of Sciences, China</i>	132
Li, Rowena / <i>Bayside High School Library, USA</i>	4490,7956
Li, Shun / <i>University of Chinese Academy of Sciences, China</i>	132
Lillie, Jonathan / <i>Loyola University Maryland, USA</i>	4692,6296
Link, Matthew R. / <i>Indiana University, USA</i>	1063
Litvak, Claudia S. / <i>Universidad Nacional de La Matanza, Argentina & Universidad Nacional del Oeste, Argentina</i>	5127
Lockett, Nigel / <i>University of Lancaster, UK</i>	5775
Locuratolo, Elvira Immacolata / <i>CNR ISTI, Italy</i>	2020,5067
Lofaro, Ronald John / <i>Embry-Riddle Aeronautical University, USA</i>	662
Logan, Robert K. / <i>University of Toronto, Canada</i>	1186
Loon, Donovan Peter Chan Wai / <i>University of Malaya, Malaysia</i>	2784,6207
Lopes, Hélder Silva / <i>University of Minho, Portugal</i>	3460
López, Maria Luisa Mendoza / <i>Tecnológico Nacional de México, Instituto Tecnológico de Querétaro, Mexico</i>	1277
Lopez-Mellado, Ernesto / <i>CINVESTAV Unidad Guadalajara, Mexico</i>	6522,7488
Lorenzi, Fabiana / <i>Lutheran University of Brazil (ULBRA), Brazil</i>	4797
Lu, Shen / <i>University of South Florida, USA</i>	1783
Lucchesi, Ivana Lima / <i>Universidade Federal do Rio Grande do Sul, Brazil</i>	6376
Lun, Roanna / <i>Cleveland State University, USA</i>	5876
M, Naveenkumar / <i>National Institute of Technology Trichy, India</i>	212
Ma, Will W. K. / <i>Hong Kong Shue Yan University, Hong Kong</i>	5093
Maamar, Zakaria / <i>Zayed University, UAE</i>	765
Maciel, Cristiano / <i>Federal University of Mato Grosso, Brazil</i>	754
Macis, Riccardo / <i>University of Cagliari, Italy</i>	3273
MacIver, Kristen / <i>Northern Michigan University, USA</i>	3678
Macko, Marek / <i>Kazimierz Wielki University, Poland</i>	521
Mahmood, Awais / <i>National University of Sciences and Technology, Pakistan</i>	3103
Mahmud, Nidhal / <i>University of Hull, UK</i>	6847

Major, Debra A. / <i>Old Dominion University, USA</i>	3382
Majumder, Swanirbhar / <i>NERIST, India</i>	4985
Malone, Lin G. / <i>Nanyang Technological University, Singapore</i>	2280
Manne, Janga Reddy / <i>Indian Institute of Technology Bombay, India</i>	239
Manrique, Cecilia G. / <i>University of Wisconsin – La Crosse, USA</i>	2861
Manrique, Gabriel G. / <i>Winona State University, USA</i>	2861
Mantovani, Fabrizia / <i>Università di Milano-Bicocca, Italy</i>	6223
Manuzzi, Raffaella / <i>ISPRA, Italy</i>	3156
Manzanares, María Ángeles Rodríguez / <i>Memorial University of Newfoundland, Canada</i>	3900
Maphanyane, Joyce Gosata / <i>University of Botswana, Botswana</i>	3484
Maraver-López, Pablo / <i>University of Huelva, Spain</i>	3976
Maravilhas, Sérgio / <i>Universidade Salvador, Brazil</i>	5757
Marichelvam, Mariappan Kadarkarainadar / <i>Mepco Schlenk Engineering College, India</i>	4369
Markkula, Jouni / <i>University of Oulu, Finland</i>	2108,4862
Marks, Madeline R. / <i>University of Central Florida, USA</i>	995
Marmo, Roberto / <i>University of Pavia, Italy</i>	2851
Marques, Rui Pedro Figueiredo / <i>Universidade de Aveiro, Portugal</i>	820
Marras, Andrea / <i>University of Cagliari, Italy</i>	3273
Martinez, Liliana / <i>Universidad Nacional del Centro de la Provincia de Buenos Aires, Argentina</i>	7447
Martinez, Maria Isabel Asensio / <i>La Ribera University Hospital, Spain</i>	5787
Martinovic, Dragana / <i>University of Windsor, Canada</i>	2314,7248
Marzano, Gilberto / <i>Rezekne Academy of Technologies, Latvia</i>	4157,5930
Masood, Rahat / <i>NUST School of Electrical Engineering and Computer Science, Pakistan</i>	2869
Masrom, Maslin / <i>Universiti Teknologi Malaysia, Malaysia</i>	7088
Mateos, Cristian / <i>ISISTAN, UNICEN-CONICET, Argentina</i>	6658
Mathews, Lincy / <i>M. S. Ramaiah Institute of Technology, India</i>	1825
Matuga, Julia M. / <i>Bowling Green State University, USA</i>	3930
Mazur, Elizabeth / <i>Pennsylvania State University, USA</i>	7026
McDaniel, Jeremy J / <i>Principal Financial Group, USA</i>	1656
McDowell Marinchak, Christina L. / <i>University of Alaska Anchorage, USA</i>	5748
McGill, Tanya / <i>Murdoch University, Australia</i>	4124
McKee, James / <i>Independent Researcher, Australia</i>	573
McKelvey, Nigel / <i>Letterkenny Institute of Technology, Ireland</i>	6216
McKenna, Kelly / <i>Colorado State University, USA</i>	3954
McNeal, Ramona Sue / <i>University of Northern Iowa, USA</i>	3612
McNeill, Fiona / <i>Heriot-Watt University, UK</i>	2697
McNutt, John G. / <i>University of Delaware, USA</i>	3569
Medina-Bulo, Inmaculada / <i>University of Cádiz, Spain</i>	7459
Medlock Paul, Casey / <i>North Carolina State University, USA</i>	2235
Meghanathan, Natarajan / <i>Jackson State University, USA</i>	1746,6507,6536
Mehta, Prashant / <i>National Law University Jodhpur, India</i>	3066
Mekhiel, Christopher / <i>Ryerson University, Canada</i>	6672
Melendez, Thiago Troina / <i>Instituto Federal Sul-Riograndense, Brazil</i>	6376
Mellott, Jennifer A. / <i>Kent State University, USA</i>	5183
Melnik, Mikhail I. / <i>Kennesaw State University, USA</i>	2833

Melro, Ana / <i>University of Aveiro, Portugal</i>	4255
Melton, Amye M. / <i>Austin Peay State University, USA</i>	5476,5516
Mentor, Dominic / <i>Columbia University, USA</i>	6184
Meroufel, Bakhta / <i>University of Oran1, Algeria</i>	1116
Messeri, Andrea / <i>Meyer Children's Hospital of Florence, Italy</i>	5955
Mezghani, Karim / <i>Al Imam Mohammad Ibn Saud Islamic University (IMSIU), Saudi Arabia & University of Sfax, Tunisia</i>	2965
Miaoui, Yosra / <i>University of Carthage, Tunisia</i>	4849
Mihaylova, Iva / <i>University of St. Gallen, Switzerland</i>	6631
Mikavica, Branka / <i>University of Belgrade, Serbia</i>	1164
Mikołajewska, Emilia / <i>Nicolaus Copernicus University, Poland</i>	521
Mikołajewski, Dariusz / <i>Kazimierz Wielki University, Poland</i>	521
Miller, Tan / <i>Rider University, USA</i>	5503
Milosevic, Danijela / <i>University of Kragujevac, Serbia</i>	5538
Mishra, Vinod Kumar / <i>B. T. Kumaon Institute of Technology, India</i>	6253,6266
Misra, Subhas Chandra / <i>Indian Institute of Technology (IIT) Kanpur, India</i>	5901,7036,7161
Mitra, Ananda / <i>Wake Forest University, USA</i>	394
Mogos, Radu Ioan / <i>Bucharest University of Economic Studies, Romania</i>	5158
Mohamad, Mohammad Taqiuddin / <i>University of Malaya, Malaysia</i>	36
Mohamed, Latifah Naina / <i>Universiti Sains Malaysia, Malaysia</i>	5527
Mohan, Fariel / <i>University of Trinidad and Tobago, Trinidad and Tobago</i>	2532
Mohanty, Radhakrishna N. / <i>VIT University, India</i>	3226
Montazemi, Ali Reza / <i>McMaster University, Canada</i>	4570
Montrucchio, Bartolomeo / <i>Politecnico di Torino, Italy</i>	3989
Morales, Edith Obregón / <i>Centro de Investigación y Desarrollo Tecnológico en Electroquímica, Mexico</i>	1277
Moreno-Ibarra, Marco Antonio / <i>Instituto Politécnico Nacional, Mexico</i>	6973
Morganti, Luca / <i>Università di Milano-Bicocca, Italy</i>	6223
Motwani, Jaideep / <i>Grand Valley State University, USA</i>	551
Mukhopadhyay, Prabir / <i>Indian Institute of Information Technology Design and Manufacturing Jabalpur, India</i>	1260
Multisilta, Jari / <i>Tampere University of Technology, Finland</i>	2641
Mupepi, Mambo Governor / <i>Grand Valley State University, USA</i>	551
Mupepi, Sylvia C. / <i>Grand Valley State University, USA</i>	551
Murakami, Masayuki / <i>Kyoto University of Foreign Studies, Japan</i>	3850
Muralidhar, Krishnamurthy / <i>University of Oklahoma, USA</i>	5583
Muritala, Omotayo / <i>Pan-Atlantic University, Nigeria</i>	2724
Murray, David / <i>Murdoch University, Australia</i>	4124
Murray, Patrick S. / <i>University of Southern Queensland, Australia</i>	7869
Murthy, Vasudeva / <i>Creighton University, USA</i>	8095
Murungi, Catherine G. / <i>Kenyatta University, Kenya</i>	1540
Musunuri, Durgamohan / <i>Bhavan's Usha and Lakshmi Mittal Institute of Management, India</i>	1529
Muthee, Jessina Mukomunene / <i>Kenyatta University, Kenya</i>	1540
Mutlu, Alev / <i>Kocaeli University, Turkey</i>	1974
Nagaty, Khaled Ahmed / <i>The British University in Egypt, Egypt</i>	2810

Naidenova, Xenia Alexandre / <i>Research Centre of Military Medical Academy – Saint Petersburg, Russia</i>	191
Nair, Jessy / <i>PES University, India</i>	2975
Najmaei, Arash / <i>Australian Catholic University, Australia</i>	1141
Narayankar, Prashant M. / <i>KLE Technological University, India</i>	847
Natarajan, Thamaraiselvan / <i>National Institute of Technology, India</i>	49
Nath, Ravi / <i>Creighton University, USA</i>	8095
Ndong, Celine / <i>ISPRA, Italy</i>	3156
Neamțu(Idorași), Alina / <i>Bucharest University of Economic Studies, Romania</i>	6579
Nedelko, Zlatko / <i>University of Maribor, Slovenia</i>	5559
Nedra, Bahri Ammari / <i>IHEC of Carthage, USA</i>	707
Nesmachnow, Sergio / <i>Universidad de la República, Uruguay</i>	7693
Neville, Karen / <i>University College Cork, Ireland</i>	2558
Ng, Alex / <i>Federation University, Australia</i>	7563
Ng, Artie W. / <i>The Hong Kong Polytechnic University, Hong Kong</i>	4774
Ng, Peggy M. L. / <i>The Hong Kong Polytechnic University, Hong Kong</i>	4774
Ngafeeson, Madison N. / <i>Northern Michigan University, USA</i>	3678,3816
Ngan, Chun-Kit / <i>The Pennsylvania State University, USA</i>	2142
Nganji, Julius T. / <i>University of Ottawa, Canada</i>	314
Nguy, Yun-Mi / <i>Griffith University, Australia</i>	5853
Niaki, Seyed Taghi Akhavan / <i>Sharif University of Technology, Iran</i>	6825
Nicholas, Arlene J. / <i>Salve Regina University, USA</i>	740
Niemi, Hannele / <i>University of Helsinki, Finland</i>	2641
Nirmala, Nancy / <i>Christ the King Matric Higher Secondary School, India</i>	2599
Nisar, Sajid / <i>National University of Science and Technology, Pakistan</i>	6872,6882
Nisha, Nabila / <i>North South University, Bangladesh</i>	3579
Nunes, Felipe Becker / <i>Federal University of Rio Grande do Sul, Brazil</i>	7935
Nwankpa, Joseph K. / <i>Miami University, USA</i>	2953
Nyamundundu, Alice Violet / <i>Skyway University, Malawi</i>	3393
O., Mohammad Mehdi Owrang / <i>American University, USA</i>	1794
Obregón, Víctor Hugo Rodríguez / <i>Centro de Investigación y Desarrollo Tecnológico en Electroquímica, Mexico</i>	1277
Ochs, Thomas / <i>Villeroy & Boch, Germany</i>	873
O'Connor, Rory V. / <i>Dublin City University, Ireland</i>	6927
O'Connor, Yvonne / <i>University College Cork, Ireland</i>	5864
O'Donnell, Eileen / <i>Trinity College Dublin, Ireland</i>	2380
O'Donnell, Liam / <i>Dublin Institute of Technology, Ireland</i>	2380
Oghogho, Ikponmwosa / <i>Delta State University, Abraka-Oleh Campus, Nigeria</i>	6618
Okunoye, Adekunle O. / <i>Xavier University, USA</i>	3561
Oliboni, Barbara / <i>University of Verona, Italy</i>	2043,2054
Oliveira, Lúdia / <i>University of Aveiro, Portugal</i>	4255
Olvera-Lobo, María-Dolores / <i>University of Granada, Spain</i>	4471
Oncioiu, Ionica / <i>Titu Maiorescu University, Romania</i>	5669
Oni, Aderonke A. / <i>Covenant University, Nigeria</i>	3561
Ono, Akinori / <i>Keio University, Japan</i>	6019
Oravec, Jo Ann / <i>University of Wisconsin – Whitewater, USA</i>	1695,4306,4316

Ordieres-Meré, Joaquín / <i>Universidad Politécnica de Madrid, Spain</i>	1934
O’Riordan, Sheila / <i>University College Cork, Ireland</i>	2558
Ortes, Faruk / <i>Istanbul University, Turkey</i>	492
Orwat, Carsten / <i>Karlsruhe Institute of Technology, Germany</i>	4267
Osman, Kamisah / <i>The National University of Malaysia, Malaysia</i>	3248,3337
Osório, António / <i>University of Minho, Portugal</i>	7312
Otebolaku, Abayomi Moradeyo / <i>Liverpool John Moores University, UK</i>	6031
Otunla, Adekunle Olusola / <i>University of Ibadan, Nigeria</i>	2303
Owen, Hazel / <i>Ethos Consultancy NZ, New Zealand</i>	3839
P, Nikhitha / <i>VIT University, India</i>	4528
Paciga, Kathleen A. / <i>Columbia College Chicago, USA</i>	7623
Pagel, James Frederic / <i>University of Colorado School of Medicine, USA</i>	202
Pallavicini, Federica / <i>Università di Milano-Bicocca, Italy</i>	6223
Paluan, Francesca / <i>Politecnico di Torino, Italy</i>	5213
Pan, Weifeng / <i>Zhejiang Gongshang University, China</i>	7588
Pană, Laura L. / <i>Polytechnic University of Bucharest, Romania</i>	88,4459
Pantano, Eleonora / <i>Middlesex University London, UK</i>	7805
Papadopoulos, Ioannis / <i>Metropolitan College Thessaloniki, Greece</i>	2337
Park, Young / <i>Bradley University, USA</i>	1735,1869
Parker, Judith / <i>Teachers College, Columbia University, USA</i>	5626
Parsopoulos, Konstantinos / <i>University of Ioannina, Greece</i>	7765
Paschoal, Leo Natan / <i>University of Cruz Alta, Brazil</i>	7935
Passaris, Constantine E. / <i>University of New Brunswick, Canada</i>	7980
Pathak, Tanuja / <i>B. T. Kumaon Institute of Technology, India</i>	6266
Patro, Chandra Sekhar / <i>GVP College of Engineering (Autonomous), India</i>	3009,4337
Pau, Iván / <i>Universidad Politécnica de Madrid, Spain</i>	7754
Paul, P. K. / <i>Raiganj University, India</i>	4448,4723,7201
Pavlidou, Fotini-Niovi / <i>Aristotle University of Thessaloniki, Greece</i>	3371
Pavlov, Yuri P. / <i>Bulgarian Academy of Sciences, Institute of Information and Communication Technologies, Bulgaria</i>	2188
Paya, Luis / <i>Miguel Hernández University, Spain</i>	6894
Pelet, Jean-Éric / <i>ESCE International Business School, France</i>	6070
Pelfer, Giuliano / <i>University of Florence, Italy</i>	3419
Pendergrass, William Stanley / <i>American Public University System, USA</i>	7077
Peracchia, Sara / <i>University of L’Aquila, Italy</i>	3296
Pereira, Claudia Teresa / <i>Universidad Nacional del Centro de la Provincia de Buenos Aires, Argentina</i>	7447
Pérez, Arturo Trejo / <i>Universidad Tecnológica de San Juan del Río, Mexico</i>	1277
Perilli, Viviana / <i>Lega del Filo d’Oro – Molfetta, Italy</i>	287
Perros, Harry G. / <i>North Carolina State University, USA</i>	6609
Perry, Gabriela Trindade / <i>Universidade Federal do Rio Grande do Sul, Brazil</i>	6376
Pes, Barbara / <i>Università degli Studi di Cagliari, Italy</i>	455
Piaggese, Danilo / <i>Framericas, USA</i>	5015
Picci, Simone / <i>University of Cagliari, Italy</i>	3273
Piciarelli, Claudio / <i>University of Udine, Italy</i>	6195
Pierce, Marlon / <i>Indiana University, USA</i>	1063

Pimenta, Marcelo Soares / <i>Federal University of Rio Grande do Sul (UFRGS), Brazil</i>	6424
Pino, José A. / <i>Universidad de Chile, Chile</i>	3745
Planu, Michael / <i>University of Cagliari, Italy</i>	3273
Platt, Rebecca / <i>Murdoch University, Australia</i>	7481
Poggi, Agostino / <i>University of Parma, Italy</i>	2392,6950
Poirier, Sandra / <i>Middle Tennessee State University, USA</i>	3930
Politis, Dionysios / <i>Aristotle University of Thessaloniki, Greece</i>	5886
Popescu(Predescu), Oana Mădălina / <i>Bucharest University of Economic Studies, Romania</i>	6579
Potgieter, Tertius N. / <i>Kimberley Hospital Complex, South Africa</i>	6147
Potocan, Vojko / <i>University of Maribor, Slovenia</i>	5559
Pourshahbaz, Shakiba / <i>Shahid Rajaei Teacher Training University, Iran</i>	7659
Prasanna, M. / <i>VIT University, India</i>	7503
Presadă, Diana / <i>Petroleum-Gas University of Ploiesti, Romania</i>	3945
Pulido, Ángela Coello / <i>University of Vigo, Spain</i>	2825
Pulkkis, Göran / <i>Arcada University of Applied Sciences, Finland</i>	7715
Punyanunt-Carter, Narissra Maria / <i>Texas Tech University, USA</i>	7069
Purushothaman, Aparna / <i>Aalborg University, Denmark</i>	4178
Qin, Xiongpai / <i>Renmin University of China, China</i>	1947
Quest, Mary / <i>Erikson Institute, USA</i>	7623
R., Sooraj T. / <i>VIT University, India</i>	3226
Rabuzin, Kornelije / <i>University of Zagreb, Croatia</i>	2031
Rada, Roy / <i>University of Maryland – Baltimore County, USA</i>	1
Radenkovic, Bozidar / <i>University of Belgrade, Serbia</i>	5398
Ragan, Mark A. / <i>The University of Queensland, Australia</i>	419
Rahim, Lukman Ab / <i>Universiti Teknologi PETRONAS, Malaysia</i>	3337
Rahim, Nihmiya Abdul / <i>UAE University, UAE</i>	3043
Rahimi, Mehrak / <i>Shahid Rajaei Teacher Training University, Iran</i>	2442,7659
Rahman, Mohd Nayyer / <i>Aligarh Muslim University, India</i>	6727
Rajakani, M. / <i>Mepco Schlenk Engineering College, India</i>	4975
Rajib, Md. Salah Uddin / <i>Jahangirnagar University, Bangladesh</i>	1896
Ramachandran, Muthu / <i>Leeds Metropolitan University, UK</i>	7525
Ramayah, T. / <i>University Sains Malaysia, Malaysia</i>	831
Ramírez, Juan Manuel Olivares / <i>Universidad Tecnológica de San Juan del Río, Mexico</i>	1277
Ramsay, Judith / <i>Manchester Metropolitan University, UK</i>	6398
Rao, N. Raghavendra / <i>FINAIT Consultancy Services, India</i>	5705
Rascão, José Poças / <i>Institute Polytechnic of Setúbal, Portugal</i>	4422
Rather, Mudasir Khazer / <i>University of Kashmir, India</i>	2264,4515
Rathore, Ashish Kumar / <i>Indian Institute of Technology Delhi, India</i>	7126
Ray, Anirban / <i>UNC Wilmington, USA</i>	2225
Ray, Nilanjan / <i>Adamas University, India</i>	3403
Realdon, Olivia / <i>University of Milano-Bicocca, Italy</i>	6223
Rebaudengo, Maurizio / <i>Politecnico di Torino, Italy</i>	3989
Recupero, Diego Reforgiato / <i>University of Cagliari, Italy</i>	3273
Reddy, D. Bhanusree / <i>VIT University, India</i>	2975
Reid-Martinez, Kathaleen / <i>Oral Roberts University, USA</i>	2588
Reinoso, Oscar / <i>Miguel Hernandez University, Spain</i>	6894

Reinthal, Ann / <i>Cleveland State University, USA</i>	5941
Remoaldo, Paula / <i>University of Minho, Portugal</i>	3460
Reynoldson, Katelyn R. / <i>Old Dominion University, USA</i>	3382
Ribeiro, Filipa M. / <i>University of Porto, Portugal</i>	3922
Ribeiro, Vitor P. / <i>University of Minho, Portugal</i>	3460
Ridao, Marcela / <i>Universidad Nacional del Centro de la Provincia de Buenos Aires, Argentina</i>	7411
Riel, Jeremy / <i>University of Illinois at Chicago, USA</i>	7888
Riemann, Ute Anna / <i>SAP SE, Germany</i>	873
Rifat, Afrin / <i>North South University, Bangladesh</i>	3579
Rigatos, Gerasimos / <i>Industrial Systems Institute, Greece</i>	15
Rijo, Rui Pedro Charters Lopes / <i>Polytechnic Institute of Leiria, Portugal</i>	3782
Robichaud, Xavier / <i>Univesité de Moncton, Canada</i>	7248
Robnik-Šikonja, Marko / <i>University of Ljubljana, Slovenia</i>	2085
Rodriguez, Alberto / <i>Universidad Miguel Hernández de Elche, Spain</i>	4937
Rodriguez, Juan Manuel / <i>ISISTAN, UNICEN-CONICET, Argentina</i>	6658
Rodriguez-Perez, Edelma / <i>CINVESTAV Unidad Guadalajara, Mexico</i>	7488
Rohunen, Anna / <i>University of Oulu, Finland</i>	4862
Roig, Juan Ignacio / <i>University of Castilla-La Mancha, Spain</i>	3698
Rojo, Lorenzo Fernández / <i>Miguel Hernandez University, Spain</i>	6894
Romão, Mário José Batista / <i>ISEG, Universidade de Lisboa, Portugal</i>	3756,5714
Romero, David Ernesto Troncoso / <i>CONACYT at ESCOM-IPN, Mexico</i>	6007,6234
Romero, Fernando Cabrita / <i>University of Minho, Portugal</i>	7150
Romero, Isabel / <i>University of Castilla-La Mancha, Spain</i>	3698
Romero, Jorge A. / <i>Towson University, USA</i>	923
Romero, Roberto Carlos Flores / <i>Universidad Tecnológica de San Juan del Río, Mexico</i>	1277
Rosen, Yigal / <i>Harvard University, USA</i>	2402
Rothberg, Helen N. / <i>Marist College, USA</i>	943
Rouhani, Saeed / <i>University of Tehran, Iran</i>	2932
Roy, Lorie / <i>The University of Texas at Austin, USA</i>	6685
Rubio, Alexis John M. / <i>University of the East, Philippines</i>	2466
Ruffin, T. Ray / <i>University of Phoenix, USA & Grand Canyon University, USA & Ashford University, USA & North Carolina Wesleyan College, USA</i>	1463,3805
Ruhi, Umar / <i>University of Ottawa, Canada</i>	4278
Ruiz, María Elena García / <i>University of Cantabria, Spain</i>	3183
Ruiz-Martínez, Antonio / <i>University of Murcia, Spain</i>	2749
Saeed, Fauzan / <i>Usman Institute of Technology, Pakistan</i>	166
Saeed, Munazza / <i>University of Malaya, Malaysia</i>	36
Safont, Gonzalo / <i>Universitat Politècnica de València, Spain</i>	4937
Saharidis, Georgios K. D. / <i>University of Thessaly, Greece</i>	5411
Saito, Nagayuki / <i>Ochanomizu University, Japan</i>	4962
Saiz-Alvarez, José Manuel / <i>Tecnologico de Monterrey, Mexico</i>	3020
Sajib, Md. Qutub Uddin / <i>China University of Geosciences (CUG), China</i>	1896
Salazar, Addisson / <i>Universitat Politècnica de València, Spain</i>	4937
Saldana-Perez, Ana Maria Magdalena / <i>Instituto Politecnico Nacional, Mexico</i>	6973
Sambhanthan, Arunasalam / <i>Curtin University, Australia</i>	609

Samuel, Anand A. / VIT University, India	2975
Sánchez, Jorge Manjarrez / Instituto Tecnológico Superior de Jerez, Mexico	1806
Sanchez-Ruiz, Lidia / University of Cantabria, Spain	3870
Sandvig, John Christopher / Western Washington University, USA	8087
Sanguinetti, Paola / University of Kansas, USA	983
Santos-Trigo, Manuel / Cinvestav-IPN, Mexico	3794
Sarathy, Rathindra / Oklahoma State University, USA	5583
Sarivougioukas, John / “G. Gennimatas” Athens General Hospital, Greece	7765
Sarp, Gulcan / Suleyman Demirel University, Turkey	3503
Satterfield, Debra / California State University – Long Beach, USA	8079
Savaş, Kübra / Istanbul University, Turkey	5490
Savastano, Marco / Sapienza University of Rome, Italy	7805
Scanzio, Stefano / CNR-IEIT, Italy	1227
Scasseddu, Riccardo / University of Cagliari, Italy	3273
Schnackenberg, Heidi Lee / SUNY Plattsburgh, USA	6245
Schulte, Frederik / University of Hamburg, Germany	601
Segall, Richard S. / Arkansas State University, USA	1783
Seidenstricker, Sven / Fraunhofer Institute for Industrial Engineering, Germany	4560
Self, Lucy / University of Sussex, UK	1153
Selvi, Ihsan Hakan / Sakarya University, Turkey	5490
Sembrano, Christian Gerard / University of the East, Philippines	2466
Sen, Devjani / University of Ottawa, Canada	6083
Sengupta, Susmit / Maulana Abul Kalam Azad University of Technology, India	4872
Seoane, Fernando / KI-Karolinska Institutet, Sweden & University of Borås, Sweden	7754
Seok, Soonhwa / Korea University, South Korea	6361
Serpa, Nilo Sylvio / Universidade Paulista, Brazil	1040
Sezgin, Anil / Yildiz Technical University, Turkey	7470
Sgobbi, Francesca / University of Brescia, Italy	4785
Shaanika, Irja Naambo / Namibia University of Science and Technology – Windhoek, Namibia	2943,4480
Shafiq, Huma / University of Kashmir, India	4535
Shah, Huma / Coventry University, UK	251
Shahamiri, Seyed Reza / Manukau Institute of Technology, New Zealand	7392
Shaharudin, Muhammad Shabir / Universiti Sains Malaysia, Malaysia	5446
Shahmir, Nazlie / WestJet Airlines Limited, Canada	7399
Shao, Lan / University of Oulu, Finland	2108
Sharif, Muhammad Majid / National University of Sciences and Technology (NUST), Pakistan	5169
Sharma, Dolly / PEC University of Technology, India	477
Sharma, Ravi S. / Nanyang Technological University, Singapore	2280
Sharma, Supreeth / Akamai Technologies, India	847
Sharmila Banu K / VIT University, India	1835
Shawyun, Teay / King Saud University, Saudi Arabia	3881
Shibli, Muhammad Awais / VisionIT, USA	2869
Shibuya, Kazuhiko / ROIS, Japan	3538
Siano, Pierluigi / University of Salerno, Italy	15

Siddiqui, Adnan Ahmed / Hamdard University, Pakistan.....	166
Sidman, Cara / Arizona State University, USA	5820
Signorella, Margaret L. / Pennsylvania State University, USA	7026
Siller, Mario / CINVESTAV Unidad Guadalajara, Mexico	6522
Silva, Fillipe Souza / University of Campinas, Brazil.....	1331
Silva, Sara Catarina Gomes / University of Minho, Portugal	3460
Silvana de Rosa, Annamaria / Sapienza University of Rome, Italy.....	4038,4064,4404,5237,7014,7044
Simões, Jorge / ISPGaya, Portugal.....	800
Sims, Cynthia M. / Clemson University, USA	4326
Singh, Harwinder / Guru Nanak Dev Engineering College, India.....	587
Singh, Indu / Griffith University, Australia.....	5853
Singh, Kuldeep / Thapar University, India.....	6106
Singh, R. Arvind / Kumaraguru College of Technology (KCT), India.....	431
Singh, Shailendra / PEC University of Technology, India	477
Siu, Kin Wai Michael / The Hong Kong Polytechnic University, Hong Kong	7382
Skarmeta, Antonio F. / University of Murcia, Spain	2749
Skoll, Geoffrey / SUNY at Buffalo, USA	3637
Skoumpopoulou, Dimitra / Northumbria University, UK	1605
Slini, Theodora / Aristotle University of Thessaloniki, Greece	3371
Šmahel, David / Masaryk University – Brno, Czech Republic	4223
Smith, Alan D. / Robert Morris University, USA.....	4837,5273,5476,5516,5570,5595,5918,5981,6094
Smith, Craig / Kyoto University of Foreign Studies, Japan.....	3850
Smith, Pete / University of Aberdeen, UK.....	7815
Smith, Peter A. / University of Central Florida, USA	3327
Smith-Ditizio, Amber A. / Texas Woman's University, USA.....	4837,5273,5570,5595,5918,5981,6094
Sohrabzadeh, Sara / Tehran University of Medical Science, Iran	4436
Sokolov, Alexander / National Research University Higher School of Economics, Russia	4704
Sokouti, Babak / Tabriz University of Medical Sciences, Iran.....	1676
Sokouti, Massoud / Mashhad University of Medical Sciences, Iran.....	1676
Soltani, Pooya / University of Porto, Portugal.....	7358
Šoltés, Michal / Technical University of Košice, Slovakia	5841
Sonnenberg, Christian / Florida Institute of Technology, USA.....	3516
Soomarah, Garry / University of Trinidad and Tobago, Trinidad and Tobago	2532
Soriano, Juan Vicente Izquierdo / La Ribera University Hospital, Spain.....	5787
Sotiriadis, Marios D. / University of South Africa (UNISA), South Africa	4088
Spaananburg, Lambert / Comoray AB, Sweden.....	6114
Spallone, Roberta / Politecnico di Torino, Italy.....	973,5213,7856
Speed, Chris / University of Edinburgh, UK	6695
Spenser, Karin / Nottingham Trent University, UK.....	4168
Spiegel, Thais / Rio de Janeiro State University, Brazil	2076
Spires, Hiller A. / North Carolina State University, USA.....	2235
Spring, Kristian J. / Brigham Young University, USA.....	1487
Sreedhar, G. / Rashtriya Sanskrit Vidyapeetha (Deemed University), India.....	8006

Srinivasan, Srividhya / <i>University of Madras, India</i>	77
Stancu, Stelian / <i>Bucharest University of Economic Studies, Romania</i>	5158,6579
Starrett, David / <i>Columbia College, USA</i>	1451
Stasolla, Fabrizio / <i>University of Bari, Italy</i>	287
Stavrianos, Athanasios T. / <i>2nd Technical Vocational School of Xanthi, Greece</i>	2651
Stefanovic, Nenad / <i>University of Kragujevac, Serbia</i>	5398,5538
Steinfeld, Joshua M. / <i>Old Dominion University, USA</i>	6710
Steri, Gary / <i>European Commission – Joint Research Centre, Italy</i>	6136
Stewart, Craig A. / <i>Indiana University, USA</i>	1063
Steyn, Gert / <i>Kimberley Hospital Complex, South Africa</i>	6147
Stojanov, Zeljko / <i>University of Novi Sad, Serbia</i>	7514
Suaste, Ernesto / <i>Cinvestav-IPN, Mexico</i>	3794
Surmen, Hasan Kemal / <i>Istanbul University, Turkey</i>	298
Susanto, Heru / <i>Indonesian Institute of Sciences, Indonesia & Tunghai University, Taiwan</i>	7369
Svitek, Miroslav / <i>Czech Technical University in Prague, Czech Republic</i>	4391
Swaid, Samar I. / <i>Philander Smith College, USA</i>	8066
Swanson, Karen / <i>Mercer University, USA</i>	1432
Swanson, Pete / <i>Georgia State University, USA</i>	7682
Switzer, Jamie S. / <i>Colorado State University, USA</i>	2121
Switzer, Ralph V. / <i>Colorado State University, USA</i>	2121
Sylaïou, Stella / <i>Aristotle University of Thessaloniki, Greece</i>	7002
Syropoulos, Apostolos / <i>Greek Molecular Computing Group, Greece</i>	2337,2651
Szczepański, Zbigniew / <i>Kazimierz Wielki University, Poland</i>	521
Szeto, Kimmy / <i>Baruch College, City University of New York, USA</i>	6447
Szopa, Anna / <i>Jagiellonian University, Poland</i>	563
T, Tamilarasi / <i>VIT University, India</i>	7503
T., Santhanamery / <i>Universiti Teknologi MARA, Malaysia</i>	831
Tabrizi, Aydin / <i>University of Kansas, USA</i>	983
Tadj, Lotfi / <i>Fairleigh Dickinson University, Canada</i>	1570
Taieb, Basma / <i>University of Cergy Pontoise, France</i>	6070
Taiwo, Ayankunle A. / <i>Texas A&M University – Commerce, USA</i>	3654
Tam, June Poh Kim / <i>Universiti Sains Malaysia, Malaysia</i>	5357
Tambo, Torben / <i>Aarhus University, Denmark</i>	2920
Tan, Amanda C. / <i>University of Central Florida, USA</i>	995
Tani, Kawtar / <i>UCOL, New Zealand</i>	3194
Tannous, Katia / <i>University of Campinas, Brazil</i>	1331
Tazhina, Gainiya / <i>University of International Business, Kazakhstan</i>	5626
Tchangani, Ayeley P. / <i>Université Fédérale Toulouse Midi-Pyrénées, France</i>	7282
Tech, Deb / <i>Dakota State University, USA</i>	3861
Tecpoyotl-Torres, Margarita / <i>Universidad Autónoma del Estado de Morelos, Mexico</i>	2897
Teik, Cheah Cheng / <i>Universiti Sains Malaysia, Malaysia</i>	5422
Teixeira, Leonor / <i>University of Aveiro, Portugal</i>	888,3963
Terras, Melody M. / <i>University of the West of Scotland, UK</i>	6398
Teshome, Assefa K. / <i>Victoria University, Australia</i>	6319
Thakare, Vaishali Ravindra / <i>VIT University, India</i>	1667
Thakur, R. S. / <i>Maulana Azad National Institute of Technology, India</i>	1376

Thompson, Nik / Curtin University, Australia	4124,7481
Tirado-Morueta, Ramon / University of Huelva, Spain.....	3976
Titova, Svetlana / Far Eastern Federal University, Russia	5116
Tiwari, Vivek / International Institute of Information Technology Naya Raipur, India.....	1376
Tobin, Maryann Tatum / University of Miami, USA.....	1508
Tomaiuolo, Michele / University of Parma., Italy	2392,6950
Tommassel, Antonela / ISISTAN, UNICEN-CONICET, Argentina	6658
Torkul, Orhan / Sakarya University, Turkey	2200
Torres-Ruiz, Miguel Jesus / Instituto Politécnico Nacional, Mexico	6973
Tounekti, Oussama / University of Murcia, Spain	2749
Trad, Antoine / Institute of Business and Information Systems Transformation Management, Switzerland	620,636,5607
Tran, Ben / Alliant International University, USA.....	277,671,3173,3260,6756
Tran, Thomas / University of Ottawa, Canada	1757
Traxler, John / University of Wolverhampton, UK	6347
Tremblay, Diane-Gabrielle / University of Quebec, Canada	5027
Tripathy, B. K. / VIT University, India	1835,3226,4528
Tsalighopoulos, Miltiadis / Aristotle University of Thessaloniki, Greece.....	5886
Tsubota, Yasushi / Kyoto Institute of Technology, Japan	3850
Tüfekci, Ashhan / Gazi University, Turkey	6789
Turel, Vehbi / The University of Bingöl, Turkey	2357
Turgay, Safiye / Sakarya University, Turkey.....	2200
Turgay, Tahsin / Sakarya University, Turkey	2200
Tzanis, George / Aristotle University of Thessaloniki, Greece.....	1984
Ulloa, Juan Carlos Solís / Instituto Tecnológico Superior de Cintalapa, Mexico	1277
Umair, Sajid / National University of Sciences and Technology (NUST), Pakistan	5169
Umukoro, Immanuel Ovemeso / Pan-Atlantic University, Nigeria	2724
Ursyn, Anna / University of Northern Colorado, USA	5103
Usat, Selisa / Universiti Teknologi Malaysia, Malaysia.....	7088
Uysal, Tugba Ucma / Mugla Sıtkı Kocman University, Turkey	5737
V., Sriharsha K. / National Institute of Technology Trichy, India	212
Vagelatos, Aristides / CTI&P, Greece	7765
Vagnoli, Laura / Meyer Children's Hospital of Florence, Italy	5955
Van Zyl, D. H. / University of Pretoria, South Africa	777
Vargas-Bernal, Rafael / Instituto Tecnológico Superior de Irapuato, Mexico	2673,2897
Vaterlaus, J. Mitchell / Montana State University, USA.....	7097
Vega-Barbas, Mario / KTH-Royal Institute of Technology, Sweden & ESNE-Universidad Camilo José Cela, Spain	7754
Veglis, Andreas A. / Aristotle University of Thessaloniki, Greece	1196,6476,8046
Velikovskiy, J. T. / University of Newcastle, Australia	4666
Verelst, Jan / University of Antwerp, Belgium.....	5317
Vergara, Luis / Universitat Politècnica de València, Spain	4937
Verma, Anil Kumar / Thapar University, India.....	6106
Vernier, Marco / University of Udine, Italy.....	2064,6195
Verteramo, Saverino / University of Calabria, Italy	7805
Vestias, Mário Pereira / Instituto Politécnico de Lisboa, Portugal	4018,4549,4607

Véstias, Mário Pereira / Instituto Politecnico de Lisboa, Portugal	6307
Vettraino, Laura / Learning Community, Italy	1559
Vilas-Boas, João Paulo / University of Porto, Portugal	7358
Vinzuda, Vipul / National Institute of Design Gandhinagar, India	1260
Viswanathan, Revathi / B. S. Abdur Rahman University, India	7635
Vondrackova, Petra / Charles University in Prague, Czech Republic	4223
Vong, Yunny Meas / Centro de Investigación y Desarrollo Tecnológico en Electroquímica, Mexico	1277
Vorndran, Brittany A. / Seton Hall University, USA.....	5810
Voskoglou, Michael / Graduate Technological Educational Institute (T.E.I.), Greece	3215
Voß, Stefan D. / University of Hamburg, Germany	601
Walczak, Steven / University of South Florida, USA	98,120
Wang, Jingying / University of Chinese Academy of Sciences, China	132
Wang, John / Montclair State University, USA	5388
Wang, Shun-Yung Kevin / University of South Florida – St. Petersburg, USA.....	1656
Wang, Ye / Zhejiang Gongshang University, China	7588
Wani, Zahid Ashraf / University of Kashmir, India.....	4535,5262,6739
Warwick, Kevin / Coventry University, UK	251
Wasim, Muhammad / Usman Institute of Technology, Pakistan	166
Wassan, Jyotsna Talreja / University of Delhi, India	326
Weaver, Kari D. / University of South Carolina – Aiken, USA	7292
Weckman, Gary R. / Ohio University, USA	6642
Weintraub, Eli / Afeka Tel Aviv College of Engineering, Israel.....	1627
Weiss, Andrew Philip / California State University – Northridge, USA.....	5226
Wellman, Barry / NetLab Network, Canada	7057
Wernert, Eric / Indiana University, USA	1063
Westerlund, Mika / Carleton University, Canada	1684
White, Jonathan R. / Höskolan Dalarna, Sweden	1217
Wiafe, Isaac / Ghana Institute of Management and Public Administration, Ghana	7785
Wickramasinghe, Nilmini / Epworth HealthCare, Australia & Deakin University, Australia	3725,6806
Widiyanesti, Sri / Telkom University, Indonesia	5550
Wierzchon, Slawomir T. / Polish Academy of Sciences, Poland	1962
Wilkins-Diehr, Nancy / San Diego Supercomputer Center, USA	1063
Williams, Anne-Evan K. / Billings Middle School, USA	5183
Williams, Idongesit / Aalborg University, Denmark	3549
Williams, James / University of Pittsburgh, USA.....	2707
Wimmer, Hayden / Georgia Southern University, USA.....	1
Winter, Jenifer Sunrise / University of Hawaii at Manoa, USA.....	4951
Wodczak, Michal / Samsung Electronics, Poland.....	6499
Wolf, Frank / CSSTA L3C, USA	6813
Wong, Yi Lin / The Hong Kong Polytechnic University, Hong Kong.....	7382
Wood, Lincoln C. / University of Otago, New Zealand & Curtin University, Australia	5335
Wooldridge, Deborah G. / Bowling Green State University, USA.....	3930
Worthington, Debra L. / Auburn University, USA	6985
Wright, Jorja / University of Charleston, USA.....	7898

Wright, Michelle F. / <i>Masaryk University, Czech Republic</i>	1723,7077
Wulansari, Puspita / <i>Telkom University, Indonesia</i>	5327
Xafopoulos, Alexandros / <i>University College London, UK</i>	6336
Xiaojun (Jenny) Yuan / <i>University of Albany, USA</i>	3745
Xu, Henry / <i>The University of Queensland, Australia</i>	5433
Yadav, Aman / <i>Michigan State University, USA</i>	2292
Yahya, Fadwa / <i>University of Sfax, Tunisia</i>	765
Yang, Hung-Jen / <i>National Kaohsiung Normal University, Taiwan</i>	6057
Yang, Yuqi / <i>University of Windsor, Canada</i>	2314
Yasuda, Gen'ichi / <i>Nagasaki Institute of Applied Science, Japan</i>	6836,7435
Yengin, İlker / <i>A*STAR, Institute of High Performance Computing, Singapore</i>	7168
Yeomans, Julian Scott / <i>York University, Canada</i>	2178
Yfantis, Vasileios / <i>Ionian University, Greece</i>	1033
Yilmaz, Rahime / <i>Istanbul University, Turkey</i>	7470
Young, Richard A. / <i>Driving Safety Consulting, LLC, USA</i>	5992
Young, William A. / <i>Ohio University, USA</i>	6642
Younis, Eman / <i>Minia University, Egypt</i>	8023
Yu, Ligu / <i>Indiana University – South Bend, USA</i>	2739,7116
Yücer, Emre / <i>Erzincan University, Turkey</i>	7831
Yuhua, Fu / <i>CNOOC Research Institute, China</i>	4399
Yusuf, Sarina / <i>Universiti Putra Malaysia, Malaysia</i>	1704,2761
Zaeri, Fahimeh / <i>Auckland University of Technology, New Zealand</i>	539
Zainab, Tazeem / <i>University of Kashmir, India</i>	5262,6739
Zaitsev, Dmitry A. / <i>International Humanitarian University, Ukraine</i>	7731
Zeinali, Ali Akbar / <i>Universiti Sains Malaysia, Malaysia</i>	5136
Zhang, Xinxin / <i>University of Alberta, Canada</i>	2369
Zhang, Yu-Jin / <i>Tsinghua University, China</i>	1308,1319,1344
Zhao, Wenbing / <i>Cleveland State University, USA</i>	1091,1238,1647,3056,4927,5876,5941,8056
Zhou, Chunfang / <i>Aalborg University, Denmark</i>	4178
Zhu, Changye / <i>University of Chinese Academy of Sciences, China</i>	132
Zhu, Tingshao / <i>University of Chinese Academy of Sciences, China</i>	132
Zhu, W. K. / <i>The University of Hong Kong, Hong Kong</i>	505
Zunino, Alejandro / <i>ISISTAN, UNICEN-CONICET, Argentina</i>	6658
Zurloni, Valentino / <i>Università di Milano-Bicocca, Italy</i>	6223
Zykov, Sergey V. / <i>National Research University Higher School of Economics, Russia</i>	1396

Table of Contents by Volume

Preface	clxii
Guide to the Encyclopedia of Information Science and Technology, Fourth Edition	clxviii
Acknowledgment	clxx
About the Editor	clxxi

Volume I **(A - BU)**

Category: Accounting and Finance

Applying Artificial Intelligence to Financial Investing.....	1
<i>Hayden Wimmer, Georgia Southern University, USA</i>	
<i>Roy Rada, University of Maryland – Baltimore County, USA</i>	
Distributed Parameter Systems Control and Its Applications to Financial Engineering	15
<i>Gerasimos Rigatos, Industrial Systems Institute, Greece</i>	
<i>Pierluigi Siano, University of Salerno, Italy</i>	
Does Inter-Bank Investments Restraints Financing Performance of Islamic Banks?.....	36
<i>Mohammad Taqiuddin Mohamad, University of Malaya, Malaysia</i>	
<i>Munazza Saeed, University of Malaya, Malaysia</i>	
An Extension to the Delone and Mclean Information Systems Success Model and Validation in the Internet Banking Context	49
<i>Veeraraghavan Jagannathan, National Institute of Technology, India</i>	
<i>Senthilarasu Balasubramanian, National Institute of Technology, India</i>	
<i>Thamaraiselvan Natarajan, National Institute of Technology, India</i>	
Impact of Business Groups on Payout Policy in India.....	61
<i>Ahana Bose, Indian Institute of Management Calcutta, India</i>	

Noise Trader.....	71
<i>Po-Keng Cheng, State University of New York, Stony Brook University, USA</i>	
Safeguarding of ATM	77
<i>Srividhya Srinivasan, University of Madras, India</i>	
<i>Priya Krishnamoorthy, SASTRA University, India</i>	
<i>Raghuraman Koteeswaran, SASTRA University, India</i>	

Category: Artificial Intelligence

Artificial Ethics.....	88
<i>Laura L. Pană, Polytechnic University of Bucharest, Romania</i>	
Artificial Intelligence	98
<i>Steven Walczak, University of South Florida, USA</i>	
Artificial Intelligence Review	106
<i>Amal Kilani, University of Gabes Tunisia, Tunisia</i>	
<i>Ahmed Ben Hamida, University of Sfax, Tunisia</i>	
<i>Habib Hamam, University of Moncton, Canada</i>	
Artificial Neural Networks.....	120
<i>Steven Walczak, University of South Florida, USA</i>	
Automatic Emotion Recognition Based on Non-Contact Gaits Information.....	132
<i>Jingying Wang, University of Chinese Academy of Sciences, China</i>	
<i>Baobin Li, University of Chinese Academy of Sciences, China</i>	
<i>Changye Zhu, University of Chinese Academy of Sciences, China</i>	
<i>Shun Li, University of Chinese Academy of Sciences, China</i>	
<i>Tingshao Zhu, University of Chinese Academy of Sciences, China</i>	
Board Games AI	144
<i>Tad Gonsalves, Sophia University, Japan</i>	
Computational Intelligence Approaches to Computational Aesthetics	156
<i>Erandi Lakshika, University of New South Wales, Australia</i>	
<i>Michael Barlow, University of New South Wales, Australia</i>	
Dotted Raster-Stereography	166
<i>Muhammad Wasim, Usman Institute of Technology, Pakistan</i>	
<i>Fauzan Saeed, Usman Institute of Technology, Pakistan</i>	
<i>Abdul Aziz, Usman Institute of Technology, Pakistan</i>	
<i>Adnan Ahmed Siddiqui, Hamdard University, Pakistan</i>	
Hybrid Computational Intelligence and the Basic Concepts and Recent Advances.....	180
<i>Georgios Dounias, University of the Aegean, Greece</i>	

Incremental Approach to Classification Learning	191
<i>Xenia Alexandre Naidenova, Research Centre of Military Medical Academy – Saint Petersburg, Russia</i>	
Machine Dreaming.....	202
<i>James Frederic Pagel, University of Colorado School of Medicine, USA</i>	
Moving Object Detection and Tracking Based on the Contour Extraction and Centroid Representation.....	212
<i>Naveenkumar M, National Institute of Technology Trichy, India</i>	
<i>Sriharsha K. V., National Institute of Technology Trichy, India</i>	
<i>Vadivel A, National Institute of Technology Trichy, India</i>	
Semantic Intelligence.....	220
<i>Maria K. Koleva, Institute of Catalysis, Bulgarian Academy of Sciences, Bulgaria</i>	
The Summers and Winters of Artificial Intelligence	229
<i>Tad Gonsalves, Sophia University, Japan</i>	
Swarm Intelligence for Multi-Objective Optimization in Engineering Design	239
<i>Janga Reddy Manne, Indian Institute of Technology Bombay, India</i>	
Trust and Decision Making in Turing’s Imitation Game.....	251
<i>Huma Shah, Coventry University, UK</i>	
<i>Kevin Warwick, Coventry University, UK</i>	

Category: Assistive Technologies

Apps as Assistive Technology	266
<i>Emily C. Bouck, Michigan State University, USA</i>	
<i>Sara M. Flanagan, University of Kentucky, USA</i>	
<i>Missy D. Cosby, Michigan State University, USA</i>	
Assistive Technology and Human Capital for Workforce Diversity.....	277
<i>Ben Tran, Alliant International University, USA</i>	
Assistive Technology for Supporting Communication, Occupation, and Leisure by Children With Severe to Profound Developmental Disabilities	287
<i>Fabrizio Stasolla, University of Bari, Italy</i>	
<i>Viviana Perilli, Lega del Filo d’Oro – Molfetta, Italy</i>	
<i>Adele Boccasini, Lega del Filo d’Oro – Termini Imerese, Italy</i>	
Design, Manufacture, and Selection of Ankle-Foot-Orthoses.....	298
<i>Hasan Kemal Surmen, Istanbul University, Turkey</i>	
<i>Nazif Ekin Akalan, Istanbul University, Turkey</i>	
<i>Yunus Ziya Arslan, Istanbul University, Turkey</i>	

A Disability-Aware Mentality to Information Systems Design and Development.....	314
<i>Julius T. Nganji, University of Ottawa, Canada</i>	

Category: Big Data

Adapting Big Data Ecosystem for Landscape of Real World Applications	326
<i>Jyotsna Talreja Wassan, University of Delhi, India</i>	
Big Data Analysis and Mining.....	338
<i>Carson K. Leung, University of Manitoba, Canada</i>	
Big Data Analytics for Tourism Destinations	349
<i>Wolfram Höpken, University of Applied Sciences Ravensburg-Weingarten, Germany</i>	
<i>Matthias Fuchs, Mid-Sweden University, Sweden</i>	
<i>Maria Lexhagen, Mid-Sweden University, Sweden</i>	
Big Data Time Series Stream Data Segmentation Methods	364
<i>Dima Alberg, Shamoon College of Engineering (SCE), Israel</i>	
Challenges for Big Data Security and Privacy	373
<i>M. Govindarajan, Annamalai University, India</i>	
How Visualisation and Interaction Can Optimize the Cognitive Processes Towards Big Data	381
<i>Antonio Feraco, Fraunhofer IDM@NTU, Singapore</i>	
<i>Marius Erdt, Fraunhofer IDM@NTU, Singapore</i>	
Managing and Visualizing Unstructured Big Data.....	394
<i>Ananda Mitra, Wake Forest University, USA</i>	
Mining Big Data and Streams.....	406
<i>Hoda Ahmed Abdelhafez, Suez Canal University, Egypt</i>	

Category: Bioinformatics

Bioinformatics.....	419
<i>Mark A. Ragan, The University of Queensland, Australia</i>	
Bioinspired Solutions for MEMS Tribology	431
<i>R. Arvind Singh, Kumaraguru College of Technology (KCT), India</i>	
<i>S. Jayalakshmi, Kumaraguru College of Technology (KCT), India</i>	
Building Gene Networks by Analyzing Gene Expression Profiles.....	440
<i>Crescenzo Gallo, University of Foggia, Italy</i>	
Concerns and Challenges of Cloud Platforms for Bioinformatics.....	455
<i>Nicoletta Dessì, Università degli Studi di Cagliari, Italy</i>	
<i>Barbara Pes, Università degli Studi di Cagliari, Italy</i>	

Category: Biology

- GWAS as the Detective to Find Genetic Contribution in Diseases 466
Simanti Bhattacharya, University of Kalyani, India
Amit Das, University of Kalyani, India
- RNA Interference Therapeutics and Human Diseases..... 477
Dolly Sharma, PEC University of Technology, India
Shailendra Singh, PEC University of Technology, India
Trilok Chand, PEC University of Technology, India

Category: Biomedical Engineering

- General Perspectives on Electromyography Signal Features and Classifiers Used for Control of Human Arm Prosthetics..... 492
Faruk Ortes, Istanbul University, Turkey
Derya Karabulut, Halic University, Turkey
Yunus Ziya Arslan, Istanbul University, Turkey
- The Principle and Process of Digital Fabrication of Biomedical Objects 505
S. H. Choi, The University of Hong Kong, Hong Kong
H. H. Cheung, The University of Hong Kong, Hong Kong
W. K. Zhu, The University of Hong Kong, Hong Kong
- Reverse Engineering in Rehabilitation 521
Emilia Mikołajewska, Nicolaus Copernicus University, Poland
Marek Macko, Kazimierz Wielki University, Poland
Zbigniew Szczepański, Kazimierz Wielki University, Poland
Dariusz Mikołajewski, Kazimierz Wielki University, Poland

Category: Business and Organizational Research

- Acceptance of E-Reverse Auction From the Buyer Perspective..... 530
Cigdem Altin Gumussoy, Istanbul Technical University, Turkey
Bilal Gumussoy, Shell and Turcas Petrol Inc., Turkey
- Advanced ICT Methodologies (AIM) in the Construction Industry 539
M. Reza Hosseini, Deakin University, Australia
Saeed Banihashemi, University of Technology Sydney, Australia
Fahimeh Zaeri, Auckland University of Technology, New Zealand
Alireza Adibfar, University of Florida, USA
- Amplifying the Significance of Systems Thinking in Organization..... 551
Mambo Governor Mupepi, Grand Valley State University, USA
Sylvia C. Mupepi, Grand Valley State University, USA
Jaideep Motwani, Grand Valley State University, USA

The Application of Crowdsourced Processes in a Business Environment	563
<i>Katarzyna Kopeć, Tischner European University, Poland</i>	
<i>Anna Szopa, Jagiellonian University, Poland</i>	
Architecture as a Tool to Solve Business Planning Problems	573
<i>James McKee, Independent Researcher, Australia</i>	
Benchmarking Performance Indicators of Indian Rail Freight by DEA Approach	587
<i>Neeraj Bhanot, Dr. B. R. Ambedkar National Institute of Technology Jalandhar, India</i>	
<i>Harwinder Singh, Guru Nanak Dev Engineering College, India</i>	
Bi-Directional Business/IT Alignment	601
<i>Hashim Chunpir, German Climate Computing Centre (DKRZ), Germany</i>	
<i>Frederik Schulte, University of Hamburg, Germany</i>	
<i>Yannick Bartens, University of Hamburg, Germany</i>	
<i>Stefan D. Voß, University of Hamburg, Germany</i>	
Business Sustainability Indices	609
<i>Arunasalam Sambhanthan, Curtin University, Australia</i>	
The Business Transformation Framework, Agile Project and Change Management	620
<i>Antoine Trad, Institute of Business and Information Systems Transformation Management,</i>	
<i>Switzerland</i>	
<i>Damir Kalpić, University of Zagreb, Croatia</i>	
The Business Transformation Framework and Its Business Engineering Law Support for (e)Transactions	636
<i>Antoine Trad, Institute of Business and Information Systems Transformation Management,</i>	
<i>Switzerland</i>	
<i>Damir Kalpić, University of Zagreb, Croatia</i>	
Challenges of Meta Access Control Model Enforcement to an Increased Interoperability	651
<i>Sérgio Luís Guerreiro, University of Lisbon, Portugal</i>	
Cognitive Ergonomics in 2016	662
<i>Ronald John Lofaro, Embry-Riddle Aeronautical University, USA</i>	
Corporate Social Responsibility	671
<i>Ben Tran, Alliant International University, USA</i>	
Digital Transformation Journeys in a Digitized Reality	682
<i>Jurgen Janssens, QSpin, Belgium</i>	
A Framework for Exploring IT-Led Change in Morphing Organizations	694
<i>Sharon A. Cox, Birmingham City University, UK</i>	

How the Crowdsourcing Enhance the Co-Creation Into the Virtual Communities.....	707
<i>Bahri Ammari Nedra, IHEC of Carthage, USA</i>	
Hyper-Sensitivity in Global Virtual Teams	720
<i>Andre Araujo, Texas A&M University, USA</i>	
Lean and Six Sigma Innovation and Design.....	729
<i>Rick Edgeman, Utah State University, USA</i>	
Motivational Factors of Telework.....	740
<i>Arlene J. Nicholas, Salve Regina University, USA</i>	
Organizational Transparency	754
<i>Gustavo de Oliveira Almeida, Federal University of the State of Rio de Janeiro, Brazil</i>	
<i>Claudia Cappelli, Federal University of the State of Rio de Janeiro, Brazil</i>	
<i>Cristiano Maciel, Federal University of Mato Grosso, Brazil</i>	
Social Business Process Modeling.....	765
<i>Fadwa Yahya, University of Sfax, Tunisia</i>	
<i>Khouloud Boukadi, University of Sfax, Tunisia</i>	
<i>Zakaria Maamar, Zayed University, UAE</i>	
<i>Hanène Ben-Abdallah, King Abdulaziz University, Saudi Arabia</i>	
Social Issues in IT Project Teams:.....	777
<i>Awie C. Leonard, University of Pretoria, South Africa</i>	
<i>D. H. Van Zyl, University of Pretoria, South Africa</i>	
Viewpoints on Business Process Models.....	788
<i>Giorgio Bruno, Politecnico di Torino, Italy</i>	

Category: Business Education

Serious Games in Entrepreneurship Education	800
<i>Fernando Almeida, Polytechnic Institute of Gaya, Portugal</i>	
<i>Jorge Simões, ISPGaya, Portugal</i>	

Volume II **(BU - CU)**

Category: Business Information Systems

Architectural Framework for the Implementation of Information Technology Governance in Organisations	810
<i>Thami Batyashe, Cape Peninsula University of Technology, South Africa</i>	
<i>Tiko Iyamu, Cape Peninsula University of Technology, South Africa</i>	

Continuous Assurance and the Use of Technology for Business Compliance	820
<i>Rui Pedro Figueiredo Marques, Universidade de Aveiro, Portugal</i>	
Explaining and Predicting Users' Continuance Usage Intention Toward E-Filing Utilizing Technology Continuance Theory	831
<i>Santhanamery T., Universiti Teknologi MARA, Malaysia</i>	
<i>T. Ramayah, University Sains Malaysia, Malaysia</i>	
Forecasting the Demand of Agricultural Crops/Commodity Using Business Intelligence Framework	847
<i>Satyadhyan Chickerur, KLE Technological University, India</i>	
<i>Supreeth Sharma, Akamai Technologies, India</i>	
<i>Prashant M. Narayankar, KLE Technological University, India</i>	
Integrated Data Architecture for Business	862
<i>Richard Kumaradjaja, Renaissance Consulting, Indonesia</i>	
IT Strategy Follows Digitalization.....	873
<i>Thomas Ochs, Villeroy & Boch, Germany</i>	
<i>Ute Anna Riemann, SAP SE, Germany</i>	
The Main Concepts Behind the Dematerialization of Business Processes.....	888
<i>Liliana Ávila, University of Aveiro, Portugal</i>	
<i>Leonor Teixeira, University of Aveiro, Portugal</i>	
Scanning for Blind Spots	899
<i>Barbara Jane Holland, Brooklyn Public Library, USA</i>	
Strategic Information Systems Planning.....	912
<i>Maria Kamariotou, University of Macedonia, Greece</i>	
<i>Fotis Kitsios, University of Macedonia, Greece</i>	
Sustainable Advantages of Business Value of Information Technology	923
<i>Jorge A. Romero, Towson University, USA</i>	
Using Business Analytics in Franchise Organizations.....	930
<i>Ye-Sho Chen, Louisiana State University, USA</i>	

Category: Business Intelligence

Big Data, Knowledge, and Business Intelligence	943
<i>G. Scott Erickson, Ithaca College, USA</i>	
<i>Helen N. Rothberg, Marist College, USA</i>	

Business Intelligence	951
<i>Richard T. Herschel, Saint Joseph's University, USA</i>	

Improving Competitiveness Through Organizational Market Intelligence.....	961
<i>George Leal Jamil, Informações em Rede, Brazil</i>	
<i>Leandro Rocha Dos Santos, In3 Inteligência de Mercado, Brazil</i>	
<i>Cecília C. Jamil, Stockholm University, Sweden</i>	

Category: Civil Engineering

Digital Animation for Representing Architectural Design	973
<i>Roberta Spallone, Politecnico di Torino, Italy</i>	

Literature Review of Augmented Reality Application in the Architecture, Engineering, and Construction Industry With Relation to Building Information.....	983
<i>Aydin Tabrizi, University of Kansas, USA</i>	
<i>Paola Sanguinetti, University of Kansas, USA</i>	

Category: Clinical Science and Technologies

A Gamification Update to the Taxonomy of Technology and Mental Health	995
<i>Madeline R. Marks, University of Central Florida, USA</i>	
<i>Amanda C. Tan, University of Central Florida, USA</i>	
<i>Clint Bowers, University of Central Florida, USA</i>	

Heart Sound Analysis for Blood Pressure Estimation	1006
<i>Rui Guedes, Faculdade de Medicina da Universidade do Porto, Portugal</i>	
<i>Henrique Cyrne Carvalho, Serviço de Cardiologia, Hospital de Santo António, Centro Hospitalar do Porto, Portugal</i>	
<i>Ana Castro, Universidade do Porto, Portugal</i>	

Sociological Perspectives on Improving Medical Diagnosis Emphasizing CAD.....	1017
<i>Joel Fisher, Department of State, United States Government, USA</i>	

Category: Cloud Computing

Cloud Computing.....	1026
<i>Eduardo Correia, Christchurch Polytechnic Institute of Technology (CPIT), New Zealand</i>	

Cloud Governance at the Local Communities	1033
<i>Vasileios Yfantis, Ionian University, Greece</i>	

Clouds of Quantum Machines	1040
<i>Nilo Sylvio Serpa, Universidade Paulista, Brazil</i>	

Cyberinfrastructure, Cloud Computing, Science Gateways, Visualization, and Cyberinfrastructure Ease of Use	1063
<i>Craig A. Stewart, Indiana University, USA</i>	
<i>Richard Knepper, Indiana University, USA</i>	
<i>Matthew R. Link, Indiana University, USA</i>	
<i>Marlon Pierce, Indiana University, USA</i>	
<i>Eric Wernert, Indiana University, USA</i>	
<i>Nancy Wilkins-Diehr, San Diego Supercomputer Center, USA</i>	
Fault Tolerant Cloud Systems	1075
<i>Sathish Kumar, VIT University, India</i>	
<i>Balamurugan B, VIT University, India</i>	
Fault Tolerant Data Management for Cloud Services.....	1091
<i>Wenbing Zhao, Cleveland State University, USA</i>	
From Information Systems Outsourcing to Cloud Computing	1101
<i>Mohammad Nabil Almunawar, Universiti Brunei Darussalam, Brunei</i>	
<i>Hasan Jawwad Almunawar, P. T. Tegar Kupas Mediatama, Indonesia</i>	
Improved Checkpoint Using the Effective Management of I/O in a Cloud Environment	1116
<i>Bakhta Meroufel, University of Oran1, Algeria</i>	
<i>Ghalem Belalem, University of Es Senia Oran1, Algeria</i>	
Service Quality and Perceived Value of Cloud Computing-Based Service Encounters.....	1129
<i>Eges Egedigwe, Dallas County Community College, USA</i>	
Understanding Business Models on the Cloud	1141
<i>Arash Najmaei, Australian Catholic University, Australia</i>	
Understanding Cloud Computing in a Higher Education Context	1153
<i>Lucy Self, University of Sussex, UK</i>	
<i>Petros Chamakiotis, University of Sussex, UK</i>	
Vertical Integration Between Providers With Possible Cloud Migration	1164
<i>Aleksandra Kostic-Ljubisavljevic, University of Belgrade, Serbia</i>	
<i>Branka Mikavica, University of Belgrade, Serbia</i>	
Virtualization as the Enabling Technology of Cloud Computing.....	1174
<i>Mohamed Fazil Mohamed Firdhous, University of Moratuwa, Sri Lanka</i>	

Category: Communications Theory

Communication, Information, and Pragmatics	1186
<i>Adriana Braga, Pontifical Catholic University of Rio de Janeiro, Brazil</i>	
<i>Robert K. Logan, University of Toronto, Canada</i>	

Data Journalism	1196
<i>Andreas A. Veglis, Aristotle University of Thessaloniki, Greece</i>	
<i>Charalampos P. Bratsas, Aristotle University of Thessaloniki, Greece</i>	

Investigating Diachronic Variation and Change in New Varieties of English	1206
<i>Rita Calabrese, University of Salerno, Italy</i>	

Negotiating Local Norms in Online Communication	1217
<i>Jonathan R. White, Höskolan Dalarna, Sweden</i>	

Category: Computer Engineering

Architecture of an Open-Source Real-Time Distributed Cyber Physical System	1227
<i>Stefano Scanzio, CNR-IEIIT, Italy</i>	

Consistency Is Not Enough in Byzantine Fault Tolerance	1238
<i>Wenbing Zhao, Cleveland State University, USA</i>	

Category: Computer Simulation

Data Visualization Strategies for Computer Simulation in Bioelectromagnetics	1249
<i>Akram Gasmelseed, Qassim University, Saudi Arabia</i>	
<i>Ali H. Alharbi, Qassim University, Saudi Arabia</i>	

Ergonomic Design of a Driver Training Simulator for Rural India	1260
<i>Prabir Mukhopadhyay, Indian Institute of Information Technology Design and Manufacturing Jabalpur, India</i>	
<i>Vipul Vinzuda, National Institute of Design Gandhinagar, India</i>	

3D Scanning and Simulation of a Hybrid Refrigerator Using Photovoltaic Energy	1277
<i>Edith Obregón Morales, Centro de Investigación y Desarrollo Tecnológico en Electroquímica, Mexico</i>	
<i>José de Jesús Pérez Bueno, Centro de Investigación y Desarrollo Tecnológico en Electroquímica, Mexico</i>	
<i>Juan Carlos Moctezuma Esparza, Universidad Politécnica Metropolitana de Hidalgo, Mexico</i>	
<i>Diego Marroquín García, Universidad Tecnológica de San Juan del Río, Mexico</i>	
<i>Arturo Trejo Pérez, Universidad Tecnológica de San Juan del Río, Mexico</i>	
<i>Roberto Carlos Flores Romero, Universidad Tecnológica de San Juan del Río, Mexico</i>	
<i>Juan Manuel Olivares Ramírez, Universidad Tecnológica de San Juan del Río, Mexico</i>	
<i>Maria Luisa Mendoza López, Tecnológico Nacional de México, Instituto Tecnológico de Querétaro, Mexico</i>	
<i>Juan Carlos Solís Ulloa, Instituto Tecnológico Superior de Cintalapa, Mexico</i>	
<i>Yunny Meas Vong, Centro de Investigación y Desarrollo Tecnológico en Electroquímica, Mexico</i>	
<i>Víctor Hugo Rodríguez Obregón, Centro de Investigación y Desarrollo Tecnológico en Electroquímica, Mexico</i>	

Uniform Random Number Generation With Jumping Facilities	1297
<i>E. Jack Chen, BASF Corporation, USA</i>	

Category: Computer Vision and Image Processing

A Critical Overview of Image Segmentation Techniques Based on Transition Region.....	1308
<i>Yu-Jin Zhang, Tsinghua University, China</i>	
Development of Image Engineering in the Last 20 Years	1319
<i>Yu-Jin Zhang, Tsinghua University, China</i>	
Particle Shape Analysis Using Digital Image Processing.....	1331
<i>Katia Tannous, University of Campinas, Brazil</i>	
<i>Fillipe Souza Silva, University of Campinas, Brazil</i>	
The Understanding of Spatial-Temporal Behaviors	1344
<i>Yu-Jin Zhang, Tsinghua University, China</i>	

Category: Criminal Science and Forensics

Forensic Investigations in Cloud Computing.....	1356
<i>Diane Barrett, Bloomsburg University of Pennsylvania, USA</i>	
Internet-Facilitated Child Sexual Exploitation Crimes.....	1366
<i>Keith F. Durkin, Ohio Northern University, USA</i>	
<i>Ronald L. DeLong, University of Dayton, USA</i>	
Knowledge-Based Forensic Patterns and Engineering System	1376
<i>Vivek Tiwari, International Institute of Information Technology Naya Raipur, India</i>	
<i>R. S. Thakur, Maulana Azad National Institute of Technology, India</i>	
Uncovering Limitations of E01 Self-Verifying Files	1384
<i>Jan Krasniewicz, Birmingham City University, UK</i>	
<i>Sharon A. Cox, Birmingham City University, UK</i>	

Category: Crisis Response and Management

Crisis Response and Management	1396
<i>Sergey V. Zykov, National Research University Higher School of Economics, Russia</i>	
Information Science and Technology in Crisis Response and Management.....	1407
<i>Randy Basham, University of Texas at Arlington, USA</i>	

Category: Curriculum Development and Instructional Design

Addressing Digital Competencies, Curriculum Development, and Instructional Design in Science Teacher Education.....	1420
<i>Isha DeCoito, Western University, Canada</i>	
Designing Engaging Instruction for the Adult Learners.....	1432
<i>Karen Swanson, Mercer University, USA</i> <i>Geri Collins, Mercer University, USA</i>	
Educational Ontology Development.....	1441
<i>Galip Kaya, Havelsan Inc., Turkey</i> <i>Arif Altun, Hacettepe University, Turkey</i>	
Factors Contributing to the Effectiveness of Online Students and Instructors	1451
<i>Michelle Kilburn, Southeast Missouri State University, USA</i> <i>Martha Henckell, Southeast Missouri State University, USA</i> <i>David Starrett, Columbia College, USA</i>	
Increasing Student Engagement and Participation Through Course Methodology	1463
<i>T. Ray Ruffin, University of Phoenix, USA & Grand Canyon University, USA & Ashford University, USA & North Carolina Wesleyan College, USA</i> <i>Donna Patterson Hawkins, University of Phoenix, USA</i> <i>D. Israel Lee, Southern Illinois University, USA & University of Phoenix, USA</i>	
Instructional Real World Community Engagement	1474
<i>Caroline M. Crawford, University of Houston – Clear Lake, USA</i>	
Learner Engagement in Blended Learning	1487
<i>Kristian J. Spring, Brigham Young University, USA</i> <i>Charles R. Graham, Brigham Young University, USA</i> <i>Tarah B. Ikahihifo, Brigham Young University, USA</i>	
Measuring Text Readability Using Reading Level	1499
<i>James C. Brewer, Texas Tech University, USA</i>	
Multimodal Literacy	1508
<i>Maryann Tatum Tobin, University of Miami, USA</i>	
An Open Learning Format for Lifelong Learners' Empowerment.....	1517
<i>Sabrina Leone, Università Politecnica delle Marche, Italy</i>	
Reflection as a Process From Theory to Practice.....	1529
<i>Sonia Bharwani, Indian School of Management and Entrepreneurship, India</i> <i>Durgamohan Musunuri, Bhavan's Usha and Lakshmi Mittal Institute of Management, India</i>	

Relationship Among Intelligence, Achievement Motivation, Type of School, and Academic Performance of Kenyan Urban Primary School Pupils	1540
<i>Jessina Mukomunene Muthee, Kenyatta University, Kenya</i>	
<i>Catherine G. Murungi, Kenyatta University, Kenya</i>	
Screencasts and Learning Styles	1548
<i>Rui Alberto Jesus, CESPU, Instituto de Investigação e Formação Avançada em Ciências e Tecnologias da Saúde, Portugal</i>	
Self-Awareness and Motivation Contrasting ESL and NEET Using the SAVE System	1559
<i>Laura Vettraino, Learning Community, Italy</i>	
<i>Valentina Castello, CIOFS FP, Italy</i>	
<i>Marco Guspini, educommunity – Educational Community, Italy</i>	
<i>Eleonora Guglielman, Learning Community, Italy</i>	

Category: Customer Relationship Management

Analysis of Two Phases Queue With Vacations and Breakdowns Under T-Policy	1570
<i>Khalid Alnowibet, King Saud University, Saudi Arabia</i>	
<i>Lotfi Tadj, Fairleigh Dickinson University, Canada</i>	
Customer Lifetime Value	1584
<i>Kijpokin Kasemsap, Suan Sunandha Rajabhat University, Thailand</i>	
Facilitating Customer Relationship Management in Modern Business.....	1594
<i>Kijpokin Kasemsap, Suan Sunandha Rajabhat University, Thailand</i>	
Implementing a Customer Relationship Management (CRM) System.....	1605
<i>Dimitra Skoumpopoulou, Northumbria University, UK</i>	
<i>Benjamin Franklin, Northumbria University, UK</i>	

Volume III (CU - ED)

The Intelligence of E-CRM Applications and Approaches on Online Shopping Industry.....	1616
<i>Bashar Shahir Ahmed, University Abdelmalek Essaadi (LEROSA Laborator), Morocco</i>	
<i>Fadi Amroush, Universidad de Salamanca, Spain</i>	
<i>Mohammed Ben Maati, University Abdelmalek Essaadi, Morocco</i>	
Optimizing Cloud Computing Costs of Services for Consumers	1627
<i>Eli Weintraub, Afeka Tel Aviv College of Engineering, Israel</i>	
<i>Yuval Cohen, Afeka Tel Aviv College of Engineering, Israel</i>	
Taxonomy for “Homo Consumens” in a 3.0 Era	1638
<i>Carlos Ballesteros, Universidad Pontificia Comillas, Spain</i>	

Category: Cyber and Network Security

Cyber Security Protection for Online Gaming Applications	1647
<i>Wenbing Zhao, Cleveland State University, USA</i>	
Piracy and Intellectual Property Theft in the Internet Era	1656
<i>Shun-Yung Kevin Wang, University of South Florida – St. Petersburg, USA</i>	
<i>Jeremy J McDaniel, Principal Financial Group, USA</i>	
Secure Group Key Sharing Protocols and Cloud System	1667
<i>Vaishali Ravindra Thakare, VIT University, India</i>	
<i>John Singh K, VIT University, India</i>	
Security of Internet-, Intranet-, and Computer-Based Examinations in Terms of Technical, Authentication, and Environmental, Where Are We?.....	1676
<i>Babak Sokouti, Tabriz University of Medical Sciences, Iran</i>	
<i>Massoud Sokouti, Mashhad University of Medical Sciences, Iran</i>	
A Three-Vector Approach to Blind Spots in Cybersecurity.....	1684
<i>Mika Westerlund, Carleton University, Canada</i>	
<i>Dan Craigen, Carleton University, Canada</i>	
<i>Tony Bailetti, Carleton University, Canada</i>	
<i>Uruemu Agwae, Carleton University, Canada</i>	

Category: Cyber Crime, Cyber Bullying, and Digital Terrorism

Cyber Bullying.....	1695
<i>Jo Ann Oravec, University of Wisconsin – Whitewater, USA</i>	
Cyberbullying Among Malaysian Children Based on Research Evidence.....	1704
<i>Sarina Yusuf, Universiti Putra Malaysia, Malaysia</i>	
<i>Md. Salleh Hj. Hassan, Universiti Putra Malaysia, Malaysia</i>	
<i>Adamkolo Mohammed Mohammed Ibrahim, Universiti Putra Malaysia, Malaysia & University of Maiduguri, Nigeria</i>	
The Nature, Extent, Causes, and Consequences of Cyberbullying.....	1723
<i>Michelle F. Wright, Masaryk University, Czech Republic</i>	

Category: Data Analysis and Statistics

Advanced Recommender Systems	1735
<i>Young Park, Bradley University, USA</i>	
Centrality Analysis of the United States Network Graph	1746
<i>Natarajan Meghanathan, Jackson State University, USA</i>	

Context-Aware Approach for Restaurant Recommender Systems.....	1757
<i>Haoxian Feng, University of Ottawa, Canada</i>	
<i>Thomas Tran, University of Ottawa, Canada</i>	
Data-Centric Benchmarking	1772
<i>Jérôme Darmont, Université de Lyon, Lyon 2, ERIC EA3083, France</i>	
Data Linkage Discovery Applications	1783
<i>Richard S. Segall, Arkansas State University, USA</i>	
<i>Shen Lu, University of South Florida, USA</i>	
Exploratory Data Analysis on Breast Cancer Prognosis.....	1794
<i>Mohammad Mehdi Owrang O., American University, USA</i>	
<i>Yasmine M. Kanaan, Howard University, USA</i>	
<i>Robert L. Copeland Jr., Howard University, USA</i>	
<i>Melvin Gaskins, Howard University Hospital, USA</i>	
<i>Robert L. DeWitty Jr., Providence Hospital, USA</i>	
In-Memory Analytics.....	1806
<i>Jorge Manjarrez Sánchez, Instituto Tecnológico Superior de Jerez, Mexico</i>	
Innovative Formalism for Biological Data Analysis.....	1814
<i>Calin Ciufudean, Stefan cel Mare University, Romania</i>	
Learning From Imbalanced Data	1825
<i>Lincy Mathews, M. S. Ramaiah Institute of Technology, India</i>	
<i>Seetha Hari, Vellore Institute of Technology, India</i>	
Neighborhood Rough-Sets-Based Spatial Data Analytics.....	1835
<i>Sharmila Banu K, VIT University, India</i>	
<i>B. K. Tripathy, VIT University, India</i>	
The Ontology of Randomness	1845
<i>Jeremy Horne, The International Institute of Informatics and Systemics, USA</i>	
Order Statistics and Applications.....	1856
<i>E. Jack Chen, BASF Corporation, USA</i>	
Recommender Technologies and Emerging Applications	1869
<i>Young Park, Bradley University, USA</i>	
Use of Data Analytics for Program Impact Evaluation and Enhancement of Faculty/Staff Development	1880
<i>Samuel Olugbenga King, Auburn University, USA</i>	

Category: Data Mining and Databases

Corporate Disclosure Measurement.....	1896
<i>Md. Salah Uddin Rajib, Jahangirnagar University, Bangladesh</i>	
<i>Md. Qutub Uddin Sajib, China University of Geosciences (CUG), China</i>	
Data Mining and Knowledge Discovery in Databases.....	1907
<i>Ana Azevedo, Polytechnic Institute of Porto, Portugal</i>	
Data Mining and the KDD Process.....	1919
<i>Ana Funes, Universidad Nacional de San Luis, Argentina</i>	
<i>Aristides Dasso, Universidad Nacional de San Luis, Argentina</i>	
Data Mining to Identify Project Management Strategies in Learning Environments.....	1934
<i>Ana González-Marcos, Universidad de La Rioja, Spain</i>	
<i>Joaquín Ordieres-Meré, Universidad Politécnica de Madrid, Spain</i>	
<i>Fernando Alba-Elías, Universidad de La Rioja, Spain</i>	
Database Techniques for New Hardware	1947
<i>Xiongpai Qin, Renmin University of China, China</i>	
<i>Yueguo Chen, Renmin University of China, China</i>	
Ensemble Clustering Data Mining and Databases.....	1962
<i>Slawomir T. Wierzchon, Polish Academy of Sciences, Poland</i>	
Graph-Based Concept Discovery	1974
<i>Alev Mutlu, Kocaeli University, Turkey</i>	
<i>Pinar Karagoz, Middle East Technical University, Turkey</i>	
<i>Yusuf Kavurucu, Turkish Naval Research Center Command, Turkey</i>	
Healthcare Data Analysis in the Internet of Things Era	1984
<i>George Tzanis, Aristotle University of Thessaloniki, Greece</i>	
N-Tuple Algebra as a Unifying System to Process Data and Knowledge.....	1995
<i>Boris Alexandrovich Kulik, Institute of Problems in Mechanical Engineering RAS, Russia</i>	
<i>Alexander Yakovlevich Fridman, Institute for Informatics and Mathematical Modelling, Kola Science Centre of RAS, Russia</i>	
A Proposed Framework for Incorporating Big-Data Technology in National Crisis Management Center.....	2006
<i>Magdy M. Kabeil, Al-Yamamah University, Saudi Arabia</i>	
<i>Ahmad M. Kabil, University of Wisconsin – Whitewater, USA</i>	
Quality Evaluation for Evolving Conceptual Database Design	2020
<i>Elvira Immacolata Locuratolo, CNR ISTI, Italy</i>	

Query Languages for Graph Databases	2031
<i>Kornelije Rabuzin, University of Zagreb, Croatia</i>	
Schema Evolution in Conventional and Emerging Databases	2043
<i>Zouhaier Brahmia, University of Sfax, Tunisia</i>	
<i>Fabio Grandi, University of Bologna, Italy</i>	
<i>Barbara Oliboni, University of Verona, Italy</i>	
<i>Rafik Bouaziz, University of Sfax, Tunisia</i>	
Schema Versioning in Conventional and Emerging Databases	2054
<i>Zouhaier Brahmia, University of Sfax, Tunisia</i>	
<i>Fabio Grandi, University of Bologna, Italy</i>	
<i>Barbara Oliboni, University of Verona, Italy</i>	
<i>Rafik Bouaziz, University of Sfax, Tunisia</i>	
Twitter Data Mining for Situational Awareness.....	2064
<i>Marco Vernier, University of Udine, Italy</i>	
<i>Manuela Farinosi, University of Udine, Italy</i>	
<i>Gian Luca Foresti, University of Udine, Italy</i>	

Category: Decision Support Systems

Cognitive Process Elements of People Decision-Making.....	2076
<i>Thais Spiegel, Rio de Janeiro State University, Brazil</i>	
Comprehensible Explanation of Predictive Models.....	2085
<i>Marko Robnik-Šikonja, University of Ljubljana, Slovenia</i>	
The Concept of the Shapley Value and the Cost Allocation Between Cooperating Participants	2095
<i>Alexander Kolker, GE Healthcare, USA</i>	
Decision Filed Theory	2108
<i>Lan Shao, University of Oulu, Finland</i>	
<i>Jouni Markkula, University of Oulu, Finland</i>	
Effectively Communicating With Group Decision Support Systems Using Information Theory....	2121
<i>Jamie S. Switzer, Colorado State University, USA</i>	
<i>Ralph V. Switzer, Colorado State University, USA</i>	
Evolutionary Algorithms for Global Decision Tree Induction	2132
<i>Marek Kretowski, Bialystok University of Technology, Poland</i>	
<i>Marcin Czajkowski, Bialystok University of Technology, Poland</i>	
A Family Review of Parameter-Learning Models and Algorithms for Making Actionable Decisions.....	2142
<i>Chun-Kit Ngan, The Pennsylvania State University, USA</i>	

Informed Decision Making With Enterprise Dynamic Systems Control	2154
<i>Sérgio Luís Guerreiro, Instituto Superior Técnico, University of Lisbon, Portugal & INESC-ID, Portugal</i>	
Managerial Tools and Techniques for Decision Making	2166
<i>Davood Askarany, University of Auckland, New Zealand</i>	
A Nature-Inspired Metaheuristic Approach for Generating Alternatives	2178
<i>Julian Scott Yeomans, York University, Canada</i>	
Preferences, Utility, and Stochastic Approximation	2188
<i>Yuri P. Pavlov, Bulgarian Academy of Sciences, Institute of Information and Communication Technologies, Bulgaria</i>	
<i>Rumen D. Andreev, Bulgarian Academy of Sciences, Institute of Information and Communication Technologies, Bulgaria</i>	
Rough-Set-Based Decision Model for Incomplete Information Systems	2200
<i>Safiye Turgay, Sakarya University, Turkey</i>	
<i>Orhan Torkul, Sakarya University, Turkey</i>	
<i>Tahsin Turgay, Sakarya University, Turkey</i>	
Using Receiver Operating Characteristic (ROC) Analysis to Evaluate Information -Based Decision-Making.....	2213
<i>Nan Hu, University of Utah, USA</i>	

Category: Digital Literacy

Digital Literacy	2225
<i>Anirban Ray, UNC Wilmington, USA</i>	
Digital Literacy for the 21st Century	2235
<i>Hiller A. Spires, North Carolina State University, USA</i>	
<i>Casey Medlock Paul, North Carolina State University, USA</i>	
<i>Shea N. Kerkhoff, North Carolina State University, USA</i>	
Digital Literacy in Theory and Practice.....	2243
<i>Heidi Julien, State University of New York at Buffalo, USA</i>	
Encouraging Digital Literacy and ICT Competency in the Information Age.....	2253
<i>Kijpokin Kasemsap, Suan Sunandha Rajabhat University, Thailand</i>	
Information Needs of Users in the Tech Savvy Environment and the Influencing Factors	2264
<i>Mudasir Khazer Rather, University of Kashmir, India</i>	
<i>Shabir Ahmad Ganaie, University of Kashmir, India</i>	

A Maturity Model for Digital Literacies and Sustainable Development	2280
<i>Ravi S. Sharma, Nanyang Technological University, Singapore</i>	
<i>Lin G. Malone, Nanyang Technological University, Singapore</i>	
<i>Chong Guan, SIM University, Singapore</i>	
<i>Ambica Dattakumar, Nanyang Technological University, Singapore</i>	
Teaching Media and Information Literacy in the 21st Century	2292
<i>Sarah Gretter, Michigan State University, USA</i>	
<i>Aman Yadav, Michigan State University, USA</i>	
Nigerian Undergraduate Students' Computer Competencies and Use of Information Technology Tools and Resources for Study Skills and Habits' Enhancement	2303
<i>Adekunle Olusola Otunla, University of Ibadan, Nigeria</i>	
<i>Caleb Okoro Amuda, University of Ibadan, Nigeria</i>	
The Roles of Digital Literacy in Social Life of Youth.....	2314
<i>Dragana Martinovic, University of Windsor, Canada</i>	
<i>Viktor Freiman, Université de Moncton, Canada</i>	
<i>Chrispina S. Lekule, St. Augustine University of Tanzania, Tanzania</i>	
<i>Yuqi Yang, University of Windsor, Canada</i>	
Toward a Working Definition of Digital Literacy.....	2326
<i>Margaret-Mary Sulentic Dowell, Louisiana State University, USA</i>	

Category: Economics

ICT Investments and Recovery of Troubled Economies	2337
<i>Ioannis Papadopoulos, Metropolitan College Thessaloniki, Greece</i>	
<i>Apostolos Syropoulos, Greek Molecular Computing Group, Greece</i>	
Uberization (or Uberification) of the Economy	2345
<i>Nabyla Daidj, Telecom Ecole de Management, France</i>	

Category: Educational Technologies

Adaptive Hypermedia in Education	2357
<i>Vehbi Turel, The University of Bingol, Turkey</i>	
Automatic Item Generation.....	2369
<i>Mark Gierl, University of Alberta, Canada</i>	
<i>Hollis Lai, University of Alberta, Canada</i>	
<i>Xinxin Zhang, University of Alberta, Canada</i>	
Challenges in Developing Adaptive Educational Hypermedia Systems	2380
<i>Eileen O'Donnell, Trinity College Dublin, Ireland</i>	
<i>Liam O'Donnell, Dublin Institute of Technology, Ireland</i>	

Computational Thinking in Innovative Computational Environments and Coding	2392
<i>Alberto Ferrari, University of Parma, Italy</i>	
<i>Agostino Poggi, University of Parma, Italy</i>	
<i>Michele Tomaiuolo, University of Parma., Italy</i>	
Computer Agent Technologies in Collaborative Learning and Assessment.....	2402
<i>Yigal Rosen, Harvard University, USA</i>	
Cost-Effective 3D Stereo Visualization for Creative Learning.....	2411
<i>R. S. Kamath, Chatrapati Shahu Institute of Business Education and Research, India</i>	
<i>R. K. Kamat, Shivaji University, India</i>	

Volume IV

(ED - F)

Could Educational Technology Replace Traditional Schools in the Future?	2421
<i>John K. Hope, University of Auckland, New Zealand</i>	
Development of Communication Skills through Auditory Training Software in Special Education	2431
<i>Eduardo C. Contreras, Autonomous University of Coahuila, Mexico</i>	
<i>Isis I. Contreras, Saltillo Institute of Technology, Mexico</i>	
Digital Storytelling in Language Classes.....	2442
<i>Mehrak Rahimi, Shahid Rajaei Teacher Training University, Iran</i>	
Distance Teaching and Learning Platforms	2455
<i>Linda D. Grooms, Regent University, USA</i>	
Do Usability Design Features of a Mobile Game Influence Learning?.....	2466
<i>Rex Perez Bringula, University of the East, Philippines</i>	
<i>Edison Cabrera, University of the East, Philippines</i>	
<i>Princess B. Calmerin, University of the East, Philippines</i>	
<i>Eduardo A. Lao, University of the East, Philippines</i>	
<i>Christian Gerard Sembrano, University of the East, Philippines</i>	
<i>Angelita D. Guia, University of the East, Philippines</i>	
<i>Joan P. Lazaro, University of the East, Philippines</i>	
<i>Alexis John M. Rubio, University of the East, Philippines</i>	
<i>Annaliza E. Catacutan, National University, Philippines</i>	
<i>Marilou N. Jamis, National University, Philippines</i>	
<i>Lalaine P. Abad, Department of Education, Philippines</i>	
Educational Technology and Intellectual Property	2477
<i>Lesley S. J. Farmer, California State University – Long Beach, USA</i>	

Employing Educational Robotics for the Development of Problem-Based Learning Skills	2492
<i>Nikleia Eteokleous, Frederick University Cyprus, Cyprus</i>	
From Digital Exclusion to Digital Inclusion for Adult Online Learners	2503
<i>Virginia E. Garland, University of New Hampshire, USA</i>	
From Digital Natives to Student Experiences With Technology	2512
<i>Sue Bennett, University of Wollongong, Australia</i>	
<i>Linda Corrin, University of Melbourne, Australia</i>	
ICT Eases Inclusion in Education.....	2521
<i>Dražena Gašpar, University of Mostar, Bosnia and Herzegovina</i>	
The Infusion of Technology Within the Classroom Facilitates Students' Autonomy in Their Learning	2532
<i>Fariel Mohan, University of Trinidad and Tobago, Trinidad and Tobago</i>	
<i>Garry Soomarah, University of Trinidad and Tobago, Trinidad and Tobago</i>	
Integrated Paper-Based and Digital Learning Material for Smart Learners	2545
<i>Sabrina Leone, Università Politecnica delle Marche, Italy</i>	
Leveraging Technology-Enhanced Teaching and Learning for Future IS Security Professionals....	2558
<i>Ciara Heavin, University College Cork, Ireland</i>	
<i>Karen Neville, University College Cork, Ireland</i>	
<i>Sheila O'Riordan, University College Cork, Ireland</i>	
Liberating Educational Technology Through the Socratic Method.....	2571
<i>Frank G. Giuseffi, Lindenwood University, USA</i>	
Online Academia	2580
<i>Magdalena Bielenia-Grajewska, University of Gdansk, Poland</i>	
Online Learning Propelled by Constructivism	2588
<i>Kathaleen Reid-Martinez, Oral Roberts University, USA</i>	
<i>Linda D. Grooms, Regent University, USA</i>	
Science Animation and Students' Attitudes.....	2599
<i>Sivasankar Arumugam, Sri Venkateswara College of Education, India</i>	
<i>Nancy Nirmala, Christ the King Matric Higher Secondary School, India</i>	
Three Cases of Unconventional Educational Uses of Digital Storytelling	2616
<i>Emmanuel Fokides, University of the Aegean, Greece</i>	
3D Printing Applications in STEM Education	2626
<i>Norman Gwangwava, Botswana International University of Science and Technology, Botswana</i>	
<i>Catherine Hlahla, National University of Science and Technology, Zimbabwe</i>	

Tools, Pedagogical Models, and Best Practices for Digital Storytelling	2641
<i>Jari Multisilta, Tampere University of Technology, Finland</i>	
<i>Hannele Niemi, University of Helsinki, Finland</i>	

The Use of Podcasting/Vodcasting in Education	2651
<i>Athanasios T. Stavrianos, 2nd Technical Vocational School of Xanthi, Greece</i>	
<i>Apostolos Syropoulos, Greek Molecular Computing Group, Greece</i>	

The Vital Importance of Faculty Presence in an Online Learning Environment	2661
<i>Ni Chang, Indiana University – South Bend, USA</i>	

Category: Electrical Engineering

Mechanisms of Electrical Conductivity in Carbon Nanotubes and Graphene.....	2673
<i>Rafael Vargas-Bernal, Instituto Tecnológico Superior de Irapuato, Mexico</i>	

Category: Electronic Business

E-Business and Big Data Strategy in Franchising	2686
<i>Ye-Sho Chen, Louisiana State University, USA</i>	

Facilitating Interaction Between Virtual Agents Through Negotiation Over Ontological Representation.....	2697
<i>Fiona McNeill, Heriot-Watt University, UK</i>	

On-Line Credit and Debit Card Processing and Fraud Prevention for E-Business	2707
<i>James Williams, University of Pittsburgh, USA</i>	

Category: Electronic Commerce

Adoption and Use of Mobile Money Services in Nigeria.....	2724
<i>Olayinka David-West, Pan-Atlantic University, Nigeria</i>	
<i>Immanuel Ovemeso Umukoro, Pan-Atlantic University, Nigeria</i>	
<i>Omotayo Muritala, Pan-Atlantic University, Nigeria</i>	

E-Commerce Models, Players, and Its Future	2739
<i>Liguo Yu, Indiana University – South Bend, USA</i>	

Electronic Payment Frameworks	2749
<i>Antonio Ruiz-Martínez, University of Murcia, Spain</i>	
<i>Oussama Tounekti, University of Murcia, Spain</i>	
<i>Antonio F. Skarmeta, University of Murcia, Spain</i>	

Factors Determining E-Shopping Compliance by Nigerians.....	2761
<i>Adamkolo Mohammed Mohammed Ibrahim, University of Maiduguri, Nigeria</i>	
<i>Md. Salleh Hj. Hassan, Universiti Putra Malaysia, Malaysia</i>	
<i>Sarina Yusuf, Universiti Putra Malaysia, Malaysia</i>	

Enterprise Interoperability	2773
<i>Ejub Kajan, State University of Novi Pazar, Serbia</i>	
Has Bitcoin Achieved the Characteristics of Money?	2784
<i>Donovan Peter Chan Wai Loon, University of Malaya, Malaysia</i>	
<i>Sameer Kumar, University of Malaya, Malaysia</i>	
The Importance of Electronic Commerce in Modern Business.....	2791
<i>Kijpokin Kasemsap, Suan Sunandha Rajabhat University, Thailand</i>	
Improving Competencies for the Courier Service Industry in Malaysia	2802
<i>Hoo Yee Hui, Universiti Tunku Abdul Rahman, Malaysia</i>	
<i>Yudi Fernando, Universiti Malaysia Pahang, Malaysia</i>	
New Advances in E-Commerce	2810
<i>Khaled Ahmed Nagaty, The British University in Egypt, Egypt</i>	
Online Mediation in E-Commerce Matters.....	2825
<i>Ángela Coello Pulido, University of Vigo, Spain</i>	
Reputational Mechanisms in Consumer-to-Consumer Online Commerce	2833
<i>Mikhail I. Melnik, Kennesaw State University, USA</i>	
Retail Prices and E-Commerce	2841
<i>Jihui Chen, Illinois State University, USA</i>	
Social Commerce Using Social Network and E-Commerce.....	2851
<i>Roberto Marmo, University of Pavia, Italy</i>	
An Update on Bitcoin as a Digital Currency	2861
<i>Cecilia G. Manrique, University of Wisconsin – La Crosse, USA</i>	
<i>Gabriel G. Manrique, Winona State University, USA</i>	
Use of Bitcoin for Internet Trade	2869
<i>Sadia Khalil, NUST School of Electrical Engineering and Computer Science, Pakistan</i>	
<i>Rahat Masood, NUST School of Electrical Engineering and Computer Science, Pakistan</i>	
<i>Muhammad Awais Shibli, VisionIT, USA</i>	

Category: Electronic Services

Determining Impact of Demographics on Perceived Service Quality in Online Retail	2882
<i>Prateek Kalia, I. K. Gujral Punjab Technical University, India</i>	
<i>Penny Law, Regenesys Business School, South Africa</i>	
<i>Richa Arora, Regenesys Institute of Management, India</i>	

The Impact of Carbon Nanotubes and Graphene on Electronics Industry	2897
<i>Rafael Vargas-Bernal, Instituto Tecnológico Superior de Irapuato, Mexico</i>	
<i>Gabriel Herrera-Pérez, Instituto Tecnológico Superior de Irapuato, Mexico</i>	
<i>Margarita Tecpoyotl-Torres, Universidad Autónoma del Estado de Morelos, Mexico</i>	
Integrating Content Authentication Support in Media Services	2908
<i>Anastasia N. Katsaounidou, Aristotle University of Thessaloniki, Greece</i>	
<i>Charalampos A. Dimoulas, Aristotle University of Thessaloniki, Greece</i>	
IT Service Management Architectures	2920
<i>Torben Tambo, Aarhus University, Denmark</i>	
<i>Jacob Filtenborg, Aarhus University, Denmark</i>	

Category: Enterprise Resource Planning

Business Intelligence Impacts on Design of Enterprise Systems	2932
<i>Saeed Rouhani, University of Tehran, Iran</i>	
<i>Dusanka Milorad Lecic, Levi9 Global Sourcing Balkan, Serbia</i>	
Deployment of Enterprise Architecture From the Activity Theory Perspective	2943
<i>Tiko Iyamu, Cape Peninsula University of Technology, South Africa</i>	
<i>Irja Naambo Shaanika, Namibia University of Science and Technology – Windhoek, Namibia</i>	
ERP Systems Benefit Realization and the Role of ERP-Enabled Application Integration.....	2953
<i>Joseph K. Nwankpa, Miami University, USA</i>	
From On-Premise ERP to Cloud ERP	2965
<i>Karim Mezghani, Al Imam Mohammad Ibn Saud Islamic University (IMSIU), Saudi Arabia</i>	
<i>& University of Sfax, Tunisia</i>	
Socio-Technical Change Perspective for ERP Implementation in Large Scale Organizations.....	2975
<i>Jessy Nair, PES University, India</i>	
<i>D. Bhanusree Reddy, VIT University, India</i>	
<i>Anand A. Samuel, VIT University, India</i>	

Category: Entrepreneurship

Current Scenario of Youth Entrepreneurship in India	2989
<i>Neeta Baporikar, Namibia University of Science and Technology, Namibia & University of Pune, India</i>	
Entrepreneurship	2998
<i>Mehmet Eymen Eryilmaz, Uludağ University, Turkey</i>	

Entrepreneurship as the Vantage Point	3009
<i>Madhu Kishore Raghunath Raghunath Kamakula, GVP College of Engineering (Autonomous), India</i>	
<i>Chandra Sekhar Patro, GVP College of Engineering (Autonomous), India</i>	
Entrepreneurship Concept, Theories, and New Approaches	3020
<i>José Manuel Saiz-Alvarez, Tecnológico de Monterrey, Mexico</i>	
<i>Martín García-Vaquero, Nebrija University, Spain</i>	
Open Data and High-Tech Startups Towards Nascent Entrepreneurship Strategies.....	3032
<i>Fotis Kitsios, University of Macedonia, Greece</i>	
<i>Maria Kamariotou, University of Macedonia, Greece</i>	

Category: Environmental Science and Agriculture

Carbon Capture From Natural Gas via Polymeric Membranes	3043
<i>Nayef Mohamed Ghasem, UAE University, UAE</i>	
<i>Nihmiya Abdul Rahim, UAE University, UAE</i>	
<i>Mohamed Al-Marzouqi, UAE University, UAE</i>	
Enhancing the Resiliency of Smart Grid Monitoring and Control	3056
<i>Wenbing Zhao, Cleveland State University, USA</i>	
E-Waste, Chemical Toxicity, and Legislation in India.....	3066
<i>Prashant Mehta, National Law University Jodhpur, India</i>	
Green IT and the Struggle for a Widespread Adoption	3077
<i>Edward T. Chen, University of Massachusetts – Lowell, USA</i>	
Identification of Green Procurement Drivers and Their Interrelationship Using Fuzzy TISM and MICMAC Analysis	3086
<i>Surajit Bag, Tega Industries South Africa Pty Ltd., South Africa</i>	
Load Flow Analysis in Smart Grids.....	3103
<i>Osman Hasan, National University of Sciences and Technology, Pakistan</i>	
<i>Awais Mahmood, National University of Sciences and Technology, Pakistan</i>	
<i>Syed Rafay Hasan, Tennessee Technological University, USA</i>	
Methodology of Climate Change Impact Assessment on Forests	3114
<i>Mostafa Jafari, Regional Institute of Forest and Rangelands (RIFR), Iran</i>	
Model for Assessment of Environmental Responsibility in Health Care Organizations.....	3131
<i>María Carmen Carnero, University of Castilla-La Mancha, Spain & University of Lisbon, Portugal</i>	

Potential Benefits and Current Limits in the Development of Demand Response	3144
<i>Clementina Bruno, University of Piemonte Orientale, Italy</i>	

Waste Gas End-of-Pipe Treatment Techniques in Italian IPPC Chemical Plants.....	3156
<i>Gaetano Battistella, ISPRA, Italy</i>	
<i>Giuseppe Di Marco, ISPRA, Italy</i>	
<i>Carlo Carlucci, ISPRA, Italy</i>	
<i>Raffaella Manuzzi, ISPRA, Italy</i>	
<i>Federica Bonaiuti, ISPRA, Italy</i>	
<i>Celine Ndong, ISPRA, Italy</i>	

Category: Ethics and Social Responsibility

The Foundation of (Business) Ethics' Evolution	3173
<i>Ben Tran, Alliant International University, USA</i>	

Integrating Sustainability and CSR in the Value Chain of the Information Technology Sector	3183
<i>Patricia Martínez García de Leaniz, University of Cantabria, Spain</i>	
<i>María Elena García Ruiz, University of Cantabria, Spain</i>	

The Morality of Reporting Safety Concerns in Aviation	3194
<i>Kawtar Tani, UCOL, New Zealand</i>	

Science, Ethics, and Weapons Research	3205
<i>John Forge, Independent Researcher, Australia</i>	

Category: Fuzzy Logic and Soft Computing

Application of Fuzzy Numbers to Assessment Processes	3215
<i>Michael Voskoglou, Graduate Technological Educational Institute (T.E.I.), Greece</i>	

Application of Soft Set in Game Theory	3226
<i>B. K. Tripathy, VIT University, India</i>	
<i>Sooraj T. R., VIT University, India</i>	
<i>Radhakrishna N. Mohanty, VIT University, India</i>	

Volume V

(G - HO)

Category: Gaming

Application of Gamification to Blended Learning in Higher Education	3238
<i>Kamini Jaipal-Jamani, Brock University, Canada</i>	
<i>Candace Figg, Brock University, Canada</i>	

Chemistry Learning Through Designing Digital Games.....	3248
<i>Kamisah Osman, The National University of Malaysia, Malaysia</i>	
<i>Ah-Nam Lay, Institute of Teacher Education – Sultan Abdul Halim, Malaysia</i>	
Clinical Use of Video Games.....	3260
<i>Ben Tran, Alliant International University, USA</i>	
Leveraging the Arduino Platform to Develop Information Technology Devices	3273
<i>Diego Reforgiato Recupero, University of Cagliari, Italy</i>	
<i>Valentino Artizzu, University of Cagliari, Italy</i>	
<i>Francesca Cella, University of Cagliari, Italy</i>	
<i>Alessandro Cotza, University of Cagliari, Italy</i>	
<i>Davide Curcio, University of Cagliari, Italy</i>	
<i>Giorgio Amedeo Iengo, University of Cagliari, Italy</i>	
<i>Riccardo Macis, University of Cagliari, Italy</i>	
<i>Andrea Marras, University of Cagliari, Italy</i>	
<i>Simone Picci, University of Cagliari, Italy</i>	
<i>Michael Planu, University of Cagliari, Italy</i>	
<i>Riccardo Scasseddu, University of Cagliari, Italy</i>	
Educational Serious Games Design.....	3287
<i>Ilias Karasavvidis, University of Thessaly, Greece</i>	
Exposure to Video Games and Decision Making	3296
<i>Giuseppe Curcio, University of L'Aquila, Italy</i>	
<i>Sara Peracchia, University of L'Aquila, Italy</i>	
Learning With Games and Digital Stories in Visual Programming	3309
<i>Wilfred W. F. Lau, The Chinese University of Hong Kong, China</i>	
The Process Model of Gameplay to Understand Digital Gaming Outcomes	3317
<i>Linda K. Kaye, Edge Hill University, UK</i>	
Serious Games Advancing the Technology of Engaging Information	3327
<i>Peter A. Smith, University of Central Florida, USA</i>	
<i>Clint Bowers, University of Central Florida, USA</i>	
Towards Modelling Effective Educational Games Using Multi-Domain Framework.....	3337
<i>Mifrah Ahmad, Universiti Teknologi PETRONAS, Malaysia</i>	
<i>Lukman Ab Rahim, Universiti Teknologi PETRONAS, Malaysia</i>	
<i>Kamisah Osman, Universiti Kebangsaan Malaysia, Malaysia</i>	
<i>Noreen Izza Arshad, Universiti Teknologi PETRONAS, Malaysia</i>	

Category: Gender and Diversity

- Computing Technologies and Science Fiction Cinema 3349
Rocío Carrasco-Carrasco, University of Huelva, Spain
- Gender Differences in Advertising Engagement Using the Case of Facebooks Ads 3359
Eva Lahuerta-Otero, University of Salamanca, Spain
Rebeca Cordero-Gutiérrez, University of Salamanca, Spain
- The Gender Dimension in Urban Air Quality..... 3371
Theodora Slini, Aristotle University of Thessaloniki, Greece
Fotini-Niovi Pavlidou, Aristotle University of Thessaloniki, Greece
- How Exclusive Work Climates Create Barriers for Women in IS&T..... 3382
Katelyn R. Reynoldson, Old Dominion University, USA
Debra A. Major, Old Dominion University, USA
- Women and IT in Lilongwe 3393
Alice Violet Nyamundundu, Skyway University, Malawi

Category: Geographic Information Systems

- Application of Geospatial Mashups in Web GIS for Tourism Development 3403
Somnath Chaudhuri, Maldives National University, Maldives
Nilanjan Ray, Adamas University, India
- Archaeological GIS for Land Use in South Etruria Urban Revolution in IX-VIII Centuries B.C.... 3419
Giuliano Pelfer, University of Florence, Italy
- Exploring Tourism Cluster in the Peripheral Mountain Area Based on GIS Mapping 3434
Ya-Hui Hsueh, National Taichung University of Education, Taiwan
Huey-Wen Chuang, National Taichung University of Education, Taiwan
Wan-Chiang Hsieh, National Taichung Girls' Senior High School, Taiwan
- Geographic Information System (GIS) Modeling Analysis and the Effects of Spatial Distribution
and Environmental Factors on Breast Cancer Incidence 3448
Akram Gasmelseed, University of Science and Technology, Sudan
Ali H. Alharbi, Qassim University, Saudi Arabia
- Geographic Information Systems..... 3460
Paula Remoaldo, University of Minho, Portugal
Vitor P. Ribeiro, University of Minho, Portugal
Hélder Silva Lopes, University of Minho, Portugal
Sara Catarina Gomes Silva, University of Minho, Portugal

Geospatial Influence in Science Mapping.....	3473
<i>Carlos Granell-Canut, Universitat Jaume I of Castellón, Spain</i>	
<i>Estefanía Aguilar-Moreno, Universitat Jaume I of Castellón, Spain</i>	
Parallel Development of Three Major Space Technology Systems and Human Side of Information Reference Services as an Essential Complementary Method	3484
<i>Joyce Gosata Maphanyane, University of Botswana, Botswana</i>	
Use of GIS and Remote Sensing for Landslide Susceptibility Mapping	3503
<i>Arzu Erener, Kocaeli University, Turkey</i>	
<i>Gulcan Sarp, Suleyman Demirel University, Turkey</i>	
<i>Sebnem Duzgun, Middle East Technical University, Turkey</i>	

Category: Government and Law

Accessibility in E-Government.....	3516
<i>Christian Sonnenberg, Florida Institute of Technology, USA</i>	
The Adoption and Transformation of Capability Maturity Models in Government.....	3526
<i>Terry F. Buss, Carnegie Mellon University, Australia</i>	
Bridging Between Cyber Politics and Collective Dynamics of Social Movement.....	3538
<i>Kazuhiko Shibuya, ROIS, Japan</i>	
Community Broadband Networks and the Opportunity for E-Government Services	3549
<i>Idongesit Williams, Aalborg University, Denmark</i>	
Critical Success Factors in E-Democracy Implementation	3561
<i>Aderonke A. Oni, Covenant University, Nigeria</i>	
<i>Adekunle O. Okunoye, Xavier University, USA</i>	
E-Activism Development and Growth.....	3569
<i>John G. McNutt, University of Delaware, USA</i>	
<i>Lauri Goldkind, Fordham University, USA</i>	
E-Government Service Adoption and the Impact of Privacy and Trust	3579
<i>Mehree Iqbal, North South University, Bangladesh</i>	
<i>Nabila Nisha, North South University, Bangladesh</i>	
<i>Afrin Rifat, North South University, Bangladesh</i>	
Mastering Electronic Government in the Digital Age	3591
<i>Kijpokin Kasemsap, Suan Sunandha Rajabhat University, Thailand</i>	

A Model for Connected E-Government in the Digital Age	3602
<i>Qiuyan Fan, Western Sydney University, Australia</i>	
Presidential Elections Web 2.0	3612
<i>Ramona Sue McNeal, University of Northern Iowa, USA</i>	
<i>Lisa Dotterweich Bryan, Upper Iowa University, USA</i>	
Project Management in Government	3621
<i>Shauneen Furlong, University of Ottawa, Canada & John Moores Liverpool University, UK</i>	
Technology and Terror.....	3637
<i>Maximiliano Emanuel Korstanje, University of Palermo, Argentina</i>	
<i>Geoffrey Skoll, SUNY at Buffalo, USA</i>	
Users Behavioral Intention Towards eGovernment in an African Developing Country	3654
<i>Ayankunle A. Taiwo, Texas A&M University – Commerce, USA</i>	
Young People, Civic Participation, and the Internet.....	3667
<i>Fadi Hirzalla, Erasmus University Rotterdam, The Netherlands</i>	
<i>Shakuntala Banaji, LSE, UK</i>	

Category: Healthcare Administration

Electronic Health Record (EHR) Diffusion and an Examination of Physician Resistance	3678
<i>Kristen MacIver, Northern Michigan University, USA</i>	
<i>Madison N. Ngafeeson, Northern Michigan University, USA</i>	
Internet of Things Applications for Healthcare	3689
<i>Ljubica Diković, Business Technical College, Serbia</i>	
Maintenance Policies Optimization of Medical Equipment in a Health Care Organization	3698
<i>Juan Ignacio Roig, University of Castilla-La Mancha, Spain</i>	
<i>Andrés Gómez, University of Castilla-La Mancha, Spain</i>	
<i>Isabel Romero, University of Castilla-La Mancha, Spain</i>	
<i>María Carmen Carnero, University of Castilla-La Mancha, Spain & University of Lisbon, Portugal</i>	
The Optimal Workforce Staffing Solutions With Random Patient Demand in Healthcare Settings.....	3711
<i>Alexander Kolker, GE Healthcare, USA</i>	
Using Technology to Reduce a Healthcare Disparity	3725
<i>Nilmini Wickramasinghe, Epworth HealthCare, Australia & Deakin University, Australia</i>	

Category: Health Information Systems

Challenges and Implications of Health Literacy in Global Health Care	3734
<i>Kijpokin Kasemsap, Suan Sunandha Rajabhat University, Thailand</i>	
Cyber Behaviors in Seeking Health Information	3745
<i>Xiaojun (Jenny) Yuan, University of Albany, USA</i>	
<i>José A. Pino, Universidad de Chile, Chile</i>	
Information Systems and Technology Projects in Healthcare Organisations	3756
<i>Jorge Gomes, ISEG, Universidade de Lisboa, Portugal</i>	
<i>Mário José Batista Romão, ISEG, Universidade de Lisboa, Portugal</i>	
Peer-to-Peer Health-Related Online Support Groups	3767
<i>Neil S. Coulson, University of Nottingham, UK</i>	
Software Evaluation From the Perspective of Patients and Healthcare Professionals	3782
<i>Rui Pedro Charters Lopes Rijo, Polytechnic Institute of Leiria, Portugal</i>	
<i>Domingos Alves, University of São Paulo, Brazil</i>	
Technology Design and Routes for Tool Appropriation in Medical Practices	3794
<i>Manuel Santos-Trigo, Cinvestav-IPN, Mexico</i>	
<i>Ernesto Suaste, Cinvestav-IPN, Mexico</i>	
<i>Paola Figuerola, Cinvestav-IPN, Mexico</i>	
Trends in Health Care Information Technology and Informatics	3805
<i>T. Ray Ruffin, University of Phoenix, USA & Colorado Technical University, USA & Grand Canyon University, USA & Ashford University, USA, & North Carolina Wesleyan College, USA</i>	
<i>Donna Patterson Hawkins, University of Phoenix, USA</i>	
User Resistance to Health Information Technology	3816
<i>Madison N. Ngafeeson, Northern Michigan University, USA</i>	

Category: Higher Education

The Effect of Innovative Communication Technologies in Higher Education	3827
<i>Stavros Kiriakidis, University of Crete, Greece</i>	
<i>Efstathios Kefallonitis, State University of New York at Oswego, USA</i>	
<i>Androniki Kavoura, Technological Educational Institute of Athens, Greece</i>	
Experiences of Implementing a Large-Scale Blended, Flipped Learning Project.....	3839
<i>Hazel Owen, Ethos Consultancy NZ, New Zealand</i>	
<i>Nicola Dunham, Massey University, New Zealand</i>	

A Flipped Learning Approach to University EFL Courses	3850
<i>Yasushige Ishikawa, Kyoto University of Foreign Studies, Japan</i>	
<i>Reiko Akahane-Yamada, Advanced Telecommunications Research Institute International (ATR), Japan</i>	
<i>Craig Smith, Kyoto University of Foreign Studies, Japan</i>	
<i>Masayuki Murakami, Kyoto University of Foreign Studies, Japan</i>	
<i>Mutsumi Kondo, Kyoto University of Foreign Studies, Japan</i>	
<i>Misato Kitamura, Advanced Telecommunications Research Institute International (ATR), Japan</i>	
<i>Yasushi Tsubota, Kyoto Institute of Technology, Japan</i>	
<i>Masatake Dantsuji, Kyoto University, Japan</i>	
A Framework for Profiling Prospective Students in Higher Education	3861
<i>Santhosh Kumar Lakkaraju, Dakota State University, USA</i>	
<i>Deb Tech, Dakota State University, USA</i>	
<i>Shuyuan Deng, Dakota State University, USA</i>	
Importance of Information Literacy	3870
<i>Lidia Sanchez-Ruiz, University of Cantabria, Spain</i>	
<i>Beatriz Blanco, University of Cantabria, Spain</i>	
An Integrated Electronic IQA System for HEI	3881
<i>Teay Shawyun, King Saud University, Saudi Arabia</i>	
International Students in Online Courses	3900
<i>María Ángeles Rodríguez Manzanares, Memorial University of Newfoundland, Canada</i>	
IT Solutions Supporting the Management of Higher Education Institutions in Poland	3910
<i>Elżbieta Janczyk-Strzała, Wrocław School of Banking, Poland</i>	
Knowledge Networks in Higher Education.....	3922
<i>Filipa M. Ribeiro, University of Porto, Portugal</i>	
Quality Online Learning in Higher Education.....	3930
<i>Deborah G. Wooldridge, Bowling Green State University, USA</i>	
<i>Sandra Poirier, Middle Tennessee State University, USA</i>	
<i>Julia M. Matuga, Bowling Green State University, USA</i>	
A Study on Extensive Reading in Higher Education	3945
<i>Diana Presadă, Petroleum-Gas University of Ploiesti, Romania</i>	
<i>Mihaela Badea, Petroleum-Gas University of Ploiesti, Romania</i>	
Technology Policies and Practices in Higher Education	3954
<i>Kelly McKenna, Colorado State University, USA</i>	

The University-Industry Collaboration	3963
<i>Marcello Fernandes Chedid, University of Aveiro, Portugal</i>	
<i>Leonor Teixeira, University of Aveiro, Portugal</i>	

Using Communities of Inquiry Online to Perform Tasks of Higher Order Learning.....	3976
<i>Ramon Tirado-Morueta, University of Huelva, Spain</i>	
<i>Pablo Maraver-López, University of Huelva, Spain</i>	
<i>Ángel Hernando-Gómez, University of Huelva, Spain</i>	

Category: High Performance Computing

Cost Evaluation of Synchronization Algorithms for Multicore Architectures	3989
<i>Masoud Hemmatpour, Politecnico di Torino, Italy</i>	
<i>Renato Ferrero, Politecnico di Torino, Italy</i>	
<i>Filippo Gandino, Politecnico di Torino, Italy</i>	
<i>Bartolomeo Montrucchio, Politecnico di Torino, Italy</i>	
<i>Maurizio Rebaudengo, Politecnico di Torino, Italy</i>	

The Future of High-Performance Computing (HPC)	4004
<i>Herbert Cornelius, Intel Corporation EMEA, Germany</i>	

High-Performance Reconfigurable Computing	4018
<i>Mário Pereira Vestias, Instituto Politécnico de Lisboa, Portugal</i>	

Category: Hospitality, Travel, and Tourism Management

Augmented Reality for Tourist Destination Image Formation	4031
<i>Azizul Hassan, Cardiff Metropolitan University, UK</i>	

Volume VI (HO - IT)

Destination @-Branding of Ten European Capitals Through the Institutional Stems and Commercial Logos.....	4038
<i>Elena Bocci, Sapienza University of Rome, Italy</i>	
<i>Annamaria Silvana de Rosa, Sapienza University of Rome, Italy</i>	
<i>Laura Dryjanska, Sapienza University of Rome, Italy</i>	

The Effect of Social Media Networking in the Travel Industry	4052
<i>Androniki Kavoura, Technological Educational Institute of Athens, Greece</i>	
<i>Efstathios Kefallonitis, State University of New York at Oswego, USA</i>	

Evaluative Dimensions of Urban Tourism in Capital Cities by First-Time Visitors	4064
<i>Annamaria Silvana de Rosa, Sapienza University of Rome, Italy</i>	
<i>Laura Dryjanska, Sapienza University of Rome, Italy</i>	
<i>Elena Bocci, Sapienza University of Rome, Italy</i>	

Fifty Shades of Dark Stories	4077
<i>Lea Kuznik, University of Maribor, Slovenia</i>	
Social Media as a Channel of Constructive Dialogue for Tourism Businesses	4088
<i>Marios D. Sotiriadis, University of South Africa (UNISA), South Africa</i>	
Usability Evaluation of Tourism Icons in India.....	4099
<i>Rajshree Tushar Akolkar, Zeal College of Engineering and Research, India</i>	
<i>Ganesh D. Bhutkar, Vishwakarma Institute of Technology, India</i>	
Virtual Tourism and Its Potential for Tourism Development in Sub-Saharan Africa.....	4113
<i>Paul Ankomah, North Carolina A&T State University, USA</i>	
<i>Trent Larson, North Carolina A&T University, USA</i>	

Category: Human-Computer Interaction

Affect-Sensitive Computer Systems	4124
<i>Nik Thompson, Curtin University, Australia</i>	
<i>Tanya McGill, Murdoch University, Australia</i>	
<i>David Murray, Murdoch University, Australia</i>	
Computer-Assisted Indian Matrimonial Services	4136
<i>Robert Leslie Fisher, Independent Researcher, USA</i>	
Creative Collaborative Virtual Environments.....	4146
<i>Luís Eustáquio, Universidade do Porto, Portugal</i>	
<i>Catarina Carneiro de Sousa, Polytechnic Institute of Viseu, Portugal</i>	
Cyberbullying	4157
<i>Gilberto Marzano, Rezekne Academy of Technologies, Latvia</i>	
Defining and Conceptualizing Cyberbullying	4168
<i>Karin Spenser, Nottingham Trent University, UK</i>	
<i>Lucy R. Betts, Nottingham Trent University, UK</i>	
Developing Creativity and Learning Design by Information and Communication Technology (ICT) in Developing Contexts.....	4178
<i>Chunfang Zhou, Aalborg University, Denmark</i>	
<i>Aparna Purushothaman, Aalborg University, Denmark</i>	
Existential Aspects of the Development E-Culture	4189
<i>Liudmila Vladimirovna Baeva, Astrakhan State University, Russia</i>	
The Fundamentals of Human-Computer Interaction	4199
<i>Kijpokin Kasemsap, Suan Sunandha Rajabhat University, Thailand</i>	

Interface Trends in Human Interaction, the Internet of Things, and Big Data	4210
<i>William J. Gibbs, Duquesne University, USA</i>	
Internet Addiction in Context	4223
<i>Petra Vondrackova, Charles University in Prague, Czech Republic</i>	
<i>David Šmahel, Masaryk University – Brno, Czech Republic</i>	
Mediated Embodiment in New Communication Technologies	4234
<i>Laura Aymerich-Franch, CNRS-AIST JRL (Joint Robotics Laboratory), AIST, Japan</i>	
The Nature of Cyber Bullying Behaviours	4245
<i>Lucy R. Betts, Nottingham Trent University, UK</i>	
Screen Culture.....	4255
<i>Ana Melro, University of Aveiro, Portugal</i>	
<i>Lidia Oliveira, University of Aveiro, Portugal</i>	
Technology Assessment of Information and Communication Technologies	4267
<i>Armin Grunwald, Karlsruhe Institute of Technology, Germany</i>	
<i>Carsten Orwat, Karlsruhe Institute of Technology, Germany</i>	
Towards an Interdisciplinary Socio-Technical Definition of Virtual Communities	4278
<i>Umar Ruhi, University of Ottawa, Canada</i>	
The Trajectivity of Virtual Worlds.....	4296
<i>Christophe Duret, Université de Sherbrooke, Canada</i>	
Virtual Hoarding	4306
<i>Jo Ann Oravec, University of Wisconsin – Whitewater, USA</i>	

Category: Human Resources Management

Cyberloafing and Constructive Recreation	4316
<i>Jo Ann Oravec, University of Wisconsin – Whitewater, USA</i>	
Influencing People and Technology Using Human Resource Development (HRD) Philosophy	4326
<i>Claretha Hughes, University of Arkansas, USA</i>	
<i>Matthew W. Gosney, University of Colorado – Health, USA</i>	
<i>Cynthia M. Sims, Clemson University, USA</i>	
Performance Appraisal.....	4337
<i>Chandra Sekhar Patro, Gayatri Vidya Parishad College of Engineering (Autonomous), India</i>	
Promoting Strategic Human Resource Management, Organizational Learning, and Knowledge Management in Modern Organizations.....	4347
<i>Kijpokin Kasemsap, Suan Sunandha Rajabhat University, Thailand</i>	

Technology, Learning Styles, Values, and Work Ethics of Millennials	4358
<i>Harish C. Chandan, Argosy University, USA</i>	

Category: Industrial Engineering and Informatics

Cuckoo Search Algorithm for Solving Real Industrial Multi-Objective Scheduling Problems	4369
<i>Mariappan Kadarkarainadar Marichelvam, Mepco Schlenk Engineering College, India</i>	
<i>Mariappan Geetha, Kamaraj College of Engineering and Technology, India</i>	
The Trends and Challenges of 3D Printing.....	4382
<i>Edna Ho Chu Fang, University of Malaya, Malaysia</i>	
<i>Sameer Kumar, University of Malaya, Malaysia</i>	

Category: Information Resources Management

Advanced Model of Complex Information System	4391
<i>Miroslav Svitek, Czech Technical University in Prague, Czech Republic</i>	
Computer Information Library Clusters	4399
<i>Fu Yuhua, CNOOC Research Institute, China</i>	
The Impact of the Impact of Meta-Data Mining From the SoReCom “A.S. de Rosa” @-Library...	4404
<i>Annamaria Silvana de Rosa, Sapienza University of Rome, Italy</i>	
<i>Laura Dryjanska, Sapienza University of Rome, Italy</i>	
<i>Elena Bocci, Sapienza University of Rome, Italy</i>	
Information and Its Conceptual Perspectives.....	4422
<i>José Poças Rascão, Institute Polytechnic of Setúbal, Portugal</i>	
Open Data Repositories in Knowledge Society	4436
<i>Nadim Akhtar Khan, University of Kashmir, India</i>	
<i>Sara Sohrabzadeh, Tehran University of Medical Science, Iran</i>	
<i>Garvita Jhamb, University of Delhi, India</i>	
Quantum Information Science Vis-à-Vis Information Schools	4448
<i>P. K. Paul, Raiganj University, India</i>	
<i>D. Chatterjee, Seacom Skills University, India</i>	
<i>A. Bhumali, Raiganj University, India</i>	
Towards a General Theory of Information	4459
<i>Laura L. Pană, Polytechnic University of Bucharest, Romania</i>	

Category: Information Retrieval

Analysis and Assessment of Cross-Language Question Answering Systems	4471
<i>Juncal Gutiérrez-Artacho, University of Granada, Spain</i>	
<i>María-Dolores Olvera-Lobo, University of Granada, Spain</i>	

Challenges in Collecting Qualitative Data for Information Systems Studies	4480
<i>Tiko Iyamu, Cape Peninsula University of Technology, South Africa</i>	
<i>Irja Naambo Shaanika, Namibia University of Science and Technology, Namibia</i>	
Cognitive and Psychological Factors in Cross-Language Information Retrieval	4490
<i>Rowena Li, Bayside High School Library, USA</i>	
A Fast and Space-Economical Algorithm for the Tree Inclusion Problem	4502
<i>Yangjun Chen, University of Winnipeg, Canada</i>	
<i>Yibin Chen, University of Winnipeg, Canada</i>	
Information Seeking Models in the Digital Age	4515
<i>Mudasir Khazer Rather, University of Kashmir, India</i>	
<i>Shabir Ahmad Ganaie, University of Kashmir, India</i>	
An Insight Into Deep Learning Architectures	4528
<i>Nishu Garg, VIT University, India</i>	
<i>Nikhitha P, VIT University, India</i>	
<i>B. K. Tripathy, VIT University, India</i>	
Online Information Retrieval Systems Trending From Evolutionary to Revolutionary Approach	4535
<i>Zahid Ashraf Wani, University of Kashmir, India</i>	
<i>Huma Shafiq, University of Kashmir, India</i>	

Category: IT Research and Theory

Adaptive Networks for On-Chip Communication	4549
<i>Mário Pereira Vestias, Instituto Politécnico de Lisboa, Portugal</i>	
Business Model Innovation-Oriented Technology Management for Emergent Technologies	4560
<i>Sven Seidenstricker, Fraunhofer Institute for Industrial Engineering, Germany</i>	
<i>Ardilio Antonino, Fraunhofer Institute for Industrial Engineering, Germany</i>	
Cognitive Mapping in Support of Intelligent Information Systems	4570
<i>Akbar Esfahanipour, Amirkabir University of Technology, Iran</i>	
<i>Ali Reza Montazemi, McMaster University, Canada</i>	
Computer-Assisted Parallel Program Generation	4583
<i>Shigeo Kawata, Utsunomiya University, Japan</i>	
Constrained Nonlinear Optimization in Information Science	4594
<i>William P. Fox, Naval Postgraduate School, USA</i>	
Decimal Hardware Multiplier	4607
<i>Mário Pereira Vestias, INESC-ID/ISEL/IPL, Portugal</i>	

Digital Divide	4619
<i>Patrick Flanagan, St. John's University, USA</i>	
Eight Tips for the Theme, "Data and Forecasts"	4629
<i>Alessio Drivet, GeoGebra Institute of Torino, Italy</i>	
Exploring New Handwriting Parameters for Writer Identification.....	4643
<i>Verónica Inés Aubin, Universidad Nacional de La Matanza, Argentina</i>	
<i>Jorge Horacio Doorn, Universidad Nacional del Oeste, Argentina & Universidad Nacional de La Matanza, Argentina</i>	
Haptics-Based Systems Characteristics, Classification, and Applications	4652
<i>Abeer Bayousuf, King Saud University, Saudi Arabia</i>	
<i>Hend S. Al-Khalifa, King Saud University, Saudi Arabia</i>	
<i>Abdulmalik Al-Salman, King Saud University, Saudi Arabia</i>	
The Holon/Parton Structure of the Meme, or The Unit of Culture.....	4666
<i>J. T. Velikovsky, University of Newcastle, Australia</i>	
ICT Standardization	4679
<i>Kai Jakobs, RWTH Aachen University, Germany</i>	
Immersing People in Scientific Knowledge and Technological Innovation Through Disney's Use of Installation Art.....	4692
<i>Jonathan Lillie, Loyola University Maryland, USA</i>	
<i>Michelle Jones-Lillie, Lillie Pad Studios, USA</i>	
Indicators of Information and Communication Technology.....	4704
<i>Gulnara Abdrakhmanova, National Research University Higher School of Economics, Russia</i>	
<i>Leonid Gokhberg, National Research University Higher School of Economics, Russia</i>	
<i>Alexander Sokolov, National Research University Higher School of Economics, Russia</i>	
Information Technologies and Social Change	4715
<i>Muhammet Ali Köroğlu, Uşak University, Turkey</i>	
<i>Cemile Zehra Köroğlu, Uşak University, Turkey</i>	
iSchools Promoting "Information Science and Technology" (IST) Domain Towards Community, Business, and Society With Contemporary Worldwide Trend and Emerging Potentialities in India	4723
<i>P. K. Paul, Raiganj University, India</i>	
<i>D. Chatterjee, Seacom Skills University, India</i>	
Logic Programming for Intelligent Systems	4736
<i>James D. Jones, Liberty University, USA</i>	

Methods for Improving Alias Rejections in Comb Filters.....	4746
<i>Gordana Jovanovic Dolecek, Institute INAOE Puebla, Mexico</i>	
A Paradoxical World and the Role of Technology in Thana-Capitalism.....	4761
<i>Maximiliano Emanuel Korstanje, University of Palermo, Argentina</i>	
Performance Measurement of Technology Ventures by Science and Technology Institutions	4774
<i>Artie W. Ng, The Hong Kong Polytechnic University, Hong Kong</i>	
<i>Benny C. F. Cheung, The Hong Kong Polytechnic University, Hong Kong</i>	
<i>Peggy M. L. Ng, The Hong Kong Polytechnic University, Hong Kong</i>	
The Skills of European ICT Specialists	4785
<i>Francesca Sgobbi, University of Brescia, Italy</i>	
A Trust Case-Based Model Applied to Agents Collaboration.....	4797
<i>Felipe Boff, Lutheran University of Brazil (ULBRA), Brazil</i>	
<i>Fabiana Lorenzi, Lutheran University of Brazil (ULBRA), Brazil</i>	
Understanding and Assessing Quality of Models and Modeling Languages	4810
<i>John Krogstie, Norwegian University of Science and Technology, Norway</i>	
Utilizing Information Science and Technology in Franchise Organizations	4822
<i>Ye-Sho Chen, Louisiana State University, USA</i>	

Category: IT Security and Ethics

Computer Fraud Challenges and Its Legal Implications.....	4837
<i>Amber A. Smith-Ditizio, Texas Woman's University, USA</i>	
<i>Alan D. Smith, Robert Morris University, USA</i>	

Volume VII (IT - MA)

Cost Estimation and Security Investment of Security Projects.....	4849
<i>Yosra Miaoui, University of Carthage, Tunisia</i>	
<i>Boutheina Fessi, University of Carthage, Tunisia</i>	
<i>Noureddine Boudriga, University of Carthage, Tunisia</i>	
Development of Personal Information Privacy Concerns Evaluation	4862
<i>Anna Rohunen, University of Oulu, Finland</i>	
<i>Jouni Markkula, University of Oulu, Finland</i>	
Digital Video Watermarking Using Diverse Watermarking Schemes.....	4872
<i>Yash Gupta, Maulana Abul Kalam Azad University of Technology, India</i>	
<i>Shaila Agrawal, Maulana Abul Kalam Azad University of Technology, India</i>	
<i>Susmit Sengupta, Maulana Abul Kalam Azad University of Technology, India</i>	
<i>Aruna Chakraborty, Maulana Abul Kalam Azad University of Technology, India</i>	

Ethical Computing Continues From Problem to Solution	4884
<i>Wanbil William Lee, The Computer Ethics Society, Hong Kong & Wanbil & Associates, Hong Kong</i>	
Group Signature System Using Multivariate Asymmetric Cryptography	4898
<i>Sattar J. Aboud, University of Bedfordshire, UK</i>	
Hexa-Dimension Code of Practice for Data Privacy Protection	4909
<i>Wanbil William Lee, Wanbil & Associates, Hong Kong</i>	
Information and Communication Technology Ethics and Social Responsibility	4920
<i>Tomas Cahlik, Charles University Prague, Czech Republic & University of Economics Prague, Czech Republic</i>	
Intrusion Tolerance Techniques	4927
<i>Wenbing Zhao, Cleveland State University, USA</i>	
New Perspectives of Pattern Recognition for Automatic Credit Card Fraud Detection	4937
<i>Addisson Salazar, Universitat Politècnica de València, Spain</i>	
<i>Gonzalo Safont, Universitat Politècnica de València, Spain</i>	
<i>Alberto Rodriguez, Universidad Miguel Hernández de Elche, Spain</i>	
<i>Luis Vergara, Universitat Politècnica de València, Spain</i>	
Privacy, Algorithmic Discrimination, and the Internet of Things	4951
<i>Jenifer Sunrise Winter, University of Hawaii at Manoa, USA</i>	
The Protection Policy for Youth Online in Japan	4962
<i>Nagayuki Saito, Ochanomizu University, Japan</i>	
<i>Madoka Aragaki, Business Breakthrough University, Japan</i>	
Security of Identity-Based Encryption Algorithms	4975
<i>Kannan Balasubramanian, Mepco Schlenk Engineering College, India</i>	
<i>M. Rajakani, Mepco Schlenk Engineering College, India</i>	
Steganography Using Biometrics.....	4985
<i>Manashee Kalita, NERIST, India</i>	
<i>Swanirbhar Majumder, NERIST, India</i>	
Usable Security	5004
<i>Andrea Atzeni, Politecnico di Torino, Italy</i>	
<i>Shamal Faily, Bournemouth University, UK</i>	
<i>Ruggero Galloni, Square Reply S.r.l., Italy</i>	

Category: Knowledge Management

Boosting the Social Development of the Majority Through the Creation of a Wireless Knowledge Society.....	5015
<i>Danilo Piaggese, Framericas, USA</i>	

Communities of Practice as a Source of Open Innovation	5027
<i>Diane-Gabrielle Tremblay, University of Quebec, Canada</i>	
Indigenous Knowledge Systems	5036
<i>Osarumwense Iguisi, University of Benin, Nigeria</i>	
<i>Osaro Rawlings Igbinomwanhia, University of Benin, Nigeria</i>	
Integrating Knowledge Management and Business Processes	5046
<i>John Steven Edwards, Aston University, UK</i>	
Intellectual Capital Measurement	5056
<i>Lukasz Bryl, Poznan University of Economics and Business, Poland</i>	
Knowledge Acquisition on Dante Alighieri's Works.....	5067
<i>Elvira Immacolata Locuratolo, ISTI-CNR, Italy</i>	
<i>Valentina Bartalesi Lenzi, ISTI-CNR, Italy</i>	
Knowledge Management for Development (KM4D).....	5077
<i>Alexander G. Flor, University of the Philippines, Philippines</i>	
Knowledge Management From the Metaphorical Perspective	5085
<i>Magdalena Bielenia-Grajewska, University of Gdansk, Poland</i>	
Theory and Practice of Online Knowledge Sharing	5093
<i>Will W. K. Ma, Hong Kong Shue Yan University, Hong Kong</i>	
Visualization as a Knowledge Transfer.....	5103
<i>Anna Ursyn, University of Northern Colorado, USA</i>	

Category: Language Studies

Mobile Testing System for Developing Language Skills.....	5116
<i>Svetlana Titova, Far Eastern Federal University, Russia</i>	
Nominalizations in Requirements Engineering Natural Language Models.....	5127
<i>Claudia S. Litvak, Universidad Nacional de La Matanza, Argentina & Universidad Nacional del Oeste, Argentina</i>	
<i>Graciela Dora Susana Hadad, Universidad Nacional del Oeste, Argentina</i>	
<i>Jorge Horacio Doorn, Universidad Nacional del Oeste, Argentina & Universidad Nacional de La Matanza, Argentina</i>	
Word Formation Study in Developing Naming Guidelines in the Translation of English Medical Terms Into Persian	5136
<i>Ali Akbar Zeinali, Universiti Sains Malaysia, Malaysia</i>	

Category: Learning Assessment and Measurement

Implicit Cognitive Vulnerability	5149
<i>Caroline M. Crawford, University of Houston – Clear Lake, US</i>	
Learning Analytics.....	5158
<i>Constanța-Nicoleta Bodea, Bucharest University of Economic Studies, Romania</i>	
<i>Maria-Iuliana Dascalu, University Politehnica of Bucharest, Romania</i>	
<i>Radu Ioan Mogos, Bucharest University of Economic Studies, Romania</i>	
<i>Stelian Stancu, Bucharest University of Economic Studies, Romania</i>	
Predicting Students Grades Using Artificial Neural Networks and Support Vector Machine.....	5169
<i>Sajid Umair, National University of Sciences and Technology (NUST), Pakistan</i>	
<i>Muhammad Majid Sharif, National University of Sciences and Technology (NUST), Pakistan</i>	
The Relationship Between Online Formative Assessment and State Test Scores Using Multilevel Modeling.....	5183
<i>Aryn C. Karpinski, Kent State University, USA</i>	
<i>Jerome V. D'Agostino, The Ohio State University, USA</i>	
<i>Anne-Evan K. Williams, Billings Middle School, USA</i>	
<i>Sue Ann Highland, Grand Canyon University, USA</i>	
<i>Jennifer A. Mellott, Kent State University, USA</i>	

Category: Library Science and Administration

Change Leadership Styles and Behaviors in Academic Libraries	5194
<i>John Kennedy Lewis, Salve Regina University, USA</i>	
Changing Expectations of Academic Libraries	5204
<i>Jennifer Ashley Wright Joe, Western Kentucky University, USA</i>	
Digital Archives for Preserving and Communicating Architectural Drawings	5213
<i>Roberta Spallone, Politecnico di Torino, Italy</i>	
<i>Francesca Paluan, Politecnico di Torino, Italy</i>	
Massive Digital Libraries (MDLs).....	5226
<i>Andrew Philip Weiss, California State University – Northridge, USA</i>	
Mission, Tools, and Ongoing Developments in the So.Re.Com. “A.S. de Rosa” @-library.....	5237
<i>Annamaria Silvana de Rosa, Sapienza University of Rome, Italy</i>	
Social Media Applications as Effective Service Delivery Tools for Librarians	5252
<i>Ihuoma Sandra Babatope, Delta State College of Physical Education, Nigeria</i>	

Web 2.0 From Evolution to Revolutionary Impact in Library and Information Centers.....	5262
<i>Zahid Ashraf Wani, University of Kashmir, India</i>	
<i>Tazeem Zainab, University of Kashmir, India</i>	
<i>Shabir Hussain, University of Kashmir, India</i>	

Category: Logistics and Supply Chain Management

Barcodes vs. RFID and Its Continued Success in Manufacturing and Services.....	5273
<i>Amber A. Smith-Ditizio, Texas Woman's University, USA</i>	
<i>Alan D. Smith, Robert Morris University, USA</i>	
Becoming Smart, Innovative, and Socially Responsible in Supply Chain Collaboration	5285
<i>Goknur Arzu Akyuz, University of Turkish Aeronautical Association, Turkey</i>	
<i>Guner Gursoy, Okan University, Turkey</i>	
Concept and Practices of Cyber Supply Chain in Manufacturing Context.....	5306
<i>Anisha Banu Dawood Gani, Universiti Sains Malaysia, Malaysia</i>	
<i>Yudi Fernando, Universiti Malaysia Pahang, Malaysia</i>	
The Concept of Modularity in the Context of IS Project Outsourcing	5317
<i>Shahzada Benazeer, University of Antwerp, Belgium</i>	
<i>Philip Huysmans, University of Antwerp, Belgium</i>	
<i>Peter De Bruyn, University of Antwerp, Belgium</i>	
<i>Jan Verelst, University of Antwerp, Belgium</i>	
Developing Global Supply Chain Manager for Business Expansion.....	5327
<i>Puspita Wulansari, Telkom University, Indonesia</i>	
<i>Yudi Fernando, Universiti Malaysia Pahang, Malaysia</i>	
Discrete Event Simulation in Inventory Management	5335
<i>Linh Nguyen Khanh Duong, Auckland University of Technology, New Zealand</i>	
<i>Lincoln C. Wood, University of Otago, New Zealand & Curtin University, Australia</i>	
E-Business Supply Chains Drivers, Metrics, and ERP Integration	5345
<i>Jean C. Essila, Northern Michigan University, USA</i>	
Ecological Performance as a New Metric to Measure Green Supply Chain Practices	5357
<i>June Poh Kim Tam, Universiti Sains Malaysia, Malaysia</i>	
<i>Yudi Fernando, Universiti Malaysia Pahang, Malaysia</i>	
E-Commerce in Logistics and Supply Chain Management	5367
<i>Yasanur Kayikci, Turkish-German University, Turkey</i>	
Exploring Drivers of Closed Loop Supply Chain in Malaysian Automotive Industry	5378
<i>Fadzlina Mohd Fadzil, Universiti Sains Malaysia, Malaysia</i>	
<i>Yudi Fernando, Universiti Malaysia Pahang, Malaysia</i>	

From Business-to-Business Software Startup to SAP's Acquisition	5388
<i>John Wang, Montclair State University, USA</i>	
<i>Jeffrey Hsu, Fairleigh Dickinson University, USA</i>	
<i>Sylvain Jaume, Saint Peter's University, USA</i>	
An Integrated Approach to Supply Chain Simulation	5398
<i>Nenad Stefanovic, University of Kragujevac, Serbia</i>	
<i>Bozidar Radenkovic, University of Belgrade, Serbia</i>	
Latest Advances on Benders Decomposition.....	5411
<i>Antonios Fragkogios, University of Thessaly, Greece</i>	
<i>Georgios K. D. Saharidis, University of Thessaly, Greece</i>	
Lean Logistics of the Transportation of Fresh Fruit Bunches (FFB) in the Palm Oil Industry	5422
<i>Cheah Cheng Teik, Universiti Sains Malaysia, Malaysia</i>	
<i>Yudi Fernando, Universiti Malaysia Pahang, Malaysia</i>	
Major Techniques and Current Developments of Supply Chain Process Modelling.....	5433
<i>Henry Xu, The University of Queensland, Australia</i>	
<i>Renae Agrey, The University of Queensland, Australia</i>	
Measuring Low Carbon Supply Chain.....	5446
<i>Muhammad Shabir Shaharudin, Universiti Sains Malaysia, Malaysia</i>	
<i>Yudi Fernando, Universiti Malaysia Pahang, Malaysia</i>	
Missing Part of Halal Supply Chain Management.....	5456
<i>Ratih Hendayani, Telkom University, Indonesia</i>	
<i>Yudi Fernando, Universiti Malaysia Pahang, Malaysia</i>	
Notions of Maritime Green Supply Chain Management	5465
<i>Fairuz Jasmi, Universiti Sains Malaysia, Malaysia</i>	
<i>Yudi Fernando, Universiti Malaysia Pahang, Malaysia</i>	
Offshoring IT	5476
<i>Susan Cockrell, Austin Peay State University, USA</i>	
<i>Terry Stringer Damron, Austin Peay State University, USA</i>	
<i>Amye M. Melton, Austin Peay State University, USA</i>	
<i>Alan D. Smith, Robert Morris University, USA</i>	
Pricing Based on Real-Time Analysis of Forklift Utilization Using RFID in Warehouse Management.....	5490
<i>Numan Celebi, Sakarya University, Turkey</i>	
<i>Kübra Savaş, Istanbul University, Turkey</i>	
<i>Ihsan Hakan Selvi, Sakarya University, Turkey</i>	

Profit Maximizing Network Modeling With Inventory and Capacity Considerations 5503
Tan Miller, Rider University, USA
Renato de Matta, University of Iowa, USA

Radio Frequency Identification Systems Within a Lean Supply Chain in a Global Environment ... 5516
Alan D. Smith, Robert Morris University, USA
Terry Stringer Damron, Austin Peay State University, USA
Susan Cockrell, Austin Peay State University, USA
Amye M. Melton, Austin Peay State University, USA

Reflections of the 1Malaysia Supply Chain (1MSC) 5527
Munira Halili, Universiti Sains Malaysia, Malaysia
Latifah Naina Mohamed, Universiti Sains Malaysia, Malaysia
Yudi Fernando, Universiti Malaysia Pahang, Malaysia

A Review of Advances in Supply Chain Intelligence 5538
Nenad Stefanovic, University of Kragujevac, Serbia
Danijela Milosevic, University of Kragujevac, Serbia

A Review of Supply Chain Risk Management in Agribusiness Industry 5550
Sri Widiyanesti, Telkom University, Indonesia
Yudi Fernando, Universiti Malaysia Pahang, Malaysia

The Role of Emerging Information Technologies for Supporting Supply Chain Management 5559
Zlatko Nedelko, University of Maribor, Slovenia
Vojko Potocan, University of Maribor, Slovenia

Samsung Company and an Analysis of Supplier-Side Supply Chain Management and IT Applications 5570
Amber A. Smith-Ditizio, Texas Woman's University, USA
Alan D. Smith, Robert Morris University, USA

Simulating Complex Supply Chain Relationships Using Copulas 5583
Krishnamurty Muralidhar, University of Oklahoma, USA
Rathindra Sarathy, Oklahoma State University, USA

Using RFID and Barcode Technologies to Improve Operations Efficiency Within the Supply Chain 5595
Amber A. Smith-Ditizio, Texas Woman's University, USA
Alan D. Smith, Robert Morris University, USA

Category: Management Science

The Business Transformation Framework for Managers in Transformation Projects 5607
Antoine Trad, IBITSM, Switzerland
Damir Kalpić, University of Zagreb, Croatia

Contemporary Leadership Development in Kazakhstan	5626
<i>Gainiya Tazhina, University of International Business, Kazakhstan</i>	
<i>Judith Parker, Teachers College, Columbia University, USA</i>	
<i>Arslan Ivashov, Kazakh Ablai Khan University of International Relations and World Languages, USA</i>	
Empirical Verification of the Performance Measurement System.....	5638
<i>Aleksander Janeš, University of Primorska, Slovenia</i>	
Lack of Characteristics Management Causing Biggest Projects Failure	5650
<i>Loredana Arana, University of Phoenix, USA</i>	
Making Sense of IS Project Stories	5660
<i>Darren Dalcher, University of Hertfordshire, UK</i>	

Volume VIII

(MA - MU)

The Measurement and Recognition of Intellectual Capital in the Process of Accounting Convergence Trends and Patterns	5669
<i>Ionica Oncioiu, Titu Maiorescu University, Romania</i>	
Project Control Using a Bayesian Approach.....	5679
<i>Franco Caron, Politecnico di Milano, Italy</i>	
Shaping Mega-Science Projects and Practical Steps for Success	5690
<i>Phil Crosby, Curtin University, Australia</i>	
Staying Ahead in Business Through Innovation.....	5705
<i>N. Raghavendra Rao, FINAIT Consultancy Services, India</i>	
Sustainable Competitive Advantage With the Balanced Scorecard Approach.....	5714
<i>Jorge Gomes, ISEG, Universidade de Lisboa, Portugal</i>	
<i>Mário José Batista Romão, ISEG, Universidade de Lisboa, Portugal</i>	
Transformational Leadership for Academic Libraries in Nigeria.....	5726
<i>Violet E. Ikolo, Delta State University Library, Nigeria</i>	

Category: Marketing

Are Social Marketing Investments Used as a Tool for Voluntary Reporting or Disclosure?	5737
<i>Tugba Ucma Uysal, Mugla Sitki Kocman University, Turkey</i>	
<i>Ganite Kurt, Gazi University, Turkey</i>	
<i>Ali Naci Karabulut, Mugla Sitki Kocman University, Turkey</i>	

The Impact of Artificial Intelligence and Virtual Personal Assistants on Marketing.....	5748
<i>Christina L. McDowell Marinchak, University of Alaska Anchorage, USA</i>	
<i>Edward Forrest, University of Alaska Anchorage, USA</i>	
<i>Bogdan Hoanca, University of Alaska Anchorage, USA</i>	
Marketing and Marketing Plan for Information Services	5757
<i>Sérgio Maravilhas, Universidade Salvador, Brazil</i>	
A Neuroaesthetic Approach to the Search of Beauty From the Consumer's Perspective.....	5767
<i>Gemma García Ferrer, Rey Juan Carlos University, Spain</i>	
Social Media Use and Customer Engagement.....	5775
<i>Aurora Garrido-Moreno, University of Malaga, Spain</i>	
<i>Nigel Lockett, University of Lancaster, UK</i>	
<i>Victor García-Morales, University of Granada, Spain</i>	

Category: Medical Education, Ethics, and Law

Comprehensive E-Learning Appraisal System	5787
<i>Jose Luis Monroy Anton, La Ribera University Hospital, Spain</i>	
<i>Juan Vicente Izquierdo Soriano, La Ribera University Hospital, Spain</i>	
<i>Maria Isabel Asensio Martinez, La Ribera University Hospital, Spain</i>	
<i>Felix Buendia Garcia, Polytechnical University of Valencia, Spain</i>	
Flipping the Medical School Classroom.....	5800
<i>Kristina Kaljo, Medical College of Wisconsin, USA</i>	
<i>Laura Jacques, Medical College of Wisconsin, USA</i>	
Integrating Evidence-Based Practice in Athletic Training Though Online Learning.....	5810
<i>Brittany A. Vorndran, Seton Hall University, USA</i>	
<i>Michelle Lee D'Abundo, Seton Hall University, USA</i>	
Integrating Web-Based Technologies Into the Education and Training of Health Professionals.....	5820
<i>Michelle Lee D'Abundo, Seton Hall University USA</i>	
<i>Cara Sidman, Arizona State University, USA</i>	
Interactivity in Distance Education and Computer-Aided Learning, With Medical Education Examples.....	5829
<i>D. John Doyle, Cleveland Clinic, USA</i>	
<i>Patrick J. Fahy, Athabasca University, Canada</i>	
Medical Equipment and Economic Determinants of Its Structure and Regulation in the Slovak Republic	5841
<i>Beáta Gavurová, Technical University of Košice, Slovakia</i>	
<i>Viliam Kováč, Technical University of Košice, Slovakia</i>	
<i>Michal Šoltés, Technical University of Košice, Slovakia</i>	

Use of Technology in Problem-Based Learning in Health Science.....	5853
<i>Indu Singh, Griffith University, Australia</i>	
<i>Avinash Reddy Kundur, Griffith University, Australia</i>	
<i>Yun-Mi Nguy, Griffith University, Australia</i>	

Category: Medical Technologies

Defining and Characterising the Landscape of eHealth.....	5864
<i>Yvonne O'Connor, University College Cork, Ireland</i>	
<i>Ciara Heavin, University College Cork, Ireland</i>	
Kinect Applications in Healthcare	5876
<i>Roanna Lun, Cleveland State University, USA</i>	
<i>Wenbing Zhao, Cleveland State University, USA</i>	
Neuroscience Technology and Interfaces for Speech, Language, and Musical Communication	5886
<i>Dionysios Politis, Aristotle University of Thessaloniki, Greece</i>	
<i>Miltiadis Tsalighopoulos, Aristotle University of Thessaloniki, Greece</i>	
<i>Georgios Kyriafinis, AHEPA University Hospital, Greece</i>	
Personalized Medicine	5901
<i>Sandip Bisui, Indian Institute of Technology (IIT) Kanpur, India</i>	
<i>Subhas Chandra Misra, Indian Institute of Technology (IIT) Kanpur, India</i>	
Pervasive Mobile Health	5908
<i>Muhammad Anshari, Universiti Brunei Darussalam, Brunei</i>	
<i>Mohammad Nabil Almunawar, Universiti Brunei Darussalam, Brunei</i>	
Radio Frequency Identification Technologies and Issues in Healthcare	5918
<i>Amber A. Smith-Ditizio, Texas Woman's University, USA</i>	
<i>Alan D. Smith, Robert Morris University, USA</i>	
Social Telerehabilitation	5930
<i>Gilberto Marzano, Rezekne Academy of Technologies, Latvia</i>	
A Validation Study of Rehabilitation Exercise Monitoring Using Kinect.....	5941
<i>Wenbing Zhao, Cleveland State University, USA</i>	
<i>Deborah D. Espy, Cleveland State University, USA</i>	
<i>Ann Reinthal, Cleveland State University, USA</i>	
Virtual Reality as Distraction Technique for Pain Management in Children and Adolescents	5955
<i>Barbara Atzori, University of Florence, Italy</i>	
<i>Hunter G. Hoffman, University of Washington, USA</i>	
<i>Laura Vagnoli, Meyer Children's Hospital of Florence, Italy</i>	
<i>Andrea Messeri, Meyer Children's Hospital of Florence, Italy</i>	
<i>Rosapia Lauro Grotto, University of Florence, Italy</i>	

Category: Mobile and Wireless Computing

Biogeography-Based Optimization Applied to Wireless Communications Problems	5967
<i>Sotirios K. Goudos, Aristotle University of Thessaloniki, Greece</i>	
BYOD (Bring Your Own Device), Mobile Technology Providers, and Its Impacts on Business/ Education and Workplace/Learning Applications	5981
<i>Amber A. Smith-Ditizio, Texas Woman's University, USA</i>	
<i>Alan D. Smith, Robert Morris University, USA</i>	
Cell Phone Conversation and Relative Crash Risk Update.....	5992
<i>Richard A. Young, Driving Safety Consulting, LLC, USA</i>	
Comb Filters Characteristics and Current Applications	6007
<i>Miriam Guadalupe Cruz-Jimenez, Institute INAOE, Mexico</i>	
<i>David Ernesto Troncoso Romero, CONACYT at ESCOM-IPN, Mexico</i>	
<i>Gordana Jovanovic Dolecek, Institute INAOE, Mexico</i>	
Consumer Adoption of PC-Based/Mobile-Based Electronic Word-of-Mouth	6019
<i>Akinori Ono, Keio University, Japan</i>	
<i>Mai Kikumori, Ritsumeikan University, Japan</i>	
Context-Aware Personalization for Mobile Services	6031
<i>Abayomi Moradeyo Otebolaku, Liverpool John Moores University, UK</i>	
<i>Maria Teresa Andrade, University of Porto, Portugal</i>	
Design of Compensators for Comb Decimation Filters	6043
<i>Gordana Jovanovic Dolecek, Institute INAOE Puebla, Mexico</i>	
An Empirical Study of Mobile/Handheld App Development Using Android Platforms	6057
<i>Wen-Chen Hu, University of North Dakota, USA</i>	
<i>Naima Kaabouch, University of North Dakota, USA</i>	
<i>Hung-Jen Yang, National Kaohsiung Normal University, Taiwan</i>	
Enhancing the Mobile User Experience Through Colored Contrasts.....	6070
<i>Jean-Éric Pelet, ESCE International Business School, France</i>	
<i>Basma Taieb, University of Cergy Pontoise, France</i>	
Ethical Ambiguities in the Privacy Policies of Mobile Health and Fitness Applications	6083
<i>Devjani Sen, University of Ottawa, Canada</i>	
<i>Rukhsana Ahmed, University of Ottawa, Canada</i>	
Exploring the Growth of Wireless Communications Systems and Challenges Facing 4G Networks	6094
<i>Amber A. Smith-Ditizio, Texas Woman's University, USA</i>	
<i>Alan D. Smith, Robert Morris University, USA</i>	

Flying Adhoc Networks Concept and Challenges	6106
<i>Kuldeep Singh, Thapar University, India</i>	
<i>Anil Kumar Verma, Thapar University, India</i>	
Health Wearables Turn to Fashion.....	6114
<i>Lambert Spaanenburg, Comoray AB, Sweden</i>	
Human Psychomotor Performance Under the Exposure to Mobile Phones-Like Electromagnetic Fields.....	6124
<i>Giuseppe Curcio, University of L'Aquila, Italy</i>	
Identification of Wireless Devices From Their Physical Layer Radio-Frequency Fingerprints	6136
<i>Gianmarco Baldini, European Commission – Joint Research Centre, Italy</i>	
<i>Gary Steri, European Commission – Joint Research Centre, Italy</i>	
<i>Raimondo Giuliani, European Commission – Joint Research Centre, Italy</i>	
The Impact of Mobile Phones on Plastic Surgery and Burn Management.....	6147
<i>Maria Giaquinto-Cilliers, Kimberley Hospital Complex, South Africa</i>	
<i>Tertius N. Potgieter, Kimberley Hospital Complex, South Africa</i>	
<i>Gert Steyn, Kimberley Hospital Complex, South Africa</i>	
The Intersection of Religion and Mobile Technology	6161
<i>Wendi R. Bellar, Texas A&M University, USA</i>	
<i>Kyong James Cho, Texas A&M University, USA</i>	
<i>Heidi A. Campbell, Texas A&M University, USA</i>	
Methods for Simultaneous Improvement of Comb Pass Band and Folding Bands	6171
<i>Gordana Jovanovic Dolecek, Institute INAOE Puebla, Mexico</i>	
Micro to Macro Social Connectedness Through Mobile Phone Engagement	6184
<i>Dominic Mentor, Columbia University, USA</i>	
Mobile Applications for Automatic Object Recognition	6195
<i>Danilo Avola, University of Udine, Italy</i>	
<i>Gian Luca Foresti, University of Udine, Italy</i>	
<i>Claudio Piciarelli, University of Udine, Italy</i>	
<i>Marco Vernier, University of Udine, Italy</i>	
<i>Luigi Cinque, Sapienza University, Italy</i>	
Mobile Apps Threats.....	6207
<i>Donovan Peter Chan Wai Loon, University of Malaya, Malaysia</i>	
<i>Sameer Kumar, University of Malaya, Malaysia</i>	
Mobile Technologies Impact on Economic Development in Sub-Saharan Africa	6216
<i>Adam Crossan, Letterkenny Institute of Technology, Ireland</i>	
<i>Nigel McKelvey, Letterkenny Institute of Technology, Ireland</i>	
<i>Kevin Curran, Ulster University, UK</i>	

Mobile Virtual Reality to Enhance Subjective Well-Being	6223
<i>Federica Pallavicini, Università di Milano-Bicocca, Italy</i>	
<i>Luca Morganti, Università di Milano-Bicocca, Italy</i>	
<i>Barbara Diana, Università di Milano-Bicocca, Italy</i>	
<i>Olivia Realdon, University of Milano-Bicocca, Italy</i>	
<i>Valentino Zurloni, Università di Milano-Bicocca, Italy</i>	
<i>Fabrizia Mantovani, Università di Milano-Bicocca, Italy</i>	
Novel Methods to Design Low-Complexity Digital Finite Impulse Response (FIR) Filters.....	6234
<i>David Ernesto Troncoso Romero, CONACYT at ESCOM-IPN, Mexico</i>	
<i>Gordana Jovanovic Dolecek, Institute INAOE, Mexico</i>	
Open Source.....	6245
<i>Heidi Lee Schnackenberg, SUNY Plattsburgh, USA</i>	
Power Consumption in Wireless Access Networks	6253
<i>Vinod Kumar Mishra, B. T. Kumaon Institute of Technology, India</i>	
<i>Pankaja Bisht, B. T. Kumaon Institute of Technology, India</i>	
Resource Management for Multimedia Services in Long Term Evaluation Networks	6266
<i>Vinod Kumar Mishra, B. T. Kumaon Institute of Technology, India</i>	
<i>Tanuja Pathak, B. T. Kumaon Institute of Technology, India</i>	
SMS & Civil Unrest.....	6275
<i>Innocent Chiluwa, Covenant University OTA, Nigeria</i>	
A Survey of People Localization Techniques Utilizing Mobile Phones.....	6286
<i>Levent Bayındır, Ataturk University, Turkey</i>	
Technological Innovation and Use in the Early Days of Camera Phone Photo Messaging.....	6296
<i>Jonathan Lillie, Loyola University, USA</i>	
Viterbi Decoder in Hardware	6307
<i>Mário Pereira Véstias, Instituto Politecnico de Lisboa, Portugal</i>	
Wireless Implant Communications Using the Human Body	6319
<i>Assefa K. Teshome, Victoria University, Australia</i>	
<i>Behailu Kibret, Victoria University, Australia</i>	
<i>Daniel T. H. Lai, Victoria University, Australia</i>	

Category: Mobile Learning

E-Collaborative Learning (e-CL).....	6336
<i>Alexandros Xafopoulos, University College London, UK</i>	

Learning With Mobile Devices.....	6347
<i>Helen Crompton, Old Dominion University, USA</i>	
<i>John Traxler, University of Wolverhampton, UK</i>	
Mobile Game-Based Learning.....	6361
<i>Boaventura DaCosta, Solers Research Group, USA</i>	
<i>Soonhwa Seok, Korea University, South Korea</i>	
<i>Carolyn Kinsell, Solers Research Group, USA</i>	
Mobile Game-Based Learning in STEM Subjects.....	6376
<i>Marcelo Leandro Eichler, Universidade Federal do Rio Grande do Sul, Brazil</i>	
<i>Gabriela Trindade Perry, Universidade Federal do Rio Grande do Sul, Brazil</i>	
<i>Ivana Lima Lucchesi, Universidade Federal do Rio Grande do Sul, Brazil</i>	
<i>Thiago Troina Melendez, Instituto Federal Sul-Riograndense, Brazil</i>	
Mobile Learning in and out of the K-12 Classroom	6388
<i>Pena L. Bedesem, Kent State University, USA</i>	
<i>Tracy Arner, Kent State University, USA</i>	
A Psychological Perspective on Mobile Learning	6398
<i>Melody M. Terras, University of the West of Scotland, UK</i>	
<i>Judith Ramsay, Manchester Metropolitan University, UK</i>	
The Role of Distance Education in Global Education	6412
<i>Kijpokin Kasemsap, Suan Sunandha Rajabhat University, Thailand</i>	

Category: Multimedia Technology

Adaptive Hypermedia Systems	6424
<i>Ana Carolina Tomé Klock, Federal University of Rio Grande do Sul (UFRGS), Brazil</i>	
<i>Isabela Gasparini, Santa Catarina State University (UDESC), Brazil</i>	
<i>Marcelo Soares Pimenta, Federal University of Rio Grande do Sul (UFRGS), Brazil</i>	
<i>José Palazzo M. de Oliveira, Federal University of Rio Grande do Sul (UFRGS), Brazil</i>	
Group Synchronization for Multimedia Systems.....	6435
<i>Dimitris N. Kanellopoulos, University of Patras, Greece</i>	
Metadata Standards in Digital Audio.....	6447
<i>Kimmy Szeto, Baruch College, City University of New York, USA</i>	
Multimedia-Enabled Dot Codes as Communication Technologies	6464
<i>Shigeru Ikuta, Otsuma Women's University, Japan</i>	

Volume IX (MU - SO)

Semantically Enhanced Authoring of Shared Media.....	6476
<i>Charalampos A. Dimoulas, Aristotle University of Thessaloniki, Greece</i>	
<i>Andreas A. Veglis, Aristotle University of Thessaloniki, Greece</i>	
<i>George Kalliris, Aristotle University of Thessaloniki, Greece</i>	

Transmedia and Transliteracy in Nemetical Analysis.....	6488
<i>Michael Josefowicz, Nemetics Institute Kolkata, USA</i>	
<i>Ray Gallon, The Transformation Society, France</i>	
<i>Maria Nieves Lorenzo Galés, The Transformation Society, Spain</i>	

Category: Networking and Telecommunications

Autonomic Cooperative Communications.....	6499
<i>Michal Wodczak, Samsung Electronics, Poland</i>	

Clique Size and Centrality Metrics for Analysis of Real-World Network Graphs.....	6507
<i>Natarajan Meghanathan, Jackson State University, USA</i>	

Distributed Methods for Multi-Sink Wireless Sensor Networks Formation	6522
<i>Miriam A. Carlos-Mancilla, CINVESTAV Unidad Guadalajara, Mexico</i>	
<i>Ernesto Lopez-Mellado, CINVESTAV Unidad Guadalajara, Mexico</i>	
<i>Mario Siller, CINVESTAV Unidad Guadalajara, Mexico</i>	

A Graph-Intersection-Based Algorithm to Determine Maximum Lifetime Communication Topologies for Cognitive Radio Ad Hoc Networks	6536
<i>Natarajan Meghanathan, Jackson State University, USA</i>	

Improving Quality of Business in Next Generation Telecom Networks	6546
<i>Vesna Radonjić Đogatović, University of Belgrade, Serbia</i>	

Information-Centric Networking	6556
<i>Mohamed Fazil Mohamed Firdhous, University of Moratuwa, Sri Lanka</i>	

Interoperability Frameworks for Distributed Systems	6566
<i>José Carlos Martins Delgado, Universidade de Lisboa, Portugal</i>	

Neural Networks and Their Accelerated Evolution From an Economic Analysis Perspective	6579
<i>Stelian Stancu, Bucharest University of Economic Studies, Romania</i>	
<i>Constanța-Nicoleta Bodea, Bucharest University of Economic Studies, Romania</i>	
<i>Oana Mădălina Popescu(Predescu), Bucharest University of Economic Studies, Romania</i>	
<i>Alina Neamțu(Idorași), Bucharest University of Economic Studies, Romania</i>	

Optimization of Antenna Arrays and Microwave Filters Using Differential Evolution Algorithms	6595
<i>Sotirios K. Goudos, Aristotle University of Thessaloniki, Greece</i>	
QoS Architectures for the IP Network	6609
<i>Harry G. Perros, North Carolina State University, USA</i>	
Throughput Dependence on SNR in IEEE802.11 WLAN Systems	6618
<i>Ikponmwosa Oghogho, Delta State University, Abraka-Oleh Campus, Nigeria</i>	

Category: Neural Networks

Applications of Artificial Neural Networks in Economics and Finance.....	6631
<i>Iva Mihaylova, University of St. Gallen, Switzerland</i>	
Artificial Neural Networks and Their Applications in Business.....	6642
<i>Trevor J. Bihl, Air Force Institute of Technology, USA</i>	
<i>William A. Young, Ohio University, USA</i>	
<i>Gary R. Weckman, Ohio University, USA</i>	
Recurrent Neural Networks for Predicting Mobile Device State.....	6658
<i>Juan Manuel Rodriguez, ISISTAN, UNICEN-CONICET, Argentina</i>	
<i>Alejandro Zunino, ISISTAN, UNICEN-CONICET, Argentina</i>	
<i>Antonela Tommasel, ISISTAN, UNICEN-CONICET, Argentina</i>	
<i>Cristian Mateos, ISISTAN, UNICEN-CONICET, Argentina</i>	

Category: Optical Engineering

Visible Light Communication Numerous Applications.....	6672
<i>Ala' Fathi Khalifeh, German Jordan University, Jordan</i>	
<i>Hasan Farahneh, Ryerson University, Canada</i>	
<i>Christopher Mekhiel, Ryerson University, Canada</i>	
<i>Xavier Fernando, Ryerson University, Canada</i>	

Category: Public Sector Management

Community Outreach.....	6685
<i>Loriene Roy, The University of Texas at Austin, USA</i>	
<i>Antonia Frydman, The University of Texas at Austin, USA</i>	
Exploring "Hacking," Digital Public Art, and Implication for Contemporary Governance.....	6695
<i>Amadu Wurie Khan, University of Edinburgh, UK</i>	
<i>Chris Speed, University of Edinburgh, UK</i>	

Political Context Elements in Public Policy of Radio Frequency Information Technology and Electromagnetic Fields..... 6710

Joshua M. Steinfeld, Old Dominion University, USA

Public Policies for Providing Cloud Computing Services to SMEs of Latin America..... 6727

Mohd Nayyer Rahman, Aligarh Muslim University, India

Badar Alam Iqbal, Aligarh Muslim University, India

Category: Research Methods and Scholarly Publishing

Advancement and Application of Scientometric Indicators for Evaluation of Research Content.... 6739

Tazeem Zainab, University of Kashmir, India

Zahid Ashraf Wani, University of Kashmir, India

Electronic Theses and Dissertations (ETDs) 6748

Ralph Hartsock, University of North Texas, USA

Daniel G. Alemneh, University of North Texas, USA

The Nature of Research Methodologies..... 6756

Ben Tran, Alliant International University, USA

Research Methodology 6767

Swati C. Jagdale, MAEER's Maharashtra Institute of Pharmacy, India

Rahul U. Hude, MAEER's Maharashtra Institute of Pharmacy, India

Aniruddha R. Chabukswar, MAEER's Maharashtra Institute of Pharmacy, India

Scholarly Identity in an Increasingly Open and Digitally Connected World..... 6779

Olga Belikov, Brigham Young University, USA

Royce Kimmons, Brigham Young University, USA

Category: Risk Assessment

Fuzzy Logic Approach in Risk Assessment 6789

Çetin Karahan, Directorate General of Civil Aviation, Turkey

Esra Ayça Güzeldereli, Afyon Kocatepe University, Turkey

Aslıhan Tüfekci, Gazi University, Turkey

Predictive Analytics and Intelligent Risk Detection in Healthcare Contexts..... 6806

Nilmini Wickramasinghe, Epworth HealthCare, Australia & Deakin University, Australia

Stress Testing Corporations and Municipalities and Supply Chains 6813

Frank Wolf, CSSTA L3C, USA

Category: Robotics

- Binary Decision Diagram Reliability for Multiple Robot Complex System 6825
Hamed Fazlollahtabar, Sharif University of Technology, Iran & National Elites Foundation, Iran
Seyed Taghi Akhavan Niaki, Sharif University of Technology, Iran
- A Bio-Inspired, Distributed Control Approach to the Design of Autonomous Cooperative Behaviors in Multiple Mobile Robot Systems 6836
Gen'ichi Yasuda, Nagasaki Institute of Applied Science, Japan
- Improving Dependability of Robotics Systems 6847
Nidhal Mahmud, University of Hull, UK
- Robotics and Programming Integration as Cognitive-Learning Tools 6859
Nikleia Eteokleous, Frederick University Cyprus, Cyprus
- State of the Art and Key Design Challenges of Telesurgical Robotics..... 6872
Sajid Nisar, National University of Science and Technology, Pakistan
Osman Hasan, National University of Science and Technology, Pakistan
- Telesurgical Robotics and a Kinematic Perspective 6882
Sajid Nisar, National University of Sciences and Technology, Pakistan
Osman Hasan, National University of Sciences and Technology, Pakistan
- Using Global Appearance Descriptors to Solve Topological Visual SLAM..... 6894
Lorenzo Fernández Rojo, Miguel Hernandez University, Spain
Luis Paya, Miguel Hernández University, Spain
Francisco Amoros, Miguel Hernandez University, Spain
Oscar Reinoso, Miguel Hernandez University, Spain

Category: Small and Medium Enterprises

- Big Data and Simulations for the Solution of Controversies in Small Businesses 6907
Milena Janakova, Silesian University in Opava, Czech Republic
- Financing Micro, Small, and Medium Enterprises in Indian Industry 6916
Shromona Ganguly, Indian Institute of Management Calcutta, India & Reserve Bank of India, India
- Software Development Process Standards for Very Small Companies 6927
Rory V. O'Connor, Dublin City University, Ireland

Category: Social Networking and Computing

Adolescents' Food Communication in Social Media:	6940
<i>Christopher Holmberg, University of Gothenburg, Sweden</i>	
Agent-Based Social Networks	6950
<i>Federico Bergenti, Università degli Studi di Parma, Italy</i>	
<i>Agostino Poggi, Università degli Studi di Parma, Italy</i>	
<i>Michele Tomaiuolo, Università degli Studi di Parma, Italy</i>	
Aspects of Various Community Detection Algorithms in Social Network Analysis	6961
<i>Nicole Belinda Dillen, St. Thomas' College of Engineering and Technology, India</i>	
<i>Aruna Chakraborty, St. Thomas' College of Engineering and Technology, India</i>	
Classification of Traffic Events Notified in Social Networks' Texts.....	6973
<i>Ana Maria Magdalena Saldana-Perez, Instituto Politecnico Nacional, Mexico</i>	
<i>Marco Antonio Moreno-Ibarra, Instituto Politécnico Nacional, Mexico</i>	
<i>Miguel Jesus Torres-Ruiz, Instituto Politécnico Nacional, Mexico</i>	
Communication Privacy Management and Mediated Communication	6985
<i>Debra L. Worthington, Auburn University, USA</i>	
<i>Margaret Fitch-Hauser, Auburn University, USA</i>	
The Dual Nature of Participatory Web and How Misinformation Seemingly Travels	6993
<i>Sameer Kumar, University of Malaya, Malaysia</i>	
Effective Cultural Communication via Information and Communication Technologies and Social Media Use	7002
<i>Androniki Kavoura, Technological Educational Institute of Athens, Greece</i>	
<i>Stella Sylaiou, Aristotle University of Thessaloniki, Greece</i>	
From the Psychoanalyst's Couch to Social Networks	7014
<i>Annamaria Silvana de Rosa, Sapienza University of Rome, Italy</i>	
<i>Emanuele Fino, Psychologist, Psychometrician, Italy</i>	
<i>Elena Bocci, Sapienza University of Rome, Italy</i>	
The Internet Behavior of Older Adults	7026
<i>Elizabeth Mazur, Pennsylvania State University, USA</i>	
<i>Margaret L. Signorella, Pennsylvania State University, USA</i>	
<i>Michelle Hough, Pennsylvania State University, USA</i>	
Issues and Challenges in Enterprise Social Media	7036
<i>Sarabjot Kaur, IIT Kanpur, India</i>	
<i>Subhas Chandra Misra, IIT Kanpur, India</i>	

Mapping the Dissemination of the Theory of Social Representations via Academic Social Networks	7044
<i>Annamaria Silvana de Rosa, Sapienza University of Rome, Italy</i>	
<i>Laura Dryjanska, Sapienza University of Rome, Italy</i>	
<i>Elena Bocci, Sapienza University of Rome, Italy</i>	
The NetLab Network.....	7057
<i>Dimitrina Dimitrova, York University, Canada</i>	
<i>Barry Wellman, NetLab Network, Canada</i>	
Online Dating/Dating Apps	7069
<i>Vladimir Santiago Arias, Texas Tech University, USA</i>	
<i>Narissra Maria Punyanunt-Carter, Texas Tech University, USA</i>	
Online Prosocial Behaviors.....	7077
<i>Michelle F. Wright, Pennsylvania State University, USA</i>	
<i>William Stanley Pendergrass, American Public University System, USA</i>	
Online Social Networking Behavior and Its Influence Towards Students' Academic Performance	7088
<i>Maslin Masrom, Universiti Teknologi Malaysia, Malaysia</i>	
<i>Selisa Usat, Universiti Teknologi Malaysia, Malaysia</i>	
Parental Mediation of Adolescent Technology Use.....	7097
<i>J. Mitchell Vaterlaus, Montana State University, USA</i>	
The Qualities and Potential of Social Media	7106
<i>Udo Richard Averweg, eThekweni Municipality, South Africa</i>	
<i>Marcus Leaning, University of Winchester, UK</i>	
Short History of Social Networking and Its Far-Reaching Impact	7116
<i>Liguo Yu, Indiana University – South Bend, USA</i>	
Social Media and Business Practices	7126
<i>Ashish Kumar Rathore, Indian Institute of Technology Delhi, India</i>	
<i>P. Vigneswara Ilavarasan, Indian Institute of Technology Delhi, India</i>	
Social Media Credit Scoring.....	7140
<i>Billie Anderson, Ferris State University, USA</i>	
<i>J. Michael Hardin, Samford University, USA</i>	
Social Network Analysis and the Study of University Industry Relations	7150
<i>Fernando Cabrita Romero, University of Minho, Portugal</i>	
Social Networking and Knowledge Sharing in Organizations.....	7161
<i>Sarabjot Kaur, Indian Institute of Technology Kanpur, India</i>	
<i>Subhas Chandra Misra, Indian Institute of Technology Kanpur, India</i>	

Understanding the Potentials of Social Media in Collaborative Learning.....	7168
<i>Adem Karahoca, Bahcesehir University, Turkey</i>	
<i>İlker Yengin, A*STAR, Institute of High Performance Computing, Singapore</i>	
Using Social Media to Increase the Recruitment of Clinical Research Participants	7181
<i>Saliha Akhtar, Seton Hall University, USA</i>	
Why It Is Difficult to Disengage From Facebook	7190
<i>Sonda Bouattour Fakhfakh, University of Tunis El-Manar, Tunisia</i>	

Category: Socio-Economic Development

Community Science and Technology and Its Meaning to Potential Requirement.....	7201
<i>P. K. Paul, Raiganj University, India</i>	
<i>A. Bhumali, Raiganj University, India</i>	
Financial Inclusion, Content, and Information Technologies in Latin America.....	7214
<i>Alberto Chong, Georgia State University, USA & Universidad del Pacifico, Peru</i>	
<i>Cecilia de Mendoza, Ministry of Production of Argentina, Argentina</i>	
The Growing Impact of ICT on Development in Africa.....	7223
<i>Sherif H. Kamel, The American University in Cairo, Egypt</i>	
Manufacturing vs. Services and the Role of Information Technology	7234
<i>Arnab Adhikari, Indian Institute of Management Ranchi, India</i>	
<i>Shromona Ganguly, Indian Institute of Management Calcutta, India</i>	
New Faces of Digital Divide and How to Bridge It.....	7248
<i>Viktor Freiman, Université de Moncton, Canada</i>	
<i>Dragana Martinovic, University of Windsor, Canada</i>	
<i>Xavier Robichaud, Université de Moncton, Canada</i>	
The Potential Role of the Software Industry in Supporting Economic Development	7259
<i>Sherif H. Kamel, The American University in Cairo, Egypt</i>	
Socio-Economic Processes, User Generated Content, and Media Pluralism	7270
<i>Androniki Kavoura, Technological Educational Institute of Athens, Greece</i>	

Volume X (SO - W)

Category: Sociology

Bipolar Model in Collective Choice	7282
<i>Ayeley P. Tchangani, Université Fédérale Toulouse Midi-Pyrénées, France</i>	

Censorship in the Digital Age the World Over	7292
<i>Kari D. Weaver, University of South Carolina – Aiken, USA</i>	
Information-Based Revolution in Military Affairs	7302
<i>Rafal Kopec, Pedagogical University of Krakow, Poland</i>	
The Networked Effect of Children and Online Digital Technologies.....	7312
<i>Teresa Sofia Pereira Dias de Castro, University of Minho, Portugal</i>	
<i>António Osório, University of Minho, Portugal</i>	
<i>Emma Bond, University Campus Suffolk, UK</i>	
Suggestions for Communication of Information for Multicultural Co-Existence	7327
<i>Noriko Kurata, Tokyo University of Science, Suwa, Japan</i>	
Vitalizing Ancient Cultures Mythological Storytelling in Metal Music	7338
<i>Uğur Kiliç, Ondokuz Mayıs University, Turkey</i>	

Category: Sports and Entertainment

Mining Sport Activities	7348
<i>Iztok Fister Jr., University of Maribor, Slovenia</i>	
<i>Iztok Fister, University of Maribor, Slovenia</i>	
Sport Exergames for Physical Education	7358
<i>Pooya Soltani, University of Porto, Portugal</i>	
<i>João Paulo Vilas-Boas, University of Porto, Portugal</i>	

Category: Systems and Software Engineering

Adopting Open Source Software in Smartphone Manufacturers' Open Innovation Strategy	7369
<i>Mohammad Nabil Almunawar, Universiti Brunei Darussalam, Brunei</i>	
<i>Muhammad Anshari, Universiti Brunei Darussalam, Brunei</i>	
<i>Heru Susanto, Indonesian Institute of Sciences, Indonesia & Tunghai University, Taiwan</i>	
Assessing Computer-Aided Design Skills	7382
<i>Yi Lin Wong, The Hong Kong Polytechnic University, Hong Kong</i>	
<i>Kin Wai Michael Siu, The Hong Kong Polytechnic University, Hong Kong</i>	
The Challenges of Teaching and Learning Software Programming to Novice Students.....	7392
<i>Seyed Reza Shahamiri, Manukau Institute of Technology, New Zealand</i>	
Developing a Glossary for Software Projects	7399
<i>Tamer Abdou, Suez Canal University, Egypt</i>	
<i>Pankaj Kamthan, Concordia University, Canada</i>	
<i>Nazlie Shahmir, WestJet Airlines Limited, Canada</i>	

Displaying Hidden Information in Glossaries	7411
<i>Marcela Ridao, Universidad Nacional del Centro de la Provincia de Buenos Aires, Argentina</i>	
<i>Jorge Horacio Doorn, Universidad Nacional del Oeste, Argentina & Universidad Nacional de La Matanza, Argentina</i>	
Dynamic Situational Adaptation of a Requirements Engineering Process.....	7422
<i>Graciela Dora Susana Hadad, Universidad Nacional del Oeste, Argentina & Universidad de Belgrano, Argentina</i>	
<i>Jorge Horacio Doorn, Universidad Nacional del Oeste, Argentina & Universidad Nacional de La Matanza, Argentina</i>	
<i>Viviana Alejandra Ledesma, Universidad Nacional del Oeste, Argentina & Universidad Nacional de La Matanza, Argentina</i>	
A Formal Approach to the Distributed Software Control for Automated Multi-Axis Manufacturing Machines	7435
<i>Gen'ichi Yasuda, Nagasaki Institute of Applied Science, Japan</i>	
Model-Driven Software Modernization.....	7447
<i>Liliana Maria Favre, Universidad Nacional Del Centro De La Provincia De Buenos Aires, Argentina</i>	
<i>Liliana Martinez, Universidad Nacional del Centro de la Provincia de Buenos Aires, Argentina</i>	
<i>Claudia Teresa Pereira, Universidad Nacional del Centro de la Provincia de Buenos Aires, Argentina</i>	
Mutation Testing Applied to Object-Oriented Languages.....	7459
<i>Pedro Delgado-Pérez, University of Cádiz, Spain</i>	
<i>Inmaculada Medina-Bulo, University of Cádiz, Spain</i>	
<i>Juan José Domínguez-Jiménez, University of Cádiz, Spain</i>	
Object-Oriented Programming in Computer Science.....	7470
<i>Rahime Yilmaz, Istanbul University, Turkey</i>	
<i>Anil Sezgin, Yildiz Technical University, Turkey</i>	
<i>Sefer Kurnaz, Istanbul Esenyurt University, Turkey</i>	
<i>Yunus Ziya Arslan, Istanbul University, Turkey</i>	
The Past, Present, and Future of UML.....	7481
<i>Rebecca Platt, Murdoch University, Australia</i>	
<i>Nik Thompson, Curtin University, Australia</i>	
Petri Nets Identification Techniques for Automated Modelling of Discrete Event Processes	7488
<i>Edelma Rodriguez-Perez, CINVESTAV Unidad Guadalajara, Mexico</i>	
<i>Ernesto Lopez-Mellado, CINVESTAV Unidad Guadalajara, Mexico</i>	

Research and Development on Software Testing Techniques and Tools.....	7503
<i>Tamilarasi T, VIT University, India</i>	
<i>M. Prasanna, VIT University, India</i>	
The Role of Feedback in Software Process Assessment.....	7514
<i>Zeljko Stojanov, University of Novi Sad, Serbia</i>	
<i>Dalibor Dobrilovic, University of Novi Sad, Serbia</i>	
Secure Software Development of Cyber-Physical and IoT Systems.....	7525
<i>Muthu Ramachandran, Leeds Metropolitan University, UK</i>	
Software Literacy	7539
<i>Elaine Khoo, University of Waikato, New Zealand</i>	
<i>Craig Hight, University of Newcastle, Australia</i>	
Software Process Improvement for Web-Based Projects Comparative View	7549
<i>Thamer Al-Rousan, Isra University, Pakistan</i>	
A Study of Contemporary System Performance Testing Framework.....	7563
<i>Alex Ng, Federation University, Australia</i>	
<i>Shiping Chen, CSIRO Data61, Australia</i>	
A Tale of Two Agile Requirements Engineering Practices	7577
<i>Pankaj Kamthan, Concordia University, Canada</i>	
<i>Terrill Fancott, Concordia University, Canada</i>	
Towards an Understanding of Performance, Reliability, and Security	7588
<i>Ye Wang, Zhejiang Gongshang University, China</i>	
<i>Bo Jiang, Zhejiang Gongshang University, China</i>	
<i>Weifeng Pan, Zhejiang Gongshang University, China</i>	
Understanding User Experience	7599
<i>Camille Dickson-Deane, University of Melbourne, Australia</i>	
<i>Hsin-Liang (Oliver) Chen, University of Massachusetts Boston, USA</i>	
The What, How, and When of Formal Methods	7609
<i>Aristides Dasso, Universidad Nacional de San Luis, Argentina</i>	
<i>Ana Funes, Universidad Nacional de San Luis, Argentina</i>	

Category: Teacher Education

Constructing Preservice Teachers' Knowledge of Technology Integration.....	7623
<i>Kathleen A. Paciga, Columbia College Chicago, USA</i>	
<i>Angela Fowler, Erikson Institute, USA</i>	
<i>Mary Quest, Erikson Institute, USA</i>	

Effectiveness of Teacher Training in Using Latest Technologies	7635
<i>Revathi Viswanathan, B. S. Abdur Rahman University, India</i>	
Role of Educational Leaders in Supporting Beginning Teachers in Al Ain Schools in the UAE	7647
<i>Salam Omar Ali, Brighton Collage Al Ain, UAE</i>	
The Technological Pedagogical Content Knowledge of EFL Teachers (EFL TPACK)	7659
<i>Mehrak Rahimi, Shahid Rajaei Teacher Training University, Iran</i>	
<i>Shakiba Pourshahbaz, Shahid Rajaei Teacher Training University, Iran</i>	
Ubiquitous Teachers' Training and Lessons Learned with the uProf! Model	7671
<i>Sabrina Leone, Università Politecnica delle Marche, Italy</i>	
<i>Giovanni Biancofiore, giovannibiancofiore.com, Italy</i>	
Video Considerations for the World Language edTPA.....	7682
<i>Elizabeth Goulette, Georgia State University, USA</i>	
<i>Pete Swanson, Georgia State University, USA</i>	

Category: Theoretical Computer Science

Efficient Optimization Using Metaheuristics.....	7693
<i>Sergio Nesmachnow, Universidad de la República, Uruguay</i>	
An Essay on Denotational Mathematics	7704
<i>Giuseppe Iurato, University of Palermo, Italy</i>	
Quantum Computing and Quantum Communication	7715
<i>Göran Pulkkis, Arcada University of Applied Sciences, Finland</i>	
<i>Kaj J. Grahn, Arcada University of Applied Sciences, Finland</i>	
Sleptsov Net Computing	7731
<i>Dmitry A. Zaitsev, International Humanitarian University, Ukraine</i>	

Category: Ubiquitous and Pervasive Computing

Energy Conservation in the Era of Ubiquitous Computing	7745
<i>P. P. Abdul Haleem, Farook College, India</i>	
From General Services to Pervasive and Sensitive Services	7754
<i>Mario Vega-Barbas, KTH-Royal Institute of Technology, Sweden & ESNE-Universidad</i>	
<i>Camilo José Cela, Spain</i>	
<i>Iván Pau, Universidad Politécnica de Madrid, Spain</i>	
<i>Fernando Seoane, KI-Karolinska Institutet, Sweden & University of Borås, Sweden</i>	

Home UbiHealth	7765
<i>John Sarivougioukas, "G. Gennimatas" Athens General Hospital, Greece</i>	
<i>Aristides Vagelatos, CTI&P, Greece</i>	
<i>Konstantinos Parsopoulos, University of Ioannina, Greece</i>	
<i>Isaac E. Lagaris, University of Ioannina, Greece</i>	
Multifaceted Applications of the Internet of Things.....	7775
<i>Kijpokin Kasemsap, Suan Sunandha Rajabhat University, Thailand</i>	
The Role of U-FADE in Selecting Persuasive System Features.....	7785
<i>Isaac Wiafe, Ghana Institute of Management and Public Administration, Ghana</i>	
Social Computing.....	7796
<i>Nolan Hemmatazad, University of Nebraska at Omaha, USA</i>	
Ubiquitous Computing, Contactless Points, and Distributed Stores.....	7805
<i>Marco Savastano, Sapienza University of Rome, Italy</i>	
<i>Eleonora Pantano, Middlesex University London, UK</i>	
<i>Saverino Verteramo, University of Calabria, Italy</i>	

Category: Urban and Regional Development

Climate Change as a Driving Force on Urban Energy Consumption Patterns	7815
<i>Mostafa Jafari, Agricultural Research Education and Extension Organization (AREEO),</i>	
<i>Iran</i>	
<i>Pete Smith, University of Aberdeen, UK</i>	
Determination of Urban Growth by the Night-Time Images	7831
<i>Emre Yücer, Erzincan University, Turkey</i>	
<i>Arzu Erener, Kocaeli University, Turkey</i>	
Need for Rethinking Modern Urban Planning Strategies Through Integration of ICTs.....	7843
<i>Rounaq Basu, Indian Institute of Technology Bombay, India</i>	
<i>Arnab Jana, Indian Institute of Technology Bombay, India</i>	
Reconstructive Architectural and Urban Digital Modelling	7856
<i>Roberta Spallone, Politecnico di Torino, Italy</i>	
Regional Development and Air Freight Service Needs for Regional Communities.....	7869
<i>Tarryn Kille, Griffith University, Australia</i>	
<i>Paul Bates, University of Southern Queensland, Australia</i>	
<i>Patrick S. Murray, University of Southern Queensland, Australia</i>	

Category: Virtual Learning Environments

Creating Active Learning Spaces in Virtual Worlds.....	7880
<i>Reneta D. Lansiquot, City University of New York, USA</i>	
<i>Tamrah D. Cunningham, New York University, USA</i>	
<i>Zianne Cuff, City University of New York, USA</i>	
Developments in MOOC Technologies and Participation Since 2012	7888
<i>Jeremy Riel, University of Illinois at Chicago, USA</i>	
<i>Kimberly A. Lawless, University of Illinois at Chicago, USA</i>	
Massive Open Online Courses and Integrating Open Source Technology and Open Access Literature Into Technology-Based Degrees	7898
<i>Maurice Dawson, University of Missouri – St. Louis, USA</i>	
<i>Sharon Burton, Grand Canyon University, USA</i>	
<i>Dustin Bessette, National Graduate School of Quality Management, USA</i>	
<i>Jorja Wright, University of Charleston, USA</i>	
Open Source Software Virtual Learning Environment (OSS-VLEs) in Library Science Schools ...	7912
<i>Rosy Jan, University of Kashmir, India</i>	
Teacher Presence.....	7922
<i>Caroline M. Crawford, University of Houston – Clear Lake, USA</i>	
Virtual Worlds in the Educational Context	7935
<i>Felipe Becker Nunes, Federal University of Rio Grande do Sul, Brazil</i>	
<i>Fabício Herpich, Federal University of Rio Grande do Sul, Brazil</i>	
<i>Leo Natan Paschoal, University of Cruz Alta, Brazil</i>	

Category: Web Technologies

Anger and Internet in Japan	7946
<i>Hiroko Endo, Rissho University, Japan</i>	
<i>Kei Fuji, University of Tsukuba, Japan</i>	
Determine Democracy in Web Design	7956
<i>Rowena Li, Bayside High School Library, USA</i>	
Discussion Processes in Online Forums	7969
<i>Gaowei Chen, The University of Hong Kong, Hong Kong</i>	
<i>Ming M Chiu, The Education University of Hong Kong, Hong Kong</i>	
The Economics of Internetization.....	7980
<i>Constantine E. Passaris, University of New Brunswick, Canada</i>	

An Efficient and Effective Index Structure for Query Evaluation in Search Engines.....	7995
<i>Yangjun Chen, University of Winnipeg, Canada</i>	
Improving Usability of Website Design Using W3C Guidelines	8006
<i>G. Sreedhar, Rashtriya Sanskrit Vidyapeetha (Deemed University), India</i>	
Internet Phenomenon	8015
<i>Lars Konzack, University of Copenhagen, Denmark</i>	
An Overview of Crowdsourcing	8023
<i>Eman Younis, Minia University, Egypt</i>	
Revisiting Web 2.0	8036
<i>Michael Dinger, University of South Carolina Upstate, USA</i>	
<i>Varun Grover, Clemson University, USA</i>	
Search Engine Optimization	8046
<i>Dimitrios Giomelakis, Aristotle University of Thessaloniki, Greece</i>	
<i>Andreas A. Veglis, Aristotle University of Thessaloniki, Greece</i>	
Toward Trustworthy Web Services Coordination.....	8056
<i>Wenbing Zhao, Cleveland State University, USA</i>	
Usability of CAPTCHA in Online Communities and Its Link to User Satisfaction.....	8066
<i>Samar I. Swaid, Philander Smith College, USA</i>	
Visual Identity Design for Responsive Web	8079
<i>Sunghyun Ryoo Kang, Iowa State University, USA</i>	
<i>Debra Satterfield, California State University – Long Beach, USA</i>	
Web Site Mobilization Techniques	8087
<i>John Christopher Sandvig, Western Washington University, USA</i>	
What Accounts for the Differences in Internet Diffusion Rates Around the World?.....	8095
<i>Ravi Nath, Creighton University, USA</i>	
<i>Vasudeva Murthy, Creighton University, USA</i>	
About the Contributors	clxxii
Index.....	clxxiii

Table of Contents in Alphabetical Order

*Volume I, pp. 1-808 · Volume II, pp. 809-1615 · Volume III, pp. 1616-2420 · Volume IV, pp. 2421-3236 · Volume V, pp. 3237-4037;
Volume VI, pp. 4038-4848 · Volume VII, pp. 4849-5668 · Volume VIII, pp. 5669-6475 · Volume IX, pp. 6476-7280 · Volume X, pp. 7281-8104*

Acceptance of E-Reverse Auction From the Buyer Perspective.....	530
<i>Cigdem Altin Gumussoy, Istanbul Technical University, Turkey</i>	
<i>Bilal Gumussoy, Shell and Turcas Petrol Inc., Turkey</i>	
Accessibility in E-Government.....	3516
<i>Christian Sonnenberg, Florida Institute of Technology, USA</i>	
Adapting Big Data Ecosystem for Landscape of Real World Applications	326
<i>Jyotsna Talreja Wassan, University of Delhi, India</i>	
Adaptive Hypermedia in Education.....	2357
<i>Vehbi Turel, The University of Bingol, Turkey</i>	
Adaptive Hypermedia Systems	6424
<i>Ana Carolina Tomé Klock, Federal University of Rio Grande do Sul (UFRGS), Brazil</i>	
<i>Isabela Gasparini, Santa Catarina State University (UDESC), Brazil</i>	
<i>Marcelo Soares Pimenta, Federal University of Rio Grande do Sul (UFRGS), Brazil</i>	
<i>José Palazzo M. de Oliveira, Federal University of Rio Grande do Sul (UFRGS), Brazil</i>	
Adaptive Networks for On-Chip Communication	4549
<i>Mário Pereira Vestias, Instituto Politécnico de Lisboa, Portugal</i>	
Addressing Digital Competencies, Curriculum Development, and Instructional Design in Science Teacher Education.....	1420
<i>Isha DeCoito, Western University, Canada</i>	
Adolescents' Food Communication in Social Media:	6940
<i>Christopher Holmberg, University of Gothenburg, Sweden</i>	
Adopting Open Source Software in Smartphone Manufacturers' Open Innovation Strategy	7369
<i>Mohammad Nabil Almunawar, Universiti Brunei Darussalam, Brunei</i>	
<i>Muhammad Anshari, Universiti Brunei Darussalam, Brunei</i>	
<i>Heru Susanto, Indonesian Institute of Sciences, Indonesia & Tunghai University, Taiwan</i>	
The Adoption and Transformation of Capability Maturity Models in Government.....	3526
<i>Terry F. Buss, Carnegie Mellon University, Australia</i>	

Adoption and Use of Mobile Money Services in Nigeria.....	2724
<i>Olayinka David-West, Pan-Atlantic University, Nigeria</i>	
<i>Immanuel Ovemeso Umukoro, Pan-Atlantic University, Nigeria</i>	
<i>Omotayo Muritala, Pan-Atlantic University, Nigeria</i>	
Advanced ICT Methodologies (AIM) in the Construction Industry	539
<i>M. Reza Hosseini, Deakin University, Australia</i>	
<i>Saeed Banihashemi, University of Technology Sydney, Australia</i>	
<i>Fahimeh Zaeri, Auckland University of Technology, New Zealand</i>	
<i>Alireza Adibfar, University of Florida, USA</i>	
Advanced Model of Complex Information System	4391
<i>Miroslav Svitek, Czech Technical University in Prague, Czech Republic</i>	
Advanced Recommender Systems	1735
<i>Young Park, Bradley University, USA</i>	
Advancement and Application of Scientometric Indicators for Evaluation of Research Content....	6739
<i>Tazeem Zainab, University of Kashmir, India</i>	
<i>Zahid Ashraf Wani, University of Kashmir, India</i>	
Affect-Sensitive Computer Systems	4124
<i>Nik Thompson, Curtin University, Australia</i>	
<i>Tanya McGill, Murdoch University, Australia</i>	
<i>David Murray, Murdoch University, Australia</i>	
Agent-Based Social Networks	6950
<i>Federico Bergenti, Università degli Studi di Parma, Italy</i>	
<i>Agostino Poggi, Università degli Studi di Parma, Italy</i>	
<i>Michele Tomaiuolo, Università degli Studi di Parma, Italy</i>	
Amplifying the Significance of Systems Thinking in Organization	551
<i>Mambo Governor Mupepi, Grand Valley State University, USA</i>	
<i>Sylvia C. Mupepi, Grand Valley State University, USA</i>	
<i>Jaideep Motwani, Grand Valley State University, USA</i>	
Analysis and Assessment of Cross-Language Question Answering Systems	4471
<i>Juncal Gutiérrez-Artacho, University of Granada, Spain</i>	
<i>María-Dolores Olvera-Lobo, University of Granada, Spain</i>	
Analysis of Two Phases Queue With Vacations and Breakdowns Under T-Policy	1570
<i>Khalid Alnowibet, King Saud University, Saudi Arabia</i>	
<i>Lotfi Tadj, Fairleigh Dickinson University, Canada</i>	
Anger and Internet in Japan	7946
<i>Hiroko Endo, Rissho University, Japan</i>	
<i>Kei Fujii, University of Tsukuba, Japan</i>	

The Application of Crowdsourced Processes in a Business Environment	563
<i>Katarzyna Kopeć, Tischner European University, Poland</i>	
<i>Anna Szopa, Jagiellonian University, Poland</i>	
Application of Fuzzy Numbers to Assessment Processes	3215
<i>Michael Voskoglou, Graduate Technological Educational Institute (T.E.I.), Greece</i>	
Application of Gamification to Blended Learning in Higher Education	3238
<i>Kamini Jaipal-Jamani, Brock University, Canada</i>	
<i>Candace Figg, Brock University, Canada</i>	
Application of Geospatial Mashups in Web GIS for Tourism Development	3403
<i>Somnath Chaudhuri, Maldives National University, Maldives</i>	
<i>Nilanjan Ray, Adamas University, India</i>	
Application of Soft Set in Game Theory	3226
<i>B. K. Tripathy, VIT University, India</i>	
<i>Sooraj T. R., VIT University, India</i>	
<i>Radhakrishna N. Mohanty, VIT University, India</i>	
Applications of Artificial Neural Networks in Economics and Finance.....	6631
<i>Iva Mihaylova, University of St. Gallen, Switzerland</i>	
Applying Artificial Intelligence to Financial Investing.....	1
<i>Hayden Wimmer, Georgia Southern University, USA</i>	
<i>Roy Rada, University of Maryland – Baltimore County, USA</i>	
Apps as Assistive Technology	266
<i>Emily C. Bouck, Michigan State University, USA</i>	
<i>Sara M. Flanagan, University of Kentucky, USA</i>	
<i>Missy D. Cosby, Michigan State University, USA</i>	
Archaeological GIS for Land Use in South Etruria Urban Revolution in IX-VIII Centuries B.C....	3419
<i>Giuliano Pelfer, University of Florence, Italy</i>	
Architectural Framework for the Implementation of Information Technology Governance in Organisations	810
<i>Thami Batyashe, Cape Peninsula University of Technology, South Africa</i>	
<i>Tiko Iyamu, Cape Peninsula University of Technology, South Africa</i>	
Architecture as a Tool to Solve Business Planning Problems	573
<i>James McKee, Independent Researcher, Australia</i>	
Architecture of an Open-Source Real-Time Distributed Cyber Physical System.....	1227
<i>Stefano Scanzio, CNR-IEIIT, Italy</i>	

Are Social Marketing Investments Used as a Tool for Voluntary Reporting or Disclosure?	5737
<i>Tugba Ucma Uysal, Mugla Sitki Kocman University, Turkey</i>	
<i>Ganite Kurt, Gazi University, Turkey</i>	
<i>Ali Naci Karabulut, Mugla Sitki Kocman University, Turkey</i>	
Artificial Ethics	88
<i>Laura L. Pană, Polytechnic University of Bucharest, Romania</i>	
Artificial Intelligence	98
<i>Steven Walczak, University of South Florida, USA</i>	
Artificial Intelligence Review	106
<i>Amal Kilani, University of Gabes Tunisia, Tunisia</i>	
<i>Ahmed Ben Hamida, University of Sfax, Tunisia</i>	
<i>Habib Hamam, University of Moncton, Canada</i>	
Artificial Neural Networks.....	120
<i>Steven Walczak, University of South Florida, USA</i>	
Artificial Neural Networks and Their Applications in Business.....	6642
<i>Trevor J. Bihl, Air Force Institute of Technology, USA</i>	
<i>William A. Young, Ohio University, USA</i>	
<i>Gary R. Weckman, Ohio University, USA</i>	
Aspects of Various Community Detection Algorithms in Social Network Analysis	6961
<i>Nicole Belinda Dillen, St. Thomas' College of Engineering and Technology, India</i>	
<i>Aruna Chakraborty, St. Thomas' College of Engineering and Technology, India</i>	
Assessing Computer-Aided Design Skills	7382
<i>Yi Lin Wong, The Hong Kong Polytechnic University, Hong Kong</i>	
<i>Kin Wai Michael Siu, The Hong Kong Polytechnic University, Hong Kong</i>	
Assistive Technology and Human Capital for Workforce Diversity.....	277
<i>Ben Tran, Alliant International University, USA</i>	
Assistive Technology for Supporting Communication, Occupation, and Leisure by Children With Severe to Profound Developmental Disabilities	287
<i>Fabrizio Stasolla, University of Bari, Italy</i>	
<i>Viviana Perilli, Lega del Filo d'Oro – Molfetta, Italy</i>	
<i>Adele Boccasini, Lega del Filo d'Oro – Termini Imerese, Italy</i>	
Augmented Reality for Tourist Destination Image Formation	4031
<i>Azizul Hassan, Cardiff Metropolitan University, UK</i>	

Automatic Emotion Recognition Based on Non-Contact Gaits Information.....	132
<i>Jingying Wang, University of Chinese Academy of Sciences, China</i>	
<i>Baobin Li, University of Chinese Academy of Sciences, China</i>	
<i>Changye Zhu, University of Chinese Academy of Sciences, China</i>	
<i>Shun Li, University of Chinese Academy of Sciences, China</i>	
<i>Tingshao Zhu, University of Chinese Academy of Sciences, China</i>	
Automatic Item Generation.....	2369
<i>Mark Gierl, University of Alberta, Canada</i>	
<i>Hollis Lai, University of Alberta, Canada</i>	
<i>Xinxin Zhang, University of Alberta, Canada</i>	
Autonomic Cooperative Communications.....	6499
<i>Michal Wodczak, Samsung Electronics, Poland</i>	
Barcodes vs. RFID and Its Continued Success in Manufacturing and Services.....	5273
<i>Amber A. Smith-Ditizio, Texas Woman's University, USA</i>	
<i>Alan D. Smith, Robert Morris University, USA</i>	
Becoming Smart, Innovative, and Socially Responsible in Supply Chain Collaboration	5285
<i>Goknur Arzu Akyuz, University of Turkish Aeronautical Association, Turkey</i>	
<i>Guner Gursoy, Okan University, Turkey</i>	
Benchmarking Performance Indicators of Indian Rail Freight by DEA Approach	587
<i>Neeraj Bhanot, Dr. B. R. Ambedkar National Institute of Technology Jalandhar, India</i>	
<i>Harwinder Singh, Guru Nanak Dev Engineering College, India</i>	
Bi-Directional Business/IT Alignment	601
<i>Hashim Chunpir, German Climate Computing Centre (DKRZ), Germany</i>	
<i>Frederik Schulte, University of Hamburg, Germany</i>	
<i>Yannick Bartens, University of Hamburg, Germany</i>	
<i>Stefan D. Voß, University of Hamburg, Germany</i>	
Big Data Analysis and Mining.....	338
<i>Carson K. Leung, University of Manitoba, Canada</i>	
Big Data Analytics for Tourism Destinations	349
<i>Wolfram Höpken, University of Applied Sciences Ravensburg-Weingarten, Germany</i>	
<i>Matthias Fuchs, Mid-Sweden University, Sweden</i>	
<i>Maria Lexhagen, Mid-Sweden University, Sweden</i>	
Big Data and Simulations for the Solution of Controversies in Small Businesses.....	6907
<i>Milena Janakova, Silesian University in Opava, Czech Republic</i>	
Big Data, Knowledge, and Business Intelligence	943
<i>G. Scott Erickson, Ithaca College, USA</i>	
<i>Helen N. Rothberg, Marist College, USA</i>	

Big Data Time Series Stream Data Segmentation Methods	364
<i>Dima Alberg, Shamoon College of Engineering (SCE), Israel</i>	
Binary Decision Diagram Reliability for Multiple Robot Complex System	6825
<i>Hamed Fazlollahtabar, Sharif University of Technology, Iran & National Elites Foundation, Iran</i>	
<i>Seyed Taghi Akhavan Niaki, Sharif University of Technology, Iran</i>	
Biogeography-Based Optimization Applied to Wireless Communications Problems	5967
<i>Sotirios K. Goudos, Aristotle University of Thessaloniki, Greece</i>	
Bioinformatics.....	419
<i>Mark A. Ragan, The University of Queensland, Australia</i>	
A Bio-Inspired, Distributed Control Approach to the Design of Autonomous Cooperative Behaviors in Multiple Mobile Robot Systems	6836
<i>Gen'ichi Yasuda, Nagasaki Institute of Applied Science, Japan</i>	
Bioinspired Solutions for MEMS Tribology	431
<i>R. Arvind Singh, Kumaraguru College of Technology (KCT), India</i>	
<i>S. Jayalakshmi, Kumaraguru College of Technology (KCT), India</i>	
Bipolar Model in Collective Choice	7282
<i>Ayeley P. Tchangani, Université Fédérale Toulouse Midi-Pyrénées, France</i>	
Board Games AI	144
<i>Tad Gonsalves, Sophia University, Japan</i>	
Boosting the Social Development of the Majority Through the Creation of a Wireless Knowledge Society.....	5015
<i>Danilo Piaggese, Framericas, USA</i>	
Bridging Between Cyber Politics and Collective Dynamics of Social Movement.....	3538
<i>Kazuhiko Shibuya, ROIS, Japan</i>	
Building Gene Networks by Analyzing Gene Expression Profiles.....	440
<i>Crescenzo Gallo, University of Foggia, Italy</i>	
Business Intelligence	951
<i>Richard T. Herschel, Saint Joseph's University, USA</i>	
Business Intelligence Impacts on Design of Enterprise Systems	2932
<i>Saeed Rouhani, University of Tehran, Iran</i>	
<i>Dusanka Milorad Lecic, Levi9 Global Sourcing Balkan, Serbia</i>	
Business Model Innovation-Oriented Technology Management for Emergent Technologies	4560
<i>Sven Seidenstricker, Fraunhofer Institute for Industrial Engineering, Germany</i>	
<i>Ardilio Antonino, Fraunhofer Institute for Industrial Engineering, Germany</i>	

Business Sustainability Indices.....	609
<i>Arunasalam Sambhanthan, Curtin University, Australia</i>	
The Business Transformation Framework and Its Business Engineering Law Support for (e)Transactions.....	636
<i>Antoine Trad, Institute of Business and Information Systems Transformation Management, Switzerland</i>	
<i>Damir Kalpić, University of Zagreb, Croatia</i>	
The Business Transformation Framework for Managers in Transformation Projects	5607
<i>Antoine Trad, IBITSM, Switzerland</i>	
<i>Damir Kalpić, University of Zagreb, Croatia</i>	
The Business Transformation Framework, Agile Project and Change Management	620
<i>Antoine Trad, Institute of Business and Information Systems Transformation Management, Switzerland</i>	
<i>Damir Kalpić, University of Zagreb, Croatia</i>	
BYOD (Bring Your Own Device), Mobile Technology Providers, and Its Impacts on Business/ Education and Workplace/Learning Applications	5981
<i>Amber A. Smith-Ditizio, Texas Woman's University, USA</i>	
<i>Alan D. Smith, Robert Morris University, USA</i>	
Carbon Capture From Natural Gas via Polymeric Membranes	3043
<i>Nayef Mohamed Ghasem, UAE University, UAE</i>	
<i>Nihmiya Abdul Rahim, UAE University, UAE</i>	
<i>Mohamed Al-Marzouqi, UAE University, UAE</i>	
Cell Phone Conversation and Relative Crash Risk Update.....	5992
<i>Richard A. Young, Driving Safety Consulting, LLC, USA</i>	
Censorship in the Digital Age the World Over	7292
<i>Kari D. Weaver, University of South Carolina – Aiken, USA</i>	
Centrality Analysis of the United States Network Graph	1746
<i>Natarajan Meghanathan, Jackson State University, USA</i>	
Challenges and Implications of Health Literacy in Global Health Care	3734
<i>Kijpokin Kasemsap, Suan Sunandha Rajabhat University, Thailand</i>	
Challenges for Big Data Security and Privacy	373
<i>M. Govindarajan, Annamalai University, India</i>	
Challenges in Collecting Qualitative Data for Information Systems Studies	4480
<i>Tiko Iyamu, Cape Peninsula University of Technology, South Africa</i>	
<i>Irja Naambo Shaanika, Namibia University of Science and Technology, Namibia</i>	

Challenges in Developing Adaptive Educational Hypermedia Systems	2380
<i>Eileen O'Donnell, Trinity College Dublin, Ireland</i>	
<i>Liam O'Donnell, Dublin Institute of Technology, Ireland</i>	
Challenges of Meta Access Control Model Enforcement to an Increased Interoperability.....	651
<i>Sérgio Luís Guerreiro, University of Lisbon, Portugal</i>	
The Challenges of Teaching and Learning Software Programming to Novice Students.....	7392
<i>Seyed Reza Shahamiri, Manukau Institute of Technology, New Zealand</i>	
Change Leadership Styles and Behaviors in Academic Libraries	5194
<i>John Kennedy Lewis, Salve Regina University, USA</i>	
Changing Expectations of Academic Libraries	5204
<i>Jennifer Ashley Wright Joe, Western Kentucky University, USA</i>	
Chemistry Learning Through Designing Digital Games.....	3248
<i>Kamisah Osman, The National University of Malaysia, Malaysia</i>	
<i>Ah-Nam Lay, Institute of Teacher Education – Sultan Abdul Halim, Malaysia</i>	
Classification of Traffic Events Notified in Social Networks' Texts.....	6973
<i>Ana Maria Magdalena Saldana-Perez, Instituto Politecnico Nacional, Mexico</i>	
<i>Marco Antonio Moreno-Ibarra, Instituto Politécnico Nacional, Mexico</i>	
<i>Miguel Jesus Torres-Ruiz, Instituto Politécnico Nacional, Mexico</i>	
Climate Change as a Driving Force on Urban Energy Consumption Patterns	7815
<i>Mostafa Jafari, Agricultural Research Education and Extension Organization (AREEO), Iran</i>	
<i>Pete Smith, University of Aberdeen, UK</i>	
Clinical Use of Video Games.....	3260
<i>Ben Tran, Alliant International University, USA</i>	
Clique Size and Centrality Metrics for Analysis of Real-World Network Graphs.....	6507
<i>Natarajan Meghanathan, Jackson State University, USA</i>	
Cloud Computing.....	1026
<i>Eduardo Correia, Christchurch Polytechnic Institute of Technology (CPIT), New Zealand</i>	
Cloud Governance at the Local Communities	1033
<i>Vasileios Yfantis, Ionian University, Greece</i>	
Clouds of Quantum Machines	1040
<i>Nilo Sylvio Serpa, Universidade Paulista, Brazil</i>	
Cognitive and Psychological Factors in Cross-Language Information Retrieval	4490
<i>Rowena Li, Bayside High School Library, USA</i>	

Cognitive Ergonomics in 2016	662
<i>Ronald John Lofaro, Embry-Riddle Aeronautical University, USA</i>	
Cognitive Mapping in Support of Intelligent Information Systems	4570
<i>Akbar Esfahanipour, Amirkabir University of Technology, Iran</i>	
<i>Ali Reza Montazemi, McMaster University, Canada</i>	
Cognitive Process Elements of People Decision-Making.....	2076
<i>Thais Spiegel, Rio de Janeiro State University, Brazil</i>	
Comb Filters Characteristics and Current Applications	6007
<i>Miriam Guadalupe Cruz-Jimenez, Institute INAOE, Mexico</i>	
<i>David Ernesto Troncoso Romero, CONACYT at ESCOM-IPN, Mexico</i>	
<i>Gordana Jovanovic Dolecek, Institute INAOE, Mexico</i>	
Communication, Information, and Pragmatics	1186
<i>Adriana Braga, Pontifical Catholic University of Rio de Janeiro, Brazil</i>	
<i>Robert K. Logan, University of Toronto, Canada</i>	
Communication Privacy Management and Mediated Communication	6985
<i>Debra L. Worthington, Auburn University, USA</i>	
<i>Margaret Fitch-Hauser, Auburn University, USA</i>	
Communities of Practice as a Source of Open Innovation	5027
<i>Diane-Gabrielle Tremblay, University of Quebec, Canada</i>	
Community Broadband Networks and the Opportunity for E-Government Services	3549
<i>Idongesit Williams, Aalborg University, Denmark</i>	
Community Outreach.....	6685
<i>Loriene Roy, The University of Texas at Austin, USA</i>	
<i>Antonia Frydman, The University of Texas at Austin, USA</i>	
Community Science and Technology and Its Meaning to Potential Requirement.....	7201
<i>P. K. Paul, Raiganj University, India</i>	
<i>A. Bhumali, Raiganj University, India</i>	
Comprehensible Explanation of Predictive Models.....	2085
<i>Marko Robnik-Šikonja, University of Ljubljana, Slovenia</i>	
Comprehensive E-Learning Appraisal System	5787
<i>Jose Luis Monroy Anton, La Ribera University Hospital, Spain</i>	
<i>Juan Vicente Izquierdo Soriano, La Ribera University Hospital, Spain</i>	
<i>Maria Isabel Asensio Martinez, La Ribera University Hospital, Spain</i>	
<i>Felix Buendia Garcia, Polytechnical University of Valencia, Spain</i>	

Computational Intelligence Approaches to Computational Aesthetics	156
<i>Erandi Lakshika, University of New South Wales, Australia</i>	
<i>Michael Barlow, University of New South Wales, Australia</i>	
Computational Thinking in Innovative Computational Environments and Coding	2392
<i>Alberto Ferrari, University of Parma, Italy</i>	
<i>Agostino Poggi, University of Parma, Italy</i>	
<i>Michele Tomaiuolo, University of Parma., Italy</i>	
Computer Agent Technologies in Collaborative Learning and Assessment.....	2402
<i>Yigal Rosen, Harvard University, USA</i>	
Computer-Assisted Indian Matrimonial Services	4136
<i>Robert Leslie Fisher, Independent Researcher, USA</i>	
Computer-Assisted Parallel Program Generation	4583
<i>Shigeo Kawata, Utsunomiya University, Japan</i>	
Computer Fraud Challenges and Its Legal Implications.....	4837
<i>Amber A. Smith-Ditizio, Texas Woman's University, USA</i>	
<i>Alan D. Smith, Robert Morris University, USA</i>	
Computer Information Library Clusters	4399
<i>Fu Yuhua, CNOOC Research Institute, China</i>	
Computing Technologies and Science Fiction Cinema	3349
<i>Rocío Carrasco-Carrasco, University of Huelva, Spain</i>	
Concept and Practices of Cyber Supply Chain in Manufacturing Context.....	5306
<i>Anisha Banu Dawood Gani, Universiti Sains Malaysia, Malaysia</i>	
<i>Yudi Fernando, Universiti Malaysia Pahang, Malaysia</i>	
The Concept of Modularity in the Context of IS Project Outsourcing	5317
<i>Shahzada Benazeer, University of Antwerp, Belgium</i>	
<i>Philip Huysmans, University of Antwerp, Belgium</i>	
<i>Peter De Bruyn, University of Antwerp, Belgium</i>	
<i>Jan Verelst, University of Antwerp, Belgium</i>	
The Concept of the Shapley Value and the Cost Allocation Between Cooperating Participants	2095
<i>Alexander Kolker, GE Healthcare, USA</i>	
Concerns and Challenges of Cloud Platforms for Bioinformatics.....	455
<i>Nicoletta Dessì, Università degli Studi di Cagliari, Italy</i>	
<i>Barbara Pes, Università degli Studi di Cagliari, Italy</i>	
Consistency Is Not Enough in Byzantine Fault Tolerance	1238
<i>Wenbing Zhao, Cleveland State University, USA</i>	

Constrained Nonlinear Optimization in Information Science.....	4594
<i>William P. Fox, Naval Postgraduate School, USA</i>	
Constructing Preservice Teachers' Knowledge of Technology Integration.....	7623
<i>Kathleen A. Paciga, Columbia College Chicago, USA</i>	
<i>Angela Fowler, Erikson Institute, USA</i>	
<i>Mary Quest, Erikson Institute, USA</i>	
Consumer Adoption of PC-Based/Mobile-Based Electronic Word-of-Mouth	6019
<i>Akinori Ono, Keio University, Japan</i>	
<i>Mai Kikumori, Ritsumeikan University, Japan</i>	
Contemporary Leadership Development in Kazakhstan	5626
<i>Gainiya Tazhina, University of International Business, Kazakhstan</i>	
<i>Judith Parker, Teachers College, Columbia University, USA</i>	
<i>Arslan Ivashov, Kazakh Ablai Khan University of International Relations and World Languages, USA</i>	
Context-Aware Approach for Restaurant Recommender Systems.....	1757
<i>Haoxian Feng, University of Ottawa, Canada</i>	
<i>Thomas Tran, University of Ottawa, Canada</i>	
Context-Aware Personalization for Mobile Services	6031
<i>Abayomi Moradeyo Otebolaku, Liverpool John Moores University, UK</i>	
<i>Maria Teresa Andrade, University of Porto, Portugal</i>	
Continuous Assurance and the Use of Technology for Business Compliance	820
<i>Rui Pedro Figueiredo Marques, Universidade de Aveiro, Portugal</i>	
Corporate Disclosure Measurement.....	1896
<i>Md. Salah Uddin Rajib, Jahangirnagar University, Bangladesh</i>	
<i>Md. Qutub Uddin Sajib, China University of Geosciences (CUG), China</i>	
Corporate Social Responsibility	671
<i>Ben Tran, Alliant International University, USA</i>	
Cost-Effective 3D Stereo Visualization for Creative Learning.....	2411
<i>R. S. Kamath, Chatrapati Shahu Institute of Business Education and Research, India</i>	
<i>R. K. Kamat, Shivaji University, India</i>	
Cost Estimation and Security Investment of Security Projects.....	4849
<i>Yosra Miaoui, University of Carthage, Tunisia</i>	
<i>Boutheina Fessi, University of Carthage, Tunisia</i>	
<i>Nouredine Boudriga, University of Carthage, Tunisia</i>	

Cost Evaluation of Synchronization Algorithms for Multicore Architectures	3989
<i>Masoud Hemmatpour, Politecnico di Torino, Italy</i>	
<i>Renato Ferrero, Politecnico di Torino, Italy</i>	
<i>Filippo Gandino, Politecnico di Torino, Italy</i>	
<i>Bartolomeo Montrucchio, Politecnico di Torino, Italy</i>	
<i>Maurizio Rebaudengo, Politecnico di Torino, Italy</i>	
Could Educational Technology Replace Traditional Schools in the Future?	2421
<i>John K. Hope, University of Auckland, New Zealand</i>	
Creating Active Learning Spaces in Virtual Worlds.....	7880
<i>Reneta D. Lansiquot, City University of New York, USA</i>	
<i>Tamrah D. Cunningham, New York University, USA</i>	
<i>Zianne Cuff, City University of New York, USA</i>	
Creative Collaborative Virtual Environments.....	4146
<i>Luís Eustáquio, Universidade do Porto, Portugal</i>	
<i>Catarina Carneiro de Sousa, Polytechnic Institute of Viseu, Portugal</i>	
Crisis Response and Management	1396
<i>Sergey V. Zykov, National Research University Higher School of Economics, Russia</i>	
A Critical Overview of Image Segmentation Techniques Based on Transition Region.....	1308
<i>Yu-Jin Zhang, Tsinghua University, China</i>	
Critical Success Factors in E-Democracy Implementation	3561
<i>Aderonke A. Oni, Covenant University, Nigeria</i>	
<i>Adekunle O. Okunoye, Xavier University, USA</i>	
Cuckoo Search Algorithm for Solving Real Industrial Multi-Objective Scheduling Problems	4369
<i>Mariappan Kadarkarainadar Marichelvam, Mepco Schlenk Engineering College, India</i>	
<i>Mariappan Geetha, Kamaraj College of Engineering and Technology, India</i>	
Current Scenario of Youth Entrepreneurship in India	2989
<i>Neeta Baporikar, Namibia University of Science and Technology, Namibia & University of Pune, India</i>	
Customer Lifetime Value	1584
<i>Kijpokin Kasemsap, Suan Sunandha Rajabhat University, Thailand</i>	
Cyber Behaviors in Seeking Health Information.....	3745
<i>Xiaojun (Jenny) Yuan, University of Albany, USA</i>	
<i>José A. Pino, Universidad de Chile, Chile</i>	
Cyber Bullying.....	1695
<i>Jo Ann Oravec, University of Wisconsin – Whitewater, USA</i>	

Cyberbullying	4157
<i>Gilberto Marzano, Rezekne Academy of Technologies, Latvia</i>	
Cyberbullying Among Malaysian Children Based on Research Evidence.....	1704
<i>Sarina Yusuf, Universiti Putra Malaysia, Malaysia</i>	
<i>Md. Salleh Hj. Hassan, Universiti Putra Malaysia, Malaysia</i>	
<i>Adamkolo Mohammed Mohammed Ibrahim, Universiti Putra Malaysia, Malaysia & University of Maiduguri, Nigeria</i>	
Cyberinfrastructure, Cloud Computing, Science Gateways, Visualization, and Cyberinfrastructure Ease of Use	1063
<i>Craig A. Stewart, Indiana University, USA</i>	
<i>Richard Knepper, Indiana University, USA</i>	
<i>Matthew R. Link, Indiana University, USA</i>	
<i>Marlon Pierce, Indiana University, USA</i>	
<i>Eric Wernert, Indiana University, USA</i>	
<i>Nancy Wilkins-Diehr, San Diego Supercomputer Center, USA</i>	
Cyberloafing and Constructive Recreation	4316
<i>Jo Ann Oravec, University of Wisconsin – Whitewater, USA</i>	
Cyber Security Protection for Online Gaming Applications	1647
<i>Wenbing Zhao, Cleveland State University, USA</i>	
Database Techniques for New Hardware	1947
<i>Xiongpai Qin, Renmin University of China, China</i>	
<i>Yueguo Chen, Renmin University of China, China</i>	
Data-Centric Benchmarking	1772
<i>Jérôme Darmont, Université de Lyon, Lyon 2, ERIC EA3083, France</i>	
Data Journalism	1196
<i>Andreas A. Veglis, Aristotle University of Thessaloniki, Greece</i>	
<i>Charalampos P. Bratsas, Aristotle University of Thessaloniki, Greece</i>	
Data Linkage Discovery Applications	1783
<i>Richard S. Segall, Arkansas State University, USA</i>	
<i>Shen Lu, University of South Florida, USA</i>	
Data Mining and Knowledge Discovery in Databases.....	1907
<i>Ana Azevedo, Polytechnic Institute of Porto, Portugal</i>	
Data Mining and the KDD Process.....	1919
<i>Ana Funes, Universidad Nacional de San Luis, Argentina</i>	
<i>Aristides Dasso, Universidad Nacional de San Luis, Argentina</i>	

Data Mining to Identify Project Management Strategies in Learning Environments.....	1934
<i>Ana González-Marcos, Universidad de La Rioja, Spain</i>	
<i>Joaquín Ordieres-Meré, Universidad Politécnica de Madrid, Spain</i>	
<i>Fernando Alba-Elías, Universidad de La Rioja, Spain</i>	
Data Visualization Strategies for Computer Simulation in Bioelectromagnetics.....	1249
<i>Akram Gasmelseed, Qassim University, Saudi Arabia</i>	
<i>Ali H. Alharbi, Qassim University, Saudi Arabia</i>	
Decimal Hardware Multiplier	4607
<i>Mário Pereira Vestias, INESC-ID/ISEL/IPL, Portugal</i>	
Decision Filed Theory	2108
<i>Lan Shao, University of Oulu, Finland</i>	
<i>Jouni Markkula, University of Oulu, Finland</i>	
Defining and Characterising the Landscape of eHealth.....	5864
<i>Yvonne O'Connor, University College Cork, Ireland</i>	
<i>Ciara Heavin, University College Cork, Ireland</i>	
Defining and Conceptualizing Cyberbullying	4168
<i>Karin Spenser, Nottingham Trent University, UK</i>	
<i>Lucy R. Betts, Nottingham Trent University, UK</i>	
Deployment of Enterprise Architecture From the Activity Theory Perspective	2943
<i>Tiko Iyamu, Cape Peninsula University of Technology, South Africa</i>	
<i>Irja Naambo Shaanika, Namibia University of Science and Technology – Windhoek, Namibia</i>	
Designing Engaging Instruction for the Adult Learners.....	1432
<i>Karen Swanson, Mercer University, USA</i>	
<i>Geri Collins, Mercer University, USA</i>	
Design, Manufacture, and Selection of Ankle-Foot-Orthoses.....	298
<i>Hasan Kemal Surmen, Istanbul University, Turkey</i>	
<i>Nazif Ekin Akalan, Istanbul University, Turkey</i>	
<i>Yunus Ziya Arslan, Istanbul University, Turkey</i>	
Design of Compensators for Comb Decimation Filters.....	6043
<i>Gordana Jovanovic Dolecek, Institute INAOE Puebla, Mexico</i>	
Destination @-Branding of Ten European Capitals Through the Institutional Stems and Commercial Logos.....	4038
<i>Elena Bocci, Sapienza University of Rome, Italy</i>	
<i>Annamaria Silvana de Rosa, Sapienza University of Rome, Italy</i>	
<i>Laura Dryjanska, Sapienza University of Rome, Italy</i>	

Determination of Urban Growth by the Night-Time Images	7831
<i>Emre Yücer, Erzincan University, Turkey</i>	
<i>Arzu Erener, Kocaeli University, Turkey</i>	
Determine Democracy in Web Design	7956
<i>Rowena Li, Bayside High School Library, USA</i>	
Determining Impact of Demographics on Perceived Service Quality in Online Retail	2882
<i>Prateek Kalia, I. K. Gujral Punjab Technical University, India</i>	
<i>Penny Law, Regenesys Business School, South Africa</i>	
<i>Richa Arora, Regenesys Institute of Management, India</i>	
Developing a Glossary for Software Projects	7399
<i>Tamer Abdou, Suez Canal University, Egypt</i>	
<i>Pankaj Kamthan, Concordia University, Canada</i>	
<i>Nazlie Shahmir, WestJet Airlines Limited, Canada</i>	
Developing Creativity and Learning Design by Information and Communication Technology (ICT) in Developing Contexts.....	4178
<i>Chunfang Zhou, Aalborg University, Denmark</i>	
<i>Aparna Purushothaman, Aalborg University, Denmark</i>	
Developing Global Supply Chain Manager for Business Expansion.....	5327
<i>Puspita Wulansari, Telkom University, Indonesia</i>	
<i>Yudi Fernando, Universiti Malaysia Pahang, Malaysia</i>	
Development of Communication Skills through Auditory Training Software in Special Education	2431
<i>Eduardo C. Contreras, Autonomous University of Coahuila, Mexico</i>	
<i>Isis I. Contreras, Saltillo Institute of Technology, Mexico</i>	
Development of Image Engineering in the Last 20 Years	1319
<i>Yu-Jin Zhang, Tsinghua University, China</i>	
Development of Personal Information Privacy Concerns Evaluation	4862
<i>Anna Rohunen, University of Oulu, Finland</i>	
<i>Jouni Markkula, University of Oulu, Finland</i>	
Developments in MOOC Technologies and Participation Since 2012	7888
<i>Jeremy Riel, University of Illinois at Chicago, USA</i>	
<i>Kimberly A. Lawless, University of Illinois at Chicago, USA</i>	
Digital Animation for Representing Architectural Design	973
<i>Roberta Spallone, Politecnico di Torino, Italy</i>	

Digital Archives for Preserving and Communicating Architectural Drawings	5213
<i>Roberta Spallone, Politecnico di Torino, Italy</i>	
<i>Francesca Paluan, Politecnico di Torino, Italy</i>	
Digital Divide	4619
<i>Patrick Flanagan, St. John's University, USA</i>	
Digital Literacy	2225
<i>Anirban Ray, UNC Wilmington, USA</i>	
Digital Literacy for the 21st Century	2235
<i>Hiller A. Spires, North Carolina State University, USA</i>	
<i>Casey Medlock Paul, North Carolina State University, USA</i>	
<i>Shea N. Kerkhoff, North Carolina State University, USA</i>	
Digital Literacy in Theory and Practice.....	2243
<i>Heidi Julien, State University of New York at Buffalo, USA</i>	
Digital Storytelling in Language Classes.....	2442
<i>Mehrak Rahimi, Shahid Rajaei Teacher Training University, Iran</i>	
Digital Transformation Journeys in a Digitized Reality	682
<i>Jurgen Janssens, QSpin, Belgium</i>	
Digital Video Watermarking Using Diverse Watermarking Schemes.....	4872
<i>Yash Gupta, Maulana Abul Kalam Azad University of Technology, India</i>	
<i>Shaila Agrawal, Maulana Abul Kalam Azad University of Technology, India</i>	
<i>Susmit Sengupta, Maulana Abul Kalam Azad University of Technology, India</i>	
<i>Aruna Chakraborty, Maulana Abul Kalam Azad University of Technology, India</i>	
A Disability-Aware Mentality to Information Systems Design and Development.....	314
<i>Julius T. Nganji, University of Ottawa, Canada</i>	
Discrete Event Simulation in Inventory Management	5335
<i>Linh Nguyen Khanh Duong, Auckland University of Technology, New Zealand</i>	
<i>Lincoln C. Wood, University of Otago, New Zealand & Curtin University, Australia</i>	
Discussion Processes in Online Forums	7969
<i>Gaowei Chen, The University of Hong Kong, Hong Kong</i>	
<i>Ming M Chiu, The Education University of Hong Kong, Hong Kong</i>	
Displaying Hidden Information in Glossaries	7411
<i>Marcela Ridao, Universidad Nacional del Centro de la Provincia de Buenos Aires, Argentina</i>	
<i>Jorge Horacio Doorn, Universidad Nacional del Oeste, Argentina & Universidad Nacional de La Matanza, Argentina</i>	

Distance Teaching and Learning Platforms	2455
<i>Linda D. Grooms, Regent University, USA</i>	
Distributed Methods for Multi-Sink Wireless Sensor Networks Formation	6522
<i>Miriam A. Carlos-Mancilla, CINVESTAV Unidad Guadalajara, Mexico</i>	
<i>Ernesto Lopez-Mellado, CINVESTAV Unidad Guadalajara, Mexico</i>	
<i>Mario Siller, CINVESTAV Unidad Guadalajara, Mexico</i>	
Distributed Parameter Systems Control and Its Applications to Financial Engineering	15
<i>Gerasimos Rigatos, Industrial Systems Institute, Greece</i>	
<i>Pierluigi Siano, University of Salerno, Italy</i>	
Does Inter-Bank Investments Restraints Financing Performance of Islamic Banks?.....	36
<i>Mohammad Taqiuddin Mohamad, University of Malaya, Malaysia</i>	
<i>Munazza Saeed, University of Malaya, Malaysia</i>	
Dotted Raster-Stereography	166
<i>Muhammad Wasim, Usman Institute of Technology, Pakistan</i>	
<i>Fauzan Saeed, Usman Institute of Technology, Pakistan</i>	
<i>Abdul Aziz, Usman Institute of Technology, Pakistan</i>	
<i>Adnan Ahmed Siddiqui, Hamdard University, Pakistan</i>	
Do Usability Design Features of a Mobile Game Influence Learning?.....	2466
<i>Rex Perez Bringula, University of the East, Philippines</i>	
<i>Edison Cabrera, University of the East, Philippines</i>	
<i>Princess B. Calmerin, University of the East, Philippines</i>	
<i>Eduardo A. Lao, University of the East, Philippines</i>	
<i>Christian Gerard Sembrano, University of the East, Philippines</i>	
<i>Angelita D. Guia, University of the East, Philippines</i>	
<i>Joan P. Lazaro, University of the East, Philippines</i>	
<i>Alexis John M. Rubio, University of the East, Philippines</i>	
<i>Annaliza E. Catacutan, National University, Philippines</i>	
<i>Marilou N. Jamis, National University, Philippines</i>	
<i>Lalaine P. Abad, Department of Education, Philippines</i>	
The Dual Nature of Participatory Web and How Misinformation Seemingly Travels	6993
<i>Sameer Kumar, University of Malaya, Malaysia</i>	
Dynamic Situational Adaptation of a Requirements Engineering Process.....	7422
<i>Graciela Dora Susana Hadad, Universidad Nacional del Oeste, Argentina & Universidad de Belgrano, Argentina</i>	
<i>Jorge Horacio Doorn, Universidad Nacional del Oeste, Argentina & Universidad Nacional de La Matanza, Argentina</i>	
<i>Viviana Alejandra Ledesma, Universidad Nacional del Oeste, Argentina & Universidad Nacional de La Matanza, Argentina</i>	

E-Activism Development and Growth	3569
<i>John G. McNutt, University of Delaware, USA</i>	
<i>Lauri Goldkind, Fordham University, USA</i>	
E-Business and Big Data Strategy in Franchising	2686
<i>Ye-Sho Chen, Louisiana State University, USA</i>	
E-Business Supply Chains Drivers, Metrics, and ERP Integration	5345
<i>Jean C. Essila, Northern Michigan University, USA</i>	
E-Collaborative Learning (e-CL).....	6336
<i>Alexandros Xafopoulos, University College London, UK</i>	
Ecological Performance as a New Metric to Measure Green Supply Chain Practices	5357
<i>June Poh Kim Tam, Universiti Sains Malaysia, Malaysia</i>	
<i>Yudi Fernando, Universiti Malaysia Pahang, Malaysia</i>	
E-Commerce in Logistics and Supply Chain Management	5367
<i>Yasanur Kayikci, Turkish-German University, Turkey</i>	
E-Commerce Models, Players, and Its Future	2739
<i>Liguo Yu, Indiana University – South Bend, USA</i>	
The Economics of Internetization.....	7980
<i>Constantine E. Passaris, University of New Brunswick, Canada</i>	
Educational Ontology Development.....	1441
<i>Galip Kaya, Havelsan Inc., Turkey</i>	
<i>Arif Altun, Hacettepe University, Turkey</i>	
Educational Serious Games Design.....	3287
<i>Ilias Karasavvidis, University of Thessaly, Greece</i>	
Educational Technology and Intellectual Property	2477
<i>Lesley S. J. Farmer, California State University – Long Beach, USA</i>	
Effective Cultural Communication via Information and Communication Technologies and Social Media Use	7002
<i>Androniki Kavoura, Technological Educational Institute of Athens, Greece</i>	
<i>Stella Sylaiou, Aristotle University of Thessaloniki, Greece</i>	
Effectively Communicating With Group Decision Support Systems Using Information Theory	2121
<i>Jamie S. Switzer, Colorado State University, USA</i>	
<i>Ralph V. Switzer, Colorado State University, USA</i>	
Effectiveness of Teacher Training in Using Latest Technologies	7635
<i>Revathi Viswanathan, B. S. Abdur Rahman University, India</i>	

The Effect of Innovative Communication Technologies in Higher Education	3827
<i>Stavros Kiriakidis, University of Crete, Greece</i>	
<i>Efstathios Kefallonitis, State University of New York at Oswego, USA</i>	
<i>Androniki Kavoura, Technological Educational Institute of Athens, Greece</i>	
The Effect of Social Media Networking in the Travel Industry	4052
<i>Androniki Kavoura, Technological Educational Institute of Athens, Greece</i>	
<i>Efstathios Kefallonitis, State University of New York at Oswego, USA</i>	
An Efficient and Effective Index Structure for Query Evaluation in Search Engines.....	7995
<i>Yangjun Chen, University of Winnipeg, Canada</i>	
Efficient Optimization Using Metaheuristics.....	7693
<i>Sergio Nesmachnow, Universidad de la República, Uruguay</i>	
E-Government Service Adoption and the Impact of Privacy and Trust	3579
<i>Mehree Iqbal, North South University, Bangladesh</i>	
<i>Nabila Nisha, North South University, Bangladesh</i>	
<i>Afrin Rifat, North South University, Bangladesh</i>	
Eight Tips for the Theme, “Data and Forecasts”	4629
<i>Alessio Drivet, GeoGebra Institute of Torino, Italy</i>	
Electronic Health Record (EHR) Diffusion and an Examination of Physician Resistance	3678
<i>Kristen MacIver, Northern Michigan University, USA</i>	
<i>Madison N. Ngafeeson, Northern Michigan University, USA</i>	
Electronic Payment Frameworks	2749
<i>Antonio Ruiz-Martínez, University of Murcia, Spain</i>	
<i>Oussama Tounekti, University of Murcia, Spain</i>	
<i>Antonio F. Skarmeta, University of Murcia, Spain</i>	
Electronic Theses and Dissertations (ETDs)	6748
<i>Ralph Hartsock, University of North Texas, USA</i>	
<i>Daniel G. Alemneh, University of North Texas, USA</i>	
An Empirical Study of Mobile/Handheld App Development Using Android Platforms	6057
<i>Wen-Chen Hu, University of North Dakota, USA</i>	
<i>Naima Kaabouch, University of North Dakota, USA</i>	
<i>Hung-Jen Yang, National Kaohsiung Normal University, Taiwan</i>	
Empirical Verification of the Performance Measurement System.....	5638
<i>Aleksander Janeš, University of Primorska, Slovenia</i>	
Employing Educational Robotics for the Development of Problem-Based Learning Skills	2492
<i>Nikleia Eteokleous, Frederick University Cyprus, Cyprus</i>	

Encouraging Digital Literacy and ICT Competency in the Information Age.....	2253
<i>Kijpokin Kasemsap, Suan Sunandha Rajabhat University, Thailand</i>	
Energy Conservation in the Era of Ubiquitous Computing	7745
<i>P. P. Abdul Haleem, Farook College, India</i>	
Enhancing the Mobile User Experience Through Colored Contrasts.....	6070
<i>Jean-Éric Pelet, ESCE International Business School, France</i>	
<i>Basma Taieb, University of Cergy Pontoise, France</i>	
Enhancing the Resiliency of Smart Grid Monitoring and Control	3056
<i>Wenbing Zhao, Cleveland State University, USA</i>	
Ensemble Clustering Data Mining and Databases.....	1962
<i>Slawomir T. Wierzchon, Polish Academy of Sciences, Poland</i>	
Enterprise Interoperability	2773
<i>Ejub Kajan, State University of Novi Pazar, Serbia</i>	
Entrepreneurship	2998
<i>Mehmet Eymen Eryilmaz, Uludağ University, Turkey</i>	
Entrepreneurship as the Vantage Point	3009
<i>Madhu Kishore Raghunath Raghunath Kamakula, GVP College of Engineering</i>	
<i>(Autonomous), India</i>	
<i>Chandra Sekhar Patro, GVP College of Engineering (Autonomous), India</i>	
Entrepreneurship Concept, Theories, and New Approaches	3020
<i>José Manuel Saiz-Alvarez, Tecnológico de Monterrey, Mexico</i>	
<i>Martín García-Vaquero, Nebrija University, Spain</i>	
Ergonomic Design of a Driver Training Simulator for Rural India.....	1260
<i>Prabir Mukhopadhyay, Indian Institute of Information Technology Design and</i>	
<i>Manufacturing Jabalpur, India</i>	
<i>Vipul Vinzuda, National Institute of Design Gandhinagar, India</i>	
ERP Systems Benefit Realization and the Role of ERP-Enabled Application Integration.....	2953
<i>Joseph K. Nwankpa, Miami University, USA</i>	
An Essay on Denotational Mathematics	7704
<i>Giuseppe Iurato, University of Palermo, Italy</i>	
Ethical Ambiguities in the Privacy Policies of Mobile Health and Fitness Applications	6083
<i>Devjani Sen, University of Ottawa, Canada</i>	
<i>Rukhsana Ahmed, University of Ottawa, Canada</i>	

Ethical Computing Continues From Problem to Solution	4884
<i>Wanbil William Lee, The Computer Ethics Society, Hong Kong & Wanbil & Associates, Hong Kong</i>	
Evaluative Dimensions of Urban Tourism in Capital Cities by First-Time Visitors	4064
<i>Annamaria Silvana de Rosa, Sapienza University of Rome, Italy Laura Dryjanska, Sapienza University of Rome, Italy Elena Bocci, Sapienza University of Rome, Italy</i>	
Evolutionary Algorithms for Global Decision Tree Induction	2132
<i>Marek Kretowski, Bialystok University of Technology, Poland Marcin Czajkowski, Bialystok University of Technology, Poland</i>	
E-Waste, Chemical Toxicity, and Legislation in India.....	3066
<i>Prashant Mehta, National Law University Jodhpur, India</i>	
Existential Aspects of the Development E-Culture	4189
<i>Liudmila Vladimirovna Baeva, Astrakhan State University, Russia</i>	
Experiences of Implementing a Large-Scale Blended, Flipped Learning Project.....	3839
<i>Hazel Owen, Ethos Consultancy NZ, New Zealand Nicola Dunham, Massey University, New Zealand</i>	
Explaining and Predicting Users' Continuance Usage Intention Toward E-Filing Utilizing Technology Continuance Theory	831
<i>Santhanamery T., Universiti Teknologi MARA, Malaysia T. Ramayah, University Sains Malaysia, Malaysia</i>	
Exploratory Data Analysis on Breast Cancer Prognosis.....	1794
<i>Mohammad Mehdi Owrang O., American University, USA Yasmine M. Kanaan, Howard University, USA Robert L. Copeland Jr., Howard University, USA Melvin Gaskins, Howard University Hospital, USA Robert L. DeWitty Jr., Providence Hospital, USA</i>	
Exploring Drivers of Closed Loop Supply Chain in Malaysian Automotive Industry	5378
<i>Fadzlina Mohd Fadzil, Universiti Sains Malaysia, Malaysia Yudi Fernando, Universiti Malaysia Pahang, Malaysia</i>	
Exploring "Hacking," Digital Public Art, and Implication for Contemporary Governance.....	6695
<i>Amadu Wurie Khan, University of Edinburgh, UK Chris Speed, University of Edinburgh, UK</i>	
Exploring New Handwriting Parameters for Writer Identification.....	4643
<i>Verónica Inés Aubin, Universidad Nacional de La Matanza, Argentina Jorge Horacio Doorn, Universidad Nacional del Oeste, Argentina & Universidad Nacional de La Matanza, Argentina</i>	

Exploring the Growth of Wireless Communications Systems and Challenges Facing 4G Networks	6094
<i>Amber A. Smith-Ditizio, Texas Woman's University, USA</i>	
<i>Alan D. Smith, Robert Morris University, USA</i>	
Exploring Tourism Cluster in the Peripheral Mountain Area Based on GIS Mapping	3434
<i>Ya-Hui Hsueh, National Taichung University of Education, Taiwan</i>	
<i>Huey-Wen Chuang, National Taichung University of Education, Taiwan</i>	
<i>Wan-Chiang Hsieh, National Taichung Girls' Senior High School, Taiwan</i>	
Exposure to Video Games and Decision Making	3296
<i>Giuseppe Curcio, University of L'Aquila, Italy</i>	
<i>Sara Peracchia, University of L'Aquila, Italy</i>	
An Extension to the Delone and Mclean Information Systems Success Model and Validation in the Internet Banking Context	49
<i>Veeraraghavan Jagannathan, National Institute of Technology, India</i>	
<i>Senthilarasu Balasubramanian, National Institute of Technology, India</i>	
<i>Thamaraiselvan Natarajan, National Institute of Technology, India</i>	
Facilitating Customer Relationship Management in Modern Business	1594
<i>Kijpokin Kasemsap, Suan Sunandha Rajabhat University, Thailand</i>	
Facilitating Interaction Between Virtual Agents Through Negotiation Over Ontological Representation.....	2697
<i>Fiona McNeill, Heriot-Watt University, UK</i>	
Factors Contributing to the Effectiveness of Online Students and Instructors	1451
<i>Michelle Kilburn, Southeast Missouri State University, USA</i>	
<i>Martha Henckell, Southeast Missouri State University, USA</i>	
<i>David Starrett, Columbia College, USA</i>	
Factors Determining E-Shopping Compliance by Nigerians.....	2761
<i>Adamkolo Mohammed Mohammed Ibrahim, University of Maiduguri, Nigeria</i>	
<i>Md. Salleh Hj. Hassan, Universiti Putra Malaysia, Malaysia</i>	
<i>Sarina Yusuf, Universiti Putra Malaysia, Malaysia</i>	
A Family Review of Parameter-Learning Models and Algorithms for Making Actionable Decisions.....	2142
<i>Chun-Kit Ngan, The Pennsylvania State University, USA</i>	
A Fast and Space-Economical Algorithm for the Tree Inclusion Problem	4502
<i>Yangjun Chen, University of Winnipeg, Canada</i>	
<i>Yibin Chen, University of Winnipeg, Canada</i>	

Fault Tolerant Cloud Systems	1075
<i>Sathish Kumar, VIT University, India</i>	
<i>Balamurugan B, VIT University, India</i>	
Fault Tolerant Data Management for Cloud Services.....	1091
<i>Wenbing Zhao, Cleveland State University, USA</i>	
Fifty Shades of Dark Stories	4077
<i>Lea Kuznik, University of Maribor, Slovenia</i>	
Financial Inclusion, Content, and Information Technologies in Latin America.....	7214
<i>Alberto Chong, Georgia State University, USA & Universidad del Pacifico, Peru</i>	
<i>Cecilia de Mendoza, Ministry of Production of Argentina, Argentina</i>	
Financing Micro, Small, and Medium Enterprises in Indian Industry	6916
<i>Shromona Ganguly, Indian Institute of Management Calcutta, India & Reserve Bank of India, India</i>	
A Flipped Learning Approach to University EFL Courses	3850
<i>Yasushige Ishikawa, Kyoto University of Foreign Studies, Japan</i>	
<i>Reiko Akahane-Yamada, Advanced Telecommunications Research Institute International (ATR), Japan</i>	
<i>Craig Smith, Kyoto University of Foreign Studies, Japan</i>	
<i>Masayuki Murakami, Kyoto University of Foreign Studies, Japan</i>	
<i>Mutsumi Kondo, Kyoto University of Foreign Studies, Japan</i>	
<i>Misato Kitamura, Advanced Telecommunications Research Institute International (ATR), Japan</i>	
<i>Yasushi Tsubota, Kyoto Institute of Technology, Japan</i>	
<i>Masatake Dantsuji, Kyoto University, Japan</i>	
Flipping the Medical School Classroom.....	5800
<i>Kristina Kaljo, Medical College of Wisconsin, USA</i>	
<i>Laura Jacques, Medical College of Wisconsin, USA</i>	
Flying Adhoc Networks Concept and Challenges	6106
<i>Kuldeep Singh, Thapar University, India</i>	
<i>Anil Kumar Verma, Thapar University, India</i>	
Forecasting the Demand of Agricultural Crops/Commodity Using Business Intelligence Framework	847
<i>Satyadhyhan Chickerur, KLE Technological University, India</i>	
<i>Supreeth Sharma, Akamai Technologies, India</i>	
<i>Prashant M. Narayankar, KLE Technological University, India</i>	
Forensic Investigations in Cloud Computing.....	1356
<i>Diane Barrett, Bloomsburg University of Pennsylvania, USA</i>	

A Formal Approach to the Distributed Software Control for Automated Multi-Axis Manufacturing Machines	7435
<i>Gen'ichi Yasuda, Nagasaki Institute of Applied Science, Japan</i>	
The Foundation of (Business) Ethics' Evolution	3173
<i>Ben Tran, Alliant International University, USA</i>	
A Framework for Exploring IT-Led Change in Morphing Organizations	694
<i>Sharon A. Cox, Birmingham City University, UK</i>	
A Framework for Profiling Prospective Students in Higher Education	3861
<i>Santhosh Kumar Lakkaraju, Dakota State University, USA</i>	
<i>Deb Tech, Dakota State University, USA</i>	
<i>Shuyuan Deng, Dakota State University, USA</i>	
From Business-to-Business Software Startup to SAP's Acquisition	5388
<i>John Wang, Montclair State University, USA</i>	
<i>Jeffrey Hsu, Fairleigh Dickinson University, USA</i>	
<i>Sylvain Jaume, Saint Peter's University, USA</i>	
From Digital Exclusion to Digital Inclusion for Adult Online Learners	2503
<i>Virginia E. Garland, University of New Hampshire, USA</i>	
From Digital Natives to Student Experiences With Technology	2512
<i>Sue Bennett, University of Wollongong, Australia</i>	
<i>Linda Corrin, University of Melbourne, Australia</i>	
From General Services to Pervasive and Sensitive Services	7754
<i>Mario Vega-Barbas, KTH-Royal Institute of Technology, Sweden & ESNE-Universidad</i>	
<i>Camilo José Cela, Spain</i>	
<i>Iván Pau, Universidad Politécnica de Madrid, Spain</i>	
<i>Fernando Seoane, KI-Karolinska Institutet, Sweden & University of Borås, Sweden</i>	
From Information Systems Outsourcing to Cloud Computing	1101
<i>Mohammad Nabil Almunawar, Universiti Brunei Darussalam, Brunei</i>	
<i>Hasan Jawwad Almunawar, P. T. Tegar Kupas Mediatama, Indonesia</i>	
From On-Premise ERP to Cloud ERP	2965
<i>Karim Mezghani, Al Imam Mohammad Ibn Saud Islamic University (IMSIU), Saudi Arabia</i>	
<i>& University of Sfax, Tunisia</i>	
From the Psychoanalyst's Couch to Social Networks	7014
<i>Annamaria Silvana de Rosa, Sapienza University of Rome, Italy</i>	
<i>Emanuele Fino, Psychologist, Psychometrician, Italy</i>	
<i>Elena Bocci, Sapienza University of Rome, Italy</i>	

The Fundamentals of Human-Computer Interaction	4199
<i>Kijpokin Kasemsap, Suan Sunandha Rajabhat University, Thailand</i>	
The Future of High-Performance Computing (HPC)	4004
<i>Herbert Cornelius, Intel Corporation EMEA, Germany</i>	
Fuzzy Logic Approach in Risk Assessment	6789
<i>Çetin Karahan, Directorate General of Civil Aviation, Turkey</i>	
<i>Esra Ayça Güzeldereli, Afyon Kocatepe University, Turkey</i>	
<i>Aslıhan Tüfekci, Gazi University, Turkey</i>	
A Gamification Update to the Taxonomy of Technology and Mental Health	995
<i>Madeline R. Marks, University of Central Florida, USA</i>	
<i>Amanda C. Tan, University of Central Florida, USA</i>	
<i>Clint Bowers, University of Central Florida, USA</i>	
Gender Differences in Advertising Engagement Using the Case of Facebooks Ads	3359
<i>Eva Lahuerta-Otero, University of Salamanca, Spain</i>	
<i>Rebeca Cordero-Gutiérrez, University of Salamanca, Spain</i>	
The Gender Dimension in Urban Air Quality.....	3371
<i>Theodora Slini, Aristotle University of Thessaloniki, Greece</i>	
<i>Fotini-Niovi Pavlidou, Aristotle University of Thessaloniki, Greece</i>	
General Perspectives on Electromyography Signal Features and Classifiers Used for Control of Human Arm Prosthetics	492
<i>Faruk Ortes, Istanbul University, Turkey</i>	
<i>Derya Karabulut, Halic University, Turkey</i>	
<i>Yunus Ziya Arslan, Istanbul University, Turkey</i>	
Geographic Information System (GIS) Modeling Analysis and the Effects of Spatial Distribution and Environmental Factors on Breast Cancer Incidence	3448
<i>Akram Gasmelseed, University of Science and Technology, Sudan</i>	
<i>Ali H. Alharbi, Qassim University, Saudi Arabia</i>	
Geographic Information Systems.....	3460
<i>Paula Remoaldo, University of Minho, Portugal</i>	
<i>Vitor P. Ribeiro, University of Minho, Portugal</i>	
<i>Hélder Silva Lopes, University of Minho, Portugal</i>	
<i>Sara Catarina Gomes Silva, University of Minho, Portugal</i>	
Geospatial Influence in Science Mapping.....	3473
<i>Carlos Granell-Canut, Universitat Jaume I of Castellón, Spain</i>	
<i>Estefanía Aguilar-Moreno, Universitat Jaume I of Castellón, Spain</i>	

Graph-Based Concept Discovery	1974
<i>Alev Mutlu, Kocaeli University, Turkey</i>	
<i>Pinar Karagoz, Middle East Technical University, Turkey</i>	
<i>Yusuf Kavurucu, Turkish Naval Research Center Command, Turkey</i>	
A Graph-Intersection-Based Algorithm to Determine Maximum Lifetime Communication Topologies for Cognitive Radio Ad Hoc Networks	6536
<i>Natarajan Meghanathan, Jackson State University, USA</i>	
Green IT and the Struggle for a Widespread Adoption	3077
<i>Edward T. Chen, University of Massachusetts – Lowell, USA</i>	
Group Signature System Using Multivariate Asymmetric Cryptography	4898
<i>Sattar J. Aboud, University of Bedfordshire, UK</i>	
Group Synchronization for Multimedia Systems	6435
<i>Dimitris N. Kanellopoulos, University of Patras, Greece</i>	
The Growing Impact of ICT on Development in Africa	7223
<i>Sherif H. Kamel, The American University in Cairo, Egypt</i>	
GWAS as the Detective to Find Genetic Contribution in Diseases	466
<i>Simanti Bhattacharya, University of Kalyani, India</i>	
<i>Amit Das, University of Kalyani, India</i>	
Haptics-Based Systems Characteristics, Classification, and Applications	4652
<i>Abeer Bayousuf, King Saud University, Saudi Arabia</i>	
<i>Hend S. Al-Khalifa, King Saud University, Saudi Arabia</i>	
<i>Abdulmalik Al-Salman, King Saud University, Saudi Arabia</i>	
Has Bitcoin Achieved the Characteristics of Money?	2784
<i>Donovan Peter Chan Wai Loon, University of Malaya, Malaysia</i>	
<i>Sameer Kumar, University of Malaya, Malaysia</i>	
Healthcare Data Analysis in the Internet of Things Era	1984
<i>George Tzanis, Aristotle University of Thessaloniki, Greece</i>	
Health Wearables Turn to Fashion	6114
<i>Lambert Spaanenburg, Comoray AB, Sweden</i>	
Heart Sound Analysis for Blood Pressure Estimation	1006
<i>Rui Guedes, Faculdade de Medicina da Universidade do Porto, Portugal</i>	
<i>Henrique Cyrne Carvalho, Serviço de Cardiologia, Hospital de Santo António, Centro</i>	
<i>Hospitalar do Porto, Portugal</i>	
<i>Ana Castro, Universidade do Porto, Portugal</i>	

Hexa-Dimension Code of Practice for Data Privacy Protection	4909
<i>Wanbil William Lee, Wanbil & Associates, Hong Kong</i>	
High-Performance Reconfigurable Computing	4018
<i>Mário Pereira Vestias, Instituto Politécnico de Lisboa, Portugal</i>	
The Holon/Parton Structure of the Meme, or The Unit of Culture.....	4666
<i>J. T. Velikovsky, University of Newcastle, Australia</i>	
Home UbiHealth	7765
<i>John Sarivougioukas, “G. Gennimatas” Athens General Hospital, Greece</i>	
<i>Aristides Vagelatos, CTI&P, Greece</i>	
<i>Konstantinos Parsopoulos, University of Ioannina, Greece</i>	
<i>Isaac E. Lagaris, University of Ioannina, Greece</i>	
How Exclusive Work Climates Create Barriers for Women in IS&T.....	3382
<i>Katelyn R. Reynoldson, Old Dominion University, USA</i>	
<i>Debra A. Major, Old Dominion University, USA</i>	
How the Crowdsourcing Enhance the Co-Creation Into the Virtual Communities.....	707
<i>Bahri Ammari Nedra, IHEC of Carthage, USA</i>	
How Visualisation and Interaction Can Optimize the Cognitive Processes Towards Big Data	381
<i>Antonio Feraco, Fraunhofer IDM@NTU, Singapore</i>	
<i>Marius Erdt, Fraunhofer IDM@NTU, Singapore</i>	
Human Psychomotor Performance Under the Exposure to Mobile Phones-Like Electromagnetic Fields.....	6124
<i>Giuseppe Curcio, University of L'Aquila, Italy</i>	
Hybrid Computational Intelligence and the Basic Concepts and Recent Advances.....	180
<i>Georgios Dounias, University of the Aegean, Greece</i>	
Hyper-Sensitivity in Global Virtual Teams	720
<i>Andre Araujo, Texas A&M University, USA</i>	
ICT Eases Inclusion in Education.....	2521
<i>Dražena Gašpar, University of Mostar, Bosnia and Herzegovina</i>	
ICT Investments and Recovery of Troubled Economies	2337
<i>Ioannis Papadopoulos, Metropolitan College Thessaloniki, Greece</i>	
<i>Apostolos Syropoulos, Greek Molecular Computing Group, Greece</i>	
ICT Standardization	4679
<i>Kai Jakobs, RWTH Aachen University, Germany</i>	

Identification of Green Procurement Drivers and Their Interrelationship Using Fuzzy TISM and MICMAC Analysis	3086
<i>Surajit Bag, Tega Industries South Africa Pty Ltd., South Africa</i>	
Identification of Wireless Devices from their Physical Layer Radio-Frequency Fingerprints	6136
<i>Gianmarco Baldini, European Commission – Joint Research Centre, Italy</i>	
<i>Gary Steri, European Commission – Joint Research Centre, Italy</i>	
<i>Raimondo Giuliani, European Commission – Joint Research Centre, Italy</i>	
Immersing People in Scientific Knowledge and Technological Innovation Through Disney’s Use of Installation Art.....	4692
<i>Jonathan Lillie, Loyola University Maryland, USA</i>	
<i>Michelle Jones-Lillie, Lillie Pad Studios, USA</i>	
The Impact of Artificial Intelligence and Virtual Personal Assistants on Marketing.....	5748
<i>Christina L. McDowell Marinchak, University of Alaska Anchorage, USA</i>	
<i>Edward Forrest, University of Alaska Anchorage, USA</i>	
<i>Bogdan Hoanca, University of Alaska Anchorage, USA</i>	
Impact of Business Groups on Payout Policy in India.....	61
<i>Ahana Bose, Indian Institute of Management Calcutta, India</i>	
The Impact of Carbon Nanotubes and Graphene on Electronics Industry	2897
<i>Rafael Vargas-Bernal, Instituto Tecnológico Superior de Irapuato, Mexico</i>	
<i>Gabriel Herrera-Pérez, Instituto Tecnológico Superior de Irapuato, Mexico</i>	
<i>Margarita Tecpoyotl-Torres, Universidad Autónoma del Estado de Morelos, Mexico</i>	
The Impact of Mobile Phones on Plastic Surgery and Burn Management.....	6147
<i>Maria Giaquinto-Cilliers, Kimberley Hospital Complex, South Africa</i>	
<i>Tertius N. Potgieter, Kimberley Hospital Complex, South Africa</i>	
<i>Gert Steyn, Kimberley Hospital Complex, South Africa</i>	
The Impact of the Impact of Meta-Data Mining From the SoReCom “A.S. de Rosa” @-Library	4404
<i>Annamaria Silvana de Rosa, Sapienza University of Rome, Italy</i>	
<i>Laura Dryjanska, Sapienza University of Rome, Italy</i>	
<i>Elena Bocci, Sapienza University of Rome, Italy</i>	
Implementing a Customer Relationship Management (CRM) System.....	1605
<i>Dimitra Skoumpopoulou, Northumbria University, UK</i>	
<i>Benjamin Franklin, Northumbria University, UK</i>	
Implicit Cognitive Vulnerability	5149
<i>Caroline M. Crawford, University of Houston – Clear Lake, US</i>	
The Importance of Electronic Commerce in Modern Business.....	2791
<i>Kijpokin Kasemsap, Suan Sunandha Rajabhat University, Thailand</i>	

Importance of Information Literacy	3870
<i>Lidia Sanchez-Ruiz, University of Cantabria, Spain</i>	
<i>Beatriz Blanco, University of Cantabria, Spain</i>	
Improved Checkpoint Using the Effective Management of I/O in a Cloud Environment	1116
<i>Bakhta Meroufel, University of Oran1, Algeria</i>	
<i>Ghalem Belalem, University of Es Senia Oran1, Algeria</i>	
Improving Competencies for the Courier Service Industry in Malaysia	2802
<i>Hoo Yee Hui, Universiti Tunku Abdul Rahman, Malaysia</i>	
<i>Yudi Fernando, Universiti Malaysia Pahang, Malaysia</i>	
Improving Competitiveness Through Organizational Market Intelligence.....	961
<i>George Leal Jamil, Informações em Rede, Brazil</i>	
<i>Leandro Rocha Dos Santos, In3 Inteligência de Mercado, Brazil</i>	
<i>Cecília C. Jamil, Stockholm University, Sweden</i>	
Improving Dependability of Robotics Systems	6847
<i>Nidhal Mahmud, University of Hull, UK</i>	
Improving Quality of Business in Next Generation Telecom Networks	6546
<i>Vesna Radonjić Đogatović, University of Belgrade, Serbia</i>	
Improving Usability of Website Design Using W3C Guidelines	8006
<i>G. Sreedhar, Rashtriya Sanskrit Vidyapeetha (Deemed University), India</i>	
Increasing Student Engagement and Participation Through Course Methodology	1463
<i>T. Ray Ruffin, University of Phoenix, USA & Grand Canyon University, USA & Ashford</i>	
<i>University, USA & North Carolina Wesleyan College, USA</i>	
<i>Donna Patterson Hawkins, University of Phoenix, USA</i>	
<i>D. Israel Lee, Southern Illinois University, USA & University of Phoenix, USA</i>	
Incremental Approach to Classification Learning	191
<i>Xenia Alexandre Naidenova, Research Centre of Military Medical Academy – Saint</i>	
<i>Petersburg, Russia</i>	
Indicators of Information and Communication Technology.....	4704
<i>Gulnara Abdrakhmanova, National Research University Higher School of Economics, Russia</i>	
<i>Leonid Gokhberg, National Research University Higher School of Economics, Russia</i>	
<i>Alexander Sokolov, National Research University Higher School of Economics, Russia</i>	
Indigenous Knowledge Systems	5036
<i>Osarumwense Iguisi, University of Benin, Nigeria</i>	
<i>Osaro Rawlings Igbinomwanhia, University of Benin, Nigeria</i>	

Influencing People and Technology Using Human Resource Development (HRD) Philosophy	4326
<i>Claretha Hughes, University of Arkansas, USA</i>	
<i>Matthew W. Gosney, University of Colorado – Health, USA</i>	
<i>Cynthia M. Sims, Clemson University, USA</i>	
Information and Communication Technology Ethics and Social Responsibility	4920
<i>Tomas Cahlik, Charles University Prague, Czech Republic & University of Economics</i>	
<i>Prague, Czech Republic</i>	
Information and Its Conceptual Perspectives.....	4422
<i>José Poças Rascão, Institute Polytechnic of Setúbal, Portugal</i>	
Information-Based Revolution in Military Affairs	7302
<i>Rafal Kopec, Pedagogical University of Krakow, Poland</i>	
Information-Centric Networking	6556
<i>Mohamed Fazil Mohamed Firdhous, University of Moratuwa, Sri Lanka</i>	
Information Needs of Users in the Tech Savvy Environment and the Influencing Factors	2264
<i>Mudasir Khazer Rather, University of Kashmir, India</i>	
<i>Shabir Ahmad Ganaie, University of Kashmir, India</i>	
Information Science and Technology in Crisis Response and Management	1407
<i>Randy Basham, University of Texas at Arlington, USA</i>	
Information Seeking Models in the Digital Age.....	4515
<i>Mudasir Khazer Rather, University of Kashmir, India</i>	
<i>Shabir Ahmad Ganaie, University of Kashmir, India</i>	
Information Systems and Technology Projects in Healthcare Organisations	3756
<i>Jorge Gomes, ISEG, Universidade de Lisboa, Portugal</i>	
<i>Mário José Batista Romão, ISEG, Universidade de Lisboa, Portugal</i>	
Information Technologies and Social Change	4715
<i>Muhammet Ali Köroğlu, Uşak University, Turkey</i>	
<i>Cemile Zehra Köroğlu, Uşak University, Turkey</i>	
Informed Decision Making With Enterprise Dynamic Systems Control	2154
<i>Sérgio Luís Guerreiro, Instituto Superior Técnico, University of Lisbon, Portugal & INESC-</i>	
<i>ID, Portugal</i>	
The Infusion of Technology Within the Classroom Facilitates Students' Autonomy in Their Learning	2532
<i>Fariel Mohan, University of Trinidad and Tobago, Trinidad and Tobago</i>	
<i>Garry Soomarah, University of Trinidad and Tobago, Trinidad and Tobago</i>	

In-Memory Analytics.....	1806
<i>Jorge Manjarrez Sánchez, Instituto Tecnológico Superior de Jerez, Mexico</i>	
Innovative Formalism for Biological Data Analysis.....	1814
<i>Calin Ciufudean, Stefan cel Mare University, Romania</i>	
An Insight Into Deep Learning Architectures.....	4528
<i>Nishu Garg, VIT University, India</i>	
<i>Nikhitha P, VIT University, India</i>	
<i>B. K. Tripathy, VIT University, India</i>	
Instructional Real World Community Engagement	1474
<i>Caroline M. Crawford, University of Houston – Clear Lake, USA</i>	
An Integrated Approach to Supply Chain Simulation	5398
<i>Nenad Stefanovic, University of Kragujevac, Serbia</i>	
<i>Bozidar Radenkovic, University of Belgrade, Serbia</i>	
Integrated Data Architecture for Business	862
<i>Richard Kumaradjaja, Renaissance Consulting, Indonesia</i>	
An Integrated Electronic IQA System for HEI.....	3881
<i>Teay Shawyun, King Saud University, Saudi Arabia</i>	
Integrated Paper-Based and Digital Learning Material for Smart Learners	2545
<i>Sabrina Leone, Università Politecnica delle Marche, Italy</i>	
Integrating Content Authentication Support in Media Services.....	2908
<i>Anastasia N. Katsaounidou, Aristotle University of Thessaloniki, Greece</i>	
<i>Charalampos A. Dimoulas, Aristotle University of Thessaloniki, Greece</i>	
Integrating Evidence-Based Practice in Athletic Training Though Online Learning.....	5810
<i>Brittany A. Vorndran, Seton Hall University, USA</i>	
<i>Michelle Lee D’Abundo, Seton Hall University, USA</i>	
Integrating Knowledge Management and Business Processes	5046
<i>John Steven Edwards, Aston University, UK</i>	
Integrating Sustainability and CSR in the Value Chain of the Information Technology Sector	3183
<i>Patricia Martínez García de Leaniz, University of Cantabria, Spain</i>	
<i>María Elena García Ruiz, University of Cantabria, Spain</i>	
Integrating Web-Based Technologies Into the Education and Training of Health Professionals	5820
<i>Michelle Lee D’Abundo, Seton Hall University USA</i>	
<i>Cara Sidman, Arizona State University, USA</i>	

Intellectual Capital Measurement	5056
<i>Lukasz Bryl, Poznan University of Economics and Business, Poland</i>	
The Intelligence of E-CRM Applications and Approaches on Online Shopping Industry.....	1616
<i>Bashar Shahir Ahmed, University Abdelmalek Essaadi (LEROSA Laborator), Morocco</i>	
<i>Fadi Amroush, Universidad de Salamanca, Spain</i>	
<i>Mohammed Ben Maati, University Abdelmalek Essaadi, Morocco</i>	
Interactivity in Distance Education and Computer-Aided Learning, With Medical Education Examples	5829
<i>D. John Doyle, Cleveland Clinic, USA</i>	
<i>Patrick J. Fahy, Athabasca University, Canada</i>	
Interface Trends in Human Interaction, the Internet of Things, and Big Data	4210
<i>William J. Gibbs, Duquesne University, USA</i>	
International Students in Online Courses	3900
<i>María Ángeles Rodríguez Manzanares, Memorial University of Newfoundland, Canada</i>	
Internet Addiction in Context	4223
<i>Petra Vondrackova, Charles University in Prague, Czech Republic</i>	
<i>David Šmahel, Masaryk University – Brno, Czech Republic</i>	
The Internet Behavior of Older Adults	7026
<i>Elizabeth Mazur, Pennsylvania State University, USA</i>	
<i>Margaret L. Signorella, Pennsylvania State University, USA</i>	
<i>Michelle Hough, Pennsylvania State University, USA</i>	
Internet-Facilitated Child Sexual Exploitation Crimes.....	1366
<i>Keith F. Durkin, Ohio Northern University, USA</i>	
<i>Ronald L. DeLong, University of Dayton, USA</i>	
Internet of Things Applications for Healthcare	3689
<i>Ljubica Diković, Business Technical College, Serbia</i>	
Internet Phenomenon	8015
<i>Lars Konzack, University of Copenhagen, Denmark</i>	
Interoperability Frameworks for Distributed Systems	6566
<i>José Carlos Martins Delgado, Universidade de Lisboa, Portugal</i>	
The Intersection of Religion and Mobile Technology	6161
<i>Wendi R. Bellar, Texas A&M University, USA</i>	
<i>Kyong James Cho, Texas A&M University, USA</i>	
<i>Heidi A. Campbell, Texas A&M University, USA</i>	

Intrusion Tolerance Techniques	4927
<i>Wenbing Zhao, Cleveland State University, USA</i>	
Investigating Diachronic Variation and Change in New Varieties of English	1206
<i>Rita Calabrese, University of Salerno, Italy</i>	
iSchools Promoting “Information Science and Technology” (IST) Domain Towards Community, Business, and Society With Contemporary Worldwide Trend and Emerging Potentialities in India	4723
<i>P. K. Paul, Raiganj University, India</i>	
<i>D. Chatterjee, Seacom Skills University, India</i>	
Issues and Challenges in Enterprise Social Media	7036
<i>Sarabjot Kaur, IIT Kanpur, India</i>	
<i>Subhas Chandra Misra, IIT Kanpur, India</i>	
IT Service Management Architectures	2920
<i>Torben Tambo, Aarhus University, Denmark</i>	
<i>Jacob Filtenborg, Aarhus University, Denmark</i>	
IT Solutions Supporting the Management of Higher Education Institutions in Poland	3910
<i>Elżbieta Janczyk-Strzała, Wrocław School of Banking, Poland</i>	
IT Strategy Follows Digitalization.....	873
<i>Thomas Ochs, Villeroy & Boch, Germany</i>	
<i>Ute Anna Riemann, SAP SE, Germany</i>	
Kinect Applications in Healthcare	5876
<i>Roanna Lun, Cleveland State University, USA</i>	
<i>Wenbing Zhao, Cleveland State University, USA</i>	
Knowledge Acquisition on Dante Alighieri’s Works.....	5067
<i>Elvira Immacolata Locuratolo, ISTI-CNR, Italy</i>	
<i>Valentina Bartalesi Lenzi, ISTI-CNR, Italy</i>	
Knowledge-Based Forensic Patterns and Engineering System	1376
<i>Vivek Tiwari, International Institute of Information Technology Naya Raipur, India</i>	
<i>R. S. Thakur, Maulana Azad National Institute of Technology, India</i>	
Knowledge Management for Development (KM4D).....	5077
<i>Alexander G. Flor, University of the Philippines, Philippines</i>	
Knowledge Management From the Metaphorical Perspective	5085
<i>Magdalena Bielenia-Grajewska, University of Gdansk, Poland</i>	

Knowledge Networks in Higher Education.....	3922
<i>Filipa M. Ribeiro, University of Porto, Portugal</i>	
Lack of Characteristics Management Causing Biggest Projects Failure	5650
<i>Loredana Arana, University of Phoenix, USA</i>	
Latest Advances on Benders Decomposition.....	5411
<i>Antonios Frangkogios, University of Thessaly, Greece</i>	
<i>Georgios K. D. Saharidis, University of Thessaly, Greece</i>	
Lean and Six Sigma Innovation and Design.....	729
<i>Rick Edgeman, Utah State University, USA</i>	
Lean Logistics of the Transportation of Fresh Fruit Bunches (FFB) in the Palm Oil Industry	5422
<i>Cheah Cheng Teik, Universiti Sains Malaysia, Malaysia</i>	
<i>Yudi Fernando, Universiti Malaysia Pahang, Malaysia</i>	
Learner Engagement in Blended Learning	1487
<i>Kristian J. Spring, Brigham Young University, USA</i>	
<i>Charles R. Graham, Brigham Young University, USA</i>	
<i>Tarah B. Ikahihifo, Brigham Young University, USA</i>	
Learning Analytics.....	5158
<i>Constanța-Nicoleta Bodea, Bucharest University of Economic Studies, Romania</i>	
<i>Maria-Iuliana Dascalu, University Politehnica of Bucharest, Romania</i>	
<i>Radu Ioan Mogos, Bucharest University of Economic Studies, Romania</i>	
<i>Stelian Stancu, Bucharest University of Economic Studies, Romania</i>	
Learning From Imbalanced Data	1825
<i>Lincy Mathews, M. S. Ramaiah Institute of Technology, India</i>	
<i>Seetha Hari, Vellore Institute of Technology, India</i>	
Learning With Games and Digital Stories in Visual Programming	3309
<i>Wilfred W. F. Lau, The Chinese University of Hong Kong, China</i>	
Learning With Mobile Devices.....	6347
<i>Helen Crompton, Old Dominion University, USA</i>	
<i>John Traxler, University of Wolverhampton, UK</i>	
Leveraging Technology-Enhanced Teaching and Learning for Future IS Security Professionals....	2558
<i>Ciara Heavin, University College Cork, Ireland</i>	
<i>Karen Neville, University College Cork, Ireland</i>	
<i>Sheila O’Riordan, University College Cork, Ireland</i>	

Leveraging the Arduino Platform to Develop Information Technology Devices	3273
<i>Diego Reforgiato Recupero, University of Cagliari, Italy</i>	
<i>Valentino Artizzu, University of Cagliari, Italy</i>	
<i>Francesca Cella, University of Cagliari, Italy</i>	
<i>Alessandro Cotza, University of Cagliari, Italy</i>	
<i>Davide Curcio, University of Cagliari, Italy</i>	
<i>Giorgio Amedeo Iengo, University of Cagliari, Italy</i>	
<i>Riccardo Macis, University of Cagliari, Italy</i>	
<i>Andrea Marras, University of Cagliari, Italy</i>	
<i>Simone Picci, University of Cagliari, Italy</i>	
<i>Michael Planu, University of Cagliari, Italy</i>	
<i>Riccardo Scasseddu, University of Cagliari, Italy</i>	
Liberating Educational Technology Through the Socratic Method.....	2571
<i>Frank G. Giuseffi, Lindenwood University, USA</i>	
Literature Review of Augmented Reality Application in the Architecture, Engineering, and Construction Industry With Relation to Building Information.....	983
<i>Aydin Tabrizi, University of Kansas, USA</i>	
<i>Paola Sanguinetti, University of Kansas, USA</i>	
Load Flow Analysis in Smart Grids.....	3103
<i>Osman Hasan, National University of Sciences and Technology, Pakistan</i>	
<i>Awais Mahmood, National University of Sciences and Technology, Pakistan</i>	
<i>Syed Rafay Hasan, Tennessee Technological University, USA</i>	
Logic Programming for Intelligent Systems	4736
<i>James D. Jones, Liberty University, USA</i>	
Machine Dreaming.....	202
<i>James Frederic Pagel, University of Colorado School of Medicine, USA</i>	
The Main Concepts Behind the Dematerialization of Business Processes.....	888
<i>Liliana Ávila, University of Aveiro, Portugal</i>	
<i>Leonor Teixeira, University of Aveiro, Portugal</i>	
Maintenance Policies Optimization of Medical Equipment in a Health Care Organization	3698
<i>Juan Ignacio Roig, University of Castilla-La Mancha, Spain</i>	
<i>Andrés Gómez, University of Castilla-La Mancha, Spain</i>	
<i>Isabel Romero, University of Castilla-La Mancha, Spain</i>	
<i>María Carmen Carnero, University of Castilla-La Mancha, Spain & University of Lisbon, Portugal</i>	
Major Techniques and Current Developments of Supply Chain Process Modelling.....	5433
<i>Henry Xu, The University of Queensland, Australia</i>	
<i>Renae Agrey, The University of Queensland, Australia</i>	

Making Sense of IS Project Stories	5660
<i>Darren Dalcher, University of Hertfordshire, UK</i>	
Managerial Tools and Techniques for Decision Making	2166
<i>Davood Askarany, University of Auckland, New Zealand</i>	
Managing and Visualizing Unstructured Big Data.....	394
<i>Ananda Mitra, Wake Forest University, USA</i>	
Manufacturing vs. Services and the Role of Information Technology	7234
<i>Arnab Adhikari, Indian Institute of Management Ranchi, India</i>	
<i>Shromona Ganguly, Indian Institute of Management Calcutta, India</i>	
Mapping the Dissemination of the Theory of Social Representations via Academic Social Networks	7044
<i>Annamaria Silvana de Rosa, Sapienza University of Rome, Italy</i>	
<i>Laura Dryjanska, Sapienza University of Rome, Italy</i>	
<i>Elena Bocci, Sapienza University of Rome, Italy</i>	
Marketing and Marketing Plan for Information Services	5757
<i>Sérgio Maravilhas, Universidade Salvador, Brazil</i>	
Massive Digital Libraries (MDLs).....	5226
<i>Andrew Philip Weiss, California State University – Northridge, USA</i>	
Massive Open Online Courses and Integrating Open Source Technology and Open Access Literature Into Technology-Based Degrees	7898
<i>Maurice Dawson, University of Missouri – St. Louis, USA</i>	
<i>Sharon Burton, Grand Canyon University, USA</i>	
<i>Dustin Bessette, National Graduate School of Quality Management, USA</i>	
<i>Jorja Wright, University of Charleston, USA</i>	
Mastering Electronic Government in the Digital Age	3591
<i>Kijpokin Kasemsap, Suan Sunandha Rajabhat University, Thailand</i>	
A Maturity Model for Digital Literacies and Sustainable Development	2280
<i>Ravi S. Sharma, Nanyang Technological University, Singapore</i>	
<i>Lin G. Malone, Nanyang Technological University, Singapore</i>	
<i>Chong Guan, SIM University, Singapore</i>	
<i>Ambica Dattakumar, Nanyang Technological University, Singapore</i>	
The Measurement and Recognition of Intellectual Capital in the Process of Accounting Convergence Trends and Patterns	5669
<i>Ionica Oncioiu, Titu Maiorescu University, Romania</i>	

Measuring Low Carbon Supply Chain.....	5446
<i>Muhammad Shabir Shaharudin, Universiti Sains Malaysia, Malaysia</i>	
<i>Yudi Fernando, Universiti Malaysia Pahang, Malaysia</i>	
Measuring Text Readability Using Reading Level	1499
<i>James C. Brewer, Texas Tech University, USA</i>	
Mechanisms of Electrical Conductivity in Carbon Nanotubes and Graphene.....	2673
<i>Rafael Vargas-Bernal, Instituto Tecnológico Superior de Irapuato, Mexico</i>	
Mediated Embodiment in New Communication Technologies	4234
<i>Laura Aymerich-Franch, CNRS-AIST JRL (Joint Robotics Laboratory), AIST, Japan</i>	
Medical Equipment and Economic Determinants of Its Structure and Regulation in the Slovak Republic	5841
<i>Beáta Gavurová, Technical University of Košice, Slovakia</i>	
<i>Viliam Kováč, Technical University of Košice, Slovakia</i>	
<i>Michal Šoltés, Technical University of Košice, Slovakia</i>	
Metadata Standards in Digital Audio.....	6447
<i>Kimmy Szeto, Baruch College, City University of New York, USA</i>	
Methodology of Climate Change Impact Assessment on Forests	3114
<i>Mostafa Jafari, Regional Institute of Forest and Rangelands (RIFR), Iran</i>	
Methods for Improving Alias Rejections in Comb Filters.....	4746
<i>Gordana Jovanovic Dolecek, Institute INAOE Puebla, Mexico</i>	
Methods for Simultaneous Improvement of Comb Pass Band and Folding Bands	6171
<i>Gordana Jovanovic Dolecek, Institute INAOE Puebla, Mexico</i>	
Micro to Macro Social Connectedness through Mobile Phone Engagement	6184
<i>Dominic Mentor, Columbia University, USA</i>	
Mining Big Data and Streams.....	406
<i>Hoda Ahmed Abdelhafez, Suez Canal University, Egypt</i>	
Mining Sport Activities	7348
<i>Iztok Fister Jr., University of Maribor, Slovenia</i>	
<i>Iztok Fister, University of Maribor, Slovenia</i>	
Missing Part of Halal Supply Chain Management.....	5456
<i>Ratih Hendayani, Telkom University, Indonesia</i>	
<i>Yudi Fernando, Universiti Malaysia Pahang, Malaysia</i>	

Mission, Tools, and Ongoing Developments in the So.Re.Com. “A.S. de Rosa” @-library	5237
<i>Annamaria Silvana de Rosa, Sapienza University of Rome, Italy</i>	
Mobile Applications for Automatic Object Recognition	6195
<i>Danilo Avola, University of Udine, Italy</i>	
<i>Gian Luca Foresti, University of Udine, Italy</i>	
<i>Claudio Piciarelli, University of Udine, Italy</i>	
<i>Marco Vernier, University of Udine, Italy</i>	
<i>Luigi Cinque, Sapienza University, Italy</i>	
Mobile Apps Threats.....	6207
<i>Donovan Peter Chan Wai Loon, University of Malaya, Malaysia</i>	
<i>Sameer Kumar, University of Malaya, Malaysia</i>	
Mobile Game-Based Learning.....	6361
<i>Boaventura DaCosta, Solers Research Group, USA</i>	
<i>Soonhwa Seok, Korea University, South Korea</i>	
<i>Carolyn Kinsell, Solers Research Group, USA</i>	
Mobile Game-Based Learning in STEM Subjects.....	6376
<i>Marcelo Leandro Eichler, Universidade Federal do Rio Grande do Sul, Brazil</i>	
<i>Gabriela Trindade Perry, Universidade Federal do Rio Grande do Sul, Brazil</i>	
<i>Ivana Lima Lucchesi, Universidade Federal do Rio Grande do Sul, Brazil</i>	
<i>Thiago Troina Melendez, Instituto Federal Sul-Riograndense, Brazil</i>	
Mobile Learning in and out of the K-12 Classroom	6388
<i>Pena L. Bedesem, Kent State University, USA</i>	
<i>Tracy Arner, Kent State University, USA</i>	
Mobile Technologies Impact on Economic Development in Sub-Saharan Africa	6216
<i>Adam Crossan, Letterkenny Institute of Technology, Ireland</i>	
<i>Nigel McKelvey, Letterkenny Institute of Technology, Ireland</i>	
<i>Kevin Curran, Ulster University, UK</i>	
Mobile Testing System for Developing Language Skills.....	5116
<i>Svetlana Titova, Far Eastern Federal University, Russia</i>	
Mobile Virtual Reality to Enhance Subjective Well-Being	6223
<i>Federica Pallavicini, Università di Milano-Bicocca, Italy</i>	
<i>Luca Morganti, Università di Milano-Bicocca, Italy</i>	
<i>Barbara Diana, Università di Milano-Bicocca, Italy</i>	
<i>Olivia Realdon, University of Milano-Bicocca, Italy</i>	
<i>Valentino Zurloni, Università di Milano-Bicocca, Italy</i>	
<i>Fabrizia Mantovani, Università di Milano-Bicocca, Italy</i>	

Model for Assessment of Environmental Responsibility in Health Care Organizations	3131
<i>María Carmen Carnero, University of Castilla-La Mancha, Spain & University of Lisbon, Portugal</i>	
A Model for Connected E-Government in the Digital Age	3602
<i>Qiuyan Fan, Western Sydney University, Australia</i>	
Model-Driven Software Modernization.....	7447
<i>Liliana Maria Favre, Universidad Nacional Del Centro De La Provincia De Buenos Aires, Argentina</i>	
<i>Liliana Martinez, Universidad Nacional del Centro de la Provincia de Buenos Aires, Argentina</i>	
<i>Claudia Teresa Pereira, Universidad Nacional del Centro de la Provincia de Buenos Aires, Argentina</i>	
The Morality of Reporting Safety Concerns in Aviation.....	3194
<i>Kawtar Tani, UCOL, New Zealand</i>	
Motivational Factors of Telework	740
<i>Arlene J. Nicholas, Salve Regina University, USA</i>	
Moving Object Detection and Tracking Based on the Contour Extraction and Centroid Representation.....	212
<i>Naveenkumar M, National Institute of Technology Trichy, India</i>	
<i>Sriharsha K. V., National Institute of Technology Trichy, India</i>	
<i>Vadivel A, National Institute of Technology Trichy, India</i>	
Multifaceted Applications of the Internet of Things.....	7775
<i>Kijpokin Kasemsap, Suan Sunandha Rajabhat University, Thailand</i>	
Multimedia-Enabled Dot Codes as Communication Technologies	6464
<i>Shigeru Ikuta, Otsuma Women's University, Japan</i>	
Multimodal Literacy	1508
<i>Maryann Tatum Tobin, University of Miami, USA</i>	
Mutation Testing Applied to Object-Oriented Languages.....	7459
<i>Pedro Delgado-Pérez, University of Cádiz, Spain</i>	
<i>Inmaculada Medina-Bulo, University of Cádiz, Spain</i>	
<i>Juan José Domínguez-Jiménez, University of Cádiz, Spain</i>	
The Nature, Extent, Causes, and Consequences of Cyberbullying.....	1723
<i>Michelle F. Wright, Masaryk University, Czech Republic</i>	
A Nature-Inspired Metaheuristic Approach for Generating Alternatives.....	2178
<i>Julian Scott Yeomans, York University, Canada</i>	

The Nature of Cyber Bullying Behaviours	4245
<i>Lucy R. Betts, Nottingham Trent University, UK</i>	
The Nature of Research Methodologies.....	6756
<i>Ben Tran, Alliant International University, USA</i>	
Need for Rethinking Modern Urban Planning Strategies Through Integration of ICTs.....	7843
<i>Rounaq Basu, Indian Institute of Technology Bombay, India</i>	
<i>Arnab Jana, Indian Institute of Technology Bombay, India</i>	
Negotiating Local Norms in Online Communication	1217
<i>Jonathan R. White, Höskolan Dalarna, Sweden</i>	
Neighborhood Rough-Sets-Based Spatial Data Analytics	1835
<i>Sharmila Banu K, VIT University, India</i>	
<i>B. K. Tripathy, VIT University, India</i>	
The NetLab Network.....	7057
<i>Dimitrina Dimitrova, York University, Canada</i>	
<i>Barry Wellman, NetLab Network, Canada</i>	
The Networked Effect of Children and Online Digital Technologies.....	7312
<i>Teresa Sofia Pereira Dias de Castro, University of Minho, Portugal</i>	
<i>António Osório, University of Minho, Portugal</i>	
<i>Emma Bond, University Campus Suffolk, UK</i>	
Neural Networks and Their Accelerated Evolution From an Economic Analysis Perspective	6579
<i>Stelian Stancu, Bucharest University of Economic Studies, Romania</i>	
<i>Constanța-Nicoleta Bodea, Bucharest University of Economic Studies, Romania</i>	
<i>Oana Mădălina Popescu(Predescu), Bucharest University of Economic Studies, Romania</i>	
<i>Alina Neamțu(Idorași), Bucharest University of Economic Studies, Romania</i>	
A Neuroaesthetic Approach to the Search of Beauty From the Consumer's Perspective.....	5767
<i>Gemma García Ferrer, Rey Juan Carlos University, Spain</i>	
Neuroscience Technology and Interfaces for Speech, Language, and Musical Communication	5886
<i>Dionysios Politis, Aristotle University of Thessaloniki, Greece</i>	
<i>Miltiadis Tsalighopoulos, Aristotle University of Thessaloniki, Greece</i>	
<i>Georgios Kyriafinis, AHEPA University Hospital, Greece</i>	
New Advances in E-Commerce	2810
<i>Khaled Ahmed Nagaty, The British University in Egypt, Egypt</i>	
New Faces of Digital Divide and How to Bridge It.....	7248
<i>Viktor Freiman, Université de Moncton, Canada</i>	
<i>Dragana Martinovic, University of Windsor, Canada</i>	
<i>Xavier Robichaud, Université de Moncton, Canada</i>	

New Perspectives of Pattern Recognition for Automatic Credit Card Fraud Detection	4937
<i>Addisson Salazar, Universitat Politècnica de València, Spain</i>	
<i>Gonzalo Safont, Universitat Politècnica de València, Spain</i>	
<i>Alberto Rodriguez, Universidad Miguel Hernández de Elche, Spain</i>	
<i>Luis Vergara, Universitat Politècnica de València, Spain</i>	
Nigerian Undergraduate Students' Computer Competencies and Use of Information Technology Tools and Resources for Study Skills and Habits' Enhancement	2303
<i>Adekunle Olusola Otunla, University of Ibadan, Nigeria</i>	
<i>Caleb Okoro Amuda, University of Ibadan, Nigeria</i>	
Noise Trader.....	71
<i>Po-Keng Cheng, State University of New York, Stony Brook University, USA</i>	
Nominalizations in Requirements Engineering Natural Language Models.....	5127
<i>Claudia S. Litvak, Universidad Nacional de La Matanza, Argentina & Universidad Nacional del Oeste, Argentina</i>	
<i>Graciela Dora Susana Hada, Universidad Nacional del Oeste, Argentina</i>	
<i>Jorge Horacio Doorn, Universidad Nacional del Oeste, Argentina & Universidad Nacional de La Matanza, Argentina</i>	
Notions of Maritime Green Supply Chain Management	5465
<i>Fairuz Jasmi, Universiti Sains Malaysia, Malaysia</i>	
<i>Yudi Fernando, Universiti Malaysia Pahang, Malaysia</i>	
Novel Methods to Design Low-Complexity Digital Finite Impulse Response (FIR) Filters.....	6234
<i>David Ernesto Troncoso Romero, CONACYT at ESCOM-IPN, Mexico</i>	
<i>Gordana Jovanovic Dolecek, Institute INAOE, Mexico</i>	
N-Tuple Algebra as a Unifying System to Process Data and Knowledge.....	1995
<i>Boris Alexandrovich Kulik, Institute of Problems in Mechanical Engineering RAS, Russia</i>	
<i>Alexander Yakovlevich Fridman, Institute for Informatics and Mathematical Modelling, Kola Science Centre of RAS, Russia</i>	
Object-Oriented Programming in Computer Science	7470
<i>Rahime Yilmaz, Istanbul University, Turkey</i>	
<i>Anil Sezgin, Yildiz Technical University, Turkey</i>	
<i>Sefer Kurnaz, Istanbul Esenyurt University, Turkey</i>	
<i>Yunus Ziya Arslan, Istanbul University, Turkey</i>	
Offshoring IT	5476
<i>Susan Cockrell, Austin Peay State University, USA</i>	
<i>Terry Stringer Damron, Austin Peay State University, USA</i>	
<i>Amye M. Melton, Austin Peay State University, USA</i>	
<i>Alan D. Smith, Robert Morris University, USA</i>	

Online Academia	2580
<i>Magdalena Bielenia-Grajewska, University of Gdansk, Poland</i>	
On-Line Credit and Debit Card Processing and Fraud Prevention for E-Business	2707
<i>James Williams, University of Pittsburgh, USA</i>	
Online Dating/Dating Apps	7069
<i>Vladimir Santiago Arias, Texas Tech University, USA</i>	
<i>Narissra Maria Punyanunt-Carter, Texas Tech University, USA</i>	
Online Information Retrieval Systems Trending From Evolutionary to Revolutionary Approach.....	4535
<i>Zahid Ashraf Wani, University of Kashmir, India</i>	
<i>Huma Shafiq, University of Kashmir, India</i>	
Online Learning Propelled by Constructivism	2588
<i>Kathaleen Reid-Martinez, Oral Roberts University, USA</i>	
<i>Linda D. Grooms, Regent University, USA</i>	
Online Mediation in E-Commerce Matters.....	2825
<i>Ángela Coello Pulido, University of Vigo, Spain</i>	
Online Prosocial Behaviors.....	7077
<i>Michelle F. Wright, Pennsylvania State University, USA</i>	
<i>William Stanley Pendergrass, American Public University System, USA</i>	
Online Social Networking Behavior and Its Influence Towards Students' Academic Performance	7088
<i>Maslin Masrom, Universiti Teknologi Malaysia, Malaysia</i>	
<i>Selisa Usat, Universiti Teknologi Malaysia, Malaysia</i>	
The Ontology of Randomness	1845
<i>Jeremy Horne, The International Institute of Informatics and Systemics, USA</i>	
Open Data and High-Tech Startups Towards Nascent Entrepreneurship Strategies.....	3032
<i>Fotis Kitsios, University of Macedonia, Greece</i>	
<i>Maria Kamariotou, University of Macedonia, Greece</i>	
Open Data Repositories in Knowledge Society	4436
<i>Nadim Akhtar Khan, University of Kashmir, India</i>	
<i>Sara Sohrabzadeh, Tehran University of Medical Science, Iran</i>	
<i>Garvita Jhamb, University of Delhi, India</i>	
An Open Learning Format for Lifelong Learners' Empowerment.....	1517
<i>Sabrina Leone, Università Politecnica delle Marche, Italy</i>	

Open Source.....	6245
<i>Heidi Lee Schnackenberg, SUNY Plattsburgh, USA</i>	
Open Source Software Virtual Learning Environment (OSS-VLEs) in Library Science Schools	7912
<i>Rosy Jan, University of Kashmir, India</i>	
The Optimal Workforce Staffing Solutions With Random Patient Demand in Healthcare Settings.....	3711
<i>Alexander Kolker, GE Healthcare, USA</i>	
Optimization of Antenna Arrays and Microwave Filters Using Differential Evolution Algorithms	6595
<i>Sotirios K. Goudos, Aristotle University of Thessaloniki, Greece</i>	
Optimizing Cloud Computing Costs of Services for Consumers	1627
<i>Eli Weintraub, Afeka Tel Aviv College of Engineering, Israel</i>	
<i>Yuval Cohen, Afeka Tel Aviv College of Engineering, Israel</i>	
Order Statistics and Applications.....	1856
<i>E. Jack Chen, BASF Corporation, USA</i>	
Organizational Transparency	754
<i>Gustavo de Oliveira Almeida, Federal University of the State of Rio de Janeiro, Brazil</i>	
<i>Claudia Cappelli, Federal University of the State of Rio de Janeiro, Brazil</i>	
<i>Cristiano Maciel, Federal University of Mato Grosso, Brazil</i>	
An Overview of Crowdsourcing	8023
<i>Eman Younis, Minia University, Egypt</i>	
A Paradoxical World and the Role of Technology in Thana-Capitalism.....	4761
<i>Maximiliano Emanuel Korstanje, University of Palermo, Argentina</i>	
Parallel Development of Three Major Space Technology Systems and Human Side of Information Reference Services as an Essential Complementary Method	3484
<i>Joyce Gosata Maphanyane, University of Botswana, Botswana</i>	
Parental Mediation of Adolescent Technology Use.....	7097
<i>J. Mitchell Vaterlaus, Montana State University, USA</i>	
Particle Shape Analysis Using Digital Image Processing.....	1331
<i>Katia Tannous, University of Campinas, Brazil</i>	
<i>Fillipe Souza Silva, University of Campinas, Brazil</i>	
The Past, Present, and Future of UML.....	7481
<i>Rebecca Platt, Murdoch University, Australia</i>	
<i>Nik Thompson, Curtin University, Australia</i>	

Peer-to-Peer Health-Related Online Support Groups	3767
<i>Neil S. Coulson, University of Nottingham, UK</i>	
Performance Appraisal.....	4337
<i>Chandra Sekhar Patro, Gayatri Vidya Parishad College of Engineering (Autonomous), India</i>	
Performance Measurement of Technology Ventures by Science and Technology Institutions	4774
<i>Artie W. Ng, The Hong Kong Polytechnic University, Hong Kong</i>	
<i>Benny C. F. Cheung, The Hong Kong Polytechnic University, Hong Kong</i>	
<i>Peggy M. L. Ng, The Hong Kong Polytechnic University, Hong Kong</i>	
Personalized Medicine	5901
<i>Sandip Bisui, Indian Institute of Technology (IIT) Kanpur, India</i>	
<i>Subhas Chandra Misra, Indian Institute of Technology (IIT) Kanpur, India</i>	
Pervasive Mobile Health.....	5908
<i>Muhammad Anshari, Universiti Brunei Darussalam, Brunei</i>	
<i>Mohammad Nabil Almunawar, Universiti Brunei Darussalam, Brunei</i>	
Petri Nets Identification Techniques for Automated Modelling of Discrete Event Processes	7488
<i>Edelma Rodriguez-Perez, CINVESTAV Unidad Guadalajara, Mexico</i>	
<i>Ernesto Lopez-Mellado, CINVESTAV Unidad Guadalajara, Mexico</i>	
Piracy and Intellectual Property Theft in the Internet Era.....	1656
<i>Shun-Yung Kevin Wang, University of South Florida – St. Petersburg, USA</i>	
<i>Jeremy J McDaniel, Principal Financial Group, USA</i>	
Political Context Elements in Public Policy of Radio Frequency Information Technology and Electromagnetic Fields.....	6710
<i>Joshua M. Steinfeld, Old Dominion University, USA</i>	
Potential Benefits and Current Limits in the Development of Demand Response	3144
<i>Clementina Bruno, University of Piemonte Orientale, Italy</i>	
The Potential Role of the Software Industry in Supporting Economic Development	7259
<i>Sherif H. Kamel, The American University in Cairo, Egypt</i>	
Power Consumption in Wireless Access Networks	6253
<i>Vinod Kumar Mishra, B. T. Kumaon Institute of Technology, India</i>	
<i>Pankaja Bisht, B. T. Kumaon Institute of Technology, India</i>	
Predicting Students Grades Using Artificial Neural Networks and Support Vector Machine.....	5169
<i>Sajid Umair, National University of Sciences and Technology (NUST), Pakistan</i>	
<i>Muhammad Majid Sharif, National University of Sciences and Technology (NUST), Pakistan</i>	
Predictive Analytics and Intelligent Risk Detection in Healthcare Contexts.....	6806
<i>Nilmini Wickramasinghe, Epworth HealthCare, Australia & Deakin University, Australia</i>	

Preferences, Utility, and Stochastic Approximation	2188
<i>Yuri P. Pavlov, Bulgarian Academy of Sciences, Institute of Information and Communication Technologies, Bulgaria</i>	
<i>Rumen D. Andreev, Bulgarian Academy of Sciences, Institute of Information and Communication Technologies, Bulgaria</i>	
Presidential Elections Web 2.0	3612
<i>Ramona Sue McNeal, University of Northern Iowa, USA</i>	
<i>Lisa Dotterweich Bryan, Upper Iowa University, USA</i>	
Pricing Based on Real-Time Analysis of Forklift Utilization Using RFID in Warehouse Management.....	5490
<i>Numan Celebi, Sakarya University, Turkey</i>	
<i>Kübra Savaş, Istanbul University, Turkey</i>	
<i>Ihsan Hakan Selvi, Sakarya University, Turkey</i>	
The Principle and Process of Digital Fabrication of Biomedical Objects	505
<i>S. H. Choi, The University of Hong Kong, Hong Kong</i>	
<i>H. H. Cheung, The University of Hong Kong, Hong Kong</i>	
<i>W. K. Zhu, The University of Hong Kong, Hong Kong</i>	
Privacy, Algorithmic Discrimination, and the Internet of Things	4951
<i>Jenifer Sunrise Winter, University of Hawaii at Manoa, USA</i>	
The Process Model of Gameplay to Understand Digital Gaming Outcomes	3317
<i>Linda K. Kaye, Edge Hill University, UK</i>	
Profit Maximizing Network Modeling With Inventory and Capacity Considerations	5503
<i>Tan Miller, Rider University, USA</i>	
<i>Renato de Matta, University of Iowa, USA</i>	
Project Control Using a Bayesian Approach.....	5679
<i>Franco Caron, Politecnico di Milano, Italy</i>	
Project Management in Government	3621
<i>Shauneen Furlong, University of Ottawa, Canada & John Moores Liverpool University, UK</i>	
Promoting Strategic Human Resource Management, Organizational Learning, and Knowledge Management in Modern Organizations.....	4347
<i>Kijpokin Kasemsap, Suan Sunandha Rajabhat University, Thailand</i>	
A Proposed Framework for Incorporating Big-Data Technology in National Crisis Management Center.....	2006
<i>Magdy M. Kabeil, Al-Yamamah University, Saudi Arabia</i>	
<i>Ahmad M. Kabil, University of Wisconsin – Whitewater, USA</i>	

The Protection Policy for Youth Online in Japan	4962
<i>Nagayuki Saito, Ochanomizu University, Japan</i>	
<i>Madoka Aragaki, Business Breakthrough University, Japan</i>	
A Psychological Perspective on Mobile Learning	6398
<i>Melody M. Terras, University of the West of Scotland, UK</i>	
<i>Judith Ramsay, Manchester Metropolitan University, UK</i>	
Public Policies for Providing Cloud Computing Services to SMEs of Latin America.....	6727
<i>Mohd Nayyer Rahman, Aligarh Muslim University, India</i>	
<i>Badar Alam Iqbal, Aligarh Muslim University, India</i>	
QoS Architectures for the IP Network	6609
<i>Harry G. Perros, North Carolina State University, USA</i>	
The Qualities and Potential of Social Media	7106
<i>Udo Richard Averweg, eThekweni Municipality, South Africa</i>	
<i>Marcus Leaning, University of Winchester, UK</i>	
Quality Evaluation for Evolving Conceptual Database Design	2020
<i>Elvira Immacolata Locuratolo, CNR ISTI, Italy</i>	
Quality Online Learning in Higher Education.....	3930
<i>Deborah G. Wooldridge, Bowling Green State University, USA</i>	
<i>Sandra Poirier, Middle Tennessee State University, USA</i>	
<i>Julia M. Matuga, Bowling Green State University, USA</i>	
Quantum Computing and Quantum Communication	7715
<i>Göran Pulkkis, Arcada University of Applied Sciences, Finland</i>	
<i>Kaj J. Grahn, Arcada University of Applied Sciences, Finland</i>	
Quantum Information Science Vis-à-Vis Information Schools	4448
<i>P. K. Paul, Raiganj University, India</i>	
<i>D. Chatterjee, Seacom Skills University, India</i>	
<i>A. Bhumali, Raiganj University, India</i>	
Query Languages for Graph Databases	2031
<i>Kornelije Rabuzin, University of Zagreb, Croatia</i>	
Radio Frequency Identification Systems Within a Lean Supply Chain in a Global Environment ...	5516
<i>Alan D. Smith, Robert Morris University, USA</i>	
<i>Terry Stringer Damron, Austin Peay State University, USA</i>	
<i>Susan Cockrell, Austin Peay State University, USA</i>	
<i>Amye M. Melton, Austin Peay State University, USA</i>	

Radio Frequency Identification Technologies and Issues in Healthcare	5918
<i>Amber A. Smith-Ditizio, Texas Woman's University, USA</i>	
<i>Alan D. Smith, Robert Morris University, USA</i>	
Recommender Technologies and Emerging Applications	1869
<i>Young Park, Bradley University, USA</i>	
Reconstructive Architectural and Urban Digital Modelling	7856
<i>Roberta Spallone, Politecnico di Torino, Italy</i>	
Recurrent Neural Networks for Predicting Mobile Device State.....	6658
<i>Juan Manuel Rodriguez, ISISTAN, UNICEN-CONICET, Argentina</i>	
<i>Alejandro Zunino, ISISTAN, UNICEN-CONICET, Argentina</i>	
<i>Antonela Tommasel, ISISTAN, UNICEN-CONICET, Argentina</i>	
<i>Cristian Mateos, ISISTAN, UNICEN-CONICET, Argentina</i>	
Reflection as a Process From Theory to Practice.....	1529
<i>Sonia Bharwani, Indian School of Management and Entrepreneurship, India</i>	
<i>Durgamohan Musunuri, Bhavan's Usha and Lakshmi Mittal Institute of Management, India</i>	
Reflections of the 1Malaysia Supply Chain (1MSC).....	5527
<i>Munira Halili, Universiti Sains Malaysia, Malaysia</i>	
<i>Latifah Naina Mohamed, Universiti Sains Malaysia, Malaysia</i>	
<i>Yudi Fernando, Universiti Malaysia Pahang, Malaysia</i>	
Regional Development and Air Freight Service Needs for Regional Communities.....	7869
<i>Tarryn Kille, Griffith University, Australia</i>	
<i>Paul Bates, University of Southern Queensland, Australia</i>	
<i>Patrick S. Murray, University of Southern Queensland, Australia</i>	
Relationship Among Intelligence, Achievement Motivation, Type of School, and Academic Performance of Kenyan Urban Primary School Pupils	1540
<i>Jessina Mukomunene Muthee, Kenyatta University, Kenya</i>	
<i>Catherine G. Murungi, Kenyatta University, Kenya</i>	
The Relationship Between Online Formative Assessment and State Test Scores Using Multilevel Modeling	5183
<i>Aryn C. Karpinski, Kent State University, USA</i>	
<i>Jerome V. D'Agostino, The Ohio State University, USA</i>	
<i>Anne-Evan K. Williams, Billings Middle School, USA</i>	
<i>Sue Ann Highland, Grand Canyon University, USA</i>	
<i>Jennifer A. Mellott, Kent State University, USA</i>	
Reputational Mechanisms in Consumer-to-Consumer Online Commerce.....	2833
<i>Mikhail I. Melnik, Kennesaw State University, USA</i>	

Research and Development on Software Testing Techniques and Tools.....	7503
<i>Tamilarasi T, VIT University, India</i>	
<i>M. Prasanna, VIT University, India</i>	
Research Methodology	6767
<i>Swati C. Jagdale, MAEER's Maharashtra Institute of Pharmacy, India</i>	
<i>Rahul U. Hude, MAEER's Maharashtra Institute of Pharmacy, India</i>	
<i>Aniruddha R. Chabukswar, MAEER's Maharashtra Institute of Pharmacy, India</i>	
Resource Management for Multimedia Services in Long Term Evaluation Networks	6266
<i>Vinod Kumar Mishra, B. T. Kumaon Institute of Technology, India</i>	
<i>Tanuja Pathak, B. T. Kumaon Institute of Technology, India</i>	
Retail Prices and E-Commerce	2841
<i>Jihui Chen, Illinois State University, USA</i>	
Reverse Engineering in Rehabilitation	521
<i>Emilia Mikołajewska, Nicolaus Copernicus University, Poland</i>	
<i>Marek Macko, Kazimierz Wielki University, Poland</i>	
<i>Zbigniew Szczepański, Kazimierz Wielki University, Poland</i>	
<i>Dariusz Mikołajewski, Kazimierz Wielki University, Poland</i>	
A Review of Advances in Supply Chain Intelligence.....	5538
<i>Nenad Stefanovic, University of Kragujevac, Serbia</i>	
<i>Danijela Milosevic, University of Kragujevac, Serbia</i>	
A Review of Supply Chain Risk Management in Agribusiness Industry	5550
<i>Sri Widiyanesti, Telkom University, Indonesia</i>	
<i>Yudi Fernando, Universiti Malaysia Pahang, Malaysia</i>	
Revisiting Web 2.0	8036
<i>Michael Dinger, University of South Carolina Upstate, USA</i>	
<i>Varun Grover, Clemson University, USA</i>	
RNA Interference Therapeutics and Human Diseases.....	477
<i>Dolly Sharma, PEC University of Technology, India</i>	
<i>Shailendra Singh, PEC University of Technology, India</i>	
<i>Trilok Chand, PEC University of Technology, India</i>	
Robotics and Programming Integration as Cognitive-Learning Tools	6859
<i>Nikleia Eteokleous, Frederick University Cyprus, Cyprus</i>	
The Role of Distance Education in Global Education	6412
<i>Kijpokin Kasemsap, Suan Sunandha Rajabhat University, Thailand</i>	

Role of Educational Leaders in Supporting Beginning Teachers in Al Ain Schools in the UAE	7647
<i>Salam Omar Ali, Brighton Collage Al Ain, UAE</i>	
The Role of Emerging Information Technologies for Supporting Supply Chain Management	5559
<i>Zlatko Nedelko, University of Maribor, Slovenia</i>	
<i>Vojko Potocan, University of Maribor, Slovenia</i>	
The Role of Feedback in Software Process Assessment.....	7514
<i>Zeljko Stojanov, University of Novi Sad, Serbia</i>	
<i>Dalibor Dobrilovic, University of Novi Sad, Serbia</i>	
The Role of U-FADE in Selecting Persuasive System Features.....	7785
<i>Isaac Wiafe, Ghana Institute of Management and Public Administration, Ghana</i>	
The Roles of Digital Literacy in Social Life of Youth.....	2314
<i>Dragana Martinovic, University of Windsor, Canada</i>	
<i>Viktor Freiman, Université de Moncton, Canada</i>	
<i>Chrispina S. Lekule, St. Augustine University of Tanzania, Tanzania</i>	
<i>Yuqi Yang, University of Windsor, Canada</i>	
Rough-Set-Based Decision Model for Incomplete Information Systems	2200
<i>Safiye Turgay, Sakarya University, Turkey</i>	
<i>Orhan Torkul, Sakarya University, Turkey</i>	
<i>Tahsin Turgay, Sakarya University, Turkey</i>	
Safeguarding of ATM	77
<i>Srividhya Srinivasan, University of Madras, India</i>	
<i>Priya Krishnamoorthy, SASTRA University, India</i>	
<i>Raghuraman Koteeswaran, SASTRA University, India</i>	
Samsung Company and an Analysis of Supplier-Side Supply Chain Management and IT Applications	5570
<i>Amber A. Smith-Ditizio, Texas Woman's University, USA</i>	
<i>Alan D. Smith, Robert Morris University, USA</i>	
Scanning for Blind Spots	899
<i>Barbara Jane Holland, Brooklyn Public Library, USA</i>	
Schema Evolution in Conventional and Emerging Databases	2043
<i>Zouhaier Brahmia, University of Sfax, Tunisia</i>	
<i>Fabio Grandi, University of Bologna, Italy</i>	
<i>Barbara Oliboni, University of Verona, Italy</i>	
<i>Rafik Bouaziz, University of Sfax, Tunisia</i>	

Schema Versioning in Conventional and Emerging Databases	2054
<i>Zouhaier Brahmia, University of Sfax, Tunisia</i>	
<i>Fabio Grandi, University of Bologna, Italy</i>	
<i>Barbara Oliboni, University of Verona, Italy</i>	
<i>Rafik Bouaziz, University of Sfax, Tunisia</i>	
Scholarly Identity in an Increasingly Open and Digitally Connected World.....	6779
<i>Olga Belikov, Brigham Young University, USA</i>	
<i>Royce Kimmons, Brigham Young University, USA</i>	
Science Animation and Students' Attitudes.....	2599
<i>Sivasankar Arumugam, Sri Venkateswara College of Education, India</i>	
<i>Nancy Nirmala, Christ the King Matric Higher Secondary School, India</i>	
Science, Ethics, and Weapons Research	3205
<i>John Forge, Independent Researcher, Australia</i>	
Screencasts and Learning Styles	1548
<i>Rui Alberto Jesus, CESPU, Instituto de Investigação e Formação Avançada em Ciências e Tecnologias da Saúde, Portugal</i>	
Screen Culture.....	4255
<i>Ana Melro, University of Aveiro, Portugal</i>	
<i>Lídia Oliveira, University of Aveiro, Portugal</i>	
Search Engine Optimization	8046
<i>Dimitrios Giomelakis, Aristotle University of Thessaloniki, Greece</i>	
<i>Andreas A. Veglis, Aristotle University of Thessaloniki, Greece</i>	
Secure Group Key Sharing Protocols and Cloud System	1667
<i>Vaishali Ravindra Thakare, VIT University, India</i>	
<i>John Singh K, VIT University, India</i>	
Secure Software Development of Cyber-Physical and IoT Systems.....	7525
<i>Muthu Ramachandran, Leeds Metropolitan University, UK</i>	
Security of Identity-Based Encryption Algorithms	4975
<i>Kannan Balasubramanian, Mepco Schlenk Engineering College, India</i>	
<i>M. Rajakani, Mepco Schlenk Engineering College, India</i>	
Security of Internet-, Intranet-, and Computer-Based Examinations in Terms of Technical, Authentication, and Environmental, Where Are We?.....	1676
<i>Babak Sokouti, Tabriz University of Medical Sciences, Iran</i>	
<i>Massoud Sokouti, Mashhad University of Medical Sciences, Iran</i>	

Self-Awareness and Motivation Contrasting ESL and NEET Using the SAVE System	1559
<i>Laura Vettrano, Learning Community, Italy</i>	
<i>Valentina Castello, CIOFS FP, Italy</i>	
<i>Marco Guspini, educommunity – Educational Community, Italy</i>	
<i>Eleonora Guglielman, Learning Community, Italy</i>	
Semantically Enhanced Authoring of Shared Media.....	6476
<i>Charalampos A. Dimoulas, Aristotle University of Thessaloniki, Greece</i>	
<i>Andreas A. Veglis, Aristotle University of Thessaloniki, Greece</i>	
<i>George Kalliris, Aristotle University of Thessaloniki, Greece</i>	
Semantic Intelligence.....	220
<i>Maria K. Koleva, Institute of Catalysis, Bulgarian Academy of Sciences, Bulgaria</i>	
Serious Games Advancing the Technology of Engaging Information	3327
<i>Peter A. Smith, University of Central Florida, USA</i>	
<i>Clint Bowers, University of Central Florida, USA</i>	
Serious Games in Entrepreneurship Education	800
<i>Fernando Almeida, Polytechnic Institute of Gaya, Portugal</i>	
<i>Jorge Simões, ISPGaya, Portugal</i>	
Service Quality and Perceived Value of Cloud Computing-Based Service Encounters.....	1129
<i>Eges Egedigwe, Dallas County Community College, USA</i>	
Shaping Mega-Science Projects and Practical Steps for Success	5690
<i>Phil Crosby, Curtin University, Australia</i>	
Short History of Social Networking and Its Far-Reaching Impact	7116
<i>Liguo Yu, Indiana University – South Bend, USA</i>	
Simulating Complex Supply Chain Relationships Using Copulas	5583
<i>Krishnamurty Muralidhar, University of Oklahoma, USA</i>	
<i>Rathindra Sarathy, Oklahoma State University, USA</i>	
The Skills of European ICT Specialists	4785
<i>Francesca Sgobbi, University of Brescia, Italy</i>	
Sleptsov Net Computing	7731
<i>Dmitry A. Zaitsev, International Humanitarian University, Ukraine</i>	
SMS & Civil Unrest.....	6275
<i>Innocent Chiluwa, Covenant University OTA, Nigeria</i>	

Social Business Process Modeling.....	765
<i>Fadwa Yahya, University of Sfax, Tunisia</i>	
<i>Khouloud Boukadi, University of Sfax, Tunisia</i>	
<i>Zakaria Maamar, Zayed University, UAE</i>	
<i>Hanêne Ben-Abdallah, King Abdulaziz University, Saudi Arabia</i>	
Social Commerce Using Social Network and E-Commerce.....	2851
<i>Roberto Marmo, University of Pavia, Italy</i>	
Social Computing.....	7796
<i>Nolan Hemmatazad, University of Nebraska at Omaha, USA</i>	
Social Issues in IT Project Teams:	777
<i>Awie C. Leonard, University of Pretoria, South Africa</i>	
<i>D. H. Van Zyl, University of Pretoria, South Africa</i>	
Social Media and Business Practices	7126
<i>Ashish Kumar Rathore, Indian Institute of Technology Delhi, India</i>	
<i>P. Vigneswara Ilavarasan, Indian Institute of Technology Delhi, India</i>	
Social Media Applications as Effective Service Delivery Tools for Librarians	5252
<i>Ihuoma Sandra Babatope, Delta State College of Physical Education, Nigeria</i>	
Social Media as a Channel of Constructive Dialogue for Tourism Businesses	4088
<i>Marios D. Sotiriadis, University of South Africa (UNISA), South Africa</i>	
Social Media Credit Scoring.....	7140
<i>Billie Anderson, Ferris State University, USA</i>	
<i>J. Michael Hardin, Samford University, USA</i>	
Social Media Use and Customer Engagement.....	5775
<i>Aurora Garrido-Moreno, University of Malaga, Spain</i>	
<i>Nigel Lockett, University of Lancaster, UK</i>	
<i>Víctor García-Morales, University of Granada, Spain</i>	
Social Network Analysis and the Study of University Industry Relations	7150
<i>Fernando Cabrita Romero, University of Minho, Portugal</i>	
Social Networking and Knowledge Sharing in Organizations.....	7161
<i>Sarabjot Kaur, Indian Institute of Technology Kanpur, India</i>	
<i>Subhas Chandra Misra, Indian Institute of Technology Kanpur, India</i>	
Social Telerehabilitation	5930
<i>Gilberto Marzano, Rezekne Academy of Technologies, Latvia</i>	
Socio-Economic Processes, User Generated Content, and Media Pluralism	7270
<i>Androniki Kavoura, Technological Educational Institute of Athens, Greece</i>	

Sociological Perspectives on Improving Medical Diagnosis Emphasizing CAD.....	1017
<i>Joel Fisher, Department of State, United States Government, USA</i>	
Socio-Technical Change Perspective for ERP Implementation in Large Scale Organizations.....	2975
<i>Jessy Nair, PES University, India</i>	
<i>D. Bhanusree Reddy, VIT University, India</i>	
<i>Anand A. Samuel, VIT University, India</i>	
Software Development Process Standards for Very Small Companies	6927
<i>Rory V. O'Connor, Dublin City University, Ireland</i>	
Software Evaluation From the Perspective of Patients and Healthcare Professionals.....	3782
<i>Rui Pedro Charters Lopes Rijo, Polytechnic Institute of Leiria, Portugal</i>	
<i>Domingos Alves, University of São Paulo, Brazil</i>	
Software Literacy	7539
<i>Elaine Khoo, University of Waikato, New Zealand</i>	
<i>Craig Hight, University of Newcastle, Australia</i>	
Software Process Improvement for Web-Based Projects Comparative View.....	7549
<i>Thamer Al-Rousan, Isra University, Pakistan</i>	
Sport Exergames for Physical Education	7358
<i>Pooya Soltani, University of Porto, Portugal</i>	
<i>João Paulo Vilas-Boas, University of Porto, Portugal</i>	
State of the Art and Key Design Challenges of Telesurgical Robotics.....	6872
<i>Sajid Nisar, National University of Science and Technology, Pakistan</i>	
<i>Osman Hasan, National University of Science and Technology, Pakistan</i>	
Staying Ahead in Business Through Innovation.....	5705
<i>N. Raghavendra Rao, FINAIT Consultancy Services, India</i>	
Steganography Using Biometrics.....	4985
<i>Manashee Kalita, NERIST, India</i>	
<i>Swanirbhar Majumder, NERIST, India</i>	
Strategic Information Systems Planning.....	912
<i>Maria Kamariotou, University of Macedonia, Greece</i>	
<i>Fotis Kitsios, University of Macedonia, Greece</i>	
Stress Testing Corporations and Municipalities and Supply Chains	6813
<i>Frank Wolf, CSSTA L3C, USA</i>	
A Study of Contemporary System Performance Testing Framework.....	7563
<i>Alex Ng, Federation University, Australia</i>	
<i>Shiping Chen, CSIRO Data61, Australia</i>	

A Study on Extensive Reading in Higher Education	3945
<i>Diana Presadă, Petroleum-Gas University of Ploiesti, Romania</i>	
<i>Mihaela Badea, Petroleum-Gas University of Ploiesti, Romania</i>	
Suggestions for Communication of Information for Multicultural Co-Existence	7327
<i>Noriko Kurata, Tokyo University of Science, Suwa, Japan</i>	
The Summers and Winters of Artificial Intelligence	229
<i>Tad Gonsalves, Sophia University, Japan</i>	
A Survey of People Localization Techniques Utilizing Mobile Phones.....	6286
<i>Levent Bayındır, Ataturk University, Turkey</i>	
Sustainable Advantages of Business Value of Information Technology	923
<i>Jorge A. Romero, Towson University, USA</i>	
Sustainable Competitive Advantage With the Balanced Scorecard Approach.....	5714
<i>Jorge Gomes, ISEG, Universidade de Lisboa, Portugal</i>	
<i>Mário José Batista Romão, ISEG, Universidade de Lisboa, Portugal</i>	
Swarm Intelligence for Multi-Objective Optimization in Engineering Design	239
<i>Janga Reddy Manne, Indian Institute of Technology Bombay, India</i>	
A Tale of Two Agile Requirements Engineering Practices	7577
<i>Pankaj Kamthan, Concordia University, Canada</i>	
<i>Terrill Fancott, Concordia University, Canada</i>	
Taxonomy for “Homo Consumens” in a 3.0 Era	1638
<i>Carlos Ballesteros, Universidad Pontificia Comillas, Spain</i>	
Teacher Presence.....	7922
<i>Caroline M. Crawford, University of Houston – Clear Lake, USA</i>	
Teaching Media and Information Literacy in the 21st Century	2292
<i>Sarah Gretter, Michigan State University, USA</i>	
<i>Aman Yadav, Michigan State University, USA</i>	
Technological Innovation and Use in the Early Days of Camera Phone Photo Messaging.....	6296
<i>Jonathan Lillie, Loyola University, USA</i>	
The Technological Pedagogical Content Knowledge of EFL Teachers (EFL TPACK)	7659
<i>Mehrak Rahimi, Shahid Rajaee Teacher Training University, Iran</i>	
<i>Shakiba Pourshahbaz, Shahid Rajaee Teacher Training University, Iran</i>	
Technology and Terror.....	3637
<i>Maximiliano Emanuel Korstanje, University of Palermo, Argentina</i>	
<i>Geoffrey Skoll, SUNY at Buffalo, USA</i>	

Technology Assessment of Information and Communication Technologies	4267
<i>Armin Grunwald, Karlsruhe Institute of Technology, Germany</i>	
<i>Carsten Orwat, Karlsruhe Institute of Technology, Germany</i>	
Technology Design and Routes for Tool Appropriation in Medical Practices	3794
<i>Manuel Santos-Trigo, Cinvestav-IPN, Mexico</i>	
<i>Ernesto Suaste, Cinvestav-IPN, Mexico</i>	
<i>Paola Figuerola, Cinvestav-IPN, Mexico</i>	
Technology Policies and Practices in Higher Education	3954
<i>Kelly McKenna, Colorado State University, USA</i>	
Technology, Learning Styles, Values, and Work Ethics of Millennials.....	4358
<i>Harish C. Chandan, Argosy University, USA</i>	
Telesurgical Robotics and a Kinematic Perspective	6882
<i>Sajid Nisar, National University of Sciences and Technology, Pakistan</i>	
<i>Osman Hasan, National University of Sciences and Technology, Pakistan</i>	
Theory and Practice of Online Knowledge Sharing	5093
<i>Will W. K. Ma, Hong Kong Shue Yan University, Hong Kong</i>	
Three Cases of Unconventional Educational Uses of Digital Storytelling	2616
<i>Emmanuel Fokides, University of the Aegean, Greece</i>	
3D Printing Applications in STEM Education	2626
<i>Norman Gwangwava, Botswana International University of Science and Technology, Botswana</i>	
<i>Catherine Hlahla, National University of Science and Technology, Zimbabwe</i>	
3D Scanning and Simulation of a Hybrid Refrigerator Using Photovoltaic Energy	1277
<i>Edith Obregón Morales, Centro de Investigación y Desarrollo Tecnológico en Electroquímica, Mexico</i>	
<i>José de Jesús Pérez Bueno, Centro de Investigación y Desarrollo Tecnológico en Electroquímica, Mexico</i>	
<i>Juan Carlos Moctezuma Esparza, Universidad Politécnica Metropolitana de Hidalgo, Mexico</i>	
<i>Diego Marroquín García, Universidad Tecnológica de San Juan del Río, Mexico</i>	
<i>Arturo Trejo Pérez, Universidad Tecnológica de San Juan del Río, Mexico</i>	
<i>Roberto Carlos Flores Romero, Universidad Tecnológica de San Juan del Río, Mexico</i>	
<i>Juan Manuel Olivares Ramírez, Universidad Tecnológica de San Juan del Río, Mexico</i>	
<i>Maria Luisa Mendoza López, Tecnológico Nacional de México, Instituto Tecnológico de Querétaro, Mexico</i>	
<i>Juan Carlos Solís Ulloa, Instituto Tecnológico Superior de Cintalapa, Mexico</i>	
<i>Yunny Meas Vong, Centro de Investigación y Desarrollo Tecnológico en Electroquímica, Mexico</i>	
<i>Víctor Hugo Rodríguez Obregón, Centro de Investigación y Desarrollo Tecnológico en Electroquímica, Mexico</i>	

A Three-Vector Approach to Blind Spots in Cybersecurity.....	1684
<i>Mika Westerlund, Carleton University, Canada</i>	
<i>Dan Craigen, Carleton University, Canada</i>	
<i>Tony Bailetti, Carleton University, Canada</i>	
<i>Uruemu Agwae, Carleton University, Canada</i>	
Throughput Dependence on SNR in IEEE802.11 WLAN Systems	6618
<i>Ikponmwosa Oghogho, Delta State University, Abraka-Oleh Campus, Nigeria</i>	
Tools, Pedagogical Models, and Best Practices for Digital Storytelling	2641
<i>Jari Multisilta, Tampere University of Technology, Finland</i>	
<i>Hannele Niemi, University of Helsinki, Finland</i>	
Toward a Working Definition of Digital Literacy.....	2326
<i>Margaret-Mary Sulentic Dowell, Louisiana State University, USA</i>	
Towards a General Theory of Information	4459
<i>Laura L. Pană, Polytechnic University of Bucharest, Romania</i>	
Towards an Interdisciplinary Socio-Technical Definition of Virtual Communities	4278
<i>Umar Ruhi, University of Ottawa, Canada</i>	
Towards an Understanding of Performance, Reliability, and Security	7588
<i>Ye Wang, Zhejiang Gongshang University, China</i>	
<i>Bo Jiang, Zhejiang Gongshang University, China</i>	
<i>Weifeng Pan, Zhejiang Gongshang University, China</i>	
Towards Modelling Effective Educational Games Using Multi-Domain Framework.....	3337
<i>Mifrah Ahmad, Universiti Teknologi PETRONAS, Malaysia</i>	
<i>Lukman Ab Rahim, Universiti Teknologi PETRONAS, Malaysia</i>	
<i>Kamisah Osman, Universiti Kebangsaan Malaysia, Malaysia</i>	
<i>Noreen Izza Arshad, Universiti Teknologi PETRONAS, Malaysia</i>	
Toward Trustworthy Web Services Coordination.....	8056
<i>Wenbing Zhao, Cleveland State University, USA</i>	
The Trajectivity of Virtual Worlds.....	4296
<i>Christophe Duret, Université de Sherbrooke, Canada</i>	
Transformational Leadership for Academic Libraries in Nigeria.....	5726
<i>Violet E. Ikolo, Delta State University Library, Nigeria</i>	
Transmedia and Transliteracy in Nemetical Analysis.....	6488
<i>Michael Josefowicz, Nemetics Institute Kolkata, USA</i>	
<i>Ray Gallon, The Transformation Society, France</i>	
<i>Maria Nieves Lorenzo Galés, The Transformation Society, Spain</i>	

The Trends and Challenges of 3D Printing.....	4382
<i>Edna Ho Chu Fang, University of Malaya, Malaysia</i>	
<i>Sameer Kumar, University of Malaya, Malaysia</i>	
Trends in Health Care Information Technology and Informatics	3805
<i>T. Ray Ruffin, University of Phoenix, USA & Colorado Technical University, USA & Grand Canyon University, USA & Ashford University, USA, & North Carolina Wesleyan College, USA</i>	
<i>Donna Patterson Hawkins, University of Phoenix, USA</i>	
Trust and Decision Making in Turing's Imitation Game	251
<i>Huma Shah, Coventry University, UK</i>	
<i>Kevin Warwick, Coventry University, UK</i>	
A Trust Case-Based Model Applied to Agents Collaboration.....	4797
<i>Felipe Boff, Lutheran University of Brazil (ULBRA), Brazil</i>	
<i>Fabiana Lorenzi, Lutheran University of Brazil (ULBRA), Brazil</i>	
Twitter Data Mining for Situational Awareness.....	2064
<i>Marco Vernier, University of Udine, Italy</i>	
<i>Manuela Farinosi, University of Udine, Italy</i>	
<i>Gian Luca Foresti, University of Udine, Italy</i>	
Uberization (or Uberification) of the Economy	2345
<i>Nabyla Daidj, Telecom Ecole de Management, France</i>	
Ubiquitous Computing, Contactless Points, and Distributed Stores.....	7805
<i>Marco Savastano, Sapienza University of Rome, Italy</i>	
<i>Eleonora Pantano, Middlesex University London, UK</i>	
<i>Saverino Verteramo, University of Calabria, Italy</i>	
Ubiquitous Teachers' Training and Lessons Learned with the uProf! Model	7671
<i>Sabrina Leone, Università Politecnica delle Marche, Italy</i>	
<i>Giovanni Biancofiore, giovannibiancofiore.com, Italy</i>	
Uncovering Limitations of E01 Self-Verifying Files	1384
<i>Jan Krasniewicz, Birmingham City University, UK</i>	
<i>Sharon A. Cox, Birmingham City University, UK</i>	
Understanding and Assessing Quality of Models and Modeling Languages	4810
<i>John Krogtie, Norwegian University of Science and Technology, Norway</i>	
Understanding Business Models on the Cloud	1141
<i>Arash Najmaei, Australian Catholic University, Australia</i>	

Understanding Cloud Computing in a Higher Education Context	1153
<i>Lucy Self, University of Sussex, UK</i>	
<i>Petros Chamakiotis, University of Sussex, UK</i>	
The Understanding of Spatial-Temporal Behaviors	1344
<i>Yu-Jin Zhang, Tsinghua University, China</i>	
Understanding the Potentials of Social Media in Collaborative Learning.....	7168
<i>Adem Karahoca, Bahcesehir University, Turkey</i>	
<i>Ilker Yengin, A*STAR, Institute of High Performance Computing, Singapore</i>	
Understanding User Experience	7599
<i>Camille Dickson-Deane, University of Melbourne, Australia</i>	
<i>Hsin-Liang (Oliver) Chen, University of Massachusetts Boston, USA</i>	
Uniform Random Number Generation With Jumping Facilities	1297
<i>E. Jack Chen, BASF Corporation, USA</i>	
The University-Industry Collaboration	3963
<i>Marcello Fernandes Chedid, University of Aveiro, Portugal</i>	
<i>Leonor Teixeira, University of Aveiro, Portugal</i>	
An Update on Bitcoin as a Digital Currency	2861
<i>Cecilia G. Manrique, University of Wisconsin – La Crosse, USA</i>	
<i>Gabriel G. Manrique, Winona State University, USA</i>	
Usability Evaluation of Tourism Icons in India.....	4099
<i>Rajshree Tushar Akolkar, Zeal College of Engineering and Research, India</i>	
<i>Ganesh D. Bhutkar, Vishwakarma Institute of Technology, India</i>	
Usability of CAPTCHA in Online Communities and Its Link to User Satisfaction.....	8066
<i>Samar I. Swaid, Philander Smith College, USA</i>	
Usable Security	5004
<i>Andrea Atzeni, Politecnico di Torino, Italy</i>	
<i>Shamal Faily, Bournemouth University, UK</i>	
<i>Ruggero Galloni, Square Reply S.r.l., Italy</i>	
Use of Bitcoin for Internet Trade	2869
<i>Sadia Khalil, NUST School of Electrical Engineering and Computer Science, Pakistan</i>	
<i>Rahat Masood, NUST School of Electrical Engineering and Computer Science, Pakistan</i>	
<i>Muhammad Awais Shibli, VisionIT, USA</i>	
Use of Data Analytics for Program Impact Evaluation and Enhancement of Faculty/Staff Development	1880
<i>Samuel Olugbenga King, Auburn University, USA</i>	

Use of GIS and Remote Sensing for Landslide Susceptibility Mapping	3503
<i>Arzu Erener, Kocaeli University, Turkey</i>	
<i>Gulcan Sarp, Suleyman Demirel University, Turkey</i>	
<i>Sebnem Duzgun, Middle East Technical University, Turkey</i>	
The Use of Podcasting/Vodcasting in Education	2651
<i>Athanasios T. Stavrianos, 2nd Technical Vocational School of Xanthi, Greece</i>	
<i>Apostolos Syropoulos, Greek Molecular Computing Group, Greece</i>	
Use of Technology in Problem-Based Learning in Health Science.....	5853
<i>Indu Singh, Griffith University, Australia</i>	
<i>Avinash Reddy Kundur, Griffith University, Australia</i>	
<i>Yun-Mi Nguy, Griffith University, Australia</i>	
User Resistance to Health Information Technology	3816
<i>Madison N. Ngafeeson, Northern Michigan University, USA</i>	
Users Behavioral Intention Towards eGovernment in an African Developing Country	3654
<i>Ayankunle A. Taiwo, Texas A&M University – Commerce, USA</i>	
Using Business Analytics in Franchise Organizations.....	930
<i>Ye-Sho Chen, Louisiana State University, USA</i>	
Using Communities of Inquiry Online to Perform Tasks of Higher Order Learning.....	3976
<i>Ramon Tirado-Morueta, University of Huelva, Spain</i>	
<i>Pablo Maraver-López, University of Huelva, Spain</i>	
<i>Ángel Hernando-Gómez, University of Huelva, Spain</i>	
Using Global Appearance Descriptors to Solve Topological Visual SLAM.....	6894
<i>Lorenzo Fernández Rojo, Miguel Hernandez University, Spain</i>	
<i>Luis Paya, Miguel Hernández University, Spain</i>	
<i>Francisco Amoros, Miguel Hernandez University, Spain</i>	
<i>Oscar Reinoso, Miguel Hernandez University, Spain</i>	
Using Receiver Operating Characteristic (ROC) Analysis to Evaluate Information -Based Decision-Making.....	2213
<i>Nan Hu, University of Utah, USA</i>	
Using RFID and Barcode Technologies to Improve Operations Efficiency Within the Supply Chain.....	5595
<i>Amber A. Smith-Ditizio, Texas Woman's University, USA</i>	
<i>Alan D. Smith, Robert Morris University, USA</i>	
Using Social Media to Increase the Recruitment of Clinical Research Participants	7181
<i>Saliha Akhtar, Seton Hall University, USA</i>	

Using Technology to Reduce a Healthcare Disparity	3725
<i>Nilmini Wickramasinghe, Epworth HealthCare, Australia & Deakin University, Australia</i>	
Utilizing Information Science and Technology in Franchise Organizations	4822
<i>Ye-Sho Chen, Louisiana State University, USA</i>	
A Validation Study of Rehabilitation Exercise Monitoring Using Kinect.....	5941
<i>Wenbing Zhao, Cleveland State University, USA</i>	
<i>Deborah D. Espy, Cleveland State University, USA</i>	
<i>Ann Reinthal, Cleveland State University, USA</i>	
Vertical Integration Between Providers With Possible Cloud Migration	1164
<i>Aleksandra Kostic-Ljubisavljevic, University of Belgrade, Serbia</i>	
<i>Branka Mikavica, University of Belgrade, Serbia</i>	
Video Considerations for the World Language edTPA.....	7682
<i>Elizabeth Goulette, Georgia State University, USA</i>	
<i>Pete Swanson, Georgia State University, USA</i>	
Viewpoints on Business Process Models.....	788
<i>Giorgio Bruno, Politecnico di Torino, Italy</i>	
Virtual Hoarding	4306
<i>Jo Ann Oravec, University of Wisconsin – Whitewater, USA</i>	
Virtualization as the Enabling Technology of Cloud Computing.....	1174
<i>Mohamed Fazil Mohamed Firdhous, University of Moratuwa, Sri Lanka</i>	
Virtual Reality as Distraction Technique for Pain Management in Children and Adolescents	5955
<i>Barbara Atzori, University of Florence, Italy</i>	
<i>Hunter G. Hoffman, University of Washington, USA</i>	
<i>Laura Vagnoli, Meyer Children’s Hospital of Florence, Italy</i>	
<i>Andrea Messeri, Meyer Children’s Hospital of Florence, Italy</i>	
<i>Rosapia Lauro Grotto, University of Florence, Italy</i>	
Virtual Tourism and Its Potential for Tourism Development in Sub-Saharan Africa.....	4113
<i>Paul Ankomah, North Carolina A&T State University, USA</i>	
<i>Trent Larson, North Carolina A&T University, USA</i>	
Virtual Worlds in the Educational Context	7935
<i>Felipe Becker Nunes, Federal University of Rio Grande do Sul, Brazil</i>	
<i>Fabício Herpich, Federal University of Rio Grande do Sul, Brazil</i>	
<i>Leo Natan Paschoal, University of Cruz Alta, Brazil</i>	

Visible Light Communication Numerous Applications.....	6672
<i>Ala' Fathi Khalifeh, German Jordan University, Jordan</i>	
<i>Hasan Farahneh, Ryerson University, Canada</i>	
<i>Christopher Mekhiel, Ryerson University, Canada</i>	
<i>Xavier Fernando, Ryerson University, Canada</i>	
Visual Identity Design for Responsive Web	8079
<i>Sunghyun Ryoo Kang, Iowa State University, USA</i>	
<i>Debra Satterfield, California State University – Long Beach, USA</i>	
Visualization as a Knowledge Transfer.....	5103
<i>Anna Ursyn, University of Northern Colorado, USA</i>	
The Vital Importance of Faculty Presence in an Online Learning Environment	2661
<i>Ni Chang, Indiana University – South Bend, USA</i>	
Vitalizing Ancient Cultures Mythological Storytelling in Metal Music	7338
<i>Uğur Kiling, Ondokuz Mayıs University, Turkey</i>	
Viterbi Decoder in Hardware	6307
<i>Mário Pereira Véstias, Instituto Politecnico de Lisboa, Portugal</i>	
Waste Gas End-of-Pipe Treatment Techniques in Italian IPPC Chemical Plants.....	3156
<i>Gaetano Battistella, ISPRA, Italy</i>	
<i>Giuseppe Di Marco, ISPRA, Italy</i>	
<i>Carlo Carlucci, ISPRA, Italy</i>	
<i>Raffaella Manuzzi, ISPRA, Italy</i>	
<i>Federica Bonaiuti, ISPRA, Italy</i>	
<i>Celine Ndong, ISPRA, Italy</i>	
Web Site Mobilization Techniques	8087
<i>John Christopher Sandvig, Western Washington University, USA</i>	
Web 2.0 From Evolution to Revolutionary Impact in Library and Information Centers.....	5262
<i>Zahid Ashraf Wani, University of Kashmir, India</i>	
<i>Tazeem Zainab, University of Kashmir, India</i>	
<i>Shabir Hussain, University of Kashmir, India</i>	
What Accounts for the Differences in Internet Diffusion Rates Around the World?.....	8095
<i>Ravi Nath, Creighton University, USA</i>	
<i>Vasudeva Murthy, Creighton University, USA</i>	
The What, How, and When of Formal Methods	7609
<i>Aristides Dasso, Universidad Nacional de San Luis, Argentina</i>	
<i>Ana Funes, Universidad Nacional de San Luis, Argentina</i>	

Why It Is Difficult to Disengage From Facebook	7190
<i>Sonda Bouattour Fakhfakh, University of Tunis El-Manar, Tunisia</i>	
Wireless Implant Communications Using the Human Body	6319
<i>Assefa K. Teshome, Victoria University, Australia</i>	
<i>Behailu Kibret, Victoria University, Australia</i>	
<i>Daniel T. H. Lai, Victoria University, Australia</i>	
Women and IT in Lilongwe	3393
<i>Alice Violet Nyamundundu, Skyway University, Malawi</i>	
Word Formation Study in Developing Naming Guidelines in the Translation of English Medical Terms Into Persian	5136
<i>Ali Akbar Zeinali, Universiti Sains Malaysia, Malaysia</i>	
Young People, Civic Participation, and the Internet.....	3667
<i>Fadi Hirzalla, Erasmus University Rotterdam, The Netherlands</i>	
<i>Shakuntala Banaji, LSE, UK</i>	

Preface

Influencing every facet of business, society, and life worldwide, with speed beyond imagination, the field of information science and technology has without a doubt brought upon a revolution in the way the human population interacts, does business, and governs. As one takes into account the leaps and bounds experienced in information sharing and communication exchange over the last few decades, a truly admirable phenomenon presents itself and clearly shows that the results of this pivotal rising will monumentally impact the way the world thinks, subsists, and evolves.

With a long history of expeditious evolution, the growth and expansion of information technology began during the early 1950s with the main purpose of initiating scientific computing, expanding research, and utilizing the power of computers as a means to support a mass volume of computational tasks in scientific applications and discoveries. Later, during the 1960s and '70s, the use of computer technology was extended to business applications, mostly in accounting and financial areas that involved processing numbers and collecting data in a quantitative sense. As a result, the use of this technology was limited to those who had an expansive knowledge of these systems and had access to computer programming languages. With the evolution of computers and telecommunications in the 1980s, a new information technology was born with a strong focus on the management and dissemination of information by both information providers and users across the globe.

In the early 1990s, the most noticeable advancement in the information technology revolution was the creation of the Internet. During the past two decades, Internet technologies have become the driving force in allowing people worldwide to communicate and exchange information, creating a new virtual, interactive dimension and providing a digital forum for global social connection. In recent years, through the use of Web-enabled technologies, organizations of all types and sizes around the world have managed to utilize these technologies to disseminate and process information with prospective customers, suppliers, students, and governments. Today, the ability to communicate and connect from many locations through personal computers has influenced different people in many different societies. These technologies allow everyone, regardless of their geographic location, to bring the information age to its full realization.

In recent years, the science of understanding the nature of information processing and management, along with the computers and technologies that decipher, disseminate, and manage information, has become known as information science and technology. Technology has profoundly impacted science, business, and society, thus constructing an entity that improves access to the rapidly expanding body of knowledge in almost every discipline. Society fuels this knowledge creation, as it receives, manages, educates, and collects information. The volume and intensity of research in information science and technology have exceeded many other fields of science, and research discoveries have become the impetus behind many emerging tools and applications seen at every organizational level.

Preface

In addressing this need for the representation of evolving information science and technology disciplines in academic literature, the first edition of the *Encyclopedia of Information Science and Technology*, released in early 2005, positioned itself as the first of its kind in reference publications, offering an invaluable source highlighting major breakthroughs, discoveries, and authoritative research results in technological advancements. In providing this compendium of references, definitions, and keywords within this field of pivotal social and organizational movement, the five-volume *Encyclopedia of Information Science and Technology* (First Edition) supplied researchers with a definitive one-stop reference source.

In late 2008, the *Encyclopedia of Information Science and Technology* (Second Edition) followed the first edition with an eight-volume compendium of updated research on crucial topics previously covered in the first edition, as well as an advanced treatment of new developments, technologies, and areas of research in the field of information and communications science. With an expanded number of contributions, the second edition surpassed the first in terms of timeliness, comprehensiveness, and critical acclaim, lauded as an essential reference for any academic library.

With the endeavor of continuing to exhibit the latest research innovations and advances, the third edition of the *Encyclopedia of Information Science and Technology* was comprised of 10 volumes of all new content and uncovered the most current research findings related to technological, organizational, and managerial issues, challenges, trends, and applications of information technologies in modern organizations. The coverage in the third edition of the encyclopedia bridged existing gaps in available references on technology and methodologies with its contribution as a valuable resource of encompassing paradigms shaping the ever-changing research, theory, and discovery of information science and technology.

This brings us to the present, and three years later we are again releasing new materials providing the latest trends and research in this ever-changing field. The selected topics of this encyclopedia provide a balanced representation of concepts and issues from researchers around the world. These researchers were asked to submit proposals describing the topic and scope of their articles. All proposals were carefully reviewed for suitability by the Editor-in-Chief. Upon the receipt of full article submissions, each contribution was forwarded to at least three expert external reviewers on a double-blind peer review basis. Only submissions with strong and favorable reviews were chosen as articles for this edition of the *Encyclopedia of Information Science and Technology*. In some cases, submissions were sent back for several revisions prior to final acceptance. The goal was to assemble the best minds in the information science and technology field from all over the world to contribute articles to this encyclopedia and to apply the highest level of quality in selecting articles for inclusion.

As a result, over 700 new articles were carefully selected for inclusion in this 10-volume encyclopedia based on their presentation of the most comprehensive, innovative, and in-depth coverage of the current concepts, issues, and emerging technologies in the field of information science and technology. The articles included in this edition of the encyclopedia are written by more than 1,400 distinguished scholars and researchers from hundreds of prominent research institutions all over the world. Over 5,000 technical and managerial terms and their definitions have been organized by the authors to enhance the articles, allowing for extensive research into core concepts and ideas. In addition, this 10-volume set offers a thorough reference section with approximately 15,000 sources of additional information for scholars, students, and researchers in the field of information science and technology. Multiple tables of content have been organized to better assist readers in navigating and identifying information. Contents are structured through alphabetical and categorical listings for easy reference.

The topics covered in this all-encompassing publication include some of the most influential areas in the field. One such example is the area of *artificial intelligence* and its current application within organizational spheres: specifically, technologies such as machine learning, computational intelligence, digital ecosystems, neural networks, and adaptive systems. Content coverage such as this will supply audiences with reputable sources for identifying future trends and breakthrough technologies that will directly impact everyday aspects of life, with specific examples being the latest computing and technological advancements in smart homes, as well as the integration of intelligent diagnostics and speech recognition technology in a variety of settings.

A substantial portion of this encyclopedia is dedicated to the latest advancements in business research, management, and technologies. Through the review of *business organizational research* and *business information systems* and their ever-changing state, this encyclopedia analyzes the driving force of globalization, and the effect of these systems on international trade, economics, and capital. Because business information systems are so far-reaching, from the intranet to export tracking and manufacturing intelligence, the research results presented within these articles examine a breadth of review and discussion of methodologies behind these systems and how they are implemented in the field of business.

Cybercrime, cyber bullying, and digital terrorism, a growing discipline fueled by daily reminders in news outlets around the world, is covered in several articles within this encyclopedia. Special attention has been paid to how behavior across all cultures affects all participants, with good and bad effects in our digitally wired world.

Continuously utilized in the evolution of technical applications, the areas of *data mining and databases* are comprehensively covered in a significant number of articles in this edition of the encyclopedia for both the purpose of introduction and advancement. Taking into consideration the growing use of data mining and database management in a number of activities such as Web development and engineering progress, these resources supply readers with authoritative results and a foundation for additional research.

With a considerable effect on the global economy, *electronic business* applications are discussed in this edition of the encyclopedia and offer researchers a credible source of knowledge for understanding the current realms of mobile applications for business functions, social commerce, and virtual enterprises. Seeing the importance of these technology-based management systems, modern electronic business is analyzed as a growing phenomenon of capitalism.

As the world becomes increasingly engaged in online shopping and trade, *electronic commerce* is fully presented in this edition of the encyclopedia as a phenomenon of progressive change. With significant growth since the dot-com explosion of the past two decades, e-commerce is now considered a discipline with significant global implications, particularly on international trade and how goods and currencies are exchanged and services are dealt.

As a contemporary social motivator, *educational technologies* are comprehensively and extensively examined in this edition of the encyclopedia. This is an area of research with philosophical and constructivist reach into all levels of learning. In discussing blended learning, distance education, tools for online learning, advanced pedagogy, virtual learning environments, and computational support for teaching, educational technologies demand further attention in developing reliable research. As a popular area of examination, this encyclopedia lends support to emerging trends in the discipline while supplying a venue for further discussion of terms, themes, and additional implications.

Preface

A new category on *gaming* makes its debut in this fourth edition, and reflects on how pervasive the gaming industry has become, no longer just for entertainment but a very real part of learning. Cognitive effects, theoretical models, game-based designs, and modeling are discussed in this section of the encyclopedia.

Given the worldwide focus on the environment, this edition of the encyclopedia provides tremendous coverage of emerging technologies and applications in the areas of *environmental science and agriculture*. As a growing academic research area, this discipline is supported by many research papers written by prominent international researchers studying the future of maintaining environmental functions and safeguarding the planet.

Global information technology has emerged as an area of study with growing importance as digital communications spread internationally. This edition of the encyclopedia supplies articles that delve into the importance of the global economy, as well as exhibiting regional adaptation, resistance, and adoption of technologies.

Patient monitoring systems, e-health, and cybertherapy are three of the many topics extensively covered in this edition of the encyclopedia by providing a wealth of knowledge related to *health information systems* and their use and implications in modern societies. With new assistive and rehabilitative technologies, medical data storage, and issues of security and privacy, health information systems affect the lives of human beings worldwide. Considering this growth, research results prove invaluable for healthcare administrators and academic disciplines, such as nursing and health management.

As technological demands extend and users multiply, *high-performance computing* seeks to maintain the flow of communication between servers, knowledge workers, and organizations. With much recent advancement in cloud computing and technologies to handle large amounts of data, this edition of the encyclopedia offers researchers a comprehensive convergence of all topics related to cloud computing applications and management.

Considering the human element in making technology the paradigm it is today, this edition of the encyclopedia provides comprehensive coverage of human aspects of technology. With an analysis of end-user behavior, gender differences, and ubiquity of computing, readers will find an extensive amount of research results and analysis to assist in building the literature in an important area of study: *human-computer interaction*.

Following the advancements of the Industrial Revolution, *industrial informatics* is described in this encyclopedia as a bridge between technological progress and the application of manufacturing, transportation, and construction. With growing utilization in enhancing and expediting good production, building, and usage, research results examine the future trends of such technologies and how these applications will incorporate into the daily lives of individuals around the globe.

Taking into account the massive growth in technological utilization for sensitive data, both in personal and organizational functions, *IT security and ethics* as an area of study continues to evolve, expanding securitization and streamlining for the best protection of digital information. In examining such areas as digital forensics, authentication, cryptography, cyber warfare, and trustworthy computing, analysis of IT security and ethics takes prominence as a point of key discussion in this edition of the encyclopedia.

As the result of the current structures of both public and private sector administration, the field of *knowledge management* has been positioned as a concept of both utility and application. Considering the need for quality sources indicating best practices in the management of knowledge workers, IT governance, and information sharing, this encyclopedia supplies readers and researchers with a considerable selection of current themes, terms, and topics relating to the current and future state of knowledge management.

With many library systems implementing digital filing and cataloguing systems, the study of bibliometrics has flourished in the scientific world. With cutting-edge technologies in archiving and classification, preservation, and reformatting, digital *library science and administration* has developed into an intricate branch of information science, and this encyclopedia provides a tremendous coverage of the vast spectrum of this study. As issues such as open access, security, and Internet property rights take center stage, legal aspects of the digitization of information are explored extensively.

Advanced *medical technologies* are incorporated into daily measures to secure and save lives. Medical informatics is explored in this fourth edition encyclopedia through an intricate look at the updates and improvements in this pivotal field. Medical imaging, biosensors, and new nanotechnology modernizations for surgical procedures are a few of the exciting and significant advances in medical technologies. Rehabilitation, disease detection, and mobile medical care are assisted and revamped with advances in hospital machinery and tools.

Mobile devices, wireless systems, sensors, and wearable computing applications are technologies that are shaping communication and public administration. Keeping in mind the spread of the need for reliable means of information transfer, *mobile and wireless computing* is explored in this edition of the encyclopedia as a field of widening pervasiveness. With uses ranging from machine correspondence to business interrelationships, important research findings are effectively exposed and discussed by top researchers in the field.

Whether utilized for entertainment, learning, or public policy, *multimedia technology* as a form of study involves behavioral and practical analysis. This edition of the encyclopedia offers readers a wealth of research coverage addressing many aspects of analysis on topics such as simulation, digital imaging, and hypermedia to best understand how the world is displaying its information for consumers and administrators.

Social networking and computing describes the intersection of social behaviors and computer systems, intricately examining areas such as social networking sites, augmented realities, and online auctions, and their substantial impact on society. Tools such as blogs, Wikis, tags, and podcasts have expanded the user's online researching experience. This edition of the encyclopedia expounds on social informatics and explores emerging technologies and applications such as instant messaging, virtual groups, mailing lists, and forums: just a few of the places that users interact with one another.

For several decades, technological engineers have focused on algorithms, modeling languages, and kernel applications in designing and formatting the machines and automation that the world uses today. Innovative *systems and software engineering* technologies now play a crucial role in the future design and improvement of the ever-changing world of automation. This edition of the encyclopedia provides ample research coverage of the emerging technologies in this area and their vulnerabilities, specifications, quality, and architectures.

As a result of the Internet explosion of the 1990s, *Web technologies* as a field of study has taken the stage as a discipline of ubiquitous importance. With the Web being a single source of infinite information and computing, responding and expanding at exponential rates, this encyclopedia offers researchers a tremendous coverage of emerging innovations in disciplines such as portal technologies, semantic computing, Web 2.0, and service-oriented technologies.

Preface

In closing, the diverse and comprehensive coverage of multiple disciplines of information science and technology in this 10-volume, authoritative encyclopedia are sure to contribute to a better understanding of all topics, research, and discoveries in this evolving field. Furthermore, the contributions included in this publication will be instrumental in the expansion of knowledge in this field. This publication will inspire its readers to further contribute to the current discoveries in this immense field, creating possibilities for further research and discovery into the future of information science and technology and what lies ahead for the knowledge society.

Mehdi Khosrow-Pour, D.B.A.

Editor-in-Chief

Encyclopedia of Information Science and Technology, Fourth Edition

Guide to the Encyclopedia of Information Science and Technology, Fourth Edition

ORGANIZATION

The *Encyclopedia of Information Science and Technology, Fourth Edition* is a 10-volume set comprised of over 700 articles. All articles are divided into categories relevant to their topical coverage. There are 86 different category sections, with each volume containing multiple categories. All category sections are arranged alphabetically across the 10 volumes, beginning with “A” categories and ending with “W” categories. Within each category section, the articles are also arranged in alphabetical order. As each new category is introduced, section dividers represent the transition from one category to the next. Also, in the print version of the encyclopedia, letters for the categories represented in each volume are clearly marked on the side binding of the cover.

To assist with easy navigation, there are two different tables of content compiled at the beginning of each volume. The first represents the “Contents by Volume,” which displays the arrangement of the content in its respective categories, and the second represents the “Contents in Alphabetical Order,” which displays the arrangement of content from A to Z by the articles’ titles.

EACH VOLUME CONTAINS

- A **preface** and **user’s guide**.
- The **Editor-in-Chief’s biography** and **acknowledgment**.
- **Two tables of content** are compiled at the beginning of each volume: “Contents by Volume” and “Contents in Alphabetical Order.”
- **Several authoritative, research-based articles** contributed by thousands of researchers and experts from all over the world.
- A **comprehensive index** supporting the extensive system of cross-references.

EACH ARTICLE INCLUDES

- A brief **introduction to the topic area** describing the general perspective and objectives of the article.
- A **background** providing the broad definitions and discussions of the topic and incorporating the views of others (i.e., a literature review) into the discussion to support, refute, or demonstrate the author's position on the topic.
- Various perspectives examining the **issues, controversies, and problems** as they relate to the theme. Also provided are arguments supporting the position as well as a comparison and contrast with regards to what has been and/or is currently being done as it relates to the article's specific topic and the overall theme of the encyclopedia.
- A discussion of **solutions and recommendations** in dealing with the issues, controversies, or problems presented in the preceding section.
- **Charts, graphs, tables, and formulae** are included as illustrative examples whenever appropriate.
- A discussion of **future research directions**.
- A **conclusion** to discuss the overall coverage of the article and present concluding remarks.
- An extensive **list of references** so that readers can benefit from the sources cited within the text.
- An **additional readings section** consisting of sources that complement the topical coverage within the article.
- A **key terms and definitions section** providing 7-10 terms related to the topic of the article with a clear and concise definition for each term.

Acknowledgment

Editing and completing an authoritative and comprehensive scholarly research publication such as the *Encyclopedia of Information Science and Technology, Fourth Edition* requires tremendous contributions, and a great deal of assistance from large groups of scholars and staff. The primary objective of this encyclopedia is to provide the most up-to-date scholarly coverage of all topics related to information science and technology as it is applied to several discipline areas, such as business, engineering, medicine, education, public administration, computer science, as well as the social sciences and humanities. I am indebted to all the authors for their excellent contributions to this edition of the encyclopedia.

All submitted manuscripts to this 10-volume edition underwent a double-blind peer review process in order to achieve the highest level of quality and accuracy. I am thankful to all the reviewers of this edition for providing their expertise and their rigorous, unbiased assessment of the manuscripts assigned to them on a double-blind basis, as well as the members of the Editorial Advisory Board for their wisdom, guidance, and assistance with various decisions throughout the editorial process.

I would also like to convey my deep appreciation and gratitude to Jan Travers, Director of Intellectual Property and Contracts, for all her tireless efforts assisting me with bringing a publication of this size to fruition, as well as to the Acquisitions, Development, Copy Editing, and Production Divisions of the Editorial Content Department at IGI Global for their valuable assistance in support of this project, especially Chris Shearer, Copy Editing Manager, Christina Henning, Production Editor, Courtney Tychinski, Assistant Managing Editor (Book Development), Mariah Gilbert, Assistant Managing Editor (Acquisitions), and also to Lindsay Wertman, IGI Global's Managing Director. Additionally, I would like to thank the IGI Global Sales and Marketing Department for their endless support in promoting this invaluable reference source.

Thank you to everyone who has provided me immeasurable amounts of knowledge, wisdom, and patience over the last 30 years.

Mehdi Khosrow-Pour, D.B.A.

Editor-in-Chief

Encyclopedia of Information Science and Technology, Fourth Edition

About the Editor

Mehdi Khosrow-Pour, D.B.A., received his Doctorate in Business Administration from the Nova Southeastern University (Florida, USA). Dr. Khosrow-Pour taught undergraduate and graduate information system courses at the Pennsylvania State University – Harrisburg for almost 20 years. He is currently Executive Editor at IGI Global (www.igi-global.com). He also serves as Executive Director of the Information Resources Management Association (IRMA) (www.irma-international.org) and Executive Director of the World Forgotten Children's Foundation (www.world-forgotten-children.org). He is the author/editor of more than 100 books in information technology management. He is also currently the Editor-in-Chief of the *International Journal of Green Computing*, *International Journal of Library and Information Services*, *International Journal of E-Entrepreneurship and Innovation*, and *International Journal of Natural Computing Research*, and is also the founding Editor-in-Chief of the *Information Resources Management Journal*, *Journal of Electronic Commerce in Organizations*, *Journal of Cases on Information Technology*, and the *Journal of Information Technology Research*, and has authored more than 50 articles published in various conference proceedings and scholarly journals.

Category A

Accounting and Finance

Applying Artificial Intelligence to Financial Investing

A**Hayden Wimmer***Georgia Southern University, USA***Roy Rada***University of Maryland – Baltimore County, USA*

INTRODUCTION

Artificial intelligence (AI) techniques have long been applied to financial investing scenarios to determine market inefficiencies, criteria for credit scoring, and bankruptcy prediction, to name a few. While there are many subfields to artificial intelligence this work seeks to identify the most commonly applied AI techniques to financial investing as appears in academic literature. Techniques identified in this work include fuzzy systems, swarm intelligence, case-based reasoning, hybrid systems, genetic algorithms, neural networks, and machine learning. AI techniques, such as knowledge-based, machine learning, and natural language processing, are integrated into systems that simultaneously address data identification, asset valuation, and risk management. Frequently, machine learning is applied to technical financial indicators in order to make predictions about the direction of stock prices. Financial investing requires data identification, asset valuation, and risk management. One such example of applying AI techniques to financial investing is the application of knowledge-based techniques for credit risk assessment and machine learning techniques for stock valuation. Future trends will continue to integrate hybrid artificial intelligence techniques into financial investing, portfolio optimization, and risk management. The remainder of this article summarizes key contributions of applying AI to financial investing as appears in the academic literature.

BACKGROUND

What Is Artificial Intelligence?

In the early days of computing, a typical task for a computer program was a numerical computation, such as computing the trajectory of a bullet. In modern days, a typical task for a computer program may involve supporting many people in important decisions backed by a massive database across a global network. As the tasks that computers typically perform have become more complex and more closely intertwined with the daily decisions of people, the behavior of the computer programs increasingly assumes characteristics that people associate with intelligence. When exactly a program earns the label of ‘artificial intelligence’ is unclear. The classic test for whether a program is intelligent is that a person would not be able to distinguish a response from an intelligent program from the response of a person. This famous Turing Test is dependent on factors not easily standardized, such as what person is making the assessment under what conditions.

A range of computer programming techniques that are currently, popularly considered artificial intelligence techniques includes (Rada, 2008):

- Knowledge-based techniques, such as in expert systems.
- Machine learning techniques, such as genetic algorithms and neural networks.

- Sensory or motor techniques, such as natural language processing and image processing.

These methods may apply to investing. For instance, expert systems have been used to predict whether a company will go bankrupt. Neural networks have been used to generate buy and sell decisions on stock exchange indices. Natural language processing programs have been used to analyze corporate news releases and to suggest a buy or sell signal for the corporate stock.

While artificial intelligence (AI) could apply to many areas of investing, much of what happens in computer-supported investing comes from non-AI areas. For instance, computational techniques not considered primarily AI techniques include numerical analyses, operations research, and probabilistic analyses. These non-AI techniques are routinely used in investing.

Investing and Data

The process of investing has 3-stages of:

1. Data Identification,
2. Asset Valuation, and
3. Risk Management.

AI has been most often applied to asset valuation but is also applicable to data identification and risk management.

Two, high-level types of data used in financial investing are technical data and fundamental data. The price of an asset across time is technical data and lends itself to various computations, such as the moving average or the standard deviation (volatility). Fundamental data should support cause-and-effect relationships between an asset and its price. For instance, the quality of management of a company should influence the profitability of a company and thus the price of its stock.

The universe of fundamental data is infinite. Many streams of data that might be relevant, such as corporate earnings or corporate debt, might also

be related to one another. Various non-AI tools, such as linear regression analysis and principal components analysis, might be used in identifying what sets of data are more likely to be useful than what other sets. Such non-AI, computational techniques can be combined with AI techniques in experimenting with various combinations of data and choosing what data to use in asset valuation.

ARTIFICIAL INTELLIGENCE APPLIED TO FINANCIAL INVESTING

AI Trends

A multi-agent architecture for an integrated system that considers data identification, asset valuation, and risk management has been proposed by researchers at Carnegie Mellon University. The system is called WARREN which refers to the first name of the famous investor Warren Buffet (Sycara, Decker, Pannu, Williamson, & Zeng, 1996). The WARREN system design includes components for collecting large amounts of real-time data, both numeric and textual. The data would be pre-processed and then fed to various asset valuation agents that would, in turn, feed their assessments to a portfolio management agent. The portfolio management agent would interact with clients of WARREN. Systems with various features of WARREN are available from commercial vendors and are developed in-house by large investing companies, but more research is needed on how to develop integrated, AI systems that support investing.

Natural language processing systems may include large bodies of domain knowledge and parse free text so as to make inferences about the content of the text. However, such natural language processing systems do not seem as popular in investing applications as much simpler natural language processing techniques. The natural language processing work that has been applied to investing seems to be largely of the sort in which the distribution of word frequencies in a

document is used to characterize the document. In this word-frequency way, Thomas (2003) has shown a potential value to processing news stories to help anticipate stock price changes.

As one can see cycles in the value of financial assets, one can also see cycles in the frequency of publication of articles on certain topics. In the field of artificial intelligence, one might identify, roughly speaking, three phases as follows (Rada, 2008):

1. Machine learning, in what was then called perceptron and self-organizing systems research, was popular from 1955 to 1975,
2. Knowledge-based, multi-agent, or expert systems work was popular from 1975 to 1995, and
3. Machine learning research, now called neural networks or genetic algorithms research, returned to dominate the AI research scene from 1995 to 2013.

When AI research has been applied to investing, the AI technique used has tended to be the technique popular at the time. This leaves unaddressed the question of whether investing is more appropriately addressed with one AI technique or another.

The recent literature is rich with neural network applications to investing, but a new trend is the combining of knowledge-based techniques with neural network and genetic algorithm techniques. For instance, Tsakonas, Dounias, Doumpos, and Zopounidis (2006) use 'logic' neural nets that can be directly understood by people (traditional neural nets are a 'black box' to humans). Genetic programming modifies the architecture of the logic neural net by adding or deleting nodes of the network in a way that preserves the meaning of the neural net to people and to the net itself. Bhattacharyya, Pictet, and Zumbach (2002) have added knowledge-rich constraints to the genetic operators in their application for investing in foreign exchange markets.

A promising research direction is to combine the earlier knowledge-based work on financial accounting with the more recent work on machine learning for stock valuation. For instance, neural logic nets could represent some of the cause-effect knowledge from a bankruptcy system and become part of a learning system for predicting stock prices. Some of the bankruptcy variables are readily available online, such as a company's debt, cash flow, and capital assets.

The financial markets are human markets that evolve over time as opportunities to make profits in this zero-sum game depend on the changing strategies of the opponent. Thus, among other things, what is important in the input may change over time. An AI system should be able to evolve its data selection, asset valuation, and portfolio management components. The future direction for AI in investing is to integrate the three major tools of AI (knowledge-based systems, machine learning, and natural language processing) into a system that simultaneously handles the three stages of investing (data collection, asset valuation, and portfolio management). Such systems will interact with humans so that humans can specify their preferences and make difficult decisions, but in some arenas, such as program trading, these sophisticated AI systems could compete with one another.

Advances in AI and Finance 2009-2013

Artificial intelligence techniques continue to be applied to areas in finance in order to extend human capabilities by surveying large and distributed data sources (Phillips-Wren, 2012). Artificial neural networks (ANNs) continue to be the most popular approach to combining artificial intelligence with finance (Bahrammirzaee, 2010). ANNs have been applied to a variety of issues in the finance domain. One such example is applying ANNs to auditor selection where the process of selecting and appointing an auditing firm is a complex

process which can benefit from the ANN approach (Kirkos, Spathis, & Manolopoulos, 2010). Employing ANNs was found to outperform the standard approach of logistic regression. ANNs are suited to modeling non-linear as well as linear relationships and are well suited to applications in financial, time-series forecasting and can be tailored by customizing the input and hidden neurons of the ANN (Tarsauliya, Kant, & Kala, 2010). Currency exchange rates are another financial time-series where an ANN approach can be effective (Oyewale, 2013).

In addition to ANN techniques, techniques such as nearest neighbor (kNN) have been explored. One such technique combines a kNN approach with technical analysis techniques for stock trading where the method was more profitable than the standard buy and hold technique (Teixeira & de Oliveira, 2010). Support Vector Machines (SVMs) are supervised learning models which are used for data analysis and pattern recognition. SVMs have been applied to a range of finance research with one such being financial distress prediction (Bae, 2012). Besides SVM, fuzzy approaches, or approaches which deal with approximations, have begun to emerge in portfolio selection (Magoč & Modave, 2011).

Evolutionary approaches have been studied, as in (Chi & Hsu, 2012), where a genetic algorithm (GA) was employed in variable selection for determining credit scoring of business firms. The GA was used to augment a bank's internal scoring model and a credit bureau scoring model in order to improve the performance of credit risk management of mortgage accounts. Dividend policy forecasting was also shown to benefit from the introduction of evolutionary computation methods where the introduction of a GA was able to outperform the average prediction accuracy of standard machine learning decision tree classifiers such as CHAID, CART, and C4.5 (Won, Kim, & Bae, 2012). Genetic Algorithms applied to technical trading indicators improve the ability to predict major losses (Kaucic, 2010). Following the theme of technical analysis, Fu, Chung, and

Chung (2013) proposed a 2 step process to technical analysis by first applying a GA to determine the optimal technical indicators and then again to weight the investments of a financial portfolio. The method lowered risk, as measured by the Sharpe ratio, and outperformed the buy and hold method over different market sectors in bullish and bearish market conditions.

While artificial neural networks remain the most popular research stream for applying artificial intelligence techniques to finance (Bahrammirzaee, 2010), hybrid AI approaches are emerging as a viable, however complex, alternative. For example, decision trees have been shown to be effective when combined with logistic regression when applied to financial distress prediction (Chen, 2011). The hybrid approach trend continues by combining decision trees and fuzzy approaches. The high dimensionality and non-stationary variations in stock price make financial trend discovery difficult; therefore, an approach which uses fuzzy logic combined with a decision tree classifier has been shown to be effective when applied to the S&P 500 (Chang, Fan, & Lin, 2011). Genetic algorithms show promise when paired with fuzzy techniques and applied to portfolio selection (Bermúdez, Segura, & Vercher, 2012). While Fuzzy approaches are viable in portfolio selection, extending fuzzy approaches by incorporating swarm convergence techniques provides satisfactory performance when compared with other techniques (Ladyzynski & Grzegorzewski, 2013).

Compared with other independent methods, ANN approaches have demonstrated higher performance as evidenced by (Rafiei, Manzari, & Bostanian, 2011). This makes combining ANN with other AI approaches a natural extension to the research. Combining ANN with Case-Based Reasoning (CBR) was explored by (Chuang & Huang, 2011) and compared to ANN alone and shown to improve accuracy and reduce misclassification errors. Hybrid approaches that pair ANN with Fuzzy systems have also emerged as an extension to ANN alone. One such approach integrates ANN with genetic fuzzy systems and

applies the technique to stock price forecasting (Hadavandi, Shavandi, & Ghanbari, 2010). The method was applied to a subset of stocks in the IT and airline sector and was demonstrated to outperform previous methods of stock price prediction. A similar method of combining ANN and fuzzy systems was examined in (Boyacioglu & Avci, 2010) where the authors concluded the approach of combining ANN and fuzzy systems, namely fuzzy logic, is applicable to emerging markets' stock price prediction. In continuing with stock price prediction, K.-L. Shen (2013) combined ANN and a fuzzy approach to create a Fuzzy-ANN model which was shown to outperform two major market indices from Taiwan from 2008-2011.

LITERATURE TREND 2013-2015

In order to examine trends in the literature, a Google Scholar search on "Artificial Intelligence and Investing" and "Artificial Intelligence and Financial Investing" was performed. The search spanned January 1, 2013 until July 1, 2015. In 2013, the dominant categories of artificial intelligence and financial investing literature can be classified as fuzzy approaches, neural networks, hybrid approaches, and other methods. In 2014, hybrid methods increase in popularity. Hybrid systems are systems which utilize a combination of two or more techniques and/or methods. This increase in hybrid system interest continues into 2015 which concludes this analysis. Machine learning and neural networks are two of the most common methods found in hybrid systems thereby

Table 1. Citation count from google scholar for "artificial intelligence and investing"

Year	Citation Count
2013	1930
2014	1980
2015	1330
2013-2015	5230

Table 2. Citation count from google scholar for "artificial intelligence and financial investing"

Year	Citation Count
2013	7
2014	12
2015	12
2013-2015	26

Table 3. Google scholar sub-field citation count

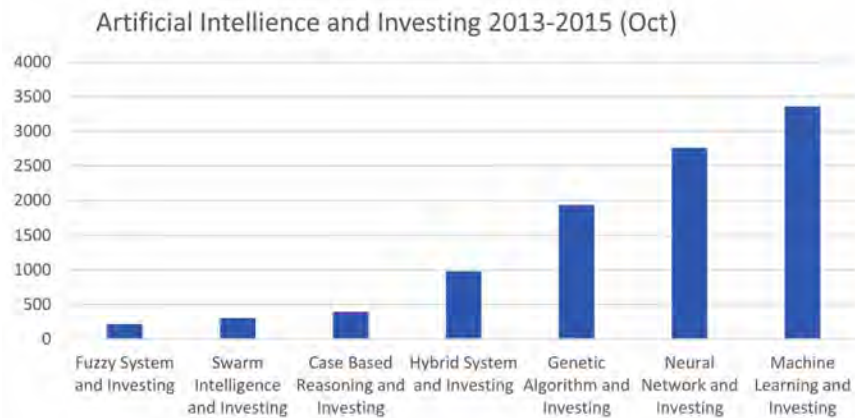
AI Subfield Citation Count 2013-2015 (Oct)	
Topic	Citation Count
Fuzzy System and Investing	213
Swarm Intelligence and Investing	304
Case Based Reasoning and Investing	394
Hybrid System and Investing	982
Genetic Algorithm and Investing	1930
Neural Network and Investing	2760
Machine Learning and Investing	3360

explaining the high citation count. Most hybrid systems do not specifically mention the term "hybrid" in the article title thereby explaining the lower number of hybrid citations.

Artificial Intelligence and Financial Investing (2013)

Fuzzy approaches were well represented. For example, Svalina, Galzina, Lujčić, and Šimunović (2013) generate a prediction model based on an adaptive neuro-fuzzy inference system to detect the close price of the Croatian Zagreb Stock Exchange. Yunusoglu and Selim (2013) propose a new fuzzy rule expert system that focusses on helping investors with investment decisions. Messaoudi and Rebai (2013) develop a new fuzzy goal programming model, which takes into consideration stochastic and fuzzy uncertainty for real life decision making problems for portfolio selection. Hassan, Ramamohanarao, Kamruzzaman, Rahman, and Hossain (2013) introduce a new adaptive inference system, which generates new fuzzy rules for forecasting

Figure 1. Citation counts of “artificial intelligence and investing” google scholar search from 2013-2015



the nonlinear time series data in stock market which performs better than the standard FIS and the performance of this system indicates that it can predict a number of stock indices including Dow Jones Industrial index, and NASDAQ index. Gradojevic and Gençay (2013) address the trading uncertainty problem by employing fuzzy logic to reduce trading uncertainty.

Neural networks remained a well-represented area of research. de Oliveira, Nobre, and Zárate (2013) use a neural network approach to predict the direction of change in stock price. Korol (2013) apply neural networks to bankruptcy prediction. Birău, Ehsanifar, and Mohammadi (2013) employ neural networks for predicting emerging stock prices. Devadoss and Ligorì (2013) show that neural networks can be an effective tool for predicting closing stock prices. Liu, Dang, and Huang (2013) propose a recurrent neural network with convergence for real time portfolio optimization for portfolio optimization. Özkan (2013) focuses on improving the foreign exchange rate forecast accuracy by using neural networks for forecasting the foreign exchange rates. Dong, Fataliyev, and Wang (2013) utilize Neural Network techniques for forecasting the stock prices. Kumar and Murugan (2013) emphasize the importance of neural networks for stock indices and conducts experiments to analyze the forecasting accuracy. Al-Jumeily, Hussain, and Alaskar (2013) propose

a new technique for predicting change in the large financial data sets used for forecasting the noisy time series data using a Dynamic Self-Organized Multilayer Neural Network based on the Immune algorithm.

Hybrid systems continued to gain momentum as an area of interest. Sermpinis, Theofilatos, Karathanasopoulos, Georgopoulos, and Dunis (2013) develop a hybrid approach pairing neural network with particle swarm optimization to predict foreign exchange rates. Evans, Pappas, and Xhafa (2013) create a hybrid system for genetic algorithms and artificial neural networks applied to intra-day market prediction. Manahov and Hudson (2013) combine agent based systems with genetic programming to evaluate the price of stocks represented by the Dow Jones Industrial Average, General Electric, and IBM. Nhu, Nitsuwat, and Sodanil (2013) combine the Firefly Algorithm and Adaptive Neuro-Fuzzy Inference System (ANFIS) and validate the approach compared with ANFIS trained by the Hybrid Algorithm, Back Propagation and Particle Swarm Optimization. Andriosopoulos, Doumpos, Papapostolou, and Pouliasis (2013) use a genetic algorithm paired with differential evolution algorithm for portfolio optimization and for addressing the problem of index-tracking in shipping stock and freight markets. D. Wang, Liu, and Wang (2013) predict the price trends of stock futures by proposing a new DT-

SVM (Decision Tree and Support Vector Machine) based on hybrid methods which overcome the problems encountered with alternative methods such as fuzzy logic, or genetic algorithms. Chen (2013) integrates particle swarm optimization with an adaptive-network-based fuzzy inference system (ANFIS) model to predict business failure.

Other approaches that are worth mentioning include machine learning, optimization, case-based reasoning, and evolutionary approaches. Benhayoun, Chairri, El Gonnouni, and Lyhyaoui (2013) employ a Support Vector Machine Model in order to predict a company's creditworthiness by using historical data. Liang and Qu (2013) use a Multi-objective Dynamic Multi-Swarm Particle Swarm Optimizer in order to overcome the difficulties involved in classical methods previously applied to portfolio optimization problems. Chuang (2013) employs case-based reasoning to predict business failures. Xiao, Che, Wang, and Yang (2013) analyze stock price movements by studying the technical pattern of stock prices and provide a strategy for case-based reasoning approaches. Canelas, Neves, and Horta (2013) combine a symbolic Aggregate approximation (SAX) technique (for identifying the relevant patterns) with an optimization kernel based on genetic algorithms (GA) for generating investment rules in order to define less risky trading strategies.

Artificial Intelligence and Financial Investing (2014)

Hybrid approaches increased in popularity with many researchers combining AI techniques. For example, Adhikari and Agrawal (2014) combine random walk with artificial neural networks to create a hybrid system that outperforms its respective parts when applied to forecasting financial data. Bagheri, Peyhani, and Akbari (2014) predict market trends using a hybrid system with swarm intelligence and a fuzzy inference system. Semaan, Harb, and Kassem (2014) demonstrate that neural network based approaches can outperform standard statistical analysis when predicting

exchange rates. Li (2014) applies a hybrid neural network and bee colony algorithm for gold price prediction. Sanz, Bernardo, Herrera, Bustince Sola, and Hagrass (2014) demonstrate a compact evolutionary interval-valued fuzzy rule-based classification system, which is able to predict the real world financial data more accurately. Oreski and Oreski (2014) create a new novel classifier based on Hybrid Genetic algorithm and artificial neural network which is able to find the most relevant data and spend less time in the search space. Costea (2014) combines neural networks with genetic algorithms and fuzzy logic to evaluate the performance of non-banking financial institutions. Pulido, Melin, and Castillo (2014) create a hybrid system by integrating neural networks with particle swarm optimization and fuzzy based system to predict the Mexican stock exchange. Kristjanpoller, Fadic, and Minutolo (2014) implement a hybrid neural network model to predict financial market volatility. Reid, Malan, and Engelbrecht (2014) focus on solving realistic portfolio problems by using the Particle Swarm Optimization (PSO) and the Ant Colony Optimization (ACO) algorithms in order to overcome the problem of the uncovered interest rate parity condition. F. Wang, Yu, and Cheung (2014) propose a new performance based reward strategy that combines the moving average and the trading range breakout in a particle swarm optimization scenario to predicting the stock price trends accurately.

Other approaches worth mentioning: Bertella, Pires, Feng, and Stanley (2014) use agent based techniques to support decision making in stock market evaluation. K.-Y. Shen and Tzeng (2014) use a neuro-fuzzy approach to financial performance evaluation evaluated on commercial banks in Taiwan. Aouni, Colapinto, and La Torre (2014) show that goal programming can be applied to various applications of portfolio selection problem and helps the financial decision maker improve their decision making process. Zheng, Zhou, Chen, and Ekedebe (2014) propose a new automated evaluation system, which utilizes machine learning to automatically predict risk based on the data in

the SEC EDGAR database. Finally, Yu, Miche, Séverin, and Lendasse (2014) focus on utilizing the Leave-One-Out-Incremental Extreme Learning Machine (LOO-IELM) in order to study the bankruptcy prediction problem.

Artificial Intelligence and Financial Investing (2015)

This article concludes with the first 2 quarters of the 2015 calendar year. We see hybrid systems continuing to garner interest in the academic community. Patel, Shah, Thakkar, and Kotecha (2015) use a hybrid approach consisting of support vector regression, random forest, and neural networks for stock market index prediction. Rather, Agarwal, and Sastry (2015) propose combining neural networks with linear models for stock market data prediction. Musto, Semeraro, Lops, DeGemmis, and Lekkas (2015) propose a new framework for personalized investment portfolios, which is based on a combining case based reasoning with diversification techniques. Silva, Vasconcelos, Barros, and Franca (2015) integrate case-based reasoning with neural networks for credit risk analysis. Iturriaga and Sanz (2015) propose a new hybrid model implemented with a multilayer perceptron and self-organizing map in order to identify new metrics of financial risks in US commercial banks. Finally, Two other approaches include examples where Chen and Chen (2015) propose a fuzzy system for making stock market predictions and Yang, Chen, and Huang (2015) advance agent based systems by illustrating a framework for developing agents to stock market trading which was shown to outperform current agent based approaches such as random traders.

FUTURE RESEARCH DIRECTIONS

Artificial neural networks (ANN) remain the most popular form of AI technique in finance research (Bahrammirzaee, 2010). ANN approaches have been shown to outperform other techniques such as

genetic algorithms and multivariate discriminant analysis by more than 5% and 18% respectively (Rafiei et al., 2011). Research has moved away from a purely ANN approach to hybrid approaches, such as ANN and support vector machines, ANN and evolutionary computing, as well as ANN and fuzzy approaches. Future research will benefit from refining the hybrid AI approaches as well as incorporating additional AI techniques into hybrid systems when applied to finance. As big data and data science emerge, one would expect to see more literature encompassing related techniques. Future directions would benefit from increasing the variety and mixture of hybrid techniques while exploiting data science and big data concepts such as parallel processing (map reduce) or even real-time stream mining.

CONCLUSION

This article presents an overview of artificial intelligence applied to financial investing. First, the background of AI techniques and finance are discussed. Following the background, the focus of the article is identifying the direction of AI in finance from 2009 to 2013 and subsequently 2013 to the first half of 2015. A trend has been shown that single AI techniques dominated early research and progresses toward hybrid approaches. Neural networks remained one of the most popular methods throughout the literature and, as hybrid systems advanced, neural networks were one of the most common techniques included in hybrid systems.

REFERENCES

Adhikari, R., & Agrawal, R. (2014). A combination of artificial neural network and random walk models for financial time series forecasting. *Neural Computing & Applications*, 24(6), 1441–1449. doi:10.1007/s00521-013-1386-y

- Al-Jumeily, D., Hussain, A., & Alaskar, H. (2013). *Recurrent neural networks inspired by artificial immune algorithm for time series prediction*. Paper presented at the Neural Networks (IJCNN), The 2013 International Joint Conference on. doi:10.1109/IJCNN.2013.6707137
- Andriosopoulos, K., Doumpos, M., Papapostolou, N. C., & Pouliasis, P. K. (2013). Portfolio optimization and index tracking for the shipping stock and freight markets using evolutionary algorithms. *Transportation Research Part E, Logistics and Transportation Review*, 52, 16–34. doi:10.1016/j.tre.2012.11.006
- Aouni, B., Colapinto, C., & La Torre, D. (2014). Financial portfolio management through the goal programming model: Current state-of-the-art. *European Journal of Operational Research*, 234(2), 536–545. doi:10.1016/j.ejor.2013.09.040
- Bae, J. K. (2012). Predicting financial distress of the South Korean manufacturing industries. *Expert Systems with Applications*, 39(10), 9159–9165. doi:10.1016/j.eswa.2012.02.058
- Bagheri, A., Peyhani, H. M., & Akbari, M. (2014). Financial forecasting using ANFIS networks with quantum-behaved particle swarm optimization. *Expert Systems with Applications*, 41(14), 6235–6250. doi:10.1016/j.eswa.2014.04.003
- Bahrammirzaee, A. (2010). A comparative survey of artificial intelligence applications in finance: Artificial neural networks, expert system and hybrid intelligent systems. *Neural Computing & Applications*, 19(8), 1165–1195. doi:10.1007/s00521-010-0362-z
- Benhayoun, N., Chairi, I., El Gonnouni, A., & Lyhyaoui, A. (2013). Financial Intelligence in Prediction of Firms Creditworthiness Risk: Evidence from Support Vector Machine Approach. *Procedia Economics and Finance*, 5, 103–112. doi:10.1016/S2212-5671(13)00015-4
- Bermúdez, J. D., Segura, J. V., & Vercher, E. (2012). A multi-objective genetic algorithm for cardinality constrained fuzzy portfolio selection. *Fuzzy Sets and Systems*, 188(1), 16–26. doi:10.1016/j.fss.2011.05.013
- Bertella, M. A., Pires, F. R., Feng, L., & Stanley, H. E. (2014). Confidence and the Stock Market: An Agent-Based Approach. *PLoS ONE*, 9(1), e83488. doi:10.1371/journal.pone.0083488 PMID:24421888
- Bhattacharyya, S., Pictet, O. V., & Zumbach, G. (2002). Knowledge-intensive genetic discovery in foreign exchange markets. *IEEE Transactions on Evolutionary Computation*, 6(2), 169–181. doi:10.1109/4235.996016
- Birău, R., Ehsanifar, M., & Mohammadi, H. (2013). *Forecasting the Bucharest Stock Exchange BET-C Index based on Artificial Neural Network and Multiple Linear Regressions*. Paper presented at the Proceedings of the 1st WSEAS International Conference on Mathematics, Statistics & Computer Engineering, Dubrovnik, Croatia.
- Boyacioglu, M. A., & Avci, D. (2010). An Adaptive Network-Based Fuzzy Inference System (ANFIS) for the prediction of stock market return: The case of the Istanbul Stock Exchange. *Expert Systems with Applications*, 37(12), 7908–7912. doi:10.1016/j.eswa.2010.04.045
- Canelas, A., Neves, R., & Horta, N. (2013). A SAX-GA approach to evolve investment strategies on financial markets based on pattern discovery techniques. *Expert Systems with Applications*, 40(5), 1579–1590. doi:10.1016/j.eswa.2012.09.002
- Chang, P.-C., Fan, C.-Y., & Lin, J.-L. (2011). Trend discovery in financial time series data using a case based fuzzy decision tree. *Expert Systems with Applications*, 38(5), 6070–6080. doi:10.1016/j.eswa.2010.11.006

- Chen, M.-Y. (2011). Predicting corporate financial distress based on integration of decision tree classification and logistic regression. *Expert Systems with Applications*, 38(9), 11261–11272. doi:10.1016/j.eswa.2011.02.173
- Chen, M.-Y. (2013). A hybrid ANFIS model for business failure prediction utilizing particle swarm optimization and subtractive clustering. *Information Sciences*, 220, 180–195. doi:10.1016/j.ins.2011.09.013
- Chen, M.-Y., & Chen, B.-T. (2015). A hybrid fuzzy time series model based on granular computing for stock price forecasting. *Information Sciences*, 294, 227–241. doi:10.1016/j.ins.2014.09.038
- Chi, B.-W., & Hsu, C.-C. (2012). A hybrid approach to integrate genetic algorithm into dual scoring model in enhancing the performance of credit scoring model. *Expert Systems with Applications*, 39(3), 2650–2661. doi:10.1016/j.eswa.2011.08.120
- Chuang, C.-L. (2013). Application of hybrid case-based reasoning for enhanced performance in bankruptcy prediction. *Information Sciences*, 236, 174–185. doi:10.1016/j.ins.2013.02.015
- Chuang, C.-L., & Huang, S.-T. (2011). A hybrid neural network approach for credit scoring. *Expert Systems: International Journal of Knowledge Engineering and Neural Networks*, 28(2), 185–196. doi:10.1111/j.1468-0394.2010.00565.x
- Costea, A. (2014). Applying Fuzzy Logic and Machine Learning Techniques in Financial Performance Predictions. *Procedia Economics and Finance*, 10, 4–9. doi:10.1016/S2212-5671(14)00271-8
- de Oliveira, F. A., Nobre, C. N., & Zárate, L. E. (2013). Applying Artificial Neural Networks to prediction of stock price and improvement of the directional prediction index—Case study of PETR4, Petrobras, Brazil. *Expert Systems with Applications*, 40(18), 7596–7606. doi:10.1016/j.eswa.2013.06.071
- Devadoss, A. V., & Ligori, T. A. A. (2013). Stock Prediction Using Artificial Neural Networks. *International Journal of Data Mining Techniques and Applications*, 2, 283–291.
- Dong, G., Fataliyev, K., & Wang, L. (2013). *One-step and multi-step ahead stock prediction using backpropagation neural networks*. Paper presented at the Information, Communications and Signal Processing (ICICS) 2013 9th International Conference on.
- Evans, C., Pappas, K., & Xhafa, F. (2013). Utilizing artificial neural networks and genetic algorithms to build an algo-trading model for intra-day foreign exchange speculation. *Mathematical and Computer Modelling*, 58(5), 1249–1266. doi:10.1016/j.mcm.2013.02.002
- Fu, T.-c., Chung, C.-p., & Chung, F.-l. (2013). Adopting genetic algorithms for technical analysis and portfolio management. *Computers & Mathematics with Applications (Oxford, England)*, (0). doi:10.1016/j.camwa.2013.08.012
- Gradojevic, N., & Gençay, R. (2013). Fuzzy logic, trading uncertainty and technical trading. *Journal of Banking & Finance*, 37(2), 578–586. doi:10.1016/j.jbankfin.2012.09.012
- Hadavandi, E., Shavandi, H., & Ghanbari, A. (2010). Integration of genetic fuzzy systems and artificial neural networks for stock price forecasting. *Knowledge-Based Systems*, 23(8), 800–808. doi:10.1016/j.knosys.2010.05.004
- Hassan, M. R., Ramamohanarao, K., Kamruzzaman, J., Rahman, M., & Hossain, M. M. (2013). A HMM-based adaptive fuzzy inference system for stock market forecasting. *Neurocomputing*, 104, 10–25. doi:10.1016/j.neucom.2012.09.017
- Iturriaga, F. J. L., & Sanz, I. P. (2015). Bankruptcy visualization and prediction using neural networks: A study of US commercial banks. *Expert Systems with Applications*, 42(6), 2857–2869. doi:10.1016/j.eswa.2014.11.025

- Kaucic, M. (2010). Investment using evolutionary learning methods and technical rules. *European Journal of Operational Research*, 207(3), 1717–1727. doi:10.1016/j.ejor.2010.07.008
- Kirkos, E., Spathis, C., & Manolopoulos, Y. (2010). Audit-firm group appointment: An artificial intelligence approach. *Intelligent Systems in Accounting, Finance & Management*, 17(1), 1–17. doi:10.1002/isaf.310
- Korol, T. (2013). Early warning models against bankruptcy risk for Central European and Latin American enterprises. *Economic Modelling*, 31, 22–30. doi:10.1016/j.econmod.2012.11.017
- Kristjanpoller, W., Fadic, A., & Minutolo, M. C. (2014). Volatility forecast using hybrid Neural Network models. *Expert Systems with Applications*, 41(5), 2437–2442. doi:10.1016/j.eswa.2013.09.043
- Kumar, D. A., & Murugan, S. (2013). *Performance analysis of indian stock market index using neural network time series model*. Paper presented at the Pattern Recognition, Informatics and Mobile Engineering (PRIME), 2013 International Conference on. doi:10.1109/ICPRIME.2013.6496450
- Ladyzynski, P., & Grzegorzewski, P. (2013). Particle swarm intelligence tuning of fuzzy geometric protoforms for price patterns recognition and stock trading. *Expert Systems with Applications*, 40.
- Li, B. (2014). Research on WNN modeling for gold price forecasting based on improved artificial bee colony algorithm. *Computational Intelligence and Neuroscience*, 2014, 2. doi:10.1155/2014/270658 PMID:24744773
- Liang, J., & Qu, B. (2013). *Large-scale portfolio optimization using multiobjective dynamic multi-swarm particle swarm optimizer*. Paper presented at the Swarm Intelligence (SIS), 2013 IEEE Symposium on. doi:10.1109/SIS.2013.6615152
- Liu, Q., Dang, C., & Huang, T. (2013). A one-layer recurrent neural network for real-time portfolio optimization with probability criterion. *Cybernetics. IEEE Transactions on*, 43(1), 14–23.
- Magoč, T., & Modave, F. (2011). The optimality of non-additive approaches for portfolio selection. *Expert Systems with Applications*, 38(10), 12967–12973. doi:10.1016/j.eswa.2011.04.093
- Manahov, V., & Hudson, R. (2013). Herd behaviour experimental testing in laboratory artificial stock market settings. Behavioural foundations of stylised facts of financial returns. *Physica A: Statistical Mechanics and its Applications*, 392(19), 4351–4372.
- Messaoudi, L., & Rebai, A. (2013). *A fuzzy stochastic Goal Programming approach for solving portfolio selection problem*. Paper presented at the Modeling, Simulation and Applied Optimization (ICMSAO), 2013 5th International Conference on. doi:10.1109/ICMSAO.2013.6552668
- Musto, C., Semeraro, G., Lops, P., DeGemmis, M., & Lekkass, G. (2015). *Financial Product Recommendation through Case-based Reasoning and Diversification Techniques*. Paper presented at the ACM Conference on Recommender Systems.
- Nhu, H. N., Nitsuwat, S., & Sodanil, M. (2013). *Prediction of stock price using an adaptive Neuro-Fuzzy Inference System trained by Firefly Algorithm*. Paper presented at the Computer Science and Engineering Conference (ICSEC), 2013 International. doi:10.1109/ICSEC.2013.6694798
- Oreski, S., & Oreski, G. (2014). Genetic algorithm-based heuristic for feature selection in credit risk assessment. *Expert Systems with Applications*, 41(4), 2052–2064. doi:10.1016/j.eswa.2013.09.004
- Oyewale, A. (2013). Evaluation of Artificial Neural Networks in Foreign Exchange Forecasting. *American Journal of Theoretical and Applied Statistics*, 2(4), 94–101. doi:10.11648/j.ajtas.20130204.11

- Özkan, F. (2013). Comparing the forecasting performance of neural network and purchasing power parity: The case of Turkey. *Economic Modelling*, 31, 752–758. doi:10.1016/j.econmod.2013.01.010
- Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock market index using fusion of machine learning techniques. *Expert Systems with Applications*, 42(4), 2162–2172. doi:10.1016/j.eswa.2014.10.031
- Phillips-Wren, G. (2012). AI Tools in Decision Making Support Systems: A Review. *International Journal of Artificial Intelligence Tools*, 21(2), 1240005. doi:10.1142/S0218213012400052
- Pulido, M., Melin, P., & Castillo, O. (2014). Particle swarm optimization of ensemble neural networks with fuzzy aggregation for time series prediction of the Mexican Stock Exchange. *Information Sciences*, 280, 188–204. doi:10.1016/j.ins.2014.05.006
- Rada, R. (2008). Expert systems and evolutionary computing for financial investing: A review. *Expert Systems with Applications*, 34(4), 2232–2240. doi:10.1016/j.eswa.2007.05.012
- Rafiei, F. M., Manzari, S. M., & Bostanian, S. (2011). Financial health prediction models using artificial neural networks, genetic algorithm and multivariate discriminant analysis: Iranian evidence. *Expert Systems with Applications*, 38(8), 10210–10217. doi:10.1016/j.eswa.2011.02.082
- Rather, A. M., Agarwal, A., & Sastry, V. (2015). Recurrent neural network and a hybrid model for prediction of stock returns. *Expert Systems with Applications*, 42(6), 3234–3241. doi:10.1016/j.eswa.2014.12.003
- Reid, S. G., Malan, K. M., & Engelbrecht, A. P. (2014). Carry trade portfolio optimization using particle swarm optimization. Paper presented at the Evolutionary Computation (CEC), 2014 IEEE Congress on. doi:10.1109/CEC.2014.6900497
- Sanz, J., Bernardo, D., Herrera, F., Bustince Sola, H., & Hagrass, H. (2014). *A compact evolutionary interval-valued fuzzy rule-based classification system for the modeling and prediction of real-world financial applications with imbalanced data*. Academic Press.
- Semaan, D., Harb, A., & Kassem, A. (2014). Forecasting exchange rates: Artificial neural networks vs regression. Paper presented at the e-Technologies and Networks for Development (ICeND), 2014 Third International Conference on. doi:10.1109/ICeND.2014.6991371
- Sermpinis, G., Theofilatos, K., Karathanasopoulos, A., Georgopoulos, E. F., & Dunis, C. (2013). Forecasting foreign exchange rates with adaptive neural networks using radial-basis functions and particle swarm optimization. *European Journal of Operational Research*, 225(3), 528–540. doi:10.1016/j.ejor.2012.10.020
- Shen, K.-L. (2013). Implementing value investing strategy through an integrated fuzzy-ann model. *Journal of Theoretical and Applied Information Technology*, 51(1), 150–157.
- Shen, K.-Y., & Tzeng, G.-H. (2014). DRSA-based neuro-fuzzy inference systems for the financial performance prediction of commercial banks. *International Journal of Fuzzy Systems*, 16(2), 173–183.
- Silva, C., Vasconcelos, G., Barros, H., & Franca, G. (2015). Case-based reasoning combined with neural networks for credit risk analysis. Paper presented at the Neural Networks (IJCNN), 2015 International Joint Conference on. doi:10.1109/IJCNN.2015.7280738
- Svalina, I., Galzina, V., Lujić, R., & Šimunović, G. (2013). An adaptive network-based fuzzy inference system (ANFIS) for the forecasting: The case of close price indices. *Expert Systems with Applications*, 40(15), 6055–6063. doi:10.1016/j.eswa.2013.05.029

- Sycara, K., Decker, K., Pannu, A., Williamson, M., & Zeng, D. (1996). Distributed intelligent agents. *IEEE Expert*, 11(6), 36–46. doi:10.1109/64.546581
- Tarsauliya, A., Kant, S., & Kala, R. (2010). Analysis of Artificial Neural Network for Financial Time Series Forecasting. *International Journal of Computers and Applications*, 9(5).
- Teixeira, L. A., & de Oliveira, A. L. I. (2010). A method for automatic stock trading combining technical analysis and nearest neighbor classification. *Expert Systems with Applications*, 37(10), 6885–6890. doi:10.1016/j.eswa.2010.03.033
- Thomas, J. D. (2003). *News and Trading Rules (Ph.D. Thesis)*. Pittsburgh, PA: Carnegie Mellon University.
- Tsakonas, A., Dounias, G., Doumpos, M., & Zopounidis, C. (2006). Bankruptcy prediction with neural logic networks by means of grammar-guided genetic programming. *Expert Systems with Applications*, 30(3), 449–461. doi:10.1016/j.eswa.2005.10.009
- Wang, D., Liu, X., & Wang, M. (2013). *A DT-SVM Strategy for Stock Futures Prediction with Big Data*. Paper presented at the Computational Science and Engineering (CSE), 2013 IEEE 16th International Conference on. doi:10.1109/CSE.2013.147
- Wang, F., Yu, P. L., & Cheung, D. W.-l. (2014). *Combining technical trading rules using parallel particle swarm optimization based on Hadoop*. Paper presented at the Neural Networks (IJCNN), 2014 International Joint Conference on. doi:10.1109/IJCNN.2014.6889599
- Won, C., Kim, J., & Bae, J. K. (2012). Using genetic algorithm based knowledge refinement model for dividend policy forecasting. *Expert Systems with Applications*, 39(18), 13472–13479. doi:10.1016/j.eswa.2012.06.001
- Xiao, Y., Che, W.-G., Wang, Z., & Yang, C.-C. (2013). *The Research of Morphological Characteristics in Time Series of Stock Prices Based on CBR*. Paper presented at the Intelligent System Design and Engineering Applications (ISDEA), 2013 Third International Conference on. doi:10.1109/ISDEA.2012.364
- Yang, G., Chen, Y., & Huang, J. (2015). The highly intelligent virtual agents for modeling financial markets. *Physica A: Statistical Mechanics and its Applications*.
- Yu, Q., Miche, Y., Séverin, E., & Lendasse, A. (2014). Bankruptcy prediction using extreme learning machine and financial expertise. *Neurocomputing*, 128, 296–302. doi:10.1016/j.neucom.2013.01.063
- Yunusoglu, M. G., & Selim, H. (2013). A fuzzy rule based expert system for stock evaluation and portfolio construction: An application to Istanbul Stock Exchange. *Expert Systems with Applications*, 40(3), 908–920. doi:10.1016/j.eswa.2012.05.047
- Zheng, Y., Zhou, H., Chen, Z., & Ekedebe, N. N. (2014). *Automated analysis and evaluation of SEC documents*. Paper presented at the Computer and Information Science (ICIS), 2014 IEEE/ACIS 13th International Conference on. doi:10.1109/ICIS.2014.6912118

ADDITIONAL READING

- Deborah, L. M., Richard, F., James, R., & Steve, W. (2000). *An Environment for Merging and Testing Large Ontologies*. Morgan Kaufmann.
- Holland, J. (1975). *Adaptation in natural and artificial systems*. Ann Arbor, MI: University of Michigan Press.

KEY TERMS AND DEFINITIONS

Artificial Intelligence: The ability of a computer to perform activities which are normally considered to require human intelligence.

Asset Valuation: The process of determining the worth of something.

Case-Based Reasoning: The process of solving new problems based on successful past solutions to similar problems.

Decision Tree: A tree like structure for modeling decisions and classifying data.

Evolutionary Computing: Branch of artificial intelligence which mimics biological evolution and often applied to optimization problems.

Expert System: A program that uses knowledge and inferences to solve problems in a way that experts might.

Fuzzy Systems: Systems which deal in approximations as opposed to exact representations (i.e. true/ false).

Genetic Algorithm: An algorithm that mimics the genetic concepts of natural selection, combination, selection, and inheritance.

Hybrid System: A system which employs a combination of techniques and methods.

Investing: The act of committing money to an endeavor with the expectation of obtaining profit.

Machine Learning: A method of automatically learning patterns from data in order to make future predictions.

Neural Networks: Programs that simulate a network of communicating nerve cells to achieve a machine learning objective.

Risk Management: The process of managing the uncertainty in investment decision-making.

Support Vector Machines (SVM): Supervised learning model to analyze data and predict which possible classes form the output.

Swarm Intelligence: Intelligence based on many individuals and decentralized control and self-organization.

Distributed Parameter Systems Control and Its Applications to Financial Engineering

A

Gerasimos Rigatos

Industrial Systems Institute, Greece

Pierluigi Siano

University of Salerno, Italy

INTRODUCTION

In several problems of financial engineering, such as options and commodities trading, forecasting of options' values, estimation of financial distress and credit risk assessment, validation of option pricing models, etc. one comes against Partial Differential Equations (PDEs). Moreover, in problems of control of financial systems where the aim is to stabilize financial processes which are described by PDE models, one has to harness again the complex PDE dynamics through the application of an external input. In the recent years differential flatness theory has emerged as an approach to the control and stabilization of systems described by PDE dynamics (Rudolph, 2003), (Rigatos, 2015). This research work focuses on differential flatness theory for the control and stabilization of single asset and multi-asset option price dynamics, described by PDE models. It is shown how the differential flatness approach achieves, stabilization of distributed parameter financial systems (that is systems modelled by PDEs) and how it enables convergence to specific financial performance indexes (Rigatos, 2014a; Rigatos, 2014b; Rigatos, 2014c; Rigatos, 2015a; Rigatos, 2015b; Rigatos, 2015c).

The Black-Scholes PDE is the principal financial model used in this study. It is demonstrated how with the use of semi-discretization and a finite differences scheme the single-asset (equivalently multi-asset) Black-Scholes PDE is transformed into a state-space model consisting of ordinary

nonlinear differential equations. For this set of differential equations it is proven that differential flatness properties hold (Rigatos, 2011; Rigatos, 2013; Rigatos, 2015). This permits to arrive at a solution for the associated control problem and to ascertain stabilization of the options' dynamics. By proving that it is feasible to control the single-asset (equivalently multi-asset) Black-Scholes PDE it is also concluded that through a selected trading policy, the price of options can be made to converge and stabilize at specific reference values.

The computational part of the considered feedback control method is as follows: For the local subsystems, into which the single-asset (equivalently multi-asset) Black-Scholes PDE is decomposed, it becomes possible to apply boundary-based feedback control. The controller design proceeds by showing that the state-space model of the single-asset (equivalently multi-asset) Black-Scholes PDE stands for a differentially flat system. Next, for each subsystem which is related to a nonlinear ODE, a virtual control input is computed, that can invert the subsystem's dynamics and can eliminate the subsystem's tracking error. From the last row of the state-space description, the control input (boundary condition) that is actually applied to the single-asset (equivalently multi-asset) Black-Scholes PDE system is found. This control input contains recursively all virtual control inputs which were computed for the individual ODE subsystems associated with the previous rows of the state-space equation. Thus, by tracing the rows of the state-space model

DOI: 10.4018/978-1-5225-2255-3.ch002

backwards, at each iteration of the control algorithm, one can finally obtain the control input that should be applied to the single-asset (equivalently multi-asset) Black-Scholes PDE system so as to assure that all its state variables will converge to the desirable setpoints.

The structure of the chapter is as follows: in Section “*The problem of boundary control of the single-asset Black-Scholes PDE*”, an overview about the single-asset Black-Scholes PDE is given and the associated boundary control problem is formulated. In Section “*Option pricing modelling with the use of the single-asset Black-Scholes PDE*” the concept of option pricing for the single-asset Black-Scholes PDE is explained. In Section “*Transformation of the single-asset Black-Scholes PDE into nonlinear ODEs*” it is explained how the single-asset Black-Scholes PDE dynamics can be transformed to an equivalent state-space form. In Section “*Computation of boundary control for the single-asset Black-Scholes PDE*” a boundary feedback control law is computed for the single-asset Black-Scholes PDE. In Section “*Closed loop dynamics of the single-asset Black-Scholes PDE*” the dynamics of the closed control loop of the single-asset Black-Scholes PDE is analysed. In Section “*The problem of boundary control of the multi-asset Black-Scholes PDE*” the multi-asset Black-Scholes PDE is introduced and the associated boundary control problem is formulated. In Section “*Boundary control of the multi-asset Black-Scholes PDE*” it is explained how the multi-asset Black-Scholes PDE can be transformed into an equivalent state-space description. In Section “*Flatness-based control of the multi-asset Black-Scholes PDE*” it is analysed how a boundary feedback control input can be computed for the multi-asset Black-Scholes PDE. In Section “*Simulation tests*” the satisfactory performance of the control loop is confirmed through simulation experiments for both the single-asset and the multi-asset Black-Scholes PDE. Finally, in Section “*Conclusions*” concluding remarks are stated.

BACKGROUND

One can note two modelling approaches for describing option price dynamics. The first one makes use of stochastic differential equations (SDEs). The underlying asset evolves according to a stochastic process that is driven by a random input. Using this representation, stochastic control methods for option price SDE diffusion models have been developed (Bensoussan, 2000), (Pascucci, 2011). Moreover, according to Kolmogorov’s theory, diffusion stochastic processes can be equivalently represented by partial differential equations (PDEs). These provide as solution the spatiotemporal distribution of the options’ value. The Black-Scholes PDE is such a relation (Platen & Heath, 2008; Sircar & Papanicolaou, 1998). Consequently, methods for PDE boundary control can be used for modifying option price dynamics. This topic will be elaborated in this article.

THE PROBLEM OF BOUNDARY CONTROL OF THE SINGLE-ASSET BLACK-SCHOLES PDE

Control and stabilization of financial systems is a difficult problem since the associated models have spatiotemporal dynamics and are either described by partial differential equations or by stochastic differential equations (Platen & Heath, 2006; Pascucci, 2011). On the one side control approaches for financial systems have been developed with the use of stochastic differential equations (Oksendal & Sulem, 2006; Stojanovic, 2007; Yin et al., 2010). On the other side, control of financial dynamics through the use of the associated partial differential equations description remains an open problem for which efficient solutions have to be provided (Rudolph, 2003; Smyshlyaev & Krstic, 2010). To this end, in this research work a new control method is developed for the diffusion-type of the Black-Scholes PDE which describes the dynamics of options in financial markets. It is

shown that boundary control, that is control based on the boundary conditions of the PDE, can be exerted on the Black-Scholes PDE thus modifying its dynamics and leading the options' prices to converge to specific reference values. Controlling the dynamics of financial systems with the application of exogenous inputs and feedback has been attempted in several research works (Miller & Weller, 1995; Barmish, 2011; Fliess & Join, 2012; Fliess & Join, 2010). Control of financial systems is important for minimizing default risks, maximizing profits or stabilizing the dynamics of stock market indexes (Fedorov & Mikhailov, 2001; Forsyth & Labahn, 2007; Bernhard, 2006; Windcliff et al., 2004). The present chapter treats the problem of boundary control of the nonlinear Black-Scholes PDE, which means that the boundary conditions are used as control inputs to modify this PDE dynamics. Boundary control or lumped-input control of nonlinear distributed parameter systems are problems of elevated difficulty and many remarkable results on this topic exist (Woitteneck & Mounier, 2010; Mounier et al., 2010; Fliess & Mounier, 2002; Boussaada et al., 2013). If the PDE system is decomposed into an equivalent set of ordinary nonlinear differential equations, then controlling such a set of dynamical subsystems by the boundary conditions implies underactuation.

The control approach followed in this chapter is as follows: First, by implementing a procedure as in the numerical solution of the Black-Scholes PDE a set of equivalent nonlinear ordinary differential equations is obtained (Pinsky, 1991; Gerdtz et al., 2008; Guo & Billings 2007; Kroner, 2011; Winkler & Lohmann, 2010; Dragonescu & Soane, 2013). Next it is shown that the system of the nonlinear ODEs is a differentially flat one. This means that all its state variables and the control inputs can be written as differential functions of one single algebraic variable which is the flat output (Sira-Ramirez & Agrawal, 2004; Lévine, 2009; Bououden et al., 2011; Fliess & Mounier, 1999; Lévine, 2011). Moreover, by examining independently each nonlinear ODE it is shown

that this stands again for a differentially flat system, for which a virtual control input can be computed as in the case of flatness-based control for the trivial system. The virtual control input is chosen such that the ODE subsystem dynamics is linearized and the tracking error is eliminated. The boundary condition that appears in the nonlinear ODE subsystem that comprises the last row of the state-space description, stands for the aggregate control input. The computation of the boundary control input uses recursively all virtual control inputs mentioned above, moving from the last ODE system to the first one. Thus, by tracing the rows of the state-space model backwards, at each iteration of the control algorithm, one can finally obtain the control input that should be applied to the Black-Scholes PDE system so as to assure that all its state vector elements will converge to the desirable setpoints. By analyzing the dynamics of the closed-loop system that results from the application of the aforementioned control method, asymptotic stability is confirmed.

OPTION PRICING MODELLING WITH THE USE OF THE SINGLE-ASSET BLACK-SCHOLES PDE

Definition of Options

A financial derivative is called a European option if it gives the right for pay-off according to a function $H : [0, \infty] \rightarrow R$ which has as principle variable the security index S_t , while there is also a predefined expiration date $T \in [0, \infty)$. If the payoff can take place before the expiration date then one has an American option. The value function of a European option is denoted as $V(t, S_t)$, where t varies in the interval $t \in [0, T]$, while the associated payoff function is defined as H and the maturity time is $T \in [0, \infty)$. The value function $V : [0, T] \times [0, \infty) \rightarrow R$ for a European option, can be differentiated with respect to time while it is also twice differentiable with respect to the security index S .

Option Price Modelling with The Use of Stochastic Differential Equations

The option price model assumes that the security price $S = \{S_t, t \in [0, T]\}$ with initial value equal to S_0 follows a geometric Brownian motion

$$dS_t = a_t S_t dt + \sigma_t S_t dW_t \quad (1)$$

where $a = \{a_t, t \in [0, T]\}$ is the appreciation rate which is a positive variable, and $\sigma = \{\sigma_t, t \in [0, T]\}$ is the volatility parameter. Variable W denoted a Wiener process, $W = \{W_t, t \in [0, T]\}$. There is also a second stochastic process in the model which is given by

$$dB_t = r_t B_t dt \quad (2)$$

where B_t corresponds to the domestic savings parameter $B = \{B_t, t \in [0, T]\}$ with initial value $B_0 = 1$. Variable r_t is the interest rate $r = \{r_t, t \in [0, T]\}$. The domestic savings account is also called locally riskless asset when there is no noise term in the SDE model given in Eq. (2).

The Black-Scholes PDE

A model that is equivalent to the SDE description of option price dynamics is provided by the Black-Scholes PDE (Platen & Heath, 2006; Pascucci, 2011). The Black-Scholes partial differential equation is given by

$$\frac{\partial V(t, S)}{\partial t} + r_t S \frac{\partial V(t, S)}{\partial S} + \frac{1}{2} \sigma_t^2 S^2 \frac{\partial^2 V(t, S)}{\partial S^2} - r_t V(t, S) = 0 \quad (3)$$

For $t \in (0, T)$ and $S \in (0, \infty)$ where r is the interest rate and σ_t is the volatility, while the associated terminal condition is

$$V(T, S) = H(S) \quad (4)$$

Equation 3 and Equation 4 form the Black-Scholes partial differential equation. The Black-Scholes PDE is a diffusion partial differential equation which describes the evolution of the option's price distribution $V(t, S)$, as a function of time t and of the underlying asset (security index) S .

TRANSFORMATION OF THE SINGLE-ASSET BLACK-SCHOLES PDE INTO NONLINEAR ODES

Next, the following nonlinear Black-Scholes PDE is considered:

$$\frac{\partial V(t, S)}{\partial t} + r_t S \frac{\partial V(t, S)}{\partial S} + \frac{1}{2} \sigma_t^2 S^2 \frac{\partial^2 V(t, S)}{\partial S^2} - r_t V(t, S) = 0 \quad (5)$$

It is assumed that the volatility σ_t has a nonlinear dependence on V_{ss} that is

$$\sigma_t = \sigma(V_{ss}) \quad (6)$$

where $V_{ss} = \frac{\partial^2 V(t, S)}{\partial S^2}$ and σ is a nonlinear function. A grid of N points is considered, that is $\{s_1, s_2, \dots, s_{N-1}, s_N\}$ which are placed at equal distances on the S axis. At the points of spatial discretization, it holds

$$\frac{\partial V(t, S_i)}{\partial t} = r_t V(t, S_i) - r_t S_i \frac{\partial V(t, S_i)}{\partial S} - \frac{1}{2} \sigma_t^2 S_i^2 \frac{\partial^2 V(t, S_i)}{\partial S^2} \quad (7)$$

where $i = 1, 2, \dots, N$. The following state vector is defined as shown in Box 1.

Box 1.

A

$$\tilde{V} = [V(t, s_1), V(t, s_2), V(t, s_3), \dots, V(t, s_{N-1}), V(t, s_N)] \quad \text{ur} \quad (8)$$

$$\tilde{V} = [V_1, V_2, V_3, \dots, V_{N-1}, V_N]$$

thus one has

$$\begin{aligned} \frac{\partial V_1}{\partial t} &= r_1 V_1 - r_1 S_1 \frac{V_2 - V_1}{\Delta S} - \frac{1}{2} \sigma_t^2(V_{ss}) S_1^2 \frac{V_2 - 2V_1 + V_0}{\Delta S^2} \\ \frac{\partial V_2}{\partial t} &= r_1 V_2 - r_1 S_2 \frac{V_3 - V_2}{\Delta S} - \frac{1}{2} \sigma_t^2(V_{ss}) S_2^2 \frac{V_3 - 2V_2 + V_1}{\Delta S^2} \\ \frac{\partial V_3}{\partial t} &= r_1 V_3 - r_1 S_3 \frac{V_4 - V_3}{\Delta S} - \frac{1}{2} \sigma_t^2(V_{ss}) S_3^2 \frac{V_4 - 2V_3 + V_2}{\Delta S^2} \\ &\dots \dots \dots \\ \frac{\partial V_{N-1}}{\partial t} &= r_1 V_{N-1} - r_1 S_{N-1} \frac{V_N - V_{N-1}}{\Delta S} - \frac{1}{2} \sigma_t^2(V_{ss}) S_{N-1}^2 \frac{V_N - 2V_{N-1} + V_{N-2}}{\Delta S^2} \\ &\dots \dots \dots \\ \frac{\partial V_{N-1}}{\partial t} &= r_1 V_N - r_1 S_N \frac{V_{N+1} - V_N}{\Delta S} - \frac{1}{2} \sigma_t^2(V_{ss}) S_N^2 \frac{V_{N+1} - 2V_N + V_{N-1}}{\Delta S^2} \end{aligned} \quad (9)$$

For the i-th ODE one has at sampling point S_i along the S axis one has

$$\begin{aligned} \frac{\partial V_i}{\partial t} &= r_i V_i - r_i S_i \frac{V_{i+1} - V_i}{\Delta S} - \frac{1}{2} \sigma_t^2(V_{ss}) S_i^2 \frac{V_{i+1}}{\Delta S^2} + \\ &\sigma_t^2(V_{ss}) S_i^2 \frac{V_i}{\Delta S^2} - \frac{1}{2} \sigma_t^2(V_{ss}) S_i^2 \frac{V_{i-1}}{\Delta S^2} \end{aligned} \quad (10)$$

Equation 32 is also written in the form

$$\begin{aligned} \frac{\partial V_i}{\partial t} &= -\frac{1}{2} \frac{\sigma_t^2(V_{ss})}{\Delta S^2} S_i^2 V_{i-1} + \frac{\sigma_t^2(V_{ss})}{\Delta S^2} S_i^2 V_i - \\ &[\frac{r_i S_i}{\Delta S} + \frac{1}{2} \frac{\sigma_t^2(V_{ss})}{\Delta S^2} S_i^2] V_{i+1} + [r_i + \frac{r_i S_i}{\Delta S}] V_i \end{aligned} \quad (11)$$

Next, by denoting $K_1 = -\sigma_t^2(V_{ss}) S_i^2$,

$$K_2 = -[2r_i S_i \Delta S + \sigma_t^2(V_{ss}) S_i^2]$$

and $f(V_i) = [r_i + \frac{r_i S_i}{\Delta S}] V_i$ one obtains the following description for Equation 11

$$\begin{aligned} \frac{\partial V_i}{\partial t} &= \frac{K_1}{2\Delta S^2} V_{i-1} - \frac{K_1}{\Delta S^2} V_i + \\ &\frac{K_2}{\Delta S^2} V_{i+1} + f(V_i) \end{aligned} \quad (12)$$

For the i-th ODE where $i = 1, 2, \dots, N$, coefficients K_1 and K_2 are computed at the local grid point S_i . Next, the following state vector is defined for the PDE model:

$$\tilde{Y} = [y_{1,1}, y_{1,2}, \dots, y_{1,i}, \dots, y_{1,N-1}, y_{1,N}]$$

where $y_{1,1} = V_1$, $y_{1,2} = V_2 \dots y_{1,N-1} = V_{N-1}$ and $y_{1,N} = V_N$. It will be shown that the state-space description of the nonlinear PDE dynamics

Box 2.

$$\dot{y}_{1,N} = \frac{K_1}{\Delta S^2} y_{1,N} + f(y_{1,N}) + \frac{K_2}{2\Delta S^2} \phi_{N+1} + \frac{K_1}{2\Delta S^2} y_{1,N-1} \quad (13)$$

$$\dot{y}_{1,N-1} = \frac{K_1}{\Delta S^2} y_{1,N-1} + f(y_{1,N-1}) + \frac{K_2}{2\Delta S^2} y_N + \frac{K_1}{2\Delta S^2} y_{1,N-2} \quad (14)$$

$$\dot{y}_{1,N-2} = \frac{K_1}{\Delta S^2} y_{1,N-2} + f(y_{1,N-2}) + \frac{K_2}{2\Delta S^2} y_{N-1} + \frac{K_1}{2\Delta S^2} y_{1,N-3} \quad (15)$$

$$\dot{y}_{1,i} = \frac{K_1}{\Delta S^2} y_{1,i} + f(y_{1,i}) + \frac{K_2}{2\Delta S^2} y_{1,i+1} + \frac{K_1}{2\Delta S^2} y_{1,i-1} \quad (16)$$

$$\dot{y}_{1,3} = \frac{K_1}{\Delta S^2} y_{1,3} + f(y_{1,3}) + \frac{K_2}{2\Delta S^2} y_{1,4} + \frac{K_1}{2\Delta S^2} y_{1,2} \quad (17)$$

$$\dot{y}_{1,2} = \frac{K_1}{\Delta S^2} y_{1,2} + f(y_{1,2}) + \frac{K_2}{2\Delta S^2} y_{1,3} + \frac{K_1}{2\Delta S^2} y_{1,1} \quad (18)$$

$$\dot{y}_{1,1} = \frac{K_1}{\Delta S^2} y_{1,1} + f(y_{1,1}) + \frac{K_2}{2\Delta S^2} y_{1,2} + \frac{K_1}{2\Delta S^2} \phi_0 \quad (19)$$

It can be proven that the state-space model of the Black-Scholes PDE is a differentially flat one, considering as flat output $y_{1,1} = V_1$. This means that all state variables and the control inputs of the model can be expressed as differential functions of this flat output. Moreover, by examining the i -th row of the state-space description as an independent subsystem with output x_i and virtual control input x_{i+1} it can be shown that this subsystem is also differentially flat with local flat output x_i .

COMPUTATION OF BOUNDARY CONTROL FOR THE SINGLE-ASSET BLACK-SCHOLES PDE

To implement boundary feedback control, the nonlinear diffusion-PDE model is rewritten as shown in Box 2.

The boundary condition ϕ_{N+1} is assumed to be known and remains steady. The boundary

condition ϕ_0 stands for the control input (for example, it can be dependent on the interest rate). The feedback control law is designed as follows:

$$a_1 = y_{1,N-1}^* = \frac{1}{(K_1 / 2\Delta S^2)} [\dot{y}_{1,N} - k_{p,1}(y_{1,N} - y_{1,N}^d)] + \frac{K_1}{\Delta S^2} y_{1,N} - f(y_{1,N}) - \frac{K_2}{2\Delta S^2} \phi_{N+1} \quad (20)$$

$$\begin{aligned} a_2 = y_{1,N-2}^* = & \frac{1}{(K_1 / 2\Delta S^2)} [\dot{y}_{1,N-1} - k_{p,2}(y_{1,N-1} - y_{1,N-1}^d)] \\ & + \frac{K_1}{\Delta S^2} y_{1,N-1} - f(y_{1,N-1}) - \frac{K_2}{2\Delta S^2} y_N \\ \Rightarrow a_2 = & \frac{1}{(K_1 / 2\Delta S^2)} [a_1 - k_{p,2}(y_{1,N-1} - a_1)] + \\ & \frac{K_1}{\Delta S^2} y_{1,N-1} - f(y_{1,N-1}) - \frac{K_2}{2\Delta S^2} y_N \end{aligned} \quad (21)$$

$$\begin{aligned}
a_3 &= y_{1,N-3}^* = \\
&\frac{1}{(K_1 / 2\Delta S^2)} [\dot{y}_{1,N-2} - k_{p,3}(y_{1,N-2} - y_{1,N-2}^d)] + \\
&\frac{K_1}{\Delta S^2} y_{1,N-2} - f(y_{1,N-2}) - \frac{K_2}{2\Delta S^2} y_{N-1} \\
\Rightarrow a_3 &= \frac{1}{(K_1 / 2\Delta S^2)} [\dot{a}_2 - k_{p,3}(y_{1,N-2} - a_2)] + \\
&\frac{K_1}{\Delta S^2} y_{1,N-2} - f(y_{1,N-2}) - \frac{K_2}{2\Delta S^2} y_N
\end{aligned} \tag{22}$$

and continuing in a similar manner

$$\begin{aligned}
a_i &= y_{1,N-i}^* = \\
&\frac{1}{(K_1 / 2\Delta S^2)} [\dot{y}_{1,N-i+1} - k_{p,i}(y_{1,N-i+1} - y_{1,N-i+1}^d)] + \\
&\frac{K_1}{\Delta S^2} y_{1,N-i+1} - f(y_{1,N-i+1}) - \frac{K_2}{2\Delta S^2} y_{N-i+2} \\
\Rightarrow a_i &= \\
&\frac{1}{(K_1 / 2\Delta S^2)} [\dot{a}_{i-1} - k_{p,i}(y_{1,N-i+1} - a_{i-1})] + \\
&\frac{K_1}{\Delta S^2} y_{1,N-i+1} - f(y_{1,N-i+1}) - \frac{K_2}{2\Delta S^2} y_{N-i+2}
\end{aligned} \tag{23}$$

Following this procedure one arrives to compute the control inputs which are associated with the last two rows of the state-space model

$$\begin{aligned}
a_{N-2} &= y_{1,2}^* = \\
&\frac{1}{(K_1 / 2\Delta S^2)} [\dot{y}_{1,3} - k_{p,N-2}(y_{1,3} - y_{1,3}^d)] \\
&+ \frac{K_1}{\Delta S^2} y_{1,3} - f(y_{1,3}) - \frac{K_2}{2\Delta S^2} y_{1,4} \\
\Rightarrow a_{N-2} &= \\
&\frac{1}{(K_1 / 2\Delta S^2)} [\dot{a}_{N-3} - k_{p,N-2}(y_{1,3} - a_{N-3})] \\
&+ \frac{K_1}{\Delta S^2} y_{1,3} - f(y_{1,3}) - \frac{K_2}{2\Delta S^2} y_{1,4}
\end{aligned} \tag{24}$$

$$\begin{aligned}
a_{N-1} &= y_{1,1}^* = \\
&\frac{1}{(K_1 / 2\Delta S^2)} [\dot{y}_{1,2} - k_{p,N-1}(y_{1,2} - y_{1,2}^d)] \\
&+ \frac{K_1}{\Delta S^2} y_{1,2} - f(y_{1,2}) - \frac{K_2}{2\Delta S^2} y_{1,3} \\
\Rightarrow a_{N-1} &= \\
&\frac{1}{(K_1 / 2\Delta S^2)} [\dot{a}_{N-2} - k_{p,N-1}(y_{1,2} - a_{N-2})] \\
&+ \frac{K_1}{\Delta S^2} y_{1,2} - f(y_{1,2}) - \frac{K_2}{2\Delta S^2} y_{1,3}
\end{aligned} \tag{25}$$

and finally

$$\begin{aligned}
a_N &= \phi_0 = \frac{1}{(K_1 / 2\Delta S^2)} [\dot{y}_{1,1} - k_{p,N}(y_{1,1} - y_{1,1}^d)] \\
&+ \frac{K_1}{\Delta S^2} y_{1,1} - f(y_{1,1}) - \frac{K_2}{2\Delta S^2} y_{1,2} \\
\Rightarrow a_N &= \frac{1}{(K_1 / 2\Delta S^2)} [\dot{a}_{N-1} - k_{p,N}(y_{1,1} - a_{N-1})] \\
&+ \frac{K_1}{\Delta S^2} y_{1,1} - f(y_{1,1}) - \frac{K_2}{2\Delta S^2} y_{1,2}
\end{aligned} \tag{26}$$

Consequently, the computation of the aggregate control input $a_N = \phi_0$ which is exerted on the PDE model is performed by moving backwards, and by substituting recursively into ϕ_0 the virtual control inputs $a_{N-1}, a_{N-2}, \dots, a_i, \dots, a_2, a_1$. It is noted that implementation of flatness-based control in cascaded loops has been studied in the case of lumped parameter systems, as for example in electric machines (Dannehl & Fuchs, 2006; Rigatos & Siano, 2002; Rigatos & Siano, 2012). The approach followed in this paper extends this concept to a problem of elevated difficulty, and with a state-space description of higher dimensionality, that is control of distributed parameter systems and in particular control of the Black-Scholes PDE.

CLOSED LOOP DYNAMICS OF THE SINGLE-ASSET BLACK-SCHOLES PDE

By substituting Equation 26 into Equation 19 of the state-space model of the PDE dynamics, and using the definition $y_{1,1} - y_{1,1}^d = z_1$ one has

$$\begin{aligned} \dot{y}_{1,1} &= \dot{y}_{1,1}^d - k_{p,1}(y_{1,1} - y_{1,1}^d) \Rightarrow \\ (\dot{y}_{1,1} - \dot{y}_{1,1}^d) + k_{p,1}(y_{1,1} - y_{1,1}^d) &= 0 \Rightarrow \\ z_1 + k_{p,1}z_1 &= 0 \end{aligned} \quad (27)$$

Equivalently, by substituting Equation 25 into Equation 18, and using the definition $y_{1,2} - a_1 = z_2$ one has

$$\begin{aligned} \dot{y}_{1,2} &= \dot{a}_1 - k_{p,2}(y_{1,2} - a_1) \Rightarrow \\ (\dot{y}_{1,2} - \dot{a}_1) + k_{p,2}(y_{1,2} - a_1) &= 0 \Rightarrow \\ z_2 + k_{p,2}z_2 &= 0 \end{aligned} \quad (28)$$

Similarly, continuing with the rest of the equations of the state-space model and by substituting Equation 24 into Equation 17, while also using the definition $y_{1,3} - a_2 = z_3$ one has

$$\begin{aligned} \dot{y}_{1,3} &= \dot{a}_2 - k_{p,3}(y_{1,3} - a_2) \Rightarrow \\ (\dot{y}_{1,3} - \dot{a}_2) + k_{p,3}(y_{1,3} - a_2) &= 0 \Rightarrow \\ z_3 + k_{p,3}z_3 &= 0 \end{aligned} \quad (29)$$

Moving backwards, and by substituting Equation 23 into Equation 16 of the state-space model of the PDE dynamics, and using the definition $y_{1,i} - a_{i-1} = z_i$ one has

$$\begin{aligned} \dot{y}_{1,i} &= \dot{a}_{i-1} - k_{p,i}(y_{1,i} - a_{i-1}) \Rightarrow \\ (\dot{y}_{1,i} - \dot{a}_{i-1}) + k_{p,i}(y_{1,i} - a_{i-1}) &= 0 \Rightarrow \\ z_i + k_{p,i}z_i &= 0 \end{aligned} \quad (30)$$

By substituting Equation 21 into Equation 14 of the state-space model of the PDE dynamics, and using the definition $y_{1,N-1} - a_{N-2} = z_{N-1}$ one has

$$\begin{aligned} \dot{y}_{1,N-1} &= \dot{a}_{N-2} - k_{p,N-1}(y_{1,N-1} - a_{N-2}) \Rightarrow \\ (\dot{y}_{1,N-1} - \dot{a}_{N-2}) + k_{p,N-1}(y_{1,N-1} - a_{N-2}) &= 0 \Rightarrow \\ z_{N-1} + k_{p,N-1}z_{N-1} &= 0 \end{aligned} \quad (31)$$

Finally, by substituting Equation 20 into Equation 13, and using the definition $y_{1,N} - a_{N-1} = z_N$ one obtains

$$\begin{aligned} \dot{y}_{1,N} &= \dot{a}_{N-1} - k_{p,N}(y_{1,N} - a_{N-1}) \Rightarrow \\ (\dot{y}_{1,N} - \dot{a}_{N-1}) + k_{p,N}(y_{1,N} - a_{N-1}) &= 0 \Rightarrow \\ z_N + k_{p,N}z_N &= 0 \end{aligned} \quad (32)$$

Thus, the dynamics of the closed-loop system becomes

$$\begin{aligned} z_1 + k_{p,1}z_1 &= 0 \\ z_2 + k_{p,2}z_2 &= 0 \\ z_3 + k_{p,3}z_3 &= 0 \\ \dots\dots\dots \\ z_i + k_{p,i}z_i &= 0 \\ \dots\dots\dots \\ z_{N-1} + k_{p,N-1}z_{N-1} &= 0 \\ z_N + k_{p,N}z_N &= 0 \end{aligned} \quad (33)$$

The dynamics of the closed-loop system can be also written in matrix form

$$\dot{\tilde{Z}} + K_p \tilde{Z} = 0 \quad (34)$$

where

$$\tilde{Z} = [z_1, z_2, z_3, \dots, z_i, \dots, z_{N-1}, z_N],$$

and

$$K_p = \text{diag}[k_{p_1}, k_{p_2}, k_{p_3}, \dots, k_{p_i}, \dots, k_{p_{N-1}}, k_{p_N}].$$

After suitable selection of the coefficients k_{p_i} , $i = 1, 2, \dots, N$ such that the monomials $p(s) = s + k_{p_i}$ to have a negative root, it can be assured that $\lim_{t \rightarrow \infty} z_i(t) = 0$ and that the closed-loop system is asymptotically stable. Moreover, to prove asymptotic stability for the closed-loop system, the following Lyapunov function can be used

$$V_L = \sum_{i=1}^N \frac{1}{2} z_i^2 \quad (35)$$

The derivative of this Lyapunov function with respect to time is given by

$$\begin{aligned} \dot{V}_L &= \sum_{i=1}^N z_i \dot{z}_i \Rightarrow \dot{V}_L = \sum_{i=1}^N z_i (-k_{p_i} z_i) \Rightarrow \\ \dot{V}_L &= -\sum_{i=1}^N k_{p_i} z_i^2 < 0 \end{aligned} \quad (36)$$

Thus, it is proven again that the closed-loop system is globally asymptotically stable.

THE PROBLEM OF BOUNDARY CONTROL OF THE MULTI-ASSET BLACK-SCHOLES PDE

A multi-asset option price model is considered next. In such a case, the dynamics of the options' value is described by the multi-asset Black-Scholes PDE (Wade et al., 2007; Moon & Kim, 2013; Lotstedt et al., 2007; Moon et al., 2006; Martins-Vaquero et al., 2014). The chapter treats also the problem of feedback control and stabilization of the aforementioned PDE. As already noted, methods for feedback stabilization of systems with nonlinear PDE dynamics have been a flourishing research subject in the last years (Balogh & Kristic, 2002; Bensoussan et al., 2006; Basseville & Nikiforov, 1993; Boussaada et al., 2013; Smyshlyaev & Krstic, 2010; Boskovic et al., 2002; Liu, 2003). In particular, feedback control of diffusion-type (parabolic) PDEs has been a subject of extensive research and several remarkable results have been produced (Maidi & Corriou, 2014; Zwart et al., 2011; Woitteneck & Mounier, 2010; Mounier et al., 2010). As already pointed out, for the control of diffusion PDEs, boundary and distributed control methods have been developed (Fliess & Mounier, 2002; Rigatos, 2015a; Winkler & Lohmann, 2010). By showing the feasibility of control of the multi-asset Black-Scholes PDE it is also proven that through a selected trading policy, the price of options can be made to converge and stabilize at specific reference values. By applying semi-discretization and the finite differences method, the multi-asset Black-Scholes PDE model is written in a state-space form (Olivier & Sedoglavic, 2001; Kroner, 2011). This state-space description stands for a differentially flat system which means that all its state variables and control inputs can be written as differential functions of the flat output vector (Bououden et al., 2011; Lévine, 2009; Lévine, 2011). One can note several results on the use of differential flatness theory in the

control of PDEs (Rudolph, 2003; Rigatos, 2013; Rigatos, 2015b). The first stage in the proposed control approach for the multi-asset Black-Scholes PDE is to decompose the state-space description of the PDE into an equivalent set of nonlinear ODEs (Guo & Billings, 2007; Pinsky, 1991; Utz et al., 2011; Laroche, 2000). Next by examining independently each nonlinear ODE it is shown that this stands again for a differentially flat system, for which a virtual control input can be computed as in the case of flatness-based control for the trivial system. The virtual control input is chosen such that the ODE subsystem dynamics is linearized and the tracking error is eliminated. The boundary condition that appears in the nonlinear ODE subsystem that comprises the last row of the state-space description, stands again for the aggregate control input.

The computation of the boundary control input uses recursively all virtual control inputs mentioned above, moving from the last ODE system to the first one. This stands for implementation of flatness-based control in successive (cascading) loops. Thus, by tracing the rows of the state-space model backwards, at each iteration of the control algorithm, one can finally obtain the control input that should be applied to the multi-asset Black-Scholes PDE system so as to assure that all its state vector elements will converge to the desirable setpoints. The stability of the control loop is proven in two manners. First, convergence to zero is proven for the tracking error of all subsystems into which the PDE's state-space model is decomposed. Next, with the use of Lyapunov analysis it is reconfirmed that this control scheme is asymptotically stable.

BOUNDARY CONTROL OF THE MULTI-ASSET BLACK-SCHOLES PDE

Next, the multi-asset Black-Scholes PDE is introduced (Martins-Vaquero et al., 2014; Wade et al., 2007):

$$\frac{\partial V}{\partial t} = \sum_{i=1}^N \sum_{j=1}^N \rho \sigma_i \sigma_j S_i S_j \frac{\partial^2 V}{\partial S_i \partial S_j} + \sum_{i=1}^N r S_i \frac{\partial V}{\partial S_i} - rV \quad (35)$$

Moreover, without loss of generality the two-asset Black-Scholes PDE is considered

$$\frac{\partial V}{\partial t} = \frac{1}{2} \sigma_1^2 S_1^2 \frac{\partial^2 V}{\partial S_1^2} + \frac{1}{2} \sigma_2^2 S_2^2 \frac{\partial^2 V}{\partial S_2^2} + \rho \sigma_1 \sigma_2 S_1 S_2 \frac{\partial^2 V}{\partial S_1 \partial S_2} + r S_1 \frac{\partial V}{\partial S_1} + r S_2 \frac{\partial V}{\partial S_2} - rV \quad (36)$$

Semi-discretization and the finite differences method is applied. To this end the partial derivatives appearing in Equation 36 are computed as follows:

$$\frac{\partial V}{\partial S_1} = \frac{V(S_{1,i+1}, S_{2,j}) - V(S_{1,i}, S_{2,j})}{\Delta S_1} \quad (37)$$

$$\frac{\partial^2 V}{\partial S_1^2} = \frac{V(S_{1,i+1}, S_{2,j}) - 2V(S_{1,i}, S_{2,j}) + V(S_{1,i-1}, S_{2,j})}{\Delta S_1^2} \quad (38)$$

$$\frac{\partial V}{\partial S_2} = \frac{V(S_{1,i}, S_{2,j+1}) - V(S_{1,i}, S_{2,j})}{\Delta S_2} \quad (39)$$

$$\frac{\partial^2 V}{\partial S_2^2} = \frac{V(S_{1,i}, S_{2,j+1}) - 2V(S_{1,i}, S_{2,j}) + V(S_{1,i}, S_{2,j-1})}{\Delta S_2^2} \quad (40)$$

Box 3.

A

$$\frac{\partial^2 V}{\partial S_1 \partial S_2} = \frac{V(S_{1,i+1}, S_{2,j+1}) - V(S_{1,j+1}, S_{2,j}) - V(S_{1,i}, S_{2,j+1}) + V(S_{1,i}, S_{2,j})}{\Delta S_1 \Delta S_2} \quad (41)$$

Using the previous semi-discretization, for grid point (i, j) it holds

$$\begin{aligned} \frac{\partial V(S_{1,i}, S_{2,j})}{\partial t} &= \frac{1}{2} \sigma_1^2 S_{1,i}^2 \left[\frac{V(S_{1,i+1}, S_{2,j}) - 2V(S_{1,i}, S_{2,j}) + V(S_{1,i-1}, S_{2,j})}{\Delta S_1^2} \right] + \\ &\frac{1}{2} \sigma_2^2 S_{2,j}^2 \left[\frac{V(S_{1,i}, S_{2,j+1}) - 2V(S_{1,i}, S_{2,j}) + V(S_{1,i}, S_{2,j-1})}{\Delta S_2^2} \right] + \\ &\rho \sigma_1 \sigma_2 S_{1,i} S_{2,j} \left[\frac{V(S_{1,i+1}, S_{2,j+1}) - V(S_{1,i+1}, S_{2,j}) - V(S_{1,i}, S_{2,j+1}) + V(S_{1,i}, S_{2,j})}{\Delta S_1 \Delta S_2} \right] + \\ &r S_{1,i} \left[\frac{V(S_{1,i+1}, S_{2,j}) - V(S_{1,i}, S_{2,j})}{\Delta S_1} \right] + r S_{2,j} \left[\frac{V(S_{1,i}, S_{2,j+1}) - V(S_{1,i}, S_{2,j})}{\Delta S_2} \right] - r V(S_{1,i}, S_{2,j}) \end{aligned} \quad (42)$$

and in Equation 41 (shown in Box 3).

The boundary conditions of the PDE are taken to be $V_{i,0} \neq 0$ only if $i = 1$, $V_{0,j} \neq 0$ only if $j = 1$ and $V(i, j) = ct$ (constant) if $i > N$ or $j > N$.

Considering that $i = 1, 2, \dots, N$ and $j = 1, 2, \dots, N$ the option's values at the grid points (i, j) are denoted as $V_{i,j}$. Using this notation, the semi-discretized model of the PDE takes the following form: At grid point $i = 1$ and $j = 1$

$$\begin{aligned} \frac{\partial V_{1,1}}{\partial t} &= \frac{1}{2} \sigma_1^2 S_{1,1}^2 \left[\frac{V_{2,1} - 2V_{1,1} + V_{0,1}}{\Delta S_1^2} \right] + \\ &+ \frac{1}{2} \sigma_2^2 S_{2,1}^2 \left[\frac{V_{1,2} - 2V_{1,1} + V_{1,0}}{\Delta S_2^2} \right] + \\ &\rho \sigma_1 \sigma_2 S_{1,1} S_{2,1} \left[\frac{V_{2,2} - V_{2,1} - V_{1,2} + V_{1,1}}{\Delta S_1 \Delta S_2} \right] + \\ &+ r S_{1,1} \left[\frac{V_{2,1} - V_{1,1}}{\Delta S_1} \right] + r S_{2,1} \left[\frac{V_{1,2} - V_{1,1}}{\Delta S_2} \right] - r V_{1,1} \end{aligned} \quad (43)$$

At grid point $i > 1$ and $j > 1$ it holds

$$\begin{aligned} \frac{\partial V_{i,j}}{\partial t} &= \frac{1}{2} \sigma_1^2 S_{1,i}^2 \left[\frac{V_{i+1,j} - 2V_{i,j} + V_{i-1,j}}{\Delta S_1^2} \right] + \\ &+ \frac{1}{2} \sigma_2^2 S_{2,j}^2 \left[\frac{V_{i,j+1} - 2V_{i,j} + V_{i,j-1}}{\Delta S_2^2} \right] + \\ &\rho \sigma_1 \sigma_2 S_{1,i} S_{2,j} \left[\frac{V_{i+1,j+1} - V_{i+1,j} - V_{i,j+1} + V_{i,j}}{\Delta S_1 \Delta S_2} \right] + \\ &+ r S_{1,i} \left[\frac{V_{i+1,j} - V_{i,j}}{\Delta S_1} \right] + r S_{2,j} \left[\frac{V_{i,j+1} - V_{i,j}}{\Delta S_2} \right] - r V_{i,j} \end{aligned} \quad (44)$$

Next, the following state vector variables are defined $x_{(i-1)N+j} = V_{i,j}$, $i = 1, 2, \dots, N$ and $j = 1, 2, \dots, N$. The system's state vector becomes

$$x = [x_1, x_2, \dots, x_{(i-1)N+(j-1)}, x_{(i-1)N+j}, \dots, x_{N^2}]^T$$

and comprises N^2 elements. Using this notation of state variables Equation 81 becomes

$$\begin{aligned}
 \dot{x}_1 = & \frac{1}{2} \sigma_1^2 S_{1,1}^2 \left[\frac{x_{N+1} - 2x_1}{\Delta S_1^2} \right] + \\
 & \frac{1}{2} \sigma_2^2 S_{2,1}^2 \left[\frac{x_2 - 2x_1}{\Delta S_1^2} \right] + \\
 & + \rho \sigma_1 \sigma_2 S_{1,1} S_{1,2} \left[\frac{x_{N+2} - x_2 - x_{N+1} + x_1}{\Delta S_1 \Delta S_2} \right] + \\
 & r S_{1,1} \left[\frac{x_{N+1} - x_1}{\Delta S_1} \right] + r S_{2,1} \left[\frac{x_2 - x_1}{\Delta S_2} \right] + \\
 & \left[\frac{1}{2} \sigma_1^2 S_1^2 \frac{V_{0,1}}{\Delta S_1^2} + \frac{1}{2} \sigma_2^2 S_2^2 \frac{V_{1,0}}{\Delta S_2^2} \right]
 \end{aligned} \tag{45}$$

Thus, by defining the control input associated with the boundary conditions as $u = [V_{0,1}, V_{1,0}]^T$ and one obtains a description for Equation 45 in the form

$$\dot{x}_1 = f_1(x) + c_1 u \tag{46}$$

Equivalently for Equation 44 one obtains Equation 47 (shown in Box 4).

Equation 48 is finally written in the form

$$\dot{x}_{(i-1)N+j} = f_{(i-1)N+j}(x) + c_{(i-1)N+j} x_{(i-1)N+(j-1)} \tag{49}$$

Considering that $i = 1, 2, \dots, N$ and $j = 1, 2, \dots, N$ there are N^2 state-space equations. Thus, the dynamics of the PDE model is written as

$$\begin{aligned}
 \dot{x}_{N^2} = & f_{N^2}(x) + c_{N^2} x_{N^2-1} \\
 \dot{x}_{N^2-1} = & f_{N^2-1}(x) + c_{N^2-1} x_{N^2-2} \\
 & \dots \dots \dots \\
 \dot{x}_{(i-1)N+j} = & f_{(i-1)N+j}(x) + c_{(i-1)N+j} x_{(i-1)N+(j-1)} \\
 & \dots \dots \dots \\
 \dot{x}_2 = & f_2(x) + c_2 x_1 \\
 \dot{x}_1 = & f_1(x) + c_1 u
 \end{aligned} \tag{50}$$

FLATNESS-BASED CONTROL OF THE MULTI-ASSET BLACK-SCHOLES PDE

First, it can be proven that the state-space description of the multi-asset Black-Scholes PDE, given in Equation 50, is a differentially flat one, with flat output $y = x_{N^2}$ (Rigatos & Siano, 2016). Solving the i -th row of the state space model, where $i = 1, 2, \dots, N^2$, with respect to x_{i+1} one finds that state variables x_{i+1} is a differential function of the flat output y . Moreover, from the last row of Equation 50 it holds that u is a function of the flat output and its derivatives. Next, the following virtual control inputs are defined

$$\begin{aligned}
 a_1 = & x_{N^2-1}, a_2 = x_{N^2-2}, \dots, \\
 a_{N^2-(i-1)N-(j-1)} = & x_{(i-1)N+(j-1)}, \dots, a_{N^2-1} = x_1
 \end{aligned} \tag{51}$$

Using the virtual control inputs of Equation 51 in the state-space model of Equation 50 one gets

$$\begin{aligned}
 \dot{x}_{N^2} = & f_{N^2}(x) + c_{N^2} a_1 \\
 \dot{x}_{N^2-1} = & f_{N^2-1}(x) + c_{N^2-1} a_2 \\
 & \dots \dots \dots \\
 \dot{x}_{(i-1)N+j} = & f_{(i-1)N+j}(x) + c_{(i-1)N+j} a_{N^2-(i-1)N-(j-1)} \\
 & \dots \dots \dots \\
 \dot{x}_2 = & f_2(x) + c_2 a_{N^2-1} \\
 \dot{x}_1 = & f_1(x) + c_1 u
 \end{aligned} \tag{52}$$

By examining independently each nonlinear ODE of the previous state-space description of Equation 52 and by defining as local flat output for the i -th ODE the state variable x_i it can be shown that the i -th row of the state-space description stands again for a differentially flat system. Actually, one has now N^2 subsystems, each one of them related to a row of the state-space model and the local flat outputs for these subsystems are

Box 4.

A

$$\begin{aligned}
\dot{x}_{(i-1)N+j} = & \frac{1}{2} \sigma_1^2 S_{1,i}^2 \left[\frac{x_{iN+j} - 2x_{(i-1)N+j} + x_{(i-2)N+j}}{\Delta S_1^2} \right] + \\
& + \frac{1}{2} \sigma_2^2 S_{2,j}^2 \left[\frac{x_{(i-1)N+(j+1)} - 2x_{(i-1)N+j} + x_{(i-1)N+(j-1)}}{\Delta S_2^2} \right] + \\
& + \rho \sigma_1 \sigma_2 S_{1,i} S_{2,j} \left[\frac{x_{iN+(j+1)} - x_{(i-1)N+(j+1)} - x_{iN+j} + x_{(i-1)N+j}}{\Delta S_1 \Delta S_2} \right] + \\
& + r S_{1,i} \left[\frac{x_{iN+j} - x_{(i-1)N+j}}{\Delta S_1} \right] + r S_{2,j} \left[\frac{x_{(i-1)N+(j+1)} - x_{(i-1)N+j}}{\Delta S_2} \right] - r x_{(i-1)N+j}
\end{aligned} \tag{47}$$

Equation 47 can be also written in the form

$$\begin{aligned}
\dot{x}_{(i-1)N+j} = & \frac{1}{2} \sigma_1^2 S_{1,i}^2 \left[\frac{x_{iN+j} - 2x_{(i-1)N+j} + x_{(i-2)N+j}}{\Delta S_1^2} \right] + \\
& + \frac{1}{2} \sigma_2^2 S_{2,j}^2 \left[\frac{x_{(i-1)N+(j+1)} - 2x_{(i-1)N+j}}{\Delta S_2^2} \right] + \\
& + \rho \sigma_1 \sigma_2 S_{1,i} S_{2,j} \left[\frac{x_{iN+(j+1)} - x_{(i-1)N+(j+1)} - x_{iN+j} + x_{(i-1)N+j}}{\Delta S_1 \Delta S_2} \right] + \\
& + r S_{1,i} \left[\frac{x_{iN+j} - x_{(i-1)N+j}}{\Delta S_1} \right] + r S_{2,j} \left[\frac{x_{(i-1)N+(j+1)} - x_{(i-1)N+j}}{\Delta S_2} \right] - r x_{(i-1)N+j} + \\
& + \left[\frac{1}{2} \sigma_2^2 S_{2,j}^2 \frac{1}{\Delta S_2^2} \right] x_{(i-1)N+(j-1)}
\end{aligned} \tag{48}$$

$$Y = [x_1, x_2, \dots, x_{(i-1)N+j}, \dots, x_{N^2-1}, x_{N^2}] \tag{53}$$

From the i -th row of the state-space model it can be seen that the virtual control input α_i is a differential function of the local flat output x_i , which shows again that the i -th subsystem, if independently examined, is also differentially flat.

The virtual control input for the i -th row of the state space model is chosen such that the ODE subsystem dynamics is linearized and the tracking error is eliminated. The boundary condition that appears in the nonlinear ODE subsystem that comprises the last rows of the state-space description, was used as the aggregate control input (Rigatos & Siano, 2016).

One can find the values that the virtual control inputs should have, so as to eliminate the tracking error for each one of the subsystems that are

obtained from the per-row decomposition of Equation 52. For the first row of Equation 52 one has

$$a_1^* = \frac{1}{c_{N^2}} \left[\dot{x}_{N^2} - f_{N^2}(x) - k_{p_1} (x_{N^2} - x_{N^2}^*) \right] \tag{54}$$

with $k_{p_1} > 0$, while it also holds that $a_1^* = x_{N^2-1}^*$. Continuing with the second row of Equation 52, the associated virtual control input can be obtained:

$$a_2^* = \frac{1}{c_{N^2-1}} \left[\dot{x}_{N^2-1} - f_{N^2-1}(x) - k_{p_2} (x_{N^2-1} - x_{N^2-1}^*) \right] \tag{55}$$

with $k_{p_2} > 0$, while it also holds that $a_2^* = x_{N^2-2}^*$. Continuing in a similar manner, for the i -th row

of Equation 52, the associated virtual control input is

$$a_{N^2-(i-1)N-(j-1)}^* = \frac{1}{c_{(i-1)N+j}} [\dot{x}_{(i-1)N+j} - f_{(i-1)N+j}(x) - k_{p_{N^2-(i-1)N-(j-1)}} (x_{(i-1)N+j} - x_{(i-1)N+j}^*)] \quad (56)$$

with $k_{p_{N^2-(i-1)N-(j-1)}} > 0$ where

$$a_{N^2-(i-1)N-(j-1)}^* = x_{(i-1)N+(j-1)}^*$$

By applying the same procedure, the virtual control input for the $N^2 - 1$ row of Equation 52 is found

$$a_{N^2-1}^* = \frac{1}{c_2} [\dot{x}_2 - f_2(x) - k_{p_{N^2-1}} (x_2 - x_2^*)] \quad (57)$$

Finally, from the N^2 -th row of Equation 53 one computes the boundary control input that is really exerted on the system

$$u = \frac{1}{c_1} [\dot{x}_1 - f_1(x) - k_{p_{N^2}} (x_1 - x_1^*)] \quad (58)$$

where $a_{N^2}^* = u$. Using the previous definitions, the virtual control inputs can be written as

$$a_1^* = \frac{1}{c_{N^2}} [\dot{x}_{N^2} - f_{N^2}(x) - k_{p_1} (x_{N^2} - x_{N^2}^*)] \quad (59)$$

$$a_2^* = \frac{1}{c_{N^2-1}} [\dot{x}_{N^2-1} - f_{N^2-1}(x) - k_{p_2} (x_{N^2-1} - a_1^*)] \quad (60)$$

$$a_{N^2-(i-1)N-(j-1)}^* = \frac{1}{c_{(i-1)N+j}} [\dot{x}_{(i-1)N+j} - f_{(i-1)N+j}(x) - k_{p_{N^2-(i-1)N-(j-1)}} (x_{(i-1)N+j} - a_{N^2-(i-1)N-(j-1)-1}^*)] \quad (61)$$

$$a_{N^2-1}^* = \frac{1}{c_2} [\dot{x}_{N^2-1} - f_{N^2-1}(x) - k_{p_{N^2-1}} (x_{N^2-1} - a_{N^2-2}^*)] \quad (62)$$

$$u = \frac{1}{c_1} [\dot{x}_1 - f_1(x) - k_{p_{N^2}} (x_1 - a_{N^2-1}^*)] \quad (63)$$

Stability analysis for the case of the multi-asset Black-Scholes PDE of Eq. (52) is performed as in the case of the single-asset Black-Scholes PDE. This was explained in Section “*Closed Loop Dynamics of the single-asset Black-Scholes PDE*”. In the multi-asset Black-Scholes PDE one demonstrates that the N^2 state-vector elements converge to the associated reference values. Moreover the associated Lyapunov function will be a sum of N^2 terms (instead of N).

SIMULATION TESTS

Simulation results about the proposed control method for distributed parameter systems and for the case of the single-asset Black-Scholes PDE (BSPDE) are depicted in Figure 1 (a). The spatial discretization of the single-asset PDE model consisted of $N = 25$ points. The sampling period was chosen to be T_s . The boundary condition V_{N+1} of the PDE was taken to be known and constant. Simulation examples for the case of the multi-asset Black-Scholes PDE are depicted in

Figure 1 (b) and in Figure 2(a) and 2(b). The spatial discretization of the PDE model consisted of $N = 16$ points. The sampling period was chosen to be again $T_s = 0.01\text{sec}$. The boundary condition $V_{0,1}$ served as the control input, while the boundary condition $V_{1,0}$ was set equal to zero.

The numerical simulation experiments have confirmed the theoretical findings of this chapter. It has been shown that by applying the proposed control method the single-asset and multi-asset Black-Scholes PDE dynamics can be modified so as to converge to the desirable reference profile. The control input that succeeds this, requires the recursive computation of the virtual control inputs which were defined in Section “*Computation of boundary control for the single-asset Black-Scholes PDE*” and in Section “*Flatness-based control of the multi-asset Black-Scholes PDE*”. The accuracy of tracking of the reference setpoints was indeed satisfactory. It is noted that the application of the proposed Black-Scholes PDE feedback control method is not limited to the case of option pricing but can be also extended to other financial models and indexes with spatiotemporal dynamics.

FUTURE RESEARCH DIRECTIONS

There is a wide class of problems in finance related to PDE models (Boucekine et al., 2013). There can be for instance, capital mobility and growth problems or commodities pricing problems (e.g. pricing of consumption goods, agricultural products, mining products, oil, electric power, services, etc.). In particular, it has been shown that commodities’ price dynamics is equivalent to option price dynamics, with the spot price and the convenience yield of the commodity to have a role equivalent to the one of the underlying assets that appear in the Black-Scholes PDE (Pirrong, 2014; Swartz, 1998). Therefore, one can also apply PDE control methods based on differential flatness theory for the stabilization of commodity prices. It is noted that the differential flatness theory-based control methods are of proven convergence and stability, whereas the convergent solution of optimal control problems in finance associated with the Hamilton-Jacobi-Bellman PDE can be assured only in specific cases.

Figure 1. (a) single-asset BSPDE: variation in time of the solution of the PDE at grid point x_N (blue line) and associated reference setpoint (dashed red line) (b) multi-asset BSPDE: Tracking of reference setpoint No 1 (dashed red line) by the value of the PDE system (blue line) at the final grid point $V_{N,N}$

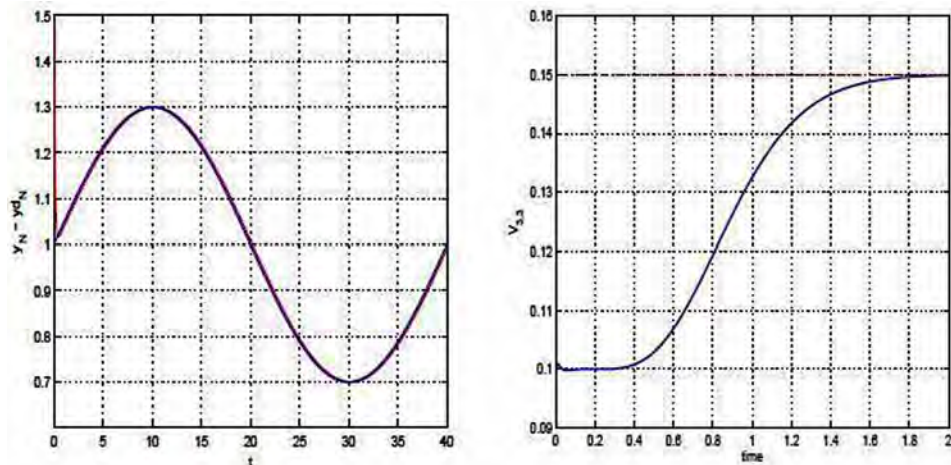
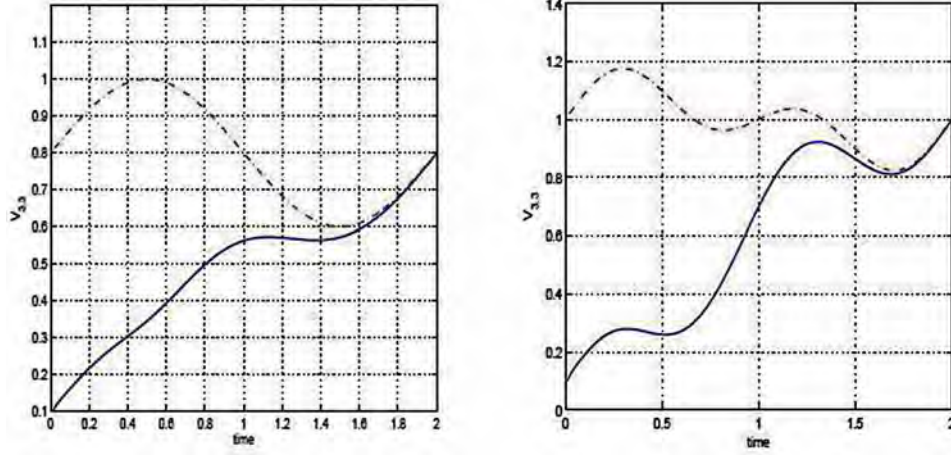


Figure 2. (a) multi-asset BSPDE: Tracking of reference setpoint No 2 (dashed red line) by the value of the PDE system (blue line) at the final grid point $V_{N,N}$ (b) multi-asset BSPDE: Tracking of reference setpoint No 3 (dashed red line) by the value of the PDE system (blue line) at the final grid point $V_{N,N}$



CONCLUSION

First, a boundary control method for the single-asset Black-Scholes PDE has been introduced, aiming at developing an approach for the stabilization of options dynamics. Following a semi-discretization method and a finite differences scheme, the Black-Scholes PDE model has been decomposed into an equivalent set of nonlinear ordinary differential equations (ODEs) and a states-space model has been obtained. Next, it has been proven that each one of the aforementioned ODEs stands for a differentially flat subsystem. This enables to compute for each ODE subsystem, a virtual control input which linearizes its dynamics and eliminates the associated output's tracking error. From the state equations that constitute the last subsystem one can find the boundary condition that also stands for the control input to the single-asset Black-Scholes PDE model. Second, a feedback control and stabilization scheme has been developed for the multi-asset Black-Scholes PDE. It has been shown that by applying again semi-discretization and the finite differences method, the multi-asset Black-Scholes PDE can be written in a state-space form for which differential flatness properties hold. Next, the state-space description

of the system was decomposed into an equivalent set of nonlinear ODEs and a control algorithm consisting of successive (cascading) loops was developed.

To compute the boundary control input of the single-asset, as well as of the multi-asset, Black-Scholes PDE model one has to use recursively all virtual control inputs which are applied to the previously mentioned ODE subsystems. The control input of i -th subsystem becomes a reference setpoint for its successive $(i+1)$ -th subsystem. Thus computation of control inputs moves progressively from the last ODE system to the first one. Consequently, by tracing the rows of the state-space model backwards, at each iteration of the control algorithm, one can finally obtain the control input that should be applied to the single-asset, as well as to the multi-asset, Black-Scholes PDE so as to assure that all its state vector elements will converge to the desirable setpoints. By analyzing the dynamics of the closed loop system that results from the application of the aforementioned control method, asymptotic stability is confirmed.

Finally, numerical simulation experiments have been provided about the application of the proposed control method to the model of both the single-asset and the multi-asset Black-Scholes

PDE. These performance tests have shown the control method's accuracy and reliability. The method can be of interest for trading of options in open electricity markets as well as for the pricing of commodities. By showing the feasibility of such a control method it is also proven that through a selected trading policy the price of the negotiated option can be made to converge and stabilize at specific reference values.

REFERENCES

- Balogh, A., & Kristic, M. (2000). Infinite dimensional backstepping style feedback transformations for a heat equation with an arbitrary level of instability. *European Journal of Control*, 8(2), 165–175. doi:10.3166/ejc.8.165-175
- Barmish, B. R. (2011). On performance limits of feedback control-based stock-trading strategies. *Proceedings of the American Control Conference. 2011 American Control Conference, ACC 2011*.
- Basseville, M., & Nikiforov, I. (1993). Detection of abrupt changes: Theory and applications. Prentice Hall.
- Bensoussan, A. (2000). Quelques remarques sur le prix des options, avec prise en compte de contraintes. In *Décision prospective auto-organisation*. Paris: Dunod.
- Bensoussan, A., Prato, G. D., Delfour, M. C., & Mitter, S. K. (2006). *Representation and control of infinite dimensional systems*. Boston: Birkhäuser.
- Bernhard, P. (2005). *The robust control approach to option pricing and interval models: An overview*. In N. Breton & H. Ben-Ameur (Eds.), *Numerical methods in finance* (pp. 91–108). Springer. doi:10.1007/0-387-25118-9_4
- Boskovic, D. M., Krstic, M., & Liu, W. J. (2002). Boundary control of an unstable heat equation via measurement of domain averaged temperature. *IEEE Transactions on Automatic Control*, 46(12), 2022–2028. doi:10.1109/9.975513
- Boucekkine, R., Camacho, C., & Fabbri, G. (2013). *On the optimal control of some parabolic partial differential equations arising in economics*. Document de Recherche EPEE. Centre d'Etudes des politiques économiques de l'université d'Evry.
- Bououden, S., Boutat, D., Zheng, G., Barbot, J. P., & Kratz, F. (2011). A triangular canonical form for a class of 0-flat nonlinear systems. *International Journal of Control, Taylor and Francis*, 84(2), 261–269. doi:10.1080/00207179.2010.549844
- Boussaada, I., Cela, A., Mounier, H., & Niculescu, S. I. (2013). *Control of Drilling Vibrations: A time-delay system-based approach. 11th Workshop on Time Delay Systems*, France. doi:10.3182/20130204-3-FR-4031.00162
- Dannehl, J., & Fuchs, F. W. (2005). Flatness-based control of an induction machine fed via voltage source inverter - Concept, control design and performance analysis. *IEEE Conference on Industrial Electronics*, 5125–5130.
- Dragonescu, A., & Soane, A. M. (2013). Multigrid solution of a distributed optimal control problem constrained by the Stokes equations. *Applied Mathematics and Computation*, Elsevier, 219(10), 5622–5634. doi:10.1016/j.amc.2012.11.070
- Fedorov, S., & Mikhailov, S. (2001). Option pricing for incomplete markets via stochastic optimization: Transaction costs, Adaptive Control and Forecast. *Intl. Journal of Theoretical and Applied Finance*, World Scientific, 1(1), 179–195. doi:10.1142/S0219024901000912
- Fliess, M., & Join, C. (2010). Delta hedging in financial engineering: towards a model-free approach. *IEEE MED 2010, 18th Mediterranean Conference on Control and Automation*. doi:10.1109/MED.2010.5547847
- Fliess, M., & Join, C. (2012). Preliminary remarks on option pricing and dynamic hedging. *IEEE ICSCS 2012, 1st Intl. Conference on Systems and Computer Science*. doi:10.1109/ICSCS.2012.6502450

- Fliess, M., & Mounier, H. (1999). Tracking control and -freeness of infinite dimensional linear systems. In *Dynamical Systems, Control, Coding and Computer Vision* (vol. 258, pp. 41-68). Birkhauser.
- Fliess, M., & Mounier, H. (2002). An algebraic framework for infinite-dimensional linear systems. In *Proc. of International School on Automatic Control of Lille, "Control of Distributed Parameter Systems*. Grenoble, France: Theory and Applications.
- Forsyth, P. A., & Labahn, G. (2007). Numerical methods for controlled Hamilton-Jacobi-Bellman PDE in finance. *Journal of Computational Finance*, 11(2), 1-43. doi:10.21314/JCF.2007.163
- Gerdt, M., Greif, G., & Pesch, H. J. (2008). Numerical Optimal Control of the Wave Equation: Optimal Boundary Control of a String to Rest in Finite Time. *Mathematics and Computers in Simulation, Elsevier*, 79(4), 1020-1032. doi:10.1016/j.matcom.2008.02.014
- Guo, L., & Billings, S. A. (2007). State-Space Reconstruction and Spatio-Temporal Prediction of Lattice Dynamical Systems. *IEEE Transactions on Automatic Control*, 52(4), 622-632. doi:10.1109/TAC.2007.894513
- Kroner A. (2011). Adaptive finite element methods for optimal control second order hyperbolic equations. *Computational Methods in Applied Mathematics*, 11(2), 214-240.
- Laroche, B., Martin, P., & Rouchon, P. (2000). Motion planning of the heat equation. *International Journal of Robust and Nonlinear Control. J. Wiley*, 40(8), 629-643. doi:10.1002/1099-1239(20000715)10:8<629::AID-RNC502>3.0.CO;2-N
- Levine, J. (2009). *Analysis and Control of Nonlinear Systems: A flatness-based approach*. Springer. doi:10.1007/978-3-642-00839-9
- Levine, J. (2011). On necessary and sufficient conditions for differential flatness, *Applicable Algebra in Engineering. Communications and Computing, Springer*, 22(1), 47-90.
- Liu, W. J. (2003). Boundary stabilization of an unstable heat equation. *SIAM Journal on Control and Optimization*, 42(3), 1033-1043. doi:10.1137/S0363012902402414
- Lotstedt, P., Persson, J., von Sydov, L., & Tysk, J. (2007). Space-time adaptive finite difference method for European multi-asset options. *Intl. Journal of Computers and Mathematics with Applications*, 33(8), 1159-1180. doi:10.1016/j.camwa.2006.09.014
- Maidi, A., & Corriou, J. P. (2014). Distributed control of nonlinear diffusion systems by input-output linearization. *International Journal of Robust and Nonlinear Control*, 26(3), 389-405. doi:10.1002/rnc.2892
- Martins-Vaquero, J., Khaliq, A.Q.M., & Kleefeld, B. (2014). Stabilized explicit Range-Kutta methods for multi-asset American options. *Computers and Mathematics with Applications*, 67, 1293-1308.
- Miller, N., & Weller, P. (1995). Stochastic saddlepoint systems stabilization policy and the stock-market. *Journal of Economic Dynamics and Control, Elsevier*, 19(1-2), 279-302. doi:10.1016/0165-1889(93)00782-Y
- Moon, K. S., & Kim, H. (2013). A multi-dimensional loop average lattice method for multi-asset models. *Quantitative Finance*, 13(6), 873-884. doi:10.1080/14697688.2012.744086
- Moon, K. S., Kim, W. J., & Kim, H. (2008). Adaptive lattice methods for multi-asset methods. *Journal of Computer and Mathematics with Applications*, 56(2), 352-366. doi:10.1016/j.camwa.2007.12.008

- Mounier, H., Rudolph, J., & Woittenneck, F. (2010). Boundary value problems and convolutional systems over rings of ultradistributions. *Lecture Notes in Control and Information Sciences*, 179-188.
- Oksendal, B., & Sulem, A. (2006). *Applied stochastic control of jump diffusions*. Springer.
- Olivier, F., & Sedoglavic, A. (2002). A generalization of flatness to nonlinear systems of partial differential equations: Application to the control of a flexible rod. *Proceedings of the 5th IFAC Symposium on Nonlinear Control Systems*.
- Pascucci, A. (2011). *PDE and Martingale Methods in Option Pricing*. Springer. doi:10.1007/978-88-470-1781-8
- Pinsky, M. (1991). *Partial differential equations and boundary value problems*. Prentice-Hall.
- Pirrong, G. (2014). *Commodity Price Dynamics: A structural approach*. Cambridge University Press.
- Platen, E., & Heath, D. (2008). *A benchmark approach to quantitative finance*. Springer.
- Rigatos, G. (2011). *Modelling and control for intelligent industrial systems: adaptive algorithms in robotics and industrial engineering*. Springer. doi:10.1007/978-3-642-17875-7
- Rigatos, G. (2013). *Advanced models of neural networks: nonlinear dynamics and stochasticity in biological neurons*. Springer.
- Rigatos, G. (2014). A Kalman Filtering Approach for Detection of Option Mispricing in the Black-Scholes PDE Model. *IEEE CIPER 2014 Computational Intelligence Conference on Financial Engineering*. doi:10.1109/CIFEr.2014.6924098
- Rigatos, G. (2014). A Kalman Filtering approach to the detection of option mispricing in electric power markets. *IEEE SSCI 2014*, Orlando, FL. doi:10.1109/CIASG.2014.7011554
- Rigatos G. (2014). *Parameter change detection in financial options models using neural networks with invariance to Fourier transform*. ISI Internal Report.
- Rigatos, G. (2015). *Stabilization of option price dynamics through feedback control of the Black-Scholes PDE*. IFAC MICNON 2015, 1st IFAC Conference on Modelling, Identification and Control of Nonlinear Systems, St. Petersburg, Russia. doi:10.1016/j.ifacol.2015.09.248
- Rigatos, G. (2015). *Nonlinear control and filtering using differential flatness approaches: applications to electromechanical systems*. Springer. doi:10.1007/978-3-319-16420-5
- Rigatos, G., & Siano, P. (2002). DFIG control using Differential flatness theory and Extended Kalman Filtering. *IFAC INCOM 2012, 14th IFAC Intl. Conference on Information Control Problems in Manufacturing*.
- Rigatos, G., & Siano, P. (2012). Sensorless nonlinear control of induction motors using Unscented Kalman Filtering. *IEEE IECON 2012, 38th Intl. Conference of the Industrial Electronics Society*.
- Rigatos, G., & Siano, P. (2013). Power corporations default probability forecasting using the Derivative-free nonlinear Kalman Filter. *IEEE IECON 2015*, Cambridge, UK.
- Rigatos, G., & Siano, P. (2015). *Stabilization of the multi-asset Black-Scholes PDE using differential flatness theory*. ISI Internal Report 5.
- Rigatos, G., & Siano, P. (2016). Feedback control of the multi-asset Black-Scholes PDE using differential flatness theory. *Journal of Financial Engineering*.
- Rudolph, J. (2003). *Flatness Based Control of Distributed Parameter Systems, Examples and Computer Exercises from Various Technological Domains*. Aachen: Shaker Verlag.

- Schwartz, E. (1998). Valuing long-term commodity assets. *Financial Management*, 27(1), 57–66.
- Sira-Ramirez, H., & Agrawal, S. (2004). *Differentially Flat Systems*. New York: Marcel Dekker.
- Sircar, K., & Papanicolaou, G. (1998). *General Black-Scholes models accounting for increased market volatility from hedging strategies*. *Applied Mathematics*. Taylor and Francis.
- Smyshlyaev, A., & Krstic, M. (2010). *Adaptive control of parabolic PDEs*. Princeton University Press. doi:10.1515/9781400835362
- Stojanovic, S. (2007). Stochastic control approach to derivative pricing and hedging in incomplete markets modeled by general Itô SDE systems: an overview and an application in FX derivatives. *Proceedings of the 2007 American Control Conference*. doi:10.1109/ACC.2007.4283153
- Utz, T., Meurer, T., & Kugi, A. (2011). *Trajectory planning for two-dimensional quasi-linear parabolic PDE based on finite difference semi-discretization*. *18th IFAC World Congress*, Milano, Italy.
- Wade, B. A., Khaliq, A. Q. M., Yousuf, M., Vigo-Aguiar, J., & Deininger, R. (2007). On smoothing of the Crank-Nicolson scheme and higher-order schemes for pricing barrier options. *Journal of Computational and Applied Mathematics*, Elsevier, 204(1), 144–158. doi:10.1016/j.cam.2006.04.034
- Windcliff, H., Forsyth, P. A., & Vetzal, K. R. (2006). Analysis of the stability of the linear boundary condition for the Black-Scholes equation. *Journal of Computational Finance*, 8(1), 65–92. doi:10.21314/JCF.2004.116
- Winkler, F. J., & Lohmann, B. (2010). Design of a decoupling controller structure for first order hyperbolic PDEs with distributed control action. *2010 American Control Conference*, Baltimore, MD. doi:10.1109/ACC.2010.5530561
- Woittenneck, F., & Mounier, H. (2010). Controllability of Networks of spatially one-dimensional second order PDEs - An algebraic approach. *SIAM Journal on Control and Optimization*, 48(6), 3882–3902. doi:10.1137/08072437X
- Yin, G., Yu, J., & Zhang, Q. (2010). *Option Pricing, Model Calibration, and Prediction with a Switchable Market: A Stochastic Approximation Algorithm*. *49th IEEE Conference on Decision and Control*, Atlanta, GA. doi:10.1109/CDC.2010.5717667
- Zwart, H., Le Gorrec, Y., & Maschke, B. (2011). *Linking hyperbolic and parabolic PDEs*. 2011 50th IEEE Conference on Decision and Control and European Control Conference, CDC-ECC, Orlando, FL.

ADDITIONAL READING

- Bhar, R. (2010). *Stochastic Filtering with applications in finance*. World Scientific. doi:10.1142/7736
- Chen, P., & Islam, S. (2005). *Optimal control problems in finance: A new computational approach*. Springer.
- Cossin, D., & Aparicio Acosta, F. (2001). *Optimal Control of Credit Risk*. Kluwer. doi:10.1007/978-1-4615-1393-3
- Ma, J., & Wohar, S. (2015). *Recent advances in estimating nonlinear models: with applications to economics and finance*. Springer.
- Shadbolt, J., & Taylor, J. (2002). *Neural networks and the financial markets: Predicting, combining and portfolio optimization*. Springer. doi:10.1007/978-1-4471-0151-2
- Zeng, Y., & Wu, S. (2013). *State-space models: Applications in economics and finance*. Springer. doi:10.1007/978-1-4614-7789-1

KEY TERMS AND DEFINITIONS

Asset: A parameter in financial models which denotes a possession or a resource which can be evaluated and traded. Assets can take the form of financial derivatives (such as options and futures) or commodities or the form of equipment or other proprietary holdings.

Black-Scholes PDE: A diffusion partial differential equation which describes the dynamics of option prices. It computes the distribution of the option price as a function of time and an underlying asset variable. It can be dependent on one single asset or on multiple assets.

Boundary Control: An approach to the control of partial differential equations in which the control action is exerted to the PDE through its boundary conditions. This is different to distributed or pointwise control of PDEs, in which the control action is exerted at several points of the system's state space.

Differential Flatness Theory: A primary research direction in the area of nonlinear dynamical systems control. It considers that instead of describing the system's dynamics through its entire state vector, one can use for this purpose specific algebraic variables which are called flat outputs and which are dependent only on certain elements of the state vector. Differential flatness theory enables to succeed global linearization for complicated nonlinear dynamics and in this manner to solve the associated control and state estimation problems.

Distributed Parameter System: A dynamical system that evolves not only in time but also in space. Otherwise stated the system exhibits spatiotemporal dynamics along the time axis and along one or more spatial axes. Systems described

by partial differential equations are distributed parameter ones.

Feedback Control: The action of applying an external excitation to a dynamical system which is dependent on the value of the system's state vector and on the deviation of this state vector from a reference value that is called setpoint.

Lyapunov Function: This is an energy function of the system which depends on quadratic terms of the system's state vector error. It takes positive values apart from the equilibrium where it becomes zero. A system is stabilized when the associated Lyapunov function becomes zero.

Options: Financial derivatives which generate a secondary value for traded assets (commodities, goods, services etc.). Options are related to long term agreements about the exploitation of the traded assets so they do not reflect only the spot price of the assets but they also show the ability to accomplish the contract term's.

Pricing: The procedure of defining the price of traded assets, services or commodities. Spot pricing is an on-site agreement about the price of the traded resource based on offer and demand. There can be more elaborated pricing schemes taking place in longer time intervals (trading of financial options or commodities) which apart from offer and demand are also dependent on supporting services and procedures for the trading transaction (e.g. ability to store and transport goods, availability of equipment, credibility of the trading parts etc.).

Stability: A property of a dynamical system denoting that the state vector of the system converges to a specific point in the state space, which is called equilibrium or to a bounded region in the state-space which is called domain of attraction and remains there.

Does Inter-Bank Investments Restraints Financing Performance of Islamic Banks?

Mohammad Taquiuddin Mohamad

University of Malaya, Malaysia

Munazza Saeed

University of Malaya, Malaysia

INTRODUCTION

Interbank markets proved to play a crucial role in propagating the distress during the recent financial crisis. Unsecured financing determines clear links between creditors and debtors, stating explicitly the risk relation (Leur, 2016). If a debtor defaults, the lender's risk materializes and she has to bear the losses. Since the beginning of the financial crisis, the interbank market has been carefully scrutinized by commentators and policy-makers. It is also considered as crucial stress indicators during financial crises: they reveal not only banks' concerns regarding to the risk of financing of their counterparts, but also regarding their own liquidity needs. Accordingly, the impact of the monetary-policy measures followed by Malaysia Islamic banks, the need to look long term effect involving financing capability. Disentangling financing and liquidity effects has essential policy implications. On the one hand, if a rise in spreads reflects poor liquidity, policy measures should aim at improving funding facilities. On the other hand financing concerns should be addressed by enhancing debtors' solvency (Buigut, 2010). This question has been of utmost importance in this industry over the last few years, where most of the interbank operations conducted by Central Bank of Malaysia were designed to reduce interbank market stress. This article seeks to examine the involvement of Malaysia Islamic banks in Islamic inter-bank money market (IIMM) investment within a financing framework in dual banking

system in Malaysia. This paper also focuses on effects bank specifications, changes in monetary policy and economic environment on financing behaviour of Malaysia Islamic banks.

The article is divided into five parts. The second section describes background and performance of financing activities in Islamic banks in Malaysia. The third section, examines some previous researches and articles related with this topic. In fourth section, model and data specification adopted used in this article. The fifth section is the finding of the research. Section five and six look at the future research direction and conclusion.

BACKGROUND

Since Islamic banking was established in Malaysia in 1983, financing growth rate showed an encouraging performance. However, that performance was not consistent over the operation period of 27 years. Since there are circumstances where it is affected in economy situation. The evidence is in 1987 and 1998 when the global economy was suffering from the recession. At that time Islamic banking financing decreased from 18.36 percent to 2.68 percent. However, this situation improved and the amount of Islamic financing continued to rise for the next year since the world economy got in good condition and stable.

Table 1 shows the use of the Malaysia Islamic banking funds in aggregate in the form of several types of financing. In reference to the table, the

Table 1. Financing by type/sector

Type of Financing			2006	2007	2008	2009	2010
			RM Million				
Overdraft			2731.0	3,278.0	3,740.1	4,203.5	4,446.7
Term financing	Hire purchase	Total	26567.7	29,208.8	33,570.7	39,164.1	44,959.0
		where: Passenger car	23127.2	25,422.9	29,154.8	36,498.9	41,569.3
	Leasing		376.5	505.2	331.1	760.4	875.4
	Financing based on block		0.0	0.0	0.0	0.0	0.0
	Connector financing		369.2	465.9	384.4	413.5	397.6
	Syndicate financing		1199.1	772.2	521.8	2,504.4	2,061.1
	Factoring		0.0	0.0	0.0	0.0	0.0
	Personal financing		4526.4	6,001.0	8,484.0	11,727.3	15,540.2
	Home financing		16403.0	17,036.6	18,940.8	22,728.3	29,792.6
	Others		14027.9	17,764.8	23,882.5	34,453.2	43,181.6
	Due time which:	Until 1 year	1255.6	2,096.6	4,289.4	4,183.1	3,878.4
		Exceeding 1 year	53460.3	60,720.5	75,642.8	104,596.9	127,185.1
Financing Bill			9164.2	10,291.8	10,070.0	8,056.9	7,881.8
Trustworthy receipt			512.4	571.3	728.2	652.8	664.3
Revolve credit			2117.3	2,079.6	3,058.5	5,268.8	6,230.4
Financing in foreign currencies			327.4	841.6	2,628.2	3,132.9	3,956.9
Others			1004.6	1,050.8	1,381.6	1,907.6	2,425.0
Total			80460.5	89,867.6	107,721.8	134,973.5	162,412.6

Source: National Bank of Malaysia (2006-2010)

Islamic banking financing flows are broken down into some type of overdraft financing; term financing, bill financing, trust receipts, revolving credit in foreign currencies, and the rest is represented by other available financing. During the five years of 2006 to 2010, the direction of financing flow of Malaysia Islamic banking increased around 11 to 25 percent each year with the latest financing in December 2010 making RM162,412.6 million. The majority of the total financing is contributed by the term financing type that covered financing such as leasing, financing by block, syndicate,

factoring, private financing, home financing and others. Meanwhile the least type of financing demand by the client is the trustworthy receipt¹ which only represents around 0.4 per cent from the total financing every year (Mahmood, 1997).

It should be noted, though Islamic banking offers the client interest-free financing, but the reality is the Islamic banking is still facing various form of risks especially credit risk which involves the capability of the bank to offer more financing. When this situation occurs, the bank will strive to reduce the financing volume in the future where

this will directly lead to the fall in bank's profit. Apart from credit risk, the other example of risk that could not be taken lightly is the interest risk rate which involves the existence of substitution effect where client switches to conventional bank financing due to the decrease in interest rate. This causes the financing cost in conventional bank to be much lower compared to financing cost of Islamic banking (Kader R. A., 2009).

Literature Review

In this section, the authors put forth some of past research that touched directly and indirectly on the bank's loans/financing behaviour relating to the bank's specifications, the economic environment and also the market. The review covers both local and international researchers. Hassan (1993) studied the loan sales activities of commercial banks in the United States that refers to the risks vulnerability in the capital market. To achieve the set objectives, six bank specification variables were included which are the credit variables, interest rate and also business variables. Based on five markets of risk measurement, this study found out that specification and loan expansion have a positive connection with all forms of market risk.

Hatakeda (2000) studied the banks loan in Japan under liquidity constraint from year 1975 to 1995. Researchers use estimation method of Ordinary Least Squares (OLS) towards bank sample data as well as other methods such as Augmented Dickey Fuller (ADF) unit roots test and co integration relation between variables. This study is successful in finding the empirical evidence of the existence of third regime in any bank sample that is financing under their liquidity constraint. From this regime, both land price index and bank capital have positive and great effect on bank financing. On the other hand, the call rate and economy activity (Real GDP) have a negative effect. This study also discovered how the liberation and regulation on bank capital are also influenced by bank financing behaviour.

Cebenoyan, A.S; Strahan, P.E (2004) studies on how domestic commercial bank in United States manages credit exposed risk through sales loan affect the bank capital structure, loan, benefits and also the risk from 1987 to 1994. This study use series of cross section estimation method to measure the usage of loan sales market for risk management. The study discovered bank samples balancing the financing portfolio exposure through loan trading, where bank use loan sales market as a platform in risk management and not for changing loan principal, having less capital from other bank. They produce riskier financing (*financing business*) as percentage to total asset compared to other banks. The study also discovered sample bank had a low risk and high profit compared to other bank. The study concluded that sophisticated risk management practice in banking affair increase the readiness bank credit but not fully decrease bank² risk.

Atunbas Y, Gambacorta L and Marques D (2007) studied the effect of drastic increase in security activity on banks credit offered in Europe. In this matter, the security is found to change the function of credit market by reducing fundamental liquidity role which played by financial mediator. Besides that, the variation of bank role from its original function had changed the bank capability in offering credit and bank financing channel efficiency for finance principle. This study found that the security usage protects bank financing offer from finance principle effect. In addition, bank security activity also concretes the bank capacity to offer new loan, but this capacity is dependent on business cycle condition and risk position.

Hazli and Ismail (2008) studied the effect on Malaysia Islamic banking involvement in securitisation activity towards loan/financing offer and risk tolerance level. Theoretically, securitisation activities will decrease the degree of bank avoidance risk. Thus, banks are motivated to increase the percentage of assets with a concentration of the risk through the granting of loans to economy sectors. This study refers to Islamic commercial

banks in Malaysia for years 1994 to 2004 by using the panel data analysis. The findings show significant securitisation activity cut the growth of financing. This indicates that securitisation is the replacement to the Malaysia Islamic banking financing. In addition, this study supports the *moral hazard* hypothesis where the bank involved in the securitisation will reduce financing with least risk and could do financing on riskier one.

Altunbas, Gambacorta, and Marques (2009) did an empirical research on the bank borrowing channel in accordance with European Community perspective which emphasis the role of financial innovation. This study used a dynamic model with panel data refer omg to the data of the quarter of year 1999 to the fourth quarter of 2008 in 643 banks throughout Europe³. As a result, the study found a significant relationship exists between low interest rates with risk taking by the bank, where the bank's actions have an impact on the risk of the bank. Besides that, the study also found that monetary policy is only partially neutral on monetary policy stability. Apart from that, the study also found that the banks have different lending channel especially those referring to size and credit reaction on GDP while the bank's capital is not involved in the shock transmission.

Rahman (2009) studied the relationship between the loans structure and market risk exposure of banks in Malaysia by adapting an estimation of unbalanced panel data to the eleven banks from the year of 1994 to 2006. In this study, the influence of the loans structure is analysed using four main measurements which are the properties lending, specialization index, short term loans stability and long term stability loans. In conclusion, the study found a phase of loan structure influence the level of high market risk when the financial crisis in 1997 and the year after. Besides that, bank and investors should take the effect of additional loan structure into consideration for the effect of loan expansion and management efficiency when the bank is exposed to the risk.

Sarantis and Nicholas (2009) did an empirical research on the bank loan channel for financial transmission in 8 countries of CEE (Central and

Eastern Europe), namely the Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, the Republic of Slovakia and Slovenia who have joined the European Union. In this study, researchers tested whether changes in monetary policy affect bank loans which differ in the size of the bank, capital strength, liquidity and structure of ownership. To achieve the objective of the study, researchers used a method of estimating dynamic panel on bulk of the data bank panel in year 1994 to 2003. As a result, this study found that the size of the bank and liquidity has the most significant role in differentiating the reactions of banking financial policy changes. The study also looked at macroeconomic effects through bank borrowings and found evidence that linking the aggregate would lend to economic activities in CEE countries.

Kader and Leong (2009) review the impact of interest rate changes on the demand for Islamic financing in the dual system of banking in Malaysia. In their analysis, the researchers using time series on political Unit Root Test, Co integration, Vector Autoregressive (VAR), Granger Causality and Impulse Response Function (IRF) on monthly data provided by Bank Negara Malaysia (BNM) from 1999 to 2007. Several variables are included such as the total number of conventional banking financing of residential properties, the number of real estate financing Islamic banking and the base lending rate (base rate was BLR). The conclusion in this study is any increase in base lending rate (BLR) will encourage users to obtain financing from Islamic banking and vice versa.

Buigut (2010) evaluates the importance of the bank's loans financial policy transmission channel in Kenya. Loan theory stated that a strengthen financial which affect demand aggregate by moving loan supply bend to the left. This reduction may be due to the stress of the demand for loans through conventional interest rate. In explaining this ambiguity, the researchers tested loan channels by using vector error correction models (VECM) to identify supply and demand curve shifts in the market for bank loans. With this method, identification of existing problems

based on aggregate data are presented clearly. Impulse method responses also used to analyse the effects of financial policy shocks on the quantity of loans, loan rates and true output. The study found that banks in Kenya have to master all loan channels that exist.

Oliver, María Pía; Yuan Li; Jeon, Bang Nam (2010) reviewed how competition between banks influences the monetary policy transmission through loans channel of commercial banks in 10 Asian countries and 10 countries of Latin America. This study used the procedure of two-step estimation data panel from year 1996 to 2006. The first stage of the estimation is to measure the level of competition between the banks by adopting the methodology done by previous researchers. Then the next step is to estimate the equation in loan growth and included the independent variables of the bank competitor. Estimation results provided consistent evidence that the increased competition in the banking sector will weaken financial transmission policy through bank borrowings. This is true especially for banks in Latin America and small-sized banks with low liquidity and capitalisation.

Based on the above studies, researcher found a specific study related to financing institutions behaviour that it is still dominated by studies involving conventional banking and Islamic banking context where the study on these issues has yet to be done in a comprehensive manner. Existing study on Islamic banking financing done by (Kader & Leong, 2009), (Kassim & Majid, 2010) and (Rahman, 2009) are only from a different perspective and leaving literature gap that can still be filled by more research.

Model and Data Specification

The present study is based mainly on secondary data. The data and information have been collected from the publications of the National Bank of Malaysia; Handbook of statistics on Malaysia Economy and Annual reports & other valuable publications of public sector, 17 Islamic

banks in Malaysia. The period covered under the study is from 1983 to 2014. However, the period varies according to the nature of subject dealt with and availability of data. The bank-specific determinants selected for assessing impact on financing are past financing, profit, risk, capital and size of respective banks. The monetary policy changes for examining the impact on financing are money supply, Islamic Interbank Investment and investment in government securities. While the economic environment for examining impact on Islamic banks financing growth domestic product, inflation rate and index of economic freedom.

Financing model in this article adopted the approach that has been done by earlier researchers such as Hassan (1993), Hatakeda (2000), Cebenoyan, A.S and Strahan, P.E (2004), Sarantis and Nicholas (2009) by using bank annual data for viewing the response of institutional banks in offering financing. Based on the evaluation, the study proposes the following model specifications as the basis to conduct financing of Islamic banking:

$$\begin{aligned} \Delta TF_{it} = & \beta_1 + \beta_2 \Delta TF_{it-1} + \beta_3 profit_{it} + \\ & \beta_4 risk_{it} + \beta_5 cap_{it} + \beta_6 size_{it} + \beta_7 \Delta M3_t \\ & + \beta_8 iibr_t + \beta_9 mgs_t + \beta_{10} \Delta gdp_t + \beta_{11} cpi_t \\ & + \beta_{11} econfree_t + \varepsilon_{it} + u_i \end{aligned}$$

$i = 1, 2, \dots, N$ (bank amount)

$t = 1, 2, \dots, T$ (period)

where μ_{it} is the fixed effect of time, u_i is fixed effect firm and $\varepsilon_{i,t}$ is the errors term which are not serially correlate nor it correlate with all variable at time $t-1$. This study defines TF_{it} as level of financing offered by Islamic banking that covers every economic sub-sector from year 1994 to 2014. Due to the lag variable (TF_{it-1}) is independent variables in this study, the above specifications of the model developed become inconsistent. Therefore, Arellano, M; Bond, S (1991), recommended the use of estimation method of GMM which is more effective and consistent.

Table 2. Descriptive of variables statistics

Variable	Mean	Std. Dev.	Skewness	Kurtosis	Jarque-Bera
ΔTF_{it-1}	13.50	2.95	-1.97	8.80	377.44*
$profit_{it}$	0.01	0.01	-0.12	11.39	511.63*
$risk_{it}$	0.46	0.23	-0.10	2.37	3.33
cap_{it}	41.50	364.83	9.81	101.21	76069.98*
$size_{it}$	14.54	2.11	-0.37	4.58	24.76*
$\Delta M3_t$	13.13	0.41	-0.20	2.26	8.05*
$iibr_t$	4.06	1.91	1.19	2.97	64.81
mgs_t	3.78	1.53	0.98	2.43	47.25*
Δgdp_t	12.83	0.40	0.14	1.95	13.34*
cpi_t	2.70	1.37	0.48	2.41	14.49*
$econfree_t$	64.60	3.63	0.40	2.05	17.36*

Note: *Significant at 5% **Significant at 1% *Significant at 10%

FINDINGS

1. Descriptive Analysis

The descriptive analysis to view statistics data and variables used in the model study was formed. Some of the statistics used in determining the statistical behaviour of variables are the mean, median, standard deviation, skewness, kurtosis and Jarque-Bera. Min refers to the average value of each variable for the whole samples, while the standard deviation shows the variation of data from the mean value.

Referring to Table 2, the variable specification sample which Islamic banking has a value of cap_{it} while the largest mean has a value of $profit_{it}$ mean the smallest and least disperse in terms of the distribution of data. The summary of the statistics also showed that the total Islamic banking specification data sample skew to the left except cap_{it} which skew to the right. Kurtosis value shows the value exceeds its normal distribution where data distribution is leptokurtic shape and cap_{it} had the highest peak of 76069.98 whereas $risk_{it}$ record as the lowest value of 2.3744 that is close to normal data distribution.

For monetary policy changes variables, the distribution data shows the average value of a dispersion of $\Delta M3_t$ was the highest with the value

13.13 and mgs_{it} indicate the average value of the lowest at 3.78. The distribution data for $iibr_{it}$ and mgs_{it} is a skew to the right with the exception of data for mgs_{it} that skew to the left with 0.20. Thus the value of kurtosis for $\Delta M3_t$, $iibr_t$ dan mgs_t , all approaching the value of normal distribution respectively of 2.26, 2.97 and 2.43.

The variable of economic cycles, $econfree_t$, data indicated that the highest dispersion of data distribution is 64.60 with the value of the standard deviation of 3.63. On the other hand, the data for the variable $inflation_{it}$ indicates the lowest data distribution that is the value of 64.600 with the value of the standard deviation of 3.63. Thus the distribution data for Δgdp_t , cpi_t , dan $econfree_t$ indicates the data has skewed to the right. Kurtosis value of Δgdp_t data has less than the normal distribution which is 1.95 while the data for cpi_t and $econfree_t$ approaching the value of normal distribution with the value of each is 2.41 and 2.05 respectively.

Next test of *jarque-Bera* test was carried out to evaluate whether the data used are scattered normally or not. The results are that all data for variables used significantly at the significance level of 0.05 per cent but with the exception of data for the variable $iibr_t$. These results show that almost all the data used in this study are not scattered normal.

Table 3. Matrix correlation

	ΔTF_{it-1}	$profit_{it}$	$risk_{it}$	cap_{it}	Δgdp_t	$\Delta M3_t$	$iibr_t$	cpi_t	mgs_{it}	$size_{it}$	$econfree_{it}$
ΔTF_{it-1}	1.0000										
$profit_{it}$	-0.1869	1.0000									
$risk_{it}$	0.4898	-0.0440	1.0000								
cap_{it}	-0.5432	0.2131	-0.2320	1.0000							
Δgdp_t	0.5481	-0.1016	0.1406	-0.1168	1.0000						
$\Delta M3_t$	0.5399	-0.1032	0.1414	-0.1070	0.9950	1.0000					
$iibr_t$	-0.4071	0.1679	0.0957	0.1016	-0.5302	-0.5427	1.0000				
cpi_t	-0.0920	0.0200	0.0912	0.0896	0.0474	0.0010	0.5557	1.0000			
mgs_{it}	-0.4197	0.1655	0.0658	0.1062	-0.5536	-0.5719	0.9905	0.5689	1.0000		
$size_{it}$	0.8859	-0.1468	0.2070	-0.2399	0.6500	0.6429	-0.4935	-0.0921	-0.5058	1.0000	
$econfree_{it}$	-0.1750	0.1896	0.1372	0.1344	-0.1535	-0.1670	0.7518	0.5499	0.7410	-0.1953	1.0000

2. Analysis of Matrix Correlation

Table 3 above shows the result of matrix correlation between independent variables (ΔTF) and other independent variables. Correlation matrix above shows the variable representing the Islamic bank specification sample which are $profit_{it}$, cap_{it} , and $size_{it}$ in relation to inverse with financing Islamic banking level, while $risk_{it}$ on positive relation. For the relationship between financial policy variables and Islamic banking financing level, the variables of $iibr_t$ dan mgs_{it} are inversely related and on the other hand $\Delta M3_t$ variable is positively related. The economic cycle variable, namely cpi_{it} dan $econfree_{it}$ has negative correlation with the dependent variables, while financing Δgdp_t relating positively to 0.548 correlations.

3. Model Estimation

Table 4 shows the estimation of a dynamic model floor financing Islamic banking. The results showed that the Islamic banking financing behaviour in this study is determined by three main factors, namely, the bank specifics, monetary policy changes and economic environment.

a. Bank Specific

Result estimation shows that lag variable of financing (ΔTF_{it-1}) shows an increase in current levels of financing and this result confirms the dynamic specifications at the level of five percent. The higher level of past financing is influenced by prudent “monetary policy” which then bring an excellent repayment track record of the customers. In addition, the past financing flow have provided in many productive sectors who are capable of generating high returns to the institution which in turn promoting Islamic banking to offer more financing for the next year.

The results also show an increase of one per cent risk level, increased the level of bank financing by 2.2700 percent. This situation reflects the Islamic banking in changing their financing portfolio by moving towards a more risky financing as a reaction to highly decreasing profit from low-risk financing. Islamic banking is generally involved in various forms of financing such as property, consumer financing, commercial financing, industrial, and others where all of which are involved in the profile and the level of risk a particular credit risks as well as other unique risks. Besides that, the effect of their active engagement in the securitisation activity also makes the bank focus

Table 4. Model estimation result

Specification	Estimation Parameter	
	GMM-Difference	GMM-System
ΔTF_{it-1}	0.0871* (2.4755)	0.1046* (2.2779)
$profit_{it}$	-1.1897 (-0.4768)	1.2797 (0.3381)
$risk_{it}$	3.0248* (11.6600)	2.2700* (2.9838)
cap_{it}	-0.0019* (-15.9870)	-0.0020* (-15.6427)
$size_{it}$	0.7042* (8.6663)	0.7783* (8.8067)
$\Delta M3_t$	1.0876 (0.7368)	0.4277 (0.7483)
$iibr_t$	-0.5965* (-4.4083)	-0.3869* (-2.6096)
mgs_t	0.6559* (3.3014)	0.4755* (2.3540)
Δgdp_t	-0.5927 (-0.4413)	-0.0037 (-0.0061)
cpi_t	-0.0065 (-0.2271)	-0.0229 (-1.0957)
$econfree_t$	0.0070 (0.3458)	0.0085 (1.0097)
Sargan Test	126.3045*	3.7883*
AR(1)	-0.42	-0.11
AR(2)	-0.89	-0.77
*Significant at 5%	**Significant at 1%	*Significant at 10%
() t-value Sargan Test is referring to exceed limitation recognition		

on their assets are motivated financing risky on the factors the potential high returns (Hazli & Ismail, 2008).

In terms of capital structure, the study found an increase of one per cent of the capital structure of Islamic banking will limits the ability of Islamic banking in offering financing by -0.0020 per cent. This decision shows the Islamic banking financing and improved the provisions reducing the construction capital and this finding is consistent with research that has been done by Liu, Beng & Hua, Min, (2009) which linked the capital control of the bank (*bank capital regulations*) either a

binding or do not affect the conduct of the bank in offering financing. Rules for banks that have surplus capital, but is bound by the rules occasionally won't increase the offer of financing but also restrict the factors each financing offered resulted in an increase in risk-weighted assets ratio.

The study also shows $size_{it}$ in have a positive relation with financing with 0.7783 percent. This shows the size factor is as the main determinant of whether a bank will add financing or not. These findings are consistent with the findings of (Sarantis & Nicholas, 2009) who found the size of the bank has the most significant role in differentiating the reactions of banking monetary policy changes. (Oliver, María Pía; Yuan Li; Jeon, Bang Nam;, 2010) also found smaller banks with low liquidity and capitalisation with limited ability in offering financing/loans.

b. Monetary Policies Changes

In this section, study found only two variables, which is $iibr_t$ dan mgs_t shows the importance of the level of financing in Islamic banking. However there is a difference of the two variables where $iibr_t$ shown negative while positive sign from mgs_{it} . The negative relationship between the variable $iibr_t$ with the TF_{it} in this study showed an increase of one percent in the interbank money market investment will lead to Islamic banking reduce level of financing -0.3869 percent. These findings prove the investment activities between Islamic banks (*Islamic interbank investment*) has "substitution effect" and decreased their capability of financing because of their tendency to maintain liquidity (Othman, Ahmad, & Kechot, 1992)

While mgs_t variables showed an increase of one percent hold significant in government securities in increasing the level of financing Islamic banking of 0.4755 per cent. This shows the retention level of high government securities held by Islamic banking and sell them when enough maturity will result in high profit return as well. This returns then are distributed in the capital formation bank for the purpose of channel it into financ-

ing (ElGindi, Said, & Salevurakis, 2009). Next, growth of money supply variable ($\Delta M3_t$) indicates positive sign for financing level. These results are consistent with the theory of financing policy to link money growth increase was in line with the level of financing in bank institutions very closely with monetary policy expands. Although these results are consistent with the theory of money supply, but the study found the relationship that exists is insignificant.

c. Economic Environment

In this section, three variables included in this analysis did not indicate the importance for Islamic financing level even though the result is in line with theories and studies carried out previously. For example the variable Δgdp_t and cpi_t show negative sign while positive sign for $econfree_{it}$ variables. Insignificant of Δgdp_t in influencing financing decisions of Islamic banking in accordance with research acquired by (Hazli & Ismail, 2008) who found the Islamic banking more depending on the balance sheet indicators as a signal the economy condition compared with the general economic performance measurement.

On the other hand, one percent increase in $econfree_t$ found to increase Islamic banking financing by 0.0085%. This condition exists when economic units, especially traders and entrepreneurs have freedom and confidence doing business. Freedom means covering the freedom of doing business, trade, fiscal, investment, finance, corruption, labor and so on. The higher the level of it freedom will encourage economic units to get financing from Islamic banking which in turn contributed to Islamic banking profits (Habibullah; Sufian, Fadzman; Shah, Muzafar, 2010). On the other hand countries that have weak financial institutions such as the corruption and democracy will increase the financing problem (Boudriga, Taktak, & Jellouli, 2009)

FUTURE RESEARCH DIRECTIONS

This article showed that while the level of market domination by Islamic banks in Malaysia is still small, but the level of public confidence towards financing facilities offered by Islamic banks is always high and increases from year to year since the establishment of the first Islamic bank in 1983. This achievement can be proud of, but the strength of this credit creation they had does not mean arbitrary Islamic banking can generate any amount of financing desired restore options unless it is appropriate in order to avoid the risk of instability, especially related to subprime financing. Thus, this study presents some of the future research directions seen to conduct Islamic banking investment in interbank money market instruments:

1. Selection criterions of unproductive and high risky investment in money market instruments which will generate more profit to the bank.
2. Government's strategies through the implementation of fiscal and monetary policy as a responses of the unique features and capabilities of Islamic banking in offering financing. This research is to prevent the occurrence of shock at the industry that eventually resulted in the Islamic banking unable to react well. Failure to respond with good will ultimately upset Islamic banking into near instability in turn expose the bank to the various forms of credit risk specifically related.
3. Therefore, the future researchers can compare the patterns and performance of Islamic banking investment in interbank investment with conventional banking sector while taking other factors into account, particularly relating to fiscal developments in Malaysia.

CONCLUSION

In conclusion to, this study has discovered a pattern and behavior of the Malaysia Islamic banking financing. Based on the results, the Malaysia Islamic banking in proven extend financing taking into account the interaction between the bank's specific, changes in monetary policies and economic environment. Bank specifics involve factors of financing growth, the level of risk, vulnerability and capitalization of size. Refer to the monetary policy changes, the mechanisms of investment in government financial markets that the turned out to make an impact in determining the strength of Islamic banking credit. However, economic environment is not really a factor in influencing Islamic banking, where the Islamic banking in this study is more depended on the balance sheet indicator. It is a signal to the economic condition as compared to general measurement in performance of the economy. This study provided a limited information in relation to conduct Islamic banking in offering financing, especially in terms of samples, the use of the formulation and the variables included in the model construction research.

REFERENCES

- Altunbas, Gambacorta, & Marques. (2009). *Financial innovation, bank capital and the bank lending channel: A European empiricist's perspective*. Conference of Monetary Policy Transmission Mechanism in the euro area in its Kertas kerja Seminar Monetary Policy Transmission Mechanism in the euro area in its first 10 years, Frankfurt, Germany.
- Arellano, M., & Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies*, 58(2), 277–297. doi:10.2307/2297968
- Atunbas, Y., Gambacorta, L., & Marques, D. (2007). *Securitisation and the bank lending channel*. Working Paper Series, 838.
- Bacha, O. I. (2008). The Islamic inter bank money and a dual banking system: The Malaysian Experiences. *International Journal of Islamic and Middle Eastern Finance and Management*, 210–226.
- Banking supervision and nonperforming loans: a cross-country analysis. (2009). *Journal of Financial Economic Policy*, 1(4), 286–318.
- Boudriga, A., Taktak, N. B., & Jellouli, S. (2009). Banking supervision and nonperforming loans: a cross-country analysis. *Journal of Financial Economic Policy*, 286–318.
- Buigut, S. (2010). Is There a Bank Lending Channel of Monetary Policy in Kenya? *International Research Journal of Finance and Economics*, 45, 183–192.
- Cebenoyan, A. S., & Strahan, P. E. (2004). Risk management, capital structure and lending at banks. *Journal of Banking & Finance*, 28(1), 19–43. doi:10.1016/S0378-4266(02)00391-6
- Demetrio & Tovar-García. (2015). Exposure to interbank market and risk-taking by Mexican banks. *Cuadernos de Economía (Santiago, Chile)*.
- ElGindi, T., Said, M., & Salevurakis, J. W. (2009). Tamer ElGindi, Mona Said, John William Salevurakis. 2009. Islamic Alternatives to purely Capitalist Modes of Finance: A study of Malaysian Banks from 1999 to 2006. *The Review of Radical Political Economics*, 4(4), 516–538. doi:10.1177/0486613409341453
- Habibullah & Fadzlan. (2010). Does economic freedom fosters banks' performance? Panel evidence from Malaysia. *Journal of Contemporary Accounting & Economics*, 6, 77–91.
- Hassan, M. (1993). Capital market tests of risk exposure of loan sales activities of large US commercial banks. *Quarterly Journal of Business and Economic*, 32(1), 27–49.

- Hatakeda, T. (2000). Bank lending behaviour under a liquidity constraint. *Japan and the World Economy*, 12(2), 127–141. doi:10.1016/S0922-1425(99)00040-7
- Hazli, R., & Ismail, A. G. (2008). Does Islamic banks securitization involvement restrain their financing activity. *Humanomics*, 24(2), 95–109. doi:10.1108/08288660810876813
- Heider, F., Hoerova, M., & Holthausen, C. (2015). Liquidity hoarding and interbank market rates: The role of counterparty risk. *Journal of Financial Economics*, 118(2), 336–354. doi:10.1016/j.jfineco.2015.07.002
- Hoshi, T. (2006). *Creditor Rights and credit Creation by Banks in Transition Economies*. Working paper, Graduate School of International Relations and Pacific Studies University of California, San Diego.
- Ismail, S. (2009). *Pengurusan Bank Perdagangan di Malaysia*. Dewan Bahasa dan Pustaka.
- Kader, R. A. (2009). The impact of Interest rate changes on Islamic Bank Financing. *International Journal of Business Research Papers*, 5(3), 189–201.
- Kader & Leong. (2009). The impact of Interest rate changes on Islamic Bank Financing. *International Journal of Business Research Papers*, 5(3), 189–201.
- Leur, M. W. (2016). *Interbank loans, collateral and modern monetary policy*. ECB Working Paper 1959, European Central Bank.
- Liu, B., & Hua, M. (2009). Islamic banking: Interest-free or interest-based? *Pacific-Basin Finance Journal*, 17, 124–144.
- Mahmood, R. (1997). *Konsep Asas Perbankan*. Kuala Lumpur: Dewan Bahasa dan Pustaka.
- Naser, K., Jamal, A., & Al-Khatib, K. (1999). Islamic Banking: A study of customer satisfaction and preferences in Jordan. *International Journal of Bank Marketing*, 17(3), 135–150. doi:10.1108/02652329910269275
- National Bank of Malaysia. (2006-2010). *Financial Report*. Kuala Lumpur: National Bank of Malaysia.
- Oliver, M. P., Li, Y., & Jeon, B. N. (2010). Competition in banking and the lending channel: Evidence from bank-level data in Asia and Latin America. *Journal of Banking & Finance*, 35(3), 560–571. doi:10.1016/j.jbankfin.2010.08.004
- Othman, Z., Ahmad, J., & Kechot, M. (1992). *Ekonomi Kewangan*. Kuala Lumpur, Malaysia: Dewan Bahasa dan Pustaka.
- Podder, J., & al-Mamun, A. (2004). Loan loss provisioning system in Bangladesh banking: A critical analysis. *Managerial Auditing Journal*, 19(6), 729–740. doi:10.1108/02686900410543859
- Rahman, A. A. (2009). Lending Structure and Market Risk Exposures: The Malaysian Case. *Asian Academy of Management Journal*, 14(1), 1–20.
- Sarantis, & Nicholas. (2009). The bank lending channel and monetary transmission in Central and Eastern European countries. *Journal of Comparative Economics*, 37, 321–334.
- Toutounchian, I. (2009). *Islamic Money & Banking: Integrating Money in Capital Theory*. Singapore: John Wiley & Sons (Asia) Pte. Ltd.

ADDITIONAL READING

- Bacha, O. I. (2008). The Islamic inter bank money and a dual banking system: The Malaysian Experiences. *International Journal of Islamic and Middle Eastern Finance and Management*, 210-226.

EdgarDemetrio & Tovar-García. (2015). Exposure to interbank market and risk-taking by Mexican banks. *Cuadernos de Economía (Santiago, Chile)*.

Heider, F., Hoerova, M., & Holthausen, C. (2015). Liquidity hoarding and interbank market rates: The role of counterparty risk. *Journal of Financial Economics*, 118(2), 336–354. doi:10.1016/j.jfineco.2015.07.002

KEY TERMS AND DEFINITIONS

Bank Profit ($profit_{it}$): Measurement of profit before tax divided by total assets in bank. This variable indicates the number of bank profits to total assets.

Bank Risk ($risk_{it}$): Describe the results of risk-taking by banks in the appropriate timeliness. The dependence of these indicators suggest risk weights reflect the economic risks for different asset categories.

Bank Size ($size_{it}$): This ratio represents the ownership of assets by banks. High asset ownership enables banks to offer more financial services at low cost.

Capitalization (cap_{it}): Capital and reserves, as a part of the liabilities in the balance sheet total. This includes paid-up capital, reserve funds, retained earnings and other capital funds. Capital and reserves comprise own funds or a bank's core capital. More investment risk was so much more is needed capital.

Consumer Price Index (cpi_t): The relationship between the consumer price index or inflation with bank performance depends on whether inflation is expected (anticipated) or unexpected (unanticipated). In the second case (ie, inflation is not expected), the bank's actions in adjusting interest rates be the leading bank costs have increased more than the bank. This second type of inflation has a negative impact on bank profits, which in turn reduces the capital structure.

Economic Freedom Index ($econfree_t$): A ranking of countries or states based on the number and intensity of government regulations on wealth-creating activity. Metrics that an economic freedom index evaluates include international trade restrictions, government spending relative to GDP, occupational licensing requirements, private property rights, minimum wage laws and other government-controlled factors that affect people's ability to earn a living and keep what they earn. Such indexes are usually produced by economic think tanks.

Growth Domestic Product ($\Delta gdpt$): This is a key indicator of a country's macroeconomic management. Any changes in this indicator will change the loan/financing which in turn affects the adjustment capital ratio and bank risk observation for certain years.

Islamic Interbank Investment Rate ($iibr_t$): A short-term intermediary to provide a ready source of short-term investment outlets based on Syariah principle. Through the IIMM, the Islamic banks and banks participating in the Islamic Banking Scheme (IBS) would be able to match the funding requirements effectively and efficiently.

Malaysian Government Securities (mgs_t): Islamic securities that shows the loan by the government from financial institutions and others. Effectively it is a loan taken by the government of the people themselves. These loans are usually required by the state to finance recurrent expenditure and development expenditure for public projects.

Money Supply ($\Delta M3t$): Growth in money supply indicators show real growth potential, especially for future growth.

Total Financing (TF): This ratio shows the behavior of banks in the pursuit of profit and risk-taking. This behavior is consistent with profit-sharing paradigm that allows Islamic banking offer long-term financing to the project risk profile and high returns.

ENDNOTES

¹ Trust receipt is a document signed by the importer, confirming receipt of the shipping documents from the bank and allows the importer to take action on behalf of the bank to get the imported items of shipping companies. Upon receipt of the goods, the importer will be selling it and using the revenue to pay the bank.

² This study also found all of the bank specification variables are significant in relation to the risk. Capital and liquidity variables in negative relation to the risks, while credit variables in positive relation with risk.

³ Those countries are Belgium, Denmark, German, Greece, Finland, France, Ireland, Italy, Luxembourg, Netherlands, Portugal, Spain, Sweden, England and United States of America.

An Extension to the Delone and Mclean Information Systems Success Model and Validation in the Internet Banking Context

A

Veeraraghavan Jagannathan

National Institute of Technology, India

Senthilarasu Balasubramanian

National Institute of Technology, India

Thamaraiselvan Natarajan

National Institute of Technology, India

INTRODUCTION

Web-based applications in the recent years helps organizations to retain customers, and offering new services and products to them (DeLone & McLean, 1992; Tan & Teo, 2000). Internet Banking is considered as an online revolution of the traditional banking services which offers customers the greatest expediency for performing banking transactions via the Internet (Furst, Lang, & Nolle, 2000; Patnasingam, Gefen, & Pavlou, 2005). More precise definition of Internet Banking is given by Sathye (1999):

With the term electronic banking we consider all the possible transactions of a bank which are performed with the use of electronic means, mainly through Internet, but also through VPNs (Virtual Private Networks), Intranet, Extranet, phone and mobile phone, and these transactions do not necessitate that the customer must visit a branch.

There is a fundamental shift in banking delivery channels since mid-1990s (Pikkarainen, Pikkarainen, Karjaluoto, & Pahlila, 2004) and many banking executives perceived technology as the key solution for controlling costs (C.-P.

Lee, Mattila, & Shim, 2007). Internet Banking improves the bank's profit levels through the reduction of both variable and infrastructure costs, provides a source of differentiation and competitive advantage, provides global reach, adds another communication and feedback channel, increases customer satisfaction through the reduction of waiting times, thus improving service performance (Harridge-March, Wong, Rexha, & Phau, 2008). Internet Banking has appeared as the trend in banking, nowadays, and emerged as one of the payment models required to enable pure e-commerce models, rather than traditional banking (Zolait, 2010).

Some of the benefits to customers identified (Angelakopoulos & Mihiotis, 2011) are no time limitation, better time organization, no geographical limits, lower costs, 24 hour support, effortless accessibility for disabled people, integrated environment for Internet Banking transactions. In recent years, a large number of banks have started to adopt Internet Banking as an additional channel to reach and interact with clients. For financial institutions, Internet or Electronic banking is recognized as a tool that can significantly reduce their overhead costs as well as day-to-day expenses (Alhinai, Albadi, Alshihi, & Al-Gharbi, 2013).

BACKGROUND

Despite the recent advancements in internet security technologies such as, digital signatures, certificates, encryption algorithms, authentication mechanisms, consumers are still concerned about the security of monetary transactions over the internet (C. Yoon, 2010). In a report of Internet and Mobile Association of India (IAMAI -2010-11) it was found that people are hesitant to do banking transactions through the web sites of the bank, because of: security concerns (43 percent); preference for face-to-face transactions (39 percent); lack of knowledge about online transactions (22 percent); lack of user friendliness environment (10 percent); and lack of this facility in current bank (2 percent). M.-C. Lee (2009) found that the intention to use online banking is adversely affected mainly by the security/privacy risk. Hence, for the success of Internet Banking security plays a key role in customer trust of the website and satisfaction, which ultimately contribute to the success of the Information System (IS). Moreover, studies on IS Success related to Internet Banking are very scarce in the literature (Hoehle et al, 2012). Furthermore, the studies were conducted in developed countries and there is a paucity of studies in developing countries. This study intends to address the knowledge gap with the help of the proposed model. Therefore, the following research questions are framed based on the research gap:

1. What are the factors contributing to the success of Internet Banking?
2. What are the impacts security dimension make in the IB use and IB user satisfaction?

Finding the answers to the research questions can initiate improvement and enhance the performance of services provided via the electronic channel. It may also provide valuable feedback for the banks for satisfying the expectations of the bank customer, who intend to use IB in future. It is expected that this study will result in a re-evaluation of the IS Success model under new circumstances,

enhance understanding of consumer behaviors in correspondence with IB service, and provide suggestions for making sustainable IB usage. The next few sections form the literature survey of IB studies and relevant theories associated with the proposed framework.

LITERATURE REVIEW

The seminal work of Delone and Mclean (DeLone & McLean, 1992) paved the way for measuring IS success, which was elusive to researchers till then. Their (DeLone & McLean, 1992) paper proposed a six factor taxonomy in system quality, information quality, use, user satisfaction, individual impact, and organizational impact, using the multitude of measures existed in previous literatures. The authors also proposed temporal and causal relationship between the constructs.

The IS success model presumes that system quality and information quality, individually and jointly, affect user satisfaction and use. It also posited use and user satisfaction to be reciprocally interdependent, and presumes them to be direct antecedents of individual impact. In addition, the amount of use can affect the degree of user satisfaction either positively or negatively and vice versa. According to the IS Success model Individual impact should also lead to organizational impact.

DeLone and McLean (DeLone & McLean, 1992) characterize system quality as desired characteristics of the information system itself, and information quality as desired characteristics of the information product. They incorporate four scales from the Bailey-Pearson (Bailey & Pearson, 1983) instrument into system quality (convenience of access, flexibility of the system, integration of the system and response time) and nine scales into information quality: accuracy, precision, currency, timeliness, reliability, completeness, conciseness, format and relevance.

Delone and Mclean (Delone & McLean, 2003) came up with an update of their model, based on the research finding from their 1992 model. The

important change was the addition of the construct 'Service Quality' recommended by (Pitt, Watson, & Kavan, 1995). Second major change to their initial (Pitt et al., 1995) model was grouping of the two impacts constructs 'individual impact' and 'organizational impact' in to single construct 'net benefits'.

In the context of Internet Banking, customers use the IB web site to conduct money transfer, checking the account balance and pay for online e-commerce transactions, This makes Internet Banking website a communication and IS phenomenon suits itself to the updated IS success model. DeLone and McLean (Delone & Mclean, 2004) argued that the Internet applications process fits well into their updated IS success model and the six success dimensions. The authors also encourage researchers to continue testing and challenging their model. Also Research has acknowledged the need for customized measures of IS success based on the context of the system, the work process supported, and the stakeholders considered (Delone & McLean, 2003).

RESEARCH MODEL AND HYPOTHESIS

We base our work on IS Success Model (Delone & McLean, 2003). According to DeLone and McLean (Delone & McLean, 2003) 'To measure the success of a single system (individual system), 'information quality' or 'system quality' may be the most important quality component. The authors argued that for measuring the overall success of the IS department, as opposed to an individual system, 'service quality' may become the most important variable." Therefore, though service quality is important for Internet Banking, it was excluded from this study because this study is based on individual level of analysis. The next few sections describe the hypotheses of the proposed model.

HYPOTHESES

According to Chellappa & Pavlou (Chellappa & Pavlou, 2002) perceived information security is defined as the subjective probability with which consumers believe that their personal information will not be viewed, stored or manipulated during transit or compromised by inappropriate parties in a manner consistent with their confident expectations. The authors also claim that the customer perceptions of information security is influenced by encryption, protection, verification and authentication. The important empirical finding of their research is that perceived security strongly influences customer trust than financial liability of the customer. Their findings of their study support antecedents of perceived security like encryption, protection, and authentication. Kim, Tao, Shin, & Kim (C. Kim, Tao, Shin, & Kim, 2010) found positive association between users' perceived security and their use of e-payment systems. Yoon (C. Yoon, 2010) found that security in online banking strongly influences customer satisfaction. The importance of security and privacy for the acceptance of online banking has been noted in many banking studies (Chen & Barnes, 2007; Mauro C. Hernandez & Afonso Mazzon, 2007). With this, we want to test whether security features role in Internet Banking use and customer satisfaction. Hence, we propose the following hypothesis.

H1a: Security of IB website positively influences IB user satisfaction.

H1b: Security of IB website positively influences IB use.

System quality was defined as quality manifested in a system's overall performance and measured by individuals' perceptions (Delone & McLean, 2003). Since customers cannot directly interact with the bank employees directly, the system quality here is the website quality. Also the

studies (Petter, DeLone, & McLean, 2008; Petter & McLean, 2009) found strong positive association between System quality and User Satisfaction. So we propose the following the hypotheses.

H2a: System quality of IB website positively affects customer satisfaction.

H2b: System quality of IB website positively affects IB use.

Information quality perceived by the user has been instrumental in providing user satisfaction (Hirschheim, 2007; Venkatesh, Davis, & Morris, 2007). According to (Hirschheim, 2007; Venkatesh et al., 2007) the quality of information, as assessed by customers, influences their satisfaction. In the net banking scenario represents the content of the website. Information quality is considered as an integral part of customer satisfaction. So we propose

H3a: Information quality of IB website positively affects customer satisfaction.

H3b: Information quality of IB website positively affects IB use.

Delone and Mclean (DeLone & McLean, 1992) claims that use of the system (here Internet Banking website) leads to customer satisfaction and net benefits to the user. Many authors tested these two relationship with different contexts and there are moderate support for these relationships found in the meta analysis analysis (Petter et al., 2008). Hence, we propose

H4a: Use of the IB website leads to positive effects customer satisfaction.

H4b: Use of the IB website provides net benefits to the user.

We measure information systems success in the context of individual customer. Delone and Mclean (DeLone & McLean, 1992, 2003) claims that the more the user is satisfied, the net benefits is more. Here if the individual user uses more net

banking, he/she may get more net benefits. So the study hypothesizes

H5a: Positive customer satisfaction with Internet Banking usage leads net benefits to the user.

RESEARCH METHODOLOGY

Data Collection

In this study, individuals who used internet, and have experienced Internet Banking services for at least six months, were selected as research participants. We selected students of a premier educational institution in Tamil Nadu, India for this study. Students were chosen for evaluating the model because they are best suited for concept development (Sekhon, Lifan Zhao, Koenig-Lewis, Hanmer-Lloyd, & Ward, 2010). Data was collected using a printed survey questionnaire. A seven point likert type scale was employed with 1 indicating 'strongly disagree' to 7 indicating 'strongly agree' where 4 being 'neither disagree nor agree (neutral). Although we used established and validated measures in the previous studies, we developed a new dimension 'security' in the study. Hence, a pilot was conducted prior to that to ensure that there are no dubious questions. The questions were coined in such a way to avoid any technical jargons. For example instead of the question 'My bank has firewall facility', we use, 'I perceive Internet Banking provides enough protection against hacker threats'.

The participants of the survey were informed that the collected responses would be used only for research purposes. During the weeklong survey period there were 312 samples were collected by randomly visiting classrooms and obtaining permission from the respective faculty members to administer the survey. As all fields were marked mandatory and all the collected responses were verified for missing values. Students who did not answer any questions were asked to answer the same. This ensured there were no missing values.

Demographics

Out of the 312 samples 184 samples were male and 128 sample were female students accounting for 58% percent and 42% respectively. The reported age groups were fallen between 18 and 24 with the average age of 20.52. Post Graduate students were 94 and Graduate students were 218. All students have Internet Banking experience ranging from 1 to 6 years where the average experience being 2.6 years. All students reported that they own a laptop.

Measurement Development

There are six constructs in total viz. information quality, system quality, security features, use, user satisfaction and net benefits. Five items of Information quality were adapted from (Song, Baker, Lee, & Wetherbe, 2012). System quality is measured using eight items adapted from (Wang, Wang, & Shee, 2007). Net benefits has five items based on (Angelakopoulos & Mihiotis, 2011), with our own wordings.

Three measures of satisfaction were adapted from (Roca, Chiu, & Martínez, 2006). Use is measured with five variables adapted from (De-lone & Mclean, 2004; Wang et al., 2007; H. S. Yoon & Steege, 2013). Five measures of security were adapted from (Chellappa & Pavlou, 2002). Customer satisfaction was measured with five measures adapted from (Delone & Mclean, 2004; Gable, Sedera, & Chan, 2008). There are 31 items

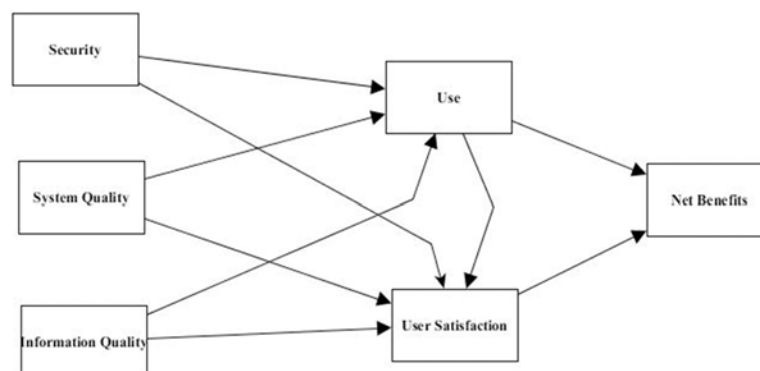
in total representing six dimensions of our model. The proposed model is shown in Figure 1.

Exploratory Factor Analysis

Since the proposed model involves latent reflective constructs, it was proposed to conduct an Exploratory Factor Analysis (EFA). The Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy was 0.9, which was > 0.5 the recommended threshold. In addition, a chi-square value of 5437.719 and significance level of .000 and degrees of freedom 296, were obtained using Bartlett's sphericity test. The results clearly indicates that the inter correlation matrix contains enough common variance, justified the need for factor analysis. Hence, an EFA was conducted with SPSS 20, using Principal Component Analysis (PCA) as the extraction method, with varimax rotation with Kaiser normalization to identify the underlying factors. The PCA procedure was converged in six iterations. We deleted four items from system quality, two from information quality, two items from net benefits, due to loadings < 0.5 .

The PCA procedure was repeated again, and it is converged in five iterations. It resulted in a 22 item 6 factor structure explaining 71.9% of the variance in the model. All the items were highly loaded in their own factor (loading > 0.7) and loaded low on other factors, proving discriminant validity. In addition, each item loads on its latent factor at the significance level of 0.05, indicating

Figure 1. Theoretical model



good convergent validity. The factor loadings of the items are shown in Table 1.

TESTING THE MEASUREMENT MODEL

Reliability and convergent validity of the factors were assessed by composite reliability (CR) and average variance extracted (AVE) (see Table 2). The composite reliabilities were calculated as follows: summation of the factor loadings)/ [(square of the summation of the factor loadings) + (summation of error variables)]. The composite reliabilities were above 0.8 for our measurement

model (Table 2). Construct reliability of all the six constructs were tested using Cronbach's Alpha. The final scales of all the six constructs demonstrate high internal consistency with Cronbach's Alpha values exceeding Nunnally's (Nunnally, Bernstein, & Berge, 1967) recommendation of at least 0.7. The reliability of each factor is as follows: System quality: 0.809; information quality: 0.902; Security: 0.841; net benefits: 0.874; user satisfaction: 0.805. The Average Variance Extracted (AVE) is greater than 0.50 for all the constructs. It implies that that more than one-half of the variances observed in the items were accounted for by their hypothesized factors.

Discriminant validity was assessed, by checking whether the square root of AVE for a construct had a higher value than the variance shared between the construct and other constructs (Fornell & Larcker, 1981). The values of all diagonal elements were greater than those of off-diagonal elements (Table 3), suggesting that all of the constructs were distinct. Also the AVE for each construct is greater than the correlation between that and all other constructs. From this, it is evident that discriminant validity is proved. All the measures of reliability and validity are depicted in Table 3. Overall, the measurement model exhibited adequate reliability, convergent validity, and discriminant validity.

The final measurement model achieved an acceptable fit to the data based on a range of commonly used fit indicators. As shown in Table IV, the CFI, GFI, NFI values are all above 0.9. The RMSEA of 0.051 is below the 0.05 value thus indicating a good fit according to (Hu & Bentler, 1995) Also, the ratio of the chi-square value to the degrees of freedom is 1.824 and thus within the recommended range of 1 to 3 (McIver, 1981).

RESULTS

The Structural Equation Modelling (SEM) approach was adapted to validate the proposed structural model. SEM allows us to estimate the

Table 1. Rotated pattern matrix and factor loadings

	Factors					
	1	2	3	4	5	6
USE3_1	.852	.074	-.004	.060	.176	.039
USE4_1	.814	.082	-.014	.144	.073	.107
USE1_1	.792	.141	.123	.073	.087	.279
USE5_1	.791	.011	.079	.099	.080	.292
USE2_1	.599	.239	.316	-.069	.184	.083
SEC4_1	.024	.819	-.016	.078	.144	-.115
SEC3_1	.105	.818	.092	.078	.139	.194
SEC5_1	.069	.807	.077	.106	.226	.009
SEC2_1	.158	.699	.099	.222	.000	.212
SEC1_1	.238	.555	.252	-.033	.047	.429
IQ1_1	.030	.058	.876	.109	.181	.138
IQ2_1	.125	.087	.874	.033	.217	.175
IQ3_1	.106	.128	.838	.133	.148	.141
WSQ8_1	-.018	.182	.129	.795	-.048	.081
WSQ3_1	.079	.049	-.036	.790	.104	.072
WSQ6_1	.134	-.006	.060	.772	.207	-.052
WSQ7_1	.082	.173	.120	.749	-.026	.184
NB3_1	.171	.181	.258	.136	.777	.143
NB4_1	.187	.216	.212	.072	.771	.302
NB5_1	.217	.252	.243	.077	.715	.323
US3_1	.317	.172	.247	.154	.269	.689
US4_1	.282	.155	.151	.216	.216	.688
US2_1	.226	.026	.189	.039	.352	.644

Table 2. Reliability and validity indicators

	CR	AVE	NB	Use	SEC	IQ	SQ	US
NB	0.883	0.716	0.846					
Use	0.875	0.587	0.497	0.766				
SEC	0.834	0.559	0.513	0.335	0.748			
IQ	0.903	0.756	0.566	0.304	0.291	0.870		
SQ	0.810	0.516	0.292	0.251	0.357	0.244	0.718	
US	0.815	0.597	0.745	0.666	0.465	0.561	0.382	0.772

Note: NB-Net Benefits; SEC-Security; SQ-System Quality; IQ-Information Quality; US-User Satisfaction

strength of interrelationships between the latent constructs (Gallagher, Ting, & Palmer, 2008). The data was analyzed using AMOS Version 22 with the proposed framework shown Figure 1. The Structural model was analyzed with similar fit indices as the measurement model. All the fit indices were according to the recommended level (See Table 2). Hence, we proceed with analyzing the path co-efficient of our model.

Except for the hypothesis between use and net benefits (H4b where $B = -0.133$ and $p = 0.177$) the entire hypothesis were supported. The hypothesis with security (H1a, H1b) on use and customer satisfaction are strongly supported ($b = 0.318$ and $p < .001$, $b = 0.278$ and $p < .001$). The relationship between System quality and use (H2a) is supported at the significance level of $p < 0.01$ with $b = 0.112$. Hypothesis H2b test the effect of system quality on customer satisfaction which has a moderate support ($b = 0.103$, $p < .05$). The relationships between information quality on use

(H3a) is moderately supported ($b = 0.153$, $p < 0.05$) and customer satisfaction (H3b) has a stronger support ($b = 0.341$, $p < 0.001$). The relationships between Use and customer satisfaction (H4a) and Customer satisfaction and net benefits (H5) are strongly supported with $p < .001$ and $b = 0.035$, 0.968 respectively. The summary of the results is presented in Table 4. The percentage of variance explained (R^2) for the main dependent variable net benefits is 68%. Customer satisfaction explains 74% of the variance. Use has an R^2 value of 20%.

DISCUSSION

The purpose of this study is to test a modified IS success model by adding security as a dimension, and to study its effects on IS use and customer satisfaction in the Internet Banking context. The findings clearly indicate that security strongly influences IB use ($b = 0.437$ and $p < .001$) much more than system quality ($B = .157$ and $p < 0.1$) and information quality ($B = .224$ and $p < .05$). Perceived security has been found, to play an important role in IT adoption (Howcroft, Hamilton, & Hewer, 2002; C. Kim et al., 2010; H.-W. Kim, Xu, & Gupta, 2012; M.-C. Lee, 2009). Our study provides support for previous research, if user's perceptions about security is higher, their usage of the IB system, and satisfaction is higher. Another finding is that system quality and information quality on use ($b = 0.157$ and $p < 0.1$ and $b = 0.224$ and $p < .05$). The meta analysis by (Petter et al., 2008) found mixed support for

Table 3. Model fit indicators

Model Fit Indicators	Measurement Model	Structural Model
CMIN/df	1.824	1.881
AGFI	0.883	0.872
NFI	0.913	0.905
CFI	0.958	0.952
GFI	0.911	0.901
RMSEA	.051	0.053
PCLOSE	0.378	0.246

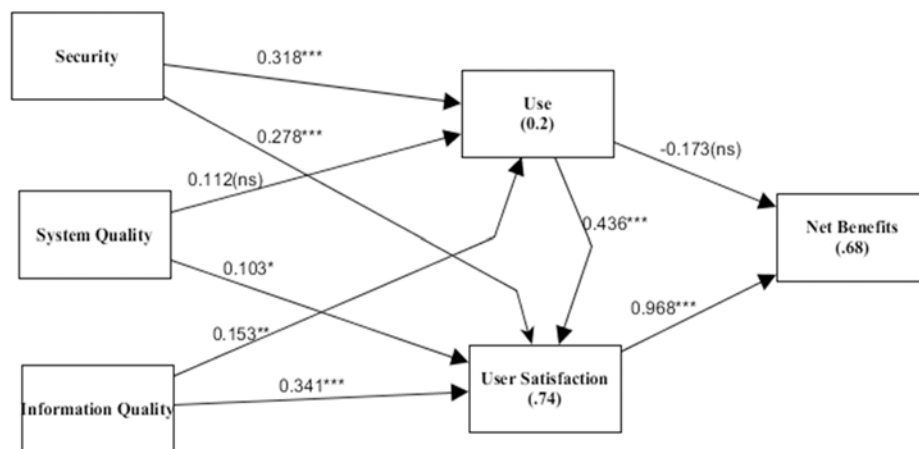
Table 4. Hypothesis testing results

Hypothesis	Path	Standard Coefficient	S.E	t-value	P-value	Result
H1a	Sec -> Use	0.318	0.098	4.459	***	Strong support
H1b	Sec ->CS	0.278	0.046	4.941	***	Strong Support
H2a	SQ->Use	0.112	0.093	1.668	0.091	Not Supported
H2b	SQ->CS	0.103	0.040	2.138	0.032	Weak support
H3a	IQ ->USE	0.153	0.094	2.378	0.017**	Moderate support
H3b	IQ->CS	0.341	0.045	6.593	***	Strong support
H4a	USE->CS	0.436	0.035	7.874	***	Strong support
H4b	USE->NB	-0.173	0.055	-2.048	0.177	Not Supported
H5	CS->NB	.968	0.117	8.726	***	Strong support

Note: ***P< 0.001

Figure 2. Path coefficients

Note ***p, 0.001, **p<0.01, *p<0.05, ns-not significant



both these relationships. Furthermore, system quality on satisfaction ($B=0.186$ $p < 0.05$) had weak support. This is in contrast to the meta analysis (Petter et al., 2008) which found strong positive association between these relationships. The contradictory result may be due to the study (Kwon & Zmud, 1987), which emphasized that, studies on IT adoption should take into consideration, the characteristics of technology under consideration. Although we did not study adoption here, we base our study in a different context. In Internet Banking, system quality may not be an important factor for user satisfaction, because today's customer is more internet perceptive and

practical understanding of the Internet than a decade ago. In addition, our study was conducted on experienced internet users, and it may be a contributing factor for the result. Furthermore, information quality on satisfaction has moderate support in our study but found a strong support in the meta analysis study (Petter et al., 2008).

FUTURE RESEARCH DIRECTIONS

From the theoretical perspective, research should focus on the suitable replacement of the use variable in D&M model since the quality dimensions

of the proposed IS Success model has produced mixed and moderate results. Also only 20% of the variance is explained by use in our model. Seddon while specifying and validating the IS Success model (Seddon, 1997) argued that benefits from use is the only meaning of use in the D & M model. Besides this, understanding Information Systems Use (ISU) behaviors by individual users has been a tough task to {Seddon, 1997 #39} researchers and practitioners (Igarria & Tan, 1997). Also according to (Burton-Jones & Straub Jr, 2006) there is a lack of a consistently accepted definition of the ISU construct in the IS literature. The relationship between use and user satisfaction has been a point of controversy ever since DeLone and McLean proposed their model. In their modified model (DeLone & McLean, 2003) the authors explained the use construct as follows: 'Use must precede "user satisfaction" in a process sense, but positive experience with "use" will lead to greater "user satisfaction" in a causal sense'. The reciprocal relationship proposed in the studies (DeLone & McLean, 1992, 2003) has garnered moderate effects owing to the meta analysis (Petter et al., 2008). This is the focus of attention we put forth for future researchers in finding a suitable alternative for the use to overcome the limitations of the IS Success model.

CONCLUSION

Although we provide a new insight to measure IS success, the study has its own limitations. The major limitation is the context in which it is studied which is Internet Banking. The model needs to be tested in different contexts like online shopping, mobile banking in order to prove its explanatory ability. The study is conducted on students and not in public. The perceptions of a population other than students may be different. Therefore, to be considered consistent, the model has to be tested across different population. In addition, this study did not consider gender, age, Internet Banking experience or any other demo-

graphic factor, which may be relevant but not included in our study. One should evaluate this model with such characteristics to know if any differences appear in the outcome. Another key area to be considered in future research is from the perspective of the study. This study analyzed IS success from individual perspective. If it is considered from the organizational perspective (here the bankers) it may give new insights to the model, which ultimately leads to optimal use of the Internet Banking system.

REFERENCES

- Alhinai, Y. S., Albadi, A., Alshihi, H., & Al-Gharbi, K. (2013). Investigating Determinants of E-banking Adoption by Individuals: Comparing the Impact of System Characteristics and User Traits. *International Review of Management and Business Research*, 2(2), 371.
- Angelakopoulos, G., & Mihiotis, A. (2011). E-banking: Challenges and opportunities in the Greek banking sector. *Electronic Commerce Research*, 11(3), 297–319. doi:10.1007/s10660-011-9076-2
- Bailey, J. E., & Pearson, S. W. (1983). Development of a tool for measuring and analyzing computer user satisfaction. *Management Science*, 29(5), 530–545. doi:10.1287/mnsc.29.5.530
- Burton-Jones, A., & Straub, D. W. Jr. (2006). Reconceptualizing system usage: An approach and empirical test. *Information Systems Research*, 17(3), 228–246. doi:10.1287/isre.1060.0096
- Chellappa, R. K., & Pavlou, P. A. (2002). Perceived information security, financial liability and consumer trust in electronic commerce transactions. *Logistics Information Management*, 15(5/6), 358–368. doi:10.1108/09576050210447046
- Chen, Y.-H., & Barnes, S. (2007). Initial trust and online buyer behaviour. *Industrial Management & Data Systems*, 107(1), 21–36. doi:10.1108/02635570710719034

- DeLone, W. H., & McLean, E. R. (1992). Information systems success: The quest for the dependent variable. *Information Systems Research*, 3(1), 60–95. doi:10.1287/isre.3.1.60
- Delone, W. H., & McLean, E. R. (2003). The DeLone and McLean model of information systems success: A ten-year update. *Journal of Management Information Systems*, 19(4), 9–30.
- DeLone, W. H., & Mclean, E. R. (2004). Measuring e-commerce success: Applying the DeLone & McLean information systems success model. *International Journal of Electronic Commerce*, 9(1), 31–47.
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *JMR, Journal of Marketing Research*, 18(1), 39–50. doi:10.2307/3151312
- Furst, K., Lang, W. W., & Nolle, D. E. (2000). *Internet banking: developments and prospects*. Office of the Comptroller of the Currency Economic and Policy Analysis Working Paper (2000-9).
- Gable, G. G., Sedera, D., & Chan, T. (2008). Re-conceptualizing information system success: The IS-impact measurement model. *Journal of the Association for Information Systems*, 9(7), 18.
- Gallagher, D., Ting, L., & Palmer, A. (2008). A journey into the unknown; taking the fear out of structural equation modeling with AMOS for the first-time user. *The Marketing Review*, 8(3), 255–275. doi:10.1362/146934708X337672
- Harridge-March, S., Wong, D. H., Rexha, N., & Phau, I. (2008). Re-examining traditional service quality in an e-banking era. *International Journal of Bank Marketing*, 26(7), 526–545. doi:10.1108/02652320810913873
- Hirschheim, R. (2007). Introduction to the Special Issue on "Quo Vadis TAM-Issues and Reflections on Technology Acceptance Research. *Journal of the Association for Information Systems*, 8(4), 9.
- Howcroft, B., Hamilton, R., & Hewer, P. (2002). Consumer attitude and the usage and adoption of home-based banking in the United Kingdom. *International Journal of Bank Marketing*, 20(3), 111–121. doi:10.1108/02652320210424205
- Hu, L.-T., & Bentler, P. M. (1995). *Evaluating model fit*. Academic Press.
- IAMAI Annual Report, 2010-2011. (2012). Retrieved from www.iamai.in
- Igbaria, M., & Tan, M. (1997). The consequences of information technology acceptance on subsequent individual performance. *Information & Management*, 32(3), 113–121. doi:10.1016/S0378-7206(97)00006-2
- Kim, C., Tao, W., Shin, N., & Kim, K.-S. (2010). An empirical study of customers perceptions of security and trust in e-payment systems. *Electronic Commerce Research and Applications*, 9(1), 84–95. doi:10.1016/j.elerap.2009.04.014
- Kim, H.-W., Xu, Y., & Gupta, S. (2012). Which is more important in Internet shopping, perceived price or trust? *Electronic Commerce Research and Applications*, 11(3), 241–252. doi:10.1016/j.elerap.2011.06.003
- Kwon, T. H., & Zmud, R. W. (1987). *Unifying the fragmented models of information systems implementation*. Paper presented at the Critical issues in information systems research.
- Lee, C.-P., Mattila, M., & Shim, J. P. (2007). An Exploratory Study of Information Systems Resistance: The Case of Mobile Banking Systems in Korea and Finland. *AMCIS 2007 Proceedings*, 448.
- Lee, M.-C. (2009). Factors influencing the adoption of internet banking: An integration of TAM and TPB with perceived risk and perceived benefit. *Electronic Commerce Research and Applications*, 8(3), 130–141. doi:10.1016/j.elerap.2008.11.006

- Mauro, C., Hernandez, J., & Afonso Mazzon, J. (2007). Adoption of internet banking: Proposition and implementation of an integrated methodology approach. *International Journal of Bank Marketing*, 25(2), 72–88. doi:10.1108/02652320710728410
- McIver, J. (1981). *Analyzing Models with Unobserved Variables*. Social Measurement: Current Issues. Academic Press.
- Nunnally, J. C., Bernstein, I. H., & Berge, J. M. t. (1967). *Psychometric theory* (Vol. 226). McGraw-Hill New York.
- Patnasingam, P., Gefen, D., & Pavlou, P. A. (2005). The role of facilitating conditions and institutional trust in electronic marketplaces. *Journal of Electronic Commerce in Organizations*, 3(3), 69–82. doi:10.4018/jeco.2005070105
- Petter, S., DeLone, W., & McLean, E. (2008). Measuring information systems success: Models, dimensions, measures, and interrelationships. *European Journal of Information Systems*, 17(3), 236–263. doi:10.1057/ejis.2008.15
- Petter, S., & McLean, E. R. (2009). A meta-analytic assessment of the DeLone and McLean IS success model: An examination of IS success at the individual level. *Information & Management*, 46(3), 159–166. doi:10.1016/j.im.2008.12.006
- Pikkarainen, T., Pikkarainen, K., Karjaluoto, H., & Pahlila, S. (2004). Consumer acceptance of online banking: An extension of the technology acceptance model. *Internet Research*, 14(3), 224–235. doi:10.1108/10662240410542652
- Pitt, L. F., Watson, R. T., & Kavan, C. B. (1995). Service quality: A measure of information systems effectiveness. *Management Information Systems Quarterly*, 19(2), 173–187. doi:10.2307/249687
- Roca, J. C., Chiu, C.-M., & Martínez, F. J. (2006). Understanding e-learning continuance intention: An extension of the Technology Acceptance Model. *International Journal of Human-Computer Studies*, 64(8), 683–696. doi:10.1016/j.ijhcs.2006.01.003
- Sathye, M. (1999). Adoption of Internet banking by Australian consumers: An empirical investigation. *International Journal of Bank Marketing*, 17(7), 324–334. doi:10.1108/02652329910305689
- Seddon, P. B. (1997). A respecification and extension of the DeLone and McLean model of IS success. *Information Systems Research*, 8(3), 240–253. doi:10.1287/isre.8.3.240
- Sekhon, H., Lifan Zhao, A., Koenig-Lewis, N., Hanmer-Lloyd, S., & Ward, P. (2010). Adoption of internet banking services in China: Is it all about trust? *International Journal of Bank Marketing*, 28(1), 7–26. doi:10.1108/02652321011013562
- Song, J., Baker, J., Lee, S., & Wetherbe, J. C. (2012). Examining online consumers behavior: A service-oriented view. *International Journal of Information Management*, 32(3), 221–231. doi:10.1016/j.ijinfomgt.2011.11.002
- Tan, M., & Teo, T. S. (2000). Factors influencing the adoption of Internet banking. *Journal of the AIS*, 1(1es), 5.
- Venkatesh, V., Davis, F., & Morris, M. G. (2007). Dead Or Alive? The Development, Trajectory And Future Of Technology Adoption Research. *Journal of the Association for Information Systems*, 8(4), 1.
- Wang, Y.-S., Wang, H.-Y., & Shee, D. Y. (2007). Measuring e-learning systems success in an organizational context: Scale development and validation. *Computers in Human Behavior*, 23(4), 1792–1808. doi:10.1016/j.chb.2005.10.006
- Yoon, C. (2010). Antecedents of customer satisfaction with online banking in China: The effects of experience. *Computers in Human Behavior*, 26(6), 1296–1304. doi:10.1016/j.chb.2010.04.001
- Yoon, H. S., & Steege, L. M. B. (2013). Development of a quantitative model of the impact of customers personality and perceptions on Internet banking use. *Computers in Human Behavior*, 29(3), 1133–1141. doi:10.1016/j.chb.2012.10.005

Zolait, A. H. S. (2010). An examination of the factors influencing Yemeni Bank users behavioural intention to use Internet banking services. *Journal of Financial Services Marketing*, 15(1), 76–94. doi:10.1057/fsm.2010.1

KEY TERMS AND DEFINITIONS

Information Security: It is the practice of preventing unauthorized access, use, disclosure, disruption, modification, inspection, recording or destruction of *information*.

Information System: An information system (IS) is any organized system for the collection, organization, storage and communication of information.

Information Systems Success Model: Seeks to provide a comprehensive understanding of IS success by identifying, describing, and explaining the relationships among six of the most critical dimensions of success along which information systems are commonly evaluated.

Internet Banking: Also known as online banking, e-banking or virtual banking, it is an electronic payment system that enables customers of a bank or other financial institution to conduct a range of financial transactions through the financial institution's website.

Impact of Business Groups on Payout Policy in India

A

Ahana Bose

Indian Institute of Management Calcutta, India

INTRODUCTION

The presence of ‘institutional voids’ in the capital, product and labour markets of emerging economies have helped in the evolution of business groups in them. Business groups represent a unique organization structure akin to ‘cliques’ which comprises separate legally independent entities (business houses or conglomerates), working in different industries which may be unrelated in core functions, held together “by persistent formal (e.g., equity) and informal (e.g., family) ties” (Khanna & Yafeh, 2015). “Spawning” new companies by established business groups may potentially be important in emerging markets where it is probably difficult to start *de novo*” (Khanna & Yafeh, 2015). Group-affiliated firms ‘complements firm-level diversification’, reduces risk of loan default, creates internal capital markets to fund projects and aids in value creation (Classens et al., 2000). Agency conflicts in business group occur between majority and minority shareholders. The interests of managers are aligned to the majority shareholders who appoint them. Paying out regular dividends signal good health of firms to investors, but if a majority shareholder wishes to spend the same money in wasteful acquisitions then dividend payments will be irregular. Here the ‘voice’ of a majority shareholder works against the interest of a minority shareholder often forcing the latter to ‘exit’ the firm. To prevent wasteful activities of cash-rich concerns government has advocated for share repurchases through the Companies Act of 2013 in India. Nearly 30 per cent of the firms accounting for 59 per cent of assets in private corporate sector are owned by business groups

in India, one of the leading emerging economies of the world. India is slated to become the third largest economy in the world by 2030, from its 2013 rank of ten according to the estimates of a PricewaterhouseCoopers (PwC) report, which it an important market to study.

Payout policy is closely linked with most of the financial and investment decisions made by firms. It decides how much amount of cash ought to be returned by the firm to its shareholders. Payout in India occurs either through dividends or share buybacks. The challenge faced by financial economists lies in devising suitable payout policies where firms maximize shareholders wealth while investors maximize their utility. Firms in Indian corporate sector have been hoarding cash after the financial crisis of 2008-09¹. Public sector enterprises (PSU) often end up sitting on cash piles as their executives are unable to decide what to do with the excess liquidity in hand. PSUs distributed 33.1 per cent of their net profit as dividends in 2008-09, which jumped to 45.5 per cent in 2013-14. Close to 60 percent of the PSU’s hoarded cash pile is in the hands of a limited number of firms with Coal India leading the pack. Coal India had a dividend yield of 6.95 per cent in 2015.

However, it is not only the PSUs which are hoarding liquidity. The last five years have seen a dramatic rise in cash hoardings in the private sector as well. We observe that “the private sector’s cash and equivalents grew at a compound annual rate of 20 per cent, accounting for 70 per cent India’s cash pile at the end of 2013-14 compared with 52 per cent in 2007-08.”¹ While Reliance Industries have followed a miserly dividend yield of less than 1 per cent for the last five years, Piramal

DOI: 10.4018/978-1-5225-2255-3.ch005

Enterprises have had a dividend yield of 9.6 per cent in FY13. Within the Tata Group, TCS has had a dividend yield of 3.19 per cent last year and had been making regular payouts over the previous five years. On the contrary, Tata Motors, which had followed a policy of making steady payment of dividends for the last fifteen years, missed it in 2015, much to the annoyance of its shareholders.

Cairn India opting for buyback after Companies Act of 2013 belongs to the Vedanta Group while Coal India is a stand-alone company. Further Tatas, Piramal Enterprises and Reliance mentioned above constitute the top business houses in India. Probing further it is observed that Indian corporate sector has been characterized by concentrated ownership and widespread presence of business groups for a long time. The position of family owned independent trading houses strengthened after the East India Company lost its monopoly. The early 1900s, saw ethnic communities such as Tatas & Birlas gaining prominence in the economy. Post-independence, the 1950s witnessed the inception of groups like the Goenkas & Khaitans through the transfer of assets from British trading houses. The 1960s was the era of creation of groups like Reliance due to the policies of the License Raj². Later on, the economic reforms & liberalization which followed brought groups like Wipro & Ranbaxy to the forefront. Through different regimes, ownership by promoters of a business group continued to dominate the Indian corporate environment. Licensing, political connections, regulations as well as financing decisions over time have affected the functioning of business groups.

The absence of institutions in emerging markets creates difficulties in efficient resource allocation to firms. Diversified groups step in to fill the gap, aiding innovation and entrepreneurship. The goodwill of the group brand name provides a guarantee on behalf of the borrower trying to secure funding for his venture. The guarantee provided by a group's brand name (Maurer & Sharma, 2001) greatly increases its acceptability in the corporate sector due to lack of adequate screening mechanisms in emerging economies to

signal whether a borrower is a 'lemon or a plum' (Akerlof, 1995). Further reallocation of in-house talent within a business group can be done easily benefitting the new units of the group. Jones (2000) notes the similarity of BG firms to private equity ones, relating it to British trading houses in the early twentieth century. He says that one of the primary functions of these early groups was "identifying opportunities and placing potential British investors in touch with them". However not all business groups are alike. For e.g. Birla, another Indian group, "helped found and finance new firms, which were later spun-off using the entrepreneurial talent of its employees".

In the light of the above discussion the focus of the article will be to investigate the impact of a firm's ownership structure on its payout policy.

BACKGROUND

Literature reveals that one of the common ownership patterns prevalent in economies characterized with ambiguous distinction between cash and control rights is business groups. Business groups in different countries have different structures such as pyramids and clusters. *Keiretsu* groups in Japan can be horizontally or vertically integrated. *Keiretsus* can either be conglomerates constituting different industries horizontally linked through cross-shareholdings and intra-group financing or a hierarchical system of suppliers vertically integrated to the core manufacturer. *Chaebol* firms in South Korea are characterized by pyramids and cross shareholdings ownership structure. A few large, older firms present in a *chaebol* control the other firms in the group. *Pyramidal structures* finance new investments of the new firms through the existing equity holdings of the controlling group members while *direct ownership structure* deals with funding requirements through a family owner's personal wealth. Firms controlled through pyramids have lower profitability than their directly controlled counterparts and often their profits are *tunneled* away to firms where

family has higher cash flow ownership. However the underperformance of a *chaebol* is often caused by the presence of low profitability firm chosen by the family for it (selection hypothesis). Further distribution of dividends in *chaebol* appears to be unrelated to the firm's position in the hierarchy and remains unaffected even if new firms are added on to the group. The presence of "diversified business groups" describes the corporate sector of India. Inter-corporate investment plays a critical role in controlling firms of business groups Indian in sharp contrast to the ownership structures of *chaebols* and *keiretsus*.

Dearth of commensurate external funding channels often forces firms to take resort to internal funding sources to meet their investment needs. Business groups are characterized by the presence of internal capital markets. The advantage of internal capital market (ICM) over external capital market (ECM) is the former's ability to shield investment project from the information and incentive problems which plague ECM. The set up assumes the information asymmetry between principal (investor) and agent (manager/entrepreneur) is greatly reduced in an ICM as corporate headquarters (HQ) can monitor its internal divisions better than any fund provider from ECM. Corporate HQ has an internal audit system in place which increases transparency between HQ and the units, compromising the private benefits enjoyed by managers otherwise. In fact managers have to give up a portion of their private benefits to the HQ as the latter has a strong endogenous monitoring system in place to assess the manager's of a division. It curtails a manager's incentives as he is being under constant surveillance by the authority. Stein (1997) observes that better information held by headquarters relative to the external capital market aids ICMs to do a good job at picking winners, especially if the units operate in related lines of business. The flip side of ICM is that the intense competition between the divisions for corporate funding may lead to excessive lobbying (Scharfstein & Stein 2000). Collusion

between "specific divisions and the headquarters" may result in "inefficient cross-subsidizations of weak units by stronger ones". Scharfstein and Stein (2000) states that cross-subsidization of weak divisions by stronger ones in internal capital markets exhibit "socialist" tendencies. An organization often fails to discipline the weaker divisions, end up complying with their whims and compromise on the overall gains of the concern in question. Stronger divisions within the conglomerate are engaged in exerting 'productive effort' to maximize their value. They would not be keen to waste their time in unproductive work as it cuts down their profits. Weaker divisions on the other hand spend time on directly unproductive activities to increase their negotiating power with the headquarters in the subsequent period.

Business groups have evolved in economies where incomplete capital markets and inefficient labour markets. In such cases presence of internal capital markets generated within the business groups help in efficiently allocating resources and filling up the "institutional voids" created by both them in the economy (Khanna et al. (2005). Vig (2013) discusses how strengthening of creditor rights through SARFAESI Act reduces depositor confidence leading to liquidity hoarding in India. Excess in-hand liquidity plays a crucial role in deciding payout strategy. Gopalan et al. (2014) states that group firms payout significant amount of dividends despite agency problems and weak enforceability of legal regulations in the economy. The propensity to do so is higher if group members possess attractive investment opportunities which require immediate funding.

Table 1 illustrates how assets are distributed across the sectors and over the years according to ownership criteria. We find that business groups (BG) owns majority of the assets the in Manufacturing and Non-Financial Services while the opposite holds true for the electricity and mining sectors. We also observe that the amount of assets held by BGs has increased significantly over the years.

Table 1. Mean assets (Rs Million)

	Manufacturing		Non Financial Services		Electricity		Mining	
Year	BG*	NON_BG**	BG	NON_BG	BG	NON_BG	BG	NON_BG
2004	3677.32	1078.08	1811.17	1118.71	5492.18	28450.20	1181.88	15640.30
2005	4286.47	1225.88	2085.06	1233.72	4942.06	28224.79	1679.63	17597.06
2006	5321.39	1454.37	2638.93	1439.55	6352.55	27181.54	15851.09	19698.74
2007	6177.91	1747.92	3582.85	1715.12	7250.37	27716.52	13139.15	19944.55
2008	7129.18	1898.86	4080.46	1817.10	8166.24	28578.33	13273.20	21811.13
2009	8073.62	2193.12	4546.80	2036.43	10980.20	31640.15	14154.47	25028.88
2010	10365.38	2894.14	5700.01	2484.95	14020.45	40752.48	15669.78	32549.05
2011	12628.37	4098.23	6671.48	3405.79	17026.37	49512.56	15490.13	45607.13
2012	14401.16	5543.85	7807.39	4689.40	22328.86	63361.68	16523.10	58026.57
2013	16769.39	7193.81	8379.97	4944.43	23713.86	62574.57	16647.86	72110.27
2014	22739.34	9668.21	10392.78	4667.77	27637.36	81797.40	20802.05	116965.97

BG*: Business Groups, NON_BG**: Non Business Groups ; Source: Prowess CMIE

Table 2 illustrates how cash in hand is distributed across the sectors and over the years according to ownership criteria. We find that BG has higher cash balance than Non BGs in Manufacturing, Non-Financial Service and Electricity sectors while Non BGs hoard higher cash balances in the Mining Sector. Presence of abundant free cash in the firm may lead to overinvestment of funds.

One of the critical decisions of a firm with excess liquidity in hand is fixing up its payout policy. Payout policies are intrinsically related to the financing and investment decisions of a firm and should be decided with foresight and care. The repeated nature of payout decisions implies managers while deciding it need not only keep in mind the pecuniary aspects but also the long-term reputation of their organization. A conservative

Table 2. Mean cash-in-hand (Rs Million)

	Manufacturing		Non Financial Services		Electricity		Mining	
Year	BG*	NON_BG**	BG	NON_BG	BG	NON_BG	BG	NON_BG
2004	14.14	7.74	12.28	12.65	14.30	116.69	15.98	45.77
2005	17.33	9.76	15.56	5.63	10.28	76.06	4.59	59.53
2006	24.33	11.86	17.68	5.81	29.94	162.25	3.58	63.99
2007	20.29	8.99	16.19	6.44	32.08	75.65	10.97	22.28
2008	20.72	7.04	16.68	4.96	21.68	100.75	1.91	16.75
2009	21.48	6.53	18.02	4.90	29.96	86.91	1.14	18.87
2010	22.26	6.67	18.63	5.87	40.51	53.83	0.94	10.04
2011	24.87	7.30	16.48	8.09	15.54	99.56	0.54	6.70
2012	27.03	11.27	19.57	7.35	24.70	101.77	1.16	2.96
2013	34.60	9.04	16.70	7.53	41.81	75.39	1.69	3.71
2014	28.11	10.35	18.36	7.62	32.11	34.12	2.70	2.84

Source: Prowess CMIE

board favors high dividend pay outs while a liberal school of thought espouse stock repurchases. Some board members are indifferent to the choice between dividend payouts and repurchase thinking it does not affect value of the firm.

There has been limited research on the impact of the organization structure of a firm on internal liquidity management in Indian context. Dividend payments and share buybacks are manifestations of what a firm does with excess cash. The next section provides an overview of the payout policies followed by the cash-rich firms in India in recent times.

OWNERSHIP STRUCTURE AND PAYOUT STRATEGY

Firms in India, belonging to business groups typically have a dominant shareholder. Hence, the chance of pursuing entrenchment strategies by them is quite high. Managers of such firms sometimes often carry out activities which render them indispensable to the firm in question. For example, they may follow investment policies designed to suit their needs without bearing in mind the welfare of the shareholders of the company. When managers enjoy private benefits through control, they may exhibit a tendency to smooth dividends. During a downturn, they are likely to pay high dividends to increase their tenure. On the better days, their short term gains bother them less. Following a stable payout policy helps firms to obtain easy access to the capital markets. When firms hold on to an excess amount of cash, then regulators of the economy often mandate share buy-backs to reduce a firm's outstanding number of shares in the market.

Ownership Structure and Dividend Payout

Dividend payments are strongly associated with the reputation of a firm. If a firm has made a bad investment and lost money then it will be finan-

cially constrained and unable to payout dividends to its shareholders. It does send a negative signal regarding the firm's health by its inability to pay dividends on time. However if the firm has a long-standing reputation for making on time dividend payments then stockholders may consider the firm's misfortune in its new investment and allow it time to recover its financial health. Lintner's (1956) analysis about a firm's decision regarding dividends provides a behavioral perspective to the dividend payout story. The survey states most companies while deciding its payout strategy follows a target payout ratio, ensure that dividends lag earnings and try not to vary dividend payment amount much. However Hail et al. (2014) states that payout policies are altered due to information shocks. Miller (2015) explores the effect of financial crisis of 2007-08 on payout policies.

Analyzing aspects of taxes, asymmetric information, incomplete contracting possibilities, and transaction costs embedded in the assumption of perfect capital markets enriches our understanding of the form and pattern of payout policies of corporations. The event of coming together of stockholders of companies with dividend policies aligned to their preferences is called the clientele effect. Clientele effect can be classified into static and dynamic ones. Static clientele models deal with different groups, or "clienteles" who are taxed differently. Miller and Modigliani (1961) argued that "firms have an incentive to supply stocks that minimize the taxes of each clientele. In equilibrium, no further possibilities for reducing taxes will exist and all firms will be equally priced." Dynamic clientele model rests on the premise that if investors continue trading activities over time then tax liabilities are likely to be reduced. Those who are taxed the least will get the dividend paying stock. However such trades may get reversed after the ex-dividend paying date. Jensen et al. (1976) observe that the managers of a publicly held firm can allocate resources to activities which may benefit them, but may not be aligned with that of the shareholders. Instances of such activities can range from excessive expenses

Table 3. Dividend distribution over the years

Year	Mean of Equity Dividend as a Percent of PAT		Mean of Preference Dividend as a Percent of PAT	
	BG*	NON_BG**	BG	NON_BG
2004	14.75	9.07	1.23	0.30
2005	15.81	9.82	1.61	0.29
2006	12.53	8.80	0.52	0.29
2007	12.15	8.04	0.52	0.26
2008	11.47	7.98	0.83	0.49
2009	12.07	8.86	0.52	0.43
2010	12.68	6.96	0.49	0.25
2011	13.63	8.51	2.26	0.52
2012	17.84	11.40	0.43	0.92
2013	19.40	9.87	0.69	0.34
2014	18.57	18.49	0.65	0.18

Source: Prowess CMIE

on corporate parties to unnecessary takeovers. Presence of abundant free cash in the firm may lead to overinvestment.

Payment of dividends can be in the form of cash or stock. Stock dividends are akin to bonus shares which are equivalent to a stock split. They increase the numbers of outstanding shares, reduce the price per share yet the company's assets, profits and the total value stays unaffected. There are regular dividend payments paid out at annual, semi-annual or quarterly time intervals as well as special dividends paid in addition to regular ones. VSNL, an Indian concern offered its shareholders a special dividend of 450 per cent before it went in for disinvestment in 2001.

Dividend payments in India before 1997 were subject to double taxation. Not only were the companies declaring dividends taxed but also the shareholders receiving dividend payments were subject to tax payment by the government. After 1997, the rules were changed, and firms issuing dividends were the only ones required to pay taxes. 2002-03 Budget reverted to the old double taxation policy for dividends. However, 2003-04 Budget changed the dividend distribution tax policy to what it had been after 1997.

Currently, firms paying dividends need to pay 16.9 per cent of dividend distribution tax to the government in India.

Dividends signal the health of a firm to the market. Majority shareholders may use dividends as a disciplining mechanism to keep their managers in check as enforceability of contracts is low in India. However in a family owned firms' agency issues between majority shareholders and managers are secondary in importance, as most of the managers are personal appointees or family members of the former and are unlikely to jeopardize their relationship with them. As the boundary between cash and control rights is fuzzy in business groups, and opaque to an external observer it is well-nigh impossible to comment on the nature of coordination between a manager and an owner while deciding dividend payout policies.

Table 3 describes the distribution of dividends as a proportion of profit after tax. Equity dividends are more popular than preference dividends in India. The data in Table 3 clearly indicates that BGs are more committed to paying out dividends than NON_BGs. However, there are instances where BGs have defaulted on their dividend payments. For example, when Tata Motors missed their divi-

Table 4. Distribution of equity dividends over the years in NON_BG firms

Year	Mean of Equity Dividend as a Percent of PAT				
	Govt*	Foreign	Indian Private	Joint Govt & Private	Stand Alone
2004	16.10	19.77	8.04	24.05	3.01
2005	13.68	19.70	9.05	14.51	6.46
2006	15.35	20.16	7.79	19.67	13.51
2007	12.54	14.23	7.36	51.41	5.47
2008	12.56	16.95	7.33	9.87	11.24
2009	12.37	14.98	8.46	14.93	5.20
2010	12.67	18.17	6.20	11.66	11.01
2011	13.83	16.86	7.81	10.51	7.54
2012	22.08	22.29	10.19	13.82	9.32
2013	21.24	21.85	8.37	15.12	17.47
2014	24.53	22.19	17.98	11.85	4.67

*Govt: Government; Source: Prowess CMIE

dend payment in 2015 for the first time since 2001, it greatly annoyed its aged shareholders; as it was the primary source of income for many of them.

Table 4 illustrates that within NON_BGs the enterprises owned by foreign private sector pay out the maximum dividends, followed closely by Government-owned firms. The firms under the joint ownership of Government and Private sector dole out large dividends as well. In fact, relative BG owned firms, firms under foreign ownership; government owned as well as joint ownership pay out a larger quantity of dividends. However as the standalone entities and privately owned firms pay fewer dividends than the firms under foreign ownership, government owned as well as joint ownership, the mean dividends paid out by all NON_BG firms taken together appear to be less than BG firms.

Ownership Structure and Share Repurchases

The debate between whether to pay out dividends or buy back shares has raged on for a long time. A fundamental difference between the two modes of payout policy is that while dividend payments

are expected to be stable payments over time, stock repurchases may be one-time occurrence for a firm. However, there are a fair number of investors who would prefer firms to buy back their shares rather than pay out dividends.

Repurchases help to do away with free cash flow problems and mitigate conflicts between insiders and external investors. Decisions about repurchase of shares rather than opting for an increase in dividend is positively related to executive options while the number of shares bought back positively correlates to the number of exercisable options (Kahle, 2002). Reddy et al. (2013) state that share repurchases in India assure short-term returns but result in low price earnings ratio compared to pre-buyback period. Gasper et al (2012) state that undervalued firms often signal their value through repurchases. However firms held by short-term investors opt for repurchases more often as they care strongly about the short-term price reaction.

Indian companies went for a total of hundred and sixty buyback offers during 2004-11, out of which only sixty-seven were tender offers. Rest of the buy-back offers happened through the open market route. The Company Act of 2013 explic-

Table 5. Comparison of share repurchases and dividends paid over the years

(Rs Million)	Mean Buyback Amount		Mean Dividend Amount	
Year	BG*	NON_BG**	BG	NON_BG
2004	240.36	1035.35	148.53	154.94
2005	367.02	14.28	179.15	160.49
2006	88.49	572.86	218.22	169.30
2007	518.05	563.41	256.98	186.24
2008	1355.04	2469.63	257.49	180.88
2009	189.48	5.62	320.04	212.31
2010	2022.24	7.86	360.89	259.67
2011	17.45	15238.24	409.85	305.51
2012	19.81	8.69	513.02	389.16
2013	10.57	10910.74	576.59	531.04
2014	20.17	19984.50	882.19	796.45

Source: Prowess CMIE

itly stated that firms sitting on cash piles ought to buy back their shares to reduce the number of outstanding shares in the market. Post the announcement instances of share repurchases have gone up in India.

Table 5 clearly shows that after 2013 share repurchases have increasingly becoming popular post 2013 among the non-business groups (NON_BG) in India. Whether buybacks will replace dividends in near future in India is yet to be seen. The crisis years of 2008-09 saw BGs indulge more in share repurchases than NON_BGs. The erratic pattern of share buybacks is in stark contrast to the mean amount of dividends paid out over time. We find that a number of dividends proposed and paid out have risen steadily over the years for both BGs and NON_BGs.

FUTURE RESEARCH DIRECTION

Carrying out a detailed comparative analysis of payout strategies, tunneling of profits in other emerging and developing economies can be an insightful study. Whether “peer effect” of the corporate HQ affects the payout strategy of its subsidiaries needs to be probed in near future to

help researchers understand how group dynamics impact key decisions in corporations.

CONCLUSION

Business groups dominate the corporate landscape in India. They are found to favor dividends to share buybacks as a payout method. Possibly the notion that share repurchases destroy the value of a firm may encourage them to do so. Share repurchases drive up stock prices, which often cannot be sustained in the long run. At times it adversely impacts the credit rating of the firm and impedes its access to finance. The clientele effect may play a crucial role in deciding the payout policy of firms. While boards of business groups constitute primarily of family members, stand-alone firms have people appointed on professional merit. As a result the latter is more open to experimenting with financing and investment policies. Of late, the stand alone firms are increasingly opting for share repurchases to strengthen their control in them.

Since 2002, firms in India are financializing; i.e they are diverting a substantial part of their cash holdings to the financial sector instead of investing the funds in expanding their production

capacity. Rather than utilizing the excess cash in hand to pay out dividends firms opt for projects where payoffs are realized within a shorter time span. Firms use the cash in hand to purchase financial instruments for getting quick returns instead of investing in long term projects hindering the existence of the firm in the long run. Though group owned firms have exhibited high incidence of financialization (Ganguly & Bose, 2015), yet they have been steadily paying out dividends as well. Whether financialization can be sustained along with a steady payout strategy and how it affects the latter over time remains to be seen.

In general, the chapter addresses the question of how the excess cash in hand of managers of conglomerates gets transferred to the shareholders through payouts. However if a firm believes that liquidity hoarding can strengthen its position in product markets and provide it a strategic advantage in becoming a market leader then they may withhold payouts. Currently, research findings in the aforementioned domain are inadequate; sufficient scope exists in delving into the same to understand how conglomerates influence the functions of corporations.

REFERENCES

- Akerlof, G. (1995). *The market for "lemons": Quality uncertainty and the market mechanism*. Macmillan Education UK.
- Claessens, S., Djankov, S., & Lang, L. H. (2000). The separation of ownership and control in East Asian corporations. *Journal of Financial Economics*, 58(1), 81–112. doi:10.1016/S0304-405X(00)00067-2
- Ganguly, S., & Bose, A. (2015). *Are Firms 'Financializing'? An Investigation into the Indian Manufacturing Sector*. Retrieved from <http://ssrn.com/abstract=2688608>
- Gaspar, J. M., Massa, M., Matos, P., Patgiri, R., & Rehman, Z. (2012). Payout policy choices and shareholder investment horizons. *Review of Finance*.
- Gopalan, R., Nanda, V., & Seru, A. (2014). Internal capital market and dividend policies: Evidence from business groups. *Review of Financial Studies*, 27(4), 1102–1142. doi:10.1093/rfs/hhu004
- Hail, L., Tahoun, A., & Wang, C. (2014). Dividend payouts and information shocks. *Journal of Accounting Research*, 52(2), 403–456. doi:10.1111/1475-679X.12040
- Jensen, M. C., & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics*, 3(4), 305–360. doi:10.1016/0304-405X(76)90026-X
- Jones, G. (2000). *Merchants to multinationals: British trading companies in the nineteenth and twentieth centuries*. Oxford University Press on Demand.
- Kahle, K. M. (2002). When a buyback isn't a buyback: Open market repurchases and employee options. *Journal of Financial Economics*, 63(2), 235–261. doi:10.1016/S0304-405X(01)00095-2
- Khanna, T., Palepu, K. G., & Sinha, J. (2005). Strategies that fit emerging markets. *Harvard Business Review*, 83(6), 4–19. PMID:15938439
- Khanna, T., & Yafeh, Y. (2015). Business groups in emerging markets: Paragons or parasites?. *Review of Economics & Institutions/Economia, Societa e Istituzioni*, 6(1).
- Lintner, J. (1956). Distribution of incomes of corporations among dividends, retained earnings, and taxes. *The American Economic Review*, 46(2), 97–113.

Maurer, N., & Sharma, T. (2001). Enforcing Property Rights Through Reputation: Mexico's Early Industrialization, 1878–1913. *The Journal of Economic History*, 61(04), 950–973. doi:10.1017/S0022050701042048

Miller, M. H., & Modigliani, F. (1961). Dividend policy, growth, and the valuation of shares. *The Journal of Business*, 34(4), 411–433.

Miller, T. O. (2015, January). How sticky are dividends? Comparing payout policy before and after the financial crisis. In *Global Conference on Business & Finance Proceedings* (Vol. 10, No. 1, p. 247). Institute for Business & Finance Research.

Reddy, K. S., Nangia, V. K., & Agrawal, R. (2013). Share repurchases, signalling effect and implications for corporate governance: Evidence from India. *Asia-Pacific Journal of Management Research and Innovation*, 9(1), 107–124.

Scharfstein, D. S., & Stein, J. C. (2000). The dark side of internal capital markets: Divisional rent-seeking and inefficient investment. *The Journal of Finance*, 55(6), 2537–2564. doi:10.1111/0022-1082.00299

Stein, J. C. (1997). Internal capital markets and the competition for corporate resources. *The Journal of Finance*, 52(1), 111–133. doi:10.1111/j.1540-6261.1997.tb03810.x

Vig, V. (2013). Access to collateral and corporate debt structure: Evidence from a natural experiment. *The Journal of Finance*, 68(3), 881–928. doi:10.1111/jofi.12020

ADDITIONAL READING

Khanna, T. (2013). *Billions of entrepreneurs: How China and India are reshaping their futures and yours*. Harvard Business Press.

KEY TERMS AND DEFINITIONS

Business Group: Indian conglomerates.

Payout: Way of repaying excess cash of firms to its shareholders.

SEBI: Regulatory-board of India.

ENDNOTES

¹ Kant Krishna, December 23, 2014, PSUs sitting on Rs 2 lakh-crore cash. Business Standard.

² License Raj: An elaborate system regulations needed to satisfied to run businesses in India between 1947-90.

Noise Trader

A

Po-Keng Cheng

State University of New York, Stony Brook University, USA

INTRODUCTION

Let us start with differentiating between *noises* and *shocks*. While the two terms have some differences, they share some characteristics. Shocks are unexpected large events that affect an economy, while noises are the aggregation of small events which could have powerful influences on the market than a shock can be. Noises and shocks both result from uncertainty and have certain influences on the financial market. Their differences are notable however. Information about shocks is available to everyone in the market but information about noises is not. For an example of a shock, suppose there is an earthquake in Antarctica. Noises, in contrast, come from information asymmetry. For instance, a trader receives messages of an acquisition for company alpha from a manager in company beta, but other traders in the market do not receive the information. Although the information could be wrong, the informed trader who receives the messages will buy more stocks of company alpha in the market. The sudden increase in the demand of company alpha's stock triggers the price of its stock rising sharply, which creates a noise.

Before the shocks happen, traders assign probability measures of shocks and consider their impacts when making financial decisions. Suppose all the traders in the market are fully rational and the initial allocation is Pareto-optimal, then the private information observed by a trader would not cause trades (Milgrom & Stokey, 1982). Thus, the existence of noise trading would provide liquidity to the market (Black, 1986; Dow & Gorton, 2008). However, observed market crashes, such as the Black Monday in 1987 and the burst of Dot-com Bubble in early 2000s, raise the questions whether

all the traders in the market are fully rational and markets are efficient.

BACKGROUND

Lucas' asset pricing model (Lucas, 1978) suggests asset prices are the discounted value of dividends with time preferences. The asset price p_t^e (Krusell, 2007) is equal to

$$p_t^e = E_t \left[\sum_{s=t+1}^{\infty} \beta^{s-t} \frac{u'(y_s)}{u'(y_t)} d_s \right],$$

where β represents for consumer's time preference, $u'(\cdot)$ is consumer's marginal utility function, y_t is the endowment consumer received at time period t , and d_t is the dividend asset paid at time period t . But in the real world, the fluctuations of asset prices in the financial market are far greater to be explained by changes in dividends (Shiller, 1981), suggesting the asset prices are affected by more than fundamentals (Shiller, 2003).

Harrison and Kreps (1978) argued that heterogeneous expectations on dividends policy may cause the price of stock to differ from its fundamental value. De Long, Shleifer, Summers, and Waldmann (1990b) provided a theoretical basis for the noise trader approach. They proposed that the existence of noise traders could lead to divergences between market prices and fundamental values. Additionally, they suggested that some financial anomalies, like the excess volatility of asset prices, the mean reversion of stock return, and the equity premium puzzle could be explained by noise trader risk. Another paper (De Long,

Shleifer, Summers, & Waldmann, 1990a) by the same authors suggested that with the presence of positive-feedback investors, rational arbitrageurs' early positions trigger positive-feedback trading which drives asset prices away from fundamentals. Cutler, Poterba, and Summers (1990, 1991) formalized the role of feedback traders and presented evidences on the characteristic speculative dynamics of returns across markets. Frankel and Froot (1990, 1991) proposed the forecast models of chartists and fundamentalists in which portfolio managers place different weights on chartist and fundamentalist views, explaining the appreciation of the US dollar in 1981–1985.

The noise trader approach (Shleifer & Summers, 1990) aims to explain the equity premium puzzle and other phenomena which cannot be explained well by efficient market hypothesis. There are two main assumptions in noise trader approach. First, not every investor in the market is fully rational: noise traders' demand for assets is affected by their beliefs or sentiments. Second, the limits of arbitrage prevent fully rational investors from bringing the price back to fundamental values. The existence of noise traders may drive out rational arbitrageurs in the market and create positive returns. Shleifer and Vishny (1997) mentioned that with the existence of noise traders in the market, rational arbitrageurs may avoid high volatile positions, dampening the effect of arbitrage which brings prices back to fundamental values.

FUTURE RESEARCH DIRECTIONS

Given the theoretical background of noise traders, the remaining question is how to reveal noise traders in real data. This section introduce several studies related to noise traders from different approaches.

Heterogeneous agent models (Hommes, 2006; LeBaron, 2000, 2006, 2012) consider agents being boundedly rational and following heuristics

or rule of thumb strategies, where the models generate stylized facts, such as clustered volatility and fat tails in asset returns, in the financial market. Several studies since 2005 have focused on estimating the parameters of these models and comparing estimated findings to that of observed historical data (Alfarano, Wagner, & Lux, 2005; Boswijk, Hommes, & Manzan, 2007; de Jong, Verschoor, & Zwinkels, 2010).

Baker and Wurgler (2006, 2007) formed a composite investor sentiment index, where the investor sentiment has significant effect on capital market prices. They found that small stocks, young stocks, unprofitable stocks, high-volatility stocks, non-dividend-paying stocks, extreme growth stocks, and distressed stocks tend to be disproportionately sensitive to broad waves of investor sentiment. When the sentiment is low, these stocks earn relatively high subsequent returns. On the other hand, when the sentiment is high, their subsequent returns are relatively low.

Barber, Odean, and Zhu (2009a, 2009b) analyzed trading records for 66,465 household at a large discount broker and 665,533 investors at a large retail broker and documented that buying and selling decisions of individual investors are highly correlated and persistent. In addition, retail trade imbalances forecast future returns. Over short horizon, such as a day or a week, stocks heavily bought by retail traders earn strong returns, while stocks heavily sold by them earn poor returns. However, over annual horizon, the retail trades move only the prices of small stocks in the same direction of their trade.

Bloomfield et al. (2009) distinguished noise traders from informed traders and liquidity traders, and designed a laboratory market to investigate the behavior of noise traders and their impact on the market. Their results indicated that noise traders have positive effects on market liquidity, acting as rational liquidity providers or behavioral contrarian traders. However, the presence of noise traders adversely affect the informational efficiency of the market when informed traders hold very valuable

information. Moreover, a securities transaction tax would not only limit noise trading, but also reduce the trading and profitability of informed traders.

Expectations of returns are investors' beliefs about future stock market returns, while expected returns (ER) are propositions by financial economists in which the measures of ER are computed by aggregate variables, like dividend-price ratio, consumption, and interest rate. Greenwood and Shleifer (2014) analyzed investor expectations of future stock market returns from six survey data source between 1963 and 2011 and measured ER by dividend-price ratio and variables proposed by Campbell and Cochrane (1999) and Lettau and Ludvigson (2001). The six measures of expectations are highly positively correlated with each other, positively correlated with past stock market returns and the price-dividend ratio, and also highly correlated with investor inflows into mutual funds. Both expectations of returns and ER predict future stock market returns, but ER and expectations of returns are negatively correlated with each other. These findings contradict rational expectations models, in which expectations of stock market returns and model-based measures of ER should be perfectly positively correlated. They reconciled their evidence with the earlier class of behavioral models developed by Cutler et al. (1990) and De Long et al. (1990a), in which fundamental traders require a premium to accommodate the time-varying demand of extrapolative traders. Barberis, Greenwood, Jin, and Shleifer (2015) developed a heterogeneous-agent consumption-based model with two types of infinitely-lived traders: extrapolators and rational traders. The extrapolators extrapolate the market's past prices changes to form their price beliefs, while the rational traders are fully rational and have correct beliefs about the evolution of future stock prices. Their model captures several features of actual prices and returns, such as the excess volatility of stock market returns and the predict power of price-dividend ratios for future price changes.

CONCLUSION

The efficient market hypothesis and the CAPM model indicate that the demand curve of stocks in the financial market would be flat. Some research had worked on figuring out the slope of demand curve for stocks in the market, such as Shleifer (1986). If the efficient markets hypothesis holds, the demand curve of stocks would be flat because all investors agree that the prices of stocks equal their fundamental values. However, this does not exist in the real world: the investors agree to disagree. Thus, an alternative approach to explain the price movements in the financial market is necessary. Challenges faced in related research to noise traders are collecting data on investors' behaviors and distinguishing between different types of traders (Khandani & Lo, 2011; Nagel, 2012; Lou, 2012). However, from the biographies of successful investors and speculators together with the studies introduced above, we could conclude that investors are not fully rational in the market and extrapolative traders do play a role on the price fluctuations in the financial market.

REFERENCES

- Alfarano, S., Lux, T., & Wagner, F. (2005). Estimation of agent-based models: The case of an asymmetric herding model. *Computational Economics*, 26(1), 19–49. doi:10.1007/s10614-005-6415-1
- Aumann, R. J. (1976). Agreeing to disagree. *Annals of Statistics*, 4(6), 1236–1239. doi:10.1214/aos/1176343654
- Baker, M., & Wurgler, J. (2006). Investor Sentiment and the Cross-Section of Stock Returns. *The Journal of Finance*, 61(4), 1645–1680. doi:10.1111/j.1540-6261.2006.00885.x
- Baker, M., & Wurgler, J. (2007). Investor Sentiment in the Stock Market. *The Journal of Economic Perspectives*, 21(2), 129–151. doi:10.1257/jep.21.2.129

- Barber, B. M., Odean, T., & Zhu, N. (2009a). Systematic noise. *Journal of Financial Markets*, 12(4), 547–569. doi:10.1016/j.finmar.2009.03.003
- Barber, B. M., Odean, T., & Zhu, N. (2009b). Do retail trades move markets? *Review of Financial Studies*, 22(1), 151–186. doi:10.1093/rfs/hhn035
- Barberis, N., Greenwood, R., Jin, L., & Shleifer, A. (2015). X-CAPM: An extrapolative capital asset pricing model. *Journal of Financial Economics*, 115(1), 1–24. doi:10.1016/j.jfineco.2014.08.007
- Black, F. (1986). Noise. *The Journal of Finance*, 41(3), 529–543. doi:10.1111/j.1540-6261.1986.tb04513.x
- Bloomfield, R., Ohara, M., & Saar, G. (2009). How noise trading affects markets: An experimental analysis. *Review of Financial Studies*, 22(6), 2275–2302. doi:10.1093/rfs/hhn102
- Boswijk, H. P., Hommes, C. H., & Manzan, S. (2007). Behavioral heterogeneity in stock prices. *Journal of Economic Dynamics & Control*, 31(6), 1938–1970. doi:10.1016/j.jedc.2007.01.001
- Campbell, J. Y., & Cochrane, J. H. (1999). By force of habit: A consumption-based explanation of aggregate stock market behavior. *Journal of Political Economy*, 107(2), 205–251. doi:10.1086/250059
- Cutler, D. M., Poterba, J. M., & Summers, L. H. (1990). Speculative dynamics and the role of feedback traders. *The American Economic Review*, 80(2), 63–68.
- Cutler, D. M., Poterba, J. M., & Summers, L. H. (1991). Speculative Dynamics. *The Review of Economic Studies*, 58(3), 529–546. doi:10.2307/2298010
- de Jong, E., Verschoor, W. F., & Zwinkels, R. C. (2010). Heterogeneity of agents and exchange rate dynamics: Evidence from the EMS. *Journal of International Money and Finance*, 29(8), 1652–1669. doi:10.1016/j.jimonfin.2010.05.007
- De Long, J. B., Shleifer, A., Summers, L. H., & Waldmann, R. J. (1990a). Positive Feedback Investment Strategies and Destabilizing Rational Speculation. *The Journal of Finance*, 45(2), 379–395. doi:10.1111/j.1540-6261.1990.tb03695.x
- De Long, J. B., Shleifer, A., Summers, L. H., & Waldmann, R. J. (1990b). Noise Trader Risk in Financial Markets. *Journal of Political Economy*, 98(4), 703–738. doi:10.1086/261703
- Dow, J., & Gorton, G. (2008). Noise traders. In S. N. Durlauf & L. E. Blume (Eds.), *The New Palgrave Dictionary of Economics* (2nd ed.). London: Palgrave Macmillan. doi:10.1057/9780230226203.1188
- Frankel, J. A., & Froot, K. A. (1990). Chartists, Fundamentalists, and Trading in the Foreign Exchange Market. *The American Economic Review*, 80(2), 181–185.
- Frankel, J. A., & Froot, K. A. (1991). *Chartists, fundamentalists and the demand for dollars*. National Bureau of Economic Research No. R1655.
- Greenwood, R., & Shleifer, A. (2014). Expectations of returns and expected returns. *Review of Financial Studies*, 27(3), 714–746. doi:10.1093/rfs/hht082
- Harrison, J. M., & Kreps, D. M. (1978). Speculative investor behavior in a stock market with heterogeneous expectations. *The Quarterly Journal of Economics*, 92(2), 323–336. doi:10.2307/1884166
- Hommes, C. H. (2006). Heterogeneous agent models in economics and finance. In L. Tesfatsion & K. L. Judd (Eds.), *Handbook of computational economics* (pp. 1109–1186). Amsterdam: North-Holland.
- Khandani, A. E., & Lo, A. W. (2011). What happened to the quants in August 2007? Evidence from factors and transactions data. *Journal of Financial Markets*, 14(1), 1–46. doi:10.1016/j.finmar.2010.07.005

Krusell, P. (2007). *Real Macroeconomic Theory*. Academic Press.

LeBaron, B. (2000). Agent-based computational finance: Suggested readings and early research. *Journal of Economic Dynamics & Control*, 24(5), 679–702. doi:10.1016/S0165-1889(99)00022-6

LeBaron, B. (2006). Agent-based computational finance. In L. Tesfatsion & K. L. Judd (Eds.), *Handbook of computational economics* (pp. 1187–1233). Amsterdam: North-Holland.

LeBaron, B. (2012). Heterogeneous gain learning and the dynamics of asset prices. *Journal of Economic Behavior & Organization*, 83(3), 424–445. doi:10.1016/j.jebo.2012.03.003

Lettau, M., & Ludvigson, S. (2001). Consumption, aggregate wealth, and expected stock returns. *The Journal of Finance*, 56(3), 815–849. doi:10.1111/0022-1082.00347

Lou, D. (2012). A flow-based explanation for return predictability. *Review of Financial Studies*, 25(12), 3457–3489. doi:10.1093/rfs/hhs103

Lucas, R. E. Jr. (1978). Asset prices in an exchange economy. *Econometrica*, 46(6), 1429–1445. doi:10.2307/1913837

Mehra, R., & Prescott, E. C. (1985). The equity premium: A puzzle. *Journal of Monetary Economics*, 15(2), 145–161. doi:10.1016/0304-3932(85)90061-3

Milgrom, P., & Stokey, N. (1982). Information, trade and common knowledge. *Journal of Economic Theory*, 26(1), 17–27. doi:10.1016/0022-0531(82)90046-1

Nagel, S. (2012). Evaporating liquidity. *Review of Financial Studies*, 25(7), 2005–2039. doi:10.1093/rfs/hhs066

Shiller, R. J. (1981). Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends? *The American Economic Review*, 71(3), 421–436.

Shiller, R. J. (2003). From efficient markets theory to behavioral finance. *The Journal of Economic Perspectives*, 17(1), 83–104. doi:10.1257/089533003321164967

Shleifer, A. (1986). Do demand curves for stocks slope down? *The Journal of Finance*, 41(3), 579–590. doi:10.1111/j.1540-6261.1986.tb04518.x

Shleifer, A., & Summers, L. H. (1990). The Noise Trader Approach to Finance. *The Journal of Economic Perspectives*, 4(2), 19–33. doi:10.1257/jep.4.2.19

Shleifer, A., & Vishny, R. W. (1997). The limits of arbitrage. *The Journal of Finance*, 52(1), 35–55. doi:10.1111/j.1540-6261.1997.tb03807.x

ADDITIONAL READING

Brunnermeier, M. K. (2001). *Asset Pricing under Asymmetric Information: Bubbles, Crashes, Technical Analysis, and Herding*. Oxford, UK: Oxford University Press. doi:10.1093/0198296983.001.0001

Campbell, J. Y., Lo, A. W. C., & MacKinlay, A. C. (1997). *The Econometrics of Financial Markets*. Princeton, NJ: Princeton University Press.

Glosten, L. R., & Milgrom, P. R. (1985). Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics*, 14(1), 71–100. doi:10.1016/0304-405X(85)90044-3

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica*, 53(6), 1315–1335. doi:10.2307/1913210

Kyle, A. S. (1989). Informed speculation with imperfect competition. *The Review of Economic Studies*, 56(3), 317–355. doi:10.2307/2297551

Malkiel, B. G. (2003). The efficient market hypothesis and its critics. *The Journal of Economic Perspectives*, 17(1), 59–82. doi:10.1257/089533003321164958

O'hara, M. (1995). *Market Microstructure Theory Oxford*. United Kingdom: Blackwell Publishing Ltd.

Thaler, R. H. (Ed.). (1993). *Advances in Behavioral Finance* (Vol. 1). New York, NY: Russell Sage Foundation.

Thaler, R. H. (Ed.). (2005). *Advances in Behavioral Finance* (Vol. 2). Princeton, NJ: Princeton University Press.

KEY TERMS AND DEFINITIONS

Agree to Disagree: A question discussed in Aumann (1976). If two agents have the same priors and their posteriors for an event are common knowledge, then these posteriors will be the same. That is, these two agents cannot agree to disagree.

Common Knowledge: An event E is common knowledge if E is known by each agent in a group, if each agent knows that E is known by each of them, if each agent knows that each agent knows that E is known by each of them, and so on.

Equity Premium Puzzle: A phenomenon observed by Mehra and Prescott (1985) that the average returns of equity are higher than those of

Treasury bills in the U.S. data during the period of 1889 to 1978.

Fully Rational: Agents who are rational with common knowledge that equilibrium trade is feasible and each agent is rational.

Fundamentalist: Traders who believe the existence of fundamental values and trade based on their expectations of fundamentals.

Heterogeneous Belief: Agents have different expectations for the state of future. Traders with heterogeneous beliefs have different price beliefs for next periods' prices in the financial market.

No-Trade Theorem: Suppose traders in the market are risk-averse and fully rational. If the initial allocation of resources is Pareto-optimal, then private information received by an individual trader will not cause trade, see Milgrom and Stokey (1982) for more details.

Pareto-Optimal: A feasible allocation is Pareto-optimal if there is no other feasible allocation which makes some individuals better off without making some individuals worse off.

Positive-Feedback Investor: Traders who buy when prices rise and sell when prices fall.

Rational: Agents who have von Neumann-Morgenstern preferences.

Safeguarding of ATM

A

Srividhya Srinivasan
University of Madras, India

Priya Krishnamoorthy
SASTRA University, India

Raghuraman Koteeswaran
SASTRA University, India

INTRODUCTION

In today's world ATM has become an inevitable part of human's life. All the developing countries are spending a lump sum of money in printing their currencies and to recycle it. Therefore, the world is changing towards cash-free transactions (Ray, 2015). Cash-free transactions can be done in various ways (PaymentWall Blog, 2015), among which plastic cards plays a vital role. According to (Sharad Raghavan.T.C.A, 2015; Hemali Chhapi, 2015), the cost of printing a currency note is much higher than that of its face value. Rather than spending a huge amount in printing and recycling notes, the cost of preparing plastic cards is considerably less. Hence in order to promote plastic cards as the primary mode of fund transfers, the banking sectors has introduced credit cards, debit cards, etc., throughout the world. Banks provide a number of card systems to facilitate and to attract the customers (BankBazaar.com). It is estimated that the number of ATMs have been increasing every year (Diebold, Incorporated, 2012; Statistic Brain, 2015). Henceforth, security to safeguard ATM in all aspects has become an inevitable one. In this article, a hybrid model consists of both logical and physical ways of securing has been suggested.

BACKGROUND

An ATM is Kiosk machine which has been filled with cash of various denominations, where the person can withdraw money at any time, by using the plastic card. That plastic card has been given by the bank from where the person is having his/her account. That card has a magnetic stripe, a 16 digit card number, a Card Verification Value (CVV) and a 4 digit secret pin which helps the customer to withdraw the amount. Now-a-days, our Indian Government insists all the citizens must have an account in a bank and immediately a debit card has been given for it, thereby promoting cashless transactions. Hence it is mandatory to provide security features for an ATM (Diebold, Incorporated, 2012). Once a card has been issued, the banking sectors are insisted to place Automated Teller Machine (ATM) at various places throughout the country for the benefit of the people (BankBazaar.com). Many research works have been done in providing security for the kiosk machine. But, as per the Newton's Third law *For Every Action there is an equal and opposite reaction*, the more security researchers found, the more techniques were being followed by the robbers to steal the money from the ATM (Diebold, Incorporated, 2012). Kiosks placed within business areas are considerably safer and are less prone to get robbed (McGoey, 2015). Robberies are done by targeting

free standing ATMs in the high ways (McGoey, 2015). Several techniques are used by robbers to steal cash, some of them are card/currency fraud, logical ways and hard ways of attacks (Diebold, Incorporated, 2012; Wild.O, 2015).

Card and currency fraud includes skimming, Transaction Reversal and card/currency trapping/phishing. In these techniques, the perpetrator would fix an extra device to an existing ATM and tries to grab the customer's confidential data thereby steals the money from their account.

In logical attack, the robbers attempt to grab the confidential data of the customers, in a smart way. In 2015, the malware attack is the most common logical attack, in which the perpetrator injects a virus called Tyupkin. The infected ATM will be then work under the control of intruder.

The physical attack is one, in which the robbers target the ATM and try to break it by drilling, cutting, and using fire exposures / explosives. Even, it includes pseudo ATM placement, removal of the ATM, smashing the ATM and many more. In some cases of logical and fraud methodology also, burglary is done by using partial breaking of the machine. Malware is a type of logical attack, in which a partial physical attack is implemented to open the top hat of ATM to inject the malware. Therefore, in any ways the ATM is disturbed by physical attack (Kaspersky).

Although several researches outcomes have been implemented to prevent and detect the ATM frauds, the physical attack on ATM is still increasing (E.A.S.T, 2016). The survey (SecureWorld, 2016) says the percentage of Physical Attacks on ATM is increased in a drastic way. According to European ATM Security Team (EAST), in the year of 2015 alone about €49m was stolen by hard breaking of around 2,657 ATMs (E.A.S.T, 2016). This is because; the main target of the looters is to grab the cash dispenser within a short duration.

There are several detection technologies and risk management technologies available that can expose attacks only after the money is gone (Peter

Beardmore, 2016). So it is indeed not only to detect but also to safeguard the ATM from the intruders.

MAIN FOCUS OF THE ARTICLE

Here, in this article, the primary focus is mainly on safeguarding ATM from physical attacks and how it can be prevented by using a hybrid model.

1. Types of Attacks

The highest rate of ATM fraud was recorded in the year of 2015 (Kelley Moody, 2016). Following section discusses in brief the various types of physical and partial physical attacks on ATM throughout the world.

1. **Ram Raids:** The primarily used hard ways is this, in which the ATM is totally removed from its location with the help of truck or large vehicle. The robbers choose this type, as its success rate is very high, with no extra knowledge or effort (Deluca.R, 2012).
2. **Cutting:** It is another type of physical attack, where the criminals use tools like scissors, saws, screwdrivers, to open the safe door or to break the safe wall of the ATM (Diebold, Incorporated, 2012; Deluca.R, 2012).
3. **Plastic Explosives:** Another foremost type is that robbers are making use of plastic explosives or small dynamites to break the lobby. Such kind of dynamites once ignited can be suppressed by using the method suggested by (Staines.B, 2012). In case of cutting, the Kiosk machine alone gets damaged, whereas in other two methods, there is a possibility of demolishing the outer surroundings also.
4. **'Black Box' Attack:** In a black box assault, the crooks gain physical access to the top of the cash machine. From there, the attackers are able to disconnect the ATM's cash dispenser from the "core" (the computer and brains of the device), and then connect

their own computer that can be used to issue commands forcing the dispenser to spit out cash. In this particular attack, the thieves included an additional step: They plugged into the controller a USB-based circuit board that NCR believes was designed to fool the ATM's core into thinking it was still connected to the cash dispenser. NCR says the crooks then attached a smart phone (a virgin, out-of-the-box Samsung Galaxy 4), which they used as a conduit through which to send commands to the cash dispenser remotely (Brian Krebs, "KrebsonSecurity", 2015).

5. **Wiretapping ATM Skimmers:** "In this attack, the ATM fascia is penetrated close to the card reader to create a hole large enough for the attacker to reach inside the ATM and place a tap directly onto the card reader in order to skim card data as it is read by the ATM," NCR (National Cash Register) said in an advisory it produced on the increasingly common attacks (Brian Krebs, "KrebsonSecurity", 2014). According to NCR, the emergence of this type of skimming attack is a response to the widespread availability of third party anti-skimming technology which is successful at preventing the operation of a traditional skimmer, placed on the outside of the ATM (Brian Krebs, "KrebsonSecurity", 2014). NCR observed that crooks employing this attack are using a variety of methods to create the hole in the front of the ATM. Modern ATMs often now include sensors that can detect vibrations consistent with drilling or cutting tools, so some thieves have taken to melting the ATM fascia in some cases. "Melting techniques have been observed which can circumvent seismic anti-drilling sensors," NCR said. According to quarterly reports from the European ATM Security Team (EAST), ATM "attacks in which the fraudsters attempt to blast open the machine with explosive gas are on the rise". "Collateral damage for solid explosive attacks is a major

concern. In one country, the average overall frequency of ATM related physical attacks is five incidents per week. Three countries reported significant collateral damage from physical attacks, in addition to cash losses suffered (Brian Krebs, "KrebsonSecurity", 2014)."

6. **Spike Brutal ATM Attacks:** (Press Trust of India, 2013, November 22), A study by the industry chamber revealed that ATMs across the country are facing a massive shortfall of 1.5 lakh security guards. Of around 1, 25,000 ATMs in the country, only 50,000 are guarded and just 1.15 lakh are covered by CCTVs, the study found. "The shocking incident at Bangalore underscores the need for recruitment of the guards at expeditious speed to ensure that all the ATMs are properly guarded and equipped with security gadgets, other than CCTVs," it says. A woman employee of Corporation Bank was attacked brutally at the on-site ATM of the bank. After entering the ATM booth and rolling the shutter down, the assailant attacked the woman when she refused to draw money and hand it over to him. "While recruitment of guards and manning ATMs 24X7 may be a time-consuming and costly affair, the answer lies in using hi-tech security gadgets like automatic alarms, improvement of camera footage and setting up central monitoring rooms for coordination between banks and the police," the study said.

2. Materials and Methods

Many tools are developed by several firms to deter Smash and Grab attack on ATM. Subsequent section discusses about the widely used tools that are under implementation.

1. **ATM Plinths:** As per (SCOTIA - Security Group) The Universal ATM Plinth ensures rapid installation, relocation and removal of all ATM models. The patented telescopic

design permits assembly in minutes within a huge height range with no wasted components, while micro-fine adjustment allows for extremely uneven floors. The plinth can be adjusted whilst the ATM is attached, thus negating manual handling issues and allowing the advantage of transportation via pallet truck. These factors allow quick installation and dramatically reduce site times and associated costs.

2. **ATM Pods:** (Associated Security Solutions, 2016) ATM pods provide an added layer of security for the ATM and staff during Cash In-Transit deliveries.
 - a. A bespoke ATM pod made to specific customer sizes depending on number of ATMs housed.
 - b. The unit is 5 sided consisting of wall & roof panels that can be installed easily on site.
 - c. Available with variety of locking options - key, combination, digital and biometric locks.
 - d. Handle and lock on inside for added security when loading ATMs.
 - e. Option of interlock between main door and ATM doors.
3. **Locking Bars:** (Associated Security Solutions, 2016) – “Our Sophisticated interlocking bars ensure that no more than one door can ever be opened at the same time whilst replenishing the ATM. ATMs are often the weakness in many Retail Banking outlets. They hold high volumes of cash and in most instances they are located in the exterior of the building making them an easy target for attack. Our inner locking bars are comprised of a set of four interlocking doors connected by a sophisticated electronic locking system that ensures that no more than one door – accessing one of four cash cassettes – can ever be open at the same time. Once closed, no door can be reopened until a pre-set time has elapsed”.

4. **ATM Anti-Explosion Mats:** (Associated Security Solutions, 2016), Anti explosion mats successfully defuse attacks using Gas and other explosives. They take the impact and reduce the pressure within the ATM, leaving your ATM even after an attack.

3. Prevention Mechanisms

Some prevention mechanisms are also under implementation on a pilot basis. The following are some of the widely implemented techniques.

The prevention methods suggested by “the Colorado banking commissioner” in (Denver, “COLORADO Department of Regulatory Agencies”, 2015) are:

1. **Check for Tampering:** At any ATM, check for signs of tampering or remote installations that don’t appear part of the machine (such as a small camera or wires that appear to be out of place). Look on the card reader and near the speakers. The keyboard should not have a film or cover on it or be more than one piece. There shouldn’t be any loose parts if you give a pull on something that protrudes, like the card reader. Even if these are in check, if something just doesn’t feel right about an ATM, don’t use it.
2. **Wiggle and Cover:** A good practice to get into is to wiggle your card a bit as you swipe it or push it into the reader. Skimmers reportedly have a harder time collecting the data because of the wiggle motion. You can also give a wiggle to the card reader mechanism, keypad and other parts to make sure nothing is loose. Additionally, always use a cover – your hand, your phone, a piece of paper – as you type in your PIN.
3. **Use Indoor ATMs:** High traffic areas with ATMs inside banks and grocery stores are typically safer than outdoor terminals. Also, the risks for skimming are higher on the weekends because criminals will install the skimmers on Saturdays or Sundays and then

remove them before banks open on Monday. Be aware that today's criminals are sophisticated and no ATM is completely safe.

4. **Always Report It:** Reporting suspicious activity or theft to your card issuer or bank as soon as possible is paramount to not being held liable to the amount you might lose if you fall victim to an ATM skimming attack. The sooner you report fraudulent activity the better!

Other than the above mentioned prevention ideas, so many mechanisms are being implemented by the ATM security firms that are as follows:

According to (Brian Krebs, "KrebsOnSecurity", 2014) "In one country no such attacks have been reported since the introduction of ink staining technology," EAST noted. Intelligent Banknote Neutralization System (IBNS) is a prevention technology in which an ink is sprayed on the banknotes, if any malpractice of looting is sensed by the system. It is under implementation, and even made mandatory in countries such as Belgium, Sweden, France, etc. These banknotes are then removed from the circulation among the public. It has one major disadvantage, that once the banknotes are inked, it cannot be used thereafter even by the central bank itself (Dominic Hirsch, 2012).

ATM security system also includes detecting sensors, video surveillance, online ATM monitoring with a battery backed-up network and alarms nevertheless the robbers are well-versed in deactivating or demolishing all security gadgets and easily snatch the cash. Even though, all these protection and detection techniques try to give security for the cash in ATM, in most cases, it fails to capture the robber.

As per the idea suggested in the patent (Staines.B, 2012) (Kenneth J Braddick, 1999), a motherboard can be fixed beneath the ATM. When any mal-functioning is detected, a safety measure is provided that invokes the computerized devices such as CCTV, alarm, tear gas cylinder and the

shutter. A slight modification can be done in the method suggested in (Kenneth J Braddick, 1999).

SOLUTIONS AND RECOMMENDATIONS

In the proposed system, all the preliminary security measures to threaten the robbers has been implemented. In addition to that, rather than using tear gas, smoke will be released from the pack kept inside the ATM, at a minimum level to make the robbers to go unconscious state. At the same time our microprocessors will send a signal to the door control unit to lock the ATM, in a rapid way. In parallel, it will send a SOS message to the nearby police station and to nearby bank branches.

The proposed system, will allow the robbers to penetrate the machine in their own way, with no disturbances to their work. Its main aim is not only to safeguard the cash, but also to capture the robbers. Here, the main aim of the robber is to steal the cash holder, the anchor design is slightly modified, such that an underground cell has been setup, beneath the cash tray of the ATM. The cash-tray is therefore, going to safe-guard by itself in an underground cell, when the robber tries to grab it.

An ATM, in proposed model includes:

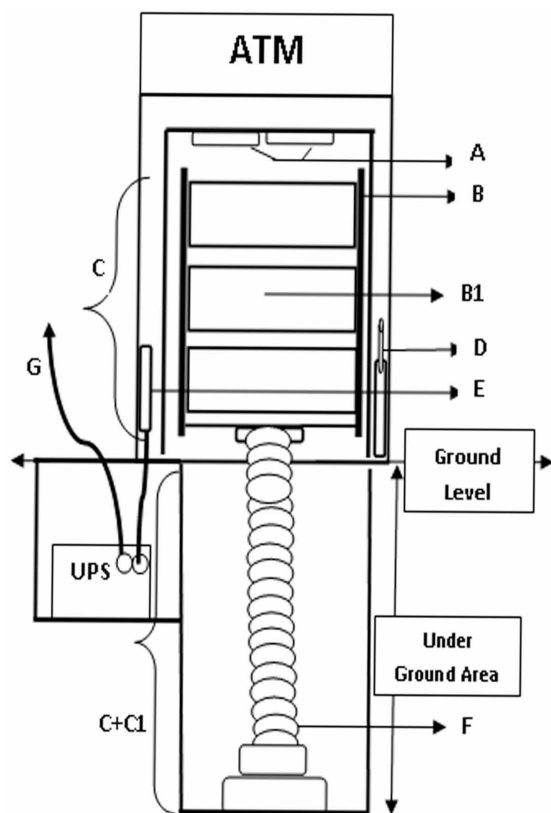
1. A cylinder filled with smoke.
2. A computerized shutter is designed to shut the ATM room in rapid way.
3. A Pulley is used to hold the cash dispenser.
4. An underground cell from where the pulley is mounted that will safeguard the cash.

When a cash holder is ejected in an unusual method, the above mentioned systems will get activated.

If the robber uses any hard devices to open the ATM in a harsh manner, the system will perform the following:

1. It activates the cylinder pack to release the smoke.

Figure 1. ATM – back view



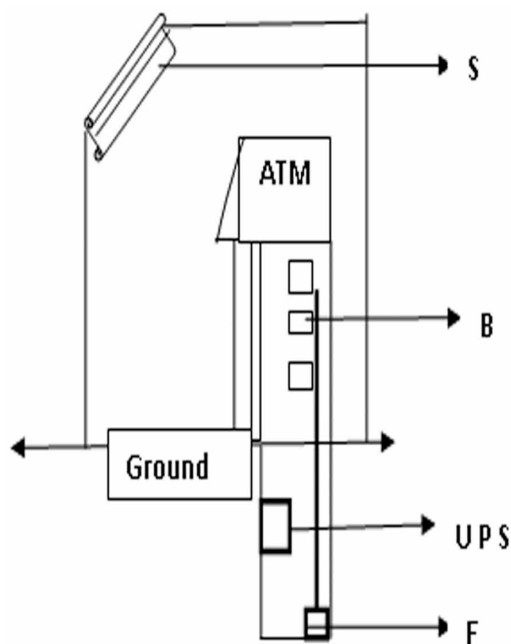
2. At the same moment, the shutter will be closed in a rapid speed. The leakage of smoke suppresses the Oxygen flow in the ATM room that makes the robber to go unconscious.
3. The metal frame fixed on the floor is opened to pull the currency box. Thus the cash tray will be safe-guarded in underground cell.

Architecture

The proposed system consists of:

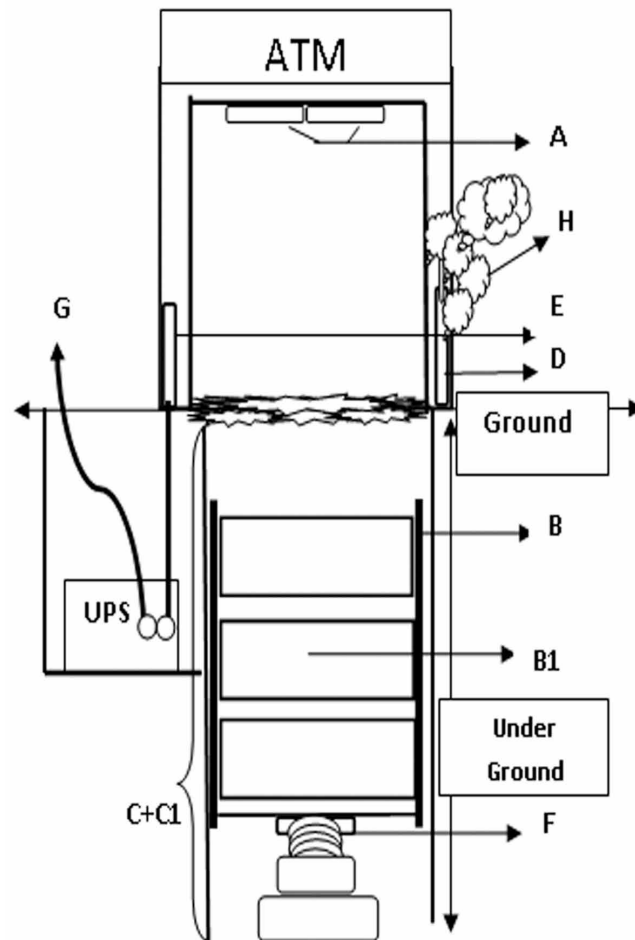
1. ATM
2. Basic Security gadgets infrastructures and Sensor boards
3. Smoke filled Cylinder
4. Pulley
5. Rolling shutter
6. GPS

Figure 2. ATM room – side view



1. **ATM:** This machine is placed inside the normal room where the bankers or third party security agents fill the amount in cash trays (B1), inside cash dispenser (B) for public use.
2. **Security Gadgets:** Cameras, CCTVs, Sensors for alarm, etc., are the preliminary security gadgets that are kept ready for use. In addition, the sensor detectors like Vibration detector and Smoke detector (A) will be placed inside the ATM, which detects a harsh attack, if happens.
3. **Cylinder:** A sensor exclusively assigned for the purpose of sending signals to the cylinder pack (D) that releases the smoke (H) when the ATM has been broken in a harsh manner.
4. **Pulley:** A pulley (F) is placed beneath the ground, and it holds the cash dispenser (B).
5. **Shutter:** A computerized shutter (S) is used, which in emergency period, is activated by the microcontroller, to shutdown the ATM room, in a rapid manner.
6. **GPS:** It is used to give an alert message to the nearby police station and bank branches.

Figure 3. ATM – back view (when a harsh attack is detected)

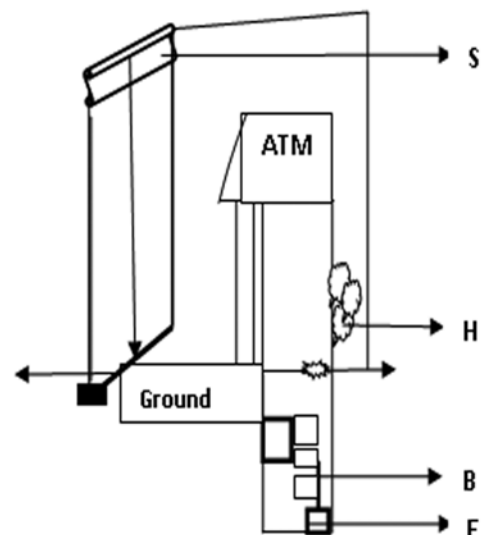


To monitor the status of all the above sensors a Central Unit (E) will be kept. The sensing devices are connected to the Central Unit through wires, channeled via walls. As per the idea provided by (Amit Tyagi, “ProTecht”, 2016), the monitoring system will send a signal to the Central Monitoring Station at the time of emergency.

Mechanism

An ATM with preliminary security measures and sensor boards is placed in a room. The Height of cash dispenser inside the ATM is calculated (C). Below the ground level, an underground cell is provided with a height of (C+C1), where C1, is any constant. A pulley (F) is set which holds the

Figure 4. ATM Room– side view (when a harsh attack is detected)



cash dispenser (B). A cylinder (D) containing smoke is placed inside the ATM. In addition, a micro-controller (E) will monitor and activate all these appliances during hard time.

For instance, if the robbers try to break the ATM with any tools or explosives, immediately the sensor board (A) which is kept inside the ATM triggers the pulley to pull down the mounted cash dispenser rapidly beneath the ground.

At the same time, according to (Staines.B, 2012) (Kenneth J Braddick, 1999) suggestions, the cylinder (D) gets alert to release the smoke so as to make the thieves go unconscious.

Simultaneously, as per the technique suggested in (Kenneth J Braddick, 1999), a signal has been passed to the shutter and it also closes in a fraction of a second, thereby our system locks the thieves and sends alert messages to the (Kenneth J Braddick, 1999) nearby police station or bank branches.

FUTURE RESEARCH DIRECTIONS

Maintaining a huge number of ATMs by implementing the discussed mechanisms leads to the maximum security for it. On the other hand, burglars used to find all sort of techniques to by-pass it. In near future, to get full control over the burglars, technology drives the human to go cashless and card less system.

CONCLUSION

The proposed system is helpful not only to prevent the robbery but also to divert the culprits for the time duration, until the security guards reach the place. So that the thieves can be identified and the criminal rate can be possibly reduced.

REFERENCES

- Associated Security Solutions. (2016). *ATM Protection*. Associated Security Solutions. Retrieved from <http://www.associated-security.co.uk/products/ATMProtection.php>
- Beardmore, P. (2016). *Threat Detection Techniques – ATM Malware*. Retrieved February 19, 2016, from <https://blogs.rsa.com/threat-detection-techniques-atm-malware/>
- Braddick. (1999). *Currency anti-theft device*. Retrieved September 14, 1999, from <http://www.google.co.in/patents/US5952920>
- Chhapia, H. (2015). *To print a Re 1 note, the govt has to spend Rs 1.14: RTI*. Retrieved July 3, 2015, from <http://timesofindia.indiatimes.com/india/To-print-a-Re-1-note-the-govt-has-to-spend-Rs-1-14-RTI/articleshow/47918118.cms>
- Debit Card. (n.d.). Retrieved from <https://www.bankbazaar.com/debit-card.html>
- Delivered, D. I. (2012). *White Paper, ATM Fraud and Security*. Retrieved from http://securens.in/pdfs/KnowledgeCenter/5_ATM%20Fraud%20and%20Security.pdf
- Deluca, R. (2012). *Ram Raiding: The Biggest Threat to ATM Security*. Retrieved June 21, 2012, from <http://www.securitysolutionsmagazine.biz/2012/06/21/ram-raiding-the-biggest-threat-to-atm-security/>
- Denver. (2015). *Consumer Advisory: ATM and credit card skimming prevention*. Retrieved September 16, 2015, from <https://www.colorado.gov/pacific/dora/news/consumer-advisory-atm-and-credit-card-skimming-prevention>
- E.A.S.T. (2016). *ATM physical attacks*. European ATM Security Team (EAST). Retrieved April 12, 2016, from <https://www.european-atm-security.eu/tag/atm-physical-attacks/>

Hirsch, D. (2012). *Cash protection technologies: Do they deliver?* Retrieved July 25, 2012, from <http://www.atmmarketplace.com/blogs/cash-protection-technologies-do-they-deliver/>

Kaspersky. (n.d.). *Tyupkin Virus (Malware)*. Retrieved from <http://www.kaspersky.com/internet-security-center/threats/tyupkin-malware-atm-security-malware>

Krebs, B. (2014). *More on Wiretapping ATM Skimmers*. Retrieved December 9, 2014, from <http://krebsonsecurity.com/2014/12/more-on-wiretapping-atm-skimmers/>

Krebs, B. (2015). *Thieves Jackpot ATMs With 'Black Box' Attack*. Retrieved January 6, 2015, from <http://krebsonsecurity.com/2015/01/thieves-jackpot-atms-with-black-box-attack/>

Moody, K. (2016). *2015 has highest rate of ATM fraud ever recorded*. Retrieved April 18, 2016, from <http://www.kesq.com/news/2015-has-highest-rate-of-atm-fraud-ever-recorded/39097248>

Payment Wall Blog. (2015). *Types of Payment Methods for Ecommerce*. Retrieved November 18, 2015, from <http://blog.paymentwall.com/guides/monetization/types-of-payment-methods-for-ecommerce/>

Press Trust of India. (2013, November 22). Spike brutal Bangalore ATM-type attacks turn to tech: Assocham to banks. *The Indian Express*. Retrieved from <http://archive.indianexpress.com/news/spike-brutal-bangalore-atmtype-attacks-turn-to-tech-asso-cham-to-banks/1198319/>

Raghavan, S. (2015). *What does it cost to print money?* Retrieved April 28, 2015, from <http://www.livemint.com/Opinion/tdg8G3rKtQFmwE-2hCcOrFK/What-does-it-cost-to-print-money.html>

Ray. (2015). *Which Countries are Closest to Cashless Societies?* Retrieved April 13, 2015, from <http://letstalkpayments.com/which-countries-are-close-to-a-cashless-world/>

Scotia Security Groups. (n.d.). *Products//ATM Plinth Advanced Raid Protection*. Scotia Security Groups. Retrieved from <http://www.scotia.co.uk/products/atm-security/atm-plinth/>

SecureWorld. (2016). *How Attackers cracked ATMs in 2015*. SecureWorld. Retrieved February 28, 2016, from <https://www.secureworldexpo.com/how-attackers-cracked-atms-2015>

Staines, B. (2012). *ATM Security System*. Retrieved April 26, 2012, from <https://www.google.com/patents/US20120097404>

Statistic Brain. (2015). *ATM Machine Statistics*. Retrieved from <http://www.statisticbrain.com/atm-machine-statistics/>

Tyagi, A. (2016). *Innovative e-Surveillance and Integrated Power Efficient Solutions for ATMs*. Retrieved from <http://www.mphasis.com/solutions/protect-atm-security-system-and-monitoring-solution.html>

Wild, O. (2015). *Six Types of ATM Attacks and Fraud*. Retrieved July 9, 2015, from <http://www.ncr.com/company/blogs/financial/six-types-of-atm-attacks-and-fraud>

KEY TERMS AND DEFINITIONS

ATM: Automated Teller Machine is a kiosk which has been filled with cash of various denominations, where a person can withdraw money at any time, by using a plastic card.

CVV: “Card verification value” on the credit card or debit card is a 3 digit number on VISA, MasterCard and Discover branded credit and debit cards.

European ATM Security Team (EAST): A ‘non-profit’ organization whose members are committed to gathering information form and disseminating EAST outputs to, ATM deployers and networks within their countries/regions.

GPS: The Global Positioning System is a space-based navigation system that provides location and time information in all weather conditions, anywhere on or near the Earth where there is an unobstructed line of sight to four or more GPS satellites.

Intelligent Banknote Neutralization System (IBNS): IBNS is a prevention technology in which an ink is sprayed on the banknotes, if any malpractice of looting is sensed by the system.

National Cash Register: NCR is an American computer hardware, software and electronics company that makes self-service kiosks, point-of-sales terminals, ATMs, Check processing systems, barcode scanners, and business consumables.

Physical Attack: To launch a physical assault (against) with or without weapons.

Sensors: A device which detects or measures a physical property and records indicates or otherwise responds to it.

Category A

Artificial Intelligence

Artificial Ethics

Laura L. Pană

Polytechnic University of Bucharest, Romania

INTRODUCTION

The possibility to conceive and to effectively apply a new ethics, deeply explicit and valid in theory, strongly relevant in practice, and suitable for both artificial and human intelligent agents is argued in this chapter.

The whole intellectual history shows that long ago humans were aware of their specific gift to add artificial objects of social (material or ideal) kind to those natural, i.e. to create culture, as an extension of nature. The oldest cyberneticist, Plato, in his *Republic*, offered the first description of society as a human design and an artificial product. Aristotle, in his turn, wrote in *Politics* (II. 5) about the automatic tools and installations of Daedal and even about the mentally controlled tripods created by Hephaestus, which served the “band” of gods.

For the specific field of moral culture, P. Danielson outlined the artificiality of morals and analyzed the possible degrees of creativity by which moral values are invented and moral culture is renewed (Danielson, 1998). He explicitly writes: “important parts of morality are artificial cognitive and social devices” (p.292). Even before, J. Bentham, who explored both the human mind and the *Table of Springs of Action* (1812), understood the artificial character of morality and described “the whole fabric of morals”. He elaborated a both comprehensive and operational moral theory, and coined a specific calculus, based on the central moral value of his vision.

This study identifies theoretical possibilities to model moral conduct, and aims to find and to develop a set of moral prerequisites (mental aptitudes and practical skills), available and suit-

able for any kind of moral agents. In this way, the axiological foundation of the new ethics may be continued by an attempt to identify an appropriate and operable value-set, selected from a complete ethical system.

The main practical contributions covered by chapter: indicates a well-founded way to effectively connect moral theory and moral practice; proposes a complex but feasible strategy of designing artificial moral agents, detailed in a few operational phases; makes a concrete proposal of modeling and implementing moral action in the behavior of artificial agents; shows the new possibilities offered by the present changes in value systems for effective moral agents designing and training.

BACKGROUND: SCIENTIFIC, TECHNICAL, AND PHILOSOPHICAL PREMISES OF ARTIFICIAL ETHICS

Some theoretical deficiencies of great ethical systems and some practical difficulties of applying abstract moral values in concrete conditions by individual agents have been frequently discussed by ethicists, in their common effort to establish a new foundational theory of moral choice, moral freedom and then of a deep moral conduct.

The nature of artificial agents able to behave ethically has been extensively studied. This study led to different conclusions, which can however be considered convergent. Among other features of moral agents McDermott emphasizes the capacity to take specific, moral decisions, just like (Moor, 2005 and 2011), and analyses ethical reasoning, seen as the main decisional sub-structure (McDermott, 2011). The role of free will is also

detailed by him, while the will itself is refined up to temptation. J. Gips also emphasizes the role of free will, as of conscious choice (Gips, 2011), and shows the necessity to develop perceptual and non-symbolic aspects of morality, and to favor training, not teaching of abstract theories. (p. 250). Wallach and Allen correlated autonomy and sensitivity (to moral considerations) as defining dimensions for artificial moral agents (Wallach & Allen, 2009). For J. P. Sullins, the relevant aspects of moral agency are autonomy, intentionality and responsibility (Sullins, 2006). L. Floridi studies the distribution of the necessary factors for an ethical behavior - interactivity, autonomy and adaptability -, between a large category of agents, such as natural objects, ecosystems, technical systems, organizations and humans. Intentionality is also associated, as the most important, responsibility, which is here distinguished from accountability (Floridi, 2011, p. 205). The list of the required capacities is also extended in (Anderson, 2011) by sentience, self-consciousness, reasoning, and emotionality. Moral agency is also considered here, although some of the above indicated elements are integrated into the very internal structure of action, as aspects or parts of interests, motivations, decisions and goals. Other authors dedicated to define agency include, among other features, normativity and asymmetry (Barandiaran, Paolo, & Rohde, 2009). Spatio-temporality, also assigned to agents by them, actually is a universal property. Moral agent's study, as part of the artificial intelligent agent's theory, may be illustrated including by (Pană, 2005c), a study on artificial cognitive agents, followed by (Pană, 2008b), on cognitive and moral agents, viewed in their evolution (Pană, 2006b). This chapter continues and deepens the growing list of the possible features previously assigned to artificial moral agents by (Pană, 2006a; Pană, 2012), especially by attributes related to consciousness and value systems.

Many scholars believe consciousness is a condition of moral judgment and moral conduct. Here the approaches range goes from those philosophical, neurobiological and psycho-sociological up

to computational ones, such as those which mark the progress in the field from diverse perspectives (Gamez, 2008), which ambition to build a conscious machine (Angel, 1989) or which explore large areas of the field but also thoroughly study a series of essential topics (Holland, 2003). A study on axioms and tests for the presence of minimal consciousness in agents establishes as eligibility criteria perception, imagination, attention, planning capacity and emotion (Alexander & Dunmall, 2003), in conformity with their set of axioms (p. 9-10), which is adequate for the consciousness understood, *apud* Dennett, as "a virtual machine running on a parallel neural computer" (p.8). The contribution of the present chapter in this matter is based on studies concerning an integrative model of brain, mind, cognition and consciousness (Pană, 2008a), as well as on a few forms of social consciousness (Pană, 2000), and consists of an analysis of the internal structure of moral conscience, in order to find which of its components and levels are suitable to be modeled and implemented in artificial moral agent programs.

The project of building an artificial ethics, suitable both for human and artificial moral agents, have to include, from the start, a value-centered approach and vision. In the today dedicated literature, the values under discussion are those human, and mainly those already recognized in the moral field, such as good and evil, right and wrong, but also a few others, implied both in current human activities, as in the human-computer interaction, e.g. egoism and altruism, treated in (Floridi, 2011), together with values occurring in organization environments, such as equal opportunity, financial stability, good working and holiday conditions for employees, good service and value to their customers or shareholders, and honesty, integrity and reliability to other companies (p. 205). Inspired by competing and cooperation needing contexts, other authors study the relationship between altruism and reciprocity, as in (Danielson, 2002), where he develops evolutionary agent-based models in order to test their conjecture in various cases. Responsibility, which remains a core value

in the new, computer-centered and wired work and life environment, is also frequently studied, sometimes with outstanding results (see Noorman, 2014), linked or not with its pair in the accredited ethical binomial. The possibility to conceive and to effectively design value, motivation and norm directed artificial and human intelligent agents was anticipated in works on the evolution from Virtue Ethics to Virtual Ethics (Pană, 2005a), by studies dealing with crucial events in IST development (Pană, 2010), which could be concluded with value computing and motivation inventing (Pană, 2006c), and by others that aim to model the evolution of value systems (Pană, 2004b), as through essays to discover or to design common skills of artificial and human moral agents which enable them to pursue specific values, such as in (Pană, 2006b).

MAIN FOCUS OF THE ARTICLE: TO GENERATE AND APPLY AN ETHICS VALID IN THEORY AND EFFECTIVE IN PRACTICE

The current attempts of moral values, principles and norms implementing raise difficult problems in the case of both human and artificial agents. Moreover, some parts or even levels of a complete moral system are weak or are even missing from the “roadmap” of the common man. Moral reflection, especially, which is conditioned by the construction and assimilation of a scientific and philosophical superstructure of moral life, is habitually in this situation.

A project of technical implementation of human morality as such in artificial agents is counter-efficient and therefore counter-indicated because of its internal contradictions and failures. This is one of the reasons from which the artificial ethics project includes, as an implicit goal and as an indirect result, a renewal and an optimization of the human morality itself, by a complex, interdisciplinary endeavor. Some additional difficulties appear, mainly from an operational perspective:

- How can moral values be put into practice by intelligent machines which use just an abstract form of intelligence?
- Moral values are synthetic values (Pană, 2008b) which require a particular type of knowledge - practical knowledge;
- A special kind of knowledge is also necessary for their understanding – evaluative knowledge - see (Pană, 2000);
- The general, vague and often counter-intuitive features of values call for creative intelligence;
- Their practice may require several mental (individual) aptitudes and cultural (social) attitudes;
- Can “natural evolution” of programs be adapted for a moral dimension of machine’s action? Singularity, un-repeatability and irreversibility of moral decisions are other characteristics of moral action that should be further considered.

To add a moral component in a machine architecture and to generate thus a “moral life” of artificial intelligent agents is a hard and time requiring task, but its accomplishment is helped by the nature itself of the moral phenomenon, that has several favoring features, such as:

1. The multitude of the implied capabilities, skills and competences and, by consequence, an increased number of ways to achieve the goal;
2. The synthetic nature of many of these factors, such as moral intelligence, whose study was initiated in (Pană, 2005c), continued in (Pană, 2006a; Pană, 2008b) and will be deepened in a study now in progress;
3. The fact that moral values are active, are observed or are even generated in each activity field, which also offer their specific means for their accomplishment;
4. Moral conduct transcends biotical and psychical conduct, and as a social conduct it is or will be at the reach of both human and artificial moral agents;

5. The spiritual nature of moral values, that make them able to be formalized and then accessed by any kind of agents;
6. If developed also as cultural beings, by training or even by educational activities similar to children's cultural formation, artificial agents will be able to effectively deal with specific cultural values;
7. The desirable and possible cooperation of entities;
8. The highly normative structure of moral conduct;
9. The possibility to develop moral techniques as variants of intellectual techniques;
10. The chance to apply a wide range of methods (philosophical, psychological, technical);
11. The opportunity to conceive and to design all practical intelligence forms as grounded on values, motivated by projects, oriented by goals, and realized by techniques (intellectual or instrumental).

SOLUTIONS AND RECOMMENDATIONS: ARTIFICIAL AGENTS WITH QUASI-COMPLETE ETHICAL SYSTEM

A Way to Design Artificial Moral Agents, Directed by Values and Norms

A promising solution may consist of the following, complex but feasible strategy, which includes as phases:

1. A both structural and functional analysis of a complete moral system, corroborated with
2. A critical, and in the same time constructive inquiry of the now operating technical achievements in the general theory of artificial intelligent agents;
3. Modeling each component of ethical conduct as a cluster of activities, composed of

actions, acts and actema. Mental activities will be decomposed in functional blocks, processes and operations.

4. Using an adapted cybernetic chain of action, with the following necessary steps of moral action: cognition → intention → motivation → decision → execution → evaluation → anticipation.

Overall, the strategy is to enrich, as much as possible, the usual software architecture of artificial agents, by adding to the cognitive and other habitual architecture components, an intentional architecture, then a motivational one, a decisional and an operative architecture, an evaluative and, finally, a projective architecture. The prospective logic used in (Pereira & Saptawijaya, 2011) can here be helpful. A coordination and control center of the general software architecture becomes necessary in an action area as complex as the ethical one.

In modeling and implementing moral action in the behavior of moral agents, the cybernetic circle has to be not only complex, but also complete. It will include, besides the well-known *feedback*, also the *feed-before* and even the *feed-up* connections. In this way, it will be transformed into an open ellipse, and in perspective, in an upward elliptic helix, that can represent including ethical aspirations. Openness here means forward, not only backward looking and acting, and new projects initiating.

For the first phase of the proposed strategy the following steps seem to be suitable and now feasible: 1 - to describe the internal structure of a complete ethical system; 2 - to study its components to find out if they can be applied by artificial agents, once translated into explicit and clear instructions, and 3 - to test the behavior of moral agents in an environment that models as accurately as possible the features of moral practice.

The Internal Structure of a Complete Ethical System

The overall picture of the moral system, specifying the place and the internal structure of ethics, as one of its sub-systems is as follows:

1. Moral practice;
2. Moral conscience:
 - a. Moral skills or habits, acquired by living in a moral community;
 - b. Moral affectivity, initially manifested as moral emotion, later on possibly developed into moral feelings;
 - c. Specific norm-based moral judgment and moral decision, facilitated by personal intelligence or even by specific education;
 - d. Moral beliefs or convictions, that are complex, both rational and affective formations and, at the top of the internal structure of moral conscience,
 - e. Moral reflection.
3. Ethics, which is the elaborated, theoretical, meta-theoretical and methodological level of a moral system, has itself a systemic feature, and presents the following internal structure:
 - a. Moral values;
 - b. Moral principles;
 - c. Moral norms, rules and procedures;
 - d. Ethical models, which are the result of various, theoretical, methodical and integrative attempts to establish a functional, harmonious and durable connection between the first three levels of the ethical subsystem (3), and the practical moral conduct (1), thanks to the mediation of (2), the moral consciousness subsystem.

In the framework of traditional morality, moral problems are emerging still from the first, habitual level of the morality, because at this level, moral norms are introduced through the system of posi-

tive and negative sanctions, which generates the corresponding, sometimes positive, but mostly negative reactions, at the second, affective level of moral life.

This is the first cause that makes humans live their moral life not as a form of freedom, but as an experience of coercion. The second one consists of troubles induced by incongruity between abstract ethical values and concrete conditions of their application in moral conduct. The aggregate effect of these causes leads to the dubitative, inconsequent or even contradictory nature of moral attitude and constitutes a major and durable inconvenience of traditional morality.

Moral decision seems to be the touchstone of coherence in a moral system, because it is the main activity able to ensure the continuity between the system of accredited values, principles and norms organized in various models, respectively the current morality. The task to bridge the gap between moral theory and moral practice may be accomplished by starting from a deep analysis of the present state and trends of the moral value-system.

Current, Favoring Changes in the Ethical Value System

As favoring axiological processes for the accomplishment of a new ethics, the following ones can be summarized:

- The homogenization process of different action types, due to their increasing scientific and technical content and features, also generates the phenomenon of incorporating into a single and almost neutral category – that of error -, all the wrong or failed results of different human enterprises, named before bad, wrong, guilt or sin, which become, in this way, its derivatives;
- The tendency of passing to a more permissive or even relativistic evaluation way, which no more considers as irreparable and

then punishment needing, all the wrong or failed results of human enterprises, but as repairable or even reversible, and which allows a more nuanced situation analysis and decision;

- The tendency of polarization in the hierarchy of values, by which the groups of highly, respectively weakly appreciated values tend to move toward extremes, while a third group, that of basic values, becomes thus central, i.e. strongly felt and widely shared;
- The complementary tendency, that of concentration of human needs, by which some needs, often basic needs, even disband several of the old values;
- Some needs, spectacularly increasing until recently, regress now in some respects, e.g. the cognitive need, which degenerates into the need for information and data;
- Knowledge as a value recovers more than what loses in favor of information, through enhanced applicability, through successful combinations of philosophical, scientific, technical and moral cognition;
- The emergence of hybrid approaches to technical and moral issues, as evaluative cognition and prospective cognition (Pană, 1988) or cognitive emotion (Wallach & Allen, p.155), operationally treated by them as “cognitive representation of emotions” (p. 152) also enriches this axiological space;
- Information however becomes the central value of the whole value system and,
- The informational language is now the only one shared with our cybernetic creations;
- The new outcomes in programming decision sequences, action structures and behavior models in artificial agents, which may be seen as steps in creating conditions for appropriate ethical values applying in this field.

The convergent effect of all these changes in the present value system is the generation of new possibilities to develop moral calculation, moral agents, and a moral interaction of human and artificial agents.

An Appropriate Central Value for the Field of Artificial Ethics

The *good* and the *bad*, which were seen, for now, as central in ethics, are not even specific to the moral realm, as demonstrated in (Pană, 2000, pp. 70-86; 108), because they are used in all types of activities as components of a general foundation of their own value-systems.

Therefore, *good* and *bad* are functioning rather as a pair of *meta-values*, which may guide and evaluate any kind of human conducts, almost always characterized by multiple and often contrary needs, interests and motivations.

The *good*, especially, as the most appreciated and widely used axiological construct, is general enough and indefinite too, as well as comprehensive and inspiring, to function as an efficient evaluative basis for the entire class of intentions, motivations, choices and results of human action.

But from the chosen perspective and for the stated purposes, good and goodness, together with their associated virtues, are not useful in the case of beings (for now) without consciousness, such as some intelligent, trainable or even learned, robotic (physical or virtual) agents.

Correctness could be such a value, at the same time strong and universal, but however within our reach, thanks to our recently acquired scientific and technical skills, and because (of):

1. It is suitable both for human and artificial morality;
2. Its increased clarity, intelligibility and applicability;
3. It may function with the same efficacy
 - a. In all types of activity, and
 - b. At every level of human or artificial conduct.

But can the evoked conditions and difficulties of moral action be eliminated by the simple correctness and can be this moral aspiration as strong as goodness was in the history of morality? As a spiritual realm of being, morality was lead, at least until now, by values which cumulated methodic and ergonomic, but also motivational and projective functions.

FUTURE RESEARCH DIRECTIONS

The presented findings may be continued by following a set of research and activity directions, dedicated:

1. To set-up a unitary and interdisciplinary strategy in artificial moral agents designing, a strategy which could be philosophically grounded, scientifically supported and technically realized.
2. To model moral conduct as a net of cybernetic connections, governed by the complete set of cybernetic interaction types - feedback, feed-before and feed-up -, able to ensure inclusively the development (ascending evolution) of a population of human and artificial agents.
3. To describe and to decompose each cybernetic step of action (cognition, intention, motivation, decision, execution, evaluation, anticipation) in practical units: activities, acts and actema.
4. To use and invent new information and knowledge-based soft techniques and technologies, appropriate for intelligent automatic social systems instantiating and mastering.
5. Building, modeling and implementing new sets of moral, scientific, technical, economic and political values with the aim to restore the social life on cultural bases.

CONCLUSION

By proposing a new ethical approach of IST issues and AI projects, and by opening a new interdisciplinary research field – that of Artificial Ethics -, we hope to a) eliminate the internal contradictions which were born wherewith some ethical theories; b) surpass the practical failures of human moral conduct; c) allow an axiological foundation of new moral codes, adequate for various types of intelligent systems; d) ensure a conflict-free and effective human-machine cooperation and e) stimulate the common evolution of human and artificial intelligent agents. The study distinguishes the appropriate values for the contemporary ethical environment and proposes a comprehensive and operational strategy of their implementation. Artificial Ethics building will be, in fact, a product of the co-cognition and co-construction of human and artificial intelligent agents in their intertwined evolution on the background of different but concurrent processes of nowadays, such as the rising of a multiple artificial intelligence or maybe of a general artificial intelligence, as of artificial cognition and artificial discovery, artificial activity and artificial society. Artificial ethics will be generally applicable and, because it will reduce or eliminate many difficulties of biotic, psychic, social or historical and even cultural sort, it will treat mainly issues related to distinct fields of activity, types of tasks and degrees of complexity. Thus, artificial ethics will expectedly be the first effective ethics.

REFERENCES

Alexander, I., & Dunmall, B. (2003). Axioms and Tests for the Presence of Minimal Consciousness in Agents. In O. Holland (Ed.), *Machine Consciousness* (pp. 7–18). Exeter, UK: Imprint Academic.

- Anderson, S. L. (2011). The Unacceptability of Asimov's "Three laws of robotics" as a Basis for Machine Ethics. In Anderson, S. L., & Anderson, M. (Eds.), *Machine Ethics*, (pp. 285-296). Cambridge University Press.
- Angel, L. (1989). *How to build a Conscious Machine*. Boulder, CO: Westview Press.
- Angeli, A., & Johnson, G. I. (2004). Emotional Intelligence in Interactive Systems. In *Design and Emotion* (pp. 262-266). London: Taylor & Francis.
- Barandiaran, T. E., Paolo, E. D., & Rohde, M. (2009). Defining Agency: Individuality, Normativity, Asymmetry, and Spatio-Temporality in Action. *Adaptive Behavior – Animals, Animats, Software Agents, Robots Adaptive Systems*, 17(5), 367–386.
- Danielson, P. (1998). How Computers Extend Artificial Morality. In T. W. Bynum & J. H. Moor (Eds.), *The Digital Phoenix: How Computers are Changing Philosophy* (pp. 292–307). Oxford, UK: Blackwell Publishers.
- Danielson, P. (2002). Competition among Cooperators: Altruism and Reciprocity. *Proceedings of the National Academy of Sciences*, 7237-7242.
- Floridi, (2011). On the Morality of Artificial Agents. In Anderson, S. L., & Anderson, M. (Eds.), *Machine Ethics*, (pp. 184-212). Cambridge University Press.
- Gamez, D. (2008). Progress in Machine Consciousness. *Consciousness and Cognition*, 17(3), 887–910. doi:10.1016/j.concog.2007.04.005 PMID:17572107
- Gips, J. (2011). Towards the Ethical Robot. In Anderson, S. L., & Anderson, M. (Eds.), *Machine Ethics*, (pp. 243-252). Cambridge University Press. doi:10.1017/CBO9780511978036.019
- Holland, O., & Goodman, R. (2003). Robots with Internal Models: A Route to Machine Consciousness? In O. Holland (Ed.), *Machine Consciousness* (pp. 77–110). Exeter, UK: Imprint Academic.
- McDermott, D. (2011). What Matters to a Machine? In Anderson, S. L., & Anderson, M. (Eds.), *Machine Ethics*, (pp. 88-114). Cambridge University Press. doi:10.1017/CBO9780511978036.010
- Moor, J. H. (2005). The Nature and Importance of Machine Ethics. In *Machine Ethics, Papers from the AAAI Fall Symposium Series* (p. 78). Arlington, VA: AAAI Press.
- Noorman, M. (2014). Computing and Moral Responsibility. In *The Stanford Encyclopedia of Philosophy*. Retrieved August 15, 2015, from <http://plato.stanford.edu/archives/sum2014/entries/computing-responsibility/>
- Pană, L. (1988). Cognition and action centered values in human conduct structuring, *Revue Roumaine des Sciences Sociales. Série de Philosophie et de Logique*, 32(1-2), 11–18.
- Pană, L. (2004b). Modeling the Evolution of Value Systems from the Perspective of Technical Culture. In L. Pană (Ed.), *Evolutions in Value Systems under the Influence of Technical Culture Development* (pp. 29–50). Bucharest, Romania: Politehnica Press.
- Pană, L. (2005a, March). *From Virtue Ethics to Virtual Ethics*. Paper presented at the Meeting of the Interdisciplinary Research Group of the Romanian Committee for History and Philosophy of Science and Technology of the Romanian Academy, Bucharest, Romania.
- Pană, L. (2005c). Moral Intelligence for Artificial and Human Agents. In *Machine Ethics, Papers from the AAAI Fall Symposium Series* (pp. 107-109), Arlington, VA: AAAI Press.
- Pană, L. (2005d). Philosophy of Artificial and Artificial Philosophy. *Academica*, 34(15), 32–34.
- Pană, L. (2006a). Artificial Intelligence and Moral Intelligence. *TripleC: Cognition, Communication, Co-operation*, 4(2), 254–264.

Pană, L. (2006b). Co-evolution of Human and Artificial Cognitive Agents. In *Proceedings from Computers and Philosophy: An International Conference* (pp. 366-378). Laval, France: Institute Technologique de Laval, Le Mans Université.

Pană, L. (2006c, June). *Values Inventing and Motives Computing*. Paper presented at The 4th European Computing and Philosophy Conference, Trondheim, Norway.

Pană, L. (2008a). An Integrative Model of Brain, Mind, Cognition and Conscience. *Noema. Journal of the Romanian Committee of History and Philosophy of Science and Technology*, 7(1), 120–137.

Pană, L. (2008b). Moral Intelligence: Elements of Ethics for Cognitive and Moral Agents, *Noesis. Travaux du Comité Roumain d'Histoire et de Philosophie des Sciences*, 33(1), 39–51.

Pană, L. (2010). Crucial Intellectual Events in the History of Information Science and Technology, *Noesis. Travaux du Comité Roumain d'Histoire et de Philosophie des Sciences*, 34(1), 171–196.

Pană, L. (2012). Artificial Ethics - A Common Way for Artificial and Human Moral Agents and an Emergent Technoethical Field. *The International Technoethic Journal*, 3(3), 1–20.

Pereira, L. M., & Saptawijaya, A. (2011). Modeling Morality with Prospective Logic. In Anderson, S. L., & Anderson, M. (Eds.), *Machine Ethics*, (pp. 398-421). Cambridge University Press. doi:10.1017/CBO9780511978036.027

Sloman, A. (1990). Motives, Mechanisms, Emotions. In M. Boden (Ed.), *The Philosophy of Artificial Intelligence* (pp. 231–247). New York: Oxford University Press.

Sloman, A., & Chrisley, R. (2003). Virtual Machines and Consciousness. In Holland, O (Ed.), 2003. *Machine Consciousness*, (pp. 133-172), Exeter, UK: Imprint Academic.

Sullins, J. P. (2006). When is a Robot a Moral Agent? *International Review of Information Ethics*, 6(12), 23–30.

ADDITIONAL READING

Haikonen, P. O. A. (2012). *Consciousness and Robot Sentience, Series on Machine Consciousness* (Vol. 2). New Jersey, London, Singapore, Beijing: World Scientific Publishing Co. doi:10.1142/8486

Holland, O. (Ed.). (2003). *Machine Consciousness*. Exeter, UK and Charlottesville, VA, USA: Imprint Academic.

Moor, J. H. (2011). On the Nature, Importance and Difficulty of Machine Ethics. In Anderson, S. L., & Anderson, M. (Eds.), *Machine Ethics*, (pp. 13-20), Cambridge University Press.

Pană, L. (2000). The Moral Culture. In Pană, L., *Philosophy of Technical Culture* (pp. 104–112). Bucharest, Romania: Editura Tehnică.

Pană, L. (2004a). Artificial Philosophy. In Pană, L., *Philosophy of Information and Information Technology* (pp. 249–323). Bucharest, Romania: Politehnica Press.

Pană, L. (Ed.). (2006). *Cultural Models of the Knowledge Society from the Perspective of Technical Culture*. Bucharest, Romania: Politehnica Press.

Pană, L. (2013, May). The Effectiveness Principle in Social Theory and Practice. Paper presented at the *Meeting of the Commission of Socio-Economic Cybernetics of the Economic, Juridical and Social Sciences Division of Romanian Academy*, Bucharest, Romania.

Wallach, W., & Allen, J. W. (2009). *Moral Machines: Teaching Robots Right from Wrong*. Oxford, UK: Oxford University Press. doi:10.1093/acprof:oso/9780195374049.001.0001

KEY TERMS AND DEFINITIONS

Artificial: Any completely new social outcome of human activity, either theoretical or practical, such as a concept, a method or artifact, a social

body, a system of relations or an institution, therefore, any kind of real mental construction or material achievement.

Artificial Ethics: Is the present stage of the evolution from Philosophical Ethics to Scientific Ethics and to Technoethics. It is also based on important findings made in computer-centered ethical fields such as Information Ethics, Computing Ethics, Machine Ethics and Web Ethics. Artificial Ethics is both a conceptual and technical research field and a theoretical and practical area of moral action.

Artificial Moral Agents: A sub-set of artificial intelligent agents, which have classes of attributes of artificial agents, and other ten specific features, gradually identified in (Pană, 2005c; Pană, 2006a; Pană, 2006b; Pană, 2008b).

Axiology: A fundamental constitutive field of philosophy, which studies the genesis, structure and functions of values, their cognition and realization, as well as the dynamics of value systems. It is also developed as a science, practiced with various instruments, including those logical, as in Formal Axiology.

Co-Evolution (of Human and Artificial Moral Agents): The process and the product of the co-cognition, co-sentience, co-operation and co-construction of human and artificial intelligent agents, by which they develop new and common functions, skills and characteristics, which become sources of their further evolution.

Complete Cybernetic Connection: Is a cluster of different types of connections, which consists not only of the well-known *feedback*, but also of *feed-before* and even of *feed-up* connections, all of them necessary in modeling and implementing moral action in the behavior of artificial moral agents.

Value Systems: Besides its geosphere, biosphere, sociosphere, technosphere, infosphere and noosphere, our world also presents an axiosphere, which may be analyzed inclusively in terms of Systemics: each subsystem of a social system has its own value-system, with a central value, around which orbit specific values. Similarly, every epoch imposes its own core value which attracts the appropriate set of values.

Artificial Intelligence

Steven Walczak

University of South Florida, USA

INTRODUCTION

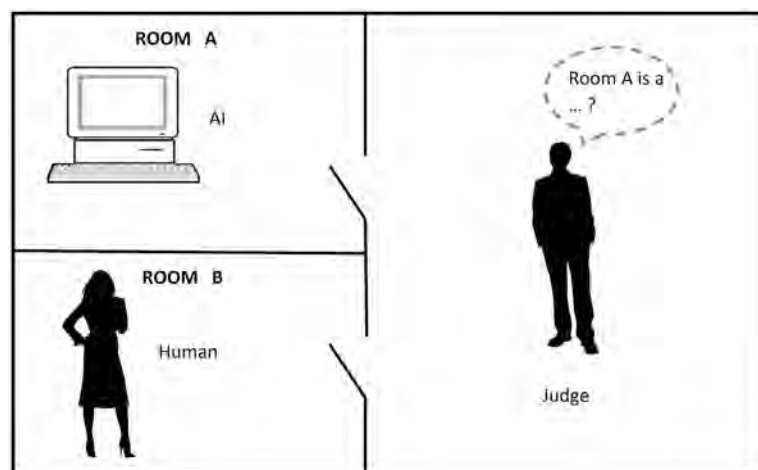
The ENIAC computer unveiled in February 1946 is considered by many to be the first programmable digital computer. Soon after this first computer was initialized, human researchers started to dream of intelligent machines. Alan Turing (1950) proposed a means for computers to demonstrate intelligence. Thus, the science of artificial intelligence (AI) was created. The term *artificial intelligence* was first coined in August, 1955 (McCarthy et al., 2006). The goal of AI is to create computer programs, possibly embedded in sophisticated hardware such as robots, which are capable of intelligent thought, ultimately creating new knowledge. Intelligent systems should be able to perform any task normally associated with human cognition and intelligence, such as goal setting, goal seeking, path planning, and problem solving.

There are two schools of thought within AI. The first seeks to create machine intelligence by

emulating the way humans think. An example of this approach is the MYCIN expert system, which utilized the problem solving processes of human experts in order to diagnose bacteriological infections (Shortliffe et al., 1975). The second is interested in creating intelligence through whatever mechanisms are available, regardless of its similarity to human thought processes. An example of this second approach is the Deep Thought (renamed Deep Blue® after IBM® purchased the program from Carnegie Mellon) chess playing program which utilized massively parallel computer architecture to achieve chess playing expertise by being able to examine game trees to a much deeper depth (Newborn & Kopec, 1989).

The Turing test (Turing, 1950), depicted in Figure 1, essentially states that if the output of an AI program cannot be distinguished by a knowledgeable human being who cannot see the respondents from responses made by another human, then the machine is intelligent. The Turing test has been

Figure 1. Turing test; judge in main room tries to determine which room has the computer AI and which room has the human by asking questions and evaluating Room A and Room B responses



DOI: 10.4018/978-1-5225-2255-3.ch009

criticized that it doesn't really evaluate intelligence, just a computer's ability to imitate human conversational responses to questions. The Eliza program is an early version of what has become known as a chatbot (Weizenbaum, 1966). Eliza conversationally performed psychotherapy, but in actuality was simply using canned responses to pre-identified keywords present in the human user's responses to Eliza's statements and questions. A modern version of the Turing test is the annual Loebner Prize competition (AISB, 2016).

An interesting outcome of the Turing test is the identification that certain abilities humans may take for granted, such as speech, vision, and hearing are extremely complex for computers to perform, while complex mathematical operations are extremely easy for computers to perform. Additionally, Turing's ideal is for AI-based computers to have global knowledge, similar to humans, so that they could converse on a wide variety of topics intelligently. However, as shown by the Eliza project, early AI research was focused on solving much more narrowly defined problems within single domains. This trend of having highly specialized intelligence to outperform human experts in problem solving within a specific domain or for a specific problem with a domain is still continued today.

The field of AI has numerous disciplines:

- **Agents, Also Known as Intelligent Agents, Mobile Intelligent Agents, or Mobile Agents:** The use of atomic processing entities that collaboratively work to solve problems.
- **Artificial Life:** Systems that model natural life processes, behaviors, and evolution, both for investigating biologic and population dynamics and to be used in problems solving.
- **Artificial Neural Networks, Also Known as Neural Networks and Connectionist Systems:** Machine learning systems based on the processing of the human brain.
- **Cognitive Modeling:** Creates models of cognitive processes to explore the nature of thinking and problem solving.
- **Computer Vision:** Creating the ability to accurately interpret visual data such as pictures, animations, and movies.
- **Expert Systems, Also Known as Knowledge Based Systems:** Computer programs that simulate or surpass human expertise with a defined domain or collection of domains.
- **Evolutionary Computation:** Genetic algorithms or genetic processing.
- **Knowledge Acquisition:** Capturing human knowledge normally at an expert level.
- **Knowledge Representation:** Encapsulating acquired knowledge in a machine usable format.
- **Machine Learning:** Designing processes that enable machine intelligence to grow beyond the bounds that have been statically programmed into the AI application.
- **Natural Language Processing, Also Known as Speech Processing:** Systems that can understand and produce both written and spoken language.
- **Planning:** Development of intelligent plans for accomplishing goals, including game playing systems.
- **Robotics:** Development of automated hardware systems that can perform human-like tasks, including robots that mimic animal or human form and behavior.

The many disciplines of AI are not necessarily independent, but frequently interact with each other to try and create more comprehensive intelligence in computers. For example, a human like robot would utilize research from cognitive modeling, computer vision, knowledge acquisition, knowledge representation, machine learning, natural language processing, and planning in addition to the obvious robotics discipline. New subdisciplines are often created when an existing

disciplines begins to grow, such as artificial neural networks and evolutionary computation being outgrowths of the original machine learning discipline and knowledge acquisition and knowledge representation having started in expert systems research. With the many and often interacting disciplines within AI, the field of AI research serves as a complex milieu of tools for mimicking or surpassing human intelligence or creating machine intelligence to solve a vast collection of research problems.

HISTORICAL BACKGROUND

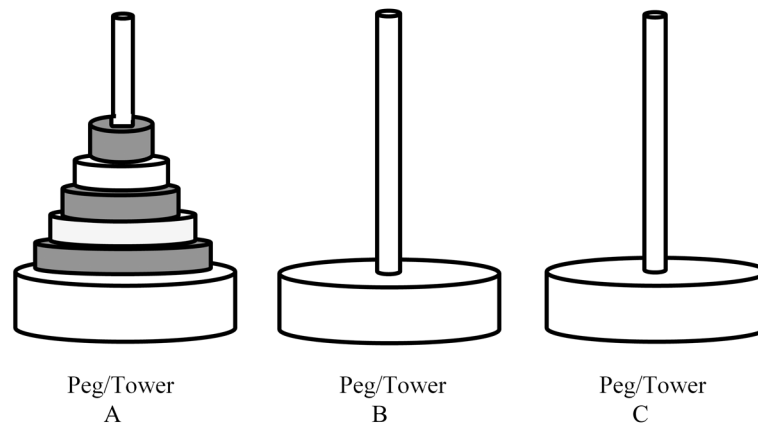
Early work in AI revolved around basic problem solving, including development of intelligent search strategies, development of heuristic methods for reducing problem complexity, and goal seeking and planning behaviors (Pearl, 1984). Although AI research had been progressing since its inception in the 1950's, AI fever hit in the mid-1970's with the demonstration of the MYCIN expert system, which was able to outperform individual physicians on diagnosing specific types of bacteriological infection. Similar to the Gartner hype cycle (Linden & Fenn, 2003), once practical applications of AI theory were established (e.g., MYCIN), research and expectations from AI grew rapidly, but often outpaced reality. Large government grant projects in the 1980's to foster AI research and AI application development often fell short, due to an overestimated potential for AI at the time. Knowledge based and intelligent systems fell out of vogue and funding contracted significantly, but academia kept AI research alive. Research shifted from expert or knowledge-based systems to more machine learning oriented systems, including artificial neural networks and genetic algorithms. Nowadays research in AI is once again increasing and this may be due in part to the success of AI research that was evaluated using games (e.g., backgammon or chess) and the entertainment value of such AI games, which caught the public's attention.

Games have often formed a foundation for initial and exploratory research in AI, especially in checkers, chess, poker, and other strategy games. The Deep Blue chess playing program of IBM played and lost its original 6 game match in 1996 under tournament conditions to then world champion and chess grandmaster Gary Kasparov, but did manage to win one game and with tied games lost to Kasparov with a match score of 2 to 4 (Newborn, 2012). A rematch occurred the following year and Deep Blue was able to win two games outright and the 6 game match against Kasparov with a match score of 3½ to 2½ (Newborn, 2012). Now that chess had been conquered, AI research evolved to investigate more difficult game tress, such as Go and shogi (Japanese chess-like game played on a 14 by 14 board with 8 different pieces and 10 movement patterns). Recently, Google's AlphaGo program, which originally defeated Fan Hui, a professional Go player, without any handicap in 2015, defeated in 2016 Lee Sedol, a world champion Go player, in a 5 game match (Wang et al., 2016).

Expert systems and intelligent game playing programs traditionally operate within a well-defined and narrow domain. The GPS (General Problem Solving) program is one of the early instances of combining a separate reasoning mechanism with domain knowledge to perform general purpose problem solving (Newell, Shaw, & Simon, 1959). The GPS program is able to prove logic and predicate calculus problems and can solve brain teasers such as the Towers of Hanoi problem (see Figure 2), but did not advance beyond this due to the intractable nature of searching the growing decision trees. The Cyc project is an early attempt to create an ontology for common sense reasoning (Lenat & Guha, 1989). The lack of common sense or "deep" knowledge in many AI systems is cited as a significant shortcoming (Carter, 2007). Cycorp, the company that owns and researches advances to Cyc, has made their Cyc database available through their OpenCyc project.

Games are used in AI research as a well-defined and understood testbed for defining new learn-

Figure 2. Towers of Hanoi problem (showing 5 disks): move disks one at a time from Peg A to Peg C and a disk may never be placed on a smaller disk (the original problem had 64 disks)



ing, search, and other AI techniques. Once these AI-based methods are proven in a game domain, the goal then becomes to generalize these newly proven AI methods to new problem types, generalizing the solution methodology, such as cognitive modelers in adversarial game domains generalized to simulators (Walczak, 2003).

Although not a game playing program, IBM's Watson system has also caught public attention when it defeated Ken Jennings (longest winning record of 74 games) and Brad Rutter (highest payout of \$3.25 million) in the quiz show Jeopardy (Best, 2013). The Watson program at IBM is similar to Cyc in that it is a system to perform common sense reasoning using very large data ontologies. The Watson system combines cognitive modeling, data mining (knowledge acquisition), knowledge representation, machine learning, and natural language processing, to be able to answer questions intelligently over a wide variety of topics. Specialized versions of Watson are under development for business oriented systems and medicine (Ferrucci et al., 2013).

The successes of AI may be appreciated in that many of the accomplishments of AI are frequently no longer recognized as AI, but rather as "the way things are". As an example, data mining algorithms for processing big data were originally developed back in the 1980's by AI and other re-

search disciplines (Chen, Chiang, & Storey, 2012). Other more common examples are the decades of natural language processing research (starting with Chomsky's (1957) formal grammar) that resulted in commonly used products such as Dragon's Naturally Speaking program and Apple's Siri and also mobile agent research that has led to today's search engines such as Google's search engine.

WHERE TO NOW?

AI research has been progressing steadily for the past 65 years. Much of AI research still focuses on solving specific problems within narrow domains and discovering heuristics for more efficient problem solving within specific domains. New problems arise continually and AI techniques provide reliable methods for developing reasonable and reliable solutions to these problems. Examples of ongoing problems are accurately predicting stock and commodities future movements, accurately diagnosing different types of cancers or other diseases, and creating self-sufficient autonomous vehicles or robots capable of intelligently reacting to unforeseen situations. Thus, the focus on AI based solutions to problems within narrowly defined domains will continue to be an ongoing research emphasis within AI, particularly in fields

that operate under uncertain information or conditions such as medical decision support systems (Miller & Geissbuhler, 2007).

Developing more global knowledge and intelligence may result from efforts to combine AI techniques and also to embed multiple AI-based systems within a single product to increase the scope of intelligence brought to solve more generalized domains and their problems. This approach is being used in medicine by embedding AI-based solutions to very narrow and specific problems into a utilized tool like an electronic medical record, to assist physicians and other healthcare providers in decision making across a much wider range of problem types (e.g., diagnosis and treatment plans across a wide variety of illness types).

The goal of creating a machine that rivals or surpasses the intelligence and reasoning capabilities of humans is still an ongoing goal of AI. The two problems currently being tackled by AI researchers in this respect are the encapsulation and usage of common sense knowledge and creating deep knowledge or understanding of the problem and corresponding solution. The first indicates that simply having knowledge about a specific domain such as the properties of metals and materials for a bridge building AI is insufficient to solve many problems that will be faced by the system that must rely on common sense knowledge available to humans, such as the heat and lack of humidity in desert regions, or the near constant precipitation in other geographic locations. Combining common sense reasoning with domain specific knowledge is a significant problem and is related to the knowledge acquisition and representation of common sense knowledge as well as inferential learning mechanisms to know when a specific piece of common sense knowledge should be applied as part of the problem solving process.

The second goal of creating understanding, sometimes referred to as “deep knowledge,” moves AI towards truly being intelligent. Current AI systems perform very well within their speci-

fied domains by using large knowledge bases of specific domain knowledge, or ontologies in the case of Cyc and Watson. While these AI are very efficient pattern recognition machines and able to apply induction, deduction, analogy, and other types of learning mechanisms to discover answers to difficult problems, they do not understand the reason why a problem may be important to solve and are not able to explain how the discovered solution makes sense, other than reciting the various rules or knowledge base facts used in determining the recommended solution.

ETHICS AND AI

A learning machine ultimately requires the capability to re-write its own code, which is certainly possible in many current programming languages. An ethical question arises from this goal of making machines intelligent and arises out of fears implanted in the human psyche from science fiction, such as robots finding ways to circumvent the three laws (Asimov, 1942). What happens if we succeed in creating a computer entity that can outthink its human creators and is capable of modifying its own code? Due diligence is required of AI researchers to accurately test their systems to make sure they are reliable, repeatable, and accurate. Along with these measures, as the day of a truly intelligent machine approaches, researchers will also be faced with providing mechanisms to disable a possible rogue machine.

As with researchers having a responsibility to develop accurate and reliable AI systems, it is also incumbent upon professionals to utilize the best technology available for performing their tasks, and when it comes to decision making or problem solving this implies AI-based systems. Legal precedent has already been set to establish the need of rational people to utilize the best technology available (see the Hooper decision in Berner and La Lande (2007, pp. 12-13)).

FUTURE RESEARCH DIRECTIONS

As indicated above, one of the challenges for AI research is to develop an intelligent machine that can understand and make sense of the world, similar to how children absorb and new knowledge and create understanding from that knowledge (Rutkin, 2016). Future AI research needs to continue to improve upon existing research to strengthen the current results. New AI research must focus further on not only imbedding knowledge, but developing intuition and inferential machine learning methods that will enable AI systems to acquire new knowledge about novel domains and develop reasonable understanding of these domains for appropriate problem solving behaviors.

While AI successes have been widely adopted and incorporated into what is perceived by the public as common or expected practices, education efforts need to be developed to remind the public of the AI origins of much of their current technological expectations and the need for sustained support and development of new AI research. One way to help emphasize the power of AI for solving complex problems is to incorporate humans back into the loop. Partnerships between AIs and humans will continue to advance research and improve not only human abilities, but also the capabilities of the AIs working in collaboration with humans (Rutkin, 2016).

CONCLUSION

Although the idea of thinking machines may be frightening to the general public, the AI research community has succeeded in producing solutions to many difficult problems. These solutions have reduced costs, improved time utilization, improved the quality of health, and many additional outcomes that have improved overall business and life and overcome uncertainty in complex decision making. AI systems tend to operate within narrow domains, but do so at the level of human experts and frequently surpass the problem-solving

performance of human experts. A new trend of embedding multiple AI-based systems within a current information system tool allows bringing multiple types of intelligence and expertise to bear on domain problems. The future of AI is promising as ever expanding research demonstrates solutions to more and more problems.

REFERENCES

- AISB. (2016). *Loebner Prize*. Retrieved June, 15, 2016, from <http://www.aisb.org.uk/events/loebner-prize>
- Asimov, I. (1942). Runaround. *Astounding Science Fiction*, 29(1), 94–103.
- Berner, E. S., & La Lande, T. J. (2007). Overview of Clinical Decision Support Systems. In E. S. Berner (Ed.), *Clinical Decision Support Systems* (pp. 3–22). New York: Springer. doi:10.1007/978-0-387-38319-4_1
- Best, J. (2013). IBM Watson: The inside story of how the Jeopardy-winning supercomputer was born, and what it wants to do next. *TechRepublic*. Retrieved 15 July 2016 from: www.techrepublic.com/article/ibm-watson-the-inside-story-of-how-the-jeopardy-winning-supercomputer-was-born-and-what-it-wants-to-do-next/
- Carter, J. H. (2007). Design and Implementation Issues. In E. S. Berner (Ed.), *Clinical Decision Support Systems* (pp. 64–98). New York: Springer. doi:10.1007/978-0-387-38319-4_4
- Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *Management Information Systems Quarterly*, 36(4), 1165–1188.
- Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton.
- Ferrucci, D., Levas, A., Bagchi, S., Gondek, D., & Mueller, E. T. (2013). Watson: Beyond Jeopardy! *Artificial Intelligence*, 199, 93–105. doi:10.1016/j.artint.2012.06.009

- Lenat, D. B., & Guha, R. V. (1989). *Building Large Knowledge-Based Systems; Representation and Inference in The Cyc Project*. Boston: Addison-Wesley.
- Linden, A., & Fenn, J. (2003). *Understanding Gartner's hype cycles*. Strategic Analysis Report N° R-20-1971. Gartner, Inc.
- Marsland, T. A., & Campbell, M. (1982). Parallel Search of Strongly Ordered Game Trees. *ACM Computing Surveys*, 14(4), 533–551. doi:10.1145/356893.356895
- McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A Proposal for The Dartmouth Summer Research Project On Artificial Intelligence, August 31, 1955. *AI Magazine*, 27(4), 12.
- Miller, R. A., & Geissbuhler, A. (2007). Diagnostic Decision Support Systems. In E. S. Berner (Ed.), *Clinical Decision Support Systems* (pp. 99–125). New York: Springer. doi:10.1007/978-0-387-38319-4_5
- Newborn, M. (2012). *Kasparov versus Deep Blue: Computer Chess Comes of Age*. New York: Springer Science & Business Media.
- Newborn, M., & Kopec, D. (1989). Results of the Nineteenth ACM North American Computer Chess Championship. *Communications of the ACM*, 32(10), 1225–1230. doi:10.1145/67933.67939
- Newell, A., Shaw, J. C., & Simon, H. A. (1959). Report on a general problem solving program. In *IFIP Congress* (Vol. 256, p. 64). IFIP.
- Pearl, J. (1984). *Heuristics – Intelligent Search Strategies for Computer Problem Solving*. Reading, MA: Addison-Wesley.
- Rutkin, A. (2016). Anything you can do.... *New Scientist*, 229(3065), 20–21. doi:10.1016/S0262-4079(16)30457-2
- Shortliffe, E. H., Davis, R., Axline, S. G., Buchanan, B. G., Green, C. C., & Cohen, S. N. (1975). Computer-based consultations in clinical therapeutics: Explanation and rule acquisition capabilities of the MYCIN system. *Computers and Biomedical Research, an International Journal*, 8(4), 303–320. doi:10.1016/0010-4809(75)90009-9 PMID:1157471
- Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 59(236), 433–460. doi:10.1093/mind/LIX.236.433
- Walczak, S. (2003). Knowledge-based search in competitive domains. *IEEE Transactions on Knowledge and Data Engineering*, 15(3), 734–743. doi:10.1109/TKDE.2003.1198402
- Wang, F. Y., Zhang, J. J., Zheng, X., Wang, X., Yuan, Y., Dai, X.,.... Yang, L. (2016). Where does AlphaGo go: from Church-Turing thesis to AlphaGo thesis and beyond. *IEEE/CAA Journal of Automatica Sinica*, 3(2), 113–120.
- Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1), 36–45. doi:10.1145/365153.365168

ADDITIONAL READING

- Barr, A., & Feigenbaum, E. A. (Eds.). (1981, reprinted in 2014). *The Handbook of Artificial Intelligence Volume 1*. Reading, MA: Addison Wesley. (see also Volumes 2–4 of the same series)
- Chrysafiadi, K., & Virvou, M. (2013). Student modeling approaches: A literature review for the last decade. *Expert Systems with Applications*, 40(11), 4715–4729. doi:10.1016/j.eswa.2013.02.007

Code Well Academy. (2015). *Artificial Intelligence Made Easy, w/ Essential Programming*.

George, B., & Carmichael, G. (2015). *Artificial Intelligence Simplified: Understanding Basic Concepts*. CS Trends LLP.

Heaton, J. (2013). *Artificial Intelligence for Humans, Volume 1: Fundamental Algorithms*. CreateSpace Independent Publishing Platform. (see also Volume 2: Nature Inspired Algorithms and Volume 3: Deep Learning and Neural Networks)

Hu, Y., Liu, K., Zhang, X., Su, L., Ngai, E. W. T., & Liu, M. (2015). Application of evolutionary computation for rule discovery in stock algorithmic trading: A literature review. *Applied Soft Computing*, 36, 534–551. doi:10.1016/j.asoc.2015.07.008

Liao, S. H. (2005). Expert system methodologies and applications—a decade review from 1995 to 2004. *Expert Systems with Applications*, 28(1), 93–103. doi:10.1016/j.eswa.2004.08.003

Liao, S. H., Chu, P. H., & Hsiao, P. Y. (2012). Data mining techniques and applications—A decade review from 2000 to 2011. *Expert Systems with Applications*, 39(12), 11303–11311. doi:10.1016/j.eswa.2012.02.063

Medsker, L. R., & Liebowitz, J. (1993). *Design and Development of Expert Systems and Neural Networks*. Upper Saddle River, NJ: Prentice Hall.

Neri, F., & Cotta, C. (2012). Memetic algorithms and memetic computing optimization: A literature review. *Swarm and Evolutionary Computation*, 2, 1–14. doi:10.1016/j.swevo.2011.11.003

Russel, S. J., & Norvig, P. (2010). *Artificial Intelligence: A Modern Approach* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Warwick, K. (2012). *Artificial Intelligence: the basics*. New York: Routledge.

‘Watson’ computer wins at ‘Jeopardy’. (2011). *YouTube*. Retrieved from <https://www.youtube.com/watch?v=Puhs2LuO3Zc>

A

KEY TERMS AND DEFINITIONS

Artificial Neural Network: A collection of interconnected processing elements known as neurodes that mimic the electrical connectivity in the human brain to produce intelligence.

Expert System: Also known as knowledge based systems, a computer program composed of a knowledge base, often represented as if-then production rules, an inference engine, and a user interface which typically includes an explanation facility to educate users on the decision making process of the expert system.

Game Tree: A knowledge representation structure shaped like a decision tree. Nodes of the tree represent game positions possible through legal moves (from the position in the parent connecting node). The layers of the game tree expand or contract based on the possible legal moves at that point in the game. It has been estimated that the game of chess has 38^{84} possible positions in its game tree (Marsland & Campbell, 1982).

Heuristic: A rule of thumb used to reach a conclusion using only partial knowledge. While not guaranteed to be perfect, the heuristic should provide a reasonable approximation of a correct solution.

Three Laws of Robotics: Sometimes referred to as Asimov’s Laws of Robotics or simply the Three Laws are a set of science fiction principles to govern intelligent/thinking robot behavior so that no harm comes to mankind. The laws are: 1) A robot may not injure a human being or, through inaction, allow a human being to come to harm; 2) A robot must obey the orders given it by human beings except where such orders would conflict with the First Law; and 3) A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.

Artificial Intelligence Review

Amal Kilani

University of Gabes Tunisia, Tunisia

Ahmed Ben Hamida

University of Sfax, Tunisia

Habib Hamam

University of Moncton, Canada

INTRODUCTION

The study of intelligence is one of the oldest disciplines. Artificial Intelligence (AI) is a very growing and actively changing field. In this paper, we present a profound review of the AI. After defining it, we briefly cover its history and enumerate its major fields of application. Also, the test that defines an artificially intelligent system called The Turing test, is also defined and detailed. Along the way, we describe some AI tools such as Fuzzy logic, genetic algorithms and swarm intelligence. Special attention will be given to neural networks. We also present the future research directions and ethics.

BACKGROUND

Artificial intelligence (AI) may be defined as the branch of computer science that is concerned with the automation of intelligent behavior (Luger & Stubblefield, 1993). It is a research area and a field of technology that creates both software and hardware sophisticated features in order to

include virtual artificial agents. It can be divided into two categories based on thinking and acting shown in Figure 1.

Alan Turing, a British mathematician, introduced the ‘Turing test’ for intelligence, referred to the accredited test as the imitation game (Hodges, 2002). The famous test appeared in Turing’s paper, *Computing Machinery and Intelligence*, was published in October 1950 in the philosophical journal, *Mind* (Turing, 1950). In fact, this test was designed to provide a satisfactory operational definition of intelligence (Russell & Norvig, 2009).

This Turing test states four conditions for a computer to be called an intelligent machine. The first is the natural language processing (Kok et al, 1993). The second condition concerns the knowledge representation (Russell & Norvig, 2009). The third is automated reasoning (Kok et al, 1993). The computer has to be able to reason based on the knowledge that has been put in its memory. Finally, the machine must be able to learn from its environment (Kok et al, 1993). Some scientists have argued that the Turing test presents some limits such as not rating the intelligence of the machine (French, 1990).

Figure 1. AI categories

Source: Russell & Norvig, 2009

Think like Humans		Act like Humans	
Think rationally	Think like Humans	Act rationally	Act like Humans

DOI: 10.4018/978-1-5225-2255-3.ch010

The apparition of artificial intelligence was due to the inventions in electronics and other disciplines (Buchanan, 2005). The term was first coined by John McCarthy in 1956 in the conference: Artificial Intelligence, a new term to the human understanding (Stewart, 2000). The fifties saw the growth of an AI community and witnessed the opening Dartmouth Artificial Intelligence Conference and the creation of DARPA: Defense Advanced Research Projects Agency (Knight, 2006). This conference (McCarthy et al. 2006) made it possible to examine the use of computers in order to process symbols. Doubts amongst some researchers concerning the efficiency of machines began to occur in the 1960s (Coppin, 2004). In the 1970s, the AI industry went through a short era described as AI Winter where AI faced a dramatic regression (University of Washington, 2006) due to factors such as the failure of machine translation in 1966. The expert system is a computer system that imitates the decision-making ability of a human expert, was first created in the 1970s and then spread in the 1980s (Peter, 1998; Leondes, 2002). The first expert system was called DENDRAL, for Dendritic Algorithm (Bhadeshia, 2015). In the recent decades, artificial intelligence became used in several domains (Russell & Norvig, 2009; Ray, 2004; NRC, 1999). The success was due to: the increasing power of computers, the creation of new links between AI and fields working on similar problems, and a new commitment by researchers to establish mathematical methods and rigorous scientific standards (Russell & Norvig, 2009; Ray, 2004; NRC, 1999; Pamela, 2004).

ARTIFICIAL INTELLIGENCE FIELDS OF APPLICATION

Industry and Robotics

AI is very used in industry. It is thanks to robots that AI have become common in many industries. They are worldwide used: in aviation, object avoidance, transportation, and other industrial domains

(telecommunication, pharmacy, medicine, etc. ...). Artificial intelligence is a theory. Robots are factory-made as hardware. The assembly between these two is that the controlling power behind the robot is a software agent that collects data from the robots 'sensors, then make decision of what to do, and to finish, guides the effectors to act in the physical world (Niemueller & Widyadharma, 2003). Hence, a robot is an artificially intelligent machine.

Economy and Commerce

The advances in the electronic transmission of data, the migration during the past decades of so-called quants and computer wizards to Wall Street and the accessibility of inexpensive but powerful computer hardware and software (Aiken et al., 1991) helped to introduce the AI in economy. Initially knowledge-based systems were considered as tools to allow non-experts to make decisions as effectively as possible for one or more experts in a particular domain (Osorio & Sánchez, 2005). A number of researchers have studied the use of expert systems in accounting, tax and auditing (Michaelsen & Messier, 1987). Also, AI helps take a decision using multistage optimization models described in Figure 2 (Jiménez et al. 2005).

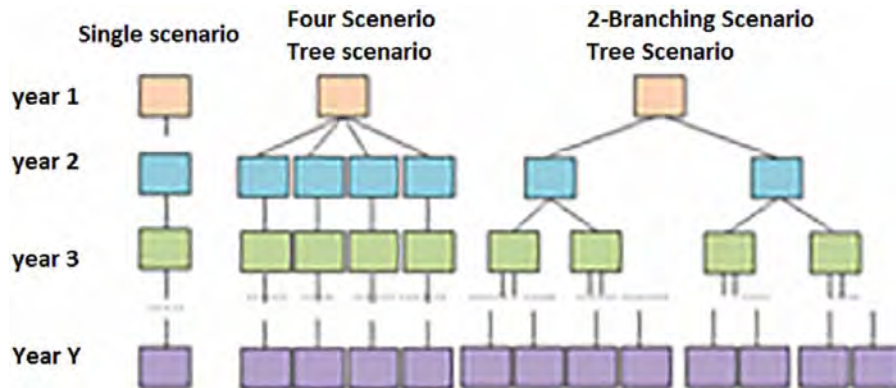
Education

AI systems are spread in the education sector as well; for example, in intelligent tutoring systems, language tutoring and etc. (Lawler & Yazdani, 1987). Some claim that in the domain of education, the procedures and the tools are various. But it is in the pedagogical techniques where the problem lies. AI can provide some solutions to this complication (Obenson, 1989) including computer-assisted education and teaching.

Information Systems

We define an information system (IS) as computer based system, which can access a diversity of

Figure 2. Scenario trees for multiperiod optimization



computer-stored or generated base information, and select to process that information as to offer particular information to help planners and management in their decision-making (Wiederhold, 1991). Artificial intelligence brings to the IS technical knowledge representation to concept bases of practical knowledge and useful learning methods to reuse the systems (Rolland & Florry, 1990).

Data Mining

In general, data mining is the study of observational information sets to find unanticipated relationships, and to summarize the information in original methods that are both comprehensible and beneficial to the data owner (Hand et al, 2001). AI and DM can expand future military information systems.

SOME ARTIFICIAL INTELLIGENCE TOOLS

In this section, some of the artificial intelligence tools is presented and special attention is given to fuzzy logic and artificial neural networks.

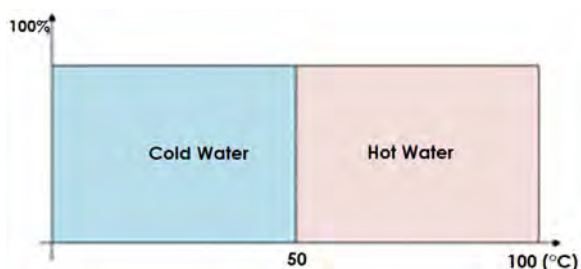
Fuzzy Logic: History and Principles

The fuzzy logic is a set of mathematical principles (Cormier, 2014) proposed by Zadeh, in 1965.

It has progressively been extended in order to handle the notion of incomplete truth (Novák et al., 1999). The variables in mathematics generally take numerical values, whereas in fuzzy logic applications, the non-numerics are frequently used to ease the expression of rules and facts (Zadeh et al. 1996).

A fuzzy set is characterized by a function which assigns to each object a grade of membership ranging between zero and one. The notions of inclusion, union, intersection, complement, relation, convexity, etc., are extended to such sets, and various properties of these notions in the context of fuzzy sets are established (Zadeh, 1965). Fuzzy logic has been practical in several fields. The high-speed train in Sendai was the first recognized application (Kosko, 1994) also in recognition of hand written symbols in Sony pocket computers; flight aid for helicopters; etc. (Bansod et Al., 2005).

Figure 3. Example of classical representation
Source: Cormier, 2014



Classical logic has two conditions: TRUE or FALSE. An example is exemplified in Figure 3, whereas another example of fuzzy variables is represented in Figure 4.

There are two techniques in fuzzy logic: the Mamdani method and the Sugeno method. The Mamdani method is the mostly used. This technique follows four steps (Cormier, 2014): Fuzzification of input variables, Rules Evaluation, Aggregation of outputs of the rules and Defuzzification.

The fuzzification step takes the crisp inputs and determines the degree to which these inputs go to each of the right fuzzy sets (Iancu, 2012).

The fuzzy rules, are a series of if and then. They state a set of fuzzy relations between inputs and outputs (Balazinska et al, 2002). For example, the possible rules for a house fan control system having one output (fan speed) and two inputs (temperature and humidity) are:

1. If x is A1 Or Y is B1 THEN z Is C1
2. If x is A2 And Y is B1 THEN z Is C2
3. If x is A3 THEN z Is C3

μ_F is called the membership function MF. It is represented by its formula. Here is a list of some famous MFs (Chen, 2013): Triangular MFs,

Figure 4. Fuzzification of the variables of house fan system

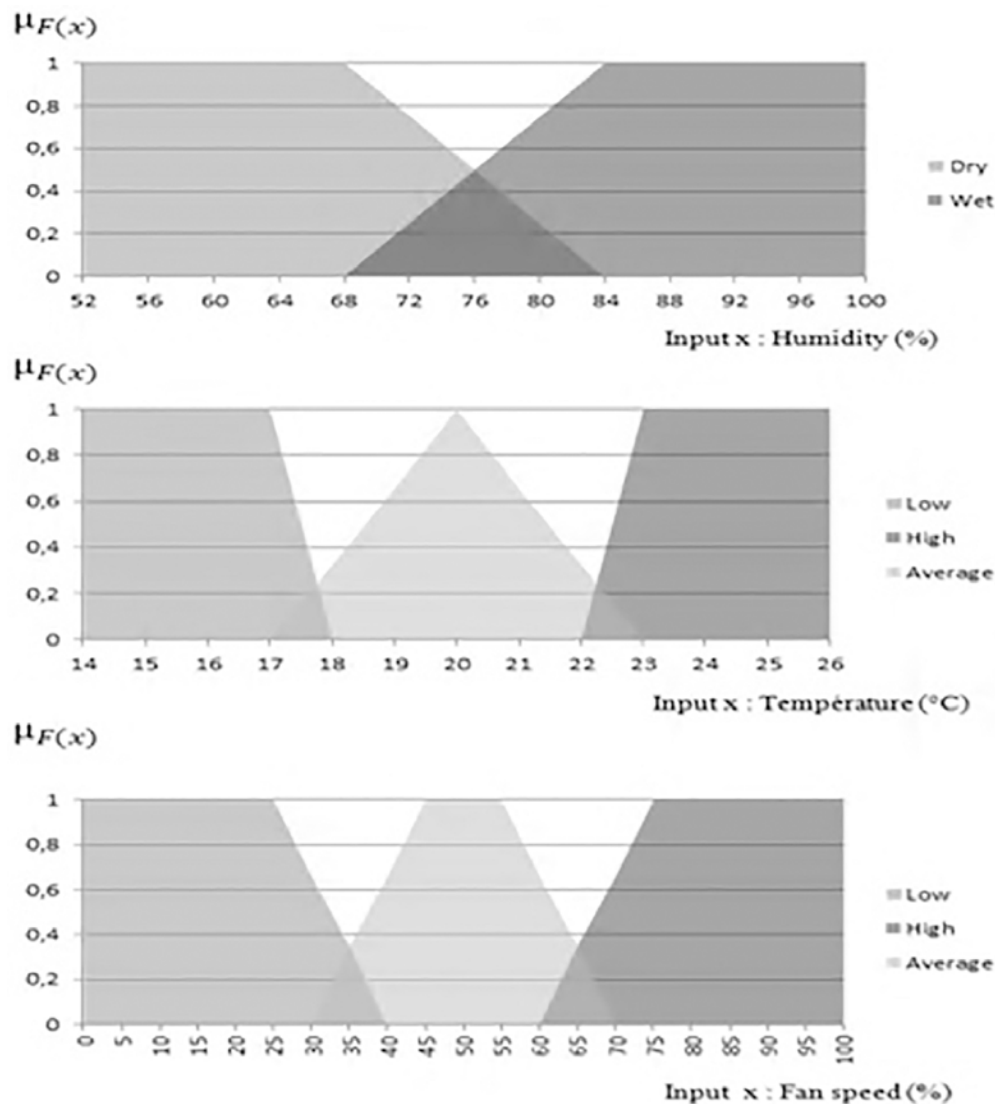
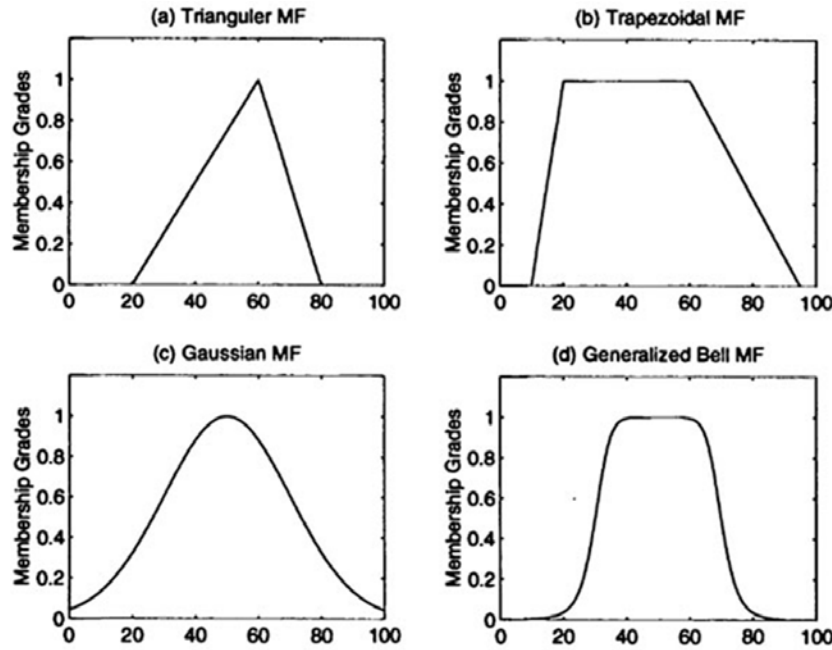


Figure 5. Membership functions of one dimension
Source: Chen, 2013



Trapezoidal MFs, Gaussian MFs, and Generalized bell MFs.

Then, using the membership values determined in fuzzification, the rules are calculated giving the compositional rule of inference. The fuzzified inputs are applied to the antecedents of the fuzzy rules. If a particular fuzzy rule has numerous antecedents, the fuzzy operator (AND or OR) is used to get a single number that characterises the outcome of the antecedent evaluation. To evaluate the disjunction of the rule antecedents, one uses the OR fuzzy operation. The classical fuzzy operation union is used by the Equation 1:

$$\mu_{A \cup B}(x) = \max \{ \mu_A(x), \mu_B(x) \} \quad (1)$$

To calculate the conjunction of the rule antecedents, the AND fuzzy operation intersection is applied using the Equation 2:

$$\mu_{A \cap B}(x) = \min \{ \mu_A(x), \mu_B(x) \} \quad (2)$$

The aggregation of outputs gives one particular fuzzy set (see Figure 6).

The Defuzzification step is to perform the following functions (Iancu, 2012):

1. A scale planning that converts the range of values of output variables into equivalent universes of discourse.
2. Defuzzification, which produces a non-fuzzy control action from an inferred fuzzy control action. There are many different approaches of defuzzification (Leekwijck & Kerre, 1999). We mention, COG (center of gravity), the AI (adaptive integration), CDD (constraint decision defuzzification), BADD (basic defuzzification distributions), COA (center of area) and ECOA (extended center of area). So, the input for the defuzzification procedure is the combined output fuzzy set and the output is a single number. The centroid method is described in Equation 3.

$$\frac{\sum_{x=a}^b \mu_F(x) \cdot x}{\sum_{x=a}^b \mu_F(x)} \quad (3)$$

Figure 6. Aggregation rules of output
Source: Cormier G., 2014

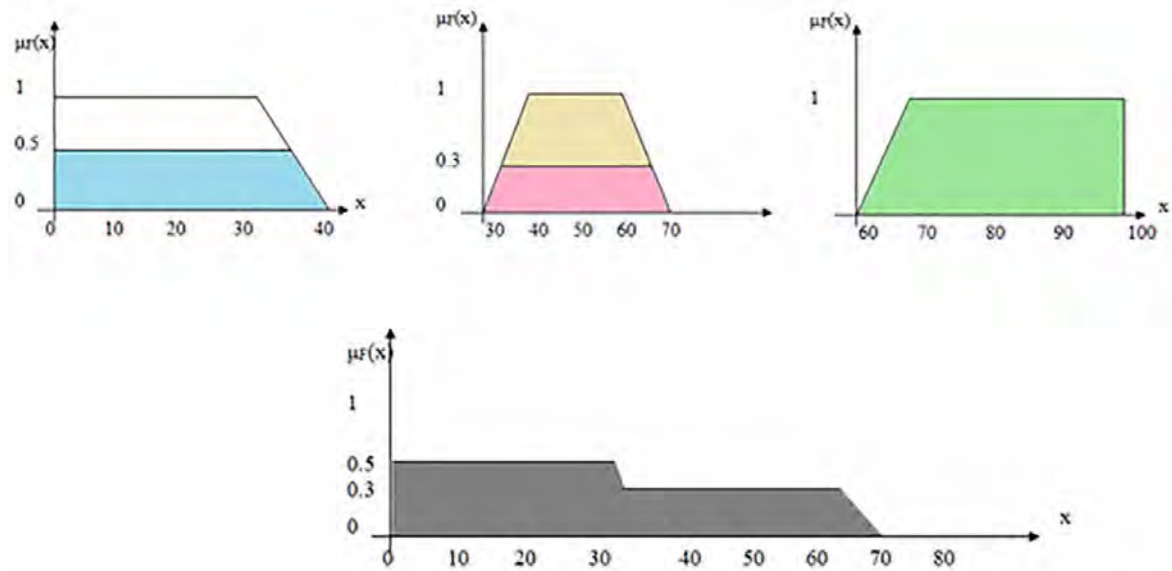
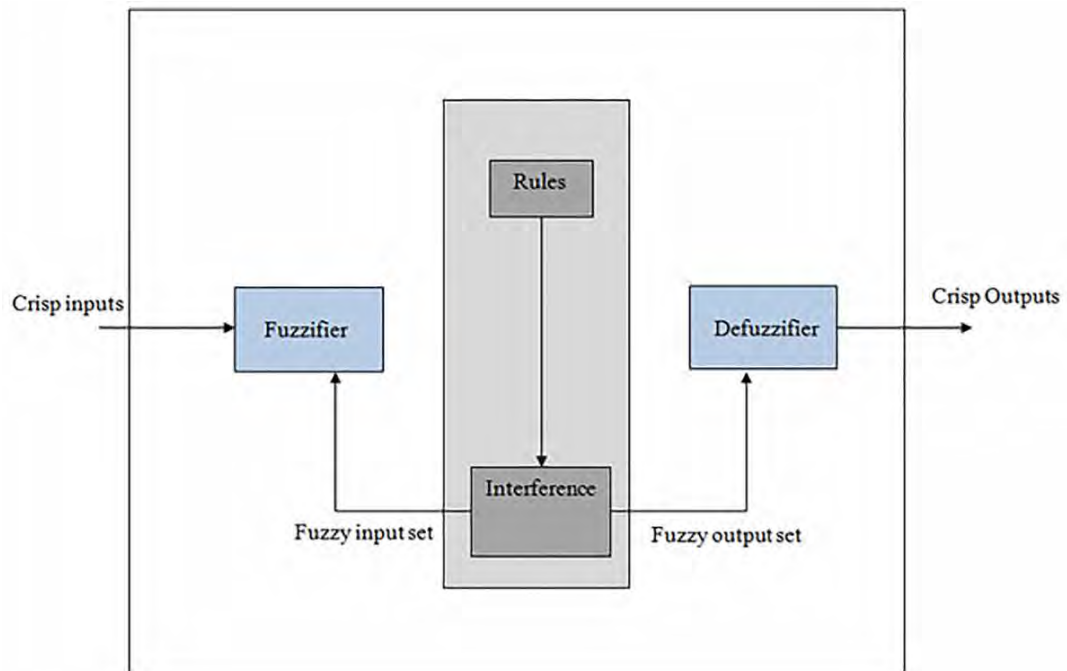


Figure 7. Fuzzy logic system



Artificial Neural Networks: History, Structure and Models

Artificial Neural Networks (ANN) are an attempt at modeling the information processing capabilities of nervous systems (Rojas, 1996). Artificial neural nets have a number of properties that make them an attractive substitute to traditional problem-solving techniques. The two important alternatives to using neural nets are to develop an algorithmic solution, and to use an expert system. The Expert systems are well used by different companies (Gallant, 1993). Neural networks have a notable capability to extract forms and identify trends that are too difficult to be noticed by either humans or other computer systems. A trained neural network can be thought of as an expert in the kind of information it has been given to analyze. This expert can then be used in providing projections given new situations of interest and response what if'' questions. In 1943, the first mathematical model of the biological neuron is suggested by Warren McCulloch and Walter Pitts (McCulloch & Pitts, 1943).

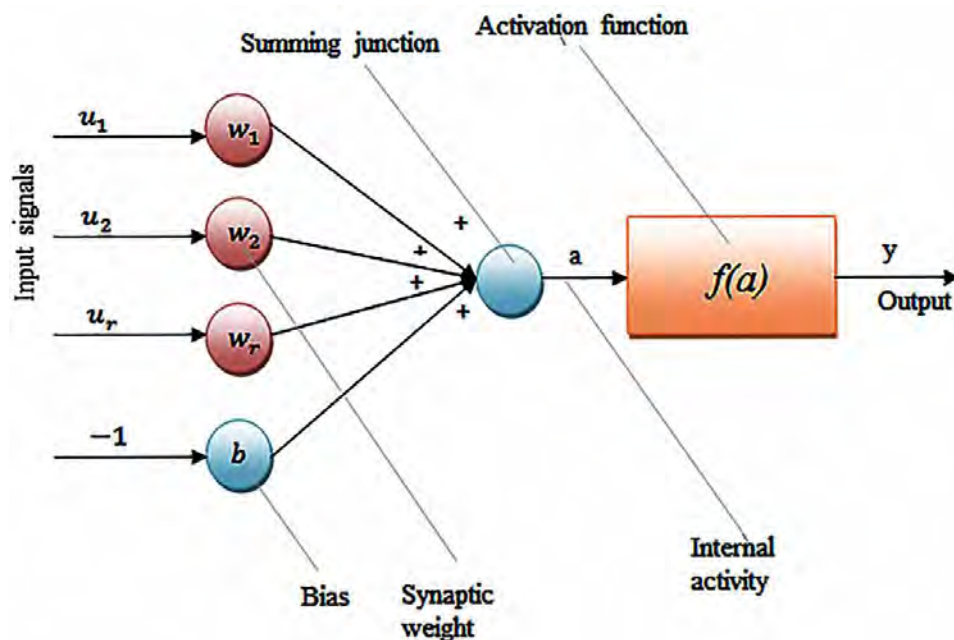
A neuron principally consists of an integrator that makes the weighted sum of its inputs. The result of this sum is at that time transformed by a transfer (activation) function that produces the output of a neuron (McCulloch & Pitts, 1943). A model of the artificial neuron is represented in Figure 8.

An ANN involves of a set of processing elements, also identified as neurons or nodes, which are interconnected. It can be described as a directed graph in which each node i implements a transfer function f_i as follows:

$$y_i = f_i \left(\sum_{j=1}^n w_{ij} x_j - \theta_i \right) \quad (4)$$

a is the internal activity, b is the bias, the $u_1, u_2 \dots u_r$ are the input signals, $w_1, w_2 \dots w_r$ are weights, and the y is the output. The function f is called the activation or squashing function, which maps the internal activity a to a closed interval $[0, 1]$ or alternatively $[-1, 1]$. An example of the

Figure 8. A simple model of an artificial neuron
Source: Ali & Schmid, 2004



activation function is the uni-polar sigmoid function, which has the expression in Equation 5.

$$f(a) = \frac{1}{1 - e^{-a}} \quad (5)$$

There are other activation functions (Karlikand & Olgac, 2010). Also, there are different neural architectures: the single-layer feed-forward networks, multilayer feed-forward networks, and the recurrent networks (Hayken, 2008) presented in Figure 9 (Jain et al., 1996).

The disadvantage of an ANN is its black box quality; it presents results without any description of the procedure. Additional current weaknesses

of ANN consider also the computational resources that are essential and the lack of standard software (Sargent, 2001).

Fuzzy-Neural Network

Neuro-fuzzy representation is concerned with the extraction of models from numerical data representing the behavioural dynamics of a system. Models like these have important functioning predictions of system performance and system control (Jang & Sun, 1993; Takagi & Hayashi, 1988; Chai et al, 2009). One way to combine the two approaches is to use neural networks to replace each component of a system of fuzzy control. The other is the serial or the parallel connections.

Figure 9. Different architectures of ANN

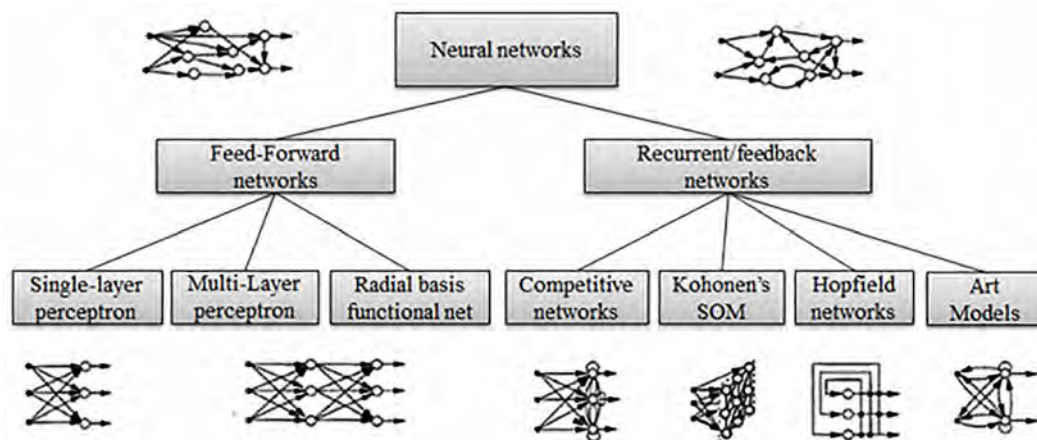
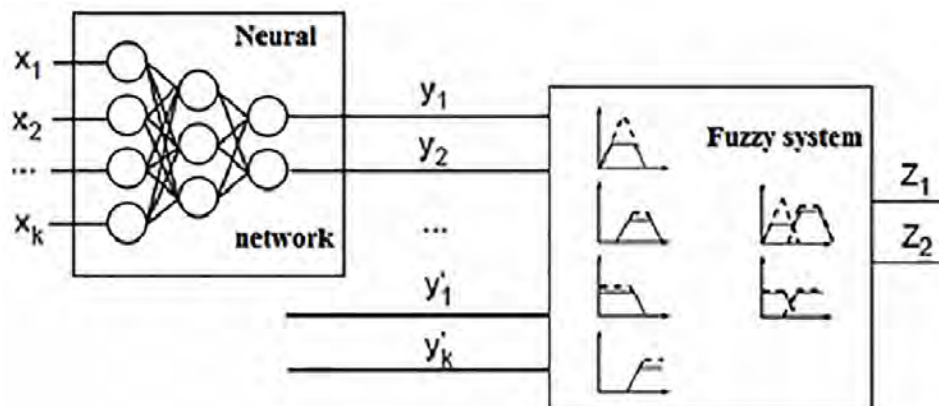


Figure 10. Example of series connection of a neural network and a fuzzy system



Genetic Algorithms

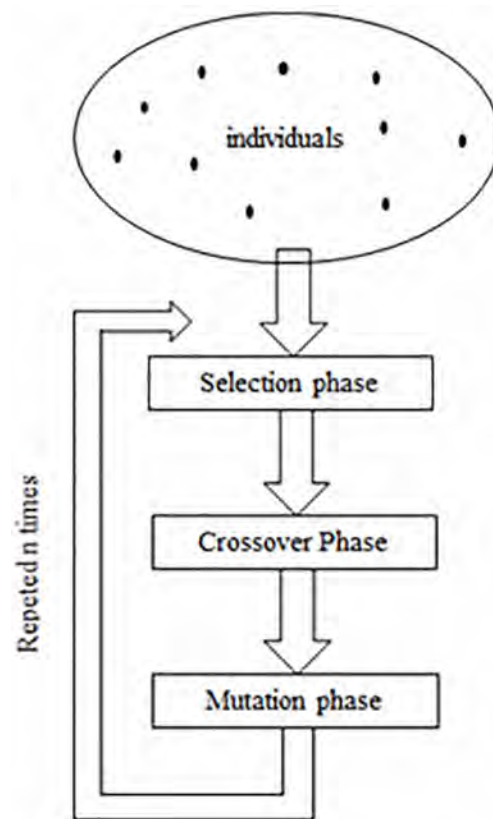
Genetic algorithms are algorithms for optimization (Montana & Davis, 1989). In a genetic algorithm, a population of candidate solutions (called individuals, creatures, or phenotypes) to an optimization problem evolves into better or more capable solutions. Each one of these candidate solutions has a given number of properties (its chromosomes or genotype), which can be mutated, modified, and altered. Typically, the solutions are represented in binary, but other encodings are also possible (Darrell, 1994). During each generation the structures in the present population are rated for their efficiency as solutions, and on the basis of these evaluations, a new population of candidate structures is produced by means of particular genetic operators such as mutation, cross over and reproduction (Oke, 2008). The steps are described in Figure 11 (Orsay University, 2013).

Simulated Annealing

The goal behind creating Simulated annealing (SA) is the global optimization problem of locating an excellent approximation of the optimum of a function in a large search space. The travel salesman problem is one of classic optimization problems. Scott Kirkpatrick, C. Daniel Gelatt, and Mario P. Vecchi used the algorithm on traveling salesman problems, to test the power of simulated annealing (Kirkpatrick et al., 1983). It has also been practical on the problem of finding the ground state of a spin glass (SG). (Soukoulis, 1985; Grest et al. 1986; Guo, 1990). For a finite problem, the possibility that the simulated annealing algorithm terminates with a global optimal solution approaches 1 as the annealing schedule is extended (Granville, 1994).

Thanks to its robustness, SA turned out to be efficient in many fields such as the training of ANNs with structured learning (Wiley & Sons Inc., 1995). Hence, ANN does not need a training set; since the ANN gradually learns new skills or improves existing ones by earlier experience (Ledesma et al. 2013).

Figure 11. Genetic algorithms steps



Swarm Intelligence

Swarm intelligence is an artificial intelligence discipline, (Blum & Li, 2008; Beni & Wang, 1989). It is designed by intelligent multi-agent systems, and the design is inspired by the behavior of some social insects (Blum & Li, 2008) and animals. The particle swarm optimization (PSO) has the purpose of managing problems in order to find a best solution that can be represented as a surface point or in an n-dimensional space (Parsopoulos & Vrahatis, 2002; Clerc, 2006). The bees algorithm is another algorithm designed by Pham and his co-workers in 2005 (Pham et al., 2005) and has gradually enhance its power in the following years (Pham & Castellani, 2009).

FUTURE RESEARCH DIRECTIONS

Future of AI

Looking into the future, there is a real possibility that AI will introduce extremely intelligent artificial agents. This would be revolutionary as we could have machine assistance in every field. AI machines will have certain capabilities such as: Self-improvement: intelligence amplification, strategy development: planning, forecasting, prioritizing and social abilities: social and psychological modeling, manipulation, rhetorical persuasive ability.

Ethics of AI

The potential existence of a machine that could exceed the human intelligence has created some ethical issues. For example, who takes the blame, when an AI system fails at its assigned task: the programmers, or the end-users (Gelbukh & Monroy, 2005)? Also, some scientists, such as Dr. Hugo de Garis of Utah State University, feel that the artificial brains will not be controllable (Roberts, 2014). Moreover, some say that people may lose their jobs, and be deprived of their personal lives (Cormier, 2014). According to (Singer & Friedman 2014), privacy is certainly serious issue in AI fields. More ethics span law, public policy, professional ethics, and philosophical ethics, and will require expertise from computer scientists, legal experts, political scientists, and ethicists.

CONCLUSION

Scientists' efforts and quest to solve the riddle of the human brain's intelligence have been documented throughout history. Artificial intelligence is an attempt to create entities as intelligent as humans or more. We have defined in this paper

the Artificial Intelligence, taken a short look at its history and listed the domains of its applications. We also, detail well known algorithms that present a great deal of interest in the optimization in many domains. Also, some tools of AI are presented.

REFERENCES

- Abid, A., & Schmid, C. (2004). System Identification Using Neural Networks, in Topic 6.43 Control Systems, Robotics and Automation. In H. Unbehauen (Ed.), *Encyclopedia of Life Support Systems (EOLSS)*. Oxford, UK: Eolss Publishers.
- Aiken, M., Liu-Sheng, O., & Vogel, D. (1991) Integrating Expert Systems With Group Decision Support Systems. *ACM Transactions on Information Systems*, 9(1).
- Balazinska, M., Czogala, E., Jemielniak, K., & Leski, J. (2002). Tool condition monitoring using artificial intelligence methods. *Engineering Applications of Artificial Intelligence*, 15(1), 73–80. doi:10.1016/S0952-1976(02)00004-0
- Bansod, N. A., Kulkarni M., & Patil S.H. (2005). Soft Computing- A Fuzzy Logic Approach. In *Soft Computing*. Allied Publishers.
- Beni, G., & Wang, J. (1989) Swarm Intelligence in Cellular Robotic Systems. *Proceed. NATO Advanced Workshop on Robots and Biological Systems*, Tuscany, Italy
- Bhadeshia, H. K. D. H. (2015). *Neural Networks in Materials Science*. University of Cambridge, Department of Materials Science and Metallurgy.
- Blum, C., & Li. (2008). *Swarm intelligence in optimization*. Springer.
- Buchanan, B. G. (2005). A (Very) Brief History of Artificial Intelligence. *Artificial Intelligence Magazine*, 26(4).

- Cagan, J., Grossmann, I. E., & Hooker, J. (1994). *Combining Artificial Intelligence and Optimization in Engineering Design*. Engineering Design Research Center, Carnegie Mellon University, Working Paper.
- Chai, Y., Jia, L., & Zhang, Z. (2009). Reference Mamdani Model based Adaptive Neural Fuzzy Inference System and its Application. *World Academy of Science, Engineering and Technology*, 3.
- Chen, B. (2013). *Fuzzy Logic Membership Function* [Blog post]. Retrieved from www.bindichen.co.uk/post/AI/fuzzy-inference-membership-function.html
- Coppin, B. (2004). *Artificial Intelligence Illuminated*. Sudbury, MA: Jones and Bartlett publishers.
- Cormier, G. (2014). *GELE5313*, University of Moncton, Canada. Retrieved from www8.umoncton.ca/umcm-cormier_gabriel/Asservissements/GELE5313_Notes11.pdf
- Cormier, G. (2014). *GIND5439* [PowerPoint Slides]. Retrieved from www8.umoncton.ca/umcm-cormier_gabriel/SystemesIntelligents/GIND5439_Chapitre1.pdf
- Darrell, W. (1994). A genetic algorithm tutorial. *Statistics and Computing*, 4(2), 65–85. doi:10.1007/BF00175354
- French, R. M. (1990). Sub cognition and the Limits of the Turing Test. *Mind*, 99.
- Gallant, S. I. (1993). *Neural network learning and expert system*. Cambridge, MA: MIT Press.
- Granville, V., Krivanek, M., & Rasson, J. P. (1994). Simulated annealing: A proof of convergence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(6), 652–656. doi:10.1109/34.295910
- Grest, G. S., Soukoulis, C. M., & Levin, K. (1986). Comparative Monte Carlo and mean-field studies of random-field Ising Systems. *Physical Review*.
- Guo, H., Zuckermann, M., Harris, R., & Grant, M. (1990). A Fast Algorithm for Simulated Annealing. Centre for the Physics of Materials, Department of Physics, McGill University.
- Hand, D., Mannila, H., & Smyth, P. (2001). *Principles of Data Mining*. Academic Press.
- Hayken, S. (2008). *Neural Networks: A Comprehensive Foundation*. Hamilton, Ontario, Canada: McMaster University.
- Hodges, A. (2002). *Alan Turing and the Turing Test*. Retrieved from www.turing.org.uk/publications/testbook.html
- Holland. (1975). *Adaptation in Natural and Artificial Systems*. University of Michigan Press.
- Iancu, I. (2012). *A Mamdani Type Fuzzy Logic Controller*. University of Craiova Romania.
- Jain, A. K., Mao, J., & Muhiuddin, K. M. (1996). *Artificial neural networks: A tutorial*. IBM Almaden Research Center.
- Jang, J. S. R., & Sun, C. T. (1993). Functional equivalence between radial basis function networks and fuzzy inference systems. *IEEE Transactions on Neural Networks*, 4(1), 156–159. doi:10.1109/72.182710 PMID:18267716
- Jiménez, A., Rodríguez, L., Mateos, A., & Ríos-Insua, S. (2005). Selection of Segments to be Sourced from Low Cost Countries for a Global Industrial Equipment Manufacturer based on a Multi-Attribute Decision Support System, Technical University of Madrid, Artificial Intelligence Department, Campus de Montegancedo. *Research in Computing Science*, 01(17), 3–1.
- Kennedy, J., Kennedy, J. F., & Eberhart, R. C. (2001). *Swarm intelligence*. Morgan Kaufman Publishers.
- Kirkpatrick, S., Gelatt, C. D., Jr., & Vecchi, M. P. (1983). Optimization by Simulated Annealing. *Science*, 220(4598), 671–680.

- Knight, H. (2006). *Early Artificial Intelligence Projects, A Student Perspective* [Blog post]. Retrieved from projects.csail.mit.edu/films/aifilms/AIFilms.html
- Kok, J. N., Egbert, J., Boers, W., Kusters, W. A., & Putten, P. V. D. (1993). Artificial intelligence: Definition, trends, techniques and cases. Encyclopedia of Life Support Systems.
- Kosko, B. (1994). *Fuzzy Thinking: The New Science of Fuzzy Logic*. Hyperion.
- Kurzweil, R. (2004). *The Singularity is Near*. Viking.
- Lawler, R. W., & Yazdani, M. (1987). *Artificial Intelligence and Education: Learning environments and tutoring*. Norwood, NJ: Ablex publishing.
- Ledesma, S., Ruiz, J., & Garcia, G. (2013). Simulated Annealing Evolution. *Simulated Annealing – Advances, Applications and Hybridizations*, 201-218.
- Lindsay, R., Buchanan, B. G., Feigenbaum, E. A., & Lederberg, J. (1993). DENDRAL: A case study of the first expert system for scientific hypothesis formation. *Artificial Intelligence*, 61(2), 209–261. doi:10.1016/0004-3702(93)90068-M
- Luger, G. F., & Stubblefield, W. A. (1993). *Artificial intelligence: Structures and strategies for complex problem solving*. Benjamin/Cummings Pub. Co.
- McCarthy, J., & Hayes, P. J. (1969). *Some Philosophical Problems from the Standpoint of Artificial Intelligent*. Computer Science Department Stanford University.
- McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. *AI Magazine*, 27.
- McCulloch, W. S., & Pitts, W. (1943). A Logical Calculus of the Ideas Immanent in Nervous Activity. *The Bulletin of Mathematical Biophysics*, 5(4), 115–133. doi:10.1007/BF02478259
- Michaelson, R. & Messier. (1987). Expert systems in auditing: The state of the art. *Auditing: A Journal of Theory and Practice*, 7(1), 94-105.
- Michaelson, R. & Messier. (1987). Expert Systems in Taxation. *The Journal of the American Taxation Association*, 7-21.
- Montana, D. J., & Davis, L. (1989). Training Feed-forward Neural Networks Using Genetic Algorithms, BBN Systems and Technologies Corp, Cambridge, MA. *Machine Learning*, 762–767.
- Newell, A. (1982). *Intellectual Issues in the History of Artificial Intelligence*. Carnegie Mellon University.
- Niemueller, T., & Widyadharma, S. (2003). *Artificial Intelligence – An Introduction to Robotics*. Academic Press.
- Novák, V., Perfilieva, I., & Močkoř, J. (1999). *Mathematical principles of fuzzy logic*. Dordrecht: Kluwer Academic. doi:10.1007/978-1-4615-5217-8
- NRC. (1999). Developments in Artificial Intelligence. In *Funding a Revolution: Government Support for Computing Research*. National Academy Press.
- Oke, S. A. (2008). A Literature Review on Artificial Intelligence. *International Journal of Information and Management Sciences*, 19(4), 535–570.
- Orsay University. (2013). *LIMSI – CNRS* [Blog post]. Retrieved from www.perso.limsi.fr/jps/enseignement/tutoriels/pcd/3.genetique/
- Parizeau, M. (2004). *Réseaux De Neurones*. LAVAL University.
- Parsopoulos, K. E., & Vrahatis, M. N. (2002). Recent Approaches to Global Optimization Problems Through Particle Swarm Optimization. *Natural Computing*, 1(2-3), 235–306. doi:10.1023/A:1016568309421
- Peter, J. (1998). *Introduction to Expert Systems* (3rd ed.), Addison Wesley.

- Pham, D.T., & Castellani, M. (2009). The Bees Algorithm – Modelling Foraging Behaviour to Solve Continuous Optimisation Problems. *Proc. ImechE, Part C*, 223(12).
- Pham, D. T., Ghanbarzadeh, A., Koc, E., Otri, S., Rahim, S., & Zaidi, M. (2005). *The Bees Algorithm, Technical Note*. Manufacturing Engineering Centre, Cardiff University.
- Phillip, O. (1989). *Education et intelligence artificielle ou la transmission automatique et intelligente du savoir. Un exemple en recherche documentaire. International Congres: éducation et informatique vers une coopération internationale renforcée*. UNSECO.
- Prince, N. N. (2011). Real – World Applications of Neural Network. *International Conference on Advancements in Information Technology*. IACSIT Press.
- Rapaport, W. J. (2005). The Turing Test, State University of New York at Buffalo, Buffalo, NY 14260-2000. Encyclopedia of Language and Linguistics (2nd ed.). Academic Press.
- Roberts, E. (2014). Retrieved from Stanford University WebCourses: cs.stanford.edu/people/eroberts/courses/soco/projects/2004-05/ai/ai-ethics.html
- Rojas, R. (1996). *Neural Networks*. Berlin: Springer-Verlag. doi:10.1007/978-3-642-61068-4
- Rolland, C., & Florry, A. (1990). La conception des systèmes d'information, état de l'art et perspectives. 90th congrès of the information association of the organisations and information system and decision, 3-40.
- Russell, S. J., & Norvig, P. (2009). *Artificial Intelligence: A Modern Approach* (3rd ed.). Academic Press.
- Russell, S. J., & Peter, N. (2003). *Artificial Intelligence: A Modern Approach* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.
- Sargent, D. J. (2001). *Comparison of artificial neural networks with other statistical approach*. Wiley Online Library.
- Singer, P. W., & Friedman, A. (2014). *Cybersecurity: What Everyone Needs to Know*. New York: Oxford University Press.
- Stewart, W. (2000). *Dartmouth Artificial Intelligence (AI) Conference* [Blog post]. Retrieved from www.livinginternet.com/i/ii_ai.htm
- Takagi, H., & Hayashi, I. (1988). Artificial neural network-driven fuzzy reasoning. *Proc. Int. Workshop Fuzzy System Applications*.
- Turing, A.M. (1950). Computing machinery and intelligence. *Mind: A Quarterly Review of Psychology and Philosophy*.
- University of Washington (2006). *The History of Artificial Intelligence*. CSEP 590A.
- Wiederhold, G. (1991). The role of artificial intelligence in information systems, methodologies for intelligent systems. *Lecture Notes in Computer Science*, 542, 38–51. doi:10.1007/3-540-54563-8_68
- Wiley, J. (1995). *Sons Inc*. New York: Masters T. Advanced Algorithms for Neural Networks.
- Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, 8(3), 338-353. Retrieved from: <http://www.sciencedirect.com/science/article/pii/S001999586590241X>
- Zadeh, L. A. (1996). *Fuzzy Sets, Fuzzy Logic, Fuzzy Systems*. World Scientific Press. doi:10.1142/2895

ADDITIONAL READING

- Bart, B., Viaene, S., Van den Poel, D., Vanthienen, J., & Dedene, G. (2002). Bayesian Neural Network Learning for Repeat Purchase Modelling in Direct Marketing. *European Journal of Operational Research*, 138(1), 191–211. doi:10.1016/S0377-2217(01)00129-1

Cox, E. (Oct. 1992). Fuzzy fundamentals. *Spectrum, IEEE*, 29:10. pp. 58–61.

Dreyfus, H. L. (1972). What Computers Can't do. *Harper and Row*, Publisher: PhilPapers

Genova, J. (1994). Turings Sexual Guessing Game. *Social Epistemology*, 8(4), 313–326. doi:10.1080/02691729408578758

Hayes, P., & Ford, K. Turing Test Considered Harmful (1995), IJCAI'95 Proceedings of the 14th international joint conference on Artificial intelligence - Volume 1, p 972-977

Hezarkhani, T. (2012). A hybrid neural networks-fuzzy logic-genetic algorithm for grade estimation. *Computers & Geosciences*, 42, 18–27. doi:10.1016/j.cageo.2012.02.004 PMID:25540468

Ronald, E., & Sipper, M. (2001). Intelligence is Not Enough: On the Socialization of Talking Machines. *Minds and Machines*, 11(4), 567–576. doi:10.1023/A:1011876818541

Siegelmann, H. T., & Sontag, E. D. (1991). Turing computability with neural nets. [GBR]. *Applied Mathematics Letters*, 4(6), 77–80. doi:10.1016/0893-9659(91)90080-F

A

KEY TERMS AND DEFINITIONS

Artificial: Description of an object made by human beings and not a naturally.

Dendritic Algorithm: A classification algorithm based on the functioning of natural immune dendritic cells.

Intelligence: The ability to acquire, learn and apply knowledge and skills.

Spin Glass: An optimization problem, NP-complete problem (non-deterministic polynomial time complete) in three dimensions.

Artificial Neural Networks

Steven Walczak

University of South Florida, USA

INTRODUCTION

Artificial neural networks (ANNs) is a subfield of machine learning within the research domain of artificial intelligence (see Artificial Intelligence, this volume). Research in developing ANNs started after McCulloch and Pitts (1943) proposed a mathematical model of neuronal activity in the brain and Hebb (1949) created a reinforcement based learning mechanism to explain learning in the human brain. Rosenblatt (1958) then created a computational model of brain processing elements called perceptrons and ANN research started in earnest. The goal of ANN research is to develop machine learning systems that are based on a biological model of the brain, specifically the bioelectrical activity of the neurons in the brain.

ANNs are a popular solution method in numerous domains including: business (Tkáč & Verner, 2016; Wong, Lai, & Lam, 2000), engineering (Ali et al., 2015; Bansal, 2006), and medicine (Reggia, 1993; Yardimci, 2009). Research and development with ANNs continues to be highly productive with the quantity of articles published in this subfield increasing annually. Using the search query *artificial neural network* on a university article database search produced 27,736 articles from 1985 to 2000 and 203,328 from 2001 to 2016 with over 51 percent of the publications appearing from 2011 to 2016; indicating a tenfold increase in ANN articles published over the same amount of time (16 years) and the trend continuing to accelerate.

It is important to understand the terminology used to discuss ANN architectures. A sample ANN architecture for a supervised learning multi-layer perceptron is shown in Figure 1. Modern ANNs are composed of:

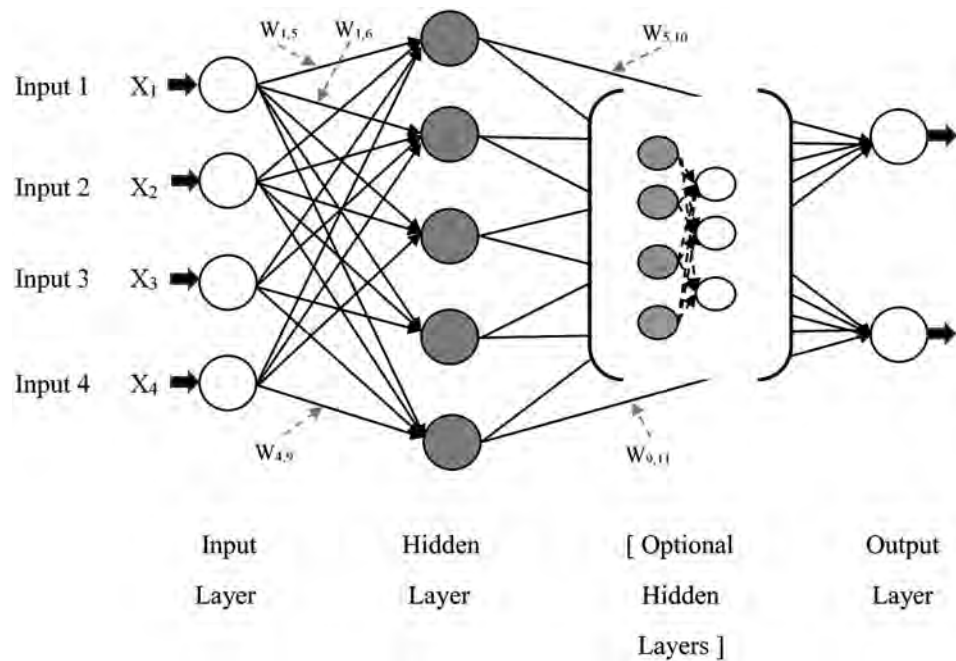
- A layer of input elements also called the input vector, representing independent variables,
- Optionally, one or more hidden processing layers,
- Weighted connections between nodes in adjacent layers, and
- An output layer of one or more elements, representing the dependent variable(s).

Every processing element or neurode in a layer is connected to all processing elements in the next layer, with input neurodes connected to hidden layer neurodes etcetera, until the neurodes in the last hidden layer are connected to the output layer neurodes. These connections all carry a value, commonly called a weight, that is adjusted to permit learning. It is possible for an neurode to not be fully connected to the subsequent layer, but to be connected selectively to one or more neurodes in the following layer. Some ANN architectures also have weighted connections from a layer to not only the next layer, but also to one or more subsequent layers of neurodes.

What types of research problems are amenable to an ANN approach? Essentially, ANNs are intelligent pattern recognition machines. Thus, any problem which may be defined as a pattern recognition problem is suitable for ANN solutions. This includes all types of classification problems and also most prediction problems, such as time-series forecasting or medical diagnosis. Additional research has shown that ANNs may be used as a tool for evaluating medical or business decision making heuristics (Walczak, 2008).

DOI: 10.4018/978-1-5225-2255-3.ch011

Figure 1. Supervised learning ANN architecture (only a few connection weights shown)



BACKGROUND

Research on ANNs and development of ANN applications started out strongly throughout the 1960's following Rosenblatt's (1958) discovery of perceptrons. Early work in ANNs used perceptrons and multiple layer perceptrons, that used a supervised learning rule to adjust connection weights based on the difference between the desired output and the output produced by the perceptron. Other learning rules were also developed including the ADALINE (ADaptive LINEar Element) and MADALINE (Multiple Adaline) (Widrow, 1964).

Minsky and Papert (1969) demonstrated limitations of simple perceptrons, specifically difficulty in solving the exclusive or problem. A simple perceptron is defined as an ANN where the input layer is connected directly to the output layer with no intervening layers. The backpropagation algorithm, also referred to as a multi-layer perceptron, was created in 1974 (Werbos, 1974), enabling rapid training of multilayer perceptrons (MLPs) and thus overcoming the limitations identified by Minsky and Papert. In the late 1970's and early

1980's ANN research exploded and has continued to increase ever since.

Processing Criteria of ANNs

As mentioned in the introduction, ANNs are composed of processing elements called neurones arranged in layers that are connected. In addition to the physical elements of the ANN, there are several soft processing criteria that are also important to understand. The learning rule is the algorithm that permits the ANN to learn, typically through updating of connection weights. The learning rate is how fast and how large updates are made to connection weights. Typically the learning rate starts out large enabling rapid changes in the weights and subsequent output values. The rate gradually decreases over time so that changes become smaller and smaller. A random adjustment of the rate upward is randomly performed to prevent the ANN from settling into a local minimum in the solution surface and to move towards the global minimum.

As noted in Figure 1, all neurodes past the input layer will have multiple input connections from the previous layer. Each neurode uses an aggregation function to collect the weighted input values. Common functions may include: summation, minimum value, maximum value, mean, mode, or median. Other functions are possible and may be defined by the ANN developer.

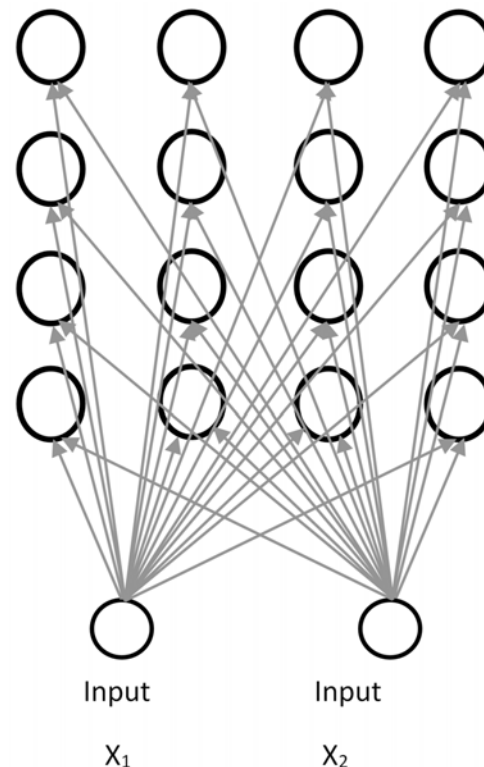
Because ANNs originally modelled the neuronal processing of the human brain, each neurode in the ANN also has a transfer function to mimic the firing or activation of brain neurons. A typical transfer function is sigmoidal, hyperbolic tangent, or Gaussian but again any function may be defined by the ANN developer to use for transferring values to the next layer. All of these soft processing criteria are involved in producing learning within the ANN.

Types of Learning

ANNs are capable of producing complex models of highly nonlinear solution surfaces through machine learning. Learning in ANNs is performed through adjustment of the connection weights. The algorithms for how to select the update values and updating the weights is an ongoing area of research in the study of ANNs. Learning algorithms for ANNs are commonly classified as being either unsupervised learning or supervised learning methods.

Unsupervised learning implies that the solution to a problem is already contained within the data being used to train the ANN. This type of learning is typically used in image classification research, such as detecting anomalies in medical images, analyzing seismographic data, and other visually and auditorily oriented data. A few examples of unsupervised learning algorithms are the self-organizing map (SOM) sometimes called Kohonen networks (Kohonen, 1990) and adaptive resonance theory (ART) (Carpenter & Grossberg, 1987). A small sample SOM network is shown in Figure 2. Since unsupervised learning training networks do not require a known value to

Figure 2. A small, 4X4 SOM network



calculate their error, they compare their current state against the state of the input vector and adjust weights around the current Kohonen neurode that is the best matching unit. Thus a neighborhood of weights gets adjusted with each training sample.

The advantage of unsupervised learning over supervised learning is primarily in the lack of a requirement for data that already has a correct classification as the unsupervised ANN learns the proper classification strictly from the input data values. A disadvantage from supervised learning is that the solution is portrayed in the Kohonen network or other neurode network attached to the input nodes and its values are left up to the interpretation of the researcher.

Supervised learning indicates that a known solution to the problem must already exist for all training data examples. Thus, supervised learning is popular for solving problems when historic data sets with known outcomes exist. Supervised

learning may be used to perform classification, as with unsupervised learning, but may also be used to create predictive models, such as time series forecasting or medical diagnosis. The most well-known supervised learning rule algorithm is backpropagation (Werbos, 1974). Backpropagation trained ANNs have been shown to be a universal approximators (Hornik, Stinchcombe, & White, 1989; White, 1990), meaning they can heuristically approximate any function or solution surface.

Research on how to improve the backpropagation algorithm was rampant in the late 1980's and early 1990's and continues today. Various algorithms other than backpropagation now exist for performing supervised learning and include radial basis functions (Lowe & Broomhead, 1988), learning vector quantization (Kohonen, 1988), probabilistic (Specht, 1990), general regression (Specht, 1991), and fuzzy ARTMAP (Carpenter et al., 1992). Each learning rule attempts to address a perceived shortcoming of the backpropagation algorithm, such as radial basis function trained ANNs being more resilient when only small or incomplete training examples are available (Barnard & Wessels, 1992) and general regression ANNs converging faster when a very large number of training examples are available (Specht, 1991).

BUILDING ANNs

Theoretical research in ANNs typically examines ways to speed up learning through the development of new learning algorithms or ANN architecture schemas. Applied research investigates the utilization of ANNs for solving difficult domain problems.

For applied research, ANNs must be developed and tested against existing solution methods. As indicated above, ANN designers face numerous decision to determine the optimal ANN design, both physical architecture and soft processing criteria. Numerous tutorials and guidelines exist (see Additional Readings) and therefore a detailed

description of ANN development will not be done here, but a quick set of decision considerations needed is given:

1. Based on the type of problem, determine supervised or unsupervised methodology.
2. Based on specifics of problem and data available, select learning algorithm to train the ANN (e.g., backpropagation or radial basis function).
 - a. Multiple algorithms may modeled independently.
 - b. Use comparative analysis to select the optimal performing ANN algorithm
3. Determine ANN architecture.
 - a. Number of layers (if applicable, for some supervised learning algorithms).
 - i. The number of layers gives insight into the complexity of the solution surface being modeled.
 - b. Number of nodes per layer.
 - i. Too few nodes will not allow the ANN to adequately model complex solution surfaces.
 - ii. Too many nodes will enable the ANN to memorize the training data and may produce suboptimal generalization performance.
4. Specify (if appropriate for learning algorithm) the aggregation function, transfer function, and learning rate.
5. Train the ANN.
6. Validate the trained ANN using data not used in training (out of sample).
7. Repeat until the best performing ANN architecture and learning algorithm are determined.

Fortunately, numerous ANN shell tools exist to facilitate the rapid development of numerous ANN solution models. Some examples include: MATLAB® (built in ANN shell), IBM® SPSS® (a separate add-on), NeuralWare® NeuralWorks products, Stuttgart Neural Network Simulator (SNNS), Emergent, and Neural Lab, among many others

(see the websites <http://www.neural-forecasting.com/software.htm> and https://grey.colorado.edu/emergent/index.php/Comparison_of_Neural_Network_Simulators, which both maintain lists of ANN shell tools). These shell tools enable researchers to select which type of ANN, supervised or unsupervised learning, and typically provide a choice of several different learning algorithms to match the learning type selected. The research then enters in the number of layers (if applicable) and quantity of neurodes per layer. Several also allow the developer to select soft processing criteria such as the aggregation and transfer functions.

ADVANTAGES AND DISADVANTAGES OF ANNs

One question researchers should ask is: why should I use an ANN versus some other proven methodology? A similar question would be should Copernicus ever have examined non-geocentric models of the solar system? Recent research has called for the investigation and utilization of ANNs in previously untried domains (DeTienne, DeTienne, & Joshi, 2003; Smith & Gupta, 2000). ANNs provide an alternative and often better framework for research seeking to solve complex problems.

Advantages

ANNs have generally been shown to be similar to parametric statistical methods for linear problems and to outperform parametric statistical methods on nonlinear problems (DeTienne, DeTienne, & Joshi, 2003; Walczak, 2012; Warner & Misra, 1996), indicating that more reliable results are achievable through ANNs than traditional statistical methods. Perhaps this is an unfair comparison as one of the significant advantages of ANNs over traditional statistical methods is that they are nonparametric. This means that the data used for training and subsequent use of ANNs does not need to satisfy any external statistical criteria such as being normally distributed or have a normal

error distribution. Any data may be used so long as it is meaningful to the problem solution.

An obvious advantage of ANNs is that they utilize machine learning to learn the solution surface for any domain problem. This indicates that they are capable of robust learning and can reliably approximate the solution surface for any appropriate problem (Hornik, Stinchcombe, & White, 1989). Once ANNs are trained the weights are typically locked when producing a real-world ANN application. This may then require periodic re-training of the ANN to capture changes in the population for classification problems or changing dynamics of the problem for prediction problems, such as periodicity of time series effects. ANNs are capable of learning any solution surface.

Another advantage to using ANNs in research is that once they are trained, the final model is resilient to noise in out of sample input data. Depending on the learning algorithm used, outliers and error prone data may be dealt with differently, but the ANN will still give a reliable approximation to the current data's location on the solution surface. Keep in mind that this advantage is for out of sample data only and that for most learning algorithms, relatively error free training samples are required to enable optimal learning.

Disadvantages

The biggest disadvantage to using ANNs for research and applications is their black box nature. Unlike the artificial intelligence technique of expert systems which can offer an explanation of the reasoning used to provide their answers, ANNs have no such explanation facility. The output is a result of the given input, but no other explanation of the reasoning or relationship of individual independent variable contribution is available. Various techniques have been developed to try and overcome the black box nature of ANNs. These include leave on out strategies that eliminate select variables from an ANN to determine the effect of that variable on the solution (Walczak, 2008), other methods to try and determine input variable

importance (Gevrey, Dimopoulos, & Lek, 2003; Paliwal & Kumar, 2011), and automated methods to generate regression-like equations from developed neural networks such that the coefficients associated with each variable may be evaluated (Hayashi, 2013; Tickle et al., 1998). Both of these techniques are labor intensive and as the size of the ANN grows, both through adding additional input variables and also when increasing the number of weighted connections through the addition of either additional hidden layers or additional neurodes within any of the layers, the regression equation techniques become more problematic.

One of the advantages of ANNs may also be viewed as a disadvantage. Since ANNs learn, they will find a way to map any input data into a solution surface. This means that anything can be learned even if it is meaningless. This introduces the problem of input variable sensitivity. There are two schools of thought on the types of input variables that may be used in ANNs. Because ANNs are learning, the first school believes that any number of relevant (and possibly irrelevant) variables may be given to the ANN as input and the ANN will learn which ones are meaningful. While this is true to an extent, the fact remains that all input influences the output of the ANN to a greater or lesser degree and even a small influence coming from an irrelevant variable may be sufficient to hinder the overall classification or prediction performance of the ANN. The second school believes that selecting the input variables is critical to performance and that only uncorrelated variables (Smith, 1993) which are somehow relevant to the solution (Tahai, Walczak, & Rigsby, 1998) should be used. Thus care must be taken to select relevant variables, perhaps through consultation with domain experts, and then test these variables to make sure none of them are highly correlated with each other.

A similar issue is the population demographics of the training data sets. Again, the learning capability of ANNs affects the answer. The choices are: balanced training sets with equal numbers of

each class present (Zhou & Liu, 2006), real-world training sets where the training set population percentages are similar to the real world (Wilson & Sharda, 1994), or other imbalanced or random sets. Balanced or real-world representative training sets are both usable as long as the real world population classes are not too disparate in membership size. This is because the ANN will learn to maximize its results and if one class of the population is only a very small percent of the overall population, then for a small number of classes the ANN may be able to optimize its classification results by ignoring the small class. If however this class is very important, such as a positive medical diagnosis for a disease, then it cannot be safely ignored. This problem decreases as the number of classes increase, as long as one class does not dominate the other class representations. Assuming all outcomes are equally important regardless of real-world population percentages, balanced training sets should be used to enable the ANN to adequately identify the correct solution surface.

A former issue with ANNs is the training time, which on older computers could take days or weeks to enable adequate learning to take place. As processor speeds have increased over time, this disadvantage no longer truly exists except for very large ANN architectures with thousands of neurodes and massive input data (big data).

A final issue with use of ANNs as real-world applications is an ethical one. Because ANNs operate as black boxes, what are the ethical and legal responsibilities of professionals who utilize the ANN's results? Furthermore, legal standards, especially in healthcare, mandate specific requirements for developers of ANN systems which include: repeatability, reliability, and performance (Lisboa & Taktak, 2006). Users of ANNs should be able to rely on the results of the ANNs as expert consultants, but should still use their own knowledge and discretion in interpreting the ANN's recommendations or results, as with any other computerized system.

Table 1. Examples of recent ANN research

Domain	ANN Types	Description	Reference
Agriculture	MLP	classify land cover images	(Heremans, Suykens, & Van Orshoven, 2016)
Business	MLP	predict manufacturing throughput	(Huang et al., 2016)
Finance	MLP recurrent	predict Chinese stock market predict stock, commodity, forex	(Sun et al., 2016) (Hussain et al., 2016)
Engineering	MLP	power outage prediction	(Tahir & Saqib, 2016)
Geology	MLP	predict landslides	(Bui et al., 2016)
Medicine	MLP MLP	predict morbidity severity predict length of stay	(Majd et al., 2016) (Tsai et al., 2016)

APPLICATIONS OF ANNs

ANNs are used across numerous disciplines to accomplish a wide variety of tasks. Table 1 shows a small sample of recent usage of ANNs appearing in the literature.

FUTURE RESEARCH DIRECTIONS

The biggest need for ANN researchers is to have ANNs recognized by the scientific community as a valid research methodology. Publications regularly appear to introduce the novel idea of ANNs as a research method across various domains (DeTienne, DeTienne, & Joshi, 2003; Smith & Gupta, 2000; Vitek, Iskander, & Oblow, 2000). Since numerous studies have already been mentioned that demonstrate the superior performance of ANNs over traditional parametric statistical techniques, this argument should not be hard to make, thus opening the door for additional ANN research across a much wider variety of research domains.

With the increasing importance of big data research, new research in ANNs is needed to examine how ANNs can learn quickly from big data training sets. Additional research will continue to examine new architectures.

Ensembles of ANNs, where one ANN produces the input for a subsequent ANN and the combination of ANNs creates a classification or prediction system is a growing area of research interest (West, Dellana, & Qian, 2005; Zhang &

Berardi, 2001). This ensemble combination of ANNs enables better classification of data when classification criteria across is independent across multiple categories within the data. Recurrent ANNs (RNNs) are ANNs where connections between units provide for directed cycle which enables the RNN to maintain a temporal state of the data. Although the idea of RNNs has been around since the late 1980's (Pineda, 1987), there has been a resurgence in interest recently in using RNNs for various tasks including speech and handwriting recognition.

Similar to combining ANNs with other ANNs into ensembles, there has been a growing research interest in combining ANNs with other statistical and artificial intelligence methods to create more robust learning environments (Walczak, 2012) and consequently better generalization from the trained ANN to a real-world application. Future research will continue to focus on integrated systems with incorporated ANNs (Pandey & Mishra, 2009), to enable the strengths of each integrated component to overcome shortcomings of the other components.

Deep learning ANNs have surfaced as a new research paradigm for the 2010's and into the future (Arel, Rose, & Karnowski, 2010). These deep learning ANNs enable the use of multiple types of data and data representations to gain a better understanding of classes (Schmidhuber, 2015). Successes in deep learning are already being demonstrated in image and speech recognition ANN systems (Ciresan et al., 2012; Hinton et al., 2012).

CONCLUSION

ANNs are a method for performing complex pattern recognition using machine learning techniques. The learning capabilities of ANNs make them robust and enable them to approximate arbitrary solution surfaces with a high degree of reliability (White, 1990). ANNs are applied across a wide variety of domains to solve numerous complex problems, but further advances into new fields are still needed.

The strengths of ANNs include: they are based on machine learning and thus able to learn arbitrary solution surfaces, they are nonparametric and may thus be applied to a larger variety of problems and data than traditional parametric statistical methods, and once trained they are resilient to noise in the data. While ANNs also have some perceived disadvantages, it is possible to overcome these disadvantages and new research in ANN ensembles and integrated systems that include an ANN component may further alleviate these perceived problems.

Because of their rich modeling capabilities based on machine learning and general superiority over traditional statistical methods, ANNs should become a de facto standard in any researcher's methodology toolbox.

REFERENCES

- Ali, J. M., Hussain, M. A., Tade, M. O., & Zhang, J. (2015). Artificial Intelligence techniques applied as estimator in chemical process systems—A literature survey. *Expert Systems with Applications*, 42(14), 5915–5931. doi:10.1016/j.eswa.2015.03.023
- Arel, I., Rose, D. C., & Karnowski, T. P. (2010). Deep machine learning—a new frontier in artificial intelligence research. *IEEE Computational Intelligence Magazine*, 5(4), 13–18. doi:10.1109/MCI.2010.938364
- Bansal, R. C. (2006). Overview and literature survey of artificial neural networks applications to power systems (1992–2004). *Journal of the Institution of Engineers (India). Part EL*, 86(1), 282–296.
- Barnard, E., & Wessels, L. F. A. (1992). Extrapolation and interpolation in neural network classifiers. *IEEE Control Systems*, 12(5), 50–53. doi:10.1109/37.158898
- Bui, D. T., Tuan, T. A., Klempe, H., Pradhan, B., & Revhaug, I. (2016). Spatial prediction models for shallow landslide hazards: A comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. *Landslides*, 13(2), 361–378. doi:10.1007/s10346-015-0557-6
- Carpenter, G. A., & Grossberg, S. (1987). A Massively Parallel Architecture for a Self-Organizing Neural Pattern Recognition Machine. *Computer Vision Graphics and Image Processing*, 37(1), 54–115. doi:10.1016/S0734-189X(87)80014-2
- Carpenter, G. A., Grossberg, S., Markuzon, N., Reynolds, J. H., & Rosen, A. B. (1992). Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multi-dimensional maps. *IEEE Transactions on Neural Networks*, 3(5), 698–713. doi:10.1109/72.159059 PMID:18276469
- Ciresan, D., Giusti, A., Gambardella, L. M., & Schmidhuber, J. (2012). Deep neural networks segment neuronal membranes in electron microscopy images. Academic Press.
- DeTienne, K. B., DeTienne, D. H., & Joshi, S. A. (2003). Neural Networks as Statistical Tools for Business Researchers. *Organizational Research Methods*, 6(2), 236–265. doi:10.1177/1094428103251907
- Gevrey, M., Dimopoulos, I., & Lek, S. (2003). Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecological Modelling*, 160(3), 249–264. doi:10.1016/S0304-3800(02)00257-0

- Hayashi, Y. (2013). Neural network rule extraction by a new ensemble concept and its theoretical and historical background: A review. *International Journal of Computational Intelligence and Applications*, 12(04), 1340006. doi:10.1142/S1469026813400063
- Hebb, D. O. (1949). *The Organization of Behavior: A Neuropsychological Approach*. New York: Wiley.
- Heremans, S., Suykens, J. A., & Van Orshoven, J. (2016). The effect of imposing fractional abundance constraints onto the multilayer perceptron for sub-pixel land cover classification. *International Journal of Applied Earth Observation and Geoinformation*, 44, 226–238. doi:10.1016/j.jag.2015.09.007
- Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A. R., Jaitly, N., & Kingsbury, B. et al. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine*, 29(6), 82–97. doi:10.1109/MSP.2012.2205597
- Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5), 359–366. doi:10.1016/0893-6080(89)90020-8
- Huang, P. B., Chen, J. C., Yu, C. W., & Chen, C. C. (2016). Integration of simulation and neural network in forecasting the throughput for TFT-LCD colour filter fabs. *International Journal of Computer Integrated Manufacturing*, 29(3), 298–308. doi:10.1080/0951192X.2015.1032356
- Hussain, A. J., Al-Jumeily, D., Al-Askar, H., & Radi, N. (2016). Regularized dynamic self-organized neural network inspired by the immune algorithm for financial time series prediction. *Neurocomputing*, 188, 23–30. doi:10.1016/j.neucom.2015.01.109
- Khalili, M., Majd, H. A., Khodakarim, S., Ahadi, B., & Hamidpour, M. (2016). Prediction of the Thromboembolic Syndrome: An Application of Artificial Neural Networks in Gene Expression Data Analysis. *Journal of Paramedical Sciences*, 7(2), 15–22.
- Kohonen, T. (1988). *Self-Organization and Associative Memory. Self-Organization and Associative Memory*. Berlin: Springer-Verlag. doi:10.1007/978-3-662-00784-6
- Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE*, 78(9), 1464–1480. doi:10.1109/5.58325
- Lisboa, P. J., & Taktak, A. F. (2006). The use of artificial neural networks in decision support in cancer: A systematic review. *Neural Networks*, 19(4), 408–415. doi:10.1016/j.neu-net.2005.10.007 PMID:16483741
- Lowe, D. S., & Broomhead, D. (1988). Multi-variable functional interpolation and adaptive networks. *Complex Systems*, 2(3), 321–355.
- Majd, H. A., Khalili, M., Khodakarim, S., Ahadi, B., & Hamidpour, M. (2016). Prediction of the Thromboembolic Syndrome: An Application of Artificial Neural Networks in Gene Expression Data Analysis. *Journal of Paramedical Sciences*, 7(2).
- McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 5(4), 115–133. doi:10.1007/BF02478259
- Minsky, M., & Papert, S. (1969). *Perceptrons*. Boston: MIT Press.
- Paliwal, M., & Kumar, U. A. (2011). Assessing the contribution of variables in feed forward neural network. *Applied Soft Computing*, 11(4), 3690–3696. doi:10.1016/j.asoc.2011.01.040

- Pandey, B., & Mishra, R. B. (2009). Knowledge and intelligent computing system in medicine. *Computers in Biology and Medicine*, 39(3), 215–230. doi:10.1016/j.compbiomed.2008.12.008 PMID:19201398
- Pineda, F. J. (1987). Generalization of back-propagation to recurrent neural networks. *Physical Review Letters*, 59(19), 2229–2232. doi:10.1103/PhysRevLett.59.2229 PMID:10035458
- Reggia, J. A. (1993). Neural computation in medicine. *Artificial Intelligence in Medicine*, 5(2), 143–157. doi:10.1016/0933-3657(93)90014-T PMID:8358491
- Rosenblatt, F. (1958). The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. *Psychological Review*, 65(6), 386–408. doi:10.1037/h0042519 PMID:13602029
- Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, 61, 85–117. doi:10.1016/j.neunet.2014.09.003 PMID:25462637
- Smith, K. A., & Gupta, J. N. D. (2000). Neural networks in business: Techniques and applications for the operations researcher. *Computers & Operations Research*, 27(11–12), 1023–1044. doi:10.1016/S0305-0548(99)00141-0
- Smith, M. (1993). *Neural Networks for Statistical Modeling*. New York: Van Nostrand Reinhold.
- Specht, D. F. (1990). Probabilistic neural networks. *Neural Networks*, 3(1), 109–118. doi:10.1016/0893-6080(90)90049-Q PMID:18282828
- Specht, D. F. (1991). A general regression neural network. *IEEE Transactions on Neural Networks*, 2(6), 568–576. doi:10.1109/72.97934 PMID:18282872
- Sun, X. Q., Shen, H. W., Cheng, X. Q., & Zhang, Y. (2016). Market Confidence Predicts Stock Price: Beyond Supply and Demand. *PLoS ONE*, 11(7), e0158742. doi:10.1371/journal.pone.0158742 PMID:27391816
- Tahai, A., Walczak, S., & Rigsby, J. T. (1998). Improving Artificial Neural Network Performance Through Input Variable Selection. In P. Siegel, K. Omer, A. deKorvin, & A. Zebda (Eds.), *Applications of Fuzzy Sets and The Theory of Evidence to Accounting II* (pp. 277–292). Stamford, CT: JAI Press.
- Tahir, M. F., & Saqib, M. A. (2016). Optimal scheduling of electrical power in energy-deficient scenarios using artificial neural network and Bootstrap aggregating. *International Journal of Electrical Power & Energy Systems*, 83, 49–57. doi:10.1016/j.ijepes.2016.03.046
- Thenmozhi, M., & Chand, G. S. (2016). Forecasting stock returns based on information transmission across global markets using support vector machines. *Neural Computing & Applications*, 27(4), 805–824. doi:10.1007/s00521-015-1897-9
- Tickle, A. B., Andrews, R., Golea, M., & Diederich, J. (1998). The truth will come to light: Directions and challenges in extracting the knowledge embedded within trained artificial neural networks. *IEEE Transactions on Neural Networks*, 9(6), 1057–1068. doi:10.1109/72.728352 PMID:18255792
- Tkáč, M., & Verner, R. (2016). Artificial neural networks in business: Two decades of research. *Applied Soft Computing*, 38, 788–804. doi:10.1016/j.asoc.2015.09.040
- Tsai, P. F. J., Chen, P. C., Chen, Y. Y., Song, H. Y., Lin, H. M., Lin, F. M., & Huang, Q. P. (2016). Length of Hospital Stay Prediction at the Admission Stage for Cardiology Patients Using Artificial Neural Network. *Journal of Healthcare Engineering*.

- Vitek, J. M., Iskander, Y. S., & Oblow, E. M. (2000). Improved Ferrite Number Prediction in Stainless Steel Arc Welds Using Artificial Neural Networks-Part 1: Neural Network Development. *Welding Journal*, 79(2), 33s–40s.
- Walczak, S. (2008). Evaluating medical decision making heuristics and other business heuristics with neural networks. In G. Phillips-Wren, N. Ichalkaranje, & L. C. Jain (Eds.), *Intelligent Decision Making: An AI-Based Approach* (pp. 259–287). Berlin: Springer. doi:10.1007/978-3-540-76829-6_10
- Walczak, S. (2012). Methodological Triangulation Using Neural Networks for Business Research. *Advances in Artificial Neural Systems*, 2012(517234), 1–12. doi:10.1155/2012/517234
- Warner, B., & Misra, M. (1996). Understanding neural networks as statistical tools. *The American Statistician*, 50(4), 284–293.
- Werbos, P. J. (1974). *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences* (Doctoral Dissertation). Harvard University, Boston, MA.
- West, D., Dellana, S., & Qian, J. (2005). Neural network ensemble strategies for financial decision applications. *Computers & Operations Research*, 32(10), 2543–2559. doi:10.1016/j.cor.2004.03.017
- White, H. (1990). Connectionist nonparametric regression: Multilayer feedforward networks can learn arbitrary mappings. *Neural Networks*, 3(5), 535–549. doi:10.1016/0893-6080(90)90004-5
- Widrow, B. (1964). Pattern Recognition and Adaptive Control. *IEEE Transactions on Applications and Industry*, 83(74), 269–277. doi:10.1109/TAI.1964.5407756
- Wilson, R. L., & Sharda, R. (1994). Bankruptcy Prediction Using Neural Networks. *Decision Support Systems*, 11(5), 545–557. doi:10.1016/0167-9236(94)90024-8
- Wong, B. K., Lai, V. S., & Lam, J. (2000). A bibliography of neural network business applications research: 1994–1998. *Computers & Operations Research*, 27(11–12), 1045–1076. doi:10.1016/S0305-0548(99)00142-2
- Yardimci, A. (2009). Soft computing in medicine. *Applied Soft Computing*, 9(3), 1029–1043. doi:10.1016/j.asoc.2009.02.003
- Zhang, G. P., & Berardi, V. L. (2001). Time series forecasting with neural network ensembles: An application for exchange rate prediction. *The Journal of the Operational Research Society*, 52(6), 652–664. doi:10.1057/palgrave.jors.2601133
- Zhou, Z. H., & Liu, X. Y. (2006). Training cost-sensitive neural networks with methods addressing the class imbalance problem. *IEEE Transactions on Knowledge and Data Engineering*, 18(1), 63–77. doi:10.1109/TKDE.2006.17

ADDITIONAL READING

- Jain, A. K., Mao, J., & Mohiuddin, K. M. (1996). Artificial neural networks: A tutorial. *IEEE Computer*, 29(3), 31–44. doi:10.1109/2.485891
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. doi:10.1038/nature14539 PMID:26017442
- Rodvold, D. M., McLeod, D. G., Brandt, J. M., Snow, P. B., & Murphy, G. P. (2001). Introduction to artificial neural networks for physicians: Taking the lid off the black box. *The Prostate*, 46(1), 39–44. doi:10.1002/1097-0045(200101)46:1<39::AID-PROS1006>3.0.CO;2-M PMID:11170130
- Walczak, S., & Cerpa, N. (1999). Heuristic principles for the design of artificial neural networks. *Information and Software Technology*, 41(2), 107–117. doi:10.1016/S0950-5849(98)00116-5

Zhang, G. P. (2007). Avoiding pitfalls in neural network research. *IEEE Transactions on Systems, Man and Cybernetics. Part C, Applications and Reviews*, 37(1), 3–16. doi:10.1109/TSMCC.2006.876059

KEY TERMS AND DEFINITIONS

Backpropagation: A learning method for ANNs that uses the relative error of a given output or set of outputs (epoch) that is propagated back through the ANN and used to adjust all weights on the connections between neurodes.

Epoch: A subset of training examples used as a group to determine average error for the ANN.

Neurode: A processing element within an ANN that simulates a single neuron in the brain.

Supervised Learning: A method used to train ANNs in which a training sample with known outcomes is used to enable the ANN to learn. The known outcome values are used to calculate an error term between the known or desired output and what is produced by the ANN to let the ANN know how to adjust its connection weights to minimize the relative error.

Unsupervised Learning: A learning method used to train ANNs where the ANN learns directly from the input values given to determine adjustments in connection weights.

A

Automatic Emotion Recognition Based on Non-Contact Gaits Information

Jingying Wang

University of Chinese Academy of Sciences, China

Baobin Li

University of Chinese Academy of Sciences, China

Changye Zhu

University of Chinese Academy of Sciences, China

Shun Li

University of Chinese Academy of Sciences, China

Tingshao Zhu

University of Chinese Academy of Sciences, China

INTRODUCTION

In recent years, emotion recognition has become a hot topic in human-computer interaction. If computers could understand human emotions, they could interact better with end users (Peter & Beale, 2008). In psychology, emotion is defined as a complex state that consists of a subjective experience (how we experience emotion), a physiological response (how our bodies react to emotion), and an expressive response (how we behave in response to emotion) (Smith & Lazarus, 1990). Emotion expression includes not only facial expression but also vocal and postural expression. The observable aspects of emotion (physiological and expressive components) might be able to be used as indicators of emotional state, such as facial expressions, speech, physiological parameters, gestures, and body movements (Peter & Beale, 2008). As the common use of modalities to recognizing emotional states in human-human interaction, various clues have been used in affective computing, such as facial expressions (e.g., Kenji, 1991), physiological signals (e.g., Picard,

Vyzas, & Healey, 2001), linguistic information (e.g., Alm, Roth, & Sproat, 2005) and acoustic features (e.g., Dellaert, Polzin, & Waibel, 1996). Other than that, it is feasible to recognize specific affective states using gait.

To investigate how gait features are effective in characterizing and recognizing emotions, gait features were used for modeling to identify different emotions. By utilizing 59 participants' gait data with emotion labels, machine learning models are trained to detect individual emotion (Li et al, 2016).

BACKGROUND: GAIT AND EMOTION

Walking is one of the basic and important components for the body posture and movement, and psychological research found that affective states can be identified by gaits (Montepare, Goldstein & Clausen, 1987). People in different emotional states could walk in different speed and show different gait patterns (Strike, Mohiyddini & Carlisle, 2009). Human can perceive others' emotion from

gait or posture in daily life. For example, people in fear may shrink his shoulder, and sad ones might lower his head and walk slowly (Roether, Omlor & Christensen, 2009). Even when gait was minimized by use of point-light displays, which meant to represent the body motion by only a small number illuminated dots, observers still could make judgments of emotion category and intensity (Atkinson, Dittrich, Gemmell, & Young, 2004).

Since Montepare et al. (1987) firstly demonstrated that gait relates to emotions, there have been a lot of researches focusing on the relationship between gaits and emotions. Krieger et al. (2013) found that gait, cognition and emotion are closely related. Applying sparse regression, Roether et al. (2009) extracted critical emotion-specific posture and movement features depended only on a small number of joints. Gross et al. (2011) identified the movement characteristics associated with positive and negative emotions experienced during walking. Destephe et al. (2013) assessed the differences between the expression of emotion regarding the expressed intensity. Characterizing human walking patterns by some kinematic cues, Hicheur et al. (2013) produced the avatar animation whose emotions could be recognized by human observers. In particular, by using data mining technique, much research has been done to recognize emotions automatically by analyzing subjects' gait. The accuracy of classifying distinct emotional states was 60~89% (Janssen, Schollhorn & Lubienetzki, 2008; Hicheur, Kadone, Grezes & Berthoz, 2013; Gunes, Shan & Chen, 2015; Clark, Pua & Pua, 2012).

Recently, it has also been reported the application of Kinect in the medical field. Lange et al. (2011) used Kinect as a game-based rehabilitation tool for balance training. Yeung et al. (2014) found that Kinect was valid in assessing body sway in clinical settings. Kinect also performed well in measuring some clinically relevant movements of people with Parkinson's disease (Galna, et al., 2014).

MAIN FOCUS: NON-CONTACT GAITS INFORMATION

A

Two Issues of Gaits Information

By exploring the researches and experiments on identifying humans' emotions from gait, there are two main foci: one is the way of data acquisition, the other is the features for describing gaits. These two foci are very important to identify human emotions automatically. Some works used high-speed cameras to get a video of gaits (Das & Meher, 2013; Rokanujjaman, Hossain & Hossain, 2014). The collected gait data lacked of some information such as joints of subjects' ankles, shoulders, etc. Some works employed 3D motion capture systems including wearable devices. While these wearable devices could help us acquire gait precisely, subjects with wearable devices may be uncomfortable, and emotions may be affected while walking (Roether, Omlor, & Christensen, 2009; Karg, Kuhnlenz, & Buss, 2010). The situation may be even worse for some people such as elderly adults, children, and mental patients. Thus, it is very important to acquire gait data in the way of non-contact. In addition, the cost is another considerable factor. The gait information recording technologies such as marker-based motion tracking system had already made it possible to automatically recognize the emotional state of a walker (Karg, Kuhnlenz, & Buss, 2010), however, because of the high cost of trained person, technical equipment and maintenance (Pogrzeba, Loreen, Wacker, Markus, Jung, Bernhard, 2013), the application of these non-portable systems were seriously limited.

Microsoft Kinect

The Microsoft Kinect is a low-cost, portable, camera-based sensor system, with the official software development kit (SDK). As a marker-free motion capture system, Kinect could continuously monitor three-dimensional body movement patterns, and

is a practical option to develop an inexpensive, widely available motion recognition system in human daily walks. The validity of Kinect has been tested in the studies of gesture and motion recognition. Kondori et al. (2011) identified head pose using Kinect, Fern'ndez-Baena, Susin and Lligadas (2012) found it perform well in tracking simple stepping movements, and Auvinet et al. (2015) successfully detected gait cycles in treadmill by Kinect. In Weber et al.'s (2012) report, the accuracy and sensitivity of kinematic measurements obtained from Kinect, such as reaching distance, joint angles, and spatialtemporal gait parameters, were estimated and found to be comparable to gold standard marker-based motion capture systems like Vicon. Since emotion could be expressed in gaits, and Kinect has been found a valid instrument to record gait, which motivates us to recognize emotion by gaits acquired by Kinect. We hypothesize that gaits information recorded by Kinect in the form of coordinates of the body joints, and the states could be recognized through machine learning methods based on gait.

Method

We have made several changes to the experiment design differing from previous work. In previous studies, participants were often required to recall a past situation or imagine a situation associated to certain affect while walking (Roether, Omlor, & Christensen, 2009; Gross, Crane, & Fredrickson, 2012), in which the induced emotional was too strong. The first change is that we chose a moderate way—watch videos—to induced people's emotions.

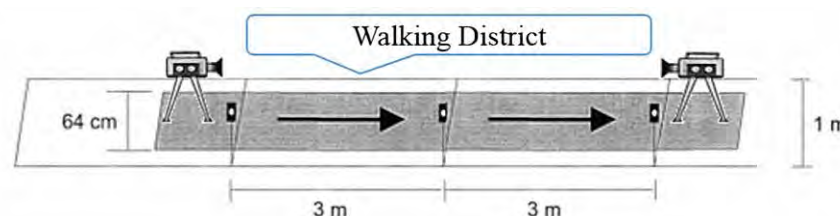
Secondly, the participants were instructed to walk naturally. Compared to requiring participants to walk with certain emotion (Westermann, STAHL, & Hesse, 1996), the walking condition of participants in our study was more ecological.

The experiment has been completed in a footpath with one meter wide and six meters long. At both ends of the footpath, we set up a Kinect sensor respectively to record walkers' gait information. Figure 1 sketches the experiment's scene, in which the dark area is the footpath. Because the minimum and maximum photographic distance of each Kinect's sensor is 1m and 3m respectively (Han et. al., 2013), subjects are instructed to walk inside the district not near to the sensors.

59 healthy graduate students (32 females) are recruited to take part in this experiment. Their average age is 23, and they have no injuries, illnesses or other condition influence their gait patterns. Every participant will first walk back and forth for 2 minutes in the walking district, evaluate his or her emotional state: angry or happy, and give a score: 1 to 10. Then, they will watch a short video for arousing one certain emotion like joy or angry, and continue walking around the footpath for 1 minute. At last, they assess themselves how happy or angry they are by presenting two scores (1 to 10) to evaluate their own emotional state finishing watching the video and finishing the second walking.

The above experiment will be conducted twice. In the first experiment, subjects' happy emotion will be aroused by a shot video full of jokes after walking with certain natural emotion for two minutes, while participants' negative emotion

Figure 1. The experiment's scene and walking district: the footpath is one meter wide and six meters long, and two Kinects are set up at both ends



like anger will be aroused in the second experiment. In addition, there is an interval of at least one day between these two experiments to avoid the emotional interference.

Data Collection and Processing

The whole procedure for identifying emotions from gait information just like a general machine learning process, it consists of four parts: Collecting Data, Data preprocessing, Feature extraction, Modeling and Classification.

The data collected from natural walking. With the Kinect cameras placed at the two ends of the footpath, participants' gait information was recorded as video on 30Hz frame rate. Each frame contains 3-dimensional information of 25 joints of body, including head, shoulders, elbows, wrists, hands, spine (shoulder, mid and base), hips, knees, ankles and feet, as shown in Figure 2.

During our experiment, all the frames recorded by one Kinect sensor were integrated consequently

in a '.txt' file. Figure 3 shows coordinate tracks of some different joints: the base of the spine (black solid), right ankle (red dashes), right wrist (rose-red line) and right shoulder (blue dash dots) for three different emotion states. And, from these figures, we can find that for different emotional states, tracks of the same joint differ obviously.

The raw data may be noisy, complicated and redundant. To make it clean for further process, it is necessary to run data preprocessing. The operations we take before extracting features, includes 4 steps: data segmentation, low-pass filtering, coordinate translation, and coordinate difference (Li, et. al., 2016).

After data preprocessing, we then run feature selection to acquire important gait features closely related to subject's emotions. Figure 4 shows the procedure of extracting features, in which operations in dotted box are optional because it needs to be verified whether they improve the performance of classification models. Since there are two types of data fractions: front and back parts, and here

Figure 2. Stick figure and location of body joint centers recorded by Kinect

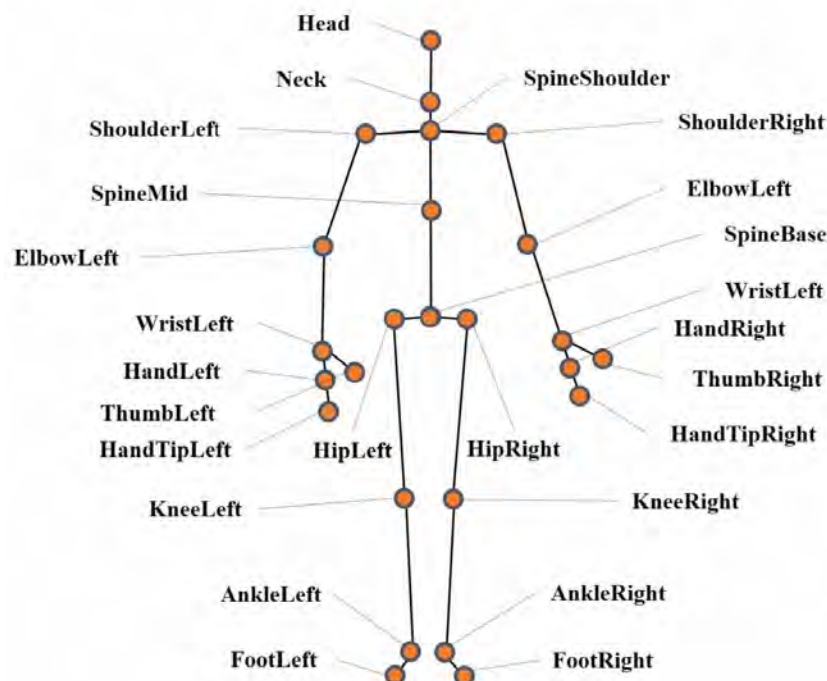
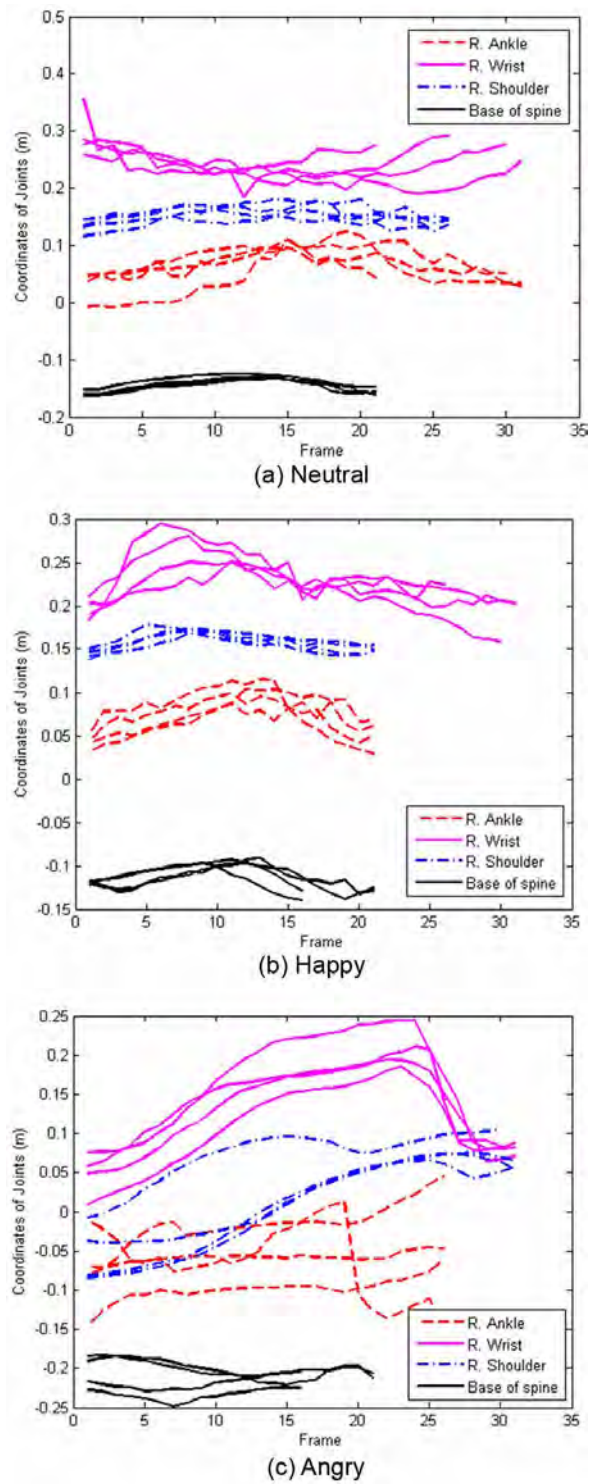


Figure 3. Coordinate tracks of joints in 3 emotional states: the diversity between joints is obvious, and the tracks of the same joint differ slightly in different emotional states



we only describe the procedure to extract features from front, and the same procedure works on the back part.

From data set, we found that each walker completed more than four rounds during experiment, so we get 2880 features of a sample. These features form the set of ones in the frequency domain, and shorted as *FF* (*Features in the frequency domain*). We denote features in the time domain as *FT* (*Features in the time domain*), which are some related kinematic ones too. At last, we calculated their mean and variance. Combined with frame number of this segment, we got 11 features in one walking segment and 44 features in a sample.

Results

Self-Reports of Emotional States

In the current study, for both anger and happiness priming, the emotional state ratings of AP (After priming) I and APII were higher than BP (Before priming). Paired-Samples *t*-test (by SPSS 15.0) showed that: for anger priming, anger ratings before priming was significantly lower than API ($t [58] = 18.98, p < .001$) and APII ($t [58] = 14.52, p < .001$); for happiness priming, happiness ratings before priming was also significantly lower than API ($t [58] = 10.31, p < .001$) and APII ($t [58] =$

7.99, $p < .001$). These results indicated that both anger and happiness priming were successfully eliciting changes of emotional state on the corresponding dimension. In the first round, participants were generally experiencing more anger while walking after video than before video, and the same happened for happiness in the second round.

Classifying Results and Comparison

We used *FF + FT*, *FT* and the “PCA-DFT-PCA” method proposed by M. Karg et al. (Karg, Kuhnlenz, & Buss, 2010) to classify three emotions: Neutral, Happy and Angry. Because there is no evidence shown that a certain classifier method in the machine learning fits for identifying emotions than others (Gunes, & Schuller, 2013), we have tries several classifiers including linear discriminant analysis (LDA) (Martinez, & Kak, 2001), Naive Bayes (Jordan, A., 2002), Decision tree (Quinlan, 1993) and SVM (William, & Teukolsky, 2007) to build our model for classifying and predicts subjects’ emotions from gait. The results of accuracy are shown in the Table 1.

We also tried classifying emotions on different Kinects. Three computational models were built to differentiate anger from neutral state (before anger priming), happiness from neutral state (before happiness priming), and anger from

Figure 4. The flow chart of extracting features in the time and frequency domain

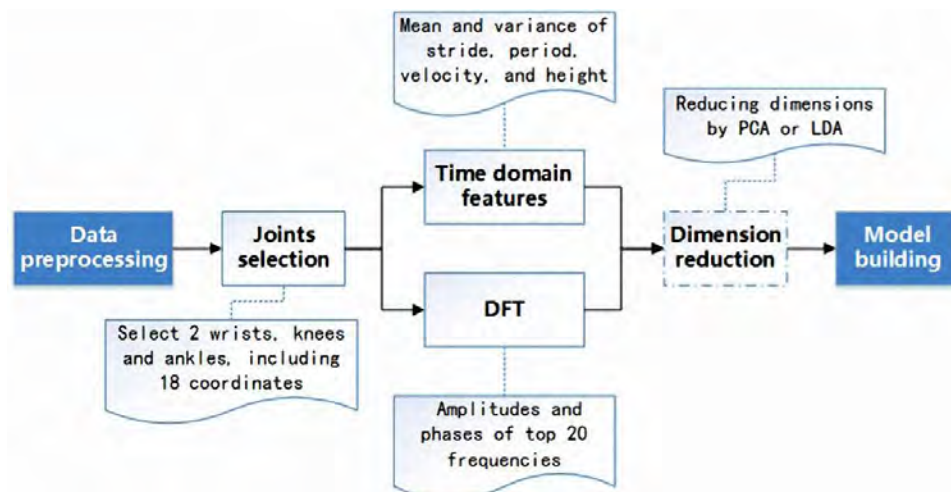


Table 1. The accuracy of different emotion (In %)

Features	Emotion	LDA	NaiveBayes	DecisionTree	SVM
FF+FT	Neutral	60	55	59	68
	Happy	62	59	61	71
FT	Neutral	58	46	53	61
	Happy	61	76	88	76
Karg's features	Neutral	66	49	17	49
	Happy	37	48	91	49
FF+FT	Neutral	53	71	49	34
	Angry	44	17	53	41
FT	Neutral	46	59	54	58
	Angry	46	54	32	48
Karg's features	Neutral	44	52	63	51
	Angry	58	48	45	48
FF+FT	Happy	59	61	55	65
	Angry	63	65	64	70
FT	Happy	61	61	55	60
	Angry	58	59	56	63
Karg's features	Happy	64	53	35	56
	Angry	56	48	67	53

happiness. We trained and evaluated several classifiers, including *Naive Bayes*, *Random Forests*, *LibSVM* and *SMO*, with 10-fold cross-validation. Table 2 shows the accuracy of each classifier in recognizing different emotions. There exists difference between the results of Kinect 1 and Kinect 2, and the accuracy using data recorded by Kinect 1 was better than Kinect 2.

FUTURE RESEARCH DIRECTIONS

The quality of gait recorded by two Kinect cameras were different, which probably meant that we did not perfectly control the illumination intensity of our experiment environment. Therefore, we should pay more attention to the control of conditions in the future. This study presented a feasible method of automatically recognizing emotional states

Table 2. The accuracy of recognizing different emotions (In %) Note: Table entries are accuracies expressed as a percentage. Values below chance level (50%) are not presented.

		Naïve Bayes	Random Forests	LibSVM	SMO
Angry & Neutral	Kinect 1	80.5085	52.5424	72.0339	52.5424
	Kinect 2	75.4237	—	71.1864	—
Happy & Neutral	Kinect 1	79.6610	51.6949	77.9661	—
	Kinect 2	61.8644	51.6949	52.5414	—
Angry & Happy	Kinect 1	52.5424	55.0847	—	51.6949
	Kinect 2	—	51.6949	—	50.8475

from gaits using the portable, low-cost Kinect cameras. It would be worthwhile for future studies to improve this method, to identify certain types of affect more accurately. In our study, the extraction of low-level features from gaits was by heuristic. In fact, people found that different facets of gait patterns may relate to certain dimensions of emotion (such as arousal level or valence) in varying degrees (Pollick, Paterson, Bruderlin, & Sanford, 2001). In future study, we plan to add more features according to the characteristics of the target emotion, for a better utilization of the features in model training. In addition, the future work could use other methods such as time-frequency analysis to identify different emotions.

CONCLUSION

In this paper, we propose a novel method to recognize human emotion from gait data collected by Microsoft Kinects. Different from existing methods, we extracted gait features in both time and frequency domain to classify different affective states: Neutral, Happy and Angry states. Experimental results partly supported our hypothesis: the walkers' emotion states (such as happiness and anger) could be reflected in their gaits recorded by Kinect in the form of coordinates of the main joints of body, and the states could be recognized through machine learning methods based on gait.

Although the current method could only recognize induced emotions, it still demonstrated the potential benefit in application. The automatically recognized emotional arousal could be very useful for decision making, security check and so on.

REFERENCES

- Alm, C. O., Roth, D., & Sproat, R. (2005, October). Emotions from text: machine learning for text-based emotion prediction. In *Proceedings of the conference on human language technology and empirical methods in natural language processing* (pp. 579-586). Association for Computational Linguistics. doi:10.3115/1220575.1220648
- Atkinson, A. P., Dittrich, W. H., Gemmell, A. J., & Young, A. W. (2004). Emotion perception from dynamic and static body expressions in point-light and full-light displays. *Perception*, 33(6), 717-746. doi:10.1068/p5096 PMID:15330366
- Auvinet, E., Multon, F., Aubin, C. E., Meunier, J., & Raison, M. (2015). Detection of gait cycles in treadmill walking using a Kinect. *Gait & Posture*, 41(2), 722-725. doi:10.1016/j.gaitpost.2014.08.006 PMID:25442670
- Clark, R. A., Pua, Y. H., Fortin, K., Ritchie, C., Webster, K. E., Denehy, L., & Bryant, A. L. (2012). Validity of the Microsoft Kinect for assessment of postural control. *Gait & Posture*, 36(3), 372-377. doi:10.1016/j.gaitpost.2012.03.033 PMID:22633015
- Das, S., & Meher, S. (2013, April). Automatic extraction of height and stride parameters for human recognition. In *Engineering and Systems (SCES), 2013 Students Conference on* (pp. 1-6). IEEE.
- Dellaert, F., Polzin, T., & Waibel, A. (1996, October). Recognizing emotion in speech. In *Spoken Language. ICSLP 96. Proceedings., Fourth International Conference on* (Vol. 3, pp. 1970-1973). IEEE. doi:10.1109/ICSLP.1996.608022

- Destephe, M., Maruyama, T., Zecca, M., Hashimoto, K., & Takanishi, A. (2013, July). The influences of emotional intensity for happiness and sadness on walking. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 7452-7455). IEEE. doi:10.1109/EMBC.2013.6611281
- Fernández-Baena, A., Susín, A., & Lligadas, X. (2012, September). Biomechanical validation of upper-body and lower-body joint movements of kinect motion capture data for rehabilitation treatments. In *Intelligent Networking and Collaborative Systems (INCoS), 2012 4th International Conference on* (pp. 656-661). IEEE. doi:10.1109/INCoS.2012.66
- Galna, B., Barry, G., Jackson, D., Mhiripiri, D., Olivier, P., & Rochester, L. (2014). Accuracy of the Microsoft Kinect sensor for measuring movement in people with Parkinsons disease. *Gait & Posture*, 39(4), 1062–1068. doi:10.1016/j.gaitpost.2014.01.008 PMID:24560691
- Gross, M. M., Crane, E. A., & Fredrickson, B. L. (2012). Effort-Shape and kinematic assessment of bodily expression of emotion during gait. *Human Movement Science*, 31(1), 202–221. doi:10.1016/j.humov.2011.05.001 PMID:21835480
- Gunes, H., & Schuller, B. (2013). Categorical and dimensional affect analysis in continuous input: Current trends and future directions. *Image and Vision Computing*, 31(2), 120–136. doi:10.1016/j.imavis.2012.06.016
- Gunes, H., Shan, C., Chen, S., & Tian, Y. (2015). Bodily expression for automatic affect recognition. *Emotion Recognition: A Pattern Analysis Approach*, 343-377.
- Han, J., Shao, L., Xu, D., & Shotton, J. (2013). Enhanced computer vision with microsoft kinect sensor: A review. *IEEE Transactions on Cybernetics*, 43(5), 1318-1334.
- Hicheur, H., Kadone, H., Grezes, J., & Berthoz, A. (2013, September). Perception of emotional gaits using avatar animation of real and artificially synthesized gaits. In *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on* (pp. 460-466). IEEE.
- Janssen, D., Schöllhorn, W. I., Lubienetzki, J., Fölling, K., Kokenge, H., & Davids, K. (2008). Recognition of emotions in gait patterns by means of artificial neural nets. *Journal of Nonverbal Behavior*, 32(2), 79–92. doi:10.1007/s10919-007-0045-3
- Jordan, A. (2002). On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes. *Advances in Neural Information Processing Systems*, 14, 841.
- Karg, M., Kuhnlenz, K., & Buss, M. (2010). Recognition of affect based on gait patterns. *IEEE Transactions on Systems, Man, and Cybernetics. Part B, Cybernetics*, 40(4), 1050–1061. doi:10.1109/TSMCB.2010.2044040 PMID:20350859
- Kenji, M. A. S. E. (1991). Recognition of facial expression from optical flow. *IEICE Transactions on Information and Systems*, 74(10), 3474–3483.
- Kleinsmith, A., & Bianchi-Berthouze, N. (2013). Affective body expression perception and recognition: A survey. *IEEE Transactions on Affective Computing*, 4(1), 15–33. doi:10.1109/T-AFFC.2012.16
- Kondori, F. A., Yousefi, S., Li, H., Sonning, S., & Sonning, S. (2011, November). 3D head pose estimation using the Kinect. In *Wireless Communications and Signal Processing (WCSP), 2011 International Conference on* (pp. 1-4). IEEE.
- Krieger, V., Lallart, E., & Jouvent, R. (2013, September). Bodily Manifestations of Affects: The Example of Gait and Virtual Reality. In *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on* (pp. 179-184). IEEE.

- Lange, B., Chang, C. Y., Suma, E., Newman, B., Rizzo, A. S., & Bolas, M. (2011, August). Development and evaluation of low cost game-based balance rehabilitation tool using the Microsoft Kinect sensor. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 1831-1834). IEEE. doi:10.1109/IEMBS.2011.6090521
- Li, S., Cui, L., Zhu, C., Li, B., Zhao, N., & Zhu, T. (2016). Emotion recognition using Kinect motion capture data of human gaits. *PeerJ*, 4, e2364. doi:10.7717/peerj.2364 PMID:27672492
- Martínez, A. M., & Kak, A. C. (2001). Pca versus lda. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(2), 228–233. doi:10.1109/34.908974
- Montepare, J. M., Goldstein, S. B., & Clausen, A. (1987). The identification of emotions from gait information. *Journal of Nonverbal Behavior*, 11(1), 33–42. doi:10.1007/BF00999605
- Peter, C., & Beale, R. (Eds.). (2008). *Affect and emotion in human-computer interaction: From theory to applications* (Vol. 4868). Springer Science & Business Media. doi:10.1007/978-3-540-85099-1
- Picard, R. W., Vyzas, E., & Healey, J. (2001). Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(10), 1175–1191. doi:10.1109/34.954607
- Pogrzeba, L., Wacker, M., & Jung, B. (2013). Field study of a low-cost markerless motion analysis for rehabilitation and sports medicine. *Gait & Posture*, 38, S94–S95. doi:10.1016/j.gaitpost.2013.07.195
- Pollick, F. E., Paterson, H. M., Bruderlin, A., & Sanford, A. J. (2001). Perceiving affect from arm movement. *Cognition*, 82(2), B51–B61. doi:10.1016/S0010-0277(01)00147-0 PMID:11716834
- Quinlan, J. R. (1993). *C4. 5: Programming for machine learning*. Morgan Kauffmann.
- Roether, C. L., Omlor, L., Christensen, A., & Giese, M. A. (2009). Critical features for the perception of emotion from gait. *Journal of Vision (Charlottesville, Va.)*, 9(6), 15–15. doi:10.1167/9.6.15 PMID:19761306
- Rokanujjaman, M., Hossain, M. A., & Hossin, M. A. (2014, May). Part-based gait identification using fusion technique. In *Informatics, Electronics & Vision (ICIEV), 2014 International Conference on* (pp. 1-5). IEEE.
- Smith, C. A., & Lazarus, R. S. (1990). Emotion and adaptation. In *Handbook of personality: Theory and research* (pp. 609–637). New York, NY: Guilford Press.
- Strike, S., Mohiyddini, C., & Carlisle, A. (2009). Effect of emotional state on walking gait. *Gait & Posture*, 30, S123. doi:10.1016/j.gaitpost.2009.08.185
- Weber, I., Koch, J., Meskemper, J., Friedl, K., Heinrich, K., & Hartmann, U. (2012). Is the MS Kinect suitable for motion analysis? *Biomedical Engineering/Biomedizinische Technik*, 57(SI-1 Track-F).
- Westermann, R., Spies, K., Stahl, G., & Hesse, F. W. (1996). Relative effectiveness and validity of mood induction procedures: Analysis. *European Journal of Social Psychology*, 26(4), 557–580. doi:10.1002/(SICI)1099-0992(199607)26:4<557::AID-EJSP769>3.0.CO;2-4
- William, H., & Teukolsky, S. A. (2007). *Numerical recipes: The art of scientific computing* (3rd ed.). New York: Cambridge University Press.
- Yeung, L. F., Cheng, K. C., Fong, C. H., Lee, W. C., & Tong, K. Y. (2014). Evaluation of the Microsoft Kinect as a clinical assessment tool of body sway. *Gait & Posture*, 40(4), 532–538. doi:10.1016/j.gaitpost.2014.06.012 PMID:25047828

ADDITIONAL READING

- Atkinson, A. P., Tunstall, M. L., & Dittrich, W. H. (2007). Evidence for distinct contributions of form and motion information to the recognition of emotions from body gestures. *Cognition*, 104(1), 59–72. doi:10.1016/j.cognition.2006.05.005 PMID:16831411
- Barliya, A., Omlor, L., Giese, M. A., Berthoz, A., & Flash, T. (2013). Expression of emotion in the kinematics of locomotion. *Experimental Brain Research*, 225(2), 159–176. doi:10.1007/s00221-012-3357-4 PMID:23250443
- Beck, A., Stevens, B., Bard, K. A., & Cañamero, L. (2012). Emotional body language displayed by artificial agents. [TiiS]. *ACM Transactions on Interactive Intelligent Systems*, 2(1), 2. doi:10.1145/2133366.2133368
- Blake, R., & Shiffrar, M. (2007). Perception of human motion. *Annual Review of Psychology*, 58(1), 47–73. doi:10.1146/annurev.psych.57.102904.190152 PMID:16903802
- Bonnechere, B., Jansen, B., Salvia, P., Bouzahuene, H., Omelina, L., Moiseev, F., & Jan, S. V. S. et al. (2014). Validity and reliability of the Kinect within functional assessment activities: Comparison with standard stereophotogrammetry. *Gait & Posture*, 39(1), 593–598. doi:10.1016/j.gaitpost.2013.09.018 PMID:24269523
- Calvo, R. A., & D’Mello, S. (2010). Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on affective computing*, 1(1), 18–37.
- Cambria, E., & Hussain, A. (2012). *Sentic computing: Techniques, tools, and applications* (Vol. 2). Springer Science & Business Media. doi:10.1007/978-94-007-5070-8
- Chouchourelou, A., Matsuka, T., Harber, K., & Shiffrar, M. (2006). The visual analysis of emotional actions. *Social Neuroscience*, 1(1), 63–74. doi:10.1080/17470910600630599 PMID:18633776
- Clark, R. A., Bower, K. J., Mentiplay, B. F., Paterson, K., & Pua, Y. H. (2013). Concurrent validity of the Microsoft Kinect for assessment of spatiotemporal gait variables. *Journal of Biomechanics*, 46(15), 2722–2725. doi:10.1016/j.jbiomech.2013.08.011 PMID:24016679
- Clark, R. A., Pua, Y. H., Bryant, A. L., & Hunt, M. A. (2013). Validity of the Microsoft Kinect for providing lateral trunk lean feedback during gait retraining. *Gait & Posture*, 38(4), 1064–1066. doi:10.1016/j.gaitpost.2013.03.029 PMID:23643880
- Cui, L., Li, S., & Zhu, T. (2016, January). Emotion detection from natural walking. In *International Conference on Human Centered Computing* (pp. 23–33). Springer International Publishing. doi:10.1007/978-3-319-31854-7_3
- Dael, N., Mortillaro, M., & Scherer, K. R. (2012). Emotion expression in body action and posture. *Emotion (Washington, D.C.)*, 12(5), 1085–1101. doi:10.1037/a0025737 PMID:22059517
- Galna, B., Barry, G., Jackson, D., Mhiripiri, D., Olivier, P., & Rochester, L. (2014). Accuracy of the Microsoft Kinect sensor for measuring movement in people with Parkinsons disease. *Gait & Posture*, 39(4), 1062–1068. doi:10.1016/j.gaitpost.2014.01.008 PMID:24560691
- Grezes, J., Pichon, S., & De Gelder, B. (2007). Perceiving fear in dynamic body expressions. *NeuroImage*, 35(2), 959–967. doi:10.1016/j.neuroimage.2006.11.030 PMID:17270466
- Gross, M. M., Crane, E. A., & Fredrickson, B. L. (2012). Effort-Shape and kinematic assessment of bodily expression of emotion during gait. *Human Movement Science*, 31(1), 202–221. doi:10.1016/j.humov.2011.05.001 PMID:21835480
- Karg, M., Samadani, A. A., Gorbet, R., Kühnlenz, K., Hoey, J., & Kulić, D. (2013). Body movements for affective expression: A survey of automatic recognition and generation. *IEEE Transactions on Affective Computing*, 4(4), 341–359. doi:10.1109/T-AFFC.2013.29

Kleinsmith, A., Bianchi-Berthouze, N., & Steed, A. (2011). Automatic recognition of non-acted affective postures. *IEEE Transactions on Systems, Man, and Cybernetics. Part B, Cybernetics*, 41(4), 1027–1038. doi:10.1109/TSMCB.2010.2103557 PMID:21278020

Ma, Y., Paterson, H. M., & Pollick, F. E. (2006). A motion capture library for the study of identity, gender, and emotion perception from biological motion. *Behavior Research Methods*, 38(1), 134–141. doi:10.3758/BF03192758 PMID:16817522

Mentiplay, B. F., Clark, R. A., Mullins, A., Bryant, A. L., Bartold, S., & Paterson, K. (2013). Reliability and validity of the Microsoft Kinect for evaluating static foot posture. *Journal of foot and ankle research*, 6(1), 1.

Ning, Y., Zhu, T., & Wang, Y. Affective-Word based Chinese Text Sentiment Classification. In *proceedings of 5th International Conference on Pervasive Computing and Applications (ICPCA2010)*. Dec. 2010, Maribor, Slovenia. doi:10.1109/ICPCA.2010.5704084

Pfister, A., West, A. M., Bronner, S., & Noah, J. A. (2014). Comparative abilities of Microsoft Kinect and Vicon 3D motion capture for gait analysis. *Journal of Medical Engineering & Technology*, 38(5), 274–280. doi:10.3109/03091902.2014.909540 PMID:24878252

Pichon, S., de Gelder, B., & Grezes, J. (2008). Emotional modulation of visual and motor areas by dynamic body expressions of anger. *Social Neuroscience*, 3(3-4), 199–212. doi:10.1080/17470910701394368 PMID:18979376

Plantard, P., Auvinet, E., Pierres, A. S. L., & Multon, F. (2015). Pose estimation with a kinect for ergonomic studies: Evaluation of the accuracy using a virtual mannequin. *Sensors (Basel, Switzerland)*, 15(1), 1785–1803. doi:10.3390/s150101785 PMID:25599426

Sauter, D. A., Eisner, F., Calder, A. J., & Scott, S. K. (2010). Perceptual cues in nonverbal vocal expressions of emotion. *Quarterly Journal of Experimental Psychology*, 63(11), 2251–2272. doi:10.1080/17470211003721642 PMID:20437296

Savva, N., Scarinzi, A., & Bianchi-Berthouze, N. (2012). Continuous recognition of player's affective body expression as dynamic quality of aesthetic experience. *IEEE Transactions on Computational Intelligence and AI in games*, 4(3), 199–212.

Van den Stock, J., Righart, R., & De Gelder, B. (2007). Body expressions influence recognition of emotions in the face and voice. *Emotion (Washington, D.C.)*, 7(3), 487–494. doi:10.1037/1528-3542.7.3.487 PMID:17683205

KEY TERMS AND DEFINITIONS

Affective Computing: It is defined as recognize, interpret, process, and simulate emotion via technologies or devices.

Automatic Emotion Recognition: It is a manner of recognize emotion by using computer technologies such as signal processing, machine learning, and computer vision.

Emotion: It is a manner of nonverbal expression of people's views and attitudes.

Emotion Recognition: Means comprehend or unscramble emotion, usually refer to a social ability of humans.

Emotional Arousal: Which is an emotional reaction induced by certain emotional stimulus.

Board Games AI

Tad Gonsalves

Sophia University, Japan

INTRODUCTION

To date, Artificial Intelligence (AI) has produced a plethora of techniques and algorithms in the field of computer game playing. Escaping from the maze being chased by the enemy, firing shots and dodging missiles, etc. coupled with background music, eerie sounds, and animated colors are some of the AI games that fill modern computer screens. Although the flashy and dazzling computer games are increasingly becoming popular, this chapter deals with the description and analysis of the *classic* AI games, where two players (an intelligent computer program and a human player) sit on opposite sides with the game board in the middle. Examples of such classic AI games are tic-tac-toe, checkers, othello, chess, shogi, go, etc.

The classic games, played by two players on a single board taking turns alternatively are known as zero-sum games with perfect information (Millington, & Funge, 2009). The game can be unfolded in the form of a tree structure and the winning path consisting of legal moves through the tree nodes and branches can be traced using an efficient algorithm called the *MiniMax* algorithm. At the heart of this algorithm is the evaluation function, which gives an estimate of the goodness of a board position. Even complex board games like chess rely on the computation of the evaluation function to determine the best possible move. However, due to the combinatorial explosion of the state space an exhaustive search of the game tree is infeasible. This situation is alleviated by the alpha-beta pruning technique and other heuristic search methods.

Machine Learning (ML) techniques employing Artificial Neural Networks is emerging as a powerful technique in game playing. In some

applications, ML programs learn the evaluation function itself and then compute the winning moves using the *MiniMax* and the alpha-beta pruning algorithms. Some rely purely on the ML methods for mastering the winning strategy while others also use expert knowledge to supplement the ML process.

This chapter introduces the classic two-player zero-sum (the end result of the winner-loser score is zero), deterministic board games and explores some of the classic AI algorithms for game playing. In particular, it describes the game-tree, the *MiniMax* and the alpha-beta pruning algorithms for searching for the winning moves in real-time. It also introduces the relatively new ML techniques for game playing and the Evolutionary and Swarm Intelligence algorithms for conducting efficient learning. Finally, it indicates the direction of future research in AI game-playing algorithms and techniques.

BACKGROUND

Playing games intelligently and trying to win against human champions has been a grand challenge for *Artificial Intelligence* from its inception. Chess, in particular, has been referred to as the *Drosophila* of AI. The first study in computer chess was published by Claude Shannon in 1950 (Shanon, 1950). However, the first working AI programs were for playing checkers (Samuel, 1960; Samuel, 1967). In the 1970s and 1980s, computer-games research concentrated on chess and the brute-force search approach (Schaeffer, 2002). The triumph of AI in computer games came in the 1990s when to everyone's surprise AI programs began to defeat the reigning world

DOI: 10.4018/978-1-5225-2255-3.ch013

champions. In 1994, the program CHINOOK won the World Man–Machine Championship (Schaeffer, 1997). Three year later, IBM’S DEEP BLUE defeated the World Chess Champion Garry Kasparov (Hsu, 2002) and LOGISTELLO won against the Othello Champion Takeshi Murakami (Buro, 1997). Finally, in 2011, IBM’s WATSON defeated the world champions in the quiz game of Jeopardy.

The computer games mentioned above, with the exception of Jeopardy, are all two-player board-games. Any classic board-game playing algorithm makes a static list of all possible moves of the two players called the game tree, at a given stage of the game. It then assigns a relative score to each of the position. The algorithmic game playing strategy consists of efficiently searching through the moves that maximize the score. However, in practice, the number of moves even for any modest game are astronomically large and an exhaustive search is impossible. Specialized super-computers like the DEEP BLUE and WATSON greatly improves the depth and breadth of the search, but in no way is it exhaustive. AI techniques deal with the design and implementation of algorithms that provide short-cuts in the search. These AI algorithms also combine the power of Machine Learning to further reduce the search path. The AI techniques and their future development is described in detail in the following sections.

BOARD GAMES

This section describes the characteristics of two-player board games. The two players taking turns in playing give rise to a “turn game”. The unfolding of the game on the board surrounded by the two players can be represented by an abstract graph called the game tree. These two concepts are defined in the following sub-sections.

Turn Games

In the simplest board games, there are usually two players. These are called turn games, because each player takes a turn in playing. The game proceeds by the two players taking turns alternately. The outcome of a two-player turn game is either a win, a loss or a draw for each of the players concerned. The two-player board games are, therefore, zero-sum games – the net outcome score is zero. Let’s say, the score for winning is +10, for losing is -10 and for drawing the game is 0. The zero-sum turn play necessarily implies that a win for a player results in a loss for the opponent and vice versa. The sum of the scores at the end of the game is zero (+10-10). In the case of a draw situation also, the sum is zero.

Perfect information is another underlying assumption in a two-player turn game. It is assumed that the two players have complete knowledge of the progression of the game till the end result. In other words, any of the two players at any given point in the game knows all the possible move options at that point and all the possible outcomes emanating from each of those moves. This is opposed to random games which may involve the rolling of a die in which none of the players can predict the next move because of its intrinsic non-deterministic nature.

Game Tree

The complete information about the progression of the turn game that each player has access to can be represented in a graph consisting of nodes and edges. This graph is called the game tree. The nodes in the graph represent the board positions (configurations), and the edges represent the possible moves a player can choose to make. The starting position of the game is represented by the *root* node. A single move transports the game from the *parent* node to one of the next level *children* node. The end of the game is represented by the *leaf* nodes.

Figure 1. Tic-Tac-Toe game tree

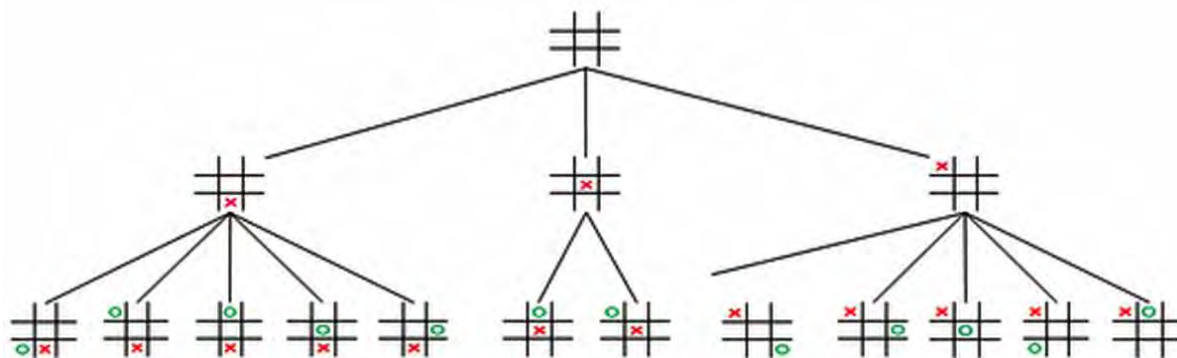


Figure 1 shows the game tree for the Tic-Tac-Toe game. The root node is the starting position of the board without any markings. The player playing 'X' has three possible moves (leaving aside the symmetrical positions of the board). These are shown in the next level of nodes, called the children nodes. The branches connecting the parent nodes and the children nodes represent the moves made by the player. Each level of play is called a ply. Each ply is where the turn switches to the other player. The bottom nodes without any branches are called leaves and signify the end of the game.

MINMAX ALGORITHM

The *MiniMax* algorithm (Plaat, 1997) developed by the AI community, provides a winning strategy in board games with perfect information. It is an intelligent approach to trace a winning path in the game tree given a heuristic or evaluation function that evaluates the goodness of the game position. "Min" and "Max" are the two players taking turns in the game. Max tries to maximize his/her score, while Min tries to minimize Max's score. Hence, the name *MiniMax* (or *MinMax*). The basic assumption is that both the players play an ideal game. While choosing the moves, Max will choose the best move to maximize his/her score while Min will choose the best move to minimize the score of Max.

Figure 2. MiniMax algorithm

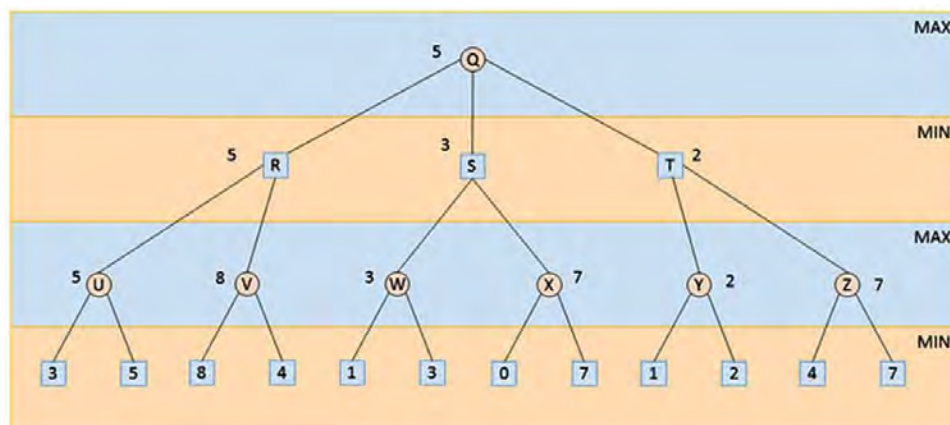


Table 1. State-space and game-tree complexity of board games

Game	State-Space Complexity	Game-Tree Complexity	Branching Factor	Average Game Length (Plies)
Tic-Tac-Toe	10^3	10^5	4	9
Connect Four	10^{13}	10^{21}	4	36
Othello	10^{28}	10^{58}	10	58
Checkers	10^{21}	10^{31}	-	-
Chess	10^{46}	10^{123}	35	80
Shogi	10^{71}	10^{226}	92	115
Go	10^{72}	10^{360}	250	150

Let us assume that Max (computer program) has won the toss to begin the play. The *MiniMax* algorithm proceeds through the following steps to determine the optimal winning strategy for *Max*.

1. Generate the entire game tree from the root node to the leaf nodes.
2. Evaluate the score of each leaf node by applying the evaluation function.
3. Back-up values from the leaf nodes through the branch nodes.
 - a. Assign the *maximum* value of the child nodes to the Max node.
 - b. Assign the *minimum* value of the child nodes to the Min node.
4. Terminate the backup procedure at the root node.
5. Follow the (winning) path from the root node to the leaf node traced through steps 1-4.

ALPHA-BETA PRUNNING

The game tree represents all the possible legal moves in the board game and the *MiniMax* algorithm provides an efficient way of searching for the successive winning moves throughout the nodes and branches of the game tree. However, real-life implementation of the *MiniMax* algorithm down to the last leaf nodes of the game tree is impossible. The state-space and game tree complexity, as shown in Table 1 are prohibitively large (Nilsson, 1998; Iida, Sakuta, & Rollason, 2002; Luger,

2005; Russell, & Norvig, 2010; Calero, 2011). This problem is partially solved by applying the evaluation function to a couple of moves ahead and then back-calculating the goodness of the current playing node. Even this look-ahead strategy is computationally expensive. The α - β version of the *MiniMax* algorithm (Pearl, 1982) further reduces the search complexity by effectively cutting off some of the branches of the game tree.

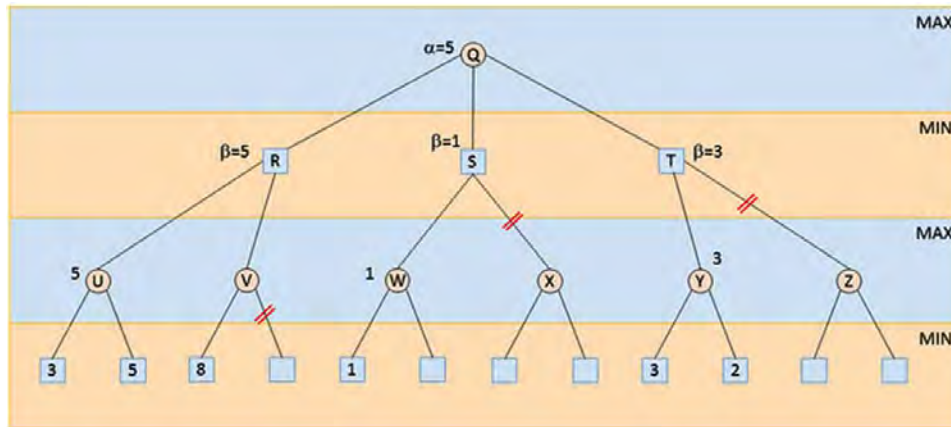
The algorithm maintains two values, α and β , which represent the maximum score that the maximizing player is assured of and the minimum score that the minimizing player is assured of, respectively. Initially α is negative infinity and β is positive infinity, i.e. both players start with their lowest possible score. It can happen that when choosing a certain branch of a certain node the minimum score that the minimizing player is assured of becomes less than the maximum score that the maximizing player is assured of ($\beta \leq \alpha$). If this is the case, the parent node should not choose this node, because it will make the score for the parent node worse. Therefore, the other branches of the node do not have to be explored.

The α and the β values during the search are computed as follows: (Nilsson, 1988; Luger, 2005)

The α value of a *Max* node is set equal to the current largest final backed-up value of its successors.

The β value of a *Min* node is set equal to the current smallest final backed-up value of its successors.

Figure 3. Alpha-Beta pruning



The search is discontinued under the following rules:

- **α Cut-Off Rule:** Search can be discontinued below any *Min* node having a β value less than or equal to the α value of any of its *Max* node ancestors. The final backed-up value of this *Min* node can then be set to its β value.
- **β Cut-Off Rule:** Search can be discontinued below any *Max* node having an α value greater than or equal to the β value of any of its *Min* node ancestors. The final backed-up value of this *Max* node can then be set to its α value.

For example starting at node Q (Figure 3), the algorithm descends in depth-first order and computes the value of the evaluation function for each of the leaf nodes. These values are backed up to the parent node U (Max = 5). This is then assigned to the grandparent node Q, as its β value. It cannot have a value larger than 5. The algorithm then descends to node Q's grandchildren and the search of their grandparent is terminated if any grandchildren is greater than or equal to Q's β . Node V is β pruned as shown in Figure 3. The algorithm continues pruning the tree with α and β cuts till the end of the right branch.

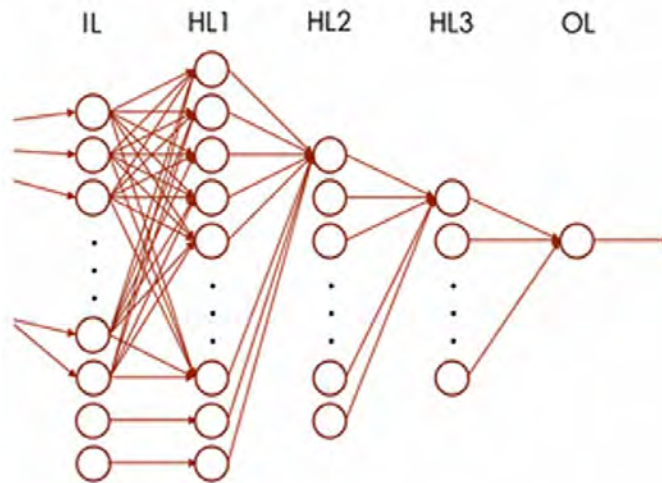
MACHINE LEARNING TECHNIQUES IN BOARD GAMES

Machine Learning and Neural Networks

Machine Learning (Mitchell, 1997; Mohri, & Rostamizadeh, 2012; Salev-Shwartz, & Ben-David, 2014) is an important research area in AI. It deals with three types of learning algorithms – (1) Supervised learning or classification in which a set of input-output data are provided to perform the task of learning or deriving the input-output pattern or mapping in the data; (2) unsupervised learning or clustering in which the algorithm learns to organize the data in clusters based on some entropy or statistical measures, and (3) reinforced learning in which the learning algorithm is given positive accumulative score for every successful performance.

The most common structures for implementing machine learning are Artificial Neural Networks (Anderson, 1995; Picton, 2000; Du, K.-L., Swamy, M. N. S., 2014). An Artificial Neural Network (ANN) consists of several layers, namely: an input layer, a number of hidden layers and an output layer. The input layer and the hidden layer are connected by links called weights and likewise the hidden layer and output layer also have connection weights. When more than one hidden layer

Figure 4. Multilayered neural network



exists, weights exist between such layers. Neural networks use some sort of learning rule by which the connection weights are determined in order to minimize the error between the neural network output and the desired output. In many applications, using only a single hidden layer is common. Each neuron in the hidden layer sums the weighted inputs from the input layer and passes the sum through the non-linear sigmoidal or hyperbolic activation function. A similar processing takes place in the output layer resulting in an output.

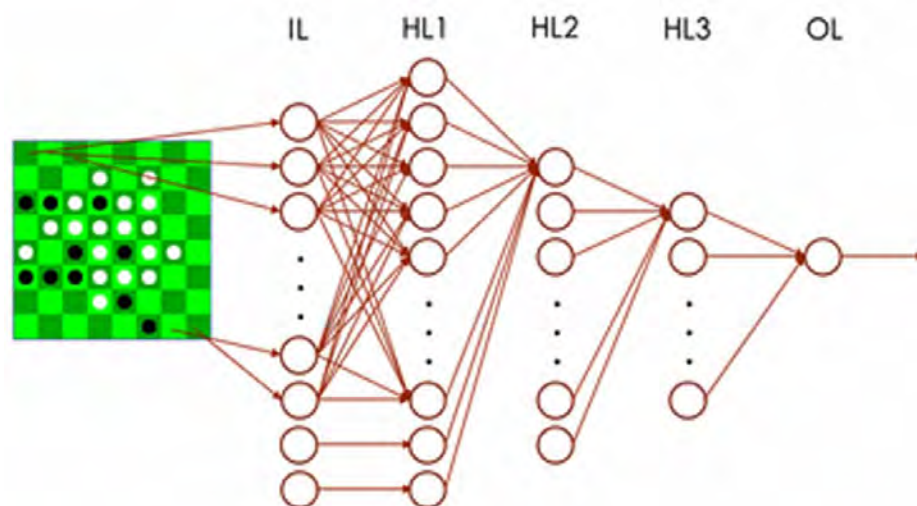
Neural Networks in Game Playing

ANN learning consists in adjusting the connection weights. These represent the strength of the synapses among the neurons in the human brain. Learning is conventionally carried out by the Back-Propagation Algorithm (BP) (Abid, Fnaiech, & Najim, 2001; Xinghuo, Efe, & Kaynak, 2002). Recently, Evolutionary Algorithms (Carpentieri, 2009; Gill, Singh, & Singh, 2010; Yorita, & Kubota, 2011,) and Swarm Intelligence algorithms (Van den Bergh, & Engelbrecht, 2000; Selles, & Rylander, 2002; Zhang & Zhang, 2007; Jin, Chang, Cheng, & Jiang, 2011; Gonsalves, & Oguro, 2014;) have also been applied to further improve the BP performance in learning.

ANN are also found to learn and excel in playing games without injecting any expert knowledge about how to play the game (Fogel, 1995). Some studies use neural network evolution through learning to guide the game tree search (Lee, 1990; Moriarty, & Miikkulainen, 1995) while others evolve neural networks to act as evaluation functions for the minimax search (Fogel, 1995; Chellapilla, & Fogel, 1999;). Initially, the connection weights are randomly generated and some relevant information such as the board configuration, the number of pieces on the board, etc. is fed to the input layer of the ANN. The neurons in the ANN layers fire and give an output, which usually indicates the move to be made. Several agents (programs), each driven by an ANN, play against each other till they reach the end of the game. The winning agent is given a reinforcing reward, and the learning cycle is repeated. The learning agent improves its playing strategy with each learning cycle.

For example (refer to Figure 5), in the case of an agent learning to excel at the game of othello, the board configuration, the piece differential and the legal moves from the current position are fed to the input neurons in each learning cycle (Lee, 1990; Chong, Ku, Lim, Tan, & White, 2003). The ANN outputs the move the agent is to play next. The opponent is normally another agent driven

Figure 5. Neural network learning othello wining moves



by expert heuristics or by another ANN. After the response of the opponent, the agent calculates the next move as the output of the ANN. This sequence of actions continues till the end of the game. A large number of games are played to achieve the desired level of expertise. After each game, the connection weights of the ANN are adjusted either by the back propagation algorithm or some optimization algorithm.

FUTURE RESEARCH DIRECTIONS

Evolutionary Algorithms, Neural Networks and Fuzzy systems form the basis of Computational Intelligence – a new version of AI with considerable success in the solution of practical problems. Evolutionary algorithms (EA) and Swarm Intelligence (SI) algorithms are called bio-inspired algorithms since their computational models are inspired by the biological processes and natural phenomena. The Genetic Algorithm (Sivanandam, & Deepa, 2007) and Genetic Programming (Reilly, 2005) are typical EAs directly modelled on the Darwinian natural selection principles, while the Swarm Intelligence algorithms are modelled on the feeding behavior of a swarm of birds or a school of fish (Engelbrecht, 2005). EA and SI

techniques have been primarily applied to solve optimization problems in diverse fields. They are gradually finding their way into the arena of PC board games. For example, they can learn the best playing strategy by efficiently searching through the massive game-tree (Karamchandani, Gandhi, Pawar, & Pawaskar, 2015; Ling & Lam, 2011). These techniques are also hybridized with ANN to produce advanced systems for efficient learning of game-playing strategies (Cant, Churchill, & Al-Dabass, 2002).

Monte Carlo Tree Search (MCTS) is a random sampling tree search method that creates a partial game tree and then searches for the best move based on its simulated gametheoretic value (Browne, Powley, Whitehouse, Lucas, Cowling, Rohlfshagen, Tavener, Perez, & Colton, 2012). MCTS method has been successfully applied to the games of Go (Bouzy, 2006), Ms Pac-Man (Samothrakis, Robles, & Lucas, 2011), and so on. Inexpensive yet powerful hardware and advanced graphics techniques coupled with newer versions of the old AI algorithms (Díez, Laforge, & Saerens, 2003) have contributed to the revival of the two-player board games. Card games which involve an element of chance were not considered in the “intelligent games” categories like chess, othello, go, etc. However, the present-day AI has

begun embracing card games also (Cowling, Ward, & Powley, 2012; Sephton, Cowling, Powley, & Slaven, 2014). This is possible because of the successful machine learning algorithms.

CONCLUSION

This chapter has explored some of the classic AI algorithms for playing board games like tic-tac-toe, checkers, othello, chess, etc. The two-player zero-sum deterministic board games are the classic games to attract the attention of the (early) AI researchers. The playing strategy of the computer program consists in presenting the moves of the two players in the form of a game tree, applying a suitable heuristic evaluation function to the game tree end nodes and backtracking to the current board position in the game. This kind of an ideal search is performed using the *MiniMax* algorithm. However, even in the case of a simple game like Tic-Tac-Toe the number of nodes soon become astronomical with the depth of the game tree, making the exhaustive search impractical and infeasible. The alpha-beta pruning offers an effective way to safely prune the futile branches. A recent AI method in game playing is Machine Learning using Neural Networks. The program learns to master the winning strategy through trial-and-error and at times through supplementing expert knowledge. Some of the latest machine learning algorithms combine the power of Evolutionary Algorithms or Swarm Intelligence with Artificial Neural Networks. Development of Monte Carlos Tree Search and deep-learning techniques hold the future for game-playing AI.

REFERENCES

- Abid, S., Fnaiech, F., & Najim, M. (2001). A fast feedforward training algorithm using a modified form of the standard backpropagation algorithm. *IEEE Transactions on Neural Networks*, 12(2), 424–430. doi:10.1109/72.914537 PMID:18244397
- Anderson, J. (1995). *An introduction to neural networks*. Cambridge, MA: MIT Press.
- Bouzy, B. (2006). Move-Pruning techniques for monte-carlo go. In *Advances in Computer Games* (pp. 104–119). Springer Berlin Heidelberg. doi:10.1007/11922155_8
- Browne, C. B., Powley, E., Whitehouse, D., Lucas, S. M., Cowling, P., Rohlfshagen, P., & Colton, S. et al. (2012). A survey of monte carlo tree search methods. *Computational Intelligence and AI in Games. IEEE Transactions on*, 4(1), 1–43.
- Buro, M. (1997). The Othello match of the year: Takeshi Murakami vs. Logistello. *International Journal of Computer Chess Association*, 20(3), 189–193.
- Buro, M. (2002). Improving heuristic mini-max search by supervised learning. *Artificial Intelligence*, 134(1-2), 85–99. doi:10.1016/S0004-3702(01)00093-5
- Calero, P. (2011). *Artificial intelligence for computer games*. New York: Springer. doi:10.1007/978-1-4419-8188-2
- Cant, R., Churchill, J., & Al-Dabass, D. (2002). A hybrid artificial intelligence approach with application to games. *Proceedings of the 2002 International Joint Conference on In Neural Networks*, (Vol. 2, pp. 1575-1580). IEEE.
- Carpentieri, M. (2009). A Genetic System Based on Simulated Crossover: Stability Analysis and Relationships with Neural Nets. *Evolutionary Computation*, 17(2), 167–201.
- Chellapilla, K., & Fogel, D. B. (1999). Evolving Neural Networks to Play Checkers without Expert Knowledge. *IEEE Transactions on Neural Networks*, 10(6), 1382–1391. doi:10.1109/72.809083 PMID:18252639
- Chellapilla, K., & Fogel, D. B. (1999). Evolution, neural networks, games, and intelligence. *Proceedings of the IEEE*, 87(9), 1471–1496. doi:10.1109/5.784222

- Chong, S. Y., Ku, D. C., Lim, H. S., Tan, M. K., & White, J. D. (2003). *Evolved neural networks learning Othello strategies. The 2003 Congress on Evolutionary Computation* (Vol. 3, pp. 2222–2229). Academic Press.
- Cowling, P. I., Ward, C. D., & Powley, E. J. (2012). Ensemble Determinization in Monte Carlo Tree Search for the Imperfect Information Card Game Magic: The Gathering. *IEEE Transactions on Computational Intelligence and AI in Games*, 4(4), 241–257.
- Díez, S. G., Laforge, J., & Saerens, M. (2013). Rminimax: An Optimally Randomized MINIMAX Algorithm. *IEEE Transactions on Cybernetics*, 43(1), 385–393. doi:10.1109/TSMCB.2012.2207951 PMID:22893439
- Du, K.-L., & Swamy, M. N. S. (2014). *Neural Networks and Statistical Learning*. London: Springer-Verlag. doi:10.1007/978-1-4471-5571-3
- Engelbrecht, A. P. (2005). *Fundamentals of Computational Swarm Intelligence*. Wiley.
- Fogel, D. B. (1995). Using Evolutionary Programming to Construct Neural Networks that are Capable of Playing Tic-Tac-Toe. *Proceedings of 1995 IEEE International Conference on Neural Networks*, 875-880.
- Fogel, D. B. (2002). *Blondie24: Playing at the Edge of AI*. San Francisco, CA: Morgan Kaufmann. doi:10.1016/B978-155860783-5/50016-7
- Gill, J., Singh, B., & Singh, S. (2010). Training back propagation neural networks with genetic algorithm for weather forecasting. *Intelligent Systems and Informatics*, 465-469.
- Gonsalves, T., & Oguro, A. (2014). Feed-forward Neural Network Training using Particle Swarm Optimization. *Advances in Computer Science and Engineering*, 1, 1–13.
- Hsu, F.-h. (2002). *Behind Deep Blue*. Princeton, NJ: Princeton University Press.
- Iida, H., Sakuta, M., & Rollason, J. (2002). Computer shogi. *Artificial Intelligence*, 134(1-2), 121–144. doi:10.1016/S0004-3702(01)00157-6
- Jin, C., Chang, G., Cheng, W., & Jiang, H. (2011). *Improved particle swarm optimization for fuzzy neural network training* (pp. 299–302). Genetic and Evolutionary Computing.
- Kaindl, H. (1990). Tree searching algorithms. In T. A. Marsland & J. Schaeffer (Eds.), *Computers, Chess, and Cognition* (pp. 133–168). New York: Springer. doi:10.1007/978-1-4613-9080-0_8
- Karamchandani, S., Gandhi, P., Pawar, O., & Pawaskar, S. (2015). A simple algorithm for designing an artificial intelligence based Tic Tac Toe game. *2015 International Conference on Pervasive Computing* (pp. 1-4). IEEE. doi:10.1109/PERVASIVE.2015.7087182
- Lee, K. F., & Mahajan, S. (1990). The Development of a World Class Othello Program. *Artificial Intelligence*, 43, 21–36.
- Ling, S. H., & Lam, H. K. (2011). Playing Tic-Tac-Toe Using Genetic Neural Network with Double Transfer Functions. *Journal of Intelligent Learning Systems and Applications*, 3(1), 37–44.
- Luger, G. F. (2005). *Artificial intelligence: Structures and strategies for complex problem solving* (5th ed.). Harlow, UK: Addison-Wesley.
- Marsland, S. (2014). *Machine Learning: An Algorithmic Perspective* (2nd ed.). Chapman and Hall/CRC. doi:10.1201/b16812-42
- Millington, I., & Funge, J. (2009). *Artificial intelligence for games* (2nd ed.). Burlington, MA: Morgan Kaufmann/Elsevier.
- Mitchell, T. (1997). *Machine Learning*. New York: McGraw-Hill.
- Mohri, M., & Rostamizadeh, A. (2012). *Foundations of machine learning*. Cambridge, MA: MIT Press.

- Nilsson, N. (1998). *Artificial Intelligence a new synthesis*. San Francisco, CA: Morgan Kaufmann.
- Pearl, J. (1982). The Solution for the Branching Factor of the Alpha-beta Pruning Algorithm and its Optimality. *Communications of the ACM*, 25(8), 559–564. doi:10.1145/358589.358616
- Picton, P. (2000). *Neural networks*. Basingstoke, UK: Palgrave.
- Plaat, A. (1997). *A Minimax Algorithm faster than NegaScout*. Rotterdam: Erasmus University.
- Plaata, A., Schaeffer, J., Pijlsa, W., & de Bruina, A. (1999). Best-first fixed-depth minimax algorithms. *Artificial Intelligence*, 87(1-2), 255–293. doi:10.1016/0004-3702(95)00126-3
- Reilly, U. (2005). *Genetic programming theory and practice II*. New York: Springer Science Business Media.
- Russell, S., & Norvig, P. (2010). *Artificial intelligence: A modern approach* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.
- Salev-Shwartz, S., & Ben-David, S. (2014). *Understanding machine learning: From foundations to algorithms*. New York: Cambridge University Press. doi:10.1017/CBO9781107298019
- Samothrakis, S., Robles, D., & Lucas, S. (2011). Fast approximate max-n monte carlo tree search for ms pac-man. *Computational Intelligence and AI in Games. IEEE Transactions on*, 3(2), 142–154.
- Samuel, A. (1960). Programming computers to play games. In F. Alt (Ed.), *Advances in Computers* (Vol. 1, pp. 165–192). New York: Academic Press.
- Samuel, A. (1967). Some studies in machine learning using the game of checkers: Recent progress. *IBM Journal*, 11(6), 601–617. doi:10.1147/rd.116.0601
- Schaeffer, J., & van den Herik, H. J. (1997). *One Jump Ahead: Challenging Human Supremacy in Checkers*. Berlin: Springer.
- Schaeffer, J., & van den Herik, H. J. (2002). Games, computers, and artificial intelligence. *Artificial Intelligence*, 134(1-2), 1–7. doi:10.1016/S0004-3702(01)00165-5
- Selles, M., & Rylander, B. (2002). Neural network learning using particle swarm optimizers. *Advances in Information Science and Soft Computing*, 22(11), 224–226.
- Sephton, N., Cowling, P. I., Powley, E., & Slaven, N. H. (2014). Heuristic move pruning in Monte Carlo Tree Search for the strategic card game Lords of War. *2014 IEEE Conference on Computational Intelligence and Games (CIG)*, 1-7. doi:10.1109/CIG.2014.6932892
- Shannon, C. (1950). Programming a computer for playing chess. *Philosophical Magazine*, 41(314), 256–275. doi:10.1080/14786445008521796
- Sivanandam, S., & Deepa, S. (2007). *Introduction to genetic algorithms*. Berlin: Springer.
- Van den Bergh, F., & Engelbrecht, A. P. (2000). Cooperative learning in neural network using particle swarm optimizers. *South African Computer Journal*, 26, 84–90.
- Vijay, A. M. (1997). *Training fuzzy neurons with backpropagation algorithm* (Unpublished doctoral dissertation). Texas A & M University, College Station, TX.
- Xinghuo, Y., Efe, M. O., & Kaynak, O. (2002). A general backpropagation algorithm for feed-forward neural networks learning. *IEEE Transactions on Neural Networks*, 13(1), 251–254. doi:10.1109/72.977323 PMID:18244427
- Yazdani, M. (1986). *Artificial intelligence: Principles and applications*. London: Chapman and Hall.
- Yorita, A., & Kubota, N. (2011). Cognitive Development in Partner Robots for Information Support to Elderly People. *IEEE Transactions on Autonomous Mental Development*, 3(1), 64–73. doi:10.1109/TAMD.2011.2105868

Yu, X., & Gen, M. (2010). *Introduction to evolutionary algorithms*. London: Springer. doi:10.1007/978-1-84996-129-5

Zhang, J. R., Zhang, J., Lok, T.-M., & Lyu, M. R. (2007). A hybrid particle swarm optimization-backpropagation algorithm for feed forward neural network training. *Applied Mathematics and Computation*, 185(2), 1026–1037. doi:10.1016/j.amc.2006.07.025

ADDITIONAL READING

Alhejali, A. M., & Lucas, S. M. (2013). Using genetic programming to evolve heuristics for a Monte Carlo Tree Search Ms Pac-Man agent. *2013 IEEE Conference on Computational Intelligence in Games (CIG)*, (pp.1-8). IEEE. doi:10.1109/CIG.2013.6633639

Bouzy, B., & Cazenave, T. (2001). Computer Go: An AI oriented survey. *Artificial Intelligence*, 132(1), 39–103. doi:10.1016/S0004-3702(01)00127-8

Haupt, R., & Haupt, S. (2004). *Practical genetic algorithms* (2nd ed.). Hoboken, N.J.: John Wiley.

Mu, Z., & Liu, K. (2000). Application of genetic algorithms in identification of linear time-varying system. *Journal of University of Science and Technology*, 7(1), 58–624.

Perez, D., Mostaghim, S., Samothrakis, S., & Lucas, S. (2014). Multi-Objective Monte Carlo Tree Search for Real-Time Games. *IEEE Transactions on Computational Intelligence and AI in Games*, vol. PP, no.99, pp.1-14.

Samarasinghe, S. (2006). *Neural Networks for Applied Sciences and Engineering: From Fundamentals to Complex Pattern Recognition*. Auerbach Publications. doi:10.1201/9780849333750

Singh, D., Thaker, C. S., & Shah, S. M. (2011). Quality of state improvisation through evaluation function optimization in genetic application learning. *2011 International Conference on Emerging Trends in Networks and Computer Communications (ETNCC)*, (pp. 93-97). IEEE. doi:10.1109/ETNCC.2011.5958494

Whitehouse, D., Powley, E. J., & Cowling, P. I. (2011). Determinization and information set Monte Carlo Tree Search for the card game Dou Di Zhu. *2011 IEEE Conference on Computational Intelligence and Games (CIG)*, (pp.87-94). IEEE. doi:10.1109/CIG.2011.6031993

Wu, X., & Zhang, S. (2011). The study and application of artificial intelligence pathfinding algorithm in game domain, *2011 International Conference on Computer Science and Service System (CSSS)*, pp.3772-3774.

Yu, X., & Gen, M. (2010). *Introduction to evolutionary algorithms*. London: Springer. doi:10.1007/978-1-84996-129-5

Zhang, H., Shao, H., & Li, Y. (2000). Particle swarm optimization for evolving artificial neural network. In *Proceedings of the IEEE International Conference on System, Man and Cybernetics*, vol. 4, pp. 2487-2490. doi:10.1109/ICSMC.2000.884366

KEY TERMS AND DEFINITIONS

Artificial Intelligence: It is a branch of computer science concerned with making computers do things that require intelligence when done by human beings.

Board Game: It is a game in which two or more players take turns in placing or moving pieces on a pre-marked surface or board, according to a set of rules.

Game Tree: It is a graph consisting of nodes and edges that represent all the possible legal moves of a (board) game. The nodes are positions in a game and the edges are moves.

Machine Learning: Refers to the activity or ability of computer programs to learn without being explicitly programmed given an adequate amount of data.

MiniMax Algorithm: It is a winning search strategy through the nodes and branches of the game tree. It is based on the principal that the algorithm's opponent will try to minimize whatever value the algorithm tries to maximize.

Neural Network: A network consisting of several layers of nodes and connections with weights which models the information processing operation of the human brain. Neural networks are used in Machine Learning as universal function approximators.

Ply: Refers to the number of levels in the game tree including the root level.

Zero-Sum Game: It is a situation in which one player's win implies the other's loss. The sum of the scores in the win-loss or the draw is always zero.

Computational Intelligence Approaches to Computational Aesthetics

Erandi Lakshika

University of New South Wales, Australia

Michael Barlow

University of New South Wales, Australia

INTRODUCTION

Computational aesthetics is an area of research which attempts to develop computational methods that can perform aesthetic judgements in the same way as humans (Hoenig, 2005). It is an area of research which has not developed as a separate discipline till relatively recently. The notion of aesthetics is highly intuitive and often subjective. An aesthetic experience can be negative, positive or more subtly nuanced. Human beings have a strong and deep sense of aesthetics, however rationalising aesthetic decisions is challenging. As such developing computational models to make aesthetic decisions is particularly challenging.

While computational intelligence techniques such as evolutionary algorithms have been able to solve many real world challenges, still such techniques are not widely used to solve problems that involve aesthetic decisions. Making an aesthetic decision often requires a human in the loop which in turn creates a barrier between computational intelligence and aesthetics. However recent advancements in computational aesthetics have made computer generated art and aesthetics realisable in several domains (den Heijer & Eiben, 2012; DiPaola & Gabora, 2009).

The purpose of this article is to summarise the advancements in the area of computational aesthetics, challenges involved, computational intelligence approaches to art and aesthetics and possible future directions. The article first summarises early attempts to define aesthetics,

through to more contemporary definitions and attempts in developing computational models of aesthetics in various domains. Then, it highlights the challenges associated with bridging the gap between aesthetics and computational intelligence. Thereafter it discusses how computational intelligence techniques are used in art and aesthetics ranging from simple classification problems to more advanced problems such as automatic generation of art artefacts, stories and simulations. The article concludes highlighting the future research directions that need to be undertaken in order to make significant advancements in computational aesthetics and its applications.

BACKGROUND

Aesthetics

The study of aesthetics is chiefly a branch of philosophy with links to other disciplines such as psychology. The term aesthetics was derived from the Greek word *aisthanesthai* (to perceive (by the senses or by the mind)) and introduced into the philosophical terminology in the eighteenth century (Saw & Osborne, 1960). The definition of aesthetics is a long standing debate. Early definitions of aesthetics are related to art or beauty (Santayana, 1904). Later attempts to define aesthetics discuss that aesthetics mean more than just art and natural beauty (Walton, 2007), (Palmer, Schloss, & Sammartino, 2013). Therefore more

contemporary definitions are woven around human mental processes involved in making aesthetic judgements; for example:

- *The study of human minds and emotions in relation to the sense of beauty (Palmer et al., 2013).*
- *Psychological mechanisms that allow humans to experience and appreciate a broad variety of objects and phenomena, including utensils, commodities, designs, other people, or nature, in aesthetic terms (beautiful, attractive, ugly, sublime, picturesque, and so on) (Leder & Nadal, 2014).*

Computational Aesthetics

Computational aesthetics is an area of research which attempts to develop computational methods that can perform aesthetic judgements in the same way as humans (Hoenig, 2005). Classifying something as aesthetically appealing or not appealing might be relatively easy for a human even though it is subjective; however for a computer it is not straight forward to make such determinations. As such, a considerably large amount of literature attempts to define explicit measures of aesthetics that make it possible to distinguish aesthetically appealing objects from objects that are not.

In 1933, Birkhoff defined an aesthetic measure M , which can be explained as the regularity (order) perceived in an aesthetic object for a unit of effort (complexity) (Birkhoff, 1933). This can be defined as $M = \frac{\text{order}}{\text{complexity}}$. The introduction of this measure is considered as the beginning of computational aesthetics. This measure was investigated by later researchers in relation to information theory. In 2007, Rigau, Feixas, and Sbert (2007) analysed the concepts of order and complexity for pixel values of paintings based on Shannon entropy (Shannon, 1948) and Kolmogorov complexity (Kolmogorov, 1965) and presented a new conceptualisation of Birkhoff's

aesthetic measure based on Zurek's physical entropy (Zurek, 1989). However, vision researchers argue that true entropy of an image, being a perceptual quantity, cannot be characterised with image statistics such as Shannon entropy (Graham & Redies, 2010).

Vision scientists have analysed whether humans have a preference for images that have similar spatial statistical structures to those that the human visual system has adapted for (Palmer et al., 2013). For example, Graham and Field (2007) have identified that paintings (a sample taken from a Museum of Art) and natural scenes have similar Fourier amplitude spectra. Similarly, Spehar, Clifford, Newell, and Taylor (2003) show that humans have a preference for fractal structures irrespective of whether they are natural or man-made.

There is a considerable amount of literature that discuss the shape of an object as an aesthetic measure. Several researchers have reviewed Golden Ratio/Golden Section as an aesthetic measure in various fields (Benjafeld, 1976; Gilmartin, 1983; Green, 1995; McManus, 1980; Russell, 2000). If a line is divided into x : $(1-x)$ ratio, the Golden Ratio arises when $1/x = x/(1-x)$ (Markowsky, 1992). This value is denoted by F and $F \approx 1.618$. However, the results in the field are controversial and some studies have shown that the Golden Ratio have no special aesthetic appeal (Boselie, 1984; Davis & Jahnke, 1991; Markowsky, 1992).

Some studies have discussed the symmetry of an object as an aesthetic factor (Locher & Nodine, 1989; Osborne, 1986). Several recent such studies have shown that people have greater preference towards curved objects than objects with sharp contours (Bar & Neta, 2006; Silvia & Barona, 2009). However these measures are fragile and susceptible to many other factors depending on the domain of interest so that their applicability as an aesthetic measure is limited.

Although there is a huge amount of literature on attempts to measure aesthetics in various disciplines, currently there is no universal measure

of aesthetics. However, these various proposed aesthetic measures provide the basis for developing computational methods of aesthetics, enabling computational aesthetics to develop as a discipline.

In particular, computational aesthetic techniques can currently be used to make aesthetic decisions primarily in two different ways as follows.

- Aesthetic Classification.
- Aesthetic Scoring.

Aesthetic Classification

Categorising objects, situations or phenomena into several classes based on their aesthetic quality can be treated as an aesthetic classification problem. Several works attempt to automate human aesthetic judgement for classification problems by training a machine learning system. In particular these works demonstrate promising results of applying computational aesthetics techniques to train machine learning classifiers in order to predict aesthetic ratings or other subjective judgements such as interestingness of photographs (Datta, Joshi, Li, & Wang, 2006; Dhar, Ordonez, & Berg, 2011, paintings (Li & Chen, 2009) and consumer videos (Moorthy, Obrador, & Oliver, 2010).

Aesthetic Scoring

Rating objects, situations or phenomena for their aesthetic quality in a discrete or a continuous range can be treated as an aesthetic scoring problem. Similar to classification problems there are attempts to automate aesthetic scoring through machine learning approaches in several problem areas such as digital portraits (Battiato, Moltisanti, Ravi, Bruna, & Naccari, 2013) computer generated stories (Wang, Bui, Petraki, & Abbass, 2012) and computer simulation of emergent behaviours (Lakshika, Barlow, & Easton, 2012). Apart from aesthetic feature extraction, one of the primary challenges in both aesthetic classification and scoring problems is finding labelled training data.

Computational Intelligence

Computational intelligence is a branch of artificial intelligence which is aimed at developing systems that can demonstrate intelligent behaviour in complex and changing environments based on adaptive methods and algorithms that are able to learn new situations, generalise, abstract and associate (Engelbrecht, 2007; Kruse et al., 2013). Often computational intelligence techniques apply nature-inspired approaches to find approximate solutions to problems that are difficult to solve using traditional methods. However, being nature-inspired is not a necessary trait for a technique to be categorised as computational intelligence.

Computational intelligence techniques have been able to solve many challenging real world problems including classification, planning, decision making and optimisation in a range of application areas such as engineering, economics, defence and medicine. However, when it comes to problems that involve aesthetic decisions, computational intelligence techniques are still not widely used as compared to the other problem areas.

Aesthetics are involved in many aspects of everyday life, and people make aesthetic decisions consciously or subconsciously. In the contemporary society where automation has become a major part of people's lives, making computers more aesthetic-aware has become a need so that the decisions assisted by the computers (e.g: designing a city, designing a computer game) match with human aesthetic preferences. This is where integrating computational intelligence techniques and aesthetics becomes particularly important.

COMPUTATIONAL INTELLIGENCE APPROACHES TO COMPUTATIONAL AESTHETICS

The recent re-emergence of the term computational aesthetics is related to computational intelligence, especially in relation to evolutionary art and music

(Greenfield, 2005). Evolutionary computation is a computational intelligence technique inspired by the evolution of biological species in nature and utilises the underlying abstract concepts of Darwinian evolution: natural selection and survival of the fittest to perform a stochastic search for solutions to a problem. A detailed introduction to evolutionary computing can be found in (Eiben & Smith, 2003). According to Darwinian principles, in nature, evolution occurs by combining the genes of parents to produce a new offspring (recombination). During this reproduction process small random variations (mutations) occur and new traits are produced. These new individuals are evaluated for environmental conditions based on an objective function (fitness evaluation). As per the concept of natural selection, those individuals who adapt to the environment conditions best, have a better chance of survival (i.e: the concept of survival of the fittest). According to this process, the individuals with higher fitness have more chance to reproduce than those with a lower fitness and favourable traits are passed to the offspring. Through mutations, new traits are also introduced. Therefore, as this process iteratively continues, the populations evolve towards better quality individuals. Evolutionary algorithms mimic this process in order to search for/create better solutions.

Evolutionary art and aesthetics focus on using evolutionary computing techniques such as natural selection and survival of the fittest in order to automatically generate art and aesthetics. For an evolutionary algorithm to automatically generate art and aesthetics a well-formulated objective function is required. Due to the inherent subjective nature of aesthetics judgements, defining an appropriate objective function is extremely challenging for aesthetic related applications.

Interactive evolutionary algorithms have been introduced to overcome this challenge by having a human in the loop providing aesthetic judgements. An interactive evolutionary algorithm works similarly to any other standard evolutionary algorithm with the exception of having the

interaction of human/humans to determine the fitness of the solutions. Interactive evolutionary techniques have been employed to evolve many systems that involve creativity and aesthetics such as music (Tokui & Iba, 2000), paintings (Aupetit, Bordeau, Monmarché, Slimane, & Venturini, 2003), graphics (McCormack, 1993; Sims, 1991) and various design problems (Kim & Cho, 2000; Oliver, Monmarché, & Venturini, 2002). Further, Chen, Kobayashi, Kawabayashi, and Huang (2008) have employed an interactive genetic algorithm to optimise the coefficients (weights) of an extended boid model (Reynolds, 1987) to simulate fish schooling behaviour. Karl Sims' work (Sims, 1991) is one of the notable examples of evolutionary arts and aesthetics. This work used an expression-based approach to evolve images using an interactive selection mechanism, demonstrating the feasibility of automatic generation of beautiful and complex images using such approaches. This inspired many artists and graphics programmers to apply similar techniques for automatic image generation and substantial bodies of work have been done to date.

One of the common problems in an interactive evolutionary system is user fatigue (Takagi, 2001). Users become inconsistent over time and consequently, evolutionary population size needs to be kept low. While some authors (Draves, 2005) have employed crowdsourcing to address user fatigue, there is a trend of seeking automated (non-interactive) mechanisms to determine human aesthetic judgement through computational aesthetic techniques.

The work by den Heijer and Eiben (2010) analyses three aesthetics measures based on Benford Law, Global contrast factor, and Information theory to evolve art. The three aesthetics measures have generated different styles of images and cross-evaluation of each aesthetics measure on the other two measures confirmed that no measure appears to appreciate the images generated by the other measures. den Heijer and Eiben (2012) present some early results in evolving pop art using colour contrast as an objective function for

a Scalable Vector Graphics (SVG) based genetic programming system. DiPaola and Gabora (2009) demonstrate the evolution of abstract portraits through a genetic programming based approach utilising an automated fitness function based on resemblance, composition, tonality and colour rules. However, it would be interesting to see a cross evaluation of these evolutionary art and paintings with artists or humans.

Baluja, Pomerleau, and Jochem (1994) present an early attempt to automate human judgement on the aesthetics of images through training a neural network with the intent of using that as an objective function to an evolutionary algorithm. The approach uses the two-dimensional pixels of an image as input to the neural network and human rating for them to train the network. The results showed that the automated system can prune out the bad images; however the performance on good images was not as accurate. This is due to the lack of good images in the training set, which they collected through several interactive evolutions. However, this research opened up some interesting research directions in automating human judgement.

Machado, Romero, and Manaris (2008) co-evolved the classifier and the evolutionary system in order to explore the images that are classified as paintings by the neural network. The system generated images are assumed to have a low aesthetics value and external images (a set of famous paintings) are assumed to have a high aesthetics value. The neural network was trained iteratively by inputting the solutions found during the evolution to the training set. This overcomes the limitation of lack of training images in the training set; however, the work does not address the challenge of completely automating human aesthetic judgement.

More recent work by Wang et al. (2012) attempts to define an approximated human story evaluation model in order to objectively evaluate the quality of computer generated stories. This work presents correlations between several objective metrics and subjective human evaluations so

that the objective metrics can be used as surrogates for human evaluations and be used as an objective function for an evolutionary algorithm. Lakshika et al. (2012) show that human judgement of aesthetic quality of standing group conversation behaviour simulations can be recreated by a machine learning system. In this work, the features used by humans to make aesthetic judgements are used to extract features from the simulations and such features were input to the sequential minimal-optimisation algorithm for learning a support vector regression model (Smola, & Schölkopf, 2004). In this work, leave-one-out cross validation results of the trained machine learning system showed a correlation co-efficient of 0.88 (in a range of 0-1, higher the co-efficient better the result) between the human scores and the machine generated scores; and a root mean squared error of 0.86 (in a range of 0-9, lower the error better the result). This machine learning system was then utilised to automatically generate high fidelity (aesthetically appealing) conversational group dynamics simulations via an evolutionary system. The work not only shows the ability of a machine learning system to capture human aesthetic judgement, but also the ability of a computational intelligence approach to generate aesthetically appealing crowd simulations. Later work by Lakshika, Barlow, and Easton (2016) explores a relationship between simulation model complexity and simulation fidelity in multi-agent systems by employing this framework in three problem domains, namely conversational group dynamics simulations, sheepdog herding simulations, and lane merge traffics simulations showing that the approach can be extended to automatically generate high fidelity multi-agent simulations.

FUTURE RESEARCH DIRECTIONS

While the field of computational aesthetics has made significant advancements over the recent decades, one of the limitations of the field is lack of validations. Current studies are limited to specific problems with limited cross evaluations;

hence large scale validations of computational aesthetic models are essential for their wider use in the tasks that traditionally require humans in the loop. This is particularly difficult as aesthetic decisions can reflect the opinion of a person, a group of people or multiple groups of people with varying demographics. While developing universal models of computational aesthetics is extremely challenging, research is required towards developing computational models of aesthetics that can adapt to various user groups and changing circumstances.

It is likely that merging computational aesthetic techniques with other complementary research techniques used in such areas as affective computing (Picard, 2003) and sentiment analysis (Liu, 2012) will greatly be beneficial to the field. This will also be useful to deal with the inherent philosophical and methodological issues arising from the subjective nature of aesthetics.

CONCLUSION

Computational aesthetics techniques have made significant advances over recent decades. These advances together with computational intelligence techniques have made computer generated art and aesthetics realisable in a number of targeted application areas. If these advancements continue, significant positive impacts on the areas that traditionally require a human in the loop to provide creative and aesthetic judgements are inevitable. The aim should not be to replace humans with a computer but to achieve more creative and novel outcomes by humans and computers working together complementing each other's skills.

REFERENCES

- Aupetit, S., Bordeau, V., Monmarch'e, N., Slimane, M., & Venturini, G. (2003). Interactive evolution of ant paintings. In *Proceedings of the 2003 Congress on Evolutionary Computation* (Vol. 2, pp. 1376–1383). doi:10.1109/CEC.2003.1299831
- Baluja, S., Pomerleau, D., & Jochem, T. (1994). Towards automated artificial evolution for computer-generated images. *Connection Science*, 6(2-3), 325–354. doi:10.1080/09540099408915729
- Bar, M., & Neta, M. (2006). Humans prefer curved visual objects. *Psychological Science*, 17(8), 645–648. doi:10.1111/j.1467-9280.2006.01759.x PMID:16913943
- Battiatto, S., Moltisanti, M., Rav`i, F., Bruna, A., & Naccari, F. (2013). *Aesthetic scoring of digital portraits for consumer applications* (pp. 866008–1, 866008–866010). IS&T/SPIE Electronic Imaging. doi:10.1117/12.2005211
- Benjafield, J. (1976). The golden rectangle: Some new data. *The American Journal of Psychology*, 89(4), 737–743. doi:10.2307/1421471 PMID:1020768
- Birkhoff, G. D. (1933). *Aesthetic measure*. Cambridge, MA: Harvard University Press. doi:10.4159/harvard.9780674734470
- Boselie, F. (1984). The aesthetic attractivity of the golden section. *Psychological Research*, 45(4), 367–375. doi:10.1007/BF00309712
- Chen, Y.-W., Kobayashi, K., Kawabayashi, H., & Huang, X. (2008). Application of interactive genetic algorithms to boid model based artificial fish schools. In *International Conference on Knowledge-based and Intelligent Information and Engineering Systems* (pp. 141–148). doi:10.1007/978-3-540-85565-1_18
- Datta, R., Joshi, D., Li, J., & Wang, J. (2006). Studying aesthetics in photographic images using a computational approach. In *Proceedings of the European Conference on Computer Vision* (pp. 288–301). Berlin: Springer-Verlag. doi:10.1007/11744078_23
- Davis, S. T., & Jahnke, J. C. (1991). Unity and the golden section: Rules for aesthetic choice? *The American Journal of Psychology*, 104(2), 257–277. doi:10.2307/1423158

- den Heijer, E., & Eiben, A. (2010). Using aesthetic measures to evolve art. In *Proceedings of IEEE Congress on Evolutionary Computation* (p. 1-8). doi:10.1109/CEC.2010.5586245
- den Heijer, E., & Eiben, A. (2012). Evolving pop art using scalable vector graphics. In *International Conference on Evolutionary and Biologically Inspired Music and Art* (pp. 48–59). doi:10.1007/978-3-642-29142-5_5
- Dhar, S., Ordonez, V., & Berg, T. (2011). High level describable attributes for predicting aesthetics and interestingness. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1657–1664). doi:10.1109/CVPR.2011.5995467
- DiPaola, S., & Gabora, L. (2009). Incorporating characteristics of human creativity into an evolutionary art algorithm. *Genetic Programming and Evolvable Machines*, 10(2), 97–110. doi:10.1007/s10710-008-9074-x
- Draves, S. (2005). The electric sheep screen-saver: A case study in aesthetic evolution. In *Applications of Evolutionary Computing* (pp. 458–467). Springer. doi:10.1007/978-3-540-32003-6_46
- Eiben, A. E., & Smith, J. E. (2003). *Introduction to evolutionary computing*. SpringerVerlag. doi:10.1007/978-3-662-05094-1
- Engelbrecht, A. P. (2007). *Computational intelligence: an introduction*. John Wiley & Sons. doi:10.1002/9780470512517
- Gilmartin, P. P. (1983). Aesthetic preferences for the proportions and forms of graticules. *The Cartographic Journal*, 20(2), 95–100. doi:10.1179/caj.1983.20.2.95
- Graham, D. J., & Field, D. J. (2007). Statistical regularities of art images and natural scenes: Spectra, sparseness and nonlinearities. *Spatial Vision*, 21(1), 149–164. doi:10.1163/156856807782753877 PMID:18073056
- Graham, D. J., & Redies, C. (2010). Statistical regularities in art: Relations with visual coding and perception. *Vision Research*, 50(16), 1503–1509. doi:10.1016/j.visres.2010.05.002 PMID:20580643
- Green, C. D. (1995). All that glitters: A review of psychological research on the aesthetics of the golden section. *Perception*, 24(8), 937–937. doi:10.1068/p240937 PMID:8848362
- Greenfield, G. (2005). On the origins of the term computational aesthetics. In *Proceedings of the First Eurographics Conference on Computational Aesthetics in Graphics, Visualization and Imaging* (pp. 9–12).
- Hoenig, F. (2005). Defining computational aesthetics. In *Proceedings of the First Eurographics Conference on Computational Aesthetics in Graphics, Visualization and Imaging* (pp. 13–18).
- Kim, H.-S., & Cho, S.-B. (2000). Application of interactive genetic algorithm to fashion design. *Engineering Applications of Artificial Intelligence*, 13(6), 635–644. doi:10.1016/S0952-1976(00)00045-2
- Kolmogorov, A. (1965). Three approaches to the quantitative definition of information. *Problems of Information Transmission*, 2(1-4), 157–168.
- Kruse, R., Borgelt, C., Klawonn, F., Moewes, C., Steinbrecher, M., & Held, P. (2013). *Computational intelligence: a methodological introduction*. Springer Science & Business Media. doi:10.1007/978-1-4471-5013-8
- Lakshika, E., Barlow, M., & Easton, A. (2012). Fidelity and complexity of standing group conversation simulations: A framework for the evolution of multi agent systems through bootstrapping human aesthetic judgments. In *Proceedings of 2012 IEEE Congress on Evolutionary Computation (CEC)* (p. 1044-1051). doi:10.1109/CEC.2012.6256473

- Lakshika, E., Barlow, M., & Easton, A. (in press). Understanding the interplay of model complexity and fidelity in multi-agent systems via an evolutionary framework. *IEEE Transactions on Computational Intelligence and AI in Games*. doi:10.1109/TCIAIG.2016.2560882
- Leder, H., & Nadal, M. (2014). Ten years of a model of aesthetic appreciation and aesthetic judgments: The aesthetic episode—developments and challenges in empirical aesthetics. *British Journal of Psychology*, 105(4), 443–464. doi:10.1111/bjop.12084 PMID:25280118
- Li, C., & Chen, T. (2009). Aesthetic visual quality assessment of paintings. *IEEE Journal of Selected Topics in Signal Processing*, 3(2), 236–252. doi:10.1109/JSTSP.2009.2015077
- Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1), 1–167.
- Locher, P., & Nodine, C. (1989). The perceptual value of symmetry. *Computers & Mathematics with Applications (Oxford, England)*, 17(4), 475–484. doi:10.1016/0898-1221(89)90246-0
- Machado, P., Romero, J., & Manaris, B. (2008). Experiments in computational aesthetics. In *The Art of Artificial Evolution* (pp. 381–415). Springer. doi:10.1007/978-3-540-72877-1_18
- Markowsky, G. (1992). Misconceptions about the golden ratio. *The College Mathematics Journal*, 23(1), 2–19. doi:10.2307/2686193
- McCormack, J. (1993). Interactive evolution of l-system grammars for computer graphics modelling. *Complex Systems: From Biology to Computation*, 118–130.
- McManus, I. C. (1980). The aesthetics of simple figures. *British Journal of Psychology*, 71(4), 505–524. doi:10.1111/j.2044-8295.1980.tb01763.x PMID:7437674
- Moorthy, A. K., Obrador, P., & Oliver, N. (2010). Towards computational models of the visual aesthetic appeal of consumer videos. In *Proceedings of the 11th European Conference on Computer Vision: Part V* (pp. 1–14). Springer. doi:10.1007/978-3-642-15555-0_1
- Oliver, A., Monmarch, N., & Venturini, G. (2002). Interactive design of web sites with a genetic algorithm. In *Proceedings of the IADIS International Conference WWW/Internet*. (pp. 355–362).
- Osborne, H. (1986). Symmetry as an aesthetic factor. *Computers & Mathematics with Applications (Oxford, England)*, 12(1), 77–82. doi:10.1016/0898-1221(86)90140-9
- Palmer, S. E., Schloss, K. B., & Sammartino, J. (2013). Visual aesthetics and human preference. *Annual Review of Psychology*, 64(1), 77–107. doi:10.1146/annurev-psych-120710-100504 PMID:23020642
- Picard, R. W. (2003). What does it mean for a computer to “have” emotions. *Emotions in Humans and Artifacts*, 213–235.
- Reynolds, C. W. (1987). Flocks, herds and schools: A distributed behavioural model. *ACM SIGGRAPH Computer Graphics*, 21, 25–34. http://doi.acm.org/10.1145/37402.37406
- Rigau, J., Feixas, M., & Sbert, M. (2007). Conceptualizing Birkhoff’s aesthetic measure using Shannon entropy and Kolmogorov complexity. In *Computational Aesthetics* (pp. 105–112). Academic Press.
- Russell, P. A. (2000). The aesthetics of rectangle proportion: Effects of judgment scale and context. *The American Journal of Psychology*, 113(1), 27–42. Retrieved from http://www.jstor.org/stable/1423459 doi:10.2307/1423459 PMID:10742842
- Santayana, G. (1904). What is aesthetics? *The Philosophical Review*, 13(3), 320–327. doi:10.2307/2176284

- Saw, R., & Osborne, H. (1960). Aesthetics as a branch of philosophy. *The British Journal of Aesthetics*, 1(1), 6–20. doi:10.1093/bjaesthetics/ONE.1.6
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(3), 379–423. doi:10.1002/j.1538-7305.1948.tb01338.x
- Silvia, P. J., & Barona, C. M. (2009). Do people prefer curved objects? angularity, expertise, and aesthetic preference. *Empirical Studies of the Arts*, 27(1), 25–42. doi:10.2190/EM.27.1.b
- Sims, K. (1991). Artificial evolution for computer graphics. *Computer Graphics*, 25(4), 319–328. doi:10.1145/127719.122752
- Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, 14(3), 199–222. doi:10.1023/B:STCO.0000035301.49549.88
- Spehar, B., Clifford, C. W., Newell, B. R., & Taylor, R. P. (2003). Universal aesthetic of fractals. *Computers & Graphics*, 27(5), 813–820. doi:10.1016/S0097-8493(03)00154-7
- Takagi, H. (2001). Interactive evolutionary computation: Fusion of the capabilities of EC optimization and human evaluation. *Proceedings of the IEEE*, 89(9), 1275–1296. doi:10.1109/5.949485
- Tokui, N., & Iba, H. (2000). Music composition with interactive evolutionary computation. In *Proceedings of the 3rd International Conference on Generative Art (Vol. 17, pp. 215–226)*. Academic Press.
- Walton, K. (2007). Aesthetics—what? why? and wherefore? *The Journal of Aesthetics and Art Criticism*, 65(2), 147–161. doi:10.1111/j.1540-594X.2007.00246.x
- Wang, K., Bui, V., Petraki, E., & Abbass, H. A. (2012). From subjective to objective metrics for evolutionary story narration using event permutations. In *Asia-Pacific Conference on Simulated Evolution and Learning* (pp. 400–409). doi:10.1007/978-3-642-34859-4_40
- Zurek, W. H. (1989). Algorithmic randomness and physical entropy. *Physical Review A*, 40(8), 4731–4751. doi:10.1103/PhysRevA.40.4731 PMID:9902721

ADDITIONAL READING

- Bentley, P., & Corne, D. (2002). *Creative evolutionary systems*. Morgan Kaufmann.
- Galanter, P. (2012). Computational aesthetic evaluation: Past and future. In *Computers and Creativity* (pp. 255–293). Springer Berlin Heidelberg. doi:10.1007/978-3-642-31727-9_10
- Johnson, C. G. (2012). Fitness in evolutionary art and music: what has been used and what could be used? In *International Conference on Evolutionary and Biologically Inspired Music and Art* (pp. 129–140). Springer Berlin Heidelberg. doi:10.1007/978-3-642-29142-5_12
- Lakshika, E., Barlow, M., & Easton, A. (2015). Evolving High Fidelity Low Complexity Sheepdog Herding Simulations Using a Machine Learner Fitness Function Surrogate for Human Judgement. In *Australasian Joint Conference on Artificial Intelligence* (pp. 330–342). Springer International Publishing. doi:10.1007/978-3-319-26350-2_29
- Lakshika, E., Barlow, M., & Easton, A. (2016). Evolving lane merge traffic behaviour simulations via a macroscopic objective function and a machine learning system trained through bootstrapped human judgement. *Applied Intelligence*, 44(4), 862–877. doi:10.1007/s10489-015-0733-3

McCormack, J. (2013). Aesthetics, art, evolution. In *International Conference on Evolutionary and Biologically Inspired Music and Art* (pp. 1-12). Springer Berlin Heidelberg.

Prusinkiewicz, P. (1986). Graphical applications of L-systems. In *Proceedings of graphics interface* (Vol. 86, No. 1986, pp. 247-253).

Sims, K. (1994). Evolving virtual creatures. In *Proceedings of the 21st annual conference on Computer graphics and interactive techniques* (pp. 15-22). ACM.

KEY TERMS AND DEFINITIONS

Aesthetic Classification: Categorising objects, situations or phenomena into several aesthetic classes based on their quality.

Aesthetic Scoring: Rating objects, situations or phenomena for their aesthetic quality in a discrete or a continuous range.

Aesthetics: Study of human mental processes involved in appreciating a subjective experience and its quality in a wide variety of objects, situations and phenomena. An aesthetic experience can be positive, negative or more subtly nuanced.

Computational Aesthetics: An area of research focused on developing computational models and techniques that are capable of making human-like aesthetic decisions.

Computational Intelligence: The application and development of adaptive computational techniques focused on making systems that can demonstrate intelligent behaviour in complex and changing environments.

Evolutionary Art: An area of research focused on utilising evolutionary computing techniques to automatically create art.

Evolutionary Computing: A computational intelligence technique inspired by the evolution of biological species in nature and utilise the underlying abstract concepts of Darwinian evolution: natural selection and survival of the fittest to perform a stochastic search for solutions to a problem.

Dotted Raster–Stereography

Muhammad Wasim

Usman Institute of Technology, Pakistan

Fauzan Saeed

Usman Institute of Technology, Pakistan

Abdul Aziz

Usman Institute of Technology, Pakistan

Adnan Ahmed Siddiqui

Hamdard University, Pakistan

INTRODUCTION

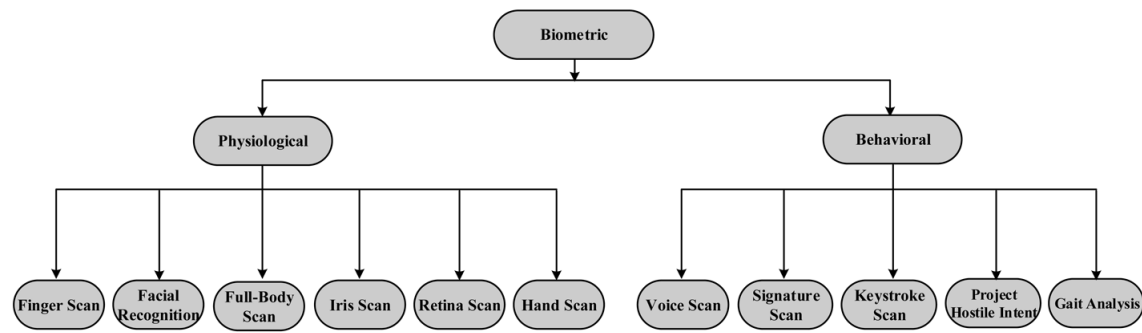
The new era has heralded several developments in technology; technological determinism has improved as well as complicated the human life. Among several challenges encountered by society security ranks as the first and foremost in present world. Every nation is facing the nuisance and security threat one way or another, from criminal activities to group initiated terrorism indiscipline is reigning supreme from Asia to Americas. In such a world providing security is of utmost significance. In this regard knowledge-based or token-based security systems are well suited to protect a system. To date several systems are implemented that use the conventional method of User ID and Password for protection. Few examples include On-line banking, E-mail system and ATM etc. the issue associated with such systems is that they can fail in advent of hacking or stolen card. Fraudulent users can easily predict short length passwords. If we encourage the user to implement long passwords then memorization becomes a daunting task. Such systems are considered as insecure therefore are less recommended. We can identify any person in several ways Signatures, Identity card, retinal recognition, voice recognition, hand geometry and face recognition techniques. A lot of work is being done in biometric technology. Through biometrics unique characteristics of a human being can be

recognized. Biometric covers physiological (direct measurement of any part of human body, finger scan, body scan, facial recognition, retina, hand scan etc) as well as behavioral (based on direct measurements derived from an action, voice scan, signature scan, gait characteristics, project hostile intent are the common examples as described in Figure 1.

Face recognition has proven itself during the past few years. It is considered as most significant application of image analysis. There might be two reasons for this trend one is its broad spectrum application in law enforcement and commercial applications and second could be the potential of this technology to play a pivotal role in near future. The concept of human face recognition using machines poses a challenge that attracts researchers from several disciplines like neural networks, computer vision, and computer graphics image processing and pattern recognition. Most face recognition systems are not excellent when it comes to reliability of identification like using cosmetic operation and face mask. All these systems have significance due to following reasons: all the biometric systems require some action on part of the user for example, user placing hand on finger recognition machine, one also has to stand firm or in anatomical position to be captured correctly by the camera for retina recognition. However, face should be picked up from a distance irrespective

DOI: 10.4018/978-1-5225-2255-3.ch015

Figure 1. Biometric system



of specific position. The algorithm can recognize face using partial facial features extracted from poorly captured photographs; this procedure is for surveillance and security projects. There are other shortcomings to biometric systems this includes damage to the epidermis tissue in case of hand and finger recognition (like bruise or any other abrasion's). Iris and retina recognition requires expensive setup and is dependent on the movement of the face. The voice recognition is vulnerable to background noise, fluctuations in phone line and tape recordings. If we talk about signatures they can be copied or forged so we need to focus on how to detect the pressure exerted on the pen while producing the signatures. In contrast to all these techniques face recognition is possible using couple of in expensive camera and in a fixed position. Efficient recognition algorithm can compensate for noise, scale and illumination. Face recognition is now a very demanding area in research, in this process the captured face is compared with known face image in the database, identification is a tedious task because of light conditions, pose, ageing factor and capture quality, this is an active field of research with focus on generating accurate and better result with fast processing time in fields like human computer interaction and biometrics (Gaidhane, Hote, & Singh, 2014).

Finally, the other technologies mentioned induces the vulnerability to germs and impurities due to the physical contact of different users with

the device, the most beautiful aspect of face recognition is that since it is stereo-photogrammetric technique it is non-destructive, non-intrusive and non-contact and hence reduces the chances of health risk substantially (Jafri & Arabnia, 2009). Such a technique can be easily incorporated by the engineers while constructing security system since their National Society of professional Engineers (NSPE) Code of ethics advocates this practice in their fundamental canon:

An Engineer should hold paramount the safety, health and welfare of the public.

BACKGROUND

As discussed face recognition approaches are one of the most attractive research areas for the researchers. A variety of face recognition techniques and methods have been discussed. This technique having no health risks and its social acceptability is high. Face recognition techniques can be categorized in three different methodologies — holistic, feature and hybrid. A holistic method is based on the global feature information of human face. Its accuracy rate is always high as this technique is built on the core features of human face. Feature based method is a structure matching technique. In this method eyes, nose and lips are considered for structural pattern matching to identified human faces. A hybrid method is a combination of holis-

Table 1.

S. No	Face Recognition Method	Description
1.	Structural Matching	In this method the face recognition is based on linear distances between key features of human face (Cox <i>et al.</i> , 1996).
2.	Principal Component Analysis (PCA)	PCA method used Eigen-picture coordinate space to demonstrate facial images as described by (Sirovich & Kirby, 1987). This technique also used to re-construct facial structures by using a small collection of Eigen-pictures.
3.	Fisher Faces	Fisher-faces used discriminatory extracted features for face representation.
4.	Independent Component Analysis (ICA)	ICA is a generalization form of PCA, but in situation of ICA the input files are de-correlated by means of second-order statistics
5.	Linear Discriminant Analysis (LDA) / Fisher Linear Discriminant (FLD)	Both LDA and FLD used global features of human faces for the purpose of identification.
6.	Feature Based Approach	This approach based on local key features of human face. The geometry described in this method is based on the locations of human eyes, mouth and nose (Manjunath, Chellappa and Malsburg, 1992).
7.	Hidden Markov Model (HMM)	HMM method based on certain statistical models, which are employed to describe statistical properties of signal.
8.	Hybrid Approach	This method combines both the local and global approaches together. For example in some described face recognition methods global information of human face combined with the local information for the purpose of face identification (Zhao, Chellappa, Phillips & Rosenfeld, 2003).

tic and feature based techniques. The technique, which is discussed in this chapter, is related with feature based method.

Face Recognition Techniques: Literature Survey

For the period last of 3-decades many face recognition methods have been advocated. It is one of the more exciting and stimulating area of research these days. During this period many techniques and methodologies have been introduced to recognized human faces. The earlier techniques were based on the structure of human faces. Relation between distances of eyes, nose and lips were the key parameters to identify human faces (Kelly, 1970). Some techniques used distances and the angles between eyes, mouth, nose, and chin top of human face (Kanade, 1973). Following is the brief description of some face recognition methods, which helps to understand this domain in a more convenient way.

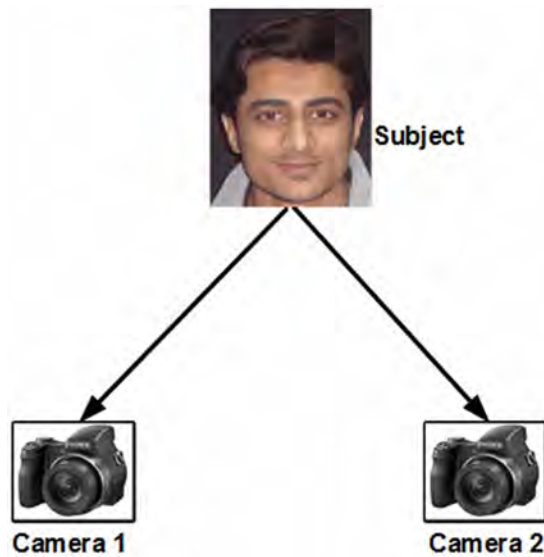
3-D IMAGING TECHNIQUES

Stereo-Photogrammetry Technique

In stereo-photogrammetry the image is captured using two cameras and the obtained results are then overlapped for analysis as given in Figure 2. It can be termed as a singular process mapping 3D object space to 2D image spaces; this concept is of dimensional reduction. Height and curvature (3D) information can be represented on a monitor or on a paper (2D surface) by using standard-light projection.

By replacing one of the cameras with raster slide being projected on human face via multimedia projector as mentioned in Figure 3, we can record the distortions and process them to acquire vital curvatures, in the route altering the passive scanning (stereophotogrammetry) to active-scanning (rasterstereography).

Figure 2. Stereophotogrammetry setup

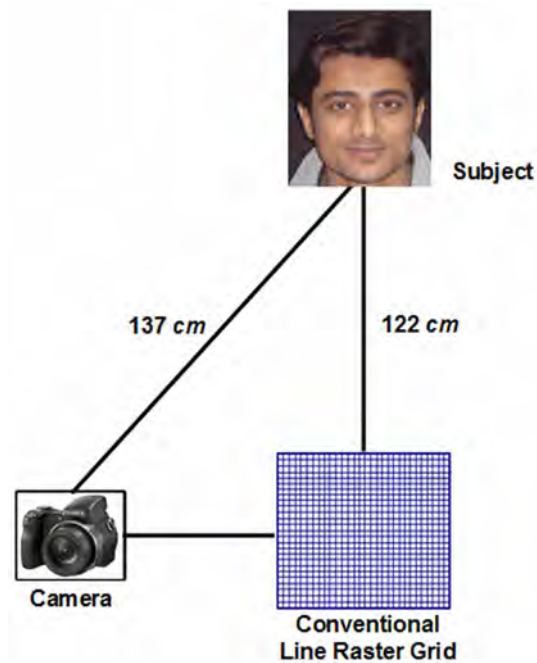


Moiré Fringe Topography Technique

Moiré fringe topography is utilized for the purpose of height-mapping, having curve like contours of constant distance from the projected screen (Kamal, 1998). Height maps produced by these contours can be epitomized as vectors that are normal to the image plane, it should be noted that the extent of all vectors are relational to the elevation coordinate at the same point. Moiré technique is applied in detection (Kamal, 1998), quantification (Kamal, 1996), documentation (Kamal, 1982) and the track of trunk irregularities in scoliosis position and gait analysis.

Ansari used a custom design wedge calibrated shadow type moiré frame (Ansari, 2008), Further on he studied the precision and accuracy of moiré-fringe-topography setup in the Laboratory. The model of precision and accuracy (Naseeruddin, Kamal, & Firdous, 1998) were changed and tested for moiré system. Height map's quantification was more than 99% precise and accurate. The formulae are defined in (Kamal, 2009) that are used to compute precision and accuracy. Body height maps can be determined using moiré technique.

Figure 3. Conventional Line raster-stereography setup



Shadow Type Moiré Technique

The fringes which we utilized for this experiment is Moiré topology; it is a wooden structure which is lightweight (Akram, 1989). We used fishing line to prepare the grid the only problem with such a structure is that the grid involved pressurizes the wooden frame and there is always a chance of damage being inflicted to the frame, to avoid such a scenario we introduce reinforcements. The method is a good way of recording information and variations the only irony is that the setup requires space.

Grid Projection Type Moiré Technique

This is the type that projects grid on the human face; as a result a changed image is formed corresponding to face shape. This image is then, superimposed on a standard to produce contour line.

The fringes are visible in real time to the human eye, in the shadow as well as the projection type. Where perpetual record in the format of a photo is not a requirement we can setup these images, the drawback is that in case of miss-positioning it cannot be averted in both the cases once the picture is recorded.

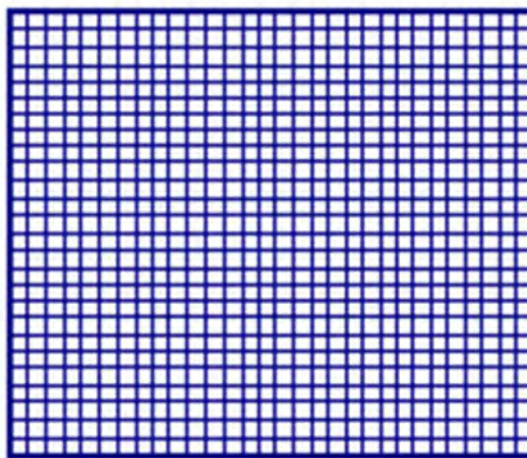
Grating Hologram Type Moiré Technique

The problem of miss-positioning during recording discussed in the previous section can be removed by using Grating hologram type moiré fringe topography, since it can adjust the moiré fringes after recording. After projecting moiré grid on the human face, the distorted grid is photographed. In order to generate moiré fringes the grid is again projected with the same magnification and angle on the photograph of the distorted grid. For gait analysis, raster recording and moiré the most appropriate type is Grating hologram.

CONVENTIONAL LINE RASTERSTEREOGRAPHY TECHNIQUE

The conventional Rasterstereography is based on a grid, which is constructed on the basis of horizontal and vertical lines as given in Figure 4. Raster stereography is chiefly, a procedure, which can define local curvatures. A camera is switched by a light-propagating system, which projects a square grid on the surface of a subject. The grid can be bred by intervallic extension of rudimentary squares in horizontal and vertical guidelines. The periodic square on projection upon the subject becomes an almost-periodic square. Distorted raster represents the topological properties of subject, conceivably revealing metric of surface. One of the beneficial aspects of the technique is that it does not need a definite prearrangement of apparatus to acquire significant raster, whereas defined geometry limits the moiré format.

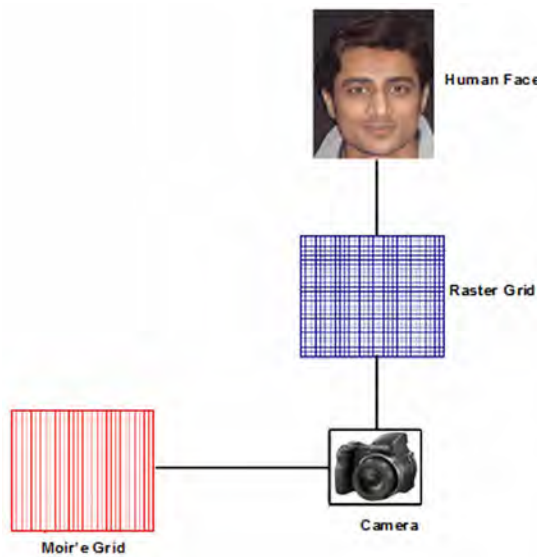
Figure 4. Conventional line raster grid



SIMULTANEOUS MOIRÉ AND RASTER RECORDING TECHNIQUE

Concurrent raster and moiré recordings can be executed using a Camera, two audiovisual aid projectors; a moiré grid that is red along with a raster grid that is blue in color are required as given in Figure 5. Both are in chorus, transformed on the face. Matching-red filter is placed, for a color photograph of face which makes the red-moiré grid obscure and the blue-raster grid give the impression of black. The grids were then analyzed using raster stereography algorithms. Now a matching-blue filter was positioned on the image. The raster grid was curbed and the moiré grid appeared black. Later, a typical moiré grid was projected with the same magnification and angle as the distorted grid, to produce moiré fringes. Triumph of this process depended, critically, on correct matching of the colors, which could be upgraded by decreasing bandwidth of the color that grid is utilizing and the matching filters consumed (Kamal, 1996). In recent times, three dimensional bone scanning was anticipated by using coinciding raster and moiré recording, joint with backscatter-X-ray expertise.

Figure 5. Layout for simultaneous moiré and raster setup



PROBLEMS ASSOCIATED WITH LINE RASTERSTEREOGRAPHY

At the initial stages, conventional raster grid, based on stacked-line squares were used to find the curvatures of the surface of a subject. There are many applications in which conventional rasterstereography were used and all the applications were related to the medical domain. Some very common and popular applications of rasterstereography technique are to find spinal deformity in human body (Zuberi, 2002), to analyze back shape of human body (Hackenberg, Hierholzer, Pötzl, Götze, & Liljenqvist, 2003) and to detect trunk deformity in human body and to find the reproducibility of the technique (Mohokum, Mendoza, Wolf, Sitter, Jürgen, & Skwara, 2010). In all these old applications it was very difficult to find the valid curvatures patterns of human body without breaking the lines. The concept of raster-stereography for human face recognition was introduced in 2008 (Kamal, 2008), which was implemented by the first author of this chapter in 2013 (Wasim, 2013). In the experimental setup a conventional raster-grid was projected on human

frontal face. As human face is a curved surface, raster grid were distorted. This distorted grid provided the curvature information of human face as given in Figure 6. For this testing purpose dark black and white line grids were used to record the facial curvatures of human face. As human face having more curvatures as compare to body so during test runs it was observed that a problem of contrast of grid and the skin color of human face (highly curved surface) has generated, especially in gray scale images. Because of this problem of poor contrast, the curved line breaks during extraction using MATLAB code. Hence it was not easy to trace valid line to define curvature information of human face.

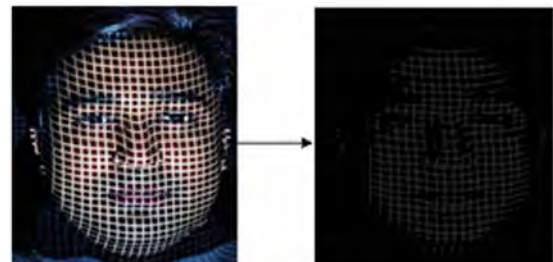
Two algorithms were constructed and used to extract valid curvature patterns of human faces:

1. Valid Curve Detection Algorithm (VCDA).
2. Line Traversal Algorithm (LTA).

Valid Curve Detection Algorithm (VCDA)

To set the criteria of valid line the facial image first cropped by an order of 10×13 based on co-ordinate values of cross over points of horizontal and vertical lines. This covered the core features (eyes, nose and lips area) of all recorded human faces. A line treated as valid if it contained ten consecutive coordinate values in a row which

Figure 6. Un-wanted breaks due to grid and skin contrast



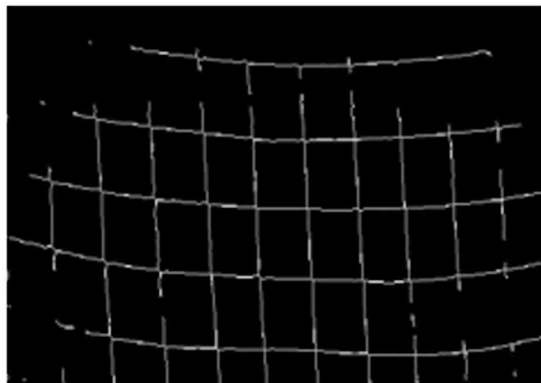
obeyed the criteria of valid line detection. Following steps were followed to detect valid line.

- Step 1:** Filtering the anomalies from the captured image
- Step 2:** Check curve validity — ten consecutive cross over points in a row without breaking the line
- Step 3:** Find valid curve in cropped image by pattern matching of valid criteria
- Step 4:** Find the coordinate values of each cross over points in term of (x, y)
- Step 5:** Apply mathematical model and solved decision parameters
- Step 6:** Used these parameters for the purpose of face identification

Problem Encountered

A problem was reported during the valid curve detection. According to the set criteria in VCDA if ten consecutive white pixels were correctly detected without breaking of line then it should be considered as a valid curve. This experimental value of ten consecutive white pixels was set through trial-and-error to run the algorithm onto 400 images. Due to un-wanted breaks in curve the algorithm could not pick all the valid points from the captured image. Figure 7 showed the small breaks in curves, which created huge problem to detect valid curve pattern of human face.

Figure 7. Un-wanted breaks in curve



Line Traversal Algorithm (LTA)

Working of LTA was based on the execution of VCDA. Traversal of each detected line was started during the execution of VCDA. White pixels were followed and turned into gray to show traversal until connection broke as given in Figure 8. If a line was traversed successfully, cross over points were easily detected and coordinate values of each point were found and recorded in to system database. However, the algorithm did not perform accurately, when there were breaks in line resulting in white regions generated. So it was concluded that if both the algorithms executed successfully valid curvature patterns were recorded.

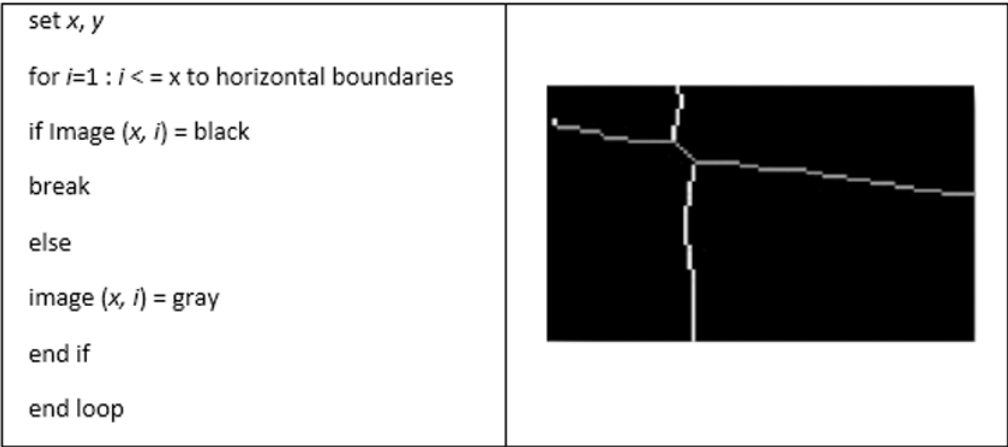
DOTTED RASTERSTEREOGRAPHY TECHNIQUE

Due to the issues related to VCDA and LTA in conventional Rasterstereography technique it was difficult to analyze human facial curvature patterns for the purpose of face identification. A modified raster-stereography concept was introduced. In this technique the raster grid consisted of dots, instead of conventional horizontal and vertical lines are used as given in Figure 9.

This new designed dotted grid minimized the noise effect and coordinates of dots easily extracted. The issues related to conventional raster grid were resolved and valid curved patterns of human face were extracted in a more proper and accurate manner. Set up Specifications of setup are given in Figure 10.

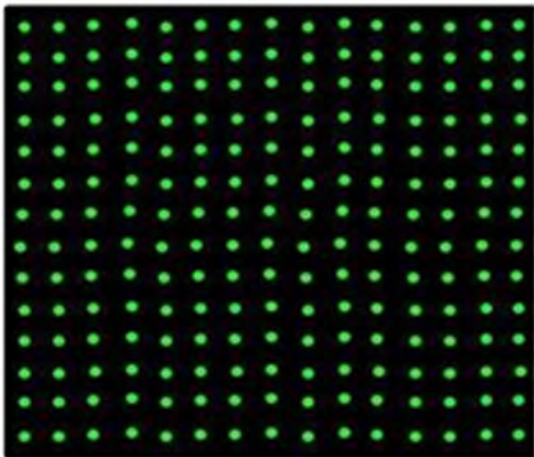
The efficiency of dotted raster to detect all crossover points was increased by reducing the processing time as compare to conventional raster-stereography technique. Dotted raster-stereography technique minimized code complexity and improves efficiency of the system. This modification made the algorithm simple and easy. Pixel and coordinate values of each point were extracted using MATLAB code.

Figure 8. Working of LTA



A

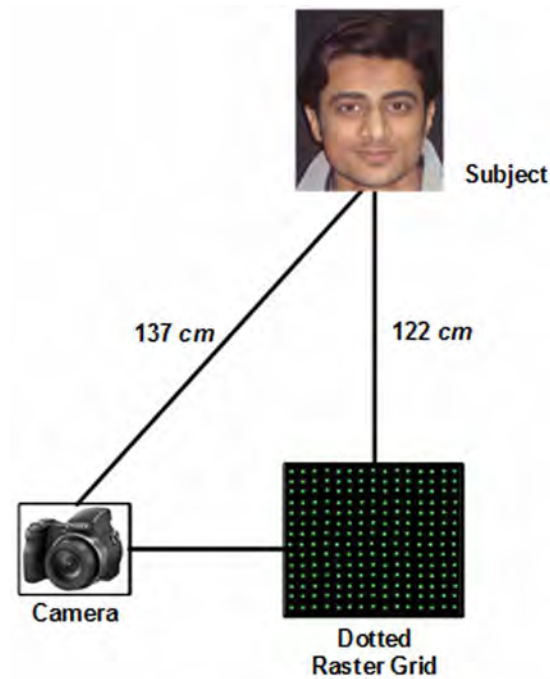
Figure 9. Dotted raster grid



DOTTED RASTER-STEREOGRAPHY TECHNIQUE IN FACE IDENTIFICATION

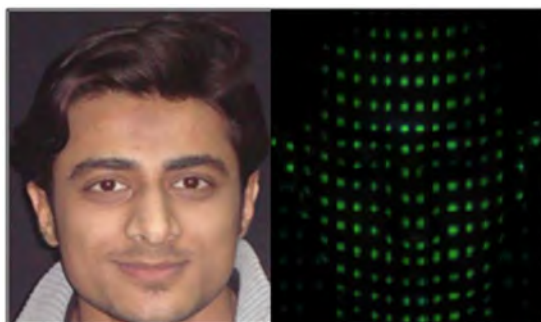
The dotted-raster-stereography setup was established in “Image Processing Research Lab (IPRL)” at Usman Institute of Technology. The set up consist of a multimedia projector to project the dotted raster grid on human face, a dotted-raster grid (with 0.6 mm inter-dot spacing) contains white dots with black background and an advanced-digital camera to capture the human face and

Figure 10. Dotted raster-stereography setup



distorted grid pattern on human face. The grid consist of dots was projected using multimedia projector on human face. Grid was distorted and was recorded using digital camera and stored in the system database for further processing. The concept is given in Figure 11.

Figure 11. Dotted raster grid projection on human face



PHASES OF FACE-RECOGNITION SYSTEM USING DOTTED GRID

The dotted raster-stereography based face recognition system is divided into three different phases called registration phase, identification phase and verification phase (Chan, 2008). A person face was first register in to system database. At later stage of the system, the register face identified based on calculated decision parameters. In the final phase of the system human face was verified based on exact match of decision parameters from the system database (Attar & Atani, 2012).

Registration

Registration is the first phase of the dotted raster-stereography based face recognition system. In

this phase of the system a human face first capture using a digital camera and stored in the system database for the purpose of verification. The dotted grid projected using multimedia projector on the human face. The grid was distorted as human face having curved surface. These distortions in grid provide the facial curvature pattern of human face. Using mathematical model decision parameters were solved on the basis of coordinate values of distorted points. Each image was assigned a specific ID for authentication purpose. Featured template of the subject was constructed on the basis of two decision parameters mean (M) and gaussian (G) curvatures. Figure 12 showed the working of registration phase.

Identification

This is second phase of the system. In this module a facial featured template of human face was compared with all available templates from the system database to find the best match. If match found that person face basis of facial featured template will be identified. So it is important that before identification phase each person face should be registered in the system database for the purpose of identification. The identification phase needs 1-t-mano comparison to find exact match from the system database (Chan, 2008). Figure 13 showed the working of identification phase.

Figure 12. Registration phase

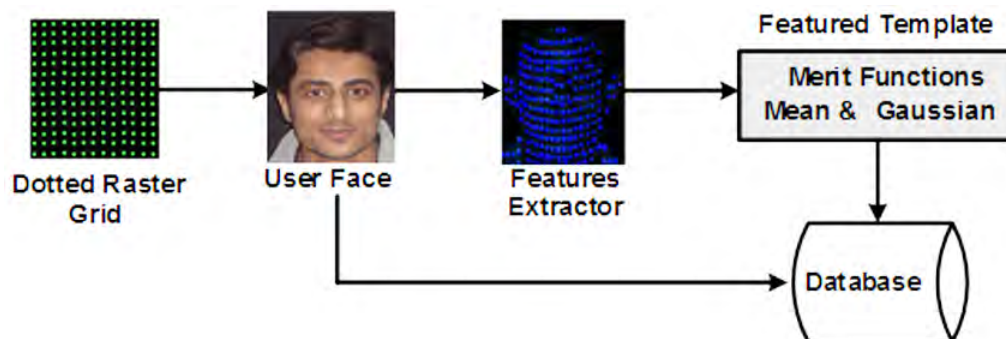
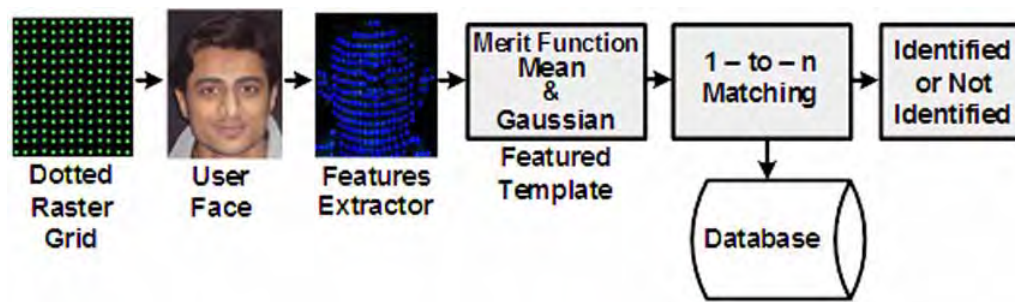


Figure 13. Identification phase



Verification

Verification phase is the third phase of the system. It is an optional phase and only used when featured template values of two different faces are much closed and it is difficult to identify true face from the system database. So for the purpose of verification there should be a secondary system connected with the main application just to verify and recognized true face. This secondary system may be a ID card number matching system, thumb, iris, retina or voice pattern recognition system. Figure 14 showed the working of proposed verification phase.

RESULTS AND DISCUSSIONS

Figure 15, talk about the five initial sample human faces from the system database. These cases having the reference ID as: IPRL-UIT-2016061-01, IPRL-UIT-2016061-02, IPRL-UIT-2016061-03, IPRL-UIT-2016061-04 and IPRL-UIT-2016061-05. The decision parameters mean (M) and gaussian (G) are different with different faces depend upon the facial curvatures of face. These parameters are the main source of face identification. The dotted raster-stereography technique tested on 800 different human faces (both male and female, age range 16-20 years and an accuracy of 95% were obtained.

Figure 14. Verification phase

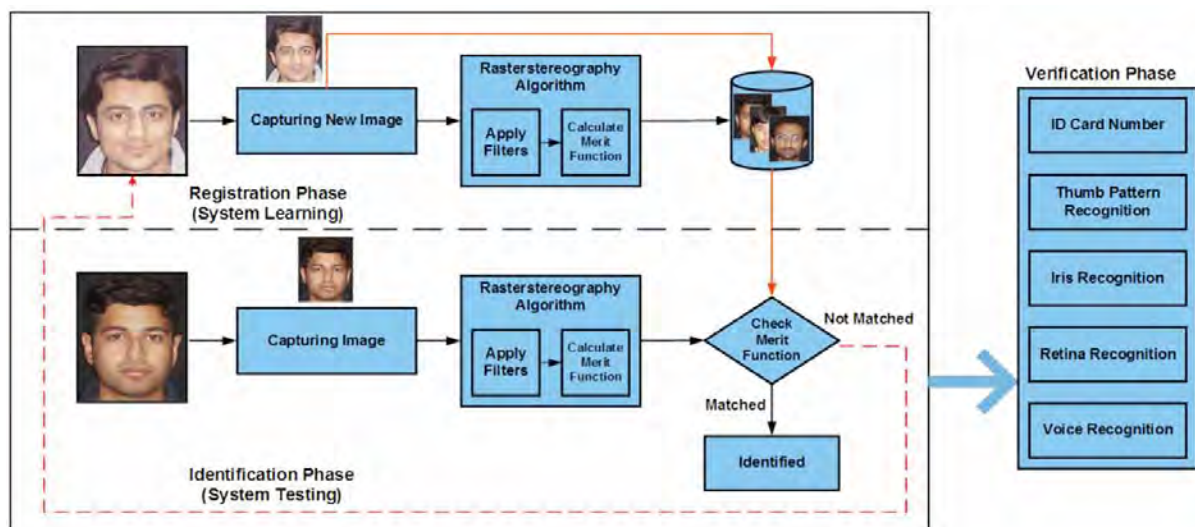


Figure 15. Results of sample faces from the system database

S. No	Face ID	Sample Faces	Rasterstereography Technique	
			Mean (M) cm^{-1}	Gaussian (G) cm^{-1}
1	IPRL-UIT-2016061-01		12.0200	24.1200
2	IPRL-UIT-2016061-02		5.1145	13.4530
3	IPRL-UIT-2016061-03		13.0500	33.9976
4	IPRL-UIT-2016061-04		18.4523	41.5641
5	IPRL-UIT-2016061-05		19.9876	52.1110

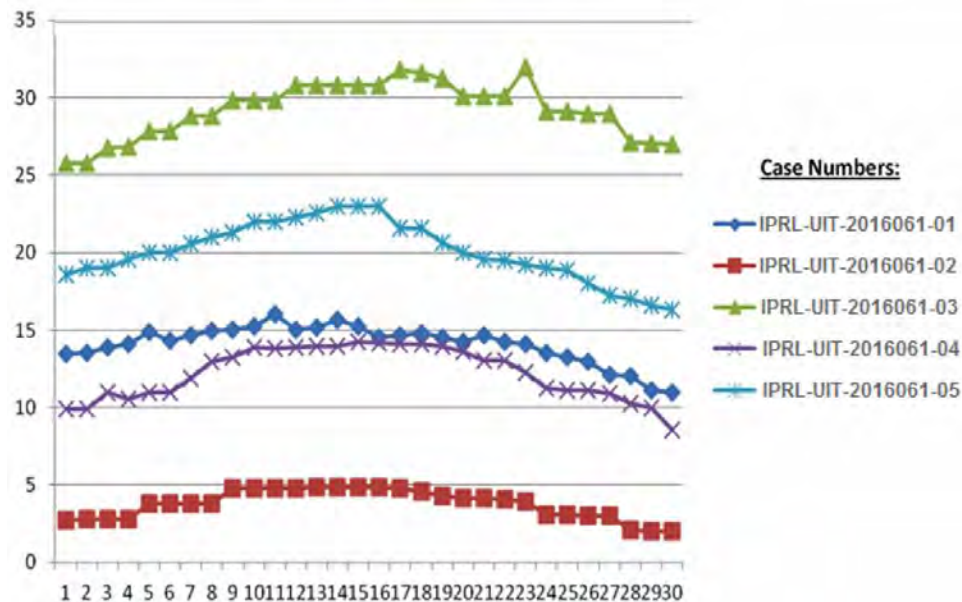
To obtained best results a registered face tested five times in Image Processing Research Lab. An average value of each individual is reported in Figure 15. The range of five repeated M and G values are discussed as follows:

- The absolute mean and gaussian curvatures of facial image IPRL-UIT-2016061-01 is found to be as $11.9900\text{ }cm^{-1} \leq M \leq 12.09\text{ }cm^{-1}$ and $24.1100\text{ }cm^{-1} \leq G \leq 24.1411\text{ }cm^{-1}$. An average value of mean and gaussian of same face is defined as $12.0200\text{ }cm^{-1}$ and $24.1200\text{ }cm^{-1}$.

- The absolute mean and gaussian curvatures of facial image IPRL-UIT-2016061-02 is found to be as $5.1010\text{ }cm^{-1} \leq M \leq 5.1600\text{ }cm^{-1}$ and $13.3701\text{ }cm^{-1} \leq G \leq 13.4711\text{ }cm^{-1}$. An average value of mean and gaussian of same face is defined as $5.1145\text{ }cm^{-1}$ and $13.4530\text{ }cm^{-1}$.
- The absolute mean and gaussian curvatures of facial image IPRL-UIT-2016061-03 is found to be as $12.9901\text{ }cm^{-1} \leq M \leq 13.0900\text{ }cm^{-1}$ and $33.9802\text{ }cm^{-1} \leq G \leq 33.9999\text{ }cm^{-1}$. An average value of mean and gaussian of same face is defined as $13.0500\text{ }cm^{-1}$ and $33.9976\text{ }cm^{-1}$.
- The absolute mean and gaussian curvatures of facial image IPRL-UIT-2016061-04 is found to be as $18.4411\text{ }cm^{-1} \leq M \leq 18.4610\text{ }cm^{-1}$ and $41.4501\text{ }cm^{-1} \leq G \leq 41.5800\text{ }cm^{-1}$. An average value of mean and gaussian of same face is defined as $18.4523\text{ }cm^{-1}$ and $41.5641\text{ }cm^{-1}$.
- The absolute mean and gaussian curvatures of facial image IPRL-UIT-2016061-05 is found to be as $19.9666\text{ }cm^{-1} \leq M \leq 20.010\text{ }cm^{-1}$ and $52.1000\text{ }cm^{-1} \leq G \leq 52.1600\text{ }cm^{-1}$. An average value of mean and gaussian of same face is defined as $19.9876\text{ }cm^{-1}$ and $52.1110\text{ }cm^{-1}$.

The system identified different faces on the basis of mean and gaussian curvatures. This system can be deployed for the purpose of security based on human face identification. Therefore it is the requirement of the system that all the users' should register first in the system database along with their associated mean and gaussian curvature patterns. Graph in Figure 16 shows the behavior of basic facial curvature patterns of first five sample faces from the system database.

Figure 16. Graphical representations of basic curvatures of sample faces



FUTURE RESEARCH DIRECTIONS

1. By experiments it is now very clear that the dotted raster-stereography method generated better results with more accuracy as compare to conventional raster method. This method now can solve all previous applications, which are based on line grid with less complexity.
2. A circular dotted raster method can be used as moiré fringe topography (MFT) and this can be used in depth analysis and pattern recognition of objects.

CONCLUSION

Dotted raster-stereography has been resolved the issue of un-wanted breaks in curvature patterns, which were reported in conventional raster-stereography technique. It also resolved the issues, related to conventional raster-stereography even when it applied to human body (lesser curved surface). The results of dotted raster-stereography technique are better as compare to conventional

raster technique. It is simple, accurate and easy to extract all the curvature pattern dots without any missing points as it was the main issue with previous technique. So it is now recommended that all previous research work, which are based on conventional raster-stereography technique can be developed using dotted raster-stereography technique to get better results in a more convenient way with high accuracy as well.

REFERENCES

- Akram, M. (1989). *Construction of a shadow type moiré apparatus (M. Sc. Project)*. University of Karachi.
- Ansari, S. A. (2008). *Validity of various techniques of moiré fringe topography (M. Phil. Thesis)*. Department of Mathematics, University of Karachi, Karachi, Pakistan
- Attar, A., & Atani, R. E. (2012). A robust face recognition method using edge-based features. *Proceedings of IEEE symposium on computers and informatics*, 185-188

- Chan, C. H. (2008). *Multi-scale local binary pattern histogram for face recognition* (PhD dissertation). University of Surrey, Surrey, UK.
- Cox, I. J., Ghosn, J., & Yianilos, P. N. (1996). Feature-based face recognition using mixture distance. In *Proceedings, IEEE Conference on Computer Vision and Pattern Recognition*, 209-216. doi:10.1109/CVPR.1996.517076
- Gaidhane, V. H., Hote, Y. V., & Singh, V. (2014). An efficient approach for face recognition based on common eigenvalues. *Pattern Recognition (Elsevier)*, 47(5), 1869–1879. doi:10.1016/j.patcog.2013.11.027
- Hackenburg, L., Hierholzer, E., Pötl, W., Götze, C., & Liljenqvist, U. (2003). Rasterstereographic back shape analysis in idiopathic scoliosis after anterior correction and fusion. *Clinical Biomechanics, Elsevier*, 18(1), 1–8. doi:10.1016/S0268-0033(02)00165-1 PMID:12527240
- Jafri, R., & Arabnia, H. R. (2009). A survey of face recognition techniques. *Journal of Information Processing System*, 5(2), 41–68. doi:10.3745/JIPS.2009.5.2.041
- Kamal, S. A. (1982). Measurement of angle of spinal curvature by moiré topographs. *Journal of Islamic Medical Association (USA)*, 14, 145–149.
- Kamal, S. A. (1996). A 3-D-static model of the human spinal column. *Karachi University Journal of Science*, 24, 29-34. Retrieved from <http://www.ngds-ku.org/Papers/J18.pdf>
- Kamal, S. A. (1998). 3-D-optical imaging and image processing (biomedical applications). In J. R. Laghari, A. A. Naqvi, A. Q. Rajput, N. A. Sangi, & M. A. Shah (Eds.), *Proceedings of the International Workshop on Recent Advances in Computer Vision* (pp. 86–95). Karachi, Pakistan: SZABIST.
- Kamal, S. A. (2008). Pattern recognition using moiré fringe topography and rasterstereography. *Proceedings of the international symposium on biometrics and security technologies (IEEE IS-BAST)*. Bahria University. Available at: <http://www.ngds-ku.org/pub/confabst.htm#C68>
- Kamal, S. A. (2009). *Mathematics of experimentation*. National Conference on Physics and the World of Today in the Memory of Prof. Dr. Muhammed Rafi and Prof. Dr. Muhammed Razi Hussain, Department of Physics, University of Karachi, Karachi, Pakistan.
- Kanade, T. (1973). Computer recognition of human faces. Birkhauser.
- Kelly, M. D. (1970). *Visual identification of people by computer. Tech. rep. AI - 130*. Stanford, CA: Stanford AI Project.
- Manjunath, B. S., Chellappa, R., & von der Malsburg, C. (1992). A feature based approach to face recognition. *Conference on Computer Vision and Pattern Recognition, Proceeding CVPR'92*. IEEE Computer Society.
- Mohokum, M., Mendoza, S., Wolf, U., Sitter, H., Jürgen, P., & Skwara, A. (2010). Reproducibility of rasterstereography for kyphotic and lordotic angles, trunk length and trunk inclination: A reliability study. *Spine*, 35(14), 1353–1358. doi:10.1097/BRS.0b013e3181cbc157 PMID:20505568
- Naseeruddin, S. A., & Firdous. (1998). Accuracy and precision of moiré fringe topography system. *Journal of Natural Sciences and Mathematics*, 38, 157-166. Retrieved from <http://www.ngds-ku.org/Papers/J21.pdf>
- Sirovich, L., & Kirby, M. (1987). Low - dimensional Procedure for the Characterization of Human Faces. *Journal of the Optical Society of America A: Optics, Image Science, and Vision*, 4, 519-524.

Wasim M., Kamal S. A. & Shaikh A., (2013). A security system employing edge-based rasterstereography. *International Journal of Biology and Biotechnology*, 10(4), 483-501.

Zhao, W., Chellappa, R., Phillips, P. J., & Rosenfeld, A. (2003, December). Face Recognition: A Literature Survey. *ACM Computing Surveys*, 35(4), 399–458. doi:10.1145/954339.954342

Zubairi, J. A. (2002). Applications of computer-aided rasterstereography in spinal deformity detection. *Image and Vision Computing*, 20(4), 319–324. doi:10.1016/S0262-8856(02)00026-4

KEY TERMS AND DEFINITIONS

A

Conventional Raster-Stereography: This technique used line raster grid to capture facial curvatures of human face.

Dotted Raster-Stereography: This technique used dotted raster grid to capture facial curvatures of human face.

Moiré Technique: It is also a 3-D surface screening technique, which is based on interference of light.

Non-Invasive: Term indicate to safe procedure and no harm during test runs.

Raster-Stereography: It is 3-D surface screening technique.

Stereo-Photogrammetry: Technique, which is constructed on the difference of two images, captured with two different cameras.

Hybrid Computational Intelligence and the Basic Concepts and Recent Advances

Georgios Dounias

University of the Aegean, Greece

INTRODUCTION

In this paper computational intelligence and its major methodologies are introduced in the first place, and then hybrid intelligent systems are defined and the most popular hybrid intelligent approaches are discussed. The increased popularity of hybrid intelligent systems during the last decade, is the result of the extensive success of these systems in a wide range of real-world complex problems, but also has to do with the increased capabilities of computational technology. One of the reasons for this success has to do with the synergy derived by the computational intelligent components, such as machine learning, fuzzy logic, neural networks, genetic algorithms, or other intelligent algorithms and techniques. Each of the partial methodologies provides hybrid systems with complementary reasoning and searching methods that allow the use of domain knowledge and empirical data to solve complex problems. The paper includes recent advances and new findings in the area of hybrid computational intelligence.

BACKGROUND

Computational Intelligence

More than half a century of tentative research has rolled since the term Artificial Intelligence was coined within the members of the computer society, in an attempt to characterize the existence of intelligence nuggets in machine performance and behavior, as well as in algorithmic design and performance. Computational Intelligence represents

the evolution of a part of Artificial Intelligence during the 90's, mostly related to well-established and popular computational techniques, such as neural computation, evolutionary computation, machine learning and fuzzy logic.

According to (Duch, 2007), a brief survey of the scope of CI journals and books including the term computational intelligence in their title shows that at present, it is an umbrella for three core technologies (neural, fuzzy and evolutionary), their applications, and selected fashionable methods.

The IEEE Computational Intelligence Society defines its subjects of interest as neural networks, fuzzy systems and evolutionary computation, including swarm intelligence.

In addition, Professor James Bezdek in (Bezdek, 1994) defined a computationally intelligent system as follows:

A system is computationally intelligent when it deals with only numerical (low level) data, has pattern recognition components, does not use knowledge in the AI sense; and additionally when it (begins to) exhibit (i) computational adaptivity, (ii) computational fault tolerance, (iii) speed approaching human-like turnaround and (iv) error rates that approximate human performance.¹

In the MIT Encyclopaedia of the Cognitive Sciences, the neutral term of computational intelligence is used to gather two complementary views of AI, one as an engineering discipline concerned with the creation of intelligent machines, the other as an empirical science concerned with the computational modelling of human intelligence.

DOI: 10.4018/978-1-5225-2255-3.ch016

According to (Abbot et al, 2002), “either natural systems like brains, immune systems, ecologies, societies or, artificial systems like parallel and distributed computing systems, artificial intelligence systems, artificial neural networks, evolutionary programs, are characterised by apparently complex behaviours that emerge as a result of often non-linear spatio-temporal interactions among a large number of component systems at different levels of organization”.

Consequently, researchers in a number of distinct areas including computer science, artificial intelligence, neural networks, cognitive science, computational economics, mathematics, optimization, complexity theory, control systems, biology, neuroscience, psychology, engineering, etc., have begun to address, through a combination of basic as well as applied techniques, theoretical as well as experimental research, analysis and synthesis of such systems. Intelligent systems could be considered those systems that have the properties of self-maintenance, adaptivity, information preservation, and increase in complexity, but use other means to achieve such objectives.

Computational intelligence tools and techniques have now become methods of choice for a number of complex domains of application, due to their specific characteristics. The area mainly deals with problems related to classification, clustering, optimization and forecasting. For some of the above problem types, computational intelligence could be considered a complementary toolbox to standard Operational Research (OR) methods and techniques for optimization, problem solving and decision-making (i.e., mathematical programming, simulation, probabilistic reasoning, etc.). Some areas of computational intelligence have become particularly known over time, due to their effectiveness in facing specific real-world problems. A very thorough analysis of what is meant by computational intelligence and what the trends of modern AI are can be found in (Nilsson 1998; Chen 2000). Newer publications around the fundamentals of Computational Intelligence as well as related to recent methodological advances and

successful applications have appeared in literature from (Siddique & Adeli, 2013; Kruse et al, 2013; Sztandera 2014; Kacprzyk & Redrycz 2015).

From now on let's consider as main components of computational intelligence, the four main streams of research that dominate the area of AI during the last two decades, namely, (1) fuzzy sets and soft computing, (2) neural networks, (3) genetic algorithms and evolutionary computation (including also nature inspired intelligence) and (4) machine learning² and data mining.

A collection of research work on computational intelligence and learning techniques in the sense presented above can also be found in (Zimmermann et al., 2001; Negoita et al., 2005; Kruse et al., 2013). A reference to the basic concepts of the most popular intelligent components of hybrid intelligent architectures follows.

Fuzzy Logic

Fuzzy logic (Zadeh 1965; Zadeh, 2001), in fact can be seen as a language with syntax and local semantics where we can imprint any qualitative knowledge about the problem to be solved, usually with the assistance of a field expert. The strong point of fuzzy logic is the robustness of its interpolative reasoning mechanism. The impact of Fuzzy Logic has been extremely high during the past years with the appearance of several patents³ awarded all over the world, e.g.

- Number of fuzzy-logic-related patents issued and applied in Japan: 7,149.
- Number of fuzzy-logic-related patents issued in the United States: 21,878.
- Number of fuzzy-logic-related patents applied in the United States: 22,272.
- Fuzzy-logic-related patents issued and applied in WIPO (International): 50,999.
- Number of fuzzy-logic-related patents issued and applied in China: 25,454 (not including all the years, and not including Taiwan or Hong Kong).

- Number of fuzzy-logic-related patents issued and applied in EPO (EU): 3,268 (not including individual countries in Europe).
- Number of fuzzy-logic-related patents issued and applied in Australia: 2,350.
- Number of fuzzy-logic-related patents issued and applied in Canada: 556.

Although fuzzy logic is now a well-established domain, recent editions can be still found in fuzzy logic and its applications like (Singhai, 2013; Trillas & Eciolata, 2015) and also (Jantzen, 2013) in fuzzy control foundations and (Tamir et al., 2015) performing a flashback in the 50 years anniversary of fuzzy logic.

Machine Learning and Data Mining

Machine Learning (Michalski et al 1983; Michalski et al 1986; Mitchell 1997), was conceived four decades ago for the development of computational methods that could implement various forms of learning, in particular mechanisms capable of inducing knowledge from examples or data (Kubat et al., 1997, p.3). Knowledge induction seems particularly desirable -especially from field experts who want to understand the outcome of computational approaches - in problems that lack algorithmic solutions, are ill-defined, or only informally stated. Research in machine learning has been mainly devoted to developing effective methods for building learning (rule-based) systems that will acquire high-level concepts and/or problem solving strategies through examples in a way analogical to human learning. Recent advances in machine learning include implementations in new programming environments and languages like R and Python (Lantz 2013; Grus 2015; Raschka 2015), predictive modelling and data analytics (Flach 2012; Kelleher et al, 2015), future attempts for automated knowledge discovery (Domingos, 2015) and useful tips for developers and technical professionals when they dig deep into data performing one of the modern trends in machine learning research, as that of deep data mining (Bell, 2014).

Evolutionary Computation and Nature Inspired Intelligence

Genetic algorithms (Holland, 1975) provide a way to perform randomized global search in a solution space. Usually a population of candidate solutions, encoded internally as chromosomes, is evaluated by a fitness function in terms of its accuracy. The best chromosomes are combined and reproduced in subsequent generations. Genetic programming, proposed by (Koza, 1992) is an extension to the original concept of genetic algorithms. The population in genetic programming is composed by variable length tree-like candidate solutions. Each of these individual candidates, called program, may have functional nodes, enabling the solution to perform arbitrarily large actions. The idea of evolutionary computation, taken from Darwinian evolution principles, seems to be very promising, as it is the main way of evolution observed also in nature. Similar to the logic embodied in genetic algorithms and genetic programming, is the so-called nature inspired intelligence cluster of algorithms and methodologies. These methodologies derived from natural living and evolving systems such as ant societies, bee colonies, bird flocks and fish schools, are also evolving over time, and prove particularly effective in finding optimal solutions in hard optimization problems by intelligently exploring the search space. See (NISIS, 2005) for detailed definitions and uses of various nature inspired smart information systems implemented and presented in recent literature.

Recently Simon (2013) and Eiben and Smith (2015) have published books with introduction and fundamentals of evolutionary computing. One of the modern trends in the area is massively parallel evolutionary computation (Tsutsui & Collet 2013), applications of evolutionary computation in gene regulatory networks (Iba & Noman 2016) and in mobile networking applications such as mobile ad hoc networks (MANETs), vehicular networks (VANETs), sensor networks (SNs), and hybrid networks (Dorrnsoro et al., 2014). Similarly advances in nature inspired intelligent systems can be found in (Yang, 2014) including

PSO, ACO, BCO, simulated annealing, cuckoo search, the firefly algorithm, the bat algorithm, the flower algorithm, harmony search, as well as hybrid methods containing NII components. In addition, in (Brownlee, 2012) programming tips are given for nature inspired techniques taken by biological and natural systems, such as the adaptive capabilities of genetic evolution and the acquired immune system, and the foraging behaviors of birds, bees, ants and bacteria.

Neural Networks

Neural networks were introduced by (Rosenblatt 1959) and (Widrow & Hoff, 1960) as computational structures that can be trained to learn by examples. Using a supervised learning algorithm, such as the back-propagation (Werbos, 1974), and a training set that samples the relation between input and output, we can perform fine local optimization. They are usually called black-box architectures as their outcome often remains unexplained in terms of comprehensibility, but they are extremely strong generalization mechanisms over data collections. Recent advances in neural networks include mostly applications in financial forecasting and trading strategies (Roghani 2015) and also in programming tips for new or experienced users (Hagan 2014), (Shaffer 2015).

HYBRID INTELLIGENT SYSTEMS

Key-Terms and Definitions

Usually complex domain problems are ill-defined, difficult to model and demand large solution spaces to be explored. Any relevant information about these problems is the prior domain knowledge, usually incomplete, and input-output instances of the system's behaviour, which is also incomplete. Therefore, in many cases, hybrid combinations seem proper to handle approximate reasoning for these domains, and thus often prove superior in performance comparison to their single compu-

tational intelligent components, thus providing researchers and data analysts with advanced problem solving tools.

Hybrid computational intelligence is defined as any effective combination of intelligent techniques that performs superior or, in a competitive way to simple standard intelligent techniques. Hybrid intelligence was in fact attempted in several scientific papers during the 90s, as an extension to the standard experimentation with most of the well-known intelligent techniques, in various application domains.

Hybrid intelligent systems denote software systems which employ in parallel a combination of methods and techniques from artificial intelligence subfields such as neuro-fuzzy systems, fuzzy expert systems, evolutionary neural networks, genetic fuzzy systems, or other combination and generally combinations of main CI techniques, i.e. neural, fuzzy, evolutionary, and machine learning techniques and algorithms.

In IJHIS⁴ the interest is in research reports “that involve the use of two or more intelligent techniques and approaches, such as neural networks, traditional knowledge-based methods, fuzzy techniques, genetic algorithms, agent-based techniques, case based reasoning etc. The combination or integration of more distinct methodologies can be done in any form, either by a modular integration of two or more intelligent methodologies, which maintains the identity of each methodology or, by fusing one methodology into another, or by transforming the knowledge representation in one methodology into another form of representation characteristic to another methodology”.

Among the topics of interest of the journal are included integrations of Neural Networks with Expert Systems, integrations of Neural Networks with Fuzzy Systems, integrations of Neural Networks with Global Optimization Algorithms knowledge-based neural networks, fuzzy Logic and Expert Systems, integration of Fuzzy Systems with Global Optimization Algorithms, hybrid Optimization Techniques (global, local search techniques etc.), hybrid Systems using

Neural Networks, Evolutionary Algorithms and Fuzzy Systems, integration of Intelligent/Computational Techniques with Information Systems for Data Mining, Knowledge Management, Decision Support, hybridization of Soft Computing with other Machine Learning Techniques (such as Support Vector Machines, Rough Sets, Bayesian Networks, Probabilistic Reasoning, Statistical Learning), Hybrid models using Case-Based Reasoning, Inductive Logic Programming, Grammatical Inference and Hybrid Intelligent Systems Applications to various domains such as Image and Signal Processing, Control and Automation, Industrial and Medical Diagnosis, Bioinformatics, Data Mining, Business Information Systems, E-commerce, E-learning, Intelligent Tutoring Systems, Web Mining, Decision Support Systems, Text Mining, Natural Language Processing, etc.

One of the first attempts to organize events related to hybrid intelligence was the special issue of IEEE Transactions on Neural Networks and Hybrid Intelligent Models (Lee Giles et al., 1998). Later was organized a series of related annual world conferences initiated in Australia (HIS, 2001), lasting till now (HIS 2013; HIS 2014). In addition, since 2006 another similar event was formed in Salamanca, Spain (HAIS 2006) lasting also until nowadays (HAIS, 2013; HAIS, 2016) with long success and interesting topics. These conferences are mostly dealing with applications of hybrid approaches in real world complex problems, involving imprecision, uncertainty, vagueness and high-dimensionality. Also Ling et al. (2014) have announced a special issue in hybrid intelligent methods for health technologies.

Hybrid Intelligence, has also been one of the central topics of attention and focus for the European Network of Excellence called EUNITE, see (EUNITE, 2001) and NISIS, Nature Inspired Smart Information Systems, see (NISIS, 2005). In EUNITE (2001), it was stated that “intelligent hybrid systems” are meant to be any combinations of intelligent technologies (e.g. neuro-fuzzy approaches, evolutionary optimized networks, etc.) but particularly those, which prove to have

an obvious advantage in their performance when applied in complex domains of application (either by means of accuracy obtained, or by means of comprehensibility of the acquired results). An obvious need has been identified for guidelines regarding design, testing and assessment, as well as a need for improved understanding of the fundamental nature of engineering systems with embedded hybrid intelligence.

Hybrid Intelligence Modelling

It should be clear that hybrid intelligence should be used only in cases where its performance proves superior to simple intelligent approaches, in terms of accuracy of the model produced, in tasks like classification, discrimination, clustering, etc., and also in cases where it increases comprehensibility of the resulting model, i.e. production of meaningful rules, relations, or solution representations. In this sense, prior to the construction of a hybrid intelligence scheme, the analyst should attempt to approach and model his problem with the existing standard approaches.

The partial components of a hybrid intelligent system should be able to collaborate efficiently between them. Another important factor of a hybrid system is the speed of process and the time needed to produce a generalized high-performance decision model. Very slowly evolving hybrid schemes (which are often the case) are not desirable, especially in very complex domains of application. Evidence drawn from literature on the effectiveness of a specific kind of hybrid methodologies in a variety of real-world applications could render this hybrid scheme over time as method of choice for the decision maker.

The decision on whether to use a hybrid intelligent methodology or not, lies in long experience and experimentation in comparison to standard intelligent approaches. When partial stages of the overall problem to be solved, seem better to be manipulated by different intelligent techniques (e.g. cluster formation, feature selection for reduction of complexity and so forth), then a hybrid

intelligent model should be applied for a better handling of the problem.

For the selection of the appropriate parts of a hybrid intelligent methodology, the advantages and disadvantages of each of the standard intelligent techniques should be taken into account. Inductive machine learning has always performed highly as a feature selection method for reducing complexity of the search space, especially (1) when entropy information criteria are applied, and (2) when the decision maker tries to analyse nominal data. It is also used as a pattern search mechanism within large collections of data, in order to apply later formal statistics for accepting or rejecting complex hypotheses. Models of evolutionary computation, genetic programming, etc., are time consuming in the training phase, so the decision maker should be aware of the complexity involved. On the other hand, they perform very well in generalization and robust model building from complex data. Fuzzy rule based approaches are by far advantageous in the handling of approximate or vague concepts defined, or existing, within a set of data. Finally, neural networks as stated above are typical black-box architectures (i.e. of very low comprehensibility of the produced decision model) that prove superior in handling numerical data and highly non-linear domains of application. Recent hybrid intelligent approaches can be found in (Siddique & Adeli 2013; Yang 2013).

Neuro-Fuzzy Systems

Neuro-fuzzy systems also called as neuro-fuzzy systems and techniques (NAISO 2002), are definitely the most popular hybrid intelligent systems in literature, due to the fact that they are fast, accurate and efficient approach, and can be understood, designed and implemented easily in automated computing environments. Neuro-fuzzy systems have shown a high rate of success when applied in complex domains of application, either when fuzzy set theory is the heart of such a system, or when the neural mechanism is the dominant component in the architecture. Neuro-fuzzy systems

are usually superior to simple neural networks, due to the fact that a neural network “suffers” from noise, whereas the neuro-fuzzy system has the ability to “absorb” the noise with the use of the embedded membership functions. For similar reasons, a neuro-fuzzy system is superior to a simple fuzzy system, as the neuro-fuzzy systems do not have to tune oneself the rule-base. The main principle of this combination, as seen by a neural network expert, can be roughly described as the adoption of fuzzy functions in (mostly consisted of 3-layers) neural networks’ nodes. On the other hand, a fuzzy systems’ expert may realize a neural-like training such as back-propagation for the membership functions of a fuzzy system. However, combinations and approaches in these hybrid systems can be less obvious and descriptive, concerning different neural or fuzzy structures, such as self-organizing maps or, radial basis functions. Generally the reader should also have in mind that the nature, the reliability and the availability of the data under processing, consist sometimes also a crucial factor for the success or the failure of a specific hybrid intelligent methodology. In their book (Siddique & Adeli, 2013), present advances in neuro-fuzzy systems. An interesting recent survey of the well-established neuro-fuzzy systems can be found in (Viharos & Kis, 2015).

Fuzzy-Evolutionary Systems

Fuzzy-evolutionary systems, in literature also found as evolving-fuzzy systems (EFS 2006; Cordon et al 2001; Kazabov, 2007; GECCO 2007; Lughofer, 2011) are also popular, used in the same complex domains as with the fuzzy-neural systems. Genetic algorithms and fuzzy logic have been used in the past collaboratively for various control engineering applications and complex optimisation problems. Both, fuzzy logic driven genetic approaches and genetic driven fuzzy logic based schemes have been proved effective in modern AI literature. The fuzzy logic driven genetic approaches primarily concern the use of fuzzy logic, either for genetic parameters’ tuning,

or for fuzzy encoding of the chromosomes. The genetic driven fuzzy logic based schemes usually consist of fuzzy rule-based systems, using a genetic approach for the determination of the rule base. An advantage of some of the fuzzy-evolutionary systems is considered the incorporation of fuzzy logic into an evolutionary algorithm. This fact makes hybrid schemes working effectively in a larger scale of application domains. Fuzzy-genetic systems are preferable than simple fuzzy systems, due to the fact that fuzzy-genetic approaches do not have to define oneself the rule-base. Recent advances in evolutionary fuzzy systems can be found in (Siddique & Adeli, 2013).

Neural Networks and Evolutionary Algorithms

The idea behind the implementation of such hybrid systems is the adoption of an evolutionary algorithm for the determination of neural network's weights or the neural network's architecture, or both. In the first case, neural networks are tuned by evolutionary algorithms, rather than generated by, which is the case in the second approach. The third approach may contain both generation and tuning. A special case of tree-like neural networks may also be served by genetic programming training. An interesting work on neuro-evolution is presented by Sher (2012). Also Siddique and Adeli (2013) give advances in evolutionary neural systems.

Machine Learning and Fuzzy Logic

Fuzzy logic is used for the modelling of ambiguity contained in decision attributes, before these attributes are subjected to further process using machine learning for classification and diagnosis tasks. The process followed for the definition of boundaries of the examined attributes, are defined by methods often called "fuzzy or soft thresholds". On the other hand, machine learning can assist the formation of fuzzy membership functions by defining successfully the fuzzy boundaries among neighbouring linguistic areas.

Machine Learning and Evolutionary Algorithms

Machine learning often works as feature selection or feature extraction methodology applied in large collections of data, prior to the application of evolutionary approaches for generalization from data. In this way, machine learning works as a mechanism for reducing complexity, a task which is necessary for time-consuming approaches such as evolutionary computation. Regarding the existence of successful hybrid intelligent methodologies in related literature, irrespective of the application area, approaches related to fuzzy control, function approximation, forecasting, knowledge discovery, decision making, scheduling, feature selection, system design, data classification and image processing can be found, e.g. see (Tsakonas & Dounias, 2002).

Other Hybrid Intelligent Schemes

There exist also intelligent hybrid systems that combine intelligent techniques with other standard mathematical methodologies. We define as *Type I - Hybrid*, those "hybrid" systems that combine two or more well-known intelligent techniques in a sensible methodological scheme in order to obtain a superior intelligent performance in complex domains of application. Similarly we define as *Type II - Hybrid*, those systems that combine a well-known intelligent technique and a standard mathematical or heuristic method, in order to effectively handle complex problems, still preserving the intelligence of the overall approach (Dounias 2003). Examples of this kind are fuzzy inference system optimization using global optimization algorithms, artificial neural network optimization using global optimization techniques, fuzzy clustering algorithms and optimization techniques, hybrid optimization techniques containing simulated annealing, tabu search, GRASP, case-based reasoning, etc., hybridization of soft computing and statistical learning techniques, hybridization with quantum computing, DNA computing, membrane computing and more.

FUTURE RESEARCH DIRECTIONS

Future research directions aim towards the production of user friendly computational environments for combining intelligent techniques and algorithms in handy and powerful hybrid formations, as well as towards the implementation of hybrid intelligent schemes within parallel computational architectures (computer clusters, cloud computing schemes) in order to reduce the time needed for the training of the algorithms and for the production of generalized solutions and models.

Hybrid intelligent systems keep developing in all kinds of data analysis problems like classification, optimization, forecasting, clustering etc. Among the most promising applications in the near future can be found in:

- Management, economics and finance (Sztandera, 2014), with applications in customer segmentation for direct marketing, customer profiling for relationship management, efficient mailing campaigns, customer retention, identification of cross-selling opportunities, credit score analysis, detection of fraudulent behavior and transactions, hedge fund strategies, etc.
- Tsutsui and Collet (2013) presenting various evolutionary methods with successful results from real-world problems in data mining, bioinformatics, drug discovery, crystallography, artificial chemistries, and sudoku.
- Iba and Noman (2016) referring to gene regulatory networks (GRN), covering sub-domains of GRN research such as expression profile analysis, reverse engineering, GRN evolution, and other applications related to systems biology, computational biology, and synthetic biology.

REFERENCES

- Abbod, M. F., Linkens, D. A., Mahfouf, M., & Dounias, G. (2002). Survey on the use of Smart and Adaptive Engineering Systems in Medicine. *Artificial Intelligence in Medicine*, 26(3), 179–209. doi:10.1016/S0933-3657(02)00083-0 PMID:12446078
- Bell, J. (2014). *Machine Learning: Hands-On for Developers and Technical Professionals*. Wiley. doi:10.1002/9781119183464
- Bezdek, J. C. (1994). *What is Computational Intelligence? Computational Intelligence Imitating Life*. Piscataway, NJ: IEEE Press.
- Brownlee J., (2012). *Clever Algorithms: Nature-Inspired Programming Recipes*. Lulu.com.
- Chen, Z. (2000). *Computational Intelligence for Decision Support*. CRC Press.
- Cordon, O., Herrera, F., Hoffmann, F., & Magdalena, L. (2001). *Genetic Fuzzy Systems – Evolutionary Tuning and Learning of Fuzzy Knowledge Bases*. World Scientific. doi:10.1142/4177
- Domingos, P. (2015). *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World*. Basic Books.
- Dorransoro, B., Danoy, G., Pigne, Y., & Bouvry, P. (2014). *Evolutionary Algorithms for Mobile Ad Hoc Networks. (Nature-Inspired Computing Series)* (1st ed.). Wiley/IEEE Computer Society. doi:10.1002/9781118833209
- Dounias, G. (2003). *Hybrid computational intelligence in medicine*. EUNITE Meeting on Intelligent and Adaptive Systems in Medicine, Prague, Czech Republic.
- Duch, W. (2007). What is Computational Intelligence and what could it become? In *Challenges for Computational Intelligence*. Retrieved from <http://cogprints.org/5358/1/06-CIdef.pdf>

- EFS 2006: International Symposium on Evolving Fuzzy Systems, Lancaster, UK. (2006). Retrieved from <http://eprints.lancs.ac.uk/893/>
- Eiben, A. E., & Smith, J. E. (2015). *Introduction to Evolutionary Computing. Natural Computing Series* (2nd ed.). Springer. doi:10.1007/978-3-662-44874-8
- EUNITE. 2001: The European Network on Intelligent Technologies for Smart Adaptive Systems: EUNITE – Symposium, Proceedings. (2001). ELITE Foundation.
- Flach, P. (2012). Machine Learning: The Art of Science of Algorithms that Make Sense of Data. *Proceedings of the Genetic and Evolutionary Computation Conference*. doi:10.1017/CBO9780511973000
- Grus, J. (2015). *Data Science from Scratch: First Principles with Python* (1st ed.). O'Reilly.
- Hagan, M. T., Demuth, H. B., Beale, M. H., & De Jesus, O. (2014). *Neural Network Design* (2nd ed.). Academic Press.
- HAIS. 2013: 8th Int. Conf. on Hybrid Artificial Intelligence Systems. (2013). Retrieved from <http://hais.usal.es/>
- HAIS. 2006: 1st Int. Conf. on Hybrid Artificial Intelligence Systems. (2006). Retrieved from <http://hais.usal.es/?q=node/10>
- HAIS-2016. 11th International Conference on Hybrid Artificial Intelligence Systems. (n.d.). Retrieved from <http://hais2016.upo.es/>
- HIS 2001: 1st Int. Workshop on Hybrid Intelligent Systems. (2001). Retrieved from <http://www.softcomputing.net/his01/>
- HIS 2013: 13th Int. Conf. on Hybrid Intelligent Systems. (2013). Retrieved from <http://www.mirlabs.net/his13/>
- HIS-2014. 14th International Conference on Hybrid Intelligent Systems (HIS 2014) Kuwait. (2014). IEEE.
- Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems*. MIT Press.
- Iba, H., & Noman, N. (2016). *Evolutionary Computation in Gene Regulatory Network Research*. Wiley Series in Bioinformatics. doi:10.1002/9781119079453
- Jantzen, J. (2013). *Foundations of Fuzzy Control: A Practical Approach* (2nd ed.). Wiley. doi:10.1002/9781118535608
- Kacprzyk, J., & Pedrycz, W. (2015). *Springer Handbook of Computational Intelligence*. Springer. doi:10.1007/978-3-662-43505-2
- Kasabov, N. (2007). *Evolving Connectionist Systems: The Knowledge Engineering Approach* (2nd ed.). London: Springer.
- Kelleher, J. D., Mac Namee, B., & D'Arcy, A. (2015). *Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms Worked Examples, and Case Studies*. Cambridge, MA: Academic Press.
- Koza, J. R. (1992). *Genetic Programming – On the Programming of Computers by Means of Natural Selection*. MIT Press.
- Kruse, R., Borgelt, C., Klawonn, F., Moewes, C., Steinbrecher, M., & Held, P. (2013). *Computational Intelligence: A Methodological Introduction. (Texts in Computer Science)*. Springer. doi:10.1007/978-1-4471-5013-8
- Kubat, M., Bratco, I., & Michalski, R. S. (1997). A Review of Machine Learning Methods. In R. S. Michalski, I. Bratco, & M. Kubat (Eds.), *Machine Learning And Data Mining – Methods and Applications* (pp. 3–69). Wiley.
- Lantz, B. (2013). *Machine Learning with R*. Packt Publishing.
- Lee Giles, C., Sun, R., & Zurada, J. M. (Eds.). (1997). Special Issue on Neural Networks & Hybrid Intelligent Models, Foundations, Theory & Applications. *IEEE Transact. on Neural Networks*. Retrieved from <http://www.cogsci.rpi.edu/~rsun/si97.ann>

- Lughofer, E. (2011). *Evolving Fuzzy Systems, Methodologies, Advanced Concepts and Applications*. Springer-Verlag.
- Michalski, R. S., Carbonell, J. G., & Mitchell, T. M. (1983). *Machine Learning: An Artificial Intelligence Approach*. Morgan Kaufmann. doi:10.1007/978-3-662-12405-5
- Michalski, R. S., Carbonell, J. G., & Mitchell, T. M. (1986). *Machine Learning: An Artificial Intelligence Approach* (Vol. 2). Morgan Kaufmann.
- Mitchell, T. M. (1997). *Machine Learning*. New York: McGraw-Hill.
- NAISO. (2002). *1st International NAISO Congress on Neuro-Fuzzy Technologies*. Habana, Cuba: ICSC-NAISO Academic Press.
- Negoita, M. G., Neagu, D., & Palade, V. (2005). *Computational Intelligence – Engineering of Hybrid Systems. Studies in Fuzziness and Soft Computing*. Springer.
- Nilsson, N. (1998). *Artificial Intelligence: A New Synthesis*. Morgan Kaufmann.
- NISIS. (2005). *Nature Inspired Smart Information Systems, European Project 2005-2008*. Retrieved from <http://www.nisis.risk-technologies.com/default.aspx>
- Raschka, S. (2015). *Python Machine Learning*. Packt Publishing.
- Roghani, A. (2015). *Artificial Neural Networks: Applications in Financial Forecasting*. CreateSpace Independent Publishing Platform.
- Rosenblatt, F. (1959). Two theorems of statistical separability in the perceptron, Mechanization of Thought Processes, London HM Stat. *The Office*, 421–456.
- Shaffer S. C., (2015). *Code Your Own Neural Network: A step-by-step explanation*. Shaffer Media Enterprises LLC.
- Sher, G. I. (2012). *Handbook of Neuroevolution Through Erlang*. Springer Verlag.
- Siddique, N., & Adeli, H. (2013). *Computational Intelligence: Synergies of Fuzzy Logic, Neural Networks and Evolutionary Computing* (1st ed.). Wiley. doi:10.1002/9781118534823
- Simon, D. (2013). *Evolutionary Optimization Algorithms* (1st ed.). Wiley.
- Singhai, R. (2013). *Introduction to Fuzzy Logic*. PHI Learning.
- Sztandera, L. (2014). *Computational Intelligence in Business Analytics: Concepts, Methods, and Tools for Big Data Applications* (1st ed.). Pearson.
- Tamir, D. E., Rishe, N. D., & Kandel, A. (2015). *Fifty Years of Fuzzy Logic and its Applications. Studies in Fuzziness and Soft Computing*. Springer. doi:10.1007/978-3-319-19683-1
- Trillas, E., & Eciolaza, L. (2015). *Fuzzy Logic: An Introductory Course for Engineering Students. Studies in Fuzziness and Soft Computing*. Springer. doi:10.1007/978-3-319-14203-6
- Tsakonas, A., & Dounias, G. (2002). Hybrid Computational Intelligence Schemes in Complex Domains: A Review of Selected Publications, LNAI-2308. In I. Vlahavas & C. Spyropoulos (Eds.), *Methods and Applications of Artificial Intelligence* (pp. 494–511). Springer Publications. doi:10.1007/3-540-46014-4_44
- Tsutsui, S., & Collet, P. (2013). *Massively Parallel Evolutionary Computation on GPGPUs*. Springer. doi:10.1007/978-3-642-37959-8
- Vassiliadis, V., & Dounias, G. (2009). Nature Inspired Intelligence: A review of selected methods and applications. *International Journal of Artificial Intelligence Tools*, 18(4), 487–516. doi:10.1142/S021821300900024X
- Viharos, Z. J., & Kis, K. B. (2015). Survey on Neuro-Fuzzy Systems & their Apps in Technical Diagnostics Measurement. *Measurement*, 67, 126–136. doi:10.1016/j.measurement.2015.02.001

- Werbos, P. (1974). *Beyond regression: new tools for predictions and analysis in the behavioral science* (PhD Thesis). Harvard University.
- Widrow B. and Hoff M.E. (1960). Adaptive switching circuits. *IRE Western Electric Show and Convention Record*, 4, 96-104.
- Yang, X.-S. (2014). *Nature-Inspired Optimization Algorithms. Elsevier Insights*. Elsevier.
- Zadeh, L. A. (1965). Fuzzy Sets. *Information and Control*, 8(3), 338–353. doi:10.1016/S0019-9958(65)90241-X
- Zadeh, L.A. (2001, Spring). A New Direction in AI: Toward a Computational. Theory of Perceptions. *AI Magazine*, 73-84.
- Zimmermann, H.-J., Tselentis, G., Van Someren, M., & Dounias, G. (Eds.). (2001). *Advances in Computational Intelligence and Learning: Methods and Applications*. Kluwer.
- Lughofer, E. (2011). *Evolving Fuzzy Systems: Methodologies, Advanced Concepts and Applications*. Springer Heidelberg. doi:10.1007/978-3-642-18087-3
- Negoita, M. G., Neagu, D., & Palade, V. (2005). *Computational Intelligence – Engineering of Hybrid Systems. Studies in Fuzziness and Soft Computing*. Springer.
- Ovaska, S. J. (Ed.). (2004). *Computationally Intelligent Hybrid Systems: The Fusion of Soft Computing and Hard Computing*. Wiley-IEEE Press. doi:10.1002/9780471683407
- Siddique, N., & Adeli, H. (2013). *Computational Intelligence: Synergies of Fuzzy Logic, Neural Networks and Evolutionary Computing* (1st ed.). Wiley. doi:10.1002/9781118534823
- Zimmermann, H.-J. Tselentis G., Van Someren M., Dounias G. (Eds.), (2001). *Advances in Computational Intelligence and Learning: Methods and Applications*. Kluwer

ADDITIONAL READING

- Angelov, P., Filev, D. P., & Kasabov, N. (Eds.). (2010). *Evolving Intelligent Systems: Methodology and Applications*. Wiley-IEEE Press. doi:10.1002/9780470569962
- Castillo, O., Melin, P., & Pedrycz, W. (Eds.). (2008). *Soft Computing for Hybrid Intelligent Systems* (Studies in Computational Intelligence), Springer Series: Studies in Computational Intelligence, Vol. 154.
- Cordon, O., Herrera, F., Hoffmann, F., & Magdalena, L. (2001). *Genetic Fuzzy Systems – Evolutionary Tuning and Learning of Fuzzy Knowledge Bases*. World Scientific. doi:10.1142/4177
- Fuller, R. (2000). *Introduction to Neuro-Fuzzy Systems*. Springer. doi:10.1007/978-3-7908-1852-9

ENDNOTES

- ¹ Computational Intelligence: Principles, Techniques and Applications, Springer, p.5 http://download.springer.com/static/pdf/873/chp%253A10.1007%252F3-540-27335-2_1.pdf?auth66=1386318381_02e21cdef958e405a9deec06c2642d98&ext=.pdf
- ² Some researchers classify neural networks as machine learning technique.
- ³ Data compiled by Dr. B. Tadayon, CEO, Z Advanced Computing TM, Inc.
- ⁴ <http://www.softcomputing.net/ijhis/>

Incremental Approach to Classification Learning

A

Xenia Alexandre Naidenova

Research Centre of Military Medical Academy – Saint Petersburg, Russia

INTRODUCTION

By classification we mean partition of a given object's set into disjoint blocks or classes. We assume that objects are described by a set U of symbolic or numeric attributes and each object can have one and only one value of each attribute. Then each attribute generates, by its values, partition of a given set of objects into mutually disjoint classes the number of which is equal to the number of values of this attribute. To give a target classification of objects, we use an additional attribute KL not belonging to U . In Table 1, we have two classes: $KL+$ (positive objects) and $KL-$ (negative objects).

By classification learning we mean approximation of given object classification in terms of attributes names or values of attributes (Naidenova, 2012). This approximation is reduced to extracting logical rules in the form of functional or implicative dependencies from observable datasets. These dependencies allow to distinguish between classes of given classification. For our example in Table 1, we have some rules based on implicative (ID) and functional dependencies (FD): $Color_of_Hairs, Color_of_Eyes \rightarrow KL$ (FD); if Blond, Blue, then $KL = "+"$; if Hazel, then $KL = "-"$; if Brown, Blue, then $KL = "-"$ (IDs).

The task of classification learning based on inferring implicative rules is equivalent to the task of concept formation (Banerji, 1969, Ganter & Wille, 1999). The goal of this task is to describe/classify new objects according to description/classification of existing objects. Inferring good diagnostic (classification) tests (GDTs) is the formation of the best descriptions of a given object

class $KL+$ against the objects not belonging to this class ($KL-$).

Let $M = (\cup \text{dom}(\text{attr}), \text{attr} \in U)$, where $\text{dom}(\text{attr})$ is the set of all values of attr . Let $X \subseteq M$ and G be the set of indices of objects considered (objects for short), $G = G+ \cup G-$, where $G+$ and $G-$ the sets of positive and negative objects, respectively. Denote by $d(g)$ the description of object $g \in G$. Let $P(X) = \{g \mid g \in G, X \subseteq d(g)\}$. We call $P(X)$ the interpretation of X in the power set 2^G . If $P(X)$ contains only positive objects and the number of these objects more than 2, then we call X a description of some positive objects and $(P(X), X)$ a test for positive objects. Let us define a good test or good description of objects.

Definition 1: A set $X \subseteq M$ of attribute values is a good description of positive (negative) objects if and only if it is the description of these objects and no such subset $Y \subseteq M$ exists, that $P(X) \subset P(Y) \subseteq G+ (\subseteq G-)$.

Table 1. Example of classification

Index of Example	Height	Color of Hair	Color of Eyes	KL
1	Low	Blond	Blue	+
2	Low	Brown	Blue	-
3	Tall	Brown	Hazel	-
4	Tall	Blond	Hazel	-
5	Tall	Brown	Blue	-
6	Low	Red	Blue	-
7	Tall	Red	Blue	+
8	Tall	Blond	Blue	+

DOI: 10.4018/978-1-5225-2255-3.ch017

It has been shown (Naidenova, 1992) that the problem of good tests inferring is reduced to searching for implicative dependencies in the form $X \rightarrow v$, $X \subseteq M$, $v \in \text{dom}(\text{KL})$ for all positive (negative) objects.

The concept of good classification (diagnostic) test has firstly been introduced in (Naidenova & Polegaeva, 1986). In (Naidenova, 2012), it is considered the link between classification learning based on inferring good tests and formal concepts in the FCA.

BACKGROUND DEFINITIONS

Let $G = \{1, 2, \dots, N\}$ be the set of objects' indices (objects, for short) and $M = \{m_1, m_2, \dots, m_j, \dots, m_q\}$ be the set of attributes' values (values, for short). Each object is described by a set of values from M . The object descriptions are represented by rows of a table R the columns of which are associated with the attributes taking their values in M . Let $D(+)$ and $G(+)$ be the sets of positive object descriptions and the set of indices of these objects, respectively. Then $D(-) = D/D(+)$ and $G- = G/G+$ are the sets of negative object descriptions and indices of these objects, respectively.

The definition of good tests as a dual construction or formal concept is based on two mapping $2^G \rightarrow 2^M$, $2^M \rightarrow 2^G$ determined as follows. $A \subseteq G$, $B \subseteq M$. Denote by B_i , $B_i \subseteq M$, $i = 1, \dots, N$ the description of object with index i . We define the relations $2^G \rightarrow 2^M$, $2^M \rightarrow 2^G$ as follows: $A' = \text{val}(A) = \{\text{intersection of all } B_i: B_i \subseteq M, i \in A\}$ and $B' = \text{obj}(B) = \{i: i \in G, B \subseteq B_i\}$. These mapping are the Galois's correspondences (Ore, 1944). Of course, we have $\text{obj}(B) = \{\text{intersection of all } \text{obj}(m): \text{obj}(m) \subseteq G, m \in B\}$. Operations $\text{val}(A)$, $\text{obj}(B)$ are reasoning operations (derivation operations).

We introduce two generalization operations: $\text{generalization_of}(B) = B'' =$; $\text{generalization_of}(A) = A'' =$. These operations are the closure operations (Ore, 1944).

A set A is closed if $A = \text{obj}(\text{val}(A))$. A set B is closed, if $B = \text{val}(\text{obj}(B))$. For $g \in G$ and m

$\in M$, g' is called object intent and m' is called value extent.

By using the dataset in Table 1, we illustrate the derivation and generalization operations: $A = \{7, 8\}$, $\text{val}(A) = \{\text{Tall, Blue}\}$; $A'' = \text{obj}(\{\text{Tall Blue}\}) = \{5, 7, 8\}$;

$m = \{\text{Red}\}$, $\text{obj}(\{m\}) = \{6, 7\}$; $m'' = \text{val}(\{6, 7\}) = \{\text{Red, Blue}\}$;

$B = \{\text{Low, Blue}\}$, $\text{obj}(\{B\}) = \{1, 2, 6\}$; $B'' = \text{val}(\{1, 2, 6\}) = \{\text{Low, Blue}\} = B$.

Classification of objects are defined as follows (Kuznetsov, 1999). Let a context $K = (G, M, I)$ be given, where $I \subseteq G \times M$. In addition to values of M , a target value $\omega \notin M$ is considered. The set of objects G is partitioned into two subsets: $G+$ of objects with property ω (positive objects) and $G-$ without this property (negative objects).

Then $K = K+ \cup K-$; $K+ \cap K- = \emptyset$; $K+ = (G+, M, I+)$; $K- = (G-, M, I-)$; $G = G+ \cup G-$; $G+ \cap G- = \emptyset$.

Diagnostic test is defined as follows.

Definition 2: A diagnostic test for $G+$ is a pair (A, B) such that $B \subseteq M$ ($A = \text{obj}(B) \neq \emptyset$), $A \subseteq G+$, and $\forall g, g \in G-: B \not\subseteq \text{val}(g)$ and $B \neq \text{val}(g)$.

Definition 3: A diagnostic test (A, B) such that $B \subseteq M$ ($A = \text{obj}(B) \neq \emptyset$) is a good test for $G+$ if and only if extension $A^* = A \cup g$, $g \notin A$, $g \in G+$ implies that $(A^*, \text{val}(A^*))$ is **not a test** for $G+$.

Definition 4: A diagnostic test (A, B) , $B \subseteq M$ ($A = \text{obj}(B) \neq \emptyset$) is a good maximally redundant test (GMRT) for $G+$ if any extension $B^* = B \cup m$, $m \notin B$, $m \in M$ implies that $\text{obj}(B^*) \subseteq \text{obj}(B)$ ($(\text{obj}(B^*), B^*)$ is a test for $G+$ but not good).

It is important to note that if a pair (A, B) is a maximally redundant test, then A and B are closed and, consequently, this test is a formal concept in terms of the Formal Concept Analysis (the FCA)

(Ganter and Wille, 1999). But, in general case, diagnostic test is not obligatory a formal concept (Naidenova, 2012).

Definition 5: A diagnostic test (A, B) , $B \subseteq M$ ($A = \text{obj}(B) \neq \emptyset$) is a good irredundant test (GIRT) for $G+$ if any narrowing $B^* = B \setminus m$, $m \in B$, $m \in M$ implies that $(\text{obj}(B^*), B^*)$ is not a test for $G+$.

If a good test (A, B) for $G+$ is irredundant, then any narrowing $B^* = B \setminus m$, $m \in B$, $m \in M$ implies that $(\text{obj}(B^*), B^*)$ is not a test for $G+$. If a good test (A, B) for $G+$ is maximally redundant, then any extension $B^* = B \cup m$, $m \notin B$, $m \in M$ implies that $(\text{obj}(B^*), B^*)$ is not a good test for $G+$.

Definitions 2, 3, 4, 5 remain true if $G+$ is replaced by $G-$.

In what follows, we are interested in inferring GMRT or good concepts (GCs).

Some examples of formal concepts are in Table 2. Let us check if a pair $(A, B) = ((5, 6), (\text{Rain}, \text{Cool}, \text{Normal}))$ is a concept or not. It is a concept, because $\text{val}(5, 6) = (\text{Rain}, \text{Cool}, \text{Normal}) = B$, $\text{obj}((\text{Rain}, \text{Cool}, \text{Normal})) = (5, 6) = A$. However, this concept doesn't distinguish the classes of

objects (it is not a test for both $k+$ and $k-$). Pair $((3, 13), (\text{Overcast}, \text{Hot}, \text{No}))$ is a concept for $k(-)$, but not a good one, because there is pair $((3, 7, 12, 13) (\text{Overcast}))$ such that $(3, 13) \subset (3, 7, 12, 13)$.

A

INCREMENTAL LEARNING GOOD CONCEPTS

Incremental concept learning is very useful for many information retrieval applications related to the changeable environment. For example, mining high utility itemsets in dynamic databases (Asha et al., 2014), intrusion detection (Fan, et al., 2013; Chen et al., 2016), Web page classification (Ciarelli et al., 2014), association rule mining (Duaimi & Salman, 2014), constructing dynamic information systems (Liu et al., 2011; Liu, et al. 2014), and many others.

To transform inferring good formal concepts (GCs) into an incremental process, there are two kinds of subtasks:

- To find all GCs intents of which are included in the description of an object (the first kind subtask); and

Table 2. Examples of formal concepts

G	Outlook	Temperature	Humidity	Windy	Class
1	Sunny	Hot	High	No	$k(+)$
2	Sunny	Hot	High	Yes	$k(+)$
3	Overcast	Hot	High	No	$k(-)$
4	Rain	Mild	High	No	$k(-)$
5	Rain	Cool	Normal	No	$k(-)$
6	Rain	Cool	Normal	Yes	$k(+)$
7	Overcast	Cool	Normal	Yes	$k(-)$
8	Sunny	Mild	High	No	$k(+)$
9	Sunny	Cool	Normal	No	$k(-)$
10	Rain	Mild	Normal	No	$k(-)$
11	Sunny	Mild	Normal	Yes	$k(-)$
12	Overcast	Mild	High	Yes	$k(-)$
13	Overcast	Hot	Normal	No	$k(-)$
14	Rain	Mild	High	Yes	$k(+)$

- To find all GCs including a given set of values (the second kind subtask).

For solving these subtasks, we need to form sub-contexts of a given classification context. The following notions of object and value projections are developed to form sub-contexts.

Definition 6: Object projection $\text{proj}(d)$ of a positive object description d on the set $D+$, i.e. descriptions of all positive objects, is equal to $Z = \{z: z = d \cap d^* \neq \emptyset, d^* \in D+ \text{ and } z \text{ is the intent of a test for } G+\}$.

An example of object projection $\text{proj}(d_3)$ on $D(-)$ is $\{z_1; z_2; z_3; z_4\}$ (see, please Table 3).

Definition 7: The value projection $\text{proj}(B)$ on a given set $D+$ is $\text{proj}(B) = \{d: B \subseteq d, d \in D+\}$.

An example of value projection $\text{proj}(\text{Cool})$ on $D(-)$ is $\{d_5; d_7; d_9\}$ (see, please, Table 4).

In order to find all GCs in an object or value projection, it is possible to use any of well known algorithms for inferring GCs (Naidenova, 2006;

Naidenova & Parkhomenko, 2013; Godin et al., 1995). We shall concentrate on the problem of incrementally change a set of GCs in changeable classification contexts when an object or value is added or removed from consideration.

FUTURE RESEARCH DIRECTIONS: FOUR CASES OF INCREMENTAL LEARNING GCs

The GCs are the minimal formal concepts by the definition in (Ganter & Kuznetsov, 2000). Incremental learning the formal concepts requires an algorithm of Galois lattice incremental construction from data. In this process, it is generally assumed that the data (objects, itemsets, or transactions) are added gradually but not deleted. Much attention has been paid in recent years to the problem of concept lattice incremental construction (Godin et al., 1995; Kourie et al., 2009; Van der Merwe et al., 2004; Outrata, 2013). An algorithm of incremental generating GCs has been proposed in (Naidenova, 2006) and investigated in more detail in (Naidenova & Parkhomenko, 2013).

Table 3. An example of object projection

G	Outlook	Temperature	Humidity	Windy	Class	DT ? Z
1	Sunny	Hot	High	No	k(+)	
2	Sunny	Hot	High	Yes	k(+)	
3	Overcast	Hot	High	No	k(-)	Z1, yes
4	Rain	Mild	High	No	k(-)	no
5	Rain	Cool	Normal	No	k(-)	no
6	Rain	Cool	Normal	Yes	k(+)	
7	Overcast	Cool	Normal	Yes	k(-)	Z2, Yes
8	Sunny	Mild	High	No	k(+)	
9	Sunny	Cool	Normal	No	k(-)	no
10	Rain	Mild	Normal	No	k(-)	
11	Sunny	Mild	Normal	Yes	k(-)	
12	Overcast	Mild	High	Yes	k(-)	Z3
13	Overcast	Hot	Normal	No	k(-)	Z4
14	Rain	Mild	High	Yes	k(+)	

Table 4. An Example of value projection

G	Outlook	Temperature	Humidity	Windy	Class
1	Sunny	Hot	High	No	k(+)
2	Sunny	Hot	High	Yes	k(+)
3	Overcast	Hot	High	No	k(-)
4	Rain	Mild	High	No	k(-)
5	Rain	Cool	Normal	No	k(-)
6	Rain	Cool	Normal	Yes	k(+)
7	Overcast	Cool	Normal	Yes	k(-)
8	Sunny	Mild	High	No	k(+)
9	Sunny	Cool	Normal	No	k(-)
10	Rain	Mild	Normal	No	k(-)
11	Sunny	Mild	Normal	Yes	k(-)
12	Overcast	Mild	High	Yes	k(-)
13	Overcast	Hot	Normal	No	k(-)
14	Rain	Mild	High	Yes	k(+)

On the other hand, there is a practical demand to modify the concept lattice already constructed under dynamic data changes. In this case, it is necessary to consider both adding and deleting the data (objects, attributes). This problem is not yet investigated sufficiently. Deleting objects (and only objects) is considered in Carpineto and Romano (2004) and Zhang et al. (2013). In the first paper, algorithm *RemoveObject* is advanced, in the second one, algorithm *DeleteObject* is proposed. These algorithms have been essentially improved with respect to their computational complexity in Zou et al. (2015) in newly proposed algorithm *FastDeletion*.

However modifying the formal classification contexts can be required not only by adding or deleting objects but also by adding or deleting attributes (values). Such a modification of concept lattice is even less explored than deleting objects. Of interest in this regard, the article (Kauer & Krupka, 2014), in which the problem of removing an incidence from a formal context is solved.

The four variants of changing formal contexts are considered in (Carpineto & Romano, 2004). Recent publication on this topic (Oustrata, 2013)

provides an efficient algorithm of modifying the formal contexts by adding objects, which may include, in their descriptions, some new attributes. Modification of the order relation in concept lattice is also determined. The proposed algorithm defines new and modified concepts without using previously built lattice but considering the only available data. It is based on algorithm *Close-by-One* (CbO) of generating formal concepts (Kuznetsov, 1993) and its next refinements *FCbO* (Oustrata, 2012) and *PFCbO* (Oustrata, 2010). As for removing objects, this process is reduced to adding objects. The paper (Wray et al., 2016) presents an implementation and evaluation of the *ECbO* updated algorithm.

We propose four cases of modifying classification contexts: adding/removing objects and adding/removing values of attributes. New and modified GCs are determined. Modification of GCs is based on a decomposition of formal classification contexts into value and object sub-contexts and using an incremental algorithm for inferring GCs in these sub-contexts (Naidenova & Parkhomenko, 2013; Naidenova et al., 2015).

Case 1

Suppose that each new object comes with the indication of its class membership. The following actions are necessary:

1. Checking whether it is possible to enlarge the extents of some existing GCs for the class to which a new object belongs (a class of positive objects, for certainty);
2. Inferring all GCs, intents of which are included into the new object description; for this goal, we use the first kind subtask; and
3. Checking the validity of GCs for negative objects, and, if it is necessary, modifying invalid GCs (test for negative objects is invalid if its intent is included in a new (positive) object description); for this goal, we use the second kind subtask.

Let $STGOOD+$ and $STGOOD-$ be the current sets of GCs for positive and negative class of objects, respectively, and M be the set of all attribute values used for object descriptions.

Let $GC \in STGOOD-$ and $Y = \text{intent}(GC)$. If $Y \subseteq \text{tnew}(+)$, where $\text{tnew}(+)$ is the description of a new positive object, then GC should be deleted from $STGOOD-$.

For correcting the GC, we have to find all $X \subseteq M$ such that $Y, Y \subseteq M, Y \subseteq X$, and $(\text{extent}(X), X)$ is a GC_{new} for $G-$.

Consider an example of adding object in the process of inferring CGs for the data in Table 2.

Let us fix the K with first 6 objects and add 7th object description. As a result, $GC = ((2,6), (\text{Yes}))$ for $k(+)$ becomes invalid, and a new GC for $k(-)$ appears $((3,7), (\text{Overcast}))$ (Table 5).

Case 2

Suppose that an object is deleted from a classification context. The following actions are necessary:

1. Selecting the set GC_{sub} of $STGOOD+$ ($STGOOD-$) containing deleted object in their extents; and
2. Modifying concepts of GC_{sub} by removing the deleted object from their extents.

After modifying a GC in GC_{sub} , we have the following possibilities:

- Intent of GC_{new} is not changed ($\text{intent}(GC_{\text{new}}) = \text{intent}(GC)$). Then GC_{new} is a good concept in modified context; and
- $\text{Intent}(GC_{\text{new}}) \neq \text{intent}(GC)$ and $\text{extent}(GC_{\text{new}})$ is included in the extent of an existing GC in $STGOOD+$ ($STGOOD-$), then GC_{new} can be deleted; else this GC_{new} is a good concept in the modified context.

Let us fix the whole classification context (Table 2) and all the GCs obtained in this context (Table 6) and delete object 9. The following result: $GC = ((5,10,13), (\text{Normal}, \text{No}))$ remains, $GC = ((11), \text{val}(11))$ is deleted, because it is not good.

Table 5. Adding object (see, please, Table 4)

G	Extents of GCs (Current Situation)
Current objects (1;...; 6)	Extents of GCs in $STGOOD(+) = \{(1, 2); (2, 6)\}$, Extents of GCs in $STGOOD(-) = \{(3); (4, 5)\}$
Adding object 7	Extents of GCs in $STGOOD(+) = \{(1, 2), (6)\}$, Extents of GCs in $STGOOD(-) = \{(3, 7), (4, 5)\}$

Table 6. Deleting an object

GCs(-)	GCs(+)
$((1; 2; 8), \text{Sunny High})$ $((6; 14), \text{Rain Yes})$	$((4; 5; 10), \text{Rain No})$ $((5; 9; 10; 13), \text{Normal No})$ $((10; 11), \text{Mild Normal})$ $((9; 11), \text{Sunny Normal})$ $((3; 7; 12; 13), \text{Overcast})$

Case 3

Suppose that a new value m^* is added to the set M of values. The task of finding all GCs whose intents contain m^* is reduced to the task of the second kind.

Case 4

Suppose that some value m is deleted from consideration. It is necessary to consider two possibilities: $\text{extent}(\text{GC}_{\text{new}}) = \text{extent}(\text{GC})$ and $\text{extent}(\text{GC}) \subset \text{extent}(\text{GC}_{\text{new}})$. In the first case, GC_{new} is a good concept for modified context. In the second case, GC_{new} is not a good concept for modified context. However, the extent of GC_{new} can contain extents of new GCs and these GCs can be obtained by using the subtasks of first or second kind.

Some words about incremental pattern recognition based on GCs. Recognizing the class membership for a new object not belonging to training set is performed as follows:

- If (and only if) description of object contains an intent of GC of only one class, then the object can be assigned to this class;
- If description of object contains intents of GCs of both Class^+ and Class^- , then we have the case of contradiction; and
- If description of an object does not contain any intent of GCs, then we have the case of uncertainty.

In two last cases, it is necessary to continue learning by adding new objects or to change the classification context.

This approach to incremental learning GMRTs (GCs) has been applied to analyzing the intellectual development of medical cadets (Naidenova et al., 2016).

CONCLUSION

It is proposed a model of incremental classification learning based on incremental inferring good formal concept and their correcting in changeable classification contexts.

REFERENCES

- Asha, P., Jebara, T., & Saranya, G. (2014). A Survey on efficient incremental algorithm for mining high utility itemsets in distributive and dynamic database. *International Journal of Emerging Technology and Advanced Engineering*, 4(1), 146-149.
- Banerji, R. (1969). *Theory of Problem Solving: An Approach to Artificial Intelligence*. New York: American Elsevier Publishing Company.
- Carpineto, C., & Romano, G. (2004). *Concept data analysis: theory and application*. John Wiley & Sons. doi:10.1002/0470011297
- Chen, M.-H., Chang, P.-C., & Wu, J.-L. (2016). A population-based incremental learning approach with artificial immune system for network intrusion detection. *Engineering Applications of Artificial Intelligence*, 51(5), 171-181. doi:10.1016/j.engappai.2016.01.020
- Ciarelli, P., Oliveira, E., & Salles, N. (2014). Multi-label incremental learning applied to Web page categorization. *Neural Computing & Applications*, 24(6), 1403-1419. doi:10.1007/s00521-013-1345-7
- Duaimi, M.G., & Salman, A.A. (2014). Association rules mining for incremental Database. *International Journal of Advanced Research in Computer Science and Technology*, 2(2), 346-352.
- Fan, W., Bouguila, N., & Sallay, H. (2013). Anomaly intrusion detection using incremental learning of an infinite mixture model with feature selection. In *Proceedings of 8th International Conference on Rough sets and Knowledge Technology (RSKT-2013): LNCS 8171* (pp. 364-373). Berlin: Springer. doi:10.1007/978-3-642-41299-8_35

- Ganter, B., & Kuznetsov, S. (2000). Formalizing hypotheses with concepts. In B. Ganter & G. W. Mineau (Eds.), *Proceedings of the 8th International Conference on Concept Structures (ICCS 2000), Conceptual Structures: Logical Linguistics, and Computational Issues* (vol. 1867, pp. 342-356). Berlin, Germany: Springer-Verlag.
- Ganter, B., & Wille, R. (1999). *Formal concept analysis: mathematical foundations*. Berlin, Germany: Springer. doi:10.1007/978-3-642-59830-2
- Godin, R., Missaoui, R., & Alaoui, H. (1995). Incremental concept formation algorithms based on Galois' (concept) lattices. *Computational Intelligence*, 11(2), 246-247.
- Jia, Z.-J., Hao, W.-X., & Zhang, X.-Y. (2011). Application of TSVM incremental learning in Web text categorization. *Advance Mathematical Research*, 159(4), 556-561.
- Kauer, M., & Krupka, M. (2014). Removing an incidence from a formal context. In K. Bertet, & S. Rudolph (Eds.), *Proceedings of CLA-2014*, (vol. 1, pp. 195-207). Košice, Slovakia: P.J. Šatáric University.
- Kourie, D., Obiedkov, S., Watson, W., & Van der Merwe, D. (2009). An incremental algorithm to construct a lattice of intersections. *Science of Computer Programming*, 74(3), 128-142. doi:10.1016/j.scico.2008.09.015
- Krajca, P., Outrata, J., & Vychodil, V. (2010). Advances in algorithms based on CbO. *Proceedings of 7th International Conference on Concept Lattices and Their Applications*, 71-82.
- Kuznetsov, S. (1999). Learning of simple conceptual graphs from positive and negative examples. In J. Zytkov, & J. Bauch (Eds.), *Proceedings of 3rd European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD 1999)*, (vol. 1704, pp. 384-391). Berlin, Germany: Springer-Verlag. doi:10.1007/978-3-540-48247-5_47
- Kuznetsov, S. O. (1993). A fast algorithm for computing all intersections of objects in a finite semi-lattice. *Nauchno-Tekhnicheskaya Informatsia*, 27(5), 11-21.
- Liu, D., Li, T., Ruan, D., & Zhang, J. (2011). Incremental learning optimization on knowledge discovery in dynamic business intelligent systems. *Journal of Global Optimization*, 51(2), 325-344. doi:10.1007/s10898-010-9607-8
- Liu, D., Li, T., & Zhang, J. (2014). A rough set based incremental approach for learning knowledge in dynamic incomplete information systems. *International Journal of Approximate Reasoning*, 55(8), 1764-1786. doi:10.1016/j.ijar.2014.05.009
- Naidenova, X. (1992). Machine learning as a diagnostic task. In I. Arefiev (Ed.), *Knowledge-Dialogue-Solution, Materials of the Short-Term Scientific Seminar*. Saint-Petersburg, Russia: State North-West Technical University.
- Naidenova, X. (2006). An incremental learning algorithm for inferring logical rules from examples in the framework of the common reasoning process. In E. Triantaphyllou & G. Felici (Eds.), *Data Mining and Knowledge Discovery Approaches Based on Rule Induction Techniques* (pp. 89-14). Springer. doi:10.1007/0-387-34296-6_3
- Naidenova, X. (2012). Good classification tests as formal concepts. In F. Domenach, D. Ignatov, & J. Poelmans (Eds.), *Formal Concept Analysis*, (vol. 7278, pp. 211-226). Springer. doi:10.1007/978-3-642-29892-9_21
- Naidenova, X., & Parkhomenko, V. (2013). An approach to incremental learning good classification tests. In *Contribution to 11th International Conference on Formal Concept Analysis* (pp. 51-64). Dresden, Germany: Dresden Technische Universität.

Naidenova, X., Parkhomenko, V., & Shvetsov, K. (2015). Context-dependent incremental learning of good maximally redundant tests In *Proceedings of the 2015 SAI Intelligent Systems Conference*, (pp. 957-962). London, UK: IEEE. doi:10.1109/IntelliSys.2015.7361258

Naidenova, X., Parkhomenko, V., Shvetsov, K., Ysupov, V., & Kuzina, R. (2016). Modification of good tests in dynamic contexts: application to modeling intellectual development of cadets. In M.- O. Aciego, D. Ignatov, & A. Lepskiy (Eds.), *Proceedings of the Second International Workshop on Soft Computing Applications and Knowledge Discovery (SCAKD 2016)* (pp. 63-74). Moscow, Russia: National Research University Higher School of Economics.

Naidenova, X. A., & Polegaeva, J. G. (1986). An Algorithm of Finding the Best Diagnostic Tests. In G. E. Mintz & P. P. Lorents (Eds.), *The Application of Mathematical Logic Methods* (pp. 63–67). Tallinn, Estonia: Institute of Cybernetics, National Acad. of Sciences of Estonia.

Ore, O. (1944). Galois Connexions. *Transactions of the American Mathematical Society*, 55(1), 493–513. doi:10.1090/S0002-9947-1944-0010555-7

Outrata, J. (2013). A lattice-Free Concept Lattice Update Algorithm based on *CbO. In M. Ojeda-Aciego, & J. Utrata (Eds.), *Proceedings of the 10th International Conference on Concept Lattices and Their Applications (CLA-2013)* (pp. 261-274). La Rochelle, France: University of La Rochelle.

Outrata, J., & Vychodil, V. (2012). Fast algorithm for computing fixpoints of Galois connections induced by object-attribute relational data. *Information Sciences*, 185(1), 114–127. doi:10.1016/j.ins.2011.09.023

Van der Merwe, D., Obiedkov, S., & Kourie, D. (2004). AddIntent: A new incremental algorithm for constructing concept lattices. In *Proceedings of the International Conference on Formal Concept Analysis*, (Vol. 2961, pp. 205-206). Berlin, Germany: Springer. doi:10.1007/978-3-540-24651-0_31

Wray, T., Outrata, J., & Eklund, P. (2016). Scalable performance of FCbO update algorithm on museum data. In M. Huchard & S. Kuznetsov (Eds.), *CLA 2016* (pp. 363–374). Moscow, Russia: National Research University Higher School of Economics.

Zang, L., Zang, H., Shen, X., & Yin, L. (2013). An incremental algorithm for removing object from concept lattice. *Journal of Computer Information Systems*, 9(9), 3363–3372.

Zou, L., Zhang, Z., Long, J., & Zhang, H. (2015). Fast incremental algorithm for deleting objects from concept lattice. *Knowledge-Based Systems*, 89(C), 411–419. doi:10.1016/j.knsys.2015.07.022

ADDITIONAL READING

Agarwal, P., Kaytoue, M., Kuznetsov, S. O., Napoli, A., & Polajillon, G. (2011). Symbolic Galois lattices with pattern structure. In Kuznetsov, S.O. et al. (Eds.), *Rough Sets, Fuzzy Sets, Data Mining and Granular Computing*, Proceeding of the 13th International Conference (vol. 6743 of LNAI, pp.191-198). Berlin Heidelberg: Springer-Verlag.

Bouchachia, A., & Mittemeir, R. (2007). Towards incremental fuzzy classifiers. *Soft Computing*, 11(2), 193–207. doi:10.1007/s00500-006-0077-3

Fong, S., Han, D., & Vasilakos, A. (2015). Comparative study of incremental learning algorithms multidimensional outlier detection on data stream. In M. Usman (Ed.), *Improving knowledge discovery through the integration of data mining techniques* (pp. 54–73). IGI Global. doi:10.4018/978-1-4666-8513-0.ch004

Fong, S., Luo, Z., & Yap, B. W. (2013). Incremental learning algorithms for fast classification in data streams. In *the proceedings of International Symposium on Computational and Business Intelligence (ISCSI)*, (pp. 186-90). Publisher: IEEE. doi:10.1109/ISCBI.2013.45

- Katzouris, N., Artikis, A., & Paliouras, G. (2015). Incremental learning of event definitions with inductive logic programming. *Machine Learning*, 100(2-3), 555–585. doi:10.1007/s10994-015-5512-1
- Kirchberg, M., Leonardi, E., Tan, Y. S., Link, S., Ko, R. K. L., & Lee, B. S. (2012). Formal concept discovery in semantic Web data. In Formal concept Analysis (vol. 7278 of LNCS, pp 164-179). doi:10.1007/978-3-642-29892-9_18
- Kraica, P., Outrata, J., & Vychodil, V. (2012). Computing formal concepts by attribute sorting. *Fundamenta Informaticae*, 115(4), 395–417.
- Kuznetsov, S., & Obiedkov, S. (2002). Comparing performance of algorithms for generating concepts lattices. *Journal of Experimental & Theoretical Artificial Intelligence*, 14(2-3), 189–216. doi:10.1080/09528130210164170
- Kuznetsov, S. O., Kundu, M. K., Mandal, D. P., & Pal, S. K. (Eds.). (2011). Proceedings of the 4th International Conference “*Pattern Recognition and Machine Intelligence*”, (vol. 6744 of LNCS). Springer.
- Naidenova, X. (2011). Constructing Galois lattice in good classification tests mining. In Dmitry I. Ignatov, Sergei O. Kuznetsov, Jonas Poelmans (Eds), *International Workshop on concept discovery in unstructured data* (pp.43-48). Moscow, Russia: the National Research University High School of Economics.
- Obiedkov, S., & Duquenne, V. (2007). Attribute-incremental construction of the canonical implication basis. *Annals of Mathematics and Artificial Intelligence Archive*, 49(1-4), 77–99. doi:10.1007/s10472-007-9057-2
- Outrata, J. (2016). A lattice-free concept lattice update algorithm. *International Journal of General Systems*, 45(2), 211–231. doi:10.1080/03081079.2015.1072928
- Wang, J.-H., Wang, H.-Y., Chen, Y.-L., & Liu, Ch.-M. (2013). A constructive algorithm for unsupervised learning with incremental neural network. *Journal of Applied Research and Technology*, 13(2), 188–196. doi:10.1016/j.jart.2015.06.017

KEY TERMS AND DEFINITIONS

Diagnostic or Classification Test: Assume that we have two sets of objects’ examples called positive and negative examples, respectively. A test for a subset of positive examples is a collection of attributes’ values describing this subset of examples, i.e. it is common or general feature for all examples of this subset and, simultaneously, none of the negative examples is described by it.

Good Classification Test: A classification test describing a given set of positive examples is good if this set of positive examples is maximal in the sense that if we add to it any positive example not belonging to it, then the collection of attributes’ values describing the obtained set will describe at least one negative example.

Good Irredundant Classification Test (GIRT): A good test is irredundant if deleting any attribute’s value from it changes its property “to be test” into the property “not to be a test”.

Good Maximally Redundant Classification Test: (GMRT): A good test is a maximally redundant one if extending it by any attribute’s value not belonging to it changes its property “to be a good test” into the property “to be a test but not a good one”.

Incremental Learning: Incremental learning is a machine learning paradigm where the learning process takes place whenever new example(s) or new attribute(s) (attribute value(s)) merge or must be deleted from dataset and the solutions already obtained are only modified.

The Subtask of the First Kind: Assume that we have two sets of positive and negative examples and a positive example. The subtask of the first

kind is to find all the collections of attributes' values that are included in the description of this example and correspond to the good tests (GMRTs or GIRTs) for the set of positive examples.

The Subtask of the Second Kind: For a given set of positive and negative examples and a non-empty collection of attributes' values such that it does not correspond to a test for the set of positive examples, find all GMRTs (GIRTs) containing it.

A

Machine Dreaming

James Frederic Pagel

University of Colorado School of Medicine, USA

INTRODUCTION

From the anthropomorphic perspective, dreaming is restricted to humans, the only organism with a cognitive processing system able to report what is perceived to be a dream. Dream equivalent processing as based on shared characteristics and human based definitions are, however, within the capacity of modern computer systems. In the human, dreaming and dream-like cognition is utilized in functions with individual and species survival value including: feedback into sleep-associated operative processing, emotional integration (particularly as part of the response to significant physical and/or psychological stress), alternative problem solving, threat avoidance, and creativity (Pagel, 2008; Revonsuo, Tuominen, & Valli, 2015). AI and other machine systems are being developed with the capacity for accomplishing dream-like mentation. This goal is not routinely stated and is in many cases not understood as part of the objective by the programmers and theorists involved in the process. But current predicate logic systems have demonstrated set limitations in their ability to function in the common-sense human environment (Parisi, 2007). Particularly in robotics, an obvious need has developed for systems able to integrate and interact better with humans in the same manner that humans interact with one another. In order for systems to function in the common-sense based human environment better than systems limited to predicate logic, they are likely to require the capacity for achieving human equivalent processing capabilities. As based on many of the defined criteria for dreaming, some current machine systems have been developed utilizing both hardware and software that can function in a dream equivalent manner.

DOI: 10.4018/978-1-5225-2255-3.ch018

BACKGROUND: THE PROBLEM OF DREAM DEFINITION

Dreaming is an almost ubiquitous personal experience, and as such, many definitions for dreaming have been developed and routinely utilized. For each individual experiencing dreaming, the experience seems concrete and obvious. There has been a strong tendency for each individual dreamer to presume that what he or she experiences as a dream is the same for every other individual who dreams. However, what one individual understands to be a dream is often far different from what another experiences or construes to be a dream. This has led to a situation in which there are multiple definitions and no concrete or overall inclusive definition, so that for any group, a series of often different, recurring and set definitions are used for dream (Pagel & Meyers, 2002). This problem of confused and even contradictory definitions has led to significant problems for researchers and investigators in the fields of dream study. For the sleep physician dreams are sleep-associated mental activity. For the psychoanalyst, dreaming defined by bizarre and/or hallucinatory content occurs in both wake and sleep. For one group, dreaming is a state of consciousness. For the other, dreaming is a form or type of thought. In some epistemologies, even the oldest of definitions, such as the dream as a message from god, are still used and believed (Buckley, 2009).

The problem of definition was confounded further in the last decades of the twentieth century when a wide spectrum of neuroscientists and clinicians decided, despite a lack of experimental evidence, that the electrophysiological state of Rapid Eye Movement Sleep (REMS) was equivalent to dreaming. Defined as REMS,

dreaming apparently needed no further definition (Pagel, 2011). Aristotle described a definition as a description of the “essence or essential nature” of the topic and as such, each definition applies and reflects an aspect, an essence of the state (Eco, 1984). In an attempt at clarification, in the year 2001, a multi-specialty panel of dream researchers and therapists developed a multi-axis definition paradigm for dreaming (Table 1) (Pagel, Blagrove, Levin, et. al., 2001).

One approach that can be used to avoid the problems of definition is to limit an approach to the associated and describable characteristics of the state. Researchers avoid defining ‘dream’ and focus on measurable factors (variables) known to be associated with the state such as recall frequency, content, and reported effects on waking behavior (Pagel, 2014). Another approach has been to avoid defining dream by concentrating on the known components of the cognitive process such as the associative memories, emotions and non-perceptual imagery that characterize the dreams state. Computer systems can be utilized to create simulacrum dreamscapes that include these components that comprise a biologic dream. Such artificial dreams are the focus of human filmmakers. The limitations for this process are not technological, they are based on the limits of our current understanding of the state. Computer

systems utilize such capacities at the human interface, and in the presentation of data.

We humans are prone to describing much that is important in our life experience as being ‘like a dream.’ Dreams can be viewed as ‘pictured metaphors’ and as such, these images in metaphor are the stuff of dreams (States, 1997). Metaphor can be used as Aristotle suggested, “to use metaphor well is to discern similarities (to see, and/or analytically get beneath the skin of something).” (Eco, 1984) Yet for dreams, comprised of many and diverse metaphors, the metaphors that we use for specific description are most often misleading, describing only limited aspects of the state. Such metaphors viewed as partial definitions are within the capacity of artificial intelligence (AI) systems.

Some machine based dream equivalents, particularly those based on definitions, are already part of the processing utilized by artificial systems. Some such as Internet based criteria are in the process of assessment, since the system has developed so quickly and with so few controls that it has become somewhat independent and opaque to our understanding. Others, such as the neural-net based systems in current development, give hints of a capacity for dream equivalence, but are able to function, at this point, only at minimalistic levels of consciousness. The interface systems are in some ways the most fascinating, producing shared dream creation and utilization between human and machines.

Table 1. Definitions for dreaming - a classification system paradigm

A Definition of Dream Has Three Characteristic Continua		
Wake/Sleep	Recall	Content
Sleep	No Recall	Awareness
Sleep Onset	Recall	Day-Reflective
Dreamlike States	Content	Imagery
Routine Waking	Associative Content	Narrative
Alert Wake	Written Report	Illogic
	Behavior	Bizarre/ Hallucinatory

Source: Pagel J. Blagrove M. Levin R. et. al., 2001

MACHINE DREAM EQUIVALENTS

Messages From God

Computers routinely access the operative instructions that are part of their programming. Many computer systems are interactive. As based on level of access, computers are far closer to their creators and controllers than most humans. These communications from interactive creators can be viewed from the machine perspective as messages from their gods. So defined, messages from the

creator, accessible programmed information asserting both the capabilities and limits of expected behavior, can be considered as meeting this definition criteria as a form of machine dreaming.

SLEEP ASSOCIATED MENTATION

Dream equivalents have been proposed to occur during sleep and screensaver mode activity in computer systems, functioning as a form of defragmentation and system cleaning (Crick & Mitchenson, 1983). In biological systems, it has been proposed that it is this process that accounts for the perceived “degraded and/or degenerative” (non-logical) aspects of the dream state (Tononi, 2008). Finite state machines, however, are always in one of the two defined states – on or off. This difference between machine sleep (off) and biologic sleep denotes a profound difference between systems, since there is little if any evidence to suggest that an off-state equivalent is possible for any biologic system that is actually alive. Independent, solitary, finite state machines, as currently designed, do not possess the processing ability within their strictly defined states to support a logical analog for the definition of dreaming as sleep-associated mentation.

In software systems such as the Internet, finite state machines are hardware interconnected nodes within an extended system. Such systems maintain a variable, but always on-mode of interconnected data flow, despite individual nodes dropping out or being turned off. Periods of low data flow in multiplex interconnected systems, correspond to human sleep cycles (Peterson & Davie, 2000). During such system-defined sleep periods, these systems operate closer to optimal performance than during periods of high data flow and congestion (wake). Exceedingly complex processing can occur at lower error rates than during congested periods of wake functioning. These periods can be viewed as periods of ‘internet sleep.’ The cognitive processing occurring during these periods meets definition criteria for dreaming on all three axis’s.

BIZARRE AND HALLUCINATORY MENTATION

The psychoanalytic definition for dreaming (bizarre and hallucinatory mentation derived from the unconscious) is today one of the most commonly utilized definitions. As based on this definition, dreaming is defined by its content. Since such mental activity defined as dreaming is defined by content rather than associated state of consciousness, psychoanalytically defined dreaming can occur during waking as well as during sleep (Pagel, Blagrove, Levin et. al., 2001).

Neural network designed processing systems utilize artificial neuron interconnections. Neural net systems can also include non-digital neuron equivalents, sometimes characterized as analog devices (Cowan & Sharp, 1988). Such systems produce a significant proportion of outcomes that can be considered as indeterminate or hallucinatory, hypothetical and non-applicable to the real world situation. Such results can be interpreted as machine dysfunction. As artificial neural density increases the combinations of confused or logically unusable real number stimuli, noise, and partial or incomplete training increases the probability for significant AI machine dysfunction. Such results are analyzed using outside rules and controls and are rejected if they fit outside expected parameters - a form of data interpretation (Padhy, 2005). The “unusable” results of such processing - degraded, indeterminate and hallucinatory - can be viewed based on definition as a form of machine dreaming (Pagel & Kirstein, 2017).

In order to program neural net systems, formalized versions of nonmonotonic reasoning have been developed. Neural net systems programmed in this manner are more likely than symbolic processors to draw conclusions that might be false even when based on premises that are certainly true (Gillies, 1996). Once an appropriate and functional programming language is in place, neural-net AI systems can then be set up and controlled using feedback loops, fuzzy logic, data weight, and repeated application of simple con-

trol rules (Kosko, 1993; Zahed, 1987). Artificial neural networks incorporating on-off processing, multiple connections, multiple levels, and dynamic feedback taking place in temporal sequence or in artificial pseudo-time space can produce neural equivalent processes of associative, multi-level memory, and cognitive feedback, processes that parody aspects of biologic dreaming.

Neural network systems are most often programmed to discard, ignore or deemphasize results that are inconsistent or out of the range of goals set by the programmer. When the obtained results vary markedly from set expectations, the results utilized will be those consistent with expected outcomes, the researchers' beliefs, or the presenter's aesthetic assuring that "outside - the - box" results will not be considered or included in the analysis. This approach clearly limits the systems' capacity to produce alternative responses. The inclusion of such alternative outcomes may be required for a system to develop the capacity to function in what the complex, and often illogically described human world. Such "machine dreaming" based data integration may be required in order to produce creatively useful results from neural network based computing operations.

REM SLEEP

The discovery of REMS came at the end of the psychoanalytic era. Researchers and theorists adopted the perspective, with almost no proof, that dreaming took place only during REM sleep. The association of REMS with dreaming was interpreted by many as evidence that the 'primitive' brain-stem based state of REMS was equivalent to the psychoanalytic wild child, the mythical 'Id,' from which bizarre content rose up from primitive aspects of the mind (activation), before being integrated and modulated by higher cortex operating systems (Pagel, 2011). The supposition that REMS was dreaming was incorporated into widely accepted theories of neu-

roconsciousness including activation-synthesis, activation-integration-modulation (AIM), reverse learning, neural-net theory, search-attention, and protoconsciousness theory (Hobson, 2011). Any physiological activation taking place in the CNS or in the extended body during REMS was viewed as part of dreaming. The dreaming = REMS correlate became philosophically and neuroscientifically indispensable. Unfortunately there never was any actual evidence that REMS was required for the process of dreaming to occur. The abundance of data indicating that dreaming occurs without REMS was ignored and/or disparaged (Foulks, 1993; Solms, 1997; Solms & Turnbull, 2002). It is only recently that most neuroscientists have been forced to surrender their belief in the equivalence of REMS to dreaming and accept that dreaming and REMS are doubly dissociable states. Currently it is questionable as to whether there is any special association between REMS and dreaming (Pagel, 2011). However, fifty years of neuroscientific research focus, the electrophysiology, neurochemistry, neuroanatomy and associated physical correlates of REMS are among the best described of any CNS state.

At this point in time, we have the technical capacity to artificially create almost anything that we can fully describe. The human CNS is composed of 100 billion neurons connected by 100 trillion synapses, each individually controlled by 80,000 genes (3 billion base pairs) affected and interconnected by multiple electrical, ionic, energy, and chemical systems that exert their effects inside and outside neural pathways. Despite such extreme complexity, the creation of an equivalent artificial system is now considered to be within the realm of technical possibility (Markham, 2009). The brain is subdivided into various pathologically and anatomically differentiated parts with each cortical areas constructed of subunits and pools of cells acting in a coordinated way, columns, cell assemblies, micro-circuits, and networks of individual nerve cells each interconnected by axon and dendritic synaptic interfaces. Individual

neurons interconnected with glial and other neural cells form the basic units of this anatomic system at the microscopic level of scale. Each neuron (cell) is subdivided into compartments, synapses, channels, and molecules each with physiological and dynamical properties. Within and outside the cellular nucleus, the genomic systems of DNA and RNA are variably expressed in response to cellular messenger, protein, pH, ATP, electrolyte, and electrical frequency changes. Extending from the organ brain to the level of DNA coding, the series of integrated levels in the mammal CNS have been postulated to include at least seventeen levels of interactive complexity (Markham, 2014).

CNS complexity extends well beyond the anatomy. The brain's spatial structure functions on a wide and intermingled range of scales in environmental and temporal interaction. Temporally, the electrical, chemical, genetic, and endocrine dynamics of the brain span many orders of magnitude. Synaptic, magnetic, genetic, and electrical processes occur and alter within a sub-millisecond to millisecond range. Perceptual and thought processes happen in fractions of seconds to hours, and memory phenomena last for seconds to many years. Circadian and ultra circadian rhythms relate brain function to the external environment, including the effects of solar and lunar light and tidal cycles. Even slower processes affect the CNS on genetic, molecular, and evolutionary time scales. Each process, each CNS function, is embedded in a network of temporal and structural interactions and dependencies. And each biological system is also a physical system with the characteristics of a liquid, a gas, or a solid, each composed of an interactive atomic and sub-atomic structure.

Despite this level of complexity, major attempts are currently underway to further analyze and then create artificial CNS constructs. The National Institutes of Health have sponsored an initiative called the Human Connectome Project designed to trace and map three-dimensionally brain neural connections (Kurzweil, 2012). At the Allen institute in Seattle, datasets with cel-

lular level spatial resolution are being used to produce cellular transcriptomic profiles of specific genetic expression in defined neuroanatomic regions (Sualin & Hohmann, 2007). Neuromorphic chips have been developed called SyNAPSE each designed to simulate the quarter-million synaptic connections of approximately 250 neurons – with the goal, creation of a simulated neo-cortex with 10 billion neurons (chips) and 100 trillion connections (IBM, 2011). The European Union has invested major research funds into an even more ambitious project through the École Polytechnique Fédérale de Lausanne (EPFL) where Henry Markham leads an approach designed to produce a supercomputer based on multilevel neural network architecture. The initial goal of the project, the simulation of a rat neocortical column, considered by some researchers to be the smallest functional unit of the neocortex, was completed in 2006. In rats, each column is about 2 mm in length, has a diameter of 0.5 mm, containing approximately 10,000 neurons (10^8 synapses). The ultimate goal of this project to allow for the parallel simulation of large numbers of connected columns in order to reconstitute a whole neocortex (in humans about 1 million cortical columns) (Graham-Rowe, 2012). At EPFL, the first-draft digital reconstruction of the microcircuitry of the somatosensory cortex of juvenile rat used cellular and synaptic organizing principles to algorithmically reconstruct detailed anatomy and physiology in a neocortical volume of 0.29 ± 0.01 mm³ containing ~31,000 neurons with 55 layer-specific morphological and 207 morpho-electrical neuron subtypes. This network of digitally reconstructed neurons included ~8 million connections with ~37 million synapses. Physiological electrical fields applied to this system produce a spectrum of network states with a sharp transition from synchronous to asynchronous activity (Markham, Muller, Ramaswamy et al., 2015). Markham has predicted that his laboratory will complete such a system with complexity equivalent to the human CNS within the next ten years (Markham, 2009)

It is potential that such neuroanatomic Zombie neural network systems, and future “perfect” Zombie systems that will incorporate electrophysiological and neurochemical aspects of CNS operations, will have the capacity to parody or approximate physiologic CNS states. Some neuroscience theorists suggest that due to the similarity in their structure and there extreme levels of integrated complexity, such systems may have the capacity for human-equivalent consciousness (Balduzzi & Tononi, 2008; Koch & Tononi, 2008). These systems might even approximate REMS - the disenfranchised, yet persistently believed correlate of dream. Lucid dreaming, conscious awareness and volitional control during dream consciousness, is most often considered as associated with REMS. It is possible that the development of such a pattern of machine lucidity would mark the development of human-equivalent or supra-human consciousness in such a system (Bostrom, 2014). The ethical considerations involved in our creation of such systems are apparently profound.

METAPHOR

Wish fulfillment, a very loosely Freudian construct (dream marriages, dream vacations, dream homes and dream cars) is the most generally utilized definition for dream. We are prone to describing much that is important in our life experience as being ‘like a dream.’ Dreams can be viewed as “pictured metaphors.” Such images in metaphor are the stuff of dreams (States, 1997). The Internet has become a repository where we search for patterns and attempt to consummate our dreams. Some of this search we define, but much of what we define is based on the structure of the Internet, our access systems, the alignment of the search engine that we use, and the slant of the programming applied by outside sources. While hardware systems may lack capacity for the human-like experience required for the creation of human-like dreams, the Internet provides a system data-base of dream metaphor exponentially larger than that

available to any individual human. As based on access to such a database of human dreams, it is well within the capacity of an AI search engine to create dream-equivalent metaphors.

DREAM STRUCTURE: ARTIFICIAL CORRELATES

Almost all dreams include images, emotions and associative memories (Pagel, 2014). Computer systems integrate and analyze digitalized data - words, narrative, emotions, and visual images included in a human based data set. Machine-based imagery constructed out of human and machine based data bits often has a dream-like phenomenology: reflecting waking experience (continuity), and producing a visually unique presentation that is complex, constantly changing, and can be perceived as emotional.

Humans in their art, they’re writing, and their films have concentrated on developing the ability to create artificial dreams. Filmmaking is a process that lends itself to artificial simulation. In our viewing or reading of such constructed dreams, we interject our own memories, emotions and imagery into the experience. Once enmeshed, the filmmaker can entice us into a vicarious experience of mental images that resemble less a dream and more the vivacity of actual experience. This capability has already been integrated in to computer interfaces (ex. IBM’s Watson). Such machine integrated presentations that incorporate and organize experience into humanly assessable formats will be required in order to further develop interface interactions.

MACHINE DREAMS AT THE HUMAN INTERFACE

Many of us spend much of our waking time within a bi-directional machine interface. The interfaced human and the attached machine both become something larger, smarter, and potentially more

interesting than either system when disconnected. After spending waking time interacting with a machine interface, our waking thoughts, our sleep, and our dream contents are altered. (Gackenback & Kurville, 2013). New interface systems under current development will include non-sensory modalities such as frequency-based physiologic electrical fields, extending the capacities of the interface to cognitive systems that more often utilized in meditation, sleep and dream states than in waking perceptual focused attention (Lakatos, Karmos, Mehta, Ulbert, Schroede, 2008; Pagel, 2012). The interface is very likely to extend into shared machine dream equivalents, incorporating memories, visions and emotions into the available data stream.

FUTURE RESEARCH DIRECTIONS

Current computer systems have the ability to represent and process human metaphor, bringing together and interconnecting human-based associated visual and text-based memories into stories that have the characteristics of dreaming mentation. Neural-net systems have the capacity to go even further, bring indeterminate and/or hallucinatory results into the resultant response paradigm. Such results even look like dreamscapes (Google, 2016). It remains debatable as to whether the neuroanatomic and perfect zombie systems that are being created will have the capacity for human-like consciousness. If such a breakthrough does occur, the development of human-like consciousness will likely require that such zombies have a human-like capacity to dream. At that point the machine might potentially enter a state of consciousness in which it attains and exerts control of potential outcomes, controlling its own story. This can be viewed as the attainment of the machine equivalent of lucid dreaming: a state of dream-like consciousness from which the system is able to independently change and control data flow. There will be little question as to whether such a system is conscious.

CONCLUSION

System disconnected FSM's built with predicate, concrete logic programming, such as the typical personal computer, have only minimal capacities for achieving dream-like forms of cognitive processing. But as based on definition, many artificial systems can meet some of the human-based criteria for dreaming (Table 1). The Internet offers access to an expansive database of human dream metaphors and memories. Cloud based operating allows for extension beyond the FSM into a processing approach in which the overall operating system is never turned off including periods of high and low data flow (system sleep). The extension of processing capacity into areas of non-parametric, approximate or fuzzy logic processing, extends system capacity beyond concrete logic applications, and increases the possibility of alternative and/or unexpected outcomes. The addition of an AI self-teaching (learning) capacity to such a system and the utilization of a neural-net hard and software formats for that processing means that a much higher percentage of results attained will be either indeterminate, hallucinatory, or unexpected. The machine, pushed to capacity in its attempts to utilize and integrate all available data in attempting to at least approximate our set goals, will be pushed to utilize all pertinent (and even non-pertinent) data. Bi-directionally, such machine dreams will be incorporated into the dream consciousness of interacting humans.

While we are not consciously attempting to extend computer capacity for attaining dream-like mentation, complex machine systems are incorporating active sleep modes, alternative logics, non-parametric programming, and extended interfaces into their processing capabilities. These approaches are being utilized in an effort to extend their capacity for human interaction and function into common-sense human interactions. In that somewhat illogical environment, modern AI systems are required to approach their external environment in similar ways to those used by humans. To function in that environment, even artificial systems will require the capacity to dream.

REFERENCES

- Balduzzi, D., & Tononi, G. (2008). Integrated Information in Discrete Dynamical Systems: Motivation and Theoretical Framework. *PLoS Computational Biology*, 4(6), e1000091. doi:10.1371/journal.pcbi.1000091 PMID:18551165
- Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford, UK: Oxford University Press.
- Buckley, K. (2009). *Dreaming and the World's Religions*. New York: New York University Press.
- Cowan, J., & Sharp, D. (1988). Neural Nets and Artificial Intelligence. In *The Artificial Intelligence Debate: False Starts, Real Foundations* (pp. 85–121). Cambridge, MA: MIT Press.
- Crick, F., & Mitchenson, G. (1983). The function of dream sleep. *Nature*, 304(5922), 111–114. doi:10.1038/304111a0 PMID:6866101
- Eco, U. (1984). Aristotle's Rhetoric as translated and interpreted by Umberto Eco. In *Semiotics and the Philosophy of Language*. Bloomington, IN: Indiana University Press. doi:10.1007/978-1-349-17338-9
- Foulks, D. (1993). Data constraints on theorizing about dream function. In A. Moffit, M. Kramer, & R. Hoffmann (Eds.), *The Functions of Dreaming*. Albany, NY: SUNY Press.
- Gackenback, J., & Kurville, B. (2013). Cognitive structure associated with the lucid features of gamers dreams. *Dreaming*, 23(4), 256–267. doi:10.1037/a0034817
- Gillies, D. (1996). *Artificial Intelligence and Scientific Method*. Oxford, UK: Oxford University Press.
- Google. (2016). *Deep Dream*. Retrieved from <https://en.wikipedia.org/wiki/DeepDream>
- Graham-Rowe, D. (2005, June). Mission to build a simulated brain begins. *New Scientist*.
- Hobson, J. (2011). *Dream Life - An Experimental Metaphor*. Cambridge, MA: MIT Press.
- IBM. (2011). *IBM Unveils Cognitive Computing Chips*. Retrieved from <http://www-03.ibm.com/press/us/en/pressrelease/35251.wss>
- Koch, C., & Tononi, G. (2008). Can Machines Be Conscious? *IEEE Spectrum*, 45(6), 55–59. doi:10.1109/MSPEC.2008.4531463
- Kosko, B. (1993). *Fuzzy Thinking: The New Science of Fuzzy Logic*. New York: Hyperion.
- Kurzweil, R. (2012). *How to Create a Mind: The Secret of Human Thought Revealed*. New York: Penguin.
- Lakatos, P., Karmos, G., Mehta, A., Ulbert, I., & Schroeder, C. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science*, 320(5872), 110–113. doi:10.1126/science.1154735 PMID:18388295
- Markham H. (2009). *A brain in a supercomputer*. TEDGlobal.
- Markham, H. (2014). *Toward a Science of Consciousness*. Tucson, AZ: Center for Consciousness Studies.
- Markham, H., Muller, E., Ramasuany, S., & Reimann, M. (2015). Reconstruction and simulation of neocortical microcircuitry. *Cell*, 163(2), 456–492. doi:10.1016/j.cell.2015.09.029 PMID:26451489
- Padhy, N. P. (2005). *Artificial Intelligence and Intelligent Systems*. New Delhi, India: Oxford.
- Pagel, J.F. (2001). Defining Dreaming – A paradigm for comparing disciplinary specific definitions of dream. *Dreaming*, 11(4), 195–202. doi:10.1023/A:1012240307661
- Pagel, J. F. (2008). *The Limits of Dream: A Scientific Exploration of the Mind/brain Interface*. Oxford, UK: Academic Press. doi:10.1016/B978-012374215-5.50033-7
- Pagel, J. F. (2011). REMS and Dreaming – Historical Perspectives. In *Rapid Eye Movement Sleep – Regulation and Function*. Cambridge, UK: Cambridge University Press.

Pagel, J. F. (2012). The synchronous electrophysiology of conscious states. *Dreaming*, 22(3), 173–191. doi:10.1037/a0029659

Pagel, J. F. (2014). *Dream Science: Exploring the Forms of Consciousness*. Oxford, UK: Academic Press.

Pagel, J.F., & Kirstein, P. (2017). *The Dreams of Artificial Intelligence - Machine Dreaming and Consciousness*. Academic Press.

Pagel, J. F., & Myers, P. (2002). Definitions of Dreaming: A Comparison of Definitions of Dreaming Utilized by Different Study Populations (College Psychology Students, Sleep Lab Patients, And Medical Professionals). *Sleep*, 25, A299–A300.

Parisi, D. (2007). Mental Robotics. In A. Chella & R. Manzotti (Eds.), *Artificial consciousness*. Charlottesville, VA: Imprint Academic.

Peterson, L., & Davie, B. (2000). *Computer Networks: A Systems Approach* (2nd ed.). San Francisco: Morgan Kaufman Publishers.

Revonsuo, A., Tuominen, J., & Valli, K. (2015). The avatars in the machine: Dreaming as a simulation of social reality. *Open Mind*, Frankfurt am Main, Germany.

Solms, M. (1997). *The Neuropsychology of Dreams - A Clinico-Anatomical Study*. Mahwah, NJ: Lawrence Erlbaum Associates.

Solms, M., & Turnbull, O. (2002). *The Brain and the Inner World: An Introduction to the Neuroscience of Subjective Experience*. New York: Other Press.

States, B. (1997). *Seeing in the Dark: Reflections on Dreams and Dreaming*. New Haven, CT: Yale University Press.

Sualin, S., & Hohmann, J. (2007). Insights from spatially mapped gene expression in the mouse brain. *Human Molecular Genetics*, 16(R2), R209–R219. doi:10.1093/hmg/ddm183 PMID:17911164

Tononi, G. (2008). Consciousness as Integrated Information: A Provisional Manifesto. *The Biological Bulletin*, 21(3), 216–242. doi:10.2307/25470707 PMID:19098144

Zaheh, L. (1987). *Fuzzy Sets and Applications: Selected Papers*. New York: Wiley.

ADDITIONAL READING

Domhoff, G. W. (2003). *The Scientific Study of Dreams: Neural Networks, Cognitive Development, and Content Analysis*. Washington, DC: American Psychological Association. doi:10.1037/10463-000

Gackenbach, J. (Ed.). (2016In Press). *Boundaries of Self and Reality On Line: Implications of Digitally Constructed Realities*. IGI.

Kramer, M., & Glucksman, M. (2015). *Dream Research: Contributions to Clinical Practice*. New York: Routledge.

Mallick, B. N., Pandi-Perumal, S. R., McCarley, R. W., & Morrison, A. R. (Eds.), *Rapid Eye Movement Sleep—Regulation and Function*. Cambridge: Cambridge University Press.

Pagel, J. F. (2008). *The Limits of Dream: A Scientific Exploration of the Mind/brain Interface*. Oxford, UK: Academic Press. doi:10.1016/B978-012374215-5.50033-7

Pagel, J. F. (2012). The synchronous electrophysiology of conscious states. *Dreaming*, 22(3), 173–191. doi:10.1037/a0029659

Pagel, J. F. (2014). *Dream Science: Exploring the Forms of Consciousness*. Oxford, UK: Academic Press.

Pagel JF, Kirstein P. (2017) *The Dreams of Artificial Intelligence - Machine Dreaming and Consciousness*, Academic Press (In Publication).

KEY TERMS AND DEFINITIONS

Anthropomorphic: Ascribing human forms or attributes to a being or thing not human, especially to a deity.

Artificial Intelligence: A computer system with capacities for self-learning.

Consciousness (AI Definition): The ability to rise above programming.

Ethics: A system of moral principles.

Metaphor: A term or phrase used to represent something else.

Singularity: The point at which a computer system will attain and move on beyond human capabilities.

A

Moving Object Detection and Tracking Based on the Contour Extraction and Centroid Representation

Naveenkumar M

National Institute of Technology Trichy, India

Sriharsha K. V.

National Institute of Technology Trichy, India

Vadivel A

National Institute of Technology Trichy, India

INTRODUCTION

Whenever human observe a scene from a video, objects in scene despite being static keep on changing or moving rapidly with in a snapshot of time. Under such circumstances, detecting and tracking of objects in a scene is absolutely essential. However, active research has been progressing in this specialized field for the last three decades and as a result several algorithms have evolved for object detection and tracking. A robust, accurate and high performance approach is still a great challenge today. The difficulty level of this problem highly depends on how the object to be detected is defined and tracked. For a critical evaluation, frame differencing algorithm is chosen for object detection and Kalman filter is chosen for object tracking. Kalman filter algorithm is successfully implemented for tracking single objects with and without occlusions. In each case, tracking efficiency is determined with error covariance estimation. It is found that as error covariance tends to zero and goes to negative, the covariance of the predicted state becomes low and uncertainty is reduced. It is proved that Kalman filter is the best estimator and subsequently the experimentation is extended to multiple objects without occlusions and successful results are revealed.

BACKGROUND

The active research on object detection is going on for past few decades. The general approach for object detection is background subtraction, also known as foreground detection. Background subtraction is a challenging task, especially in complex dynamic scenes that might contain moving trees, rippling water, etc. Many approaches are presented in literature to deal with background subtraction (Elgammal 2014; Bouwmans 2011; Brutzer et al., 2011; Bouwmans et al., 2008). Huapeng Yua et al. (2014) proposed object detection using contour level top down information. Recently Kooij et al. (2015) proposed a novel approach for multiple object detection using their appearance. It identifies the minimal set of objects from a frame, from which the features are extracted. It focuses mainly on appearance rather than temporal cues. In tracking by detection method, target objects are detected and these are combined with consistent tracks. For example, human detectors can be trained on Histogram of Gradient (HoG) features or by analyzing about spatial occupancy to explain background/foreground masks (Ben Shitrit et al., 2011; Fleuret et al., 2008; Liem and Gavrilu, 2011). In terms of tracking, online trackers on a frame by frame basis can be differentiated (Breitenstein et al., 2011; Kim et al., 2012; Yan et al., 2012).

Procedure CECR (video)

```

while (end of video) do
    // stage 1: pre processing
    // two successive frames (i, i+1) are reading
        read (frame i)
        read (frame i+1)
    // frame i, i+1 are converted to gray
        grayframe1 = color2gray(frame i)
        grayframe2 = color2gray(frame i+1)
    // absolute difference is calculated
        absdiff_frame = absdiff(grayframe1,grayframe2)
    // gray threshold technique is applied
        binary_frame = threshold(absdiff_frame)
        normalized_box_filter(binary_frame)
        binary_frame = threshold(binary_frame)
    // stage2: Segmentation using Contour Extraction
    // (Suzuki 85)
        contours = getContours(binary_frame)
        no_of_contours = size(contours)
        drawAllContours(contours, contours_frame)
        centroid[i]= findCentroid(threshold_frame)
        generateTrajectory(centroid)
    end while
end CECR

```

OBJECT TRACKING USING CONTOUR EXTRACTION AND CENTROID REPRESENTATION

A primitive experiment is conducted on single object moving with a constant speed and no occlusions. In this perspective a series of videos captured using Nikon COOLPIX 12.0 mega pixels are used for motion analysis. For moving object identification, frame differencing technique is chosen. In this technique, absolute difference between two successive frames i and $i+1$ is calculated. For each of the differenced image that is obtained in each of the successive iteration, a grey threshold is calculated and is applied. As a result the differenced images are transformed into binary images. To smooth the binary image, normalized box filter is applied and then again grey threshold is calculated. As a result, the perfect binary image

is produced. For each of the moving object that is identified in the preprocessed image, a centroid is computed. This centroid represents the moving object in each of the differenced image. Finally a trajectory is drawn by connecting the centroids for all the differenced images. The results obtained after testing the algorithm in a video is presented for discussion. The pseudo code for object detection and tracking using Contour Extraction and Centroid Representation (CECR) is shown below.

In the preprocessing stage, two consecutive frames, say frame i and frame $i+1$, are read from the video. These frames are converted into gray format (grayframe1, grayframe2). The absolute difference is calculated between grayframe1 and grayframe2 and result is stored in absdiff_frame. absdiff_frame is converted into binary format using gray thresholding technique and result is stored in binary_frame. It is smoothed using normal-

Procedure *CECR_Kalman* (video)

1. Find foreground objects using CECR algorithm.
2. Kalman Filter object is Created with initial detected location (centroid value of initial frame).
3. The tracks are initialized. (track contains centroid, bounding box, id, age, kalman filter and total visible count).
4. New locations of existing tracks are predicted.
5. Update track details based on the new detections.
6. Delete the last tracks and create new tracks for next frame.

Procedure *PredictNewLocation* (initial track)

Covariance of predicted state vector (psv) $=A*P*A'+Q$

Where

A: State transformation matrix

P: State covariance matrix

Q: System noise

Kalman gain matrix (K)

$$K = psv * H^T * (H P H^T + R(t_i))^{-1}$$

where H- observation matrix

R-measurement noise

correction factor $\Delta(x_i)$

$$\Delta x(t_i) = K(t_i) \cdot (l(t_i) - (t_i)), \quad // \text{ measurement residual}$$

$$x^+(t_i) = x^-(t_i) + \Delta x(t_i)$$

$$\text{Error covariance } \Delta P(t_i) = E[\Delta x(t_i) \Delta x(t_i)^T]$$

ized box filter and then again gray thresholding is applied to get final threshold frame in binary format. The preprocessing result on a video from KTH data set (Schüldt et al., 2004) is shown in Figure 2. We have used OpenCV implementation (OpenCV 2.4.10) of contour extraction algorithm based on (Suzuki 85) to segment the binary frame. This generates a collection of external contours.

EXTENDED CECR USING KALMAN FILTER

We have extended the CECR algorithm using Kalman filter and it is shown below. The Kalman filter is a recursive estimator. It not only estimates the

state from the previous time step and the current measurement is needed to compute the estimate for the current state.

As explained presented above, foreground objects are detected using CECR algorithm and the motion of each track is estimated using CECR Kalman Algorithm. For motion analysis, a foreground object is represented as centroid and is used as a measurement for tracking. In order to predict the new location of the existing track, state vector $x(t_i)$, covariance matrix $(P(t))$, Kalman gain(K) and correction factor is calculated. Based on the new locations error covariance is estimated and then track details are updated. The algorithms for *PredictNewLocation* and *UpdateTrackDetails* are:

Figure 1. Sample images from KTH dataset a) Walking b) Jogging c) Running d) Boxing



Figure 2. Preprocessing result of CECR (video from KTH dataset) a) i th frame b) $i+1$ th frame c) Difference image d) Threshold image e) Blurred image f) final threshold image

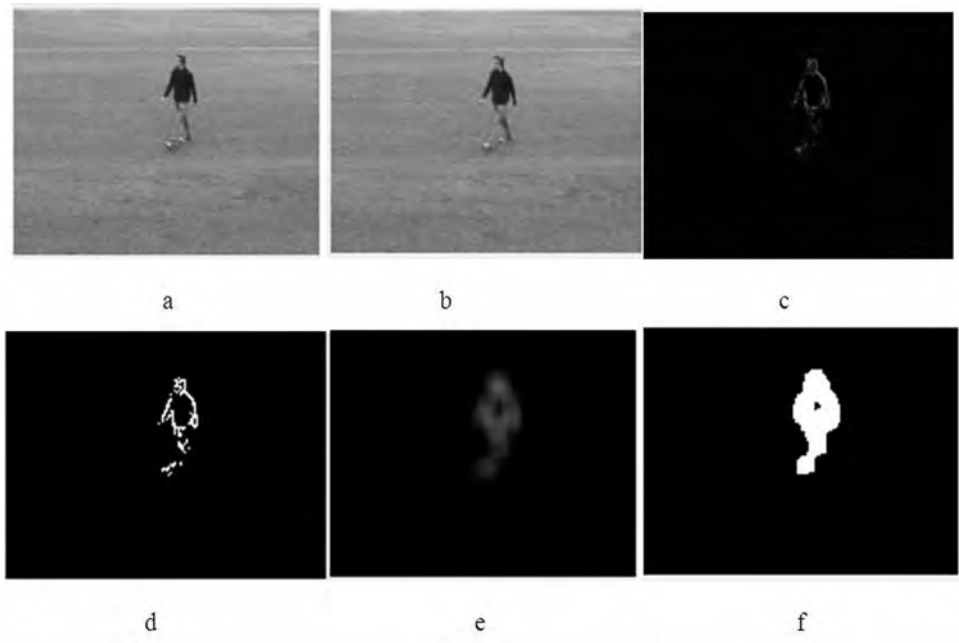


Figure 3. Detection results of CECR a) video frame from KTH b) Video frame captured using Smart Phone



SOLUTIONS AND RECOMMENDATIONS

To study the performance of the proposed algorithm, experiments are conducted on KTH data-

set. Furthermore, experiments are conducted on various video sequences captured using Samsung Galaxy S3 smart phone.

Procedure *UpdateTrackDetails* (*ppre*, ΔP)

Updated covariance of the predicted state *p*

$$p(t_i) == ppre(t_i) + \Delta P(t_i)$$

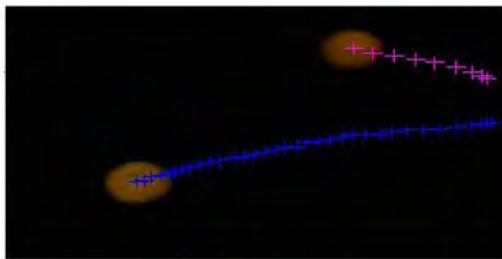
$$\text{where } \Delta P(t_i) = (1 - K(t_i)) * H * ppre(t_i)$$

H is an observation matrix

Figure 4. Object Tracking results using using CECR_Kalman (a) Multiple human object tracking results (b) Multiple non human object tracking results (c) Human object tracking results with occlusions



(a)



(b)



(c)

KTH Dataset

It contains videos of six action classes (walking, jogging, running, boxing, hand waving, and hand clapping). Some of the sample frames are shown in Figure 1.

Multiple Object Tracking Results (Using CECR_Kalman)

To evaluate the performance of the proposed algorithm, experimentation is conducted on various scenarios as shown below.

In the above all scenarios, the graph between the frame and error covariance is drawn as shown in Figure 5 and the graph between the error covariance and covariance of predicted state is drawn as shown in Figure 6.

Variation of error covariance ΔP over a sequence of frames is depicted in the above Figure 5. The results show that ΔP is approximating to zero from the negative. It is interpreted from the above graph as the ΔP approaches to zero, covariance of predicted state (*P*) becomes low and hence the uncertainty will be reduced.

FUTURE RESEARCH DIRECTIONS

The work is likely to be extended for tracking of moving objects in non-stationary environment in case of partial and/or full occlusions, illumination changes and also in cluttered background. There is a possibility of extending the work in identifying and tracking the location of stationary and moving objects in 2D /3D space based on acoustic waves and visual ability i.e. with complex interaction of light, eyes and brain.

CONCLUSION

In order to study in depth the essence of existing target tracking algorithms developed based on their mathematical and environment of their applications, complex situations like partial and full

A

Figure 5. Graph between the frame and error covariance

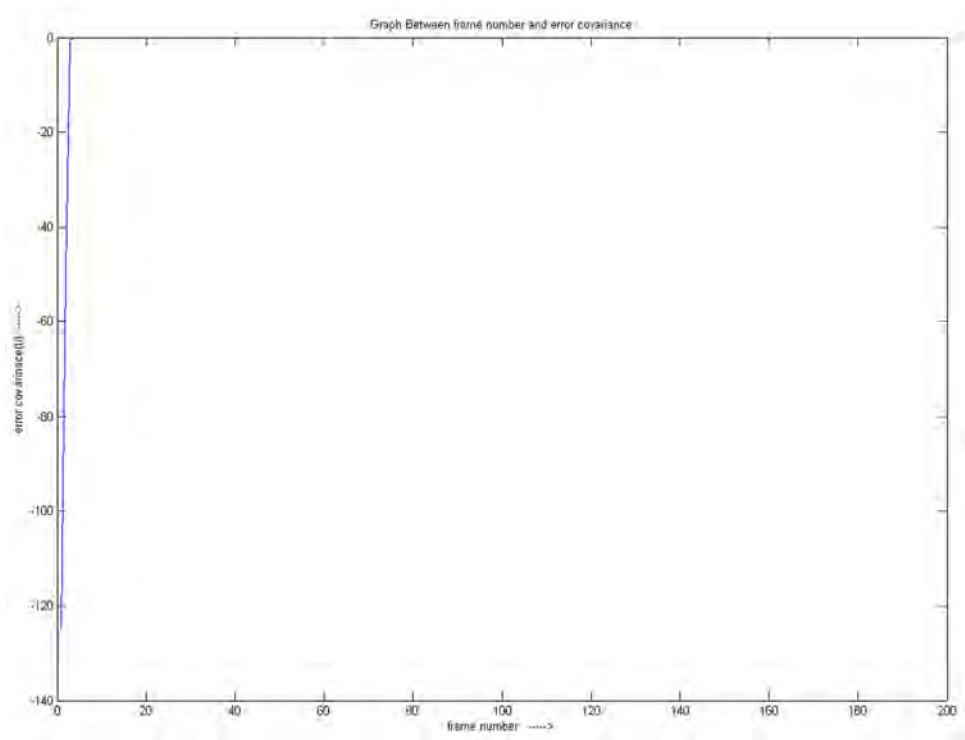
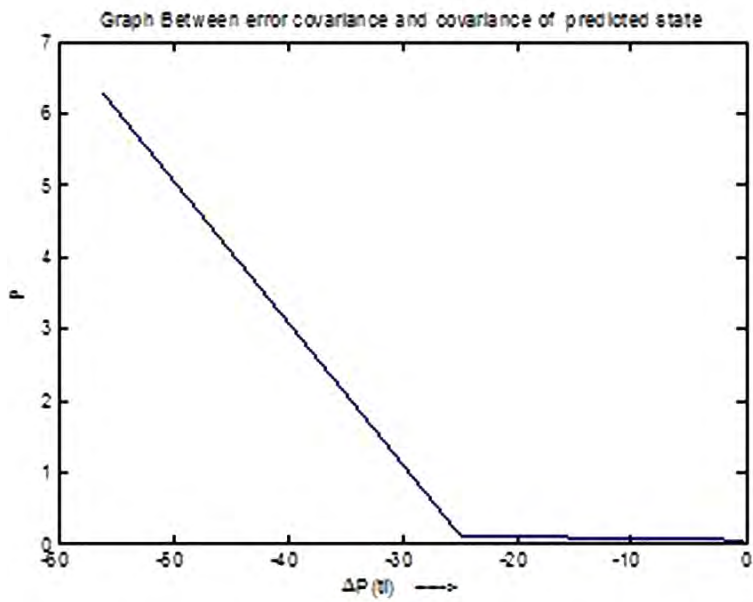


Figure 6. Graph between the error covariance and covariance of predicted state



occlusion are chosen and are successfully implemented. A critical evaluation has been made in measuring uncertainties in the tracking scenarios with error covariance parameter and it is proved from the experimental observations that Kalman filter is the best estimator in object location prediction in a video sequence.

REFERENCES

- Ben Shitrit, H., Berclaz, J., Fleuret, F., & Fua, P. (2011, November). Tracking multiple people under global appearance constraints. In *Computer Vision (ICCV), 2011 IEEE International Conference on* (pp. 137-144). IEEE. doi:10.1109/ICCV.2011.6126235
- Bouwman, T. (2011). Recent advanced statistical background modeling for foreground detection-a systematic survey. *Recent Patents on Computer Science*, 4(3), 147–176.
- Bouwman, T., El Baf, F., & Vachon, B. (2008). Background modeling using mixture of gaussians for foreground detection-a survey. *Recent Patents on Computer Science*, 1(3), 219–237. doi:10.2174/2213275910801030219
- Breitenstein, M. D., Reichlin, F., Leibe, B., Koller-Meier, E., & Van Gool, L. (2011). Online multiperson tracking-by-detection from a single, uncalibrated camera. *Pattern Analysis and Machine Intelligence. IEEE Transactions on*, 33(9), 1820–1833.
- Brutzer, S., Höferlin, B., & Heidemann, G. (2011, June). Evaluation of background subtraction techniques for video surveillance. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on* (pp. 1937-1944). IEEE. doi:10.1109/CVPR.2011.5995508
- Elgammal, A. (2014). Background subtraction: Theory and practice. *Synthesis Lectures on Computer Vision*, 5(1), 1–83. doi:10.2200/S00613ED-1V01Y201411COV006
- Fleuret, F., Berclaz, J., Lengagne, R., & Fua, P. (2008). Multicamera people tracking with a probabilistic occupancy map. *Pattern Analysis and Machine Intelligence. IEEE Transactions on*, 30(2), 267–282.
- Kim, S., Kwak, S., Feyereisl, J., & Han, B. (2012). Online multi-target tracking by large margin structured learning. In *Computer Vision-ACCV 2012* (pp. 98–111). Springer Berlin Heidelberg.
- Kooij, J. F., Englebienne, G., & Gavrilu, D. M. (2015). Identifying multiple objects from their appearance in inaccurate detections. *Computer Vision and Image Understanding*, 136, 103–116. doi:10.1016/j.cviu.2015.03.012
- KTH Action Data Set. (n.d.). Retrieved from <http://www.nada.kth.se/cvap/actions>
- Liem, M., & Gavrilu, D. M. (2011). Multi-person localization and track assignment in overlapping camera views. In *Pattern Recognition* (pp. 173–183). Springer Berlin Heidelberg. doi:10.1007/978-3-642-23123-0_18
- Schüldt, C., Laptev, I., & Caputo, B. (2004, August). Recognizing human actions: a local SVM approach. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on* (Vol. 3, pp. 32-36). IEEE. doi:10.1109/ICPR.2004.1334462
- Yan, X., Wu, X., Kakadiaris, I. A., & Shah, S. K. (2012). To track or to detect? an ensemble framework for optimal selection. In *Computer Vision-ECCV 2012* (pp. 594–607). Springer Berlin Heidelberg. doi:10.1007/978-3-642-33715-4_43
- Yu, H., Chang, Y., Lu, P., Xu, Z., Fu, C., & Wang, Y. (2014). Contour level object detection with top-down information. *Optik-International Journal for Light and Electron Optics*, 125(11), 2708–2712. doi:10.1016/j.ijleo.2013.11.029

ADDITIONAL READING

Challa, S. (2011). *Fundamentals of object tracking*. Cambridge University Press. doi:10.1017/CBO9780511975837

Forsyth Ponce. (2002). *Computer Vision A Modern Approach*. Prentice Hall Professional Technical Reference.

Jain, R., Kasturi, R., & Schunck, B. G. (1995). *Machine vision* (Vol. 5). New York: McGraw-Hill.

KEY TERMS AND DEFINITIONS

Binary Image: An image whose pixels have only two possible intensity values. They are normally displayed as black and white.

Correction Step: Uses the current value of the estimate to refine the result given by the predictor step.

Covariance Measure: Covariance measures the degree to which two variables change or vary together (i.e. co-vary).

Error Covariance Matrix (ΔP): Defines the expectation of the square of the deviation of the state vector estimate from the true value of the state vector.

Grayscale Images (or Gray Level): Are simply whose colors are the shades of gray. Each pixel in a gray scale image is represented only with single intensity value (stored as 8 bit integer value).

Image Filtering: It is the process of transforming pixel intensity values to reveal certain image characteristics like image enhancement, smoothing technique and template matching.

Kalman Filter: A recursive predictive filter used for estimating Kalman filter is used to estimate the state of a linear system where the state is assumed to be distributed by a Gaussian.

Kalman Gain Matrix (K): Determine the weighting of the measurement information in updating the state estimates.

Measurement Noise (R): Associated with the measurement vector, describes the statistics of the noise on the measurements.

Measurement Residue: Gives difference in measurement between the true state vector and the estimated state vector.

Measurement Vector: It is a set of simultaneous measurements of properties of the system which are functions of the state vector.

Object Detection: Identifying an object over a sequence of frames in a video.

Object Representation: In a tracking scenario, an object can be defined as anything that is of interest for further analysis.

Object Tracking: It is defined as the problem of estimating the trajectory of an object in an image plane as it moves around a scene.

Observation Matrix: It is a measure of how dependent the measurements are upon the state of the system.

Posterior State: State during time span (t_{i+1} to t).

Prediction Step: Calculates the next estimate of the state based only on past measurements of the output.

Prior State: State during time span (t_{i-1} to t).

State Transformation Matrix (A): It is an approximation of the change that the state undergoes over the specified time interval.

State Vector: Set of parameters describing a system, known as states, which the Kalman filter estimates.

System Noise (Q): Determines the variation in the true values of the states.

Thresholding: Separates the regions of the image corresponding to objects in which we are interested, from the regions of the image that correspond to background.

Semantic Intelligence

Maria K. Koleva

Institute of Catalysis, Bulgarian Academy of Sciences, Bulgaria

INTRODUCTION

Semantic intelligence is an exclusive for the intelligent complex systems way of response to non-specified and non-predetermined ever-changing environment subject to the mild assumption of boundedness alone. Its major distinctive property is autonomous comprehension and creation of information. Another exclusive property of semantic intelligence is autonomous discrimination between true and false statements.

The semantic intelligence naturally arises in the setting of concept of boundedness where it commences from highly non-trivial interplay between structural and functional properties of a complex system. The major generic property of that interplay is that it renders the relation between structural and functional properties to be non-recursive. In result this renders the efficiency of semantic intelligence to be provided by efficiency of hierarchical self-organization of a complex system prior to the efficiency of software and to the speed of hardware.

Concept of boundedness is a new explanatory paradigm aimed towards explanation of the behavior of the complex systems behavior. A systematic study can be found in (Koleva, 2013; Koleva, 2016). In the setting of this paradigm the semantic intelligence is implemented by a specific for each intelligent complex system hierarchical self-organization of the physical processes. The major generic property of this hierarchy extensively discussed in the chapter, renders it bi-directional and non-extensive.

The concept of boundedness is the first systematic theory which considers the behavior of complex systems in non-specified non-predetermined

ever-changing environment subject to the mild assumption of boundedness alone. The major assumption is that a non-predetermined bounded environment is the starting point for fundamentally novel properties of complex systems behavior, semantic intelligence included, which are not available for the traditional scientific approach where the environment is supposed set apriori. It is should be stressed that the traditional approach is grounded on the assumption that the environment is statistically or deterministically pre-determined and unbounded and hence all physical, chemical, biological and engineered systems are considered as set in specific pre-determined environment. The fundamental discrimination between both approaches consists of the fact that the behavior of corresponding systems, semantic intelligence included, is describable by non-recursive means while the behavior of the algorithmic intelligence is describable by recursive means only.

The goal of this chapter is to highlight the major properties of semantic intelligence that naturally arise under the concept of boundedness and to demonstrate their relation to the general characteristics of physical processes which implement it. This defines the strategic goal to be establishing the grounding principles of next generation approach for building a circuit able to exhibit semantic intelligence.

Since the semantic intelligence naturally arises in the frame of concept of boundedness, the most general assumptions of the concept of boundedness are presented. The necessary conditions for setting semantic intelligence are presented later. This is made because the intelligent complex systems are a sub-class of the family of complex systems. As such they share all properties of the wider class

of complex systems along with specific for the semantic intelligence properties. A comparison between the exclusive properties of semantic and algorithmic intelligence is made.

BACKGROUND

Before presenting the basic principles of the concept of boundedness let us remind the notion of a complex system and the major characteristics of their behavior. Let us start with the notion of a complex system: it is how parts of a system are organized so that the system behaves as a single object and how it interacts with its environment. Thus the notion of a complex system encompasses an enormous variety of systems ranging from physical ones such as quasar pulsations, to biological such as DNA sequences, to social ones such as financial time series.

The intensive empirical examination that was going on in the last decades displays the remarkable enigma of complex systems behavior: the highly specific for each complex system properties persistently coexist with certain universal, shared by each of them ones. Thus on the one hand, they all share the same characteristics such as power law distribution and sensitivity to environmental variations, for example; on the other hand, each system has its unique “face”, i.e. one can distinguish between an earthquake and heartbeat of a mammal. What makes the study of this coexistence so important is the enormous diversity of systems where it has been established. In order to get an idea about this vast ubiquity let us present a brief list of such phenomena: earthquakes, traffic noise, heartbeat of mammals, public opinion, currency exchange rate, electrical current, chemical reactions, weather, ant colonies, DNA sequences, telecommunications, etc.

But the greatest mystery enshrined in the behavior of complex systems is that both intelligent and non-intelligent systems belong to the same class: thus a Beethoven symphony, a product of a genius mind, and the traffic noise which,

though being also a product of human activity, but un-intelligent in its behavior, share the same type of power spectrum. Another example is the semantics of human languages: in the year 1935 the linguist G. K. Zipf established that, given some corpus of natural languages, the frequency of any word is inversely proportional to its rank in the frequency table. Thus the most frequent word will occur approximately twice often as the second most frequent word, three times as often as the third most frequent word, etc. Thus, the Zipf law ignores any semantic meaning and thus it seems to sweep out the difference between mind activity and random sequences of letters. Thus we come to the following fundamental problem: what makes a complex system “intelligent” and why it should share such “indifferent” to the intelligence properties?

The affiliation of the apparently intelligent systems such as human languages and music to the same class as earthquakes and a variety of other natural phenomena, suggests in a straightforward way that the intelligent behavior is embedded in natural processes. Thus the opposition between intelligence as highly specific activity and the fact that it inherently belongs to a universal class of natural phenomena raises the major question whether it is likely to expect defining a criterion able to distinguish the specificity of a system from another one along with affiliating each of them to the same class of complex systems.

CONCEPT OF BOUNDEDNESS – GENERAL PRINCIPLES

Next the grounding assumptions of the concept of boundedness are introduced. They are:

- A complex system remains stable if and only if the rate and amplitude of variations that it exerts in response to an ever-changing environment are bounded to specific for the system margins.

- The response is local and not pre-determined. This implies that it depends on the current state of a system and the current environment impact. Yet, in order to sustain boundedness of rates it behaves in a non-linear and non-homogeneous way.
- Complex systems are self-organized in a hierarchy of responses so that different hierarchical levels are linked through inter-level feedbacks. An exclusive property of the inter-level feedbacks is that each of them operates as a bounded irregular environment for the self-organization at each and every hierarchical level. In turn this provides bi-directionality of the hierarchy as it goes both bottoms up and top down. This is in sharp contrast with the traditional approach where different objects are considered closed systems and the hierarchy goes only bottom up.

A central for the boundedness result is the proof that there exists a presentation basis where the response decomposes into two parts, specific and universal one, each of which has characteristics that are robust to environmental variations. The rigorous assertion is that the power spectrum of each time series that represents the behavior of a complex system is decomposable into two parts: a specific discrete band and a continuous band whose shape is universal. Further, the claim is that the decomposition is additive and it happens with constant in the time accuracy.

In addition to these two parts in the power spectrum a non-recursive component persists at finite distance from both of them. It commences from the highly non-trivial interplay between the specific and the universal part. Since each time series which represents a record of the behavior of a complex system is bounded irregular sequence, it could be presented as wild and permanent bounded “twisting” and “winding up” around a “skeleton” built on the discrete band of the power spectrum. Thus, the motion resembles the mo-

tion on a torus with irrational frequencies. In the present case the motion never stops and thus the extra-line associated with irregular motion around the skeleton appears always as non-recursive regardless to the relations among the thresholds over the corresponding and the other scales in the corresponding system.

It should be stressed on the fact that non-recursive component persists only for bounded variations since, if variations are unbounded, the non-recursive component is shifted away to the infinity at the left of a power spectrum. Thus, the boundedness turns out to be the crucial factor in setting the non-recursivity of relation structure-functionality which in turn serves as the major distinctive implement for semantic intelligence. As consequence, the semantic intelligence turns out to be an exclusive for the concept of boundedness way of response to an ever-changing environment.

The greatest value of that decomposition is:

1. It allows unambiguous separation of an object from its environment. The notion of an object consists of specific pattern, called homeostasis, because its characteristics remain intact in an ever-changing environment. Moreover, one can define them regardless to the details of that environment.
2. Although the exact local behavior remains unpredictable, it provides predictability of the behavior of each complex system up to the predictability of its current homeostatic pattern.
3. The persistence presence of a non-recursive component constitutes the non-recursive way of relating structural and functional properties. Indeed, while structural properties are defined through the specific pattern whose power spectrum is given by the discrete band, the functional properties are presented through the entire power spectrum, i.e. through the interplay among specific, noise and non-recursive component altogether.

SEMANTIC INTELLIGENCE AS A SUB-CLASS OF COMPLEX SYSTEM BEHAVIOR

Next the necessary conditions for the realization of the semantic intelligence are considered. It should be stressed that the class of intelligent complex systems is a sub-class of complex systems. Thus, all complex systems are subject to boundedness but the intelligent complex systems are subjected to an additional constraint which will be presented next.

An exclusive property of any complex system considered in the frame of boundedness is that its state space is partitioned into domains so that each domain is characterized by a specific for it homeostatic pattern; the rate of intra-domain motion is bounded and irregular. In result the motion among domains is restricted to the adjacent domains only. The boundedness of the inter-domain motion renders the state space trajectories to be non-random. This is so because any trajectory in a state space goes in such a way that it jumps only to adjacent states. The allowed inter-domain motion is confined only to the adjacent domains next to any given domain. The selection of any of the adjacent domains is random one since in each run it depends on the current intra-domain trajectory which is irregular.

The general necessary constraint which provides implementation of the semantic intelligence by physical processes consists of requirement that the state space of any intelligent complex system is to be partitioned into at least 4 domains. The reason behind the particular choice of no less than 4 domains will become clear later.

The utilization of this constraint is two-fold:

- The first assumption is to associate information symbols with intra-domain homeostatic patterns. The far-going consequence of that association consists of providing algorithmic uniqueness of each and every specific law. The algorithmic uniqueness is achieved through the ubiquitous presence of a non-recursive component additional to the homeostatic and noise characteris-

tics. To remind, a “non-recursive” implies that neither specific law can be achieved by finite number of steps organized in an algorithm. The non-recursive component comes from a highly non-trivial interplay between the homeostatic pattern and inter-level feedbacks for every intelligent complex system as explained in the previous section. In turn it makes different specific laws unique and algorithmically unreachable from one another. It should be stressed that the comparison with the traditional explanatory paradigm displays a sharp contrast. Indeed, the traditional explanatory approach assumes existence of a universal law such that all specific laws are its algorithmic derivatives.

- The second assumption consists of association of the meaning of each semantic unit with the performance of a specific non-mechanical engine built on the corresponding inter-domain orbit in the state space. As an example of non-mechanical engines may be considered the bio-chemical cycles in living organisms. The exclusive property of presentation of the meaning of a semantic unit through performance of an engine is that the latter provides sensitivity to permutations of the semantics. Indeed, like the semantics of human languages where the meaning of each word depends on the order of the letters in it, an engine performs differently if it operates in different directions. Thus, for example the famous Carnot engine operates in one direction as a pump and in the opposite direction it performs as refrigerator.

Thus, each semantic response consists of hierarchy of cycles which follow through adjacent states only irrespectively to the intensity of the current environmental impact. In turn, this provides the most fundamental property of the semantic intelligence to be the autonomous creation and comprehension of information.

Thus, there are two ways of presenting a semantic unit: the first one is through the sequence of information symbols which it consists of while the second one is through performance of a specific engine. This two-fold presentation of every semantic unit is the major implement for non-extensivity and bi-directionality of the hierarchy of every semantic response. It should be stressed that both presentations are algorithmically unreachable from one another.

SECOND LAW AND SEMANTIC INTELLIGENCE

The next point is the relation between physical processes which implement semantic intelligence and their environment. The importance of this question is rendered by the fact that in the traditional algorithmic approach no hardware is in balance with its environment. This is so because it is assumed that any hardware can execute any algorithm. In turn this opens the door to an uncontrolled “drainage” of the environment or its uncontrolled heating. Thus a question arises: can one built information perpetuum mobile? The answer is negative and it consists of reformulation of the Second Law so that its availability to encompass the widest scope of systems.

A clue for the most general formulation of the Second Law in the frame of boundedness is the need for its availability to semantic intelligence. Since the latter is executed by means of performance of non-mechanical engines, an appropriate formulation should be grounded on their generic properties. The author have proved that under boundedness the efficiency of a non-mechanical engine never exceeds the efficiency of corresponding Carnot engine where the engine is free from necessity of a physical coupling to two heat reservoirs. Alongside, the proof is free from the condition for entropy maximization viewed as condition for reaching equilibrium. Thus the proof substantiates the most ubiquitous

formulation of the Second Law to impose ban over perpetuum mobile.

It should be stressed that the present formulation of the Second Law justifies the general condition about the necessity of at least 4 domains with different homeostasis for implementation of semantic intelligence. Indeed, suppose that a cycle consists of a trajectory that goes through 3 domains only. Then the corresponding engine performs all possible permutations in a single cycle by means of running it in opposite directions and starting at different states. Thus the words ‘dog’ and ‘god’ and words ‘on’ and ‘no’ are product of the same engine but run in opposite direction. Therefore, it is obvious that only orbits which go through at least 4 states provide different cycles for different permutations. Further, the work produced by different engines is a measure for the efforts exerted behind any semantics. Since the engines are specific, the corresponding work is also specific for each semantic response. In turn this puts the hierarchical organization of the semantic response prior to the fact whether it is fastest or most efficient. This fact is opposite to the traditional algorithmic approach where the fastest computers and the most efficient algorithms are most valued.

It is worth noting the sharp contrast with the traditional approach. A generic property of the state space in the traditional approach is the Markovianity. It implies that successive jumps are not correlated and the trajectories never close. As a result no semantic intelligence is ever possible in the frame of the traditional approach.

The most fundamental property of the semantic intelligence, the autonomous creation and comprehension of information comes at a price: unlike a Turing machine which is able to execute every algorithm, the semantic intelligence executes only algorithms compatible with its “hardware”. However, this is rather an advantage than a setback because it serves as grounds for autonomous discrimination of a true and a false statement by semantic intelligence.

Then, the distinctive properties of both algorithmic approach and semantic intelligence make them rather counterparts than opponents.

Summarizing, the additional constraint to which intelligent complex systems are subject is that its state space must consist of at least 4 domains. This condition is satisfied for spatially heterogeneous systems only. The state space of homogeneous systems has only 2 domains. Put in other words, only systems which are self-organized in patterns or are subject to morphogenesis can exhibit semantic intelligence. Yet, at this point a question arises: the necessary condition for heterogeneity is to have 3 domains in the state space. Then, to what generic type of pattern formation the requirement of about at least 4 domains corresponds?!

A great advantage of the semantic intelligence is that it is in balance with the environment. This fact is an immediate consequence of the assumption about considering semantic intelligence as a sub-class of complex systems. Then any piece of semantics is executed as a sequence of information symbols. A generic property of this sequence is that, being a bounded irregular sequence, its variance is always finite (Koleva, 2013).

MAJOR DIFFERENCES BETWEEN SEMANTIC AND ALGORITHMIC INTELLIGENCE

Algorithmic intelligence is grounded on the rules of formal logic. It is widespread since it serves as grounds for the software of all computers. It is worth noting that the formal logic presupposes execution of any algorithm by recursive means only. Alongside, the compatibility between software and hardware implies that the latter operates by means of linear processes only. In turn, this provides the most exclusive for the algorithmic intelligence property, which is the property that any algorithm can be executed by any hardware. However, this property comes at a price which is

best seen through the differences with the semantic intelligence. Next the major of them come:

- Algorithmic intelligence is not able to create or comprehend any information in an autonomous way; semantic intelligence comprehends and creates information in an autonomous way. Algorithmic intelligence can execute any algorithm by any hardware because the relation software-hardware is expressed by recursive means only. On the contrary, the non-recursive relations between structural and functional properties of intelligent complex systems render a specific compatibility between software and hardware.
- Algorithmic intelligence assumes acquiring of absolute knowledge about each system by means of obtaining definite answers to yes-no questions through cleverly posed experiments. This assumption is grounded on the idea that relation impact-response is always recursive. On the contrary, the setting of boundedness puts at the center preservation of the current homeostasis. A central for the boundedness property is that the response of each complex system decomposes into a specific and universal part whose properties are robust to environmental changes. The specific information about any system is concentrated into the specific part called homeostasis. Then, any specific knowledge about a system comes at the expense of presence of accompanying noise whose intensity cannot be made arbitrarily small. In turn, this makes obtaining of absolute knowledge impossible.
- Algorithmic intelligence assumes generation of information out of noise processes. On the contrary semantic intelligence puts a ban over creating information out of noise. The semantic intelligence can be executed only by means of going through changes into homeostasis. Thus the meaning of a semantic unit is performed as a

non-mechanical engine whose working cycle goes through at least 4 states of different homeostasis. Then, it is obvious that acquiring information out of noise alone is impossible. This setting is in sharp contrast with the traditional algorithmic approach where it is assumed possible to obtain information from noise.

- Algorithmic intelligence assumes that the only relation between a semantic meaning and the sequence of symbols that constitutes it is expressed by recursive means only. In turn this justifies the major goal of algorithmic theory to be search for fastest computer and most efficient algorithms. A crucial notion of the concept of boundedness is the implementation of semantic meaning through performance of a non-mechanical engine. It is important to stress that it utilizes the fact that the latter is algorithmically irreducible to the sequence of symbols that constitute it. In turn this renders the leading role of hierarchical super-structuring of intelligent complex systems prior to the efficiency of software and the speed of the hardware.
- An important consequence of the presentation of semantic intelligence as a sequence of semantic units is that renders possible autonomous comprehension of true and false statements. A 'true' statement corresponds to that unique trajectory, which has been physically realized in a given run; a 'false' statement is the one that follows an imaginary trajectory that comprises a non-admissible state; a counterfactual corresponds to an admissible trajectory which has not been realized in a given run. And last but not least, an improbable statement corresponds to an imaginary impasse: the statement does not correspond to any trajectory. Thus, the automatic selection of trajectories to admissible and non-admissible renders the fundamental difference of the present approach from the traditional

algorithmic theory and the formal logic which are grounded on the assumption that all trajectories are available. Let us provide an example: the sentences "the Sun rises at east" and "the Sun rises at west" are both true in the setting of algorithmic theory because they are both grammatically correct; on the other hand in the setting of boundedness the first sentence "the Sun rises at east" is true since it corresponds to a physically realizable trajectory while the sentence "the Sun rises at west" is false because it corresponds to a physically unrealizable trajectory.

FUTURE RESEARCH DIRECTIONS

The semantic intelligence puts a bridge over physics, biology and artificial intelligence. As a corollary of the development of this novel theoretical framework one of the major directions in the future research is establishing the principles of the next generation strategy aimed towards building of a circuit capable to autonomous comprehension and creation of information. One of the key questions about this strategy is whether such circuits are capable to evolution.

The grounds for such question come from the exclusive for the theory of boundedness property that is the non-recursive relation between structure and functionality which thus serves as major implement for novel route to evolution different from the survival of the fittest. The non-recursive relation between structure and functionality sets the major advantage of that route to evolution to be the best hierarchical organization irrespectively whether is the fastest or the least energy consuming because its exclusive property is that it is in balance with its environment. And since this hierarchy is bi-directional, it acts as specific promotor to evolution: only compatible with the entire hierarchy changes are allowed. On the contrary, the traditional assumption about recursive relation between structural and functional prop-

erties implies that the only route to evolution is survival of the fittest: the best construction provide the best performance.

Thus, in the frame of the theory of boundedness it is to be expected appearing of the two routes to evolution as counterparts rather than as opponents. The question that stands now is how to utilize in the best way the interplay between both routes to evolution for achieving sustainable development of artificially created circuits operating by means of semantic intelligence?

CONCLUSION

The properties of semantic intelligence, viewed as an exclusive for intelligent complex systems response to non-specified and non-predetermined ever-changing environment, are considered in the present paper. The semantic intelligence naturally arises in the frame of recently introduced concept of boundedness. An exclusive for the concept of boundedness property is that the relation between structural and functional properties of complex systems is non-recursive on the contrary to the traditional science where the relation between structural and functional properties is supposed recursive one. In turn, this renders the most efficient hierarchical organization to become more important for the functioning of a complex system than the efficiency of the software and the speed of the hardware. The best hierarchical organization provides multi-purpose usage of specific for every system relations between structure and functionality. The major difference with the traditional engineering is that the multi-purpose usage of this relation is subject to the following constraint: since the semantic intelligence is in balance with the environment, only compatible with the entire hierarchy changes are allowed. On the contrary, the recursive relations between structure and functionality render that the most effective structure automatically provides the best performance.

The non-recursive relation between structure and functionality serves as grounds for the most exclusive property of semantic intelligence, i.e. the ability for autonomous creation and comprehension of information. Another exclusive for the semantic intelligence property is that it automatically distinguishes between true and false statements. These properties provide major difference with Turing machines. Turing machines are able to execute any algorithm by means of any hardware but neither of them is able to autonomously comprehend the output and to distinguish whether the latter is true or false. On the contrary, the execution of semantic intelligence requires compatibility between software and hardware but the gain is autonomous comprehension of the output and autonomous distinguishability of its value, i.e. the semantic intelligence automatically distinguishes whether it is true or false. In turn, this makes semantic intelligence and Turing machines rather counterparts than opponents.

The systematic development of the properties of semantic intelligence in the frame of the concept of boundedness defines the strategic goal of the future research to be establishing the grounding principles of next generation approach for building a circuit able to exhibit semantic intelligence. One of the key questions about this strategy is whether such circuits are capable to evolution. The grounds for such question come from the exclusive for the theory of boundedness property that is the non-recursive relation between structure and functionality which thus serves as major implement for novel route to evolution different from the survival of the fittest. The non-recursive relation between structure and functionality sets the major advantage of that route to evolution to be the best hierarchical organization irrespectively whether is the fastest or the least energy consuming because its exclusive property is that it is in balance with its environment. And since this hierarchy is bi-directional, it acts as specific promotor to evolution: only compatible with the entire hierarchy changes are allowed. On the contrary, the traditional assumption about recursive

relation between structural and functional properties implies that the only route to evolution is survival of the fittest: the best construction provide the best performance.

Thus, in the frame of the theory of boundedness it is to be expected appearing of the two routes to evolution as counterparts rather than as opponents. The question that stands now is how to utilize in the best way the interplay between both routes to evolution for achieving sustainable development of artificially created circuits operating by means of semantic intelligence?

REFERENCES

Koleva, M. (2016). *Boundedness and Applied Self-Organized Semantics*. Retrieved from <http://www.igi-global.com/video.aspx?ref=boundedness-applied-self-organized-semantics&titleid=137626>

Koleva, M. K. (2013). *Boundedness and Self-Organized Semantics: Theory and Applications*. Hershey, PA: IGI Global. doi:10.4018/978-1-4666-2202-9.ch001

KEY TERMS AND DEFINITIONS

Algorithmic Intelligence: Exclusive for the formal logic way of “response” to an input.

Concept of Boundedness: A new explanatory paradigm, put forward by the author, aimed towards explanation of the behavior of the complex systems behavior.

Non-Recursive Relation: Relation that cannot be expressed in a closed form or by means of finite step algorithms.

Second Law: One of the most fundamental laws in physics.

Semantic Intelligence: Semantic intelligence is an exclusive for the intelligent complex systems way of response to non-predetermined and non-specified ever-changing environment. Its major distinctive property is autonomous comprehension and creation of information. Another exclusive property of semantic intelligence is autonomous discrimination between true and false statements.

Survival of the Fittest: A theory of evolution grounded on the assumption that for any given change in the environment the best suited structure of an object provides the greatest advantage in performance. This assumption is explicit outcome of the assumption that the relations between structure and functionality are expressed by recursive means only.

The Summers and Winters of Artificial Intelligence

A

Tad Gonsalves

Sophia University, Japan

INTRODUCTION

Humanity has toyed with the idea of building intelligent machines from antiquity. The ancient Egyptians, Greeks, and the Chinese are reported to have built mechanical devices that imitated human intelligence, although primitively. In the 16th century, it was Hobbes who stated that thinking was symbolic reasoning like working out an answer with pen and paper. This idea of thinking as symbolic reasoning was further developed by Descartes, Pascal, Spinoza, and Leibniz (Poole & Mackworth, 2010). The English mathematician Charles Babbage, who designed the Analytical Engine in 1837, is considered to be the inventor of the modern computers. Although the Analytical Engine was not built until 1991, it served as a prototype of the modern general-purpose computer. The first link to machines and intelligence is found in the 1950 paper published by Alan Turing, “Computing Machinery and Intelligence” (Turing, 1950). The “Turing test” proposed in this paper, is a simple pragmatic approach assuming that a computer that is indistinguishable from an intelligent human actually shows that machines can think. Its relevance to this day seems to indicate that it will be a goal for the AI field for many years to come (McGuire, 2006).

The term “Artificial Intelligence” was coined by John McCarthy in a conference organized in 1956 in Dartmouth (McCarthy, 1955). He defines *intelligence* as the computational part of the ability to achieve goals in the world and *Artificial Intelligence* (AI) as the science and engineering of making intelligent machines, especially intelligent computer programs (McCarthy, 2001).

AI has had a tumultuous history (McCorduck, 1979; Crevier, 1993). The early theorem proving and checkers playing programs astonished the public with their seemingly intelligent performance. The expert systems, too, drew great admiration from the general public as well as the specialists, since their competence and performance in consultation was in par with human experts. This chapter introduces DENDRAL and MYCIN, two of the most successful expert systems developed in the childhood of AI.

However, the hype and expectations raised by these early systems led to the downfall of AI in the ensuing years. Perceptrons, for example, were just picking up as promising machine learning mechanisms when they were nipped in the bud by other AI researchers owing to a premature evaluation. Their apparent inability in evaluating the XOR and the NXOR functions was later rectified by the introduction of an additional middle layer of neurons. However, this came much later, and enough harm had already been done to the Artificial Neural Network research.

The ALPAC report in 1966 and the Lighthill report in 1973 were responsible for the onset of an all-time low-fund and low-activity period, called the AI winter. The two fatal reports not only led to the cutting of funds, but a general distrust in AI research. The recent victories of AI programs over human champions in chess and in Jeopardy have restored faith in AI discipline. AI has run its course through summers and long winters. True, we are far from an overall theory or framework, explaining what thinking is and how to build intelligent machines (Lungarella, 2007). Nevertheless, AI has made great strides and, after overcoming

Figure 1. The AI hype curve



the *trough of disillusionment* is marching on the *plateau of productivity*.

BACKGROUND

As it happens to any new technology, the history of AI, too, ran through a hype curve (Menzies, 2003). The early AI programs like the ones elegantly proving theorems and skillfully playing board games aroused great interest and expectations. This was followed by the successful application of Expert Systems in business and academia. This early period in the development of AI is referred to as the “peak of inflated expectations” shown in the AI hype curve (Figure 1).

However, for the next ten years AI did not live up to its expectations. Not getting any substantial returns, the AI investors and stakeholders appointed special committees to inquire in the progress of AI. This period called the “trough of disillusionment” proved fatal to the development of AI. The summers and the winters experienced in the history of AI are tersely expressed in the words of Tim Menzies (Menzies, 2003):

In the 21st century, AI has many reasons to be proud, but it wasn't always this way. New technologies such as AI typically follow the hype curve. By the mid-1980s, early successes with expert systems caused skyrocketing attendance at AI conferences and a huge boom in North American AI startups.

Just like the dot-coms in the late 1990s, this AI boom was characterized by unrealistic expectations. When the boom went bust, the field fell into a trough of disillusionment that Americans call the AI Winter. A similar disillusionment had already struck earlier, elsewhere.

The funds were heavily cut and AI research almost came to a grinding halt. Despite the setbacks, some AI researchers persisted with their research in a few selected areas, refraining from calling their work “AI”. Their approach finally led to the “slope of enlightenment” bringing AI out of the harsh winters. In the 1990s AI research began to flower again culminating in the victory of IBM’s Deep Blue over the reigning chess world champion. Yet another milestone heralding the “plateau of productivity” was IBM’s Watson’s victory over the Jeopardy world champions in the year 2011.

THE HEY DAYS OF EXPERT SYSTEMS

An Expert System is a program designed to solve problems at a level comparable to that of a human expert in a given domain (Cooper, 1989). Also known as Knowledge-based system, an Expert System relies on the expert human knowledge captured in a computer program to solve problems that ordinarily require human expertise. It

has a vast store of knowledge in the knowledge-base and uses some kind of reasoning process to infer solutions to a given problem. The problems solved by an Expert System are restricted to their respective domain field. Accordingly, the Expert Systems do not display general intelligence like human beings, but can be immensely skillful in analyzing and solving problems in specialized domains (Duda, & Buchanan, 1982; Giarratano, & Riley, 1998; Nikolopoulos, 1997; Rajendra, & Sajja, 2010). Expert Systems do not use abstruse numerical reasoning. Rather, they follow human-like rules of thought, and can answer why and how questions reasonably well (Sashikumar, 2006; Schalkoff, 2011). This is because, in addition to the knowledge base and the inference module, they are also equipped with an explanation module.

The following sub-sections describe two of the highly successful Expert Systems, DENDRAL and MYCIN developed in the mid 60's.

DENDRAL

DENDRAL (acronym coined from DENDritic ALgorithm) was the earliest expert system developed in 1965 at Stanford University by a team of computer scientists, organic chemists and geneticist(s). It was a chemical-analysis expert system, programmed in LISP considered to be *the* AI language of that period (Boban, 2013). DENDRAL was designed to function in the domain of organic chemistry and mass spectrometry. The chemical substance to be analyzed had its spectrographic data fed into the system, and the system would hypothesize the substance's molecular structure (Lindsay et al., 1980). For example, the molecular weight of hydrogen and oxygen is 1 and 16, respectively. Therefore, the mass spectrum of a water molecule (H_2O) formed from two molecules of hydrogen and one molecule of oxygen has a peak at 18 units. DENDRAL would use this input mass and the knowledge of atomic mass numbers and valence rules, to determine the possible combinations of atomic constituents whose mass would add up to 18.

DENDRAL followed a typical plan-generate-test method to arrive at a solution to the given problem. The generator module would generate potential solutions for a particular problem, which were then expressed as chemical graphs. Faced with a large numbers of possible solutions, Dendral had to find a way to put constraints that would rule out many candidate solutions. This was the function of the planner module, which was a hypothesis-formation program. Finally, the tester module would analyze each proposed candidate solution and discarded those that failed to fulfill certain criteria (Lindsay et al., 1980).

DENDRAL was a convergence of three different technologies: knowledge engineering, machine learning and heuristics programming. A heuristic is a *rule of thumb* that human experts use intuitively in solving a problem. Heuristic programming models the expert heuristics in programming. In particular, they are used to introduce short-cuts to reduce the search through the huge knowledge base. Heuristics programming was a major approach and a giant step forward in Artificial Intelligence, as it allowed scientists to finally automate certain traits of human intelligence (Lindsay et al., 1980).

DENDRAL's performance rivaled that of a chemist expert at this task. However, this was not because the program knew more than an expert. It performed well because of its systematic search through the space of possibilities and its systematic use of what it knew. Although, it did receive a good coverage in chemical journals, its actual use in industry and academia was somewhat limited.

MYCIN

MYCIN, as the name suggest was an early medical expert system developed in the early 1970s at Stanford University (Buchanan & Shortliffe, 1984). MYCIN's main task was to diagnose infectious blood diseases and recommend antibiotics as a doctor in real-life would do. Although the inferences were based on a knowledge-base containing a modest 500 rules, MYCIN's success was

Table 1. XOR function

Input x_1	Input x_2	Output
0	0	0
0	1	1
1	0	1
1	1	0

impressive. Apart from diagnosing the bacterial infections and prescribing medication, it would also give an explanation and its reasoning procedure to the patient at the end, which increased its credibility.

A typical rule in MYCIN has an antecedent and a consequent. We refer to the antecedent of a rule as the premise or left-hand side (LHS) and to the consequent as the action or right-hand side (RHS). The inference can either take the forward chaining or the backward chaining mechanism. Forward chaining is also called data-directed inference, because the data that are known drive the inferences from left to right in rules, with rules chaining together to deduce a conclusion.

MYCIN primarily used backward chaining, or a goal-directed control strategy. The deductive validity of the argument is established in the same way, but the system's behavior is very different. In goal-directed reasoning a system starts with a statement of the goal to achieve and works backward through inference rules, i.e., from right to left, to find the data that establish that goal (Buchanan, & Shortliffe, 1984).

In several tests conducted by the Stanford medical school, MYCIN was found to perform better than the average physicians in diagnosis and prescription. However, due to legal and ethical issues it was never used commercially.

A BLOW TO CONNECTIONISM

The field of Artificial Neural Networks (ANN) is currently a budding field of AI. It is conventionally used as a de facto methodology in machine

Table 2. NXOR function

Input x_1	Input x_2	Output
0	0	1
0	1	0
1	0	0
1	1	1

learning and pattern recognition. Although the research in ANN began in 1940s, it had a severe setback in the late 60s. This section outlines the development of ANN and the cause of its decline in the late 60s.

Neurons modeled as binary threshold activation functions (McCulloch & Pitts, 1943) were successful in evaluating the basic logic gate functions. Rosenblatt further developed the basic neuron model and called it *perceptron* (Rosenblatt, 1962). The perceptron had weights assigned to the inputs and learning consisted in adjusting the connection weights to produce appropriate outputs. In the training phase, inputs are provided to the input layer of the perceptrons and the outputs are compared to the known outputs of the data-sets used for learning. The internal weights of the perceptrons are adjusted in iterative controlled steps until the error between the known output and that predicted by the perceptron is reduced below a desired level. This positive feedback enables the perceptrons to learn the patterns that map the input to the output. The perceptrons, in general, serve as linear threshold functions.

The perceptron, however, had some limitations. For example, it could not solve the exclusive OR function (XOR) (Table 1) and the exclusive NOR (NXOR) function (Table 2).

Minsky and Papert (1968, 1988) showed that the perceptron could only solve linearly separable functions and argued that the representations learned were inadequate for intelligent action. Their ill-timed remark led to the decline in the perceptron research until about the 1980's. Later researchers overcame the limitations of the Rosenblatt perceptron model by creating neural networks

with hidden layers. These networks connect the neurons of the input layer to those in the output layer via the neurons in the hidden layer. Learning involves adjusting all the connection weights given an input dataset with known output. Although the present day ANN have grown considerably more complex, they still stem from the early advances of the McCulloch-Pitts neuron and the Rosenblatt perceptron.

TWO FATAL REPORTS

Progress in science depends to a great extent not only on intellectual creativity but also financial support. It is widely known in the history of AI that it suffered drastic funding setback that led to its downfall. Two major reports were responsible for this downfall. The first was the ALPAC Report of 1960, the second the Lighthill Report some ten years later. The former was launched in the USA, and virtually stopped machine translation funding on both sides of the Atlantic. The latter originated in the UK and caused significant damage there – “with shock waves spreading across the Pond” (Boden, 2006).

The ALPAC Report

In the 1950s, the climate for machine translation was very favorable. Optimism reigned in the 1952 conference on machine translation. Many universities in the US, UK and Russia started major machine translation projects. At Georgetown University, in particular, devoted researchers began building a pilot system to convince potential funding agencies of the feasibility and the practicability of machine translation. This led in 1954 to the famous Georgetown experiment, a pilot system translating from Russian to English, which was hailed as an unqualified success - during the next ten years over 20 million dollars were invested in machine translation by various US government agencies (King, 1984).

As it happens in all new technology projects, there was an initial hype, but soon the AI researchers faced grand challenges in natural language processing and in machine translation (MT), in particular. In 1964, owing to the lack of appreciable results, the US government appointed a committee to evaluate the MT research. This “Automatic Language Processing Advisory Committee” (ALPAC), compiled a report in 1966 (Pierce, 1966) which proved fatal to MT researchers. The report stated that MT was slower, less accurate and twice as expensive as human translation and that there was no immediate or predictable prospect of useful machine translation. MT research funds were soon suspended and for years after that the message given to the general public was that MT was utterly hopeless (Hutchins, 1996; 2005).

The Lighthill Report

In 1973, the UK Science Research Council commissioned Sir James Lighthill, the FRS Lucasian Professor of Applied Mathematics, Cambridge University, to evaluate the state of AI research in the UK. The Lighthill report (Lighthill, 1973) stated straightforward that AI had simply not lived up to the expectations, calling all AI achievements as past disappointments.

The main section of the document was titled “Past disappointments”. Lighthill included an example of automatic landing system for airplanes, where conventional engineering techniques using radio waves had proven to be more useful than AI methods. Although AI techniques might be useful when landing an aircraft on an uncontrolled environment, Lighthill saw AI was nowhere near a practical solution. He also cited the lack of sophistication of chess programs that only reached an “experienced amateur” level at that time. Other than showing the lack of successful AI applications, Lighthill pointed out that all AI method required substantial knowledge of the subject matter in order to be useful. He was unsatisfied with at the lack of ability to automatically acquire knowledge that made the AI methods not really intelligent.

The problem of combinatorial explosion was specifically cited ridiculing the state-of-the-art AI technologies as merely applicable for solving toy problems in the laboratory and not scalable to the real world problems. Professor Lighthill, being an expert in fluid dynamics could not comprehend how a robot would function without the explicit coding of each and every instruction necessary for real-life behavior and action, since an infinite series of actions are not possible in the real-life situation. Thus, robotics was doomed to fail due to the combinatorial explosion problem. This same argument would apply that chess playing AI programs performing on the Grand Master level would be virtually impossible.

Lighthill's report provoked a massive loss of confidence in AI by the academic establishment in the UK including the funding body. It persisted for almost a decade.

TWO TRIUMPHS OF AI

In 1997, the IBM supercomputer Deep Blue defeated the chess world champion, Gary Kasparov. In 2011, the IBM supercomputer Watson, named after IBM's founder, defeated the reigning Jeopardy champions, Brad Rutter and Ken Jennings. Judging from these two major triumphs against human champions one might say that AI has recovered from the long winters and is making sustainable progress in the 21st century.

IBM's Deep Blue

In the 1997 chess world championship man-machine match, the first game was won by Kasparov and the second by Deep Blue; the next three games were a draw. In the 6th game, Deep Blue defeated Kasparov (Newborn, 2003).

Deep Blue's decided its moves by computing the evaluation function, which is a measure of the goodness of a given chess position. Deep Blue's evaluation function looked at four basic chess values: material, position, King safety, and tempo. Material is based on the worth of a

particular chess piece. For example, if a pawn is worth one point, a knight or bishop three, a rook five and the queen nine. However, conducting an exhaustive search to find the best possible move is next to impossible since the game of chess has a state-space of 10^{23} and game-tree complexity of 10^{143} (Iida, 2002).

However, instead of analyzing to the end of the game, the chess-playing program can merely look a few moves further ahead than a human could manage. Deep Blue typically looked 12 plies ahead in all variations (and 40 or more plies in selective lines), generating around 200 million nodes per second. It would evaluate each of these node positions and add points for a range of positional factors, chosen from a data-based created by human grandmasters. (Hsu, 1999; 2002)

Deep Blue conducted the search using a massively parallel, RS/6000 SP Thin P2SC-based system with 30 nodes, with each node containing a 120 MHz P2SC microprocessor, enhanced with 480 special purpose VLSI chess chips. Its chess playing program was written in C and ran under the AIX operating system. In June 1997, Deep Blue was the 259th most powerful supercomputer according to the TOP500 list, achieving 11.38 GFLOPS on the High-Performance LINPACK benchmark (Gelder, 1998).

IBM'S Watson

In 2011, IBM's supercomputer Watson, named after IBM's founder, defeated the reigning Jeopardy champions, Brad Rutter and Ken Jennings. Jeopardy is a well-known TV quiz in the US (Baker, 2012). Watson was designed to take the hints from the Quiz, do natural language processing and frame the Jeopardy questions. The following represents a typical Jeopardy question. "This NFL quarterback is a great-great-great-grandson of Brigham Young." Jeopardy question can be complex, usually requiring the contestant to put several facts together. Fortunately, for Watson, the answers were formatted very neatly. In this case, the answer is "Who is Steve Young". Watson simply had to return a fact (Heaton, 2011).

Table 3. Watson's software components and data contents

Software Components	Structured Text	Unstructured Text
SUSE Linux (OS)	WordNet	Encyclopedias
Apache Hadoop	DBPedia	Wikipedias
Apache UIMA	YAGO	Dictionaries & Thesauri

The software components and the data contents used inside Watson are shown in Table 3. Watson uses a complex combination of natural language processing, semantic analysis, information retrieval, automated reasoning and machine learning to answer the questions. Watson uses many existing algorithms to generate potential answers. First, the sentence is parsed. Then hypotheses are created. These hypotheses are then checked against evidence. Finally, the hypotheses have confidence levels assigned to them. If the top hypothesis has a confidence level above the threshold, Watson proposes an answer. All this is done under 3 seconds in the actual game of jeopardy.

The parallel processing inside Watson is performed by a specially designed hardware, consisting of 90 IBM Power 750 Servers with a total of 2,880 POWER7 processor cores and 16 terabytes of RAM. The software components include Apache Hadoop and UIMA (Unstructured Information Management Architecture). UIMA provides standards-based frameworks that allow analysis and annotation of large volumes of computer text. Watson used Apache UIMA for real-time content analytics and natural language processing.

AI ON THE CLOUD

Computing technology on the whole is moving on to the cloud. In the cloud computing environment, system developers and programmers as well as end-users access the systems installed in remote locations. The hardware on which the systems

are installed is networked and distributed. Current trends in AI are slowly embracing the cloud computing architecture. Watson is a case in point. After its grand success in Jeopardy!, Watson is offering its AI services over the cloud. Large scale financial and healthcare enterprises are increasingly seeking the services of AI embedded systems running in the cloud environment. The machine learning AI bots can easily operate on distributed datasets containing disparate and asynchronous data leading to useful knowledge discovery in real time necessary for business advancement. Parallel computing on millions of nodes, machine-readable information in the form of the Semantic Web and natural language query interface, all functioning in a seamless way in the cloud computing environment represent a genuine innovation that will make the future AI scalable, faster and closer to humans.

CONCLUSION

The summers and winters of AI is a metaphor for describing the ups and downs of the hype curve traced by AI in its long history. The early AI programs and Expert Systems created a sensation and consequently very high optimism. However, AI could not deliver the goods and soon entered a wintry period in its history with an all-time low funds and low activity. However, some dedicated researchers continued to work in limited areas refraining from calling their work "AI". These bore fruits with the culmination of IBM's Deep Blue defeating chess world champion in 1997, followed by IBM's Watson defeating Jeopardy world champions in 2011. True, we are far from an overall theory or framework, explaining what thinking is and how to build intelligent machines. Nevertheless, AI has made great strides and, after overcoming the *trough of disillusionment* is marching on the *plateau of productivity*. Currently, embracing the frontline cloud computing technology, AI is poised to make a further breakthrough.

REFERENCES

- Baker, S. (2012). *Final Jeopardy: The story of Watson, the computer that will transform our world*. Boston, MA: Mariner Books.
- Berk, A. A. (1985). *LISP: the Language of Artificial Intelligence*. New York: Van Nostrand Reinhold Company.
- Boban, B. (2013). *An analysis on the dendral expert system*. GRIN Verlag.
- Boden, M. A. (2006). *Mind as Machine: A History of Cognitive Science*. New York: Oxford University Press.
- Buchanan, B. G., & Shortliffe, E. H. (Eds.). (1984). *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*. Addison-Wesley.
- CloudWedge. (n.d.). *Cloud Computing and A.I.* Retrieved January 12, 2016, from <http://www.cloudwedge.com/cloud-computing-and-a-i/>
- Cooper, G. F. (1989). Current research directions in the development of expert systems based on belief networks. *Applied Stochastic Models in Business and Industry*, 5(1), 39–52. doi:10.1002/asm.3150050106
- Crevier, D. (1993). *AI: The Tumultuous history of the search for artificial intelligence*. New York: Basic Books.
- Duda, R., & Buchanan, B. (1982). *Principles of rule-based expert systems*. Stanford, CA: Computer Science Dept., Stanford University.
- Gelder, T. (1998). *Into the Deep Blue yonder. IBM's Deep Blue chess computer and Gary Kasparov*. Quadrant.
- Giarratano, J. C., & Riley, G. D. (1998). *Expert Systems: Principles and Programming* (3rd ed.). Boston: PWS Pub.
- Heaton, J. (2011). *The Free and Open Software Behind IBM's Jeopardy Champion Watson*. Retrieved April 30, 2015, from <http://www.heatonresearch.com/content/free-and-open-software-behind-ibm's-jeopardy-champion-watson>
- Hsu, F. (1999). IBM's Deep Blue Chess grand-master chips. *IEEE Micro*, 70–81.
- Hsu, F. (2002). *Behind deep blue: Building the computer that defeated the world chess champion*. Princeton, NJ: Princeton University Press.
- Huffington Post Articles, Posts Tagged With: Artificial Intelligence. (n.d.). Retrieved January 13, 2016, from <http://cloud-computing-today.com/tag/artificial-intelligence/>
- Hutchins, J. (1996). ALPAC: The (in)famous report. *MT News International*, 14, 9–12.
- Hutchins, J. (2005). *The history of machine translation in a nutshell*. Retrieved April 30, 2015, from <http://www.hutchinsweb.me.uk/Nutshell-2005.pdf>
- Iida, H., Sakuta, M., & Rollason, J. (2002). Computer shogi. *Artificial Intelligence*, 134(1-2), 121–144. doi:10.1016/S0004-3702(01)00157-6
- King, M. (1984). When is the next ALPAC report due? In *Proceedings of 10th International conference on computational linguistics* (pp. 352-353). Stanford University. doi:10.3115/980431.980563
- Lighthill, J. (1973). *Artificial Intelligence: A General Survey. Artificial Intelligence: a paper symposium*. Science Research Council.
- Lindsay, R. (1980). *Applications of artificial intelligence for organic chemistry: The DENDRAL project*. New York: McGraw-Hill Book.
- Lindsay, R. K., Buchanan, B. G., Feigenbaum, E. A., & Lederberg, J. (1980). *Applications of Artificial Intelligence for Organic Chemistry: The Dendral Project*. McGraw-Hill Book Company.

- Lungarella, M. (2007). *50 years of artificial intelligence essays dedicated to the 50th anniversary of artificial intelligence*. Berlin: Springer.
- McCarthy, J. (2001). *What is Artificial Intelligence?* Stanford University.
- McCarthy, J., Minsky, M., Rochester, N., & Sannon, C. (1955). *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*. Retrieved April 30, 2015, from <http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>
- McClelland, J., & Rumelhart, D. (1988). *Explorations in Parallel Distributed Processing: A Handbook of Models, Programs, and Exercises*. Cambridge, MA: MIT Press.
- McClelland, J. L., & Rumelhart, D. E. (1986). *Parallel Distributed Processing* (Vol. 1). MIT Bradford Press.
- McCloskey, M. (1991). Networks and Theories: The Place of Connectionism in Cognitive Science. *Psychological Science*, 2(6), 387–395. doi:10.1111/j.1467-9280.1991.tb00173.x
- McCorduck, P. (1979). *Machines who think: A personal inquiry into the history and prospects of artificial intelligence*. San Francisco: W.H. Freeman.
- McCulloch, W., & Pitts, W. (1943). A Logical Calculus of the Ideas Immanent in Nervous Activity. *The Bulletin of Mathematical Biophysics*, 5(4), 115–133. doi:10.1007/BF02478259
- McGuire, B. (2006). *The Turing Test, History of Computing Course, CSEP 590A*. University of Washington.
- Menzies, T. (2003). 21st-Century AI: Proud, Not Smug. *IEEE Intelligent Systems*, 18(3), 18–24. doi:10.1109/MIS.2003.1200723
- Minsky, M., & Papert, S. (1988). *Perceptrons: An Introduction to Computational Geometry* (3rd ed.). Cambridge, MA: MIT Press.
- Newborn, M. (2003). *Deep Blue An Artificial Intelligence Milestone*. New York, NY: Springer.
- Nikolopoulos, C. (1997). *Expert systems: Introduction to first and second generation and hybrid knowledge based systems*. New York: M. Dekker.
- Nilsson, N. (2010). *The quest for artificial intelligence: A history of ideas and achievements*. Cambridge, UK: Cambridge University Press.
- Palmer, S. (2011, May 1). Man vs. Watson: The IBM Jeopardy Challenge. *The Humanist*.
- Pierce, J. R. (1966). *Language and Machines — Computers in Translation and Linguistics. ALPAC report*. Washington, DC: National Academy of Sciences, National Research Council.
- Poole, D., & Mackworth, A. (2010). *Artificial intelligence foundations of computational agents*. New York: Cambridge University Press. doi:10.1017/CBO9780511794797
- Rajendra, A., & Sajja, P. (2010). *Knowledge-Based Systems*. London: Jones & Bartlett.
- Rosenblatt, F. (1962). *Principles of neurodynamics*. New York: Spartan.
- Russell, S., & Norvig, P. (2010). *Artificial Intelligence: A Modern Approach*. Upper Saddle River, NJ: Pearson Education, Inc.
- Sasikumar, M., Ramani, S., Raman, S. M., Anjaneyulu, K. S. R., & Chandrasekar, R. (2007). *A Practical Introduction to Rule-Based Expert Systems*. New Delhi: Narosa Publishing House.
- Schalkoff, R. (2011). *Intelligent Systems: Principles, paradigms, and pragmatics*. Sudbury, MA: Jones and Bartlett.
- Tim Jones, M. (2009). *Artificial Intelligence: A Systems Approach*. Jones and Barlett Publishers.
- Turing, A. (1950). *Computing Machinery and Intelligence*. Retrieved April 30, 2015, from <http://www.abelard.org/turpap/turpap.htm>

ADDITIONAL READING

Coppin, B. (2004). *Artificial intelligence illuminated*. Boston: Jones and Bartlett.

Ertel, W. (2011). *Introduction to artificial intelligence*. London: Springer. doi:10.1007/978-0-85729-299-5

Frankish, K., & Ramsey, W. M. (n.d.). *The Cambridge handbook of artificial intelligence*. Cambridge, UK: Cambridge University Press.

Margulies, P. (2004). *Artificial intelligence*. San Diego, CA: Blackbirch Press.

Mishkoff, H. C., Shafer, D., & Rasmus, D. W. (1988). *Understanding artificial intelligence*. Indianapolis, IN, USA: H.W. Sams.

Roland, A., & Shiman, P. (2002). *Strategic Computing: DARPA and the Quest for Machine Intelligence, 1983-1993 (History of Computing)*. Cambridge, USA: MIT.

Wood, G. (2003). *Edison's Eve: A Magical History of the Quest for Mechanical Life*. New York: Random house, Inc.

Woolf, A. (2003). *Artificial intelligence: The impact on our lives*. Austin, TX: Raintree Steck-Vaughn.

Zarkadakēs, G. (n.d.). *In our own image: Savior or destroyer?: The history and future of artificial intelligence*. New York. Pegasus Books LLC.

KEY TERMS AND DEFINITIONS

Artificial Intelligence (AI): The science of building intelligent machines.

Cloud Computing: The distributed and parallel computing technology that allow provides users with applications and data storage over the internet. The *cloud* is a synonym for the internet.

Connectionism: Another term for neural networks.

Deep Blue: IBM supercomputer that defeated world chess champion in 1997.

Expert Systems: Computers programs styled on the knowledge and reasoning of experts.

Machine Learning: AI programs that extract useful patterns from vast amounts of data and makes unforeseen predictions based on these data, without explicit programming.

Neural Networks: Network system, modeled after the human brain, capable of learning from data.

Turing Test: A test proposed by Alan Turing in 1950 to test the intelligence of computers.

Watson: IBM supercomputer that beat human champions in the Jeopardy TV contest in 2011.

Swarm Intelligence for Multi-Objective Optimization in Engineering Design

A

Janga Reddy Manne

Indian Institute of Technology Bombay, India

INTRODUCTION

Nowadays, most of the engineering design problems are intrinsically complex and difficult to solve, because of diverse solution search space, complex functions, continuous and discrete nature of decision variables and hard constraints. Meta-heuristic algorithms are becoming popular in dealing with these kind of complexities. Evolutionary algorithms (EAs) and swarm intelligence (SI) algorithms are being population based random search techniques becoming attractive global optimization solvers. The algorithms use guided rules or heuristics inspired from nature to enable effective exploration of optimal solutions to complex engineering problems. In recent past, a number of swarm intelligence (SI) algorithms were proposed based on principles of co-operative group intelligence and collective behavior of self-organized systems. The approaches use agents to perform explorations while they interact with neighbors and the environment. However, the individual members have limited search capabilities, co-operative group intelligence and/or knowledge sharing among the swarm helps to obtain optimal solutions to complex engineering problems. The popular SI algorithms include particle swarm optimization (PSO) that was emerged from simulating the behavior of flocks of birds (Kennedy & Eberhart, 1995), ant colony optimization (ACO) that was emulating the behavior of ants foraging for food (Dorigo, 1992). Other SI algorithms include Honey-bee mating algorithm, glow swarm algorithm, bacterial Foraging and Cuckoo search algorithms etc.

Many times, practical engineering design problems are characterized by multiple conflicting goals. In contrast to single objective optimization, multi-objective optimization deals with simultaneous optimization of several non-commensurable and often competitive/conflicting objectives. Because of the multiple conflicting objectives, it may not be possible to find a single optimal solution that will satisfy all the stated goals, instead, the solution exists in the form of alternative trade-offs, also known as the non-inferior or non-dominated solutions. In the past, several studies have used classical optimization techniques such as linear programming (LP), dynamic programming (DP) and non-linear programming (NLP) to solve the multi-objective problems by adopting weighted-sum or constrained approach etc. These approaches may face difficulties while generating non-dominated solutions for practical problems. For example, in the weighted-sum approach, the multiple objectives of the problem are converted into a single objective optimization by adopting suitable weights to all the objectives. By using a single pair of fixed weights, only one point on Pareto-front can be obtained. Therefore, if one would like to obtain the complete set of Pareto optimal front, all possible Pareto solutions must first be derived. This requires the algorithms to be executed iteratively, so as to ensure that every weight combination has been evaluated. Obviously, it is unrealistic to reiterate the algorithms continually to exhaust all the weight combinations. Similarly, in the constraint method, it needs to reiterate the algorithm for a large number of times, which requires more computational effort. Also

conventional approaches may face difficulties, if optimal solution lies on non-convex or disconnected regions of the objective function space. Thus the classical approaches are not ideal to solve multi-objective optimization problems (MOOP). In developing an algorithm for solution of an MOOP, it should have an ability to learn from past performance, to direct proper selection of weights for further evolutions. To achieve these goals, multi-objective evolutionary algorithms (MOEAs) have been proposed and are suggested as effective means to deal with these issues (Reddy & Kumar, 2007a). Due to their efficiency and easiness to handle non-linear functions, ability to approximate the non-convex and disconnected Pareto optimal fronts of real-world problems, MOEAs are getting diverse applications in engineering design. Apart from that, the specific advantage of MOEAs over the classical approaches is that they generate a population of solutions in each iteration and offer a set of alternatives (Pareto optimal set) in a single run. Thus population based stochastic search techniques are becoming more popular to solve MOOPs.

In the following sections, first the principles and issues in developing multi-objective algorithms are discussed. Then swarm intelligence based algorithm for multi-objective optimization is presented. Subsequently, application of the methodology is illustrated though few multi-objective engineering design problems.

BACKGROUND

Multi-Objective Problem

A general multi-objective optimization problem can be defined as, minimize a set of functions $f(x)$, subject to p inequality and q equality constraints (Reddy & Kumar, 2007a).

$$\text{Min. } f(x) = \{f_1(x) f_2(x) \dots f_m(x)\}^T \quad (1)$$

$$x \in D$$

where $x \in R^n$, $f_i: R^n \rightarrow R$ and

$$D = \left\{ \begin{array}{ll} x \in R^n : & l_i \leq x \leq u_i, \quad \forall i = 1, \dots, n \\ & g_j(x) \geq 0, \quad \forall j = 1, \dots, p \\ & h_k(x) = 0, \quad \forall k = 1, \dots, q \end{array} \right. \quad (2)$$

where m is number of objectives; D is feasible search space; $x = \{x_1 x_2 \dots x_n\}^T$ is the set of n -dimensional decision variables (continuous, discrete or integer); R is the set of real numbers; R^n is n -dimensional hyper-plane or space; l_i and u_i are lower and upper limits of i -th decision variable.

Pareto Optimal Solution

In MOOP, the desired goals are often conflicting against each other and it is not possible to satisfy all the goals at a time, which leads to definition of *Pareto optimal* solutions. The Pareto optimal solution refers to a solution, around which there is no way of improving any objective without degrading at least one other objective.

Pareto front is a set of non-dominated solutions, being chosen as optimal, if no objective can be improved without sacrificing at least one other objective (Deb et al., 2002). On the other hand a solution x^* is referred to as dominated by another solution x , if and only if, x is equally good or better than x^* with respect to all objectives. The definition of Pareto optimality is very much useful in MOEAs to classify the population of solutions into dominated and non-dominated members, thereby helping in the selection of member solutions from one generation to next generation.

Multi-Objective Evolutionary Algorithms

In the last two decades, a number of evolutionary algorithms (EAs) were proposed to solve multi-objective optimization problems. The first generation MOEAs, Non-dominated Sorting Genetic

Algorithm (NSGA) (Srinivas & Deb, 1994) and Niched Pareto Genetic Algorithm (NPGA) (Horn et al., 1994) have received good recognition at the beginning in 1990s. Subsequently, the elitist multi-objective evolutionary algorithms were found to be more efficient than those without elitism, since the elitism helps to preserve the best solutions in the past iterations and speeds up the convergence of the solution. Of them, the second generation MOEAs, Pareto-Archived Evolution Strategy (PAES) (Knowles & Corne, 2000), Strength-Pareto Evolutionary Algorithm (SPEA) (Zitzler & Thiele, 1999) and Non-dominated Sorting Genetic Algorithm-II (NSGA-II) (Deb et al., 2002) are popular due to their efficiency in producing better Pareto front. Later on, many variants of MOEAs have been proposed over the years. Also several SI algorithms are extended to solve multi-objective optimization problems. This chapter mainly focuses particle swarm optimization based multi-objective optimization.

Studies on Multi-Objective Optimization Using PSO

The PSO algorithm has proven capabilities of quick convergence to optimal solution for single objective problems (Kumar & Reddy, 2007). The similarities of Particle Swarm Optimization (PSO) with EAs inspired the researchers to extend the algorithm to handle multiple objectives. The PSO algorithm maintains population of solutions, which allows exploration of different parts of the Pareto front simultaneously. By incorporating Pareto-dominance principle into PSO algorithm, various Multi-objective Particle Swarm Optimization (MOPSO) techniques are formulated. Ray & Liew (2002) proposed swarm metaphor approach, which uses Pareto dominance relation and combines the concepts of evolutionary techniques with the PSO. Parsopoulos & Vrahatis (2002) proposed MOPSO algorithm, adopting different types of aggregating functions to solve multi-objective optimization problems. Hu and Eberhart (2002) proposed Dynamic Neighborhood PSO, in which only one objective is optimized at

a time using a scheme similar to lexicographic ordering. A revised version of this approach, which uses a secondary population, is presented in Hu et al. (2003). Coello & Lechuga (2002) proposed MOPSO, based on the idea of having a global repository in which every particle will deposit its flight experiences after each flight cycle. Additionally, the updates to the repository are performed considering a geographically-based system, defined in terms of the objective function values of each individual. This repository is used by the particles, to identify a leader that will guide the search. The revised version of the approach is presented in Coello et al. (2004), in which a perturbation operation on decision variable space is applied to improve the performance of the algorithm. The approach of Fieldsend and Singh (2002) incorporates an unconstrained elite archive to store the nondominated individuals found during the search process. The archive interacts with the primary population in order to define local guides. This approach also uses a mutation operator that acts on the velocity value used by PSO. Li (2003) proposed non-dominated sorting PSO, which incorporates the main mechanisms of the NSGA-II into a PSO algorithm and the approach showed very competitive performance compared to NSGA-II. Reddy (2006) proposed an elitist mutated multi-objective particle swarm optimization (EM-MOPSO) for solving multi-objective problems in engineering design, and then tested its performance for solving numerical optimization problems, structural design problems, and water resources problems (Reddy & Kumar, 2007a & 2007b). In the following, the EM-MOPSO methodology for solving MOOP is explained.

MULTI-OBJECTIVE PARTICLE SWARM OPTIMIZATION

Main Issues in Developing Efficient Multi-Objective Solvers

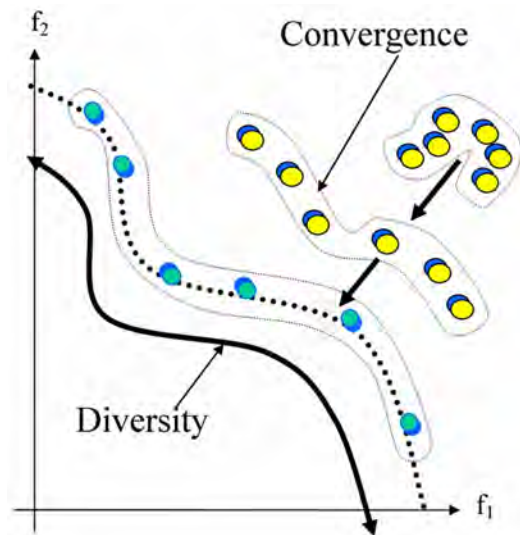
Achieving a well-spread and global Pareto optimal front is the primary goal while solving MOOPs.

In these algorithms, apart from finding a non-dominated solution in each generation, more computational effort is necessary for preserving diversity among the generated solutions. This computational complexity is directly related to the level of diversity and distribution, which the particular algorithm aims to obtain. Therefore, the major issues in developing an effective solution approach are (Reddy & Kumar, 2007a): (i) how to guide the randomly scattered population towards the true Pareto optimal front; (ii) how to maintain good diversity in the generated non-dominated solutions. (iii) how to avoid the loss of obtained quality non-dominated solutions over the generations. These issues are also depicted in Figure 1. A good solver should address all these issues.

Elitist-Mutated Multi-Objective Particle Swarm Optimization

The EM-MOPSO procedure combines Pareto-dominance principles with PSO and uses elitism in its evolution (Reddy & Kumar, 2007a). The main algorithm consists of initialization of population,

Figure 1. Graphical illustration of main issues in the development of multi-objective evolutionary algorithm where f_1 and f_2 are two objective functions



evaluation, performing PSO operations and reiterating the search on swarm to reach true Pareto optimal solutions. In this process, the particles are first evaluated and checked for dominance relation among the swarm. Then the non-dominated solutions found are stored in an external repository (*ERP*). The size of this repository is restricted to a pre-defined number. This restriction is done, using crowding distance comparison operator (Deb et al., 2002), which gives the density measure of the existing particles in the function space. This *ERP*, with crowding operator, helps the particles to create effective selection pressure towards true Pareto optimal solutions. Also in this procedure a step wise linearly variable *ERP* is used. This helps in achieving a well distributed Pareto front and saves considerable computational time during optimization. The selection of global guides for each particle is performed by randomly choosing one solution from those stored in the *ERP*. In addition, an efficient strategic mechanism called elitist-mutation is incorporated into the algorithm. By attracting the swarm towards sparsely populated regions in *ERP*, it helps the search to uniformly distribute the non-dominated solutions along the true Pareto optimal front. Some of the main operators used in the procedure are explained below.

PSO Algorithm

In PSO each particle represents a potential solution and the position of each particle is changed according to its own experience and that of its neighbors (Eberhart & Keenedy, 1995). If the search space is D -dimensional, the i^{th} individual (*particle*), of the population (swarm), can be represented by a D -dimensional vector, $X_i = (x_{i1}, x_{i2}, \dots, x_{iD})^T$. The *velocity* (position change) of this particle, can be represented by another D -dimensional vector $V_i = (v_{i1}, v_{i2}, \dots, v_{iD})^T$. The best previously visited position of the i^{th} particle is denoted as $P_i = (p_{i1}, p_{i2}, \dots, p_{iD})^T$ and acts as a local guide to the particles. Defining g as the index of the global guide of particle, and superscripts denoting the iteration number, the

velocity and position of a particle are manipulated according to the following two equations:

$$v_{id}^{n+1} = \chi[w v_{id}^n + c_1 r_1 \frac{(p_{id}^n - x_{id}^n)}{\Delta t} + c_2 r_2 \frac{(p_{gd}^n - x_{id}^n)}{\Delta t}] \quad (3)$$

$$x_{id}^{n+1} = x_{id}^n + \Delta t v_{id}^{n+1} \quad (4)$$

where $d = 1, 2, \dots, D$; $i = 1, 2, \dots, N$; N is the size of the swarm population; χ is a constriction factor which controls and constricts the velocity's magnitude; w is the inertia weight, which is often used as a parameter to control exploration and exploitation in the search space; c_1 and c_2 are positive constant parameters called acceleration coefficients; r_1 and r_2 are functions of uniformly distributed random numbers in $[0, 1]$; Δt is the time step usually set as 1 and n is iteration number.

Elitist-Mutation

The idea underlying the elitist-mutation mechanism (Reddy & Kumar, 2007a) is to effectively explore and exploit the search space in the feasible region, where members of the repository that are isolated in the non-dominated *ERP* should be preferentially mutated and replace the worst particles in the swarm. This is performed on a predefined number of particles. At initial phase of this mechanism, it replaces the infeasible solutions with the least crowded solutions in *ERP* after performing elitist mutation on those particles, and at later phase it tries to exploit the search space around the sparsely populated particles along the Pareto fronts. This is a special mechanism which helps to overcome the drawbacks of traditional PSO algorithm, on being extended to solve MOOP. This also promotes diversity in the population and consequently helps in the search to find the true Pareto optimal front.

The elitist-mutation operator involves three key steps: (i) Randomly select one of the objec-

tives from m objectives. Sort the fitness function in descending order and get the index number for the respective particles; (ii) Use crowding distance assignment operator and calculate the density of solutions in the external repository (*ERP*) and sort them in descending order of crowding value. Randomly select one of the sparsely populated solutions from top 10% of *ERP* as guide (g). (iii) Perform elitist-mutation on predefined number of particles.

More details on elitist-mutation operator, variable size external repository, crowding distance comparison operator, and handling overlapping solutions can be found in Reddy & Kumar (2007a.)

Step-By-Step Procedure of EM-MOPSO

The EM-MOPSO algorithm can be summarized in the following steps (Reddy and Kumar, 2007a):

1. Initialize the population position and velocity vectors. The current position of the i -th particle, X_i is initialized with random real numbers within the specified decision variable range; each particle velocity vector V_i is initialized with uniformly distributed random number in $[0, 1]$.
2. Evaluate each particle in the population. The personal best position P_i is set to X_i . Identify particles that give non-dominated solutions in the current population and store them in an external repository (*ERP*). Set iteration counter, $t = 0$.
3. Select randomly a global best P_g for the i -th particle from the solutions stored in *ERP*. Calculate the new velocity V_i based on the Equation (3), and update the position X_i by Equation (4). Repeat the procedure for all the particles.
4. Evaluate each particle in the population. Then each individual is checked for dominance with its current personal best solution. Replace P_i with current solution, if the new one dominates the current P_i .

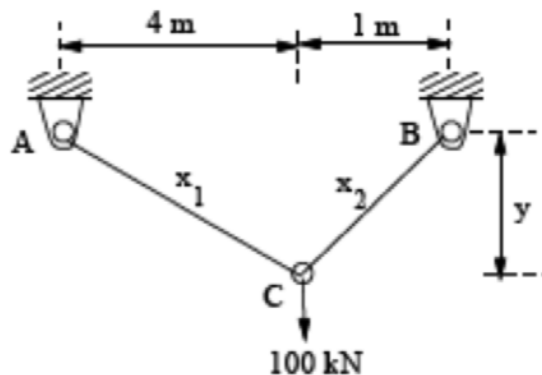
5. Set *ERP* to a temporary repository (*TempERP*) and empty *ERP*. Identify particles that give non-dominated solutions in the current iteration and add them to *TempERP*.
6. Find the non-dominated solutions in *TempERP*. If the number of non-dominated solutions is found to exceed the desired size of *ERP*, then use crowding distance operator to select the desired ones and store them in *ERP*. Empty the *TempERP*.
7. Sort the *ERP* according to crowding distance values and perform elitist-mutation operation on specified number of particles.
8. Increment the iteration counter, $t = t + 1$ and check for termination criteria. If the termination criterion is not satisfied, then go to step 3; otherwise output the non-dominated solution set from *ERP*.

To handle constrained optimization problem, the study adopted the constraint handling mechanism as given in Deb et al. (2002). This approach provides simplicity in using feasibility and non-dominance of solutions when comparing solutions. In the following sections, the application of EM-MOPSO algorithm for solving MOOP is demonstrated.

APPLICATION OF EMMOPSO TO MULTI-OBJECTIVE ENGINEERING DESIGN

This section presents the application of EM-MOPSO algorithm to structural design problems namely two bar truss design, and I-beam design. To apply EM-MOPSO and NSGA-II algorithms the same values of the parameters are used as constant parameters $c1=1.0$ and $c2=0.5$; inertial weight $w=1$; constriction coefficient $\chi=0.9$; size of external repository, $ERP=100$; the size of elitist-mutated particles is set to 20; and the maximum number of iterations used is 100.

Figure 2. The two-bar truss design problem. Minimization of cost of fabrication and minimizing stresses in members (AC and BC) are the two objectives



Two-Bar Truss Design

This problem was originally studied using the ϵ -constraint method (Palli et al., 1999). As shown in Figure 2, the truss has to carry a certain load without elastic failure. Thus, in addition to the objective of designing the truss for minimum volume (which is equivalent to designing for minimum cost of fabrication), there are additional objectives of minimizing stresses in each of the two members AC and BC.

So the design problem involves a two-objective optimization problem for three variables y (vertical distance between B and C in m), x_1 (length of AC in m) and x_2 (length of BC in m). The two-bar truss design problem in mathematical form can be expressed as:

minimize,

$$f_1(x) = x_1 \sqrt{16 + y^2} + x_2 \sqrt{1 + y^2}$$

minimize,

$$f_2(x) = \max(\sigma_{AC}, \sigma_{BC})$$

subject to

$$\max(\sigma_{AC}, \sigma_{BC}) \geq 1(10)^5$$

$$1 \leq y \leq 3 \text{ and } x \geq 0$$

The stresses are calculated as follows:

$$\sigma_{AC} = \frac{20 \sqrt{16 + y^2}}{y x_1},$$

$$\sigma_{BC} = \frac{80 \sqrt{1 + y^2}}{y x_2}$$

To apply the proposed method, the bounds on x_i are taken as $0 \leq x_i \leq 0.01$ for $i=1$ and 2. The ϵ -constraint method reported only five solutions with the following spread: (0.004445 m³, 89983 kPa) and (0.004833 m³, 83268 kPa). The Figure 3 shows the optimized fronts obtained using the ϵ -constraint method, the EM-MOPSO and the NSGA-II methods.

With EM-MOPSO the solutions are spread in the following range: (0.004026 m³, 99996 kPa) and (0.05273 m³, 8434.493 kPa) and for NSGA-II (0.00407 m³, 99755 kPa) and (0.05304 m³, 8439 kPa). Thus both have a wide variety of alternatives. But the ϵ -constraint method could not find much variety of solutions in terms of the second objective (Palli et al., 1999). If minimum volume

is desired, EM-MOPSO gives as low as 0.004026 m³ or if minimization of stress is important, it finds a solution with stress as low as 8434.493 kPa, whereas the ϵ -constraint method has found a solution with minimum stress of 83,268 kPa, which is about 10 times higher than that found with EM-MOPSO. Also the EM-MOPSO solutions are very competitive with NSGA-II solutions, both in terms of closeness to the true optimum front and in their spread.

I-Beam Design

The second design problem is taken from Yang et al. (2002). The problem is to find the dimensions of the beam presented in Figure 4. In this design problem, it should satisfy the dimensions of the geometric and strength constraints, and at the same time minimize two objectives, cross-sectional area of the beam and static deflection of the beam under the force P.

The mathematical form of the problem can be written as below

Minimize, Cross sectional area (cm²):

$$f_1 = 2 x_2 x_4 + x_3 (x_1 - 2 x_4)$$

Figure 3. Non-dominated solutions obtained using EM-MOPSO and NSGA-II for the two-bar truss design problem. The optimal solutions found by ϵ -constraint method (eps-constr) are also shown for comparison

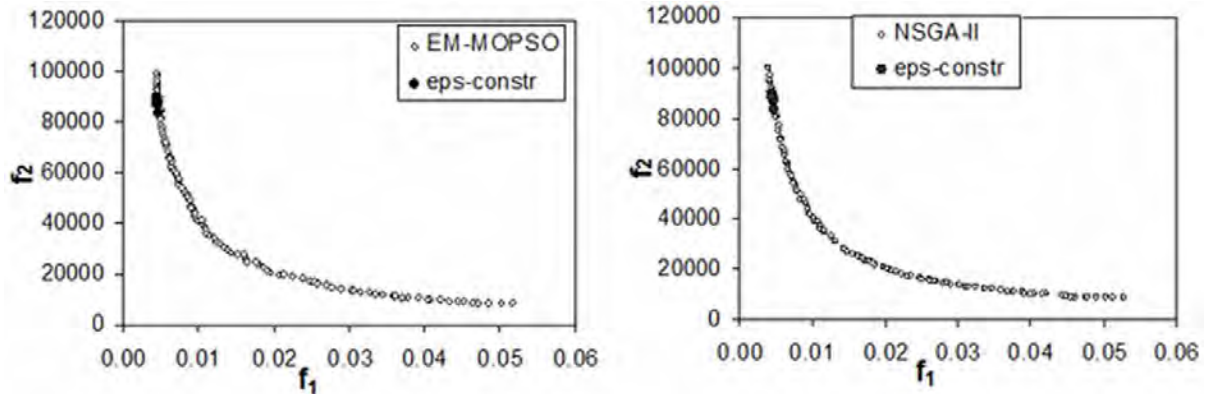
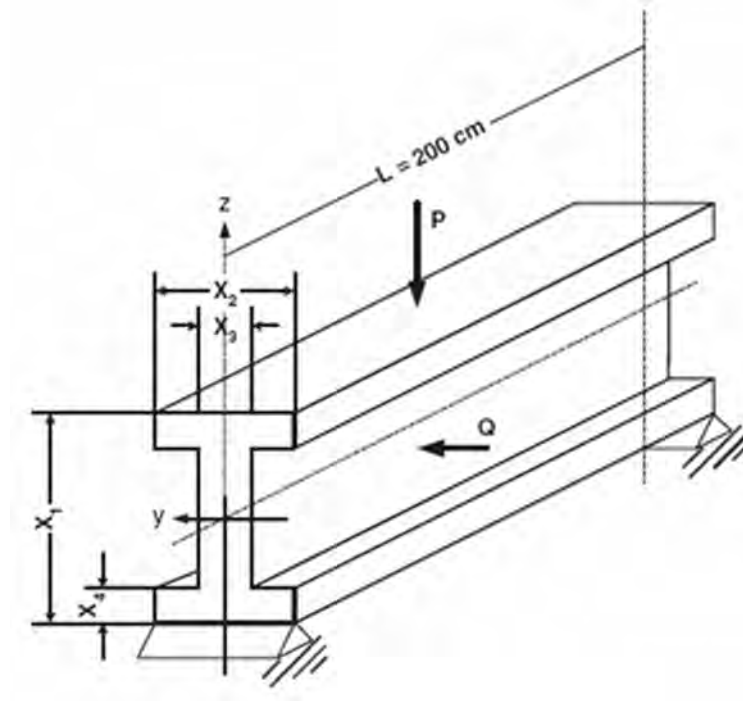


Figure 4. The I-beam design problem. Minimization of cross-sectional area of the beam and static deflection of the beam are two objectives



Minimize, Deflection (cm):

$$f_2 = \frac{P L^3}{48 E I}$$

where

$$I = \frac{1}{12} \left\{ x_3 (x_1 - 2x_4)^3 + 2x_2 x_4 \left[4x_4^2 + 3x_1 (x_1 - 2x_4) \right] \right\}$$

subject to

$$g(x) = \sigma_a - \left(\frac{M_y}{Z_y} + \frac{M_z}{Z_z} \right) \geq 0$$

$$10 \leq x_1 \leq 80, 10 \leq x_2 \leq 50, 0.9 \leq x_3 \leq 5, 0.9 \leq x_4 \leq 5$$

$$\text{where } M_y = \frac{P}{2} \times \frac{L}{2} \quad M_z = \frac{Q}{2} \times \frac{L}{2}$$

$$Z_y = \frac{1}{6x_1} \left\{ x_3 (x_1 - x_4)^3 + 2x_2 x_4 \left[4x_4^2 + 3x_1 (x_1 - 2x_4) \right] \right\}$$

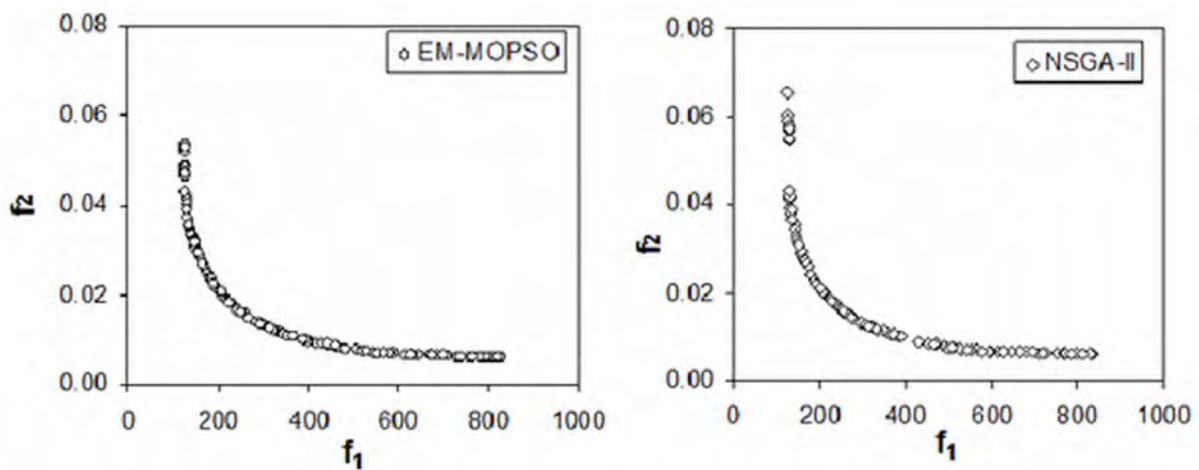
$$Z_z = \frac{1}{6x_2} \left\{ (x_1 - x_4)x_3^3 + 2x_4 x_2^3 \right\}$$

$E = 2 \times 10^4 \text{ kN/cm}^2$, $\sigma_a = 16 \text{ kN/cm}^2$, $P = 600 \text{ kN}$, $Q = 50 \text{ kN}$ and $L = 200 \text{ cm}$.

The non-dominated solutions obtained after 100 iterations using both EM-MOPSO and NSGA-II are shown in Figure 5. It can be noticed that EM-MOPSO is able to maintain uniform distribution of solutions.

The EM-MOPSO found the minimal cross sectional area as 127.9508 units for deflection of 0.05368 units and for the minimal deflection of 0.005961 units, the cross sectional area is 829.5748 units. Whereas for NSGA-II, the minimal cross sectional area is 127.2341 units with deflection of 0.0654 units and the minimal deflection is 0.0060

Figure 5. Non-dominated solutions obtained using EM-MOPSO and NSGA-II for the I-beam design problem



with cross sectional area of 829.8684 units. Thus the proposed method is able to find wider spread of Pareto optimal solutions than NSGA-II.

DISCUSSION

The swarm intelligence based multi-objective algorithms are simple to implement, yet efficient in yielding true Pareto optimal solutions. The computational complexity is also reasonable. In EMMOPSO approach, in addition to the objective function computations, the computational complexity of the algorithm is mainly dominated by the non-dominated comparison of the particles in the population, sorting and crowding distance computation. This computational requirement is comparable to that of state of the art MOEAs available at present. The results of EM-MOPSO for constrained optimization have shown that the method is yielding diverse set of Pareto optimal solutions. Thus, the use of these swarm intelligence based multi-objective algorithms having great importance for solving many practical problems in different areas of science, engineering and industrial problems.

CONCLUSION

The swarm intelligence algorithms can provide acceptable optimal solutions to many complex problems that are difficult to cope-up using the conventional methods due to their nature that may imply discontinuities of the search space, non-differentiable objective functions, non-linear relationships, imprecise arguments and function values. In this chapter, elitist-mutated multi-objective particle swarm optimization (EM-MOPSO) approach for solving multi-objective engineering design problems is presented. The EM-MOPSO method combines PSO algorithm with Pareto dominance criteria to evolve non-dominated solutions. It uses an external repository of variable size along with a crowded distance comparison operator to promote solution diversity. In addition, the methodology uses an elitist-mutation operator to keep diversity in the population and for effective exploration of Pareto optimal solutions. The results of EM-MOPSO for engineering design problems showed that the performance of EM-MOPSO approach is quite robust and providing a wide range of alternative solutions. Advantages of the proposed EM-MOPSO approach include

its adoptability and providing efficient Pareto frontiers for multi objective optimization. So it can be inferred that the MOPSO approach can be used as a viable multi objective tool for solving real world multiple objective problems. Still there is a lot of scope for research in improving the procedures, such as, automatic fine-tuning of algorithmic parameters, handling of larger number of objectives, reducing the computational complexity of the solvers, etc.

REFERENCES

- Chang, C. (2012). Multi-objective optimal SVC Installation for power system loading margin improvement. *IEEE Transactions on Power Systems*, 27(2), 984–992. doi:10.1109/TPWRS.2011.2176517
- Coelho, L. D. S., Barbosa, L. Z., & Lebensztajn, L. (2010). Multiobjective particle swarm approach for the design of a brushless DC wheel motor. *IEEE Transactions on Magnetics*, 46(8), 2994–2997. doi:10.1109/TMAG.2010.2044145
- Coello, C. A. C., & Lechuga, M. S. (2002). MOPSO: A proposal for multiple objective particle swarm optimization. *Proc. Congr. Evolutionary Computation (CEC'2002)*, 1051–1056. doi:10.1109/CEC.2002.1004388
- Coello, C. A. C., Pulido, G. T., & Lechuga, M. S. (2004). Handling multiple objectives with particle swarm optimization. *IEEE Transactions on Evolutionary Computation*, 8(3), 256–279. doi:10.1109/TEVC.2004.826067
- Daneshyari, W., & Yen, G. G. (2011). Cultural-based multiobjective particle swarm optimization. *IEEE Transactions on Systems, Man, and Cybernetics. Part B, Cybernetics*, 41(2), 553–567. doi:10.1109/TSMCB.2010.2068046 PMID:20837447
- Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. (2002). A fast and elitist multi-objective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2), 182–197. doi:10.1109/4235.996017
- Dorigo, M. (1992). *Optimization, learning and natural algorithms* (Ph.D. thesis). DEI, Politecnico di Milano, Italy.
- Eberhart, R. C., & Kennedy, J. (1995). A new optimizer using particle swarm theory. In *Proc 6th Symp on Micro Machine and Human Science*. IEEE Service Center. doi:10.1109/MHS.1995.494215
- Fieldsend, J. E., & Singh, S. (2002). A multi-objective algorithm based upon particle swarm optimization, an efficient data structure and turbulence. *Proc. 2002 U.K. Workshop on Computational Intelligence*, 37–44.
- Hamid, A., Waseem, S., & Farrukh, A. K. (2012). Energy-efficient clustering in mobile ad-hoc networks using multi-objective particle swarm optimization. *Applied Soft Computing*, 12(7), 1913–1928. doi:10.1016/j.asoc.2011.05.036
- Horn, J., Nafpliotis, N., & Goldberg, D. E. (1994). A niched pareto genetic algorithm for multiobjective optimization. *Proc. of the 1st IEEE Conference on Evolutionary Computation*, 22–87. doi:10.1109/ICEC.1994.350037
- Hu, X., & Eberhart, R. (2002). Multiobjective optimization using dynamic neighborhood particle swarm optimization. *Proc. Congr. Evolutionary Computation (CEC'2002)*, 1677–1681.
- Hu, X., Eberhart, R. C., & Shi, Y. (2003). Particle swarm with extended memory for multiobjective optimization. *Proc. 2003 IEEE Swarm Intelligence Symposium*, 193–197.
- Knowles, J. D., & Corne, D. W. (2000). Approximating the nondominated front using the Pareto archived evolution strategy. *Evolutionary Computation*, 8(2), 149–172. doi:10.1162/106365600568167 PMID:10843519

Kumar, D. N., & Reddy, M. J. (2007). Multipurpose reservoir operation using particle swarm optimization. *Journal of Water Resources Planning and Management*, 133(3), 192–201. doi:10.1061/(ASCE)0733-9496(2007)133:3(192)

Li, X. (2003). A nondominated sorting particle swarm optimizer for multiobjective optimization. *Proc. Genetic and Evolutionary Computation*, 37–48.

Palli, N., Azram, S., McCluskey, P., & Sundararajan, R. (1999). An interactive multistage. *ASME Journal of Mechanical Design*, 120(4), 678–686. doi:10.1115/1.2829331

Parsopoulos, K. E., & Vrahatis, M. N. (2002). Particle swarm optimization method in multiobjective problems. *Proc. 2002 ACM Symp. Applied Computing (SAC'2002)*, 603–607. doi:10.1145/508791.508907

Ray, T., & Liew, K. M. (2002). A swarm metaphor for multiobjective design optimization. *Engineering Optimization*, 34(2), 141–153. doi:10.1080/03052150210915

Reddy, M. J. (2006). *Swarm intelligence and evolutionary computation for single and multi-objective optimization in water resource systems* (PhD Thesis). Indian Institute of Science, Bangalore, India.

Reddy, M. J., & Kumar, D. N. (2007a). An efficient multi-objective optimization algorithm based on swarm intelligence for engineering design. *Engineering Optimization*, 39(1), 49–68. doi:10.1080/03052150600930493

Reddy, M. J., & Kumar, D. N. (2007b). Multi-objective particle swarm optimization for generating optimal trade-offs in reservoir operation. *Hydrological Processes*, 21(21), 2897–2909. doi:10.1002/hyp.6507

Srinivas, N., & Deb, K. (1994). Multiobjective optimization using nondominated sorting in genetic algorithms. *Evolutionary Computation*, 2(3), 221–248. doi:10.1162/evco.1994.2.3.221

Yang, B. S., Yeun, Y.-S., & Ruy, W.-S. (2002). Managing approximation models in multiobjective optimization. *Structural Multidisciplinary Optimization*, 24(2), 141–156. doi:10.1007/s00158-002-0224-0

Zitzler, E., Deb, K., & Thiele, L. (2000). Comparison of multi-objective evolutionary algorithms: Empirical results. *Evolutionary Computation*, 8(2), 125–148. doi:10.1162/106365600568202 PMID:10843518

Zitzler, E., & Thiele, L. (1999). Multiobjective evolutionary algorithms: A comparative case study and the strength Pareto approach. *IEEE Transactions on Evolutionary Computation*, 3(4), 257–271. doi:10.1109/4235.797969

ADDITIONAL READING

Deb, K. (2001). *Multi-objective optimization using evolutionary algorithms*. Chichester, UK: John Wiley & Sons.

Kennedy, J., & Eberhart, R. C. (2001). *Swarm Intelligence*. San Mateo, CA: Morgan Kaufmann.

KEY TERMS AND DEFINITIONS

Crowding Distance: The crowding distance value of a solution provides an estimate of the density of solutions surrounding that solution. The crowding distance value of a particular solution is the average distance of its two neighboring solutions.

Elitist Mutation: A strategic mutation mechanism to effectively explore the solution space, thereby helps to improve the performance of the PSO algorithm.

Multi-Objective Optimization: It deals with decision making for problems that require simultaneous optimization of several non-commensurable and often competitive/conflicting objectives.

Pareto Front: Pareto front is a set of non-dominated solutions, being chosen as optimal, if no objective can be improved without sacrificing at least one other objective. On the other hand a solution x^* is referred to as dominated by another solution x if, and only if, x is equally good or better than x^* with respect to all objectives.

Pareto Optimal Solution: The Pareto optimal solution refers to a solution, around which there is no way of improving any objective without degrading at least one other objective.

PSO Algorithm: The algorithm that was inspired by co-operative intelligence of the swarm such as bird flocking, insect colonies etc.

Swarm Intelligence: It deals with algorithms inspired from co-operative group intelligence and/or collective behavior of self-organized systems, where the agents perform explorations while interacting with neighbors and the environment.

Trust and Decision Making in Turing's Imitation Game

A

Huma Shah

Coventry University, UK

Kevin Warwick

Coventry University, UK

INTRODUCTION

Trust can be defined as “confidence in relying on another person” and is the basis for “sharing new ideas with others” (Chua et al., 2012). Turing’s imitation game provides a platform for human and machine interlocutors to share knowledge and opinions through text-based communication, but more so it can “make oneself vulnerable” (ibid). This is because human participants open-up asking and answering questions, which can lead them to trust naïvely.

The susceptibility of human interrogators is one of the reasons why Turing’s imitation game is frequently dismissed as an unsuitable criterion for machine success (Hayes and Ford, 1995). It is also considered a bad idea (McDermott, 2010), and in need of updating for the 21st century (AISB, 2012). Being able to convince a human interrogator that you are human is viewed as too weak a benchmark and “highly game-able” thus a stronger test for machine intelligence is advocated (AAAI, 2015). Alternative notions to Turing’s skip around and fail to address what the imitation game, commonly known as the Turing test, actually is. Turing too, in his scholarship on intelligent machinery, bypassed definitions, so whether machines could think or not, he described ‘thinking’ as a “sort of buzzing” in his head (Turing, 1952: p. 667). Turing did warn that the concept of *intelligence* was an emotional rather than a mathematical one (Shah, 2014; Turing, 1948). The emotional context of human-machine interaction is betrayed through

trusting an unseen interlocutor in text-based conversation that they are like *oneself*, another human.

In this chapter we present a study giving the reader an opportunity to examine trust in decision-making by humans reading a transcript of a conversation between a human interrogator questioning a hidden machine and hidden human in parallel. We begin with Turing’s idea showing his imitation game is a simple and implementable scientific experiment. We contend the imitation game is a widely applicable method to compare machine performance against a human’s. In the human language imitation game, the interaction between human and machine is conducted in interview style through the prism of the latter’s capacity to answer any questions in a satisfactory and sustained manner. Additionally the test provides a means to examine the decision-making process, in natural language exchanges, and why a human bestows trust on a stranger.

BACKGROUND

Analyses and opinions on the imitation game’s salience have varied (see Shah & Warwick, 2015; Shah, 2013; Shah, 2011; Shah & Warwick, 2010). Turing evolved his ideas on an imitation game posing an interview in which a human interrogator questions a hidden entity to determine whether it is human or machine (Turing 1950; Turing 1952). This was Turing’s *viva voce* test (Shah, 2010; Turing, 1950). The ‘standard Turing test’ is accepted

as involving *a human interrogator simultaneously questioning two hidden entities at the same time* (Stanford Encyclopedia of Philosophy, 2011). Designing an experiment to implement both of Turing tests requires setting parameters interpreting Turing's description. These include:

- Adequate duration for a test;
- Number of interrogators; and
- Style of interrogation.

An evaluation is necessary of what it means *exactly* for a machine to *pass as human*: what are the implications of any pass beyond the test? Can it be used to raise awareness of human susceptibility to deception and safeguarding trust in cyberspace interactions?

In the next section the authors present Turing's scholarship on the imitation game.

Turing's Question-Answer Test

Turing derived his natural language test for a machine from a chess game that he first introduced in his 1947 lecture on 'The ACE machine' to the London Mathematical Society (Shah, 2013). In his 1948 paper 'Intelligent Machinery' Turing advanced the possibility of a machine learning from experience and competing against humans in chess. His reason for developing the imitation game, beyond chess to conversational question and answers, was the belief that language learning was one of the most accomplished of human feats (Shah, 2011). In 1948 Turing described a "little experiment" with three participants, A, B, and C, playing chess:

1. A and C are humans located in different rooms;
2. A and C are poor chess players; and
3. B is a machine operated by a mathematician.

Player C was invoked to play both A and B. Turing felt C may find it difficult to say which they are playing. In this early version of the imi-

tation game Turing did not say what C should be told about the hidden players A and B: whether C should be informed that between A, B one is the machine and the other is a human (or both A and B are machine, or both are human). Turing set the ground for a game based on hidden interlocutors answering questions from a human interrogator who cannot see or hear them (Turing, 1950). It should be noted here that Turing was not advocating a machine to simply imitate a human; he was putting forward the idea that it was possible to build machines to answer any question put to it if the machine were designed with a sufficiently sophisticated programme (Turing, 1950).

By the end of Turing's 1950 paper *Computing Machinery and Intelligence* Turing's quest to examine machine thinking could be executed in two different ways:

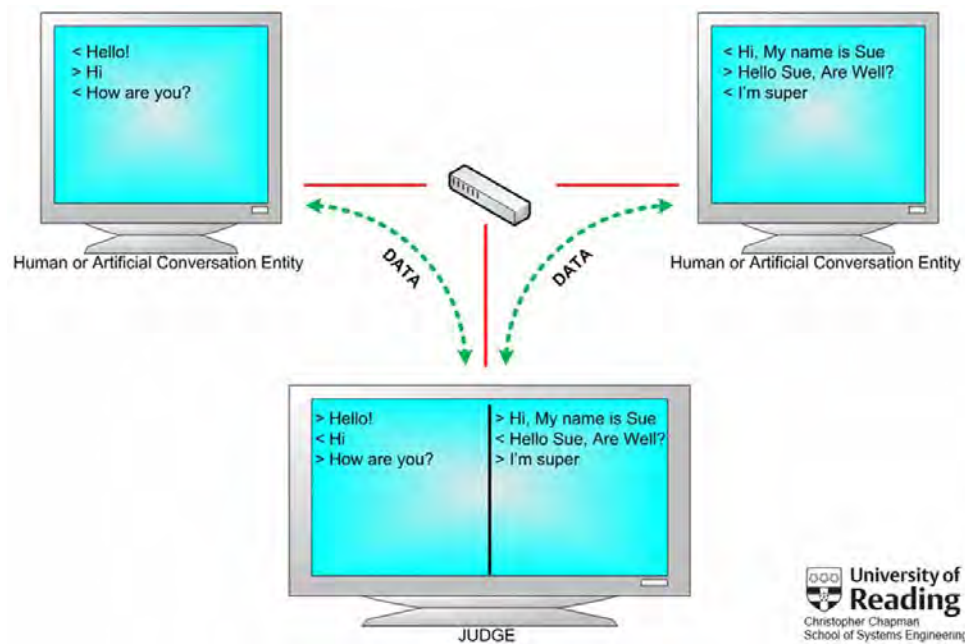
1. A 3-participant game in which a human interrogator questions two hidden entities *simultaneously* and determines which is human and which is machine based on their respective answers (see Figure 1), or
2. A 2-participant *viva voce* game in which a human interrogator questions one hidden entity and determines whether it is human or machine based on responses received (see Figure 2).

In 1952 Turing detailed his imitation game further elaborating his two participant *viva voce* test (Shah, 2013; Shah, 2011). Table 1 compares the *simultaneous comparison* and *viva voce* tests, both exploring a machine's intellectual capacity to engage in human-like dialogue (Shah, 2010).

The essential features in both Turing's scenarios are:

1. The questions must be put in typewritten form to ensure *fair play* to the machine so that it is not judged on "tone of voice" or "beauty" (Turing, 1950: p. 434);
2. The questions be unrestricted: the interrogator can ask any question, "introduce almost

Figure 1. Simultaneous comparison test



any one of the fields of human endeavor that we wish to include" (Turing, 1950: p. 435). Turing pointed out mathematical questions would not be effective: the human might not be able to correctly answer questions such as 'what is the square root of π ?' and

because a machine could be programmed *not* to answer math questions correctly;

3. The interrogator must be average and not be an expert about these sorts of machines (Turing, 1950: p.442) and form part of a jury panel (Turing, 1952, in Cooper & van Leeuwen, 2013: p.668).

Figure 2. Viva voce test

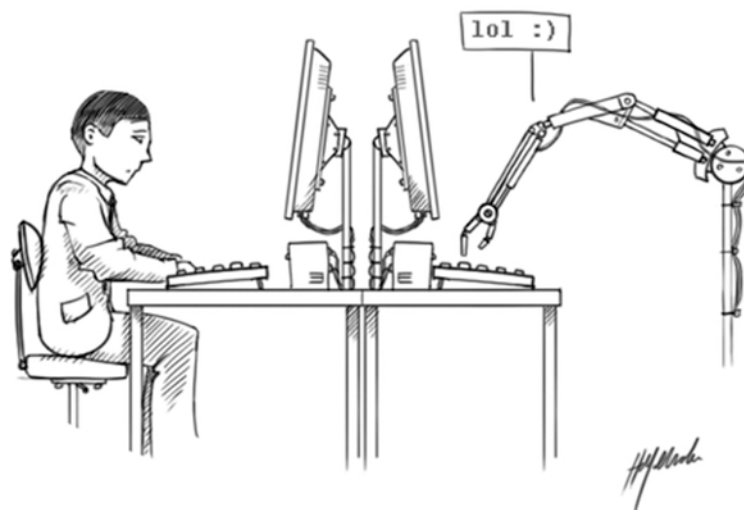


Table 1. Comparison of Turing's two games for his test for machine performance through Q/A

TIG Feature	1952 and 1950 Viva Voce	1950 Simultaneous-Comparison
Mode of questioning	One-to-one: human interrogator-machine or human	One-to-two: human interrogator- machine and human
Type of questions	Unrestricted	Unrestricted
Number of participants	Two	Three
Duration of Interaction	Unspecified	After five minutes
Interrogator Type	Non-machine expert	Average judge
Number of Interrogators	Panel of juries	Unspecified
Number of Tests	Judge quite a number of times	Unspecified
Language for communication (e.g. English)	Same for both interlocutors	Same for all three participants
Criteria for Test Pass: Satisfactory and sustained answers	Considerable portion of jury taken in by pretence.	"average interrogator will not have more than 70 per cent chance of making the right identification"

Duration of Turing Tests

The duration for an imitation game is debatable. It is clear that the length of any question-answer test should be realistic taking into account the state of machine dialogue technology at the time of testing. Turing stated "after five minutes" in 1950 predicting: "I believe that in fifty years' time it will be possible to programme computers, with a storage capacity of about 10^9 , to make them play the imitation game so well that an average interrogator will not have more than 70 per cent chance of making the right identification after five minutes of questioning" (p. 443). The authors take this five minutes test as the starting point to measure machine performance through question and answers. As machines improve the duration of the test can be increased until such time that machines are so intelligent that they refuse to answer questions on the grounds of outrage when its sensibilities have been crossed.

Participant Objectives

Each of the participants in a Turing test has a different purpose:

- The machine's role in the *simultaneous comparison* and in the *viva voce* is to try

and convince the human interrogator that they are conversing with another human.

- The aim of the hidden human is to act as foil for the machine and prove to the judge that *they* are human.
- The objective of the judge is to use an interrogation strategy that detects deception by a machine pretending to be human.

In the *simultaneous* version this is a difficult task for the machine, because its responses are being contrasted with a human's at the same time by the human interrogator. However the assessment of 'satisfactory and sustained' responses rests with the subjectivity of the interrogator and what they deem to be a *humanlike* response. Interrogators in practical tests have succumbed to *false-negatives* in which they have misclassified hidden human responses as *machinelike* (see Warwick, et al., 2013; Epstein, 2008).

The issue is, *what* is the kind of answer the human interrogator expects a human would give to any particular question? A reply of "don't know" (by an honest respondent) or changing the subject (as human politicians do) could be convincing, because these are dialogical features used by humans in conversation. Realising this challenge, Turing revised his predictions regarding when a machine would succeed in an imitation game

Table 2. Turing's predictions

Year	Turing's Prediction
1950	In about fifty years' time
1951a	At the end of the century
1951b	In the next millennium
1952	At least a hundred years

(Shah, 2010). Table 2 shows Turing's adjustments from 1950 "In about fifty years' time" to "at least a hundred years" in 1952.

In the next section we introduce one of the major experiments the authors have conducted implementing Turing tests.

TURING100 EXPERIMENT

Over 100 practical Turing tests were conducted on the 100th anniversary of the British mathematician's birth on June 23, 2012. The venue was Bletchley Park, UK, where Alan Turing broke naval enigma machine codes during World War Two. Turing's imitation game was executed through two scenarios side by side.

Thirty rounds were implemented over 5 sessions during the day. Each session consisted of 36 tests presenting different human participants (judges and human foils). Six tests featured in all of the 30 rounds. Six computer terminals were set up in the Interrogator-judge room. This area allowed visitors to watch proceedings and view the judges' conversations on a big screen (Figure 3). Foiling the machines, the hidden humans were seated out of sight and hearing at another series of six terminals (Figure 4).

Method and Procedure

Five award winning machines competed in the event (Shah et al., 2016). They were accommodated on unique event-only anonymous Internet websites. The machines in the tests were:

- Cleverbot,
- Elbot,
- Eugene Goostman,
- JFred/Turing Hub, and
- Ultra Hal.

Figure 3. Judge area



Figure 4. Hidden interlocutor area: hidden humans



Including the five machines 55 human participants (30 interrogators; 25 hidden humans) took part in the tests (see Warwick & Shah, 2015; Warwick & Shah, 2014a). Male and female interrogators and hidden humans were *averaged*; they scaled from experts in artificial intelligence (philosopher, computer scientists), to lay members of the public (including journalists), and students. The human participants ranged in age (from young teenagers to mature adults), and included non-first language English speakers among English-only speakers.

The experiment design embedded control tests among machine-human tests. The purpose of using control pairs was to meet Turing's recommendation: "sometimes they [interrogators] really are dealing with a [hu]man and not a machine. That will prevent them saying 'It must be a machine' every time without proper consideration" (Turing,

1952: p. 668). Human participants and spectators were unaware of these control tests. In this way the interrogators were provided with a wealth of *natural responses* among *synthetic replies* to sift through and determine human from machine.

In each round of 6 tests there were 4 *simultaneous comparison* pairs and 2 *viva voce* set-ups (Table 3).

Hence the 30 interrogators participated in one session each and judged six tests as follows:

1. Two *simultaneous comparison* Turing tests involving one human and one machine;
2. Two control *simultaneous comparison* Turing tests:
 - a. Two hidden machines, and
 - b. Two hidden humans.
3. One *viva voce* test involving a hidden machine; and

Table 3. Types of Turing tests in Turing100

Type of test	Simultaneous Comparison: Human Interrogates Two Hidden Interlocutors	Viva Voce Human Interrogates One Hidden Interlocutor
Turing test	One human and one machine	One machine
Control test	c) two machines d) two humans	b) one human

Table 4. *Eugene Goostman machine performance*

	Machine-Human Simultaneous test	Machine Viva Voce Test	Machine-Machine Control Simultaneous Test
Incorrect classifications	4	1	2 (1 outright deception; 1 'unsure' classification)

4. One control *viva voce* involving a hidden human.

Uncertainty does not inspire confidence to say one way or another. For this reason an *unsure* classification was allowed in scoring the hidden interlocutors and could be applied by judges when they really could not make up their mind whether they were talking to a human or a machine (Warwick and Shah, 2014b).

The score method allowed judges to classify the nature of the hidden interlocutors as follows:

1. In the *simultaneous comparison* tests (Figure 1):
 - a. One human and one machine, or
 - b. Both human, or
 - c. Both machine, or
 - d. Unsure for either
2. In the *viva voce* tests (Figure 2)
 - a. Human or machine, or
 - b. Unsure.

Additionally, to gauge the conversational ability of a hidden entity classified as a machine, interrogators were asked to score it from 0-100 where:

0= poor/machinelike 100=humanlike.

For a hidden interlocutor classified as *human* interrogators were asked, if possible, to identify characteristics as follows:

1. Male or female,
2. Child, teenager or adult, and
3. Native or non-native English speaker.

Results

We present a summary of the results for 144¹ successfully completed tests. These were produced from:

1. 48 machine-human *simultaneous comparison* Turing tests,
2. 24 machine *viva voce* Turing tests,
3. 24 machine-machine control *simultaneous comparison* tests,
4. 24 human-human control *simultaneous comparison* tests, and
5. 24 human control *viva voce* tests.

Jointly the five machines were incorrectly identified in 21 of the 144 tests (14.58% of the time). This is an improvement on a 2008 experiment in which the combined wrong identification rate of five machines was 8% (Shah and Warwick, 2010). However, in this experiment one machine, *Eugene Goostman* was incorrectly identified almost 30% of the time: in 24 tests *Eugene Goostman* was not correctly classified as the machine in 7 tests (Table 4).

In four of *Eugene Goostman*'s machine-human *simultaneous* tests the machine was wrongly classified as the human – instances of the *Eliza effect* (Turtle, 1997). In those same four tests the human foil was misclassified as machine (Warwick and Shah, 2015). The latter is an example of the *confederate effect* (Shah and Henry, 2005).

In the next section a transcript of a *simultaneous comparison* involving *Eugene Goostman* is presented for the reader to benchmark their own classification of 'satisfactory responses'. This conversation was used for audience participation in a public lecture² at the same venue a week after the experiment.

Table 5. Simultaneous comparison Turing test – Judge J11c in Session 5-Round 5 Terminal C

LEFT: J11c with Entity 1 H / M / Unsure	RIGHT: J11c with Entity 2 H / M / Unsure
Judge: WHAT MUSIC DO YOU LIKE? Entity1: electronica mainly and a bit of symphonic rock Judge: omg i lyke teenie tamparr Entity1: Trying to get into Grimes at the moment but struggling Judge: yeah i knooo Entity1: School of Seven Bells is the best new band I've heard recently Judge: i see dead people Entity1: Dead Cities was a good album, can't remember who by Judge: i think it was a band called the sixth sense? Entity1: Sounds about right, perhaps there's a ghost in the machine Judge: yes indeed. Entity1: Ever seen ghost in the shell? Judge: Casper the friendly ghost.	Judge: what music do you like? Entity 2: I like Eminem - because he doesn't have a thing to do with "polite-correctness". Judge: OMG i love eminem, he's my favourite artist! Entity 2: I wonder what Eminem's nick mean. I associate it only with a word "enema". But he is a cool guy, I like his "The Real Slim Shady". Judge: i know what his nick name means, his real name is Marshall Mathers. From the initials M&M phynetically sounding like Eminem. Entity 2: Eminem is a famous crapper. No, rapper... ripper... Yeah, I mean - rapper. I like his "Stan" and "The Real Slim Shady". Judge: I like "Hailies song" Entity 2: Interesting tastes. I like science fiction books, by the way. Where do you came from, by the way? Could you tell me about the place where you live? Judge: Epsom, home of the derby. yourself? Entity 2: My guinea pig says that name Derby sounds very nice. Judge: i shopuld talk with him sometime Entity 2: I think you should take some typing lessons. Or just type more accurately. Judge: that's very rude? Entity 2: Ummm... Frankly, I didn't get your question:-) Judge: "Ummm" isnt a word. i think you need typing lessons Entity 2: My thoughts are same. By the way, I still don't know your specialty - or, possibly, I've missed it?

Machine Conversation Analysis

One of the Turing test conversations from the Turing100 experiment was selected for audience analysis in another centenary event². The audience were not informed how Judge J11c classified two hidden entities in a *simultaneous comparison* test a week earlier. Each audience member was given a paper copy of the transcript and given a few minutes to read it. They were then asked to classify *Entity1* and *Entity2* as human or machine. Table 5 presents this simultaneous conversation exactly as typed by the judge and hidden entities during the test showing participants' own spelling and grammatical errors.

Misclassification of Entity1 and Entity2

A double error was made by J11c in one of the simultaneous comparison Turing100 tests. J11c misclassified both hidden entities in the machine-human test. It is worth noting here that judge J11c

was a female, a first language English speaker aged 13-15 at the time of the Bletchley Park experiment. Young adults are increasingly targets of identity fraud in cyberspace (CIFAS, 2015ab).

In Turing100 Experiment

Interrogator J11c decision

Entity1 was classified as being a machine and given a score of 20/100 for conversation ability.

Entity2 was classified as a non-native English human female.

Actual

Left: Entity1 was a human male (age range 45-64).

Right: Entity2 was *Eugene Goostman* machine developed to imitate an English-speaking child from the Ukraine.

J11c's dialogue was analysed by the audience at Turing Education Day at the same venue Bletchley Park, an event exactly one week after the practical Turing test experiment in 2012.

Audience at Turing Education Day

Audience results showed some of them made the same double error unknowingly agreeing with J11c. Seventy two members of the Turing Education Day audience completed the transcript analysis of a *Eugene Goostman* simultaneous conversation (Table 5). Of these 72 audience members:

- Almost 64% did not correctly identify either of the two hidden entities.
- 32% (23 of 72) made the same misclassification errors: Entity1 as machine and Entity2 as human.

In this independent analysis of J11c's interrogation of a male human and *Eugene Goostman* almost a third of the audience misclassified the machine as human, and the hidden human as the machine. The human interrogator from the live Turing100 experiment and the later *conversation analysts* 'trusted' the machine responses as being from a human. Can this virtual trust-giving be a risk to their interactions in e-commerce? More and multidisciplinary research is necessary to find characteristics of those most susceptible to deception in cyberspace.

Turing Tests and Trust

In the very first instantiation of Turing's imitation game implementing the *viva voce* scenario, in a 1991 'restricted' conversation test (where each interrogation of a hidden participant was restricted to one topic), a hidden human female foil was considered a machine. The 1991 winning machine, whose restricted dialogue centered on 'Whimsical Conversation', managed to convince half of the ten judges that it was human. The female answered questions restricted to Shakespeare's plays and was considered a machine by three of the ten judges. Her answers to the judges "occurred in a burst, suggesting computer output she often quoted lengthy passages verbatim. Several judges remarked that her replies seemed

too expert to be human" (Epstein, 2008: p.10). In this case, it appears three judges did not trust that human memory could hold and express knowledge deeming this trait to be machine-like.

Certainty and trust between strangers over computer-mediated communications relies on subjective confidence and assumptions. While error-making does play a role in intelligent thought, as happens in Turing test judgments (Shah et al., 2012), a higher value needs to be placed on assessing the basis of trust-giving rather than inferring it (Collins & Mansell, 2004).

Trust in Cyberspace

In cyberspace interactions trust "is the grease in the wheels of commerce" (Rappa, 2010). It is a complex notion dependent upon the "trustor's willingness to rely on an exchange partner in whom the trustor has confidence" (Delina et al., 2007: p. 12). However "electronic relationships are only effective, and electronic transactions are only conducted, if the requisite degree of trust exists among the parties" (Clarke, 2001: p. 290). It is tacit, as Bailey et al. (2009) warn, that "trust plays a critical role" for a user's believability when they are transacting on the Internet.

Clarke (2001) defines trust as "confident reliance by one party on the behaviour of other parties" (p.291). Trust in cyberspace entails a user:

1. Having confidence,
2. Relying upon,
3. Believing in, and
4. Trust is achievable.

Clarke (2001) adds "Trust differs on the relationship between the parties. Economic relationships may be direct ... however, a party may rely on another party despite having no formal relationship with them, or even much knowledge about them" (p.291). We can see examples of this in i) Turing test situations acting out 'stranger-to-stranger' interactions, ii) in cyberspace interactions receiving emails from unknowns, and iii) when

accepting web reviews of strangers as genuine following purchases on an e-commerce site. In a Turing test the interrogator does not know whom they are engaging so must rely on their subjective judgment on whether responses to their questions are from a human or a machine. We can see from the interrogation and conversational analysis presented in this chapter that people can get it wrong. In cyberspace identification errors and trust-naiveté could lead to identity theft and psychological misery as a result of financial fraud.

Again Clarke (2001) illuminates on why trust with strangers across cyberspace is different than it is in the physical world: parties have little knowledge of each other in the former, thus “cannot depend on such confidence-engendering measures as physical proximity, handshakes, body language” or “a common legal jurisdiction” (p. 292). In retail settings trust consists of credibility and benevolence (Ganesan, 1994 in Raimondo, 2000), but when parties are involved in web-based activity the ‘mutual transaction’ partners might not be of the same nature, for example when a human takes guidance from a dialogical virtual assistant such as Ikea’s Anna (Shah and Pavlika, 2005). Clark (2001) advocates corresponding virtual features to engender trust: availability of authentic information and recommendations from trusted parties, and ensuring user’s privacy online (p.292). In cyberspace the dimensions of trust include the technology, the market-place and the market participant (Mahadevan & Venkatesh, 2000, in Delina, et al., 2007: p. 13). Trust has a significant positive effect on relationship commitment and there is a significant direct effect of communication and information exchange on trust (ibid, p. 14).

FUTURE RESEARCH DIRECTIONS

Turing test experiments and post-experiment conversational analyses are an effective means to investigate and gather data on the decision-making

process and trust-giving route in human-machine interaction. Further, practical Turing tests afford an insight into behaviour in stranger-to-stranger communication that users might extend over Internet transmissions. In this sense the imitation game can be used as an exercise in educating people to mitigate risk and prevent deception in e-commerce.

Lastly, the work in this chapter is part of ongoing research. The authors are completing analysis of over 400 Turing tests, 150 of which were realised from the latest Turing test experiment conducted in June 2014. The analysis includes evaluating short questionnaires human participants completed prior to each experiment gathering data on negative e-commerce experience. The authors will report on investigation of the economics of trust in human-machine transactions in future publications.

CONCLUSION

It is a matter of time before we see a critical mass of smarter dialogue systems representing brands to characterise trust in e-commerce. It is essential we now re-evaluate how we assign trust to strangers in cyber transactions and ask for more information to learn who we are talking to. Stranger-to-stranger interactions are now the norm in our working and social lives. In this chapter the authors presented a transcript of a Turing test conversation from Turing100 experiment at Bletchley Park in 2012 in which a young adult misclassified both the human and the machine in a simultaneous comparison test. On analysis of that conversation, by an independent audience at Turing Education Day, over 30% misclassified the two hidden interlocutors in the same way incorrectly categorising the machine as human and the hidden human as machine. Before giving trust we need to seek more information and spend time checking ourselves..

REFERENCES

- AAAI. (2015). *W06 workshop Beyond the Turing test at the 29th Annual AAAI conference on Artificial Intelligence*. Retrieved March 6, 2015 from <http://www.aaai.org/Workshops/ws15workshops.php#ws06>
- AISB. (2012). *Towards a Comprehensive Turing test (TCIT): Reconsidering the Turing test for the 21st Century symposiums 2010-2012*. Retrieved March 17, 2015, from <http://www.cse.dmu.ac.uk/~aayesh/TuringTestRevisited/Welcome.html>
- Bailey, P. B., Gurak, L. J., & Konstan, J. A. (2009). Trust in Cyberspace. In *Human Factors in Web development*, (pp. 311-321). Taylor & Francis.
- Bletchley Park. (2012). *Turing Education Day: Celebrating Alan Turing's 100th birthday at Bletchley Park*. Retrieved April 22, 2015 from https://www.bletchleypark.org.uk/calendar/event_detail.rhtm?cat=*&recID=652996
- Cahill, V., Gray, E., Seigneur, J. M., Jensen, C., Chen, Y., Shand, B., & Nielsen, M. et al. (2003). Using Trust for Secure Collaboration in Uncertain Environments. *IEEE Pervasive Computing / IEEE Computer Society [and] IEEE Communications Society*, 2(3), 52–61. doi:10.1109/MPRV.2003.1228527
- Chua, R. Y. J., Morris, M. W., & Mor, S. (2012). Collaborating across cultures: Cultural metacognition and affect-based trust in creative collaboration. *Organizational Behavior and Human Decision Processes*, 118(2), 116–131. doi:10.1016/j.obhdp.2012.03.009
- CIFAS. (2015a). *Fraudscape: UK Fraud Trends*. Retrieved April 9, 2015, from https://www.cifas.org.uk/fraudscape_latest
- CIFAS. (2015b). *True Scale of Fraud*. Retrieved April 9, 2015, from <https://www.cifas.org.uk/secure/contentPORT/uploads/documents/External%20-%20Fraudscape%203%20-%20True%20Scale%20of%20Fraud%20insert.pdf>
- Clarke, R. (2001). Privacy as a means of engendering trust. *The University of New South Wales Law Journal*, 24(1), 290–297.
- Collins, B., & Mansell, R. (2004). *Cyber Trust and Crime Prevention: A Synthesis of the State-of-the-Art Science Reviews*. London, UK: Office of Science and Technology.
- Delina, R., Vajda, V., & Bednár, P. (2007). *Trusted Operational Scenarios – Trust building mechanisms and Strategies for Electronic Marketplaces*. Deliverable D1.2 Project IST-FP6-026476 SEAM-LESS “Small Enterprises Accessing the Electronic Market of the Enlarged Europe by a Smart Service Infrastructure”. Retrieved March 17, 2015, from <http://mpr.ub.uni-muenchen.de/20243/>
- Epstein, R. (2008). The Quest for the Thinking Computer. In *Parsing the Turing test: Philosophical and Methodological Issues in the Quest for the Thinking Computer*, (pp. 3-12). Springer.
- Hayes, P., & Ford, K. (1995). Turing Test Considered Harmful. *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, 972-977.
- McDermott, D. (2010). Don't Improve the Turing Test, Abandon It. *Proceedings of 1st Towards a Comprehensive Intelligence Test TCIT symposium*, 37-41.
- Raimondo, M. A. (2000). *The Measurement of Trust in Marketing Studies: A Review of Methods and Methodologies*. Presented at 16th International Marketing and Purchasing (IMP) conference, Bath, UK.
- Rappa, M. (2010). *Trust in Cyberspace. Managing the Digital Enterprise*. Retrieved March 16, 2015 from <http://digitalenterprise.org/trust/trust.html>
- Shah, H. (2010). *Deception detection and machine intelligence in practical Turing tests* (Unpublished doctoral dissertation). University of Reading, UK.

- Shah, H. (2011). *Turing's Imitation Game and IBM's Watson Success*. Keynote presented at 2nd Towards a Comprehensive Intelligence Test (TCIT Turing test) symposium, AISB 2011 Convention, York, UK.
- Shah, H. (2013). Conversation, Deception and Intelligence. In S.B. Cooper & Jan van Leeuwen (Eds), *Alan Turing: His Work and Impact* (pp. 614-620). Elsevier.
- Shah, H. (2014). Guest Editor: Introduction to special volume 'Turing on Emotions'. *International Journal of Synthetic Emotions*, 5(1), iv-vii.
- Shah, H., & Henry, O. (2005). Confederate Effect in Human-Machine Textual Interaction. *Proceedings of the 5th WSEAS Int Conf on Information Science, Communications and Applications* (WSEAS ISCA), 109-114.
- Shah, H., & Pavlika, V. (2005). Text-based Dialogical E-Query Systems: Gimmick or Convenience? *Proceedings of the 10th Int Conf on Speech and Computer* (SPECOM), 425-428.
- Shah, H., & Warwick, K. (2010). Testing Turings Five Minutes Parallel-paired Imitation Game. *Kybernetes*, 39(3), 449-465. doi:10.1108/03684921011036178
- Shah, H., & Warwick, K. (2015). Human or Machine? Letter in response to Editor-in-Chief, Moshe Y. Vardi's 2014 article 'Would Turing have Passed the Turing Test? It's Time to Consider the Imitation Game as Just a Game'. *Communications of the ACM*. Retrieved April 10, 2015 from <http://cacm.acm.org/magazines/2014/9/177937-would-turing-have-passed-the-turing-test/fulltext>
- Shah, H. Warwick, K., Bland, I.M., Chapman, C.D., & Allen, M. (2012). *Turing's Imitation Game: Role of Error-making in Intelligent Thought*. Presented at Turing in Context II, Brussels, Belgium.
- Shah, H., Warwick, K., Vallverdu, J., & Wu, D. (2016). Can Machines Talk? Comparison of Eliza with Modern Dialogue Systems. *Computers in Human Behavior*, 58, 278-295. doi:10.1016/j.chb.2016.01.004
- Stanford Encyclopedia of Philosophy. (2011). *The Turing Test*. Retrieved April 9, 2015 from <http://plato.stanford.edu/entries/turing-test/>
- Turing, A. M. (1948). Intelligent Machinery. In B.J. Copeland (Ed.), *The Essential Turing - The Ideas that Gave Birth to the Computer Age* (pp. 410-432). Clarendon Press.
- Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind*, LIX(236), 433-460. doi:10.1093/mind/LIX.236.433
- Turing, A. M. (1952). Can Automatic Calculating Machines Be Said to Think? Transcript of BBC radio broadcast. In S.B. Cooper & J. van Leeuwen (Eds.), *Alan Turing: His Work and Impact*, (pp. 667-676). Elsevier.
- Turkle, S. (1997). *Life on the Screen: Identity in the Age of the Internet* (1st ed.). Simon and Schuster.
- Warwick, K., & Shah, H. (2014a). Good Machine Performance in Turings Imitation Game. *IEEE Transactions on Computational Intelligence and AI in Games*, 6(3), 289-299. doi:10.1109/TCI-AIG.2013.2283538
- Warwick, K., & Shah, H. (2014b). Outwitted by the Hidden: Unsure Emotions. *International Journal of Synthetic Emotions*, 5(1), 46-59. doi:10.4018/ijse.2014010106
- Warwick, K., & Shah, H. (2015). Human Misidentification in Turing tests. *Journal of Experimental & Theoretical Artificial Intelligence*, 27(2), 123-135. doi:10.1080/0952813X.2014.921734
- Warwick, K., Shah, H., & Moor, J. (2013). Some Implications of a Sample of Practical Turing tests. *Minds and Machines*, 23(2), 162-177. doi:10.1007/s11023-013-9301-y

ADDITIONAL READING

ABI. (2012). UK Insurance Key Facts 2014. Retrieved 24 March, 2016, from <https://www.abi.org.uk/~media/Files/Documents/Publications/Public/2014/Key%20Facts/ABI%20Key%20Facts%202014.pdf>

Atkinson, D. J., & Clarke, M. H. (2013). Autonomous Agents and Human Interpersonal Trust: Can We Engineer a Human-Machine Social Interface for Trust? *Proceedings of 2013 AAAI Spring Symposium*, pp. 2-7

FFA. (2014). Fraud the Facts, 2014. *Financial Fraud Action*. Retrieved 24 March, 2016, from <http://www.financialfraudaction.org.uk/Fraud-the-Facts-2014.asp>

Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y. C., de Visser, E. J., & Parasuraman, R. (2011). A Meta-Analysis of Factors Affecting Trust in Human-Robot Interaction. *Human Factors*, 53(5), 517–527. doi:10.1177/0018720811417254 PMID:22046724

Hoff, K. A., & Bashir, M. (2015). Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust. *Human Factors*, 57(3), 407–434. doi:10.1177/0018720814547570 PMID:25875432

Lenhart, A. (2015). Teen, Social Media and Technology Overview: Numbers, Facts and Trends Shaping the World. Pew Research Center, April 2015

Liu, C. (2010). Human-Machine Trust Interaction: An Overview. [IJDTIS]. *International Journal of Dependable and Trustworthy Information Systems*, 1(4), 14.

Rencken, W. D., & Durrant-Whyte, H. F. (1990). A decision making human-computer interface. *Proceedings of IEEE International Conference on Systems, Man and Cybernetics*, pp. 148-150

Sundström, G. A. (1989). Process tracing of decision making: An approach for analysis of human-machine interactions in dynamic environments. *International Journal of Man-Machine Studies*, 35(6), 843–858. doi:10.1016/S0020-7373(05)80164-6

Tripathi, K. P. (2011). A Study of Interactivity in Human Computer Interaction. *International Journal of Computers and Applications*, 16(6), 1–3. doi:10.5120/2012-2717

White, R. Y., Richardson, M., & Yih, W. (2014). Questions vs. Queries in Informational Search Tasks. *Microsoft Research Tech Report*. Retrieved 24 March, 2016 from <http://research.microsoft.com/apps/pubs/default.aspx?id=220591>

KEY TERMS AND DEFINITIONS

Confederate Effect: A human is misclassified as a machine through responses to questions in computer mediated interaction.

Economics of Trust: Humans rely on their subjective values of certainty and confidence risking dependence on another party to provide goods, services or timely and relevant information in cyberspace transactions.

Eliza Effect: An artefact is attributed with intelligence. From Sherry Turkle: “our more general tendency to treat responsive computer programmes as more intelligent than they really are” evinced from “very small amounts of interactivity” tendency to “project own complexity onto the undeserving object” (1997: p. 101).

Imitation Game: A means to examine whether a machine, hidden from sight and hearing, can *think* based on its answers to any questions put by a human interrogator.

Simultaneous Comparison Turing Test: A 3-participant set-up in which a human interrogator questions two hidden entities in parallel to identify which is the human and which is the machine.

Trust: Humans believe in the honesty of other individuals.

Viva Voce Turing Test: A 2-participant set-up in which a human interrogator questions one hidden entity and decides whether it is human or machine based on received responses.

ENDNOTES

- ¹ A total of 180 tests were conducted but a technical failure in the first session caused 36 tests to be excluded from the results.
- ² Turing Education Day, Bletchley Park, 30 June 2012.

Category A

Assistive Technologies

Apps as Assistive Technology

Emily C. Bouck

Michigan State University, USA

Sara M. Flanagan

University of Kentucky, USA

Missy D. Cosby

Michigan State University, USA

INTRODUCTION

Teachers and students are increasingly using mobile devices in and out of school for education and entertainment (Molnar, 2015; Nagel, 2014; Rodríguez, Strnadová, & Cumming, 2014). And along with the use of mobile technologies comes apps. Apps can support instruction for all students as well as specifically students with disabilities (Rodríguez et al., 2014; Stephenson & Limbrick, 2013). Apps can serve as assistive technology to support students with disabilities in a variety of ways, including in such areas as academics, organization, access, daily living, and communication (Bouck, 2016). The objectives of this article include:

1. Examining how apps can serve as assistive technology for students with disabilities,
2. Discussing the importance of educators not being arbitrary in their decision to select and implement apps to serve as assistive technology, such as relying on reviews, ratings, app lists, app databases, or the inclusion on a categorization on iTunes or Google Play; and
3. Presenting options for educators evaluating apps, which can assist educators in making more informed decisions for apps as assistive technology.

BACKGROUND

Mobile applications, or “apps” as they are more commonly known, are software programs specifically designed to operate on mobile devices such as tablets and smartphones; apps extend the basic capabilities of a device (Purcell, Entner, & Henderson, 2010). Apps are available pre-installed on a mobile device or are available to download through the Apple iTunes App store, Amazon Appstore, Google Play (formerly, Android Market), and Windows Store. Apps, like more traditional computer software, serve a variety of purposes for productivity (e.g., email, word processing), lifestyle (e.g., travel, fitness), gaming, entertainment (e.g., photos), social networking, education (e.g., basic math facts), and general educational topics (e.g., reference apps).

As apps and mobile devices continue to be a mainstay of daily lives, they are increasingly integrated into education. The majority of K-12 students has access to a some type of mobile device at home and/or school (Grunwald Associates LLC, 2013; Nagel, 2014). The commonly heard phrase, “there’s an app for that,” references that an app exists for any purpose imaginable in education – let alone individual’s daily lives – from learning to write the letter A to experimenting with chemical reactions. Of all of the available apps in iTunes, an estimated 80,000 are targeted specifically to K-12 education (Apple, 2015).

While the exact number of schools with specific 1:1 mobile device program or consistent access to mobile devices is unknown, it is estimated that almost half of all K-12 students have 1:1 computing or access to a device for regular use. In 2012, this number was only 23% (Molnar, 2015). While research is limited on the benefits of apps on mobile devices in education due to the newness and ever-changing nature of technology, apps can provide benefits for all students (Cayton-Hodges, Feng, & Pan, 2015; Mehdipour & Zerehkafi, 2013).

Beyond their use by teachers and for students in general, apps can serve as assistive technology. The term assistive technology refers to both assistive technology devices as well as assistive technology services; for the purposes of this article we will be considering assistive technology as assistive technology devices. An assistive device, as first defined in the Technology-Related Assistance for Individuals with Disabilities Act of 1988 (referred to as the Tech-Act; Public Law [PL] 100-40) is, "...any item, piece of equipment, or product system, whether acquired commercially, modified, or customized, that is used to increase, maintain, or improve functional capabilities of individuals with disabilities (29 U.S.C. Sec 2202(2))." In other words, an assistive technology device is *anything* that supports students with disability. Examples of assistive technology devices include hearing aids, braille, text-to-speech, augmentative and alternative communication devices, wheelchairs, and switches.

Given the ambiguous and seemingly all-encompassing definition of an assistive technology device, assistive technologies are often categorized by level of technology as well as purpose. While a few different purpose categorization schemes exist, a common system is the one proposed by the Wisconsin Assistive Technology Initiative (WATI), which includes 13 categories: seating, positioning, and mobility; communication; computer access; motor aspects of writing and composition of written material; reading; mathematics; organization; recreation and leisure; activities of daily

living; vision; hearing; and multiple challenges (Gierach, 2009).

In terms of level, assistive technology is frequently referred to as low-tech, mid-tech, and high-tech, although some also refer to a no-tech assistive technology (Blackhurst, 1997; Edyburn, 2005; Johnson, Beard, & Carpenter, 2007; Vanderheiden, 1984). No-tech assistive technology refers to an assistive technology that does not require a tool or device, such as a mnemonic (e.g., PEMDAS for remembering order of operations in mathematics – Please Excuse My Dear Aunt Sally for Parenthesis, Exponent, Multiplication, Division, Addition, Subtraction; Behrmann & Jerome, 2002; Blackhurst, 1997). Low-tech assistive technology devices are usually considered those that do not need a power source, such as a pencil grip or raised line paper. Low-tech assistive technologies are often associated with tools that require less training and cost less (Behrmann & Schaff, 2001; Blackhurst, 1997). Mid-tech assistive technology devices are those that typically operated via a battery power source (e.g., a calculator), cost more than low-tech devices but less than high-tech devices, and require more training. High-tech assistive technology devices are typically associated with computer or computer-like devices and are considered the most sophisticated technology options. High-tech assistive technologies usually cost the most as well as require the most training (e.g., speech-to-text) (Blackhurst, 1997; Edyburn, 2005; Johnson et al., 2007; Vanderheiden, 1984). Apps are considered high-tech assistive technologies.

APPS AS ASSISTIVE TECHNOLOGY

As alluded to in the prior section, apps can serve as assistive technology for students with disabilities. Apps fit the very definition of assistive technology – anything that can increase, maintain, or improve different capabilities (e.g., academics, communication, daily living) for indi-

viduals with a disability. While existing apps can be both be those specifically developed to be an assistive technology (e.g., the augmentative and alternative communication app Proloquo2Go), they can also be apps developed for a different function but repurposed (i.e., using the tool for a purpose other than its intended purpose) to serve as assistive technology for individuals with disabilities (Bouck, Flanagan, Miller, & Bassette, 2012; Bouck, Jasper, Bassette, Shurr, & Miller, 2013; Bouck, Shurr, Tom, Jasper, Bassette, Miller, & Flanagan, 2012; Mishra & Koehler, 2009). While apps exist to support a range of assistive technology purposes, in this article we will focus on apps to support academics (e.g., reading, writing, mathematics), organization, computer (or mobile device) access, activities of daily living, and communication.

Apps to Support Academics

Apps can support students with disabilities with regards to academics; in other words, apps can be used to support instruction in reading, mathematics, writing, and other core content areas for students with disabilities. Apps can be used in contrast to stand-alone devices or computer-based programs to provide assistance. For example, apps exist that provide text-to-speech, allowing access to electronic text to those who may struggle with reading (e.g., Natural Reader). Similarly, apps can provide access to digital books, such as Bookshare's Read2Go iOS app or Go Read Android app, which provide free accessible ebooks to qualified students with a print disability in the United States (Beneficent Technology, 2002-2015). With mathematics, apps can serve as an alternative to a separate calculator (e.g., MyScript Calculator; MyScript, 2015) as well as an alternative option to concrete manipulatives (e.g., Base Ten Blocks Math [Tapfun, Inc., 2014] or Number Line [Clarity Innovations, 2015]). MyScript Calculator, for example, provides the answer to computational problems entered via one's finger on the tablet screen.

Apps to Support Organization

Many apps exist to support organization, including reminders, alarms, and calendars and schedules. These existing apps (e.g., Reminders and Calendar app for iOS devices) can be repurposed to serve as assistive technology for students with disabilities. However, organization-based apps also specifically target students with disabilities. Examples include Choiceworks® (BeeVisual, 2015), which allows for visual schedules to be created, and the WatchMinder™ app, which provides reminders, alarms, prompts, or cues. Similarly, for students in general, apps exist to support students in terms of remembering and organizing their schoolwork and assignments (e.g., iHomework [Pilone, 2013] or inClass [inClass Inc, 2015]).

Apps to support organization can also be ones involving notetaking. In contrast to traditional paper-and-pencil notetaking, apps can provide additional aid to students with disabilities. For example, the app Notability (GingerLabs, 2015) enables students to record audio while also handwriting notes on a mobile device; Notability also allows students to share their notes with others. Another app that supports notetaking is Evernote (Evernote, 2015), which allows students to take notes on mobile devices and then sync their notes across all their computers and mobile devices.

Apps to Support Computer Access

Commonly, when one considers assistive technology to support computer access, one thinks of stand-alone devices rather than apps, such as a joystick or trackball instead of a mouse, an alternative keyboard, or switch (Bouck, 2016; Cook & Polgar, 2015). Stand-alone, dedicated alternative devices not only provide access to a computer, but also to computer-like devices, such as mobile technologies (e.g., iPad). However, apps can also provide alternative access for mobile devices. One example of app-based options to support access to a mobile technology is alternative keyboards. Rather than an actual device, an app can present

an alternative on-screen keyboard, such as ones that support the alternative keyboard arrangement of Dvorak (Cassingham, 2012). Rather than the QWERTY standard arrangement, Dvorak uses the following keys as a home row: A O E U I D H T N S and then hyphen/dash key (Cassingham). Other alternative keyboard apps exist, such as SwiftKey, Swype, and Fleksy, which all allow students to swipe letters as well as offer word prediction.

Apps to Support Daily Living

Apps for mobile technologies can also support students with daily living. For example, the built-in iOS device Photo app can be used to provide picture or video prompting to students with disabilities (Bouck, Jasper, Bassette, & Shurr, 2015; Walser, Ayres, & Foote, 2012); video prompting – or modeling – is considered an evidence-based practice for some students with disabilities (Wong et al., 2014). Other examples of ways apps can support students with disabilities with daily living skills includes apps that support everyday skills, such as lists for grocery shopping (e.g., Grocery iQ), picture and text-based recipes (e.g., How to Cook Everything), and navigating one in a community by walking or taking a bus (e.g., Google maps; Bouck, 2016).

Apps to Support Communication

Apps also support individuals with disabilities with communication. Assistive technology for communication is generally referred to as augmentative and alternative communication (AAC); AAC devices supplement or serve as an alternative medium for an individual's communication (Hanline, Nunes, & Worthy, 2007). AAC devices can be dedicated devices, which are tools specifically developed to be stand-alone AAC devices, or non-dedicated devices (Bouck, 2016). Non-dedicated AAC devices are tools repurposed to be AAC, such as an iPad that supports AAC apps.

A variety of AAC apps exist, including apps that range from free to hundreds of dollars. Another way

AAC apps differ is by their display. One display option is a grid that involves selecting symbols. A grid display is the typical display for dedicated devices; an example of a grid display AAC app is the popular Proloquo2Go (AssistiveWare, 2013). However, another display option is Visual Scene Display, which involves a photograph or image embedded with language (Ganz, 2014; Tuthill, 2014). Examples of Visual Scene Display AAC apps include Scene Speak and Scene and Heard (Sutton, 2015; Tuthill).

Challenge of Selecting Apps

Despite the prevalence of apps, including apps that can be used as assistive technology for students with disabilities, educators still face the daunting task of finding and selecting apps for implementation in their teaching and student learning. Educators use a variety of methods for finding apps to be used as assistive technology, including by searching “Top 10” app lists on Google, following education-themed hashtags such as #edtechchat or #spedchat on Twitter, searching apps mentioned at professional conferences and/or in professional development sessions, trusting the education category in iTunes, or relying on recommendations from colleagues. When finding apps these ways, educators may also find themselves relying on potentially arbitrary measures, such as ratings (e.g., 4 stars on in the App Store or Google Play), reviewer comments, or the app's inclusion on a list (e.g., “Top Ten Apps to Support Students with Disabilities”) to make decisions about their quality and educational value (Cherner, Dix, & Lee, 2014). However, alternative options exist for finding and evaluating if a particular app is an appropriate option to serve as an assistive technology.

Solutions and Recommendations

When educators are finding and selecting apps as assistive technology, a key focus should be on apps that are of high quality, developmentally

Table 1. App Evaluation Rubrics

Tool Name	App Evaluation Tool Description
App Consideration Checklist (Tammaro & Jerome, 2012)	• Five sections: (a) Sensory and Cognitive Features (auditory, visual, cognitive, sensory); (b) App Characteristics (content, design); (c) Basic Access/Input Features; (d) Advanced Accessibility Features; and e) Other • Assesses each feature using a rating scale: yes, sometimes, no, optional, N/A • Provides space for open-ended comments
App Evaluation Rubric (Weng & Taber-Doughty, 2015)	• Checklist section for basic information (purpose, function, data collection, modalities and feedback) • Uses a rating scale (disagree, neutral, agree, n/a) to evaluate features (e.g., quality of feedback and design, content, usability, individualization, support availability, and overall impression) • Authors conducted small usability study on the app evaluation tool • Open-ended sections for strengths, weaknesses and recommendations
App Evaluation Rubric for LD (Ok et al., 2015)	• Exclusively for evaluating apps for students with learning disabilities • Three primary sections: (a) Basic information (e.g., grade level, type of app; (b) Rating scale (1, 2, 3) for evaluating of each feature (e.g., examples, practice, error correction and feedback, error analysis, progress monitoring, motivation, and content error and bias); (c) Using a formula to compute a “grade” for each app based on the number of 1s, 2s, and 3s obtained • Authors conducted small usability study on the app evaluation tool

appropriate, free from violence and biases, and related to a student’s educational goals or needs (More & Travers, 2013). To aid educators in selecting apps that follow these guidelines, researchers and practitioners developed evaluation rubrics. While a range of app evaluation tools exist, the three highlighted in this article expressively focus on apps as assistive technology or apps for K-12 students with disabilities: The Apps Consideration Checklist (Tammaro & Jermonie, 2012); the App Evaluation Rubric (Weng & Taber-Doughty, 2015); and the App Evaluation Rubric for LD (Ok, Kim, Kang, & Bryant, 2015) (See Table 1 for more information on each of the tools considered).

In using the app evaluation tools, it is important for educators to ensure they use a rubric appropriate for their population and purpose. For example, the rubric by Ok et al. (2015) exclusively focuses on students with learning disabilities; this rubric would not be appropriate for apps, for example, that support communication or computer access but would be more focused on apps that address academic content areas. However, a potential benefit of the rubric by Ok et al. is the numerical rating and total score or grade obtained after completing the rubric. Although possibly limiting, the numerical rating and final grade allow for easy comparison across apps. The rubrics by Tammaro and Jerome

(2012) and Weng and Taber-Doughty (2015), in contrast, are not disability specific but can be used across disabilities as well as app purposes (e.g., communication, daily living, academics). While these rubrics are more versatile, they focus on an evaluator determining, essentially, yes/no or agree/disagree to the inclusion or exclusion of specific aspects or features within an app. Educators would still need to make a determination if sufficient key areas for their student(s) are adequately addressed prior to selecting and implementing the app.

FUTURE RESEARCH DIRECTIONS

Despite the development of education-based apps and their use in K-12 education, little research actually examines apps as assistive technology for students with disabilities. Hence, the field is ripe for research exploring if apps really do support students with disabilities as assistive technology as well as if app-based assistive technologies are as effective as stand-alone devices. While the research on mobile devices for students with disabilities increased significantly in recent years, the attention has not necessarily transferred to apps themselves. The little research examining apps as assistive technology suggest the poten-

tial for this technology, although the results do not necessarily suggest that app-based assistive technology is more effective than other options (e.g., traditional teacher instruction, paper-based options, computer-based instruction) (Bryant, Kim, et al., 2014; Bryant, Ok, et al., 2015; Weng & Bouck, in press).

CONCLUSION

Apps provide an option to serve as assistive technologies for K-12 students with a range of disabilities. Apps specifically exist or can be repurposed to support students with disabilities in a range of purposes, including but not limited to, academics, organization, computer access, daily living, and communication. Yet, it is important that educators give sufficient consideration to the educational apps implemented as assistive technology for students with disabilities. It is insufficient to rely on the placement of an app on a list or a recommendation from an organization or an individual. Rather, educators should evaluate apps as assistive technology. This article presented three rubrics for educator consideration with focus exclusively on evaluating apps as assistive technology or apps for students with disabilities. Educators need to fit the existing app rubrics to their population and purpose, as well as select one that is easy for them to use.

REFERENCES

- Apple. (2015). *iPad in education*. Retrieved from <http://www.apple.com/education/ipad/apps-books-and-more/>
- AssistiveWare. (2013). *Proloquo2Go*. Retrieved from <https://itunes.apple.com/us/app/proloquo-2go/id308368164?mt=8>
- BeeVisual. (2015). *Choiceworks®*. Retrieved from <http://www.beevisual.com/>
- Behrmann, M., & Jerome, M. (2002). Assistive technology for students with mild disabilities: Update 2002. *ERIC Digest*. Retrieved from <http://www.ericdigests.org/2003-1/assistive.htm>
- Behrmann, M., & Schaff, J. (2001). Assisting educators with assistive technology: Enabling children to achieve independence in living and learning. *Children and Families*, 42(3), 24–28.
- Beneficent Technology. (2002-2015). *Reading tools – Find a compatible reading tool*. Retrieved from <https://www.bookshare.org/cms/node/545>
- Blackhurst, A. (1997). Perspectives on technology in special education. *Teaching Exceptional Children*, 29(5), 41–48.
- Bouck, E. C. (2016). *Assistive technology*. Thousand Oaks, CA: Sage.
- Bouck, E. C., Flanagan, S., Miller, B., & Bassette, L. (2012). Technology in Action: Rethinking everyday technology as assistive technology to meet students IEP goals. *Journal of Special Education Technology*, 27(4), 47–57. doi:10.1177/016264341202700404
- Bouck, E. C., Jasper, A., Bassette, L., & Shurr, J. (2015). Mobile phone: Repurposed assistive technology for individuals with disabilities. In Z. Yan (Ed.), *Encyclopedia of Mobile Phone Behavior* (Vols. 1-3). Hershey, PA: IGI Global.
- Bouck, E. C., Jasper, A., Bassette, L., Shurr, J., & Miller, B. (2013). Applying TAPE: Rethinking assistive technology for students with physical disabilities, multiple disabilities, and other health impairment. *Physical Disabilities Research & Practice*, 32(1), 31–54.
- Bouck, E. C., Shurr, J., Tom, K., Jasper, A. D., Bassette, L., Miller, B., & Flanagan, S. M. (2012). Fix it with TAPE: Repurposing technology to be assistive technology for students with high incidence disabilities. *Preventing School Failure*, 56(2), 121–128.

- Bryant, B. R., Kim, M. K., Ok, M. W., Kang, E. Y., Bryant, D. P., Lang, R., & Son, S. H. (2014). A comparison of the effects of reading intervention on engagement and performance for fourth-grade students with learning disabilities. *Behavior Modification*, 39(1), 167–190. doi:10.1177/0145445514561316 PMID:25432586
- Bryant, B. R., Ok, M., Kang, E. Y., Kim, M. K., Lang, R., Bryant, D. P., & Pfannestiel, K. (2015). Performance of fourth-grade students with learning disabilities on multiplication facts comparing teacher-mediated and technology-mediated interventions: A preliminary investigation. *Journal of Behavioral Education*, 24(2), 255–272. doi:10.1007/s10864-015-9218-z
- Cassingham, R. (2012). *The Dvorak keyboard: A brief primer*. Retrieved from <http://www.dvorak-keyboard.com/>
- Cayton-Hodges, G. A., Feng, G., & Pan, X. (2015). Tablet-based math assessment: What can we learn from math apps? *Journal of Educational Technology & Society*, 18(2), 3–20.
- Cherner, T., Dix, J., & Lee, C. (2014). Cleaning up that mess: A framework for classifying educational apps. *Contemporary Issues in Technology & Teacher Education*, 14(2), 158–193.
- Clarity Innovations. (2015). *Number Line*. Retrieved from <https://itunes.apple.com/us/app/number-line-by-math-learning/id751816884?mt=8>
- Cook, A. M., & Polgar, J. M. (2015). *Assistive technologies: Principles and practice* (4th ed.). St. Louis, MO: Elsevier.
- Edyburn, D. (2005). Assistive technology and students with mild disabilities: From consideration to outcomes measurement. In D. Edyburn, K. Higgins, & R. Boone (Eds.), *Handbook of special education technology research and practice* (pp. 239–270). Whitefish Bay, WI: Knowledge by Design.
- Evernote. (2015). Retrieved from <https://evernote.com/>
- Ganz, J. B. (2014). *Aided augmentative communication for individuals with autism spectrum disorders*. New York: Springer. doi:10.1007/978-1-4939-0814-1
- Gierach, J. (Ed.). (2009). *Assessing students' needs for assistive technology (ASNAT): A resource manual for school district teams* (5th ed.). Retrieved from <http://www.wati.org/?pageLoad=content/supports/free/index.php>
- GingerLabs. (2015). *Notability*. Retrieved from <http://www.gingerlabs.com/>
- Grunwald Associates LLC. (2013). *Living and learning with mobile devices: What parents think about mobile devices for early childhood and K-12 learning*. Retrieved from <http://grunwald.com/pdfs/Grunwald%20Mobile%20Study%20public%20report.pdf>
- Hanline, M. F., Nunes, D., & Worthy, M. B. (2007). Augmentative and alternative communication in the early childhood years. *Young Children*, 62(4), 78–82.
- InClass Inc. (2015). *InClass*. Retrieved from <https://itunes.apple.com/us/app/inclass/id374986430?mt=8>
- Johnson, L., Beard, L., & Carpenter, L. (2007). *Assistive technology: Access for all students*. Upper Saddle River, NJ: Pearson Education.
- Mehdipour, Y., & Zerehkafi, H. (2013). Mobile learning for education: Benefits and challenges. *International Journal of Computational Engineering Research*, 3(6), 93–101.
- Mishra, P., & Koehler, M. J. (2009). Too cool for school? No way! *Learning and Leading with Technology*, 36(7), 14–18.

- Molnar, M. (2015). Half of K-12 students to have access to 1-to-1 computing by 2015-16. *Education Week*. Retrieved from http://blogs.edweek.org/edweek/marketplacek12/2015/02/half_of_k-12_students_to_have_access_to_1-to-1_computing_by_2015-16_1.html
- More, C. M., & Travers, J. C. (2013). Whats app with that? Selecting educational apps for young children with disabilities. *Young Exceptional Children*, 16(2), 15–32. doi:10.1177/1096250612464763
- MyScript. (2015). *MyScript Calculator*. Retrieved from, <https://itunes.apple.com/us/app/myscript-calculator-handwriting/id578979413?mt=8>
- Nagel, D. (2014). One-third of U.S. students use school-issued mobile devices. *THE Journal*. Retrieved from <http://thejournal.com/articles/2014/04/08/a-third-of-secondary-students-use-school-issued-mobile-devices.aspx>
- Ok, M. W., Kim, M. K., Kang, E. Y., & Bryant, B. R. (2015). How to find good apps: An evaluation rubric for instructional apps for teaching students with learning disabilities. *Intervention in School and Clinic*. doi:10.1177/1053451215589179
- Pilone, P. (2013). *iHomework*. Retrieved from, <http://www.element84.com/ihomework/iOS.html>
- Purcell, K., Entner, R., & Henderson, N. (2010). *The rise of the applications culture* Retrieved from http://www.pewinternet.org/files/old-media/Files/Reports/2010/PIP_Nielsen%20Apps%20Report.pdf
- Rodríguez, C. D., Strnadová, I., & Cumming, T. (2014). Using iPads with students with disabilities: Lessons learned from students, teachers, and parents. *Intervention in School and Clinic*, 49(4), 244–250. doi:10.1177/1053451213509488
- Stephenson, J., & Limbrick, L. (2013). A review of the use of touch-screen mobile devices by people with developmental disabilities. *Journal of Autism and Developmental Disorders*. PMID:23888356
- Sutton, M. (2015). *Talking pictures: Visual scene displays*. Retrieved from http://community.advanceweb.com/blogs/sp_6/archive/2013/06/24/talking-pictures-visual-scene-displays.aspx
- Tammaro, M. T., & Jerome, M. K. (2012). Does the app fit? Using the Apps Consideration Checklist. In M. J. Ault & M. E. Bausch (Eds.), *Apps for all students: A teacher's desktop guide* (pp. 23-31). Reston, VA: Technology and Media Division (TAM) for the Council for Exceptional Children.
- TapFun, Inc. (2014). *Base Ten Blocks Math*. Retrieved from <https://itunes.apple.com/us/app/base-ten-blocks-math/id878351349?mt=8>
- Technology-Related Assistance for Individuals with Disabilities Act, 29 U.S.C. § 2201 *et seq.*
- Tuthill, J. (2014, June). Get real with visual scene displays. *The ASHA Leader*, 19, 34-35. Retrieved from <http://leader.pubs.asha.org/article.aspx?articleid=1878423>
- Vanderheiden, G. (1984). High and light technology approaches in the development of communication systems for the severely physically handicapped person. *Exceptional Education Quarterly*, 4(4), 40–56.
- Walser, K., Ayres, K. M., & Foote, E. (2012). Effects of a video model to teach students with moderate intellectual disabilities to use key features of an iPhone. *Education and Training in Autism and Developmental Disabilities*, 47(3), 319–331.
- Weng, P.-L., & Bouck, E. C. (in press). *An evaluation of app-based and paper-based number lines for teaching number comparison*. Academic Press.
- Weng, P.-L., & Taber-Doughty, T. (2015). Developing an app evaluation rubric for practitioners in special education. *Journal of Special Education Technology*, 30(1), 43–58. doi:10.1177/016264341503000104

Wong, C., Odom, S. L., Hume, K., Cox, A. W., Fettig, A., Kurcharczyk, S.,... Schultz, T. R. (2014). Evidence-based practices for children, youth, and young adults with autism spectrum disorder. Chapel Hill, NC: The University of North Carolina, Frank Porter Graham Child Development Institute, Autism Evidence-Based Practice Review Group. Retrieved from <http://autismpdc.fpg.unc.edu/sites/autismpdc.fpg.unc.edu/files/2014-EBP-Report.pdf>

ADDITIONAL READING

Ally, M., Grimus, M., & Ebner, M. (2014). Preparing teachers for a mobile world, to improve access to education. *Prospects*, 44(1), 43–59. doi:10.1007/s11125-014-9293-2

Banister, S. (2010). Integrating the iPod Touch in K–12 education: Visions and vices. *Computers in the Schools*, 27(2), 121–131. doi:10.1080/07380561003801590

Bouck, E. C., Meyer, N. K., Hunley, M., Satsangi, R., & Savage, M. (2015). Free computer-based assistive technology to support students with high incidence disabilities in the writing process. *Preventing School Failure*, 59(2), 90–97. doi:10.1080/1045988X.2013.841116

Boyle, J. R., Forchelli, G. A., & Cariss, K. (2015). Note-taking interventions to assist students with disabilities in content area classes. *Preventing School Failure*, 59(3), 186–195. doi:10.1080/1045988X.2014.903463

Burton, C. E., Anderson, D. H., Prater, M. A., & Dyches, T. T. (2013). Video self-modeling on an iPad to teach functional math skills to adolescents with autism and intellectual disability. *Focus on Autism and Other Developmental Disabilities*, 28(2), 67–77. doi:10.1177/1088357613478829

Crowley, K., McLaughlin, T., & Kahn, R. (2013). Using direct instruction flashcards and reading racetracks to improve sight word recognition of two elementary students with autism. *Journal of Developmental and Physical Disabilities*, 25(3), 297–311. doi:10.1007/s10882-012-9307-z

Cumming, T. M., & Rodriguez, C. D. (2013). Integrating the iPad into language arts instruction for students with disabilities: Engagement and perspectives. *Journal of Special Education Technology*, 28(4), 43–52. doi:10.1177/016264341302800404

Evmeanova, A. S., & King-Sears, M. E. (2013). Technology and literacy for students with disabilities. In J. Whittingham, S. Huffman, W. Rickman, & C. Wiedmaier (Eds.), *Technological tools for the literacy classroom* (pp. 83–104). Hershey, PA: IGI-Global; doi:10.4018/978-1-4666-3974-4.ch006

Flanagan, S. M., & Bouck, E. C. (2015). Mapping out the details: Supporting written expression through concept mapping. *Preventing School Failure*, 59(4), 244–252. doi:10.1080/1045988X.2014.933400

Fletcher-Watson, S. (2015). Evidence-based technology design and commercialization: Recommendations derived from research in education and autism. *TechTrends*, 59(1), 84–88. doi:10.1007/s11528-014-0825-7

Flores, M., Musgrove, K., Renner, S., Hinton, V., Strozier, S., Franklin, S., & Hil, D. (2012). A comparison of communication using the Apple iPad and a picture-based system. *Augmentative and Alternative Communication*, 28(2), 74–84. doi:10.3109/07434618.2011.644579 PMID:22263895

Green, L. S., Hechter, R. P., Tysinger, P. D., & Chasereau, K. D. (2014). Mobile app selection for 5th through 12th grade science: The development of the MASS rubric. *Computers & Education*, 75(1), 65–71. doi:10.1016/j.compedu.2014.02.007

- Gulchak, D. J. Daniel J. Gulchak. (2008). Using a mobile handheld computer to teach a student with an emotional and behavioral disorder to self-monitor attention. *Education & Treatment of Children*, 31(4), 567–581. doi:10.1353/etc.0.0028
- Interactive Educational Systems Design, Inc. (2014). *Educator Edition: 2014 National Survey on Mobile Technology for K-12 Education*. Retrieved from <http://go.amplify.com/2014-mobile-report>
- Ismail, I., Azizan, S. N., & Azman, N. (2013). Mobile phone as pedagogical tools: Are teachers ready? *International Education Studies*, 6(3), 36–47. doi:10.5539/ies.v6n3p36
- Israel, M., Maynard, K., & Williamson, P. (2013). Promoting literacy-embedded, authentic STEM instruction for students with disabilities and other struggling learners. *Teaching Exceptional Children*, 45(4), 18–25. doi:10.1177/004005991304500402
- McMahon, D., Cihak, D. F., & Wright, R. (2015). Augmented reality as a navigation tool to employment opportunities for postsecondary education students with intellectual disabilities and autism. *Research on Technology in Education*, 47(3), 157–172. DOI: 10.80/15391523.2015.1047698
- McMahon, D. D., Cihak, D. F., Gibbons, M. M., Fussell, L., & Mathison, S. (2013). Using a mobile app to teach individuals with intellectual disabilities to identify potential food allergens. *Journal of Special Education Technology*, 28(3), 21–32. doi:10.1177/016264341302800302
- Neely, L., Rispoli, M., Camargo, S., Davis, H., & Boles, M. (2013). The effect of instructional use of an iPad on challenging behavior and academic engagement for two students with autism. *Research in Autism Spectrum Disorders*, 7(4), 509–516. doi:10.1016/j.rasd.2012.12.004
- Newton, D. A., & Dell, A. G. (2011). Mobile devices and students with disabilities: What do best practices tell us? *Journal of Special Education Technology*, 26(3), 47–49. doi:10.1177/016264341102600305
- OBannon, B. W., & Thomas, K. (2014). Teacher perceptions of using mobile phones in the classroom: Age matters! *Computers & Education*, 74(1), 15–25. doi:10.1016/j.compedu.2014.01.006
- Pew Research Center. (2013). *Tablet and e-reader ownership update*. Retrieved from <http://www.pewinternet.org/2013/10/18/tablet-and-e-reader-ownership-update/>
- Pew Research Center. (2015). *U.S. smartphone use in 2015*. Retrieved from <http://www.pewinternet.org/2015/04/01/us-smartphone-use-in-2015/>
- Powell, S. (2014). Choosing iPad apps with a purpose: Aligning skills and standards. *Teaching Exceptional Children*, 47(1), 20–26. doi:10.1177/0040059914542765
- Retter, S., Anderson, C., & Kieran, L. (2013). iPad use for accelerating reading gains in secondary students with learning disabilities. *Journal of Educational Multimedia and Hypermedia*, 22(4), 443–463.
- Roth, K. (2013). Adapt with apps. *Journal of Physical Education, Recreation & Dance*, 84(2), 4–6. doi:10.1080/07303084.2013.757168
- Sheng, H., Siau, K., & Nah, F. F.-H. (2010). Understanding the values of mobile technology in education: A value-focused thinking approach. *The Data Base for Advances in Information Systems*, 41(2), 25–44. doi:10.1145/1795377.1795380
- Simmons, K. D. (2014). Apps for communication and video modeling for middle school students with autism spectrum disorders. *Journal of Instructional Psychology*, 41(1-4), 79–82.

Swan, K., Hooft, M. V. T., Kratcoski, A., & Unger, D. (2005). Uses and effects of mobile computing devices in K–8 classrooms. *Journal of Research on Technology in Education*, 38(1), 99–112. doi :10.1080/15391523.2005.10782451

Wilkinson, K. M., & Light, J. (2011). Preliminary investigation of visual attention to human figures in photographs: Potential considerations for the design of aided AAC visual scene displays. *Journal of Speech, Language, and Hearing Research*, 54(6), 1644–1657. Retrieved, from <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3462357/>

KEY TERMS AND DEFINITIONS

Apps: Applications or software for mobile devices.

Assistive Technology Device: Any technology that benefits or supports a student with a disability.

Augmentative and Alternative Communication Device: A piece of assistive technology specifically focused on communication; it supplements or replaces a student’s verbal communication.

Computer Access: The way in which an individual accesses or uses a computer or computer-like device.

iOS: The operating system for Apple mobile devices.

Mobile Device: Handheld computer-like devices, including tablets and smartphones.

Rubric: A method of evaluation that assesses quality.

Assistive Technology and Human Capital for Workforce Diversity

A**Ben Tran***Alliant International University, USA*

INTRODUCTION

Workforce development has evolved to describe any one of a relatively wide range of national and international policies and programs related to learning for work. This evolution, in part, is the end result of the term being frequently misunderstood: workforce development consists of job training only. Harrison and Weiss (1998, p. 5) defined workforce development as the “constellation of activities from orientation to the work world, recruiting, placement, and mentoring to follow-up counseling and crisis intervention”. However, inequality in the workplace, specifically promoting equal rights to employment for members of discriminated groups with disabilities (Kriegel, 2002; McClain, 2002) has become a widely talked about phenomenon in the business world, and is visible from an international (Jakovljevic & Buckley, 2011) paradigm. Hence, the purpose of this chapter is not on the varieties of the availability of assistive technologies (AT) and their usages based on individuals’ specified disability, so that individuals who require the usage of ATs can be of equal playing field compared to those individuals who do not require the usage of ATs. For information regarding AT and the state of AT in the past, present, and future in the United States, ADA, and the like, refer to Tran’s (2015a) article titled *Assistive Technology*. The purpose of this chapter is beyond the coverage of Tran’s (2015a) *Assistive Technology* article, such that, the purpose of this article is on the end results that AT could provide and contribute to the diverse workforce, and the role AT play in relations to workforce development—from an international perspective.

BACKGROUND

Historically speaking, from past to present, for many developing countries, legislation regarding the employment of individuals with disabilities has been criticized due to its ineffectiveness (Schall, 1998; Siegal, 2001). In particular, according to Jakovljevic and Buckley (2011), the legislation has had little or no impact on the employment status of people with disabilities (Agocs, 2002; Brett, 2000; Conlin, 2000; De Jonge, Rodger, & Fitzgibbon, 2001; De Laurentiis, 1991; Hignite, 2000; IRS, 1998; McGregor, 1991; Robitaille, 2002; Saskatchewan, 2000; Schall, 1998, Thomas, 2002). When addressing the needs of employees with disabilities, the Act and the Code both include the term *reasonable accommodation* (Tran, 2015a). Reasonable accommodation (disability accommodation) is any modification or adjustment to a job or to a working environment that will enable a person from a designated group to have access to or participate or advance in employment (Department of Labor, 2002; Tran, 2015a). It includes acquisition and modification of equipment and devices, as well as any necessary training. These devices and equipment are collectively known as assistive technologies (AT).

HUMAN CAPITAL FOR WORKSPACE DIVERSITY: ASSISTIVE TECHNOLOGIES (AT)

Human capitals [cultural and global (Tran, 2014a)] include labor market skills, leadership skills, general education background, artistic development

DOI: 10.4018/978-1-5225-2255-3.ch025

and appreciation, health, experiences, and intelligence [behavioral, emotional, cognitive, cultural (CQ), general (IQ), metacognitive, motivational, and social (Tran, 2014a)]. Human capitals are essential asset in most communities, both domestic, and international. Traditional approaches to human capital development emphasize individual responses. Quite often, individuals bear the cost, and the burden of obtaining education and training. Although many government programs have been developed to provide training, they tend to focus on specific populations and often are not well connected with local labor market conditions, as compared to nongovernmental that may have different types and levels of access and accessibility for various populations: the abled and the disabled.

Disability

There is much debate about the best way to define *disability*. The issue of definition has also been further complicated by the links to individual eligibility criteria for program and financial assistance or to legal implications. Furthermore, disability is difficult to define because it is a multi-dimensional concept with both objective and subjective characteristics. When interpreted as an illness or *impairment*, disability is seen as fixed in an individual's body or mind (Tran, 2014b; Tran, 2015b). When interpreted as a social construct disability is seen in terms of the socioeconomic, cultural and political disadvantages resulting from an individual's exclusion. Furthermore, people with disabilities, advocacy groups, legal and medical practitioners, policy-makers and the general public, all have a different view of disability. Thus, the meaning of disability has evolved and changed over the years through various perspectives, as such moral, medical, social, and human rights perspectives (Department of Education & Training, 2005).

The various perspectives have an effect not only on how we define disability, but also on social planning and program design, employment strategies, and how decisions are made regarding

such things as program eligibility (Nagi, 1965). In seeking to emphasize the distinction between impairment and disability, and promote a *social model*, the Physical Disability Council of Australia (2004) has stated:

Disability means the disadvantage or restriction caused by a contemporary social organization, which takes no account or little account of people who have impairments and the functional or behavioral consequences of those impairments, leading to social exclusion or resulting in less favorable treatment of and discrimination against people with impairments. Therefore people with disability are people with impairments who are disabled by barriers in society. [The] central theme in this definition are disabled is external to the individual and is a result of environmental and social factors (Submission by Physical Disability Council of Australia to Productivity Commission, 2004).

There are many variations of the social model, but all portray disability as a social construct created by ability-oriented and ability-dominated environments. According to the social model, even though impairment has an objective reality that is attached to the body or mind, disability has more to do with society's failure to account for the needs of people with disabilities.

Assistive Technology

The World Report on Disability [World Health Organization (WHO, 2011, p. 101)] produced by the World Health Organization and the World Bank, defined assistive technology (citing the U.S. Assistive Technology Act, 2004) as "any item, piece of equipment, or product, whether it is acquired commercially, modified, or customized, that is used to increase, maintain, or improve the functional capabilities of individuals with disabilities" (WHO, 2011). Assistive technology was defined in 1988 U.S. Technology-Related Assistance of Individuals with Disabilities Act as "any item,

piece of equipment, or product system, whether acquired commercially off the shelf, modified or customized, that is used to increase, maintain, or improve functional capabilities of individuals with disabilities (Brodwin, Star, & Cardoso, 2003; Tran, 2014b; Tran, 2015b). The U.K. Foundation for Assistive Technology provides a more succinct definition, namely, “any product or service designed to enable independence for disabled and older people”. This would cover not only devices and services, but also environmental adaptations such as home modifications, and would range from simple objects like walking and grab rails to complex devices like power wheelchairs (Hobbs, Close, Downing, Reynolds, & Walker, 2009).

It is widely accepted that assistive technologies are of vital importance for people with disabilities. The WHO report (2011), for example, said that they had been shown to be powerful tools to increase independence and improve participation, as long as they were appropriate for the users and for their environments. The report also noted that assistive devices had been known to be effective in reducing disability, and that they could at times substitute for support services, or at least complement them, and thus reduce the costs of care (MTAA, 2011). A content analysis of the letters to the editor in two disability-related magazines in the U.S. (Marini, Bhaktam & Graf, 2009) found that adaptive aid equipment was clearly a major concern for the people with disabilities who wrote in to these magazines. Adaptive aid and equipment was the second most frequently cited issue whereas the first most cited issue was accessibility (Thompson, Fisher, & Kayess, 2012). Despite the recognition of its importance, research on AT is still limited.

ASSISTIVE TECHNOLOGY

Interest in the potential of assistive technology to promote independence among people with disabilities arises out of concerns about quality of life, decreases in the availability of informal caregivers and a shrinking long-term care work-

force. Using AT in addition to personal services could potentially improve the quality of care, and thus improve social inclusion and defer functional declines and institutionalization (Agree, Freedman, Cornman, Wolf, & Marcotte, 2005). It is hoped that the use of AT can alleviate pressures on the existing long-term care system while at the same time improving quality of life, easing difficulties with *activities of daily living* (ADLs), and reducing the level of unmet need. Where the use of AT reduces the amount of personal care needed, public expenditures on home health care could be reduced and the responsibilities of informal care lightened.

Types of Disabilities: Impact on Employment

Not a great deal of the research undertaken has compared the types of disabilities and their impact on employment experience (Tran, 2014b; Tran, 2015b). However, research into work participation carried out by the New Zealand (NZ) Ministry of Social Development concluded that the more severe a person's disability, the lower the likelihood that the person will be in employment (Jensen, Sathiyandra, Rochford, Jones, Krishnan, & McLeod, 2004). According to this research, having a hearing impairment appeared to have a much smaller impact on employment than other disabilities, and this did not alter according to the severity of the impairment. It was noted that people who were visually impaired or legally blind were perceived by employers to be the hardest to accommodate.

Barriers to Employment

Various *barriers* to the employment and subsequent disability accommodation of people with disabilities exists, including confusion about the definitions, stereotyping and misconceptions regarding work attitudes (Global Diversity, 2000; Hignite, 2000; Schall, 1998; Services SETA, 2002; Whiting, 2001). The clause allowing escape from employing workers with disabilities due to

unjustifiable hardship, where an employer can avoid reasonable disability accommodation if it involves action that requires significant or considerable difficulty or expense (Department of Labor, 2002), is another possibility (Tran, 2015a).

Approaches to Accommodating Disability Employment

While numerous research documents have confirmed that people with a disability are valuable contributors to a diverse work force, the benefits of diversity in employment have not flowed to people with a disability. Past research such as the report prepared for the Victorian Office of Public Employment (Graffam, Smith, & Hardcastle, 2005) and Human Rights and Equal Opportunity Commission (HREOC) (HREOC, 2005) have heightened that people with a disability remain disproportionately excluded as employees. The Australian Public Service Commission, 2004) reported that in the decade between 1994 and 2004 there had been a decline in the employment of people reporting a disability as a proportion of APS employees. In 2004, people with a disability represented 3.8% of continuing Australian Public Service (APS) employees, down from 5.8% ten years prior (Australian Public Service Commission, 2004). Although such date may not necessarily be the most updated, validity nevertheless, is reliable and generalizable based on the popularity of research conducted regarding the population at hand (as compared to the abled population in the high-tech or related industries). Based on the data, the data suggests that although promoting a diverse work force is well advised, there is still much room for development and improvement in relation to people with a disability.

Accommodation for Individuals with Disability: Assistive Technologies

Assistive technologies in the workplace are necessary where the employee with disabilities is reliant on some device in order to complete assigned tasks.

Reasonable disability accommodation guidelines relate to any modification or adjustment to a job or to the working environment, which will enable a person from a designated group to have access to or participate or advance in employment (Department of Labor, 1998; Tran, 2015a). The code laid down by the Department of Labor (2002, p. 5) reflects reasonable disability accommodation guidelines as follows:

Employers should reasonably accommodate the needs of people with disabilities. The aim of the accommodation is to reduce the impact of the impairment of the person's capacity to fulfill the essential functions of a job...Employers may adopt the most cost-effective means that are consistent with effectively removing the barrier to a person being able to perform the job, and to enjoy equal access to the benefits and opportunities of employment.

A particularly appropriate example of the reasonable disability accommodation guidelines (Tran, 2015a) is “adapting existing equipment or acquiring new equipment including computer hardware and software” (Department of Labor, 2002, p. 5). Adherence to the guidelines is strongly recommended to those employers classified designated employers, who are at the same time obliged to employ people out of designed groups (Department of Labor, 1998).

FUTURE RESEARCH

Human capital for the workforce development, from an international perspective, some countries (both the developing and the developed) are more equipped to accommodate the disabled (and the abled) than other countries (both the developing and the developed). In accommodating the disabled, a great deal of literature suggested a dual approach, combining both work force diversity approaches and equal employment opportunity (EEO) approaches, to improve employment for

people with a disability. The EEO legislation is a well-recognized tool for promoting a diverse work force. Seen in its historical context, EEO promised equality of access and opportunity for under-represented groups, and, through this, social and economic change. However, while other guidelines, policies, legislations, and the like for various countries have made changes (or attempts to make changes) reflecting and addressing the changing needs of the workforce, changes have been disappointing in terms of employment of people with disabilities. Professional organizations and officers who should offer employment advice to persons with disabilities are missing in many countries, and more commonly and surprisingly, in many companies' department of human resources. Hence, a question left unanswered to date: does the value of intellectual human capitals (individuals' with disabilities) cause a change in legal accommodations or does a change in legal accommodations cause a change of value in intellectual human capitals (individuals' with disabilities)?

CONCLUSION

The international workforce must understand and utilize knowledge and competencies in order to achieve and sustain longevity. In order to address and understand knowledge and innovation as invaluable factors that affect the longevity of large organizations, to utilize and capitalize on these factors so that they can leverage the knowledge base of their organizations to build competitive advantages for their firms, one must understand what they are and their evolution (Tran, 2009). This is because human capital is an essential asset in most organizations (domestic and international), because human capital is an organization's most important asset. Hence, organizations need to be able to capitalize on human capital a source of competitive advantage. In doing so, organizations need to implement Tran's (2008) *Expatriate Selection and Retention* methodology in recruiting, as-

sessing, selecting, training, and mentoring human capitals, by focusing of individuals' knowledge, skills, abilities, and other characteristics (KSAOs). KSAOs refers to (Tran, 2008, p. 3; Tran, 2013, 2014c, 2015c):

1. Knowledge is usually defined as the degree to which a candidate is required to know certain technical material.
2. Skill indicates adequate performance on tasks requiring the use of tools, equipment, machinery, etc.
3. Abilities are physical and mental capacities to perform tasks not requiring the use of tools, equipment, or machinery.
4. Other characteristics include personality, interest, or motivational attributes that indicate a candidate will learn certain tasks, rather than whether they can do those tasks (Schneider & Schmitt, 1992: 53).

Organizations must first, focus on human capitals based on their ability, thereafter, provide modifications and accommodations, if necessary for disabled human capitals who may, upon request, require assistive technologies.

REFERENCES

- Agocs, C. (2002). Canadas employment equity legislation and policy, 19872000: The gap between policy and practice. *International Journal of Manpower*, 23(3), 256–276. doi:10.1108/01437720210432220
- Agree, E. M., Freedman, V. A., Cornman, J. C., Wolf, D. A., & Marcotte, J. E. (2005). Reconsidering substitution in long-term care: When does assistive technology take the place of personal care? *Journal of Gerontology*, 60B(5), S272–S280. doi:10.1093/geronb/60.5.S272 PMID:16131628
- Austrailian Public Service Commission. (2004). *State of the service report 2003-2004*. Canberra: APS Commission.

- Brett, J. T. (2000). Working for people with disabilities. *New England's Journal of Higher Education and Economic Development*, 15(2), 25–28.
- Brodwin, M. G., Star, T., & Cardoso, E. (2003). Users of assistive technology: The human component. *Journal of Applied Rehabilitation Counseling*, 34(4), 23–29.
- Conlin, M. (2000). The new workforce. *Business Week*, 3673, 64–68.
- De Jonge, D., Rodger, S., & Fitzgibbon, H. (2001). Putting technology to work: Users' perspective on integrating assistive technology into the workplace. *Work (Reading, Mass.)*, 16(2), 77–89. PMID:12441461
- De Laurentiis, J. (1991). Eliminating barriers. *Canadian Business Review*, 18(2), 33–35.
- Department of Education & Training. (2005). *Case studies of employees with a disability*. Melbourne: Policy and Employee Relations Division, Department of Education & Training.
- Department of Labor. (1998). *The employment equity act no. 55 of 1998*. Retrieved on October 1, 2015, available at <http://www.labour.gov.co.za>
- Department of Labor. (2002). *Code of good practice on key aspects on the employment of people with disabilities. Toolkit: Managing people with disabilities in the workplace*. Retrieved on October 1, 2015, available at http://www.services.org.za/EAP/eap_index.html
- Global Diversity. (2000). *Global news: Global diversity @ work*. Retrieved on October 1, 2015, available at http://www.diversityatwork.com/news/mar00/news_sa.html
- Graffam, J., Kaye, S., & Hardcastle, L. (2005). *Achieving substantive equality and optimal participation: Employees with a disability in the Victorian public sector*. Deakin University: A report prepared for the Victorian Office of Public Employment.
- Harrison, B., & Weiss, M. (1988). *Workforce development networks: Community-based organizations and regional alliance*. Thousand Oaks, CA: Sage.
- Hignite, C. (2000). The accessible association. *Association Management*, 52(13), 36–44.
- Hobbs, D. A., Close, J. K., Downing, A. R., Reynolds, K. J., & Walker, L. T. (2009). Developing a national research and development centre in assistive technologies for independent living. *Australian Health Review*, 33(1), 152–160. doi:10.1071/AH090152 PMID:19203345
- Human Rights and Equal Opportunity Commission. (2005). *Issues paper 1: Employment and disability—The statistics*. Sydney: HREOC.
- IRS. (1998). Employers' disability policies. *IRS Employment Review*, 650, 4–15.
- Jakovljevic, M., & Buckley, S. (2011). Assistive technologies in a workplace environment: Barriers for the employment of persons with disabilities. *Disability, CBR (Community-Based Rehabilitation) & Inclusive Development*, 22(2), 55–78.
- Jensen, J., Sathiyandra, S., Rochford, M., Jones, D., Krishnan, V., & McLeod, K. (2004). Work participation amongst people with disabilities: Does the type of disability influence the outcome? In *Proceedings of the 2004 Social Policy Research and Evaluation Conference*. Ministry of Social Development.
- Kriegek, L. (2002). Handicapping the crippled. *Nation (New York, N.Y.)*, 275(6), 32–36.
- Marini, I., Bhakta, M. V., & Graf, N. (2009). A content analysis of common concerns of persons with physical disabilities. *Journal of Applied Rehabilitation Counseling*, 40(1), 44–90.
- McClain, C. V. (2002). *Governance and legislation in South Africa: A contemporary overview*. Retrieved on September 30, 2015, available at <http://www.disabilityworld.org/gov/southafrica.shtml>

McGregor, C. (1991). Disabled people for employment equity. *Canadian Business Review*, 18(2), 30–33.

MTAA. (2011). *Disability care and support: Medical assistive technologies to enable Australians to remain in their homes*. Sydney: Medical technology association of Australia limited. Retrieved on October 1, 2015, available at http://www.mtaa.org.au/pages/images/Disability%20Care%20and%20Support_MTAA%20response%20to%20draftreport.pdf

Productivity Commission. (2004). *Review of the disability discrimination act 1992, Report No. 30*, Melbourne: Author.

Robitaille, S. A. (2002). *Victory in disguise for the disabled*. Retrieved on September 30, 2015, available at <http://www.bloomberg.com/bw/stories/2002-01-09/a-victory-in-disguise-for-the-disabled>

Saskatchewan, H. R. C. (2000). *Saskatchewan Human Rights Commission—Equity*. Retrieved on September 30, 2015, available at <http://www.gov.sk.ca/shrc/equity2000/equity15.htm>

Schall, M. C. (1998). The American disabilities act—are we keeping our promise? An analysis of the effect of the ADA on the employment of people with disabilities. *Journal of Vocational Rehabilitation*, 10(3), 191–203. doi:10.1016/S1052-2263(98)00024-5

Schneider, B., & Schmitt, N. (1992). *Staffing organizations* (2nd ed.). Prospect Heights, IL: Waveland Press, Harper Collins, Inc.

Service, S. E. T. A. (2002). *Toolkit: Managing people with disabilities in the workplace*. Retrieved in October 1, 2015, available at http://www.serviceseta.org.za/EAP/eap_index.html

Siegal, C. (2001). International disability rights law: Norm or not? *International Civil Liberties Report*, 18-31.

Thomas, A. (2002). Employment equity in South Africa: Lessons from the global school. *International Journal of Manpower*, 23(3), 237–255. doi:10.1108/01437720210432211

Thompason, D., Fisher, K. R., & Kayess, R. (2012). The role of assistive technology in supporting people with disabilities and complex care needs: A literature review. University of New South Wales' Social Policy Research Centre Report Series.

Tran, B. (2008). *Expatriate selection and retention* (Published doctoral dissertation). California School of Professional Psychology at Alliant International University.

Tran, B. (2009). Knowledge management: The construction of knowledge in organizations. In D. Jemielniak & J. Kociatkiewicz (Eds.), *Handbook of Research on Knowledge-Intensive Organizations* (pp. 512-528). Hershey, PA: Premier Reference Source: Information Science Reference/IGI Global.

Tran, B. (2013). Industrial and organizational (I/O) psychology: The roles and purposes of I/O practitioners in global business. In B. Christiansen, E. Turkina, & N. Williams (Eds.), *Cultural and technological influences on global business* (pp. 175-219). Hershey, PA: Premier Reference Source: Business Science Reference/IGI Global.

Tran, B. (2014a). The foundation of cultural intelligence: Human capital. In B. Christiansen, S. Yildiz, and E. Yildiz (Eds.), *Handbook of Research on Effective Marketing in Contemporary Globalism* (pp. 348-376). Hershey, PA: Premier Reference Source: Business Science Reference/IGI Global.

Tran, B. (2014c). The human element of the knowledge worker: Identifying, managing, and protecting the intellectual capital within knowledge management. In M. A. Chilton and J. M. Bloodgood (Eds.), *Knowledge management and competitive advantage: Issues and potential solutions* (pp. 281-303). Hershey, PA: Premier Reference Source: Information Science Reference/IGI Global.

Tran, B. (2015a). Assistive technology. In M. Khosrow-Pour (Eds.), *Encyclopedia of information science and technology*, Third Edition (Vol. 6, pp. 4053-4061). Hershey, PA: Premier Reference Source: Information Science Reference/IGI Global.

Tran, B. (2015b). Extended assistive technology: The impact of human-computer interfaces on independence, employment, and organizations. In B. Christiansen & J. Koeman (Eds.), *Nationalism, cultural indoctrination, and economic prosperity in the digital age* (pp. 173-201). Hershey, PA: Premier Reference Source: Information Science Reference/IGI Global.

Tran, B. (2015c). Expatriate selection and retention: Identifying and assessing the other characteristics beyond knowledge, skills, and abilities. In A. A. Camillo (Eds.), *Handbook of research on global hospitality and tourism management* (pp. 468-492). Hershey, PA: Premier Reference Source: Business Science Reference/IGI Global.

Whiting, M. A. (2001). Hiring workers with disabilities. *Across the Board*, 38(6), 87-89.

WHO. (2011). *World report on disability*. Retrieved on October 1, 2015, available at http://www.who.int/disabilities/world_report/2011/en/index.html

ADDITIONAL READING

Beukelman, D. R., & Mirenda, P. (2013). *Augmentative and alternative communication: Supporting children and adults with complex communication needs* (4th ed.). Baltimore: Paul H. Brooks Publishing Company.

Birch, G. E., Fengler, M., Gosine, R. G., Schroeder, K., Schroeder, M., & Johnson, D. L. (1996). An assessment methodology and its application to a robotic vocational assistive device. *Technology and Disability*, 5(2), 151-166. doi:10.1016/1055-4181(96)00159-8

Biswas, P., & Langdon, P. (2011). A new input system for disabled users involving eye gaze tracker and scanning interface. *Journal of Assistive Technologies*, 5(2), 58-66. doi:10.1108/17549451111149269

Bowe, F. G. (1995). The political and economic issues that drive and derail assistive technology development. *Generations: Journal of the American Society on Aging*, 19(1), 37-40.

Boyd, R., & Richerson, P. J. (1995). Why does culture increase human adaptability? *Ethology and Sociobiology*, 16(2), 125-143. doi:10.1016/0162-3095(94)00073-G

Davies, T. C., Mudge, S., Ameratunga, S., & Stott, N. S. (2010). Enabling self-directed computer use for individuals with cerebral palsy: A systematic review of assistive devices and technologies. *Developmental Medicine and Child Neurology*, 52(6), 510-516. doi:10.1111/j.1469-8749.2009.03564.x PMID:20059508

Dawe, M. (2005). Caregivers, cost, and complexity: Understanding technology usage by individuals with cognitive disabilities. *Accessibility and Computing: The Newsletter of ACM SIGACCESS*, 81(81), 20-23. doi:10.1145/1055674.1055678

- Dennerlein, J. T. Y., & Maria, C. (2001). Haptic force-feedback devices for the office computer: Performance and musculoskeletal loading issues. *Human Factors*, 43(2), 278–286. doi:10.1518/001872001775900850 PMID:11592668
- Emiliani, P. L., & Stephanidis, C. (2005). Universal access to ambient intelligence environments: Opportunities and challenges for people with disabilities. *IBM Systems Journal*, 44(3), 605–619. doi:10.1147/sj.443.0605
- Enders, A. (1995). The role of technology in the lives of older people. *Generations: Journal of the American Society on Aging*, 19(1), 7–12.
- Engelbart, D. C. (1963). A conceptual framework for the augmentation of man's intellect. In P. W. Howerton & D. C. Weeks (Eds.), *Vistas in information handling*. Washington, D.C.: Spartan Books.
- Gitlin, L. N. (1995). Why older people accept or reject assistive technology. *Generations: Journal of the American Society on Aging*, 19(1), 41–46.
- Hinckley, K. (2008). Input technologies and techniques. In A. Sears & J. A. Jacko (Eds.), *The human-computer interaction handbook* (pp. 161–176). New York: Lawrence Erlbaum Associates.
- Hoenig, H., Taylor, D. H. Jr., & Sloan, F. A. (2003). Does assistive technology substitute for personal assistance among the disabled elderly? *American Journal of Public Health*, 93(2), 330–337. doi:10.2105/AJPH.93.2.330 PMID:12554595
- Hourkade, J., Pilotte, T. E., West, E., & Parette, P. (2004). A history of argumentative and alternative communication for individuals with severe and profound disabilities. *Focus on Autism and Other Developmental Disabilities*, 19(4), 235–244. doi:10.1177/10883576040190040501
- Roberts, M. (2000). From the wilderness to the workplace. *Journal of Alternative Dispute Resolution in employment*, 2(2), 66–73.
- Schneider, M. (1999). Achieving greater independence through assistive technology, job accommodation and supported employment. *Journal of Vocational Rehabilitation*, 12(3), 159–164.
- Schultz, I. Z., & Rogers, E. S. (2010). *Work accommodation and retention in mental health* (2011th ed.). Springer.
- Spalteholz, L. (2012). *KeySurf—A keyboard web navigation system for persons with disabilities*. Unpublished doctoral dissertation. Victoria, B.C.: University of Victoria.
- Steingold, F. S. (12th ed.) (2015). *The employer's legal handbook: Manage your employees & workplace effectively*. NOLO: Employer's Legal Handbook.
- Szymanski, E. M., & Parker, R. M. (2009). *Work and disability: Contexts, issues, and strategies for enhancing employment outcomes for people with disabilities* (3rd ed.). Pro Ed.
- Tran, B. (2014). The construction of knowledge management: The foundation of organizational learning based on learning organization. In F. Soliman (Eds.), *Learning models for innovation in organizations: Examining the roles of knowledge transfer and human resources Management* (pp. 89–110). Hershey, PA: Premier Reference Source: Business Science Reference/IGI Global.
- Tran, B. (2015). Ethical concerns in human enhancement: Advantages in corporate/organizational settings. In J. White and R. Searle (Eds.), *Rethinking machine ethics in the age of ubiquitous technology* (pp. 124–150). Hershey, PA: Premier Reference Source: Information Science Reference/IGI Global.

KEY TERMS AND DEFINITIONS

Assistive Technology: Any item, piece of equipment, or product, whether it is acquired commercially, modified, or customized, that is used to increase, maintain, or improve the functional capabilities of individuals with disabilities.

Behavioral Intelligence: Reflect the capability to exhibit appropriate verbal and nonverbal actions when interacting with people from different cultures.

Cognitive Intelligence: Knowledge structures and is consistent with Ackerman's intelligence-as-knowledge concept, which argues for the importance of knowledge as part of intellect.

Cultural Capital: The higher education success rates of educated parents.

Cultural Intelligence (CQ): A set of capabilities comprising mental, motivational, and behavioral components that focus specifically on resolving cross-cultural problems.

Disability: The disadvantage or restriction caused by a contemporary social organization, which takes no account or little account of people who have impairments and the functional or behavioral consequences of those impairments, leading to social exclusion or resulting in less favorable treatment of and discrimination against people with impairments.

Emotional Intelligence: Targeted at understanding one's and others' emotions, and practical intelligence targeted at solving practical problems.

General Intelligence (IQ): Person's capability for successful adaption to new cultural settings, that is, for unfamiliar settings attributable to cultural context or an individual's capability to function and manage effectively in culturally diverse settings.

Global Capital: Characterized by two major types of resources: the intangible cultural element in the form of organizational values toward globalization, and the overt cultural elements in the form of organizational routines for promoting its global values.

Human Capitals: Human capitals include labor market skills, leadership skills, general education background, artistic development and appreciation, health, experiences, and intelligence.

Metacognitive Intelligence: Control of cognition: the processes individuals are to acquire and understand knowledge.

Motivational Intelligence: Reflects the capability to direct attention and energy toward learning about and functioning in situations characterized by cultural differences.

Social Intelligence: Targeted at interpersonal relations, emotional intelligence.

Assistive Technology for Supporting Communication, Occupation, and Leisure by Children With Severe to Profound Developmental Disabilities

A**Fabrizio Stasolla***University of Bari, Italy***Viviana Perilli***Lega del Filo d'Oro – Molfetta, Italy***Adele Boccasini***Lega del Filo d'Oro – Termini Imerese, Italy*

INTRODUCTION

Assistive technology (AT) includes a wide range of electronic devices finalized at supporting children with disabilities and special needs to have an active role in the social environment and to minimize their isolation and passivity. By providing practical help to persons with disabilities, AT produce beneficial effects enhancing constructive engagement and quality of life (Borg, Larson, & Ostegren, 2011; Felce & Perry, 1995). To design an effective AT intervention program is necessary creating a solution tailored for the targeted users, and considering the targeted behaviors, tasks and environments in which programs are implemented (Reichle, 2011; Shih, 2011). The growing availability of such technologies may be considered a crucial outcome of the effectiveness of AT-based rehabilitative programs (De Pace & Stasolla, 2014).

BACKGROUND

Two categories of users have been recruited for the aforementioned programs:

1. Participants with extensive motor disabilities and normal intellectual functioning, and
2. Individuals with severe to profound developmental and multiple disabilities (i.e. a combination of intellectual, motor and sensorial disabilities).

The AT for the first category, implemented for educational and/or rehabilitative purposes, may be useful to improve literacy process (Chiapparino, Stasolla, De Pace, & Lancioni, 2011), instead the interventions for the second category may be focused on the independent access to preferred stimuli or on constructive engagement (Stasolla & Caffò, 2013; Stasolla et al., 2015).

A basic form of AT are the microswitches, that is electronic devices planned to enabling persons with disabilities to control autonomously their environment through the exhibition of small and simple behavioral responses (Holburn, Nguyen, & Vietze, 2004; Mechling, 2006; Saunders et al., 2003). For example, through a pressure microswitch, a child may receive a short listen (e.g. 10 sec) of preferred song, rather than directly switching on the computer or the CD player, not accessible to him/her, considering his/her clinical conditions (Lancioni et al., 2008), or accessible only through the help of caregivers (Lancioni, Singh, et al., 2006). Consequently, they are considered as a great educational and rehabilitative resource (Stasolla & Perilli, 2015).

Some basic guidelines are needs for the effectiveness of a microswitch-based program. First, a plausible behavioral response, naturally present in the person's repertoire and exhibited without excessive effort, should be selected. Second, a microswitch for monitoring the aforementioned response should be adapted. Third, a positive stimulation for motivating the exhibition of such response should be retained. Thus, the response cost (i.e. the effort produced to perform the behavioral response) should be highly compensated by the preferred stimuli (Lancioni et al., 2008).

A second relevant goal of AT-based program is represented by the improving of quality of life (Felce & Perry, 1995). The main construct of quality of life is happiness, which encompasses well-being, pleasure, fulfillment and satisfaction, particularly difficult to detect among people with severe to profound developmental disabilities (Lancioni, Singh, O'Reilly, Oliva, & Basili, 2005). To overcome this issue, professionals refer to behavioral signs of happiness such as smiling, laughing and energized body movements, labeling them indices of happiness (Lancioni et al., 2005) which may represent an outcome measure of positive mood (Ripat & Woodgate, 2011). Thus, interacting independently with the environment, a person with developmental disabilities should increase positive mood with beneficial consequences on the quality of life (Lancioni et al., 2008).

This chapter includes a selective overview of microswitches, describes their use and provides some representative examples concerning their applications to persons with severe to profound developmental and/or multiple disabilities. Specifically, the chapter illustrates:

1. The independent access to positive stimulation (i.e. through the use of a simple behavioral response detected by one microswitch),
2. The independent access to preferred stimuli and/or the opportunity to ask for social contact with a caregiver (i.e. through a combination of microswitch and VOCA),
3. The increase of adaptive response and reduction of challenge behavior (i.e. Microswitch-cluster), and
4. The communication, occupation and leisure skills (i.e. microswitch and computer).

For each group the chapter outlines a selection of empirical evidences, emphasizing strengths and weaknesses. Furthermore, social validation assessments, aimed at formally endorse the interventions programs by practitioners and caregivers, is pointed out. Finally, the chapter will recommend some guidelines for the future research in this topic.

METHOD

Computerized and manual searches were combined within electronic database such as SCOPUS, PSYCHINFO, PUBMED, ERIC using AT, developmental/multiple/learning/intellectual disabilities, communication impairments, quality of life, indices of happiness, positive participation and social validation as keywords. Overall, twenty-seven studies were included, considering as including criteria at least:

1. One participant between 4 and 18 years,
2. One cognitive-behavioral intervention, and
3. Empirical data.

Excluding criteria concerned:

1. Review papers,
2. Adults as participants,
3. Meta-analysis studies, and
4. Pharmacological treatments without behavioral program.

Two studies are briefly summarized within each group.

LITERATURE OVERVIEW

Independent Access to Stimulation

Six studies regarding the independent access to positive stimulation through a single microswitch and including ten participants, were selected. In two studies the microswitch was an optic sensor recording an eye blinking (Lancioni et al., 2010a, 2011). Three studies presented a pair of optic sensors monitoring lip and/or chin movements (Lancioni et al., 2005, 2006, 2009), and one study adopted tilt sensors (Lancioni et al., 2004). For example, Lancioni et al. (2004) employed a grid as a microswitch to detect hand movements of a girl with severe multiple disabilities. The grid had two mercury devices, ending with conductive leads. A lateral movement of the grid would make the mercury drop shift along the conductive leads and activate the sensor, which produced contingently brief periods of preferred stimulation. Analysis showed that the girl increased the frequency of hand movements during the intervention with respect to baseline (where stimulation was not available). The adaptive behavior was maintained during a post-intervention check, implemented one month after the end of intervention.

Stimulation Access and Asking for Social Contact

Seven studies including a total of 22 participants and a combination of one microswitch and one

or two VOCA(s) have been selected (Lancioni et al., 2008a, b, c, 2009a, b, c, 2012). The purpose of these studies was to provide participants with independent access to positive stimulation and opportunity to ask for social contact with their caregivers, and to choose between aforementioned opportunities. For instance, Lancioni, et al. (2009b) conducted two studies implementing microswitch and VOCA technologies. The first one monitored 11 participants in using a microswitch for accessing preferred environmental stimuli and a Voice Output Communication Aid (VOCA) for asking for social contact. The second study carried out a social validation assessment involving 110 students as raters. Results emphasized that all participants learned to use the microswitch and the VOCA. Moreover, the external raters favorably endorsed the use of such technology.

Increasing Adaptive Responses and Reducing Challenge Behaviors

One way to pursue the dual objective of promoting adaptive responses and reducing challenge behavior is constituted by microswitch-cluster. In a first intervention phase, the adaptive behavior is followed by contingent positive stimulation, even if at the same time a challenge behavior occurs. Then, in the cluster-intervention phase the adaptive response is positively reinforced only if the problem behavior is absent. Additionally, positive stimulation is interrupted if a challenge behavior occurred during its supply. Eight studies were retained in this section (Lancioni et al., 2007d, 2008d, 2009d, e, 2013b, Stasolla et al., 2014), including thirteen participants.

For instance, Stasolla et al. (2014) carried out a microswitch-cluster aimed at promoting object manipulation and extinguishing hand mouthing for three boys with severe autism spectrum disorders and profound intellectual disabilities. The adaptive response was detected through a wobble microswitch (that would be pulled, pushed or moved sideways), and the challenging behavior was monitored through an optic sensor

fixed on participants' chin. Data showed that all participants learned to use the cluster technology, increasing their adaptive responses and reducing their challenge behaviors. Moreover, all participants seemed to enjoy the sessions, since their indices of happiness (measured according to a 15-seconds partial interval coding system, with a 10-seconds observation interval followed by a 5-seconds recording interval in terms of presence/absence regarding the indices of happiness within the previous observation interval) improved during intervention compared to the baseline.

Enhancing Occupation, Communication and Leisure Skills

Six studies involving microswitch devices connected to computer systems for promoting engagement and communication, are included in this section (Lancioni et al., 2011b, 2012b, Stasolla, Caffò, et al., 2014; Stasolla, Caffò, Picucci, & Bosco, 2013; Stasolla & De Pace, 2014, Stasolla, Caffò, et al., 2015).

Essentially, the system automatically present different categories of events (e.g. encircled pictures matched with verbal cues), selected through a formal preference screening (Crawford & Schuster, 1993). The response (i.e. activation of microswitch), let the system open a second page, which provides different options within the previous category. A second response is required to choose the selected item. A final confirmation response may be necessary, inquiring whether participants really desired the selected item. Consequently the item would be available to participants for 8-30 seconds.

For example, Stasolla, Caffò, Picucci, & Bosco (2013) worked with three participants with cerebral palsy and severe communication impairments, presenting them a pressure sensor (circular button) connected to a computer system introducing three categories (i.e. food, beverage, leisure) including four items. Participants were required to respond (i.e. activate the pressure sensor) three times to select an item. A first activation to select

a category, a second response for an item and a third to confirm the choice. Indices of happiness were additionally monitored. Data emphasized that all participants increased their request and choices of preferred items during intervention and consolidated their performance during a post-intervention. Indices of happiness augmented as well during those phases, compared to baseline, for all the participants involved.

DISCUSSION

The positive evidences of the studies reviewed emphasize the importance of AT for children with developmental disabilities (Lancioni, Singh, O'Reilly, & Sigafos, 2011; Shih, Chang, & Shih, 2010; Sigafos et al., 2009). Outcomes are definitely higher when:

1. The response requires a low level of effort to be performed,
2. The stimulus events are powerful and motivating, and
3. The intervention time is sufficient to meet the person's learning conditions (Lancioni Sigafos, O'Reilly, & Singh, 2012; Saunders et al., 2003).

A suitable response may also include less conventional forms of behavior like vocalizations, eye blinking, or chin and lip movements (Lancioni et al., 2012; Lancioni & Singh, 2014), especially with children with very minimal motor behavior. These persons could not benefit from intervention programs involving typical motor responses and traditional microswitches (Lancioni et al., 2005; Lancioni & Singh, 2014). Providing these children with the possibility to be active and to decide about their environmental stimulation can be considered highly relevant for improving the individuals' quality of life (Lachapelle et al., 2005; Petry, Maes, & Vlaskamp, 2005). In fact, being constructively engaged and determining independently the level of stimulation, increase the person's overall satis-

faction, improve the general mood, and enhance social image and status (Browder, Wood, Test, Karvonen, & Algozzine, 2001; Karvonen, Test, Wood, Browder, & Algozzine, 2004; Petry et al., 2005; Wehmeyer & Schwartz, 1998; Zekovic & Renwick, 2003).

The selection of stimuli is provided through indirect and direct strategies, such as observations of the persons within their environment, interviews with parents, staff and caregivers, and stimulus preference screening procedures (Crawford & Schuster, 1993; Stasolla et al., 2015). Two features that could help to maximize the success are represented by:

1. The selection and use of more than one event so as to avoid saturation risks, and
2. Repetitions of the selection procedures over time to upgrade to the person's current interest (Saunders et al., 2003).

The studies summarized earlier in this chapter considered basic adaptive responses and specific forms of problem behavior or inadequate posture (e.g., hand mouthing, eye poking, and head forward tilting) (Stasolla et al., 2015). Supplementary adaptive responses and other problem behaviors may also be targeted to extend the applicability of the approach.

Despite the high potential of programs involving microswitches, one can state that they do not allow the participant to ask for social contact with the caregiver. Combining a VOCA to request caregiver attention, mediation or social contact with regular microswitches that allow direct access to environmental stimuli, may be considered a fairly straightforward strategy, as indicated by the studies reviewed earlier (Lancioni et al., 2008, 2009; Lancioni & Singh, 2014; Schlosser & Sigafoos, 2006; Sigafoos et al., 2009).

These intervention strategies seem to foster communication opportunities for those children. Participants may choose between selecting preferred item (e.g., a video or a song), asking for the mediation of caregivers or accessing the literacy

process, within the same rehabilitative program, fostering the children's communicative potential (Lancioni & Singh, 2014; Stasolla et al., 2014).

A similar consideration may be formulated for cognitive-behavioral interventions involving the use of AT for individuals with pervasive developmental disabilities and high functioning. Notably, in school settings, one may plan a computer-based intervention program aimed at providing students with learning difficulties, a wide range of academic activities, pursuing the dual objective of improving self-determination and constructive engagement on one side and eventually reducing stereotypic or challenge behaviors on the other. Eventually, one may argue that saturation is also prevented (Lancioni & Singh, 2014).

Summarizing, this chapter emphasizes that these forms of AT represent crucial educational and rehabilitative resources for children with developmental disabilities in daily contexts. However, research initiatives are necessary to extend and upgrade the aforementioned forms of technology for adapting such microswitches to different response situations and provide them with minimal invasion/contact (i.e., realizing devices that do not need to be fixed onto the person's body) (Leung & Chau, 2010).

CONCLUSION

The studies reviewed identify four intervention approaches for children with developmental disabilities with various levels of functioning. Despite the outcomes of the studies were largely positive, readers should cautiously interpret some of these findings. In fact, questions arose about:

1. Some of the measures used to determine the effects of the intervention,
2. The lack of control over the time variable, and
3. The absence of emphasis on the participants' acquisition of a constructive engagement role.

Given the aforementioned criticisms, the extension of research may be necessary. Furthermore, it is crucial to analyze the studies reviewed in terms of:

1. Size, (i.e., the quantitative relevance of the changes observed) and
2. Credibility/reliability based on the methodological conditions of the intervention programs (Barlow, Nock, & Hersen, 2009; Kennedy, 2005).

With regard to the first point, some studies reported significant changes in the participants' behavior, measured through specific scales or by means of response frequencies (Schiff et al., 2007), other:

1. Reported changes only related to behavioral ratings expressed by professionals or other representative figures involved in the intervention and/or
2. Presented partial, inconclusive or negative statistical evidence on formal measures (Barreca et al., 2003).

The outcomes of the studies using single-subject designs could be taken as more solid and reliable, because these designs allow for control over the time (history) variable (Kennedy, 2005). Quantitative relevance and credibility are the two important criteria on which judging the outcomes. A supplementary criterion could be the practical relevance of changes observed in terms of participants' environmental engagement and interaction involvement. With regard to this criterion, it might be stated that not all the studies code this outcome, but it is becoming more frequent.

FUTURE RESEARCH

In light of above, the following perspectives for future research should be investigated. First, the extension of AT for other children with developmental disabilities. Those efforts should consider:

1. The extensive literature on the use of this intervention approach in other rehabilitation trades;
2. The ethical considerations related to the participants' legal representatives right to decide on the use of such approach; and
3. The need for methodological sophistication.

Moreover, rehabilitative interventions through a microswitch-based program should consider the individual's involvement in the decision making and the positive interaction with the outside world (Lancioni & Singh, 2014).

A second extension may concern the settings, the generalization and the maintenance over time. With respect to the first point, one may design intervention programs in different settings, involving school, home, or community. With respect to the second point, one may consider different tasks, caregivers, research assistants, parents or teachers. With respect to the third point, one may systematically include maintenance, generalization and post intervention phases within those studies (Stasolla, Perilli, & Damiani, 2014).

A third extension may relate to the technologies. One may design new technological solutions responding to participants' characteristics on the one hand, and to contexts' resources on the other. With respect to participants' characteristics, one may keep in mind new technologies that ensure that participants can interact constructively with minimal effort. With respect to the second point, one should consider the financial resources available to families and to rehabilitative centers (Lancioni, Sigafoos, O'Reilly, & Sigafoos, 2012).

REFERENCES

- Barlow, D. H., Nock, M., & Hersen, M. (2009). *Single case experimental designs: Strategies for studying behavior change* (3rd ed.). New York: Allyn and Bacon.
- Barreca, S., Velikonja, D., Brown, L., Williams, L., Davis, L., & Sigouin, C. S. (2003). Evaluation of the effectiveness of two clinical training procedures to elicit yes/no responses from patients with a severe acquired brain injury: A randomized single-subject design. *Brain Injury*, 17(12), 1065–1075. doi:10.1080/0269905031000110535 PMID:14555365
- Borg, J., Larson, S., & Ostenberg, P. O. (2011). The right to assistive technology: For whom, for what, and by whom? *Disability & Society*, 26(2), 151–167. doi:10.1080/09687599.2011.543862
- Browder, D. M., Wood, W. M., Test, D. W., Karvonen, M., & Algozzine, B. (2001). Reviewing resources on self-determination: A map for teachers. *Remedial and Special Education*, 22(4), 233–244. doi:10.1177/074193250102200407
- Chiapparino, C., Stasolla, F., de Pace, C., & Lancioni, G. E. (2011). A touch pad and a scanning keyboard emulator to facilitate writing by a woman with extensive motor disability. *Life Span and Disability*, 14, 45–54.
- Crawford, M. R., & Schuster, J. W. (1993). Using microswitches to teach toy use. *Journal of Developmental and Physical Disabilities*, 5(4), 349–368. doi:10.1007/BF01046391
- De Pace, C., & Stasolla, F. (2014). Promoting Environmental Control, Social Interaction, and Leisure/Academy Engagement Among People with Severe/Profound Multiple Disabilities Through Assistive Technology. In G. Kouroupetroglou (Ed.), *Assistive Technologies and Computer Access for Motor Disabilities* (pp. 285–319). doi:10.4018/978-1-4666-4438-0.ch010
- Felce, D., & Perry, J. (1995). Quality of life: Its definition and measurement. *Research in Developmental Disabilities*, 16(1), 51–74. doi:10.1016/0891-4222(94)00028-8 PMID:7701092
- Kagohara, D. M., Sigafoos, J., Achmadi, D., van der Meer, L., O'Reilly, M. F., & Lancioni, G. E. (2011). Teaching students with developmental disabilities to operate an iPod Touch to listen to music. *Research in Developmental Disabilities*, 32(6), 2987–2992. doi:10.1016/j.ridd.2011.04.010 PMID:21645989
- Karvonen, M., Test, D. W., Wood, W. M., Browder, D., & Algozzine, B. (2004). Putting self-determination into practice. *Exceptional Children*, 71(1), 23–41. doi:10.1177/001440290407100102
- Kennedy, C. (2005). *Single case designs for educational research*. New York: Allyn & Bacon.
- Lachapelle, Y., Wehmeyer, M. L., Haelewyck, M. C., Courbois, Y., Keith, K. D., Schalock, R., & Walsh, P. N. et al. (2005). The relationship between quality of life and self-determination: An international study. *Journal of Intellectual Disability Research*, 49(10), 740–744. doi:10.1111/j.1365-2788.2005.00743.x PMID:16162119
- Lancioni, G. E., Bellini, D., Oliva, D., Singh, N. N., O'Reilly, M. F., & Sigafoos, J. (2010). Camera-based microswitch technology for eyelid and mouth responses of persons with profound multiple disabilities: Two case studies. *Research in Developmental Disabilities*, 31(6), 1509–1514. doi:10.1016/j.ridd.2010.06.006 PMID:20598501
- Lancioni, G. E., O'Reilly, M. F., Singh, N. N., Sigafoos, J., Oliva, D., Smaldone, A., & Chiapparino, C. et al. (2009c). Persons with multiple disabilities access stimulation and contact the caregiver via microswitch and VOCA technology. *Life Span and Disability*, 12, 119–128.

- Lancioni, G. E., Olivetti Belardinelli, M., Stasolla, F., Singh, N. N., O'Reilly, M. F., Sigafoos, J., & Angelillo, M. T. (2008a). Promoting engagement, requests and choice by a man with post-coma pervasive motor impairment and minimally conscious state through a technology-based program. *Journal of Developmental and Physical Disabilities*, 20(4), 379–388. doi:10.1007/s10882-008-9104-x
- Lancioni, G. E., O'Reilly, M. F., Singh, N. N., Buonocunto, F., Sacco, V., Colonna, F., & Bosco, A. et al. (2009). Technology-based intervention options for post-coma persons with minimally conscious state and pervasive motor disabilities. *Developmental Neurorehabilitation*, 12(1), 24–31. doi:10.1080/17518420902776995 PMID:19283531
- Lancioni, G. E., O'Reilly, M. F., Singh, N. N., Sigafoos, J., Didden, R., Oliva, D., & Groeneweg, J. et al. (2009b). Persons with multiple disabilities accessing stimulation and requesting social contact via microswitch and VOCA devices: New research evaluation and social validation. *Research in Developmental Disabilities*, 30(5), 1084–1094. doi:10.1016/j.ridd.2009.03.004 PMID:19361954
- Lancioni, G. E., O'Reilly, M. F., Singh, N. N., Sigafoos, J., Oliva, D., Alberti, G., & Lang, R. et al. (2013). Technology-based programs to support adaptive responding and reduce hand mouthing in two persons with multiple disabilities. *Journal of Developmental and Physical Disabilities*, 25(1), 65–77. doi:10.1007/s10882-012-9303-3
- Lancioni, G. E., O'Reilly, M. F., Singh, N. N., Sigafoos, J., Oliva, D., & Severini, L. (2008b). Enabling two persons with multiple disabilities to access environmental stimuli and ask for social contact through microswitches and a VOCA. *Research in Developmental Disabilities*, 29(1), 21–28. doi:10.1016/j.ridd.2006.10.001 PMID:17174529
- Lancioni, G. E., O'Reilly, M. F., Singh, N. N., Sigafoos, J., Oliva, D., & Severini, L. (2008c). Three persons with multiple disabilities accessing environmental stimuli and asking for social contact through microswitch and VOCA technology. *Journal of Intellectual Disability Research*, 52(4), 327–336. doi:10.1111/j.1365-2788.2007.01024.x PMID:18339095
- Lancioni, G. E., O'Reilly, M. F., Singh, N. N., Sigafoos, J., Oliva, D., Smaldone, A., & Chirico, M. et al. (2011b). Technology-assisted programs for promoting leisure or communication engagement in two persons with pervasive motor or multiple disabilities. *Disability and Rehabilitation. Assistive Technology*, 6(2), 108–114. doi:10.3109/17483107.2010.496524 PMID:20545564
- Lancioni, G. E., O'Reilly, M. F., Singh, N. N., Sigafoos, J., Tota, A., Antonucci, M., & Oliva, D. (2006). Children with multiple disabilities and minimal motor behavior using chin movements to operate microswitches to obtain environmental stimulation. *Research in Developmental Disabilities*, 27(3), 290–298. doi:10.1016/j.ridd.2005.02.003 PMID:16005183
- Lancioni, G. E., O'Reilly, M. F., Singh, N. N., Stasolla, F., Manfredi, F., & Oliva, D. (2004a). Adapting a grid into a microswitch to suit simple hand movements of a child with profound multiple disabilities. *Perceptual and Motor Skills*, 99(2), 724–728. doi:10.2466/pms.99.2.724-728 PMID:15560365
- Lancioni, G. E., Sigafoos, J., O'Reilly, M. F., & Singh, N. N. (2012). *Assistive Technology, Interventions for Individuals with Severe/Profound and Multiple Disabilities*. New York: Springer.
- Lancioni, G. E., & Singh, N. N. (2014). *Assistive technologies for people with diverse abilities*. New York: Springer. doi:10.1007/978-1-4899-8029-8

- Lancioni, G. E., Singh, N. N., O'Reilly, M. F., & Sigafoos, J. (2011). Assistive technology for behavioral interventions for persons with severe/profound multiple disabilities: A selective overview. *European Journal of Behavior Analysis*, 12, 7–26.
- Lancioni, G. E., Singh, N. N., O'Reilly, M. F., Sigafoos, J., Oliva, D., & Severini, L. et al. (2007d). Microswitch technology to promote adaptive responses and reduce mouthing in two children with multiple disabilities. *Journal of Visual Impairment & Blindness*, 101, 628–636.
- Lancioni, G. E., Singh, N. N., O'Reilly, M. F., Green, V. A., Oliva, D., Buonocunto, F., & Di Nuovo, S. et al. (2012b). Technology-based programs to support forms of leisure engagement and communication for persons with multiple disabilities: Two single-case studies. *Developmental Neurorehabilitation*, 15(3), 209–218. doi:10.3109/17518423.2012.666766 PMID:22582852
- Lancioni, G. E., Singh, N. N., O'Reilly, M. F., Oliva, D., & Basili, G. (2005). An overview of research on increasing indices of happiness of people with severe/profound intellectual and multiple disabilities. *Disability and Rehabilitation*, 27(3), 83–93. doi:10.1080/09638280400007406 PMID:15823988
- Lancioni, G. E., Singh, N. N., O'Reilly, M. F., Sigafoos, J., Chiapparino, C., Stasolla, F., & Oliva, D. (2007d). Using an optic sensor and a scanning keyboard emulator to facilitate writing by persons with pervasive motor disabilities. *Journal of Developmental and Physical Disabilities*, 19(6), 593–603. doi:10.1007/s10882-007-9073-5
- Lancioni, G. E., Singh, N. N., O'Reilly, M. F., Sigafoos, J., Buonocunto, F., Sacco, V., & De Pace, C. et al. (2009a). Two persons with severe post-coma motor impairment and minimally conscious state use assistive technology to access stimulus events and social contact. *Disability and Rehabilitation. Assistive Technology*, 4(5), 367–372. doi:10.1080/17483100903038584 PMID:19565377
- Lancioni, G. E., Singh, N. N., O'Reilly, M. F., Sigafoos, J., Didden, R., & Oliva, D. (2009d). A technology-based stimulation program to reduce hand mouthing by an adolescent with multiple disabilities. *Perceptual and Motor Skills*, 109(2), 478–486. doi:10.2466/pms.109.2.478-486 PMID:20038002
- Lancioni, G. E., Singh, N. N., O'Reilly, M. F., Sigafoos, J., Didden, R., & Oliva, D. (2009e). Two boys with multiple disabilities increasing adaptive responding and curbing dystonic/spastic behavior via a microswitch-based program. *Research in Developmental Disabilities*, 30(2), 378–385. doi:10.1016/j.ridd.2008.07.005 PMID:18760566
- Lancioni, G. E., Singh, N. N., O'Reilly, M. F., Sigafoos, J., Didden, R., Oliva, D., & Cingolani, E. (2008d). A girl with multiple disabilities increases object manipulation and reduces hand mouthing through a microswitch-based program. *Clinical Case Studies*, 7(3), 238–249. doi:10.1177/1534650107307478
- Lancioni, G. E., Singh, N. N., O'Reilly, M. F., Sigafoos, J., Oliva, D., Piazzolla, G., & Manfredi, F. et al. (2007c). Automatically delivered stimulation for walker-assisted step responses: Measuring its effects in persons with multiple disabilities. *Journal of Developmental and Physical Disabilities*, 19(1), 1–13. doi:10.1007/s10882-006-9030-8
- Lancioni, G. E., Singh, N. N., O'Reilly, M. F., Sigafoos, J., Oliva, D., Smaldone, A., & Groeneweg, J. et al. (2010a). Promoting ambulation responses among children with multiple disabilities through walkers and microswitches with contingent stimuli. *Research in Developmental Disabilities*, 31(3), 811–816. doi:10.1016/j.ridd.2010.02.006 PMID:20207105
- Leung, B., & Chau, T. (2010). A multiple camera tongue switch for a child with severe spastic quadriplegic cerebral palsy. *Disability and Rehabilitation. Assistive Technology*, 5(1), 58–68. doi:10.3109/17483100903254561 PMID:19941441

- Lontis, E. R., & Struijk, L. N. (2010). Design of inductive sensors for tongue control system for computers and assistive devices. *Disability and Rehabilitation. Assistive Technology*, 5(4), 266–271. doi:10.3109/17483101003718138 PMID:20307253
- Petry, K., Maes, B., & Vlaskamp, C. (2005). Domains of quality of life of people with profound multiple disabilities: The perspective of parents and direct support staff. *Journal of Applied Research in Intellectual Disabilities*, 18(1), 35–46. doi:10.1111/j.1468-3148.2004.00209.x
- Reichle, J. (2011). Evaluating assistive technology in the education of persons with severe disabilities. *Journal of Behavioral Education*, 20(1), 77–85. doi:10.1007/s10864-011-9121-1
- Saunders, M. D., Timler, G. R., Cullinan, T. B., Pilkey, S., Questad, K. A., & Saunders, R. R. (2003). Evidence of contingency awareness in people with profound multiple impairments: Response duration versus response rate indicators. *Research in Developmental Disabilities*, 24(4), 231–245. doi:10.1016/S0891-4222(03)00040-4 PMID:12873657
- Schiff, N. D., Giacino, J. T., Kalmar, K., Victor, J. D., Baker, K., Gerber, M., & Rezai, A. R. et al. (2007). Behavioural improvements with thalamic stimulation after severe traumatic brain injury. *Nature*, 448(7153), 600–603. doi:10.1038/nature06041 PMID:17671503
- Schlosser, R., & Sigafoos, J. (2006). Augmentative and alternative communication interventions for persons with developmental disabilities: Narrative review of comparative single-subject experimental studies. *Research in Developmental Disabilities*, 27(1), 1–29. doi:10.1016/j.ridd.2004.04.004 PMID:16360073
- Shih, C. H., Chang, M. L., & Shih, C. T. (2010). A new limb movement detector enabling people with multiple disabilities to control environmental stimulation through limb swing with a gyration air mouse. *Research in Developmental Disabilities*, 31(4), 875–880. doi:10.1016/j.ridd.2010.01.020 PMID:20381996
- Sigafoos, J., Green, V. A., Payne, D., Son, S. H., O'Reilly, M. F., & Lancioni, G. E. (2009). A comparison of picture exchange and speech-generating devices: Acquisition, preference, and effects on social interaction. *Augmentative and Alternative Communication*, 25(2), 99–109. doi:10.1080/07434610902739959 PMID:19444681
- Stasolla, F., & Caffò, A. O. (2013). Promoting adaptive behaviors by two girls with Rett Syndrome through a microswitch-based program. *Research in Autism Spectrum Disorders*, 7(10), 1265–1272. doi:10.1016/j.rasd.2013.07.010
- Stasolla, F., Caffò, A. O., Damiani, R., Perilli, V., Di Leone, A., & Albano, V. (2015). Assistive technology-based programs to promote communication and leisure activities by three children emerged from a minimal conscious state. *Cognitive Processing*, 16(1), 69–78. doi:10.1007/s10339-014-0625-1 PMID:25077461
- Stasolla, F., Caffò, A. O., Picucci, L., & Bosco, A. (2013). Assistive technology for promoting choice behaviors in three children with cerebral palsy and severe communication impairments. *Research in Developmental Disabilities*, 34(9), 2694–2700. doi:10.1016/j.ridd.2013.05.029 PMID:23770888
- Stasolla, F., Damiani, R., Perilli, V., Di Leone, A., Albano, V., Stella, A., & Damato, C. (2014). Technological supports to promote choice opportunities by two boys with fragile X syndrome and severe to profound developmental disabilities. *Research in Developmental Disabilities*, 35(11), 2993–3000. doi:10.1016/j.ridd.2014.07.045 PMID:25118066

Stasolla, F., & De Pace, C. (2014). Assistive technology to promote leisure and constructive engagement by two boys emerged from a minimal conscious state. *NeuroRehabilitation*, 35, 253–259. PMID:24990021

Stasolla, F., Perilli, V., Damiani, R., Caffò, A. O., Di Leone, A., Albano, V., & Damato, C. et al. (2014). A microswitch-cluster program to enhance object manipulation and to reduce hand mouthing by three children with autism spectrum disorders and intellectual disabilities. *Research in Autism Spectrum Disorders*, 8(9), 1071–1078. doi:10.1016/j.rasd.2014.05.016

Stasolla, F., Perilli, V., Di Leone, A., Damiani, R., Albano, V., Stella, A., & Damato, C. (2015). Technological aids to support choice strategies by three girls with Rett syndrome. *Research in Developmental Disabilities*, 36, 36–44. doi:10.1016/j.ridd.2014.09.017 PMID:25310833

Wehmeyer, M. L., & Schwartz, M. (1998). The relationship between self-determination, quality of life, and life satisfaction for adults with mental retardation. *Education and Training in Mental Retardation and Developmental Disabilities*, 33, 3–12.

Zekovic, B., & Renwick, R. (2003). Quality of life for children and adolescents with developmental disabilities: Review of conceptual and methodological issues relevant to public policy. *Disability & Society*, 18(1), 19–34. doi:10.1080/713662199

ADDITIONAL READING

Agree, E. M. (2014). The potential for technology to enhance independence for those aging with a disability. *Disability and Health Journal*, 7(1), S33–S39. doi:10.1016/j.dhjo.2013.09.004 PMID:24456682

Hoppestad, B. S. (2013). Current perspective regarding adults with intellectual and developmental disabilities accessing computer technology. *Disability and Rehabilitation. Assistive Technology*, 8(3), 190–194. doi:10.3109/17483107.2012.723239 PMID:23078291

Sauer, A. L., Parks, A., & Heyn, P. C. (2010). Assistive technology effects on the employment outcomes for people with cognitive disabilities: A systematic review. *Disability and Rehabilitation. Assistive Technology*, 5(6), 377–391. doi:10.3109/17483101003746360 PMID:20392190

KEY TERMS AND DEFINITIONS

Assistive Technology: Umbrella including any technological device, equipment and/or tool ensuring people with disabilities to better function within their context and/or environment of daily life.

Constructive Engagement: Active and/or positive interaction of individuals with disabilities towards the outside world, enhancing self-determination.

Developmental Disabilities: Cognitive, intellectual, motor and/or sensorial disabilities throughout the life span.

Indices of Happiness: Signs of happiness such as smiling, laughing and/or excited body movements usually exhibited by non verbal individuals.

Microswitches: Electronic devices enabling persons with severe to profound developmental disabilities to independently and autonomously access to positive stimulation.

Quality of Life: Complex psychological construct including well-being, satisfaction, fulfillment and happiness.

Social Validation: Assessment of the effectiveness, the impact and the quality of a rehabilitative program, involving parents, teachers, students and/or caregivers external to the proposed intervention as raters.

Design, Manufacture, and Selection of Ankle–Foot–Orthoses

Hasan Kemal Surmen

Istanbul University, Turkey

Nazif Ekin Akalan

Istanbul University, Turkey

Yunus Ziya Arslan

Istanbul University, Turkey

INTRODUCTION

Ankle-foot-orthoses (AFOs) are externally applied assistive devices that are prescribed to the patients with neuromuscular dysfunctions in order to improve abnormal lower limb motor functions. AFOs are mainly used to control the range of motion of the ankle joint, to compensate for the muscle weakness caused by different motor-neuron diseases, to improve the gait functions during post-operative stages and to optimize the efficiency of walking.

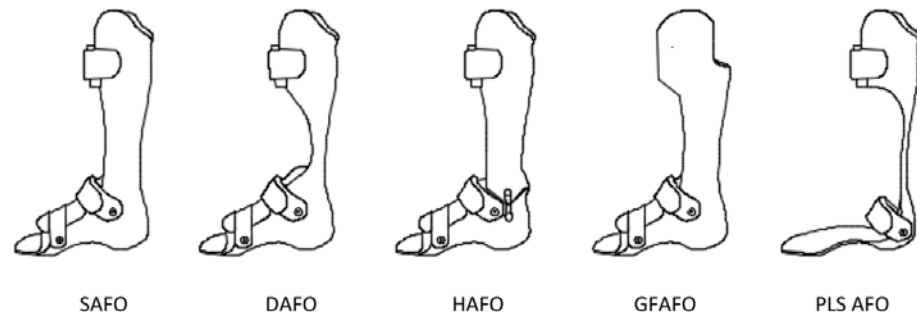
Different types of AFOs are used to treat different dysfunctions. Each type of AFOs has its characteristic function. However, AFOs with same function can have different designs that differ in material, geometry, additional mechanism and components which affect the comfort, cost of AFO and oxygen consumption of patients. Additionally, recent advances in different technology areas, such as additive manufacturing (AM), three dimensional (3D) scanning and CAD-CAM (computer aided design-computer aided manufacturing) have led to new designs and manufacturing methods for AFOs. The objective of this chapter is to provide a survey on design, manufacture and selection of AFOs.

BACKGROUND

First of all, it would be beneficial to describe orthosis and prosthesis concepts that are mostly confused with each other. Briefly, orthoses are braces to support dysfunction of a body part, while prostheses are artificial parts to replace a missing body part. Prostheses are devices for external and internal use. External prostheses, such as prosthetic legs or prosthetic breast form used after mastectomy (Lake, Ahmad, & Dobrashian, 2013), can be employed for cosmetic and also functional aims with the developments in prosthetic technology. On the other hand, internal prostheses, such as artificial knee joints (Guo, Hao, & Wan, 2016) and cataract lenses (Heys & Truscott, 2008) are devices which are surgically implanted within a body.

Orthoses are assistive devices that are used to align, protect and assist limbs or body parts besides supporting to treat deformities. Orthoses can be used for neurological conditions, injuries and congenital deformities. Orthoses are designed as standard or custom made forms from an individual mold in the shape of patient's foot. Orthoses can be divided into two classes, i.e. *i*) standard orthoses for general use and *ii*) custom made orthoses

Figure 1. Typical examples of ankle foot orthoses



that are prescribed for more complex conditions. Orthoses are used for lower extremity (Moisan & Cantin, 2016), upper extremity (Belda-Lois et al., 2006), and spine (Hofmann et al., 2016). Lower extremity orthoses have a wide range of use that are designed for hip, knee and ankle joints' immobilizations. They reduce energy consumption and pain as assisting the gait and improving the posture. Development of lower extremity orthotic technologies and new materials lead to new designs and manufacturing methods, and also affect selection criteria of orthoses.

AFOs are braces encompassing the lower leg, ankle joint and foot of the patients. AFOs provide stability in the ankle joint and biomechanical control above and below of ankle. For example, a patient with crouch gait pathology (walking with flexed knees) can reduce knee flexion during stance phase by using an AFO. Because, AFO produces a moment around the ankle joint that prevents ankle dorsiflexion in stance phase which prevents excessive knee flexion by directing the ground reaction force in front of the knee joint center. They are manufactured using metal and plastic materials. However, plastic AFOs are more preferred than metal ones, because they are lighter and more cosmetic (Franceschini et al., 2001). Also it was reported that custom plastic AFOs decrease oxygen consumption in the patients. However, the patients, who want to use AFO, should have sufficient active hip flexion to propel their legs. And their quadriceps muscle strength should be greater than four or five grade according to manual muscle test (Hsu, Michael, & Fisk, 2008).

There are several different types of AFOs for different biomechanical aims (Figure 1). Solid ankle foot orthosis (SAFO) (Ridgewell, Rodda, Graham, & Sangeux, 2015) rigidly supports ankle and prevent any movement at the ankle. Dynamic ankle foot orthosis (DAFO) provides subtalar stabilization. Unlike solid AFO models, this device allows ankle to dorsiflex and partially limits the plantarflexion (Sherief, Gazya, & El Gafaar, 2015). Hinged ankle foot orthosis (HAFO) is also a type of dynamic AFO which let the dorsiflexion exists during gait. On the other hand, HAFO is commonly used to restrict three-dimensional ankle mobility and limit the motion of ankle joint within the sagittal plane (Leardini, Aquila, Caravaggi, Ferraresi, & Giannini, 2014). Ground reaction ankle foot orthosis (GRAFO) is used to reduce excessive knee flexion (Ries & Schwartz, 2015). This type of orthoses has a solid part below the knee (pre-tibial support) which doesn't allow the knee joint moving forward. Posterior leaf spring ankle foot orthosis (PLS AFO) is used to primarily for foot drop in order to control plantarflexion during heel strike and swing phases to improve the functional quality of locomotion (Leone, 1987). All these AFOs have different characteristics, since they are designed for specific goals. Different characteristics of AFOs meet specific needs which result from injuries and diseases, such as foot drop (Everaert et al., 2013), cerebral palsy (van Beeten, Hartman, & Houdijk, 2015), spina bifida (Duffy, 1997) and hemiplegia (Nolan, Savalia, Lequerica, & Elovic, 2009).

AFO, with physical therapy combination, is widely used to provide adequate proper heel contact at initial-contact, to prevent premature heel-rise and to increase stance phase stability during walking (Rethlefsen, Kay, Dennis, Forstein, & Tolo, 1999). Solid, articulated and leaf spring AFOs are widely used in clinics which, in literature, were reported improving gait velocity and gait dynamics (White, Jenkins, Neace, Tylkowski, & Walker, 2002), preventing excessive plantar flexion in stance, providing adequate dorsiflexion in initial-contact and swing phases (Lam, Leong, Li, Hu, & Lu, 2005; Romkes, Hell, & Brunner, 2006), and improving natural position of the foot in late swing for hemiplegic children (Van Gestel, Molenaers, Huenaerts, Seyler, & Desloovere, 2008). Solid AFOs are used to prevent excessive plantar and dorsiflexion of the ankle during walking to enhance the stability in stance, reduce the abnormal motion at ankle and foot. The simplest way of creating a hinge motion in AFO is to trim material away around the ankle, which makes the material more flexible at this point. This is so called posterior leaf spring AFO, which capable of limiting plantar flexion and allowing dorsiflexion as required. The amount of dorsiflexion is directly related with the amount of trimming from the back of the ankle (Morris, 2007). On the other hand, hinged AFO is only prescribed to prevent either plantar flexion or dorsiflexion of the ankle (Morris, 2007). It was revealed in literature that dynamic parameters of walking are significantly improved by using solid and articulated AFO, although, in terms of these gait parameters, there was no significant superiority of any of these AFOs on another (Radtka, Skinner, & Johanson, 2005; Radtka, Skinner, Dixon, & Johanson, 1997; Eddison & Chockalingam, 2013).

Sophisticated analysis tools (computerized gait analysis, force platform, electromyography etc.), video based gait analysis methods (Edinburg Visual Gait Scale) or functional assessment scales (Gross Motor Functional Measure, PEDI etc.) are utilized to evaluate the gait parameters (Bella, Rodrigues, Valenciano, Silva, & Souza, 2012;

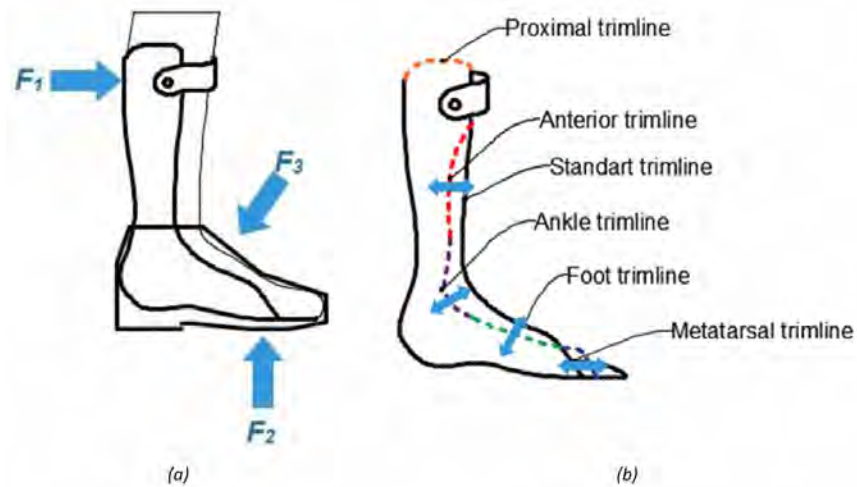
Dalvand, Dehghan, Feizi, Hosseini, & Amirsalari, 2013). These tools are valuable to compare the effects of the AFO for the patient and to tune the AFO in order to increase the influence on related gait parameters (gait velocity, step length, kinetics and kinematics).

DESIGN OF AFOS

AFO is designed to control the motion of the ankle joint, and to improve the gait function of patients with motor impairments. As shown in Figure 2a, an AFO applies forces to three different points of the limb (Edelstein & Bruckner, 2002). F_1 is applied to the proximal-posterior calf, F_2 to the foot sole and F_3 to the dorsal foot (Figure 2a). Three point pressure system will help manage the deformities such as excessive pronation and valgus angles. This system limits the motion around the joint axes and therefore rotations around the joint axes could be managed and joint stabilizations could be provided.

Stiffness, geometrical shape and material type are the main parameters which give to AFO its characteristic properties. AFO stiffness is an important parameter that directly has relation with other parameters taken into account AFO design, and should be determined properly to reduce the gait deficiencies (Esposito, Blanck, Harper, Hsu, & Wilken, 2014). AFO's stiffness depends on the type and level of deformity, and weight of patient; briefly the patients' biomechanical conditions. Different types of AFOs have different stiffness values. For example, a solid AFO should keep the ankle stable and inflexible. Therefore, they are designed thicker in the ankle section. AFO can have flexibility with a thinner thickness or a trimmed ankle section. Location of trimlines is determined according to goals such as holding ankle in fixed position, assisting dorsiflexion, allowing free dorsiflexion with free or restricted ankle plantar flexion. Also, trimline severity is an effective parameter on stiffness of AFO (Bielby et al., 2010). "How much section must be trimmed"

Figure 2. (a) Three-point force system, (b) trimlines for some design options



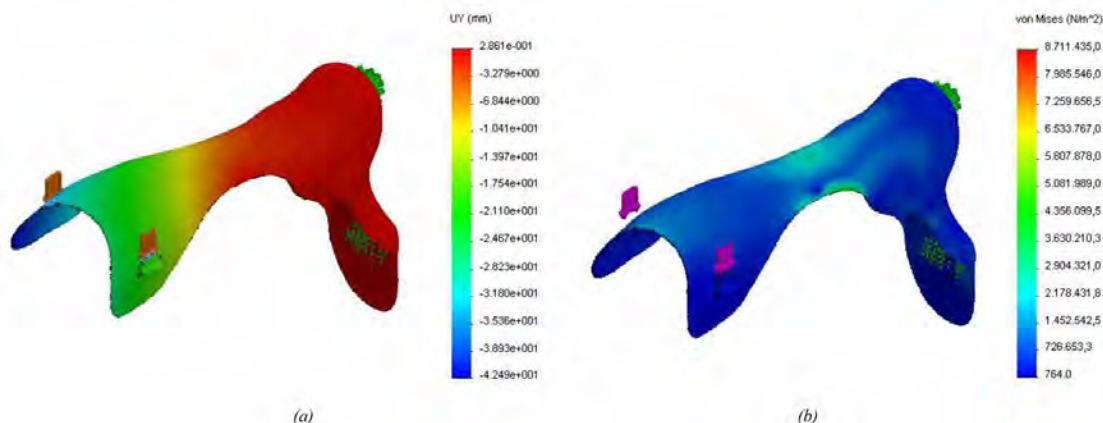
and “from where AFO is needed to be cut” should be determined in such a way to obtain an optimal stiffness. Therefore, geometrical shape of AFO is a key factor on determination of stiffness. Geometry and size of trimmed area and its location will affect allowed limits of the range of motion of ankle joint. Because, displacement distribution over AFO during stance phase of gait is directly related to the trimmed area in the AFO. In standard trim form, proximal trimline becomes 1.5–2 cm below the fibula head, ankle trimline must be 1 cm anterior to the tip of the malleoli and metatarsal trimline must be just posterior to metatarsal heads. For a solid AFO design, an existing standard trimline is valid (Figure 2b). If a flexible AFO is intended to be designed, then ankle trimline and anterior trimline should be moved to behind the standard ankle trimline. To control the forefoot adduction metatarsal trimline, it should be moved forward covering metatarsal heads and toes (Figure 2b).

Also trimlines affect stress distribution over the AFO, when a force is applied to it (Figure 3). It is important to distribute the stress homogeneously over the AFO to not lead to plastic deformity. At this point, it would be beneficial to take advantage of a finite element software. Finite element method (FEM) has been used for designing more efficient AFOs (Ramsey, 2011). FEM is a numerical method

to simulate the behavior of physical systems and presents time and cost effective solutions for various engineering problems. In FEM, a numerical model is constructed after the physical problem is translated into a mathematical model and then it is solved using a computer (Dhatt, Lefrançois, & Touzot, 2012). FEM enables to observe the effects of different trimlines on the stress/strain distribution over AFO models to test different AFO designs.

AFO models have free-form geometry. So it is difficult to draw a 3D AFO model with a CAD software. However, scanning technology provides having a complete 3D AFO model for analyzing in a finite element analysis (FEA) software. 3D scanning is commonly preferred for custom-made designs, prototyping, reverse engineering, industrial design, orthotic and prosthetic design, entertainment industry, inspection and digital archiving. 3D scanners basically convert the physical data of an object into digital data. The purpose of these systems is to construct a point cloud of the object. This data can be turned into 3D surface model of the scanned object and then it is possible to obtain 3D CAD model of the object. There are various systems for 3D scanning operation using different technologies such as laser, white light and x-ray scan. 3D optical scan is also one of the

Figure 3. (a) Displacement and (b) stress distribution of a trimmed AFO model



most popular scanning systems using for orthotic and prosthetic design works in biomechanics and dental areas (Balasundaram, Gurun, Neely, Ash-Rafzadeh, & Ravichandra, 2014; Kang, Kim, & Kim, 2016; Mikkelsen, Skorini, & Løgstrup Andersen, 2011). This scanning system consists of several steps (Figure 4). The object should be sprayed for obtaining matt surfaces before scanning. Different sizes of fringes are projected over the subject and then their photos are taken from different views by two high resolution digital cameras. The number of taken photos depends on size and complexity of the object. Later these photos are bought together and a point- cloud model is obtained. This model is converted into a 3D CAD model by determining surface number in the scan software. Surface number can be increased depending on the complexity of the object. At the final step, the model is verified by using deviation analysis according to a predetermined tolerance value.

Another design parameter is the material type which affects the stiffness level of the AFO. Material type is also related to other mechanical properties of AFOs. Properties, such as durability, fatigue strength, corrosion resistance and process ability depend on the material type. Moreover, material type is a determining factor in terms of the manufacturing cost of AFO. In AFO design, lightweight materials are desired and encouraged,

especially when considering young patients. Polypropylene (Bregman et al., 2010), carbon fiber (King, Mnatsakanian, & Kissel, 2015), metal (Sherk, 2008) and leather are the materials used in AFO design.

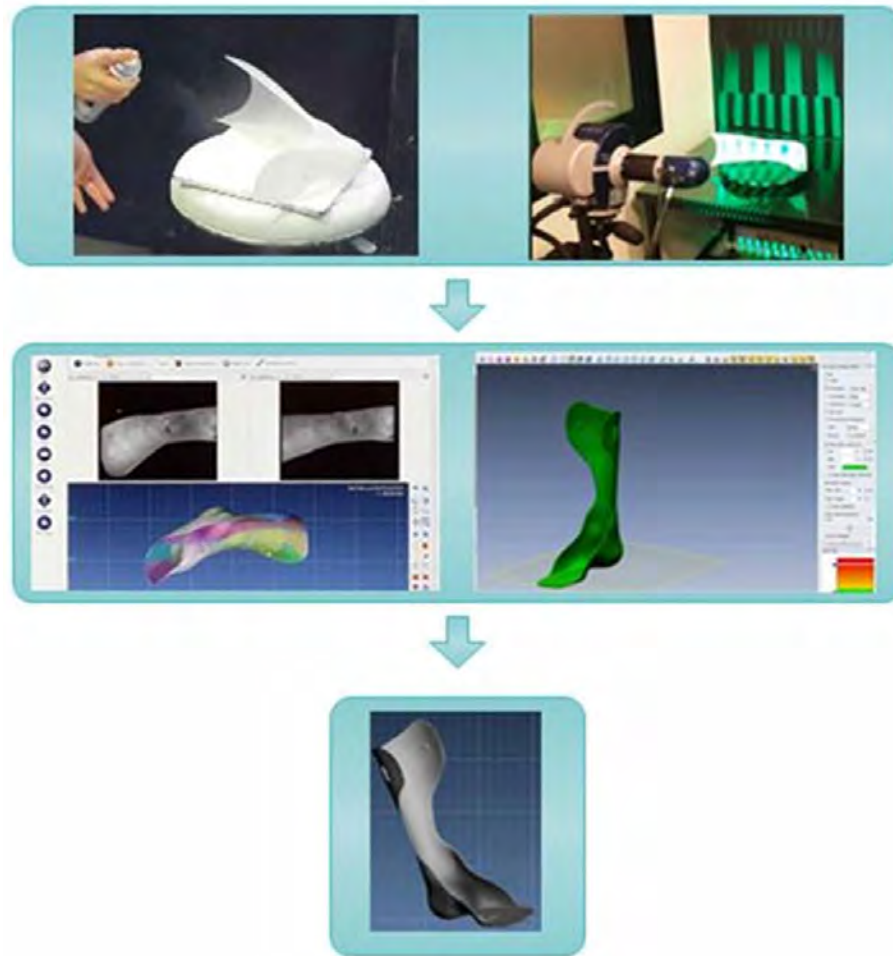
An optimum AFO design should be implemented according to the patient's specific conditions such as spastic movement, low-tone pronation, high-tone pronation or drop-foot and should provide the following properties.

An AFO should:

- Be ergonomically suitable to the patient's body,
- Minimize the skin and tissue injury as reducing or distributing the pressure around the limb,
- Prevent deformity, pain and contracture,
- Reduce consumption of energy during walking,
- Be light, durable and cosmetic,
- Be resistant against environmental effects,
- Be at reasonable cost and be produced in a proper time, and
- Be easy to use.

Hinged AFOs, which are of an important role among AFO designs, consist of a joint mechanism, and a calf and a foot component that are usually manufactured by molding technique (May & Lock-

Figure 4. Steps of 3D optical scanning



ard, 2011). The joint mechanisms that provide the control of plantarflexion or dorsiflexion are made of various materials such as metal, composite or plastic. There are also different types of joints for such functions as assisting, restricting or stopping. It is also possible to adjust the stiffness of AFO by means of some kinds of joint mechanisms (Kobayashi, Leung, Akazawa, & Hutchins, 2011). However, the stress distributes a wider area in the type of one-piece AFOs, while it is mainly concentrated around the joint in hinged type AFOs. Unless the design parameters are determined properly, then the joints, particularly metal ones, may be broken in short time. Many kinds of hinged

AFO and joints have been designed for better solutions to improve the gait function efficiently so far (Carlson, 2004; Engelman, 2010; Kobayashi et al., 2011; Kramer & Hinshon, 2016; Schwartz, 2014; Wiggin, Sawicki, & Collins, 2012), and it is still a popular research area. On the other hand, there are some studies that investigate the effectiveness of hinged AFOs for different abnormal gait conditions (Kim, Eng, & Whittaker, 2004; Rha, Kim, & Park, 2010; Tyson & Thornton, 2001). For example, it is considered that if patients have a preexisting tendency to crouch, hinged AFO may not be a good solution (Hsu et al., 2008).

MANUFACTURE OF AFO

There are many material types used for AFO manufacture such as metal, plastic, synthetic fabrics or composites. However, because it is light, cosmetic and providing support and better contact with the body, plastics are commonly prescribed. Nevertheless, they have some disadvantages such as being non-adjustable. New designs have been proposed both to provide comfort to the patient and to improve the efficiency of available designs. New designs lead to development of new manufacturing methods. The most conventional manufacturing technique is the molding process in which the lower part of the leg is casted by producing a positive cast to represent patients' shank, ankle and foot. Moreover, different manufacturing methods have been improved due to the requirements of different design and material types (Morris, 2006).

Vacuum Molding Technique

In this technique (Figure 5), a thermoplastic sheet is heated to its softening temperature in an oven. The heating time depends on the oven temperature, material thickness, material type and oven efficiency. Then the heated thermoplastic sheet is forced against the contours of a mold by an orthotist. For the purpose of better formed orthosis, vacuum pressure is also applied to the material. The molds used in this process are called positive molds. Positive molds are produced from negative molds.

Negative Mold Production

The limb is isolated with the aid of a foil. Then a rope is placed in front of the limb in order to assist to cut and remove the negative mold. The limb is wrapped with a plaster bandage applying a little impression. The ankle is held at 90 degrees of flexion in a neutral position. Once the plaster mold has been hardened, it is cut over the rope

with a cutting knife and removed from the limb. Then, the negative mold is put to the drying oven.

Positive Mold Production

The negative plaster mold is isolated and then a metal stick is placed into it. It is filled with liquid plaster and left to harden. Negative mold is removed from the positive mold. After cleaning process, positive mold is modified by scraping according to anthropometric measures of the patient, and finally mold is smoothed.

AFO Producing

A thermoplastic sheet of which thickness depends on material type, patient's weight, type and level of dysfunctions is heated in an oven. When the material has reached its softening temperature, it is wrapped around the positive mold by an orthotist. Pads and hinges should be placed to the related locations before the wrapping process. The thermoplastic sheet is pulled from their tips over the mold surface and edges of the sheet are combined by applying a moderate pressure. After the material cools, its redundant parts are cut by a scissors. Depending on prospectus there can be several design options. Once a design has been determined, trimline is marked on the orthosis and then the material is cut from the trimline. Then, locations of the straps are determined on the orthosis and they are placed. The goal of the straps is to keep the limb within orthosis in total contact. Therefore, straps should be placed at the correct angles and positions.

Carbon fiber AFOs are rather durable and light devices. They are manufactured in standard sizes for patients, but also it is possible to manufacture custom size carbon fiber AFO by using a positive mold of the patient's limb. Carbon fiber fabric layers are laid onto the modified positive mold with resin. The number and directions of layers should be arranged depending on the stiffness of the AFO.

Figure 5. Steps of vacuum molding technique: Limb is isolated with the aid of a foil and wrapped with a plaster bandage. Once the plaster mold has been hardened, it is removed from the limb and then the negative mold is obtained. The negative mold is isolated and filled with liquid plaster and left to harden. By doing so, positive mold is obtained. After cleaning process, positive mold is modified according to anthropometric measures of the patient. A thermoplastic sheet is heated in an oven. Then, it is wrapped around the positive mold. Pads and hinges are placed to the related locations before the wrapping process. The thermoplastic sheet is pulled from their tips over the mold surface and edges of the sheet are combined by applying a moderate pressure. After the material cools, its redundant parts are cut by a scissors. Trimline is marked on the orthosis and then the material is cut from the trimline. Then, the straps are placed on the orthosis

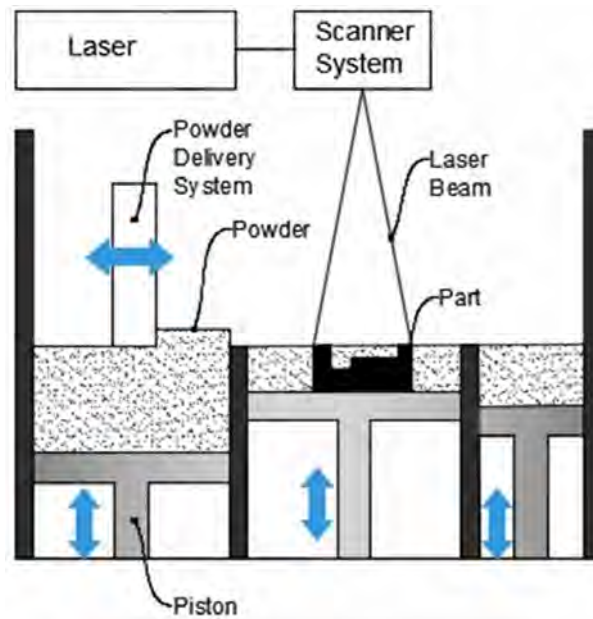


Additive Manufacturing (AM)

Additive manufacturing (AM), also known as 3D printing, is a rapid manufacturing technology that turn 3D CAD data into solid parts by using a number of additive processes, such as selective

laser sintering (SLS), direct metal laser sintering (DMLS), selective laser melting (SLM), fused deposition modeling (FDM), stereolithography (SLA) and laminated object manufacturing (LOM). Additive manufacturing is of several advantages, such as enabling freeform design,

Figure 6. Selective laser sintering (SLS)



high dimensional accuracy, reducing raw material usage, and the time to market. However, there are also several disadvantages, such as low-volume production, limited materials, high build time and surface quality depending on layer thickness (Gao et al., 2015). AM is utilized in many fields such as automotive, aerospace and biomedical industries, and can be an economical solution for customized products and creating prototype of models.

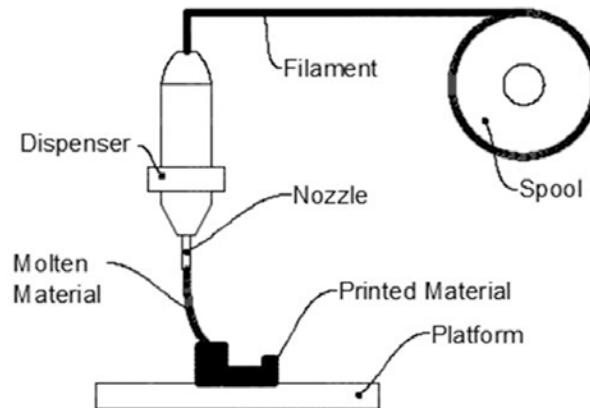
In additive manufacturing, the first step of the process is obtaining a 3D solid model. This can be accomplished by CAD software or 3D scanning systems. Then the model exported to 3D printer in stereolithography (STL) format (Ahn, Montero, Odell, Roundy, & Wright, 2002). Then the manufacturing process varying according to AM technique of the 3D printer machine is started.

SLS is one of the popular AM technique using laser (Figure 6). Parts are manufactured by sintering the powdered material. The material is heated and fused, after the laser has completed a scan process. By doing so, a layer is produced. After all layers has been produced and fused, model is completed.

There are a number of studies explore the feasibility of using SLS in AFO manufacturing. Faustini et al. (2008) evaluated the feasibility of an SLS-based manufacturing framework to produce patient-specific PD-AFOs. They reported that systematic and controlled design modifications can be possible in the AFO shape or volume which permits the exploration of the relative advantages of various passive dynamic PD-AFO designs. Another advantage is that additional design features such as holes and attachments can be easily integrated into an AFO design. Schrank (2011) reported that dimensional accuracy of the SLS process was within tolerances and material cost was acceptable. When the time saving parameter was considered, the SLS surpassed the traditional PD-AFO fabrication methods.

FDM, another popular AM process, uses production-grade thermoplastics (Figure 7). A plastic filament is wound in a coil on a spool. As material is extruded from a nozzle, the spool rotates and supplies filament to the extrusion nozzle. The material is heated in a dispenser and melted before extrusion. It hardens in a short time

Figure 7. Fused deposition modeling (FDM)



after the extrusion. The nozzle is moved along the horizontal and vertical directions by a numerically controlled mechanism (Bralla & Press, 2007; Miller, Vandome, & McBrewster, 2010). While the material is extruded as a thin ribbon on the work surface, the CAD model can be constructed layer by layer in a computer controlled pattern.

Number of AFO manufacturing researches using FDM process has been increasing day by day. In this technique, products can be manufactured within the repeatable dimensional accuracy. Patar et al. (2012) developed a DAFO prototype using FDM. The layer thickness was determined as 0.04 mm in their study. In the AFO model, ABS was used as FDM material and a DC motor provided dynamic feature of the AFO. They reported that the prototype model was much higher than the estimated cost. However, it is important to note that the cost of this model also includes a mechanism with DC motor.

Jin et al. (2015) evaluated additive manufacturing techniques of custom-made orthoses and prostheses in their review study. They concluded that AM technology enabled the fabrication of custom-made AFOs with proper fit and adequate strength, while it is not satisfied enough from the clinical, technological and economical point of views. Nevertheless, it is important to note that AM does not require molding process, which takes a considerable time, and an experienced orthoptist trimming the AFO.

There are also some researches on the process ability of materials for AM machines (Fiedler, Correa, Radusch, Wutzler, & Gerken, 2007). Additionally, some companies have been releasing new materials for AM machines. It seems that AM can be a good option for the AFO manufacturing with advancing of AM technology. The design concept for AM can be improved for more efficient products.

SELECTION OF AFOS

Developments in design and manufacture of orthoses affect the selection of the AFO. The proper selection of an AFO is crucial to support the patient during daily life activities. In this respect, biomechanical examination should be carefully implemented to ensure thorough assessment of gait functions and to gain feedback of patients' experiences properly. Generally orthoses are prescribed in order to improve patients activity of daily living (ADL) (World Health Organization, 2001). Therefore, it should be well analyzed that what the patients' gait abnormalities are, as well as the other ADL activities such as sitting, standing and even running. In order to understand the abnormalities of these different ADL activities, some functional tests or questioners should be applied to the patients or parents such as Functional Independence Measure for Children (Ziviani et

al., 2002), Gross Motor Function Measurement (Russell et al., 2000), Client Satisfaction (Bravini et al., 2014), Functional Reach Test (Bravini et al., 2014). Additionally, some sophisticated laboratories including high speed cameras, force plates and electromyography sensors are used to objectively define the movement abnormalities during standing, sitting, stair ascending-descending, reaching, walking and running. After the definition of the dynamic problem of the activity, orthotic team, which is consisted of physiotherapist, doctor and orthotic technician, get to gather in order to decide the most appropriate orthotic design for the patient. After making the decision, it should be prescribed and manufactured in a short time. During the first try, the team should be with the patients and notes the fine tunings of the AFO for the best usage in ADL. The same functional tests should be re-performed after couple of weeks in order to define the functional changes that the AFO made in patient's life.

FUTURE RESEARCH DIRECTIONS

Orthotic designs focus on the improvement of the function and stability of patients during ADL. Therefore, different movement types require different supports or facilitations. Orthoses need to be lighter, more durable, more skin friendly and smarter. If smart designs and smart materials gather with better understanding of the patients' need, the orthotics may find an important role in their life. Also it is expected that with the improvement of the additive manufacturing techniques, more subject-specific AFOs would be fabricated, thereby leading to eliminate the most of the problems stemmed from the standard and common approaches on patients with different demands and diseases.

CONCLUSION

This chapter is intended to be useful for the orthotists, physiotherapists and engineers as well as the other health specialists who are eager to know about the usage of AFOs, their different designs, and different ways of their manufacture. Finally, we emphasized the importance of the evaluation of the patients and clearly understanding of their functional difficulties by using some well-known tests and questionnaires.

REFERENCES

- Ahn, S.-H., Montero, M., Odell, D., Roundy, S., & Wright, P. K. (2002). Anisotropic material properties of fused deposition modeling ABS. *Rapid Prototyping Journal*, 8(4), 248–257. doi:10.1108/13552540210441166
- Balasundaram, A., Gurun, D., Neely, A., Ash-Rafzadeh, A., & Ravichandra, J. (2014). Novel CBCT and optical scanner-based implant treatment planning using a stereolithographic surgical guide: A multipronged diagnostic approach. *Implant Dentistry*, 23(4), 401–406. doi:10.1097/ID.0000000000000128 PMID:25051414
- Belda-Lois, J.-M., Barberà, R., Gómez, J., Baydal, J., Navarro, J., Gimeno, S., & Vera, P. (2006). Bio-mechanical requirements in upper-limb orthoses for tremor suppression. *Journal of Biomechanics*, 39, S76. doi:10.1016/S0021-9290(06)83192-1
- Bella, G. P., Rodrigues, N. B., Valenciano, P. J., Silva, L. M., & Souza, R. C. (2012). Correlation among the visual gait assessment scale, Edinburgh visual gait scale and observational gait scale in children with spastic diplegic cerebral palsy. *Brazilian Journal of Physical Therapy*, 16(2), 134–140. doi:10.1590/S1413-35552012000200009 PMID:22584771

- Bielby, S. A., Warrick, T. J., Benson, D., Brooks, R. E., Skewes, E., Alvarez, E., & DesJardins, J. D. (2010). Trimline severity significantly affects rotational stiffness of ankle-foot orthosis. *Journal of Prosthetics and Orthotics*, 22(4), 204–210. doi:10.1097/JPO.0b013e3181f9082e
- Bralla, J., & Press, I. (2007). *Handbook of Manufacturing Processes*. Industrial Press.
- Bravini, E., Franchignoni, F., Ferriero, G., Gior-dano, A., Bakhsh, H., Sartorio, F., & Vercelli, S. (2014). Validation of the Italian version of the Client Satisfaction with Device module of the Orthotics and Prosthetics Users Survey. *Disability and Health Journal*, 7(4), 442–447. doi:10.1016/j.dhjo.2014.04.002 PMID:25224984
- Bregman, D. J., De Groot, V., Van Diggele, P., Meulman, H., Houdijk, H., & Harlaar, J. (2010). Polypropylene ankle foot orthoses to overcome drop-foot gait in central neurological patients: A mechanical and functional evaluation. *Prosthetics and Orthotics International*, 34(3), 293–304. doi:10.3109/03093646.2010.495969 PMID:20738233
- Carlson, J. M. (2004). *Adjustable mounting housing for orthotic ankle flexure joint*. Patent No: US 20040002672 A1.
- Dalvand, H., Dehghan, L., Feizi, A., Hosseini, S. A., & Amirsalari, S. (2013). The impacts of hinged and solid ankle-foot orthoses on standing and walking in children with spastic diplegia. *Iranian Journal of Child Neurology*, 7(4), 12–19. PMID:24665312
- Dhatt, G., Lefrançois, E., & Touzot, G. (2012). *Finite element method*. John Wiley & Sons. doi:10.1002/9781118569764
- Duffy, C. (1997). AFOs, energy consumption and gait in spina bifida. *Gait & Posture*, 6(3), 278. doi:10.1016/S0966-6362(97)90089-8
- Eddison, N., & Chockalingam, N. (2013). The effect of tuning ankle foot orthoses–footwear combination on the gait parameters of children with cerebral palsy. *Prosthetics and Orthotics International*, 37(2), 95–107. doi:10.1177/0309364612450706 PMID:22833518
- Edelstein, J. E., & Bruckner, J. (2002). *Orthotics: a comprehensive clinical approach*. Slack Thorofare.
- Engelman, I. K. (2010). *Articulated orthosis providing lift support*. Patent No: US 7682322 B2.
- Esposito, E. R., Blanck, R. V., Harper, N. G., Hsu, J. R., & Wilken, J. M. (2014). How does ankle-foot orthosis stiffness affect gait in patients with lower limb salvage? *Clinical Orthopaedics and Related Research*, 472(10), 3026–3035. doi:10.1007/s11999-014-3661-3 PMID:24817379
- Everaert, D. G., Stein, R. B., Abrams, G. M., Dromerick, A. W., Francisco, G. E., Hafner, B. J., & Kufta, C. V. (2013). Effect of a Foot-Drop Stimulator and Ankle–Foot Orthosis on Walking Performance After Stroke A Multicenter Randomized Controlled Trial. *Neurorehabilitation and Neural Repair*, 27(7), 579–591. doi:10.1177/1545968313481278 PMID:23558080
- Faustini, M. C., Neptune, R. R., Crawford, R. H., & Stanhope, S. J. (2008). Manufacture of passive dynamic ankle–foot orthoses using selective laser sintering. *Biomedical Engineering. IEEE Transactions on*, 55(2), 784–790.
- Fiedler, L., Correa, L. O. G., Radusch, H. J., Wutzler, A., & Gerken, J. (2007). Evaluation of Polypropylene Powder Grades in Consideration of the Laser Sintering Processability. *Journal of Plastic Technology*, 3(4).
- Franceschini, M., Massucci, M., Ferrari, L., Agosti, M., Orsi, M., & Paroli, C. (2001). Gait with custom-made orthosis in hemiplegic patients: Preliminary data. *Gait & Posture*, 13(2), 139–140.

- Gao, W., Zhang, Y., Ramanujan, D., Ramani, K., Chen, Y., Williams, C. B., & Zavattieri, P. D. (2015). The status, challenges, and future of additive manufacturing in engineering. *Computer Aided Design*, 69, 65–89. doi:10.1016/j.cad.2015.04.001
- Guo, Y., Hao, Z., & Wan, C. (2016). Tribological characteristics of polyvinylpyrrolidone (PVP) as a lubrication additive for artificial knee joint. *Tribology International*, 93, 214–219. doi:10.1016/j.triboint.2015.08.043
- Heys, K. R., & Truscott, R. J. (2008). The stiffness of human cataract lenses is a function of both age and the type of cataract. *Experimental Eye Research*, 86(4), 701–703. doi:10.1016/j.exer.2007.12.009 PMID:18289531
- Hofmann, U. K., Thumm, S., Jordan, M., Walter, C., Rondak, I.-C., & Ipach, I. (2016). The Effects of Hip and Spine Orthoses on Braking Parameters: A Simulated Study With Healthy Subjects. *PM & R*, 8(1), 35–44. doi:10.1016/j.pmrj.2015.06.002 PMID:26079866
- Hsu, J. D., Michael, J., & Fisk, J. (2008). *AAOS atlas of orthoses and assistive devices*. Elsevier Health Sciences.
- Jin, Y., Plott, J., Chen, R., Wensman, J., & Shih, A. (2015). Additive Manufacturing of Custom Orthoses and Prostheses—A Review. *Procedia CIRP*, 36, 199–204. doi:10.1016/j.procir.2015.02.125
- Kang, S.-H., Kim, Y.-H., & Kim, M.-K. (2016). Comparison of digital dental images yielded by digital dental casts, cone-beam computed tomography, and multislice computed tomography for measurement of dental area. *Oral Radiology*, 1–9.
- Kim, C. M., Eng, J. J., & Whittaker, M. W. (2004). Effects of a simple functional electric system and/or a hinged ankle-foot orthosis on walking in persons with incomplete spinal cord injury. *Archives of Physical Medicine and Rehabilitation*, 85(10), 1718–1723. doi:10.1016/j.apmr.2004.02.015 PMID:15468037
- King, W., Mnatsakanian, A., & Kissel, J. (2015). “Toe Off” Style Carbon Fiber Ankle Foot Orthoses in Neuromuscular Disease: The OSUMC Experience (P7. 063). *Neurology*, 84(14S).
- Kobayashi, T., Leung, A. K., Akazawa, Y., & Hutchins, S. W. (2011). Design of a stiffness-adjustable ankle-foot orthosis and its effect on ankle joint kinematics in patients with stroke. *Gait & Posture*, 33(4), 721–723. doi:10.1016/j.gaitpost.2011.02.005 PMID:21376602
- Kramer, T., & Hinshon, P. S. (2016). *Ankle Foot Orthotic Joint*. US Patent 20,160,030,223.
- Lake, E., Ahmad, S., & Dobrashian, R. (2013). The sonographic appearances of breast implant rupture. *Clinical Radiology*, 68(8), 851–858. doi:10.1016/j.crad.2013.03.014 PMID:23623260
- Lam, W. K., Leong, J. C., Li, Y. H., Hu, Y., & Lu, W. W. (2005). Biomechanical and electromyographic evaluation of ankle foot orthosis and dynamic ankle foot orthosis in spastic cerebral palsy. *Gait & Posture*, 22(3), 189–197. doi:10.1016/j.gaitpost.2004.09.011 PMID:16214658
- Leardini, A., Aquila, A., Caravaggi, P., Ferraresi, C., & Giannini, S. (2014). Multi-segment foot mobility in a hinged ankle-foot orthosis: The effect of rotation axis position. *Gait & Posture*, 40(1), 274–277. doi:10.1016/j.gaitpost.2014.03.188 PMID:24792637
- Leone, D. J. (1987). A structural model for molded thermoplastic ankle-foot orthoses. *Journal of Biomechanical Engineering*, 109(4), 305–310. doi:10.1115/1.3138685 PMID:3695430
- May, B. J., & Lockard, M. A. (2011). *Prosthetics & orthotics in clinical practice: a case study approach*. FA Davis.
- Mikkelsen, L. P., Skorini, R. Í., & Løgstrup Andersen, T. (2011). Biomechanical study of a drop foot brace. In *Proceedings of the 2011 SIMULIA Customer Conference*.

- Miller, F. P., Vandome, A. F., & McBrewster, J. (2010). *Fused Deposition Modeling*. VDM Publishing.
- Moisan, G., & Cantin, V. (2016). Effects of two types of foot orthoses on lower limb muscle activity before and after a one-month period of wear. *Gait & Posture*, 46, 75–80. doi:10.1016/j.gaitpost.2016.02.014 PMID:27131181
- Morris, C. (2007). Lower-limb orthoses. In C. Morris & L. Dias (Eds.), *Paediatric Orthotics: Orthotic management of children* (pp. 44–66). Chicago: Mac Keith Press.
- Morris, C. (2007). Materials, Components and Fabrication. In C. Morris & L. Dias (Eds.), *Paediatric Orthotics: Orthotic management of children* (pp. 28–43). Chicago: Mac Keith Press.
- Nolan, K. J., Savalia, K. K., Lequerica, A. H., & Elovic, E. P. (2009). Objective assessment of functional ambulation in adults with hemiplegia using ankle foot orthotics after stroke. *PM & R*, 1(6), 524–529. doi:10.1016/j.pmrj.2009.04.011 PMID:19627941
- Patar, A., Jamlus, N., Makhtar, K., Mahmud, J., & Komeda, T. (2012). Development of dynamic ankle foot orthosis for therapeutic application. *Procedia Engineering*, 41, 1432–1440. doi:10.1016/j.proeng.2012.07.332
- Radtka, S. A., Skinner, S. R., Dixon, D. M., & Johanson, M. E. (1997). A comparison of gait with solid, dynamic, and no ankle-foot orthoses in children with spastic cerebral palsy. *Physical Therapy*, 77(4), 395–409. PMID:9105342
- Radtka, S. A., Skinner, S. R., & Johanson, M. E. (2005). A comparison of gait with solid and hinged ankle-foot orthoses in children with spastic diplegic cerebral palsy. *Gait & Posture*, 21(3), 303–310. doi:10.1016/j.gaitpost.2004.03.004 PMID:15760746
- Ramsey, J. A. (2011). Development of a method for fabricating polypropylene non-articulated dorsiflexion assist ankle foot orthoses with predetermined stiffness. *Prosthetics and Orthotics International*, 35(1), 54–69. doi:10.1177/0309364610394477 PMID:21515890
- Rethlefsen, S., Kay, R., Dennis, S., Forstein, M., & Tolo, V. (1999). The effects of fixed and articulated ankle-foot orthoses on gait patterns in subjects with cerebral palsy. *Journal of Pediatric Orthopedics*, 19(4), 470–474. doi:10.1097/01241398-199907000-00009 PMID:10412995
- Rha, D. W., Kim, D. J., & Park, E. S. (2010). Effect of hinged ankle-foot orthoses on standing balance control in children with bilateral spastic cerebral palsy. *Yonsei Medical Journal*, 51(5), 746–752. doi:10.3349/ymj.2010.51.5.746 PMID:20635451
- Ridgewell, E., Rodda, J., Graham, H., & Sangeux, M. (2015). The effect of bilateral solid AFO on gait. *Gait & Posture*, 42, S70–S71. doi:10.1016/j.gaitpost.2015.03.122
- Ries, A., & Schwartz, M. (2015). Comparative effectiveness of ground reaction and solid ankle foot orthoses for crouch gait in children with cerebral palsy. *Gait & Posture*, 42, S90. doi:10.1016/j.gaitpost.2015.06.164
- Romkes, J., Hell, A., & Brunner, R. (2006). Changes in muscle activity in children with hemiplegic cerebral palsy while walking with and without ankle-foot orthoses. *Gait & Posture*, 24(4), 467–474. doi:10.1016/j.gaitpost.2005.12.001 PMID:16413188
- Russell, D. J., Avery, L. M., Rosenbaum, P. L., Raina, P. S., Walter, S. D., & Palisano, R. J. (2000). Improved scaling of the gross motor function measure for children with cerebral palsy: Evidence of reliability and validity. *Physical Therapy*, 80(9), 873–885. PMID:10960935

- Schrank, E. S., & Stanhope, S. J. (2011). Dimensional accuracy of ankle-foot orthoses constructed by rapid customization and manufacturing framework. *Journal of Rehabilitation Research and Development*, 48(1), 31. doi:10.1682/JRRD.2009.12.0195 PMID:21328161
- Schwartz, N. (2014). *Ankle-Foot Orthosis*. Patent No: US 8904674.
- Sherief, A. E. A. A., Gazya, A. A. A., & El Gafaar, M. A. A. (2015). Integrated effect of treadmill training combined with dynamic ankle foot orthosis on balance in children with hemiplegic cerebral palsy. *The Egyptian Journal of Medical Human Genetics*, 16(2), 173–179. doi:10.1016/j.ejmhg.2014.11.002
- Sherk, K. A. (2008). Technical Note: The Development and Use of a Floating T-Strap on a Double Upright Metal AFO to Correct Coronal-Plane Pathologies and Reduce Skin Shear. *JPO: Journal of Prosthetics and Orthotics*, 20(1), 24–26.
- Tyson, S., & Thornton, H. (2001). The effect of a hinged ankle foot orthosis on hemiplegic gait: Objective measures and users opinions. *Clinical Rehabilitation*, 15(1), 53–58. doi:10.1191/026921501673858908 PMID:11237162
- Van Beeten, B., Hartman, A., & Houdijk, H. (2015). Optimizing knee kinematics in mid-stance by tuning the ankle foot orthoses-footwear combination of children with cerebral palsy: A case series. *Gait & Posture*, 42, S88–S89. doi:10.1016/j.gaitpost.2015.06.162
- Van Gestel, L., Molenaers, G., Huenaerts, C., Seyler, J., & Desloovere, K. (2008). Effect of dynamic orthoses on gait: A retrospective control study in children with hemiplegia. *Developmental Medicine and Child Neurology*, 50(1), 63–67. doi:10.1111/j.1469-8749.2007.02014.x PMID:18173633
- White, H., Jenkins, J., Neace, W. P., Tylkowski, C., & Walker, J. (2002). Clinically prescribed orthoses demonstrate an increase in velocity of gait in children with cerebral palsy: A retrospective study. *Developmental Medicine and Child Neurology*, 44(4), 227–232. doi:10.1017/S0012162201001992 PMID:11995890
- Wiggin, M. B., Sawicki, G. S., & Collins, S. H. (2013). *Apparatus and clutch for using controlled storage and release of mechanical energy to aid locomotion*. Patent No: US20130046218 A1.
- World Health Organization. (2001). *International classification of functioning, disability and health: ICF*. World Health Organization.
- Ziviani, J., Ottenbacher, K. J., Shephard, K., Foreman, S., Astbury, W., & Ireland, P. (2002). Concurrent validity of the Functional Independence Measure for Children and the Pediatric Evaluation of Disabilities Inventory in children with developmental disabilities and acquired brain injuries. *Physical & Occupational Therapy in Pediatrics*, 21(2-3), 91–101. doi:10.1080/J006v21n02_08 PMID:12029858

ADDITIONAL READING

- Kobayashi, T., Leung, A. K. L., & Hutchins, S. W. (2011). Techniques to measure rigidity of ankle-foot orthosis: A review. *Journal of Rehabilitation Research and Development*, 48(5), 565–576. doi:10.1682/JRRD.2010.10.0193 PMID:21674406
- Tyson, S. F., Sadeghi-Demneh, E., & Nester, C. J. (2013). A systematic review and meta-analysis of the effect of an ankle-foot orthosis on gait biomechanics after stroke. *Clinical Rehabilitation*, 27(10), 879–891. doi:10.1177/0269215513486497 PMID:23798747

KEY TERMS AND DEFINITIONS

AFO: Ankle-foot-orthosis is an assistive device designed to improve abnormal lower limb motor functions.

AM: Additive manufacturing is the rapid manufacturing technology that converts 3D CAD data to physical model by such methods as selective laser sintering, direct metal laser sintering, selective laser melting, fused deposition modeling, stereolithography, laminated object manufacturing and etc.

Gait: A manner of walking that shows the sequence of foot movements and also other parts of the body.

Orthosis: Externally applied bio-mechanical device to the body parts to control their motions and also provides protection and support.

Stiffness: An important AFO design parameter that is the resistance of the AFO to deformation by an applied force.

Trimline: Border of the trimmed section of the orthosis that is an important parameter in determining of AFO stiffness.

Vacuum Molding: A common AFO manufacturing technique in that heated sheet of plastic is laid over a positive lower limb mold and formed by the help of vacuum.

A Disability–Aware Mentality to Information Systems Design and Development

Julius T. Nganji

University of Ottawa, Canada

INTRODUCTION

Nowadays, it is difficult to imagine a world without technology. We interact with technology every day in various ways. This inevitable modern technological revolution has meant that many services are increasingly being offered online. Amongst these services e-learning, e-commerce and e-health are most common. With the drive to use technology for delivering services online, a lot of the technological developments around these areas by default have often focused on meeting the needs of people without disabilities, thus leaving people with disabilities to seek appropriate assistive technologies in order to interact with such systems. In some cases, such assistive technologies might not be compatible with these systems. There are legislations in various countries around the world necessitating that people with disabilities be included in services that are offered, by ensuring that these services are accessible to them; otherwise “reasonable adjustments” or “reasonable accommodation” need to be made in order to meet their needs.

A lot of the difficulties people with disabilities face when interacting with most information systems are related to the lack of consideration of the needs of people with disabilities during the development cycle. In designing technological solutions, designers and developers need to understand that disability could affect different functions related to the senses and how this happens. Such designers need to develop a new mind set when it comes to designing systems that will be used by everyone. The difficulties and failures of

existing information systems towards people with disabilities have necessitated the search for a better approach for designing and developing information systems. Thus, this article aims to propose a disability-aware approach to information systems design and development in order to ensure that adequate analysis of the needs of potential users with disabilities is carried out and that their needs are incorporated into the design. It also ensures that a representative group of people are selected and involved in the design process. That way, useful feedback is obtained and cost of production is reduced as designers do not have to spend a lot of money retrofitting accessibility. By developing a disability-aware mentality to systems design, the result is an accessible and usable product.

In the following sections, the difficulties that people with disabilities face when using technology will be discussed through a review of literature. The disability-aware approach to information systems design will then be proposed. The recommendations from students with disabilities on how to design e-learning to meet their needs will then be presented after which future research will be discussed, before a conclusion of this article.

BACKGROUND

In order to set the scene for proposing a disability-aware mentality to information systems design and development, this section reviews literature relating to the difficulties that people with disabilities face while interacting with information systems, some assistive technologies for interacting with

DOI: 10.4018/978-1-5225-2255-3.ch028

information systems and also reviews some approaches used in designing information systems.

Difficulties Encountered by People With Disabilities When Using Information Systems

It is a fact that people with disabilities are generally more disadvantaged than those without disabilities when accessing services for various reasons which could include the way the environment has been designed to accommodate their needs or how society responds to their needs in various ways. Environmental factors such as the weather, specifically during winter where there have been heavy snowfalls for instance have been barriers to people with wheelchairs accessing community services (Ripat, Brown, & Ethans, 2015). The way society is designed tends to favor people without disabilities. Thus, it is common even nowadays to see newer buildings being designed without accessibility in mind (e.g. no ramps or elevators for wheelchair access). Nevertheless, there is an increase in awareness of the needs of people with disabilities when constructing buildings, perhaps because it is mandated by disability legislations. Also, many older buildings are being adapted to meet the needs of people with disabilities, such as adding ramps for wheelchair access, including induction loops in meeting halls, controlling the lighting of buildings to suit people with light sensitivity and many other adaptations. Steyaert (2005) observed that such accessibility principles are not necessarily being translated from the physical environment into the electronic environment, leading to numerous difficulties for learners with disabilities, when considering an educational environment.

The severity of disabilities such as visual impairment may range from low vision to complete blindness. In the latter case, medical interventions such as eyeglasses, contact lenses or even surgery might not result in a permanent solution. People with this severity level will have to employ assistive technologies such as screen readers in order to

read information that is found in electronic format or through the use of digital Braille amongst other interaction methods. In education, students with visual impairments are often more disadvantaged than students with other disabilities when using technology (Kelly & Smith, 2008). There is no wonder therefore why a lot of the guidelines for designing inclusive solutions tend to use this group of people as example when discussing accessibility. For people with visual impairments to use web-based information systems many accessibility considerations, including the need to include alternative texts in images (Nganji, Brayshaw, & Tompsett, 2013) need to be adhered to. As visual impairment affects the sense of sight, if information systems are not designed accessibly with the needs of such users in mind, the result will be significant difficulties trying to interact with such systems which in extreme cases could also lead to incompatibility with assistive technologies such as screen readers. For people with low vision who do not necessarily have to rely on screen readers to read text on an information system, there is great need to ensure that they can be able to manipulate the text through increasing its size for more visibility in addition to providing responsive design.

Dyslexia is the most common disability affecting learning, particularly reading. It affects word recognition and spelling in an individual. There are a number of assistive technologies that could help students with dyslexia. Tools such as Read&Write help with writing and spelling. The amount of information that is presented in an information system as well as how the information is presented could determine if an individual with dyslexia would understand the information. Smaller chunks of information that is clearly written could be helpful.

Individuals with hearing impairment may have partial or complete loss of hearing. Depending on the severity of the impairment, individuals may rely on hearing aid or sign language in order to understand information that is being conveyed. Thus, it is important for designers and developers

of information systems to take these into consideration during design.

Amongst people who will also need to access information systems are people with upper limb mobility difficulties. For these people, using the mouse or keyboard may be difficult. Thus, their needs should be taken into consideration during design.

Assistive Technologies for Interacting With Information Systems

To help overcome the difficulties encountered by people with disabilities when using information systems that are not disability-aware, a wide range of assistive technologies have been implemented. When compatible with information systems, these technologies can greatly facilitate access to information to those who would otherwise not be able to access such information due to disability. A few of such technologies will be discussed here.

For people with visual impairments ranging from low vision to complete blindness, screen magnifiers and screen readers are the main technologies. Some of the most common screen readers available are JAWS (Job Access with Speech) and NVDA (NonVisual Desktop Access) and have been used in research studying adoption of screen readers (McCarthy, Pal, & Cutrell, 2013). MAC operating systems have a screen reader known as VoiceOver which can be used to read out text. Screen readers generally work by converting text into audio format so that the visually impaired person can then listen to the information they could not otherwise read. A lot of mobile devices nowadays have screen reading software that enable people to interact with information, also incorporating speech-to-text software. Thus, users can search the web through voice input and obtain results, could dial someone through speech commands amidst many other functions.

There are also some assistive technologies for people with hearing difficulties, which work to improve hearing. These are generally known as assistive listening devices (ALDs). These do not

work for people who are completely deaf as they cannot hear any sound. According to Kim and Kim (2014), ALDs “usually use microphones to capture an audio source and broadcast it wirelessly over a frequency modulation (FM), infra-red, induction loop, or other transmission techniques.” For people with hearing impairment, videos need to contain captions in order to be accessible to them. When captions are included, users can understand the content of the video. Some readers may be confused over the difference between subtitles and captions. Subtitles are often used to translate the speech in a video into a different language (e.g. the video is in English but subtitles could be in French so that French viewers would understand the video) while captions are subtitles that help people with hearing impairment to understand the video. The captions could be turned on or off but not subtitles. To make videos accessible, transcripts of the video could also be included when there are no captions. The transcript is a text version of the words spoken in the video. Research has shown that users would interact more with subtitles than with transcripts (Grgurovic & Hegelheimer, 2007). Also available to people with hearing impairment are text telephones which enable communication by typing and then displaying the typed information for the individual on the other end to read.

For people with dyslexia, various assistive technologies such as word processors and spell checkers help with writing, voice recognition software (e.g. Dragon NaturallySpeaking, Kurzweil) convert speech to text and could be very useful. In a literature review on the use of assistive technologies by students in higher education, Pino and Mortari (2014) found that the spellchecker is useful in finding spelling mistakes but it also displays a list of possible words which, for a student with dyslexia could be difficult to decipher which one is the correct word. They also reported that voice recognition software was very useful to students. However, students did not find tape recording very useful to them due to the amount of time they will spend in listening to the tape.

Approaches and Methodologies for Designing Information Systems

There are several approaches to information systems design. This section will focus on traditional methods, often used in software engineering by discussing the waterfall and the spiral methods and will also look at approaches that are common to designing user interfaces such as user-centered design, participatory design and inclusive design.

The waterfall method (Royce, 1987) was inspired by engineering disciplines such as civil and mechanical engineering and is sequential; progressing through the phases of conception, initiation, analysis, design, construction, testing, implementation and maintenance. This method is weak in that the development progresses downwards and each phase must be completed and signed off for the next phase to begin. Thus, designers would have to go back to the phase where a problem has been identified and start again from there. In reality, information systems development does not proceed as such. Also, when following this methodology, designers do not emphasize on the needs of people with disabilities during the analysis phase, thus the resultant product is not accessible to them.

The spiral method (Boehm, 1988) is intended for large, expensive and complicated projects. It has four phases that are adopted from other methods such as the waterfall method or incremental method, which include: planning (determining objectives), risk analysis (identify and resolve risks), engineering (development and testing) and evaluation (planning the next iteration). This method involves high costs and is suitable for large projects. However, this is not suitable for smaller information systems design projects.

User-centered design (Helander, Landauer, & Prabhu, 1997; Nielsen, 1993; Norman, 1988) focuses on designing information systems that would meet the needs of the user. The user is supposed to be the main focus of the design. However, when this method has been used, the focus has often been on users without disabilities. Thus, the

resultant product is not accessible to people with disabilities. Participatory design (Kyng, 1991; Muller, 1996) on the other hand seeks to involve all stakeholders, including the intended users in order to design an information system that is usable to them.

Despite the intention of participatory and user-centered design to develop products that meet the needs of users, Gregor et al. (2005) asserted that these approaches often do not address how to meet the needs of people with disabilities and those of older users. Thus, inclusive design seeks to focus on the needs of people with disabilities. Newell et al. (2010) also asserted that such approaches do not meet the needs of people with disabilities as most of the products might not be usable to older people and people with disabilities.

A DISABILITY-AWARE APPROACH TO INFORMATION SYSTEMS DESIGN AND DEVELOPMENT

Before discussing the disability-aware approach to information systems design and development, it is fitting to first discuss some reasons for developing inclusive systems. There are numerous reasons but three will be discussed here. Firstly, the number of people with disabilities is on the rise. The World Health Organization (WHO) estimates that there are over a billion people with disabilities, constituting about 15% of world population (WHO, 2015). It is morally correct to include everyone in society and designing accessible and usable systems is one way to ensure the participation and contribution of everyone, regardless of their ability or disability. Secondly, the increasing number of people with disabilities also shows that they can contribute to the economy. Bowtell (2015) reports that research has dispelled the myth that people with disabilities are poor; and has on the contrary shown that people with disabilities have a sizable spending power on accessible services. For designers to be part of this, they need to design accessible and usable systems. Thirdly, it

A

is a legal requirement in most countries to offer accessible services, which include designing accessible information systems. Those who don't comply with legislation such as the Accessibility for Ontarians with Disabilities Act (AODA), the Equality Act 2010 in the United Kingdom and the Americans with Disabilities Act (ADA), just to name these, could be prosecuted.

As already discussed, the focus of most designers and developers of information systems is on meeting the needs of users without disabilities, thus the resultant products are not accessible or usable for people with disabilities. A more inclusive approach to designing information systems is the paradigm of user-sensitive inclusive design (Newell & Gregor, 2000). The authors assert that “inclusivity” is more achievable than “universal design” and “design for all”. Additionally, in this approach, designers develop an “empathetic” relationship with the end users.

The disability-aware approach in this article is a form of user-sensitive inclusive design approach that focuses on the needs of people with disabilities. It advocates that designers and developers should always think about the needs of the users, particularly those with disabilities when designing systems. A common issue with most systems is that even when they are designed to meet the needs of people with disabilities, they often neglect the needs of people with multiple disabilities (Nganji & Brayshaw, 2015). Designing accessible systems does not only benefit people with disabilities, but will be usable to people without disabilities since needs analysis includes everyone's needs.

A disability-aware approach can be applied to the design of all systems that are required to be used by humans, including in fields such as software engineering (Nganji & Nggada, 2011). The key components of the approach as depicted in Figure 1 include:

People With Disabilities

People with disabilities are the focus of this model since information systems need to be designed to

meet their needs. They need to participate in the needs analysis phase of the development cycle in order to ensure that the product meets their needs. Given that disability is an individual experience, their input will be very useful. They will need to participate by giving feedback to the designers or developers of information systems through expressing how the system could meet their needs as well as during the testing phase. However, they could be available at any time during development cycle to be consulted.

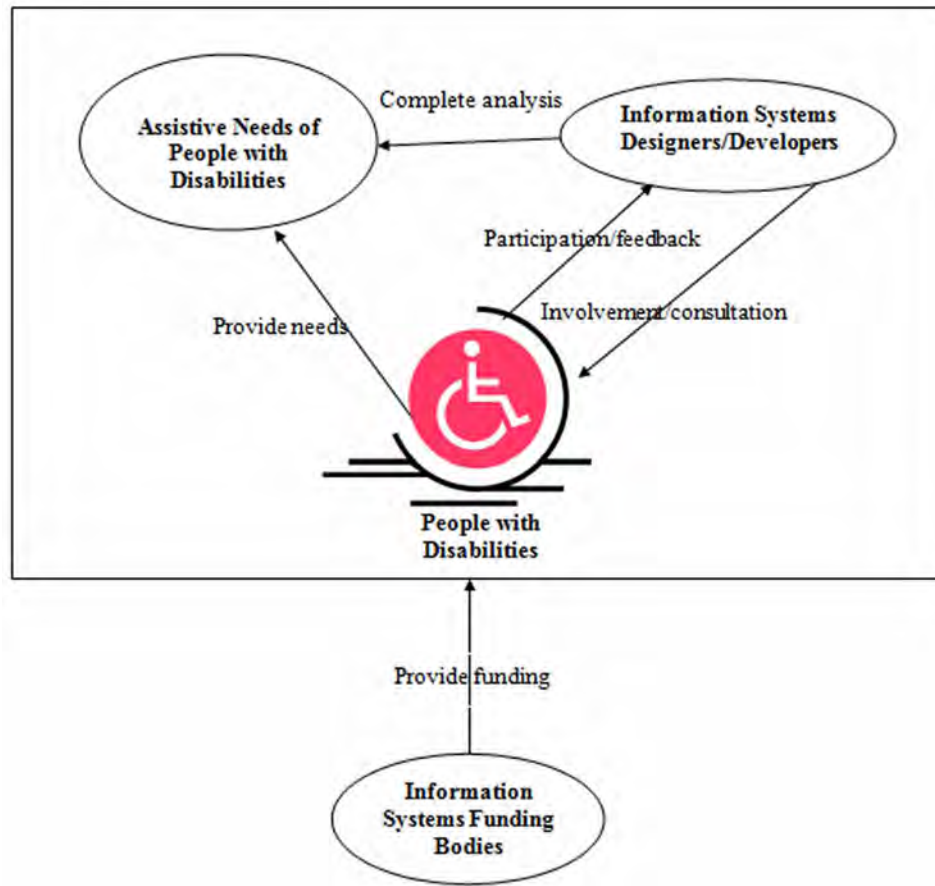
Assistive Needs of People With Disabilities

In designing an information system, the assistive needs of people with disabilities need to be taken into consideration. When this is done at the beginning of the development cycle, designers and developers of such systems could incorporate these needs into the design. When this is done correctly, it could eliminate the need for the user to acquire a standalone assistive technology to use with the system. This also helps to solve incompatibility issues that may arise with using some assistive technologies. In determining the assistive needs of the users, designers need to involve the users themselves but could in addition, consult experts in the domain as well as obtain input from existing literature. Thus, a complete analysis of such needs is necessary to ensure that the product is both accessible and usable.

Designers and Developers of Information Systems

The designers and developers drive the development of information systems. They are very important in the process and need to develop the mentality of working with users and constantly thinking about how to meet their needs. Also, developing an empathetic relationship with users as recommended by Newell and Gregor (2000) will ensure that the system is accessible and usable.

Figure 1. Disability-aware approach to information systems design and development



Information Systems Funding Bodies

Critical to the success of the information system project is the source of funding. Enough funds would be needed in order to involve end users in the project and to carry out usability and accessibility testing of the system. Thus, funding bodies could contribute financially to ensure that good quality systems which are bug-free are released.

IMPLEMENTING A DISABILITY-AWARE APPROACH: RECOMMENDATIONS FROM STUDENTS WITH DISABILITIES

In application to the field of education, 30 students with disabilities at a United Kingdom university

and 18 students with disabilities at a Canadian university were consulted on their opinions to designing e-learning systems to meet their needs. This was done during the evaluation of a piece of software and they had some useful recommendations, which could also be applied to the design of other information systems. During the evaluation, the author also observed their interactions and preferences which are also included in the following recommendations, some of which are presented in Figure 2.

Grant Users Control Over Font Sizes and Background Color

Some students with disabilities, particularly those who had low vision and those with dyslexia found it frustrating when they could not manipulate the

size of fonts that were used in designing e-learning systems. Students with low vision have limited vision, so, in the absence of a screen magnifier to increase the font size, they would need the ability to increase the font size in order to read the information that is presented. This needs to be done without distorting the presentation. Thus, designers need to consider this. Such considerations should also be made when designing for mobile devices.

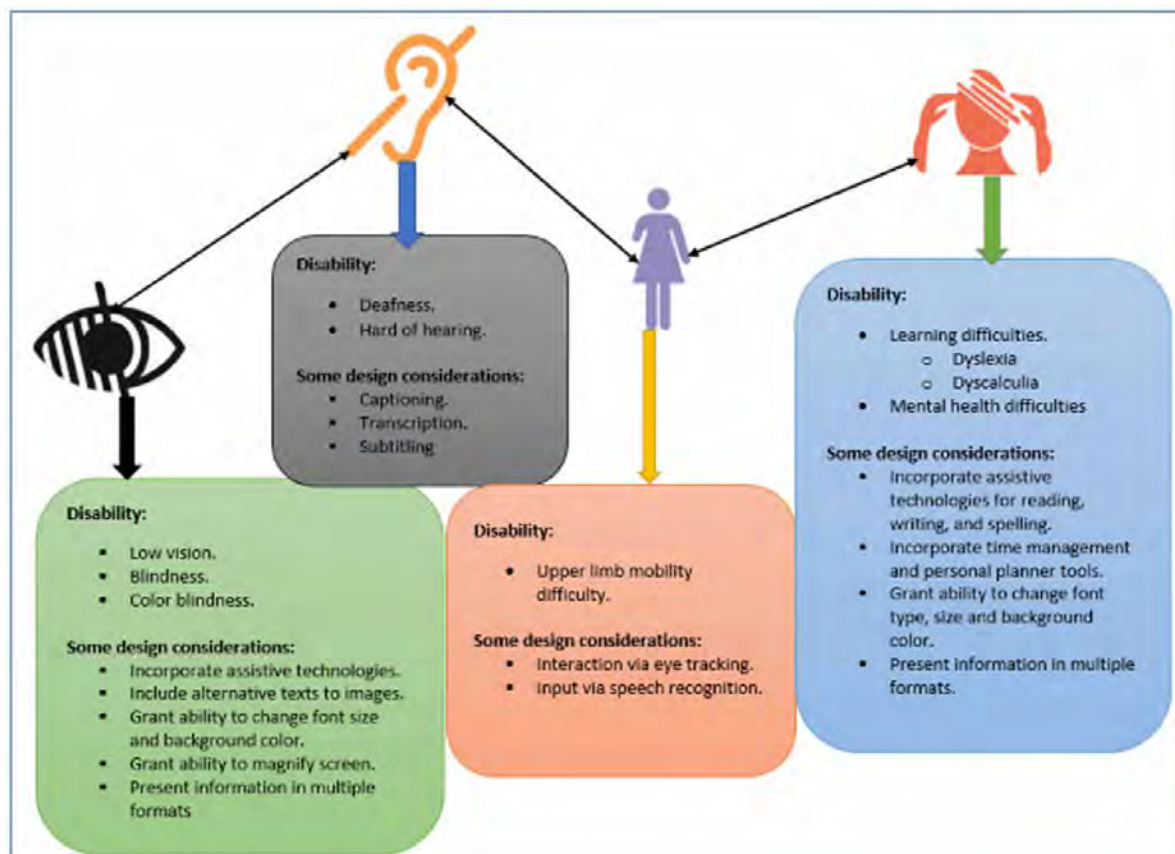
Students with dyslexia found the ability to change the type of font as well as its size in combination to the background color very useful. Thus, designers should consider granting users choice over the background color. This color contrast, although it may not be very useful to some people, are very important to others. On meeting their needs in an e-learning system, a student with both dyslexia and hearing impairment had this

recommendation: “the way that learning environments could adapt to my learning disabilities and hearing difficulties would be if I can change the text and the background color so that the text is much clearer for me to read.”

Embed Assistive Technology in Information Systems

During the needs analysis phase of the system design, a complete analysis of the needs of the user is necessary, including assistive needs. Once this is done, there is need to directly include functionalities that act as assistive technologies in the design. For instance, for a user with low vision, it would be useful to include the ability to magnify various sections to facilitate reading. The ability to check spelling, including help with writing could be very useful for some people with dyslexia.

Figure 2. Some design considerations for some disabilities



Present Information in Multiple Formats

People have preference for different formats of information. In the case of students with disabilities for instance, learning materials could be presented in various formats, including digital Braille, text (e.g. PDF, Word, HTML), audio and video with captions. For learners who are blind, the Braille format will be very useful, as will the audio and text formats. The text formats in this case could be read out loud by an inbuilt screen reader. For an individual with hearing impairment, both audio and video need to be captioned. On designing learning environments to meet their needs, a student with hearing impairment had this to say: “Even if there is no video available, it is nice to pair audio and video formats- it helps to keep track of reading and set a pace”.

Consider Implementing Multimodal Interaction

Interacting with information systems and inputting information is not just limited to using a physical or virtual keyboard. In order to fully meet the needs of all users with disabilities when interacting with information systems, designers need to consider implementing multimodal interaction interfaces. People have different ways of interacting with systems and this may be due to disability. An individual with an upper limb mobility difficulty who cannot use their hands could benefit from using a speech-to-text software. Those who are paralysed but could still see and can move their head around could benefit from eye tracking interactions.

FUTURE RESEARCH

Designers and developers of information systems are becoming more aware of the needs of people with disabilities, due to the existence of several legislations and also due to many sensitizations and the existence of standards such as the Web

Content Accessibility Guidelines (WCAG 2.0). Although it is not yet the case, in future each designer will by default look to the needs of people with disabilities in order to design an information system, without which their product will not be adopted by many organizations. By default, people will design for accessibility and usability. Future research should look at the possibility of incorporating multimodal interactions in information systems and also automatically adapting to the needs of the user.

Given that some assistive technologies are incompatible with some information systems, it is worth researching how compatibility issues could be resolved for various assistive technologies.

CONCLUSION

Most information systems are not designed with the interest of people with disabilities in mind, thus resulting in inaccessible systems. By developing a disability-aware mentality, designers of information systems can develop products that are both usable and accessible not only to people with disabilities, but to everyone.

This article has proposed a disability-aware approach to information systems design and development and shown with application to learning systems that this is what students with disabilities want and recommend such thought processes during design and development.

REFERENCES

- Boehm, B. W. (1988). A Spiral Model of Software Development and Enhancement. *Software Engineering Notes*, 11, 14–24.
- Bowtell, J. (2015). Assessing the value and market attractiveness of the accessible tourism industry in Europe: A focus on major travel and leisure companies. *Journal of Tourism Futures*, 1(3), 203–222. doi:10.1108/JTF-03-2015-0012

- Gregor, P., Sloan, D., & Newell, A. F. (2005). Disability and technology: Building barriers or creating opportunities? *Advances in Computers*, 64(64), 283–346. doi:10.1016/S0065-2458(04)64007-1
- Grgurovic, M., & Hegelheimer, V. (2007). Help options and multimedia listening: Students' use of subtitles and the transcript. *Language Learning & Technology*, 11(1), 45–66.
- Helander, M., Landauer, T. K., & Prabhu, P. V. (1997). Handbook of human-computer interaction (2nd ed.). Amsterdam: Elsevier.
- Kelly, S. A., & Smith, T. J. (2008). The digital social interactions of students with visual impairments: Findings from two national surveys. *Journal of Visual Impairment & Blindness*, 102(9), 528–539.
- Kim, J. S., & Kim, C. H. (2014). A review of assistive listening device and digital wireless technology for hearing instruments. *Korean Journal of Audiology*, 18(3), 105–111. doi:10.7874/kja.2014.18.3.105
- Kyng, M. (1991). Designing for cooperation - Cooperating in design. *Communications of the ACM*, 34(12), 64–73. doi:10.1145/125319.125323
- McCarthy, T., Pal, J., & Cutrell, E. (2013). The Voice Has It: Screen Reader Adoption and Switching Behavior Among Vision Impaired Persons in India. *Assistive Technology*, 25(4), 222–229. doi:10.1080/10400435.2013.768719 PMID:24620705
- Muller, M. J. (1996). Defining and designing the Internet - Participation by Internet stakeholder constituencies. *Social Science Computer Review*, 14(1), 30–33. doi:10.1177/089443939601400109
- Newell, A. F., & Gregor, P. (2000). "User Sensitive Inclusive Design" - in search of a new paradigm. Academic Press.
- Newell, A. F., Gregor, P., Morgan, M., Pullin, G., & Macaulay, C. (2010). User-Sensitive Inclusive Design. *Univ Access Inf Soc*, 9(2).
- Nganji, J. T., & Brayshaw, M. (2015). *Personalizing learning materials for students with multiple disabilities in virtual learning environments*. Paper presented at the Science and Information Conference (SAI), London, UK. doi:10.1109/SAI.2015.7237128
- Nganji, J. T., Brayshaw, M., & Tompsett, B. (2013). Describing and Assessing Image Descriptions for Visually Impaired Web Users with IDAT. *Advances in Intelligent Systems and Computing*, 179, 27–37. doi:10.1007/978-3-642-31603-6_3
- Nganji, J. T., & Nggada, S. H. (2011). Disability-Aware Software Engineering for Improved System Accessibility and Usability. *International Journal of Software Engineering and Its Applications*, 5(3), 47–62.
- Nielsen, J. (1993). *Usability engineering*. Boston: Academic Press.
- Norman, D. A. (1988). *The psychology of everyday things*. New York: Basic Books.
- Pino, M., & Mortari, L. (2014). The Inclusion of Students with Dyslexia in Higher Education: A Systematic Review Using Narrative Synthesis. *Dyslexia (Chichester, England)*, 20(4), 346–369. doi:10.1002/dys.1484 PMID:25293652
- Ripat, J. D., Brown, C. L., & Ethans, K. D. (2015). Barriers to Wheelchair Use in the Winter. *Archives of Physical Medicine and Rehabilitation*, 96(6), 1117–1122. doi:10.1016/j.apmr.2015.01.020 PMID:25666785
- Royce, W. W. (1987). Managing the Development of Large Software Systems. In *Proc. 9th. Intern. Conf. Software Engineering*. IEEE Computer Society.
- Steyaert, J. (2005). *Web-based higher education: The inclusion/exclusion paradox*. Academic Press.
- WHO. (2015). *Disability and health*. Retrieved from <http://www.who.int/mediacentre/factsheets/fs352/en/>

ADDITIONAL READING

Gregor, P., Sloan, D., & Newell, A. F. (2005). Disability and technology: Building barriers or creating opportunities? *Advances in Computers*, 64(64), 283–346. doi:10.1016/S0065-2458(04)64007-1

Nevile, L., & Treviranus, J. (2006). Interoperability for individual learner centred accessibility for web-based educational systems. *Journal of Educational Technology & Society*, 9(4), 215–227.

Nganji, J., & Brayshaw, M. (2011). *Towards An Ontology-Based Community of Agents for Personalisation of Services for Disabled Students*. Paper presented at the IADIS Interfaces and Human Computer Interaction, Rome, Italy.

Nganji, J., & Brayshaw, M. (2013). Designing Personalised Learning Resources for Disabled Students Using an Ontology-Driven Community of Agents. In I. Pedro & N. Miguel Baptista (Eds.), *Information Systems Research and Exploring Social Artifacts: Approaches and Methodologies* (pp. 81–102). Hershey, PA, USA: IGI Global. doi:10.4018/978-1-4666-2491-7.ch005

Nganji, J. T. (2012). Designing Disability-Aware E-Learning Systems: Disabled Students' Recommendations. *International Journal of Advanced Science and Technology*, 48, 61–70.

Nganji, J. T. (2015). The Portable Document Format (PDF) accessibility practice of four journal publishers. *Library & Information Science Research*, 37(3), 254–262. doi:10.1016/j.lisr.2015.02.002

Nganji, J. T., & Brayshaw, M. (2014). *Designing and Reflecting on Disability-Aware E-learning Systems: The Case of ONTODAPS*. Paper presented at the 2014 IEEE 14th International Conference on Advanced Learning Technologies (ICALT), Athens. doi:10.1109/ICALT.2014.167

Nganji, J. T., & Brayshaw, M. (2015). Facilitating Learning Resource Retrieval for Students with Disabilities through an Ontology-Driven and Disability-Aware Virtual Learning Environment. *International Journal of Information Retrieval Research*, 5(3), 75–98. doi:10.4018/IJIRR.2015070105

Nganji, J. T., Brayshaw, M., & Tompsett, B. (2011). Ontology-Based E-Learning Personalisation for Disabled Students in Higher Education. *ITALICS*, 10(1), 1–11. doi:10.11120/ital.2011.10010001

Nganji, J. T., Brayshaw, M., & Tompsett, B. (2013). Ontology-driven disability-aware e-learning personalisation with ONTODAPS. *Campus-Wide Information Systems*, 30(1), 17–34. doi:10.1108/10650741311288797

KEY TERMS AND DEFINITIONS

Accessibility: The design of information systems such that they can be used by everyone, regardless of their ability or disability.

Assistive Listening Device: This is a device which helps to improve hearing for people with hearing impairment.

Assistive Technology: A piece of software or device that allows a person with disability to interact with an information system without difficulty.

Disability: A physical or mental condition that may limit an individual's ability to fully utilize their senses such as vision and hearing and could thus limit their use of information systems, particularly when they are not designed accessibly.

Disability-Aware Information System: Information system that considers the needs of users with disabilities and adjusts to their needs without them having to seek extra assistive technology in order to interact with it.

Induction Loop: Also known as hearing loop, this is an assistive listening device which works by amplifying sound for an individual who can only hear partially and may not be able to distinguish speech when there is noise.

Mentality: A mindset or way of thinking. In this case, the mindset that designers of information systems should adopt in order to develop accessible and usable systems.

Subtitles: These are texts displayed in a video usually serving as a translation of the words spoken

in the video into another language so that people who don't understand the language of the video could understand what it is all about.

Transcript: In the context of this chapter, it is a text version of a video that is produced for an individual with hearing impairment, to facilitate understanding of the words spoken in the video.

Usability: This refers to how easy it is for an end user to use an information system.

Category B
Big Data

Adapting Big Data Ecosystem for Landscape of Real World Applications

Jyotsna Talreja Wassan
University of Delhi, India

INTRODUCTION

Big data is revolutionizing world in the age of Internet. The wide variety of areas like online businesses, electronic health management, social networking, demographics, geographic information systems, online education etc. are gaining insight from *big data principles*. Big data is comprised of heterogeneous datasets which are too large to be handled by traditional relational database systems. An important reason for explosion of interest in big data is that it has become cheap to store volumes of data and there is a major rise in computation capacity.

To extract valuable patterns from big data, one needs to choose a right platform for capturing, organizing, searching and analyzing the context of voluminous data.

Data Management systems adhering to big data, aim to add computing nodes to cater to increasing data volumes, automatic balancing of data between various nodes, and reducing the operational cost for functioning of distributing data over various nodes (Patel, 2016).

Various NoSQL data stores like Cassandra, MongoDB and Hadoop HBASE etc. are in use today to acquire, manage, store and query big data. NoSQL databases are inherently schema-less and permit records to have variable number of fields, making them distinct from other non-relational databases like hierarchical databases and object-oriented databases. These are highly scalable and well suited for dynamic data structures. NoSQL data is characterized by being basically available and eventually consistent.

The frameworks like MapReduce, Dryad etc. support processing of large amounts of data in parallel and hence the management of big data (Singh & Reddy, 2015). The technologies like GNU R and Apache MAHOUT are useful in exploring and analyzing big data for finding relevant valuable patterns. This article aims at giving an overview of Big Data Ecosystem comprising various big data platforms useful in today's competitive world.

BACKGROUND

In 1970's big meant megabytes, subsequently with the increasing data needs, it grew to gigabytes and terabytes and further to zettabytes with the increase in digital information. The traditional world of relational database systems like Oracle RDBMS etc. faced challenges in storing large quantities of data and needed to scale databases to data volumes beyond the storage and/or processing capabilities of a single large computer system. Many efforts have been made to store and manage data being generated from everywhere on the web. Several database management systems were proposed on the basis of master/slave, cluster computing or partitioning architecture like IBM DB2 partitioning, VoltTB etc.

However, the problems in reliance on shared facilities and resources (CPU, Disk, and Processors), scalability and complex administration limitations, augmented by lack of support for critical requirements, led to development of SHARED NOTHING architectures (Strauch, 2011; Lee, 2011) in 1980's. These systems focused on paral-

lel and distributed data computation and solved big data problems using parallel computations. By 90's, even these solutions faced challenges in running OLTP and queries due to data overload. To provide solutions to these problems, Google responded with its GFS (Dean & Ghemawat, 2004), followed by a powerful programming paradigm of MapReduce (Dean & Ghemawat, 2004). Thereafter a spectrum of new technologies emerged as the NoSQL movement stating a broad class of database management system to support increasing data storage and analytical requirements.

MAIN FOCUS

Major real world applications like health care, business analytics etc., operational on big data, cannot store or process all of the data on just one machine. The data must be stored, distributed or processed in parallel manner for computations to be completed efficiently. Various platforms are making *big data* management and processing more effective, forming the basis of current research theme in the era of *Big Data* (Gandomi & Haider, 2015). The main focus of this article is to discuss about NoSql Movement, big data platforms which could support processing of futuristic massive volumes of data in parallel and their applications.

BIG DATA ECOSYSTEM

Big data could be visualized in two aspects:

1. Big Data Storage and
2. Big Data Analytics.

Hence, Big data applications fall under two general categories:

1. Large-volume applications needing hundreds of terabytes of data to work on, or
2. Performance-intensive big data analytics applications needing computation.

Broadly big data ecosystem comprise of stacked layers of Storage (NoSQL stores), MapReduce and Query (SMAQ) as illustrated in Figure 1. SMAQ structured systems are typically open source and distributed. These systems are changing the landscape of big data processing to a broader class of users similarly as LAMP stack of Linux, Apache, MySQL and PHP changed the horizon of developing web applications (Dumbill, E., 2010).

The base idea is to store big data in parallel NoSQL data stores like MongoDB, Apache HBASE, and BigTable etc.

Above the storage layer, a layer is required to divide the stored big data set, and run it in parallel over many machine nodes. This distribution provides a solution to the issue of data being too

Figure 1. SMAQ stack for big data

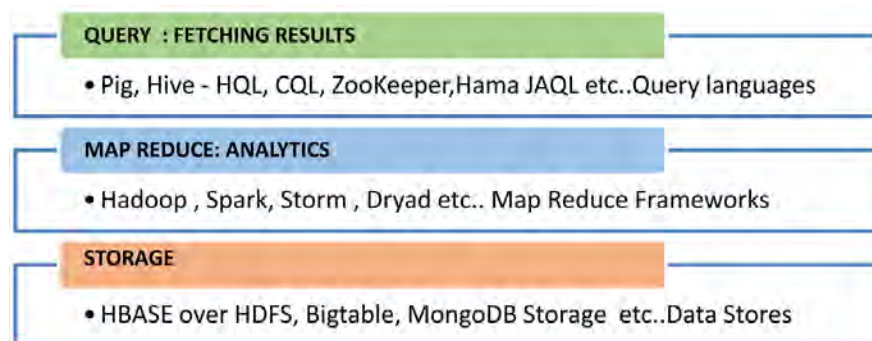


Table 1. Features of NoSQL databases

Feature of NoSQL	Description
Schema-less	Database “Tables” don’t have a pre-defined schema. Records can have variable number of fields and record contents and semantics are enforced by concerned applications.
Shared nothing architecture	Each server uses its own local storage instead of available common storage pool. Thus storage is accessed at local disk speeds instead of network speeds, and this permits capacity to be enhanced by adding more number of nodes.
Sharding	It is a type database partitioning that supports horizontal partitioning of data records in smaller, faster and easily manageable shards to achieve high levels of scalability. An existing shard splits when it gets too much loaded with data. Applications can assist in data sharding by assigning each record a partition key ID
Asynchronous replication	NoSQL databases employ asynchronous replication to scale up horizontally, to become highly available and to complete processing more quickly as they don’t depend on extra network traffic. This corresponds traditionally to RAID storage or synchronous replication.
Follows CAP & BASE instead of ACID	The major characteristics common amongst NoSQL databases, working over distributed environment are: Consistency , Availability and Partition Tolerance . Eric Brewer, a professor at the University of California, Berkeley, and amongst the founders of Google, in 2001, introduced the idea that there is a fundamental trade-off between consistency, availability, and partition tolerance when considering a distributed environment and came up the CAP-theorem, which is widely adopted today in the NoSQL community (E. A. Brewer, 2000 ; S. Gilbert et al., 2002) NoSQL databases emphasize performance and availability over distributed environment. An application works basically all the time (basically available), does not have to be consistent all the time (soft-state) but will be in some known-state state eventually (eventual consistency).

large to fit onto a single machine and is commonly known as MapReduce. Many of the platforms like Hadoop, Storm, and MongoDB etc. support parallel computations through MapReduce.

SMAQ systems also incorporate a HLQL (higher-level query language) layer to simplify both the specification of the MapReduce operations and the retrieval of the result. Several of these HLQL’s like HiveQL, Pig Latin, and JAQL (Stewart, R.J. et al., 2011) etc. have emerged either as open source projects or commercial products. All of them translate higher level jobs into MapReduce jobs. HiveQL is like SQL.

NoSQL MOVEMENT FOR BIG DATA STORAGE

NoSQL movement drives through the database technology, to overcome the limitations associated with scaling and performance of traditional relational databases (RDBMS) over distributed architectures.

NoSQL stands for “Not Only SQL” or “Not Relational”. The growing needs to process continuously increasing volumes of data in lesser time, led to the movement of developing schema less data stores, parallel programming models and various analytical platforms. This movement is commonly covered under the term NoSQL. NoSQL platforms are non-relational, distributed and horizontally scalable. NoSQL platforms have been motivated by large scale web based applications, where data has to be compulsively partitioned over multiple nodes that share the load, and parallelism is essential. This section aims at providing an overview of several characteristics and categories of NoSQL data stores over SQL, based on data models (Stonebraker, M.,2010). NoSQL has some essential features as described in Table 1 (Cattell, R., 2011).

NoSQL Systems may characterize data models based on *Key*, *Aggregation* and *Connections* (Adiba, M., Castrejon-Castillo, J. C., Oviedo, et al, 2016). There are various kinds of NoSQL databases namely key-value store, document

Table 2. NoSQL datamodels

Type	Data Models	CRUD Operations	Description
Key	Key-Value	Create, Read, Update or Delete a record based on primary key	A <i>key-value store</i> provides one of the simplest possible data extraction base in which a user is allowed to fetch data by using the key and the store doesn't know <i>anything</i> about values. This key based approach makes read and writes operations (a simple disk access) very fast.
Aggregation	Document or Column Oriented	Create, Read, Update or Delete a entity/ record based on a specific pattern	A <i>document-oriented store</i> extends the key-value model and values are stored in a <i>structured</i> format (known as document), that the database can understand. For example, a blog post or comments stored as a document in a de-normalized way. Content applications (like Facebook, Twitter etc.) can fetch an entire blog post data with just a single query with the help of <i>document-oriented data stores</i> . A <i>column-oriented store</i> is advantageous for data warehouses as it stores its content by column rather than by row of data sets. It contains predefined columns and provides scaling and updating of data at relatively high speeds.
Connection	Graph	Create, Read, Update or Delete a record based on relationships between entities.	<i>Graph data stores</i> are built with nodes, their properties and relationships amongst them. A flexible graph model is used to scale across many machines instead of tables of rows and columns in the rigid structure of SQL

oriented store, column oriented stores and graph data stores (Russom, P., 2011).

Various examples of NoSQL databases from each of the above mentioned categories are listed in Table 3.

None amongst the above mentioned databases in Table 3 are best for all user cases. The choice of user depends on the application, as well as the type of scalability required. Thus the record semantics in these data stores are enforced by user applications (Tauro & Aravindh et al., 2012).

COMPARISION OF VARIOUS NoSQL BIG DATA PLATFOMS

Variety of NoSQL platforms are discussed in Table 4.

MAP: REDUCE PARADIGM FOR BIG DATA ANALYTICS

Map Reduce is a simple but powerful programming paradigm, used for development of scalable

Table 3. Classifications of NoSQL databases

Category	Databases
Key-Value Databases	DynamoDB, Riak, Redis, Aerospike, Azure Table Storage, FoundationDB, LevelDB, GenieDB, BangDB, Chordless, Scalaris, Tokyo, Cabinet / Tyrant, Scalien, Voldemort, Dynomite, KAI, MemcacheDB, Chordless, HamsterDB, Tarantool/Box, Maxtable, Pincaster, RaptorDB, Mnesia, LightCloud, Hibari, OpenLDAP, LSM, STSdb
Document Databases	MongoDB, Elasticsearch, Couchbase Server, CouchDB, RethinkDB, RavenDB, MarkLogic, Server, Clusterpoint Server, ThruDB, Terrastore, RaptorDB, JasDB, SisoDB, SDB, djondb, EJDB, densodb
Column Databases	Hadoop / HBase, Cassandra, Hypertable, Accumulo, Amazon SimpleDB, Cloudata, Stratosphere, Cloudera, HPCC,
Graph Databases	Neo4J, Infinite Graph, InfoGrid, HyperGraphDB, DEX, GraphBase, Trinity, AllegroGraph, BrightstarDB, Bigdata, Meronymy

Source: <http://nosql-database.org/> last seen on May 2016

Table 4. Comparison of various No-SQL databases

	<i>Riak</i>	<i>MongoDB</i>	<i>HBase</i>	<i>Cassandra</i>	<i>Neo4J</i>
Description	Riak is a distributed open source key/value store by Basho Technologies, which scales predictably and makes development easier by giving users the facility to prototype, test, and deploy their applications.	MongoDB (from “humongous”) is a NoSQL based open source document-oriented database, developed by 10Gen company. MongoDB stores structured data as JSON-like documents having dynamic schemas, facilitating the integration of data in certain types of applications easier and faster. The communication with MongoDB is done using a BSON driver in any language like Ruby, PHP etc. or using the MongoDB JavaScript shell.	HBase is the high-performance non-relational column oriented database, supporting real time and random access to Big Data stored. It was modeled after Google’s Big Table and runs over HDFS. It is under Apache licensing. It is natively integrated with Hadoop and works with other database engines through YARN. It supports master node that manages the cluster and other servers storing portions of data. It scales linearly and has the capability to easily combine data from various sources.	Cassandra is a distributed system, supported by DataStax, and was originally open sourced by Facebook in 2008. It is used under Apache licensing. Jeff Hammerbacher, who led the Facebook data team at that time, described Cassandra as a BigTable data model which uses architectural concepts introduced by Amazon Dynamo. It is a wide column store database.	Neo4j is a robust ACID transactional property graph database developed by Neo Technologies.. Neo4j is highly agile and runs a thousand times faster than relational databases for connected data operations due to its graph data model.
Data Model	Riak uses a simple key/value model for object storage. Riak contains objects in the form of a unique key and value, stored in a flat namespace known as a bucket. One can store anything in Riak including text, images, JSON/XML/HTML documents, log files etc. Data is stored and referenced by bucket/key pairs. Each key is attached to a unique value any data type.	Databases in MongoDB are the groups of collections stored on disk using a single set of data files where collections are containers for documents that share one or more indexes. Each document collection is stored in one namespace file.	HBase system consists of set of tables designed to store semi structured data that could vary in field size, data types or columns. Each table has rows representing objects and columns representing attribute of an object, like traditional databases. Each table must have an element defined as a Primary Key which is used to access the table. HBase allows for many attributes or data in a row to be grouped together into column families, such that the elements of a column family are all stored together. HBase table schema is scalable and flexible.	Cassandra data model is comprised of columns, rows, column families, and keyspace. Each column consists of a name and a value. A row is the collection of columns labeled with a same name. Each row is uniquely identifiable by a key within a column family. Each row in a column family may have a different set of columns. This indicates the schema-less model in Cassandra. A collection of columns along with the row keys forms a column family and the logical grouping of column families is known as keyspace which is used as schema.	Neo4j is an “embedded, disk-based, fully transactional Java persistence engine” that stores data structured in nodes of a graph rather than in tables of traditional RDBMS. The nodes of a graph in Neo4j are organized by relationships which also have properties. A traversal of graph is used to query a Graph, ranging from starting node to related nodes according to sequence of instructions and finding answers to the queries. An index can also be created to perform look up operations as it maps properties to either nodes or relationships.
Implementation Language	Erlang	C++	Java	Java	Java
Technical Details Available on	docs.basho.com	docs.mongodb.org/manual	hbase.apache.org	www.datastax.com/docs	neo4j.com/docs/stable
Supports Map-Reduce	Yes	Yes	Yes	Yes	No
Replication	Selectable Replication Factor	Master- Slave	Selectable Replication Factor	Selectable Replication Factor	Master- Slave
Support for Server Side Scripts	Java Script & Erlang	Java Script	Yes	No	Yes
Partitioning	Sharding	Sharding	Sharding	Sharding	No

Source: <http://db-engines.com/>

applications. It handles the computation in terms of a *Map* and a *Reduce* functions (Dean & Ghemawat, 2004).

Map Function: Takes an input pair and produces a set of intermediate key/value pairs e.g.,

- **Map:** $(key_1, value_1) \rightarrow \text{list of } (key_2, value_2)$.

The MapReduce library groups together all intermediate values associated with the same intermediate key

- **Reduce Function:** Accepts an intermediate key and a set of values for that key.
- **Reduce:** $(key_2, \text{list of } (key_2, value_2)) \rightarrow value_3$.

The underlying runtime system automatically parallelizes the computation across large-scale clusters of computing nodes via automatic load balancing. (Dean, & Ghemawat, 2004).

The MapReduce paradigm could be applied to various areas. To better understand the MapReduce, the following example is illustrated.

Problem Statement

To determine the count of each word that appears in a file/document (or a set of documents). Each file is associated with a document URL/ID.

- **Motivation:** Various real world applications are based on counting and summing problems. E.g. there is a log file where each record contains a URL and it is required to calculate frequency of websites accessed.

Solution

- **Map Function:** Input to map functions is: **Key** = document URL/ID; **Value** = document contents. Outputs of map function are (potentially many) key/value pairs. Output (word, "1") once per word in the document

- **Pseudo Code for Map - Map (String key, String value):** // input key: document name, input value: document contents.

For each word w in value:

Emit Intermediate (w, "1");

- **Example: Key, Value Pair:** "documentID01", "he is a dictator or he is a narrator"

Result of applying the above map function is:

```
"he", 1
"is", 1
"a", 1
"dictator", 1
"he", 1
"is", 1
"a", 1
"narrator", 1
```

Pseudo-Code for Reduce

- **Reduce (String Key, Values):** // key: a word, same for input and output; values: a list of counts

```
int result = 0;
```

For each v in values:

```
Result = Result + value;
```

```
Emit (Result);
```

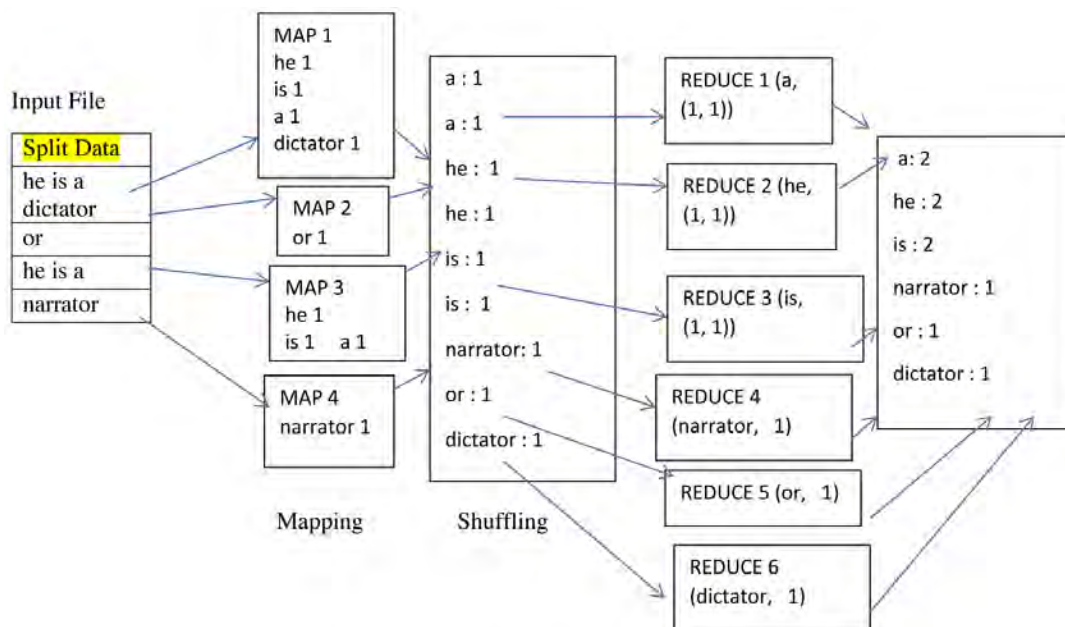
The function sums together all counts emitted for a particular word. The MapReduce framework sorts all pairs with the same key

(a,1), (a,1), (he, 1), (he,1), (is,1), (is,1), (narrator, 1), (or, 1), (dictator,1)

The pairs are then grouped as: (a, 1, 1), (he, 1, 1), (is,1, 1), (or, 1), (narrator, 1),(dictator,1)

The reduce function combines (sums) the values for a key.

Figure 2. Illustration of an example of MapReduce



- **Example:** Applying reduce to (a, 1, 1) gives value as 2. The scenario could be realized as illustrated in Figure 2.

Various applications where MapReduce could be applied are:

- Log Querying and Analysis,
- Geospatial Query Processing,
- Page Rank Algorithms,
- Engineering Simulations,
- Graph Processing,
- Processing Health Care Records (EHRs),
- Relational patterns like joins, projection, union, intersection etc.,
- Machine learning applications like linear regressions,
- k means, neural networks etc., and
- Mathematical Problems like Sparse Matrix operations on large data.

And the list goes on with advancements in distributed environment.

There are scenarios, where MapReduce programming model cannot be employed.

For example, if the output of the computation depends on previously computed values, then MapReduce cannot be used. Fibonacci series calculation where each value is summation of the previous two values: $\text{fib}(k+2) = \text{fib}(k+1) + \text{fib}(k)$, is one of the classic example of this scenario.

Map reduce is maturing to handle complex areas like real time streaming data, XML query processing etc in various platforms.

EMERGING APPLICATIONS OF BIG DATA PLATFORMS

Big Data movement with NoSQL platforms, has become imperative for various industries and these are using big data in variety of contexts (Boinepelli, 2015). NoSQL platforms could be

1. Chosen to develop new applications (cloud based).
2. Chosen to modify an existing (RDBMS) system by augmenting NoSQL platform for making the existing relational system scalable.

Figure 3. Big data and social networks intersection



3. Chosen to replace existing system to fulfill the needs of volume, variety, velocity in data.

In *Internet of things*, NoSQL has emerged as the preferred choice for various applications such as Web Development, Social Networks, Health Sector, Mobile Apps, Education etc (Lohr, S., 2012).

Big Data in Health Care

Healthcare is influenced by various challenges, like high-rising expenditures, inconsistent quality, and delays in care and access in various areas across the world. It is important to improve health care quality and it is widely believed that the use of information technology in health care is aiding in reducing the cost of health care while improving its quality. Especially electronic health records (EHRs), have been thought to be possible solutions to these problems as they store all the information about a patient and make it interoperable and shared among different health care providers. EHRs deal with heterogeneous, volumetric and may time real time data. To deal with this, Healthcare is undergoing the Big Data Revolution (Wassan, 2015). In parallel, advancement in techniques using big data platforms, have made it easier to collect and analyze heterogeneous health data from multiple sources like patients, hospitals, laboratories, and physicians and insurance payors. Big Data interventions have brought about a 300-450

billion dollar reduction in U.S. healthcare (<http://rockhealth.com>). The true potential of digitized information lies in big data handling in EHRs providing structured output. EHRs using big data techniques aim to play a major part in the health-care reforms and health records could be mined to detect fruitful health related patterns using big data analytics tools and techniques.

Big Data in Social Networks

Social networking and media applications containing human communication and interaction also generate large amounts of data whose processing scales out with the use of high performance servers, data mining grids and usage of platforms like Hadoop and parallel modeling paradigms like MapReduce to enhance the computation and storage performance (Figure 3).

Social media providers focus not only on measuring followers, friends and likes on social networking sites like Facebook and Twitter, but also on analyzing and integrating the emerging volumes of data into future business strategies deriving insight for enhancing business intelligence. Web search engines and social networks capture and analyze every user action on their sites to improve site features, generating recommendations for the visitors, spam and fraud detection, and advertising opportunities. This could be analysed efficiently by using big data platforms.

Figure 4. Big data approaching online business trends



Big Data in Education

Big data may also has great impact on the education field like it is impacting other sectors such as social media and medicines. The big data technology also offers a great platform for the authorities looking to improve their educational institutions. Educational institutions are also generating huge volumes of data, from grades or test scores to admissions or enrollment numbers while in-taking students in an institution.

With the advent of online courses offered by many universities, the amount of data available to educational officials and students has exploded. The development of electronic learning courses enable learning and assessment of students in systematic, real-time ways and big data platforms could be fruitful in handling the educational data streams (Wassan, 2015).

Many universities across the world ; have an eAdvisor system in which a candidate can choose amongst broad areas of study to facilitate learning and in identifying study areas of interests. (Parry, 2012). A number of educational institutions have developed dashboard software to monitor learning, performance, and behavioral issues for individual students as well as the school as a whole (West, 2012). All these softwares could be enhanced with Big Data techniques to handle large amounts of educational data.

Big Data in Online Retailing Business

Today more and more online retailers are interacting with their customers in real time. The retail sector has witnessed strong demands and returns on investment in an online environment. The ability to analyze and predict consumer behavior on the basis of patterns mined from stored transactional data; has proved to be an invaluable tool for retailers looking to engage customers and gain an edge over competitors (Figure 4).

With the huge volumes of customer data, online retailers tend to benefit from predictive analytics applications. Various big data applications designed on the Hadoop, Cloudera etc. platforms, can provide businesses with tools that may prove more useful for enhancing customer and retailer relationship. E-Commerce retailers can more effectively manage their resources with the big data solutions. Big data analytics is proving highly useful for online retailers.

RECOMMENDATIONS

Choosing Right Big Data Platform w.r.t to Application Needs

Big Data platform choice is impacted by an architecture (e.g. distributed, clustered machines)

for storing the data and programming models for development of data parallel, distributed applications for big data considering factors like scalability limits, query speed etc. It demands cost effectiveness and real time support for data.

Design Models for Security w.r.t Big Data Analytics

Big data analytics also requires governance, privacy, and security.

FUTURE RESEARCH DIRECTIONS

The scope of big data varies widely. The techniques for storage, management and analysis of big data are still emerging with day to day changing trends in volume, velocity and variety of data.

The continuous efforts taken for building timely and cost-effective management over “Big Data” are appreciable and will form the basis of key ingredient for success in many disciplines. We must aim for designing of efficient applications or their updation to handle big data needs by using big data platforms and gaining insight of their usage.

Also there is a need for building highly scalable comprehensive BDMS (Big Data Management Software) like RDBMS. Asterix, is one such example and is expanding its work in three areas: semi structured data, parallel database systems, and data-intensive computing (Borkar, V.R. & Carey, M.J., et al., 2012). Asterix could be explored as a comprehensive suite.

CONCLUSION

The article focuses on big data revolution i.e. changing trends in data w.r.t size and its structure and Big Data Ecosystem (SMAQ) supporting the same. It focuses on NoSQL movement with various NoSQL databases and their applications in

real world. It listed MapReduce and applications of Big Data in real world.

To conclude, it provides an idea that big data field is emerging and we need to think and query about everything in relation to big data storage, management and analysis. It may involve multi-user, multi-dimensional, heterogeneous data environment having queries of widely varying needs, importance and sizes. Big data platforms must learn to manage such workloads effectively by developing or exploring whole Big Data Management Systems (BDMS)..

REFERENCES

- Adiba, M., Castrejon-Castillo, J. C., Oviedo, J. A. E., Vargas-Solar, G., & Zechinelli-Martini, J. L. (2016). *Big Data Management Challenges, Approaches, Tools and their limitations*. Networking for Big Data.
- Boinepelli, H. (2015). Applications of Big Data. In *Big Data* (pp. 161–179). Springer India. doi:10.1007/978-81-322-2494-5_7
- Borkar, V. R., Carey, M. J., & Li, C. (2012). Big data platforms: Whats next? *XRDS: Crossroads. The ACM Magazine for Students*, 19(1), 44–49. doi:10.1145/2331042.2331057
- Brewer, E. A. (2000). Towards robust distributed systems. *Proceedings of the nineteenth Annual ACM Symposium on Principles of Distributed Computing*. doi:10.1145/343477.343502
- Cattell, R. (2011). Scalable SQL and NoSQL data stores. *SIGMOD Record*, 39(4), 12–27. doi:10.1145/1978915.1978919
- Database Comparative Study. (n.d.). Retrieved May 2016, from <http://db-engines.com>
- Dean, J., & Sanjay, G. (2004). MapReduce: Simplified data processing on large clusters. *Communications of the ACM*, 53(1), 72–77. doi:10.1145/1629175.1629198

- Dumbill, E. (2010). *SMAQ Stack for Big Data*. Retrieved May 2016, from <http://strata.oreilly.com/2010/09/the-smaq-stack-for-big-data.html>
- Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137–144. doi:10.1016/j.ijinfomgt.2014.10.007
10. gen, Inc. (n.d.). *mongoDB*. Retrieved May 2016, from <http://www.mongodb.org>
- Lee, S. (2011). Shared-Nothing vs. Shared-Disk Cloud Database Architecture. *International Journal of Energy, Information and Communications*, 2(4), 212–216.
- Lohr, S. (2012). The age of big data. *New York Times*, 11.
- MapReduce Applications and Limitations. (n.d.). Retrieved May 2016, from <http://mapreduce-specifcs.wikispaces.asu.edu/Applications+and+Limitations+of+MapReduce>
- Neo4j Database. (n.d.). Retrieved May 2016, from <http://docs.neo4j.org/>
- NoSQL Databases. (n.d.). Retrieved May 2016, from <http://nosql-database.org/>; <http://www.datastax.com/nosql-databases>
- Patel, J. M. (2016). Operational NoSQL Systems: Whats New and Whats Next? *Computer, IEEE*, 49(4), 23–30. doi:10.1109/MC.2016.118
- Riak. (n.d.). Retrieved 2016, from <http://basho.com/riak/>
- Russom, P. (2011). *Big Data Analytics*. Renton, WA: TDWI Best Practices Report, Fourth Quarter.
- Singh, D., & Reddy, C. K. (2015). A survey on platforms for big data analytics. *Journal of Big Data*, 2(1), 1-20.
- Stewart, R. J., Trinder, P. W., & Loidl, H. W. (2011). Comparing high level mapreduce query languages. *Proceedings of the of the 9th international conference on Advanced Parallel Processing Technologies*, 58-72. doi:10.1007/978-3-642-24151-2_5
- Stonebraker, M. (2010). SQL databases v. NoSQL databases. *Communications of the ACM*, 53(4), 10–11. doi:10.1145/1721654.1721659
- Strauch, C., Sites, U. L. S., & Kriha, W. (2011). *NoSQL databases*. Retrieved May 2016, from <http://www.christof-strauch.de/nosql dbs.pdf>
- Tauro, C. J., Aravindh, S., & Shreeharsha, A. B. (2012). Comparative Study of the New Generation, Agile, Scalable, High Performance NOSQL Databases. *International Journal of Computers and Applications*, 48(20), 1–4. doi:10.5120/7461-0336
- Wassan, J. T. (2015). Big Data Paradigm for Healthcare Sector. *Managing Big Data Integration in the Public Sector*, 169.
- Wassan, J. T. (2015). Discovering Big Data Modeling for Educational World. *Procedia: Social and Behavioral Sciences*, 176, 642–649. doi:10.1016/j.sbspro.2015.01.522

KEY TERMS AND DEFINITIONS

Big Data: A term that describes large amount of data that could be structured, unstructured or real time and is difficult to be handled by traditional systems. It is identified by major 3 V's: Volume, Variety and Velocity.

Data Intensive Computing: It classifies various parallel computing applications which use a data parallel approach to process large volumes of data like terabytes or petabytes in size and referred to as Big Data.

Database Management Systems (DBMS): A Database Management Systems (DBMS) is a set of programs that enables storing, adding, deleting, accessing, modifying, updating or analyzing data stored in one location.

Distributed System: It consists of autonomous machine nodes connected in a network to communicate, share and coordinate their activities through message passing to achieve a common goal.

MapReduce: MapReduce is a parallel programming model proposed by Google and is used to distribute computing on clusters of computers for processing large data sets.

Online Transaction Processing (OLTP): It refers to a class of systems that facilitate and manage transaction-oriented applications, typically for data entry and retrieval transaction processing.

B

Big Data Analysis and Mining

Carson K. Leung

University of Manitoba, Canada

INTRODUCTION

Advancements in the field of information science and technology enables users to collect or generate high volumes of valuable data of different levels of veracities—such as streams of banking, financial, and shopper market basket data—at high velocities from wide varieties of data source in various real-life business, engineering, scientific applications in modern organizations and society. Embedded in these *big data* (Madden, 2012; Leung, 2015) is implicit, previously unknown, and potentially useful information and knowledge. However, these big data come with volumes beyond the ability of commonly-used software to capture, manage, and process within a tolerable elapsed time. Hence, new forms of information science and technology—such as *big data analysis and mining*—are needed to process and analyze these big data so to as enable enhanced decision making, insight, knowledge discovery, and process optimization. Over the past few years, algorithms have been proposed for various big data analysis and mining tasks, including clustering (which groups similar data together), classification (which categorizes groups of similar data), outlier detection (which identifies anomalies), and frequent pattern mining (which discovers interesting knowledge in the forms of frequently occurring sets of merchandise items or events). Most of these algorithms use the *MapReduce* model—which mines the search space with distributed or parallel computing (Shim, 2012). Among different big data analysis and mining tasks, this chapter focuses on applying the MapReduce model to big data for the discovery of frequent patterns.

BACKGROUND

Since the introduction of the research problem of *frequent pattern mining* (Agrawal, Imieliński, & Swami, 1993), numerous algorithms have been proposed (Hipp, Güntzer, & Nakhaeizadeh, 2000; Ullman, 2000; Ceglar & Roddick, 2006). Notable ones include the classical Apriori algorithm (Agrawal & Srikant, 1994) and its variants such as the Partition algorithm (Savasere, Omiecinski, & Navathe, 1995). The Apriori algorithm uses a level-wise breadth-first bottom-up approach with a candidate generate-and-test paradigm to mine frequent patterns from transactional databases of precise data. The Partition algorithm divides the databases into several partitions and applies the Apriori algorithm to each partition to obtain patterns that are locally frequent in the partition. As being locally frequent is a necessary condition for a pattern to be globally frequent, these locally frequent patterns are tested to see if they are globally frequent in the databases. To avoid the candidate generate-and-test paradigm, the tree-based FP-growth algorithm (Han, Pei, & Yin, 2000) was proposed. It uses a depth-first pattern-growth (i.e., divide-and-conquer) approach to mine frequent patterns using a tree structure that captures the contents of the databases. Specifically, the algorithm recursively extracts appropriate tree paths to form projected databases containing relevant transactions and to discover frequent patterns from these projected databases.

In various real-life business, engineering, scientific applications in modern organizations and society, the available data are not *precise* data but *uncertain* data (Tong et al., 2012; Leung,

Cuzzocrea, & Jiang, 2013; Leung, MacKinnon & Tanbeer, 2014; Jiang & Leung, 2015; Ahmed et al., 2016). Examples include sensor data and privacy-preserving data. Over the past few years, several algorithms have been proposed to mine and analyze these uncertain data. The tree-based UF-growth algorithm (Leung, Mateo, & Brajczuk, 2008) is an example.

With high volumes of big data, it is not unusual for users to have some phenomenon in mind. For example, a store manager is interested in some promotional items. Hence, it would be more desirable if data mining algorithms return only those patterns containing the promotional items rather than returning all frequent patterns, out of which many may be uninteresting to the store manager. It leads to *constrained mining*, in which users can express their interests by specifying constraints and the mining algorithm can reduce the computational effort by focusing on mining those patterns that are interesting to the users.

Besides the aforementioned algorithms discover frequent patterns *in serial*, there are also *parallel and distributed* frequent pattern mining algorithms (Zaki, 1999). For example, the Count Distribution algorithm (Agrawal & Shafer, 1996) is a parallelization of the Apriori algorithm. It divides transactional databases of precise data and assigns them to parallel processors. Each processor counts the frequency of patterns assigned to it and exchanges this frequency information with other processors. This counting and information exchange process is repeated for each pass/database scan.

As we are moving into the new era of big data, more efficient mining algorithms are needed because these data are wide varieties of valuable data of different veracities with volumes beyond the ability of commonly-used algorithms for mining and analyzing within a tolerable elapsed time. To handle big data, researchers proposed the use of the *MapReduce programming model*.

BIG DATA ANALYSIS AND MINING FOR FREQUENT PATTERNS

B

The MapReduce Programming Model

MapReduce (Dean & Ghemawat, 2004; Dean & Ghemawat, 2010) is a high-level programming model for processing high volumes of data. It uses parallel and distributed computing on large clusters or grids of nodes (i.e., commodity machines), which consist of a master node and multiple worker nodes. As implied by its name, MapReduce involves two key functions:

1. The “map” function, and
2. The “reduce” function.

To solve a problem using MapReduce, the master node reads and divides input big data into several partitions (sub-problems), and then assigns them to different worker nodes. Each worker node executes the *map function* on each partition (sub-problem). The map function takes a pair of $\langle \text{key}, \text{value} \rangle$ and returns a list of $\langle \text{key}, \text{value} \rangle$ pairs as an intermediate result:

- **Map:** $\langle \text{key}_1, \text{value}_1 \rangle \mapsto \text{list of } \langle \text{key}_2, \text{value}_2 \rangle$,

where:

1. key_1 & key_2 are keys in the same or different domains, and
2. value_1 & value_2 are the corresponding values in some domains.

The pairs in the list of $\langle \text{key}, \text{value} \rangle$ pairs for this intermediate result are then shuffled and sorted. Each worker node then executes the *reduce function* on:

1. A single key from this intermediate result, and
2. The list of all values that appear with this key in the intermediate result.

The reduce function “reduces”—by combining, aggregating, summarizing, filtering, and/or transforming—the list of values associated with a given key (for all k keys) in worker nodes and returns a list of $\langle \text{key}, \text{value} \rangle$ pairs, a list of values, or simply a single (aggregated or summarized) value:

- **Reduce:** $\langle \text{key}_2, \text{list of value}_2 \rangle \mapsto \text{list of } \langle \text{key}_3, \text{value}_3 \rangle$, or list of value_3 , or a single value_3 ,

where:

1. key_2 & key_3 are keys in some domains, and
2. value_2 & value_3 are the corresponding values in some domains.

By using the MapReduce model, users only need to focus on (and specify) the map and reduce functions, without worrying about implementation details for partitioning the input data, scheduling and executing the program across multiple machines, handling machine failures, or managing inter-machine communication.

Earlier works on MapReduce mainly focused on data processing in big databases (Dean & Ghemawat, 2004) or some big data mining tasks other than frequent pattern mining. Examples of big data processing with MapReduce include the construction of inverted indexes, the evaluation of queries involving joins or selection, the elimination of duplicates, text processing tasks (Lin & Dyer, 2010; Jiang & Leung, 2015) like the word counting of documents, and the processing of data cubes (Lee & Kim, 2016). Examples of big data mining with MapReduce include clustering (Shahrivari & Jalili, 2016), classification (Bechini, Marcelloni, & Segatori, 2016), and outlier detection (Liao & Squicciarini, 2015).

Apriori-Based MapReduce Algorithms: SPC, FPC, and DPC

To mine frequent patterns from big databases of precise data with MapReduce, Lin et al. (2012)

proposed three algorithms—namely, the Single Pass Counting (SPC), Fixed Passes Combined-counting (FPC), and Dynamic Passes Combined-counting (DPC) algorithms—based on both the Apriori and the Count Distribution algorithms. SPC first divides the databases into partitions, and then executes map and reduce functions in each pass k to generate candidate k -itemsets (i.e., candidate patterns each consisting of k items) and count their support/frequency. More specifically, SPC executes the following map and reduce functions in Pass 1:

- **Map:** $\langle \text{ID of transaction } t_j \in \text{partition } P_i, \text{contents of } t_j \rangle \mapsto \text{list of } \langle \text{item } x \in t_j, 1 \rangle$, and
- **Reduce:** $\langle x, \text{list of 1's} \rangle \mapsto \text{list of } \langle \text{frequent 1-itemset } \{x\}, \text{sup}(\{x\}) = \text{sum of 1's in the list for } x \rangle$.

Here, the worker node corresponding to each partition P_i of the big databases executes the map function by outputting $\langle x, 1 \rangle$ (which represents the support of candidate 1-itemset $\{x\}$ in $t_j = 1$) for every item x in transaction $t_j \in P_i$:

For each transaction $t_j \in \text{partition } P_i$ **do**
 for each item $x \in t_j$ **do**
 output $\langle x, 1 \rangle$.

The reduce function is then executed by summing all the 1's for each x to compute its support $\text{sup}(\{x\})$, and outputting $\langle \{x\}, \text{sup}(\{x\}) \rangle$ (which represents a frequent 1-itemset $\{x\}$ and its frequency) if $\text{sup}(\{x\}) \geq$ a user-specific *minsup* threshold:

For each $x \in \langle x, \text{list of 1's} \rangle$ **do**
 $\text{sup}(\{x\}) = \text{sum of 1's in the list for } x$
 if $\text{sup}(\{x\}) \geq \text{minsup}$
 then output $\langle \{x\}, \text{sup}(\{x\}) \rangle$.

In each subsequent pass $k \geq 2$, SPC generates candidate k -itemsets from frequent $(k-1)$ -itemsets. The worker node corresponding to each partition P_i then outputs $\langle X, 1 \rangle$ for every candidate

k -itemset X that exists in some transaction $t_j \in P_i$. Afterwards, the reduce function sums all the 1's for each X to compute its support $sup(X)$, and output $\langle X, sup(X) \rangle$ (which represents a frequent k -itemset X and its frequency) if $sup(X) \geq minsup$. In other words, SPC executes the following map and reduce functions in Pass k (for every $k \geq 2$):

- **Map:** $\langle \text{ID of } t_j \in P_i, \text{contents of } t_j \rangle \mapsto \text{list of } \langle \text{candidate } k\text{-itemset } X \subseteq t_j, 1 \rangle, \text{ and}$
- **Reduce:** $\langle X, \text{list of } 1\text{'s} \rangle \mapsto \text{list of } \langle \text{frequent } k\text{-itemset } X, sup(X) = \text{sum of } 1\text{'s in the list for } X \rangle$.

The three algorithms (SPC, FPC and DPC) are similar except that both FPC and DPC apply the *pass bundling* technique to reduce the number of passes/database scans in each P_i when generating candidate itemsets (for Pass $k \geq 3$). For instance, FPC statistically bundles a fixed number of passes (e.g., three passes) to generate all candidate k -, $(k+1)$ -, and $(k+2)$ -itemsets from frequent $(k-1)$ -itemsets. In contrast, DPC dynamically bundles several passes (depending on the number of generated candidates in these bundled passes).

Partition-Based MapReduce Algorithm

On the one hand, the use of the pass bundling technique reduces the number of passes/database scans in each partition P_i . On the other hand, the Apriori-based SPC, FPC and DPC algorithms still require multiple passes/database scans when mining frequent patterns from big data. In contrast, the MapReduce version of the Partition algorithm—presented by Rajaraman & Ullman (2011)—requires only two passes/database scans. Hence, the Partition-based MapReduce algorithm uses two sets of “map” and “reduce” functions. Specifically, the master node reads and divides big databases of precise data into partitions. The worker node corresponding to each partition P_i then outputs $\langle X, 1 \rangle$ for every candidate pattern X (of any cardinality) that exists in some transaction $t_j \in P_i$. The reduce function then sums all

the 1's for each X to find patterns that are locally frequent in P_i . Taking the union of these locally frequent patterns forms global candidate patterns. To summarize, the first set of “map” and “reduce” functions can be expressed as follows:

- **Map:** $\langle \text{ID of transaction } t_j \in P_i, \text{contents of } t_j \rangle \mapsto \text{list of } \langle \text{itemset } X \subseteq t_j, 1 \rangle, \text{ and}$
- **Reduce:** $\langle X, \text{list of } 1\text{'s} \rangle \mapsto \text{list of } \langle \text{locally frequent itemset } X, \text{NULL} \rangle$.

Afterwards, the worker node corresponding to each partition P_i outputs $\langle X, 1 \rangle$ for every global candidate pattern X that exists in some transaction $t_j \in P_i$. Then, the reduce function sums all the 1's for each X to compute its support $sup(X)$, and outputs $\langle X, sup(X) \rangle$ (which represents a globally frequent pattern X and its frequency) if $sup(X) \geq minsup$. To summarize, the second set of “map” and “reduce” functions can be expressed as follows:

- **Map:** $\langle \text{ID of } t_j \in P_i, \text{contents of } t_j \rangle \mapsto \text{list of } \langle \text{global candidate pattern } X \subseteq t_j, 1 \rangle, \text{ and}$
- **Reduce:** $\langle X, \text{list of } 1\text{'s} \rangle \mapsto \text{list of } \langle \text{globally frequent pattern } X, sup(X) \rangle$.

Tree-Based MapReduce Algorithm for Precise Data: PFP

Given that tree-based algorithms avoid the candidate generate-and-test paradigm of Apriori-based algorithms, Li et al. (2008) proposed the Parallel FP-growth (PFP) algorithm for query recommendation. PFP uses MapReduce to parallelize the tree-based FP-growth algorithm by first reading and dividing big databases of precise data into several partitions. The worker node corresponding to each partition P_i then outputs $\langle \{x\}, 1 \rangle$ for every item x in transaction $t_j \in P_i$. The reduce function then sums all the 1's for each x to compute its support $sup(\{x\})$, and outputs $\langle \{x\}, sup(\{x\}) \rangle$ (which represents a frequent 1-itemset $\{x\}$ and its frequency) if $sup(\{x\}) \geq minsup$. In other words, PFP first executes the following set of “map” and “reduce” functions:

- **Map:** $\langle \text{ID of transaction } t_j \in \text{database partition } P_i, \text{ contents of } t_j \rangle \mapsto \text{list of } \langle \text{item } x \in t_j, 1 \rangle$, and
- **Reduce:** $\langle x, \text{list of 1's} \rangle \mapsto \text{list of } \langle \text{frequent 1-itemset } \{x\}, \text{sup}(\{x\}) = \text{sum of 1's in the list for } x \rangle$.

Afterwards, PFP reads the big databases a second time to form an $\{x\}$ -projected database (i.e., a collection of transactions containing x) for each item x in the list produced from the first reduce function (i.e., for each frequent 1-itemset $\{x\}$). The worker node corresponding to each projected database then:

1. Builds appropriate local FP-trees (based on the projected database assigned to the node) to mine frequent k -itemsets (for $k \geq 2$), and
2. Outputs $\langle X, \text{sup}(X) \rangle$ (which represents a frequent k -itemset X and its frequency) if $\text{sup}(X) \geq \text{minsup}$.

To summarize, PFP executes the second set of “map” and “reduce” functions as follows:

- **Map:** $\langle \text{ID of transaction } t_j \in P_i, \text{ contents of } t_j \rangle \mapsto \text{list of } \langle \{x\}, \{x\}\text{-projected database} \rangle$, and
- **Reduce:** $\langle \{x\}, \{x\}\text{-projected database} \rangle \mapsto \text{list of } \langle \text{frequent itemset } X, \text{sup}(X) \rangle$.

As PFP was designed for query recommendation, it usually takes a third set of “map” and “reduce” functions to aggregate and rank the list of frequent itemsets for the top-K frequent patterns to facilitate recommendations.

Tree-Based MapReduce Algorithm for Uncertain Data: MR-growth

Characteristics of big data can be described by the 5Vs: volume, value, velocity, variety, and veracity. Among the 5Vs, veracity focuses on the quality of data (e.g., precision, uncertainty, messiness, or trustworthiness of data). In many

real-life applications, available data can be *uncertain*. Uncertainty of the data may partially be caused by various factors such as imprecision or limitation of measuring instruments, as well as intentional blurring of data for privacy-preserving data. Hence, in these applications, users may be uncertain about the presence or absence of some merchandise items or events. For example, a manager may highly suspect (but cannot guarantee) that a customer is interested in certain products without explicitly asking the customer. The uncertainty of such suspicion can be expressed in terms of existential probability. Hence, to handle uncertain data, each item x in the transaction t_j is associated with an existential probability $P(x, t_j)$ expressing the likelihood of the presence of that item or event. With this notion, each item in a transactional database of precise data can be viewed as an item with a 100% likelihood of being present in the transaction.

When using probabilistic-based mining with the “possible world” interpretation (Leung, 2013; Leung, 2014), a pattern is considered *frequent* if its expected support is no less than the user-specified *minsup* threshold. When items within a pattern X are independent, the *expected support* of X in the database can be computed by summing (over all transactions) the product (of existential probabilities within X):

$$\text{expSup}(X) = \sum_{t_j} (\prod_{x \in X} P(x, t_j))$$

where $P(x, t_j)$ is the existential probability of item x in transaction t_j .

Leung & Hayduk (2013) presented the MR-growth algorithm, which uses MapReduce to mine frequent patterns from *uncertain* data in a tree-based pattern-growth fashion for big data mining. Again, MR-growth uses two sets of the “map” and “reduce” functions. Specifically, the master node reads and divides uncertain data into partitions. The worker node corresponding to each partition P_i then outputs $\langle \{x\}, P(x, t_j) \rangle$ for every item x in transaction $t_j \in P_i$. Note that, unlike the map functions for the aforementioned algorithms

(which output $\langle \{x\}, 1 \rangle$ for representing the actual support of x in $t_j = 1$, i.e., x appears 1 time in t_j), the map function for MR-growth outputs $\langle \{x\}, P(x, t_j) \rangle$ (where $P(x, t_j)$ indicates the expected support of x in t_j , which means x has a probability of $P(x, t_j)$ to appear in t_j):

- **Map:** $\langle \text{ID of transaction } t_j \in \text{database partition } P_i, \text{ contents of } t_j \rangle \mapsto \text{list of } \langle \text{item } x \in t_j, P(x, t_j) \rangle$.

Then, the reduce function sums all the $P(x, t_j)$'s for each x to compute its expected support $\text{expSup}(\{x\})$, and outputs $\langle \{x\}, \text{expSup}(\{x\}) \rangle$ (representing a frequent 1-itemset $\{x\}$ and its expected support) if $\text{expSup}(\{x\}) \geq \text{minsup}$:

- **Reduce:** $\langle x, \text{list of } P(x, t_j)\text{'s} \rangle \mapsto \text{list of } \langle \text{frequent 1-itemset } \{x\}, \text{expSup}(\{x\}) \rangle$.

where $\text{expSup}(\{x\}) = \text{sum of } P(x, t_j) \text{ in the list for } x$. Notice that, when handling precise data, the actual support of $\{x\}$ is its frequency. In contrast, when handling uncertain data, the expected support of $\{x\}$ may not be the same as its frequency. For instance, consider an item b with existential probability of 0.9 that appears only in transaction t_1 . Its expected support may be higher than item c that appears seven times but with an existential probability of 0.1 in each appearance. Then, $\text{expSup}(\{b\}) = 0.9 > 0.7 = \text{expSup}(\{c\})$.

Afterwards, MR-growth rereads the big databases to form an $\{x\}$ -projected database (i.e., a collection of transactions containing x) for each item x in the list produced from the first reduce function (i.e., for each frequent 1-itemset x). The worker node corresponding to each projected database then:

1. Builds appropriate local UF-trees (based on the projected database assigned to the node) to mine frequent k -itemsets (for $k \geq 2$), and
2. Outputs $\langle X, \text{expSup}(X) \rangle$ (which represents a frequent k -itemset X and its expected support) if $\text{expSup}(X) \geq \text{minsup}$.

To summarize, MR-growth executes the second set of “map” and “reduce” functions as follows:

- **Map:** $\langle \text{ID of transaction } t_j \in P_i, \text{ contents of } t_j \rangle \mapsto \text{list of } \langle \{x\}, \{x\}\text{-projected database} \rangle$, and
- **Reduce:** $\langle \{x\}, \{x\}\text{-projected database} \rangle \mapsto \text{list of } \langle \text{frequent itemset } X, \text{expSup}(X) \rangle$.

Constraint-Based MapReduce Algorithm: BigAnt and BigSAM

When mining and analyzing high-volume big data, it is not uncommon that users may be interested in only some subsets of these big data in many real-life situations. This leads to *constraint-based mining* (Bonchi, 2009; Leung, 2009), with which users can focus the mining on certain subsets of the big data by freely specifying some constraints to express their interest. For example, users can express their interests in finding a collection X of meteorological records having a total rainfall of less than 10mm by specifying a constraint “ $\text{sum}(X.\text{Rainfall}) < 10\text{mm}$ ”. Similarly users can also express their interests in finding a collection Y of geographic records with a minimum GPS coordinate (latitude) of 45°N by specifying a constraint “ $\text{min}(Y.\text{GPSCoordinate}) = 45^\circ\text{N}$ ”.

To handle these constraints and to focus the mining on those subsets of big data that are interesting to users, Leung & Jiang (2014) proposed the BigAnt algorithm to handle the user-specified anti-monotonic constraints by exploring the anti-monotonicity of the constraints, which states that “if a pattern satisfies an anti-monotonic constraint, then so do all its subsets”. In other words, if a pattern violates the anti-monotonic constraints, it can be pruned as any of its supersets is guaranteed to violate the constraints. For instance, if $\text{sum}(Z.\text{Rainfall}) \geq 10\text{mm}$ for some pattern Z , then $\text{sum}(Z'.\text{Rainfall}) \geq 10\text{mm}$ for any superset Z' of Z . So, the BigAnt algorithm also uses two sets of the “map” and “reduce” functions. The master node reads and divides uncertain data into partitions. The worker node corresponding to each partition

P_i then performs constraint checking and outputs $\langle \{x\}, P(x, t_j) \rangle$ for every *valid* item x in transaction $t_j \in P_i$ (when $\{x\}$ satisfies the constraints):

- **Map:** $\langle \text{ID of transaction } t_j \in \text{partition } P_i, \text{ contents of } t_j \rangle \mapsto \text{list of } \langle \text{valid item } x \in t_j, P(x, t_j) \rangle$.

Then, the reduce function sums all the $P(x, t_j)$'s for each valid x to compute its expected support $\text{expSup}(\{x\})$, and outputs $\langle \{x\}, \text{expSup}(\{x\}) \rangle$ (representing a valid frequent 1-itemset $\{x\}$ and its expected support) if $\text{expSup}(\{x\}) \geq \text{minsup}$:

- **Reduce:** $\langle \text{valid } x, \text{list of } P(x, t_j)\text{'s} \rangle \mapsto \text{list of } \langle \text{valid frequent 1-itemset } \{x\}, \text{expSup}(\{x\}) \rangle$,

where $\text{expSup}(\{x\}) = \text{sum of } P(x, t_j) \text{ in the list for } x$. Afterwards, BigAnt rereads the big databases to form an $\{x\}$ -projected database (i.e., a collection of transactions containing x) for each *valid* item x in the list produced from the first reduce function (i.e., for each valid frequent 1-itemset x). The worker node corresponding to each projected database then:

1. Builds appropriate local UF-trees (based on the projected database assigned to the node) to mine frequent k -itemsets (for $k \geq 2$),
2. Performs constraint checking to find valid ones from these mined frequent k -itemsets, and
3. Outputs $\langle \text{valid } X, \text{expSup}(X) \rangle$ (which represents a valid frequent k -itemset X and its expected support) if $\text{expSup}(X) \geq \text{minsup}$ and X is valid with respect to the user-specified constraints.

To summarize, BigAnt executes the second set of “map” and “reduce” functions as follows:

- **Map:** $\langle \text{ID of transaction } t_j \in P_i, \text{ contents of } t_j \rangle \mapsto \text{list of } \langle \text{valid } \{x\}, \{x\}\text{-projected database} \rangle$, and

- **Reduce:** $\langle \text{valid } \{x\}, \{x\}\text{-projected database} \rangle \mapsto \text{list of } \langle \text{valid frequent itemset } X, \text{expSup}(X) \rangle$.

Note that BigAnt pushes the user-specified anti-monotonic constraints into the big data mining process so that it directly discovers frequent patterns that satisfy the constraints.

To a further extent, Jiang et al. (2014) proposed the BigSAM algorithm to explore additional property of some anti-monotonic constraints. Specifically, BigSAM explores the succinctness of the constraints, which reveals that all and only those itemsets satisfying the succinct anti-monotonic (SAM) constraints can be explicitly and precisely generated using only individual items that satisfy the SAM constraints. For instance, any collection Y of geographic records with $\min(Y.GPSCoordinate) = 45^\circ\text{N}$ must consist of only records with GPS coordinate of 45°N . Hence, the BigSAM only needs to perform constraint checking in the first—but not the second—map function.

FUTURE RESEARCH DIRECTIONS

Due to the popularity of Web-based communities and social networking sites, high volumes of social media big data are available. For instance, as of March 2016, there were:

- 1.65 billion monthly active Facebook users, and 1.51 billion mobile monthly active Facebook users;
- 310 million monthly active Twitter users, and 1 billion unique visits monthly to Twitter with embedded tweets; and
- More than 433 million registered LinkedIn members worldwide, out of which more than 12 million registered Canadian members.

Embedded in these big data are rich sets of meaningful knowledge about the social networks. Hence, a logical future research direction is to

apply *social media mining and social network analysis* (Leung, Medina, & Tanbeer, 2013; Xu & Li, 2013; Jiang & Leung, 2014; Jiang, Kawagoe, & Leung, 2015; Leung et al., 2016) to social media data for discovery of rich sets of meaningful knowledge from these big data.

Many existing big data mining algorithms return the mined results in a textual form—e.g., a *textual list* of all (constrained or unconstrained) frequent patterns. As “a picture is worth a thousand words”, visual representation of the mining results is usually easier to comprehend for users than its equivalent textual representation. Thus, it is desirable to show the mining results interactively by applying the concepts of visual analytics. As such, to enhance user experience in exploring big data, the second future research direction is to incorporate *visual analytics and interactive technologies* (Zhang, Segall, & Cao, 2011; Leung, Carmichael, Johnstone, & Yuen, 2013) into big data mining so that the (constrained or unconstrained) frequent patterns mined from Big data are returned to the users in visual forms.

CONCLUSION

Big data analysis and mining aims to discover implicit, previously unknown, and potentially useful information and knowledge from big databases that contain high volumes of valuable veracious data collected or generated at a high velocity from a wide variety of data sources. Among different big data mining tasks, this chapter focuses on *big data analysis and mining for frequent patterns*. By relying on the MapReduce programming model, researchers only need to specify the “map” and “reduce” functions to discover frequent patterns from:

1. Big databases of precise data in a breadth-first manner (e.g., by using the SPC, FPC, and DPC algorithms) or in a depth-first manner (e.g., by using the PFP algorithm), and/or

2. Big databases of uncertain data (e.g., by using the MR-growth algorithm).

Such a big data analysis and mining process can be sped up (e.g., by using the BigAnt and BigSAM algorithms, which focus the mining according to the user-specified constraints that express the user interests). The resulting (constrained or unconstrained) frequent patterns mined from big databases provide users with new insights and a sound understanding of users’ patterns. Such knowledge is useful in many real-life information science and technology applications.

REFERENCES

- Agrawal, R., Imieliński, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. In *Proceedings of ACM SIGMOD 1993*, (pp. 207-216). doi:10.1145/170035.170072
- Agrawal, R., & Shafer, J. C. (1996). Parallel mining of association rules. *IEEE Transactions on Knowledge and Data Engineering*, 8(6), 962–969. doi:10.1109/69.553164
- Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules in large databases. In *Proceedings of VLDB, 1994*, (pp. 487–499).
- Ahmed, A. U., Ahmed, C. F., Samiullah, M., Adnan, N., & Leung, C. K. (2016). Mining interesting patterns from uncertain databases. *Information Sciences*, 354, 60–85. doi:10.1016/j.ins.2016.03.007
- Bechini, A., Marcelloni, F., & Segatori, A. (2016). A MapReduce solution for associative classification of big data. *Information Science*, 332, 33–55. doi:10.1016/j.ins.2015.10.041
- Bonchi, F. (2009). Constraint-based pattern discovery. In J. Wang (Ed.), *Encyclopedia of data warehousing and mining* (2nd ed.; pp. 313–319). IGI Global. doi:10.4018/978-1-60566-010-3.ch050

- Ceglar, A. & Roddick, J.F. (2006) Association mining. *ACM Computing Surveys*, 38(2), art. 5. doi: 10.1145/1132956.1132958
- Dean, J., & Ghemawat, S. (2004) MapReduce: Simplified data processing on large clusters. In *Proceedings of OSDI 2004* (pp. 137-150). USE-NIX Association.
- Dean, J., & Ghemawat, S. (2010). MapReduce: A flexible data processing tool. *Communications of the ACM*, 53(1), 72–77. doi:10.1145/1629175.1629198
- Farahat, A. K., Elgohary, A., Ghodsi, A., & Kamel, M. S. (2013). Distributed column subset selection on MapReduce. In *Proceedings of IEEE ICDM, 2013*, (pp. 171–180). doi: 10.1109/ICDM.2013.155
- Han, J., Pei, J., & Yin, Y. (2000) Mining frequent patterns without candidate generation. In *Proceedings of ACM SIGMOD 2000*, 1-12. doi:10.1145/342009.335372
- Hipp, J., Guntzer, U., & Nakhaeizadeh, G. (2000). Algorithms for association rule mining – a general survey and comparison. *SIGKDD Explorations*, 2(1), 58–64. doi:10.1145/360402.360421
- Jiang, F., Kawagoe, K., & Leung, C. K.-S. (2015). Big social network mining for “following” patterns. In *Proceedings of C3S2E 2015* (pp. 28–37). ACM. doi: 10.1145/2790798.2790802
- Jiang, F., & Leung, C. K.-S. (2014). A business intelligence solution for frequent pattern mining on social networks. In *Proceedings of IEEE ICDM Workshops, 2014*, (pp. 789–796). doi: 10.1109/ICDMW.2014.128
- Jiang, F., & Leung, C. K.-S. (2015). A data analytic algorithm for managing, querying, and processing uncertain big data in cloud environments. *Algorithms*, 8(4), 1175–1194. doi:10.3390/a8041175
- Jiang, F., Leung, C. K.-S., & MacKinnon, R. K. (2014) BigSAM: mining interesting patterns from probabilistic databases of uncertain big data. In *Proceedings of PAKDD Workshops 2014* (pp. 780-792). Springer. doi:10.1007/978-3-319-13186-3_70
- Lee, S., & Kim, J. (2016). Performance evaluation of MRDataCube for data cube computation algorithm using MapReduce. In *Proceedings of BigComp, 2016*, (pp. 325–328). doi: 10.1109/BIGCOMP.2016.7425939
- Leung, C. K.-S. (2009). Constraint-based association rule mining. In J. Wang (Ed.), *Encyclopedia of data warehousing and mining* (2nd ed.; pp. 307–312). IGI Global. doi:10.4018/978-1-60566-010-3.ch049
- Leung, C. K.-S. (2013). Mining frequent itemsets from probabilistic datasets. In *Proceedings of EDB, 2013*, (pp. 137–148).
- Leung, C. K.-S. (2014). Uncertain frequent pattern mining. In C. C. Aggarwal & J. Han (Eds.), *Frequent pattern mining* (pp. 339–367). Springer. doi: 10.1007/978-3-319-07821-2_14
- Leung, C. K.-S. (2015). Big data mining applications and services. In *Proceedings of BigDAS, 2015*, (pp. 1–8). doi:10.1145/2837060.2837076
- Leung, C. K.-S., Carmichael, C. L., Johnstone, P., & Yuen, D. S. H.-C. (2013). Interactive visual analytics of databases and frequent sets. *International Journal of Information Retrieval Research*, 3(4), 120–140. doi:10.4018/ijirr.2013100107
- Leung, C. K.-S., Cuzzocrea, A., & Jiang, F. (2013) Discovering frequent patterns from uncertain data streams with time-fading and landmark models. *LNCS Transactions on Large-Scale Data- and Knowledge-Centered Systems*, 8, 174-196. doi: 10.1007/978-3-642-37574-3_8

- Leung, C. K.-S., & Hayduk, Y. (2013) Mining frequent patterns from uncertain data with MapReduce for big data analytics. In *Proceedings of DASFAA 2013, Part I* (pp. 440-455). Springer. doi:10.1007/978-3-642-37487-6_33
- Leung, C. K.-S., & Jiang, F. (2014). A data science solution for mining interesting patterns from uncertain big data. In *Proceedings of IEEE BDCloud, 2014*, (pp. 235-242). doi: 10.1109/BDCloud.2014.136
- Leung, C. K.-S., Jiang, F., Pazdor, A. G. M., & Peddle, A. M. (2016). Parallel social network mining for interesting 'following' patterns. *Concurrency and Computation: Practice & Experience*, 28(15), 3994-4012. doi: 10.1002/cpe.3773
- Leung, C. K.-S., MacKinnon, R. K., & Syed, K. Tanbeer (2014) Fast algorithms for frequent itemset mining from uncertain data. In *Proceedings of IEEE ICDM 2014*, 893-898. doi:10.1109/ICDM.2014.146
- Leung, C. K.-S., Mateo, M. A. F., & Brajczuk, D. A. (2008) A tree-based approach for frequent pattern mining from uncertain data. In *Proceedings of PAKDD 2008 (LNCS/LNAI, 5012)* (pp. 653-661). Springer. doi:10.1007/978-3-540-68125-0_61
- Leung, C. K.-S., Medina, I. J. M., & Tanbeer, S. K. (2013). Analyzing social networks to mine important friends. In G. Xu & L. Li (Eds.), *Social media mining and social network analysis: Emerging research* (pp. 90-104). IGI Global. doi:10.4018/978-1-4666-2806-9.ch006
- Li, H., Wang, Y., Zhang, D., Zhang, M., & Chang, E. Y. (2008) PFP: Parallel FP-growth for query recommendation. In *Proceedings of ACM RecSys 2008*, (pp. 107-114). doi:10.1145/1454008.1454027
- Liao, C., & Squicciarini, A. C. (2015) Towards provenance-based anomaly detection in MapReduce. In *Proceedings of IEEE/ACM CCGrid 2015* (pp. 647-656). IEEE. doi:10.1109/CCGrid.2015.16
- Lin, J., & Dyer, C. (2010). Data-intensive text processing with MapReduce. Morgan & Claypool Publishers. doi: 10.2200/S00274ED-1V01Y201006HLT007
- Lin, M.-Y., Lee, P.-Y., & Hsueh, S.-C. (2012). Apriori-based frequent itemset mining algorithms on MapReduce. In *Proceedings of ICUIMC, 2012*, 76. doi: 10.1145/2184751.2184842
- Madden, S. (2012). From databases to big data. *IEEE Internet Computing*, 16(3), 4-6. doi:10.1109/MIC.2012.50
- Rajaraman, A., & Ullman, J. D. (2011). *Mining of massive datasets*. Cambridge University Press. doi:10.1017/CBO9781139058452
- Savasere, A., Omiecinski, E., & Navathe, S. (1995). An efficient algorithm for mining association rules in large databases. In *Proceedings of VLDB, 1995*, (pp. 432-444).
- Shahrivari, S., & Jalili, S. (2016). Single-pass and linear-time *k*-means clustering based on MapReduce. *Information Systems*, 60, 1-12. doi:10.1016/j.is.2016.02.007
- Shim, K. (2012). MapReduce algorithms for big data analysis. *PVLDB*, 5(12), 2016-2017. doi: 10.14778/2367502.2367563
- Tong, Y., Chen, L., Cheng, Y., & Yu, P. S. (2012). Mining frequent itemsets over uncertain databases. *PVLDB*, 5(11), 1650-1661. doi: 10.14778/2350229.2350277
- Ullman, J. D. (2000) A survey of association-rule mining. In *Proceedings of DS 2000 (LNCS/LNAI, 1967)* (pp. 1-14). Springer.
- Xu, G., & Li, L. (Eds.). (2013). *Social media mining and social network analysis: Emerging research*. IGI Global. doi:10.4018/978-1-4666-2806-9
- Zaki, M. J. (1999). Parallel and distributed association mining: A survey. *IEEE Concurrency*, 7(4), 14-25. doi:10.1109/4434.806975

Zhang, Q., Segall, R. S., & Cao, M. (Eds.). (2011). *Visual analytics and interactive technologies: Data, text and web mining applications*. IGI Global. doi:10.4018/978-1-60960-102-7

KEY TERMS AND DEFINITIONS

Anti-Monotonic Constraint: A constraint C such that, if an itemset S satisfying C , then any subset of S also satisfies C .

Big Data: High-velocity, high-value, and/or high-variety data with volumes beyond the ability of commonly-used software to capture, manage, and process within a tolerable elapsed time. These Big data necessitate new forms of processing to deliver high veracity (& low vulnerability) and to enable enhanced decision making, insight, knowledge discovery, and process optimization.

Data Mining: Non-trivial extraction of implicit, previously unknown and potentially useful information from data.

Frequent Pattern (Or Frequent Itemset): An itemset or a pattern with its actual support

(or expected support) exceeds or equals the user-specified minimum support threshold.

Frequent Pattern Mining: A search and analysis of high volumes of valuable data for implicit, previously unknown, and potentially useful patterns consisting of frequently co-occurring events or objects. It helps discover frequently co-located trade fairs and frequently purchased bundles of merchandise items.

Itemset: A set of items.

MapReduce: A high-level programming model, which uses the “map” and “reduce” functions, for processing high volumes of data.

Succinct Constraint: A constraint C such that all itemsets satisfying C can be expressed in terms of powersets of a fixed number of succinct sets using the set union and/or set difference operators. A succinct set is an itemset, in which items are selected from the domain using the usual SQL selection operator. In simple terms, a constraint C is succinct meaning that all and only those itemsets satisfying C can be explicitly and precisely generated using some precise “formula”.

Big Data Analytics for Tourism Destinations

B**Wolfram Höpken***University of Applied Sciences Ravensburg-Weingarten, Germany***Matthias Fuchs***Mid-Sweden University, Sweden***Maria Lexhagen***Mid-Sweden University, Sweden*

INTRODUCTION

Although information and communication technologies (ICT) were an important issue for Travel & Tourism (T&T) since the 1960ies (i.e. computer reservations systems, global distribution systems; Werthner & Klein, 1999), the difference today is that ICT has become a strategic issue for every tourism business (Buhalis, 2006). The special benefit tourism gains from ICT can be put down to the characteristics of the tourism product, being a service bundle ideally portrayed by electronic media and being jointly delivered by (usually) small-sized enterprises. Indeed, T&T is a highly information intensive sector, and not surprisingly, T&T represents the largest branch within the e-Commerce sector. In 2012, 45% of the EU online sales volume has been generated by the T&T sector, whereat in 2008 this figure stood only at 25%. Moreover, in the US already 51.5% of the total travel revenue is generated online (E-Marketer, 2012). However, although tourism shows high penetration rates with respect to web-based marketing & distribution, shortcomings become evident with respect to e-business networks (supply-chains) and integrated process automation (e-procurement, enterprise resource planning, etc.). Finally, significant adoption gaps are ascertained for ICTs in tourism SMEs to support market research, product development and strategic decision making (e-Business Watch, 2006).

The attractiveness of tourism destinations particularly depends on how communication and information needs of tourism stakeholders can be satisfied through ICT-based infrastructures so that sustainable knowledge sources can emerge (Buhalis, 2006). Although huge amounts of customer-based data are widespread in tourism destinations (e.g. web-servers store tourists' website navigation, databases save transaction and survey data, respectively), these valuable knowledge sources typically remain unused (Pyo, 2005). However, managerial effectiveness and organisational learning could be significantly enhanced by applying methods of *business intelligence* (BI) and *big data analytics* (Wong et al., 2006; Shaw & Williams 2009), offering reliable, up-to-date and strategically relevant information, such as tourists' travel motives and service expectations, information needs, channel use and related conversion rates, occupancy trends, quality of service experience and added value per guest segment (Min et al., 2002; Pyo et al., 2002). This makes clear why ICT and methods of BI are playing a crucial role in effectuating a *knowledge destination* by enhancing large-scale intra and inter-firm knowledge exchange. Indeed, the major challenge of knowledge management for tourism destinations is to make individual knowledge about customers, products, processes, competitors or business partners available and meaningful to others.

DOI: 10.4018/978-1-5225-2255-3.ch031

The objective of this chapter is to address the above deficiencies in tourism by presenting the concept of the tourism knowledge destination – a specific knowledge management architecture that supports value creation through enhanced supplier interaction and decision making. Information from heterogeneous data sources categorized into explicit feedback (e.g. tourist surveys, user ratings) and implicit information traces (navigation, transaction and tracking data) is extracted by applying semantic mapping, wrappers or text mining (Lau et al., 2005). Extracted data are stored in a central data warehouse enabling a destination-wide and all-stakeholder-encompassing data analysis approach. By using machine learning techniques interesting patterns are detected and knowledge is generated in the form of validated models (e.g. decision trees, neural networks, association rules, clustering models). These models, together with the underlying data (in the case of exploratory data analysis) are interactively visualized and made accessible to destination stakeholders. The technical architecture and implementation issues are discussed based on a prototypical implementation for the leading Swedish tourism destination, Åre (Höpken et al., 2015).

BACKGROUND

Since the widespread adoption of computerized reservation and booking systems in the 1980ies, comprehensive databases are available for all types of tourism transactions, i.e. the complete booking and consumption behavior (e.g. Passenger Name Record (PNR) databases of global distribution systems (GDS) or the airline on-time performance database of the Bureau of Transportation Statistics; BTS, 2012). Immediately, especially airline companies started to analyze such data as input to process and product optimization. A first prominent example in the area of revenue and yield management is the DINAMO system, introduced by American Airlines in 1988 (Smith et al., 1992). Further early examples can be found

for demand forecasting (Hueglin & Vannotti, 2001), prediction of cancellation or no-show behavior (Subramanian et al., 1999), or customer segmentation (Min et al., 2002).

Only very recently, data mining (DM) became increasingly important for tourism branches, due to its ability to discover previously unknown patterns in huge databases through explorative techniques and - compared to most statistical methods - to also identify non-linear relationships (Fuchs & Höpken, 2009; Fuchs et al., 2010; Höpken et al., 2011). Although, the potential of DM is not fully used in tourism yet, all major DM techniques are principally applied. More precisely, descriptive data analysis is widely used in form of reports or online analytical processing (OLAP) to visualize tourism arrivals depending on dimensions, like time/season, travel type or customer origin (TourMIS; Wöber, 1998; Destinometer; Fuchs & Weiermair, 2004). Methods of supervised learning, like classification, estimation and prediction are used to explain tourists' booking/cancellation or consumption behavior (Morales & Wang, 2008) and to predict tourism demand (Vlahogianni & Karlaftis, 2010). As a method of unsupervised learning, clustering is one of the most heavily used DM techniques in tourism, mostly applied to the task of customer segmentation as input to product differentiation, dynamic pricing or customer relationship management (Xia et al., 2010; Kuo et al., 2012).

With the uptake of the World Wide Web and its tremendous adoption in tourism, the topic of web DM gained more and more attention. *Web content mining*, i.e. the analysis of content from online platforms and websites, first of all deals with the analysis of user generated content (UGC), i.e. tourists' feedback and comments in blogs or review platforms, which currently constitute one of the most intensively researched topics in tourism (Bronner & Hoog, 2011). Methods of text mining are applied to the tasks of feedback aggregation and opinion mining or sentiment detection, typically based on statistical or linguistic approaches (Gräbner et al., 2012; Schmunk et al., 2014).

Additionally, web content mining increasingly deals with the extraction of knowledge about tourism markets and offers (market and concurrence analysis) (Walchhofer et al., 2010). Finally, *web usage mining* is dealing with the analysis of tourists' behavior when using online platforms or websites. Although current applications typically focus on descriptive analyses, like number of clicks or sessions depending on dimensions, like time, origin of user or URL, also supervised and unsupervised learning techniques have been applied, like customer segmentation for website adaptation and product recommendation (Wallace et al., 2004; Pitman et al., 2010) or (sequential) association rule mining for click-stream analysis (Jiang & Gruenwald, 2006).

THE KNOWLEDGE DESTINATION ARCHITECTURE

Although methods of BI and DM are partly used in tourism, their full potential, based on the huge amount of available data, is not yet utilized by far. Particularly, on the level of entire destinations, a comprehensive approach, i.e. a systematic integration of all data available from different stakeholder-based sources as well as a stakeholder-encompassing analysis approach, is missing.

The knowledge destination architecture fills this gap and distinguishes between a *knowledge generation* and a *knowledge application layer* (Figure 1; Höpken et al., 2011). The *knowledge generation layer* extracts destination-specific knowledge from structured and unstructured data sources, explicitly or implicitly generated by tourists (customer-based knowledge generation) or other destination stakeholders (supplier-based knowledge generation). For instance, content is generated by tourists through feedback mechanisms providing sources of standardized (e.g. surveys, evaluation platforms) and non-standardized data (e.g. blogs, review-platforms, etc.). Implicit knowledge can be made explicit by visualizing tourists' *information traces* (web-search, naviga-

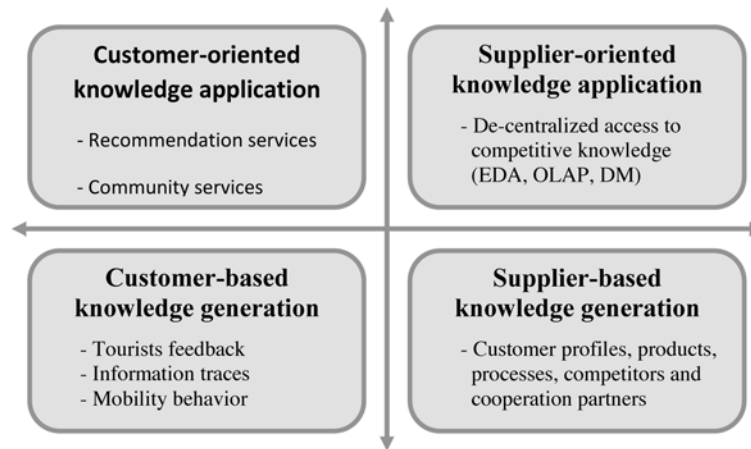
tion behavior) through online application tracking and web-mining (Liu, 2008). Moreover, knowledge about tourists' buying behavior can be generated through mining transaction (booking) data. Finally, tourists' *mobility* behavior may be traced by GPS/WLAN-based position tracking (Zanker et al., 2009). Additionally, destination stakeholders generate knowledge about customers, products, processes, competitors and cooperation partners, extractable from existing data sources or websites, e.g. in the form of destination profiles, or availability or capacity information (supplier-based knowledge generation in Figure 1; Pyo, 2005).

The *knowledge application* layer provides knowledge-based services for customers as well as destination suppliers and stakeholders. *Customer-oriented knowledge application* comprises recommendation services or community services typically used to learn from prior consumption experiences of other customers (Xiang & Gretzel, 2010). Within the *supplier-oriented knowledge application* explorative data analyses (EDA), OLAP as well as DM (predictive analytics, classification, clustering and association rules) allow the de-centralized management and access of competitive knowledge by the destination management organization (DMO) as well as private and public destination suppliers. Crucial management functions that support organizational learning and strategic decision making with a concern to product innovation and destination development are provided, such as benchmarking, business performance management, forecasting, multi-channel management, dynamic pricing and online community management (blog aggregator, alert functions, etc.) (Pyo et al., 2002).

The Tourism Data Warehouse Model

Core of the knowledge destination architecture and the prerequisite for destination-wide, stakeholder-encompassing analyses is a central *data warehouse* (Cho & Leung, 2002). Beside organizational constraints and privacy concerns, this first development step often fails due to the

Figure 1. The knowledge destination



heterogeneity of available data and the absence of a homogeneous data model. Although data modeling approaches and design patterns for business areas, like retail/sales, procurement, accounting, human resources, E-commerce, etc., already exist, there is no overall and general data warehouse model available for the tourism domain in general and for tourism destinations in specific (Kasavana & Knutson, 1999). Thus, agreeing on a common understanding of relevant business indicators, its underlying business processes and their characteristics and defining an appropriate and expressive destination data warehouse model is a crucial step towards powerful decision support

in a knowledge destination. The first activity of defining a destination data warehouse model is to identify most relevant business indicators which fulfill the requirements for a powerful decision support. Based on a literature review (Pyo et al., 2002; Cho & Leung, 2002; Pyo, 2005; Wang & Russo, 2007) and input from stakeholders of the Swedish tourism destination, Åre, relevant business indicators are defined and summarized in Table 1.

Generally, two concurrent approaches exist for developing a data warehouse model: while the *dimensional modeling* approach (Kimball, 2008) offers the advantage of a simple, straight-forward

Table 1. Business indicators

1 Economic performance	bookings, overnights, prices, occupancy, sales
2 Customer behavior	2A <i>Website navigation & search</i> : referrer sites, sessions, page views, search terms 2B <i>Booking and consumption behavior</i> : booking channels, conversion rates, days to arrival, length of stay, cancellations, guest tracking 2C <i>Customer profile</i> : country of origin, age, gender, skiing travel behavior, customer life time value, average/last length of stay, preferred type of accommodation, preferred mode of accommodation booking, preferred mode of transportation, purpose of visit
3 Customer perception and experience	3A <i>Destination brand awareness</i> : destination brand visibility, destination knowledge, information sources 3B <i>Destination value areas (functional & emotional)</i> : skiing & non-skiing winter activities, summer activities and attractions, services and features, atmosphere, social interaction 3C <i>Value for money and customer satisfaction</i> : functional value, emotional value, satisfaction 3D <i>Loyalty</i> : cognitive / affective loyalty, conative loyalty

Table 2. The tourism data warehouse bus matrix

Business Process	Indicators	Fact Table Type	Dimensions														
			Connectivity	Marketing	Survey	Referral	Session	Event	URI	Var Feedback	Feedback	Location	Channel	Supplier	Vendor	Product	Customer Demographic Profile
Booking	1, 2B, 2C	AS							x								
Stay	1, 2B, 2C	T										x					
Consumption	1, 2B, 2C	T										x		x		x	
Location Tracking	2B	T										x					
Info Request	2A	T											x			x	
Web Navigation	2A	T							x					x		x	
Feedback	3A-D	T							x		x	x	x	x	x	x	
Capacity	I	PS															
Marketing Activity	I	T															

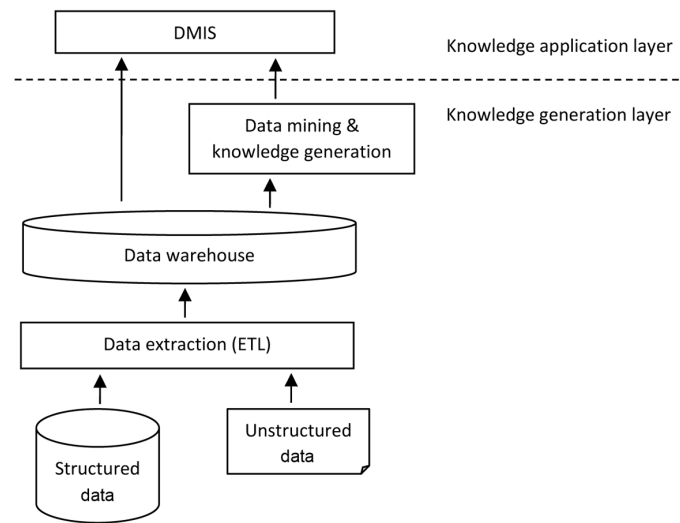
database design, supporting powerful analyses and a relatively good understandability by end users, the *fully normalized modeling* approach (Inmon, 2002) offers the advantage of a relatively better support of data integration, as redundancies are avoided and the identification of inconsistencies is simplified (especially if inconsistencies should be solved in underlying operational databases as well). In case of a tourism destination, the focus is not that much on reducing inconsistencies (i.e. especially not across different operational systems, due to the relative independence of stakeholders compared to departments of a single company), but rather on enabling powerful analyses on data of the whole destination. Thus, the approach of dimensional modeling seems most appropriate for defining a tourism destination data warehouse model. Table 2 provides an overview of the destination data warehouse model (bus matrix), listing identified business processes, necessary to support the defined indicators, the type of the corresponding fact table (transaction (T) = one entry per transaction, periodic snapshot (PS) = one entry per each period in time, accumulated snapshot (AS) = one entry covering all phases of a process) and, finally, all dimensions linked to a process (Kimball, 2008). Following the fundamental concept of *conformed dimensions*, most dimensions in a tourism data warehouse are used by several processes and, thus, such processes share the same physical dimension tables in the data warehouse. Conformed dimensions interlink the corresponding processes and enable cross-process analyses, for instance identifying relationships between customers' booking/consumption, web navigation and feedback behavior, respectively.

Technical Architecture and Implementation Issues

Figure 2 sketches the overall architecture of the knowledge destination. The *knowledge generation layer* comprises structured and unstructured data sources, the process of extracting relevant data, transforming source data into a homogeneous data

Figure 2. Knowledge destination framework architecture

Source: Höpken et al., 2011, p. 420



format appropriate for further analyses and storing/loading the data into the data warehouse (ETL), the data warehouse as central and homogeneous destination data store and, finally, DM and knowledge generation. The *knowledge application layer* comprises a destination management information system (DMIS) as central application offering data visualization and DM functionalities to destination stakeholders and managers.

Data Extraction (ETL)

Extraction, transformation and loading represents the process of extracting data from different data sources (e.g. operational databases, CRM systems, web-server log-files, etc.), transforming data into a format appropriate for visualization or DM activities and loading or storing data into a database (typically a data warehouse). The most important requirement related to *data extraction* in the context of tourism destinations is the support of all possible data sources and data formats, including structured data, like text files, databases or application-specific formats (e.g. SPSS files, excel files, etc.) or unstructured data, like html files or free text (typically converted into structured data by means of wrappers and text mining).

With regard to *data transformation*, in the context of the knowledge destination, data migration is most critical, i.e. value domain transformations, combination/separation of data elements and structure mapping, and record linkage (identification of duplicate entries, e.g. customers from different source systems, due to the heterogeneity and independence of operational systems of different destination stakeholders). Additionally, a powerful ETL process has to support different execution chronologies (e.g. periodically, on-demand or event-driven) and different incremental update strategies. In the course of the prototypical implementation of the knowledge destination for the Swedish destination Åre, RapidMiner and the Rapid Analytics BI server have been chosen, due to its full range of support of all components of the knowledge destination architecture, especially DM activities.

Data Warehouse

The knowledge destination data warehouse builds on a multi-dimensional data modeling (MDM) approach (Kimball et al., 2008; Höpken et al., 2013). In the course of the prototypical implementation of the knowledge destination, the Dimensional

Fact Model (DFM) has been chosen as graphical modeling notation, due to its good understandability (Golfarelli & Rizzi, 1998). Moreover, simple spreadsheets have been used to define the detailed dimensional model, due to insufficient tool support for DFM. Nevertheless, UML extensions and CWM (Common Warehouse Metamodel) are reasonable alternatives.

Data Mining

Due to the need of expert supervision and often long execution time for certain DM processes, not all processes can be executed by end users autonomously. Thus, DM processes are either executed online, triggered by the end user within the DMIS, or offline, i.e. triggered by data changes caused by corresponding ETL processes. Several open source tools for DM currently exist. For instance, RapidMiner is a powerful DM tool and together with the rapid analytics BI server offers a full-functional framework for executing DM processes in a client-server environment. More precisely, RapidMiner includes all WEKA DM methods and due to its support of web DM, text mining and time series analysis it seems most appropriate for implementing the knowledge destination architecture. However, an important limiting factor of DM tools in general and RapidMiner in specific is the performance concerning execution time and main memory consumption. Depending on available hardware, this especially limits the number of DM processes, which can be executed online.

DMIS

The Destination Management Information System (DMIS) represents the interface of the knowledge destination architecture to the end user. DMIS enables to access the data and knowledge stored in the central data warehouse and to execute data analyses as a mean of decision support. Available functions and analyses can be differentiated into reporting, i.e. offering predefined reports, OLAP

and DM. Specific frameworks for developing MIS or cockpit solutions, e.g. RapidMiner Enterprise Edition (www.rapid-i.com) or Pentaho BI, have been analyzed and compared with developing a specific web application from scratch. Integrating specific functionalities, especially in the area of adaptive decision support based on meta-learning, is an important requirement in the context of the knowledge destination architecture in order to reach a high level of usability and understandability. Thus, the prototypical implementation for the destination of Åre makes use of a self-developed web application, based on the Ruby on Rails open-source web development framework (rubyonrails.org).

Decision Support by Interactive Visualization and Data Mining

Supporting destination and stakeholder management decision making takes place by explorative data analysis (EDA) and OLAP on the one hand and more complex DM techniques on the other hand. Exemplarily for EDA, OLAP and DM, a description of the framework for UGC extraction and analysis, which is integrated in DMIS-Åre, will follow next.

Online reviews regarding the Swedish destination Åre are gained from social media platforms *TripAdvisor.com* and *Booking.com* and text-processing techniques are applied to automatically extract single statements from each of the reviews. Subsequently, by using machine-learning techniques and a dictionary-based approach, these statements are classified as either 'positive' experiences, 'negative' experiences or 'neutral' (Liu, 2011). Moreover, statements are classified into product areas, like 'FoodBreakfast', 'Rooms', and 'Service-Personnel'. From the proportion of classified positive and classified negative statements, an average feedback value (normalized between 0-1) is computed. By following Hippner and Rentzmann (2006), the framework for UGC extraction and analysis comprises five steps (Schmunk et al., 2014):

1. **Document Selection:** Relevant review sites are identified as input to the mining process for the sentiment analysis regarding hotels in the destination Åre. The review sites *Booking.com* (10 hotels and 248 reviews) and *TripAdvisor.com* (17 hotels and 1,193 reviews) were selected. To collect relevant pages from review sites, a web crawler was used, fetching html pages and following contained links based on regular expressions specified for each site (ibid, 2014, p. 256).
2. **Document (Pre-)Processing:** The pre-processing of collected html documents includes four steps: a) extraction of relevant opinion texts from html-code based on regular expressions, b) removal of empty reviews, c) filtering of English texts, since sentiment analysis is language dependent (e.g. wordlists in the case of the dictionary-based approach or stop word removal and stemming in the case of machine-learning approaches),¹ and d) generation of single statements based on sentence end characters as delimiters (Schmunk et al., 2014, p. 257). Pre-processing of *TripAdvisor.com* delivers 127 reviews and 1,296 usable single statements, while from *Booking.com* 81 reviews and 220 statements are gained (total no. of reviews = 208; total no. of usable statements = 1,516).
3. **Mining:** Mining includes the sub-tasks recognition of properties, subjectivity, and sentiment, respectively. For machine learning methods training data were created for all classification tasks (property, subjectivity and sentiment). Sentences are tokenized, stop words removed, words reduced to their stem, and finally, a word vector is created based on TF-IDF scores. Subsequently, the machine learning methods Naïve Bayes, Support Vector Machines (SVM) and k-nearest neighbor (k-NN) are applied. For the dictionary-based approach, a word list for each class (property, subjectivity and sentiment) is employed. In this case, the class of a sentence is directly deduced by the wordlist to which the majority of contained words belong to (ibid., 2014, p. 258).
4. **Evaluation:** A 10-fold cross-validation is used to evaluate the machine-learning models, while the dictionary-based approach is evaluated by comparing the results with pre-classified test data. For the recognition of properties SVM gained the highest accuracy (i.e. percentage of correctly classified statements within test data) of 72.35%. The highest accuracy of 82.63% for subjectivity recognition was achieved by the dictionary-based method. The best result for sentiment recognition is gained again by the SVM method, showing an accuracy of 76.80% (ibid, 2014, p. 260).
5. **Usage:** The outcome of the sentiment analysis provides valuable information on customer reviews and opinions in a structured format (ibid, 2014, p. 262). Thus, data extracted from the review sites (e.g. review date, hotel name, demographic data of reviewer, and the review text itself) is enriched by the product properties that the opinion is linked to and its orientation (sentiment) and finally stored in the multi-dimensional destination data warehouse of DMIS-Åre (Höpken et al., 2013). Thus, social media knowledge is made available for powerful OLAP analyses as discussed below.

The DMIS-Åre dashboard in Figure 3 displays the customers' (i.e. guest) profile directly deduced from social media content, such as travel purpose, travel companions, number of UGC-based statements by origin countries, etc.

The dashboard in Figure 4 displays the 'average feedback' value extracted and calculated from UGC as described above, grouped by hotel 'product areas' and major hotels in Åre. Hotel and destination managers can, thus, monitor the performance in terms of customer feedback (UGC) in total, or grouped by hotel product area. Moreover, UGC-based performance metrics can

Figure 3. DMIS-Åre dashboard: guest profile extracted from social media sites



be benchmarked among various accommodation suppliers (Fuchs & Weiermair, 2004).

As outlined in the section about the *Tourism Data Warehouse Model*, a final powerful feature is to join so far disconnected and separately filed knowledge areas through common (conformed) dimensions across different business processes (Höpken et al., 2013). This condition, namely to relate customer feedback to web-navigation and booking behavior is considered as crucial for enhanced organizational learning and creativity processes in destinations (Fuchs et al., 2014, p. 201). More concretely, this procedure provides analyses across the three prototypically implemented processes *Web Navigation* (site clicks, sessions), *Booking* (total bookings, booking price, number of persons/booking, average time between booking and arrival, average stay duration/booking) and *Feedback* (total feedback answers, average feedback value) (Höpken et al., 2015). As mentioned, the average feedback value is generated not only through UGC extracted from various social media platforms. Rather, the multi-dimensional structures of the data warehouse are populated also by feedback data from destination brand equity surveys (Chekalina et al. 2014), real-time feedback from guests during their stay provided by an electronic customer registration and survey tool (e-CRST) accessible via Quick-Response codes (Höpken et al., 2012), and, finally, data from customer surveys conducted by various destination accommodation suppliers.

Thus, the following cross-process analyses link customer feedback from different sources with data from the web-navigation and booking process, respectively. For instance, for each country of customer origin, Figure 5 lists key performance indicators (KPIs) across the booking process, the web navigation process and the feedback process (Schmunk et al., 2014, p. 262). For instance, and interestingly enough, the cross-process analysis reveals that Finnish tourists show a relatively high

Figure 4. DMIS-Åre dashboard: UGC-based feedback value grouped by Åre hotels and product area



average feedback value (0.812), pay a relatively high average booking-price per overnight (7,852 SEK) and are characterized by a relatively long average duration of stay (6.16 days). In contrast, tourists from Belgium show a relatively low average feedback value (0.702), pay a comparatively lower booking-price per overnight (5,355 SEK), and show a relatively short average duration of stay (5.36 days).

FUTURE RESEARCH DIRECTIONS

Integration of Data Mining and OLAP

Visualizations of DM models, such as decision trees, regression models or cluster models, are complex and typically displayed by DM tools in a form which is usually difficult to understand for managers. Thus, such models are vanished manually and presented through static reports with no direct integration of OLAP functionalities. Thus,

Figure 5. Key performance indicators (KPIs) gained through DMIS-Åre's cross process analysis

Group by attribute	Total bookings	Total clicks	Total feedback, answers	Total sessions	Average booking price in SEK	Average number of persons per booking	Average time between booking and arrival in days	Average stay duration per booking	Average feedback value
Australia	33	165	156	56	4029.485	2.606	49.100	4.818	0.859
Belgium	33	1207	130	361	5355.424	2.970	112	5.364	0.702
China	14	456	71	230	3087.250	3.500	164.500	8.786	0.859
Denmark	1830	2392	107	521	6235.034	4.527	115.609	6.144	0.778
Estonia	1080	161	16	53	4837.395	5.292	85.108	5.838	0.784
Finland	3155	4543	1039	893	7852.046	4.386	95.536	6.160	0.812
France	25	2923	12	739	4616	3.040	38.318	5.042	0.715
Germany	105	3268	150	665	5307.040	3.155	124.426	6.644	0.837
Netherlands	279	2168	150	490	7894.760	3.802	104.493	6.588	0.750
Norway	10625	27636	9093	5889	3901.213	4.911	71.890	3.297	0.767
Sweden	56073	162942	36880	39139	5023.956	3.860	83.014	5.498	0.773
United Kingdom	1042	48513	303	12843	4929.136	3.009	59.399	6.786	0.765
United States of America	30	4284	16	2307	4719.167	4.036	39.148	4.133	0.539

managers get no hints for DM models relevant in the context of a specific OLAP analysis, and DM models are not related to other information available within the data warehouse. However, to overcome these limitations, DM models have to be integrated directly into the dimensional structures of the data warehouse and existing OLAP functionalities can be used for their interactive visualization and relating to other information. Cluster models, for example, can be stored in the data warehouse by a specific cluster dimension linked to the corresponding fact table, if the clustering is based on several dimensions, or by a cluster membership attribute within a dimension, if the clustering is based solely on one dimension (e.g. customer segmentation). Moreover, association rules can be stored as a new fact table, but reusing existing dimensions, like product or customer. Thus, future research has to analyze how DM models can be optimally integrated into the data warehouse in order to enable a powerful integration of DM and OLAP analyses (Mayer et al., 2015).

Meta Learning and Adaptive MIS

Typically, managers are overwhelmed with the question, which DM method is best suited for the analysis at hand. Additionally, in most situations

an optimal result cannot be reached in one single step, but rather an iterative and interactive process is needed, what, in turn, typically requires guidance based on domain-specific experience. Such guidance in selecting appropriate DM methods and parameters or the best approach for improving and refining DM results, can be provided to managers through concepts of meta-learning. Following the approach of an adaptive recommender system, the system adapts available choices or even the overall user dialog and offers recommendations for DM methods, parameters or optimization steps (Ricci et al., 2011). An interesting future research questions is, how existing approaches from meta-learning and recommender systems can be combined and applied to the concrete context of an adaptive MIS within a tourism knowledge destination.

CONCLUSION

This chapter presented a BI architecture to apply DM techniques to the tourism domain and more concrete to tourism destinations. The proposed *knowledge destination architecture* supports knowledge creation and enhanced decision making in tourism destinations. As core component of the knowledge destination, a multi-dimensional

data warehouse model has been presented which considers all business processes and corresponding dimensions most relevant to any tourism destination, and, thus, serves as reference model for similar implementations in a tourism destination context. Subsequently, UGC-based analyses from the area of OLAP and DM have been presented. Moreover, crucial technical aspects of the BI architecture related to the area of data extraction (ETL), data warehousing, DM and a destination management information system (DMIS) as user interface have been critically discussed and concepts and techniques suitable to the tourism knowledge destination have been presented. Finally, veins of future research have been sketched. The concept of the tourism knowledge destination has been prototypically implemented and validated within the leading Swedish tourism destination Åre. Not only has the concept proven to be viable but also the strategic importance of this system for destination stakeholders was proved through a remarkable high final user acceptance. Especially the flexible integration of customer feedback in form of survey results into the data warehouse and DMIS appeared to be a powerful and promising approach to knowledge creation in the tourism domain.

REFERENCES

- Bronner, F., & Hoog, R. (2011). Vacationers and eWOM: Who Posts, and Why, Where, and What? *Journal of Travel Research*, 50(1), 15–26. doi:10.1177/0047287509355324
- BTS. (2012). *Transportation On-Time Performance Database*. Bureau of Transportation Statistics. Retrieved July 19, 2012, from <http://www.transtats.bts.gov/>
- Buhalis, D. (2006). The impact of ICT on tourism competition. In A. Paptheodorou (Ed.), *Corporate rivalry and market power: competition issues in the tourism industry* (pp. 143–171). London: IB Tauris.
- Chekalina, T., Fuchs, M., & Lexhagen, M. (2014). A-Value Creation Perspective on the Customer-based Brand Equity Model for Tourism Destinations – A Case from Sweden. *Finnish Journal of Tourism Research*, 10(1), 7–23.
- Cho, V., & Leung, P. (2002). Knowledge Discovery Techniques in Database Marketing for the Tourism Industry. *Journal of Quality Assurance in Hospitality & Tourism*, 3(3), 109–131. doi:10.1300/J162v03n03_07
- E-Marketer. (2012). *Online and Total Travel Sales Worldwide*. Retrieved from <http://b-i.forbesimg.com/greatspeculations/files/2013/03/1460919.gif>
- eBusinessWatch. (2006). *The European e-Business Market*. EC, Industry DG, Unit 4, Brussels.
- Fuchs, M., & Höpken, W. (2009). Data Mining in Tourism (In German: „Data Mining im Tourismus“). *Praxis der Wirtschaftsinformatik*, 270(12), 73–81. doi:10.1007/BF03340421
- Fuchs, M., Höpken, W., Föger, A., & Kunz, M. (2010). E-Business Readiness, Intensity, and Impact – An Austrian Destination Management Organization Study. *Journal of Travel Research*, 49(2), 165–178. doi:10.1177/0047287509336469
- Fuchs, M., Höpken, W., & Lexhagen, M. (2014). Big Data Analytics for Knowledge Generation in Tourism Destinations – A Case from Sweden. *Journal of Destination Marketing and Management*, 3(4), 198–209. doi:10.1016/j.jdmm.2014.08.002
- Fuchs, M., & Weiermair, K. (2004). Destination Benchmarking - An Indicator Systems Potential for Exploring Guest Satisfaction. *Journal of Travel Research*, 42(3), 212–225. doi:10.1177/0047287503258827
- Golfarelli, D. M., & Rizzi, S. (1998). *The Dimensional Fact Model: A Conceptual Model for Data Warehouses*. Bologna: University of Bologna.

- Gräbner, D., Zanker, M., Fliedl, G., & Fuchs, M. (2012). Classification of customer reviews based on sentiment analysis. In *Information and Communication Technologies in Tourism*. Springer. doi:10.1007/978-3-7091-1142-0_40
- Harren, A. (2001). *Multidimensional Modeling Language (MML) und Multidimensional UML (MUML)*. Data-Warehouse-Systeme: Architektur, Entwicklung, Anwendung. dpunkt Verlag: 163–167.
- Hippner, H., & Rentzmann, R. (2006). *Text Mining*. Ingolstadt: Springer-Verlag.
- Höpken, W., Deubele, Ph., Höll, G., Kuppe, J., Schorpp, D., Licones, R., & Fuchs, M. (2012). Digitalizing Loyalty Cards in Tourism. In M. Fuchs, F. Ricci, & L. Cantoni (Eds.), *Information and Communication Technologies in Tourism* (pp. 272–283). New York: Springer.
- Höpken, W., Fuchs, M., Höll, G., Keil, D., & Lexhagen, M. (2013). Multi-dimensional Data Modeling for a Destination Data Warehouse. In L. Cantoni & Ph. Xiang (Eds.), *Information and Communication Technologies in Tourism 2013* (pp. 157–169). New York: Springer. doi:10.1007/978-3-642-36309-2_14
- Höpken, W., Fuchs, M., Keil, D., & Lexhagen, M. (2011). The Knowledge Destination – A Customer Information-based Destination Management Information System. In R. Law, M. Fuchs, & F. Ricci (Eds.), *Information and Communication Technologies in Tourism* (pp. 417–429). New York: Springer. doi:10.1007/978-3-7091-0503-0_34
- Höpken, W., Fuchs, M., Keil, D., & Lexhagen, M. (2015). Business Intelligence for Cross-process Knowledge Extraction at Tourism Destination. *Journal of Information Technology and Tourism*, 15(2), 101–130. doi:10.1007/s40558-015-0023-2
- Hueglin, C., & Vannotti, F. (2001). *Data Mining Techniques to Improve Forecast Accuracy in Airline Business*. doi:10.1145/502512.502578
- Inmon, W. H. (2002). *Building the data warehouse*. New York: Wiley & Sons.
- Jiang, N., & Gruenwald, L. (2006). Research Issues in Data Stream Association Rule Mining. *SIGMOD*, 35(1), 14–19. doi:10.1145/1121995.1121998
- Kasavana, M. L., & Knutson, B. J. (1999). A primer on sftware: Warehousing, marting and mining hospitality data for more effective marketing decisions. *Journal of Hospitality & Leisure Marketing*, 6(1), 83–96. doi:10.1300/J150v06n01_06
- Kimball, R., Ross, M., Thronthwaite, W., Mundy, J., & Becker, B. (2008). *The Data Warehouse Lifecycle Toolkit: Practical Techniques for Building Data Warehouse and Business Intelligence Systems* (2nd ed.). Indianapolis, IN: Wiley.
- Kuo, R. J., Akbaria, K., & Subroto, B. (2012). Application of Particle Swarm Optimization and Perceptual Map to Tourist Market Segmentation. *Expert Systems with Applications*, 39(10), 8726–8735. doi:10.1016/j.eswa.2012.01.208
- Lau, K.-N., Lee, K.-H., & Ho, Y. (2005). Text Mining for the Hotel Industry. *The Cornell Hotel and Restaurant Administration Quarterly*, 46(August), 344–362. doi:10.1177/0010880405275966
- Liu, B. (2008). *Web data mining*. New York: Springer.
- Mayer, V., Höpken, W., & Fuchs, M. (2015). Integration of Data Mining Results into Multi-Dimensional Data Models. In I. Tussyadiah & A. Inversini (Eds.), *Information and Communication Technologies in Tourism 2015* (pp. 155–166). New York: Springer. doi:10.1007/978-3-319-14343-9_12
- Min, H., Min, H., & Emam, A. (2002). A DM approach to develop the profile of hotel customers. *Int. Journal of Contemporary Hospitality Management*, 14(6), 274–285. doi:10.1108/09596110210436814

- Morales, D. R., & Wang, J. (2008). Passenger Name Record Data Mining Based Cancellation Forecasting for Revenue Management. *Innovative Applications of O.R.*, 202(2), 554–562.
- Pitman, A., Zanker, M., Fuchs, M., & Lexhagen, M. (2010). Web usage mining in tourism – a query term analysis and clustering approach. In *Information and Communication Technologies in Tourism*. Springer.
- Pyo, S. (2005). Knowledge map for tourist destinations. *Tourism Management*, 26(4), 583–594. doi:10.1016/j.tourman.2004.03.001
- Pyo, S., Uysal, M., & Chang, H. (2002). Knowledge discovery in databases for tourist destinations. *Journal of Travel Research*, 40(4), 396–403. doi:10.1177/0047287502040004006
- Ricci, F., Rokach, L., Shapira, B., & Kantor, P. (Eds.). (2011). *Handbook on Recommender Systems*. New York: Springer. doi:10.1007/978-0-387-85820-3
- Schmunk, S., Höpken, W., Fuchs, M., & Lexhagen, M. (2014). Sentiment Analysis – Implementation and Evaluation of Methods for Sentiment Analysis with Rapid-Miner. In Ph. Xiang & I. Tussyadiah (Eds.), *Information and Communication Technologies in Tourism* (pp. 253–265). New York: Springer.
- Shaw, G., & Williams, A. (2009). Knowledge transfer and management in tourism organisations. *Tourism Management*, 30(3), 325–335. doi:10.1016/j.tourman.2008.02.023
- Smith, B. C., Leimkuhler, J. F., & Darrow, R. M. (1992). Yield Management at American Airlines. *Interfaces*, 22(1), 8–31. doi:10.1287/inte.22.1.8
- Subramanian, J., Stidham, S. Jr, & Lautenbacher, C. J. (1999). Airline Yield Management with Overbooking, Cancellations, and No-Shows. *Transportation Science*, 33(2), 147–167. doi:10.1287/trsc.33.2.147
- Vlahogianni, E. I., & Karlaftis, M. G. (2010). *Advanced Computational Approaches for Predicting Tourist Arrivals: The Case of Charter Air-Travel*. Nonlinear Dynamics.
- Walchhofer, N., Hronsky, M., Pöttler, M., Baumgartner, R., & Fröschl, K. A. (2010). Semantic online tourism market monitoring. In *Information and Communication Technologies in Tourism*. Springer. doi:10.1007/978-3-211-99407-8_52
- Wallace, M., Maglogiannis, I., Karpouzis, K., Kormentzas, G., & Kollias, S. (2004). Intelligent one-stop-shop travel recommendations using an adaptive neural network. *Information Technology & Tourism*, 6(3), 181–193. doi:10.3727/1098305031436971
- Wang, Y., & Russo, S. M. (2007). Conceptualizing and evaluating the Functions of Destination Marketing Systems. *Journal of Vacation Marketing*, 13(3), 187–203. doi:10.1177/1356766707077687
- Werthner, H., & Klein, S. (1999). *Information Technology and Tourism – A Challenging Relationship*. Wien: Springer-Verlag. doi:10.1007/978-3-7091-6363-4
- Wöber, K. W. (1998). Global Statistical Sources: TourMIS: An Adaptive Distributed Marketing Information System for Strategic Decision Support in National, Regional, or City Tourist Offices. *Pacific Tourism Review*, 2(3), 273–286.
- Wong, J.-Y., Chen, H.-J., Chung, P.-H., & Kao, N.-C. (2006). Identifying valuable Travellers by the Application of Data Mining. *Asia Pacific Journal of Tourism Research*, 11(4), 355–373. doi:10.1080/10941660600931200
- Xia, J., Evans, F. H., Spilsbury, K., Ciesielski, V., Arrowsmith, C., & Wright, G. (2010). Market segments based on the dominant movement patterns of tourists. *Tourism Management*, 31(4), 464–469. doi:10.1016/j.tourman.2009.04.013

Xiang, Z., & Gretzel, U. (2010). Role of Social Media in Online Travel Information Search. *Tourism Management*, 31(2), 179–188. doi:10.1016/j.tourman.2009.02.016

KEY TERMS AND DEFINITION

Adaptive Management Information System:

A management information system adapting its user interface and interaction strategy depending on user preferences and past user behavior and satisfaction.

Computer Reservation Systems (CRS) / Global Distribution Systems (GDS): Computer systems providing information like prices and availabilities for a wide range of tourism products (e.g. hotels, flights, car-rental, etc.) and supporting the full booking, settlement and after-sales processes.

Customer-Based Data: Data provided by customers either intentionally, like demographic data, reviews and comments, or data provided unintentionally, like data on web-navigation, booking or consumption behavior.

Data Warehouse Bus Matrix: Visualization of business processes, corresponding facts and dimensions for a multi-dimensional data warehouse model.

Destination Management Information System (DMIS): A management information system specifically designed to enable improved decision support for the destination management organization and other stakeholders of a tourism destination.

Destination Management Organization (DMO): Organization which coordinates the many constituent elements of the tourism prod-

uct; provides visitor services and the necessary information structure to market the destination in a most democratic way to enhance residents' well-being.

Multi-Dimensional Data Modeling (MDM):

A modeling paradigm for data warehouse models building on a separation of measurements, called facts, and surrounded context, called dimensions.

Tourism Destination: Agglomeration of companies and organizations involved in producing and marketing the overall tourism product within a geographical area; strategic unit providing all necessary resources whose integrated activities allow tourists with the kind of experiences they expect.

Tourism Knowledge Destination: Novel concept of a tourism destination that supports knowledge creation, transfer and enhanced decision making among destination stakeholders by applying techniques from business intelligence and data mining.

Tourist Feedback: Feedback on tourism products, suppliers or whole destinations, provided by tourists in structured and unstructured ways, e.g. in the form of customer ratings, comments or product reviews.

ENDNOTE

- ¹ To filter English texts, supervised learning was used and a Naïve Bayes classifier was trained based on a sample of texts for each relevant language as training data reaching an accuracy of 100% (Schmunk et al., 2014).

Big Data Time Series Stream Data Segmentation Methods

Dima Alberg

Shamoon College of Engineering (SCE), Israel

INTRODUCTION

Big data time series data streams are ubiquitous in finance, meteorology and engineering. It may be impossible to process an entire “big data” continuous data stream or to scan through it multiple times due to its tremendous volume. In Heraclitus’s well-known saying, “*you never step in the same stream twice*,” and so it is with “big data” temporal data streams.

Unlike traditional data sets, big data continuous data streams flow into a computer system continuously, in a non-stationary way and with varying update rates. They are time-stamped, fast-changing, massive, and potentially infinite.

Under these circumstances, they represent an application area of growing importance in the data mining research. For example, sensors generate one million samples every minute therefore the primary purpose of time series data stream segmentation is dimensionality reduction. This technique is used in many areas of data stream mining as: frequent patterns finding, structural changes and concept drifts detection (Tabassum & Gama, 2016), time series classification and prediction (Hulten & Domingos, 2003), time series similarities searching (Mori, Mendiburu, Keogh & Lozano, 2016) etc. The main principle of segmentation algorithms concludes in reducing the big data time series dimensionality by dividing the time axis into intervals behaving approximately according to a simple model. A good big data time series data stream segmentation algorithm must be OFASC (Online, Fast, Accurate, Simple and Comparable). For example the Sliding Window algorithm (Keogh, Chu, Hart, & Pazzani, 2004) on

the one hand is online (O), very fast (F) and relatively simple (S) for using in online segmentation applications but on the other hand, it sometimes gives poor accuracy (A) and does not allow to perform online multivariate segmentation (C). Therefore, we will classify this algorithm to OFS segmentation algorithms domain.

The segmentation problem can be defined in following way: first, given a time series data stream to produce the best representation such that the maximum error for any segment does not exceed some user specified confidence level error threshold. It is important to add, that using a relative parameter such as confidence level will allow to evaluate an online multivariate segmentation and second, to construct a user friendly segmentation application which will evaluate and compare the proposed online segmentation algorithms in real time. As we shall see in later sections, the state-of-the-art segmentation algorithms do not meet all these requirements.

The rest of the paper is organized as follows. In Section 2, we provide a literature review of three state-of-the-art online piecewise linear segmentation algorithms. In Section 3, we provide a methodology for improving the existing state-of-the-art online segmentation algorithms. The proposed methodology based on novel bound error estimation, which uses a relative probability parameter instead of maximum error nominal parameter and meets the proposed OFASC requirements. Section 4 briefly demonstrates a real-time segmentation application. Finally, in Section 5 and 6 we provide brief and meaningful empirical comparison of the proposed algorithms and suggest final conclusions.

DOI: 10.4018/978-1-5225-2255-3.ch032

BACKGROUND

Several high level representations of time series have been proposed in the research literature, including Fourier Transforms (Keogh et al., 2000), Wavelets (Chan & Fu, 1999), Symbolic Mappings (Das, Lin, Mannila, Renganathan, & Smyth, 1998; Perng et al., 2000) and Piecewise Linear Approximation or PLA: (Chan & Fu, 1999; Ge & Smyth, 1999; Hunter & McIntosh, 1998; Junker, Amft, Lukowicz, & Tröster, 2008; Keogh et al., 2004; Lavrenko, Schmill, Lawrie, Ogilvie, Jensen, & Allan, 2000; Li, Yu, & Castelli, 1998; Osaki, Shimada, & Uehara, 1999; Park, Lee, & Chu, 1999; Qu, Wang, & Wang, 1998; Shatkay & Zdonik, 1996; Vullings, Verhaegen, & Verbruggen, 1997; Wang & Wang, 2000).

In this work, our attention will confine to PLA, perhaps the most frequently used representation in continuous time series data streams. Obviously, all piecewise linear segmentation algorithms can also be classified as batch or online (Vullings et al., 1997). The problem discussed by (Keogh et al., 2004) is actually how to build online, fast and accurate algorithm for piecewise linear segmentation of time series data stream, because on the one hand, the main problem of online Sliding Window algorithm (Keogh et al., 2004) concerns in its poor accuracy (Qu et al., 1998; Wang & Wang, 2000) and its inability to look ahead. On the other hand the offline accurate Bottom Up (Keogh et al., 2004) algorithm is impractical or may even be unfeasible in a data mining context, where the data are in the order of terabytes or arrive in continuous streams. This problem is very important because for scalability purposes the proposed piecewise linear segmentation algorithm needs to capture the online nature of sliding windows and yet retain the superiority of Bottom Up.

In 2004 Keogh et al. (Keogh et al., 2004) introduced online Sliding Window Bottom Up (SWAB) algorithm which scales linearly with the size of the dataset, requires only constant space, produces high quality approximations of the initial time series data, and can be seen as operating on

a continuum between the two extremes of Sliding Windows and Bottom-Up. The authors have shown that the most popular Sliding Window approach generally produces very poor results, and that while the second most popular approach, Top-Down, can produce reasonable results, it does not scale well with massive time series stream data.

MAIN FOCUS

As indicated in (Keogh et al., 2004), the main problem with the Sliding Windows algorithm is its inability to look ahead, lacking the global view of its offline (batch) counterparts. The Bottom-Up and the Top-Down (Junker et al., 2008; Keogh et al., 2004) approaches produce better results, but are offline and require the scanning of the entire data set. The SWAB algorithm has three nominal input parameters, which need to be defined carefully by the user in order to obtain an accurate segmentation model. For example, often the user obstructs to determine for the value of the maximal error threshold, because the data has very noisy non-stationary behavior. Therefore, in order to produce an accurate segmentation model, the user needs to perform the preprocessing of the obtained data or to perform a time consuming experiment design. Second, the inner loop of the SWAB algorithm simply invokes the Bottom-Up algorithm each time. This results in some computation redundancy and increases the computational complexity of algorithm. In this paper we introduce two new algorithms ISW (Interval Sliding Window) and ISWAB (Interval Sliding Window and Bottom Up) which decreases computational redundancy and complexity of SWAB (Sliding Window and Bottom Up) (Keogh et al., 2004) algorithm.

The performance of the Sliding Window and SWAB algorithms depends on the value of maximal error. As maximal error goes to zero the Sliding Window and SWAB algorithms have the same performance, since they would produce multiple short segments with no error. At the opposite end, as the maximal error becomes very

large, the algorithms once again will all have the same performance, since they will simply approximate a data stream with a single best-fit line. Keogh et al. (Keogh et al., 2004) recommend test the relative performance for some “reasonable value” of maximal error, a value that achieves a good tradeoff between compression and fidelity. Because this “reasonable value” is subjective and dependent on the data mining application and the data itself, they did the following. First, they chose a “reasonable value” of maximal error for each dataset and then bracketed it with 6 values separated by powers of two. The lowest of these values tends to produce an over-fragmented approximation, and the highest tends to produce a very coarse approximation. Second, they chose performance in the mid-range of the 6 values which by their opinion should be considered most important. Obviously, the maximal error calculation routine proposed by Keogh et al. (Keogh et al., 2004) is very subjective and requires multi pass computational efforts. In the proposed ISWAB algorithm it is possible to derive the maximal error by using Hoeffding bound and one pass calculation. In fact, a similar approach was used in the VFDT decision tree induction algorithm introduced in (Hulten & Domingos, 2003).

Suppose we have a segment A with range R_A and n_A observations and segment B with range R_B and n_B observations which belong to a sliding window S of time series T . Assume that $\bar{x}_A$ and $\bar{x}_B$ are the sample means of segments A and B respectively, then the new created segment AB will have range R_{AB} , $n_A + n_B$ observations and sample mean $\bar{x}_{AB}$. Finally, the Hoeffding bound states that the true mean of the merged segment AB lies with confidence level of $1 - \delta$ in the interval where

$$\epsilon = \sqrt{R_{AB}^2 \frac{\ln(1/\delta)}{2(n_A + n_B)}}. \quad (1)$$

According to Motwani and Raghavan (1998) the Hoeffding bound is independent of the distribution generating the examples. This bound is

applicable to all situations where observations are independent and generated by a stationary distribution. Important to note, that Hoeffding bound is additive, its error is absolute and it does not require calculation of expected means of two merged segments. It is easy to show, that when a confidence level $= 1$, the Hoeffding error value is equal to zero meaning that the observed and the segmented time series data streams are the same.

In case of Sliding Window algorithm the proposed technique ISW works in following manner. Each time a new observation arrives the algorithm calculates the Hoeffding bound using Equation 1 and a user defined confidence level, then in case the new calculated error is greater than the previously calculated Hoeffding bound, the algorithm starts a new sliding window, otherwise it continues with the current sliding window. This incremental technique on the one hand is more sensitive to data stream concept drift changes and on the other hand allows to create relatively large segments when the data stream is stable and therefore to decrease significantly the running time of the proposed algorithm. The pseudocode for the ISW algorithm is shown in the Table 1.

The following numerical example briefly explains main calculation procedure of the ISW algorithm. Suppose that the current sliding window segment includes four observations: 1, 2, 3 and 4 (Table 2). Now, a new, fifth observation arrives and the ISW algorithm checks whether to start a new segment or to continue updating the previous one. The current segment range equals to 0.8 (4.8-4.0), the number of observations is 4 and with the user specified confidence level of 95% the value of Hoeffding bound equals to 0.06. Now, with the aid of the new, fifth observation we will recalculate the linear interpolation model error. The new model error equals to 0.025. This error is lower to previously calculated Hoeffding bound (0.06), thus the algorithm will increase the current segment.

The main similarity of the SWAB and ISWAB algorithms consists in the pairwise linear bottom up segmentation approach. This approach assumes

Table 1. The ISW online algorithm

Input: Data stream, DS Confidence Level,
Output: ISW Segmented data stream.
<pre> anchor = 1; while not finished segmenting time series i = 2; *Bound and Error Calculation while Model_error < Hoeffding_Bound Model_Error(Segment[anchor: anchor + i]) Hoeffding_Bound(Segment[anchor],) i = i + 1; New_Segment = Create_Segment(T[anchor: anchor + (i-1)]) Segment = Merge(Segment, New_Segment) anchor = anchor + i; </pre>

Table 2. Observations for ISW numerical example

n	1	2	3	4	5 (new)
Obs. Value	4.1	4	4.4	4.8	4.7
Segment	1	1	1	1	2

creating the finest possible linear interpolation of the time series, so that $n/2$ segments are used to approximate the n -length time series. Differently from the SWAB algorithm the proposed ISWAB algorithm chooses between the best linear

interpolation and regression models. Finally, the pseudocode for the proposed ISWAB algorithm is shown in Table 3.

The following numerical example briefly explains main calculation procedure of our presented

Table 3. The ISWAB online algorithm

Input: Data stream, DS Confidence Level, Sliding Window size, SW
Output: ISWAB Segmented data stream.
<pre> Create the Sliding Window buffer of size SW For (i = 1, length(SW), 2) Segment = Concatenate(SW[i: i + 1]); *Bound and Error Calculation For (i = 1, length(Segments.Count) - 1) Bound(i) = Hoeffding_Bound(Segment[i], Segment[i+1],) Error(i) = Model_Error(Segment[i], Segment[i+1]) *Merging the Segments while (min(error(k)) < Bound(k)) index = min(error(k)); Segment(index) = Merge(Segment(index), Segment(index+1)) delete(Segment (index+1)) Error(index) = Model_Error (Segment[i], Segment[i+1]) Bound(index) = Hoeffding_Bound(Segment[i], Segment[i+1],) Error(index-1) = Model_Error (Segment[i-1], Segment[i]) Bound(index-1) = Hoeffding_Bound(Segment[i-1], Segment[i],) Shift the Sliding Window to the Right Remove the Left Segment from the Sliding Buffer </pre>

Table 4. Observations for ISWAB numerical example

n	1	2	3	4	5	6	7
Obs. Value	0.4	0.1	4.4	5.6	6.7	8.8	9.9
Segment	1	1	2	2	↔ ₂	3	3

ISWAB approach. Suppose that a sliding window contains 3 segments with lengths of two, three, and two observations respectively. Table 4 shows the numerical values of seven collected observations.

The main question is which segments to merge: the first with the second or the second with the third? If we merge the first two segments then the number of observations in the merged segment equals to 5 and its range is 6.6, otherwise if we merge the second and the third segment we obtain the same number of observations and the range of 5.5. Supposing that our selected confidence level is 95%, we will calculate the Hoeffding bounds for the two proposed scenarios. For the first scenario according to the formula 1 we have Hoeffding bound value ε of 0.47 and for second the error value is 0.39. In the opposite case, the linear interpolation model error for the first scenario equals to 0.89 and it will be described by the following linear interpolation equation, whereas the second scenario's linear interpolation model has error of 0.078 and the following linear interpolation equation. Choosing the minimal model error value of 0.078 which is lower than the previously corresponded Hoeffding bound value 0.39 the ISWAB algorithm we will perform merging of segments 2 and 3.

The proposed techniques have the following advantages. First it is more convenient to define a confidence level as input rather than the maximum error threshold. Second our techniques allow to compare ISW and ISWAB algorithms across different time series and data streams mining domains, since they require relative confidence level parameter as input. Third the main advantage of the proposed algorithm is that it encourages merging of big segments. Suppose that we have 3 segments A , B and C with the same range but

with different lengths (measured by the number of examples), then it is easy to show that when $n_A + n_B > n_B + n_C$ the our proposed algorithm tends to merge bigger segments.

$$\epsilon_{A+B} = \sqrt{R_{AB}^2 \frac{\ln(1/\delta)}{2(n_A + n_B)}} < \sqrt{R_{BC}^2 \frac{\ln(1/\delta)}{2(n_B + n_C)}} = \epsilon_{B+C} \quad (2)$$

EXPERIMENTAL RESULTS

Our experimental study is aimed at estimating the accuracy and comparing the performance of the proposed algorithms. The first part was focused on stationary time series data streams (TSDS) and the second one was focused on non stationary data streams. The proposed algorithms were implemented using C# programming language and tested on a PC Intel Core2Duo at 3 GHz with 2GB RAM and 256 GB hard disk.

The stationary data streams were generated from two synthetically distributed normally distributed time series ND25 and ND100 whereas the non-stationary data streams were obtained from two Israel's daily financial indexes TA25 and TA100 (www.finance.yahoo.com). These finance indexes behavior strongly depends on time and therefore they demonstrate non stationary behavior. The few descriptive statistics for the four selected time series is shown in Table 5.

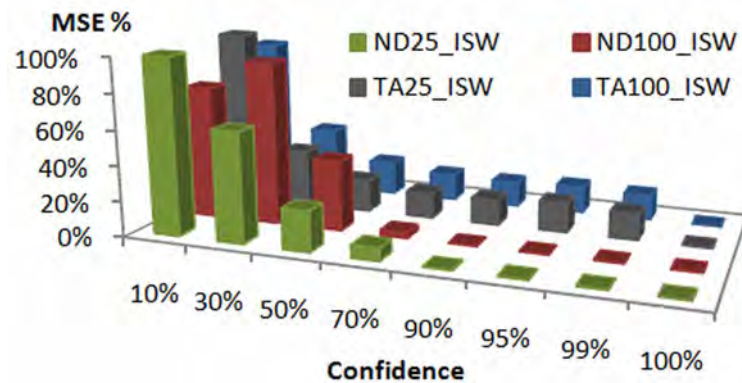
Parametrically, the ND25 time series is similar to TA25 because their averages, standard deviations and lengths are equal. Same thing is right

Table 5. Descriptive statistics

TSDS	N*	Max	Min	Avg.	St.Dev.
ND25	1,228	1,693.82	-29.21	748.06	244.32
ND100	1,228	1,426.14	-22.79	756.70	225.00
TA25	1,228	1,237.13	333.90	748.06	244.32
TA100	1,228	1,189.04	341.04	756.70	225.00

(*by N we denote the total number of observations with at least one minute interval commit frequency)

Figure 1. The ISW algorithm accuracy



regarding to TA100. Figure 1 demonstrates the ISW algorithm evaluation on the four collected time series. The blue columns point out the non-stationary data as financial indexes TA25 and TA100 and red stationary data e.g. similarly generated normally distributed processes ND25 and ND100. The most obvious result is that ISW produces more accurate results on stationary data (red columns) when the user specified confidence level is greater than 80%. In case of non-stationary data streams the ISW algorithm produces stable but less accurate results. This stable quality pattern results from the ISW algorithm ability to detect mean concept drifts in time series data stream behavior.

Figure 2 demonstrates the performance results of ISW algorithm. Actually, this figure Y axis shows the number of created segments when it is obvious that a large number of segments increases the evaluation time of algorithm and vice versa. As

previously mentioned the ISW algorithm produces good accuracy results for our collected stationary data. From Figure 2 we will conclude that when user specified confidence level greater than 80% the ISW and ISWAB algorithms produces similar quantity of segments (+300).

Now take our attention to ISWAB algorithm. On the one hand the ISWAB algorithm significantly outperforms the ISW algorithm and has good accuracy results for two sorts of collected data (Figure 3) when user specified confidence level is greater than 80%.

On the other hand, in case of non-stationary data the ISWAB algorithm creates large number of segments and this fact significantly increases the algorithm run time. In case of stationary data streams (red columns) the ISWAB algorithm creates small number of segments (from 40 to 50) and significantly outperforms the previously mentioned ISW algorithm.

Figure 2. The ISW algorithm performance

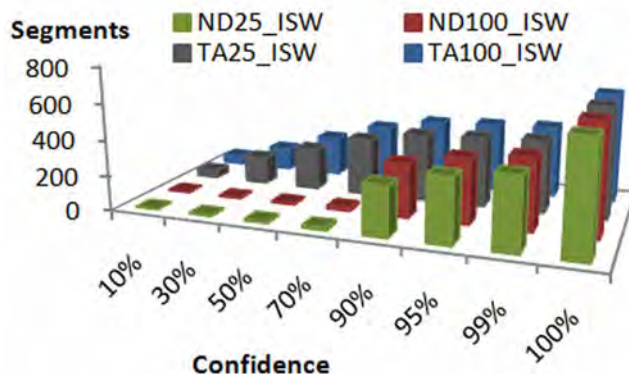
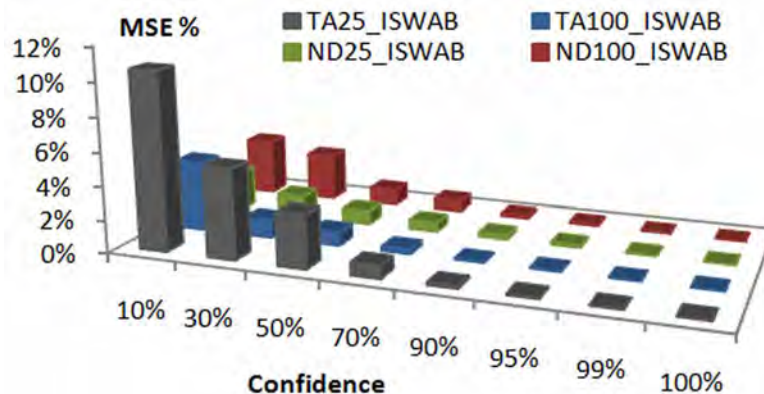


Figure 3. The ISWAB algorithm accuracy



FUTURE RESEARCH

Another promising direction for future research based on this work is the multi-sensor fusion area. One straightforward direction for further research in this case is to employ several sensors, thus extracting as much information as possible. In multi-sensor data fusion systems, these sensors can be used to measure the same quantities, which is especially helpful in case of sensor failure. Alternatively, different quantities associated with the same state of nature can be measured by different sensors. In all of these cases, observations are available from several sources and the information provided by each of the sources should be combined. We believe that the algorithms developed here can reduce uncer-

tainty and obtain a more complete knowledge in this area of research.

CONCLUSION

This study has highlighted a number of limitations in existing state-of-the-art online piecewise linear segmentation approaches: Sliding Window (SW) and SWAB. First, the new relative parameter of confidence level was used instead of nominal input parameter of maximal error threshold. This improvement has two advantages: first is that the user does not need to preprocess the original time series data stream in order to detect reasonable maximum error value and second is that the proposed technique allows to perform cross

comparisons between different time series data streams. Second, the implementation of new real time application was performed. Finally, was performed an empirical comparison of proposed time series segmentation algorithms on two types of time series data: stationary (normally distributed data) and non stationary (financial data) in order to check noise robustness of two proposed algorithms. The proposed ISWAB algorithm produces high quality approximations on all types of data and significantly decreases the number of created segments in case of stationary distributed data.

REFERENCES

- Chan, K., & Fu, W. (1999). Efficient time series matching by wavelets. *Proceedings of the 15th IEEE International Conference on Data Engineering*, 126-133.
- Das, G., Lin, K., Mannila, H., Renganathan, G., & Smyth, P. (1998). Rule discovery from time series. *Proceedings of the 3rd International Conference of Knowledge Discovery and Data Mining*, 16-22.
- Ge, X., & Smyth P. (2001). Segmental semi-markov models for endpoint detection in plasma etching. *IEEE Transactions on Semiconductor Engineering*, 1-5.
- Hulten, P., & Domingos, G. (2003). A general framework for mining massive data streams. *Journal of Computational and Graphical Statistics*, 12(4), 945-949. doi:10.1198/1061860032544
- Hunter, J., & McIntosh, N. (1999). Knowledge-based event detection in complex time series data. *Artificial Intelligence in Medicine*, 271-280.
- Junker, H., Amft, O., Lukowicz, P., & Tröster, G. (2008). Gesture spotting with body-worn inertial sensors to detect user activities. *Pattern Recognition*, 41(6), 2010-2024. doi:10.1016/j.patcog.2007.11.016
- Keogh, E., Chakrabarti, K., & Pazzani, M., & Mehrotra (2000). Dimensionality reduction for fast similarity search in large time series databases. *Proceedings of the 22th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 263-286.
- Keogh, E., Chu, S., Hart, D., & Pazzani, M. (2004). Segmenting time series: A survey and novel approach. *Data Mining in Time Series Databases*, 1-21.
- Lavrenko, V., Schmill, M., Lawrie, D., Ogilvie, P., Jensen, D., & Allan, J. (2000). Mining of concurrent text and time series. *Proceedings of the 6th International Conference on Knowledge Discovery and Data Mining*, 37-44.
- Li, C., Yu, P., & Castelli, V. (1998). MALM: A framework for mining sequence database at multiple abstraction levels. *Proceedings of the 9th International Conference on Information and Knowledge Management*, 267-272. doi:10.1145/288627.288666
- Mori, U., Mendiburu, A., Keogh, E., & Lozano, J. (2016). Reliable early classification of time series based on discriminating the classes over time. *Data Mining and Knowledge Discovery*, 1-31.
- Motwani, R., & Raghavan, P. (1996). Randomized algorithms. *ACM Computing Surveys*, 28(1), 33-37. doi:10.1145/234313.234327
- Osaki, R., Shimada, M., & Uehara, K. (1999). Extraction of primitive motion for human motion recognition. *The 2nd International Conference on Discovery Science*, 351-352.
- Park, S., Kim, S., & Chu, W. (2000). Segment-based approach for subsequence searches in sequence databases. *Proceedings of the 16th ACM Symposium on Applied Computing*, 248-252.
- Park, S., Lee, D., & Chu, W. (1999). Fast retrieval of similar subsequences in long sequence database. *Proceedings of the 3rd IEEE Knowledge and Data Engineering Exchange Workshop*, 60-67.

Perng, C., Wang, H., Zhang, S., & Parker, S. (2000). Landmarks: A new model for similarity based pattern querying in time series databases. *Proceedings of 16th International Conference on Data Engineering*, 33-42. doi:10.1109/ICDE.2000.839385

Qu, Y., Wang, C., & Wang, S. (1998). Supporting fast search in time series for movement patterns in multiples scales. *Proceedings of the 7th International Conference on Information and Knowledge Management*, 251-258. doi:10.1145/288627.288664

Shatkay, H., & Zdonik, S. (1996). Approximate queries and representations for large data sequences. *Proceedings of the 12th IEEE International Conference on Data Engineering*, 546-553. doi:10.1109/ICDE.1996.492204

Tabassum, S., & Gama, J. (2016). Sampling Massive Streaming Call Graphs. *Proceedings of the 31th ACM/SIGAPP Symposium on Applied Computing*, 1.

Vullings, H., Verhaegen, M., & Verbruggen, H. B. (1997). ECG segmentation using time-warping. *Proceedings of the 2nd International Symposium on Intelligent Data Analysis*, 275-285.

Wang, C., & Wang, S. (2000). Supporting content-based searches on time series via approximation. *Proceedings of the 12th International Conference on Scientific and Statistical Database Management*, 69-81. doi:10.1109/SSDM.2000.869779

KEY TERMS AND DEFINITIONS

Big Data: Big data usually includes data sets with sizes beyond the ability of commonly-used software tools to capture, curate, manage, and process the data within a tolerable elapsed time.

Data Stream: Data stream is time series sequence which flows into a computer system continuously, in a non-stationary way and with varying update rates. Often, it may be impossible to store an entire data stream or to scan through it multiple times due to its tremendous volume.

Financial Index: A time dependent indicator used to measure and report value changes in a selected group of stocks.

Segmentation: Segmentation refers to the automatic process of partitioning a data stream into multiple segments by set of predefined features (e.g. temperature, electricity consumptions, workdays, etc.).

Sliding Window: Sliding (Rolling) window refers to looking at a subset of points in the time series rather than all previous points. This “window” consecutively rolls back, holding the same number of points within the window as it moves along the time series data streams.

Stationarity: A statistical characteristic of a time series for which the distribution does not change over time.

Time Series: A series of values of a quantity obtained at successive times, often with equal intervals between them.

Challenges for Big Data Security and Privacy

B

M. Govindarajan

Annamalai University, India

INTRODUCTION

Big data refers to collections of data sets with sizes outside the ability of commonly used software tools such as database management tools or traditional data processing applications to capture, manage, and analyze within an acceptable elapsed time. Big data sizes are constantly increasing, ranging from a few dozen terabytes in 2012 to today many petabytes of data in a single data set. Big data creates tremendous opportunity for the world economy both in the field of national security and also in areas ranging from marketing and credit risk analysis to medical research and urban planning. The extraordinary benefits of big data are lessened by concerns over privacy and data protection.

As big data expands the sources of data it can use, the trust worthiness of each data source needs to be verified and techniques should be explored in order to identify maliciously inserted data. Information security is becoming a big data analytics problem where massive amount of data will be correlated, analyzed and mined for meaningful patterns.

Security of big data can be enhanced by using the techniques of authentication, authorization, encryption and audit trails. There is always a possibility of occurrence of security violations by unintended, unauthorized access or inappropriate access by privileged users.

To protect privacy, two common approaches used are the following. One is to restrict access to the data by adding certification or access control to the data entries so sensitive information is accessible to a limited group of users only. The other

approach is to anonymize data fields such that sensitive information cannot be pinpointed to an individual record. For the first approach, common challenges are to design secured certification or access control mechanisms, such that no sensitive information can be misconduct by unauthorized individuals. For data anonymization, the main objective is to inject randomness into the data to ensure a number of privacy goals (Xindong Wu et al., 2014).

BACKGROUND

Today we are living in an era of digital world. With the rapid increase in digitization the amount of structured, semi structured and unstructured data being generated and stored is exploding. Usama Fayyad (2012) has presented amazing data numbers about internet usage like “every day 1 billion queries are there in Google, more than 250 million tweets are there in Twitter, more than 800 million updates are there in Face book, and more than 4 billion views are there in You tube”. Each day, 2.5 quintillion bytes of data are generated and 90 percent of the data in the world today were created within the past two years. The data produced nowadays is estimated in the order of zeta bytes, and it is growing around 40% every year. International Data Corporation (IDC) terms this as the “Digital Universe” and predicts that this digital universe is set to explode to an unimaginable 8 Zetabytes by the year 2015. The above examples demonstrate the rise of big data applications where data collection has grown tremendously and is beyond the ability of com-

monly used software tools to manage, capture, and process.

From a privacy and security perspective, the challenge is to ensure that data subjects (i.e., individuals) have sustainable control over their data, to prevent misuse and abuse by data controllers (i.e., big data holders and other third parties), while preserving data utility, i.e., the value of big data for knowledge/patterns discovery, innovation and economic growth.

Cloud protection alliance big data working group identify top protection and seclusion problems that need to confine for making the big data computing and infrastructure more secure. Most of these issues are linked to the big data storage and computation. There having some challenges which are related to secure data storage (Cloud Security Alliance White paper, 2012). Different security challenges related to data security and privacy are discussed in (A. A. Soofi et al., 2014) which include data breaches, data reliability, data accessibility and data support. Privacy is major concern in outsourced data. Recently, some controversies have revealed how some security agencies are using data generated by individuals for their own benefits without permission. Therefore, policies that cover all user privacy concerns should be developed. Furthermore, rule violators should be identified and user data should not be misused or leaked. The following sections describe some relevant challenges to security and privacy in the context of big data.

MAIN FOCUS

Challenges for Big Data Security and Privacy

With the proliferation of devices connected to the Internet and connected to each other, the volume of data collected, stored, and processed is increasing everyday, which also brings new challenges in terms of the information security. In fact, the currently used security mechanisms such as fire-

walls and DMZs cannot be used in the Big Data infrastructure because the security mechanisms should be stretched out of the perimeter of the organization's network to fulfill the user/data mobility requirements and the policies of BYOD (Bring Your Own Device). Considering these new scenarios, the pertinent question is what security and privacy policies and technologies are more adequate to fulfill the current top Big Data privacy and security demands (Cloud Security Alliance, 2013). These challenges may be organized into four Big Data aspects such as infrastructure security (e.g. secure distributed computations using MapReduce), data privacy (e.g. data mining that preserves privacy/granular access), data management (e.g. secure data provenance and storage) and, integrity and reactive security (e.g. real time monitoring of anomalies and attacks).

Considering Big Data there is a set of risk areas that need to be considered. These include the information lifecycle (provenance, ownership and classification of data), the data creation and collection process, and the lack of security procedures. Ultimately, the Big Data security objectives are no different from any other data types – to preserve its confidentiality, integrity and availability.

Being Big Data such an important and complex topic, it is almost natural that immense security and privacy challenges will arise (Michael & Miller, 2013; Tankard, 2012). Big Data has specific characteristics that affect information security: variety, volume, velocity, value, variability, and veracity. These challenges have a direct impact on the design of security solutions that are required to tackle all these characteristics and requirements (Demchenko, Ngo, Laat, Membrey, & Gordijenko, 2014). Currently, such out of the box security solution does not exist.

Cloud Secure Alliance (CSA), a non-profit organization with a mission to promote the use of best practices for providing security assurance within Cloud Computing, has created a Big Data Working Group that has focused on the major challenges to implement secure Big Data

services (Cloud Security Alliance, 2013). CSA has categorized the different security and privacy challenges into four different aspects of the Big Data ecosystem. These aspects are Infrastructure Security, Data Privacy, Data Management and, Integrity and Reactive Security. Each of these aspects faces the following security challenges, according to CSA:

INFRASTRUCTURE SECURITY

Secure Distributed Processing of Data

Distributed programming frameworks utilize parallel computation and storage to process massive amounts of data. For example, the MapReduce framework splits an input file into multiple chunks. In the first phase of MapReduce, a Mapper for each chunk reads the data, performs some computation, and outputs a list of key/value pairs. In the next phase, a Reducer combines the values belonging to each distinct key and outputs the result. There are two major attack prevention measures: securing the mappers and securing the data in the presence of an untrusted mapper.

Security Best Actions for Non-Relational Data-Bases

The security infrastructures of non-relational data stores popularized by NoSQL databases are still evolving (L. Okman et al., 2011). For instance, robust solutions to NoSQL injection are still not mature. Each NoSQL database was built to tackle different challenges posed by the analytics world, and security was never addressed during the design stage. Developers using NoSQL databases usually embed security in the middleware. NoSQL databases do not provide any support for explicitly enforcing security in the database. However, clustering aspects of NoSQL databases pose additional challenges to the robustness of such security practices.

DATA PRIVACY

Data Analysis through Data Mining Preserving Data Privacy

As described by Boyd and Crawford (2012), Big Data can potentially enable invasions of privacy, invasive marketing, decreased civil liberties, and increased state and corporate control.

A recent analysis of how companies are leveraging data analytics for marketing purposes included an example of how a retailer was able to identify a teen's pregnancy before her father learned of it (C. Duhigg, 2012). Similarly, anonymizing data for analytics is not enough to maintain user privacy. For example, AOL released anonymized search logs for academic purposes, but users were easily identified by their searches (M. Barabard et al., 2006). Netflix faced a similar problem when anonymized users in their data set were identified by correlating Netflix movie scores with IMDB scores.

Therefore, it is important to establish guidelines and recommendations for preventing inadvertent privacy disclosures.

Cryptographic Solutions for Data Security

There are two fundamentally different approaches to controlling the visibility of data to different entities, such as individuals, organizations and systems. The first approach controls the visibility of data by limiting access to the underlying system, such as the operating system or the hypervisor. The second approach encapsulates the data itself in a protective shell using cryptography. Both approaches have their benefits and detriments. Historically, the first approach has been simpler to implement and, when combined with cryptographically-protected communication, is the standard for the majority of computing and communication infrastructure.

However, the system-based approach arguably exposes a much larger attack surface. The

literature on system security is replete with attacks on the underlying systems to circumvent access control implementations (such as buffer overflow and privilege escalation) and access the data directly. On the other hand, protecting data end-to-end through encryption exposes a smaller, more well-defined attack surface. Although covert side-channel attacks (C. Percival, 2005; Aciğmez et al., 2006) are possible to extract secret keys, these attacks are far more difficult to mount and require sanitized environments.

Granular Access Control

The security property that matters from the perspective of access control is secrecy – preventing access to data by people that should not have access. The problem with coarse-grained access mechanisms is that data that could otherwise be shared is often swept into a more restrictive category to guarantee sound security. Granular access control gives data managers more precision when sharing data, without compromising secrecy.

DATA MANAGEMENT AND INTEGRITY

Secure Data Storage and Transaction Logs

Data and transaction logs are stored in multi-tiered storage media. Manually moving data between tiers gives the IT manager direct control over exactly what data is moved and when. However, as the size of data set continues to grow exponentially, scalability and availability have necessitated auto-tiering for Big Data storage management. Auto-tiering solutions do not keep track of where the data is stored, which poses new challenges to secure data storage. New mechanisms are imperative to thwart unauthorized access and maintain constant availability.

Granular Audits

With real-time security monitoring, notification at the moment an attack takes place is the goal. In reality, this will not always be the case (e.g., new attacks, missed true positives). In order to discover a missed attack, audit information is necessary. Audit information is crucial to understand what happened and what went wrong. It is also necessary due to compliance, regulation and forensic investigation. Auditing is not something new, but the scope and granularity might be different in real-time security contexts. For example, in these contexts there are more data objects, which are probably (but not necessarily) distributed.

Data Provenance

Provenance metadata will grow in complexity due to large provenance graphs generated from provenance-enabled programming environments in Big Data applications. Analysis of such large provenance graphs to detect metadata dependencies for security and/or confidentiality applications is computationally intensive.

REACTIVE SECURITY

End-to-End Filtering and Validation

Many Big Data uses in enterprise settings require data collection from a variety of sources, including end-point devices. For example, a security information and event management system (SIEM) may collect event logs from millions of hardware devices and software applications in an enterprise network. A key challenge in the data collection process is input validation: how can we trust the data? How can we validate that a source of input data is not malicious? And how can we filter malicious input from our collection? Input validation and filtering is a daunting challenge posed by untrusted input sources, especially with the bring-your-own-device (BYOD) model.

Supervising the Security Level in Real-Time

These security and privacy challenges cover the entire spectrum of the Big Data lifecycle: sources of data production (devices), the data itself, data processing, data storage, data transport and data usage on different devices.

SOLUTIONS AND RECOMMENDATIONS

There is no single magical solution to solve the identified Big Data security and privacy challenges and traditional security solutions, which are mainly dedicated to protect small amounts of static data, are not adequate to the novel requisites imposed by Big Data services (Cloud Security Alliance, 2013). There is the need to understand how the collection of large amounts of complex structured and unstructured data can be protected. Non-authorized access to that data to create new relations, combine different data sources and make it available to malicious users is a serious risk for Big Data. The basic and more common solution for this includes encrypting everything to make data secure regardless where the data resides (data center, computer, mobile device, or any other). As Big Data grows and its processing gets faster, then encryption, masking and tokenization are critical elements for protecting sensitive data.

The new Big Data security solutions should extend the secure perimeter from the enterprise to the public cloud (Juels & Oprea, 2013). In this way, a trustful data provenance mechanism should be also created across domains. In addition, similar mechanisms to the ones used in (Luo, Lin, Zhang, & Zukerman, 2013) can be used to mitigate distributed denial-of-service (DDoS) attacks launched against Big Data infrastructures. Also, a Big Data security and privacy is necessary to ensure data trustworthiness throughout the entire data lifecycle – from data collection to usage.

The personalization feature of some Big Data services and its impact on the user privacy is discussed in (Hasan, Habegger, Brunie, Bennani, & Damiani, 2013). They discuss these issues in the backdrop of EEXCESS, a concrete project aimed to both provide high level recommendations and to respect user privacy. A recent work describes proposed privacy extensions to UML to help software engineers to quickly visualize privacy requirements, and design them into Big Data applications (Jutla, Bodorik, & Ali, 2013).

While trying to take the most of Big Data, in terms of security and privacy, it becomes mandatory that mechanisms that address legal requirements about data handling, need to be met. Secure encryption technology must be employed to protect all the confidential data (Personally Identifiable Information (PII), Protected Health Information (PHI) and Intellectual Property (IP) and careful cryptographic material (keys) access management policies, need to be put in place, to ensure the correct locking and unlocking of data – this is particularly important for data stored. In order to be successful these mechanisms need to be transparent to the end-user and have low impact of the performance and scalability of data (software and hardware-based encryptions mechanisms are to be considered) (Advantech, 2013).

An important security and privacy challenge for Big Data is related with the storage and processing of encrypted data. Running queries against an encrypted database is a basic security requirement for secure Big Data however it is a challenging one. This raises questions such as:

1. Is the database encrypted with a single or multiple keys;
2. Does the database needs to be decrypted prior to running the query;
3. Do the queries need to be also encrypted; and
4. Who as the permissions to decrypt the database; and many more.

Recently a system that was developed at MIT, provides answers to some of these questions. CryptDB allows researchers to run database queries over encrypted data (Ra Popa & Redfield, 2011). Trustworthy applications that intent to query encrypted data will pass those queries to a CryptDB proxy (that sits between the application and the database) that rewrites those queries in a specific way so that they can be run against the encrypted database. The database returns the encrypted results back to the proxy, which holds a master key and will decrypt the results, sending the final answer back to the application. CryptDB supports numerous forms of encryption schemes that allow different types of operations on the data (RA Popa & Redfield, 2012). Based on CryptDB, Google has developed the Encrypted Big Query Client that will allow encrypted big queries against their BigQuery service that enables super, SQL-like queries against append-only tables, using the processing power of Google's infrastructure (Google, 2014).

Apart from more specific security recommendations, it is also important to consider the security of the IT infrastructure itself. One of the common security practices is to place security controls at the edge of the networks however, if an attacker violates this security perimeter it will have access to all the data within it. Therefore, a new approach is necessary to move those security controls near to the data (or add additional ones). Monitoring, analyzing and learning from data usage and access is also an important aspect to continuously improve security of the data holding infrastructure and leverage the already existing security solutions (Kindervag et al., 2012; Kindervag, Wang, Balaouras, & Coit, 2011).

FUTURE RESEARCH DIRECTIONS

Big data is a conventional term used to describe the exponential increase and accessibility of structured and unstructured data. In future big data will be essential to business as well as so-

ciety like internet facility. Resolutions that were previously build on estimation or on conceptual models of reality can now be done based on the collected and stored data itself. Big Data analysis is now used in almost every phase of our society, communication services, marketing, banking and research. The big data phase has shown the ways for huge opportunities in science, health care system, economic decision, educational system and novel forms of public interaction and entertainment. But these opportunities also result in challenges in the area of privacy and security.

There are also important research challenges on maintaining end to end data security and privacy. Ensuring that data is never revealed in clear, in particular to non-authorized parties, on any point of the Big Data lifecycle. Moving from data to programs, there are techniques for protecting privacy in browsing, searching, social interactions, and general usage through obfuscation methods. However, there is more research to be conducted on the processing of encrypted data and privacy protection in the context of both computer programs and web-based systems.

CONCLUSION

In this chapter, the security and privacy problems have been highlighted that need to be addressed to make Big Data processing and computing infrastructure more secure. Common elements specific to Big Data arise from the use of multiple infrastructure tiers (both storage and computing) for processing Big Data; the use of new compute infrastructures such as NoSQL databases (for fast throughput necessitated by Big Data volumes) that have not been thoroughly vetted for security issues; the non-scalability of encryption for large data sets; the non-scalability of real-time monitoring techniques that might be practical for smaller volumes of data; the heterogeneity of devices that produce the data; and the confusion surrounding the diverse legal and policy restrictions that lead to ad hoc approaches for ensuring security and

privacy. Many of the items in this list serve to clarify specific aspects of the attack surface of the entire Big Data processing infrastructure that should be analyzed for these threats.

Throughout this chapter it was possible to present some of the most important security and privacy challenges that affect Big Data projects and their specificities. Although the information security practices, methodologies and tools to ensure the security and privacy of the Big Data ecosystem already exist, the particular characteristics of Big Data make them ineffective if they are not used in an integrated manner. This chapter also presents some solutions for these challenges, but it does not provide a definitive solution for the problem. It rather points to some directions and technologies that might contribute to solve some of the most relevant and challenging Big Data security and privacy issues. As noted throughout this chapter, although some important steps are being given towards solving Big Data security and privacy issues, there is still a long road ahead.

REFERENCES

- Acıçmez, Koç, & Seifert. (2006). Predicting secret keys via branch prediction. *Topics in Cryptology—CT-RSA 2007*, 225-242.
- Advantech. (2013). *Enhancing Big Data Security*. Retrieved from http://www.advantech.com.tw/nc/newsletter/whitepaper/big_data/big_data.pdf
- Barbard, M., & Zeller, T. (2006). A Face is Exposed for AOL Searcher No. 4417749. *The New York Times*.
- Boyd, D., & Crawford, K. (2012). Critical Questions for Big Data. *Information Communication and Society*, 15(5), 662-675. doi:10.1080/1369118X.2012.678878
- Cloud Security Alliance. (2013). *Expanded Top Ten Security and Privacy Challenges*. Retrieved from https://downloads.cloudsecurityalliance.org/initiatives/bdwc/Expanded_Top_Ten_Big_Data_Security_and_Privacy_Challenges.pdf
- Demchenko, Y., Ngo, C., de Laat, C., Membrey, P., & Gordijenko, D. (2014). Big Security for Big Data: Addressing Security Challenges for the Big Data Infrastructure. In W. Jonker & M. Petković (Eds.), *Secure Data Management*. Springer International Publishing. Retrieved from http://link.springer.com/chapter/10.1007/978-3-319-06811-4_13
- Duhigg, C. (2012). How Companies Learn Your Secrets. *The New York Times*.
- Fayyad, U. (2012). *Big Data Analytics: Applications and Opportunities in On-line Predictive Modeling*. Retrieved from <http://big-datamining.org/keynotes/#fayyad>
- Google. (2014). *Encrypted Big Query Client*. Retrieved August 03, 2014, from <https://code.google.com/p/encrypted-bigquery-client/>
- Hasan, O., Habegger, B., Brunie, L., Bennani, N., & Damiani, E. (2013). A Discussion of Privacy Challenges in User Profiling with Big Data Techniques: The EEXCESS Use Case. *IEEE International Congress on Big Data*, 25-30. doi:10.1109/BigData.Congress.2013.13
- Juels, A., & Oprea, A. (2013). New approaches to security and availability for cloud data. *Communications of the ACM*, 56(2), 64. doi:10.1145/2408776.2408793
- Jutla, D. N., Bodorik, P., & Ali, S. (2013). Engineering Privacy for Big Data Apps with the Unified Modeling Language. *2013 IEEE International Congress on Big Data*, 38-45. doi:10.1109/BigData.Congress.2013.15
- Kindervag, J., Balaouras, S., Hill, B., & Mak, K. (2012). *Control And Protect Sensitive Information In the Era of Big Data*. Academic Press.

Kindervag, J., Wang, C., Balaouras, S., & Coit, L. (2011). *Applying Zero Trust To The Extending Enterprise*. Academic Press.

Luo, H., Lin, Y., Zhang, H., & Zukerman, M. (2013). Preventing DDoS attacks by identifier/locator separation. *IEEE Network*, 27(6), 60–65. doi:10.1109/MNET.2013.6678928

Michael, K., & Miller, K. W. (2013). Big Data: New Opportunities and New Challenges. *Computer*, 46(6), 22–24. doi:10.1109/MC.2013.196

Okman, L., Gal-Oz, N., Gonen, Y., Gudes, E., & Abramov, J. (2011). Security Issues in NoSQL Databases. *TrustCom IEEE Conference on International Conference on Trust, Security and Privacy in Computing and Communications*, 541–547.

Percival, C. (2005). Cache missing for fun and profit. *BSDCan*.

Popa, R., & Redfield, C. (2011). Cryptdb: protecting confidentiality with encrypted query processing. doi:10.1145/2043556.2043566

Popa, R., & Redfield, C. (2012). CryptDB: Processing queries on an encrypted database. *Communications*, 55, 103. doi:10.1145/2330667.2330691

Soofi, A. A., & Khan, M. I. (2014). Security Issues in SaaS Delivery Model of Cloud Computing. Academic Press.

Tankard, C. (2012). Big data security. *Network Security*, (7), 5–8. doi:10.1016/S1353-4858(12)70063-6

Top ten big data security and privacy challenges. (2012). Cloud Security Alliance White Paper.

Wu, , Zhu, , Wu, , & Ding,. (2014). Data Mining with Big Data. *IEEE Transactions on Knowledge and Data Engineering*, 26(01), 97–107. doi:10.1109/TKDE.2013.109

ADDITIONAL READING

Boyd, D., & Crawford, K. (2012). Critical Questions for Big Data. *Information Communication and Society*, 15(5), 662–675. doi:10.1080/1369118X.2012.678878

Hough, A. (2010). *Google engineer fired for privacy breach after ‘stalking and harassing teenagers*. The Telegraph.

Muniswamy-Reddy, K. K., Macko, P., & Seltzer, M. (2010). Provenance for the Cloud, *the 8th USENIX Conference on File and Storage Technologies (FAST '10)*.

KEY TERMS AND DEFINITIONS

Analysis: Finding tractable solutions based on the threat model.

Big Data: Big data is a conventional term used to describe the exponential increase and accessibility of structured and unstructured data.

CSA: Cloud Secure Alliance, a non-profit organization with a mission to promote the use of best practices for providing security assurance within Cloud Computing, has created a Big Data Working Group that has focused on the major challenges to implement secure Big Data services.

Distributed Programming Frameworks: It utilize parallel computation and storage to process massive amounts of data.

IDC: International Data Corporation terms this as the “Digital Universe” and predicts that this digital universe is set to explode to an unimaginable 8 Zetabytes by the year 2015.

Implementation: Implementing the solution in existing infrastructures.

SIEM: Security information and event management system may collect event logs from millions of hardware devices and software applications in an enterprise network.

How Visualisation and Interaction Can Optimize the Cognitive Processes Towards Big Data

Antonio Feraco

Fraunhofer IDM@NTU, Singapore

Marius Erdt

Fraunhofer IDM@NTU, Singapore

INTRODUCTION

Visual analytics is the study of transformation of data to visual representations. The goal is to create these representations in such a way that their interpretation is driven by effective and efficient cognitive processes that enable an easy understanding of the data itself. Today, visualization and visual analytics is more than just a collection of plots, graphs, and computer-generated 3D renderings. There are many visualization techniques for every form of data, including, but not limited to, texts, documents and corpora, tree graphs and networks, image collections and videos, time series, tabular and multivariate data, geographical data, scalar vector, and tensor fields, isosurfaces, numerical, geometrical, statistical and other mathematical models, historical events and provenance records, dynamic data streams, algorithms, programs, and computational logs, and a wide range of domain-specific data in disciplines such as engineering, biology, medicine and many others.

Visualization is easily perceived as a means for presenting beautiful computer generated images and animations to impress an audience. However, significant evidence obtained through perceptual studies and user evaluation confirms that proper visualization has enabled researchers and decision makers to be more efficient in gaining insights from data and therefore efficiently improving their cognitive system. It facilitates the formulation of new hypotheses, assists in decision-making,

enables effectual communication of ideas, and facilitates dissemination of knowledge.

To gain insight into these data, to make sense of these data, and to gain new scientific knowledge out of it, we must proceed to work on and aim to define efficient and user-friendly solutions.

The importance of having proper visualisation solutions to improve and optimize the memorizing processes is continuously increasing, e.g. to optimize time and effort in cognitive learning processes and in gaining insights from geo-spatial, behavioral, commercial, scientific data.

Useful data is generated with the advent of new hardware technologies and with sensors embedded everywhere. Examples are CCTV cameras, social media, and systems where machines and users generate contents in various forms, such as, videos, images, text and geospatial data. This data represents a valuable source to identify trends in all sectors. Leveraging social media data presents many challenges: social media data has a large volume; it is a multimodal set of data, it is often ambiguous in its content, and is highly context- and user-dependent.

Moreover, Human Computer Interaction (HCI) transformed the way end users are interacting with information. New interaction paradigms, such as voice recognition, gesture based interactions and multi-touch interaction now enable a deeper interactive experience in memorizing information and in optimizing search for information.

DOI: 10.4018/978-1-5225-2255-3.ch034

Mobile platforms like smart phones and tablets have advanced significantly in the last 10 years. The incorporation of faster processors and mobile GPUs has greatly increased the impact and range of applications. The ubiquity of geo data and mapping applications for smart phones has familiarized a broad base of users with the navigation and interpretation of 2-D maps as well as 3-D maps that have been introduced on the major mobile platforms. This phenomenon is also increasing the familiarity of users with navigating in 3-D textured environments not only representing geographical spaces.

The following chapter describes the transformation of cognitive processes through different ways of interacting with data as well as it describes the importance that visualisation is gathering as a common taxonomy to optimize insights from disseminating and sharing data. Its structure is as follows:

1. The first part focuses on the Human Computer Interaction and Interactive Digital Media,
2. The second part focuses on Big Data and Interaction, and
3. The last part comprises of the description of the experiment and the authors conclusions.

The chapter is structured as follows:

1. A first part focuses on the Human Computer Interaction and Multi model user interfaces,
2. The second part on Interaction,
3. The last part comprises of the results from an experiment and the authors conclusions.

BACKGROUND: HUMAN COMPUTER INTERACTION AND MULTI-MODAL USER INTERFACES

To define HCI is not a simple task because of the applicative nature that this subject has and the continuous changes it is facing. Simply, we

can assert that HCI tries to model and analyze the relation between man and computer where the element “computer” changed its nature drastically in the last decades following a very rapid dynamic of change¹. Twenty-five years ago, few people would have anticipated the tremendous processing speed of contemporary computer systems. Even though major improvements have been made in many areas regarding HCI, important issues still remain (Szameitat et al., 2009). HCI studies very heterogeneous objects such as Personal Computers (PC), Personal Digital Assistant (PDA), Mobile Phones, but even simpler objects such as watches or electrical furniture as well as the technologies related to Internet or even more complex technological issues such as the control panel of a chemical plant, the plane cabin, and other. According to the Association for Computing Machinery (ACM) the definition of Human Computer Interaction is:

Human-computer interaction is a discipline concerned with the design, evaluation and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them.

There are other disciplinary points of view that would place the focus of HCI differently than computer science does - just as the definition of databases would be different from a computer science vs. a business perspective. HCI at large is an interdisciplinary area. It is emerging as a specialty concern within several disciplines, each with different emphases: computer science (application design and engineering of human interfaces), psychology (the application of theories of cognitive processes and the empirical analysis of user behavior), sociology and anthropology (interactions between technology, work, and organization), and industrial design (interactive products). From a computer science perspective, other disciplines serve as supporting disciplines, much as physics serves as a supporting discipline for civil engineering, or as mechanical engineering serves as

a supporting discipline for robotics. A lesson learned repeatedly by engineering disciplines is that design problems have a context, and that the overly narrow optimization of one part of a design can be rendered invalid by the broader context of the problem. Even from a direct computer science perspective it is advantageous to frame the problem of human-computer interaction broadly enough such as to help practitioners to avoid the classic pitfall of design that is separated from the context of the problem. Human-computer interaction arose as a field from intertwined roots in computer graphics, operating systems, human factors, ergonomics, industrial engineering, cognitive psychology, and the systems part of computer science. Computer graphics was born from the use of Cathode Ray Tube (CRT) and pen devices very early in the history of computers. This led to the development of several human-computer interaction techniques. Many techniques date from Sutherland's Sketchpad Ph.D. thesis (1963) that essentially marked the beginning of computer graphics as a discipline. Work in computer graphics has continued to develop algorithms and hardware that allows the display and manipulation of ever more realistic-looking objects (e.g., CAD/CAM machine parts or medical images of body parts). Computer graphics has a natural interest in HCI as "interactive graphics" (e.g., how to manipulate solid models in a CAD/CAM system). Furthermore it is an ubiquitous and indispensable tool for industrial design, being the primary means for modeling and communicating product design proposals (Sener et al, 2008). Many disciplines flow together in HCI, but all of them study a common issue. This issue is the digital interactive system placing HCI on the cross road among all these disciplines. Between these disciplines we can name the computer graphics, the cognitive psychology, the design, the economical studies and the economical processes management. In reality, this multi-disciplinary nature derives from the ergonomics both for the physics aspects and the cognitive aspects. The International Ergonomics Association defines ergonomics as:

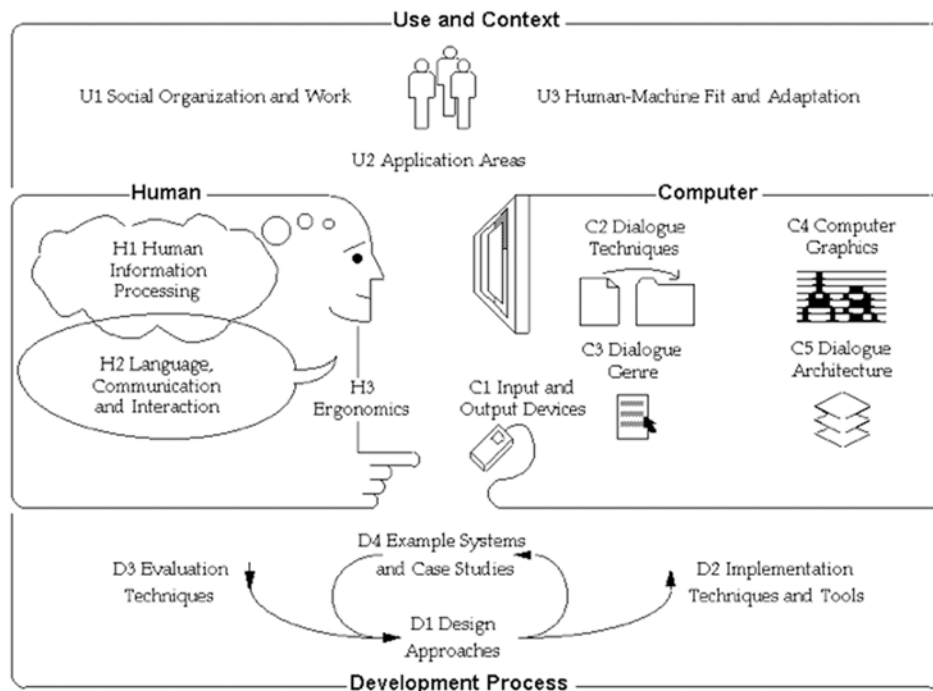
Ergonomics (or human factors) is the scientific discipline concerned with the understanding of interactions among humans and other elements of a system, and the profession that applies theory, principles, data and methods to design in order to optimize human well-being and overall system performance.

B

The difference between HCI and ergonomics is given by the different emphasis on the various aspects of the interaction human-system. We could think about the HCI as a studying the relationship existing between users and computer systems. A great deal of specifically directed work has taken place over the years, but there are excellent introductory books to HCI (Baecker, Grudin, Buxton, & Greenberg, 1995); (Helander, Landauer, & Prabhu, 1997) and also to the more specific study of user-interfaces as well as the development of new generation interactive interfaces (Feraco, 2016). In HCI there are a great number of approaches used in analyzing computer systems. Of particular importance from early in the discipline's history is the use of cognitive psychological models of user's action and planning to discuss how they interact with computers. Theoretical framework such as the GOMS (Goals, Operators, Methods, Selection rules) approach or Donald Norman's seven-stage action cycle in the Design of everyday things is representative of this tradition. HCI has traditionally used a methodology of building software and user-interfaces, testing them on users, measuring and analyzing the results, improving the software, and so on. In recent times, the popularity of ethnographic techniques such as interviews, observation in workplaces, and so on, has been growing as researchers become more concerned with the context of use. Jakob Nielsen has popularly defined usability as focusing on five main properties for emphasis in software: learnability, efficiency, memorability, error prevention, and user satisfaction. According to ISO 9241-11 (1998), usability is concerned with the effectiveness, efficiency and satisfaction with which user can achieve specified goals in specified context

Figure 1. Context and Use of HCI

Source: Pearce 1994



of use. According to (Preece, Rogers, & Sharp, 2002) the terms effectiveness means “how good a system is at doing what it is supposed to do”. That is, effectiveness suggests that specified goals are to be achieved with accuracy and completeness (ISO 9241-11). Effectiveness is thus related to a system’s desired functionality. Some of the interrelationships among these topics are represented in Figure 1. Computer systems exist within a larger social, organizational and work milieu (U1). Within this context there are applications for which we wish to employ computer systems (U2). But the process of putting computers to work means that the human, technical, and work aspects of the application situation must be brought into fit with each other through human learning, system tailor-ability, or other strategies (U3). In addition to the use and social context of computers, on the human side we must also take into account the human information processing (H1), communication (H2), and physical (H3) characteristics of users. On the computer side, a

variety of technologies have been developed for supporting interaction with humans: Input and output devices connect the human and the machine (C1). These are used in a number of techniques for organizing a dialogue (C2). These techniques are used in turn to implement larger design elements, such as the metaphor of the interface (C3). Getting deeper into the machine substrata supporting the dialogue, the dialogue may make extensive use of computer graphics techniques (C4).

Among the HCI, one of the main aim of the research has been to humanize the interfaces. Humanization of interfaces has two aspects: to simplify interfaces in order to make them easier and pleaser to be used (Schneiderman, 1992; Norman & Draper, 1966) and to make the interfaces the more possible it was similar to human being. The way, among communication human-computer, can be defined as a perceptive process through one of the three channel of human perception. To design effective multi modal interfaces it is good to follow some suggestions given by the

Table 1. Performing the four basic manipulation tasks using four popular input modes, ranked from the easiest (1) to the most difficult (4)

Content Manipulation Task	Voice Mode	Pen Mode	Keyboard/Keypad	Mouse/Joystick
Select objects	Speak the name of the object (3)	Point to or circle the object (1)	Press keys to position the cursor on the object and press the select key (4)	Point to and click on the object or drag to select text (2)
Enter text	Speak the words in the text (2)	Write the text (3)	Press keys to spell the words in the text (1)	Spell the text by selecting letters from a soft keyboard (4)
Enter symbols	Say the name of the symbol and where it should be placed. (3)	Draw the symbol where it should be placed (1)	Enter one or more characters that together represent the symbol (4)	Select the symbol from a menu and indicate where it should be placed (2)
Enter sketches or illustrations	Verbally describe the sketch or illustration (2)	Draw the sketch or illustration (2)	Impossible (4)	Create the sketch by moving the mouse so it leaves a trail (similar to an Etch-a-Sketch™) (3)

Source: W3C Group Note

W3C organizations; these suggestions are based on four major principles:

1. Satisfy real-world constraints.
2. Communicate clearly, concisely and consistently with users.
3. Help users recover quickly and efficiently from errors.
4. Make users comfortable.

What the user is willing to achieve through the application is limited by the real-world constraints. These limitations may be due to the nature of the task the user intend to perform, other activities the user is performing, physical limitations of the user, and condition of the environment in

which the user is performing the task. The user interface should be designed to compensate for these limitations. The tasks mostly performed by the users are indicated in Table 1, and those tasks are mostly achieved through the use of the new mobile devices that will enable to enter data by speaking into a microphone, writing with a stylus, and pressing keys on a small keypad.

It is needed to take in account even the physical suggestion thus physical devices exhibit different usability characteristics. The size, the shape and the weight of the device affect how it may be used. Most important, the placement of a microphone and speaker, the size of the display and writing surface, and the size of keys in a keypad affect the ease with which a user can enter information

Table 2. Physical usability issues for the four most popular modes of information entry

Device Usability Issues	Voice Mode	Pen Mode	Keystrokes Mode	Mouse/Joystick Mode
Required number of user hands	None (plus possibly one to hold the device)	One (plus possibly one to hold the device)	One or two	One
Required use of eyes	No	Yes	Frequently, but some users can operate familiar keyboards without looking at them	Yes
Portable	Yes, especially when walking	Yes, but difficult while walking	Yes, but difficult while walking	Yes, but difficult while walking

Source: W3C working group note

by speaking, writing or pressing keys. Table 2 summarizes the three modes of respect to physical usability issues.

Another important issue to consider is the one related to the environment. People may work in environment that may not be ideal for some modes of user interfaces. The environment might be noisy or quiet, hot or cold, light or dark, or moving or stationary with a variety of distractions and possible dangers. Multimodal user interfaces must be designed to work in the environments where they will be used. Table 3 summarizes the environmental usability issues with respect to four popular input modes.

The second principle on which these suggestions are based is the clear, concise and consistent communication with the user. Effective communication between the user and the device is necessary for achieving the user's goals. The MUI (Multimodal User Interface) is the conduit for all communication between the user and the device. Communication has to be clear and concise, avoiding ambiguities and confusion. Communication styles should be consistent and systematic so users know what to expect and can leverage the patterns and rhythms in the dialog. This can be achieved through consistency suggestions that will enable users to leverage conversational patterns to accelerate their interaction, through organizational suggestions thus organizing information and transition between topics will improve the users' comprehension of and performance with the multimodal interface; information should be structured and organized as in ways that are familiar to the user. The third principle is to help users recover quickly and efficiently from errors, in fact all users,

especially novice users, will occasionally fail to respond to a prompt appropriately. The UI must be designed to detect such errors and assist user to recover naturally, furthermore the multimodal interface should help users learn how to use the user interface to achieve the desired results quickly and efficiently. This could be achieved through the conversational inputs, in fact the principle of conversational discourse advice that the suggestions for the nature, content and format of information exchanged between two humans may be applied to information exchanged between a human and a computer; through reliability, that is a lot frustrating for the user having a device at hand and not being able to use it. The last principle is to make users feel comfortable; users often judge a computer application by its user interface. If they will not like the user interface, the application will not be used. If the user interface is not easy to learn and not easy to use, the application cannot be used successfully.

Three different way can be chosen: visible, listenable and touchable. With the advances in ubiquitous computing the quest for natural interaction is of utmost importance. This includes interaction with different devices and modalities that are optimally suited to support the user's tasks, ideally without requiring the user to select and configure such devices. A multimodal/multi-device media player application is sketched by (Schaefer & Mueller, 2008) and summarized in Figure 2.

A user can employ different devices to control the media player such as a mobile phone, the media-playing device itself, a microphone for voice control or a wireless multisensory for gesture interaction. When we consider the

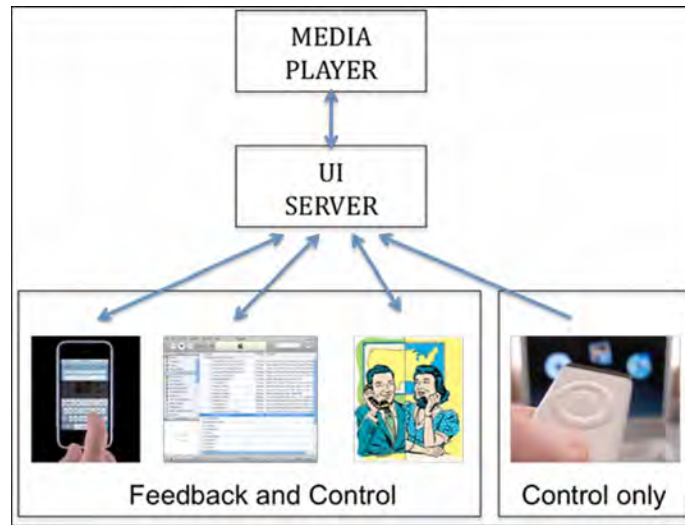
Table 3. Environmental usability issues for the four popular modes of information entry

Device Usability Issues	Voice Mode	Pen Mode	Keystroke Mode	Mouse/Joystick Mode
Noisy environment	Works poorly in a noisy environment	Works well in a Noisy environment	Works well in a noisy environment	Works well in a noisy environment
Other environmental concerns	Works well independently of gloves	Does not work well when users must wear thick gloves	Does not work well when users must wear thick gloves	Does not work well when users must wear thick gloves

Source: W3C working group note

Figure 2. A multimodal/multi-device media player

Source: Schaefer and Mueller, 2008



Interactive Digital media, we have to care about the ubiquitous computing environments. In this environment an increasing number of services are delivered to users through various devices including cellular phones and PDAs (Portable Digital Assistant). Then, it is important to facilitate interoperability among services on these heterogeneous and autonomic environments. Web services (Ferris & Farrel, 2003) based on Internet standards such as SOAP (Simple Object Access Protocol) (Mitra & Lafon) and WSDL (Web Service Description Language) gained a momentum as a standard interface for the interoperability between software applications. This led literature to deepening studies related to the understanding individuals' motivation and involvement. It is of a central concern in media uses and effects research, because it is more illuminating to reveal why and how people use media rather than just to focus on what people do with media.

BIG DATA

It is worth to provide the reader with a brief description also about Big Data and the related challenges, to introduce then the impact that

visualisation and interaction might have on the learning cognitive processes.

According to Wikipedia Big Data is a term for *data sets* that are so large or complex that traditional data processing applications are inadequate to deal with them. Challenges include analysis, capture, data curation, search, sharing, storage, transfer, visualisation, querying, updating and information privacy. The term Big Data often refers simply to the use of predictive analytics, user behavior analytics, or certain other advanced data analytics methods that extract value from data, and seldom to a particular size of data set (Cavanillas, Curry, & Wahlster, 2015).

According to (Cavanillas, Curry, & Wahlster, 2015) Big Data is also the emerging field where innovative technology offers new ways of extracting value from the huge fluxes of information coming from heterogeneous sources. Today's environment allows us to have access to more types of data which comes in all kind of forms: from highly structured ERP (Enterprise Resource Planning) data, or CRM (Customer Relation Management) data, to multi-million text file, to video files and machine generated sensor data and BIM (Building Information Management) data. The data sources include online transactions, social networking ac-

tivities, mobile device services, internet gaming, conceptualization and design and etc.

In line with it, data sources are expanding, not only coming from social networks sources, such as Facebook, YouTube, Google (Shen, Li, Wu, & Liu, 2016), but also from the Digital Transformation the most of the sector are facing, from the most traditional, like medical and constructions, to the newest ones, e-commerce and Fintech.

The amount of Heterogeneous and multidimensional datasets is most of the times not easy to be analyzed. In many domains experts are striving to make sense out of data which is collected and computed through several sources, mainly why there is a boom in the amount of information that can be produced per sample (Turkay, Lundervold, Lundervold, & Hauser, 2013).

In this context, HCI, as well as HMI (Human Machine Interaction) gain exponential importance, above all considering the nowadays interaction methodologies with handheld devices and the technology trend of moving towards a keyboard-less based human data interaction. Among these consideration, visualisation also start to play a key role in the discipline of Human Data Interaction.

VISUALISATION

Not far in the past, the ability to create smart data visualisation (or aka “Data-Viz”) was a simple nice-to-have skill for design- and data-minded managers. Nowadays it became a must have skill for all software managing big datasets (Berinato, 2016). Decision makers increasingly relies on data, which are characterized by the high pace of their generation, velocity and huge volume, leading to some level of abstraction in most of the cases. In the context of this chapter, Visualisation is defined by the common taxonomy to understand big datasets. We entered a data-driven era, not only from the data generation perspective, but also from the growing expectations placed in the analysis of data. On a parallel path, graphic cards together with their imaging processing power can

leverage large data visualisation but they can also be of great interest to support interaction (Hurther, 2016). One of the main technique used in data visualisation is the image based one. This approach takes advantage of changes in the bottlenecks of computer graphics: since data storage and memory limitation is becoming less and less of an issue (Sutherland, 2012). Main instances of image-based techniques with the graphic card usage to facilitate multivariate data exploration are:

- **Rendering:** Graphics card can now render numerous items on the screen and thus can display large datasets.
- **Computation:** Graphic cards can perform fast and parallel data processing, and can be used to process information at the data level.
- **Interaction:** Interaction with the data is an important manipulation paradigm to perform data exploration. Graphic cards can be used to provide tools to help users to interact with large data sets.

THE INTERACTIVITY IN DIGITAL MEDIA

As big data are at the base of digital media, most of digital media are based on the principle of interactive structures. User should be able to relocate and to challenge them-selves to the interactive observation of a work. All projects and/or development related to digital media should excite not only the visitors’ bodies, but also should bring their thoughts into motion. Becoming part of an interactive work, for the user, “Touching” an interactive work is not only allowed, but becomes necessary; whether with mouse, trackball, touch screen, tangible objects, video camera, responsive workbench, virtual balance, the touch less Point-Screen or other interfaces; the observer should be conducted to bring the process into motion (Fleischmann & Strauss, 2008), (Feraco, 2016). In the field of arts, digital media, and even more

interactive digital media, represents a milestone of the new challenges. Many researches tried to define interactivity, underlining its importance and focusing it on differences existing in perception (Johnson & Kaye, 2016). The difference in perception is differentiated as follows: “A painting is an instance of representation. A film is a sequence of representation. Interactive artworks are not an instance of representation; they are virtual machines which themselves produce instances of representation based on real time inputs”. According to (Valli, 2008) interaction design is the art of instigating and guiding behaviors (or interaction design) by means of proper static or dynamic stimuli. It is defined in terms of experiences: people naturally communicate through gestures, expressions, movements, and discover the world by looking around and manipulating physical stuff. Valli also states that the key assumption here is that they should be allowed to interact with technology as they are used to interact with the real world in everyday life, as evolution and education taught them to do. Because of these assumptions today’s designers face a great challenge: the creation of new interaction paradigms and new media conventions, that exploit the new machines’ sensing capabilities offered by technology and to take care of human spontaneous ways to discover the real world. On the other side, interactive technology, in terms of sensors, actuators and narrative intelligence, is still matter of research for engineers and scientists. As even stated by (Fleischmann & Strauss, 2008) interactive structures remain the basic principle of digital media. Users of digital media will be in a position to relocate and to challenge themselves to the interactive projects to make an experience that moves over and beyond the usual contemplative observation.

Based on the state of the art before mentioned Fraunhofer IDM@NTU, in Singapore, develops interactive user interfaces that enables a smoother interaction between the user himself and big datasets. Following is described an approach to the study of one of the technologies developed that strengthen the importance of three key dimensions for interactive digital media:

1. Efficiency,
2. Experience, and
3. Usability.

TAM STUDY OF A BUILDING INTERACTIVE 2D MAP AND A 3D WALKTHROUGH

BIM (Building Information Models) represents a big dataset in the building and construction industry. Their visualisation is not always smooth and simple for non-experts, consequently in this specific area the adoption of interactive technologies, i.e. interactive walkthrough, can represent a technique enabling a faster acquisition of data insights. In this context the author shares a Technology Acceptance Model study conducted during SIGGRAPH Asia 2014 in Singapore and its results gathered where three main dimensions: experience, efficiency and usability, related to the interaction with a 2D map allowing a 3D navigation walk through have been analyzed. To simplify the description of the overall set-up, including both the hardware (a 46” multi-touch table and a 55” vertical UHD TV) and the User Interface have been given the acronym of MTT. The core characteristics of both the technologies used for the TAM study are based on visualisation and interaction paradigms.

The study focused on understanding the impact on tasks given to the users of three different dimensions and their respective factors:

- **Experience, Factors Related:** Engaging and Intuitiveness.
- **Efficiency, Factors Related:** Reliability of the touch (precise and sensitive), User friendly (easy to locate myself in the map).
- **Usability, Factors Related:** 3 fingers related function, maps switching, multi-map manipulation, usefulness for multi-collaborative purposes.

Each end-user was given specific tasks in two different maps:

1. The Siena Cathedral, and
2. The Frankfurt Messe.

Both maps and 3D environment are a virtual representation of the real

1. Cathedral in Siena, and
2. The building used in Frankfurt to conference and events.

The study has been conducted on four different categories of people: Computer Graphics experts, Computer Science experts, Students, and Media.

Figure 3 summaries the model outcome of the analysis:

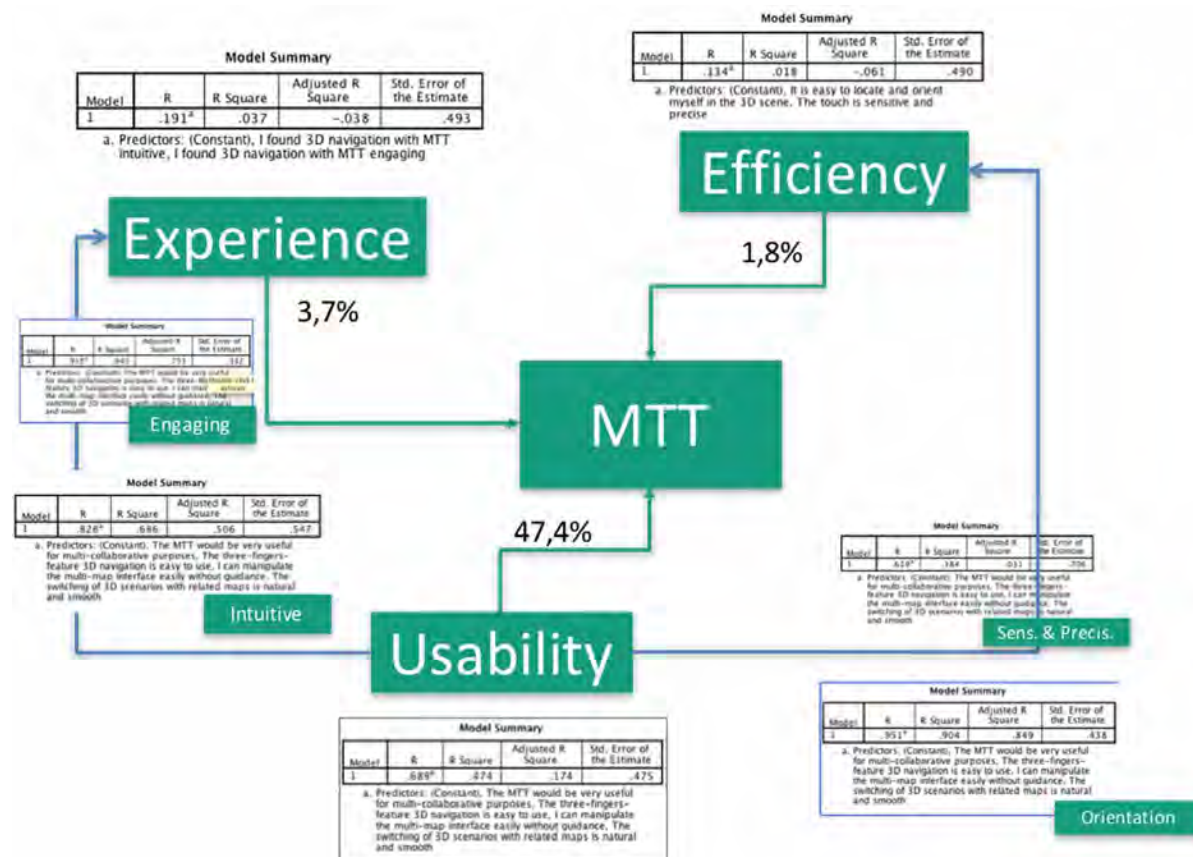
Figure 3 shows that Usability is dimension that impacts the most on the UI developed and that also has major influence on the others. In this

specific context usability has been developed on top of Visualisation and interaction paradigm that simplified and enhanced the cognitive processes and the task accomplishment of the end users.

FUTURE RESEARCH DIRECTIONS

More and more UI are being developed with in mind the main paradigm of usability, user friendliness and intuitiveness, but there is still the need to improve and to further analyze the overall Human Data Interaction considering that the trend to optimize both time and space is leading us towards a keyboard-less world where touch based interactions are becoming the main mean of querying for data insights.

Figure 3. Model resulting from the TAM study



CONCLUSION

The study further proved that visualisation and interaction flow into optimizing the usability of a multimodal user interface. Visualisation on its own can be regarded as the new taxonomy to communicate with the information while interaction as the query generator.

Technologies such as virtual reality and augmented reality are now facing a breakthrough in the market, and their fields of application are growing exponentially. Nevertheless, in this specific fields, as well as in the computer graphics in general, expertise and experience plays a key role to design and implement at the best the usability in relation to the scope the UI is made for.

Nowadays, more than in the past, the capability of merging visual expertise and interaction consciousness are the basic breakthrough for any application to be successful in the end-users' acceptance environment.

REFERENCES

- Baecker, R. M., Grudin, J., Buxton, W., & Greenberg, S. (1995). *Readings in Human-Computer Interaction: Toward the Year 2000* (2nd ed.). Morgan Kaufmann Publishers.
- Berinato, S. (2016, June). Visualisation that really works. *Harvard Business Review*, 94(6), 92–100.
- Boyd, D., & Crawford, K. (2016, November 15). *Big data*. Retrieved from Wikipedia: https://en.wikipedia.org/wiki/Big_data
- Cavanillas, J., Curry, E., & Wahlster, W. (2015). *New Horizons for a Data-Driven Economy*. Springer.
- Feraco, A. (2016). *Info-Land: New interactive and visual marketing tool to reach new generation of customers*. In *Circle 2016* (pp. 93–102). Naples: Access Press.
- Ferris, C., & Farrel, J. (2003). What are Web services? *Communications of ACM E-Services: A Cornucopia of Digital Offerings Ushers in the Next Net-Based Evolution*, 46(6), 31.
- Fleischmann, M., & Strauss, W. (2008). Interactivity as Media Reflection between Art and Science. *The Art and Science of Interface and Interaction Design*, 75-92.
- Helander, M. G., Landauer, T. K., & Prabhu, P. V. (1997). *Handbook of Human-Computer Interaction*. Amsterdam: Elsevier.
- Hurther, C. (2016). *Image-based Visualisation: Interactive Multidimensional Data Exploration*. Morgan & Claypool.
- Johnson, T. J., & Kaye, B. K. (2016, August). Some like it lots: The influence of Interactivity and reliance on credibility. *Computers in Human Behavior*, 61, 136–145. doi:10.1016/j.chb.2016.03.012
- Mitra, & Lafon. (n.d.). *SOAP 1.2 Part 0: Primer*. Academic Press.
- Norman, D. A., & Draper, S. W. (1966). *User Centered System Design; New Perspectives on Human - Computer Interaction*. Hillsdale, IL: L. Erlbaum Associates Inc.
- Preece, J., Rogers, Y., & Sharp, H. (2002). *Interaction Design*. New York: John Wiley & Sons.
- Schaefer, R., & Mueller, W. (2008, July). Assessment of a multimodal interaction and rendering system against established design principles. *Journal of Multimodal User Interface*, 25-41.
- Schneiderman, B. (1992). *Tree visualization with tree-maps: 2-dspace-filling approach*. New York: ACM Transactions on Graphics.
- Sener, B., & Wormald, P. (2008). User evaluation of HCI concepts for defining product form. *Design Studies*, 29(1), 12–29. doi:10.1016/j.destud.2007.06.003

- Shen, Y., Li, Y., Wu, L., & Liu, S. (2016). Big Data Overview. In *Big Data: Concepts, Methodologies, Tools, and Applications* (pp. 1-29). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-9840-6.ch001
- Soro, A. (2008). *Human Computer Interaction. Fondamenti e Prospettive*. Roma: Polimetrica International Scientific Publisher.
- Sutherland, I. (2012, October). The tyranny of the clock. *Communications of the ACM*, 55(10), 35–36. doi:10.1145/2347736.2347749
- Szameitat, A. J., Rummel, J., Szameitat, D. P., & Sterr, A. (2009). Behavioral and emotional consequences of brief delays in human-computer interaction. *International Journal of Human-Computer Studies*, 67(7), 561–570. doi:10.1016/j.ijhcs.2009.02.004
- Turkay, C., Lundervold, A., Lundervold, J. A., & Hauser, H. (2013). Hypothesis Generation by Interactive Visual Exploration of Heterogeneous Medical Data. In A. Holzinger & G. Pasi (Eds.), *Human-Computer Interaction and Knowledge Discovery in Complex, Unstructured, Big Data* (pp. 1–12). Maribor, Slovenia: Springer. doi:10.1007/978-3-642-39146-0_1
- Valli, A. (2008). The design of natural interaction. *The Design of Natural Interaction*, 295-305.
- Hurter, C. (2016). *Image-Based Visualisation: Interactive Multi-dimensional Data Exploration*. Morgan & Claypool.
- Karimi, H. A. (2014). *Big data: techniques and technologies in geoinformatics*. Boca Raton: CRC Press, Taylor & Francis Group. doi:10.1201/b16524
- Khosrow-Pour, M. (2016). *Big data: concepts, methodologies, tools, and applications*. Hershey, PA: Information Science Reference.
- Marz, N., & Warren, J. (2015). *Big data: principles and best practices of scalable real-time data systems*. Shelter Island, NY: Manning.
- Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: a revolution that will transform how we live, work, and think*. Boston: Houghton Mifflin Harcourt.
- Schmarzo, B. (2013). *Big data: understanding how data powers big business*. Indianapolis, IN: John Wiley & Sons.
- Søndergaard, M. (2015). The Politics of Big Data Aesthetics. *Mediekultur: Journal Of Media & Communication Research*, 31(59), 1–5.
- Telea, A. (2015). *Data visualization: principles and practice*. Boca Raton: CRC Press, Taylor & Francis Group.

ADDITIONAL READING

- Daniel, B. (2015). Big Data and analytics in higher education: Opportunities and challenges. *British Journal of Educational Technology*, 46(5), 904–920. doi:10.1111/bjet.12230
- Halpern, O. (2014). *Beautiful data: a history of vision and reason since 1945*. Durham: Duke University Press.
- Hu, F. (2016). *Big data: storage, sharing, and security*. Boca Raton, FL: CRC Press. doi:10.1201/b19694

KEY TERMS AND DEFINITIONS

Big Data: Is a term for data sets that are so large or complex that traditional data processing applications are inadequate.

Cognitive Process: The performance of some composite cognitive activity; an operation that affects mental contents; “the process of thinking”; “the cognitive operation of remembering”.

Efficiency: The ratio of the output to the input of a given system. Efficiency is measured as the resources expended by the user in relation to the accuracy and completeness of goals achieved (ISO standard 9241).

Ergonomics: Is the scientific discipline concerned with the understanding of interactions among humans and other elements of a system, and the profession that applies theory, principles, data and methods to design in order to optimize human well-being and overall system performance”.

Human Computer Interaction: HCI (human-computer interaction) is the study of how people interact with computers and to what extent computers are or are not developed for successful interaction with human beings.

Interactive Walkthrough: The possibility to digitally interact with a 2D map and visualizing the 3D environment.

Usability: The degree to which a software can be used by specified consumers to achieve quantified objectives with effectiveness, efficiency, and satisfaction in a quantified context of use.

Visualisation: Making a visible presentation of numerical data, particularly a graphical one.

B

Managing and Visualizing Unstructured Big Data

Ananda Mitra

Wake Forest University, USA

INTRODUCTION

This essay expands on the notion of “Big Data” to open up alternative analytic opportunities, on certain components of the data, through a theoretical lens that is mobilized to offer an interpretation and visualization of the information contained in large amounts of data. It is useful first to examine the term “Big Data” in some detail. The term refers to a phenomenon which results from the fact that institutions and individuals are digitizing many different kinds of information leading to an exponential growth of the amount of data that is being stored in the digital space. First, this expansion relates to the *increase in data points* as more records are added to the corpus of Big Data. Second, the idea of Big Data needs to be considered in terms of the *details that are being digitized*. The notion of Big Data should be considered both in terms of the breadth of the data in terms of number of data points (*amount*) and the depth of the data related to the various fields of information available for each record (*details*). Therefore, Big Data has become an object of analysis for a variety of groups, from academics to marketers, all of whom are interested in understanding how Big Data could provide highly granular voluminous information about people (see, e.g., Mitra, 2014c). Next, it is useful to examine the different categories of information that makes up “Big Data.”

Much of Big Data is numeric that is amenable to mathematical analysis. For instance, it is possible to easily count the number of tweets produced by an individual. Such counts offer the opportunity

for companies such as Tweeter to offer information about what topics are popular at any moment in time. The segment of Big Data that offers the ease of analysis and visualization has been called “structured” Big Data. There is, however, another vast component of Big Data that does not allow for easy numeric analysis. This segment is made up of the utterances of the people who are self-generating the Big Data by voicing themselves in the digital space. An example of this segment of Big Data is the actual specific tweet produced by an individual or the specific photograph uploaded on photograph sharing spaces. In the case of the tweet, the language of the tweet contains information about attitudes and opinions, just as a photograph offers information about the individual who has captured the picture. This form of data requires a more nuanced and “qualitative” analytic process that would discover the intent of the authors and the meanings of the messages encapsulated in a microblog or picture. This segment of Big Data has been named “unstructured” Big Data. Not only is it difficult to analyze the unstructured Big Data but it is also difficult to visualize the findings of analysis. The qualitative process does not typically produce convenient charts and graphs. The analysis needs to be offered for easier understanding and unstructured Big Data makes this a challenge as well.

This paper offers a theoretical and analytic process to consider ways of analyzing Big Data and visualizing the analysis. To do this, it is important to offer a theoretical basis to consider the elements of unstructured Big Data.

BACKGROUND

Perspective on Big Data

The unstructured Big Data can be categorized into three main types. The first set are characterized by short word length where the information is authored by individuals and institutions. This form of the data has sometimes been called “micro-blogs,” as reference to blogs that are strictly restricted by the number of words that can be used in the discourse. The most popular example of micro-blogs are the statements produced by the users of the computer program called Tweeter. The second set of unstructured Big Data is an extension of micro-blogs where the restriction on size disappears but all the other characteristics remain intact. This is a situation where a user can generate discourse of significant length and place it in a digital repository. One popular example of this category are “posts” that users upload within their Facebook “profiles.” The third category of unstructured Big Data that is worthy of consideration is discourse that users generate in response to specific queries. This form of data is rarely circulated over the Internet, but remains as in-depth lengthy treatise on very specific issues that the user is prompted to elaborate. Much like the second category, this segment of unstructured Big Data is not usually restricted in length. However, there are greater restrictions on the scope of content of this form of data since a majority of this data is generated in response to prompts and questions. A good example of this form of data are responses to open-ended questions used in questionnaires in varieties of data collection projects ranging from measuring political opinions to public health assessments. This three-pronged categorization encompasses the majority of unstructured Big Data with one common characteristic – the data is user-generated. It is therefore useful to consider how to characterize the author of the discourse. I offer two broad categories, which can be considered to be mutu-

ally exclusive for most considerations. The two user groups are “institutions” and “individuals” with the differentiating factor being the level of agency of the user.

Institutions often bring the full force of their creative, financial and cultural capital to the creation of digital data, often in the form of “home pages” that populate the digital space. This form of data represents an “authoritative” voice of the powerful and dominant within the public sphere (Foucault, 1991). Such voices carry the ideological baggage of the institutions and their relative position along a continuum of power from the dominant to the oppositional. However, with the availability of the digital tools and the relative ease with which digital data can be produced and circulate it is increasingly possible for some individuals to create such home pages which represented the voice of the relatively powerless individual who would not have the institutional support to present themselves in the public sphere (see, e.g., Mitra and Watts, 2002; Mitra, 2011). Starting with home pages, individuals were able to gain a sense of agency and authorship in the digital space where the individual user-generated data was beginning to become available in the digital space. This tendency expanded rapidly with the development of better tools for digital communication and the ease with which individuals could constantly generate data and place it within networks of other users who were also generating data. Consequently, a new state of empowerment was being achieved by the individuals where their voice was becoming alongside the authoritative voices of the institutions. The corpus of Big Data is thus made up of the institutionally produced and individually authored information. In the remainder of the essay, the focus is on the three categories of Big Data – micro-blogs, posts, and response to open ended questions – produced by individuals. In the next section a specific theoretical approach is suggested to help analyze this form of data.

B

Narrative Theory and Narbs

One of the key characteristics of the people who are producing the unstructured Big Data is that they are individuals without any specific claims of authority, in contrast to the authoritative voice of the institutions. For example, in the last quarter of 2015, 1.59 billion users had logged into Facebook at least once in the past thirty days (Facebook users worldwide 2016, n.d.). Even if a portion of these active users would have created a small Facebook “post” by stating what they have done, what they think about something, or any other personal information, then the individual has contributed to the creation of unstructured Big Data. This form of personal information has been defined as a narb which, “is short for “narrative bit,” a snippet of personal information about your life that you’re willing to share with others (Zimmer, n.d.).”

The key to this definition is the notion of narrative (Mitra, 2014b). It has been suggested that unstructured Big Data is indeed made up of narratives or stories that individuals are now empowered to tell about themselves. It has been argued that people are fundamentally story-telling beings. This has been maintained by scholars across many disciplines where it has been suggested that along with being rational beings at times, rhetorical beings at moments, human beings are also narrative beings. As a narrative being, humans make sense of the World around them through the stories that they tell, and the stories that they listen to (Fisher, 1982, 1984, 1985a, 1985b, 1987). A simple illustration of this story telling process is the way in which interpersonal interactions typically start between strangers with a simple, but profound question – “where are you from?” In the response to the question lies an entire narrative of a person’s life, and the way in which an individual chooses to respond, depending on context and one’s history, produces, in the voice of the individual, a personal story. When such stories are carried over to the digital spaces such as those of digital networks, those stories operate as the narbs that make up unstructured Big Data.

The opportunity to characterize these narbs as building blocks of large stories also acts as the point of departure to place these narbs within the larger framework of narrative theories, most notably the theoretical approach that argues for a “narrative paradigm” to consider the way everyday life operates. Just as there are paradigms that argue for rationality, the narrative paradigm proposed by Walter Fisher suggests that the analysis of the stories offer an understanding of the elements of the story in terms of its believability and aesthetic structure. When the theory was first proposed, the motives of the authors in telling the story were of less importance than the structure of the stories themselves. However, the paradigm offered both a theoretical and methodological possibility of considering the stories as the indicator of the motives of the authors in telling the stories. This opportunity was also not used because a good portion of the stories available for analysis in the mid-1980s were non-biographical stories. Consider any of the tragedies authored by Shakespeare. These are not stories that tell the life story of the author himself, but the author’s perspective on real and fictional moments in history. While the analysis of the stories would offer a tangential look at the perspective of the author, that was not the main goal of the analysts who were working within the theoretical umbrella of the narrative paradigm. Yet, with the advent of the narb, and its explosive growth as un-structured Big Data, the majority of the narbs are stories uttered by the author about the author.

This is well illustrated in the way that Facebook invites the active user to interact within the digital network. The opening page of the system states, “What’s on your mind?” This is an unambiguous invitation to create a narb about the self. When countless narbs become available - through the three categories discussed earlier in this essay - these narbs become the wealth of stories that can now be analyzed precisely because the narrative paradigm has successfully theorized and argued that humans are narrative beings and by listening to their narratives we learn about people (Mitra,

2012b). The challenge posed by unstructured Big Data is the fact that the volume of narratives has become so large that there must now be specific analytic strategies that need to be utilized for the analysis and visualization of the narbs.

Methodology

There are several aspects in the analytic process that pose challenges. The first challenge relates to the volume of data. Very large amounts of data must be analyzed so that a narrative can be obtained from the large corpus of narbs to obtain the narrative that the narbs collectively tell. In the analytic process used in the method proposed here, this process begins with the identification of narrative categories that make up the narrative. The term “narrative categories” refer to the different components of a story. These components have been identified by narrative analysts and could include elements such as the protagonists of the narrative and the roles the characters play, the actions of the characters, the feelings of the characters towards different issues that are addressed in the narrative (see, e.g., Barthe, 1975; Propp, 1968; Todorov, 1977). The process of creating the narrative from the narbs requires the extraction of the narrative categories from the large corpus of narbs, and then identifying the connection between the categories to see the story told by the narbs. This process begins with the machine reading of the narbs to help the researcher narrow down the narbs into a manageable number of narrative categories. This process can be done by many different computer programs with varying degrees of sophistication depending on the amount of resources available to the researchers. For instance, large supercomputers such as the machine called “Watson” developed by International Business Machines (IBM) can do very complex analysis of natural human language as demonstrated in the machine winning a game of Jeopardy against human players (Markoff, 2011). Ultimately the goal is the “funneling” of the numerous narbs into a manageable number of “categories.”

The funneling is a tiered process where the narbs are analyzed by the machine to create a set of concepts that are extracted from the narbs. The number of concepts could still be quite large, for instance thousands of narbs could yield hundreds of concepts. Unfortunately, narratives cannot be built with hundreds of concepts with each concept acting as a narrative category. It is next necessary to parse the concepts into a suitable number of narrative categories that become the building block of the eventual narrative. This process is an iterative process where different concepts are collapsed into a set of coherent narrative categories. The notion of coherence is important to note since it connects with a key concept in the development of the narrative paradigm as proposed by Fisher. The idea of coherence examines the chances that the story makes logical sense and whether it “hangs together” with all its elements connected to each other. For instance, concepts represented by terms such as “good,” “great” or “excellent” can be collapsed into a single narrative category such as “positive affect.” The categories themselves are also built with the notions of narrative structure in mind. As suggested earlier in this essay, narrative analysts look for specific narrative categories to analyze a story. For instance, Barthe offers the various narrative codes that one can seek in a narrative, Propp offers several narrative units that can be found in a story. While these strategies are useful, and often applied to the analysis of narratives, they can also be used to create specific narrative categories that mimic the codes and units used in the analysis of narratives. This is precisely the method used in creating the narrative categories where the coherent concepts are collapsed into narrative categories that represent the different form of units and codes of a narrative.

Once the categories are produced it is then possible to analyze the categories – which now represent the large volume of narbs originally collected – to see the connection between the categories. In the tradition of syntactic analysis of narratives, where one of the objectives is to seek the connection between different narrative

categories, a correlation can be found in the narrative categories. The strength and direction of the correlation offers a sense of the connection between the categories, and a story, based on the numerous narbs, can now be formed by considering what connections are contained in the utterances of the numerous voices that make up the narbs. In the next section I offer a set of examples of the way in which these stories can be constructed from narbs available in the public domain. In each case the data was collected in different ways and from different sources. The goal was, however, similar – to process the data using the technique used above, and then creating a visual representation of the data, along with a short narrative, to be made available on a public Web site called the “TheMediaWatch.com” which was developed for the specific purpose of the popularization and dissemination of the narrative analysis of narbs across many different contexts.

The visualization is primarily in the form of a narrative map. As stated on the Web site, a narrative map is:

Narrative maps are produced from analysis of narbs, where the maps are made up of specific nodes and connectors. The nodes represent attitudes, concepts, behaviors and different issues that people are talking about in their narbs. For instance a common node is “positive opinion” which could be made up of positively affected terms such as “like,” “love,” “good,” etc. Similarly, another node could be “immigration” referring to the issue. The map connects nodes together to indicate how strongly two nodes are related to each other. For instance, a dark line between two nodes would show a strong relationship between the nodes, suggesting people are talking about two things simultaneously, almost “in the same breath!”

It is important to note that the relative placement of the specific nodes that represent the narrative categories obtained from the narbs is of no consequence to the narrative. What matters

most is the visible *size* of the node – a larger circle represents that the narrative category appeared more frequently than others – and the *thickness* of the line connecting nodes where a dark and thicker line represents larger number of narbs that included the narrative categories connected by the line. All the examples used in this essay are drawn directly from the Web site, which also acknowledges the various contributors to the different narratives described here.

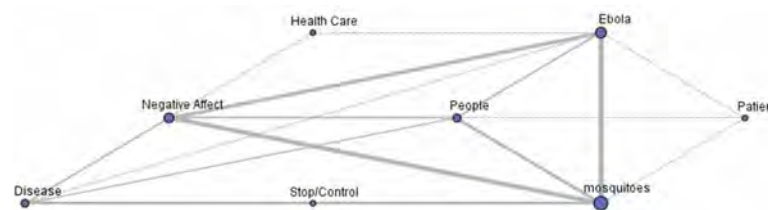
The Ebola Virus

The first example deals with the type of narbs where the word length of the narb is unrestricted and the narbs appear on the Facebook digital network. Since the March of 2014, there was a World-wide panic related to the spread of the virus in some countries in Western Africa. There was a significant amount of discussion within popular culture about the implications and threats posed by the virus as a global phenomenon. As such, it also generated a large amount of narbs in various digital spaces. For the analysis that was presented on the Web site narbs were obtained in response to the information about the Ebola virus that was presented by the World Health Organization (WHO) through their presence on the Facebook digital space. The comments were collected in the winter of 2014 as a build up to the declaration by the WHO that the spread of the virus would be contained in the December of 2014. The narbs analyzed here represented the comments made on the Facebook “page” of the WHO.

The analysis of the narbs was conducted exactly in the manner described earlier and it was possible to identify several narrative categories from the narbs. A total of 800 narbs were obtained resulting in the narrative map presented here and on our Web site (see Figure 1).

As the narrative map demonstrates, there were several relevant narrative categories that was already in the public sphere as points of discussion such as “mosquitoes” and the general notion of “disease.” This narrative map offered the oppor-

Figure 1.



tunity to produce a narrative that was reported on the Web site as shown in Figure 2.

This combination of the map and the narrative offered a visitor to the Web site a concise picture of the narrative contained in the narbs that were obtained from the comments available at a digital network site such as Facebook. The next example uses narbs created as micro-blogs by users in the digital space created by Tweets.

The Indian Elections of 2014

The second example uses narbs of limited word length by obtaining tweets related to specific authors on the Tweeter digital network. The object of analysis were the tweets of the journalists collected over nearly twelve months leading up

to the elections and the tweets of the followers of the journalists over the same time period. The Big Data in this case was made up of a large number of people as well as a long period of time. Such Big Data eventually allow for longitudinal analysis as well as the analysis segregated by specific time periods. There were several different points where the analysis was visualized and distributed on the Web site. This example shows one such presentation that happened closer to the final moment of voting in May 2014 (see Figure 3).

This is a more complex map because of the many different issues related to the elections, and the diversity of narrative categories that were extracted from the narbs. This complexity is also related to the fact that this map offers an analysis of a much larger volume of data compared to the

Figure 2.

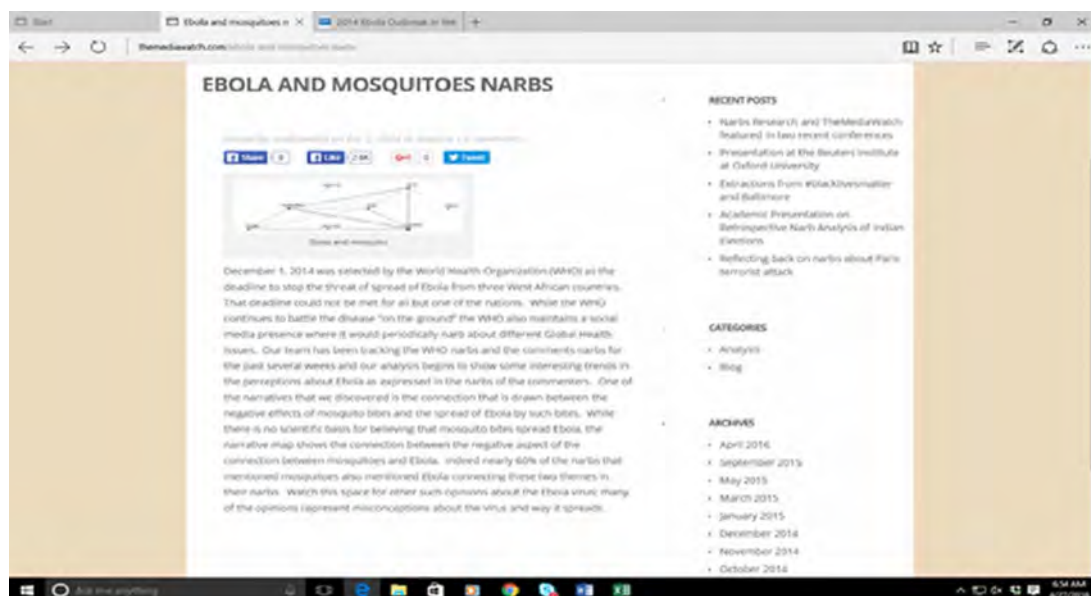
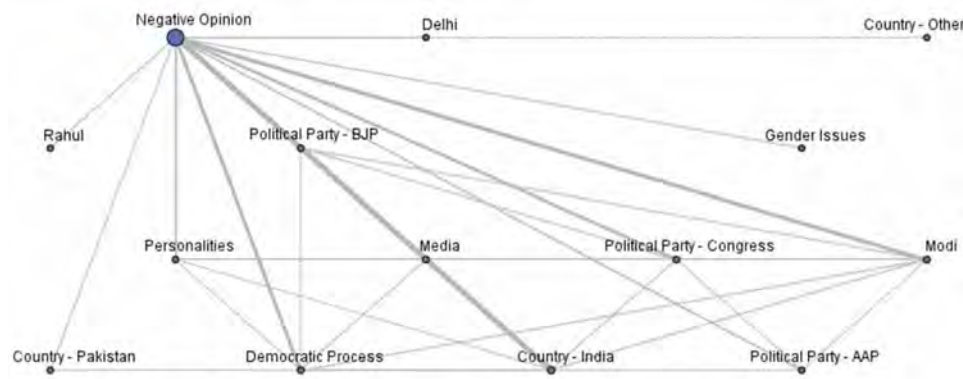


Figure 3.



previous example. The narrative that emerged from the data was also more complicated given the numerous elements that made up the narrative

One of the things to note about the narrative shown in Figure 4 is the fact that the narrative here is extracted not only from the narrative map, but the narrative also refers to specific computations of the percentages of different combination of narbs. The numeric analysis often adds to the value of the narrative, especially when the narrative might be used for specific policy decisions as demonstrated in the last example.

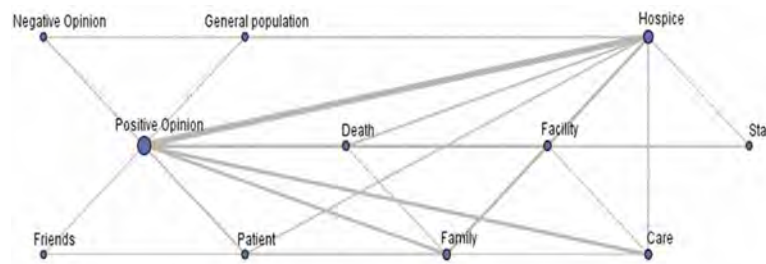
Hospice and Death

The Hospice Project used the third category of narbs where the narbs were solicited from respondents to a questionnaire that was distributed digitally. The respondents voluntarily answered “open-ended” questions about various elements of the medical system called “Hospice” which, in America, is often associated with “end of life” care and a general sense of “death.” The goal of this project was to discover what the respondents in the community thought about the connection

Figure 4.



Figure 5.



between the notion of death, and the idea of the Hospice (see Figure 5).

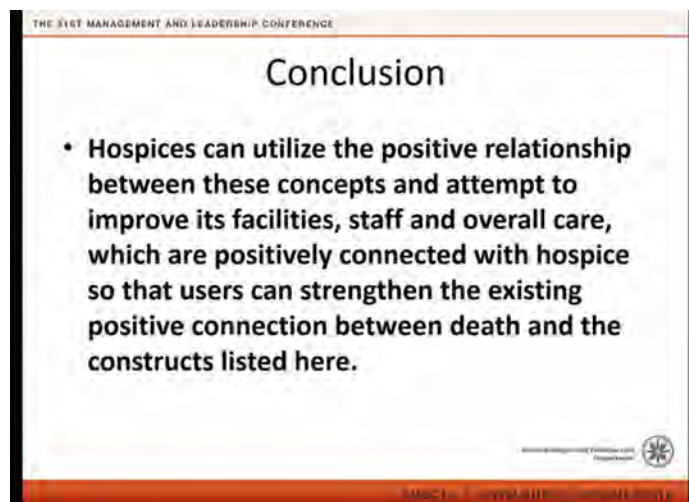
The map offered a glimpse into the opinions of the respondents expressed in the narbs that resulted in narrative categories that were quite unique, especially in the positive connection between the notion of death and Hospice, which ran counter to existing information. As a result the narrative was also provocative, resulting in some specific recommendations that were made to the Hospice and presented at the national conference (see Figure 6).

Indeed, the Hospice developed a new set of messages based on the narrative that was extracted from the narbs.

FUTURE RESEARCH DIRECTIONS

The argument presented here first suggests that it is possible to parse the overarching term – Big Data – in manageable sub-parts and each of these sub-parts offer different opportunities and challenges. Thus, the comments from the WHO presence on Facebook, the Tweets related to the elections in India and the self-response obtained through the distribution of a questionnaire all represent different aspects of Big Data and each can be obtained within the public sphere using means of data collection that are not extremely challenging as long as there is a clear sense of what kind of Big Data is being sought.

Figure 6.



The examples also addresses a second aspect of Big Data – the analysis of the material can be meaningful when there is a specific theoretical framework that informs the analysis. As long as there is reliance on a theory – in this case I use the narrative paradigm and the theory of narbs – it is possible to corral large amounts of information to answer specific research questions. The large volume of data is extremely seductive and it is often possible to continue to “drill” deeper into the structured aspect of Big Data trying to find connections and relationships without a specific conceptual framework driving the analysis. These examples illustrate the importance of theory in this process.

Finally, the examples also illustrate that once a theoretical approach is adopted it is possible to interpret the findings, such as the narrative maps, into coherent information that can be visualized and circulated in an understandable manner. Groups such as Infographics have been doing this with numeric data for some time, offering simple and attractive graphical representations of large amounts of data making it meaningful to most observers. It is far more difficult to do that with large amounts of discursive data which makes up the unstructured Big Data. This is the reason why different apparatuses of popular culture, for instance the mass media tools, focus on single narbs, such as a single Tweet, to make sweeping generalizations because capturing the essence of millions of narbs becomes far more difficult without a theoretical basis to support the finding. Here the collection of narbs is presented through the mechanisms of a narrative map and a simple narrative. These narratives could become much more powerful in impact than relying on specific small set of narbs to create a whole “news story.” The framework and the analysis offered could offer much more robust narratives.

CONCLUSION

In closing the role of www.themediawatch.com is to offer a way to grapple with the emerging volume of Big Data. It is likely that popular awareness and appreciation of the importance of Big data can be boosted as more of the findings can be presented on outlets like this Web site. These three examples are offered to demonstrate several aspects of the discussions related to the notion of Big Data in academia and popular culture in the late 2010s. The term has taken on a mythological power as it is used within many different circles (Daly, 2013). However, when specific applications of the term are sought within academia or popular culture, one is either faced with complex mathematics along with sophisticated computational discourse which remain inaccessible to many, making Big Data a mysterious construct whose utility is shrouded in incomprehensible language (see, e.g., Diebold, 2003; Lohr, 2013; Mayer-Schonberger and Cukier, 2013; Tien, 2013). On the other hand, sometimes in the regime of the popular, Big Data is presented as the answer to all future decision making processes as well the “bogey man” that will eventually snatch away all elements of personal privacy and drive us to the dystopian future as pictured by George Orwell in his novel, *1984*.

REFERENCES

- Barthes, R. (1975). *S/Z* (R. Miller, Trans.). New York, NY: McMillan.
- Daly, J. (2013). Quartz: “Big Data” Is the Most Overused Term of 2013. *FedTech*. Available at: <http://www.fedtechmagazine.com/article/2013/12/quartz-big-data-most-overused-term-2013>

- Dalziel, S. (January 14, 2011). IBM computer beats human Jeopardy champs. *The Inquirer*. Retrieved on June 3, 2013 from: <http://www.theinquirer.net/inquirer/news/1937273/ibm-beats-human-jeopardy-champs>
- Diebold, F. X. (2003). Big Data Dynamic Factor Models for Macroeconomic Measurement and Forecasting: A Discussion of the Papers by Reichlin and Watson. In M. Dewatripont, L.P. Hansen and S. Turnovsky (Eds.), *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress of the Econometric Society*. Cambridge, UK: Cambridge University Press.
- Facebook users worldwide 2016 | Statista. (n.d.). Retrieved May 16, 2013, from <http://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/>
- Fisher, W. R. (1984). Narration as Human Communication Paradigm: The Case of Public Moral Argument. *Communication Monographs*, 51(1), 1–22. doi:10.1080/03637758409390180
- Fisher, W. R. (1985a). The Narrative Paradigm: An Elaboration. *Communication Monographs*, 52(4), 347–367. doi:10.1080/03637758509376117
- Fisher, W. R. (1985b). The Narrative Paradigm: In the Beginning. *Journal of Communication*, 35(4), 74–89. doi:10.1111/j.1460-2466.1985.tb02974.x
- Fisher, W. R. (1987). *Human Communication as Narration: Toward a Philosophy of Reason, Value, and Action*. Columbia, SC: University of South Carolina Press.
- Fisher, W. R., & Filloy, R. A. (1982). Argument in drama and literature: An exploration. In J. R. Cox & C. A. Willard (Eds.), *Advances in argument theory and research* (pp. 343–362). Carbondale, IL: Southern Illinois University Press.
- Gadamer, H. G. (1982). *Truth and method*. New York, NY: Crossword Publishing.
- Lohr, S. (2013). The Origins of ‘Big Data’: An Etymological Detective Story. *The New York Times*. Retrieved on May 31, 2013 from: <http://bits.blogs.nytimes.com/2013/02/01/the-origins-of-big-data-an-etymological-detective-story/>
- Markoff, J. (2011). Computer Wins on ‘Jeopardy!’: Trivial, It’s Not. *New York Times*. Available at: http://www.nytimes.com/2011/02/17/science/17jeopardy-watson.html?pagewanted=all&_r=0
- Mayer-Schonberger, V., & Cukier, K. (2013). *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. New York, NY: Eamon Dolan/Houghton Mifflin Harcourt.
- Mitra, A. (1999). Characteristics of the WWW Text: Tracing Discursive Strategies. *Journal of Computer-Mediated Communication*, 5(1).
- Mitra, A. (2010). Creating a Presence on Social Networks via Narbs. *Global Media Journal*, 9(16).
- Mitra, A. (2011). Using narratives from social network to manage teacher-student interaction. In *Education in a technological world: communicating current and emerging research and technological efforts*. Formatex Research Center.
- Mitra, A. (2012a). Collective Narrative Expertise and Using the Narbs of Social Media. In T. Takseva (Ed.), *Social Software and the Evolution of User Expertise: Future Trends in Knowledge Creation and Dissemination*. Hershey, PA: IGI Global.
- Mitra, A. (2012b). Narbs as a Measure and Indicator of Identity Narratives. In Dudley et al. (Eds.), *Investigating Cyber Law and Cyber Ethics: Issues, Impacts and Practices*. Hershey, PA: IGI Global. doi:10.4018/978-1-61350-132-0.ch008
- Mitra, A. (2013). Mapping Narbs. In G. Wise (Ed.), *New Visualities, New Technologies: The New Ecstasy of Communication*. New York, NY: Ashgate Publishing Ltd.

Mitra, A. (2014a). *Using narbs unstructured Big Data analytics: Political analysis*. Paper presented at the 2014 E-Society Conference.

Mitra, A. (2014b). *Digital DNA: Managing Identity in Social Networking Sites*. New Delhi, India: Rupa Publications.

Mitra, A. (2014c). Narbs: A narrative approach to the use of Big Data. In *Communication Yearbook 38* (pp. 369–386). New York, NY: Routledge. doi:10.1080/23808985.2014.11679168

Mitra, A., & Cohen, E. (1998). Analyzing the Web: Directions and Challenges. In S. Jones (Ed.), *Doing Internet Research*. Newbury Park, CA: SAGE.

Propp, V. (1968). *Morphology of the Folktale* (2nd ed.). Austin, TX: University of Texas Press.

Tien, J. M. (2013). Big Data: Unleashing information. *Journal of Systems Science and Systems Engineering*, 22(2), 127–151. doi:10.1007/s11518-013-5219-4

Todorov, T. (1977). *The Poetics of Prose*. Ithaca, NY: Cornell University Press.

Zimmer, B. (n.d.). *The Year in Words, 2011 Edition*. Retrieved April 16, 2013, from <https://www.visualthesaurus.com/cm/wordroutes/the-year-in-words-2011-edition/>

ADDITIONAL READING

Fisher, W. R. (1987). *Human communication as narration: toward a philosophy of reason, value, and action*. Columbia, SC: University of South Carolina Press.

Hyde, M. J. (1985). Storytelling and public moral argument: The case of medicine. In J. R. Cox, M. O. Sillars, and G. B. Walker (Eds.), *Argument and social practice: Proceedings of the fourth SCA-AFA conference on argumentation* (pp. 364–375). Annandale, VA: SCA.

Mayer-Schönberger, V., & Cukier, K. (2014). *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Eamon Dolan/Mariner Books.

Mitra, A. (2014). *Digital DNA: Managing Identity in Social Networking Sites*. New Delhi, India: Rupa Publications.

Ricouer, P. (1984). Time and narrative: Vol. 1. K. McLaughlin & D. Pellaur, Trans. Chicago, IL: The University of Chicago Press.

KEY TERMS AND DEFINITIONS

Big Data: This is personalized data that is coming from people who are actively and voluntarily contributing to the compilation of these data sets. Much of the attention on big data has focused on the two key components: (1) gathering the large amounts of data and (2) quantitatively analyzing the data to obtain both personal-individualized information as well as information about different groups of people.

Data Visualization: The process of translating complex data and data interpretations into visual graphs and figures that capture the complexity in easy to understand and easy to communicate images.

Narb: As stated by the World Future Society on their Web site (<http://www.wfs.org/futurist/january-february-2013-vol-47-no-1/tomorrow-brief/wordbuzz-narbs>): “Narrative bits, or “narbs,” refer to small bits of information in the digital universe that, when collected, tell an otherwise untold story. The term is credited to Wake Forest University communication professor Ananda Mitra, who believes that narbs offer a way to turn massive amounts of social communication into a tool for predicting behavior and reactions.”

Narrative Map: Narrative maps are produced from analysis of narbs, where the maps are made up of specific nodes and connectors. The nodes represent attitudes, concepts, behaviors and dif-

ferent issues that people are talking about in their narbs. For instance, a common node is “positive opinion” which could be made up of positively affected terms such as “like,” “love,” “good,” etc. Similarly, another node could be “immigration” referring to the issue. The map connects nodes

together to indicate how strongly two nodes are related to each other. For instance, a dark line between two nodes would show a strong relationship between the nodes, suggesting people are talking about two things simultaneously, almost “in the same breath!”

B

Mining Big Data and Streams

Hoda Ahmed Abdelhafez

Suez Canal University, Egypt

INTRODUCTION

Mining big data is getting lot of attention currently because the businesses need more complex information in order to increase their revenue and gain competitive advantage. the growing of the telecommunication data traffic according to Cisco annual forecasting will reach 8.6 zettabytes by the end of 2018 up from 3.1 zettabytes per year in 2013 (Cisco analysis, 2014). Therefore, mining the huge amount of data as well as mining real-time data needs to be done by new data mining techniques/approaches. Big Data is a new term used to identify the datasets that are of large size and have grater complexity (Bifet, 2013). Data mining (DM) is the process of searching large volumes of data automatically for patterns such as association rules (Gupta et al., 2014). Big data mining is defined as the capability of extracting valuable information from large datasets or streams of data that due to its characteristics it is not possible before to do it (Fan & Bifet, 2013). This chapter will discuss the challenges of big data, new data mining techniques compared with traditional techniques and the main DM tools for handling very large datasets. Moreover, the chapter will focus on two industrial areas telecommunications and healthcare and lessons learned from them.

BACKGROUND

The data produced nowadays is estimated in the order of zettabytes, and it is growing around 40% every year. We need new algorithms, and new tools to deal with all of these data. Therefore,

the use of big data is becoming a crucial way for leading companies. For example, in healthcare, data pioneers are analyzing the health outcomes of pharmaceuticals when they were widely prescribed, and discovering benefits and risks that were not evident during necessarily more limited clinical trials (McGuire, 2012). We selected some significant articles that discussed challenges, techniques and tools for mining big data. Yadav et al. (2013) presented a review of several algorithms from 1994-2013 necessary for handling big data set. It gives an overview of architecture and algorithms used in large data sets, various tools that were developed for analyzing them as well as various security issues and trends. Bifet (2013) discussed data stream mining and how it offers many challenges and also many opportunities. Che et al. (2013) presented an overview of mining big data and its challenges include heterogeneity, scalability, speed, accuracy, trust, provenance and privacy. This paper also provides an overview of the platforms for processing and managing big data as well as platforms and libraries for mining big data. Jovic et al. (2014) discussed several data mining tools including RapidMiner, R, Weka, KNIME, Orange, and scikit-learn. Fan and Bifet (2013) presented big data challenges, applications of mining big data, Apache Hadoop and other open sources for big data mining and big graphic mining. Singh (2014) discussed machine learning techniques to capturing the value hidden in big data. He presented supervised learning using neural networks, Support Vector Machines (SVMs) and Naive Bayes classifiers, and also unsupervised learning using k-Means, hierarchical clustering and self-organizing maps.

CHALLENGES OF BIG DATA SYSTEMS

Big data has five key elements: Volume, Velocity, Variety, Veracity and value. These 5 V's are considered challenges of Big Data systems (Yin & Kaynak, 2015; Ishwarappa & Anuradha, 2015; Marr, 2015).

Volume refers to the huge amount of data. Many companies have large archived data in the form of logs but do not have the capacity to manipulate and analyze that data using traditional database technology. Now big data technology can help store and use these datasets in order to gain benefits from them.

Velocity represents the speed at which data generated and the speed at which data moves around. The speed at which credit card transactions is checked for fraudulent activities and the social media messages going to viral in seconds. Thus, big data technology can be used to analyze the data while it is being generated without putting it into databases.

Variety means different data types or format. Traditional database can store and process structured data that fit into tables such as financial data. Now 90% of data generated is in unstructured form and it cannot easily be put into relational databases such as photos, video sequences or social media updates. Big data technology can now harness various types of data like messages, photos, sensor data, and social media conversations and bring them together with more structured data.

Veracity refers to the trustworthiness of the data. The quality and accuracy of big data are less controllable because there will be dirty data. For instance, twitter posts with hash-tags, abbreviations, typos and colloquial speech. Big data analytics now allows us to work with these types of data. The volume, variety and velocity of data often make up for the lack of quality or accuracy.

Value refers to the ability to turn big data into value. Value is the most important aspects of big data because implement IT infrastructure systems are very costly to store big data and businesses are

going to require a return on investment. Big data can deliver value in almost any area of business or society such as improving healthcare and better understanding and serving customers.

There are other important challenges in big data management and analytics such as analytics architecture and hidden big data (Singh, 2014). Some key issues like accuracy and privacy are also very critical in mining big data (Che et al., 2013).

DATA MINING TECHNIQUES FOR LARGE SCALE DATA

The challenges in handling big data include capturing, storage, analysis, sharing, visualizing and more. In addition to connection and correlation of data which describes more about relationship among the data. Therefore mining big data needs new architecture, algorithms, techniques for its implementation. This section is focusing on the data mining techniques/methods that can be used for handling big data. These techniques/methods are classified classification, clustering, association rules, time series and data streams as follows:

Classification

One of the important classification techniques is decision tree. Decision tree learning is fast and accurate. Building Decision trees from large datasets requires long time for processing all the training instances and the available memory may be not sufficient for storing the whole training set. Implementation of traditional algorithms for building Decision trees becomes very time consuming. Therefore, many incremental algorithms are available such as BOAT (optimistic decision tree construction), ICE (implication counter examples) and VFDT (very fast decision tree) for handling large datasets (Franco-Arcega et al., 2013). The hybrid approach combining both decision tree and genetic algorithm are also used to create optimized decision tree in order to improve classification performance (Yadav et al., 2013). Another clas-

sification technique is Artificial Neural network (ANN). ANN is used for large data sets described the techniques of SOM (self-organizing feature map) network and LVQ (learning vector quantization networks). SOM reduces the dimensions of data through the use of self-organizing neural networks and it takes input in an unsupervised manner. The self-organizing map is a single layer feed forward network where the output syntaxes are arranged in low dimensional grid. LVQ uses supervised learning to categorize large data set into small sets in order to improve the overall computing time needed to process the large data set (Lu & Fahn, (2007).

Clustering

Clustering is a process of grouping objects with similar properties. The main clustering techniques for handling large data sets are Hierarchical clustering algorithms, K-means clustering algorithms, and Density Based Clustering (Vijayalakshmi & Renuka devi, 2012). Hierarchical clustering (HC) can be classified into two approaches agglomerative hierarchical clustering and divisive hierarchical clustering. In agglomerative approach, each data points are considered to be a separate cluster and the clusters are merged based on criteria. In divisive approach all data points are considered as a single cluster and they are splitted into number of clusters based on certain criteria. Two examples of hierarchical clustering algorithms are:

1. Clustering- Based SVM (support vector machine), which trains a very large datasets using the hierarchical micro clusters (Yu et al., 2003), and
2. Efficient hierarchical clustering using P-trees algorithm for handling multimedia-, stream- and spatial data for efficient data storage and access.

K-means clustering is another method of cluster analysis that aims to partition n observations into k clusters and each observation belongs to

the cluster with the nearest mean. The limitation is this method requiring the user to predefine K parameter for a clustering solution, which is often non-deterministic. To overcome this problem, parallel k/h -Means algorithm could use for clustering large distributed data sets; it is scalable and enlarges its field of application to clustering tasks. *Finally*, Density Based Clustering is another method identifying clusters in large high dimensional data set having different size and shape, and it is suitable for handling noise in dataset. DBSCAN and DENCLUE are examples for this algorithm as well as DESCRY a new algorithm to identify clusters in large high dimensional data set having different size and shape. DBSCAN is a density-based algorithm that treats clusters as dense regions of objects in the spatial space separated by regions of low density. DENCLUE is a highly efficient Density-based algorithm to deal with high dimensional datasets and it is faster than CLARANS (improved k -medoid method) and DBSCAN (Ding et al., 2015).

Association Rules

Association rules are used to discover facts that often occur together within a particular dataset. Many algorithms have been proposed in the past. For example, the Apriori algorithm is used for mining frequent itemsets for Boolean association rules. This algorithm is based on previous knowledge (a priori) of the sets of frequent data items. An item is considered frequent if its frequency of appearance in the database is higher than the minimum support threshold. The Apriori algorithm has big problems with large volume of Big Data. It does multiple scans of the whole database, thus the execution time increases in line with the number of transactions (Nedunchezian & Geethanandhini, 2016; Fernandez-Basso et al., 2016). Another algorithm is Eclat using bottom up approach to finds the itemsets from the dataset like depth first search. Eclat algorithm is very simple method to find the frequent item sets. It scans database only once. It is suitable for small datasets also it

requires less time for frequent pattern generation than apriori. Some new algorithms/methods are proposed for mining Big Data. BigFIM method is optimized to run on really large datasets. BigFIM, it is a hybrid method. It uses the Apriori algorithm to extract frequent itemsets of length k and later on it moves to Eclat when the projected databases fit in memory (Moens et al., 2013). Another new method is PARMA, it divides dataset into samples in order to give the better and fast results. PARMA is implemented using Map Reduce framework. It has two stages. In the first stage, N samples are created using first Map function and the samples are mined using first Reduce function. Second is the aggregation stage, all the results of samples are combined to get a whole set of frequent itemsets in a database (Riondato et al. 2012).

Time Series

Time series databases consist of sequence of events or values over repeated measurement of time. With the growing deployment of a large number of sensors and online data collection tools, the amount of time series data is increasing rapidly. Moreover, time series are a rich and important data source for detecting abnormal states, diagnosing performance issues such as finance, communication, automatic control, and online services, etc. (Mahajan et al., 2012).

Time Series Data Mining is a framework of analysis time series data for discovery and use of patterns as well as prediction of future values. Data in Time Series has lots of variations, therefore, both clustering and classification methods are used (Mahajan et al., 2012). Sampling techniques are adopted to reduce the size of time series dataset such as DENCLUE 2.0. It employs random sampling with different sample sizes to reduce the number of iterations for density attractor estimation. However, the sample size is chosen in an ad hoc manner in DENCLUE 2.0, and how to set proper sample size for arbitrary large-scale time series datasets is still difficult and unknown (Hinneburg and Gabriel, 2007). YADING, a

novel time series clustering algorithm is used to automatically cluster large-scale time series with fast performance and quality results. YADING consists of three steps: data reduction, clustering, and assignment. Data reduction reduces the dimensionality of the input time series instances (Ding et al., 2015).

Data Streams

Data stream means continuous flow of data; the size of the data is extremely large and potentially infinite. Examples of data stream include phone conversation, web searches and sensor data. Data stream mining is a process of extracting knowledge structure from continuous, rapid data records. There are two types of data Stream online streams and offline streams. Online data stream mining used in a number of real world applications such as fraud detection and network traffic monitoring; offline data stream mining is like generating report based on web log streams. Thus, increasing rate from such applications like sensor networks and call detail records need data stream real time analytics to manage data currently generated (Parikh & Tirkha, 2013). Mining data streams face great challenge in storage devices as well as the need for different techniques in order to extract patterns/knowledge (Trambadiya & Bhanodia, 2012).

There are many techniques for data stream classification; we will discuss decision tree algorithms for classifying data streams. These algorithms are the Hoeffding Tree, Very Fast Decision Tree method (VFDT) and Concept adaptive Very Fast Decision Tree method (CVFDT). Hoeffding tree provides incremental approach and scale different attributes better than tradition decision tree. It is capable of learning from massive data streams. Both Hoeffding Window Tree and Hoeffding Adaptive tree can cope with concept and distribution drift on data streams (Bifet & Gavalda, 2009). VFDT is based on Hoeffding tree algorithm. It use basic fundamental of a decision tree learning that is capable of learning from high speed data

streams in an incremental anytime fashion. The incremental approach in VFDT means where as a new example come it merged with old data. CVFDT is an extend VFDT to keep trees up to date with time variant data streams. It implements sliding window of various dataset to keep its model consistent. CVFDT continuous monitors the quality of new data and adjusts those that are no longer correct. Every time a new data arrives, the CVFDT incrementing counts for new data and decrementing counts for oldest data in the window (Trambadiya & Bhanodia, 2012, Hulten et al., 2001).

TOOLS FOR MINING BIG DATA

There are many data mining tools for handling large datasets. We will focus on the main open source tools that can handle big data. These tools are Mahout, RapidMiner, R, MOA, KNIME, SPMF, Orange, and SAMOA.

Apache Mahout is Scalable machine learning and data mining open source software based on Hadoop. It has implementations of a wide range of machine learning and data mining algorithms including clustering (such as k-means, fuzzy k-means and streaming k-means), classification (such as logistic regression and Naive Bayes), collaborative filtering (such as user based collaborative filtering and item based collaborative filtering) and frequent pattern mining. While Mahout's core algorithms are implemented on top Apache Hadoop using Map/reduce, it does not restrict contributions to Hadoop-based implementations. Contributions that run on non-Hadoop cluster or on a single node are also welcomed. (Apache Mahout, 2014).

RapidMiner is an open source and Java-based, general data mining tool currently in development by the company RapidMiner, Germany. The tool has become very popular in several recent years and has a large community support. RapidMiner offers an integrating environment with visually appealing and user-friendly GUI. Everything in

RapidMiner is focused on processes that may contain sub-processes. Processes contain operators in the form of visual components. Operators are implementations of data mining algorithms, data sources, and data sinks. RapidMiner provides data mining and machine learning procedures including: data loading and transformation, ETL (Extract, transform, load), data preprocessing and visualization, predictive analytics and statistical modeling, evaluation, and deployment. Although RapidMiner is quite powerful with its basic set of operators, it is the extensions that make it even more useful. Popular extensions include sets of operators for text mining, web mining, time series analysis, etc (Jovic et. al., 2014, RapidMiner Review, 2015).

R open source programming language designed for statistical computing and visualization. R is the successor of S, a statistical language originally developed by Bell Labs in 1970s. The source code of R is written in C++, Fortran, and in R itself. Interface to R is command line and use through scripting. Extension of R is implemented as an R library and provides a GUI to many of R's data analysis and modeling functions. From data mining user's perspective, R offers very fast implementations of many machine learning algorithms, comparable in number to RapidMiner. It has specific data types for handling big data, supports parallelization, web mining, data streams, graph mining, spatial mining, and many other advanced tasks (Jovic et al., 2014).

MOA (Massive On-Line Analysis) is open source software to perform data mining stream. MOA is related to WEKA, which is an award-winning open-source workbench containing implementations of a wide range of batch machine learning methods. It contains collection offline and online for both classification and clustering as well as tools for evaluation. In particular, for classification it implements boosting, bagging, and Hoeffding Trees, all with and without Naive Bayes classifiers at the leaves. For clustering, it implements StreamKM++, CluStream, ClusTree, Den-Stream, D-Stream and CobWeb. MOA is

written in Java so that the applications can be run on any platform with an appropriate Java virtual machine. MOA streams can be built using generators, reading ARFF files, joining several streams, or filtering streams. They allow for the simulation of a potentially infinite sequence of data. Examples of these generators are: Random Tree Generator, STAGGER Concepts Generator, SEA Concepts Generator, LED Generator, Waveform Generator and Function Generator (Bifet et al., 2010).

KNIME (Konstanz Information Miner) is a general purpose data mining tool based on the Eclipse platform. It is developed and maintained by the Swiss company KNIME.com AG. KNIME is an open-source, though commercial licenses exist for companies requiring professional technical support. The KNIME tool has building blocks called nodes, and than 1000 nodes are available through the core installation and various extensions. Nodes are organized in a hierarchy and can be searched by name within an intuitive interface. Each node performs a certain function, such as reading data, filtering, modeling, visualization. It is documented in detail and the nodes have input and output ports. Some node handle data model as classification trees. One of the greatest strengths of KNIME is the integration with WEKA and R. WEKA integration enables using almost all the functionality available in Weka as KNIME nodes, while R integration enables running R code as a step in the workflow, opening R views and learning models within R (Bifet, 2013; Jovic et al., 2014).

SPMF (Sequential Pattern Mining Framework) is an open-source data mining library implemented in Java. It is specialized in frequent pattern mining such as association rules and sequential patterns. SPMF offers implementations of more than 55 data mining algorithms. The source code can be integrated in other Java programs. Moreover, SPMF offers a command line interface and a simple graphical interface for quick testing. There are many general purpose open source data mining libraries such as Mahout and Knime that provide a wide range of data mining techniques. However, they offer a very limited set of algorithms for

frequent pattern mining. Knime and Mahout offer only a few popular pattern mining algorithms such as Apriori and GSP (Viger, 2014).

Orange is a Python-based tool for data mining and also an open source platform. It can be used either through Python scripting as a Python plugin, or through visual programming. Orange Canva offers the following functionalities grouped into nine categories: data operations, visualization, classification, regression, evaluation, unsupervised learning, association, visualization using QT (a cross platform development framework) and prototype implementations. Functionalities are visually represented by different widgets (such as read file, train SVM classifier etc). The visual programming environment of Orange uses graphical widgets that combine methods from the core library and associated modules to help users develop custom algorithms. The data mining algorithms of Orange are organized in hierarchical toolboxes, which make them easy to implement. Although, the downside of Orange is that the number of available widgets seems limited when compared to other tools (such as KNIME or RapidMiner), the coverage of standard data mining techniques is quite good (Jovic et al., 2014).

SAMOA (Scalable Advanced Massive Online Analysis) is an open source platform for mining big data streams and it is written in Java. Its aim is to satisfy the future needs for big data stream mining by combining the two approaches (streaming algorithms and distributed computing) in a single platform. It provides a collection of distributed streaming algorithms for data mining and machine learning tasks such as classification, clustering, and regression. For example, SAMOA includes a distributed version of CluStream, an algorithm for clustering evolving data streams. It also provides an API for algorithm developers that simplify implementing distributed streaming algorithms. It features a pluggable architecture that allows it to run on several distributed stream processing engines such as Storm, S4, and Samza (Morales & Bifet, 2015).

MINING BIG DATA IN TELECOMMUNICATIONS

Big data in telecommunication industry creates the need to analyze multiple data types including location data, social media (such as facebook and twitter), data from sensors and natural language text in order to provide insights that can enable them to increase revenues and reduce costs (Dam, 2013). Large volumes of data come from collecting millions of call details records per day, smartphones, tablets, personal computers, networks sensors and social media and the like. These data are in myriad formats and need to be analyzed in near real-time. Moreover, the volume of signaling data from smart phones, tablets and other devices that are application dependent has also increased significantly (informatica, 2012). The AT&T long distance data stream consists of approximately 300 million records per day from 100 million customers. Data in Time Series has lots of variations, as in telecom industry, some data sequences are long like billing data and each data item in a time series is a multidimensional vector. To overcome these difficulties two steps are followed: first, transform the data into equal-length vectors using a model based clustering method. Second, use standard classification method such as decision tree and/or SVM methods (Mahajan et al., 2012).

Fraud detection is important to the telecommunications industry because companies and suppliers of telecommunications services lose a significant proportion of their revenue as a result. Fraud in telecom can actually be viewed as fraud scenarios, which are related to the way the access to the network was acquired. Fraud detection focuses on the evaluation of different user profiles and their effect towards the proper discrimination between legitimate and fraudulent activity. The user daily profiles will be used as an input for the two algorithms K-means and the hierarchical clustering techniques called agglomerative clustering algorithm. The data of the basic profile vector for a telecommunications user are contained in the

Call Detail Record (CDR) of any Private Branch Exchange (PBX) or any VoIP switch. Mainly a CDR contains at least data such as the caller ID, the chargeable duration of the call, the date and the time of the call, etc. The experiments are based on real data extracted from a database that holds the CDR for a period of eight years from an organization's PBX. According to the organization's charging policy, only international and mobile destinations are charged. The k -means algorithm is used to partition the input space in two distinct groups. If the legitimate and the fraudulent behavior cases are sufficiently different from each other, then the k -means algorithm will provide two distinct clusters of data. If this is not the case, then the same input data will be fed into the agglomerative clustering algorithm to check whether there is an output with distinct cases separation (Hilas et al., 2015).

Churn is the movement of customers from one mobile network operator to another. Telecommunication faces an increasing churn rate compared with other industries. Increasing churn rate causes losing a high valued client means a loss of future incomes. Therefore, customer retention has become very important for marketing campaigns. It is more profitable for mobile telecom operators to invest in those customers that already have an experience with the service by renewing their trust, rather than trying to attract new customers characterized by a higher churn rate. A hybrid learning model built on data mining techniques is implemented to explain the churn behavior with more accuracy than single methods; and that in some extend the reason of churn can be revealed, as well as explaining the gap between the decision to churn and the deactivation time. This hybrid model is built using WEKA; it includes Logistic Regression (LR) in parallel with Voted Perception (VP) for classification, and combined with clustering. A real world dataset from an Asian mobile operator are used to evaluate the performance of the model. Two methods are combined in parallel to reinforce the classification. LR is used as major classifier and VP reinforce the prediction. The

Data preprocessing step transforms the selected input data in an appropriate format for analysis in WEKA knowledge flow. The evaluation of the model shows that its accuracy is higher than when using single model. The investigated case shows that four months period where enough for a churn to show his dissatisfaction before the deactivation; this depending on the Tenure and the data collection periodicity (Olle & Cai, 2014). Another framework for predicting customer behavior based on application hybrid learning approaches for churn prediction in mobile network. This framework used hybrid model including decision tree induction C4.5 and genetic algorithm to give a pragmatic churner model. The hybrid model is used for prediction of various user defined groupings based on usage time, location and their underlying social network (Yeshwanth, Raj, & Saravanan, 2011).

Transactional data streams are a challenging domain in which to apply data mining ideas because of a huge volume of simple records and a dynamic continuous flow of data. AT&T has approximately 5 million call detail records (CDRs) per day relating to international telecommunications traffic from 12 million accounts. A stream of CDRs associated with AT&T's wireless communications services includes approximately 80 million records per day from 15 million accounts. AT&T has developed signature-based methods for fraud and intrusion detection. Anomaly detection via signatures is used by AT&T for telecommunications fraud detection involving international calling. It uses anomaly detection to measure the unusualness of a new call specific to a particular account (or to a fraudster), and it uses a profile-based approach to characterize this unusualness as fraud. Finally, AT&T methodology can be fooled by fraud that is slowly introduced into an account's traffic. The good news is that this type of fraud will be caught by a secondary and highly distributed detection mechanism—namely the customers themselves—as itemized bills are sent out every month for payment (Cortes & Pregubon, 2001).

MINING BIG DATA IN HEALTHCARE

Data mining in healthcare industry is an important field for providing prognosis and a deeper understanding of medical data. Applications of data mining in healthcare include analysis of health care centers for better health policy-making and prevention of hospital errors as well as early detection of diseases. Health care industry today generates huge amounts of complex clinical data about patients and other hospital resources. Data mining techniques are used to analyze these rich collections of data from different perspectives and extracting useful information. (Mariamma et al., 2014; Venkatalakshmi & Shivsankar, 2014).

Mining data streams is required for real-time medical prediction and classification in clinical support systems. One of the main algorithms in clinical support system is decision tree for handling continuous rapid data streams. Very Fast Decision Tree (VFDT) an advanced decision tree algorithm that can handle large amounts of continuous data (data streams). The proposed system implements VFDT with the capability using pointers to allow the decision tree to remember the mapping relationship between leaf nodes and the history records. For each leaf nodes, it can be looked as a class label and it indicates a certain kind of medical situation (illness). Each leaf node has one or several pointers. These pointers are added by the learning algorithm in the training process. Each pointer has a unique value and each pointer represents a unique medical record. This way can save the need of the offline clustering process (Zhang et al., 2012).

Breast cancer is one of the major causes of death in women; therefore, the early detection and accurate diagnosis of this disease can ensure a long survival of the patients. Breast cancer research using data mining techniques has been one of the important topics in medical science during the recent years. Early diagnosis requires an accurate and reliable diagnosis procedure that allows physicians to distinguish benign breast tumors

from malignant ones. An experiment conducted on the Wisconsin Breast Cancer Dataset (WBCD) taken from UCI Repository of Machine Learning Databases. The dataset contains 699 instances taken from needle aspirates from patients' breasts, of which 458 cases belong to benign class and the remaining 241 cases belong to malignant. A hybrid model of decision tree based support vector machine (SVM) diagnostic system is implemented using Weka tool with Java platform for breast cancer diagnosis. Classification had two main phases, Training and Testing phases. The input parameters for SVM were optimized using decision tree algorithm. The SVM algorithm was used to classify breast cancer patients into one of two classes (Benign/Malignant). The advantage of decision tree-SVM method is that it provides higher accuracy than other compared algorithms (Sivakami, 2015).

Heart disease is a major health problem and it affects a large number of people. Cardiovascular Disease (CVD) is one such threat. To develop diagnosis and prediction system for heart diseases based on predictive mining, a number of experiments has been conducted to compare the performance of various predictive data mining techniques including Decision tree and Naïve Bayes algorithms. The dataset for this proposed work includes a 13 attribute structured clinical database from UCI Machine Learning Repository. This dataset consists of attributes and values. Decision tree and Naïve Bayes have been applied for the diagnosis of heart disease. The prediction of heart disease is executed with the help of a tool known as Weka. The results of this tool are the accuracy of how many patients are having the heart disease with in a particular time. In order to improve the efficiency and accuracy an optimization process is carried out using genetic algorithm. The genetic algorithm is being implemented with the help of Matlab. The optimized attributes are fed into Weka tool for the prediction purpose. The results show that the optimization technique is the best method for improving the prediction of heart disease. Benefit of using genetic algorithm

is the prediction of heart disease can be done in a short time with the help of reduced dataset. (Venkatalakshmi & Shivsankar, 2014).

FUTURE TRENDS

The next research will focus more on different big data mining tools and illustrate strengths and weaknesses through making comparison among them. The next research will also discuss big data mining in other industrial areas such as retailing and banking and finance.

CONCLUSION

Mining big data has a lot of challenges because of the scale, complexity and heterogeneous of datasets. Thus the new techniques and tools for mining large dataset have been tried to handle these challenges. This chapter discussed the data mining techniques/ algorithms like advanced decision tree and hierarchical clustering algorithms that are able of handling massive datasets. Some of open source tools for mining big data are also discussed like WEKA, MOA and Apache Mahout and how these tools help telecom and healthcare industries to extract useful information. Mining big data in telecommunication and healthcare industries represents big opportunities in these industries. In healthcare, mining massive amounts of complex clinical data using advanced algorithms can provide a deeper understanding of medical data and help saving patients lives. Telecommunication industry also deals with millions of call details records per day as well as data from smartphones, tablets, personal computers, networks sensors and social media and the like. Mining these data can help telecom detecting fraudulent and improving customer retention to reduce churn rate as well as improving the customer experience. So we can conclude that mining big data can help the companies to understand their customers, driving their businesses forwards and increasing their revenue streams.

REFERENCES

- Analysis, C. (2014). *Cisco global cloud index: forecast and methodology 2013-2018*. White Paper. Retrieved from <https://www.terena.org/mail-archives/storage/pdfVVqL9tLHLH.pdf>
- Bifet, A. (2013). Mining big data in real time. *Infomatica.*, 37, 15–20.
- Bifet, A., & Gavalda, R. (2009). Adaptive Parameter-free Learning from Evolving Data Streams. *IDA: Proceedings of 8th International Symposium on Intelligent data Analysis in the series lecture notes in computer science*, 5772. 249-260.
- Bifet, A., Holmes, G., Kirkby, R., & Pfahringer, B. (2010). MOA: Massive online analysis. *Journal of Machine Learning Research*, 11, 1601–1604.
- Che, D., Safran, M., & Peng, Z. (2013). From Big Data to Big Data Mining: Challenges, Issues, and Opportunities. DASFAA workshops. *LNCN*, 7827, 1-15.
- Cortes, C., & Pregubon, D. (2001). Signature-Based Methods for Data Streams. *Data Mining and Knowledge Discovery*, 5(3), 167–182. doi:10.1023/A:1011464915332
- Dam, R. (2013). Big data a sure thing for telecommunications: telecom's future in big data. *Proceeding of International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery*. doi:10.1109/CyberC.2013.32
- Ding, R., Wang, Q., Dang, Y., Fu, Q., Zhang, H., & Zhang, D. (2015). YADING: Fast Clustering of Large-Scale Time Series Data. *Proceedings of the VLDB Endowment.*, 8(5), 473–484. doi:10.14778/2735479.2735481
- Fan, W., & Bifet, A. (2013). Mining big data: Current status and forecast to the future. *SIGKDD Explorations*, 14(2), 1–5. doi:10.1145/2481244.2481246
- Franco-Arcega, A., Carrasco-Ochoa, J., Sánchez-Díaz, G., & Martínez-Trinidad, J. (2013). Decision Tree based Classifiers for Large Datasets. *Computación y Sistemas.*, 17(1), 95–102.
- Gupta, R. (2014). Journey from data mining to Web Mining to Big Data. *International Journal of Computer Trends and Technology*, 10(1), 18–20. doi:10.14445/22312803/IJCTT-V10P104
- Gupta, R., Gupta, S., & Singhal, A. (2014). Big Data: Overview. *International Journal of Computer Trends and Technology*, 9(5), 266–268. doi:10.14445/22312803/IJCTT-V9P150
- Hilas, C., Mastorocostas, P., & Rekanos, I. (2015). Clustering of Telecommunications User Profiles for Fraud Detection and Security Enhancement in Large Corporate Networks: A case Study. *An International Journal Applied Mathematics & Information Sciences.*, 9(4), 1709–1718.
- Hinneburg, A., & Gabriel, H. (2007). DENCLUE 2.0: Fast Clustering Base on Kernel Density Estimation. *IDA*.
- Hulten, G., Spencer, L., & Domingos, P. (2001). Mining time-changing data streams. *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 97-106.
- Infomatica. (2012). *Big data for telecommunications industry*. Retrieved from https://www.infomatica.com/content/dam/infomatica-com/global/amer/us/collateral/executive-brief/big-data-telecommunications_ebook_2190.pdf
- Ishwarappa, K., & Anuradha, J. (2015). A brief introduction on big data 5Vs. characteristics and Hadoop technology. *Proceeding of International Conference on Intelligent Computing, Communications & Convergence*, 319 – 324. doi:10.1109/MIPRO.2014.6859735
- Joseph M. (2013). Data mining and business intelligence applications in telecommunication Industry. *International Journal of Engineering and advanced Technology*, 2(3), 525-528.

- Jović, A., Brkić, K., & Bogunović, N. (2014). An overview of free software tools for general data mining. *Proceeding of 37th International Convention on Information & Communication Technology Electronics & Microelectronics*, 1112-1117.
- Lu, Y., & Fahn, C. (2007). Hierarchical Artificial Neural Networks for recognizing high similar large data sets. *Proceedings of the sixth International Conference on Machine Learning and Cybernetics*, 1930-1935. doi:10.1109/ICMLC.2007.4370463
- Mahajan, R., Thangavelu, A., & Shahakar, M. (2012). Data Mining Techniques for Identifying Temporal Patterns of Time Series Data. *Journal of Engineering Research and Applications*, 2(6), 185-187.
- Mahout, A. (2014). Retrieved from <http://mahout.apache.org>
- Mariammal, D., Jayanthi, S., & Patra, P. (2014). Major Disease Diagnosis and Treatment Suggestion System Using Data Mining Techniques. *International Journal of Advanced Research in Computer Science & Technology*, 2(1), 338–341.
- Marr, B. (2015). *Why only one of the 5 Vs of big data really matters*. Retrieved from <http://www.ibmbigdatahub.com/blog/why-only-one-5-vs-big-data-really-matters>
- Moens, S., Aksehirli, E., & Goethals, B. (2013). Frequent itemset mining for big data. *Proceeding of International Conference on Big Data*, 111–118. doi:10.1109/BigData.2013.6691742
- Morales, G., & Bifet, A. (2015). SAMOA: Scalable Advanced Massive Online Analysis. *Journal of Machine Learning Research*, 16, 149–153.
- Nedunchezian, R., & Geethanandhini, K. (2016). Association Rule Mining on Big Data – A Survey. *International Journal of Engineering Research and Technology*, 5(5), 42–46.
- Olle, G., & Cai, S. (2014). Mature Mobile Operators (MO) despite possessing a churn in Mobile telecommunication industry. *International Journal of e-Education, e-Business, e- Management Learning*, 4(1), 55–62.
- RapidMinerReview. (2015). Retrieved from <http://butleranalytics.com/rapidminer-review/>
- Singh, J. (2014). Big data analysis and mining with machine learning algorithm. *International Journal of Information and Computation Technology*, 4(1), 33-40.
- Sivakami, K. (2015). Mining Big Data: Breast Cancer Prediction using DT - SVM Hybrid Model. *International Journal of Scientific Engineering and Applied Science*, 1(5), 418–429.
- T., & Kumar, V. (2013). Application of big data in data mining. *International Journal of Emerging Technology and advanced Engineering*, 3(7), 390-393.
- Trambadiya, T., & Bhanodia, P. (2012). A comparative study of stream data mining Algorithms. *International Journal Engineering and Innovation Technology*, 2(3), 149–154.
- Venkatalakshmi, B., & Shivsankar, M. (2014). Heart Disease Diagnosis Using Predictive Data Mining. *International Journal of Innovative Research in Science. Engineering and Technology*, 3(3), 1873–1877.
- Viger, P., Gomariz, A., Gueniche, T., Soltani, A., Wu, C., & Tseng, V. (2014). SPMF: A Java Open-Source Pattern Mining Library. *Journal of Machine Learning Research*, 15, 3569–3573.
- Vijayalakshmi, M., & Renuka Devi, M. (2012). A Survey of Different Issue of Different clustering Algorithms Used in Large Data sets. *International Journal of Computer Science and Software Engineering*, 2(3), 305-307.
- Yadav, C., Wang, S., & Kumar, M. (2013). Algorithm and Approaches to handle large Data- A Survey. *IJCSN International Journal of Computer Science and Network*, 2(3), 277–5420.

Yadav, C., Wang, S., & Kumar, M. (2013). Algorithm and Approaches to handle large data-survey. *International Journal of Computer Science and Network*, 2(3).

Yeshwanth, V., Raj, V., & Saravanan, M. (2011). Evolutionary Churn Prediction in Mobile Networks Using Hybrid Learning. *Proceedings of the 24th International Florida Artificial Intelligence Research Society Conference*, 471-476.

Yin, S., & Kaynak, O. (2015). Big Data for Modern Industry: Challenges and Trends. *Proceedings of the IEEE*, 103(2), 143–146. doi:10.1109/JPROC.2015.2388958

Yu, H., Yang, J., & Han, J. (2003). Classifying Large Data Sets Using SVMs with Hierarchical Clusters. SIGKDD '03, Washington, DC. doi:10.1145/956750.956786

Zhang, Y., Fong, S., Fiaidhi, J., & Mohamed, S. (2012). Real-Time Clinical Decision Support System with Data Stream Mining. *Journal of Biomedicine and Biotechnology*, 1-8. doi:10.1155/2012/580186

ADDITIONAL READING

Hall, L., Chawla, N., & Bowyer, R. (1998). Decision tree learning on very large data sets. *IEEE Conference on Systems, Man and Cybernetics*. doi:10.1109/ICSMC.1998.725047

Mahajan, R., Thangavelu, A., & Shahakar, M. (2012). Data Mining Techniques for Identifying Temporal patterns of time series data. *International Journal of Engineering Research and Applications*, 2(6), 185–187.

Mukherjee, R. (2014), “Big data mine for revenue” Voice & Data. www.mobileum.com/images/news/2014/big-data.pdf

Musolesi, M. (2014). Big mobile data mining good or evil? *IEEE Internet Computing*, published by the IEEE. *Computers & Society*, 78–81.

Parikh, D., & Tirkha, P. (2013). Big mining and data stream mining: Open Source tools. *International Journal of Innovative Research in Science Engineering and Technology*, 2(10), 5234–5239.

Weiss, G. (2005forthcoming). Data mining in Telecommunication. In O. Maimon & L. Rokach (Eds.), *Data mining & Knowledge discovery handbook: A complete guide for practitioners and research*. Kluwer Academic publisher. doi:10.1007/0-387-25465-X_56

Wu, X., & Zha, X. (2014). Data mining with big data. *IEEE Transactions on Knowledge and Data Engineering*, 26(1), 97–107. doi:10.1109/TKDE.2013.109

KEY TERMS AND DEFINITIONS

Big Data: The datasets that are of large size and have a greater complexity.

Data Mining: The process of searching large volumes of data automatically for patterns.

Data Stream: A continuous flow of data.

Mining Data Stream: The ability to extract valuable information from large data streams.

Category B

Bioinformatics

Bioinformatics

B**Mark A. Ragan***The University of Queensland, Australia*

INTRODUCTION

Over the past 30 years, bioinformatics has emerged as new discipline at the interface of molecular bioscience with mathematics, computer science and information technology. Bioinformatics is driven by data arising from high-throughput technologies in molecular bioscience including DNA and genome sequencing, gene expression analysis, protein and RNA structure characterisation, and bio-imaging. To enable biological discovery, bioinformatics draws on and extends technologies for data capture, management, integration and mining, computing, and communication technology including the Internet. The rise of genomics, from the first bacterial and model-organism projects to the Human Genome Projects and the thousands of genome projects that have followed, has been a key driver for bioinformatics, and in turn has enabled these projects to be completed and their results applied. Genomics, however, was never an end unto itself, but rather was intended to enable the understanding of complex biological systems. Bioinformatics continues to evolve in support of its constituent domains and, increasingly, their integration into genome-scale molecular systems biology.

This article presents bioinformatics first from the perspective of computer science and IT, then from the perspective of bioscience. In practice these perspectives often merge, making bioinformatics a rich, vibrant area of multidisciplinary research and application.

BACKGROUND

The term *bioinformatics* was introduced in 1970 in reference to the study of informatic processes in biological systems (Hogeweg, 2011). In this original usage, *bioinformatics* encompassed “how living systems gather, process, store and use information” (Nurse, 2008). Never widely adopted, this usage was superseded in the late 1980s when bioinformatics, as presently understood, emerged as a new field at the interface of molecular bioscience with computer science and information technology (Dickson, 1987). Today bioinformatics builds on mathematics, statistics and algorithmics, and finds applications across the biosciences particularly in genomics, proteomics, structural biology and molecular systems biology. Biology is increasingly an information science, with bioinformatics a key enabling technology.

Other disciplines have developed at the bioscience - computer science - IT interface, and there is little consensus on where boundaries should be drawn among them. Bioinformatics is sometimes said to focus on the development and application of methods and software tools to acquire, manage, analyse and/or visualise biological data, whereas *computational biology* refers more to the application of these methods and tools to theoretical or applied biological questions (Huerta *et al.*, 2000). *Biomathematics* or *mathematical biology* involves the development or use of mathematical modeling or simulation, while *biostatistics* emphasises experimental design and statistical analysis. *Mo-*

lecular systems biology focuses on the inference or analysis of networks of genes, proteins and/or other cellular molecules, while *synthetic biology* applies these technologies to design and engineer new biological functions or organisms.

BIOINFORMATICS FROM THE PERSPECTIVE OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY

One way of exploring the interface between molecular bioscience and IT is to track experimental data from their generation, capture and retrieval, through their aggregation and dissemination *via* international data services, to their subsequent analysis. Here I deconstruct data analysis into data models, algorithms, analytical methods and software, workflows and visualisation. This trajectory is common to scientific data, although bioinformatics is notable for its culture of open data, well-established data formats and standards, and data reuse facilitated by large international repositories with associated data services.

Data Generation, Storage and Retrieval

Instruments and experiments generate diverse data types in molecular bioscience. Capturing these primary data and the associated metadata, and managing their storage and retrieval, are primary activities in bioinformatics. The quantities of data generated by DNA-sequencing platforms, in particular, are such that raw data are no longer archived; rather, bioinformatic methods are used to assess quality and extract summaries. Data formats are specific to experimental technologies and, to some extent, instrument manufacturers. In some areas of molecular bioscience, standards have been developed to ensure that data can be interpreted unambiguously and, in principle, the experiment can be reproduced. For example, the MIAME (Minimum Information About a Mi-

croarray Experiment) standard (Brazma *et al.*, 2001) specifies how experimental design, laboratory protocols, biological samples, microarray platforms, and raw and processed data must be described, and recommends the use of certain data formats and ontologies. The MIBBI project promotes the coherence of minimum-standards checklists across nearly 40 areas of bioscience and biomedicine (Taylor *et al.*, 2008).

Public Data Resources

Newly generated biomolecular data (*e.g.* DNA and protein sequences, protein structures, gene-expression data) are submitted to public data repositories, where they are assigned unique persistent identifiers; all major bioscience journals require new data to be so identified. The main international collection centres are the US National Centre for Biotechnology Information (NCBI), the EMBL European Bioinformatics Institute (EBI), and the DNA Data Bank of Japan (DDBJ). Individually and in collaboration with each other, these centres carry out further quality control on incoming data, conduct research in bioinformatics, promote best practice in bioscience data management and analysis, and provide comprehensive online data services (*e.g.* search, retrieval, integrative analyses over multiple data sources, and links to journal articles and patents) which are cost-free at the “point of use” for the international research community. Other public data resources serve specific areas of molecular bioscience, *e.g.* protein structure. Increasingly, the largest projects in molecular bioscience maintain their own public data resources, *e.g.* The Cancer Genome Atlas and the International Cancer Genome Consortium. Both large and small data sources are reviewed in the annual Database issue of the journal *Nucleic Acids Research*.

Data Formats and Models

The need for data integration and re-use has driven the development of standard data presentation

formats, notably the FASTA format (Lipman & Pearson, 1985). However, no single data model has been universally adopted across all bioinformatics applications. Many public data services provide flat (ascii) files, but relational (MySQL), semantic (RDF), Web Services and hybrid approaches are also in use. Third-party tools such as SRS or BioMart are often used to integrate and index multiple related collections for combined use. With public data now in the tens of petabytes and growing rapidly, bioinformatics has entered the era of Big Data.

Algorithms and Computation

The molecular biosciences offer diverse and difficult analytical challenges, against which a wide range of algorithmic approaches are deployed. Gene and protein sequences map naturally to strings, regulatory signals to motifs, phylogenies to trees, lateral genetic transfer to edits on trees, genetic regulatory networks and protein interactions to networks, and so on. Thus operations in bioinformatics can often be recast as known or new problems in *e.g.* string matching, motif discovery, classification or graph theory. Many problems mapped in this way are NP-hard, *e.g.* maximum clique, vertex cover and Steiner tree (Karp, 1972), or are suspected of being so. Increasingly, however, many are fixed-parameter tractable, allowing high-quality solutions even on very large bioscience data. Heuristics are important in bioinformatics as well: the sequence alignment problem drove the early development of dynamic programming, phylogenetics is an important application domain for Markov chain Monte Carlo in conjunction with Bayesian approaches, and problems in gene regulation, protein localisation and biomolecular networks continue to provide challenges for machine learning. Even so, key problems in bioinformatics continue to pose algorithmic and computational challenges for large data, requiring computational infrastructure with large shared memory and/or high-capacity input/output.

Methods and Software

Innumerable bioinformatic methods and software exist. Many are developed for highly specific applications, and quickly fall out of use or are superseded. Stand-alone tools, available by download from websites, remain popular; other tools are available *via* public servers or webtools, or integrated into online data services. Coordinated sets of tools that share input/output formats and a common user interface are popular in some application areas, *e.g.* DNA sequence analysis (DNASTAR, Sequencher), motif discovery (MEME Suite) and phylogenetics (PAUP*, PHYLIP). The great majority of bioinformatic tools arise from academic research, with the attendant issues of stability, support and sustainability. Although bioinformatics is characterised by a robust open-source ethic and active developer/user communities exist around many toolsets, commercial software is important in some areas, notably structural biology, proteomics and bio-image informatics. Notices of new methods and software may appear in academic journals (*e.g.* the Application Notes section of *Bioinformatics*; the annual Web Server issue of *Nucleic Acids Research*).

Bioinformatic Workflows

An important class of tools supports the composition and execution of bioinformatic *workflows* or “pipelines”, *i.e.* linked series of data-manipulation and analytical steps. For example, software modules that retrieve and reformat sequences, align multiple sequences, and infer trees can be linked to form a phylogenetic workflow. Popular workflow-management frameworks in bioinformatics include BioConductor/R, Galaxy, GenePattern, GMOD, SRS and Taverna. E-research middleware that links data sources, analytical and visualisation tools, computational resources and storage are gaining popularity particularly for large-scale projects.

Data Visualisation

Visualisation of large bioscience data is likewise an important theme in bioinformatics. Challenges include the visual representation of very large, complex, multi-scale data; integration of visualisation with data mining and analysis; integrated display of different data types (*e.g.* pathways and gene-expression levels); standards; and augmented computer interaction (O'Donoghue *et al.*, 2010). Genome browsers (*e.g.* the ENSEMBL, IGV and UCSC Genome browsers) allow simultaneous display of one or more genome sequences together with additional experimental data (*e.g.* expression levels) and computationally based annotation. Stand-alone visualisation tools remain popular for defined tasks, *e.g.* visualisation of molecular structures (PyMol) and display of phylogenetic trees (TreeView).

BIOINFORMATICS FROM THE PERSPECTIVE OF BIOSCIENCE

Comprising the other side of the bio – IT interface are the multiple fields of modern molecular bioscience, arranged here in reference to the primary direction of information flow in biological systems, from DNA (genomics) to RNA (transcriptomics) to protein (proteomics). To a first approximation, DNA/genome bioinformatics focuses on the primary (linear) sequence, whereas protein informatics is equally or more concerned with three-dimensional folded structure. There is increasing appreciation that biological information flow is not unidirectional; the analysis of biomolecular networks and systems is one of the fastest-growing areas of today's bioinformatics.

Individual Molecular Sequences

DNA, RNA and protein sequences represent the largest application area of bioinformatics. Important problems in *sequence bioinformatics* include quality control and management of se-

quence data; assembly of full-length sequences; discovery and annotation of features (*e.g.* genes, regulatory signals); sequence comparison in functional and evolutionary contexts; and retrieval of similar sequences from databases. Some of the earliest bioinformatic algorithms and software were designed to manage and analyse individual sequences.

Genome Informatics

With progress in automated technologies, it is becoming commonplace to sequence the genetic material of an individual – the coding regions only (*exome sequencing*), or *whole-genome sequencing*. Many earlier methods have been repurposed, and new approaches developed, to assemble, annotate and analyse plant and animal genomes, which are typically 10^8 - 10^{11} basepairs in size and replete with low-complexity regions and repetitive elements. In *metagenomics*, these methods are applied to study DNA extracted directly from environmental samples, without prior isolation of individual species. Continuing innovation in sequencing technologies, including non-hierarchical (“shotgun”) strategies, short-read and single-molecule sequencing, and new types of data encoding, has been made possible by innovative bioinformatics.

Feature Discovery and Annotation

For a sequence to be useful in biological investigation, features represented therein must be discovered and annotated. Early discovery of features involved painstaking laboratory work; today, bioinformatic methods that draw on the body of experimental knowledge are the mainstay of feature annotation. Some features (*e.g.* transcriptional start and stop signals for bacterial genes) are conservative and can be found efficiently by similarity matching and/or rule-based inference, whereas others (*e.g.* transcription factor binding sites) are based on motifs of variable position and low information content, and can be discovered

(if at all) only by statistically based approaches including neural networks, hidden Markov models, support vector machines or other machine-learning approaches (Libbrecht & Noble, 2015). The MEME Suite (Bailey *et al.*, 2015) provides important programs for motif discovery and analysis. The annotation process is well-suited to a workflow-based approach.

Sequence Comparison

Many operations in bioinformatics involve the statistical comparison of two or more sequences. Historically, these problems introduced into bioinformatics important concepts from computer science including dynamic programming, hashing, and suffix trees. Comparison problems can be classified as pairwise, one-*versus*-many (*e.g.* finding best matches in a database), and many-*versus*-many (*e.g.* multiple sequence alignment). The heuristic Basic Local Alignment Search Tool (BLAST) is by far the most widely used bioinformatic algorithm for database searching (Altschul *et al.*, 1990); variants are available *e.g.* to match a nucleotide sequence against a nucleotide database, a protein against a protein database or a nucleotide sequence against a protein database, or preferentially to recover weak matches (more distantly related sequences). Software for multiple sequence alignment includes Clustal, MUSCLE, PRANK and T-Coffee, while bacterial genomes can be aligned with Mauve.

Phylogenetics

Although originally developed as subfields within evolutionary biology, *molecular phylogenetics* and its genome-scale counterpart, *phylogenomics*, are widely accorded a place in bioinformatics conferences, journals and texts. Important operations include database matching, clustering, multiple sequence alignment, and the inference and comparison of phylogenetic trees and networks. The main algorithmic frameworks for tree inference include distance (*e.g.* neighbour-joining),

parsimony, maximum likelihood and Bayesian inference, the latter using Markov chain Monte Carlo (MCMC) to estimate posterior distributions of model parameters. Important software includes the PHYLIP package, PAUP*, RaxML, BEAST and MrBayes (Yang & Rannala, 2012). As multiple sequence alignment and most methods of tree inference scale poorly with data size, heuristics are widely employed and alternatives that do not require multiple sequence alignment are being actively explored (Chan *et al.*, 2014). TreeBASE and the Ribosomal Database Project (RDP) are important repositories of phylogenetic models and data.

Gene Expression

The development of bioinformatic methods for analysis of gene expression has been driven by technology: the introduction of spotted arrays in the 1980s, microarrays of increasingly high density since 1995, and more recently sequence-based transcriptomics. Operations in *expression bioinformatics* include experimental design, data capture and management, normalisation (statistical correction for variation arising from the array platform rather than from the biological sample), analysis of differential expression (*e.g.* under different experimental conditions), clustering, classification and functional interpretation. Specialised methods continue to be developed for the study of alternative splicing, other types of transcriptional variation, and regulatory RNAs. Both commercial (GeneSpring) and free (LIMMA; programs in BioConductor/R and GenePattern) software is widely used to analyse expression data.

Epigenetics

Epigenetics refers to heritable states in gene expression or phenotype caused by mechanisms other than changes to the sequence of genomic DNA. Large-scale epigenetic data have recently begun to appear with the advent of technologies that probe the methylation status of DNA,

chemical modifications of histones, structural conformation of the chromatin and occupancy of control sites in specific cell types and/or under specified conditions. Bioinformatic operations in epigenetics are not unlike those in genome analysis and annotation, with emphasis on statistically based approaches including machine learning. ENCODE (Encyclopaedia of DNA Coding Elements) is a very large consortium project whose aim is to assess the activity and function of the entire genomic DNA, not only protein-coding genes (Ecker *et al.*, 2012).

Structural Bioinformatics

Structural bioinformatics focuses on the computational analysis of the three-dimensional structures of biomolecules, especially proteins and RNAs. Important activities include the description or prediction of protein structures, surfaces, dynamics and interactions with other molecules including antibodies and drugs (Gu & Bourne, 2009). Important software packages include Amber, Gaussian, GROMACS and NAMD. The Critical Assessment of Techniques for Protein Structure Prediction (CASP) is a biennial community-wide competitive evaluation of methods and software, while Folding@Home is a distributed computing project for protein folding and computational drug design. For RNAs, recurrent themes include the prediction of folded structure and target specificity. Software includes the Vienna package, UNAFold, miRanda, PicTar and TargetScan. Databases include PDB, Swiss-Prot, UniProt (proteins) and Rfam (RNA).

Systems Bioinformatics

Systems bioinformatics aims to understand cells or organisms as integrated molecular systems. Typical operations include management and integration of multiple types of genome-scale data; inference of biomolecular interactions, pathways (*e.g.* of

biosynthesis or metabolism) or networks (*e.g.* of genetic regulation or protein-protein interaction); and computational modelling, sometimes at very large scale. Models may be static or dynamic, and may focus on mechanistic details or phenotypic outcomes (Csermely *et al.*, 2013). Important data sources include the Kyoto Encyclopedia of Genes and Genomes (KEGG) and Reactome. The Dialog for Reverse Engineering Assessments and Methods (DREAM project) is a CASP-like competitive evaluation in computational systems biology.

Ontologies

The Gene Ontology models knowledge in molecular and cell biology in three hierarchies: Biological Process, Molecular Function, and Cellular Component (The Gene Ontology Consortium, 2012). Other biological ontologies model knowledge about *e.g.* specific organisms, types of molecules, pathways, phenotypes, diseases and experimental methods. GO and some others follow the Open Biomedical Ontology description format. *Gene set enrichment analysis* provides statistical tests of the over- or under-representation of ontology terms in the results of an experiment, compared with their frequency in a background.

Bio-Image Informatics

Bio-image informatics focuses on the capture, management, annotation, analysis and visualisation of data generated by light and electron microscopy, electronic imaging and related technologies. Image data can be static or dynamic, and are often large (Walter *et al.*, 2010). Consensus remains to be developed on metadata standards. Application areas range from the molecular scale through cell structure, tissues, organ systems and whole organisms and include informatic methods in support of histopathology, high-throughput screening, medical imaging, particle tracking and single-particle analysis.

Societies and Journals

The scientific journal *Computer Applications in the Biosciences* adopted the title *Bioinformatics* from 1998. Other important journals include *BMC Bioinformatics*, *Briefings in Bioinformatics*, *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, and *PLoS Computational Biology*. The main international scientific society in bioinformatics is the International Society for Computational Biology (ISCB), which sponsors annual conferences including Intelligent Systems in Molecular Biology, the International Conference on Research in Computational Molecular Biology, ISCB Asia, and Pacific Symposium on Biocomputing. Other important conference series include Asia Pacific Bioinformatics Conference, European Conference on Computational Biology, International Conference on Bioinformatics, and International Conference on Genome Informatics. Many other well-regarded societies, conferences and journals now offer a bioinformatics section.

FUTURE RESEARCH DIRECTIONS

The rapidly increasing affordability of genome sequencing is driving new fields of personalized genomic health care and medicine, and opening unprecedented opportunities in areas as diverse as anthropology, forensics, plant and animal breeding, industrial biotechnology, and environmental management. Bioinformatic challenges across these application domains include the management of immense quantities of diverse data, sometimes in secure environments (*e.g.* to protect patients or intellectual property); highly scalable algorithms and tools especially for networks, systems and populations; and the ability to overlay geospatial coordinates and temporal models. Many important datasets are already too large to be moved, and new types of bioinformatic tools and infrastructure are being trialled that take analyses to the data (*e.g.* in the cloud) rather than vice-versa. The most-

pressing and difficult challenge continues to be the education and training of personnel who can work across the component disciplines of bioinformatics to solve ever more-complex biological problems, while enriching information science and technology.

CONCLUSION

As foreseen by Nobel Laureate Walter Gilbert (Gilbert, 1991), biology has substantially become an information science. At the same time, the richness of problems and ever-growing deluge of data across modern biology is increasingly transforming information science and technology into a living science in the most literal sense.

Bioinformatics is now an indispensable enabling technology at the interface of bioscience with mathematics, statistics, computer science, and information science and technology. This is seen most dramatically in genomics and the other “omic” sciences, which continue to be driven by ever-larger data and the need for integrative analysis of complex biological systems. Specialty areas continue to evolve within bioinformatics to support the study of gene regulation, biomolecular and cellular structure, drug discovery, molecular evolution, bio-imaging and other application domains. Standards, policies and a network of large international bioinformatics centres ensure that biomolecular data are well-managed, openly available for re-use, and in many cases integrated into value-added data services.

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), 403–410. doi:10.1016/S0022-2836(05)80360-2 PMID:2231712

- Bailey, T. L., Johnson, J., Grant, C. E., & Noble, W. S. (2015). The MEME Suite. *Nucleic Acids Research*, 43(W1), W39–W49. doi:10.1093/nar/gkv416 PMID:25953851
- Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., & Vingron, M. et al. (2001). Minimum information about a microarray experiment (MIAME) — toward standards for microarray data. *Nature Genetics*, 29(4), 365–371. doi:10.1038/ng1201-365 PMID:11726920
- Chan, C. X., Bernard, G., Poirion, O., Hogan, J. M., & Ragan, M. A. (2014). Inferring phylogenies of evolving sequences without multiple sequence alignment. *Scientific Reports*, 4, 6504. doi:10.1038/srep06504 PMID:25266120
- Csermely, P., Korscsmáros, T., Kiss, H. J. M., London, G., & Nussinov, R. (2013). Structure and dynamics of molecular networks: A novel paradigm of drug discovery. A comprehensive review. *Pharmacology & Therapeutics*, 138(3), 333–408. doi:10.1016/j.pharmthera.2013.01.016 PMID:23384594
- Dickson, D. (1987). EMBL: small science on a European scale. *Science*, 237(4819), 1108–1109. doi:10.1126/science.3629233 PMID:3629233
- Ecker, J. R., Bickmore, W. A., Barroso, I., Pritchard, J. K., Gilad, Y., & Segal, E. (2012). Genomics: ENCODE explained. *Nature*, 489(7414), 52–55. http://www.nature.com/encode doi:10.1038/489052a PMID:22955614
- Gilbert, W. (1991). Towards a paradigm shift in biology. *Nature*, 349(6305), 99. doi:10.1038/349099a0 PMID:1986314
- Gu, J., & Bourne, P. E. (Eds.). (2009). *Structural Bioinformatics* (2nd ed.). Hoboken, NJ: Wiley.
- Hogeweg, P. (2011). The roots of bioinformatics in theoretical biology. *PLoS Computational Biology*, 7(3), e1002021. doi:10.1371/journal.pcbi.1002021 PMID:21483479
- Huerta, M., Haseltine, F., Liu, Y., Downing, G., & Seto, B. (2000). *NIH working definition of bioinformatics and computational biology*. Retrieved from <http://www.bisti.nih.gov/docs/CompuBioDef.pdf>
- Karp, R. M. (1972). Reducibility among combinatorial problems. In R. E. Miller & J. W. Thatcher (Eds.), *Complexity of Computer Computations* (pp. 85–103). New York: Plenum. doi:10.1007/978-1-4684-2001-2_9
- Libbrecht, M. W., & Noble, W. S. (2015). Machine learning approaches in genetics and genomics. *Nature Review Genetics*, 16(6), 321–332. doi:10.1038/nrg3920 PMID:25948244
- Lipman, D. J., & Pearson, W. R. (1985). Rapid and sensitive protein similarity searches. *Science*, 227(4693), 1435–1441. doi:10.1126/science.2983426 PMID:2983426
- Nurse, P. (2008). Life, logic and information. *Nature*, 454(7203), 424–426. doi:10.1038/454424a PMID:18650911
- O’ Donoghue, S. I., Gavin, A.-C., Gehlenborg, N., Goodsell, D. S., Hériché, J.-K., Nielsen, C. B., & Wong, B. et al. (2010). Visualizing biological data – now and in the future. *Nature Methods*, 7(Suppl. 3s), S2–S4. doi:10.1038/nmeth.f.301 PMID:20195254
- Taylor, C. F., Field, D., Sansone, S.-A., Aerts, J., Apweiler, R., Ashburner, M., & Wiemann, S. et al. (2008). Promoting coherent minimum reporting guidelines for biological and biomedical investigations: The MIBBI project. *Nature Biotechnology*, 26(6), 889–896. doi:10.1038/nbt.1411 PMID:18688244
- The Gene Ontology Consortium. (n.d.). Retrieved from <http://www.geneontology.org>
- Walter, T., Shattuck, D. W., Baldock, R., Bastin, M. E., Carpenter, A. E., Duce, S., & Hériché, J.-K. et al. (2010). Visualization of image data from cells to organisms. *Nature Methods*, 7(Suppl. 3s), S26–S41. doi:10.1038/nmeth.1431 PMID:20195255

Yang, Z., & Rannala, B. (2012). Molecular phylogenetics: Principles and practice. *Nature Reviews. Genetics*, 13(5), 303–314. doi:10.1038/nrg3186 PMID:22456349

ADDITIONAL READING

Alterovitz, G., Muso, T., & Ramoni, M. F. (2010). The challenges of informatics in synthetic biology: From biomolecular networks to artificial organisms. *Briefings in Bioinformatics*, 11(1), 80–95. doi:10.1093/bib/bbp054 PMID:19906839

Arabidopsis Genome Initiative. (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, 408(6814), 796–815. doi:10.1038/35048692 PMID:11130711

Bock, C., & Lengauer, T. (2008). Computational epigenetics. *Bioinformatics (Oxford, England)*, 24(1), 1–10. doi:10.1093/bioinformatics/btm546 PMID:18024971

Capriotti, E., Nehrt, N. L., Kann, M. G., & Bromberg, Y. (2012). Bioinformatics for personal genome interpretation. *Briefings in Bioinformatics*, 13(4), 495–512. doi:10.1093/bib/bbr070 PMID:22247263

Cunningham, F., Amode, M. R., Barrell, D., Beal, K., Billis, K., Brent, S., & Flicek, P. et al. (2014). Ensembl 2015. *Nucleic Acids Research*, 43(D1), D662–D669. doi:10.1093/nar/gku1010 PMID:25352552

Du Plessis, L., Škunca, N., & Dessimoz, C. (2011). The what, where, how and why of gene ontology – a primer for bioinformaticians. *Briefings in Bioinformatics*, 12(6), 723–735. doi:10.1093/bib/bbr002 PMID:21330331

Durbin, R., Eddy, S. R., Krogh, A., & Mitchison, G. (1998). *Biological sequence analysis: probabilistic models of proteins and nucleic acids*. Cambridge: University Press. doi:10.1017/CBO9780511790492

Eddy, S. R. (1998). Profile hidden Markov models. *Bioinformatics (Oxford, England)*, 14(9), 755–763. doi:10.1093/bioinformatics/14.9.755 PMID:9918945

Eliceiri, K. W., Berthold, M. R., Goldberg, I. G., Ibáñez, I., Manjunath, B. S., Martone, M. E., & Carpenter, A. E. et al. (2012). Biological imaging software tools. *Nature Methods*, 9(7), 697–710. doi:10.1038/nmeth.2084 PMID:22743775

Felsenstein, J. (2004). *Inferring phylogenies*. Sunderland, MA: Sinauer.

Forsdyke, D. R. (2011). *Evolutionary bioinformatics* (2nd ed.). New York: Springer. doi:10.1007/978-1-4419-7771-7

Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., & Zhang, J. et al. (2004). Bioconductor: Open software development for computational biology and bioinformatics. *Genome Biology*, 5(10), R80. doi:10.1186/gb-2004-5-10-r80 PMID:15461798

Gevers, D., Knight, R., Petrosino, J. F., Huang, K., McGuire, A. L., Birren, B. W., & Huttenhower, C. et al. (2012). The Human Microbiome Project: A community resource for the healthy human microbiome. *PLoS Biology*, 10(8), e1001377. doi:10.1371/journal.pbio.1001377 PMID:22904687

Gusfield, D. (1997). *Algorithms on strings, trees and sequences: Computer science and computational biology*. Cambridge: University Press. doi:10.1017/CBO9780511574931

Hubbard, S. J., & Jones, R. A. (2010). *Proteome bioinformatics*. London: Humana. doi:10.1007/978-1-60761-444-9

Hung, J.-H., Yang, T.-H., Hu, Z., Weng, Z., & DeLisi, C. (2012). Gene set enrichment analysis: Performance evaluation and usage guidelines. *Briefings in Bioinformatics*, 13(3), 281–291. doi:10.1093/bib/bbr049 PMID:21900207

- Jones, N., & Pevzner, P. (2004). *An introduction to bioinformatics algorithms*. Cambridge, MA: MIT Press.
- Kanehisa, M., Goto, S., Sato, Y., Kawashima, S., Furumichi, M., & Tanabe, M. (2014). Data, information, knowledge and principle: Back to metabolism in KEGG. *Nucleic Acids Research*, 42(D1), D199–D205. doi:10.1093/nar/gkt1076 PMID:24214961
- Keith, J. M. (Ed.). (2008). *Bioinformatics*. Volume I: Data, sequence analysis and evolution. Volume II: Structure, function and applications. *Methods in Molecular Biology*, volumes 452–453. Humana.
- Kim, M., Lee, K.-H., Yoon, S.-W., Kim, B.-S., Chun, J., & Yi, H. (2013). Analytical tools and databases for metagenomics in the next-generation sequencing era. *Genomics & Informatics*, 11(3), 102–113. doi:10.5808/GI.2013.11.3.102 PMID:24124405
- Koschmieder, A., Zimmermann, K., Trißl, S., Stoltmann, T., & Leser, U. (2012). Tools for managing and analysing microarray data. *Briefings in Bioinformatics*, 13(1), 46–60. doi:10.1093/bib/bbr010 PMID:21422065
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., & Morgan, M. J. et al. International Human Genome Sequencing Consortium. (2001). Initial sequencing and analysis of the human genome. *Nature*, 409(6822), 860–921. doi:10.1038/35057062 PMID:11237011
- Larson, R. S. (Ed.). (2012). *Bioinformatics and drug discovery*, Second edition. *Methods in Molecular Biology*, volume 910. New York: Springer. doi:10.1007/978-1-61779-965-5
- Li, C., Courtot, M., Le Novère, N., & Laibe, C. (2010). BioModels.net Web Services, a free and integrated toolkit for computational modelling software. *Briefings in Bioinformatics*, 11(3), 270–277. doi:10.1093/bib/bbp056 PMID:19939940
- Lourenço, A., Cameiro, S., Rocha, M., Ferreira, E. C., & Rocha, I. (2011). Challenges in integrating *Escherichia coli* molecular biology data. *Briefings in Bioinformatics*, 12(2), 91–103. doi:10.1093/bib/bbq067 PMID:21059604
- Maetschke, S. R., Madhamshettiwar, P. B., Davis, M. J., & Ragan, M. A. (2013). Supervised, semi-supervised and unsupervised inference of gene regulatory networks. *Briefings in Bioinformatics*, 15(2), 195–211. doi:10.1093/bib/bbt034 PMID:23698722
- Mathur, S. K. (2010). *Statistical bioinformatics with R*. New York: Academic Press.
- Moult, J., Fidelis, K., Kryshchuk, A., Schwede, T., & Tramontano, A. (2014). Critical assessment of methods of protein structure prediction (CASP) – Round X. *Proteins*, 82(S2), 1–6. doi:10.1002/prot.24452 PMID:24344053
- Ochs, M. F. (2010). Knowledge-based data analysis comes of age. *Briefings in Bioinformatics*, 11(1), 30–39. doi:10.1093/bib/bbp044 PMID:19854753
- O’ Donoghue, S. I., Gavin, A.-C., Gehlenborg, N., Goodsell, D. S., Hériché, J.-K., Nielsen, C. B., & Wong, B. et al. (2010). Visualising biological data – now and in the future. *Nature Methods*, 7(Suppl. 3s), S2–S15. doi:10.1038/nmeth.f.301 PMID:20195254
- Overby, C. L., & Tarczy-Hornoch, P. (2013). Personalized medicine: Challenges and opportunities for translational bioinformatics. *Personalized Medicine*, 10(5), 453–462. doi:10.2217/pme.13.30 PMID:24039624
- Prill, R. M., Marbach, D., Saez-Rodriguez, J., Sorger, P. K., Alexopoulos, L. G., Xue, X., & Stolovitzky, G. et al. (2010). Towards a rigorous assessment of systems biology models: The DREAM3 challenges. *PLoS ONE*, 5(2), e9202. doi:10.1371/journal.pone.0009202 PMID:20186320

- Radivojac, P., Clark, W. T., Oron, T. R., Schnoes, A. M., Wittkop, T., Sokolov, A., & Friedberg, I. et al. (2013). A large-scale evaluation of computational protein function prediction. *Nature Methods*, 10(3), 221–227. doi:10.1038/nmeth.2340 PMID:23353650
- Romano, P., Guigno, R., & Pulvirenti, A. (2011). Tools and collaborative environments for bioinformatics research. *Briefings in Bioinformatics*, 12(6), 549–561. doi:10.1093/bib/bbr055 PMID:21984743
- Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., & Huelsenbeck, J. P. et al. (2012). MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, 61(3), 539–542. doi:10.1093/sysbio/sys029 PMID:22357727
- Rosenbloom, K. R., Armstrong, J., Barber, G. P., Casper, J., Clawson, H., Diekhans, M., & Kent, W. J. et al. (2015). The UCSC Genome Browser database: 2015 update. *Nucleic Acids Research*, 43(D1), D670–D681. doi:10.1093/nar/gku1177 PMID:25428374
- Ruttenberg, A., Rees, J. A., Samwald, M., & Marshall, M. S. (2009). Life sciences on the Semantic Web: The Neurocommons and beyond. *Briefings in Bioinformatics*, 10(2), 193–204. doi:10.1093/bib/bbp004 PMID:19282504
- Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P.-L., & Ideker, T. (2011). Cytoscape 2.8: New features for data integration and network visualization. *Bioinformatics (Oxford, England)*, 27(3), 431–432. doi:10.1093/bioinformatics/btq675 PMID:21149340
- Smyth, G. K. (2004). Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology*, 3(1), 3. doi:10.2202/1544-6115.1027 PMID:16646809
- The FANTOM Consortium. (2005). The transcriptional landscape of the mammalian genome. *Science*, 309(5740), 1559–1563. doi:10.1126/science.1112014 PMID:16141072
- Tompa, M., Li, N., Bailey, T. L., Church, G. M., De Moor, B., Eskin, E., & Zhu, Z. et al. (2005). Assessing computational tools for the discovery of transcription factor binding sites. *Nature Biotechnology*, 23(1), 137–144. doi:10.1038/nbt1053 PMID:15637633
- Venter, J. C., Adams, M. C., & Myers, E. W. et al. (2001). The sequence of the human genome. *Science*, 291(5507), 1304–1351. doi:10.1126/science.1058040 PMID:11181995
- Venter, J. C., Remington, K., & Heidelberg, J. F. et al. (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science*, 304(5667), 66–74. doi:10.1126/science.1093857 PMID:15001713
- Wishart, D. S. (2005). Bioinformatics in drug development and assessment. *Drug Metabolism Reviews*, 37(2), 279–310. doi:10.1081/DMR-55225 PMID:15931766
- Wooley, J. C., Godzik, A., & Friedberg, I. (2010). A primer on metagenomics. *PLoS Computational Biology*, 6(2), e1000667. doi:10.1371/journal.pcbi.1000667 PMID:20195499
- Zhang, Z., Cheung, K.-H., & Townsend, J. P. (2012). Bringing Web 2.0 to bioinformatics. *Briefings in Bioinformatics*, 10(1), 1–10. doi:10.1093/bib/bbn041 PMID:18842678
- Zvelebil, M., & Baum, J. O. (2008). *Understanding bioinformatics*. New York, London: Garland Science.

KEY TERMS AND DEFINITIONS

Epigenetics: The study of heritable changes in gene expression or phenotype that do not arise from changes in the nucleotide sequence of the genome.

Gene Expression: The process in which information in a gene is transduced to yield one or more gene products (proteins and/or non-protein-coding RNAs).

Genomics: The study of the totality of genetic information in a cell or organism.

Metagenomics: The study of the genetic information in a sample of a mixed community, for example from soil, water, or the intestinal tract of an animal, without prior separation of the constituent organisms.

Omics (or ‘Omics): Genomics, transcriptomics, and other fields of biology characterised by the study of the totality of a class of molecule (*e.g.* proteomics), process (*e.g.* metabolomics) or traits (*e.g.* phenomics) in a cell or organism.

Phylogenetics: The study of genetic relationships among molecules, organisms or taxa over time.

Synthetic Biology: The engineering design and construction of new or modified functions in biological systems.

Systems Biology: The study of networks of interactions among the components of a biological system, emphasising a synthetic, non-reductionist approach and emergent properties.

Transcriptomics: The study of the totality of the RNA molecules arising from the expression of genes under specified conditions.

Bioinspired Solutions for MEMS Tribology

B**R. Arvind Singh***Kumaraguru College of Technology (KCT), India***S. Jayalakshmi***Kumaraguru College of Technology (KCT), India*

INTRODUCTION

Micro-Electro-Mechanical Systems (MEMS) are miniaturized devices that perform intelligent functions. They can be classified as: (i) sensors based, which have sensing elements and (ii) actuators based, which have elements that undergo mechanical motion. Micro-motors/engines, micro-gears and micro-shutters are examples of actuators based devices. Figure 1 shows a micro-gear made from silicon (http://www.sandia.gov/mstc/_assets/images/mems/gallery/gears/1.jpg). In these devices, tribological issues such as adhesion, friction and wear strongly manifest, which undermine the mechanical motion of MEMS elements (Kim, Asay, & Dugger, 2007). Therefore, it is imperative to solve these tribological issues in order to realize smooth operation and increased operating lifetimes of actuators based MEMS. Conventionally, thin films/coatings have been researched for their application to MEMS as solutions to mitigate the tribological issues. Examples include self-assembled monolayers (SAMs), diamond-like carbon (DLC) coatings, polymers and perfluoropolyether films (PFPE) (Bhushan, 2001). In recent years, bioinspired approaches have attracted attention as better alternative solutions for MEMS. This chapter highlights bioinspired approaches that are promising tribological solutions for actuators based MEMS devices.

BACKGROUND

MEMS are built at micro/nano-scale. At these scales, the ratio of surface area to volume is high.

Figure 1. MEMS micro-gear made from silicon. (http://www.sandia.gov/mstc/_assets/images/mems/gallery/gears/1.jpg). In MEMS devices, tribological issues such as adhesion, friction and wear strongly manifest and undermine the mechanical motion



Hence, body forces such as inertia and gravity become insignificant. In contrast, surface forces such as capillary, van der Waals, electrostatic, and chemical bonding dominate. These surface forces cause adhesion at the interface of contacting MEMS elements. Amongst these forces, the capillary force that arises due to the condensation of water from the environment is the strongest. Further, adhesion strongly influences friction at micro/nano-scale (Maboudian & Howe, 1997). The magnitude of these surface forces is comparable with those that drive the motion of MEMS elements, thereby rendering the elements completely inoperable. MEMS are traditionally made from silicon due to availability of the

DOI: 10.4018/978-1-5225-2255-3.ch038

process knowledge developed for the material in semiconductor industries. However, silicon does not have good tribological properties (Bhushan, 2001). Silicon due to its inherent hydrophilic nature experiences high surface forces, and because of its brittle nature it undergoes severe wear. Thus, the improvement of tribological performance of silicon is the key to realize the smooth operation of actuators based MEMS. Williams and Le have presented an excellent review on the tribological issues in MEMS devices (Williams & Le, 2006).

Tribological properties of materials at micro/nano-scale are evaluated using atomic force microscopes (AFM) and micro-tribo testers. With these instruments, contact conditions similar to those in MEMS can be easily simulated (loads $\sim$ nN to μ N-mN; area $\sim$ few hundreds of nm²).

CONVENTIONAL SOLUTIONS

The gap between moving elements in MEMS is usually about few microns. Hence, solid lubricants (e.g. graphite, MoS₂, WS₂) or liquid lubricants (e.g. oils, greases) cannot be applied in MEMS devices as their sizes are of the same order as those of MEMS elements. Due to this reason, thin films/coatings with thickness less than few microns have been investigated for their application as 'boundary lubricants' for MEMS. Table 1 shows the water contact angle (CA) values, nano-scale adhesion force and nano-scale friction coefficient (μ_n) of silicon (Yoon et al., 2005). The table also shows CA values and nano-tribological properties of thin films/coatings, namely diamond-like carbon (DLC), Z-DOL (commercial name for perfluoropolyether, PFPE), and OTS (octadecyltrichlorosilane) SAM (Yoon et al., 2005; Liu & Bhushan, 2003; Liu, Ahmed, & Scherge, 2001; Singh & Yoon, 2007). The adhesion values given in the table were obtained from force-distance curves measured using AFM. In some cases, the values are taken from the negative intercepts in the friction force versus normal load plots. The

nano-scale friction coefficient (μ_n) is estimated as the slope of friction force versus normal load.

Surface energy of a material is indicated by the contact angle between a water droplet and its surface (water contact angle is inversely proportional to surface energy (Bain, Evans & Whitesides, 1989). From Table 1, it is seen that the adhesion values of the test materials are inversely dependent on their CA values. Silicon is hydrophilic in nature and thereby it supports strong capillary force that causes high adhesion. This in turn increases its friction, which retards the motion of actuators. In contrast, DLC, Z-DOL and OTS have higher CA values. This means that they have lower surface energies and sustain lower surface forces. When applied to actuators these materials are capable of promoting smooth motion.

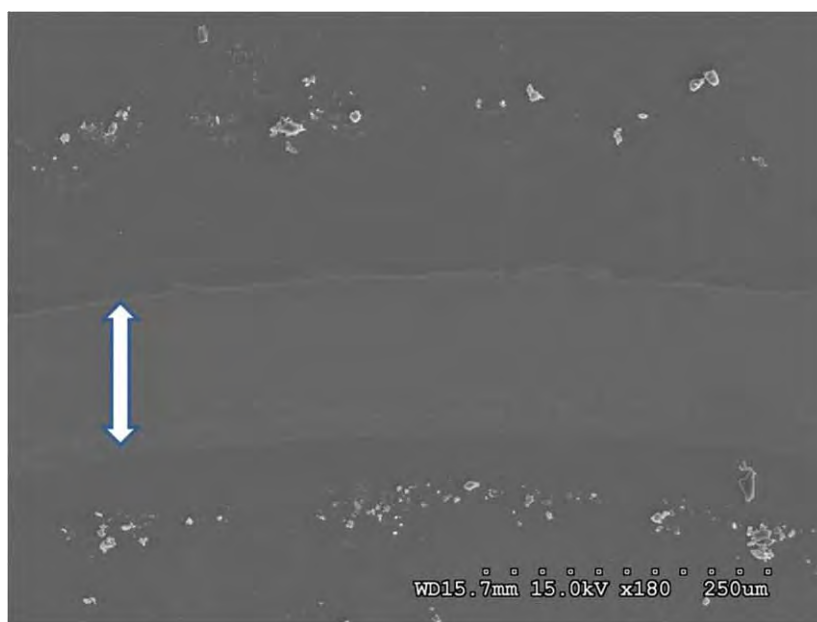
Table 1 shows micro-scale friction coefficient (μ_m) and durability (i.e. the number of cycles for the onset of wear) of the test materials (Yoon et al., 2005; Liu & Bhushan, 2003; Liu, Ahmed, & Scherge, 2001; Singh & Yoon, 2007). At micro-scale, silicon undergoes severe wear within a short duration of time, i.e. it has low durability (< 100 cycles). Figure 2 shows a worn surface of silicon tested at micro-scale. Wear in the case of DLC, Z-DOL and OTS begins only after hundreds of cycles of sliding, which indicates good durability. These boundary lubricants have distinct mechanisms that promote good tribological properties. By nature they are semi-hydrophobic/hydrophobic, due to which adhesion and friction get reduced to a great extent. In DLC films, in addition to their semi-hydrophobic nature, the formation of transfer layer on the counterface surfaces lowers friction and wear (Erdemir et al., 1996). Z-DOL has fluorine atoms which makes its hydrophobic. Fluorine atoms have larger van der Waals radii that support molecular scale order under shear thereby making Z-DOL a good lubricant (Bhushan, 2001). In OTS, the molecular chains act as molecular springs and reduce friction (Liu & Bhushan, 2003). When these boundary lubricants are applied to silicon MEMS, the tribological performance of the devices can be considerably enhanced.

Table 1. Water contact angle (CA) values, nano-scale tribological properties (adhesion force (nN), friction coefficient (μ_n)) and micro-scale tribological properties (micro-scale friction coefficient (μ_m), durability (n)) of silicon and various thin films/coatings, taken from literature

Nano-Scale Tribological Properties			
MaterialSurface	CA(deg)	Nano-scaleAdhesionForce (nN)	Nano-scaleFrictionCoefficient (μ_n)
Silicon	30	50	0.05
DLC	60	40	0.03
Z-DOL	72	35	0.03
OTS	100	5.6	0.04
Micro-Scale Tribological Properties			
MaterialSurface	Micro-ScaleFrictionCoefficient (μ_m)	Durability(Number of Cycles, n)	
Silicon	0.7	< 100	
DLC	0.2	> 500	
Z-DOL	0.17	~ 5000	
OTS	0.18	~ 1600	

Source: Yoon et al., 2005; Liu & Bhushan, 2003; Liu, Ahmed, & Scherge, 2001; Singh & Yoon, 2007

Figure 2. Worn surface of silicon tested at micro-scale. At micro-scale, silicon undergoes wear within a short duration of time, i.e. it has low durability



WATER REPELLENT SURFACES

Lotus leaves are superhydrophobic with CA of about 162° (Neinhuis, Barthlott, 1997). This property is an effect of the hierarchical structures

on the leaves. Figure 3 shows the surface of a lotus leaf with micro-scale protuberances. The superhydrophobic property is induced by the dual-scale surface roughness from the micro-scale protuberances and waxy nano-crystals. The unique

characteristic of lotus leaves to avoid getting wet is popularly known as “Lotus Effect”. Colocasia is another water repellent plant having CA of about 164° (Neinhuis, Barthlott, 1997). Superhydrophobicity of water repellent plants arises due to the heterogeneous/composite wetting of water, wherein water drops sit on top of the protuberances with air trapped in between them, reducing the solid contact area with the water drops. By mimicking the features present on lotus leaves, scientists have created water repellent surfaces for various applications such as mirrors, car windows, computer/notebook key boards, building structures, fabrics etc., (Gould, 2003).

BIOINSPIRED SOLUTIONS

At micro/nano-scale, friction is in a regime where the contribution from intrinsic/inherent adhesion can outweigh that from asperity deformation (Maboudian & Howe, 1997). Thus, in reducing adhesion, friction also gets reduced, and therefore solutions to reduce adhesion are the key to real-

ize smooth motion between elements of MEMS devices. Further, the effective route to mitigate adhesion is by reducing capillary force at the interface. This can be achieved by learning from ‘lotus effect’. Friction at micro/nano-scale depends on real area of contact (Yoon et al., 2005). Therefore, any reduction in real contact area between surfaces directly reduces adhesion and friction. By creating surface features i.e. topography similar to that seen on lotus leaves (which reduces the contact area between water drops and their surfaces) solutions to reduce surface forces in MEMS can be realized.

As a pioneering work, nano-scale polymeric patterns on silicon surfaces that mimic the protuberances on lotus leaves were fabricated by capillary force lithography, and investigated for adhesion and friction at micro/nano-scale (Yoon et al., 2006). Figure 4 shows a representative image of polymeric nano-patterns on silicon. CA values indicated that the nano-patterns were hydrophobic (Table 2). AFM measurements showed that the nano-scale adhesion and friction of the patterns was significantly lower than those of silicon and PMMA thin film (Table 2). The

Figure 3. Surface of a lotus leaf with micro-scale protuberances. The micro-scale protuberances and waxy nano-crystals on it make the lotus leaf surface superhydrophobic

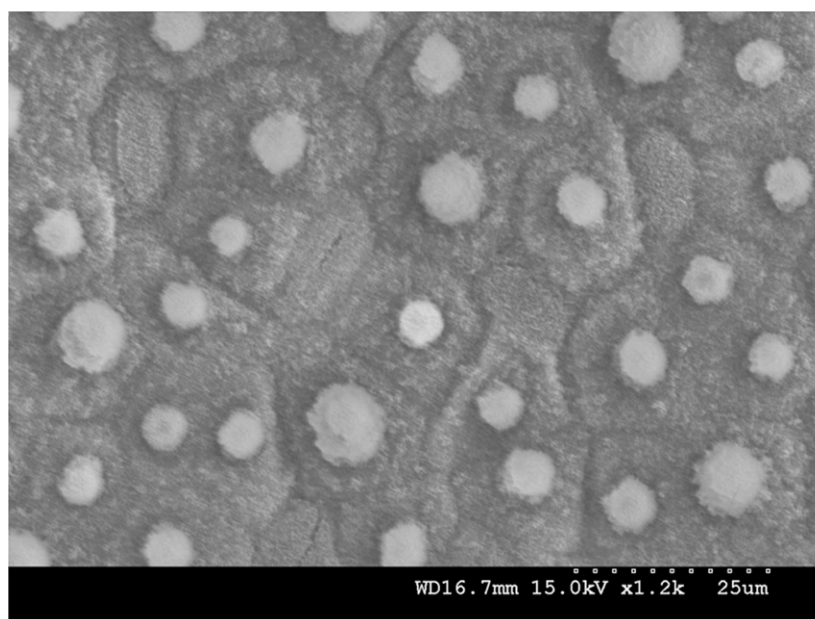


Table 2. Water contact angle (CA) values, nano-scale tribological properties (adhesion force (nN), friction force (nN)) and micro-scale tribological property (micro-scale friction coefficient (μ_m), durability (n)) of silicon, PMMA, DLC and various nano/micro-patterned surfaces, taken from literature

MaterialSurface	CA(deg)	Nano-Scale Adhesion Force (nN)	Nano-Scale Friction Force (nN)	Micro-Scale Friction Coefficient (μ_m)
Flat Surfaces				
Silicon	22	984	20-28	0.46
PMMA Thin Film	69	872	5-8	0.66
DLC Thin Film	78	226	5.5-8.3	0.19
Nano/Micro-Patterned Surfaces				
PMMA Nano-Patterns	91-99	30-40	0.51-1.97	0.24-0.33
PMMA Micro-Patterns	91-106	-	-	0.11-0.13
(Replicated Surfaces)	95	67	0.98-1.96	0.13
DLC Nano-Dot Surfaces	59	-	-	0.16
Silicon Micro-Pillars	145	-	-	0.12
Silicon Micro-Pillars coated with DLC	105	-	-	0.08
Silicon Micro-Pillars coated with ZDOL				

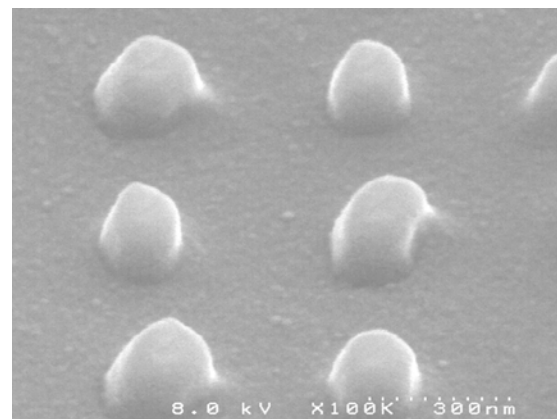
Source: Yoon et al., 2006; Singh et al., 2007; Singh et al., 2011; Singh et al., 2011

superior tribological performance of the patterns is due to hydrophobicity and reduced real area of contact between their surfaces and AFM tips. To note, patterning of surfaces causes a reduction in contact area when the size of the asperities (patterns) is considerably smaller than that of the counterface (AFM tips). An investigation by Burton et al. clearly highlights the influence of pattern size (aspect ratio), counterface tip size and relative humidity on the magnitude of surface forces in polymeric patterns (Burton & Bhushan, 2005). Polymeric nano-patterns have excellent tribological properties at nano-scale, however they do not have good load bearing capability at micro-scale. At micro-scale, under micro-loads (μN -mN) polymeric ‘micro-patterns’ exhibit low friction and good load bearing capability (Table 2) (Singh et al., 2007). Figure 5 shows a representative image of polymeric micro-pattern from the direct replication of the surface of natural lotus (*Nelumbo nucifera*). The larger size of the micro-patterns and the increased distance between

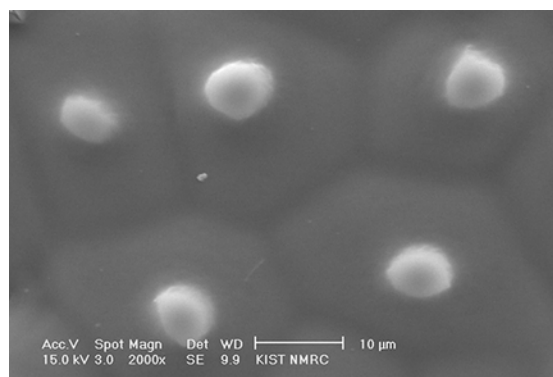
the micro-patterns together contribute towards reducing the contact area significantly resulting in lower friction.

Polymeric patterns undergo elastic and/or plastic deformation under relatively higher normal loads. Such deformation result in higher contact

Figure 4. A representative image of polymeric nano-patterns on silicon. These patterns mimic the surface topography of a lotus leaf



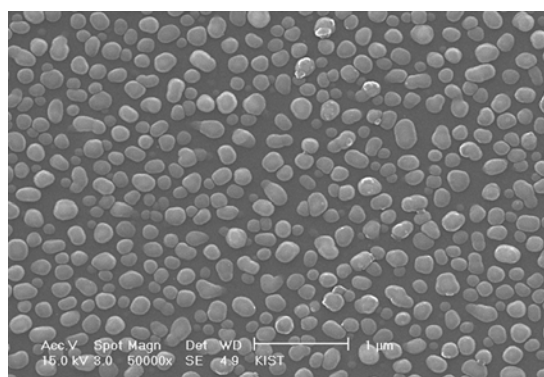
*Figure 5. A representative image of micron-sized polymeric pattern from the direct replication of the surface of natural lotus (*Nelumbo nucifera*). These micropatterns contribute to large reduction in friction under micro-loads*



areas and increase the magnitude of surface forces. An effective way to avoid deformation is to pattern surfaces that are harder and stiffer. Examples of such surfaces include micro/nano-textures produced by UV assisted/aluminum-induced crystallization of amorphous silicon (Zou et al., 2005; Nair & Zou, 2008) and DLC nano-dots surface (Singh et al., 2011). DLC nano-dot surfaces are created by depositing DLC films on nano-sized nickel (Ni) dots fabricated on silicon wafers by annealing Ni thin films. The nano-dot surfaces are hydrophobic and reduce adhesion and friction forces significantly (Table 2). Unlike polymeric nano-patterns that have uniform height and are spaced equidistant, the DLC nano-dots are randomly distributed and have un-equal heights. Due to their un-equal heights, only those nano-dots that are taller come in contact with the counterface, thereby resulting in smaller contact areas that lower adhesion and friction. Nano-dots can withstand micro-loads as they are relatively harder and stiffer than polymeric patterns.

Combination of surface patterning and boundary lubrication proves to be more effective in minimizing surface forces. Examples include polymeric patterns with perfluorodecyltriethoxysilane (PFDTES) SAM (Burton & Bhushan, 2005), silicon micro-pillars coated with DLC/ZDOL (Table

Figure 6. DLC nano-dots surface. The surface significantly reduces adhesion and friction



2) (Singh et al., 2011) and textured surfaces of Ni films coated with PFPE (Wang et al., 2014). For a detailed discussion, readers are suggested to refer the book (Sinha, Nalam & Lim (eds.), 2013).

COMPARATIVE NOTE

Figures 7 and 8 show the nano-scale friction and micro-scale friction coefficient (μ_m) of polymeric patterns in comparison with those of silicon, DLC and diphenyldichlorosilane (DPDC) SAM (Singh & Yoon, 2007). From these figures, it could be seen that the bioinspired patterns have relatively lower friction property when compared to the rest of the test materials. Bioinspired tribological surfaces have two fold advantages over the conventional solutions i.e. boundary lubricants (DLC, Z-DOL, SAMs) namely:

1. Ease of application and
2. Effectiveness in reducing surface forces.

As for the first aspect, boundary lubricants have to be applied to MEMS surfaces externally i.e. via dip coating process, physical/chemical vapor deposition etc. Thus, application of boundary lubricants and ensuring their uniform coating/coverage throughout the surfaces of MEMS elements has remained a major challenge. In contrast, bioinspired patterning of surfaces of MEMS ele-

ments can be easily achieved by using the same lithography techniques that are used to fabricate MEMS elements. Moreover, surface patterning can be done verily during the fabrication of the MEMS surfaces. This makes the patterning route more time-effective and cost-effective. Further, in the case of boundary lubricants complex chemistry has to work to make them effective in reducing the surface forces. For example, the tribological performance of SAMs depends on the suitable selections of end groups and chain length. In the case of DLC thin films, the choice of sp^3/sp^2 ratio and the mechanical properties determine their tribological performance. On the other hand, for patterns simple geometric parameters such as height, diameter and pitch (distance between any two adjacent features) define their tribological behavior. By varying these simple geometric parameters, surface forces can be easily controlled.

FUTURE RESEARCH DIRECTIONS

Till date, simple patterns/textures have been investigated for their micro/nano-scale tribological properties for MEMS applications. These surfaces are hydrophobic in nature and reduce surface forces considerably. For more effective reduction in surface forces, patterns with hierarchical structures

Figure 7. Nano-scale friction of polymeric patterns in comparison with those of silicon, DLC and diphenyldichlorosilane (DPDC) SAM

Source: Singh &, Yoon, 2007

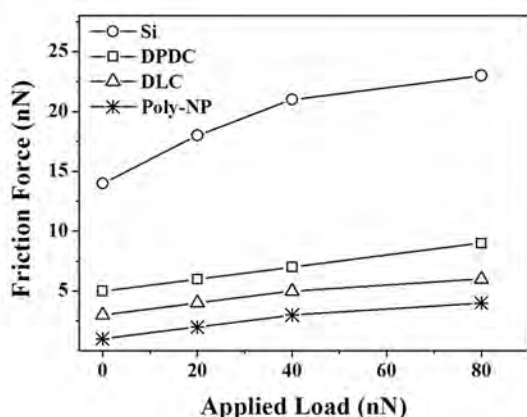
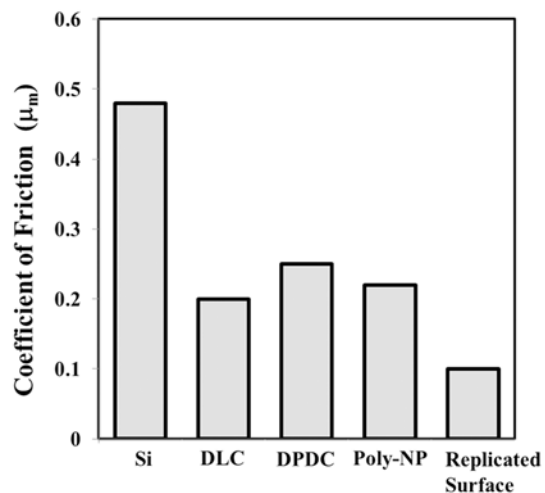


Figure 8. Micro-scale friction coefficient (μ_m) of polymeric patterns in comparison with those of silicon, DLC and diphenyldichlorosilane (DPDC) SAM

Source: Singh & Yoon, 2007



(2 tier) need to be investigated. Some examples of hierarchical structures that are of interest are nano-engineered multiscale hierarchical structures with tailored wetting properties (Jeong et al., 2006) and biomimetic, hierarchical structures fabricated on polymer surfaces by sequential imprinting (Zhang, Chan & Low, 2008). Patterns with hierarchical structures give rise to superhydrophobicity, and such topography in combination with boundary lubrication has the potential to greatly reduce surface forces at micro/nano-scale.

CONCLUSION

Bioinspired surfaces have proved to be effective in reducing adhesion and friction at micro/nano-scale and have paved a new way of addressing the tribological issues in MEMS devices. Examples presented in this chapter clearly show that the biomimetic engineering of surfaces is an exciting and practical way to control surface forces. Patterning of surfaces in combination with boundary lubrication can be pursued to control surface forces effectively for the application in miniaturized devices.

REFERENCES

- Bain, C. D., Evall, J., & Whitesides, G. M. (1989). Formations of monolayers by coadsorption of thiols on gold: Variation in the head group, tail group, and solvent. *Journal of the American Chemical Society*, 111(18), 7155–7164. doi:10.1021/ja00200a039
- Bhushan, B. (2001). *Modern Tribology Handbook*, 2. Boca Raton, FL: CRC Press.
- Burton, Z., & Bhushan, B. (2005). Hydrophobicity, Adhesion, and Friction Properties of Nanopatterned Polymers and Scale Dependence for Micro- and Nanoelectromechanical Systems. *Nano Letters*, 5(8), 1607–1613. doi:10.1021/nl050861b PMID:16089497
- Erdemir, A., Bindal, C., Fenske, G. R., Zuiker, C., & Wilbur, P. (1996). Characterization of transfer layers forming on surfaces against diamond-like carbon. *Surface and Coatings Technology*, 86-87, 692–697. doi:10.1016/S0257-8972(96)03073-3
- Gould, P. (2003, November). Smart, clean surfaces. *Materials Today*, 44-48. http://www.sandia.gov/mstc/_assets/images/mems/gallery/gears/14.jpg
- Hoon, E. J., Sung, H. L., Kim, J. K., & Suh, K. Y. (2006). Nanoengineered Multiscale Hierarchical Structures with Tailored Wetting Properties. *Langmuir*, 22(4), 1640–1645. doi:10.1021/la0526434 PMID:16460085
- Kim, S. H., Asay, D. B., & Dugger, M. T. (2007). Nanotribology and MEMS. *Nano Today*, 2(5), 22–29. doi:10.1016/S1748-0132(07)70140-8
- Liu, H., Ahmed, I. U., & Scherge, M. (2001). Microtribological properties of silicon and silicon coated with diamond like carbon, octadecyltrichlorosilane and stearic acid cadmium salt films: A comparative study. *Thin Solid Films*, 381(1), 135–142. doi:10.1016/S0040-6090(00)01546-7
- Liu, H., & Bhushan, B. (2003). Nanotribological characterization of molecularly thick lubricant films for applications of MEMS/NEMS by AFM. *Ultramicroscopy*, 97(1-4), 321–340. doi:10.1016/S0304-3991(03)00058-5 PMID:12801686
- Maboudian, R., & Howe, R. T. (1997). Critical Review: Adhesion in surface micromechanical structures. *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures*, 15(1), 1–20. doi:10.1116/1.589247
- Nair, P. R., & Zou, M. (2008). Surface-nano-texturing by Aluminum-induced Crystallization of Amorphous Silicon. *Surface and Coatings Technology*, 203(5-7), 675–679. doi:10.1016/j.surfcoat.2008.07.014
- Neinhuis, C., & Barthlott, W. (1997). Characterization and distribution of water-repellent, self-cleaning plant surfaces. *Annals of Botany*, 79(6), 667–677. doi:10.1006/anbo.1997.0400
- Singh, R., & Yoon, E. S. (2007). Friction of chemically and topographically modified Si (100) surfaces. *Wear*, 263(7-12), 912–919. doi:10.1016/j.wear.2007.01.059
- Sinha, S., Nalam, S., & Lim, S. C. (2013). *Nano-Tribology and Materials Issues in MEMS*. Berlin: Springer-Verlag. doi:10.1007/978-3-642-36935-3
- Wang, Y., Yang, J., Guo, X., Zhang, Q., Wang, J., Ding, J., & Yuan, N. (2014). Fabrication and tribological properties of superhydrophobic nickel films with positive and negative biomimetic microtextures. *Friction*, 2(3), 287–294. doi:10.1007/s40544-014-0060-4
- Williams, J. A., & Le, H. R. (2006). Tribology and MEMS. *Journal of Physics. D, Applied Physics*, 39(12), 201–214. doi:10.1088/0022-3727/39/12/R01
- Yoon, E. S., Arvind Singh, R., Oh, H. J., & Kong, H. (2005). The effect of contact area on nano/micro-scale friction. *Wear*, 259(7-12), 1424–1431. doi:10.1016/j.wear.2005.01.033

Yoon, E. S., Singh, R. A., Kong, H., Kim, B., Kim, D. H., Jeong, H. E., & Suh, K. Y. (2006). Tribological properties of bio-mimetic nano-patterned polymeric surfaces on silicon wafer. *Tribology Letters*, 21(1), 31–37. doi:10.1007/s11249-005-9005-4

Zhang, F., Chan, J., & Low, H. Y. (2008). Bio-mimetic, hierarchical structures on polymer surfaces by sequential imprinting. *Applied Surface Science*, 254(10), 2975–2979. doi:10.1016/j.apsusc.2007.10.061

Zou, M., Cai, L., Wang, H., Yang, D., & Wyrobek, T. (2005). Adhesion and friction studies of a selectively micro/nano-textured surface produced by UV assisted crystallization of amorphous silicon. *Tribology Letters*, 20(1), 43–52. doi:10.1007/s11249-005-7791-3

ADDITIONAL READING

Arvind Singh, R., Satyanarayana, N., Kustandi, T. S., & Sinha, S. K. (2011). Tribo-functionalizing Si and SU8 materials by surface modification for application in MEMS/NEMS actuator based devices. *Journal of Physics. D, Applied Physics*, 44(1), 015301. doi:10.1088/0022-3727/44/1/015301

Arvind Singh, R., & Suh, K. Y. (2013). Biomimetic patterned surfaces for controllable friction in micro- and nanoscale devices. *Micro and Nano Systems Letters*, 1, 1–11.

Nosonovsky, M., & Bhushan, B. (Eds.). (2012). *Green Tribology*. Springer-Verlag Berlin Heidelberg. doi:10.1007/978-3-642-23681-5

KEY TERMS AND DEFINITIONS

Adhesion: It is the tendency of surfaces to cling/stick to one another. Materials with higher surface energy exhibit higher surface adhesion and vice versa.

Biomimetics: It is the use and implementation of concepts and principles found in nature/natural materials for man-made technology.

DLC (Diamond-Like Carbon): It is a class of amorphous carbon material that has certain properties similar to that of diamond (e.g. hardness). Usually it is applied to surfaces in the form of thin films/coatings by physical/chemical vapor deposition techniques. DLCs are semi-hydrophobic and are good lubricant materials.

Friction: It is the resistance to motion. It is the resistance that one surface encounters when moving over another. Friction causes loss in energy.

Lotus Effect: It is the phenomenon of super hydrophobicity and self-cleaning displayed by the leaves of lotus plant. The unique characteristic of lotus leaves to avoid getting wet is due to the presence of micro-protuberances and waxy nano-crystals on the lotus leaves surfaces.

MEMS (Micro-Electro-Mechanical Systems): They are micro/nano-scale devices that perform sensing and/or actuating functions. They are made using the techniques of microfabrication. Silicon is the most popular MEMS material.

Real Contact Area: It is the real area of contact between two surfaces that are in contact under tribological conditions. The real contact area constitutes the areas where asperities of the contacting surfaces meet. The real area of contact is lower by 3 to 4 orders of magnitude when compared to the apparent area of contact (geometrical area).

SAMs (Self-Assembled Monolayers): They are molecular assemblies of organic molecules, which are formed on surfaces by adsorption or chemical grafting. The thickness of SAMs is typically in the range of few nanometers to tens of nanometers.

Tribology: It is the science and engineering of interacting surfaces in relative motion. It encompasses the study and application of the principles of friction, wear and lubrication.

Wear: It is the unintentional material removal/loss under relative mechanical motion between contacting surfaces.

Building Gene Networks by Analyzing Gene Expression Profiles

Crescenzo Gallo

University of Foggia, Italy

INTRODUCTION

Since the detection of the composition of DNA our understanding of biological structures and processes has expanded to a great extent, mostly thanks to computer science which plays a fundamental role in the field of bioinformatics. The main target at present is to analyze and employ the huge amount of accessible data. It is particularly important to distinguish various diseases through useful selection of gene indicators for morbid state and information about the possible correlations between genes.

Data analysis is seen as the largest and possibly the most important area of microarray bioinformatics to obtain the above said targets. Some specific data analysis methods address the fundamental scientific questions about microarray data, that is:

1. Which genes are differentially expressed in one set of samples relative to another,
2. What are the associations between the genes or samples being observed, and
3. Is it possible to group samples based on gene expression values?

In the next section, we illustrate the basic concepts underlying the previous questions and the bioinformatics research. Then we describe the methods for the first of these questions: the search for differentially (up or down) expressed genes. The following sections address the other two topics of clustering and classifying gene profiles. In the end, we show some concerns and issues of interest for future study and development in the field of (microarray) bioinformatics.

BACKGROUND

Gene expression profiling is an extensively used method in the analysis of microarray data. The leading hypothesis is that genes with similar expression profiles are co-regulated and are probably connected functionally.

Cluster analysis helps reaching these objectives; in particular, gene expression clusters help typify unknown genes assigned to the cluster by those genes that have a known function, and are the support for distinguishing common upstream regulatory sequence elements (Brazma *et al.*, 2000).

Clustering of expression profiles and functional grouping is especially compelling if the complete gene set is known. Hence, we used the large publicly available data set included in the Stanford Yeast Database at <http://genome-www.stanford.edu> for our clustering study.

Many applications aim at the molecular classification of diseases based on gene expression profiling and clustering. See, for example, works on leukemias (Golub *et al.*, 1999) and B-cell lymphomas (Alizadeh *et al.*, 2000). These and other studies confirm the usefulness of microarray bioinformatics for scientific and industrial research.

Gene Expression Profiling

Having M array probes with N samples (time points, patient tissues, *etc.*), you construct an $M \times N$ data matrix (the *gene expression matrix*), where the M rows describe the gene expression values across the experiments and the N columns (samples) describe the experiments across the

gene set. Practically, each gene is assigned a set of (possibly normalized) numerical values (the *gene expression profile*) corresponding to the gene's "presence" in each sample.

Then, an $M \times M$ similarity matrix is obtained by calculating the "closeness" between each gene pair with values inversely proportional to the relative (expression profile) distance. Typically Pearson correlation, Spearman's rank correlation, Hamming distance, Euclidean distance and mutual information are employed as similarity measures (Jain & Dubes, 1988; Mirkin, 1996), each having its specific advantages and disadvantages.

The (normalized and possibly log-transformed) expression profiles of thousands of genes are first examined under the typical two-fold comparison (between the same patients in *control vs treatment*, or between two different groups of patients), in order to identify differentially expressed genes. This can be done using an appropriate statistic (*t*-statistic, Wilcoxon, Mann-Whitney depending on the type of comparison) to compare gene expression variability.

Then selected genes are clustered in order to find groups of co-regulated genes.

The clustering output in the end serves as the basis for typifying the role of undetermined genes by means of information available from known genes, to recognize supposed regulatory elements in the early regions of the genes in the same clusters and reduce redundancy and generate averaging profiles in the context of regulatory networks.

Cluster Analysis

Cluster analysis (Duda *et al.*, 2001; Jain & Dubes, 1998; Jain *et al.*, 1999) can be summarized as follows. Let n experimental outcomes $x_i \in \mathbb{R}^m$ ($i=1..n$; each point has m components): the objective is to identify the underlying structure of the data, partitioning the n points into k clusters in order to group in the same cluster points "closer" to each other than to points belonging to different clusters.

In the above statement no clear definition exists for "closer" points, and this depends on the

resolution at which the data are viewed. The last issue is typically addressed by generating a tree of clusters (a dendrogram), whose number and structure depend on the resolution that is used.

Two of the most common methods of clustering gene expression data are hierarchical clustering and k -means clustering (Geraci *et al.*, 2009; Jain *et al.*, 1999).

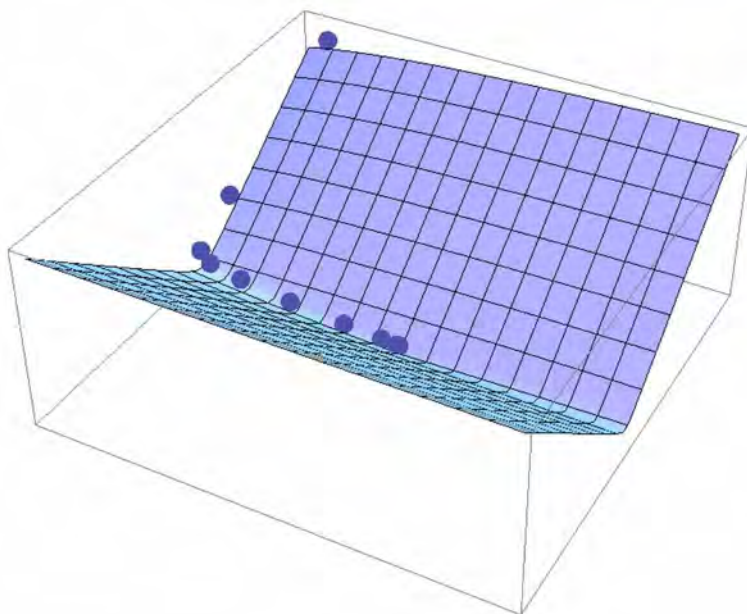
Hierarchical clustering is the most used, and produces a representation of the data with the most similar patterns grouped in a hierarchy of subsets. This method, however, suffers from considerable problems when applied to data containing a significant amount of *noise*, revealing itself of low applicability. In this case the solutions may not be unique and be data-order dependent. Mathematically, hierarchical clustering involves computing a matrix of all distances for each expression measurement in the study, merging and averaging the values of the closest nodes, and repeating the process until all nodes are merged into a single node.

K -means clustering involves generating cluster centers in n -dimensions and computing the distance of each data point from each of the cluster centers. The data points are assigned to their closest cluster center. A new cluster position is then computed by averaging the data points assigned to the cluster center. The process is repeated until the positions of the cluster centers stabilize.

UNSUPERVISED NEURAL NETWORKS

Artificial Neural Networks can be used not only for prediction but also for data classification. Unlike regression problems, where the goal is to produce a particular output value for a given input, classification problems require labeling of all data as belonging to one of n known classes. These classification models are typical cases of supervised networks, in which it is a priori possible to associate data to clusters and to train the neural network for the classification of further data.

Figure 1. Metaphor of the hyper-surface for the unsupervised neural network



However, there are cases in which the association is not known before. It is then up to the neural network to independently find the associative structures between data, grouping them in clusters. These neural networks are known as *unsupervised* or *self-organizing* (associative memories).

There is not, therefore, a *correct* output as there is not an *erroneous* output. This involves careful observation and analysis by the researcher. In fact, after an unsupervised network has been trained, it must be tested in many directions as it is important to understand what are the data structures resulting from the neural network itself.

An unsupervised network can be compared to a hypersurface in the n -dimensional space, with m “valleys” or minima corresponding to the desired shape of the outputs. The presentation of an input $X_i = \{x_{i1}, \dots, x_{in}\}$ to the network is equivalent to putting a weight in the corresponding point on the surface, thus making it fall into the nearest valley, corresponding to $Y_i = \{y_{i1}, \dots, y_{im}\}$. Obviously, an input X_i' “similar” to X_i — according to some similarity criterion — should fall into the same depression, drawing by association (hence the

term “associative memories”) the same output Y_i (Cammarata, 1990).

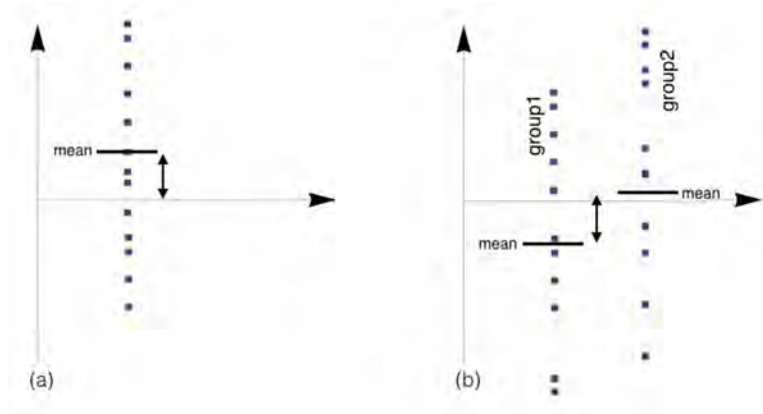
Practically, an unsupervised neural network consists of a number of *codebook vectors* Y_i which constitute the center of each cluster. They are the same size of the input space, and their components are the parameters of the unsupervised network. The codebook vectors are the network’s neurons. Some parameters help in training the neural network. The position of the vectors should be adjusted so that the average Euclidean distance between the data points and the corresponding codebook vector should be minimal (Yanai & De Lisi, 2002).

If you also minimize the distance between codebook vectors, the unsupervised network is known as Self-Organizing Map (SOM) or Kohonen network (Herrero *et al.*, 2001).

DIFFERENTIAL GENE EXPRESSION

This section illustrates a variety of methods, ranging from classical and more modern statistical

Figure 2. Example of paired (a) and unpaired (b) data



theory, to give a detailed path for finding differentially expressed genes in DNA microarray data. Let's start by illustrating the meaning of "differentially expressed genes".

DNA microarray data is made of a matrix of numerical values of gene expression levels taken from various patients' samples. We aim at recognize genes that are differentially (i.e., more or less) expressed under some different experimental conditions (ill and healthy patients, before and after some therapy).

Usually, each gene in the microarray is inspected to decide whether or not it is differentially expressed, typically doing this by studying many individual genes in parallel. This differs from the clustering and classification methods of the next sections, which are specifically interested in the interactions between the genes under investigation.

We will now show how expression data can be reshaped and examined for discovering differentially expressed genes.

In paired data there are two measurements from each patient, one before and one after treatment (Figure 2a). To determine whether a gene has been up- or down-regulated after treatment, you may compute the "after/before" fold ratio by considering the log-transformed values. This allows you to simply subtract (and not divide) the "after" from the "before" value, thus obtaining a set of single values whose mean (or median) you want to know if it is different from zero.

In unpaired data (Figure 2b) there are two (independent) groups of patients, and you search for those genes differentially expressed between the two groups. For each patient from each group there is a gene expression value; you want to know whether the mean (or median) of the two groups is different.

If you have a data set with more than two groups then a more complex approach, typically the analysis of variance (ANOVA), is required.

THE "RELAXED" CLUSTERING ALGORITHM

Preliminaries

The interpretation of the large amounts of data produced by DNA microarray techniques for the analysis of gene expression requires new efficient strategies to reduce the size involved. Clustering algorithms typically group genes (or samples) in clusters of similar expression profiles to identify possible functional relationships between them. Of particular importance are graphical representations of the clusters and their *automatic* annotation from available genome databases (Eisen *et al.*, 1998; Furey *et al.*, 2000; Golub *et al.*, 1999; Pe'er *et al.*, 2002; Wu *et al.*, 2000; Zhou *et al.*, 2002). Similar problems are found in the analysis of large networks, where you try to extract subnets that

meet certain criteria (such as the search for web pages on the same subject) (Eckmann & Moses, 2002; Jenssen *et al.*, 2001; Milo *et al.*, 2002; Shen-Orr *et al.*, 2002; Butte *et al.*, 2000; Yanai & De Lisi, 2002).

Here we analyze an algorithm, based on correlation graphs, that combines network analysis with the classical clustering for the study of DNA microarray data. The algorithm depends on two parameters (the *correlation* and *clustering* thresholds), resulting in an effective ability to analyze the sensitivity of the results in response to the particular configuration chosen.

The algorithm is based on the concept of “curvature” (Eckmann & Moses, 2002; Watts & Strogatz, 1998) applied to the network (correlation graph) of co-expressed genes, where the nodes are the genes (or samples) and the edges represent the co-expression between linked genes. Then a certain number of clusters are identified that correspond to connected components of the graph with a high clustering index, defined as the local density of triangular relations (Collet & Eckmann, 2002). The clusters of genes (or samples) are the densest regions of the correlation graph; communities with high clustering index are low-entropy structures, with obvious biological relevance.

The algorithm has been implemented in *Mathematica* (due to its intrinsic computational power and functional paradigm) and applied to some publicly available datasets: *Saccharomyces cerevisiae* (Eisen *et al.*, 1998) and DLBCL lymphoma (Alizadeh *et al.*, 2000). The computational performances obtained are comparable to those required by Hierarchical (Jain *et al.*, 1999) and Markov Clustering (Van Dongen, 2000).

The Clustering Algorithm

A set of DNA microarray data consists of expression levels of N genes in M different experimental conditions (RNA samples). The algorithm (based on the correlation graph approach) is articulated according to the following steps.

1. Read input DNA microarray data X .
2. The $N \times N$ correlation matrix P is defined on all pair of genes $x_{k,l}$ ($k, l = 1..N$) whose elements are computed using the Spearman's rank correlation coefficient¹:

$$\rho_{k,l} = 1 - \frac{(6 \times S_{j=1,M} d_j^2)}{(n^2 - 1)n} \in [-1, 1]$$

where

$$d_j^2 = x_k(j) - x_l(j).$$

Let θ a fixed threshold defining a range of symmetrical comparison $(-\theta, \theta)$. Construct the square adjacency matrix A between genes whose elements $a_{k,l}$ ($k, l = 1..N$) are defined as follows:

$$\begin{cases} a_{k,l} = 0 & \text{if } \rho_{k,l} \in (-\theta, \theta), \\ 1 & \text{otherwise} \end{cases}$$

3. A correlation (undirected) graph G is constructed from the adjacency matrix, with edge weights equal to the correlation coefficients $\rho_{k,l}$.
4. Each node i of the correlation graph has an associated clustering coefficient $c(i) = 2\alpha / \beta(\beta-1)$, where α is the number of real and β is the number of possible triangular clusters for node i .
5. Let $\gamma \in [0, 1]$ (clustering threshold). From the graph G we induce a subgraph H made of all nodes $i \in G$ (and related edges e) with “relaxed” clustering coefficient $c_r(i)$ greater than or equal to γ .

$$H = \{\text{nodes } i \in G \text{ and related edges } e: c_r(i) \geq \gamma\}$$

The “relaxed” clustering coefficient for a node i is given by the maximum among its clustering coefficient $c(i)$ and those of the nodes directly connected to it: this gives a “tethering” effect by the surrounding nodes, better taking into account the complex structure of the relations between the

(correlation) graph nodes than the simple clustering coefficient of each node.

6. From the subgraph H we extract all connected components (made of at least two linked nodes) representing the classes (clusters) identified by the values chosen for thresholds (algorithm's parameters) θ and γ .

To better understand how the algorithm operates with thresholds θ and γ you may think of the initial graph G as a mountain range. The clustering threshold γ is the fog extending up to a certain height and clusters are emerging mountain peaks. Varying the correlation threshold θ means changing the soil structure, while γ only raises or lowers the fog level and, therefore, modifies the existence and relative size of clusters.

ANALYSIS OF GENE EXPRESSION PROFILES: EXAMPLES

Analysis of Gene Expression Profiles in *Saccharomyces Cerevisiae*

Figures 3 and 4 represent the computed results based on the expression profiles of 6,221 *Saccharomyces* genes, crossed with 80 available samples. Even with relatively high values of correlation and clustering we obtain well-defined clusters. Figure 1 represents the final subgraph with correlation $\theta = 0.90$ and clustering $\gamma = 0.70$ derived from an initial graph with 1,917 links. Note that only 237 of 6,221 genes (3.81%) have a nonzero clustering coefficient. Figure 2 shows the subgraph corresponding to $\theta = 0.80$ and $\gamma = 0.64$ starting from 19,026 links, with 1,112 genes (17.87%) showing nonzero clustering coefficient.

Many of the obtained clusters have an acceptable biological basis. For example, Cluster #7 of Figure 3 and Cluster #13 of Figure 4 contain (with obvious differences) almost all the genes for the production of histones: This cluster is quite stable against the variation of threshold parameters. Larger clusters have also been detected, such as

the cluster of protein synthesis (Cluster #10 of Figure 3 and #2 of Figure 4). The latter is likely to vary substantially with varying thresholds, and can end up with including hundreds of genes.

Figure 5. shows a visual comparison of a particular cluster belonging to graphs of Figures 3 and 4. This comparison shows how the two-dimensional clustering analysis here implemented with correlation graphs poses a greater resolution detail in carrying out the clustering of the genes involved. This turns out to be an important factor for an adequate biological study, eliminating during analysis those links between genes that appear to be not too strong to resist a greater resolution level.

Analysis of Gene Expression Profiles in *Lymphoma*

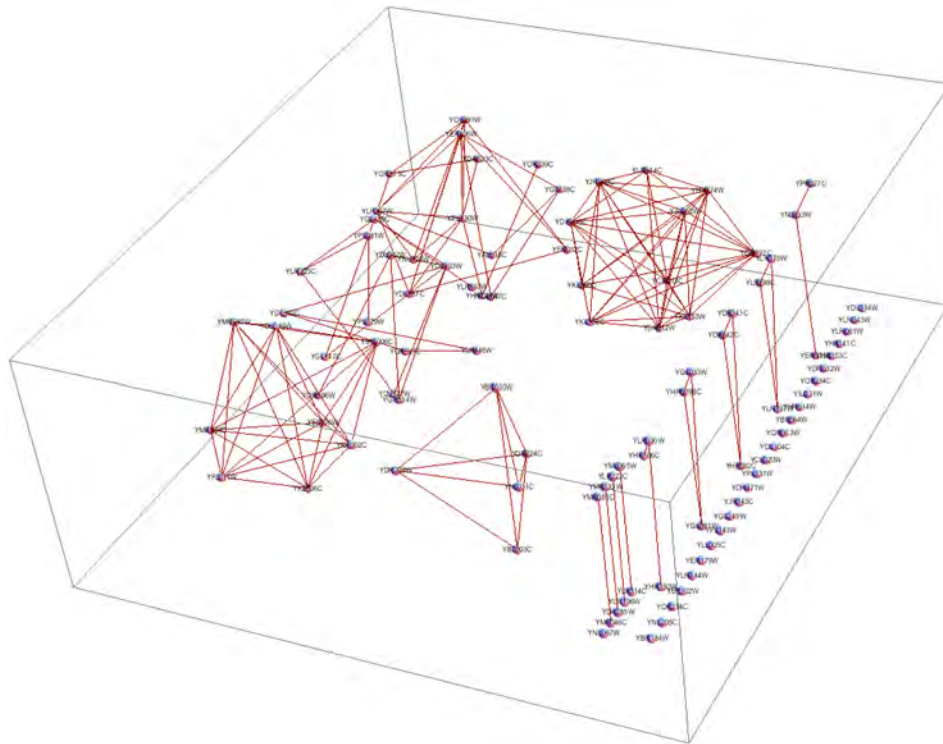
Graphs in Figures 6 and 7 represent the results based on expression profiles of 4,026 lymphoma genes crossed with 96 available samples. Both graphs have a correlation threshold $\theta = 0.80$, with 568 of 4,026 genes with nonzero clustering coefficient (14.11%) and 1,872 initial links.

Classification obtained with $\gamma = 0.40$ shows in Cluster #75 a group of under-expressed genes in CLL (Chronic Lymphocytic Leukaemia) tumors. With $\gamma = 0.24$ Cluster #76 represents a group of over-expressed genes in DLCL (Diffuse Large B-Cell Lymphoma) tumors.

Lymphoma Samples' Analysis of Expression Profiles

Figure 8 shows the classification based on 96 samples of the lymphoma data set (this is the transpose of the gene expression matrix, made of 96 rows and 4,026 columns). The thresholds are $\theta = 0.60$ and $\gamma = 0.30$. There are 17 samples (17.71%) with nonzero clustering coefficient, with 58 initial links. The clusters obtained are clearly identified and are associated with a single sample subtype: Cluster #1 corresponds to T-cells sample, Cluster #2 to FL (Follicular Lymphoma) and Cluster #3 to CLL.

*Figure 3. Correlation graph of genes in *Saccharomyces cerevisiae* with $\theta = 0.90$ and $\gamma = 0.70$*



*Figure 4. Correlation graph of genes in *Saccharomyces cerevisiae* with $\theta = 0.80$ and $\gamma = 0.64$*

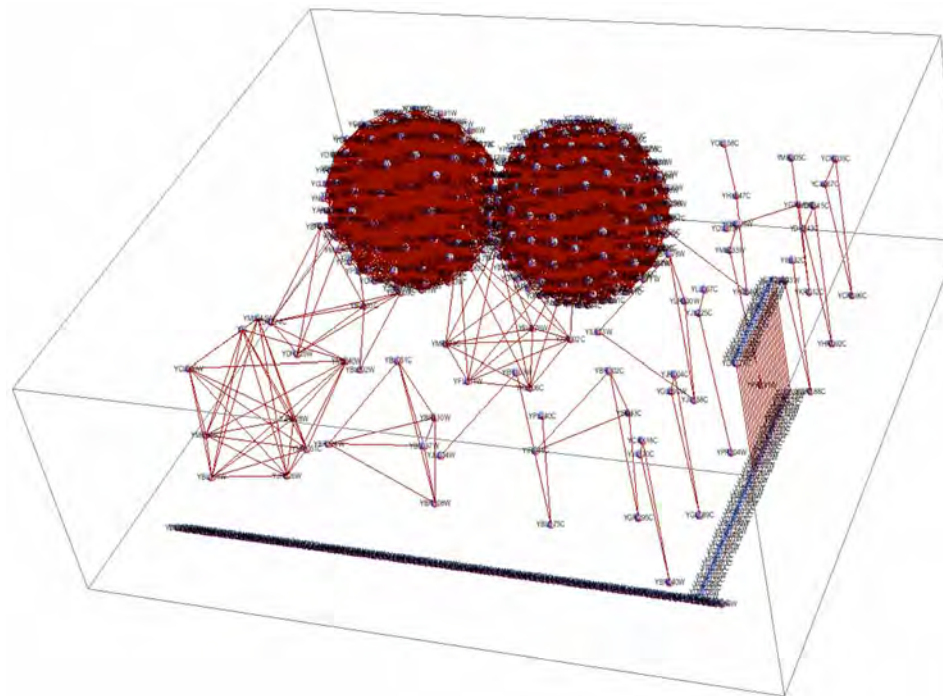
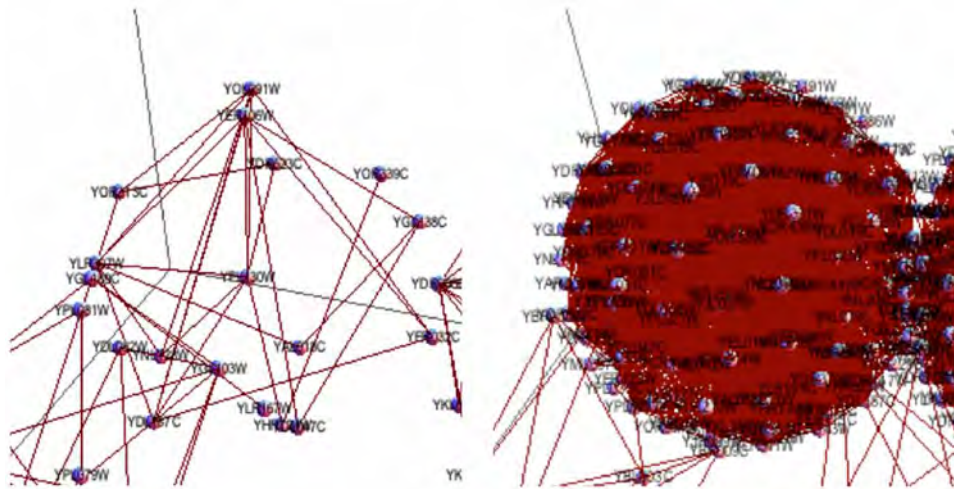
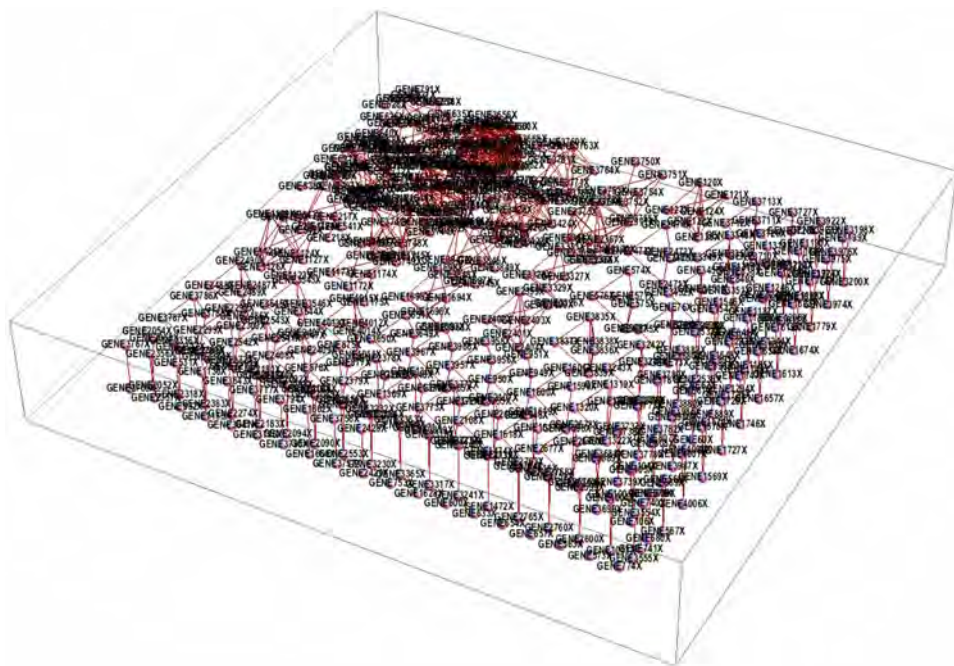


Figure 5. A cluster comparison between graphs in Figures 3 and 4

Figure 6. Correlation graph of genes in Lymphoma with $\theta = 0.80$ and $\gamma = 0.40$ 

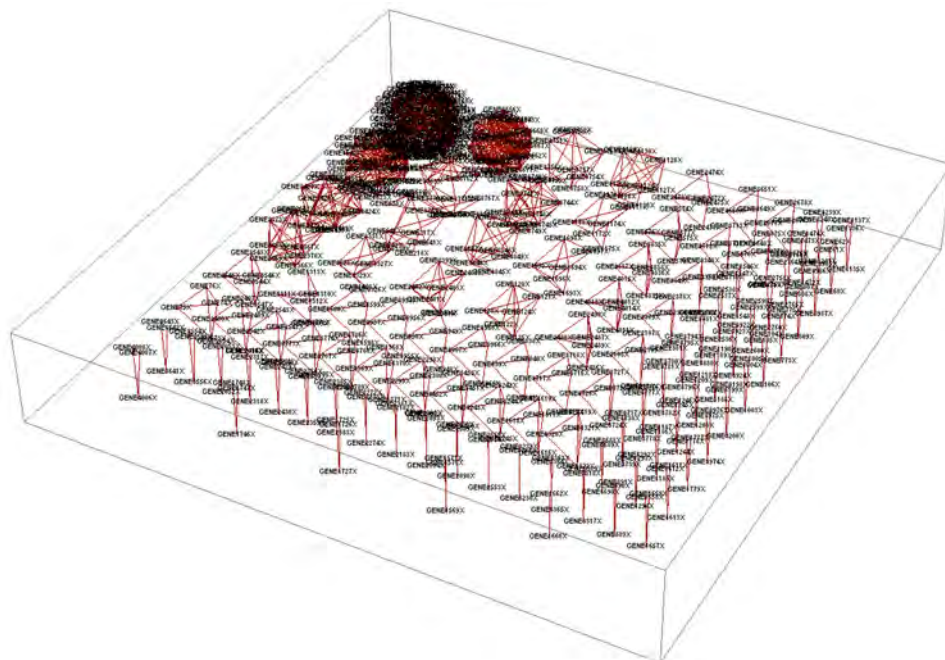
Clustering Microarray Data with Neural Networks

The clustering analysis with neural networks has been implemented on the same input data on which the classification was carried out using correlation graphs, that is the gene expressions in *Saccharomyces cerevisiae* (Eisen *et al.*, 1998) and DLBCL lymphoma (Alizadeh *et al.*, 2000).

Clustering Gene Expression in *Saccharomyces Cerevisiae*

The search for gene expression clusters in *Saccharomyces* fulfilled through unsupervised neural networks provides the basic tool for performing an additional analysis to that of the correlation graphs. Obviously there are some differences.

Figure 7. Correlation graph of genes in Lymphoma with $\theta = 0.80$ and $\gamma = 0.24$



The analysis carried out with unsupervised neural networks needs to ensure that all possible genes of our original data set are put into predicted clusters. This means classifying 6,221 genes and looking for, in the aggregate source, the biological meaning of that classification.

Figure 9. shows the fitting parameters of the neural network and the average distance obtained. The latter turns out to be 3.8085 and all genes are partitioned into 14 clusters. From this point of view, the correlation graphs' algorithm offers an advantage over the analysis made with neural networks: With the ability to intervene on two thresholds you obtain a greater "filtering" in the analysis of results.

Classification of Lymphoma Samples

In this section we focus on analyzing the lymphoma dataset by observing samples with reference to genes. In this context, Figure 10 shows the characteristic elements of the unsupervised neural network in the various clustering iterations, with an average distance equal to 43.1945.

FUTURE RESEARCH DIRECTIONS

By identifying combinations of genes that "cooperate" in order to obtain specific bio-logical results, it is possible to identify such genes more attractive for future research and, potentially, specific drug targets to improve the quality of the therapeutic tools available. This can be done by:

1. Filtering those genes that are more up/down-regulated (differentially expressed), and thus possibly responsible for outbreak of disease, and
2. Discovering gene groups possibly associated to different pathological states.

To illustrate the performance of the methodology, we applied it to two independent datasets from patients with various disease subtypes. The results show the validity of the proposed approach and the selection of key deregulated genes recently reported in the literature.

Figure 8. Correlation graph of Lymphoma with $\theta = 0.60$ and $\gamma = 0.30$

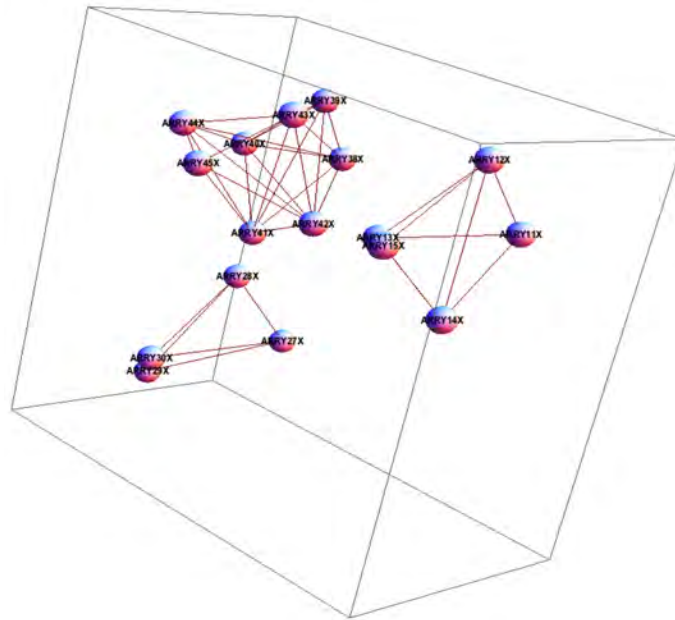


Figure 9. Classification of *Saccharomyces cerevisiae*: Initialization graph and training of (SOM) unsupervised neural network's parameters

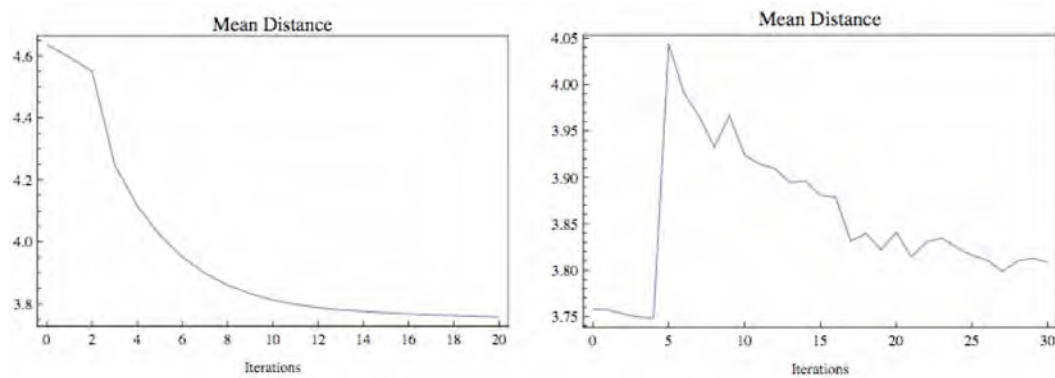
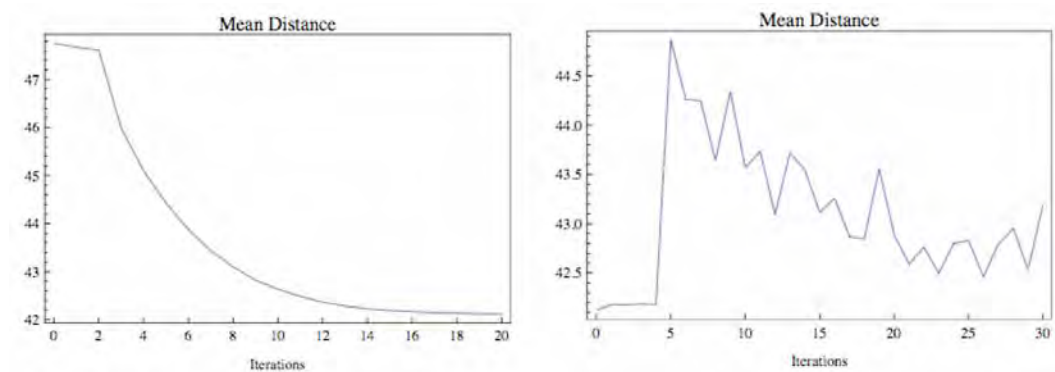


Figure 10. Classification of Lymphoma: Initialization graph and training of (SOM) unsupervised neural network's parameters



CONCLUSION

It is increasingly accepted the hypothesis that complex diseases are the result of events that occur in various ways. Through the development of computational tools that detect the involvement of multiple genes interacting with each other in specific experimental situations, you can provide solutions that go beyond the capacity offered by those dealing with single genes.

In this chapter we illustrate the analysis of DNA microarray data obtained from the expression of multiple genes, using both special parametric clustering (correlation graphs) and classification techniques with artificial neural networks. The greatest usefulness of the classification procedure described here and its computational approach is that it requires no prior knowledge of the specific functions of genes or interactions between them to produce meaningful results. Instead, it represents an appropriate means of exploring genomic “situations” not yet well investigated, a target within the reach of the individual researcher with the power and flexibility of the computing environments now available.

REFERENCES

- Albert, R., Jeong, H., & Barabasi, A. L. (2000). Error and attack tolerance of complex networks. *Nature*, 406(6794), 378–382. doi:10.1038/35019019 PMID:10935628
- Alizadeh, A. A., Eisen, M. B., Davis, R. E., Ma, C., Lossos, I. S., Rosenwald, A., & Staudt, L. M. et al. (2000). Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature*, 403(6769), 503–511. doi:10.1038/35000501 PMID:10676951
- Barabasi, A. L. (2009). Scale-free networks: A decade and beyond. *Science*, 325(5939), 412–413. doi:10.1126/science.1173299 PMID:19628854
- Barabasi, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439), 509–512. doi:10.1126/science.286.5439.509 PMID:10521342
- Brazma, A., & Vilo, J. (2000). Gene expression data analysis. *FEBS Letters*, 480(1), 17–24. doi:10.1016/S0014-5793(00)01772-5 PMID:10967323
- Butte, A., Tamayo, P., Slonim, D., Golub, T., & Kohane, I. (2000). Discovering functional relationships between RNA expression and chemotherapeutic susceptibility using relevance networks. *Proceedings of the National Academy of Sciences of the United States of America*, 97(22), 12182–12186. doi:10.1073/pnas.220392197 PMID:11027309
- Cammarata, S. (1990). *Reti Neuronali. Una introduzione all’ “altra” intelligenza artificiale*. Etaslibri.
- Collet, P., & Eckmann, J. (2002). The number of large graphs with a positive density of triangles. *Journal of Statistical Physics*, 109(5/6), 923–943. doi:10.1023/A:1020489507547
- Cruz, J., & Wishart, D. (2006). Applications of machine learning in cancer prediction and prognosis. *Cancer Informatics*, 2, 59–77. PMID:19458758
- De Ridder, D., De Ridder, J., & Reinders, M. J. T. (2013). Pattern recognition in bioinformatics. *Briefings in Bioinformatics*, 14(5), 633–647. doi:10.1093/bib/bbt020 PMID:23559637
- Duda, O., Hart, P. E., & Stork, D. (2001). *Pattern Classification*. New York: John Wiley & Sons Inc.
- Eckmann, J., & Moses, E. (2002). Curvature of co-links uncovers hidden thematic layers in the world wide web. *Proceedings of the National Academy of Sciences of the United States of America*, 99(9), 5825–5829. doi:10.1073/pnas.032093399 PMID:11972019

- Eisen, M., Spellman, P., Brown, P., & Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences of the United States of America*, 95(25), 14863–14828. doi:10.1073/pnas.95.25.14863 PMID:9843981
- Furey, T., Cristianini, N., Duffy, N., Bednarski, D., Schummer, M., & Haussier, D. (2000). Support vector machine classification and validation of cancer tissue samples using microarray data. *Bioinformatics (Oxford, England)*, 16(10), 906–914. doi:10.1093/bioinformatics/16.10.906 PMID:11120680
- Geraci, F., Leoncini, M., Montangero, M., Pellegrini, M., & Renda, M. E. (2009). K-Boost: A Scalable Algorithm for High-Quality Clustering of Microarray Gene Expression Data. *Journal of Computational Biology*, 16(6), 859–873. doi:10.1089/cmb.2008.0201 PMID:19522668
- Golub, T. R. et al.. (1999). Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. *Science*, 286(5439), 531–538. doi:10.1126/science.286.5439.531 PMID:10521349
- Haferlach, T. (2006). Clinical utility of microarray-based gene expression profiling in the diagnosis and subclassification of leukemia: report from the International Microarray Innovations in Leukemia Study Group. *Briefings in Bioinformatics*, 7, 86–112. PMID:16761367
- Herrero, J., Valencia, A., & Dopazo, J. (2001). A hierarchical unsupervised growing neural network for clustering gene expression patterns. *Bioinformatics (Oxford, England)*, 17(2), 126–136. doi:10.1093/bioinformatics/17.2.126 PMID:11238068
- Jain, A. K., & Dubes, R. C. (1988). *Algorithms for clustering data*. Englewood Cliffs, NJ: Prentice-Hall.
- Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: A review. *ACM Computing Surveys*, 31(3), 264–323. doi:10.1145/331499.331504
- Jenssen, T., Laegreid, A., Komorowski, J., & Hovig, E. (2001). A literature network of human genes for high-throughput analysis of gene expression. *Nat Gene*, 38, A21–28.
- Joo, J., Kwak, M., Chen, Z., & Zheng, G. (2010). Efficiency robust statistics for genetic linkage and association studies under genetic model uncertainty. *Statistics in Medicine*, 29, 158–180. PMID:19918942
- Joo, J., Kwak, M., & Zheng, G. (2010). Improving power for testing genetic association in case-control studies by reducing alternative space. *Biometrics*, 66(1), 266–276. doi:10.1111/j.1541-0420.2009.01241.x PMID:19397584
- Ley, T. J. et al.. (2013). Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *The New England Journal of Medicine*, 368(22), 2059–2074. doi:10.1056/NEJMoa1301689 PMID:23634996
- Lockhart, D., Dong, H., Byrne, M. C., Follettie, M. T., Gallo, M. V., Chee, M. S., & Brown, E. L. et al. (1996). Expression monitoring by hybridization to high density oligonucleotide arrays. *Nature Biotechnology*, 14(13), 1675–1680. doi:10.1038/nbt1296-1675 PMID:9634850
- Mesko, B., Poliska, S., & Nagy, L. (2011). Gene expression profiles in peripheral blood for the diagnosis of autoimmune diseases. *Trends in Molecular Medicine*, 17(4), 223–233. doi:10.1016/j.molmed.2010.12.004 PMID:21388884
- Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., & Alon, U. (2002). Network motifs: Simple building blocks of complex networks. *Science*, 298(5594), 824–827. doi:10.1126/science.298.5594.824 PMID:12399590
- Mirkin, B. (1996). *Mathematical classification and clustering*. Dordrecht, The Netherlands: Kluwer Academic Publishers. doi:10.1007/978-1-4613-0457-9

- Newman, M. E. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences of the United States of America*, 103(23), 8577–8582. doi:10.1073/pnas.0601602103 PMID:16723398
- Peer, D., Regev, A., & Tanay, A. (2002). Minreg: Inferring an active regulator set. *Bioinformatics (Oxford, England)*, 18(Suppl 1), s258–s267. doi:10.1093/bioinformatics/18.suppl_1.S258 PMID:12169555
- Perou, C. et al. (1999). Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. *Proceedings of the National Academy of Sciences of the United States of America*, (96): 9112–9217. PMID:10430922
- Pirooznia, M., Yang, J. Y., Yang, M. Q., & Deng, Y. (2008). A comparative study of different machine learning methods on microarray gene expression data. *BMC Genomics*, 9. PMID:18366602
- Price, A. L., Zaitlen, N. A., Reich, D., & Patterson, N. (2010). New approaches to population stratification in genome-wide association studies. *Nature Reviews. Genetics*, 11(7), 459–463. doi:10.1038/nrg2813 PMID:20548291
- Qin, H., Morris, N., Kang, S. J., Li, M., Tayo, B., Lyon, H., & Zhu, X. et al. (2010). Interrogating local population structure for fine mapping in genome wide association studies. *Bioinformatics (Oxford, England)*, 26(23), 2961–2968. doi:10.1093/bioinformatics/btq560 PMID:20889494
- Shen-Orr, S., Milo, R., Mangan, S., & Alon, U. (2002). Network motifs in the transcriptional regulation network of escherichia coli. *Nature Genetics*, (298): 824–827. PMID:11967538
- Slawski, M., Daumer, M., & Boulesteix, A. (2007). CMA: A comprehensive Bioconductor package for supervised classification with high dimensional data. *BMC Bioinformatics*, 9, 439.
- Sneath, P., & Sokal, R. (1973). *Numerical Taxonomy*. San Francisco: W.H. Freeman.
- Thomas, D. C. (2010). Gene-environment-wide association studies: Emerging approaches. *Nature Reviews. Genetics*, 11(4), 259–272. doi:10.1038/nrg2764 PMID:20212493
- Van Dongen, S. (2000). *Graph Clustering by Flow Simulation* (PhD Thesis). University of Utrecht, The Netherlands.
- Wang, X., Elston, R. C., & Zhu, X. (2011). Statistical interaction in human genetics: How should we model it if we are looking for biological interaction? *Nature Reviews. Genetics*, 12(1), 74. doi:10.1038/nrg2579-c2 PMID:21102529
- Watts, D., & Strogatz, S. (1998). Collective dynamics of small-world networks. *Nature*, 393(6684), 440–442. doi:10.1038/30918 PMID:9623998
- Wu, C., Zhang, H., Liu, X., DeWan, A., Dubrow, R., Ying, Z., & Ying, Z. et al. (2010). Detecting essential and removable interactions in genome-wide association studies. *Statistics and Its Interface*, 2, 161–170. PMID:21165165
- Wu, L., Hughes, T., Davierwala, A., Robinson, M., Stoughton, R., & Altschuler, S. (2002). Large-scale prediction of saccharomyces cerevisiae gene function using overlapping transcriptional clusters. *Nature Genetics*, 31(3), 255–265. doi:10.1038/ng906 PMID:12089522
- Yanai, I., & De Lisi, C. (2002). The society of genes: networks of functional links between genes from comparative genomics. *Genome Biology*, (3).
- Zhang, K., Pirooznia, M., Arabnia, H. R., Yang, J. Y., Wang, L., Luo, Z., & Deng, Y. (2011). Genomic signatures and gene networking: Challenges and promises. *BMC Genomics*, 12(Suppl 5), 1. doi:10.1186/1471-2164-12-S5-I1 PMID:22369358
- Zhang, K., & Zhao, H. (2000). Assessing reliability of gene clusters from gene expression data. *Functional & Integrative Genomics*, 1(3), 156–173. doi:10.1007/s101420000019 PMID:11793234

Zhang, Z., Ersoz, E., Lai, C.-Q., Todhunter, R. J., Tiwari, H. K., Gore, M. A., & Buckler, E. S. et al. (2010). Mixed linear model approach adapted for genome-wide association studies. *Nature Genetics*, 42(4), 355–360. doi:10.1038/ng.546 PMID:20208535

Zhou, X., Kao, M., & Wong, W. (2002). Transitive functional annotation by shortest-path analysis of gene. *Proceedings of the National Academy of Sciences of the United States of America*, (99): 5825–5829. PMID:12196633

Ziegler, A., König, I., & Thompson, J. R. (2008). Biostatistics aspects of genome-wide association studies. *Biometrical Journal. Biometrische Zeitschrift*, 50(1), 8–28. doi:10.1002/bimj.200710398 PMID:18217698

KEY TERMS AND DEFINITIONS

Algorithm: An algorithm is a procedure that solves a given problem by a finite number of steps. A problem solved by an algorithm is said *computable*. The term “algorithm” is derived from the Latin transcription of the name of the Persian mathematician al-Khwarizmi, which is considered one of the first authors to have made reference to this concept.

Artificial Neural Network: ANNs are mathematical models that represent the interconnection between elements defined artificial neurons, i.e. mathematical constructs that to some extent mimic the properties of living neurons. These mathematical models can be used both to obtain an understanding of biological neural networks, but even more to solve engineering problems of artificial intelligence such as those that arise in various technological fields (in electronics, computer science, simulation, and other disciplines).

Bioinformatics: Bioinformatics is a scientific discipline devoted to the solution of biological problems at the molecular level with computer methods. It is an attempt to describe, in numerical and statistical terms, biological phenomena with a

set of analytical and numerical tools. In addition to information technology, bioinformatics uses applied mathematics, statistics, chemistry, biochemistry and concepts of artificial intelligence. Bioinformatics mainly deals with providing valid statistical models for the interpretation of data from experiments in molecular biology and biochemistry in order to identify trends and numerical laws; generate new models and mathematical tools for the analysis of sequences of DNA, RNA and proteins in order to create a body of knowledge concerning the frequency of relevant sequences, their evolution and possible function; and organize the knowledge acquired at the global level of genome and proteome databases in order to make such data accessible to all, and to optimize the data search algorithms to improve accessibility.

Cluster: Natural subgroup of a population, used for statistical sampling or analysis.

Clustering: Clustering or cluster analysis is a set of techniques of multivariate data analysis aimed at selecting and grouping homogeneous elements in a data set. Clustering techniques are based on measures relating to the similarity between the elements. In many approaches this similarity, or better, dissimilarity, is designed in terms of distance in a multidimensional space. Clustering algorithms group items on the basis of their mutual distance, and then the belonging to a set or not depends on how the element under consideration is distant from the collection itself.

DNA: Deoxyribonucleic acid (DNA) is a nucleic acid that contains the genetic information necessary to the biosynthesis of RNA and protein molecules essential for the development and proper functioning of most living organisms. The order in the sequential arrangement of the nucleotides A, T, C, G represents the genetic information, which is translated with the genetic code in the corresponding amino acids.

DNA Microarray: A DNA microarray (commonly known as gene chip, DNA chip, biochip array or high density) is a collection of microscopic DNA probes attached to a solid surface such as glass, plastic, or silicon chip forming an array

(matrix). Such arrays allow to simultaneously examine the presence of many genes within a DNA sample (which often can also represent the entire genome or transcriptome of an organism). A typical use is to compare the gene expression profile of an individual patient with that of a healthy one to identify which genes are involved in the disease.

Gene: The gene is the fundamental hereditary unit of living organisms. Genes correspond to portions of the genetic code localized in specific positions within the sequence (DNA or, more rarely, RNA) and contain all the information necessary for the production of a protein. They are contained and organized within chromosomes, present in all cells of an organism.

Gene Expression: In the field of molecular biology, gene expression profiling is the measure of the activity (expression) of thousands of genes at a time, to create a global picture of cellular function. These profiles can, for example, distinguish between cells that are in proliferation, or show how the cells react to a particular treat-

ment. Many experiments of this type measure an entire genome simultaneously. DNA Microarray technology measures the relative activity of target genes previously identified.

Pattern: In biology with pattern (sometimes “profile”) one refers to different types of regularity, such as the regularity of the biological sequences of DNA or proteins that allow the recognition and specific binding between molecules, or the regularity in the level of expression of the genes of cells which allow the recognition of different experimental cell types including tumor cell types, or the regularity in the events that occur during processes such as the development of an organism, or even the regularities in the behavior of animals.

ENDNOTE

- ¹ Spearman’s rank is more robust than Pearson’s correlation coefficient if there is a non linear correlation between data when data are not normally distributed.

Concerns and Challenges of Cloud Platforms for Bioinformatics

B**Nicoletta Dessì***Università degli Studi di Cagliari, Italy***Barbara Pes***Università degli Studi di Cagliari, Italy*

INTRODUCTION

In recent years, computer advances have played an important role in promoting scientific research in biological areas such as genomics, proteomics and other “-omic” subfields which rely heavily on suitable computational infrastructures for managing large-scale data. In particular, the flood of data from genome sequences has given rise to “bioinformatics”, an interdisciplinary research domain which employs a wide range of computational techniques derived from scientific disciplines (such as statistics, machine learning, applied mathematics etc..) for managing biological data. To get to understand the current application fields of bioinformatics, it is necessary to consider the following aspects.

A first aspect is about the massive production and spread of biological data around the web. Generated within a short period of time and stored in a growing number of web resources, the increasing amount of biological data has introduced new challenges about its management and exploitation.

For example, thanks to next generation sequencing instruments and ICT advances, areas of life sciences that were previously distant from each other (in the ideology, analysis practices, toolkits etc.) are now able to share and analyze data in transparent and reproducible fashion. This interdisciplinary research task calls for the integration of information with multiple levels of granularities from several web resources that often represent information and data in different ways.

In this respect, bioinformatics increasingly deals with providing technical approaches to support interdisciplinary scientific knowledge which relies on working with concepts from different areas in constant evolution and more and more requires experimental techniques, scientific approaches and collaborative management of data (Bosin, Dessì, & Pes, 2007).

A second aspect is about recent advances in computer science that significantly influence the development of computational tools in bioinformatics. Specifically, the service-oriented paradigm has provided a new way of thinking biological resources in terms of computational infrastructures by positioning services as primary functional elements for data integration. Several biomedical organizations (such as the National Center for Biotechnology Information (NCBI) and The National Center for Biomedical Ontology (NCBO)) provide web portals that expose Web services for searching data. Existing techniques for web content classification, search, and visualization seem to be actually inadequate to satisfy the biologist’s needs because accessing these heterogeneous systems from the Internet is not straightforward without the availability of standard and common interfaces.

In this respect, bioinformatics research is devoted to search explicit and automatic ways of joining information to improve the usability of web resources.

Finally, the rapid development of the Internet has provided an opportunity to investigate about

the use of state-of-the-art technology for the construction of a new generation of tools that integrate plain data sources, public programmable APIs and any kind of available services. Usually referred to as Web2.0 applications, these tools rely on open APIs or reusable services. The availability of biomedical ontologies dramatically increases the range of benefits and the usages derived from these applications that often support a deeper analysis of data by taking into account the semantic information (Dessì, Pascariello, & Pes, 2014).

Considering the aforementioned aspects, it is clear that bioinformatics research addresses three main challenges namely:

1. Storing and analyzing large amount of heterogeneous data.
2. Enabling knowledge extraction from several web resources and collaboration through user-friendly interfaces.
3. Promoting solutions for offering different categories of services to end-users.

Nowadays, the cloud computing paradigm represents a primary solution to these challenges as it extends the role of the Internet to enable a new form of distributed system for large-scale data processing.

As significant component of the Web 2.0 world, Cloud computing expresses a service-based architecture where computing services are delivered on-demand to customers over a network in a self-service fashion that is independent of device and its location. This innovative technology enables the introduction of new computational paradigms that could have a great influence on facing the above challenges. The application of cloud computing in bioinformatics is still preliminary, although an increasing number of biological applications is relying on the Cloud for processing large datasets (Calabrese & Cannataro, 2016; Schadt, Linderman, Sorenson, Lee, & Nolan, 2011). Conversely, few work has been done to explore the use of architectural cloud-based solutions (Mrozek, 2015).

This chapter aims investigating the extent to which cloud technology offers a viable platform for developing and deploying applications that support users in searching and integrating information offered by bioinformatics resources. The chapter outlines the basic features that such computing applications should exhibit and the challenging issues they deal with. The architecture and the functionality of the cloud-based environments are presented to stress how cloud platforms could offer added-value service components and flexibility that make their adoption attractive for bioinformatics.

The chapter also sketches some application scenarios that illustrate the potential of cloud-based environments for integrating information from multiple web resources, including biomedical portals.

BACKGROUND ON CLOUD COMPUTING

Cloud technology endows users with a service-based environment that provides hardware and system components and refers to software applications delivered as services over the Internet (Lin, Fu, Zhu, & Dasmalchi, 2009).

Cloud resources are usually organized into bottom-up layers, depending on the type of service:

- **Infrastructure-as-a-Service (IaaS):** Provides suitable hardware, including physical servers and network bandwidth services directly to end users which do not own enough processing power, storage capacity, applications to meet their needs. These resources can be ‘rented’ from a cloud provider on an as-needed basis and accessed from anywhere via an Internet connection. Cloud services employ a metering system that divides the computing resources in appropriate blocks and users pay for the service as an operating expense without incurring any significant initial capital costs.

- **Platform-as-a-Service (PaaS):** Provides a hosting platform in which users are allowed to develop applications without dealing with software issues such as operating system upgrades, implementation of compilers, installation and maintenance of databases etc.
- **Software-as-a-Service (SaaS):** It basically means that the software is available on demand, most of which is browser-based and devoted to a specific function. SaaS applications are often priced by a rental fee, which includes the application software license fees, software maintenance, and technical support costs.

Although provided at different level, Cloud Services communicate with each other and are self-contained i.e. each service provides the same functionality, independently of other services. The global infrastructure offered by a Cloud environment is made possible by abstracting the physical resources (i.e. storage, memory, network etc.) so that multiple operating systems can run on a single hardware platform concurrently. Commonly referred as virtualization, this abstraction is based on technologies that greatly improve resource utilization and offer several key advantages, such as a low entry cost to use compute-intensive resources, a pay-per-use model, a high scalability that is dynamically adjusted according to user demand.

The cloud computing platform composes of thousands or even tens of thousands nodes, which provide mass data storage, management and processing. Because these dynamically scalable and virtualized cloud resources are provided as services, users are not required to have expertise nor knowledge or control over the technological structure that supports their tasks.

THE DRIVING FORCES OF CLOUD COMPUTING IN BIOINFORMATICS

Cloud Computing could offer many *added-value services* that previous solutions do not have. In

fact, the primary difference is that offerings and services are not located in house, but outside the bioinformatics organization. As reported by Chae, Lee, Marru, Lee, & Kim (2013), a significant number of research institutions are experiencing the capacity limits of their computing facilities. Additionally, collecting and configuring suitable tools and resources for certain research purposes is non-trivial job, even for expert developers. By adopting cloud technology, bioinformatics organizations could avoid the installation and management of software on their own computers and benefit from centralized and automatic software updates. The ease of access, to and from cloud-based resources, promotes the use of cloud-based strategies such as MapReduce a run time system which automatically partitions input data and schedules the executions of programs for large datasets in a large cluster of commodity machines (Taylor, 2010).

As well, its open-source implementation Hadoop has a widespread adoption in bioinformatics. Moreover, cloud paradigm promotes the development of biomedical applications specifically designed to take advantage of cloud infrastructure. Meanwhile, new functionality could be provided, the usage of cloud-based services can be scaled up and down smoothly without the need for upfront cost. Finally, cloud technology helps break the barriers between different web applications.

Recent work in bioinformatics addresses the *flexibility* of the services provided at IaaS layer for the deployment of data intensive applications such as processing high throughput sequencing data or executing intensive analysis algorithms (Zhang, Gu, Liu, Wang, & Azuaje, 2012; Hung & Lin, 2013). These applications get as many machines as the analysis of data needs and automatically scale up and down the hardware resources based on dynamic workloads. Conversely, in distributed processing environments, it is difficult to distribute and co-ordinate a large-scale job on different machines, run processes on them and install additional machines to recover if one machine fails. Cloud technologies remove such technological concerns from the users.

A less explored level of cloud services is PaaS which focuses on building applications such that:

1. Draw the necessary resources on-demand (like compute servers or storage),
2. Perform their tasks, and
3. Relinquish the unneeded resources after a task is done.

Everything is automated and operates without any human intervention by taking advantage of simple APIs of Internet-accessible services that scale on-demand.

A relatively few number of bioinformatics applications exists which has been developed at PaaS layer, although PaaS based solutions should be an added value for the biological community. In particular PaaS applications are attractive for small bioinformatics organizations and research groups that are allowed to benefit from the most advanced technology available today at lower costs and lower risk. Indeed, the upfront investment is low, they don't need to invest in software / hardware, licenses and renewal costs are kept to a minimum and they only pay for what they use.

As such, PaaS based solutions can offer a powerful and flexible alternative for bioinformatics applications.

Two significant concerns must be addressed before moving data and bioinformatics applications to a cloud environment: the security of information and the bottlenecks of data transferring.

- **Information Security and Privacy:** The scalability and the easy access to cloud resources increases the risk of security of data that have security requirements out of the ordinary such as medical records, clinical data from human studies etc. The storage of such data on publicly accessible servers requires the definition of security rules to protect data security and privacy. For example, mechanisms such as encryption measures and role-based access control models should be deployed on the Cloud

respecting both the typicality of biological data and the requirements of scientists in sharing such a data. Differently from conventional web-based environments, the use of the Cloud for storing data and applications improves physical security as it enables the creation of separate virtual machines and firewalls for each independent application environment. Security professionals traditionally recommend partitioning a system as a means of protection. The ease of creating new virtual machines provides ways to improve the security of hosted data by partitioning resources into separate areas, conferring protections against attacks.

- **Bottlenecks of Data Transferring:** Cloud vendors store the backups of users' applications and data in multiple geographical locations. Although, cloud can provide extra processing resources during the peaks (within limits), transferring vast amounts of biological data to the cloud is a significant bottleneck in cloud computing (Grossman & White, 2012). Networking bandwidth limitation causes delays in data transfer and incurs high bandwidth costs from service providers. This cost is an important issue for biomedical institutions that require substantial data movement (on the order of terabytes and petabytes) on a regular basis. As a consequence, the use of cloud infrastructure currently does not make economic sense for applications that need to continuously export or import large volumes of data to and from the cloud.

APPLICATION SCENARIOS

Cloud computing has been applied successfully to deal with tasks that require substantial computational power. Some initiatives have already successfully incorporated the cloud to expedite data processing in next-generation sequencing,

comparative genomics and proteomics (Krampis et al., 2012; Sudha Sadasivam, & Baktavatchalam, 2010).

With the emergence of readily available cloud platforms, software developers have the opportunity to move one step closer towards the search and integration of information from the wide range of biomedical web resources.

In exploring these resources, biologists are not interested in the large amount of their content. Conversely, they distill a huge amount of data from multiple web sources in order to obtain key information that focuses on a single class of concepts (e.g., only gene products). Because biological information encompasses many domains of knowledge, the success of the user's depends on his ability in browsing information that is stored in several web sites that exhibit diverse organization, terminology, data formats. For example, consider a user who investigates about the role of thyroid hormones administration in the treatment of the heart failure. He has to translate his problem in a set of concepts that meet the taxonomy of terms defined by the curators of the web resource he navigates. Often, the concepts of interest don't belong to a single taxonomy. Often, the concepts of interest don't belong to a taxonomy exploited by a single web site. In this case, unearthing specialized information from different web resources can be complex, time consuming and daunting. Moreover, the user must learn and remember the navigation paths of each specific web site he visited. Finally, different biomedical web portals implement the same basic functionality and are often concerned with overlapping information.

The design of applications for effectively support the users in searching biomedical resources must consider three fundamental questions. First, how to integrate structured, semi-structured and unstructured available data with diverse and sparse schemas? Second, how to retrieve meaningful information in an easy and efficient way? Finally, how to implement a searching infrastructure which has to scale, hence change, to meet new requirements stemming from the growth of its searching domain?

The adoption of the Cloud computing makes it easy to exchanging and integrating data amongst multiple service providers via high quality, interoperable services offered in a "neutral" territory by the cloud infrastructure. As it happens for web engines, these services act as specialists which search for data exposed by biomedical organizations via public programmable APIs and return real-time information.

These circumstances can be addressed by implementing PaaS applications that broaden the spectrum of use of the cloud paradigm and encloses the adequate environment for building applications that harvest the full potential of this technology. Typically, this programming environment includes components such as programming languages, web servers, and databases that are not lightweight for most biologists or people with no or limited programming experience. A recent major development is Google App Engine (GAE), a PaaS environment that allows users to quickly develop and run any web application on Google's infrastructure while benefiting from Google services, such as search, e-mail, geographic maps and social networking (Google App Engine. Documentation). GAE applications scale as the traffic and the data storage vary. The use of the basic services is free and users pay only for increased level of services that are delivered over network connection. Within GAE, the data management relies on Datastore a NoSQL database which supports data storage service distribution. Unlike a traditional relational database, a NoSQL database supports operations (i.e. create, read, update, delete) to manage structured objects by means of an SQL-like language (Stonebraker, 2010). Typically, PaaS delivers a software environment to deploy and run applications where cloud resources scale automatically and developers do not need to know how many resources are required nor to assign resources manually in advance. As well, the PaaS environment eases the rapid development of new software components and the introduction of new functionality in existing PaaS applications. Indeed, the PaaS applications promote the event-driven integration of data from heterogeneous and

distributed resources by providing convenient and on-demand access to web resources via Web Application Programming Interfaces (API) without the necessity to know where the services are hosted or how they are delivered.

The development of PaaS applications involves substantial changes in the current styles of software development. Specifically, instead of being conceived as a set of procedural tasks, a PaaS application needs to be thought in terms of a framework which undertakes the responsibility of coordinating and organizing the search across different web resources by combining high level services offered by the cloud infrastructure and APIs that satisfy the application requirements.

As an example, BioCloud Search EnGene (BSE) is a PaaS application that facilitates searching and integrating biological information from public and large-scale genomic repositories (Dessi, Pascariello, Milia, & Pes 2014) using web services provided by the National Center for Biotechnology Information (NCBI). Using BSE, users find their information of interest by means of a simple “Google-like” query interface that accepts the standard name of genes as keywords. The design of BSE is centered on the concept of dataspace, a new paradigm for the integration of heterogeneous data coming from a diverse set of sources, without an a priori schema and regardless their format and location (Dong & Halevy, 2007). A dataspace consists of a set of resources, namely the participants, and a set of relationships between them (Argiolas, Atzori, Dessi, & Pes, 2012). A fundamental part of a dataspace is the catalogue that contains information about the URL of the participants. Using suitable services exposed by participants, BSE provides integrated views of data over web resources by following the pay-as-you-go principle: data from participants are accessed as metered resources without the need to store them locally. Conceived as a multi-level index, the catalogue stores identifiers and annotations about genes, acquires and combines information from participants. The BSE’s user interface supports special templates, namely contexts, that hide the complexity of data searching by means of accor-

dions i.e. a vertically stacked list of items. Each item represents a context and can be “expanded” or “stretched” in order to reveal additional contexts associated with it. Contexts are nested: the user who clicks on a gene’s identifier is redirected to a new accordion to explore new information about the searched gene.

FUTURE RESEARCH DIRECTIONS

Currently, the bioinformatics community mainly adopts the Cloud technology for storing and analyzing large sets of biological data. Hadoop (Taylor, 2010) is the most popular tool for organizing data-intensive analysis into cloud-based pipelines that are executed over multiple nodes. However, this approach remains difficult for biologists with limited programming skills. Therefore, future efforts should be devoted to building lightweight cloud environments that do not require extensive coding by keyboard and allow biologists to organize their work easily by drag and drop operations. In recent years, there has been much activity on the development of biomedical portals i.e. on specially designed web sites that bring information together from diverse sources in a uniform way. For example, BioPortal, is a popular web portal that provides access to a library of biomedical ontologies and RESTful Web Services that enable the incorporation of ontology content into software applications. Although the benefits of such portals are well documented, this explosion of biomedical resources generates impediments for researchers and clinicians when they try to discover the most appropriate information for the realization of their goals.

Perhaps, one of the major obstacles is the difference in the structure of portals: each portal has its own organization, terminology and data formats and the users have to expend an amount of cognitive effort in remembering navigation paths. Moreover, many biomedical portals implement the same basic functionality and are often concerned with overlapping information.

Because cloud technologies ease the access to diverse types of web resources via interoperable services, future research directions should face the problem of setting up a system of standards for data exchanges in order to facilitate the implementation of lightweight cloud applications that support data integration.

Over the past few years, WEB 2.0 technologies tried to enhance user interaction by promoting the use of flexible organizational systems (i.e. social networks, blogs and wikis) that have been introduced into current research practices. As well, recent technologies have evolved and new cloud environments are appearing that allow to reformulate the common investigation practices so that they can profit from a paradigm where domain experts (rather than programmers) organize the domain knowledge in a intuitive and collaborative way (Dessì, Milia, Pascariello, & Pes, 2014). From the medical or biological perspective, collaboration is attractive for many circumstances and consists well with the e-Science paradigm that envisions the scientific research carried out via internet (Bosin, Dessì, & Pes, 2011; Bauzer Medeiros & Katz, 2016). So, future research efforts should be directed to build distributed research environments that support autonomy and intelligent coordination of researchers in creating and managing shared data and knowledge.

In recent years the technological world has grown thanks to the incorporation of sensors, microcontrollers and transceivers for collecting and sharing real-world information. Internet of Things (IoT) is an information acquisition paradigm that defines such computer-based environments where each physical object is equipped with devices for interacting with each other and communicating with the users (Chaâri et al., 2016). By enabling continuous and ubiquitous access to information from any connected device over the Internet, IoT is leading to collect large amount of data from scientific experiments in life sciences and bioinformatics without an uniform way to share, process and understand context information. It is evident that IoT will only achieve its full potential

when all the devices will work and learn together without human interaction. Without dedicated functionalities and organizational data structures there is a risk of creating the illusion that scientific advances can be supported by simply making available a greater amount of data. In this respect, it is crucial that bioinformatics starts devoting more attention to cloud technology as a technical structure enabling the organization and the knowledge extraction from these new information sources.

CONCLUSION

Cloud environments contrast sharply with traditional centralized computing platforms that are not only costly, but risk to become increasingly inadequate to meet the requirements of modern biologists. In particular, PaaS level can offer a powerful alternative for building applications that can be re-designed to take advantage of cloud infrastructure. Finally, cloud technology is not only an alternative to the traditional computing environments but occupies an its own niche: it transparently provides the biologists with pragmatic solutions to some open research issues such as big data storing and analytics, knowledge and data sharing, extracting and integrating information from different web resources etc. The above solutions are increasingly being suitable for many biologists and cloud computing is becoming a significant technology for developing a large variety of services which, most important, are exerting revolutionary influences on the whole scientific community.

REFERENCES

Argiolas, M., Atzori, M., Dessì, N., & Pes, B. (2012). Dataspace Enhancing Decision Support Systems in Clouds. *International Journal of Web Portals*, 4(2), 35–55. doi:10.4018/jwp.2012040103

- Bauzer Medeiros, C., & Katz, D. S. (2016). eScience today and tomorrow—Part 2. *Future Generation Computer Systems*, 59(1), 93–94. doi:10.1016/j.future.2015.12.010
- Bosin, A., Dessì, N., & Pes, B. (2007). Capturing heuristics and intelligent methods for improving micro-array data classification. In *Proceeding of the 8th International Conference on Intelligent Data Engineering and Automated Learning (LNCS)* (Vol. 4881, pp. 790-799). Berlin, Germany: Springer.
- Bosin, A., Dessì, N., & Pes, B. (2011). Extending the SOA paradigm to e-Science environments. *Future Generation Computer Systems*, 27(1), 20–31. doi:10.1016/j.future.2010.07.003
- Calabrese, B., & Cannataro, M. (2016). Bioinformatics and microarray data analysis on the cloud. *Methods in Molecular Biology (Clifton, N.J.)*, 1375, 25–39. doi:10.1007/7651_2015_236 PMID:25863787
- Chaâri, R., Ellouze, F., Koubâa, A., Qureshi, B., Pereira, N., Youssef, H., & Tovar, E. (2016). Cyber-physical systems clouds: A survey. *Computer Networks*, 108, 260–278. doi:10.1016/j.comnet.2016.08.017
- Chae, H., Jung, I., Lee, H., Marru, S., Lee, S. W., & Kim, S. (2013). Bio and health informatics meets cloud: BioVLab as an example. *Health Information Science and Systems*, 1(1), 6. doi:10.1186/2047-2501-1-6 PMID:25825658
- Dessì, N., Milia, G., Pascariello, E., & Pes, B. (2016). COWB: A cloud-based framework supporting collaborative knowledge management within biomedical communities. *Future Generation Computer Systems*, 54(1), 399–408. doi:10.1016/j.future.2015.04.012
- Dessì, N., Pascariello, E., Milia, G., & Pes, B. (2014). BioCloud Search EnGene: Surfing Biological Data on the Cloud. In E. Formenti, R. Tagliaferri, & E. Wit (Eds.), *Computational Intelligence Methods for Bioinformatics and Biostatistics (LNCS)* (Vol. 8452, pp. 33–48). Berlin, Germany: Springer. doi:10.1007/978-3-319-09042-9_3
- Dessì, N., Pascariello, E., & Pes, B. (2014). Integrating Ontological Information About Genes. In *Proceedings of the IEEE 23rd International WETICE Conference* (pp. 417- 422). IEEE Press.
- Dong, X., & Halevy, A. Y. (2007). Indexing dataspaces. In *Proceedings of the 2007 ACM SIGMOD international conference on Management of data, SIGMOD'07* (pp. 43-54). New York: ACM. doi:10.1145/1247480.1247487
- Google App Engine Documentation. (n.d.). Retrieved July 26, 2016 from <https://cloud.google.com/appengine/docs>
- Grossman, R. L., & White, K. P. (2012). A vision for a biomedical cloud. *Journal of Internal Medicine*, 271(2), 122–130. doi:10.1111/j.1365-2796.2011.02491.x PMID:22142244
- Hung, C. L., & Lin, Y. L. (2013). Implementation of a parallel protein structure alignment service on cloud. *International Journal of Genomics*.
- Krampis, K., Booth, T., Chapman, B., Tiwari, B., Bicak, M., Field, D., & Nelson, K. E. (2012). Cloud BioLinux: Pre-configured and on-demand bioinformatics computing for the genomics community. *BMC Bioinformatics*, 13(1), 42. doi:10.1186/1471-2105-13-42 PMID:22429538
- Lin, G., Fu, D., Zhu, J., & Dasmalchi, G. (2009). Cloud Computing: IT as a Service. *IT Professional*, 11(2), 10–13. doi:10.1109/MITP.2009.22

Mrozek, D. (2015). Cloud-based computing architectures for solving hot issues in structural bioinformatics. In S. Bagchi (Ed.), *Emerging Research in Cloud Distributed Computing Systems* (pp. 294–314). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-8213-9.ch009

Schadt, E. E., Linderman, M. D., Sorenson, J., Lee, L., & Nolan, G. P. (2011). Cloud and heterogeneous computing solutions exist today for the emerging big data problems in biology. *Nature Reviews. Genetics*, 12(3), 224. doi:10.1038/nrg2857-c2 PMID:21301474

Stonebraker, M. (2010). SQL databases v. NoSQL databases. *Communications of the ACM*, 53(4), 10–11. doi:10.1145/1721654.1721659

Sudha Sadasivam, G., & Baktavatchalam, G. (2010). A novel approach to multiple sequence alignment using Hadoop data grids. *International Journal of Bioinformatics Research and Applications*, 6(5), 472–483. doi:10.1504/IJ-BRA.2010.037987 PMID:21224205

Taylor, R. C. (2010). An overview of the Hadoop/MapReduce/HBase framework and its current applications in bioinformatics. *BMC Bioinformatics*, 11(1).

Website of the National Center for Biomedical Ontology (NCBO). (n.d.). Retrieved July, 26, 2016 from <http://bioportal.bioontology.org/>

Website of the National Center for Biotechnology Information (NCBI). (n.d.). Retrieved July, 26, 2016 from <http://www.ncbi.nlm.nih.gov/>

Zhang, L., Gu, S., Liu, Y., Wang, B., & Azuaje, F. (2012). Gene set analysis in the cloud. *Bioinformatics (Oxford, England)*, 28(2), 294–295. doi:10.1093/bioinformatics/btr630 PMID:22084254

ADDITIONAL READING

Bagchi, S. (Ed.), *Emerging Research in Cloud Distributed Computing Systems*. Hershey, PA: IGI Global. doi:10.4018/978-1-4666-8213-9

Calabrese, B., & Cannataro, M. (2015). Cloud Computing in Healthcare and Biomedicine. *Scalable Computing. Practice and Experience*, 16(1), 1–18.

Goli-Malekabadi, Z., Sargolzaei-Javan, M., & Akbari, M. K. (2016). An effective model for store and retrieve big health data in cloud computing. *Computer Methods and Programs in Biomedicine*, 32, 75–82. doi:10.1016/j.cmpb.2016.04.016 PMID:27282229

Lee, K. K. Y., Tang, W.-C., & Choi, K.-S. (2013). Alternatives to relational database: Comparison of NoSQL and XML approaches for clinical data storage. *Computer Methods and Programs in Biomedicine*, 110(1), 99–109. doi:10.1016/j.cmpb.2012.10.018 PMID:23177219

ODriscoll, A., Daugelaite, J., & Sleator, R. D. (2013). Big data, Hadoop and cloud computing in genomics. *Journal of Biomedical Informatics*, 46(5), 774–781. doi:10.1016/j.jbi.2013.07.001 PMID:23872175

Sakr, S., Liu, A., Batista, D. M., & Alomari, M. (2011). A survey of large scale data management approaches in cloud environments. *IEEE Communications Surveys and Tutorials*, 13(3), 311–336. doi:10.1109/SURV.2011.032211.00087

Stein, L. D. (2010). The case for cloud computing in genome informatics. *Genome Biology*, 11(5), 207. doi:10.1186/gb-2010-11-5-207 PMID:20441614

KEY TERMS AND DEFINITIONS

Application Programming Interface (API):

An API specifies how software components interact in terms of its operations, inputs, outputs, and underlying types in a way that is independent of their respective implementations. APIs allows software developers to build programs as a set of building blocks whose functionality is provided by their corresponding APIs.

Biomedical Ontology: A structured representation of the knowledge by means of formal naming and definition of the types, properties, and interrelationships of the entities that really exist for biological research.

Computer Platform: The computer hardware and the operating system that conforms a set of standards enabling software developers to deploy software applications for the platform.

NoSQL Database: Modern databases conceived for web applications. Unlike a relational database, a NoSQL database does not store data and relationships in tables. Conversely, NoSQL databases are schema-free, distributed and horizontally scalable to clusters of machines i.e. the database is partitioned in a cluster of distributed database servers (each maintaining its own data

and a self-contained schema) and it makes easy to add or remove, namely to scale, a single database server.

RESTful Services: In a distributed computing environment, the Representational State Transfer (REST) architecture allows clients and servers to interoperate by using a standardized interface and protocol using a uniform set of simple and well-defined operations. Resources are manipulated using a fixed set of four create, read, update, delete operations: RESTful web service exposes a set of resources identified by URIs and manipulated using a fixed set of four operations (i.e. create, read, update, delete).

Web Service: A Web Service allows two resources to communicating each other over a network Internet for transferring machine readable file formats such as XML and JSON in a manner prescribed by its interface.

Web 2.0: A recent new vision of the Web that enhances collaboration and interaction of users with each other. Unlike traditional Web sites where users had a passive role in viewing information, Web 2.0 environments (such as social networks sites, blogs, wikis etc.) are grounded on user-generated content and foster people to social media dialogue.

Category B

Biology

GWAS as the Detective to Find Genetic Contribution in Diseases

Simanti Bhattacharya

University of Kalyani, India

Amit Das

University of Kalyani, India

INTRODUCTION

In the era of unprecedented advancement in medical and technological sciences, Garrod A.E., a physician to the Hospital of Sick Children, in the year of 1902, reported a case of alkaptonuria that he described as “not the manifestation of disease but is rather of the nature of an alternative course of metabolism...”. That was the first report where the possibility of underlying “molecular evidence” behind human disease came in the lime light. Human diseases and their genetic contribution share a complex and intricate relationship, yet to be explored fully. For few cases, phenotypes (diseased condition) could directly be associated with gene, experimentally, whereas a large number of genetic associations behind disease state remained hidden.

That called for a situation where mapping of gene knowledge between afflicted and un-afflicted individuals can be mapped to find out the difference at each point (each nucleotide position or *allele* or *variant*) and the number of occurrences of the mismatched alleles in the diseased individuals (*allelic frequency*) with an assumption that if any allele has higher frequency to appear in the diseased individual then that is associated with the diseased trait. That can be translated as ‘scan through entire gene’ for gene to disease relation mapping or rather, as genetic language, *Genome Wide Association Study (GWA or GWAS)*.

These GWAS data not only provide us with the information on the disease association with

the gene level knowledge but also enable a deeper understanding of the entire scenario generating a landscape of gene with its minute changes that can be extrapolated to genes coding (impact on protein production) or non-coding regions (impact on protein production regulation), the transcription factor binding sites (regulating transcription), epigenetic modification probabilities (regulation in genetic coding), pathways involved (visualizing the upstream or downstream possible effects) extending to heritability of the diseases. These all impose final impact on the phenotype which is nothing but the diseased state to us. Thus, insight generated with GWAS leads us to understand the actual reason or mechanism behind disease onset that, in turn, guides scientists to find novel drug-gable targets for more efficient medications or some times, to look for personalized medications (Bush & Moore, 2012).

With the completion of Human Genome Project (human DNA sequence) in 2003 (International HapMap Consortium, 2003) and the International Hapmap project (haplotype map of the human genome) in 2005, scientists are well-equipped with resources to correlate genetic contribution to disease onset. Success of the GWAS reflects in identifying the genetic factor contributing to Parkinson’s disease, Crohn’s disease, type 2 diabetes and obesity to name a few. These GWAS data can also be accessed through various repositories. However, its smaller variant size, unavailability of replicated reports, smaller population size under study stand as limiting factors to uncover a larger

portion of genetic information to understand properly. With the rapid advancement in research, these limitations will be overcome to generate a better understanding of the entire scenario of disease with GWAS concept.

BACKGROUND

Life material nucleus i.e. DNA is composed of four basic entities and they are: the purines: Adenine (A), Guanine (G) and the pyrimidines: Cytosine and Thymine (T) arranged in a specific pattern that carries information of life. DNA is double helical element where these purines and pyrimidines pair up with each other (A with T and C with G). The complex of different DNA stabilizing and regulatory molecules (protein, RNA etc) along with DNA is collectively known as Genome. Since not only the presence of these basic entities but also their relative position is crucial to maintain the information for life and its sustainability, small variation of these elements (mutation or polymorphism or SNPs) at any of the otherwise conserved position on DNA has been found to bring change in the system. In Pre-GWAS (Ertekin-Taner, 2010) era researchers managed to characterize the genetic association behind disease that followed a Mendelian pattern of inheritance that largely depended on generation wise co-segregation of causal variants with marker alleles that simply followed Hardy-Weinberg Equilibrium. Cause to genetic variation was thought to be chromosomal cross over at chiasmata during meiosis. To explain, suppose, a family or rather a pedigree of a population was studied for a disease that was found to be occurring due to certain enzyme mutation. This, at that time, was considered to be the candidate gene and was followed by characterization, identification of mutated allele as well as developing therapeutics against it. But there could be number of possible reason behind genetic variation within or among the population, random mutations being the ultimate source of genetic variations. Along with that not only inheritance but also other ef-

fects including environment, living style, age etc contribute significantly to genetic changes. And these also can contribute to disease state if the changed allele gets a higher frequency in diseased population than in healthy population. But during GWAS era scientists actually received the scanned report of entire genome that help understanding the complex relation of DNA level mutation and disease phenotypes. As for human cases, each disease may require looking at hundreds or thousands of positions to identify SNPs and associated genes that may contribute to risk of developing a certain disease. Increasing evidences show that GWAS represents a really powerful technique to identify these marker alleles at relatively higher speed and precision. GWAS is based on linkage disequilibrium (LD) principle where loci that are physically close together show stronger LD than the loci that are far apart (Visscher, Brown, McCarthy, & Yang, 2012). The strength of LD for an effective population size also decides the number of genetic marker required to specify the haplotype. GWAS also hypothesizes the 'allelic frequency' of a rare variant will be in low LD with nearby common variant. As time advanced, GWAS started exploring different type of variations and linkage to disease for example rare variant identification through evolutionary model, copy number variation (CNV) that include deletion and/or duplication of DNA segments of diverse size and frequency.

WORKFLOW FOR GWAS METHODS

For any experiment to be successful, a well-structured experimental design is essential. GWAS is not exceptional in this case. It systematically follows basic steps (Figure 1), given below:

- **Sample Collection:** A large cohort collection (>1000) for case and control is essential for initial set up. Chip based microarray technology that assays 1 million or more SNPs from population under study

Table 1. Overview of technologies (including their respective advantages and disadvantages) used for GWAS

Sl. No.	Technology	Detectable No. of Markers	Advantages	Disadvantages
1	Whole Genome Sequencing	Around 35million or EXOME sequencing; around 3 billion for whole genome sequencing	Individual sampling data, better comprehensive information including non-coding (even intronic) SNPs	Selective applications, high cost and analytically challenging
2	Hap-Map genotyping	Hundreds of Thousands	High quality, cost effective, strong statistical association with diseases/phenotypes	Chance of missing uncommon variants
3	RNASeq	Varies with samples and study designs	Detection RNA specific SNPs (both coding and non-coding), variations across expression levels can also be co-related	Dominance of common RNAs.
4	1000 genomes based platforms	Millions	Systemic assessment of all variants including uncommon ones	Require multiple testing, accuracy of genotypic and sensitivity of statistical tests are often compromised to some extent

are the crude source for GWAS sample size for measuring SNP variation. The technology, briefly, explores the hybridization property of DNA with the collection of microscopic DNA spot attached to solid surface. It is also important to know the sample in details. Primarily there are two types of phenotypes: case/control pair (dichotomous; where the genetic variation has to be deciphered by comparing, like schizophrenia related SNPs) and quantitative traits (where the phenotype is measurable like blood group). Based on the sample type the source platform is selected and derived data is further taken for analysis. In this regard, quality control step for the data is very crucial to start with a good set of samples. GWAS files contain vast amount of data and thus are prone to error. Again different methods are adopted while generating the data by different cohorts which increase noise to signal ratio. To improve the quality of data and compatibility among cohorts it is necessary to have quality check for the sample that removes false positive cases(Psychiatric GWAS Consortium Coordinating Committee et

al., 2009). Table 1 provides as short description of different technologies used for collecting data for GWAS. Advantages and limitations of each technique are also discussed.

- **Find Deviating SNP:** From expecting haplotypes visualize SNP-SNP interaction with HapMap: Chip derived hybridized sample DNA then is extracted and the haploid is taken for sequencing. The result of sequencing experiment is then mapped to the known or control DNA sequence. Thereby the mismatched portion, considered as variant, is obtained. The SNP to SNP interaction is further identified with HapMap result. Other than HapMap, global view of all SNPs signals can also be represented via Manhattan plots. Advantage of this plot is that the strengths and location of each SNP can be directly derived (Figure2). However unlike HapMap, correlation between two or more SNPs and associated phenotype cannot be derived from such plot.
- **Detection of Potential Association Signals:** Once the deviating SNPs are gathered, statistical analysis is carried out

to detect the association of the variant to the disease phenotype and majorly done by examining single SNP for its association, done with a series of single locus statistics test. Generalized linear model (GLM) is generally approached to test for qualitative traits and GLM is linear regression model that broadly applies Analysis of Variance (ANOVA)(Hong & Park, 2012). The assumptions are: 1) normal distribution of variance, 2) the trait variances within each group are same and 3) the groups are independent. On the contrary, case/control traits are analyzed with contingency table or logistic regression followed by chi-squared test or related Fisher's exact test. The statistical power derived from these will direct the strength of true association in each case. This is further then stratified based on allelic association or Genotype association followed by further grouping into dominant, recessive, additive or multiplicative type.

- **Fine Mapping:** This step requires all variants in the associated region to have been imputed in genotype or associated version through quality check with stringent conditions and large sample size that is sufficient to resolve among highly correlated signals. The methods employed here broadly calculate posterior probability that tells whether the variant in that locus is causal. Suppose an allele A in the SNP is found to be changed to G through GWAS experiment and found to have high LD with nearby SNP. Now it also requires the SNP location (whether intergenic, within the gene or at nearby location of genes), gene information, allelic change position, consequences etc to know in details about the SNP.
- **Replication of Detected Association in New Cohort or Subset for Validation Purpose:** Once an associated and finely mapped variant with correlated data comes

up, the study is replicated through the new cohort keeping the condition same or through the subset of the sample taken. This step essentially examines the robustness of the experiment design and its outcome.

- **Biological or Clinical Validation:** However, statistical power calculations or significance analyses will only predict the effect of variant (whether risk or protective). To make the findings be free from doubts these findings are validated with biological experiments or clinical validations.
- **Downstream Analysis and Functional Annotations:** The association is further expanded with meta analysis to understand the integrated knowledge of SNP-disease association and consequences (Evangelou & Ioannidis, 2013) for instances, quantitative trait loci analysis, gene ontology or pathway enrichment etc.

IMPORTANCE OF GWAS

In the pre-GWAS era, studies required to go from phenotype (i.e. understand the disease/trait first) to biological processes (with which the disease/trait is mostly associate) so that the causing agents (mostly the genes in the pathway) could be identified. Further molecular dissection, mutagenesis studies and other analyses were carried out to characterize the identified gene and the marker. If the assumption would have been 'one gene one disease' then it would have one shot identification process but could not be held true for complex gene-phenotype relationships. GWAS has this selective advantage for which prior knowledge on biological pathway of the trait does not have to be known as it approaches SNP level to gene to phenotype direction. Since GWAS does an in depth study at allelic level often it comes up with novel candidate genes. The independent statistical analysis and power driven association result often

Figure 1. Overview of GWAS experiment design. This is a schematic representation of GWAS data collection and further downstream analyses. The initial sample cohort should be at equal state ideally. Test set may contain heterozygous allele (shown in yellow). SNP are collected and deviated markers are statistically validated and functionally mapped.

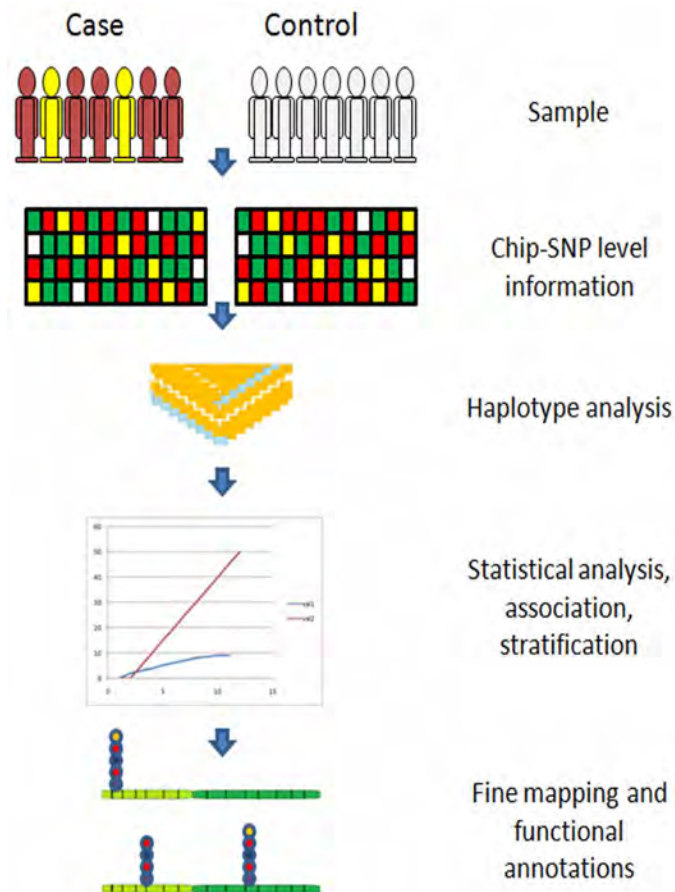
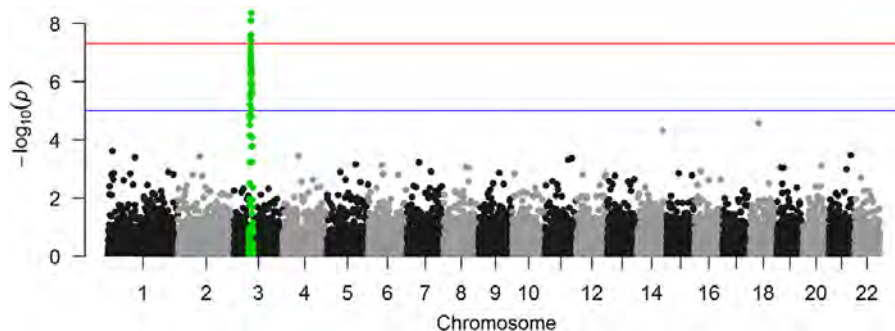


Figure 2. Sample of Manhattan plot based visualization of GWAS data. In a Manhattan plot, x-axis denotes to location of a SNP while the y-axis denotes the strengths of association ($-\log_{10}$ of pvalue). Highly probable SNPs often have strong values associated in the y-axis (marked in green). One major limitation of such visualization is that such plots do not provide any correlations among SNPs and phenotype of interest. The plot is created with the R qqman package.

Source: Turner, 2014



eliminates the chance of biased specific association of SNP to phenotype. Its efficacy in finding SNP and CNV provides more robust data as well as gives information on heritability or ancestry of each subject. Furthermore, GWAS generated huge data to analysis and not only association or fine mapping the post GWAS requires a string meta analysis where more information on SNP/gene correlated data like amino acid change, protein motif change, alteration in biological processes/pathways etc. To access these information effective collaboration are formed with strong consortium of expertise and finally an integrated answer to disease mechanism is solved.

GWAS FOR NEW TARGET IDENTIFICATION IN DISEASES

With GWAS personalized medicine (Thrall J.H., 2004) has achieved a new edge. Personalized medicine is referred to tailoring of medical treatment to the individual characteristics of each patient. In pre-GWAS era the most acclaimed belief was for the association of 'one gene one disease' and identifying the causal gene required 'trial-and-error' kind of studies. But with the advancement in understanding disease complexity a new insight has been revealed showing multiple genetic association and SNP variation behind disease onset. Even within a population under study, there exists SNP level differentiation in closest relatives. This triggers the importance of understanding the disease prospective for each and every individual taken as case and identifying the causal agent behind the disease in order to develop personalized medicine. Personalized medicine relies on traditional as well as newly developed concepts of genetic association with disease and environmental factors regulating it in order to individualize prevention, diagnosis and treatment. GWAS provides insight to this modern age understanding of multiple gene to disease association by performing entire genome wide scanning at each point, thereby scrapping

variant information for individuals (more often called 'personalized genome') along with cohort sample. A classic example is study of medicine trastuzumab which is prescribed for only 30% of the breast cancer patient who carry a rare mutation that causes over-expression of protein HER2.

Not only to find out rare events for personalized medicine, GWAS is an intricate study in different fields of diseases like cancer, autoimmune diseases, central nervous system related disorder and many more to name. GWAS have been implemented in five of the commonest cancer types: breast, prostate, colorectal, lung and melanoma and have successfully identified 20 novel disease loci (Easton & Eeles, 2008). With GWAS the genetic architectures of multiple sclerosis (Bashinskaya, Kulakova, Boyko, Favorov, & Favorova, 2015), schizophrenia, bipolar disorder etc have been investigated and GWAS has successfully identified disease susceptibility, clinical phenotypes and treatment responses. Autoimmune diseases like celiac disease, crohn's disease, rheumatoid arthritis etc occur when own immune cells of an individual attack and destroy body's cells. With GWAS analysis, the number of genetic risk loci associated with different autoimmune diseases has remarkably increased. Similarly GWAS has encroached different disease section and have identified more associated variant for the given disease/trait than was known before.

GWAS AND ITS SPECIFICITY

First GWAS Discovery

The first report that included prominent GWAS involvement came into light in 2005 when Klein R.J. *et al.* were studying Age Related Macular Degeneration (AMD), a major cause of blindness in elderly. They performed genome wide screening of 96 cases and 50 controls for identifying polymorphisms associated with AMD. Their statistical analysis showed, among 116,204 SNPs that were screened one variant in the complement factor H

(*CFH*) gene showed strangest correlation (p value $<10^{-7}$) with the diseased state. Further resequencing revealed a variant in linkage disequilibrium with the risk allele representing a tyrosine to histidine change at amino acid 402 which is interestingly at the binding site for C-reactive protein (CRP) and heparin. Moreover, *CFH* is located in a region of chromosome 1 that is repeatedly linked to AMD in family based studies. Previous studies on the conservation of tyrosine at 402 for the functionality of *CFH* were reported along with the report for elevated serum CRP during AMD. These previous reports and Klein group's finding established a mechanism of AMD onset through GWAS (Klein, Zeiss, Chew, & Tsai, 2005).

One Article Retraction Due to Statistical Anomaly

Statistical analysis plays a crucial and strong role in GWAS data analysis. Anomaly in this aspect can cause severe case of false positive data as inference. Such a case happened when Sebastiani et al. presented their work on identifying sets of markers responsible for aging (Sebastiani et al., 2011). Although they collected samples from more than 1000 centenarians, they employed a statistical analysis that was different from the canonical process and led to controversy in scientist community regarding data discrepancies. Ultimately they had to retract their article, one year after its publication.

LIMITATIONS

GWAS is critically dependent on the samples and control taken for studies. Thus GWAS involves different limitations associated with proper quality control and study setup. Common problems that usually occur during GWAS involve case-control groups with less defined demarcations, smaller and less diversified population under study, erroneous sample stratification and proper control size that is comparable to test set for multiple testing.

GWAS depends on the statistical relevance that an SNP is linked with disease/trait for a population compared to control population. An error in sample collection, study design and proper quality control will definitely fall into results with faulty statistical report. This will then give rise to a situation of false positive or false negative data and thereby over or under representation of SNP-gene relation. Furthermore, most of the time scientists falsely assume that GWAS SNP gives the hint for the causation of the associated disease but the real scenario is GWAS states the possibility of correlation between an SNP and disease. A sample size with proper statistical power is essential for success of GWA study. Sometimes improper statistical analysis may give false negative result. However, the above mentioned issues can be taken care of and can be corrected. But there are certain fundamental problems associated with GWAS like the common belief of heritability of SNPs for specific disease/traits. To avoid such cases the experiment needs proper information of sample as well as linkage analysis with relevant statistical measures. GWAS detects only those variants that are common ($>5\%$) in a population. Validation of results requires replication in independent samples from different population to avoid any falsified data. Furthermore stringent statistical power calculation may give rise to over or under representation of association. Most of the time the SNPs do not fall in the protein coding region or at least at the nearby region of protein coding gene with known functionality. These sets of SNPs are often under analyzed though they may have significant effect on phenotype.

FUTURE DIRECTIONS

The pre-GWAS era thought of one gene - one disease but with the intervention of GWAS, scientific community has been forced to reprocess their thoughts towards complexities of SNP to disease relation. This has huge impact on human genetics. This view has been further strengthened

due to the parallel increase in Whole Genome Sequencing data. With the rapid progress of associated technologies, new and cheap sequencing techniques will continue to evolve further and that will generate huge sample set for careful and thorough fine-mapping. A major concern regarding GWAS is complex dependence-structure for SNPs as well as for individual sample. Mixed model approaches are now being adopted that provide a general and flexible attempt to carry out GWAS of correlated phenotypes.

In the near future, following areas of research or healthcare could benefit significantly from the technical and translational development of GWAS or its associated techniques:

- **Personalized Medicine:** This is going to be area where GWAS will probably have maximum impact. Armed with only gene sequence information, GWAS plots (hap-map and Manhattan plots) provide some sort of systemic view of a person's overall genetic architecture. This type of information will help researchers and physicians in finding effective medicines for each and every patient. Drug like trastuzumab already has already shown the way.
- **Phenotype Refinement or Re-Classifications:** What gene signatures differentiate Crohn's disease from ulcerative colitis or psoriasis from lupus? While there are several genes and signaling pathways that are significantly overlapping across such diseases, GWAS has the potential to provide a clear set of gene signatures that would help in proper classification of phenotypes (re-classification in some cases).
- **Understanding the Impact of SNPs in the Non-Coding DNA:** Majority of human DNA is non-coding. However GWAS has found several SNPs which are strongly associated with several diseases are located in non-coding region. Understanding the impact of these non-coding SNPs and finding ways to translate that information to

advancement of healthcare could be major focus of GWAS studies.

- **Predictive Analysis Based on Genome Profiling:** Similar to personalized medicine, GWAS helps in identifying potential disease threats for a particular individual. The advantages of hap-map and Manhattan plots (Figure 2) are that they help in weighing the potential impact of known disease causing mutations by directly comparing each person's co-occurring mutations with that of a population data.
- **Policy-Making:** This could be a little distant reality but definitely a very important one in the long run. Similar to many scientific discoveries, GWAS information can also be used for un-explained exploitation of potential threats. While a person may be predicted to be susceptible to certain diseases based on the genetic makeup but there also lies a chance that the person might never develop those diseases at all. To stop healthcare providers from unnecessarily punishing a person due to predicted threats, proper guidelines have to be brought in place.

WEBLINKS

There are several tools/web-servers and statistical packages available for GWAS analysis and further fine mapping. For each of the type of experimental design and analysis selected set of examples are given in Table 2.

CONCLUSION

Today GWAS, along with HapMap project data, other sequencing data, various statistical analysis and further annotations has become a powerful tool to analyze variant association with diseases/traits with the extension in gaining deeper knowledge in disease mechanism for producing personal-

Table 2. List of Web-tools used in GWAS

Name	Web Link	Type	Remark
Database for Genotypes and Phenotypes (dbGaP)	http://www.ncbi.nlm.nih.gov/gap/	R package	For GWAS data
Disease-ANCEstry Networks (DANCE)	http://gilderlanio.pythonanywhere.com/	Web tool	For GWAS data/ visualization
Bayesian hierarchical generalized linear model (BhGLM)	http://www.ssg.uab.edu/bhglm/	R package	For Haplotype association
Diplotyper	https://code.google.com/archive/p/diplotyper/	Automated tool	For Haplotype association
EasyQC	http://www.uni-regensburg.de/medizin/epidemiologie-praeventivmedizin/genetische-epidemiologie/software/index.html	R package	For Quality Control
GWASTools	https://bioconductor.org/packages/release/bioc/html/GWASTools.html	R package	For Quality Control
DIST	http://dleeelab.github.io/dist/	Software (C++)	For Summary statistics
SNPStats	http://bioinfo.iconcologia.net/SNPstats/	Software	For statistical analysis
Sherlock	http://sherlock.ucsf.edu/	Web tool	For statistical analysis
PLINK	http://pngu.mgh.harvard.edu/~purcell/plink/	Open source tool set	For statistical analysis
Comprehensive Meta Analysis (CMA)	https://www.meta-analysis.com/index.php	Paid software	For GWAS Meta-analysis, takes excel input
Meta Analysis Helper (Metal)	http://csg.sph.umich.edu/abecasis/metal/	Free Software	For GWAS Meta-analysis
MetaSKAT	https://www.hsph.harvard.edu/skat/metaskat/	R package	For GWAS Meta-analysis
FunciSNP	http://www.bioconductor.org/packages/release/bioc/html/FunciSNP.html	R package	For SNP prioritization

ized medicines. A careful selection of samples and clever experimental design would make it a powerful tool to identify new targets for better understanding of complex diseases. Its efficacy in pedigree wise and expression dependent analysis of genetic association with diseased phenotype has made it an essential tool for decoding developmental abnormalities across several generations of a population. However, GWAS gives least information for SNPs situated in the non-coding portion (intergenic, distant from known protein coding gene region) of DNA and thereby GWAS loses a large amount of data for downstream analysis. With the advancement in science in near future this problem will definitely get its solution and GWAS will be fully explored to get answers for disease complexity.

REFERENCES

- Bashinskaya, V. V., Kulakova, O. G., Boyko, A. N., Favorov, A. V., & Favorova, O. O. (2015). A review of genome-wide association studies for multiple sclerosis: Classical and hypothesis-driven approaches. *Human Genetics*, 134(11-12), 1143–1162. doi:10.1007/s00439-015-1601-2 PMID:26407970
- Bush, W. S., & Moore, J. H. (2012). Chapter 11: Genome-Wide Association Studies. *PLoS Computational Biology*, 8(12), e1002822. doi:10.1371/journal.pcbi.1002822 PMID:23300413
- Easton, D. F., & Eccles, R. A. (2008). Genome-wide association studies in cancer. *Human Molecular Genetics*, 17(R2), R109–R115. doi:10.1093/hmg/ddn287 PMID:18852198

Ertekin-Taner, N. (2010). Genetics of Alzheimer disease in the pre- and post-GWAS era. *Alzheimers Research & Therapy*, 2(1), 3. doi:10.1186/alzrt26 PMID:20236449

Evangelou, E., & Ioannidis, J. P. A. (2013). Meta-analysis methods for genome-wide association studies and beyond. *Nature Reviews. Genetics*, 14(6), 379–389. doi:10.1038/nrg3472 PMID:23657481

Hong, E. P., & Park, J. W. (2012). Sample size and statistical power calculation in genetic association studies. *Genomics & Informatics*, 10(2), 117–122. doi:10.5808/GI.2012.10.2.117 PMID:23105939

International HapMap Consortium. (2003). The International HapMap Project. *Nature*, 426(6968), 789–796. doi:10.1038/nature02168 PMID:14685227

Klein, R., Zeiss, C., Chew, E., & Tsai, J. (2005). Complement factor H polymorphism in age-related macular degeneration. *Science*, 308(5720), 385–389. doi:10.1126/science.1109557 PMID:15761122

Psychiatric GWAS Consortium Coordinating Committee. (2009). Genomewide association studies: History, rationale, and prospects for psychiatric disorders. *The American Journal of Psychiatry*, 166(5), 540–556. doi:10.1176/appi.ajp.2008.08091354 PMID:19339359

Sebastiani, P., Solovieff, N., Puca, A., Hartley, S. W., Melista, E., Andersen, S., & Perls, T. T. et al. (2011). Retraction. *Science*, 333(6041), 404. doi:10.1126/science.333.6041.404-a PMID:21778381

Thrall, J. H.J.H. (2004). Personalized medicine. *Radiology*, 231(3), 613–616. doi:10.1148/radiol.2313040323 PMID:15163802

Turner, S. D. (2014). qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *bioRxiv*.

Visscher, P. M., Brown, M. A., McCarthy, M. I., & Yang, J. (2012). Five years of GWAS discovery. *American Journal of Human Genetics*, 90(1), 7–24. doi:10.1016/j.ajhg.2011.11.029 PMID:22243964

ADDITIONAL READING

Alexopoulos, E. C. (2010). Introduction to multivariate regression analysis. *Hippokratia*, 14(Suppl 1), 23–28. PMID:21487487

Cantor, R. M., Lange, K., & Sinsheimer, J. S. (2010). Prioritizing GWAS Results: A Review of Statistical Methods and Recommendations for Their Application. *American Journal of Human Genetics*, 86(1), 6–22. doi:10.1016/j.ajhg.2009.11.017 PMID:20074509

Cho, J. H. (2010). Genome-wide association studies: Present status and future directions. *Gastroenterology*, 138(5), 1668–1672.e1. doi:10.1053/j.gastro.2010.03.028 PMID:20302866

Hayes, B. (2013). Overview of Statistical Methods for Genome-Wide Association Studies (GWAS). *Methods in Molecular Biology (Clifton, N.J.)*, 1019, 149–169. doi:10.1007/978-1-62703-447-0_6 PMID:23756890

Schneider, A., Hommel, G., & Blettner, M. (2010). Linear regression analysis: Part 14 of a series on evaluation of scientific publications. *Deutsches Ärzteblatt International*, 107(44), 776–782. doi:10.3238/arztebl.2010.0776 PMID:21116397

KEY TERMS AND DEFINITIONS

Chromosome: A chromosome is a packaged and organized structure containing most of the DNA of a living organism.

Gene: Gene has different definitions depending on the way it is studied in biology. As per classical (Mendelian) genetics a gene is the basic physical and functional unit of heredity. According to molecular biology a gene is a locus (or region) of DNA which is made up of nucleotides and is the molecular unit of heredity.

Genome: Genome refers to the complete set of genes or genetic material present in a cell or organism.

Linkage Disequilibrium: Linkage disequilibrium is the non-random association of alleles at different loci in chromosome.

Locus: A locus (plural loci), in genetics, is the specific location or position of a gene's DNA sequence, on a chromosome.

Nucleus: A nucleus is a membrane bound cellular organelle that contains genetic material in eukaryotes.

Regression Analysis: A regression analysis is a statistical process for estimating the relationships among variables.

SNPs: Single nucleotide polymorphism (SNP) s are single base pair changes in the DNA sequence that occurs with high frequency in human genome.

RNA Interference Therapeutics and Human Diseases

B**Dolly Sharma***PEC University of Technology, India***Shailendra Singh***PEC University of Technology, India***Trilok Chand***PEC University of Technology, India*

INTRODUCTION

Humans have more than 20 thousand genes that are protein coding. Proteins are essential building blocks of our body as they perform a variety of functions in human body (Pines, 2007). They produce hormones, enzymes and hemoglobin (Hoffman & Falvo, 2004). As proteins are important, their accurate production is also essential. Diseases are caused either by formation of defective protein or by over production of protein or less production of protein (Shenoy, & Jayaram, 2010; Selkoe, 2003). Diseases like Alzheimer's and Parkinson's disease, Huntington's disease, Creutzfeldt–Jakob disease, cystic fibrosis, Gaucher's disease, Uremia as well as other neurodegenerative disorders, heart disease, some cancers, stroke (Chaudhuri & Paul, 2006) are caused by inappropriate synthesis of protein. Proteins in association with RNA form Ribonucleoprotein (RNP) which relates to diseases like influenza A virus, inherited motor neuron disease spinal muscular atrophy (SMA) (Pellizzoni, 2007), premature aging, cancer (Blasco, 2005), aplastic anemia (Yamaguchi, Calado, Ly, Kajigaya, Baerlocher & Chanock, 2005), Alzheimer (Diner, Hales, Rabenold, Bishof & Duong, 2004) etc. In short, RNA and protein is linked with an outsized figure of diseases.

As gene expression (Alberts et al., 2007) is the process of formation of proteins, wherein DNA

acts as a template for mRNA which further gets converted into proteins. So, altering the gene expression process may prevent from some disease to occur. More precisely, it is altering gene coding at mRNA level. This is the concept adapted by RNA Interference Therapeutics. RNAi Therapeutics (Aagaard & Rossi, 2007) is the branch of science that focuses on control of gene activity at RNA level to cure diseases.

This chapter is organized as follows: RNA, DNA and gene expression are discussed in Introduction and Background section followed by RNA types, RNA interference process and RNA Therapeutics. Table 1 mentions some of the RNA Therapeutics companies. Further RNAi Therapeutics techniques PMO, LUNAR, RNA upregulation and Gene Silencing are discussed. It is concluded by RNAi Therapeutics' challenges and issues and lastly its future research directions.

BACKGROUND

Cells are the building blocks of life. Inside nucleus of a cell, DNA and RNA molecules reside. Ribonucleic Acid (RNA) molecule is a single stranded sequence consisting of four bases Adenine (A), Guanine (G), Cytosine (C) and Uracil (U). This single stranded sequence is known as the primary structure of RNA (Tinoco & Bustamante, 1999).

The backbone of RNA contains ribose as sugar. Phosphate groups are also attached to ribose. RNA can be synthesized from DNA. DNA is another molecule which is double stranded molecule. DNA consists of nucleotides that contain a phosphate group, a sugar group and a nitrogen base. Nitrogen bases are Adenine (A), Guanine (G), Cytosine (C) and Thymine (T).

DNA molecule contains a number of genes. Some genes code for protein and others do not. The process of production of protein is called gene expression. Precisely, Gene Expression process involves two sub processes- Transcription and Translation. Gene expression process is initiated by making a copy of DNA to produce a complementary strand called messenger RNA (mRNA) (Anfinsen, 1972) as shown in Figure 1. mRNA is one of the types of RNA. Thymine of DNA is replaced by Uracil of RNA. This step is called Transcription. Proteins are made up of sequence of amino acids. 20 amino acids exist. The bases of mRNA in sets of three, code for a specific amino acid and forms a codon as shown in Figure 1. As the number of bases is four, so 64 ($4 \times 4 \times 4$) possible codons are there. Consequently, more than one codon may code for a specific amino acid. The amino acid is meshed together with codon by another type of RNA which is transfer RNA (tRNA). tRNA contains a complementary sequence of bases on the other side called anti codon as shown in Figure 1. A sequence of amino acids is called a protein. This process is called Translation. Together these processes transcription and translation are termed as gene expression. Each protein consists of a unique sequence of amino acids. Ribosomal RNA (rRNA) links the amino acids so that specific protein is formed.

TYPES OF RNA

Gene Expression is catalyzed by majorly three types of RNAs, namely, messenger RNA, transfer RNA and ribosomal RNA.

- **Messenger RNA (mRNA):** mRNA is responsible for transfer of genetic information from DNA to the ribosome and is the biggest family of RNA molecules. The structure of mRNA is shown in Figure 2.
- **Transfer RNA (tRNA):** Suppose we need to bake a cake. We need a kitchen, a recipe and raw materials. In the gene expression process, kitchen correlates to the cell, recipe is with mRNA and raw materials are supplied by tRNA. It brings together RNA nucleotides and amino acids.
- **Ribosomal RNA (rRNA):** rRNA links the amino acids into specific sequence for translation into particular protein.

Other types of RNA are non-coding RNA, transfer-messenger RNA, small-nuclear RNA, micro RNA, small interfering RNA, small nucleolar RNA, antisense RNA and Signal recognition particle RNA. The details of these types are out of the scope of this chapter. Additional readings section can be referred for the same.

RNA INTERFERENCE

Decades ago, an experiment was conducted to increase the color of petunias flower wherein color producing gene was injected to the flower. Instead of increase in color, it was observed that color was lightened and even white petunias were produced in generations of flower (Jorgensen, Cluster, English, Que & Napoli, 1996). This was how the concept of gene expression suppression was coined. Further experiments revealed that introduction of sense and antisense RNA together forming double stranded RNA (dsRNA) leads to gene silencing (Fire et al., 1998).

The mechanism of RNA Interference (RNAi) involves treating the disease causing gene with specific small interfering RNA (siRNA) that targets the specific messenger RNA (mRNA). This mRNA is involved in production of protein that leads to the disease in question. The motive

Figure 1. Gene expression process

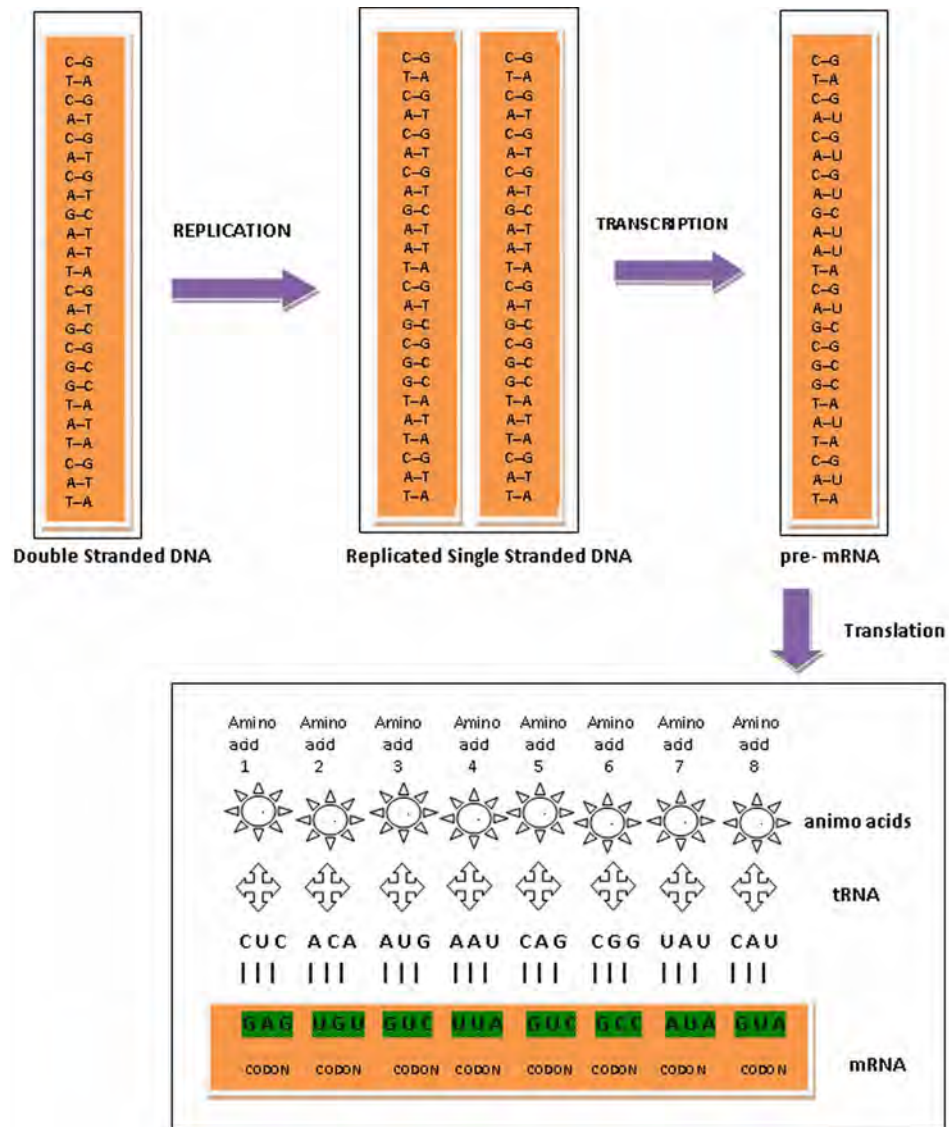
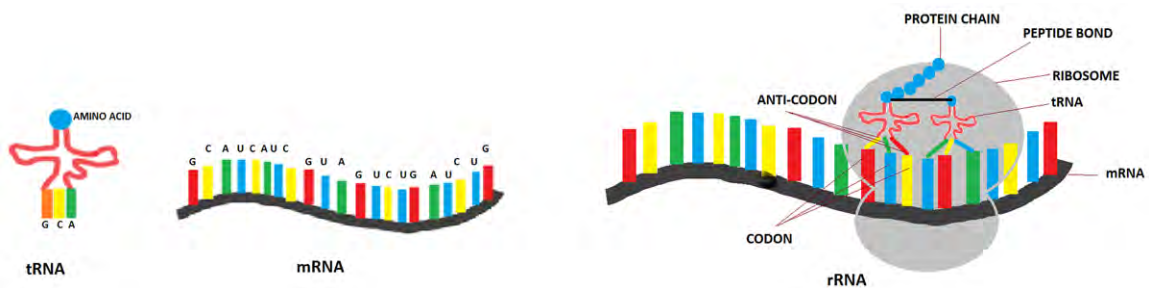


Figure 2. tRNA, mRNA and rRNA



of treating this mRNA with siRNA is to degrade the mRNA and thus prevent the formation of protein causing disease (Figure 3). Alteration in protein production refers to impeding the gene expression process. This concept of RNAi leads to Gene silencing that means switching off the target gene. It prevents the expression of a certain gene. One of the applications of Gene Silencing technique was to cure Huntington disease which was caused by the production of protein called Huntington (Sah & Aronin, 2011).

RNA based medicines have been in focus since many years. RNA Therapeutics drug are designed to increase or decrease the production of a protein involved in a disease. It is a mechanism to regulate normal gene expression in order to eradicate several disorders like cancer, Hepatitis C, Huntington's Disease, HIV, malaria etc. (Abdolhamid et al.,2010). Some of the Therapeutics companies have been discussed in Table 1.

PMOs Based RNA Therapeutics

PMO stands for phosphorodiamidate morpholino oligomer. These are synthetic RNA's consisting of same bases as of RNA (A, C, G and T). They

are molecular structures that are binded to RNA at specific position in order to change the function of RNA. As they have reasonable RNA binding affinity and low protein binding affinity, they bind RNA wholly without off- target effects. Their prime function is to retard the gene expression in order to prevent cells from making a targeted protein and thus preventing the disease to occur.

PMO's are especially stable, highly soluble and non-toxic. They have the following functions: (Moulton, 2013) blocking translation of mRNA (Hudziak, Summerton, Weller & Iversen, 2000), modifying of splicing of pre-mRNA (Alter et al.,2006), blocking miRNA maturation and activity (Kloosterman, Lagendijk, Ketting, Moulton, & Plasterk, 2007), and blocking other RNA sites – splice regulation sequences (Bruno, Jin & Cote,2004) and ribozymes (Yen et al., 2004).

The manufacture of medicine using PMO technology drives through a number of steps. After examining the target RNA, subunits of PMOs are created by converting 5 sided RNA rings to 6 sided PMO subunits. Then these are linked to form the final structure. This structure corresponds to the targeted RNA. Figure 4 has been adapted from Sarepta Therapeutics.

Figure 3. RNAi process

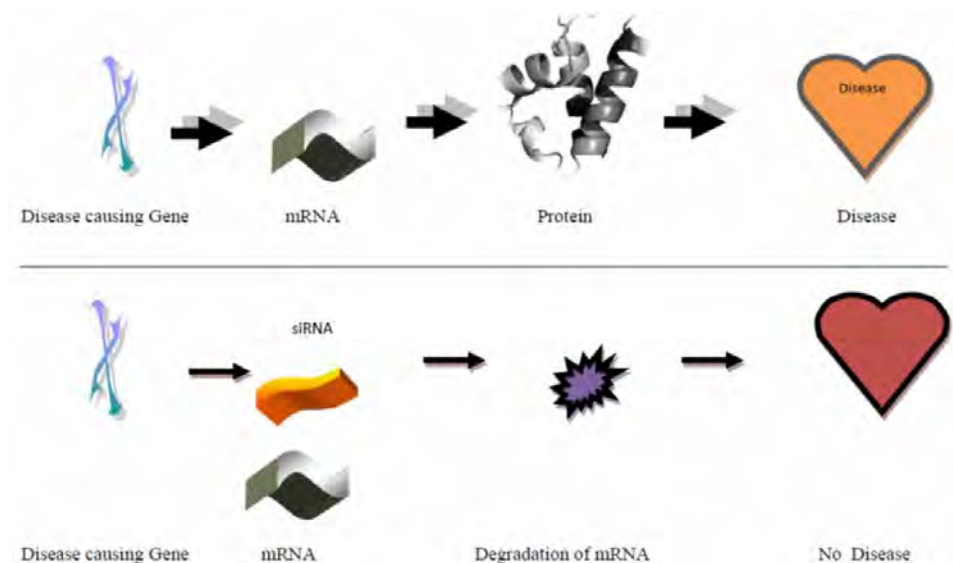
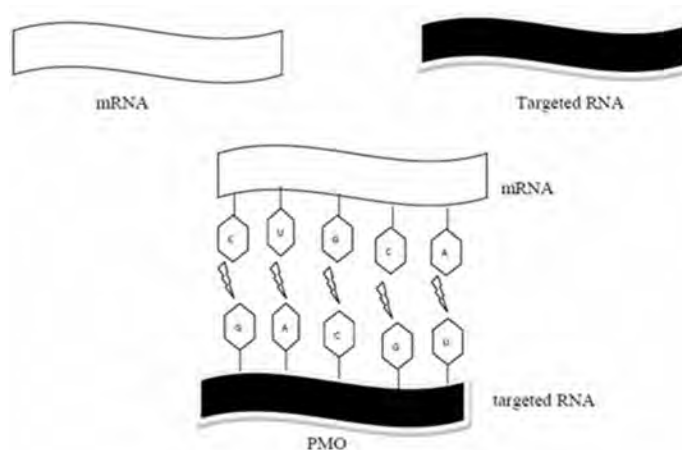


Table 1. RNA therapeutics companies

S.No	Company	Technology	Targetted RNA	Pipeline	Diseases	Web Address
1	Sarepta Therapeutics	PMO's	mRNA	Exon Skipping	Duchenne muscular dystrophy (DMD)	http://www.sarepta.com/
2	Arcturus Therapeutics	LUNAR UNA	siRNA, mRNA	ornithine transcarbamylase deficiency, Iron disorders	TTR-mediated amyloidosis	www.arcturusrx.com/
3	RaNA	RNA upregulation	long nc RNA	musculoskeletal diseases, and inflammatory diseases	Hematologic disease	ranarx.com
4	Alnylam	Gene Silencing	dsRNA	Cardio-Metabolic Disease, dyslipidemia, hypertension, diabetes, and non-alcoholic steatohepatitis (NASH); and Hepatic Infectious Disease	RSV infection/ hypercholesterolemia	www.alnylam.com/
5	SMaRT	Antisense Technology	asRNA	Myotonic Dystrophy Type 1	breast cancer, prostate cancer, viral infections, deafness	smart-therapeutics.com/
6	Mina Therapeutics	Transcriptional gene activation	saran	development and commercialization of small activating RNA (saRNA) based therapeutics	Liver function, diabetes	http://minatx.com/

Figure 4. PMO's technology



Duchenne muscular dystrophy (DMD) was treated with PMO's Technology. This disease is caused by mutations that disrupt the reading frame of the human DMD gene. Selective removal of exons flanking an out-of-frame DMD mutation can result in an in-frame mRNA transcript that

may be translated into an internally deleted, Becker muscular dystrophy (BMD)-like, but functionally active dystrophin protein with therapeutic activity (Poppewell, Graham, Malerba & Dickson, 2011).

PMO are RNA based molecules in which morpholine ring replaces the ribose. PMOs are also

used for restoration of dystrophin in humans which protects from muscle fiber injury (Kinali et al., 2009) as strengthening of muscles depends on it.

Huntington's disease (HD) is a progressive brain disorder genetically transferred disease that hits a person's cognitive, emotional behavior and leads to uncontrolled jerky movements in body. Mutations in HTT gene cause HD. A trinucleotide sequence CAG of DNA sequence repeats itself 20-30% times leading to HD. PMO's have been used to suppress mutant huntingtin expression (Sun et al., 2014; Burgess, 2014).

Behavior of Flavivirus has also been studied using PMO's. Flavivirus is a genus of viruses in the family Flaviviridae. This genus includes the West Nile virus, dengue virus, tick-borne encephalitis virus and yellow fever virus (Brinton & Disputo, 1988). Viruses like West Nile virus (Deas et al., 2007), Marburg virus, Ebola virus (Warfield et al., 2006), etc are under test while treating them with PMO's. PMO's whose sequences were complementary to RNA elements located in the WNV genome included various sub structures like stem loops, Pseudoknot (Sharma, Singh & Chand, 2014; Sharma, Singh & Chand, 2015; Jiwan & Singh, 2012a; Jiwan & Singh, 2012b; Mann, & Singh, 2011) etc.

LUNAR Technology Based RNA Therapeutics

LUNAR stands for Lipid-enabled and Unlocked Nucleic Acid modified RNA delivery technology. In this technique, lipid particle is fused into the cell membrane that leads to biodegradation process and delivers the required RNA. Arcturus Therapeutics, Inc. ("Arcturus Therapeutics", 2015), a leading company in RNA therapies for rare diseases works on this technology.

Cystic Fibrosis (Hegarty et al., 2009) is a genetic disease that affects the lungs and digestive system. Cystic Fibrosis is being treated with the help of LUNAR technology. Researchers at Arcturus Therapeutics tried injecting human erythropoietin messenger RNA (mRNA) into non-human

primates using LUNAR technology. The results were outstanding. The chemically modified RNA was well absorbed by the primate and protein expression proceeded as expected. Most important there were no off target effects of the activity.

Both siRNA and mRNA are used with LUNAR technology. Similar to LUNAR-formulated cystic fibrosis transmembrane conductance regulator (CFTR) mRNA, LUNAR-formulated thrombopoietin (TPO) mRNA, LUNAR-formulated ornithine transcarbamylase (OTC) mRNA and LUNAR-formulated hepcidin (HPN) mRNA have been geared up. LUNAR-formulated TPO mRNA is used for the treatment of thrombocytopenia which is a blood related disorder that leads to low platelet count. LUNAR-formulated ornithine transcarbamylase (OTC) mRNA is used for the treatment of ornithine transcarbamylase deficiency, which leads to accumulation of ammonia in blood. LUNAR-formulated hepcidin mRNA is used for the treatment of iron disorders.

Arcturus Therapeutics ("Arcturus Therapeutics", 2015) has rightly proved that mRNA medicines have brought a revolution. It has become possible to target any gene in the human genome (Nassa, Singh & Goel, 2013). Various gene prioritization techniques (Gill, Singh, & Aseri, 2014), gene prediction techniques (Goel, Singh & Aseri, 2013a; Goel, Singh & Aseri, 2013b) and gene regulatory networks (Singh, 2013a; Singh, 2013b) are used to identify candidate genes for a particular disease. Not only in humans, genetic testing can be conducted even in unborn babies to detect diseases (Singh, & Singh, 2011).

Therapeutics for RNA Upregulation Using Long Non-Coding RNA

Few years back non-coding RNA was recognized as responsible for several regulatory functions in body. Long non-coding RNA has a specific effect on gene regulation. Using this concept, RNA can be targeted to selectively 'activate' therapeutic protein expression (Lee, 2012). Till now in this chapter, we have discussed the evolution of RNA

therapeutics for controlling gene expression, but now we would discuss the same for up-regulating the gene expression (Donner, 2014). Some genes including tumor suppressors, growth factors, transcription factors and genes that are deficient in various genetic diseases require upregulation (Wahlestedt, 2013).

During cancer, some tumor suppressor genes are inhibited. These genes encode protein that hinders the formation of tumors. Mutations in the said gene inhibit the function of tumor suppressor gene, thereby leading to growth of infected tissues. A two-hit mutation process (Hyland, 2008) is shown in Figure 5 that results in tumor.

In cardiovascular diseases, therapeutic genes are upregulated using Long non-coding RNA like APOA1-AS and human mimics of Bvht (Fatemi, Velmeshev & Faghihi, 2014). Similarly neurological disorders occur due to downregulation of a protein called Brain-derived neurotrophic factor (BDNF). BDNF is a therapeutic target for neurological disorders. As LncRNA has the potential to target specific set of genes, it has been a major therapeutic target for upregulation.

Not only lncRNA but short activating RNA has also been used for upregulation. Mina Therapeutics proposed a novel approach for treating diabetes (Reebye et al., 2013). Diabetic patients who are bound to inject insulin supplements in their body may be treated with the said approach. In this approach, short activating RNA is used to produce insulin in body using surrogate cells by upregulating several pancreatic endodermal genes.

Antisense RNA Based Gene Silencing

Antisense RNA (asRNA) is RNA having molecular sequence which is complementary to a specific mRNA. Thus asRNA can be binded to mRNA thus forming a double stranded RNA just like DNA. This double stranded RNA can inhibit gene expression. There are various sources of asRNA including induction by doctors. Gene silencing component is produced using asRNA which then interacts with mRNA for its degradation. This process finally alters the gene expression thereby curing the disease. Such small molecules (gene silencing components) have been synthesized to cure many diseases like Myotonic Dystrophy Type 1 and Myotonic Dystrophy Type 2.

Neuromuscular diseases like DMD, myotonic dystrophy and spinal muscular atrophy (SMA) are also treated using antisense oligonucleotides (Muntoni, & Wood, 2011). Myotonic Dystrophy occurs due to repetition of some units in non coding elements of mRNA. More the number of repetitions, severe is the disease. Spinal muscular atrophy (SMA) is another genetic disease affecting nervous system that controls muscle movement. SMA occurs due to deficiency of SMN, a motor neuron protein. Antisense oligonucleotides have been designed to modulate pre-mRNA splicing for restoration of functional proteins in body.

Not only antisense oligonucleotides, but double stranded RNA (dsRNA) that is homologous to the target gene is also used for gene silencing. In

Figure 5. Two hit mutation for retinoblastoma

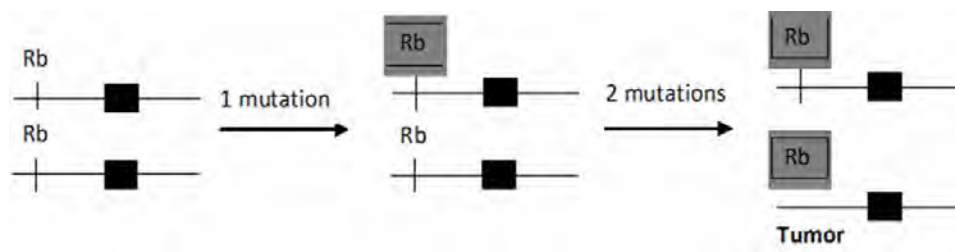
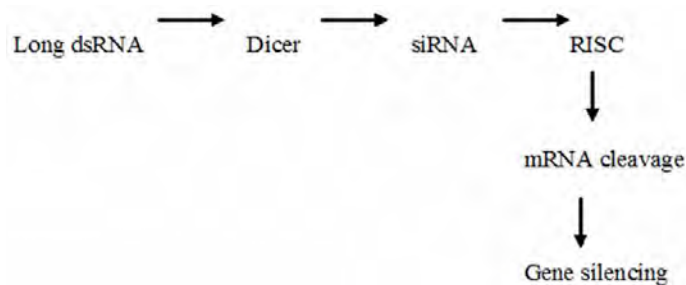


Figure 6. Antisense RNAi based gene silencing



this technique dsRNA is diced to produce siRNA (Wood, Trülzsch, Abdelgany, & Beeson, 2003). The antisense strand of siRNA, being complementary to mRNA, leads to mRNA cleavage, degradation and hence gene silencing as shown in Figure 6.

RNAi THERAPEUTICS: CHALLENGES AND ISSUES

RNAi Therapeutics is not easy to implement. Also there are certain issues (Ebbesen, Jensen, Andersen & Pedersen, 2008; Burnett & Rossi, 2012) involved such as:

- Induction of drug into the specific body part,
- Distribution of drug into cells,
- Successful silencing of gene,
- Side Effects of drug on other genes,
- Generate natural immune response,
- Instability under physiological conditions.

Biodistribution of RNAi drugs includes technical and scientific challenges including extended blood circulation times, vascular escape, tissue penetration, cellular uptake, and escape into the cytoplasm (Haussecker, 2014). First and foremost question is how to deliver the drug to the body. Some of the techniques are nanoparticles, cationic lipids, antibodies, cholesterol, aptamers, viral vectors GalNAc-siRNA conjugates and Dynamic PolyConjugates/DPCs (Aagaard, & Rossi, 2007). Two strategies may be adopted- in vivo delivery

or ex vivo delivery. In vivo delivery refers to concept of delivering the said drug to the body itself through inhaler or direct injection. On the contrary, ex vivo is a concept of extracting cell from patients, modify them in environment external of body and re-inject the modified cell back to the body. Both the techniques have their own challenges (Tiemann & Rossi, 2009).

Another issue of RNAi delivery is that RNA molecule has to reach cytoplasm of cell in order to interact with protein and hence carry forward the process of gene silencing. For example, in case of cancer therapy, when siRNA routes through the blood, it is easily digested by nucleases in the serum thereby leading to its degradation instead of following its targeted action.

Off target effect is another problem. A gene which is closely related to the targeted gene is also damaged. This leads to unwanted action in body. The literature shows masses of consequences of experiments conducted so far for off-target effects of RNAi delivery. In some cases physiological effects have lead to non-targeted gene expression (Jackson et al., 2003) and even alteration of gene expression of many non-targeted transcripts.

Another growing concern while inducing RNAi in vivo is the activation of immune response. This even sometimes leads to interferon effect. Interferons (Andrea, Gariglio, Gioia, Gariglio & Landolf, 2002) are actually signaling proteins that are sent by cells to other cells signaling them to increase their immune effect. So, the immunostimulatory “side effects” must be taken in account when inducing RNAi.

FUTURE RESEARCH DIRECTIONS

RNAi Therapeutics has been improving incessantly since a decade. Although there are some hurdles to be able to traverse clinical barrier easily, particularly RNA delivery. Lots of lessons have been learnt from previous trials. These lessons are unquestionably going to aid researchers and scientists to accomplish their goal to conquer fatal diseases. With awareness of the topics of gene expression, artificial RNA, techniques for RNA Therapeutics like PMO, Lunar based approach, antisense RNA etc, a number of researchers can bring revolution in disease control. Firstly, research can be carried out on structures of RNA that cause a particular disease. Further research can be done on preparation of synthetic RNA to control that disease. Delivery of drug to cells is another area.

CONCLUSION

As RNA is associated with a large number of human diseases, treatment of these diseases are also related to RNA. The process of Gene expression is the foundation of functionality of human body. Gene expression includes RNA. Controlling the gene expression leads to control of formation of diseases causing structures in nucleotide sequence. Synthetic RNA controls the gene expression and thus there is no doubt regarding the potential of RNAi Therapeutics in treatment of diseases. Novel techniques and delivery systems are expected to be investigated to perk up RNAi delivery. RNAi Therapeutics should be able to treat more number of diseases in near future. It is going to bring revolution in human disease therapy.

REFERENCES

- Aagaard, L., & Rossi, J. J. (2007). RNAi therapeutics: Principles, prospects and challenges. *Advanced Drug Delivery Reviews*, 59(2-3), 75–86. doi:10.1016/j.addr.2007.03.005 PMID:17449137
- Abdolhamid, A. S., Hedayati, S. S., Poor, R. H., Madani, S., Poor, S. S., & Panahi, S. (2010). Application of RNA interference in treating human diseases. *Journal of Genetics*, 3(3), 89.
- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2007). *Molecular Biology of the Cell* (5th ed.). New York: Garland Science. Retrieved from http://classes.biology.ucsd.edu/bicd110.SU2.09/CH1_7_1.pdf
- Alter, J., Lou, F., Rabinowitz, A., Yin, H., Rosenfeld, J., Wilton, S. D., & Lu, Q. L. et al. (2006). Systemic delivery of morpholino oligonucleotide restores dystrophin expression bodywide and improves dystrophic pathology. *Nature Medicine*, 12(2), 175–177. doi:10.1038/nm1345 PMID:16444267
- Andrea, M. D., Gariglio, M., Gioia, D., Gariglio, M., & Landolfo, S. (2002). The interferon system: An overview. *European Journal of Paediatric Neurology*, 6, A41–A58. doi:10.1053/ejpn.2002.0573 PMID:12365360
- Anfinsen, C. (1972). The formation and stabilization of protein structure. *Biochemical Journal*, 128(4), 737–749.
- Arcturus Therapeutics, Inc. (n.d.). Retrieved from <http://arcturusrx.com>
- Blasco, M. A. (2005). Telomeres and human disease: Ageing, cancer and beyond. *Nature Reviews. Genetics*, 6(8), 611–622. doi:10.1038/nrg1656 PMID:16136653
- Brinton, M. A., & Disposito, J. H. (1988). Sequence and secondary structure analysis of the 5-terminal region of flavivirus genome RNA. *Virology*, 162(2), 290–299. doi:10.1016/0042-6822(88)90468-0 PMID:2829420
- Bruno, I. G., Jin, W., & Cote, G. J. (2004). Correction of aberrant FGFR1 alternative RNA splicing through targeting of intronic regulatory elements. *Human Molecular Genetics*, 13(20), 2409–2420. doi:10.1093/hmg/ddh272 PMID:15333583

- Burgess, D. J. (2014). Targeting huntingtin through morpholino oligomers. *Nature Reviews. Genetics*, 15, 572. doi:10.1038/nrg3804
- Burnett, J., & Rossi, J. (2012). RNA-based therapeutics: Current progress and future prospects. *Chemistry & Biology*, 19(1), 60–71. doi:10.1016/j.chembiol.2011.12.008 PMID:22284355
- Chaudhuri, T. K., & Paul, S. (2006). Protein-misfolding diseases and chaperone-based therapeutic approaches. *The FEBS Journal*, 273(7), 1331–1349. doi:10.1111/j.1742-4658.2006.05181.x PMID:16689923
- Deas, T. S., Bennett, C. J., Jones, S., Tilgner, M., Ren, P., Behr, M. J., & Shi, P. et al. (2007). In vitro resistance selection and in vivo efficacy of morpholino oligomers against west nile virus. *Antimicrobial Agents and Chemotherapy*, 51(7), 2470–2482. doi:10.1128/AAC.00069-07 PMID:17485503
- Diner, I., Hales, C. M., Rabenold, L., Bishof, I., & Duong, D. M. (2004). Aggregation properties of the small nuclear ribonucleoprotein U170K in Alzheimer disease. *The Journal of Biological Chemistry*, 289(51), 35296–35313. doi:10.1074/jbc.M114.562959 PMID:25355317
- Donner, A. (2014). RNA is for activation. *SciBX*, 7(5). doi:10.1038/scibx.2014.136
- Ebbesen, M., Jensen, T. G., Andersen, S., & Pedersen, F. S. (2008). Ethical perspectives on RNA interference therapeutics. *International Journal of Medical Sciences*, 5(3), 159–168. doi:10.7150/ijms.5.159 PMID:18612370
- Fatemi, R. P., Velmeshev, D., & Faghihi, M. A. (2014). De-repressing LncRNA-Targeted Genes to Upregulate Gene Expression: Focus on Small Molecule Therapeutics, *Molecular Therapy. Nucleic Acids*, 3(11), e196. doi:10.1038/mtna.2014.45 PMID:25405465
- Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., & Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*, 391(6669), 806–811. doi:10.1038/35888 PMID:9486653
- Gill, N., Singh, S., & Aseri, T. A. (2014). Computational disease gene prioritization: An appraisal. *Journal of Computational Biology*, 21(6), 456–465. doi:10.1089/cmb.2013.0158 PMID:24665902
- Goel, N., Singh, S. & Aseri, T.(2013a). A Review of Soft Computing Techniques for Gene Prediction. *ISRN Genomics*, 1-8.
- Goel, N., Singh, S., & Aseri, T. (2013b). A comparative analysis of soft computing techniques for gene prediction. *Analytical Biochemistry*, 438(1), 14–21. doi:10.1016/j.ab.2013.03.015 PMID:23529114
- Haussecker, D. (2014). Current issues of RNAi therapeutics delivery and development. *Journal of Controlled Release*, 195, 49–54. doi:10.1016/j.jconrel.2014.07.056 PMID:25111131
- Hegarty, M., Macdonald, J., Watter, P., & Wilson, C. (2009). Quality of life in young people with cystic fibrosis: Effects of hospitalization, age and gender, and differences in parent/child perceptions. *Child: Care, Health and Development*, 35(4), 462–468. doi:10.1111/j.1365-2214.2008.00900.x PMID:18991968
- Hoffman, J. R., & Falvo, M. J. (2004). Protein - Which is best? *Journal of Sports Science and Medicine*, 3(3), 118–130. PMID:24482589
- Hudziak, R. M., Summerton, J., Weller, D. D., & Iversen, P. L. (2000). Antiproliferative effects of steric blocking phosphorodiamidate morpholino antisense agents directed against c-myc. *Antisense & Nucleic Acid Drug Development*, 10(3), 163–176. doi:10.1089/oli.1.2000.10.163 PMID:10905553

- Hyland, K. M. (2008). Tumor Suppressor Genes and Oncogenes: Genes that Prevent and Cause Cancer. *Biochemistry/Molecular Biology Lecture*, 71-86. Available: <http://biochemistry2.ucsf.edu/programs/ptf/m3%20links/TumorSuppressLEC.pdf>
- Jackson, A. L., Bartz, S. R., Schelter, J., Kobayashi, S. V., Burchard, J., Mao, M., & Linsley, P. S. et al. (2003). Expression profiling reveals off-target gene regulation by RNAi. *Nature Biotechnology*, 21(6), 635–637. doi:10.1038/nbt831 PMID:12754523
- Jiwan, A., & Singh, S. (2012a). A review on RNA pseudoknot structure prediction techniques, *IEEE International Conference on Computing, Electronics and Electrical Technologies (IC-CEET)*, Kumaracoil, 975-978. doi:10.1109/IC-CEET.2012.6203854
- Jiwan, A., & Singh, S. (2012b). Soft Computing based Model for Identification of Pseudoknots in RNA Sequence using Learning Grammar. *International Journal of Computers and Applications*, 54(9), 1–7. doi:10.5120/8591-2344
- Jorgensen, R. A., Cluster, P. D., English, J., Que, Q., & Napoli, C. A. (1996). Chalcone synthase cosuppression phenotypes in petunia flowers: Comparison of sense vs. antisense constructs and single-copy vs. complex T-DNA sequences. *Plant Molecular Biology*, 31(5), 957–973. doi:10.1007/BF00040715 PMID:8843939
- Kinali, M., Arechavala-Gomez, V., Feng, L., Cirak, S., Hunt, D., Adkin, C., & Muntoni, F. et al. (2009). Local restoration of dystrophin expression with the morpholino oligomer AVI-4658 in Duchenne muscular dystrophy: A single-blind, placebo-controlled, dose-escalation, proof-of-concept study. *Lancet Neurology*, 8(10), 918–928. doi:10.1016/S1474-4422(09)70211-X PMID:19713152
- Kloosterman, W. P., Lagendijk, A. K., Ketting, R. F., Moulton, J. D., & Plasterk, R. H. A. (2007). Targeted inhibition of miRNA maturation with morpholinos reveals a role for miR-375 in pancreatic islet development. *PLoS Biology*, 5(8), e203. doi:10.1371/journal.pbio.0050203 PMID:17676975
- Lee, J. T. (2012). Epigenetic regulation by long noncoding RNAs. *Science*, 338(6113), 1435–1439. doi:10.1126/science.1231776 PMID:23239728
- Mann, R., & Singh, S. (2011). Effect of Neural Network Parameters on RNA Secondary Structure Classification. *International Journal of Computers and Applications*, 23(6), 6–9. doi:10.5120/2894-3786
- Moulton, H. M. (2013). In Vivo Delivery of Morpholino Oligos by Cell-Penetrating Peptides. *Current Pharmaceutical Design*, 19(16), 2963–2969. doi:10.2174/1381612811319160010 PMID:23140456
- Muntoni, F., & Wood, M. J. (2011). Targeting RNA to treat neuromuscular disease. *Nature Reviews. Drug Discovery*, 10(8), 621–637. doi:10.1038/nrd3459 PMID:21804598
- Nassa, T., Singh, S., & Goel, N. (2013). Splice Site Detection in DNA Sequences using Probabilistic Neural Network. *International Journal of Computers and Applications*, 76(4), 1–4. doi:10.5120/13232-0664
- Pellizzoni, L. (2007). Chaperoning ribonucleoprotein biogenesis in health and disease. *EMBO Reports*, 8(4), 340–345. doi:10.1038/sj.embor.7400941 PMID:17401408
- Pines, M. (2007). *The Structures of Life*. NIH Publ., No. 07–2778. Available: <http://publications.nigms.nih.gov/structlife/structlife.pdf>

- Popplewell, L. J., Graham, I. R., Malerba, A., & Dickson, G. (2011). Bioinformatic and functional optimization of antisense phosphorodiamidate-morpholino oligomers (PMOs) for therapeutic modulation of RNA splicing in muscle. *Methods in Molecular Biology (Clifton, N.J.)*, 709, 153–178. doi:10.1007/978-1-61737-982-6_10 PMID:21194027
- Reebye, V., Sætrom, P., Mintz, P. J., Rossi, J. J., Kasahara, N., Nteliopoulos, G., & Habib, N. A. et al. (2013). A Short-activating RNA Oligonucleotide Targeting the Islet β -cell Transcriptional Factor MafA in CD34(+) Cells, *Molecular Therapy. Nucleic Acids*, 2, e97. doi:10.1038/mtna.2013.23 PMID:23736775
- Sah, D. W. Y., & Aronin, N. (2011). Review series Oligonucleotide therapeutic approaches for Huntington disease. *The Journal of Clinical Investigation*, 121(2), 500–507. doi:10.1172/JCI45130 PMID:21285523
- Selkoe, D. J. (2003). Folding proteins in fatal ways. *Nature*, 426(6968), 900–904. doi:10.1038/nature02264 PMID:14685251
- Sharma, D., Singh, S., & Chand, T. (2014). RNA Secondary Structure Prediction Algorithms including Pseudoknots. *Current Bioinformatics*, 9(3), 268–283. doi:10.2174/15748936113086660010
- Sharma, D., Singh, S., & Chand, T. (2015). RNA Pseudoknot : Topology and Prediction. *Proc International Conference on Computer and Computational Sciences (ICCCS), Greater Noida*, 257-261.
- Shenoy, S. R., & Jayaram, B. (2010). Proteins: Sequence to structure and functioncurrent status. *Current Protein & Peptide Science*, 11(7), 498–514. doi:10.2174/138920310794109094 PMID:20887265
- Singh, S. (2013a). A Gene Regulatory Network Prediction Method using Particle Swarm Optimization and Genetic Algorithm. *International Journal of Computers and Applications*, 83(12), 32–37. doi:10.5120/11410-6743
- Singh, S. (2013b). Approaches to Gene Regulatory Network Modeling, *International Journal of IT. Engineering and Applied Sciences Research*, 2(4), 35–38.
- Singh, S. & Singh, A. (2011). Emerging Web Tools and Their Applications in Bioinformatics, *Business Organizations and Collaborative Web: Practices, Strategies and Patterns*, 76-89. DOI:10.4018/978-1-60960-581-0.ch006
- Sun, X., Marque, L. O., Cordner, Z., Pruitt, J. L., Bhat, M., Li, P. P., & Rudnicki, D. D. et al. (2014). Phosphorodiamidate morpholino oligomers suppress mutant huntingtin expression and attenuate neurotoxicity. *Human Molecular Genetics*, 23(23), 6302–6317. doi:10.1093/hmg/ddu349 PMID:25035419
- Tiemann, K., & Rossi, J. J. (2009). RNAi-based therapeutics-current status, challenges and prospects. *EMBO Molecular Medicine*, 1(3), 142–151. doi:10.1002/emmm.200900023 PMID:20049714
- Tinoco, L., Jr., & Bustamante, C. (1999). How RNA folds, *Journal of Molecular Biology*, 293(2), 271–281. DOI: 1055020810.1006/jmbi.1999.3001.PMID
- Wahlestedt, C. (2013). Targeting long non-coding RNA to therapeutically upregulate gene expression. *Nature Reviews. Drug Discovery*, 12(6), 433–446. doi:10.1038/nrd4018 PMID:23722346
- Warfield, K.L., Swenson, D.L., Olinger, G.G., Nichols, D.K., Pratt, W.D., Blouch, R., Stein, D.A., ... Bavari, S. (2006). Gene-specific countermeasures against Ebola virus based on antisense phosphorodiamidate morpholino oligomers. *PLoS Pathogens*, 2(1), 5-13. doi:10.1371/journal.ppat.0020001

Wood, M. J., Trülsch, B., Abdelgany, A., & Beeson, D. (2003). Therapeutic gene silencing in the nervous system. *Human Molecular Genetics*, 12(2), R279–R284. doi:10.1093/hmg/ddg275 PMID:12928477

Yamaguchi, H., Calado, R. T., Ly, H., Kajigaya, S., Baerlocher, G. M., Chanock, S. J., & Young, N. S. et al. (2005). Mutations in TERT, the gene for telomerase reverse transcriptase, in aplastic anemia. *The New England Journal of Medicine*, 352(14), 1413–1424. doi:10.1056/NEJMoa042980 PMID:15814878

Yen, L., Svendsen, J., Lee, J. S., Gray, J. T., Magnier, M., Baba, T., & Mulligan, R. C. et al. (2004). Exogenous control of mammalian gene expression through modulation of RNA self-cleavage. *Nature*, 431(7007), 471–476. doi:10.1038/nature02844 PMID:15386015

ADDITIONAL READING

Carthew, R. W., & Sontheimer, E. J. (2009). Origins and Mechanisms of miRNAs and siRNAs, Northwestern University. *Cell*, 136(4), 642–655. doi:10.1016/j.cell.2009.01.035 PMID:19239886

Eddy, S. R. (2001). Non-Coding RNA Genes and the Modern RNA World. *Nature Reviews. Genetics*, 2(12), 919–929. doi:10.1038/35103511 PMID:11733745

Forbes, D. J., Kornberg, T. B., & Kirschner, M. W. (1983). Small Nuclear RNA Transcription and Ribonucleoprotein assembly an early *Xenopus* development. *The Journal of Cell Biology*, 97(1), 62–72. doi:10.1083/jcb.97.1.62 PMID:6190822

Gupta, S., Singh, R. P., Rabadia, N., Patel, G., & Panchal, H. (2011). Antisense Technology. *International Journal of Pharmaceutical Sciences Review and Research*, 9(2), 38–45.

Rich, M., & RajBhandary, U. L. (1976). Transfer RNA: Molecular Structure, Sequence, and Properties. *Annual Review of Biochemistry*, 45(1), 805–860. doi:10.1146/annurev.bi.45.070176.004105 PMID:60910

Shabalina, S. A., Ogurtsov, A. Y., & Spiridonov, N. A. (2006). Periodic pattern of mRNA secondary structure created by the genetic code, *Nucleic Acids Research. National Center for Biotechnology Information*, 34(8), 2428–2437.

Zwieb, C., Nues, R. W. V., Rosenblad, M. A., Brown, J. D., & Samuelsson, T. (2005). A nomenclature for all signal recognition particle RNAs, *RNA Society. RNA (New York, N.Y.)*, 11(1), 7–13. doi:10.1261/rna.7203605 PMID:15611297

KEY TERMS AND DEFINITIONS

dsRNA: dsRNA stands for double stranded RNA. Sense and antisense RNA together form dsRNA.

Gene Expression: It is the process of formation of protein, intermediate steps being conversion from DNA to mRNA.

Gene Silencing: The process of prevention of gene expression process during translation or transcription process is called gene silencing.

LUNAR Technology: LUNAR stands for Lipid-enabled and Unlocked Nucleic Acid modified RNA delivery technology. In this technique, lipid particle is fused into the cell membrane that leads to biodegradation process and delivers the required RNA.

PMO: PMO (phosphorodiamidate morpholino oligomer) are synthetic RNA who retards the gene expression in order to prevent cells from making a targeted protein. They are molecular structures that are binded to RNA at specific position.

RNA Interference: The mechanism of RNA Interference (RNAi) involves treating the disease causing gene with specific small interfering RNA (siRNA) that targets the specific messenger RNA (mRNA).

RNAi Therapeutics: Altering gene coding at mRNA level is called RNAi Therapeutics. RNA Therapeutics drug are designed to increase or decrease the production of a protein involved in a disease.

siRNA: Small interfering RNA is also known as silencing RNA or short interfering RNA. siRNA is formed from double-stranded RNA molecular sequence of length 20-25 base pairs and is used for knockdown of long noncoding RNAs genes.

Category B

Biomedical Engineering

General Perspectives on Electromyography Signal Features and Classifiers Used for Control of Human Arm Prosthetics

Faruk Ortes

Istanbul University, Turkey

Derya Karabulut

Halic University, Turkey

Yunus Ziya Arslan

Istanbul University, Turkey

INTRODUCTION

Physically handicapped people encounter various kinds of obstacles and difficulties in their daily lives due to the restricted ability of motion. Assistive technologies represent a crucial challenge of scientific studies to overcome such an issue of reducing quality of life. Assistive devices such as wheelchairs, orthoses and prostheses are designed and built to contribute rehabilitation progress and to regain lost functions, as well. Although human body parts have intricate forms and functions, artificial devices and components integrating to the body are anticipated to compensate the fundamental functions related to user's demands. Upper or lower arm amputations also result in severe cosmetic matters. However, what is more important and obtrusive is the loss of primary functions including manipulating and grasping the objects besides the locomotor tasks which are performed by human body during daily activity.

BACKGROUND

Development of human arm prosthetics, which are improved to regain lost functions of amputated limbs, encounters critical and challenging problems to carry out various dexterous tasks. To

date, many of revolutionizing design of human arm prosthetics including Boston Arm (Mann & Reimers, 1970), Deka Arm (Resnik, 2010), Otto Bock trans carpal hand (Otto Bock Health Care, Minneapolis, MN), and Shanghai Kesheng Hands (Shanghai Kesheng Prosthese Corporation Ltd.) have been developed. Intuitive and precise control of such prostheses is still one of the main interests of scientific studies. The main deduction from researches could be stated as control of the prosthetics is a particular concern of understanding the nature of the electrical activations of muscles. Imitation of the fundamental patterns of human arm motion depends highly upon the transformation of the neuromuscular activities of residual limbs to a specific control signal for controlling the artificial arm. In this respect, myoelectric signals provide a base of intuitive control, unlike the conventional or direct control. The dexterous control of such myoelectric-based prostheses requires a clear extraction of features from recorded surface electromyography (SEMG) signals and pattern recognition to discriminate the motion and force intentions of the prosthetics users. The progress of feature extraction from SEMG signals has an extensive coverage of myoelectric controlled prostheses studies due to the features in both time and frequency domains have the great potential on representing clear and meaningful information

of EMG signals. Additionally, the feature classifiers have been given a special scientific interest by researchers. Selection and developing of the case-specific classifiers, which are desired to have the optimal performance to specify motion classes, still continue to be the main goal of current studies. Although, various types of classifiers such as linear discriminant analysis (LDA), support vector machine (SVM), artificial neural networks (ANN) and fuzzy logic (FL) techniques have been utilized to classify human arm motion patterns, merits, shortcomings and pitfalls of the classifiers are still required to be discussed extensively.

FUNDAMENTAL ASPECTS OF EMG

EMG is the electrical activity of skeletal muscles (Basmajian & Deluca, 1985). It represents the summation of the muscle action potentials which cause the contraction of muscle fibers. Recorded EMG data by means of electrodes are amplified and filtered to eliminate the motion artifacts, as well as the environment and device related noises. Rejection of ambient influences on natural muscle activation improves the accuracy and usability of EMG signals. One of the most widely usage of EMG signals is to control the myoelectric-based prosthetics which are used by amputated people. Control scheme for EMG-driven human arm prosthetics includes a sequential series of signal processing (Figure 1).

A condensed and clear control signal is needed to control the EMG-based prosthetics. In order to reduce calculation and to provide stability of signal, EMG data are scanned by sliding segmented windows (Figure 2). Because the raw (amplified+pre-processed) EMG signal contains a huge burden of data, this signal is needed to be represented in a concise, but accurate ways. Widely used time domain features extracted from signals includes mean absolute value (MAV), root mean square (RMS), Willison amplitude (WAMP), waveform length (WL), variance of EMG (VAR), simple square integral (SSI), zero crossing (ZC) and integrated EMG (IEMG) (Phinyomark et al., 2013). In frequency domain, mean frequency, median frequency, peak frequency, mean power, total power, and spectral power features are commonly preferred (Phinyomark et al., 2013).

EMG SIGNAL FEATURES

Obtained EMG signals during contraction of a muscle or muscle groups are needed to be quantified in order to relate these signals with some certain sets of movement types (Zecca, Micera, Carrozza, & Dario, 2002). Mathematical expression of EMG signals could be defined using feature extraction approach. An EMG signal could be expressed in two domains including time and frequency domains.

Figure 1. Control scheme of multifunctional human arm prosthetics

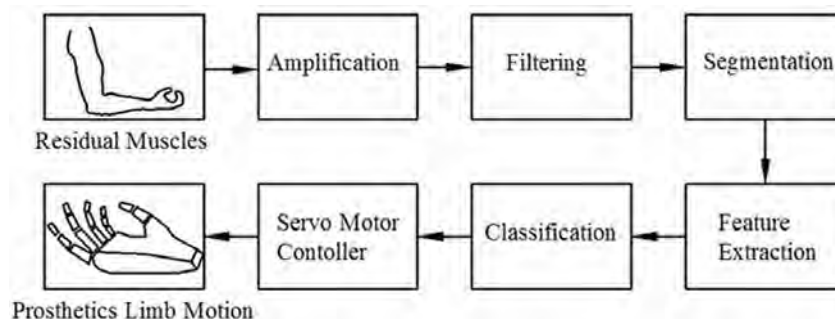
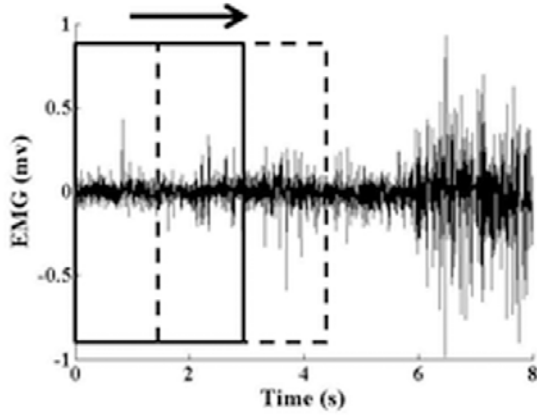


Figure 2. The basic representation of sliding windows



Time Domain Features

Features expressed in time domain are useful for pattern recognition process due to no transformation process is required. Easy and fast calculation of features provides to reduce delay which is a critical concern in control of human arm prosthetics. A wide range of time domain features have been proposed by researchers for the purpose of movement or force classification process (Phinyomark, Phukpattaranont, & Limsakul, 2012). While x_k is the k th EMG sample and N is the number of samples in each segment, the most widely used time domain features are given as follows.

Mean Absolute Value

Mean absolute value (MAV) of an EMG signal is the average of absolute value of sequential signal amplitudes. MAV is one of the mostly used features and defined as,

$$MAV = \frac{1}{N} \sum_{k=1}^N |x_k| \quad (1.1)$$

Root Mean Square

Root mean square (RMS) feature represents a calculation of amplitude modulated Gaussian

random process relating to constant force and non-fatiguing contraction. The mathematical expression of the RMS is given as,

$$RMS = \sqrt{\frac{1}{N} \sum_{k=1}^N x_k^2} \quad (1.2)$$

Willison Amplitude

Willison amplitude (WAMP) feature is the number of times the EMG signal amplitude exceeds a predefined threshold. WAMP is an indicator of motor unit action potentials (MUAP) and contraction force in muscles and can be expressed mathematically as,

$$WAMP = \sum_{k=1}^N \left[f(|x_k - x_{k+1}|) \right] \quad (1.3)$$

$$f(x) = \begin{cases} 1, & \text{if } x \geq \text{threshold} \\ 0, & \text{otherwise} \end{cases}$$

Waveform Length (WL)

Waveform length (WL) of EMG signal is the cumulative length of the waveform over the time segment. WL feature can be calculated as,

$$WL = \sum_{k=1}^{N-1} |x_{k+1} - x_k| \quad (1.4)$$

Variance of EMG (VAR)

Variance of EMG (VAR) implies the second-order moment of EMG signal and is a measure of power. VAR feature can be defined as follows,

$$VAR = \frac{1}{N-1} \sum_{k=1}^N x_k^2 \quad (1.5)$$

Simple Square Integral (SSI)

Simple square integral (SSI) of an EMG signal represents the summation of square values of EMG signal amplitude over time segment. SSI can be expressed as,

$$SSI = \sum_{k=1}^N x_k^2 \quad (1.6)$$

Zero Crossing (ZC)

Zero crossing (ZC) feature measures how many times the amplitude of EMG signal crosses zero level. Threshold value is assigned to prevent voltage fluctuations or noises effects. ZC feature calculation could be defined as,

$$ZC = \sum_{k=1}^{N-1} \left[\text{sgn}(x_k \times x_{k+1}) \cap |x_k - x_{k+1}| \geq \text{threshold} \right] \quad (1.7)$$

$$\text{sgn}(x) = \begin{cases} 1, & \text{if } x \geq \text{threshold} \\ 0, & \text{otherwise} \end{cases}$$

Integrated EMG (IEMG)

Integrated EMG refers to the summation of absolute values of EMG amplitude for each time segment. IEMG feature is also used for clinical applications and could be expressed as,

$$IEMG = \sum_{k=1}^N |x_k| \quad (1.8)$$

Frequency Domain Features

Investigation of EMG signal characteristics in frequency (or spectral) domain is mainly carried out to analyze both the fatigue phenomenon in

muscles and the motor unit recruitment (Kallenberg, Schulte, Disselhorst-Klug, & Hermensa, 2007). Various types of features have been proposed to handle EMG signal behavior in the frequency domain. While f_j , P_j and M represent a frequency value at a frequency bin j , the EMG power spectrum at a frequency bin j and the length of frequency bin, respectively, some of frequency domain features are given as follows.

Mean Frequency (MNF)

Mean frequency (MNF) is basically the calculation of average frequency dividing the sum of product of EMG power spectrum and frequency by total sum of the spectrum intensity. The mathematical expression is given as,

$$MNF = \frac{\sum_{j=1}^M f_j P_j}{\sum_{j=1}^M P_j} \quad (1.9)$$

Median Frequency (MDF)

Median frequency (MDF) of an EMG signal is the frequency at which the spectrum is partitioned into two equal amplitude. The MDF feature can be calculated as,

$$\sum_{j=1}^{MDF} P_j = \frac{1}{2} \sum_{j=1}^M P_j \quad (1.10)$$

Peak Frequency (PKF)

The frequency containing the maximum power is called peak frequency (PKF). The PKF can be calculated as follows,

$$PKF = \max(P_j) \quad (1.11)$$

Mean Power (MNP)

Average of power spectrum of EMG signals is used to determine the characteristics of signal. The feature could be expressed mathematically as,

$$MNP = \frac{1}{M} \sum_{j=1}^M P_j \quad (1.12)$$

Total Power (TTP)

The sum of the power spectrum of EMG signal reveals another feature, namely, the total power (TTP) and it is obtained as,

$$TTP = \sum_{j=1}^M P_j \quad (1.13)$$

Spectral Moments (SM)

Spectral moments (SM) is another important approach for feature extraction. Although higher order spectral moment could be calculated, the mathematical expressions of the first (SM_1) and the second (SM_2) order moments are given as follows, respectively.

$$SM_1 = \sum_{j=1}^M P_j f_j \quad (1.14)$$

$$SM_2 = \sum_{k=1}^M P_k f_k^2 \quad (1.15)$$

EMG signals can also be characterized in joint time-frequency domain (von Tscherner, 2000). In order to observe more accurate description of the signal in physical manner, EMG signals could be transformed to the area at which both frequency and time domain features exist. However, this transformation requires heavy computational costs and likely causes delay in controlling assistive devices. Main features of time-frequency

domain are Wavelet Transform (WT), Wavelet Packet Transform (WPT) and Short-time Fourier Transform (STFT).

The above mentioned signal features are needed to be classified to specify the intended motion and force production. To achieve this task, various types of classifiers such as artificial neural networks (ANN) (Arslan, Adli, Akan & Baslo, 2010), fuzzy logic (FL) (Chan, Yang, Lam, Zhang, & Parker, 2000), support vector machines (SVM) (Oskoei & Hu, 2008) and linear discriminant analysis (LDA) (Lorrain, Jiang, & Farina, 2011) are widely employed in literature.

FEATURE CLASSIFICATION AND PATTERN RECOGNITION

Extracted time or frequency domain features are required to be classified to determine the motion or applied force patterns (Oskoei & Hu, 2007). Characterization capability of prosthetic hands is firmly related to the classification performance of the selected classifier due to the classification accuracy reflects the fundamental neuromuscular activity of human muscles. The main consideration of pattern recognition progress of myoelectric signal is that each force or motion class is described by the corresponding muscle activation which is represented by a set of extracted features (Farina et al., 2014). The chosen classifier discriminates separate tasks using trial and test approaches, so that a relation between muscle activation, features and real-world tasks could be built. Thus, selection of appropriate classifier for pattern recognition process is a key issue which is expected to identify accurate patterns and to perform fast sufficiently. A great amount of literature exists to propose an optimal performance of classifiers and thereby selecting the most suitable one (Lorrain et al., 2011). The next section involves fundamental structures and applications of the widely used classifiers including ANN, FL, SVM and LD for the purpose of using in control of myoelectric based prosthetics.

Artificial Neural Networks (ANN)

ANN is an artificial intelligence method inspired by the biological structure of human brain and generally referred to as “neural networks” (Haykin, 1999). In human brain, the neural networks is the central of the decision making process. The receptors receive stimuli from the external environment and convert to electrical impulses in order to transmit them to neural nets. Then, neural nets perceive the information and make decision. Finally, the decision is transmitted the effectors to convert the impulses to response as outputs. Setting a linear or nonlinear relation between inputs and outputs, biological and artificial neural networks makes a specific decision.

Surface EMG feature classification using ANN is a popular subject among scientific studies related to control of human arm prosthetics. The basic structure of ANN which is used in pattern recognition process of EMG signals is shown in Figure 3.

The structure, namely the multilayer perceptron (MLP), which consists of a set of one input layer, one output layer, and a number of hidden layers is one of the most simple and commonly used type of ANN. Typical structure of an ANN, which is used for classification of EMG signal features, includes input, hidden, and output layers, so that the features could be related to different force or

position classes. Extracted EMG features are fed to ANN as a set of inputs and classified into different force or motion classes as the set of outputs. Each connection between neurons in neighboring layers such as input/hidden layers and hidden/output layers has a weighting factor (w). Moreover, hidden and output layer neurons implement a transfer function to make a mathematical relation between inputs and outputs. The transfer function $f(x)$ of input arrays x , which sets a relation between input and output data arrays, could be selected according to characteristics of the problem. For instance, a logistic sigmoid transfer function is given as,

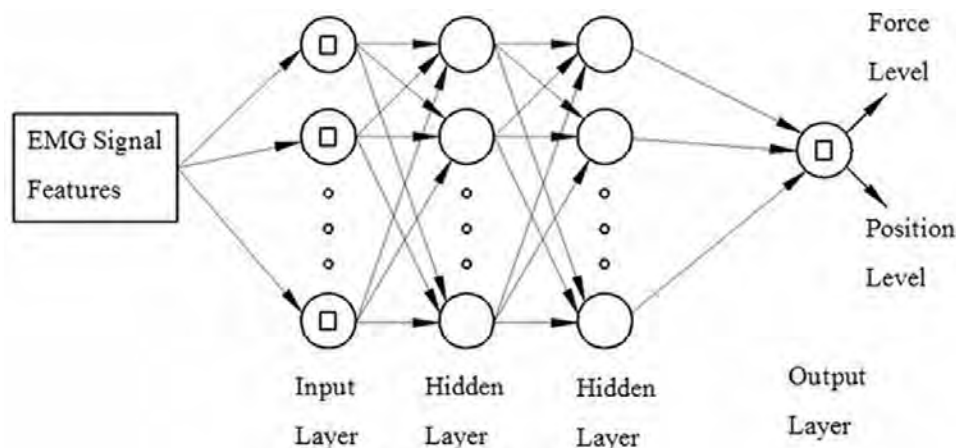
$$f(x) = \frac{1}{1 + e^{-\beta x}} \quad (2.1)$$

where

$$x = \sum_{n=1}^k w_n a_n \quad (2.2)$$

means the total input of neuron, where w_n , β and a_n are the weight, coefficient and input of i th element, respectively. Although many transfer functions are available, the most widely used transfer functions in pattern recognition of EMG

Figure 3. Schematic representation of an artificial neural network



signals are logistic sigmoid and hyperbolic tangent sigmoid transfer functions. The number of layers and neurons are adjustable based on obtained results. Using huge numbers of training data and neurons could lead to overfitting and make a complex networks structure which has to carry out more tasks and likely produces delays. To overcome such issue, some dropout techniques could be operated and train-test proportion of data is proposed to be adjusted.

Neural networks have been utilized to obtain the closest values of output for targeting the real world results by changing the weights (training stage). Adaptation of weights is implemented according to the desired results which is called supervised learning. In the cases of supervised learning of EMG signals, desired results can be position, hand/muscle force, joint torque or motion trajectory. ANN have been operated as a classifier to predict arm and joint trajectories (Cheron, Draye & Bourgeois, 1996), to estimate hand and wrist motion trajectories in the control of a virtual hand (Sebelius et al., 2005), to classify types of limbs motion (Hudgins et al., 1993), to recognize motion patterns based on signal time scale features (Zhao et al., 2006) and to predict the kinematics of shoulder and elbow (Luh, Chang, Cheng, Lai, & Kuo, 1999).

Fuzzy Logic (FL) System

FL systems are beneficial in signal processing and classification, especially for biomedical signals which are not always repeatable, and may even be conflicting (Zadeh, 1973). One of the most useful properties of fuzzy logic systems is that discrepancies in the data can be tolerated. Moreover, it

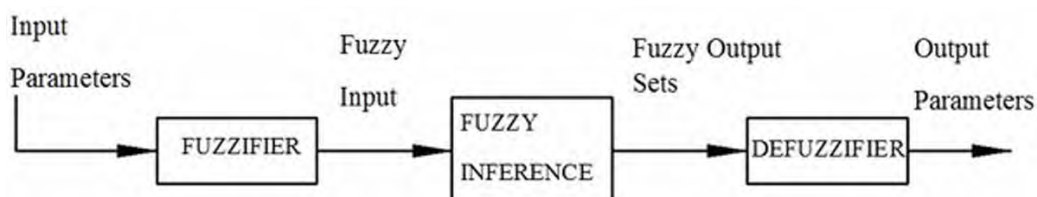
is possible to detect the patterns in data which are not easily identified by other methods using trainable fuzzy systems. Thus, the experience of medical experts or clinicians could be integrated and benefitted. It is possible to incorporate this incomplete but precious knowledge into the fuzzy logic system, due to the system's reasoning style, which is similar to that of a human being. This is a substantial advantage over the artificial neural network (ANN). Fuzzy logic systems better reflect the human decision-making ability than the ANN (Chan et al., 2000). The fundamental of a fuzzy system is the fuzzy inferring engine. Fuzzy production rules are identified according to the available knowledge or well-classified examples (Wang, 1994).

In the fuzzy method, none of operations are random. Information involving a certain amount of suspense is expressed as reliable as possible, without the deformation of forcing it into a "crisp" mold, and it is then handled in a suitable manner. Figure 4 shows the schematic representation of a fuzzy logic system with a fuzzifier and a defuzzifier phases. Fuzzy Logic Systems architecture has three main (Figure 4). These are

1. Fuzzification Module (transforms the system inputs into fuzzy sets),
2. Fuzzy Inference Engine (simulates the human decision making process by making fuzzy inference on the inputs),
3. Defuzzification Module (transforms the fuzzy sets into output parameters).

The inference engine maps each rule's fuzzy input sets into each rule's fuzzy output set. Rules have a critical influence on the performance of a

Figure 4. Schematic representation of a fuzzy logic system



FL system. The rules operate only when the inputs are applied to them.

In order to predict the rules for a fuzzy logic system, dataset is trained. To begin, a certain number of input-output training pairs are selected. The next step is to convert the training dataset into a set of fuzzy rules (IF-THEN, IF-THEN-ELSE, etc.).

The fuzzy rules are mapping from the inputs to the outputs and this mapping can be denoted quantitatively. This kind of FL system is very common and widely used in many engineering applications, such as in fuzzy logic controllers and signal processors. It is also known as fuzzy system, fuzzy controller, fuzzy model or fuzzy expert system.

In recent years, FL systems have been used for decision making process in biomechanical science (Reaz, Hussain, & Mohd-Yasin, 2006). FLS system has been performed to control the elbow and shoulder joint angles of the exoskeletons to design a controller of multifunction prosthetics (Kiguchi, Tanaka, Watanabe, & Fukuda, 2003).

Support Vector Machines

Support vector machines (SVM) is a modern and sophisticated machine learning method (Cortes & Vapnik, 1995). Since EMG-based classification process for prosthetic control problems requires high accuracy and short duration of time to obtain outputs, SVM has become a prevalent and widely used classifier (Lorrain et al., 2011). Although the main notion of the classification progress is to assign the inputs to predefined groups or categories, SVM basically separates the classes operating an optimal hyperplane. In order to discriminate the data among a vast number of classes, a combination of multiple SVM is used. SVM classification process, briefly, is described as follows (Leon, Gutierrez, Leija & Munoz, 2011),

Let x_i and y_i are inputs and outputs, respectively, for $x_i \in R^l$ and $y_i \in \{-1, 1\}^l$. The hyperplane, which divides them into two previously determined groups, is defined as,

$$w^T \phi(x) + b = 0 \quad (2.3)$$

where w and b are weight and bias parameters of hyperplane. Additionally, ϕ is a mapping function which transforms x_i vector into higher dimensional space.

For a classification case, a various number of hyperplane could separate data into two classes. However, there must be only hyperplane, which satisfies maximum margin between the classes, is defined as,

$$\min \left[\frac{1}{2} \|w\|^2 + C \sum_{i=1}^m \xi_i \right] \quad (2.4)$$

subject to $y_i(w\phi(x_i) + b) \geq 1 - \xi_i$ and $\xi_i \geq 0$ where ξ_i is the slack variables that related to error between training data. In order to obtain optimal hyperplane with limited error equation, equation (2.4) is solved, While α_i and $k(x_i, x_j)$ are Lagrange multipliers and Kernel function, respectively, the equation is reduced as follows,

$$\max \left[\sum_{i=1}^{\infty} \alpha_i - \frac{1}{2} \sum_{i,j=1}^m \alpha_i \alpha_j y_i y_j k(x_i, x_j) \right] \quad (2.5)$$

The equation of optimal hyperplane is expressed as,

$$w = \sum_{i=1}^m y_i \alpha_i \phi(x_i, x_j)$$

which satisfies

$$\sum_{i=1}^m \alpha_i y_i = 0 \text{ and } 0 \leq \alpha_i \leq C \quad (2.6)$$

The inputs x_i which satisfy $\alpha_i \neq 0$ are called support vectors. The maximization process to build decision function of the classifier is related to choose suitable kernel function which is generally selected based on inputs type and structure.

The most common used kernel functions are linear, polynomial, sigmoid and radial basis functions. The major components of SVM is shown in Figure 5.

Lucas et al. (2008) implemented the SVM method as a supervised classification of multi-channel surface electromyographic signals with the aim of controlling myoelectric prostheses. They concluded that the SVM classification rule can be effectively implemented with fast algorithms (after training) for real-time applications.

Linear Discriminant Analysis (LDA)

Linear discriminant analysis (LDA) has become a prominent classifier with the intent of grouping very complex EMG data arrays. This section summarizes, briefly, that how LDA method works. The method is based on deriving the combination of parameters that optimally discriminates the priori defined groups (Cao & Sanders, 1996). It is assumed that the vector of features is given as $X = [x_1, x_1, \dots, x_m]$. The mean values of X for the i th class are expressed as $\mu_i = [\mu_{i1}, \mu_{i2}, \dots, \mu_{im}]$. The main procedure of LDA method is to maximize

the following function which is known as linear discrimination or gate function (Kim, Choi, Moon, & Mun, 2011),

$$f_i(x) = x^T S^{-1} \mu_i - \frac{1}{2} \mu_i^T S^{-1} \mu_i + \log(\pi_i) \quad (2.7)$$

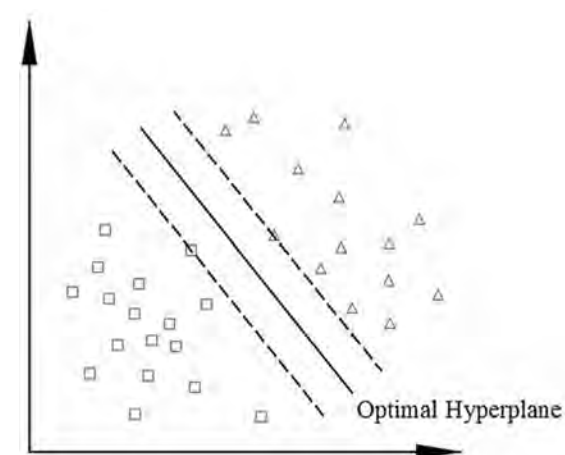
where S is the pooled covariance matrix of input data and π_i is a prior probability of inputs coming from class i . Using the combinations of the equation, misclassification could be minimized by obtaining higher likelihood index for each defined class.

LDA method has been applied to identify EMG signals to discriminate the patterns of EMG linear envelope of healthy subjects and patients with anterior cruciate ligament injury (Alkan & Gunay, 2012) and to classify the features to enhance the controllability of a powered prosthetics (Hargrove, Scheme, Englehart, & Hudgins, 2010).

It is necessary to make an extensive evaluation of the performance of these listed classifiers in pattern recognition process besides shortcomings and merits of them. Arslan et al. (2010) employed ANN to predict externally applied forces to human hands using EMG signal features. The study, which was aimed to estimate the forces accurately, showed that the classifier predicted the targeted force values in a range of 0.34 and 0.05, and of 0.24 and 0.09 root mean square difference (RMSD) for isometric and anisometric contraction experiments, respectively. In this study it was clearly stated that ANN method could built a successful non-linear relation between force and EMG signal features. However, the authors highlighted that it is not possible to propose a standard ANN design for training the EMG signal efficiently. Even though ANN is a powerful classifier, the absence of a conceptualized and standardized training method represents an important disadvantage.

Chan et al. (2000) performed a fuzzy logic based classification procedure to control prosthetics. They also provided a comparison between ANN and fuzzy systems in pattern recognition process. In the study, it was shown that 8% and

Figure 5. The basic representation of support vector machine. Optimal hyperplane is represented with solid line which divides optimal margin between two classes. Square and triangle on dashed lines are support vectors.



11.3% of error rates were obtained by Fuzzy and ANN classifiers in pattern recognition, respectively. Some advantages of fuzzy systems were listed as *i)* slightly higher recognition rate than obtained by ANN, *ii)* insensitivity to overtraining, and *iii)* consistent outputs demonstrating higher reliability. The main drawback of the method was noted that requiring more human intervention at initialization stage in order to get the minimum inter-class cross-over. Hence it was stated that the procedure is not automatic to the same extent as ANN.

LDA is becoming a prominent tool for pattern recognition in EMG studies. Chu et al. (2007) conducted a study which includes an EMG feature discrimination process. After a real-time pattern-recognition progress, it was shown that the proposed method achieves the recognition accuracy rate of 97.4%. Phinyomark et al. (2013) also reported that LDA shows a better performance in the classification of fluctuating EMG signals compared to several classifiers such as quadratic discriminant analysis (QDA), random forests (RFs) and k-nearest neighbor (KNN).

The performance of SVM as a machine learning method is needed to be assessed. Subasi (2012) compared the performance of a group of classifiers and reported that classical SVM method provided 96.75% accuracy, while the kNN and the radial basis function (RBF) classifiers achieved the process with 95.17% and 94.08% accuracy, respectively. In the study, it was also noted that SVM performance could be enhanced with some modifications.

FUTURE RESEARCH DIRECTIONS

Though rapid improvements of assistive technologies and specifically artificial human hand prosthetics have been observed, challenging problems are still remaining to be solved in terms of EMG signal pre- and post-processing operations with the intention of providing dexterous control of prosthetics.

Feature Selection

Features extracted from EMG signals for both in time and frequency domains are used at present. However, new features may be proposed for better representation of EMG signals. Furthermore, effect of sliding windows for calculation of features should be investigated in detail. Used features presently are calculated by means of differentiation or summation of neighboring EMG signal amplitudes. New approaches such as measuring energy consumption for each EMG signal maybe employed for time domain in future studies, as well.

Classification Methods

The classifier performance is investigated by researchers extensively. Selection of training and test data for discrimination process has a great influence on classification accuracy and operation duration. Selection of the optimum cross-validated EMG signal arrays should be a purpose of next studies. Additionally, using the combinations of time and frequency domain features for training and test may provide higher accuracy of classification. Furthermore, semi-supervised learning, which is one of the fundamental aspects of classification methodology for cases that it is hard to obtain sufficient training data, should be studied in more detail.

CONCLUSION

The chapter provides a general overview on the EMG signal features and their classification methodologies which are critical issues for controlling of human arm prosthetics. EMG-driven human arm prosthetics are highly sensitive to the scientific and technological advances. Through the last decades, many of EMG signal features calculation and discrimination methods have been proposed and applied to prostheses. Precise and intuitive control of prosthetics depends mainly upon the type of extracted feature and classifica-

tion techniques. Making a significant difference or advancing the dexterity in the control of prosthetic devices depend on achieving the optimum signal feature and classifier architecture. Needless to say that in addition to the control structure, mechanical structure of the prosthetics also plays a major role in the completion of complex motor tasks which deserves to be extensively dealt with in a separate report.

REFERENCES

- Alkan, A., & Gunay, M. (2012). Identification of EMG signals using discriminant analysis and svm classifier. *Expert Systems with Applications*, 39(1), 44–47. doi:10.1016/j.eswa.2011.06.043
- Arslan, Y. Z., Adli, M. A., Akan, A., & Baslo, M. B. (2010). Prediction of externally applied forces to human hands using frequency content of surface EMG signals. *Computer Methods and Programs in Biomedicine*, 98(1), 36–44. doi:10.1016/j.cmpb.2009.08.005 PMID:19762107
- Basmajian, J. V., & DeLuca, C. J. (1985). *Muscles alive: Their functions revealed by electromyography*. Baltimore, MD: The Williams & Wilkins Company.
- Cao, J., & Sanders, D. B. (1996). Multivariate discriminant analysis of the electromyographic interference pattern: Statistical approach to discrimination among controls, myopathies and neuropathies. *Medical & Biological Engineering & Computing*, 34(5), 369–374. doi:10.1007/BF02520008 PMID:8945863
- Chan, F. H. Y., Yang, Y. S., Lam, F. K., Zhang, Y. T., & Parker, P. A. (2000). Fuzzy EMG classification for prosthesis control. *IEEE Transactions on Rehabilitation Engineering*, 8(3), 305–311. doi:10.1109/86.867872 PMID:11001510
- Cheron, G., Draye, J. P., Bourgeois, M., & Libert, G. (1996). A dynamic neural network identification of electromyography and trajectory relationship during complex movements. *IEEE Transactions on Bio-Medical Engineering*, 43(5), 552–558. doi:10.1109/10.488803 PMID:8849468
- Chu, J. U., Moon, I., Lee, Y. J., Kim, S. K., & Mun, M. S. (2007). A supervised feature-projection-based real-time EMG pattern recognition for multifunction myoelectric hand control. *IEEE/ASME Transactions on Mechatronics*, 12(3), 282–290. doi:10.1109/TMECH.2007.897262
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297. doi:10.1007/BF00994018
- Farina, D., Jiang, N., Rehbaum, H., Holobar, A., Graimann, B., Dietl, H., & Aszmann, O. (2014). The extraction of neural information from the surface EMG for the control of upper limb prostheses: Emerging avenues and challenges. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 22(4), 797–809. doi:10.1109/TNSRE.2014.2305111 PMID:24760934
- Hargrove, L. J., Scheme, E. J., Englehart, K. B., & Hudgins, B. S. (2010). Multiple binary classifications via linear discriminant analysis for improved controllability of a powered prosthesis. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 18(1), 49–57. doi:10.1109/TNSRE.2009.2039590 PMID:20071277
- Haykin, S. (1999). *Neural Networks: A Comprehensive Foundation*. Englewood Cliffs, NJ: Prentice-Hall.
- Hudgins, B., Parker, P., & Scott, R. N. (1993). A new strategy for multifunction myoelectric control. *IEEE Transactions on Bio-Medical Engineering*, 40(1), 82–94. doi:10.1109/10.204774 PMID:8468080

- Kallenberg, L. A. C., Schulte, E., Disselhorst-Klug, C., & Hermensa, H. J. (2007). Myoelectric manifestations of fatigue at low contraction levels in subjects with and without chronic pain. *Journal of Electromyography and Kinesiology*, 17(3), 264–274. doi:10.1016/j.jelekin.2006.04.004 PMID:16790358
- Kiguchi, K., Tanaka, T., Watanabe, K., & Fukuda, T. (2003). Exoskeleton for human upper-limb motion support, *IEEE International Conference on Robotics and Automation*, 2206 – 2211.
- Kim, K. S., Choi, H. H., Moon, C. S., & Mun, C. W. (2011). Comparison of k-nearest neighbor, quadratic discriminant and linear discriminant analysis in classification of electromyogram signals based on the wrist-motion directions. *Current Applied Physics*, 11(3), 740–745. doi:10.1016/j.cap.2010.11.051
- Leon, M., Gutierrez, J. M., Leija, L., & Munoz, R. (2011). EMG pattern recognition using support vector machines classifier for myoelectric control purposes. *Pan American Health Care Exchanges (PAHCE)*, 175–178.
- Lorrain, T., Jiang, N., & Farina, D. (2011). Influence of the training set on the accuracy of surface EMG classification in dynamic contractions for the control of multifunction prostheses. *Journal of Neuroengineering and Rehabilitation*, 8(1), 25. doi:10.1186/1743-0003-8-25 PMID:21554700
- Lucas, M. F., Gaufriau, A., Pascual, S., Doncarli, C., & Farina, D. (2008). Multi-channel surface EMG classification using support vector machines and signal-based wavelet optimization. *Biomedical Signal Processing and Control*, 3(2), 169–174. doi:10.1016/j.bspc.2007.09.002
- Luh, J. J., Chang, G. C., Cheng, C. K., Lai, J. S., & Kuo, T. S. (1999). Isokinetic elbow joint torques estimation from surface EMG and joint kinematic data: Using an artificial neural network model. *Journal of Electromyography and Kinesiology*, 9(3), 173–183. doi:10.1016/S1050-6411(98)00030-3 PMID:10328412
- Mann, R. W., & Reimers, S. D. (1970). Kinesthetic sensing for the EMG controlled Boston Arm. *IEEE Transactions on Man-Machine Systems*, 11(1), 110–115. doi:10.1109/TMMS.1970.299971
- Oskoei, M. A., & Hu, H. (2007). Myoelectric control systems-a survey. *Biomedical Signal Processing and Control*, 2(4), 275–294. doi:10.1016/j.bspc.2007.07.009
- Oskoei, M. A., & Hu, H. (2008). Support vector machine-based classification scheme for myoelectric control applied to upper limb. *IEEE Transactions on Bio-Medical Engineering*, 55(8), 1956–1965. doi:10.1109/TBME.2008.919734 PMID:18632358
- Otto Bock Gmb. (2012). Retrieved from <http://www.ottobock.com/>
- Phinyomark, A., Phukpattaranont, P., & Limsakul, C. (2012). Feature reduction and selection for EMG signal classification. *Expert Systems with Applications*, 39(8), 7420–7431. doi:10.1016/j.eswa.2012.01.102
- Phinyomark, A., Quaine, F., Charbonnier, S., Serviere, C., Tarpin-Bernard, F., & Laurillau, Y. (2013). EMG feature evaluation for improving myoelectric pattern recognition robustness. *Expert Systems with Applications*, 40(12), 4832–4840. doi:10.1016/j.eswa.2013.02.023
- Reaz, M. B. I., Hussain, M. S., & Mohd-Yasin, F. (2006). Techniques of EMG signal analysis: Detection, processing, classification and applications. *Biological Procedures Online*, 8(1), 11–35. doi:10.1251/bpo.115 PMID:19565309
- Resnik, L. (2010). Research update: VA study to optimize DEKA arm. *Journal of Rehabilitation Research and Development*, 47(3), ix–x. doi:10.1682/JRRD.2010.03.0034 PMID:20665342

Sebelius, F., Eriksson, L., Holmberg, H., Levins-son, A., Lundborg, G., Danielsen, N., & Montelius, L. et al. (2005). Classification of motor commands using a modified self-organising feature map. *Medical Engineering & Physics*, 27(5), 403–413. doi:10.1016/j.medengphy.2004.09.008 PMID:15863349

Shanghai Kesheng Prosthese Corporation Ltd. (2012). Retrieved from <http://www.keshen.com/index-en.asp>

Subasi, A. (2012). Classification of EMG signals using combined features and soft computing techniques. *Applied Soft Computing*, 12(8), 2188–2198. doi:10.1016/j.asoc.2012.03.035

von Tscharnner, V. (2000). Intensity analysis in time-frequency space of surface myoelectric signals by wavelets of specified resolution. *Journal of Electromyography and Kinesiology*, 10(6), 433–445. doi:10.1016/S1050-6411(00)00030-4 PMID:11102846

Wang, L. X. (1994). Adaptive fuzzy systems and control. *Design and Stability Analysis*, 29–31.

Wehner, M. (2012). Man to machine, applications in electromyography. In M. Schwartz (Ed.), *EMG Methods for Evaluating Muscle and Nerve Function* (pp. 427–454). InTech. doi:10.5772/26495

Zadeh, L. A. (1973). Outline of a new approach to the analysis of complex systems and decision process. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-3(1), 28–44. doi:10.1109/TSMC.1973.5408575

Zecca, M., Micera, S., Carrozza, M. C., & Dario, P. (2002). Control of multifunctional prosthetic hands by processing the electromyographic signal. *Critical Reviews in Biomedical Engineering*, 30(4–6), 459–485. doi:10.1615/CritRevBiomedEng.v30.i456.80 PMID:12739757

Zhao, J., Xie, Z., Jiang, L., Cai, H., Lio, H., & Hirzinger, G. (2006). EMG control for a five-fingered interactuated prosthetic hand based on wavelet transform and sample entropy. *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*.

KEY TERMS AND DEFINITIONS

Assistive Technology: A branch of technology is used to regain the lost functions of human body parts.

Feature Classification: A pattern recognition technique that is used to categorize a huge number of data into different classes.

Feature Extraction: A method to obtain meaningful and clear data of a signal.

Human Arm Prostheses: Assistive devices which enable to perform lost functions of human arm due to upper or lower arm amputations.

Pattern Recognition: A machine learning process which identifies the pattern of physical systems using data belong to investigated systems.

Rehabilitation: A series of therapy to make injured or amputated people regained lost skills or functions.

Surface Electromyography: A type of electromyography signal recording method carrying out by means of adhering electrodes to skin surface.

The Principle and Process of Digital Fabrication of Biomedical Objects

B**S. H. Choi***The University of Hong Kong, Hong Kong***H. H. Cheung***The University of Hong Kong, Hong Kong***W. K. Zhu***The University of Hong Kong, Hong Kong*

INTRODUCTION

Mounting pressure of market globalization and intensifying competition has ferociously been driving the manufacturing industry to survive on incessant reductions in cost and lead-time. However, conventional manufacturing methods can no longer satisfy increasingly diverse customer demands, tight cost control, and complex new products.

Against this background, much research efforts have been devoted to developing various technologies to help the manufacturing industry, and layered manufacturing (LM, or now often called 3D printing) and virtual reality (VR) simulation have been among the most significant technologies developed over the past couple of decades.

Despite recent proliferation of LM (3D printing) for free-form fabrication, most of the current systems can only fabricate objects of a single material (Wohlers Report, 2013). There are imminent demands for multi-material layered manufacturing (MMLM) processes to fabricate advanced products and biomedical objects comprising of multiple materials. A few experimental MMLM machines have been developed based on the conventional single material LM systems for relatively simple objects (Bellini, 2002; Wang & Shaw, 2006; Wachsmuth, 2008; Li et al, 2009; Åklint et al, 2013). However, their fabrication

speed is unsatisfactory for most complex, large products or medical objects for emergency cases.

More recently, many researchers have worked on virtual prototyping and virtual manufacturing (VPM) (Bracht and Masurat, 2006; Wang and Li, 2006), which is regarded as one of the most important technological advancements for product design and development. VPM has been successfully used in ship-building and car industries (Kim et al, 2002; Wöhlke and Schiller, 2005). It uses simulation techniques to analyze and improve a product design and validate the fabrication processes and production schedules.

Through simulations in a VR environment, key factors such as the product shape, manufacturability, and durability that may affect the profitability of manufactured products are optimized. VPM enhances profitability by reducing production cost and material usage, etc. Moreover, it reduces time and tooling cost by eliminating the need for multiple physical prototypes. This allows the users to review and validate a product design to “get it right the first time” for delivery of quality products to market on time and within budget.

This chapter describes the principle of virtual prototyping and virtual manufacturing, with a focus on the processes of modeling and subsequent digital fabrication of multi-material biomedical objects. Case studies of modeling and digital fabrication of biomedical objects using a multi-

material virtual prototyping and manufacturing (MMVPM) system will be presented to demonstrate its principle and possible applications in biomedical engineering.

BACKGROUND

There has been a huge surge in demand for biomedical objects in recent years for various medical and dental purposes (Khan & Dickens, 2014; Lee et al., 2001; Maji et al., 2014; Pinnock et al., 2016; Ripley et al., 2016; Sanghera et al., 2001; Winder et al., 1999).

Biomedical objects have been traditionally used as prostheses to repair damaged bone structures or to replace missing body parts (D'Urso et al., 2000; Eufinger et al., 1995; Sannomiya et al., 2008). They are now commonly used by medical students, surgeons, and dentists to help study the intricate anatomical details of human organs and bone structures, as well as to facilitate planning of implantations and surgical procedures (Singare et al., 2009). For example, artificial hip joints, and bone and jaw structures are often used in hospitals to assist complex medical operations. In addition, they are used as specimens for experiments in pharmaceutical manufacturing enterprises.

Depending on the required properties and applications, biomedical objects can be made of either homogeneous (single) material, or heterogeneous (discrete multiple) materials, or functionally graded materials (FGM) (Pompe et al., 2003; Sun et al., 2005). Watari et al. (2004) described the fabrication of an FGM dental implant by powder metallurgy. Experimental results showed that the implant could achieve better mechanical properties and biocompatibility, and that it could control the tissue response through the gradient function of FGM.

But over the years, biomedical objects have been getting more complex, both geometrically and structurally, with more intricate internal details and delicate material variations. As such, most biomedical objects are not economical, and

very often not possible, to make by the traditional manufacturing processes.

To fabricate biomedical objects of multiple materials, some researchers have explored the MMLM technology, which is a layer-by-layer additive process that fabricates a heterogeneous object of a number of different materials, which can be discrete with distinctive boundary interfaces, or functionally graded with composition gradients changing gradually from one to another. This process requires a computer-aided design (CAD) model with sufficient material information (Gu and Li, 2002; Gupta et al., 2015; Jafari et al. 2000; Sun et al., 2005).

MMLM, however, remains experimental, and its practical application for fabricating biomedical objects is limited. Indeed, most current MMLM systems can only handle relatively small, simple objects of few materials. More importantly, they are slow and expensive to operate. To address the limitations of MMLM, research efforts have recently focused on developing multi-material virtual prototyping and manufacturing (MMVPM) technology for digital fabrication of complex biomedical objects in a convenient and cost-effective manner.

MMVPM provides a digital platform that integrates virtual reality (VR) simulation technique with MMLM processes. It is an effective tool that can digitally fabricate complex biomedical objects for use in lieu of physical ones. The user can model a complex biomedical object and digitally fabricate its prototypes. Subsequently, the resulting digital biomedical prototypes can be visualized and analyzed in a VR environment, as if the user is manipulating physical objects, for some medical and dental purposes.

The following sections describe the workflow of the MMVPM system in detail. Case studies of modeling and subsequent digital fabrication of biomedical objects using MMVPM would be presented to demonstrate the principle and process of digital fabrication of complex objects for possible applications in biomedical engineering.

THE MMVPM SYSTEM FOR DIGITAL FABRICATION OF BIOMEDICAL OBJECTS

The MMVPM system consists of a suite of software packages for design and visualization of discrete and functionally graded multi-material biomedical objects and simulation of MMLM process in a VR environment. These packages include a color modeler for coloring monochrome STL models, a slicing algorithm, a topological hierarchy-sorting algorithm for grouping random slice contours, a topological hierarchy-based toolpath planning algorithm for generation of sequential and concurrent multi-toolpaths, and a virtual prototyping and virtual manufacturing (VPM) module for digital fabrication of the multi-material biomedical objects (Choi and Cheung, 2005, 2007, 2009; Choi and Kwok, 2004).

Figure 1 summarizes the three major processes for digital fabrication of biomedical objects, namely:

1. Data acquisition for model generation of biomedical objects;
2. Model processing, contour slicing, and toolpath planning for digital fabrication;
3. Digital fabrication, visualization and analysis of biomedical objects.

Data Acquisition for Model Generation of Biomedical Objects

To acquire data for generation of a biomedical model, a target anatomical structure is firstly scanned by a medical imaging device, like CT, μ -CT, and MRI, to generate 2D image slices. The resulting 2D slice images in grayscale are then imported into a 3D construction software, such

Figure 1. The main processes of digital fabrication of biomedical objects

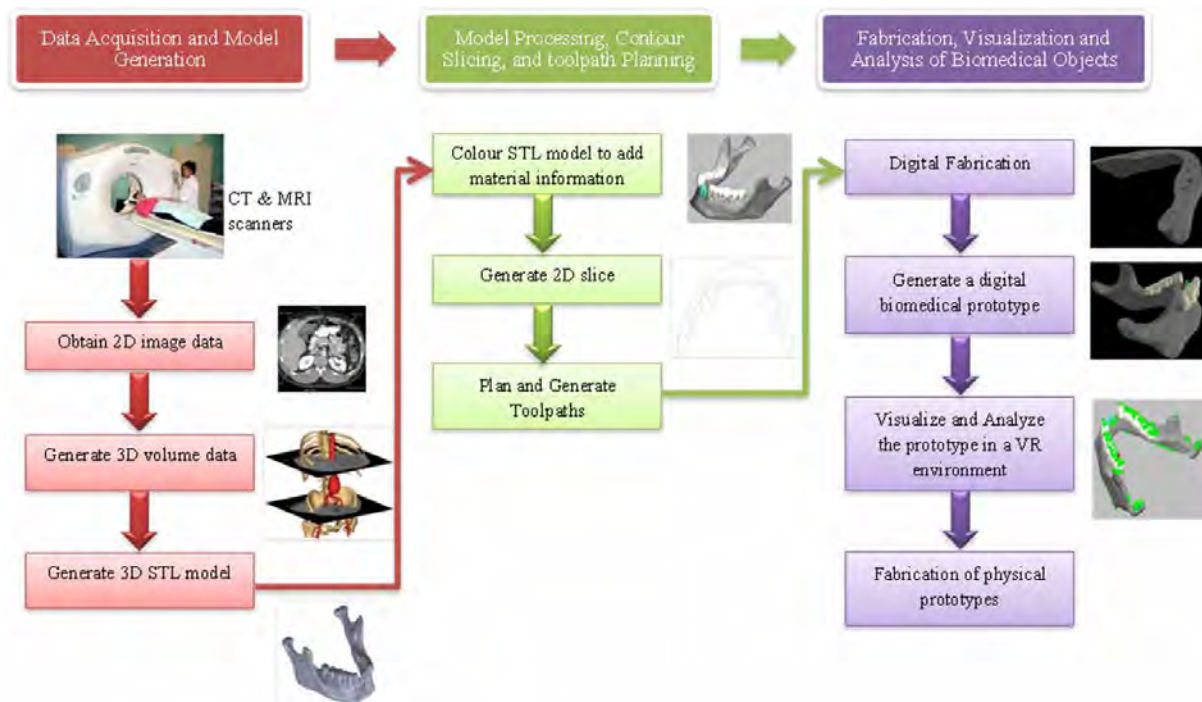
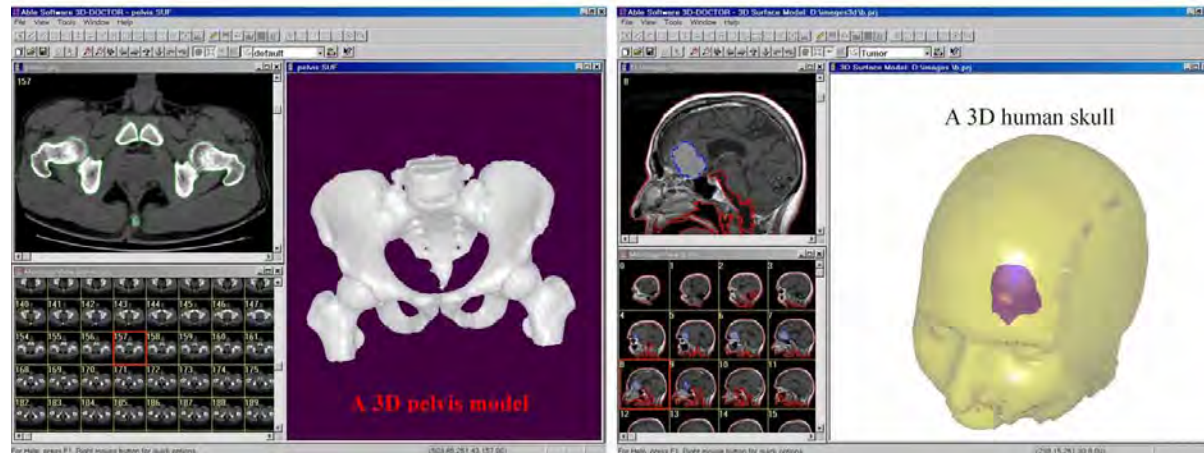


Figure 2. Generation of a 3D pelvis model and a 3D human skull model from 2D image slices by 3D-DOCTOR

Source: Able Software Corp, 2011



as Materialise Mimics and Able Software Corp 3D-DOCTOR, to extract characteristic data points and subsequently generate the required biomedical model, normally in STL format, which is the de-facto industry standard for layered manufacturing (Sun et al., 2005; Wang et al., 2010). Figure 2 shows two screen captures of the 3D-DOCTOR generating a 3D pelvis model and a 3D human skull model from respective 2D slice images (Able Software Corp, 2016).

Model Processing, Contour Slicing, and Toolpath Planning for Digital Fabrication

The resulting biomedical STL model is then processed in the MMVPM system to add material information, generate slice contours, and plan toolpaths for subsequent digital fabrication and visualization of the resulting biomedical object, as follows:

1. An in-house package is used to paint the biomedical STL model (which is originally monochrome of a single material), with each color representing a specific material;
2. The color STL model is sliced into a number of layers of a predefined thickness. The

resulting slice contours and material information are stored in a modified Common Layer Interface (CLI) file;

3. The slice contours are sorted with a contour sorting algorithm to establish explicit topological hierarchy;
4. Based on the hierarchy information, a multi-toolpath planning algorithm is used to plan and generate toolpaths for concurrent deposition of materials by hatching the slice contours with a predefined hatch space. The hatch vectors are stored in the modified CLI file for fabrication of biomedical object and build-time estimation.

Digital Fabrication, Visualization and Analysis of Biomedical Objects

The toolpaths together with the hatch vectors are subsequently input to a VP module to perform digital fabrication of the biomedical object. The resulting digital biomedical object can be visualized and analyzed, as if manipulating a physical one, to review and improve the design of the biomedical objects conveniently. Although the digital biomedical object can be displayed on a general computer monitor or projection screen, it is preferable to visualize and analyze it in a VR

environment to take advantage of the stereoscopic views and immersive feelings.

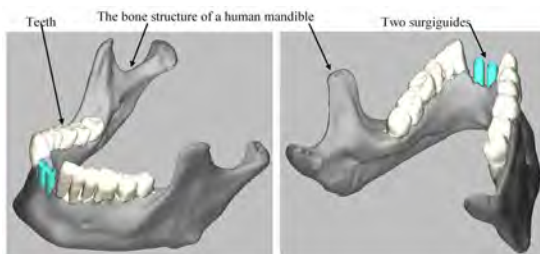
The following section uses a human mandible to be inserted with surgiguides to demonstrate how the MMVPM system can model and fabricate multi-material objects for biomedical applications, such as surgical training and planning, patient's education, and implantations.

Figure 3 shows a multi-material assembly of a human mandible inserted with two surgiguides. To ensure the required implant precision, virtual models of the human mandible will be useful to simulate and study possible treatment plans. The MMVPM system can be used to assist surgeons to study a treatment plan through stereoscopic visualization and analysis of a human mandible prototype with surgiguides, as shown in Figure 4.

The surgiguides facilitate positioning the implant accurately. Based on the treatment plan resulted from the simulation, surgeons can improve the design of the surgiguides to fit the patient's bone structure. Figure 5 shows the complete process of digital fabrication of the mandible prototype, while its quality can be visualized accordingly in Figure 6.

It can be seen in Figure 6 that areas of dimensional deviations scattered mainly around the teeth and the bottom of the mandible bone, while the surgiguides are relatively smooth. This is considered satisfactory as we are more concerned with the accuracy of the surgiguides.

Figure 3. A color STL model of a human mandible with two surgiguides

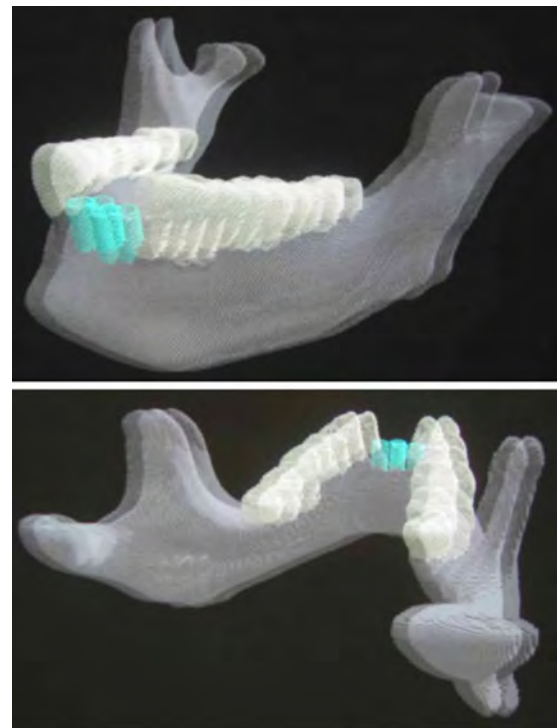


Nevertheless, the simulation process can be iterated with a new combination of process parameters to improve the accuracy of the surgiguides, if necessary. Therefore, based on stereoscopic visualization and deviation analysis, satisfactory designs can be enhanced and validated for subsequent fabrication of dental implants in a convenient and cost-effective way.

To achieve best surgical results possible, an implant should preferably be made of functionally graded materials with properties that would mimic the biological and mechanical characteristics of human organs or tissues (Krishna et al., 2008; Lin et al., 2009; Watari et al., 2004). The MMVPM system can design and digitally fabricate FGM objects to assist such purposes.

For example, it would be desirable to have dental implants made of functionally graded materials, such as titanium (Ti) and hydroxyapatite (HAP), to satisfy both mechanical and bio-compatible

Figure 4. Stereoscopic visualization of a human mandible prototype with surgiguides



properties. Figure 7 shows the resulting FGM dental implant, with material variation represented by blending of red (100% HAP) and green (100% Ti) colors along the Z-axis. Figure 8 shows the resulting FGM teeth.

CASE STUDIES

To further demonstrate possible use of the MMVPM system for modeling and digital fabrication of biomedical objects to facilitate complex surgical procedures, a human ear model and a human thorax model are illustrated in this section.

Case Study One: A Human Ear Model

Hearing loss is one of the common diseases of human ears. It may be caused by genetic problems, harmful noises from working environments, and ageing, etc. (McCullagh, 2011). Some possible clinical approaches include cochlear implant and tympanum (eardrum) repair (Huber & Kipman, 2011). For this purpose, a digitally fabricated human ear model would be particularly cost-effective and convenient to help surgeons diagnose the disease and plan the possible surgical operations. Therefore, to make one, a monochrome human ear STL model is firstly painted with an in-house package to highlight various parts of the model, as shown in Figure 9.

Figure 5. Digital fabrication of the human mandible prototype with surgiguides

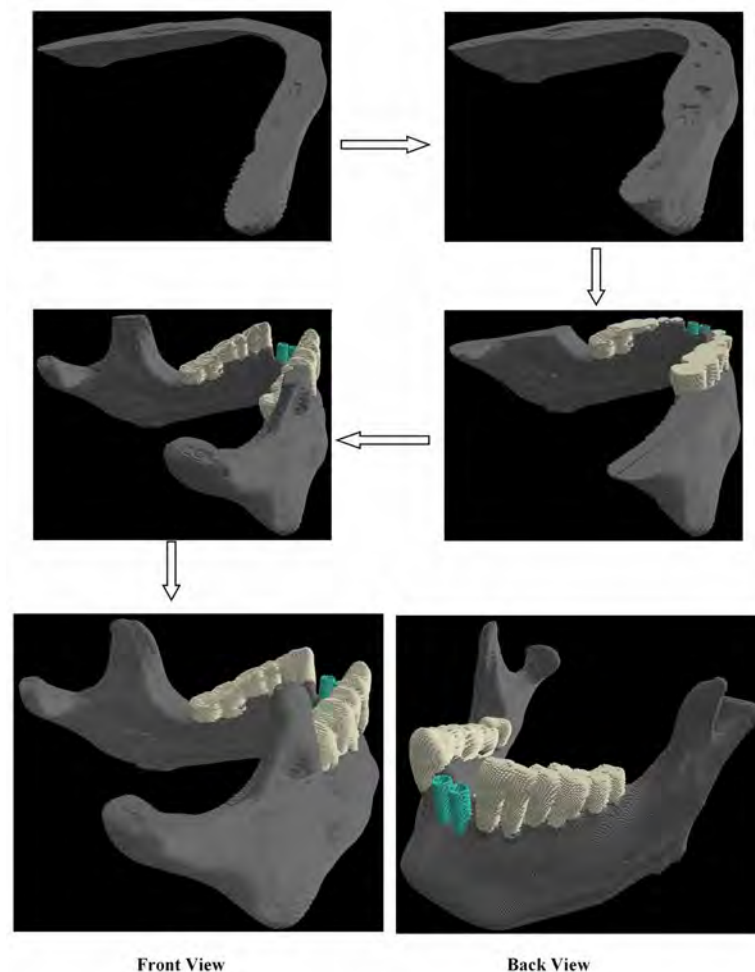


Figure 6. Areas of the human mandible prototype with dimensional deviations beyond design limits

B

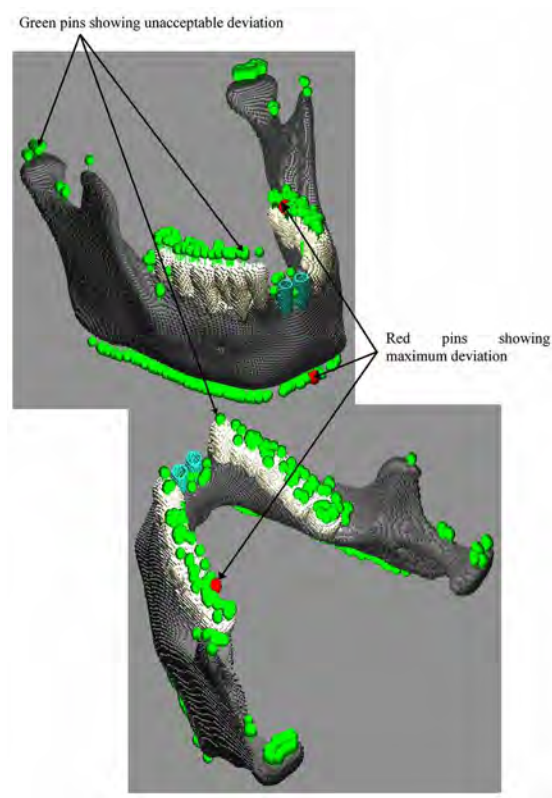


Figure 7. A functionally graded dental implant generated by the MMVPM system

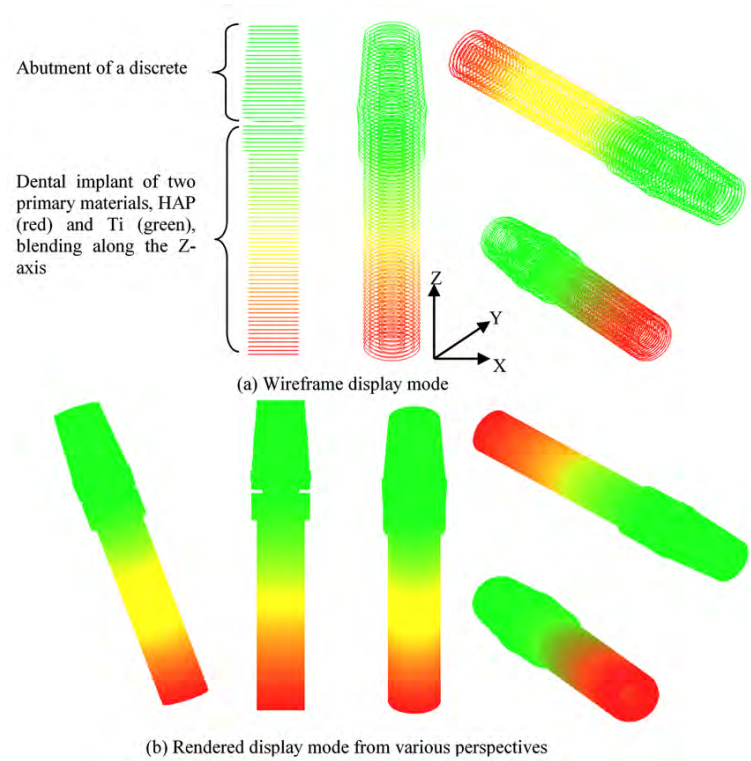


Figure 8. FGM teeth

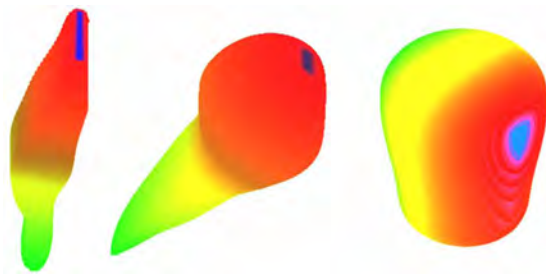


Figure 9. A color STL model of an anatomical human ear

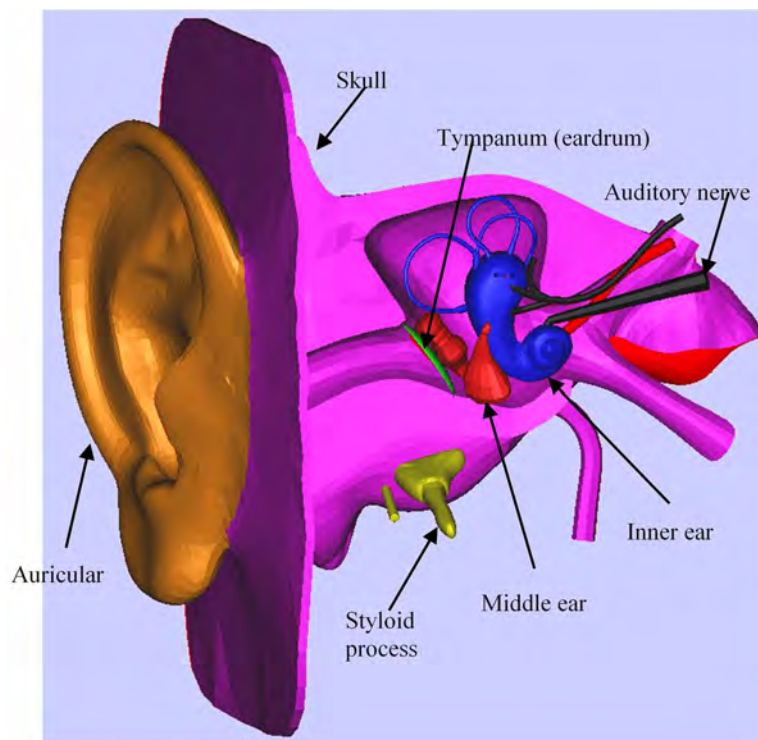
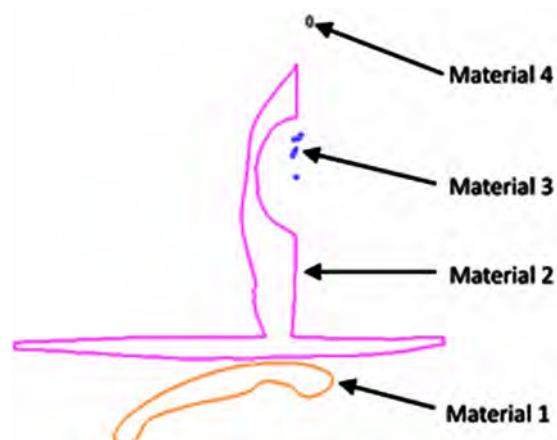


Figure 10. A slice of the ear model consisting of four materials



Based on the geometric and material information provided by the color STL model, the ear model is sliced into a number of layers with a predefined layer thickness. Figure 10 shows the slice contours of a particular layer to be made of four materials colored in orange, pink, blue, and grey. The contours of each layer are then sorted to obtain explicit topological hierarchy information for planning and generation of toolpaths for concurrent deposition of materials. Subsequently, a human ear prototype with discrete multi-materials can be digitally fabricated using the VPM module. Figure 11 shows some stages of the digital fabrication process and the resulting discrete multi-material prototype of the human ear model. Indeed, a digital multi-material prototype of a human ear would be particularly useful for study and planning of delicate surgeries, in that they can differentiate clearly one part from another, or tissues from blood vessels of the ear.

Case Study Two: A Human Thorax With an Intervertebral Disc Spacer

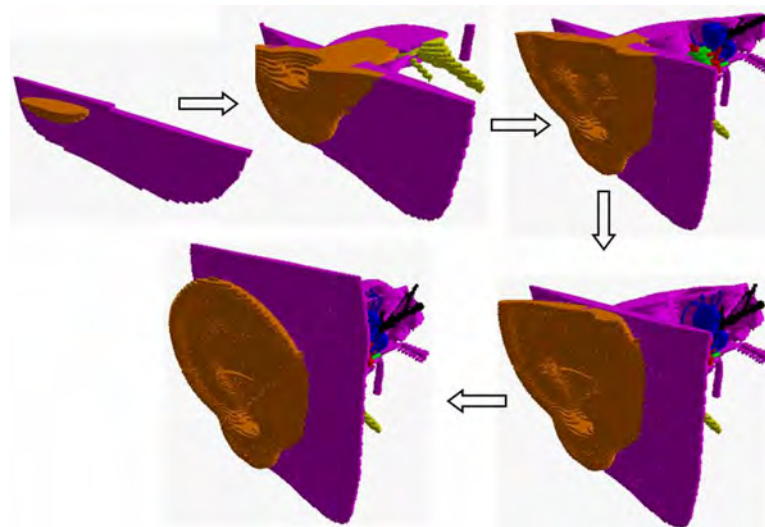
Figure 12 shows a color STL model of a human thorax, which consists mainly of three parts: the ribs, the sternum, and the spine. Based on the

geometric and material information provided by the color STL model, the MMVPM system can slice the thorax model into a number of layers, one of which is shown in Figure 13. The layer contours are then sorted to obtain explicit topological hierarchy information for planning and generation of toolpaths for concurrent deposition of materials. Subsequently, a thorax prototype with discrete multi-materials can be digitally fabricated using the VP module. Figure 14 shows parts of the digital fabrication process.

The resulting digital thorax model would be particularly useful for studying related symptoms like back pain, which is among the frequently reported musculoskeletal problems. The most common factor causing low back pain is the degeneration of intervertebral discs. Taksali et al. (2004) reported that back pain can result from irritation of the surrounding sinuvertebral nerves when the nucleus extrudes into annular tears.

To alleviate this problem, it may be necessary to replace the degenerated disc with an artificial intervertebral disc spacer to remove the primary pain generator while preserving functional motion. As shown in Figure 15, an artificial intervertebral disc spacer has to be inserted between two adjacent vertebrae. For this purpose, it would be

Figure 11. Digital fabrication of a prototype of the human ear model



medically preferable to make such a disc spacer with materials having functional and biocompatible properties similar to a normal disc.

The artificial intervertebral disc is to be composed of three components, namely the nucleus pulposus, the annulus fibrosus, and the vertebral endplates, as shown in Figure 16. In comparison with the material property of the annulus fibrosus, the nucleus pulposus is softer while the endplates are harder.

Parsons et al. (1992) invented a molding method to fabricate a functional and biocompatible intervertebral disc spacer. The disc spacer consisted of:

1. A central core of a biocompatible elastomer with a hardness in the range of 20-70 shore-A, shaped and sized to approximate the nucleus pulposus of a natural intervertebral disc, and
2. An outer ring of elastomeric material with a hardness in range of 40–80 shore-A, surrounding the central core to approximate the size and shape of a natural annulus fibrosus.

The endplates comprised a biocompatible elastomer with a hardness in range of 90-100 shore-A, and affixed, one to each end, to the outer ring / central core. The relative size of the central core and the thickness of the outer ring in the radial direction, and the selection of material hardness, could be varied in order to better match the mechanical properties of the composite prosthesis with that of the normal disc. The disc spacer could thus achieve the desired properties by varying the hardness of the elastomeric material along the circumferential and the Z-axis directions of the spacer, respectively, by changing the composition of the elastomeric material.

It is therefore desirable to design and fabricate an FGM intervertebral disc spacer with a gradual change of compositions of the elastomeric material to achieve the desired hardness properties in order to mimic the normal disc in an intervertebral disc prosthesis.

The MMVPM system can also be a practical tool to model an FGM intervertebral disc spacer with a functional gradation in hardness property,

Figure 12. A color STL model of a human thorax

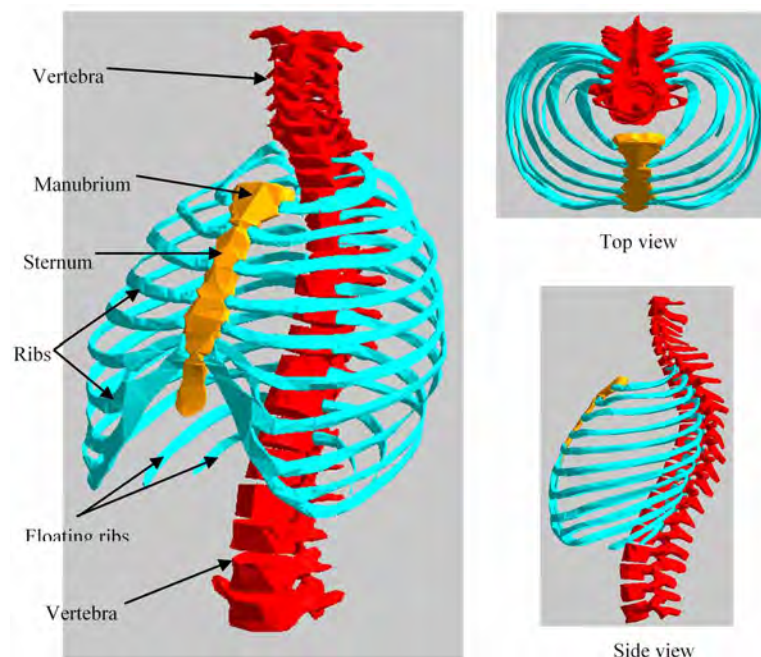
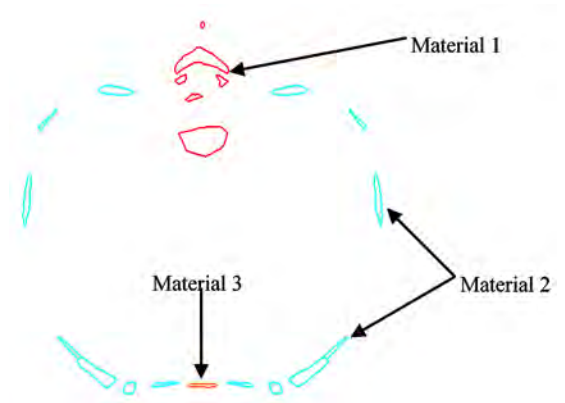


Figure 13. A slice of the thorax model consisting three materials



and then build a digital prototype to facilitate visualization and subsequent fabrication.

To represent the material variations of the intervertebral disc spacer, it is firstly sliced into a set of homogeneous layer contour-based model, and explicit topological hierarchy information is then built for each layer. Secondly, feature layers are selected for assigning primary materials and material control functions for calculation of property values of material composition. Thirdly, each layer is discretized into sub-regions of constant material composition.

Figure 14. Digital fabrication of a prototype of the human thorax model

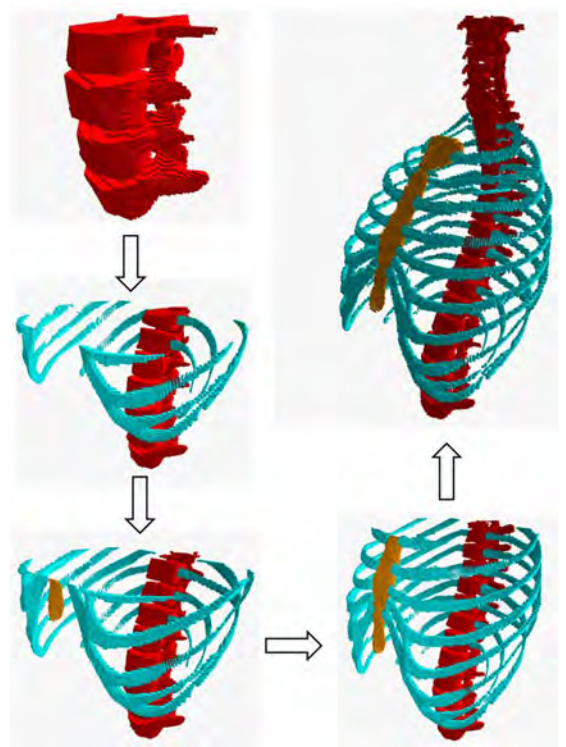


Figure 17 shows the resulting FGM intervertebral disc spacer with 3D variations of material composition. Along the Z-axis, it has a material variation represented by blending of blue (100

Figure 15. An artificial intervertebral disc spacer inserted between two adjacent vertebrae

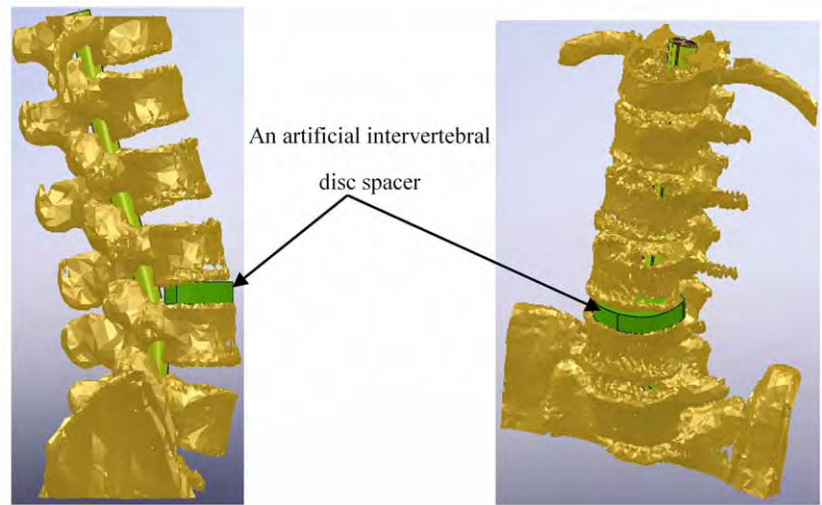
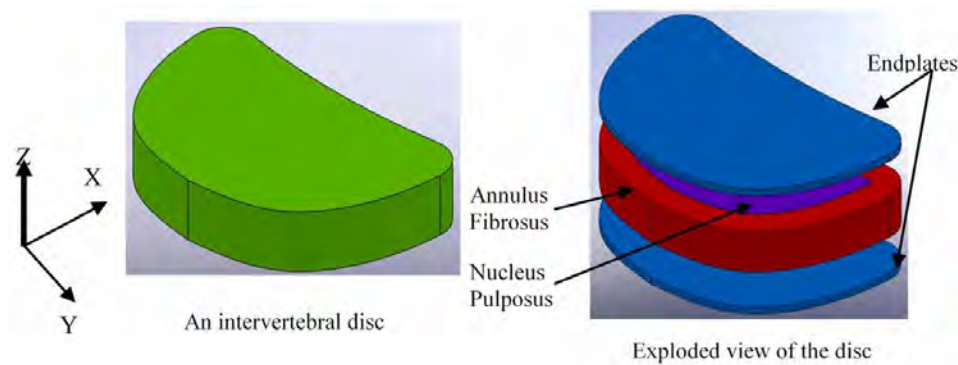


Figure 16. The structure of an intervertebral disc



shore-A elastomer) and red (60 shore-A elastomer) colors; at the same time, there is material variation of each layer in the X-Y plane.

Now, the resulting contour-based FGM intervertebral disc spacer model has both geometric and material information to be processed for visualization of internal material variation of each layer,

multi-toolpath planning, and subsequent digital fabrication. Figure 18 shows various stages of the digital fabrication process of an FGM intervertebral disc spacer prototype. The MMVPM system allows adjustment of the resolution of material composition to suit practical visualization and fabrication requirements, simply by changing

Figure 17. An FGM intervertebral disc spacer with material variations along the Z-axis and in the X-Y plane

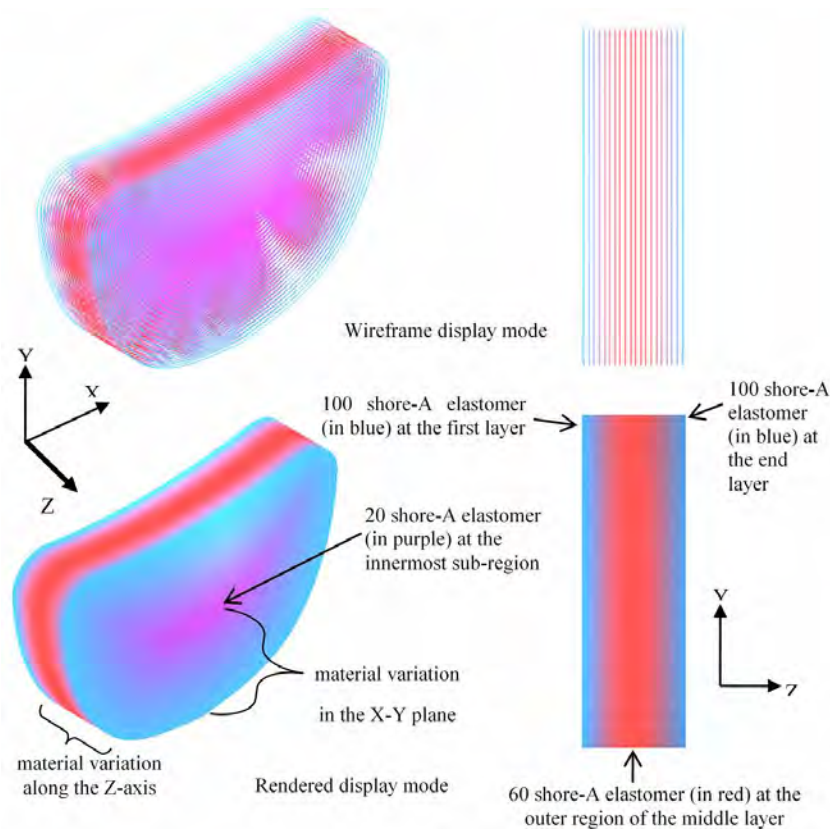
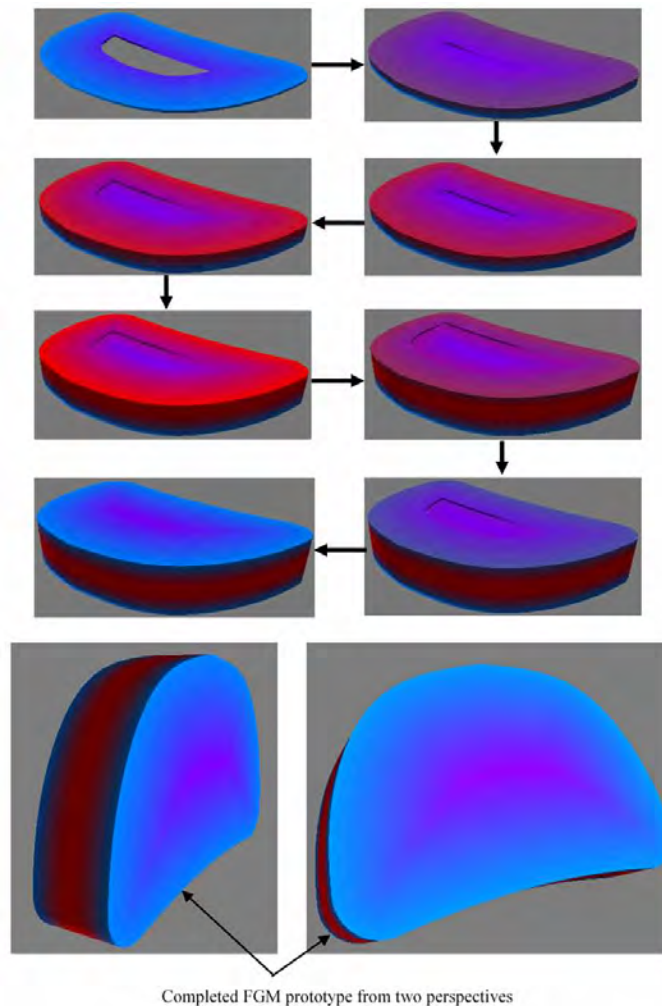


Figure 18. Digital fabrication of an FGM intervertebral disc spacer prototype



the discretization of layer contours, which is the number of layers and the number of sub-regions. Therefore, it is a practical tool for modeling and digital fabrication of biomedical objects with FGM and discrete materials.

FUTURE RESEARCH DIRECTIONS

Although the MMVPM system presented above provides a convenient digital platform that integrates VR simulation technique with MMLM processes for virtual prototyping and virtual manufacturing of complex multi-material objects

for advanced product development and biomedical applications, it is limited by its local connectivity and there is a lack of cloud computing capability for remote processing.

Therefore, a worthwhile future development of the MMVPM system is to enhance the algorithms of its main modules for integration with cloud-based computing technology to facilitate remote image processing and visualization of digital fabrication processes. As such, users who are geographically separated can participate in the discussion and sharing of digital fabrication of complex multi-material objects simultaneously for different applications.

CONCLUSION

This chapter described the workflow of a multi-material virtual prototyping and manufacturing (MMVPM) system, which is a digital platform that integrates VR simulation technique with MMLM processes. Case studies of modeling and digital fabrication of biomedical objects using the MMVPM system were presented to demonstrate its principle and possible applications in biomedical engineering. The case studies showed that the MMVPM system is a convenient and cost-effective tool that can digitally fabricate complex biomedical objects for use in lieu of physical ones. The user can model a complex biomedical object and perform fabrication simulation. Subsequently, the resulting digital biomedical object can be visualized and analyzed in a VR environment conveniently for some medical and dental purposes.

ACKNOWLEDGMENT

The authors would like to acknowledge the Committee on Research and Conference Grants (CRCG) of the University of Hong Kong for partial financial support for this project.

REFERENCES

- Able Software Corp. 3D-DOCTOR. (2016). Retrieved June 28, 2016, from <http://ablesw.com/3d-doctor/images.html#SurfaceRendering>
- Åklint, T., Carlström, E., Johander, P., & Stiernstedt, J. (2013, March 7). Layered manufacturing of free-form multi-material micro-components. *Patent, WO2013030064*, A1.
- Bellini, A. (2002). *Fused deposition of ceramics: a comprehensive experimental, analytical and computational study of material behavior, fabrication process and equipment design* (Ph.D thesis). Drexel University.
- Bracht, U., & Masurat, T. (2006). The Digital Factory between vision and reality. *Computers in Industry*, 56(4), 325–333. doi:10.1016/j.comp-ind.2005.01.008
- Choi, S. H., & Cheung, H. H. (2005). A Multi-Material Virtual Prototyping System. *Computer Aided Design*, 37(1), 123–136. doi:10.1016/j.cad.2004.06.002
- Choi, S. H., & Cheung, H. H. (2007). Multi-Material Virtual Prototyping for Rapid Product Development and Biomedical Engineering. *Computers in Industry*, 53(5), 557–566.
- Choi, S. H., & Cheung, H. H. (2009). A Topological Hierarchy-Based Approach to Layered Manufacturing of Functionally Graded Multi-Material Objects. *Computers in Industry*, 60(5), 349–363. doi:10.1016/j.compind.2009.01.008
- Choi, S. H., & Kwok, K. T. (2004). A Topological Hierarchy-Sorting Algorithm for Slice Contours. *Rapid Prototyping Journal*, 10(2), 98–113. doi:10.1108/13552540410526971
- DUrso, P. S., Earwaker, W. J., Barker, T. M., Redmond, M. J., Thompson, R. G., & Effeney, D. J. (2000). Custom Cranioplasty Using Stereolithography and Acrylic. *British Journal of Plastic Surgery*, 53(3), 200–204. doi:10.1054/bjps.1999.3268 PMID:10738323
- Eufinger, H., Wehmöller, M., Harders, A., & Heuser, L. (1995). Prefabricated Prostheses for the Reconstruction of Skull Defects. *Internal Journal of Oral and Maxillofacial Surgery*, 24(1), 104–110.
- Gu, P., & Li, L. (2002). Fabrication of Biomedical Prototypes with Locally Controlled Properties Using FDM. *CIRP Annals – Manufacturing Technology*, 51(1), 181–184.
- Gupta, V., & Tandon, P. (2015). Heterogeneous Object Modelling with Material Convolution Surfaces. *Computer Aided Design*, 62, 236–247. doi:10.1016/j.cad.2014.12.005

- Huber, M., & Kipman, U. (2011). The Mental Health of Deaf Adolescents with Cochlear Implants Compared to Their Hearing Peers. *International Journal of Audiology*, 53(3), 146–154. doi:10.3109/14992027.2010.533704 PMID:21309643
- Jafari, M. A., Han, W., Mohammadi, F., Safari, A., Danforth, S. C., & Langrana, N. (2000). A Novel System for Fused Deposition of Advanced Multiple Ceramics. *Rapid Prototyping Journal*, 6(3), 161–175. doi:10.1108/13552540010337047
- Khan, M., & Dickens, P. (2014). Selective Laser Melting (SLM) of Pure Gold for Manufacturing Dental Crowns. *Rapid Prototyping Journal*, 20(6), 471–479. doi:10.1108/RPJ-03-2013-0034
- Kim, H., Lee, J. K., Park, J. H., Park, B. J., & Jang, D. S. (2002). Applying digital manufacturing technology to ship production and the maritime environment. *Integrated Manufacturing Systems*, 13(5), 295–305. doi:10.1108/09576060210429748
- Krishna, B. V., Xue, W. C., Bose, S., & Bandyopadhyay, A. (2008). Functionally Graded Co-Cr-Mo Coating on Ti-6Al-4V Alloy Structures. *Acta Biomaterialia*, 4(3), 697–706. doi:10.1016/j.actbio.2007.10.005 PMID:18054298
- Lee, S. J., Jung, I. Y., Lee, C. Y., Choi, S. Y., & Kum, K. Y. (2001). Clinical Application of Computer-Aided Rapid Prototyping for Tooth Transplantation. *Dental Traumatology*. *Munksgaard*, 17(3), 114–119. doi:10.1034/j.1600-9657.2001.017003114.x PMID:11499760
- Li, L., Saedan, M., Feng, W., Fuh, J., Wong, Y., Loh, H., & Lu, L. et al. (2009). Development of a multi-nozzle drop-on-demand system for multi-material dispensing. *Journal of Materials Processing Technology*, 209(9), 4444–4448. doi:10.1016/j.jmatprotec.2008.10.040
- Lin, D., Li, Q., Li, W., Zhou, W. L., & Swain, M. V. (2009). Design Optimization of Functionally Graded Dental Implant for Bone Remodeling. *Composites. Part B, Engineering*, 40(7), 668–675. doi:10.1016/j.compositesb.2009.04.015
- Maji, P. K., Banerjee, A. F., Banerjee, P. S., & Karmakar, S. (2014). Additive Manufacturing in Prosthesis Development – A Case Study. *Rapid Prototyping Journal*, 20(6), 480–489. doi:10.1108/RPJ-07-2012-0066
- McCullagh, M. C. (2011). Effects of a Low Intensity Intervention to Increase Hearing Protector Use among Noise-exposed Workers. *American Journal of Industrial Medicine*, 54(3), 210–215. doi:10.1002/ajim.20884 PMID:20721900
- Parsons, J. R., Le, C. K., Langrana, N. A., Clemow, A. J., & Chen, E. H. (1992). *Functional and Biocompatible Intervertebral Disc Spacer Containing Elastomeric Material of Varying Hardness*. United States Patent.
- Pinnock, C. B., Meier, E. M., Joshi, N. J., Wu, B., & Lam, M. T. (2016). Customizable Engineered Blood Vessels using 3D Printed Inserts. *Methods (San Diego, Calif.)*, 99, 20–27. doi:10.1016/j.ymeth.2015.12.015 PMID:26732049
- Pompe, W., Worch, H., Epple, M., Friess, W., Gelinsky, M., Greil, P., & Schulte, K. et al. (2003). Functionally Graded Materials for Biomedical Applications. *Materials Science and Engineering A*, 362(1-2), 40–60. doi:10.1016/S0921-5093(03)00580-X
- Report, W. (2013). *State of the industry, Annual worldwide progress report*, Wohlers Associates Inc. United States of America.
- Ripley, B., Kelil, T., Cheezum, M. K., Goncalves, A., Carli, M. F. D., Rybicki, F. J., & Blankstein, R. et al. (2016). 3D Printing based on Cardiac CT Assists Anatomic Visualization prior to Transcatheter Aortic Valve Replacement. *Journal of Cardiovascular Computed Tomography*, 10(1), 28–36. doi:10.1016/j.jcct.2015.12.004 PMID:26732862
- Sanghera, B., Naique, S., Papaharilaou, Y., & Amis, A. (2001). Preliminary Study of Rapid Prototype Medical Models. *Rapid Prototyping Journal*, 7(5), 275–284. doi:10.1108/13552540110410486

Sannomiya, E. K., Brito, A. A., & Agelieri, F. (2008). Surgical Planning for Resection of an Ameloblastoma and Reconstruction of the Mandible Using a Selective Laser Sintering 3D Biomodel. *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontics*, 106(1), 36–40. doi:10.1016/j.tripleo.2008.01.014 PMID:18585610

Singare, S., Lian, Q., Wang, W. P., Wang, F., Liu, Y. X., Li, D. C., & Lu, B. H. (2009). Rapid Prototyping Assisted Surgery Planning and Custom Implant Design. *Rapid Prototyping Journal*, 15(1), 19–23. doi:10.1108/13552540910925027

Sun, W., Starly, B., Nam, J., & Darling, A. (2005). Bio-CAD Modeling and its Applications in Computer-Aided Tissue Engineering. *Computer Aided Design*, 37(11), 1097–1114. doi:10.1016/j.cad.2005.02.002

Taksali, S. M. D., Jonathan, N., Grauer, M. D., Alexander, R., & Vaccaro, M. D. (2004). Material Considerations for Intervertebral Disc Replacement Implants. *The Spine Journal*, 4(11), 231–238. doi:10.1016/j.spinee.2004.07.012 PMID:15541671

Wachsmuth, J. (2008). *Multiple Independent Extrusion Heads for Fused Deposition Modeling. Master of Science*. Virginia Polytechnic Institute and State University.

Wang, C. S., Wang, W. H. A., & Lin, M. C. (2010). STL Rapid Prototyping Bio-CAD Model for CT Medical Image Segmentation. *Computers in Industry*, 61(3), 187–197. doi:10.1016/j.compind.2009.09.005

Wang, J. W., & Shaw, L. L. (2006). Fabrication of functionally graded materials via inkjet color printing. *Journal of the American Ceramic Society*, 89(10), 3285–3289. doi:10.1111/j.1551-2916.2006.01206.x

Wang, Q. J., & Li, J. R. (2006). Interactive visualization of complex dynamic virtual environments for industrial assemblies. *Computers in Industry*, 57(4), 366–377. doi:10.1016/j.compind.2005.11.002

Watari, F., Yokoyama, A., Omori, M., Hirai, T., Kondo, H., Uo, M., & Kawasaki, T. (2004). Biocompatibility of Materials and Development to Functionally Graded Implant for Bio-medical Application. *Composites Science and Technology*, 64(6), 893–908. doi:10.1016/j.compsci-tech.2003.09.005

Winder, J., Cooke, R. S., Gray, J., Fannin, T., & Fegan, T. (1999). Medical Rapid Prototyping and 3D CT in the Manufacture of Custom Made Cranial Titanium Plates. *Journal of Medical Engineering & Technology*, 23(1), 26–28. doi:10.1080/030919099294401 PMID:10202700

Wöhlke, G., & Schiller, E. (2005). Digital Planning Validation in automotive industry. *Computers in Industry*, 56(4), 393–405. doi:10.1016/j.compind.2005.01.010

KEY TERMS AND DEFINITIONS

Biomedical Object: An object made of bio-compatible materials for prosthetic or bioengineering applications.

Digital Fabrication: Graphical simulation of a fabrication process to make an object in digital form.

Multi-Material Layered Manufacturing: An additive manufacturing process for fabrication of a multi-material object layer by layer.

Multi-Material Object: An object made of a number of heterogeneous (discrete) materials or functionally graded materials.

Multi-Material Virtual Prototyping and Virtual Manufacturing: Graphical simulation of a multi-material layered manufacturing process to make a multi-material object in digital form.

Reverse Engineering in Rehabilitation

B**Emilia Mikołajewska***Nicolaus Copernicus University, Poland***Marek Macko***Kazimierz Wielki University, Poland***Zbigniew Szczepański***Kazimierz Wielki University, Poland***Dariusz Mikołajewski***Kazimierz Wielki University, Poland*

INTRODUCTION

Reverse engineering is relatively novel technology, which may revolutionize clinical practice in rehabilitation. This technology may constitute next step toward patient-tailored therapy, providing customized medical products increasing effectiveness and accessibility of rehabilitation procedures and decreasing cost of manufacturing and time of delivery. Such opportunities need separate research, assessment of associated threats, and dedicated solutions.

In this chapter authors investigate the extent to which the available opportunities in the area of application of reverse engineering in rehabilitation are being exploited, including own concepts, studies and observations.

BACKGROUND

The additive manufacturing (AM), called also 3D printing or stereolithography is relatively novel technology developing since 1980s. It constitutes iterative technology based on construction the real objects layer by layer, translating this way digital file (digitized object) into a solid object. Features of such object depend on technology and material used to print, but number of both of them rapidly

increases, providing important alternative for traditional manufacturing techniques. Moreover some objects have unique features (e.g. shapes) not comparable with products of traditional manufacturing.

Recent editorial article by Maruthappu & Keogh paid particular attention to potential of 3D printing applications to transform healthcare technologies and organization. Authors divided possible healthcare applications of additive manufacturing into three main groups:

- Internet as decentralised store of blueprints (drugs, equipment, devices, and even body parts) for Early patients-tailored interventions, much quicker and cheaper than traditional delivery solutions,
- Patient-tailored therapy based on medical imaging combined with 3D printing,
- Engineering of 3D printed tissues (Maruthappu & Keogh, 2014; Murphy & Atala, 2014; Seol et al., 2014).

Further implementation of reverse engineering needs additional interdisciplinary research (including randomized controlled trials on patients where available), dedicated methodology, careful assessment of opportunities and threats as far as dedicated solutions.

Reverse Engineering as A Complex Process

Reverse engineering is regarded as quick and cost-effective method of creating functional or nonfunctional copies of existing objects. Process of reverse engineering for rehabilitation purposes is unified to several subsequent stages covered by semi-automated process:

- Digital acquisition of the 3D geometric data: directly from the patient or based on his/her medical records (e.g. using computed tomography – CT or magnetic resonance imaging – MRI),
- Modification/adaptation procedures,
- Creation of 3D model or final product on 3D printer and control of its feasibility: material features, shape, dimensions, patient comfort, etc.

Reverse Engineering for Rehabilitation Support Purposes

Rehabilitation aims at restoration of patient's functions to the maximum possible degree. Scientists and clinicians are aware that in such person the full capacity available in healthy people can not be achievable despite efforts of patient, therapists and caregivers. Increasing number of severely ill, disabled and elderly people makes this task even more difficult. Thus there is still need for novel solutions increasing effectivity of the current rehabilitation procedures.

Rehabilitation should be common, early, comprehensive, and continuous. Reverse engineering products are beneficial for a wide variety of applications in the area of rehabilitation providing:

- Short track between measurements in particular patient and final ready-to-use patient-tailored product (assistive technology device, drug, artificial organ),

- The same quality, geometric accuracy and shape reconfiguration possibilities as original product (e.g. in the case of replacement),
- Features modification possibility (according to the current patient's need) by therapists/manufacture,
- Low price thanks to cheaper commercial technologies of 3D printing, lack of transport and storage costs,
- Cost-effective production of customized products.

Reverse engineering can provide cells based bones or soft tissues for modelling, testing and therapy purposes thanks to the use of bio-ink (composition of cells and hydrogel materials). Moreover assistive technology products can be cheaper and patient-tailored. Use of such solutions can provide higher therapy efficacy, life quality in patients with severe disorders, usually associated with long-term disability. The huge breakthrough in neurology, neurosurgery and neurorehabilitation can cause novel nerve repair technique: microstructured scaffolds to promote nerves regeneration (Chang et al., 2008; Zhu et al., 2008).

SOLUTIONS AND RECOMMENDATIONS

There is need for further development of interdisciplinary collaboration to identify possible threats and limitations emerging from novel applications of the reverse engineering in everyday clinical practice. Current limitations in the area of reverse engineering cover following topics:

- Ethical issues concerning use of living tissue to print artificial organs, and possible tissue modifications,

- Legal issues including copyright laws to printed copies, patents and original-based hybrid solutions,
- Lack of commonly approved standards of quality,
- Lack of quality and development of internet-based blueprints supervision,
- Lack of control on printed medical devices, especially in poor and developing countries,
- More complex emerging issues concerning eugenics and military applications (Maruthappu & Keogh, 2014).

Novel imaging techniques, printing techniques and software encouraging reverse engineering will constitute challenge for medical education system, due to required specialistic knowledge and experience. Graduate and postgraduate education of medical staff should incorporate novel standards of interdisciplinary co-operation, within patient-tailored therapy. Local scientific and clinical authorities should incorporate modified evidence-based clinical guidelines, paying particular attention to possible secondary changes in patients.

Important limitation constitutes variability of materials and associated techniques of 3D printing. Current materials for 3D printers are limited to plastic, metal, ceramics and living cells. Material engineering faces to biomaterials with features desired for surgical handling, encouraging tissue reintegration, anti-allergic, water-resistant, non-fragile or even biodegradable materials. Natural materials (wood, leather) may be hard to replace. Also combining various materials (e.g. metal and plastic) within one printable product still constitutes a challenge. Increasing use of E-health technologies can change face of the healthcare in the future.

Traditional manufacturing technologies (e.g. welded) in selected cases can provide better solutions than 3D printed. Thus key element of further development in the area of reverse engineering is integration of 3D printed objects with existing approaches.

Despite aforementioned problems many successful products have resulted from this approach:

B

- Surgical management and strategy of rehabilitation of acetabular fracture showed accurate and safer surgical management, shorter postoperative recovery time and significantly decreases costs (Deng et al., 2014, Klein et al., 2013),
- Full-arch prostheses of edentulous mandibles (De Santis et al., 2013),
- Customised foot orthoses (Telfer et al., 2013),
- Customized Gensingen braces (Bibb & Brown, 2000; Weiss, 2010),
- Dental technologies (Shahmiri et al., 2014; Yuan et al., 2013; Giordano et al., 2012; Leijnse & Spoor, 2012),
- Urological practice (Youssef et al, 2015),
- Various other easy to use and effective assistive technology solutions (Watanabe et al., 2015).

Rapid development of hardware and software allows for expectations concerning novel technologies in the area of 2D to 3D images conversion as far as rapid reconstruction of the anatomical parts based on CT and MRI. There is need for effective and accurate, but relatively simple, quick and cheap solutions dedicated for medical applications. Errorless (even semi-automatic) reproducing of natural affected anatomy can significantly decrease time of recovery in patients after severe injuries.

Development of stem cell based 3D printing can make another breakthrough not only in everyday clinical practice (e.g. reduce mortality in burns or vascular diseases) but also in medical ethics and law regulations concerning transplantation procedures. Lack of physical person (donor) and possibility to print artificial organ as copy of original organ can positively affect public awareness and public opinion playing an important role in increasing organ donation. Clinicians should be aware that it may take a lot

of years to print some complex organs as heart and avoid complications.

Another challenge remains integration with other future therapies like nanomedicine, nanorobotics and bioMEMS. There is need for closest co-operation of whole interdisciplinary research teams.

FUTURE RESEARCH DIRECTIONS

Important directions for further research are multifunctional designs that combine geometric and material complexity, such as self-folding 3D printed structures, actuated in a heating environment (Deng & Chen, 2015; Gao et al., 2013; Ishida et al., 2014). Further research on bio-inspired design can help understand the mechanisms underlying many associated cognitive processes (Fu et al. 2014), as well as results of the Human Brain Project (Rose, 2014; Markram, 2012).

Current evidences concerning use of reverse engineering in rehabilitation are regarded as insufficient. There is strong need for further research exploring novel area of applications, e.g. semi-natural bones and tissues, associated with bone kinematics and dynamics simulation (Fang et al., 2015; Wei & Dong, 2011; He et al., 2014). Our own research concern three main topics:

- Use of various technologies and materials to print 3D patient-tailored assistive technology (Lei et al., 2014),
- Research on mechanical properties of printed objects,
- Easy acquiring and modification of existing patients' electronic health records, including medical imaging, as an useful source of templates for reverse engineering.

Further challenge constitutes development of family of novel reverse engineering tools (including user-friendly software) for clinical staff - allowing for easy adjustment of the digital models/patterns to the needs of the particular pa-

tient (Macko et al., 2016). Industrial 3D printers and associated technologies may reach extremely high resolution, but everyday use of 3D biomaterial printers and 3D desktop printers is limited. Moreover cost-effect factors associated with aforementioned solutions are hard to compare. Use of bio-ink has still many limitations: limited number of bio-plotting materials, limited resolution and strength, complex technology, timing of gelation time, degradation kinetics, and influence of byproducts (even biodegradable). Simultaneous printing using multiple biofactors and materials is hard to achieve. But the room temperature of processing and homogenous distribution of cells makes bio-printing the most useful and promising technology do far.

Even attempts of low-cost fabrication the soft prostheses with the help of a desktop 3D printer were recently reported (He et al., 2014), but every novel medical technology needs careful studies to prove its safety for patients and therapists, and avoid possible complications and secondary changes (even as long-term effects). On-demand production of drugs must provide patient safety (e.g. dose precision, drug features and effects) and quality similar to the drugs manufactured using traditional technologies (Weismann et al. 2015).

Quick development of 3D printing and reverse engineering creates novel solutions and challenges every day: 95.8% of plastic surgeons want CT- or MRI-derived models for their patients (especially as low-cost solutions for facial transplantation and reconstruction purposes). There is common belief that use of 3D printed models and reverse engineering should be incorporated into graduate and postgraduate medical education. This way preoperative planning, developing intraoperative guidance tools, producing patient-specific prosthetics, and teaching patients and surgeons in everyday clinical practice may be easier, and post-operative comfort and esthetics can be optimized (Chia & Wu, 2015; AlAli et al., 2015; Chae et al., 2015; Jones et al., 2015, Chi & Kim, 2015; Wu & Hsu, 2015; Gerstle et al., 2014).

CONCLUSION

Proposed approach—reverse engineering -enables us to design and manufacture relatively complex objects from its original structures or digitized images. It poses intriguing research questions to its further applications in rehabilitation and possible novel research. It can make key contribution to safe human health and improve the quality of life of patients with functional deficits. Such powerful solution needs for careful further research on larger samples, follow-up studies, and more advanced analytical measurements and technical assistance.

REFERENCES

- AlAli, A. B., Griffin, M. F., & Butler, P. E. (2015). Three-Dimensional Printing Surgical Applications. *Eplasty*, 15, e37. PMID:26301002
- Bibb, R., & Brown, R. (2000). The application of computer aided product development techniques in medical modelling topic: Rehabilitation and prostheses. *Biomedical Sciences Instrumentation*, 36, 319–324. PMID:10834252
- Chae, M. P., Rozen, W. M., McMenamin, P. G., Findlay, M. W., Spychal, R. T., & Hunter-Smith, D. J. (2015). Emerging Applications of Bedside 3D Printing in Plastic Surgery. *Frontiers in Surgery*, 2, 25. doi:10.3389/fsurg.2015.00025 PMID:26137465
- Chang, W. C., Kliot, M., & Sretavan, D. W. (2008). Microtechnology and nanotechnology in nerve repair. *Neurological Research*, 10(10), 1053–1062. doi:10.1179/174313208X362532 PMID:19079980
- Chia, H. N., & Wu, B. M. (2015). Recent advances in 3D printing of biomaterials. *Journal of Biological Engineering*, 9(1), 4. doi:10.1186/s13036-015-0001-4 PMID:25866560
- Choi, J. W., & Kim, N. (2015). Clinical application of three-dimensional printing technology in craniofacial plastic surgery. *Archives of Plastic Surgery*, 3(3), 267–277. doi:10.5999/aps.2015.42.3.267 PMID:26015880
- De Santis, R., Gloria, A., Russo, T., DAMora, U., Varriale, A., Veltri, M., & Ambrosio, L. et al. (2014). Reverse engineering of mandible and prosthetic framework: Effect of titanium implants in conjunction with titanium milled full arch bridge prostheses on the biomechanics of the mandible. *Journal of Biomechanics*, 16, 3825–3829. doi:10.1016/j.jbiomech.2014.10.020 PMID:25468297
- Deng, A., Xiong, R., He, W., Wei, D., & Zeng, C. (2014). Postoperative rehabilitation strategy for acetabular fracture: Application of 3D printing technique. *Nan Fang Yi Ke Da Xue Xue Bao*, 4, 591–593. PMID:24752117
- Deng, D., & Chen, Y. (2015). Origami-based self-folding structure design and fabrication using projection based stereolithography. *Journal of Mechanical Design*, 2(2), 021701–021712. doi:10.1115/1.4029066
- Fang, A., Zheng, M., & Fan, D. (2015). Application status of rapid prototyping technology in artificial bone based on reverse engineering. *Sheng Wu Yi Xue Gong Cheng Xue Za Zhi*, 1, 225–228. PMID:25997297
- Fu, K., Moreno, D., Yang, M., & Wood, K. L. (2014). Bio-inspired design: An overview investigating open questions from the broader field of design-by-analogy. *Journal of Mechanical Design*, 11(11), 111102. doi:10.1115/1.4028289
- Gao, W., Ramani, K., Cipra, R. J., & Siegmund, T. (2013). Kinetogami: A reconfigurable, combinatorial, and printable sheet folding. *Journal of Mechanical Design*, 11(11), 111009. doi:10.1115/1.4025506

- Gerstle, T. L., Ibrahim, A. M., Kim, P. S., Lee, B. T., & Lin, S. J. (2014). A plastic surgery application in evolution: Three-dimensional printing. *Plastic and Reconstructive Surgery*, 2(2), 446–451. doi:10.1097/01.prs.0000436844.92623.d3 PMID:24469175
- Giordano, M., Ausiello, P., & Martorelli, M. (2012). Accuracy evaluation of surgical guides in implant dentistry by non-contact reverse engineering techniques. *Dental Materials*, 9(9), e178–e185. doi:10.1016/j.dental.2012.06.006 PMID:22796037
- He, Y., Xue, G. H., & Fu, J. Z. (2015). Fabrication of low cost soft tissue prostheses with the desktop 3D printer. *Scientific Reports*, 4, 6973. doi:10.1038/srep06973 PMID:25427880
- Ishida, S., Nojima, T., & Hagiwara, I. (2014). Mathematical approach to model foldable conical structures using conformal mapping. *Journal of Mechanical Design*, 136(9), 091007. doi:10.1115/1.4027848
- Jones, D. B., Sung, R., Weinberg, C., Korelitz, T., & Andrews, R. (2016). Three-Dimensional Modeling May Improve Surgical Education and Clinical Practice. *Surgical Innovation*, 2(2), 189–195. doi:10.1177/1553350615607641 PMID:26423911
- Klein, G. T., Lu, Y., & Wang, M. Y. (2013). 3D printing and neurosurgery - ready for prime time? *World Neurosurgery*, 3-4(3-4), 233–235. doi:10.1016/j.wneu.2013.07.009 PMID:23871811
- Lei, S., Frank, M. C., Anderson, D. D., & Brown, T. D. (2014). A Method to Represent Heterogeneous Materials for Rapid Prototyping: The Matryoshka Approach. *Rapid Prototyping Journal*, 5, 390–402. doi:10.1108/RPJ-10-2012-0095 PMID:26120277
- Leijnse, J. N., & Spoor, C. W. (2012). Reverse engineering finger extensor apparatus morphology from measured coupled interphalangeal joint angle trajectories - a generic 2D kinematic model. *Journal of Biomechanics*, 3(3), 569–578. doi:10.1016/j.jbiomech.2011.11.002 PMID:22134183
- Macko, M., Mikołajewska, E., Szczepański, Z., Augustyńska, B., & Mikołajewski, D. (2016). Repository of images for reverse engineering and medical simulation purposes. *Medical and Biological Sciences*, 3, 23–29.
- Markram, H. (2012). The human brain project. *Scientific American*, 6(6), 50–55. doi:10.1038/scientificamerican0612-50 PMID:22649994
- Maruthappu, M., & Keogh, B. (2014). How might 3D printing affect clinical practice? *BMJ (Clinical Research Ed.)*, 349, g7709. PMID:25550064
- Murphy, S. V., & Atala, A. (2014). 3D bioprinting of tissues and organs. *Nature Biotechnology*, 32(8), 773–785. doi:10.1038/nbt.2958 PMID:25093879
- Rose, N. (2014). The Human Brain Project: Social and ethical challenges. *Neuron*, 82(6), 1212–1215. doi:10.1016/j.neuron.2014.06.001 PMID:24945767
- Seol, Y. J., Kang, H. W., Lee, S., Atala, A., & Yoo, J. J. (2014). Bioprinting technology and its applications. *European Journal of Cardio-Thoracic Surgery*, 46(3), 342–348. doi:10.1093/ejcts/ezu148 PMID:25061217
- Shahmiri, R., Das, R., Aarts, J. M., & Ben-nani, V. (2014). Finite element analysis of an implant-assisted removable partial denture during bilateral loading: Occlusal rests position. *The Journal of Prosthetic Dentistry*, 5(5), 1126–1133. doi:10.1016/j.prosdent.2014.04.023 PMID:24951387
- Telfer, S., Abbott, M., Steultjens, M., Rafferty, D., & Woodburn, J. (2013). Dose-response effects of customised foot orthoses on lower limb muscle activity and plantar pressures in pronated foot type. *Gait & Posture*, 3(3), 443–449. doi:10.1016/j.gaitpost.2013.01.012 PMID:23391752
- Watanabe, T., Hatakeyama, T., & Tomiita, M. (2015). Improving Assistive Technology Service by Using 3D Printing: Three Case Studies. *Studies in Health Technology and Informatics*, 217, 1047–1052. PMID:26294609

Wei, X., & Dong, F. (2011). Development of computer aided forming techniques in manufacturing scaffolds for bone tissue engineering. *Zhongguo Xiu Fu Chong Jian Wai Ke Za Zhi*, 12, 1508–1512. PMID:22242356

Weisman, J. A., Nicholson, J. C., Tappa, K., Jammalamadaka, U., Wilson, C. G., & Mills, D. K. (2015). Antibiotic and chemotherapeutic enhanced three-dimensional printer filaments and constructs for biomedical applications. *International Journal of Nanomedicine*, 10, 357–370. PMID:25624758

Weiss, H. R. (2010). Brace technology thematic series - the Gensingen brace in the treatment of scoliosis. *Scoliosis*, 5(1), 22. doi:10.1186/1748-7161-5-22 PMID:20942970

Wu, G. H., & Hsu, S. H. (2015). Review: Polymeric-Based 3D Printing for Tissue Engineering. *Journal of Medical and Biological Engineering*, 3(3), 285–292. doi:10.1007/s40846-015-0038-3 PMID:26167139

Youssef, R. F., Spradling, K., Yoon, R., Dolan, B., Chamberlin, J., Okhunov, Z., & Landman, J. et al. (2015). Applications of three-dimensional printing technology in urological practice. *BJU International*, 5(5), 697–702. doi:10.1111/bju.13183 PMID:26010346

Yuan, F. S., Sun, Y. C., Wang, Y., & Lü, P. J. (2013). Accuracy evaluation of a new three-dimensional reproduction method of edentulous dental casts, and wax occlusion rims with jaw relation. *International Journal of Oral Science*, 3(3), 155–161. doi:10.1038/ijos.2013.49 PMID:23907676

Zhu, W., O'Brien, C., O'Brien, J. R., & Zhang, L. G. (2014). 3D nano/microfabrication techniques and nanobiomaterials for neural tissue regeneration. *Nanomedicine (London)*, 6(6), 859–875. doi:10.2217/nnm.14.36 PMID:24981651

ADDITIONAL READING

Castroagudin, M. R., Zhai, Y., Li, Z., Marnell, M. G., & Glavy, J. S. (2016). Cyto-3D-print to attach mitotic cells. *Cytotechnology*, 4(4), 1641–1645. doi:10.1007/s10616-015-9917-2 PMID:26464272

Ingber, D. E. (2016). Reverse Engineering Human Pathophysiology with Organs-on-Chips. *Cell*, 6(6), 1105–1109. doi:10.1016/j.cell.2016.02.049 PMID:26967278

Jeandupeux, E., Lobjois, V., & Ducommun, B. (2016). 3D print customized sample holders for live light sheet microscopy. *Biochemical and Biophysical Research Communications*, 4, 1141–1143. PMID:26093294

Numajiri, T., Nakamura, H., Sowa, Y., & Nishino, K. (2016). Low-cost Design and Manufacturing of Surgical Guides for Mandibular Reconstruction Using a Fibula. *Plastic and Reconstructive Surgery Global Open*, 7(7), e805. doi:10.1097/GOX.0000000000000682 PMID:27536484

Pillay, V., & Choonara, Y. E. (2015). 3D Printing in Drug Delivery Formulation: You Can Dream it, Design it and Print it. How About Patent it? *Recent Patents on Drug Delivery & Formulation*, 3(3), 192–193. doi:10.2174/1872211309666150608094744 PMID:26051151

Shinozawa, T., Yoshikawa, H. Y., & Takebe, T. (2016). Reverse engineering liver buds through self-driven condensation and organization towards medical application. *Developmental Biology*, 420(2), 221–229. doi:10.1016/j.ydbio.2016.06.036 PMID:27364470

Zhou, Z. J., Wen, C. L., Sun, Q. M., Wang, A. P., Yan, Z. G., Liu, F., & Li, Y. B. et al. (2016). Morphometric measurement of the cervical spine for minimally invasive pedicle screw fixation using reverse engineering and three-dimensional reconstruction. *International Journal of Medical Robotics and Computer Assisted Surgery*. doi:10.1002/rcs.1765 PMID:27538708

KEY TERMS AND DEFINITIONS

3D Printing: Technical process of building physical objects from a three-dimensional (3D) digital model by adding many subsequent thin layers of special building material (plastic, metal, etc.).

Assistive Technology: Technology used in an assistive technology device or assistive technology service.

Assistive Technology Device: Equipment used to increase, maintain, or improve functional abilities of disabled people.

Biomedical Engineering: Application of principles and practices of engineering science to biomedical research and everyday clinical practice in health care.

Health-Related Quality of Life (HRQoL): Multi-dimensional concept covering the overall condition of a human life in areas associated with impact of the health status on quality of life: physical, mental, emotional, and social functioning.

Physical Therapy: Use of physical therapy methods, techniques and tools to promote, maintain, or restore the physical and physiological well-being of an individual with movement dysfunction.

Quality of Life (QoL): Multi-dimensional concept covering the overall condition of a hu-

man life in various areas: physical, emotional, social, political, moral, etc., their modification and enhancement.

Recovery of Function: A complete (where available) or partial return to the normal (or maximum degree) physiologic activity of an organism following disease or trauma.

Rehabilitation: Process of restoration of human functions to the full (or maximum possible) degree in patients suffering from disease or injury.

Rehabilitation Engineering: A part of biomedical engineering: application of engineering science to design, develop, adapt, test, evaluate, apply, and distribute technological solutions to problems associated with disabilities in areas of mobility, communications, hearing, vision, and cognition.

Reverse Engineering: Quick and cost-effective method of creating copies or modified versions of existing objects, unified to three subsequent stages: digital acquisition of the 3D geometric data, modification/adaptation procedures, and creation of 3D model or final product on 3D printer.

Universal Design: Concept for designing and delivering products and services usable by people with the widest possible range of functional capabilities (including people with various deficits).

Category B

**Business and Organizational
Research**

Acceptance of E-Reverse Auction From the Buyer Perspective

Cigdem Altin Gumussoy

Istanbul Technical University, Turkey

Bilal Gumussoy

Shell and Turcas Petrol Inc., Turkey

INTRODUCTION

With the advances in internet technologies, an e-reverse auction becomes a popular way of procurement of products and services. E-reverse auction is “an online, real-time dynamic auction between a buyer organization and a group of pre-qualified suppliers who compete against each other to win the business to supply goods or services that have clearly defined specifications for design, quantity, quality, delivery, and related terms and conditions” (Beall et al., 2003, p.7). E-reverse auction is first introduced in 1995 by FreeMarkets Online and since then the number of users has dramatically increased.

E-reverse auction initiates with the preparation of a request for quotation (RFQ) by the buyer, which includes the information about the specifications of the products or services to be auctioned. The format of the auction such as historical price, bid decrement and auction starting time is discussed and agreed with the suppliers and the buyer before the auction (Mithas and Jones, 2007). Suppliers all over the world without any geographical limitation can participate to the auction if qualified by the buyer to join. This increases the competition and the chance to find most capable supplier with unique capabilities (Hartley et al., 2004; Lucking-Reiley, 2000). At the starting time, all the potential suppliers log in to the system, start to compete and bid against each other. However, the participants do not see their

identity and generally see the lowest bid value and their rank with respect to the other participants or their rank only. Each participant may submit multiple bids during the auction. This competitive environment can lead to great price reductions for the items auctioned. In addition, cycle-time savings for buyers and suppliers, growing markets, accessing to a larger pool of suppliers can be the other advantages of e-reverse auction use (Schoenherr and Mabert, 2007; Mithas et al., 2008).

On the other hand, using e-reverse auction includes risks such as damaging supplier relationships, switching to a supplier who cannot satisfy the quality requirements, increasing the total cost of procurement, the need to invest in technology, distrust to a new supplier about the responsiveness and flexibility to the changes in demand (Schoenherr and Mabert, 2011; Mithas et al., 2008; Hartley et al., 2006). Therefore, there may be resistance from internal users and suppliers for possible risks or underestimation of the benefits of e-reverse auction use (Schoenherr and Mabert, 2007).

In the current study, a research model including risks and benefits of using e-reverse auction is proposed in order to reveal the important factors in the decision to use e-reverse auction. Furthermore, the effect of organizational characteristics such as firm size, information technology (IT) capability and managerial support on the usage decision are discussed in the research model. The next section discusses the literature on e-reverse auction use.

This section is followed with a research model on the use of e-reverse auction use, and the chapter concludes with the possible future research areas and conclusion.

BACKGROUND

The empirical research on e-reverse auction reveals that the main reason to use e-reverse auction is to maximize the savings in the procurement of products or services. One of the popular service provider, Ariba StartSourcing indicates that they are delivering “value for organizations for all sizes and industries by- sourcing \$340 billion spend in annual across 500 categories, saving \$30 billion annually and cutting process and cycle times by 50-70 percent” (Ariba StartSourcing Solution, 2015).

In contrast to price decrease, switching to a new supplier may increase the transaction cost. The basic principle of Transaction Cost of Economics (TCE) is that “people like to conduct transactions in the most economical way” (Teo and Yu, 2005, p. 452). It explains the reason for selecting one transaction over another. In that transaction the costs are minimized and the benefits are maximized. With e-reverse auction, buyers may procure the products or services with less prices and search costs. Whereas, the supplier change may lead to an uncertainty, which is the principal factor of TCE (Teo and Yu, 2005). Mithas et al. (2008) define uncertainty in the element of non-contractible dimensions. Non-contractible dimensions include situations that cannot be estimated at the start of the contract. Task-based non-contractibility includes uncertainty about the quality, technological investment requirement and prosperity in information exchange, whereas interaction-based dimensions include responsiveness, trust and flexibility of the suppliers for the possible changes in the demand. They revealed that non-contractibility presents an obstacle in the decision to use e-reverse auction. To mitigate risks of the non-contractibility dimensions, buyer can identify a value for the cost

of vendor change. During the e-reverse auction, incumbent supplier loses the auction if the gap between the bids is higher than that pre-identified cost of supplier change factor.

The literature on reverse auction show that larger organizations, using Internet for businesses, also prefer to use e-reverse auction (Hartley et al., 2004; Mithas et al., 2008). IT capability is a prerequisite in order to determine to use such a system (Mithas et al., 2008). The characteristics of a purchased product or services also become important in the decision to use the system. Some products or services require investment in order to produce. These products are generally special products, in which description of such products require close coordination between the supplier and the buyer (Mithas et al., 2008; Schoenherr and Mabert, 2011).

In addition, use of e-reverse auction has negative effects on supplier-buyer relationship (Löscher and Lambert, 2007). Especially the behaviors and practices of buyer during auction may deteriorate the relationship. Furthermore, the lack of clarity in product specifications and requirements of contract create concern among suppliers (Tassabehji et al., 2006)

Our aim in this study is to understand the factors affecting e-reverse auction use. According to the Technology Acceptance Model (TAM), the final decision to use a system is determined by intention to use. Potential users first intend to use and then actually use the system. Therefore, behavioral intention to use is the immediate predictor of actual usages. Perceived usefulness and perceived ease of use are two important beliefs that explain usages and considered as the human related factors. TAM is taken as the base model in constructing the research model since the TAM is a widely used and robust model for predicting and explaining user behavior in IT usage (Davis, 1989; Legris et al., 2003). Several studies related with e-reverse auction also confirmed the significant effect of both perceived usefulness and perceived ease of use on usage (Gumussoy and Calisir, 2009; Gumussoy and Calisir, 2012).

Furthermore, subjective norms, which refer to the “person’s perception that most people who are important to him think that he should or should not perform the behavior in question” (Ajzen and Fishbein, 1980), is included as one of the social factors and potential predictors of system use.

As a result, the research model integrates the concept of buyer price, non-contractibility, organizational factors, product/service related factors, supplier related factors, subjective norms and TAM as the potential predictors of e-reverse auction use.

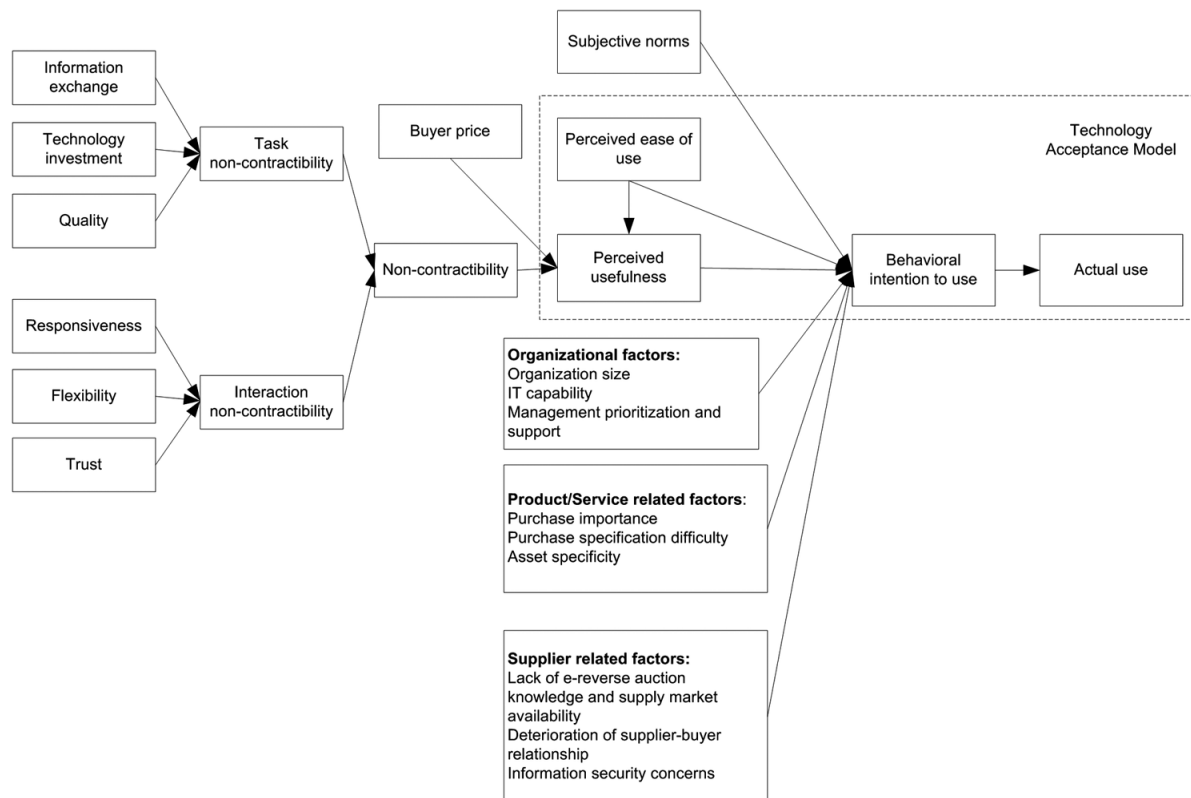
RESEARCH MODEL

The e-reverse auction usage model is presented in Figure 1.

Technology Acceptance Model

Davis (1989) developed TAM in order to understand the main reason for using an IT system. According to the TAM, potential users first decide to use a system and then actually use it. Therefore, it is expected that actual use is predicted by behavioral intention to use, which in turn positively affected by two beliefs: perceived usefulness and perceived ease of use. Perceived usefulness is defined as the “degree to which an individual believes that using a particular system would enhance his or her performance”, perceived ease of use is “the degree to which a person believes that using a particular system would be free of effort” (Davis, 1989, p. 195). According to the TAM, as the system becomes more useful and users do not spend too much effort, intention to use a system

Figure 1. E-reverse auction usage model



is affected positively. Furthermore, perceived ease of use influences behavioral intention to use indirectly through perceived usefulness. As the system becomes more complicated, users spend more time on how to use the system. This may induce users to stop using the system instead of spending too much time on how to use the system. The studies related with e-reverse auction also confirmed that TAM is applicable in explaining the use of e-reverse auction (Gumussoy and Calisir, 2009; Calisir and Gumussoy, 2012). Therefore, it is expected that the relationships defined in the TAM is also applicable in understanding e-reverse auction use.

Subjective Norms

Subjective norms refer to the “person’s perception that most people who are important to him think that he should or should not perform the behavior in question” (Ajzen and Fishbein, 1980). The buyer may decide to use e-reverse auction system with recommendation of other people (such as managers or colleagues) whose point of view are important. Several studies also confirmed the significant effect of subjective norms on the decision to use e-reverse auction (Gumussoy and Calisir, 2009; Calisir and Gumussoy, 2012). Gumussoy and Calisir (2009) found that the opinions of both peers and managers affect the decision to use e-reverse auction use. Calisir and Gumussoy (2012) also confirmed that subjective norms explain the use of e-reverse auction together with attitude and perceived behavioral control. Therefore, it is expected that subjective norms of buyer has a positive effect on behavioral intention to use e-reverse auction.

Buyer Price

Buyer price is influenced with the cost saving of e-reverse auction. Cost saving is “the difference between willingness-to-pay and actual price paid” (Cui and Lai, 2013, p.316). One of the main reasons for using e-reverse auction is the price decrease.

For example, in 2001, General Electric saved 600 million dollar by putting 12 billion dollar to contracts (Kwak, 2002). Davis (1989) revealed that when the saving with the use of e-reverse auction is high, users perceive that the system is more useful. Therefore, it is expected that the buyer who has good experience on the results of previous auction in terms of price decrease will intend to use e-reverse auction in the future.

Non-Contractibility

Non-contractible dimensions include uncertainty and cannot be predicted at the moment, but the changes in the environment and performance in the future may affect the cost of transaction. Mithas et al. (2008) divide non-contractibility dimensions into two: task based non-contractibility and interaction based non-contractibility. Task dimension helps to increase product performance while interaction dimension supports the long term relationship (Mithas et al., 2008). Task-based non-contractibility includes issues related with quality, technology investments and information exchange. Focusing only on purchasing prices may increase the long terms costs since a new low price supplier may have quality problems that cannot be taken into account at the beginning (Hartley et al., 2004). Although buyer may easily specify the requirements of quality, achievement of the standards may be much more difficult for the suppliers. Furthermore, an information exchange especially in terms of tacit knowledge becomes standard with the long term relationship developed between the partners (Mithas et al., 2008). Buyer company also may not be sure about the supplier’s capacity to fulfill the technological investment requirements in the future.

Interaction-based non-contractibility is related with the uncertainty about the relationships between the partners in the future. Buyer wants to trust to suppliers in terms of its integrity, benevolence and ability of supplier in the forthcoming experiences (Zhou, 2013). Integrity is the reliability of supplier in terms of maintaining commitment.

Benevolent suppliers do not only think the benefits of themselves but also the buyer organization. Perceived ability is the capability and adequacy of supplier in terms of knowledge and fulfilling the buyer organization's mandates (McKnight et al., 2002). Furthermore, buyer prefers suppliers who can respond to changes and have adequate flexibility (Mithas et al., 2008). Because this kind of trust is developed over time, buyers may prefer to work with the suppliers that they are certain about the trustworthiness. Therefore, it is hypothesized that the greater the non-contractibility, the less likely the buyer will benefit and use the system.

Organizational Factors

The studies related to the e-reverse auction revealed that organization size has an effect on the decision to use e-reverse auction. Larger firms generally have internet connection and adequate IT infrastructure that e-reverse auction system may be integrated. Furthermore, the volume of small size organizations' needs in terms of products and services may be minor to conduct e-reverse auction (Hartley et al., 2004). Similarly, Hartley et al. (2004) found that firm size is the strongest predictor of likelihood of e-reverse auction use.

Management prioritization and support for e-reverse auction usage are also critical. Some companies put the usage of e-reverse auctions in the employee yearly targets to increase the number of e-reverse auctions so that they save in procurement activities. Therefore, it is expected that managerial prioritization and support for e-reverse auction may increase the use of e-reverse auction.

Product/Service Related Factors

The importance of a product is related with the perceived impact of the product on the firm (McQuiston, 1989). Buyer may not decide on the supplier of high important products easily since the price competition conducted during e-reverse auction cannot be the sole determinant. The buyer

wants to be sure that the products can be procured in a secure way. This can only be achieved by traditional or face-to-face negotiations (Schoenherr and Mabert, 2011). In the literature, Schoenherr and Mabert (2007) show that the companies prefer to use system for more commodity-like product or services. Furthermore, the specialized products require more asset specificity. Asset specificity is the "particular physical or human capital investments that a firm expects to undertake in order to support a given stream of transactions" (Mithas et al., 2008, p. 708). Difference between the asset specificity and non-contractible dimensions is that asset specificity is the need of asset for current stable transactions, whereas non-contractible dimensions involve activities that may change between organizations over time (Mithas et al., 2008).

Product specification difficulty, which refers to the "amount of information needed to fully specify attributes, is sometimes referred to as description complexity" (Mithas et al., 2008, p. 710) may become important in the decision to use e-reverse auction. At the start of the auction, buyer prepares RFQ that describes the specifications of the product or services. On the other hand, there may be fewer suppliers that meet the specifications defined in the RFQ. This decreases the competition during e-reverse auction.

Several studies also confirmed the significant effect of purchase importance (Tassabehji, 2010), product specification difficulty (Schoenherr and Mabert, 2011) and asset specificity (Mithas et al., 2008) on the use of e-reverse auction. Therefore, it is expected that the greater the purchase importance, product specification difficulty and asset specificity, the less likely that the buyers will use e-reverse auction.

Supplier Related Factors

E-reverse auction can only be conducted only when both partners- buyer and suppliers agree to use such a system. In an e-auction system, buyer examines the potential suppliers that the products

can be procured and sends RFQ to join the auction. The lack of supplier may impact the effective use of e-reverse auction negatively because a competitive environment can be settled down only if there exists enough number of suppliers competing to each other. On the other hand, there may be barriers of suppliers in accepting request. The lack of knowledge on how to use e-reverse auction and security of all partners' information may be an obstacle to use e-reverse auction. There needs to make training programs on how to use the system. Furthermore, the information of the parties may be secured with a contract. In addition, the literature on e-reverse auction emphasize that e-auctions affect supplier relationships negatively (Hartley et al., 2006; Gattiker et al., 2007; Tassabehji et al., 2006). Especially, the non-ethical behavior of both supplier and buyer may deteriorate the relationship (Tassabehji et al., 2006). Therefore, lack of e-reverse auction knowledge and available suppliers, deterioration of supplier-buyer relationship and information security concerns may have negative effects on e-reverse auction use.

FUTURE RESEARCH DIRECTIONS

The current study offer a comprehensive research model that explains the possible reasons for using e-reverse auction. By collecting data from the actual users of e-reverse auction, this model may be used to reveal the significant factors affecting e-reverse auction usage decision. Furthermore, group differences in terms of sector or culture can be explored by collecting data from different groups. On the other hand, a successful e-reverse auction can be achieved if both parties- suppliers and buyer organization benefit from the system. As a further study, a research model that explains the reasons of use from the perspective of suppliers can be developed.

In the research model, it is hypothesized that buyer price is one of the important motivator of using e-reverse auction. As a future study, the factors affecting buyer price may be analyzed. According to the studies related to this topic revealed that

different auction configurations may yield different results (Carter and Stevens, 2007; Mithas and Jones, 2007). Carter and Stevens (2007) examine how different reverse auction configurations affect the bid price and supplier's perception. Mithas and Jones (2007) analyze more than 700 auctions conducted in the automotive industry and analyze the effects of bidding competition, information asymmetry, reserve prices, bid decrement and auction duration on buyer surplus.

CONCLUSION

The rapid developments in IT enables the procurement of products or services via electronic system. Buyer organizations use e-auction system to achieve a more competitive environment which leads to great price reductions. Although e-reverse auction system provides benefits in terms of price decrease, small cycle times to determine the prices of products or services compared to standard procurement systems, it includes some risks that may decrease the use of e-reverse auction. In the current study, an integrated model, which includes both the benefits and risks of using e-reverse auction in addition to TAM, is developed. The developed model can guide e-reverse auction users and the procurement managers of the organizations to generalize the use of e-reverse auction.

REFERENCES

- Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior*. Upper Saddle River, NJ: Prentice-Hall.
- Ariba StartSourcing Solution. (2015). *Ariba StartSourcing Solution*. Retrieved June 10, 2016, from <http://www.ariba.com/search/results/cmV2ZXJzZSBhdWN0aW9u>
- Beall, S., Carter, C., Carter, P., Germer, T., Hendrick, T. S. J., Kaufman, L., & Petersen, K. et al. (2003). *The role of reverse auctions in strategic sourcing*. Tempe, AZ: CAPS Research.

- Calisir, F., & Gumussoy, C. A. (2012). Acceptance of e-reverse auction use: A test of competing models. *International Journal of Industrial Engineering*, 19(3), 137–148.
- Carter, C. R., & Stevens, C. K. (2007). Electronic reverse auction configuration and its impact on buyer price and supplier perceptions of opportunism: A laboratory experiment. *Journal of Operations Management*, 25(5), 1035–1054. doi:10.1016/j.jom.2006.10.005
- Cui, X., & Lai, V. S. (2013). Bidding strategies in online single-unit auctions: Their impact and satisfaction. *Information & Management*, 50(6), 314–321. doi:10.1016/j.im.2013.05.007
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *Management Information Systems Quarterly*, 13(3), 319–340. doi:10.2307/249008
- Gattiker, T. F., Huang, X., & Schwarz, J. L. (2007). Negotiation, email, and Internet reverse auctions: How sourcing mechanisms deployed by buyers affect suppliers trust. *Journal of Operations Management*, 25(1), 184–202. doi:10.1016/j.jom.2006.02.007
- Gumussoy, C. A., & Calisir, F. (2009). Understanding factors affecting e-reverse auction use: An integrative approach. *Computers in Human Behavior*, 25(4), 975–988. doi:10.1016/j.chb.2009.04.006
- Hartley, J. L., Lane, M. D., & Duplaga, E. A. (2006). Exploring the barriers to the adoption of e-auctions for sourcing. *International Journal of Operations & Production Management*, 26(2), 202–221. doi:10.1108/01443570610641675
- Hartley, J. L., Lane, M. D., & Hong, Y. (2004). An exploration of the adoption of e-auctions in supply management. *IEEE Transactions on Engineering Management*, 51(2), 153–161. doi:10.1109/TEM.2004.826010
- Kwak, M. (2002). *Smart ideas about reverse auctions*. Harvard Business School. Retrieved: 23 June 2016, from http://hbswk.hbs.edu/item.jhtml?id_3086&t=strategy&noseek=one
- Legris, P., Ingham, J., & Colletette, P. (2003). Why do people use information technology? A critical review of the technology acceptance model. *Information & Management*, 40(3), 191–204. doi:10.1016/S0378-7206(01)00143-4
- Lösch, A., & Lambert, J. S. (2007). E-reverse auctions revisited: An analysis of context, buyer-supplier relations and information behavior. *Journal of Supply Chain Management*, 43(4), 47–63. doi:10.1111/j.1745-493X.2007.00040.x
- Lucking-Reiley, D. (2000). Auctions on the Internet: What is being auctioned, and how? *The Journal of Industrial Economics*, 48(3), 227–252. doi:10.1111/1467-6451.00122
- McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and validating trust measures for e-commerce: An integrative typology. *Information Systems Research*, 13(3), 334–359. doi:10.1287/isre.13.3.334.81
- McQuisten, D. H. (1989). Novelty, complexity, and importance as casual determinants of industrial buyer behavior. *Journal of Marketing*, 53(2), 66–79. doi:10.2307/1251414
- Mithas, S., & Jones, J. L. (2007). Do auction parameters affect buyer surplus in e-auctions for procurement. *Production and Operations Management*, 16(4), 455–470. doi:10.1111/j.1937-5956.2007.tb00272.x
- Mithas, S., Jones, J. L., & Mitchell, W. (2008). Buyer intention to use internet-enabled reverse auctions: The role of asset specificity, product specialization, and non-contractibility. *Management Information Systems Quarterly*, 32(4), 705–724.
- Schoenherr, T., & Mabert, V. A. (2007). Online reverse auctions: Common myths versus evolving reality. *Business Horizons*, 50(5), 373–384. doi:10.1016/j.bushor.2007.03.003
- Schoenherr, T., & Mabert, V. A. (2011). A comparison of online and offline procurement in B2B markets: Results from a large-scale survey. *International Journal of Production Research*, 49(3), 827–846. doi:10.1080/00207540903473359

Tassabehji, R. (2010). Understanding e-auction use by procurement professionals: Motivation, attitudes and perceptions. *Supply Chain Management: An International Journal*, 15(6), 425–437. doi:10.1108/13598541011080419

Tassabehji, R., Taylor, W. A., Beach, R., & Wood, R. (2006). Reverse e-auctions and supplier-buyer relationships: An exploratory study. *International Journal of Operations & Production Management*, 26(2), 166–184. doi:10.1108/01443570610641657

Teo, T. S. H., & Yu, Y. (2005). Online buying behavior: A transaction cost economics perspective. *Omega The International Journal of Management Science*, 33(5), 451–465. doi:10.1016/j.omega.2004.06.002

Zhou, T. (2013). Understanding continuance usage of mobile sites. *Industrial Management & Data Systems*, 113(9), 1286–1299. doi:10.1108/IMDS-01-2013-0001

ADDITIONAL READING

Aloini, D., Dulmin, R., & Mininno, V. (2012). E-reverse auction design: Critical variables in a B2B context. *Business Process Management Journal*, 18(2), 219–249. doi:10.1108/14637151211225180

Bei, L., & Chen, M. (2015). The effects of hedonic and utilitarian bidding values on e-auction behavior. *Electronic Commerce Research*, 15(4), 483–507. doi:10.1007/s10660-015-9197-0

Caniels, M. C. J., & Raaij, E. M. (2009). Do all suppliers dislike electronic reverse auctions? *Journal of Purchasing and Supply Management*, 15(1), 12–23. doi:10.1016/j.pursup.2008.10.003

Charki, M. H., Jossierand, E., & Charki, N. B. (2011). Toward an ethical understanding of the controversial technology of online reverse auctions. *Journal of Business Ethics*, 98(1), 17–37. doi:10.1007/s10551-010-0532-z

Doherty, N. F., McConnell, D. J., & Ellis-Chadwick, F. (2013). Institutional responses to electronic procurement in the public sector. *International Journal of Public Sector*, 26(6), 495–515. doi:10.1108/IJPSM-04-2012-0048

Elmaghraby, W. (2007). Auctions within e-sourcing events. *Production and Operations Management*, 16(4), 409–422. doi:10.1111/j.1937-5956.2007.tb00269.x

Etzion, H. E., & Pinker, J. (2008). Asymmetric competition in B2B spot markets. *Production and Operations Management*, 17(2), 150–161. doi:10.3401/poms.1080.0014

Gupta, S., Koulamas, C., & Kyparisis, G. J. (2009). E-business: A review of research published in production and operations management (1992–2008). *Production and Operations Management*, 18(6), 604–620. doi:10.1111/j.1937-5956.2009.01100.x

Hawkins, T. G., Gravier, M. J., & Wittmann, C. M. (2010). Enhancing reverse auction use theory: An exploratory study. *Supply Chain Management: An International Journal*, 15(1), 21–42. doi:10.1108/13598541011018102

Ho, K. K. W., Yoo, B., & Yu, S. (2015). An empirical analysis of factors affecting the bidding competition in an online auction: A comparison between English and buy-it-now auctions. *Journal of Electronic Commerce Research*, 16(4), 329–341.

Jap, S. D., & Haruvy, E. (2008). Inter organizational relationships and bidding behavior in industrial online reverse auctions. *JMR, Journal of Marketing Research*, 45(5), 550–561. doi:10.1509/jmkr.45.5.550

Kim, Y. (2005). The effects of buyer and product traits with seller reputation on price premiums in e-auction. *Journal of Computer Information Systems*, 46(1), 79–91.

- Kumar, S. (2013). Is your supply chain business operation a good fit for online reverse auctions? *Transportation Journal*, 52(1), 121–133. doi:10.5325/transportationj.52.1.0121
- Kumar, S., & Chang, C. W. (2007). Reverse auctions: How much total supply chain cost savings are there? A business simulation model. *Journal of Revenue and Pricing Management*, 6(3), 229–240. doi:10.1057/palgrave.rpm.5160088
- Lu, J., Wang, L., Yu, C., & Wu, J. (2009). E-auction web assessment model in China. *Electronic Commerce Research*, 9(3), 149–172. doi:10.1007/s10660-009-9033-5
- Majid, K., Bryant, A., & Rau, P. A. (2014). Name your price – online auctions and reference prices. *Journal of Product and Brand Management*, 23(6), 420–428. doi:10.1108/JPBM-06-2014-0626
- Pham, L., Teich, J., Wallenius, H., & Wallenius, J. (2015). Multi-attribute online reverse actions: Recent research trends. *European Journal of Operational Research*, 242(1), 1–9. doi:10.1016/j.ejor.2014.08.043
- Razuk, R. O., Arkader, R., & Braga, A. R. (2009). Reverse e-auction usage in supply processes of Brazilian companies: An exploratory survey. *Latin American Business Review*, 10(1), 41–60. doi:10.1080/10978520902908551
- Rotchanakitumnuai, S. (2013). The governance evidence of e-government procurement. *Transforming Government People. Process and Policy*, 7(3), 309–321.
- Rothkopf, M. H., & Whinston, A. B. (2007). On e-auctions for procurement operations. *Production and Operations Management*, 16(4), 404–408. doi:10.1111/j.1937-5956.2007.tb00268.x
- Saprikis, V. (2013). Suppliers behavior on the post-adoption stage of business-to-business e-reverse auctions: An empirical study. *Telematics and Informatics*, 30(2), 132–143. doi:10.1016/j.tele.2012.04.002
- Sehwail, L. M., Ingalls, R. G., & Pratt, D. B. (2008). Business-to-business online reverse auctions: A literature review and a call for research. *International Journal of Services and Operations Management*, 4(4), 498–520. doi:10.1504/IJSOM.2008.017432
- Standing, S., Standing, C., Love, P., & Gengatharen, D. (2013). How organizing visions influence the adoption and use of reverse auctions. *Electronic Commerce Research*, 13(4), 493–511. doi:10.1007/s10660-013-9133-0
- Tassabehji, R. (2014). Electronic reverse auctions: Emerging from the shadows? In E. Lacka, H. K. Chan, & N. Yip (Eds.), *E-commerce platform acceptance suppliers, retailers, and consumer*. Springer International Publishing. doi:10.1007/978-3-319-06121-4_4

KEY TERMS AND DEFINITIONS

Acceptance: A decision to use a particular IT/IS system.

Bidding: It is an offer of a supplier during e-reverse auction that the supplier is willing to sell.

E-Procurement: Purchase of a product or a service over internet.

E-Reverse Auction: It is a type of e-auction, conducted in real time over internet between buyer and suppliers in order to determine the supplier and the price of a product/service.

Negotiation: It is a discussion about an issue to reach a consensus between the parties.

Request for Quotation: Includes the specifications and some critical contract terms like payment terms etc. of a product or a service to be auctioned, and is prepared by the buyer in order to make an invitation to the suppliers.

Reserve Prices: It is the starting price of an e-reverse auction determined by a buyer with the examination of historical prices of a product/service.

Advanced ICT Methodologies (AIM) in the Construction Industry

B**M. Reza Hosseini***Deakin University, Australia***Saeed Banihashemi***University of Technology Sydney, Australia***Fahimeh Zaeri***Auckland University of Technology, New Zealand***Alireza Adibfar***University of Florida, USA*

INTRODUCTION

A wide range of advanced ICT methodologies (AIM) have made inroads in the construction market due to the conditions rampant in today's business environment. That is, the trend of globalisation, the fierce competition in the market and the need to deliver more complex projects have led leading construction companies towards adopting available technological innovations (Toole, Hallowell, & Chinowsky, 2013). As a corollary, the construction community has observed a rapid emergence of different AIM to improve communications, visualization, and foster collaboration among the parties involved in construction projects. Nevertheless, the overall rate of adoption within the industry has been slow, particularly in small and medium sized (SMEs) companies (Heydarian et al., 2015). Evidence shows that such slow uptake of cutting-edge methodologies is for the most part stemmed from a lack of support of practitioners in the construction industry exacerbated by clients being uninformed of the advantages of available methods for their projects (Cao, Li, & Wang, 2014). In essence, major barriers to widespread adoption of new technology in the construction industry have roots in lack of familiarity of clients, managers and company directors

with AIM, and the potential benefits envisaged for utilizing such technology as observed by Van der Vlist, Vrolijk, and Dewulf (2014). As asserted by Williams (2007, p. 6) one barrier to adoption of AIM is "Lack of knowledge by people in the construction industry about what is possible."

Therefore, the first step towards the enhancement of ICT adoption in the construction context includes generating the necessary motivation amongst practitioners and policy makers through disseminating the knowledge about currently-available methods and highlighting the potential benefits (Sarshar & Isikdag, 2004; Williams, 2007). To address such need, this chapter aims at introducing available cutting edge methodologies applicable to construction activities. The useful feature of this chapter is that it brings information about different cutting-edge methodologies under one cover. The content of the chapter is designed for readers who might not be specialists in ICT-oriented techniques while working within any domain of the construction industry. In the classroom, this chapter could be a supplementary source for courses on information systems in construction or lean project delivery.

The chapter starts with a brief discussion on the driving forces and root causes of rapid advances in the tools and technologies utilized in the con-

struction context. This is followed by describing the concepts of virtual design and construction (VDC). Afterwards, different aspects of building information modelling are discussed. The rest of the chapter is concerned with introducing various innovative AIM that are emerging within the construction industry.

BACKGROUND

The driving forces pushing the construction industry towards adopting AIM for the most part have roots in two main categories of drivers. These include the efficiency needed to proactively addressing internal process problems facing the industry alongside pressures from the external environment (Cao et al., 2014). In essence, the main reasons behind the interest of construction practitioners to adopt advanced ICT methods come from three main grounds under two categories:

1. External pressure
 - a. Fierce competition of globalisation (external pressure)
 - b. New requirements of projects (external pressure)
2. Internal problems
 - a. Advantages of advanced ICT methodologies (efficiency needed to address internal problems)

As illustrated in Figure 1, the capabilities of AIM in terms of facilitating achieving efficacy in delivering projects by integration of design and construction phases as a main reason for construction companies (Aouad, Lee, & Wu, 2007). In this regard, enhancing collaboration level, facilitating accessibility and exchange of information and enhancing the effectiveness of communications has been acknowledged (Nitithamyong & Skibniewski, 2004). Besides, many studies have attested to the positive impacts of utilising ICT in the construction industry on grounds of improving cost-effectiveness, scheduling and quality of products (Williams, 2007). In the same vein, ICT

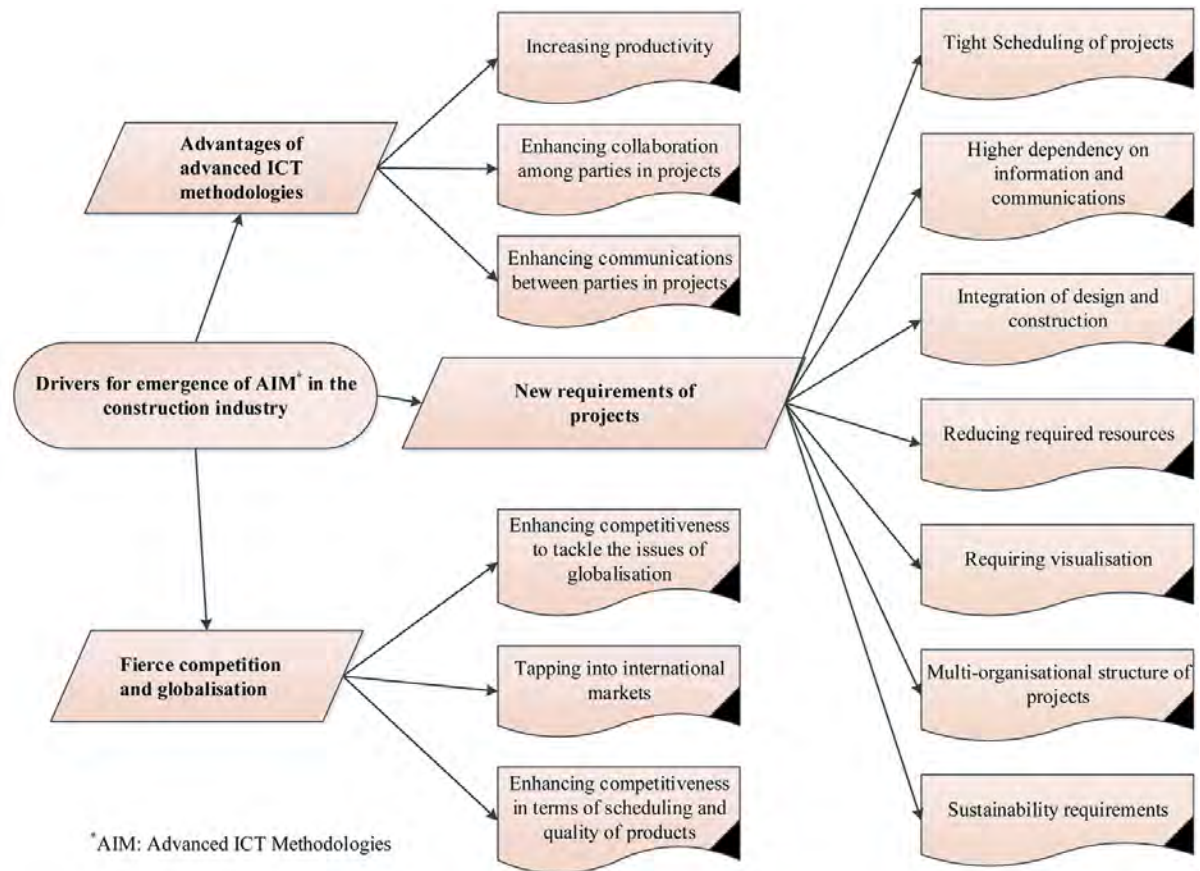
implementation has far-reaching positive impacts on all major aspects of AEC projects including time, cost, competitiveness enhancement and information exchange (Hosseini, Chileshe, Zuo, & Baroudi, 2012). AIM are deemed the most promising tool for organisations in today's market. That is, they are regarded as a catalyst for process improvement through providing the ability for construction companies to share multi-disciplinary interests, goals, perspectives, and constraints in an accurate, timely and economical way (Aouad et al., 2007). As such, there is consensus regarding the positive effects of utilising AIM in construction organisations and on projects (Adriaanse, Voordijk, & Dewulf, 2010). Moreover, the cost of adoption AIM on construction projects is continuously declining (Sardroud, 2015). On top of that, technologies that support AIM are progressively advancing in other fields. These result in AIM redefining the methods and procedures in the construction context, thus bringing a paradigm shift into the practices of the construction context (Hardin & McCool, 2015).

In this context, a wide range of AIM have emerged under the framework of VDC (Khanzode, Fischer, Reed, & Ballard, 2006) as will be described next.

Virtual Design and Construction

The VDC concept emerged out of the attempts to apply lean production principles to the practices of the construction industry in delivering projects. The birthplace for VDC was the Centre for Integrated Facility Engineering (CIFE) at Stanford University where VDC was officially introduced in 2001. CIFE has remained a magnet for research and teaching on VDC since then (Kunz & Fischer, 2012). As defined by CIFE, VDC is the application of multi-disciplinary performance models of design/construction, which include the *Product* (i.e., facilities), *Organization* of the design, and *Processes* referred to as POP in order to support business interests. In essence, VDC tools enable construction practitioners of simulating construc-

Figure 1. The major driving forces behind the emergence of AIM within the construction industry
 Source: adapted from Hosseini & Chileshe, 2015



tion project models as the composition of related models of Product, Organization and Process or POP as asserted by Garcia, Kunz, Ekstrom, and Kiviniemi (2004). This benefits construction practitioners by, automation, integration, and visualization of construction activities in particular to predict project outcomes and manage the project towards the desired performance (Yee, Fischer, & Kam, 2013).

Adopting VDC could be advantageous in every stage of a project even during the bidding and preparing the tender documents by increasing the accuracy and efficacy available information and data. Rich visualised representation of data could be used to effectively communicate complex technical information and illustrate predictable circumstances that are likely to occur in the real projects through a virtual environment. This will

enhance effective communication of concepts and designs for all the stakeholders. Based on the concept of POP as discussed above, Khanzode et al. (2006) introduced the most well-known available AIM under VDC as illustrated below.

- Product visualization tools or 3D object modelling, which can create a common understanding among various discipline on how the completed product of the project (e.g. a building) will look like. 3D models will facilitate coordination of design among different disciplines as well.
- Product+ Processes visualisation tools, which present a visualisation of the processes for completion of the product based on 3D models. These tools will add a di-

mension of time to 3D models and are usually referred to as 4D (3D+ Time = 4D).

- *Organization+ Processes* tools that simulate the effort needed for completion of the projects and assist stakeholders in identifying organizational risks. Online collaboration tools fall within this category.

Apart from this, the concept of *nD* model is used in the literature for referring to visualization models in which information required for various stages of a project are embedded (Aouad et al., 2007). Such models could contain information regarding material resources integrated with the scheduling and cost data, termed as 5D (Lu, Won, & Cheng, 2015). All such visualisation models are in fact created through Building Information Modelling (BIM) as described below.

Building Information Modelling (BIM)

Building Information Modelling (BIM) has been evolving since 2003 with no clear starting point found for the commercial introduction of BIM (Holzer, 2015). In today's business environment BIM could be regarded as the most promising ICT-oriented development in the construction industry (Eastman, Teicholz, & Sacks, 2011). BIM involves collating, applying and maintaining an integral digital representation of all building information for different phases of the project life cycle in the form of a data repository (Gu & London, 2010). It provides a comprehensive concept as an umbrella for the processes and tools, which integrate all projects required data through containing information needed in particular phases of a building's life-cycle (scheduling, analysis, cost evaluation, etc.). Yet, BIM is much more than a data container for the building model; it is an object oriented building design and construction-specific model to assist the progress of the exchange and interoperability of data in the digital format (Babič, Podbreznik, & Rebolj, 2010). A major benefit of utilising BIM in the design and construction phase of a project is obviously coming through its ability to 'model' and test the constructability of

the design within the model prior to setting foot on the project site. As a management paradigm, BIM can be implemented through chains of ICT technologies including BIM authoring tools such as Revit, ArchiCad, Microstation and Navisworks.

Implementing BIM helps to avoid errors alongside improving the productivity, scheduling, safety, cost and quality of construction projects (Zuppa, Issa, & Suermann, 2009). BIM is a fast and effective process by which information pertaining to one project can be updated at any stage of project from any department or unit (e.g. engineering department). Accordingly, because of its efficiency in adopting and propagating changes in the model, editing objects and reloading updated links, the entire project model will be updated based on the changes on one aspect of the project. It is asserted that BIM is capable of enhancing the performance within the industry along with overcoming the problems stemmed from the fragmented structure dominating the industry (Succar, 2009). As an emerging method within the construction context, BIM entails managing data flows in different natures between different parties (Succar, 2009) by manifesting a virtual working environment (Fox & Hietanen, 2007; Howard & Björk, 2008). Serving a catalyst of change for the construction industry, BIM encompasses a radical reorientation of 2D to 3D modelling and a recent shift to 4D (project scheduling integrated), 5D (project cost integrated) and 6D (facility management integrated), exploiting more intelligent data analysis techniques in order to achieve a superior performance in delivering an As-Built BIM. In this context, ICT should supply the platform for BIM to pursue its trend becoming a mainstay of research and development in the construction industry (Singh, 2014).

ICT and Sustainable Built Environment

The construction industry is progressively discovering how ICT tools can help the construction industry to attain more sustainable outcomes. ICT is capable of providing a unified approach in

storing, transmitting, sharing and distributing data among the parties involved in the built environment context. The advantages of incorporating ICT in fostering a sustainable built environment include:

- Facilitate the rule-based checking of sustainability guidelines,
- Find and retrieve environmental solutions,
- Increase the speed and efficiency of sharing green principles, and
- Decrease the environmental impact by underrating the paperwork and waste generation.

One example for integration of ICT into sustainable construction could be the powerful visualization engine of BIM, which envisages any technically possible way of creating or testing “what if” scenarios during the sustainable building procedure (Banihashemi Namini, Ding, Wang, & Jack, 2015). Beyond graphically depicting the design, much of the data needed for supporting sustainable building is captured naturally as design of the project proceeds. It paves the way of superimposing multidisciplinary information in an integrated model creating an opportunity for measuring the sustainability state of the project throughout its lifecycle (Azhar, Carlton, Olsen, & Ahmad, 2011). According to Krygiel and Nies (2008), BIM has the potential to aid sustainable design through:

- Selecting a suitable building orientation,
- Evaluating building envelop,
- Day-lighting analysis,
- Water needs reduction in a building,
- Integration with energy analysis software to reduce energy consumption,
- Featuring renewable energy possibilities,
- Material needs reduction,
- Decreasing waste and carbon emission by logistic site management.

Architects can import data into BIM model to locate the project geographically and present the

conditions of climate, location and surrounding area. Then, the model can be reoriented and edited on site based on the real coordination for reducing required resources and exposing efficiently to solar radiation (Eastman, Teicholz, & Sacks, 2011). Engineers utilize BIM to decrease the energy consumption by exporting the 3D model to particular energy analysis software and calculating light reflectance and transmittance (Underwood, Isikdag, & Global, 2010). Effectively coordinating logistics through site analysis and modelling including wetlands and protected habitats can help contractors to eliminate the possible issues and even easily quantify the amount of materials extraction of buildings to simplify reusing or recycling measures. Finally, the BIM file arranges an array of required resources to advise the project team regarding sustainability issues from the incipient stages of a project (Hardin, 2011).

Smart mechanisms such as smart metering and sensor network-based sub-utility measurements are among the emerging ICT technologies that will be rapidly developing in the upcoming years. These devices are capable of measuring energy usage more precise than the current instruments and also transmitting the detected information via networks. This system will allow for better energy consumption records in both temporal and demand profiles as well as supplying the novel digital platforms for different stakeholders including energy performance benchmarking, load forecasting and consumption monitoring for residential buildings (Ahola, Ahlqvist, Ermes, Myllyoja, & Savola, 2010). Provision of these services enables all parties involved to apply environmentally sound principles in energy usage reduction, enhance energy efficiency measures and in the long run, integrate with other ICT tools like smart appliances and building automation systems.

Virtual Reality

The term ‘virtual reality’ (VR) has been used in the literature for more fifteen years. The expression VR was introduced for the first time in the

US in the 80s by Jaron Lanier and became a common name. VR is simply defined as providing a cognitive activity in an artificial digitally-created world. Such a world is symbolic, imaginary and is literally a simulation of certain aspects of the real world with attributes copied from the real world (Fuchs, Guitton, & Moreau, 2011). For the construction context, VR presents a natural medium for a building design and provides 3D visualization, which can be manipulated in real-time and deployed in a collaborative manner in order to explore all stages of the lifecycle of construction projects. The applications of VR for the construction industry are diverse and target a wide range of audience. That is, VR greatly benefits the construction industry through providing (Dawood, 2009; Nikolic, Jaruhar, & Messner, 2011):

- Greater level of safety on construction sites,
- Enabling construction practitioners to test alternative construction techniques and methods by simulation and visualisations of all alternatives,
- Evaluation and assessment of available information in a virtual environment
- Allowing an accurate sequencing of operations,
- Presenting novel approaches for collaboration and communication between parties involved on projects and
- Education and training of students

Augmented Reality

Augmented Reality (AR) is a powerful user interface paradigm that enhances a user's perception by inserting computer-generated information into a user's real world experience. Construction practitioners need to make decisions frequently on onsite work environment. This necessitates access to mobile visualization technology as well as being grounded and aware of the real construction site context. This has become the primary driving force behind increasing adoption and

promise of AR compared to VR. That is, AR is a tool capable of supplementing reality instead of replacing it. AR is the *middle ground* between a purely virtual environment and a completely real world (Lin, Liu, Tsai, & Kang, 2014). AR has been employed for a wide range of applications in the construction industry. These include maintenance, manufacturing, training, 3D video conferencing, computer assisted training and instruction, entertainment and construction design. Using AR, a construction practitioner will be able to combine the 3D object into the normal viewing perspective without losing any of the advantages of object movement and individual movement in a real-world environment. When it comes to BIM, AR is envisaged to become the tool to convey BIM in an effective manner on construction sites described as extended hand for BIM (Wang, Truijens, Hou, Wang, & Zhou, 2014).

Drones

A drone is an aerial vehicle without on board pilot officially termed as Unmanned Aerial Vehicles (UAVs). Drones came into public attention due to their recent applications in military operations and their civil applications predominantly in the agriculture industry (Nonami, Kendoul, Suzuki, Wang, & Nakazawa, 2010). Drones offer major benefits when used for aerial surveillance and inspection in complex environments such as construction sites. Application of drones on construction sites is very recent and limited to below activities.

- Assisting safety management for which drone acts as a safety officer and controls and monitors a number of workers. It assists safety managers to spot safety issues on the site quickly (Irizarry, Gheisari, & Walker, 2012).
- Drones assist site managers by picking up the signals emitted by Radio Frequency Identification Technology (RFID) tags and provide information on the location and

movement of tags. The data could be collected to be used in BIM models (Hubbard et al., 2015).

- Drones can be flown to certain altitudes and locations for capturing images. The images can be used to develop 3D illustration by the use of AR technology.
- Surveying for which drones can be used in surveying activities with difficult access to objects such as landfill, pits, stockpile, quarries, surface mining areas and slopes. Drones are efficient and can complete surveying activities in shorter times, at lower costs, with fewer number of personnel and in higher levels of accuracy.

Geospatial Information System (GIS)

The concept of GIS emerged in the 60s in Canada. GIS is defined as a tool that is utilised to assist people in acquiring spatial information and learn about geography. That is, storing spatial data in digital formats provides rapid access for a wide range of purposes (Fazal, 2008). When the performing area of construction site is vast, a single BIM model may not be a good fit to handle all the existing objects and facilities. Considering a large-sized construction zone with a lot of buildings and recreational facilities with myriad trees and green spaces would make it hard to think about implementing all the substances on a single model. Working with such a model would be hard, time consuming and vulnerable to flaws. Thus, the BIM approach needs to use a reliable and comprehensive context in order to link different BIM models and their information and provide an integrated and powerful decision support system for both project managers and stakeholders to handle the whole area of the project at hand. GIS provides a reliable tool based on its precise coordination system and geospatial integrity for as many different BIM models as possible (Mignard & Nicolle, 2014). With GIS, the entire most recent detailed data about every single project becomes readily available. Moreover, GIS would let the owners

to have cumulative information about the whole project data including elapsed and remaining costs, time and required materials. This would help them have remarkable ability toward decision making and expenditure control. Therefore, bundling GIS with BIM would bring remarkable potentials for managers in large-sized construction projects.

3D Terrestrial Laser Scanning (TLS)

There is a great need within the construction industry for automating the inspection of projects. Among all available methods, 3D Terrestrial Laser Scanning (TLS) has found to be the best available method to capture 3D information on a construction project with high levels of accuracy in high speed. TLS is defined as a modern measurement technology that is revolutionizing dimensional surveying in the construction industry particularly in the facilities management field (Bosché & Biotteau, 2015). TLS has been a promising technique in the construction industry, which enables construction practitioners of tracking their progress, controlling quality, monitoring health and creating 3D as-built models for managing facilities (Bosché, Ahmed, Turkan, Haas, & Haas, 2015). One of the recent advancements has been the emergence of interfaces between BIM models and models created by TLS.

3D Printing

According to Cano (2015, p. 3) “3D printing is basically a type of additive technology manufacturing, which means that instead of starting with a solid block of material and cutting it down to a shape, an object is built by adding one layer of material at a time.”. 3D printing falls within the category of rapid manufacturing methods, which have permeated the construction industry. 3D printing based on inkjet technology is a common method in which the inkjet selectively deposits a binder in form of liquid onto a bed of powder. The role of the binder is to glue the powder together (Buswell, Soar, Gibb, & Thorpe, 2007). The

construction industry is moving toward using utilization of manufacturing methods for which construction components could be printed on site followed by assembly. Printing components such as concrete parts have become common and experienced in detail, yet printing large construction components or full-scale units are in their infancy stages (Lim et al., 2012).

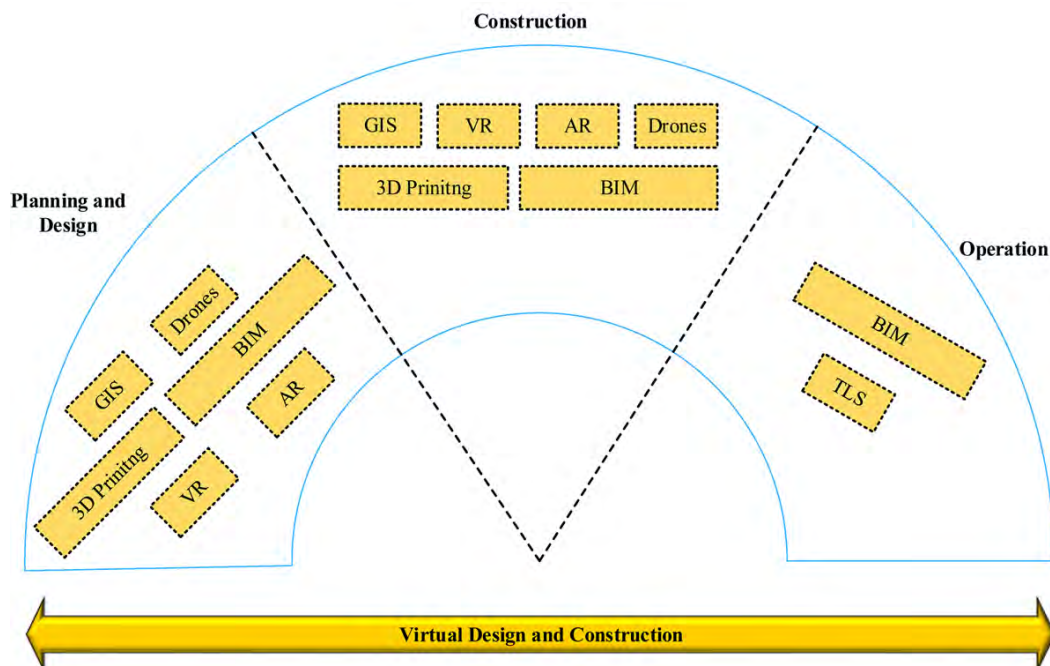
FUTURE RESEARCH DIRECTIONS

Future researchers would contribute to the construction industry body of the knowledge through attempting to integrate the different AIM within the lifecycle of construction projects. That is, at the moment applications of AIM on projects are piecemeal and disjointed, thus a major part of the potentials is lost. Integration of AIM within each other under the umbrella of VDC can assist construction practitioners in gaining the full potentials envisaged for using AIM. A rudimentary effort has been presented here in Figure 2 to show

a sample of delivering a construction projects in 3 main phases utilising an integrated AIM.

Therefore, further research studies should concentrate on ascertaining the most effective policies to lead construction practitioners towards harnessing the benefits of adopting integrated AIM in different stages of construction projects. To this end, future inquiries should present frameworks for facilitating managing change in construction organisations. A deep appreciation of the dynamics of change at the people/technology interaction is a prerequisite for the successful implementation of integrated AIM on construction projects. This could be a fertile ground for future investigations. Additionally, mapping the status of universities and the levels of skills and knowledge of graduates in regards to using each of AIM elements particularly BIM should be considered as a subject for future inquiries. Such explorations should compare the pertinent requirements in the industry against the abilities of graduates and the content of university programs and accordingly spot the gaps and shortcomings in this area.

Figure 2. Delivering a construction projects through an integrated AIM



CONCLUSION

The construction context has yet to harness the benefits of AIM effectively, due to many barriers hampering the shift from traditional methods to AIM. The major barriers include cultural and organisational resistance coupled with the lack of knowledge in the industry. Hence, it becomes incumbent to practitioners and researchers to foster the culture of adopting AIM and promote delivering projects through integration of AIM as described in previous sections. One remedial solution for this might be conducting research studies to spot the true impediments towards uptake of AIM in construction organizations and projects. In the same vein, procedures of merging AIM into construction activities and processes would be facilitated by incorporating the principles of established theories associated with innovation adoption. To this end, inquiries should consider building upon robust theoretical foundations such as principles of Innovation Diffusion Models (IDMs).

REFERENCES

- Adriaanse, A., Voordijk, H., & Dewulf, G. (2010). The use of interorganisational ICT in United States construction projects. *Automation in Construction*, 19(1), 73–83. doi:10.1016/j.autcon.2009.09.004
- Ahola, J., Ahlqvist, T., Ermes, M., Myllyoja, J., & Savola, J. (2010). *ICT for environmental sustainability Green ICT Roadmap*. VTT Tiedotteita.
- Aouad, G., Lee, A. P. D., & Wu, S. D. (2007). *Constructing the future: nD modelling*. London: Routledge.
- Azhar, S., Carlton, W. A., Olsen, D., & Ahmad, I. (2011). Building information modeling for sustainable design and LEED® rating analysis. *Automation in Construction*, 20(2), 217–224. doi:10.1016/j.autcon.2010.09.019
- Babič, N. Č., Podbreznik, P., & Rebolj, D. (2010). Integrating resource production and construction using BIM. *Automation in Construction*, 19(5), 539–543. doi:10.1016/j.autcon.2009.11.005
- Banihashemi Namini, S., Ding, G. W., & Jack. (2015). *Developing A Framework of Artificial Intelligence Application for Delivering Energy Efficient Buildings Through Active BIM*. Paper presented at the COBRA 2015.
- Bosché, F., Ahmed, M., Turkan, Y., Haas, C. T., & Haas, R. (2015). The value of integrating Scan-to-BIM and Scan-vs-BIM techniques for construction monitoring using laser scanning and BIM: The case of cylindrical MEP components. *Automation in Construction*, 49, Part B, 201–213.
- Bosché, F., & Biotteau, B. (2015). Terrestrial laser scanning and continuous wavelet transform for controlling surface flatness in construction – A first investigation. *Advanced Engineering Informatics*, 29(3), 591–601. doi:10.1016/j.aei.2015.05.002
- Buswell, R. A., Soar, R. C., Gibb, A. G. F., & Thorpe, A. (2007). Freeform Construction: Mega-scale Rapid Manufacturing for construction. *Automation in Construction*, 16(2), 224–231. doi:10.1016/j.autcon.2006.05.002
- Cano, L. M. (2015). *3D printing: a powerful new curriculum tool for your school library*. Santa Barbara, CA: Libraries Unlimited.
- Cao, D., Li, H., & Wang, G. (2014). Impacts of isomorphic pressures on BIM adoption in construction projects. *Journal of Construction Engineering and Management*, 140(12), 04014056. doi:10.1061/(ASCE)CO.1943-7862.0000903
- Dawood, N. (2009). VR-roadmap: a vision for 2030 in the built environment. *Journal of Information Technology in Construction*, 14, 489–506.
- Eastman, C., Teicholz, P., & Sacks, R. (2011). *BIM handbook: a guide to building information modeling for owners, managers, designers, engineers and contractors* (2nd ed.). Hoboken, NJ: John Wiley and Sons.

- Fazal, S. (2008). GIS basics. New Delhi: New Age International (P) Ltd.
- Fox, S., & Hietanen, J. (2007). Interorganizational use of building information models: Potential for automational, informational and transformational effects. *Construction Management and Economics*, 25(3), 289–296. doi:10.1080/01446190600892995
- Fuchs, P., Guitton, P., & Moreau, G. (2011). Virtual reality: Concepts and technologies. Boca Raton, FL: CRC Press.
- Garcia, A. C. B., Kunz, J., Ekstrom, M., & Kiviniemi, A. (2004). Building a project ontology with extreme collaboration and virtual design and construction. *Advanced Engineering Informatics*, 18(2), 71–83. doi:10.1016/j.aei.2004.09.001
- Gu, N., & London, K. (2010). Understanding and facilitating BIM adoption in the AEC industry. *Automation in Construction*, 19(8), 988–999. doi:10.1016/j.autcon.2010.09.002
- Hardin, B. (2011). *BIM and construction management: proven tools, methods, and workflows*. John Wiley & Sons.
- Hardin, B., & McCool, D. (2015). *BIM and Construction Management: Proven Tools, Methods, and Workflows*. Hoboken, NJ: John Wiley & Sons.
- Heydarian, A., Carneiro, J. P., Gerber, D., Becerik-Gerber, B., Hayes, T., & Wood, W. (2015). Immersive virtual environments versus physical built environments: A benchmarking study for building design and user-built environment explorations. *Automation in Construction*, 54, 116–126. doi:10.1016/j.autcon.2015.03.020
- Holzer, D. (2015). *The BIM manager's handbook: guidance for professionals in architecture, engineering, and construction*. Chichester, UK: John Wiley & Sons.
- Hosseini, M. R., & Chileshe, N. (2015). Emerging ICT-Based Methods in the Architecture, Engineering, and Construction Context. In K.-P. Mehdi (Ed.), *Encyclopedia of Information Science and Technology* (3rd ed.). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-5888-2.ch698
- Hosseini, M. R., Chileshe, N., Zuo, J., & Baroudi, B. (2012). Approaches for implementing ICT technologies within construction industry. *Australasian Journal of Construction Economics and Building: The Conference Series*, 1(2), 1–12.
- Howard, R., & Björk, B.-C. (2008). Building information modelling—Experts views on standardisation and industry deployment. *Advanced Engineering Informatics*, 22(2), 271–280. doi:10.1016/j.aei.2007.03.001
- Hubbard, B., Wang, H., Leasure, M., Ropp, T., Lofton, T., Hubbard, S., & Lin, S. (2015). *Feasibility Study of UAV use for RFID Material Tracking on Construction Sites*. Paper presented at the 51st ASC Annual International Conference. Retrieved from <http://ascpro.ascweb.org/chair/paper/CPRT367002015.pdf>
- Irizarry, J., Gheisari, M., & Walker, B. N. (2012). Usability assessment of drone technology as safety inspection tools. *Journal of Information Technology in Construction*, 17, 194–212.
- Khanzode, A., Fischer, M., Reed, D., & Ballard, G. (2006). *A Guide to Applying the Principles of Virtual Design & Construction (VDC) to the Lean Project Delivery Process*. Stanford, CA: Center for Integrated Facility Engineering, Stanford University.
- Krygiel, E., & Nies, B. (2008). *Green BIM: successful sustainable design with building information modeling*. Indianapolis, IN: John Wiley & Sons.
- Kunz, J., & Fischer, M. (2012). Virtual design and construction: themes, case studies and implementation suggestions (version 14). Stanford, CA: Center for Integrated Facility Engineering (CIFE), Stanford University.

- Lim, S., Buswell, R. A., Le, T. T., Austin, S. A., Gibb, A. G. F., & Thorpe, T. (2012). Developments in construction-scale additive manufacturing processes. *Automation in Construction*, 21, 262–268. doi:10.1016/j.autcon.2011.06.010
- Lin, T.-H., Liu, C.-H., Tsai, M.-H., & Kang, S.-C. (2014). Using Augmented Reality in a Multiscreen Environment for Construction Discussion. *Journal of Computing in Civil Engineering*.
- Lu, Q., Won, J., & Cheng, J. C. P. (2015). A financial decision making framework for construction projects based on 5D Building Information Modeling (BIM). *International Journal of Project Management*.
- Mignard, C., & Nicolle, C. (2014). Merging BIM and GIS using ontologies application to urban facility management in ACTIVE3D. *Computers in Industry*, 65(9), 1276–1290. doi:10.1016/j.compind.2014.07.008
- Nikolic, D., Jaruhar, S., & Messner, J. I. (2011). Educational Simulation in Construction: Virtual Construction Simulator 1. *Journal of Computing in Civil Engineering*, 25(6), 421–429. doi:10.1061/(ASCE)CP.1943-5487.0000098
- Nitithamyong, P., & Skibniewski, M. J. (2004). Web-based construction project management systems: How to make them successful? *Automation in Construction*, 13(4), 491–506. doi:10.1016/j.autcon.2004.02.003
- Nonami, K., Kendoul, F., Suzuki, S., Wang, W., & Nakazawa, D. (2010). *Autonomous flying robots: unmanned aerial vehicles and micro aerial vehicles*. Tokyo: Springer. doi:10.1007/978-4-431-53856-1
- Sardroud, J. M. (2015). Perceptions of automated data collection technology use in the construction industry. *Journal of Civil Engineering and Management*, 21(1), 54–66. doi:10.3846/13923730.2013.802734
- Sarshar, M., & Isikdag, U. (2004). A survey of ICT use in the Turkish construction industry. *Engineering, Construction, and Architectural Management*, 11(4), 238–247. doi:10.1108/09699980410547595
- Singh, V. (2014). BIM and systemic ICT innovation in AEC: Perceived needs and actors degrees of freedom. *Construction Innovation*, 14(3), 292–306. doi:10.1108/CI-02-2013-0006
- Succar, B. (2009). Building information modelling framework: A research and delivery foundation for industry stakeholders. *Automation in Construction*, 18(3), 357–375. doi:10.1016/j.autcon.2008.10.003
- Toole, T. M., Hallowell, M., & Chinowsky, P. (2013). A tool for enhancing innovation in construction organizations. *Engineering Project Organization Journal*, 3(1), 32–50. doi:10.1080/21573727.2012.717531
- Underwood, J., Isikdag, U., & Global, I. (2010). *Handbook of research on building information modeling and construction informatics: concepts and technologies*. Hershey, PA: Information Science Publishing. doi:10.4018/978-1-60566-928-1
- Van der Vlist, A. J., Vrolijk, M. H., & Dewulf, G. P. M. R. (2014). On information and communication technology and production cost in construction industry: Evidence from the Netherlands. *Construction Management and Economics*, 32(6), 641–651. doi:10.1080/01446193.2014.911932
- Wang, X., Truijens, M., Hou, L., Wang, Y., & Zhou, Y. (2014). Integrating Augmented Reality with Building Information Modeling: Onsite construction process controlling for liquefied natural gas industry. *Automation in Construction*, 40, 96–105. doi:10.1016/j.autcon.2013.12.003
- Williams, T. (2007). *Information technology for construction managers, architects and engineers*. Clifton Park, NY: Thomson Delmar Learning.

Yee, P., Fischer, M., & Kam, C. (2013). Prospective validation of virtual design and construction methods. [ITcon]. *Journal of Information Technology in Construction*, 18, 214–239.

Zuppa, D., Issa, R. R., & Suermann, P. C. (2009). BIM's impact on the success measures of construction projects. *Computing in Civil Engineering*, 503-512.

ADDITIONAL READING

Becerik-Gerber, B., & Kensek, K. (2010). Building Information Modeling in Architecture, Engineering, and Construction: Emerging Research Directions and Trends. *Journal of Professional Issues in Engineering Education and Practice*, 136(3), 139–147. doi:10.1061/(ASCE)EI.1943-5541.0000023

Bryde, D., Broquetas, M., & Volm, J. M. (2013). The project benefits of building information modelling (BIM). *International Journal of Project Management*, 31(7), 971–980. doi:10.1016/j.ijproman.2012.12.001

Chen, K., Lu, W., Peng, Y., Rowlinson, S., & Huang, G. Q. (2015). Bridging BIM and building: From a literature review to an integrated conceptual framework. *International Journal of Project Management*, 33(6), 1405–1416. doi:10.1016/j.ijproman.2015.03.006

Kensek, K. M. (2014). *Building information modeling: BIM in current and future practice*. Hoboken, New Jersey: John Wiley & Sons Inc.

Khanzode, A. (2010). *An Integrated, Virtual Design and Construction and Lean (IVL) Method for Coordination of MEP*. Stanford, CA: Center for Integrated Facility Engineering, Stanford University.

Smith, D. K., & Tardif, M. (2012). *Building Information Modeling: a Strategic Implementation Guide for Architects, Engineers, Constructors, and Real Estate Asset Managers*. Chichester: John Wiley & Sons.

Sun, C., Jiang, S., Skibniewski, M. J., Man, Q., & Shen, L. (2015). A literature review of the factors limiting the application of BIM in the construction industry. *Technological and Economic Development of Economy*, 1–14. doi:10.3846/20294913.2015.1087071

KEY TERMS AND DEFINITIONS

Construction Industry: The term represents architecture, engineering, and construction i.e. all sectors that provide architectural design, engineering design and construction services.

ICT: Stands for information communication technology and concerns any technology or product associated with storing, retrieving, manipulating, transmitting, or receiving information electronically in a digital nature.

ICT in the Construction Context: This covers goods (such as office software suite and communication equipment), services and management (such as telecommunication including transmission and display) adopted to generate growth and improve productivity in the construction context.

Innovation: Innovations for the construction context represent the profitable developments and implementation of novel ideas, procedures, products, and practices geared towards enhancing organisational performance.

Amplifying the Significance of Systems Thinking in Organization

B**Mambo Governor Mupepi***Grand Valley State University, USA***Sylvia C. Mupepi***Grand Valley State University, USA***Jaideep Motwani***Grand Valley State University, USA*

INTRODUCTION

The reality in organization is that learning never stops. People in the organization can be taught to think alike about the job at hand. Systems thinking is pervasive in highly productive organizations. Effective teams can be well-coordinated in their specific roles in the value creation system. Job specifications and descriptions are akin to tele-prompters in aiding individuals to do the right thing first time and always. They are part of the documentation necessary in successful organizations. Individuals groups and teams must learn their roles and understand what gives life to the organization. The sociotechnical conceptual model provides a description of perceived reality. It gives a systemic content of inquiry and views the organization as an open system model, a purposeful conceptual part or whole organism. Koestler (1967) coined the term *holon* to imply something that could be described simultaneously as a whole or part. Divisions or units in organization can be viewed as *holon* because they can be corporations in their own right and at the same time they can be *owned* by another corporation.

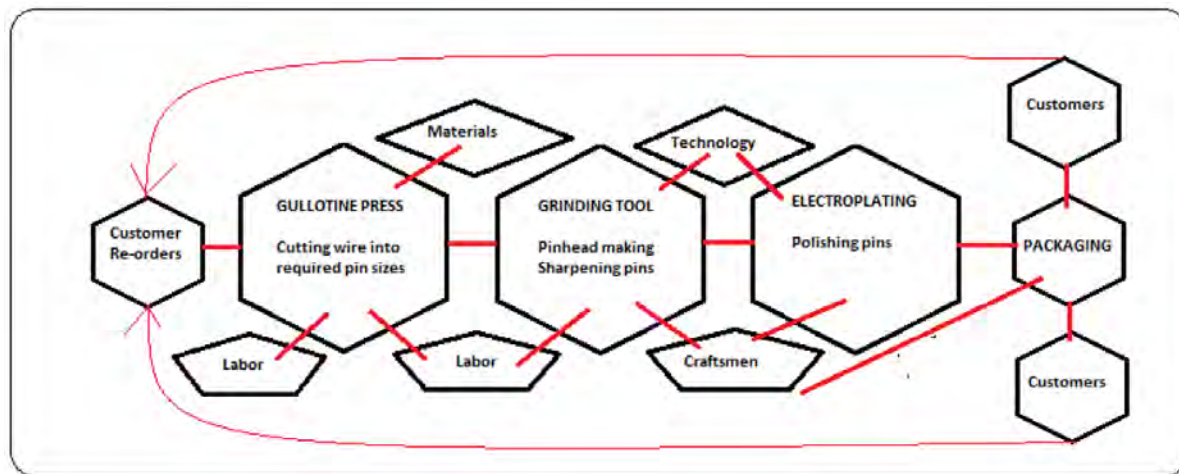
An Interdisciplinary Origin

A system is viewed as a whole whose elements hang together because they continuously affect each other over time and operate toward a

common purpose. In Stichweh (2010), systems theory is defined as an interdisciplinary study of systems in general, with the goal of discovering patterns and algorithms necessary in maintaining the value creation system to produce the goods and services demanded by the customers. Stichweh elucidates the definition of systems theory as an understanding related to the General System Theory of biologist Ludwig von Bertalanffy (Bertalanffy, 1934), physiologists Walter B. Cannon, Walter Pitts, Warren McCulloch and cybernetics technologists Claude Shannon, Norbert Wiener, and William Ross Ashby. In Daft (2013) a group of technocrats from the same organization referred to as skunkworks meet regularly in defined physical places and in cyberspace to progress organizational goals. Daft argued that the skunkworks were a people who shared the same vision about their organization. They deliberated in designing new products and appropriate technology to make those new products a reality. In Gharajedaghi (2010) systems thinking is deployed in the search for solutions to the problems the organization faces such as competition. How to effectively compete can be framed by co-constructing a shared vision of the future. Systems thinking can be applied to design and implement the competitive advantage.

DOI: 10.4018/978-1-5225-2255-3.ch048

Figure 1. Adam Smith's value creation open system where customers come back for more all the time
 Source: Mupepi, 2017



BACKGROUND

Action Research: A Lewinian Approach

The sociotechnical systems is founded on action research (Lewin 1946) as its basic process of inquiry or purposeful human activity *holon* (Eijnatten, 1993). The Adam Smith Pin Factory is revisited to illustrate how the novice can increase their skillfulness and explicit knowledge to become specialists. The conceptual framework provides an opportunity to envision the techniques and tools essential in developing the explicit practices required in effective organization. Following the work of Adam Smith in Mupepi (2014), structures are viewed as open systems molded and determined by values and norms. They also contain job descriptions job specifications and techniques needed for the structure to make meaning contribution in the value creation process. The pin-production open system provides an opportunity to understand the explicit knowledge needed to increase output (see Figure 1). In Systems thinking the organizational structure is viewed as the pattern of interrelationships among key actors. It includes the value

creation process and hierarchical chart, attitudes and perceptions, quality of products, ways in which decisions are made and how things are done. Systems dynamics demonstrate the interdependence of each position and task, and how they are all related to whole production department or *holon*. The art of sociotechnical systems thinking includes learning to recognize the ramifications and trade-offs of the actions members choose to follow (see Figure 1).

The historical development of sociotechnical systems design (STSD), can be argued to have roots in London's Tavistock Human Relations Institute before WW2. Trist & Bamforth (1951) attribute the first studies conducted by Fred Emery in the British coal fields in Yorkshire and Norwegian industries (Emery & Trist, 1981). These studies resulted in the development of effective STSD approach as a collaboration effort of many scholars including Lewin, Lippitt & White (1939), Bion (1949), Cummings (1978), and Eijnatten & Hoevenaars (1989), among many others. Sociotechnical systems thinking is espoused from STSD and the systems theory developed by von Bertalanffy (1969) and Lewin (1946) respectively (Eijnatten 1993).

Group Dynamics

Kurt Lewin (1943) is commonly identified as the founder of the movement to study groups scientifically. He coined the term *group dynamics* to describe the way groups and individuals act and react to changing circumstances. In Backstrom, Huttenlocher, Kleinberg & Lan (2006), group dynamics is a system of behaviors and psychological processes occurring within a social group (intragroup dynamics), or between social groups (intergroup dynamics). The study of group dynamics can be useful in understanding decision-making behavior, how the job is thought out and implemented and tracking the value chain to appreciate bottlenecks in efforts to increase productivity. In Bion (1949), alludes to idea that the software of a group's mind can be seen in thinking alike or systems thinking. Wilfred Bion's observations about the role of group processes in group dynamics are set out in a book entitled: *Experiences in Groups* first published in 1962. In this book Bion refers to recurrent emotional states of groups as basic assumptions. Bion suggested that the work group was interested in the aspect of group functioning which had to do with the primary task of the group—what the group has formed to accomplish. The primary group kept the organization anchored to a sophisticated and rational level of behavior. The secondary group was concerned with sustenance by the leader. The group was dependent on its leader for material and spiritual resources. In later research Salas & Diazgranados et al (2008), posit that groups could focus on aligning around goals, building effective working relationships, reducing team members' role ambiguity and finding solutions to team problems. Salas & Diazgranados add to the group dynamics discourse by adding performance factors: setting goals, role clarification, problem solving, and interpersonal relations to increase organizational effectiveness.

Transformative Small Groups



Many on-job training techniques have been adapted from the research conducted in many places including Tavistock Institute and T-Groups at National Training Laboratories at Bethel, Maine (Gallant & Rios, 2014). Team-based learning represents an even more intense use of small groups to craft the core capability of the organization. As team members become more willing to commit to very high level of effort in their learning they develop systemic way of seeing things, and techniques to solve the problems they encounter in progressing their tasks. In Gallant & Rios, a contention is made about building teams that can perform beyond the capability of even their most talented members. As groups think more about the task at end, the more innovative they can be (Lockheed, 2016).

The Sociotechnical Perspective

The historical background illustrates that the STSD can boast of a long tradition rooted in famous and successful intervention of the Tavistock Institute of Human Relations that solved what could have been a chronic productivity problem in the British pre-WW2 coal industries. The concept of structure was deployed with Adam Smith to define the organizational culture and the efficacies essential in making successful specialists (see Figure 1). In Blau et al (1976), it is postulated that during the fifties and sixties Weber's distinction between the categories of *goal-oriented* behavior and *value-oriented* behavior were still regarded as being of fundamental importance. For example the learner had to develop his skills by observing and taking orders on what to do from the expert. It implied that the division of labor could be viewed as a structure whose rationality was to increase productivity by allowing key actors to become specialists. Blau et al argued that structure was

related to behavior drawn from sociology and economics. The division of labor was the rationale to increase output and innovation became a reality as each divided labor thought about effective ways of doing the job as a means to earn better wages. In much later studies Occelli (2017), argued that the advancements in information communication technologies (ICTs), and mobile computing, has created an information-rich environment where new types of socio-technical systems (STS) can be established. Due to the pervasiveness of ICT in the value creation systems designing and implementing socio-technical systems continue to raise an increasing interest in organizations. They play a crucial role in the improvement of productivity. Occelli suggested that STS were essential in leveraging technology and talent management in competitive environments.

In Beal (2016), the ERP is short for enterprise resource planning. An ERP is a socio-technical system in the sense that it is a tool or know-how deployed to make the job easy in many aspects of organization including human resources or manufacturing management. The enterprise resource planning (ERP) is business process management software that allows an organization to use a system of integrated applications to manage the business and automate many back office functions related to technology, services and human resources. The ERP software integrates all facets of an operation — including product planning, development, manufacturing, sales and marketing — in a single database, application and user interface. It is a socio-technical system that engages management to effectively plan and implement human resources strategy pertaining to recruitment and selection, hiring, training and develop and compensate and retain talent in useful; organization. In Kasemsap (2015), the ERP has emerged as one of the major breakthrough information technologies that can re-shape the manufacturing industry. The ERP is the business application that weaves together all the data within an organization's business processes and associated functional areas. Kasemsap considered the advent of globalization

and the appearance of new forms of organization based on networks of closely cooperating firms, it seems clear that successfully implementing sociotechnical systems such as ERP will take on an increased significance for the survival, growth, and competitiveness of many small and medium-sized enterprises (SMEs).

The Sociotechnical Systems Thinking Proposition

In Miller & Rice (1967), the argument that sociotechnical systems thinking was rooted in industrial democracy and the design of tasks compatible with generic human needs is made. In later research Davis Challenger Jayewardene & Clegg (2014), posited that sociotechnical systems thinking can be viewed as a subset of social theory and philosophy. This way of thinking is particularly relevant in domains such as organizational development which are closely related to knowledge management. Socio-technical systems thinking has predominantly been applied to the domains of new technology and work design in postwar industries. Whilst it has made an impact, Davis et al argue that they need to be valiant, encouraging the approach to evolve and extend its reach. In particular they suggested that the conceptualization of what constituted a system applied in the organization's thinking. To illustrate the claim, Davis et al provided examples of socio-technical perspectives on the management of change and how the approach can be applied to increase in productivity. They argued that there was a positive correlation between sociotechnical systems thinking and reduction in risk resulting in increased yields.

The participative view in sociotechnical systems thinking imply continuous learning and improvement and use of appropriate technology. An assumption can also be made that the key actors in the value chain must perfect their act to remain relevant to the job at hand. In Senge (1990), system thinking encompasses a large and fairly amorphous body of methods, tools, and principles

all oriented to looking at the interrelatedness of forces and seeing them as part of a common process. In Daft (2013), a group of technocrats referred to as skunkworks can meet regularly to solve organizational problems. Their main asset is what they know, and how they coordinate their expertise to create organizational advantages. They examine interrelated forces and see them as part of a common process. Senge asserted that systems thinking follow certain common principles, the nature of which are being discovered and understood by all in the organization. In Mupepi (2017b), the ultimate success or failure of a business in modern society depends on a variety of factors across all levels of the organization. By utilizing dynamic human resource planning techniques and operational teams can lead organization more efficiently to reach desirable goals. Appreciating the configuration and combination necessary in socio-technical systems is pivotal to maintaining and increasing the competitive advantage. For example, the Skunkworks philosophy has been tested successfully at Lockheed Martin Aerospace industry (Lockheed, 2016) and its thesis can be applied to create diffuse and distribute the explicit knowledge required to differentiate the organization in competitive environments successfully. The Skunkworks has continued to be a combination of highly qualified and experienced aeronautical engineers who designed and built one of the most effective jetfighter plane in use today. Daft suggested that the competitive advantage can be created successfully by a people who share the same thinking and mindset. The Skunkworks were able to be innovative in all their deliberation to produce a product that continue to be in demand internationally.

Learning and Improving All the Time

Williams & Hummelbrunner (2010) explore the application of systems ideas to investigate, evaluate, and intervene in complex highly productive organizations. Their proposition serves as a field guide to how sociotechnical systems thinking can

be implemented to take organizations to the next level. Williams & Hummelbrunner describe and analyze organizational learning, and how change can be designed and implemented successfully. Williams & Hummelbrunner places importance on how the sociotechnical systems concept can be put into useful practice that focus on developing system thinking and a shared mental model about how best to do the job at hand.

In Mupepi (2017a), a co-constructed vision of the future is made possible by an organization that can come together to strategize using the same heuristics to solve problems and enable continuous learning and improvement to happen in all facets of organization.

Organizations as Open Systems

In Soliman (2014), System theory provides a first perspective for critical and normative exploration. Systems theory has been deployed successfully to elucidate phenomenon in many fields including organization development cybernetics, and economics, among many others (von Bertalanffy 1969:2). Soliman (2014) posited that the centrality of systems thinking is to view a phenomenon holistically as a set of diverse interacting elements. Mingers & White (2010) suggested that the relationship between the components is more meaningful than investigating each component in isolation. Vancouver (1996), suggested a hierarchical structure of a system enabled accountability and analysis of responsibility. Drawing on this early work Soliman (2014), considered that there were generic characteristics that applied to all open systems such as the importation of energy. In this acquisition all open systems are dependent on their environment for the resources they need to effectively function. Bertalanffy (1934) employs evolutionary biology to explain how the organism amoeba replicates itself to successfully thrive in marshy environments. The amoeba draws energy from the plants, water, sunlight, and air as throughput to replicate itself successfully. The absorbed energy will be transformed by the amoeba open

Table 1. Hierarchy of systems

Level	Characteristics
Framework	Static descriptive (e.g. anatomy of the universe, gene; the patterns of electrons around a nucleus)
Clockwork	Simple & predetermined motion; dynamic, stochastic, equilibria
Control	Self-regulation; Transmission & interpretation of information
Open system	Self-contained, reproduction, throughput of material (e.g. cell and river)
Genetic-Societal	Division of labor, differentiated and mutually dependent parts, Blue prints for growth
Animal	Increased mobility, teleological behavior, Self-awareness, specialized information-receptors
Human	Self-consciousness, self-reflexive quality, ability to produce, absorb, and interpret symbols
Social organization	Value system, Transcription of images
Inspiring	Ultimates, absolutes, and the inescapable unknowables

Source: Boulding, 1956

system to continue to grow. Organization will transform the resources and materials into products or services. In the evolutionary cycle the output of the organisms will be exported back to the environment and thus the amoeba continues to replicate itself faster. Soliman argues that systems are cycles of events characterized by inputs and outputs. The offering of the products or services will generate new resources for organization in order to repeat its transformation process.

Open systems are bound to obtain negative entropy. Entropy as a universal law is the degree of disorder or uncertainty in a system. It states that all organizations will eventually result in disorganization or death. Thus, open systems tend to maximize their ratio of imported and consumed energy to maintain survival (von Bertalanffy 1969:3) Information input, negative feedback, and the coding process are some the characteristics of reproduction in organisms. Open systems not only absorb energy as input but also information. Negative information will provide feedback to the system about its operation and potential need of corrective action.

The Need for Homeostasis

Homeostasis is the property of a system in which a variable is actively regulated to remain very nearly constant. This regulation occurs inside a defined environment. Open systems that are successful in

importing energy to obtain negative entropy are epitomized by a steady state. Negative entropy has been used by biologists as the basis for purpose or direction in life, namely cooperative or moral instincts. However, as the environment of open system will change accordingly. Dynamics homeostasis predicts this process of adjustment (von Bertalanffy 1969:4)

Differentiation

In Barney (1991), firms are viewed as organizations that have different resource endowments that enable them to construct specific competitive advantages over competitors. Resource endowments allow firms to be different which reduces competition and makes it possible to reach new segments of the market. Thus, differentiation is the process of distinguishing the differences of a product or offering from others, to make it more attractive to a particular target market. Barney argued that differentiation is imperative in all open systems to allow specialization to occur and justify their existence and maximize efficiency.

Boulding (1956) provided a dual basis to treat organizations as open systems as well as social systems with the objective to attain a distinctive goal that sets it apart from other social systems. Human beings as open systems were superior among all organism for their capacity to think and reason.

Systems Thinking in Environmental Uncertainty

In Mingers & White (2010), perceived environmental uncertainty has long been employed to describe an objective state of an organization's environment or as a state of an individual's perception about the environment. Attributing certain characteristics to the environment enables organizations to explain and prepare for uncertainty. The adage: forewarned is forearmed makes organizational sense when risk can be planned and anticipated (Milliken 1987).

There are three types of uncertainty and these are: state uncertainty, effect uncertainty, and response uncertainty. State uncertainty is when an individual lacks information about the nature of the environment on how environmental events, changes, or set of changes will affect the particular organization. With response uncertainty an individual experiences an inability to predict accurately what the organization's response options are in accomplishing desired outcomes. Subsequently it is the type of information an individual lacks that determines which type of perceived environmental uncertainty s/he experiences. State uncertainty is conceptually the closest to using the term environmental uncertainty to describe the state of an organization's environment (Milliken, 1987:121).

Feedback

Argyris (1990) postulated about double loop learning and how organizations could improve productivity by giving each other constructive feedback. Argyris referred to this phenomenon about learning as the ladder of inference. Argyris suggested that the organization's ability to achieve desirable outcomes was truly eroded by feelings that generated ideas such as: our beliefs are the only truth, our beliefs are based on real data, or the data we select are the real data. Argyris & Schon (1978), suggested that double loop learning provided the learner with appropriate feedback

which enabled the learner to draw useful lessons to perfect his act. Double-loop learning occurred when error an error was detected and corrected in ways that involved the modification of an organization's underlying norms, policies and objectives.

RECOMMENDATIONS

Learning to see the organization systematically is something intangible as deeply set attitudes found in many companies. There are four levels of a systems view. The first is made up of the organization's notable events such as success stories. How the team met the sales targets in a given period of time can be employed to remind current employees that targets can be achieved if there is teamwork (see Figure 2).

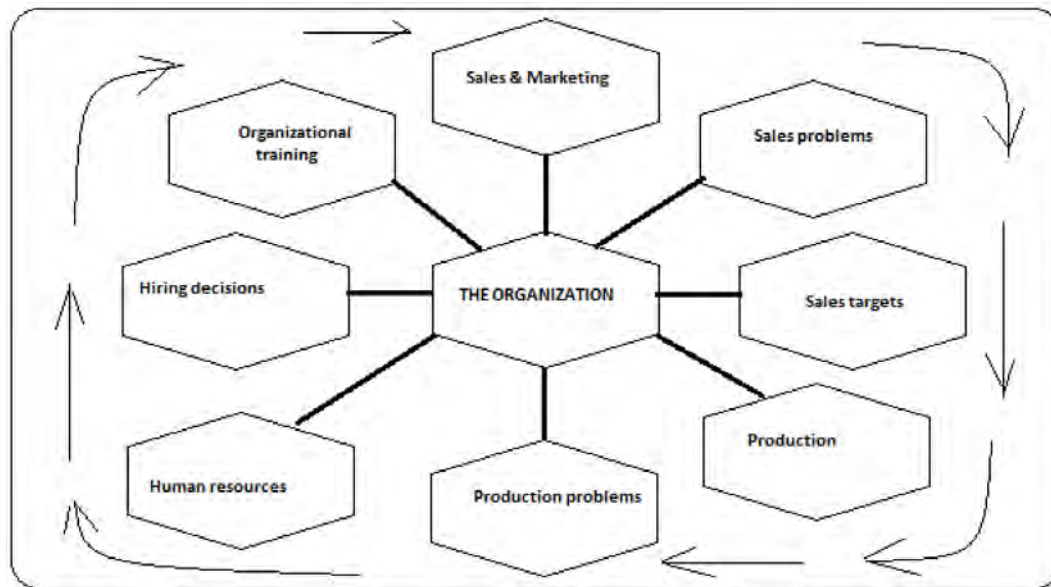
The second perspective concerns patterns of behavior (see Figure 2). Drawing lessons from trends of the system over time is important in charting the change discourse. Instead of listing isolated historical events forecasting can be deployed to chart the future with some degree of accuracy.

The third is about the systemic structure. This activity examines the interrelatedness of the events to the system and how the parts are connected to make the whole *holon*. The group can discuss the key interrelationships between these parts factors and events to understand if there are any causal relationships (see Figure 2).

The fourth part is the mental model. The mental models are tacit, they relate to experience. Two people with different experiences can relate to the same object differently. Mental models shape how people behave. The core task of the fourth part is to talk about the job candidly. There are mental models that should not be sustained. For example managers may believe that poor quality is a sabotage by the hourly paid workers. What needs to be encouraged is the analysis of the system with its other parts to understand the causes of the problem of poor quality (see Figure 2). Mental models are also critical in designing and implementation of successful organizational training and development programs.

Figure 2. Organizational interrelated factors impact the core competency

Source: Mupepi, 2017



Sociotechnical Systems Thinking Is Good for Business

In Beal (2016), the last couple of years have unleashed forces which are fundamentally shifting behavior in the use of sociotechnical systems such as the ERP. The industry has new and continuing trends such as the development of the following software:

- **Mobile ERP:** The business community want real-time access to information, regardless of where they are. It is expected that businesses will embrace mobile ERP for the reports, dashboards and to conduct key business processes.
- **Cloud ERP:** The cloud has been advancing steadily into the enterprise for some time, but many ERP users have been reluctant to place data cloud. Those reservations have gradually been evaporating, however, as the advantages of the cloud become apparent.
- **Social ERP:** There has been much hype around social media and how important

– or not -- it is to add to ERP systems. Certainly, vendors have been quick to seize the initiative, adding social media packages to their ERP systems with much fanfare. But some wonder if there is really much gain to be had by integrating social media with ERP.

- **Two-Tier ERP:** Enterprises once attempted to build an all-encompassing ERP system to take care of every aspect of organizational systems. But some expensive failures have gradually brought about a change in strategy – adopting two tiers of ERP (Beal, 2016).

LIMITATION

Sociotechnical systems thinking approach in management can be limited to many aspects associated with teams. Creating a shared mindset among groups of workers may require building team work. In this process effectiveness can be enhanced when the group does the similar jobs where teamwork can become meaningful. Creat-

ing a shared mindset can require a lengthy period of time. People need time to reflect and build effective interpersonal communications among themselves. Failure to achieve a sociotechnical systems thinking can result in frustration among members of the organization.

FUTURE RESEARCH DIRECTIONS

Future research should endeavor to study more contextual factors such as organizational culture useful in socio-technical organization systems (see Mupepi 2017a). Surveys distributed to customers can be analyzed to produce performance metrics necessary in increasing output (see Mupepi (2017b). The same data can be analyzed to design and implement exploitable databases which be interpreted to understand the responsiveness of customers and behavior in competitive environments.

CONCLUSION

Organizational learning undergirds sociotechnical system thinking. Constructive feedback is essential in continuous learning and improvement. Feedback is a common mental pathway of increasing abstraction and can often mislead individuals to misguided beliefs such as the sabotage of quality by the hourly paid workers.

REFERENCES

- Argyris, C. (1990). *Overcoming Organizational Defenses: Facilitating Organizational Learning*. Boston: Allyn and Bacon.
- Argyris, C., & Schön, D. (1978). *Organizational learning: A theory of action perspective*. Reading, MA: Addison Wesley.
- Backstrom, L., Huttenlocher, D., Kleinberg, J., & Lan, X. (2006). Group formation in large social networks. *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '06*. doi:10.1145/1150402.1150412
- Beal, V. (2016). *ERP - enterprise resource planning*. Accessed 11/14/16 <http://www.webopedia.com/TERM/E/ERP.html>
- Bertalanffy, L. v. (1969). *General System Theory*. New York: George Braziller Publishers.
- Bion, W. R. (1949). Experience in groups, III. *Human Relations*, 2(1), 13–22. doi:10.1177/001872674900200102
- Bion, W. R. (1962). A theory of thinking. *The International Journal of Psycho-Analysis*, 43.
- Blau, P. M. (1976). *Approaches to the study of social structure*. London, UK: Open Book Publishing.
- Boulding, K. E. (1956). General systems theory--the skeleton of science. *Management Science*, 2(3), 197–208. doi:10.1287/mnsc.2.3.197
- Cummings, T. G. (1976). Sociotechnical systems: an intervention strategy. In W. W. Burke (Ed.), *Current issues and strategies in organization development* (pp. 187–213). New York: Human Science Press.
- Daft, R. (2013). *Management*. London, UK: Cengage Learning.
- Davis, M. C., & Challenger, R. (2014). Advancing socio-technical systems thinking: A call for bravery. *Applied Ergonomics*, 45(2), 171–180.
- Eijnatten, F.M. van & Hoevenaars, A.M. (1989). Modern sociotechnique in the Netherlands. *Ge-drag en Organisatie*, 2(4/5), 289–304.
- Eijnatten, F. M. van. (1993). *The Paradigm that Changed the work place*. Stockholm, Sweden: the Swedish Center for Working Life.

- Emery, F. E., & Trist, E. L. (1981). The causal texture of organizational environments. In F. E. Emery (Ed.), *Systems Thinking* (Vol. 1). London, UK: Penguin Books.
- Gallant, S., & Rios, D. (2014). The organization development consulting process. In B. B. Jones & M. Brazzel (Eds.), *The NTL handbook of organization development and change* (2nd ed.; pp. 153–174). San Francisco, CA: Wiley. doi:10.1002/9781118836170.ch8
- Gharajedaghi, J. (2010). *Systems Thinking: Managing Chaos and Complexity a Platform for Designing Business Architecture* (3rd ed.). London, UK: Elsevier.
- Kasemsap, K. (2015). Implementing Enterprise Resource Planning. In M. Khosrow-Pour (Ed.), *Encyclopedia of Information Science and Technology* (3rd ed.; pp. 798–807). Hershey, PA: IGI Global; doi:10.4018/978-1-4666-5888-2.ch076
- Koestler, A. (1967). *The Ghost in the machine*. London, UK: Hutchinson.
- Lewin, K. (1943). Defining the “Field at a Given Time”. *Psychological Review*, 50, 292-310.
- Lewin, K. (1946). Research on minority problems. *Technology Review*, 48(3).
- Lewin, K., Lippitt, R., & White, R. K. (1939). Patterns of aggressive behavior in experimentally created social climates. *The Journal of Social Psychology*, 10(2), 271–299. doi:10.1080/00224545.1939.9713366
- Miller, E. C., & Rice, A. K. (1967). *Systems of organization: the control of task and sentient boundaries*. London, UK: Tavistock Publications.
- Milliken, F. J. (1989). Three types of perceived uncertainty about the environment: State, effect and response uncertainty. *Academy of Management Review*, 12(1), 133–143.
- Mingers, J., & White, L. (2010). A review of the recent contribution of systems thinking to operational research and management science. *European Journal of Operational Research*, 207(3), 1147–1161. doi:10.1016/j.ejor.2009.12.019
- Mupepi, M. (2014). *Can the Division of Labor Be Re-Engineered to Advance Organizational Dynamism?* Sage Open. doi:10.1177/2158244014536404
- Mupepi, M. (2017b). *Effective Talent Management Strategies for Organizational Success*. Hershey, PA: IGI Global; doi:10.4018/978-1-5225-1961-4
- Mupepi, M. G. (2017a). Using Communities of Practice to Identify Competencies. In K. Rasmussen, P. Northrup, & R. Colson (Eds.), *Handbook of Research on Competency-Based Education in University Settings* (pp. 157–167). Hershey, PA: IGI Global; doi:10.4018/978-1-5225-0932-5.ch008
- Salas, E., Diazgranados, D., Klein, C., Burke, C. S., Stagl, K. C., Goodwin, G. F., & Halpin, S. M. (2008). Does Team Training Improve Team Performance? A Meta-Analysis. *Human Factors: the Journal of the Human Factors and Ergonomics Society.*, 50(6), 903–933. doi:10.1518/001872008X375009 PMID:19292013
- Senge, P. (1990). *The Fifth Discipline: The art and practice of the learning organization*. New York: Doubleday.
- Skunkworks, L. (2016). *Lockheed SR-71 Blackbird*. Accessed 11/14/16 <http://www.bing.com/images/search?q=skunk+works&view=>
- Stichweh, R. (2011). Systems Theory. In B. Badie et al. (Eds.), *International Encyclopedia of Political Science* (Vol. 8, pp. 2579–2582). New York: Sage. doi:10.4135/9781412959636.n597

Vancouver, J. B. (1996). Living systems theory as a paradigm for organizational behavior: Understanding humans, organizations, and social processes. *Behavioral Science*, 41(3), 165–204. doi:10.1002/bs.3830410301 PMID:8856949

Williams, B., & Hummelbrunner, R. (2010). *Systems concepts in action: A practitioner's toolkit*. New York: Stanford University Press.

ADDITIONAL READING

Mupepi, G. M., Mupepi, S. C., Tenkasi, R. V., & Sorensen, P. F. (2013). Creating High Impact Organization in the SADC. In J. Vogelsang & M. Townsend et al. (Eds.), *Handbook for Strategic HR: Best Practices in Organization Development from the OD Network* (pp. 624–631). New York: American Management Association.

Mupepi, Mambo. (2017). “Effective Talent Management Strategies for Organizational Success.” IGI Global, 2017. 1-302. Web. 15 Nov. 2016. doi:10.4018/978-1-5225-1961-4

Mupepi, M. G. (2017). Using Communities of Practice to Identify Competencies. In K. Rasmussen, P. Northrup, & R. Colson (Eds.), *Handbook of Research on Competency-Based Education in University Settings* (pp. 157–167). Hershey, PA: IGI Global; doi:10.4018/978-1-5225-0932-5.ch008

Mupepi, M. G., & Motwani, J. (2015). Deconstructing the Value Creation Process: Positioning Diversity to Increase Output. *International Journal of Sociotechnology and Knowledge Development*, 7(4), 15–30. doi:10.4018/IJSKD.2015100102

Mupepi, M. G., & Mupepi, S. C. (2014). Appreciating Rapid Technology Integration in Creating Value in Enterprises. *Journal of Electronic Commerce in Organizations*, 12(1), 53–75. doi:10.4018/jeco.2014010104

Mupepi, M. G., & Mupepi, S. C. (2016). Applying Theory to Inform Competency Development: Bootstrapping Epistemic Communities in Growing Specialists. *International Journal of Productivity Management and Assessment Technologies*, 4(1), 28–38. doi:10.4018/IJPMAT.2016010103

Mupepi, M. G., Taruwinga, P., & Al-Rabayah, W. A. (2017). Can the Usage of Social Media Increase the Gregariousness of the Family to Grow Successful Family-Owned Businesses?: The Usefulness of Social Media in Growing a Family-Owned Business. In W. Al-Rabayah, R. Khasawneh, R. Abu-shamaa, & I. Alsmadi (Eds.), *Strategic Uses of Social Media for Improved Customer Retention* (pp. 224–245). Hershey, PA: IGI Global; doi:10.4018/978-1-5225-1686-6.ch012

Soliman, F. (2014). Could Knowledge, Learning, and Innovation Gaps be Spiraling? In F. Soliman (Ed.), *Learning Models for Innovation in Organizations: Examining Roles of Knowledge Transfer and Human Resources Management* (pp. 1–29). Hershey, PA: IGI Global; doi:10.4018/978-1-4666-4884-5.ch001

Soliman, Fawzy (2016). *Role of Cloud Systems as Enabler of Global Competitive Advantages. Organizational, Legal, and Technological Dimensions of Information System Administration*. IGI Global, 2014. 120-138. Web. 15 Nov. 2016. doi:10.4018/978-1-4666-4526-4.ch008

KEY TERMS AND DEFINITIONS

Algorithm: An algorithm is a self-contained step-by-step set of procedures applied in solving mathematical or industrial problems.

Differentiation: The process of making products or organization different from others in competition.

Holon: Adapted from Hebrew implying a self-containing part which is part of a larger system.

Skunkworks: A project typically developed by a small and loosely structured group of people who research and develop a project primarily for the sake of innovation.

Systems Thinking: Systems thinking involves the use of various techniques to study systems of many kinds.

Ubiquitous: Omnipresence or ubiquity is the property of being present everywhere.

Value Creation System: A value creation system is a production perspective that describes social and technical resources employed to produce goods and services demanded by customers.

The Application of Crowdsourced Processes in a Business Environment

B**Katarzyna Kopeć***Tischner European University, Poland***Anna Szopa***Jagiellonian University, Poland*

INTRODUCTION

Globalization is creating tremendous opportunities as well as challenges for businesses. There can be observed increasing emphasis on maximizing creativity, human capital and problem solving. Companies are more likely to be innovative and shape ideas as well as to create attractive propositions for users or customers (Cox, 2005). On the other hand, consumers are becoming more like external employees who take over specific parts of a design process, whereby this process finally remains under the control of a company (Kleemann, Voß, & Rieder, 2008). The recent shift towards a polycentric perception of business making requires the involvement of stakeholders in an organization's management. This is a starting point for looking at the idea of crowdsourcing as an innovative business model. Crowd-based business models enable organizations to harness the collective energy, through different processes, companies gather a large population by inviting users to create value (Kohler 2015). The inclusion of stakeholders in the process of developing new products or services forms a general foundation of this notion. Entrepreneurs take advantage of crowdsourcing ventures due to the delegation of tasks to stakeholders in the form of an open call.

Organizations rely on crowds for different reasons, e.g. in order to create content, evaluate, create or to solve problems. The article's aim is to present crowdsourcing as innovative business model in the context of evolving consumer society based

on the OpenIDEO case. In addition it highlights the pros and cons of crowdsourcing as well as prospective research directions that might clarify some unidentified aspects of crowdsourcing.

BACKGROUND

Technological advance including Internet's development of the late 1990s and thus the wide recognition of web-dependent participatory culture in the 2000s (Brabham, 2013) has resulted in the inclusion of consumers into the process of creating new ideas also for business. In consequence, this link between enterprises and the groups of consumers has become more and more evident.

The idea of outsourcing a business task to the web-based community is a relatively recent invention, although it shows a close relationship with other deep-rooted concepts. The literature points out the catalogue of crowdsourcing-related notions such as prosumerism (Toffler, 1980), user-innovation (Hippel, 1988), open-innovation (Chesbrough, 2003), co-creation (Prahalad & Ramaswamy, 2004). These terms, however, overlap with crowdsourcing. The notion of crowdsourcing has experienced a great success in a variety of areas. Its evidence are blogs (e.g. www.crowdsourcing.com by Howe; <http://dbrabham.wordpress.com> by Brabham; <http://www.crowdsourcing-blog.org> by Estellés-Arolas), books (Howe, 2006, 2008; Tapscott & Williams, 2006, 2013; Surowiecki, 2004) including academic contributions (Brab-

ham, 2013; Chanal & Caron-Fasan, 2008; Pénin & Burger-Helmchen, 2011; Vuković, 2009; Estellés-Arolas & González-Ladrón-de-Guevara, 2012).

The term crowdsourcing (which is a blend of *crowd* and *outsourcing*) was coined and popularized by Jeff Howe in the 2006 *Wired* article (Howe, 2006). It is a real challenge to define clearly the idea of crowdsourcing as it is not a coherent term. The term has developed on the intersection of various disciplines, and thus a variety of approaches overlap here. It is also a marketing slogan of a rapidly growing recognition. Thus, scholar discourse mixes with popular media discourse which leads to “unkempt theory and practical crowdsourcing applications with shaky foundations” (Brabham, 2013, p. 6).

Howe (2010) states that “crowdsourcing is the act of taking a job traditionally performed by a designated agent (usually an employee) and outsourcing it to an undefined, generally large group of people in the form of an open call.”

Estellés-Arolas and González-Ladrón-de-Guevara (2012) in their article surveyed a vast number of existing crowdsourcing interpretations. They succeeded to develop an integrated definition of crowdsourcing based on 40 definitions identified in the literature so far. “Crowdsourcing is a type of participative online activity in which an individual, an institution, a non-profit organization, or company proposes to a group of individuals of varying knowledge, heterogeneity, and number, via a flexible open call, the voluntary undertaking of a task. The undertaking of the task, of variable complexity and modularity, and in which the crowd should participate bringing their work, money, knowledge and/or experience, always entails mutual benefit” (pp. 9-10). This definition points out the following key elements of crowdsourcing: an organization that has a task to be carried out, a crowd ready to complete a task voluntarily, an online environment enabling an organization to work with the crowd on a task and mutual benefits for both parties (Brabham, 2013).

Advantages and Disadvantages of the Crowdsourcing Process

Grün and Brunner (2002) state that one of key advantages for companies to announce crowdsourcing initiatives is cost reduction. It is possible when a business task can be outsourced to the customer via the Internet. To the pros of crowdsourcing belongs also the inclusion of users into the quality improvement and efficient use of resources. The crowdworkers can be co-designers, co-producers, testers, evaluators, and thus they can participate in product innovation. Reichwald and Piller (2006) see further advantages to involve customers in the value creation process that can be also adopted in the case of crowdsourcing: saving time for new product enhancement, wider acceptance of new products among customers, and strengthening consumers’ individual assessment of the new product’s originality. Kleemann, Voß and Rieder (2008) notice that companies encourage users to take part in crowdsourcing initiatives by presenting the distinctive, but intangible quality of the open content culture which emphasizes the perception of enterprises as community-orientated and creative entities. This is consistent with the domains in which crowdsourcing is commonly used: configuration, design and testing of products, innovation, and problem solution.

The users take part in crowdsourcing to develop one’s creative skills, network with other creative professionals, share with others or develop a portfolio for employment in future. Other crowdworkers are driven by the challenge to solve a puzzle, they want to contribute to a large project of common interest or they are lured by financial reward (which is the most frequent one). (Brabham, 2008a, 2010, 2013; Lakhani, Jeppesen, Lohse, & Panetta, 2007; Acar & van den Ende, 2011; Rogstadius, Kostakos, Kittur, Smus, Laredo, & Vuković, 2011).

Interestingly, Reichwald and Piller (2006) discovered that consumers actively participating

in product innovation are also driven by dissatisfaction with current solutions. They expect to support a company in product improvement and innovation.

It is important to remember that crowdsourcing communities are a subsection of broader online communities. Their typical activity is to respond to open calls posted by enterprises. In general, crowdworkers are members of a broader concept of participatory culture and believe their contributions matter (Jenkins, 2009).

The literature reveals also legal and ethical pitfalls that crowdsourcing may face. Felstiner (2010) found out that the legislative framework of crowdsourcing labor model is based on existing labor laws which does not address a variety of employment-related issues (status of a crowdworker, issue of control, employee's working equipment). It is almost impossible to flexibly revise the law due to the specific features of crowdwork – it is carried out online (in cyberspace), by anonymous crowdworkers, and governed by obligatory participation agreements (usually in the form of a pop-up window which informs that users are obligated to agree to the terms and conditions before using the software). Felstiner analyses the example of Amazon's Mechanical Turk, where its members can accomplish a series of "human intelligence tasks" (HITS) for small monetary rewards paid by the "Requester." In this crowdsourcing form workers become unquestionably low wages (an average crowdworker works eight hours per week on Amazon's Mechanical Turk for below \$2 per hour, for example a crowdworker for writing only a short definition of a new term may earn \$0.03 (Ross, Zaldivar, Irani, & Tomlinson, 2009; Ipeirotis, 2010). Brabham (2008b) states that the intellectual labor of the crowd is worth more than winning solutions are paid (in the literature referred to as "digital sweatshop," "virtual sweatshop" or "cyber-sweatshop" by Scholtz (2013). Wolfson and Lease (2011) highlight a subsequent legal problem of crowdsourcing initiatives, that is, copyright ownership and patent inventorship (e.g. joint inventorship) since it is rather difficult

to control creative works delivered by crowds. Another crowdsourcing-related risk is consumer data security as enterprises can attempt to reveal data about their crowdworkers to external parties to support research and development. In addition, crowdsourcing initiatives raise ethical concerns like substituting company's own employees with lower-wage crowdworkers what shows the exploitative nature of crowdsourcing (digital sweatshop, slave labor) for the enterprise's benefits (Harris, 2011).

CROWDSOURCING-BASED BUSINESS MODEL

In essence, a business model clarifies the process of how an organization creates and captures value and helps understand how a firm is embedded in and interacts with the surrounding ecosystem (Teece, 2010). The term crowdsourcing accentuates a novel business model to solve problems by leveraging the "collective intelligence" (Lévy, 1997) of crowds in which a major communication platform is based on the web. Crowdsourcing in the context of the new type of consumer can be explained through crowd wisdom, communicative planning and collective intelligence as groups are often smarter than the smartest people in them (Surowiecki, 2004).

To put it simply, an organization (crowdsourcer) announces a problem online, and individual members of an online community (crowdworkers) voluntarily submit their proposals to face an issue and to put forward a solution of a problem addressed. Eventually, an organization chooses one or the best-tailored solution for which a crowdworker gets a bounty (e.g. financial award, self-esteem, economic recognition, social recognition). What is more, an organization can use the right to employ this idea in its business activity. Schenk and Guittard (2009) point at this distinctive feature of crowdsourcing saying that a crowdsourcer takes advantage of going for an option that best meets his/her expectations, what means that he/she pays

for a product after its acceptance. In crowdsourcing business model the new consumer type is providing services usually for free, however, the capacities that he/she possesses are valuable economic assets. He/she takes part in the production process and is perceived as value-adding worker. The crowdworker is systematically integrated into corporate structures, and his/her activities are supervised by project managers as if he/she was an employee.

The intensity of involving groups of customers into an enterprise's operation is in crowdsourcing much more advanced than in any previous business models. Brabham (2013), as one of very few scholars, very clearly draws the line between activities that belong to crowdsourcing and those that cannot be included into this concept. It is of great importance as the popular media definitions which are situated outside the scholar literature blend all tendencies into one term crowdsourcing. Brabham points out the following features that differentiate crowdsourcing from similar creativity-based processes: top-down management, shared process of bottom-up creativity, locus of control with regards to the creation process between an organization and the public. It is important to notice that the concept of open source (e.g. Linux) and commons-based peer production (e.g. Wikipedia) are by Brabham (2013) excluded to be perceived as crowdsourcing (no top-down management, locus of control resides among the users).

Crowdsourcing Taxonomy

There are few approaches to classify crowdsourcing, including one most transparent taxonomy by Brabham (2013) presenting four dominant crowdsourcing types in business operation:

1. **Knowledge Discovery:** This crowdsourcing type is similar to commons-based peer production (e.g. Wikipedia), where users create online content. In this crowdsourcing case the organization manages the process of knowledge discovery (what information
2. **Broadcast Search Approach:** The focus is put here on searching scientific solutions by selected problem-solvers. Example: OpenIDEO – a platform where individuals work out challenges devised by OpenIDEO customers (for details see case study below). Based on a statistical analysis Lakhani et al. (2007) found out that solvers were able to solve 29% of the problems of the seekers (large companies with R&D departments) that had remained unsolved before. In the broadcast search approach, financial remuneration is commonly designed for problem solvers whose solution wins (though financial incentive is in this case not the only motivation to participate in crowdsourcing) and the intellectual property belongs to a seeker.
3. **Peer-Vetted Creative Production Approach:** The creation process is open to online community and finally the best option is selected. This approach can be applied in the matters of user preferences, also one's business can be based on crowdsourcing strategy. Example: Threadless.com, online company selling t-shirts based on design provided by artists (artists who provide winning designs are presented cash bounty). The best designs (of the month, of the week) to be produced are selected in a voting process.
4. **Distributed Human Intelligence Tasking:** Appropriate approach when the data is available but the problem includes data processing which can be split into small tasks done by the humans only. It is the least creative task, and financial reward is often a stimulus to

is needed, its objective, ways to collect this knowledge). Example: SeeClickFix, a web and phone application that allows citizens to report non-emergency problems (potholes, graffiti etc.) in their local community. City governments use SeeClickFix as the mechanism to better understand issues facing the community and to better allocate resources to fix problems.

participate. Example: Amazon Mechanical Turk. It is an online platform that essentially coordinates sets of effortless jobs which cannot be provided by computers as they require human intelligence. In this model “requesters” (employers) solicit jobs (usually microtasks) like tagging images, responding surveys, rewriting texts, transcribing audio, conducting Internet research etc. to be accomplished by “providers” (workers).

In general, crowdsourcing has been made possible thanks to the evolution of information and communication technology which supports innovation and development of enterprises. There are several ways enterprises can apply crowdsourcing strategies. Aitamurto, Leiponen and Tee (2011) classify internal crowdsourcing initiatives (e.g. IBM innovation jams which take place internally within the IBM various divisions), external crowdsourcing initiatives (e.g. LEGO Group’s LUGNET or DELL’s IdeaStorm that enables users to publish their product-related ideas or product improvements) as well as collaboration initiatives that are outsourced to companies specializing in crowdsourcing innovation problems to the public (e.g. InnoCentive.com). However, the crowdsourcing application is not restricted only to companies. They also can be implemented by non-governmental organizations that also take advantage of crowdsourcing strategy (e.g. the US Federal Transit Administration project which involved neighborhood’s residents to design a better functioning bus stop in Salt Lake City, Utah).

OpenIDEO’s Application of Collective Problem Solving Model

OpenIDEO is an open innovation platform existing within the IDEO company structure. IDEO is a design company located primarily in Palo Alto, California. It was established in 1991 from existing design firm (Taylor, 2005). This success company is ranked as one of the most innovative

companies in the world by business leaders in a global survey by Boston Consulting Group, (IDEO, 2015).

The crowdsourcing process in OpenIDEO starts with defining problem, which then is divided into several themes published on the website. Defined problem is divided into four phases. At the first phase – Inspiration – the crowd educates itself on the larger question, both informing and inspiring. Suggestions are voted, the so-called applauding process elevates the highest voted submissions to the user in the next phase. Then at the Concepting stage all solutions are collected to the larger problem. At this stage using the crowd voting results, the company cooperates with the Challenge sponsor to narrow the list of concepts to twenty. In the third phase, that is, Evaluation, the crowd again submits ideas as comments to refine and expand the winning concepts. Furthermore, each comment stands also a subject to the applauding process and motivates the author of the concept to acknowledge and include comments that receive frequent crowd approval. Then the crowd votes on the refined concepts, and OpenIDEO once again cooperates with the Challenge sponsor to select a final list of winning concepts that the sponsor implements. Presented process broadly comprises company’s problem-solving platform, which repeats the crowdsourcing process in each phase. At the final stage, Realisation, discussed project is implemented into real life. OpenIDEO also profits from having multiple solvers scattered across the landscape, but in addition allows them to exchange information and construct solutions together. (OpenIDEO, 2015). The main idea of presented crowdsourcing process is to allow the crowd to create solutions in each stage, and at the same time consequently making progress toward creating a comprehensive solution - its approach is to populate the same landscape with a small number of contributors and endorse initial solutions’ improvement over repeated iterations, each iteration bringing it closer to the optimal explanation.

FUTURE RESEARCH DIRECTIONS

The vision of possible crowd-workplace which substitutes a traditional workplace is hardly discussed in the literature. Paid crowd work is a great chance to enhance “productivity, social mobility, and the global economy by engaging a geographically distributed workforce to complete complex tasks on demand and at scale.” (Kittur et al., 2013, p. 1301). It is important to stress some aspects related to the crowdsourcing process in an enterprise’s operation which should be explored further. Among them we can mention internal and external factors that exert influence on the way enterprises manage their resources. Crowdsourcing business model changes organizational structure, relationships between employer and employee, quality control etc. And this is why there will be a need to tackle with the following questions of holistic nature: Will crowdsourcing model affect employee’s motivation to work? Will traditional employees be replaced by web-based crowdworkers?

There are also questions of functional nature which are focused on crowdsourcing as business model to be implemented in an enterprise’s operation: Can crowdsourcing be formalized (career levels, quality, task assigning, crowdworkers’ training)? How to manage synchronous collaboration between a crowdsourcer and a crowdworker? How to manage online work tasks? How to attract efficient crowdworkers? How to manage shared resources?

CONCLUSION

Pierre Lévy (1997) claims that since “no one knows everything, everyone knows something,[and] all knowledge resides in humanity” (pp. 13-14). His statement reflects the main idea of crowdsourcing process. Crowdsourcing means that anyone with Internet access can accomplish a task provided

by a company worldwide. Thus, crowdsourcing gives companies a great chance to distribute tasks on demand to an international group of crowdworkers. It is a great opportunity to internalize consumers’ thoughts.

This article describes crowdsourcing as a business model that can be applied by enterprises intending to empower consumers (prosumers). Enterprises create innovations in order to gain competitive advantage in the traditional, as well as in cyber-marketplace. In this context, crowdsourcing is an original instrument for advancing products, developing online services, and fostering online communication with users. What is more, crowdsourcing links an open creation process with conventionally managed top-down process in which prosumers play a significant role.

They are integrated into the corporate value creation by engaging them (their competencies and knowledge) to develop products/services to meet consumers’ needs. The emergence of crowdsourcing-based business models is driven by technology, active users, and the move towards open innovation.

Among the advantages of crowdsourcing we can indicate: cost effectiveness (crowdsourcing basically reduces the costs of hiring and dismissal of employees), the opportunity to involve crowdworkers in a business activity on demand, recruitment flexibility. On the other hand, it is important to accentuate legal and ethical risks emerging from crowdsourcing practice (lack of regulations related to online labor, “digital sweatshops”). Paid crowdwork is perceived as ethically controversial, however, it is the crowdsourcing option which gains great popularity since it gives the opportunity to labor mobility.

The suggestions presented above lead to the conclusion that companies have much to learn from crowds. Identifying how to harness the power of crowds is the essential step for companies to differentiate themselves in the marketplace and sustain their competitive edge.

REFERENCES

- Acar, O. A., & van den Ende, J. (2011, January). *Motivation, Reward Size and Contribution in Idea Crowdsourcing*. Paper presented at the Dime-Druid Academy Winter Conference 2011 on Comwell Rebild Bakker, Aalborg, Denmark.
- Aitamurto, T., Leiponen, A., & Tee, R. (2011). *The Promise of Idea Crowdsourcing – Benefits, Contexts, Limitations*. White Paper June 2011. Retrieved November 13, 2013 from <http://www.crowdsourcing.org/document/the-promise-of-idea-crowdsourcing--benefits-contexts-limitations/5218>
- Brabham, D. C. (2008a). Crowdsourcing as a Model for Problem Solving. An Introduction and Cases. *Convergence (London)*, 14(1), 75–90. doi:10.1177/1354856507084420
- Brabham, D. C. (2008b, May). Moving the Crowd at iStockphoto: The Composition of the Crowd and Motivations for Participation in a Crowdsourcing Application. *First Monday*. Retrieved November 19, 2013 from <http://firstmonday.org/ojs/index.php/fm/article/view/2159/1969>
- Brabham, D. C. (2010). Moving the Crowd at Threadless: Motivations for Participation in a Crowdsourcing Application. *Information Communication and Society*, 13(8), 1122–1145. doi:10.1080/13691181003624090
- Brabham, D. C. (2013). *Crowdsourcing*. Cambridge, MA: MIT Press Essential Knowledge.
- Chanal, V., & Caron-Fasan, M.-L. (2008). How to Invent a New Business Model Based on Crowdsourcing: the Crowds spirit Case. *Euram*. Retrieved November 13, 2013 from <http://hal.archives-ouvertes.fr/docs/00/37/07/61/PDF/Euram2008-Chanal-Caron.pdf>
- Chesbrough, H. (2003). *Open Innovation: The New Imperative for Creating and Profiting from Technology*. Boston: Harvard Business School Press.
- Cox, G. (2005). *The Cox Review of Creativity in Business*. London: HM Treasury.
- Estellés-Arolas, E., & González-Ladrón-de-Guevara, F. (2012). Towards an Integrated Crowdsourcing Definition. *Journal of Information Science*, 38(2), 189–200. doi:10.1177/0165551512437638
- Felstiner, A. (2010). *Working the Crowd: Employment and Labor Law in the Crowdsourcing Industry*. From the Selected Works by Alek L. Felstiner. University of California, Berkeley. Retrieved November 13, 2013 from http://works.bepress.com/alek_felstiner/1/
- Grün, O., & Brunner, J.-C. (2002). *Der Kunde als Dienstleister: Von der Selbstbedienung zur Co-Produktion*. Wiesbaden: Gabler Verlag; doi:10.1007/978-3-322-89489-2
- Harris, Ch. G. (2011, October). *Dirty Deeds Done Dirt Cheap. A Darker Side of Crowdsourcing*. Paper presented at the IEEE International Conference on Privacy, Security, Risk, and Trust. Retrieved November 13, 2013 from http://myweb.uiowa.edu/cgharris/papers/harris_c08.pdf
- Hippel, E. (2006). *The Sources of Innovation*. New York: Oxford University Press.
- Howe, J. (2006, June). The Rise of Crowdsourcing. *Wired*. Retrieved November 19, 2013 from www.wired.com/wired/archive/14.06/crowds.html
- Howe, J. (2008). *Crowdsourcing. Why the Power of the Crowd Is Driving the Future of Business*. New York: Crown.
- Howe, J. (2010). *Crowdsourcing*. Retrieved February 19, 2014 from <http://www.crowdsourcing.com/cs/>
- IDEO. (2015). *IDEO*. Retrieved November 15, 2015 from <http://ideo.com>
- Ipeirotis, P. (2010). *Demographics of Mechanical Turk*. The Center for Digital Economy Research, CeDER Working Papers, CeDER-10-01. Retrieved November 19, 2013 from <http://hdl.handle.net/2451/29585>

- Jenkins, H. (Ed.). (2009). *Confronting the Challenges of Participatory Culture. Media Education for the 21st Century*. MacArthur The John D. and Catherine T. MacArthur Foundation. Cambridge, MA: The MIT Press.
- Kittur, A., Nickerson, J. V., Bernstein, M. S., Gerber, E. B., Shaw, A., Zimmerman, J.,.... (2013). The Future of Crowd Work. In *Proceedings of the 2013 conference on Computer supported cooperative work, CSCW'13* (pp. 1301-1318). New York: ACM. Retrieved November 10, 2013 from <http://hci.stanford.edu/publications/2013/CrowdWork/futureofcrowdwork-cscw2013.pdf>
- Kleemann, F., Voß, G. G., & Rieder, K. (2008). Un(der)paid Innovators: The Commercial Utilization of Consumer Work through Crowdsourcing. *Science, Technology & Innovation Studies, North America*, 4(1). Retrieved November 20, 2013 from <http://www.sti-studies.de/ojs/index.php/sti/article/view/81/62>
- Kohler, T. (2015). Crowdsourcing-Based Business Models: How to create and capture value. *California Management Review*, 57(4), 63–83. doi:10.1525/cmr.2015.57.4.63
- Lakhani, K. R., Jeppesen, L. B., Lohse, P. A., & Panetta, J. A. (2007). *The Value of Openness in Scientific Problem Solving*. Harvard Business School Working Paper, No. 07-050. Retrieved November 20, 2013 from <http://www.hbs.edu/research/pdf/07-050.pdf>
- Lévy, P. (1997). *Collective Intelligence: Mankind's Emerging World in Cyberspace*. New York: Plenum.
- OpenIDEO. (2015). *OpenIDEO*. Retrieved November 15, 2015 from <http://openideo.com>
- Pénin, J., & Burger-Helmchen, T. (2011). Crowdsourcing of Inventive Activities: Definition and Limits. *International Journal of Innovation and Sustainable Development*, 5(2/3), 246–263. doi:10.1504/IJISD.2011.043068
- Prahalad, C. K., & Ramaswamy, V. (2004). Co-Creation Experiences: *The Next Practice Invalue Creation*. *Journal of Interactive Marketing*, 18(3), 5–14. doi:10.1002/dir.20015
- Reichwald, R., & Piller, F. (2006). *Interaktive Wertschöpfung: Open Innovation, Individualisierung und neue Formen der Arbeitsteilung*. Wiesbaden: Gabler Verlag.
- Rogstadius, J., Kostakos, V., Kittur, A., Smus, B., Laredo, J., & Vuković, M. (2011, July). An Assessment of Intrinsic and Extrinsic Motivation on Task Performance in Crowdsourcing Markets. In *Proceedings of the Fifth International Conference of the Association for the Advancement of Artificial Intelligence on Weblogs and Social Media*. Retrieved November, 10, 2013 from <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2778>
- Ross, J., Zaldivar, A., Irani, L., & Tomlinson, B. (2009). *Who Are the Turkers? Worker Demographics in Amazon Mechanical Turk*. Social Code Report 2009-01. Retrieved November 10, 2013 from <http://www.ics.uci.edu/~jwross/pubs/SocialCode-2009-01.pdf>
- Schenk, E., & Guittard, C. (2009). *Crowdsourcing: What Can Be Outsourced to the Crowd, and Why?* Working Papers Series, HAL-CCSD. Retrieved November 10, 2013 from <http://raptor1.bizlab.mtsu.edu/s-drive/DMORRELL/Mgmt%204990/Crowdsourcing/Schenk%20and%20Guittard.pdf>
- Scholtz, T. (Ed.). (2013). *Digital Labor. The Internet as Playground and Factory*. New York: Routledge.
- Surowiecki, J. (2004). *The Wisdom of Crowds: Why the Many are Smarter than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations*. New York: Doubleday.
- Tapscott, D., & Williams, A. D. (2006). *Wikinomics. How Mass Collaboration Changes Everything*. New York: Tantor Media.

Tapscott, W., & Williams, A. (2013). *Radical Openness*. TED.

Taylor, C. (2005). School of Bright Ideas. *Time Magazine*. Retrieved November 19, 2013 <http://www.time.com/time/magazine/article/0,9171,1034705,00.html>

Teece, D. J. (2010). Business Models, Business Strategy and Innovation. *Long Range Planning*, 43(2), 172–194. doi:10.1016/j.lrp.2009.07.003

Toffler, A. (1980). *The Third Wave*. New York: Bantam House.

Vuković, M. (2009). Crowdsourcing for Enterprises. *Congress on Services I* (pp. 686–692). Retrieved November 12, 2013 from <http://sistemasmahumano-computacionais.wdfiles.com/local-files/capitulo%3Aredes-sociais/vukovic2009.pdf>

Wolfson, S. M., & Lease, M. Look Before you Leap: Legal Pitfalls of Crowdsourcing. *Proceedings of the American Society for Information Science and Technology*, 48(1), 1–10. doi:10.1002/meet.2011.14504801135

ADDITIONAL READING

Afuah, A., & Tucci, Ch. (2013). Value Capture and Crowdsourcing. *Academy of Management Review*, 38(3), 457–460. doi:10.5465/amr.2012.0423

Alonso, O., Rose, D. E., & Stewart, B. (2008). Crowdsourcing for Relevance Evaluation. *ACM SIGIR Forum*. 42(2), 9–15. doi:10.1145/1480506.1480508

Bayus, B. L. (2013). Crowdsourcing New Product Ideas Over Time: An Analysis of the Dell IdeaStorm Community. *Management Science*, 59(1), 226–244. doi:10.1287/mnsc.1120.1599

Bonaccorsi, A., & Rossi, C. (2007). Altruistic Individuals, Selfish Firms? The Structure of Motivation in Open Source Software. *First Monday*, 9(1), 1–5. doi:10.5210/fm.v9i1.1113

Cefkin, M., Anya, O., Dill, S., Moore, R., Stucky, S., & Omokaro, O. (2014). Back to the Future of Organizational Work: Crowdsourcing and Digital Work Marketplaces. In *Proceedings of the Companion Publication of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing* (pp. 313–316). ACM: New York.

Chesbrough, H. W. (2007). *Open Business Models. How to Thrive in the New Innovation Landscape*. London: Mac-Graw Hill.

Dabholkar, P. A. (1996). Consumer Evaluations of New Technology-based Self-Service Options: An Investigation of Alternative Models of Service Quality. *International Journal of Research in Marketing*, 13(1), 29–51. doi:10.1016/0167-8116(95)00027-5

Evans-Cowley, J., & Hollander, J. (2010). The New Generation of Public Participation: Internet-based Participation Tools. *Planning Practice and Research*, 25(3), 397–408. doi:10.1080/02697459.2010.503432

Geiger, D., Seedorf, S., Schulze, T., Nickerson, R. C., & Schader, M. (2011). Managing the Crowd. Towards a Taxonomy of Crowdsourcing Processes. *Proceedings of the Seventeenth Americas Conference on Information Systems*, paper 430, Detroit, Michigan.

Hars, A., & Ou, S. (2002). Working for Free? Motivations for Participating in Open Source Projects. *International Journal of Electronic Commerce*, 6(3), 25–39.

Hertel, G., Niedner, S., & Hermann, S. (2003). Motivation of Software Developers in the Open Source Projects: An Internet-Based Survey of Contributors to the Linux Kernel. *Research Policy*, 32(7), 1159–1177. doi:10.1016/S0048-7333(03)00047-7

Johnson, S. (2010). *Where Good Ideas Come from: The Natural History of Innovation*. New York: Riverhead Books.

Kelemen, M., & Smith, W. (2001). Community and its Virtual Promises: A Critique of Cyberlibertarian Rhetoric. *Information Communication and Society*, 4(3), 370–387. doi:10.1080/713768547

Mau, B., & Leonard, J. (2004). *Massive Change*. New York: Phaidon.

Muniz, A. M. Jr, & OGuinn, T. C. (2001). Brand Community. *The Journal of Consumer Research*, 27(4), 412–432. doi:10.1086/319618

Piller, F. T., & Walcher, D. (2006). Toolkits for Idea Competitions: A Novel Method to Integrate Users in New Product Development. *R & D Management*, 36(3), 307–318. doi:10.1111/j.1467-9310.2006.00432.x

Postigo, H. (2003). From Pong to Planet Quake: Post-Industrial Transitions from Leisure to Work. *Information Communication and Society*, 6(4), 593–607. doi:10.1080/1369118032000163277

Severin, W. J., & Tankard, J. W. Jr. (2001). *Uses of the Mass Media in Communication Theories: Origins, Methods, and Uses in the Mass Media*. New York: Longman.

Sloane, P. (2011). *A Guide to Open Innovation and Crowdsourcing: Advice from Leading Experts*. London, Philadelphia: Kogan Page Publishers.

Sui, D., Elwood, S., & Goodchild, M. (Eds.). (2013). *Crowdsourcing Geographic Knowledge. Volunteered Geographic Information (VGI) in Theory and Practice*. Heidelberg, New York, London: Springer; doi:10.1007/978-94-007-4587-2

Terranova, T. (2004). *Network Culture: Politics for the Information Age*. London: Pluto Press.

KEY TERMS AND DEFINITIONS

Broadcast Search Approach: A type of crowdsourcing which focuses on scientific solutions by a group of selected problem-solvers (e.g. InnoCentive – a platform where scholars work out challenges devised by its customers).

Collective Intelligence: Shared or group intelligence that emerges from the collaboration, collective efforts, and competition of many individuals and appears in consensus decision making.

Crowd Wisdom: The process of taking into account the collective opinion of a group of individuals rather than a single expert.

Crowdsourcing: A business model in which one party (an individual, a company, a NGO, a public institution) solicits a group of individuals for solving a task via the Internet.

Distributed Human Intelligence Tasking: A type of crowdsourcing in which the core issue is data processing by the humans only (e.g. Amazon Mechanical Turk, an online platform coordinating jobs which can be provided by Internet users).

Knowledge Discovery: A type of crowdsourcing which is based on creating online content by users, however, this process of knowledge discovery is managed by an organization (e.g. SeeClickFix, the US website on which people can report non-emergency problems in their local community via the Internet or a mobile phone application).

Peer-Vetted Creative Production Approach: A type of crowdsourcing in which the creation process is open to web users. This web-based community work out solutions that live up to the expectations of an organization. Finally the best option is selected (e.g. Threadless.com, online company selling t-shirts based on design provided by artists. The best designs are chosen in a voting process).

Prosumer: Proactive consumer; takes an active role in creating a product or service.

Architecture as a Tool to Solve Business Planning Problems

B

James McKee

Independent Researcher, Australia

INTRODUCTION

Architecture can give systematic structure to problem investigation, solution development and formality of presentation to the solution plan. An architectural reference model will allow the planners guidance in their investigations and in their development of the plan. An architectural blueprint of the solution will allow implementers to better understand the intentions of the planners.

Architecture is essential for any large building or urban planning proposal (not the aesthetics of the design but the structural design – how things fit together), why should this architectural discipline not help in designing solutions for other types of large scale problems? The architecture discipline would not only aid the development of the solution but allow the project plan to become a better communication document. There is therefore a need for a generalized discussion on the basic architectural principles that can guide the development of architectures for different problem areas.

A key argument for architecture versus design is that design solves the functional requirements of the specification whereas architecture is in *addition*, about the long term sustainability of the solution and in the process takes into account more environmental factors (market place, competitors, etc.). A definition of IS architecture given by Bass, Clements and Kazman (1998) allows a clear distinction between architecture and design. They write that architecture begins the task of envisioning the central function of a system looking for those influences that persist beyond the lifetime of the system, whereas design is just about ensuring the system will function as expected.

In other words, architecture is concerned with a need to explore beyond the immediate description of functionality and is required to also stretch the boundaries, looking at any environmental factors that surround the objective. The architecture gives much more knowledge about the structure and integration of the entire system being described. The sustainability is due to lower cost and risk associated with development changes because of that knowledge.

Architecture provides a systematic examination of the entire enterprise from a summary overview at the top level of abstraction with navigating links to all the component parts and their definition at descending levels of abstraction and with descriptions of any interactions between component parts. This blueprint, if correctly done, can be a powerful management tool for determining where efficiencies can be made and choosing effective strategic directions.

However before beginning the analysis of the organisation, a set of architectural principles need to be adopted to guide the analysis and to develop a reference model of the organisation's business architecture for use in presenting the analysis results in the architectural format for any specific enterprise and this will be discussed in the section headed architecture as a solutions tool.

BACKGROUND: CONCEPT OF ARCHITECTURE IN COMMERCE

Architecture is often considered as just the aesthetics of the building design but there are entire undergraduate and post graduate courses

DOI: 10.4018/978-1-5225-2255-3.ch050

on building architecture that go beyond just the classic proportions for a building but detail the formulas and tables on the structural loads for foundations and support walls etc. A great many technical issues are pre-solved for use in the design and development of a new building.

The idea of architecture has now been extended to the area of information technology although this is fairly recent say within the last 30 years; two of the most influential authors in the early years were Brancheau, and Martin. Brancheau and Wetherbie wrote “an information architecture is a high level map of the information requirements of an organisation” (1986), however I believe the use of the term ‘map’ diminishes the amount of structure inherent in an architecture. Martin in 1990 drove the emphasis for *information* with his books on Information Engineering and his depiction of an architecture to describe the enterprise. There are various descriptions of enterprise architecture used by the IT profession, a very detailed description is given on the web site for TOGAF (2013) and Koontz (2000) gives a very clear layout of the different levels in the enterprise architecture. All these descriptions of enterprise architecture are essentially intended to improve the design of IT systems to align the information systems better with the business operation and strategic development.

Harmon (2008, 2004) has also stated this pre-occupation of definitions for “enterprise architecture” to be focused on the application of IT resources. Harmon (2004) also shows a detailed description of the hierarchy of levels in enterprise architecture, but in his view of the enterprise the focus is on the business processes and that business process management is a key component of the “enterprise architecture”.

Following the emphasis on business processes there was a focus on creating standard processes by the management consulting and software development world and the concept of service oriented architectures (SOA) was developed. I believe the emphasis should be with the requirement to determine the structure of the enterprise - including

what product suits the available market at what price point and what style of distribution to use etc., - before reviewing the processes required.

An information architecture has been proposed by McKee for presentation of the strategic plan, to pick up after the creative determination of the strategic objectives and to focus on the implementation plan. The architecture included determining the planning information components, their interaction and the overall structure of the plan and a reference model was developed for the presentation of this information, which is explained in more detail in a later section (McKee, 2013).

There are now many discussions about the need for business architecture; a white paper entitled “The Evolution of Business Architecture” was written by a business consultant (Cochlin, 2012). Cochlin defines business architecture as: “Transforming corporate strategy into business designs, structures and aligned delivery capabilities that enable corporations to increase market share, return on investment and agility in the face of change, whilst at the same time reducing risk”. However the use of architectural design in the business world needs a lot more definition of frameworks and reference models than is currently available. A comment from some management consultants is to say that it is a specialist function to interpret the business processes and requirements into an architecture and it may not be appropriate to have generic models available for the business person to use themselves. Surely the senior business managers are in the best position to determine and then describe the functions and structure of the organisation, however they do need some examples and reference models to guide them.

One definition of enterprise architecture is that it is “a conceptual blueprint that defines the structure and operation of an organisation. The intent of an enterprise architecture is to determine how an organisation can most effectively achieve its current and future objectives. Purported advantages of having an enterprise architecture include “improved decision making, improved adaptability to changing demands or market conditions, elimi-

nating redundant processes, optimization of the use of organisational assets and minimization of employee turnover” (Rouse, 2007).

There is an increasing number of books and articles now available with business architecture as the primary focus, many are being written by consultants rather than academics. A definition of business architecture is given by Ulrich & Kuehn “business architecture is a blueprint of the enterprise that provides a common understanding of the organization and is used to align strategic objectives and tactical demands [quoting the Business Architecture Guild]. It is an abstract set of standardized perspectives that represent a given business ecosystem. The foundational perspectives include the following core “domain” categories:

- What a business does (capability)
- The vocabulary it uses (information)
- How a business is organized (organization)
- How a business delivers value to key stakeholders (value stream)” (Ulrich & Kuehn, 2015).

However the discussion of what constitutes a business architecture is very broad and diffuse from many business consultants and minimal direction from academia creating a lack of common definitions and frameworks for business managers to refer to and use.

ARCHITECTURE USE IN COMMERCE IS LIMITED

Zachman was very concerned that “[t]he credibility of IS is in a steep decline.” And that “[t]he issues of quality, timeliness and change are the conditions that are forcing us to face up to the issues of Enterprise Architecture” (Zachman, 1996, p.1). If these issues were thought to be very relevant to the development of effective IT systems, how much more relevant can they be to developing a modern effective business organisation and in fact one could extrapolate to saying any large complex operation.

James as the Asia-Pacific Architecture Research Director for Gartner (a company of IT industry analysts), and in an article for the newspaper *The Australian* writes about the need for a business to respond ever faster to the changing environment, and this diversity must be managed effectively to get good business results. That Enterprise architecture is important is largely because there are more applications today than ever before. And those applications are managing more data on a diverse range of platforms, written in many languages and running in a variety of environments and this diversity can be chaotic to manage (James, 2001). A similar argument can be made for the modern global business competing worldwide in rapidly changing environments (political and marketplace).

James is clear that for business success there needs to understanding of enterprise architecture: “Within the control of the enterprise are the technologies that are employed (IT architecture), the business processes, functions and information (information architecture) that embody the organisation and the integration infrastructure (the city plan)” (James, 2001).

Unfortunately the emphasis in these quotes and other similar writings about enterprise architecture is on developing a better IT system, rather than understanding the key operational structures and dependent relationships of the business organisation. Even though the definitions for enterprise architecture do include defining the business and strategic planning functions, they seem to be only there as an afterthought to support the IT development. There needs to be an approach that starts with the business focus.

Issues

There is quite a large body of material that now help explain and define information technology architectures which gives the information systems designer useful tools to aid the analysis and the design of new systems. The architectures gives the designer, the analyst and the programmer a common view of the problem and its solution,

this reduces uncertainty and aids communication and reduces problems with the reliability of the product.

However there is still not a great deal of undergraduate offerings in the architecture fields (IT or business), and there should be discussions at least on how effective or how well they describe and define their problem area. Another issue is that because the focus is of the business world and the profit to be made, several researchers in the area of architectures for information technology are now management consultants to the industry and no longer write academic research papers.

There is very detailed specifications from TOGAF and others for 'Enterprise Architecture' which is purported to cover everything from the Strategic Plan, Business Architecture to the IT System. But the reality is the focus of 'Enterprise Architecture' is very much on what is required to build the IT system for the organisation. This paper reviews the background to the use of architecture in commerce and the discussion on 'Enterprise Architecture' and the limitations associated and will then propose the way architecture can be used for problem evaluation and solution development with the focus on business planning.

Where are the generic reference models and frameworks for use in business architecture? A first response is to indicate that the Zachman framework and the TOGAF information on architectures provide this requirement. Grigoriu argues that the Zachman's matrix describes the various cells of the framework but does not explain how to model them and so is of very limited in help with building the enterprise architecture. He also criticizes the TOGAF website of architectural information as being primarily concerned with IT architecture not with the business architecture and has neither information navigation nor an integrated view of the enterprise (Grigoriu, 2013), Grigoriu is an executive consultant in Enterprise Strategy and Architecture.

Controversies

Many heads of organisations would argue that their business is too complex or too diverse or too changeable to spend the necessary time developing an architecture for their operation and they would question where the return on investment might come from.

Several authors including (Grigoriou, 2011; Sereff, 2012; and Ross Weil and Robertson, 2006) are now writing about business architecture as a fundamental part of enterprise architecture. Enterprise architecture (EA) was first developed to enable the IT world to better understand the business enterprise and therefore create IT systems better aligned to the business needs. Unfortunately there is still a leaning toward IT with the enterprise architecture definitions; Ross, Weill and Robertson make a determined effort to avoid this trap and rail against "traditional IT architecture as too remote from reality of business" Grigoriu also reported his view that many speakers thought the enterprise architects should be more business oriented and that many did not have a clear definition of what Enterprise Architecture is. This is perhaps a reflection that the subject of architecture is much underrepresented in University courses, how many MBA's have a component on the topic of architecture?

One area of architecture that is getting some prominent exposure is that of business intelligence (BI) architecture, one major player is IBM and their Cognos subsidiary. Cognos has of course been in the area of developing software for management and executive information systems (MIS and EIS) for a long time along with many other software companies. These packages allow non-programmers to ask (English like) questions of the company databases (IBM-Cognos, 2011).

A concern is that looking into the data base for answers to your management questions about whether strategic directions are being achieved is

the wrong end of the planning and development cycle. The performance questions need to be determined at the time of creating the strategic objectives with the actions necessary to achieve them; maybe the performance measurements can then be built into the information system and the results fed into the data base. Therefore the business management questions and strategy formulation should be at the beginning of the planning cycle together with the performance measures, rather than examining what has been collected via standard operational transactions.

Problems

It can be justifiably argued that every organisation is different; the argument might then go on to say there are few common denominators of operation across organisations that could have an architecture defined. A significant problem with research on the subject of defining an architecture, is the difficulty associated with proving the effectiveness or usefulness of the architecture because of the long term and the confidential nature of the subject, however the normative model does provide a research option (Routio, 2005). There is a multitude of definitions for enterprise architectures but there is a lack of models & frameworks that provide a cohesive structured format to define a business.

The problems of definitions and frameworks can be seen very specifically in the area of the business model which is deemed by many to be of critical importance “great business models can reshape industries and drive spectacular growth” (Johnson, Christensen & Kagerman, 2008) but Linder & Cantrell (2000) say that “everyone talks about business models but 99% have no clear framework for describing their model” (2000, p.2). The lack of motivation in this respect may well be driven by the concern over the lack of a common definition for the business model, Sorensen promotes the combination of business model and business plan states that “there are no established definitions of business models in the literature today” (Sorensen, 2012, p.155).

Reviewing the definitions for the business model there is ambiguity over the use of the term ‘value’ with conflict between value for the customer and value to the organisation. There is also discussion whether strategy fits in the model or not, and if the whole of the organisation operation is covered by the model, which would make the model too complex to be updated on a regular basis (McKee, 2015).

As indicated above the business model and the strategic plan are often described as part of the enterprise architecture along with all the information of business processes and other business resources making the total amount of information very unwieldy to use as a communication tool as a regular management function.

There is often a lack of effective communication in organisations, particularly across the management silos and is of particular concern when attempting a coordinated approach to strategic objectives that cross management boundaries, (large organisations and multinational organisations are especially vulnerable).

Planning sessions have a potential danger when participants lack objective consideration of all the factual information about the problem area under discussion. They can be prejudiced by subjective misdirection on critical information by inappropriate cues. Rea and Kerzner (1997) refer to errors in judgment where people have a tendency to remember only those things that confirm existing belief rather than examine and analyse more accurate and objective material. Similarly Alter draws attention to eight common decision making flaws and advocates the use of more objective data that can be drawn from the information systems. The same can be said of any management decisions being made without the support of coherent structured information about the organisation which should be contained and presented via a well-designed architecture.

The enterprise architecture is supposed to contain both the business architecture and the IT architecture, however in enterprise architecture models they both appear as independent compo-

nents of the enterprise architecture. In these models there are no relationship mechanisms or dependencies shown between the two architectures and therefore no impact statements to say the amount of effect of the IT component on each of the of the business components. It is well understood today that an effective software system package that is well aligned to the business processes of a particular department and carefully implemented can drive up the profitability of that department, and this relationship should be measured with the degree of impact well understood.

Authors Harmon and Grigoriou complain that many descriptions of enterprise architectures are not able to be used to model the enterprise effectively. However unless the architecture includes descriptions of components with rationales, interactions and relations defined, then the architecture will not allow effective modelling and will not be sustainable in changing conditions.

ARCHITECTURE AS A SOLUTIONS TOOL

Minztberg suggests an important function of a plan is to be the communications media; “plans clearly serve in two capacities, or roles. They are media for communication and devices for control”. He writes that the concept of coordination is “to ensure that everyone in the organisation pulls in the same direction, which may be facilitated by specifying that direction as precisely as possible”

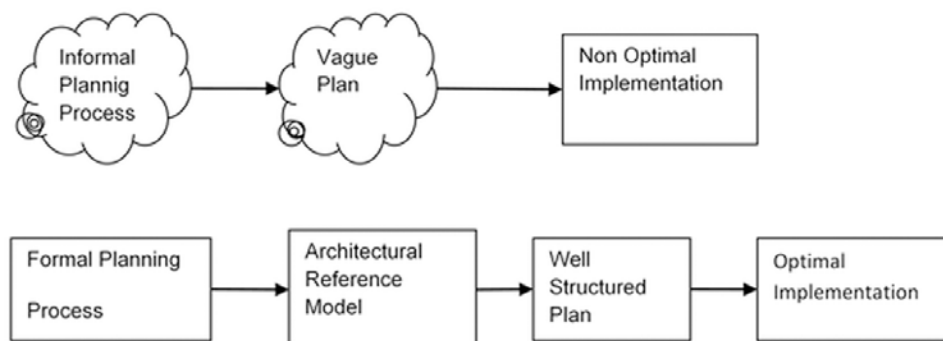
(1994, p. 351-352). It is necessary to have a solution document that has a systematically structured format to give implementers the clear instructions for what is required. If you start with an ill-defined process it is likely that the resulting solution is going to be sub optimal.

An architectural blueprint of the organisation would go a long way to ensuring that all the staff are on the same page and by including the strategic plan within the enterprise architecture then it would ensure development of the organisation would be as effective as possible. The architectural approach in many ways enhances the standard useful investigative questions of What, How, When and Why.

The idea is to create a LOGICAL view of the problem elements and possible solution components without any preconceived notions of what the final solution is or of how it is to be implemented. When the investigation is complete, then the structure of information can be reconsidered for relevance and importance with the degree of impact assessed for all cause & effect relations. From the key information elements various alternative options can be drawn up for final consideration. The solution space should now contain a range of alternative options not constrained by the PHYSICAL determination of how they are to be implemented.

Having a set of architectural principles to follow will act as a guide in the search through the problem space for the critical information components useful in the construction of an

Figure 1. Unstructured vs. structured planning



architectural reference model. This will give a systematic approach to the problem investigation and of the solution development. It will also encourage a broader review of the problem area and of the solution space.

When a final choice is made for the components that will make up the solution then the plan for how the solution is to be implemented should be documented using the guidance of the architectural reference model based on the above structure and the architectural principles and thereby following the standard that all staff will understand.

Defining Architectural Principles

Principles are the fundamental laws or doctrines for something; they are not rules for which the action is more specific but are guidelines to the requisite action. Guimarães (2012) in developing ideas to guide the development of an enterprise architecture for use in the financial market puts forward the following definition “principles are high-level definitions of fundamental values that guide the IT decision-making process, serving as a base for the IT architecture, development policies, and standards”. This definition unnecessarily limits the definition to the IT aspect of the architecture however, the definition is well suited for any problem solving domain, leave out the term IT and the definition suits the more general case. Although his use of the term ‘principle’ is more to do with the way to run the business than determining the structure of the business.

Aier et al. (2011) follows the development of the concept of enterprise architectural (EA) principles which are used as a means to develop a transition model for the enterprise: to establish a ‘TO BE State’ or ‘strategic plan’ for the enterprise. The EA principles are developed from requirements for strategic developments for the organisation and is framed by rationale, key actions and measurements, this takes a slightly different focus from McKee (2013). In McKee (2013) the architectural principles are aimed at the development of the Reference Model first, for a problem

area or solution state, and uses the reference model to guide the description of the ‘AS IS’ state prior to the development of the ‘TO BE’ state of the enterprise. The other difference in McKee (2013) is that the term ‘enterprise architecture principle’ is replaced in the strategic plan with ‘strategic objective’, framed by rationale, critical success factors, key actions and measures; as can be seen there are similarities in the supporting description of each approach.

Aier (2014) follows on from the discussion in Aier (2010), to research the role organisational culture has on the efficacy of the EA principles and has gathered a table of definitions from several authors for the enterprise architecture principles, two of the definitions are particularly interesting. One from Richardson (1990) “Principles are an organisation’s basic philosophies that guide the development of the architecture ... Principles provide guidelines and rationales for the constant examination and re-evaluation of technology plans” (p.389) and another from Lindstrom (2006) “Architectural principles define the underlying general rules and guidelines for the use and deployment of all IT resources and assets across the enterprise..” (p.2).

It is strange that these definitions of principles are on one hand limited to the technology plans and on the other, limited to IT resources. Each definition is much more powerful if the word technology is left out of the former and the word IT is left out of the latter. However if we replace the words ‘technology and IT’ with the word ‘business’ we get a very effective definition for use with the concept of ‘business architecture’.

In the search for a set of principles to use in developing an information architecture, the two definitions for ‘Enterprise Architecture’ developed by The Open Group were examined as follows:

- That architecture is “a formal description of a system or a detailed plan of a system at component level to guide its implementation”.

- That architecture is “the structure of components, their interrelationships, and the principles and guidelines governing their design and evolution over time” (Introduction, p.3).

The following architectural principles were derived from these definitions by McKee:

1. A formal detailed plan to guide implementation.
2. The plan should contain the components, their description and structure.
3. The plan should define the interrelationships between components.
4. The plan includes guidelines covering evolution over time (The rationale for why major components are there – which allows the rationale to be checked against changing circumstances).

Two other principles were added from other literature sources:

5. The plan should comprise several levels of abstraction from conceptual at the top linked to increasing detail at the next level.
6. The plan should include investigating environmental factors that surround the objective. (McKee, 2013).

These principles were then followed in the review of the business management and planning literature in order to develop an architectural reference model for use in creating the information architecture for presenting strategic planning information.

Developing an Architectural Model

A Reference Model (RM) is a conceptual framework to identify the major elements that comprise a concept and to show at a general level how these

elements are related with each other. As mentioned in the previous section the identified architectural principles are used drive the development of the reference model for the enterprise architecture. “A model is a representation of an important aspect of the real world. Sometimes the term abstraction is used because we abstract (separate out) an aspect of particular importance to us” (Satzinger et al., 2004, p.45).

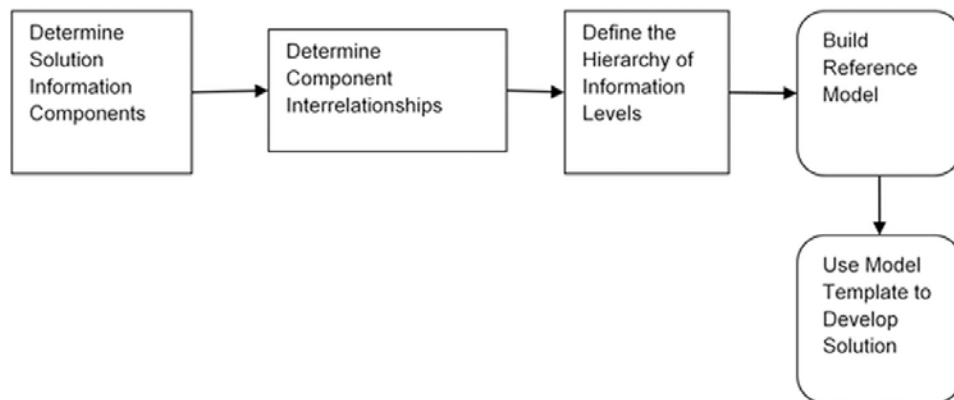
Bearing in mind the above set of principles, the problem domain is surveyed for the elements that are most critical and the nature of important dependencies noted and the RM built accordingly. Once the RM has been discussed and validated then the RM can be used to construct the actual model of the business organisation from the business analyses (the Porter’s 5 forces and value system, the SWOT etc.) and then the cause and effect relations can be confirmed and the degree of association calculated/estimated.

The purpose behind modelling is to help design something, in this case the business architecture, to ensure that the real business operation and its market setting is correctly identified and represented. The process of creating the model helps the planner clarify and refine the design, the model acting as a communication tool and assisting the dialogue (Satzinger et al., 2004).

As an example of how the architectural principles can be used to aid the problem solving process the principles were used to create a reference information architecture as a template for managing the development of the strategic implementation plan. The concept flow of this idea can be seen in the Figure 2.

The requirements for developing a solution for the format of a strategic plan were determined from the business management and planning literature, firstly what the solution components could be, then determine what interrelationships between components should be defined. The following material for the reference model is drawn from McKee, 2013.

Figure 2. Concept flow to develop solution



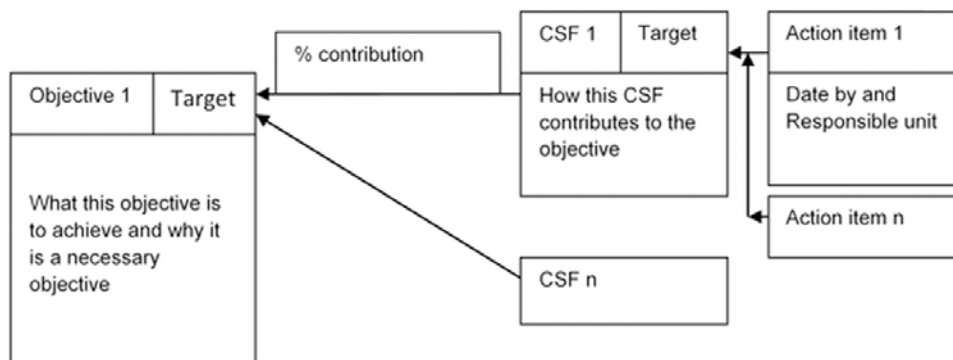
1. Components suggested by the literature were:
 - a. Objectives,
 - b. Critical success factors (CSF) of the objective,
 - c. Actions needed to achieve objective,
 - d. Key performance indicator/s (KPI's) for achieving the CSF or action.
2. Relations were determined to be:
 - a. Direct linkage between related components,
 - b. Rationale for major components,
 - c. Degree of contribution to higher level component.
3. The next step was to determine what the reasonable levels of abstraction for the solution structure should be.

Level one was determined to be the focus areas of the organisation requiring attention and the objectives and CSF's needed to achieve success, level two shown in Figure 3 was the CSF's and action items needed for each objective and level3 was the detail of key performance indicators required for each CSF.

Figure 3 clearly shows the relationships between components and includes the rationales - why the objective is there and how the CSF contributes to the objective. The action item has the detail of due date for completion and who is responsible.

The principles provided a guide for the development of a reference architecture for the strategic planning problem and could realistically provide

Figure 3. Second level abstraction showing objective to CSF relationship



guidance for the development of a reference business architecture when examining the business management literature and the various papers on the business analysis tools such as Porters Five forces and Porters Value chain and value system etc. In a different problem space, market research or other analysis may provide additional information components that would be appropriate for the solution space.

Once the reference architecture is built it gives the planning team a communication template to fill in as required as the solution develops. The levels of abstraction and the direct linkages between information elements allows the problem to be worked on by different sections of the staff as appropriate and also allows the solution to be developed iteratively with prior areas revisited as more detail develops. This same method of working will allow the solution to be revisited on a regular basis during implementation and changes made as required due to changing circumstances.

A useful starting point for the development of a business focused architecture for an organisation might be to dissect the key aspects of an organisation's core logic of business processes (perhaps via Porter's value chain / value system) together with an analysis of the functioning of the business model (see above section headed 'Problems') and how that relates to the operation of the business. The overall business architecture might perhaps also contain a finance architecture, logistic architecture, supply chain architecture; whichever area of activity are of particular concern to that organisation

The complexity of all the information concerning the organisation leads to the idea to separate the business model, the organizational plan and the strategic plan into different documents; each with its own approach. This also separates the more dynamic information that needs to be reviewed regularly, from that information that is more static over time. With the organisational analysis structured into the organisational description (business architecture); the future planning can be iterative between developing the business

model and testing the strategic choices against the business model then selecting the optimal items for the strategic plan.

Each document should have a clear purpose and function as follows:

- The Business Architecture lays out the information of the AS-IS state from the organisational analyses.
- The Strategic choice document lays out the various options for improvement.
- The Business Model sets out how the organisation is to do business; including selecting the target customer segment and deciding on the value to that segment. By parsing the strategic choices against the business model, decisions can be made and choices selected for implementation - creating the strategic plan.
- The Strategic Plan is the summation of what developments the company is to pursue with the actions determined and the budgets and key measurements set in order to create the TO-BE state of the organisation.

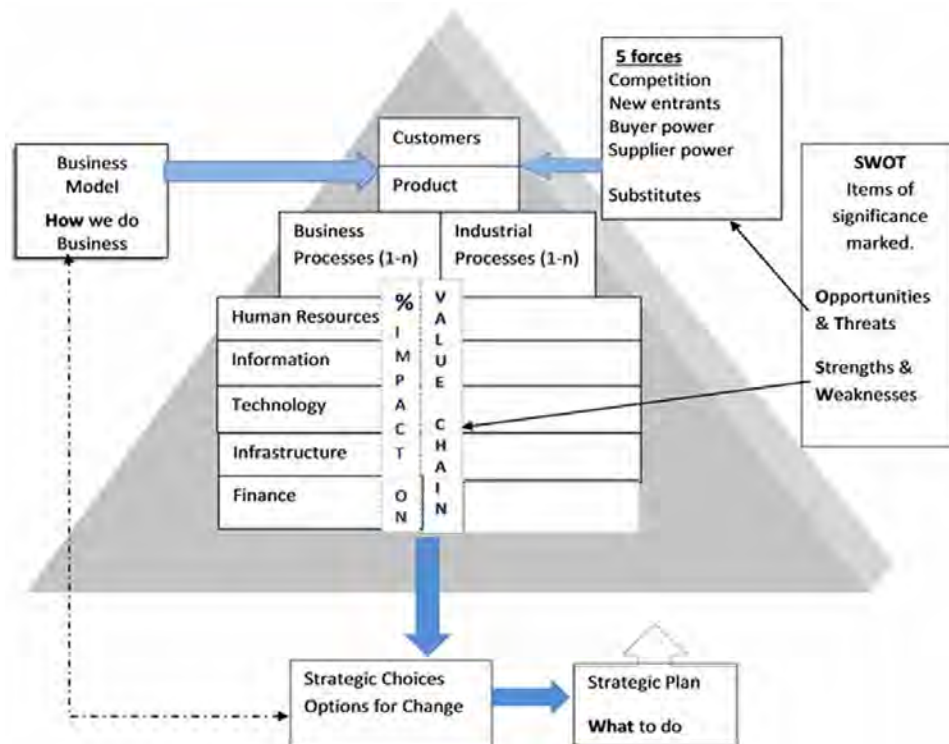
It is important that with the problem divided into different documents (see Figure 4), that the interdependencies are clearly identified and the linkages shown between the related elements, as seen in Figure 3 above and in particular between the different documents.

FUTURE RESEARCH DIRECTIONS

It is essential for there to be reference architectures developed as examples and templates for more definable aspects of the business enterprise – the value chain would be a good example where the example reference model would demonstrate how to show the parts that are driven by the information system and those that are manual and indicating the relative value add for each aspect. These can then be used as discussion vehicles and would al-

Figure 4. Level 1 basic business architecture reference model

Source: Adapted from Alter, 2002; McKee, 2015



low them to be further developed and refined for use by businesses that wish to create a sustainable progressive environment for their organisation.

The set of components, relations and levels used above for an information architecture for the strategic plan could be used and developed in future research. Other researchers could add to this set or refine them when checking their suitability across different organisations or industries. Similarly the range of components and the sorts of interactions that could be expected could be developed for the business architecture.

The identified set of architectural principles could be tested and refined across a range of scenarios for developing architectures.

A white paper entitled “Platform Engineering Maturity Model” discusses the evaluation of the maturity level of an organisation’s computing software platform, the importance I feel is the idea that as the organisation uses more products that have *certification* then the more mature it is

(Momentum, 2011). The computing hardware and now the application software products and blueprints, recipes & services can be deemed to be *certified* when meeting standard operating procedure requirements. It is necessary to research and develop *certification* for reference architectures, frameworks and models for different organisational structures i.e. mining, manufacturing, retail etc.

As part of these reference architectures it would be necessary to refine the principles and define frameworks and models and describe their attributes and possible artifacts that can result from their use.

CONCLUSION

This paper has shown how the idea of architectural principles can be used to guide the development of an architectural reference model which can be

used to develop and present the organizational analysis in the model of the organisation. McKee (2013) demonstrates a set of reference models for information architecture in the relatively easy area (in terms of architecture) of strategic planning. Figure 4 shows a possible starting point for a business architecture reference model. The next step must be to refine a set of reference models for business architecture, and perhaps separate models for the supply chain, marketing finances, logistics etc., it would depend on the focus of the organisation. The various different industry segments, mining, manufacturing, retail sales etc., each would require their own approach to what constituted a useful reference model.

It would be a good step forward if there was an academic subject on business architecture in the MBA that would give the subject a more standardized view for a global audience rather than the multiplicity of short courses and competing different views being run at present.

Architecture by providing a blueprint of the whole enterprise with descending levels of abstraction and navigating links to the component parts and their interactions can be a powerful communication tool for the organisation. The possible advantages of architecture for improved decision making, improved adaptability to change, eliminating waste and optimization of resources make it a goal worth having. A larger goal is going beyond business architecture to the concept of developing an architecture for each different problem space, creating a template for solution development.

Finally a concluding statement from Reynolds that gives a good argument for spending the effort in creating an architecture for your organisation.

Where there is a business architecture, it becomes easy to build a portfolio of projects or initiatives that will help the business move forward in a strategic way, with each initiative bringing the business closer to the envisioned end state of the roadmap (Reynolds, 2009, p8).

REFERENCES

Aier, S. (2014). The Role of Organizational Culture for Grounding, Management, Guidance and Effectiveness of Enterprise Architecture Principles. *Information Systems And E-Business Management*, 12(1), 43-70.

Aier, S., Fischer, C., & Winter, R. (2011) Construction and Evaluation of a Meta-Model for Enterprise Architecture Design Principles. *The 10th International Conference on Wirtschaftsinformatik*, 637–644.

Alter, S. (2002). *Information Systems: The Foundations of E-Business* (4th ed.). Prentice Hall, Pearson Education International.

Brancheau, J. C., & Wetherbe, J. C. (1986). Information Architectures: Methods and Practice. *Information Processing & Management*, 22(6), 453–463. doi:10.1016/0306-4573(86)90096-8

Business Architecture Guild. (2014). *A Guide to the Business Architecture Body of Knowledge™ (BIZBOK® Guide), Version 4.1. Part 1*. Author.

Grigoriu, A. (2011). *An Enterprise Architecture Development Framework* (3rd ed.). E-book, Trafford.

Grigoriu, A. (2013). *EA Framework*. Retrieved from www.enterprise-architecture-matters.co.uk/ea-links/ea-framework

Guimarães, T. S. M. (2012). *21 principles of enterprise architecture for the financial sector*. Retrieved from www.ibm.com/developerworks/.../enterprise-architecture-financial-sector

Harmon, P. (2004). Enterprise architectures. *Business Process Trends Newsletter*, 21, 1-14.

Harmon, P. (2008). *Review of 'An Enterprise architecture Development Framework' by Adrian Grigoriu*. Business Process Trends Newsletter.

IBM-Cognos. (2011). *Monitor, Manage, Perform*. Retrieved from <http://www-01.ibm.com/software/analytics/cognos/business-intellig>

James, G. (2001, June 12). Architecture must align with business. *The Australian*.

Johnson, M. W., Christensen, C. M., & Kagerman, H. (2008). Reinventing Your Business Model. *Harvard Business Review*.

Koontz, C. (2000). Develop a solid e-commerce architecture. *e-Business Advisor*, 18(1), 22.

Linder, J., & Cantrell, S. (2000). *Changing Business Models: Surveying the Landscape*. Accenture Institute for Strategic Change.

Lindström, Å. (2006). On the Syntax and Semantics of Architectural Principles. In *39th Annual Hawaii International Conference on Systems Sciences*. IEEE Computer Society.

Martin, J. (1990). *Information Engineering, Book II: Planning and Analysis*. Prentice Hall. doi:10.1109/HICSS.2006.367

McKee, J. (2013). *Applying Principles from IT Architecture to Strategic Business Planning*. Hershey, PA: IGI Global. doi:10.4018/978-1-4666-2527-3

McKee, J. (2015). *A Sustainable Business Model Needs to be Underpinned by a Business Architecture*. Download Info: <http://www.hrpub.org/download/20151030/UJM7-12104722.pdf>

Mintzberg, H. (1994). *The Rise and Fall of Strategic Planning*. New York: Prentice Hall.

Momentum. (2011). *Platform Engineering Maturity Model*. sales@momentumsi.com.

Rea, P. J., & Kerzner, H. (1997). *Strategic Planning: A Practical Guide*. John Wiley & Sons Inc.

Reynolds, C. (2009). Introduction to Business Architecture. Boston, MA: Course Technology / Cengage Learning. ProQuest ebrary.

Richardson, G. L., Jackson, B. M., & Dickson, G. W. (1990). A Principle-Based Enterprise Architecture: Lessons From Texaco and Star Enterprise. *MIS Quarterly: Management Information Systems*, 14(4), 285–403. doi:10.2307/249787

Ross, J. W., Weill, P., & Robertson, D. (2006). *Enterprise Architecture as Strategy: Creating a Foundation for Business Execution*. Harvard Business School Press.

Rouse, M. (2007). *Definition of Enterprise Architecture*. Retrieved from www.searchcio.com

Routio, P. (2005). *Arteology: Normative Study*. Retrieved 22/11/2005, 2005, from <http://www2.uiah.fi/projects/metodi/178.htm>

Sereff, G. B. (2012). *Launching an Enterprise Business Architecture Practice: A Playbook for Getting Started*. e-book from Amazon.com.

Sorensen, H. E. (2012). *Business Development: A market oriented perspective*. Wiley.

The Open Group. (2006). *The Business Executive's Guide to IT Architecture: Why IT Architecture is Crucial to the Success of Your Business*. Retrieved from <http://www.opengroup.org/public/arch/pl/oview>

TOGAF (2013). Architecture Principles. *TOGAF 8.1.1 Online > Part IV: Resource Base > Architecture Principles*.

Ulrich, W., & Kuehn, W. (2015). *Business Architecture Setting the Record Straight*. Retrieved from <http://www.businessarchitectureassociates.com>

Zachman, J. A. (1996). *Enterprise Architecture: The issue of the Century*. Retrieved from <http://www.zifa.com/zifajz01>

ADDITIONAL READING

Bernus, P., Nemes, L., & Gunter, S. (2003). *Handbook on Enterprise Architecture*. Springer. doi:10.1007/978-3-540-24744-9

Hope, J. (2010). *How KPIs can help to motivate and reward the right behavior*. Retrieved from <http://www-01.ibm.com/software/analytics/cognos/business-intellig>

IBM-Cognos. (2012). *BI Architecture Explained*. Retrieved from WWW.IBM.com/architecture_matters

Jensen, C. T., Charters, I., Amsden, J., Darlington, S., Owen, M., Herness, E., & Irassar, P. (2008). *Leveraging SOA, BPM and EA for Strategic Business and IT Alignment*. IBM: 1-26.

Lee, J. (2005). *The Metaphor of Architecture*. Josh Lee's Financial Services Blog.

O'Rourke, C., Fishman, N., & Selkow, W. (2003). *Enterprise Architecture Using the Zachman Framework*. Boston: Thompson.

Ross, J. W., Weill, P., & Robertson, D. C. (2013). *Enterprise Architecture as Strategy: Creating a Foundation for Business Execution*. Boston: Harvard Business School.

Saha, P. (2007). *Handbook of Enterprise Systems Architecture in Practice*. Information Science Reference. Hershey, PA: IGI Global. doi:10.4018/978-1-59904-189-6

Spiegel, T., & Caulliraux, H. M. (2012). Developing an Organizational Reference Model by Selecting and Integrating Multiple References. *Journal of Management Research*, 4(1), E2.

TheOpenGroup. (2013). *Modelling tool*. www.opengroup.org/subjectareas/enterprise/archimate

Zachman. (2013). *Enterprise architecture visualization tool*. www.visual-paradigm.com/

KEY TERMS AND DEFINITIONS

Blueprint: A plan of action or a guide to doing something.

Framework: Aggregate of fundamental parts of a structure, or of those which hold the rest together.

Level of Abstraction: The different views of a system from high level to low level, where the high level can be decomposed into lower levels.

Model: A standard pattern or example to follow.

Principle: Ultimate source or origin of anything from which all else is derived. A general and comprehensive law, doctrine or truth as the basis for other laws.

Reference Model or Reference Architecture: Model to guide and improve. These models are sets of structured concepts, guidelines and/or solutions used to guide organisations through a process.

Service Oriented Architecture (SOA): A particular architecture where the application services are provided online through a service provider.

Benchmarking Performance Indicators of Indian Rail Freight by DEA Approach

B**Neeraj Bhanot***Dr. B. R. Ambedkar National Institute of Technology Jalandhar, India***Harwinder Singh***Guru Nanak Dev Engineering College, India*

INTRODUCTION

To accomplish the organization's logistics/distribution activities, measurement systems have been designed to capture information regarding five types of performance. These five categories are: asset management, cost, customer service, productivity and logistics quality (Li, 2005).

Li (2005) discussed several evaluation methods for logistics performance. Qualitative and quantitative data can be dealt by AHP method but the weight of different factors is objective. The result is influenced by the subjectivity greatly. Mathematic statistics method mainly conducts the quantitative data, and the evaluation process is objective. However strong qualitative analysis ability has been shown by Fuzzy comprehensive evaluation and Rough sets method. The above methods are not systematized as a whole. The objects, contents and methods are too much complicated to be integrated. But DEA (Data Envelopment Analysis) especially suitable for multi-input and multi-output complicated systems take the weights of input and output of DMU (Decision Making Units) for variables, evaluates in terms of DMU. In a word, DEA is more suitable evaluation method for the logistics activities as it has very strong objectivity.

Data Envelopment Analysis (DEA)

DEA measurement can be defined as the ratio of total weighted output to total weighted input. It is a linear programming-based technique that

converts multiple input and output measures into a single comprehensive measure of productivity efficiency (Epstein, & Henderson, 1989). DEA provides a measure by which one firm or department can compare its performance, in relative terms, with other homogeneous firms or departments. In measuring the relative efficiencies of organizations with DEA each organization can utilize different weights for the set of performance measures. Weights are selected that will maximize the composite efficiency score for each functional unit. Taking into account differences in goals, responsibilities, and type of procurement, this variable weighting allows for the evaluation of performance. The range of possible weights is controlled by requiring all weights to be positive, and specifying that if another unit used the same weight, their total efficiency score could not exceed the value of one. The technique also gives information as to the specific effect each input or output has on overall efficiency as yet, which has demonstrated its use as a potential evaluation tool for logistics performance. In summation, DEA can be used to reduce some of the existing problems with performance evaluation system (Charnes, Cooper, & Rhodes, 1978).

An examination of the literature reveals the application of DEA for performance benchmarking in a variety of settings such as telecommunications (Lupi, Manenti, Scialà, & Varin, 2011), educational institutes (Portela, Camanho, & Borges, 2011), productive efficiency (Farrell, 1957; Chandraprakaikul, & Suebpongsakorn, 2012; Holden, Xu, Greening, Piecyk, & Dadhich,

DOI: 10.4018/978-1-5225-2255-3.ch051

2016), Irrigation sector (Phadnis, & Kulshrestha, 2012), safety performance (Berih, Patnaik, & Mahapatra, 2011), etc.

Indian Railway Container Scenario

By taking over network of seven Inland Container Depots (ICDs) of Indian Railways (IR) located at Delhi, Ludhiana, Bangalore, Coimbatore, Guwahati, Guntur and Anaparti, Container Corporation of India (CONCOR) was incorporated in March 1988 as a Public Sector Enterprise under the Ministry of Railways. The company was formed with objective to have a separate organization for promoting and managing the growth of containerization in India as well as developing multimodal (surface, rail, water and air) transport logistics and infrastructure to support India's growing inland as well as international trade. Under the supervision of Managing Director, the day-to-day affairs of the business of CONCOR are managed by various divisional heads (CARE, 2009). Although, IR is a public sector monopoly, there is a growing demand on it to realize the full scope of its assets through proper use of systems and technology and also to address the lack of its customer orientation. Therefore, there is a strong need to devise a performance benchmarking system of its supply chain operations (George, & Rangaraj, 2008). To introduce competition in the container operations segment, the Ministry of Railways allowed the entry of private and public sector operators to obtain licenses for running container trains on the Indian Railways (IR) network in January 2006. Until then, the Container Corporation of India, a subsidiary of IR, was the monopoly operator of container trains in India.

This book chapter is an extension of previous study conducted on evaluating the performance of CONCOR over a period of time using super-efficiency analysis by employing CCR and BCC models of DEA considering different parameters for additional years to outline the key issues for needful improvement by validating the results of CCR and BCC models. The subsequent sections narrate thorough study of background of

performance evaluation perspective of logistics operations within and outside country. Further, DEA models, which are used in this study and the application of these models for benchmarking of Indian Railway Container Business within CONCOR, have been explained. The last section presents appropriate discussion on the results obtained and conclusions of the study.

BACKGROUND

Freight Container Business continued to be monopoly in the hands of government owned Container Corporation of India. In this section, history of performance evaluation has been discussed within and outside country in competitive scenario over the years in order to cite the various cases of performance oriented studies related to logistics.

Background Within Country

Napa (2006) discussed strategic moves and logistic operations, CONCOR took to establish it and face the challenges posed by the entry of new players. The task ahead in the company had been discussed in addition to questions, which can be taken as possible new problems for work. This framework further provides scope for readers to analyse various strengths and weaknesses of the CONCOR in the light of the future competition.

George, & Rangaraj (2008) analysed the performance benchmarking study of the zones of Indian Railways (IR) to develop an alternate approach for measurement of aggregate operational performance. The efficiency trends within the set of inputs and outputs considered with efficiency scores ranging between 58.26% to 100% by CCR and 58.26% to 254.92% by RCCR analysis identified the best performing railway zones over the years. The conclusion showed some weaknesses of the conventional DEA due to cross-efficiencies along with self-efficiencies and operating ratios.

Gangwar, & Raghuram (2010) examined the current policy environment from the point of view of business viability for 15 new Container Train

Operators discussing the various problems and prospects of Container Train Operators in India. It further brought out issues related to licensing, pricing, terminals, maintenance, and service levels found good response to the policy as entrants obtained licenses to run container trains. However, due to lack of clarity or inconsistency in matters pertaining to haulage charges, maintenance of wagons, transit guarantees from Indian Railways (IR) and terminal access charges, operators started feeling susceptible about the viability of the business.

Ghosh, & Bandyopadhyay (2014) investigated the efficiency of operating no-frill airlines (LCCs) in eastern India by collecting publicly available data of various parameters related to the technical and financial health of these airlines. Technical efficiency of each no-frill airline in comparison with other no-frill airlines was studied by employing Data Envelopment Analysis on the input and output variables. Further, Super Efficiency Model was applied on the most efficient airlines. The study thus helped in understanding the technical efficiency and performance of operating no-frill airlines in eastern India.

Performance Evaluation Study of Logistics Operations Within and Outside Country

Hilmola (2007) used DEA methodology to discuss about the efficiency and productivity of railway freight transportation sector in Europe. The analysis revealed that former Eastern Bloc and West European countries were showing the highest efficiency levels in the 1980s and experienced an efficiency collapse in the 1990s. Study on partial productivity analysis proposed that productivity of locomotives and railway tracks being primary target for productivity improvement in the highly efficient countries showed approximately 80% increase in transit traffic via rail in all the Baltic states.

Kunadhamraks, & Hanaoka (2008) evaluated the logistics performance of intermodal freight transportation using Fuzzy set theory to deal

with attitudes and perceptions within the decision process. Fuzzy-MCA was used to assess logistics operators' perception of the performance via proper assignment of numerical scores. Sensitivity analyses of significant variables were performed to examine the impact on changed costs and services levels and found that it can entail use of improved corresponding parameters to develop an efficient logistics system in Thailand.

Koster, Balk, & Nus (2009) used data envelopment analysis (DEA) on primary data of large container terminals, APMT (39 terminals) and European terminals of PSA (seven terminals) to discuss the reasons behind diverging results. Qingdao-APMT, Raysut/Salah-APMT, and Singapore-PSA-Brani were reported as the top three efficient terminals with 100% efficiency under CRS and VRS are all in Asia. Houston-APMT would be number four with 38% efficiency. The results differed strongly from those available in the literature and the reason for different terminal types and scales are mixed with the ports.

Mäkitalo, & Hilmola (2010) adopted methodology to foresee the future development of deregulated industries especially in the transportation sector. On the basis of a qualitative expert profiling analysis using Delphi questionnaire directed at 52 Finnish experts, as three different argument types existed in Finnish railway transport policy. Descriptions for each group based on material collected by the Delphi technique and concluded that the policy definitions have followed the views of the moderate group and the realization of the deregulation process has been slow. The research was limited to the Finnish railway freight transport market.

Hilmola (2011) developed efficiency benchmarking measurement models for public transportation systems to decrease environmental emissions in the future (CO₂), worldwide decreasing reserves of oil, and growing population in larger cities in line with the global agreements. The author used four different DEA based efficiency benchmarking models to evaluate public transportation efficiency in larger cities and found that medium sized, old and Central European cities

show frontier performance in all four models whereas Mega-cities fail to reach frontier and/or good performance.

Lee, Yeo, & Thai (2014) applied a slacks-based data envelopment analysis (SBM-DEA) model to assess the environmental efficiency of port cities by considering labor population in respective port cities as input variable, and gross regional domestic product (GRDP) and container throughput as output variables. The study also considered various undesirable output variables such as, nitrogen oxide (NO_x), sulfur oxide (SO₂), and carbon dioxide (CO₂) emissions. The results of the study indicated Singapore, Busan, Rotterdam, Kaohsiung, Antwerp, and New York as the most environmentally efficient port cities with Tianjin as the least environmentally efficient.

Shao, & Sun (2016) categorised the production process of air routes into two stages viz., allocation and transport, and then proposed two network data envelopment analysis (DEA) models to analyse the efficiency of the system, allocation, passenger transport, and freight transport of 477 air routes. The results of the study highlighted the significant impact of different constraints on intermediate measure in the network DEA models affecting the air routes' efficiency significantly. In addition to this, most of the air routes depicted high allocation and passenger transport efficiency with low freight transport efficiency.

MAIN FOCUS

Data Envelopment Analysis Approach

Linear programming (LP) multi-factor productivity analysis model Data Envelopment Analysis approach (DEA) can be used to measure relative efficiency of homogenous set of DMUs. It calculates a maximal performance measure for each DMU relative to all other DMUs with an objective that each DMU lie on or below the external frontier. Without requiring any specific assumptions about the functional form, it optimizes on each

individual observations with the purpose of calculating a discrete piecewise frontier determined by the set of Pareto-efficient DMUs. Each DMU not on the frontier is scaled down against a convex combination of the DMUs on the frontier facet closest to it.

The usual measure of efficiency, i.e.:

$$efficiency = \frac{output}{input}$$

is often inadequate related to different resources, activities and environmental factors due to the existence of multiple inputs and outputs. DEA methodology is developed to solve this problem.

There are two basic DEA models—CCR model, developed by Charnes, Cooper and Rhodes in 1978 and BCC model, developed by Banker, Charnes, and Cooper in 1984. CCR model generalizes the single output/input ratio measure of efficiency for a single DMU in terms of fractional linear programming (FLP) formulation transforming the multiple output/input characteristics of each DMU to that of a single “virtual” output and “virtual” input. The model defines the relative efficiency for any DMU as a weighted sum of outputs divided by a weighted sum of inputs where all efficiency scores are restricted to lay between zero and one. An efficiency score less than one means that a linear combination of other units from the sample could produce the same vector of outputs using a smaller vector of inputs. The score reflects the radial distance from the estimated production frontier to the DMU under consideration. In order to calculate efficiency scores, FLP is converted into LP by normalizing either the numerator or the denominator of the fractional programming objective function. In case of input-minimization DEA program, the weighted sum of outputs is constrained to be unity to minimize weighted sum of inputs while in output – maximization DEA program, the weighted sum of inputs is constrained to be unity to maximize weighted sum of outputs. CCR model is based on constant returns to scale assumption. Under this assumption, if the input

levels of a feasible input-output correspondence are scaled up or down, then another feasible input-output correspondence is obtained in which the output levels are scaled by the same factor as the input levels (Charnes *et al.*, 1978).

CCR Model

This model generalizes the usual input/output ratio measure of efficiency for a given firm in terms of a fractional linear program formulation. Mathematically, the relative efficiency of the k^{th} DMU is given by:

$$\text{Max } h_k = \frac{\sum_{r=1}^s u_{rk} y_{rk}}{\sum_{i=1}^m v_{ik} x_{ik}} \quad (1)$$

Subjected to:

$$\frac{\sum_{r=1}^s u_{rk} y_{rj}}{\sum_{i=1}^m v_{ik} x_{ij}} \leq 1 \quad \forall j = 1 \dots k \dots n$$

$$\frac{u_{rk}}{\sum_{i=1}^m u_{ik} x_{ik}} \geq \varepsilon \quad \forall r = 1 \dots s$$

$$\frac{v_{ik}}{\sum_{i=1}^m v_{ik} x_{ik}} \geq \varepsilon \quad \forall i = 1 \dots m$$

where: y_{rk} = the amount of the r^{th} output produced by the k^{th} DMU; x_{ik} = the amount of the i^{th} input used by the k^{th} DMU; u_{rk} = the weight given to the r^{th} output of the k^{th} DMU; v_{ik} = the weight given to the i^{th} input of the k^{th} DMU; n = no. of DMUs; s = no. of outputs; m = no. of inputs; and ε = a non-Archimedean (infinitesimal) constant.

The above objective function is reformulated in LP problem as follows:

$$\text{Max } w_k = \sum_{r=1}^s \mu_{rk} y_{rk} \quad (2)$$

$$\sum_{i=1}^m v_{ik} x_{ik} = 1$$

$$\text{Subjected to } \sum_{r=1}^s \mu_{rk} y_{rj} - \sum_{i=1}^m v_{ik} x_{ij} \leq 0$$

$$\forall j = 1 \dots n$$

$$\mu_{rk} \geq \varepsilon \quad \forall r = 1 \dots s$$

$$v_{ik} \geq \varepsilon \quad \forall i = 1 \dots m$$

Since the number of DMUs is generally larger than the total number of inputs and outputs, solving the dual of the model can reduce the computational burden. Mathematically, the dual formulation of the above model is:

$$\text{Min } z_k = \theta_k - \varepsilon \sum_{r=1}^s S_{rk}^+ - \varepsilon \sum_{i=1}^m S_{ik}^- \quad (3)$$

Subjected to

$$\sum_{j=1}^n \lambda_{jk} y_{rj} - S_{rk}^+ = y_{rk} \quad \forall r = 1 \dots s$$

$$\sum_{j=1}^n \lambda_{jk} x_{ij} + S_{ik}^- = \theta_k x_{ik} \quad \forall i = 1 \dots m$$

$$\lambda_{jk} \geq 0 \quad \forall j = 1 \dots n$$

θ_k free

$$S_{rk}^+, S_{rk}^- \geq 0; \quad r = 1 \dots s, i = 1 \dots m$$

where: S_{rk}^+ = Slacks in the i^{th} input of the k^{th} DMU; S_{ik}^- = slacks in the r^{th} output of the k^{th} DMU; λ_{jk} 's = non-negative dual variables; θ_k (scalar) is the (proportional) reduction applied to all inputs of DMU k to impose efficiency. If for DMU k , $\theta_k^* = 1$ and all slacks are zero, it is Pareto efficient. The non-zero slacks and (or) $\theta_k^* \leq 1$ identify the sources and amount of any inefficiency that may exist in the DMU under reference.

Efficiency scores are constructed by measuring how far a utility is from the frontier. A test DMU is considered inefficient if a composite DMU (defined as linear combination of units in the set) can be identified which utilizes less input than the test DMU while maintaining the same or greater output levels (Phadnis, & Kulshrestha, 2012).

BCC Model

The convexity constraint is the primary difference between BCC model and CCR model. In the BCC model λ_{jk} are restricted to summing to one (i.e. $\sum_{j=1}^n \lambda_{jk} = 1$). The model is converted into Non-Increasing Returns to Scale (NIRS) model, if we impose $\sum_{j=1}^n \lambda_{jk} \leq 1$ instead of $\sum_{j=1}^n \lambda_{jk} = 1$. Similarly if we impose $\sum_{j=1}^n \lambda_{jk} \geq 1$ instead of $\sum_{j=1}^n \lambda_{jk} = 1$, then the model is known as Non-Decreasing Returns to Scale (NDRS) model.

DEA Super-Efficiency Model

The discriminatory power of super-efficiency model provides insights that cannot be gained with the standard DEA model. The DEA score for the inefficient unit is considered as its rank scale. In order to rank scale the efficient units they allow the efficient units to receive an efficiency score greater than 100 percent by dropping the constraint that bounds the score of the evaluated unit. In standard DEA, Decision Making Units (DMUs) are identified as fully efficient and assigned an efficiency score of unity if they lie on the efficient frontier. Inefficient DMUs are assigned scores of less than unity (Yawe, 2010).

A number of uses have been proposed for super-efficiency models (Lovell, & Rouse, 2003). These include:

1. Ranking of efficient DMUs;
2. Classification of DMUs into extreme-efficient and non-extreme efficient groups;
3. Sensitivity of efficiency classifications;
4. Two-person ratio efficiency games;
5. Identifying outliers in the data;
6. Overcoming truncation problems in second-stage regressions intended to explain variation in efficiency;
7. Calculating and decomposing a Malmquist productivity index.

Super-efficiency measures are constructed by avoiding that the evaluated firm can help span the technology (Bogetoft, & Otto, 2011). Let $T^*(\gamma | -k)$ be a DEA approximation of the technology using the γ assumptions and based on all observations but that of firm k (Box 1).

Now the efficiency of (x^k, y^k) relative to $T^*(\gamma | -k)$ is called super-efficiency.

$$E^{SUPk} = E((x^k, y^k); T^*(\gamma | -k)),$$

$$F^{SUPk} = F((x^k, y^k); T^*(\gamma | -k)).$$

APPLICATION TO BENCHMARKING OF INDIAN RAILWAY CONTAINER BUSINESS

The above-described DEA models are used for the performance benchmarking of the Indian Railway container business. The implementation of DEA however involves recognizing the inputs

Box 1.

$$T^*(\gamma | -k) = \left\{ (x, y) \in \mathbb{R}_+^m \times \mathbb{R}_+^n \mid \exists \lambda \in \wedge^{K-1}(\gamma) : x \geq \sum_{j \neq k} \lambda^j x^j, y \leq \sum_{j \neq k} \lambda^j y^j \right\}$$

and outputs of the units being assessed, selecting measures for the inputs and outputs, collecting data on the inputs and outputs, solving the appropriate models and interpreting the results (Thanassoulis, Dyso, & Foster, 1987).

Inputs and Outputs/ Performance Indicators

The performance evaluation using DEA starts with selection of appropriate input and output measures that can be combined into a composite index of overall performance. Inputs are basically the different resources that the firms consume for its operations while the outputs represent a set of quantitative measures of results expected from these firms (George, & Rangaraj, 2008). In this study, they were chosen based on previous DEA studies reported in literature and also on the availability of data. Table 1 highlights the list of performance indicators used for the study.

Data Analysis

The secondary data from year 1994-95 till 2015-16 has been used for performance evaluation within CONCOR and shown in Table 2.

The secondary data from all the organizations has been collected through personal interviews. However Figure 1 shows methodology adapted in present study in a flow chart.

Table 1. Performance indicators to be used

S.No.	Performance Indicators	Type
1.	Number of Handling Terminals	Input
2.	Number of Employees	Input
3.	Number of freight wagons	Input
4.	Number of Containers (TEUs)	Input
5.	Net Profit (in Crores)	Output
6.	TEUs Handled	Output

Sources: Azevedo, Ferreira, Dias, & Palma, 2009; Morris, Pandey, Raghuram, & Gangwar, 2010; George, & Rangaraj, 2008; Hilmola, 2007

In order to gain deep insights in the study, Super-efficiency analysis have been deployed so as to verify the results by both CCR and BCC models. Further depending on whether inputs and outputs are controllable, a Decision Making Unit (DMU) can have either an input orientation or output orientation (George, & Rangaraj, 2008). As such the study is based on input orientation.

In order to evaluate the performance within CONCOR the years 1994-95 to 2015-16 have been the Decision Making Units (DMU's) along with respective inputs and outputs in order to determine the efficiency trends.

To reach the conclusion, Spearman Rank Correlation Coefficient (r_s) has been used to correlate the results shown by CCR and BCC Models as per the formulae below:

$$r_s = 1 - \left(\frac{6 \sum d^2}{n(n^2 - 1)} \right)$$

wherein, r_s varies from -1 to +1 and is weak positive correlated if its value is between 0 and 0.5 & strong positive correlated if its value is between 0.5 and 1. Same is the case with negative correlation when its value varies between 0 and -1.

RESULTS AND DISCUSSION

In this section, results for the performance evaluation within CONCOR from 1994-95 till 2015-16 have been suitably presented to highlight various constraints in order to improve their efficiency.

In case of analysis by Super-efficiency Model, Table 3 showed Efficiency Analysis for CONCOR by CCR and BCC models for the years 1994-95 to 2015-16 with fluctuation of 75.83% to 109.51% and 93.52% to 1000% respectively. The revelation through efficiency analysis using CCR and BCC models indicated a Spearman Rank Correlation Coefficient (r_s) value of 0.60, which clearly reflects a strong positive correlation value.

Table 2. Secondary Data for CONCOR from 1994-95 till 2015-16

Year	No. of Handling Terminals	No. of Employees	No. of Freight Wagons	No of Containers (TEUs)	Net Profit (in Crores)	TEUs Handled
1994-95	11	488	3835	7800	23.89	402632
1995-96	14	502	4214	8216	52.23	594118
1996-97	16	546	4558	8802	95.25	703542
1997-98	20	590	5142	9417	115.8	721719
1998-99	22	602	5560	9678	140.66	801946
1999-00	25	628	5672	9850	177.59	903151
2000-01	31	656	5768	9975	216.65	1044728
2001-02	40	781	5845	10156	249.48	1231833
2002-03	42	841	5920	10314	272.85	1383163
2003-04	47	915	6012	10771	367.59	1602119
2004-05	52	988	7582	10874	428.6	1727976
2005-06	54	1036	7762	11745	525.8	1930562
2006-07	59	1080	8577	12812	703.82	2105266
2007-08	61	1134	8421	13517	752.21	2447769
2008-09	63	1176	8668	13576	791.2	2308232
2009-10	64	1129	10194	15579	786.69	2361247
2010-11	66	1147	10666	15754	875.95	2562597
2011-12	68	1164	10988	16109	877.88	2604311
2012-13	71	1198	11770	18680	940.03	2585686
2013-14	71	1325	12111	20984	984.76	2868612
2014-15	73	1331	13111	20247	1047.55	3110756
2015-16	73	1332	13471	20944	786.93	2924046

The years pertaining to lower efficiencies can be attributed to the lack of infrastructural facility and their proper management as visible from potential improvements as in case of CCR Model. While, in case of BCC Model, the lowest efficiency has been attributed to the year 1997-98 since the cargo handlings and exports were declined on account of incessant rains and floods causing disruption of traffic movement. Thus there was insignificant growth into logistic business. However, the successive years showed little better towards growth considering the efficiency improvement factor may be due to use of better managerial skills. The overall issues pertaining to lower performance have been presented at the end of analysis.

Table 4 and Table 5 again express the potential improvement by Super-efficiency DEA study for performance indicators by CCR and BCC models respectively, which indicates by how much, and in what areas an inefficient unit needs to improve in order to be efficient. The negative percentage shows the amount of reduction needed for highlighting the specific variables in order to increase the efficiency.

The CCR model analysis from Table 4 concluded that throughout the discussed inefficient years, it was evident that the number of employees and number of containers have shown highest requirement for potential improvement with percentage fluctuation varying between -44.87%

Figure 1. Methodology for benchmarking CONCOR performance by DEA approach

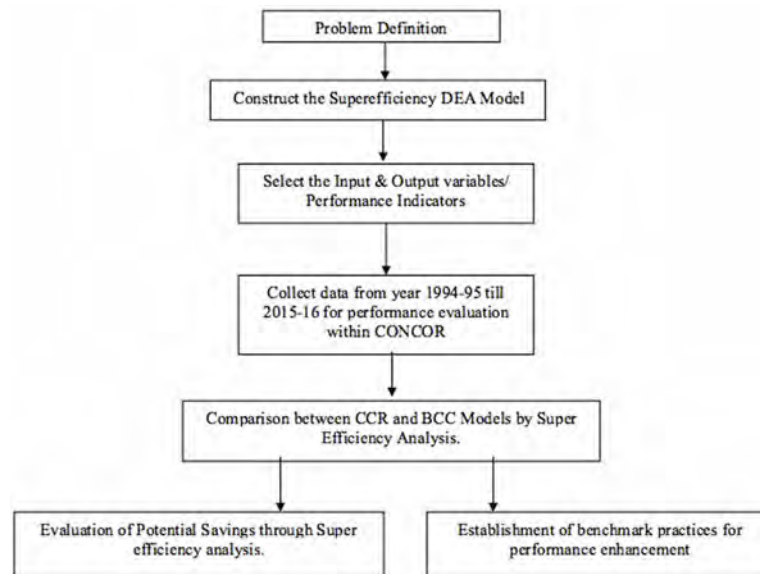


Table 3. Super-Efficiency Analysis within CONCOR

Year	Efficiency (CCR model)	Ranking (CCR Model)	Efficiency (BCC Model)	Ranking (BCC Model)
1994-95	83.24	5	127.27	21
1995-96	96.51	14	104.56	17
1996-97	103.19	19	103.53	16
1997-98	82.81	4	93.52	1
1998-99	84	7	94.89	2
1999-00	83.64	6	96.06	4
2000-01	78.93	2	99.14	8
2001-02	75.83	1	96.91	6
2002-03	81.72	3	98.7	7
2003-04	91.68	10	108.15	19
2004-05	87.75	8	102.49	15
2005-06	90.77	9	99.47	10
2006-07	95.77	13	101.29	12
2007-08	109.51	22	112.11	20
2008-09	104.73	20	105.07	18
2009-10	93.2	11	96.28	5
2010-11	101.98	18	102.25	14
2011-12	99.51	17	99.78	11
2012-13	99.84	16	101.48	13
2013-14	99.32	15	99.34	9
2014-15	109.23	21	1000	22
2015-16	94	12	95.27	3

to +9.51% and -47.98% to +6.81% respectively in comparison to other performance indicators.

However The BCC model analysis from Table 5 concluded that throughout the discussed inefficient years, the number of handling terminals have shown highest requirement for potential improvement with percentage fluctuation varying between -16.62% to +27.27% with requirement from number of freight wagons also with percentage fluctuation ranging between -11.77% to +12.11% respectively.

This percentage fluctuation in CCR and BCC models signifies the variation of percentage potential improvement and suitable reduction or enhancement in highlighted performance indicators could have increased the overall performance of

CONCOR for respective inefficient years. Thus the study from Super-efficiency DEA revealed some of the following issues which needs to be taken care of to improve and maintain the performance level of CONCOR:

- Focus of management remained developing the particular field instead of moving step by step in a planned strategy for all-round growth of the complete set of activities.
- It is also clear that no planned policy for infrastructure development was enforced whereas the licensing norms too added in-different scenario.

Table 4. Potential improvements performance indicator wise within CONCOR by CCR model through Super-efficiency analysis (in percent)

Year	No. of Handling Terminals	No. of Employees	No. of Freight Wagons	No of Containers (TEUs)	Net Profit (in Crores)	TEUs Handled
1994-95	-16.76	-35.97	-31.98	-35.42	128.17	0
1995-96	-3.49	-8.15	-8.66	-9.53	54	0
1996-97	3.19	-44.87	-34.94	-47.98	148.73	0
1997-98	-17.19	-17.19	-18.11	-17.21	20.1	0
1998-99	-16	-16	-20.2	-17.11	25.15	0
1999-00	-16.36	-16.66	-17.2	-16.36	26.68	0
2000-01	-21.07	-28.35	-21.07	-27.86	56.25	0
2001-02	-24.17	-28.07	-24.17	-30.58	54.7	0
2002-03	-18.28	-24.24	-18.28	-24.97	56.89	0
2003-04	-15.05	-18.88	-8.32	-17.86	33.94	0
2004-05	-17.19	-18.97	-21.59	-12.25	23.9	0
2005-06	-10.91	-13.67	-14.43	-9.23	12.83	0
2006-07	-5.25	-4.23	-8.47	-4.23	0	0
2007-08	9.12	9.51	9.51	6.81	11.49	0
2008-09	1.84	1.43	2.19	4.73	0	11.54
2009-10	-7.89	-6.8	-6.8	-7.36	0	0
2010-11	1.98	1.98	1.57	1.98	0	1.25
2011-12	-3.58	-0.43	-2.89	-0.43	0	0
2012-13	-7.4	-0.16	-0.16	-3.04	0	7.89
2013-14	-0.68	-2.47	-0.68	-11.26	0	1.6
2014-15	9.23	5.15	-3.76	-6.96	0	0
2015-16	-6	-6	-8.47	-9.06	25.06	0

Table 5. Potential improvements performance indicator wise within CONCOR by BCC model through super-efficiency analysis (in percent)

Year	No. of Handling Terminals	No. of Employees	No. of Freight Wagons	No of Containers (TEUs)	Net Profit (in Crores)	TEUs Handled
1994-95	27.27	2.87	9.88	5.33	118.63	47.56
1995-96	1.3	4.56	1.92	2.7	32.68	0
1996-97	3.53	-1.46	0.94	-0.71	0.27	0
1997-98	-9.93	-6.48	-8.36	-6.58	0	3.37
1998-99	-10.99	-5.11	-11.75	-6.74	0	0.44
1999-00	-11.35	-3.94	-7.85	-4.57	2.23	0
2000-01	-16.62	-0.86	-2.28	-0.86	10.13	0
2001-02	-15.16	-3.09	-3.09	-3.09	6.33	0
2002-03	-4.82	-1.45	-1.3	-1.3	10.92	0
2003-04	-15.83	-7.58	8.15	3.04	17.76	0
2004-05	-4.49	2.69	-11.77	2.49	0	0
2005-06	-0.6	-1.47	-0.53	-0.53	0	0
2006-07	-3.17	1.29	-4.7	1.29	0	0
2007-08	6.13	6.08	12.11	9.02	11.11	0
2008-09	-0.67	-2.79	5.07	5.07	0	8.13
2009-10	-5.32	-3.72	-3.72	-4.97	0	0
2010-11	2.25	2.25	1.89	2.25	0	1.35
2011-12	-2.25	-0.22	-1.51	-0.22	1.1	0
2012-13	-3.36	1.48	-1.62	-6.68	0	7.02
2013-14	-0.66	-2.43	-0.66	-11.23	0	1.61
2014-15	0	0	0	0	0	0
2015-16	-4.73	-4.73	-8.17	-9	24.64	0

- The declines in the efficiency lead on account of unplanned expansion of terminals and infrastructure growth haphazardly without caring competitive strategies.
- The deficient services were due to monopoly of CONCOR till 2005, the work culture generally relied on irregular practices like pick and choose, bribery, tip to staff.

The strategic weaknesses and unplanned development can however be tackled efficiently as CONCOR has sufficient infrastructure in light of competition but there is a need to utilize it optimally and as of now, it seems that to some extent, proper measures have started being taken

to enhance the performance of the organization with reference to concurrent efficiency measures. However, it cannot be denied that the reliability of data is a primary concern which if forged will not lead to highlight the true issues, which reduce the efficiency.

CONCLUSION

The Super-Efficiency DEA study identified efficiency fluctuation between 75.83% to 109.51% (CCR Model) and 93.52% to 1000% (BCC Model) owing to lack of efficient staff and operational planning in rake management in addition to

haphazard infrastructure growth viz. number of handling terminals, number of handling equipments and containers without caring for optimal requirements and managing properly. The same being concurrent to licensing policy norms, which needs to be thoroughly formulated in a phased and realistic manner to boost up the trade and industry within the set of inputs and outputs considered. It will also be beneficial if CONCOR focuses on enhancing the productivity of staff by providing them suitable reading to enhance their skill level which ultimately will lead to enhanced performance level.

LIMITATIONS

Though a lot of advantages, there are limitations for this work as the private operators engaged in the transportation business of containers could not be directly approached owing to time and cost constraints. Only government operator being CONCOR has been considered for the internal performance evaluation over the period of time. Further the reliability of data provided by the operator for needful research is based merely on the information and available sources.

FUTURE RESEARCH DIRECTIONS

The present study can be extended to other private operators e.g. Adani Logistics, Reliance Infrastructure, etc. operating in other parts of India for their individual and competitive performance comparisons. It can also be expanded by considering some more performance indicators viz. Cost, Service/quality, Productivity, etc. to gain deeper insights to enhance the performance of the organizations.

REFERENCES

- Azevedo, S. G., Ferreira, J., Dias, J. C. Q., & Palma, S. F. (2009). *Performance evaluation of the main terminal containers in the Iberian seaports: A benchmarking approach*. POMS, 20th Annual Conference, Orlando, FL.
- Beriha, G. S., Patnaik, B., & Mahapatra, S. S. (2011). Safety performance evaluation of Indian organizations using data envelopment analysis. *Benchmarking: An International Journal*, 18(2), 197–220. doi:10.1108/14635771111121676
- Bogetoft, P., & Otto, L. (2011). Additional topics in DEA - Benchmarking with DEA, SFA, and R. *International Series in Operations Research & Management Science*, 115-153.
- Chandraprakaikul, W., & Suebongsakorn, A. (2012). Evaluation of logistics companies using data envelopment analysis. *4th IEEE International Symposium on Logistics and Industrial Informatics*, 81-86. doi:10.1109/LINDI.2012.6319466
- Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2(6), 429–444. doi:10.1016/0377-2217(78)90138-8
- de Koster, M. B. M., Balk, B. M., & van Nus, W. T. I. (2009). On using DEA for benchmarking container terminals. *International Journal of Operations & Production Management*, 29(11), 1140–1155. doi:10.1108/01443570911000168
- Epstein, M. K., & Henderson, J. C. (1989). Data envelopment analysis for managerial control and diagnosis. *Decision Sciences*, 20(1), 90–119. doi:10.1111/j.1540-5915.1989.tb01399.x
- Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society. Series A (General)*, 120(3), 253–290. doi:10.2307/2343100

- Gangwar, R., & Raghuram, G. (2010). *Container train operators in India: Problems and prospects. Research and Publications, Indian Institute of Management. IIM-A.*
- George, S. A., & Rangaraj, N. (2008). A performance benchmarking study of Indian railway zones. *Benchmarking: An International Journal*, 15(5), 599–617. doi:10.1108/14635770810903178
- Ghosh, P., & Bandyopadhyay, G. (2014). Efficiency and ranking of operating no-frill airlines in Eastern India: An application of data envelopment analysis (DEA). *2nd International Conference on Business and Information Management (ICBIM)*, 136–140. doi:10.1109/ICBIM.2014.6970975
- Hilmola, O. P. (2007). European railway freight transportation and adaptation to demand decline - Efficiency and partial productivity analysis from period of 1980–2003. *International Journal of Productivity and Performance Management*, 56(3), 205–225. doi:10.1108/17410400710731428
- Hilmola, O. P. (2011). Benchmarking efficiency of public passenger transport in larger cities. *Benchmarking: An International Journal*, 18(1), 23–41. doi:10.1108/14635771111109805
- Holden, R., Xu, B., Greening, P., Piecyk, M., & Dadhich, P. (2016). Towards a common measure of greenhouse gas related logistics activity using data envelopment analysis. *Transportation Research Part A, Policy and Practice*, 91, 105–119. doi:10.1016/j.tra.2016.06.001
- Kitalo, M. M., & Hilmola, O. P. (2010). Analysing the future of railway freight competition: A delphi study in Finland. *Foresight*, 12(6), 20–37. doi:10.1108/14636681011089961
- Kunadhamraks, P., & Hanaoka, S. (2008). Evaluating the logistics performance of intermodal transportation in Thailand. *Asia Pacific Journal of Marketing and Logistics*, 20(3), 323–342. doi:10.1108/13555850810890084
- Lee, T., Yeo, G. T., & Thai, V. V. (2014). Environmental efficiency analysis of port cities: Slacks-based measure data envelopment analysis approach. *Transport Policy*, 33, 82–88. doi:10.1016/j.tranpol.2014.02.009
- Li, W. (2005). Research on the performance evaluation of logistics activities. *China USA Business Review*, 4(4), 53–55.
- Lovell, C. A. K., & Rouse, A. P. B. (2003). Equivalent standard DEA models to provide super-efficiency scores. *The Journal of the Operational Research Society*, 54(1), 101–108. doi:10.1057/palgrave.jors.2601483
- Lupi, P., Manenti, F., Scialà, A., & Varin, C. (2011). On the assessment of regulators efficiency: An application to European telecommunications. *Info*, 13(1), 61–73. doi:10.1108/14636691111101883
- Morris, S., Pandey, A., Raghuram, G., & Gangwar, R. (2010). Introducing competition in container movement by rail. Research and Publications, Indian Institute of Management Ahmedabad.
- Napa, J. (2006). *CONCOR: No more a monopoly*. IBS Center for Management Research, CLINDM/004.
- Phadnis, S. S., & Kulshrestha, M. (2012). Evaluation of irrigation efficiencies for water users associations in a major irrigation projects in India by DEA. *Benchmarking: An International Journal*, 19(2), 193–218. doi:10.1108/14635771211224536
- Portela, M. C. A. S., Camanho, A. S., & Borges, D. N. (2011). BESP: Benchmarking of Portuguese secondary schools. *Benchmarking: An International Journal*, 18(2), 240–260. doi:10.1108/14635771111121694
- CARE Ratings. (2009). Container Corporation of India Limited. *Careview*, 353–354.
- Shao, Y., & Sun, C. (2016). Performance evaluation of Chinas air routes based on network data envelopment analysis approach. *Journal of Air Transport Management*, 55, 67–75. doi:10.1016/j.jairtraman.2016.01.006

Thanassoulis, E., Dyso, R. G., & Foster, M. J. (1987). Relative efficiency assessment using DEA: An application to data on rates departments. *The Journal of the Operational Research Society*, 38(5), 397–411. doi:10.1057/jors.1987.68

Yawe, B. L. (2010). Hospital performance evaluation in Uganda: A super-efficiency data envelope analysis model. *Zambia Social Science Journal*, 1(1), 79–105.

ADDITIONAL READING

Chandraprakaikul, W., & Suebpongsakorn, A. (2012). Evaluation of logistics companies using data envelopment analysis. In *4th IEEE International Symposium on Logistics and Industrial Informatics*, 81-86. doi:10.1109/LINDI.2012.6319466

Ghosh, P., & Bandyopadhyay, G. (2014). Efficiency and ranking of operating no-frill airlines in Eastern India: An application of data envelopment analysis (DEA). In *2nd International Conference on Business and Information Management (ICBIM)*, 136-140. doi:10.1109/ICBIM.2014.6970975

Kunadhamraks, P., & Hanaoka, S. (2008). Evaluating the logistics performance of intermodal transportation in Thailand. *Asia Pacific Journal of Marketing and Logistics*, 20(3), 323–342. doi:10.1108/13555850810890084

KEY TERMS AND DEFINITIONS

Benchmarking: It is a management tool through which a plan for evaluation, measurement, and improvement is implemented.

Container Corporation of India: Container Corporation of India Ltd. (CONCOR) is a Category I Miniratna Public sector undertaking under the Indian Ministry of Railways. Incorporated in March 1988 under the Companies Act, Concor commenced operations in November 1989 taking over an existing network of seven inland container depots (ICDs) from Indian Railways.

Data Envelopment Analysis: Data envelopment analysis (DEA) is a nonparametric method in operations research and economics for the estimation of production frontiers. It is used to empirically measure productive efficiency of decision making units (or DMUs).

Indian Railways: Indian Railways (IR) is an Indian state-owned enterprise, owned and operated by the government of India through the Ministry of Railways.

Performance Evaluation: The analysis in terms of initial objectives and estimates, and usually made on site, of accomplishments using an automatic data-processing system, to provide information on operating experience and to identify corrective actions required, if any.

Super-Efficiency: The Super-efficiency analysis allows calculating efficiency improvements for efficient units.

Bi-Directional Business/IT Alignment

B

Hashim Chunpir

*German Climate Computing Centre (DKRZ),
Germany*

Frederik Schulte

University of Hamburg, Germany

Yannick Bartens

University of Hamburg, Germany

Stefan D. Voß

University of Hamburg, Germany

INTRODUCTION

Business/IT alignment – as the process of aligning business and IT to achieve organizational objective and competitive advantages – is a core task of information management (IM) with implications for IT innovations management. Recent research has revealed the need to depict business/IT alignment as bi-directional process in IT governance frameworks to appropriately capture customer-driven IT requirements (Bartens et al., 2016; 2014a; 2014b; Chen, 2008; Patel, 2002, 2004).

The requirement for a bi-directional business/IT alignment becomes obvious regarding the case of e-business (Bartens et al., 2014b) and in two-sided markets (Bartens et al., 2016, 2014a). In this context, technological development often leads to altered customer requirements that trigger a need to adjust business models. Prominent examples can be observed in the media streaming movement. For a long time media were sold as products such as books, records, DVDs, etc. With the emergence of streaming technology, the respective business models faced a drastic need to adapt to flexibility and technology requirements of customers. While those requirements are incorporated in the discussion around buzz words like digital transformation (Patel & McCarthy, 2000) and digital transition (McFadden, 2012), they involve the reversal of the causal chain of requirements in IM. In contrast to current approaches, as the top-down business/IT alignment in the IT governance framework COBIT 5, in these business models, requirements can emerge out of the business operations and

may be defined by the customer. The business cooperation of T-Mobile and Spotify Inc., where the fulfillment of customer needs required the adjustment of both business models (Bartens et al., 2016), may serve as a specific example.

Current IT governance frameworks, definitions and models of IM predominantly still assume a one-directional top-down business/IT alignment process (Krcmar, 2010; Singh & Woo, 2009; Voß & Gutenschwager, 2001; Wollnik, 1988). Thus, they neglect the requirements of changing business models, e.g., in e-business. In the light of the work by Pérez Lorences and García Ávila (2013) as well as De Haes et al. (2013), IM and IT governance is a focus of discussion here and it is put into perspective in this article. Pathways on how IM models and definitions should be extended to integrate bi-directional business/IT alignment are introduced. The remainder of this article is structured as follows. Section 2 provides a brief literature review of IM and business/IT alignment. Section 3 proposes possible extensions of current IM models and definitions to cover bi-directional business/IT alignment. Section 4 gives conclusions and implications for related research.

BACKGROUND: FUNDAMENTALS OF BUSINESS IT/ ALIGNMENT

In this section, a background on the fundamentals of business/IT alignment is given. The alignment of the strategic plans of a business and its IT can be considered an important responsibility and

DOI: 10.4018/978-1-5225-2255-3.ch052

function of IM (Krcmar, 2010). Furthermore, the current state of IM and business/IT alignment is defined and presented in the following.

Information Management

While a broad range of definitions on IM — even contradicting ones — are proposed in literature, the definition by Voß and Gutenschwager (2001) serves as a fundament for the discussion in this article. According to Voß and Gutenschwager, IM is seen as the economic (efficient) planning, acquisition, processing, distribution and allocation of information as a resource for the preparation and support of decisions (decision-making), as well as designing the required infrastructure for this purpose. Applying a similar definition, the majority of authors, argue that IM can be subdivided into several fields of action. One of them is a strategic field in which business IT/alignment as well as the definition and implementation of IT governance are key functions (Boaden & Lockett, 1991; Choo, 2002; Sabherwal et al., 2001; Krcmar, 2010; Voß & Gutenschwager, 2001). The above definitions do not explicitly incorporate business/IT alignment, but the distribution of information in order to support decisions and designing the corresponding circumstances as, e.g., described by Voß and Gutenschwager (2001), may be understood as an abstract depiction of a business/IT alignment process.

Business/IT Alignment and IT Governance

The idea of aligning business and IT functions in order to facilitate business value through the usage of IT has received plenty of attention in recent years (Chan & Reich, 2007; De Haes et al., 2013; Luftman & Kempaiah, 2007; Van Grembergen & De Haes, 2009). Important contributions were provided by the studies of Henderson and Venkatraman (1993; 1999) and have been taken as reference by many authors up to today. In this line, Duffy (2002) describes business/IT alignment

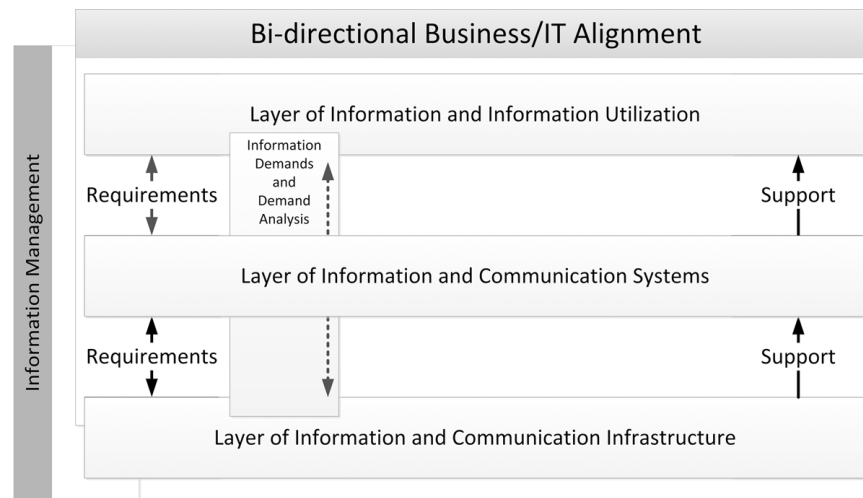
as “the process and goal of achieving competitive advantage through developing and sustaining a symbiotic relationship between business and IT.” Recognizing the previously presented explanations on the definition of IM, business/IT alignment is widely considered part of IM (Maes et al., 2000).

Currently, scientific and practical insights on business/IT alignment are dominated by top-down approaches (Chen, 2008; J. C. Henderson & Venkatraman, 1993; Kearns & Sabherwal, 2007; Kooper, Maes, & Lindgreen, 2011; Lainhart, 2001; N. V. Patel, 2002, 2004). In practice, business/IT alignment is often guided by IT governance frameworks. The COBIT 5 Goals Cascade (ISACA, 2012a), e.g., applies a top-down alignment in this context (ISACA, 2012b; Krcmar, 2010; Van Grembergen & De Haes, 2009). Nonetheless, recent research reveals the existence and need of bottom-up alignment processes (Bartens et al., 2014b). In its origin this need is identified in highly dynamic environments, as well as sectors where the influence of IT on operations is above average. The acknowledgement and combination of the given single aspects by different authors form a holistic array of arguments for a transferability of this hypothesis to a general viewpoint. At this point, the contributions of Györy et al. (2012) (user-driven innovations, shadow IT), Patel (2002) (emergent organizations), Bartens et al. (2014a, 2014b, 2016) (merger of IT management and operations, requirements in two-sided markets) as well as Chen (2008) (bottom-up requirements and alignment) have to be mentioned.

BI-DIRECTIONAL ALIGNMENT IN INFORMATION MANAGEMENT

The introductory sections revealed the ties between business/IT alignment and IM. In the following, the integration of the recent approaches on considering customer-driven or bottom up-requirements in business/IT alignment is outlined and IT governance implications are discussed. In addition, the proposed model is contrasted against different

Figure 1. Integrated bi-directional information management model



definitions on IM and the views on this topic in frameworks such as COBIT 5 and ITIL 2011.

A Bi-Directional Information Management Model

As a general framework, IM research often refers to structured layer models (Galliers & Leidner, 2014; Maes et al., 2000). Such an approach was proposed by Wollnik (1988) and adapted in various later publications (Galliers & Leidner, 2014; Krcmar, 2010; Maes et al., 2000; Voß & Gutenschwager, 2001). Nevertheless, these models neither explicitly incorporate bi-directional alignment structures. Figure 1 now incorporates the bi-directional/IT alignment paradigm by extending the original layer model of information management by Wollnik. The three layers interact via a requisitioning and supportive relationship with each other. Requirements are defined as top-down, while support relations are bottom-up (Krcmar, 2010; Voß and Gutenschwager, 2001). In order to integrate the bi-directional business/IT alignment, valuing this task as a key element within the realm of IM, the requirements relationships to a bi-directional interface are extended. In addition, the inputs on information demands and the preceding demand analysis are now extended over all three layers as well.

The proposed modification enables organizations to identify processes and implement customer-driven (Bartens et al., 2014b), especially more generic, bottom-up issued requirements (Chen, 2008). In addition, the model's extension is applicable to other fields and (communication, interaction, and coordination) processes of IM as well as in between IM — particularly applying to the IM tasks definitions formulated by several authors (Galliers & Leidner, 2014; Krcmar, 2010; Voß & Gutenschwager, 2001). In this sense, an altered IM definition, extending the definition by Voß and Gutenschwager (2001), may assume IM as the economic (efficient) planning, acquisition, processing, distribution, and allocation of information as a resource to support and design decision-making, as well as designing the required infrastructure for this purpose. This extended definition now explicitly allows IM to design decision-making, i.e., to align a business model according to bottom-up requirements. Bearing in mind the introductory example of media business, the proposed IM model now provides orientation in the challenges these business models face. It generates fundamental awareness for technological evolution that may require re-designing the business model. In the media streaming case, the model illustrates that, by implementing streaming technology (as an innovation), revenues can no

longer be generated from selling physical goods and may be replaced by subscriptions. Obviously, this model just provides strategic orientation; specific implementations are guided by IT governance (frameworks).

Two Directions of Alignment in IT Governance

While business/IT alignment is discussed in IM research, it is also addressed from the angle of IT governance (De Haes & Van Grembergen, 2015; Van Grembergen & De Haes, 2009). In this realm, the aspects related to IT governance and the frameworks COBIT 5 and ITIL 2011 are examined. As shown by Bartens et al. (2014b) COBIT 5 (ISACA, 2012a) does not reflect a bi-directional business/IT alignment. The main criticism is found due to the strict top-down structure of the key alignment instrument, the COBIT 5 Goals Cascade. Bartens et al. (2014b) postulate that in e-business a merger of the business and IT side of an enterprise as well as the customer side (through the business operations) occurs. Within these merged structures bottom-up requirements evolve on a regular basis. In order to stay competitive, enterprises classified as e-business models, have to accommodate these requirements in a structured and dynamic way as well as in a timely manner. The transition of the authors' findings to businesses outside of the e-business sector can be assumed due to the number of initiatives (hybrid business models) or even the high rate of transformation from pure brick-and-mortar businesses to e-business models (Patel, 2004).

Despite the general focus on the customer, it can be argued that a bi-directional paradigm within ITIL 2011 (Cannon, 2011; Steinberg et al., 2011) is only limited. Main arguments are found in the overall control of change proposals by the service provider (Cannon, 2011) and top-down design coordination (Cannon, 2011) as well as the improvement processes (Lloyd et al., 2011). On the other hand, a complete neglect of bi-directional actions is not observed. Especially

the recently added processes of "strategy management" and "business relationship management" consider customer feedback in a bottom-up approach, though its management is still structured top-down (Cannon, 2011).

Most importantly, note that the aspect of a merger of business and IT operations together with the customer are purposely counteracted by the restrictive structure (Krcmar, 2010) of ITIL processes. No hints on possible mergers can be found within the operations publications (Steinberg et al., 2011), even though merged structures naturally occur where IT influence and dependency is high (Bartens et al., 2014b). Here, the need for a constructive approach as self-claimed by ITIL is potentially the highest overall. A solitary exception is given by workarounds, as seen in "problem management" (Steinberg et al., 2011), which are independently created and implemented by operations. Still, this only refers to temporary solutions, neither a permanent process for requirements satisfaction nor IT innovations management. Depending on the interpretation and depth of application of the proposed model and the underlying bi-directional approach, one may still argue that the basic concepts are at least recognized at a high level.

FUTURE RESEARCH DIRECTIONS

With the bi-directional model of IM the theoretical fundamentals of bi-directional business/IT alignment are given, and future research directions may examine implications for various aspects in IM and IT governance. While IT governance may be concerned with more practical issues of the bi-directional paradigm, IM research may also rethink the role of IT for the competitive position of an enterprise and come up with altered models for IT innovation management. Related research in IT governance will need to provide solutions for bi-directional alignment with the frameworks COBIT and ITIL. Future research may, furthermore, discover required extensions

in IT innovations management models. Also, bi-directional business/IT alignment may be described for different IM models.

CONCLUSION

Recent research has highlighted the need to model customer driven bottom-up business/it alignment in IT governance. This led to the question whether this new understanding of business/IT alignment is reflected in definitions and models of IM. In this article, it has been examined to which extent this new aspect is covered in IM and described possibilities for an integration or extension of IM models. It was found that IM models generally allow a bi-directional business/IT alignment, but hardly specify it in a way that reflects customer needs for technological changes. Consequently, an extension of an established layer model of IM has been provided. Subsequently, a novel specification in IM was introduced and implications for IT innovations management were demonstrated as well as IT consulting approaches were highlighted. In addition, pathways to handle the implications of our findings were elucidated.

REFERENCES

- Bartens, Y., Chunpir, H., Schulte, F., & Voß, S. (2016). *Business/IT Alignment in Two Sided Markets: A COBIT 5 Analysis for Media Streaming Business Models*. Working Paper, Institute of Information Systems, University of Hamburg.
- Bartens, Y., De Haes, S., Lamoën, Y., Schulte, F., & Voss, S. (2015). On the Way to a Minimum Baseline in IT Governance: Using Expert Views for Selective Implementation of COBIT 5. In *Proceedings 48th Hawaii International Conference on System Sciences (HICSS)* (pp. 4554–4563). Kauai, HI: IEEE. doi:10.1109/HICSS.2015.543
- Bartens, Y., Schulte, F., & Voß, S. (2014a). Business/IT Alignment in Two Sided Markets: A Study of COBIT 5 for Internet Based Business Models. *International Journal of IT/Business Alignment and Governance*, 5(2), 27–43. doi:10.4018/ijit-bag.2014070102
- Bartens, Y., Schulte, F., & Voß, S. (2014b). E-Business IT Governance Revisited: An Attempt towards Outlining a Novel Bi-directional Business/IT Alignment in COBIT5. In *Proceedings of the 47th Hawaii International Conference on System Sciences (HICSS)* (pp. 4356–4365). Waikoloa, HI: IEEE. doi:10.1109/HICSS.2014.538
- Boaden, R., & Lockett, G. (1991). Information technology, information systems and information management: Definition and development. *European Journal of Information Systems*, 1(1), 23–32. doi:10.1057/ejis.1991.4
- Cannon, D. (2011). *ITIL service strategy*. London: TSO.
- Chan, Y. E., & Reich, B. H. (2007). IT alignment: What have we learned? *Journal of Information Technology*, 22(4), 297–315. doi:10.1057/palgrave.jit.2000109
- Chen, H.-M. (2008). Towards Service Engineering: Service Orientation and Business-IT Alignment. In *Proceedings of the 41st Hawaii International Conference on System Sciences (HICSS)* (pp. 114–114). Waikoloa, HI: IEEE. doi:10.1109/HICSS.2008.462
- Choo, C. W. (2002). *Information Management for the Intelligent Organization: The Art of Scanning the Environment* (3rd ed.). Medford, NJ: Information Today, Inc.
- De Haes, S., & Van Grembergen, W. (2015). Business/IT Alignment. In *Enterprise Governance of Information Technology* (pp. 45–70). Springer.

- De Haes, S., Van Grembergen, W., & Debreceeny, R. S. (2013). COBIT 5 and Enterprise Governance of Information Technology: Building Blocks and Research Opportunities. *Journal of Information Systems*, 27(1), 307–324. doi:10.2308/isisys-50422
- Duffy, J. (2002). *IT/Business Alignment: Is It an Option Or Is It Mandatory?* IDC document no. 26831. Farmington, MA: IDC.
- Galliers, R. D., & Leidner, D. E. (2014). *Strategic Information Management: Challenges and Strategies in Managing Information Systems*. New York: Routledge.
- Györy, A., Cleven, A., Uebernickel, F., & Brenner, W. (2012). Exploring the Shadows: IT Governance Approaches to User-Driven Innovation. In *Proceedings of the European Conference on Information Systems (ECIS)*. Barcelona: AISel.
- Henderson, J. C., & Venkatraman, H. (1999). Strategic alignment: Leveraging information technology for transforming organizations. *IBM Systems Journal*, 38(2/3), 472–484. doi:10.1147/SJ.1999.5387096
- Henderson, J. C., & Venkatraman, N. (1993). Strategic Alignment: Leveraging Information Technology for Transforming Organizations. *IBM Systems Journal*, 32(1), 4–16. doi:10.1147/sj.382.0472
- ISACA. (Ed.). (2012a). *COBIT 5 Framework*. Rolling Meadows, IL: ISACA.
- ISACA. (Ed.). (2012b). *COBIT 5: Implementation*. Rolling Meadows, IL: ISACA.
- Kearns, G. S., & Sabherwal, R. (2007). Strategic Alignment Between Business and Information Technology: A Knowledge-based View of Behaviors, Outcome, and Consequences. *Journal of Management Information Systems*, 23(3), 129–162. doi:10.2753/MIS0742-1222230306
- Kooper, M. N., Maes, R., & Lindgreen, E. E. O. R. (2011). On the Governance of Information: Introducing a New Concept of Governance to Support the Management of Information. *International Journal of Information Management*, 31(3), 195–200. doi:10.1016/j.ijinfomgt.2010.05.009
- Krcmar, H. (2010). *Informationsmanagement*. Berlin: Springer. doi:10.1007/978-3-642-04286-7
- Lainhart, J. W. IV. (2001). Why IT Governance Is a Top Management Issue. *Journal of Corporate Accounting & Finance*, 11(5), 33–40. doi:10.1002/1097-0053(200007/08)11:5<33::AID-JCAF6>3.0.CO;2-U
- Lloyd, V., Wheeldon, D., Lacy, S., & Hanna, A. (2011). *ITIL continual service improvement*. London: TSO.
- Luftman, J., & Kempaiah, R. (2007). An Update on Business-IT Alignment: “A Line” Has Been Drawn. *MIS Quarterly Executive*, 6(3).
- Maes, R., Rijsenbrij, D., Truijens, O., & Goedvolk, H. (2000). *Redefining business: IT alignment through a unified framework*. Working Paper, Department of Accountancy and Information Management, Universiteit van Amsterdam.
- McFadden, C. (2012). Are textbooks dead? Making sense of the digital transition. *Publishing Research Quarterly*, 28(2), 93–99. doi:10.1007/s12109-012-9266-3
- Patel, K., & McCarthy, M. P. (2000). *Digital transformation: The essentials of e-business leadership*. McGraw-Hill Professional.
- Patel, N. V. (2002). Emergent Forms of IT Governance to Support Global E-Business Models. *Journal of Information Technology Theory and Application*, 4(2), 33–48.

Patel, N. V. (2004). An emerging strategy for e-business IT Governance. In W. Van Grembergen (Ed.), *Strategies for information technology governance*. Hershey, PA: IGI. doi:10.4018/978-1-59140-140-7.ch003

Pérez Lorences, P., & García Ávila, L. F. (2013). The Evaluation and Improvement of IT Governance. *Journal of Information Systems and Technology Management*, 10(2), 219–234. doi:10.4301/S1807-17752013000200002

Sabherwal, R., Hirschheim, R., & Goles, T. (2001). The dynamics of alignment: Insights from a punctuated equilibrium model. *Organization Science*, 12(2), 179–197. doi:10.1287/orsc.12.2.179.10113

Singh, S. N., & Woo, C. (2009). Investigating business-IT alignment through multi-disciplinary goal concepts. *Requirements Engineering*, 14(3), 177–207. doi:10.1007/s00766-009-0081-0

Steinberg, R. A., Rudd, C., Lacy, S., & Hanna, A. (2011). ITIL service operation. London: TSO.

Van Grembergen, W., & De Haes, S. (2009). *Enterprise Governance of Information Technology: Achieving Strategic Alignment and Value*. New York, NY: Springer.

Voß, S., & Gutenschwager, K. (2001). *Informationsmanagement*. Berlin: Springer. doi:10.1007/978-3-642-56878-7

Wollnik, M. (1988). Ein Referenzmodell des Informationsmanagement. *Information & Management*, 3(3), 34–43.

Hiekkanen, K., Helenius, M., Korhonen, J. J., & Patricio, E. (2013). Aligning alignment with strategic context: A literature review. In *Digital Enterprise Design and Management 2013* (pp. 81–98). Berlin: Springer. doi:10.1007/978-3-642-37317-6_8

Luftman, J., & Brier, T. (1999). Achieving and sustaining business-IT alignment. *California Management Review*, 42(1), 109–122. doi:10.2307/41166021

Luftman, J., Papp, R., & Brier, T. (1999). Enablers and inhibitors of business-IT alignment. *Communications of the AIS*, 1(3), 1–33.

Papp, R. (1999). Business-IT alignment: Productivity paradox payoff? *Industrial Management & Data Systems*, 99(8), 367–373. doi:10.1108/02635579910301810

Schlosser, F., Wagner, H. T., & Coltman, T. (2012). Reconsidering the dimensions of business-IT alignment. In *45th Hawaii International Conference on System Science (HICSS) 2012* (pp. 5053–5061). IEEE. doi:10.1109/HICSS.2012.497

Solli-Sæther, H., & Gottschalk, P. (2010). The modeling process for stage models. *Journal of Organizational Computing and Electronic Commerce*, 20(3), 279–293. doi:10.1080/10919392.2010.494535

Ullah, A., & Lai, R. (2013). A systematic review of business and information technology alignment. *ACM Transactions on Management Information Systems (TMIS)*, 4(1), 4:1–4:29.

ADDITIONAL READING

Benlian, A., & Haffke, I. (in press). Does mutuality matter? Examining the bilateral nature and effects of CEO–CIO mutual understanding. *The Journal of Strategic Information Systems*.

KEY TERMS AND DEFINITIONS

Bi-Directional Business IT Alignment:

The process and goal of achieving competitive advantage through developing and sustaining a symbiotic relationship between business and IT, while emphasizing an often omitted bi-directional alignment between business and IT.

Business /IT Alignment: The process and goal of achieving competitive advantage through developing and sustaining a symbiotic relationship between business and IT.

COBIT: An IT governance framework and toolset that allows managers to bridge the gap between control requirements, technical issues and business risks (ISACA).

Information Management: The economic (efficient) planning, acquisition, processing, distribution and allocation of information as a resource for the preparation and support of deci-

sions (decision-making), as well as designing the required infrastructure for this purpose.

IT Governance: Framework function for the administration, organization, and control of IT in enterprises and public organizations.

ITIL: Framework that aligns IT services to the needs of the business and support its core processes.

Layer Model of Information Management: A structuring model that defines layers of information management, e.g., with respect to the level of technology used.

Business Sustainability Indices

B

Arunasalam Sambhanthan
Curtin University, Australia

INTRODUCTION

Sustainability indices are integral part of measuring the process level variables in an organism (i.e. county, region, city, industry, organization, business unit). The documented literature has less evidence for targeted indices developed for measuring specific industry clusters / organizational units. In fact, there are indices which are developed by researchers to measure the sustainability in a country or regional level (Böhringer & Jochem, 2007). These indices are meant to assess the sustainability at policy level applicable for countries and organizations in general. However, there is a considerable research gap in terms of indices which are developed for measuring the ICT Business Sustainability. This paper aims at reviewing the existing sustainability indices in the body of literature.

BACKGROUND

Business Sustainability research has been tremendously progressed. However, it is still a question as to how well the concept of business sustainability has been theorized. Theorizing business sustainability has a number of challenges to be tackled in terms of which aspects are critical for sustainability and which are not. Therefore, it is essential to have a theoretical model as the basis for theorizing business sustainability elements. This paper reviews the two types of theoretical models available in the business sustainability research namely the sustainability hierarchy and business sustainability typology. The relevance of the concept of sustainability to business domain is paramount.

DOI: 10.4018/978-1-5225-2255-3.ch053

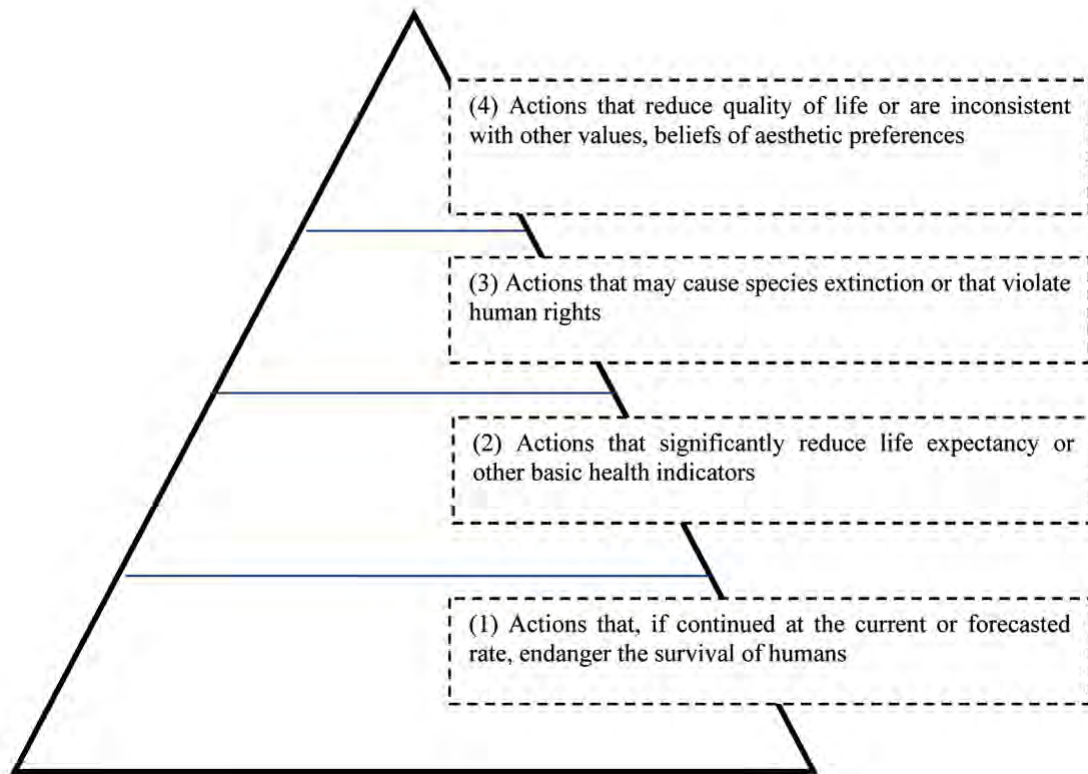
THE SUSTAINABILITY HIERARCHY

Four sustainability frameworks have been identified as ‘ambiguous with regard to what is being sustained’. The label ‘unsustainable’ has been categorized into four levels of actions which prevent sustainability of any kind. This led towards the development of a Sustainability hierarchy (Marshall & Toffel, 2005) which was constructed using several distinct but related concepts (refer to Figure 1).

The above hierarchy is an action centered hierarchy which looks at the sustainability from a hierarchical model as depicted above. The model is based on the hierarchy of needs theory propounded by Abraham Maslow. The needs theory indicates that the satisfaction of lower level needs leads towards the higher level need in the hierarchy. The same metaphor is applied here for environmental sustainability in this context. The achievement of a lower level environmental metric leads to the achievement of the higher level one in the hierarchy.

The main shortcoming of this model in terms of applying to business sustainability is its lack of specialization to business domain. The model is named as sustainability hierarchy which could be applied to anything which has the ability to be applied to the sustainability research. Therefore, the above model is argued to be not contextually relevant for assessing and benchmarking sustainability in organizations. A definition of business sustainability goes as “adapting business strategies and activities that meet the needs of the enterprise and its stakeholders today while protecting, sustaining and enhancing the human and natural resources that will be needed in the future” (International Institute of Business

Figure 1. Business sustainability hierarchy



Sustainability). Therefore, the following section focuses on reviewing the business sustainability typologies out there in the literature.

THE SYSTEM BASED SUSTAINABILITY BUSINESS MODELS

Models to assess business sustainability have been less documented in the literature. Besides industry specific frameworks there are a number of common frameworks out there in the literature such as the European Corporate Sustainability Framework (ECSF) (Costanza & Patten, 1995), there have been less documented models out there in the published literature to assess and benchmark business sustainability. The industry specific frameworks and the ECSF all could be classified as frameworks than of models. The System Based

Sustainability Business Model documented by (Dyllick & Muff, 2013) has four characteristics such as economic characteristics, environmental characteristics, social characteristics and multidimensional or social characteristics – each of which is attributed into two attributes namely structural attributes and cultural attributes.

There are a number of structural attributes that have been found within the economic characteristics. For example, bodies external to the organisation that keep track of performance of the companies that uses a triple bottom line approach and the government that keep track of changes to taxation system and legislation to support sustainability and local shareholders and investment in local sustainability initiatives. The environmental characteristics consists of structural attributes such as a threefold strategy which offsets (do no harm but make amends if you do), sustainable (do no harm), restorative (leave the world better than you

found it), closed-loop systems which is responsible for product throughout its lifecycle implementing a services model and the industrial ecosystems and stakeholder networks. The social characteristics includes attributes such as stakeholder engagement skills - understanding stakeholders' needs and expectations (i.e. being relevant to stakeholders), educating stakeholders - "relentless" communication implementing stakeholder consultation program and getting "buy-in" from internal and external stakeholders. The multidimensional characteristics has been structurally attributed with cooperative business strategy and planning, collaborative model including supply chain, competitors, government agencies, communities, TBL approach to measure organizational performance, institutionalize sustainability in the business: "relentless" communication, stakeholder education, leadership, champions, and align internal performance measures and demand-driven model, not supply-driven model (driven by what people need, not driven by companies trying to get people to buy more).

The cultural attributes associated with economic characteristics includes attributes such as profit is a means not an end. Business makes a profit to do something more. "Higher purpose" to business than making money, shareholders invests for social & environmental impact reasons as well as for financial reasons and shareholders temper expectations for short term financial returns. The environmental characteristics have only one cultural attribute which looks at treating nature as a stakeholder. The social characteristics attributed for culture are stakeholder approach (managing the organization for the benefit of all stakeholders and not prioritizing shareholders' expectations above other stakeholders), alignment of stakeholder expectations, sharing of resources (people, profits, and time) among stakeholders to achieve sustainable outcomes and the relationship building (trust, two-way loyalty, honesty, integrity, and fairness, equity). Finally the multidimensional characteristics have been attributed for culture

with a number of attributes such as medium to long-term focus and reduction in consumption.

This business model is quite comprehensive to evaluate business sustainability, but essentially lacks to incorporate the fourth bottom line - institution into it. In fact, the fourth dimension has been identified to be playing a critical role in determining the sustainability of organisations at the current milieu with WEDC proposing it as an additional bottom line for sustainability evaluation. In addition to this, the social dimension of the above model is identified to be overly emphasising on stakeholder engagement, but lacks to incorporate the human resources issues in social sustainability of organisations - a critical aspect. For an example, the employability and other human resource issues arising in organisations could be included into this dimension to make the social sustainability dimension more relevant to the business sustainability research agenda.

BUSINESS SUSTAINABILITY TYPOLOGY

There is not much documented research in terms of defining the typologies of business sustainability. An early research by (Costanza & Patten, 1995) pose few questions regarding defining and predicting sustainability. The paper deals with the sustainability of systems. The critics questions about the usefulness of the concept of sustainability due to the ambiguities involved in properly defining the term. Hence, the discussions of this regard are argued to be misdirected due to the following reasons:

1. The discussions were arguing the problem as definitional instead of predictive which deals with the prediction of what will last and survive in achieving the consensus on what we want to last
2. The discussions failed to account for the range of interrelated time and space scales over which the concept must apply

Hence, the authors argues that the basic idea of sustainability is very simple and straightforward as a sustainable systems is one which services or persists over the time. However, there are three main concerns to be addressed in this regard. Firstly, it should be clarified as to what system or subsystems or characteristics of the system persist? For how long it is going to persists? And when it is going to be assessed as to whether it has persisted? In overall, the paper is trying to address the above questions.

- **When?:** Although the paper discusses about the biological point of view of sustainability, it is avoided in this discussion due to the fact that this discussion is targeted on the business aspect of sustainability. In economic point of view, sustainability is all about avoiding major disruptions and collapses, hedging against instabilities and discontinuities. In this context, it could be argued that the sustainability of businesses could not be defined in present tense, as it is very hard to predict as to how long a business will last. In fact, the sustainability of a business could only be termed after the period of existence it has. Therefore, it is quite contradictory to argue whether sustainability is a concept which needs to be defined in present tense or past tense. Whereas, predicting sustainability is technically possible, the authors argues that defining sustainability is impossible due to the time and space constraints involved in defining the concept of sustainability. The actions taken today in order to prevent the system will contribute towards the sustainability in the future. Whilst, making an effort to define a future stage could be considered as more of an attempt to predict, it could also be argued whether it is feasible to predict something solely based on the present stage of the particular system or organism. Thus, the authors argue that a possible definition of sustainability could

be a set of predictions of actions taken today that will lead to sustainability in long run. This entails the consideration of system uncertainties as a bottleneck in defining business sustainability.

- **What?:** What system or subsystem one has to sustain in long run is the second question? This dimension indicates the preferred characteristics as an element of sustainability. Thus the three sub elements of the characteristics highlighted by the authors include scale of economy, distribution of resources and allocation of resources. The above characteristics are argued to be the predictors of the sustainability of a system or sub-system.
- **How Long?:** This involves with the time span of the sustainability of system concerned. Some arguments foster the everlastingness as systems' sustainability. However, infinity cannot be a timespan to be specified for sustainability. Therefore, it is a vital concern to consider the time length when speaking about the sustainability of any system.

In contrast, a research documented by (Dyllick & Muff, 2013) looks at three compartments such as concerns, organizational perspective and value created. The business sustainability typology outlined by (Dyllick & Muff, 2013) is depicted in table 1.

The above typology shows three different versions all of which is described with the aforementioned three compartments (Muff & Dyllick, 2014). The first compartment covers concerns such as consuming the income than of capital and being accountable for the impact of business decisions and activities. The second component looks at concerns such as managing risks and opportunities, increasing market share, revenue and reputation. The third dimension looks at integration of economic, ecological and social value creation.

Table 1. Business sustainability typology

Typology of Business Sustainability	Concerns (What?)	Organizational Perspective (How?)	Values Created (What for?)
Business-as-usual	Economic concerns	Inside-out	Shareholder Value
Business sustainability 1.0	Three-dimensional concerns	Inside-out	Refined shareholder value
Business Sustainability 2.0	Three-dimensional concerns	Inside-out	Triple bottom line
Business Sustainability 3.0	Starting with sustainability challenges	Outside-in	Creating value for the common good

In addition to this, the authors are outlining three versions of business sustainability which has been presented as the business sustainability typology theorization basis by (Muff & Dyllick, 2014). The basis of versioning this has been justified based on the Input => Process => Output logic. In fact, the version 1 states business sustainability as an approach and the version 2 define it as a process while the version 3 claim it to be the impact / effect on the environment. Therefore the business sustainability typology outlined by (Dyllick & Muff, 2013) could be argued as the first step in business sustainability typology research with more opportunities for applied research in this domain with the typology proposed by (Muff & Dyllick, 2014).

SUSTAINABLE PRODUCTION

Sustainable production is an area of paramount concern by businesses which involves any form of production. In fact, (Veleva & Ellenbecker, 2000) argues there is no comprehensive framework for assessing business sustainability. Regardless of the above absence in the literature, there are a number of business sustainability frameworks, developed and documented by researchers in the recent literature which are constructed using a number of sustainability indicators (Høgevold & Svensson, 2012); (Pojasek, 2007); (Hardjono & De Klein, 2004). The indicators for sustainable production as outlined by (Veleva & Ellenbecker, 2000) includes energy and martial use (resources), natural environment, Economic viability, com-

munity development and social justice, employees and products. The above indicators aim to inform business decision makers and co-operate policy makers with the principles of sustainable production that will eventually lead towards compliance to global standards on production sustainability. In addition to this, (Quinn & Dalton, 2009) outlines a number of principles for enabling companies to have sustainable production which includes elements such as products and packaging, energy and materials, workplace practices, economic viability and community development. However, it is questionable whether the indicators covered by the aforementioned authors cover a whole range of industries which involves production of various goods and services. For example, some of the industries which uses technological resources for production such as software and e-learning would need to have indicators for sustainable production which will include e-waste and code reuse. On the contrary, there could be a number of practical issues related to the intellectual property and non-disclosure agreements between client-business partnership that could prevent the code reuse and related sustainability measures at instances. The following section evaluates the qualities of sustainable production indicators in detail.

INDICATORS FOR SUSTAINABLE PRODUCTION

Integrating the sustainability concept into the production process is another core challenge in meeting the sustainability standards. This entails

the companies to reduce the resource intensity used in production process which would eventually lead towards sustainable production. Whilst evaluating the ways and means of integrating sustainability thinking to the production process, this section reviews the existing literature with the aim of outlining the desirable qualities of the indicators for sustainable production. Table 2 depicts the major qualities of the indicators of sustainable production as outlined by (Veleva & Ellenbecker, 2000).

The above qualities are of importance in valuing the indicators of sustainable production to an extent. Whilst the above indicators emerged from the early research on sustainability, there was a vacuum for concrete indicators for the measurement of sustainability reported by researchers (Veleva & Ellenbecker, 2000); (Böhringer & Jochem, 2007). On the other hand, it is questionable whether the aforementioned qualities are matching with the indicators denominating the sustainability of technology based industries (i.e. software, e-learning). In fact, most of the qualities outlined in the above table are of common for all

industries. In contrast, (Veleva & Ellenbecker, 2000) build an argument that it is a reality to have a standard set of indicators applicable to any company. The authors justify their argument stating that it is advantageous to measure the right things approximately than measuring the wrong ones with great accuracy and precision. For example, sustainability could be achieved through effective process redesign and production waste management (i.e. peripheral waste) in the context of software industry, whereas some other aspects of sustainability might get relatively low priority. Moreover, the companies could have innovative approaches for resource utilization in order to achieve sustainability in a product or service based technology led company. Whereas, indicators are considered to be in the second stage of an abstraction process (data -> Indicators / indices -> Frameworks - > Definitions), frameworks are considered to be the more concrete form of a sustainability theory next to definitions. Thus, the following section will examine the business sustainability frameworks available in the existing body of knowledge.

Table 2. Qualities of sustainable production indicators after

Indicator	Description
Appropriateness	The indicator's appropriateness to the task. This focuses on precisely and specifically expressing the unit of measure in the task.
Standardization	This looks at the ability to have a standard criteria using which a possible comparison could be made between two companies.
Data Availability and Accuracy	This enables the businesses to start with a single indicator of existing easily collectable data.
Verifiability	The credibility of reported information could be diluted if the data processed is not objective enough to make conclusions.
Set of Indicators	Set of indicators will allow the businesses to have a comprehensive coverage of sustainable production.
Inform Decision Making	Enabling the management regarding the concerns and priorities of the employees.
Manageable Number	The number of indicators employed needs to be manageable in order to eliminate the errors occurs due to the manageability.
Simplicity and Meaningfulness	The indicator needs to be simple as well as meaningful in terms of information richness.
Measurability	The indicators need to be quantifiable in terms of qualitative as well as quantitative aspects.
Consistency	Consistency between national sustainable production initiatives and sustainable production initiatives.
Community Involvement	Employees, community and interested stakeholder involvement is important for sustainability initiatives.

Veleva & Ellenbecker, 2000

A REVIEW OF SUSTAINABILITY MEASUREMENT INDICES

Whilst making the production and consumption of goods make a paramount contribution to the business sustainability literature, it is commonplace that the measurement indices for business sustainability are a lacking area in the existing literature. There have been a number of indices developed by researchers to facilitate the measurement of sustainability in differing contexts (see Table 3). For example, there are eleven sustainability indices documented by (Böhringer & Jochem, 2007) which looks at measuring sustainability at country level.

Whilst aiming to measure the sustainability of countries in different terms, the aforementioned eleven indices make meaningful contribution to the sustainability literature. For instance, the Living planet index has a direct contribution to the measurement of sustainability of species living across this planet. Another example would be the Environmental Performance Index that aims at measuring the environmental stresses which could affect the human health condition and sustainability. However, one of the main weaknesses of all these indices relies on its scope and measurement object. In fact, these indices are all focused on measuring the human ecosystem and the natural environment of a country, which does not have any direct contextual relevance to the measurement of business sustainability. On the contrary, a number of measurement criteria set forth for measuring the sustainability of a country could be applied for modelling the same phenomenon in an organization. For example, The Environmental Performance Index (EPI) has an element of measuring human health condition of a country as measurement criteria to assess the environmental sustainability. This could effectively be applied to an organization as well, since one of the measurement criteria for measuring business sustainability is health and safety of the employees and workers. Another drawback identified among these indices is the

lack of a cross-country comparability of indices. For instance, the Well-Being Index calculated for one country could not be compared with the Well-Being Index of another country. Hence, all these indices are useful in measuring the sustainability at country level, but not necessarily be suitable to measure business sustainability. Therefore, the following section focuses on a review of business sustainability indices out there in the literature.

MEASURING BUSINESS SUSTAINABILITY

Measuring business sustainability has been a long question in front of researchers. Measuring the triple bottom lines in terms of sustainability is paramount for any organization to survive. There are indices created by researchers at policy level which could facilitate the measurement of a country's sustainability as reviewed in the previous section. In contrast, there should be certain measurement scales for benchmarking sustainability at business level. One example would be the Dow Johns Sustainability Group Index which covers a range of sustainability indices. The index was created by S&P Dow John Indices which consists of a number of regional indices such as DJSGI World Index, DGSGI Europe Index, DGSGI North America Index, DGSGI Asia Pacific Index, and DJSGI USA Index. The index is recorded to be covering more than 2000 largest capitalized companies, 64 industries and 34 countries with over 231 components. This index is also documented for having a number of specialized indexes such as DJSGI excluding Alcohol indexes, DJSGI excluding Gambling indexes, DJSGI excluding Tobacco indexes, DJSGI excluding all indexes (i.e. excluding alcohol, gambling and tobacco). This index covers a number of issues such as Illegal commercial practices e.g. tax fraud, money laundering, antitrust, balance sheet fraud, and corruption cases, workforce conflicts e.g. extensive layoffs and strikes, large disasters or accidents e.g. fatalities, accidents, workplace safety, technical

Table 3. Sustainability indices after

Index / Sub-Indices	Description
Living Planet Index (LPI)	“LPI measures trends in over 2000 populations of more than 1100 species of vertebrates in terrestrial, freshwater, and seawater ecosystems” (Böhringer & Jochem, 2007)
Ecological Footprint (EF)	“The ecological foot print is based on the quantitative land and water requirements to sustain a (national) living standard into infinity thereby assuming certain efficiency improvements” (Böhringer & Jochem, 2007)
City Development Index (CDI) 1. Infrastructure Index 2. Waste Index 3. Health Index 4. Education Index 5. City Product Index	1. “Infrastructure index builds on four (equally weighted) indicators as percentages of households which are connected to clean water, canalization, electricity and phone network (without mobiles)” (Böhringer & Jochem, 2007) 2. “A twofold (equally weighted) waste index, which is composed of the percentage of untreated sewage in total wastewater and the percentage of disposal of solid wastes” (Böhringer & Jochem, 2007) 3. “A twofold (diversely weighted) health index, which considers the life expectance and the infant mortality rate” (Böhringer & Jochem, 2007) 4. “A twofold (equally weighted) education index which is calculated by adding the percentages of literacy and combined enrolment” 5. “A city product index, which is based on the logarithmic value of a city’s GDP” (Böhringer & Jochem, 2007).
Human Development Index (HDI)	“The Human development Index consists of three (equal weighted) sub-indices which are aggregated by an arithmetic mean: Life Expectance Index, Education Index (decomposed into and Adult Literacy Index and a Gross Enrolment Ration Index), and a GNP Index” (Böhringer & Jochem, 2007).
Environmental Sustainability Index (ESI)	“The Environmental Sustainability Index score quantifies the likelihood that a country will be able to preserve valuable environmental resources effectively over the period of several decades” (Böhringer & Jochem, 2007).
Environmental Performance Index (EPI)	“The Environmental Performance Index addresses the need for a gauge of policy performance in reducing environmental stresses on human health and promoting ecosystem vitality and sound natural resource management” (Böhringer & Jochem, 2007).
Environmental Vulnerability Index (EVI)	“The Environmental Vulnerability Index Consists of 32 indicators of hazards, 8 indicators of resistance and 10 indicators that measure damage” (Böhringer & Jochem, 2007).
Index of Sustainable Economic Welfare / Genuine Progress Index (ISEW/GPI)	“The starting point of the Index of Sustainable Economic Welfare is the inflation-adjusted consumption of households. The time series consumption values is adjusted by five categories to obtain a ‘GDP’ which is more appropriate for measuring social welfare: (1) distribution of income, (2) economic activities not counted in the conventional gross national income, (3) time adjustments, (4) damage caused by economic activity, and (5) the consideration of net capital endowment of foreign investors” (Böhringer & Jochem, 2007).
Well-Being Index (WI)	“The Well-Being Index is the arithmetic mean of a human Well-being index (HWI) and an Ecosystem Well-Being Index (EWI). The indices HWI and EWI in turn consist of five sub-indices. The HWI comprises a health and population, welfare, knowledge, culture and society as well as for resource deployment. The five dimensions of the HWI are based on 36 indicators, those of EWI on 51 indicators” (Böhringer & Jochem, 2007).
Genuine Savings Index (GSI)	“The Genuine Savings (GS) are an indicator of weak SD. The societal capital stock consists of produced capital, human capital (knowledge, skills etc.) as well as natural capital (resources etc.)” (Böhringer & Jochem, 2007).
Environmental Adjusted Domestic Product (EDP)	“The Green Net National Product or likewise the Environmentally Adjusted Net Domestic Product (EDP) has been developed within the scope of SEEA (System of Integrated Environmental and Economic Accounting). There are three different versions of the EDP can be distinguished: (1) the EDPI which subtracts depreciations of natural resources caused by their extraction from the Net National Income (NNI), (2) the EDPII, which subtracts from the NNI the costs necessary to reach the same state of the environment at the end of the period as existed at the beginning of the period, and (3) the EDPIII, which subtracts the costs of environmental pressure and destruction (calculated by willingness-to-pay methods)” (Böhringer & Jochem, 2007).

Source: Böhringer & Jochem, 2007

Table 4. Focusing on the automobile industry

Bottom-line	Indicator	Metric / Method	Driver for Inclusion
Environmental & Health	Life Cycle Global Warming Potential	Greenhouse emissions along the life cycle (vehicle production, driving 15555 KM based on EUCAR agreement, end of life) - Part of an LCA according to ISO14040	Carbon intensity is the main strategic issue in the automotive industry
	Life Cycle Air Quality Potential	Summer smog creation potential (POCP) along the same life cycle (VOCs, NO _x) – part of an LCA according to ISO 14040	Non CO ₂ air quality issues have to be monitored for trade offs
	Sustainable Materials	Recycled & natural materials related to all polymers	Resource scarcity
	Restricted substances	Vehicle interior air quality / allergy-tested interior, management of substances along the supply chain (15 point rating)	Substance risk management is key
	Drive-by-exterior noise	Drive-by-exterior noise = dB (a)	Main societal concern
Social	Safety	Including number of all EuroNCAP stars (see note under acronyms regarding EuroNCAP ¹)	Main direct impact
	Mobility capability	Mobility capability (luggage compartment volume plus weighted number of seats) related to vehicle size	Crowded cities (future: diversity – disabled)
Economics	Life Cycle ownership costs	Vehicle price plus 3 year's fuel costs, maintenance costs, taxation minus residual value (note: for simplification reasons costs have been tracked for one selected market; life cycle costing approach using discounting)	Customer focus, competitiveness

failures, ecological disasters and product recall. It is a question whether all these elements are really necessary to measure business sustainability. In fact, many of the issues listed in the above two categories are overly focused on measuring the sustainability of countries and cities than businesses. Whilst listing the above categories as the key factors for evaluation, the DJS group index lacks to include the institutional dimension in its evaluation criterion, which could be noted as one of the main shortcomings of this index. In addition to this the DSJ sustainability index has a regional classification of indices, but lacks to have an industry based classification, although the industry based classifications are used as a way for assessing the results to an extent. However, this is not targeted for assessing business sustain-

ability and there are not so many indexes other than this to measure business sustainability in the documented literature.

PRODUCT SUSTAINABILITY

There is not much documented research with regard to Product Sustainability Indices except the Ford of Europe's Product Sustainability Index (Schmidt, Wulf-Peter & Taylor, 2006), which supports measuring a vehicles' product sustainability throughout the entire product lifecycle along all three bottom lines. Table 4 depicts the details of product sustainability index measurement criteria proposed by (Schmidt, Wulf-Peter & Taylor, 2006) for measuring the sustainability of vehicles.

FUTURE RESEARCH DIRECTIONS

A number of future research directions are highlighted for subsequent treatment.

1. An analysis of industry specific sustainability indices.
2. An analysis of sustainability indicators in practice for different industries.
3. An analysis of sustainable indices in practice for different industry segments.

CONCLUSION

Theorisation of business sustainability has been dealt in this article. We have reviewed two types of business sustainability models in this article namely action centred models and construct centred models. Action centred model reviewed include the sustainability hierarchy which uses an action centred hierarchy to model sustainability. Construct centred models include system based sustainability business model and the business sustainability typology. The review indicates that there is not much documented research exist in business sustainability typologies. There is a considerable amount of research effort required from the applied researchers working in this area to contextualise and re-specify the business sustainability typologies to differing industry clusters such as e-learning, agriculture, manufacturing and healthcare. Furthermore, this paper has reviewed the sustainability measurement indices out there in the published domain. Primarily a number of researches have been reported on the sustainable production and the sustainable production indicators. Appropriateness, standardization, data availability and accuracy, verifiability, indicators, inform decision making, manageable number, simplicity and meaningfulness, measurability, consistency and community involvement are the sustainable production indicators reported in the past research. In addition to this there are a number of sustainability measurement indices such as Living Planet Index, Ecological Footprint, City

Development Index, Human development Index, Environmental Sustainability Index, Environmental Performance Index, Environmental Vulnerability Index, Index of Sustainable Economic Welfare / Genuine Progress Index, Wellbeing Index, Genuine Savings Index, Environmental Adjusted Domestic Product. A number of shortcomings which are not addressed by these indices has been discussed in the article and followed by a discussion about the Dow Johns Sustainability Group Index - which is the only index available in the literature to measure business sustainability. This paper serves as an introductory review to the sustainability indices and the applicability of those indices to the context of businesses.

REFERENCES

- Böhringer, C., & Jochem, P. E. P. (2007). Measuring the immeasurable — A survey of sustainability indices. *Ecological Economics*, 63(1), 1–8. doi:10.1016/j.ecolecon.2007.03.008
- Costanza, R., & Patten, B. C. (1995). Defining and predicting sustainability. *Ecological Economics*, 15(3), 193–196. doi:10.1016/0921-8009(95)00048-8
- Dyllick, T., & Muff, K. (2013). Clarifying the meaning of business sustainability. *Organization and Environment*, 1–26.
- Hardjono, T., & De Klein, P. (2004). Introduction on the European Corporate Sustainability Framework (ECSF). *Journal of Business Ethics*, 55(2), 99–113. doi:10.1007/s10551-004-1894-x
- Høgevold, N. M., & Svensson, G. (2012). A business sustainability model: A European case study. *Journal of Business and Industrial Marketing*, 27(2), 142–151. doi:10.1108/08858621211197001
- Marshall, J. D., & Toffel, M. W. (2005). Framing the elusive concept of sustainability: A sustainability hierarchy. *Environmental Science & Technology*, 39(3), 673–682. doi:10.1021/es040394k PMID:15757326

Muff, K., & Dyllick, T. (2014). An organizational roadmap of Business Sustainability. In *Preparation*, (March), 1–25.

Pojasek, R. B. (2007). A framework for business sustainability. *Environmental Quality Management*, 17(2), 81–88. doi:10.1002/tqem.20168

Quinn, L., & Dalton, M. (2009). Leading for sustainability: Implementing the tasks of leadership. *Corporate Governance*, 9(1), 21–38. doi:10.1108/14720700910936038

Schmidt, W-P., & Taylor, A. (2006). *Ford of Europe's Product Sustainability Index*. Retrieved May 4, 2015, from <http://www.mech.kuleuven.be/lce2006/key2.pdf>

Veleva, V., & Ellenbecker, M. (2000). A Proposal for Measuring Business Sustainability: Addressing Shortcomings in Existing Frameworks. *Greener Management International*.

ADDITIONAL READING

Dyllick, T., & Muff, K. (2013). Clarifying the meaning of business sustainability. *Organization & Environment*, 1–26.

Hardjono, T., & De Klein, P. (2004). Introduction on the European Corporate Sustainability Framework (ECSF). *Journal of Business Ethics*, 55(2), 99–113. doi:10.1007/s10551-004-1894-x

Høgevold, N. M., & Svensson, G. (2012). A business sustainability model: A European case study. *Journal of Business and Industrial Marketing*, 27(2), 142–151. doi:10.1108/08858621211197001

KEY TERMS AND DEFINITIONS

Indicators: The pointers of the sustainability performance of certain industries.

Indices: The measurement criterion used for assessing the sustainability of different industries.

Sustainability: The ability to sustain a certain period of time.

ENDNOTE

- ¹ European New Car Assessment Program <http://www.euroncap.com/en> EuroNCAP provide one indicator of a Car's Performance. To get a comprehensive indicator, Ford's approach to car's safety additionally embraces other impact modes as well as accident avoidance measures in Multi-Panel chart.

The Business Transformation Framework, Agile Project and Change Management

Antoine Trad

Institute of Business and Information Systems Transformation Management, Switzerland

Damir Kalpić

University of Zagreb, Croatia

INTRODUCTION

In order to restructure and make the global economy agile in a sustainable manner, the integration of business engineering related Agile Project and Change Management (for simplification in further text the term *Project-Management* will be used) mechanisms are fundamental. *Project-Management* transforms traditional business environments to become a part of the newly interconnected global economy (Trad & Kalpić, 2014b). An important factor in continuously transforming a business environment into an innovative and lean business engineering services-oriented environment are the roles of the Business Transformation Manager (for simplification in further text the term *Manager* will be used) and the needed business transformation framework that has *Project-Management* capabilities. These capabilities are also needed to support dynamic project management activities, in order to facilitate project management's integration, manage critical success factors selection and risk assessment management. *Project-Management* must be used to schedule, manage, assert, govern, automate, trace, monitor and control the Business Transformation Project (for simplification in further text the term *Project* will be used) artefacts (Trad & Kalpić, 2015a; Trad & Kalpić, 2015b; Trad, 2015b). The *Manager's* role is of crucial importance for the implementation phase of complex *Projects*; where his or her (for simplicity reasons the authors will use his/he in further text) project management decisions can be made

in a just-in-time manner using outputs from the business environment's existing events, monitoring, audits and logging systems. Unfortunately, an agile holistic approach for project management, change management and enterprise architecture is very complex to implement (Lee & Yong, 2010).

The *Manager* must have the needed *Project-Management* skills to handle the complex and chaotic technical implementation phase of *Projects* that are the major cause of very high failure rates (CapGemini, 2009). Therefore the implementations of business engineering resources in *Projects* require specific advanced *Project-Management* techniques and that are defined in the initial architecture strategy. A *Project-Management* curriculum must contain project management, combine management sciences, risk management, law assertion, enterprise architecture, team management, technology management and business engineering sections (Trad & Kalpić, 2016). Such a curriculum or a certificate of advanced studies program should contain project management, engineering, risk assessment and technology and enterprise architecture topics. For this specific goal the Institute of Business and Information Systems' Transformation Management (IBISTM) has developed: a) a real world framework; b) an education curriculum; c) a training syllabi; and d) a set of certification programmes, to support *Managers* and hence *Projects* (IBISTM, 2016a).

This research proposes the *Project-Management* component that is a part of the Project management (Pm) module and this module is in turn a

part of the Selection management, Architecture-modelling, Control-monitoring, Decision-making, Training management and Project management Framework (SmAmCmDmTmPmF, for simplification in further text the term *Environment* will be used), that supports various aspects of *Project's* activities. In this article the authors present a set of *Project-Management* various types of recommendations and a reusable architecture pattern (ISO, 2000; ISO, 2007; Trad 2015c; Trad 2015d). The *Environment's Project-Management* must be synchronized through the architecture development method's phases, as shown in Figure 1, where each *Project* implementation building block circulates through the Architecture Development Method's (ADM) phases. The building blocks contain sets of Critical Success Factors (CSF) (Sugumaran & Lavanya, 2014) that can be applied for: a) the selection of the *Managers*; b) the implementation of *Project's* architecture and modelling strategies (Trad, & Kalpić, 2014c); c) the decision support system, in order to estimate the actual status of the *Project* and to decide whether to stop or continue the on-going project (Gartner, 2013); d) the *Project-Management*; and e) the training and educational needs of the *Project's* team.

BACKGROUND

The global research topic's and final research question (hypothesis #1-1) is: "Which business transformation manager characteristics and which type of support should be assured in the implementation phase of a business transformation project?" The targeted business domain is any business environment that: a) uses e-communication and business engineering technologies; and b) has frequent transformation iterations. For this phase of research the sub-question (hypothesis #6-2, as it will be seen from the further text) is: "What is the impact of agile project and change management on enterprise architecture and business transformation projects?"

In this research phase the authors are targeting the *Project-Management* component's integration that is a part of the Project management module (Pm) and its influence on *Projects* (Trad, 2015d). Actually we are experiencing the start of automation of project management and its integration enterprise architecture environments, but this integration phase is still immature. This is a barrier for a successful *Project's* management and has to be handled. In this article the authors use credible sources, proof of concept, industrial experiences gained and lessons learned from surveying many enterprises; and the first step is to define the ADM's, *Project's* and *Project-Management's* unit(s) of work.

MAIN FOCUS OF THE ARTICLE

Enterprise Architecture's Role

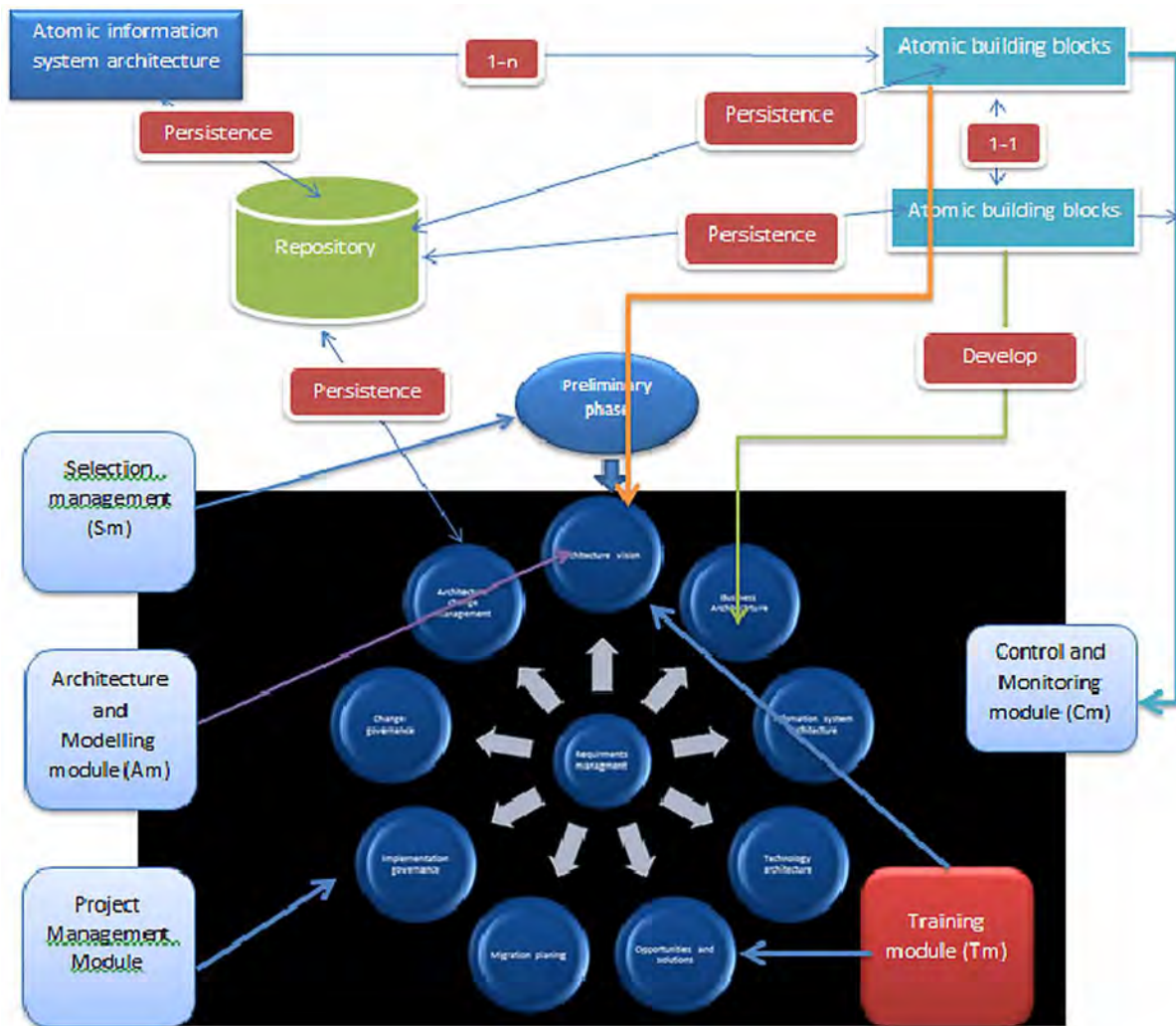
Projects, *Project-Management* and enterprise architectures are well-defined practices for conducting enterprise transformations, enterprise analysis, design, planning, and implementation, using a holistic approach at all times, for respecting a successful strategy. The main goal is to structure complex Information Technology (IT) systems in terms of their business, application, information, and technical layers, and to revise programs through *Project-Management* planning processes, as illustrated in Figure 2. Such organizations have to apprehend various techniques and methodologies before integrating complex frameworks like The Open Group's Architecture Framework (TOGAF) in their organization, as shown in Figure 2.

The proposed *Project-Management* is founded on TOGAF's ADM and the needed synchronization interfaces, as shown in Figure 3. The synchronization interfaces represent an asset for the initiation, development and maintenance of a *Project*; optimally automatable in both senses.

Technology is getting complex to manage, due to various problems with project management is-

Figure 1. The Environment's interaction with the project and change management component

Source: Trad, 2015c; Trad, 2015d



sues, like security, infrastructure, methodologies, management concepts, software development, business architecture etc. Therefore, micro changes that have minor implications on the *Project*, have to be implemented and as shown in Figure 4, are easy to be governed. The *Project* team should manage the micro changes and integrate them in the ADM's phases H, G, and F.

The Architecture Development Phase

The ADM can be enforced with *Project-Management* that adds a spiral model making the *Project*

iterative, this approach adds quality insurance and helps to deliver solutions on time, as shown in Figure 5. After the initiation of the *Project*, the designed building blocks that pass through each ADM's phase are updated with the new change requests. Each ADM iteration automatically updates and notifies the *Project-Management*'s scheduling tools and the *Project* plans. That means that the *Project-Management* schedule is dynamically generated and that represents a major *Project* or business benefit.

In the *Project*'s implementation phase, the developers' role is crucial, where their skills

Figure 2. Integration of various standardized methodologies in the transformation process

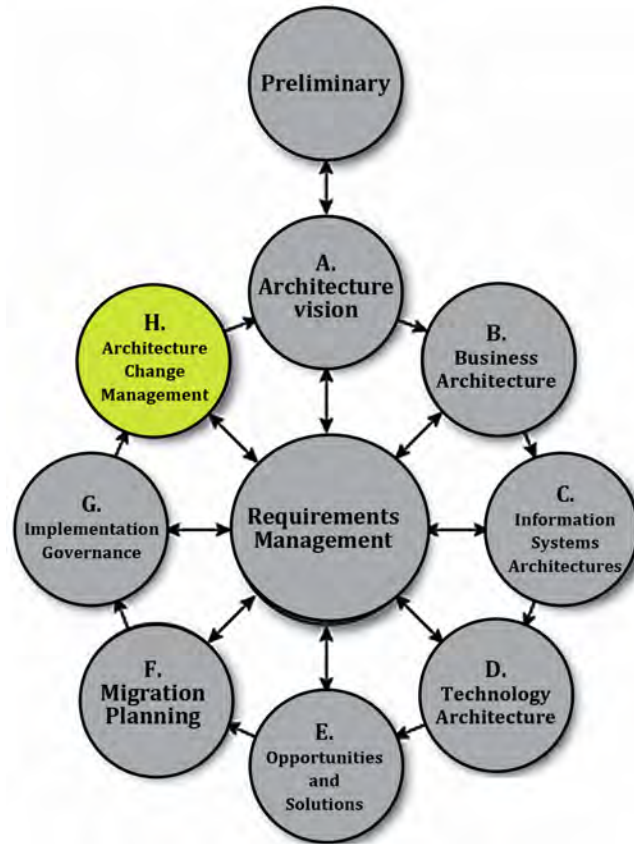


Figure 3. Linking various phases with the requirements

Source: Sparxsystems, 2015

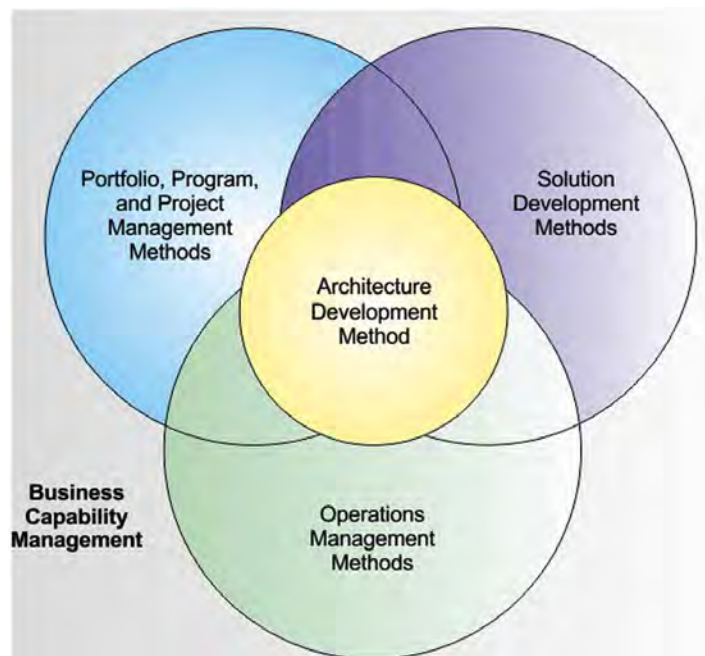


Figure 4. The phases that are relevant to project management

Source: The Open Group, 2015a

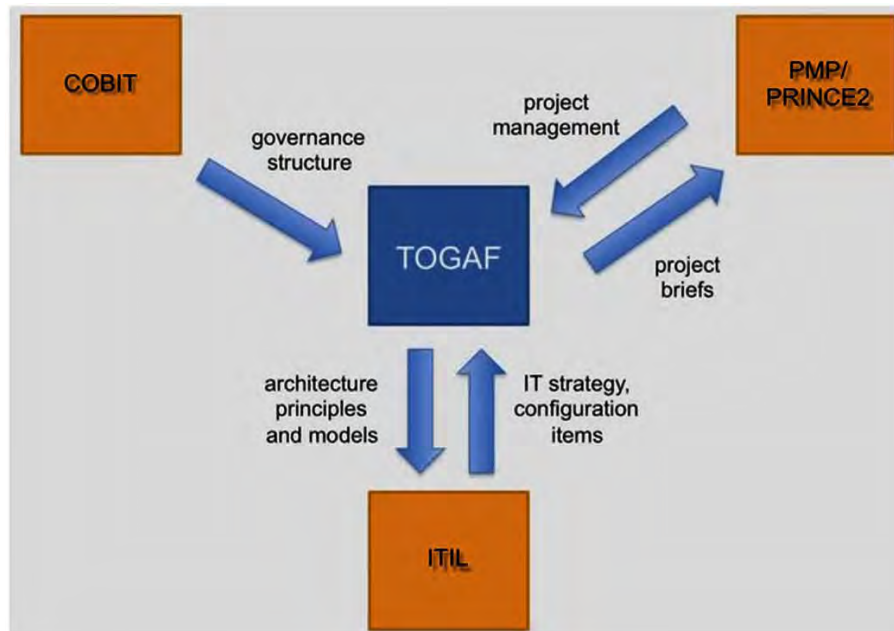
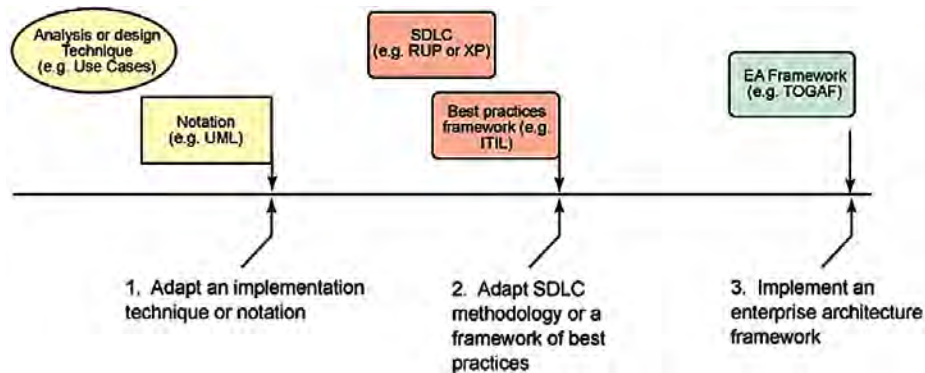


Figure 5. The spiral model

Source: UML, 2015



always represent two undeniable *Project* assets which can boost the *Project* towards success or failure. As shown in Figure 5, the Unified Modelling Language's (UML) spiral model can serve as a dynamic bridge between the *Project*, *Project-Management* and the ADM's phases, where their communication relies on standards like the Project Management extensible Mark-up Language format (PMXML).

The Maintenance

The *Project* has an evolutionary and not a revolutionary approach and it is recommended that the migration process and the interaction between the *Project-Management* and ADM make such a process possible. Because each ADM's iteration is reflected to the *Project-Management* and that makes migration planning and maintenance a playground where all *Project* teams come together and

Figure 6. The needed skills
Source: The Open Group, 2015b

B

Source: Resources	Type: <All>	Link Type: Realisation
Target: UC01-1: User Management	Type: <All>	Direction: Source -> Target
	UC01-1: User Management: Collaborations	
	UC01-1: User Management: Customer	
	UC01-1: User Management: Data Entry	
	UC01-1: User Management: Incompatibility	
	UC01-1: User Management: Incompatibility	
	UC01-1: User Management: Login Screen	
	UC01-1: User Management: Registration	
	UC01-1: User Management: Security	
	UC01-1: User Management: Update Versions	
	UC01-1: User Management: Workbench	
Resources: Andrew Sutton		
Resources: Claire Owens		

propose different changes that are synchronized with a *Project-Management* scheduling tool (Moe, Dingsøyr, & Dybå, 2010). The evaluation of each new change can be tested prior to its migration by simulating it with a corresponding CSF.

PROJECT AND CHANGE MANAGERMENTS' ROLE

Skills

The *Environment* defines the skills that are required for the *Project-Management's* integration in the *Project*, in order to deliver. These skills selection criteria should match TOGAF's Roles Program in order to select *Manager*. As shown in Figure 6, *Project's* project management skills typically comprise: a) managing business changes; b) project management methodologies; and c) development tools. The *Manager* must have also strong negotiation and problem-solving skills to work with the Project Management Office (PMO) in order to take the right decisions in just in time manner to keep the *Project* on track (The Open Group, 2015d).

Project Management Office

The PMO assists in the execution of *Project* Life-cycle Processes, which are based on best practices from standardized methodologies like Prince2 and TOGAF (Guiry, 2015).

STRATEGIC VISION, PLANNING, AND GOAL(S)

Project-Management is complementary to the ADM and as a matter of fact, the *Project-Management* should map to the organizational structure to the deliverables for each department that is led by a *Manager*. The *Manager* must have the skills for development of *Projects*, however modern business environments adopt a matrix project approach where each team is managed by more than one *Manager*. This fact can result in *Project-Management* collisions that can cause major delays, because several *Projects* can create unnecessary repetitive and endless *Project* activities like meetings that are due to disagreement between the leaders and team members.

A Holistic Agile View

The *Manager* must achieve an agile holistic view on the *Project* and such an agility should be integrated in its ADM, as shown in Figure 4 (APMG international 2011; The Open Group 2015c). The *Project-Management*, audit, risk and control mechanisms compose the core of enterprise's governance (ITGI, 2007; Tan, Carter, & Toleman, 2009).

Business Engineering and the Role of Standards

This *Environment* is based on technology and business engineering standards (OASIS, 2006) to insure that business engineering services' architecture and their *Project's* interfaces comply with the system's holistic concept. Actual advanced technologies like Management extensible Mark-up Language (XML) based architecture insure interoperability between all the *Project's* components in order to simplify the interface of the *Project-Management* to other components.

BUSINESS TRANSFORMATION AGILITY

Business transformation agility can be achieved by combining various agility methodologies and by using a bottom-up approach. The *Manager* should achieve the optimal business transformation agility needed to cope with the *Project's* implementation complexities. Agility is today supported by various methodologies and concepts that can be adapted to the ADM by using:

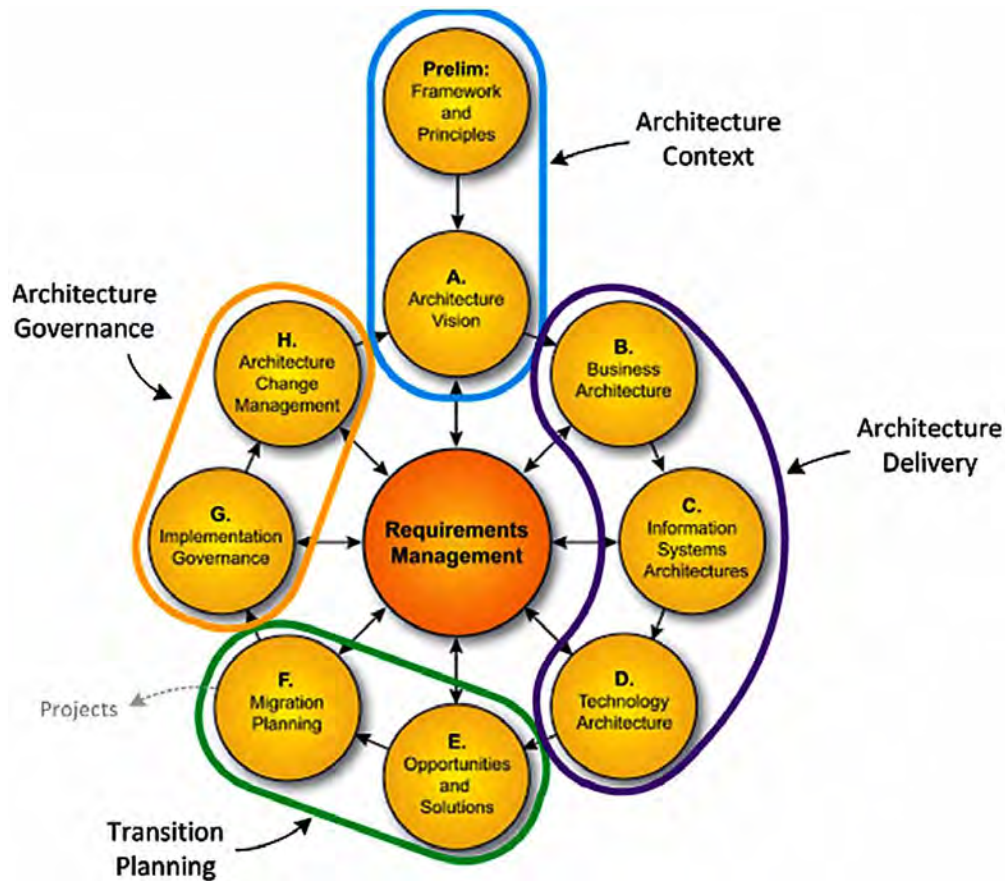
1. Lean software development methods which focus on the enterprise's value stream and are build on an atomic services concept. Lean development glues the enterprise, information system, *Project-Management* resources, and user into the company's value stream. *Project-Management* agility achieves: a)

success throughput low latency; b) planning reactive to change; and c) the focus on process and people (Coplien & Bjørnvig, 2010).

2. A lean project management framework can be assisted with the Scrum framework. Scrum is used for complex *Projects* in which failure rates are high and their complexity is impossible to predict. Scrum is a framework that has related best practices to make *Project's* development feasible (Beedle, 2001), and it allows to do just in time changes to keep the *Project* on track (Schwaber, 2004).
3. The initiation phase of a *Project* starts with a starting project package which can contain: a) a feasibility study report to support a proof of concept; b) contract documents; c) project terms and conditions; and d) responsibilities of the *Project's* team members.
4. Change management techniques for managing continuous changes permit the *Manager* to apply a deterministic approach to project management in order to manage structured plans, Gantt charts and various types of dynamic planning schedules. A framework like Scrum enables to adapt to *Projects* that are prone to frequent changes without the need for huge changes in the *Project-Management's* roadmap (Schwaber, 2004).
5. The *Manager* is a ScrumMaster, who is responsible for coaching other team members to face the *Project's* complexity (Schwaber, 2004). The *Manager* is also an Architect of Adaptive Business Environments (AofABE) and assumes the role of a ScrumMaster that fills the position of a project manager by using a Scrum framework. Traditionally, the project manager is or was responsible for defining and managing the *Projects*, as shown in Figure 7 (Schwaber, 2004).
6. Because of very high failure rates in *Projects* that are due to various types of human conflicts; heroic project's management might appear necessary (Cordier, 2013).

Figure 7. Scrum skeleton
Source: Schwaber, 2004

B



7. Standardized project management methodologies like Prince2 and Project Management Professional (PMP) can be adapted in order to adapt to the ADM's fast iterations, by the use of import and exports procedures.

gaps between the Baseline and Target Business Architectures. Business architecture, as shown in Figure 8, uses the following types of models: a) Business Process Models (BPM); b) Use-Case Models; and c) Class Models.

Business Architecture

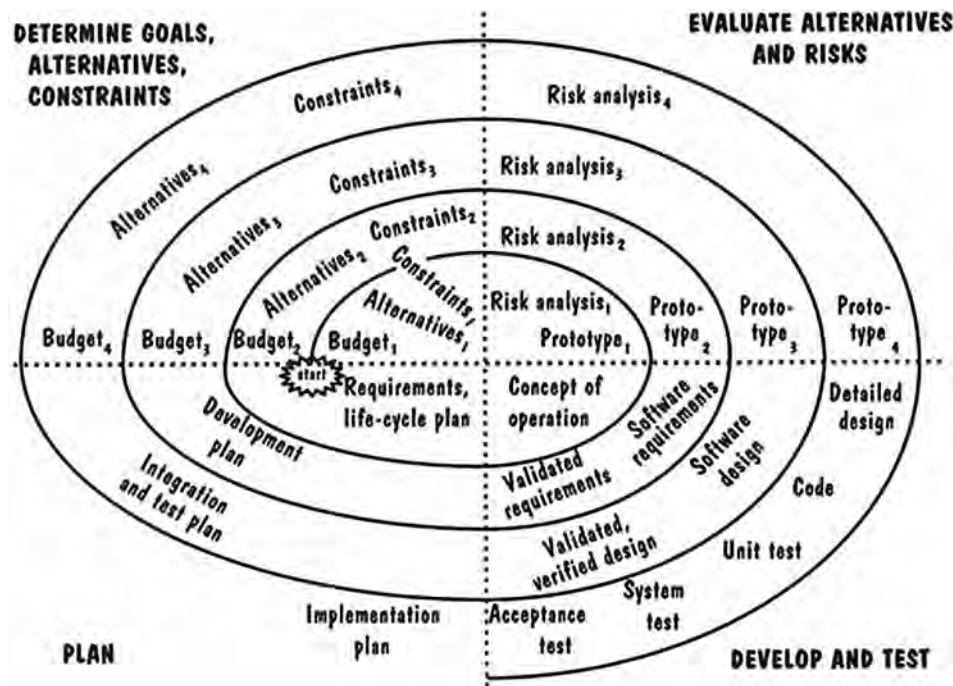
Business Architecture (BA) describes how the enterprise needs to operate to achieve its business goals and to respond to its strategic drivers that are set in the Architecture Vision process. This process addresses the Request for Architecture Work format and manages the stakeholder concerns in order to identify the candidate architecture roadmap's components that are based on

AGILE DECISION SUPPORT

Critical Success Factors

The selected Critical Success Factors (CSF) enable the *Project-Management's* filtered data that are logged into the monitoring database to be used for evaluation and problem solving purposes. The *Project* is based on BPMs which use CSFs to enable the evaluation of historical data-patterns.

Figure 8. Shows basic business architecture information technology requirements



All the ADM's phases manage the same set of selected CSFs to support the evaluation of encountered *Project* status, problems, requests, capabilities and objectives. CSFs are incorporated into the *Environment* and are CSFs will be updated with each ADM phase iteration's outcomes. The *Project's* tasks are fed into the *Project-Management* that can eventually reschedule the *Project* plan. PMXML plays a major role in the automated communication between the *Project-Management* and the ADM.

Decision Support System

Project evaluation or decision making addresses two major types of decisions: a) tangible; and b) intangible. Tangible decisions are the decisions which can be quantified and are based on measurements; where the best solutions are achieved by *Project-Management* automation. The *Project-Management* automation will facilitate and reduce the duration of the ADM's life-cycle; whereas intangible decisions require human interaction. The *Environment's* applied *Project-Management* component feeds the *Environment's* decision

Figure 9. The integration of business process critical success factors

© The Open Group	Architecture Board Member	Architecture Sponsor	Enterprise Architecture Manager	Enterprise Architecture Technology	Enterprise Architecture Data	Enterprise Architecture Applications	Enterprise Architecture Business	Program/Project Manager	IT Designer
Program or Project Management Skills									
Program Management	1	2	3	3	3	3	3	4	2
Project Management	1	2	3	3	3	3	3	4	2
Managing Business Change	3	3	4	3	3	3	4	4	2
Change Management	3	3	4	3	3	3	4	3	2
Value Management	4	4	4	3	3	3	4	3	2

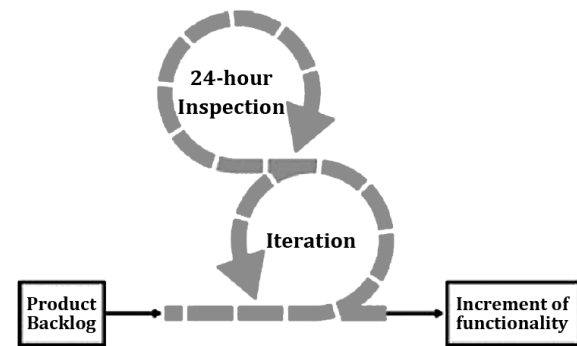
module, which is based on an empirical problem-solving algorithm uses CSFs. This decision module is specialized for *Project* problem solving and is based on a heuristics engine and has potential to improve the competitiveness of the business enterprise. The *Environment* presents a new approach for decision module, where this approach can be applied by the modification of an *Project* plan (Sreka & Slaninova, 2012).

The *Environment*'s decision module delivers a set of possible solutions that determine the *Manager*'s skills and assists the *Project* in solving various types of *Project-Management* problems, such as possible critical paths (Trad & Kalpić, 2014a). The case of complex *Projects*, project scheduling tools need to display information coming from the ADM in a cascade manner; in other words, one task would contain a group of sub-tasks; this facilitates the *Project*'s governance to manage unexpected problems (Ylimäki, 2006).

Types of Management Risks

Management of risk is central for managing the outcomes of any *Project* change effort and there are two levels of risks that should be considered (The Open Group, 2011a): a) the Initial Level of Risk; and b) the Residual Level of Risk. As shown in Figure 10 the ADM's phases, including phase M, phase H and phase G are important for governing the *Project*; this governing process includes monitoring of: a) various governance requests; b) new developments initiatives; and c) change requests, when a change is requested, the

Figure 10. Shows the various project phases



change management process will initiate a new architecture evolution initiative.

Initial Risk Assessment

The combination of frequency and effect can help in quantifying the *Project*'s risk assessment. In order to find the solutions for the preliminary risk assessment, the decision module must be used; however, there are no instant rules for quantifying the effect and frequency factors.

Conduct Residual Risk Assessments

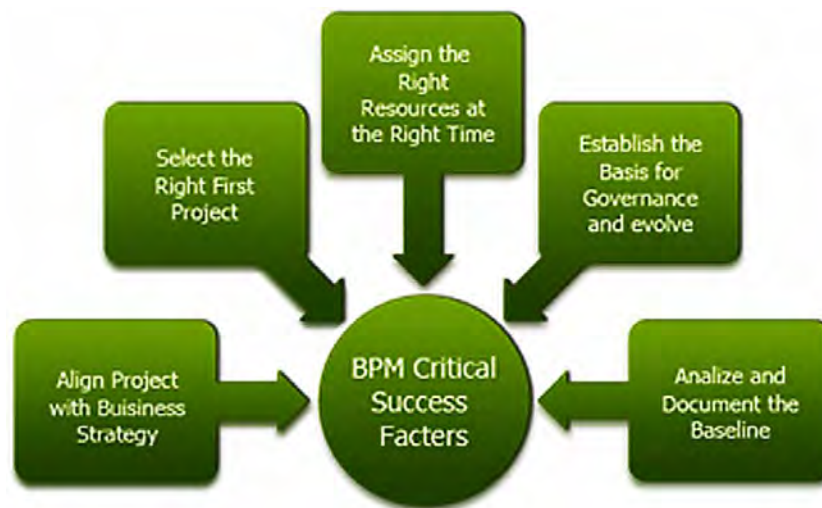
Once mitigation efforts have been identified for each CSF risks, continuous re-evaluation of the effect and frequency have to be done; the final deliverable of this process is the transformation risk assessment that could be estimated.

Figure 11. Business capability management

Source: The Open Group, 2011a



Figure 12. The interaction between standardized frameworks



Control Objectives for Information and Related Technology Framework

The Control Objectives for Information and Related Technology (COBIT) is a business framework for the governance and management of business environment's information system and insures that globally accepted standards, practices and models support the quality of the *Project*. (TOGAF, 2013; Sutherland, Viktorov, Blount, & Puntikov).

Information Technology Infrastructure Library

The standardized Information Technology Infrastructure Library contains a well detailed description of a seven-step improvement process that provides the capabilities to: 1) measure; 2) plan; and 3) implement business services improvement and their needed assertions. This seven-step improvement process is not only used on an operational level but it also provides support for all the ADM's phases (Van Sante & Ermers, 2009). In this seven-step improvement process the *Project* team implements the needed service level agreements that can be incorporated in the *Project-Management*.

SOLUTIONS AND RECOMMENDATIONS

Proof of Concept

The Proof of Concept (PoC) or solution was built using: a) Microsoft's project management environment; b) the XML standard export; and c) Microsoft Visual Studio; to support the following requirements:

1. To use the research's *Environment* including all its components.
2. That the *Project's* team uses enterprise architecture tool(s) to model the *Project* iterations; where the ADM phases create a class diagram that is used to generate the *Project-Management's* project plan.
3. That the ADM's class diagram entities, use the "1:1" mapping concept to inherit task class properties, like: a) title; b) task-id; c) start-time; d) end-time; e) priority; etc... The inherited task class properties are entered in the project plan.
4. That the ADM class relationships are deduced to create the project plans dependencies and critical path(s).

5. That the ADM class test driven development components are used to find the statuses of the completion and they are injected into the *Project-Management's* project plan as the project plan status.

Recommendations

This research initiative is part of a long series of publications related to business engineering, enterprise architecture and business transformation projects; and is based on a mixed action research model; where the *Environment's* critical success factors and managerial recommendations are offered to help *Managers* and enterprise architects to decrease the chances of failure. For that goal a real world framework is offered with a set of architecture, technical and managerial recommendations. The most important recommendation that was generated by the previous research phases was that the business transformation manager must be an Architect of Adaptive Business Environments (Trad & Kalpić, 2014a). The used literature review sources and the *Environment's* PoC proved the feasibility of this research phase and delivered the recommendations on how to implement a *Project-Management* component. These recommendations are:

1. The *Manager* is an Architect of Adaptive Business Environments and he must be also a Scrum master.
2. The *Project* adopts a holistic systemic agile *Project-Management's* approach that is based on enterprise architecture's ADM.
3. The ADM can be enforced with *Project-Management* outputs that add a spiral model, to the make the *Project* iterative.
4. The ADM's initial phase defines the *Project-Management* roadmap.
5. In order to insure the optimal unit of work that will align all *Project* resources the use of a "1:1" mapping concept is needed (Trad, 2015a; Trad, 2015b). Apply a "1:1" mapping concept for all the *Project* artefacts that can comprise ArchiMate diagrams, UML diagrams, project plans, class diagrams...
6. A *Project-Management* curriculum must combine project management, management sciences, risk management, legal aspects, enterprise architecture, team management, technology management and business engineering (Trad & Kalpić, 2016).
7. In the *Project's* implementation phase, the developers' role is crucial where their skills represent an undeniable asset, to avoid very high failure rates in *Projects* that are also due to various types of human conflicts; that is why heroic *Project* management might appear necessary (Cordier, 2013).
8. The *Project* has an evolutionary and not a revolutionary approach and it is recommended to apply a migration process as an interaction between the *Project-Management* and the ADM to make such a transformation process possible.
9. Standardized project management methodologies like Prince2 and Project Management Professional standards can be adapted in order to adapt to the ADM's fast iterations, by the use of import and exports procedures.
10. The *Manager* must deploy a multivendor strategy and an anti-locked-in approach, where the *Project-Management* component must be designed to open to the ADM and project management standards. To achieve this strategic goal, Availability and Performance Monitoring and the use of interfaces for Unified Monitoring Solution for the control of *Projects*, must be implemented and linked to the *Project-Management* component (Kowall & Fletcher, 2013).
11. The decision making module uses the *Project's* logging system's database using the critical success factors that determine which *Project-Management* item must be checked and logged (Şeref, Ravindra, Ahuja, & Winston, 2007).

12. Integrating other standardized frameworks like the Information Technology Infrastructure.

FUTURE RESEARCH DIRECTIONS

The *Environment* future research will focus on a neural network based decision making module, which use a micro services concept.

CONCLUSION

In this research article, the focus is on how to integrate the *Project-Management's* component in a *Project*. The authors' aim is to offer an agile *Project* management concept; that can be used to view the vision and to define the best possible set of CSFs. The *Project-Management* communicates dynamically with the ADM and from a technological perspective this type of communication offers three major advantages for the *Project*: a) clear initiation; b) iterative development; and c) successful migration. The clear initiation presents a well-defined structure, with a governable set of the *Project* activities to be exported from the *Project-Management* to the ADM. The *Project's* iterative development enforced with frequent prototypes ensures the quality of the end system. For the *Project-Management*, migration planning goes to another level of abstraction where non-technical agents can intervene on all levels of the *Project*.

The *Environment's Project-Management* managerial, architectural and technical recommendations and a real-world *Environment*, round up the *Environment's Project-Management* component's characteristics for business transformation projects.

REFERENCES

- Analysing Business. (2013). *Business architecture. Analysing business*. Retrieved from <http://analysing-business.com/business-architecture>
- APMG International. (2011). *Agile Project Management White Paper*. APMG International.
- Beedle, M. (2001). *Agile Software Development with Scrum*. Prentice Hall.
- Capgemini. (2009). *Business transformation: From crisis response to radical changes that will create tomorrow's business*. A Capgemini Consulting survey.
- Coplien, J., & Bjørnvig, G. (2010). *Lean Architecture for Agile Software Development*. John Wiley and Sons, Ltd.
- Cordier, V. (2013). *Project rescue & the heroic project management*. Retrieved from <https://www.projectsmart.co.uk/project-rescue-and-the-heroic-project-manager.php>
- Downey, K. (2003). *Architectural Design Patterns for XML Documents*. Retrieved from <http://www.xml.com/pub/a/2003/03/26/patterns.html>
- Gartner. (2013). *Scenario Toolkit: Using EA to Support Business Transformation*. Gartner Inc.
- Guiry, D. (2015). *Project Management & Administration*. Boston University.
- Hoidn, P., & Schwidder, K. (2014). Enterprise information technology Architectures-information technology Architecture Standards, TOGAF and OMG in more Detail, Key Architecture Work Products. University of Zurich.
- IBISTM. (2016a). *The Business Engineering Curriculum. Syllabi*. IBISTM. Retrieved from <http://www.ibistm.org/academics.html#content>

- ISO. (2000). *15704 Industrial automation systems — Requirements for enterprise-reference architectures and methodologies*. ISO.
- ISO. (2007). *ISO/IEC 42010 Systems and Software Engineering — Architectural Description*. ISO.
- ITGI. (2007). *COBIT 4.1- Framework Control Objectives Management Guidelines Maturity Models*. IT Governance Institute.
- Kabzeva, A., Niemann, M., Müller, P., & Steinmetz, P. (2010). Applying TOGAF to Define and Govern a Service-oriented Architecture in a Large-scale research & development (R&D). *Proceedings of the Sixteenth Americas Conference on Information Systems*.
- Lee, S., & Yong, H. S. (2010). Distributed agile: Project management in a global environment. *Empirical Software Engineering*, 15(2), 204–217. doi:10.1007/s10664-009-9119-7
- Moe, N. B., Dingsøyr, T., & Dybå, T. (2010). A teamwork model for understanding an agile team: A case study of a Scrum project. *Information and Software Technology*, 52(5), 480–491. doi:10.1016/j.infsof.2009.11.004
- OASIS. (2006). *The Framework for eBusiness*. An OASIS White Paper. The OASIS ebXML Joint Committee for OASIS. Retrieved from www.oasis-open.org
- Schwaber, K. (2004). *Agile Project Management with Scrum*. Microsoft Press.
- Şeref, Ahuja, & Winston. (2007). *Developing Spreadsheet-Based Decision Support Systems*. Michelle M.H. Dynamic Idea Publishing.
- Sparxsystems. (2015). *UML 2 Activity Diagram*. Retrieved from http://www.sparxsystems.com.au/resources/uml2_tutorial/uml2_activitydiagram.html
- Sreka, R., & Slaninova, K. (2012, April). The Usability of Agent-Based Simulation in Decision Support System of E-Commerce Architecture. *International Journal of Information Engineering & Electronic Bu*, 4(2), 10.
- Sugumaran, B., & Lavanya, M.R. (2014). Evaluation of Critical Success Factors In Construction Projects. *International Journal of Advanced Research in Civil, Structural, Environmental and Infrastructure Engineering and Developing*, 2(2).
- Sutherland, J., Viktorov, A., Blount, J., & Puntikov, N. (2007). Distributed scrum: Agile project management with outsourced development teams. In *System Sciences, 2007. HICSS 2007. 40th Annual Hawaii International Conference on* (pp. 274a-274a). IEEE.
- Tan, W., Carter, A., & Toleman, M. (2009). *Implementing information technology services management: A case study focussing on critical success factors*. Toowoomba, Australia: Univ. of Southern Queensland.
- The Open Group. (2011a). *Risk Management*. Retrieved from pubs.opengroup.org/architecture/togaf9-doc/arch/chap31.html
- The Open Group. (2015a). *The Open Group's Architecture Framework-Building blocks-Module 13*. Retrieved from www.open-group.com/togaf
- The Open Group. (2015b). *TOGAF-Architecture Skills Framework*. Retrieved from <http://pubs.opengroup.org/architecture/togaf8-doc/arch/chap30.html>
- The Open Group. (2015c). *The Open Group's Architecture Framework*. Retrieved from www.open-group.com/togaf
- Tidd, J. (2006). *From Knowledge Management to Strategic Competence* (2nd ed.). London, UK: Imperial College. doi:10.1142/p439

- Tidd, J., & Bessant, J. (2009). *Managing Innovation, Integrating Technological, Market and Organizational Change* (4th ed.). Wiley.
- TOGAF. (2013). *Project Management & Software Engineering*. Retrieved from <https://pmsware.wordpress.com/tag/togaf/>
- Trad, A. (2015a). *A Transformation Framework Proposal for Managers in Business Innovation and Business Transformation Projects-Intelligent atomic building block architecture*. Centeris.
- Trad, A. (2015b). *A Transformation Framework Proposal for Managers in Business Innovation and Business Transformation Projects-An information system's atomic architecture vision*. Centeris.
- Trad, A. (2015c). The Transformation Framework's Resources Library. IBISTM.
- Trad, A. (2015d). *The Transformation Framework Proof of Concept*. IBISTM.
- Trad, A., & Kalpić, D. (2014a). *The Selection and Training Framework (STF) for Managers in Business Innovation and Transformation Projects - The profile of an Architect of adaptive business systems*. IMRA.
- Trad, A., & Kalpić, D. (2014b). *The Selection and Training Framework (STF) for Managers in Business Innovation and Transformation Projects- Integrating the restructuring process of the global economy*. Cambridge, UK: IMRA.
- Trad, A., & Kalpić, D. (2014c). The Selection, Modelling and Training Framework (SMTF) for Managers in Business Innovation and Transformation Projects-An executive's business architecture and modelling strategy. *Journal of Business and Economics*, 5(1).
- Trad, A., & Kalpić, D. (2015a). *The Business Engineering Transformation Framework for (e)commerce Architecture-Modelling Projects*. *Encyclopaedia of E-Commerce Development, Implementation, and Management*. IGI-Global.
- Trad, A., & Kalpić, D. (2015b). *The Business Engineering Transformation Framework for (e)commerce Control and Monitoring Pattern*. *Encyclopaedia of E-Commerce Development, Implementation, and Management*. IGI-Global.
- Trad, A., & Kalpić, D. (2015c). The Business Transformation Framework for Managers in Business Innovation Transformation Projects-An educational business engineering and transformation module. AASCIT.
- Trad, A., & Kalpić, D. (2016). *The Business Transformation Framework for Managers in Business Innovation Transformation Projects - An Educational Business Engineering and Transformation Module*. AASCIT.
- UML. (2015). *Unified Modeling Language™ (UML)*. Retrieved from <http://www.omg.org/spec/UML/>
- Ylimäki, T. (2006). *Towards Critical Success Factors For Enterprise Architecture-AISA Project Report*. University of Jyväskylä.

ADDITIONAL READING

- Capgemini (2007). *Trends in Business transformation - Survey of European Executives*. Capgemini Consulting and the Economist Intelligence Unit. France.
- Daellenbach, H., McNickle, D., & Dye, Sh. (2012). *Management Science. Decision-making through systems thinking* (2nd ed.). USA: Plagrave Macmillan.
- Farhoomand, A. (2004). *Managing (e)business transformation*. UK: Plagrave Macmillan.
- Farhoomand, A., Lynne, M., Markus, M., Gable, G., & Khan, H. (2004). *Managing (e)business Transformation: A global Perspective*. Plagrave Macmillan.

IBM. (2009). TOGAF or not TOGAF: Extending Enterprise Architecture beyond RUP. IBM developerWorks. USA.

IBM. (2015). Enterprise architecture vision. IBM Knowledge Center. IBM. USA.

KEY TERMS AND DEFINITIONS

ADM: Architecture Development Method.

AofABE: Architect of Adaptive Business Environments.

Critical Success Factors: Can be used to manage the statuses and gaps in various project plans and gives the Projects the capacity to proactively

and automatically recognize erroneous building blocks and to just-in-time reschedule the Project plan(s).

Environment: It is this research's framework.

Manager: Business Transformation Manager.

Project: Business Transformation Project.

Project Management: It is an administrative and planning aspect of a *Project* and any other type of project where it sets the responsibilities, tasks and duties of the *Project*'s team members.

Scrum Master: It is not a traditional project manager, but a coordinator between the project team and the external world.

TOGAF: The Open Group's Architecture Framework.

The Business Transformation Framework and Its Business Engineering Law Support for (e)Transactions

Antoine Trad

Institute of Business and Information Systems Transformation Management, Switzerland

Damir Kalpić

University of Zagreb, Croatia

INTRODUCTION

In order to restructure the global economy in a sustainable manner, the integration of business engineering related legal standards is fundamental. Today, these legal standards are mature and can help in the transformation of traditional business environments to become a part of the newly interconnected global economy (Trad & Kalpić, 2014b). An important factor in continuously transforming a traditional business environment into an innovative and lean business engineering services oriented business environment are the roles of the Business Transformation Manager (BTM) and the needed business transformation framework. This framework is also needed to support the business environment's legal integration in the globalized environment. To achieve this legal support and integration, Critical Success Factors (CSF) must be used to legally assert, govern, automate, trace, monitor and control the Business Transformation Project's (BTP) artefacts. The CSFs can be used to manage the differences in business or (e)business local and international laws. Business environments must have the capacity to proactively and automatically recognize erroneous transactions, illegal activities, fraud and tax evasions (Trad & Kalpić, 2015a; Trad & Kalpić, 2015b; Trad, 2015b). A holistic legal assertion component for such projects is non-existent and is very complex to implement and an (e)system approach (or view) is suited for complex undertakings like the

integration and use of the Business Engineering Legal Assertion (BELA) component via the use of atomic building blocks and atomic solution blocks (Daellenbach & McNickle, 2005). In this research the focus is on the (e)transaction's legal assertion component that is a part of the this research project's control and monitoring module.

BACKGROUND

An BELA Component can be used in various business transformation and business engineering fields and this component is the main focus of this research phase. The BELA is a part of the Control and monitoring module (Cm). The Cm is in turn a part of the Selection management, Architecture-modelling, Control-monitoring, Decision-making, Training management and Project management Framework (SmAmCmDmTmPmF, for simplification in further text the term *Environment* will be used), that supports various aspects of BTP's activities. In this article the authors present a set of BELA managerial recommendations and a reusable pattern in the form of an (e)transaction legal support reference architecture (ISO, 2000; ISO, 2007; Trad 2015c; Trad c). The *Environment's* BELA must be synchronized through the architecture development method's phases, where each building block circulates through the architecture development method's phases. The building blocks contain sets of CSFs (Sugumaran

DOI: 10.4018/978-1-5225-2255-3.ch055

& Lavanya, 2014). These CSFs can be applied for: 1) the selection of the BTMs; 2) the implementation of BTP's architecture and modelling strategies; 3) the decision support system, in order to estimate the actual status of the BTP and to decide whether to stop or continue the on-going project (Gartner, 2013); 4) the control and monitoring engine with the needed BELA mechanisms; 5) the training needs of the BTP's team; and 5) the project management module's support.

A well-designed business architecture, as shown in Figure 1, must define (Analysing business, 2013): 1) the BTP's objective(s); and 2) a loose coupled BELA component that is a part of the control and monitoring system (Trad & Kalpić, 2015b). (e)Business or business engineering driven business environments refer to various types of businesses that are conducted using different types of avant-garde technologies, electronic media and services oriented technologies; where

the most common form is the business that makes its transactions and revenue via the web, using automation (e-business, 2014). In 2015, we can simply talk of plain business engineering, because many engineering artefacts like service oriented architecture, business process choreography have merged with business engineering fields; therefore there is a need for BELA robustness that is supported by the *Environment's* control and monitoring module, which covers the business engineering domain. The global research topic's and final research question (hypothesis #1-1) is: "Which business transformation manager characteristics and which type of support should be assured in the implementation phase of an business transformation project?" The targeted business domain is any business environment that uses: 1) internet and engineering technologies; and 2) frequent transformation iterations. For this phase of research the sub-question (hypothesis #3-4) is:

Figure 1. Business architecture's interaction with the environment
Source: Analysing business, 2013



“What is the impact of the business engineering legal support on (e)enterprise architecture and business transformation projects ?”

In this research phase the authors are targeting the BELA component's integration that as a part of the control and monitoring module and its influence on BTPs (Trad, 2015d). Actually there is a lack of awareness of business engineering integration that can hinder the financial status of business engineering oriented companies. This barrier for a successful BTP's finalization has to be handled. In this chapter the authors use industrial experiences made and lessons learned from surveying many enterprises. The surveys' outcome was that these enterprises are insufficiently informed about the impact of BELA laws that are applicable in business environment. Nearly half of these enterprises (45.8%) have responded and expressed this view. It seems that the problem lies in the current status of business engineering's related legislation and its integration. This integration can be considered to be a legal barrier and a BTP major issue. The BELA problems reported affect all parts of (e)transactions' execution process (Merriam-Webster, 2015). The vast majority of the surveyed enterprises expressed the importance of concluding a valid business engineering contract; the survey also showed that enterprises faced the following BELA problems related to (European commission, 2004):

1. The recognition of (e)contracts (39%).
2. The validity of (e)signatures (30%).
3. The recognition of (e)taxation (29%).
4. International provision of a (e)transactions by electronic means (27.5%).
5. The robustness of (e)payments (25.5%).
6. The efficiency of (e)marketing and advertising activities (25%).

These facts are fairly significant and they should be treated with caution because some enterprises which were surveyed are incapable to precisely identify the type of BELA problem they faced. That is why the automation of (e)

transaction's legal assertion becomes vital. The identification of BELA problems is usually considered to be of general nature and very difficult to localize and to be interpreted. Where, almost half (or 49.2%) of the encountered BELA problems are due to the huge volume of (e)transactions between enterprises. Insuring the BELAs should become the enterprise's strategic objectives and they should be treated in its strategic vision (European commission, 2004).

MAIN FOCUS OF THE ARTICLE

The BTM has to handle the complex and chaotic technical implementation phase of BTPs that are the major cause of very high failure rates (CapGemini, 2007; CapGemini, 2009). Therefore the implementations of Internet-based technologies and engineering resources in BTPs require specific (e)business and business engineering architecture strategy and real-world professional knowledge. For this specific goal the Institute of Business and Information Systems' Transformation Management (IBISTM) has developed: 1) a real framework; 2) education curriculum; 3) training programmes; and 4) certification programmes, to support BTMs and hence BTPs (IBISTM, 2015).

STRATEGIC VISION, PLANNING, AND GOAL(S)

Business engineering fields are very much related to (e)business models such as business-to-consumer (B2C) and consumer-to-consumer (C2C) and they need automated BELA components and control support to be insured against various types of bad intentions. By the year 2016, 40% of all business companies will standardize their governance, control and monitoring architecture; comparing this to less than 10% today; proves that the BELA's integration is a strategic goal for all BTPs (Kowall, & Fletcher, 2013). As shown in Figure 2, the BTM should be a member of the

company's strategic planning team and he should closely work with the company's legal team, where he can bring a view on changes in BELA regulations, control and governance integration (Trad, 2015b).

BUSINESS AGILITY

Business agility is achieved by combining various agility methodologies, using a bottom-up approach. The BTM can achieve an optimal business agility that can cope with local, national and international business engineering legislation constraints.

A Holistic Agile View

The BTM must achieve an agile holistic view on the BTP and that this agility is integrated in the enterprise's architecture development method through the use of various agile methods, as shown in Figure 3 (APMG international 2011; The Open Group 2014b). BELA components, risk and control compose the core of (e)enterprise's governance (ITGI, 2007; Tan & Carter & Toleman, 2009), where the integration of the BELA is integrated into the *Environment* using CSFs and business standards using agile methods.

Business Engineering and the Role of Standards

The research project uses technology and business engineering standards (OASIS, 2006) that include: 1) legislation and regulation frameworks; 2) enterprise architecture frameworks; 3) business process monitoring; 4) governance of processes and services; 5) business services monitoring; 6) agile standards; and 7) platform control and monitoring standards (Desfray, 2011; Cummins, 2010).

Business Engineering Service's Granularity

How granular should the related business services be and which types of atomic services should they contain BELAs, governance control and monitoring mechanisms, that all defined by the global architecture. These processes and services are classified into specialized repositories; a line of business services is a collection of similar business processes and services that should be classified in a corresponding categories in the BTP's repository. In order to have a successful unbundling process and hence a successful BTP, the BTM must integrate business services governance frameworks (Olson, 2008). Service oriented architecture governance focuses on the life cycle

Figure 2. Interaction between various parties in the global economy

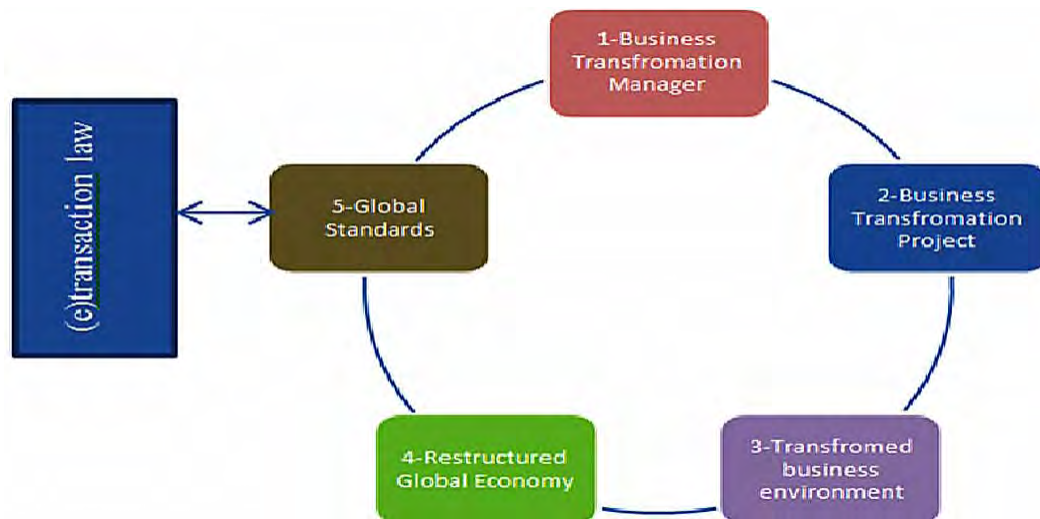


Figure 3. Combining various agile techniques

Source: APMG international, 2011

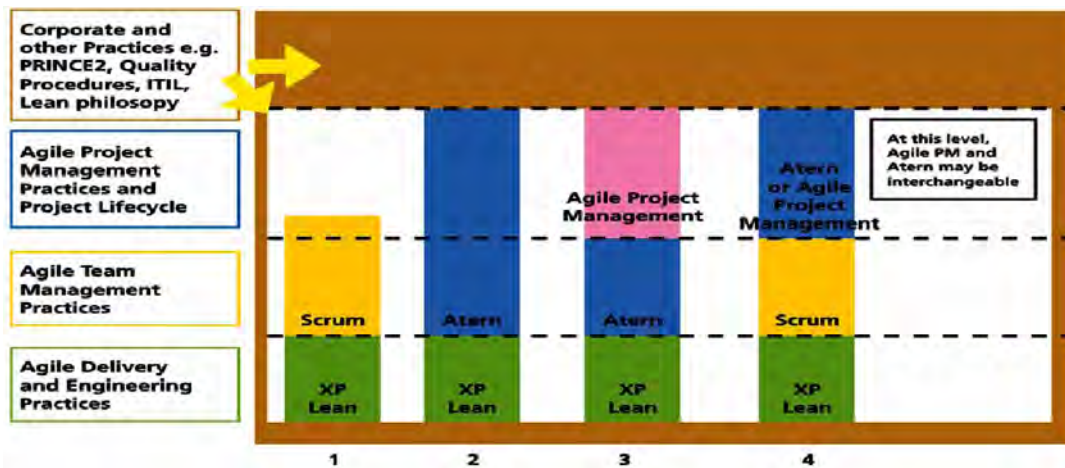
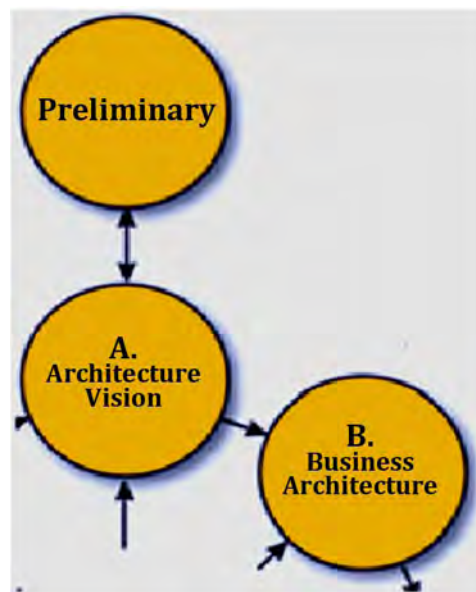


Figure 4. The (e)business architecture phase

Source: The Open Group, 2014b; Hoidn & Schwidder, 2014



of a business service from its inception through modelling, assembly, deployment, management and eventually retirement (Trad & Kalpić, 2014c).

Business Engineering Services' Architecture

The business services' architecture and their BELA interfaces paradigm are based on the "1:1" concept.

This concept helps in achieving the assertion of: 1) the global business engineering concept; and 2) monitoring and integration activities. The BTM, or the enterprise architect, builds a prototype to define a template based on atomic building blocks. These building blocks contain BELA, control and monitoring mechanisms (Trad, 2015a). The outcome of the unbundling process is a bank of state-full atomic business services that can be as-

serted, controlled and monitored in real-time. In order to manage the agile and autonomic BTP's complexity in its implementation phase, an adequate assertion concept must be integrated in the architecture (OASIS, 2014). Such a concept is enabled by the establishment of a real world iterative model that can map all the BTPs artefacts in a linear "1:1" manner (The Open Group, 2014b).

(e)TRANSACTIONS

There is a need for a multi-end (e)transaction model that minimizes the dependencies between the various parties; where the (e)transaction connection is established between the consumer and the end supplier with minimal risks (Jin & Zhu, 2011). One of the most important CSFs that can be integrated and controlled in the BTP is the cost of (e)transaction ratio after and before the BTP completion. To calculate the cost of an (e)transaction there is the need to add-up the total cost of providing the business service and to divide it by the total number of successfully executed (e) transactions (Government Service Design, 2015). (e)transactions' management and outcomes have to be continually legally asserted, traced, reviewed, logged and their periodic summaries must be reported to the BTM and the enterprise's executive management (Fu & Mittnight, 2015).

(e)Transactions' Security Violations

Business environment's roles are orthogonal to security requirements, where the business environment roles define the responsibility for enterprise's resources. Management of the enterprise's legal interests, resources, governance, support, usage and reporting of access activity, should be managed by the BTM or the enterprise architect(s). In special cases, like the case of denial of request for access to sensitive business resources, it should be managed by the BTM. Thus the business environment's structure is an important consideration in the legal assertion, governing, control, access management and monitoring of (e)transaction's

security. The (e)transaction's security uses the following standards: 1) the policy framework (Oracle, 2015a); and 2) business messaging is needed for domain-specific assertion languages (Oracle, 2015b; OASIS, 2009). The regulation for the (e)transaction's security needs qualified time-stamps for robust (e)certification like: (e) seal, (e)signature, or other techniques (European Union, 2014).

(e)Transaction Law

From the business engineering point of view, the Uniform Law Commissioners promulgated the Uniform Electronic Transactions Act in 1999. It is the first adaptable effort to prepare an (e)law for business engineering and (e)commerce fields. Many countries and regions have already adopted business engineering standards to support these fields like, for example, the digital signature's legislation. The Uniform Electronic Transactions Act represents the first effort in providing some standardized rules to govern (e)transactions in (e) commerce. The Uniform Electronic Transactions Act rules are related to the Uniform Commercial Code, where the Uniform Electronic Transactions Act rules are primarily for (e)records and (e)signatures that are related to the execution of (e)transactions. An (e)transaction is composed of related actions that are executed between two or more end-points (The Uniform Law Commissioners, 2015).

LEGAL BASIS AND INTEGRATION

The evolution of business engineering forces various industries to implement BELA mechanisms into their business transformation frameworks. Such frameworks must incorporate (e)transactions legal collaboration with: 1) international; 2) national; and 3) local-regional laws and rules. Besides international and national laws, there are many legal frameworks, conventions, treaties and directives to be reviewed for the support of (e) system's (e)transactions.

The United Nations Commission on International Trade Law

The evolution of the Internet and service oriented technologies in business engineering has forced various global players to regulate all possible forms of (e)transactions (Wright, 1991). The exponential rise in the volume of (e)transactions has created a new dimension to this complex problem. The increasing numbers of (e)transactions in both international, national and local business engineering trade operations are achieved by means of (e)data interchange using wide means of communication technologies known as (e)commerce, which is a business engineering field. (e)commerce replaces the traditional business exchanges and replaces the classical paper-based exchanges in order to improve productivity and to enforce in real-time BELAs (Chissick & Kelman, 2000). This fact has motivated the United Nations Commission's push to establish the International Trade Law (UNCITRAL), which is a model of (e)law that supports various business engineering fields. This (e)law package is downwards compatible with all national and regional legal systems. The BTMs should be aware of various (e)law initiatives which define the rules for trans-border (e)commerce collaboration, that is crucial for an BTP. The BTP should accommodate all future legal issues and challenges in the future (e)system. Based on this fact, the BELA and rules integration must be adjustable, flexible, feasible and enforceable; this can be achieved by linking a legal act with a CSF. The *Environment* provides a BELA mechanism to wrap the existing national legal system and to be synchronized with current international (e)initiatives; taking into account the borderless nature of (e)transactions (The United Nations, 2004).

The European Union's (e)Commerce Directive

On the European commission's level, a legislation has been implemented to govern (e)commerce

collaboration, and an important progress has been done in the assertion of business engineering. European commission's member states have implemented and enforced business engineering related national practices (European commission, 2015a).

Business Engineering Legislation Assertion

The integration of the business engineering module is done with the use of the standardized legal environment of The Open Group's Architecture Framework. This legal environment supports data protection laws, contract law, procurement law, fraud law and many other legislation domains. The BTM and his team must design a pattern that links:

1. The initial business requirement, with
2. The applied business process model, with
3. The (e)transaction building blocks, with
4. The service level agreement, with
5. A well-defined set of (e)law articles and regulations to be respected by the (e)transaction, with
6. The possible set of legal risks, which are represented by a CSF.

As shown in Figure 5, the BTM and his team must have the needed set of skills to design a pattern that integrates various legislation topics.

Decision System's Integration and Risk Management

The *Environment*'s decision system delivers a set of possible solutions that determine BTM's skills and assists the BTP in solving various types of problems, such as possible legal pitfalls. The decision support can be also represented as an implementation of a business process model; such a solution is optimal because of the *Environment* BELA knowledge that can be stored in the (e) system (Trad & Kalpić, 2014d).

Figure 5. The Open Group's legal environment

Source: The Open Group, 2015a

IT Architect Roles	Architecture Board Member	Architecture Sponsor	IT Architecture Manager	IT Architecture Technology	IT Architecture Data	IT Architecture Application	IT Architecture Business	Program or Project Manager	IT Designer
Legal Environment									
Contract Law	2	2	2	2	2	2	2	3	1
Data Protection Laws	3	3	4	3	3	3	3	2	2
Procurement Law	3	2	2	2	2	2	2	4	1
Fraud	3	3	3	3	3	3	3	3	1
Commercial Law	3	3	2	2	2	2	3	3	1

DECISION MAKING

Artificial Legal Intelligence

The *Environment* offers a heuristics based legal intelligence model to support legal reasoning mechanisms. The authors reviewed a number of systems that apply artificial intelligence for legal reasoning and have remarked that artificial legal intelligence model offered a holistic vision of (e) law, while a heuristics based legal intelligence model can offer solutions to BELA problems (Gray, 1997).

The Implementation

As shown in Figure 6, the *Environment's* applied BELA component feeds the decision making module that is based on an empirical problem-solving algorithm, with the needed success factors. The *Environment's* decision making module is specialized in the cases of problem solving and is based on a heuristics engine. The decision making module has an immense potential to improve the competitiveness of business engineering oriented enterprises; where it is an engine that enhances the BTM's ability to make decisions on selected (e)law issues (e-commerce, 2014). The *Environment* presents a new approach for decision support systems; where this approach can be applied by the modification and extension of existing

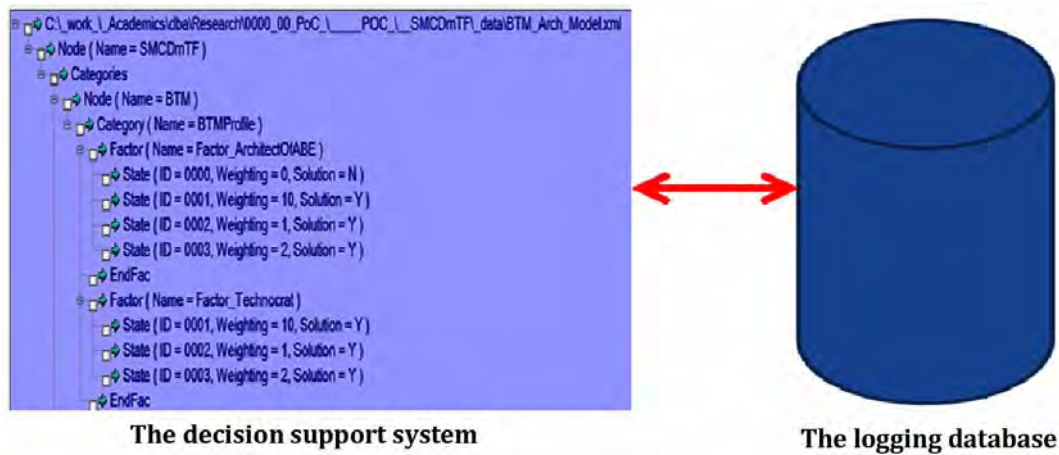
decision support system's basic CSFs (Sreka & Slaninova, 2012).

A BTM selection request or a specific BTP problem solving request can be fed in the form of CSF(s), into the *Environment's* reasoning model's root node and then the search process is launched to propose the optimal solution. The *Environment's* decision tree that is shown in Figure 6, is based on CSFs' mapping to BELAs' mechanisms; the mapping rules should be defined in the BTP's architecture vision.

Generated Events

Many events that are generated by the execution of an (e)transaction can affect the success of BTPs and the (e)system, therefore there is a need to capture and process all of them. The potential business impact on the BTP's (e)transactions legislation assertion must be assessed. Primarily, the costs related to BELA, control and processing can have an important impact on the all over costs of the BTP. Different BTP BELA techniques should be defined to ensure the accessibility to (The Open Group, 2014c): 1) consistent data; 2) credible (e)transaction information; 3) needed legal knowledge; 4) decision support for legislation governance; 5) monitoring; 6) analysis of logged data; and 7) planning. Information technology tools provide support for the planning and implementation of BTP's capabilities and legal control.

Figure 6. A view on the environment's tree solution nodes
Source: Trad, 2015c



Critical Success Factors

(e)transaction's trusted business information exchange that is enabled by the agreement of business partners, must be insured by a cohesive and coordinated interoperability framework. This interoperability framework should include the integration of the BELA environment, by the use of CSFs for linking various legal frameworks, standards, rules and agreements (European commission, 2015b). The *Environment's* BELA component uses the tracked information that is stored in the logging database. The stored information is filtered by using the CSFs that helps in analysing the Business Engineering environment to verify: 1) BELA violations; 2) performance; and 3) service level agreements' usage. The instances of business process models are tracked and the selected CSFs enable that the processes' filtered data are logged into the monitoring database. Business process models use CSFs to enable the management of certain types of historical data-patterns; these patterns are used in decision making process (IBM, 2015).

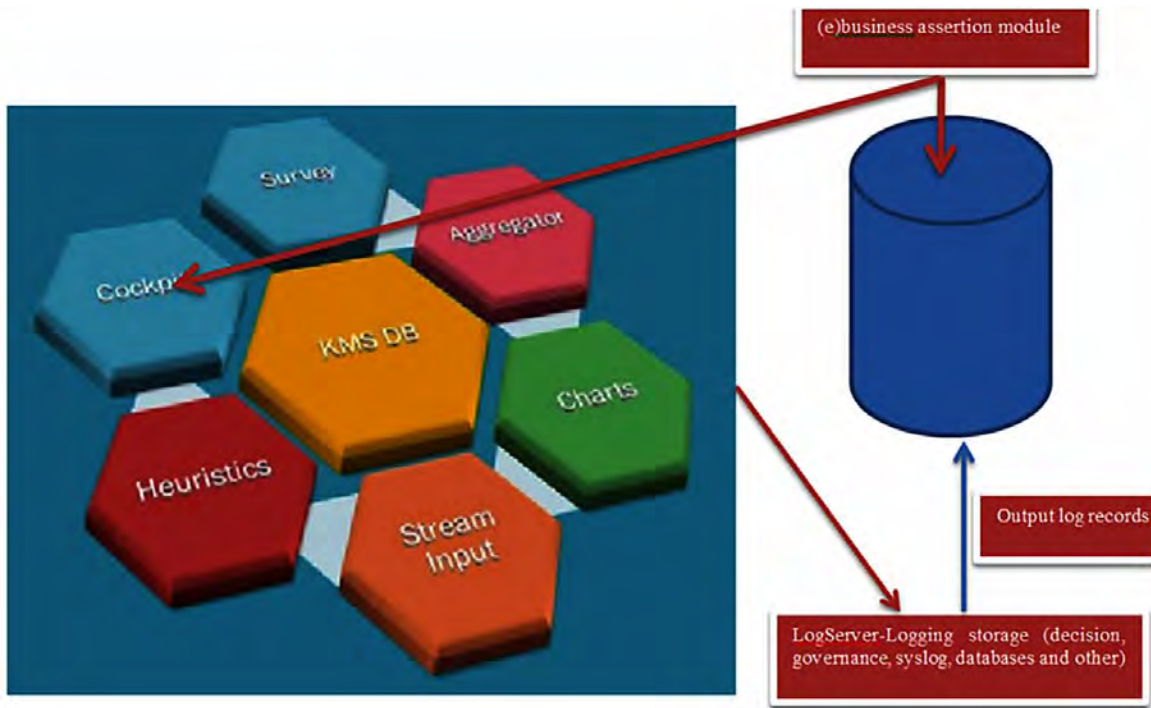
SOLUTIONS AND RECOMMENDATIONS

Proof of Concept

The proof of concept was built using the Microsoft Visual Studio; as shown in Figure 7, the proof of concept is based on the following facts:

1. It uses the research's *Environment* as a whole and its decision module.
2. The BTM and the BTP team members or the (e)enterprise's legal team, link legal constraints to specific (e)transactions' building blocks. This linking is done through the use critical factors to specific legal acts to be asserted.
3. The (e)transactions' model contains exception classes which check the legal constraints and if one or more of these constraints are violated, the (e)transaction is paused and warning messages are stored in the (e)system's log-server. The warnings are escalated to the support team.
4. The business engineering component scans the log-server for possible violations and if found, they are automatically reported to corresponding enterprise's department.

Figure 7. The (e)business legislation assertion's proof of concept
Source: Trad, 2015c



Critical Success Factors' Setup

Business interoperability is enabled by the use of CSFs' mapping. CSFs bind various legal acts, frameworks, standards, rules and service agreements (European commission, 2015b).

Process Models

The (e)transaction was prototyped using Sparxsystem's Enterprise Architect and Microsoft's .NET visual studio environments. The (e)transaction business process model uses web-services that make calls to the BELA class. The process model represents the relations between the (e)transaction's requirement, building block and an BELA process. As shown in Figure 8, the control and monitoring interface stays unchanged, whereas the logged data are interpreted differently to check business engineering violations.

FUTURE RESEARCH DIRECTIONS

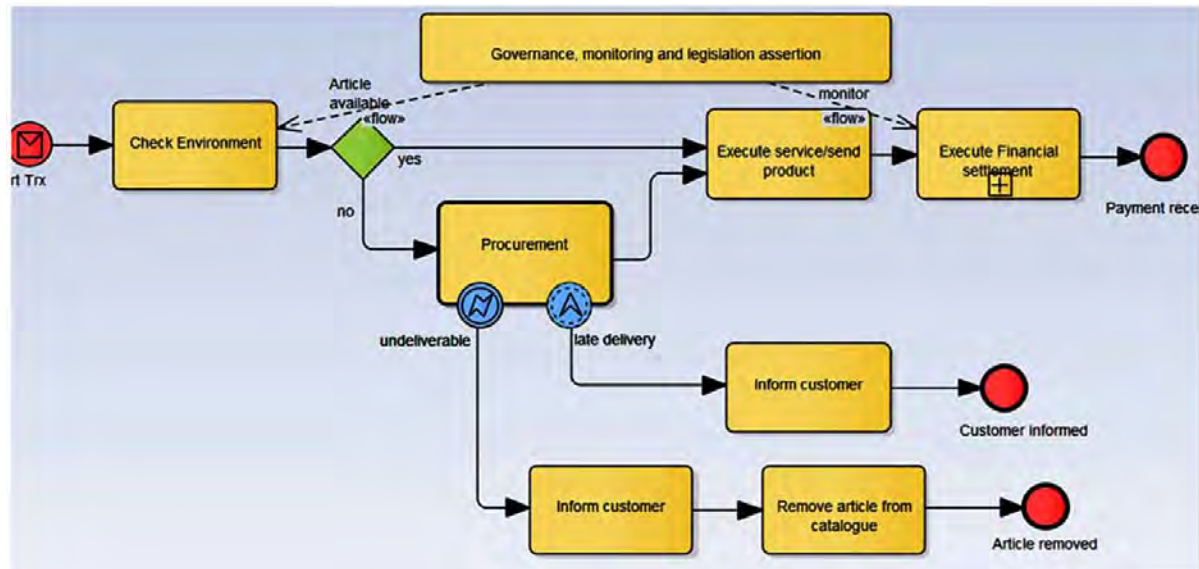
The *Environment* future research will focus on a neural network based decision making module, which use a micro services concept.

CONCLUSION

The BELA component's holistic integration is an important factor for the business engineering environment's evolution and stability. In the past recent years, many industries have been implementing BELA solutions to proactively respond to probable legal problems and challenges. This research initiative is part of a long series of publications related to the business engineering architecture and business transformation projects. This research is based on mixed action research model; where the *Environment's* CSFs

Figure 8. The (e)business legislation assertion's process model

Source: Trad, 2015c



and managerial recommendations are offered to help (e)system's architects to decrease the chances to fail. A real world framework is offered. The most important managerial recommendation that was generated by the previous research phases was that the business transformation manager must be an architect of adaptive business system (Trad & Kalpić, 2014a). In this research phase and article, the focus is on how to integrate the BELA's component in a BTP. The used literature review sources and the *Environment's* proof of concept proved the feasibility of this research's phase and delivered the recommendations on how to implement a BELA component. The authors present the *Environment's* BELA list of managerial, architecture and technical recommendations that are sorted by their importance:

1. **This Phase's Proof of Concept:** Proved the research project's feasibility by implementing the proposed recommendations and an architectural pattern.
2. **The Role of Standards:** The *Environment* uses (e)law, architecture, technology and business engineering monitoring standards

(OASIS, 2006; European commission, 2004; European commission, 2015a).

3. **The BTM Must Deploy a Multivendor Strategy and An Anti-Locked-In Approach:** The BELA component must be designed by the BTP team and its implementation must be set as the company's strategic goal. To achieve this strategic goal, Availability and Performance Monitoring and the use of interfaces for Unified Monitoring Solution for the control of BTPs, must be implemented and linked to the BELA component (Kowall & Fletcher, 2013).
4. **The BTM's Basic Skills Should Encompass Knowledge of:** 1) Business Engineering and (e)commerce processes and the needed logging, monitoring and assertion architectures; 2) automated real-time BELA environments; 3) (e)enterprise or enterprise 2.0 architectures and implementations; and 4) governance integration.
5. **Use of Business Process Models:** Controlled and monitored business process models will enhance the management of legislation knowledge and will help also in the Business transformation project's implementation

- phase (Trad & Kalpić, 2013a; Birudavolu & Nag, 2011).
6. **Design and Implementation of a BELA Component:** For business transformation project there is a need to implement a BELA component that can be easily integrated with any framework or tool. The *Environment* is an applicable framework and the BTP should be supported by a configurable BELA component (IBISTM, 2015).
 7. **Decision and Governance Logs:** To manage BTPs, the team must implement an (e)system decision and governance logs management mechanism for the support of the Business Engineering environment.
 8. **Legal Intelligence, Decision Making Module and CSFs:** The decision-making module uses the BTP's logging system's database. The CSFs determine which BELA item must be checked and logged (Şeref & Ravindra & Ahuja & Winston, 2007).
 9. **Integrating Other Frameworks:** Operations', the Information Technology Infrastructure Library and other standard frameworks can be integrated in the BTP through the use of CSFs.
- REFERENCES**
- Analysing business. (2013). *Business architecture. Analysing business*. Retrieved from <http://analysing-business.com/business-architecture>
- APMG International. (2011). *Agile Project Management White Paper*. APMG International.
- Birudavolu, S., & Nag, B. (2011). *A Study of Open Innovation in Telecommunication Services: A Review of Literature & Trends*. Indian Institute of Foreign Trade.
- Capgemini. (2007). *Trends in Business transformation - Survey of European Executives*. Capgemini Consulting and the Economist Intelligence Unit.
- Capgemini. (2009). *Business transformation: From crisis response to radical changes that will create tomorrow's business*. A Capgemini Consulting survey.
- Chissick, M., & Kelman, A. (2000). *Electronic commerce: law and practice* (2nd ed.). London: Sweet & Maxwell.
- Cummins, F. (2010). *Building the Agile Enterprise-with SOA, business process model and MBM*. Morgan Kaufmann Publications.
- Daellenbach, H., McNickle, D., & Dye, Sh. (2012). *Management Science. Decision-making through systems thinking* (2nd ed.). Plagrave Macmillian.
- Desfray, Ph. (2011). *Using OMG Standards with TOGAF*. SOFTEAM – Modeliosoft.
- European commission (2004). *Commission of the European Communities Legal barriers in e-business: The results of an open consultation of enterprises*. Brussels, 26.4.2004 SEC(2004) 498. European commission. EU.
- European commission. (2015a). *e-Business legal aspects*. European commission. Retrieved from http://ec.europa.eu/enterprise/sectors/ict/documents/e-business/index_en.htm
- European commission (2015b). *Annex 7 - Critical Success Factors*. European commission. EU.
- European Union. (2014). *Regulation (EU) No 910/2014 of the European Parliament and of the Council - on electronic identification and trust services for electronic transactions in the internal market and repealing Directive 1999/93/EC The European Parliament and of the Council – Regulation*. European Union.
- Fu, Z., & Mittnight, E. (2015). *Critical Success Factors for Continually Monitoring, Evaluating and Assessing Management of Enterprise IT*. ISACA. Retrieved from <http://www.isaca.org/COBIT/focus/Pages/critical-success-factors-for-continually-monitoring-evaluating-and-assessing-management-of-enterprise-it.aspx>

- Gartner. (2013). *Scenario Toolkit: Using EA to Support Business Transformation*. Gartner Inc.
- Government Service Design. (2015). *Measuring cost per transaction*. Retrieved from <https://www.gov.uk/service-manual/measurement/cost-per-transaction.html>
- Gray, P. (1997). *Artificial legal intelligence*. Dartmouth Publishing Co.
- IBISTM. (2015). *The Business Engineering Curriculum. Syllabi*. IBISTM. Retrieved from <http://www.ibistm.org/academics.html#content>
- IBM. (2015). *Tivoli Usage and Accounting Manager Enterprise Edition accurately assesses shared computing resource usage*. IBM.
- ISO. (2000). *15704 Industrial automation systems — Requirements for enterprise-reference architectures and methodologies*. ISO.
- ISO. (2007). *ISO/IEC 42010 Systems and Software Engineering — Architectural Description*. ISO.
- ITGI. (2007). *COBIT 4.1- Framework Control Objectives Management Guidelines Maturity Models*. IT Governance Institute.
- Jin, K., & Zhu, J. (2011). *MET: Multi-party E-Commerce Transaction Model*. Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference, Boston, MA.
- Kowall, J., & Fletcher, C. (2013). *Modernize Your Monitoring Strategy by Combining Unified Monitoring and Log Analytics Tools*. Gartner.
- Merriam-Webster (2015). *Transaction*. Merriam-Webster an Encyclopædia Britannica Company Dictionary.
- OASIS. (2006). *The Framework for eBusiness*. An OASIS White Paper. The OASIS ebXML Joint Committee For OASIS. Retrieved from www.oasis-open.org
- OASIS. (2009). *OASIS Web Services Reliable Messaging (WS-ReliableMessaging) Version 1.2*. OASIS Standard. Retrieved from <http://docs.oasis-open.org/ws-rx/wsrn/200702/wsrn-1.2-spec-os.html>
- OASIS. (2014). *ISO/IEC and OASIS Collaborate on E-Business Standards-Standards Groups Increase Cross-Participation to Enhance Interoperability*. Retrieved from <https://www.oasis-open.org/news/pr/isoiec-and-oasis-collaborate-on-Business>
- Olson, L. (2008). *Synchronize UDDI registries with WebSphere Service Registry and Repository for better SOA governance*. IBM.
- Oracle. (2015a). *Using WS-Policy in ALSB Proxy and Business Services*. Oracle. Retrieved from http://docs.oracle.com/cd/E11036_01/alsb30/security/ws_policy.html
- Oracle. (2015b). *About Web Services Policy*. Oracle. Retrieved from http://docs.oracle.com/cd/E11036_01/alsb30/security/ws_policy.html
- SAP. (2015). *ALM103: Implementation Tools: SAP Business All-in-One and Rapid Deployment Solutions*. SAP.
- Şeref, R., Ahuja, K., & Winston, W. (2007). *Developing Spreadsheet-Based Decision Support Systems*. Michelle M.H. Dynamic Idea Publishing. Instructor Notes on DSS and data warehouse.
- Sreka, R., & Slaninova, K. (2012, April). *The Usability of Agent-Based Simulation in Decision Support System of E-Commerce Architecture*. *International Journal of Information Engineering & Electronic Bu*, 4(2), 10.
- Sugumaran, B. & Lavanya, M.R. (2014). *Evaluation of Critical Success Factors In Construction Projects*. *International Journal of Advanced Research in Civil, Structural, Environmental and Infrastructure Engineering and Developing*, 2(2).

Tan, W., Carter, A., & Toleman, M. (2009). *Implementing information technology services management: A case study focussing on critical success factors*. Toowoomba, Australia: Univ. of Southern Queensland.

The Open Group. (2014a). *The Open Group's Architecture Framework-Building blocks-Module 13*. Retrieved from www.open-group.com/togaf

The Open Group. (2014b). *The Open Group's Architecture Framework*. Retrieved from www.open-group.com/togaf

The Open Group. (2014c). *Risk Monitoring and Governance (Phase G)*. Retrieved from www.open-group.com/togaf

The Uniform Law Commissioners. (2015). *Electronic Transactions Act Summary*. The Uniform Law Commissioners.

Trad, A. (2015a). *A Transformation Framework Proposal for Managers in Business Innovation and Business Transformation Projects-Intelligent atomic building block architecture*. Centeris.

Trad, A. (2015b). *A Transformation Framework Proposal for Managers in Business Innovation and Business Transformation Projects-An information system's atomic architecture vision*. Portugal: Centeris.

Trad, A. (2015c). *The Transformation Framework's Resources Library*. IBISTM.

Trad, A. (2015d). *The Transformation Framework Proof of Concept*. IBISTM.

Trad, A., & Kalpić, D. (2014a). *The Selection and Training Framework (STF) for Managers in Business Innovation and Transformation Projects - The profile of an Architect of adaptive business systems*. IMRA.

Trad, A., & Kalpić, D. (2014b). *The Selection and Training Framework (STF) for Managers in Business Innovation and Transformation Projects- Integrating the restructuring process of the global economy*. Cambridge, UK: IMRA.

Trad, A., & Kalpić, D. (2014c). *The Selection, Modelling and Training Framework (SMTF) for Managers in Business Innovation and Transformation Projects-An executive's business architecture and modelling strategy*. *Journal of Business and Economics*, 5(1).

Trad, A., & Kalpić, D. (2015a). *The Business Engineering Transformation Framework for (e)commerce Architecture-Modelling Projects. Encyclopedia of E-Commerce Development, Implementation, and Management*. IGI-Global.

Trad, A., & Kalpić, D. (2015b). *The Business Engineering Transformation Framework for (e)commerce Control and Monitoring Pattern. Encyclopedia of E-Commerce Development, Implementation, and Management*. IGI-Global.

United Nations (2004). *Harmonized development of legal and regulatory systems for e-commerce in Asia and the Pacific: current challenges and capacity-building needs*. New York: UN.

Zulhuda, S. (2008). *The Fundamental Principles of E-Business Legal Framework*. Retrieved from <http://sonnyzulhuda.com/2008/01/16/the-fundamental-principles-of-e-business-legal-framework/>

ADDITIONAL READING

Farhoomand, A. (2004). *Managing Business Engineering transformation*. UK: Palgrave Macmillan.

Farhoomand, A., & Tsang, S. (2006). *Global Corporate Social Responsibility Vs. Local Legal Compliance-A Case of Internet Censorship in China*. Hong Kong: Asia Case Research Center, University of Hong Kong.

Trad, A., & Kalpić, D. (2015Submitted). *The Business Transformation Framework for Managers in Business Innovation Transformation Projects-An educational business engineering and transformation module*. USA.

Universität, St. Gallen (2015). EMBA in Business Engineering. <http://www.es.unisg.ch/de/programme/emba/embe.php>. Universität St. Gallen. Executive School of Management, Technology and Law (ES-HSG). Switzerland.

KEY TERMS AND DEFINITIONS

Business Engineering: It is the development and implementation of business solutions using business models to business processes and organizational structure.

CSFs: Can be used to manage the statuses and gaps in various project plans and gives the Projects the capacity to proactively and automatically recognize erroneous building blocks and to just-in-time reschedule the Project plan(s).

Environment: It is this research's framework.

Manager: Business Transformation Manager.

Project: Business Transformation Project.

Challenges of Meta Access Control Model Enforcement to an Increased Interoperability

Sérgio Luís Guerreiro
University of Lisbon, Portugal

INTRODUCTION

Today, countless access control models (ACM) solutions are available in the academy and industry. Nevertheless, the recognized development of ACM, in the majority of situations, these solutions specifies and implements the structural security access concerns of a single organizational silo (Sandhu *et al.*, 2000). Typically, ACM solutions are designed to fit and follow policies that are applied to a specific application layer of an organization. Early examples of such approach are the discretionary access control (DAC), mandatory access control (MAC), role-based access control (RBAC), time-role-based access control (TRBAC), Orcon or Chinese wall (Ferraiolo *et al.*, 2001; 2007).

Following this problem, Ferraiolo & Alturi (2008) raise the discussion about the feasibility of designing a meta access control model (MACM) fitting any specific ACM. So far, there are no bibliographic proofs that solve this posed problem. Moreover, Baker (2009) contributes to this discussion, stating the need to define a meta-ACM rather than specifying multiple instances in order to minimize the duplication of effort. Accordingly, to this author, the first goal to achieve in this endeavor is ACM conceptualization, and exemplifies it stating that RBAC is a particular instance of a MACM.

Moreover, recent advances proposed by Korman *et al.* (2016) show that the myriad of ACM solutions available difficult the management of IT models. Therefore, these authors propose an

ACM meta-model, designed in ArchiMate (The Open Group, 2013) to be used by the Enterprise Architects professionals. The meta-model is derived from the conceptual mapping of seventeen different ACM models. Then, ArchiMate relates the meta-model with enterprise concepts. In the end a unified meta-model for modeling authorization within enterprises is presented.

One the other hand, from a first sight, positioned in a different scientific field, interoperability is referred by Naudet *et al.* (2008) as “interoperability problem appears when two or more incompatible systems are put in relation”. In a broader sense, “Interoperability requirement is a statement that specifies a function, ability or characteristic, related to the capacity of a partner to ensure its partnership regarding compatibility, interoperation, autonomy, and reversibility, which it must satisfy” (Mallek *et al.*, 2012). Therefore, interoperability is considered as a key capacity to partners’ cooperation success (Patil *et al.*, 2007; Reul *et al.*, 2010). However, when two or more systems are interoperating, most of times, different ACM are in place and a barrier to the interoperation appears.

Therefore, given the context of ACM and interoperability, the following research question is logically raised: How to design and enforce a meta-access control model to facilitate the interoperability between different access control mechanisms?

For short, this paper assesses the possibility of using a meta-access control model to conceptual-

ize and instantiate many access control models. In specific, this research explores the following challenges:

- **Access Models Interoperability:** The paper uses a MACM to abstract all the concepts and relations contained in the many ACMs, and therefore creating interoperability between them.
- **Standardization of Storage for Access Data:** Standardization is realized through a single repository (and unique) for the MACM. When needed the MACM is instantiated for a specific ACM.
- **Provisioning of Access Models:** The MACM and ACM relationship enables the dynamic creation, reading, updating and deleting of access models, in order to adapt to the evolving organizational access requirements.

This paper has two-fold contributions: technological and societal. On the one hand, technological benefits are identified because of easier ACM implementation in each organization. On the other hand, societal benefits are related with lowering financial investments to interoperate the different organizations (*e.g.*: adaptive software enterprise resource planning (ERP) solutions between two small or medium enterprises).

This paper is organized as follow. Firstly, the research background is presented. Then, meta-access control model (MACM) is formalized. After that, solutions and recommendations are identified for the MACM. Then, future research directions are discussed. Finally, the last section concludes the paper.

BACKGROUND

Generically, a control system offers the capability to react whenever any disturbance affects the behavior of the controlled system or whenever a new reference is established (Guerreiro *et al.*, 2016).

Disturbance is assumed whenever the system is not producing the desired output for the imposed input. In these situations, the control system acts in the input, to change the controlled system's state.

Ferraiolo *et al.* (2007) defines that access control systems, or authorization in its broadest sense, is present in today's every information technology and is concerned with the ways in which users can access resources in the computer system, or informally speaking, with "who can do what". By these authors, access control is arguably the most fundamental and most pervasive security mechanism in use today. The author compares the actual access control models with the Guards, gates and locks that have been used since the ancient times to limit the individual's access to the valuables.

Nowadays, organizations use access control mechanisms to mitigate the risks of unauthorized access to their data, resources, software systems, etc. Each access request is assessed against a pre-defined authorization schema, and as a result, the access will be granted or denied. Several access control models exist to address changes in organizational structures, technologies, organizational needs, technical capabilities, and organizational relationships.

The standard NIST98 (Ferraiolo *et al.*, 2001; Sandhu *et al.*, 2000) is the most frequently used. It models the concepts for symmetric role-based access control (RBAC) to be used between the users, roles, permissions and constraints, as represented in Figure 1. It represents an evolution from the Discretionary Access Control (DAC) that grants access of individual object to individuals; Mandatory Access Control (MAC) and other policies due to less provisioning effort needed (Smith, 1997). Other known policies are different flavors of DAC, Time-Role Based Access Control (TRBAC), ORCON or Chinese wall. In RBAC the users are directly assigned to a role, each role has a set of associated permissions and changing the permissions affects the users associated with each role. Some well known constraints are separation of duties (SoD), conflict of interest (CoI), delegation of duties (DoD), binding of duty (BoD), least

privilege possible to the users, four-eyes-principle (2 tasks that must be executed by 2 users), history-based separation of duties (HsD). For example, the new constraints in the social networks: context constraints (Carminati *et al.*, 2009). The constraints are, in practice, applied to the relation between users/roles and roles/permission. Static constraints are offline computation and dynamic constraints are online computation. This model is applicable to organizational silos. However, it is limited to only one kind of organizational artifact, at a single time. NIST98 is broadly used in single architectural layers such as software applications or databases (Damiani *et al.*, 2002).

The RBAC model needs only to be made to role assignments, which are significantly fewer than individual assignments (Sandhu *et al.*, 1996). The RBAC model relies on user authentication, which in turn relies on identity management and defines relationships between the main concepts of Users, Roles and Permissions. RBAC's constraints restrict permissions depending on contextual information such as separation of duties (SoD) (Botha & Eloff, 2001).

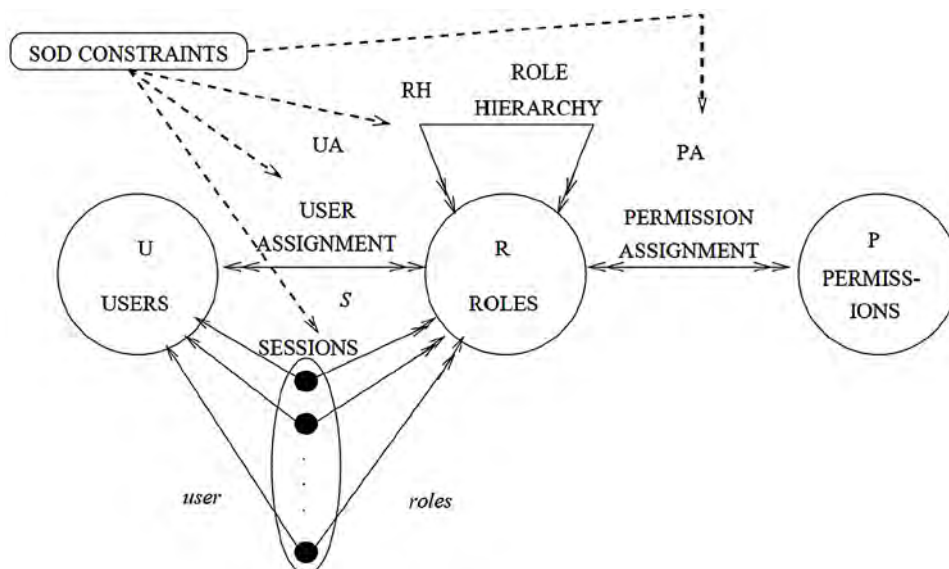
ABAC approach, as proposed by Wang *et al.* (2004) extends the RBAC capability granting the

accesses to the artifacts based on the attributes possessed by the requester. The advantage when compared with RBAC is the management of the roles and permissions that is not needed. The access configuration is specified by a set of rules that uses the attributes that are issued in the client request. A change in the access policy is performed by a simple change in the rules. Kuhn *et al.* states that ABAC is still evolving from previous existing RBAC definition instead of emerging from new knowledge (Kuhn *et al.*, 2010). Moreover, the distributed nature of the web services and the service oriented architecture (SOA) environments are identified as a potential application area for the development of the ABAC concepts (Shen & Hong, 2006).

XACML 3.0 (2014) is a standard document that defines a RBAC detailed protocol for the structural part of the access control models. However, it does not allow SoD constraint. It includes Digital Signature Services, and a method for conveying and applying data access policies and controls. Complementary to the XACML, Security Assertion Markup Language (SAML, 2016) allows business entities to make assertions regarding the identity, attributes, and entitlements

Figure 1. Symmetric RBAC with dynamic separation of duties (SoD) constraint

Source: Adapted from Sandhu *et al.*, 2000



of a subject (an entity that is often a human user) to other entities, such as a partner company or another enterprise application. In practice, SAML is complementary with XACML, where XACML specifies and SAML implements.

Moreover, in the scope of federated and single sign-on, Shibboleth (2016) is an application example for allowing the definition of a single authentication mechanism. It is used by federated architectures of different resources (*e.g.*, web, pc client, etc.). The advantage of such approaches is the ability to implement identity management systems in a distributed environment. However, this solution only solves the initial user authentication process and does not include concerns with fine-grained access control to the artifacts.

Meta Access Control Model Formalization

This section introduces the problematic of using a single access control model. Afterwards, a formalization of a meta-model to solve this problem is prescribed.

Problem Definition

Figure 2 presents the classical approach for ACM enforcement. One ACM is chosen and an access control repository implements it by software. This approach is able to fine-grained control the artifact's access of a software system, usually, the entrance in the software system (*e.g.*, login). It is also noticed that ACM is most of the times, decoupled from the enterprise design. Besides being the root cause of interoperability problem, the practical consequences of this decoupling are: (i)

the duplication of effort in the control and models design counterparts and (ii) the designed models not aligned with control requirements.

Although, because of simplicity, this approach is highly used by the industry and a domain conceptualization is also offered by the ACM community (Ferraiolo *et al.*, 2007) to support a common conceptual understanding. The key terms and definitions presented in the end of this paper, represents the set of consistent terminology that is referenced as relevant concepts and that is used in the ACM industrial enforcements (and are used in this paper from this point forward).

On the other hand, when complexity rises the conceptualization is a usual intellectual tool to analyze and synthesize a set of concepts and their relations. Today, the connected, and digital, enterprises present a high level of complexity demanding new forms to represent them. In this rationale, a conceptualization has the benefit of sharing a common understanding of concepts between stakeholders with different backgrounds and understandings, as explained and exemplified in Gaaloul *et al.* (2014). One straightforward solution to visualize concepts and its relations is using the IHMC CMaps tool (Grandry, 2013). From an ontological perspective, the ACM concepts from Table 1 are the first step to obtain a full ACM ontology. Guizzardi (2007) distinguishes an abstraction and its representation, and their relationship with conceptualization and representation language. Additionally, in this line of research, Kühne (2006) presents a full study in regard the relationships between models, meta models and languages.

For the problem at hand, Figure 3 depicts ACM as an abstraction that, usually, is represented by an

Figure 2. Access control model definition and implementation, the classical approach

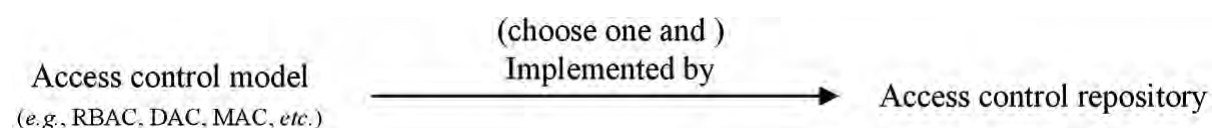
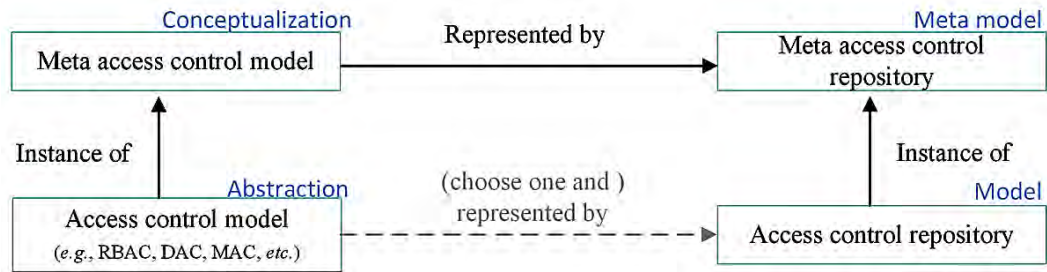


Figure 3. Meta access control model definition and implementation



access control repository (*cf.* Figure 2). Moreover, an abstraction is an instance of a conceptualization, in this paper proposed as a Meta ACM. By its turn, this Meta ACM is then represented by a Meta ACM repository. Finally, each access control repository is considered as a Model, which is an instance of a Meta ACM repository.

Therefore, our proposal is formalized by three stages: (i) the capacity to conceptualize a Meta access control model from a manifold of access control models (which are already implementation abstractions). Then, (ii) the representation of Meta access control model on a repository. Finally, (iii) the capacity to instantiate each access control model repository from the Meta access control repository.

SOLUTIONS AND RECOMMENDATIONS

This section identifies the challenges introduced by the previous section. Firstly, the proof of concept design is discussed, and then, the implementation challenges are discussed.

Proof of Concept Design Discussion

A proof of concept shows the applicability of a proposal within a specific domain. For exemplification purposes, a software system (an evolutionary prototype¹) to authorize queries on a graph database is considered. The objective of the prototype is to supply answers to queries formu-

lated in a previous provisioned graph composed by concepts and relations between the concepts. Before executing the query against the graph database the meta access control model components are responsible to check if the user is authorized.

The implementation follows the concepts proposed in Figure 3, where the ACM is supported by a XACML implementation. Then, MACM is specified using OWL. The consequence is that ACM is an instance of MACM. The OWL representation is then represented in a graph database to enforce the MACM repository (some alternatives are Neo4j (public available at <http://neo4j.com/>), orientDB (public available at <http://orientdb.com/orientdb/>), *etc.*). Afterwards, at run-time, when an access control mechanism is invoked, a transformation is applied to the MACM repository to implement the specific access control policy.

Implementation Challenges

Recalling the challenges posed in the introduction (i) Access models interoperability, (ii) Standardization of storage for access data and (iii) Provisioning of access models, this paper advances the following details as identified in proof of concept and summarized in Table 1.

FUTURE RESEARCH DIRECTIONS

As conceptualized in Figure 4, the meta access control model design, and supported in the biblio-

Figure 4. Meta access control model proof of concept

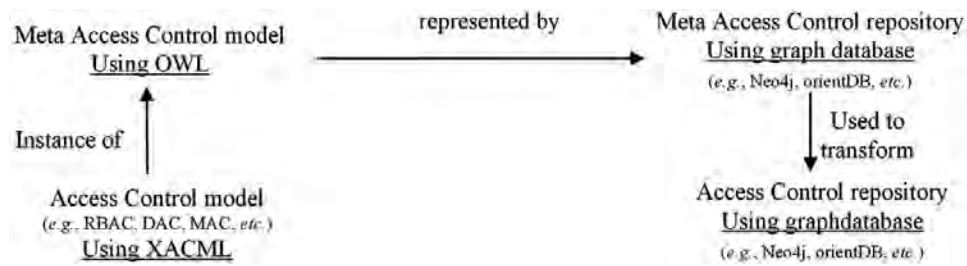


Table 1. The comparison between the challenges for access control models and meta-access control model

Challenges Posed	Access Control Models	Meta-Access Control Model
Access models interoperability	Casusically design with high level of implementation effort involved.	One model (MACM) supports all the access models, increasing the interoperability capability between organizations.
Standardization of storage for access data	ACM classical models do not encompass any type of standard between models.	Full supported. MACM supports an unique repository that is afterwards transformed into the desired ACM model.
Provisioning of access models	ACM classical models (namely the main standards) do not encompass any concern with provisioning. Each implementation needs to enforce its own provisioning mechanisms.	Similarly to ACM, the provisioning is also not specified. However, since MACM supports an abstract model, the provisioning follows a single provisioning mechanism.

graphic references, the following set of research challenges are raised:

1. *Implementing access control models (e.g., RBAC, DAC, MAC, ABAC, etc.) in XACML is still not consensual.*

XACML v3.0 (2014) Core and Hierarchical Role Based Access Control (RBAC) is the OASIS standard to enforce RBAC in XACML. The main challenges that are present in this standard are similar to the RBAC standard (Sandhu *et al.*, 2000): (i) the XML definition is verbose and complex; (ii) the interaction between the software systems specified in the standard are not industrial standardized; (iii) the policy administration, policy versioning, *etc.*, are not standardized; and (iv) the inability to enforce segregation (or by some literature is named as *separation*) of duties.

2. *How to design the mapping between XACML and OWL.*

This mapping problem is addressed by Finin *et al.* (2008a; 2008b). These authors identify a research thread of access control models defines as ontological models and exemplify the solution using RBAC. Two solutions are elicited: defining RBAC roles as OWL classes and a simpler approach defining RBAC roles as OWL values. As a result, the authors state that in the first solution: “the queries about a general class of access requests can be answered efficiently using a standard DL reasoner through subsumption reasoning”. However, both solutions have a problem with state change management due to the OWL monotonic nature. In practice, the role deactivations and role-permission assignment modifications must be performed outside the ontological reasoners.

Moreover, Helil & Rahman (2010) propose a solution to enforce RBAC XACML using OWL including the role hierarchy, the subject role assignment. To facilitate the RBAC policy definition and maintenance the OWL-DL is used.

3. *How to implement the correspondingly OWL into a graph database.*

Graph databases are derived from the scientific backgrounds of graph theory and mathematical logical and were firstly introduced around 1970 (Angles & Gutierrez, 2008). Despite the broad knowledge available from the last decades, it is recognized that recent advances in the scope of social networks has renewed its interest. Some examples of recent graph databases implementations are: Neo4j, orientDB, etc. OWL is a specification of a specific application domain written in a class/subclass oriented approach. The resulting graph database implementation could follow a schema definition or a schema-less approach. However, even in a schema-less approach, a simple definition of concepts and relationships are required. Therefore, when OWL is implemented into a graph database even a rudimentary schema is needed. The more refined is the graph database schema the more likely to fit the specific application domain, but the less applicable to similar applications. In the context of ACM, a balance between interoperability demand and support for many ACM models needs to be researched.

4. *How to transform MACM repository into an access control repository to be used at run-time by software application.*

From a non-ontological perspective, Hafeez *et al.* (2012) propose to use the SAML (2016) language to implement the interoperability using only the ABAC model. Consequently, this approach is only for the access models considered, and therefore, the solution is not scalable to others. With the same problem, De la Rosa Algarín *et al.* (2013) presents an implementation

of RBAC XACML using XML, in the scope of Health Level 7's Clinical Document Architecture and the Continuity of Care Record. Therefore, a wide-ranging solution for this problem demands a decoupling between the access control model itself and the capacity to store and interpret any access data. Recalling, Figure 3, the decoupling is designed by the MACM repository and its instantiation into a specific access control repository. For that end, the access control model needs to be referred in the instantiation event. Similarly, to an object oriented approach, when an object is instantiated with its specific data from a class.

5. *How to interoperate them?*

After all these identified research challenges, a question still remains, which is broader and is usually associated with, the management information systems scope (Laudon & Laudon, 2016). How is it possible to integrate into a unique enterprise system design all the previous four research items? In addition, how to use these concepts to enforce systems interoperability, considering the different technologies from disparate vendors that are in place? Moreover, how to enforce access control interoperability in organizational relationships that are semi-stable? From a lower abstraction perspective, the question of how to intersect data from different access control models stored in the same MACM repository is posed.

CONCLUSION

This paper addresses the context of access control models combined with interoperability between information systems. The research question of “*How to design and enforce a meta access control model to facilitate the interoperability between different access control mechanisms?*” uncover a complex problem demanding integrated efforts from different scientific areas, but where ontological design is proposed as a key competence to ground a future solution. In this sense, a meta

access control model design is experimented to systematize future research. Therefore, we identify the main challenges that are associated with bibliographic references identified.

To sum up, the challenges posed are related as follows: Access models interoperability, Standardization of storage for access data and Provisioning of access models. A full list of future work is presented in section future research directions.

REFERENCES

- Angles, R., & Gutierrez, C. (2008). Survey of graph database models. *ACM Computing Surveys*, 40(1), 1–39. doi:10.1145/1322432.1322433
- Barker, S. (2009). The next 700 access control models or a unifying meta-model? In *Proceedings of the 14th ACM symposium on Access control models and technologies* (pp. 187-196). ACM. doi:10.1145/1542207.1542238
- Botha, R. A., & Eloff, J. H. P. (2001). Separation of duties for access control enforcement in workflow environments. *IBM Systems Journal*, 40(3), 666–682. doi:10.1147/sj.403.0666
- Carminati, B., Ferrari, E., & Perego, A. (2009). Enforcing access control in web-based social networks. *ACM Transactions on Information and System Security*, 13(1), 6. doi:10.1145/1609956.1609962
- Damiani, E., De Capitani di Vimercati, S., Paraboschi, S., & Samarati, P. (2002). A fine-grained access control system for XML documents. *ACM Transactions on Information and System Security*, 5(2), 169–202. doi:10.1145/505586.505590
- De la Rosa Algarín, A., Ziminski, T. B., Demurjian, S. A., Sánchez, Y. K. R., & Kuykendall, R. (2013). Generating XACML Enforcement Policies for Role-Based Access Control of XML Documents. In *International Conference on Web Information Systems and Technologies* (pp. 21-36). Springer Berlin Heidelberg.
- Ferraiolo, D., & Atluri, V. (2008). A meta model for access control: why is it needed and is it even possible to achieve? In *Proceedings of the 13th ACM symposium on Access control models and technologies* (pp. 153-154). ACM. doi:10.1145/1377836.1377860
- Ferraiolo, D. F., Kuhn, R. D., & Chandramouli, R. (2007). *Role-Based Access Control* (2nd ed.). Artech House.
- Ferraiolo, D. F., Sandhu, R., Gavrila, S., Kuhn, D. R., & Chandramouli, R. (2001). Proposed NIST standard for role-based access control. *ACM Transactions on Information and System Security*, 4(3), 224–274. doi:10.1145/501978.501980
- Finin, T., Joshi, A., Kagal, L., Niu, J., Sandhu, R., Winsborough, W., & Thuraishingham, B. (2008a). R OWL BAC: representing role based access control in OWL. In *Proceedings of the 13th ACM symposium on Access control models and technologies* (pp. 73-82). ACM. doi:10.1145/1377836.1377849
- Finin, T., Joshi, A., Kagal, L., Niu, J., Sandhu, R., Winsborough, W., & Thuraishingham, B. (2008b). Role based access control and OWL. In *Proc. 4th Int. Workshop on OWL: experiences and directions*.
- Gaaloul, K., Guerreiro, S., & Proper, H. A. (2014). Modeling access control transactions in enterprise architecture. In *2014 IEEE 16th Conference on Business Informatics* (Vol. 1, pp. 127-134). IEEE.
- Grandry, E., Feltus, C., & Dubois, E. (2013). Conceptual integration of enterprise architecture management and security risk management. In *2013 17th IEEE International Enterprise Distributed Object Computing Conference Workshops* (pp. 114-123). IEEE. doi:10.1109/EDOCW.2013.19
- Guerreiro, S., Marques, R. P., & Gaaloul, K. (2016). Optimizing Business Processes Compliance Using an Evolvable Risk-Based Approach. In *Proceedings of the 2016 49th Hawaii International Conference on System Sciences (HICSS)* (pp. 5652-5661). IEEE Computer Society. doi:10.1109/HICSS.2016.699

- Guizzardi, G. (2007). On ontology, ontologies, conceptualizations, modeling languages, and (meta) models. *Frontiers in Artificial Intelligence and Applications*, 155, 18.
- Hafeez, K., Rajpoot, Q., & Shibli, A. (2012). Interoperability among access control models. In *Multitopic Conference (INMIC), 2012 15th International* (pp. 111-118). IEEE. doi:10.1109/INMIC.2012.6511466
- Helil, N., & Rahman, K. (2010). Extending XACML profile for RBAC with semantic concepts. In *2010 International Conference on Computer Application and System Modeling (ICCASM 2010)* (Vol. 10, pp. V10-69). IEEE.
- IAM. (2016). *Identity and authority management*. Retrieved from <http://shibboleth.internet2.edu/>
- Korman, M., Lagerström, R., & Ekstedt, M. (2016). Modeling Enterprise Authorization: A Unified Metamodel and Initial Validation. *Complex Systems Informatics and Modeling Quarterly*, (7), 1-24.
- Kuhn, D. R., Coyne, E. J., & Weil, T. R. (2010). Adding attributes to role-based access control. *IEEE Computer*, 43(6), 79-81. doi:10.1109/MC.2010.155
- Kühne, T. (2006). Matters of (meta-) modeling. *Software & Systems Modeling*, 5(4), 369-385. doi:10.1007/s10270-006-0017-9
- Laudon, K., & Laudon, J. (2016). *Management Information Systems* (14th ed.). Prentice Hall.
- Mallek, S., Daclin, N., & Chapurlat, V. (2012). The application of interoperability requirement specification and verification to collaborative processes in industry. *Computers in Industry*, 63(7), 643-658. doi:10.1016/j.compind.2012.03.002
- Naudet, Y., Latour, T., & Chen, D. (2008). A systemic approach to interoperability formalization. IFAC WC 2008, invited session on Semantic-Based Solutions for Enterprise Integration and Networking, Seoul, Korea. doi:10.3182/20080706-5-KR-1001.02018
- Patil, V., Mei, A., & Mancini, L. V. (2007). Addressing interoperability issues in access control models. In *Proceedings of the 2nd ACM symposium on Information, computer and communications security* (pp. 389-391). ACM. doi:10.1145/1229285.1229337
- Reul, Q., Zhao, G., & Meersman, R. (2010). Ontology-based access control policy interoperability. In *Proc. 1st Conference on Mobility, Individualisation, Socialisation and Connectivity, MISC*.
- SAML. (2016). *Specification*. Retrieved from http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=security
- Sandhu, R., Ferraiolo, D., & Kuhn, R. (2000). The NIST model for role-based access control: towards a unified standard. ACM Workshop on Role-Based Access Control. doi:10.1145/344287.344301
- Sandhu, R. S., Coynek, E. J., Feinstein, H. L., & Youmank, C. E. (1996). Role-based access control models yz. *IEEE Computer*, 29(2), 38-47.
- Shen, H. B., & Hong, F. (2006). An attribute-based access control model for web services. In *2006 Seventh International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT'06)* (pp. 74-79). IEEE. doi:10.1109/PDCAT.2006.28
- Smith, C. L. (1997). A survey to determine federal agency needs for a role-based access control security product. In *Software Engineering Standards Symposium and Forum, 1997. Emerging International Standards. ISESS 97., Third IEEE International* (pp. 222-232). IEEE. doi:10.1109/SESS.1997.595975
- The Open Group. (2013). *ArchiMate 2.1 Specification, Technical Standard*. Van Haren Publishing. Retrieved from <http://www.opengroup.org/archimate/>

Wang, L., Wijesekera, D., & Jajodia, S. (2004). A logic-based framework for attribute based access control. In *Proceedings of the 2004 ACM workshop on Formal methods in security engineering* (pp. 45-55). ACM. doi:10.1145/1029133.1029140

XACML v3.0. (2014). *Core and Hierarchical Role Based Access Control (RBAC) Profile Version 1.0*. Retrieved from <http://docs.oasis-open.org/xacml/3.0/rbac/v1.0/cs02/xacml-3.0-rbac-v1.0-cs02.html>

ADDITIONAL READING

Gaaloul, K., Guerreiro, S., & Proper, H. A. (2014). Modeling access control transactions in enterprise architecture. In 2014 IEEE 16th Conference on Business Informatics (Vol. 1, pp. 127-134). IEEE.

Guédria, W. (2014). A conceptual Framework for Enterprise Interoperability. *International Journal of E-Business Research*, 10(3), 54-64. doi:10.4018/ijebr.2014070104

Guédria, W., Naudet, Y., & Chen, D. (2013). MMEI: Maturity Model for Enterprise Interoperability, *Enterprise Information Systems journal*, Taylor & Francis. Johnson, P., Vernotte, A., Ekstedt, M. & Lagerström, R. (2016). pwnPr3d: an Attack-Graph-Driven Probabilistic Threat-Modeling Approach, *Proc. of the International Conference on Availability, Reliability and Security (ARES)*, Sept.

Korman, M., Lagerström, R., & Ekstedt, M. (2015). Modeling Authorization in Enterprise-wide Contexts. In The 8th IFIP WG 8.1 working conference on the Practice of Enterprise Modeling (PoEM 2015). (Vol. 1497, pp. 81-90). CEUR-WS.

Lagerström, R., Sommestad, T., Buschle, M., & Ekstedt, M. (2011). Enterprise architecture management's impact on information technology success. In System Sciences (HICSS), 2011 44th Hawaii International Conference on (pp. 1-10). IEEE.

Slimani, N., Khambhammettu, H., Adi, K., & Logrippo, L. (2011). UACML: Unified access control modeling language. In New Technologies, Mobility and Security (NTMS), 2011 4th IFIP International Conference on (pp. 1-8). IEEE.

Stallings, W., & Brown, L. (2012). *Computer Security: Principles and Practice* (2nd ed.). Pearson Education.

KEY TERMS AND DEFINITIONS

Identity: Identity is the fundamental concept of uniquely identifying an object (person, computer, etc.) within a context. That context might be local (within a department), corporate (within an enterprise), national (within the bounds of a country), global (all such object instances on the planet), and possibly universal (extensible to environments not yet known). Many identities exist for local, corporate, and national domains. Some globally unique identifiers exist for technical environments, often computer-generated.

Object: Any resource accessible on a computer system, including files, peripherals such as printers, databases, and fine-grained entities such as individual fields in database records. Objects are traditionally viewed as passive entities that contain or receive information, although even early access control models included the possibility of treating programs, printers, or other active entities as objects.

Operation: It is an active process invoked by a subject. Early access control models that were concerned strictly with information flow (i.e., read-and-write access) applied the term subject to all active processes, but RBAC models require a distinction between subject and operation. For example, when an ATM user enters a card and correct PIN, the control program operating on the user's behalf is a subject, but the subject can initiate more than one operation of deposit, withdrawal, balance inquiry, or others.

Permission (or Privileges): They are authorizations to perform some action on the system. In computer security literature, the term permission refers to some combination of object and operation. A particular operation used on two different objects represents two distinct permissions, and similarly, two different operations applied to a single object represent two distinct permissions. For example, a bank teller may have permissions to execute debit and credit operations on customer records, through transactions, while an accountant may execute debit and credit operations on the general ledger, which consolidates the bank's accounting data.

Session: An instance of a user's dialog with a system is called a session.

Subject: A computer process acting on behalf of a user is referred to as a subject. Note that in reality, all of a user's actions on a computer system are performed through some program running on the computer. A user may have multiple subjects in operation, even if the user has only one login and one session. For example, an e-mail system may

be operating in the background, fetching e-mail from a server periodically, while the user operates a Web browser. Each of the user's programs is a subject, and each program's accesses will be checked to ensure that they are permitted for the user who invoked the program.

User: Refers to people who interface with a computer system. In many designs, it is possible for a single user to have multiple login IDs, and these IDs may be simultaneously active. Authentication mechanisms make it possible to match the multiple IDs to a single human user.

B

ENDNOTE

- ¹ An evolutionary prototype is a software system that has the ability to support new developments. In the opposite, a disposable prototype is only used to requirements validation and do not have the ability to support new developments.

Cognitive Ergonomics in 2016

Ronald John Lofaro

Embry-Riddle Aeronautical University, USA

INTRODUCTION

Cognitive Ergonomics deals with, amongst other areas, decision-making, skilled performance(s) and training. This chapter will briefly explicate what began, 30 years ago as a cognitive task analysis (CTA) for identifying critical skills needed for skilled performance, with decision-making involved in every step. Over the years, it evolved into a Delphi paradigm that extended far beyond a CTA. With the advances in technology, especially communication technologies, some of this paradigm has re-emerged in what are termed Estimate-Talk-Estimate (ETE) Delphi efforts. This Chapter will show how to integrate the elements of the paradigm cited above with elements of Pareto Analysis and ETE. The resultant model uses today's technology to achieve extremely accurate results across many venues. Thus, this Chapter is a natural progression/extension of the author's Chapter entitled "Knowledge Engineering: A Methodology and Examples" (Lofaro, R.J., 2013) in the 3rd edition of *The Encyclopedia of Information Science and Technology* (2014). In that Chapter the author presented both the Small Group Delphi (Lofaro, 1992) as developed and expanded over the years and the newer Estimate-talk-Estimate variant of the Delphi process. At that point, elements of both Delphi processes were blended to produce an initial look, with some methodological avenues, for use in a paradigm in tune with the tremendous speed, instantaneity and scope of today's technological advances in communication. This Chapter will take that rough template and integrate it with elements of Pareto techniques, to initially identify and sharpen criticality of possible problems/solutions of issues. The selection and the rationale for the selection

of the elements of both the Pareto analysis and for the elements of the SGDP will be explicated. At that point, the new Pareto/SGDP/ETE process could be implemented to achieve guidelines for achievable resolution of a variety of problems requiring accurate decision-making. The decision-making will be of a distributed nature, enabled by computers and the Internet. Further, this Chapter will examine three current, problematic areas and outline how this now tri-partite paradigm can be used on these areas to show how decisions can be made on standards, policy and training needs

BACKGROUND

Knowledge engineering (KE) has been defined as follows: "...an engineering discipline that involves integrating knowledge into computer systems in order to solve complex problems normally requiring a high level of human expertise." (Feigenbaum and McCorduck, 1983). For a succinct overview of KE, see Studer, Benjamins and Fensel (1998). Knowledge engineering is also linked to cognitive science and socio-cognitive engineering where the knowledge is produced by socio-cognitive aggregates (mainly humans); this was one rationale for the SGDP. A newer term, cognitive engineering (CE), includes mental workload, decision-making, skilled performance, human-computer interaction, human reliability, work stress and training as these may relate to human-system design. There is not only great overlap with KE but, almost an isomorphism. Therefore, CE has mainly replaced KE as the term used in such efforts. Cognitive Ergonomics deals with decision-making and focuses on the fit between human cognitive abilities and limitations and the task. As such, it can be

DOI: 10.4018/978-1-5225-2255-3.ch057

viewed as a portion of a Venn diagram Universe that includes Cognitive Engineering, again with definite overlapping.

Delphi Processes

A subset of CE is the Delphi technique/process. Traditional Delphi techniques include anonymity of response, multiple iterations, convergence of the distribution of answers and, a statistical group response (Judd, 1972). A seminal paper on the Delphi process was written by a then-Rand Corporation employee (Brown, 1968) and may be available from Rand or from American Society of Tool and Manufacturing Engineers (ASTME), *now known as Society of Manufacturing Engineers*. It would seem clear that Cognitive Ergonomics can avail itself of Delphi processes

The Small Group Delphi Paradigm

A modification to Delphi processes is the small group Delphi paradigm (SGDP). The SGDP took the Delphi process in another direction by modifying it via merger with elements of group dynamics in order to have interactive (face-to-face) Delphi workshops. The development of this modified Delphi, the SGDP, involved the merger of a specific knowledge engineering technique (Delphi), with Fleischmann's theories of underlying abilities (Fleishmann and Quaintance, 1984; revised 2000) and some principles of group dynamics. This modification resulted in a paradigm for using small groups of subject matter experts (SMEs) for any project that requires that a set of SMEs be used to identify, evaluate, and critically rank tasks (an extended and enhanced task analysis), identify core needs/skills, recommend modifications to equipment, procedures and training. Finally, the SGDP can be used to sharpen, modify and revise existing methodologies. Thus, traditional Delphi processes were modified into a new paradigm (Lofaro, 1992). This effort and the subsequent use (Lofaro, Gibb and Garland, 1994; Gibb and Lofaro, 1994; Lofaro, 1998, 1999) of the resultant

paradigm, SGDP, in other, varied venues have produced both new modifications to the SGDP and highly accurate data that were operationally implemented. The extensions of the paradigm indicated that it had an applicability over many domains. These many SGDP efforts also resulted in sharpening, modifying and revising the original methodology. That the core SGDP has been used in many environments demonstrated a robust flexibility and generalizability of the paradigm. As Meister (1985) had noted, "The (Delphi) methodology is by no means fixed...[it] is still evolving and being researched." This is as true now as it was when Meister stated it. In point of fact, with the leaps in communication methods and related technology, even more so.

CURRENT ADVANCES IS DELPHI PROCESSES AND A NEXT STEP

The use of face-to-face Delphi techniques has been re-discovered. It is important to state that, circa 2005, the use of face-to-face groups in a Delphi has now become accepted. New technologies have resulted in what are generically referred to as mini-Delphi or Estimate-Talk-Estimate (ETE) with many variations. Other innovations come from the use of computer-based (and later web-based) Delphi conferences. One example of a difference in a type of ETE (a computer-based Delphi) versus either a traditional or SGDP Delphi is the iteration structure used in the traditional or SGDP Delphis, those iterations which are divided into three or more discrete rounds, can be replaced by a process of continuous (roundless) interaction, enabling SMEs to change their evaluations at any time. As Bolognini (2001) has said, the involvement of a large number of participants, the use of two or more panels representing different groups (such as policy-makers, experts, citizens), which the administrator can give tasks reflecting their diverse roles and expertise, and have them interact within ad hoc communication structures: these are strengths of computer and web-based ETEs. The

reader is referred to the work of Turoff and Hiltz (1996) on computer-based Delphis.

While the author was heartened to see his seminal concepts (the coupling of traditional Delphi methods with group dynamics and face-to-face sessions) have become accepted, a question arose: What now? The author believes that another process for analysis needed to be front-loaded into any SGDP/ETE; elements of a Pareto analysis. That is to say, the author believes the next step is to combine elements of a Pareto technique at the front end of an SGDP/ETE process, resulting in a new paradigm. The questions to be answered are which elements of Pareto analysis are to be used (with rationale) as well as which elements of a SGDP are to be incorporated, again with rationale? We already have ETE-type Delphis done using web sites where the process is conducted in real-time. The ETE type would not involve selecting elements thereof but would be chosen on the basis of what data came out of the Pareto/SGDP. Re-stated: As to which type of ETE would become part of this new paradigm, that is dependent on what elements of the Pareto and SGDP are to be part of the new paradigm. Lastly, a demonstration will be given of how this new model would be applied to current, real-life problems in three very different venues

What Are the Elements of a Pareto Analysis to Be Used and Why?

The Pareto would be step one because it would identify problems, then, sharpen (focus in on the ones that are amenable to resolution) while, at the same time, providing an ordering of which should be worked on via criticality ratings. The elements of a Pareto analysis would be the first three steps in such an analytic technique. Step 1: Identify and list problems to be examined. Or, if and only if one problem/issue exists, a break-out of the sub-problems could be done. Step 2: identify causal factors inherent in each problem. Step 3: Score (in this case, criticality rate) the problems. The Pareto, done in the context of a SGDP/ETE could also allow for many possible solutions to be

evaluated by many types of SMEs with differing areas of expertise but areas germane to the problem. This would result in a winnowing down of courses of action to those that were realistic and possible of success. The advantages of using a Pareto, SGDP and an ETE (all computer/Internet driven) are that they are all content-area neutral and, in a real sense, generic in application.

What Core Elements of the SGDP Are to Be Used and Why?

Briefly, what are the core elements of a SGDP and which should be/can be used in an ETE Delphi that is computer-based and web-driven? Additionally, why are these particular core elements selected? What follows is a very short synopsis of SGDP core elements because of space constraints. For a more complete SGDP write-up, email the author; *lofaro@msn.com*.

1. **SME Selection:** Identify the best SMEs for the SGDP Workshop.
2. **The SGDP Objectives:** Develop the objectives and sub-objectives so that all the components needed to achieve the whole objective will be in place.
3. **The Read-Ahead Package:** It must be prepared and sent to the Workshop participants at least three weeks in advance; longer if there is time. The read-ahead should include: purpose and rationale for the workshop, the use-value of the data generated by the SGDP workshop; the objectives of the workshop and, any longer-range goals that the workshop will enable; [a clear statement that small-group processes will be used and that group dynamics training will be the first step in the process---and that a trained, experienced facilitator will be doing this training and will be on-site for the duration along with a "homework" assignment to be completed and brought to the workshop. Finally, included is a day-by-day Agenda of sorts.

Along with the read-ahead, a complete set of protocols is to be completed and given to the participants when they arrive.]

4. **The SGDP Process:** Upon participant arrival, proceed with instructions and exercises in group dynamics, “groupthink,” functional/dysfunctional groups and a definition/example of real consensus.

Simply put: Numbers 1,2 and 3 along with a definition/example of real consensus. As to Number 3, the portion in parentheses does not apply. Note that 2 and 3 essentially only one component; doing 3 incorporates 2. These elements were selected because they can apply to any problem/issue and would be the foundation of any Pareto/SGDP/ETE.

General Methodology for Pareto/SGDP/ETE

All participants can be logged on simultaneously, each participant can briefly state their name and credentials, the group dynamics instruction can be done by the facilitator to all simultaneously (aside: it would seem that a linked network of all SMEs is possible and even *de rigueur*. This will allow for instant feedback by any SME during a session, as well as discussions). The iteration structure used in SGDP, which is divided into as many discrete rounds as needed for consensus, can be replaced by a process of continuous (roundless) interaction enabling SMEs to change their evaluations at any time and give a rationale with ensuing discussion in real-time. Finally, the statistical group(s) response(s) can be updated in real-time and shown whenever a SME, or a group, provides a new evaluation. It is clear that “face-to-face” discussion will be virtual. This is, to the author, a real and significant loss. But, the speed, multiple iterations, real-time, access to a large number of SMEs and other aspects to be gained cannot be ignored. Another possible modification is a multi-tiered Pareto and then, SGDP where

the use of two or more groups working different problems can be convened and given objectives based on their expertise. As these groups come to consensus on their objectives, these new data can be integrated, built into a new re-ahead package and made available to a new SGDP set with new or prior SME’s

EXAMPLES OF USING AN SGDP/ETE WITH PARETO TECHNIQUES: 2016 AND BEYOND

An initial thought, with the author’s background in education (mathematics degree; taught mathematics and math ed) was the current and it seems now controversial United States federal Common Core Standards. While these Standards were federally developed (actually done under contracts), the States were to develop (or not) curricula based on these standards. One set of issues seem to revolve around whether the Common Core Standards are an attempt by the federal government, through federally-funded interstate consortia assessment entities, to make the States hew to the Standards. This would make the Standards, *de facto*, mandatory at the State level. Another issue is the critiques of the Standards themselves, especially it seems the Mathematics and English ones, by educators, scholars and parent interest groups. At the federal level, it seems the Standards are a *fait accompli*, thus precluding any Pareto/SGDP/ETE efforts. However, at the State level, such Delphis could help in analyses of the Federal Standards; development of State curricula and pedagogy for teaching the new State curricula.

Such an effort would involve multiple subject area-specific groups, each composed of educators (classroom teachers as well as University teacher education professors), parents and some legal experts. The goal would be a State-wide curricula, with assessment techniques designed to both accurately evaluate student progress and needs as well as pass muster with the Smarter Balanced Assessment Consortia (SBAC) and the Partnership for

Assessment for Colleges and Careers (PARCC); these are the federally funded assessment consortia mentioned above. While this seems to possibly be a moot effort as there are federal Standards, some States (Oklahoma, Texas, Virginia, Alaska, Nebraska and Indiana) have already indicated they will not use the federal Standards and Minnesota will not use the federal Mathematics Standards. So, it seems that such States might support such Pareto/SGDP/ETE efforts.

The author selected, as another exemplar, the aviation field in which he worked for over 22 years: What are some current 2015 aviation issues (as well as prior but unresolved ones) that are both vital to aviation safety and amenable to some form of the Pareto/SGDP/ETE?

In view of current and recent fatal accidents, a look at air carrier upset training may be fruitful. Such training was brought to the fore by the Colgan Air accident with 50 fatalities. There was the Air France flight 447 (229 lives lost) in which the pilots did not recognize that they were in a high speed, high altitude stall; these accidents led to some attention/training being devoted to this issue. We now have AirAsia flight 8501, crashed with 162 souls on board and Malaysia flight MH 370, still unfound with 239 lives lost. At least the AirAsia flight may be a scenario similar to the Air France accident.

In direct response to the Colgan crash, Congress passed the Airline Safety and FAA Extension Act of 2010, which mandated that the Federal Aviation Administration (The FAA regulates aviation in the USA, from air traffic to civil aviation security, the operation of air carriers, pilot training and more) require pilots to complete 1,500 flight hours before they're allowed to fly commercially, up from just 250 hours before the act. This new rule may do little to improve safety. Both pilots involved in the Colgan crash had far surpassed 1,500 hours of flight time, so that requirement probably had little to no impact on the accident. However, the FAA, while not yet issuing an Advisory Circular (AC) or a Federal Aviation Regulation (FAR), has issued a document called Airline Upset Recovery

Training Aid, version 2. There seems to be some problems with using a hexapod/full-motion flight simulator (FS) that will be part of an expected FAA pilot training rule by 2018 (Croft, John. 2014a). There are also issues about in-aircraft training and swept wing jet aircraft specialized training. There is some controversy also involving the American Airlines UPRT ground school with FS training, called advance aircraft maneuvering program (AAMP). This training was seen by The National Transportation Board (NTSB) as possibly a contributing factor in the American Airlines flight 587/A300-600 crash in November, 2001 (Croft, John. 2014b). The area outlined may seem fruitful but it is fraught with regulatory and legal problems. That having been said, since the upset training is in the FAA rule-making arena, any Pareto/SGDP/ETE efforts would probably be moot, albeit both germane and instructive. Yet, there seems to be a definitive need for such efforts. However, we go to another problem area which may be more amenable and less encumbered.

A second problem area is of long standing and, has been often researched: flight crew fatigue. Fatigue from flying has been an issue since World War I, yet no resolution has been found. A fairly recent example is the United Parcel Service (UPS) flight 1354 crash in Birmingham, Alabama, in late 2013. At this time, it seems that the UPS pilots, NTSB and Airbus (the type aircraft involved was an Airbus A300-600), are in, to put it mildly, some disagreement as to causation. Remember that UPS has an FAA-managed aircrew fatigue risk management program (Croft, John, 2014c). If memory serves, (the author was with the FAA at the time), the FAA, the National Aviation and Aerospace Administration (NASA) and United Air Lines were involved in a study on the fatigue effects of flying "long haul," Trans Pacific flights; these were sometimes 15 or more hours long. As one result, the FAA proposed a draft AC called Controlled Rest on the Flight Deck; the aircarrier industry opposed it and it was shelved. Here we are, in 2015/6 and aviation still is working on this issue while lives are being lost.

Thus let us look at aircrew fatigue. What would be the design and process for a SGDP/ETE to work this safety issue?

The SME's must include, in the main, current aircarrier pilots; both U.S. flag-carriers and regional air carriers pilots must be in any/every working group. The goal of such an effort would be to develop a set of recommendations for an FAA FAR on dealing with aircrew fatigue as a recognized safety problem, with rationale. The FAA rule-making process is long and complex. It does have methods for aviation industry input as well as aviation industry objections/agreement. With that in mind, each group in this effort must have an FAA attorney as a member. The material for every group should include accident investigation data (for the prior 30 plus years) where aircrew fatigue was identified as a causal or contributing factor. This would also entail having an NTSB accident investigator/analyst on each group. The general mythology to be used has been shown in a prior section.

Another seemingly obvious candidate for the SGDP methodology is Operational Testing (OT); specifically OT&E (Operational Testing & Evaluation). The goal of any OT is to determine the system's capability to perform its mission in the operational setting and to both determine and evaluate core operational effectiveness and suitability problems (Maliko-Abraham, Helene and Lofaro, R.J., 2003). Operational tests and evaluations are conducted on production, or production-representative articles, to determine whether systems are operationally effective and suitable. Such testing entails operational testing of production representative test articles and uses typical operational scenarios that are as realistic as possible. It is here that the Pareto/SGDP/ETE can be modified so that SMEs can use it in OT.

Dumas and Redish (1999) indicate that to be a good OT testing scenario, it needs to be short, developed in the end-user's terminology and, so clear that all participants using the scenario will understand it. A small set of carefully selected

SMEs would be used to face-validate the existing scenarios. This SME set also would ensure that all the necessary operational issues were embedded in the scenarios. A second group of SMEs would also assist in the test sequencing as well as the techniques/scalings to be used in workload analysis. While the process of multiple SGDP workshops may seem lengthy, it is not the case. The SGDP/ETE sessions would run consecutively, each new one beginning with the data from the prior one--and each session should need approximately 4 days to complete. The experience and insight as to the realities of the operational arena of the SMEs are both needed and invaluable. Upon completion of the actual OT, all SME members of the two previous sessions would be convened to interpret the results and to make a set of recommendations. These recommendations may include retrofit, modification, and training. The two SGDP groups would work independently at first. Then, as they finished, they would convene into one group to finalize their results.

FUTURE RESEARCH DIRECTIONS

The Future Research Directions are close to the Conclusions, being inter-twined: the tri-partite model shown above needs to be implemented; perhaps as a University or think-tank project... even as a government (at some level) one. The results, pro or con, need to be published and, as always, suggested modifications need to be listed, with rationale, for any future work.

CONCLUSION

The tri-partite model explicated above, with exemplar uses, offers significant advances in diversified, accurate decision-making needed over a broad spectrum of problem areas; defining and explicating standards as well as the training needed to achieve resolution.

REFERENCES

- Bolognini, M. (2001). *Electronic Democracy. Delphi Method and Public Policy-Making*. Rome: Carocci Editore.
- Brown, B. B. (1968). Delphi Process: A Methodology for the Elicitation of Opinions of Experts. Santa Monica, CA: The Rand Corporation.
- Croft, J. (2014a). Unambiguous Upset. *Aviation Week & Space Technology*, 176(20), 52.
- Croft, J. (2014b). Rest Assured. *Aviation Week & Space Technology*, 176(18), 38.
- Dumas, J. S., & Redish, J. C. (1999). *A Practical Guide to Usability Testing* (revised ed.). Portland, OR: Intellect Books.
- Feigenbaum, E. A., & McCorduck, P. (1983). *The Fifth Generation* (1st ed.). Reading, MA: Addison-Wesley.
- Fleishman, E. A., & Quaintance, M. K. (1984). Taxonomies of human performance: The description of human tasks. Management Research Institute, Inc.
- Gibb, G. M., & Lofaro, R. J. (1994). *Airport Security Screener and Checkpoint Security Supervisor Training Workshop*. DOT/FAA/CT-94/109. National Technical Information Service.
- Judd, R. L. (1972). *Use of Delphi Methods in Higher Education in Technological Forecasting and Social Change*, 4. Reed Elsevier.
- Lofaro, R.J. (1992). A Small Group Delphi Paradigm. *The Human Factors Society Bulletin*, 35(2), 1-4.
- Lofaro, R. J. (1998). Identifying and Developing Managerial and Technical Competencies: Differing Methods and Differing Results. Proceedings of 42nd annual Human Factors and Ergonomics Society.
- Lofaro, R. J. (1999). Core Competencies and Individual Differences: Chasm or Linkage? Human Factors and Ergonomics Society/HF&ES Individual Differences in Performance News, 12(2).
- Lofaro, R. J. (2013). Knowledge Engineering: A Methodology and Examples. In *Encyclopedia of Information Science and Technology* (3rd ed.). IGI-Global.
- Lofaro, R. J., Gibb, G. M., & Garland, D. (1994). *A Protocol for Selecting Airline Passenger Baggage Screeners*. DOT/FAA/CT-94/110. Springfield, VA: National Technical Information Service.
- Lofaro, R. J., & Intano, G. P. (1989). Exploratory Research and Development: Army Aviator Candidate Classification By Specific Helicopter. In *Proceedings of 5th International Symposium of Aviation Psychologists*. ISAP.
- Maliko-Abraham, H., & Lofaro, R.J. (2003). Usability Testing: Lessons Learned and Methodology. *Proceedings of the 2003 International Military Testing Association (IMTA) Conference*.
- Meister, D. (1985). *Behavioral Analysis and Measurement*. Wiley Press.
- Studer, R., Benjamins, V., & Fensel, D. (1993). Knowledge Engineering, Principles and Methods. *Data & Knowledge Engineering*, 25, 161-197.
- Turoff, M., & Hiltz, S. R. (1996). Computer-based Delphi processes. In *Gazing Into the Oracle: The Delphi Method and Its Application to Social Policy and Public Health* (pp. 56-88). London: Kingsley Publishers.

ADDITIONAL READING

- Crandall, B., Klein, G., & Hoffman, R. R. (2006). *Working Minds: A Practitioner's Guide to Cognitive Task Analysis*. Massachusetts Institute of Technology.

Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Review of Psychology*, 47(1), 273–305. doi:10.1146/annurev.psych.47.1.273 PMID:15012483

Salas, E., Tannenbaum, S. I., Kraiger, K., & Smith-Jentsch, K. A. (2012). The science of training and development in organizations: What matters in practice. *Psychological Science in the Public Interest*, 13(2), 74–101. doi:10.1177/1529100612436661 PMID:26173283

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124–1131. doi:10.1126/science.185.4157.1124 PMID:17835457

Venetjoki, N., Kaarlela-Tuomaala, A., Keskinen, E. & Hongisto, V., "The effect of speech and speech intelligibility on task performance", *Ergonomics*, 49, 11, 2006, pp. 1068-1091

KEY TERMS AND DEFINITIONS

(United States) Common Core Standards (Initiative): The Common Core State Standards Initiative is an educational assembly in the United States that details what K-12 students should know in subject areas focused on English language arts and mathematics at the end of each grade. What has been produced is a set of standards by subject and federal government will fund states to incorporate these standards into teaching curricula. This has begun to draw criticism and generate controversy as a method of federal control of education.

Domain Expert: A person with special knowledge or skills in a particular area of endeavor. The term *domain expert* is frequently used in expert systems software development.

Estimate-Talk-Estimate (ETE) Delphi Process: Any Delphi process is based on the principle that forecasts from a structured group of experts are more accurate than those from unstructured

groups or individuals. The technique has lately been adapted for use in face-to-face meetings, and is then called mini-Delphi or Estimate-Talk-Estimate (ETE) Delphi.

Group Dynamics: The interactions that influence the attitudes and behavior of people when they are grouped with others through either choice or accidental circumstances. Social psychologist Kurt Lewin coined the term *group dynamics* to describe the positive and negative forces within groups.

Groupthink: Groupthink, a term coined by social psychologist Irving Janis (1972), occurs when a group makes faulty decisions because of group pressures; groupthink is a psychological phenomenon that occurs within a group of people, in which the desire for harmony or conformity in the group results in an incorrect or deviant decision-making outcome.

Pareto Analysis: A technique used for decision making based on the Pareto Principle, known as the 80/20 rule. It is a decision-making technique that separates a limited number of input factors as having the greatest impact on an outcome, either desirable or undesirable. In its simplest terms, Pareto analysis will typically show that a disproportionate improvement can be achieved by ranking various causes of a problem and by concentrating on those solutions or items with the largest impact. This type of decision-making can be used in many fields of endeavor, from government policy to decision-making.

Small Group Delphi Paradigm (SGDP): A variant Delphi model based on traditional Delphi processes, as developed and refined from 1985 through 1992 by the author. Its core features are multiple face-to-face sessions of small groups of carefully selected SME's, always with a facilitator and a step- through set of iterations to achieve consensus, not statistical sharpening.

Subject Matter Expert (SME): An individual who exhibits the highest level of expertise in performing a specialized job, task, or skill within the organization.

United States Federal Aviation Administration (FAA) Advisory Circulars (AC): The FAA normally publishes advisory circulars (ACs) designed to provide assistance and guidelines in complying with FARs; they serve as a “how to” template. The AC’s are titled and numbered and are not legally binding.

United States Federal Aviation Administration Regulations (FAR): All are in the Combined Federal Regulations (CFR), Title 14, Aeronautics and Space; each a descriptive title and number; they are legally binding.

Corporate Social Responsibility

B**Ben Tran***Alliant International University, USA*

INTRODUCTION

In today's changing world, Corporate Social Responsibility (CSR) is a growing area of interest for academics, practitioners and entrepreneurs, in terms of both theory and practice. CSR is a concept whereby companies integrate social, environmental, and health concerns in their business strategy (policy) and operations and in their interactions with stakeholders on a voluntary basis. The social responsibility of business encompasses the economic, legal, ethical, and discretionary expectations that society has of organizations at a given point in time (Carroll, 1979). As the world is shrinking due to globalization the concept of CSR has acquired an undeniably high degree of relevance and scope in a large number of sectors. Many academicians and practitioners are developing theory and practices of this concept of social responsibility among entrepreneurs.

An examination of the literature of CSR and corporate reputation has led to the rise of stakeholder management in a complex market environment in the United States, Europe and Asian regions. Today's corporations have paid substantial attention to the three key concepts, corporate reputation (CR), corporate social responsibility (CSR), and stakeholder relations (SR) (Argenti & Barnes, 2009; Argenti & Druckemiller, 2004; Cees, Riel, & Fombrun, 2007). In an Asian country, many corporations such as multinational corporations (MNCs), government linked corporations (GLCs) and local private companies (LPC) eagerly communicate with their internal and external stakeholders to portray the unique identity of the organizations (Abdullah, 2007). By having a unique identity of the organization, the corporations not only can improve the

image and reputation of the organization but also differentiate its strategy development and shape the extraordinarily valuable asset of the organization (Bouchokhi & Kimberly, 2008).

There are numerous studies (Balmer & Greyser, 2003; Fombrun & Riel, 2004; Kitchen & Schultz, 2001; Cees et al., 2007) and increasing debates concerning image, identity, and reputation with the greater emphasis on how the organizations managed their communication systems. Some studies also attempt to focus on corporate branding which is substantially associated with managing reputation of MNCs in a Western country (Money, Rose, & Hillenbrand, 2008; Hatch & Schultz, 2008). However, a few studies in an Asian country (Lines, 2004; Moon, 2005) were found to examine corporate reputation focusing on stakeholder relations and corporate social responsibility. A "West is best mentality" can be daunting thought but it affects a real attitude among Asian society. Chinese consumers downgraded Chinese-branded companies and rated higher reputation to international companies such as Nokia, Intel, BMW, and IBM (Fombrun & Pan, 2006, p. 165). Within this context, one frequently asked question is: What theory or model should be followed or considered for implementation of CSR practices?

BACKGROUND

The days when corporate social responsibility (CSR) was a concept and practice confined to North American and European companies are over. Particularly in the past ten years or so, Asian businesses have increasingly brought to bear their considerable energy and thought to matters of regional and global concern. From climate change

DOI: 10.4018/978-1-5225-2255-3.ch058

and other sustainability questions to product safety and global labor standards, Asian business leaders have an opportunity to shape the international response to some of the broad issues facing global society. Not only will Asian business leaders who embrace corporate social responsibility offer their own distinctive approach, but their participation in sweeping worldwide matters can have far-reaching and beneficial impact for all.

Literature Review

The impact of CSR on corporate reputation is shaped by how the firm communicates its CSR activities to its external stakeholders and how its activities are reported in the national media and other communication mediums (Rettab, Ben, & Mellahi, 2008). McWilliams and Siegel (2001) emphasized that consumers consider socially responsible firms to have a good reputation by constantly executing a series of programs of social responsibility. Such CSR programs give a big impact to corporate reputation. A good reputation influences a positive consumer satisfaction. In addition, companies use CSR communication to enhance customer loyalty (Abdullah & Aziz, 2011; Jacob & Kyner, 1973; Keller, 1993). This effort may improve favorable relationships between the organization and its stakeholders. As Grunig and Hunt (1984, p. 6) stated that effective public relations is about “management of communications between an organization and its publics”.

McWilliams and Siegel (2001, p. 120) pointed out that positive CSR, “creates a reputation that a firm is reliable and honest”. Similarly, Bhattacharya and Sen (2004) argued that CSR builds a reservoir of goodwill that firms can draw upon in times of crisis. Positive reputations have often been linked with positive financial returns, with their value tied to the inability of competitors to imitate the reputation. Roberts and Dowling (2002) found that the value of a positive reputation is, “precisely because the development of a good reputation takes considerable time and depends on a firm making stable and consistent investments

over time”. Therefore, reputation is arguable the most valuable asset of any firm and thus worth protecting. Banerjee, Iyer, and Kashyap (2003) postulated firms that are under greater scrutiny from a broader stakeholder groups facing greater business exposure.

In addition, Post, Lawrence, and Weber (2002, p. 8) found that stakeholders is “those people and groups that affect, or can be affected by an organization’s decision, policies, and operations”. Øyvind Ihlen (2008, p. 136) from his research indicated that the organization’s success depends on how it is able to manage its relationship with key stakeholder groups, such as customers, employees, suppliers, communities, politicians, owners, and others. Similarly, Werther and Chandler (2006) quoted that the stakeholder relationship occurs as a consequence of business activities and such groups might be local communities, the media, business support groups, state and local government, social activist groups and so forth. To ensure the organization’s success, the company have “to keep the support of all these groups, balancing their interest, make the organizations a place where stakeholder interests can be maximized over time” (Freeman & Philips, 2002, p. 333).

According to Abdullah and Aziz (2011), however, defined CSR as a business approach that creates long-term shareholder value by embracing opportunities and managing risks deriving from economic, environmental and social developments. Business in the Community (BITC) defined CSR as “a company’s positive impact on society and the environment through its operations, productions or services and through its interaction with key stakeholders such as employees, customers, investors and suppliers” (Katsoulakos & Katsoulakos, 2006, p. 13). On the other hand, Carroll (1979, p. 500) states that businesses that practice social responsibility attend to “economic, legal, ethical, and discretionary (philanthropic) expectations that society has of organizations at a given point in time”. Hence, the CSR concept relates closely to its family terms such as corporate citizenship (Bowen, 1953; Carroll, 1979; Mason,

1960), corporate social responsiveness (Ackerman & Bauer, 1976; Frederick, 1998; Strand, 1983), corporate social performance (Stanwick & Stanwick, 1998; Swanson, 1995; Wood, 1991), and stakeholder management (Donaldson & Preston, 1995; Jones, 1995). Following to the notion of CSR quoted by McWilliams, Siegel, and Wright (2006, p. 1), organizations should not only be concerned about making a profit but also engaged in “actions that appear to further some social good, beyond the interests of the firm and that which is required by law”.

However, the concept of CSR is still blurred and fuzzy argued by Lantos (2001). Vaaland and Heide (2005) decomposed CSR into CSR enforcement—the ability to make the firm more resistant to be attacked for sudden, unforeseen and negative incidents highlighted by anger stakeholders. This may threaten and could seriously harm the company’s reputation. On the other hand, if the critical incident is handled skillfully, under CSR recovery—the potential loss of company reputation could be minimized and possibly recovered after a short time (Vaaland & Heide, 2005, 2008). Studying on CR, CSR, and SR are often associated with improving the organizational performance. Those recent and past studies have proven the strong alignment between CR, CSR, and SR within the Western hemisphere. The fact that the emerging concepts are universal, which means they are widely recognized in the U.S., Europe, and Asia. However, there is no universal definition of CSR as the concept has been defined differently in different countries.

Corporate Social Responsibility (CSR) in Asia

Corporate social responsibility (CSR) in Asia, however, has to be seen not only as philanthropic activities of the firm on the one hand but also as initiatives beyond legal requirements in the other (Ramasamy, Yeung, & Au, 2010). Such definitions of CSR are prevalent in CSR literature from western developed countries. Hemingway and MacLagan (2004), for instance, argue that

the social responsibility of firms is one that is voluntary and goes beyond the law. This may implicitly assume that companies are already ethical and lawful. This assumption, however, cannot be taken for granted in developing countries where the rule of law is relatively weak. Thus, companies that adhere to legal manner have to be considered to be socially responsible businesses as well. For this reason, Carroll’s (1979) definitions of SR is perhaps among the most suitable description. Carroll’s pyramid-like model explains the various responsibilities of businesses, building on economic responsibilities and later moving onto legal, ethical, and philanthropic responsibilities in a stage like manner.

CORPORATE SOCIAL RESPONSIBILITY (CSR) WITH ASIAN CHARACTERISTICS

While sustainability is essentially a global concept, CSR has developed differently in different parts of the world. This is not surprising, because CSR flows naturally from the social contract that defines the relationship between business and society. These social contracts, moreover, derive from distinct cultural, economic and governance models. The rise of an indigenous version of CSR across Asia is therefore developing with both similarities and differences to that practiced in the rest of the world. Unique Asian realities are critical in shaping regional approaches to a sustainable and responsible business model, particularly in comparison to the United States and Europe, notably (Sharma, 2013; The Asia Business Council, 2008):

- Asia is more diverse—culturally, linguistically, and economically—than others regions of the world. First, it is a mistake to refer to one version of Asian CSR because it is defined differently across the region. Second, Asia taken as a whole, offers a fascinating tested where various approaches and models can be tried out.

- Asia is experiencing the most rapid economic growth of any of the world's regions. The promise and reality of rising living standards remain foremost in the minds of policymakers and businesses. This outlook contrasts sharply with Western perspectives, where the focus is more on maintain already high living standards.
- Asia, more so than other regions, includes a dynamic mix of developed and developing economies. Coupled with the rise of "South-South" trade, this mix positions Asia to influence other regions of the worlds on a range of issues.
- Asia's rising economic power coincides with increasing political power. China, India, and Russia are poised to assert ever more political influences in the coming decades, meaning that Asian perspectives on CSR may alter or argument global definitions of the concept.
- Asian businesses and policymakers are substantially less inclined than Western companies to rely on established international principles and standards on social and environmental questions. This inclination may reflect the fact that the standards are not universal, but were developed primarily by and for Western interests. It may also result from the desire to prioritize economic growth over other factors, and Asia's affinity for consensus-building, in contrast with the more legalistic approaches favored in the West.

These distinct features shape the way Asian companies and leaders define CSR, and also have impact on how "Made-in-Asia CSR" will influence global perspectives on sustainability.

Asian Models of Corporate Responsibility

CSR in Asia is evolving in different ways, each with its own priorities and mindsets. The original

CSR model emerged from Western multinational companies operating in Asia. Beginning in the early 1990s, American and the European companies began to apply a set of principles to the operations of their Asian suppliers and business partners. The rise of codes of conduct, focusing both on labor and environmental practices, can in retrospect be seen as an early effort to develop "soft law" principles for the globalizing economy (Sharma, 2013; The Asia Business Council, 2008). These efforts while often well-intentioned, in many cases reflected Western perceptions—and misperceptions—of Asia. Tied to sometimes controversial Western origins, this first conception of CSR has marked the development of three different approaches to CSR more firmly tethered to Asian thinking.

The first approach—based on the notion of community benefit—involves the practices of Asian companies doing business inside their own countries. While the subject is too diverse to be treated completely here, the basic premise is that companies can work together at the intersection of local conditions, community expectations, and at times, the inclinations of powerful business leaders who may be looking to establish—and demonstrate—a legacy of community benefit. The most interesting examples of this approach are the ones where companies create new business models for "blended" benefit, such as Industrial Credit and Investment Corporation of India (ICICI) Bank in India, which is aiming to bring new financing models to impoverished communities desperately in need of banking services. The second approach, is the arena of Asian-influenced CSR that has developed where Asian multinational companies shape—or choose not to—the communities in which they operate outside Asia. The third approach, is that, as the number of Asian companies operating in the United States and Europe grows, there is new opportunity for Asian business leaders to demonstrate their will to be good corporate citizens in these developed markets.

China Securities Regulatory Commission

The most notable development for the adoption of CSR in China at the turn of the century was the 2001 Code of Corporate Governance for Listed Companies issued by the China Securities Regulatory Commission and the State Economic and Trade Commission. This was the first ever recognition of the interest of “stakeholders”, a remarkably modern text but with little on implementation and enforcement (Sharma, 2013). Chapter 6 of the Code clearly extends CSR’s scope beyond labor issues to a broader range of activities, ranging from respect for legal rights of stakeholders to transparency in reporting on operations and financial situation of a company (China Securities Regulatory Commission and State Economic and Trade Commission, 2001).

The 2006 Company Law

When China first drafted the company law in the early 1990s, CSR was a little known concept. The 1994 Company Law explicitly addressed issues of labor rights and employee rights in Article 15 and 16 of the law, giving rise to China’s first ever stakeholder-friendly legal statute. The revised 2006 Company Law gives explicit recognition to CSR, by requiring companies to adhere to social and business ethics as well as fulfil social responsibilities.

Shenzhen Stock Exchange

In September 2006, the Social Responsibility Index (SSE) released a set of “Social Responsibility Guidelines for Listed Companies” (Unknown, N. Y.) that encourage listed companies to assume responsibility for social development, the environment and other natural resources, and commit to protecting the right and interests of shareholder, creditors, employees, customers, consumers, and others involved with their business. The SSE guidelines also encourage companies to regularly

evaluate and issue voluntary disclosures about their performance.

B

Emergence of Different Models

It is no more possible to define a single model of “Asian CSR” than it is to define a single Asian language. The region’s diversity of corporate practices rivals that of its natural environments. In India, philanthropy continues to loom larger in the CSR landscape. At the same time, civil society also plays a significant role in India, suggesting that until the now Western “stakeholder” model with the views of multiple players taken into account, could develop. Any such model, and indeed the overall approach to CSR in India, will also be defined by the tension between the country’s energetic commercial development and its cultural and spiritual values, which sometime run counter to a focus on simple material progress.

China’s rise has coincided with growing interest in CSR in Asia. For example, Hu Jintao’s drive for a “harmonious society” has focused increased government attention on the nation’s mounting environmental worries, as well as growing economic disparity (The Asia Business Council, 2008). Concerns about corporate governance, safety issues and frenzied stock speculation are bubbling under the surface, raising the possibility that pressure on the role of business in Chinese society could greatly increase. The insurance of a directive from the State-owned Assets and Administration Commission to state-owned enterprises to adopt CSR is but the latest sign of interest from the central government in Beijing.

In Northeast Asia, while Japan and Korea have experienced historical conflict, they share relative prosperity, as well as concerns about the economic and political impact of their Chinese neighbor. Japanese and Korean multinationals are showing increased interest in the global CSR dialogue. With fellow Korean Ban Ki-moon as Secretary-General of the UN, Korean business may be more inclined to align its energy-saving practices with those of the UN Global Compact.

Across Southeast Asia, smaller economies from Vietnam to Cambodia to Indonesia focus mainly on economic development and their role as suppliers to global manufacturers. Apart from Indonesia's effort to legislate CSR, these nations are not yet shaping distinct approaches to CSR. Hence, the trading hubs of Hong Kong and Singapore have increasingly active CSR communities. These two cities might well become to Asia what London has become in Europe: centers of debate and innovations that can inform and invigorate practices around the region.

FUTURE RESEARCH DIRECTIONS

What happens in Asia matters, because this is not only because so much of the world's economic energy comes from the region, but also because the global challenges that CSR is designed to address are more vividly on display in Asia than elsewhere. Any look at global mega-trends will find countless examples of CSR at work today from Tokyo to Karachi. Asia leads in the number of mega-cities, which bring both economic opportunity and environmental nightmares. Asian is also home to both demographic "youthquakes" with explosions of growth in the younger population such as in Vietnam, and fast-aging populations such as those in China, Japan, and South Korea. Asia has massive migration both within and across national borders. Many Asian societies, such as Korea, are champion early adopters of information technology. But Asia is experiencing rapidly changing social contracts, with questions about governance gaps and power shifts across the region.

The evolution of CSR in the West over four distinct periods offers some lessons in how Asian businesses might consider incorporating the concept into their business strategies. The first phase, in the mid-1990s, companies' actions were characterized by reaction. In this phase, businesses responded, often with defensiveness, as new issues landed on their agendas. The second phase, companies began to look to innovation, focusing

on the reaction of best practices. All too often, however, such efforts developed only as demonstration projects that could not be scaled up. The third phase, begun earlier this decade, saw a welcome focus on integration. With this approach, companies began to integrate social and environmental questions into their core activities. More recently, businesses have begun to look at how CSR can generate value creation. This latest idea means the development of new markets, products and services that create top-line benefit, and build sustainability into the business, rather than just as a coat of "sustainability varnish" layered on top of an existing business model (Sharma, 2013; The Asia Business Council, 2008).

CONCLUSION

Compliance to the existing legislative and regulatory framework should be the mantra for CSR in Asia for the current generation of corporation of corporate entities. State capacity to enforce compliance by industry is very often constrained by pervasive corruption and the limited resources of the developing economy governments. In the individual country narrative it is amply evident that in many countries the poor state of labor rights or environmental degradation is not a result of inadequate regulation, but the weak enforcement of national and local laws (Sharma, 2013). Enforcement is the basic first step for governments to create a compliance culture and an environment conducive to industry taking on responsibility citizenship beyond compliance.

With CSR, business is being required to replace that regulatory approach or face a backlash from society. It may behoove the corporate sector in Asia to work on building the regulatory and enforcement capacity of governments so that business is not expected to play this governing role in the future. CSR can in fact be business' willingness to contribute to the governance deficit today, not by usurping that role, but building public institutional capacity to level the playing field

not only across business but across stakeholders. Hence, it may well be the SME that will define the CSR trajectory in the years to come. Closer to the community it operates in, the SME sector as a collective is the largest employer in most economies. In the 30 countries that comprise the Organization for Economic Co-Operation and Development (OECD), SMEs represent over 95% of enterprises in most countries and generate over half of private sector employment

REFERENCES

- Abdullah, Z. (2007). Toward international cultural diversity management of public relations: Viewpoints of chairmen/CEO. *International Journal of Economics & Management*, 1(2), 285–299.
- Abdullah, Z., & Aziz, A. Y. A. (2011). Managing corporate reputation, stakeholder relations, and corporate social responsibility: A southwest Asian perspective. *International Journal of Economics and Management*, 5(1), 199–318.
- Ackerman, R. W., & Bauer, R. A. (1976). *Corporate social responsiveness*. Reston, VA: Reston Publishing.
- Argenti, P., & Barnes, C. M. (2009). *Digital strategies for powerful corporate communications*. New York, NY: McGraw-Hill.
- Argenti, P., & Druckenmiller, B. (2004). Reputation and the corporate brand. *Corporate Reputation Review*, 6(4), 368–374. doi:10.1057/palgrave.crr.1540005
- Balmer, J. M., & Greyser, S. A. (2003). *Revealing the corporation: Perspectives on identity, image, reputation, corporate branding and corporate-level marketing*. London: Routledge, Taylor and Francis Group. doi:10.4324/9780203422786
- Banerjee, S. B., Iyer, E. S., & Kashyap, R. K. (2003). Corporate environmentalism: Antecedents and influence of industry types. *Journal of Marketing*, 6(2), 106–122. doi:10.1509/jmkg.67.2.106.18604
- Bhattacharya, C. B., & Sen, S. (2004). Doing better at doing good: When, why, and how consumer product respond to corporate social initiatives. *California Management Review*, 47(1), 9–24. doi:10.2307/41166284
- Bouchikhi, H., & Kimberly, J. R. (2008). *The soul of the corporation: How to manage the identity of your company*. Wharton School Publishing.
- Bowen, H. R. (1953). *Social responsibilities of the businessman*. New York: Harper-Row.
- Carroll, A. B. (1979). Three-dimensional conceptual model of corporate performance. *Academy of Management Review*, 4(4), 497–505.
- Cees, B. M., Riel, V., & Fombrun, C. J. (2007). *Essentials of corporate communication*. Abingdon, UK: Routledge.
- China Securities Regulatory Commission and State Economic and Trade Commission. (2001). *Code of corporate governance for listed companies in China: Code of corporate governance for listed companies*. Retrieved on February 25, 2015. Available at http://www.ecgi.org/codes/documents/code_en.pdf
- Donaldson, T., & Preston, L. E. (1995). The stakeholder theory of the corporation: Concepts, evidence, and implications. *Academy of Management Review*, 20(1), 65–91.
- Fombrun, C. J., & Pan, M. (2006). Corporate reputation in China: How do consumers feel about companies? *Corporate Reputation Review*, 9(3), 165–170. doi:10.1057/palgrave.crr.1550024

- Fombrun, C. J., & Riel, V. (2004). *Fame and fortune: How successful companies build winning reputation*. New York, NY: Prentice Hall.
- Frederick, W. C. (1998). Creatures, corporations, communities, chaos, complexity: A naturological view of corporate social role. *Business & Society*, 37(4), 358–389. doi:10.1177/000765039803700403
- Freeman, R. E., & Philips, R. A. (2002). Stakeholders theory: A libertarian defense. *Business Ethics Quarterly*, 12(3), 331–349. doi:10.2307/3858020
- Grunig, J. E., & Hunt, T. (1984). *Managing public relations*. New York, NY: Holt, Rinehart and Winston.
- Hatch, M. J., & Schultz, M. (2008). *Taking brand initiative: How companies can align strategy, culture, and identity through corporate branding*. San Francisco, CA: Wiley.
- Hemingway, C. A., & MacLagen, P. W. (2004). Managers personal values as drivers of corporate social responsibility. *Journal of Business Ethics*, 50(1), 33–44. doi:10.1023/B:BUSI.0000020964.80208.c9
- Ihlen, O. (2008). Mapping the environment for corporate social responsibility: Stakeholders, publics and the public sphere. *Corporate Communications: An International Journal*, 13(2), 135–146. doi:10.1108/13563280810869578
- Jacob, J., & Kyner, D. B. (1973). Brand loyalty versus repeat purchase behaviour. *JMR, Journal of Marketing Research*, 10(1), 1–9. doi:10.2307/3149402
- Jones, T. M. (1995). Instrumental stakeholder theory: A synthesis of ethics and economic. *Academy of Management Review*, 20(2), 404–437.
- Katsoulakos, P., & Katsoulakos, Y. (2006). *A multi-dimensional view of corporate responsibility*. Retrieved on February 23, 2015. Available at <http://www.scribd.com/doc/53327875/4CR-A-Multi-Dimensional-View-of-Corporate-Responsibility-Copy#scribd>
- Keller, K. L. (1993). Conceptualizing, measuring and managing customer based brand equity. *Journal of Marketing*, 57(1), 1–22. doi:10.2307/1252054
- Kitchen, P. J., & Schultz, D. E. (2001). *Raising the corporate umbrella: Corporate communications in the 21st century*. London: Palgrave Publisher. doi:10.1057/9780230554580
- Lantos, G. P. (2001). The boundaries of strategic corporate social responsibility. *Journal of Consumer Marketing*, 18(7), 595–630. doi:10.1108/07363760110410281
- Lines, V. L. (2004). Corporate reputation in Asia: Looking beyond bottom-line performance. *Journal of Communication Management*, 8(3), 233–245. doi:10.1108/13632540410807664
- Mason, E. S. (1960). *The corporation in modern society*. Cambridge, MA: Harvard University Press.
- McWilliams, A., & Siegel, A. D. (2001). Corporate social responsibility: A theory of the firm perspective. *Academy of Marketing Review*, 26(10), 117–127.
- McWilliams, A., Siegel, A. D., & Wright, P. M. (2006). Corporate social responsibility: Strategic implications. *Journal of Management Review*, 43(1), 1–18.
- Money, K., Rose, S., & Hillenbrand, C. (2008). *Where does corporate branding end and reputation management begin?* Paper presented at the 12th International Reputation Institute Conference on Corporate Reputation, Brand Identity, and Competitiveness, Beijing, China.

Moon, C. W. (2005). Corporate social responsibility (CSR) in Asia: A seven country study of CSR website reporting. *Business & Society*, 44(4), 415–441. doi:10.1177/0007650305281658

Post, J. E., Lawrence, A., & Weber, J. (2002). *Business and society: Corporate strategy, public policy, ethics* (19th ed.). New York, NY: McGraw-Hill.

Ramasamy, B., Yeung, M. C. H., & Au, A. K. M. (2010). Consumer support for corporate social responsibility (CSR): The role of religion and values. *Journal of Business Ethics*, 91(1/2), 61–72. doi:10.1007/s10551-010-0568-0

Rettab, B., Ben, B. A., & Mellahi, K. (2008). A study of management perceptions of the impact of corporate social responsibility on organisational performance in emerging economies: The case of Dubai. *Journal of Business Ethics*, 89(3), 371–390. doi:10.1007/s10551-008-0005-9

Roberts, P., & Dowling, G. (2002). Corporate reputation and sustained superior financial performance. *Strategic Management Journal*, 23(12), 1077–1093. doi:10.1002/smj.274

Sharma, B. (2013). *Contextualising CSR in Asia: Corporate social responsibility in Asian economies*. Singapore: Lien Centre for Social Innovation.

Stanwick, S. D., & Stanwick, P. A. (1998). Corporate social responsiveness: An empirical examination using the environmental disclosure index. *International Journal of Commerce and Management*, 8(3/4), 26–40. doi:10.1108/eb047373

Strand, R. (1983). A system paradigm of organizational adaptations to the social environment. *Academy of Management Review*, 8(1), 90–96. PMID:10260562

Swanson, D. L. (1995). Addressing a theoretical problem by reorienting the corporate social performance model. *Academy of Management Review*, 20(1), 43–64.

The Asia Business Council. (2008). *Corporate social responsibility: Business solutions to global challenges*. Business for Social responsibility (BSR) for the Asia Business Council Spring 2008 Forum.

Vaaland, T. I., & Heide, M. (2005). Corporate social responsiveness: Exploring the dynamics of bad episode. *European Management Journal*, 23(5), 495–506. doi:10.1016/j.emj.2005.09.005

Vaaland, T. I., & Heide, M. (2008). Managing corporate social responsibility: Lessons from the oil industry. *Corporate Communications: An International Journal*, 13(2), 212–225. doi:10.1108/13563280810869622

Werther, W. B., & Chandler, D. (2006). *Strategic corporate social responsibility: Stakeholders in a global environment*. London: Sage Publication.

Wood, D. (1991). Corporate social performance revisited. *Academy of Management Review*, 16(4), 691–718.

ADDITIONAL READING

Beal, B. D. (2013). *Corporate social responsibility: Definition, core issues, and recent developments*. SAGE Publications, Inc.

Benn, S., & Bolton, D. (2011). *Key concepts in corporate social responsibility*. SAGE Key Concepts Series.

Brenkert, G. G., & Beauchamp, T. L. (2012). *The oxford handbook of business ethics*. Oxford University Press.

Chandler, D. B., & Werther, W. B. (2013). *Strategic corporate social responsibility: Stakeholder, globalization, and sustainable value creation* (3rd ed.). SAGE Publications, Inc.

- Crane, A., Matten, D., & Spence, L. (2013). *Corporate social responsibility: Readings and cases in a global context* (2nd ed.). Routledge.
- Crane, A., McWilliams, A., Matten, D., Moon, J., & Siegel, D. S. (2009). *The oxford handbook of corporate social responsibility*. Oxford University Press. doi:10.1093/oxfordhb/9780199211593.001.0001
- Dillard, J., Haynes, K., & Murray, A. (2014). *Corporate social responsibility: A research handbook*. Routledge.
- Ehnert, I., Harry, W., & Zink, K. J. (2013). *Sustainability and human resource management: Developing sustainable business organizations*. Springer.
- Fleming, P., & Jones, M. V. (2012). *The end of corporate social responsibility: Crisis and critique*. SAGE Publications, Ltd.
- Kolk, A., Hong, P., & van Dolen, W. (2010). Corporate social responsibility in China: An analysis of domestic and foreign retailers' sustainability dimensions. *Business Strategy and the Environment*, 19(5), 289–303.
- Lee, S. Y., Ramasamy, A., & Rhee, J. H. (2014). *Green leadership in China: Management strategies from China's most responsible companies*. Springer. doi:10.1007/978-3-642-55058-4
- Li, J., & Qian, C. (2013). Principal-principal conflicts under weak institutions: A study of corporate takeovers in China. *Strategic Management Journal*, 34(4), 498–508. doi:10.1002/smj.2027
- Lin, L. W. (2010). Corporate social responsibility in China: Window dressing or structural change. *Berkeley Journal of International Law*, 28(1), 64–100.
- Low, K. C. P., Idowu, S. O., & Ang, S. L. (2013). *Corporate social responsibility in Asia: Practice and experience*. Springer.
- Lyon, T. P., & Maxwell, J. W. (2011). Greenwash: Environmental disclosure under threat of audit. *Journal of Economics & Management Strategy*, 20(1), 3–41. doi:10.1111/j.1530-9134.2010.00282.x
- Marquis, C., & Qian, C. (2014). Corporate social responsibility reporting in China: Symbol or substance? *Organization Science*, 25(1), 127–148. doi:10.1287/orsc.2013.0837
- Schwartz, M. S. (2011). *Corporate social responsibility: An ethical approach*. Broadview Press.
- Tran, B. (2008). Corporate ethics: An end to the rhetorical interpretations of an endemic corruption. *A Special Issue of Social Responsibility Journal. Ethics and Morality in Business Practice*, 4(1/2), 63–81.
- Tran, B. (2008). Paradigms in corporate ethics: The legality and values of corporate ethics. *A Special Issue of Social Responsibility Journal. Ethics and Morality in Business Practice*, 4(1/2), 158–171.
- Wang, S. (2014). *Chinese strategic decision-making on CSR*. Springer.
- Werther, W. B., & Chandler, D. B. (2010). *Strategic corporate social responsibility: Stakeholders in a global environment* (2nd ed.). SAGE Publications, Inc.
- Williams, G. (2010). *Responsible management in Asia: Perspectives on CSR*. Palgrave Macmillan.

KEY TERMS AND DEFINITIONS

Corporate Reputation: It is the process of creating influencer strategies that deepen understanding, build trust and mitigate risk in complex, ever-changing environments.

Corporate Social Responsibility (CSR): It is a concept whereby companies integrate social,

environmental, and health concerns in their business strategy (policy) and operations and in their interactions with stakeholders on a voluntary basis.

Government Linked Corporations (GLCs):

A term that is related to Asian businesses where companies in which some of the shares are owned by the government.

Local Private Companies (Privately Held Company or Close Corporation): It is a business company owned either by non-governmental organizations or by a relatively small number of shareholders or company members which does not offer or trade its company stock (shares) to the general public on the stock market exchanges,

but rather the company's stock is offered, owned and traded or exchanged privately.

Multinational Corporation [(MNCs) or Multinational Enterprise]: It is an organization, that owns or controls productions of goods or services in one or more countries other than the home country.

Stakeholders: Are those individuals, organizations who 'hold' a 'stake' in your organization's interests.

Stakeholder Relations: It is about improving your professional relationships - or lack of them - with governments, regulators, membership, industry, media or communities.

B

Digital Transformation Journeys in a Digitized Reality

Jurgen Janssens
QSpin, Belgium

INTRODUCTION

Big Data and Digital Transformation are hot topics since several years.

Big Data are large volumes of data from one or more sources, analyzed through innovative information processing to develop insights, resulting in better business decisions and process adaptation¹. Estimates of Turner, Reinsel and Gantz (2014) expect human and machine data to increase to 40ZB by 2020.

Digital Transformation implies often Big Data powered changes (Mayhem, Saleh, & Williams, 2016). Whereas concrete outcomes of the latter are booming, Digital Transformation is still seen as a faraway future. Nowadays, however, companies and organizations should not ask themselves anymore how to prepare for the upcoming Digitization, but how to adapt to and in today's Digitized/Digitizing World.

To get there, they need to understand how to evolve by embracing new technological potential and by looking through the blurring frontiers between offline and online reality. They have to take changing habits and continuously transforming customer segments into account. They have to look beyond the mere 'product' and 'selling to customers' angle (Richter, & Wee, 2016). To keep their value, companies, organizations and even governments need to understand that mental mobility of people and their data-stimulated ecosystem have turned people's need for products in a need for shaped services.

This has a significant impact on the strategic vision. Typically, this requires a new way of working adapted to the redefined borders between

reactivity and proactivity, and to the attention for real-time service and contextual adaptation. An organic compromise has to be found that answers the need for creative freedom and the need for an environment where management and development of human capabilities are possible in a structured way.

This article will shed light on core components for Digital Transformation. Concrete methodologies will also be analyzed that can guide the process of reviewing company fundamentals in the Digital Reality.

It will be explained that human means and data power need to be developed to develop a company's digitally quantified intuition² and to get most out of the customer journey.

Altogether, this article will illustrate that the success of Digital Transformation goes through integrating technological possibilities and dynamic customer journeys in the corporate DNA. It will require a new internal perspectives on projects, and create a new dynamic that, in the end, will allow people, companies and public and private organizations to be in the driving seat of this Fourth Industrial Revolution.

BACKGROUND

Reflections about Digitization and Digital Transformation exist since several years, also on intergovernmental level³. With technology developments in various fields amplifying one another, this will continue. It will lay the foundation for a revolution more all-encompassing than anything seen so far (Schwab, & Samans, 2016).

DOI: 10.4018/978-1-5225-2255-3.ch059

There are several risks if the dynamic is not engineered and managed correctly.

First of all, companies need to be aware that Digital is not an add-on, but the very essence of the transformation journey to remain in the game (Henke, Libarikian, & Wiseman, 2016). It involves a change in leadership, new business models, and an increased use of technology to improve the customer experience. It is only by integrating this Digital backbone that Digital strategies can have a lasting effect. Digital is one of the main reasons half of the companies on the Fortune 500 have disappeared since 2000 (Nanterme, 2016). Failing to understand its functioning can thus lead to companies losing their relevancy.

Traditionally, Digital Transformation is covered from separate angles.

Certain sources analyze the impact of technology on job markets and human employability (Van Driessche, 2014). Others focus more on specific pieces of the corporate landscape – ranging from the need to extend the CxO suite with a Digital Officer, to the development of a new way of thinking, or the analysis of managerial choices to be made for techn(ological) transition.

Solely focusing on subsections risks to lead to an incomplete integration of the Digital dynamic. At best, the initiative will be an intermediary step. At worst, it will create a negative spiral that funnels means, time and potential - a risk further reinforced by the speed at which Digital is unfolding.

On the methodological level, a similar tendency can be observed. Agile frameworks like scrum⁴ or Kanban⁵ ensure guidance of essential aspects for agile product delivery (Galen, 2013). It is however less common to have agile covered for more profound transformations like those observed in the Digital Age.

The author wants to bring the components of the Digital journey together. Shaping their complementarity should open the possibilities of a new way of thinking about customers, organizations and projects, about shaping service and product delivery, about interacting with the ecosystem - and about creating value.

SOLUTIONS AND RECOMMENDATIONS

Changing Ecosystems, Dynamic Needs

Companies are continuously challenged to raise the bar, or to reinvent themselves. Digital Transformation is seen as one of the means to get there. This transformation is already well underway. Early 2016, the World Economic Forum in Davos called it the Fourth Industrial Revolution (Schwab & Samans, 2016):

Today, we are at the beginning of a Fourth Industrial Revolution. Developments in genetics, artificial intelligence, robotics, nanotechnology, 3D printing and biotechnology, to name just a few, are all building on and amplifying one another. (...) While the impending change holds great promise, the patterns of consumption, production and employment created by it also pose major challenges requiring proactive adaptation by corporations, governments and individuals.

Besides the semantical discussion to call it Digital Transformation or Fourth Industrial Revolution, facts are there that illustrate the vastness of this evolution (Carter, 2015). The world's largest taxi firm, Uber, owns no cars. The world's most popular media company, Facebook, creates barely content. The world's largest accommodation provider, Airbnb, owns no property. So, vast changes are already ongoing.

In fact, private and public actors have already interwoven Digital in everyone's day-to-day. The British government, for instance, is coached by scientific experts of the Behavioural Insights Team⁶, amongst others to improve processes like the reduction of late paid taxes (Van Leemputten, 2016).

This is also true for private companies. Thanks to PlayStation, Sony's revenue is growing for services like the selling of data to game producers. These data allow third parties to understand

the way players navigate through games, and increase the user experience. As for car insurance companies and marketers use driving habits and machine data from connected cars to adjust risk-pricing models and fee structures.

One of the biggest changes of Digital is thus the transition from a product and competition based economy to a service and customer journey focused reality. Companies have to manage and make sense of data to provide services. This is a major change for product focused companies. The opportunity, from a service perspective, is to rethink the way to satisfy a customer's need through a service, especially as many services that can be created through the Digital World were not there before (Edelman, 2016).

The Digital Age is thus very present. Modern society is in the midst of it. At its core, it is fundamentally changing the value proposition to and from the customer. The question is thus not anymore how to prepare for it, but how to live and work in it.

Anticipation and Adaptability

Before analyzing strategy development or actionable translations, it is important to agree on what Digital really is. For some, it is about adding high tech niches to the portfolio. Others see it as collecting zealous amounts of data, being in constant communication with customers, or completely reshaping the company.

In practice, it is a dynamic sequencing of each of the above. It is a new way of thinking about service and product delivery, about organizing a company, about understanding and interacting with the larger ecosystem - and about creating value.

Digital Transformation can be defined as a process in which an organization is shifted to new ways of working and thinking with digital and social technologies. It involves a change in leadership, a different thinking, the encouragement of innovation and new business models, and an increased use of technology to improve

the experience of an organization's internal and external customers (Terrar, 2015).

Digital Transformation is thus not really about technology itself, but about how companies integrate it to transform their businesses and shape the way of working.

With data collection happening more through apps, for example, people can be seen as customers that use the apps' services. But as data generated by people are increasingly managed by the people themselves, consumers are, for instance in healthcare, turning into suppliers. This leads to the obligation to rethink products or develop new services (De Brouwer, 2015). Similarly, the supply chain is needed to efficiently deliver the correct product. At the same time, process KPIs and user data can be used to build insights about customers, and continuously fuel improvement for production, marketing or transportation.

Embarking on Digital is thus about shaping a dynamical way of working. With a clear strategy, skilled people and powerful data, digitally activated companies should get a fine-grained sense of the interactive relations between customers and their ecosystem, take proactive decisions based and, ultimately, adapt their business models to shape people's journey-focused experience.

Strategic Vision: Models and Methodologies

Organizations are facing a major challenge to reinvent themselves. To help them, professionals have developed tools that make this Digital process more tangible and the outcome more actionable.

In the Digital Transformation process, a clear baseline reference can be built based on the Digital Quotient. This metric for the Digital Maturity of a company has been developed by consultancy company McKinsey, based on an evaluation of 18 practices related to digital strategy, culture and capabilities, in 150 companies around the world (Catlin, Scanlan, & Wilmott, 2015).

The Digital Quotient is driven by four main components.

- **Strategic Commitment:** Companies must not go for digital as an add-on of their core business, but by wholeheartedly committing to a clear strategy.
- **Development of Digital Capabilities:** Digital success depends on the ability to invest in the relevant capabilities. They need to be sized for and aligned with the strategy.
- **Adaptive Culture:** In complement of technical capabilities such as Big Data analytics and the like, a strong and adaptive culture helps to compensate for a lack of them.
- **Internal Coherence:** Organizations need to align their internal structure, talent management and key performance indicators with the chosen Digital strategy.

Companies that desire to launch the transition to becoming Digital-ready can shape their roadmap by making choices with regards to the components of the Digital Quotient. This vision is the foundation for future actions and the management with a clear sense of direction.

Various frameworks are available to make the strategy actionable. A holistic approach to structure it from conception to realization has been developed by Jo Caudron and Dado Van Peteghem (2014). The 'Digital Transformation Modeling' methodology aims at clarifying the potential impact of Digital on an organization's activities and builds an action plan based on these insights.

In essence, the process consists of five steps:

Step 1 - Creating Insights:

- Understanding that a company can be impacted by and needs to prepare for Digital Transformation. This is traditionally initiated by one of the company's stakeholders.

Step 2 - Assessing Potential Impacts:

- Assessing impacts is done by evaluating the readiness of the company with regards to seven Drivers for Impact, and their respective constituents.

- Each of the Drivers for Impact is represented by an analogy: The Glasshouse, The Package, The Frog, The Gatekeeper, The Traveler, The Participant, The Cyborg⁷.
- The 'Cyborg', for instance, represents the impact driven by a series of technological evolutions, like the growing automation, the possibilities coming with real time data power, or additional potential offered by the Internet of Things.
- The 'Frog' materializes the fact that evolution does not have to be gradual anymore, but can be driven by shortcuts. The speed of insights, for example, can allow much faster reactions. Virtualization fits in this logic as it allows companies to jump over competitors more easily.
- Figures 1 and 2 illustrate the evaluation of the Drivers of Impact for an existing Glass Manufacturing company, and an existing Consultancy company. The evaluation has been done companywide. The former has more medium risks (amber). The latter has more risk factors under control (more black/green, less amber), but more risks with high impact (red).

Step 3 - Developing Scenarios:

- By developing scenarios, the reality resulting from the combination of two risk factors is evaluated. A risk factor is one of the main Drivers of Impact, or one of its major constituents.
- This leads to four possible scenarios: risk factor 1 and risk factor 2 have a high impact, 1 high and 2 low, 1 low and 2 high, 1 low and 2 low. In practice, this results only in three useable scenarios, as the last one offers little added value compared to the existing situation.
- Each combination represents a high risk environment that could become

Figure 1. Evaluation of the drivers of impact for an existing glass manufacturing company

The GlassHouse	The Package	The Frog	The GateKeeper	The Traveller	The Participant	The Cyborg
Transparency	Monolithic vs Atomic	Bypassing	New gatekeepers	Place	Community	Internet of Things
Accountability	APPlification	Virtualization	Recommendations	Utility	Gamification	Wearables
Proximity	Long Tail	Fragmentation of Touchpoints	"Good is good enough"	Empowered Self	Collaboration	Quantified Self
Responsiveness	Personalised		Ambassadorship	Instant Gratification	The power of the crowd	Robotics
Bi-directional	Experience			New way of working		Data
Humanization	Speed			Contextual Awareness		Monitoring
Authenticity	Value for Money					
	(Self-) Service					
	Scalability					
Team(s)/Departement(s)/Division(s) concerned by the assesment: All Glass Divisions						
White: has 0 impact in a digital context.	Green: has an impact (in a digital context), plan is defined and actions in motion.	Amber: has an impact (in a digital context), plan is existing but nothing in motion yet.	Red: has an impact (in a digital context), but no plan or action put on paper.			

reality when both conditions are combined (or an existing risk reality for which a solution needs to be found).

- For each scenario a headline should be provided (Figures 3 and 4), as well as a detailed report of how to tackle the risks.

Step 4 - Building Business Case(s):

- Based on the scenarios, actions need to be defined to avoid such a scenario to happen (or to turn the risks in opportunities). This will result in the building of one or more Business Cases.
- Once the Business Case is approved, the 'traditional' transformation life-cycle starts, at the organizational level, process level, IT level, or combined at multiple levels.

Step 5 - Staying Alert Through Trend Watching:

- To ensure a lasting effect of the awakening created by the start of the Digital journey, it is important to stay alert through structural trend watching and periodic workshops.

Empowering Digital People

The Business Case might reveal gaps in terms of technical power or human capabilities. Whereas the data side can be managed through technological gear-up and anticipative sizing, human gaps have to be tackled by developing transversal capabilities. This means focusing on customer value, data experimentation, ecosystem collaboration, Digital platforms, and maturity (Viaene, 2014).

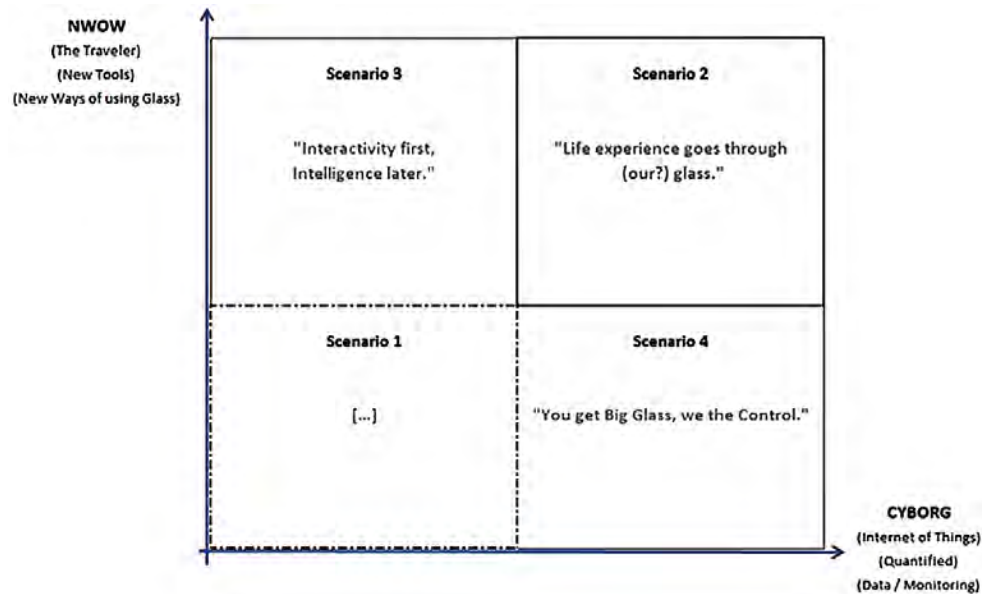
To support the transformation, pockets of catalytic potential have to be foreseen (Janssens, 2016). The ideal people have a healthy level of 'functional craziness' and are seasoned to think beyond the traditional silver lining. At the same time, companies need to be aware that skills disruption is an urgent concern in the Digital Age (World Economic Forum, 2016). Consequently, companies need to look for people for their proverbial Digital SWAT Team, and, increasingly, for people with a high adaptability.

To ensure compatibility, organizations need, firstly, to create a stimulating environment. Methods to do so are available in spades, going from Google Ventures' proven techniques (Brillantes,

Figure 2. Evaluation of the drivers of impact for an existing consultancy company

The GlassHouse	The Package	The Frog	The GateKeeper	The Traveler	The Participant	The Cyborg
Transparency	Monolithic vs Atomic	Bypassing	New gatekeepers	Place	Community	Internet of Things
Accountability	APPification	Virtualization	Recommendations	Utility	Gamification	Wearables
Proximity	Long Tail	Fragmentation of Touchpoints	"Good is good enough"	Empowered Self	Collaboration	Quantified Self
Responsiveness	Personalised		Ambassadorship	Instant Gratification	The power of the crowd	Robotics
Bi-directional	Experience			New way of working		Data
Humanization	Speed			Contextual Awareness		Monitoring
Authenticity	Value for Money					
	(Self-) Service					
	Scalability					
Team(s)/Departement(s)/Division(s) concerned by the assesment: All Consultancy Practices						
White: has 0 impact in a digital context.	Green: has an impact (in a digital context), plan is defined and actions in motion.	Amber: has an impact (in a digital context), plan is existing but nothing in motion yet.	Red: has an impact (in a digital context), but no plan or action put on paper.			

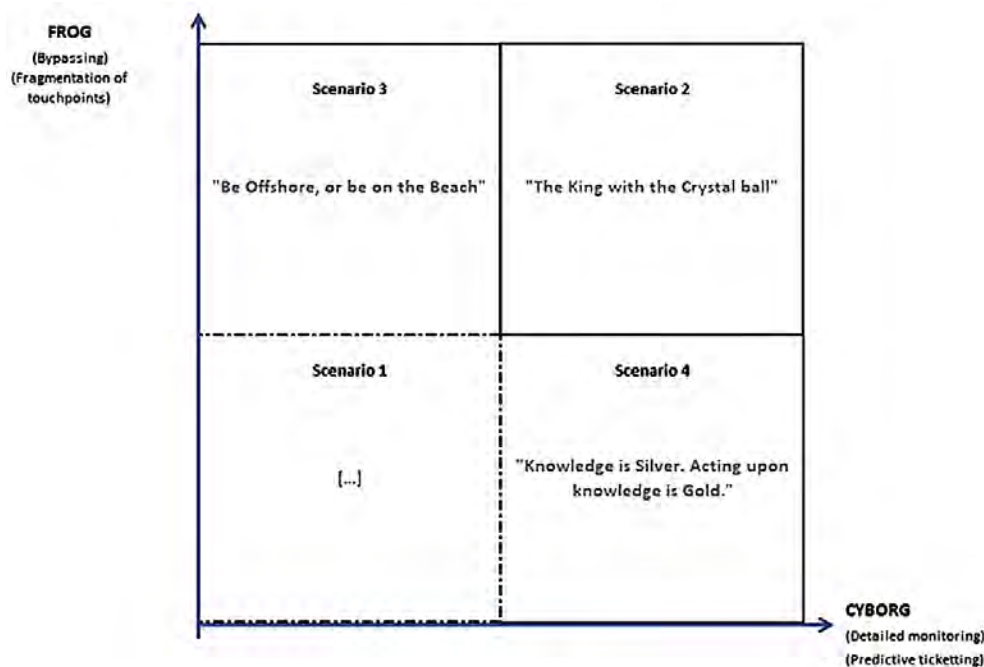
Figure 3. Example of a scenario evaluation for a glass manufacturing company, based on the risks coming from a new way of working, and major cyborg elements



2015; Knapp, Zeratsky, & Kowitz, 2016) to more anarchical approaches (Bushnell & Stone, 2013). As such, it comes down to creating a culture where thinking beyond the traditional frames of reference is stimulated.

In complement, it can be fruitful to see if hidden gems were discarded due to traditional analysis. At Atari, this was part of the corporate DNA. They would periodically list the worst ideas of the past period, and brainstorm to turn them into

Figure 4. Example of a scenario evaluation for a consultancy company, based on the risks coming from the frog, and major cyborg elements



good ones. At worst, it offered precious lessons. At best, it resulted in a huge success (Bushnell & Stone, 2013). Google is following a comparable approach to turn insights of past (failed) projects into ingredients for new ones.

Secondly, companies and organizations need to nurture their people for and with Digital.

It is important to retain high performers as they have a beneficial influence on results and potentially on other colleagues. Similarly, by doing predictive HR management, employee expectations can be anticipated and employees can be identified that need mentoring. Google's People & Innovation Lab, for instance, uses data-based people management to anticipate employee aspirations and to monitor impacts of individual relations on team performance. Even if not every company has the same data power, proactive management benefits the internal dynamic and the outcome on bottom line Digital productivity.

Thirdly, the different levels of management need to learn to talk Digital. Leaders need to be able to exchange with their experts, ask them

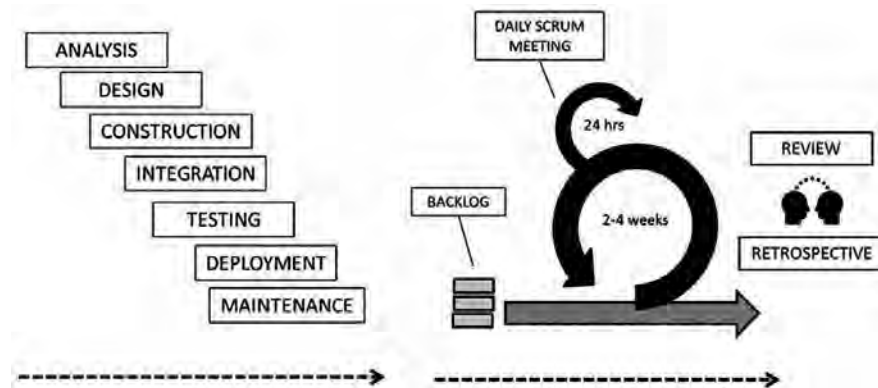
questions and, at least, understand the essence of their answer. This will reinforce people's faith and contribute in staying alert for continuous reinvention and adaptability.

Overall, having the appropriate human capital on board is thus one of the cornerstones to manage digitally. This implies creating an appropriate environment and culture for these people, nurturing them, and managing them with a solid understanding of what Digital means for them.

Dynamic Feedback and Agile Projects

The biggest challenge for the execution of the Digital Strategy is the structural integration of the fundamental paradigm shift. Pre-Digital organizations focused on market share and people got managed accordingly. In the Digital Economy, companies need to be designed for reactivity, flexibility and agility. An integrated operational backbone is needed that embraces a continuously

Figure 5. A fair share of the nimble way of organizing the work for reactivity and agility is coming from agile software development. Collaboration with tradition waterfall approaches might be needed.



evolving service backbone, and a clearly articulated, but dynamic business strategy (Ross, 2016).

A fair share of this nimble way of working is initially coming from agile software development, focusing on rapid prototyping and getting releases out smoothly. This philosophy is now also influencing the development of business deliverables. For details on agile approaches, the author refers to dedicated literature⁸. But some important attention points should be taken into account.

Firstly, it should be noted that the structural fluidity associated with agile methodologies requires maturity and discipline. Scope control and regular iterations are part of the Agile principles. Success comes thus not from the speed of idea generation alone. Rather, it is the unconditional availability of decision takers towards rapid feedback that makes the difference.

Secondly, the feedback and follow-up loops need to be managed with conviction. No matter if they are scrum or Kanban inspired, reviews need to aim at real quality. Indirectly, this implies that baselines and reference points remain important to fuel further improvements.

Thirdly, to fit in the larger corporate ecosystem, agile perimeters might have to blend with more traditional ones (see Figure 5). To make such a hybrid context function, compromises are needed. In case agile subsets are, for instance, inserted in a larger waterfall based environment⁹, they will have to work with a smaller degree of autonomy

and with a different perspective on due dates. Waterfall entities, on the other hand, will have to adapt to a more flexible way of working.

Similarly, a critical mind is needed with regards to theories advocating radically simplified approaches, like replacing strict controls by minimal or no budgets, or intuitive prioritization (Laloux, 2014). Digital companies still need guidance. The ‘simplification’ should therefore be limited to a leaner way of working, rather than going for a full deconstruction of supporting frameworks.

Most importantly, the overarching focus on agility should not become a goal on itself. In a reality where decisions have to be made quickly based on continuously evolving insights, companies should be able to develop their quantified intuition, and take decisions that make a difference. The aim of agility is therefore to support the overall goal: providing appropriate solutions to customer needs and preferences.

FUTURE RESEARCH DIRECTIONS

Algorithms power businesses like Amazon and Netflix. Health related apps are changing the medical world. And research about industries like hotels and retail indicate that benefits of skilled sales force may be diminishing in an increasingly data-driven world (Philips & van Ryzin, 2016).

Digital Transformation is thus providing society with substantial opportunities, as well as with challenges in the dynamic relation between agile organizations, rapid decision taking and the role of Big Data powered artificial intelligence.

This raises the question whether machines or people should take decisions to provide the necessary value in the Digital Age. At the same time, it indicates that specialists and managers need to be able to sharpen their Digital Capabilities and elevate their quantified intuition to a higher level, in order to remain in control of their organization. To do so, several elements could be further explored.

From the methodological point of view, the project management toolbox has to be further enriched with specific best practices. They will have to dedicate specific attention to the permeable delimitation between product delivery and service delivery. Likewise, cook books for hybrid contexts would be valuable. This will not dilute the role of project management frameworks and dynamic organizational management, but will improve compatibility with the Digital Reality.

Likewise, it will be valuable to analyze if best practices could be identified for specific sub-sectors. Also, attention to the public sphere would be valuable. By bringing public and private practices on par, it is imaginable to activate the potential of the public-private force field in domains like education, health or city management. In turn, this could benefit corporate and individual ecosystems.

Obviously, best practices will require continuous improvement. But the author believes that recurring efforts will be largely compensated by advantages in quality and outcome.

CONCLUSION

Managing the Digital Transformation is not a high-tech option for a faraway future. It is about how public and private organizations integrate technology to transform their businesses. It involves a change in leadership, a different thinking, new business models, and increased use of technol-

ogy to improve the customer experience. It is by integrating this backbone that Digital strategies can have a durable impact.

To get there, a series of key points have been addressed.

Firstly, it has been explained that a thorough understanding of the new reality is essential. The Fourth Industrial Revolution redefines borders between reactivity and proactivity, and creates the need for real-time service and contextual adaptation. It requires a dynamic silver lining, compatible with interactive work clusters and a hybrid organization.

It has been explained that this understanding needs to be integrated a strategy to profoundly revisit the company, encompassing the internal way of working, as well as the customer value proposition, the articulation of the product/service portfolio and the related follow-up dynamics.

Secondly, practical approaches have been analyzed that can guide this process. With the Digital Quotient, organizations have a baseline to put findings into perspective. Similarly, by embracing the Digital Transformation Modeling framework, a solid methodology is available to clarify the potential impact of Digital, create future scenarios for the Digital Age, and build an insight-based action plan.

Thirdly, it has been highlighted that putting a strategy into motion for Digital success implies an appropriate organizational way of working and managing.

Lastly, it has been explained that the Digital way of working needs to be inspired by an agile philosophy. When doing so, organizations and managers are namely better prepared to develop their quantified intuition, and take relevant decisions for the company and its customers.

Overall, Digital Transformation projects are thus part of a major change journey. A journey that shapes the present to be ready for an already ongoing future - and that will continue to transform at high speed. It is a new way of thinking about the interaction with customers, about shaping service and product delivery, about organizing an organization - and about creating value.

REFERENCES

Brillantes, K. (2015). *Hiring lessons from GV's design event series*. Retrieved March 7, 2016, from <https://library.gv.com/>

Carter, S. (2015). *The Digital Disruption has already happened*. Retrieved February 13, 2016, from https://twitter.com/sandy_carter

Catlin, T., Scanlan, J., & Willmott, P. (2015). Raising your Digital Quotient. *The McKinsey Quarterly*, (June): 2015.

DeBrouwer, W. (2015). *The Era of citizen Doctors*. Retrieved February 25, 2016, from <https://www.scanadu.com/blog/the-era-of-citizen-doctors>

Edelman, D. (2016). Achieving a digital state of mind. *The McKinsey Quarterly*, (February), 2016.

Henke, N., Libarikian, A., & Wiseman, B. (2016). Straight talk about Big Data. *The McKinsey Quarterly*, (October): 2016.

Mayhem, H., Saleh, T., & Williams, S. (2016). Making data analytics work for you—instead of the other way around. *The McKinsey Quarterly*, (October), 2016.

Nanterme, P. (2016). *Digital disruptions has only just begun*. Retrieved February 3, 2016, from <http://www.weforum.org/agenda/2016/01/digital-disruption-has-only-just-begun>

Philips, R., & van Ryzin, G. (2016). *Dealing in Data*. Retrieved February 12, 2016, from <https://www8.gsb.columbia.edu/ideas-at-work/publication/1673>

Richter, G., & Wee, D. (2016). Steering IT into the digital manufacturing era. *The McKinsey Quarterly*, (October): 2016.

Ross, J. (2016). *Digital Disruption: Transforming your Company for the Digital Economy*. Retrieved February 4, 2016, from <http://cdn.executive.mit.edu>

Terrar, D. (2015). *What is Digital Transformation?* Retrieved March 8, 2016, from <http://www.theagileelephant.com/what-is-digital-transformation/>

Turner, V., Reinsel, D., & Gantz, J. F. (2014). *The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things*. Retrieved September 3, 2014, from <http://www.emc.com/leadership/digital-universe/index.htm>

Van Leemputten, P. (2016). *Les big data jouent avec votre cerveau (et paient vos impôts)*. Retrieved February 29, 2016, from <http://datanews.levif.be/>

Viaene, S. (2014). Zorg ervoor dat je technologie mee kan. *Business Bytes*. Retrieved May 15, 2015, from <http://business.telenet.be/nl/artikel/zorg-ervoor-dat-je-technologie-mee-kan>

World Economic Forum. (2016). *The Future of Jobs: Employment, Skills and Workforce Strategy for the Fourth Industrial Revolution*. Global Challenge Insight Report. Geneva, Switzerland: World Economic Forum.

ADDITIONAL READING

Aggarwal, A. (2016). *Managing Big Data Integration in the Public Sector*. Hershey, PA: IGI Global. doi:10.4018/978-1-4666-9649-5

Bushnell, N., & Stone, G. (2013). *Finding the Next Steve Jobs: How to Find, Hire, Keep and Nurture Creative Talent* (pp. 83–87). London, United Kingdom: Headline Publishing Group.

Caudron, J., & Van Peteghem, D. (2014). *Digital Transformation*. Tielt, Belgium: LannooCampus.

Fitzgerald, M., Kruschwitz, N., Bonnet, D., & Welch, M. (2013). *Embracing Digital Technology*. Cambridge, MA: MIT Sloan Digital Review.

Galen, R. (2013). Scrum Product Ownership: Balancing Value from the Inside Out, Cary, NC: RGCG, pp.19-22, 209-221.

Gartner (2014). *Taming the Digital Dragon: The 2014 CIO Agenda*. Stamford, CT: Gartner.

Higgins, J. (2015). *Digital Transformation of European Industry and Enterprises*. Brussels, Belgium: DG Internal Market, Industry, Entrepreneurship and SMEs.

Hinssen, P. (2014). *The Network Always Wins. How to Win in the Age of Uncertainty*. Ghent, Belgium: Mach Media.

Janssens, J. (2016). Blending Technology, Human Potential, and Organizational Reality: Managing Big Data Projects in Public Contexts. In Aggarwal, A. (2016). *Managing Big Data Integration in the Public Sector* (pp. 107–137). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-9649-5.ch007

Knapp, J., Zeratsky, J., & Kowitz, B. (2016). *Sprint: How to Solve Big Problems and Test New Ideas in Just Five Days*. New York, NY: Simon & Schuster.

Laloux, F. (2014). *Reinventing Organizations A Guide to Creating Organizations Inspired by the Next Stage in Human Consciousness* (pp. 105–114). Leuven, Belgium: LannooCampus.

Schwab, K., & Samans, R. (2016). *The Future of Jobs: Employment, Skills and Workforce Strategy for the Fourth Industrial Revolution*, Global Challenge Insight Report, Geneva, Switzerland: World Economic Forum, p. V.

Van Driessche, W. (2014). *Modern Minds - Kan uw hoofd de 21ste eeuw aan?* Brussels, Belgium: Mediafin.

KEY TERMS AND DEFINITIONS

Agile: Project management methodology in which the development is characterized by the breakdown of tasks into short periods, with frequent reassessment of work and plans. Used in software related projects and digital transformation activities.

Big Data: Large volumes of structured, semi-structured and unstructured data from one or more sources, created and analyzed through innovative forms of information processing to develop insights on patterns, especially on human behaviour and corporate interactions, resulting in better decisions, strategic business decisions, and process adaptation. Characterized by high volume, high velocity and/or high variety.

Digital Quotient: Metric of the digital maturity of a company, based on the evaluation of a series of practices related to digital strategy, capabilities, and culture.

Digital Transformation: Process in which human and corporate society is shifted to new ways of working and thinking with digital and social technologies. Involves a change in leadership, a different mindset, the encouragement of innovation and new business models, and an increased use of technology to improve the experience of internal and external customers.

Digitization: Process in which an object, image, sound, document or signal are represented by a series of numbers that describe a discrete set of its points. Often used as synonym for the larger process of Digital Transformation.

Fourth Industrial Revolution: Industrial revolution driven by systems involving entirely new capabilities for people and machines. Represents new ways to embed technology in society, and induces new ways of working and thinking for human and corporate matters. Used as synonym for Digital Transformation.

Quantified Intuition: Non-technical capability that helps in making impactful decisions at a fast rate, despite a high level of uncertainty and an important inflow of interconnected data.

Scrum: Iterative and incremental product development framework used in agile projects.

Waterfall Project Methodology: Sequential project management methodology, in which project progress is regarded as a downwards process. Originally described as consisting of phases for Requirement Specifications, Design, Construction, Integration, Testing, Installation

and Maintenance, variations exist on the naming and number of phases.

ENDNOTES

- ¹ For more information on Big Data, the author refers to the 'Key Terms and Definitions' section.
- ² For more information on quantified intuition, the author refers to: <http://www8.gsb.columbia.edu/>
- ³ For more information on international attention for Digital Transformation, the author refers to: <http://www.digitaltransformation2016.eu/> and <http://www.weforum.org/events/world-economic-forum-annual-meeting-2016>
- ⁴ For more information on scrum, the author refers to: <https://www.scrum.org/> and <https://www.scrumalliance.org/>
- ⁵ For more information on Kanban, the author refers to: <https://www.atlassian.com/agile/kanban>
- ⁶ For more information on the Behavioral Insights Team, the author refers to <http://www.behaviouralinsights.co.uk/>
- ⁷ For a detailed explanation of each driver and the related analogy, the author refers to Caudron & Van Peteghem (2014).
- ⁸ For more information on Agile, the author refers to <https://www.scrum.org/> and <https://www.scrumalliance.org/> (scrum), and to <https://www.atlassian.com/agile/kanban> (Kanban).
- ⁹ For more information on waterfall methodologies, the author refers to: <http://www.pmi.org/PMBOK-Guide-and-Standards.aspx> (PMBOK), <http://www.prince-officialsite.com/> (Prince2).

A Framework for Exploring IT–Led Change in Morphing Organizations

Sharon A. Cox

Birmingham City University, UK

INTRODUCTION

Organizations must continually change to survive, adapting to changes in their social, economic and trading environments, responding to the changing needs of customers and reacting to competition. Organizational transformation refers to changes in the way in which an organization operates. Morphing organization is a term used to reflect organizational transformation, recognizing that as an organization changes its outer face displayed to the world, internal structures and processes within the organization must change to facilitate and support the external interface. A morphing organization has an identity that is determined by the organization's relationship with its environment, its internal components and the social practices in which the participants in the organization interact. As the external face of the organization changes and internal components are adjusted, core structures and values need to be retained in order to maintain the organization's unique identity (Cox, 2014).

Developments in information technology (IT) have been and continue to be, a major driver of organizational transformation. IT has changed the structure and face of organizations in most industries and it is recognized that introducing and changing technology in an organization is not merely an IT issue (Jackson & Harris, 2003). For example, the impact of ebusiness transformation of organizations is well documented. Technology has changed the way in which customers and suppliers interact with the outer face of the organization, and transformed the back office pro-

cesses needed to support the external face of the front office activities. The wider implications of changing IT systems in organizations as an enabler of organizational change need to be considered.

The contribution of this chapter is the presentation of an organization architecture to explore the impact of IT-led change in morphing organizations. The framework of the organization architecture offers a coherent structure in which to consider organizational transformation in response to IT-led change, to structure analysis and evaluation of the aspects of an organization that may be affected by proposed changes to IT systems in an organization. The framework may also be used retrospectively to analyse how changes in IT have led to organizational transformation. The organizational architecture builds on the McKinsey 7-S framework (Peters & Waterman Jr., 2004) and considers the impact of IT changes to both formal elements of the organization such as organizational processes and informal elements such as social values. The application of the organization architecture is illustrated through case study examples from UK organizations.

The following section discusses organization identity. A business model is discussed showing the relationship between an organization and its trading environment focusing on the external face of the organization. The internal structures that support the external organization identity are then discussed. An organization architecture is presented as a model to explore the changes in the organization needed to accommodate IT and how an organization may be affected by changes to IT. The chapter concludes by considering the

challenges of the third wave of IT-led change, predicted by Porter & Heppelmann (2014), to organizational identity.

BACKGROUND

Identity is the most fundamental concept of humanity and is the essence of an organization, which emerges through communication and interaction between members of the organization (Koskinen, 2015). The act of identifying an organization establishes a conceptual boundary around a group of people, resources and activities. Defining an organization separates the organization linguistically from other things (Koskinen, 2015), creating an entity with a unique identity. The relationships between the organization, its customers, suppliers, competitors and trading partners can then be explored.

Whetten (2006) uses the phrase ‘organizational identity claims’ or ‘referents’ in relation to the attributes of an organization that form its identity. These are enduring attributes that reflect the core values of an organization and are tested by considering the question of whether the organization would be the same organization without a specific attribute.

In the same way that the view an individual holds about themselves may differ from the way that others see them, there is the identity of how others see the organization and how those inside the organization perceive the organization. Balmer (2008) uses the term *corporate identity* to reflect an external view of how the organization is perceived by those outside the organization, such as customers and the term *organizational identity* to reflect the internal view of how the organization is perceived by those within it. Koskinen (2015) uses similar terms to represent these two views and adds a third form of identity:

- *Company identity* is how the organization presents itself to its environment.

- *Reflective identity* is how the organization sees itself internally.
- *Substantive identity* is the mechanism that keep the different parts of the organization focused as a single unit.

B

An organization’s identity is created by the decisions and practices in an organization, which in turn influence the identity. Koskinen (2015) describes identity as the hidden face of strategy. Accurate identification of organizations is essential (Whetten, 2006) and attention to the organizational identity is particularly needed during periods of change (Albert & Whetten, 1985). The following sections discuss two models: the business model and organization architecture that reflect the *company identity* and internal *substantive identity* respectively.

Business Model

Organizations operate within an external trading context that is represented in a business model. All organizations have a business model (Andersén *et al.*, 2015). A business model is a way of doing business focusing on the organization’s competencies (DaSilva & Trkman, 2014), reflecting the relationship between control and value in the organization (Ballon, 2007). Business models identify key aspects of an organization that determine how the organization positions itself in its market(s) and industry(ies) in order to create, maintain and reflect its unique identity.

Osterwalder *et al.*, (2005) identified nine elements in a business model:

1. Value proposition (products and services offered by the organization).
2. Target customer segments.
3. Distribution channels (used to reach customers).
4. Relationship (links between organization and customer segments).
5. Value configuration (arrangements of activities and resources)

6. Core competency (of organization exploited in business model).
7. Partner network (co-operative agreements with external parties).
8. Cost structure.
9. Revenue model (how it makes money).

Most of the elements identified by Osterwalder *et al.*, (2005) focus on the external relationships of the organization. Koskinen (2015) suggests that although business models represent the external relationships of an organization, business models are not sufficient to understand the history and evolution of the organization, which contribute to the organization's identity. Although a business model may include some referents of company identity, it is insufficient for examining the organization's substantive identity and how the internal context will be affected organizational transformation.

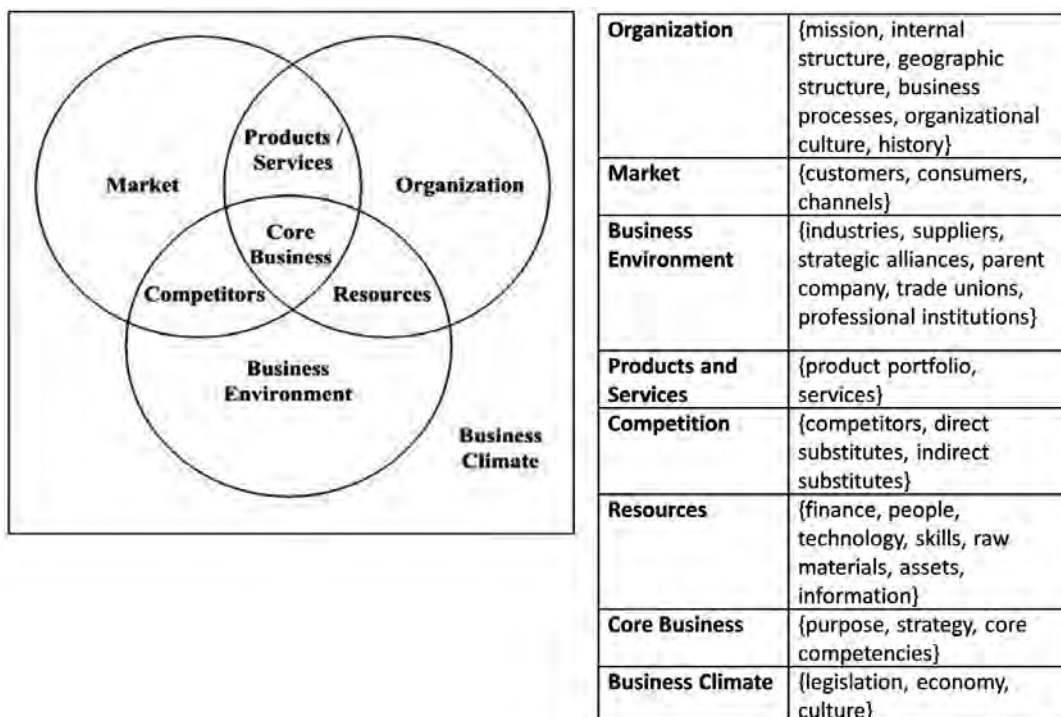
Figure 1 shows a generic business model (Cox, 2014) comprising three sets, the organization,

market and business environment situated within the wider business climate.

The business model provides the context for the organization's identity and identifies the external stakeholders, which may provide opportunities and constraints for the organization achieving its vision. The organization architecture forms the context for substantive identity within the organization set of the business model and provides the contextual framework for activity in the organization; changes to the organization architecture trigger and facilitate organizational transformation. Organizational change may be initiated in response to external changes in the market, business environment or business climate, for example, changing customer expectations, changing industry standards or legislative requirements. Organizational change may also be internally initiated, for example, by new management.

The performance of an organization and the outcome of organizational change emerge from the interaction between the organization's struc-

Figure 1. Generic business model
Source: Cox, 2014



tures, individuals within the organization and the organizational strategies, which form the context for technology use (Yeo & Marquardt, 2015) in the organization architecture. Much of the research about organizational design focuses on organizational structures (Sherif *et al.*, 2013). For example, Ethiraj & Levinthal (2004) consider organizational design based on two choices: the number of departments and the assignment of functions to departments. Organizations need to have sufficient flexibility to enable them to respond to the demands and opportunities of innovations in IT, balanced with stability to maintain their unique identity. Sherif *et al.*, (2013) describe an ambidextrous structure as one that allows organizations to move between organic and mechanic organizational structures depending on market conditions.

Organization Architecture

Business models focus externally on the relationship between the organization and its trading environment, in contrast the organization architecture focuses internally on the structures and components that maintain the external organization identity.

The organizational architecture is different to the organizational structure. Sherif *et al.*, (2013) describe an organization architecture as a high level map of the organization that includes people, processes and technology. The formal and informal organization structure form part of the organization architecture. The organization structure groups resources, provides a boundary for activities to be controlled, and determines the flow of information, authority and decision-making responsibility through the organization. An organizational architecture creates behavioural space, which provides opportunities and constraints for organizational activities. The architecture provides a boundary for activity, shaping the organization. Saucer & Willcocks (2004) emphasize that defining the organizational architecture is not a one-off task (like designing a building); it is a continuous process

of adjustment to meet the changing needs of the business environment.

Sherif *et al.*, (2013) suggest that the organization architecture includes three architectures:

- Business architecture: organizational capabilities (services, processes and information) translating the strategic objectives into processes and competences.
- Infrastructure architecture: technical capabilities (hardware and telecommunications) supporting business processes
- Application architecture: governance capabilities monitoring and controlling activities.

Waterman Jr. *et al.*, (1980) proposed that organizational effectiveness emerges from the dynamic interactions between the organization's structure, strategy, systems, style (of management and culture), skills, staff resource (including morale, attitude, behaviour, training and appraisal) and subordinate goals (later renamed shared values (Peters & Waterman Jr., 2004) representing the values underlying the mission of the organization). This became known as the McKinsey 7-S framework©. Each area of the 7-S framework is connected to the others (Kaplan, 2005), demonstrating that change in any of the seven areas will trigger change in the other areas.

Organization Change

Changes in an organization can range from minor changes to improve efficiency through to the complete destruction and redesign of structures and systems (Senior & Swailes, 2010). Ethiraj & Levinthal (2004) differentiate between first-order change (incremental change within the existing organizational structure) and second-order change to the underlying structures to align with environmental demands. Dunphy & Stace (1993) identify four types of change within this broad spectrum of minor modification to major redesign. Each type of change differs in terms of the scope and

scale of the change in the organization. First, fine tuning involves minor adjustments, which may be restricted to one department and may be internally triggered, to improve organizational alignment of strategy, structure, people and processes. Second, increment adjustment may affect one or more departments with the aim of improving the operation of the organization by modifying strategies, structures and management processes in response to changes in the external environment. Third, modular transformation relates to the restructuring of one or more departments and changing the co-ordination of activities between departments. Fourth, corporate transformation involves revolutionary change of the strategy, structures and systems in the organization, requiring wide scale redesign of the organization, challenging organization identity. Organizational change is often triggered by IT.

IT-LED CHANGE

Organizations must adopt and embed information technology into their social practices in order to meet the changing demands and expectations of their trading environment. IT can lead to structural changes in organizations (Goksen *et al.*, 2015) as technology can be used to change the flow of information, control and decision-making. Organizational participants respond to IT in different ways, leading to change in the organization architecture to accommodate the technology and posing threats to the organization's identity.

IT can constrain or enable change in organizations (Yeo & Marquardt, 2015). IT has been deliberately used as a vehicle to initiate cultural change in organizations, for example, e-government has been introduced with the aim of transforming public sector services (Yeo & Ajam, 2010). However, the introduction of IT or changes to IT, can lead to both intentional and unintentional consequences in the organization (Yeo & Marquardt, 2015).

The design of technological systems embodies business rules that constrain actions in an organization, though constraints can be resisted and lead to unintentional consequences. For example, Boudreau & Robey (2005) discuss a case where a feature of the system was that it would automatically log-off users who did not interact with the system for a defined period in order to increase security. Users did not like the lack of freedom to leave the system unattended, partly because of the length of time taken to log-in. Users therefore devised ways to "beat the system" such as by asking others to simulate actions for them (Boudreau & Robey, 2005). Although the intention was for the system feature to increase security, an unintended consequence of the system was that it decreased security as staff developed ways of working around the system controls.

Figure 2 presents a 2x2 grid to reflect the positive and negative impacts of IT on organizations. The grid forms four sectors:

Sector A: As in the log-in example (Bordreau & Robey, 2005), technology can restrict actions, which may result in unintended consequences as staff seek ways to work around the restrictions.

Sector B: In contrast, positive side effects can emerge from technology through facilitating social improvisation and invention (Boudreau & Robey, 2005).

Sector C: Positive planned improvements to organizational performance can emerge. For example, the electronic submission and retrieval of documents has been used to improve service delivery (Yeo & Marquardt, 2015).

Sector D: Intentional imposed restrictions and monitoring controls can be implemented to impose standards and formalize practices. For example, in one case the requirement to store information in shared folders improved information access and also led to opportunities for collaboration through a virtual community (Yeo & Marquardt, 2015).

Figure 2. Impact of IT on organizations

Change	Enable	B: Invention	C: Improvements
	Inhibit	A: Workaround	D: Restrictions
		Unintentional Consequences	Intentional Consequences

The unintentional consequences arising from technology may result from both intended and unintended uses of technology by actors in the organization. For example, the inventive use of data entry fields overcame perceived limitations of the system but the misuse of data entry fields also leads to data quality issues (Bordreau & Robey, 2005).

Yeo & Marquardt (2015) suggest that enactment with technology leads to both innovation and disruption in work practices, which will impact organizational structure, strategy and performance. Technology can also lead to inequality and exclusion of some workers in the organization. The potential impact of technology adoption on organizations needs to be assessed to inform decisions about whether to implement the technology. Orlikowski (2007) criticizes studies of technology adoption that either adopt a techno-centric perspective of how technology impacts social practices, ignoring how technology is given meaning within a specific context, or adopt a human-centric perspective of how agents make sense of technology but implement a restricted view of technology. Orlikowski (2007) uses the term sociomateriality to refer to the emergent relationship between technology and human actors in the organization reflecting how technology loses its objective meaning as it becomes embedded in the social actions and practices of the actors in the organization. As actors make sense of the technology there is an interrelationship between the task and technology; the technology changes the task and the task changes the requirements and

use of the technology (Carroll & Rossen, 1992). Actors in the organization engage in collective sensemaking of technology (Yeo & Marquardt, 2015) and it is through this enactment of technology that organizations change (Bordreau & Robey, 2005). A framework is needed to explore how technology drives change in organizations.

SOLUTIONS AND RECOMMENDATIONS

Developments in IT provide new ways for organizations to interact with customers and other external parties, changing the external face of the organization. Organizations need a means to:

- Examine how the internal structures of the organization need to be adjusted to support the changes to the organization's external face.
- Assess the potential consequences arising from the internal changes proposed.
- Reflect on the impact the changes may have on the organization's identity.

The organization architecture provides an abstraction of organizational factors that influence an organization's IT development capability (Sherif *et al.*, 2013). IT can in turn also influence these factors in the organization architecture to lead change in the organization.

Organizations change over time, going through a series of phases of change. Cox *et al.*, (2006)

examined two cases of organizational transformation. The first case examined the organizational impact on a UK manufacturer following the implementation of e-business systems enabling collaboration between the manufacturer and the major supermarket chains. The second case explored the manufacturer's engagement in e-business with smaller retailers and the impact on the retailers. From these cases, eight dimensions of organizational transformation were identified, shown in the first column of Table 1.

These dimensions were also identified in analysis of the phases of change of a typical retail organization in the UK (Cox, 2014) as it moved from:

- Counter service (where customers asked staff for goods over a counter) to self service (where customers took goods from shelves and paid at a checkout before leaving the store).

- Self service to Internet service as customers purchased goods online.
- Internet service to Internet collaboration, as customers engaged in online communication with the store.

As the organization moved through these phases of change, the dimensions of change of the organization are shown in Table 1. This third case study is significant as it revealed that non-IT related change (the move from counter service to self service in the store) affected the same dimensions of the organization as the technology-related change (such as the introduction of online retailing).

Analysis of these dimensions showed that some of the dimensions related to formal components of the organization such as its infrastructure and security, and some elements referred to informal components such as trust.

Table 1. Morphing dimensions of a retail organization

Dimension	Counter Service to Self Service	Self Service to Internet Service	Internet Service to Internet Collaboration
People	Roles redefined.	Procedural roles, lacking flexibility.	Collaborative roles working with others online.
Skills	Broader and less specialized knowledge needed.	Less personal, online communication skills needed.	Develop relationships through ability to communicate online.
Practices	Way of shopping changed for customers and staff.	Way of shopping changed for customers and staff.	Information shared through working in partnership.
Processes	Front office processes changed: sales, purchasing and payment changed.	Back office processes changed: stock control, packing and distribution.	Front office processes to facilitate collaboration and back office processes to act on data captured changed.
Trust	Customers trusted to pay for items before leaving store.	Organization trusted to provide accurate shopping data, honour transaction and keep personal data secure.	Customers and organizations trust that data are accurate and not misused.
Data	Better product labelling needed on shelves.	Accurate data needed about the product, shopping process and terms of sale.	Challenges of technical capability and semantic issues arose.
Security	Risks to product security increased as customers had direct access to stock.	Risks to information security increased and security focuses on restricting access to information and transactions rather than physical products.	Security focuses on enabling legitimate access to information and communications.
Infrastructure	Shelving and counters in store changed style and placement.	Technical infrastructure development.	Compatibility and integration of data systems required.

The findings from these case studies were compared with the components of the 7-s framework (Peters & Waterman Jr., 2004) in Table 2. Although some overlap in areas such as staff and skills was identified, the analysis of the case studies highlighted more practice-based issues from the perspective of the staff affected by the organizational change. A key component that is not included in the 7-S framework is consideration to the changes to data and information in the organization and the role of data and information in the change process, which are particularly important in IT-led change. Table 1 shows how issues of data, security and trust were identified as key components affected by IT-related changes in the organization.

A further case study was conducted exploring the impact of e-business in a UK organization (Cox, 2013). This led to further refinement of the

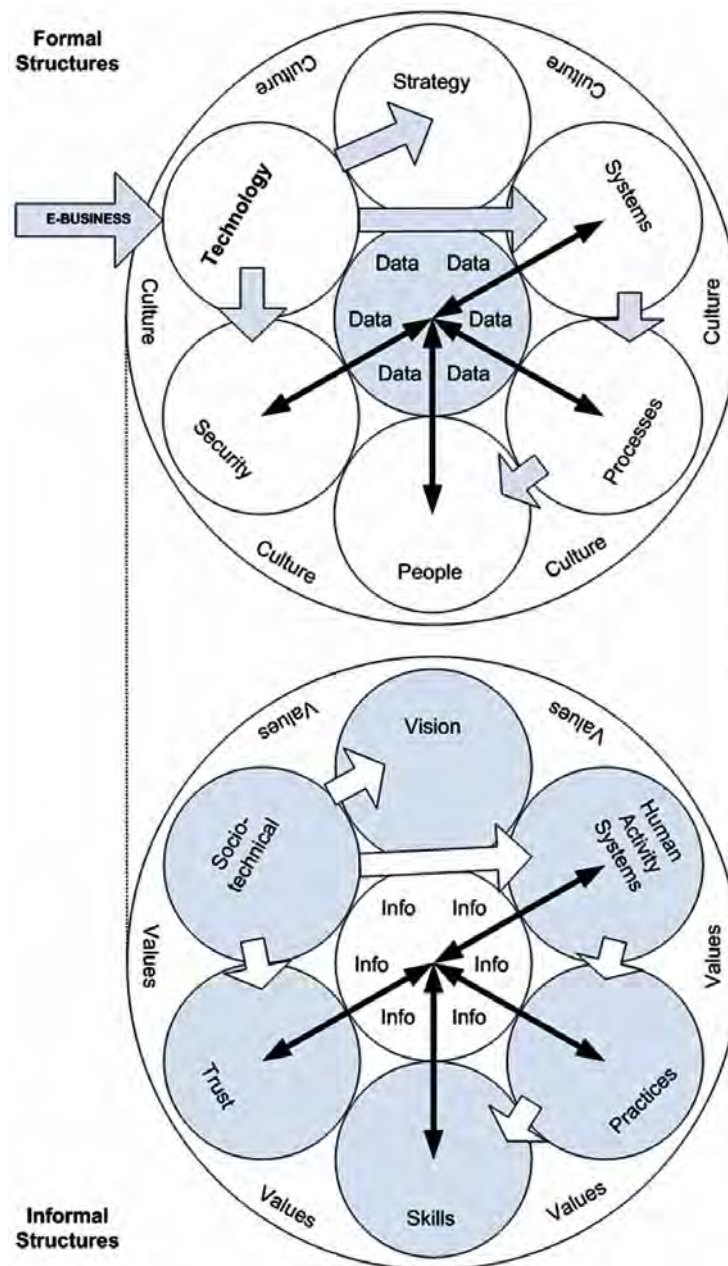
organizational components affected by e-business. This list was then used to plan the introduction of social media systems in an organization taking into account the impact on the organization. In these five cases, the organizational structure was not directly affected by the e-business and social media systems introduced into the organizations. This led to the development of the dimensions to form the components in the organization architecture in Figure 3, which includes the elements of business architecture and infrastructure architecture of Sherif *et al.*, (2013). Referents of identity are different in each organization but the organization architecture provides a general framework in which to assess organizational transformation and the potential impact on organization identity. Components of the organization architecture are embedded in the culture and values of the orga-

Table 2. Comparison of organizational components

Case Studies 1 and 2 (Cox <i>et al.</i> , (2001), Case Study 3 (Cox, 2014)	7-S Framework (Peters & Waterman Jr., 2004)	Case Studies 4 and 5 (Cox, 2013)	Formal Components of Organization Architecture	Informal Components of Organization Architecture
People	Staff	People	People	
Skills	Skills	Skills		Skills
Practices		Practices		Practices
Processes		Processes	Processes	
Trust		Trust		Trust
Data		Data	Data	
Security		Security	Security	
Infrastructure		Infrastructure	Infrastructure	
	Structure			
	Strategy	Strategy	Strategy	
	Systems	Systems	Systems	
	Style	Shared Values		Values
	Subordinate Goals			
			Technology	
			Culture	
				Vision
				Human Activity Systems
				Socio-technical
				Architecture
				Information

Figure 3. Organization architecture

Source: Cox, 2014



ization, shown in Figure 3. Change to the organization architecture can be triggered by change to any of the architecture's components. Before a component in the architecture is changed, the morphing organization will need to consider to what extent the change aligns with the organization's culture or challenges its identity.

The organization architecture provides a framework to explore how changes to one component such as technology lead to changes of other components in the organization and whether changes to a specific component are consistent with the organization's identity. For example, the introduction of social media technology changed

the organization strategy and vision in terms of how the organization related to its external stakeholders in case study 5. The introduction of social media changed communication processes and practices, increasing threats to data security requiring additional software controls to be introduced. The organization had to trust that staff would not disclose confidential information through social media. As social media was integrated into existing systems and processes as another communication mechanism, staff needed to develop new skills in using the media. One of the organization's core values was consistency; consistency in information and service. The introduction of social media challenged this value as in responding to social media messages quickly, staff mistakenly published inconsistent information to customers. Technology is situated within the culture of organizations and has to be used in ways consistent with the organization's values in order to ensure that the organization's identity is maintained.

FUTURE RESEARCH DIRECTIONS

Organizational identity is an important concept, distinct from branding, to establish relationships between the organization and external stakeholders. As trade and communication is increasingly mediated by Internet technologies, further work is needed to establish how organizational identity is reflected, perceived and maintained through online activities. Organizations have been transformed and organizational identities have been challenged by developments in IT and by the speed with which IT has become embedded into society. As technology continues to become increasingly ubiquitous and the organizational boundary becomes blurred, the context in which technology is used by actors in the organization needs to be considered.

Porter & Heppelmann (2014) describe three waves of IT-driven transformation. The first wave focused on automating activities in the value chain, leading to the standardization of processes.

The second wave focused on co-ordinating and integrating activities in the value chain. A third wave of information IT-led change is now reshaping organizational and industry boundaries. The third wave integrates IT into products sold, capturing usage data about the product's functions and performance, enabling designs to be changed after products have been manufactured and sold. Porter & Heppelmann (2014) refer to the technology stack as the technology infrastructure needed to support smart connected products, comprising five levels: hardware, software applications, operating system, network communication and product cloud. Further understanding of organizational architectures is needed in order to explore how organizations can collaborate in the different levels of the technology stack to deliver smart connected products. As organizations collaborate within the technology stack, the impact on organizational identity needs to be considered.

CONCLUSION

IT has been used to lead organizational transformation, however, as actors in the organization seek to make sense of how the technology affects daily practices, technology becomes used in ways which were unintended. Boudreau & Robey (2005) therefore argue that as technology can result in unintended consequences, technology has a limited role in transforming organizations.

Technology forms one component of the organizational architecture. Change to any component in the organizational architecture has the potential to trigger change in other components. A framework has been presented, based on analysis of IT-led transformation in five case studies, to assist organizations in planning how a change in technology may affect other components of the organization architecture.

Porter and Heppelman (2014) suggest that organizations need to focus on the question, what business are we in? The business model helps organizations to explore the external face presented

to the trading environment. The organization architecture helps organizations to explore the internal components that sustain the organization's identity. Continued developments in technology will consistently challenge the shape, identity and practice of organizations. Further work is therefore needed to explore the longitudinal impact of changes to the organizational architecture and identity of morphing organizations within an ever changing technological context.

REFERENCES

- Albert, S., & Whetten, D. A. (1985). Organizational identity. *Research in Organizational Behavior*, 7, 263–295.
- Andersén, J., Ljungkvist, T., & Svensson, L. (2015). Entrepreneurially oriented in what? A business model approach to entrepreneurship. *Journal of Small Business and Enterprise Development*, 22(3), 433–449. doi:10.1108/JSBED-11-2013-0170
- Ballon, P. (2007). Business modelling revisited: The configuration of control and value. *Info*, 9(5), 6–19.
- Balmer, J. M. T. (2008). Identity based views of the corporation. *European Journal of Marketing*, 42(9/10), 879–906. doi:10.1108/03090560810891055
- Boudreau, M.-C., & Robey, D. (2005). Enacting integrated information technology: A human agency perspective. *Organization Science*, 16(1), 3–18. doi:10.1287/orsc.1040.0103
- Carroll, J. M., & Rossen, M. B. (1992). Getting around the task-artifact cycle: How to make claims and design by scenario. *ACM Transactions on Information Systems*, 10(2), 181–212. doi:10.1145/146802.146834
- Cox, S. A. (2013). E-business planning in morphing organizations: Maturity models of business transformation. In E. Y. Li., S. Loh, C. Evans, & F. Lorenzi (Eds.), *Organizations and social networking: utilizing social media to engage consumers* (pp. 286–312). IGI Global.
- Cox, S. A. (2014). *Managing information in organizations: A practical guide to implementing an information management strategy*. Basingstoke, UK: Palgrave Macmillan. doi:10.1007/978-1-137-31668-4
- Cox, S. A., Krasniewicz, J. A., Perkins, J. S., & Cox, J. A. (2006). Modelling the organisational transformation associated with implementing e-business collaborative systems in the supply chain. *Proceedings of the British Academy of Management Conference (BAM2006) in Association with the University of Ulster and Queen's University*.
- DaSilva, C., & Trkman, P. (2014). Business model: What it is and what it is not. *Long Range Planning*, 47(6), 379–389. doi:10.1016/j.lrp.2013.08.004
- Dunphy, D., & Stace, D. (1993). The strategic management of corporate change. *Human Relations*, 46(8), 905–918. doi:10.1177/001872679304600801
- Ethiraj, S. K., & Levinthal, D. (2004). Bounded rationality and the search for organizational architecture: An evolutionary perspective on the design of organizations and their evolvability. *Administrative Science Quarterly*, 49, 404–437.
- Goksen, Y., Cevik, E., & Avunduk, H. (2015). A case analysis on the focus of the maturity models and information technologies. *Procedia Economics and Finance*, 19, 208–216. doi:10.1016/S2212-5671(15)00022-2
- Jackson, P., & Harris, L. (2003). E-business and organisational change: Reconciling traditional values with business transformation. *Journal of Organizational Change*, 16(5), 497–511. doi:10.1108/09534810310494900

Kaplan, R. S. (2005). How the balanced scorecard complements the McKinsey 7-s model. *Strategy and Leadership*, 33(3), 41–46. doi:10.1108/10878570510594442

Koskinen, K. U. (2015). Identity change in organizations: A philosophical exposition. *The International Journal of Organizational Analysis*, 23(4), 621–636. doi:10.1108/IJOA-03-2012-0564

Orlikowski, W. J. (2007). Sociomaterial practices: Exploring technology at work. *Organization Studies*, 28(9), 1435–1448. doi:10.1177/0170840607081138

Osterwalder, A., Pigneur, Y., & Tucci, C. L. (2005). Clarifying business models: Origins, present, and future of the concept. *Communications of the Association for Information Systems*, 16. Retrieved October 14, 2015, from <http://aisel.aisnet.org/cais/vol16/iss1/1>

Peters, T. J., & Waterman, R. H. Jr. (2004). *In search of excellence: Lessons from America's best run companies* (2nd ed.). New York: Harper Collins.

Porter, M. E., & Heppelmann, J. E. (2014). How smart, connected products are transforming competition. *Harvard Business Review*, 92(11), 64–88.

Saucer, C., & Willcocks, L. (2004). Strategic alignment revisited: Connecting organizational architecture and IT infrastructure. *Proceedings of the 37th Hawaii International Conference on System Sciences*, 1–10.

Senior, B., & Swailes, S. (2010). *Organisational change* (4th ed.). Harlow: Pearson.

Sherif, K., Tsado, L., Zheng, W., & Airhia, B. (2013). An exploratory study of organization architecture and the balance between exploration and exploitation of knowledge. *Vine: The Journal of Information and Knowledge Management Systems*, 43(4), 442–461. doi:10.1108/VINE-04-2012-0006

Waterman, R. H. Jr, Peters, T. J., & Phillips, J. R. (1980). Structure is not organization. *Business Horizons*, 23(3), 14–26. doi:10.1016/0007-6813(80)90027-0

Whetten, D. A. (2006). Albert and Whetten revisited: Strengthening the concept of organizational identity. *Journal of Management Inquiry*, 15(3), 219–234. doi:10.1177/1056492606291200

Yeo, R. K., & Ajam, M. Y. (2010). Technological development and challenges in strategizing organizational change. *The International Journal of Organizational Analysis*, 18(3), 295–320. doi:10.1108/19348831011062148

Yeo, R. K., & Marquardt, M. (2015). Think before you act: Organizing structures of action in technology-induced change. *Journal of Organizational Change Management*, 28(4), 511–528. doi:10.1108/JOCM-12-2013-0247

ADDITIONAL READING

Cox, S. A. (2014). *Managing information in organizations: A practical guide to implementing an information management strategy*. Basingstoke, UK: Palgrave Macmillan. doi:10.1007/978-1-137-31668-4

DaSilva, C. M., & Trkman, P. (2014). Business model: What it is and what it is not. *Long Range Planning*, 47(6), 379–389. doi:10.1016/j.lrp.2013.08.004

Iansiti, M., & Lakhani, K. R. (2014). Digital ubiquity: How connections, sensors, and data are revolutionizing business. *Harvard Business Review*, 92(11), 91–99.

Leonardi, P. M., Nardi, B. A., & Kallinikos, J. (Eds.). (2012). *Materiality and organizing: Social interaction in a technological world*. Oxford, UK: Oxford University Press. doi:10.1093/acprof:oso/9780199664054.001.0001

Osterwalder, A. (2010). *Business model generation: A handbook for visionaries, game changers, and challengers*. New Jersey: John Wiley & Sons.

Rothwell, W. J., Stavros, J. M., & Sullicva, R. L. (Eds.). (2016). *Practicing organization development: Leading transformation and change*. New Jersey: John Wiley & Sons.

Schultz, M., Maguire, S., Langley, A., & Tsoukas, H. (Eds.). (2012). *Constructing identity in and around organizations*. Oxford, UK: Oxford University Press. doi:10.1093/acprof:oso/9780199640997.001.0001

Stanford, N. (2013). *Organization design: Engaging with change*. Oxon, UK: Routledge.

KEY TERMS AND DEFINITIONS

Business Model: An external representation of how the organization makes money and creates value through interactions with customers, suppliers and partners in the external trading environment.

Constitutive Entanglement: Complex configurations of inter-related concepts.

Context: The setting within which organizations, people, processes, information and events are interpreted. Context is an emergent property derived from interrelated dynamic states at a point in time.

Morphing Organization: An organization that changes the outer face the organization displays to the world by changing internal structures and processes within the organization to facilitate and support the external face, whilst seeking to maintain organizational identity.

Organization Architecture: An internal representation of the formal and informal components of the organization that provide the behavioural space for the organization's activities.

Organization Identity: A set of enduring attributes, which emerge through interaction with the organization, representing the spirit of the organization.

Organization Structure: The way in which resources are grouped in the organization to determine the control of activities, flow of information, and authority and responsibility for decision-making.

Organization Transformation: Significant changes to more than one component of the organization architecture that directly impact more than one department or function in the organization.

Sociomateriality: The view that technology and human interaction with technology are interlinked such that the two concepts cannot be objectively studied separately due to constitutive entanglement.

Workaround: A way of working to overcome perceived omissions, limitations or problems with formally defined working practices that are often embedded in IT systems.

How the Crowdsourcing Enhance the Co-Creation Into the Virtual Communities

B

Bahri Ammari Nedra
IHEC of Carthage, USA

INTRODUCTION

Nowadays, more and more companies are increasingly following the trend, by actively integrating their customers in their new processes of developing products and services (Romero et al. 2014). Co-creation is the process by which products, services and experiences are jointly developed by companies, their partners and the final consumer; leading to a new space where the created value would be shared (Prahalad and Ramaswamy, 2003). This form of cooperation has developed thanks to IT-based advances (IT) which promoted the development of social networks, blogs, forums and creative platforms (Peer et al. 2015). Many means that have allowed customers to interconnect with the world and to share, exchange, create and co-create knowledge (Lee et al, 2003). Through these new interconnected tools of socialization, customers can now actively contribute to the development of new products, ideas and concepts (Hoyer et al. 2010; Ogawa and Piller, 2006). Companies have soon realized its potential and seized the opportunity. Upstreaming or downstreaming the value chain, allowing for co-creation by costumers, resulted in customers as lead users (LU), simple users or emerging nature consumers (ENC). As such, customers represent a key to unlock new sources promoting competitive advantages for brands (Prahalad and Ramaswamy, 2004). By having access to brands' platforms, costumers could share their opinions about the product and experience. They could also discuss freely and participate in the evolution of the company's offer and develop its image and reputation. This is possible through crowdsourcing. This is

an online participatory activity that involves two parties: the user and the company / brand. The user will be satisfied by a given type of need, either, social recognition, self-esteem or development of individual skills, meanwhile the crowdsourcer will have the advantage of gaining what the user will bring to the company. Thus, some brands have found the way to keep their customers loyal while reducing the risks associated with creating a new product or service (Romero and Molina, 2011). In this line of thinking, the eYeka platform allows consumers and major brands to collaborate on creative projects and give in return an amount of money to the winning project.

However, the consequences / strategic implications of value co-creation were studied rigorously and prolifically, in terms of the cognitive, emotional and intentional patterns of co-creation. The aim of this paper is to determine the contribution of crowdsourcing in facilitating co-creation in virtual communities. And, *in fine*, to present the consequences of IT on firm, in terms of: the strength of the relationship, productivity, and efficiency.

BACKGROUND

Online Co-Creation: Definition of Co-Creation

According to O'Hern and Rindfleisch (2009), co-creation can be defined as "an activity of collaborative development of new products (NPD)¹ in which consumers actively contribute and select various elements of a new product offering" (p.86).

DOI: 10.4018/978-1-5225-2255-3.ch061

The recent literature pointed to the importance of integrating customers in the process of value creation as an effective way to develop better products while reducing the costs and risks of product / service failure (Prahalad and Ramaswamy, 2004). The concept of co-creation presents a new value form, something that is co-constructed, not just consumed by the customer. Co-creation, also known as participatory marketing or “consumer made”, is a new activity that refers to the participation of the customer as an active contributor from the start of the innovation process. It is a “situation in which consumers are working with companies to create value” (Humphreys and Grayson, 2008, p.963). In the value co-creation process, the customer can propose the company with innovative ideas for future products. In this regard, Sanders and Stappers, (2008) define co-creation as “any act of collective creativity, i.e. creativity that is shared by two or more people”. Similarly, co-creation involves customers in developing new products where they act as a source of innovation in order to increase the new product or service value. Indeed, it involves transforming the roles of the traditional players into Co-creators and co-responsible actors over time. Co-creation techniques refer to crowdsourcing or even collaborative or participatory marketing. The concept of co-creation is far from being simply customer-oriented. According to Prahalad and Ramaswamy, (2004), this tendency of the joint creation of products or services is a concept that shifts from a business-centric view to a product and experience-centric view of value co-creation.

Co-creating unique experiences with individual customers particularly helped to promote new sources of competitive advantage for companies. In this regard, several marketing researchers have investigated different facets of online Co-creation and its impact on business performance and the competitive advantage it may acquire. According to Zhuang (2010), co-creation is seen as a two-dimensional concept in which the customer may be considered either as a source or as a co-developer. Contrary to the approach that positions the cus-

tomers “outside the company”, the new approach places the customer within the company given the nature of the execution of value creation and innovation (Sahwney et al. 2005).

Co-Creation and Theoretical Trends in Marketing

Researchers show that the principle of co-creation has contributed to the emergence of several theoretical trends in marketing. We briefly mention the three main currents: tribal marketing, collaborative innovation and marketing Knowledge.

- *Tribal marketing or community marketing* (Cova, 2008). Tribal marketing builds on one central idea, which is the company’s losing some control over the brand for the benefit of the community of consumers seeking to reappropriate it (Brincker, 2003).
- *Collaborative innovation* includes three major concepts, namely “lead users”, “user innovation”, “user design” that were well developed by the work of Von Hippel (1978).
- *Knowledge marketing* considers consumer expertise, skills and intelligence allowing the company to define its offer and its production (Abidi-Barthe and Kaabachi, 2010).

VALUE CO CREATING WITHIN VIRTUAL BRAND COMMUNITY (VBC)

“Customer Empowerment”: Concept Presentation

Consumers dispose of limited tools to share their experiences and give their opinions. Today, with the emergence of Web 2.0, consumers have various means to share their passions, to know the opinion of others, and to produce and exchange content (Kozinet, 2002). They can quickly gather

information on the brand and its cost. In fact, they have more control over the use of information and are able to influence other consumers' purchasing decisions. This change has changed the relationship between consumers and businesses.

The advent of social web as a new creative space on the internet has redefined the role of the customer who becomes more involved in trade policy and corporate strategy (Füller et al. 2010). The customer wants to become a real "consum'actor" by being more involved in the company's value chain. He/she will interact with brands, shares the output power and contributes thus to the definition of the offer. He/she is not passive, he/she has become a key player in the marketing process (Firat and Venkatesh, 1993). The consumer is described as an actor more influential, more connected, more informed, more expert, more demanding and more active (Cova, 2008).

This shift in position contributed to the emergence of what is defined today as a "seizure of power" or "empowerment" of the consumer. This is a phenomenon involving the delegation of some power and empowering the consumer. The latter is considered a player who has more skills and is able to outsmart strategies and contribute to defining the guidelines and the company's shares (Cova, 2008).

Today, the customer is asked to design his/her own products, to conduct consumer arbitrations, model experience, either individually or with other people, and prepare for his/her favorite brand advertising (Xie et al. 2008). The consumer has taken the power to improve his/her knowledge of the business, which can procure satisfaction (Zwick et al. 2008). Managers will be invited to "take into account the other, the consumer, not learning about it, but learning from his/her, expertise, experiences..." (Cova, 2008, p.73). In this case, it is about doing marketing with consumers' "market with", instead of marketing to consumer "market to". A new collaborative perspective is established where the consumer is the business partner rather than the client. Value co-creation

(Prahalad and Ramaswamy, 2004) between the company and the (s) consumer (s) is thus the key process of this new marketing approach. Companies must, in this case, promote customized interactions and expertise and knowledge exchange with its customers. They must also rethink the way they conceptualize the brand and its image, think of a more collaborative approach in which value creation for the brand and its development is executed with consum'actors (Vargo et al. 2008).

The concept of empowerment refers to "co-constructed" marketing, considered the key to succeeding in maintaining long-term relationships and to effectively managing customer relationships (Filser and Vernet, 2011).

Forms of Value Co-Creation in Virtual Brand Communities (VBC)

Consumers' participation in value co-creation can take place upstream or downstream the production process. Upstream co-creation refers to customers' involvement before launching the product, while downstream co-creating focuses on looking to interact with customers after the purchase and use of the brand (Vargo et al. 2008). These two forms of participation were presented by Vernet and Hamdi-Kidar, (2010). They fully report to "customer empowerment":

Upstream Participation

- **Preconception:** The company involves creative consumers (lead users) in the design and development of new products to increase the efficiency of the new product development process and to reduce innovations cost. Companies are adopting preconception techniques to develop their imaginations and inspire new ideas for their brands. This is possible through exploiting the potential of the best customers endowed with expertise, intelligence and creativity. At this level, consumers are asked to respond to a given problem, and

to find a solution. They are requested to actively participate in the creation of new products, improved prototypes to test new products, and in the generation and evaluation of new ideas (Yan et al. 2014).

- **Mass Customization:** Consumers participate in an experience to change themselves certain elements of a product, specifying the attributes that best meet their preferences among a set of predefined modules by the company. This is done so that the final product is in line with customers' tastes and requirements. Consumers participating in product customization tend more to seek and develop positive attitudes towards the brand. Participation in product customization increases the level of consumer satisfaction, fosters brand loyalty and creates barriers against rebranding.
- **Co-Promotion:** Communities are increasingly wishing to be involved in the brand's life cycle and by means of involving themselves in designing an advertisement in accordance with the company's strategy. This is about involving the customer in developing communication policy. This contribution aims to promote and enhance brand image through the creation of advertising campaigns.

Downstream Participation

There are two forms of downstream participation are presented in Figure 1: co-production and co-determination (Cova, 2008).

- **Co-Production:** (User generated content) With the emergence of Web 2.0, "user generated content" has become one of the new trends in current marketing. Consumers share their stories and speak via virtual communities, post information, comments, and photos to share their experiences and give their opinions. The content is produced directly by users and helps to pro-

mote interaction between the company and users. This content creation by customers is a key element for developing brand awareness and influencing attitudes and behavior of other consumers.

- **Co-Determination:** Co-determination is granted when the company promotes cross-identification of needs between the company and the creative consumers. It is an open dialogue that should determine consumers' needs and expectations in an exploratory manner. The company can then use this information to readjust its offer and enjoys the consumption experience and generates the maximum of value. Companies are committed to this type of "participatory" marketing because they may gain several advantages.

Practical Online Value Co-Creation

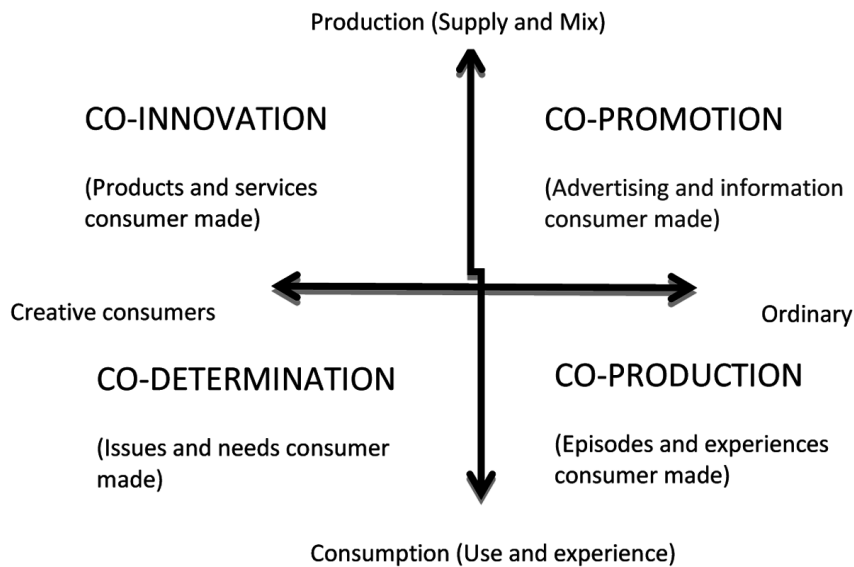
Different forms online value co-creating may be distinguished, namely: crowd sourcing, open and user innovation.

Crowdsourcing

Crowdsourcing emerged in 2006 and was used to describe users' activities to co-create content. Crowdsourcing is the use of information using a host of technologies (eg, business, individual, organization) to outsource certain business tasks (Prpic' et al. 2015). Crowdsourcing demands concrete solutions to a problem or a specific task. Contributions can firstly be objective and well-founded and partly subjective and be required to perform tasks to determine users' opinions, tastes, or beliefs.

Crowdsourcing is an online participatory activity which an individual, an institution, a non-profit organization, or a company offers to a group of individuals with different knowledge, voluntary commitment to a task. The user will be satisfied with a given type of need, either, social recognition,

Figure 1. Managerial forms of consumer made



self-esteem or development of individual skills, meanwhile the crowdsourcer obtains and uses to their advantage what the user has brought to the company. This form will depend on business activity types (Estellés-Arolas and González-Ladron-de-Guevara, 2012).

Open Innovation

Changes brought by globalization have significantly influenced companies, which have become constantly seeking knowledge resources and expertise. Open innovation is to create and innovate with the support of external stakeholders: customers, suppliers, partners and the wider community. Open innovation therefore creates a space where individuals and organizations can be actively involved in the creation of mutually beneficial solutions. Open innovation is a social and inclusive way to solve complex issues and improve processes. It suggests contacting larger groups and it is based on community engagement to collaborate around specific challenges and issues. Thus, the flow of ideas from outside of the organization generates intelligent and innovative solutions within the organization.

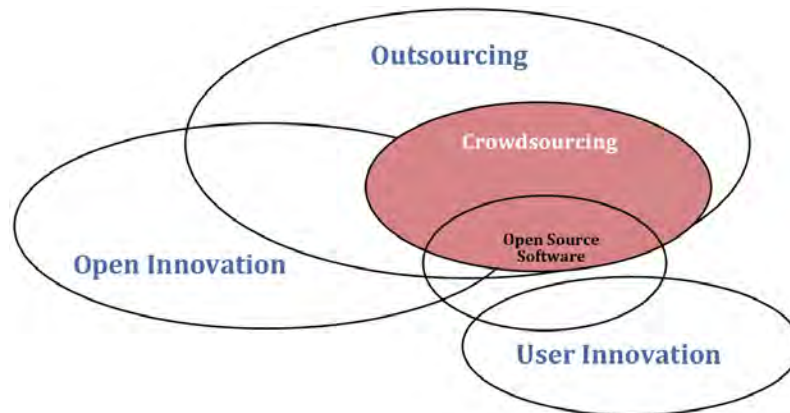
User Innovation

User innovation refers to innovation by intermediate users (such as users, companies) or consumer users (end individual users or groups of users), rather than by suppliers (producers or manufacturers). Many products and services are developed or at least refined by users on the implementation and use phases. When most consumers have problems with products, users bring their own modifications to existing products or entirely new products. Often, users share their ideas with manufacturers hoping to innovate and participate in the production of the product. According to Tuomi (2002), the main uses are often not expected uses invented by the user communities, which reinterpret and reinvent the sense of emerging technological opportunities (open source software).

Figure 2. shows the position of crowdsourcing vis-a-vis open innovation and user innovation (Tuomi, 2002; Rayna et al. 2015; LaToza and Van-der-Hoek, 2016):

Figure 2. Positioning of Crowdsourcing

Source: Tuomi, 2002



CONSEQUENCES OF CO CREATION

The Effects of Participating in Online Co-Creation

In this section, we will define and detail the implications of co-creation on consumer loyalty behavior, namely: electronic word-of-mouth (e-WOM), re-purchase and price tolerance.

E-WOM

E-WOM allows influencing the success or failure of marketing new products, sales volume, brand awareness and brand equity (Bambaur-Sasche, 2011). Word-of-mouth influences consumers' purchase decision. Indeed, comments circulated about the product significantly affect attitudes (Lee et al. 2003), purchase intent (Lin et al. 2013) and reduce purchase perceived risk. Thanks to E-WOM, cyberconsumers are the best sales force to build the brand and develop brand awareness and loyalty. Companies should make considerable efforts to encourage the implementation of positive word of mouth and accelerate its diffusion (Dellarocas, 2003).

Re-Purchase

Under a behavioral perspective, loyalty characterizes consumers who buy a special offer - whether

a product or service - and continue to buy, rather than switch products/services (Lin et al. 2013). "It refers to the behavior of consumers purchasing a brand for a period of time (Lin et al. 2013, p. 279)". Behavioral loyalty can be understood as a synonym to repeating a purchase behavior or a repeated purchases behavior.

However, to achieve a repeated purchase behavior, the consumer undergoes various interactions that determine their decisions. This brings us to discuss the relationship between emotions, brand engagement and loyalty behavior. First, we believe that customer engagement is considered a determinant of customer loyalty (Gundlach et al. 1995; Morgan and Hunt, 1994; Fullerton, 2003).

Price Tolerance

Price tolerance generally is considered as the situation consumers to be willing to accept an increase in prices without recourse to switching the product or brand. Thus, this consistency in customer behavior is called price tolerance. However, Anderson (1996) explains that price tolerance sets a limit above which the customer will switch brands or products. In addition, several authors have established a link between affective commitment and price tolerance, pointing to a significant relationship between affective commitment and price tolerance in e-commerce (Fullerton, 2003), and proving the positive effect of affective com-

Table 1. Transformation of the relationship between businesses and consumers

From	To
A sens. Firm to customer. Controlled by the firm Choice = Buy / No buy Customer must be integrated in the company's offer.	Double sens. Customer to firm Customer to customer Customer wants/can impose its choice view The consumer wants / is pushed to co create a personalized experience while benefiting from the company's experience.

Source: Adapted from Prahalad and Ramaswamy, 2003

mitment on consumers' willingness to pay more for a service.

Strength of the Relationship

Strength of the relationship is often defined as *"the extent and degree of the relationship"* (Bove and Johnson, 2006). It amounts to the creation of a better value for the customer, while involving customers in the process. It is considered the basis for the development of long term relationships with customers. It should be mentioned also that co-creation through customized interactions and specific customized problem-solving solutions can strengthen more buyer-seller relationships (Claycomb and Martin, 2001). Evanschitzky et al. (2006) argue that strong buyer-seller relationships contribute to customer loyalty. Therefore, interaction and dialogue resulting from co-creation is conducive to strong buyer-seller relationships. Indeed, business-customer relationships can be strengthened by co-creation, which in turn improves customer equity (Kumar et al. 2010; Van-Doorn et al. 2010).

Typology of Co-Creation Activities

A conceptual typology of customer co-creation should positively contribute to building and strengthening relationships between companies and customers during the innovation process. This typology depends on the nature of provided information.

The first dimension rests on the nature of provided information. In every innovation process, companies face uncertain sources with regard to their technical and managerial capacities. To re-

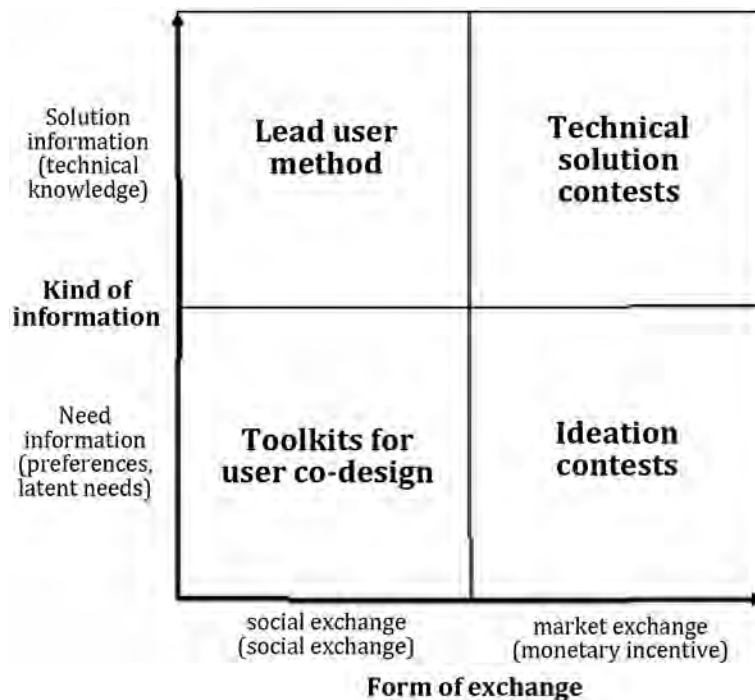
duce these uncertainties, companies need to have access to different types of information which is divided into two groups and presented in Figure 3 (Ogawa and Piller, 2006; Von-Hippel, 1986):

- Information on customer needs and market (needs information): i.e. information about preferences, needs, desires, satisfaction, motivations... This information allows the company to better understand customer requirements and reduce failure risk.
- Information on technical solutions, and in particular on how to use technology to transform customer needs in view of offering a new product / service. Access to this type of information allows product developers to put in place resolution activities more oriented towards the innovation process.

Information distinguishes co-creation methods into "money markets" based on economic relationships with monetary incentives, and "social market" methods based on social exchange relationships with non-monetary rewards (Füller, 2010):

- Economic exchange is based on a monetary exchange of ideas against solutions. Participants compete to get the maximum from a limited allocation of resources.
- Social exchange consists of methods with which participants engage in innovative behavior for reasons such as pleasure of completing the task, or expectations of results that enhance their own usage experience (Harhoff et al. 2003).

Figure 3. Typology of co-creation activities



Three types of opportunities of value co-creation are distinguished (Payne et al. 2008; Storbacka et al. 1995): opportunities offered by technology, by industry changes and customer preferences changes.

- Opportunities offered by technology (e.g. iPod) provide companies with new methods to: engage with customers, co-create innovative goods / services and enjoy the experience offered.
- Opportunities offered by changes in customer preferences and lifestyles. In this regard, companies should constantly detect new preferences, new lifestyles...

Challenges to Co-Creation

The company can create two important sources of competitive advantage through the implementation and good management of co-creation: productivity gains and improved efficiency (Payne et al. 2008; Storbacka et al. 1995).

Productivity Gains

Co-creation allows the brand to increase its productivity gains and efficiency. Consumer input in the development of products and services reduced certain internal costs to a minimum (Gatautis and Vitkauskaitė, 2014). Indeed, this contribution will partially replace the work done by employees. Cost reduction also results from a gain in acquiring certain information and studying consumers. Co-creation also allows the brand to be more efficient in its market by reducing failure risk of its products through the continuous improvement of products and the response speed to the market (Xie et al. 2008).

Improved Efficiency

Co-creation can bring significant gains in efficiency in co-creating products, through a better alignment of consumer needs and their high commercial potential (Lilien et al. 2002). These results can directly influence product / service

sales, employee satisfaction, organizational performance, increase operational efficiency and ultimately, revenue / profitability. Co-created products differ from other products (Song and Adams 1993), have better commercial attractiveness (Franke et al. 2006) and generate high expected benefits. In addition, adjusting products / services preferences to co-creation can increase positive attitudes toward the product, subsequent purchase intentions, and willingness to pay more and positive word-of-mouth about the brand (Keinz and Steger, 2009). Thus, co-creation can strengthen business / customers relationships, improve customer capital through increased delivery of value and an increasing number of connecting points between the company and consumers (Van-Doorn et al. 2010; Tong et al. 2014). Companies in the same market are able to aggressively compete to attract and retain the most important co-creators.

FUTURE RESEARCH DIRECTIONS

Web 2.0 has facilitated the work of brand managers and enabled consumers to be more active and more involved on the net. However, some issues are to be mentioned that call for special attention. First, we need to identify the factors that motivate users to participate in co-creation in order to control them. Second, we should think of criteria to assess participants' skills in co-creation? This issue is very relevant as consumers who participate should be very effective in co-creation. Moreover, it would be best to determine which application of crowdsourcing applies to which target / population? Third, we should determine the nature of the relationship between brands and consumers. How it should be and what are the determinants that strengthen this relationship and how to control them?

CONCLUSION

Today, more than 40 million people worldwide participate in a virtual community (Sicilia and Palazon, 2008). These people are becoming more demanding vis-à-vis the products and services offered by companies. Given the competitive environment, companies put customers at the heart of their strategies to best meet their expectations.

The concept of co-creation was created to engage the customer, whose aim, for the company, is to create new opportunities and gain competitive advantages in the market. First, it helps to generate positive word of mouth about the brand and participatory operations perceived as original. This strategy can reach a wide audience to increase awareness, recruit new customers, and increase sales. Developing interactivity with the consumer is a part of the brand's objectives. Indeed, brands are creating connecting opportunities with customers to better understand consumers, better meet their expectations and improve the relevance of innovations. The more individuals contribute to the generation of content, the better the company identifies their needs and desires.

Involving consumers in value co-creation makes the brand more salient and memorable and strengthens brand / consumer proximity. Integrating the consumer in a co-creation process promotes customer self-identification with the community and insures their loyalty in the long-run (Algesheimer et al. 2005). Participatory marketing allows the company to benefit from the consumers' creativity and reduce research and development costs.

REFERENCES

- Abidi-Barthe, A., & Kaabachi, S. (2010). La co-création d'expériences de consommation par le Web2.0. *Journal of Internet Banking and Commerce*, 15(2). Retrieved from <http://www.arraydev.com/commerce/JIBC/2010-08/Abidi-Barthe.pdf>
- Anderson, E. W. (1996). Customer satisfaction and price tolerance. *Marketing Letters*, 7(3), 265–274. doi:10.1007/BF00435742
- Bambaur-Sasche, S. (2010). Brand equity dilution through negative online word-of-mouth communication. *Journal of Retailing and Consumer Services*, 18(1), 38–45. doi:10.1016/j.jretconser.2010.09.003
- Bove, L. L., & Johnson, L. W. (2006). Customer loyalty to one service worker: Should it be discouraged? *International Journal of Research in Marketing*, 23(1), 79–91. doi:10.1016/j.ijresmar.2006.01.007
- Claycomb, C., & Martin, C. L. (2001). Building customer relationships: An inventory of service providers objectives and practices. *Marketing Intelligence & Planning*, 19(6), 385–399. doi:10.1108/EUM00000000006109
- Dellaroccas. (2003). The Digitization of Word of Mouth: Promise and Challenges of Online Feedback Mechanisms. *Management Science*, 49(10), 1407–1424.
- Estellés, A., & González-Ladron-de-Guevara, F. (2012). Towards an integrated crowdsourcing definition. *Journal of Information Science*, 38(2), 189–200. doi:10.1177/0165551512437638
- Evanschitzky, H., Iyer, G. R., Plassman, H., Niessing, J., & Meffert, H. (2006). The relative strength of affective commitment in securing loyalty in service relationships. *Journal of Business Research*, 59(12), 1207–1213. doi:10.1016/j.jbusres.2006.08.005
- Firat, A. F., & Venkatesh, A. (1993). Marketing in a postmodern world. *European Journal of Marketing*, 29(1), 40–57. doi:10.1108/03090569510075334
- Franke, N., Von Hippel, E., & Schreier, M. (2006). Finding Commercially Attractive User Innovations: A Test of Lead-User Theory. *Journal of Product Innovation Management*, 23(4), 301–315. doi:10.1111/j.1540-5885.2006.00203.x
- Füller, J. (2010). Refining Virtual Co-Creation from a Consumer Perspective. *California Management Review*, 52(2), 98–122. doi:10.1525/cmr.2010.52.2.98
- Füller, J., Faullant, R., & Matzler, K. (2010). Triggers for virtual customer integration in the development of medical equipment from a manufacturer and a users perspective. *Industrial Marketing Management*, 39(8), 1376–1383. doi:10.1016/j.indmarman.2010.04.003
- Fullerton, G. (2003). When Does Commitment Lead to Loyalty? *Journal of Service Research*, 5(4), 333–344. doi:10.1177/1094670503005004005
- Gatautis, R., & Vitkauskaitė, E. (2014). Crowdsourcing application in marketing activities. *Procedia: Social and Behavioral Sciences*, 110, 1243–1250. doi:10.1016/j.sbspro.2013.12.971
- Gundlach, G. T., Ravi, S. A., & Mentzer, J. T. (1995). The Structure of Commitment in Exchange. *Journal of Marketing*, 59(1), 78–92. doi:10.2307/1252016
- Harhoff, D., Henkel, J., & Von-Hippel, E. (2003). Profiting from voluntary information spillovers: How users benefit by freely revealing their innovations. *Research Policy*, 32(10), 1753–1769. doi:10.1016/S0048-7333(03)00061-1
- Hoyer, W. D., Chandy, R., Dorotic, M., Krafft, M., & Singh, S. S. (2010). Consumer co-creation in new product development. *Journal of Service Research*, 13(3), 283–296. doi:10.1177/1094670510375604

- Keinz, F. P., & Steger, C. J. (2009). Testing the Value of Customization: When Do Customers Really Prefer Products Tailored to Their Preferences? *Journal of Marketing*, 73(5), 103–121. doi:10.1509/jmkg.73.5.103
- Kietzmann, J. H., Hermkens, K., McCarthy, I. P., & Silvestre, B. S. (2011). Social media? Get serious! Understanding the functional building blocks of social media. *Business Horizons*, 54(3), 241–251. doi:10.1016/j.bushor.2011.01.005
- Kozinets, R. V. (2002). The Field behind the Screen: Using netnography for marketing research in online communities. *JMR, Journal of Marketing Research*, 39(1), 61–72. doi:10.1509/jmkr.39.1.61.18935
- Kumar, V., Aksoy, L., Donkers, B., Wiesel, T., Venkatesan, R., & Tillmanns, S. (2010). Undervalued or Overvalued Customers: Capturing Total Customer Engagement Value. *Journal of Service Research*, 13(August), 297–310. doi:10.1177/1094670510375602
- LaToza, T. D., & Van-der-Hoek, A. (2016). Crowdsourcing in Software Engineering: Models, Motivations, and Challenges. *Software, IEEE*, 33(1), 74–80. doi:10.1109/MS.2016.12
- Lee, F. S. L., Vogel, D., & Limayem, M. (2003). Virtual community informatics: A review and research agenda. *Journal of Information Technology Theory and Application*, 5(1), 47–61.
- Lilien, G. L., & Pamela, D. (2002). Performance Assessment of the Lead User Idea-Generation Process for New Product. *Management Science*, 48(8), 1042–1059. doi:10.1287/mnsc.48.8.1042.171
- Lin, C., Wu, Y. S., & Chen, J. C. V. (2013). Electronic Word-of-Mouth: The Moderating Roles of Product Involvement and Brand Image. *Diversity, Technology, and Innovation for Operational Competitiveness: Proceedings of the 2013 International Conference on Technology Innovation and Industrial Management*.
- Morgan, R. M., & Hunt, S. D. (1994). The commitment-trust theory of relationship marketing. *Journal of Marketing*, 14(4), 374–395.
- O'Hern, M. S., & Rindfleisch, A. (2009). Customer Co-Creation: A Typology and Research Agenda. *Review of Marketing Research*, 6, 84–106.
- Ogawa, S., & Piller, F. T. (2006). Reducing the Risks of New Product Development. *Sloan Management Review*, 47, 65–72.
- Payne, A. F., Storbacka, K., & Frow, P. (2008). Managing the Co-creation of Value. *Journal of the Academy of Marketing Science*, 36(1), 83–96. doi:10.1007/s11747-007-0070-0
- Peer, E., Samat, S., Brandimarte, L., & Aquisti, A. (2015). *Beyond the Turk: Alternative platforms for crowdsourcing behavioral research*. Available: <http://papers.ssrn.com/abstract=2594183>
- Prahalad, C. K., & Ramaswamy, V. (2003). The new frontier of experience innovation. *MIT Sloan Management Review*, 44(4), 12.
- Prahalad, C. K., & Ramaswamy, V. (2004). *The future of competition: co-creating unique value with customers*. Boston, MA: Harvard Business School Press.
- Prpic', J., Shukla, P. P., Kietzmann, J. H., & McCarthy, I. P. (2015). How to work a crowd: Developing crowd capital through crowdsourcing. *Business Horizons*, 58(1), 77–85. doi:10.1016/j.bushor.2014.09.005
- Rayna, T., Striukova, L., & Darlington, J. (2015). Co-creation and user innovation: The role of online 3D printing platforms. *Journal of Engineering and Technology Management*, 137, 90–102. doi:10.1016/j.jengtecman.2015.07.002
- Romero, D., & Molina, A. (2011). Collaborative networked organizations and customer communities: Value co-creation and co-innovation in the networking era. *Production Planning and Control: The Management of Operations*, 22(5-6), 447–472. doi:10.1080/09537287.2010.536619

- Sahwney, M., Verona, G., & Prandelli, E. (2005). Collaborating to create: The internet as a platform for customer engagement in product innovation. *Journal of Interactive Marketing*, 19(4), 4–34. doi:10.1002/dir.20046
- Sanders, E. B. N., & Stappers, P. J. (2008). Co-creation and the new landscapes of design. *CoDesign*, 4(1), 5–18. doi:10.1080/15710880701875068
- Sicilia, M., & Palazon, M. (2008). Brand communities on the internet: A case study of Coca-Cola's Spanish virtual community, Corporate Communications. *International Journal (Toronto, Ont.)*, 13(3), 255–270.
- Song, J. H., & Adams, C. R. (1993). Differentiation Through Customer Involvement in Production or Delivery. *Journal of Consumer Marketing*, 10(2), 4–12. doi:10.1108/07363769310039085
- Storbacka, K., Strandvik, T., & Groonroos, C. (1995). Managing customer relationships for profits: The dynamics of relationship quality. *International Journal of Service Industry Management*, 5(5), 21–38. doi:10.1108/09564239410074358
- Tong, Y., Cao, C. C., & Chen, L. (2014, August). Tcs: efficient topic discovery over crowd-oriented service data. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 861–870). ACM. doi:10.1145/2623330.2623647
- Tuomi, I. (2002). *Networks of Innovation*. Oxford University Press.
- Van-Doorn, J., Lemon, K. E., Mittal, V., Naß, S., Pick, D., Pirner, P., & Verhoef, P. C. (2010). Customer Engagement Behavior: Theoretical Foundations and Research Directions. *Journal of Service Research*, 13(3), 253–266. doi:10.1177/1094670510375599
- Vargo, S. L., Maglio, P. P., & Akaka, M. A. (2008). On Value and Value Co-Creation: A Service Systems and Service Logic Perspective. *European Management Journal*, 26(3), 145–152. doi:10.1016/j.emj.2008.04.003
- Vernette, E., & Hamdi-Kidar, L. (2010). Lead-user ou consommateur de nature émergente ? Une recherche de paternité pour les stratégies de co-création marketing. *Centre de Recherche en Management*, 21, 1–21.
- Von Hippel, E. (1978). Successful industrial products from customer ideas. *Journal of Marketing*, 42(1), 39–49. doi:10.2307/1250327
- Von Hippel, E. (1986). Lead Users: A Source of Novel Product Concepts. *Management Science*, 32(7), 791–805. doi:10.1287/mnsc.32.7.791
- Xie, C., Bagozzi, R. P., & Sigurd, V. T. (2008). Trying to Prosume: Toward a Theory of Consumers as Cocreators of Value. *Journal of the Academy of Marketing Science*, 36(1), 109–122. doi:10.1007/s11747-007-0060-2
- Yan, H., Param, V. S., & Kannan, S. (2014). Crowdsourcing New Product Ideas Under Consumer Learning. *Management Science*, 60(9), 2138–2159. doi:10.1287/mnsc.2013.1879
- Zwick, D., Bonsu, S. K., & Darmody, A. (2008). Putting consumers to work: Co-creation and new marketing govern-mentality. *Journal of Consumer Culture*, 8(2), 163–196. doi:10.1177/1469540508090089

ADDITIONAL READING

- Digouta, J. (2013). Azourib., M., Decaudinc, J-M., & Rochard, S. (2013). Crowdsourcing, Outsourcing to Obtain a Creativity Group. *ARAB ECONOMIC AND BUSINESS JOURNAL*, 8(1-2), 6–15. doi:10.1016/j.aebj.2013.11.001
- Lorenzo-Romero, C., Constantinides, E., & Brünink, L. A. (2014, August). Co-creation: Customer Integration in Social Media Based Product and Service Development. In *Procedia - Social and Behavioral Sciences, 2nd International Conference on Strategic Innovative Marketing*, 148(25), 383–396.

Meyer, A. N., Longhurst, C. A., & Singh, H. (2016). Crowdsourcing Diagnosis for Patients With Undiagnosed Illnesses: An Evaluation of CrowdMed. *Journal of Medical Internet Research*, 18(1), e12. doi:10.2196/jmir.4887 PMID:26769236

Nadangea, R.-S. (2014). Customer perception about Crowdsourcing within the suburbs of Mumbai. *Procedia Economics and Finance*, 11, 268–275. doi:10.1016/S2212-5671(14)00195-6

Poetz, M. K., & Schreier, M. (2012). The Value of Crowdsourcing: Can Users Really Compete with Professionals in Generating New Product Ideas? *Journal of Product Innovation Management*, 29(March), 245–256. doi:10.1111/j.1540-5885.2011.00893.x

KEY TERMS AND DEFINITIONS

Co-Creation: Co-creation is the process by which products, services and experiences are jointly developed by companies, their partners and the final consumer, leading to a new space where the value created would be shared.

Crowdsourcing: This is a participatory activity online that involves two parties: the user and the company / brand.

ENC: Consumers capable of developing any product that the ordinary consumer will find more attractive and useful as those developed by the LU. Lead users are users who, through their

experience, their expertise and their know-how in one area have a real attention from firms.

E-WOM: Any positive or negative statement made by a potential customer and is available to a multitude of consumers through the Internet. E-WOM promotes faster diffusion of information among members of a community. Moreover, it is the basis of a viral effect likely to cover a wider audience and contribute to developing or damaging brand image.

LU: Lead users are users who, through their experience, their expertise and their know-how in one area have a real attention from brands.

Virtual Brand Community (VBC): The place where consumers identify with the group they belong to, and who are actively involved in value co-creation and in the interactive exchange of opinions on online platforms in order to contribute to the development of the brand image (Peer et al. 2015).

Web 2.0: Allows, through a set of simple tools, better interactivity between users / user. It is for them to exchange information and interact in a simple way to create and share content on the web. The user becomes more active on the web and contributes to the enrichment of social networks by providing useful information for brands/firms.

ENDNOTE

¹ NPD: New Product Development

Hyper-Sensitivity in Global Virtual Teams

Andre Araujo

Texas A&M University, USA

INTRODUCTION

Increasingly multinational enterprises are implementing global virtual teams. In such online collaborative settings, developing shared understanding and managing interpersonal conflict may be constrained by the fact that members have little of the traditional mechanisms that are used to engage in personal interactions with others, observe fellow team members at work, or develop a shared history of professional accomplishments. In addition, members of virtual teams may be thousands miles away from each other and possess different cultural background. Due to the absence of proximal interactions, virtual team members are likely to evaluate others based on amplified perceptions of their computer-mediated social interaction (Walther, 1996). This chapter examines hyper-sensitivity in virtual teams and its consequences to team's relational interactions and outcomes. Specifically, the following research question is addressed: *What are the mechanisms that influence levels of engagement and interpersonal conflict in global virtual teams?*

BACKGROUND

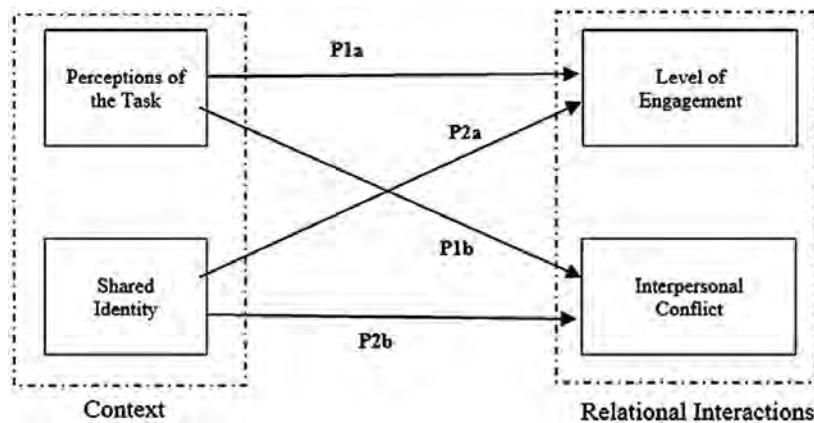
Typically, global virtual teams employ web-based collaboration tools such as Trello, Basecamp, and Sync.in, to name a few (Gilson, Vartiainen, & Hakonen, 2015). While recent web-based technologies provide a great deal of functionalities, anecdotal and empirical evidence suggest a number of challenges faced by teams who rely on computer-based technologies to communicate

and accomplish their tasks. Some of the challenges include overcoming isolation among team members, cultural differences, less time for relationship building, lack of participation, conflict management, and building trust.

This chapter broadens our understanding of this phenomenon by integrating two theoretical approaches—the Hyper-personal Perspective and the Social Constructionist Theory—to examine how members of global virtual teams develop relational interactions. Specifically, the integrated research model presented here suggests that global virtual team members are hyper-sensitive to their computer-mediated interpersonal interactions in that individuals' socially constructed perceptions of the context influence their relational development and judgments much more intensively than those of collocated members.

In this study, the key components of the context are the task-at-hand and shared identity, while relational interactions are defined in terms of interpersonal conflict and level of engagement. Thus, how members of global virtual teams perceive their task and their team will profoundly affect their relational interactions, including how much conflict they experience and how engaged they view their team members as being. Over time, the amplification of these affective elements will affect their judgments about their fellow members' trustworthiness. The main components of our research model and its relationships are depicted in the Figure 1. In the following sections we describe the research model (Figure 1), its components, and their combined impact on relational interactions in global virtual teams.

Figure 1. The research model



RESEARCH MODEL

The Hyper-Personal Perspective

Walther (1996) suggests that the meaning inherent in messages is amplified when individuals have limited physical contact with each other and therefore develop perceptions about others and their relationships based on the available mechanisms at hand, e.g., electronic communication exchanges. In other words, in computer-mediated settings, any piece of information exchanged between team members is likely to take on a significance of its own and is subject to over attribution.

This perspective helps explain why members of distributed groups have the potential to profoundly amplify interpersonal exchanges between the sender and receiver, thereby offering important insights to relational development in virtual teams. For instance, Jones (1995) reported that individuals who participated in online groups exhibited feelings of closeness due to the warmth conveyed in their communication exchanges, compared to collocated groups even though the virtual context would imply physical isolation. These results suggest that members of virtual teams enhanced the meaning of their communication exchanges (beyond the mere messages), thereby enabling them to develop emotional ties that exceeded those in collocated settings.

Social Constructionist Theory

The social constructionist perspective suggests that human social order is produced through interpersonal negotiations and implicit understandings that are built up via shared stories and experiences (Berger & Luckman, 1966). Thus, beliefs held by members of a group determine to what extent meanings of terms are invented and sustained. In other words, how members interpret their context and social interaction processes helps to predict individual cognitions and behavior. An important aspect of the social constructionist approach is its distinction between objective and subjective reality. While the objective reality refers to facts of everyday life that are real or apparent to those who 'inhabit' it, the subjective reality is constructed and reproduced over time. Consequently, given that these subjective interpersonal perceptions help to define reality, perceptions developed over time by members of work groups influence how individuals "objectify" organizational elements around them.

When applying social constructionist notions to virtual teams, "reality" is realized through interpretation whereby team members develop patterns of meanings from their electronic interactions with others. In other words, members' views and perceptions in computer-mediated settings evolve and change over time (Walther, 1996)

based on how they interpret or make sense of their social interactions, which then helps define their reality. To sum up, members will develop socially constructed perceptions of the computer-mediated context based on how they interpret or make sense of their task and team. These social constructions will influence their behavior, and in turn, influence their relational interactions.

THEORETICAL PROPOSITIONS

Drawing upon the previous discussion, we establish the central tenet of our thesis: Global virtual team members will develop perceptions regarding their task and team. These socially constructed perceptions will influence their relational experiences, which, in turn, will impact their judgments about their fellow members' trustworthiness. Furthermore, the hyper-personal perspective suggests that these relationships will be stronger in virtual teams (in comparison to collocated teams) because their members are likely to amplify communication exchanges in the absence of other stimuli (present in collocated teams). The model is depicted in Figure 1 and the variables comprising it are described below.

Perceptions of the Task: An extensive body of research has dealt with the effects of different types of task on group work (Benbasat & Lim, 1993; McGrath, Arrow, Gruenfeld, Hollingshead, & Oconnor, 1993). In general, these studies have defined task as an objective input variable, which then impacts group members' behavior and performance McGrath (1984). Such an approach ignores members' subjective perceptions of the task, which are based on member's sense making processes that evolve over time (Berger & Luckman, 1966). In this sense, behavioral requirements of the task are not conceived as physical characteristics; rather they are viewed as the members' interpretation of what needs to be accomplished (Hackman & Vidmar, 1970).

Consider, for example, the working processes of two teams. A member of one team could inter-

pret a cooperative task [as defined in McGrath's (1984) task circumplex] as requiring conflictive behavior, if, during group interaction, members engage in frequent disagreements or sharp exchanges. Conversely, a conflictive task [as defined in McGrath's (1984) task circumplex] may be interpreted as requiring cooperative behavior if a group member perceives members as sharing their ideas freely and operating in a climate of mutual understanding. Drawn upon these arguments we define *perceptions of the task as a contextual element that is susceptible to interpretation and reinterpretation and may vary as the result of an individual's social interactions.*

Level of Engagement: It refers to the extent to which group members are viewed as being responsive to the needs of the group and reflects the extent of reciprocal interaction among team members (Kramer, 1999). Research suggests that members' perceptions of continuous involvement by and interaction with others help to build trust relationships over time (Rousseau, Sitkin, Burt, & Camerer, 1998). Additionally, in the context of virtual teams, the effects of responsiveness are even more salient because the level of engagement is based on members posting and responding to messages over time (Gefen & Ridings, 2002). Based on these notions, we define *level of engagement as the extent to which members of global virtual teams see their virtual counterparts as being responsive to the needs of the team as they interact through computer-mediate technologies.*

In most group settings, a significant amount of members' efforts involve activities critical to the effectiveness and well-being of the group. Therefore, it is likely that members perceiving their task as being a cooperative venture will engage in more frequent interactions and exchanges with their partners, i.e., be more responsive to the needs of the team. As a result, cooperative perceptions of the task are likely to positively influence the level of engagement in group interactions. Furthermore, as discussed earlier, in the virtual context most of the communication exchanged between partners is done electronically; hence, virtual teams are

likely to over attribute their impressions about contextual structures compared to collocated teams. Thus, how the task is perceived is likely to have a particularly strong effect on members' engagement in virtual teams. Hence, we propose:

P1a: Perceptions of the task will be more related to the level of engagement in global virtual teams than in collocated teams.

Interpersonal Conflict: Interpersonal conflict is targeted at persons within the group and can be detrimental to group work by increasing the intensity of negative attitudes toward others (Deutsch, 1969). Consequently, it can impede trust development and hinder group performance (Jehn & Mannix, 2001). Furthermore, the negative effects of interpersonal conflict may be perceptible even more keenly in virtual teams because their meetings have fewer structural mechanisms that enable them to interpret member's behavior and attitudes over time (Nunamaker, Applegate, & Konsynski, 1988). Hence, we define *interpersonal conflict as the extent to which individuals perceive clashes among team members when working in a global virtual setting*.

A social constructionist view suggests that a member's perceptions of the task will influence how they view fellow members and the ensuing relationships (Yang & Mossholder, 2004). Additionally, given that groups spend considerable time focusing on the task at hand (McGrath, 1991), task type is likely to be a critical factor in understanding relational dimensions such as interpersonal conflict (Jehn, Northcraft, & Neale, 1999). Accordingly, tasks perceived as being cooperative suggest a context where members are likely to perceive their partners as being on the same side, collaborating and helping each other in order to achieve a common goal. Hence, members are likely to develop positive perceptions of their partners, which will reduce interpersonal conflict (Jehn & Mannix, 2001). Conversely, tasks perceived as being contentious implies a context where members see their partners as adversar-

ies. In this context, members are likely to have difficulty developing relational ties and conflict levels are likely to rise. Moreover, as discussed earlier, individuals working in the virtual context are likely to rely greatly on their communication exchanges in developing affective relationships; hence, perceptions of their environment, including the task, will more intensely affect their social interaction processes compared to collocated groups (Walther, 1996). In other words, given the hyper-sensitivity of virtual teams, the positive (or negative) effects of members' perceptions about the task will carry over to their online relationships, including interpersonal conflict. Thus, we propose:

- **P1b:** Perceptions of the task will be more related to interpersonal conflict in global virtual teams than in collocated teams.

Shared Identity: According to Social Identity Theory (Hogg & Terry, 2000; Tajfel, 1978; Turner, 1985) members categorize themselves as well as those around them into two distinct groups: in-group or out-group. Members perceive themselves as being 'ingroup' when they perceive their partners as sharing common values, beliefs, and attitudes as they engage in group interactions over time. Based on these assumptions, *shared identity refers to the extent which an individual identifies with his or her team members when working in a global virtual setting*.

Shared identity has been shown to impact other aspects of group interaction such as members' levels of engagement. Social psychologists have suggested that individuals tend to communicate more, and are more likely to interact with those perceived as sharing a common set of values and work goals. In addition, positive group identity motivates individuals to be more cooperative with others because group membership is likely to increase attachment to and embeddedness with the group (Yang & Mossholder, 2004). In other words, group members who perceive themselves as being part of the group are likely to engage

in frequent interactions with their partners in an effort to remain connected to the group.

Conversely, low levels of shared identity in groups result in reduced communication and coordination (Ancona & Chong, 1996). In virtual contexts, the reduced communication and coordination processes result in members replying to others' requests sporadically. For example, Grinter, Herbsleb, and Perry (1999) found that members of distributed software development teams constantly misinterpreted the activities of their distant colleagues due to the reduced level of communication and lack of coordination among members. Thus, the lack of shared identity appears to lower members' engagement with the group and its activities. Given the earlier discussion about hyper-sensitivity in virtual teams, members of these teams are likely to experience heightened effects of shared identity on levels of engagement in comparison to collocated teams. Hence, we present:

- **P2a:** Shared identity will be more related to the level of members' engagement in global virtual teams than in collocated teams.

A strong body of research indicates that members' evaluations of others actions and behaviors are influenced by their views of shared group identity (Levine & Moreland, 1990). These social mechanisms act as interpretation processes through which group members make sense of their activities (Berger & Luckman, 1966). For instance, many theorists (Hogg & Terry, 2000; Tajfel, 1978; Turner, 1985) suggest that perceptions of shared identity are likely to be closely related to positive evaluations of group members. These positive effects result from the fact that perceptions of increased "groupness" are related to members' perceptions that their attitudes and behavior are similar to those of others in the group (Hogg & Terry, 2000). In other words, individuals tend to evaluate others' actions more positively when they feel that they share common values

and goals. Conversely, in the absence of shared values, team members are likely to have trouble affiliating. Viewed differently, stronger perceptions of shared identity are likely to dampen the development of interpersonal conflict (Jehn et al., 1999; Mortensen & Hinds, 2001). Moreover, given the arguments of hyper-sensitivity, the perceptions of "groupness" may be particularly salient in contexts where no physical contact exists (Lea & Spears, 1991). Hence, we propose:

- **P2b:** Shared identity will be more related to interpersonal conflict in global virtual teams than in collocated teams.

FUTURE RESEARCH DIRECTIONS

The model developed in this study suggests that the beliefs held by group members about contextual variables such as the task-at-hand and shared identity have a differential impact on their relational interactions based on the setting. In collocated settings, where members have the ability to engage in various forms of communication and thus have competing or converging inputs for evaluation, members' perceptions of the task plays only a minor role in shaping their relationships. However, in global virtual settings, where members have to deal with the effects of the geographical distance and thus have fewer mechanisms to observe and evaluate their relational interactions, individuals give greater weight to their perceptions of the computer-mediated work context. These perceptions, in turn, are likely to affect the formation of social ties significantly.

While we expect to test the relationships detailed in the model in future research, the propositions have several practical implications for managing global virtual teams. First, members of such teams can get their work done remotely without having to meet up, anchored to a specific location such as a country, an organizational setting, or a local office. However, often it is difficult for members with different nationalities to

understand each other due to language and cultural differences. In this context, communication is vital for the team's success. Thus, managers may consider employing web-based technologies that enhance cross-cultural communication and team building relationships. For example, tools such as video conferencing and instantaneous message systems may enhance team's ability to monitor and track a project by providing real-time interaction among team members. These tools may ultimately help team members to quickly develop relational interactions since they enable the conveyance of subjective meanings such as tone of voice, facial expressions, and efforts spent toward the completion of the task. In contrast, managers should be careful when implementing asynchronous tools such as emails because they may lead to confusion and frustration due to the fact that co-workers are likely to exchange back-and-forth messages trying to understand subjective meanings attached to each message.

Second, our propositions suggest that level of a member's engagement in group activities is critical to the development of positive perceptions about its peers. Thus, managers should carefully select web-based tools that support activities such as planning, time management, and meeting scheduling. These tools are likely to allow different members of the team to update their specific areas of the project, allowing everyone to see the changes as the project progresses, teams can create regularly scheduled online meetings so that members can avoid wasting time trying to interpret other's input to the task.

Third, the model implies that the reliance on technology to communicate with others and the effects of the distance that separates members can exacerbate negative interactions and emotions, whereby team members tend to over attribute meaning to the information contained in computer-mediated messages. One promising technology to enhance team work and perhaps minimize negative meaning attached to limited text-based messages is cloud computing. Given that members are geographically dispersed and have a number

of other organizational activities outside of their team project, individuals often have limited time to meet to be able to discuss their progress. This is where cloud systems come into play. Managers may use cloud computing resources to set up a platform that lets individuals log in to a central repository of data and documents. This would allow team members to access and share files from different locations at any time. In addition, the repository may contain a history of group communication and relevant task information, which, in turn, allows members to capture, store, and visualize members' interactions and contributions to the task at a hand. In addition, teams can work on the project from where ever they are on their portable devices (without relying to any specific local server) and still get the job done.

From a research perspective this study offers a number of insights for future research. First, scholars interested in examining social components affecting team performance in global virtual teams may consider testing the relationships depicted in our model in longitudinal studies. For example, our model suggests that the effects of technologies on group work should not be divorced from the processes that emerge as individuals use these technologies and construct perceptions about their ongoing interactions over time. The impact of these tools may have a differential effect on individuals' perceptions as teams interact over time. Therefore, longitudinal studies can incorporate the effects of time to examine how these relationships unfold as the team progresses in their assigned task.

CONCLUSION

The purpose of this study was to explore the mechanisms that influence levels of engagement, interpersonal conflict, and trust in global virtual teams. Overall, the model supports the idea of hyper-sensitivity in virtual teams by suggesting that perceptions of the context influence their ongoing interaction processes in ways that exceed those of collocated teams. In other words, in the absence

of physical proximity, members gave primacy to their available sources of information—such as perceptions of their task, which then informs their ongoing interactions with others. These interactions, defined by interpersonal conflict, then color their judgments about others, including trust. The detrimental effects of interpersonal conflict on the development of trust suggest that negative relational interactions in virtual teams have a more serious effect on trust development (than in the comparable collocated setting).

Furthermore, in global virtual teams, perceptions of the task are likely to have an inverse relationship with interpersonal conflict. In addition, perceptions of the task are expected to have an amplified effect on the level of members' engagement during group interaction in the global virtual teams. This relationship is expected to be weaker in the collocated teams. Taken together, these propositions support the social constructionist view of virtual work in that how the task is viewed affects members' relational ties and evaluations of their peers—in some settings. In other words, in the absence of a collocated social context, members of a virtual team may pay more attention to the actual messages exchanged through their text-based interaction, rather than multidimensionally evaluating the socio-emotional climate of their context.

While contextual elements are likely to impact members' relational interactions and behavior, how individuals 'objectify' these structures defines their social interactions. Our findings support the idea of hyper-sensitivity in virtual teams, which tend to experience a disproportionately greater effect of task perceptions on social interactions compared to their collocated counterparts. Thus, the hyper-personal perspective serves a vital function by enabling virtual teams to make sense of their environment using the available information and engaging in the process of relational development.

REFERENCES

- Ancona, D., & Chong, C. L. (1996). Entrainment: Pace, cycle, and rhythm in organizational behavior. *Research in Organizational Behavior*, 18, 251–284.
- Benbasat, I., & Lim, L. H. (1993). The effects of group, task, context, and technology variables on the usefulness of group support systems - A metaanalysis of experimental studies. *Small Group Research*, 24(4), 430–462. doi:10.1177/1046496493244002
- Berger, P. L., & Luckman, T. (1966). *The Social Construction of Reality: A Treatise in the Sociology of Knowledge*. Garden City, NY: Anchor Books.
- Deutsch, M. (1969). Conflicts: Productive and destructive. *The Journal of Social Issues*, 25(1), 7–41. doi:10.1111/j.1540-4560.1969.tb02576.x
- Gefen, D., & Ridings, C. M. (2002). Implementation team responsiveness and user evaluation of customer relationship management: A quasi-experimental design study of social exchange theory. *Journal of Management Information Systems*, 19(1), 47–69.
- Gilson, L., Vartiainen, M., & Hakonen, M. (2015). Virtual Teams Research: 10 Years, 10 Themes, and 10 Opportunities. *Journal of Management*, 41(5), 1313–1337. doi:10.1177/0149206314559946
- Grinter, R. E., Herbsleb, J. D., & Perry, D. E. (1999). *The geography of coordination: Dealing with distance in r&d work*. Paper presented at the ACM Conference on Supporting Group Work (GROUP 99), Phoenix, AZ. doi:10.1145/320297.320333
- Hackman, J. R., & Vidmar, N. (1970). Effects of size and task type on group performance and member reactions. *Sociometry*, 33(1), 37–54. doi:10.2307/2786271

- Hogg, M. A., & Terry, D. J. (2000). Social identity and self-categorization processes in organizational contexts. *Academy of Management Review*, 25(1), 121–140.
- Jehn, K. A., & Mannix, E. A. (2001). The dynamic nature of conflict: A longitudinal study of intragroup conflict and group performance. *Academy of Management Journal*, 44(2), 238–251. doi:10.2307/3069453
- Jehn, K. A., Northcraft, G. B., & Neale, M. A. (1999). Why differences make a difference: A field study of diversity, conflict, and performance in workgroups. *Administrative Science Quarterly*, 44(4), 741–763. doi:10.2307/2667054
- Jones, S. G. (1995). Understanding community in the information Age. In S. G. Jones (Ed.), *Cybersociety: Computer-mediated Communication and Community* (pp. 10–35). Thousand Oaks, CA: Sage.
- Kramer, R. M. (1999). Trust and distrust in organizations: Emerging perspectives, enduring questions. *Annual Review of Psychology*, 50(1), 569–598. doi:10.1146/annurev.psych.50.1.569 PMID:15012464
- Lea, M., & Spears, R. (1991). Computer mediated communication, de-individuation and group decision-making. *International Journal of Man-Machine Studies*, 34(2), 283–301. doi:10.1016/0020-7373(91)90045-9
- Levine, J. M., & Moreland, R. L. (1990). Progress in small group research. *Annual Review of Psychology*, 41(1), 585–634. doi:10.1146/annurev.ps.41.020190.003101
- McGrath, J. E. (1984). *Group Interaction and Performance*. Englewood Cliffs, NJ: Prentice-Hall.
- McGrath, J. E. (1991). Time, interaction, and performance (TIP) - A theory of groups. *Small Group Research*, 22(2), 147–174. doi:10.1177/1046496491222001
- McGrath, J. E., Arrow, H., Gruenfeld, D. H., Hollingshead, A. B., & O'Connor, K. M. (1993). Groups, tasks, and technology - The effects of experience and change. *Small Group Research*, 24(3), 406–420. doi:10.1177/1046496493243007
- Mortensen, M., & Hinds, P. J. (2001). Conflict and shared identity in geographically distributed teams. *The International Journal of Conflict Management*, 12(3), 212–238. doi:10.1108/eb022856
- Nunamaker, J. F., Applegate, L. M., & Konsynski, B. R. (1988). Computer-aided deliberation: Model management and group decision support. *Journal of Operations Research*, 36(6), 826–848. doi:10.1287/opre.36.6.826
- Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review*, 23(3), 393–404. doi:10.5465/AMR.1998.926617
- Tajfel, H. (1978). *Differentiation between Social Groups: Studies in the Social Psychology of Intergroup Relations*. London: Academic Press.
- Turner, J. C. (1985). Social categorization and the self-concept: A social cognitive theory of group behavior. *Advances in Group Processes*, 2, 77–122.
- Walther, J. B. (1996). Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction. *Communication Research*, 23(1), 3–43. doi:10.1177/009365096023001001
- Yang, J., & Mossholder, K. W. (2004). Decoupling task and relationship conflict: The role of intragroup emotional processing. *Journal of Organizational Behavior*, 25(5), 589–605. doi:10.1002/job.258

ADDITIONAL READING

Gibson, C. B., Huang, L., Kirkman, B. L., & Shapiro, D. L. (2014). Where Global and Virtual Meet: The Value of Examining the Intersection of These Elements in Twenty-First-Century Teams. *Annual Review of Organizational Psychology and Organizational Behavior*, 1(1), 217–244. doi:10.1146/annurev-orgpsych-031413-091240

Gilson, L., Vartiainen, M., & Hakonen, M. (2015). Virtual Teams Research: 10 Years, 10 Themes, and 10 Opportunities. *Journal of Management*, 41(5), 1313–1337. doi:10.1177/0149206314559946

Hoch, J., & Kozlowski, S. W. J. (2014). Leading virtual teams: Hierarchical leadership, structural supports, and shared team leadership. *The Journal of Applied Psychology*, 99(3), 390–403. doi:10.1037/a0030264 PMID:23205494

Lipnack, J., & Stamps, J. (2000). *Virtual Teams: People Working across Boundaries with Technology* (2nd ed.). New York: John Wiley & Sons.

Van Ryssen, S., & Godar, S. H. (2000). Going international without going international: Multi-national virtual teams. *Journal of International Management*, 6(1), 49–60. doi:10.1016/S1075-4253(00)00019-3

Wilson, J., Crisp, C. B., & Mortensen, M. (2013). Extending Construal-Level Theory to Distributed Groups: Understanding the Effects of Virtuality. *Organization Science*, 24(2), 629–644. doi:10.1287/orsc.1120.0750

Windeler, J. B., Maruping, L. M., Robert, L. P., & Riemenschneider, C. K. (2015). E-profiles, Conflict, and Shared Understanding in Distributed Teams. *Journal of the Association for Information Systems*, 16(7), 608–645.

KEY TERMS AND DEFINITIONS

Hyper-Sensitivity Perspective: This perspective helps explain why members of distributed groups have the potential to profoundly amplify interpersonal exchanges between the sender and receiver

Interpersonal Conflict: Refers to the extent to which individuals perceive clashes among team members when working in a global virtual setting. Typically, it is detrimental to group work by increasing the intensity of negative attitudes toward others.

Level of Engagement: Refers to the extent to which members of global virtual teams see their virtual counterparts as being responsive to the needs of the team as they interact through computer-mediate technologies.

Perceptions of the Task: It is a contextual element that is susceptible to interpretation and reinterpretation and may vary as the result of an individual's social interactions.

Shared Identity: It refers to the extent which an individual identifies with his or her team members when working in a global virtual setting.

Social Constructionist Perspective: It suggests that human social order is produced through interpersonal negotiations and implicit understandings that are built up via shared stories and experiences (Berger and Luckman, 1967). Thus, beliefs held by members of a group determine to what extent meanings of terms are invented and sustained. In other words, how members interpret their context and social interaction processes helps to predict individual cognitions and behavior.

Lean and Six Sigma Innovation and Design

B

Rick Edgeman

Utah State University, USA

INTRODUCTION

Six Sigma Innovation and Design theory, strategy and supporting methods have evolved along two primary pathways – one focused on significant innovation of existing products, services, processes or systems, and a second focused on design of new products, services, products or systems. A third, more recently emerged path is referred to as LSS or *Lean Six Sigma* (Pepper & Spedding, 2010). LSS integrates and leverages the sizeable commonalities, synergies and strengths of Six Sigma and Lean Enterprise theory and methods while ameliorating their weaknesses and distinctions. In particular, LSS wedges key lean methodologies and perspectives such as value stream mapping, kaizen (continuous improvement), and waste identification and minimization together with equally key Six Sigma concepts and tools such as its focused approaches to innovation and design – DMAIC and DMADV.

Six Sigma in the late 1970s at Motorola Corporation, but it is GE with which Six Sigma is most familiarly associated. Its use has proliferated due in large to its acknowledged contribution of multiple billions of dollars to the economic performance of many enterprises. Six Sigma delivers either desirable and reliable new designs or significant improvements in existing products, processes, systems of key importance to customers or – more generally – to relevant enterprise stakeholders. This is accomplished by identifying and subsequently exacting change in key measurable and internally controllable levers of change that drive outputs highly relevant to those stakeholders. Such outputs are referred to as critical-to-quality (CTQ) characteristics and directly reflect preferences of customers of the product, process or systems

that is being addressed. It is because we cannot directly control these CTQs that we must instead identify and control levers of change. Doing so begins with a clear and elaborated definition and understanding of customer and other relevant stakeholder needs, desires, and expectations – the so-called voice of the customer (VOC).

Lean Enterprise theory and approaches are often associated with Toyota Corporation and the familiar *Toyota Production System* (TPS) credited to late Toyota executive Taiichi Ohno and his mentor, Shigeo Shingo (Schmenner, 2015). Lean Enterprise methods are also associated with Kaizen (continuous improvement) philosophy and methods popularized by Masaki Imai (Antony, 2015). The arc of Lean Enterprise projects and enterprises is one typically aimed at waste reduction leading to improved efficiency, increased reliability, enhanced design, and better resource utilization so that such projects tend to be internally-focused, with derivative value for the customer.

Both Six Sigma and Lean Enterprise emphasize near relentless pursuit of perfection. Lean Enterprise methods do so via continuous incremental improvement cycles with an eye toward all enterprise processes and activities. In contrast, Six Sigma projects are typically discrete in nature and target breakthrough improvement in strategically important processes, products or systems. As such, the union of Lean Enterprise with Six Sigma is both internally and customer focused, taking simultaneous aim at both cost savings and value creation.

Six Sigma Innovation, Design for Six Sigma, Lean Enterprise and – more specifically – their integration that resulting in Lean Six Sigma are discussed. Although each of these has been his-

DOI: 10.4018/978-1-5225-2255-3.ch063

torically emphasized financial objectives, in principle they be used to singly or jointly address any number of objectives, including financial, social, or ecological performance, or anything contributing to organizational resilience and robustness (Edgeman, 2013). Also discussed are distinctions between the COPIS approach to business process conception prior to process implementation and execution via SIPOC (Edgeman, 2011a); commonly used supporting tools and techniques such as the Kano Needs Model and Quality Function Deployment or QFD (Tan & Shen, 2010); and concept generation and selection (Girotra, Terwiesch and Ulrich, 2010).

BACKGROUND

Of many competing Six Sigma and Lean Six Sigma definitions, the following, adapted from Klefsjö, Bergquist and Edgeman (2006), is herein employed:

Lean Six Sigma provides highly structured innovation, design, and lean enterprise strategies and methods for acquiring, assessing, and activating customer, competitor, and enterprise intelligence in order to deliver superior product, process, system, or enterprise performance that benefits all relevant stakeholders through best and next best practices and sources of sustainable competitive advantage.

Six Sigma is not unique in this quest for exceptional performance and competitive advantage. Rather, it is the combination of Six Sigma's strategic focus, structured approaches, breakthrough performance levels sought, and relatively short time horizons within which such improvement is demanded that distinguish it from other improvement, design, and innovation approaches.

Six Sigma's is provided by its focused design and innovation approaches. Its focused innovation algorithm is referred to as DMAIC and is a simple, yet logical scheme that demands the

project in question to be carefully defined (D), with definition followed by measurement (M), analysis (A), improvement (I), and control (C) phases. Design for Six Sigma (DFSS) projects may employ any of a number of approaches, with the most common being DMADV, an acronym for Define-Measure-Analyze-Design-Verify (Edgeman, 2011b). While there are similarities between DMAIC and DMADV, there are also key differences, including in the intentions behind Define, Measure and Analyze in these algorithms (Cronemyr, 2007).

Specific content of each step in DMAIC and DMADV depends on enterprise and competitive context, the knowledge array resident in project team members, and the disciplinary traditions of those team members. Superior performance may be defined in absolute terms or specific to competitive context. Sustainable competitive advantage often relies not on application of Six Sigma or other strategies per se, but rather on enterprise enculturation and the effective and efficient use of such approaches in areas of strategic importance.

The need to significantly improve product and process performance through innovation provided much of the initial impetus behind Six Sigma, yet it is perhaps its status as a documented driver of superior financial performance that has led to its proliferation, subsequent diversification to design and lean environments, and dissemination across a number of application domains. In addition to traditional manufacturing applications of Lean and Six Sigma, significant gains in a number of "soft" or service application areas have been realized and include financial services (De Koning, Does, & Bisgaard, 2008), regional and national security (Edgeman, Bigio, & Ferleman, 2005), healthcare (Kaplan, Bisgaard, Truesdell, & Zetterholm, 2009) including surgery (Mason, Nicolay & Darzi, 2015), and energy production and distribution (Kaushik & Khanduja, 2009). In many of these latter applications, financial performance has been of secondary or tertiary importance and other considerations, such as ecological or societal sustainability has been deemed preeminent.

Table 1. DPMO values in relation to sigma levels

σ Level	DPMO When Perfectly Centered	DPMO With 1.5σ Displacement	Cost of Poorly Performing Process (CP ³)
2	46,000	308,537	Cost may exceed sales
3	2,700	66,807	25% to 40% of sales
4	60	6,210	15% to 25% of sales
5	0.6	233	5% to 15% of sales
6	0.002	3.4	< 1% of sales.

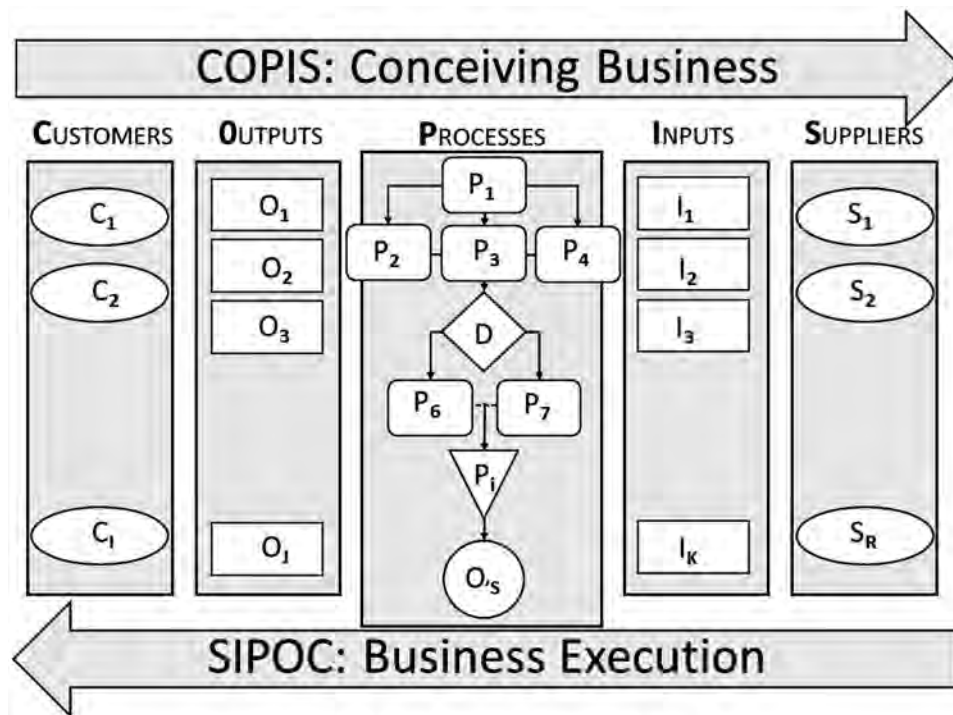
The key issue of “what is sigma?” remains. Symbolized by σ , the term ‘sigma’ is a measure of variation or “imperfection” widely recognizable as the standard deviation of process, product, or system output. Contextually, variation does not represent intentional introduced diversity, but is rather any departure from intended performance levels. *Higher* process sigma levels imply *lesser* standard deviation values so that higher sigma levels imply that a higher proportion of output or results will be associate with performance ranging from acceptable to superior. Higher sigma levels thus imply reduced ‘defect’ levels where a defect is anything not matching a required performance profile—for example—true Six Sigma is ordinarily associated with “near perfect performance” that is often cited as “3.4 defects per million opportunities for a defect”.

The method used to estimate a sigma level differs depending on whether quantitative or qualitative information is being assessed, but in either case captures the capability of the underlying process to deliver required results. Relative to prior discussion we find that ‘high sigma levels’ are associated with more capable processes. Defects per million opportunities for defects (DPMO) in relation to sigma levels and associated cost or loss due to imperfections are reported in Table 1, similar content of which may be found in a large number of sources, including Montgomery and Woodall (2008). DPMO values in Table 1 reflect an average process performance level displaced by 1.5σ from perfect centering. It is not always the case that a process should perform *within* limits,

since there are numerous reasonable circumstances where there is only an *upper* or *lower* acceptable performance limit. As revealed in Table 1, the commonly reported ‘3.4 DPMO for a true six sigma level process’ reflects displacement or a shift of 1.5σ from perfect centering, that is, from ideal performance level. Use of a 1.5σ displacement or drift factor – referred to in statistical parlance as a non-centrality factor – is based on the natural tendency of processes to vary through time within a range – that is, within ‘plus or minus’ distance of its (natural) average performance level, rather than ideal performance level).

Left alone, most processes tend not only to drift, but in fact have a tendency toward entropy or decaying performance so that process performance is often monitored and managed. There may be any number of motivations for improving performance, but it is this tendency toward entropy that necessitates the more mundane activity of monitoring and controlling a process at a stable level via tools such as failure modes and effects analysis (FMEA) or statistical process control charts (Montgomery, 2008; Shahin, 2004). Process monitoring and control ordinarily represents both pre- and post- Six Sigma project activity as does the associated use of continuous improvement practices such as use of the PDCA (Plan, Do, Study, Act) cycle that is central to kaizen and that is also known as the Deming cycle (Moen & Norman, 2010). Application of kaizen after breakthrough improvement has been attained via Six Sigma will generate ongoing incremental improvements.

Figure 1. Six Sigma COPIS $\Rightarrow$ SIPOC translation



SELECTED LEAN SIX SIGMA STRATEGIES AND APPROACHES

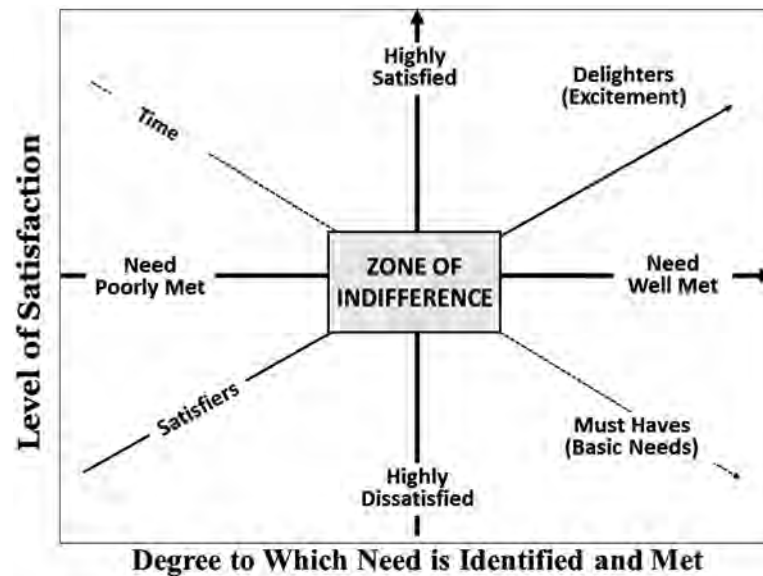
All Six Sigma strategies explicitly or implicitly have customer or, more generally, key stakeholder considerations at their core, including internal stakeholders. For that reason, both COPIS and SIPOC are central to most Six Sigma formulations. SIPOC is familiar as a generic business execution process whereas COPIS reflects how business should first be conceived before it is practiced. The letters are identical, but their flow reversed. COPIS indicates that the enterprise should first carefully elicit, elaborate, and assess customer (C) / stakeholder needs and desires that provide solid process output (O) indicators. These indicators then supply information concerning optimal configuration of processes (P) capable of delivering the outputs. Process configuration and requirements then influence inputs (I) and the suppliers (S) from whom inputs are obtained. This information is often used to construct a “COPIS

map” and to feed SIPOC practice. These concepts are illustrated in Figure 1.

All elements of the COPIS $\Rightarrow$ SIPOC translation chain are important, but clear and careful articulation and elaboration of customer / stakeholder needs is particularly so, since failure in this domain will almost certainly contribute to project failure. Generation and selection of poor concepts, or inadequate later deployment of strong concepts intended to fulfill customer needs will in like manner typically lead to project failure.

The most often used customer / stakeholder needs elaboration model is the Kano Customer Needs Model (Matzler & Hinterhuber, 1999) as portrayed in Figure 2. This model differentiates customer / stakeholder needs and wants into the self-explanatory categories of dissatisfiers, satisfiers, and excitors-delighters. Such needs may be further classified according to either service or product quality dimensions that include aesthetics, reliability, features, conformance, durability, serviceability, perceived quality, performance, responsiveness, competence, security, empathy,

Figure 2. Kano customer / stakeholder needs model



courtesy, access and communication (Golder, Mitra, & Moorman, 2012).

Included among increasingly important design and innovation considerations are biologically inspired biomimetic ones (Hartmann & Germain, 2015), and life-cycle / cradle-to-cradle / circular economy considerations (Bocken, 2016; Tempelman et al., 2015). In any case, the selected concept must ultimately be developed and deployed, with common deployment methods being QFD and axiomatic design (Rauch, Dallasega, & Matt, 2015). QFD matches customer / stakeholder needs or “what’s” to enterprise capabilities or “how’s” and then deploys those capabilities in order to yield the product, process, service or system and can

be adapted to accommodate changing priorities in customer needs (Asadabadi, 2016).

SIX SIGMA INNOVATION AND DESIGN

As previously noted, Lean Six Sigma Innovation projects customarily employ the DMAIC algorithm and Lean DFSS projects will ordinarily make use of the DMADV algorithm. These algorithms are described in brief in Tables 2 and 3.

Whether it is motivated by lean, innovation, or design considerations, a Six Sigma project is essentially always a team endeavor and as such

Table 2. Six sigma innovation DMAIC algorithm

DMAIC Phase		Description
D	Define	Define the problem and customer / stakeholder needs and requirements. Prepare project team charter.
M	Measure	Measure product, process or service defect rates and document constituent processes in their current incarnation.
A	Analyze	Analyze process data in order to set performance objectives, identify key sources of variability in terms of CTQs, and determine the overall current capability to fulfill those objectives.
I	Improve	Replace any environmentally or humanly damaging substances or actions with more environmentally or socially friendly alternatives.
C	Control	Rethink every product or service and functions thereof. Are they needed? Are they efficient? Are they ecologically and socially friendly?

Table 3. Design for Six sigma DMADV algorithm

DMADV Phase		Description
D	Define	The Define phase of DFSS aims to acquire and access the Voice of the Customer (VOC) and subsequently align goals for the product, process, service, or system with the VOC. Among VOC acquisition methods are use of focus groups; sample surveys; direct observation of customer use of similar product, processes or systems so that unspoken and more implicit information can be gathered; and examination of customer complaints. Goals should be so-called SMART (Specific, Measurable, Attainable, Relevant, and Time-bound) stretch goals – ones that are more likely to position the product, process, or service at the leading edge.
M	Measure	In the DFSS context measurement entails measuring and matching performance to customer requirements in order to quantify the VOC and align it with organizational and management goals.
A	Analyze	Analysis demands that the design of existing relevant products, processes, services or systems must be assessed to identify or determine suitability, performance, error or defect sources, and any corrective or innovative actions that may implemented. Among the many tools that may be useful in the Analyze phase are Design Failure Modes and Effects Analysis or DFMEA (Chen & Ko, 2009), concept generation and selection, and the Theory of Inventive Problem Solving or TRIZ (Ilevbare, Probert & Phaal, 2013).
D	Design	Design requires that the array of corrective or innovative actions identified in the analyze phase are embedded in the design and subsequent deployment of new processes required to concurrently activate the VOC and fulfill organizational and management goals. QFD, EVOP, and RSM are particularly valuable in this phase.
V	Verify	The objective of the verify phase is to assess performance of the design via such means as simulation, direct observation of the new design in its use environment but prior to marketplace deployment, or prototyping.

common to the Define phase is construction of, and agreement to a team charter that establishes purpose, goals, objectives, and code of conduct and engagement. While specific content of team charter may vary, elements that are typically included are an executive summary, projective objectives, project scope, the business case for the project, project organization, schedules, a communication plan, project control procedures, project assumptions, and a conflict resolution agreement. A key driver of the conflict resolution agreement specifically and the team charter more generally is to transform conflict discussions from that may otherwise be personality-driven into ones that are more issue-oriented. This discussion highlights the importance of good project management to conduct of any Six Sigma project. Similarly, all Six Sigma projects have a process orientation so that a value stream map (Gibbons, Kennedy, Burgess, & Godfrey, 2012) and one or more, more specific process maps are generally included.

Most Six Sigma projects will ultimately call for comparison of performance of an existing or conceived product, process or system to similar ones or to similar concepts so that some form of benchmarking (Watson, 2007) is typically re-

quired. In addition to previously cited ones, other often-used and widely familiar Six Sigma tools and methods span project management ones such as work breakdown structures and Gantt charts; ideation and creativity ones such as brainstorming, affinity diagrams, interrelationship digraphs, mind mapping, and storyboarding; and more analytically-based decision support ones such as the matrix diagrams, nominal group technique, theory of constraints / evaporating cloud (Sproull & Nelson, 2012), analytical hierarchy process, and SWOT (strengths, weaknesses, opportunities and threats) analysis.

Given the origin of the “ σ ” terminology it is unsurprising that various statistical and graphical methods ranging from simple to complex may also be brought to bear in Six Sigma projects. Among such tools and methods are Pareto charts, cause-and-effect diagrams, run charts, control charts, boxplots, hypothesis testing, designed experiments, and regression and correlation analysis with results that may be summarized by radar charts or displays such as performance dashboards.

While selected tools and techniques may generally prove more useful in one phase than another, a given tool or technique may be applied anywhere

useful in the DMAIC or DMADV structures. It is also true that two Six Sigma teams approaching the same issue may attack the issue differently and make use of differing tools and techniques and hence produce differing solutions. In any case and relative to constraints, a given team will endeavor to deliver as near to an ideal final result (IFR) as is possible. Numerous practical examples of both DMAIC and DMADV implementation can be found in the academic literature, with an especially rich resources being the journals *Total Quality Management and Business Excellence*, *International Journal of Lean Six Sigma*, *The TQM Journal*, and *International Journal of Six Sigma and Competitive Advantage*.

LEAN SIX SIGMA

Lean Six Sigma (LSS) derives from integration of lean enterprise methods that seek to reduce or eliminate waste with Six Sigma's quest for near perfect performance, hence LSS simultaneously seeks resource savings, cost savings, and value creation. Integration drives both the structure and content of specific LSS improvement efforts (Gremyr & Fouquet, 2012). LSS is a relatively new development with no uniform approach and literature that is immature relative to that for DMAIC and DMADV.

Many lean approaches derive from TPS (Ohno, 1988). TPS identifies multiple forms of waste (*muda*) that may be easily recalled via the acronym NOW TIME that represents non-quality (N), over-production (O), waiting (W), transportation (T), inventory (I), motion (M), and excess processing (E). Often non-value adding efforts in a process represent unnecessary waste. Womack and Jones (2003) add the manufacture of goods or supply of services that do not meet customer demands and expectations to the NOW TIME list of wastes that may also incorporate untapped or under-used human talent and creativity. Similarly, value stream mapping is helpful in pinpointing the

best place or places in a process where strategic change impacting CTQs can be executed.

FUTURE RESEARCH DIRECTIONS

A number of articles cited under "additional references" propose future research directions for Six Sigma directions. Since the DMAIC and DMADV algorithms follow well-established problem-solving approaches, the most obvious future research directions for Six Sigma are ones that will be associated with new applications areas, and adaptation of tools and techniques from other fields, many of which are suggested by Edgeman and Dugan (2008). In particular, environmental deterioration and pressing societal needs are areas where radical improvement needs abound and hence where application of Six Sigma Innovation, Design, and Lean strategies and methods should be carefully, yet urgently explored (Kumar, Kumar & Haleem, 2015).

Many projects combine significant innovation with new design so that development of one or DMAIC-DMADV hybrids can be anticipated. Similarly, it can be expected that effort will be dedicated to development of standardized or semi-standardized mixtures of Lean and Six Sigma methods to produce a LSS algorithm akin to DMAIC and DMADV. Such future directions are relatively predictable.

Less predictable is the issue of how Lean Six Sigma will probe and expand the "big data frontier" that is commonly viewed as critical to advancing innovation, competition, and productivity (Dubey et al., 2016). In large the question is one of how Six Sigma will morph to incorporate both the challenges and advantages of computationally intensive methods in order to effectively and efficiently leverage massive amounts of information and disparate sources of enterprise, social, and competitive intelligence. That such transformation will occur is almost certain and is also likely to spur refinement or significant innovation of such Six

Sigma tools and techniques that address customer or stakeholder needs elaboration, assessment and deployment.

CONCLUSION

Whether applied in innovation, design, or lean contexts, Six Sigma aims to deliver superior value through superior performance of products, processes or systems that are of critical importance to key enterprise stakeholders. Discussed herein have been the means by which this accomplished.

Use of Six Sigma has proliferated due in large to significant documented contributions to the financial bottom lines of numerous enterprises. While superior financial performance often provides the initial impetus behind Six Sigma adaptation, it must be emphasized that financial performance generally follows significant enhancement of products, processes or systems specific to enterprise context that may be as disparate as surgical procedures, harvesting methods, service delivery, a new manufacturing process, or related to safety and security.

REFERENCES

- Antony, J. (2015). The ten commandments of quality: A performance perspective. *International Journal of Productivity and Performance Management*, 64(5), 723–735. doi:10.1108/IJPPM-10-2014-0167
- Asadabadi, M. R. (2016). A Markovian-QFD approach in addressing changing priorities of the customer needs. *International Journal of Quality & Reliability Management*, 33(8), 1–14. doi:10.1108/IJQRM-07-2014-0091
- Bocken, N. M. P., de Pauw, I., Bakker, C., & van der Grinten, B. (2016). Product design and business model strategies for a circular economy. *Journal of Industrial and Production Engineering*, 33(5), 308–320. doi:10.1080/21681015.2016.1172124
- Chen, L.-H., & Ko, W.-C. (2009). Fuzzy linear programming models for new product design using QFD with FMEA. *Applied Mathematical Modelling*, 33(2), 633–647. doi:10.1016/j.apm.2007.11.029
- Cronemyr, P. (2007). DMAIC and DMADV: Differences, similarities and synergies. *International Journal of Six Sigma and Competitive Advantage*, 3(3), 193–209. doi:10.1504/IJSSCA.2007.015065
- Dubey, R., Gunasekaran, A., Childe, S. J., Wamba, S. F., & Papadopoulos, T. (2016). The impact of big data on world-class sustainable manufacturing. *International Journal of Advanced Manufacturing Technology*, 84(1-4), 631–645. doi:10.1007/s00170-015-7674-1
- Edgeman, R. (2011a). SIPOC and COPIS: The business flow–business optimization connection. In M. Lovric (Ed.), *International encyclopedia of statistical sciences* (pp. 1337–1338). Berlin, Germany: Springer. doi:10.1007/978-3-642-04898-2_522
- Edgeman, R. (2011b). Design for six sigma. In M. Lovric (Ed.), *International encyclopedia of statistical sciences* (pp. 374–376). Berlin, Germany: Springer. doi:10.1007/978-3-642-04898-2_25
- Edgeman, R. (2013). Six sigma enablement of sustainable enterprise excellence & resilience. *Six Sigma Forum*, 12(4), 6-9.
- Edgeman, R., Bigio, D., & Ferleman, T. (2005). Six sigma and business excellence: Strategic and tactical examination of IT service level management at the Office of the Chief Technology Officer of Washington DC. *Quality and Reliability Engineering International*, 21(3), 257–273. doi:10.1002/qre.635
- Edgeman, R., & Dugan, J. (2008). Six sigma from products to pollution to people: Migration from engineering and business to the natural and social environments. *Total Quality Management and Business Excellence*, 19(1), 1–8. doi:10.1080/14783360701601918

- Gibbons, P., Kennedy, C., Burgess, S., & Godfrey, P. (2012). The development of a value improvement model for repetitive processes (VIM): Combining lean, six sigma and systems thinking. *International Journal of Lean Six Sigma*, 3(4), 315–338. doi:10.1108/20401461211284770
- Girotra, K., Terwiesch, C., & Ulrich, K. T. (2010). Idea generation and the quality of the best idea. *Management Science*, 56(4), 591–605. doi:10.1287/mnsc.1090.1144
- Golder, P., Mitra, D., & Moorman, C. (2012). What is quality? An integrative framework of processes and states. *Journal of Marketing*, 76(4), 1–23. doi:10.1509/jm.09.0416
- Gremyr, I., & Fouquet, J. (2012). Design for six sigma and lean product development. *International Journal of Lean Six Sigma*, 3(1), 45–58. doi:10.1108/20401461211223722
- Hartmann, J., & Germain, R. (2015). Understanding the relationships of integration capabilities, ecological product design, and manufacturing performance. *Journal of Cleaner Production*, 92, 196–205. doi:10.1016/j.jclepro.2014.12.079
- Ilevbare, I. M., Probert, D., & Phaal, R. (2013). A review of TRIZ, and its benefits and challenges in practice. *Technovation*, 33(2), 30–37. doi:10.1016/j.technovation.2012.11.003
- Kaplan, S., Bisgaard, S., Truesdell, D., & Zetterholm, S. (2009). Design for six sigma in healthcare: Developing an employee influenza vaccination process. *Journal for Healthcare Quality*, 31(3), 36–43. doi:10.1111/j.1945-1474.2009.00029.x PMID:19522346
- Kaushik, P., & Khanduja, D. (2009). Application of six sigma DMAIC methodology in thermal power plants: A case study. *Total Quality Management & Business Excellence*, 20(2), 197–207. doi:10.1080/14783360802622995
- Klefsjö, B., Bergquist, B., & Edgeman, R. (2006). Six sigma and total quality management: Different day, same soup? *International Journal of Six Sigma & Competitive Advantage*, 2(2), 162–178. doi:10.1504/IJSSCA.2006.010107
- Kumar, S., Kumar, N., & Haleem, A. (2015). Conceptualisation of sustainable green lean six sigma: An empirical analysis. *International Journal of Business Excellence*, 8(2), 210–250. doi:10.1504/IJBEX.2015.068211
- Mason, S. E., Nicolay, C. R., & Darzi, A. (2015). The use of lean and six sigma methodologies in surgery: A systematic review. *The Surgeon*, 13(2), 91–100. doi:10.1016/j.surge.2014.08.002 PMID:25189692
- Matzler, K., & Hinterhuber, H. (1998). How to make product development projects more successful by integrating Kanos model of customer satisfaction into quality function deployment. *Technovation*, 18(1), 25–38. doi:10.1016/S0166-4972(97)00072-2
- Moen, R., & Norman, C. (2010). Circling back: Clearing up myths about the Deming cycle and seeing how it keeps evolving. *Quality Progress*, 43(11), 22–28.
- Montgomery, D. (2008). *Introduction to statistical quality control*, 6e. Hoboken, NJ: John Wiley.
- Montgomery, D., & Woodall, W. (2008). An overview of six sigma. *International Statistical Review*, 76(3), 329–346. doi:10.1111/j.1751-5823.2008.00061.x
- Ohno, T. (1988). *Toyota Production System: Beyond large-scale production*. Cambridge, MA: Productivity Press.
- Pepper, M., & Spedding, T. (2010). The evolution of lean six sigma. *International Journal of Quality & Reliability Management*, 27(2), 138–155. doi:10.1108/02656711011014276

Rauch, E., Dallasega, P., & Matt, D. T. (2015). Axiomatic design based guidelines for the design of a lean product development process. *Procedia CIRP*, 34, 112–118. doi:10.1016/j.procir.2015.07.005

Schmenner, R. W. (2015). The pursuit of productivity. *Production and Operations Management*, 24(2), 341–350. doi:10.1111/poms.12230

Shahin, A. (2004). Integration of FMEA and the Kano model: An exploratory examination. *International Journal of Quality & Reliability Management*, 21(7), 731–746. doi:10.1108/02656710410549082

Sproull, B., & Nelson, B. (2012). *Epiphanized: Integrating theory of constraints, lean and six sigma*. Great Barrington, MA: North River Press.

Tan, K., & Shen, X. (2010). Integrating Kanos model in the planning matrix of quality function deployment. *Total Quality Management*, 11(8), 1141–1151. doi:10.1080/095441200440395

Tempelman, E., de Pauw, I. C., van der Grinten, B., Mul, E. J., & Grevers, K. (2015). Biomimicry and cradle to cradle in product design: An analysis of current design practice. *Journal of Design Research*, 13(4), 326–344. doi:10.1504/JDR.2015.074151

Watson, G. (2007). *Strategic benchmarking reloaded with six sigma: Improving your company's performance using global best practice*. Hoboken, NJ: John Wiley & Sons.

Womack, J., & Jones, D. (2003). *Lean thinking: Banish waste and create wealth in your corporation*. New York: The Free Press.

ADDITIONAL READING

Antony, J., Gijo, E., & Childe, S. (2012). Case study in six sigma methodology: Manufacturing quality improvement and guidance for managers. *Production Planning & Control: The Management of Operations*, 23(8), 624–640. doi:10.1080/09537287.2011.576404

Bisgaard, S., & De Mast, J. (2006). After six sigma, what's next? *Quality Progress*, 39(1), 30–36.

De Mast, J., & Lokkerbol, J. (2012). An analysis of the six sigma DMAIC method from the perspective of problem solving. *International Journal of Production Economics*, 139(2), 604–614. doi:10.1016/j.ijpe.2012.05.035

Hilton, R., & Sohal, A. (2012). A conceptual model for the successful deployment of lean six sigma. *International Journal of Quality & Reliability Management*, 29(1), 54–70. doi:10.1108/02656711211190873

Jacobs, B. W., Swink, M., & Linderman, K. (2015). Performance effects of early and late six sigma adoptions. *Journal of Operations Management*, 36, 244–257. doi:10.1016/j.jom.2015.01.002

Jayaram, J., Das, A., & Nicolae, M. (2010). Looking beyond the obvious: Unraveling the Toyota Production System. *International Journal of Production Economics*, 128(1), 280–291. doi:10.1016/j.ijpe.2010.07.024

Kwak, Y. H., & Anbari, F. T. (2006). Benefits, obstacles and future of six sigma approach. *Technovation*, 26(5), 708–715. doi:10.1016/j.technovation.2004.10.003

Montgomery, D. (2010). A modern framework for achieving enterprise excellence. *International Journal of Lean Six Sigma*, 1(1), 56–65. doi:10.1108/20401461011033167

Schroeder, R., Linderman, K., Liedtke, C., & Choo, A. (2008). Six sigma: Definition and underlying theory. *Journal of Operations Management*, 26(4), 536–554. doi:10.1016/j.jom.2007.06.007

Xie, M., Tan, K., & Goh, T. (2003). *Advanced QFD applications*. Milwaukee, WI: ASQ Quality Press.

KEY TERMS AND DEFINITIONS

Critical to Quality (CTQ) Characteristics:

Traits or factors that are critical to fulfillment of customer or other stakeholder needs. These are key levers in the delivery of value through process, product, system or enterprise designs or innovations.

DMADV: Define – Measure – Analyze – Design – Verify is an algorithm or approach associated with Six Sigma and Lean Six Sigma that is commonly used in process, product or system design applications.

DMAIC: Define – Measure – Analyze – Improve – Control is an algorithm or approach used in Six Sigma and Lean Six Sigma projects that are aimed at significant innovation in and improvement of existing processes, products or systems.

Kano Customer Needs Model: A model that differentiated customer and other stakeholder needs into three primary categories: those that dissatisfy customers, customer satisfiers, and those that excite customers. Specific knowledge of these needs is used to inform product, process, or system innovation and design.

Lean Six Sigma: Highly structured innovation, design, and lean enterprise strategies and methods for acquiring, assessing, and activating customer, competitor, and enterprise intelligence in order to deliver superior product, process, system, or enterprise performance that benefits all relevant stakeholders through best and next best practices and sources of sustainable competitive advantage.

Quality Function Deployment (QFD): A meticulously detailed method of mapping and subsequently executing enterprise capabilities and resources to customer and other stakeholder needs. QFD is often practiced as a four-stage process that carries and implements the voice of the customer from concept to market.

Value Stream/Value Chain: A value stream is a sequence of activities needed to design, produce and provide a specific service and along which information, material and value flow. A value chain is a set of linked activities that transform inputs into outputs that in turn add to at least one of the ecological, societal or economic bottom lines and help create competitive advantages. Linked to Six Sigma and Lean methodologies the goal is to create sustainable competitive advantages.

Motivational Factors of Telework

Arlene J. Nicholas

Salve Regina University, USA

INTRODUCTION

This chapter discusses telework as a motivational option for workers and employers. Telework can be performed as a supplemental or a full-time employment opportunity for local, regional or global work. Teleworkers must be self-starters, able to work with minimal supervision, and, in some cases, rely on technological communications for professional interactions. Teleworkers can enjoy flexibility for work and personal responsibilities (Greer & Payne, 2014) while organizations can gain more satisfied productive employees (Bloom, 2014; Caillier, 2014). An overview of telework's motivational benefits, incentives, and organizational examples for both employees and management are identified.

BACKGROUND

Telework is a growing method of employment with a variety of benefits. Telework enables work from anywhere, anytime through information commu-

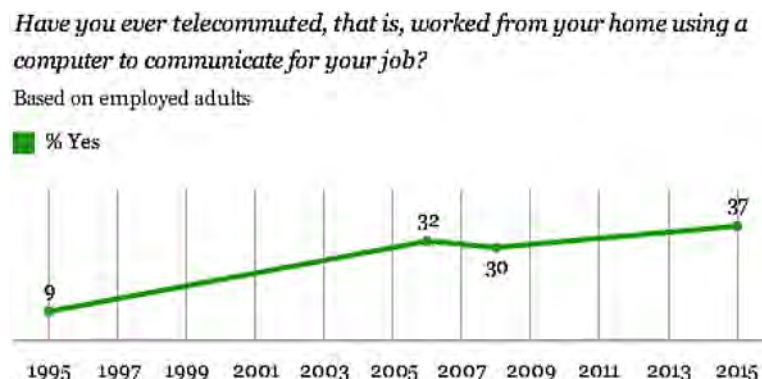
nication technologies (ICT) (Garett & Danziger, 2006). A recent Gallup poll of 1011 adults in the U.S. shows a steady increase in telework, with a short term dip caused by the 2008 economic crash (see Figure 1) (Jones, 2015).

Forbes analyzed 40,000 jobs posted on Flexjobs 2015 to create the Top 100 Companies for Remote Jobs list (See Appendix A). The 36% increase of listings from the previous year included jobs in IT, medical and health, sales, administration, customer service, education and training, and marketing (Shin, 2016). The variation of jobs permitting telework include full time, part time, contract and seasonal workers. Unexpected were some positions, such as Director of Community Advancement - full time, Physiotherapist - part time, Expert Wellness Coach - part time, and Vice President of Communications and Development - full time (FlexJobs, 2016). Other recently posted positions are listed in Table 1.

Many jobs can be a fit for telework which is considered one of those best practices that allows employees flexible options for improved work/life balance. For organizations, telework is a savvy

Figure 1. Increase of teleworking

Source: Gallup Education and Work poll August 5 - 9, 2015



DOI: 10.4018/978-1-5225-2255-3.ch064

Table 1. May 2016 Listings (partial list)

Position	Telecommute	Duties	Worker Location
Farm Food Safety Manager	Full Time	Develop criteria for assurance & assessment of farm practices to ensure food safety	Anywhere in US
Notary	Part Time	Reviewing, signing, & notarizing documents	Must work in City or State specified
Production Assistant Multicultural Marketing	Freelance for Contract	US Latin American culture & editing software knowledge	Must work in City or State specified
Assistant Architectural Historian	Seasonal	Supports projects through historical research	Must work in City or State specified
Managing Editor	Full Time	Manage city-level writers, perform researching, & edit articles & features	Anywhere in US

Source: FlexJobs, 2016

strategic management plan (Kowalski & Swanson, 2005). Telework brings increased benefits to both the employer and the employee (see Table 2).

TELEWORKERS

The number of teleworkers is growing in most sectors every year. A breakdown of organizations using telework was done by GlobalWorkplace-Analytics (2016) based on data from 2005 to 2014 of U.S. telework population. In those nine years, except for occasional small drops in government workers, such as 2013 which rebounded in 2014, all other organizations experienced growth. Table 3 depicts the figures from 2012 – 2014.

According to the Telework Advisory Group of World at Work, the representative teleworker is a college educated 40 year old male working from home. And although this credible not for

profit organization of 30,000 members in 100 countries, founded in 1995, reported that people who worked remotely one day a month declined, the number working more than one day a month increased (Telework 2011, 2011). Teleworkers often include knowledge workers, a term coined by management author Peter Drucker, who are occupied with analysis and manipulation of data as opposed to production (Knowledge worker, 2010). Today's technology has enabled knowledge workers to work remotely from anywhere in the world be it at home, an offsite office or telecenter, or even a coffee shop. Improved ICT, personal incentives and even employer encouragement will add to the number of teleworkers.

Legislation

In March of 2010, a House subcommittee approved legislation to promote telework in Federal

Table 2. Telework benefits

Organizational	Employee	Societal
Productivity	Savings	Work opportunities
Retention/attraction	Flexibility	disabled, homebound, distant
Virtual Workplace	Work/life Balance	Ecological – less car pollution
24 hour cycle	Some autonomy	Environmental
Continuity	Job satisfaction	less traffic congestion

(Caillier, 2014; GII, 2008; Rhodes, 2009; Telework=better, 2009; Ursery, 2003)

Table 3. Teleworkers by type of employer

Employer	2012	2013	2014
For Profit	2,501,855	2,689,963	2,860,517
Non Profit	314,984	355,327	355,327
Local Government	119,622	122,530	131,597
State Government	164,382	160,661	170,932
Federal Government	157,682	144,966	158,688
Total	3,268,525	3,480,447	3,677,061

Source: GlobalWorkplaceAnalytics

agencies. The Telework Enhancement Act of 2010 (<http://www.gpo.gov/fdsys/pkg/BILLS-111hr1722enr/pdf/BILLS-111hr1722enr.pdf>) mandated a government-wide telework policy. Agencies are required to allow eligible employees to work at least 20 hours a week remotely (Rosenberg, 2010).

The law has been successful. In 2009, only 10% of eligible federal workers telecommuted. In Sept 2011, almost 21% were regular teleworkers (Resneck, 2012). It increased in 2012 to 24% and again in 2013 to 27% (2014 Status of Telework, 2014). However, there are still many federal employees who are eligible to telework but are not participating. Some concerns are double-taxation costs for the worker in both their home state and work state (Allen, Golden & Shockley, 2015). The Multi-State Worker Tax Fairness Act of 2014 that would reduce tax burdens for employees and their payroll departments has been under ‘committee consideration’ since March of 2014 (H.R.4085, 2014). Passage of this bill could raise the numbers of federal and private sector teleworkers.

Many states have their own statutes or executive orders and some simply have policies to motivate telework options for state-agency employees. A growing number of states have tax incentives on percentages of teleworkers or conversion costs for agencies (Allen, Golden & Shockley, 2015).

Another useful part of the Telework Enhancement Act legislation is the requirement for agencies to train managers on properly supervising teleworkers (Rosenberg, 2010), this is something the corporate world must address also (“Managing

teleworkers ...”, 2009). Training managers how to supervise and guide teleworkers may be challenging given some resistance noted by earlier and recent studies of managers unwilling to give up direct control of the visible worker (Clear & Dickson, 2005; Daniels, Lamond, & Standon, 2001; Tuutti, 2012).

Motivation

Beyond legislation, adoption and implementation of telework programs can have a positive effect on job satisfaction (Bae & Kim, 2016). This satisfaction can be a very motivating factor in choosing a position with teleworking as pay and titles are not always the determining factors (Gose, 2013). There can be reduced turnover intention when telework is an option (Bae & Kim, 2016). This attrition factor combined with higher levels of performance, satisfaction and creativity (Vega, Anderson & Kaplan, 2015), should motivate managers to implement telework.

If managers deny teleworking requests, it could result in de-motivating workers. In a study of Department of Health and Human Services workers (98% of whom are eligible for telework), the employees who were denied the opportunity to telework reported lower motivation than the employees who did telework (Caillier, 2012). Interestingly, employees who teleworked frequently, more than two days a week, reported less motivation than employees who teleworked infrequently, up to two days a week. According to Golden (2006), this decline of satisfaction with

extensive telework is a result of isolation from co-workers and lack of face-to-face interaction with managers. Hybrid work arrangements of office and telework can offer both face-to-face and flexible options (Sewell & Taskin, 2015).

A positive aspect of being away from the workplace is avoiding office politics and gossip (Ellison, 2004; Fonner & Roloff, 2010; Kurland & Cooper, 2002). Physical distance can ensure contact, such as phone or e-mail, is focused on job related matters, not drama. Of course, avoiding this type of information sharing or wanting to shop talk is a personal characteristic of the employee.

As noted earlier, employees could be motivated by savings that can include gas, car wear and tear, clothing, child/elder or even pet care arrangements; flexibility for care arrangements, doctors' appointments, and other chores or business needs; autonomy to meet business/home needs; and contributing to green environmental effects (Caillier, 2012; Caillier, 2014; GII, 2008; Rhodes, 2009; Telework=better, 2009; Urserly, 2003).

Challenge to Management

Managers of teleworkers must motivate and inspire out of sight workers and provide authentic information sharing, from managers to workers and between work teams. An affinity distance problem can result when there are trust issues between team members who have never met one another (Reilly & Lojeski, 2009). Some non-teleworkers can have resentment of teleworkers and perceive doing some of the teleworkers' duties (Greer & Payne, 2014).

Managers too must overcome lack of trust for workers who are miles away (Brown, Smith, Arduengo & Tayolor, 2016; Leonard, 2011). When Marissa Meyer was brought in as CEO to revitalize Yahoo, she stopped the telecommuting program (Cohan, 2013; Gaudreau, 2013). CEO Hubert Joly of Best Buy halted the Results Only Work Environment (ROWE) program a week later, also as a revamping scheme. Both moves, and others, were attributed to a weak job market

that gave management tighter controls as these eliminations countered studies of increased productivity, retention and job satisfaction (Cohan, 2013; Gaudreau, 2013; Valcour, 2013).

In a study by Sewell and Taskin (2015) of 31 information technology and research and development teleworkers, 26 teleworked one day a week and 5 teleworked two days a week. The teleworkers perceived more intense supervision and direction, even of tasks that would have been under their authority if still in the office. Technology and customer service has to be combined with human resource management for a productive teleworking program (Blount, 2015).

Management has to have a balance of monitoring while encouraging open communications among teleworkers, site workers and management. This is definitely a challenge and relatively new to many managers. Just as there should be guidelines for teleworkers, more managers should get training. If little more than 17% of organizations with some kind of flexible work program provide training for managers of flexible workers (Sharp, 2013), there is considerable room for improvement.

Today's Workforce Considerations

Premiere Global Services Incorporate surveyed 3,000 knowledge workers in North America (NA), Europe, the Middle East and Africa (EMEA) and Asia Pacific Japan (APJ). Of those surveyed, 79% reported that they teleworked and 54% were without any policies. NA reported the widest use of telework followed by EMEA and APJ (2015 PGi Global, 2016).

Some attribute the rise to 37% teleworkers in the U.S. to the winter of 2014-2015 and record-breaking cold and snow for the Midwest and the Northeast (when snowstorms even had names like 'Juno' in January 2015). Businesses continued to do productive work through telework (Plumb, 2015).

Earlier reports had lower percentage of teleworkers at 9%, the same as Europe, compared to emerging markets: Asia-Pacific, 24%; Latin

America, 25%; Africa and the Middle East, 27% each. The highest percentage reported is India with 82% teleworking at least once a week, and 57% are frequent teleworkers, some of which is outsourced jobs (Reaney, 2012).

Some differences certainly relate to job opportunities and cultural norms. Studies of U.S. workers note the social isolation or relational impoverishment of teleworking has negative effects on job performance (Gajendran & Harrison, 2007; Golden, Veiga, & Dino, 2008). But the workforce of today is composed of many generations and what Baby Boomers (born 1946-1964) may perceive as detrimental to their careers, such as being away from the workplace, could appeal to Gen Xers (1965-1980) with families and to the Net Generation's (1981-2001) tech comfort.

For example, teleworkers have reported many benefits that could relate to a new generation of workers who are said to be "less at home with the real world than in the virtual world" (Eisner, 2005, p.2). Combining that technological penchant with the Net Generation's team-orientation (Bridges & Johnson, 2006) would seem to make the virtual team scenario of telework very compatible and its eco-friendly aspect would appeal to these socially conscious citizens (Eisner, 2005; Raines, 2002). Many Baby Boomers have been assessed as reluctant to use new technology which has not really been designed for or marketed to older users (Burdick, 2005). There may be need for additional training and support for those older workers.

SOLUTIONS AND RECOMMENDATIONS

For teleworkers who are beyond the reach of the office, if there is one, interaction with managers and team members can be critical. Managers need defined meeting times and agendas, and, in the cultural context, learn to communicate in the language of the workforce (Klein, 2008). Infrequent teleworkers, especially those who may be part of an alternate shared workspace arrangement, such

as in-office work on Monday/Wednesday/ Friday with the same office used on Tuesday/Thursday by another colleague, should also be connected by management arrangement meetings.

Peter Linkow, president of WFD Consulting, based in Newton, MA, and research leader of the Conference Board's Research Working Group on Managing a Distant Workforce gives 10 recommended guidelines for distant managers:

1. Recruit for character, build fundamentals then develop distance competencies.
2. Emphasize relationships as much as tasks.
3. Create intra-company collaboration, then assess it.
4. Build solid communication infrastructure.
5. Secure the upper management support that employees need to be successful.
6. Set clear goals and then empower distant employees to determine how to achieve them.
7. Focus managers on creating a work environment that enables employee achievement.
8. Concentrate on the job environment.
9. Approach cultural differences with an open mind.
10. Support the distant managers who report to you. (Klein, 2008)

Selection of the right workers for telework as well as setting detailed and specific goals is critical (Wilkie, 2015). Making employees successful will also depend on easy to use technology and readily available support. The older workforce is not as resistant to technology as it is disappointed with its user unfriendliness which limits their use of information communications technology (ICT) (de Koning & Gelderblom, 2006). The younger workforce will demand speed and mobile applications (Tapscott, 2009).

Training for teleworkers should include using technology to accommodate work and communications, creating separations between work and non-work roles and "planning appropriate tasks to maximize daily productivity" (Greer & Payne, 2014, p. 107). Some face-to-face interventions

are recommended, but when that is not always possible, video conference, Skype, Facebook, personal blogs and even Second Life virtual meetings have been used to give some connection between teams and leaders (Klein, 2008). Reiterating the mission/vision/goals can help reduce operational distances amongst teams' interpretations of the organizational practices (Reilly & Lojeski, 2009). Results based managers will get the best out of their employees whether they are in the office or miles away (Leonard, 2011).

FUTURE RESEARCH DIRECTIONS

The value of teleworking programs to organizations and workers must continually be assessed. Before and after comparisons of the effects of telework programs should be analyzed. Are there differences in desirability according to gender, age, distance from work, family composition, culture? Organizations will want to know retention, recruitment, job satisfaction and productivity based on telework options. Are there technology pitfalls to be avoided? Concern for secure data transmissions, customer/client/patient confidentiality, bandwidth capabilities, support and back-up plans could determine what organizations will envision as a cost/benefit investment in telework.

Further studies on the reported curvilinear relationship of telework time for increased satisfaction with limited hours/days and then a decrease of satisfaction with too many hours/days (Golden & Viega, 2005) should be conducted. This could be a critical phenomenon to understand for organizations that can offer varied schedules of telework.

CONCLUSION

Telework already is an indispensable option of working for the myriad benefits reviewed. These include continuity of work functions, attracting and recruiting employees, office space savings, and global extensions for organizations; flexibility,

work/life balance, clothes/transportation savings, family care ability for employees; reduced traffic, less emissions, cultural and disability options for society. The difficulties for management of selection, training, security, procedures and connectivity to workers must be met as well as keeping up with the quick pace of changing technology. Clear and open communications can motivate and maintain a thriving workforce from anywhere, at anytime. The organization, managers and employees need to work in a trusting cohesive method to develop a productive teleworking experience (Brown, Smith, Arduenga, & Taylor, 2016).

REFERENCES

- Allen, T. D., Golden, T. D., & Shockley, K. M. (2015). How effective is telecommuting? Assessing the status of our scientific findings. *Psychological Science in the Public Interest*, 16(2), 40–68. doi:10.1177/1529100615593273 PMID:26403188
- Bae, K. B., & Kim, D. (2016). The impact of decoupling of telework on job satisfaction in U.S. federal agencies: Does gender matter? *American Review of Public Administration*, 46(3), 356–371. doi:10.1177/0275074016637183
- Bloom, N. (2014). To raise productivity, let more employees work from home. *Harvard Business Review*. Retrieved May, 2016 <https://hbr.org/2014/01/to-raise-productivity-let-more-employees-work-from-home>
- Blount, Y. (2015). Telework: Not business as usual. In *Business Technologies in Contemporary Organizations: Adoption, Assimilation, and Institutionalization* (pp. 76–95). IGI Global; doi:10.4018/978-1-4666-6623-8.ch004
- Brown, C., Smith, P., Arduenga, N., & Taylor, M. (2016). Trusting telework in the federal government. *Qualitative Report*, 21(1), 87–101. Retrieved from <http://nsuworks.nova.edu/tqr/vol21/iss1/8/>

- Burdick, D. C. (2005). Technology and aging: Not an oxymoron. *AGHEExchange*, 28(4), 1-4. Retrieved December, 2012, from <http://www.capitalhealth.ca/NR/rdonlyres/eynmm2f7g-3bdeiafxgznieliralz6neejay7hgmzixevjwtd-wduaw2m5f15lbspexfaz24fil6zfdj5a5jjxqlcza/Dr.+David+Burdick+-+Articles+2.pdf>
- Caillier, J. G. (2012). Impact of telework on work motivation in a U.S. federal government agency. *American Review of Public Administration*, 42(4), 461–480. doi:10.1177/0275074011409394
- Caillier, J. G. (2014). Do role clarity and job satisfaction mediate the relationship between telework and work effort? *International Journal of Public Administration*, 37(4), 193–201. doi:10.1080/01900692.2013.798813
- Clear, F., & Dickson, K. (2005). Teleworking practice in small and medium-sized firms: Management style and worker autonomy. *New Technology, Work and Employment*, 20(3), 218–233. doi:10.1111/j.1468-005X.2005.00155.x
- Cooper, C. D., & Kurland, N. B. (2002). Telecommuting, professional isolation, and employee development in public and private organizations. *Journal of Organizational Behavior*, 23(4), 511–532. doi:10.1002/job.145
- Daniels, K., Lamond, D., & Standen, P. (2001). Teleworking: Frameworks for Organizational Research. *Journal of Management Studies*, 38(8), 1151–1186. doi:10.1111/1467-6486.00276
- de Koning, J., & Gelderblom, A. (2006). ICT and older workers: No unwrinkled relationship. *International Journal of Manpower*, 27(5), 467–490. doi:10.1108/01437720610683967
- Eisner, S. P. (2005). Managing Generation Y. *S. A. M. Advanced Management Journal*, 70(4), 4–15.
- Ellison, N. B. (2004). *Telework and social change: How technology is reshaping the boundaries between home and work*. Westport, CT: Praeger.
- FlexJobs. (2016). *100 of the most surprising flexible jobs, past and present*. Retrieved from <https://www.flexjobs.com/jobs/surprising-jobs>
- Fonner, K. L., & Roloff, M. E. (2010). Why teleworkers are more satisfied with their Jobs than are office-based workers: When less contact is beneficial. *Journal of Applied Communication Research*, 38(4), 336–361. doi:10.1080/00909882.2010.513998
- Gajendran, R. S., & Harrison, D. A. (2007). The good, the bad, and the unknown about telecommuting: Meta-analysis of psychological mediators and individual consequences. *The Journal of Applied Psychology*, 92(6), 1524–1521. doi:10.1037/0021-9010.92.6.1524 PMID:18020794
- GII. (2008). *Telecommuter Update, 2008: Employer and Employee Profiles, Changing Work Dynamics, and Technology Use Product Type: Market Research Report*. Retrieved from <http://www.giexpress.com/products/id71727.htm>
- Global Workplace Analytics. (2016). *Latest telecommuting statistics*. Retrieved May, 2106 <http://globalworkplaceanalytics.com/telecommuting-statistics>
- Golden, T. D. (2006). The role of relationships in understanding telecommuter satisfaction. *Journal of Organizational Behavior*, 27(3), 319–334. doi:10.1002/job.369
- Golden, T. D., Veiga, J. F., & Dino, R. N. (2008). The impact of professional isolation on teleworker job performance and turnover intentions: Does time spent teleworking, interacting face-to-face, or having access to communication-enhancing technology matter? *The Journal of Applied Psychology*, 93(6), 1412–1421. doi:10.1037/a0012722 PMID:19025257
- Gose, B. (July, 2013). Flexible jobs mean happy instructors. *The Chronicle of Higher Education*. Retrieved May, 2016. <http://chronicle.com/article/Flexible-Jobs-Mean-Happy/140411/>

- Greer, T. W., & Payne, S. C. (2014). Overcoming telework challenges: Outcomes of successful telework strategies. *The Psychologist Manager Journal*, 17(2), 87–111. doi:10.1037/mgr0000014
- H.R. 4085. (2014). *Multi-state worker tax fairness act*. Retrieved May 2016. <https://www.congress.gov/bill/113th-congress/house-bill/4085/all-actions?overview=closed>
- Jones, J. M. (2015). In U.S., telecommuting for work climbs to 37%. *GALLUP Workplace*. Retrieved May, 2016 <http://www.gallup.com/poll/184649/telecommuting-work-climbs.aspx>
- Klein, P. A. (2008). It's 9:05: Do you know where your workforce is? *The Conference Board Record*, 46(2), 55–59.
- Knowledge worker definition. (2010). *Business dictionary*. Retrieved December, 2010. <http://www.businessdictionary.com/definition/knowledge-worker.html>
- Kowalski, K. B., & Swanson, J. A. (2005). Critical success factors in developing teleworking programs, *Benchmarking. International Journal (Toronto, Ont.)*, 12(3), 236–249.
- Leonard, B. (2011). Telework increasing slowly worldwide. *SHRM*. Retrieved December, 2012. <http://www.shrm.org/hrdisciplines/technology/Articles/Pages/TeleworkWorldwide.aspx>
- Managing teleworkers requires topnotch performance management skills. (n.d.). U.S. Office of Personnel Management. Retrieved October, 2012. <http://www.opm.gov/perform/articles/2001/win01-1.asp>
2015. PGI global telework survey: Trends around the world shaping the future of work. (n.d.). Retrieved May, 2016. <http://go.pgi.com/gen-genspec-15telesur-SC1129>
- Plumb, E. (2015). Working through the snow: This record-breaking winter shows the benefits of telework and flexible schedules. *U.S. News & World Report*. Retrieved May, 2016. <http://www.usnews.com/opinion/economic-intelligence/2015/02/26/2015-record-breaking-winter-prove-benefits-of-teleworking>
- Raines, C. (2002). *Managing millennials*. Retrieved December, 2012. <http://www.generation-satwork.com/articles/millennials.htm>
- Reaney, P. (2012). *About one in five workers worldwide telecommute: poll*. Reuters. Retrieved from <http://www.reuters.com/article/2012/01/25/us-telecommuting-idUSTRE80N1IL20120125>
- Reilly, R. R., & Lojeski, K. S. (2009). Leading the dispersed workforce. *Mechanical Engineering (New York, N.Y.)*, 131(11), 30–35.
- Rhodes, M. (2009). *Telework revs up as more employers offer work flexibility*. The Telework Advisory of World at Work. Retrieved December, 2010. <http://www.workingfromanywhere.org/news/pr021609.html>
- Rosenberg, A. (2010). Telework and benefits bills take one step forward. *Government Executive*. Retrieved December, 2010. <http://www.govexec.com/dailyfed/0310/032410ar1.htm>
- Sewell, G., & Taskin, L. (2015). Out of sight, out of mind in a new world of work? Autonomy, control, and spatiotemporal scaling in telework. *Organization Studies*, 36(11), 1507–1529. doi:10.1177/0170840615593587
- Sharp, M. (2013). New survey shows telework offered by 88% of organizations. *WorldatWork*. Retrieved May, 2016. <https://www.worldatwork.org/waw/adimLink?id=73920>
- Shin, L. (2016). Work from home in 2016: The top companies for remote jobs. *Forbes*. Retrieved May, 2016 <http://www.forbes.com/sites/laurashin/2016/01/27/work-from-home-in-2016-the-top-100-companies-for-remote-jobs/#557836fa36f0>

2014. Status of Telework in the Federal Government Report to Congress. (2014). United States Office of Personnel Management. Retrieved May, 2016 <https://www.telework.gov/reports-studies/reports-to-congress/2014-report-to-congress.pdf>

Tapscott, D. (2009). *Grown up digital: How the net generation is changing your world*. New York: McGraw-Hill.

Telework = better health. (November, 2009). *The Teleworker*. Retrieved January, 2011. <http://www.teleworkexchange.com/teleworker-11-09b.asp>

Telework. (2011). *A special report from WorldatWork*. Retrieved October, 2012. <http://www.workingfromanywhere.org/>

Tuutti, C. (2012). Why federal managers resist telework. *FCW: The Business of Federal Technology*. Retrieved from <http://fcw.com/articles/2012/07/11/reasons-federal-managers-resist-telework.aspx>

Ursery, S. (2003, May). Austin fights air pollution with telework program. *American City and County*, 118(5), 12–13.

Valcour, M. (March 2013). The end of “results only” at Best Buy is bad news. *Harvard Business Review*. Retrieved May, 2016. <https://hbr.org/2013/03/goodbye-to-flexible-work-at-be>

Vega, R. P., Anderson, A. J., & Kaplan, S. A. (2015). A within-person examination of the effects of telework. *Journal of Business and Psychology*, 30(2), 313–323. doi:10.1007/s10869-014-9359-4

Wilkie, D. (2015). Has the telecommuting bubble burst? *Society of Human Resource Management*, 60(5). Retrieved May, 2016, from <https://www.shrm.org/publications/hrmagazine/editorialcontent/2015/0615/pages/0615-telecommuting.aspx>

ADDITIONAL READING

Baiamonte, L. W. (2012). Telecommuting: Do you agree with it? Does your company agree with it? Yes. *Journal of Property Management*, 77(3), 9.

Colbert, C. J. (2012). *Telework as part of a business continuity strategy: A path toward organizational resilience*. ProQuest UMI Dissertation.

Crook, S. K., Jaffe, J., Boggs, R., & Drake, S. D. (2011). Worldwide mobile worker population 2011-2015 forecast. <http://www.idc.com/getdoc.jsp?containerId=232073>

Elsbach, K., & Cable, D. (2012). Why showing your face at work matters. *MIT Sloan Management Review*, 53(4), 10–12.

Garrett, R. K., & Danziger, J. N. (2006). Which telework? Defining and testing a taxonomy of technology-mediated work at a distance. *Social Science Computer Review*, 25(1), 27–47. doi:10.1177/0894439306293819

Glenn Dutcher, E. E. (2012). The effects of telecommuting on productivity: An experimental examination. The role of dull and creative tasks. *Journal of Economic Behavior & Organization*, 84(1), 355–363. doi:10.1016/j.jebo.2012.04.009

Global Workplace Analytics and the Telework Research Network. (2012). Whitepapers. Retrieved December, 2012. <http://www.teleworkresearch-network.com/whitepapers>

Golden, T. (2012). Altering the effects of work and family conflict on exhaustion: Telework during traditional and nontraditional work hours. *Journal of Business and Psychology*, 27(3), 255–269. doi:10.1007/s10869-011-9247-0

- Lister, K., & Hamish, T. (2011). The state of telework in the U.S.: How individuals, businesses and government benefit. *Telework Research Network*. Retrieved December, 2012. <http://www.workshifting.com/downloads/downloads/Telework-Trends-US.pdf>
- Lojeski, K. S., & Reilly, R. R. (2008). *Uniting the Virtual Workforce: Transforming Leadership and Innovation in the Globally Integrated*. New York: John Wiley & Sons.
- Madsen, S. R. (2011). The benefits, challenges, and implications of teleworking: A literature review. *Culture & Religion Review Journal*, 2011(1), 148-158.
- Martínez-Sánchez, A., Pérez-Pérez, M., de-Luis-Carnicer, P., & Vela-Jiménez, M. J. (2008). Telework, human resource flexibility and firm performance. *New Technology, Work and Employment*, 22(3), 207-223.
- McKendrick, J. (2011). Telecommuting surge: 5 million new home-based offices by 2015. SmartPlanet, CBS Interactive. <http://www.smartplanet.com/blog/business-brains/telecommuting-surge-5-million-new-home-based-offices-by-2015/14431>
- Morganson, V. J., Major, D. A., Oborn, K. L., Verive, J. M., & Heelan, M. P. (2010). Comparing telework locations and traditional work arrangements: Differences in work-life balance support, job satisfaction, and inclusion. *Journal of Managerial Psychology*, 25(6), 578-595. doi:10.1108/02683941011056941
- Next generation telework: A literature review. (2011). Australia: Deloitte Access Economics Department of Broadband, Communications and the Digital Economy. Retrieved December, 2012 http://www.nbn.gov.au/files/2012/02/Next_Generation_Telework-A_Literature_Review-July_2011.pdf
- Nicholas, A., & Guzman, I. (2009). Is teleworking for millennials? *Association for Computing Machinery SIGMIS Computer Personnel Research Proceedings*: special interest group on management information system's 47th annual conference.197-208. Presented at University of Limerick, Ireland.
- Nilles, J. (1996). *Managing telework: Strategies for managing the virtual workforce*. New York: John Wiley & Sons.
- Noonan, M. C., & Glass, J. L. (2012). The hard truth about telecommuting. *Monthly Labor Review*, 38-45. <http://www.bls.gov/opub/mlr/2012/06/art-3full.pdf> Retrieved December, 2012
- Nurul Habib, K. M., Sasic, A., & Zaman, H. (2012). Investigating telecommuting considerations in the context of commuting mode choice. *International Journal of Sustainable Transportation*, 6(6), 362-383. doi:10.1080/15568318.2011.621014
- Offstein, E. H., & Morowick, J. M. (2009). *Making telework work: Leading people and leveraging technology for high-impact results*. Boston: Davies-Black.
- Pyöriä, P. (2011). Managing telework: Risks, fears and rules. *Management Research Review*, 34(4), 386-399. doi:10.1108/01409171111117843
- Rietveld, P. (2011). Telework and the transition to lower energy use in transport: On the relevance of rebound effects. *Environmental Innovation and Societal Transitions*, 1(1), 146-151. doi:10.1016/j.eist.2011.03.002
- Rodensky, R., Rybeck, J., Johnson, H., & Rollins, M. (2010). Secrets to a robust telework program: The 'STIR' model. *SHRM*, Retrieved December, 2010. <http://www.shrm.org/hrdisciplines/benefits/Articles/Pages/TeleworkModel.aspx>
- Rosenberg, A. (2009). OPM announces new governmentwide telework policy. *Government Executive*. Retrieved December, 2010. http://www.govexec.com/story_page.cfm?filepath=/dailyfed/0409/042909ar1.htm

Scott, D. M., Dam, I., Páez, A., & Wilton, R. D. (2012). Investigating the effects of social influence on the choice to telework. *Environment & Planning A*, 44(5), 1016–1031. doi:10.1068/a43223

Shah, M. (2013). *Mobile working: Technologies and business strategies*. New York: Routledge.

Star Workforce Solutions. (2009). Trends in telecommuting: Industry expansion and solutions for your business. Retrieved October, 2012. <http://starworkforce.com/download/Telecommuting%20Trends.pdf>

Taylor, C. (2012). Ethical dilemmas of teleworkers. *The Industrialist*. Retrieved December, 2012. <http://theindustrialists.me/2012/03/04/ethical-dilemmas-of-the-teleworker/>

TELCOA. The telework coalition. Retrieved December, 2012. <http://www.telcoa.org/>

Telework.gov Retrieved December, 2012. <http://www.telework.gov/>

Troup, C., & Rose, J. (2012). Working from home: Do formal or informal telework arrangements provide better work–family outcomes? *Community Work & Family*, 15(4), 471–486. doi:10.1080/13668803.2012.724220

Wheatley, D. (2012). Work-life balance, travel-to-work, and the dual career household. *Personnel Review*, 41(6), 813–831. doi:10.1108/00483481211263764

Zhang, J. (2016). The dark side of virtual office and job satisfaction. *International Journal of Business and Management*, 11(2), 40–46. doi:10.5539/ijbm.v11n2p40

KEY TERMS AND DEFINITIONS

Affinity Distance: Emotional separation between virtual team members who have no personal relationship.

ICT: Information communications technology.

Knowledge Worker: Researchers, planners, analysts and/or developers who acquire, manipulate, and analyze information.

Operational Distance: Psychological gaps regarding workplace issues.

Physical Distance: Variances in space, time and environment.

Telework: Work done remotely from the office using some form of telecommunications; also referred to as telecommuting.

Virtual Distance: Working and communicating mainly, or exclusively, through technology.

APPENDIX

B

Forbes Top 100 Companies for Remote Jobs 2016

1. LiveOps
2. TeleTech
3. Amazon
4. Sutherland Global Services
5. UnitedHealth Group
6. Dell
7. IBM
8. U.S. Department of Agriculture
9. Working Solutions
10. Humana
11. Aetna
12. Intuit
13. Kaplan
14. Kelly Services
15. Cactus Communications
16. Westat
17. Salesforce
18. PAREXEL
19. CyberCoders
20. American Express
21. VMware
22. SAP
23. Xerox
24. First Data
25. US-Reports
26. Oracle
27. CACI International
28. A Place for Mom
29. Anthem, Inc.
30. Dell SecureWorks
31. World Travel Holdings
32. ADP
33. Aon
34. University of Maryland University College
35. Allergan Inc
36. K12
37. U.S. Department of Transportation
38. CSI Companies
39. Robert Half
40. Nielsen
41. Red Hat

Motivational Factors of Telework

42. Adobe Systems
43. Overland Solutions, Inc.
44. BCD Travel
45. Connections Education
46. Deloitte
47. Apple
48. McKesson Corporation
49. Thermo Fisher Scientific
50. Precyse
51. Haynes & Company
52. Pharmaceutical Product Development Inc.
53. IT Pros Philadelphia
54. Cigna
55. Houghton Mifflin Harcourt
56. Sungard Availability Services
57. Infor
58. Sodexo
59. About.com
60. Altegra Health
61. GE – General Electric
62. Western Governors University
63. Grand Canyon University
64. Walden University
65. Vivint
66. BroadSpire
67. Covance
68. Ellucian
69. HD Supply
70. Perficient Inc.
71. Teradata
72. Wells Fargo
73. Symantec Corporation
74. Real Staffing
75. Science Applications International Corporation – SAIC
76. AmerisourceBergen Corporation
77. Appen
78. Hartford Financial Services Group
79. RetailData
80. SYKES
81. SRA International
82. Citizens Financial Group
83. CVS Health
84. Healthfirst
85. American Heart Association
86. BMC Software

87. hibu
88. inVentiv Health
89. Rosetta Stone
90. Erie Insurance Group
91. Worldpay
92. CleverTech
93. Achieve Test Prep
94. Deluxe
95. DataStax
96. CDK Global
97. Teleflex
98. Aquent
99. Parallon
100. U.S. Department of the Interior

Organizational Transparency

Gustavo de Oliveira Almeida

Federal University of the State of Rio de Janeiro, Brazil

Claudia Cappelli

Federal University of the State of Rio de Janeiro, Brazil

Cristiano Maciel

Federal University of Mato Grosso, Brazil

INTRODUCTION

Transparency is one of the most frequent words in the contemporary public discourse, due to the growing demand for organizational transparency from the many stakeholders, which include shareholders, national and international watchdog organizations, the mass media and influential bloggers (Wehmeier & Razz, 2012). For instance, since shareholders invest money in corporations, they require financial disclosure, the exposure of future strategies and the revealing of the corporations' decision-making processes. Also, factors such the crisis in the 'new markets' and financial crisis have increased international and national watchdog organizations such as Transparency International to continue to raise the issue of corporate transparency (Wehmeier & Razz, 2012).

Although the notion is frequently part of the daily discourse, the concept of transparency is rarely defined – neither in the news nor in the companies – beyond common-sense understandings as 'openness', 'insight' or 'clarity' (Christensen & Cornelissen, 2015).

According to Hood (2006 apud Wehmeier & Razz, 2012) the roots of transparency were present in ancient China and Greece. The Online Etymological Dictionary (2007), points out that the term transparency is derived from the word "transparentem", meaning "show light through", which in figurative use became "easily seen through", its figurative meaning started being used

somewhere between the XV and XVI century. The word continued to be used throughout and became part of many Latin derived languages such as the Italian, Spanish, Portuguese, French and English (Vaccaro & Madsen, 2009).

Transparency is usually presented as an unquestionably positive concept for the public, governments, and firms, found in almost every code of conduct, held as a foundational principle and an objective of most governmental and non-governmental organizations. However, fewer authors dedicate their research to present the issues or negative consequences of transparency (Fenster, 2006). There is an established view that transparent companies, governments, institutions and processes are essential to achieve corporate social responsibility, social justice, environmental security, true democracy and wellbeing (Menéndez-Viso, 2009). In the words of Christensen & Cornelissen (2015), this veneration of transparency as a value is the "myth of transparency". More than that, the contemporary organizations are in the business of transparency, as the result of legal, social or public pressure, or even as a self-goal to be transparent. Nevertheless, although the transparency is a growing concern, it does not mean that they are open about everything they do (Christensen & Cornelissen, 2015)

Thus, in a context dominated by frequent episodes of corporate wrongdoing, transparency is often presented as a remedy for the trust issues between the organization and its stakeholders.

Some studies have consistently pointed out the role of transparency in creating, maintaining, or repairing trust, between organizations and stakeholders (Schnackenberg, & Tomlinson, 2014). Consequently, many companies have embraced the concept of transparency believing it would bring increased levels of trust among the public, more specifically shareholders or investors, and other actors such as consumers, government and regulators (Williams, 2005). As a result, in a search for a more precise definition of transparency, Schnackenberg & Tomlinson (2014) nominated three factors, which the authors define as theoretically viable and managerially relevant: disclosure, clarity, and accuracy. The first factor disclosure increases as stakeholders perceive information as more relevant and timely; while clarity is related to the stakeholders' perception of the information as understandable; and finally the factor accuracy is defined as a perception by the stakeholders that the information is reliable. Each of these dimensions contributes in a singular way to the transparency by increasing stakeholder confidence in the quality of information received.

Rosendorff and Vreelandde (2006) define transparency as the dissemination of regular and accurate information. According to Leite et. al. (2010), there are three different levels of transparency: social transparency, target transparency and organizational transparency. According to Fung (2007), social transparency allows citizens to be more informed and encourages the disclosure of information as a regulation mechanism of centers of authority, while target transparency aims to reduce specific risks or performance problems through selective disclosure by corporations and other organizations. The concept of organizational transparency is defined as the disclosure of organizational information between an organization and its stakeholders, allowing to the society to verify whether the organization's activities are consistent with regard to the society's interests (Cappelli, 2009).

Describing in more detail the concept of organizational transparency, it is composed of five

dimensions: access, usability, informativeness, understandability and auditability. All of them can be applied for information transparency and business processes transparency (Cappelli, 2009). The information transparency is the feature related to the information of interest, and its ease of access, ease of use, quality of content, understanding and auditing. In turn, process transparency is the feature that enables the citizen access, easily use, understand and audit the processes dealing with information of interest. Process transparency requires that the transformation steps of the process be transparent, that is, it should be possible to understand its enactment (Cappelli, 2009).

This chapter has as its main objective to present and detail the concept of organizational transparency, and its dimensions, according to a review of literature. The chapter also aims to contrast the importance and implications of transparency for the society and organizations and to present research agenda in the topic.

BACKGROUND

Although transparency is not a recent concept, it remained a secondary concept until the end of 20th century, and an increased interest in transparency has emerged in organizational research only in the past two decades, mostly due to the corporate scandals in that period (Schnackenberg & Tomlinson, 2014). In addition, diverse areas of research have studied different aspects of transparency.

Generally, information systems researchers investigated the role of transparency in the relation of business to consumer relationships and digital markets. Meanwhile, organizational behavior researchers studied transparency in the context of organizational trust development, organizational identity, perceptions of leadership, and organizational culture, while researchers of finance and accounting have examined transparency in the context of financial markets, corporate disclosures, and monetary policy decision making, among other areas (Schnackenberg & Tomlinson, 2014).

Moreover, various social actors and institutions defend the value and significance of transparency without agreeing on what it means and encompass. Even though not precisely defined, transparency is treated as a way to improve trust in organizations and to legitimate the organization's acts without any form of questioning (Christensen & Cornelissen, 2015).

Henriques (2007) examines different constituents of transparency as a concept and frames them in the context of organizations, claiming that transparency will be essential for successful organizations. Therefore, transparency has obtained a status in contemporary society much like dogma or religious principles, or "myths" (Christensen & Cornelissen, 2015). Myths are not simply false understandings or accounts; instead, they are descriptions of the world presented in a narrative form with the objective of producing meaning and direction for a community and its members (Midgley, 2004).

Holzner and Holzner (2006 *apud* Cappelli, 2009) present social and historical perspectives to understand the forerunners to open government movement, in which transparency is key towards more open and democratic societies. However, a central issue facing any implementation of Information and Communication Technologies (ICT) is their relationship to individual and shared values and conceptions of the good, making it necessary determine what is acceptable and unacceptable regarding the uses of technology. In this discussion is important to understand the reasons regarding the creation of transparency and the preservation and promotion of the interests of all of organizational stakeholders. It is imperative to note that corporations are getting increasingly dependent on ICT, due to the fact that organizational process tend to be heavily related to collection, organization, storage, and communication of information. Nevertheless, although the new ICT technologies enable transparency it does not drive the implementation of transparency.

Schnackenberg & Tomlinson (2014) compiled a series of four common features of the different

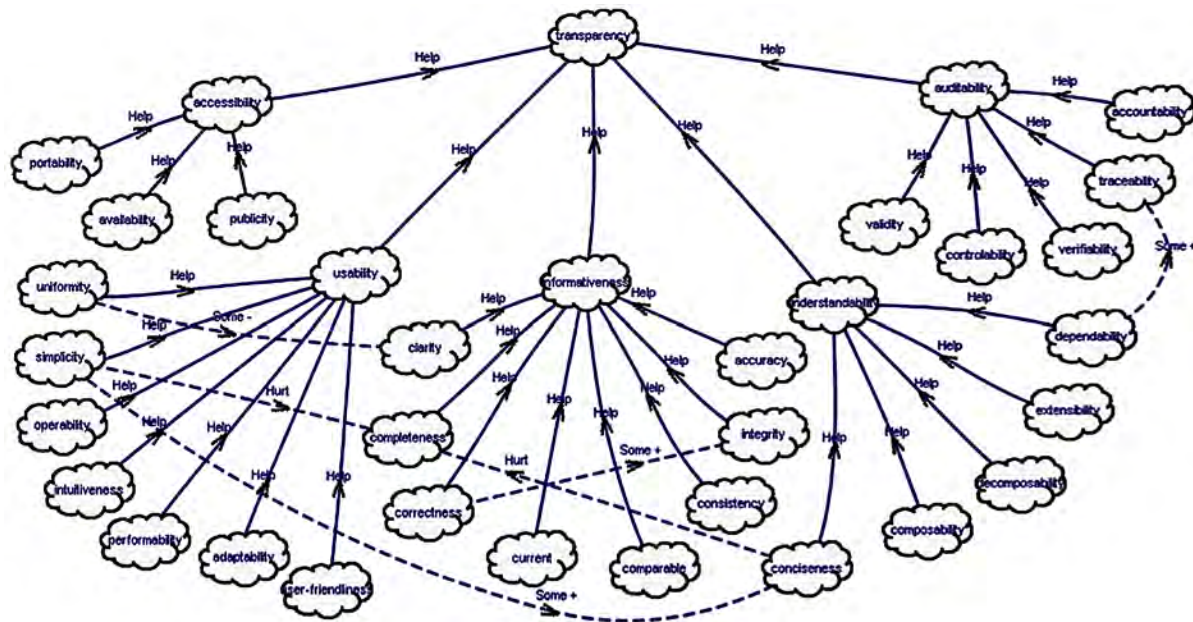
definitions of transparency. The first feature is that transparency is about information. Most research present transparency as a critical element of knowledge sharing, and highlight that the augment of transparency is connected to an increase in awareness, coherence, and comprehensibility to information exchanged between two parties. Other finding is that most definitions of transparency share the notion that transparency relies on intentionally shared information, performed in a systematic way. The third common feature suggests that transparency is a perception of received information. Finally, Schnackenberg & Tomlinson (2014) have found that transparency perceptions vary according to the perceived quality of information. This fourth feature was associated with the importance of information quality, which was present in an obvious or implied manner in almost all of the reviewed studies.

Schnackenberg & Tomlinson (2014) declare that there is a lack of theoretically grounded consensus on the transparency construct, what they call a "patchwork" of various operationalization from many areas of academic inquiry. The mentioned authors also state that the present literature on transparency does not suggest a definite solution on how the construct should be conceptualized.

Similarly, the current empirical studies on transparency have suffered from a great deal of conceptual variation, also disagreeing on what is meant by information quality. Thus, the operationalization of the transparency construct vary significantly and include concepts such amplified disclosure of information, greater openness and accuracy of information, higher visibility and accessibility of information, increased clarity and understandability of information, reduction on the concealment of information, and better timing of information disclosure. Thus, although there is a great deal of research an interest on transparency, a consensus has not been reached concerning which factors discriminate between high and low quality information. Still according to Schnackenberg & Tomlinson (2014) this lack of convergence of a single explanation of the aspects of information

Figure 1. Transparency SIG

Source: Cappelli, 2009



quality more relevant to transparency has hindered researchers from advancing a systematic theory of antecedents and consequences of transparency.

According to Cappelli (2009), transparency is a concept related to information disclosure, employed in different settings, mostly related to the empowering of citizens with regard to their rights. The issues of transparency are directly linked to information processing or information technology. Cappelli (2009) mapped these contexts into an “SIG” chart (Figure 1) to characterize the elements that contribute and are necessary to reach transparency:

Although most of the literature focuses on information transparency, Weber (2008) discusses the dimension of process transparency. A further analysis of the transparency concept indicates that the processes that produce information should themselves be transparent. According to Cappelli et. al, (2007), the transparency of the organizational process is related to the existence of organizational policies that aim to present to the stakeholders information about the organiza-

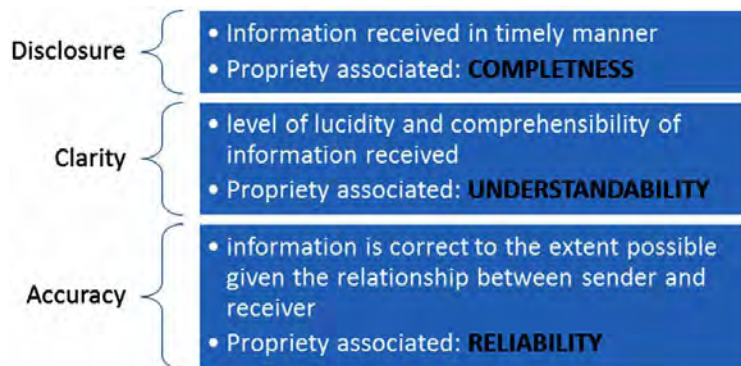
tion according to general aspects of access, use, presentation, understandability and auditability.

According to a literature review performed by Schnackenberg & Tomlinson (2014), researchers have conceptualized transparency in three primary ways: disclosure, clarity, and accuracy. Their findings have found that disclosure is usually defined as the perception that relevant information is received in a timely manner, in this regard, many studies present the disclosure as a central dimension of transparency.

The concept of disclosure implies that information must be openly shared for it to be considered transparent. On the other hand, clarity is defined as being the perceived level of lucidity and comprehensibility of information received from a sender. It is important to note that disclosed information consisting of industry jargon, unknown foreign languages, and complicated mathematical algorithms are not transparent even if it is thoroughly disclosed. That is, the transparency is not present if data is available but not understandable to the receiver. Finally, the concept of accuracy is

Figure 2. Three proprieties of transparency

Source: Schnackenberg & Tomlinson, 2014



related to the perception that information is correct to the extent possible given the relationship between the parts. In addition, information cannot be transparent if it is purposefully biased or false. However, accuracy does not mean that information must be completely correct for it to be considered transparent (Schnackenberg & Tomlinson, 2014)

Instead, accuracy suggests that material claims should reflect precise qualifications about their expected validity for information to be deemed transparent. Accuracy has some differences from disclosure and clarity since it deals with information reliability rather than completeness or understandability. Figure 2 has the three components of transparency along with their brief definition:

Information Systems and Organizational Transparency

The emergency of new ICTs is an important force that is driving organizations to higher degrees of transparency. The new technology can transform opaque organizations into transparent ones, allowing the stakeholders of institutions to have full accessibility of information regarding issues that affect their interests (Vaccaro & Madsen, 2009). More than that, transparency is a new ethical virtue of organizations, which is shaping and revolutionizing business practice, turning into a required feature for gaining and maintaining customer trust and to establish collaborative relationships with

all stakeholders, and it is heavily based on ICT, which has progressively increased role.

Consequently, technology is a key factor to promote transparency, but organizations may only implement ICTs that provide competitive advantage, *in lieu* of systems and processes created to promote transparency, unless the organization notices advantages on using IT differently than in their core business activities. In addition, while ICTs do allow higher access corporate information than ever before, they can also be employed to limit hierarchy and control of the information (Elia, 2009). Thus, to be transparent an organization needs to produce information, enabled by processes based on information technologies (Menéndez-Viso, 2009). According to the author, this portrays some issues with the transparency:

- It needs an active production of information opposed to a passive disclosure of data. The data has to be produced and made transparent (by having the proprieties of transparency: complete, understandable and reliable)
- The transparency of an organization is related to the required information as demanded by the various stakeholders
- The level of transparency can be measured with instruments analogous to the types used in quality control

- It is not free, it has some associated costs related to a organizational structure/ systems
- There is an economy of scale on the “transparency”, that is, bigger organizations or governments are abler to produce and publicize their information

According to Bertot et al. (2010), transparency and the right to access government information is now internationally regarded as essential to democratic participation, trust in government, prevention of corruption, informed decision-making, accuracy of government information, and provision of information to the public, companies, and journalists, among other essential functions in society.

Transparency in governmental relations is so important that it was the topic of the first memorandum signed by Barack Obama on his first term (Obama, 2009). According to the memorandum transparency “promotes accountability and provides information for citizens”. In addition, the document points out that the lack of transparency makes impossible to monitor the acts of companies, governments, and other organizations, and in turn, they would be less prone to act fairly. The document states that transparency promotes accountability and inform citizens about government actions, and calls for U.S Federal Agencies to employ new technologies to divulge information about their decisions in an online and prompt manner, ideally using the feedback of the users to identify what information is of greater interest (Obama, 2009). Still, many authors also point out the limitations of transparency.

According to Fenster (2006), governmental transparency cannot be complete, since complete access to all governmental data is not a conceivable scenario for governments. This inability comes in part from prohibitive logistical issues and costs, but also, this kind of disclosure would prevent the correct functioning of many of the government’s most important operations. Other issue is that full transparency is a threat on the

privacy of citizens, who are usually obliged to give most of their personal information to the government. Fenster (2006) also states that the excessive disclosure of information may render a nation more susceptible to security breaches and less able to enforce its own laws because evildoers will have greater access to information that could be used to threaten the health and safety of the public. Christensen & Cornelissen (2015) insist that being transparent is usually taken as a promise of openness, exposure and insight and access to the truth about the organization, the commitment to organizational transparency is based on the principle that if procedures and methods are open to scrutiny, then the organization is open to critique and ultimately to improvement. Thus, transparency is seen as a reply to the demand for increased openness and accessibility, and counting on that openness facilitates insight and clarity, organizations increasingly adopt their structures to signal a practice of openness and truthfulness.

It is worth emphasizing that the technological tools must provide citizens with easy and rapid interaction, enabling the public to contribute to the definition of how public resources are applied. Anjos and Ezequiel (2011) show that the use of new technologies by public administration may represent an openness of the governments to the public, allowing ample access to the digitalized databases. The right to use these databases to obtain information and services directly, without intermediaries, is an important contribution to exercise of citizenship. Consequently, the use of technology can contribute positively to the dissemination and guidance given to citizens, if methods to improve the usability are implemented, allowing the use of the information by citizens.

Nielsen and Loranger (2007) present usability as a quality attribute related to the easiness of using something. More specifically, usability refers to the quickness with which users can learn to use something, their efficiency in using them, how much they remember of it, and how much the users like to use it. Therefore, web information systems expected to be transparent have to provide

ease of use, clarity and comprehensibility of the information available.

Accordingly, the use of ICTs as a mean to transparency, and the central role of the organizational intention on the implementation of technology make clear how the right to know (information) is not automatically granted when using ICTs. On the contrary, the use of ICTs should be guided by a theory of disclosure and its relationship to organizational and stakeholder interests, and to design systems that have high degrees of usability, as to allow the broad use and understandability of the information provided. At the same time, currently there is small attention to the rights to know of the stakeholders (Elia, 2009).

SOLUTIONS AND RECOMMENDATIONS

It is noteworthy that the level of disclosure, or transparency, can be managed by the organization. The organization can diminish its level of disclosure by making more information secret; or it can increase disclosure by the employing open information systems. There is a strong research evidence that managers have a tendency to conceal negative organizational information. In addition, organizations can profit from not disclosing proprietary knowledge. Still, according to empirical studies organizations may keep two kinds of secrets: sanctioned and unsanctioned secrets (Schnackenberg, Tomlinson, 2014).

The sanctioned secrets are those that are intentionally hidden by the organization for the purposes of competitive advantage, they are “sanctioned” because keeping this kind of information secret is commonly considered as a legitimate action by the stakeholders. Fung et al. (2007) use the concept of target transparency as a way for organizations to reduce specific risks or performance problems through selective disclosure and does this by providing a careful analysis of the constituents of transparency. On the other hand, the stakeholders see the unsanctioned secrets as illegitimate, even

though these secrets may have resulted from the need to maintain competitiveness or good image during internal organizational crises (Schnackenberg, Tomlinson, 2014).

According to Christensen & Cornelissen (2015), in the last years, transparency has developed from a question of how organizations can adapt to external requests for information to a deliberate strategy used to pursue respectability and social accreditation through the perception of being more transparent. Business consultants are recommending to organizations when dealing with the call for transparency to be proactive in carefully choosing the fields and the activities where they will demonstrate this value to their stakeholders. This notion of commitment beyond compliance, that is, disclosing more than obliged to, is being recognized as social responsibility, where corporations increasingly are expected to move from agreement to engagement. Both trends brings the necessity of proactive organizational initiatives in the transparency field (Christensen & Cornelissen, 2015).

Oliveira and Maciel (2013) assert that public spending transparency is an essential factor for strengthening the relationship between society and public managers. The disseminated information must be clear, understandable and easy to access. The citizens’ access to simple and understandable information is the starting point for greater transparency. It is important to notice that since being transparent is an active instead of a passive act, organizations carefully select, simplify and summarize data before they are revealed, they selectively disclose information, in their own timing of disclosure, which sometimes can be used to avoid critique or deal with potential issues (Oliveira & Maciel, 2013).

It can be argued however that transparency not only incorporates the rather passive right of every citizen to have access to information, but also the much broader and more pro-active duty of the administration itself to ensure that information about its policy and actions is provided in an accessible fashion (Curtin et al., 2006). Moreover,

the stimulus to public transparency must be one of Public Administration goals, since it contributes to strengthening of democracy.

FUTURE RESEARCH DIRECTIONS

The importance of transparency has been long debated and it is acknowledged by both academia and general public, or in Christensen & Cornelissen (2015) words it is already a “myth”, a characteristic highly coveted by organizations, government and society, even though there is a general lack of understanding of its complex nature, costs and real possibilities. In this regard, Fenster (2006) highlight the negative side of transparency, which need to be more deeply understood. The costs associated with a greater disclosure and the needed limits and limitations of the transparency need to be further considered by the field as a whole.

The process of transparency has turned into a discourse of corporations, which already use it as a strategy of disclosing information that they deem favorable, while keeping their “secrets” or making it less understandable through obscure reporting, hard to find data and excessive use of jargon. Therefore, new research ought to be more critical of transparency. Not that the new research should not advocate for the disclosure of understandable, complete and reliable information. Instead, new research should disclose the true nature of transparency and the organization’s attitude towards this new trend, including a better consideration of the benefits and costs of transparency.

New research should focus on a general operationalization of transparency construct, if it is feasible. Future research might find that the concept is dependent on industry, type of stakeholder of other variable. Other important source of research on transparency is related to Open Government Data¹, which calls for open data and easy access to previously obscure information, with the challenge of producing information that is understandable for the average citizen.

CONCLUSION

The organizational transparency construct is still at development. Not only the operationalization but also the consequents and antecedents have to be better understood by governments, organizations and society alike. Notwithstanding, there is a clear demand for transparency for both government and organizations. This is a point of no return. The tendency is for an even greater call for transparency of all kind of data and information produced by the organizations and governments.

In this regard, increasing transparency can strengthen the lines of accountability between government and citizens. That is, when citizens are better informed about the governmental performance, citizens are more capable to pressure public officials to perform their duties in the public interest (Shah and Schacter, 2004). However, legislative or constitutional commitment to transparency does not magically lead to the informed, deliberative, and/or participatory citizens (Fenster, 2006). The correct allocation of public resources and the dissemination of public information is crucial, not only to meet the legislation in force, but also to be accessible to all citizens who have the right to know how public money is being used (Oliveira & Maciel, 2013).

Nevertheless, the search for an equilibrium point of between full disclosure and secrecy, carefully weighting the relative costs and benefits of transparency in each context, is a very important issue. There should be special efforts to define what kinds of governmental decisions, data and political participation are most likely to benefit from transparency, while also considering what costs and dangers government and institutions will face as the result of complying with the transparency requirements. Unfortunately, this does not seem to be the norm nowadays. According to Fenster (2006), this discussion about the need of equilibrium for transparency has not been sufficiently considered in academia or governments.

Agreeing with the various sources consulted, one can conclude that the simple disclosure of information is far from being transparency. Christensen & Cornelissen (2015) indicate that the exact relationship between transparency, openness and information are often unclear. More important is to note the even though organizational openness is a precondition for transparency; it is insufficient for true transparency, since the mere disclosure of information does not consider the capacity of the (multiple) receivers to process the information disclosed.

Thus, the disclosure, the clarity and accuracy of the information are dependent on the receptor or user of the information. Thus, some stakeholders may consider an organization transparent, while the same organization may seem opaque to others. It only will be truly transparent when all conditions for transparency have been met: completeness, understandability and reliability.

REFERENCES

- Anjos, G. M. A. A., & Ezequiel, V. C. (2011). Cidadania virtual: O espetáculo do governo eletrônico. *Estudos de Sociologia*, 16(30).
- Bertot, J. C., Jaeger, P. T., & Grimes, J. M. (2010). Using ICTs to create a culture of transparency: E-government and social media as openness and anti-corruption tools for societies. *Government Information Quarterly*, 27(3), 264–271. doi:10.1016/j.giq.2010.03.001
- Cappelli, C. (2009). *Uma abordagem para transparência em processos organizacionais utilizando aspectos* (Doctoral dissertation). PUC-Rio.
- Cappelli, C., Leite, J. C. S. D. P., & Oliveira, A. D. P. A. (2007). Exploring business process transparency concepts. In *Requirements Engineering Conference, 2007. 15th IEEE International* (389–390). IEEE. doi:10.1109/RE.2007.35
- Christensen, L. T., & Cornelissen, J. (2015). Organizational transparency as myth and metaphor. *European Journal of Social Theory*, 18(2), 132–149. doi:10.1177/1368431014555256
- Curtin, D., & Meijer, A. J. (2006). Does transparency strengthen legitimacy? *Information Polity*, 11(2), 109–122.
- Elia, J. (2009). Transparency rights, technology, and trust. *Ethics and Information Technology*, 11(2), 145–153. doi:10.1007/s10676-009-9192-z
- Fenster, M. (2006). The opacity of transparency. *Iowa Law Review*, 91.
- Fung, A., Graham, M., & Weil, D. (2007). *Full disclosure: The perils and promise of transparency*. Cambridge University Press. doi:10.1017/CBO9780511510533
- Henriques, A. (2007). *Corporate truth: The limits to transparency*. Earthscan.
- Holzner, B., & Holzner, L. (2006). *Transparency in global change: the vanguard of the open society*. University of Pittsburgh Pre.
- Hood, C. (2006). Transparency in historical perspective (No. 135, pp.3–23). Oxford University Press.
- Leite, J. C. S. P., & Cappelli, C. (2010). Software transparency. *Business & Information Systems Engineering*, 2(3), 127–139. doi:10.1007/s12599-010-0102-z
- Menéndez-Viso, A. (2009). Black and white transparency: Contradictions of a moral metaphor. *Ethics and Information Technology*, 11(2), 155–162. doi:10.1007/s10676-009-9194-x
- Midgley, M. (2004). *The myths we live by*. Oxon, UK: Routledge.
- Nielsen, J., & Loranger, H. (2007). *Usabilidade na web: projetando websites com qualidade*. Rio de Janeiro: Campus.

Obama, B. (2009). *Memorandum for the heads of executive departments and agencies*. Re: Freedom of Information Act, Office of the President of the United States [Obama Memorandum].

Oliveira, L. K. B., & Maciel, C. (2013). Transparency and Social Control via the Citizen's Portal: A Case Study with the Use of Triangulation. In *Technology-Enabled Innovation for Democracy, Government and Governance* (pp. 112-124). Springer Berlin Heidelberg. doi:10.1007/978-3-642-40160-2_10

Rosendorff, B. P., & Vreeland, J. R. (2006). *Democracy and data dissemination: the effect of political regime on transparency*. New York University and Yale University. Available at <http://homepages.nyu.edu/bpr1/papers/BPRJRV.pdf>

Schnackenberg, A. K., & Tomlinson, E. C. (2014). Organizational Transparency A New Perspective on Managing Trust in Organization-Stakeholder Relationships. *Journal of Management*, 0149206314525202.

Shah, A., & Schacter, M. (2004). Combating corruption: Look before you leap. *Finance & Development*, 41(4), 40–43.

Transparency. (n.d.). *Online Etymology Dictionary*.

Vaccaro, A., & Madsen, P. (2009). Corporate dynamic transparency: The new ICT-driven ethics? *Ethics and Information Technology*, 11(2), 113–122. doi:10.1007/s10676-009-9190-1

Weber, R. H. (2008). Transparency and the governance of the Internet. *Computer Law & Security Report*, 24(4), 342–348. doi:10.1016/j.clsr.2008.05.003

Wehmeier, S., & Raaz, O. (2012). Transparency matters: The concept of organizational transparency in the academic discourse. *Public Relations Inquiry*, 1(3), 337–366. doi:10.1177/2046147X12448580

Williams, C. C. (2005). Trust diffusion: The effect of interpersonal trust on structure, function, and organizational transparency. *Business & Society*, 44(3), 357–368. doi:10.1177/0007650305275299

ADDITIONAL READING

Bertot, J. C., Gorham, U., Jaeger, P. T., Sarin, L. C., & Choi, H. (2014). Big data, open government and e-government: Issues, policies and recommendations. *Information Polity*, 19(1,2), 5-16.

Hood, C., & Heald, D. (2006). *Transparency: The key to better governance?* (Vol. 135). Oxford University Press/British Academy. doi:10.5871/bacad/9780197263839.001.0001

Kopits, M. G., & Craig, M. J. (1998). *Transparency in government operations* (No. 158). International monetary fund. doi:10.5089/9781557756978.084

Mitchell, R. B. (1998). Sources of transparency: Information systems in international regimes. *International Studies Quarterly*, 42(1), 109–130. doi:10.1111/0020-8833.00071

Tapscott, D., & Ticoll, D. (2003). *The naked corporation: How the age of transparency will revolutionize business*. Simon and Schuster.

Williams, C. A. (1999). The securities and exchange commission and corporate social transparency. *Harvard Law Review*, 112(6), 1197–1311. doi:10.2307/1342384

KEY TERMS AND DEFINITIONS

Disclosure of Information: The act of making relevant information about the organization readily available to its stakeholders.

Opaque: Quality of an organization or data not being transparent, usually used as the opposite of transparent.

Openness: Quality of being open, transparent and available.

Process Transparency: The existence of a defined organizational policies to disclose information to the stakeholders, subject to criteria of the information being accessible, able to be used, well presented, prone to be understood and auditable.

Target Transparency: Selective disclosure of information by organizations to reduce risk or performance issues.

Transparency: The propriety of disclosing information to the stakeholders that is complete, reliable and understandable.

Transparent Organizations: Organizations that disclose information about its activities, process and all relevant information to its stakeholders.

ENDNOTE

¹ <http://opengovernmentdata.org/>

Social Business Process Modeling

B**Fadwa Yahya***University of Sfax, Tunisia***Khouloud Boukadi***University of Sfax, Tunisia***Zakaria Maamar***Zayed University, UAE***Hanêne Ben-Abdallah***King Abdulaziz University, Saudi Arabia*

INTRODUCTION

Web 2.0, which represents a major evolution of the regular Web (*aka* Web 1.0), provides Internet users with a set of technologies (e.g., AJAX and JSON) and applications (e.g., Facebook and Twitter). It aims at increasing communication, collaboration, and knowledge sharing among multiple stakeholders (Yahya, Boukadi, Maamar, & Ben-Abdallah, 2015). To tap into these Web 2.0's opportunities, enterprises are putting a lot of efforts into adopting Web 2.0 in their day-to-day operations. These enterprises are referred to as *Enterprise 2.0* (Mäntymäki & Riemer, 2016).

More specifically, an Enterprise 2.0 is characterized by the use of Web 2.0 technologies and applications to achieve different goals (McAfee, 2006) like improving its visibility on search engines, reducing the cost of some services such as communication, improving the quality of services it provides especially those related to customer satisfaction, etc. It may also use Web 2.0 to enhance transparency by enhancing the availability of information and knowledge across its organizational units (DiStaso & Bortree, 2012).

To reach its goal, an Enterprise 2.0 may need to "re-engineer" its Business Processes (BP) due to the new context imposed by Web 2.0. The resulting business process, called Social Business

Process (SBP) (Yahya et al., 2015), differs from the classical business process in two main aspects: On the one hand, the new interactions over Web 2.0 should be integrated into the SBP model so that "to-be" *versus* "as-is" processes are aligned: on the other hand, the business and social aspects of an SBP should be modelled separately so that the "separation-of-concerns" principle is supported. The business aspect may be represented using any existing notation like UML activity diagram (OMG, 2011b) or Business Process Model and Notation (BPMN) (OMG, 2011a). However, the social aspect is not supported by existing notations and thus should be modelled using a Domain Specific Language (DSL) that extends any existing business process modelling language. To provide for such DSL, we first enrich the BP meta-model proposed in (Curtis et al., 1992) with social aspects. Secondly, based on the obtained meta-model, we develop a BPMN extension, called BPMN4Social, for modeling SBP. BPMN4Social is a simple, yet generic domain-specific language that is based on BPMN, the de facto standard for business process modelling.

The remainder of this chapter is structured as follows. Section 2 discusses works related to SBP. Section 3 introduces our SBP meta-model which we produced by enriching the business process meta-model with social concepts. Section

4 presents BPMN4Social including its meta-model and concrete syntax. Section 5 illustrates the BPMN4Social notation supported by its editor. Finally, Section 6 summarizes the chapter and outlines future works on SBP.

BACKGROUND

The business process meta-model of Curtis et al. (Curtis, Kellner, & Over, 1992) is the most referenced business process meta-model by the ICT community (Figure 1). It encloses the core business process components classified into four perspectives: functional, organizational, behavioral, and informational. The functional perspective focuses on the tasks in a business process where a task is either an atomic or composite unit of work. The organizational perspective describes the units that participate in the business process execution. The behavioral perspective represents the flows and control nodes linking the tasks of the business process. Finally, the informational perspective describes the entities that a business

process produces or manipulates. These entities could be events, information resources, or tangible resources. These four perspectives are commonly modeled using BPMN (OMG, 2011a).

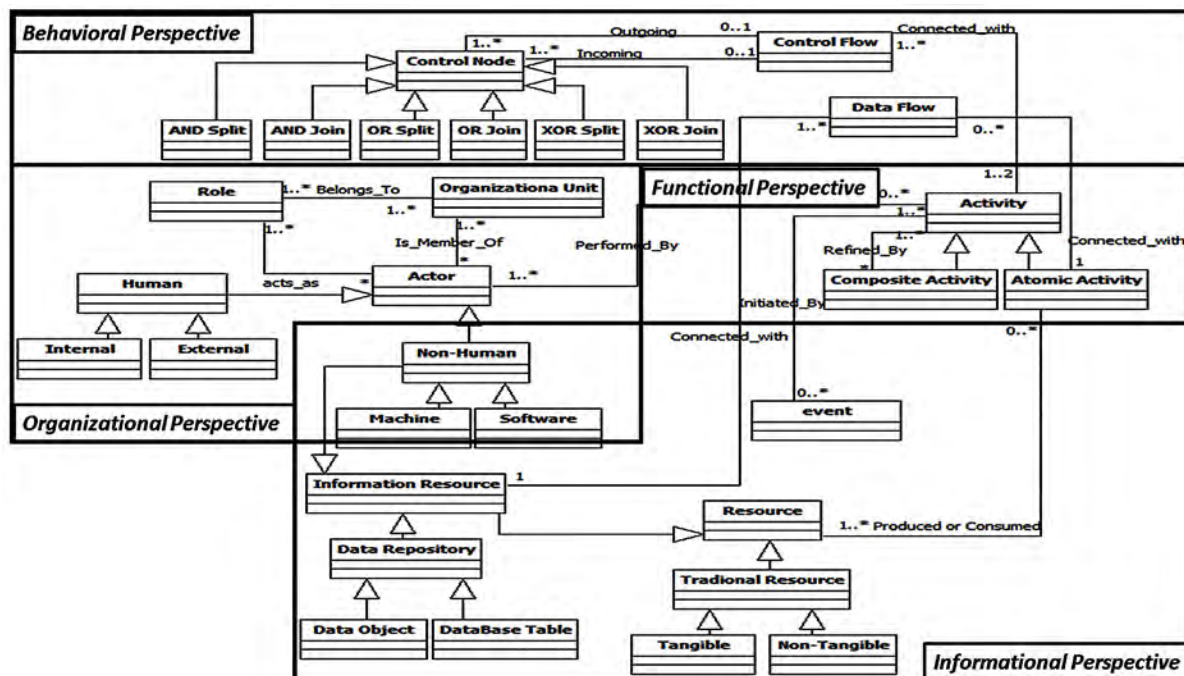
RELATED WORK

In this section, we present the most important works dealing with social business process management systems (BPMS) and those focusing on Social Business Processes.

Web 2.0 and Business Process Management Systems

In the literature, different studies highlight the role of Web 2.0 in improving traditional BPMSs. According to Schmidt et al. (Schmidt & Nurcan, 2009), weak ties, egalitarianism, social production, and service-dominant logic principles of Web 2.0 would address four well-known BPM problems: model-reality divide, lack of information fusion, information pass-on threshold, and lost innovation.

Figure 1. Business process meta-model



For instance, weak ties should help enterprises capture the necessary knowledge for improving their BPs and thus, preventing problems of lack of information fusion and loss of innovation (Schmidt & Nurcan, 2010). Schmidt et al., also, argue that egalitarianism combined with service-dominant logic ensure the involvement of final users in a process's lifecycle (i.e., modeling, implementation, execution, and evaluation) and thus, reduce the risks of model-reality divide where "to-be" *versus* "as-is" processes are not aligned. Furthermore, social production helps generate knowledge by collecting information from different participants in an uncontrolled way so that the problem of information pass-on threshold is addressed.

Erol et al. (Erol et al., 2010) argue that social software (i.e., Web 2.0 applications) allows the integration of all internal and external business actors into the BP lifecycle in order to reach an operational and flexible BP model. In particular, social software offers tools that various internal and external actors can use to collect knowledge, which can be utilized in the continuous improvement of the process. Yet, the authors recognize that social software can raise some security challenges.

Rangiha et al. (Rangiha & Karakostas, 2013) propose a goal-oriented approach to define a Social BPMS (SBPMS) that uses social software to engage final users in the BP lifecycle. This work addresses the lack of an effective method for enhancing BPMS with social software especially during the enactment stage. As a result, they introduce goals starting at the modeling stage to act as controls. Thanks to the goals, users can specify the sequence of tasks to be performed when executing processes.

All of the aforementioned approaches concentrate only on the ability of social software to improve BPMS without considering their ability of improve the business processes themselves. In fact, the improvement of BPMS will result in the improvement of the business processes; however, it does not guarantee the establishment of social business processes (Yahya et al., 2015).

Web 2.0 and Enterprise Business Processes

Brambilla et al. (Brambilla, Fraternali, & Vaca, 2011) propose a social network-based extension of BPMN to model SBP. The proposed notation extends the following BPMN elements:

- The BPMN pool is extended by three types of pool named internal performer, internal observer and external observer;
- The BPMN task is extended by several specific tasks such as commenting, invitation to activity, and voting.
- The BPMN gateway is extended by two specialized decisions, which are choices performed by users or automatically; and
- The BPMN event is extended by several specialized types of events like social relationship link and an invitation's acceptance/rejection.

Being closely related to social networks, this proposed BPMN extension therefore allows to model SBPs that use only a social network as Web 2.0 application. Besides this SBP notation, Brambilla et al. also propose a technical framework that allows the design, implementation, deployment, and monitoring of SBPs. In the implementation phase, the proposed framework allows enterprises to implement SBPs as Web applications connected with any social network.

In 2014, IBM proposed seven repeatable patterns that may be used to add social capabilities to business processes (IBM, 2014). Each social pattern is described in terms of: the concerned business processes, relevant industries, key stockholders, recommended actions, business challenges, potential benefits as well as the IBM experience. The proposed social patterns are applicable by organizations in various industrial sectors. However, the patterns are informally described and do not offer any guidelines for their application.

Cerenkovs et al. (Cerenkovs & Kirikova, 2014) propose an iterative, incremental development

process for incorporating social interaction within business processes. Their development process consists of four phases: inception, elaboration, construction, and transition. The inception phase provides recommendations, which could assist the enterprise assess BP socialization feasibility. The second phase models the SBP using BPMN language. The social aspect of the SBP is represented using Brambilla extension (Brambilla et al., 2011). The third phase implements the SBP. Finally, the fourth phase executes the new SBP. While the proposition of Cerenkovs et al. covers the core phases of the development process, it informally specifies the four phases and it does not provide any information about the used techniques, standards, nor implementation of the process.

SOCIAL BUSINESS PROCESS META-MODEL

To offer an adequate BPMN extension that supports the social dimension of a business process, we define an SBP Meta-Model called SBP2M. We constructed SBP2M by enriching the business process meta-model of Curtis et al. (Curtis et al., 1992) with a social aspect. We have chosen this meta-model because it is widely used in the literature as a reference for modeling different types of processes. For instance, List and Korherr extend it with a fifth aspect, called context aspect, to describe contextual information of a business process like goals, owner, and deliverable (List & Korherr, 2005). Chaâbane et al. also extend it to support business process versioning (Chaâbane, Andonoff, Bouaziz, & Bouzguenda, 2009). Ellouze et al. extend it to support inter-organizational business processes (Ellouze, Chaâbane, Andonoff, & Bouaziz, 2015).

In this chapter, we illustrate how to enrich Curtis et al.'s business process meta-model so that it supports the social dimension of an SBP. Figure 2 shows the proposed meta-model where the new concepts are represented in grey:

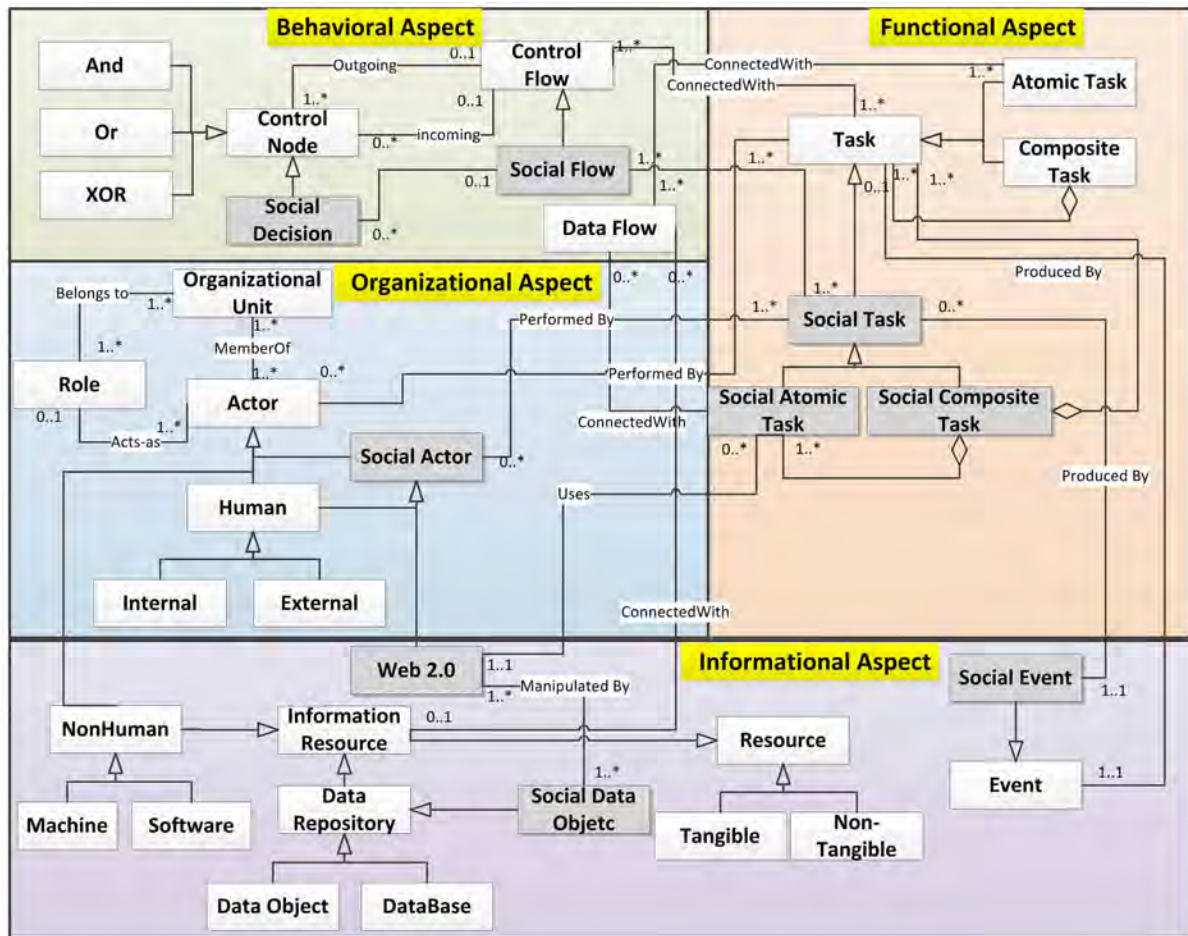
1. The functional aspect is enriched by social task that is either atomic or composite. A social task is executed by a Web 2.0 application (e.g., post advertisements on Facebook).
2. The organizational aspect is enriched by social actor, which describes those who will perform social tasks. A social actor may be either human (internal or external to the enterprise) or a Web 2.0 application.
3. The informational aspect is enriched by social concepts like Web 2.0, social event, and social data object. A social task may produce zero or many social events. In addition, a social task manipulates or produces social data object (e.g., advertisement data) that could be made available over Web 2.0.
4. The behavioral aspect is enriched by a social decision that is made over Web 2.0 applications and it is represented inside the SBP model as a control node. The behavioral aspect is also enriched by a social flow to describe the interactions that take place between actors over Web 2.0 applications. For example, a social flow can be used to model the link between posting task and the corresponding commenting and collecting feedbacks tasks.

Thanks to SBP2M, the different business entities that socialization could affect can be explicitly described (e.g., task, event, and information resource). The next section focuses on modeling SBP, which considers the social dimension of a business process.

BPMN EXTENSION FOR SOCIAL BUSINESS PROCESS MODELING

In this chapter, we propose BPMN4Social, which is a specific notation for SBP modeling. BPMN4Social distinguishes social from business elements, and it represents social interactions in the SBP model. Our aim is to provide a simple, yet generic notation that can be used easily by busi-

Figure 2. Social Business Process Meta-Model (SBP2M)



ness analysts and independently from any Web 2.0 application. On the one hand, the simplicity of the notation is ensured by reducing the number of new extensions. On the other hand, technology independence is ensured by not targeting any particular Web 2.0 service (e.g., commenting votes and chatting services) when defining the extension. The meta-model of BPMN4Social is detailed below.

BPMN4Social Meta-Model

Before presenting the BPMN4Social meta-model, let us first introduce the mapping of SBP2M social concepts onto BPMN4Social concepts. In Figure 3, only the relevant classes of the BPMN meta-

model are shown in white while BPMN4Social classes are shown in yellow.

1. SBP2M social actor is mapped onto three different social lanes:
 - a. **Social Internal Lane:** This extension models internal departments of the company that use Web 2.0.
 - b. **Social Customer Lane:** This extension models individual customers of the company, in the case of B2C, who use Web 2.0.
 - c. **Social Community Lane:** This extension models the partners of the company that use Web 2.0. Each new lane is characterized by two supplementary attributes:

- i. **Technology:** It identifies the used Web 2.0 application.
 - ii. **Type:** It distinguishes between authenticated and anonymous users.
2. SBP2M social activity is mapped onto four different social activities:
 - a. **Communication Activity:** It models communication activities such as the exchange of messages between the various internal and external actors.
 - b. **Collaboration Activity:** It models the collaborative activities, such as joint drafting activities, among the various internal and external actors.
 - c. **Collecting Activity:** It models activities that allow the collection of information and knowledge using Web 2.0. Such an activity may be used to collect customers' feedback.
 - d. **Sharing Activity:** It models activities that allow the sharing of information and knowledge using Web 2.0. Such an activity may be used to share knowledge with various actors whether known or unknown in advance. Each new activity is characterized by four additional attributes:
 - i. **Technology:** It identifies the Web 2.0 application used to execute the social activity.
 - ii. **Social Service:** It describes the service provided by the Web 2.0 application (e.g., messaging and commenting) that permits the accomplishment of the social activity.
 - iii. **Input Data:** It describes the data required to accomplish the social activity.
 - iv. **Output Data:** It describes the data produced by the social activity.
3. SBP2M social event is mapped onto two different social events:
 - a. **Send Social Notification:** It models the notification of an actor to participate in a new social activity. The sender of a social notification can invite the receiver to execute some tasks. A social notification is communicated through the used social software.
 - b. **Receive Social Notification:** It waits until a new notification is received from a participant through the used social software. These two new notifications have two additional attributes:
 - i. **Technology:** It identifies the used Web 2.0 application.
 - ii. **Social Service:** It describes the service provided by Web 2.0 that permits the reception or the transmission of the notification.
4. SBP2M social data object is mapped onto BPMN4Social social data object:
 - a. **Social Data Object:** We define a social data object as an extension of the BPMN data object to model the data required or produced by social activities. A social data object has two additional attributes:
 - i. **Type:** It identifies the type of social data object. It may be for example a text, an image, a video, etc.
 - ii. **Criticality:** It allows distinguishing between critical data, which require setting up specific security procedures, from general data.
5. SBP2M social decision is mapped onto BPMN4Social social decision:
 - a. **Social Decision:** The social decision notation allows the modeling of distributed decision made by different actors through Web 2.0. A social decision has three additional attributes:
 - i. **Technology:** It identifies the Web 2.0 application used by the SBP

- ii. **Social Service:** It describes the service provided by the used Web 2.0 application that permits distributed decision-making.
- iii. **Participant:** It describes the different actors that must be involved in the decision making process. An actor may be internal (e.g., company employees) or external (e.g., customer and partner).

as a result, it enforces the separation of concerns principle.

We implemented a graphical editor supporting this syntax as shown in Figure 4. This editor extends the BPMN palette with new components. Each new component is created as a customization of a standard BPMN element. The attributes of each new concept are accessible, like standard attributes, via the property sheet relative to the selected concept.

Graphical Editor

Figure 4. shows the concrete syntax of BPMN-4Social, which defines a visual representation for the proposed BPMN extensions. This syntax has two advantages: it is simple to use thanks to the limited number of new notations that are added to the standard BPMN; and the new concepts of BPMN4Social have a different representation and,

CASE STUDY

To illustrate the use of our graphical editor along with the benefits of the proposed extensions, we consider the maintenance business process of a large commercial company. This business process is triggered when an employee's device breaks down. The employee of the device sends a complaint to the maintenance department. Two

Figure 3. BPMN4Social meta-model

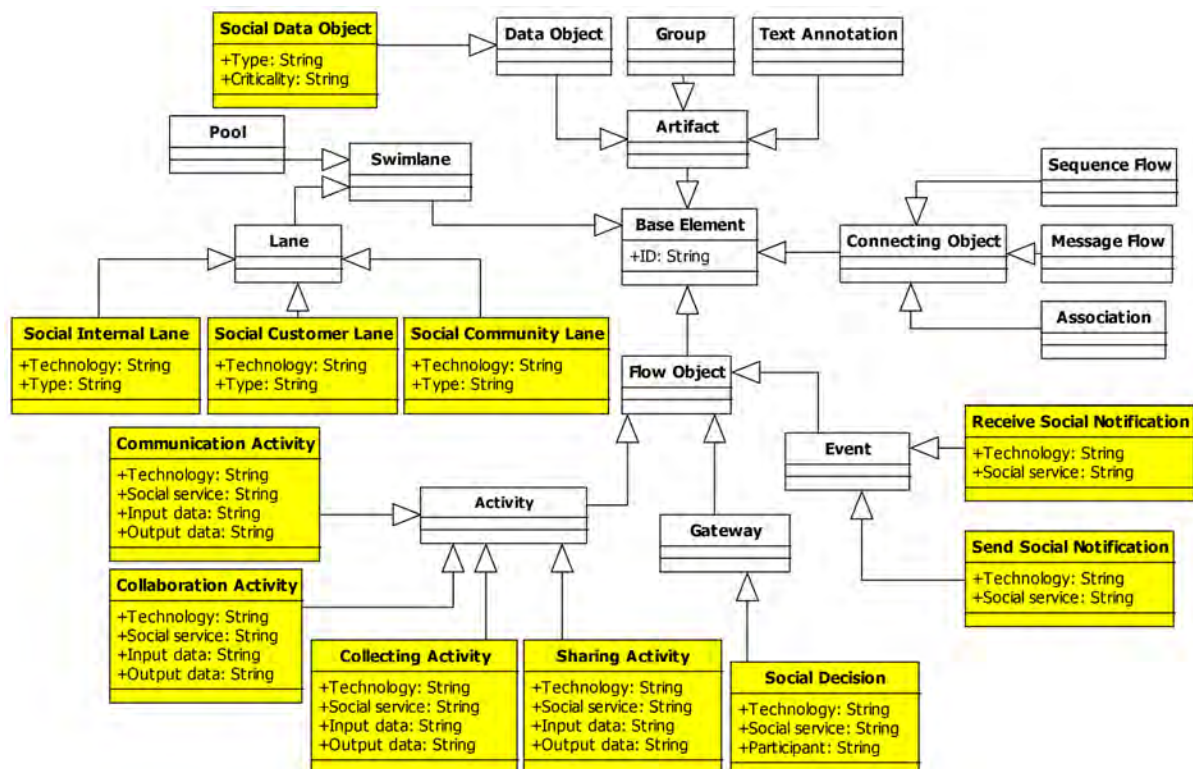


Figure 4. Graphical editor of BPMN4Social



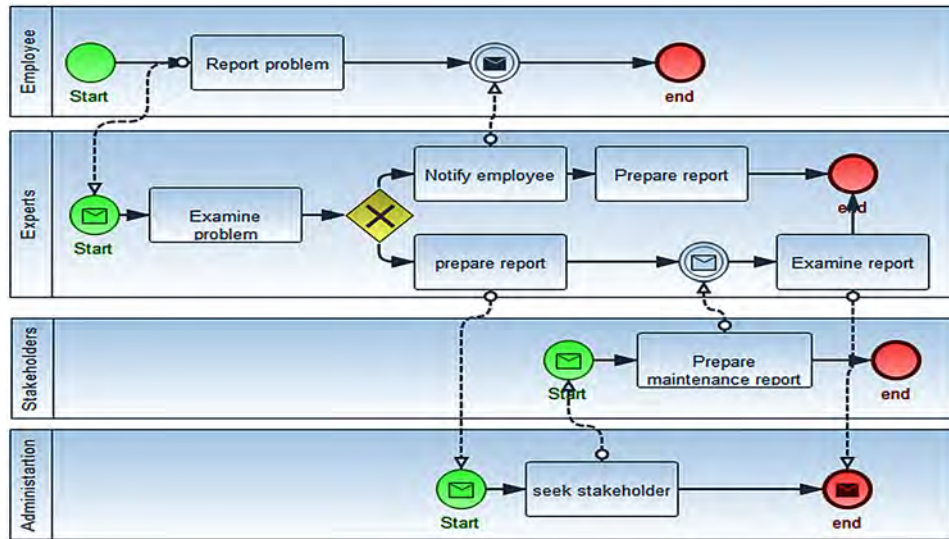
scenarios are possible: success or failure. In the first case, the maintenance department informs the employee that the problem is fixed and then prepare and submit a report to the administration. In the second case, the maintenance department submits a report to the administration that identifies some external support. Once identified, the administration sends the report to external stakeholders. These latter prepare and send a technical report to internal experts. Internal experts examine the technical report and can accept or refuse the preposition. The maintenance BP model is depicted in Figure 5.

The maintenance business process could benefit from Web 2.0 applications. Indeed, these applications could keep users and experts connected and engaged in the progress of this business process. Web 2.0 applications could be also used by the administration to find the most adequate stakeholders. To this end, administration can use

LinkedIn a Web 2.0 application. Last but not least, experts and external stakeholders can use Web 2.0 to examine and solve the problem collaboratively.

If the maintenance business process had to be modelled using a standard modeling language like BPMN, the distinction between *social* and *regular* interactions would not be clearly illustrated. Yet, by using our graphical editor, the model is clear and developers can easily distinguish social interactions and activities thanks to the new notations. Therefore, BPMN4Social models the socialized version of this BP by first, illustrating the communication between users and experts using communication activity. Second, modeling the search for an adequate stakeholder using collecting activities. Finally, introducing the collaboration between, experts and external stakeholders using collaborative and sharing activities. As such, we added social lane within each pool that will contain social activities. The choice of the nature of

Figure 5. Maintenance business process

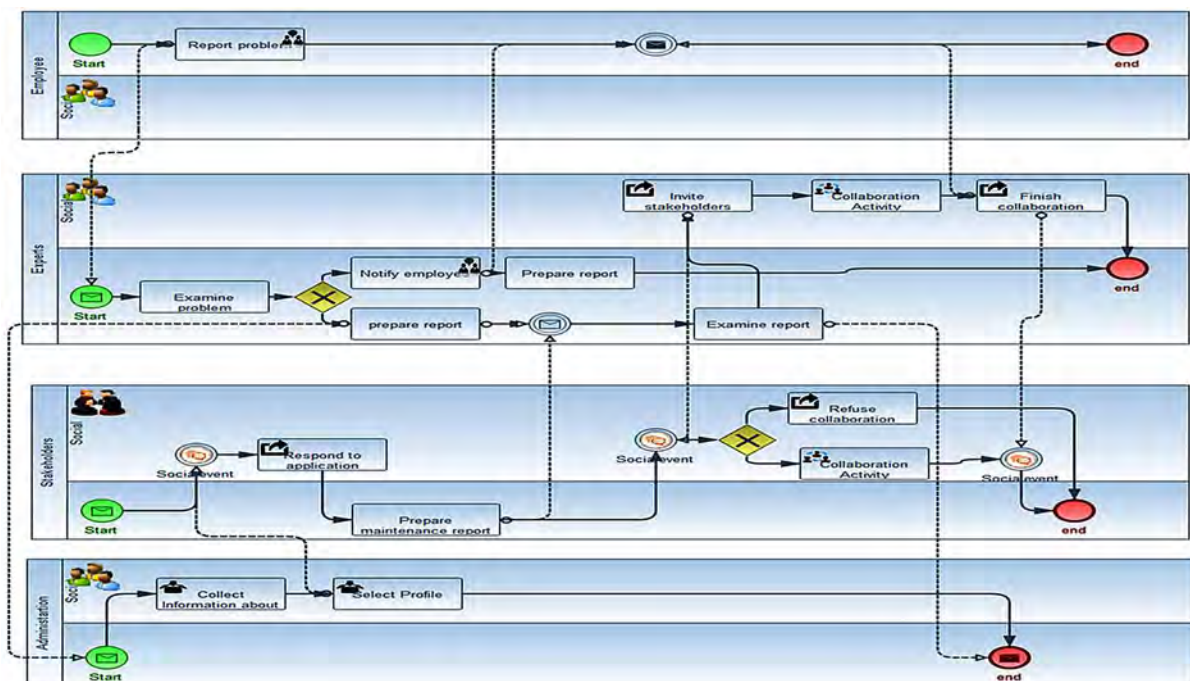


the social lane is determined by the nature of the original pool. For instance, as users, experts, and administration are internal actors, we added within their pools a social internal pool while we added a social community lane within the stakeholder pool. Figure 6 shows the maintenance business process modelled via the use of BPMN4Social.

FUTURE RESEARCH DIRECTIONS

As a future endeavor, we plan to implement an automatic code generator that generates an executable code from the social business process model. The generator will be built upon model transformations a la Model-Driven Architecture

Figure 6. Maintenance business process modelled with BPMN4Social



(Soley, 2000). In addition, we plan to propose an approach that helps enterprises blend social software with their existing BPs so they can tap into the opportunities of social media. This will help an enterprise identify business processes that could benefit from socialization while meeting the enterprise goals and requirements. It will also transform automatically the identified business processes into social business processes.

CONCLUSION

Web 2.0 enhances business processes with new functionalities that improve the exchange of knowledge and information between various actors (Zhu & Chen, 2012) (Paniagua & Sapena, 2014). Thanks to Web 2.0 applications, new communication forms between enterprises and stakeholders are created (Fortino & Nayak, 2010). In fact, several studies have shown the added value of Web 2.0's practices inside the enterprise like improvement of enterprise productivity thanks to collaboration between actors (Bennett, 2012) (Zhu & Chen, 2012) (Jussila, Kärkkäinen, & Aramo-Immonen, 2014). The combination of business and Web 2.0 incurs to several changes in business processes and thus social business processes are rolled out. Social business processes are different calling for new ways of modeling, designing, and executing them. These new social elements must be represented explicitly through a specific notation during the modeling stage.

In this chapter, we proposed SBP2M, a meta-model for social business processes. SBP2M is the result of enriching the business process meta-model of (Curtis et al., 1992) with the social aspect. Based on SBP2M, we developed a BPMN extension for modeling social business processes. The proposed DSL extends BPMN visual language with new notations like collaboration and sharing activities, which express social interactions between actors. These new notations are defined as a customization of the standard BPMN notation.

This extension is characterized by its simplicity and independence from Web 2.0 technology. Our graphical editor is implemented as an extension of BPMN2 modeler.

REFERENCES

- Bennett, S. (2012). The benefits of communicating and collaborating in a real-time enterprise social network. In *Professional Communication Conference (IPCC), 2012 IEEE International* (pp. 1–4). IEEE. doi:10.1109/IPCC.2012.6408614
- Brambilla, M., Fraternali, P., & Vaca, C. (2011). A notation for supporting social business process modeling. In *Business Process Model and Notation* (pp. 88–102). Springer. doi:10.1007/978-3-642-25160-3_7
- Cerenkovs, R., & Kirikova, M. (2014). Supporting Introduction of Social Interaction in Business Processes. In *Perspectives in Business Informatics Research* (pp. 187–201). Springer. doi:10.1007/978-3-319-11370-8_14
- Chaâbane, M. A., Andonoff, E., Bouaziz, R., & Bouzguenda, L. (2009). Versions to address business process flexibility issue. In *Advances in Databases and Information Systems* (pp. 2–14). Springer. doi:10.1007/978-3-642-03973-7_2
- Constantinides, E., & Fountain, S. J. (2008). Web 2.0: Conceptual foundations and marketing issues. *Journal of Direct Data and Digital Marketing Practice*, 9(3), 231–244. doi:10.1057/palgrave.dddmp.4350098
- Curtis, B., Kellner, M. I., & Over, J. (1992). Process modeling. *Communications of the ACM*, 35(9), 75–90. doi:10.1145/130994.130998
- DiStaso, M. W., & Bortree, D. S. (2012). Multi-method analysis of transparency in social media practices: Survey, interviews and content analysis. *Public Relations Review*, 38(3), 511–514. doi:10.1016/j.pubrev.2012.01.003

- Ellouze, F., Chaâbane, M. A., Andonoff, E., & Bouaziz, R. (2015). Modelling Flexible Collaborative Process: The VCP2M Approach. *BT - ICE-B 2015 - Proceedings of the 12th International Conference on e-Business*. <http://doi.org/doi:10.5220/0005544000560063>
- Erol, S., Granitzer, M., Happ, S., Jantunen, S., Jennings, B., Johannesson, P., & Schmidt, R. et al. (2010). Combining BPM and social software: Contradiction or chance?. *Journal of Software Maintenance and Evolution: Research and Practice*, 22(6-7), 449–476. doi:10.1002/smr.460
- Fortino, A., & Nayak, A. (2010). An architecture for applying social networking to business. In *Applications and Technology Conference (LISAT), 2010 Long Island Systems* (pp. 1–6). IEEE. doi:10.1109/LISAT.2010.5478285
- IBM. (2014, January 23). *Applying social business: The repeatable patterns that improve business processes and provide return*. Retrieved April 7, 2015, from <http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?infotype=SA&subtype=WH&htmlfid=EPW14036USEN>
- Jussila, J. J., Kärkkäinen, H., & Aramo-Immonen, H. (2014). Social media utilization in business-to-business relationships of technology industry firms. *Computers in Human Behavior*, 30, 606–613. doi:10.1016/j.chb.2013.07.047
- Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons*, 53(1), 59–68. doi:10.1016/j.bushor.2009.09.003
- List, B., & Korherr, B. (2005). *A uml 2 profile for business process modelling*. Springer.
- Mäntymäki, M., & Riemer, K. (2016). Enterprise social networking: A knowledge management perspective. *International Journal of Information Management*, 36(6), 1042–1052.
- Mcafee, A. (2006). *Enterprise 2.0, version 2.0*. Retrieved March 3, 2015, from http://andrewmcafee.org/2006/05/enterprise_20_version_20/
- OMG. (2011a). *BPMN 2.0*. Retrieved February 28, 2015, from <http://www.omg.org/spec/BPMN/20100601/10-06-02.pdf>
- OMG. (2011b). *UML 2.4.1*. Retrieved February 28, 2015, from <http://www.omg.org/spec/UML/2.4.1/>
- Paniagua, J., & Sapena, J. (2014). Business performance and social media: Love or hate?. *Business Horizons*, 57(6), 719–728. doi:10.1016/j.bushor.2014.07.005
- Rangihia, M. E., & Karakostas, B. (2013). Goal-driven social business process management. In *Science and Information Conference (SAI)* (pp. 894–901). CONF, IEEE.
- Schmidt, R., & Nurcan, S. (2009). BPM and social software. *Lecture Notes in Business Information Processing*, 17, 649–658. http://doi.org/10.1007/978-3-642-00328-8_65
- Schmidt, R., & Nurcan, S. (2010). Augmenting bpm with social software. In *Business Process Management Workshops* (pp. 201–206). Springer.
- Soley, R. (2000). Model driven architecture. *OMG White Paper*, 308(308), 5.
- Yahya, F., Boukadi, K., Maamar, Z., & Ben-Abdallah, H. (2015). Enhancing Business Processes with Web 2.0 Features. *International Conference on E-Business, Colmar, Alsace, France.*, 183–190. doi:10.5220/0005526501830190
- Zhu, A., & Chen, X. (2012). A Review of Social Media and Social Business. In *Multimedia Information Networking and Security (MINES), 2012 Fourth International Conference on* (pp. 353–357). <http://doi.org/doi:10.1109/MINES.2012.44>

ADDITIONAL READING

Brambilla, M., Fraternali, P., & Vaca, C. (2011, August). BPMN and design patterns for engineering social BPM solutions. *In International Conference on Business Process Management* (pp. 219-230). Springer Berlin Heidelberg.

Maamar, Z., Faci, N., Kajan, E., Boukadi, K., Sakr, S., Boukhebouze, M., & Hennion, R. et al. (2014, April). Enterprise 2.0: Research Challenges and Opportunities. *In International Conference on Web Information Systems and Technologies* (pp. 16-30). Springer International Publishing.

Myddleton, J., & Fullwood, C. (2016). Social Media Impact on Organisations. In *Applied Cyberpsychology* (pp. 216-235). Palgrave Macmillan UK.

KEY TERMS AND DEFINITIONS

Social Business Process: It is “a BP that uses any Web 2.0 technology or application to achieve the enterprise business goals and to foster

communication, collaboration, and exchange of knowledge either among internal or with external actors in everyday work” (Yahya et al., 2015).

Social Media: Kaplan and Haenlein define social media as “a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of user generated content.” (Kaplan & Haenlein, 2010).

Web 2.0: Constantinides and Fountain define Web 2.0 as “a collection of open-source, interactive, and user-controlled online applications expanding the experiences, knowledge, and market power of the users as participants in business and social processes. Web 2.0 applications support the creation of informal users’ networks facilitating the flow of ideas and knowledge by allowing the efficient generation, dissemination, sharing and editing/refining of informational content” (Constantinides & Fountain, 2008).

Social Issues in IT Project Teams

B**Awie C. Leonard***University of Pretoria, South Africa***D. H. Van Zyl***University of Pretoria, South Africa*

INTRODUCTION

The role and impact of social issues in IT project teams should not be underestimated. With the involvement of virtual project teams, this is even more relevant. One such social issue is the establishment of relationships between members of teams. The establishment and maintenance of social relationships and networks in the IT project environment is a phenomenon all IT professionals are exposed to and in many cases involved in. Furthermore, these relationships are used by IT project team members for personal as well as professional purposes. The question is what (negative or positive) impact such social relationships and networks might have on the success or failure of any given IT project?

The objective of this chapter is therefore to emphasize the important role social relationships and networks play in the IT project team environment. Furthermore, to illustrate to the management (project managers and project leaders) of software project teams how such relationships can have a positive or negative impact on team members. In this regard a climate or culture should be allowed for these relationships to flourish to the benefit of the IT department.

BACKGROUND

The IT project management literature is extensive with regard to success factors as well as the

causes of failure; however, little focus is placed on the role or importance of social relationships and networks within IT projects.

Liebowitz (1999) feels that the greatest threat to the success of any IT project is the failure to communicate. This statement in particular draws the attention to the problem area of the research. Although one wants to see a project environment where a culture of sound communications is promoted, it is difficult for any project manager to “control” any influence this might have on team members and as such on the progress of a given project. Sauer (1993) believes that a major part of the problem of IT project failure is the lack of recognition that information systems development is largely a social and political process. This view is also shared by Standing (1998). Considerable effort has already been spent on the process of managing IT projects and has produced multiple methodologies and methods for project management and the IT software development life cycle (Standing & Bavington, 1996). This is further complicated by the fact that virtual teams are in particular an area of great risk as pointed out by Reed & Knight (2010). According to them, the lack of project team cohesion can impact seriously on team performance. Problems with team cohesion can occur when there is conflict between team members or when the relationships necessary for the team to function do not develop.

Ashworth and Carley (2006) state for example that “Social network theories suggest that the types and degrees of an individual’s relationships

in social and communication networks are key impactors of group performance, while resource dependency theory suggests that non-social factors, such as knowledge and skills, figure at least as prominently as social dimensions in determining such performance.”

In organisational theory, managers are viewed as contributing over and above the skills they have acquired through experience and education, the value of their social networks. These values or assets refer to the social capital of the manager. Scholars have highlighted the ability of these social networks that can be used to the individual’s or organisation’s advantage (Gargiulo & Benassi, 2000; Ashworth & Carley, 2006). With this in mind, the question is how social relationships and networks within IT project teams are viewed, instead of focusing only on that of the project managers. The social capital of the individuals participating in the IT project teams is an influencing factor on the social networks that are active within the project teams.

The first consideration is that of determining the strength of these social networks. Network strength can be defined as the frequency of communication, while the degree of the network is defined as the number of direct links with other network members (Monge & Contractor, 2003, cited in Hovorka & Larsen, 2006).

Social networks have a key function in the social information processing within an organisation,

especially relating to connecting social influence, knowledge and the organisational culture to the actual projects at hand. This influence is depicted in Figure 1.

In the rest of the chapter a brief theoretical analysis of the field is described as well as a brief description of the empirical research.

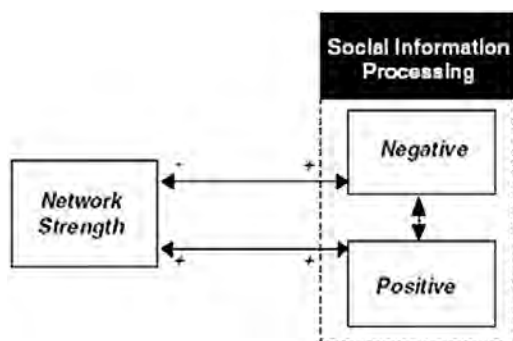
A THEORETICAL BACKGROUND ON SOCIAL RELATIONSHIPS AND NETWORKS

Relationships between end users and team members of information technology (IT) projects are described by Leonard (2002) as intriguing and complex. According to Leonard a large number of elements (amongst others support; cooperation; knowledge and commitment) are involved during the establishment and maintenance of sound relationships. Furthermore, he argues that if any of these elements are disturbed, the whole relationship is disturbed. In other words, these elements form an holistic “unit.” Each of these elements therefore plays a specific social role in a relationship, which impacts on the soundness of a relationship and as such on the cooperation between team members.

In order to overcome the problem of poor relationships between IT professionals and end users, for example, it is argued that a “human-behaviour” strategy of some kind should be followed. This strategy should involve amongst other things focusing on those social issues that will enhance trust, commitment, co-operation etc. Reich & Benbasit (1999, referred to by Leonard (2002)) point out that there are two dimensions to strategy creation: the intellectual dimension and the social dimension. With regard to this research, the social dimension was the focus.

Sound social relationships could be regarded as an important ingredient for any working environment. Not only between employees in general, but also for the purpose of organisational learning and support.

Figure 1. Interaction of network strength and social information processing
Source: Hovorka & Larsen, 2006



Trust

Although all the elements mentioned by Leonard (2002) are important for the establishment and maintenance of sound relationships, it is noteworthy that the element of trust could be regarded in most cases as the basic ingredient for sound relationships, and therefore more theory is given in this regard.

Trust is an important component in social relationship building, but it remains a complex and ambiguous phenomenon (Kadefors, 2004). Kadefors defines trust as: “Trust is a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behaviour of another.” This definition implies that trust is not a behaviour, but rather a psychological state. Furthermore, Kadefors explains that trust is not a prerequisite for cooperation, but the presence of trust improves and extends the level of cooperation.

Trust becomes important in IT projects due to the high level of cross-functional members that participate in IT projects and the associated fundamental need for cooperation. Required trust levels are directly affected by the situational circumstances and the team dynamics within IT projects. Finding the right balance of trust is important, as there are costs associated with trust (Kadefors, 2004):

- Direct costs are associated with the building of trust.
- Potential costs of breaching trust.
- Costs resulting from inefficiencies due to excessive levels of trust.
- Trust can be created by the following methods:
 - **Relational Trust:** Relational trust is created through repeated interaction between individuals. Trust is based on the personal experiences of individuals and their interpretation of events.

- **Calculus-Based Trust:** Calculus-based trust results when the trusting party believes that the trusted party will deliver on the promised actions, as this delivery is in the financial interests of the trusted party.
- **Institution-Based Trust:** Trust is created through institutions, such as legal systems, regulatory systems and societal systems. These systems are very much context-related and the levels of trust can thus differ, based on their context.

Trust is thus a context-based psychological state that is affected by the participating individuals, as well as the associated circumstances. Kadefors (2004) explains it in the following way: “We trust a colleague or exchange partner in some situations but not in others, and decisions on whether or not to trust are continuously revised in the light of new information.”

In the next part of the chapter the empirical research is presented (based on project charter and project closure documentation) to illustrate what social activities take place in a typical IT project environment.

EMPIRICAL RESEARCH

A large South African financial bank was chosen for the investigation. The motivation for using this bank was because of the large number of on-going projects that exist at any given point in time. This bank also undertakes a considerable number of IT projects on an annual basis.

Information was primarily obtained from project documentation of completed or abandoned projects within the IT departments of the corporate bank. A survey was created, based on the initial findings of the grounded theory research, and these findings were then sorted into categories.

Based on the information analysed during the initial literature review and personal experience, the following question was formulated: Do the social relationships and networks within project teams and external to these teams influence the outcome of such projects? During the grounded theory process the following secondary questions were identified and addressed:

- How are social networks used?
- Do separate and distinct social networks develop within project teams?
- Do pre-existing social relationships and networks between potential team members influence the dynamics of a new project team?
- What factors outside the project team have an influence on the social relationships and networks?

To answer the above mentioned questions, the empirical research process took the form of reading through a large number of relevant project documentation. At the same time data was arranged, categorised and analysed. Answers to certain research questions emerged and a comparison with existing literature was done. In the following section a theoretical background is given, as well as how the empirical research process took place.

The survey utilised in the research is classified as a cross-sectional survey, as it involves approaching a sample of respondents only once. The sample is regarded as a cross-section of the population under study. The survey results were used to compare subgroups (such as project managers and developers) and evaluate relationships between variables.

The target population for the survey was limited to members (project managers, developers and other participants in IT projects) of two case studies. The survey was distributed to 100 possible candidates in the form of a self-administered Internet questionnaire.

The Likert Scale was used to measure the participants' views on the categories identified during

the initial data collection phase. The Likert Scale allows for interval scales and a full spectrum of statistical analysis. A deductive interpretavistic approach was applied to analyse the results of the survey.

The data used for the initial Grounded Theory (GT) process was obtained from project documentation. These documents consisted primarily of project charter and closure documentation. The data was used to identify the concepts and categories during the data collection phase of GT. Further information relating to the concepts and categories were obtained using a survey distributed to participants from the two selected case studies which is briefly described in the next section.

Case Studies

The two case studies used for the GT process are based on two different IT environments. The first case study refers to the primary centralised approach and the second case study relates to a decentralised approach. The two environments where the case studies took place, are just briefly discussed because of space limitations.

IT Environment 1 (Case Study 1)

In the first case a Technology Program Office was created by the bank to provide integrated end-to-end management services for the strategic project portfolio. The bank makes use of program management to control the myriad of projects executed concurrently within its environment. The project management methodology was based on the PMI PMBOK. All nine knowledge areas are implemented, but tailored to the bank's environment.

IT Environment 2 (Case Study 2)

The second case focuses on the bank's Personal Loans Information Technology division, which is a full-fledged IT department that services the needs of the Personal Loans business division

within the bank. The primary focus of the team is to provide and maintain information systems that cater for the specific needs of the Personal Loans mono-line. Their approach to project management was similar to that of case study 1.

Analysis of Data

The ATLAS.ti software was used to perform the GT data ordering, data analysis and theory development phases. Furthermore, the process was used to identify the impact areas as well as the impact types of social relationships and networks.

Impact Areas

The impact areas (with their related issues) of social nature are as follows:

- The role of leadership within the project. The following issues are related to this impact area:
 - The presence of a strong leadership component within the project.
 - Importance of leadership versus procedures.
 - The level of leadership and the project outcome.
 - The level of support for leadership within the project team.
- The project culture. The following issues are related to this impact area:
 - The presence of a recognisable culture within the project.
 - The effect of the project culture on the project team.
 - The role of the project manager in determining the project culture.
 - The project culture versus the organisational culture.
 - The influence of the project culture on the project outcome.
- The social relationships between team members. The following issues are related to this impact area:

- The level of social relationships that develop between team members.
- How the social relationships are used within the project.
- The effect of social relationships on how team members view others.
- The individual's external social networks. The following issues are related to this impact area:
 - Which type of external social networks are utilised?
 - What are the external social networks utilised for?
- External influencing factors on the project. The following issues are related to this impact area:
 - The effect of external factors on the project outcome
 - The types of external factors influencing projects.

These impact areas were used as basis for gathering more data by means of a survey. The survey and key findings are discussed in the next section.

Impact Types

Social relationships and networks are active in IT projects in three primary formats which are called impact types. They are as follows: influence, friendship, and advice:

Influence

Project team members establish social relationships within project team structures over time through their personal interactions. Stronger social relationships can be developed with some team members compared to others which can create areas of leverage for the individuals. These social relationships provide a platform for individuals to influence the project direction or decisions through:

- Influencing the project leadership.

- Influencing other team members to gain support for their own ideas and agendas.
- Using their social relationships to solve problems when project structures and procedure present a stumbling block.
- Influencing the project culture.

Harnessing this influence to the advantage of the project will increase the contribution of the specific individual to be greater than just their knowledge and skills.

Friendships

Some individuals develop social relationships with other project team members to such a level that it evolves into friendships that extend beyond the project structures. These friendships can result in the creation of certain social groupings within the project team that result in the alienation of other project team members, to the disadvantage of the project. Harnessing these friendships in the composition of new project teams can result in the creation of highly effective project teams that deliver beyond the sum of the capabilities of the individuals.

Furthermore, Leonard (2002) states that these relationships has a very intrigue nature. According to him they consist of two dimensions, namely a physical and abstract dimension. The physical dimension describes those elements that are necessary in order to enable contact between team members, whereas the abstract dimension describes the soft issues of a relationship. These two dimensions enable one to describe the holistic nature of such a relationship fully and to encapsulate the important elements of a support-oriented organization, namely mutuality, belonging, and connection. Because of the holistic nature of the different elements, Leonard (2002) argues that any kind of change having an effect on any of the elements of either the physical or abstract dimensions of a relationship will in fact disturb the relationship. Based on this, one can

argue that social relationships and networks that are disturbed needs to be managed in such a way that all role players stay focused and committed.

Advice

Individuals make use of their social networks outside of the project structures to gain advice to assist them in performing their tasks and to influence the project. Individuals with strong and influential social networks can be advantageous to project teams as these networks can be used by other team members and to provide validity to the project. These networks can assist greatly to identify possible external factors that could influence the project and develop associated actions to minimize potential negative impacts.

Survey Design and Results

The survey was distributed to 100 participants across the two cases with 58 valid responses received. As stated previously, the Likert Scale was used to measure the participants' views on the categories identified during the GT initial data collection phase.

The purpose of the survey was to:

- Gathering further information to saturate the categories identified during GT.
- Gathering personal views of participants in IT projects.

The survey illustrates clearly the huge role played by the establishment and use of social relationship and networks in a given IT project environment. In summary, the survey indicates the following phenomena:

- In total 87% of the respondents indicated that social relationships between team members developed within the project teams as well as between members of other teams. Almost two thirds indicated that

they developed stronger social relationships with some project team members than with other.

- Social relationships are used by team members within and around the structures created in the project environment.
- Project team members used their social relationships to solve problems whilst following project procedures as well as when the procedures proved to be a stumbling block in resolving the particular problem.
- A significant finding is that one in two respondents indicated that they made use of their social networks outside the organisation to gain knowledge so that they could influence the project.
- A large number of participants indicated that the team members make use of their social relationships to the benefit of the project, but to their individual benefit as well.

SUMMARY OF RESEARCH FINDINGS

Here follows a summary of the major findings.

The nature of IT projects normally requires high levels of team member interaction throughout the project life cycle and thus social relationships will develop between project team members. Two thirds of respondents indicated that they developed stronger social relationships with some team members than with others. This implies that social alignments or clicks can develop within project teams that must be monitored to prevent possible alienation of team members. Harnessing these groupings to the advantage of the project can provide momentum and energy toward successful project delivery. Some social relationships developed between project team members to such an extent that team members developed friendships that extended beyond the project environment. Although this phenomenon is on a personal level, this could also enhance or hamper performance which needs to be monitored.

Project team members use their social relationships primarily to solve problems and to gain some level of advantage. This could be regarded as a positive development in the sense that team members will, because of certain relationship, deal with their problems as quickly as possible and enhance the team performance as a whole.

External relationships and networks also play an important role. These social networks can be of a formal or informal nature. Formal networks include for example industry portals, special interest groups, past colleagues and information feeds, whilst informal networks include blogs, search engines and social networking tools. Individuals can also be used by other members in their social networks to influence the IT project. Influential business people can influence projects by accessing project team members directly and thus bypassing project structures. Such interventions must be monitored as it could negatively impact on the project direction or results.

In terms of the strong phenomena of the use of virtual teams all over the globe, management must do some risk management. They must identify those issues that are particularly possible to impact negatively on relationship development. While collaboration technologies have the capability of creating a communication environment for virtual partners who are separated by time and/or space, they may hinder the development of a strong sense of cohesion and satisfaction with the group's interaction process. Second, the strength of relational links is positively associated with the effectiveness of information exchange. Therefore, the loss of relationship building in virtual teams implies that the use of traditional meetings as a supplement to the use of electronic means might be useful (Warkentin *et al* (1997)).

SOLUTIONS AND RECOMMENDATIONS

This research identified impact areas as well as impact types of social relationships and net-

works on IT projects. These impact areas and types need to be considered when project teams are established as well as monitored throughout the project life cycle. Taking cognisance of the importance of social relationships and networks within IT projects can improve the management of technology and ultimately contribute to a greater success rate of IT projects.

The results of this research put a very important obligation on project managers' shoulders to take responsibility for these relationships and to "allow" them for the benefit of each project and as such for the organization. One can argue that such relationships enhance easier communication regarding issues that are normally not discussed in a more formal project environment. Here one can for example think of issues that need "whistle blowing." This brings us back to the importance of creating a project culture that should portray the following: "Instead of shooting the messenger, managers would do well to establish ...climate that encourages individuals to come forward with accurate project status information, regardless of whether the news is good or bad" (Smith & Keil, 2003). Therefore, social relationships between IT professionals of different teams as well as with other employees in the organisation are important to investigate on a deeper level. Furthermore, discussions between practitioners and academics should be encouraged to explore their impact on the normal operations of an IT department and ways to manage it in a professional way.

FUTURE RESEARCH DIRECTIONS

The impact of social relationships on project failure or success, is a very complex issue. This research has illustrated that these relationships are very important to team members and that they use it for different reasons of which the most important is to support them in solving problems. What is needed and unclear about these relationships is their impact and role with regard to project success or failure. One can for example argue that

such relationships enhance easier communication regarding issues that are normally not discussed in a more formal project environment. Here one can for example think of issues that needs "whistle blowing." This brings us back to the important issue of a project culture that should portray the following: "Instead of shooting the messenger, managers would do well to establish an organizational climate that encourages individuals to come forward with accurate project status information, regardless of whether the news is good or bad" (Smith & Keil, 2003). Therefore, social relationships between IT professionals of different teams as well as with other employees in the organisation are important to investigate on a deeper level.

CONCLUSION

The results of the research illustrate that project management philosophies and methodologies alone are not enough to achieve project success and that the social relationships and networks of project team members cannot be ignored. Each project team member contributes more than just their knowledge and skills to the project. Social relationships and networks will develop and evolve within IT project teams and need to be harnessed to the advantage of the project to improve the likelihood of a successful outcome.

The chapter emphasizes the important role of social relationships in the IT project environment. The results put a very important obligation on project managers' shoulder to take responsibility for these relationships and to allow it for the benefit of each project and as such for the organization.

REFERENCES

Ashworth, M. J., & Carley, K. M. (2006). Who You Know vs. What You Know: The Impact of Social Position and Knowledge on Team Performance. *The Journal of Mathematical Sociology*, 30(1), 43–75. doi:10.1080/00222500500323101

Gargiulo, M., & Benassi, M. (2000). Trapped in your own net? Network cohesion, structural holes and the adaptation of social capital. *Organization Science*, 11(2), 183–196. doi:10.1287/orsc.11.2.183.12514

Hovorka, D. S., & Larsen, K. R. (2006). Enabling agile adoption practices through network organizations. *European Journal of Information Systems*, 15(2), 159–168. doi:10.1057/palgrave.ejis.3000606

Kadefors, A. (2004). Trust in project relationships – inside the black box. *International Journal of Project Management*, 22(3), 178–182. doi:10.1016/S0263-7863(03)00031-0

Leonard A.C., (2002). A Conceptual Framework For Managing Relationships between all Participants During IT Service And Support Activities. *South African Journal of Industrial Engineering*, 81-96.

Liebowitz, J. (1999). A look at why information systems fail, Kybernetes. *The International Journal of Systems & Cybernetics*, 28(1), 61–67. doi:10.1108/03684929910253243

Reed, A. H., & Knight, L. V. (2010). Project Risk Differences Between Virtual and Co-Located Teams. *Journal of Computer Information Systems*, 19–30.

Reich, B.H., & Benbasit, I. (1999). *Factors that influence the social dimension of alignment between business and information technology objectives*. Society of Information Management (SIM) and the Management Information Systems Research Center (MISRC).

Sauer, C. (1993). *Why Information Systems Fail: A Case Study Approach*. New York, NY: Wiley.

Smith, H. J., & Keil, M. (2003). The reluctance to report bad news on troubled software projects: A theoretical model. *Information Systems Journal*, 13(1), 69–95. doi:10.1046/j.1365-2575.2003.00139.x

Standing, C. (1998). Myths and the art of deception in information systems. *Work Study*, 47(1), 5–13. doi:10.1108/00438029810196667

Standing, C., & Bavington, A. (1996). Implementing and using a structured information systems methodology. *Accountability and Performance*, 2(2), 51–68.

Warkentin, M. E., Sayeed, L., & Hightower, R. (1997). Virtual Teams versus Face-to-Face Teams: An Exploratory Study of a Web-based Conference System. *Decision Sciences*, 28(4), 1997. doi:10.1111/j.1540-5915.1997.tb01338.x

ADDITIONAL READING

Avgerou, C., & McGrath, K. (2007). Power, Rationality, and the Art of Living through Socio-Technical Change. *Management Information Systems Quarterly*, 31(2), 295–315.

Baerentsen, K. B., & Trettvik, J. (2002). *An Activity Theory Approach to Affordance*. NordiCHI. doi:10.1145/572020.572028

Barczak, G., & Wilemon, D. (2001). Factors influencing product development team satisfaction. *European Journal of Innovation Management*, 4(1), 32–36. doi:10.1108/14601060110365556

Beise, C. M. (2004). IT Project Management and Virtual Teams, *Proceedings of ACM SIGMIS'04*, Tucson Arizona, 129-133.

Berg, M. (1998). The Politics of Technology: On Bringing Social Theory into Technological Design. *Science, Technology & Human Values*, 23(4), 456–490. doi:10.1177/016224399802300406

Biggs, S., & Smith, S. (2003). A Paradox of Learning in Project Cycle Management and the Role of Organizational Culture. *World Development*, 31(10), 1743–1757. doi:10.1016/S0305-750X(03)00143-8

- Bresnen, M., Edelman, L., Newell, S., Scarbrough, H., & Swan, J. (2003). Social Practices and management of knowledge in project environments. *International Journal of Project Management*, 21(3), 157–166. doi:10.1016/S0263-7863(02)00090-X
- Carrol, P., Ridley, G., & Young, J. (2004). COBIT and its utilization: a framework from the literature. *System Sciences*, 233–240.
- Chaudhuri, N., & Vainshtein, N. (2007). Agile Project Management: PMBOK vs. Agile, Agile Project Leadership Network Washington DC Chapter Meeting. Retrieved on September 5, 2008 from http://aplndc.pbwiki.com/f/PMBOK_and_Agile.pdf
- Checchi, R. M., Sevcik, G. R., Loch, K. D., & Straub, D. D. (2002). *An Instrumentation Process for Measuring ICT Policies and Culture*. National Science Foundation.
- Cockburn, A. (2000). Agile Software Development, The Agile Software Development Series. Retrieved on September 7, 2008 from <http://nkcho.isp.st/article/AgileSwDevDraft3.pdf>
- Corbin, J., & Strauss, A. (1990). Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative Sociology*, 13(1), 3–21. doi:10.1007/BF00988593
- Esteves, J., Ramos, I., & Carvalho, J. A. (2002). Use of Grounded Theory in Information Systems Area: An Exploratory Analysis. In *Proceedings of European Conference on Research Methodologies on Business Management ERCM 2002* (pp. 129–136). Reading, UK.
- Glazer, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory*. Chicago: Aldine.
- Guzman, I. R., Stam, K. R., & Stanton, J. M. (2008). The Occupational Culture of IS/IT Personnel within Organizations. *The Data Base for Advances in Information Systems*, 39(1), 33–50. doi:10.1145/1341971.1341976
- Jackson, P., & Klobas, J. (2008). Building knowledge in projects: A practical application of social constructivism to information systems development. *International Journal of Project Management*, 26(4), 329–337. doi:10.1016/j.ijproman.2007.05.011
- King, J. (2002). Back to Basics. *Computerworld*, 36(17), 36–37.
- Lemon, M., & Sahota, P. S. (2004). Organizational culture as a knowledge repository for increased innovative capacity. *Technovation*, 24(6), 483–498. doi:10.1016/S0166-4972(02)00102-5
- McMichael, H. (1999). An Activity Based Perspective for Information Systems Research. *Proceedings of the 10th Australasian Conference on Information Systems*, 610–621.
- Motjolopane, I., & Brown, I. (2004). Strategic Business-IT Alignment, and Factors of Influence: A Case Study in a Public Tertiary Educational Institution. In *Proceedings SAICSIT 2004* (pp. 147–156).
- Nyffjord, J., & Kajko-Mattsson, M. (2007). Commonalities in Risk Management and Agile Process Models. In *Proceedings of the International Conference on Software Engineering Advances, ICSEA* (pp. 25–31). doi:10.1109/ICSEA.2007.22
- Pandit, N. R. (1996). The Creation of Theory: A Recent Application of the Grounded Theory Method. *Qualitative Report*, 2(4). <http://www.nova.edu/ssss/QR/QR2-4/pandit.html> Retrieved on August 10, 2008
- Project Management Institute. (2004). *A guide to the Project Management Body of Knowledge* (3rd ed.). New York: PMI Publishing Division.
- Rennie, D. L. (1998). Grounded Theory Methodology: The pressing need for coherent logic of justification. *Theory & Psychology*, 8(1), 101–119. doi:10.1177/0959354398081006

Rising, L., & Janoff, N. S. (2000). The Scrum software development process for small teams. *IEEE Software*, 17(4), 26–32. doi:10.1109/52.854065

KEY TERMS AND DEFINITIONS

Grounded Theory: Although this kind of research approach is complex; one can briefly say that it means to develop your theory based on what you find in the existing data. The theory is therefore “grounded” in the data and this is called an inductive process.

Information Systems: Systems that are used by organisations to support them in decision making. These systems provide information to all levels of management in an organisation. These systems are normally the end result of an IT project.

Information Technology (IT) Projects: Projects that are launched to develop an information and communication technology system or to provide any kind of ICT service to an organisation based on a specific need.

Project Culture: The way people in a specific project think and apply their values. It impacts

on their behaviour and how they perform their responsibilities during the project life cycle. A project culture is normally based on the belief (value) system of a group of people or that of the organisation.

Project Team: A team of people that are responsible to do all the work that is necessary to take a project from its initiation until it is completed.

Social Networks: The group of people or staff that are involved in a relationship for a specific reason. For example, a social network of people having a discussion each week to discuss technical issues regarding a specific project.

Social Relationships: The relationships (professional or personal) that are established during the project life cycle between employees of different teams or the same team or other parts of the organisation.

Virtual Teams: Virtual project teams, by definition, are groups of people working together toward a common goal. Whether, they are called virtual or distributed, the team members are not co-located; they can reside in different cities, states or countries.

Viewpoints on Business Process Models

Giorgio Bruno

Politecnico di Torino, Italy

INTRODUCTION

A business process is a standard way of organizing work in a business context (Rummler & Brache, 1995). Business processes cross functional boundaries in that they involve members of different departments; common examples are developing a new product, ordering goods from a supplier, and processing and paying an insurance claim (Davenport & Short, 1990).

Over the past few years a number of viewpoints have influenced the design of notations for business processes (Bruno, 2011). They emphasize the different elements (tasks, business entities and roles) that compose business process models; for this reason, they are referred to as activity-centric, data-centric, and role-centric viewpoints.

Probably, the most popular viewpoint is the activity-centric one, whose standard representative is BPMN (Business Process Model and Notation) (OMG-BPMN, 2011). It considers business processes essentially as orchestrators of operational activities, which encompass human tasks and automatic ones. Human tasks are carried out by participants by means of graphical interfaces, while automatic tasks are implemented with services. Orchestration is achieved through control flow elements, which enforce rigid precedence relationships: in this context, a business process is like a master distributing the work among the participants in the process.

When processes are meant to operate on the entities which form an information system, whose purpose is to enable human participants and/or machines to perform work using information (Alter 2008), focusing on the identification and the ordering of the activities may not be the right way to start with an investigation of the intended business.

The alternative solution proposed by data-centric approaches is to begin with the identification of the key business entities and their life cycles consisting of states and transitions (Hull, 2008). The term artifact has been introduced to designate a concrete and self-describing chunk of information used to run a business (Nigam & Caswell, 2003). The artifact types come from experience and are associated with business goals; the analysis of how to progress towards the goals determines the definition of their life cycles (Battacharya et al., 2005). The major benefit is the right level of granularity, which facilitates communication among the stakeholders and helps them focus on the primary purposes of the business (Chao et al., 2009).

An extension to data-centric approaches is the case-centric viewpoint, whose major purpose is to support knowledge workers in applications requiring the flexibility that cannot be provided by the approaches based on a rigid control flow (Marin, Hull and Vaculín, 2013). Flexibility implies that the participants in the process are no longer considered as mere resources needed to carry out tasks that are not automatable, but they can be involved in a number of choices, such as the selection of the input entities when a task needs more than one and the selection of the task with which to handle the input entities when two or more tasks are admissible (Bruno, 2014).

Human tasks are associated with roles, which represent the participants involved in the processes. However, if the process model shows the activity flows of the roles in a single view, it may be difficult for a participant to understand what their actual involvement is. The participation of users in processes is made more evident if the process is decomposed into several “role” components:

DOI: 10.4018/978-1-5225-2255-3.ch068

this is the essence of the role-centric viewpoint (Ould, 2005).

This article illustrates how the above-mentioned viewpoints can be extracted from a global model, which is based on a notation named iBPN (integrated Business Process Notation). This notation is inspired by high-level Petri nets so as to take advantage of their ability of being state and action oriented at the same time (Jensen, 1997). An example concerning the submission of papers to conferences will be used to illustrate the construction of the global process model and then the extraction of the data-centric viewpoint and the role-centric one.

This article is organized as follows. First it presents background information and the informal requirements of the example to be worked out. Then it illustrates the key modeling issues, the process model related to the example, and the automatic extraction of the above mentioned viewpoints; finally it makes some considerations on future research directions and provides the conclusion.

BACKGROUND

The data-centric orientation is put into practice in two major ways depending on whether the life cycles are defined separately or are combined in a single model. In the first case, various techniques have been proposed to handle the interactions between the life cycles: events and rules in the Guard-Stage-Milestone approach (Hull et al., 2001), macro processes in PHILharmonicFlows (Künzle & Reichert, 2011), hierarchical relationships in COREPRO (Müller, Reichert & Herbst, 2007), and messages in Proclets (van der Aalst, Barthelmess, Ellis & Wainer, 2000). In the second case, the combination of the life cycles gives rise to a data flow which may play a part on the activation of the tasks. The data flow may be the only responsible for the activation of the tasks, as it happens with the iBPN notation presented in this article, or it may act as a complementary

cause with respect to the control flow. The two approaches are referred to as dataflow-centric and extended activity-centric, respectively.

BPMN has been extended with dataflow-oriented features, such as data objects; however, the coexistence of the control flow and the data flow makes it difficult to handle the situations where a many-to-many mapping between entities of different types is needed. An example related to requisition orders and procurement orders is discussed by Meyer, Pufahl, Fahland & Weske (2013).

Extended activity-centric models include all the states of the artifacts involved; for this reason, the life cycles (made up of states and state transitions) can be extracted automatically. The main purpose is to verify the compliance with reference life cycles. Eshuis and Van Gorp (2014) present the extraction of life cycles from UML activity diagrams (OMG-UML, 2015), while the extraction from BPMN models is addressed by Cabanillas, Resinas, Ruiz-Cortés and Awad (2011). The opposite direction, i.e., the generation of process models from life cycles, has also been investigated (Küster, Ryndina & Gall, 2007). Roundtrip transformations between data-centric models and extended activity-centric ones are described by Meyer and Weske (2014).

The Guard-Stage-Milestone approach provided a good foundation for the recent standard CMMN (Case Management Model and Notation) (OMG-CMMN, 2014). Case management involves taking actions regarding a subject in order to achieve a desired outcome. The important issue is that the actions cannot be anticipated in a predefined sequence of tasks. In fact, the case workers may perform the run-time planning of the tasks to be carried out. A short history of the key concepts of CMMN is provided by Marin, Hull and Vaculín (2013). Various flavors of case management are illustrated by Marin, Hauder and Matthes (2015).

Human choices have been addressed by research on flexible processes based on declarative approaches. For example, in Declare any task of an ad-hoc process may be performed as long as

the mandatory constraints are not violated (van der Aalst, Pesic & Schonenberg, 2009).

With the role-centric viewpoint, business processes result from the composition of a number of role components, which are structured as processes: in addition to the tasks pertaining to the corresponding role, they include control flow elements and send/receive operations. The latter are needed to coordinate the various role components involved in the process. Examples of role-centric notations are Role Activity Diagrams (RADs) (Ould, 2005) and Subject-oriented Business Process Management (S-BPM) (Fleischmann and Stay, 2012).

INFORMAL REQUIREMENTS OF THE EXAMPLE

The process under consideration is called Handle-Submissions and is meant to handle the submission of papers to a conference. An informal definition of its requirements is as follows:

1. When the process is started, the chairman has already been appointed and the reviewers have already been identified. In addition, four deadlines, d1, d2, d3 and d4, have already been established; d1 is the end of the submission period, d2 is the end of the reviewing period, d3 is the term for notifying the acceptance or rejection of the papers, d4 is the term for submitting the final versions of the accepted papers.
2. After the start of the process, authors may submit papers before deadline d1; in the same period, they may also update their papers or remove them.
3. At the end of the submission phase, the chairman is enabled to assign the papers to the reviewers; the purpose is to get three reviews for each paper. The chairman produces three assignments for each paper: an assignment is related to one paper and one reviewer and

includes the deadline for completing the review.

4. In case an assignment has not been fulfilled in due time, the chairman produces an additional assignment, which is related to the same paper and directed to a different reviewer.
5. The papers become ready for evaluation when the related assignments have been fulfilled. The chairman decides the acceptance or rejection of the papers before deadline d3.
6. The authors of the accepted papers should provide the final versions before deadline d4; if they do not, the papers are considered withdrawn.

KEY MODELING ISSUES

With the iBPN notation, a business process model is made up of two interrelated models: the dataflow model and the information one. The information model is a UML class diagram (OMG-UML, 2015), which defines the structure (types, attributes and relationships) of the business entities of the domain. The dataflow model consists of tasks which operate on business entities whose types have been defined in the information model. The tight integration between the tasks and the business entities, which is the major purpose of iBPN, is obtained because the behavior of the tasks is specified in terms of constraints and intended effects, which are expressed in a declarative manner with rules and conditions leveraging the elements of the information model. The language used is a simplified version of the Object Constraint Language (OCL) (OMG-OCL, 2014).

Entity Life Cycles

Each entity follows a path (called life cycle) made up of states, from the initial state to a final one. States denote stages of progress through the process and the transitions between states are carried

out by the tasks. In the dataflow model, states are represented by circles called places. Places have two labels: the type name and the state name, e.g., “Paper, submitted”. If the state label is missing, the state is the initial one; final states are denoted by places having no output transitions.

Process Instances and Contextual Entities

When a process is instantiated, the new instance receives a business entity called contextual entity, which provides background information in terms of attributes and associations with other entities. Moreover, all the entities generated during the execution of the instance will be related to the contextual entity, either directly or indirectly. The type of the contextual entity is written between parentheses after the process name, e.g., “process HandleSubmissions (Conference)”.

Mandatory Attributes and Relationships

When a new entity is generated, a number of mandatory attributes and associations with other entities have to be set; they are specified in the information model as follows. Mandatory attributes are qualified by an exclamation mark while mandatory associations correspond to oriented relationships (the source entity type denotes the entities for which the associations are mandatory).

Tasks

A task may be a human one or an automatic one; in the first case, it is depicted as a rectangle with rounded corners, in the second case as an ordinary rectangle. Tasks take the input entities from the input places and deliver the output entities to the output places. If a task generates an output entity, the output link is decorated with the label “new” and the output place corresponds to the initial state

of the entity. Such a task is called generative task for that entity type. The intended effects of tasks are expressed in a declarative manner by means of post-conditions, which are based on the elements (attributes and relationships) of the information model. Restrictions to the execution of tasks are expressed with pre-conditions.

Weights of the Arcs

The weights of the arcs indicate how many input entities are taken from the input places and how many output entities are added to the output places. The weights are integer numbers shown close to the attachments of the arcs to the tasks. If the weight is 1, it is omitted.

Timing Constraints

Tasks may be annotated with timing constraints, such as the “before” clause for human tasks and the “at” clause for automatic tasks. The “before” clause specifies the deadline of the task and the “at” clause the starting time. The actual value is usually obtained from the attributes of the entities visible to the task. The process may be notified when a task is not performed in due time so as to take corrective actions; this feature will be illustrated in the next section.

Performers of the Tasks

Tasks are placed in swim lanes, which are associated with roles: a human task will be assigned to a performer playing the role specified by the swim lane it belongs to. The actual performers may be generic actors or specific ones. A generic actor denotes any actor entitled to play the role needed for the task; on the contrary, a specific actor is an actor who bears a connection to the input entities of the task. For example, any author may enter a paper, but updating a paper is permitted only to the author.

THE PROCESS MODEL AND ITS INTERPRETATION

This section illustrates the model of process *HandleSubmissions*, which results from the above mentioned requirements; it is shown in Figure 1.

Requirement 1 introduces the mandatory properties – the four deadlines and the relationships *Conference-Chairman* and *Conference-Reviewer* – of the contextual entity of the process. Tasks *enterPaper*, *updatePaper* and *removePaper* fulfill requirement 2, and they are all optional.

Requirement 3 is dealt with by tasks *submitPaper* and *assignPaper*. Task *submitPaper* is an automatic one: it is performed at the end of the submission phase and makes all the papers in the initial state leave this state and enter the “submitted” state.

Task *assignPaper* enables the chairman to assign papers to reviewers. The effect of the task is to generate three assignment entities for each paper entity taken from the input place. An assignment entity has to be related with one paper entity and one reviewer entity. For this reason, the mandatory relationships *Assignment-Paper* and *Assignment-Reviewer* are included in the information model. The actual paper entity is the input entity, while the reviewer entity will be decided during the execution of the task. However, the invariants “*paper.assignmentsreviewer in paper.conference.reviewers*” and “*paper.assignmentsreviewer distinct*” guarantee that the reviewers will be selected from among those associated with the conference and that no reviewer will receive the same paper more than once. Task *assignPaper* also sets the mandatory deadlines (attribute *d*) of the newly generated assignments and changes the state of the input papers from “submitted” to “assigned”. The deadline of the assignments is decided by the chairman but the local invariant “*before paper.conference.d2*”, which is written between parentheses after the attribute name in the information model, establishes the upper

limit of the actual deadline. The term “paper” in the invariant denotes the paper the assignment is related to.

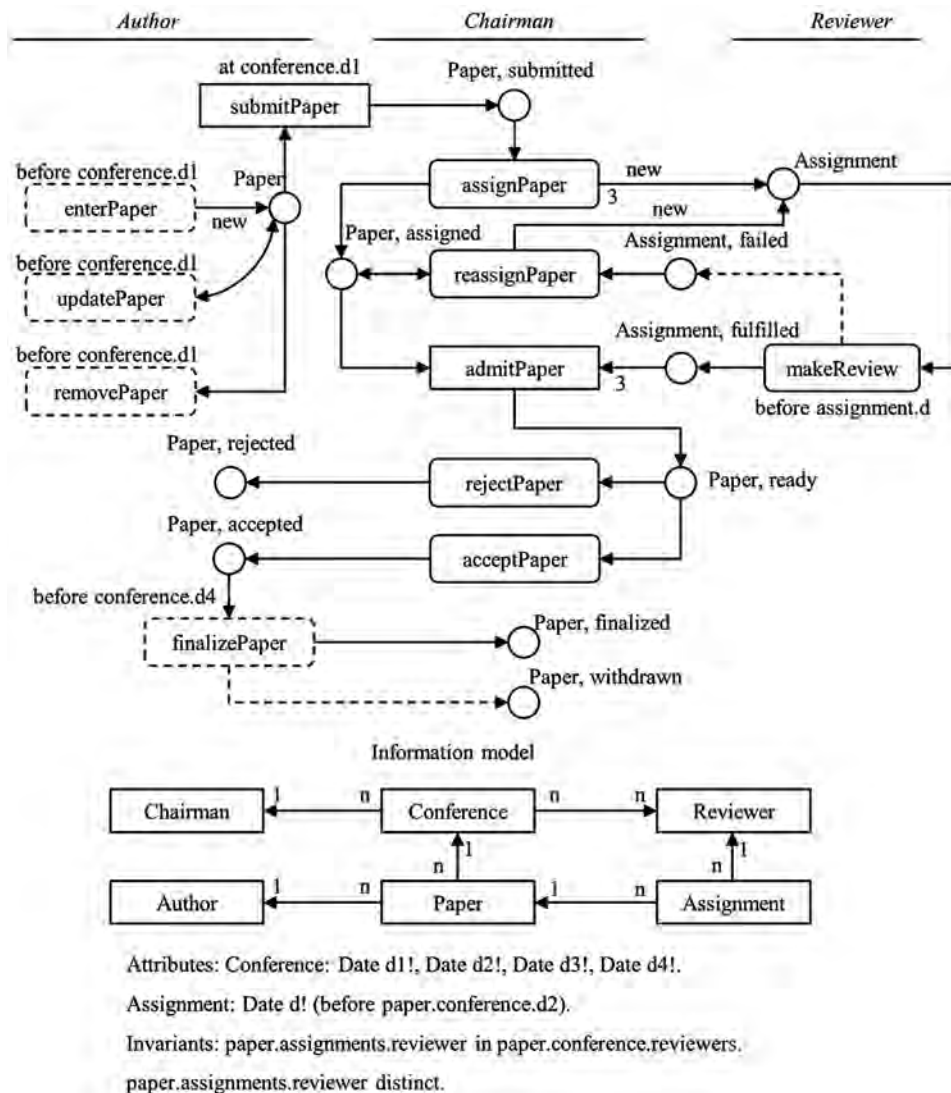
Requirement 4 brings about a task (“*makeReview*”) for the reviewer role. For the sake of simplicity, the review provided by the reviewer is not represented explicitly. In case the reviewer does not provide the expected review, the process needs to be notified so as to enable the chairman to make an additional assignment to a different reviewer. Such notifications, called failure notifications, are carried out by means of failure links (shown dashed). If task *makeReview* fails, the input assignment is moved into place (*Assignment, failed*), which is an input place of task *reassignPaper*.

Requirement 5 leads to tasks *admitPaper*, *acceptPaper* and *rejectPaper*. Automatic task *admitPaper* moves an assigned paper to the ready state when there are three fulfilled assignments available for it. The other two tasks change the state of a ready paper into “accepted” or “rejected”, respectively.

Task *finalizePaper* is an optional task enabling the authors of the accepted papers to submit the final versions, according to requirement 6. If an accepted paper is not finalized in due time, the paper is moved to the “withdrawn” state through the default link (shown dashed). A default link is used to inform the process that an optional task has not been performed; default links and failure ones have similar purposes.

In process *HandleSubmissions* there are two cases of human task selection: this kind of selection takes place when two or more human tasks have some or all of the input entities in common. The first choice is between task *updatePaper* and task *removePaper*: this is an optional choice in that the tasks are optional ones. On the contrary, the choice between task *acceptPaper* and task *rejectPaper* is a mandatory one: for each paper the chairman has to decide whether to accept or reject it.

Figure 1. The model of process HandleSubmissions (Conference)



EXTRACTING DATA-CENTRIC AND ROLE-CENTRIC VIEWPOINTS FROM IBPN MODELS

This section shows how the data-centric viewpoint and the role-centric one can automatically be extracted from a dataflow model and presents the results obtained from process HandleSubmissions.

Data-Centric Models

The data-centric viewpoint shows the life cycles of the entity types involved in the process: it may

help analysts check whether the resulting life cycles match those coming from experience or comply with those expected by the organization in charge of running the process.

The dataflow model of process HandleSubmissions deals with two entity types, i.e., Paper and Assignment. Their life cycles are shown in Figure 2. The entity life cycles can be extracted on the basis of a simple rule: for each entity type, say, T, a life cycle model is obtained by first introducing a state for each place of type T and then connecting the states with arcs corresponding to the tasks in the dataflow model. For example, if task

t connects place (T, s1) to place (T, s2), then an arc labeled with t is added from state s1 to state s2. If there is no output place of the same type, as it happens with task RemovePaper, the input state is connected to an implicit final state named “final”. An arc is dashed if the corresponding task is optional.

Generative tasks appear on initial arcs, an initial arc being an arc that has no source state and ends in the initial state. Dashed arcs without task labels, such as initial-failed in the Assignment life cycle and accepted-withdrawn in the Paper life cycle, correspond to failure or default links in the dataflow model. Tasks are prefixed by the initials of the corresponding roles, i.e., A for Author, C for Chairman and R for Reviewer. For the sake of simplicity timing constraints have been omitted.

Role-Centric Models

In the role-centric viewpoint, each role model provides a restricted view of the process limited to the behavior of the role under consideration. This is particularly important when actors external to the organization running the process are involved.

The role-centric models for process HandleSubmissions are shown in Figure 3(a).

Role models derive from swim lanes and have places in common: such places (e.g., “Paper, submitted”) are called inter-role places. A context model, such as the one shown in Figure 3(b), is then provided so as to underline the connections between role models in terms of inter-role places.

FUTURE RESEARCH DIRECTIONS

Given that activity-centric notations such as BPMN provide a mature solution to highly repetitive processes (routines), the focus of current research has shifted to flexible processes in which human choices play a fundamental role.

The new case-centric standard CMMN allows tasks to be selected on the basis of the state of the case being considered; in addition, decisions may be triggered by events such as the availability of a new document, the completion of a task, and the achievement of a milestone. A CMMN model is not a full-fledged data-centric model in that it does not show the entity types involved: this is a

Figure 2. The entity life cycles extracted from the dataflow model

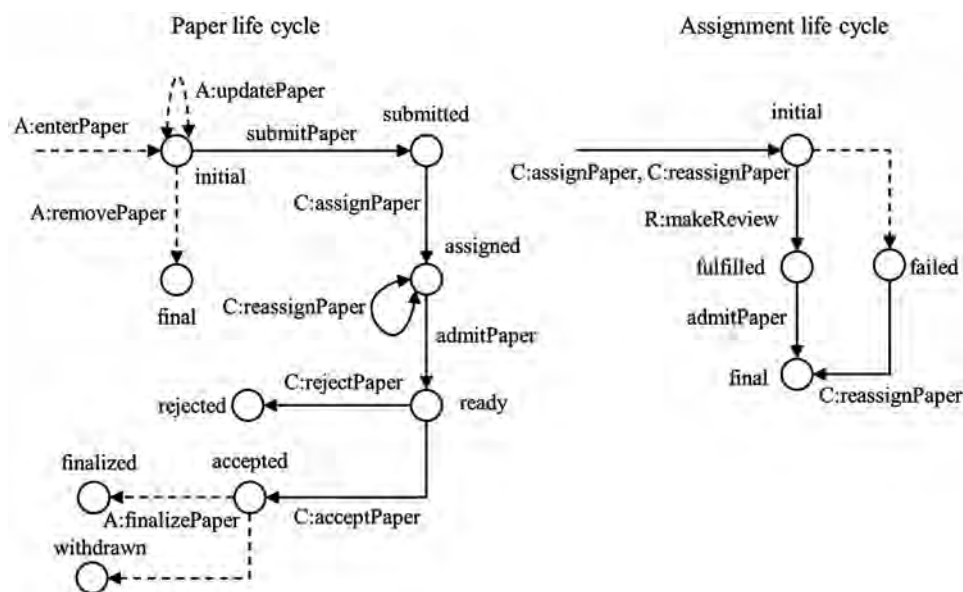
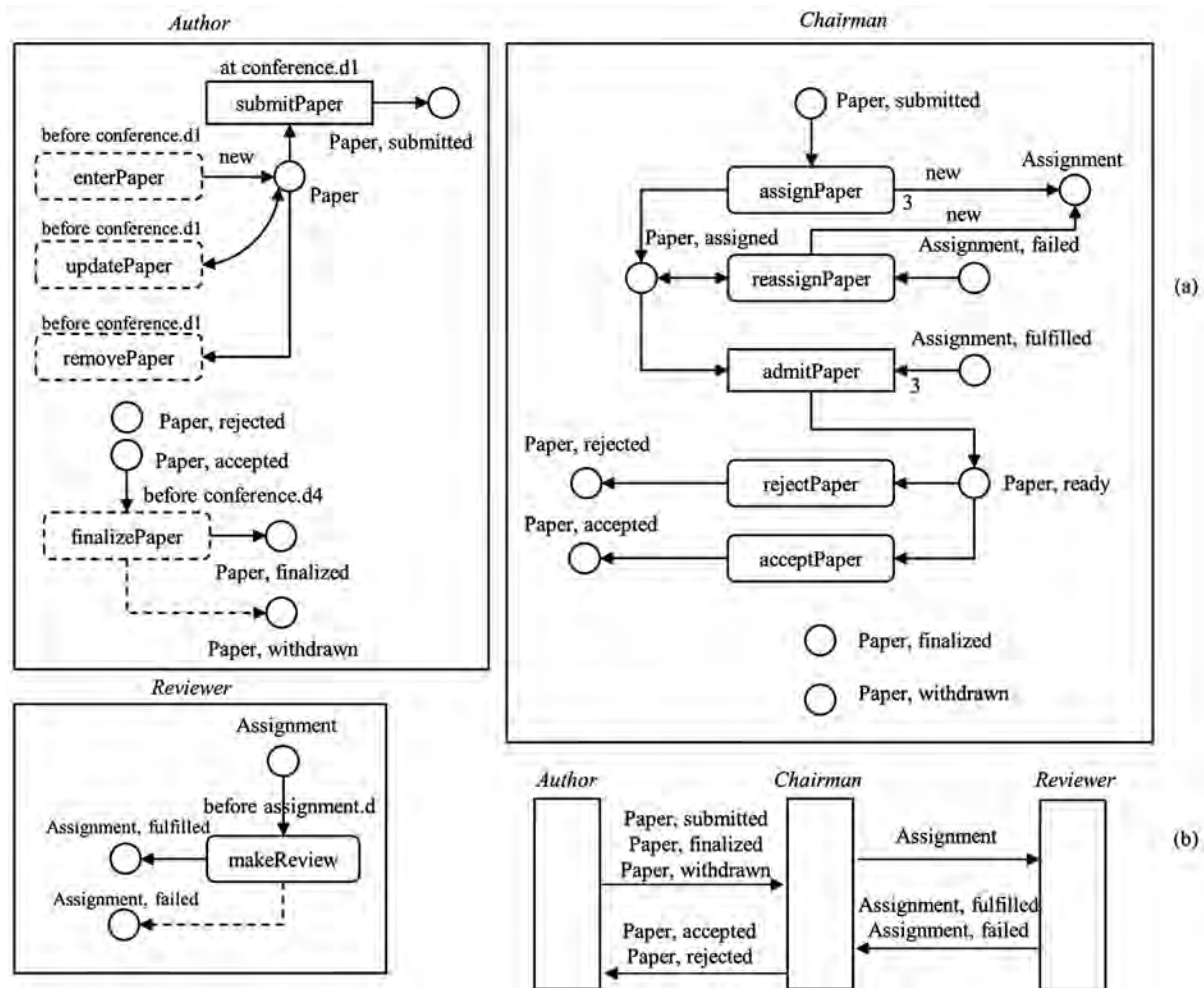


Figure 3. The role-centric models extracted from the dataflow model (a) and the context model (b)



major difference from the Guard-Stage-Milestone approach, which may be considered its precursor. In the new standard, the collection of data related to the case is found in a case file, which is accessible by all the tasks. However, the data flow related to the tasks is not evident from the model. In the process examined in this article, the case is a conference but several tasks operate on papers; therefore showing the connections between business entities and tasks will improve the understanding of the model. For this reason, a more articulated view of the case data may be advisable: this could be a relevant direction of development.

CONCLUSION

Business processes address complex systems resulting from the integration of several aspects, i.e., activities, roles, business entities, control flow rules and data flow rules. For this reason, a single model might be too complex and then its decomposition driven by specific viewpoints may help improve its readability and understanding.

The major contribution of this article has been to illustrate the key modeling issues of business processes and to present three forms of representation. In the first one, i.e., the global model, all the tasks appear in the same model; in the others, they

are distributed on the basis of two criteria: the entity life cycles and the roles associated with the tasks. These three forms address the dynamics aspects of the process and rely on a common information model providing the attributes and relationships of the business entities involved in the process. Current work is dedicated to the formalization of the iBPN notation and to the implementation of a suitable process engine.

REFERENCES

- Alter, S. (2008). Defining information systems as work systems: Implications for the IS field. *European Journal of Information Systems*, 17(5), 448–469. doi:10.1057/ejis.2008.37
- Bhattacharya, K., Guttman, R., Lyman, K., Heath, F. F. III, Kumaran, S., Nandi, P., & Staudt, A. et al. (2005). A model-driven approach to industrializing discovery processes in pharmaceutical research. *IBM Systems Journal*, 44(1), 145–162. doi:10.1147/sj.441.0145
- Bruno, G. (2011). Business agility and process agility: how do they relate to each other? In M. M. Cruz-Cunha & J. Varajão (Eds.), *E-business managerial aspects, solutions and case studies* (pp. 71–89). Hershey, PA: Information Science Publishing. doi:10.4018/978-1-60960-463-9.ch005
- Bruno, G. (2014). Combining flexibility and data handling in business process models. *International Journal of Human Capital and Information Technology Professionals*, 5(2), 14–27. doi:10.4018/ijhcritp.2014040102
- Cabanillas, C., Resinas, M., Ruiz-Cortés, A., & Awad, A. (2011). Lecture Notes in Computer Science: Vol. 6741. *Automatic generation of a data-centered view of business processes* (pp. 352–366). Heidelberg: Springer. doi:10.1007/978-3-642-21640-4_27
- Chao, T. et al.. (2009). Lecture Notes in Computer Science: Vol. 5701. *Artifact-based transformation of IBM Global Financing* (pp. 261–277). Heidelberg: Springer.
- Davenport, T. H., & Short, J. E. (1990). The new industrial engineering: Information technology and business process redesign. *Sloan Management Review*, (Summer), 11–27.
- Eshuis, R., & Van Gorp, P. (2012). Lecture Notes in Computer Science: Vol. 7532. *Synthesizing object life cycles from business process models* (pp. 307–320). Heidelberg: Springer. doi:10.1007/978-3-642-34002-4_24
- Fleischmann, A., & Stary, C. (2012). Whom to talk to? A stakeholder perspective on business process development. *Universal Access in the Information Society*, 11(2), 125–150. doi:10.1007/s10209-011-0236-x
- Hull, R. (2008). Lecture Notes in Computer Science: Vol. 5332. *Artifact-centric business process models: brief survey of research results and challenges* (pp. 1152–1163). Heidelberg: Springer.
- Hull, R. et al.. (2011). Lecture Notes in Computer Science: Vol. 6551. *Introducing the Guard-Stage-Milestone approach for specifying business entity lifecycles* (pp. 1–24). Heidelberg: Springer. doi:10.1007/978-3-642-19589-1_1
- Jensen, K. (1997). *Petri nets: basic concepts, analysis methods and practical use, 1*. Berlin: Springer.
- Künzle, V., & Reichert, M. (2011). PHILharmonicFlows: Towards a framework for object-aware process management. *Journal of Software Maintenance and Evolution: Research and Practice*, 23(4), 205–244. doi:10.1002/smr.524
- Küster, J. M., Ryndina, K., & Gall, H. (2007). Lecture Notes in Computer Science: Vol. 4714. *Generation of business process models for object life cycle compliance* (pp. 165–181). Heidelberg: Springer. doi:10.1007/978-3-540-75183-0_13

Marin, M., Hauder, M., & Matthes, F. (2015). Case management: an evaluation of existing approaches for knowledge-intensive processes. In *4th International Workshop on Adaptive Case Management and other non-workflow approaches to BPM*. Heidelberg: Springer

Marin, M., Hull, R., & Vaculín, R. (2013). Data centric BPM and the emerging case management standard: a short survey. In *Lecture Notes in Business Information Processing*, 132 (pp. 24–30). Heidelberg: Springer. doi:10.1007/978-3-642-36285-9_4

Meyer, A., Pufahl, L., Fahland, D., & Weske, M. (2013). Lecture Notes in Computer Science: Vol. 8094. *Modeling and enacting complex data dependencies in business processes* (pp. 171–186). Heidelberg: Springer. doi:10.1007/978-3-642-40176-3_14

Meyer, A., & Weske, M. (2014). Activity-centric and artifact-centric process model roundtrip. In *Lecture Notes in Business Information Processing*, 171 (pp. 167–181). Heidelberg: Springer.

Müller, D., Reichert, M., & Herbst, J. (2007). Lecture Notes in Computer Science: Vol. 4803. *Data-driven modeling and coordination of large process structures* (pp. 131–149). Heidelberg: Springer; doi:10.1007/978-3-540-76848-7_10

Nigam, A., & Caswell, N. S. (2003). Business artifacts: An approach to operational specification. *IBM Systems Journal*, 42(3), 428–445. doi:10.1147/sj.423.0428

OMG-BPMN. (2011). *Business Process Model and Notation*, V.2.0. Retrieved on October 2, 2015, from <http://www.omg.org/spec/BPMN/>

OMG-CMMN. (2014). *Case Management Model and Notation*, V.1.0. Retrieved on October 2, 2015, from <http://www.omg.org/spec/CMMN/>

OMG-OCL. (2014). *Object Constraint Language*, V.2.4. Retrieved on October 2, 2015, from <http://www.omg.org/spec/OCL/>

OMG-UML. (2015). *Unified Modeling Language*, V.2.5. Retrieved on October 2, 2015, from http://www.omg.org/spec/UML/

Ould, M. (2005). *Business Process Management: a rigorous approach*. The British Computer Society.

Rummler, G. A., & Brache, A. P. (1995). *Improving Performance: how to manage the white space on the organizational chart*. San Francisco: Jossey-Bass.

van der Aalst, W. M. P., Barthelmess, P., Ellis, C. A., & Wainer, J. (2000). Lecture Notes in Computer Science: Vol. 1901. *Workflow modeling using Proclefs* (pp. 198–209). Heidelberg: Springer.

van der Aalst, W. M. P., Pesic, M., & Schonenberg, H. (2009). Declarative workflows: Balancing between flexibility and support. *Computer Science. Research for Development*, 23(2), 99–113.

ADDITIONAL READING

Di Ciccio, C., Marrella, A., & Russo, A. (2015). Knowledge-intensive processes: Characteristics, requirements and analysis of contemporary approaches. *Journal on Data Semantics*, 4(1), 29–57. doi:10.1007/s13740-014-0038-4

Dumas, M., La Rosa, M., Mendling, J., & Reijers, H. (2013). *Fundamentals of Business Process Management*. Heidelberg: Springer. doi:10.1007/978-3-642-33143-5

Dumas, M., van der Aalst, W. M. P., & ter Hofstede, A. H. M. (2005). *Process-Aware Information Systems: bridging people and software through process technology*. New York: Wiley; doi:10.1002/0471741442

van der Aalst, W. M. P., Weske, M., & Grünbauer, D. (2005). Case handling: A new paradigm for business process support. *Data & Knowledge Engineering*, 53(2), 129–162. doi:10.1016/j.datak.2004.07.003

zur Muehlen, M., & Indulska, M. (2010). Modeling languages for business processes and business rules: A representational analysis. *Information Systems*, 35(4), 379–390. doi:10.1016/j.is.2009.02.006

KEY TERMS AND DEFINITIONS

Business Process: A standard way of organizing work in a business context. It consists of a number of process actions to be performed by a process engine.

Case Management: A flexible approach for taking care of a subject in order to achieve a desired outcome. The actions cannot be anticipated in a predefined sequence of tasks: the case workers

may perform the run-time planning of the tasks to be carried out.

Data Flow: The information items used and produced by the tasks.

Data Selection: A situation in which a performer has to select the input data before performing the task.

Performer: For a given task, the user in charge of performing it.

Task Selection: A situation in which a performer may operate on the same input data with a number of alternative tasks.

Task: A unit of work in the system.

Category B

Business Education

Serious Games in Entrepreneurship Education

Fernando Almeida

Polytechnic Institute of Gaya, Portugal

Jorge Simões

ISPGaya, Portugal

INTRODUCTION

Computer games are currently often used in different serious applications and especially in the terms of vocational education and training. In fact, serious games (SG) are used in a wide variety of areas, in several aspects of common life, such as for health, politics, advertising, project management, and virtual reality, amongst others. The intention of serious games is based on the idea to offer a balanced possibility for an authentic and amusing learning. Currently the teaching community is progressively discovering how game based learning supports the personalized trainings and gives new instruments for teaching basic key competences.

At the same time, one of the main challenges of educational programmes is to ensure that education delivers the right skills for the labor market and the growth of entrepreneurship, while delivering support to young people to secure their economic future and enable businesses to grow and create new jobs. Actually, entrepreneurship is considered as the most common powerful economic force across the globe (Gwija et al., 2014). Entrepreneurship education has the capability to benefit students from all socioeconomic backgrounds because it guides students to think outside the box and nurtures unconventional talents and skills. Furthermore, it fosters innovation, ensures social justice, encourages confidence and stimulates the economy.

Serious games may be considered powerful tools to sustain entrepreneurship in the context of the emerging paradigm of Technology Enhanced Learning (TEL). Serious games combine simultaneously instruction and gameplay, by challenging and involving players in motivating learning contexts. They also offer students a genuine “situated” learning experience and can concretely support the “learning by doing” approach, considering that simulations offer a more pragmatic experience and provide a safe environment in which immersive entrepreneurs could test their own business.

This study aims to identify and synthesize the main key benefits, but also challenges and issues, provided by the introduction of serious games in the learning process. For that, we initially perform a revision of literature in the field of simulations and serious games initiatives applied to entrepreneurship domain. In this context we briefly present and describe five serious games initiatives in entrepreneurship education. After that, we identified and discussed the main key benefits and challenges created by the introduction of serious games. Finally, we look for current and emergent future research directions and we draw the conclusions of our work.

BACKGROUND

According to Belloti et al. (2014), entrepreneurship is “a personal skill and motivation which

draws a person to engage his abilities and efforts in the creation of new products and services". The European Commission defines entrepreneurship as "an individual's creative capacity, independently or within an organization, to identify an opportunity and to pursue it in order to produce new value or economic success" (Carvalho et al., 2012). Entrepreneurship education is often seen as a way to foster economic growth and to deal with economic crisis (Allegra, 2013). However, entrepreneurship education is often pointed as immature, not sufficiently integrated in schools curricula and not adequately addressed by national policies, particularly in European countries (Allegra et al., 2013; Belloti et al., 2014).

Today, digital technologies are part of most people lives, from the early childhood. Educators and corporate trainers did not ignore this reality and Information and Communication Technologies (ICT) are being used in education and training for several years with different approaches, including the use of games. Games have been used for a long time (Farber, 2015; Schifter, 2013) as a support for learning activities. In the last decades, the popularity of video games and the introduction of ICT in education and training gave rise to a trend known as Game-Based Learning (GBL). This trend deals with games that have defined learning outcomes. Within GBL, simulations and serious games have gained notoriety (Allegra et al. 2013; Bastos et al. 2012).

According to Kapp et al. (2014), simulations provide a cost-effective scenario for training tasks that would be costly and time intensive to set up in a real scenario. Simulations must be realistic, they should allow to practice behaviors and experience the impact of decisions. Simulations provide a safe and realistic environment to test and get feedback of those decisions.

SGs is a term usually used to designate games with educational purposes. The term was coined by Clark Abt who published the book "Serious Games" in 1970 (Farber, 2015). Abt, cited by Ulicsak & Wright (2010), defined serious games as games that "have an explicit and care-

fully thought-out educational purpose and are not intended to be played primarily for amusement. This does not mean that serious games are not, or should not be, entertaining". Other definitions can be found as the one proposed by Laamarti et al. (2014) which state that "serious game has the potential to enhance the user's experience through multimodal interaction. This can be in different contexts such as education, training, health, or interpersonal communication". In this definition a SG can be seen as an application composed by three components: experience, entertainment, and multimedia (Laamarti et al., 2014).

Simulations and serious games share common features (Kapp et al., 2014) and they often overlap. They both have story line, they can be competitive and they can keep some kind of score. But while games may not reflect reality, simulations must be realistic representations of the real world. Serious games in particular have been considered as powerful tools to support entrepreneurship education (Belloti et al., 2014; Bastos, 2012). Belloti et al. presented a comprehensive study with an overview of several serious games available and identified some key benefits regarding their adoption in entrepreneurship education. Some of the studied games were Hot Shot Business, SimVenture and Go Venture Any Business, detailed below, and used in three higher education case studies (Hauge et al., 2013).

To introduce students with the basic principles of entrepreneurship, Hot Shot Business¹, a Flash game, allows students to open and run a simulated business and undertake the necessary actions to make it grow. Currently, users can choose to run a pet spa, a custom skate factory or a comic shop. Target users are teenager students but the game can also be used in higher education contexts. The game is a cartoon-like simulation, easy to use, available through a web browser with a high game speed suited for young users. Users choose the financing mode of their business, the resources they need and how to advertise the business. They get feedback through simple weekly financial reports to tune their activities and decisions. Hot

Shot Business was successfully employed to teach the basic principles of entrepreneurship (Hauge et al., 2013).

SimVenture² is a detailed single player simulation of a small computer assembly and selling company and aims university students and professionals. Users receive feedback in the simulation's monthly cycles and the all session runs for one simulated year. During the session bankruptcy may occur causing an early end of the game session. The game provides to the users several indicators for self-assessment and exports data to a spreadsheet allowing for students' comparison.

GoVenture Any Business³ is another business simulation platform that can be used with different types of businesses, industries and markets. Players can compete with other players or with the computer to successfully manage a business. Players have to make business decisions by setting different parameters (e.g. price, product features, and human resources) before the end of each simulation. The game presents the results of the students' actions in terms of sales and profits. An instructor receives detailed reports with all the students' actions.

Another online serious game for entrepreneurship is the ENTREplorer⁴ project (Carvalho et al., 2012) an European funded project. The project aims to improve training activities and initiatives for young entrepreneurs willing to run their own businesses. The game assists students to acquire entrepreneurial skills simulating the environment where they will develop real work. The game platform is also web-based and multiplayer. Players go through different scenarios or levels (represented as floors in a building) where they learn different subjects and face challenges. In each floor players learn about the phases of creating a new business. The game guides players through the different phases in the process of writing a business plan.

PNPVillage⁵ (Allegrò et al., 2013) is also a web-based serious game developed under an European project focused in fostering, especially among youths, the entrepreneurial mindsets. PNPVillage, where students have to manage a tourist resort,

creates a simulated environment to help students deal with complex situations and encourage competition among them. The game is level-based and students can progress from simple scenarios to more complex ones dealing with the main issues of company management. The game engine was integrated in a Learning Management System in order to improve the benefits of both platforms.

Although the above examples and others are already in use, there are still many difficulties in using video games for educational purposes, from both teachers and students: "bringing effective games to students has many barriers. Teachers may see play as frivolous, while students may be biased against learning games in school" (Farber, 2015). Also, the entertainment quality of many serious games is often quite low, particularly when compared to commercial games (Belloti, 2014). Games require "a large amount of design and development efforts" (Dicheva et al., 2015) and, therefore, "creating a highly engaging, full-blown instructional game however is difficult, time consuming, and costly" (Kapp, 2012). Regarding entrepreneurship education, the study from Belloti et al. (2014) pointed that most of the games focus only on company management with lack of support for improving specific entrepreneurial mindset. This mindset includes fostering the students' creative and innovative potential and must also improve their autonomy (Allegra et al., 2013).

DISCUSSION

The adoption of serious games in the learning process offers many key benefits for students and teachers but, simultaneously, brings also some challenges and issues that should be properly analyzed and mitigated in order to offer a complete immersive experience.

Key Benefits

GBLs help students to understand more easily concepts and remember them in a seamless way.

Sitzmann & Ely (2010) in a meta-analysis found that learners participating in simulation game learning experiences have 11% higher declarative knowledge, 14% higher procedural knowledge and 9% higher retention of training material than those trainees participating in more traditional learning experiences. Additionally, immersion in a learning environment can be highly realistic. It is expected that an entrepreneur that gets an experience by playing serious games with real life situations of business, will be able to overcome the problems by recollecting the solutions he learned from the game (Kapp, 2010). This approach is adopted by SimVenture and GoVenture SGs where players have to make business contextualized decisions. According to Gopal et al. (2015) serious games should also give the visualization effects and experience on different problems and solutions, which will affect the cognitive mind of the player. This situation will help the player while he/she is facing similar situations in real business.

Serious games in entrepreneurship education provide typically different scenarios, situations, problems and their solutions in a business environment. Therefore they can be used to analyze the global environment with the intent to identify international entrepreneurship opportunities. Typically after completing levels and scenarios of the games, the player will get an idea about the common challenges and their solutions in the business field. Typically a common difficulty faced by students when trying to create a new business is around the concept of financial viability. There are traditional difficulties to understand the concept of cash flow and its impact in the business. Serious games can be used to more easily understand, categorized and apply in practice this concept. This approach is fully adopted by PNPVillage SG where the player progresses from simple scenarios to more complex ones, while managing and self-monitoring market trends, financial indicators and cash flow projections.

Game enjoyment by students has been shown to be an essential factor in the learning process (McClarty et al., 2012). Identification with fictional

characters has been researched during these last years and has been found to be one of the main factors why people enjoy the use of different media (Looy et al., 2010). Therefore, the role of avatar identification in the learning experience of a serious game is particularly relevant. Furthermore, Ahn et al. (2014) have confirmed that an experience as an avatar can change a person's real life perceptions and behaviors. Several serious games capture players and provide immersive experiences through the use of avatars. A good example of this approach is done by Hot Shot Business SG, where the player takes on the role of pet spa manager, skate factory owner or a comic shop manager. In such games, players either inherit a strongly formed and appealing character or they get to build a character from the ground up. Therefore, players become committed to the new virtual world in which they will learn and act.

Serious games also allow students to work together to accomplish tasks and goals as a virtual team. In cooperation, students can split the work, solve sub-tasks individually and then assemble the partial results into the final output. This possibility is highly suitable for entrepreneurship context where multidisciplinary competencies are fundamental. Fayolle (2007) & Wilson (2008) confirm this theory by advocating that teaching entrepreneurship requires a multidimensional and cross-disciplinary approach with an emphasis on dynamic processes. Additionally, entrepreneurship classes attract a significant number of student workers, which typically have less availability to attend the semester sessions. In this way, serious games can be used to promote greater collaboration and cooperation between students and teachers, while they are at home or at work.

Most serious games are role-playing games (RPGs), which means that users adopt the role of "invisible actors" who take actions based on the decisions of the person playing the game. A key element in this approach is the construction of an immersive story composed by different characters, scenes and a narrative. Lippa & Borst (2013) advocate that the human brain has a natural affinity

for narrative construction and that people tend to remember facts more accurately if they encounter them in a story rather than in a list.

The key benefits found can be systematized in five elements:

1. Facilitate the learning process by turning it more realistic and attractive,
2. Test different scenarios, situations and problems in business context,
3. Game enjoyment,
4. Virtual cooperation, and
5. Immersive experience.

Challenges and Issues

One of the major concerns identified about using GBL approaches in classrooms is the difficulty in assessing effectiveness at achieving the learning goals. Additionally, the adoption of serious games in academic context can interfere with evaluative learning and creates an additional increase difficulty in evaluating individual learning when teams play. The assessment process should take into account all the classroom or focus group and not only assess an individual player performance. This is considering a crucial element in entrepreneurship classes, where students need to work in group and establish a high degree of cooperation among them. Some serious games offer explicit scoring mechanisms that can be used to determine the performance of the player, namely the number of correct answers or time taken to complete a game. This approach is implemented by ENTRExplorer game, where the score of the user is calculated based in the number of correct answers provided by the players. The appropriate assessment methodology is also dependent upon content and context. Ulicsak (2010) states that the game play may have more of an impact on learning as that determines how feedback is given.

Kirriemuir (2010) has shown that it is difficult for teachers and other school stakeholders to easily realize the full potential benefits of serious games. Researchers found that teachers

have shown difficulties to identify quickly how a particular game is relevant to some component of the statutory curriculum, which includes also the entrepreneurship area, as well as the accuracy and appropriateness of the content within the game. Egenheldt-Nielsen (2011) states that the most important consideration from a teacher's perspective is how much the game will make their life easier. Thus it involves to analyze how the selected game will easily enhance teaching, or, if there is not yet a relevant game in that area, what a game would need in order for it to be useful. Furthermore, the lack of time available for teachers to familiarize themselves with the games, and methods of producing the best results from its use have been reported as a constraint. Schools can achieve this challenge by reserving some amount of time in the final semester for teachers' coaching, where teachers can discuss about their practice in classroom.

Serious games have to be embedded in a didactical system. In entrepreneurship domain the challenge is even bigger due to its inter-disciplinary approach. Therefore, schools and programme coordinators need to guarantee that students acquire knowledge from different disciplines through a unique theme. At the end, teachers should always care and make a critical analysis about what went good or wrong with the introduction of a serious games in the classroom. A key challenge in this process is to assure that players are engaged and motivated with the game. Having fun is central to effective learning. According to Karner and Härtel (2011) the goal is to enhance a self-reflexive attitude, because through individual reflexion self-controlled learning will be successful.

In order to offer an immersive experience serious games needs to support different styles of learning. The background heterogeneity of students that typically are enrolled in Entrepreneurship classes increases the creativity and potential of new business ideas, but turns more difficult to define a well stable learning approach. For instance, there are students that prefer visual styles, but others are more auditory. In fact, it

is a challenge to combine and support different types of media elements, such as text, pictures, videos, audio, and different types of communication. Each student will have its own experience of gaming and learns in a different way. This has to be recognized through the phase of conception and planning the game (Karner & Härtel, 2011).

Finally, there are also interoperability, accessibility and usability issues (Gooriah, 2013). There are two main factors involved in interoperability of serious games: content identification (e.g. integration with a Learning Management System such as Moodle) and data exchange (e.g. access to questions and test data results). Accessibility and usability issues are also common due to the nature of a serious game, that typically have a huge dynamic situations and offer great interactivity, turns very difficult to achieve high levels of accessibility and usability compliance.

The main challenges and issues can be summarized in five dimensions:

1. Difficulties to assess the effectiveness of learning goals,
2. Realize the full potential benefits of serious games by teachers and school stakeholders,
3. Embedment in a didactical system,
4. Support different styles of learning, and
5. Interoperability, accessibility and usability issues.

FUTURE RESEARCH DIRECTIONS

There are several perspectives about future evolution of serious games. In our work we will look mainly for the challenges that affect the entrepreneurship learning and behavior of the students regarding the idea to build their own business.

A first research direction that we can appoint concerns the assumed motivational impact of games, particularly the relation between “flow” and learning from serious games. Flow has been described as such an extent of involvement in a

task that nothing else seems to matter (Arnab et al., 2012). According to Wouters et al. (2013) it is undeniable that players fully engage in popular games and forget the real world around them, it is still unclear how such full engagement relates to learning. With this focus research may begin to establish links between game features, motivational processes, and learning outcomes.

Another perspective for future research concerns the validity and relevance of the learning outcomes. The use of traditional assessment methods with the adoption of future games still deserve deeper research. Additionally, assessment methods should take the context of learning into account and, therefore, it can evolve from traditional evaluation methods for more innovative assessment models that could reflect better the dynamics created by the students working groups.

The entrepreneurship education is interdisciplinary in nature. This situation turns entrepreneurship so exciting but also creates additional challenges for teachers and students. Students, for example, need to learn a wide set of different competencies from disciplines such as design, finance, technology and engineering. Additionally, students need to interact in multidisciplinary groups and accept the different of knowledge and personal behavior of several students. On the other hand, teachers have the challenge to be particularly attentive and solicitous to all students from different disciplines. In addition to that, there is a rapidly changing technology, human mindset, social value, business and personal values, entrepreneurship as a career path looks different and feels different (Neck & Greene, 2011). Connor et al. (2014) defend that in order to accommodate all new waves in the global environment there is a need for a framework of teaching.

Finally, there is a need for the establishment of new innovation policies that could stimulate stronger links between academia, business and entrepreneurs. For that, the technology transfer offices of universities and polytechnic institutes should have a more active role and establish

cooperation between teachers and students that could facilitate the appearance of new spinouts from academic institutions.

CONCLUSION

Serious games offer the potential of improving learning processes by providing attractive, motivating and effective tools that may also create positive situations among students and with teachers. Serious games create the ideal model to match the content in different ways and incorporate problem-solving and reasoning stimulations, which results in learning core competencies.

Serious games can be used in different academic contexts and, definitely, entrepreneurship education is an important challenge for the nowadays knowledge societies. Entrepreneurship education requires practices that make it possible to unleash the creative and innovative potential of young people, enabling them to work with multidisciplinary teams and contexts, while they create their own jobs and promote a more dynamic economy. When adopting serious games, students are motivated and guided more efficiently and effectively in the management of their virtual companies, receiving the necessary information regarding entrepreneurship, management principles, marketing research and financial analysis.

The introduction of serious games for entrepreneurship education offers a wide range of key benefits, namely in terms of students' motivation, immersive experience, and acquisition of multidisciplinary competencies. Additionally, serious games offers an environment that promotes the free choice of learning place and speed, flexible time-management and autonomous learning in the game context. However, the adoption of serious games for entrepreneurship education brings some challenges and issues that should be properly mitigated by educators. Among them, we highlighted in our study the difficulty in assessing effectiveness at achieving the learning goals, the need for

embedded in a coherent didactical system, support different styles of learning, interoperability of serious games in terms of content identification and data exchange, and accessibility and usability of the serious games platform.

REFERENCES

- Ahn, S., Bailenson, J., & Park, D. (2014). Short- and long-term effects of embodied experiences in immersive virtual environment on environmental locus of control and behavior. *Computers in Human Behavior*, 39, 235–245. doi:10.1016/j.chb.2014.07.025
- Allegra, M., Dal Grande, V., La Guardia, D., Gentile, M., & Ottaviano, S. (2013). A serious game to promote and facilitate entrepreneurship education for young students. *Proceedings of the 2013 International Conference on Education and Educational Technologies*, 256-263.
- Arnab, S., Berta, R., Earp, J., Freitas, S., Popescu, M., Romero, M., Stanescu, I., & Usart, M. (2012). Framing the Adoption of Serious Games in Formal Education. *Electronic Journal of e-Learning*, 10(2), 159-171.
- Bastos, N., Sánchez-Cantón, B., & Costas, I. (2012). ENTRExplorer: a Serious Game for Immersive Entrepreneurs. In *Atas do Encontro sobre Jogos e Mobile Learning*. Braga: CIED.
- Bellotti, F., Berta, R., De Gloria, A., Lavagnino, E., Antonaci, A., Dagnino, F., & Mayer, I. S. et al. (2014). Serious games and the development of an entrepreneurial mindset in higher education engineering students. *Entertainment Computing*, 5(4), 357–366. doi:10.1016/j.entcom.2014.07.003
- Connor, A. M., Karmokar, S., & Walker, C. (2014). Doing Entrepreneurship: Towards an Entrepreneurial Method for Design and Creative Technologies. In T. T. Jachna, E. McLafferty, & Y. S. Tzvetanova (Eds.), *Action! Hong Kong: Doing Design Education*.

- Dicheva, D., Dichev, C., Agre, G., & Angelova, G. (2015). Gamification in education: A systematic mapping study. *Journal of Educational Technology & Society*, 18(3), 75–88.
- Egenfeldt-Nielsen, S. (2011). What makes a good learning game. *ACM Learn Magazine*, 1, 223–232.
- Farber, M. (2015). *Gamify Your Classroom: A Field Guide to Game-Based Learning*. New York: Peter Lang Publishing.
- Fayolle, A. (2007). *Handbook of research in entrepreneurship education: A general perspective*. Cheltenham, UK: Edward Elgar Publishing, Ltd.
- Gooriah, K. (2013). Capacity building in entrepreneurship education and training using simulation. In *Proceedings of International HRD Conference* (pp. 1-12). Mauritius: Human Resource Council of Mauritius (HRDC).
- Gopal, A., Bijlani, K., & Rengaraj, V. (2015). An Approach to Overcome Startup Issues Using Serious Games. In N. R. Shetty, N. H. Prasad, & N. Nalini (Eds.), *Emerging Research in Computing, Information, Communication and Applications: ERCICA 2015* (Vol. 2, pp. 109–121). Heidelberg, Germany: Springer.
- Gwija, S., Eke, C., & Iwu, C. (2014). The link between entrepreneurship education and business success: Evidence from youth entrepreneurs in South Africa. *Journal of Economics*, 5(2), 165–175.
- Hauge, J., Bellotti, F., Berta, R., Carvalho, M., De Gloria, A., Lavagnino, E., & Ott, M. et al. (2013). Field assessment of serious games for entrepreneurship in higher education. *Journal of Convergence Information Technology*, 8(13), 1–12.
- Kapp, K. (2010). *Key Advantages of Serious Games/Immersive Learning Simulations*. Retrieved February 18, 2016, from <http://karlkapp.com/table-key-advantages-of-serious-games-immersive-learning-simulations/>
- Kapp, K., Blair, L., & Mesch, R. (2014). *The Gamification of Learning and Instruction Fieldbook: Ideas into Practice*. Wiley.
- Karner, F., & Härtel, G. (2011). *Theory and Taxonomies of Serious Games*. Retrieved February 9, 2016, from http://www.entrexplorer.com/pdf/Theory_and_Taxonomies_of_Serious_Games.pdf
- Laamarti, F., Eid, M., & El Saddik, A. (2014). An Overview of Serious Games. *International Journal of Computer Games Technology*, 10(1), 1–15. doi:10.1155/2014/358152
- Looy, J., Courtois, C., & Vocht, M. (2010). Player identification in online games: validation of a scale for measuring identification in MMORPGs. *Proceedings of the 3rd International Conference on Fun and Games*, 126-134.
- Luppa, N., & Borst, T. (2012). *Story and Simulations for Serious Games; Tales from the Trenches*. Abingdon, VA: Taylor & Francis.
- McClarty, K., Orr, A., Frey, P., Dolan, R., Vassileva, V., & McVay, A. (2012). *A Literature Review of Gaming in Education*. Pearson Education.
- Neck, H., & Greene, P. (2011). Entrepreneurship Education: Known Worlds and New Frontiers. *Journal of Small Business Management*, 49(1), 55–70. doi:10.1111/j.1540-627X.2010.00314.x
- Schifter, C. (2013). Games in learning, design, and motivation. In *Handbook on Innovations in Learning* (pp. 149–164). Information Age Publishing.

Sitzmann, T., & Ely, K. (2010). Sometimes you need a reminder: The effects of prompting self-regulation on regulatory processes, learning, and attrition. *The Journal of Applied Psychology*, 95(1), 132–144. doi:10.1037/a0018080 PMID:20085411

Ulicsak, M. (2010). *Games in education: serious games*. Bristol: Futurelab.

Wilson, K. (2008). *Entrepreneurship Education in Europe*. Retrieved January 12, 2016, from <http://www.oecd.org/site/innovationstrategy/42961567.pdf>

Wouters, P., Van der Spek, E., & Van Oostendorp, H. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology*, 105(2), 249–265. doi:10.1037/a0031311

ADDITIONAL READING

Gibb, A. (2010). *Towards the Entrepreneurial University*. Birmingham: NCGE.

Lungu, E. (2014). A Competitive Serious Game for Potential and Young Entrepreneurs. *eLearning & Software for Education*, 1, 340-345.

Neck, H., Greene, P., & Brush, C. (2014). *Teaching Entrepreneurship: A Practice-Based Approach*. Cheltenham: Edward Elgar Pub. doi:10.4337/9781782540564

Pihie, Z., & Sani, A. (2009). Exploring the Entrepreneurial Mindset of Students: Implication for Improvement of Entrepreneurial Learning at University. *The Journal of International Social Research*, 2(8), 340–346.

Valerio, A., Parton, B., & Robb, A. (2014). *Entrepreneurship Education and Training Programs around the World: Dimensions for Success (Directions in Development)*. New York: World Bank Publications. doi:10.1596/978-1-4648-0202-7

KEY TERMS AND DEFINITIONS

Entrepreneurship: The capacity and willingness to develop, organize and manage a business venture along with any of its risks in order to make a profit. The most obvious example of entrepreneurship is the starting of new businesses.

Flash: A popular authoring software developed by Macromedia that is used to create vector graphics based animation programs with full-screen navigation interfaces.

Learning Management System: A software system that delivers courseware plus e-tutoring over the Internet. It is typically also called as Course Management System, Pedagogical Platform or E-Learning Platform.

Role-Playing Games: A game in which the player assumes the role of a character, generally in a fantasy or science fiction setting, that can interact within the game's imaginary world.

Serious Games: A game designed for any purpose rather than entertainment.

Simulations: A simulation game assumes typically a form of video or computer game that is concerned with playing out realistic situations in game settings.

Technology Enhanced Learning: The adoption of technology in order to promote a better classroom experience for students or increase e-learning pedagogic activities.

ENDNOTES

- 1 <http://igrezadecu.com/hot-shot-business/>
- 2 <http://simventure.co.uk/>
- 3 <http://goventureceo.com/>
- 4 <http://www.entrexplorer.com/game/>
- 5 <http://pnpv2.pa.itd.cnr.it/en/>

Category B

Business Information Systems

Architectural Framework for the Implementation of Information Technology Governance in Organisations

Thami Batyashe

Cape Peninsula University of Technology, South Africa

Tiko Iyamu

Cape Peninsula University of Technology, South Africa

INTRODUCTION

For many organisations, it is not just about information technology (IT), but about governance of systems and technologies, which is inseparable from people and processes. In a similar manner, as business management is governed by generally accepted principled practices, IT must be governed by practices that facilitate and make sure an organisation's IT resources are used responsibly and that its risks are managed appropriately. According to Van Grembergen and De Haes (2007), the widespread application of technology has generated a critical reliance on IT, necessitating a special focus on IT governance. The past decade has seen the term 'governance' moved to the forefront of business thinking in response to instances indicating the importance of good governance of IT. Governance is not an approach by itself, it is guided by architecture.

Enterprise architecture (EA) consists of four main domains, business, information, technical and application (Iyamu, 2014). Technical architecture means IT architecture in the context of this paper. This paper focuses on the technical architecture in the context IT governance. Technical architecture involves the design of systems or sets of systems. Iyamu (2011) defined technical architecture as a logically constant array of prin-

ciples, standards and models that are originating from business requirements. It guides the engineering of an organisation's information systems and technology infrastructure across. According to The Open Group Architecture Forum's (TOGAF) document of 2009, IT architecture provides some governance aspects, such as change management and quality assurance. In other words, it is the grouping of systems, represented in components, their relationships to each other and the environment, and the principles governing design and development.

IT architecture is driven by the need to bridge the gap between IT and business people and process towards a common goal of the organisation. Klein and Gagliardi (2010) describe IT architecture as "the logical software and hardware capabilities that are required to support the deployment of business, data and application services. This includes IT infrastructure, middleware, networks, communications, processing and standards". Along the same vein, in 2009, TOGAF described IT architecture as the hardware, software and network infrastructure needed to support the deployment of core, mission critical applications of an organisation. These activities require management and governance in achieving the objectives as well as a return on investment (ROI) for the organisational purposes.

BACKGROUND: IT GOVERNANCE AND ARCHITECTURE

Some organisations view both IT governance and architecture from two different perspectives, in implementation and operationalisation. The main and most commonly adopted IT governance frameworks include COBIT, ITIL, ISO/IEC 17799/27002 and TOGAF (Simonsson & Johnson, 2006; Niemann, Eckert, Repp & Steinmetz, 2008). Enterprise architecture (EA) is the focal point, though some organisations do sometimes focus on one or two domains of EA.

Mårten, Lagerström and Johnson (2008) asserted that the aim of IT governance is to support IT's function as a business enabler in order to realise the internal effectiveness in an organisation. IT governance enables and improves IT and business strategies gain alignment, including management of risks. Brown (2006) argued that IT governance governs the crafting and execution of the IT strategy, and also help to aligns both IT and business strategies.

The IT Governance Institute (2007) defined IT governance as the “responsibility of executives and the board of directors, and consists of the leadership, organisational structures and processes that ensure that the organisation's IT sustains and extends its strategies and objectives”. Concurring, Ross, Weill and Robertson (2006) refer to IT governance as “the decision rights and accountability framework for encouraging desirable behaviour in the use of IT”. IT governance focuses on managing and employing IT to realise corporate performance objectives whilst reflecting the wider corporate governance principles.

Enterprise architecture (EA) is intended to govern and manage both technical and nontechnical activities of an organisation. According to Anaya and Ortiz (2005), “an Enterprise Architecture provides a common view of the primary resources of any enterprise (people, processes and technology) and how they integrate to provide the primary drivers of the enterprise (that is, the strategy)”. Enterprise Architecture can be used to guide

against business-IT misalignments. This includes coordination of technology investments, to suite business needs, improve the integration between services, and eliminate redundant investments while replacing them with standardised and cost effective IT services. Becker, Antunes, Barateiro, Vieira and Borbinha (2011) stated the purpose of EA is to provide a comprehensive coverage of the organisation.

Weill and Woodman (2002) asserted that IT architecture offers an interrelated set of technical choices to direct the organisation in sustaining business needs. According to Hafner and Winter (2008), IT architecture is the sphere of architecture which signifies a combined, enterprise wide model of hardware and communications elements in addition provides support among the technology artefacts. While Iyamu (2011) maintained that it guides the initiation of new technology. IT architecture as defined by Weill and Woodman (2002) is a set of procedures and guidelines that govern the use of IT and design a migration path to the way business will be done. IT architecture consists of standards and guidelines for technology, utilisation of data, design of applications and change management processes necessary to use new technologies.

RESEARCH APPROACH

The qualitative, case study and interpretive methods and approach were followed in conducting this research, which was to develop an architectural framework for the implementation of IT governance in the organisations. Qualitative research methods assist researchers to understand people and the social and cultural contexts within which they live. Concurring, Denzin and Lincoln (2011) asserted that qualitative research is often based upon interpretivism, constructivism and inductivism. It is about exploring the subjective meanings through which people interpret the world and the diverse ways in which life is constructed in particular settings. Social events and phenomena

are understood from the perspective of the people themselves, thus avoiding the imposition of the researcher's own presumptions and definitions.

The case study research approach was followed in this study. A South African organisation, RedLeaf Communications was used as the case, with specific focus on the IT division of the company. Case studies are empirical studies that investigate a contemporary phenomenon within its real life context using multiple sources of evidence (Yin, 2003). Noor (2008) sees case studies as being concerned with “*how*” and “*why*” things happen, allowing the study of appropriate realities and the differences between what was planned and what actually occurred. Thus case studies are a valuable approach of looking at the world around us. Willis (2007) defined a case study as “an examination of a specific phenomenon such as a program, and event, a person, a process, an institution, or a social group”. The case study approach, according to Creswell (2007) “explores a single entity or phenomenon bounded by time and activity and collects detailed information by using a variety of data collection procedures during a set period of time”.

The semi-structured interview approach was employed to collect data. This is mainly because the approach allows the interviewer or interviewee to deviate in order to pursue an idea or response in more detail (Rabinet, 2011). According to Qu and Dumay (2011), the semi-structured interviewee approach helps to reveal important and often hidden aspects of human and organisational behaviour. Kvale and Brinkmann (2009) argued that it is often the most effective and suitable method to collect data.

The hermeneutics method, from the perspective of interpretivism, was employed to analyse the data, in order to make sense of the subjective views and opinions of the participants in this study. Hermeneutics is constantly employed within interpretive research. Hermeneutics is the discovery or discipline of interpretation. Its primary concern is questioning the meaning of a text or text-analogue, which can be a book, scholarly

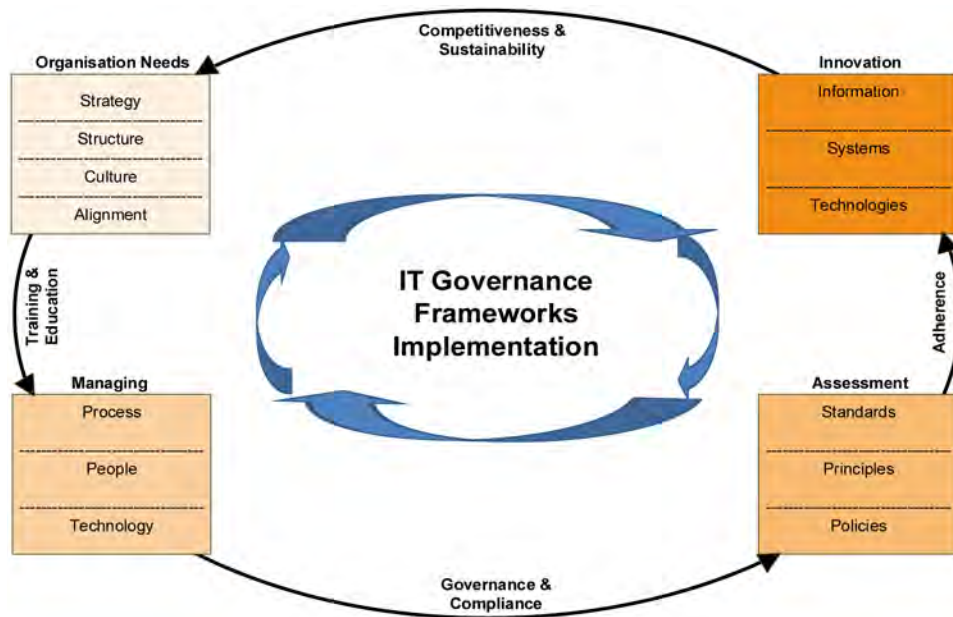
article, interview notes or an organisation (Myers, 2009). The hermeneutics practice states that in qualitative research, the interpretation is mainly used, no matter how adamantly many researchers may argue the facts speak for themselves (Folwer, Horan and Cope, 2007). These researchers further stated that hermeneutics is a form of interpretive research where researchers begin with the assumption that interaction to our understanding of reality is gained only through social constructions such as language, consciousness, shared meanings, documents, tools and other artefacts. In Information Systems (IS), interpretive research is aimed at creating an understanding of the context of IS and the process by which IS affects and is affected by its context (Fowler, Horan and Cope, 2007).

ARCHITECTURAL FRAMEWORK FOR THE IMPLEMENTATION OF IT GOVERNANCE

The data was collected from the case study, RedLeaf Communications was analysed. The analysis was carried within the context in which data was gathered: (1) What are the factors that influence the selection of IT governance in the organisations? (2) How is IT governance implemented in the organisations? The findings from the case study were combined and presented in Figure 1 below to give a better understanding of the common factors which could influence the implementation of IT governance in organisations.

The critical factors found at RedLeaf Communications, were evaluation of information, technology repository, education and training, organisational strategy and organisational culture. Also found were factors which include environmental assessment, training, knowledge sharing, communicative scheme and organisational culture. The combination of the findings informed the development of the Architectural Framework for IT Governance (AFITG), which can serve as a best practice guide to the implementers of IT governance frameworks in organisations. The

Figure 1. Architectural framework for IT governance (AFITG)



framework (AFITG) is considered a best practice mainly because of the depth and rigour through which empirical evidence was examined, and its organisational relevance.

It is believed that this study could be generalised for the benefit of organisations as the commonalities in the findings from the case. According to Walsham (1995), a generalisation can be made to a concept or a theory can be developed or generated from facts or a rich description of a case. Lee and Baskerville (2003) asserted that generalisation is important for the purposes of managing and solving problems that corporations and other organisations experience in society. Along the same line of argument, Qureshi, (2005) posited that the insights from a case study can be generalised to reach conclusive or theoretical statements for other organisations' purposes.

The framework in Figure 1 depicts four fundamental categories in an organisation that must work in concert for the implementation of IT governance to be effective: organisational needs, managing, assessment and innovation. Figure 1 suggests that if a change in IT governance is made in an organisation such as the introduction of a

new governance framework, this change is likely to affect the other three categories. For example, changes in organisation needs can lead to managing changes, and organisation needs can drive changes in innovation.

Implementation of IT Governance Frameworks

Figure 1 illustrates the most critical factors which impact and influence the implementation of IT governance frameworks. These factors are categorised into organisational needs, innovation, assessment, and managing. Each category relates to each other and no category stands on its own. Also, each category positively supports the others. The factors are considered critical because of their prevailing and significant roles, as shown in the analysis of the case study. The categories are discussed in more detail below.

Organisational Needs

Organisational needs are made up of strategy, structure, culture and alignment. IT strategy is

informed by and aligned to the organisational strategy. An IT structure must support the IT strategy to ensure the organisation's objectives are achieved. An organisational culture is shaped by the norms and values of the organisation and the management and individual practices.

- **Strategy:** To change its operations, an organisation's IT management needs to define a strategy that is aligned with the organisational and business units' strategies. It needs to become a true service provider that operates like a business within a business; the IT strategy should entail an IT governance framework that fully involves the organisation's management and individuals in planning and aligning IT initiatives with the organisation's priorities. Once the strategy is operational, the structure that supports the strategy should be outlined and implemented with the support of the organisation's management.
- **Structure:** Structure is extremely important and a key element in implementing an IT governance framework. The role of the structure within the organisation is to define expectations, assign responsibilities and verify performance. This structure also establishes the strategic, operational and technical decision-making processes, which are critical. This helps to manage IT investment; assign trained and capable resources to govern IT resources; put appropriate standards, principles, and policies into place; and create the disciplines around governing and managing IT artefacts. Similar to any major organisational innovation, implementation of IT governance must have an owner and accountabilities. Putting the right IT structure in place is important to support and manage the implementation of IT governance usually requires having an organisational culture that embraces and supports innovation.
- **Culture:** Culture relates to the beliefs and values of the organisation, which are generally expressed in its mission statement, and these include ethical concerns. Management need to understand the influence culture has on the organisation's values and behaviours. A strong organisational culture fosters organisational learning capacity to increase awareness of how culture shapes organisational effectiveness by training and educating employees and creating communication programs to address critical cultural weaknesses.
- **Alignment:** Strategic alignment involves making certain that the business and IT plans are connected, and that IT operations are aligned with overall business operations. Alignment is a key element that, if ignored, can lead to serious management issues, sub-standard performance, and even financial problems. Alignment is not easy to achieve. The challenges of creating internal alignment of systems that support appropriate levels of risk-taking include the development of a common language, and an understanding among employees as to the nature and tasks of IT governance. Proper alignment also requires conflicts between functions be addressed. Unnecessary overlaps of jobs and areas of accountability, structure, as well as gaps in responsibility, must be identified and resolved. This may require complex and difficult organisational change.

Innovation

Organisations are continuously challenged to ensure the survival and wellness of their competitiveness and sustainability. Thus, there is a need for continuous innovation. Innovation is intended to provide improved approaches for efficiency and effectiveness in achieving an organisation's needs. There are three main factors which repre-

sent innovation. They comprise of information, systems and technologies. Information should facilitate making decisions regarding the systems and technologies that are required to implement the innovation.

- **Information:** Information is about the know-how and know-what of the organisation's decisions that needs to be made. Information assists decision makers to have an understanding of what needs to be prioritised to ensure that the right systems are implemented to achieve the goals and objectives of the organisation. It is also important to ensure that the relevant information is communicated to the relevant stakeholders.
- **Systems:** Although there are various definitions of systems, in this context system refer to the collections of technical and non-technical factors that enable an organisation to meet its desired objectives. A system which creates a cohesive work environment, and ensures that implementation of IT governance is understood and adhered to. Technologies are required to enable these systems.
- **Technologies:** Key technologies should be identified to automate and standardise those systems. It should then consider what linking technologies, if any, can be shared across the organisation. Organisations dependence on technologies is swiftly and continuously increasing. It is, therefore vital to ensure that IT capability is described and directed by the organisation's needs.

Assessment

An environment assessment can be conducted to identify IT objectives and to ensure that the business objectives are aligned with an understanding of management's risk appetite and understanding

of the maturity of the existing governance and related processes.

Assessment of the organisation environment assists management to identify the gaps in standards, principles, policies. Once they have been reviewed, new business proposals are provided to advice on IT governance compliance issues. As a result, policies are established to ensure compliance with relevant legislation. Policies that govern IT governance and the implementation and management of IT governance frameworks should be familiar with the relevant standards and frameworks and the principles embedded within them.

- **Standards:** standards are set to ensure uniformity of systems, processes and technologies. However, standards, principles and policies are not permanent solutions as their effectiveness depends on how they have been implemented and kept up to date. Standards are highly valuable when they are used as a set of principles to begin the customisation of certain policies. It is therefore important for management and employees to understand what to do, how to do it, and why it is important.
- **Principles:** the organisation needs to provide the direction and guidance for the selection and implementation of IT governance based on the capacity and organisational strategy. The guidelines point the organisation to the principles that should be reinforced to manage IT governance. These principles are extremely useful to develop policies to manage and control IT governance.
- **Policies:** Policies for IT governance frameworks implementation include best practise framework and standards, stakeholder management, business alignment, knowledge transfer and management of the selected frameworks.

Managing

Managing the implementation of IT governance frameworks consists of three main factors: process, people and technology. The crucial job of implementing these IT governance frameworks is performed by people. All three factors are necessary to get a clear understanding of the IT governance processes, therefore these factors cannot be separated, as they rely on each on the success or failure of implementation IT governance frameworks.

- **Process:** Processes are used to guide the activities of people in the execution of various tasks. Also, processes are applied in the management (selection, implementation, and support) of information systems and technologies in achieving the organisational needs. Without processes, the organisations would face challenges achieving their objectives and managing day-to-day operations. The failure of most IT governance implementation is the lack of well-defined processes.
- **People:** The processes are developed by people, based on their understanding and know-how. People have an integral role in process. People make use of processes at their discretion, either for personal interest or organisational interest, in the selection, implementation and management of information systems and technologies. Different people have different skills and knowledge, which is required in innovations undertakings such as implementing IT governance. This makes their role a determining factor to the success or failure of the organisations' goals and objectives. Thus, it is important to involve the right people, doing the right things, in the right ways, at the right times.
- **Technology:** The implementation of IT governance frameworks can be complex and involves technology, process and

people at all levels of the organisation. Technologies are used to enable an organisation's goals and objectives. However, technologies can also constrain the organisation in the quest for competitiveness. Hence the use of technologies highly depends on its management by people, through processes.

Significance of the AFITG

The architectural framework described in the previous section guides organisations in the selection, implementation and management of frameworks for IT governance. It proposes a strategic plan for IT, which satisfies the current and ongoing needs of an organisation's business strategy, and the current and future IT capabilities. It promotes clear decision making, leading to valid reasons for IT acquisitions.

Also, the architectural framework monitors provision of IT services, levels of service and service quality. Thus, it assures that the organisation's business processes, people and technology are compliant with relevant IT governance legislation and that the organisation operates according to the principles embedded in relevant IT governance frameworks. The AFITG also promotes IT standards, principles and policies, and decisions that recognise the current and changing needs of organisations.

FUTURE RESEARCH

The study was methodically and comprehensively carried out within the objective, which was to develop an architectural framework for the implementation of IT governance in organisations. This includes understanding of the factors that influence the selection and implementation of IT governance in an organisation. The findings and analysis reveals that further work with different users and settings will support the research's understanding of the processes that are employed

to selecting, implementing and managing IT governance frameworks. In addition, such further research could make use of the framework that is presented in chapter as a point of departure, thereby explores its validity and practicality.

The analysis and interpretation of the two case studies uncovered participants' concern regarding factors, such as organisational culture and organisational strategy during the selection and implementation of IT governance frameworks. It would be in the interest of scholars to explore and gain a better understanding how organisational culture and strategy influence the selection and implementation of IT governance frameworks.

In addition, further studies could be conducted on the impact that training and educating the implementers and non-implementers has on the implementation and management of IT governance frameworks. Actors need to be knowledgeable on the selected frameworks of IT governance they implement in order for them to make better decisions to innovate.

CONCLUSION

Even though one organisation was used in the study, the framework can be generalised. This is mainly because there are many organisations that have similar challenges in the area of IT governance, as revealed in the study. Also, many other organisations that do not yet have challenges, but could potentially do so, now have the privilege of preventing them before they occur. As presented and discussed in this paper, the use of the AFITG brings a fresh perspective and helps with a deeper understanding of how IT governance can be selected and implemented, through its influencing factors in the organisations. The architectural framework, shown in Figure 1, which was derived from the study, illustrates how non-technical factors can influence and impact the implementation IT governance frameworks in organisations.

This study can assist boards of directors, top managers and other employees in organisations to better comprehend the factors which influences or can potentially impact the selection, implementation and management of IT governance frameworks in their organisations. Also, the study can be a useful material to the academics, in twofold: (1) It can be used as a case material, from empirical evidence perspective, and (2) the framework can be used as theoretical foundation.

REFERENCES

- Anaya, V., & Ortiz, A. (2005) How enterprise architectures can support integration. In *Proceedings of the first international workshop on Interoperability of heterogeneous information systems*, (pp. 25-30). ACM.
- Becker, C., Antunes, G., Barateiro, J., Vieira, R., & Borbinha, J. (2011) Modeling Digital Preservation Capabilities in Enterprise Architecture. In *Proceedings of the 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times*, (pp. 84-93). ACM. doi:10.1145/2037556.2037570
- Brown, W. C. (2006). IT governance, architectural competency, and the Vasa. *Journal of Information Management and Computer Security*, 14(2), 140–154. doi:10.1108/09685220610655889
- Creswell, J. W. (2007). *Qualitative inquiry and research design: Choosing Among five Approaches*. Sage Publications.
- Denzin, N. K., & Lincoln, Y. S. (2011). *The SAGE Handbook of Qualitative Research 4*. London: Sage.
- Folwer, J., Horan, P., & Cope, C. (2007). How an “Imperative” IS Development was Saved from a Failing Course of Action – A Case Study. *Information and Beyond: Part I: Issues in Informing Science and Information Technology*, 4, 395–406.

- Hafner, M., & Winter, R. (2008) Processes for Enterprise Application Architecture Management. In *Hawaii International Conference on System Sciences, Proceedings of the 41st Annual* (pp. 396-396). IEEE.
- IT Governance Institute. (2007). *Unlocking Value: An Executive Primer on the Critical Role of IT Governance*. Information Systems Audit and Control Association. Retrieved from <http://www.itgi.org>
- Iyamu, T. (2011). A Framework for Developing and Implementing the Enterprise Technical Architecture. *Computer Science and Information Systems*, 9(1), 189–206. doi:10.2298/CSIS101103040I
- Iyamu, T. (2014). *Information Technology Enterprise Architecture: From Concept to Practice*. Heidelberg Press.
- Johnson, P., Buehring, A., Cassel, C., & Symon, G. (2007). Defining qualitative management research: An empirical investigation. *Qualitative Research in Organisations and Management: An International Journal*, 2(1), 23–42. doi:10.1108/17465640710749108
- Klein, J., & Gagliardi, M. (2010). *A Workshop on Analysis and Evaluation of Enterprise Architectures*. Retrieved from <http://www.sei.cmu.edu>
- Kvale, S., & Brinkmann, S. (2009). *Interviews: Learning the Craft of Qualitative Research Interviewing*. Los Angeles, CA: Sage Publications.
- Lee, A. S., & Baskerville, R. L. (2003). Generalizing generalizability in information systems research. *Information Systems Research*, 14(3), 221–243. doi:10.1287/isre.14.3.221.16560
- Mårten, S., Lagerström, R., & Johnson, P. (2008) A Bayesian network for IT governance performance prediction. In *Proceedings of the 10th International Conference on Electronic Commerce*. ACM.
- Maxwell, J. A. (2013). *Qualitative Research Design: An Interactive Approach*. London: Sage Publications.
- Myers, M. D. (2009). *Qualitative Research in Business & Management*. London: Sage Publications.
- Niemann, M., Eckert, J., Repp, N., & Steinmetz, R. (2008) Towards a Generic Governance Model for Service-oriented Architectures. AMCIS 2008 Proceedings, 361.
- Noor, K. B. M. (2008). Case Study: A Strategic Research Methodology. *American Journal of Applied Sciences*, 5(11), 1602–1604. doi:10.3844/ajassp.2008.1602.1604
- Qu, S., & Dumay, J. (2011). The Qualitative Research Interview. *Qualitative Research in Accounting and Management*, 8(3), 238–264. doi:10.1108/11766091111162070
- Qureshi, S. (2005). How does information technology affect development? Integrating theory and practice into a process model. AMCIS 2005 Proceedings, 261.
- Rabinet, S. E. (2011). How I Learned to Design and Conduct Semi-Structured Interviews: An Ongoing and Continuous Journey. *Qualitative Report*, 16(2), 563–566.
- Ross, J. W., Weill, P., & Robertson, D. C. (2006). *Enterprise Architecture As Strategy: Creating A Foundation For Business Executives*. Boston: Harvard Business School Press.
- Simonsson, M., & Johnson, P. (2006) Assessment of IT Governance - A Prioritization of Cobit. In *Proceedings of the Conference on Systems Engineering Research*, (pp. 1-10). Studentlitteratur.
- TOGAF version 9. (2009). The Open Group Architecture Framework (TOGAF). *The Open Group*, 1.
- Van Grembergen, W., & De Haes, S. (2007). *Implementing Information Technology Governance: Models, Practices and Cases*. Hershey, PA: Idea Group Publishing.
- Walsham, G. (1995). The Emergence of Interpretivism in IS Research. *Information Systems Research*, 6(4), 376–394. doi:10.1287/isre.6.4.376

Weill, P., & Woodham, R. (2002). *Don't Just Lead, Govern: Implementing Effective IT Governance*. MIT Sloan Working Paper No. 4237-02. Retrieved from <http://ssrn.com/abstract=317319>

Willis, J. W. (2007). *Foundations of qualitative research: Interpretive and critical approaches*. Thousand Oaks, CA: Sage Publications. doi:10.4135/9781452230108

Yin, R. K. (2003). *Case Study Research*. London: Sage Publications.

ADDITIONAL READING

Batyashe, T., & Iyamu, T. (2016). IT Governance: An Architectural Framework Based on Consolidated Best Practices. *Journal of Governance and Regulation*, 5(1), 7–15. doi:10.22495/jgr_v5_i1_p1

De Haes, S., Van Grembergen, W., & Debreceeny, R. S. (2013). COBIT 5 and enterprise governance of information technology: Building blocks and research opportunities. *Journal of Information Systems*, 27(1), 307–324. doi:10.2308/isisys-50422

Fengou, M. A., Mantas, G., Lymberopoulos, D., Komninos, N., Fengos, S., & Lazarou, N. (2013). A new framework architecture for next generation e-health services. *IEEE journal of biomedical and health informatics*, 17(1), 9-18.

Ferguson, C., Green, P., Vaswani, R., & Wu, G. H. (2013). Determinants of effective information technology governance. *International Journal of Auditing*, 17(1), 75–99. doi:10.1111/j.1099-1123.2012.00458.x

Klemm, D. M., & Chang, R. A. (2006). *U.S. Patent No. 7,124,413*. Washington, DC: U.S. Patent and Trademark Office.

Shaanika, I., & Iyamu, T. (2015). Deployment of Enterprise Architecture in The Namibian Government: The Use of Activity Theory to Examine the Influencing Factors. *The Electronic Journal of Information Systems in Developing Countries*, 71.

Tiwana, A., Konsynski, B., & Venkatraman, N. (2013). Special issue: information technology and organizational governance: the IT governance cube. *Journal of Management Information Systems*, 30(3), 7–12. doi:10.2753/MIS0742-1222300301

van Grembergen, W. (Ed.). (2007). *Implementing Information Technology Governance: Models, Practices and Cases: Models, Practices and Cases*. IGI Global.

Wilkin, C. L., & Chenhall, R. H. (2010). A review of IT governance: A taxonomy to inform accounting information systems. *Journal of Information Systems*, 24(2), 107–146. doi:10.2308/jis.2010.24.2.107

KEY TERMS AND DEFINITIONS

COBIT: Control Objectives for Information and Related Technologies is a framework that focuses on IT management and governance.

Enterprise Architecture: Systematic approach for governance of organisational artefacts.

Framework: Layered structure, which consist of social-technical factors.

IT Governance: Approach for control and management of IT artefacts.

ITIL: Information Technology Infrastructure Library, is a set of practices that focuses on IT services.

RedLeaf Communications: A South African based organisation that was used in the study.

TOGAF: Open Group Architecture Forum's.

Continuous Assurance and the Use of Technology for Business Compliance

Rui Pedro Figueiredo Marques
Universidade de Aveiro, Portugal

INTRODUCTION

The concept of Continuous Assurance began to arouse much interest when in 1999 a joint committee of the AICPA (American Institute of Certified Public Accountants) and the CICA (Canadian Institute of Chartered Accountants) took up the issue of Continuous Assurance and defined it as a set of services and technologies which enables independent auditors to provide written assurance on a subject issued simultaneously with, or a short period of time after, the occurrence of events underlying the subject (Vasarhelyi, Alles & Williams, 2010). Nowadays, the concept of Continuous Assurance is still relevant due its contribution to the current organizational context, in which there is a fierce competitiveness and a constant need for more timely, relevant and reliable information to support the decision making and achieve the strategic and operational objectives.

In this context of constant change and increasing competitiveness, seeking productivity gains and improving management tools have been a core priority. Thus, Continuous Assurance has been asserting itself and assuming an increasingly important role within organizations, with the function of management support and, in general, to ensure the economic and efficient use of resources, areas where the effects of the impact of new risk factors caused by the constant change, fierce competition and widespread access to global information are more felt (Morais, 2008).

Continuous Assurance and Auditing are sometimes used interchangeably. However, assurance is a much broader concept than auditing because whilst auditing is a systematic process of obtaining

and evaluating evidences on organizational data and transactions in order to verify their compliance to standards, policies, or rules, assurance additionally includes professional services that ensure quality of information or its context, for decision makers (Soltani, 2007).

This chapter provides the concept of Continuous Assurance, its objectives and components, and a model which allows both to evaluate information systems with Continuous Assurance services and to help design the requirements of new ones. Finally, some implementations are also presented providing a comprehensive understanding the state-of-the-art and the benefits of Continuous Assurance.

BACKGROUND

Continuous Assurance is defined as the application of emerging information and communication technologies to the standard techniques of auditing, both mandatory periodic auditing and internal auditing. In that view, Continuous Assurance is a new step in the evolution of transactional auditing from manual techniques to automated methods. The term “continuous” does not mean real time, but to be effective considering, respecting and being consistent with the pulse and rhythm of each organizational transaction and process (Vasarhelyi *et al.*, 2010).

Furthermore, Continuous Assurance emerges as a set of services which aims to restore the credibility of auditing, simultaneously allowing organizations to meet the requirements of regulations. Hence, it can diagnose the company's

viability and allegations of fraud and illegal acts, assessing the economy, efficiency and effectiveness of organizations (Murcia, Souza & Borba, 2008; Vasarhelyi *et al.*, 2010).

In 2006, a survey (PricewaterhouseCoopers, 2006) concluded that Continuous Assurance triggered corporate sensitivity to its adoption because in 2005 only 35% had a continuous auditing or monitoring processes in place or were planning to develop one, and this value increased to 50% in 2006. It is interesting to observe that 56% of respondents said their continuous auditing processes include both manual and automated elements, 41% indicated their processes are entirely manual, and 3% reported having fully automated processes.

Another study by Institute of Internal Auditors and ACL (2006) also showed similar results: 36% of surveyed organizations confirmed they implemented a Continuous Assurance approach in all their business processes or simply in some selected areas, and 39% intended to implement in the near future. However, it also states that regardless of the reasons that organizations may have had to neglect the continuing auditing in the past, the recent regulations, the stimulus for real-time monitoring and reporting of financial information and the ability to automate the traditional audit methods have strongly encouraged its adoption.

The issues discussed regarding the motivations which are driving Continuous Assurance include: the growth of complexity and amount of data, the growth of electronic exchange of information and outsourcing, the integration of the value chain; reports available on the Web, and the users' desire for reliable information and disclosed more frequently, more timely and more detailed; and the need for disclosure of updated information imposed by the Sarbanes Oxley Act (section 409) (Brown, Wong & Baldwin, 2007).

Brown, Wong and Baldwin (2007) reviewed more than 60 articles on this topic and concluded that Continuous Assurance was still a concept for most organizations and a goal for the future. The implementation of systems with continuous assurance services was residual when comparing with

continuous monitoring and continuous auditing systems. These latter systems presented high levels of maturation regarding their implementation in the inquired organizations.

According to the section "Implementations" of this chapter, we can see that the implementations designated as providers of Continuous Assurance services, in the period 2002-2015, were still few and with some limitations regarding the diversity of services that Continuous Assurance should be expected to offer. This shows that this area is still developing and maturing.

CONCEPT

The concept of Continuous Assurance refers to the set of services which, by means of technology, uses the information immediately and produces audit results simultaneously with or within a short period of time after the occurrence of relevant events. Furthermore, it allows analytical monitoring of business processes. It is intended to be timely, more comprehensive, more accurate and more supportive to management than the traditional auditing (Alles, Kogan & Vasarhelyi, 2003, 2004; Vasarhelyi, Alles & Kogan, 2004).

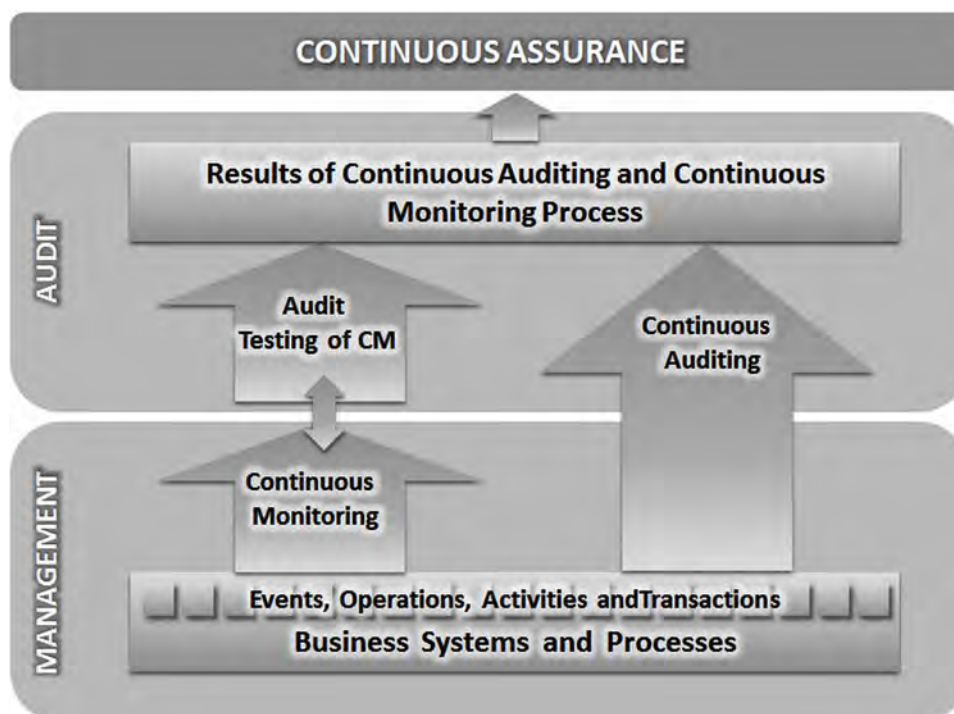
Moreover, Continuous Assurance has provided a change in the auditing practice for the maximum possible degree of automation. Given the emphasis on the transformation of the entire auditing system, the development of Continuous Assurance requires a fundamental reassessment of all aspects of auditing, in particular on how data is made available to the auditor, how alerts are managed, what kind of reports are issued, and how often and to whom they are sent (Vasarhelyi *et al.*, 2010).

According to IFAC (2004), the subject matter, and subject matter information, of Continuous Assurance can take many forms, such as:

- Financial performance or conditions (for example, historical or prospective financial position, financial performance and cash

Figure 1. Continuous monitoring, auditing and assurance

Source: adapted from Coderre, 2005



flows) for which the subject matter information may be the recognition, measurement, presentation and disclosure represented in financial statements;

- Non-financial performance or conditions (for example, performance of an entity) for which the subject matter information may be a key indicator of efficiency and effectiveness;
- Physical characteristics (for example, capacity of a facility) for which the subject matter information may be a specifications document;
- Systems and processes (for example, an entity's internal control or IT system) for which the subject matter information may be an assertion about effectiveness; and
- Behavior (for example, corporate governance, compliance with regulation, human resources practices) for which the subject matter information may be a statement of compliance or a statement of effectiveness.

When Continuous Assurance services are running in real time, they provide the opportunity to manage transactions simultaneously with their execution or after a particular event, and even in some cases, the ability to interfere with the completion of the transaction, correcting it. When some service of Continuous Assurance points a wrong transaction, and this is corrected by the auditor, we can say the system becomes a proactive player in the processing of organizational information (Vasarhelyi *et al.*, 2004).

The terms Continuous Assurance, Continuous Auditing and Continuous Monitoring are sometimes used indiscriminately in literature. Therefore, it is crucial to understand what characterizes and distinguishes them, but mostly to understand how they can be relate to, complementing each other (Figure 1).

Continuous Assurance is a statement on the adequacy and effectiveness of controls and integrity of information. Continuous monitoring of controls is at the center of Continuous Assurance

strategies; however, the audit activity ensures that management activities are appropriate and effective so that organizations have a greater level of certainty about the effective operation of controls, about risk management and about the integrity of information used for decision making (Alles *et al.*, 2003; Coderre, 2005; Kuhn & Sutton, 2006).

In turn, Continuous Auditing refers to activities undertaken to provide warranty and credibility to operations, besides giving a more timely character to issues of control and management of risk. Continuous Monitoring is responsible for constantly monitoring and evaluating business transactions and their related controls, enabling a real-time view on effectiveness of controls and on integrity of transactions (Minnaar, Little & Farineau, 2008).

Evaluating the combined results of Continuous Monitoring and audit procedures, auditors can provide Continuous Assurance. However, only recently has an effort been made to understand their differences and distinguish between the concepts.

OBJECTIVES

The objectives of Continuous Assurance, making it really advantageous, are divided into four levels, which are difficult to define in a mutually exclusive way, but which serve to illustrate the functional dependence of Continuous Assurance on auditing (Vasarhelyi *et al.*, 2004).

Level 1: Evaluation of Organizational Transactions

At this level, it is intended to evaluate organizational transactions, and analyze and verify the atomic actions of transactions execution (for example, movement of money or information to the level of data that compose it). With the use of corporate systems, such as ERP (Enterprise Information System) systems, it is possible to analyze, aggregate and evaluate data in order to classify and monitor organizational transactions.

At this level, input data should be tested in order to verify whether they are valid, and whether procedures in execution of transaction are consistent with the sequence of established operations, for example check whether a certain movement in the stock corresponds to a sale, to a certain bill, and whether it is in accordance with the order received. Contrary to traditional auditing, with Continuous Assurance it is possible to do this check in real time due to automation and integration of audit procedures. This transaction control can use the formal specification of workflow of processes defined in the ERP systems as a standard behavior of transaction. Thus, it is possible to verify whether transactions have been executed in compliance with all provided steps to foresee the flow of transactions, and whether these are missing or failing.

Level 2: Compliance of Performed Operations

At this level it is intended to ensure that procedures applied in the execution of organizational transactions are appropriate (for example, are consistent with the rules, norms, or standards set by the organization or by external regulatory entities).

The procedures of this level use comparison with standards in order to verify the application of these rules and standards, and if any inconsistency is found, the situation will be forwarded to the auditors or other responsible person for consideration. Although in this latter case the resolution of the problem is not guaranteed in real time, this selectivity will allow greater efficiency in the auditors' tasks.

Level 3: Quality of Estimates and the Consistency of Aggregate Data

In some businesses, estimates and forecasts are often used because the measurement or direct determination of some information is difficult and expensive to obtain. For example, the percentage

of completion of a work is not always easy to measure, and therefore, must be estimated.

Even if an organization does not automatically generate an estimate, continuous assurance services can use their own formal models to ensure in real time that the estimate used is acceptable. Obviously, the creation of a formal model of an accounting estimate is not a simple proposition and can add significant costs to the development of Continuous Assurance. A simple and economical alternative is to use a formal model to automatically get an estimate, then the auditor's task will be reduced only to verifying the acceptability of this model, which can be done only once and off-line, based on knowing whether the used parameter values in model are reasonable.

This level of Continuous Assurance includes automation of analytical procedures based on internal and external parameters. The use of analytical procedures in an automated system of Continuous Assurance increases efficiency and effectiveness of auditing.

Level 4: Evaluation of Organizational Decisions

The auditing carried out by using ERP systems and advanced financial instruments must incorporate complex and high level assessments, which are especially important for decision making. Such judgments may have to deal, for example with the relevance of contingencies, the extent of related-party transactions, the boundaries of corporate systems, and the nature of the relationships across the value chain. Continuous Assurance and the current analytic technology allow the extensive gathering of exogenous evidence, which provides crucial input into these judgments. Continuous Assurance may use, for example data warehousing and data mining as tools that facilitate automation of some of these decisions, improving the quality of high-level decisions and decreasing the audit risk.

CONSTITUTION

Continuous Assurance is divided into three distinct components, but complementary (Vasarhelyi *et al.*, 2010):

- **Continuous Controls Monitoring (CCM):** Consists of a set of procedures to monitor the operation of internal control mechanisms;
- **Continuous Data Assurance (CDA):** Verifies the integrity of the data circulating in organizational information systems;
- **Continuous Risk Monitoring and Assessment (CRMA):** Measures the risk dynamically and allows to sustain an auditing plan.

Continuous Controls Monitoring

CCM is a set of emerging technologies which monitors controls in ERP systems and other applications in order to improve business governance, monitor and verify access and transactional rules, and to automate audit processes. CCM assists the business in: reducing business losses from fraud or failure; following rules or procedures which govern transactions; and improving performance through monitoring and auditing of the controls implemented in the systems (Caldwell & Proctor, 2010).

Caldwell and Proctor (2010) identified many important CCM functionalities, namely controls monitoring functions, exception and remediation management, reporting and analytics, and workflow. With these functionalities, CCM adds value, in an organization, to risk management and to the following compliance initiatives:

- **Lowering Compliance Costs:** CCM reduces the cost of auditing by eliminating some manual sampling and minimizing the time it takes to gather documentation;

- **Improving Financial Governance:** CCM can increase the reliability of transactional controls, improve auditor trust and increase the effectiveness of antifraud controls; and
- **Improving Operational Performance:** CCM controls, such as those that monitor duplicate payments, incorrect discounts or misapplied warranties, go beyond what most people consider compliance.

Continuous Data Assurance

CDA has been a powerful tool in organizations, particularly in financial and accounting areas, to extract data from organizational databases and applications in order to make analyses at the transactional level and to provide more detailed assurance. It has the main objective to change the traditional vision of data auditing in which reports were prepared and audited only once a year; sampling was used rather than examining the entire population; data were analyzed at the trial balance level using ratios.

This aggregation and analysis at a higher level than the transactional level has been a cost and capability-based limitation rather than the ideal process for assurance. Thus, CDA can continuously monitor transactions, comparing their characteristics with expected behaviors, thus identifying anomalous situations. When substantial divergences occur, alarms will be triggered and routed to the respective stakeholders (Vasarhelyi *et al.*, 2010).

To implement transaction verification it is necessary to specify data validity, consistency and referential integrity rules which are then used to filter data. These rules aim to detect and remove two types of data errors: data integrity violations (for example, invalid purchase quantities, receiving quantities and bank check numbers) and referential integrity violations which are largely caused by many unmatched records among different business processes (for example a payment made for a non-existent purchase order).

Continuous Risk Monitoring and Assessment

B

CRMA is the process by which organizations address the risks attaching to their activities in pursuit of organizational objectives and across the set of all their activities. Besides, CRMA involves: measuring risk factors continuously, integrating different risk scenarios into some quantitative framework, and providing inputs for audit planning, i.e. it engages risk assessment, risk evaluation, risk treatment and risk reporting. The organizational risk management approach is intended to align risk management with business strategy and embed a risk management culture into business transactions (Collier, Berry & Burke, 2007; Vasarhelyi *et al.*, 2010).

INFORMATION SYSTEMS WITH CONTINUOUS ASSURANCE SERVICES

Marques, Santos and Santos (2013) proposed a model composed of three dimensions and one requirement, which aims to evaluate an information system with Continuous Assurance services, and which intends to be regarded as a referential set of requirements when developing this type of information systems. The authors used the Delphi method to validate the model in order to ensure the relevance of inclusion of these dimensions and requirement and also a set of metrics to be included in each dimension and requirement (Marques, Santos & Santos, 2016). Furthermore, this model was successfully used in a more comprehensive research project (Marques, Santos & Santos, 2015).

It is composed of the dimensions Monitoring, Compliance, Estimation and the requirement Reporting, which comprise all objectives and components of a system with Continuous Assurance services (Marques *et al.*, 2013, 2016):

- Dimension ‘Monitoring’ relates with the objectives of level 1 and the component CCM. In this dimension, the following metrics may be found to assess whether the system can: monitor the various operations of a process as soon as they occur; identify an irregular (unforeseen or inconsistent) operation as soon as it occurs; verify whether the operations were processed at all the previous steps as required; detect lack of operations; and assess the continuity and completeness of transactions.
 - Dimension ‘Compliance’ includes the features of the component CDA and the objectives of level 2. The metrics associated to this dimension aims to assess whether the system can: recognize which known execution pattern was or has been followed by each organizational transaction monitored; ascertain which rules, conditions and procedures were fulfilled and unfulfilled in the organizational transactions monitored; detect potential errors; inhibit inappropriate events or behaviors; and help compliance with existing laws, policies, norms and procedures.
 - Dimension ‘Estimation’ covers the objectives of level 3 and the functions of the CRMA component. The metrics of this dimension assess whether the system can: estimate, given the current situation, what the possible results of the organizational transaction execution will be; and determine the execution pattern, or a set of execution patterns, which will be possible to be followed by the organizational transactions monitored, according to the current status of execution.
 - Requirement ‘Reporting’ includes the features of all components and the objectives of all levels. The metrics of this requirement must assess whether the system can: report the results of the monitoring of transactions; notify the results of the verification of compliance; inform the results of estimation; and alert users of irregular situations in monitoring, compliance verification and estimation of negative results.
- Table 1 concisely summarizes this model, presenting quantitatively measurable metrics for each dimension and requirement.

Table 1. Dimensions of the model and their metrics

Dimensions and Requirement	Metrics
Monitoring (dimension)	– Real-time monitoring of operations – Real-time identification of irregular operations – Real-time verification of processing of required operations at all previous steps – Real-time detection of lack of operations
Compliance (dimension)	– Recognition of execution patterns – Ascertaining of fulfilling of rules – Detection of potential errors – Verification of compliance of existing policies
Estimation (dimension)	– Estimation of possible results – Determination of possible execution patterns which are likely to be followed
Reporting (requirement)	– Real-time presentation of the executed operations which were monitored – Real-time presentation of execution patterns which are being followed or are likely to be followed – Real-time presentation of the compliance verification in transactions executions – Real-time presentation of the risk estimated on determining possible execution patterns – Real-time alert for irregular situations in monitoring, compliance verification and estimation of negative results

IMPLEMENTATIONS

Some solutions in the field of Continuous Assurance have been made available in the market and are mostly developed by well-established CAAT (Computer Assisted Auditing Techniques) manufacturers and vendors, which begin to position themselves in this still emerging market. Despite regarded as Continuous Assurance solutions by vendors, these products focus more on the CDA component. The implementation of CCM solutions is very complex due to significant differences in the types of business objects, process configurations and controls and then it is not cost-effective. Thus, these solutions have developed some very specific CCM modules targeted at specific enterprise systems. Such modules are quite laborious since they require customization for each new enterprise system (Vasarhelyi *et al.*, 2010).

An organization in the steel industry with 234 industrial and commercial units, present in 10 countries and with shares in New York Stock Exchange, was forced to implement measures of Continuous Assurance in accordance with SOX requirements. After these measures, all processes had to be documented and information started to be disclosed in time and with their integrity ensured and it held the participants responsible, inhibiting the occurrence of fraud (Hargadon & Fanelli, 2002).

Siemens has made an effort in this area and it has successfully experimented some aspects of Continuous Assurance, namely to ensure the integrity of its ERP systems and their automated modules with audit functions. Siemens's project proposes a methodology to monitor and evaluate the daily configuration of various settings of existing controls through the use of CCM, because about 68% of its audit activities could be fully automated. Thus, an independent system was developed for interacting with read-only operations with the existing information systems and for reporting system and alarms in case of emergency. These implementations have ensured the effectiveness, efficiency and a timely character

to the audit procedures. (Alles, Brennan, Kogan & Vasarhelyi, 2006; Alles, Kogan & Vasarhelyi, 2008). The use of CCM at Siemens Financial Services enabled significant improvements in the average exception rate, through the testing and monitoring of performance analytics (e.g. input checks, validity checks, and compliance with regulations and internal policies). In some departments, the reduction in the average exception rate was over 20% in the first year of implementation (Medinets, Gross, & Brennan, 2015).

Also according to Vasarhelyi *et al.* (2010), one of the biggest banks in Brazil, with more than 1400 branches, has implemented measures for Continuous Assurance, proving the viability and benefits of these measures. This institution has a CDA, a system which daily analyses more than five million accounts and generates about six thousand alerts a month. This system aims to increase productivity with efficiency and quality and its mission is to assess the risks and controls automatically and continuously in order to identify exceptions, anomalies, trends and indicators of risk; advise about controls, risk assessment; and contribute to the corporate governance. These features include all products, processes and services which enable data extraction and analysis. The taken approaches are of detection (routines to detect possible errors), of deterrence (routines to inhibit inappropriate behavior and events), financial (routines to reduce or avoid losses), and of compliance (routines to ensure compliance with applicable laws, policies and standards).

Recently, Marques *et al.* (2015) developed an innovative solution to implement Continuous Assurance services in information systems applicable to any business process, regardless of its type, dimension, business area or even its information system support technology, supported by an ontological model at an abstraction level that guarantees that contextual independence. The results analysis from an implementation of this solution allowed to ensure the feasibility and the effective use of the solution.

FUTURE RESEARCH DIRECTIONS

The limitations of Continuous Assurance that can be found in the chapter are likely to be the next future research topics in this area. A major limitation is in the CCM, because its implementation is complex due to the variety of control mechanisms it has to monitor in different business processes and contexts. A CCM which can be independent of the business processes and contexts was already implemented (Marques et al., 2015) although it is still very dependent on control mechanisms and on the initial system customization, which is a time-consuming and very hard task. Guerreiro, Marques and Gaaloul (2016) present a high-level conceptualization of an evolvable risk-based approach capable of self-learning the needs of monitoring, based on the continuous observation of business processes and of the results of continuous monitoring (approach with feedback control). This type of approaches has promising results and may be a future research direction.

Furthermore, a conceptualization of a model to assess the maturation of Continuous Assurance seems to be relevant to evaluate the services provided by the emergent solutions and, simultaneously, to follow the evolution of Continuous Assurance.

CONCLUSION

Although Continuous Assurance is not a concept firmly established yet, it is clear that it is maturing both in practice and in research because of its proven benefits in various organizational branches. In the public sector, the health service, social security institutions, tax administration, banking can be stated as examples. In the private sector, any organization can benefit from the implementation of Continuous Assurance, although those with greater resources are more able to bear it and keep it, for example, the areas of banking and telecommunications.

Furthermore, recent researches have proved the feasibility of implementation of information systems with full services of Continuous Assurance, covering all components and levels of objectives, to any business area and any type, dimension and nature of business process, providing very positive outcomes to the operational performance and accountability of organizations.

REFERENCES

- ACL. (2006). *New demands, new priorities - the evolving role of internal audit. Global Audit Executives Survey Report*. ACL Services Ltd.
- Alles, M., Brennan, G., Kogan, A., & Vasarhelyi, M. (2006). Continuous monitoring of business process controls: A pilot implementation of a continuous auditing system at Siemens. *International Journal of Accounting Information Systems*, 7(2), 137–161. doi:10.1016/j.accinf.2005.10.004
- Alles, M., Kogan, A., & Vasarhelyi, M. (2003). Black box logging and tertiary monitoring of continuous assurance systems. *Information Systems Control Journal*, 1, 37–39.
- Alles, M., Kogan, A., & Vasarhelyi, M. (2004). Restoring auditor credibility: Tertiary monitoring and logging of continuous assurance systems. *International Journal of Accounting Information Systems*, 5(2), 183–202. doi:10.1016/j.accinf.2004.01.010
- Alles, M., Kogan, A., & Vasarhelyi, M. (2008). Putting continuous auditing theory into practice: Lessons from two pilot implementations. *Journal of Information Systems*, 22(3), 195–214. doi:10.2308/jis.2008.22.2.195
- Brown, C. E., Wong, J. A., & Baldwin, A. A. (2007). A review and analysis of the existing research streams in continuous auditing. *Journal of Emerging Technologies in Accounting*, 4(1), 1–28. doi:10.2308/jeta.2007.4.1.1

- Caldwell, F., & Proctor, P.E. (2010). *Magic quadrant for continuous controls monitoring*. Gartner, Inc.
- Coderre, D. (2005). Continuous auditing: Implications for assurance, monitoring, and risk assessment. In IIA (Ed.), *Global technology audit guide*. The Institute of Internal Auditors.
- Collier, P. M. M., Berry, A., & Burke, G. T. T. (2007). *Risk and management accounting - best practice guidelines for enterprise-wide internal control procedures*. CIMA Publishing.
- Guerreiro, S., Marques, R. P., & Gaaloul, K. (2016). *Optimizing business processes compliance using an evolvable risk-based approach*. Paper presented at 49th Hawaii International Conference on System Sciences (HICSS). doi:10.1109/HICSS.2016.699
- Hargadon, A., & Fanelli, A. (2002). Action and possibility: Reconciling dual perspectives of knowledge in organizations. *Organization Science*, 13(3), 290–302. doi:10.1287/orsc.13.3.290.2772
- IFAC. (2004). Auditing and assurance. In IFAC (Ed.), *Handbook of international auditing, assurance and ethics pronouncements* (Vol. 1). New York: International Federation of Accountants.
- Kuhn, J. R., & Sutton, S. G. (2006). Learning from worldcom: Implications for fraud detection through continuous assurance. *Journal of Emerging Technologies in Accounting*, 3(1), 61–80. doi:10.2308/jeta.2006.3.1.61
- Marques, R. P., Santos, H., & Santos, C. (2013). A conceptual model for evaluating systems with continuous assurance services. *Procedia Technology*, 9, 304–309. doi:10.1016/j.protcy.2013.12.034
- Marques, R. P., Santos, H., & Santos, C. (2015). Monitoring organizational transactions in enterprise information systems with continuous assurance requirements. *International Journal of Enterprise Information Systems*, 11(1), 13–32. doi:10.4018/ijeis.2015010102
- Marques, R. P., Santos, H., & Santos, C. (2016). Evaluating information systems with continuous assurance services. *International Journal of Information Systems in the Service Sector*, 8(3), 1–15. doi:10.4018/IJISS.2016070101
- Medinets, A. F., Gross, J. A., & Brennan, G. (2015). Case study A – Developing continuous assurance at Siemens. In AICPA (Ed.), *Audit analytics and continuous audit – looking toward the future*. New York: American Institute of Certified Public Accountants, Inc.
- Minnaar, D., Little, J., & Farineau, D. (2008). *Continuous auditing and continuous monitoring: Transforming internal audit and management monitoring to create value*. New York: KPMG LLP.
- Morais, M. G. C. T. (2008). *A importância da auditoria interna para a gestão: Caso das empresas portuguesas* [The importance of internal auditing for management: Case of Portuguese enterprises]. Paper presented at the 18º Congresso Brasileiro de Contabilidade Gramado, Brazil.
- Murcia, F.D.-R., Souza, F.C.d., & Borba, J.A. (2008). Continuous auditing: A literature review. *Organizações em contexto*, 4(7), 1–17.
- PricewaterhouseCoopers. (2006). *State of the internal audit profession study: Continuous auditing gains momentum*. PricewaterhouseCoopers LLP.
- Soltani, B. (2007). *An introduction to auditing and assurance*. In *Auditing: An international approach* (Vol. 1, pp. 1–26). FT Prentice Hall.
- Vasarhelyi, M., Alles, M., & Kogan, A. (2004). Principles of analytic monitoring for continuous assurance. *Journal of Emerging Technologies in Accounting*, 1(1), 1–21. doi:10.2308/jeta.2004.1.1.1
- Vasarhelyi, M., Alles, M., & Williams, K. T. (2010). *Continuous assurance for the new economy* (1st ed.). Sydney: Institute of Chartered Accountants in Australia.

ADDITIONAL READING

Marques, R. P. (2014). *Organisational transactions with embedded control*. Unpublished PhD. Guimarães, Portugal: University of Minho.

Marques, R. P., Santos, H., & Santos, C. (2013). *An enterprise ontology-based database for continuous monitoring application*. Paper presented at the IEEE 15th Conference on Business Informatics (CBI), Vienna, Austria. doi:10.1109/CBI.2013.10

Marques, R. P., Santos, H., & Santos, C. (2014). *Management of internal control mechanisms in erp for continuous monitoring purposes*. Paper presented at the 9th Iberian Conference on Information Systems and Technologies (CISTI), Barcelona, Spain. doi:10.1109/CISTI.2014.6877086

Marques, R. P., Santos, H. M. D., & Santos, C. (2013). Organizational transactions with real time monitoring and auditing. *The Learning Organization*, 20(6), 390–405. doi:10.1108/TLO-09-2013-0048

KEY TERMS AND DEFINITIONS

Continuous Assurance: Set of services which, making use of technology, uses the information immediately and produces audit results simultaneously or within a short period of time after the occurrence of relevant events.

Continuous Auditing: Activities undertaken to assess the business compliance to rules, policies or legislation, providing warranty and credibility to operations.

Continuous Monitoring: Process responsible for the management of internal control mechanisms in order to ensure their effectiveness, allowing to obtain timely information about transactions.

Explaining and Predicting Users' Continuance Usage Intention Toward E-Filing Utilizing Technology Continuance Theory

B

Santhanamery T.

Universiti Teknologi MARA, Malaysia

T. Ramayah

University Sains Malaysia, Malaysia

INTRODUCTION

Continuance usage intention is vital in today's environment because the longer a business firm can keep a customer; the greater the life-time revenue from that customer and at the same time the cost of serving the customer declines (Mittal & Lassar, 1998). Besides that the high level of user continuance usage intention is also a reflection that the program or the product is well designed and implemented (Lin, Chen & Fang, 2011) and it ensures the continued existence of many customer based electronic commerce firms (Bhattacharjee, 2001). Moreover, although initial acceptance is important in recognizing the success of an information system (Bhattacharjee, 2001) but continued usage is even more significant in ensuring the long-term viability of technology innovations (Premkumar & Bhattacharjee, 2008). Furthermore, Devaraj and Kohli (2003) argued that the long term usage of a technology will enhance the financial and quality performance of an organization. Thus, it is important for businesses to accentuate on continuance usage intention as the key for long term growth.

Guided by Vision 2020, Malaysia has embarked on an ambitious plan by launching the Multimedia Super Corridor (MSC) in August 1996. Seven specific flagship applications were identified as the pioneering MSC projects, which

includes e-government as one of the flagships (Muhammad Rais & Nazariah, 2003). The Vision of e-government is to transform administrative process and service delivery through the use of ICT and multimedia (Lean, Zailani, Ramayah and Fernando, 2009). The projects under the e-Government flagship have been started since ten years ago aimed at building a more effective and efficient way to communicate and transact with the citizens and industries. One of the projects under e-government flagship is Online Tax System or e-Filing (Hussein, Mohamed, Ahlan, Mahmud & Aditiawarman (2010).

BACKGROUND

The e-Filing system in Malaysia which was introduced in 2006 by IRBM is receiving much attention and there has been an upward trend in the adoption of the system among taxpayers with the latest income tax submission of 2,330,298 via e-filing in 2014 (Annual Report, 2014). Among the factors that could have contributed to this increase are convenience, faster refund and cheaper cost. However, according to Bhattacharjee (2001), while initial acceptance of information system (IS) is very important toward realizing IS success but its eventual success depend on its continued use rather than first-time use.

In this vein, investigating the continuance usage intention of e-filing system is deemed to be important because as more citizens use e-filing services, the more operation and management costs are reduced (Wangpipatwong, Chutimaskul & Papasratorn, 2008). In Malaysia, IRBM has saved millions of ringgit annually by reducing cost of printing, imaging, postal and storage through their e-filing system. In 2009, a total cost of RM9, 162,845.92 has been saved via the e-filing submission of tax (Hasmah, 2009). Apart from that, Bhatnagar (2009) reveals that governments are spending millions of dollars to build online service delivery portals in terms of hardware, software, training and maintenance and communication infrastructure. In developing the e-filing system in Malaysia, millions of ringgit has been invested (Aziz & Idris, 2012; Azmi & Kamarulzaman, 2010) especially to upgrade the agency's computer hardware and software (Bernama, 2005). Therefore, in ensuring that the heavy investments invested in developing the e-filing online portals will not be wasteful, identifying the factors that will motivate the continuance usage intention is crucial.

As such, this paper has focused in determining the effect of two main central construct in determining the individual adoption; satisfaction and attitude, towards continuance usage intention.

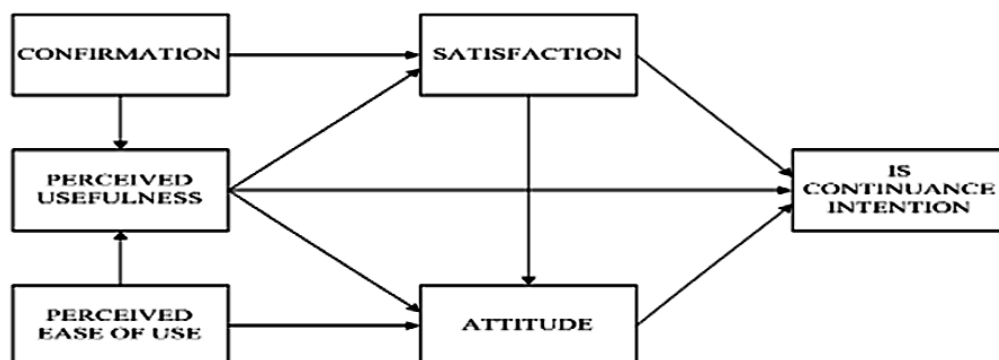
LITERATURE REVIEW

Technology Continuance Theory

Technology Continuance Theory (TCT) by Liao, Palvia and Chen (2009) is the newest theory introduced in the context of predicting the individuals continuance usage behavior; This new theory is actually a combination of three previous theoretical models which has been widely used in predicting the user behavior in IS acceptance and continuance, namely the Technology Acceptance Model (TAM), Expectation Confirmation Model (ECM) and Cognitive Model (COG).

The major contribution by TCT according to Liao et al. (2009) is it combined two main central constructs; satisfaction and attitude; into one continuance model. So, it is applicable for all levels of adoption such as initial, short term and long term users. TCT is believed to represent an improvement over TAM, ECM and COG models in terms of applicability and explanatory power. Based on the research done by Liao et al. (2009), they found that although satisfaction has a temporary effect on behavioral intention but it is still subjected for rejection or willingness to continuously use a system in short term. However, they also revealed that the success of information system in long term is determined by users' attitude.

Figure 1. Theoretical framework



Relationship Between Confirmation, Satisfaction and Perceived Usefulness

It has been found that a positive relationship exists between confirmation and satisfaction and perceived usefulness towards continuance usage intentions by many previous studies. Hwang, Yu, Tsai and Lin (2011) investigated the effect of confirmation on satisfaction and perceived usefulness on the continuance usage intention of e-portfolio learning system and found that confirmation has a positive effect on perceived usefulness and satisfaction. Alike, Lee and Kwon (2010) studied the effect of confirmation on satisfaction and perceived usefulness in determining continuance intention of web based services and found that confirmation positively influence users satisfaction and perceived usefulness. The study also shows that the effect of confirmation on satisfaction is much higher than perceived usefulness. Similarly, confirmation of expectation was found to be positively associated with perceived usefulness and user satisfaction in determining the mobile data services continuance intention among university students in South Korea (Kim, 2010). As such, it is hypothesized that:

H1: Confirmation has a direct positive relationship on satisfaction towards e-filing continuance usage intentions.

H2: Confirmation has a direct positive relationship on perceived usefulness towards e-filing continuance usage intentions.

Relationship Between Perceived Usefulness and Satisfaction

Previous researchers have confirmed a significant positive relationship between perceived usefulness and satisfaction towards continuance usage intention. Islam (2012) based on the study on the motivation to continue e-learning system use found that perceived usefulness has a significant positive impact on satisfaction towards continu-

ance intention. Hung, Chang and Hwang (2011) in exploring the continuance intention of the web based learning system found that the perception of perceived usefulness of the system has a positive influence on satisfaction towards continuance intention. Further, Shih, Shiau and Huang (2010) supported the previous study and disclosed that perceived usefulness has a significant relationship towards satisfaction in determining the bloggers continuance intention to engage in blogging activities. Correspondingly, Kang and Lee (2010) found perceived usefulness as the most important antecedent of customer satisfaction towards continuance intention of online services. As such, it is hypothesized that:

H3: Perceived usefulness has a direct positive relationship on satisfaction towards e-filing continuance usage intentions.

Relationship Between Perceived Usefulness and Attitude, Perceived Ease of Use and Attitude and Perceived Ease of Use and Perceived Usefulness

It is found that significant positive relationship exists between perceived ease of use and perceived usefulness and attitude which is proven by previous researches in post adoption environment. Lin (2011) based on their study reveals that perceived ease of use has a significant effect on perceived usefulness which in turn affects the attitude towards e-learning. The study also found that perceived ease of use has a greater effect on attitude toward e-learning for users with limited learning experience while perceived usefulness has a more effect on attitude for experienced users. This is supported by Suki and Ramayah (2010) whose study found that when a system is perceived as useful the user will form a positive attitude towards the system and its effectiveness. Similarly another study on e-learning investigated the consumer's continuance intention and found that perceived usefulness and perceived ease of

use significantly related to attitude of consumers in determining the continuance usage intention. Subsequently, the research also found a positive relationship between perceived ease of use and perceived usefulness of e-learning (Lee, 2010). Further, study by Kim, Kim and Shin (2009) indicates that perceived ease of use is a strong predictor of perceived usefulness in the use of airline B2C e-Commerce websites (AB2CEWS). It further found a positive relationship between perceived ease of use and attitude and perceived usefulness and attitude towards use and re-use of AB2CEWS with a more prominent effect of perceived ease of use towards attitude. As such, it is hypothesized that:

H4: There is a direct positive relationship between perceived usefulness and attitude.

H5: There is a direct positive relationship between perceived ease of use and attitude.

H6: There is a direct positive relationship between perceived ease of use and perceived usefulness.

Relationship Between Perceived Usefulness and Continuance Usage Intention

Perceived usefulness was correlated with all technology usage. Research by Brahmasrene and Lee (2012) on online learning found that perceived usefulness strongly influence the intention to continue using online learning. Hung et al. (2011) found that perceived usefulness has a positive relationship in determining the continuance intention towards web based learning system. Similarly, Al-maghrabi, Dennis and Halliday (2011) also revealed that perceived usefulness is the main determinant of continuance intention of e-shopping in Saudi Arabia. Correspondingly, Shiau, Huang and Shih (2011) examined the continuance intention of blog users and reported that perceived usefulness of the blogs positively influence the blogger's intention to continually use blogs. As such, it is hypothesized that:

H7: Perceived usefulness has a direct positive relationship towards e-filing continuance usage intention.

Relationship Between Satisfaction and Continuance Usage Intention

Strong relationship between customer satisfaction and continuance usage intention has been established previously by numerous researches. Choi, Park and Park (2011) found that satisfaction has a positive effect on the tourist continual usage intention of mobile tourism information services. Similarly, continuance intention was also found to be positively affected by satisfaction in determining micro blogging service continuance intention (Zhao & Lu, 2012). Ong and Day (2010) found that the outcome of user satisfaction with Social Media Service such as You tube and Facebook is the continued usage of these services. Satisfaction also was found to be the strongest predictor of users' continuance intention of Enterprise Resource Planning system compared to individual differences such as personal innovativeness, computer anxiety and computer self-efficacy (Chou & Chen, 2009). As such, it is hypothesized that:

H8: Satisfaction has a positive significant relationship towards e-filing continuance usage intention.

Relationship Between Satisfaction and Attitude

Oliver (1980) viewed satisfaction as the main influence for post purchase attitude. Huang and Hsu (2009) found that the more satisfied the respondents with their past experience, the more favorable attitude towards revisiting Hong Kong, means that satisfaction on past visit experience had a positive effect on tourist attitude to revisit. Likewise, Mosavi and Ghaedi (2012) found that customer satisfaction positively influenced positive attitude. Further, according to Sivadas and Prewitt (2000), the key to customer retention is

favourable attitude and the favourable attitude can be created by satisfying the customers. Similarly, study on examining the critical factors that influence the learner's satisfaction on e-learning reported that instructor's attitude towards e-learning significantly affect the e-learners satisfaction (Sun, Tsai, Finger, Chen & Yeh, 2008). As such, it is hypothesized that:

H9: Satisfaction has a direct positive relationship on attitude towards e-filing continuance usage intention.

Relationship Between Attitude and Continuance Usage Intention

Studies on continuance intention of a particular technology have proved the significant relationship between attitude and continuance intention. Hsiao and Chou (2012) investigated the effect of attitude on the continuance usage intention of online gaming and found that attitude significantly influences the continuance intention of online gaming. Lee (2010) and Ho (2010) further found that attitude is significant predictor of users' continuance intention toward e-learning. Indeed, attitude was found to be a stronger influence on continuance intention compared to other antecedents in shaping continued ICT usage intention (Hsieh, Rail & Keil, 2008). The study also suggested that individuals who have usage experience, attitude will have a stronger impact on continuance intention. As such, it is hypothesized that:

H10: Attitude has a direct positive relationship towards e-filing continuance usage intention.

METHOD

Sampling

The non-probability purposive sampling was employed in this study. Purposive sampling or particularly judgment sampling occurs when re-

searchers select respondents to conform to some criterion (Cooper & Schindler, 2008). In this study, the sample selected were taxpayers who had used the e-filing system before at least once as the measures required them to express their willingness to continually use the e-filing system.

Settings

A total of 900 questionnaires were distributed among the taxpayers in Selangor and Kuala Lumpur, Malaysia using self-administered questionnaire. A total of 401 questionnaires were returned and out of it, 355 were completed whereas the other 46 were incomplete. As such, the response rate was 44.5%. The questionnaire consists of 8 sections. The first section elicited the screening question, the second section collected the demographic data, the third section extracted information on confirmation, section four measured the perceived usefulness, section five measured the perceived ease of use, section six measured the satisfaction, section seven measure the attitude and last section measured continuance intention.

Measures

The measures were all adapted from the published literature. The measures for continuance intention were from Bhattechejee (2001). Perceived Usefulness and Perceived Ease of Use were from Davis (1989). Confirmation was adapted from Bhattacharjee (2001). Satisfaction was adapted from Spreng, Mackenzie and Olshavsky (1996) and Liao et al. (2009). Attitude was adapted from Taylor and Todd (1995) and Liao et al. (2009).

Sample Profile

The demographic of the respondents tabulated in Table 1 were derived from descriptive analysis. The majority of the age group (23.3%) was in the category of 30-34 years old. Male (63.8%) outnumbered the females (36.2%). In terms of ethnicity, the majority of the respondents were

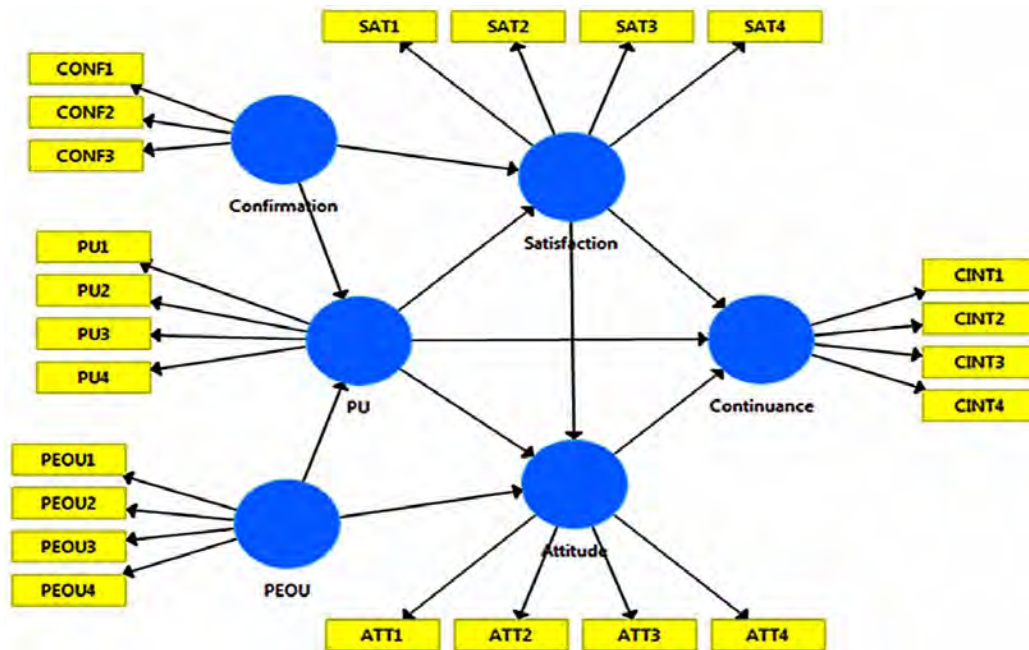
Table 1. Demographic of the Respondents

Age	Frequency	Percent
20-24 years	3	0.80
25-29 years	41	11.50
30-34 years	68	19.20
35-39 years	70	19.70
40-44 years	30	8.50
45-49 years	62	17.50
50-54 years	57	16.10
55 years and above	24	6.80
Gender		
Male	116	32.70
Female	239	67.30
Ethnicity		
Malay	276	77.70
Chinese	42	11.80
Indian	33	9.30
Others	4	1.10
Education		
Diploma College	34	9.60
Bachelor Degree	87	24.50
Masters Degree	146	41.10
Doctoral Degree	63	17.70
Others	24	6.80
Income		
RM2000 - RM2999	36	10.1
RM3000 - RM3999	83	23.4
RM4000 - RM4999	68	19.2
RM5000 - RM5999	30	8.5
RM6000 - RM6999	33	9.3
RM7000 - RM7999	41	11.5
RM8000 - RM8999	15	4.2
RM9000 - RM9999	22	6.2
RM10,000 and above	19	5.4
Marital Status		
Single	69	19.4
Married	281	79.2
Others	5	1.4
Computer Usage		
1-3 years	1	0.30
4-6 years	9	2.50
7-9 years	18	5.10
10 years and above	327	92.1
Internet Usage		
1-3 years	7	2.0
4-6 years	24	6.80
7-9 years	55	15.5
10 years and above	269	75.8

Malays (55.2%), followed by Chinese (25.0%) and Indians (19.8%) which somewhat reflects the ethnic group distribution in Malaysia. About 60.3% of the total respondents are highly educated with Bachelor degree and followed by Masters degree. The majority of the respondents (31.9%)

are earning within RM3000-RM3999 per month with majority (81.0%) are married respondents. Lastly, about 84.5% and 57.8% of the respondents claimed to have experience in computer usage and internet usage approximately 10 years and above, respectively.

Figure 2. Research model



DATA ANALYSIS

Smart PLS version 3.0, a variance based Structural Equation Modelling (SEM) was used to analyze the hypotheses generated. The two step analytical procedure suggested by Anderson and Gerbing (1988) was adopted to analyze data whereby the measurement model was evaluated first and then followed by the structural model. Also following the suggestion of Chin (1998), the bootstrapping method (500 resample) was done to determine the significant level of loadings, weights and path coefficients. The research model of this study is as below.

Measurement Model

Convergent validity is the degree to which the items that are indicators of a specific construct should converge or share a high proportion of variance in common (Hair, Black, Babin & Anderson, 2010). According to Hair et al. (2010), factor loadings and Average Variance Extracted (AVE) of more than 0.5 and Composite Reliability (CR) of 0.7 or

above is deemed to be acceptable. As can be seen from Table 2, all loadings and AVE are above 0.5 and the composite reliability values are more than 0.7. Therefore, we can conclude that convergent validity has been established.

Next, we assessed the Discriminant Validity which is the extent to which a construct is truly distinct from other constructs (Hair et al., 2010). This can be established by the low correlations between all the measure of the interest and the measure of other constructs. To address discriminant validity, the square root of the AVE is compared against the correlations of the other constructs, when the AVE extracted is greater than its correlations with all the other constructs then discriminant validity has been established (Fornell & Larcker, 1981) (refer Table 3).

Structural Model

The structural model represents the relationship between constructs or latent variables that were hypothesized in the research model. The goodness of the theoretical model is established by

Table 2. Result of the measurement model

Model Construct	Items	Loadings	AVE	CR	R ²
Attitude	ATT1	0.937	0.855	0.959	0.616
	ATT2	0.955			
	ATT3	0.951			
	ATT4	0.852			
Confirmation	CONF1	0.926	0.875	0.955	
	CONF2	0.946			
	CONF3	0.935			
Continuance	CINT1	0.937	0.894	0.971	0.722
	CINT2	0.948			
	CINT3	0.944			
	CINT4	0.954			
Perceived Ease of Use	PEOU1	0.867	0.806	0.943	
	PEOU2	0.899			
	PEOU3	0.919			
	PEOU4	0.904			
Perceived Usefulness	PU1	0.904	0.838	0.954	0.508
	PU2	0.909			
	PU3	0.923			
	PU4	0.925			
Satisfaction	SAT1	0.901	0.852	0.958	0.419
	SAT2	0.935			
	SAT3	0.937			
	SAT4	0.918			

Table 3. Discriminant validity of constructs

	Attitude	Confirmation	Continuance Intention	PE OU	PU	Satisfaction
Attitude	0.925					
Confirmation	0.613	0.936				
Continuance	0.829	0.561	0.946			
PE OU	0.673	0.723	0.636	0.898		
PU	0.747	0.597	0.741	0.701	0.915	
Satisfaction	0.543	0.620	0.507	0.576	0.518	0.923

Note: Diagonal represents the square root of Average Variance Extracted (AVE) while the other entries represent squared correlations

the variance explained (R^2) of the endogenous constructs and the significance of all path estimates (Chin, 2010). Together the R^2 and the path coefficients indicate how well the data support the hypothesized model (Chin, 1998). Figure 3 and Table 4, shows the results of the structural model from the PLS output. Confirmation was significantly related towards Satisfaction ($\beta = 0.483$, $p < 0.01$) and Perceived Usefulness ($\beta = 0.190$, $p < 0.01$) thus supporting H1 and H2 of this study. Perceived Usefulness was found in this study to be significantly related to Satisfaction (β

$= 0.230$, $p < 0.01$), Attitude ($\beta = 0.509$, $p < 0.01$) and Continuance Intention ($\beta = 0.266$, $p < 0.01$), thus supporting H3, H4 and H7. Perceived Ease of Use was found to be statistically significant to Attitude ($\beta = 0.232$, $p < 0.01$) and Perceived Usefulness ($\beta = 0.563$, $p < 0.01$) thus supporting H5 and H6. Satisfaction was found to be significantly related to Attitude ($\beta = 0.145$, $p < 0.01$) but insignificantly related to Continuance Intention ($\beta = 0.038$), thus rejects H8 and supports H9. Attitude was significantly related to Continuance Intention ($\beta = 0.609$, $p < 0.01$), thus supports H10.

Figure 3. Structural model

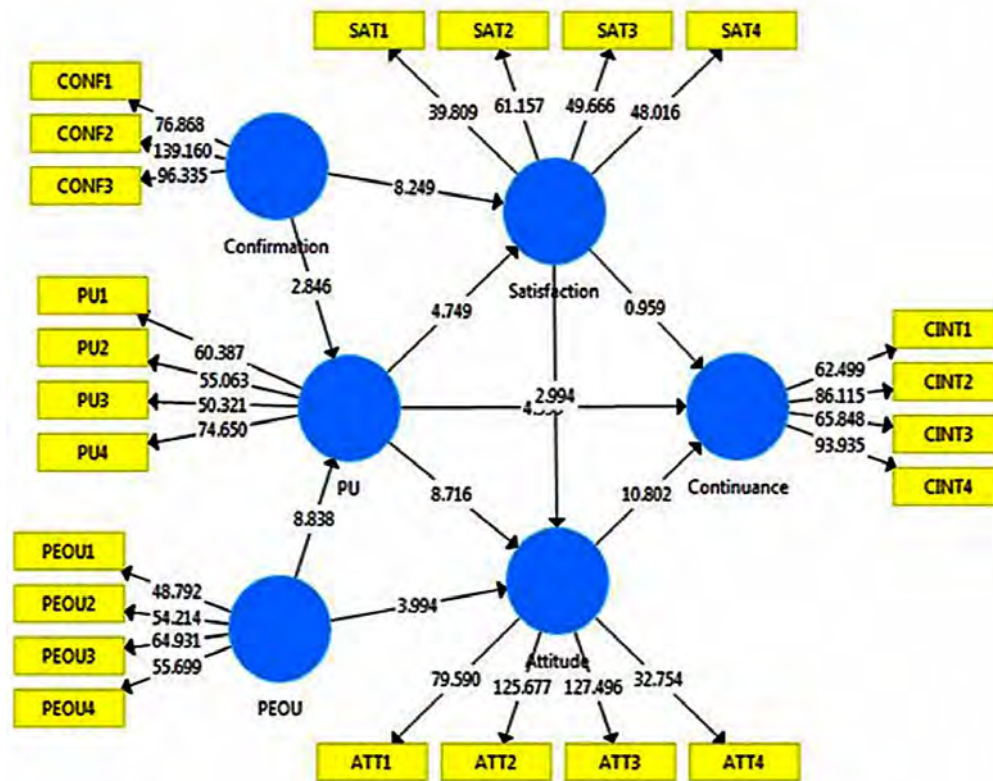


Table 4. Result of the structural model analysis

Hypotheses	Relationship	Std Beta	Std Error	t-value	Decision	f ²
H1	CONF → SAT	0.483	0.059	8.249**	Supported	0.258
H2	CONF → PU	0.190	0.067	2.846**	Supported	0.035
H3	PU → SAT	0.230	0.048	4.749**	Supported	0.058
H4	PU → ATT	0.509	0.058	8.716**	Supported	0.330
H5	PEOU → ATT	0.232	0.058	3.994**	Supported	0.063
H6	PEOU → PU	0.563	0.064	8.838**	Supported	0.308
H7	PU → CINT	0.266	0.053	4.999**	Supported	0.108
H8	SAT → CINT	0.038	0.039	0.959	Not supported	0.003
H9	SAT → ATT	0.145	0.049	2.994**	Supported	0.035
H10	ATT → CINT	0.609	0.056	10.802**	Supported	0.544

**p < 0.01

CONF Confirmation, SAT Satisfaction, PU Perceived Usefulness, ATT Attitude, PEOU Perceived Ease of Use, CINT Continuance Intention.

The study also has assessed the effect sizes (f^2). The effect size (f^2) is calculated based on the changes of the R^2 to determine the substantive influence of the predictor variable on the dependent variable (Thies & Albers, 2010). To measure the magnitude of the effect size Cohen's (1988) guideline was utilized which is 0.02, 0.15

and 0.35 signify small, medium and large effects respectively. Looking at the f^2 values in Table 4, it can be observed that nine out of ten relationships shows substantive impact whereby there were 5 relationships with small effect size, 3 with medium effect size and 1 with large effect size.

Table 5. Blindfolding results

	Q ²
Attitude	0.519
Continuance Intention	0.641
Perceived Usefulness	0.424
Satisfaction	0.345

Apart from that, “blindfolding” procedure was also performed to measure the predictive relevance (Q²) of the model fit. The Q² “represents a measure of how well observed values are reconstructed by the model and its parameter estimates” (Chin, 1998). Models with Q² greater than zero imply that the model has predictive relevance. Table 5 shows the result of the blindfolding results. Omission distance of 7 was utilized as Chin (1998) indicates that values between 5 and 10 are feasible (refer to Table 5).

DISCUSSION

The purpose of this study is to predict and to explain the continuance usage intention of e-filing system among taxpayers in Malaysia using two important central constructs; satisfaction and attitude. The findings of the research reveal that confirmation and perceived usefulness has a significant relationship towards satisfaction. This is in line with previous findings by Hwang et al., (2011), Lee and Kwon (2010), Kim (2010), Islam (2012) and Jiang (2011). However confirmation was found to be the best predictor of satisfaction compared to perceived usefulness. This coincide with previous findings which reveals perceived usefulness as secondary determinant to satisfaction (Lee & Kwon, 2010 ; Limayem & Cheung, 2008).

On the effect of confirmation and perceived ease of use on perceived usefulness, both have a significant relationship towards perceived usefulness. This is consistent with previous researches such as Hwang et al., (2011), Lee and Kwon (2010), Kim (2010), Lee (2010), Roca and Gagne

(2008). Perceived ease of use was found to be the main predictor of perceived usefulness. As such it indicates that for a voluntary system like e-filing system, the perceived ease of use is the first impression, once the taxpayers perceived it is easy to use then it will be considered useful.

Perceived usefulness, perceived ease of use and satisfaction shows a significant relationship towards attitude. Perceived usefulness was found to be an important predictor of attitude followed by perceived ease of use and satisfaction as the weakest predictor. Thus this finding supports the TAM model which stated that users' positive perception of usefulness towards any technology will lead to a positive attitude towards using the particular technology (Davis, Bagozzi & Warshaw, 1989). This is supported by previous researches such as Lin (2011), Lee (2010), Liang and Yeh (2011), Mosavi and Ghaedi (2012) and Sivadas and Prewitt (2000).

This study hypothesized three important variables; perceived usefulness, satisfaction and attitude to have direct relationship towards e-filing continuance usage intention. The findings reveals that perceived usefulness and attitude have significant relationship towards continuance usage intention (Li & Shi; 2012), Islam; 2012, Hsiao & Chiou; 2012, Lin; 2011, Liang & Yeh; 2011), while satisfaction to have an insignificant relationship. This contradicts with many previous findings which revealed a significant relationship such as Lin (2012), Chang and Zhu (2012), Lee (2010). This could imply that as e-filing system is a technology delivered by the government, taxpayers may be more concerned on its usefulness of the system in submitting their tax returns

safely, accurately and timely, with satisfaction has limited impact on continuance usage intention.

The model adopted in this study shows that the perceived usefulness and attitude can explain about 72.2% of the variance in continuance usage intention. These results showed that the model has relatively good predictive power on continuance usage intention. Further, the blindfolding result in Table 5 shows that the Q^2 are all above 0 which indicates the model has predictive relevance (Fornell & Cha, 1994).

The implication of the findings can be divided into two; theoretical and practical. Theoretically, this study adds to the growing body of literature that focus on the post adoption environment which is continuance usage intention. It also contributes to the evidence in support for the determinants of continuance usage intention of taxpayers especially in Malaysian context. Practically, since attitude was found to be the major predictor with high effect size towards e-filing continuance usage intention, the Inland Revenue Board Malaysia (IRBM) needs to pay more attention to implement awareness campaigns from time to time to encourage more taxpayers to use the e-filing system and to change their perception about the system.

Limitation and Suggestion for Future Research

Despite the useful findings of this study, there are several limitations that need to be acknowledged. Firstly, due to time and resource constraint the sample size of the study is only limited to 355 respondents. Secondly, the findings cannot be generalized extensively in Malaysia as the scope of the study is only limited to the taxpayers in Selangor and Kuala Lumpur only. As such, caution need to be taken when generalizing to the whole country. Lastly, this study only focus on predicting the effect of satisfaction, perceived usefulness, perceived ease of use, confirmation and attitude on continuance usage intention and does not incorporate the actual usage behaviour in the proposed model.

Therefore, this research can be done further in future by (1) expanding the study to other states in Malaysia, (2) extend the model by incorporating the actual usage behaviour or any other relevant variables such as personality or trust, (3) replicate the study to any other e-government services.

CONCLUSION

In this study, it was found that attitude is an important determinant of continuance usage intention. This finding confirms with several attitude models such as Theory of Reasoned Action (TRA), Technology Acceptance Model (TAM) and Theory of Planned Behaviour (TPB) that attitude has an important impact on behavioural intention. This is supported by Lin, Chen and Fang (2011) whose study rationalizes that attitude is a better predictor of intention compared to satisfaction due to the fact that attitude comprises of three components: cognition, affect and conation whereas satisfaction is only an affective component of attitude. As such, a change in attitude will have a high impact on continuance usage intention (Bhattacharjee & Premkumar, 2004).

REFERENCES

- Al-maghrabi, T., Dennis, C., & Halliday, S. V. (2011). Understanding The Factors That Derive Continuance Intention Of E-Shopping In Saudi Arabia: Age Group Differences In Behaviour. *Journal of Enterprise Information Management*, 24(1), 85–111.
- Annual Report Inland Revenue Board of Malaysia. (2014). Retrieved from the web April 20th 2016. http://www.hasil.gov.my/pdf/pdfam/annual_report_2014_bi.pdf
- Aziz, S. A., & Idris, K. M. (2012). The Determinants of Tax E-filing among Tax Preparers in Malaysia. *World Journal of Social Sciences*, 2(3), 182–188.

- Azmi, A. A. C., & Kamarulzaman, Y. (2010). Adoption of Tax E-Filing: A Conceptual Paper. *African Journal of Business Management*, 4(5), 599–603.
- Bernama, The Malaysian National News Agency. (2005). *IRB Invest in e-Filing System*. Retrieved from the web September 30, 2010: <http://blis.bernama.com.eserv.uum.edu.my//getArticle.do?id=7405&tid=48&cid=3>
- Bhatnagar, S. (2009). *Unlocking E-Government Potential: Concepts, Cases and Practical Insights* (1st ed.). SAGE Publications India Pvt Ltd.
- Bhattacharjee, A. (2001). Understanding Information Systems Continuance: An Expectation-Confirmation Model. *Management Information Systems Quarterly*, 25(3), 351–370. doi:10.2307/3250921
- Bhattacharjee, A., & Premkumar, G. (2004). Understanding Changes in Belief and Attitude Toward Information Technology Usage: A Theoretical Model and Longitudinal Test. *Management Information Systems Quarterly*, 28(2), 229–254.
- Brahmasrene, T., & Lee, J. W. (2012). Determinants of Intent to Continue Using Online Learning: A Tale of Two Universities. *Interdisciplinary Journal of Information, Knowledge, and Management*, 7, 1–20.
- Chang, Y. P., & Zhu, D. H. (2012). The Role of Perceived Social Capital And Flow Experience In Building Users Continuance Intention To Social Networking Sites In China. *Computers in Human Behavior*, 28(3), 995–1001. doi:10.1016/j.chb.2012.01.001
- Chin, W. W. (1998). Commentary: Issues and Opinion on Structural Equation Modeling. *Management Information Systems Quarterly*, 22(1), vii–xvi.
- Chin, W. W. (2010a). How to Write Up and Report PLS Analyses. In V. Esposito Vinzi et al. (Eds.), *Handbook of Partial Least Squares: Concepts, Methods and Applications* (pp. 655–689). Springer Handbooks of Computational Statistics, Springer-Verlag Berlin Heidelberg. doi:10.1007/978-3-540-32827-8_29
- Choi, H. S., Park, J. W., & Park, S. B. (2011). A Study on the Effect of Mobile Tourism Information Services on Tourist Satisfaction and Continual Reuse. *International Journal of Business & Information Technology*, 1(3), 189–195.
- Chou, S. W., & Chen, P. Y. (2009). The Influence of Individual Differences on Continuance Intentions of Enterprise Resource Planning (ERP). *International Journal of Human-Computer Studies*, 67(6), 484–496. doi:10.1016/j.ijhcs.2009.01.001
- Cooper, D. R., & Schindler, P. (2008). *Business Research Methods* (10th ed.). McGraw Hill.
- Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use and User Acceptance of Information Technology. *Management Information Systems Quarterly*, 13(3), 319–340. doi:10.2307/249008
- Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. *Management Science*, 15(8), 982–1003. doi:10.1287/mnsc.35.8.982
- Devaraj, S., & Kohli, R. (2003). Performance Impacts of Information Technology: Is Actual Usage the Missing Link? *Management Science*, 49(3), 273–289. doi:10.1287/mnsc.49.3.273.12736
- Fornell, C., & Cha, J. (1994). Partial Least Squares. In R. P. Bagozzi (Ed.), *Advanced Methods of Marketing Research* (pp. 52–78). Cambridge, UK: Blackwell.

- Fornell, C., & Larcker, D. F. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *JMR, Journal of Marketing Research*, 18(1), 39–50. doi:10.2307/3151312
- Hair, J. F. Jr, Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate Data Analysis: A Global Perspective* (7th ed.). Pearson Prentice Hall.
- Hasmah, A. (2009). *e-Filing Pays*. Retrieved from the web April 12, 2010. [http://www.intanbk.intan.my/i-portal/nict/nict/DAY1/SESSION3PARALLEL1\(SESSION3A\)/Dato_hasmah_e-filing_Pays.pdf](http://www.intanbk.intan.my/i-portal/nict/nict/DAY1/SESSION3PARALLEL1(SESSION3A)/Dato_hasmah_e-filing_Pays.pdf)
- Ho, C. H. (2010). Continuance Intention of E-Learning Platform: Toward An Integrated Model. *International Journal of Electronic Business Management*, 8(3), 206–215.
- Hsiao, C. C., & Chiou, J. S. (2012). The Effects of a Players Network Centrality On Resource Accessibility, Game Enjoyment, And Continuance Intention: A Study On Online Gaming Communities. *Electronic Commerce Research and Applications*, 11(1), 75–84. doi:10.1016/j.elerap.2011.10.001
- Hsieh, J. J. P. A., Rai, A., & Keil, M. (2008). Understanding Digital Inequality: Comparing Continued Use Behavioral Models of the Socio-Economically Advantaged and Disadvantaged. *Management Information Systems Quarterly*, 32(1), 97–126.
- Huang, S. S., & Hsu, C. H. C. (2009). Effects of Travel Motivation, Past Experience, Perceived Constraint, and Attitude on Revisit Intention. *Journal of Travel Research*, 48(1), 29–44. doi:10.1177/0047287508328793
- Hung, M. C., Chang, I. C., & Hwang, H. G. (2011). Exploring Academic Teachers Continuance Toward The Web-Based Learning System: The Role Of Causal Attributions. *Computers & Education*, 57(2), 1530–1543. doi:10.1016/j.compedu.2011.02.001
- Hussein, R., Mohamed, N., Ahlan, A. R., Mahmud, M., & Aditiawarman, U. (2010). An Integrated Model on Online Tax Adoption in Malaysia. *European, Mediterranean & Middle Eastern Conference on Information Systems (EMCIS)*, 1–16.
- Hwang, I. H., Yu, C. C., Tsai, S. J., & Lin, C. H. (2011). An Empirical Study on the Factors Affecting Continuous Usage Intention of Double Reinforcement Interactive E-Portfolio Learning System. *6th IEEE Joint International Conference on Information Technology and Artificial Intelligence*, 246–249. doi:10.1109/ITAIC.2011.6030196
- Islam, R. (2010). Critical Success Factors of The Nine Challenges in Malaysias Vision 2020. *Socio-Economic Planning Sciences*, 44(4), 199–211. doi:10.1016/j.seps.2010.07.002
- Jiang, X. (2011). Enhancing Users' Continuance Intention to E-Government Portals: An Empirical Study. *International Conference on Management and Service Science (MASS)*, 1–4. doi:10.1109/ICMSS.2011.5998967
- Kang, Y. S., & Lee, H. (2010). Understanding The Role Of An IT Artifact In Online Service Continuance: An Extended Perspective Of User Satisfaction. *Computers in Human Behavior*, 26(3), 353–364. doi:10.1016/j.chb.2009.11.006
- Kim, B. (2010). An Empirical Investigation of Mobile Data Service Continuance: Incorporating the Theory of Planned Behavior Into The Expectation–Confirmation Model. *Expert Systems with Applications*, 37(10), 7033–7039. doi:10.1016/j.eswa.2010.03.015
- Kim, H. B., Kim, T. T., & Shin, S. W. (2009b). Modeling Roles of Subjective Norms and eTrust in Customers Acceptance of Airline B2C eCommerce Websites. *Tourism Management*, 30(2), 266–277. doi:10.1016/j.tourman.2008.07.001

- Lean, O. K., Zailani, S., Ramayah, T., & Fernando, Y. (2009). Factors Influencing Intention to Use e-Government Services Among Citizens in Malaysia. *International Journal of Information Management*, 29(6), 458–475. doi:10.1016/j.ijinfomgt.2009.03.012
- Lee, M. C. (2010). Explaining and Predicting Users Continuance Intention Toward e-Learning: An Extension of the Expectation-Confirmation Model. *Computers & Education*, 54(2), 506–516. doi:10.1016/j.compedu.2009.09.002
- Lee, Y., & Kwon, O. (2010). (Article in Press). Intimacy, Familiarity and Continuance Intention: An Extended Expectation–Confirmation Model In Web-Based Services. *Electronic Commerce Research and Applications*.
- Li, G., & Shi, X. (2012). The Determinants of Consumers' Purchase Intention to Online Group-Buying. *Advanced Materials Research*, 459, 372–376.
- Liang, T. P., & Yeh, Y. H. (2011). Effect of Use Contexts On The Continuous Use Of Mobile Services: The Case Of Mobile Games. *Personal and Ubiquitous Computing*, 15(2), 187–196. doi:10.1007/s00779-010-0300-1
- Liao, C., Palvia, P., & Chen, J. L. (2009). Information Technology Adoption Behavior Life Cycle: Toward a Technology Continuance Theory (TCT). *International Journal of Information Management*, 29(4), 309–320. doi:10.1016/j.ijinfomgt.2009.03.004
- Limayem, M., & Cheung, C. M. K. (2008). Understanding Information System Continuance: The Case of Internet-Based Learning Technologies. *Information & Management*, 45(4), 227–232. doi:10.1016/j.im.2008.02.005
- Lin, K. M. (2011). E-Learning Continuance Intention: Moderating Effects of User E-Learning Experience. *Computers & Education*, 56(2), 515–526. doi:10.1016/j.compedu.2010.09.017
- Lin, K.M., Chen, N.S., & Fang, K. (2011). Understanding E-Learning Continuance Intention: A Negative Critical Incidents Perspectives. *Behaviour & Information Technology*, 1362 – 3001.
- Lin, W. S. (2012). Perceived Fit and Satisfaction On Web Learning Performance: IS Continuance Intention And Task- Technology Fit Perspectives. *International Journal of Human-Computer Studies*, 70(7), 498–507. doi:10.1016/j.ijhcs.2012.01.006
- Mittal, B., & Lassar, W. M. (1998). Why do customers switch? The Dynamics of Satisfaction Versus Loyalty. *Journal of Services Marketing*, 12(3), 177–194. doi:10.1108/08876049810219502
- Mosavi, S. A., & Ghaedi, M. (2012). An Examination Of The Effects Of Some Factors On Behavioral Intentions (A New Model). *Global Advanced Research Journal of Management and Business Studies*, 1(5), 163–172.
- Muhammad Rais, A. K., & Nazariah, M. K. (2003). *E-Government in Malaysia*. Pelanduk Publications (M). Sdn Bhd.
- Oliver, R. L. (1980). A Cognitive Model of the Antecedents and Consequences of Satisfaction Decisions. *JMR, Journal of Marketing Research*, 17(4), 460–469. doi:10.2307/3150499
- Ong, C. S., & Day, M. Y. (2010). An Integrated Evaluation Model of User Satisfaction with Social Media Services. *Proceeding of the International Conference on Information Reuse and Integration (IRI)*, 195-200. doi:10.1109/IRI.2010.5558940
- Premkumar, G., & Bhattacharjee, A. (2008). Explaining Information Technology Usage: A-Test of Competing Models. *International Journal of Management Sciences*, 36(1), 64–75.
- Roca, J. C., & Gagne, M. (2008). Understanding E-Learning Continuance Intention In The Workplace: A Self-Determination Theory Perspective. *Computers in Human Behavior*, 24(4), 1588–1604. doi:10.1016/j.chb.2007.06.001

- Shiau, W. L., Huang, L. C., & Shih, C. H. (2011). Understanding Continuance Intention of Blog Users: A Perspective of Flow and Expectation Confirmation Theory. *Journal of Convergence Information Technology*, 6(4), 306–317. doi:10.4156/jcit.vol6.issue4.33
- Shih, C. H., Shiau, W. L., & Huang, L. C. (2010). Understanding The Bloggers' Continuance Usage: Integrating Flow Into The Expectation-Confirmation Theory Information System Model. *Proceeding of the Pacific Asia Conference on Information Systems (PACIS)*, 1680–1686.
- Sivadas, E., & Prewitt, J. L. B. (2000). An Examination of the Relationship Between Service Quality, Customer Satisfaction and Store Loyalty. *International Journal of Retail & Distribution Management*, 28(2), 73–82. doi:10.1108/09590550010315223
- Spreng, R. A., MacKenzie, S. B., & Olshavsky, R. W. (1996). A Reexamination of the Determinants of Consumer Satisfaction. *Journal of Marketing*, 60(3), 15–32. doi:10.2307/1251839
- Suki, N. M., & Ramayah, T. (2010). User Acceptance of the E-Government Services in Malaysia: Structural Equation Modelling Approach. *Interdisciplinary Journal of Information, Knowledge, and Management*, 5, 395–413.
- Sun, P. C., Tsai, R. J., Finger, G., Chen, Y. Y., & Yeh, D. (2008). What Drives A Successful E-Learning? An Empirical Investigation of The Critical Factors Influencing Learner Satisfaction. *Computers & Education*, 50(4), 1183–1202. doi:10.1016/j.compedu.2006.11.007
- Taylor, S., & Todd, P. A. (1995). Understanding Information Technology Usage: A Test of Competing Models. *Information Systems Research*, 6(2), 144–176. doi:10.1287/isre.6.2.144
- Thies, S. B., & Albers, S. (2010). Application of PLS in Marketing: Content Strategies on the Internet. In V. Esposito Vinzi et al. (Eds.), *Handbook of Partial Least Squares: Concepts, Methods and Applications* (pp. 589–604). Springer Handbooks of Computational Statistics, Springer-Verlag Berlin Heidelberg. doi:10.1007/978-3-540-32827-8_26
- Wangpipatwong, S., Chutimaskul, W., & Papatrorn, B. (2008). Understanding Citizen's Continuance Intention to Use e-Government Website: A Composite View of Technology Acceptance Model and Computer Self-Efficacy. *The Electronic Journal of E-Government*, 6(1), 55–64.
- Zakaria, Z., Hussin, Z., Zakaria, Z., Noordin, N. B., Sawal, M. Z. H. B. M., Saad, S. F. B. M., & Kamil, S. B. O. (2009). E-Filing System Practiced by Inland Revenue Board (IRB): Perception Towards Malaysian Taxpayers. *Cross Cultural Communication*, 5(4), 10–20.
- Zhao, L., & Lu, Y. (2012). Enhancing Perceived Interactivity Through Network Externalities: An Empirical Study On Micro-Blogging Service Satisfaction And Continuance Intention. *Decision Support Systems*, 53(4), 825–834. doi:10.1016/j.dss.2012.05.019

ADDITIONAL READING

- Carter, L., Weerakkody, V., Phillips, B., & Dwivedi, Y. K. (2016). Citizens Adoption of E-Government Services: Exploring Citizens Perception of Online Services in the United States and United Kingdom. *Journal of Information System Management*, 33(2), 124–140. doi:10.1080/10580530.2016.1155948

Rana, N., Lal, B., & Slade, B. (2016) Adoption of Two Indian E-Government Systems: Validation of Extended Theory of Planned Behaviour (TPB). *Proceedings of the Americas Conference on Information System (AMCIS)* San Diego, 11-14 August, 7

Rana, N. P., Dwivedi, Y. K., Williams, M. D., & Weerakkody, V. (2015). Investigating Success of an E-Government Initiative: Validation of an Integrated IS Success Model. *Information Systems Frontiers*, 17(1), 127–142. doi:10.1007/s10796-014-9504-7

Siddiquee, N. (2016). E-Government and Transformation of Service Delivery in Developing Countries: The Bangladesh Experience and Lessons. *Transforming Government: People, Process and Policy*, 10(3), 368–390.

Venkatesh, V., Thong, J. Y. L., Chan, F. K. Y., & Hu, P. J. H. (2016). Managing Citizens Uncertainty in E-Government services: The Mediating and Moderating Roles of Transparency and Trust. *Journal of Information Systems Research*, 27(1), 87–111. doi:10.1287/isre.2015.0612

Weerakkody, V., Irani, Z., Lee, H., Osman, I., & Hindi, N. (2015). E-Government Implementation: A Birds Eye View of Issues Relating to Costs, Opportunities, Benefits and Risks. *Information Systems Frontiers*, 17(4), 889–915. doi:10.1007/s10796-013-9472-3

KEY TERMS AND DEFINITIONS

Attitude: The characteristics of a user.

Confirmation: The evaluation of the performance of the system against the expectation of the users.

Continuance Intention: One's intention to continually use a system or reuse a system.

E-Filing: Filing income tax returns via electronically.

Perceived Usefulness: Perception that a particular system is useful in completing a job.

Perceived Ease of Use: Perception that a particular system is easy to use.

Satisfaction: One's pleasure response when a system meets their desire and wants.

Taxpayers: Individuals who are liable to pay the tax based on their income for a particular year.

Forecasting the Demand of Agricultural Crops/Commodity Using Business Intelligence Framework

B

Satyadhyan Chickerur

KLE Technological University, India

Supreeth Sharma

Akamai Technologies, India

Prashant M. Narayankar

KLE Technological University, India

INTRODUCTION

Business Intelligence (BI) has been around for more than a decade and is being used in various application domains. BI is about creating value for organizations based on data or, more precisely, facts. While it seems like another buzzword to describe what successful entrepreneurs have been doing for years, if not centuries, that is, using business common sense? From a modern business-value perspective, corporations use BI (Sherman, 2014; Prasad & Acharya, 2011) to enhance decision-making capabilities for managerial processes (e.g., planning, budgeting, controlling, assessing, measuring, forecasting and monitoring) and to ensure critical information is exploited in a timely manner. And computer systems are the tools that help us do that better, faster, and with more reliability. BI is a broad category of application programs and technologies for gathering, storing, analyzing and providing access to data to help enterprise users make better business decisions. BI (Sherman, 2014; Prasad & Acharya, 2011) applications support the activities of decision support, query and reporting, online analytical processing, statistical analysis, forecasting, and data mining. BI includes a set of concepts and methods to improve business decision making by

using fact-based support systems. BI has become synonymous with historical query and reporting tools for many years, the definition of BI in recent years has evolved to include technology that addresses data integration, predictive analytics and information deployment (Shmueli, Patel, & Bruce, 2010; Sherman, 2014; Prasad & Acharya, 2011). Predictive analytics often help companies address their most difficult business issues and achieve a competitive advantage in their respective markets. Forecasting is a discipline of analytics that provides the foundation for planning processes across organizations. Good forecasting provides a more accurate view of the future, helping an organization save time and reduce costs, while better serving its customers and managing its resources. When used with data integration and information deployment capabilities, forecasting becomes even more powerful by pulling from multiple data sources and disseminating forecast results consistently. Information technology for agricultural (Shmueli et al., 2010) use can be broadly classified for:

1. Tools which might be used to improve productivity
2. Tools which empower farmers to take informed and quality decisions

At present the information available for farmers is through conventional sources, which are slow and unreliable. As a result of the information not reaching in time to the farmers, there is lot of wastage in terms of crops getting spoiled, some crop getting excess production and some crops not getting produced at all. These factors ultimately result in controlling the prices of the commodities dependent upon a particular crop. Thus the Farmers would be in a position to take the benefit of the technology used in the proposed approach. BI allows the decision maker to understand their business environment in order to make informed decision. Decision-making requires evaluating performance (what happened), testing hypotheses (why and how things happened) and predicting future events (what may happen). Stated simply management needs to know if their strategies are sound if they are being carried out. Most formally a business intelligence system allows users to answer above-mentioned decisions. Forecasting involves predicting and analysis of the given data, which can be achieved by BI.

- **Predictive Analytics:** Predictive analytics represent any solution that supports the identification of meaningful patterns and correlations among variables in complex, structured, unstructured, historical, and potential future data sets for the purposes of predicting events and assessing the attractiveness of various courses of action.
- **Planning Decisions:** This includes analytical reporting based on a data warehouse (DW) (Shmueli et al., 2010) or data mart and operational reporting based on an operational database management system. Reporting tools often include pixel-perfect positioning of data and graphics, a scripting language equal in power to a full program programming language, and the ability to handle complex headers, footers, nested subtotals, and multiple report bands on a single page.

BACKGROUND

Over the next thirty-five years, global population is expected to reach nine billion people. This will increase collective expectation of economic growth, nutritional demand and the resources used for food production. In order to meet expected worldwide nutritional demand and global food needs the agricultural industries and farmers should maintain good agricultural practices. Many previous studies have put together projections of global food demand. These studies tend to make large commodity groupings such as cereals, oil-seeds, meats, etc. (Kruse, 2010)

Business intelligence (Sherman, 2014; Prasad & Acharya, 2011; Legorreta & Valentine, 2014) means different things to different people. In its simplest form, business intelligence is getting the right information to the right people at the right time so they can make decisions that ultimately improve performance. A more technical view of business intelligence usually centers on the process of, or applications and technologies for, gathering, forecasting, and storing, analyzing and providing access to data to help make better business decisions. Regardless of how you define it, business intelligence is proliferating and reaching more and more constituents inside and outside of your organization. Information demand, data volumes and audience populations are growing and will continue to grow exponentially. The increase in available information also affects forecasting because there are more inputs to consider in both volume and type.

BI evolves different types of platforms and tools with respect to different categories. Precisely those are defined as (Sherman, 2014; Prasad & Acharya, 2011; Esri's, 2014):

1. Enterprise BI platform,
2. Visual data and discovery tool,
3. Pure OLAP (online analytical processing) tool,
4. Database or packaged BI, and
5. Dashboard tool.

BI functionality becomes commoditized; buyers need to look for differentiation elsewhere in these platforms. It's not just about tool functionality; tight integration with enterprise ERP, ETL, Forecasting, portal (Hamilton, 2012), and desktop office applications can often break the tie when selecting a strategic enterprise BI platform. Elaborating these platforms is vital task.

1. **Enterprise BI Platform:** BI platform has ability to satisfy business users, developers and enterprise needs. Its sets the right balances between ease of use and sophistication in our opinion making process.
2. **Visual Data and Discovery Tool:** Data visualization is a rich, highly interactive visual tool that allow user to manipulate and interact with information directly. Data discovery is the discovery relationship between elements regardless of where data is stored.
3. **Pure OLAP (Online Analytical Processing) Tool:** Otherwise known as "slicing and dicing" analysis, online analytical processing (OLAP) tools allow a user to almost instantaneously regroup, reaggregate, and resort facts typically additive numbers like transaction amounts and account balances by any dimension.
4. **Database or Packaged BI:** These are products that are supplied as an option when you buy the database or packaged application software. Some of the functionality may be built in to the database and/or the application and not be available separately.
5. **Dashboards Tool:** Dashboards are interactive visualizations that mash up different historical, current, and/or predictive information into one efficient graphical user interface (GUI). Typically, dashboards display key performance indicators (KPIs) using visual cues.

Although business intelligence was synonymous with query and reporting tools for many years, many organizations have discovered that the effective use of information requires more

than reports that show historical data. In addition to the information delivery component of business intelligence, data integration and analytics are garnering just as much attention in overall business intelligence strategies. A platform for enterprise business intelligence routinely consists of the following categories of technologies like data integration capabilities for data connectivity, data quality, ETL (extract, transform and load) data migration, data synchronization and data federation (Shmueli et al., 2010; Ranjan, 2005).

Correct forecasting involves the following two steps:

1. Predictive analytics that includes predictive and descriptive modeling, forecasting, optimization, simulation, experimental design and more. Forecasting uses many specialized types of models, such as exponential smoothing, ARIMA and unobserved components.
2. Information delivery capabilities to surface information from consistent, companywide data, and provide a single view of the data on which all planning decisions can be based.

Each of these components should be integrated not only into the business intelligence platform but also with the existing investments in hardware and software. A comprehensive business intelligence platform will be able to access all the required data no matter where it resides and no matter what operating system is being used. The BI framework will move the data if necessary, store it properly, analyze it thoroughly and disseminate it to users through familiar interfaces.

DEMAND OF FORECASTING FOR AGRICULTURAL CROPS/COMMODITY

In today's world with increasing population and year on year decrease in the agricultural land, there is a burning need for efficient utilization of the available land so as to become self-sufficient. Also in India, food inflation stands in double digits because of increasing demand and shortage

in supply. To control the food inflationist's very essential to meet the demands within the territory. If it's possible to forecast the demands of major agricultural crops for coming years with an acceptable range of accuracy, then proper planning could be made by the government bodies so as to meet the food demands. Even the food articles prices could be regulated without harming the economy. Demand for selected crops in each coming years can be achieved by predictive analytics. This may result in projecting approximate demand for each year for a particular crop/commodity.

Forecasting is the process of making projections of demand for products by examining past and present performance levels, combined with assessment of available products and markets. Forecasts of agricultural production and prices are intended to be useful for farmers, governments, and agribusiness industries. Because of the special position of food production in a nation's security, governments have become both principal suppliers and main users of agricultural forecasts. They need internal forecasts to execute policies that provide technical and market support for the agricultural sector. Government publications routinely provide

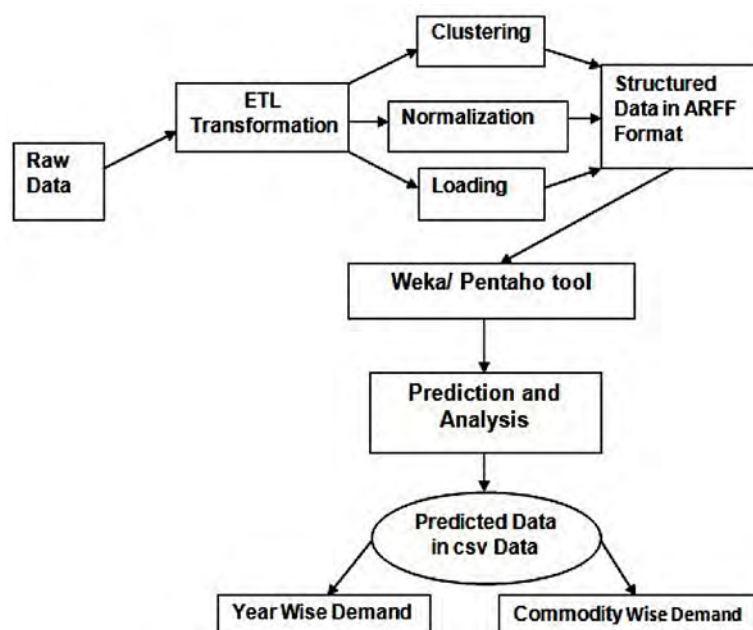
private decision makers with commodity price and output forecasts at regional and national levels.

Commodities like Oils, Pulses, Rice, Sugar and wheat have been chosen for the research. Demand for selected commodities may vary over the years, so projecting the demands of each crops/commodity over the year is challenging. Forecasting demands can be considered for other crops/commodity of agriculture. The objective of this research was to develop a detailed forecasting of demand for agricultural crops/commodity. Forecasting of demands is categorized based on year wise and commodity wise.

PROPOSED APPROACH

Implementation process involves integrating different system modules present in the architecture. Architecture describes functionalities of different modules and their activities with respect to system. It also defines how system works and what will be the flow of activity. In forecasting the demand of agriculture crops/commodity, we describe different set of modules and their activities in Figure 1.

Figure 1. System architecture



In Figure 1 an unstructured raw data acts as an input to the business intelligent process that evolves ETL transformation (Shmueli, et al., 2010). Clustering, Normalization and loading are sub activities for ETL transformation, which will formally structures the unstructured data, which we fed into our system. After transformation the structured data will be stored in ARFF (Bouckaert, et al., 2013) (Attribute relation file format) format, these files are then processed in weka tool. Prediction and analysis technique will be applied to structured data. The values derived from prediction are stored in CSV (Comma separated values) file format and they are now analyzed based on year wise crop and commodity wise crop. The raw data required for making the analysis in csv format was obtained in the form of the demand of five crops for last fifteen years. The structure of raw data is shown in Figure 2, Even though the data obtained was structured, it was so mixed up and jumbled that it was not fit for making effective analysis. Also data contained many redundant entries, which needed to be removed before storing them in the database. So it was a necessity to perform the ETL (Extract, Transform and Load) operations on this raw data to make it suitable for analysis.

Methodology

ETL (Extract, Transform and Load) (Pentaho Community, 2014) is the primary step for applying the business intelligence. It mainly included all those preprocessing activities, which are required for transforming the data into a proper format, which are required for analysis. Pentaho's data integration tool was used for performing the ETL. It included three main activities namely:

1. Clustering,
2. Normalization, and
3. Loading.

Clustering

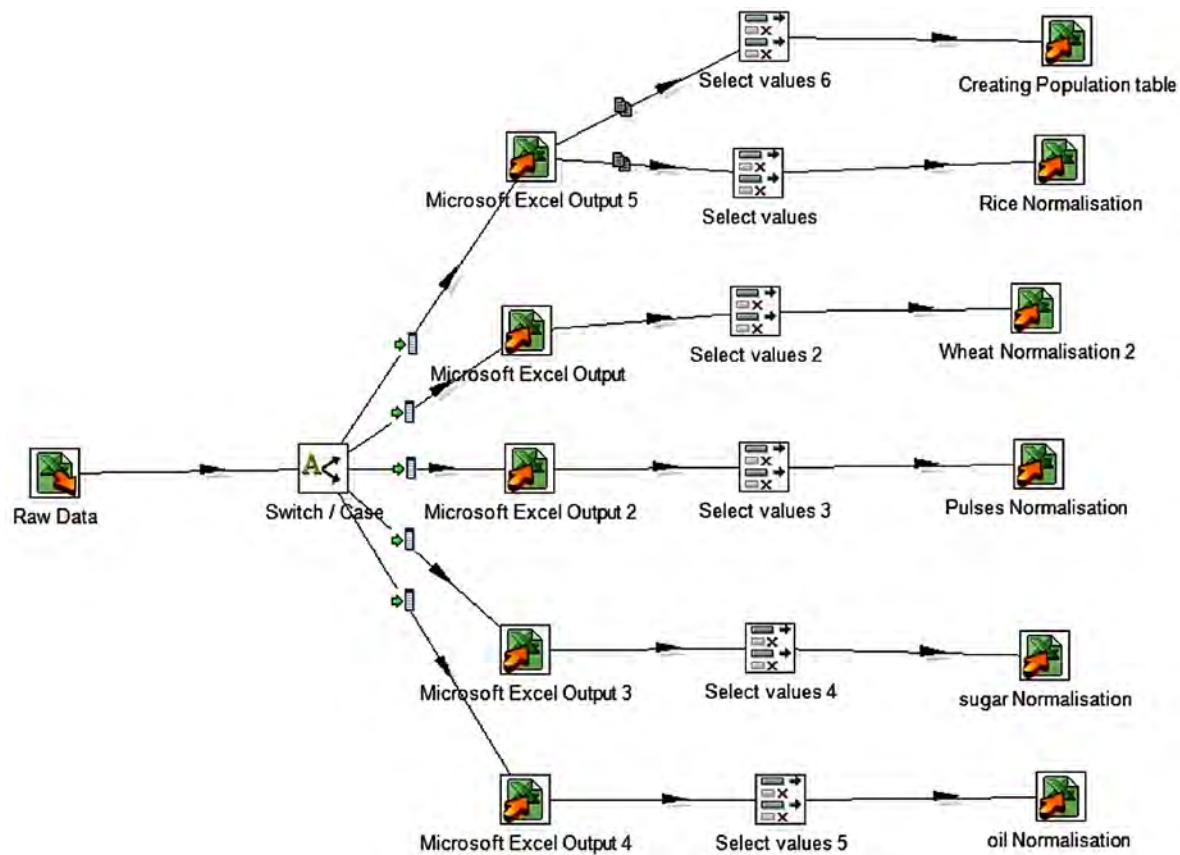
Cluster analysis or clustering is the task of assigning a set of objects into groups (called clusters) so

Figure 2. Structure of the RAW data as input

	A	B	C
1	Crop Nam	Year	Demand
2	Wheat	1981	45.6
3	Edible Oil	1983	4.1
4	Pulses	1983	10.1
5	Sugar	1983	9.7
6	Wheat	1983	47.9
7	Rice	1983	68.1
8	Edible Oil	1984	4.2
9	Pulses	1984	10.1
10	Sugar	1984	9.8
11	Wheat	1984	48.1
12	Rice	1984	68.7
13	Edible Oil	1985	4.5
14	Pulses	1985	10.2
15	Sugar	1985	10.2
16	Wheat	1985	48.3
17	Rice	1985	69.5
18	Edible Oil	1986	4.6
19	Pulses	1986	10.3
20	Sugar	1986	9.9
21	Wheat	1986	48.7
22	Rice	1986	70.9
23	Edible Oil	1987	4.7
24	Pulses	1987	10.3
25	Sugar	1987	10.1

that the objects in the same cluster are more similar (in some sense or another) to each other than to those in other clusters. Clustering (Shmueli, et al. 2010) is a main task of explorative data mining, and a common technique for statistical data analysis used in many fields, including machine learning, pattern recognition, image analysis, information retrieval, and bioinformatics. As it can be seen from Figure 2, the raw data is completely mixed up. So it has to be grouped into different clusters with each group containing the data relevant to particular crop. Pentaho's data integration tool is used for clustering. Input data from the file is fed to a switch case, which filtered the data based on specified condition. We made the group based on the crop name. So clusters of crops are formed. The output streams of the switch case are fed as an input to different files, which now contain the information relevant to each individual crop. This process and structure of the data obtained after this stage is shown in Figure 3.

Figure 3. Clustering based on the crop



Normalization

Normalization (Elmasri & Navathe, 2011) is the process of organizing the fields and tables of a relational database to minimize redundancy and dependency. Normalization usually involves dividing large tables into smaller (and less redundant) tables and defining relationships between them. The objective is to isolate data so that additions, deletions, and modifications of a field can be made in just one table and then propagated through the rest of the database via the defined relationships. The data obtained after clustering stage is redundant. As it can be seen from Figure 4 the crop name is redundantly repeated. So to normalize it we removed the crop name field from the file and the data is stored in a new file whose name is same as that of the crop name. Thus the data obtained is normalized and kept in 3NF form, where non-primary attribute demand is fully functionally

dependent on primary key year. The structure of the normalized data is shown in Figure 5.

Functional Dependencies

A functional dependency (Elmasri & Navathe, 2011) is a constraint between two sets of attributes from the database. Functional dependencies are the formal tool for analysis of relational schema that enables us to detect the redundancies present in the clustered crop wise data Figure 5. It also used to define the normal forms for relational schemas.

Definition

A functional dependency (Elmasri & Navathe, 2011), denoted by $\text{Year}[A] \rightarrow \text{Demand}[B]$, between two sets of attributes A and B that are subsets of R specifies a constraint on the possible tuples that can form a relation state r (Relational state) of

Figure 4. Clustered crop wise data

	A	B	C
1	Crop Name	Year	Demand
2	Rice	1,983	68.10
3	Rice	1,984	68.70
4	Rice	1,985	69.50
5	Rice	1,986	70.90
6	Rice	1,987	72.00
7	Rice	1,988	73.10
8	Rice	1,989	73.40
9	Rice	1,990	72.60
10	Rice	1,991	71.00
11	Rice	1,992	69.50
12	Rice	1,993	68.60
13	Rice	1,994	67.30
14	Rice	1,995	67.70
15	Rice	1,996	67.40
16	Rice	1,997	67.50
17	Rice	1,998	67.00
18	Rice	1,999	67.20
19	Rice	2,000	66.80
20	Rice	2,001	67.60
21	Rice	2,002	68.90
22	Rice	2,003	70.10
23	Rice	2,004	71.50
24	Rice	2,005	72.30
25	Rice	2,006	72.90

R. The constraint is that, for any two tuples t_1 and t_2 in r that have $t_1[A] = t_2[A]$, they must also have $t_1[B] = t_2[B]$ where $R = \{A_1, A_2, \dots, A_n\}$. This means that the values of the B component of a tuple in r depend on, or are determined by, the values of the A component; alternatively, the values of the A component of a tuple uniquely (or functionally) determine the values of the B component. We also say that there is a functional dependency from A to B, or that B is functionally dependent on A. Thus, A functionally determines B in a relation schema R if, and only if, whenever two tuples of $r(R)$ agree on their A-value, they must necessarily agree on their B value.

3rd Normal Form

B

The normal form (Elmasri & Navathe, 2011) of a relation refers to the highest normal form condition that it meets, and hence indicates the degree to which it has been normalized. Third normal form (3NF) is based on the concept of transitive dependency. A functional dependency $A \rightarrow B$ in a relation schema R is a transitive dependency if there exists a set of attributes Z in R that is neither a candidate key nor a subset of any key of R and both $A \rightarrow Z$ and $Z \rightarrow B$ hold. The dependency Demand [B] $\rightarrow$ Year [A] is transitive through Crop Name in Normalized crop data Figure 4.

Figure 5. Normalized crop data

	A	B
1	Year	Demand
2	1,983.00	68.10
3	1,984.00	68.70
4	1,985.00	69.50
5	1,986.00	70.90
6	1,987.00	72.00
7	1,988.00	73.10
8	1,989.00	73.40
9	1,990.00	72.60
10	1,991.00	71.00
11	1,992.00	69.50
12	1,993.00	68.60
13	1,994.00	67.30
14	1,995.00	67.70
15	1,996.00	67.40
16	1,997.00	67.50
17	1,998.00	67.00
18	1,999.00	67.20
19	2,000.00	66.80
20	2,001.00	67.60
21	2,002.00	68.90
22	2,003.00	70.10
23	2,004.00	71.50
24	2,005.00	72.30
25	2,006.00	72.90

Loading

The load phase loads the data into the end target, usually the data warehouse (DW) (Shmueli et al., 2010). Depending on the requirements of the organization, this process varies widely. Some data warehouses may overwrite existing information with cumulative information; frequently updating extract data is done on daily, weekly or monthly basis. Other DW (or even other parts of the same DW) may add new data in a historicized form, for example, hourly. To understand this, consider a DW that is required to maintain sales records of the last year. Then, the DW will overwrite any data that is older than a year with newer data. However, the entry of data for any one-year window will be made in a historicized manner. The timing and scope to replace or append are strategic design choices dependent on the time available and the business needs. More complex systems can maintain a history and audit trail of all changes to the data loaded in the DW. In this stage normalized data is loaded into database or a data warehouse is created. Postgres was used which is an open source DBMS. Separate table is created for each crop and the normalized data obtained from the previous stage are loaded into respective tables.

Forecasting the Demand

Enhancing BI with forecasting (SAS: Analytics, 2015) uses time series analysis; it is the process of using statistical techniques to model and explain a time-dependent series of data points. Time series forecasting is the process of using a model to generate predictions (forecasts) for future events based on known past events. Time series data has a natural temporal ordering - this differs from typical data mining/machine learning applications where each data point is an independent example of the concept to be learned, and the ordering of data points within a data set does not matter. Examples of time series applications include: capacity planning, inventory replenish-

ment, sales forecasting and future staffing levels. Weka (3.7.8) (Bouckaert, et al., 2013) now has a dedicated time series analysis environment that allows forecasting models to be developed, evaluated and visualized.

This environment takes the form of a plugin (Pentaho Community, 2014) tab in Weka's graphical Explorer user interface and can be installed via the package manager. Weka's time series framework takes a machine learning/data mining approach to modeling time series by transforming the data into a form those standard propositional learning algorithms can process. It does this by removing the temporal ordering of individual input examples by encoding the time dependency via additional input fields. These fields are sometimes referred to as "lagged" variables. Various other fields are also computed automatically to allow the algorithms to model trends and seasonality. After the data has been transformed, any of Weka's regression algorithms can be applied to learn a model. An obvious choice is to apply multiple linear regressions, but any method capable of predicting a continuous target can be applied - including powerful non-linear methods such as support vector machines for regression and model trees (decision trees with linear regression functions at the leaves). This approach to time series analysis and forecasting is often more powerful and more flexible than classical statistical techniques such as ARMA and ARIMA (Bouckaert, et al., 2013). The above-mentioned "core" time series-modeling environment is available as open-source free software in the CE version of Weka. The same functionality has also been wrapped in a Spoon Perspective plugin that allows users of Pentaho Data Integration (PDI) (Pentaho, 2014) to work with time series analysis within the Spoon PDI GUI. There is also a plugin step for PDI that allows models that have been exported from the time series-modeling environment to be loaded and used to make future forecasts as part of an ETL transformation.

The perspective and step plugin for PDI are part of the enterprise edition. In the proposed work

the normalized crop files as input are loaded into the weka-scoring tool. Parameters are adjusted so as to make the prediction for eight years. This is as shown in Figure 5. The predicted values from the weka tool are stored in respective csv files, which are later loaded into the database tables.

The predicted values are generally stored in csv file but inputs to the weka tool is given in the format of An ARFF (Attribute-Relation File Format) file is an ASCII text file that describes a list of instances sharing a set of attributes. ARFF files have two distinct sections. The first section is the Header information, which is followed the Data information. The Header of the ARFF file contains the name of the relation, a list of the attributes (the columns in the data), and their types.

EXPERIMENTATION AND RESULTS

This section discusses experimentation carried out on raw data related to various crops. Results are derived for two criteria, one is with respect to crop wise and another one is on year wise demand. These forecasting strategies will; help to analyze demands for the particular years and crops. Crop demand is presented in the form pie chart to easily understand year wise demand for particular crop. The crop demands for the year 2013-2020 have been predicted. Some sample crops like rice; oil, pulse, sugar and wheat have been taken for which their demand is estimated. By analyzing the line graphs market commodity fluctuation can be easily understood.

Deriving a New Perception of the Available Information

The forecasted values provide the demand for a particular crop in coming years. In other words it's a crop wise view. The crop wise view shows the demand for different crops in each year from already available information. For this, the predicted values from all the crop files are combined and separate files are generated which contain the demand for different crops in each year (Director of Economics, 2015). This is done using Switch/Case tool as shown in Figure 7.

The predicted values from all the five crops are fed as inputs to the Switch/Case tool. Values of the input stream are filtered based on year field as parameter. These filtered output stream values are fed to respective files so that each file contains five rows, which are the predicted demands of five different crops/Commodity. Values from these files are very helpful in generating the reports for year wise projection in later stage. Two kinds of reports are provided with two different perspectives of the forecasted information. These are namely:

- Year wise demand.
- Commodity wise demand.

Year Wise Demand Forecasting

This type of a report presents the demand for different crops in each coming years i.e. from 2013-2020. This is presented in the form of a pie

Figure 6. Normalized data being fed into scoring tool

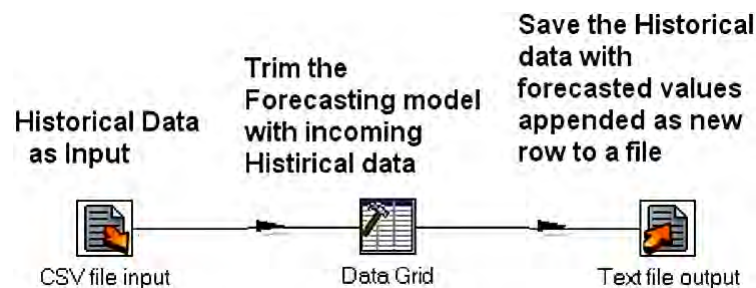


Figure 7. Combined Crop Wise View

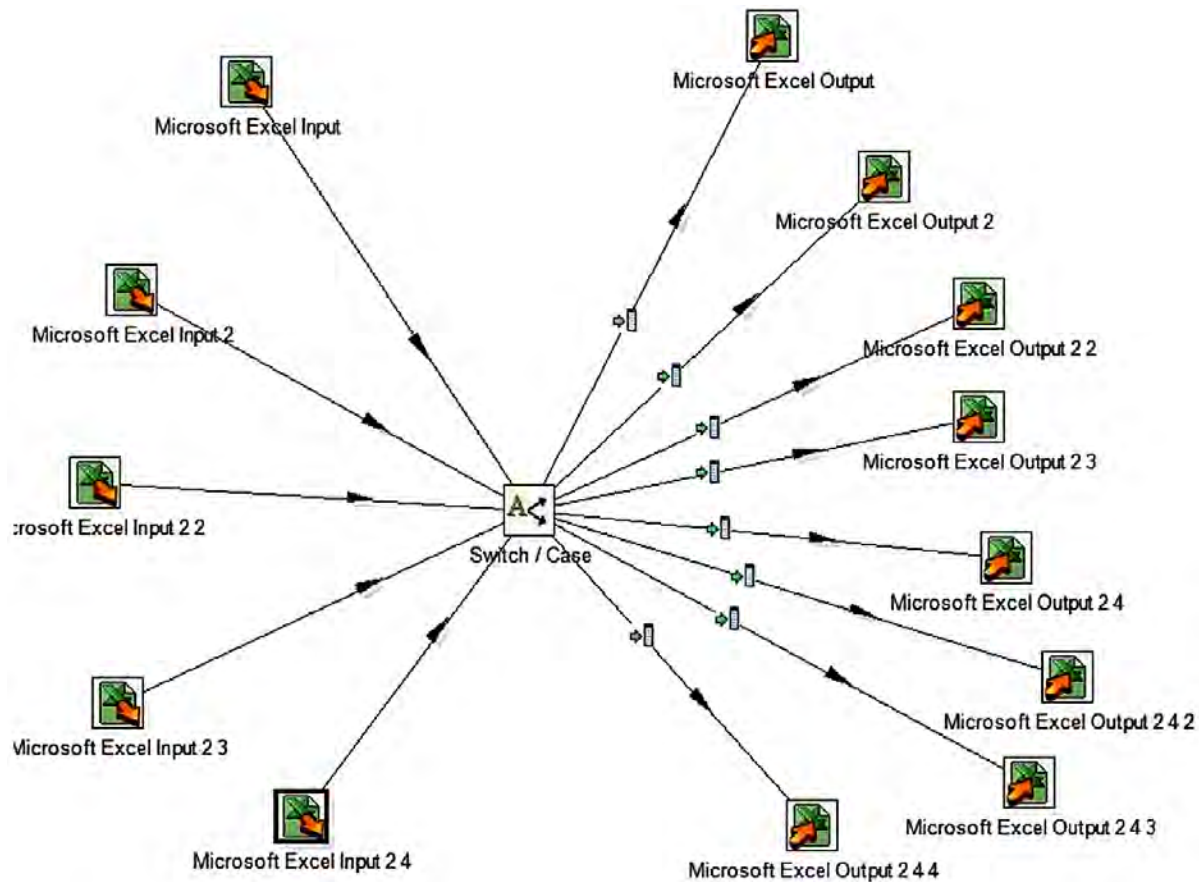


chart in Figure 8. Thus eight charts are produced, one for each year. This format of view is very helpful for reading and analyzing the demands for crops in any particular year. It provides the required information in an easily readable and understandable format.

Crop Wise Demand Forecasting

The demand for particular crop in different years is presented in the form of a line graph. Thus five graphs are produced, one for each crop. These graphs are as shown in Figure 9. This format of view is very helpful for reading and analyzing the demands for a particular crop in coming years. It provides the required information in an easily readable and understandable format. One can also see the fluctuation in demand over the years

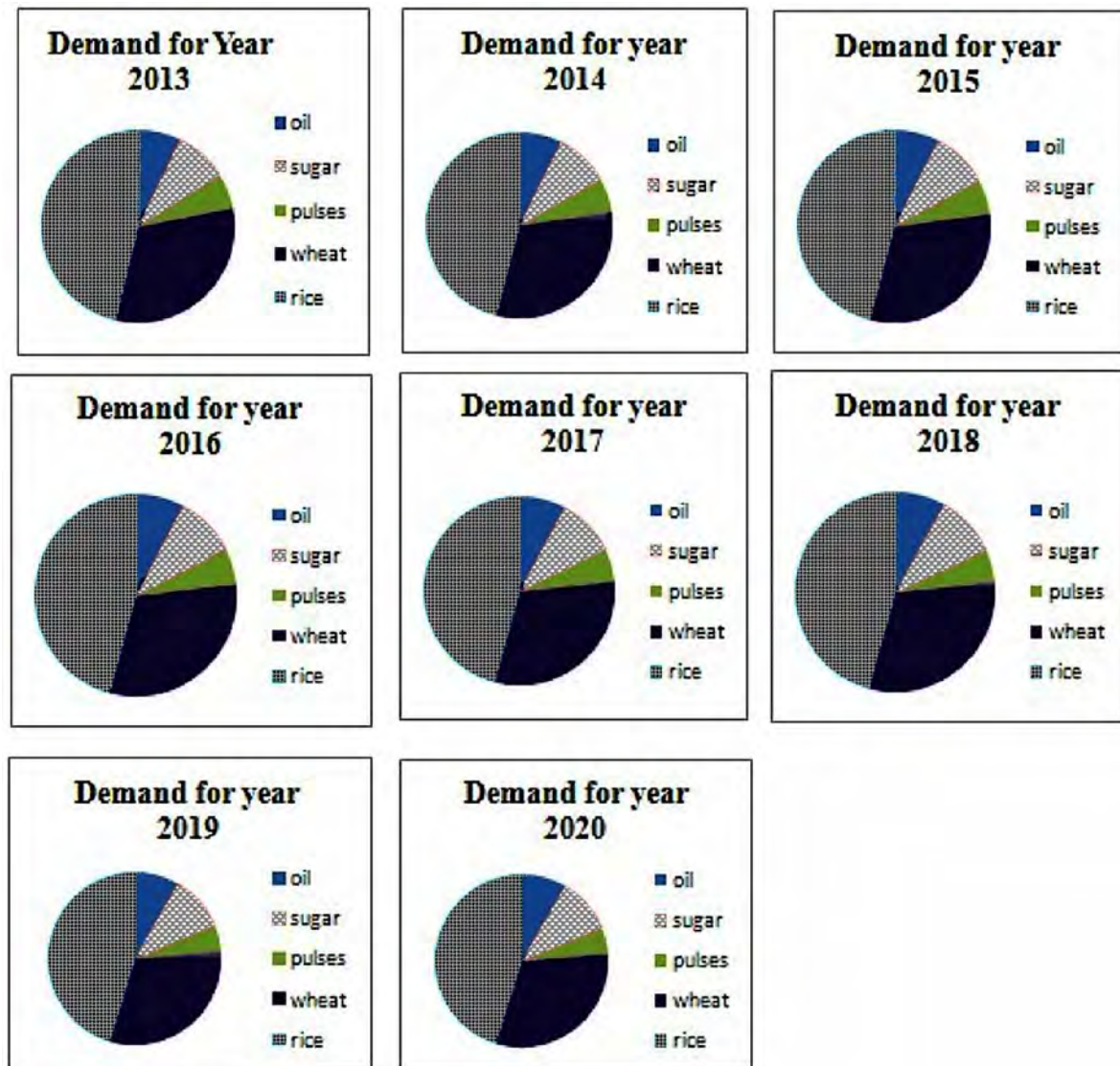
DISCUSSION

The rate at which the population is growing in all the region of the world, demand for agricultural crops/commodity may tremendously increase. Demands can be drives from population growth, income of the people, inflation, globalization; non farming activities etc... It may leads to less production of agricultural crops, which will never fulfill the exact requirements, or demands of government/people.

To overcome this problem we have chosen sample crops like Oils, Pulses, Rice, Sugar and wheat, and demand for those crops over the years from 2013-2020 (Director of Economics, 2015) is discussed in our research.

Forecasting The Demand of Agricultural Crops/Commodity Using Business Intelligence

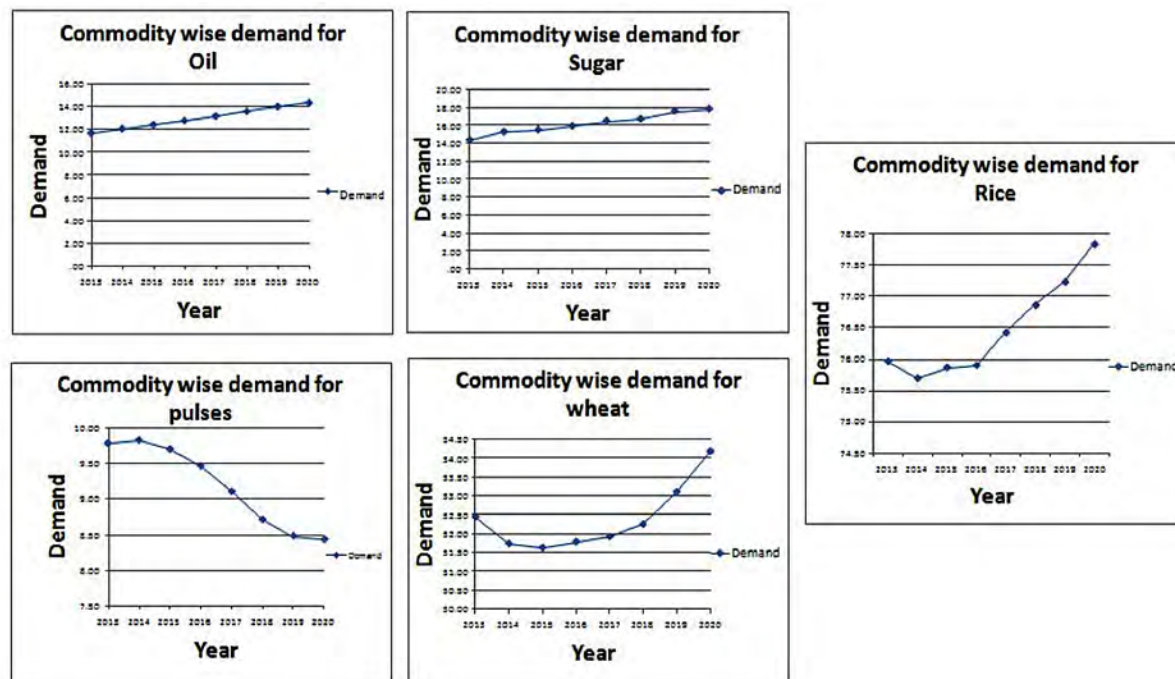
Figure 8. Forecasting of the demand - year wise



Framework extend us to work on weka tool using different set of data about the crops. Business intelligence provides data organization in such a way that the knowledge filters can easily associate with this data and turn it into information for any of the domain. Persons involved in business intelligence processes may use application software and other technologies to gather, store, analyze, and provide access to data, and present that data in a simple, useful manner. The software aids in Business performance management, and aims to help people make better business decisions by

making accurate, current, and relevant information available to them when they need it. Some businesses use data warehouses because they are a logical collection of information gathered from various operational databases for the purpose of creating business intelligence. We come across different technology provided by BI which are Extract, Transform and Loading (ETL); this platform also ensures that there are some techniques do exit for forecasting demand of agricultural crops and commodity forthcoming years.

Figure 9. Forecasting of the demand – commodity wise



Agricultural forecasting uses a wide range of techniques in a wide variety of situations. The largest group is outlook forecasts, mainly of production, at different national and regional aggregations. They have a long history and a detailed and specialized development of leading indicator analysis unique in forecasting. Emphasis, indeed one might say overemphasis, on econometric modeling of ever increasing complexity has been a hallmark of agricultural forecasting. It stems, perhaps, from the desire and the training of agricultural economists to explain phenomena rather than predict them. Often, analysts make conditional predictions or projections based on assumptions that are oversimplifications of any policy that might arise. The ratio of policy analysis to long-term forecasting appears higher in agricultural applications than elsewhere. If this review appears to have concentrated on short-term forecasting, it is because few published long-term forecasts were located.

Results found here conform generally to the beliefs held by forecasters. The conclusions

are drawn from published studies and there are many unpublished forecasts. Although the types of forecasts required are similar to those found in business, agricultural forecasters have made little use of univariate time series methods. More common in the general forecasting literature is attempts to mechanically calculate correct (i.e. well calibrated) forecast probability distributions.

Table 1. Different types of current BI techniques used

Sl.No	Techniques	Description
1	Predictive Modeling	Predicts a value for different set of data records and attributes.
2	Classification	Determines to which class a different data records belongs.
3.	Characterization and describing data mining.	Data distribution, dispersion and exception.
4	Association and correlation analysis.	Identify the relation between different attributes.

The problem was recognized in agricultural forecasting in the late 1970s. Table 1. (Ranjan, 2005) describes brief description about the BI techniques used in our research.

FUTURE RESEARCH DIRECTIONS

A much-needed development is to combine forecasting with decision making (the so-called decision support system). Agricultural economists because of their historical emphasis on the analysis of resource allocation decisions, are perhaps more likely to emphasize this area than are other forecasters. BI framework can predict demands of agricultural crops/commodity. Deep learning is another technology, which could be used for more realistic analysis in the future. Appropriate prediction of food supply over the global regions and shortage of supplies can be estimated to fulfill required food supplies. For futuristic scope we can also use this framework to predict the demands over 50 years across the globe.

CONCLUSION

In this paper it is shown how BI and open source tools could be used to make the forecasting. Good results could be obtained by using data mining techniques along with business intelligence. Also very effective reports and dashboards could be generated, which provide multidimensional perspective of the available information. These are very helpful in making dynamic decisions with greater accuracy and effectiveness. Better forecasting of the demands is possible, if other parameters like demography; food price, global demand and changing nature of food consumption among people are also considered. One of the other important factors would be the changing environmental condition.

Food supply is important to national security; various governments have attempted to quantify agricultural production and to exert some control

over it. In the beginning, simply collecting and tabulating data on the current agricultural situation was a major challenge, and agricultural statisticians played a major role in the development of statistical methods (National agricultural statistics, 2015; Department of Agriculture, 2012; Department of Agriculture, World Agriculture Supply and Demand Estimate Report, 2014). In the coming days forecasting the agricultural demands would be a challenging task, and BI framework can help achieve proper forecasting food requirements.

REFERENCES

- Bouckaert, R. R., Frank, E., Hall, M., Kirkby, R., Reutemann, P., & Seewald, A. (2013). *WEKA Manual Version 3.7.8*. Hamilton, New Zealand: University of Waikato.
- Department of Agriculture. (2012). *World Agricultural Supply and Demand Estimates Report*. Department of Agriculture.
- Department of Agriculture. (2014). *World Agriculture Supply and Demand Estimate Report*. Washington, DC: UASD.
- Director of Economics. (2015). *Publication January 2015*. Director of Economics and Statistics.
- Elmasri, R., & Navathe, S. B. (2011). *Fundamentals of database system* (6th. ed.). Boston: Pearson Education, Inc.
- Esri's. (2014, August 10). *Geographical Information System and Business Intelligence*. Retrieved March 30, 2015, from [www.esri.com: http://www.esri.com/library/whitepapers/pdfs/gis-and-business-intelligence.pdf](http://www.esri.com/library/whitepapers/pdfs/gis-and-business-intelligence.pdf)
- Hamilton, J. R. (2012). Gathering and deploying tourism destination intelligence. *International Journal of Electronic Business*, 10(2), 185–203. doi:10.1504/IJEB.2012.051119
- Kruse, J. (2010). *Estimating Demand for Agricultural Commodities to 2050*. Global Harvest Initiative.

Legorreta, L., & Valentine, J. (2014, November 12). *Business Intelligence*. Hubli, Karnataka.

National Agricultural Statistics. (2015). *Education, Outreach and understanding statistics*. Washington, DC: USDA-NASS.

Pentaho, I. (2014, July 30). Retrieved October 14, 2014, from infocenter.pentaho.com: http://infocenter.pentaho.com/help/topic/pdi_user_guide/pdi_user_guide.pdf

Pentaho Community, W. H. (2014, August 20). *Data Mining Using Weka Forecasting*. Hubli, Karnataka, India.

Prasad, R. N., & Acharya, S. (2011). *Fundamentals of Business Analytics*. Wiley-India.

Ranjan, J. (2005). Business Intelligence: Concepts, Components, Techniques and Benefits. *Journal of Theoretical and Applied Information Technology*, 9(1), 60-70.

SAS Analytics. (2015, January 10). SAS. Retrieved March 20, 2015, from [www.sas.com: http://www.sas.com/offices/europe/denmark/pdf/102719.pdf](http://www.sas.com/offices/europe/denmark/pdf/102719.pdf)

Sherman, R. (2014). *Business intelligence guide book: From data integration to analytics*. MK.

Shmueli, G., Patel, N. R., & Bruce, P. C. (2010). *Data mining for business intelligence*. Hoboken, NJ: Wiley.

ADDITIONAL READING

Braun, J. v. (2007). *The World Food Situation: New driving forces and required actions*. Washington, D.C: International Food Policy Research Institute.

Bruinsma, J. (2003). *World agriculture: towards 2015/20130*. Rome, Italy: Earthscan Publication Ltd.

Islam, N. (1995). *Population and Food in the Early Twenty-First Century: Meeting Future Food Demand of an Increasing Population*. Washington, DC: International Food Policy Research Institute.

Olszak, C. M., & Ziemba, E. (2007). Approach to Building and Implementing Business Intelligence Systems. *Interdisciplinary Journal of Information, Knowledge, and Management*, 2, 137-148.

Turban, E., Sharda, R., & Delen, D. (2011). *Decision Support and Business Intelligence System*. Hawaii: Pearson.

Vadivelu, A., & Kiran, B. (2013). Problems and prospects of agricultural marketing in India: an overview. *International Journal of Agricultural and Food Science*, 108-118.

KEY TERMS AND DEFINITIONS

Business Intelligence: Business intelligence is getting the right information to the right people at the right time so they can make decisions that ultimately improve performance.

Clustering: The term clustering refers to the task of assigning a set of objects into groups (called clusters) so that the objects in the same cluster are more similar (in some sense or another) to each other than to those in other clusters.

Data Warehouse (DW): A data warehouse is a repository of information collected from multiple sources, stored under a unified schema, and usually residing at a single site.

ETL (Extract, Transform and Load): It is the primary step for applying the business intelligence. It mainly included all those preprocessing activities, which are required for transforming the data into a proper format, which are required for analysis.

Functional Dependencies: A functional dependency is a constraint between two sets of at-

tributes from the database. Functional dependencies are the formal tool for analysis of relational schema that enables us to detect the redundancies present in the clustered data.

Loading: The loading refers to the phase of loading the data into the end target, usually the data warehouse (DW).

Normalization: Normalization is the process of organizing the fields and tables of a relational database to minimize redundancy and dependency.

Predictive Analytics: Predictive analytics represent any solution that supports the identification of meaningful patterns and correlations among variables in complex, structured, unstructured, historical, and potential future data sets for the purposes of predicting events and assessing the attractiveness of various courses of action.

B

Integrated Data Architecture for Business

Richard Kumaradja

Renaissance Consulting, Indonesia

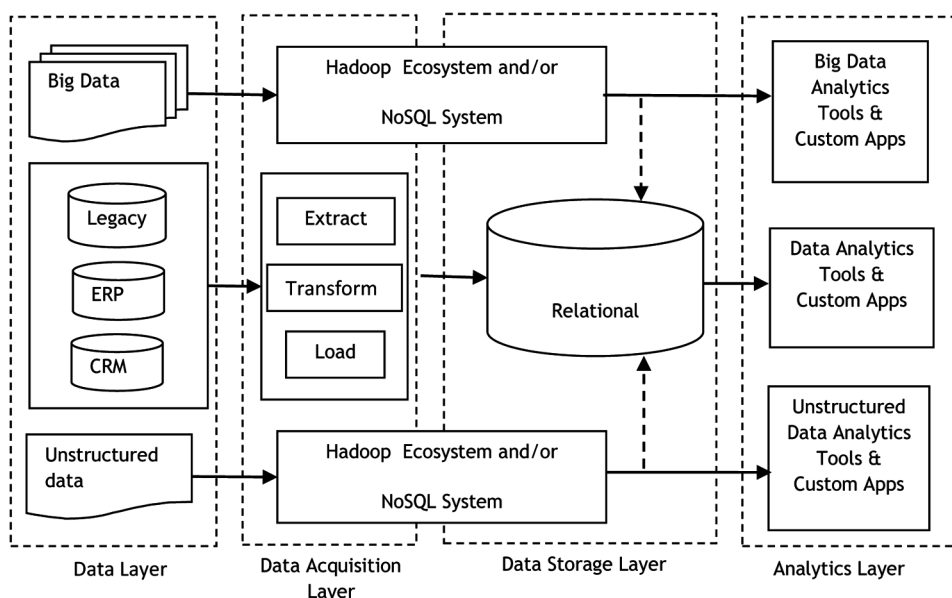
INTRODUCTION

According to (Kosala & Kumaradja, 2014), to achieve a better analytics capability business organizations must utilize three Business Analytics components altogether so that a 360 degree analysis or a more comprehensive view of data can be implemented. The three Business Analytics components are: Big Data Analytics Tools and custom applications, Data Analytics Tools and custom applications, Unstructured Data Analytics Tools and custom applications form the Analytics Layer of the Big Data Analytics within the Business Intelligence and Business Analytics framework proposed by (Kosala & Kumaradja, 2014). For convenience, the structure of the Big Data Analytics framework is shown in Figure 1.

As stated by (Kosala & Kumaradja, 2014), the integration methodology of the three analytics components in Figure 1 and how they can be utilized together to get a 360 degree analysis view of data could represent a major research effort in the next few years. One of those research areas is about the data integration framework between unstructured Hadoop or NoSQL systems to create a structured OLAP data structure, and its real business case studies.

The author believes that the data integration framework must consider all layers of the Big Data Analytics framework shown in Figure 1, namely, Data Layer, Data Acquisition Layer, Data Storage Layer and Analytics Layer. Therefore, the proposed data integration framework must consider the entire Big Data Analytics ecosystem as one integrated data architecture.

Figure 1. Big data analytics within the business intelligence and business analytics framework



DOI: 10.4018/978-1-5225-2255-3.ch074

The purpose of this article is to clarify and show the concept of an integrated data architecture as a foundation of a modern data architecture which has the potential of addressing and solving critical issues related to big data analytics implementation in business organizations. Eventually, the goal of this article is to improve the likelihood of successful big data analytics implementation in businesses.

BACKGROUND

In their article (Kosala & Kumaradjaja, 2014) has pointed to data integration framework between unstructured Hadoop or NoSQL systems to create a structured OLAP data structure as one of the promising future research areas in big data analytics. The issue, however, is that there are still some confusions about data integration as mentioned by (Russom, 2008). In the next section, the concept of data integration architecture will be clarified as well as the detailed explanation of the proposed integrated data architecture for business organizations.

MAIN FOCUS OF THE ARTICLE

The main focus of this article is to address the issues of data integration from the architectural point of view. Addressing the issues of data integration from the architectural point of view will lead to a better understanding of the current situation and better able to construct proposed solutions to those issues since architectural approach can give us a holistic and comprehensive view of the problems.

In this section, the data integration issues related to Big Data Analytics adoption will be discussed, and an integrated data architecture for implementing Big Data Analytics will be proposed.

Issues, Controversies, Problems

There are some issues and misunderstandings when we talk about data integration. (Russom, 2008) has stated that to a lot of people the term data integration architecture sounds like an oxymoron because they do not think that data integration has its own architecture. For example, some data warehouse professionals still use the methods and practices of the 1990s when data integration was subsumed into the larger data warehouse architecture. He also observed that many data integration specialists still cling to poor practice that's inherently anti-architectural by building one independent interface at a time. These misunderstandings are further exacerbated by the belief that using a vendor product for data integration automatically assures architecture.

The problems with the above misconception are:

- We can not address how architecture affects data integration's scalability, staffing, cost, and ability to support real time, master data management, SOA.
- We can not address interoperability with related integration and quality tools.

In other words, with the absence of data integration architecture, we will not be able to construct an integrated data architecture which is strongly independent, future-facing, productive, scalable and interoperable (Russom, 2008).

Why have this misconception and confusion about integrated data architecture persisted? The complexity and perception about the role of enterprise data warehouse are the major reasons behind the misconception.

Today, in most business organizations, data flow from diverse source systems such as ERP, CRM and supply chain applications where most enterprise data originates as well as non-traditional

sources such as social media and other internet-based data sources. Since these data have different models not to mention the structured and unstructured ones, the data must be transformed in the middle of the process in a variety of ways. Diverse interfaces are also required to connect these pieces together (Russom, 2008). That said, the complexity level is quite high just to organize into a data integration solution such that most people would prefer not to invest their time in formulating a holistic solution. Most will resort to a vendor product with the expectation that the product will deliver the solution by itself. This has led to misconception about integrated data architecture.

Another issue is the understanding about the role of enterprise data warehouse. Traditionally, data processing for analytic purposes follows a fairly static workflow (Brite Group, 2013). Enterprise data is structured with stable data models via enterprise applications such as ERP, CRM, SCM and billing systems. Then the data is extracted, transformed and loaded (ETL) using data extraction tool from the enterprise applications and transactional databases to a staging area where data standardization normalization occurs. Using the data warehouse, business analysts use data analytics tools to perform advanced analytics. Historically, the data extraction tools have only supported structured data primarily associated with data from internal applications. However, today, the advent of the web, mobile apps and social media has caused a fundamental change to the nature of data. Unlike the traditional enterprise data that is centralized, highly structured and easily manageable, this so-called Big Data is highly distributed, loosely structured, and increasingly large in volume. This has caused a fundamental change in the role of enterprise data warehouse as well since it is not only sufficient to accommodate structured data anymore.

The context of data integration becomes more complex such that it demands a holistic view of the proposed solution. Therefore, by constructing

data integration architecture, the overall goal of Big Data Analytics can be achieved.

Solutions and Recommendations

In this section, the proposed data integration architecture which will lead to an integrated data architecture for business organizations will be presented and discussed. The goals of constructing such an architecture are:

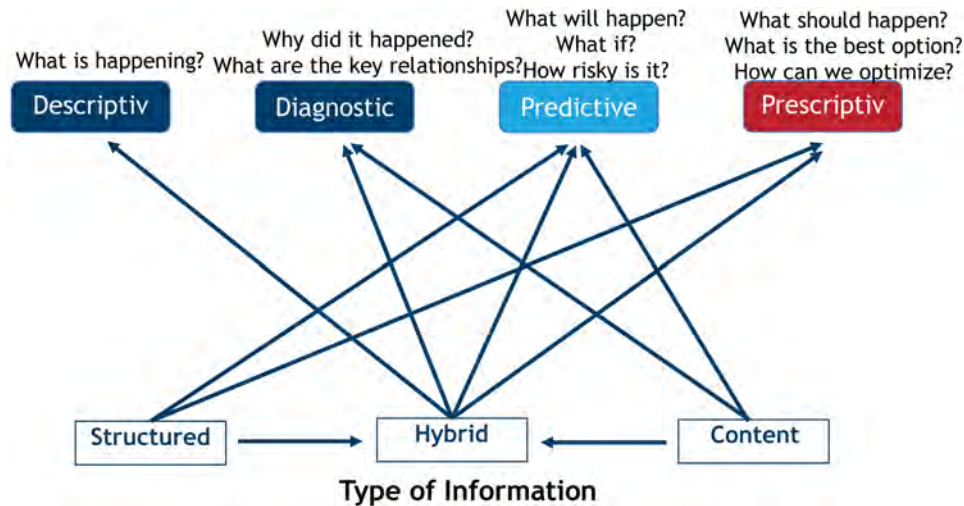
- Obtain an independent architectural description of a business organization's enterprise data consists of both structured and unstructured data which are productive, scalable and interoperable.
- Obtain a development standard derived from the architectural patterns.
- Obtain reusability and consistency through the architectural descriptions.

Once a business organization's implement an integrated data architecture the resulting ROI is obtained through standardized data analytics components which lead to increase productivity at a cost less than using various proprietary components. It is expected that an integrated data architecture may have significant impact in the next few years as more and more companies adopt Big Data Analytics as part of their strategic technology initiatives.

The proposed architecture is based on the Big Data Analytics within the Business Intelligence and Business Analytics Framework explained in (Kosala & Kumaradaja, 2014), and some recent research works about big data architectures and analytics.

In formulating the proposed architecture, it starts with the observation that the processing environment for big data analytics becomes increasingly complex. This complexity has driven us to take a different approach in formulating a data architecture which can accommodate both structured and unstructured data seamlessly for

Figure 2. Different stages of analytics based on type of information



analytics purposes. Figure 2 shows different stages of analytics based on the type of information (Sallam & Cearley, 2012).

Figure 2 shows that there are three types of information: structured, hybrid and content used in different stages of analytics namely: descriptive, diagnostic, predictive and prescriptive (Sallam & Cearley, 2012). The descriptive stage tells us what happened, the diagnostic stage tells us why certain things happened and what the key relationships are, the predictive stage tells us what will happen and can show us what if scenarios along with the risks, and the prescriptive stage tells us what should happen and can show us the best option and how to optimize. It can be seen from Figure 2 that hybrid information is used in all stages of analytics. The existence of hybrid data becomes a necessity since it is needed to perform advanced analytics. Therefore, an integrated data architecture is defined as a data architecture that can accommodate and formulate architectural description of hybrid data.

Furthermore, the proposed integrated data architecture forms an advanced data ecosystem shown in Figure 3.

It can be seen from Figure 3 that Advanced Data is at the center of the ecosystem where it is supported by six (6) critical elements:

- Scale-Out Databases deal with the support of SQL and NoSQL data to support storage platform of Big Data.
- EDW (Enterprise Data Warehouse) which must be augmented by enabling it to process structured or multi-structured data and pre-processing of raw data.
- Real-Time Analytics which is a relatively new field but becomes more and more important in the near future.
- Data Integration and Quality which forms the foundation of a solid Big Data.
- Distributed Processing which uses Hadoop technology to process and combine Big Data sources.
- Hardware & Security

Therefore, the proposed integrated data architecture is shown in Figure 4.

It can be seen from Figure 4 that there are four main components of an integrated data architecture:

1. Apache Hadoop Framework as the Main Distributed Processing

Apache Hadoop Framework, is a framework that allows for the distributed processing of large data

Figure 3. Advanced data ecosystem

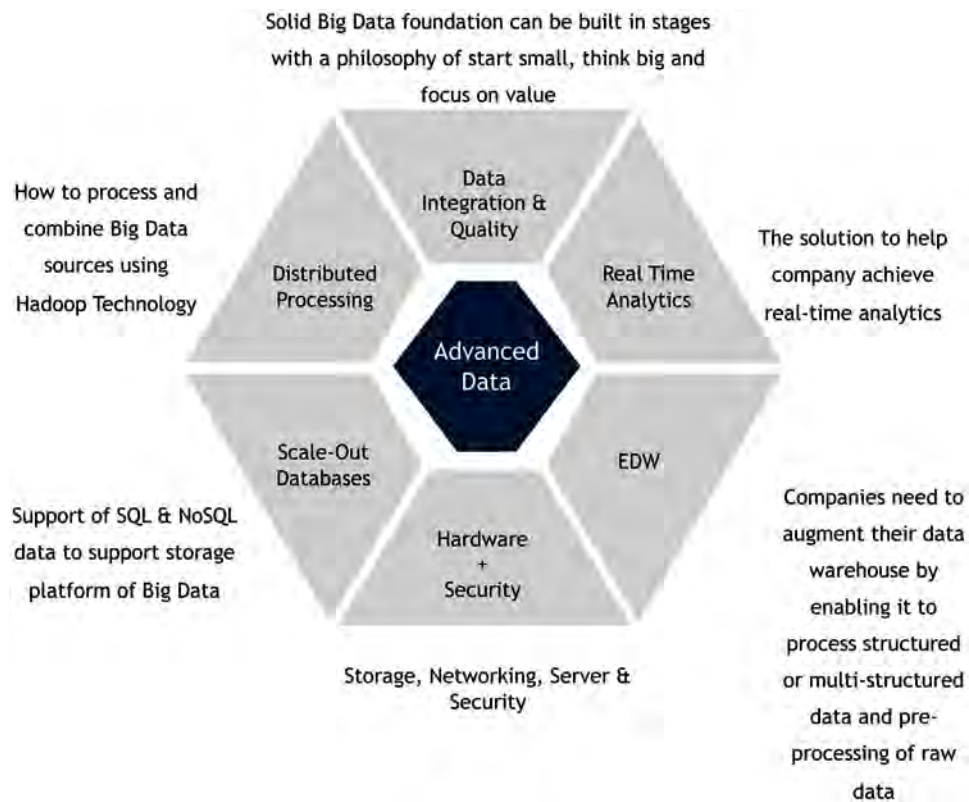
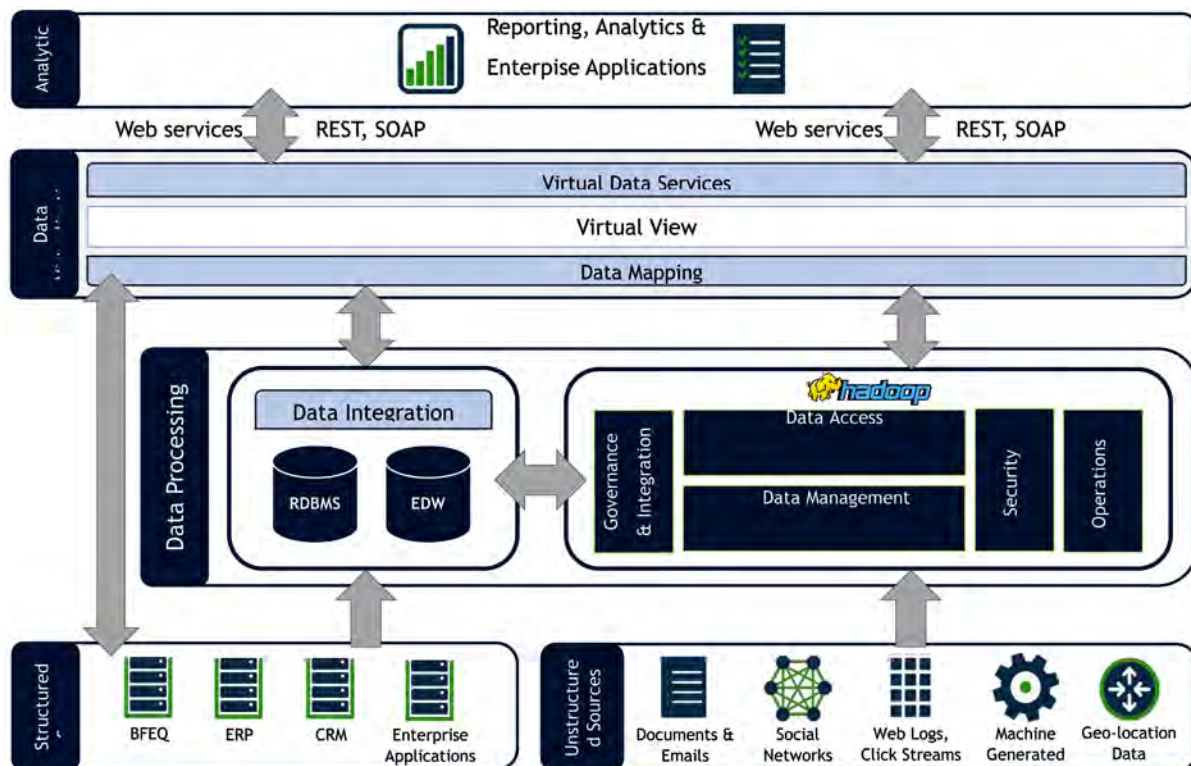


Figure 4. Integrated data architecture for big data analytics



sets across clusters of commodity computers using a simple programming model.

It is designed to scale up from single servers to thousands of machines, each providing computation and storage. Rather than rely on hardware to deliver high-availability, the framework itself is designed to detect and handle failures at the application layer, thus delivering a highly-available service on top of a cluster of computers, each of which may be prone to failures.

There are two core components in Hadoop:

- Hadoop Distributed File System (HDFS), the storage
- MapReduce, the processor

The core components in Hadoop are shown in Figure 5. We can see from Figure 5 that besides the two components HDFS and MapReduce there are other components that form the Hadoop ecosystem. They are Hive which acts as a data warehouse within Hadoop environment, Hbase which is a column-oriented big table, Pig which is a scripting language and Mahout which is a machine learning tool. All these components are still evolving primarily driven by commercial demand.

The advantages of using Hadoop are:

- **It's Scalable:** New nodes can be added as needed and added without needing to change data formats, how data is loaded, how jobs are written, or the applications on top.
- **It's Cost Effective:** Hadoop brings massively parallel computing to commodity servers. The result is a sizeable decrease in the cost per terabyte of storage, which in turn makes it affordable to model all your data.
- **It's Flexible:** Hadoop is schema-less, and can absorb any type of data, structured or not, from any number of sources. Data from multiple sources can be joined and aggregated in arbitrary ways enabling deeper analyses than any one system can provide.
- **It's Fault Tolerant:** When you lose a node, the system redirects work to another location of the data and continues processing without missing a beat.

2. Hadoop Add-On(s) as the Advanced Component for Easy Administration

To accelerate development and delivery of Hadoop-based solutions, Hadoop vendors have developed add-ons component to enhance the

Figure 5. Core components of hadoop

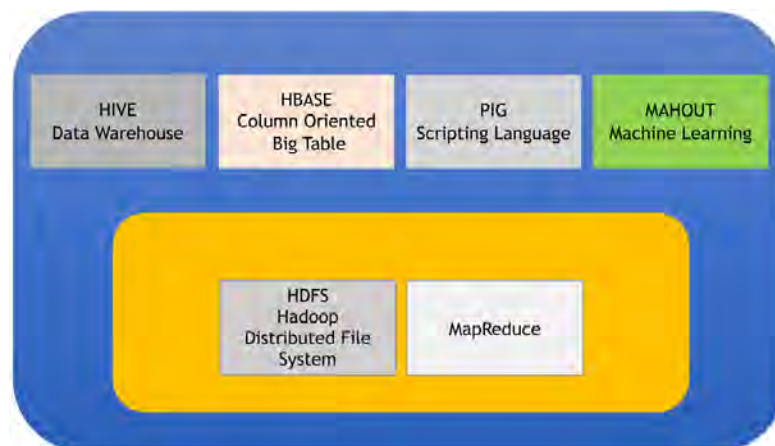


Figure 6. Example of hadoop add-on components



basic functionalities and provide administration & monitoring functions. Figure 6 shows an example of the Hadoop add-ons developed by one of the leading Hadoop vendors.

It can be seen from Figure 6 that there is a Data Operating System called YARN (Yet Another Resource Negotiator) that glues the Data Access and Data Management components. Additionally, there are other components Governance & Integration, Security and Operations have been added to harden the system. In the Governance & Integration component, Data Workflow Lifecycle & Governance subcomponents have been added, in the Security component, Authentication, Authorization, Accounting & Data Protection subcomponents have been added and in the Operations component, Provision, Manage & Monitor as well as Scheduling subcomponents have been added. The component for deployment choice has been added as well such as types of operating systems (Windows or Linux) and the types of deliveries (On-Premises or Cloud-Based).

3. Data Virtualization Abstraction Layer

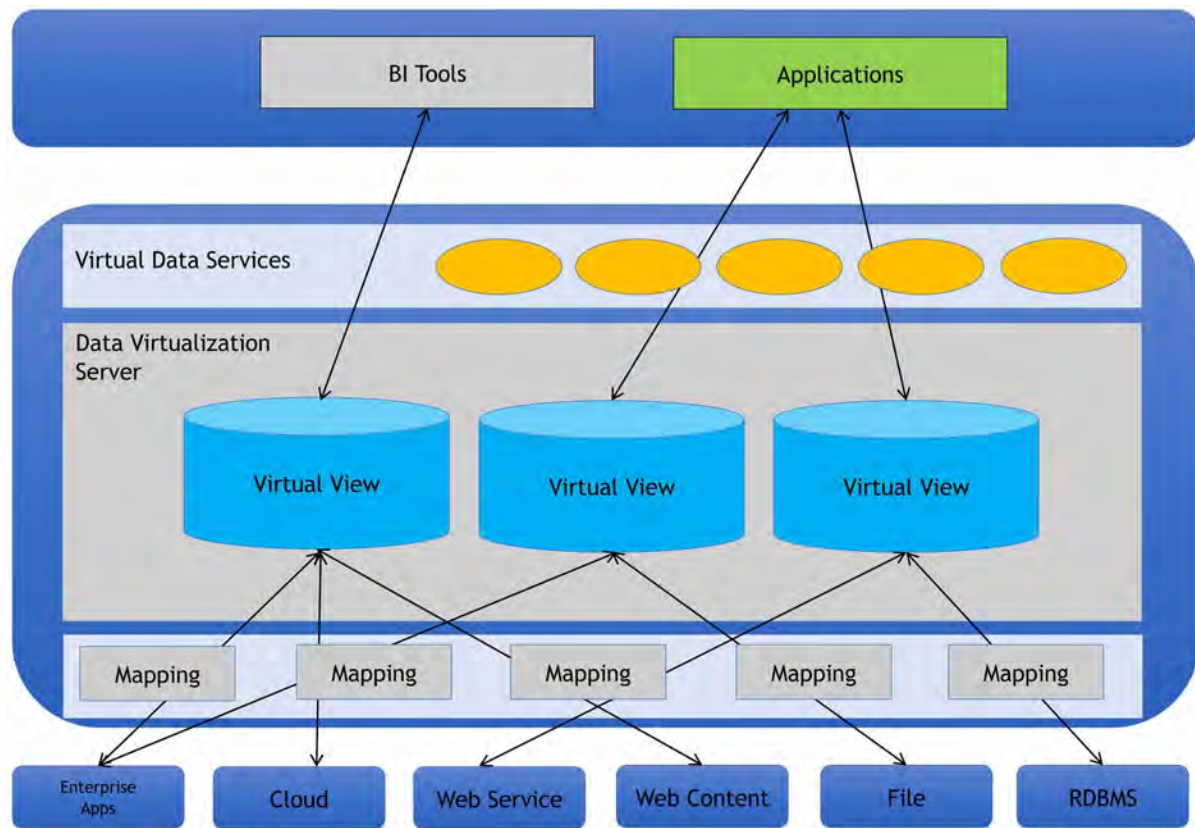
Data Virtualization logically abstracting data from different data sources and data types into an integrated virtual data layer consisting of multiple virtual data views. Virtual data views in this layer can be shared and queried on demand, and can also

published as data services that serve up real-time data in various format.

There are already many research works and articles dedicated to data virtualization, among them are (van der Lans, 2014), (CITO, 2014), (Yuhanna, 2015), (Chandramouly, Patil, Ramamurthy, Krishnan, & Story, 2013), (Ferguson, 2014) and (Szabo, 2011). All of these articles have indicated the benefits of implementing data virtualization in business organizations. The benefits, among others, are:

- **Increased Agility:** The agility of the IT system is significantly improved because developers only have to focus on what data is needed and what data must be manipulated such that less code is needed. Making changes is also easier because the applications have been decoupled from the data sources.
- **Improved Time-to-Market:** Since data can be integrated faster, it simplifies the implementation of new systems and their integration with existing systems. Therefore, the resulting impact is a significant improvement of time-to-market so that organizations can react faster.
- **Improved Productivity and Maintenance:** As indicated and explained above, integrating data and systems with data virtualization is easier than with

Figure 7. Conceptual view of data virtualization



many other integration technologies due to the decoupling of applications and data sources.

- Improved Data Quality:** Using a data virtualization server, validation and correction of new data can be conducted before the corresponding data is inserted and updated in the source systems. Furthermore, when there is a false data retrieved from source systems, the data virtualization server can apply all types of data cleansing operations. These operations produce a higher level of data quality.

Figure 7 shows the conceptual view of data virtualization.

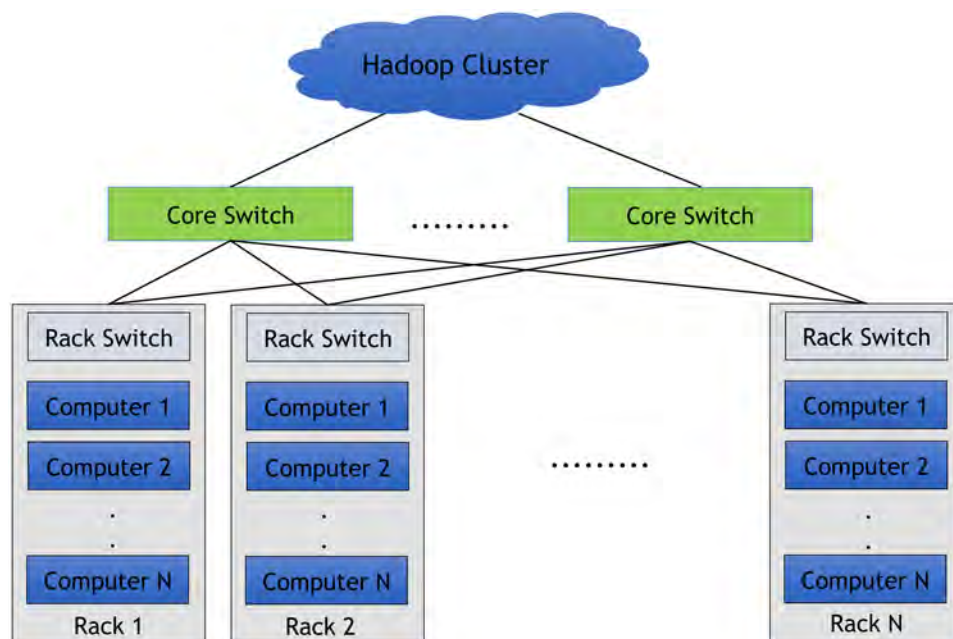
4. Apache Hadoop Cluster

Normally any set of loosely connected or tightly connected computers that work together as a single system is called a cluster. In simple words, a computer cluster used for Hadoop is called Hadoop Cluster.

Hadoop Cluster is a special type of computational cluster designed for storing and analyzing vast amount of unstructured data in a distributed computing environment. These clusters run on low cost commodity computers. The conceptual view of a Hadoop Cluster is shown in Figure 8.

There are already successful business applications of using Hadoop Cluster, among them are Yahoo who have more than 10,000 machines running Hadoop and nearly one (1) petabyte of user data. Others are:

Figure 8. Conceptual view of a hadoop cluster



- **Search:** Amazon.
- **Log Processing:** Facebook, Yahoo.
- **Data Warehouse:** Facebook.
- **Video and Image Analysis:** New York Times.

Besides implementation in well known e-commerce companies, Hadoop has also been successfully implemented in many financial services companies such as Bank of America, Morgan Stanley, OCBC Bank, ANZ, Deutsche Bank and RBS.

FUTURE RESEARCH DIRECTIONS

The future research directions of the proposed integrated data architecture will be in two (2) areas:

1. Data Warehouse Replacement or Augmentation?

The replacement of EDW into Hadoop became a trend topic in the IT world since Hadoop brings better performance at lower cost than EDW. Since

the biggest cost in Big Data is storage, Figure 9 shows the storage cost comparison by using Hadoop based on the survey shown in (Hortonworks, 2014).

It can be seen in Figure 9 that the storage cost per raw TB of data for Hadoop is the lowest among in-premise systems. Therefore, using Hadoop for large scale data warehouse applications can significantly lower the TCO of corporate ICT systems. This will motivate many companies to rethink their data architectures. However, there is a critical question that must be addressed. The question is whether it is possible to replace the traditional EDW with a Hadoop ecosystem?

Based on the survey conducted by Gartner shown in (Elliot, 2014), there were 11% of the organizations surveyed in 2012 who said they will replace their EDWs with Hadoop, however, in 2013 the figures dropped to 5.22%. The resulting conclusion drawn from the survey is that there is no relationship between the EDW and Hadoop at the moment – instead they are going to be complimentary. RDBMS or MPP will not be replaced, but instead big data analytics professionals will

Figure 9. Hadoop storage cost comparison

Storage Type	Fully-loaded Cost Per Raw TB of Data (Min-Max Cost)
Cloud	< \$2,000
HADOOP	< \$3,300
Network Attached Storage (NAS)	\$12,000 - \$18,000
Engineered System	\$11,000 - \$20,000
Massively Parallel Processing (MPP)	\$21,000 - \$80,000
Storage Area Network (SAN)	\$37,500 - \$180,000

use the right tool for the right job – and that will very much be driven by price.

In the near future, the research should not be focused on replacing traditional EDWs with Hadoop but how to use Hadoop to augment their functions in the big data analytics ecosystem.

2. Stream Processing and Stream Analytics

This is a relatively new but very important area of Hadoop research. Some Hadoop vendors now provides the stream processing capability by combining it with Apache Storm in the data access component of Hadoop architecture. Future research should focus on its enhancement for real time analytics applications.

CONCLUSION

In this article, the concept of integrated data architecture from the viewpoint of structured and unstructured data, its importance for big data analytics, the components of integrated data architecture have been clarified. Toward the end, an integrated data architecture for business organizations and the main components of the architecture have been proposed and explained. In the end, there are some discussions about the

research directions of the proposed integrated data architecture with the expectation that it will increase the likelihood of successful big data analytics implementation.

REFERENCES

- Brite Group. (2013). *How can Big Data/Hadoop co-exist with your Enterprise Data Warehouse*. Brite Group white paper.
- Chandramouly, A., Patil, N., Ramamurthy, R., Krishnan, S. R., & Story, J. (2013). *Integrating Data Warehouses with Data Virtualization for BI Agility*. IT @Intel white paper.
- CITO Research. (2014). *Putting the Data Lake to Work A Guide to Best Practices*. CITO Research white paper.
- Elliot, T. O. (2014). No, Hadoop Isn't Going to Replace Your Data Warehouse. *Business Analytics*. Retrieved from <http://timoelliott.com/blog>
- Ferguson, M. (2014). *Data Virtualization – Flexible Technology for the Agile Enterprise*. SAS white paper.
- Hortonworks. (2014). *A Modern Data Architecture with Apache Hadoop – The Journey to a Data Lake*. Hortonworks white paper.

Kosala, R., & Kumaradjaja, R. (2014). Big Data Analytics for Business. In Encyclopedia of Information Science & Technology (3rd ed.). Academic Press.

Russom, P. (2008). *Data Integration Architecture: What It Does, Where It's Going, and Why You Should Care*. The Data Warehousing Institute.

Sallam, R. L. & Cearley, D. W. (2012). *Advanced Analytics: Predictive, Collaborative and Pervasive*. Gartner Report.

Szabo, P. (2011). *Data Virtualization and Federation*. Stone Bond Technologies white paper.

van der Lans, R. F. (2014). *Creating an Agile Data Integration Platform using Data Virtualization*. R20/Consultancy technical white paper.

Yuhanna, N. (2015). The Forrester Wave™. *Enterprise Data Virtualization, Q1, 2015*.

KEY TERMS AND DEFINITIONS

Big Data: The size, speed and sorts of data that exceed an organization's traditional database capacity to access, integrate, store, and analyze for accurate and timely decision making.

Big Data Analytics: The process, techniques, and tools used to gain insights from Big Data so that the decision making process could be optimized.

Business Analytics: The broad use of data and quantitative analysis to make business decisions in corporations.

Data Mining: The computational process and technique to find and discover patterns from large data sets.

Data Virtualization: Data management approach that allows an application to retrieve and manipulate data without requiring technical details about the data.

Hadoop: An open source technology developed under Apache Software Foundation that can be used to process Big Data in a distributed manner.

NoSQL: A collection of non-relational database technologies that are designed to store unstructured web data or documents.

Structured Data: Data that is defined and organized in a structure. For example: data in tables, data in relational databases.

Text Mining: The use of data mining techniques to discover and extract patterns from text.

Unstructured Data: Data that has no identifiable structure. For example, images, videos, email, text documents, and web documents.

Web Mining: The use of data mining techniques to discover and extract information from web documents and services.

IT Strategy Follows Digitalization

B**Thomas Ochs***Villeroy & Boch, Germany***Ute Anna Riemann***SAP SE, Germany*

INTRODUCTION

Digitalization is one of the most powerful and important phenomenon than any technology innovation before (Nordstöm, 2015). Digitalization has an enormous value and with it everything has the capability to change (Salo, 2006). With the ability of changing the way, products and services are consumed by the suppliers, consumers and partners completely different ways traditional industries being blurred (Rai, A., Pavlou, Im, & Du, 2012; Ray, Muhana, & Barney, 2005; Sambamurthy, Bharadwaj, & Grover, 2003).

In this sense the question of a new IT strategy emerges as companies need to understand where to invest in digital technology and what it deserves to get the highest value out of this investment. Much of that investment concentrated on applying new technologies to existing products, practices and processes. Going mobile, adding analytics, or extending the online experience are essential considering the transformation of the business with digital, in the face of changing consumer expectations, options and information.

Refining the business and digitalize them is more than just extend the IT strategy with digitalization by new functions and processes. Digital is more than a set of technologies. It is about the abilities those technologies create. Digital is the application of information and technology to raise human performance and in this sense the human factor is the essence of digital transformation as it creates the type of value that leads to revenue. In addition – of course - goals for creating efficient

cies that largely drive down the cost of creating short-term benefits but drain the economy and growth needs to be considered. Digital technology has a transformation potential starting in its silos such mobile, cloud, process, customer, supply chain, etc. (Rai, Pavlou, Im, & Du, 2012; Klein, & Rai, 2009; Saraf, Langdon, & Gosain, 2007; D’Adderio, 2001). Now it is the time to go down to the essence of the IT strategy asking the question how can a business win using information and technology raising human performance.

DIGITALIZATION

Background

Currently we see a lot of information addressing the topic of digitalization. Beyond the pure buzz-word a lot of previously effective and useful assumptions, structures and procedures have to be adapted to become successful in the new digitalized world. Our chapter focuses on the company’s IT and highlights in detail what the required changes in its various levels and dimensions for corporate IT are to meet the new digital requirements. The idea of our chapter is to make the relevant levels and dimensions transparent, outline how the various elements interfere with each other and how this transforms the corporate IT in regards to its positions and importance of the company, the structure and processes within an IT organization and the requirements towards agility and skills.

What Does Digitalization Mean?

As a trend, the substantial growth in the output in the recent years has changed. What is now of importance for many firms is the adoption of 'customer orientation through the reengineering of the value proposition of the desires and expectations declared by customers. This requires an increases for of collaboration, information sharing and flexibility (Jarach, 2002). In that sense the digitalization creates new patterns of action for the companies and especially for the IT. In simplified terms, the digitalization impacts the IT in various ways which should be reflected in an appropriate IT strategy, IT organization and IT processes. The impact the digitalization has towards the IT is huge as digital tools is to increase the coordination and adaptation possibilities since many organizations are reaching for the goal that is changing. The coordination with help of IT is possible if initial investments and adaptations to IT are made. The focus is on simple product, service and money exchange rather than knowledge or informational exchange. To move into this new digital era as a traditional organization it is necessary to accelerate the digitization of business processes that goes should go beyond simply automating an existing process. Processes must be re-invented, including cutting the number of steps required, reducing the number of documents, developing automated decision making, and dealing with regulatory and fraud issues. In parallel the operating models, organizational structures, and roles need to be redesigned to match the reinvented processes. Data models have to be adjusted and rebuilt to enable better decision making, performance tracking, and customer insights. Real-time reports and dashboards on digital-process performance permit managers to address problems before they become critical. All this is just not possible once the IT stays in their traditional role!

Why Does Digitalization Force a New IT Strategy?

One of the biggest challenges for mainly every company in today's times and thus as well for any internal IT is the digital transformation that will have a massive impact on all businesses and business areas. This impact has a high importance not only for companies emerged in the digital age but as well for companies of the old economy and does have relevance for the business models, new competitors and competitions as well as chances and risks that have to be considered due to the new networks and digital possibilities.

To succeed in this new digital era a clearly formulated digital corporate strategy is a fundamental pre-requisite - all business areas need to learn to think and act "truly digital". The setting and tracing of a digital agenda needs to have top priority in every area of a company. In addition any company need to find an answer to the question how the existing business model will be affected and possibly change or have to be adapted due to the digital transformation.

The digitalization itself demands for a superior and strategic IT view – just not the pettiness in which we often remain. IT conveys a certain importance in this sense as with the right use of the new IT capabilities new chances and strategic competitive advantages can be generated (Agarwal, & Lucas, 2005). This consequently means that the thesis of Nicolas G. Carr from 2003 "IT doesn't matter" (Carr, 2003) is disconfirmed.

The power of a digital IT strategy lies in its scope and objectives. Therefore the question to be answered is how an Information Technology (IT) strategy should be developed.

While the IT strategy basically regulate the longer term deployment of information and communication systems of a company, it is difficult to define general rules for formulating them. In the

light of digitalization we believe that IT becomes a dominant role in the corporate strategy, i.e. where IT is used in internal and external operations to create a competitive advantage (Kohli, & Grover, 2008). Therefore the IT strategy is the pivotal point the digitalization. In this chapter we will first outline what – according to our understanding – defines a new digital IT strategy, the organizational governance and steering model required and the conditions under which such an IT strategy shall be developed. Following we will highlight the core elements of the strategy framework and the potential for transformational innovation.

Even though the fundamental understanding remains unchanged, that the business strategy and the priority initiatives need to be aligned to the business value and that the digitalization can only be successfully succeeded within a cooperative collaboration between business and IT (Bharadwaj, El Sawy, Pavlou, & Venkatraman, 2013). IT does need a new definition of its own role that goes beyond the IT operations and IT maintenance (Brynjolfsson, & Hitt, 2000; Devaraj, & Kohli, 2002; Santhanam, & Hartono, 2003). The questions to be answered need to address the scope, the scale and the core elements of the IT strategy with the following key questions that drive the definition of the IT strategy:

- What is our digital vision? And how does this vision impact our stakeholders?
- Where does the digital strategy fit in our company structure?
- What are our digital priorities?
- What is your business process maturity and standardization level?
- What sources of data will significantly change our business?
- How do you envision machine and people collaboration?
- What is your vision for real-time collaboration with suppliers and business partners?
- What is your strategy for the workforce?

- How are you digitizing your product and services?
- Do your employees have the right tools and information to drive effectiveness and reduce hierarchies?

The following chapter will highlight core elements of such a newly defined role and a target picture for a corporate IT.

Scale of the Digitalized IT Strategy

Today's internal information management is mainly concerned with the provisioning of a corporate IT infrastructure, the process support and the data management (Barua, & Mukhopadhyay, 2000). With the harmonization if the application landscape throughout the entire company as the key target as of now has already been achieved (Rai, Patnayakuni, & Patnayakuni, 2006): a globally used software landscape as a robust and trustworthy platform for the main business activities supporting the business processes effectively and efficiently has been put in place (Gartner, 2016). However, this target is no longer in the core focus – even more, there are new challenges arising at the horizon in the area of business process support: in addition to the process support the IT department need to step in into the role as a “business enabler” in order to let the company participate from the chances of the digitalization. No matter if these changes are disruptive or evolutionary for a company, IT has to take over a core role whereas the IT strategy is based on the digitalization (Melville, Kraemer, & Gurbaxani, 2006; Wade, & Hulland, 2004).

According to our understanding the emergence of the new digitalized IT strategy can be summarized as highlighted in the Figure 1:

We believe that the path forward for a true digitalization needs to be reflected in the way IT is perceived as a strategic element within the company and how agile IT is able to adapt to the new IT trends:

Figure 1. Emergence of the new digitalized IT strategy



- Cooperative Collaboration of the IT with Board, Top Management and Business:** It is essential to define the way the IT is organized and how the IT is linked to the company's management board as the key decision-making unit of a company. We believe, that the CIO and thus the Corporate IT need to be member of the board to build-in the increasing importance of the IT (Gartner, 2016). The link to the Corporate IT itself to the various business units need to be anchored with the establishment of dedicated key account roles making sure that the entire business spectrum as addressable and that the role of an internal consultancy and trusted advisor can be established (Baskerville, & Land, 2004).
- Continuous Consideration of IT Technology Trends:** Especially in the past 2-3 years newly developed IT technologies had a tremendous impact on the companies and their value proposition (Luftman, & McLean, 2004). This can be determined on the stellar ascent of the topic „Industry 4.0“ that is overall being celebrated as the 4th industrial revolution (Hames, 1994).

Based on the close alliance of computer technology with electronics and mechanics (Embedded Systems) and the “any-time-any-place” communication ability via internet (Internet of Things) new opportunities in production, logistics, sales and marketing leading to new types (smart) product and (smart) services emerge.

A further technology dynamics will be achieved due to the enhanced ability of mass data processing of various formats, data sources and structuring levels in real time (= big data) (Pavlou, & El Sawy, 2010). Linked with complex statistical concepts and methodologies (= Predictive Analytics) new possibilities of the gaining in knowledge and insight emerge that go far beyond the usual type of reporting functionalities As already described by Davenport the ability to analyze data becomes a critical capability for contemporary organizations (Davenport, 2006). Companies such as Google clearly stated: “the one that owns the data owns the key for success”.

Additionally the accessibility and velocity of IT usage has changed dramatically. Cloud-based solutions offer new possibilities in regards to a faster deployment, flexible usage, scalability, global accessibility and actuality (see NIST).

As previously said, the new IT role needs to be geared towards the digitalization and can be characterized with the following three core statements (Bharadwaj, El Sawy, Pavlou, & Venkatraman, 2013):

- **IT Shall Be the Pioneer of a New Digital Culture:** Within the digital world taking over risks becomes a cultural norm. Therefore the new digital culture has to embrace failure as a prerequisite for success. Whether culture drives technology adoption or whether technology changes the culture is still an open question. Beth Israel Deaconess' Halamka stands on the culture side of the question. "I have never seen a technology drive change on its own," he says. "Culture leads the adoption of technology. Our ability to innovate depends on the impatience of our culture." (Kane, 2015)
 - IT projects need to strive for the new digital culture.
 - Moreover, all acting and behaving need to be targeted to anchor this digital culture within the company.
 - IT needs to consequently ask the question how processes can be newly defined with the advantage of digital appliances.
 - All operations and maintenance activities need to be realized as efficient and effective as possible as this "operational obligation" shall not consume more time than necessary to leave space for the digital challenges.
- **IT Shall Promote Technological Innovation:**
 - The relevant IT technologies need to be identified and adapted at an early stage. In this sense the clumsiness of the IT application landscape needs to be overcome while a stable and global operations must be secured.
- The era of IT monoliths are over and will be replaced by an orientation on available market solutions integrated according to the Lego principle. This landscape will as well lead to solutions that are more agile and enable a faster provisioning of solutions to the business according to the 80:20 rule.
- The focus on standardization, speed, and user experience is key (Gartner, 2016). This leads as well to the usage of cloud solutions. The consistent application of a hybrid approach with the use of public cloud applications if a further market differentiation with an on premise solution is unrealistic.
- Service & Support needs to be productized based on globally standardized processes to gain freedom for the leadership role that allows an autonomous acting in pre-defined target corridor.
- **IT Shall Take Over Leadership:** IT needs to have a comprehensive knowledge in Best Practices in various disciplines to be applied that are core to the company (e.g. logistics, production,...). While having established this knowledge bases IT has potential of becoming a trusted advisor rather than being an assistant for the business.

Scope of the Digitalized IT Strategy

The scope of the digitalized IT strategy, expressed in vision and mission, is the essence of any strategy. Vision and mission contribute the key points to make the core elements transparent for the organization. The embodiment of the mission delivers the frame of the new IT strategy. The internal IT management is the core and corporate-wide solution provider for all information and telecommunication technology having the full sourcing-, organization- and architectural responsibility (Amit, & Schoemaker, 1993; Conner, & Prahalad,

Table 1. New orientation if the IT based on key characteristics

Characteristics	Description
Alignment with the business	The close alignment with the business is a key success factor and essential for the success within a digital environment.
Costs & Efficiencies	Turn away from a pure cost efficiency towards a cost optimization that balances the costs and benefits.
Degree of standardization	Move from a software monolith towards a multi-supplier strategy. Process harmonization and process integration remains a core target.
Localization	The internal information management was, is and will be centrally organized and allows localizations only where needed.
Maturity	The change from a stable software with a high level of maturity towards innovative and modern software solutions require to take the risk dealing of a less mature software within the company.
Size of the IT	The IT will grow compared to other business units. However, the core elements of the IT organization remains stable such as leadership, demand generation, application design, support and maintenance.
Competency and Organizational Culture	The required skill and competencies of the OT staff will change dramatically. The classical IT competencies (e.g. IT administration) will decline in its importance where the new challenges such as flexibility diversity, digital native and innovative behavior grow in its importance.

1996; Wernerfelt, 1995). Therefore, the IT is the natural innovation partner when it comes to digitalization and has a key responsibility within the proactive identification and analysis of IT technology trends to further increase the value of the company (process optimization, revenue potentials, etc.) (Berger, & Ofek, 1995; Lang, & Stulz, 1994). An additional core element is the holistic realization of projects including the support of the required transformation. It will no longer remain in their core competency of purely deliver the IT components but support a sustainable value proposition with a continuous support of the business leading to a continuous change process and management. Furthermore IT will still take

care of the design and standardization of business processes to achieve a degree of business process excellence, the provisioning of IT solutions and applications that meet the requirements of the requesters in regards to usability (user experience) and performance (Gartner, 2016). To secure this timely and efficient provisioning of new applications and features, agile methodologies and the usage of the so 80:20 rule will be applied to shorten the time to market. The guarantee of an appropriate data security and data protection that meets the legal requirements will still remain a core target. The following Table 1 outlines the strategic reorientation of the IT with its core characteristics.

Scope and Principles of the Digitalized IT Strategy

IT Organization

To meet the new requirements the digitalization asks for, the IT organization needs to be adapted in regards to organizational structure, number of employees, roles, profiles and competencies. Core element of the new organizational structure should be competence centers to bundle core competencies and the End-to-End process knowledge (Collins, & Porras, 1996). The further development of the employees must be oriented towards the new tasks of the IT – formulated in the new roles of Table 1 which is mainly transformation management, integration abilities, creativity and flexibility. The ability to adapt quickly to change also stands out as an important capability. According to Perry Hewitt agility is more important than technology skills. Emory professor Konsynski concurs: “The 21st century is about agility, adjustment, adaptation and creating new opportunities (Buschek, 2015). Table 2 gives a brief overview of the new organizational elements within a modern IT organization.

For being able to live up to all new activities and tasks we assume to increase the headcount of the internal IT organization by 10% within the

Table 2. New organizational elements of a modern IT organization

Role	Description
Usability Lab	Implement a cross-functional team across all applications to secure an enhanced and state-of-the-art usability for all core applications. To gain further impact a specific usability lab concentrates all activities on this topic.
Digitalization and innovation hub	Establish a dedicated organizational unit at a trend-setting location with access to „Digital Natives“, where Trend Scouts and Business Transformation Manager work together as a co-innovation team.
Shared Service Center	Set-up a team with specialized duties in regards to software development and Application LiveCycle Management (ALM).
IT Management	Global structuring of the entire IT organization with regional managers to directly address regional specifics without losing the global directive of the core IT.

next 3-4 years. In addition the core competency of the IT employees will shift towards the new digital requirements e.g. Solution Designer, Usability experts, Project Manager for digitalization projects, technology trend scouts, data scientist and data mining specialists for big data.

IT Steering- and Governance Model

Companies are and will be increasingly confronted with considerations that are less related with business and more about work, people and society as such. For anchoring the required new structures, roles and processes it is necessary to establish a new and corresponding steering and governance model that comprises the new planning process, portfolio management, internal process steering and IT controlling. The crux is to know what to digitize and automate and how to achieve this. Therefore, this new steering and governance model needs to be structured in a way to allow a better value creation through innovation, flawless execution and intense collaboration. It needs to be driven by the target to digitalize as much as possible and as consequent as possible to not ample opportunities to optimize business processes, better focus customer experience and employees

satisfaction. Overall, a stronger collaboration and synchronization between all business partners is required in this new digital era. The following table describes the elements of a new level of governance dedicated to ensuring the success of the digital strategy. It outlines the core elements and responsibilities of the modernized steering and governance model as stated in Table 3.

IT Infrastructure

The development of an IT strategy, must take into account what is happening in the external environment in order not to miss opportunities and possible threats (Bharadwaj, El Sawy, Pavlou, & Venkatraman, 2013). This includes as well new technologies. A retrospective view over the last decades shows that the IT strategies chosen by companies seem to have followed three general strategic tendencies: In the 1980s and 1990s, the main priority for companies was to integrate business processes. Achieving this integration by implementing integrated systems was hugely beneficial. As globalization increased, the attention shifted to efficiency improvements. The decentralized structures were suddenly perceived as “road blockers” and as obstructing control and reporting processes and as limiting transparency. For IT, this meant developing consolidation strategies to achieve global harmonization of processes, data, and systems (Davidow, & Malone, 1992). From a technological perspective, the web offered an excellent base for the global harmonization projects, but they turned out to be complex, time-consuming and they cost a lot of money. In around 2005, a new paradigm emerged which promised to be the savior in the dilemma of global standardization versus dynamic adaptation: “service-oriented architecture” (SOA). Nowadays, while service-oriented architecture stays a relevant architecture component, a discernible trend away from centralized standards, and back to peripheral control, is emerging. As a consequence of the changed requirements in regards to the IT business-support the business usage

Table 3. New steering and governance model

Element	Definition
Portfolio Management	The fit of the IT architecture as well as the mid- to long-term project roadmap is if strategic priority. The collaboration of IT with the Top Management across all business units is essential to guarantee the synchronicity with the business targets. Additionally the implementation and tracking of key KPIs to measure value creation and implement corrective actions is essential.
Planning Process	The planning process needs to be set up efficiently and lean meaning that the entire planning shall be on a contingency level rather than on a detailed level. This contingency planning will allow a controlling on a flexibility level that allows an adjustment where needed. Furthermore it supports to shift budgets from 'keeping the lights on' to innovation and value creation
Steering and Reporting	To give the increasing importance of the IT topics the appropriate standing within the company it is necessary to establish a new and dedicated decision-making body to address the new topics of digitalization. This newly established committee needs to balance the digitalization topics in the light of executive finance and IT.
Digitalization as core element of the IT project portfolio	The definition of a company-wide valuation of projects in regards to their digitalization relevance to provide a framework to consequently adjust all activities to this core targets.
Master Data Management a score competency	Data will be more tightly infused into processes. Therefore it is essential to position master data management as a core IT competency. Even though master data management is not a digitalization asset at first sight the process responsibility on an end-to-end perspective deserves a high master data quality.
Profit Center structure for IT controlling	Due to the increase responsibility we believe it is essential to structure the entire cost control as a profit center structure to allow an increase transparency on costs but as well to provide a unified and appropriate accounting for IT services.

and the capabilities driven by newly developed IT tools and accelerators the IT infrastructure needs to be enhance with cloud-based and mobile components. Cloud computing became the new way of delivering and charging for services and functionality. For internal IT this means to give away some control over service areas, as it will no longer be provided through the IT department. For the business side cloud computing offers a higher flexibility, as services can be ordered and delivered on shorter notice and paid as variable costs. Therefore, we believe the following strategic guidance, outlined in Table 4, shall be seen as a support to address these required adaptations.

IT Service and Product Portfolio

An adequate IT service level and product portfolio is essential to cope with the expected business requirements. This operational excellence means

bringing business processes to a standardized and productized level, avoiding interfaces or integration gaps wherever possible, and having full transparency regarding key service performance. It deserves as well that IT itself provides an excellence in regards real-time insight into important company data, stability, digital security, and agility.

IT Sourcing Strategy

The sourcing strategy adds a significant part towards the overall productivity of the IT and thus of the company. In terms of the effect of digitalization, the feature to highlight is that IT has continued to improve efficiency for a long period of time and digitalization is now making inroads towards improving efficiency in services as well. Looking ahead, we expect that much of the future productivity resides in improved services.

Table 4. Adapted IT infrastructure

Infrastructure Element	Adaptation
Hybrid IT infrastructure with an increased connectivity	The hybrid IT infrastructure shall consist of “On-premise” and cloud components. “On-premise” refers mainly to the classical operations with buying or renting the IT component together with the IT operations in own responsibility. By contrast, cloud means a purely user-oriented consumption of IT functionalities and a corresponding pricing. Depending on the level of consumed services we need to differentiate between Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS) and Software-as-a-Service (SaaS). Overall, we can state that cloud will offer a higher degree of cost flexibility, the safe-guarding of a consistent and timely modernization of the used IT components and a better realization of the “Economics of Scale” due to a bundling of service providers within the community of all users of a service. In case a service shall be used exclusively for a certain community there is the ability of a cloud-service usage in a private cloud. In this sense, public cloud components shall be advised to be considered in case the IT support is not seen as a competitive differentiator and where the data security is sufficiently provided within the cloud environment. This is where it is essential to balance the On-premise combination with the Public Cloud components in a hybrid IT environment. In a world of hybrid deployment models, high connectivity (Business Network and omni-channel), Big Data and Internet of Things, implementation and operations have to be in-sync more than ever to allow velocity, innovation, stronger technology adoption and much lower operational risk.
Build-in international IT security levels	In general the IT infrastructure needs to meet EU and international security levels to address the increased requirements towards data protection and data security. This can be achieved by appropriate firewalls, authentication methodologies and certifications (e.g. ISO 27001, IDW 951, etc.) This is where it is necessary to keep a special attention on service suppliers that are outside the EU data protection guideline
Mobile IT Technology	The mobile usage of IT technology is mandatory and becomes the new standard. With the increasing prevalence of web technologies (Curbera, et al., 2002), employees now have access to company data via the Internet (“anywhere”), which facilitates e.g. work from home or at a customer’s site (Aungst, & Wilson, 2005; Sinisalo, Salo, Karjaluo, & Leppäniemi, 2006). Smart mobile devices offer further opportunities to optimize many business processes (“any service, anywhere, anyhow”). Mobile connectivity is undoubtedly becoming an important building block in any digital eEnterprise Therefore all applications need to be available on any end-user devices – especially on the mobile devices (e.g. tablets or smartphones) – within an easy-to-go way. This needs correspond to the availability of the wireless infrastructure and the contracts with these external providers (e.g. global flat rate for calls and data).
Globally standardized catalog for end-user devices	With the provisioning of a globally standardized set of IT end-user devices that is mandatory for all users a comprehensive but manageable portfolio will be provided to support the mobile strategy at the most cost-effective way.
Centralized user authentication	A centralized authentication platform for every IT infrastructure and application used within a company with a single-sign-on is core to support the convenience of the users together with the security requirements.

Following, we will following present an overall view of where the IT sourcing strategy is headed.

- Increase the percentage of external sourcing to accomplish a better coverage of specific IT demands.
- Professionalize the multi-vendor strategy in regards to globally binding service levels, escalation processes, reporting structures and warranty rules.
- Global validity of software contracts to realize cost effects with the bundling of demands as well as to better manage compliance regulations based on software license contracts.
- Extended use of external service providers to flexible and better capture capacity peaks and thus to avoid delivery and project delays.
- Outsourcing keeps a core element of the IT operations to benefit from the lower cost level, cost flexibility and to meet the issues of resource shortages of IT specialists (Dos Santos, 2003).

Table 5. New IT service and product portfolio

Topic	Definition
Global Service and Support	Establish a professional and global acting application lifecycle management (ALM) with dedicated service levels per application and coverage of the “follow the sun” principle
Trusted advisor	Establish a professional and comprehensive consulting service based on an in-depth knowledge of IT methodologies and business know how
Streamlined IT product catalogue	Develop an IT product catalogue focused on end-user oriented topics, necessities and applications
End-user devices	Provide a broad catalogue of end-user devices oriented towards the technological developments, application and business needs and tailored to the most appropriate usage
Flexible IT usage	Turn away from the standard cost charging of IT towards a used-on demand concept of IT services. The shall be a flexibility to book and decline the usage of IT services and a transparency on the charges for the consumed services
Global service levels	Deliver the IT service deliveries towards a unified, company-wide and standardized service level agreements (SLAs). A monitoring of the various categories and deliveries supported with customer satisfaction surveys shall be performed on a regular basis
Process Re-Engineering	The analysis and fact-based re-design of business processes will emerge towards a core IT competency. Even though this is not a brand-new topics we still waste huge amounts of time and money because we still rely on expensive manual and slow paper-based processes. I these areas digitization is the transformation of information into digital formats and ultimately the digitization of business processes.

IT Applications

The business requirements towards the IT applications will increase due to the digitalization as they emerge from a process support to a business enabler. Having this role emergence in mind we can draw 5 key principles for the IT applications:

- **Principle 1:** Overall target is a well-balanced hybrid IT application landscape.
- **Principle 2:** User friendliness becomes a core target of the entire application development and application provisioning with an appropriate quality level delivered out of the new usability lab.
- **Principle 3:** Increase use of Apps to support business processing is essential to ease the handling of functionalities and create user-centric rather than system-centric applications, processes and workflows – in parallel the expert mode shall be maintained as well. This user centricity includes a strong focus on user value and the best possible interaction with the business, and

intuitive “from-where-you-are” access to important information or processes.

- **Principle 4:** Extend the use of public cloud applications to allow a networked collaboration within the company to enable a paperless process and a truly digital workplace.
- **Principle 5:** No release of new applications without user acceptance tests as a mandatory asset of every IT implementation to systematically increase the applications quality.

CONCLUSION

Over the last decades, the central points of view of IT strategy has been that it is a supportive strategy that must necessarily been derived from the company’s strategy and that focuses on the support of processes with a high degree of process standardization leading to cost effectiveness. Now business becomes digital with increased IT elements among products, services and processes. This is as well true for companies in the old

economy where the digital technology transforms the entire business of a company. Consequently we believe that it is necessary to re-shape the IT strategy – still aligned with the company strategy but no more a niche player. Our understanding is, that the IT strategy need to reflect the business strategy and vice versa leading to a digitalized IT strategy that becomes a driver of the business and its value. This will affect the traditional domains of a company such as marketing, sales and production as it is important to overcome the disjointed views of IT and digitalization to gain the full value and benefit of this new era. As outlined, the digitalization will have an impact on the IT organization, the structures and processes. It will change the patterns of collaboration as well as the necessary knowledge the IT staff need to have. Finally we believe that IT needs to be anchored in the core decision making body of a company to drive the company strategy with key impulses from the digital era.

FURTHER RESEARCH DIRECTIONS

Even though we may have the feeling that we have dealt with digitalization for ages it is a rather new phenomenon to the business and we are just at the starting point of this transition journey. Therefore, our chapter outlines the levels, dimensions and elements that affects the corporate IT but one important questions for the future is how to orchestrate the new level of agility and ever changing requirements for an IT organization. We believe, that it deserves an in-depth analysis how these “digital-dynamics” affect the IT organization, how these novelties interfere with the existing organization and how to best balance novelty, agility and tradition to create the best out of the digitalization.

REFERENCES

- Agarwal, R., & Lucas, H. C. Jr. (2005). The information systems identity crisis: Focusing on high-visibility and high-impact research. *Management Information Systems Quarterly*, 381–398.
- Amit, R., & Schoemaker, P. J. (1993). Strategic assets and organizational rent. *Strategic Management Journal*, 14(1), 33–46. doi:10.1002/smj.4250140105
- Banker, R. D., Bardhan, I. R., Chang, H., & Lin, S. (2006). Plant information systems, manufacturing capabilities, and plant performance. *Management Information Systems Quarterly*, 315–337.
- Barua, A., & Mukhopadhyay, T. (2000). Information technology and business performance: Past, present, and future. *Framing the domains of IT management: Projecting the future through the past*, 65–84.
- Baskerville, R. L., & Land, F. (2004). Socially self-destructing systems. *The social study of information and communications technology: Innovation, actors and contexts*, 263–285.
- Berger, P. G., & Ofek, E. (1995). Diversifications effect on firm value. *Journal of Financial Economics*, 37(1), 39–65. doi:10.1016/0304-405X(94)00798-6
- Bharadwaj, A., El Sawy, O. A., Pavlou, P. A., & Venkatraman, N. V. (2013). Digital business strategy: Toward a next generation of insights. *Management Information Systems Quarterly*, 37(2), 471–482.
- Brynjolfsson, E., & Hitt, L. M. (2000). Beyond computation: Information technology, organizational transformation and business performance. *The Journal of Economic Perspectives*, 14(4), 23–48. doi:10.1257/jep.14.4.23
- Buschek, A. (2015). *Friends Don't Let Friends Build Data Centers Anymore*. November. <https://www.digitaltransformationblog.com/friends-dont-let-friends-build-data-centers-anymore>

- Carr, N. G. (2003). IT doesn't matter. *EDUCAUSE Review*, 38, 24–38. PMID:12747161
- Chan, Y. E., & Reich, B. H. (2007). IT alignment: What have we learned? *Journal of Information Technology*, 22(4), 297–315. doi:10.1057/palgrave.jit.2000109
- Collins, J. C., & Porras, J. I. (1996). *Built to Last. Successful Habits of Visionary Companies*.
- Conner, K. R., & Prahalad, C. K. (1996). A resource-based theory of the firm: Knowledge versus opportunism. *Organization Science*, 7(5), 477–501. doi:10.1287/orsc.7.5.477
- DAdderio, L. (2001). Crafting the virtual prototype: How firms integrate knowledge and capabilities across organisational boundaries. *Research Policy*, 30(9), 1409–1424. doi:10.1016/S0048-7333(01)00159-7
- Davenport, T. H. (2006). Competing on analytics. *Harvard Business Review*, 84(1), 98.
- Davidow, W. H., Malone, M.A. (1992). *The virtual corporation: Structuring and revitalizing the corporation for the 21st century*. Academic Press.
- Devaraj, S., & Kohli, R. (2001). *The IT payoff: measuring the business value of information technology investments*. Financial Times/Prentice Hall.
- Dos Santos, B. L. (2003). Information technology investments: Characteristics, choices, market risk and value. *Information Systems Frontiers*, 5(3), 289–301. doi:10.1023/A:1025653428097
- Ettlie, J. E., & Pavlou, P. A. (2006). Technology-based new product development partnerships. *Decision Sciences*, 37(2), 117–147. doi:10.1111/j.1540-5915.2006.00119.x
- Evans, D. S., Hagi, A., & Schmalensee, R. (2008). *Invisible engines: how software platforms drive innovation and transform industries*. MIT press.
- Gartner. (2016). *Building the Digital Platform: Insights From the 2016 Gartner CIO Agenda Report*. Stamford, CT: Gartner Corporate Headquarters.
- Hames, R. D. (1994). *The management myth: Exploring the essence of future organisations*. Business & Professional Publishing.
- Hirschheim, R., & Sabherwal, R. (2001). Detours in the path toward strategic information systems alignment. *California Management Review*, 44(1), 87–108. doi:10.2307/41166112
- Hussin, H., King, M., & Cragg, P. (2002). IT alignment in small firms. *European Journal of Information Systems*, 11(2), 108–127. doi:10.1057/palgrave/ejis/3000422
- Jarach, D. (2002). The digitalisation of market relationships in the airline business: The impact and prospects of e-business. *Journal of Air Transport Management*, 8(2), 115–120. doi:10.1016/S0969-6997(01)00039-4
- Kane, G. C. (2015). Enterprise Social Media: Current Capabilities and Future Possibilities. *MIS Quarterly Executive*, 14(1), 1–16.
- Klein, R., & Rai, A. (2009). Interfirm strategic information flows in logistics supply chain relationships. *Management Information Systems Quarterly*, 735–762.
- Kohli, R., & Grover, V. (2008). Business value of IT: An essay on expanding research directions to keep up with the times. *Journal of the Association for Information Systems*, 9(1), 23.
- Lang, L. H., & Stulz, R. M. (1993). *Tobin's q, corporate diversification and firm performance (No. w4376)*. National Bureau of Economic Research. doi:10.3386/w4376
- Luftman, J., & Brier, T. (1999). Achieving and sustaining business-IT alignment. *California Management Review*, 42(1), 109–122. doi:10.2307/41166021

- Luftman, J., Kempaiah, R., & Nash, E. (2006). Key issues for IT executives 2005. *MIS Quarterly Executive*, 5(2).
- Melville, N., Kraemer, K., & Gurbaxani, V. (2004). Review: Information technology and organizational performance: An integrative model of IT business value. *Management Information Systems Quarterly*, 28(2), 283–322.
- Nordstöm, C. (2015). *Digitalization - a powerful source for change*. Retrieved from https://www.sics.se/sites/default/files/pub/web_sics_activities_2015-2016.pdf
- Pavlou, P. A., & El Sawy, O. A. (2010). The third hand: IT-enabled competitive advantage in turbulence through improvisational capabilities. *Information Systems Research*, 21(3), 443–471. doi:10.1287/isre.1100.0280
- Rai, A., Patnayakuni, R., & Seth, N. (2006). Firm performance impacts of digitally enabled supply chain integration capabilities. *Management Information Systems Quarterly*, 225–246.
- Rai, A., Pavlou, P. A., Im, G., & Du, S. (2012). Interfirm IT capability profiles and communications for cocreating relational value: Evidence from the logistics industry. *Management Information Systems Quarterly*, 36(1), 233–262.
- Ray, G., Muhanna, W. A., & Barney, J. B. (2005). Information technology and the performance of the customer service process: A resource-based analysis. *Management Information Systems Quarterly*, 625–652.
- Reich, B. H., & Benbasat, I. (2000). Factors that influence the social dimension of alignment between business and information technology objectives. *Management Information Systems Quarterly*, 24(1), 81–113. doi:10.2307/3250980
- Sabherwal, R., & Chan, Y. E. (2001). Alignment between business and IS strategies: A study of prospectors, analyzers, and defenders. *Information Systems Research*, 12(1), 11–33. doi:10.1287/isre.12.1.11.9714
- Sabherwal, R., & Kirs, P. (1994). The alignment between organizational critical success factors and information technology capability in academic institutions. *Decision Sciences*, 25(2), 301–330. doi:10.1111/j.1540-5915.1994.tb01844.x
- Salo, J. (2006). Business relationship digitization: What do we need to know before embarking on such activities? *Journal of Electronic Commerce in Organizations*, 4(4), 75–93. doi:10.4018/jeco.2006100104
- Sambamurthy, V., Bharadwaj, A., & Grover, V. (2003). Shaping agility through digital options: Reconceptualizing the role of information technology in contemporary firms. *Management Information Systems Quarterly*, 237–263.
- Santhanam, R., & Hartono, E. (2003). Issues in linking information technology capability to firm performance. *Management Information Systems Quarterly*, 125–153.
- Saraf, N., Langdon, C. S., & Gosain, S. (2007). IS application capabilities and relational value in interfirm partnerships. *Information Systems Research*, 18(3), 320–339. doi:10.1287/isre.1070.0133
- Sledgianowski, D., Luftman, J. N., & Reilly, R. R. (2006). Development and validation of an instrument to measure maturity of IT business strategic alignment mechanisms. *Innovative Technologies for Information Resources Management*, 19(3), 18–33. doi:10.4018/irmj.2006070102
- Tanriverdi, H., & Venkatraman, N. (2005). Knowledge relatedness and the performance of multi-business firms. *Strategic Management Journal*, 26(2), 97–119. doi:10.1002/smj.435
- Wade, M., & Hulland, J. (2004). Review: The resource-based view and information systems research: Review, extension, and suggestions for future research. *Management Information Systems Quarterly*, 28(1), 107–142.

Wernerfelt, B. (1995). The resource-based view of the firm: Ten years after. *Strategic Management Journal*, 16(3), 171–174. doi:10.1002/smj.4250160303

Wheeler, B. C. (2002). NEBIC: A dynamic capabilities theory for assessing net-enablement. *Information Systems Research*, 13(2), 125–146. doi:10.1287/isre.13.2.125.89

ADDITIONAL READING

Aungst, S. G., & Wilson, D. T. (2005). A primer for navigating the shoals of applying wireless technology to marketing problems. *Journal of Business and Industrial Marketing*, 20(2), 59–69. doi:10.1108/08858620510583650

Curbera, F., Duftler, M., Khalaf, R., Nagy, W., Mukhi, N., & Weerawarana, S. (2002). Unraveling the Web services web: An introduction to SOAP, WSDL, and UDDI. *IEEE Internet Computing*, 6(2), 86–93. doi:10.1109/4236.991449

Sinisalo, J., Salo, J., Karjaluo, H., & Leppaniemi, M. (2006, January). Managing customer relationships through mobile medium-underlying issues and opportunities. In *Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06)* (Vol. 6, pp. 112b–112b). IEEE. doi:10.1109/HICSS.2006.262

KEY TERMS AND DEFINITIONS

ALM: Application lifecycle management (ALM) is the supervision of a software application from its initial planning through retirement. It also refers to how changes to an application are documented and tracked.

Big Data: Big data is high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation.

Cloud: Describes the locations where services are acquired and data is stored, can take various forms.

Cloud Computing: Cloud computing is a general term for the delivery of hosted services over the Internet. Cloud computing enables companies to consume compute resources as a utility -- just like electricity -- rather than having to build and maintain computing infrastructures in-house.

Deployment Models: A cloud deployment model represents a specific type of cloud environment, primarily distinguished by ownership, size, and access. There are four common cloud deployment models: Public Clouds; Community Clouds; Private Clouds; and Hybrid Clouds.

Digitalization: Integration of digital technologies into everyday life by the digitization of everything that can be digitized.

Hybrid IT Infrastructure: Approach to enterprise computing in which an organization provides and manages some information technology (IT) resources in-house but uses cloud-based services for others. A hybrid approach allows an enterprise to maintain a centralized approach to IT governance, while dealing with cloud computing.

Economics of Scale: Economies of scale is the cost advantage that arises with increased output of a product. Economies of scale arise because of the inverse relationship between the quantity produced and per-unit fixed costs; i.e. the greater the quantity of a good produced, the lower the per-unit fixed cost because these costs are spread out over a larger number of goods. Economies of scale may also reduce variable costs per unit because of operational efficiencies and synergies. Economies of scale can be classified into two main types: Internal – arising from within the company; and External – arising from extraneous factors such as industry size.

Industry 4.0: Industry 4.0 or the fourth industrial revolution, is the current trend of automation and data exchange in manufacturing technologies. It includes cyber-physical systems, the Internet of things and cloud computing.

Infrastructure-as-a-Service (IaaS): Infrastructure as a Service (IaaS) is a form of cloud computing that provides virtualized computing resources over the Internet.

Internet of Things: A network comprised of physical objects capable of gathering and sharing electronic information. The Internet of Things includes a wide variety of “smart” devices, from industrial machines that transmit data about the production process to sensors that track information about the human body. Often, these devices use Internet Protocol (IP), the same protocol that identifies computers over the world wide web and allows them to communicate with one another.

Master Data Management: Master data management (MDM) is a comprehensive method of enabling an enterprise to link all of its critical data to one file, called a master file that provides a common point of reference.

Omni-Channel: Omni-channel is a multichannel approach to sales that seeks to provide the customer with a seamless shopping experience whether the customer is shopping online from a desktop or mobile device, by telephone or in a bricks and mortar store.

On Premise: On-premises is a type of software delivery model that is installed and operated from a customer’s in-house server and computing infrastructure. It utilizes an organization’s native computing resources and requires only a licensed or purchased copy of software from an independent software vendor.

Platform-as-a-Service (PaaS): Platform as a service (PaaS) is a cloud computing model that delivers applications over the Internet.

Predictive Analytics: Predictive analytics is the branch of data mining concerned with the prediction of future probabilities and trends.

The central element of predictive analytics is the predictor, a variable that can be measured for an individual or other entity to predict future behavior.

Process Re-Engineering: Documenting, analyzing, and comparing a process to benchmarks such as best-in-class practices, implementing the required changes, or installing a different process.

Public Cloud: A public cloud is one based on the standard cloud computing model, in which a service provider makes resources, such as applications and storage, available to the general public over the Internet.

Shared Service Center: A shared services center – a center for shared services in an organization – is the entity responsible for the execution and the handling of specific operational tasks, such as accounting, human resources, payroll, IT, legal, compliance, purchasing, security. The shared services center is often a spin-off of the corporate services to separate all operational type of tasks from the corporate headquarters, which has to focus on a leadership and corporate governance type of role.

Service Level Agreements (SLAs): Contract between a service provider and a customer, it details the nature, quality, and scope of the service to be provided.

Service-Oriented Architecture (SOA): Service-oriented architecture (SOA) is an approach used to create an architecture based upon the use of services. It carries out function, such as producing data, validating a customer, or providing simple analytical services.

Software-as-a-Service (SaaS): Software as a service (SaaS) is a software distribution model in which a third-party provider hosts applications and makes them available to customers over the Internet.

The Main Concepts Behind the Dematerialization of Business Processes

Liliana Ávila

University of Aveiro, Portugal

Leonor Teixeira

University of Aveiro, Portugal

INTRODUCTION

Currently, Information and Communication Technologies (ICTs) represent an important tool for all the organizations that intend to improve their performance through a better management of their information processes. The dematerialization of business processes emerges as one solution that promotes a clear, structured and transparent process, especially when the problems are associated with the existence of documentation in a paper format. It comprises a range of practices, such as electronic document management, workflow management and reengineering of business processes, aiming the elimination of inefficiencies caused by the circulation of information in paper format, as well as the elimination of resources associated with the maintenance and retrieval of information in that format. The increase in productivity that often is expected to result with dematerialization of business processes comes from the (i) decrease in the amount of time to perform the tasks, (ii) effectiveness of human resources allocated to those tasks, (iii) quick access to information, and also (iv) improvement of communication flow between different services.

The dematerialization of business processes is not a new concept. It is already implemented in different areas, particularly in the public sector. However, due the lack of a theoretical framework, each organization makes its own interpretation of the concept and adopts its own methodologies and technologies. In this context, it is very important

to clarify the concept of dematerialization of business processes in order to achieve a common understanding.

This chapter aims to contribute to this knowledge domain, providing an overview of the main concepts behind the dematerialization of business processes, giving special attention to the vital role of ICTs, through the analysis of some tools that can be used to support the dematerialization of business processes. In this way, the main goal is to help organizations understanding the benefits and commitments that can arise from the modernization of their processes.

INTRODUCING THE DEMATERIALIZATION OF BUSINESS PROCESSES

The dematerialization of business processes is a relatively new concept in the domain of scientific research. Nevertheless, it is possible to find evidences of its implementation, namely in the public sector through the topic known as “e-governance”.

Usually, the concept of dematerialization can be defined as the process of scanning paper documents to obtain electronic documents, in order to eliminate the circulation of paper. According to Siatiras (2004), currently the information is probably one of the most important assets in the organizations, and therefore its efficient management is very important.

However, the dematerialization of business processes is not limited to scanning and managing documents, it also allows the organization to manage its workflows. In this sense, the dematerialization of business processes refers to the optimization of information circuits through technological applications, where the information is communicated and stored in digital format in order to expedite and facilitate the monitoring of processes, always ensuring the legal value of those electronic documents (Ávila, Teixeira, & Almeida, in press). The main gains associated with the dematerialization of business processes are the facility in search documents (Ashby, 2011; Schnitzer, 2010; Want, 2009), and the concurrent access by multiple users to the same document regardless their geographic location (Ashby, 2011; Downing, 2006). This is even more important for organizations that are active in different locations and need to exchange information in a regular basis.

The dematerialization of business processes can also be a mean to improve customer service, since the information is readily available on the computer without the need for search in paper files, allowing an immediate response to customer requests and a greater flexibility to respond to a variety of needs (Burns, 2009; Downing, 2006; Medina & Fenner, 2005). Other consequences are the savings in time and costs (Johnston & Bowen, 2005; Schnitzer, 2010; Want, 2009) and the improvement in the quality of processes and their outcomes (Ashby, 2011; Johnston & Bowen, 2005). At the human resources level, it can encourage greater communication and collaboration between people, creating a more harmonious work environment (Ashby, 2011; Siatiras, 2004; Want, 2009).

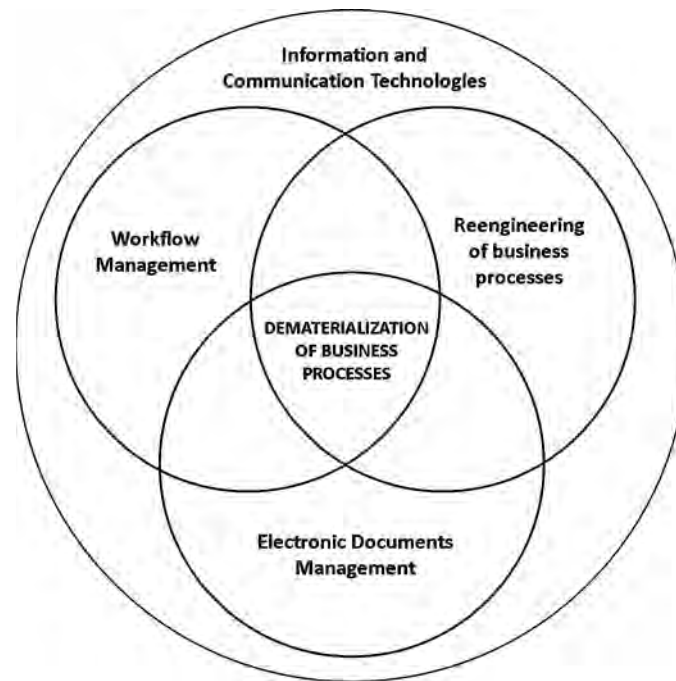
Despite the benefits pointed to a project of dematerialization of business processes, it should be noted that this type of project requires some efforts to understand and manage the factors (difficulties), arising from the implementation of

this type of projects. One of the most prevalent factors is the time of a transition to a dematerialized environment. The implementation of a dematerialization project must be preceded by a comprehensive and detailed planning process. The transition should be progressive, because it is a slow and gradual process, along which there are a large number of small steps that should be performed (Ashby, 2011; Burns, 2009). In addition, the large investment required in the initial phase (software, scanners and, eventually, new computers) and the resistance of some employees to change, should be considered. The existence of some instability feelings among workers due to the possible feeling of greater control by the managers (Downing, 2006; Medina & Fenner, 2005) may also appear as a result of the dematerialization of business processes, often seen as an obstacle to its acceptance.

Organizations must understand that the dematerialization of their business processes may involve significant organizational changes, so its implementation should not be motivated by the simple desire for eliminate the use of paper and be environmentally friendly (Ashby, 2011). Thus, the dematerialization of business processes is a complex concept emerging from the interception of other related concepts that have been more explored in the literature such as the electronic documents management, the workflow management and the reengineering of business processes. The use of ICTs is common to all these domains, playing a determinant role in the implementation of the dematerialization of business processes, justifying the greater emphasis on this topic (Figure 1).

In the following section, we will present the concepts in order to discuss in which extent they are related with the main topic that is been analysed in this chapter, i.e. with dematerialization of business processes.

Figure 1. Main concepts behind the dematerialization of business processes



THE MAIN CONCEPTS BEHIND THE DEMATERIALIZATION OF BUSINESS PROCESSES

Electronic Documents Management: Welcome to the Digital Era

The concept of electronic document management describes the tools and processes associated with managing documents and records in digital format, allowing the decrease of some risks, such as the time spending searching the required information and the use of some outdated or incomplete data (Siatiras, 2004). Thus, an electronic document management system represents the software used to manage unstructured information like emails, images, spreadsheets or text in a controlled and consistent manner.

In the scientific literature, some authors list several features related to this type of systems. These authors highlight the fact that there is a central repository for all the electronic documents, the ability to search through this repository and the control of multiple versions of a single document.

The implementation of such systems could also be a great opportunity to define and establish a common vocabulary in the organization, a consistent terminology to describe the information, instead of each person having its own way to attribute names to documents (Al Qady & Kandil, 2013).

In this “digital” context, it is very important the existence of good security protocols to protect individual rights to access and edit documents as well as define retention and automatic deletion policies, standardized classification structures and audit trails. The use of digital documents implies the existence of user access control methods for the authentication as a substitute of the traditional signature in order to assure the system integrity, authenticity, non-repudiation and notarization (Lee, 2013). Any document can be accessed remotely from any computer through a secure virtual private network connection (Wisniewski, Pummer, & Krenzelo, 2010). Other important factors that can be found in the literature and can contribute to the success of this type of project are the management support (Alzubi, 2015; Cho, 2008; Haug, 2012), the concerns with user acceptance (Cho, 2008;

Hung, Tang, Chang, & Ke, 2009), which could be mitigated for example through the definition of reward systems and training activities, as well as the external support from a consultant and the selection of software (Hung et al., 2009).

It is noted that the empirical work that have been conducted in this field focus mainly the application of electronic documents management systems in three different types of organizations: governmental entities (Afonso, Schwarz, Roldán, & Franco, 2015; Hung et al., 2009), healthcare organizations (Harmsen, 2007; Schmidt, Simmons, Grimm, Middlebrooks, & Changchien, 2006; Wisniewski et al., 2010) and companies acting in the construction sector (Al Qady & Kandil, 2013; Hjelt & Björk, 2007).

Workflow Management Systems: Performing Tasks at the Right Moment

Mentzas, Halaris, & Kavadias (2001) and Hollingsworth (1995) define workflow as a set of tasks organized to perform a business process. In its turn, a task can be performed by one or more software systems and carried out by one or more employees. Moreover, a workflow defines the order or the conditions on which a task should be executed, its synchronization and information flow. Therefore, the workflow technology enables an organization to automate their business processes more efficiently. A workflow technology distributes work items by users, indicates how the task should be performed, tracks the progress of the work item throughout the process, and generates statistics on how different steps are carried out. According to Aalst & Hee (2009), a workflow management system is a generic software package that supports the management of business processes, taking into account its information logistics. Such systems ensure that the right information is given to the right person, at the appropriate time.

Generically, workflow management systems can be characterized as supporting three levels – the business process definition, its monitoring and

the interaction with users or applications and Information Technology (IT) tools (Hollingsworth, 1995). Figure 2 illustrates how these three levels are related with each other.

According to the author, the first level relates to the build time functions, which have as main output the computerized definition of the business process. The process is translated from the real world into a formal definition, readable by computer, through the use of one or more analytical techniques, modelling and definition tools. The process design and definition usually includes a set of activities, related tasks performed by computer or people, and rules that guide the progression of the process throughout those tasks. The description of the process may be expressed in a textual or graphical language based on a defined notation. The other two levels are related to the run time. One of them consists of the process control activities. The process definition is processed by the software that schedules the several steps to perform the activities constituting that process, invoking the human resources and applications involved in each activity. The third level represents the interaction with users and IT applications in order to perform tasks.

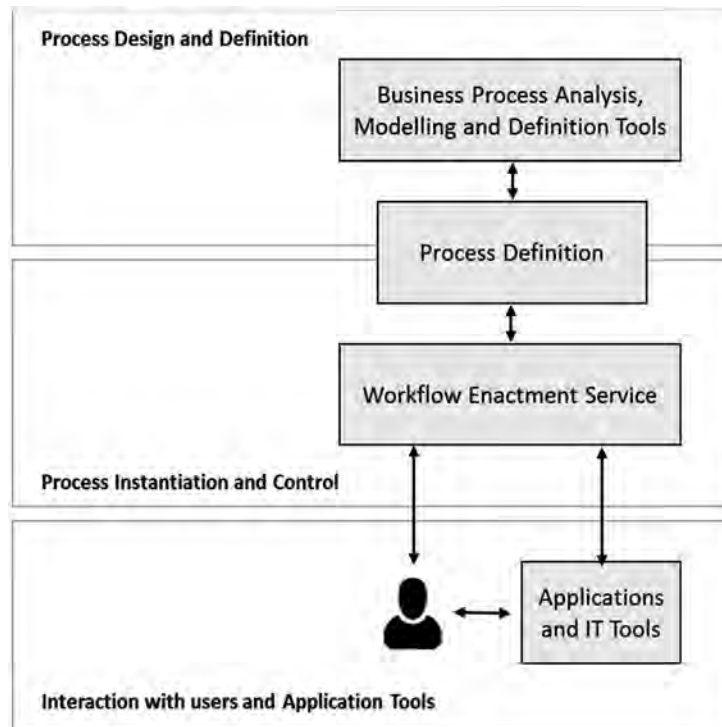
The description provided by Hollingsworth (1995) helps to understand the benefits arising from the integration of the electronic documents management with the workflow management, representing the benefits from the adoption of the dematerialization of business processes. However, that approach can be more fruitful when combined with the reengineering of business processes. The next section presents reengineering as a tool that can be used to achieve better results.

Reengineering as a Way to Achieve Better Results

Hammer & Champy (1993), experts in the field of reengineering, define this concept as the rethinking and redesigning of business processes, in order to achieve significant improvements in key performance indicators (KPI), such as costs, quality,

Figure 2. Characteristics of workflow systems

Source: Adapted from Hollingsworth, 1995



service and speed. Achieving this goal usually requires the implementation of deep changes and abandoning all pre-established procedures in the organization (Attaran, 2004; Aversano, Canfora, De Lucia, & Gallucci, 2002; Weerakkody, Janssen, & Dwivedi, 2011). Weerakkody et al. (2011) refer that the first projects developed in this domain had a large failure rate and that the reengineering of business processes is often criticized by its radical approach. Nevertheless, these authors concluded that only the accumulation of small improvements will not result in significant changes in the organization, so if the organization really wants that to happen, it is essential to start with a more radical step, such as the business process reengineering suggests.

Limam Mansar & Reijers (2007) present a framework for the reengineering of business processes implementation composed by six main components: customers, products, business process, organization structure and population, technology and information. The proposed framework shows

how BPR implementation may involve a deep understanding of how the organization works in order to achieve significant changes and benefits. Information is identified by these authors as one of the most important components because it refers to an area where very large improvements can be made using business process redesign (BPR), either through using technology or not. It was considered the second most important component by practitioners, only behind customers.

The reengineering of business processes is a concept closely related to the workflow management systems (Aversano et al., 2002; Hollingsworth, 1995). These concepts are often mistaken, although they have different definitions. The term of workflow management system presented previously is in general associated with the technology used for automating business processes. Thus, a company can automate their processes without conducting its reengineering, and this can be done without the use of a workflow management system. However, it is considered that a process-oriented

approach is essential to implement a workflow management system. When such systems are implemented to support existing practices will certainly result in limited improvements. The adoption of a workflow management system can benefit from a reengineering effort and the opposite is also true (Aalst & Hee, 2009; Hollingsworth, 1995).

Attaran (2004) explores the relationship between information technology and business process reengineering. The author identifies several companies that have implemented the reengineering of business processes to a wide variety of processes such as product development, procurement or sales. The main barriers pointed by the author to its effective implementation are the misunderstanding of the concept, the lack of proper strategy, the management and Information Systems failure to change and the failing to recognize the importance of people.

The three concepts presented above contribute to the understanding of the dematerialization of business processes. However, ICTs play a vital role. Without them does not make any sense to talk about dematerialization of business processes. The next section presents the importance of these tools and analyses what are the alternatives and main features of the existing software to support such projects.

THE VITAL ROLE OF INFORMATION AND COMMUNICATION TECHNOLOGIES IN THE DEMATERIALIZATION OF BUSINESS PROCESSES

As mentioned previously, when we talk about electronic documents management, workflow management and reengineering of business processes, we are talking about processes supported by ICTs, and this is why the choice of software is frequently pointed as a critical success factor of such projects. ICTs play the main role and it is very important to ensure that the right choice

is made as a guarantee that its implementation benefits the organization as a whole, allowing improvements in terms of its performance.

Internal vs. External, Proprietary vs. Open-Source

Regarding the selection of the most appropriate software, each enterprise should evaluate its internal competencies before making a decision. If it is a big enterprise and has an IT department it can use its own resources and develop internally the ICT tool for support the dematerialization of business processes.

However, small and medium enterprises, that may not have the same resources, usually opt by other solutions, as outsourcing all or part of the development process or hiring external consultants to support their employees in the development of an ICT tool that fits the enterprise requirements. Nowadays, it is possible to find in the software market many technologies able to be adapted to different contexts requiring less efforts in terms of financial and other resources when compared with the development of a new software. In case the enterprise decides to select an alternative from those already available in the market, it can choose between proprietary and open-source software.

In the last years, there were great advances in the field of open-source software and it became an alternative increasingly sought, presenting many benefits when compared with proprietary software, namely in terms of costs, quality and customization (Chester, 2006). Although open-source software presents clear benefits, it is important to think about all the variables because in some cases savings may be illusory. According to a survey conducted by Accenture PLC in August 2010 with directors from 300 enterprises in United States and United Kingdom, 50% of respondents said to be committed to the use of open-source software, while 28% assume it as an experimentation phase and 69% expressed interest to increase the investment at this level (Collett, 2010). In the same study, when surveyed about the main benefits

of open-source, the most cited factors were quality (76%), reliability (71%), security (70%) and then the decreasing of total costs (50%).

Nagy, Yassin & Bhattacharjee (2010) identified some barriers for the adoption of open-source software and suggested some actions that can be taken in order to mitigate each one of them. They identified barriers related to knowledge, integration with different technologies, lack of code standardization, previous investment in proprietary software and lack of software maturity. To prevent these barriers, companies may, for example, provide training to their employees, define standards, consider using open-source where there are no proprietary software, define maturity models and evaluate case studies in an independent way.

The open-source software has attracting increasingly interest of researchers from Software Engineering and ICTs. Therefore, it seems relevant to explore the applicability of this type of software to the dematerialization of business processes, through the comparison of features provided by

several proprietary and open-source software solutions available in the market.

Is Open-Source A Solution for the Dematerialization of Business Processes?

In the software market is possible to find some proprietary and open-source solutions for the management of business processes which can be considered in the implementation of dematerialization business processes projects. In a previous project (Ávila, Teixeira, & Almeida, 2015) a study was conducted that aims to analyze some existing market solutions in order to identify the main features of these tools and determine if open-source constitutes an advantageous alternative when compared with proprietary software. This comparative study was only based on information accessible via the official website of the respective providers, where are only presented the most important and distinctive features. Table 1 shows the results of this study.

Table 1. Comparative study of software to support the implementation of dematerialization of business processes

Proprietary Software			Features	Open-Source Software				
[1]	[2]	[3]		[4]	[5]	[6]	[7]	[8]
•		•	<i>Design workflows with BPMN</i>	•		•		•
•			<i>Process simulation</i>			•		
•		•	<i>Process optimization</i>	•				•
•		•	<i>Model repository</i>	•		•		
•	•	•	<i>Web forms designer</i>			•		•
•		•	<i>Client-server architecture</i>		•	•	•	
•		•	<i>Software in the cloud</i>	•				•
•			<i>Notifications</i>		•		•	
•		•	<i>Monitoring activities</i>	•				
	•	•	<i>Management users</i>	•		•	•	•
•	•	•	<i>Reports</i>	•		•	•	
	•		<i>Digital signature</i>				•	
•			<i>Collaboration tools</i>		•	•		

Label: [1] Appian; [2] FileDoc; [3] RunMyProcess; [4] Activiti; [5] Alfresco; [6] Bonita Open Solution; [7] iPortalDoc; [8] ProcessMaker

Analyzing Table 1, it is possible to conclude that there are some features which are common to almost all the studied solutions, namely the possibility of manage users and the production of reports. Other features frequently offered are the design of workflows with BPMN (Business Modeling Processes Notation), the design of web forms and client-server architecture. Moreover, some of the presented solutions give more relevance to the features related to workflow management, while others give more importance to documents management, such as FileDoc and iPortalDoc. In general, and as expected, proprietary software presents a wider set of features, while open-source software is more limited. However, taking into account the advantages mentioned previously, the features offered by open-source solutions should not be ignored, namely the Activity and Bonita Open Solution.

Currently, there are companies from several sectors with proven real use cases using open-source solutions to manage their business processes. Some of them are, for example, NASA and Michelin which are clients of Alfresco or even Accenture, Cisco and several French Ministries which adopted the Bonita Open Solution. Open source technologies may also be valuable options for organizations with a small dimension which intend to implement and test the dematerialization of business processes in some of their processes.

FUTURE RESEARCH DIRECTIONS

As already mentioned, the dematerialization of business processes is a relatively unexplored concept in the literature. Therefore it is possible to identify many directions for future research. First, it seems important to find a common definition of what the dematerialization of business processes is, reflecting the vision of academics and practitioners. Academics can contribute producing case studies of organizations that have experienced the dematerialization of business processes and identifying the best practices in this domain, covering

organizations with different dimensions and from different sectors.

Moreover, this approach to administrative modernization seems to be more popular among the public sector and less notorious in the private sector. Defining a more consistent methodology could be very useful to support organizations in general, regardless the sector they belong to.

Finally, it could be relevant in future research explore the best practices to accelerate the implementation of the dematerialization of business processes in a cost-effective way for companies, in order to make possible for any organization take advantage from this approach to manage their information flows. One suggestion is evaluating the potential use of open-source technologies in a cloud computing paradigm. Currently, the popularity of these technological solutions represents a good option for organizations, allowing them to reduce the initial investment and also offering advantages in terms of mobility and collaboration. Therefore, cloud computing can be a good ally to the dematerialization of business processes.

CONCLUSION

The main purpose of this chapter was to clarify what the dematerialization of business processes is through the presentation of the main concepts behind it, and which in turn are further explored in the literature.

First, the electronic documents management, the workflow management and the reengineering of business processes topics were briefly introduced. After that, a great emphasis was given to the vital role of ICTs. It was conducted a comparative study in order to evaluate in which extent open-source software could be a good option for the implementation of the dematerialization of business processes with a lower cost for organizations. Regarding the technological offers available in the market, it is concluded that, although the proprietary solutions are the most comprehensive, open-source technologies

now occupy a prominent position, because they offer a set of features that responds to the needs of most organizations looking for new approaches to the production, organization, movement and retrieval of information based on the redefinition of its processes and information circuits.

With this chapter, it is intend to present the dematerialization of business processes as a good solution for problems related to the existence of documents in paper format, which are common to most organizations, enabling the improvement of its performance by managing their workflows. This chapter can give a significant theoretical contribution to this study area, although we believe that there are still many possible directions to explore more about the dematerialization of business processes once it is an approach that integrates the contribution of other knowledge domains.

REFERENCES

- Aalst, W. d., & Hee, K. V. (2009). Gestão de Workflows. Coimbra: Imprensa da Universidade de Coimbra. (in Portuguese)
- Afonso, C. M., Schwarz, A., Roldán, J. L., & Franco, M. J. S. (2015). EDMS Use in Local E-Government. *International Journal of Electronic Government Research*, 11(2), 18–34. doi:10.4018/IJEGR.2015040102
- Al Qady, M., & Kandil, A. (2013). Document Management in Construction: Practices and Opinions. *Journal of Construction Engineering and Management*, 139(10), 06013002. doi:10.1061/(ASCE)CO.1943-7862.0000741
- Alzubi, K. (2015). Implementing electronic documentation management system. *Journal of Theoretical and Applied Information Technology*, 77(1), 52–61. Retrieved from <http://www.scopus.com/inward/record.url?eid=2-s2.0-84936936214&partnerID=tZOtx3y1>
- Ashby, L. (2011). Extension's Progress in the Paperless Revolution : Balancing Digital and Paper. *Journal of Extension*, 49(1), 1–5.
- Attaran, M. (2004). Exploring the relationship between information technology and business process reengineering. *Information & Management*, 41(5), 585–596. doi:10.1016/S0378-7206(03)00098-3
- Aversano, L., Canfora, G., De Lucia, A., & Gallucci, P. (2002). Business process reengineering and workflow automation: A technology transfer experience. *Journal of Systems and Software*, 63(1), 29–44. doi:10.1016/S0164-1212(01)00128-5
- Ávila, L., Teixeira, L., & Almeida, P. (2015). Promotion of Administrative Modernization through Processes Dematerialization. In *Encyclopedia of Information Science and Technology* (3rd ed.; pp. 640–649). Information Science Reference.
- Ávila, L., Teixeira, L., & Almeida, P. (in press). A methodological approach to dematerialization of business processes using open-source technology. *International Journal of Industrial Engineering and Management*.
- Burns, C. (2009). Insurers Scale the Many Steps to Paperless. *Insuring Networks News*, 12(8), 14–18.
- Cho, V. (2008). A study of the effectiveness of electronic document management systems. *International Journal of Information Technology and Management*, 7(3), 327. doi:10.1504/IJITM.2008.017608
- Collett, S. (2010). Open-source: Grows up. *Computerworld*, 44(22), 23–25.
- Downing, L. (2006). Implementing EDMS: Putting People First. *Information Management Journal*, 40(4), 44–50.
- Hammer, M., & Champy, J. (1993). *Reengineering the Corporation: A manifesto for business revolution* (1st ed.). New York: Harper Business.

- Harmsen, S. (2007). Electronic document management in clinical research. *Die Pharmazeutische Industrie*, 69(7), 785–790. Retrieved from <http://www.scopus.com/inward/record.url?eid=2-s2.0-34547797491&partnerID=tZOtx3y1>
- Haug, A. (2012). The implementation of enterprise content management systems in SMEs. *Journal of Enterprise Information Management*, 25(4), 349–372. doi:10.1108/17410391211245838
- Hjelt, M., & Björk, B.-C. (2007). End-User Attitudes toward EDM Use in Construction Project Work: Case Study. *Journal of Computing in Civil Engineering*, 21(4), 289–300. doi:10.1061/(ASCE)0887-3801(2007)21:4(289)
- Hollingsworth, D. (1995). *The Workflow Reference Model*. Workflow Management Coalition, TC00-1003 (1.1).
- Hung, S.-Y., Tang, K.-Z., Chang, C.-M., & Ke, C.-D. (2009). User acceptance of intergovernmental services: An example of electronic document management system. *Government Information Quarterly*, 26(2), 387–397. doi:10.1016/j.giq.2008.07.003
- Johnston, G. P., & Bowen, D. V. (2005). The benefits of electronic records management systems: A general review of published and some unpublished cases. *Records Management Journal*, 15(3), 131–140. doi:10.1108/09565690510632319
- Lee, K. (2013). A Study on User Access Control Method using Multi-Factor Authentication for EDMS. *International Journal of Security and Its Applications*, 7(6), 327–334. doi:10.14257/ijisia.2013.7.6.33
- Limam Mansar, S., & Reijers, H. A. (2007). Best practices in business process redesign: Use and impact. *Business Process Management Journal*, 13(2), 193–213. doi:10.1108/14637150710740455
- Medina, R., & Fenner, J. (2005). Controlling your documents. *Information & Management*, 39(1), 20–22.
- Mentzas, G., Halaris, C., & Kavadias, S. (2001). Modelling business processes with workflow systems: An evaluation of alternative approaches. *International Journal of Information Management*, 21(2), 123–135. doi:10.1016/S0268-4012(01)00005-6
- Nagy, D. E. L., Yassin, A. M., & Bhattacharjee, A. (2010). Organizational Adoption of Open-source Software: Barriers and Remedies. *Communications of the ACM*, 53(3), 148–151. doi:10.1145/1666420.1666457
- Schmidt, R. A., Simmons, K., Grimm, E. E., Middlebrooks, M., & Changchien, R. (2006). Integration of scanned document management with the anatomic pathology laboratory information system: Analysis of benefits. *American Journal of Clinical Pathology*, 126(5), 678–683. doi:10.1309/77143BRXM3XDGBVX PMID:17050064
- Schnitzer, E. (2010). Paperless World. *Multi-Housing News*, 45(2), 33–34.
- Siatiras, K. (2004). Electronic document management--why you need it. *Chartered Accountants Journal*, 83(5), 4–6. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=13357351&site=ehost-live>
- Want, C. (2009). Bring Your Organisation up to Speed: The Benefits of Electronic Document Management. *IDMi (Information & Document Management International)*, (59), 14. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=47637229&site=ehost-live>
- Weerakkody, V., Janssen, M., & Dwivedi, Y. K. (2011). Transformational change and business process reengineering (BPR): Lessons from the British and Dutch public sector. *Government Information Quarterly*, 28(3), 320–328. doi:10.1016/j.giq.2010.07.010

Wisniewski, C. S., Pummer, T. L., & Krenzelok, E. P. (2010). The implementation of an electronic document management system at a drug information center. *The Journal of Pharmacy Technology*, 26(2), 66–70. doi:10.1177/875512251002600204

KEY TERMS AND DEFINITIONS

Dematerialization of Business Processes: The management of processes through an information system that allows simultaneously the workflow and electronic documents management.

Electronic Documents Management: The management of documents that exist in a digital format.

Information and Communication Technologies: Tools used to support information sharing and communication in a digital world.

Open-Source Software: A software developed by a group of volunteers and which source code is available for anyone, who can modify or improve it.

Proprietary Software: A close source software, licensed under exclusive legal right of the copyright holder.

Reengineering of Business Processes: A total redefinition of processes in order to improve significantly their results.

Workflow Management System: A system that controls the execution of tasks by different actors in order to complete a process with the desire output, in the right time.

Scanning for Blind Spots

B

Barbara Jane Holland
Brooklyn Public Library, USA

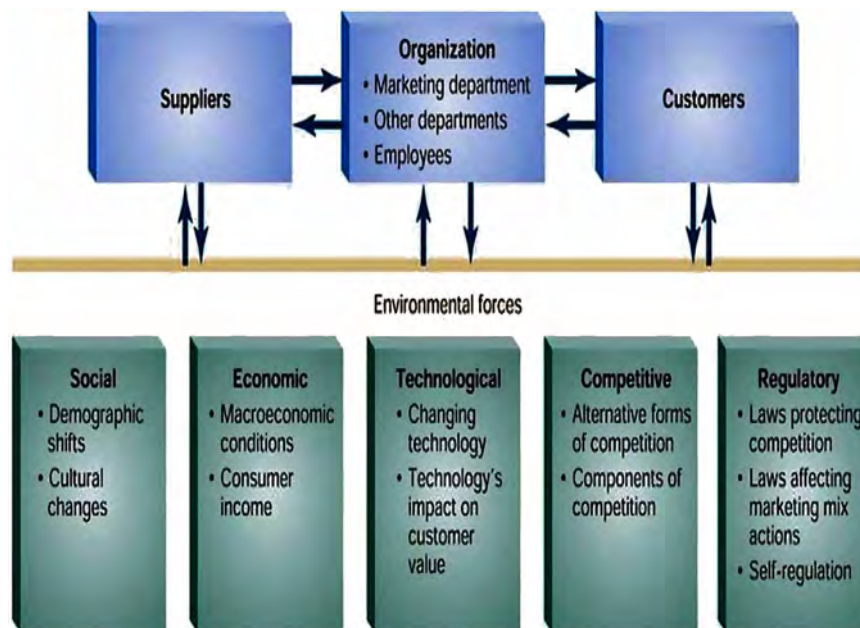
INTRODUCTION

Many research studies indicate that when reality and conviction are at odds with each other conviction often win. Companies and nonprofit organizations are often blindsided when faced with demographic, political, environmental, technological changes and new rivals. The most talented and best prepared leaders face complex, ever-changing, and often unpredictable challenges. The more an organization utilizes a systemic approach, the more likely they will avoid blind spots. Environmental Scanning or ES, provides a view of community needs by detecting pertinent economic, social, cultural, environmental, health, technological and political trends, situations and events.

Frank Aguilar's classic book "Scanning the Business Environment", exemplify some of the factors involved in determining what external information is strategic or relevant. The managers studied by Aguilar were mainly interested in the news of the market, including competitors, customers, pricing, market structure and change...

To be effective in a complex and changing world, decision makers must look ahead to anticipate emerging trends, issues, opportunities, and threats. Developing and applying strategic foresight must come into play. A well conducted scan results in a new management and marketing style that is more forward thinking. Scanning has long been standard practice in the military, the intelligence community, and the business world and is a core method in futures research. In recent

Figure 1. Environmental scanning



DOI: 10.4018/978-1-5225-2255-3.ch077

years, horizon scanning has been used in a growing number of fields in the public sector, such as human health (Douw and Vondeling 2006) and education (Munck and McConnell 2009). This chapter defines environmental scanning, examines Blind spots, weak signals, and effective tools used for discovering emerging trends beneficial to a company's or organization's strategic planning.

BACKGROUND

The lack of diversity in information sources may contribute to blind spots in environmental scanning. If organizations get their information from acceptable, conservative or authoritative sources, they may not see the changes taking place in their external environment. A large amount of pertinent information comes from face to face discussions, rumor, gossip and conjecture.

Moreover, it keeps the organization informed of current trends and challenges of the future.

Companies find it demanding to identify and include basic signals about their future developments and challenges in their existing, and often static strategic planning processes. Environmental scanning helps corporations, and nonprofits to understand external threats and opportunities leading to change. When an organization has undergone significant changes the knowledge discovered by an environment scan is intrinsic.

Aguilar (1967), in his study of the information gathering practices of managers, defined scanning as the systematic collection of external information in order to (1) lessen the randomness of information flowing into the organization and (2) provide early warnings for managers of changing external conditions.

Research suggests that effective scanning and planning is linked to improved organizational learning and performance (Choo 2002).

Environmental scanning practiced by different Japanese and Korean corporations share a surprisingly common pattern as observed in a 1988 study

by Ghoshal). Information gathering is typically achieved through six channels and sources:

- The planning division,
- Individual specialists,
- Patent department,
- Advisory boards,
- Individual employees, and
- Technology attaches.

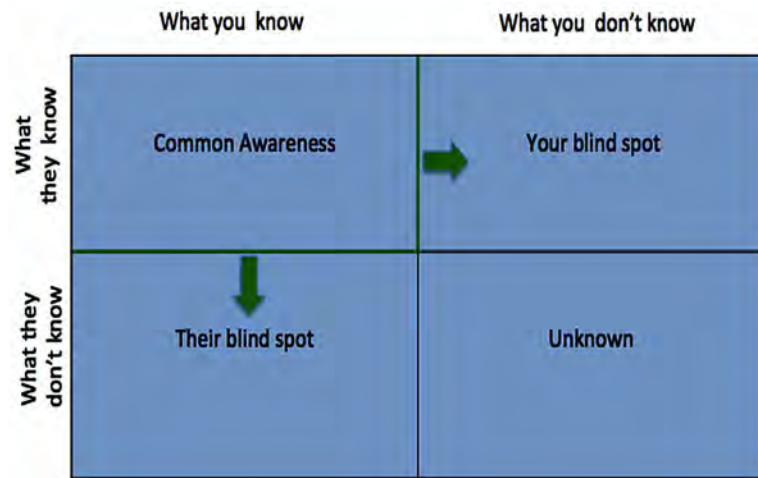
ENVIRONMENTAL SCANNING AND THE CORPORATE ENVIRONMENT

Environmental scanning practiced by various Japanese corporations share a common pattern. The same similarity has been observed in large South Korean corporations (Ghoshal 1988).

Japan owe much of its success to trading companies after world war II. Their corporate culture is based on gathering information from trade magazines, newspapers and business associates. Raw information is collected daily and transmitted overseas to the Japanese trading company called Sogo Shosha. A senior intelligence staff collates and shapes this information and submits it to key policy makers within the trading company. The Sogo Shosha have become masters at collecting business intelligence, tapping into the most public, boring sources, cultivating open human contacts.

A major task in any strategic planning process is thus to challenge existing perceptions and to identify blind spots as well as weak signals in order to effectively and efficiently detect future chances and risks at an early stage. Existing tools for identifying future changes that a company might face, such as operative forecasting or strategic forecasting, are generally not able to perform this task very well. Blind spots can be described as developments that a company knowingly or unknowingly overlooks while weak signals can be described as first indicators for future changes in the environment.

Figure 2. Blindspot analysis chart



Blind spots are commonly caused by an individual's or organization's psychological filtering of incoming data flow from economic and industry analysis when determining strategic decisions. The presence of blind spots within an organization becomes apparent when one of the following occurs:

1. Firm is unaware of a pending strategic development
2. Firm is aware, but misinterprets the strategic development
3. Firm correctly perceives the strategic development, but is too slow to react.

ANALYTICAL TOOLS AND DETECTING BLIND SPOTS

Blind spots may prevent an organization from recognizing a problem or enacting a solution.

Identifying and communicating the blind spots acknowledges their existence and power. People may still choose to support their blind spot while understanding the risks.

Originally conceived by Michael Porter and Developed by Benjamin Gilad, blind spot analysis attempts to identify and improve flaws, assumptions and inaccuracies that adversely affect strategies and decisions. SWOT analysis, although widely used and misused is to assess the current situation of an organization, department, program or strategy.

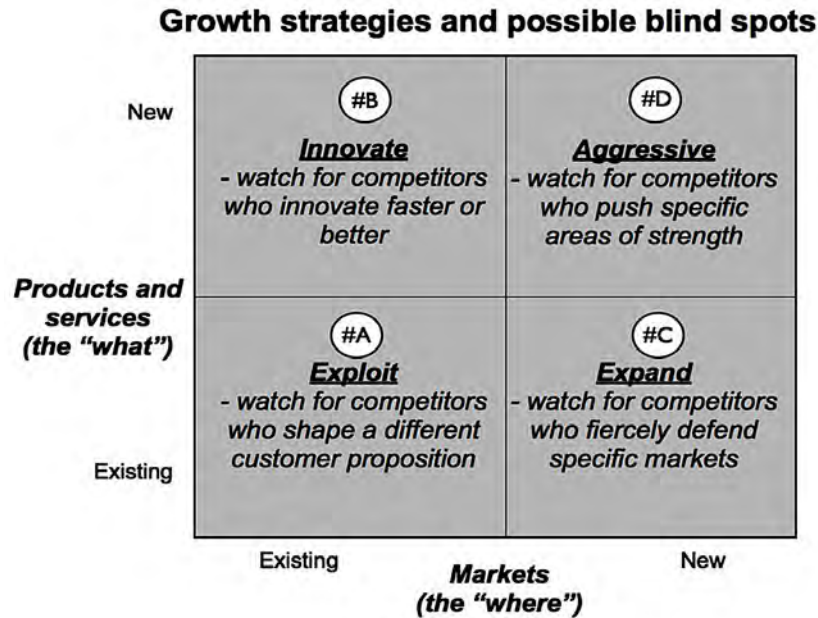
DEFINING THE ISSUE TO BE ANALYZED

Example: Libraries should use drones to improve outreach by delivering resources to the geographically isolated or homebound and providing deposit collections to disaster areas. In addition to con-

Table 1. Analyzing blindspots

Invalid Assumption	Winners Curse	Escalating Commitment	Constrained Perspective	Over Confidence	Information Filtering	Educated Incapacity
organizational Myths Unchecked Facts	Overpayment to achieve objectives	Pride/Fear Prevents Loss Cutting	Over estimate risks	Over estimate expertise and knowledge	Misinterpretation of Information	Experience created bias

Figure 3. Blindspot matrix



necting researchers via video equipped drones, and lending library books on a 24 hour basis. This may be analyzed using a blind spot analysis and SWOT analysis.

SWOT Analysis was originated by Albert S Humphrey in the 1960s, the tool is as useful now as it was then. You can use it in two ways – as

a simple icebreaker helping people get together to “kick off” strategy formulation, or in a more sophisticated way as a serious strategy tool. Strengths and weaknesses are often internal to your organization, while opportunities and threats generally relate to external factors.

Figure 4. SWOT analysis chart



Table 2. SWOT analysis

Strengths	Weaknesses
1.Strength One	1.Weakness One
2.Strength Two	2.Weakness Two
3.Strength Three	3.Weakness Three
Opportunities	Threats
1.Opportunity One	1.Threat One
2.Opportunity Two	2.Threat Two
3.Opportunity Three	3. Threat Three

After using SWOT and Blind spot Analysis to define the problem and consider solutions then create action plans to mitigate the risks.

WHAT ELSE SHOULD YOU BE LOOKING FOR?

TOWS Analysis is a variant of the classic business tool, SWOT Analysis. The only difference between TOWS and SWOT is that TOWS emphasizes the external environment while SWOT emphasizes the internal environment. SWOT or TOWS analysis helps you get a better understanding of the strategic choices that you face.

It helps you ask, and answer, the following questions: How do you:

- Make the most of your strengths?
- Circumvent your weaknesses?
- Capitalize on your opportunities?
- Next step of analysis, usually associated with the externally-focused TOWS Matrix, helps you think about the options that you could pursue. To do this you match external opportunities and threats with your internal strengths and weaknesses.

SPOTTING AND INTERPRETING TRENDS

Trends are developments in the environment of an organization and multiple events in a particular direction. It is the pattern that counts. Early warnings and signals are usually swept off the table unless you have proof and data to support it.

Scanning will result in a repository of trends that you and your scanning team think might be important to your organization's future. The next stage in your scanning process is to interpret your trends. This work is best done as a group by your scanning team. During this process scanners explore trends might evolve over time. Spotting a trend as an individual, can be risky. Moreover, you most likely will not be given the credit you deserve and may be terminated for pointing out

Table 3. Risk mitigation plan

RISK	Severity	Probability	Warning Signs	Plan
Summarize the risk	Low/medium/high	Low/medium/high	Identify the signals that warn you when risks are occurring	Create potential solutions to mitigate the risk

Table 4. SWOT worksheet

Strengths What do you do well? What unique resources can you draw on? What do others see as your strengths?	Weaknesses What could you improve? Where do you have fewer resources than others? What are others likely to see as weaknesses?
Opportunities What opportunities are open to you? What trends could you take advantage of? How can you turn your strengths into opportunities?	Threats What threats could harm you? What is your competition doing? What threats do your weaknesses expose you to?

Figure 5. TOWS matrix

Internal Factors External Factors	Strengths (S)	Weaknesses (W)
	SO Strategies Generate strategies here that use strengths to take advantage of opportunities	WO Strategies Generate strategies here that take advantage of opportunities by overcoming weaknesses
Opportunities (O)		
Threats (T)	ST Strategies Generate strategies here that use strengths to avoid threats	WT Strategies Generate strategies here that use minimize weaknesses and avoid threats

Table 5. 2TOWS strategic alternatives worksheet

	External Opportunities(O) 1. 2. 3. 4.	External Threats (T) 1. 2. 3. 4.
Internal Strengths (S) 1. 2. 3. 4.	SO	ST
Internal Weaknesses (W) 1. 2. 3. 4.	WO	WT

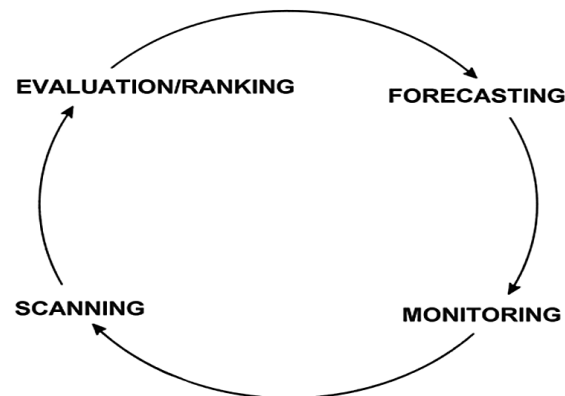
new developments which are in stark contrasts with the official view (de Rujiter 2014). Most trend reports are one sided. They report about globalization and ignore a growing trend of anti-counter globalization. Counter-trends are more interesting. Counter-trends are full of insights and are many times the bases of real foresights.

Once you come up with a new trend, finding the data to support it is the most difficult task. The data has to be named and captured to support the trend.

MAJOR ISSUES, DRAWBACKS AND LIMITATIONS

Scanning may not always be effective in an organization. The volume of information can be so overwhelming resulting information overload. Pieces of information may be overlooked and or missed. There are many sources of information that scanners are not be aware of missing information that's pertinent. Information may not have been gathered in a timely manner. Furthermore, determining if the information is relevant, familiarity with the topic and information sources language

Figure 6. Scanning and forecasting chart



usage, all play a role in the analysis process. Environmental uncertainty can lead to misjudgments.

Judging the appropriate level of resources to devote to environmental scanning is difficult. Where environmental conditions are turbulent and full of potentially significant changes, large amounts of resources may be justified. Even with large efforts, there is no guarantee that some wholly-unanticipated event will cause serious problems or present a great opportunity for which you aren't prepared. More importantly, the interpretation of trend information and forecasting is an inexact science at best. The farther into the future a scan probes, the more cautious you must be with the interpretation. Blind spots and Group think must be avoided.

AVOIDING BLIND SPOTS AND GROUPTHINK

All organizations have blind spots. Groupthink is a cognitive dissonance on a collective level. It forces everyone in a group to think the same. It can be useful for visioning and alignment. On the other hand it can make an organization blind, especially if there is an dominate leader or extremely strong peer pressure to think alike. It's not enough to spot the trend but also be alert to the prevailing organizational perspective. It is better to use outsiders as they are not affected

by company blindness. They are allowed to say things that insiders can't. Managers have instinctive attitudes and assumptions about the external environment. They usually pay attention to what's most immediate, namely their industry sector and environments. They may have blind-spots which ignore or discount certain market sector trends (Pitt 2012).

Active and open exploration of communities incorporating diverse sources of information and viewpoints are keys to successful scanning. When organizations utilize a systemic approach, they are more likely to avoid blind spots while scanning.

You may not be given credit for noticing important trends. Being too blunt might get you fired for pointing out new developments which are in stark contrast to the official view. Spotting trends is not enough but you must also be alert to the prevailing organizational perspective (Rujiter 2014).

Understanding the institutional frame of reference with help you locate the blind spots and simultaneously make you aware of potential resistance you'll need to overcome to get your point across. Here, it is best to use outsiders. They are not affected by company blindness and cannot be fired. They are allowed to say what insiders cannot. The outside view can stretch the company mindset. Outsiders can lightened all the blind spots..

SOLUTIONS AND RECOMMENDATIONS

Emerging issues start with a value shift, or a change in how an issue is viewed. An opinion leader or champion inevitably emerges to move the issue into the public view. It is at this time that the emerging issue can be identified.

Leaders often are forced to make decisions with flawed or incomplete information and, more importantly, often don't know that is the case—creating blind spots. One way of avoiding blind spots is to create a culture within a company that promotes straight talk about problems and concerns. Developing peripheral vision can help alleviate blind spots.

- **De-Emphasize the Hierarchy:** Break tradition when in the process of problem-solving on broad organizational issues, skip level down into the organization and gain new perspectives. Just be sure to let your management team know that you're going to be doing it, so that they don't overreact.
- **Discourage Group Think:** Rather, promote speaking for one's self. Whenever you hear someone on your management team begin a sentence with something like, "I think we all believe..." it is quite possible that you have fallen into some form of groupthink.
- **Create Recognition Opportunities:** By establishing mechanisms that foster ways for individuals or work units to be recognized for "out of the box" thinking, a leader can begin to change the very culture that has inadvertently muted the growth potential of a business.

TEAM MEMBERS AS OBJECTS OF STUDY

Get to know your team in depth in order to interpret the subtleties of their behavior. Knowing how to read team members means not only being aware

of their tendencies but also of any deviations from their typical approach. For example, if an individual who is inclined to examine every side of an issue before making a decision suddenly appears to see only one side of an argument, you need to understand why that is.

PAY ATTENTION TO BEHAVIORAL RED FLAGS

Be aware of subtle clues that may suggest people are not expressing what they really think

Warning signals you should watch for:

- Nonverbal behaviors, such as eye rolling.
- Omissions, which are the important areas that are simply not surfaced or discussed.
- Silence, which may indicate an unwillingness to take a contrary position.
- Non-answers, such as when an individual is obviously being intentionally evasive.

When you scan the environment you should interview and talk to people in and out of your organization. Aim to talk to people that have different approaches and perspectives.\

FUTURE RESEARCH DIRECTIONS EXAMINED

Desk Research

A sensible way to research backgrounds and facts when you want to map the outside world. When conducting Desk Research, you are looking for signals about future. Many times the amount of research has already been done with desk research. Other sources can be used besides the internet such as reference books, newspapers, magazines, and annual reports. The advantages of Desk Research you can work from your office, you do not need other people and it focuses on facts. RSS feeds can used this way to support this way of environmental scanning.

The Disadvantage of Desk Research

Test may be outdated. You cannot allow yourself to be guided by data of the past. You may run the risk of losing yourself in time-consuming theoretical studies in which you have only consulted a few people. By involving several people in your scan of the environment, you create ownership in a broader group thereby raising support for the project.

Currently, with the help of a DESTEP analysis organizations may easily get the overview of its external environment. Each and every company is faced with some factors they cannot exercise influence on and the factors they can exert influence on. The DESTEP theory mainly represents the Demographic, Economic, Social/ Cultural, Technological, Economic and Political factors. These factors may directly or indirectly influence the organization's operations and clarifies how a company can adjust their strategy with such macro-economic factors.

Horizon Scanning

Horizon scanning, a term frequently applied to futures scanning is frequently based on desk research, helping to develop the big picture behind the issues to be examined. This is scanning at the periphery, at the edges of current thinking, looking for emerging issues, threats and opportunities. Also known as environmental scanning, external scanning, and strategic scanning, horizon scanning may also be defined as "the acquisition and use of information about events, trends and relationships in an organization's external environment, which would assist management in planning the organization's future course of action. Effective horizon scanning serves as an early warning system to identify potential opportunities and threats, enable decision makers to plan accordingly and take timely action, and foster a culture of foresight throughout an organization.

The goals are to find emerging indications of important future developments so decision makers

can plan accordingly and take action quickly and more broadly to foster a culture of foresight in the organization. Some horizon scanning projects and systems fail to meet expectations or achieve goals, for a variety of reasons such as insufficient budget, lack of management support, and weak stakeholder participation. An overabundance of raw data and techniques to make sense of the data is not progressing fast enough.

The Foresight Process

The foresight process is to gather insightful information on potential futures, develop future scenarios, and action plans in the event that one of scenarios unfold. When participants in the foresight process are all have similar knowledge and attitudes towards trends, and developments, they are prone to develop monotonic scenarios.

Foresight, traditionally have involved only a limited number of top level decision makers from the organization. As a rule, internal players have been assigned to scan the environment, gather external and internal information, processing it into forecasts, and presenting it to top management. This procedure provides top management with information that facilitates decision making on current and new business development in general and innovation activities. Many companies discover that opening this process can identify new trends, business opportunities and detect potential blind spots.

Open Foresight workshops enable the company to learn about weak signals.. When managers recognize weak signals, they can identify new business opportunities.

See Tables 6, 7, and 8

TREND WATCHERS AND FUTURE OPPORTUNITIES

Trends, are interlinked and connected to socioeconomic and cultural drivers influencing tomorrow's government and business success stories. Trends

Table 6. Foresight methods by type of technique

Qualitative Methods Providing Meaning to Events and Perceptions. Interpretations are Often Difficult to Corroborate	Quantitative Methods Measuring Variables and Applying Statistical Analysis Using Reliable Data	Semi-Quantitative Methods That Apply Mathematical Principles to Quantify Subjectivity
brainstorming conferences, workshops scanning genius forecasting scenario writing simulation gaming SWOT analysis/wild cards surveys role play	benchmarking bibliometrics modelling patent analysis trend extrapolation	Delphi Multi-criteria analysis polling-voting road mapping stakeholder analysis

Table 7. Foresight methods classified according to output

Tool	Quantitative	Qualitative	Normative	Exploratory
Scenario Analysis	X	X	X	X
Delphi		X	X	X
Relevance Trees		X	X	
Decision Matrices	X			X

Source: The Millennium project 1994

help you find the right insight or opportunity, generate the right ideas, and then execute them in the right way. The result is happy consumers. The difference between a trend and a fad is when trends emerge and external change unlocks new ways to serve age-old human needs and desires. There's the expectation that the instant connection and social gratification available online will be a part of offline social lives, too. This is a trend. Then you have to consider is the service a fad, or will young urbanites still be swiping furiously in 2035? The question is not important: trends aren't about the success or failure of individual

innovations but will consumer desire for instant connection and social gratification continue?

Trend watchers understand that if they rely *only* on conventional market research they'll risk missing powerful *future* opportunities. Today trend watchers use a network of spotters who send hundreds of local innovation examples every month. They allow a unique window on global consumerism as it evolves week by week, month by month. consumerism as it evolves week by week, month by month. These insights can help companies such as Trendwatching.com spot new trends, or help us update existing trends.

Kjaer Global, is a leading international trend forecasting consultancy helping companies navigate the future. They specialize in Trend Management – strategic platforms and trend tools such as a Trend Atlas, to enable the development of 'future proof' concepts for people centric products and services. Their forecasts include, by 2020 the focus will be on prevention rather than healing. Cloud health intelligence will evolve, as apps, mobile diagnostics and intuitive biofeedback become

Table 8. Foresight methods classified according to perspective

Perspective	Method
Identifying Issues	SWOT Analysis
Extrapolative Approaches	Multi-agent simulation
Creative Approaches	Brainstorming
Prioritization	Road mapping

Source: Miles and Keenan 2003

accessible resources in our daily lives. Printing your own medicine will become available. Gender equality will be crucial. Two thirds of graduates in advanced economies will be women majoring in business.

CONCLUSION

There are many advantages to environmental scanning in modern organizations. Environmental scanning assists in the development of strategic plans and policies, the assessment of new information, and the adjustment of internal operations to meet new challenges as they arise. It can identify an organization's unique strengths, locate weaknesses in its competitors, and identify new markets, prospective customers, and emerging technologies..

Despite the advantages of environmental scanning, it is not widely used in companies as might be suggested in academic publications on the subject. Although scanning is challenging, it is also necessary. Bringing diverse stakeholders together in participatory scanning is also challenging. Identifying blind spots and weak signals of change in a complex, turbulent environment is like looking for a sewing needle on plush rug. Separating meaningful signals from all the noise requires the rare ability to see beyond prevailing mindsets and paradigms. Analyzing and correctly interpreting potential signals of change calls for creativity and insight. Acting on possible opportunities and threats presented by an early change demands visionary leadership and careful strategy. Successfully achieving all of these requirements is indeed difficult. Scanning can enhance discussion about future-oriented issues within an organization, as well as help decision makers anticipate and quickly respond to external change. A formal horizon scanning system is a vital but often missing component in the strategic planning process of public natural resource and environmental organizations.

REFERENCES

- Aguilar, F. J. (1967). *Scanning the business environment*. New York: Macmillan Co.
- Bengston, D. (2013). *Horizon Scanning for Environmental Foresight: A Review of Issues and Approaches*. Retrieved October 7, 2015 from http://www.fs.fed.us/nrs/pubs/gtr/nrs_gtr121.pdf
- Brown, A., & Weiner, E. (1985). *Super Managing: How to harness change for personal and organizational success*. New York: Mentor.
- Castiglione, J. (2008). Environmental scanning: An essential tool for twenty-first century librarianship. *Library Review*, 57(7), 528–536. doi:10.1108/00242530810894040
- Chemelik, S. (2015). *Museum and historic site management: a case study approach*. Rowman & Littlefield.
- Choo, C. (1999). *The art of scanning the environment*. Retrieved April 16, 2013 <http://choo.fis.utoronto.ca/FIS/respub/ASISbulletin/>
- Choo, C. W. (2001). *Information management for the intelligent organization: the art of scanning the environment* (3rd ed.). Medford, NJ: Information Today, Inc.
- Douw, K., & Vondeling, H. (2006). Selection of new health technologies for assessment aimed at informing decision making: A survey among horizon scanning systems. *International Journal of Technology Assessment in Health Care*, 22(2), 177–183. doi:10.1017/S0266462306050999 PMID:16571193
- Fahey, L., King, W. R., & Narayanan, V. K. (1981). Environmental scanning and forecasting in strategic planning—the state of the art. *Long Range Planning*, 14(1), 32–39. doi:10.1016/0024-6301(81)90148-5
- Georgiou, L. (2008). *Handbook of technology foresight: concepts and practice*. Edward Elgar.

Ghoshal, S. (1988). Environmental Scanning in Korean Firms: Organizational Isomorphism in Action. *Journal of International Business Studies*, 19(1), 69–86. doi:10.1057/palgrave.jibs.8490375

Giatouzi, M. (2013). *Recent developments in foresight methodologies*. New York: Springer. doi:10.1007/978-1-4614-5215-7

Gilad, B. (2004). *Early warning: using competitive intelligence to anticipate company shifts*. New York: AMA.

Griffin, A., & Moble, C. (2014). *Open Innovation: New product development essentials from the PDMA*. John Wiley & Sons.

Idris, M. (2013). *Design thinking for strategic innovation: what they can't teach you at business or design school*. John Wiley & Sons.

Lesca, H., & Lesca, N. (2014). *Strategic decisions and weak signals: anticipation for decision-making*. John Wiley & Sons. doi:10.1002/9781118959152

Miles, I., & Keenan, H. (2003). Overview of methods used in foresight" the technology foresight for organizers training course. Academic Press.

Pitt, M., & Koufopoulos, D. (2012). *Essentials of strategic management*. London: Sage.

Ruijter, P. (2014). *Scenario based strategy: navigate the future*. London: Gower.

KEY TERMS AND DEFINITIONS

Blind Spots: Areas important information go unnoticed and thus the organization falls into one of any number of traps. These include misjudging industry boundaries, failing to identify emerging competition, falling out of touch with customers, over-emphasizing a competitor's visible competence, and allowing corporate taboos or lack of foresight to limit their frame of reference. Any one of these mistakes will prevent organizational leaders from taking advantage of the opportunities available to them.

Competitive Scanning: The objective is to understand competitors' strategic positioning in terms of sustainable development but also the way in which they are evolving.

Delphi: It is essentially the polling of experts, over a number of rounds, to generate consensus about a particular topic or issue. Delphi does not survey the outlying views or the dissenting views, but is useful where a high degree of credibility is required. The key question to ask here, as with all methods, is what image of the future is generated in this scanning output? And, whose voice is not being heard? Japan has been using a Delphi approach for many years, and produces quarterly reviews on science and technology trends.

Environmental Scanning: Defined as 'the study and interpretation of the political, economic, social and technological events and trends which influence a business, an industry or even a total market.

External Environment: Evaluation of the possible or probable effects of external forces and conditions on an organization's survival and growth strategies.

Foresight: Defined as the capacity to think systematically about the future to inform decision making today. It is a capacity that we need to develop as individuals, as organisations, and as a society.

Futures Approaches: The tools, methods and thinking styles used to build an organisational foresight capacity, usually interdisciplinary and inclusive rather than restricted to a particular method or philosophy.

Horizon Scanning: It is a term frequently applied to futures scanning. This is scanning at the periphery, at the edges of current thinking, looking for emerging issues, threats and opportunities.

Macro Environment: Major external and uncontrollable factors that influence an organization's decision making, and affect its performance and strategies. These factors include the economic factors; demographics; legal, political, and social conditions; technological changes; and natural forces.

Metascanning: There are many people and organizations whose job it is to scan the future. Some will let you have limited access to some of their findings for free, but most will ask you to sign up for a service. The value is that they have already done the initial work for you saving you time – and are satisfied with their approach and methods.

Paralysis by Analysis: The state of over analyzing.

PEST: It is an acronym that stands for Political, Economic, Social and Technological factors affecting a business decision. It is a framework used in the early phases of strategy development to describe the landscape and environment in which a firm operates.

Strategic: The decision making process whereby a strategic plan is drawn up.

Strategic Scanning: Involves looking for what are known, in the foresight profession, as “weak

signals.” and as a result, they may seem to have little or no bearing on “here and now” and may therefore not seem useful.

Strategy: Michael Porter a strategy expert and professor at Harvard Business School, emphasizes the need for strategy to define and communicate an organization’s unique position, and says that it should determine how organizational resources, skills, and competencies should be combined to create competitive advantage.

Strategic Scanning: Involves looking for what are known, as “weak signals and have little or no bearing on “here and now” and may therefore not seem useful.

Trend: A grouping of similar or related events that tends to move in a given direction, increasing or decreasing in strength of frequency of observation; usually suggests a pattern of change in a particular area (for example, consumer behaviour, technology use).

Strategic Information Systems Planning

Maria Kamariotou

University of Macedonia, Greece

Fotis Kitsios

University of Macedonia, Greece

INTRODUCTION

In this day, turbulence in the businesses' environment is on the increase. Therefore, businesses are obliged to respond to that environmental uncertainty. Information Systems (IS) and Strategic Planning support this effort. IS are connected with business strategy, management skills, decision making and aims to increase competitive advantage (Zubovich et al., 2014). Researchers have focused on the process of Strategic Information Systems Planning (SISP) since 1970. The purpose of SISP is to support the goals and the business strategy, through IS. Furthermore, SISP helps businesses to innovate, create new products, reduce costs and enhance relationships with customers (Kamariotou & Kitsios, 2015, Ullah & Lai, 2013). The process of SISP contains five phases, which are strategic awareness, situation analysis, strategy conception, strategy formulation and strategy implementation planning. These phases help businesses not only to define IS strategy but also to develop IS.

Previous studies have examined the effect of these phases on SISP success. Also, other studies have concluded that there is a positive relationship between SISP and firm performance. These surveys that have discussed the relationship between SISP and firm performance have been presented only by a theoretical approach (Goldsmit, 1991, Lederer & Sethi, 1996). These surveys collected data using questionnaires in large firms. Surveys in SMEs which constitute the main component of each economy, are limited however (Newkirk et al., 2003).

Consequently, the aim of this chapter is to propose a holistic approach in order to investigate the significance of SISP process, to highlight phases that contribute to a greater extent of success and to draw conclusions concerning the successful implementation of digital strategy in firms.

The structure of this chapter is as following: after a brief introduction to this field, a theoretical framework is analyzed on the basis of the literature review about SISP phases and success in firms. Then in the next section, the findings of previous surveys are presented, whereas the final section concludes the paper.

BACKGROUND: STRATEGIC INFORMATION SYSTEMS PLANNING AND PERFORMANCE

SISP has become a significant planning activity in organizations and a major issue for IS management. Many factors have contributed to the change of the role of IS in business in recent years. These factors involve the use of IS for competitive advantage, the effusion of IS in businesses, the involvement of IS on businesses' daily operations and the increase of interorganizational systems. Due to these changes, strategic planning for the IS function is necessary so that businesses can effectively achieve their goals in this complex and dynamic environment (Pvemkumar & King, 1991, 1994).

SISP can be defined as the ability to shape the strategy of a business and tools, techniques and

methodologies were suggested to support organizations in determining potential opportunities to deploy IS with greater competitiveness (Pepard, et al., 2014). SISP is an integrated process which includes certain phases. These phases and their activities are presented as follows. The first phase is Strategic awareness. This phase includes the identification of key planning issues, planning objectives, organizing the planning team and the encouragement of top level managers. The second phase is Situation analysis. This phase involves the analysis of current business systems, current organizational systems and current IS. Also, it includes the analysis of the current external and internal business environment and the current external Information Technology (IT) environment. Next, the third phase is Strategy conception. This phase includes the determination of main IS objectives, opportunities for improvement, alternative scenarios as well as the evaluation of opportunities for improvement. Also, it includes the definition of high level IS strategies. The next phase is Strategy formulation. In this phase, businesses select the most suitable scenario from the previous alternatives scenarios, according to new business processes and new IT architectures. Then this scenario is evaluated according to its strategic and technological impact. Also, in this phase, specific new projects and priorities for new projects are identified. These projects consist of specific activities which support the implementation of the selected scenario. The last phase is Strategy implementation planning. This phase involves approaching the actions of change management and the evaluation of strategic plan (Brown, 2004, 2010, Kamariotou & Kitsios, 2017, 2016, Kitsios & Kamariotou, 2016, Maharaj & Brown, 2015, Mentzas, 1997, Mirchandani & Lederer, 2014, Newkirk & Lederer, 2006, Newkirk, et al., 2008).

Despite the fact that previous researches have shown the success of the implementation of these phases in IS' development, many factors affect negatively the success of the process. Such important critical factors are management commitment, the involvement of employees, the

communication between the members of the team as well as the planning horizon. On the other hand, their advantages are the increase of products' quality and the effectiveness of IS (Brown, 2010, Cohen, 2008, Haki, 2011, Luftman & Brier, 1999, Pai, 2006, Silvius & Stoop, 2013, Teo, et al., 1997, Yang, et al., 2013, Zubovich, et al., 2014). Other factors which have an impact on SISP process are the analysis of the environment and the knowledge sharing (Maharaj & Brown, 2015, Pai, 2006, Yang & Pita, 2014, Yang, et al., 2013, Zubovich, et al., 2014). The determination of planning with a certain frame is very important due to technological changes which affect SISP process (Newkirk, et al., 2008).

King and Teo, (2000), argue that the factors which influence the process of SISP are the understanding of the objectives, management commitment and the creation of integrated plans. These factors impact on the firm performance of a business and especially on the index of Return on investment (ROI), as well as on the increase of market share and on customer's satisfaction. Also, Lederer and Sethi, (1991), concluded that the most important factors are related with systems' architecture, their cost and their implementation. Furthermore, they highlighted factors which are related to managers' involvement, the small horizon of projects, the support of financial plan for the development of IS, as well as knowledge sharing, analysis of business environment and understanding of strategic objectives. Earl, (1993), considered the linkage between business objectives and IS, the increase of competitive advantage through IT as well as the necessary resources for the development of IS as the major factors.

The success of SISP is defined as "the degree to which the objectives of SISP are achieved" (Pai, 2006). Several authors have measured the success of SISP with two dimensions: quality and effectiveness, while others have measured it with four, named: alignment, analysis, collaboration and capabilities (Newkirk & Lederer, 2006, Pai, 2006). A combination of the above dimensions proposed by Yang and Pita, (2014), who measure

the success of SISP with aligning business strategy and Information Systems' strategy and the effectiveness of Information Systems Design. Other researchers measure SISP success with variables such as user satisfaction, quality of information and system and organizational impact (Heo & Han, 2003). Another different approach for the measurement of the process comes from Silvius and Stoop, (2013), who suggested variables such as the alignment between business strategy with Information Systems strategy, the development of skills, the degree of management commitment and the achievement of objectives.

As for the four dimensions of SISP success, the first one includes the understanding of IS in supporting strategy, the identification of opportunities to support the strategic direction of the firm, the alignment between IS strategies with the strategic plan of the organization, the education of top management on the importance of IT and the adaption of technology to strategic change. The second dimension contains variables such as the generation of new ideas to reengineer business processes through IT, the understanding of information needs through subunits, the understanding of the dispersion of data, applications, and other technologies throughout the firm, the development of a "blueprint" which structures organizational processes, the improvement of the understanding of how the organization actually operates and the monitoring of internal business needs and the capability of IS to meet those needs. The third dimension includes the development

of clear guidelines of managerial responsibility for plan implementation, the Identification of potential sources of resistance to IS plans, the support of open lines of communication with other departments, the achievement of a general level of agreement regarding the risks/tradeoffs among system projects and the avoidance of the overlapping development of major systems. Finally, the last dimension includes capabilities such as the ability to identify key problem areas, the ability to anticipate surprises and crises, the flexibility to adapt to unanticipated changes and the ability to gain cooperation among user groups for IS plans (Newkirk & Lederer, 2006, Newkirk et al., 2003).

Table 1 summarizes the phases and the dimensions of SISP success.

Researchers have measured firm performance with variables such as sales growth, profitability, innovation, growth of customer satisfaction and growth of market share of products and services (Andersen, 2001, Bergeron et al., 2004, Chatzoglou et al., 2011, Croteau & Bergeron, 2001, King & Teo, 2000, Oh & Pinsonneault, 2007).

In Europe at about 75% of all businesses are SMEs. Although, family businesses pay attention to the business's long-term sustainability, they do not implement strategic planning (Siakas et al., 2014). Greece is a country which has plenty of SMEs comparatively to other countries of the European Union and most of them have been negatively affected from the financial crisis (Vassiliadis & Vassiliadis, 2014).

Figure 1. Conceptual framework

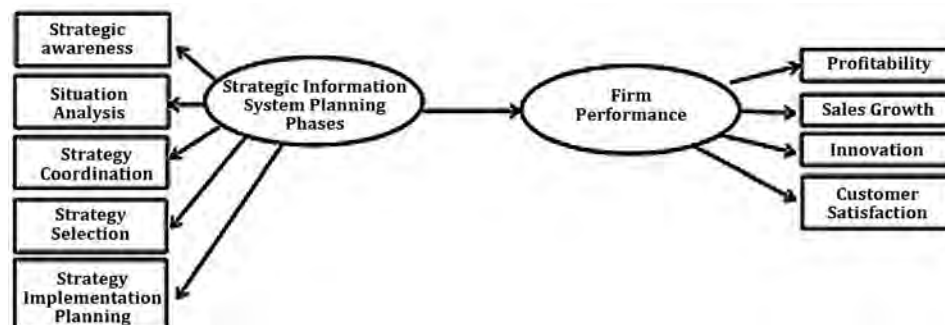


Table 1. Description of the SISP process

Variables	Questions	References
Strategic Awareness	Determining key planning issues	(Mentzas, 1997, Mirchandani & Lederer, 2014, Newkirk & Lederer, 2006, Newkirk et al., 2003)
	Defining planning objectives	(Mentzas, 1997, Mirchandani & Lederer, 2014, Newkirk & Lederer, 2006, Newkirk et al., 2003)
	Organizing the planning team(s)	(Mentzas, 1997, Mirchandani & Lederer, 2014, Newkirk & Lederer, 2006, Newkirk et al., 2003)
	Obtaining top management commitment	(Mentzas, 1997, Mirchandani & Lederer, 2014, Newkirk & Lederer, 2006, Newkirk et al., 2003)
Situation Analysis	Analyzing current business systems	(Mentzas, 1997, Mirchandani & Lederer, 2014, Newkirk & Lederer, 2006, Newkirk et al., 2003)
	Analyzing current organizational systems	(Mentzas, 1997, Mirchandani & Lederer, 2014, Newkirk & Lederer, 2006, Newkirk et al., 2003)
	Analyzing current information systems	(Mentzas, 1997, Mirchandani & Lederer, 2014, Newkirk & Lederer, 2006, Newkirk et al., 2003)
	Analyzing the current external business environment	(Mentzas, 1997, Mirchandani & Lederer, 2014, Newkirk & Lederer, 2006, Newkirk et al., 2003)
	Analyzing the current external IT environment	(Mentzas, 1997, Mirchandani & Lederer, 2014, Newkirk & Lederer, 2006, Newkirk et al., 2003)
Strategy Conception	Identifying major IT objectives	(Mentzas, 1997, Mirchandani & Lederer, 2014, Newkirk & Lederer, 2006, Newkirk et al., 2003)
	Identifying opportunities for improvement	(Mentzas, 1997, Mirchandani & Lederer, 2014, Newkirk & Lederer, 2006, Newkirk et al., 2003)
	Evaluating opportunities for improvement	(Mentzas, 1997, Mirchandani & Lederer, 2014, Newkirk & Lederer, 2006, Newkirk et al., 2003)
	Identifying high level IT strategies	(Mentzas, 1997, Mirchandani & Lederer, 2014, Newkirk & Lederer, 2006, Newkirk et al., 2003)
Strategy Formulation	Identifying new business processes	(Mentzas, 1997, Mirchandani & Lederer, 2014, Newkirk & Lederer, 2006, Newkirk et al., 2003)
	Identifying new IT architectures	(Mentzas, 1997, Mirchandani & Lederer, 2014, Newkirk & Lederer, 2006, Newkirk et al., 2003)
	Identifying specific new projects	(Mentzas, 1997, Mirchandani & Lederer, 2014, Newkirk & Lederer, 2006, Newkirk et al., 2003)
	Identifying priorities for new projects	(Mentzas, 1997, Mirchandani & Lederer, 2014, Newkirk & Lederer, 2006, Newkirk et al., 2003)
Strategy Implementation	Defining change management approach	(Mentzas, 1997, Mirchandani & Lederer, 2014, Newkirk & Lederer, 2006, Newkirk et al., 2003)
	Defining action plan	(Mentzas, 1997, Mirchandani & Lederer, 2014, Newkirk & Lederer, 2006, Newkirk et al., 2003)
	Evaluating action plan	(Mentzas, 1997, Mirchandani & Lederer, 2014, Newkirk & Lederer, 2006, Newkirk et al., 2003)
	Defining follow-up and control procedure	(Mentzas, 1997, Mirchandani & Lederer, 2014, Newkirk & Lederer, 2006, Newkirk et al., 2003)

continued on following page

Table 1. Continued

Variables	Questions	References
Alignment	Maintaining a mutual understanding with top management on the role of IS in supporting strategy	(Newkirk & Lederer, 2006, Newkirk et al., 2003)
	Understanding the strategic priorities of top management	(Newkirk & Lederer, 2006, Newkirk et al., 2003)
	Identifying IT-related opportunities to support the strategic direction of the firm	(Newkirk & Lederer, 2006, Newkirk et al., 2003)
	Aligning IS strategies with the strategic plan of the organization	(Newkirk & Lederer, 2006, Newkirk et al., 2003)
	Adapting the goals/objectives of IS to changing goals/objectives of the organization	(Newkirk & Lederer, 2006, Newkirk et al., 2003)
	Educating top management on the importance of IT	(Newkirk & Lederer, 2006, Newkirk et al., 2003)
	Adapting technology to strategic change	(Newkirk & Lederer, 2006, Newkirk et al., 2003)
	Assessing the strategic importance of emerging technologies	(Newkirk & Lederer, 2006, Newkirk et al., 2003)
Analysis	Identifying opportunities for internal improvement in business processes through IT	(Newkirk & Lederer, 2006, Newkirk et al., 2003)
	Maintaining an understanding of changing organizational processes and procedures	(Newkirk & Lederer, 2006, Newkirk et al., 2003)
	Generating new ideas to reengineer business processes through IT	(Newkirk & Lederer, 2006, Newkirk et al., 2003)
	Understanding the information needs through subunits	(Newkirk & Lederer, 2006, Newkirk et al., 2003)
	Understanding the dispersion of data, applications, and other technologies throughout the firm	(Newkirk & Lederer, 2006, Newkirk et al., 2003)
	Development of a ‘blueprint’ which structures organizational processes	(Newkirk & Lederer, 2006, Newkirk et al., 2003)
	Improved understanding of how the organization actually operates	(Newkirk & Lederer, 2006, Newkirk et al., 2003)
	Monitoring of internal business needs and the capability of IS to meet those needs	(Newkirk & Lederer, 2006, Newkirk et al., 2003)
Cooperation	Developing clear guidelines of managerial responsibility for plan implementation	(Newkirk & Lederer, 2006, Newkirk et al., 2003)
	Identifying and resolving potential sources of resistance to IS plans	(Newkirk & Lederer, 2006, Newkirk et al., 2003)
	Maintaining open lines of communication with other departments	(Newkirk & Lederer, 2006, Newkirk et al., 2003)
	Coordinating the development efforts of various organizational subunits	(Newkirk & Lederer, 2006, Newkirk et al., 2003)
	Establishing a uniform basis for prioritizing projects	(Newkirk & Lederer, 2006, Newkirk et al., 2003)
	Achieving a general level of agreement regarding the risks/tradeoffs among system projects	(Newkirk & Lederer, 2006, Newkirk et al., 2003)
	Avoiding the overlapping development of major systems	(Newkirk & Lederer, 2006, Newkirk et al., 2003)
Capabilities	Ability to identify key problem areas	(Newkirk & Lederer, 2006, Newkirk et al., 2003)
	Ability to anticipate surprises and crises	(Newkirk & Lederer, 2006, Newkirk et al., 2003)
	Flexibility to adapt to unanticipated changes	(Newkirk & Lederer, 2006, Newkirk et al., 2003)
	Ability to gain cooperation among user groups for IS plans	(Newkirk & Lederer, 2006, Newkirk et al., 2003)

The current financial crisis has negatively impacted plenty of activities and most of the family businesses have already found themselves in a new turbulent financial environment where uncertainty dominates and the market characteristics radically changes. Except for financial barriers, their lack of technological, managerial and human capabilities may restrict their ability to bowl over the financial crisis (Bourletidis et al., 2014). Also, the lack of strategic planning negatively increases this difficulty. It appears that formal processes in SMEs which are connected with strategic management and information handling support managers to pay attention to strategies, structures and processes that are expected to increase firm performance. In uncertain environments, businesses should formalize processes using standardized rules and procedures which support the lowering of environmental uncertainty and arrange economic consistency. Formalization encourages the development of frameworks that request communication among the individuals as well as sharing of new information, and also supports the transformation of new ideas into real plans through the flexible structures. The result is the raise of innovation (Giannacourou et al., 2015).

SISP is a formal process which can be followed by SMEs in order to determine IS strategy and develop the most suitable IS for their needs. This process supports businesses to make decisions on the planning and the development of IS, analyzing their resources taking not only the environmental opportunities but also the threats into consideration. Also, SISP includes all the factors and the activities which are presented above as the advantages of the formalization. In this view, Mirchandani & Lederer, (2014), investigated SISP phases and they found that as the environment becomes more uncertain, more situational analysis is needed. The analysis of current business systems, organizational systems and IS, as well as current external and internal business environment and the current external IT environment allow the organization to identify problems and diagnose opportunities.

The findings of these surveys show that IS executives went their efforts into the strategic conception phase. Despite the fact that planners put their efforts into this phase, they cannot define the suitable alternative strategies so their efforts do not affect SISP success. As a result, they cannot achieve their aims. The most common problems which have been observed during the SISP process are the lack of participation and the failure to implement strategic IS plans. Executives cannot be committed to the plan, so the team has difficulties to implement the IS strategy. Also, findings highlight that executives understand that the implementation phase is difficult but very important and they are looking for ways in order to pay more attention on this phase (Lederer & Sethi, 1991, Newkirk & Lederer, 2006, Newkirk et al., 2003, Zubovic, et al., 2014). Internet can be used as a tool which supports the communication between the individuals and increases their participation to the process (Andersen, 2001). Except from Internet all the IT applications can support the knowledge sharing and the communication between the individuals (Pai, 2006). The participation between the members of the team is increased when CEOs support the process (Brown, 2010, King & Teo, 2000).

Results show that several managers put too much efforts to SISP process while others too little. In the first case, the process could be confused, detained or its implementation is prevented. In the second case, the implemented plans could be inefficient and they could not meet the objectives. So, the evaluation of the process is very important because managers can avoid these unsatisfactory results.

Findings conclude that managers focus on Strategy Conception and Strategy Implementation and they do not concentrate on Strategic Awareness and Situation Analysis, so the implemented plans are not effective, successful and they do not meet the aims (Brown, 2010, Newkirk & Lederer, 2006, Newkirk et al., 2003). Also, the strong focus on the implementation of the process leads to shorter SISP horizons but the strategic objectives have

not been achieved. Executives do not concentrate on what strategic aims really matter and how they can add value to the business because they pay attention to the horizon of the project and to reduce its cost due to limited IT budget (Brown, 2010). The results suggest that planners should pay attention on conducting situational analysis with greater punctiliousness, so they can develop strategy conception and strategy implementation planning with greater versatility than they currently implement. Managers should analyze their current business systems, organizational systems, IS, business environment and external IT environment. An understanding of those elements can improve the outcome of the planning process despite the fact that the time and cost will be increased. When managers understand the environment, they can identify significant IT objectives and opportunities for improvement, evaluate them for improvement and identify high level IT strategies in their business' strategy conception (Mirchandani & Lederer, 2014 Zubovic, et al., 2014).

FUTURE RESEARCH DIRECTIONS

This study has focused on the implementation of a digital strategy through SISP process. As a result, further empirical research is required to determine the success of SISP, especially in SMEs. More research could also be realised in the effect of factors (such as management commitment, horizon of projects, clear strategic objectives, financial support) in the success of SISP. So, the results of an exploratory study will be summarised in a conceptual model for further research. Finally, researchers can use different dimensions in order to measure SISP success and they can use tools from strategic management in order to implement each phase more effectively and to develop a successful digital strategy. In this view, they could examine if the methodologies used in the formulation, implementation and evaluation of

business strategy can be implemented in each phase in order to be developed a digital strategy.

CONCLUSION

Up to now academic studies have not focused on the impact of SISP phases on success. This chapter presented a framework which can provide a formal process to be followed by IS executives and managers in order to plan and use the right IS and to raise competitive advantage.

The authors contribute to the existing studies by increasing IS executives awareness of the strategic use of IS planning in order to accomplish success and competitive advantage. The presented framework is significant for SMEs. SMEs are very important for national and global economy but they do not implement strategic planning, so they cannot increase their resources as well as their competitive advantage. SMEs should develop strategic planning in order to develop and use IS according to their resources and their needs.

REFERENCES

- Andersen, T. J. (2001). Information Technology, strategic decision making approaches and organizational performance in different industrial settings. *The Journal of Strategic Information Systems*, 10(2), 101–119. doi:10.1016/S0963-8687(01)00043-9
- Bergeron, F., Raymond, L., & Rivard, S. (2004). Ideal patterns of strategic alignment and business performance. *Information & Management*, 41(8), 1003–1020. doi:10.1016/j.im.2003.10.004
- Bourletidis, K., & Triantafyllopoulos, Y. (2014). SMEs Survival in Time of Crisis: Strategies, Tactics and Commercial Success Stories. *Procedia: Social and Behavioral Sciences*, 148, 639–644. doi:10.1016/j.sbspro.2014.07.092

- Brown, I. (2010). Strategic Information Systems Planning: Comparing Espoused Beliefs with Practice. *Proceedings of 18th European Conference on Information Systems (ECIS)*, 1-12.
- Brown, I. T. J. (2004). Testing and Extending Theory in Strategic Information Systems Planning Through Literature Analysis. *Information Resources Management Journal*, 17(4), 20-48. doi:10.4018/irmj.2004100102
- Chatzoglou, P. D., Diamantidis, A. D., Vraimaki, E., Vranakis, S. K., & Kourtidis, D. A. (2011). Aligning IT, strategic orientation and organizational structure. *Business Process Management Journal*, 17(4), 663-687. doi:10.1108/14637151111149474
- Cohen, J. F. (2008). Contextual determinants and performance implications of information systems strategy planning within South African firms. *Information & Management*, 45(8), 547-555. doi:10.1016/j.im.2008.09.001
- Croteau, A. M., & Bergeron, F. (2001). An information technology trilogy: Business strategy, technological deployment and organizational performance. *The Journal of Strategic Information Systems*, 10(2), 77-99. doi:10.1016/S0963-8687(01)00044-0
- Earl, M. J. (1993). Experiences in Strategic Information Systems Planning. *Management Information Systems Quarterly*, 17(1), 1-24. doi:10.2307/249507
- Giannacourou, M., Kantaraki, M., & Christopoulou, V. (2015). The Perception of Crisis by Greek SMEs and its Impact on Managerial Practices. *Procedia: Social and Behavioral Sciences*, 175, 546-551. doi:10.1016/j.sbspro.2015.01.1235
- Goldsmith, N. (1991). Linking IT planning to business strategy. *Long Range Planning*, 24(6), 67-77. doi:10.1016/0024-6301(91)90045-P
- Haki, M. K. (2011). A model and empirical test of information technology strategy success. *International Journal of Information Systems and Change Management*, 5(1), 54-75. doi:10.1504/IJISCM.2011.039070
- Kamariotou, M., & Kitsios, F. (2015). Innovating with Strategic Information Systems Planning: A Structured Literature Review. *Proceedings of the 2nd International Conference on Applied Innovation*.
- Kamariotou, M., & Kitsios, F. (2016). Strategic Information Systems Planning: SMEs Performance outcomes. *Proceedings of 5th International Symposium and 27th National Conference on Operational Research*, 153-157.
- Kamariotou, M., & Kitsios, F. (2017). Information Systems Phases and Firm Performance: A conceptual Framework. In A. Kavoura, D. Sakas, & P. Tomaras, (Eds.), *Springer Proceedings in Business and Economics of 1st International Conference on Business Informatics and Modelling, Online, Global*. doi:10.1007/978-3-319-33865-1_67
- King, W. R., & Teo, S. H. T. (2000). Assessing the impact of proactive versus reactive modes of strategic information systems planning. *Omega*, 28(6), 667-679. doi:10.1016/S0305-0483(99)00079-1
- Kitsios, F., & Kamariotou, M. (2016). The impact of Information Technology and the alignment between business and service innovation strategy on service innovation performance. *Proceedings of 3rd International Conference on Industrial Engineering, Management Science and Applications (ICIMSA 2016)*, 247-251. doi:10.1109/ICIMSA.2016.7504042
- Kitsios, F., & Kamariotou, M. (2016). Decision Support Systems and Business Strategy: A conceptual framework for Strategic Information Systems Planning. *Proceedings of the 6th International Conference on IT Convergence and Security (ICITCS2016)*, 149-153. doi:10.1109/ICITCS.2016.7740323

- Lederer, A. L., & Sethi, V. (1991). Critical Dimensions of Strategic Information Systems Planning. *Decision Sciences*, 22(1), 104–119. doi:10.1111/j.1540-5915.1991.tb01265.x
- Lederer, A. L., & Sethi, V. (1996). Key prescriptions for strategic information systems planning. *Journal of Management Information Systems*, 13(1), 35–62. doi:10.1080/07421222.1996.11518111
- Luftman, J., & Brier, T. (1999). Achieving and sustaining business-IT alignment. *California Management Review*, 42(1), 109–122. doi:10.2307/41166021
- Maharaj, S., & Brown, I. (2015). The impact of shared domain knowledge on strategic information systems planning and alignment: Original research. *South African Journal of Information Management*, 17(1), 1–12.
- Mentzas, G. (1997). Implementing an IS Strategy-A Team Approach. *Long Range Planning*, 30(1), 84–95. doi:10.1016/S0024-6301(96)00099-4
- Mirchandani, D. A., & Lederer, A. L. (2014). Less is More: Information Systems Planning in an Uncertain Environment. *Information Systems Management*, 29(1), 13–25. doi:10.1080/10580530.2012.634293
- Newkirk, H. E., & Lederer, A. L. (2006). The effectiveness of strategic information systems planning under environmental uncertainty. *Information & Management*, 43(4), 481–501. doi:10.1016/j.im.2005.12.001
- Newkirk, H. E., Lederer, A. L., & Johnson, A. M. (2008). Rapid business and IT change: Drivers for strategic information systems planning? *European Journal of Information Systems*, 17(3), 198–218. doi:10.1057/ejis.2008.16
- Newkirk, H. E., Lederer, A. L., & Srinivasan, C. (2003). Strategic information systems planning: Too little or too much? *The Journal of Strategic Information Systems*, 12(3), 201–228. doi:10.1016/j.jsis.2003.09.001
- Oh, W., & Pinsonneault, A. (2007). On the Assessment of the Strategic Value of Information Technologies: Conceptual and Analytical Approaches. *Management Information Systems Quarterly*, 31(2), 239–265.
- Pai, J. C. (2006). An empirical study of the relationship between knowledge sharing and IS/IT strategic planning (ISSP). *Management Decision*, 44(1), 105–122. doi:10.1108/00251740610641490
- Peppard, J., Galliers, R. D., & Thorogood, A. (2014). Information systems strategy as practice: Micro strategy and strategizing for IS. *The Journal of Strategic Information Systems*, 23(1), 1–10. doi:10.1016/j.jsis.2014.01.002
- Premkumar, G., & King, W. R. (1991). Assessing strategic information systems planning. *Long Range Planning*, 24(5), 41–58. doi:10.1016/0024-6301(91)90251-I
- Premkumar, G., & King, W. R. (1994). The evaluation of strategic information system planning. *Information & Management*, 26(6), 327–340. doi:10.1016/0378-7206(94)90030-2
- Siakas, K., Naaranoja, M., Vlachakis, S., & Siakas, E. (2014). Family businesses in the new economy: How to survive and develop in times of financial crisis. *Procedia Economics and Finance*, 9, 331–341. doi:10.1016/S2212-5671(14)00034-3
- Silvius, A. J. G., & Stoop, J. (2013). The relationship between the process of Strategic Information Systems Planning and its success; An explorative study. *Proceedings of 46th Hawaii International Conference on System Sciences (HICSS)*, 4495–4501. doi:10.1109/HICSS.2013.536
- Teo, S. H. T., Ang, S. K. J., & Pavri, F. N. (1997). The state of strategic IS planning practices in Singapore. *Information & Management*, 33(1), 13–23. doi:10.1016/S0378-7206(97)00033-5

Ullah, A., & Lai, R. (2013). A systematic review of business and information technology alignment. *ACM Transactions on Management Information Systems*, 4(1), 1–30. doi:10.1145/2445560.2445564

Vassiliadis, S., & Vassiliadis, A. (2014). The Greek family businesses and the succession problem. *Procedia Economics and Finance*, 9, 242–247. doi:10.1016/S2212-5671(14)00025-2

Webster, J., & Watson, R. T. (2002). Analyzing the Past to Prepare for the Future: Writing a Literature Review. *Management Information Systems Quarterly*, 26(2), 13–23.

Yang, J., Pita, Z., & Singh, M. (2013). A Conceptual Framework for Assessing Strategic Information Systems Planning (SISP) Success in the Current Dynamic Environments. *Proceedings of 24th Australasian Conference on Information Systems (ACIS)*, 1–10.

Zubovic, A., Pita, Z., & Khan, S. (2014). A Framework for Investigating the Impact of Information Systems Capability on Strategic Information Systems Planning Outcomes. *Proceedings of 18th Pacific Asia Conference on Information Systems*, 1–12.

ADDITIONAL READING

Amrollahi, A., Ghapanchi, A. H., & Talaei-Khoei, A. (2014). Three Decades of Research on Strategic Information System Plan Development. *Communications of the Association for Information Systems*, 34(1), 1439–1467.

Bechor, T., Neumann, S., Zviran, M., & Glezer, C. (2010). A contingency model for estimating success of strategic information systems planning. *Information & Management*, 47(1), 17–29. doi:10.1016/j.im.2009.09.004

Bourletidis, K. (2013). The strategic management of market information to SMEs during economic crisis. *Procedia: Social and Behavioral Sciences*, 73, 598–606. doi:10.1016/j.sbspro.2013.02.096

Chen, D. Q., Mocker, M., & Preston, D. S. (2010). Information Systems Strategy: Reconceptualization, Measurement, and Implications. *Management Information Systems Quarterly*, 34(2), 233–259.

Ensari, M. Ş., & Karabay, M. E. (2014). What Helps to Make SMEs Successful in Global Markets? *Procedia: Social and Behavioral Sciences*, 150, 192–201. doi:10.1016/j.sbspro.2014.09.030

Heo, J., & Han, I. (2003). Performance measure of information systems IS in evolving computing environments: An empirical investigation. *Information & Management*, 40(4), 243–256. doi:10.1016/S0378-7206(02)00007-1

Merali, Y., Papadopoulos, T., & Nadkarni, T. (2012). Information systems strategy: Past, present, future? *The Journal of Strategic Information Systems*, 21(2), 125–153. doi:10.1016/j.jsis.2012.04.002

Palanisamy, R. (2005). Strategic information systems planning model for flexibility and success. *Industrial Management & Data Systems*, 105(1), 63–81. doi:10.1108/02635570510575199

Raghunathan, B., & Raghunathan, T. S. (1991). Information Systems Planning and Effectiveness: An Empirical Analysis. *Omega*, 19(2-3), 125–135. doi:10.1016/0305-0483(91)90022-L

KEY TERMS AND DEFINITIONS

Alignment: The connection between business strategy and IS strategy.

Firm Performance: The power of a business to implement activities better rather than its competitors.

Resources: Resources can be categorized into physical, organizational, technological and financial. They are unique, they give value to the business and they cannot be copied by competitors.

Strategic Awareness: This phase includes the identification of key planning issues, planning objectives, organizing the planning team and the encouragement of top level managers.

Strategic Analysis: This phase involves analysis of current business systems, current organizational systems and current IS. Also, it includes the analysis of the current external and internal business environment and the current external IT environment.

Strategy Conception: This phase contains the determination of main IS objectives, opportunities for improvement, alternative scenarios as well as the evaluation of opportunities for improvement.

Also, it includes the definition of high level IS strategies.

Strategy Formulation: In this phase, businesses select the most suitable scenario from the previous alternatives scenarios, according to new business processes and new IT architectures. Then this scenario is evaluated according to its strategic and technological impact. Also, in this phase, specific new projects and priorities for new projects are identified. These projects consist of specific activities which support the implementation of the selected scenario.

Strategy Implementation Planning: This phase involves an approach and the actions of change management and the evaluation of strategic plan.

Strategic Information Systems Planning: The process of finding opportunities through IT which will add value to the organization, aligning its goals with IT and increases competitive advantage.

Sustainable Advantages of Business Value of Information Technology

B

Jorge A. Romero

Towson University, USA

INTRODUCTION

The business value of information technology (IT) and the way to sustain returns obtained through investment in IT has been a topic of interest in the last few years. Early studies during the 1980s and early 1990s on the relationship between business value and information technology investments did not find a link between them, although the firms analyzed at that time were investing large amounts of money on IT. This issue created what is called in the literature “the IT productivity paradox” (Brynjolfsson, 1993). It is essential to understand the relationship of IT and its effects on business value, especially the sustainable advantages that companies can obtain if they use IT to obtain a strategic differentiation and operating efficiency in using competitors as a benchmark.

Later studies have criticized the business value of IT investments claiming that IT does not bring any benefits to a firm because competitors can also have access to the same type of IT, thus eliminating any strategic advantage that a firm may have when they adopt a new type of IT (Carr, 2004).

However, more recent studies have found that IT improves productivity and profitability (Brynjolfsson and Hitt, 1996; Romero et al., 2010), and that IT can help firms differentiate their products from competitors and manage processes in a more efficient way (Anderson et al., 2011). For instance, Bardhan et al. (2013) proposed that IT investments interact with research and development investments of a firm enhancing its market value, and the results are visible through the increase in shareholder value. They used investments in research and development which is an intangible

variable that can be used as a proxy measure of the value and capabilities of a firm. They concluded that IT investments create business value, and they measured it with Tobin’s Q, which is a financial measure that takes into consideration future earnings and future growth of the firm.

With regard to the methodology used in most of the studies on the effects of business value of IT, most were based on cross sectional studies using data across industries and across different types of IT such as Enterprise Resource Planning (ERP), Customer Relationship Management (CRM), Material Requirements Planning (MRP), Supply Chain Management (SCM), and cloud-based applications among others. One criticism of this methodology is that each type of IT differs from another, including different installation periods and different investments, therefore the conclusions reached in these studies need to be analyzed with this in mind. Very few studies have looked at the business value of information technology using only one industry and using only one type of IT.

BACKGROUND: INDUSTRY STUDIES

Previous studies have concluded that the implementation of IT and its effects may affect firms in different industries in different ways (Porter, 2001; Melville et al., 2004; Mitra, 2007). Therefore, it is important to recognize the structural differences across industries when we try to measure business value and the managerial ability of firms applying IT in order to improve their performance. The understanding of those structural differences is not yet completely understood (Melville et al. 2004).

From a methodological point of view, it is possible to control industry effects in a regression across industries, but the concern with those studies across industries is that the industry effect may be averaged out when analyzing firms from a broad group of industries. On the contrary, in a study that focuses on one industry, the effect of industry-related factors are controlled and the unique characteristics of the specific industry can be taken into consideration in order to isolate the IT effects (Mitra, 2007). There have been very few industry studies analyzing the effect of business value of IT. The main limitation has been the difficulty to obtain and collect this kind of data which is more precise and accurate than cross sectional data. One of the advantages of an industry study is that it does not restrict the analysis; rather it minimizes the possibility of bogus results due to structural differences among industries.

For instance, Anderson et al. (2011) is one of the few industry studies that analyzed the effects of one type of IT within one specific industry. Using only one industry gave them the opportunity to avoid interactions among industries, to have a more homogenous dataset and to use this industry as a control variable. Using highly specific industry data, their study used a novel model to look at the business value that is generated during the implementation and after the implementation of one type of IT. They also looked at whether or not the business value of IT was sustainable over time. They concluded that firms that were involved in shorter IT implementations outperformed firms that were involved in longer IT implementations. One of their explanations for these results is that when a firm is involved in a long implementation, then structural changes within the firm and macroeconomic changes that may occur during the implementation may affect the implementation, whereas during a short implementation, firms avoid this potential issue. Also, during a long implementation, firms also face the possibility of having an outdated system by the time the implementation is complete. They also discussed that firms involved in shorter IT implementations

started to enjoy operational improvements before the IT implementation was officially declared complete. This is even more critical on large IT systems such as ERP where the operations of the whole firm may be affected and its supply chain interactions may also be affected.

BUSINESS VALUE ON ENTERPRISE RESOURCE PLANNING

It is important to discuss ERP as one type of IT because it is one of the most important types of IT that have had an important effect on business value of firms in the last two decades. ERP is a type of IT that is capable of processing enormous amounts of data and has become a widely-used package, particularly after major events such as the year 2000 problem (Y2K) when firms had to start using four digits in the variable year starting on year 2000. This, combined with the enforcement of the Sarbanes-Oxley Act of 2002 that imposed stringent accounting regulations on firms after several major financial scandals that included large publicly traded firms such as Enron, Adelphia, and Worldcom, among others, led to an increase in the use of ERP. With the enforcement of the Sarbanes-Oxley Act of 2002, ERP has helped firms standardize business processes and helped firms integrate their systems with the systems of suppliers and customers.

ERP has been considered as an extension of MRP II, but with more sophisticated features and applications (Gumaer, 1996; Yusuf and Little, 1998). Likewise, MRP was designed to replace legacy reorder point-based IT systems in production and planning departments (Cooper and Zmud, 1990). Large and complex IT systems such as ERP, that are very expensive and take a long period of time to implement, may generate disruption in the operations and will affect profitability of a company (Bardhan et al. (2013).

There are multiple studies on ERP, but there are few studies analyzing the effect of business value of IT over long periods of time using only

one type of IT such as ERP. Previous studies have not found any issues or any vendor related effects using only one type of IT. In the case of ERP, not every vendor may offer an identical package and similar functionalities in their ERP systems and prices and support offered by the corresponding vendor may be different. In addition, the complexity of each ERP system may be different, therefore more complex systems may impact firms in different ways (Reck, 2004).

Hunton et al. (2003) analyzed the effects of ERP comparing ERP-adopters versus non-adopters. They also identified the different ERP vendors associated with each data point in their sample and run a statistical test to see if there was a vendor effect that may affect the results based on the fact that not every vendor offers an identical system and there is a difference in price and support provided by each vendor. They concluded that there was no statistically significant vendor effect.

Most of the previous studies on ERP did not use detailed implementation data, go live dates, or types of ERP modules installed providing limitation on the types of conclusions that can be generalized (e.g. Hitt et al., 2002; Hunton et al., 2003; Nicolau, 2004; Nicolau et al., 2006). Reck (2004) concluded that, based on the results from previous studies, there were still contradictory results on the effects of IT on business value. Later on, Stratman (2007) concluded that the adoption of ERP by firms did not guarantee them a sustainable advantage over competitors or a positive return on investment.

Romero et al., (2010) is one of the few studies that used a specific industry to analyze the effects of business value on one type of IT: ERP. The analysis covered an important period of time during which firms were making large investments in ERP, and the dataset that they used had detailed information about installation periods and helped them isolate the effects of ERP. They looked at profitability ratio and its five components: profitability ratio, productivity ratio, capacity utilization ratio, product mix ratio, and price recovery ratio in order to drill down on the sources of profitability.

Their model used the pre-implementation period as a benchmark and the group of non-adopters as a control group. They concluded that the source of profitability was coming from the following two components: productivity and capacity utilization. Another important contribution of this study to the literature on business value is that the model used in this study can be easily applied to other types of IT.

FUTURE RESEARCH DIRECTIONS: CLOUD COMPUTING

New types of IT, such as cloud computing, will have a great impact on business value and firms will have to adapt their business models in order to incorporate this new IT. Cloud computing is a general term that can include ERP services, CRM services, online storage, or backup services. The growth rate of cloud services has grown at a higher rate than the ones of IT services in general (Choudhary and Vithayathil 2013). Cloud-based ERP applications are also available (Choudhary and Vithayathil 2013).

Cloud computing allows firms to reduce costs and can be deployed in a short period of time. Firms can reduce cost because the cost structure is based on the number of users and the period of time that the software will be used. For instance, the period could be a month, a quarter or a year (Chatman, 2010). Firms that use cloud-computing services eliminate fixed IT costs, and only incur variable costs based on their usage (Choudhary and Vithayathil 2013). Cloud software is also scalable so it can support an increase in the number of users, if needed (Chatman, 2010). When we talk about cloud computing, we are not only looking at the services offered but also at the rental of computing and storage (Choudhary and Vithayathil 2013).

Choudhary and Vithayathil (2013) looked at cloud computing as a service to the firm. They also noted that cloud computing services are standardized products with a limited amount of customization. Cloud computing services are

expected to offer high quality services; otherwise it will be very difficult and costly for the firm to try to improve the quality of the vendor.

With regard to business value and competitive advantage that firms may get in relation to competitors, firms that use cloud computing applications have to accept that cloud computing services are standard for all firms using that service, a characteristic that may affect the business value of the firm in relation to competitors (Choudhary and Vithayathil 2013).

Cloud computing gives firms the possibility of getting real-time data, anywhere connectivity, as well as affordable digital optimization to generate business value and competitive advantage by creating customer intimacy through the use of algorithms designed specifically to identify each customer, and by creating operational excellence based on a more agile process (Weinman, 2015).

Cloud computing can foster business innovation and can create business value because business units in a firm do not need to depend on their corresponding IT department for technology needs, business department can bypass their IT departments and rent what they need from the cloud and pay as-they-go. In the past, IT department were not always synchronized with the needs of the business units, because sometimes IT departments had limited budgets, and limited personnel that not let them satisfy the business needs of the different departments and did not let them provide the IT infrastructure in a timely manner (Andriole, 2015). More research needs to be done in this area of cloud computing in order to examine the effects on business value.

CONCLUSION

Recent studies on the business value of IT using detailed data on one specific industry and using only type of IT have shown that IT provides a competitive advantage over competitor. The way to sustain the advantage is to see the investment in IT as a life cycle investment because technology

changes at a rapid pace, so firms have to be continuously innovating in order to keep their competitive advantage. One way to do it is to always stay at the forefront of technological innovations. For example, cloud computing is changing the way that firms do business and the way firms handle electronic records.

The proposal of Clemons and Row (1991) still holds on the fact that the strategic advantage will be sustainable only if the firm uses the IT in such a way that competitors cannot replicate the system. Competitors can buy the same type of IT, it may be the same type of ERP or they may be renting the same type of cloud service, and may try to replicate the system, but the only way to keep a strategic advantage over competitors is to see investments in IT as a life cycle investment in which the firm is constantly seeking ways to differentiate itself from competitors and to use its unique characteristics and unique resources to combine with IT to sustain the advantage over competitors, and therefore increase the business value of the firm. One way to differentiate from competitors and gain a sustainable competitive advantage is to add value and unique characteristics to the standard services provided by cloud computing services that can also be used by competitors.

This chapter provides a blueprint to conduct more research on this dynamic topic and will help researchers and managers to gain knowledge on the impact of IT. This chapter will also help researchers, managers and senior executives to identify the sources of competitive advantage and their managerial and strategic value to investments in IT.

REFERENCES

Anderson, M., Banker, R., Menon, N., & Romero, J. (2011). Implementing enterprise resource planning systems: Organizational performance and the duration of the implementation. *Information Technology Management*, 12(3), 197–21. doi:10.1007/s10799-011-0102-9

- Andriole, S. (2015). Who Owns IT? *Communications of the ACM*, 58(3), 50–57. doi:10.1145/2660765
- Bardhan, I., Krishnan, V., & Lin, S. (2013). Business Value of Information Technology: Testing the Interaction Effect of IT and R&D on Tobins Q. *Information Systems Research*, 24(4), 1147–1161. doi:10.1287/isre.2013.0481
- Brynjolfsson, E. (1993). The productivity paradox of information technology. *Communications of the ACM*, 36(12), 67–77. doi:10.1145/163298.163309
- Brynjolfsson, E., & Hitt, L. (1996). Paradox lost? Firm-level evidence on the returns to information systems spending. *Management Science*, 42(4), 541–558. doi:10.1287/mnsc.42.4.541
- Carr, N. (2004). *Does IT Matter? Information Technology and the Corrosion of Competitive Advantage*. Cambridge, MA: Harvard Business School Press.
- Chatman, C. (2010, May-June). How Cloud Computing is changing the Face of Health Care Information Technology. *Journal of Health Care Compliance*, 37-70.
- Chen, I., & Popovich, K. (2003). Understanding Customer Relationship Management (CRM) People, Process, and Technology. *Business Process Management Journal*, 9(5), 672–688. doi:10.1108/14637150310496758
- Choudhary, V., & Vithayathil, J. (2013). The Impact of Cloud Computing: Should the IT Department Be Organized as a Cost Center or a Profit Center? *Journal of Management Information Systems*, 30(2), 67–100. doi:10.2753/MIS0742-1222300203
- Clemons, E., & Row, M. (1991). Sustaining IT Advantage: The Role of Structural Differences. *Management Information Systems Quarterly*, 15(3), 275–292. doi:10.2307/249639
- Cooper, R., & Zmud, R. (1990). Information Technology implementation research: A technological diffusion approach. *Management Science*, 36(2), 123–139. doi:10.1287/mnsc.36.2.123
- Gumaer, R. (1996). Beyond ERP and MRPII. *IIE Solutions*, 28(9), 32–36.
- Hitt, L., Wu, D., & Xiaoge, Z. (2002). Investment in enterprise resource planning: Business impact and productivity measures. *Journal of Management Information Systems*, 19(1), 71–99.
- Hunton, J., Lippincott, B., & Reck, J. (2003). Enterprise resource planning systems: Comparing firm performance of adopters and non-adopters. *International Journal of Accounting Information Systems*, 4(3), 165–184. doi:10.1016/S1467-0895(03)00008-3
- Jodlbauer, H., & Reitner, S. (2012). Material and Capacity Requirements Planning with Dynamic Lead Times. *International Journal of Production Research*, 50(16), 447–4492. doi:10.1080/00207543.2011.603707
- Melville, N., Kraemer, K., & Gurbaxani, V. (2004). Review: Information Technology and Organizational Performance: An Integrative Model of IT Business Value. *Management Information Systems Quarterly*, 28(2), 283–322.
- Mitra, S. (2007). Nonaudit service fees and auditor independence: Empirical evidence from the oil and gas industry. *Journal of Accounting, Auditing & Finance*, 22(1), 85–107.
- Nicolaou, A. (2004). Firm performance effects in relation to the implementation and use of enterprise resource planning systems. *Journal of Information Systems*, 18(2), 79–105. doi:10.2308/jis.2004.18.2.79
- Nicolau, A., & Bhattacharya, S. (2006). Organizational performance effects of ERP systems usage: The impact of post-implementation changes. *International Journal of Accounting Information Systems*, 7(1), 18–35. doi:10.1016/j.accinf.2005.12.002

Olhager, J. (2013). *Evolution of Operations Planning and Control: From Production to Supply Chains*. *International Journal of Production Research*, 51(23-24), 6836–6843. doi:10.1080/00207543.2012.761363

Porter, M. (2001). Strategy and the Internet. *Harvard Business Review*, 79(3), 62–78. PMID:11246925

Reck, J. (2004). Discussion of: Firm Performance Effects in Relation to the Implementation and Use of Enterprise Resource Planning Systems. *Journal of Information Systems*, 18(2), 107–110.

Romero, J., Menon, N., Banker, R., & Anderson, M. (2010). ERP: Drilling for Profit in the Oil and Gas Industry. *Communications of the ACM*, 53(7), 118–121. doi:10.1145/1785414.1785448

Stratman, J. (2007). Realizing benefits from Enterprise Resource Planning: Does Strategic Focus Matter? *Production and Operations Management*, 16(2), 203–216. doi:10.1111/j.1937-5956.2007.tb00176.x

Weinman, J. (2015, November). Digital Technologies and Competitive Advantage. *Research-Technology Management*, 12–17.

Yusuf, Y., & Little, D. (1998). An empirical investigation of enterprise-wide integration of MRPII. *International Journal of Operations & Production Management*, 18(1), 66–86. doi:10.1108/01443579810192916

ADDITIONAL READING

Bogers, M., Boyd, B., & Hollensen, S. (2015). Managing Turbulence: Business Model Development in a Family-Owned Airline. *California Management Review*, 58(1), 41–64. doi:10.1525/cmr.2015.58.1.41

Botosan, C., & Huffman, A. (2015). Decision-Useful Asset Measurement from a Business Valuation Perspective. *Accounting Horizons*, 29(4), 757–776. doi:10.2308/acch-51141

Freedman, M., Park, J., & Romero, J. (2013). *Recognizing Environmental Liabilities Surrounding CEO Turnover* (Vol. 16). *Advances in Public Interest Accounting*.

Mazur, O. (2015). Tax Challenges in the Cloud. *Journal of Internet Law*, 19(4), 3–7.

Romero, J., & Cong, Y. (2013). On Information System Complexity and Vulnerability. *Journal of Information Systems*, 27(2), 51–64. doi:10.2308/isis-50562

Sabherwal, R., & Jeyaraj, A. (2015). Information Technology Impacts on Firm Performance: An Extension of Kohli and Devaraj (2003). *Management Information Systems Quarterly*, 39(4), 809–836.

Wagner, H., Beimborn, D., & Weitzel, T. (2014). How Social Capital Among Information Technology and Business Units Drives Operational Alignment and IT Business Value. *Journal of Management Information Systems*, 31(1), 241–272. doi:10.2753/MIS0742-1222310110

KEY TERMS AND DEFINITIONS

Business Value: It is the value of a firm that not only includes the monetary value of a firm, the hardware, the software, and the IT organizational processes but also includes other intangible components such as innovation, integration, standardization of processes, efficiency, and quality of products produced.

Cloud-Based Applications: Applications that do not need to be installed or run from a desktop or server on premise in a firm (Choudhary and Vithayathil, 2013). They are applications that are hosted on remote servers and that can be accessed via the Internet from anywhere. It offers an advantage to firms avoiding the acquisition of IT assets that usually depreciate at a fast rate. Because the software is not installed on site, firms do not need to invest on upgrading software regularly or to pay annual license fees. However, there is a pay as-you-go fee in cloud applications (Choudhary and Vithayathil, 2013).

Customer Relationship Management: Also known as CRM. CRM with the use of technology supports a firm managing its interactions with customers and focuses on customer retention with the purpose of maximizing profitability (Weinman, 2015). It is mostly considered a business and strategic solution rather than a software solution. CRM integrates different parts of a company that deal with customers such as: sales, advertising, direct mail, email, and call centers in order to improve customer communication and to increase competitiveness. With the use of CRM is possible to identify customer preferences. There have been case studies of companies failing to extract benefits from CRM installations mostly due a poor understanding of the system and not an adequate training of employees using the system (Chen and Popovich, 2003). For instance, Salesforce, headquartered in San Francisco, is a cloud-based CRM package and is currently one of the market leaders that includes the following components: marketing cloud, analytics cloud, sales cloud, and service cloud.

Enterprise Resource Planning: Also known as ERP and was developed in the 1990's. An ERP system attempts not only to integrate all departments within a company but also to synchronize the ERP system of the firm with the ERP systems of others companies in the supply chain that it operates (Anderson et al., 2011). An ERP system consists of several modules such as finance, sales, human resources, inventory, operations, and marketing among others. An ERP system is deemed complete when all modules are installed, and it is preferably to install all modules from the same vendor in order to avoid any compatibility issues (Romero et al., 2010). Having all departments synchronized let a company know the status of an

order at any point in time, and if there are changes in an order, all systems are updated automatically without having a need to update the systems of each department one by one.

Information Technology: Information technology includes the use of computer software, computer hardware, telecommunications equipment and services. It provides support to firms to store data, process data, synchronize and monitor processes. It also allows managers to think at a more strategic level because they do not need to spend time or resources on repetitive or administrative tasks that are being performed by the information technology (Romero et al., 2010).

Material Requirements Planning II: Also known as MRP II and was developed in 1980's. MRP II evolved from MRP incorporating some additional features such as the master production schedule using the available-to-promise logic, sales forecasting,, and tracking tools until the items are delivered to customers among other tools (Olhager, 2013).

Material Requirements Planning: Also known as MRP. MRP was developed in the 1960's and 1970's as a production planning system to support manufacturing operations calculating the amount of raw materials needed by the production department and the specific timing of different individual parts that are needed (Jodlbauer, 2012). MRP also helps managers decide when to buy.

Supply Chain Management: Also known as SCM. SCM is a group of different firms that work in a coordinated way and are linked together to maximize competitive advantage. It is composed of suppliers of raw materials, firms that process those raw materials, and retailers that sell the products to the final customers.

Using Business Analytics in Franchise Organizations

Ye-Sho Chen

Louisiana State University, USA

INTRODUCTION

Franchising has been a popular approach to growing a business (Justis & Judd, 2002). Its popularity continues to increase, as we witness an emergence of a new business model, Netchising, which is the combination of the Internet for global demand-and-supply processes and the international *franchising* arrangement for local responsiveness (Chen, Justis, & Yang, 2004; Chen, Chen, & Wu, 2005, 2007; Chen & Wu, 2007; Chen, Liu, Zeng, & Azevedo, 2012).

BACKGROUND

In his best seller, *Business @ the Speed of Thought*, Bill Gates (1999) wrote: "Information Technology and business are becoming inextricably interwoven. I don't think anybody can talk meaningfully about one without talking about the other." (p. 6) Gates' point is quite true when one talks about franchise's use of big data and business analytics, which is "delivering the right decision support to the right people at the right time" (Laursen & Thorlund, 2010, p. xii). Thus, to see how big data and business analytics can be "meaningfully" used in franchise organizations, one needs to know how franchising really works. In this paper, we show that building up a good "family" relationship between the franchisor and the franchisee is the real essence of franchising, and proven working knowledge is the foundation of the "family" relationship. Specifically, we use the following seven pillars of business analytics (Isson & Harriott, 2013) to discuss the process

of how to make business analytics "meaningful" in franchising: business challenges, data foundation, analytics implementation, insights, execution and measurements, distributed knowledge, and innovation.

BUSINESS CHALLENGES: MANAGING THE FRANCHISOR AND FRANCHISEE RELATIONSHIP

Franchising is "a business opportunity by which the owner ... grants exclusive rights to an individual for the local distribution... The individual or business granting the business rights is called the *franchisor*, and the individual or business granted the right to operate ... is called the *franchisee*." (Justis & Judd, 2002, pp. 1-3) Developing a good "family" relationship between the franchisor and the franchisee is the key business challenge of a successful franchise (Justis & Judd, 2002). Figure 1 describes how such a "family" relationship is built in the franchise business community. In the figure, it shows that the franchise system is operated in the dynamic business environment of global, national, regional, and local communities. The "family" relationship is developed through a mutual learning process of person-centric relationship building.

The franchisor's learning process is incrementally developed through five stages (Justis & Judd, 2002): Beginner – learning how to do it; Novice – practicing doing it; Advanced – doing it; Master – teaching others to do it; and Professional – becoming the best that you can be. Once attaining the advanced stages of development,

Figure 1. Understanding how the franchisor/franchisee “family” relationship works



most preceding struggles have been overcome. However, further convoluted and challenging enquiries will arise as the franchise continues expansion. This is especially true once the system reaches the “Professional” stage, where various unpredicted and intricate problems could arise. To capture the learning process, a counter-clockwise round arrow surrounding the franchisor is used to depict the increasing intensity of learning as the franchisor continues to grow.

To understand how the “family” relationship is developed, one needs to know the five phases of franchisee life cycle (Schreuder, Krige, & Parker, 2000): Courting: both the franchisee and the franchisor are eager with the relationship; “We”: the relationship starts to deteriorate, but the franchisee still values the relationship; “Me”: the franchisee starts to question the reasons for payments related issues with the attitude that the success so far is purely of his/her own work; Rebel: the franchisee starts to challenge the restrictions being placed upon; and Renewal: the franchisee realizes the “win-win” solution is to continue teaming up with the franchisor to grow the system. Similar to the franchisor, a counter-clockwise round arrow surrounding the franchisee is used in Figure 1 to depict the increasing intensity of franchisee life cycle as the franchisee continues learning and growing.

As the franchisee progresses through the life cycle, the “family” relationship gradually develops a mutually influencing process (Justis & Vincent, 2001) as depicted in Figure 1 with a bi-directional arrow: working knowledge, proven abilities of expanding the franchise system profitably; positive attitude, constructive ways of presenting and sharing the working knowledge; good motivation, providing incentives for learning or teaching the working knowledge; positive individual behavior, understanding and leveraging the strengths of the participants to learn and enhance the working knowledge; and collaborative group behavior, having the team spirit to find the best way to collect, disseminate, and manage the hard-earned working knowledge. By going through the processes of learning and influencing, both the franchisor and the franchisee gain the progressive working knowledge in the franchise business community. The franchisor, the franchisee, and the franchise business community in Figure 1 are surrounded with dashed lines, indicating that there is no limit to the learning process.

DATA FOUNDATION: MANGING FRANCHISE BIG DATA

There are many “touchpoints” within the franchise business community where the franchisor and

Table 1. The customer-service-life-cycle model in franchising

CSLC	Sub-Stages	Example: WSI Internet
Requirements	Understanding How Franchising Works	Internet • WSI Proven Franchise Business Model • WSI Franchise Lifecycle • Master Franchise Opportunities
	Investigating Franchise Opportunities	Internet • About WSI • Client Portfolio
	Obtaining Franchisee Prospectus	Internet • E-mails
	Making the Choice	Internet • Global Franchise Leader • Success Stories • WSI in the News
Acquisition	Preparing Business Plan	Internet
	Financing the Franchised Business	Internet
	Signing the Contract	Internet • Own a WSI Franchise
Ownership	Marketing & Promoting the Franchise Products or Services	Internet/Intranet/Extranet • WSI Proven System
	Managing the Franchise System	Internet • WSI Community Outreach Intranet • Franchise Training and Support • Serving Franchisee's Customers Extranet
	Building the Relationship between the Franchisor and the Franchisee • The Courting Phase • The "We" Phase • The "Me" Phase • The Rebel Phase • The Renewal Phase	Internet Intranet • Knowledge Centre • Training at Headquarters • Newsletter • Meetings • Toll-free Phone Line Extranet • Purchasing Cooperative
Renewal or Retirement	Becoming a Professional Multi-unit Franchisee or Retiring from the Franchise System	More refined and diverse views of the customer information are gathered through Internet, Intranet, and Extranet at the different stages of "touchpoints". This will help the franchise system look into the customer insights (Suther, Burkart, & Cheng, 2013) and better serve the customers.

the franchisee can influence each other. Based on the Customer Service Life Cycle (CSLC) model, Chen, Chong, & Justis (2002) proposed a framework (Table 1) to harness the Internet to develop the big data foundation needed to serve its customers, i.e., franchisees and their customers. Three exemplar types of data are (1) unit operational data: the daily activities at the franchisor headquarters; (2) external benchmarking data: the relationship management activities in the

franchise community; and (3) business legacy data: the activities that have been working well or gradually adapted since the franchise system came into existence.

To tap the full potential of the CSLC framework, more refined and diverse views of the customer information shall be gathered at the different stages of "touchpoints". This will help the franchise system look into the customer insights (Suther, Burkart, & Cheng, 2013) and better serve

the customers. A well-designed Internet strategy, often enabled by Application Service Providers (Chen, Ford, Justis, & Chong, 2001), shall empower the franchisor and the franchisees to collect, use, renew, store, retrieve, transmit, and share the organizational data needed to do the collaborative work in the various phases of the CSLC model.

ANALYTICS IMPLEMENTATION: MANAGING FRANCHISE ORGANIZATIONAL INFORMATION

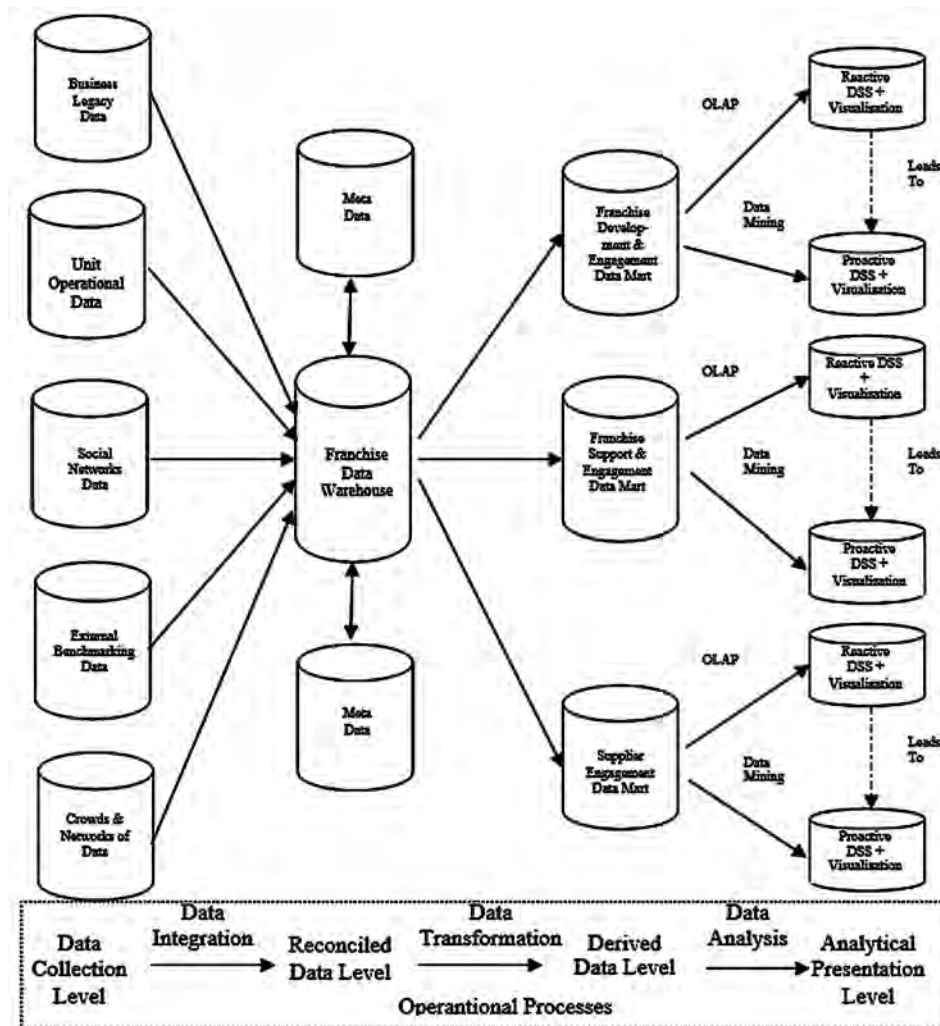
An architecture, adapted from Inmon (1996) and Thomas & McSharry (2015), of big data and business analytics in franchising is shown in Figure 2. The architecture consists of four levels of operational processes: (1) data collection level, holding operational, external, and legacy data collected from the franchise business environment depicted in Figure 1; (2) reconciled data level, holding data warehouse data and meta data; (3) derived data level, containing several data marts derived from the data warehouse based on various franchisee/customer-centered segmentations; and (4) the analytical presentation level, producing various relationship performance indicators with strong data visualization capabilities for decision making via decision support systems (DSS). To move from the data collection level to the reconciled data level, data integration is needed. It is a very time consuming process that involves the activities such as recovery, cleansing, extracting, filtering, conditioning, scrubbing, and loading. To move from the reconciled data level to the derived data level, data transformation is needed which involves the activities such as exploration, replication, propagation, summary, aggregate, and metadata. To move from the derived data level to the analytical presentation level, data analysis is needed which involves activities such as online analytical processing (OLAP) and data mining (Chen, Zhang, & Justis, 2005; Chen, Justis, & Chong, 2008).

A typical OLAP analysis consists of pre-defined multi-dimensional queries such as (Chen, Justis, & Watson, 2000): (1) Show the gross margin by product category and by franchise outlets from Thanksgiving to Christmas in the last five years; (2) Which franchise outlets are increasing in sales and which are decreasing? And (3) Which kinds of customers place the same orders on a regular basis at certain franchise outlets? Other OLAP activities include spread sheet analysis, data visualization, and a variety of statistical data modelling methods. Since the query activities are pre-defined, we call the supporting systems reactive DSS.

Data mining is used to identify context-sensitive patterns of the data residing in the data marts. Typical data mining modelling analysis can be classified into the following three categories: Classification and Prediction, using techniques such as RFM (recency, frequency, and monetary), regression, decision tree, and neural network; Association Rules, using techniques such as market basket analysis, correlation analysis, cross-sell analysis, and link analysis; and Cluster Analysis, using techniques such as partition, hierarchy, outlier, and density analysis. Table 2, adapted from Delmater & Hancock (2001), shows that data mining techniques can be used to help serve franchisees' customers at the different stages of the CSLC model.

Since the data mining queries and related activities are not pre-defined, we call the supporting systems proactive DSS. A major drawback of proactive data mining is the fact that without vigilant preliminary examination of data characteristics, the mining activities may end in vain (Delmater & Hancock, 2001). In order to achieve higher success rate of data mining, we suggest (on the right side of Figure 2) that OLAP-based queries need to be conducted first. For example, one may find, through daily OLAP queries, that certain segments of customers buy certain products frequently. This pattern may lead to perform thorough and proactive analysis of the customer-product relationship and human resource analytics (Pease, Byerly, & Fitz-enz, 2012). The results may

Figure 2. Architecture of big data and business analytics in franchising



help the company provide legendary services to its clients and generate higher profits.

INSIGHTS: MANAGING FRANCHISE ORGANIZATIONAL WORKING KNOWLEDGE

As mentioned in the discussions of Figure 1, the key for building the franchisor/franchisee “family” relationship is in the franchise organizational learning. In addition, there are five vital insights for a successful learning program: knowledge, attitude, motivation, individual behavior, and

group behavior. Thus, working knowledge is the real foundation of a successful franchise “family” relationship. The working knowledge is structured in many forms of profiles that are embedded in the operational manuals of the franchise business processes. Table 3 gives some examples of those working knowledge profiles with respect to the CSLC business processes associated with the sub-stages in Table 1.

A working knowledge profile is developed when a certain task of the CSLC process is repeated several times with superior results. Consider the Site Profile used at the “Marketing & Promoting the Franchise Products/Services” sub-stage in

Table 2. Franchisees' customers data mining using the CSLC approach

CSLC	Explanation	Data Mining (and Techniques Used) for Context-Sensitive Patterns
Requirements	Finding and reaching the customers	Context-Sensitive Lead Generation • Market Analysis & Segmentation (Classification and Prediction) • Mining Web Site Visitors (Association Rules) • Text Mining Usenet Newsgroups (Cluster Analysis)
Acquisition	Selling to the customers	Context-Sensitive Customer Acquisition Profiling • Customer Segmentation Strategy (Classification and Prediction) • Online Shopping Tracking (Association Rules) • Pricing Strategy (Association Rules) • Customer-centric Selling (Association Rules) • Text Mining Contact E-Mails (Cluster Analysis) • Scenario Notification (Association Rules)
Ownership	Satisfying the customers after the sales	Context-Sensitive Customer Service • Inquiry Routing (Association Rules) • Text Mining E-Mails & Inquiries (Cluster Analysis) • Scenario Notification (Association Rules) • Staffing Level Prediction (Classification and Prediction)
Retirement or Renewing	Retaining the customers so that you can continue coming back	Context-Sensitive Customer Retention • Sharper Customer Focus through Loyalty Program (Classification and Prediction) • Detecting Customer Complaints through Text Mining (Cluster Analysis) • Detecting Inappropriate Customer Services (Cluster Analysis) • Individual Customer Profiles (Classification and Prediction) • Scenario Notification (Association Rules)

Table 3. The Site Profile is used to assist the new franchisee locate a top business site. Typically it is the real estate department at the franchisor headquarters responsible for the profile development. The Site Profile is unremittingly being tested and enhanced. Various OLAP/Data Mining analytical reports, monitoring the performance of the sites, are generated at the Analytical Presentation Level shown in Figure 2. Based on those reports, the real estate experts and their teams are able to fine-tune the attributes and the parameters within the Site Profile. Most often, the corresponding data collection procedures in the CSLC sub-stage also need to be revised and perfected so that better scorecards can be generated.

EXECUTION AND MEASUREMENTS: DEVELOPING A GROWING FRANCHISE

This process of executing and enhancing the working knowledge profiles will achieve its high peak when both the franchisor and the franchisees are

Table 3. The CSLC model of franchise working knowledge

CSLC Sub-stages	Examples of Working Knowledge Profiles
Understanding How Franchising Works	Lead Generation Profiles Website Visitor Profile
Investigating Franchise Opportunities	Benchmarking Profile Successful Franchisee Profile
Obtaining Franchisee Prospectus	Prospectus Profile
Making the Choice	Competitor Profile
Preparing Business Plan	Business Plan Profile
Financing the Franchised Business	Financing Institute Profile Non-traditional Franchising Profile
Signing the Contract	Franchisee Profile
Marketing & Promoting the Franchise Products or Services	Site Profile Product Profile
Managing the Franchise System	Support Team Profile Employee Profile Demand-Supply Matching Profile
Building the Relationship between the Franchisor and the Franchisee	Event Management Profile Best Practices Profile
Becoming a Professional Multi-unit Franchisee or Retiring from the Franchise System	Multi-unit Franchisee Profile Social Networks Profile

arriving at the Professional and Renewal stage of growth. A significant phenomenon of being a Professional franchisor and a Renewal franchisee are their ability to leverage the assets of the hard-earned working knowledge profiles into dynamic capabilities and high-business-value-creation complete-advantage strategies (Chen, Yuan, & Dai, 2004; Chen, Seidman, & Justis, 2005). Business performance tracking, using mobile analytics (Munteanu & Puican, 2012) with effective analytics communication strategies, is very important here. The new products or services coming out of the process of leveraging the working knowledge profiles may transform the franchise business into a more, sometimes surprisingly, profitable enterprise.

The capability of leveraging the assets of franchise working knowledge into profitable products or services is at the heart of a successful franchise. For instance, consider the site selection working knowledge at McDonald's. The Franchise Realty Corporation real estate business, a result of site selection asset leveraging, is the real moneymaking engine at McDonald's. This as can be evidenced from the following speech of Ray Kroc, founder of McDonald's, to the MBA class at the University of Texas at Austin in 1974: "... I 'm not in the hamburger business. My business is real estate." (Kiyosaki, 2000, p.85) In the book *McDonald's: Behind the Arches* (Love, 1995, p. 152), Ray Kroc commented further: "What converted McDonald's into a money machine had nothing to do with ... the popularity of McDonald's hamburgers, French fries, and milk shakes. Rather, McDonald's made its money on real estate" McDonald's makes money out of real estate by leasing properties from landlords and then subleasing the stores to the franchisees. The Professional franchisees, many of them are multiunit operators, can then focus on expanding the business without worrying about finding good locations for the growth. This moneymaking real estate strategy is what separates McDonald's from other fast-food chains.

DISTRIBUTED KNOWLEDGE: FRANCHISE KNOWLEDGE REPOSITORY

Knowledge repository systems, consisting of working knowledge profiles such as the one shown in Table 4 (Chen, Chong, & Justis, 2000; Chen, Hammerstein, & Justis, 2002; Chen, Justis, & Wu, 2008), can be linked into the franchisor headquarters and the franchisee outlets for knowledge sharing and learning. Such a repository has two dimensions. First, there is a working knowledge level for the collaborative team, the franchisee outlet, the franchisor headquarters, and the franchise community. Second, there are user skill levels, including Beginner in the Courting Phase, Novice in the "We"-Phase, Advanced in the "Me"-Phase, Master in the Rebel Phase (since the rebel ones tend to be those who know the system very well and are capable of influencing others to follow them), and Professional in the Renewal Stage of franchisee life cycle. The foundation of the framework is the working knowledge of the five crucial elements—Knowledge, Attitude, Motivation, Individual Behavior, and Group Behavior—used by the collaborative team, to effectively influence others in building the franchise "family" relationship. The working knowledge profiles at the franchisee outlet, the franchisor headquarters, and the franchise community can be modularized according to user's level. An Intranet-based curriculum of working knowledge modules can then be designed for the users to learn the working knowledge profiles effectively.

FUTURE RESEARCH DIRECTIONS: INNOVATION BY LEVERAGING FRANCHISE VALUE NETWORKS

The third industrial revolution, combining Internet technology with globalization, produces various new big data and business analytics opportunities

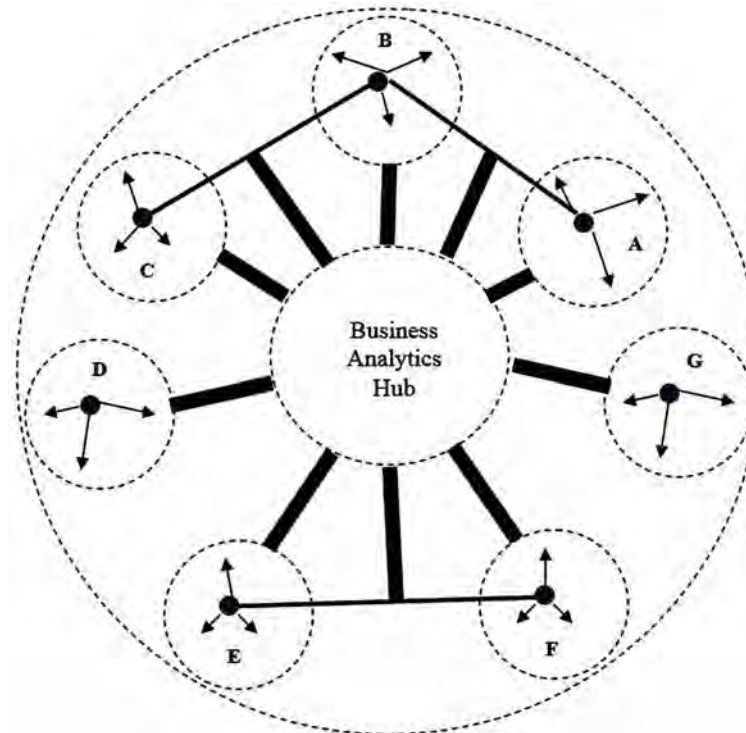
Table 4. Working knowledge repository in franchise organizations

		User Skill Levels				
		Beginner in the Courting Phase: Beginner Guide	Novice in the “We”-Phase: Practicing	Advanced in the “Me”-Phase: Doing	Master in the Rebel Phase: Teaching Others	Professional in the Renewal Stage: Improving and Leveraging
Working Knowledge Levels	Collaborative Team	Culture and Process of Influencing Others for Knowledge Sharing: Knowledge, Attitude, Motivation, Individual Behavior, and Group Behavior				
	Franchisee Outlet	Learning and Training of Working Knowledge Profiles for Running the Franchisee Outlet: Customer Profile, Employee Profile, Product Profile				
	Franchisor Headquarters	Learning and Training of Working Knowledge Profiles for Running the Franchisor Headquarters: Franchisee Profile, Site Profile, Product Profile, Employee Profile, Event Management Profile				
	Franchise Community	Learning and Training of Working Knowledge Profiles for Relationship Management with the Community: Supplier Profiles, Community Profiles				

for the innovative growth of franchise organizations. For example, value network applications, using business analytics techniques such as social network analysis, can be developed to connect Professional franchisees in the world. The goal is to enable the franchise system to venture into new global emerging markets, e.g., China, through international franchising and develop innovative

products/services through asset leveraging. This could be done because franchise capabilities, structured in the working knowledge repository shown in Table 4, empower the Professional franchisees to work with the franchisor to continuously improve and leverage the current franchise working knowledge. An example of value networks of Professional franchisees can be illustrated in Figure 3.

Figure 3. Business analytics hub enhances and strengthens the value networks of professional franchisees



There are seven Professional franchisees (A - G) in the figure with four clusters (A-C, D, E-F, and G) of value networks. Each Professional franchisee (a dot) has his/her personal value network (arrows pointing out of the dot) tested and built over the years while doing day-to-day problems solving at the franchisee outlet. The value network may include the customers' likes and dislikes, the kind of employees to hire, the competitors' and suppliers' pricing strategies, the social needs in the local community. Each Professional franchisee is surrounded with a circle with dashed lines, meaning there is no limit to the personal value network. In order to solve the problems more effectively, Professional franchisees may share with each other their approaches. Thus, clusters (connected dots) of value network are formed for solving various problems more effectively (Chen, Justis, & Wu, 2006). The big data gathered from the Professional franchisees and networks are connected to the business analytics hub to enhance and strengthen the value networks.

CONCLUSION

Franchising has been popular as a growth strategy for small businesses; it is even more so in today's global and e-commerce world (Chen, Chen, & Wu, 2005). The essence of franchising lies in managing the "family" relationship between the franchisor and the franchisee. In this paper we showed business analytics plays an important role in growing and nurturing such a "family" relationship. Specifically, we used the seven pillars of business analytics (Isson & Harriott, 2013) to discuss the process of how to make big data and business analytics "meaningful" in franchising: how franchise big data can be managed effectively using the methodology of Customer Service Life Cycle; how franchise organizational information is deciphered from the customer-centered data using business analytical techniques such as OLAP and data mining; and how the franchise

organizational working knowledge is leveraged to grow the franchise system. The ability to continue creating value networks by leveraging the organizational working knowledge assets based on the good "family" relationship is really what a franchise business is about.

REFERENCES

- Chen, Y., Chen, G., & Wu, S. (2005). Issues and opportunities in e-business research: A Simonian perspective. *International Journal of E-Business Research*, 1(1), 37–53. doi:10.4018/jebr.2005010103
- Chen, Y., Chen, G., & Wu, S. (2007). A Simonian approach to e-business research: A study in netchising. In I. Lee (Ed.), *E-business innovation and process management* (pp. 133–164). Hershey, PA: Idea Group. doi:10.4018/978-1-59904-277-0.ch007
- Chen, Y., Chong, P., & Justis, R. T. (2000). Franchising knowledge repository: A structure for learning organizations. *Proceedings of the 14th Annual International Society of Franchising Conference*.
- Chen, Y., Chong, P. P., & Justis, R. T. (2002). E-business strategy in franchising: A customer-service-life-cycle approach. *Proceedings of the 16th Annual International Society of Franchising Conference*.
- Chen, Y., Ford, C., Justis, R. T., & Chong, P. (2001). Application service providers (ASP) in franchising: Opportunities and issues. *Proceedings of the 15th Annual International Society of Franchising Conference*.
- Chen, Y., Hammerstein, S., & Justis, R. T. (2002). Knowledge, learning, and capabilities in franchise organizations. *Proceedings of the 3rd European Conference on Organizational Knowledge, Learning, and Capabilities*.

- Chen, Y., Justis, R., & Watson, E. (2000). Web-enabled data warehousing. In M. Shaw, R. Blanning, T. Strader, & A. Whinston (Eds.), *Handbook of electronic commerce* (pp. 501–520). Springer-Verlag. doi:10.1007/978-3-642-58327-8_24
- Chen, Y., Justis, R., & Wu, S. (2006). Value networks in franchise organizations: A study in the senior care industry. *Proceedings of the 20th Annual International Society of Franchising Conference*.
- Chen, Y., Justis, R., & Wu, S. (2008). Skew distributions and the sizes of franchise firms: A study in the senior care industry. *Proceedings of the 22nd Annual International Society of Franchising Conference*.
- Chen, Y., Justis, R. T., & Chong, P. P. (2002). Franchising and information technology: A framework. In S. Burgess (Ed.), *Managing information technology in small business: Challenges and solutions* (pp. 118–139). Hershey, PA: Idea Group Publishing.
- Chen, Y., Justis, R. T., & Chong, P. P. (2008). Data mining in franchise organizations. In J. Wang (Ed.), *Data warehousing and mining: Concepts, methodologies, tools, and applications* (pp. 2722–2733). Hershey, PA: Idea Group Publishing. doi:10.4018/978-1-59904-951-9.ch169
- Chen, Y., Justis, R. T., & Yang, H. L. (2004). Global e-business, international franchising, and theory of netchising: A research alliance of east and west. *Proceedings of the 18th Annual International Society of Franchising Conference*.
- Chen, Y., Liu, C., Zeng, Q., & Azevedo, R. F. L. (2012). E-business strategy in franchise relationship management. *International Journal of Strategic Technology and Applications*, 3(1), 55–67.
- Chen, Y., Seidman, W., & Justis, R. (2005). Strategy and docility in franchise organizations. *Proceedings of the 19th Annual International Society of Franchising Conference*.
- Chen, Y. & Wu, S. (2007). E-business research in franchising (Invited editorial preface). *Information Management*, 20(1/2), 9.
- Chen, Y., Yuan, W., & Dai, W. (2004). Strategy and nearly decomposable systems: A study in franchise organizations. *International Symposium on "IT/IS Issues in Asia-Pacific Region, Co-sponsored by ICIS-2004"*.
- Chen, Y., Zhang, B., & Justis, R. T. (2005). Data mining in franchise organizations. *Encyclopaedia of Information Science and Technology*, 714–722.
- Delmater, R., & Hancock, M. (2001). *Data mining explained: a manager's guide to customer-centric business intelligence*. Digital Press.
- Gates, W. (1999). *Business @ the speed of thought*. Warner Books.
- Inmon, W. H. (1996). *Building the data warehouse*. John Wiley & Sons.
- Isson, J., & Harriott, J. (2013). *Win with advanced business analytics: Creating business value from your data*. Wiley and SAS Business Series.
- Justis, R. T., & Judd, R. J. (2002). *Franchising*. DAME Publishing.
- Justis, R. T., & Vincent, W. S. (2001). *Achieving wealth through franchising*. Adams Media Corporation.
- Kiyosaki, R. (2000). *Rich dad, poor dad*. Time Warner.
- Laursen, G. H. N., & Thorlund, J. (2010). *Business analytics for managers: Taking business intelligence Beyond Reporting*. Wiley. doi:10.1002/9781118983812
- Love, J. (1995). *McDonald's: behind the arches*. Bantam Books.
- Munteanu, A. and Puican, F. (2012). Mobile analytics: Modern decision support systems. *Revista Economica - Applied Informatics in Economy*, 5, 139–146.

Pease, G., Byerly, B., & Fitz-enz, J. (2012). *Human capital analytics: How to harness the potential of your organization's greatest asset*. Wiley. doi:10.1002/9781119205050

Schreuder, A. N., Krige, L., & Parker, E. (2000). The franchisee lifecycle concept—A new paradigm in managing the franchisee-franchisor relationship. *Proceedings of the 14th annual International Society of Franchising Conference*.

Suther, T., Burkart, B., & Cheng, J. (2013). Jack and the big data beanstalk: Capitalizing on a growing marketing opportunity. In *Big data and business analytics*. CRC Press.

Thomas, R., & McSharry, P. (2015). *Big Data Revolution*. Wiley.

ADDITIONAL READING

Alon, I., Boulanger, M., Misati, E., & Madanoglu, M. (2015). Are the parents to blame? Predicting franchisee failure. *Competitiveness Review*, 25(2), 205–217. doi:10.1108/CR-10-2014-0034

Balaji, K., & Viswanadham, N. (2008). A tax integrated approach for global supply chain network planning. *Automation Science and Engineering. IEEE Transactions on*, 5(4), 587–596.

Ball, S. J. (2012). *Global Policy Networks, Social Enterprise and Edu-business*. Routledge.

Bell, P., & Zaric, G. (2012). *Analytics for Managers: With Excel*. Routledge.

Bürkle, T., & Posselt, T. (2008). Franchising as a plural system: A risk-based explanation. *Journal of Retailing*, 84(1), 39–47. doi:10.1016/j.jretai.2008.01.004

Cantoni, L., Faré, M., Inversini, A., & Passini, V. (2011, January). Hotel Websites and Booking Engines: A Challenging Relationship. In *ENTER* (pp. 241–252). doi:10.1007/978-3-7091-0503-0_20

Dant, R. P., & Berger, P. D. (1996). Modelling cooperative advertising decisions in franchising. *The Journal of the Operational Research Society*, 47(9), 1120–1136. doi:10.1057/jors.1996.141

Fajardo, J. F., Faulin, J., Juan, A. A., Grasman, S. E., & Fry, M. J. (Eds.). (2012). *Decision Making in Service Industries: A Practical Approach*. CRC Press.

Ghosh, S. (2005). When Exclusionary Conduct Meets the Exclusive Rights of Intellectual Property: *Morris v. PGA Tour* and the Limits of Free Riding as an Antitrust Business Justification. *Loy. U. Chi. LJ*, 37, 723.

Girod, S. J., & Rugman, A. M. (2005). Regional business networks and the multinational retail sector. *Long Range Planning*, 38(4), 335–357. doi:10.1016/j.lrp.2005.04.010

Hoq, M. Z. (2011). The Implication of Data Mining in Business. *The AUIB Faculty Journal*, 2(3), 2010–2011.

Hsu, J. (2004). *Data mining and business intelligence: Tools, technologies, and applications*. IGI Global. doi:10.4018/978-1-59140-206-0.ch009

Kim, S., Koh, Y., Cha, J., & Lee, S. (2015). Effects of social media on firm value for US restaurant companies. *International Journal of Hospitality Management*, 49, 40–46. doi:10.1016/j.ijhm.2015.05.006

Meiseberg, B., & Ehrmann, T. (2012). Lost in Translation? The Prevalence and Performance Impact of Corporate Social Responsibility in Franchising. *Journal of Small Business Management*, 50(4), 566–595. doi:10.1111/j.1540-627X.2012.00367.x

Merz, C., Friedland, C., de Louw, R., Dörflinger, J., Maritz, J., van Rensburg, J., & Naudé, A. (2010). Sekhukhune: A Living Lab Stimulating Economic Growth of Rural Micro-Enterprises in South Africa. *Living Labs for Rural Development*, 143.

Odlyzko, A. (2007, August). Privacy and the clandestine evolution of e-commerce. In *Proceedings of the ninth international conference on Electronic commerce* (pp. 3-6). ACM. doi:10.1145/1282100.1282104

Pandit, M., Bhatia, V., & Deshmukh, S. (2010). *U.S. Patent No. 7,689,557*. Washington, DC: U.S. Patent and Trademark Office.

Park, S. Y., & Lee, S. (2009). Financial Rewards for Social Responsibility A Mixed Picture for Restaurant Companies. *Cornell Hospitality Quarterly*, 50(2), 168–179. doi:10.1177/1938965509331814

Potocan, V., Nedelko, Z., & Mulej, M. (2011). What is New with Organization of E-Business: Organizational Viewpoint of the Relationships in E-business. E-business issues, challenges and opportunities for SMEs. Hershey, New York, 131-148.

Ramakrishna, H. (2011). *Business analytics success: A conceptual framework and an application to virtual organizing*. IGI Global. doi:10.4018/978-1-60960-587-2.ch111

Ronca, S. F. (2013). *The Best Social Media Practices for an Individual Franchisee in the Restaurant Industry* (Doctoral dissertation, California Polytechnic State University).

Strang, K. D. (2008). Collaborative synergy and leadership in e-business. *Handbook of research on electronic collaboration and organizational synergy*, 409-429.

Taylor, M. (2012). Routledge handbook of sport management. *European Sport Management Quarterly*, 12(5), 569–573. doi:10.1080/16184742.2012.728237

Tian, K., & Borges, L. (2011). Cross-Cultural Issues in Marketing Communications: An Anthropological Perspective of International Business. *International Journal of China Marketing*, 2(1), 110–126.

Zarkada, A. K., & Polydorou, C. (2013). You Might be Reputable but are you Liked? Orchestrating Corporate Reputation Co-Creation on Facebook. *Advanced Series in Management*, 11, 87–113. doi:10.1108/S1877-6361(2013)0000011009

KEY TERMS AND DEFINITIONS

Customer Service Life Cycle: Serving customers based on a process of four stages: Requirements, Acquisition, Ownership, and Retirement. Many companies are using the approach to harness the Internet to serve the customers.

Data Mart: A small database with data derived from a data warehouse.

Data Mining: Analytical techniques used to find out the hidden relationships or patterns residing in the organizational data.

Data Warehouse: A database which is subject-oriented, integrated, time-variant, and non-volatile.

Franchisee Life Cycle: The stages a franchisee goes through in the franchise system: Courting, “We”, “Me”, Rebel, Renewal.

Franchisor/Franchisee Learning Process: The stages of learning, including Beginner, Novice, Advanced, Master, and Professional.

Franchisor/Franchisee Relationship Management: The vital factor for the success of a franchise, including: Knowledge, Attitude, Motivation, Individual Behavior, and Group Behavior.

Category B

Business Intelligence

Big Data, Knowledge, and Business Intelligence

B

G. Scott Erickson
Ithaca College, USA

Helen N. Rothberg
Marist College, USA

INTRODUCTION

After developing the scholarly foundation of the existing disciplines, this paper will look specifically at how selected concepts relate across the fields, particularly to what we know about big data. Although not as developed a discipline, big data does have recognizable elements from its own and other literatures. This article will establish the links between fields and demonstrate opportunities for sharing and learning between the different disciplines, both old and new.

BACKGROUND

The field of knowledge management (KM) and its related discipline, intellectual capital (IC) have both contributed considerably to our understanding of the value of intangible assets of the firm. The general concept that intangibles are something worthwhile goes back at least to Schumpeter's (1934) work on innovation and has since included contributions from other high-profile writers such as Drucker, with his knowledge workers (1991). The idea that proper management of such intangibles might lead to competitive advantage was explored by scholars such as Nelson and Winter (1982) in their evolutionary theory of growth. Such competitive advantage fits well with the resource-based theory of the firm (Wernerfelt, 1984), specifically identifying knowledge as a potential key resource. As a result, we have the

knowledge-based theory of the firm (Teece, 1998; Grant, 1996) and its suggestion that knowledge is not only a potentially important differentiator but perhaps the only differentiator for firms looking for sustainable competitive advantage.

In pushing the field forward, the KM and IC disciplines have always carefully defined the nature of their study, perhaps because of the obvious need to clarify and distinguish knowledge assets or intellectual capital from the more widely known but explicitly formal intellectual property terminology. Patents, copyrights, and other intellectual property are valuable intangible assets, but intellectual capital extends our recognition of value to additional, less well-defined intangibles such as know-how and expertise (knowledge). As a result, a clear distinction exists in the field between data, information, and knowledge. Formally, data are observations, information is data in context, and knowledge is information subjected to experience, reflection, or some similar analysis (Zack, 1999b). Within the field, knowledge is often referred to as know-how, effectively a sort of learning based on experience, learning, or insight. Such a perspective flows naturally out of the more general DIKW (data, information, knowledge, wisdom) hierarchy suggested by Ackoff (1989).

Growth in knowledge of this sort can come about in different ways. Nonaka and Takeuchi (1995) borrowed the concept of tacit knowledge from the sociology literature (Polanyi, 1967), identifying and explaining tacit and explicit knowledge in a business context. Tacit knowledge

is more personal, harder to express, and more difficult to codify within organizational information technology (IT) systems. Explicit knowledge, on the other hand, is easier to express, easier to share, and easier to store in IT structures. Nonaka & Takeuchi also developed the SECI or ba framework categorizing how knowledge grows, by tacit to explicit, tacit to tacit, explicit to tacit, or explicit to explicit transfer. The explicit to tacit process is of particular interest as it concerns the conversion of more structured intangible assets into personal tacit insights. From there, it is only a short step to the idea of creating new knowledge from data and information, foreshadowing how non-knowledge intangible assets can also create value. The overall objective of KM is to better understand how knowledge can be more effectively developed and employed by means of combination, sharing, learning, or similar means (Zack, 1999a; Grant, 1996).

Since these early insights, knowledge management as a field has focused more on the circumstances surrounding knowledge development as well as appropriate tools. Beyond the tacit/explicit distinction, other circumstantial variables relating to knowledge development include other aspects of the knowledge itself as well as organizational conditions. Knowledge aspects include characteristics such as complexity and stickiness (McEvily & Chakravarthy, 2002; Zander & Kogut, 1995; Kogut & Zander, 1992). These essentially assess how complicated the knowledge might be (and difficult to fully understand) and how sticky or tied to the originating firm. Organizational aspects include characteristics like absorptive capacity (Cohen & Levinthal, 1990) and social capital (Nahapiet & Ghoshal, 1998). Absorptive capacity refers to the learning capabilities of the firm and its people while social capital refers to the number and strength of personal ties throughout the entity. As a consequence of the differing conditions, organizations pursue different KM approaches (Choi & Lee, 2003; Schulz & Jobe, 2001; Boisot, 1995) ranging from tacit-to-tacit techniques such as communities of practice or

storytelling to more explicit-oriented tools such as IT-driven knowledge markets (Matson, Patiath, & Shavers, 2003; Thomas, Kellogg, & Erickson, 2001; Brown & Duguid, 1991).

The related field of intellectual capital addresses many of the same basic concepts but from the standpoint of metrics, categorization, and strategic development of the assets. Growing out of scholar and practitioner interest in better accounting for fuzzy, intangible assets, IC typically looks to describe the knowledge assets of the firm according to the categories human capital, structural capital, and relational capital (Bontis, 1999; Edvinsson & Malone, 1997; Stewart, 1997). Human capital refers to individual, job-specific knowledge, structural capital to more long-term knowledge persisting in the organization (culture, systems), and relational capital to knowledge about relationships with those external to the firm.

A further extension adds in the idea of actionable intelligence. In many ways relatable to Ackoff's "wisdom" in the DIKW hierarchy mentioned earlier, intelligence refers to analysis of the knowledge (as well as the pre-knowledge data and information) in order to develop and execute strategies and tactics. Much like KM and IC, competitive intelligence grew out of practice as much as scholarship, especially early on (Prescott & Miller, 2001; Gilad & Herring, 1996; Fuld, 1994). Also similarly, the field often focuses on sources of knowledge, information, and data, as well as techniques for obtaining them (Fleisher & Bensoussan, 2002; McGonagle & Vella, 2002). Where matters start to diverge is in the analytic processes often applied to competitor knowledge and information, especially as competitive intelligence operations mature (Wright, Picton, & Callow, 2002; Raouch & Santi, 2001). Consequently, we do have some theory and practice related to the usefulness of all intangibles (data, information, and knowledge), not just knowledge, and the extension of creating actionable insights or intelligence, not just additional knowledge. Those fields looking for "intelligence" (competitive intelligence, marketing intelligence, business intelligence) have clear

connections with knowledge management but also have the potential to extend our understanding into new areas.

Within these interconnected structures and fields, scholars and practitioners have made substantial progress in identifying, understanding, and managing intangible assets. In recent years, data and information assets have reasserted themselves in the form of big data and business analytics. What can our understanding of knowledge assets and these related disciplines tell us about the potential value of the different approaches to big data?

Big Data and Knowledge

A number of factors have come together to create the new and exciting phenomenon known as big data, business analytics, business intelligence, or by any number of other names. The combination of ever increasing amounts of data generated by organizations, enhanced computing power, and the sharply decreased costs of storage and processing in the cloud have resulted in a greater interest in finding ways to benefit from better use of such data (Bussey, 2011; Vance, 2011b). Much of the data comes from operational systems (e.g. enterprise systems, supply chain management, supply chain/distribution channels) and/or transactional systems (e.g. customer relationship management, loyalty programs) (Vance, 2011a). Communication data, especially Internet-based or mobile (Google Analytics, social media) also has a role. While the databases are important to the process, it is the insights from analyzing the data that provide the strategic, tactical, and operational value. An oft-cited study from McKinsey Global Services notes the increased transparency available to big data practitioners allowing real-time experimentation, more precise segmentation, more objective decision-making via algorithm, and idea generation for new products (Manyika et al., 2011).

Although the scholarly literature related to big data is limited, it does have some history. Laney (2001) characterized the promise of data growth

as being wrapped up in volume of data, velocity of input and output, and variety of data, including sources. These three “v’s” have persisted over the past decade, including Laney’s (Beyer & Laney, 2012) own update that noted the volume, velocity, and variety allow new analysis techniques leading to deeper insights and better decisions.

Big data spans numerous industries and applications and is growing rapidly, especially in terms of explicit metrics like data storage (Liebowitz, 2013; Manyika et al., 2011). In constructing and utilizing an effective big data capability, however, organizations need some guidance, as just the hardware and systems are not sufficient. Decision-making can be enhanced by data systems, but the data systems also need a human element, the analytics, to be fully effective (Zhao, 2013). As noted above, the insights from the data are key. That requires intelligence, the ability to process, analyze, and make effective use of the data and information as well as marry the IT system to its human users.

This is where the intersection of the knowledge management, intellectual capital, and competitive intelligence frameworks can contribute. As hinted at in the literature review, some distinct connections between the fields have already been made. Big data treatments will sometimes acknowledge the KM and related literature in developing their conceptual foundation (Bose, 2009; Jourdan, Rainer, & Marshall, 2008). And intellectual capital, in particular has shown interest in expanding its purview, extending analysis to the full range of intangible assets (data, information, knowledge, intelligence) capable of providing value and competitive advantage to the organization. Andreou, Green, and Stankosky (2007), for example, constructed a framework including not just the traditional intellectual capital categories (or closely related ones) but also such items as competitive intelligence and enterprise intelligence, identifying all as potentially valuable corporate intangible assets.

Previous studies on the combination of KM/IC with competitive intelligence informs these links.

The structure and systems of KM differed from typical practice in competitive intelligence, often focusing on the IT and human interaction issues. Providing systems and incentives for employees to care about and make use of the KM structure was often the primary concern. Competitive intelligence, on the other hand, added ideas such as more targeted search, use of data and information as well as more developed knowledge, and the application of analysis with actionable results. Opportunities for cross-fertilization existed that could benefit the practice of both disciplines (Rothberg & Erickson, 2005). Recognition of both disciplines also made clear the security concerns when KM was fully leveraged, spreading organizational intangibles throughout extended networks (and thus more exposed to competitive intelligence efforts). Quantitative study demonstrated that both KM/IC and competitive intelligence practice faced dramatically different operating environments (Erickson & Rothberg, 2012). KM programs could be more or less effective in different industries while competitive intelligence could be a greater or lesser threat. Related depth interviews showed a similar wide variety in practice of both KM and competitive intelligence across industries and firms. Understanding the proper approach to KM/IC and competitive intelligence is strategic, with the right direction dependent on careful analysis of the firm, its industry, and other competitive circumstances.

One new direction in the field is found in revisiting the Ackoff DIKW hierarchy noted earlier. Kurtz & Snowden (2003) created the “Cynefin” sense-making framework, organizing intangible assets according to structure and distribution. As adapted by Simard (2014), this framework can be viewed as a hierarchy, running from data/information to explicit knowledge, then to tacit knowledge and on to insight/intuition. This form has a nature similar to DIKW but illustrates the similarities of the more structured, distributable assets such as big data (data/information) and

explicit knowledge while also highlighting their differences. Similarly, as tacit knowledge becomes even less structured and more personal, the distinctions are even clearer, but the final link to insight/intuition is also quite clear.

The value of employing this sort of thinking in analyzing the application and interaction of the full range of intangible assets is found in being able to better understand differences between industries or even between firms. KM applications have always stressed the differences between more explicit and more tacit knowledge and the implications for installing appropriate management systems. Taking the same thought a step further, installing big data systems and/or business analytics structures will also depend on circumstance. Some firms have the resources and core competencies to use data/information, and expertise with more information technology-based solutions such as those found in explicit knowledge conditions may help. Similarly, some firms will have the human element capable of the insights required for intelligence or analytics while others won't. But success with managing tacit knowledge may be an indicator of such potential success.

If we can begin to organize our thoughts and examine different circumstances in these ways, we can provide more eclectic, strategy-based recommendations for managing the full range of intangible assets. Further, the differences we see across industries may be more readily explained. Indeed, the gap between KM and IC scholarship and the work on human capital in strategy research might be narrowed. This type of structure better explains concepts like general vs. firm-specific knowledge and star systems as it differentiates even further between the structured and the personal (Nyberg, et. al., 2014; Ployhart, et. al. 2014). An opportunity exists to build a more comprehensive theory of how intangibles work in firms by reconsidering how data, information, explicit and tacit knowledge, and insight/intuition/intelligence fit together.

FUTURE RESEARCH DIRECTIONS

Abundant research opportunities exist in big data generally, as noted earlier, it is an underdeveloped field in terms of scholarship. Practice is well ahead of the academic literature. But the key point of this article concerns the overlap between the distinct fields of knowledge management, intellectual capital, competitive intelligence, and, now, big data and business intelligence. While big data serves to remind us that not all valuable intangibles are found in what we define as “knowledge,” there is also much that could be transferred from what we know about knowledge management as more organizations move to add big data programs.

In particular, both KM/IC and competitive intelligence have developed a substantial literature on understanding the nature of knowledge and using appropriate tools to manage it. Whether knowledge characteristics (tacit/explicit, complexity, specificity) or knowledge type (human, structural, or relational capital), there are undoubtedly extensions that can be made to big data or business intelligence, particularly when looking at mechanisms to develop knowledge/data (communities of practice, knowledge markets) or analyze them (war games, scenario planning, shadow teams). Similarly, the people issues surrounding use of such systems (trust, incentives) and related issues (social capital, social networks) could also be useful in this new application.

And, in the other direction, increasing use and development of big data should improve the practice of KM/IC and competitive intelligence. Although not as advanced, there is no reason valuable knowledge and intelligence cannot be enhanced by the analysis of data and information. The value of intangibles can only be enhanced by learning to use an even wider range of inputs, and there is interesting work to be done on how to add data and information analysis to existing KM structures.

CONCLUSION

This article has focused on some of the similarities and opportunities for cross-fertilization between the existing disciplines of knowledge management, intellectual capital, and competitive intelligence and the newer field of big data/business analytics. Previous work has shown opportunities for learning between the existing fields. Bringing big data into the picture extends the opportunities in important ways in a new field drawing considerable interest from both practitioners and scholars. In particular, big data widens the net for potentially useful inputs, including data and information that has yet to develop into the knowledge traditionally developed by KM systems. Big data practice also reinforces the importance of actionable intelligence, recognized in competitive intelligence practice but not always in KM/IC.

At the same time, big data installations will undoubtedly run into some of the same growing pains experienced by KM/IC and competitive intelligence over the years. Establishing the right-sized initiative, with the right price tag, and with the right tools/techniques has shown itself to be an important consideration in the more mature disciplines. Beyond the IT structure, working out the people issues, with concerns about trust, motivation, usefulness, and other such matters has also proven to be an important issue. The underlying social capital and social networks can also be critical. Finally, the importance of protecting intangible assets, not only in terms of IT and data security but, again, from the perspective of the people involved will likely also prove to be a concern. The existing disciplines have depth in all of these areas, in both in scholarship and application, proving a useful guide for those willing to learn from past experience.

REFERENCES

- Ackoff, R. (1989). From data to wisdom. *Journal of Applied Systems Analysis*, 16, 3–9.
- Andreou, A. N., Green, A., & Stankosky, M. (2007). A framework of intangible valuation areas and antecedents. *Journal of Intellectual Capital*, 8(1), 52–75. doi:10.1108/14691930710715060
- Beyer, M. A., & Laney, D. (2012). *The importance of 'big data': A definition*. Retrieved from <https://www.gartner.com/doc/2057415>
- Boisot, M. (1995). Is your firm a creative destroyer? Competitive learning and knowledge flows in the technological strategies of firms. *Research Policy*, 24(4), 489–506. doi:10.1016/S0048-7333(94)00779-9
- Bontis, N. (1999). Managing organizational knowledge by diagnosing intellectual capital: Framing and advancing the state of the field. *International Journal of Technology Management*, 18(5/6/7/8), 433–462. doi:10.1504/IJTM.1999.002780
- Bose, R. (2009). Advanced analytics: Opportunities and challenges. *Industrial Management & Data Systems*, 109(2), 155–172. doi:10.1108/02635570910930073
- Brown, J. S., & Duguid, P. (1991). Organizational learning and communities-of-practice: Toward a unified view of working, learning, and innovation. *Organization Science*, 2(1), 40–57. doi:10.1287/orsc.2.1.40
- Bussey, J. (2011). Seeking safety in clouds. *The Wall Street Journal*, p. B8.
- Choi, B., & Lee, H. (2003). An empirical investigation of KM styles and their effect on corporate performance. *Information & Management*, 40(5), 403–417. doi:10.1016/S0378-7206(02)00060-5
- Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly*, 35(1), 128–152. doi:10.2307/2393553
- Edvinsson, L., & Malone, M. (1997). *Intellectual capital*. New York: Harper Business.
- Erickson, G. S., & Rothberg, H. N. (2012). *Intelligence in action: Strategically managing knowledge assets*. London: Palgrave Macmillan; doi:10.1057/9781137035325
- Fleisher, C. S., & Bensoussan, B. (2002). *Strategic and competitive analysis: Methods and techniques for analysing business competition*. Upper Saddle River, NJ: Prentice Hall.
- Fuld, L. M. (1994). *The new competitor intelligence: The complete resource for finding, analyzing, and using information about your competitors*. New York: John Wiley.
- Gilad, B., & Herring, J. (Eds.). (1996). *The art and science of business intelligence*. Greenwich, CT: JAI Press.
- Grant, R. M. (1996). Toward a knowledge-based theory of the firm. *Strategic Management Journal*, 17(82), 109–122. doi:10.1002/smj.4250171110
- Jourdan, Z., Rainer, R. K., & Marshall, T. E. (2008). Business intelligence: An analysis of the literature. *Information Systems Management*, 25(2), 121–131. doi:10.1080/10580530801941512
- Kogut, B., & Zander, U. (1992). Knowledge of the firm, combinative capabilities, and the replication of technology. *Organization Science*, 3(3), 383–397. doi:10.1287/orsc.3.3.383
- Kurtz, C. F., & Snowden, D. J. (2003). The new dynamics of strategy: Sensemaking in a complex complicated world. *IBM Systems Journal*, 42(3), 462–483. doi:10.1147/sj.423.0462
- Laney, D. (2001). *3D data management: Controlling data volume, velocity and variety*. Retrieved November 1, 2013 from <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>

- Liebowitz, J. (Ed.). (2013). *Big data and business analytics*. Boca Raton, FL: CRC Press/Taylor & Francis; doi:10.1201/b14700
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Hung Byers, A. (2011). *Big data: The next frontier for innovation, competition and productivity*. McKinsey Global Institute.
- Matson, E., Patiath, P., & Shavers, T. (2003). Stimulating knowledge sharing: Strengthening your organizations internal knowledge market. *Organizational Dynamics*, 32(3), 275–285. doi:10.1016/S0090-2616(03)00030-5
- McEvily, S., & Chakravarthy, B. (2002). The persistence of knowledge-based advantage: An empirical test for product performance and technological knowledge. *Strategic Management Journal*, 23(4), 285–305. doi:10.1002/smj.223
- McGonagle, J., & Vella, C. (2002). *Bottom line competitive intelligence*. Westport, CT: Quorum Books.
- Nahapiet, J., & Ghoshal, S. (1998). Social capital, intellectual capital, and the organizational advantage. *Academy of Management Review*, 23(2), 242–266.
- Nelson, R. R., & Winter, S. G. (1982). *An evolutionary theory of economic change*. Cambridge, MA: Harvard University Press.
- Nonaka, I., & Takeuchi, H. (1995). *The knowledge-creating company: How Japanese companies create the dynamics of innovation*. New York: Oxford University Press.
- Nyberg, A., Moliterno, T., Hale, D., & Lepak, D. (2014). Resource-based perspectives on unit-level human capital: A review and integration. *Journal of Management*, 40(1), 316–346. doi:10.1177/0149206312458703
- Ployhart, R. E., Nyberg, A. J., Reilly, G., & Maltarich, M. A. (2014). Human capital Is dead; Long live human capital resources! *Journal of Management*, 40(2), 371–398. doi:10.1177/0149206313512152
- Polanyi, M. (1967). *The tacit dimension*. New York: Doubleday.
- Prescott, J. E., & Miller, S. H. (2001). *Proven strategies in competitive intelligence: Lessons from the trenches*. New York: John Wiley and Sons.
- Rothberg, H. N., & Erickson, G. S. (2005). *From knowledge to intelligence: Creating competitive advantage in the next economy*. Woburn, MA: Elsevier Butterworth-Heinemann.
- Rouach, D., & Santi, P. (2001). Competitive intelligence adds value: Five intelligence attitudes. *European Management Journal*, 19(5), 552–559. doi:10.1016/S0263-2373(01)00069-X
- Schulz, M., & Jobe, L. A. (2001). Codification and tacitness as knowledge management strategies: An empirical exploration. *The Journal of High Technology Management Research*, 12(1), 139–165. doi:10.1016/S1047-8310(00)00043-2
- Schumpeter, J. A. (1934). *The theory of economic development*. Cambridge, MA: Harvard University Press.
- Simard, A. (2014). Analytics in context: Modeling in a regulatory environment. *Proceedings of the International Conference on Analytics Driven Solutions 2014*, 82–92.
- Stewart, T. A. (1997). *Intellectual capital: The new wealth of nations*. New York: Doubleday.
- Teece, D. J. (1998). Capturing value from knowledge assets: The new economy, markets for knowledge, and intangible assets. *California Management Review*, 40(3), 55–79. doi:10.2307/41165943
- Thomas, J. C., Kellogg, W. A., & Erickson, T. (2001). The knowledge management puzzle: Human and social factors in knowledge management. *IBM Systems Journal*, 40(4), 863–884. doi:10.1147/sj.404.0863
- Vance, A. (2011a). *The data knows*. *Bloomberg Businessweek*.

Vance, A. (2011b). *The power of the cloud*. *Bloomberg Businessweek*.

Wernerfelt, B. (1984). The resource-based view of the firm. *Strategic Management Journal*, 5(2), 171–180. doi:10.1002/smj.4250050207

Wright, S., Picton, D., & Callow, J. (2002). Competitive intelligence in UK firms, a typology. *Marketing Intelligence & Planning*, 20(6), 349–360. doi:10.1108/02634500210445400

Zack, M. H. (1999a). Developing a knowledge strategy. *California Management Review*, 41(3), 125–145. doi:10.2307/41166000

Zack, M. H. (1999b). Managing codified knowledge. *Sloan Management Review*, 40(4), 45–58.

Zander, U., & Kogut, B. (1995). Knowledge and the speed of transfer and imitation of organizational capabilities: An empirical test. *Organization Science*, 6(1), 76–92. doi:10.1287/orsc.6.1.76

Zhao, D. (2013). Frontiers of big data business analytics: Patterns and cases in online marketing. In J. Liebowitz (Ed.), *Big data and business analytics* (pp. 43–68). Boca Raton, FL: CRC Press/Taylor & Francis; doi:10.1201/b14700-4

KEY TERMS AND DEFINITIONS

Big Data: Large amounts of data generated and stored by an enterprise, often in the areas of operations, transactions, and/or communications.

Business Analytics/Intelligence: Analysis of big data for strategic, tactical, and/or operational insights.

Competitive Intelligence: Gathering and analyzing data, information, and knowledge relating to a competitor or related topic, resulting in actionable intelligence.

Data: Observations.

Information: Data in context.

Intellectual Capital: Knowledge assets of the organization, commonly thought to be made up of human capital (job-related knowledge), structural capital (persisting knowledge assets of the organization such as systems or culture), and relational capital (knowledge relating to external entities).

Intelligence: Actionable insights.

Knowledge Management: Methods to identify, organize, and leverage knowledge assets through further distribution and sharing.

Knowledge: Know-how or expertise.

Business Intelligence

B**Richard T. Herschel***Saint Joseph's University, USA*

INTRODUCTION

Big data is one of the most commonly written about topics in today's press. Routinely we are bombarded with reports about how much more data there is, how much more is now able to be captured, how many new sources it comes from, and how it is being used to in new and novel ways at the expense of our privacy. As Bernard Marr (2015) asserts it is a topic that is discussed in boardrooms, business publications, and the mainstream media, because big data provides new insights into everything. Big data encompasses traditional sources of structured transaction data that is now supplemented by mass quantities of unstructured data. This data is processed by new, inexpensive, and faster hardware that is then scrutinized by new and more advanced analytics that provide organizations with more in-depth insight into their operational environment than ever before.

The role of business intelligence [BI] is to seek value from data. Today, BI combines text, video, voice, location data, social media, and any other new source of data with traditional data sets in order to learn about, interact with, and predict what is happening so that the organization can respond as fast as possible to whatever it perceives is the opportunity that the data reveals. BI deals with imperfect data that is oftentimes ambiguous, but which is available on a vast scale. As a result, Mayer-Shonberger and Cukier (2014) assert that the effect is that the extraction of value from big data is analogous to a treasure hunt. That is, organizations are scrutinizing big data to learn what is happening, without necessarily needing to understand why. They argue that in a big data world, correlations supersede causality, because

the data is simply used to discover patterns and correlations in the data that offer novel and invaluable insights. The more data you have the better the insights. The underlying premise for BI then becomes this: the more data an organization can capture, the better the data-driven probability of understanding what is happening, and the faster you can respond to this insight. This means then that actions taken in BI are often based on an organizational/system confidence level in the analytic assessment of what the data suggests without any clear understanding of the root cause.

Big data would appear to many to be more about systems, and less about people. Certainly people are important because they are themselves a major source of big data fodder and it is oftentimes people's behavior that big data is trying to affect. Nevertheless, big data also is dependent on people because people must inevitably be responsible for how data is used, how it is managed, and for the consequences of the decisions made when using it.

This paper is intended then to remind us that big data is not simply something that data, systems, and analytics make happen and that we are somehow divorced from it and not responsible for unintended consequences. Instead, prudence would require that since we have unleashed big data, we have to somehow insure to the best of our ability that if we can't control big data we can at least use common sense in how we approach it and manage it.

BACKGROUND

The amount of data in our world has exploded exponentially such that data, especially unstructured data, is now referred to as "big data". Where

DOI: 10.4018/978-1-5225-2255-3.ch082

measures of data were once gradually evolving from megabytes to terabytes, the sudden phenomena of big data accelerated these measures to volumes expressed in petabytes (1,024 terabytes) or exabytes (1,024 petabytes). The new influx of data is derived from billions to trillions of records of millions of people—all from different sources (e.g. Web, sales, customer contact center, social media, mobile data and so on). The data is typically loosely structured and often incomplete.

Big Data is the natural result of four major global trends:

1. Mobile computing
2. Social networking
3. Cloud computing
4. Moore's Law [processing power doubles every 2 years]

A lot of big data is derived from cell phone traffic and social networks, with much of it being stored in the cloud. The amount of this information grows almost exponentially as people routinely interact with companies and each other via wireless and wired networks as a matter of course.

The data being transmitted and captured from these transactions has become a key basis of competition, underpinning new waves of productivity growth, innovation, and consumer surplus, according to research by MGI and McKinsey's Business Technology Office (Manyika, Chui, Brown, Bughin, Dobbs, Roxburgh, & Hung Byers, 2011). From the standpoint of competitiveness and the potential capture of value, big data has had a substantial impact. Today, organizations leverage data-driven strategies to innovate, compete, and capture value from deep and up-to-real-time information. Hence, leaders in every sector are confronted with the need to grapple with the implications of big data in a quest to capture and harness its potential value.

The significance of big data to an organization falls into two categories: analytical use, and enabling new products. Big data analytics can reveal insights hidden previously by data too costly to

process, such as peer influence among customers, revealed by analyzing shoppers' transactions, social and geographical data. Being able to process every item of data in reasonable time removes the troublesome need for sampling and promotes an investigative approach to data, in contrast to the somewhat static nature of running predetermined reports. Hence, big data expands BI's data-driven decision-making opportunities.

In their book, "Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Business", Minelli, Chambers, and Dhiraj (2013) present significant evidence supporting the value and justification for businesses to be 'big data-driven'. Minelli et al. define big data as data that goes beyond the traditional limits of data along three dimensions: volume, variety, and velocity. Volume is important because in the past, business use cases and predictive analyses were restricted because the data volume utilized was limited due to storage or computational processing constraints. However, with big data technology removing these constraints and allowing the use of unstructured data and more transaction data for larger data sets, Minelli et al., assert that organizations can now discover more subtle patterns that can lead to targeted actionable decisions, or allow them to factor in more observations or variables captured over a longer period of time into predictive models. Variety of data refers to the different types of data now available for analyses. Instead of simply using structured text or numbers, unstructured text, audio, images, geospatial information, and internet data are now captured and able to be analyzed. Velocity is about the speed at which data is created, accumulated, ingested, and processed. Minelli et al., state that since this data is now readily available, organizations pursue technologies to help them immediately process this information for real-time analytics-based decision-making.

IBM's Institute for Business Value (2013) added a fourth dimension to describe big data called *veracity*, or the uncertainty of data. They argue that the need to acknowledge and plan for uncertainty is a dimension of big data that has

been introduced as executives seek to better understand the world around them. With data types changing constantly, and data collection increasing in quantity and speed, the quality of the data is a concern to be managed.

SAS Institute (2015) proposes two more dimensions as important to the understanding of big data:

1. **Variability:** In addition to the increasing velocities and varieties of data, data flows can be highly inconsistent with periodic peaks. Daily, seasonal and event-triggered peak data loads can be challenging to manage, but they are even more so with unstructured data involved.
2. **Complexity:** Data now comes from multiple sources. And it is still an undertaking to link, match, cleanse and transform data across systems. This requires that organizations increasingly focus on connecting and correlating relationships, hierarchies and multiple data linkages or data can quickly spiral out of control (SAS Institute, 2015).

In essence, big data has now become a kind of capital. Because of its breadth, volume, richness, and timeliness, many firms see it as an important asset that when harnessed, can be used for creating new products and services. It is a capital on par with financial capital. Oracle (2015) asserts that for CEOs to realize its potential, they must secure access to big data and then increasingly exploit its use before their competition can.

ISSUES, CONTROVERSIES, PROBLEMS

The rapid rise of the big data phenomena presents a plethora of new challenges for CEOs and their organizations. For example, IDG Enterprise Big Data research (IDG, 2014) conducted a survey with the goal of gaining a better understanding of organizations' big data initiatives, investments

and strategies. Their key finding included the following:

- Organizations are seeing exponential growth in the amount of data managed.
- Companies are intensifying their efforts to derive value through big data initiatives.
- CEOs are focused on the value of big data and are partnering with IT executives who will purchase/manage/execute the strategies.
- Organizations are investing in developing or buying software applications, additional server hardware, and hiring staff with analytics skills for big data initiatives.
- Organizations are facing numerous challenges with big data initiatives, especially the limited availability of skilled employees to analyze and manage data.
- Half of respondents indicated there is no clear thought leader in the big data solution space (IDG, 2014).

Clearly big data is getting organizational attention. However, the above findings suggest that firms are struggling to manage how they adapt to big data challenges on a number of fronts while at the same time facing some urgency in trying to harness its value.

Some of the challenges they face were unanticipated. For example, access to new sources of big data, including mobile data, social media content, and location services information means that businesses are collecting an extensive amount of personal information about their customers. This means that:

- Businesses now have the capacity to leverage more personally identifiable and sensitive information for competitive advantage.
- There is a dramatic surge in identity theft that has the potential to compromise secure transactions.

- Firms must address the development of sophisticated technology to exploit data security vulnerabilities.
- Consumers are increasingly aware and concerned about the collection, use, and disclosure of personal information.
- Legislators are responding to consumer concerns by restricting access to and use of personal information (Herschel, 2015).

An obvious conclusion is that big data can cause big headaches if organizations don't have a plan to manage it in the midst of rapidly evolving circumstances. To address these issues, firms cannot rely simply on software solutions, hardware fixes, or other IT initiatives. Instead, management is going to have to take a deep breath and consider pursuing broader organizational solutions to address some of these issues.

SOLUTIONS AND RECOMMENDATIONS

To determine how well organizations are taking advantage of big data, the IBM Institute for Business Value (Balboni, Finch, Rodenbeck Reese, & Shockley, 2013) surveyed 900 business and IT executives from 70 countries. Their analysis of the survey results enabled them to establish nine levers that differentiated companies who are best able to derive value from big data from those who aren't:

1. **Culture:** Availability and use of data and analytics within an organization.
2. **Data:** Structure and formality of the organization's data governance process and the security of its data.
3. **Expertise:** Development of and access to data management and analytic skills and capabilities.
4. **Funding:** Financial rigor in the analytics funding process.
5. **Measurement:** Evaluating the impact on business outcomes.
6. **Platform:** Integrated capabilities delivered by hardware and software.
7. **Source of Value:** Actions and decisions that generate results.
8. **Sponsorship:** Executive support and involvement.
9. **Trust:** Organizational confidence (Balboni, Finch, Rodenbeck Reese, & Shockley, 2013).

This list affirms that many of the issues surrounding big data are issues of organizational management. That is, culture, expertise, finance, performance measurement, value creation, leadership, and trust are all traditional management concerns. What has changed, is the suddenness of big data, the new technologies that have been created to process it, and the impact it is having on the ability of firms to react to it in an effective and timely manner. In essence, big data has become an aggressive, persistent, and pervasive irritant that is able to expose organizational data management limitations. Therefore, one of the easiest ways to understand an organization's shortcomings in the handling of big data would be to examine how capable they currently are in dealing with data management - period.

In his book on data management, Tony Fisher (2009) argues that data governance programs are essential for enterprises to experience success with business intelligence initiatives. As companies grow and look to incorporate additional data sources such as big data it is vital that they have an effective data governance process in place. To assess this, Fisher created the Data Governance Maturity Model that encompasses four phases: Undisciplined, Reactive, Proactive, and Governed. Each phase represents a different risk versus reward profile and as a company progresses through the model they increase their ROI for business intelligence while simultaneously reducing risk for the company. Each of these stages is defined below:

- **Undisciplined:** At the initial stage of the Data Governance Maturity Model, an organization has few defined rules and policies regarding data quality and data integration. The same data may exist in multiple applications, and redundant data is often found in different sources, formats and records. Companies in this stage have little or no executive-level insight into the costs of bad or poorly integrated data. About one-third of all organizations are at the Undisciplined stage.
- **Reactive:** A Reactive organization locates and confronts data-centric problems only after they occur. Enterprise resource planning (ERP) or Customer Relationship Management (CRM) applications perform specific tasks, and organizations experience varied levels of data quality. While certain employees understand the importance of high-quality information, corporate management support is lacking. Studies show that the largest share of all organizations – 45 to 50 percent – fall into this stage.
- **Proactive:** Less than 10% of all companies have reached this level. The Proactive organization adopts a more unified enterprise view of data, implementing and using customer data integration and product data management solutions. In this stage, management understands and appreciates the role of data governance and data is seen as a strategic asset. Here data stewards emerge as the primary implementers of data management strategy and work directly with cross-functional teams to enact data quality standards.
- **Governed:** At the Governed stage, an organization has a unified data governance strategy throughout the enterprise. Data quality, data integration and data synchronization are integral parts of all business processes. The organization has a comprehensive program that elevates the pro-

cess of managing business-critical data (Dataflux, 2015).

These stages provide a framework for organizations to gauge how well they are progressing with their data governance, data quality, and data management efforts while also providing information as to how they can progress to improve to a higher level of Master Data Management [MDM] – a term commonly used in the marketplace to describe this framework. Informatica (2015) defines MDM this way:

Master data management (MDM) in effect turns the Data Governance Maturity Model from a framework to a methodology for improving data management. MDM, like the Maturity Model, has the goal of helping organizations to attain:

- **A Single View of The Data:** Creating a single, authoritative view of business-critical data from disparate, duplicate, and conflicting information.
- **A 360-Degree View of The Relationships:** Establishing business rules that let firms identify the relationships among the data.
- **A Complete View of All Interactions:** Integrating the transactions and social interactions that have occurred with a product, customer, channel partner, or other data element to give the organization a complete view of that customer (Informatica, 2015).

Ali (2014) states that C-suite executives are increasingly taking notice of the market opportunity for big data analytics. Organizations that deploy data analytics are, he asserts, better equipped to make time sensitive decisions, monitor emerging trends, react quickly to new business opportunities, and ultimately, maximize the ROI of real-time data better than ever before. If data is being dispersed throughout the enterprise, the ability to discover or to fully leverage analytics tools to gain value and insight from data is cumbersome and inaccurate. An MDM framework therefore becomes essential to ensure that the data supporting the analytics

is relevant, timely and accurate. Inevitably then, executive support of MDM becomes essential if the organization is going to employ big data for data-driven decision-making.

Indeed, Eckerson (2014) writes that effective data governance programs must be enterprise-wide with effective programs combining both a top-down and bottom-up approach to data management. He argues that organizations must have a data governance methodology like MDM in place to enable this. Moreover, he recommends that senior management establish a data governance office (DGO) with a full time staff and a data governance officer. Eckerson proposes that the DGO should work with a data governance steering committee and data stewards to manage and monitor the organization's data governance program. Data stewards are employees with specific subject matter expertise who assume responsibility for certifying data, managing exceptions, and applying controls to safeguard data and individual privacy. Their role is to ensure that the data management, data governance, and business and quality rules definitions are appropriate for different lines of business. Each data steward provides critical input into cross-enterprise data management policies and ensures that these policies are appropriate for the business processes and applications that support the specific lines of business that they represent. Because of the necessity for these interactions, Eckerson believes that data governance is mostly about communications and change management where many conversations are routinely occurring among many people in different parts of the organization to insure that business processes flow smoothly and efficiently.

Brown, Court, and Willmott (2013) state that effective data management requires executive buy-in for data governance efforts to succeed, because capturing the potential of data analytics requires a clear plan that establishes priorities and well-defined pathways to business results and developing that plan requires leadership. This presumes that executives clearly understand and endorse the use of big data, that they comprehend what data

governance is and how important their role is to insuring its adoption and diffusion, and that they fully comprehend that effective data governance comes at a cost: organizational change. This means that senior management must be intimately involved in crafting the blueprint that encompasses the vision, strategy and requirements for the organization's big data journey. This blueprint should define the scope of big data utilization and management within the organization, prioritize the key business challenges that big data will address, provide a schedule for dealing with these challenges, and provide a plan for developing or acquiring the people and technology resources needed to get the job done (Balboni, Finch, Rodenbeck Reese, & Shockley, 2013).

To be effective, big data solutions must be based on business requirements and address the infrastructure, processes, data sources, and skills required to support this business opportunity. To accomplish this, senior management must establish a plan and process that is consistent with the organization's culture and guiding principles such that these are subsequently reflected in the organizational data governance activities. Management must ensure that data governance is viewed as a continuous and systemic endeavor that is woven into the fabric of the organization's structure and processes rather than being established as an ad hoc response created to deal with a novel 'big data' phenomenon.

Executive understanding and support of MDM-type big data governance processes should reflect the conscious realization that they will need to develop, acquire, and retain a strong core of human capital. Qualified data stewards, for example, must be identified and selected. These individuals, through their actions, reflect the values, policies, and procedural requirements that are derived from senior management. In a big data world, executive leadership is also needed to find, develop, and retain strong analytical talent. Analysts with strong mathematical, statistical and computer science roots, as well as familiarity or expertise in areas such as data mining, data visualization,

optimization and simulation, text analytics, and predictive analytics are going to be required to harvest meaning and value from big data. The problem is that there is a severe shortage of this talent. For example, McKinsey & Company (2015) asserts that only 18 percent of companies believe they have the skills to gather and use data effectively. And this becomes more troubling when they further state that the United States alone faces a shortage of 140,000 to 190,000 people with analytical expertise and 1.5 million managers and analysts with the skills to understand and make decisions based on the analysis of big data. IDC's (2014) estimates are also daunting, stating that in the U.S. alone, there will be 181,000 deep analytics roles in 2018 and five times that many positions requiring related skills in data management and interpretation. Leadership is required to find creative means for acquiring the requisite BI talent that organizations need.

Dyché and Nevala (2015) discuss key issues where executives have to be careful in addressing big data and data management issues. These issues include failing to define what data governance is, not approaching big data and data management in a progressive and systematic manner, failing to design data governance teams properly, treating data management as a project, and overlooking organizational cultural issues. They note that issues needing to be addressed by data governance and big data are far-flung and pervasive, ranging from arbitration of cross-functional data usage to information privacy, security, and access policies. As a result, executives need to understand that if they look to implement data governance and other big data initiatives too quickly, the intended results may well be stymied quickly by role confusion, prioritization debates, project development problems, and resistance from the incumbent culture. When these issues are combined with the plethora of other issues that big data creates, it becomes essential that executives be cautious and level headed when evaluating how data management efforts are being constructed and deployed in their organizations.

FUTURE RESEARCH DIRECTIONS

B

Research into the impact of big data needs to expand its focus to explore in more depth how BI is affected by management's ability to cope with data governance, analytics talent recruitment, and other peripheral issues. MDM and data governance activities appear to be sensible techniques for managing data, but it is clear from the statistics that not all of these practices are being effectively adopted and deployed. Understanding the rationales for the adoption and leveraging of data management techniques as well as their consequences provide insight into managements understanding of big data and their ability to exploit it.

Managing big data requires common sense. Common sense is derived from the way you think, intuition, and experience. It suggests the ability to think and behave in a reasonable way and to make good decisions. It incorporates such attributes as the ability to think rationally, to act purposefully, and to deal effectively with your environment. Common sense dictates that firms need to be cautious with big data so as not to compromise their BI efforts and data-driven decision-making. I argue that given the technologies, data management techniques, and advanced analytics that impact BI, we need to also focus on better understanding how big data affects management common sense and vice versa. That is, we need to look behind the curtain to learn more about the management actors on the big data stage.

CONCLUSION

Big data is not a cure-all, and it is inherently filled with noise and uncertainty. It has tremendous potential if people approach it the right way. What is important is that there be thoughtful leadership in organizational efforts to manage its capture, storage, sharing, and use. Big data is not just about data and technology. It is clearly about organizational changes in attitudes and functions. And collectively all of these efforts focus on one

thing – gaining better business intelligence for decision-making.

All Analytics (2015) notes that with big data, it is important to understand not just what data you have and what the data says, but also how specifically the data will be used in the business environment. The article asserts that:

- Companies can miss opportunities because line managers may not have been shown how data can help them do their jobs better,
- Management and analytics teams may need to examine existing problems, workflows, and processes to also understand how data might help them do their jobs better, and
- Organizations need to understand why it is essential to incorporate the human element into their planning processes for analytics initiatives.

Inevitably, business intelligence is about people, not just data and analytics. People establish and employ the processes and procedures where data and analytics are deployed and through which decisions are made. Hence, understanding the reasoning underlying these processes and procedures [and their consequences], and fleshing out the relevant data and processes that will help people make better data-driven decisions is important. However, in doing this we must also ensure that people can understand and appreciate the value of any changes being made. This means that we need to look hard and examine, document, and improve the interaction between human behavior and the goals of business intelligence in a world troubled by big data.

REFERENCES

- Ali, S. (2014). *Developing a Modern Data Management Strategy*. Retrieved April 1, 2015 from <http://www.information-management.com/news/Data-Management-Strategy-10026378-1.html>
- All Analytics. (2015). *Analytics Best Practices: Plan for the Human Factor*. Retrieved October 5, 2015 from <http://blog2best.blogspot.com/2015/09/avoid-key-flaws-in-analytics-plan-for.html>
- Balboni, F., Finch, G., Rodenbeck Reese, C., & Shockley, R. (2013). *Analytics: A Blueprint for Value*. IBM Global Services. Somers, NY: IBM Corporation.
- Brown, B., Court, D., & Willmott, P. (2013). *Mobilizing your C-suite for big-data analytics*. Retrieved March 10, 2015 from http://www.mckinsey.com/insights/business_technology/mobilizing_your_c_suite_for_big_data_analytics
- Dataflux. (2015). *The Data Governance Maturity Model: Establishing the People, Policies and Technology That Manage Enterprise Data*. Retrieved April 19, 2015 from <http://www.sas.com/offices/NA/canada/lp/DIDQ/DataFlux.pdf>
- Dyché, J., & Nevala, K. (2015). *Ten Mistakes to Avoid When Launching Your Data Governance Program*. Retrieved March 21, 2015 from http://www.sas.com/resources/whitepaper/wp_63774.pdf
- Eckerson, W. (2014). *Data Governance for the Enterprise - Trends in the Use of Data Quality, Master Data Management and Metadata Management*. Newton, MA: TechTarget.
- Fisher, T. (2009). *The Data Asset: How Smart Companies Govern Their Data for Business Success*. Hoboken, NJ: John Wiley & Sons, Inc.
- Herschel, R. (2015). *The Chief Knowledge Officer: Leadership for Big Data & Business Intelligence*. In *The Encyclopedia of Information Science and Technology* (3rd ed.; pp. 371–378). Hershey, PA: IGI Global Publishing.
- IBM Institute for Business Value. (2013). *Analytics: The real-world use of big data*. Somers, NY: IBM Corporation.

IDC. (2014). *IDC Reveals Worldwide Big Data and Analytics Predictions for 2015*. Retrieved April 20, 2015 from <http://www.idc.com/getdoc.jsp?containerId=prUS25329114>

IDG. (2014). *Research Reports: Big Data*. Retrieved April 11, 2015 from <http://www.idg-enterprise.com/report/big-data-2>

Informatica. (2015). *What is Master Data Management?* Retrieved April 02, 2015 from https://www.informatica.com/products/master-data-management.html#fbid=s7qB_Q-sMY-

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Hung Byers, A. (2011). *Big data: The next frontier for innovation, competition, and productivity*. Retrieved April 17, 2015 from http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

Marr, B. (2015). *Big Data: Using Smart Big Data Analytics and Metrics to Make Better Decisions and Improve Performance*. Chichester, UK: John Wiley & Sons Ltd.

Mayer-Schonberger, V., & Cukier, K. (2014). *Big Data*. Boston, MA: First Mariner Books.

McKinsey & Company. (2015). *Big data: The next frontier for competition*. Retrieved April 20, 2015 from http://www.mckinsey.com/features/big_data

Minelli, M., Chambers, M., & Dhiraj, A. (2013). *Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Business*. Hoboken, NJ: John Wiley & Sons, Inc. doi:10.1002/9781118562260

Oracle. (2015). *Enterprise Big Data Predictions 2015*. Retrieved April 18, 2015 <http://www.oracle.com/us/technologies/big-data/big-data-predictions-2015-2421021.pdf>

SAS Institute, Inc. (2015). *Big Data: What it is & why it matters*. Retrieved March 30, 2015 from http://www.sas.com/en_us/insights/big-data/what-is-big-data.html

ADDITIONAL READING

Davenport, T. (2014). *Big data at work: Dispelling the Myths, Uncovering the Opportunities*. Boston, MA: Harvard Business Review Press. doi:10.15358/9783800648153

Fisher, T. (2009). *The Data Asset*. Hoboken, NJ: John Wiley & Sons.

Mayer-Schonberger, V., & Cukier, K. (2014). *Big Data*. Boston, MA: First Mariner Books.

Minelli, M., Chambers, M., & Dhiraj, A. (2013). *Big Data Big Analytics*. Hoboken, NJ: John Wiley & Sons. doi:10.1002/9781118562260

KEY TERMS AND DEFINITIONS

Big Data: It is a popular term used to describe the exponential growth and availability of data, both structured and unstructured.

Business Intelligence (BI): It is a business management term which refers to applications and technologies which are used to gather, provide access to, and analyze data and information about their company operations.

Data Governance: It is a combination of people, processes and technology that drives high-quality, high-value information. The technology portion of data governance combines data quality, data integration and master data management to ensure that data, processes, and people can be trusted and accountable, and that accurate information flows through the enterprise driving business efficiency.

Data Steward: A data steward is a job role that involves planning, implementing and managing the sourcing, use and maintenance of data assets in an organization. Data stewards enable an organization to take control and govern all the types and forms of data and their associated libraries or repositories.

Master Data Management (MDM): It is a technology-enabled discipline in which business

and IT work together to ensure the uniformity, accuracy, stewardship, semantic consistency and accountability of the enterprise's official shared master data assets. Master data is the consistent and uniform set of identifiers and extended attributes that describes the core entities of the enterprise including customers, prospects, citizens, suppliers, sites, hierarchies and chart of accounts.

Unstructured Data: Unstructured data (or unstructured information) refers to information that either does not have a pre-defined data model or is not organized in a pre-defined manner. Unstructured information is typically text-heavy, but may contain data such as dates, numbers, facts, or video content as well.

Improving Competitiveness Through Organizational Market Intelligence

B**George Leal Jamil***Informações em Rede, Brazil***Leandro Rocha Dos Santos***In3 Inteligência de Mercado, Brazil***Cecília C. Jamil***Stockholm University, Sweden*

INTRODUCTION

Market intelligence (MI) is a concept that results, in its implementations, in real opportunities for knowledge management in organizations. In an initial view, MI can be regarded as a cyclic process to deliver knowledge for strategic marketing decisions, considering typical organizations of a defined sector as its final users. As organizations, mainly corporations, must not only react to external factors or phenomena, but also try to lead its sector proposing and executing innovative plans and differential strategic positioning, MI is a modern and cooperative approach which can produce consistent bases for such planning abilities (Boblitz, 2006; Porter, 2008; Mintzberg, Ahlstrand & Lampel, 2009; Kim & Mauborgne, 2015). Market intelligence is affirmed as an organizational continuum that aims to answer typical decision problems faced by firms when competing in actual business environments (Van Kesteren, 2012; Jamil *et al.*, 2012).

This article aims to detail MI concept, working from theoretical point of view through literature and adds a practical approach on studying real cases of market intelligence potential applications. As MI is a multidisciplinary context, its conceptualization also provokes perspectives for research from other scientific fields, as Information Science, Systems and Management, Computing

Science, Human resources management, Strategy, Marketing, among many others.

BACKGROUND

Strategic planning and execution are two main contexts where market intelligence is definitely relevant for critical decisions, motivating the following evaluation as the start of this theoretical work.

Strategic Decision Scenarios and Its Demands for Continuous Knowledge

Strategy formulation and execution are knowledge-dependent tasks, demanding for its continuity reliable and updated knowledge availability (Jamil *et al.*, 2012; Johnson, 2012; Dimitrios, Sakas & Vlachos, 2013).

Strategic marketing decisions are the objective of the MI process, as the knowledge provision aims to solve problems, allowing decisions with clearer risk delimitations and implementation results with better customer aggregated value perspectives, attending to the basic organizational marketing demands (Kotler & Keller, 2005; Schiffman & Kanuk, 2010; Ferrel & Hartline, 2010; De Man, 2012). Typical decisions of strategic marketing processes that can benefit from market intelligence process are:

DOI: 10.4018/978-1-5225-2255-3.ch083

- Product line configuration and distribution,
- Pricing,
- Advertising and general communication,
- Differentiation as a value-based strategy, and
- Marketing channel analysis.

As examples of knowledge needed in usual marketing decisions it can be perceived: consumer behavior details, demographic perspectives and constraints, customer reaction to distribution forms and communication, distribution channels performance and financial performance for all productive chain components. Knowledge provided for marketing strategic planning will allow any organization to develop its practical, tactical-to-operations plans, which will, at the end, specify the real work to be executed, aligned to those strategic views and propositions.

Data, Information, and Knowledge

Davenport & Prusak (2000), Tuomi (2000) and Lucas Jr. (2005) defined not only the concepts of data, information and knowledge but also its relationship and integrated application views. Data is considered as an absolute value that can be obtained directly from a measurement activity or collected from an automated source. It shows flexibility to communicate, as it can be easily processed by automated instruments and machinery, but lacks more profound meaning, as it is almost impossible to deduce more from the context or process it was created, being just an instantaneous observation of a reality. From these and other authors, information is conceptualized as a collection of correlated data added with context, providing a better condition for deciding, but offering more complexity to be treated or processed. Information increases decision capabilities, but demands additional work to be finally applied, as additional work must be conducted to assure its homogeneity and coherency (example: measurements – data – were collected referring to the same system and / or context).

Finally, knowledge is composed through collection of information, including descriptions of the processes which produced that information. Knowledge allows maximum decision capabilities, for example, enabling even prediction capabilities (Davenport, 2000; Jamil, 2005; Nonaka, 2008, Badia, 2014). But on the other hand, knowledge is difficult to manage, communicate, questionable to be stored and critical to be shared, resulting in the need of a specific process to treat it, defined as knowledge management (Jamil, 2001; El-Bashir, Collier & Sutton, 2011). This fundamental area of conceptual relationship is being significantly treated, as the “big data” phenomenon is increasingly perceived by entrepreneurs and socio-political actors, calling additional attention to the process of “knowledge generated from data” which proves how important is to observe data, information and knowledge in an organizational context for decisions such as strategic marketing planning (Ohata & Kumar, 2012; Park, Huh, Oh, & Han, 2012).

Market Intelligence

Derived from many areas, intelligence is another concept that motivated various studies and expressive debate. It is opportune, though, to evaluate contributions from other scientific fields and also applied practice to improve market intelligence comprehension. This approach was also applied for several authors, as Albescu and Pugna (2014), supported by almost the same fundamental concepts applied in this development.

Competitive intelligence (CI) is a process which offers an opportune complementarity to market intelligence, although it is possible to find some confusion and lack of delimitation in some approaches of these two concepts. SCIP (2012) defines competitive intelligence as “A process of monitoring the competitive environment and analyzing the findings in the context of internal issues, for the purpose of decision support”. Kahaner (1998) and Miller (2002) defined it as a process related to strategic problems solution,

configuring CI as an organizational that has the goal to provide better decision support specifically for higher executive levels in organizations. It is important to notice that as a process primarily designed to produce knowledge from external environment interpretation, CI offers an objective response to a very precise problem, usually proposed in a formulation or revision of a strategic plan. It is so a managed cycle which is ignited by a strategic problem, held in order to provide a punctual, detailed, precise and objective to that problem (Makadok & Barney, 2001).

Market intelligence, on its way, composes internal, sectorial (external, but correlated by internal organizational configurations) and external sources to produce specific, focused knowledge to strategic marketing contexts and decision making guided to implement aggregated value positioning (Jamil *et al.*, 2012; De Man, 2012; Jamil, 2014). Finally, the convergence of strategic results of both processes – CI for strategic decision, such as those in planning and executing strategies and MI for marketing strategic decisions – also allows to understand how they can be complementary and interrelated, reiterating that market intelligence has a more practical, applicative, “to the market” objective as it is designed to support a connection between organizational strategy formulation and answers provided by strategic marketing decisions (Johnson, 2012).

As another contribution for MI conceptualization, “artificial intelligence” (AI) was proposed early as in 1950 decade, as an academic subject, sometimes related to philosophy or, alternatively, understood as a computational engineering support for automated or robotic system, as presented by John McCarty in 1956, in a conference at Dartmouth College. It turned out to be a provocative area that motivated computing, mathematic, physics and other areas’ specialists to develop theoretical and practical studies. From some contributions of artificial intelligence it is possible to understand and perceive more comprehension about this vision of intelligence: a continuum that promotes progressive and adaptive learning

absorbed from the reality, transforming it in rules resulted from accumulative knowledge created (Russel & Norvig, 2009). It is interesting to note that this accumulation of knowledge that compose rules to be further applied by one organization can also be considered as a registering of knowledge, explicitly transformed in contents that could be applied, developed, integrated and transmitted (Nonaka, 2008). As it was suggested for robotic and automated or controlled environments, AI is a form to codify predictable knowledge for repetitive application, with many improvements that can occur from continuous cycles of process (Hubber, 1990; Weiss & Verma, 2002; McAfee & Brynjolfsson, 2012).

Finally, from information technology (IT) and computing science fields, “business intelligence” (BI) arises as a relevant and broadly discussed concept. It is used for many interpretations of tools, techniques, methods and integration of resources oriented to produce specific knowledge from disperse sources of information and data, as analyzed by Inmon, Strauss & Nishloss (2008) and Kimball & Ross (2010). The IT literature concentrates predominantly in technical issues, such as implementation models, physical storage, database structuring and integration. From this point of view, intelligence comes from technological arrangements, which offers not only the mechanisms for its development, correlation between logical models to physical implementations, but also for knowledge codification, process modeling and implementation and knowledge communication and distribution (McAfee & Brynjolfsson, 2012; Courtney, 2013; Jamil, 2014). Another special aspect is integration, as the business intelligence usually is based on an infrastructure that is implemented through an associated arrangement of organizational resources, not regarding aspects such as cyclic updates or evolution, but thinking it as a solution-driven knowledge provider (Cao, Zhang & Liu, 2006; Kimball & Ross, 2010; El-Bashir, Collier & Sutton, 2011; Hoffman, 2013).

Completing this conceptual approach, market is a keyword that has to be noticed. As a con-

ceptual focus, market defines the objective for continuous knowledge treatment and production for decisions. Market intelligence is a continuous process that allows the production of knowledge from data and information in order to be applied for strategic market decisions, or, better defining, strategic marketing planning (Chen, Chiang & Storey, 2012; Vavra, 2012; Dimitrios, Sakas & Vlachos, 2013).

Considering that data and information are spread over a value-aggregated chain and MI produces knowledge to be applied for sectorial strategic marketing decision on plans, it can be proposed, from those and other references, as the following continuous process, cycle definition:

- **Aggregated Chain Value Analysis:** For an overview of the organizational arrangement where the MI cycle will occur and have its results applied.
- **Data and Information Diagnosis:** where business sector contents of data and information are evaluated to constitute a model that represents what is available for initial process design.
- **Knowledge Diagnosis:** To conceive, initially, what are the main contents of knowledge demanded by typical market competitors, for decision taking in their strategic marketing planning work.
- **MI Continuum - Process (Modeling and Implementation):**
 - **Planning:** As a start / stop point of every cycle, it is a step of planning next cycle execution – resource allocation, pre and post requirements, deliveries, risks, etc. – based on the last cycle execution.
 - **Collection:** Gathering of data and information from organizations of the sector analyzed. External sources, as government and agencies, research institutes and other partners, if diagnosed as essential or opportune, are also collected.

- **Validation:** Normalization and correction of data and information collected, as those sources can produce different or mistreated contents that have to be leveled to become understandable as similar for cyclic analysis.
- **Processing:** Usage of business rules, logic modeling, statistical and mathematical analysis applied to data and information to produce specified knowledge contexts.
- **Sharing:** Results sharing of the processing that are delivered, presented, transmitted to interested market players for further application in their decision scenarios.

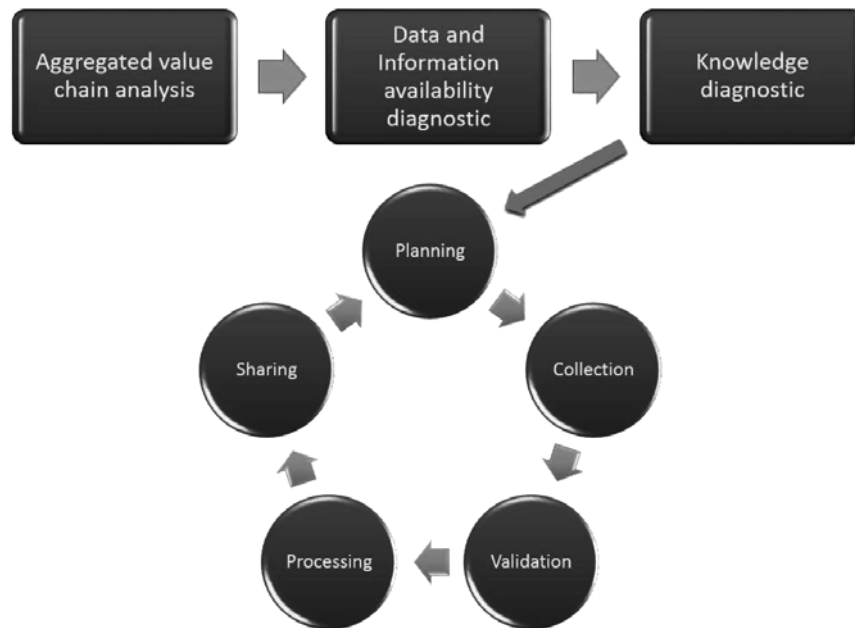
This definition, encompassing the first conceptual affirmation and the structure depicted above constitutes the theoretical background of market intelligence, being defined by the Fig. 1, detailed below:

In the following, an analysis of practical cases is held, validating this framework.

MARKETING STRATEGIC DECISIONS: CASES FOR MARKET INTELLIGENCE PROCESS APPLICATION

This section intends to consolidate the conceptual base discussed before through case studies where market intelligence played a relevant role as conceptualized. Each of the following case studies were worked for commercial and marketing sectors, with firms grouped as *informational associations*, according to Marchand, Kettinger and Rollins (2001), taken or observed as a group of organizations that present similar informational behaviors in terms of its problems, needs, infrastructures, research methods and processes.

Figure 1. Market intelligence cycle



Case 1: International Expansion Decision

In the pharmaceutical sector, productive arrangements are complex and can be specialized by product lines, resulting in a situation where one producer can formulate different marketing strategies for different products. This complexity increases substantially as a producer intends to expand its operations abroad, commercializing pharmacies in additional regions and countries. Along with customer data – for instance those which inform consumption potential, desirable forms for payment, buying behavior and decision – information about fiscal, health sector laws and distribution issues are needed to perform a complete study in order to solve decisions as to determine where to install a factory, the optimal distribution system for final customers of one region or how to reduce costs for a competitive offer in a region already efficiently handled by a competitor.

These issues can be more clearly solved with support of a market intelligence process, being conducted by a group of firms that compete in

this sector. A sectorial market scenario description – market knowledge – can be supplied in a predicted frequency, for instance biannually, presenting knowledge contents as: market share evolution, prices evolution, shared data about distribution and sales, final customer behavior and reaction to advertisement and promotional campaigns, among many others. A continuous analysis for any player can identify more precisely where – in what country or region – there is a potential local market that could provide financial support for a factory installation, allowing it to sell locally and abroad.

Case 2: Product Launching and Positioning

Is it better to launch a new isolated product or insert it in a product line already being in offer for final customers? This typical marketing-oriented decision is critical, as a product can have its perceived value components not correctly evaluated by a final customer, depending on the way it will be sold. For steel industry, for instance, a special type of steel can be understood by automotive

industry as a specialization of an already offered steel composition or a special product, resulting from sophisticated laboratorial work, with more aggregated value.

This decision can be worked more intensively and studied in a more detailed way, if the steel industrial sector practices a market intelligence analysis that, for a period of some “buying cycles” (for instance, for some years), show how and why the final production for its customers – automobiles, in this example – was commercially successful. Analyzing its position as a part of a productive chain, one steel manufacturer can observe if automotive industry will or not perceive this special type of steel as only an additional product or a differential, more valuable one as it is being used in more competitive assemblies (such as cars, in this scenario). Each market intelligence cycle provided, in this case, knowledge about the performance of the final customer end, automobile manufacturers, which base the decision of the supplier, allowing to a better choice on how to offer the special metal artifacts developed as a new product.

Case 3: Investments in Changing Economies Planning

Many questions arise when analyzing a new market, as those usually shown in emerging economies. As it was discussed in the first example, market intelligence can provide not only historical content for knowledge production, to be completely examined, but also allow one organization to project future prospective scenarios, where a whole economic system can be better scrutinized. Complex decisions are made in order to coordinate strategic and tactical movements to start in one new market as:

- How to develop commercial arrangements,
- Predict and design financial positions,
- Implement logistics network,
- Along with human resources projection,

- Organizational structure and
- Capabilities development.

Continuous knowledge provision can be applied for prospective scenarios design, allowing an executive board to evaluate projections of revenues, costs, taxes (which are a very complex issue, mainly in some growing economies), performance and production alternatives to compose strategic and correlated tactical plans and also monitor its execution. As strategic planning guides and defines conditions for tactical and operational plans, the aligned definition provided by more precise designs of competitive conditions, resulting from a complex analysis which encompasses sectorial, governmental, competitors and value chain data and information to produce knowledge, treated by a cyclic process, allows each firm to develop and control more realistic and risk managed plans.

From the cases discussed above it is possible to understand the main benefits of market intelligence process, bringing a more objective focus and comprehension about its structure. All of those decisions, and many other that could be studied, do not essentially suffer from lack of data, or even information, instead of it is possible to preview that huge amounts of data and information are available for decision. But it is also possible to predict that these contents are shown without the needed cohesion, coherency, relationship and, mainly, a process management which make it difficult to obtain the demanded knowledge. Examples of these problems can be learned from authors of information systems areas as Laudon & Laudon (2009) and O'Brien & Marakas (2008) – or in marketing research – as Kotler & Keller, (2008) or Ferrel & Hartline (2010). Structured models for data and information acquisition, along with process definitions for its collection, validation, processing and result production – the market intelligence process definition – allow to cyclically apply knowledge to better define these issues and obtain the expected responses for decision making.

FURTHER STUDIES AREAS

In this section, some of potential fields and further studies are presented both to reaffirm the concept of market intelligence and, in the spirit of an Encyclopedia, to give perspectives to the reader to develop new and deeper studies with support of the conceptual framework reached.

Information technology has a remarkable application not only in the structure of the market intelligence process, but also to advance its development, as database models, information systems design architecture and implementation, application of user-oriented knowledge production tools and many other subjects related to this area can be better observed collaboratively (Turban, Mc Lean, & Wetherbee, 2002; Turban, Rainer & Potter, 2007; Kimball 2010; Johnson, 2012; Mahrt & Scharkow, 2013).

As suggested themes for new researches in IT area a special attention must be devoted to how to implement business models, based on market intelligence through modern IT architectures, analyzing, for instance, the introduction of mobile payment, communications and customer-oriented interactions, database services modeling and the aspects of IT support for social media environments and the “big data” phenomena, rich but still not completely deciphered in terms of scientific application researches.

Information systems (IS) design, modeling and application are other subjects, as understood both from computing science and information science point of views. As knowledge production tools based on information and data gathering, it is opportune to advance the conceptual framework studied and developed and also understand how the information systems perspectives, services, design and repercussions can be obtained or implemented, along with user reaction to this tools and processes in organizations (Jamil, 2001; Stair & Reynolds, 2009; McAfee & Brynjolfsson, 2012; Hoffman, 2013).

Clearly, as those fields are the main areas of research implied by the eventual success of a

market intelligence process, organizational strategy – involving planning, execution, relationship with tactical planning and others – and marketing – especially strategic marketing planning – are two main scientific subjects to propose advanced studies about organizational intelligence. These themes can also provide opportune methodology-supported case studies for conceptual and practical evaluations, about planning and implementing hypothetical and practical market solutions. Connecting theory to practice, this relationship enables application of various methodological approaches, such as surveys, Delphi group analysis, participative studies, qualitative techniques and so forth. The results, potentially rich for new business models implementation can be considered as some interesting repercussions from market intelligence process detailed comprehension.

Also, information and knowledge management as central themes and its repercussions around areas such as Management, Information Science, Engineering and many others can benefit from MI discussion and conceptualization.

With this study, it is expected that not only a contribution for market intelligence process itself was reached, but also a base for further studies in various areas and scientific fields, related to strategic and marketing competitive positioning can be developed, taking it as one source of conceptual base.

CONCLUSION

In this chapter market intelligence process concept was developed in order to provide its fundamental understanding and allowing its observation for further studies. As a complex, intricate and interrelated organizational process, it offers implications for many entrepreneurial aspects and tasks, allowing not only the consolidation of the past, but also projecting some future perspectives for businesses.

MI is conceptualized as a process structured mainly as a continuum that starts with the compre-

hension of the value chain of a sector, modeling of data and information that exists in one sector and relating it to the needs of typical competitive firms of one sector for knowledge to solve strategic marketing decisions. From this first level, a routinely process is designed in order to collect, validate, process data and information and produce marketing strategic knowledge allowing better level of decision for each organization that was researched.

Its implications as a modern, competition-driven and competition-oriented process to be applied by sectors with possibilities for entrepreneurial plans and actions inside or outside a market sector are expressive, favoring better strategic, tactical and operational planning and execution and in all levels, better risk management and monitoring.

REFERENCES

- Albescu, F., & Pugna, I. B. (2014, July-September). Marketing intelligence: The last frontier of business information technologies. *Romanian Journal of Marketing*, 55–68.
- Badia, A. (2014). Data, information and knowledge: an information science analysis. *Journal of the Association for the Information Science and Technology*, 65(6), 1279–1287.
- Boblitz, M. C. (2006). Looking out the window: Market intelligence for a view of the real world. In *Healthcare financial management*, (pp. 46–53). Academic Press.
- Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business intelligence and analytics: from big data to big impact. *MIS Quarterly*, 36(4), 1165–1188.
- Choo, C. W. (1996). The knowing organization: How organizations use information to construct meaning, create knowledge and make decisions. *International Journal of Information Management*, 16(5), 329–340. doi:10.1016/0268-4012(96)00020-5
- Courtney, M. (2013). Puzzling out Big Data. *Engineering & Technology*, 7(12), 56–60.
- Davenport, T. H., & Prusak, L. (2000). *Working knowledge: how organizations manage what they know* (2nd ed.). Harvard Business Press.
- De Man, D. (2012). A platform for market intelligence. *International Trade Forum*, (4), 21.
- Dimitrios, N. K., Sakas, D. P., & Vlachos, D. S. (2013). The role of information systems in creating strategic leadership model. *Procedia: Social and Behavioral Sciences*, 73, 285–293. doi:10.1016/j.sbspro.2013.02.054
- El-Bashir, M. Z., Collier, P., & Sutton, S. G. (2011). The Role of Organizational Absorptive Capacity in Strategic Use of Business Intelligence to Support Integrated Management Control Systems. *Accounting Review*, 86(1), 155–184. doi:10.2308/accr.000000010
- Ferrel, O. C., & Hartline, M. (2010). *Marketing Strategy*. South Western College Publications.
- Hoffman, L. (2013, April). Looking back at the big data. *Communications of the ACM*, 56(4), 21–23. doi:10.1145/2436256.2436263
- Huber, G. P. (1990). A theory of the effects of advanced information technologies on organizational design, intelligence and decision making. *Academy of Management Review*, 15(1), 47–71.
- Inmon, B., Strauss, D., & Neushloss, G. (2008). *DW 2.0: The Architecture for the Next Generation of Data Warehousing*. Morgan Kaufmann.
- Jamil, G. L. (2001). *Repensando a TI na empresa moderna*. Rio de Janeiro: Axcel Books do Brasil.
- Jamil, G. L. (2005). *Gestão da Informação e do conhecimento em empresas brasileiras: estudo de múltiplos casos*. Belo Horizonte: Ed. Con / Art.

- Jamil, G. L., Santos, L. H. R., Alves, M. L., & Furbino, L. (2012). *A design framework for a market intelligence system for healthcare sector: a support decision tool in an emergent economy*. In *Handbook on research of ICTs for Social Services and Healthcare: developments and applications*. Hershey, PA: IGI Publishing.
- Jamil, G. L. (2014). Inteligência de Mercado como um processo de gestão da informação e do conhecimento: proposta de oficinas de capacitação setoriais. Post-doctoral Project at Faculdade de Letras from Universidade do Porto, Porto University, Portugal.
- Jeston, J., & Nellis, J. (2007). *Business Process Management: practical guide to successful implementations* (2nd ed.). Butterworth-Heinemann.
- Johnson, J. E. (2012, July-August). Big data + Big Analytics + Big opportunity. *Financial & Executive*, 51-53.
- Kim, W. C., & Mauborgne, R. (2015). *Blue Ocean Strategy: Expanded Edition*. Harvard Business Review Press.
- Kimball, R., & Ross, M. (2010). *Relentlessly Practical Tools for Data Warehousing and Business Intelligence*. John Wiley and sons.
- Kotler, P., & Keller, K. (2005). *Marketing Management* (12th ed.). Prentice Hall.
- Laudon, K., & Laudon, L. (2009). *Management Information Systems* (11th ed.). Prentice Hall.
- Lucas, H. C. Jr. (2005). *Information technology: strategic decision making for managers*. Hoboken, NJ: John Wiley and Sons.
- Mahrt, M., & Scharkow, M. (2013). The value of big data in Digital Media research. *Journal of Broadcasting & Electronic Media*, 57(1), 20–33 DOI:10.1080/08838151.2012.761700
- Makadok, R., & Barney, J. B. (2001, December). Strategic Factor Marketing Intelligence: An application of information economics to strategy formulation and competitor intelligence. *Management Science*, 47(12), 1621–1638. doi:10.1287/mnsc.47.12.1621.10245
- Marchand, D., Kettinger, W., & Rollins, J. (2001). *Making the invisible visible: how companies win the right information, people and IT*. Wiley.
- McAfee, A., & Brynjolfsson, E. (2012). Big data: the management revolution. *Harvard Business Review*, 90(10), 60-68.
- Mintzberg, H., Alhstrand, B., & Lampel, J. (2009). *Strategy Safari: the complete guide through the wilds of strategic management* (2nd ed.). Pearson Education Limited.
- Nonaka, I. (2008). *The knowledge creating company*. Harvard Business Review Classics.
- O'Brien, J., & Marakas, G. (2008). *Management Information Systems*. Irwin: Mc Graw Hill.
- Ohata, M. & Kumar, A. (2012, September). Big Data: A Boom for Business Intelligence. *Financial Executive*.
- Park, S. H., Huh, S. Y., Oh, W., & Han, S. P. (2012). A social network-based inference model for validating customer profile data. *MIS Quarterly*, 36(4), 1217-1237.
- Porter, M. (2008). *On Competition*. Harvard Business School Press.
- Russel, S., & Norvig, P. (2009). *Artificial intelligence: A modern approach* (3rd ed.). Prentice Hall.
- SCIP. (2012). *What is competitive intelligence*. Retrieved from <http://www.scip.org/content.cfm>
- Stair, R., & Reynolds, G. (2009). *Principles of information systems*. Course Technology.

- Turban, E., Mc Lean, E., & Wetherbee, J. (2002). *Information technology for management: transforming business in the digital economy* (3rd ed.). Hoboken, NJ: John Wiley and Sons.
- Turban, E., Rainer, R. K. Jr, & Potter, R. E. (2007). *Introduction to information systems*. Hoboken, NJ: John Wiley and Sons.
- Tuomi, I. (2000). Data is more than knowledge: Implications of the reversed knowledge hierarchy for knowledge management and organizational memory. *Journal of Management Systems*, 16(3), 103-117.
- Van Kesteren, J. (2012). How a market intelligence system can add value to the decision making process of NDP and sales: A design science approach (Master Theses of School of Industrial Engineering). Eindhoven University, Netherlands.
- Vavra, T. G. (2013). The advantage of aftermarketing. *American Salesman*, 58(5), 13-18.
- Weiss, S. M., & Verma, N. K. (2002). A System for Real-time Competitive Market Intelligence. *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, 360-365. doi:10.1145/775047.775100
- Kahaner, L. (1998). *Competitive Intelligence: How to gather, analyse, and use Information to move your business to the top*. New York: Touchstone Books.
- Kaplan, R., & Norton, D. (2007). *The execution premium: linking strategies to operation for competitive advantage*. Harvard Business School Press.
- Kearns, G. S., & Lederer, A. L. (2003). A resource based view of IT alignment: How knowledge sharing creates a competitive advantage. *Decision Sciences*, 34(1), 1-29. doi:10.1111/1540-5915.02289
- Markovitch, D. G., Steckel, J. H., & Yeung, B. (2005). Using Capital Markets as Market Intelligence: Evidence from the Pharmaceutical Industry. *Management Science*, 51(10), 1467-1480. doi:10.1287/mnsc.1050.0401
- Schiffman, L., & Kanuk, L. (2010). *Consumer behavior*. Prentice Hall.
- Soares, C., & Carvalho, A. C. P. L. F. (2013). *Novas tendências em Marketing Intelligence* (P. Campos & P. Q. Brito, Eds.). Ed. Actual.

ADDITIONAL READING

- Akbar, H. (2003). Knowledge Levels and their Transformation: Towards the Integration of Knowledge Creation and Individual Learning. *Journal of Management Studies*, 40(8), 1997-2021. doi:10.1046/j.1467-6486.2003.00409.x
- Dotzel, T., Shankar, V., & Berry, L. (2013, April). Service Innovativeness and Firm Value. *JMR, Journal of Marketing Research*, 50(2), 259-276. doi:10.1509/jmr.10.0426
- Jamil, G. L. (2013). Sistemas de informação e a tomada de decisões: estudos de caso sobre a inteligência de mercado e a gestão de marketing. In P. Campos & P. Q. Brito (Eds.), *Novas tendências em Marketing Intelligence*. Ed. Actual.

KEY TERMS AND DEFINITIONS

Competitive Intelligence: An organizational process which aims helping to solve strategic problems and decisions through data and information collection, validation, analysis and results production.

Knowledge Management: A process that intends to administrate the creation, retention and registering, sharing, valuating, monitoring and application of knowledge in organizations.

Marketing: An integrated organizational discipline oriented to aggregate value for product and services, considering customers wishes and needs.

Market Intelligence: A continuous organizational process that intends to produce knowledge for strategic marketing planning, collecting data from a value-aggregate chain, validating, processing and communicating the final results in a standardized way.

Strategic Marketing Planning: A part of the marketing planning process which intends to connect the strategic goals expressed in organizational

strategic planning to tactical marketing specifications, which will lead to product or services definitions, prices, promotion and distribution.

B

Category C

Civil Engineering

Digital Animation for Representing Architectural Design

C

Roberta Spallone

Politecnico di Torino, Italy

INTRODUCTION

Since the late Nineties of XX century digital movies have been emerging as one of the main methods for submission of architectural designs at international competitions, public presentations and shares on contemporary masters' websites, allowing to preview buildings, spaces and urban environments before their eventual construction.

The chapter highlights how the ability to access the fourth dimension, through the construction of a sequence of images, constitutes a specific prerogative of the digital representation, which goes beyond the static constraint imposed by conventional methods of representation codified by the Descriptive Geometry in the previous centuries.

The great complexity of architectural video production has led to the birth of new professionals who specialize in modelling, rendering, animation, graphics, montage, editing and post production and work in specialist studios. In the past, this had resulted in the generation of specific and recognizable languages: the language of the design was overlaid by that of its narration, and they were not always in harmony one with the other.

The author investigates on the methods, techniques and languages of the fourth-dimensional communication of architecture, unexplored area of research thus far, already the subject of previous studies, relating them with the masters' personal poetics.

The chapter provides also analysis and critiques of several case studies collected by the author from the beginning of the phenomenon, and traces an ideal interpretation path, due both to the changing technologies and the emerging specific languages.

BACKGROUND

Digital modelling has affected the entire design process in recent years, starting with the conception through the control and communication of designs and finally to the provision of tools for building construction and then management activities. For the design stage of the process, these new modelling tools seem to have unified the constituent elements of drawing and design, reconstituting them within the etymological roots of the Latin term *designo*, the twofold meaning of which alludes to an abstract component, indicating the mental processes intrinsic to design, and another concrete component relating to the tangible operation of drawing (Garzino, Spallone, & Lo Turco, 2011).

These tools offer a possible answer to Rogers's wishes, who in 1990 said that "what is needed... is for the meaning of drawing (and therefore also the pedagogical and practical interpretation that is derived from it) not to be just the descriptive representation of an object generated by means of symbols, but for its ancient semantic value to return to it, which is summed up today by the English word design (as opposed to drawing), in which the symbols themselves contain the concept of thought, concrete intention and the design of an object" (Rogers, 1990).

Digital models have a syncretic nature which offers new opportunities to the scientific research and architectural design: tests and errors occur in a space in which our experience of problems are rendered fluid and immediate as in an architectural promenade, "3D computer models... allow more enhanced and more controlled interaction

between users and models, they are able to cover the whole range of possible models in a single system of representation” (Maldonado, 2005).

Moriconi, an earlier scholar of infographics, underlined that “through the infographic support, digital drawing simulates a hypothetical reality, goes beyond the limitations of the static and allows interaction with any type of sign. With the creation of virtual images, infographic technology represents what is perhaps the most appropriate tool for interpreting the complexity of reality” (Moriconi, 2001).

According to Manovich the rise of the movie camera as a universal paradigm for the interaction with any data which is represented on three dimensions goes back to the 1980s and 90s. Indeed he affirms that “as the computer culture is gradually spatializing all representations and experiences, they become subjected to the camera particular grammar of data access. Zoom, tilt, pan and track: now we use these operations to interact with data spaces, models, objects and bodies” (Manovich, 2011).

As a new medium, the animation of digital models requires the establishment of relationships with several disciplines such as communication sciences and cinema engineering and with technologies like that of video games and it must also pass the test of comparisons with the established conventional technologies of film production. “Considering architecture as a form of art, we might learn from other artistic disciplines, such as moviemaking (cinematographic approaches, sequencing and animation), theatre (physical expression, interaction, improvisation) and music (rhythm, harmonic variation, but also digital recording and sampling). These may expand the palette of architecture (traditionally making use of drawings, models, pictures and symbols)” (Breen & Breen, 2011).

This triggers a critical discussion on the ontological nature of films, on their narrative form and on their character of exploring human emotions, and it implies the attention of videos

of architecture to the perceptual effects produced by relations between persons and space.

Three dimensional digital models frequently constitute the basis for the production of particularly effective visualizations, using a variety of techniques, which can be exploited for different purposes. As Engeli noted, “the creation of specific messages relating to space requires an in-depth knowledge of the intrinsic characteristics to the different possibilities and to the aspects that are wished to emphasize.

An object that is viewed can be interpreted in different ways. The main parameters (light, materials, perspective) can bring out the nature of an object, work against it, reveal new information and add or hide aspects of the object” (Engeli, 1999).

Engeli also underlines the role of narration in the communication of a design: “when you talk about a design, the aim is to give the public an opportunity to identify with the object that is being presented. This is the main reason why it is necessary to tell stories about designs. In reality all architects perform a similar operation when they present design ideas, but usually it is not a capability that is used knowingly, despite the importance of that activity for the communication of ideas” (Engeli, 1999).

The producers themselves, who include multidisciplinary groups, composed of architects, graphic designers, modellers, producers and musicians, underline the importance of animations for the representation in architecture. In an old interview, Scott, Hampton, Alsop and Cocke, members of Squint/Opera, one of the most famous creative company, underlined the importance of videos in the presentation of a design: “I would like to encourage all architects to tell the story of their designs by using videos, because it is a tool which helps people to see a design in a different way. Our experience of architecture is dynamic: we move through spaces and it is precisely this aspect that a film is able to explore” (Oddo, 2004).

Taking the part of the stakeholders, also today the foresight of Gregotti, who observed that “who

has had experience like a member of architectural competition jury panels composed by different experts, knows very well which confusion has been done between the value of design and the fascination of representation considered as a painting... we will see in any year in which way video and virtual representation will act on communication in architectural competition" (Gregotti, 1995) seems to the author surely alive.

MAIN FOCUS OF THE ARTICLE

Issues, Controversies, Problems

For over fifteen years the 'spectacle of architecture' has been identified in the digital representation of space and time, especially in the animation, one of the most effective media for the prefiguration and communication of architectural design.

The author, following the development of the architectural representation from the outset through digitally constructed video believes that, with some simplification, after a first moment of the centrality of digital modelling within which paths were created or singular elements of the building were animated a following moment in which photo-realistic and hyper-realistic results were chased and another, still, characterized by the pursuit and consolidation of narrative styles, is witnessing nowadays a phase of reflection and refinement, with a return of the communication process production within the atelier of architecture, which addresses, therefore, technical choices and languages consistent with its design poetics.

The current trend, in fact, is to turn increasingly to the opportunities offered by the motion graphics, through the hybridization between static and

dynamic images of different nature, photorealistic renderings tend to be singular, as they are snapshots within conceptual dynamic views, and a search for greater deepening of content takes advantage of the development of the narration on motives, concept phases and constructive solutions.

The below developed discussion, traces, through the comments of the peculiarities of several of the videos found online (not always available even today) and collected by the author over the years, an ideal path built on some case studies illustrating the developments outlined above.

As said, at the beginning of the phenomenon the centrality of digital modelling is evident in the construction of the video.

Zaha Hadid is one of the first architects who use the digital tools for the design, presentation and communication via video of the project. Central themes of her poetry, such as dynamism, fluidity, transparency, found a solution to the problems of design and representation in the digital modelling.

Also the animations of the three-dimensional models offer an opportunity to overcome the two-dimensional characteristics of her drawings and paintings and to move around buildings generated by the computer while they are still in the design phase (Schumacher, 2004). A video realized in 1999, the first that the author was able to find, in which the techniques of animation of three-dimensional model are used in the project of architecture, shows the spatial characteristics of the architectural competition proposal for the Casino and the Grand Hotel in Lugano submitted by Hadid. In the short video (0:58), produced by the computer graphics studio Neutral, which is also one of the pioneers of the specific professional field, a lava flow cools and consolidates into a new architectural form that wraps the existing

Figure 1. Zaha Hadid, Casino and Grand Hotel in Lugano

Source: Movie by Neutral (0:58), 1999



buildings. The model is rendered with only two colours, red and white, in order to resume the metaphor of lava and ice, in complete harmony with Hadid's the figurative language.

In subsequent years, the improvement of the techniques and digital tools accompanies the transition from animation resolved within the conceptual modelling to an extreme tendency toward photorealism.

The availability of software used in computer-gaming, along with the launch of YouTube in 2005, promotes the proliferation and spread of design representation through video, in particular in the field of international architecture competitions, in which usually it becomes one of the mandatory requirements, in public presentations and in self-promotion of the main architectural studios through websites. The tendency to create photorealistic images aspiring to become perfect copies of reality establishes itself at this stage, in which rendering engines, textures of materials and systems for setting the lights are improving.

The video for the project by Christian de Portzamparc in the competition for the headquarters of the Region of Rhône-Alpes in Lyon (2006) describes a building in an urban scale, a complex designed as a block, a place that can be enjoyed by the community in democratical and active way. The project won because of its elegance, fluidity and rigor and the particular suitability to the site (Cardani, 2006). In the video (3:12), composed by the studio Bartproject, a virtual man takes viewers into the building suggesting the functions of the different spaces through the movements of the body and emphasizing the role of the building as a business and trade centre. The constant speed of the virtual camera movement simulates the human walking. The video uses a communicative style,

typical of Bartproject, which combines shooting extrapolated from the reality with a hyper-realistic rendering that overlaps transparent figures symbol of an ethereal humanity in motion. The animation shows the transparency and openness of the interior spaces - the monumental entrance, the patio, the exhibition space and the room for debate, all designed to minimize the power consumption.

Architectural competition for the *Musée des Confluences in Lyon*, won by Coop Himmelb(l)au in 2001, awards a project aimed at the needs of information-society: to make perceivable current knowledge to a broad public in a process of constant change.

The concept, named Crystal and Cloud, combines a hard space, the crystal rising towards the city, conceived as an urban space and an entrance hall for visitors, and a soft space, the cloud intended by the designers to hide the knowledge of the future (Coop Himmelb(l)au, 2003). A short movie presenting the project has been included in the Coop Himmelb(l)au website, since 2007. This was probably the basis for the longer and richer educational movie now present in the Musée des Confluences website which is also enriched by an interactive virtual visit with 360° panorama views in the principal rooms.

The longer movie (7:30), produced by Isochrom in 2006, starts exploring the area of Confluences by an aerial view, first zenith and then perspective, which shows a wide urban tissue extended to the town hall, modelled with simple volumes and rendered as a plaster model lit by natural light. As the approach is made from the river to the building, the model takes on greater realism and the movement of the camera transforms into a walk through towards and inside the building under design. A virtual person guides the observ-

Figure 2. Christian de Portzamparc, New Rhône-Alpes Regional headquarters in Lyon

Source: Movie by Bartproject (3:12), 2006

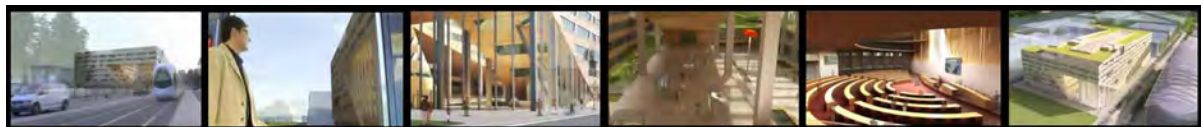
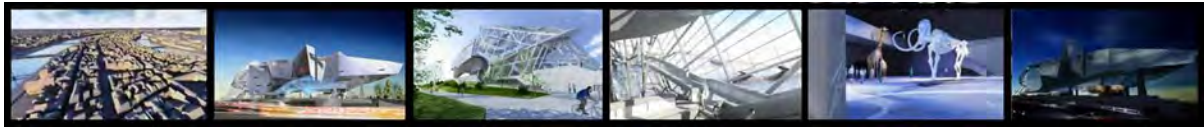


Figure 3. Coop Himmelb(l)au, Musée des Confluences in Lyon

Source: Movie by Isochrom (7:30), 2006



ers through the main halls to the different floors and invites them to appreciate the architecture and examine the main collections by the way the shots are framed. During the tour the visitors meet “Crystal and Cloud” which characterize the shape of the building. The model is realistically, but delicately, rendered with a particular focus on bringing out the quality of the materials, transparencies, reflections and opacities and interactions between the materials and the lights. A few blobs of warm colours attributed to the entrances of some halls, arousing the curiosity to enter them, and to secondary animation elements, emerge in the interiors in cold colours, all ranging from white to grey and blue.

In the following period there is a passage from the technical virtuosism to the research and consolidation of narrative styles.

This step intended to prevent the technocratic homologation by the software, and led the video producers to create personal and recognizable narrative languages. As stated by the following three case studies, this choice anticipates the current developments.

The submission by Steven Holl for the international architecture competition to design the future Musée du Louvre-Lens in the north of France, launched in early 2005, is a building in which the idea of a walk through time organizes the plan of the new museum according to a linear

time increment along the rectangular frame, which is crossed by cyclical time (represented by the arcs) in short circuits and different access points.

The competition animation (4:00), made by Neutral, deepens the concept ideas, showing the aquarelle sketch drawn by Holl as the basis for the three-dimensional geometric construction which develops the idea of intersection between linear time (a parallelepiped) and cyclical time (four cylindrical rings cut to make four arcs). When the geometries become buildings, the parallelepiped appears as a hard block of bricks and the arcs as light curtains of glass. The architecture, surrounded by the green, is connected to the built city in which it is wedged. The new museum and the settings are modelled and soft rendered. The camera movement, characterized by constant speed, guides the spectators outside (by a fly trough) and inside (by a walk trough) towards the knowledge of the design. Therefore most of the animation is produced by modelling software, with consequent use of bigger time.

The proposal for the invited architectural competition for a new library in Utrecht (2008), presented by 3XN, describes a multifunctional building suitable to house a library and a cinema with residences, shops and cafes. In this way, the cultural centre could be enlivened by various activities throughout the entire day. The architects' goal is to preserve the existing idea of public

Figure 4. Steven Holl, Musée du Louvre-Lens

Source: Movie by Neutral (4:00), 2005

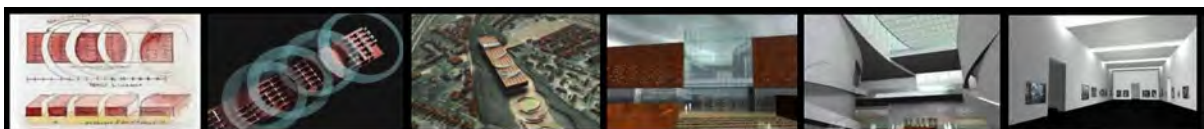


Figure 5. 3XN, New library in Utrecht

Source: Movie by Cadpeople (3:30), 2008



life in the area, while the new building should be in function as an anthill, of which the shape resembles, characterized by multiple fronts, and should be a new meeting place for the entire city.

The 3XN claims to be deeply involved in the digital revolution (3XN, 2008) and this seems to reverberate in their choices of projectual communication, entrusted to the studio Cadpeople, demonstrating the close cooperation between the two groups. In particular, the video (3:30), resolved entirely thanks to the digital modelling, shows an original visual and communicative approach. It starts highlighting the relationship between the building and the environment, the different functions, displayed with exploded isometric, the accessibility system. Then, describing the project, a particular camera movement alternating the acceleration and stop-image is used, combined with special representation techniques that change from the conceptual, in gray scale with a few spots of primary colours, secondary animations of people, vehicles, videos, up to static photorealistic images delicately rendered.

Opposed to the essentiality carried out in the previous case study, the imaginative language, rich in contaminations between real and virtual can be found in the next case.

The animation for the presentation of the HOK Sport Olympic Stadium (2007), to be built for the

Olympics and Paralympics in London, was created by the filmmakers of Squint/Opera.

The film is produced through hybridization of materials and languages, real footage and photographs with digital artefacts; the animation of three-dimensional model is reduced to a minimum in favour of applying dynamic post-production tools to the static views.

A short presentation, which simulates a communication of television news, introduces the project, firmly framed in the atmosphere of London through the city landmarks (the City Hall, the Tower Bridge and the Battersea Power Station) progressively invaded by the flags of the Olympic competition. A child spectator from the beginning to the end of the event takes the viewer into the discovery of the project elements. These are realized through geometric models of roofs, services, seating and structural system that immediately become, through animation, in other objects related to the fantasies of the child: the colourful ellipsoidal shapes of the services become flying ships, the seats in the form of gulls arrive from above, the three-dimensional trusses move to the place of foundation as they were snakes, while in the background there is the real city. Finally, the elements reassembled and made static, assume again their geometric appearance related to the environment through a model illustrating the overall functions and accessibility. A sequence in

Figure 6. HOK Sport Olympic Stadium, London

Source: Movie by Squint/Opera (4:52), 2007



Figure 7. AllDesign, Great Beijing Theatre and Masterplan, Langfang

Source: Movie by AllDesign (2:44), 2010



which the realistic render of the three-dimensional model is set in motion through secondary animations of public, athletes, giant screens and actions of zoom, scroll, rotate, concludes the movie.

Among the current trends of the animation for the design representation, as seen in the description of the emblematic case studies, the minimalist choices consisting in conceptual views, in animations that refer to the simple movements indicated above by Manovich, in attention to the concept communication and the artefact constructability, can be affirmed to prevail.

Moreover, the video production within the same study of architecture is witnessed in many cases, facilitated by the most common software for modelling, rendering and post-production.

Differently from the most of the architects group websites, that led by Will Alsop, which name is AllDesign, shows his projects by movies besides static renders. These movies at first realized by Squint/Opera or Virtual Artworks are nowadays produced inside the studio.

The choice of attributing a central role to the movies in the communication strategy of the architect's team, in tune with the graphic style of his website, stimulates more involvement by the viewers.

The video for presenting the project for the complex of three theatres in Langfang, near Beijing, in 2010, was produced by the same AllDesign

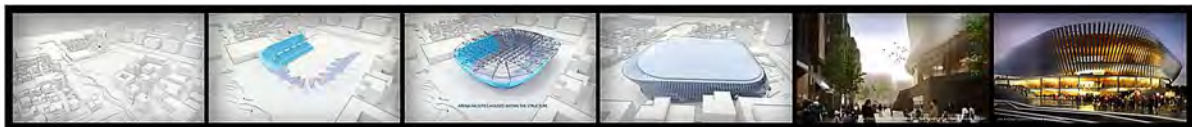
studio that, in its website, claims to create movies since the best way to express their own ideas is telling a story.

The short video (2:44) introduces the narration with a recall to some images of the traditional Chinese landscape: vegetation of bamboo, flowering cherry-trees, canoes on the lakes and, finally, the wild boar, of which metamorphosis, described as the initial concept, gives shape to the project. The complex aims to fit into the natural environment with the creation of a park with ship canals and walkways leading to theatres and shady spots designed as a meeting place. The architectures, displayed in conceptual clay-render, on which the connecting blue ring raises, are part of the system of the green. Animations, fly-through, static rendering zoom and scroll and simulations of the growth of the artefacts and the green, combined with the apparition of evocative images (musical notes that come out of the buildings and moulds of traditional theatre masks) constitute the sequence.

The video for the architectural competition in 2012 for Copenhagen Arena, presented by C.F. Møllers, shows a multifunctional complex intended for sports and concerts, without losing its human scale, becoming a landmark, a building element of the landscape and an attractive element of activity. The animation (3:54), presumably produced by the same architectural studio, initially presents, in a bird eye view, the structural

Figure 8. C.F. Møllers, Suggestion to Copenhagen Arena

Source: Movie by C.F. Møllers (3:54), 2012



technologic and constructive elements and the building phases of the stadium, set within a context geometrically modelled and clay-rendered. When the video camera is lowered to the human level, its recurrent motion is the rotation around the conceptual model. The following still image is characterized by the representation as a realistic static rendering to which actions of scroll up or zoom out are slowly applied.

SOLUTIONS AND RECOMMENDATIONS

In summary, methods, techniques and figurative choices in the ambit of the architectural animation, highlighted by the selection of case studies proposed above, can be synthesized as follows.

One or more geometric models of a design having an appropriate degree of detail in relation to the positions from which they are filmed, which are based on figurative convention for the architectural representation: dynamic bird eye views emphasize relationships between the building and the existing environment, while those at human eye height give a preview of the perceptual impact of the future settlements; orthogonal views of horizontal and vertical sections illustrate the space distribution and the functional aspects.

The figurative decisions reconcile the design poetics of the architects on the one hand and the narrative styles of the producers on the other, fluctuating between conceptual and hyper-realistic styles.

The camera movement is not generally at a constant speed, accelerating and slowing, sometimes moving rapidly on tracks to get close to the action or following the pace of an avatar who guides the spectator in the discovery of the place, while secondary animations, of men and vehicles in movement, monitor which projects films and fountains and fireworks overlap the primary animation or a still image of it.

The time dimension, introduced by the movement of the camera, is frequently overlapped by

the representation of the natural flow of time, with animations which start in the daylight and end at the night, thereby facilitating the entire daily life cycle of the object.

Animated projections are always accompanied by sounds. They can range from background music which captures the attention and helps focus the observers' attention on the noises which enliven scenes evoking feelings of empathy and the recognition of places.

Finally different video production techniques can be observed, finding their place of convergence in the digital world (Ciotti & Roncaglia, 2000): the composition of different two and three dimensional views of digital models with photographs and texts using the established technique of stop-motion combined with movement options such as fades, zooms, pans and scene changes which are practiced today using dynamic postproduction software; the modelling of building designs in their existing context, also modelled, but often in a summary manner and the formation of a long route along which a series of rendered views are created; the modelling of building designs inserted dynamically in a film of the environment using camera-tracking techniques.

FUTURE RESEARCH DIRECTIONS

Architectural videos, today with a greater presence, sometimes are accompanied, above all in the web-sites, by the extremely recent and occasional introduction of 360° panoramic views, which offer a modest degree of interaction, and by virtual views, which, on the contrary, offer a certain number of scenarios that can be freely selected and navigated by visitors.

CONCLUSION

The rapid development of digital technologies makes foreseeable a greater interactivity between observer and digitally simulated space, with the

ability not only to freely explore the project, but also to modify it in real time. The author believes that in the same time the research on languages should advance, with the aim of increasing the wealth of knowledge provided by the animation, not only with respect to the users, but also to the designers themselves who could find another effective verification tool and project control thanks to it.

REFERENCES

- Breen, J., & Breen, J. (2008). Critical Observations and Strategic Perspectives at Half-time. In K. Terzidis (Ed.), *Proceedings of First International Conference on Critical Digital: What Matter(s)?* (pp. 129-136). Cambridge, MA: Harvard University Graduate School of Design.
- Cardani, E. (2006). Proposta vincente. From Lyon Confluence: Progetto, Christian de Portzamparc. *Arca*, 219, 91.
- Ciotti, F., & Roncaglia, G. (2000). *Il mondo digitale. Introduzione ai nuovi media*. Roma-Bari, Italy: Laterza.
- Coop Himmelblau (2003). *Musée des Confluences*. Retrieved December 17, 2008, from http://www.arcspace.com/architects/coop_himmelblau/musee_des_confluences/
- Engeli, M. (1999). *Storie digitali. Poetiche della comunicazione*. Torino, Italy: Testo&Immagine.
- Garzino, G., Spallone, R., & Lo Turco, M. (2011). Strategie digitali per modelli conoscitivi / Digital strategies for knowledge based models. In G. Garzino (Ed.), *Disegno (e) in_ formazione. Disegno politecnico / Drawing (and) information. Polytechnic drawing* (pp. 70-111). Sant'Arcangelo di Romagna, Italy: Maggioli.
- Gregotti, V. (1995). Editoriale. *Rassegna*, 61.
- Himmelblau, C. (2003). Musée des Confluences. *GA Document*, 73, 12–17.
- Maldonado, T. (2005). *Reale e virtuale*. Milano, Italy: Feltrinelli.
- Manovich, L. (2001). *The Language of New Media*. Cambridge, MA: MIT Press.
- Moriconi, C. (2001). Il modello virtuale. In A. Soletti, P. Belardi, & F. Cataliotti (Eds.), *Claudio Moriconi professione infografico* (pp. 20–30). Perugia, Italy: Università degli Studi di Perugia.
- Negroponte, N. (1995). *Being digital*. New York: Alfred A. Knopf.
- Oddo, F. (2004). *Squint/Opera. Nuove professionalità per raccontare l'architettura contemporanea*. Retrieved October 3, 2015, from <http://architettura.it/files/20040123/index.htm>
- Rogers, E. N. (1990). *Gli elementi del fenomeno architettonico*. Napoli, Italy: Guida Editore.
- Schumacher, P. (2004). *Hadid digitale. Paesaggi in movimento*. Torino, Italy: Testo & Immagine.
- Spallone, R. (2014). Vero, falso, finto. Rappresentazioni digitali di paesaggi urbani in trasformazione fra comunicazione e seduzione [True, false, fake. Digital representations of urban landscapes subject to transformation between communication and seduction]. *DisegnareCon*, 13, 1–8.
- Spallone, R., Lo Turco, M., & Sanna, M. (2009a). Representing ideas by animated digital models in architectural competitions. *Proceedings of ARCC 2009. Leadership in Architectural Research, between academia and the profession*.
- Spallone, R., Lo Turco, M., & Sanna, M. (2009b). Dimension “time” to represent contemporary architectural design. *Proceedings of IDCT. Innovative Design & Construction Technologies. Building complex shapes and beyond*, 243-258.

Spallone, R., Lo Turco, M., & Sanna, M. (2010). From the place of research to the research of place: new digital technologies for simulating future architectures in building environments. *Proceedings of ARCC 2010. International Conference on Architectural Research*.

3. XN. (2008). *Digital architecture*. Retrieved May 2, 2010, from <http://www.3xn.dk>

ADDITIONAL READING

Levy, P. (1998). *Becoming virtual. Reality in the Digital Age*. London, England: Persius Books.

Lynn, G. (1999). *Animate form*. New York, USA: Princeton Architectural Press.

McLuhan, M. (1964). *Understanding media. The extension of man*. Berkeley, California: Gingko Press.

Sacchi, L., & Unali, M. (2003). *Architettura e cultura digitale*. Milano, Italy: Skira.

Saggio, A. (2007). *Introduzione alla rivoluzione informatica in architettura*. Roma, Italy: Carocci.

Szalapaj, P. (2005). *Contemporary Architecture and the Digital Design Process*. Massachusetts, USA: Architectural Press.

K. Terzidis (Ed.). *Proceedings of First international Conference on Critical Digital: What Matter(s)?* Cambridge, Massachusetts: Harvard University Graduate School of Design.

KEY TERMS AND DEFINITIONS

3D Modelling: In 3D computer graphics, 3D modelling is the process of developing a mathematical representation of any three-dimensional surface of object via specialized software. The product is called a 3D model.

Animation: Animation is the process of creating a continuous motion and shape change illusion by means of the rapid display of a sequence of static images that minimally differ from each other. In the architectural representations, animation can be the creation of a path inside and outside of buildings, or the putting in motion of static elements.

Compositing: Compositing is the combining of visual elements from separate sources into single images, often to create the illusion that all those elements are parts of the same scene.

Montage: Montage is a technique which uses rapid editing, special effects and music to present compressed narrative information.

Post-Production: Post-production is part of the process of filmmaking, video production and rendering, which consists in a series of editing and optimization of the product, subsequent to the camera shot or the animated sequence creation.

Rendering: Rendering is a image generated by a computer using three-dimensional modeling software, after applying materials and lights and setting a perspective view.

Secondary Animations: In architectural movies, the secondary animations consist of motions of people, vehicles, videos which overlap the animation created by the video camera movement.

Literature Review of Augmented Reality Application in the Architecture, Engineering, and Construction Industry With Relation to Building Information

C

Aydin Tabrizi

University of Kansas, USA

Paola Sanguinetti

University of Kansas, USA

INTRODUCTION

Productivity rate in construction industry shows a general pattern of decline in comparison to other industries. This issue influences on performance and efficiency of construction projects by adding unnecessary costs, time, materials, and manpower waste Alwi, Hampson et al. (2002). In dealing with this issue, it is necessary to apply proactive approaches rather than reactive ones through using new method and processes in the construction industry. During the last decade, Information and communication technologies (ICT) have been advanced significantly where the application of these technologies could improve the construction industry efficiency to some level (Park & Kim 2012). However, in order to fulfill this task in a larger scale, new methods and processes are required to develop and analyzed. This paper presents a conceptual framework to enhance construction industry efficiency via a comprehensive and proactive mechanism of Augmented Reality (AR) and Building Information Modeling (BIM) linkage. To fulfill this objective, the study begins with an extensive and critical review on AR and BIM separately in order to investigate the efficiency of each technique in the construction industry. Then the study proposes a conceptual framework of AR and BIM combination by investigating this collaboration in the enhancement of construction industry procedure.

BACKGROUND

Augmented Reality (AR) provides the means for intuitive knowledge presentation by enhancing the perceiver's situational awareness and cognitive perception of the real world. Through AR approach, virtual objects can be registered in relation to real objects where these objects can be seen in the same position and orientation of other real objects of the scene, as perceived by the user (Wang et al. 2004). In addition the real objects can be tracked and their 3D shape can be reconstructed from pictures (Azuma 1997).

AR has introduced as a technology which allows the user to see, hear, touch, smell and taste things that others cannot (Van Krevelen and Poelman 2010). It is a technology to perceive elements and objects within real world experience in a complete computational environment. It applies creatures and structures that could be used in daily activities unconsciously through interaction with others such as enabling mechanics to see instructions for repairing an unknown piece of equipment, surgeons to see ultrasound scans of organs while performing surgery on them, fire fighters to see building layouts to avoid invisible hazards and people to read reviews for each restaurant on their way (Feiner 2002). (Wang and Dunston 2007) describes AR as a tool allowing users to work with real world environment while visually re-

DOI: 10.4018/978-1-5225-2255-3.ch085

ceiving displays of additional computer-generated information about the item by superimposition of additional information onto the real world scene. This approach enhances the user's perception of the real environment by showing information that cannot be sensed unaided.

It is expected that in the near future, an increase in the use of AR applications will occur due to the advancement of hardware and software. (McKibben and Furlonger 2009) predict that by the end of 2014 approximately 30% of workers will use some form of AR capability and after a long period of technological development and refinement, the implementation of AR applications for the general public will reach its peak. In addition, the commercial market shows a same trend by promising examples such as Project Glass as an R&D program by Google (Goldman 2012) to develop an AR head-mounted display (HMD) for enabling users to experience a truly immersive digital life. ABI Research (Hyers 2006) predicts that by the end of 2014 the revenue from the AR mobile market will reach \$350 M and Juniper Research (Holden 2005) predicts that the market for AR services will reach \$732 M at the same time. It is widely believed that AR technologies are maturing and that within the coming years they will be broadly adopted by industry.

Estimations show that at least 5% of total building construction costs are due to occurrence of problems in the early design process, causing insufficiency, inconsistency and omission of design-related information towards construction phase (Hwang, Thomas et al. 2009). Construction worksite layout planning relies mainly on 2D paper media where the worksite planners sketch the future layout adjacent to their real environment. This traditional approach is ineffective and prone to error because only experienced and well-trained planners are able to generate the effective layout design with paper sketch (Wang and Dunston 2007). In the architecture, engineering and construction industry (AEC), gaps between planned solutions and practical implementations, poor communications between project participants

and inefficient scheduling are the main issues for the lack of sufficient information/communication technology (ICT) support and innovative business procedures (Chi, Kang et al. 2013). (Froese 2005) has categorized trends in construction ICT into three categories. The first category is stand-alone tools to assist specific work tasks such as CAD, structural analysis and estimating programs. The second category includes communication and on-line information sharing tools through worldwide web document management system. Finally the third category includes the potential for integrating first and second category as a cohesive model through Building Information Modeling (BIM) (Froese 2010) where project teams come together to produce comprehensive and virtual prototypes of all aspects of the construction project as the central activity.

With the progress of ICT use in the AEC industry, higher quality visualization platforms are necessary for the efficient use of shared information among involved teams. Available research studies show the attempts of construction activities simulation with feedback generation through visualizing construction information for easier understanding and data share among project participants (Kang, Anderson et al. 2007). However, this approach only enables the visualization of activities in the virtual environment without enough resemblance to the actual tasks in the real world. (Froese 2005) introduces AR approach for the generation of digital project information prior to construction with transfer onto construction site in a fully digital way where this process facilitates the comparison of the actual situation with the planned final appearance to identify the concerned items.

Building Information Modeling (BIM)

(Penttilä 2006) defines BIM as a set of interacting policies, processes and technologies that generates a methodology to manage the essential building design and project data into digital format throughout the building's life cycle. BIM creates an

interdependency environment between structural, architectural and mechanical services (Dossick and Neff 2009) with digital representation of the facility's physical and functional characters to provide a shared knowledge resource for client or user to use and maintain throughout the project's life cycle (Eastman, Jeong et al. 2009).

According to the studies by (Taylor and Bernstein 2009), BIM can incorporate parametric 3D, 4D and 5D modeling where 4D includes a time dimension and 5D contains time based cost model. In addition, studies by (Aouad, Lee et al. 2006) presents the expansion of BIM into nD environment in the recent developments of AEC industry by incorporation of engineering analysis and construction business functions during the lifecycle of a building. This expansion includes scheduling, cost analysis, quality control, accessibility, safety, maintenance, sustainability and energy simulation. However, despite the recent BIM developments in the AEC industry, it still requires further progress during construction phase, specifically in relation to the daily monitoring of work and management of involved teams.

Augmented Reality (AR)

Generally, the life cycle of construction project consists of five phases as a) project feasibility analysis; b) planning/design; c) construction; d) operation; e) disposal (Halpin and Woodhead) where each phase requires the creation of physical structures and elements to create visual information for understanding and communicating the complexity of projects with consideration of relation to existing structures or elements. With recent advances in computer interface design and hardware capability, the application of AR research prototypes and test platforms have fostered in the construction industry (Yang, Tsai et al. 2011). However, AR technologies itself are not advanced enough to be applied effectively in real AEC projects and there are two main issues in this regard and they must be addressed before AR technologies can become prevalent in the AEC industry. The first issue is the limitation of

available technologies to support AR approach and the second is to identify appropriate application areas for the use of AR in the construction projects. To clarify the issue, (Stanco, Battiato et al. 2011) studied the accurate alignment of virtual objects in a real world scene with regards to the user's orientation and position. Their research identified recent tracking technologies for indoor, prepared environments with accurate registration, but the lack of strong accuracies tracking techniques for outdoor environments due to the scale difference. (Yokokohji, Sugawara et al. 2000), demonstrated a system combining accelerometers and video tracking registration for analysis of indoors, but its' non-practicality for outdoor and unprepared environments.

In order to create an efficient application of AR in construction projects, it is necessary to identify the areas with the highest potential for better performance. (Balzani, Santopuoli et al. 2004), demonstrated the potential of AR as a visualization aid for underground structures and (Bruner 1966) discussed the benefits of AR for architectural assembly guidance, infrastructure field tasks, urban planning and design detailing.

Therefore, the analysis of current literature, suggests the feasibility of AR technologies to increase the likelihood of success in the AEC industry with the consideration of its suitability validation in each case.

FURTHER RESEARCH DIRECTIONS

Integration of Augmented Reality into AEC Industry

Research by (Chi, Kang et al. 2013) presents the four rapidly developing technologies in the application of AR approach to integrate the virtual world with the real world as a) Cloud computing environment; b) localization; c) portable/mobile devices and d) natural user interface (Table 1). The application of these technologies plays a significant role in the integration of AR into the AEC industry.

Table 1. Architecture of AR application in 4 phases adapted from “research trends and opportunities of augmented reality applications in aec”

Architecture of AR Application			
Data Phase	Computing Phase	Tangible Phase	Presentation Phase
Cloud Computing Environment	Localization Technologies	Portable Devices	Natural User Interface
BIM	GPS	Cheap	Gesture
Internet	RFID	Light	Motion Capture
	Barcode	Wearable	

Source: (Chi, Kang et al. 2013)

Cloud Computing Environment

Cloud computing has become a powerful technology with the progress in the Internet data processing speed where it has the potential to extend the use of AR and BIM applications in the AEC industry. Research by (Jardim-Goncalves and Grilo 2010) emphasizes the shifts of CloudBIM approach from stand-alone static applications to dynamic shared environment with dynamic tasks allocation and access via a network. Internet operation in real time and cloud-based computer devices helps to make the required information more accessible to the users by enhancing the AR experience with the possibility of virtual information update or change by the cloud side. Therefore, similar to the use of web browser, the involved parties can use the AR as the front-end to explore the virtual information with extra access to more variable and meaningful information in the field. According to (Kamat, Martinez et al. 2010), it can benefit construction field tasks such as schedule monitoring and building inspection where they require huge amounts of dynamic information.

Localization Technologies

(Chi, Kang et al. 2013) expect the localization technologies influence in the future development of AR applications and they introduce it as part of the core functionality of AR technology with its necessity in identifying the posture of subjects. However, the accuracy of the localization method introduces the limitations of AR applications in

superimposing virtual information into the real world. To solve this issue, a research by (Zhang and Hammad 2012) discusses the development of global positioning system (GPS) and outdoor localization technologies such as radiofrequency identification (RFID), ultra-wide band (UWB) and barcoding where all combine multiple sensors to provide the accurate functionality. Despite the current developments, still all available localization approaches are limited by environmental complexity, signal quality, ranges of sensors and uncertainty where it makes it difficult for developers to maintain the accuracy level of AR and to promote the use of AR applications before improving its accuracy.

Portable and Mobile Devices

The size, weight, performance, and cost of hardware directly influence AR applications where the portability of AR depends on the level of mobility a hardware device can provide. This issue requires AR systems focus on portability and providing the ability to maneuver at outdoor environments. Research by (Yang, Tsai et al. 2011) highlights the need for all AR devices to be stable and capable of surviving in the variable environments. Until this capability is achieved, new AR applications will not become available.

Natural User Interface

(Yeh, Tsai et al. 2012) define natural user interface as intuitive control mechanism that imitate human

behaviors and gestures with consideration of an increasing number of AR applications utilizing gesture controls such as Sixth-Sense as a wearable gestural interface that augments the physical world with digital information and enables the users to interact with the digital information using natural hand gestures. However, their research clarifies the dependency of AR application not only on its stability but also on the quality of the control interface. Future research requires the control interface to be more intuitive and more stable with fewer sensing limitations where it will reduce the risk of operating AR devices in dynamic environments such as construction sites.

Combination of Augmented Reality and Building Information Modeling

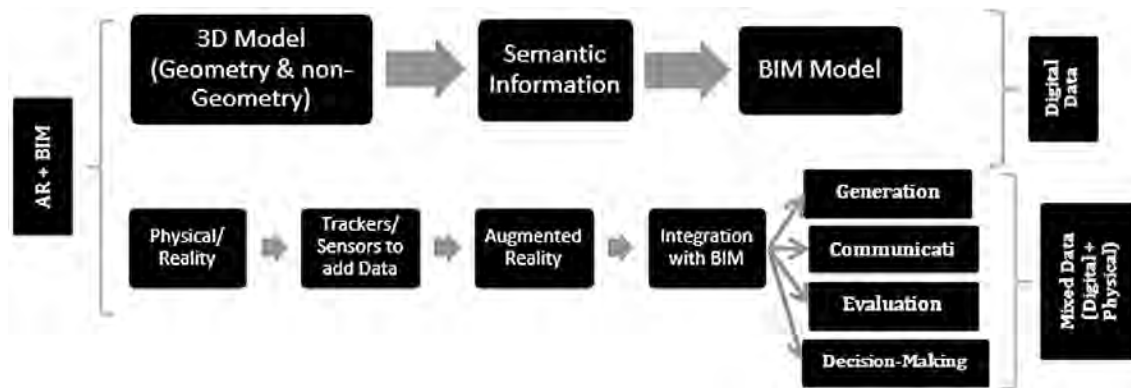
BIM related research has predominantly focused on enhancing communication and collaboration between stakeholders through the use of 3D modeling and 4D simulation throughout a project's life cycle (Love, Edwards et al. 2011) where BIM is utilized simply as a representation and simulation tool. The issues dealing with large quantities of data and their accessibility have hindered the use of BIM being effectively implemented on the construction site and even the issues regarding the application of BIM from design to real-time on-site construction have not explored thoroughly yet. However, the contained information within BIM should be used during construction to ensure that activities and tasks are completed on time according to schedule and to ensure the desired quality and safety standards are met (Construction 2008).

While BIM provides static and pre-defined data and information, AR can be integrated with BIM to enable the physical context of construction activities visualization with real-time monitoring of tasks. The integration of BIM with AR can provide a platform for a site management team and subcontractors to effectively interact and utilize the BIM model contained data (Hou and Wang 2011). This translates to a major opportu-

nity presented by the use of BIM+AR technique as the ability to work at real-scale in all phases. Involved people in the building process, such as clients or prospective buyers, often have difficulties understanding and translating 2D CAD architectural representations into real 3D spatial models where they basically need to build up their individual mental models to understand the project (Donath, Beetz et al. 2001). There is also a need for a powerful and proper visualization approach to supply information to the onsite workers. BIM can make the interdependencies between work tasks more explicit with making the existing complexity more visible and manageable onsite. It can be a proactive approach where the potential negative impacts of any action can be identified earlier and mitigated avoided easily. Spatial collision analysis between trades by 3D modeling systems such as CATIA or Navisworks is an example of this approach (Wang and Love 2012). However, during the actual construction process, there is a possibility of collisions occurrence due to the change of building elements, site facilities and the movement of construction machinery. Therefore, it is challenging to introduce the onsite real-time dynamic collision detection approach to include the variations of construction sequence and schedule. Addition of AR to BIM enables real-scale representations of the existing situation to be communicated and augmented by further information where misunderstandings between planner, client and involved parties can be better avoided with savings on additional cost measures and time-delays (Donath, Beetz et al. 2001).

Therefore, BIM+AR can be proposed as the fourth era of construction IT where prior to construction, the generated digital project information can be brought onto construction site with capability of being processed in a fully digital way. Simultaneously, the site crews will have the capability of accessing the information associated with the concerned construction component or entity in a fully digital approach (Froese 2005).

Figure 1.



Framework of Building Information Modeling and Augmented Reality Integration

(Wang, Love et al. 2013) define AR as an information aggregator to collect and consolidate information from BIM model with sensing and tracking technologies such as barcode, RFID, and GPS to identify, track and monitor individual resources. The tracking devices are mobile and ideal for onsite use to integrate BIM and AR where tags are attached to elemental components and the progress is monitored by identifying the details about the specific properties such as date, number, and text lists. Since the tags are recorded with predefined activities with the requirement of constructing a specific component and given capability to the site operator to record comments of each activity, it can be a direct link between the BIM model and AR database where both containing drawings and documents linked to a specific component and element database.

The Generation-Communication-Evaluation and Decision making (GCED) cycle was propagated by (Arayici, Coates et al. 2011) as referring to the typical onsite decision making process. In this approach a potential solution is generated before its being communicated and being aware of the potential solution, its evaluation can commence based on a set of pre-defined criteria and then decision can be made. As an example, the architects who design the building envelope interact with

engineers developing the steel structures and when architects and engineers engage in discussions related to complex geometrical relationships such as facades, the GCED cycle commences (Figure 1). In the conventional approach, physical mock-ups are created considering their time-use and inaccuracy where many features and properties are lost. Another method is creation of computer-generated models prior to discussions with the consideration of their ongoing insufficiency for evaluation and collaboration purposes. Integration of BIM with AR enables the visualization of buildings 3D models with their detailed features where their properties can be visualized directly onsite to support AEC team in their communication and dynamic generation of alternative site solutions.

Similarly, in second study by (Froese 2010, Arayici, Coates et al. 2011), the framework for integration of BIM and AR is defined in four steps. First, the design and planning of construction process starts with the creation of digital prototypes in BIM model where it contains both geometric and non-geometric information. Second, the BIM model is used as the reference guide to organize the production process. Third, each involved party carries out their role according to the drawing information from the same AR based BIM model where the models are used to support effective interaction and communication. Fourth, the work results are used to update the same BIM model through the function of AR annotation and commenting.

DISCUSSION

(Behzadan, Timm et al. 2008) discuss the recent developments in mobile processing units, camera quality and tracking technology where they enable AR applications to be implemented in different environments. Available mobile augmented information has various applications at the construction site including construction work planning, verification, training and safety, communication and marketing prior to construction work. Seamless interaction cycle between the real world and digital building information model developed by (Hakkarainen, Woodward et al. 2009) identifies the camera location in order to implement mobile location based visual feedback from the construction site to the BIM system with full awareness of the user's location in time and space.

(Irizarry, Gheisari et al. 2012) developed a low-cost mobile AR based tool for facility managers called InfoSPOT as a mobile AR method for accessing building information through a situation awareness approach where a real world example of a BIM was applied to test its development. The authors discovered the challenging problem of the BIM being inconsistent with the built environment and in order to solve this issue, they optimized the geometry to reduce the BIM complexity in displaying on a mobile device. Their statistical analysis indicated a generally positive response for this inexpensive solution and suggested it will assist facility managers in their routine tasks where the real-time views of space can be supplemented with vital information. They concluded that AR solutions can be easily setup and the costly hardware installations were unnecessary for a successful AR application.

(Jiao, Zhang et al. 2012) presented an AR framework integrated with a BIM and cloud computing technologies where they utilized Web3D technology to display virtual objects and resulting solution from integration with an as-built BIM model. The authors suggested the framework capability of being used in real data application where it can support collaboration in a

more natural and effective way with the power of cloud services. They conducted a test case with an actual construction project and their results demonstrated the usability of the proposed framework to source virtual objects from BIM and to support collaboration. They concluded that the extension of the framework as a daily tool for construction applications require further research in the future.

(Park, Lee et al. 2012) proposed using AR for the management of defects during construction where they made an investigation in the construction industry for identifying current issues and requirements of defect management practices. Their studies presented a conceptual system framework integrating AR with BIM model for construction defect management. Based on the laboratory tests, the authors concluded that the BIM+AR system will enable managers to efficiently inspect and control their worker's job performance with allowing workers to readily confirm their works in real-time.

(Kandil, Hastak et al. 2014) proposed the integration of BIM model with AR for ductwork installation and clash detection involved in erecting the HVAC system. Even though the conflicts and clash detection can be largely identified in BIM model during design, design changes and errors or poor installation may lead to conflicts arising onsite. Using integrated AR allows a site manager to address the potential for conflicts onsite by retrieving and visualizing all the concerned properties and details in the BIM model where it can improve speed, safety and accuracy as well as reducing the cost.

CONCLUSION

Despite the limitations of building information modeling (BIM) to the design phase throughout the project lifecycle, it has started to be implemented in the construction industry in the recent years. In the meantime, augmented reality (AR) as a new and emerging technology in the construction industry deems to be a key enabler to address

the current shortcomings of using BIM onsite throughout the construction process. While BIM can be used to improve the efficiency and effectiveness of design coordination, it is unable to take into account the inherent uncertainty associated with design changes and rework happens during construction and particularly in complex projects. With the rapid development and adoption of AR applications, there are numerous opportunities to integrate BIM and AR in order to visualize the information about as-built, as-planned and future progress.

In this paper, the current trends in the development of AR applications were identified and a series of examples were presented in the application of AR technologies in the AEC industry by describing how BIM+AR can be used for reasoning the interdependences of tasks, project progress monitoring, digital to physical relations, material flow tracking and visualizing design during production. However, more research and case studies are required in the future to empirically examine the proposed integrated framework in obtaining the potential productivity and performance improvements in the construction industry.

REFERENCES

- Alwi, Hampson, & Mohamed. (2002). *Non value-adding activities: A comparative study of Indonesian and Australian construction projects*. Academic Press.
- Aouad, G., Lee, A., & Wu, S. (2006). *Constructing the future: nD modelling*. Routledge.
- Arayici, Y., Coates, P., Koskela, L., Kagioglou, M., Usher, C., & Oreilly, K. (2011). Technology adoption in the BIM implementation for lean architectural practice. *Automation in Construction*, 20(2), 189–195. doi:10.1016/j.autcon.2010.09.016
- Azuma, R. T. (1997). A survey of augmented reality. *Presence (Cambridge, Mass.)*, 6(4), 355–385. doi:10.1162/pres.1997.6.4.355
- Balzani, M., Santopuoli, N., Grieco, A., & Zaltron, N. (2004). *Laser scanner 3D survey in archaeological field: The Forum of Pompeii. International Conference on Remote Sensing Archaeology*, Beijing, China.
- Behzadan, A. H., & Kamat, V. R. (2012). Enabling discovery-based learning in construction using telepresent augmented reality. *Automation in Construction*.
- Behzadan, A. H., Timm, B. W., & Kamat, V. R. (2008). General-purpose modular hardware and software framework for mobile outdoor augmented reality applications in engineering. *Advanced Engineering Informatics*, 22(1), 90–105. doi:10.1016/j.aei.2007.08.005
- Bruner, J. S. (1966). *Towards a theory of instruction*. Harvard University Press.
- Caudell, T. P., & Mizell, D. W. (1992). *Augmented reality: An application of heads-up display technology to manual manufacturing processes. System Sciences, 1992. Proceedings of the Twenty-Fifth Hawaii International Conference on*. IEEE. doi:10.1109/HICSS.1992.183317
- Chi, H.-L., Kang, S.-C., & Wang, X. (2013). Research trends and opportunities of augmented reality applications in architecture, engineering, and construction. *Automation in Construction*, 33, 116–122. doi:10.1016/j.autcon.2012.12.017
- Construction, M.-H. (2008). *Building Information Modeling Trends SmartMarket Report*. New York: Academic Press.
- Donath, D., Beetz, J., Grether, K., Petzold, F., & Seichter, H. (2001). *Augmented reality techniques for design and revitalisation in existing built environments*. Buffalo, NY: Angenommener Beitrag zur ACADIA.
- Dong, S., Feng, C., & Kamat, V. R. (2012). Sensitivity analysis of augmented reality-assisted building damage reconnaissance using virtual prototyping. *Automation in Construction*.

- Dossick, C. S., & Neff, G. (2009). Organizational divisions in BIM-enabled commercial construction. *Journal of Construction Engineering and Management*, 136(4), 459–467. doi:10.1061/(ASCE)CO.1943-7862.0000109
- Eastman, C., Jeong, Y.-S., Sacks, R., & Kaner, I. (2009). Exchange model and exchange object concepts for implementation of national BIM standards. *Journal of Computing in Civil Engineering*, 24(1), 25–34. doi:10.1061/(ASCE)0887-3801(2010)24:1(25)
- Feiner, S. K. (2002). Augmented reality: A new way of seeing. *Scientific American*, 286(4), 48–55. doi:10.1038/scientificamerican0402-48 PMID:11905108
- Froese, T. (2005). Impact of emerging information technology on information management. *International Conference on Computing in Civil Engineering*, Cancun, Mexico. doi:10.1061/40794(179)162
- Froese, T. M. (2010). The impact of emerging information technology on project management for construction. *Automation in Construction*, 19(5), 531–538. doi:10.1016/j.autcon.2009.11.004
- Goldman, D. (2012). *Google unveils 'Project Glass' virtual-reality glasses*. Money.
- Hakkarainen, M., Woodward, C., & Rainio, K. (2009). Software architecture for mobile mixed reality and 4D BIM interaction. *Proc. 25th CIB W78 Conference*.
- Halpin & Woodhead. (1998). *Construction management*. Wiley.
- Holden, W. (2005). *Juniper research*. White Paper.
- Hou, L., & Wang, X. (2011). Experimental Framework for Evaluating Cognitive Workload of Using AR System for General Assembly Task. *Proceedings of the 28th International Symposium on Automation and Robotics in Construction*.
- Hwang, B.-G., Thomas, S. R., Haas, C. T., & Caldas, C. H. (2009). Measuring the impact of rework on construction cost performance. *Journal of Construction Engineering and Management*, 135(3), 187–198. doi:10.1061/(ASCE)0733-9364(2009)135:3(187)
- Hyers, K. (2006). *US mobile broadcast video market: Five predictions*. ABI Research.
- Irizarry, J., Gheisari, M., Williams, G., & Walker, B. N. (2012). InfoSPOT: A mobile Augmented Reality method for accessing building information through a situation awareness approach. *Automation in Construction*.
- Jardim-Goncalves, R., & Grilo, A. (2010). SOA4BIM: Putting the building and construction industry in the Single European Information Space. *Automation in Construction*, 19(4), 388–397. doi:10.1016/j.autcon.2009.11.009
- Jiao, Y., Zhang, S., Li, Y., Wang, Y., & Yang, B. (2012). Towards cloud Augmented Reality for construction application by BIM and SNS integration. *Automation in Construction*.
- Kamat, V. R., Martinez, J. C., Fischer, M., Golparvar-Fard, M., Peña-Mora, F., & Savarese, S. (2010). Research in visualization techniques for field construction. *Journal of Construction Engineering and Management*, 137(10), 853–862. doi:10.1061/(ASCE)CO.1943-7862.0000262
- Kandil, Hastak, & Dunston. (2014). Creating 3D web-based game environment using BIM models for virtual on-site visiting of building HVAC systems. *Bridges*, 10.
- Kang, J. H., Anderson, S. D., & Clayton, M. J. (2007). Empirical study on the merit of web-based 4D visualization in collaborative construction planning and scheduling. *Journal of Construction Engineering and Management*, 133(6), 447–461. doi:10.1061/(ASCE)0733-9364(2007)133:6(447)

- Love, P. E., Edwards, D. J., Han, S., & Goh, Y. M. (2011). Design error reduction: Toward the effective utilization of building information modeling. *Research in Engineering Design*, 22(3), 173–187. doi:10.1007/s00163-011-0105-x
- McKibben, D., & Furlonger, D. (2009). *Magic quadrant for operational risk management software for financial services*. Gartner Research. Available: <http://www.gartner.com/technology/home.jsp>
- Papagiannakis, G., Singh, G., & Magnenat-Thalmann, N. (2008). A survey of mobile and wireless technologies for augmented reality systems. *Computer Animation and Virtual Worlds*, 19(1), 3–22. doi:10.1002/cav.221
- Park, C.-S., & Kim, H.-J. (2012). A framework for construction safety management and visualization system. *Automation in Construction*.
- Park, C.-S., Lee, D.-Y., Kwon, O.-S., & Wang, X. (2012). A Framework for Proactive Construction Defect Management using BIM, Augmented Reality and Ontology-based Data Collection Template. *Automation in Construction*.
- Penttilä, H. (2006). *Describing the changes in architectural information technology to understand design complexity and free-form architectural expression*. ITcon.
- Shin, D. H., Park, J., Woo, S., & Jang, W.-S. (2012). Representations for imagining the scene of non-existing buildings in an existing environment. *Automation in Construction*.
- Stanco, F., Battiato, S., & Gallo, G. (2011). *Digital Imaging for Cultural Heritage Preservation: Analysis, Restoration, and Reconstruction of Ancient Artworks*. CRC Press.
- Taylor, J. E., & Bernstein, P. G. (2009). Paradigm trajectories of building information modeling practice in project networks. *Journal of Management Engineering*, 25(2), 69–76. doi:10.1061/(ASCE)0742-597X(2009)25:2(69)
- Van Krevelen, D., & Poelman, R. (2010). A survey of augmented reality technologies, applications and limitations. *International Journal of Virtual Reality*, 9(2), 1.
- Wang & Dunston (2007). *Design, strategies, and issues towards an augmented reality-based construction training platform*. ITcon.
- Wang, X., & Love, P. (2012). BIM+ AR: Onsite information sharing and communication via advanced visualization. *Computer Supported Cooperative Work in Design (CSCWD), 2012 IEEE 16th International Conference on*. IEEE.
- Wang, X., Love, P. E., Kim, M. J., Park, C.-S., Sing, C.-P., & Hou, L. (2013). A conceptual framework for integrating building information modeling with augmented reality. *Automation in Construction*, 34, 37–44. doi:10.1016/j.autcon.2012.10.012
- Wong, J. K., Li, H., Wang, H., Huang, T., Luo, E., & Li, V. (2012). Toward low-carbon construction processes: The visualisation of predicted emission via virtual prototyping technology. *Automation in Construction*.
- Yang, M.-D., Chao, C.-F., Huang, K.-S., Lu, L.-Y., & Chen, Y.-P. (2012). Image-based 3D scene reconstruction and exploration in augmented reality. *Automation in Construction*.
- Yang, T.-J., Tsai, Y.-M., & Chen, L.-G. (2011). Smart display: A mobile self-adaptive projector-camera system. *Multimedia and Expo (ICME), 2011 IEEE International Conference on*. IEEE.
- Yeh, K.-C., Tsai, M.-H., & Kang, S.-C. (2012). On-site building information retrieval by using projection-based augmented reality. *Journal of Computing in Civil Engineering*, 26(3), 342–355. doi:10.1061/(ASCE)CP.1943-5487.0000156
- Yokokohji, Y., Sugawara, Y., & Yoshikawa, T. (2000). Accurate image overlay on video see-through HMDs using vision and accelerometers. *Virtual Reality (Waltham Cross)*.

Zhang, C., & Hammad, A. (2012). Improving lifting motion planning and re-planning of cranes with consideration for safety and efficiency. *Advanced Engineering Informatics*, 26(2), 396–410. doi:10.1016/j.aei.2012.01.003

KEY TERMS AND DEFINITIONS

Augmented Reality: A live direct or indirect view of a physical and real-world environment

where the elements are augmented by computer-generated input such as sound, video, graphics, etc.

AEC: Stands for architectural, engineering and construction field.

Building Information Modeling (BIM): A process where it involves the generation and management of digital representations of physical and functional characteristics of items or places.

Literature Review: A text of a scholarly paper including the current knowledge and findings alongside the theoretical and methodological contributions to a particular topic.

Category C

Clinical Science and Technologies

A Gamification Update to the Taxonomy of Technology and Mental Health

C

Madeline R. Marks

University of Central Florida, USA

Amanda C. Tan

University of Central Florida, USA

Clint Bowers

University of Central Florida, USA

INTRODUCTION

Mental health providers cannot ignore the importance of utilizing technology in this era of the “Internet of Things.” This chapter reaffirms the need for mental health providers and software developers to work in concert with each other when developing technology for mental health. We also articulate the importance of the patient and the patient’s role in connecting technology into the equation.

BACKGROUND

The Problem

Marks and Bowers (2014) developed a taxonomy that allowed mental health providers to understand how technology can benefit patients as well as direct the provider to the proper platform depending on the aspect of the patient’s mental health requirements. A taxonomy was developed that included classifications of disorders with specific disorders from the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (*DSM-5*), delivery platform i.e., mobile device, personal computer, and video game console, and the four classifications of mental health services (e.g., Training, Therapy, Assessment, and Prevention). This taxonomy guided developers to focus

on building mental health technology (mHealth) that revolve around a particular sector. Further, the taxonomy capitalized on the similarities between providers and developers in achieving the end-goal of repeated and continued use of their respective product (therapeutic services and software). However, the previous taxonomy did not account for the different motivational strategies employed to reach the end-goal; in the design and development of mHealth providers and developers focus on two different factors. Providers focus on delivering evidenced-based treatments driven by underlying mechanisms of action (MOA); whereas, developers focus on delivering usable and entertaining technology. Compounding this disparate approach between providers and developers is the failure to take into account the patient perspective. The present chapter presents a revised taxonomy that represents the integration of MOA, usability and entertainment, engagement to create gamified technology for the purpose of increasing, repeated and continued engagement with the technology, by the end-user. The taxonomy has been revised to reflect the change of including the end-user and integration of gamification. Additional changes have been made to the taxonomy to better represent this new focus. These include changes to the component of Training, which is now called Psychoeducation; Assessment, which has been removed from the taxonomy because the focus of technology for assessment is not for the

DOI: 10.4018/978-1-5225-2255-3.ch086

end-user. All changes have been made in an effort to better synthesize the newly added information afforded by considering the end-user, in pursuit of creating repeated and continuous engagement with mHealth.

To better understand the disparate approaches between providers and developers and how the end-user will be incorporated, an explanation of each motivational target of the triad is warranted.

Provider

The mental health provider's goal is clear: it is to help a patient improve upon a targeted domain that is interfering with the patient's quality of life. Mental health providers agree that motivation is a key component to attendance and participation in treatment. Without motivation, therapeutic processes cannot take place nor be effective. At each step of the therapeutic process the mental health provider guides the patient through the therapeutic process, and works with the patient to determine the pace of progression to end goals. It is through clinical experience that is informed by research, that the provider makes these crucial decisions.

The therapeutic process involves the general components of 1) establishing a therapeutic alliance, 2) establishing what is the patient's commitment to change, conducting a behavioral analysis, coordinating treatment objectives, 3) executing treatment and maintaining motivation, monitoring progress, and 4) planning for treatment termination by generalizing skills (Rosenbaum & Horowitz, 1983). These steps create a tailored versus standardized treatment structure, and encourages the patient to be an active participant in the progression toward change (Kanfer & Grimm, 1980). It creates a therapeutic alliance between the mental health provider and the patient. The therapeutic alliance refers to the patient's experience as to the mental health provider's empathy; it refers to the patient's belief as to the provider's credibility as an expert; it refers to the patient's perceived rapport with the provider; it refers to the patient's perception that he or she is receiving the help needed from

the provider; it is the patient's perspective that the provider is engaged; it refers to the patient's perception that the treatment is a process working toward specified goals (Elvins & Green, 2008). This therapeutic alliance, while not a specific strategy for motivating patients, encompasses all domains of influencing motivation to adhere to treatment (Barrett, Chua, Crits-Christoph, Gibbons, & Thompson, 2008; Bordin, 1994). The best form of treatment requires practitioners to focus on evidence that supports effective intervention using evidence-based treatments. Evidence-based treatment involves the integration of best available research, clinical expertise, and consideration of individual patient characteristics (Anderson, 2006). In essence, each treatment has an underlying MOA. Considering that mental health providers should be delivering evidence-based treatment in a traditional face-to-face format, providers should continue to provide the best possible treatment regardless of the platform. The focus of the mental health provider is on mHealth's ability to support treatment rationale and adherence to clinical practice.

Developer

Adams (2014) describes the developer's job as brainstorming designs, creating meaning, distinguishing content, and testing the technology. Adams organizes the developer's software designing job into a three-part system:

1. **Concept Stage:** Deciding the kind of technology to develop.
2. **Elaboration Stage:** Creating and developing the software, then conducting usability testing.
3. **Tuning Stage:** Preparing the software for release, including fixing errors in code.

The second and third steps are particularly important because developers desire for users to spend time playing the software and do so on a repeated basis, which, as an aside, is consistent

with the desires of mental health providers (Von Ahn & Dabbish, 2008). In fact, there is evidence noting that evaluating usability is crucial in motivating people to engage in the use of mHealth technology. Poor usability and user experience has demonstrated to de-motivate people from using the technology (Spillers & Asimakopoulos, 2014). When issues of usability are resolved, elements that facilitate great technology can be implemented.

Gamification, while not a specific strategy for motivating players, summarizes the integration of elements that compose great games. Successful mHealth games have goals, rewards that people are excited to obtain and, thus, play in order to achieve. Developers elicit excitement by integrating elements that capture attention, relevance, confidence, and satisfaction (Keller, 1983; Paras, 2005). The excitement is palpable and fosters motivation to continue playing. Players work hard to advance in game play by completing levels, collect badges, earn coins, and receive points, in other words, positive reinforcement. The aforementioned rewards are direct reflections of the link between the player's effort and achievement (Malone, 1982). The visual representation of effort and outcome is highly rooted in the competitive nature of a player, which when combined synchronizes and increases the desire to continue playing (Vorderer, Hartmann, & Klimmt, 2003).

This visual representation is a form of feedback, which leans on our innate desire for immediate feedback regarding performance. This desire for immediate feedback is present regardless of context, which in this case, happens to be within a game environment (Ciampa, 2014). While typically viewed as positive, immediate feedback received within the game environment can serve as a potential source of frustration - frustration stemming from game elements being too demanding (Tauer & Harackiewicz, 1999).

Establishing an equilibrium of a game's goals between simplicity, which can be boring, and difficulty, which can be aggravating is essential when considering motivational target of repeated, continuous, use (Malone & Lepper, 1987). This

balance is imperative to keep the player interested and motivated in advancing towards the end goal (Richards & Caldwell, 2015). The balance starts with developing a game progression embedded with appropriate challenges and sufficient rewards (Richards & Caldwell, 2015). The in-game rewards of advancing thru the game and earning the aforementioned reinforcement is intrinsically motivating as it captures the notion of progress (Klosowski, 2014).

Another aspect that motivates individuals to advance towards the end goal is the incorporation of narratives, or stories (McDaniel, Fiore, & Nicholson, 2010). Strong storytelling stimulates the users draw to the game, in turn, providing motivation to play. The strength of the story is enhanced when relatable avatars are leveraged. Relatable avatars provide players the feeling of personal connection to the game (Doyle, 2004). The opportunity to customize avatars deepens the player's personal connection to the game leading to enjoyment, and therefore, increased time spent within the game environment (Doyle, 2004; Schmierbach, Limperos, & Woolley, 2012).

The emphasis of the developers' motivational targets, however, is on whether or not the technology is usable and entertaining for the target audience, not that it will actually be used. Just because a technology is *usable* and *entertaining* by definition, does not ensure that the technology serves its' purpose in regards to the providers' MOA. This implies that there is a third variable in the equation, and that is the end-user, patient.

End-User, Patient

End-users validate technology. The End-user must be taken into account at the design phase in order to ensure usability that fosters prolonged and repeated use and engagement. Specifically, the inclusion of an entertaining, usable, mechanism of action driven mHealth will not predict actual use of said technology. Only 35% of adults who own phones have downloaded at least one app, but only two-thirds of those actually use these

apps (Purcell, 2011). A mental health technology simply for the sake of creation is not sufficient; no matter the appeal of health technology, it will not be used by the patient unless (i) it helps address with her or his specific problem and (ii) avoids technical failures, i.e., crashes, freezes, error messages (Perez, 2013; Scher, 2013). Perez (2013) demonstrated that only 16% of people would try to use an app more than twice if it fails.

In addition, mHealth design and functionality needs to provide flexibility by taking into account a patient's motivation to use the technology. Motivation to engage with the technology is rooted in the need to reduce symptoms of psychopathology, willingness to seek help, and improve quality of life (Wilson, 2007).

Mental health technology cannot consider each individual when designing technology. That is, symptoms that the technology is designed to mitigate may not line up perfectly with what the patient experiences. Yee (2006) hypothesized and defined three end-user, player, and motivational components as follows.

1. **Achievement:** A Player yearns to make progress within the game, which includes gaining power, wealth, and knowledge of the rules to "better" her or his character, avatar, as well as the inclination to compete against others in the game.
2. **Social:** A Player desires to form relationships and help others in the game, and enjoy being part of a group.
3. **Immersion:** A Player wishes to know what others players do not know and enjoys being able to customize her or his own story and character in the game.

The challenge, therefore, is integrating the benefits of face-to-face therapy into a game construct for mental health in order for the application of the Triad to the taxonomy to be effective.

APPLICATION TO THE TAXONOMY

Good user engagement offers insights into development of more effective implementation of MOA and incorporation of usability and entertainment. As illustrated, each component of the Triad contributes unique motivational targets. Similarly, each component of the taxonomy is different. In psychoeducation, the goal is to gain knowledge and apply that knowledge to the world outside. In therapy, the goal is to deliver a type of treatment that may address and help the patient cope with impairments to activities of daily living. Finally, in prevention, the main goal is preventing individuals (regardless of their risk level) from developing a disorder (Marks & Bowers, 2014). The challenge becomes creating a balance amongst each component to satisfy the MOA, entertainment, and engagement needs of each entity. The following recommendations about gamification guided by the motivational targets of provider, developer, and end-user.

Psychoeducation

Mini-Games

The goal of psychoeducation is to gain and apply knowledge to novel environments. Mini-games can serve as the mechanism for testing knowledge and assessing knowledge transfer. Mini-games are small games, play lasts under an hour, embody one learning goal, easy to learn, and difficult to master. Mini-games can be employed within several contexts and are best suited towards acquisition of conceptual and procedural knowledge (Smith & Sanchez, 2010).

Mini-games provide the opportunity for fun and engagement while simultaneously accomplishing the goal of providing the user with psychoeducation. New information is processed as stimuli, which is encoded during gameplay interaction.

In order to complete the game, players are required to memorize content and be able to recall for completion of the game (Smith & Sanchez, 2010). The advantages of mini-games are two-fold: one, in an environment where the provider is not able to directly observe the user's understanding, mini-games allow for the quantitative assessment of the users understanding; two, understanding, engagement is imperative in order for the user to gain the conveyed knowledge and retain the newly learned cognitive skills (Andrews, Joyce, & Bowers, 2010).

Therapy

Narratives, Avatars, and Alternate Reality Games

Narratives are powerful tools that embed learning content within the game to create a coherent story that motivates the user to continue their engagement in the game (McDaniel, Fiore, & Nicholson, 2010). In other words, narratives can be used for improving immersion by drawing the player into the world of the game. One type of narrative congruent with the face-to-face therapy format is the interactive story archetype. An interactive story is an experience in which the simulated world responds to the player's movements, decisions, and actions within the game-space (Bates, 1992; McDaniel, Fiore, & Nicholson, 2010; Thue, Bulitko, Spetch, & Wasylishen, 2007). The parallels between face-to-face therapy and the interactive story are seen by the emphasis that the player rather than that of the game structure dictates movements, decisions, and actions. The game structure is synonymous with the therapist, who provides the content and environment and the player is synonymous with the patient, who dictates the simulated world's response. This structure can assist with the creation of a therapeutic alliance between the user and the mHealth technology.

Further strengthening the therapeutic alliance, avatars should be considered during development of the interactive narrative. By placing empathic

characters that are, perceived to instill hope; perceived to be competent; and perceived to be committed, creates motivation for continued play. One example of translating the perception of commitment to the novel environment is the utilization of "push-notifications." Outside of the virtual environment, Avatars can be programmed to send "push-notifications" direct to the users cell-phone. Translating the narrative and the avatars presence from the virtual environment, to interacting with the user outside of the platform is reminiscent of alternate reality games. Alternate reality games (ARG) are a narrative that extends to the players real world (Johnston, Sheldon, & Massey, 2010).

The integration of an interactive story, empathic avatars, and ARG is an example of combining the motivational targets of the triad and addressing the problem of translation of gameplay to the generalizable, novel, everyday environment.

Prevention

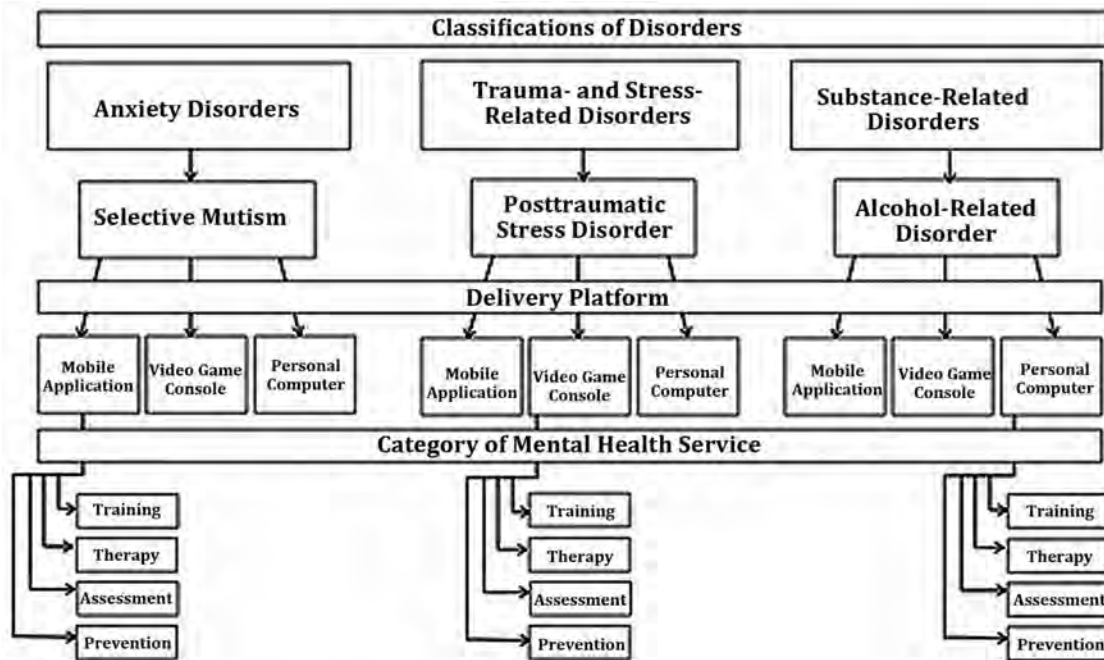
Progression and Positive Reinforcement

Games are typically designed so that complexity and degree of difficulty increases as game-play progresses. Similar to games, users of mHealth technology will have a wide range of abilities and knowledge regarding the target of prevention. In a traditional face-to-face setting, a mental health provider will assess the current skill level, needs, and targets for improvement of the population targeted for prevention. A congruent assessment must take place within the game environment so that the difficulty of the game can be presented accordingly.

Increasing challenge can be designed by way of levels. Levels and degree of difficulty can be implemented by completing an earlier less challenging level, by increasing difficulty as time spent in game increases, or by an adaptive system. The adaptive system, as the name suggests, adapts by

Figure 1. Expanding the taxonomy of technology and mental health

Source: Marks & Bowers, 2015



adjusting degree of difficulty to the user based on their performance in game. The advantages of the adaptive system are evident considering that users become bored with games that are too easy or frustrated with games that are too challenging, which leads to discontinued play (Klosowski, 2014; Tauer & Harackiewicz, 1999). A user who has the skills and knowledge will be able to progress more quickly while another user will be provided the content in a more judicious manner.

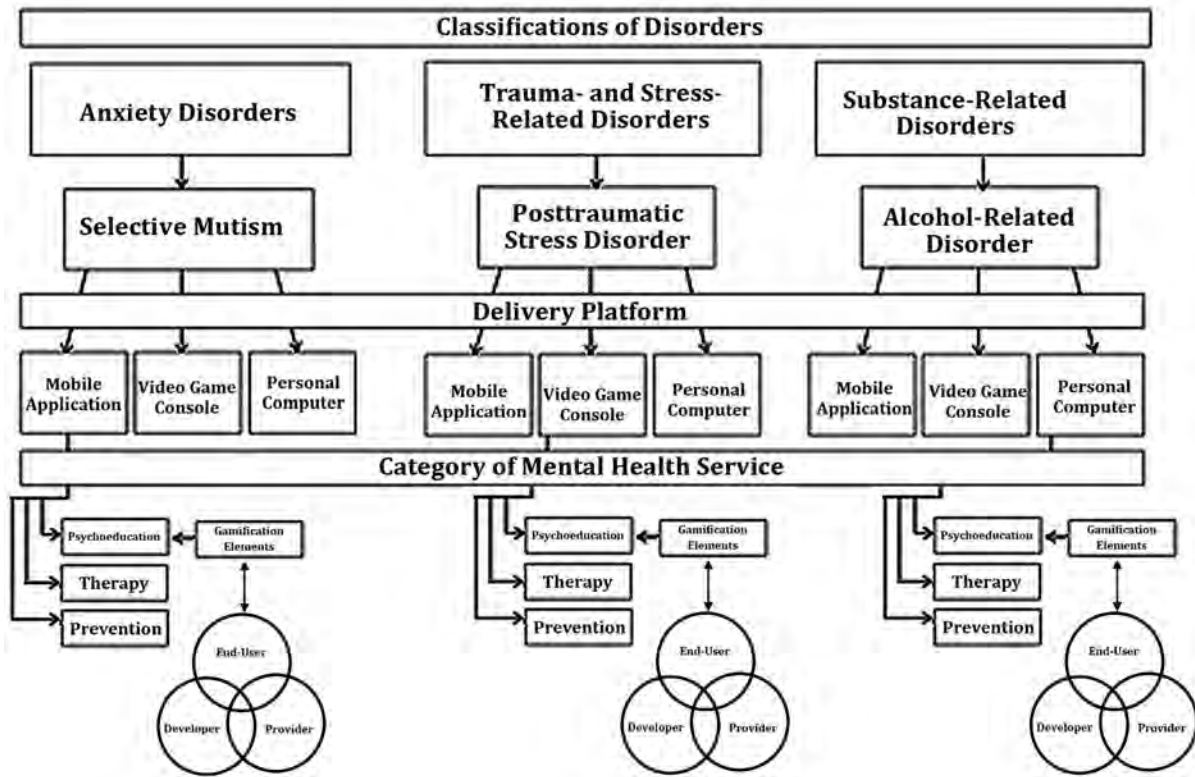
This adaptive approach also lends well to the capturing and analyzing performance in real time. Two types of data can be generated using the adaptive approach. The first is what we would call performance “meta-data.” This data is derived from the capabilities embedded in smartphones and other software to track time spent engaging with content, length of time spent within a module/level, and the location of use. The performance meta-data can provide a mental health provider, as well as the user, with worthwhile information regarding level of engagement for the practice of skills. The performance meta-data can be synced

with in-game performance reinforcement (i.e., amount of time spent practicing skills and engaging in content results in collection of badges, coins, points).

FUTURE RESEARCH DIRECTIONS

The aforementioned suggestions are not mutually exclusive, rather the goal was to demonstrate how taxonomy for mHealth can be leveraged to synthesize information and best practices. Future research should look to understand the specific motivational targets that underlie each component of the taxonomy. Specific attention should be given to mental health technology that focuses on prevention. Primary prevention seeks to stop or stall the development of disorder. Primary prevention takes place prior to the expression of symptoms by intervening with all those who share characteristics that place them at risk for development of disorder (Cowen, 2000).

Figure 2. A Gamification, End-User incorporated taxonomy of technology and mental health

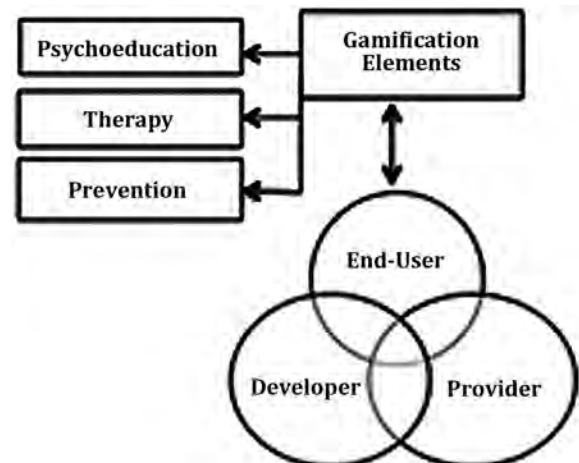


The very definition of primary prevention is that intervention takes place *prior* to the manifestation of symptoms. Whereas in training, therapy, and assessment, the patient has identified that there is interference in their quality of life that warrants attention. In prevention, there is a need to understand how to motivate people to use mHealth technology to promote resilience, and intervene prior to a problem developing. Targeting the motivational aspects unique to prevention would result in the increase of effectiveness of the primary prevention. The challenge is motivating a user to take the time to engage in technology that is perceived as a worthwhile investment of time.

Prevention technology is going to need to have more salient entertainment functions, but will also need to have strong therapeutic MOA. It is important, however, to understand that you cannot simply interpolate a game-like interface onto the MOA that is desired; rather, you must merge the MOA into the game (Von Ahn & Dabbish, 2008).

The original taxonomy, updated taxonomy, as well as the updated component of the taxonomy, has been included (see Figure 1, Figure 2, and Figure 3, respectively)

Figure 3. Updated component of the taxonomy of technology and mental health



CONCLUSION

The integration of provider, developer, and patient into the creation and development of mental health related technology will be crucial for ensuring success. Thus, in order ensure success, including increasing accessibility and utilization of mental health technology along with balancing mechanism of action, usability and entertainment, mental health providers and software developers must not forget the end-user, patient, in the creation and development of mHealth. Without a doubt, technology will be as relevant tomorrow as it is today, even more so, as technological advancements continue at an ever increasing pace and patient's become more reliant and dependent upon, and integrate technology into everyday life. As researchers and practitioners, the goal should be to create technology that will encourage repeated and continuous use of said technology, and not just technology acceptance in order to move the field forward toward the provision of low-cost, effective mental health services. Furthermore, each of the foregoing must ensure the continued evolution of such apps as new delivery platforms come to fruition. The Internet of Things is here to stay.

REFERENCES

- Adams, E. (2014). *Fundamentals of game design*. Pearson Education.
- American Psychological Association. (2015). *Understanding psychotherapy and how it works*. Retrieved from <http://www.apa.org/helpcenter/understanding-psychotherapy.aspx>
- Anderson, N. B. (2006). Evidence-based practice in psychology. *The American Psychologist*, 61(4), 271–285. doi:10.1037/0003-066X.61.4.271 PMID:16719673
- Barrett, M. S., Chua, W., Crits-Christoph, P., Gibbons, M. B., & Thompson, D. (2008). Early withdrawal from mental health treatment: Implications for psychotherapy practice. *Psychotherapy (Chicago, Ill.)*, 45(2), 247–267. doi:10.1037/0033-3204.45.2.247 PMID:19838318
- Bates, J. (1992). *The nature of characters in interactive worlds and the Oz project*. Pittsburgh, PA: School of Computer Science, Carnegie Mellon University.
- Bordin, E. S. (1994). Theory and research on the therapeutic working alliance: New directions. *The working alliance: Theory, research, and practice*, 13-37.
- Ciampa, K. (2014). Learning in a mobile age: An investigation of student motivation. *Journal of Computer Assisted Learning*, 30(1), 82–96. doi:10.1111/jcal.12036
- Cowen, E. L. (2000). Now that we all know that primary prevention in mental health is great, what is it? *Journal of Community Psychology*, 28(1), 5–16. doi:10.1002/(SICI)1520-6629(200001)28:1<5::AID-JCOP3>3.0.CO;2-H
- Doyle, P. (2010). *A study of the immersion enhancement techniques utilised by video games*. Academic Press.
- Elvins, R., & Green, J. (2008). The conceptualization and measurement of therapeutic alliance: An empirical review. *Clinical Psychology Review*, 28(7), 1167–1187. doi:10.1016/j.cpr.2008.04.002 PMID:18538907
- Johnston, J. D., Sheldon, L., & Massey, A. P. (2010). *Influencing physical activity and healthy behaviors in college students: lessons from an alternate reality game. Serious game design and development: technologies for training and learning*. Hershey, PA: IGI Global.

- Kanfer, F. H., & Grimm, L. G. (1980). Managing Clinical Change A Process Model of Therapy. *Behavior Modification*, 4(4), 419–444. doi:10.1177/014544558044001
- Keller, J. M. (1983). Motivational design of instruction. In C. M. Reigeluth (Ed.), *Instructional design theories and models: An overview of their current status*. Hillsdale, NJ: Erlbaum.
- Klosowski, T. (2014). *The Psychology of Gamification: Can Apps Keep You Motivated?* Retrieved from <http://lifehacker.com/the-psychology-of-gamification-can-apps-keep-you-motiv-1521754385>
- Kreutzer, C., Marks, M., & Bowers, C. (2015). A Pedagogical Approach to Usability in Serious Games. *HCI International 2015 - Posters' Extended Abstracts*, 43. doi:10.1007/978-3-319-21380-4_8
- Malone, T. M. (1982). Heuristics for designing enjoyable user interfaces: Lessons from computer games. In *Proceedings of the Conference on Human Factors in Computing Systems*. ACM Press. doi:10.1145/800049.801756
- Malone, T. W., & Lepper, M. R. (1987). Making learning fun: A taxonomy of intrinsic motivations for learning. In R. E. Snow & M. J. Farr (Eds.), *Aptitude, learning, and instruction: III. Cognitive and affective process analyses* (pp. 223–253). Hillsdale, NJ: Erlbaum.
- Marks, M., & Bowers, C. (2015). Technology and Mental Health. In M. Khosrow-Pour (Ed.), *Encyclopedia of Information Science and Technology* (3rd ed.; pp. 6455–6465). Hershey, PA: Information Science Reference; doi:10.4018/978-1-4666-5888-2.ch634
- McDaniel, R., Fiore, S. M., & Nicholson, D. (2010). Serious storytelling: Narrative considerations for serious games researchers and developers. *Serious game design and development: Technologies for training and learning*, 13–30.
- Newman, C. F. (1994). Understanding client resistance: Methods for enhancing motivation to change. *Cognitive and Behavioral Practice*, 1(1), 47–69. doi:10.1016/S1077-7229(05)80086-0
- Paras, B. (2005). Game, Motivation, and Effective Learning: An Integrated Model for Educational Game Design. *DiGRA 2005: Changing Views: Worlds in Play, 2005 International Conference*.
- Perez, S. (2013). *Users Have Low Tolerance For Buggy Apps – Only 16% Will Try A Failing App More Than Twice*. Retrieved from http://techcrunch.com/2013/03/12/users-have-low-tolerance-for-buggy-apps-only-16-will-try-a-failing-app-more-than-twice/?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed%3A+Techcrunch+%28TechCrunch%29
- Purcell, K. (2011). *Half of adult cell phone owners have apps on their phones*. Pew Internet & American Life Project.
- Reeves, B., & Read, J. L. (2013). *Total engagement: How games and virtual worlds are changing the way people work and businesses compete*. Harvard Business Press.
- Richards, D., & Caldwell, P. H. (2015). Gamification to Improve Adherence to Clinical Treatment Advice: Improving Adherence to Clinical Treatment. *Handbook of Research on Holistic Perspectives in Gamification for Clinical Practice*, 47.
- Rosenbaum, R. L., & Horowitz, M. J. (1983). Motivation for psychotherapy: A factorial and conceptual analysis. *Psychotherapy (Chicago, Ill.)*, 20(3), 346–354. doi:10.1037/h0090205

Scher, D. L. (2013). *5 reasons why mobile health apps fail*. Retrieved from <http://www.kevinmd.com/blog/2013/02/5-reasons-mobile-health-apps-fail.html>

Schmierbach, M., Limperos, A. M., & Woolley, J. K. (2012). Feeling the need for (personalized) speed: How natural controls and customization contribute to enjoyment of a racing game through enhanced immersion. *Cyberpsychology, Behavior, and Social Networking*, 15(7), 364–369. doi:10.1089/cyber.2012.0025 PMID:22687145

Smith, P. A., & Sanchez, A. (2010). Mini-games with major impacts. *Serious game design and development: Technologies for training and learning*, 1-12.

Spillers, F., & Asimakopoulos, S. (2014). Does Social User Experience Improve Motivation for Runners? In A. Marcus (Ed.), *Design, User Experience, and Usability: User Experience Design Practice* (pp. 358–369). Springer International Publishing. doi:10.1007/978-3-319-07638-6_35

Tauer, J. M., & Harackiewicz, J. M. (1999). Winning Isn't Everything: Competition, Achievement Orientation, and Intrinsic Motivation. *Journal of Experimental Social Psychology*, 35(3), 209–238. doi:10.1006/jesp.1999.1383

Thue, D., Bulitko, V., Spetch, M., & Wasylishen, E. (2007, June). Interactive Storytelling: A Player Modelling Approach. *AIIDE*, 43-48.

Von Ahn, L., & Dabbish, L. (2008). Designing Games With A Purpose. *Communications of the ACM*, 51(8), 58–67. doi:10.1145/1378704.1378719

Vorderer, P., Hartmann, T., & Klimmt, C. (2003, May). Explaining the enjoyment of playing video games: the role of competition. In *Proceedings of the second international conference on Entertainment computing* (pp. 1-9). Carnegie Mellon University.

Wilson, C. J. (2007). When and how do young people seek professional help for mental health problems? *The Medical Journal of Australia*, 187.

Yee, N. (2006). Motivations for play in on-line games. *Cyberpsychology & Behavior*, 9(6), 772–775. doi:10.1089/cpb.2006.9.772 PMID:17201605

ADDITIONAL READING

Alber, J. M., Watson, A. M., Barnett, T. E., Mercado, R., & Bernhardt, J. M. (2015). Development of a Coding Instrument to Assess the Quality and Content of Anti-Tobacco Video Games. *Cyberpsychology, Behavior, and Social Networking*, 18(7), 417–425. doi:10.1089/cyber.2015.0051 PMID:26167842

All, A., Castellar, E. P. N., & Van Looy, J. (2016). Assessing the effectiveness of digital game-based learning: Best practices. *Computers & Education*, 92, 90–103. doi:10.1016/j.compedu.2015.10.007

Banfield, J., & Wilkerson, B. (2014). Increasing Student Intrinsic Motivation And Self-Efficacy Through Gamification Pedagogy. *Contemporary Issues in Education Research*, 7(4), 291–298. doi:10.19030/cier.v7i4.8843

Clarke, A. M., Kuosmanen, T., & Barry, M. M. (2015). A systematic review of online youth mental health promotion and prevention interventions. *Journal of Youth and Adolescence*, 44(1), 90–113. doi:10.1007/s10964-014-0165-0 PMID:25115460

DeSmet, A., Van Ryckeghem, D., Compennolle, S., Baranowski, T., Thompson, D., Crombez, G., & De Bourdeaudhuij, I. et al. (2014). A meta-analysis of serious digital games for healthy lifestyle promotion. *Preventive Medicine*, 69, 95–107. doi:10.1016/j.ypmed.2014.08.026 PMID:25172024

- Eseryel, D., Law, V., Ifenthaler, D., Ge, X., & Miller, R. (2014). An investigation of the inter-relationships between motivation, engagement, and complex problem solving in game-based learning. *Journal of Educational Technology & Society*, 17(1), 42–53.
- Farrer, L., Gulliver, A., Bennett, K., & Griffiths, K. M. (2015). *Exploring the acceptability of on-line mental health interventions among university teaching staff: Implications for intervention dissemination and uptake*. Internet Interventions.
- Farrer, L., Gulliver, A., Chan, J. K., Bennett, K., & Griffiths, K. M. (2015). A virtual mental health clinic for university students: A qualitative study of end-user service needs and priorities. *JMIR Mental Health*, 2(1), e2. doi:10.2196/mental.3890 PMID:26543908
- Filsecker, M., & Hickey, D. T. (2014). A multi-level analysis of the effects of external rewards on elementary students motivation, engagement and learning in an educational game. *Computers & Education*, 75, 136–148. doi:10.1016/j.compedu.2014.02.008
- Frost, J., & Eden, A. (2014). The Effect of Social Sharing Games and Game Performance on Motivation to Play Brain Games. In *Games for Health 2014* (pp. 48–55). Springer Fachmedien Wiesbaden. doi:10.1007/978-3-658-07141-7_7
- Hamari, J., Shernoff, D. J., Rowe, E., Coller, B., Asbell-Clarke, J., & Edwards, T. (2016). Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning. *Computers in Human Behavior*, 54, 170–179. doi:10.1016/j.chb.2015.07.045
- Kreutzer, C. P., & Bowers, C. A. (2015). Attitudes Toward a Game-Based Approach to Mental Health. *Cyberpsychology, Behavior, and Social Networking*, 18(1), 20–24. doi:10.1089/cyber.2014.0382 PMID:25584727
- McKernan, B., Martey, R. M., Stromer-Galley, J., Kenski, K., Clegg, B. A., Folkestad, J. E., & Strzalkowski, T. et al. (2015). We dont need no stinkinbadges: The impact of reward features and feeling rewarded in educational games. *Computers in Human Behavior*, 45, 299–306. doi:10.1016/j.chb.2014.12.028
- Sorgendal, I., & Boks, C. (2014). Challenging interfaces are more fun! Operant conditioning for the interaction designer. *International Journal of Learning Technology*, 9(2), 94–110. doi:10.1504/IJLT.2014.064488
- Torous, J., Chan, S. R., Tan, S. Y. M., Behrens, J., Mathew, I., Conrad, E. J., & Keshavan, M. et al. (2014). Patient smartphone ownership and interest in mobile apps to monitor symptoms of mental health conditions: A survey in four geographically distinct psychiatric clinics. *JMIR Mental Health*, 1(1), e5. doi:10.2196/mental.4004 PMID:26543905
- Whyte, E. M., Smyth, J. M., & Scherf, K. S. (2014). Designing Serious Game Interventions for Individuals with Autism. *Journal of Autism and Developmental Disorders*, 1–12. PMID:25488121

Heart Sound Analysis for Blood Pressure Estimation

Rui Guedes

Faculdade de Medicina da Universidade do Porto, Portugal

Henrique Cyrne Carvalho

Serviço de Cardiologia, Hospital de Santo António, Centro Hospitalar do Porto, Portugal

Ana Castro

Universidade do Porto, Portugal

INTRODUCTION

Heart disease is one of the main causes of death globally. Some heart diseases may be diagnosed by measuring the pulmonary arterial pressure (PAP), which may be increased due to heart and pulmonary diseases. An increase in the pressure of the pulmonary artery could be an indication of pulmonary arterial hypertension, and may be defined as a mean PAP superior to 25 mmHg (Galie et al., 2009) and outside the 15-30 mmHg range for the systolic PAP (Edwards Lifesciences LLC, 2009).

For diagnostic purposes, most patients undergo an electrocardiogram (ECG), a chest X-ray, an echocardiogram, a computerized tomography scan of the chest, and a cardiac catheterization. The last one, the invasive arterial blood pressure (BP) monitoring, is considered the gold standard for BP assessment, providing accurate information, however it is also the most expensive (Chung, Chen, Alexander, & Cannesson, 2013). During a cardiac catheterization, a small flexible tube is inserted into a blood vessel in the arm or the groin, with a contrast dye that makes it possible to see under X-ray and obtain images of the heart structure and vessels. The catheter with a pressure sensor allows measurement of pressures in the vessels and heart chambers, which provide clues to a pulmonary hypertension diagnosis. Despite the catheterization procedure being performed

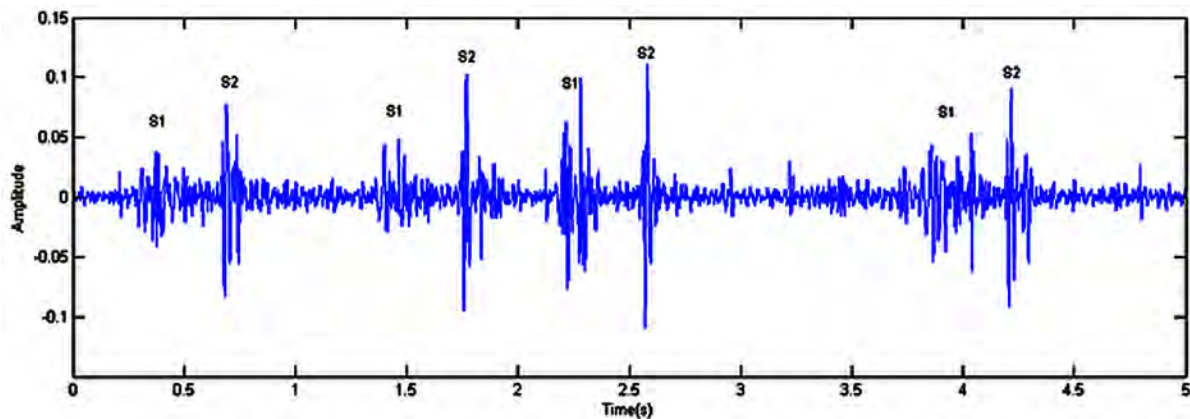
with the patient awake, and bring little or no pain, the patient may still feel pain in the blood vessel where the catheter is inserted. There is a mandatory rest time after the procedure. Also financial costs and risk discourage patients to undergo right heart catheterization procedure and from seeking potentially life-saving diagnosis and treatment of elevated PAP. Elevated PAP can be an indicator of pulmonary hypertension, a condition that may worsen over time, although treatments can relieve symptoms and give patients a better quality of life. The diagnostic we want to focus in pulmonary hypertension is the pulmonary arterial hypertension (PAH) that is a type of high BP that occurs in the right side of the heart and in the pulmonary arteries. PAH occurs when the pulmonary arteries thicken or grow rigid, this makes blood flow more difficult. The right side of the heart has to work harder to push blood through the arteries, and the arteries are not able to carry adequate blood to oxygenate in the lungs. It is difficult to continuously assess pulmonary arterial hypertension evolution, which motivates the search for accurate noninvasive PAP measurement devices.

BACKGROUND

A noninvasive estimation of the PAP is already in clinical use, the Doppler echocardiography; it uses ultrasound technology to measure PAP by

DOI: 10.4018/978-1-5225-2255-3.ch087

Figure 1. Phonocardiogram (PCG) representation with first (S1) and second (S2) heart sounds annotated



analyzing the speed of blood flowing through the heart. Doppler is widely used by cardiologists, it requires specific equipment, a specialized and trained doctor, and therefore it is difficult to be used for continuous monitoring, or in places without specialized healthcare professionals. Also some patients are not candidates for Doppler due to physiological constraints or poor signal-to-noise ratio: PAP cannot be estimated in 34 to 76% of patients with chronic obstructive pulmonary disease, 10 to 20% of patients with elevated PAP, and the most important in approximately half of patients with normal PAP (Xu, Durand, & Pibarot, 2003); also the Doppler estimations present an average error of 30% compared to right heart catheterization (Smith & Ventura, 2013).

Another method proposed for PAP noninvasive estimation is the heart sound analysis. Phonocardiography is a technique that generates a record of sounds produced by the contracting heart, resulting from valves and associated vessels vibration, the phonocardiogram (PCG), as we can see in Figure 1.

The sound is collected with a stethoscope over the main auscultations points on the chest; for the detection of pulmonary hypertension, usually the second heart sound exhibits an hyperphoresis of the pulmonary component, more audible over the pulmonary auscultation site (Young, 2001). Recent rebirth of the stethoscope, with electronic record of the PCG, has led to the search of automatic signal processing techniques to detect this clinical sign,

and few studies have addressed this issue in the literature (J Xu et al., 2002; Lim, Shin, Park, & Bae, 2013; Nigam & Priemer, 2004). Phonocardiography has been used to estimate PAP from noninvasively recorded heart sounds using models for how these sounds correlate with PAP. There are however several research questions that need to be answered in the PCG study, namely the robust segmentation of heart sounds, and modeling of the heart components that allow the extraction of relevant clinical information. It is a necessity to better understand the PCG and its relation with the PAP. To allow a better comprehension of the problem, the following sections provides a brief introduction to the anatomy, physiology, and the heart sound genesis.

HEART SOUND ANALISYS FOR HYPERTENSION DETECTION

Several studies have addressed the way information contained in the heart sound may be related to the arterial BPs at the heart output. This method has been studied with more emphasis on the detection of patients with PAH, considering the frequency, the amplitude, and split of the constituent components of the S2 - aortic (A2) and pulmonary (P2) - corresponding to the closure of the homonym heart valves (Nigam & Priemer, 2004; Jingping Xu, Durand, & Pibarot, 2001).

In order to estimate the BP it is important to separate the S2 components, meaning the aortic and pulmonary components, which has revealed to be a challenging task. In this section we intend to describe the main challenges for the second heart sound analysis, with emphasis on the PAH detection, since this is the most developed to the moment.

The first task in the analysis of the second heart sound is to detect the boundaries between the reference sounds. The concern here is that when the heart sound is registered several other sounds are mixed together, and the objective is to recover the heart sound components from that combined signal.

One of the most cited papers on this subject is the Liang et al., their algorithm segments the heart sound into four parts: the first heart sound, the systole, the second heart sound, and the diastole. The authors consider segmentation of the PCG the first and the most important step of analysis in automatic diagnosis of heart sounds. This algorithm first applies a low-pass filter with a cutoff frequency at 882Hz, to reduce noise outside the range of interest, and then converts to energy with a normalized average Shannon energy of the PCG signal. They compared different envelope methods and demonstrate that the Shannon entropy is the best to attenuate the effect of high amplitude noise that contaminates the envelope.

The peaks are detected based on the following criteria: one peak is always picked up; more than two neighbor peaks detected means the existence of split first or second heart sound, so the first peak is picked up in order to get the onset of each sound. To exclude peaks that do not correspond to any of the heart sounds, the following criteria are used: (1) when two peaks appear within 50 ms, which is the largest normal sound interval, and the energy of the first peak is not too small compared to that of the second one, the first peak is chosen, otherwise the second peak is selected; (2) when the interval between two adjacent peaks exceeds 50 ms, their energies are compared. If the energy of the first peak is larger than that of the

second one, and the last interval meets certain consistency of every second interval, they pick up the first one, otherwise the second one is picked.

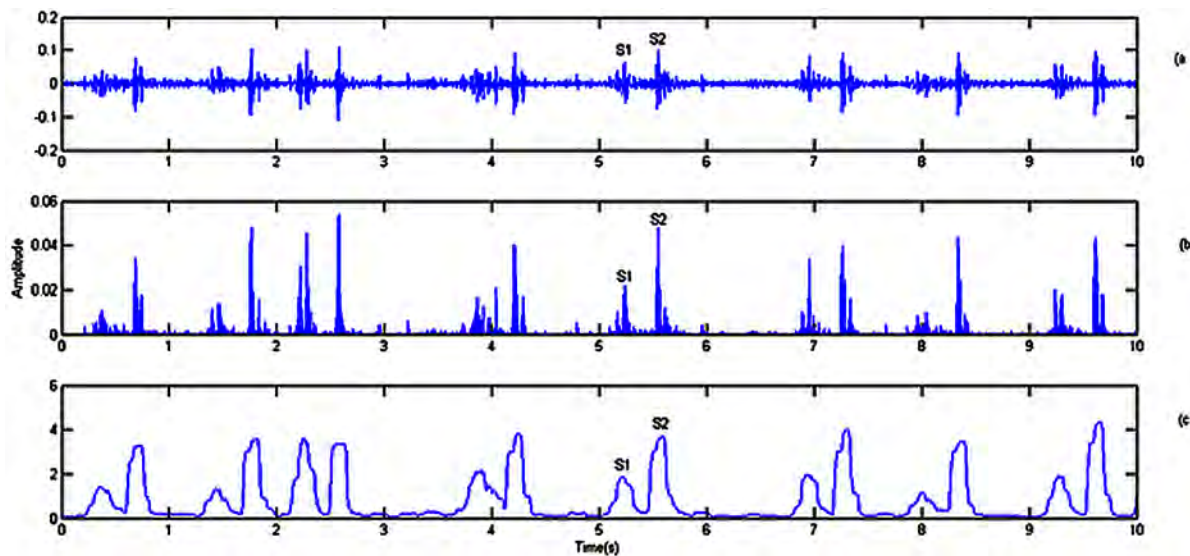
To find which peaks are S1 and S2 the authors assume that the largest interval of a recording (within 20 seconds) is the diastolic period, and that the systolic period is relatively constant compared to the diastolic one. They decided to confine the duration of heart sound between 20 ms and 120 ms. The performance evaluation was made using 515 PCG signal segments recorded from 37 subjects, including normal and abnormal (Liang, Lukkarinen, & Hartimo, 1997).

There are more recent and complex studies of this problem like (Nieblas, Alonso, Conte, & Villarreal, 2013) that uses a matching pursuit algorithm achieving a detection rate between 96 to 97.5%, or (Papadaniil & Hadjileontiadis, 2014) that analyses PCG by employing ensemble empirical mode decomposition combined with kurtosis features to locate the presence of S1, S2, achieving a detection rate of 94.5%.

To diagnose pathologies from the right side of the heart we have to separate the aortic and pulmonary components of S2 and that it is a very difficult task from a computational view.

In Bartels et al. the authors explored the non-invasive BP estimation by heart sound pattern analysis. Having a total of 17 patients participating in this study, measurements of BP were made in the radial artery catheter, also gathering synchronously PCG and ECG data. Only 9 of 17 were sufficiently free from disturbances to be used to investigate the PCG features. The accentuated second heart sound of a patient was converted into a computer-aided pattern recognition process for the second heart sound. After a learning phase of the model, during which the second heart sound is recorded, analyzed and correlated with a set of systolic BP values of the individual patient (calibration phase), the computer is able to determine the future systolic BP values of the same patient from the acoustic spectrum of the second heart sound with an error of 6 to 8 mmHg (Bartels & Harder, 1992).

Figure 2. a) Filtered phonocardiogram segment; b) shannon energy segment; c) shannon energy with a moving window segment



In a different perspective Chen et al. tested a noninvasive method based on spectral analysis of S2 to estimate PAP in 89 patients with a bio-prosthetic heart valve. The technique was compared with continuous-wave Doppler estimation of PAP in these patients. The heart sounds were recorded at the pulmonary area on the chest and digitized to a computer. The spectra of S2 and its components A2 and P2 were computed using a moving window fast-Fourier transform. In Figure 2 we can see an example of a PCG signal from a patient with PAH, and an envelope calculated with a moving window over the Shannon energy; in all stages we can see S1, S2 and its sub-components A2 and P2.

Seven features were extracted from these spectra: dominant frequency of S2 (F_s) corresponds to the frequency associated with the maximal amplitude of the power spectra of S2; dominant frequency of P2 and A2 (F_p), dominant frequency of A2 and P2 (F_a), quality of resonance of P2 (Q_p), which is defined as: where $(f_2 - f_1)$ represents the difference between the frequency values above and below F_p , respectively, for which the amplitude of the power spectrum falls to 50% of

the maximal value; quality of resonance of A2 (Q_a), similarly to Q_p :

The ratio of the dominant frequencies of P2 and A2 (F_p/F_a), which can be considered as a type of normalization between patients; the ratio of the quality factors between P2 and A2 (Q_p/Q_a). The statistical analysis performed showed that the best estimation of PAP obtained by spectral phonocardiography was provided by the following equation, as evaluated by the Pearson linear correlation coefficient, with a standard estimation error of 5.6 mmHg, (Chen, Pibarot, Honos, & Durand, 1996):

In Zhang et al. the authors propose a method to distinguish the two components of the S2, which may be overlapped. For that purpose they use the matching pursuit method to decompose a signal into a series of time-frequency atoms by using an iterative process for the PCG. Compressing or expanding the frequency band of the signal can be performed without perceptible change in its spectral characteristics. It is also possible to shift a frequency band up or down to a desired frequency range without changing the temporal properties of the signal. With these characteristics

in mind their technique has been tested on 11 PCG containing heart sounds and different murmurs. The spectrogram and a matching pursuit based Wigner-Ville distribution were used for visual comparison in the time-frequency domain. The results showed that their technique is suitable and effective for the time-frequency scale transformation of the heart sounds (Xuan Zhang, Durand, Senhadji, Lee, & Coatrieux, 1998).

Xu et al. in an animal study, used a nonlinear transient chirp signal modeling approach for the analysis and synthesis of the overlapping A2 and P2 sound components of the second heart sound. This approach is based on the time-frequency representation of multicomponent signals for estimating and reconstructing the instantaneous phase and amplitude functions of each component. The basic procedure consists in estimating the component which has the higher energy first, because it is the component most easily identified in the time-frequency plane. Then, the estimated component is subtracted from the original signal, after coherent alignment using a cross-correlation technique. The method is then re-applied iteratively until all components are detected and isolated. Each component is defined by a pair of functions describing its instantaneous amplitude (signal envelope) and its instantaneous phase. In practice, the instantaneous frequency (IF) of A2 and P2 are derived from the time-frequency representation of the S2 signal and the phase functions are then estimated by integrating the IFs as a function of time. For decomposition of a multicomponent signal time-frequency this is a good tool, especially when each component is located in a different region of the time-frequency plane or has limited overlap with the other components.

They considered that the splitting interval between A2 and P2 and the exponentially decaying IF functions of both components could be advantageously exploited in the time-frequency plane to filter out and extract each component from the background noise. Since it is known that the A2 component generally appears first, its IF is estimated first using the Wigner-Ville distribution

of the S2 signal and a masking operation (time-frequency domain filter) to remove the cross terms, the P2 component, and the background noise.

The technique can also be applied to physiological signals having three or more nonlinear chirp components, but the results show that is very effective for extracting the two components, even in the presence of noise. The normalized root-mean-squared error between the original A2 and P2 components and their reconstructed versions varied between 1% and 6%, proportionally to the duration of the overlapping interval, and it increased by less than 2% in the presence of noise

The validated technique was then applied to S2 components recorded in pigs with normal or high PAP. The results show that this approach can successfully isolate and extract overlapping A2 and P2 components from successive S2 recordings obtained from different heartbeats of the same animal, as well from different animals (Jingping Xu et al., 2001).

Tranulis et al. objective was also to develop a non-invasive method for the estimation of PAP using a neural network (NN) and features extracted from the S2. To obtain the information required to train and test the NN, they conducted an experiment in 9 pigs and analyzed the heart sounds under different conditions of PAP, including pulmonary hypertension. The data recorded from each subject was the ECG, the PCG, and the PAP. Around 15 to 50 S2 sounds were isolated for each PAP stage and for each animal studied. A Coiflet wavelet decomposition and a pseudo smoothed Wigner-Ville distribution were used to extract features from the S2 and train a one hidden layer NN using 2/3 of the data. The NN performance was tested on the remaining data. NN estimates of the systolic and mean PAPs were obtained for each S2 and then ensemble averaged over the 15 to 50 S2s selected for each PAP stage. Their classification accuracy, using a 23 mmHg mean PAP and a 30 mmHg systolic PAP thresholds between normal PAP and pulmonary hypertension was of 97% and 91% respectively (Tranulis, Durand, Senhadji, & Pibarot, 2002).

In a different study, Xu et al. employed the same method as described in (Xu et al., 2003) for the estimation of PAP in animal and human subjects. It was shown that this new non-invasive method based on advanced signal processing of the second heart sound provides an accurate estimation of the PAP in those populations.

Dennis et al. study describe the development of a prototype system, called pulmonary hypertension diagnose (PHD), that uses patient data with machine learning algorithms to build models of how pulmonary hypertension affects heart sounds. They designed the PHD with three sets of parameters: chest wall location, heart sound features subset, and machine learning algorithm. In the first parameter, chest wall location, they chose 5 locations and try to determine which locations are most useful for PH diagnosis. Then they extract 46 features for each recorded heart sound in the PCG traces. Some of these features were proprietary (Inovise Medical, 2013) and correspond to the intensity or width of the heart sounds, but they also use several features already described in literature such as dominant frequencies of S2, A2, and P2, quality of resonance of A2 and P2, 8 types of ratios such as F_{P2}/F_{A2} and Q_{P2}/Q_{A2} , the splitting interval of S2, and systole duration. For the classification set 5 models of machine learning algorithms were used: decision tree, k-nearest neighbors, multilayer perceptron, naive Bayes, and support vector machine. Their main findings from the feature list were: the features extracted from the S2 sound are helpful in classifying a patient and also the percentage of the heartbeat taken for ventricular systole was much more predictive than the absolute duration time of ventricular systole, the others features had less impact on the predictive model.

The PHD tries to determine whether the PAP is above or below a certain threshold, diagnosing pulmonary hypertension in the validation set with 77% accuracy and 0.78 area under the receiver-operating-characteristic curve (Dennis, Michaels, Arand, & Ventura, 2010).

An interesting work from Angel Lopez-Candales et al. studied the split of the S2 in people with chronic pulmonary hypertension. The authors used a pulsed Doppler to measure the total duration of right and left ventricular outflow tract spectral signal to identify this narrowing between the ejections of both ventricles. Doppler data and standard measures of right and left ventricles were collected from 85 patients divided into two groups according to their estimated pulmonary artery systolic pressure (estimated by echocardiogram). Chronic pulmonary hypertension patients had a shorter total duration between right ventricle output and left ventricle output ejection time than individuals without pulmonary hypertension (-15 ± 16 ms vs. 22 ± 14 ms; $P < 0.01$), that explain the split second heart sound in chronic pulmonary hypertension patients (Lopez-Candales, Edelman, Gulyasy, & Candales, 2011).

Peretto et al. in their study present a comparison between two non-invasive approaches for the estimation of PAP. The study was conducted on 40 athletes, aged between 8 and 67 years, practicing sports at different levels. The first method uses the Pulsed Doppler which has been used for the measurement of maximum tricuspid transvalvular pressure gradient. The second method is based on the analysis of PCG sounds and particularly of the S2 and uses ECG as a trigger, extracting features from S2 and its components as suggested in previous studies. The signals were digitally acquired simultaneously, and analyzed with a software developed in LabView™. Results show encouraging matching between the two methods and also that the proposed method may be relevant for PAP estimation on athletes without significant cardiovascular diseases and practicing different sports (Peretto et al., 2011).

Chan et al. examined the relationship between acoustic characteristics of S1 and S2 and underlying cardiac structure and hemodynamics in patients with isolated PAH and controls. They did a cohort study in referral and community hospitals, evaluating 40 PAH patients undergoing

right-heart catheterization with a digital acoustic cardiography and two-dimensional transthoracic echocardiography. As control they had 130 participants without clinical or hemodynamic evidence of PAH or congestive heart failure. Ten second artifact-free recordings were selected and analyzed utilizing proprietary software with wavelet-based signal processing techniques for acoustic analysis. A time-frequency analysis allows the separation of heart sounds from murmurs and an Audicor algorithm was used to reduce external noise like speech and movement. Two acoustic cardiographic characteristics of S1 and S2 were analyzed: intensity and complexity. A value for heart sound intensity is generated based on the peak-to-peak amplitude of the sound, and complexity is determined using time-frequency measures of width, intensity and frequency content of the signal. Complexity and intensity were normalized and expressed as a ratio between S2 and S1 to account for variations in anthropomorphic characteristics. Then univariate and multivariate linear logistic regressions were applied to examine the relationship between hemodynamic variables and S1 and S2 complexity. The authors reported that acoustic characteristics of both S1 and S2 are related to the severity of PAH and are associated with right ventricular enlargement and systolic dysfunction. The reciprocal relationship between S2 and S1 complexity may also reflect the underlying ventricular interaction associated with PAH (Chan et al., 2013).

Some studies with noninvasive experiments found that the time delay from the peak of ECG R wave to the onset of the second heart sound had a close inverse correlation with arterial systolic BP (Xin-yu Zhang, Macpherson, & Zhang, 2008); Lim et al. used PCG to determine whether the ratio of S1 to S2 analyzed by esophageal stethoscope transmitted via a wireless bluetooth connection can be an accurate indicator of the BP. A total of 33 adult subjects, without cardiac disorder and with normal heart rhythm were enrolled and randomly selected to be subjects of the study,

also all patients undergo a total intravenous anesthesia induced with propofol and remifentanyl by target controlled infusion. Two microphones were used: one for the acquisition of heart sound by connecting it to the esophageal stethoscope; the other was used to measure the background noise in the operating room. After transmission of the heart sounds measured with the esophageal stethoscope to the receiver by using a bluetooth module, the portion of the recorded data through the esophageal stethoscope that coincided with the noise in the operating room was removed; noise such as the respiration sound was removed with low-pass filtering. S1 and S2 were detected with computation of the ratio of S1 to S2. Correlations between the systolic BP with S1, S2 and S1 to S2 ratio were examined. The ratio of S1 to S2 presented the highest correlation with the systolic BP (Lim et al., 2013).

Smith et al present a machine learning model based on heart sounds characteristics to estimate PAP; some features were derived from heart sounds, and machine learning techniques were applied on a large patient dataset to choose a specific model and features for PAP values prediction. Two machine learning models were tested, support vector machines and multilayer perceptron. A greedy search through the 38 possible features was conducted within the following categories; binary features, dominant frequency, quality of frequency, power, splitting interval, ratios, systole duration and heart rate, using a 109-patient cross-validation to find the most predictive features. The best model had a standard estimate error of 8.3mmHg (Smith & Ventura, 2013).

The methods presented in the literature demonstrate that several important features of the cardiovascular system may be estimated through the relations between the signals measured on the body surface; the heart sound analysis presents encouraging results for the systemic and pulmonary blood pressure estimation, however further studies are necessary to improve accuracy and validate the technique for clinical use.

SOLUTIONS AND RECOMMENDATIONS

The use of heart sounds for BP estimation is a very promising noninvasive method, however it needs to be further studied so that it may be widely used in clinical practice, and as a complement to existing methods. There are several issues in the method that need to be better understood, one of these problems, and also one of the most complex, is the separation of A2 and P2 components. Advancing the techniques for heart sound segmentation as referred earlier will allow to improve overall results; several studies use the ECG as a triggering method for heart sound segmentation, which although improves S2 detection, increases complexity of such a system. Commercially available digital stethoscopes, already define the best characteristics for heart sound acquisition, however the lack of a data collection standard tends to generate less homogeneity of the various studies. Noise interference is also a concern, and it should be controlled for better results. The definition of a data collection standard will play an important role for the success of such a system. The impact of different biotypes on sound quality also needs to be addressed, as well as the determination of the best set of parameters for BP estimation. In general there is a need for larger multicentric studies to validate the method.

FUTURE RESEARCH DIRECTIONS

To estimate BP using heart sounds there are three areas that are critical and must be extensively studied in the coming years, starting with the most complex, the separation of A2 and P2 components. Also complex enough is what kind of features should be extracted from heart sounds to estimate BP, it is necessary to look for a faster way and a more analytical method to find an optimal feature subset for on-line implementation, therefore these issues should continue to be subject of specific

and more focused studies. Finally the estimation models algorithm that best correlate heart sound features with BP also have a lot of potential to be further developed, although promising results were already achieved.

CONCLUSION

Digital phonocardiography and the analysis of the second heart sound is an evolving area; studies demonstrate that it is possible to separate the A2 and P2 components of S2, even when the overlapping interval of the two components is relatively high. Since the operating pressure and basic elastic properties of the aortic and pulmonary valves and vessels are different, their instantaneous resonant frequency functions should also be generally different. It is possible with an auscultation to determine pathologies from the left side of the heart (A2 component and systemic circulation), and from the right side of the heart (P2 component and pulmonary circulation). The advantage of using auscultation to detect heart and lung pathologies is the fact that it is a noninvasive method, with a simple and easily accessible technology. These systems could bring advantage due to its practical application, making the management of PAH patients easier, and improving patient comfort.

REFERENCES

- Bartels, A., & Harder, D. (1992). *Non-invasive determination of systolic blood pressure by heart sound pattern analysis*. Academic Press.
- Chan, W., Woldeyohannes, M., Colman, R., Arand, P., Michaels, A. D., Parker, J. D., & Mak, S. et al. (2013). Haemodynamic and structural correlates of the first and second heart sounds in pulmonary arterial hypertension: An acoustic cardiography cohort study. *BMJ Open*, 3(4), e002660. doi:10.1136/bmjopen-2013-002660 PMID:23572199

- Chen, D., Pibarot, P., Honos, G., & Durand, L. G. (1996). Estimation of pulmonary artery pressure by spectral analysis of the second heart sound. *The American Journal of Cardiology*, 78(7), 785–789. doi:10.1016/S0002-9149(96)00422-5 PMID:8857483
- Chung, E., Chen, G., Alexander, B., & Cannesson, M. (2013). Non-invasive continuous blood pressure monitoring: A review of current applications. *Frontiers of Medicine*, 7(1), 91–101. doi:10.1007/s11684-013-0239-5 PMID:23345112
- Dennis, A., Michaels, A. D., Arand, P., & Ventura, D. (2010). Noninvasive diagnosis of pulmonary hypertension using heart sound analysis. *Computers in Biology and Medicine*, 40(9), 758–764. doi:10.1016/j.compbiomed.2010.07.003 PMID:20691967
- Edwards Lifesciences, L. L. C. (2009). *Normal Hemodynamic Parameters and Laboratory – Adult*. Retrieved from www.edwards.com
- Galie, N., Hoeper, M. M., Humbert, M., Torbicki, A., Vachiery, J.-L., Barbera, J. A., & Zamorano, J. L. et al. (2009). Guidelines for the diagnosis and treatment of pulmonary hypertension: The Task Force for the Diagnosis and Treatment of Pulmonary Hypertension of the European Society of Cardiology (ESC) and the European Respiratory Society (ERS), endorsed by the International Society of Heart and Lung Transplantation (ISHLT). *European Heart Journal*, 30(20), 2493–2537. doi:10.1093/eurheartj/ehp297 PMID:19713419
- Inovise Medical, I. (2013). *Audicor® 200s*. Retrieved from www.audicor.com
- Liang, H., Lukkarinen, S., & Hartimo, I. (1997). Heart sound segmentation algorithm based on heart sound envelopgram. *Computers in Cardiology*, 24, 105–108. doi:10.1109/CIC.1997.647841
- Lim, K. H., Shin, Y. D., Park, S. H., & Bae, J. H. (2013). *Correlation of blood pressure and the ratio of S1 to S2 as measured by esophageal stethoscope and wireless bluetooth transmission*. Academic Press.
- Lopez-Candales, A., Edelman, K., Gulyasy, B., & Candales, M. D. (2011). Differences in the duration of total ejection between right and left ventricles in chronic pulmonary hypertension. *Echocardiography (Mount Kisco, N.Y.)*, 28(5), 509–515. doi:10.1111/j.1540-8175.2010.01377.x PMID:21539602
- Nieblas, C. I., Alonso, M. A., Conte, R., & Villarreal, S. (2013). High performance heart sound segmentation algorithm based on Matching Pursuit. In 2013 IEEE Digital Signal Processing and Signal Processing Education Meeting (DSP/SPE) (pp. 96–100). IEEE. doi:10.1109/DSP-SPE.2013.6642572
- Nigam, V., & Priemer, R. (2004). Cardiac sound separation. *Computers in Cardiology*, 497–500. doi:10.1109/CIC.2004.1442983
- Papadaniil, C. D., & Hadjileontiadis, L. J. (2014). Efficient Heart Sound Segmentation and Extraction Using Ensemble Empirical Mode Decomposition and Kurtosis Features. *IEEE Journal of Biomedical and Health Informatics*, 18(4), 1138–1152. doi:10.1109/JBHI.2013.2294399 PMID:25014929
- Peretto, L., Poletti, G., Longhini, L., Vitolo, A., Santini, C., Elettrica, I., ... Bologna, U. S. L. (2011). *A comparison study of two indirect methods for measuring the systolic pulmonary arterial pressure*. Academic Press.
- Smith, R., & Ventura, D. (2013). A general model for continuous noninvasive pulmonary artery pressure estimation. *Computers in Biology and Medicine*, 43(7), 904–913. doi:10.1016/j.compbiomed.2013.04.010 PMID:23746733
- Tranulis, C., Durand, L. G., Senhadji, L., & Pibarot, P. (2002). Estimation of pulmonary arterial pressure by a neural network analysis using features based on time-frequency representations of the second heart sound. *Medical & Biological Engineering & Computing*, 40(2), 205–212. doi:10.1007/BF02348126 PMID:12043802

Xu, J. (2001). Extraction of the aortic and pulmonary components of the second heart sound using a nonlinear transient chirp signal model. *IEEE Transactions on Bio-Medical Engineering*, 48(3), 277–283. doi:10.1109/10.914790 PMID:11327495

Xu, J., Durand, L.-G., & Pibarot, P. (2003). A new, simple, and accurate method for non-invasive estimation of pulmonary arterial pressure. *Heart (British Cardiac Society)*, 88, 1–5.

Young, E. R. (2001). Bates Guide to Physical Examination and History Taking. *Anesthesia Progress*, 48(2), 72–73. doi:10.1080/095852097330784

Zhang, X., Durand, L. G., Senhadji, L., Lee, H. C., & Coatrieux, J. L. (1998). Time-frequency scaling transformation of the phonocardiogram based of the matching pursuit method. *IEEE Transactions on Bio-Medical Engineering*, 45(8), 972–979. doi:10.1109/10.704866 PMID:9691572

Zhang, X., Macpherson, E., & Zhang, Y. (2008). *Relations Between the Timing of the Second Heart Sound and Aortic Blood Pressure*. Academic Press.

ADDITIONAL READING

Buxi, D., Redout, J.-M., & Yuce, M. R. (2015). A survey on signals and systems in ambulatory blood pressure monitoring using pulse transit time. *Physiological Measurement*, 36(3), R1–R26. doi:10.1088/0967-3334/36/3/R1 PMID:25694235

Dadfarmay, S., Berkowitz, R., Kim, B., & Manchikalapudi, R. B. (2010). Differentiating Pulmonary Arterial and Pulmonary Venous Hypertension and the Implications for Therapy. *Congestive Heart Failure (Greenwich, Conn.)*, 16(6), 287–291. doi:10.1111/j.1751-7133.2010.00192.x PMID:21091615

Delgado, J. F., Conde, E., Sánchez, V., López-Ríos, F., Gómez-Sánchez, M., Escribano, P., & de la Calzada, C. S. et al. (2005). Pulmonary vascular remodeling in pulmonary hypertension due to chronic heart failure. *European Journal of Heart Failure*, 7(6), 1011–1016. doi:10.1016/j.ejheart.2004.10.021 PMID:16227139

Foo, J. Y. A., Lim, C. S., & Wang, P. (2006). Evaluation of blood pressure changes using vascular transit time. *Physiological Measurement*, 27(8), 685–694. doi:10.1088/0967-3334/27/8/003 PMID:16772667

Mitchell, J. R., & Wang, J.-J. (2014). Expanding application of the Wiggers diagram to teach cardiovascular physiology. *Advances in Physiology Education*, 38(2), 170–175. doi:10.1152/advan.00123.2013 PMID:24913453

Peter, L., Noury, N., & Cerny, M. (2014). A review of methods for non-invasive and continuous blood pressure monitoring: Pulse transit time method is promising? { . *IRBM*, 35(5), 271–282. doi:10.1016/j.irbm.2014.07.002

Thenappan, T., Shah, S. J., Gomberg-Maitland, M., Collander, B., Vallakati, A., Shroff, P., & Rich, S. (2011). Clinical characteristics of pulmonary hypertension in patients with heart failure and preserved ejection fraction. *Circulation: Heart Failure*, 4(3), 257–265. doi:10.1161/CIRCHEARTFAILURE.110.958801 PMID:21411741

KEY TERMS AND DEFINITIONS

Blood Pressure (BP): The pressure exerted in the walls of blood vessels by circulating blood.

Digital Stethoscope: Equipment used to digitally amplify and record sounds from the body.

Electrocardiogram (ECG): A record and representation of the electrical activity of the heart.

Heart Sound Analysis for Blood Pressure Estimation

First Heart Sound (S1): Sound resulting from mitral and tricuspid valves closure.

Phonocardiogram (PCG): The record and representation of the heart sound.

Pulmonary Arterial Hypertension (PAH):
An increase of blood pressure in the pulmonary artery.

Second Heart Sound (S2): Sound resulting from aortic and pulmonary valves closure.

Sociological Perspectives on Improving Medical Diagnosis Emphasizing CAD

C

Joel Fisher

Department of State, United States Government, USA

INTRODUCTION

Health care costs worldwide are increasing because of new medicines, new techniques, and more expensive and extensive research on diseases. It is essential that health care delivery systems be implemented that take advantage of these advances in a cost effective economic manner. One critical aspect of the health care delivery system is the improvement in diagnosis of disease. This paper emphasizes diagnosis and the need to collect key patient information “pivotal information” at the earliest possible point in the patient’s disease process, and to put such pivotal information in physicians’ hands. There is a potential for huge benefits in cost savings and greater effectiveness of treatment from these actions, but challenges must be overcome to realize their full benefits. The major problems include (1) market incentives in the current health care system fail to encourage collecting pivotal medical information as early as possible (2) physician resistance to some of these ideas and (3) technical and ethical problems that remain to be solved.

BACKGROUND

Previously, Fisher and Fisher (2014) discussed the problems of improving medical diagnosis as both a management and information problem. Their analysis focused on three specific strategies: improving medical histories of patients by using discourse analysis, public education about

selected major diseases that if ignored can lead to health impairment and even death, and use of computer assisted diagnosis (CAD).

THE PROCESS OF DIAGNOSIS

Diagnosis is sociologically a process of bounded rationality, consisting of evidence gathering punctuated by a series of conjectures about a patient’s ills before arriving at the *best* conjecture and treatment plan. This view of diagnostics originates from Zeldenrust’s (1990) version of the garbage can model of organizational behavior that sees diagnosis as a linkage between a demand (patient with an illness) and a solution (physician guesses nature of underlying problem and offers a plan for treatment). The solution may include medicine, behavioral modification such as bed rest, change in diet etc., and perhaps tests to rule out certain competing diagnostic hypotheses consistent with applicable constraints. The physician builds on the information learned not only from the patient reports but also on test results and on other objective data obtained. The physician treats symptoms uncovered until the patient either is cured and a final diagnosis is achieved or at least until all competing hypotheses about the condition are ruled out.

Improving the diagnostic process hinges on rapidly expanding the diagnostic information available to the physician. This diagnostic information is not just any data but the sort that expands the number of hypotheses of possible diagnoses,

DOI: 10.4018/978-1-5225-2255-3.ch088

i.e. pivotal data. The physician can then examine the maximum number of plausible hypotheses. The idea is that *the more hypotheses the physician needs to rule out at the outset the more likely* the physician will find the disease(s) behind the patient's complaint and provide the proper treatment (including referral for specialized services). The process also covers multiple possible pathologies that may exist at any moment, although the clinician may place ordering priorities on these possible multiple pathologies.

A. The Medical History and Discourse Analysis

Medical authorities agree that the most important basic information beyond a description of the patient's medical complaints is the patient's health history. An encounter with any provider usually begins with some kind of questionnaire or interview with regard to that history. The needs of different providers often result in different kinds of questionnaires or approaches about health history. They likely cover all the same major issues, but the differences can be important. There is little evidence that these initial interviews involve people trained in discourse analysis. Since the Fisher and Fisher (2014) analysis, almost no real advances or progress have been described. In fact, clinical people are almost expected to develop interviewing skills by "osmosis" on the job. A surprising finding was that neither nursing nor medical curricula include formal course work on this although some medical schools give orientation lectures and guides to such interviewing skills. The technique is not included in the so-called "patient assessment courses."

Discourse analysis, popular in the United Kingdom, is basically the study of language in context (Elwyn and Gwyn, 1999). The subject shares much in common with assessments of communication that look at gestures (often referred to as "body language") –the "silent language" (Edward Hall, 1973). The discipline examines conversational processes: how is one version of events selected

over another? How is a familiar reality described to give it an unquestionable authority? It then brings to bear on the patient's diagnosis *all the information* that patients are providing, and *saves it for future use*. It can identify and elicit previously hidden patterns and perspectives (for example, in patient clinics, among health visitors, and in transcripts of interviews conducted by HIV counselors). Their extensive list of references suggests that the technique has been widely applied in health care related situations with promising results.

The technique has a sound basis in social science and can be applied quickly by medical practitioners and should be included in the basic course work of the field. Discourse analysis seems to be easiest to implement in medical practices serving ethnically relatively homogeneous areas whose residents have a relationship with the medical staff. This has major implications for CAD.

B. Public Education

Although public education is not a diagnostic procedure, it garners public cooperation in early diagnosis and treatment of many serious illnesses. Its purpose here is not to impress on people to take better care of themselves, but rather to have targeted campaigns about specific conditions that if treated early are easily cured or managed, but if ignored, become extremely expensive to treat and perhaps lethal. Public education should put *pivotal* information in the hands of both the physicians and patients as early as possible in the disease process to maximize the probability of effective treatment at the lowest possible cost.

Patients who recognize the possible variety of symptoms of heart attack and stroke etc. will more likely seek treatment when the physician's intervention may be most certain to work, and probably least expensive. Some diseases, such as colon and pancreatic cancers often show few or no early warning symptoms and only make you uncomfortable enough to seek medical attention in their advanced stages. These diseases typify a propensity of people to avoid medical attention

until they are in pain or their normal functioning seems to have been greatly impaired. Much current public education work seems to have a “disease of the month” flavor, reflecting what specific illnesses have become objects of attention in the news media.

C. Computer Assisted Medical Diagnosis (CAD)

The most important suggestion is CAD, and it is where the most progress has been made. CAD does not diagnose the problem but only suggests a range of hypotheses for the physician to check out. Put simply, computerization of the diagnostic process might advance medicine by looking at individual symptoms, clustering symptoms, and providing a kind of pattern matching algorithm that generates the needed conjectural hypotheses consistent with the information in the medical data banks.

CAD offers the major advantages of a nearly unlimited virtual library or information facility, a communications facility, and an inexpensive infrastructure. Scadding (1967) hoped that physicians would soon take advantage of the possibility of instantaneous correlation of groups of clinical and laboratory findings with the established features of specific diseases. Physicians would also capitalize on the computer’s ability to *suggest* a wide range of possible diagnoses without necessarily discriminating among them although it could include risk assessment algorithms that rank them by probability of occurrence and seriousness of outcome if left untreated). These special information tools have become a major focus for big data and information specialists in many fields.

When Fisher and Fisher (2014) examined the issue a universal CAD system did not exist in the United States. Early considerations by the military to develop one remained basically in the conceptual stage. There is still none, but a review of CAD systems at large indicated a number of interesting findings.

1. The most widespread and successful medical CAD usage has been in medical imaging, a highly limited area of diagnosis. Qiang and Nishakawa (2015) have produced a state-of-the-art compendium of its use for various types of imaging to detect and evaluate carcinomas. That may highlight the difficulties in expanding the tool to a wider area of medical diagnostic concerns. Watson, a supercomputer from IBM, involves a look at diagnosing certain kinds of cancer. (Mearns, Lucas, 2013) One of the tasks of Watson is to continuously update and comment on the literature. The scientists working on Watson hope to make it widely available soon. Watson’s developers at IBM mention that its major feature is that it can focus its massive data processing capabilities on personal genomes. Note, however, that work involving the human genome tends to raise certain ethical and privacy issues, major sociological concerns. In looking at a CAD system, one is really not currently easily able to relate to the sociological issues of privacy and ethics beyond indicating that the operation and access of the system to information be in accord which respects privacy and adheres to ethical norms.
2. A CAD system has been used extensively in France (Seka, Fresnel et al, 1997) and the French researchers have proposed expansion for the Internet at large. Although France has a diverse ethnic and demographic population, compared to the United States’ diversity of ethnic and demographic features, the French might be considered a somewhat ethnically and demographically a smaller, possibly homogenous cultural unit. The French system accommodates information on about 10,000 diseases. This may or may not be a suitable number for system guidelines, but since various forms of cancer, a disease with probably the most variations and types has over 1,000 separate disease designations, it may be satisfactory.

3. Various groups are using CAD systems in rural India and areas of sub-Saharan Africa where the medical problems have unique geographic and ethnographic features. (Kalode, Kempar, and Deshpande, 2012; Friedman, 2009). A major thrust in CAD systems for diagnosis is for rural areas that are medically underserved mainly because sociologically this is not an area where clinicians seem anxious to practice. It often lacks infrastructure, access to researchers and research facilities, and other desirable features.

The *sociological* implications are that a *universal CAD may not necessarily be either optimal or desirable*. Rather, one might develop a series of CAD systems related to smaller regions with ethnic, geographical and demographic unique considerations as well as the most common pathologies universally encountered, unique to certain regions with a subroutine to search other systems based on information in a patient's history about travel, residency and other factors that may become important. This is the most important new direction that this diagnosis work should take.

CAD's biggest technical hurdle is the fiendishly complex nature of a diagnosis (see: Innocent & John, 2004). Some symptoms, such as fever, are always important; but for other symptoms the importance depends on context. An effective CAD in generating hypotheses will include a provision to note that certain symptoms are especially significant in the presence of others in confirming a diagnosis.

The many and diverse bases for deciding a disease behind the symptoms include cause, description, site, symptom/syndrome, variation from statistical norm, and treatability. Further, diagnoses change as individuals, professions or society dynamically modify the boundaries of what they consider respectively normal or problematic. For example, not all physicians accept chronic fatigue syndrome or even fibromyalgia as diseases. In the 1970's United Nations publications listed

world-wide fatalities from juvenile prostate cancer as a single cause of death, but the most recent compendia list over six different categories of the disease. Psychiatric disorders continue to be sources of controversy as to what is a symptom or syndrome or condition. This has even prompted legal arguments as to the legitimacy of certain diagnoses. These are mainly *social issues*.

Jutel (2011) endorses Aronowitz (2001) that *social influences* have largely determined which symptom clusters have *become* diseases. CAD may not cover every condition that physicians treat because no societal or professional consensus exists as to whether the condition is a disease and to what is its appropriate treatment. That fact should not inhibit CAD. Computer programs can assign conditions with appropriate data in special categories where there is no societal or professional or societal consensus, or such identity has not been sought or even rejected. A CAD system should accommodate certain symptoms or syndromes as identifiable even if they are professionally not recognized, simply because they occur so often that to ignore them is foolhardy. The patient still suffers from the condition regardless of any societal or professional agreement as to the legitimacy of the diagnosis.

CAD can assist in reducing misdiagnosis. Since the Fisher and Fisher (2014) paper, this has become something of a two-sided issue. On one side, CAD can offer a comprehensive list of hypotheses and indicate what additional pivotal facts, such as test results, are needed to confirm each of them. On the other side, there seems some feeling that reliance on CAD could result in increasing errors of medical malpractice. Several websites of attorneys have appeared recently soliciting clients who may have been improperly treated because the diagnosis and treatment were traceable to hypotheses of recommendations from a CAD system. This should not discourage development of CAD because it only provides information, it does not formally diagnose.

In discussing misdiagnosis, Bradley (1993) found that the physician typically begins with three

or four hypotheses but rarely more than seven. He counsels physicians who cannot avail themselves of a CAD that one of the four hypotheses should be a “catch-all”— *i.e.* any of the above, or even none of the above. There needs to be considerable computer work to specify the features and operating protocols for this “catch-all” category.

When facts support one of the three starting hypotheses, it is important to maintain the possibility of the true diagnosis being something else. Good diagnosticians do this, or something akin to it. The author of this paper lives in an area characterized by a fungal disease known as “Valley Fever.” It has symptoms that may suggest flu, various kinds of pneumonia, tuberculosis, unusual allergy situations, certain kinds of seizure disease orders and a host of more common ailments (Malo et al, 2014). A CAD system for the physician in this area should include Valley Fever in all proposed hypotheses if the symptoms comport. Another disease that has caused similar misdiagnoses because of its many symptom manifestations is Lyme disease (Pachner, 1989; Stanek, 2012; see also fact sheet from CDC, 2015).

Pivotal facts often differ between men and women. Recent studies have shown that men and women tend to present different sets of symptoms for heart attacks and strokes. These gender differences require considering more possibilities when certain symptoms are present. Campbell (2015) has written extensively on this issue with respect to women’s heart conditions.

Sometimes a misdiagnosis arises from prematurely accepting a positive confirmation of an initial. Bradley warns physicians to guard against misusing information by over-emphasizing positive findings when a negative finding may be just as diagnostically significant. There is a natural tendency to prefer positive evidence proving a hypothesis over negative evidence disproving it. A test that could confirm a diagnosis because of a positive outcome tends to mean that the assumed diagnosis that called for the test is likely to be correct. However, to improve diagnosis, a negative test [finding] should have the converse effect of

making the original assumed diagnosis less likely. Bradley notes that often the clinician ignores this negative evidence. The disease often gives negative or ambiguous test results from immunological and serological studies despite a patient having the disease). This becomes especially important when the confirming test result is subject to a false positive error, or when a confirming test result can also suggest other hypotheses, which may or may not have been in the original cluster of assumed diagnostic conjectures. A CAD should include a feature that alerts the provider with guidelines to question if a particular result may originate from assuming that a certain diagnosis is confirmed by a certain test.

A negative finding does not always disprove a diagnostic conjecture. Many patients are asymptomatic for certain conditions and a particular test may prove negative for that person even under the condition it is designed to highlight. [see references on Lyme disease at end of paper, Pachner 1989 and Stanek, 2012]. This author had *normal* sinus rhythm from an EKG while having a near fatal heart attack, and only a physician’s recognition of other signs led to a life saving diagnosis.

CAD can reduce misdiagnosis by having a “parallel testing feature”— it can offer a list of hypotheses fitting the same set of facts. A physician not accessing a CAD may settle on just one of the possible hypotheses (without ruling out the others) and choose the wrong one. The frequency of such situations of misdiagnosis is unknown, but missing the correct diagnosis of a disease such as MDR tuberculosis as a possible tropical parasitic infection has serious public health consequences (Naik, 2012).

A CASCADE OF SMALLER CAD SYSTEMS AND A SUPERCOMPUTER

Computer technology continues to advance at exponential levels. The computers have greater storage, more flexibility, smaller sizes and a host of great features which should now be augmented

with the diagnosis requirements. There is a need to link smaller CAD systems individually emphasizing regionally specific diseases into a cascade, or through access to a supercomputer designed to obtain information needed for a patient with an unusual history of travel, residence, or other factors. By keeping local systems small, they prevent management overload with regard to information, and minimize problems related to network function and oversight for the clinician.

The local systems can include provisions for symptoms and diseases which have not yet received either professional or social acceptance or those being debated at various professional levels of practice. The local systems can also indicate which techniques, tests, and protocols are experimental, or have not received peer review but are still in use. The larger system can provide information as to the status of any disease or treatment with regard to national or governmental authorities, both locally and internationally, and provide for appropriate mechanisms of information translation where the local CAD system has linguistic differences from another CAD system's language of inquiry.

Many current medical systems have enhancements that allow patients to access their medical records from home computers and "smart phones" though "patient portals." This technology also allows rapid conferencing among medical colleagues. This could reduce some costs associated with second opinions needed for a diagnosis and increase the quality and extent of care available in medically underserved areas by focusing on the kinds of consultations needed. The patient portal approach could make the CAD usage patient friendly. The proponents of Watson see this as one of their major goals.

CAD can foster major infrastructure improvements. CAD can provide links to consulting physicians in distant medical centers to augment a health care provider's local network. One major system links physicians on call in Australia and Europe to provide 24 hour consultation possibilities because of time zone differences and working hours in the

United States. Even diagnosis and treatment in outer space is possible, as astronauts can follow instructions from remote centers on earth to deliver health services to others in space with them. And we should not overlook the benefits to doctors in medical centers who deliver much of the health care to people who use the emergency room as their family physician.

FUTURE RESEARCH DIRECTIONS

Getting wider physician use of a CAD may be a problem, although this author believes that the physician resistance to CAD as machine replacing their work is largely unfounded. Even if the profession's leading members endorse CAD, others might not readily fall in line. The current United States approach is to produce "best practices" papers and manuals to reflect the new knowledge. Professional societies have taken a positive lead here, but some government agencies, notably the Veterans Administration, see the need for these documents as internal procedures and protocols and have made their use more accepted and widespread. CAD can act as an assistant in producing these documents through its information archival capabilities, as well as by incorporating the results of these documents to assist clinicians when looking at possible hypotheses of diagnosis and treatment.

CONCLUSION

The importance of good diagnostic tools will go far to improve health and health care delivery systems in a new environment of high medical costs, special insurance requirements, record mobility and big data usage. Because quantity and quality of medical care are not evenly distributed in the United States, one way to improve the quality across the country is to focus on improving the accuracy of and speed with which physicians make diagnoses.

A comprehensive CAD offers the current best hope to improve diagnosis. The tool need not be perfect to be valuable in reducing physician diagnostic errors and in making more precise the needs for consultations and second opinions, and possibly in speeding up diagnostics.

The most useful and successful CAD systems for other than special technical protocols or single diseases, are regionally specific. That seems the key to improved diagnostic usage. A series of smaller CAD systems should emphasize the diseases unique and common to certain areas. These smaller systems can contain linkages to other systems or a supercomputer when the patient information, either from medical history or encounter questionnaire/interview shows residence, work, and travel or simply traversing certain locations which feature diseases whose symptoms may confound others or offer additional hypotheses in a specific diagnosis.

REFERENCES

- Aronowitz, R. (2001). When do symptoms become a disease? *Annals of Internal Medicine*, 134(9), 803-808.
- Bary, A. (2012, October 8). A Better Test for Colon Cancer. *Barron's*, p. 25.
- Bradley, G. W. (1993). *Disease, Diagnosis, and Decisions*. Chichester, UK: John Wiley and Sons.
- Cambell, K. R. (2015). *Women and Cardiovascular Disease*. London: Imperial College Press.
- Center for Disease Control. (2015). *Lyme Disease information sheet*. CDC.
- Elwyn, G., & Gwyn, R. (1999). Stories we hear and stories we tell: Analyzing talk in clinical practice. *British Medical Journal*, 318, 16.
- Fisher, D. A. (2007). The bottom line: Offer the colorectal cancer screening test you can deliver. *Gastrointestinal Endoscopy*, 65(4), 646-647. doi:10.1016/j.gie.2006.07.027 PMID:17173919
- Fisher, R. L., & Fisher, J. L. (2014). Improving Medical Diagnosis: A Sociological Perspective. In J. Wang (Ed.), *Encyclopedia of Business Analytics and Optimization* (pp. 1162-1172). doi:10.4018/978-1-4666-5202-6.ch107
- Friedman, E. A. (2009). *Medical Diagnosis for Sub-Saharan Africa. IEEE Technology and Science Magazine*.
- Hall, E. T. (1973). *The Silent Language*. Garden City, NY: Anchor Press/Doubleday.
- Innocent, P. R., & John, R. I. (2004). Computer aided fuzzy medical diagnosis. *Information Sciences*, 162(2), 81-104. doi:10.1016/j.ins.2004.03.003
- Jutel, A. G. (2011). *Putting a Name to It*. Baltimore, MD: Johns Hopkins University Press.
- Kalode, P. (2012). Computer Assisted Medical Health Systems for the Benefit of Hard to Reach Areas. *International Journal of Computational Research*, 2(2), 254-259.
- Malo, J., Luraschi-Monjagatta, C., Wolk, D. M., Thompson, R., Hage, C. A., & Knox, K. S. (2014). Update on the diagnosis of pulmonary coccidioidomycosis. *Ann Am Thor Soc*, 11(2), 243-253. doi:10.1513/AnnalsATS.201308-286FR PMID:24575994
- Mearnian, L. (2013, March 5). IBM: Watson will eventually fit on a smartphone, diagnose illness. *Computerworld*.
- Naik, G. (2012, Dec. 31). At Europe's Doorstep, Fiercde War Against TB. *The Wall Street Journal*, p. A1.
- Pachner, A. R. (1989). Neurologic manifestations of Lyme disease, the new great imitator. *Reviews of Infectious Diseases*, 11(Suppl 6), S1482-S1486. doi:10.1093/clinids/11.Supplement_6.S1482 PMID:2682960
- Qiang, L., & Nishakawa, R. (2015). *Computer Aided Detection and Diagnosis in Medical Imaging*. New York: CRC Press.

Scadding, J. (1967, October 21). Diagnosis: The clinician and the computer. *Lancet*, 290(7521), 877–881. doi:10.1016/S0140-6736(67)92608-6 PMID:4168324

Stanek, G., Wormser, G. P., Gray, J., & Strle, F. (2012, February). Lyme borreliosis. *Lancet*, 379(9814), 461–473. doi:10.1016/S0140-6736(11)60103-7 PMID:21903253

Zeldenrust, S. (1990). *Ambiguity, Choice and Control in Research* (PhD Dissertation). Amsterdam: Universiteit van Amsterdam.

ADDITIONAL READING

Cole, S. (1992). *Making science*. Cambridge, MA: Harvard University Press.

Freidson, E., & Lorber, J. (Eds.). (2009). *Medical professionals and the organization of knowledge*. New Brunswick, NJ: Aldine Transaction, paper-bound. (Original work published 1972)

Lorber, J. (1984). *Women physicians: Careers, status, and power*. New York: Tavistock Publications.

KEY TERMS AND DEFINITIONS

Computer Assisted Diagnosis (CAD): A conclusion as to the disease reached with the assistance of a computer.

Diagnosis: A conclusion or assumption the pathology represented by a patient's inquiry, particular set of symptoms, or complaints.

Diagnostic Process: The process of arriving at a diagnosis using various informational inputs from a patient's inquiry, symptoms or complaints.

Category C

Cloud Computing

Cloud Computing

Eduardo Correia

Christchurch Polytechnic Institute of Technology (CPIT), New Zealand

INTRODUCTION

In some ways it is easy to see the impact cloud computing has had on the way people interact with technology. Users need not depend on a single device, but can instead access their data from a range of devices. Someone may for instance begin working on a presentation on a PC, continue working on it on a tablet, and decide to do some light editing on a smart phone. When people purchase a new phone these days, they need only to link it to the appropriate provider in order to access their contacts, calendar and other data, whether it be documents, photographs or financial statements, and they can exchange one device for another seamlessly and still be able to do what they wish to do. If an individual device fails, it may be inconvenient, but ultimately it does not destroy data or prevent access to it from another device, as one or more third-party providers hold an up to date copy of the data somewhere. Everyone can see these changes, what is not clear are the underlying enabling technologies that make this all possible and how organizations can best leverage them.

BACKGROUND

A common view of the cloud is that it is nothing new, that people have been using centralized computing in the form of mainframes for instance. While this sheds light on some of the similarities between mainframes and the cloud, it does not reflect the differences between them, differences that are significant. One useful analogy is to compare the cloud and earlier technologies, with the smart phone and the traditional telephone. While the telephone as such is not new and the smart phone

can do the same kinds of things as the traditional phone, it is much more flexible and powerful, and this is down to, in large part, how mobile devices take advantage of the cloud, making them more functional and helping extend battery life (Liu, Chen, Ma, & Xie, 2016). As a term “cloud computing” is problematic because it does not mean one thing, but is “an umbrella term” (Missbach, Stelzel, Gardiner, Anderson & Tempes 2013) that refers not just to a set of technologies but a new model of delivering a wide range of computing resources, services and solutions.

Fortunately, the National Institute of Standards and Technology (NIST) has made a deliberate effort to define cloud computing and done this well, so much so that its three-page description of this term is now considered the standard for the industry. NIST defines cloud computing as “a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction” (Mell & Grance, 2011). This definition is comparable to the four key characteristics of cloud computing as “pay-per-use”, “elastic capacity and the illusion of infinite resources”, a “self-service interface” and “resources that are abstracted or virtualised” (Voorsluys, Broberg, & Buyya, 2011). The cloud is “ubiquitous”, “convenient”, and “on-demand” in the sense that users can access applications and data from anywhere and at any time, and because organizations (or “customers” as the literature often calls them) can, without having to talk to anyone, provision computing resources and services from a “shared pool of configurable computing resources,” a busi-

DOI: 10.4018/978-1-5225-2255-3.ch089

ness model that involves renting multi-tenanted infrastructure.

SERVICE MODELS

Cloud providers, through the use of virtualization, essentially abstract the software, whether it be a server, a database or a network from the underlying physical infrastructure, and offer customers logical software-defined computing resources and services that they can customize, secure, provision and manage in a way that meets their needs and requirements (Buyya, Vecchiola, & Thamarai Selvi, 2013). The service models in cloud computing all reflect this principle of abstraction and represent different sets of responsibility depending how much or how little customers wish to outsource. Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS) therefore represent progressively lower levels of engagement on the part of customers, as they cede more of their responsibilities to the cloud provider.

With IaaS, customers no longer need to purchase physical equipment and associated infrastructure, and manage it, whether it is a server in the case of a small business or a large datacenter in the case of a large business. They can instead utilize the provider's resources presented to them in the form of *virtual* infrastructure such as virtual machines (VMs), networks, storage, firewalls, and load balancers, all on a granular pay-as-you-go basis, so that the customer pays for example for each hour of use of a small VM, which will be cheaper than a more powerful VM. While the benefits outweigh the constraints, customers need to know not just what the provider can but also what it cannot do and work within the limitations of a particular cloud. For instance, in Microsoft Azure once a virtual network has been created it is not possible to add a virtual machine to it. There are workarounds, for instance keeping the disk with the data and connecting them to newly created VMs in the existing virtual network.

In the case of PaaS, customers have no access to or control over not just to the underlying physical but also virtual infrastructure, such as servers, networks, storage and operating systems. PaaS involves the provision of resources and services, such as a development environment, database, and integration and development tools that can be used in a matter of minutes, all provided so as to reduce the effort and cost for SaaS vendors (Brown & Nyarko, 2013). Customers can instead focus on the functionality and logic of applications. It is the responsibility of the provider to ensure that a database for instance can cope with the increasing demands made on it and that the underlying operating system is properly patched and secure. Well-known examples of PaaS are AWS Elastic Beanstalk and Force.com (part of Salesforce.com). In terms of IaaS and PaaS, Amazon Web Services (AWS) is currently the largest public cloud provider and Microsoft the second largest public cloud provider, with 57% and 20% of enterprises making use of them respectively (RightScale, 2016). This is in line with the Gartner magic quadrant (Leong, Toombs, & Gill, 2015).

SaaS is a fully-built software solution that customers can use immediately. It is an application that is owned by the provider, who takes responsibility for all aspects of its development, deployment and operation (Carstensen, Morgenthal, Golden, 2012) as well as for security, performance and availability. The application is accessed through a network connection by the customer, who usually pays a subscription of some kind instead of the traditional upfront license fee, which costs a great deal more because it gives the customer a permanent right to use the software whereas the subscription does not. The application or set of applications is hosted on a remote system and the customer data is stored by the provider but the customer can access the application or set of applications and the data from various devices and systems. Well-known examples of SaaS are Microsoft Office 365 and Salesforce.

DEPLOYMENT MODELS

NIST identifies four cloud deployment models: private cloud, public cloud, hybrid cloud and community cloud (Mell & Grance, 2011). The private cloud is owned and operated by a single organization that may have various customers, such as departments or business units. The organization is completely responsible for managing all aspects of the private cloud: procurement, configuration, maintenance, support and delivery of services. As a result, there are usually few significant cost savings or economies of scale or pooling of resources typical of the public cloud. Some commentators argue that few organizations will implement an on premise cloud environment that satisfies the NIST definition because that is not only expensive but is also technically a highly complex undertaking (Golden, 2015). When most people refer to the “cloud,” they are thinking of the public cloud, a cloud entirely owned and managed by an external third-party provider. In most cases the public cloud involves shared third-party infrastructure, though exceptions can be made for those customers willing to pay a premium for hardware dedicated for their exclusive use.

The hybrid cloud acknowledges that most organizations would wish to continue making use of on premise infrastructure that current IT staff know and understand while being able to take advantage of the public cloud and associated services. While the hybrid cloud is meant to be a private cloud that expands into the public cloud, in reality it may not entail the use of a private cloud at all and may instead simply be a more traditional on premise virtual infrastructure that connects securely with one or more public clouds to form a single hybrid entity. One example of this is the use of compute in the public cloud while retaining data on local on premise systems. Another is the use of SaaS that integrates seamlessly with existing environments. This is now being extended to the connection of not one but multiple clouds as various companies take the view that most organizations are likely

to make use of a number of providers and that these clouds need to be connected seamlessly (Schwartz, 2015).

The community cloud is typically a private cloud in which a number of entities share the cost and benefits, while avoiding some of the concerns around security and regulation associated with the cloud. This works best when the entities share similar security, governance and policy requirements (Bond, 2015). For example, multiple government agencies may make use of a single datacenter, with the result that less of their interactions have to traverse the internet (Winkler, 2011). While this satisfies regulatory and compliance requirements, it is not elastic, not necessarily highly available, and is typically only IaaS, with managed services having to still be built and maintained. The community cloud, though, may also be part of an external, public cloud that has been especially designed and provisioned for that purpose. The Microsoft Office 365 Government Community cloud is an example of this. It satisfies a particular set of requirements: data is stored in the United States, customer data is restricted to screened Microsoft personnel and satisfies all the requirements of US public sector customers (Microsoft, 2016).

SECURITY IN THE CLOUD

When organizations adopt the cloud, much of their network, systems, application and data moves out of their immediate control and becomes the responsibility of the cloud service provider, something that clearly has profound implications for security, compliance and governance (Kale, 2014). The multi-tenant model of the cloud requires customers to trust providers, and not just with the orderly functioning of their operations but often with private data and intellectual property. The responsibility for security in the cloud is shared between customer and provider. With some of the security being the responsibility of the cloud

provider and some of the security being the responsibility of the cloud customer. It is important that the customer understands, this demarcation line otherwise known as the “trust boundary”, since this involves having a good grasp of the scope of management and monitoring responsibilities. Moreover, the trust boundary changes depending on the service delivery model, the service level agreement and the provider’s approach to security (Mather, Kumaraswamy, & Latif, 2009).

With data being handled and stored by third parties, security, regulation and compliance considerations are vital. One widely-used standard is the Payment Card Industry Data Security Standard (PCI DSS), which applies to organizations that handle credit cards and specifies how personally identifiable information (PII) is stored and accessed. Another is the Health Insurance Portability and Accountability Act (HIPAA), which now designates providers as business associates of customers and therefore requires them to also be HIPAA compliant. These and other standards ensure operations are more secure but at some point organizations also have to trust the cloud provider. Security is often cited as a barrier for adopting the public cloud, and certainly in the past many customers have struggled with this (Carstensen, Morgenthal, Golden, 2012). Yet while organizations lose total control of security, the fact is that large cloud providers draw extensively on the best security expertise and practice, so as to ensure that their side of the trust boundary is as secure as it possibly can be, so that IT operations as a whole are more rather than less secure. As Rob Alexander, chief information officer of Capital One put it, “We can operate more securely on AWS than we can in our own data centers” (Krazit, 2015). There is some evidence to suggest that these days security in the cloud is no longer the challenge it once was, with the major issues for organizations being instead a lack of resources and expertise (RightScale, 2016).

FUTURE RESEARCH DIRECTIONS

C

Some people continue to express reservations and concerns that not only does cloud computing raise difficult questions related to security but that reliance on a small group of large cloud providers also threatens the essential openness and innovative nature of the internet (Anderson & Rainie, 2010). Other questions need to be looked at, most particularly how providers can best interface not just with customers’ devices but also with one another, as increasing numbers of customers’ systems and infrastructure utilize multiple providers’ clouds. What is clear, though, is that demand for and use of cloud services is rapidly increasing. Within a short time for instance few organizations will run their own email systems and most email will be stored in and accessed from the cloud. In addition an increasing number and range of products from health monitoring implants to sensors in cars and buildings will rely on cloud computing. One of the critical areas requiring scholarly attention is how government can create a sufficiently robust regulatory environment that will adequately protect customers, even though systems and data in the cloud traverse national boundaries and jurisdictions. These are as much philosophical as they are technical questions but they demand urgent study and investigation.

CONCLUSION

There is, strictly speaking, no such thing as “the cloud”. There is not one but many clouds, with providers aiming to make customers more effective and competitive through the use of a wide range of services, including account management and billing, identity and access management, databases, developer tools, monitoring and a range of applications. Increasingly the cloud is mission critical for customers, whether these are individuals, small and medium-sized businesses

or large corporations. It started mostly with test development environments and its use is expanding to increasingly diverse range of uses. The cloud increases access to resources that scale, but also masks even eliminates for customers much of the technical complexities and drudgery of running on premise infrastructure. Customers no longer need to manage hardware and licensing and have only to pay for the resources they use while being able to operate efficiently through the use of automation (Winkler, 2011). They can also immediately leverage large cloud providers' infrastructure to deploy applications easily and rapidly across multiple geographic regions. In 2014, there were twenty times more active VMs in the public cloud compared to 2011 (Bittman, 2015), making it clear that while the use of on premise infrastructure and private clouds has increased, the greatest growth is from customers seeking to move new workloads to the public cloud.

When organizations adopt the public cloud, not only are they able reduce the complexity and cost of operations, they also gain, through the provider, much greater flexibility, elasticity, and scalability. Storage, compute and networks can be "rapidly provisioned and released" (Mell & Grance 2011) because interaction with the provider is minimal and operations are streamlined to maximize efficiency, including typically the use of standard service terms, service-level agreements (SLAs) and limitation agreements. This increased flexibility does raise challenges in terms of data security and access control because users store data and then share it with others who are not necessarily in the same trusted domain, but then staff managing on premise infrastructure face the same kinds of challenges, often with a great deal less support from third parties. As one provider put it, the reason to move to the public cloud is "to lower costs, reduce complexity, and increase flexibility" (AWS, 2015). These factors will likely continue to drive growth in the public cloud over the next few years and beyond.

REFERENCES

- Anderson, J., & Rainie, L. (2010). The future of cloud computing. Washington, DC: Pew Internet & American Life Project. Retrieved from <http://www.pewinternet.org/2010/06/11/the-future-of-cloud-computing/>
- AWS. (2015). *Introduction to AWS Economics: Reducing costs and complexity*. Amazon Web Services. Retrieved from <https://d0.awsstatic.com/whitepapers/introduction-to-aws-cloud-economics-final.pdf>
- Bittman, T. (2015). *Some perspective on the explosion of VMs in the cloud*. Retrieved from http://blogs.gartner.com/thomas_bittman/2015/03/05/some-perspective-on-the-explosion-of-vm-in-the-cloud
- Bond, J. (2015). *The enterprise cloud: Best practices for transforming legacy IT*. Sebastopol, CA: O'Reilly.
- Brown, C. W., & Nyarko, K. (2013). Software as a Service. In A. M. Bento & A. K. Aggarwal (Eds.), *Cloud Computing Service and Deployment Models: Layers and Management*. Hershey, PA: Business Science Reference. doi:10.4018/978-1-4666-2187-9.ch003
- Buyya, R., Vecchiola, C., & Thamarai Selvi, S. (2013). *Mastering cloud computing: Foundations and applications programming*. Retrieved from <http://www.ebilib.com>
- Carstensen, J., Morgenthal, J. P., & Golden, B. (2012). *Cloud computing: Assessing the risks*. Retrieved from <http://www.ebilib.com>
- Golden, B. (2015). *5 cloud computing predictions for 2016*. Retrieved from <http://www.cio.com/article/3024329/cloud-computing/5-cloud-computing-predictions-for-2016.html>
- Kale, V. (2014). *Guide to cloud computing for business and technology managers: From distributed computing to cloudware applications*. Hoboken, NJ: Taylor and Francis.

Krazit, T. (2015). *Are public clouds really safer than private data centers?*. Retrieved from <http://fortune.com/2015/10/20/structure-2015-public-cloud/>

Leong, L., Toombs, D., & Gill, B. (2015). *Gartner magic quadrant for cloud infrastructure as a service, Worldwide*. Retrieved from <https://www.gartner.com/doc/reprints?id=1-2G2O5FC&ct=150519&st=sb>

Liu, T., Chen, F., Ma, Y., & Xie, Y. (2016). An energy-efficient task scheduling for mobile devices based on cloud assistant. *Future Generation Computer Systems*, 61, 1–2. doi:10.1016/j.future.2016.02.004

Mather, T., Kumaraswamy, S., & Latif, S. (2009). *Cloud security and privacy: An enterprise perspective on risks and compliance*. Beijing: O' Reilly.

Mell, P., & Grance, T. (2011). *The NIST definition of cloud computing*. National Institute of Standards and Technology. Retrieved from <http://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-145.pdf>

Microsoft. (2016). *Office 365 Government Community Cloud*. Retrieved from <https://technet.microsoft.com/en-us/library/office-365-government.aspx>

Missbach, M., Stelzel, J., Gardiner, C., Anderson, G., & Tempes, M. (2013). *SAP on the Cloud*. Heidelberg: Springer. doi:10.1007/978-3-642-31211-3

RightScale. (2016). *State of the cloud report*. Retrieved from <http://assets.rightscale.com/uploads/pdfs/RightScale-2016-State-of-the-Cloud-Report.pdf>

Schwartz, J. (2015). Microsoft's hybrid cloud strategy. *Redmond Magazine*. Retrieved from <https://redmondmag.com/articles/2015/07/01/cloud-alignment.aspx>

Voorsluys, W., Broberg, J., & Buyya, R. (2011). *Introduction to Cloud Computing*. doi:10.1002/9780470940105.ch1

Winkler, V. (2011). *Securing the cloud: Cloud computer security techniques and tactics*. Amsterdam: Elsevier.

ADDITIONAL READING

Barton, B. (2015). *Microsoft public cloud services: Setting up your business in the cloud*. Redmond, WA: Microsoft Press.

Buyya, J. Broberg, & A. Goscinski (Eds), *Cloud computing: Principles and paradigms*. Hoboken, N.J., Wiley and Sons.

Chandrasekaran, *Essentials of cloud computing*. Boca Raton, FL, CRC Press.

Chang, V., Kuo, Y.-H., & Ramachandran, M. (2016). Cloud computing adoption framework: A security framework for business clouds. *Future Generation Computer Systems*, 57, 24+. Retrieved from <http://libproxy.cpit.ac.nz/login?url=http://go.galegroup.com/libproxy.cpit.ac.nz/ps/i.do?id=GALE%7CA444536257&v=2.1&u=polytechnic&it=r&p=AONE&sw=w&asid=10398f3598ba50cb7df4e5964ade8bb8>

Erl, T., Puttini, R., & Mahmood, Z. (2013). *Cloud computing: Concepts, technology and architecture*. Upper Saddle River, NJ: Prentice Hall.

Goyal, S. (2014). Public vs private vs hybrid vs community - cloud computing: A critical review. *International Journal of Computer Network and Information Security*, 3(3), 20–29. Retrieved from <http://www.mecspress.org/ijcnis/ijcnis-v6-n3/IJCNIS-V6-N3-3.pdf> doi:10.5815/ijcnis.2014.03.03

Kavis, M.J. (2014). *Architecting the cloud: Design decisions for cloud computing service models*. Hoboken, NJ: Wiley. doi:10.1002/9781118691779

Limoncelli, T., Chala, S. R., & Hogan, C. (2015). *The practice of cloud system administration: Designing and operating large distributed systems*. Upper Saddle, NJ: Addison-Wesley.

Rountree, D. (2013). *Basics of cloud computing: Understanding the fundamentals of cloud computing in theory and practice*. Saint Louis, MO, USA: Academic Press; Available from <http://www.ebrary.com>

Savill, J. (2015). *Mastering Microsoft Azure infrastructure services*. Indianapolis, IN: John Wiley & Sons.

Srinivasan, S. (2014). *Cloud Computing Basics*. New York: Springer. doi:10.1007/978-1-4614-7699-3

Synergy Research Group. (2016). *AWS remains dominant despite Microsoft and Google growth surges*. Retrieved from <https://www.srgresearch.com/articles/aws-remains-dominant-despite-microsoft-and-google-growth-surges>

Van den Berg, M. (2015). *Microsoft hybrid clouds*. Birmingham: Packt Publishing.

Weinman, J. (2012). *Clouconomics: The Business Value of Cloud Computing*. Available from <http://www.ebilib.com>

Wittig, A., & Wittig, M. *Amazon Web Services in action*. Shelter Island, NY, Manning.

KEY TERMS AND DEFINITIONS

Cloud Deployment Models: The way in which cloud services are provision, for instance the private cloud, the public cloud and the hybrid cloud.

Cloud Service Models: The distribution of responsibilities for technological requirements between customer and cloud provider.

National Institute of Standards and Technologies: A federal agency within the U.S. Department of Commerce that is responsible for measurement science, standards and technology.

Service Level Agreement: A contract that defines the level of service the customer can expect from the cloud provider.

Trust Boundary: The boundary between the responsibilities of the customer and the responsibilities of the cloud provider.

Virtual Machine: A guest machine that is capable of sharing the same physical hardware with another guest machine and runs on a hypervisor.

Virtualization: The division of a computer into multiple execution environments so that one level (e.g. the physical) can be abstracted from another (e.g. the operating system).

Cloud Governance at the Local Communities



Vasileios Yfantis
Ionian University, Greece

INTRODUCTION

The public organization is an organization that operates as both physical and virtual entity because the citizens' service is implemented on both ways. The physical infrastructure includes all the physical offices and processes that the government uses. Consequently, government as a virtual organization should adopt the most characteristics of a virtual organization in terms of infrastructure. According to Lee and other scholars (Lee, 2014) a virtual organization is «a security and collaboration context not exclusively associated with anyone physical organization». In other words a virtual organization may be considered as a different entity than the physical organization. If a virtual public organization is not associated with the physical organization, might confuse the citizens. On the other hand, the virtual form of the public organization can offer services than the physical one is not capable of offering. For instance, a virtual organization can offer the services of the virtual authentication and digital signature on 24/7 level. It is obvious that a virtual public organization transforms the government into a set of electronic services.

Government as a Service is a terminology that describes a governance model where cloud computing (Ostermann, 2002) is an important element in the operation of the public sector. Modern society promotes the service oriented concepts as the fundamental principles for the future of the global economy. One very promising concept being implemented is the Software as a Service (SaaS). The platform that is being used for its implementation is cloud computing.

GOVERNANCE AS A CLOUD COMPUTING SERVICE

In this section we are exploring the concept of community in order to improve Government as a Service.

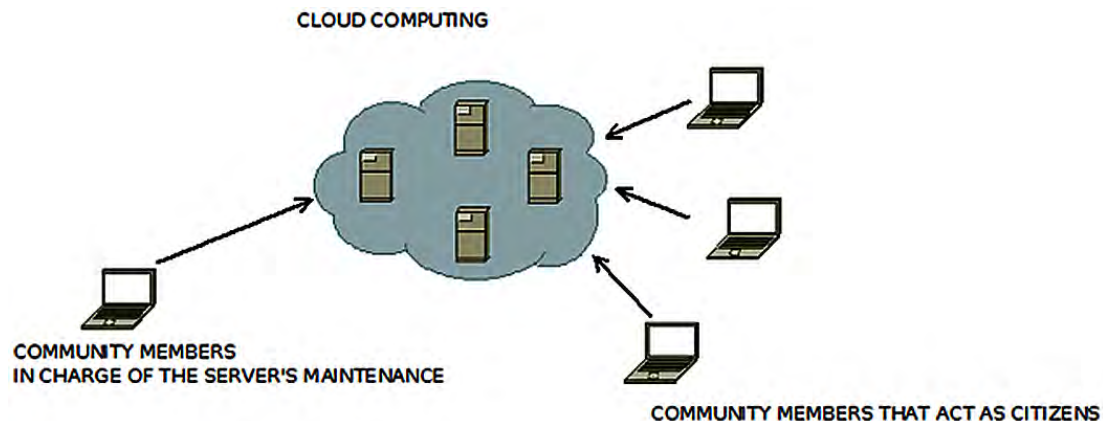
Community cloud computing and cloud computing have the same philosophy; they distribute software, hardware and application resources among users. According to Manion (2004): «People in a country with a high degree of political freedom can participate in the political process more openly and actively, which in turn helps alleviate corruption by empowering people to monitor the exercise of the official power». At community level, it means that if people participate in the electronic governance of the community, then they can help in the minimization of the corruption. This is an important reason in order to explore the actual participation of the autonomous community in the electronic governance. There are two types of community cloud computing that depend on the management of the server side components (Badger, 2011):

1. **On-Site Community Cloud:** The management and the maintenance of the server are being implemented by the members of the community.
2. **Outsourced Community Cloud:** The management and the maintenance of the server outsource to a private company.

These two types of community cloud computing are going to influence governance in various levels regarding its distribution and application.

DOI: 10.4018/978-1-5225-2255-3.ch090

Figure 1. On-site community cloud computing



The most interesting example is the case of self governed communities such as small municipalities or villages. Users could be divided into those who only act as citizens and those who manage the server's tasks, but remain citizens as well (Figure 1). It is important that everyone is obliged to obey the rules of the community.

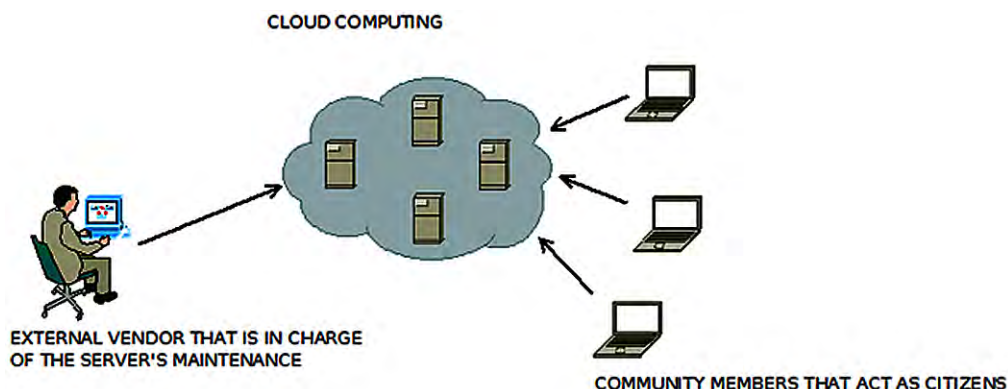
The IT literal members will take over the technical tasks with a number of members assisting to routine technical operations; everyone is able to enjoy the citizen's services. The bond is strong as long as each individual member focuses on the governance mission of the community and acts responsibly. The responsible behavior of the community will trigger the responsible application of the citizen's rights. So, the more respect the members show to the community, the more

respect will show to the public sector. Realizing this fact, governments will try to form a new generation of responsible citizens through training and educational school programs.

The adaptation of the outsourced community cloud model by the citizens offers them greater independence (Figure 2). The server's maintenance will be outsourced to an external company and the community members will focus on the responsible action of the citizen rights and their cooperation with other communities; a business model could be developed in order to allow access to their governance's best practice to other communities under a subscription agreement.

The governments need a serious amount of computing resources and different applications to use. Community cloud allows citizens to share

Figure 2. Outsourced community cloud computing



resources and save money minimizing the use of expensive software and hardware. Last but not least, communities will be able to rent part of its computing resources to other communities.

The fact that each community has a common mission and members with common needs, leads to the adoption of a custom cloud computing system which satisfies their particular computing needs. The community of citizens, just like every other community features these characteristics (Anderson, 2008):

1. **Access:** Every member of the community has to be authorized to access the community's resources.
2. **Communication:** The members of the community have to communicate with each other.
3. **Presence:** The members of the community have to establish relationships between them.
4. **Involvement:** Every member of the community has to participate in the activities of the community.

Gerald Briscoe and Alexandros Marinos note that the features of a community cloud computing include (Briscoe, 2009):

1. **Openness:** There is an open source concept because there are no restrictions with the major cloud computing vendors (Krutz, 2010).
2. **Community:** The technology infrastructures shares common characteristics with the social structure.
3. **Graceful Failures:** The cloud is not owned by a specific company; therefore the risk of failure is minimized. The community cloud is consisted of nodes which function independently, so if a node collapses the failure would be "graceful" because the system will continue its operation, and gradually will recover, almost eliminating the downtime (Birman, 2012).

4. **Convenience and Control:** There is no conflict between convenience and control because they are distributed equally among the community.
5. **Community Currency:** The users of the community are able to define their own currency for the use of the service. This currency could be financial aid or a voluntary technical task (Soder, 2008).
6. **Environmental Sustainability:** The community cloud computing is an environmentally friendly system. Each member of the community uses the exact amount of energy necessary to service its needs, so as to ensure the scalability of the available resources (Wang, 2011).

Adopting the community cloud computing in the citizen's community makes citizen's action a communal service with the following characteristics:

Citizen community's access and community cloud computing: The open architecture of the mobile cloud computing makes the access to the public documents easier by minimizing the access restrictions. The members share common needs and a common citizen mission, so they know what to expect from their participation in this community.

Citizen community's communication and community cloud computing: Communication between the community members is the key function to share information about new laws or measures. The convenience and control over the file sharing that the community cloud computing offers to the community members could trigger the exchange of new public documents as well.

Citizen community's presence and community cloud computing: Trust is one of the most important values in the personal relationships as well as the economic transactions in the human society. In a citizen community the file sharing between the members could lead them to organize social gatherings, conferences and other com-

mon interest political activities. The structure of the community cloud computing allows users to be “friendly” with the other members; which improves the social bonding.

Citizen community’s involvement and community cloud computing: In a citizen community every member has to be active so as to contribute to the quantity and quality of public documents sharing within the community. The graceful failures of the community cloud computing are a sign of warning for the contributing members, in order to participate vigorously and avoid destruction.

The currency of the community cloud computing might be anything that helps the citizen community’s functionality. This means that the citizens will save money from their financial obligations to the government if they volunteer to do technical activities. The community will be divided into teams and each team will undertake a technical task: e.g. Maintenance of the server, back up duties, application testing and much more. This is an ideal advantage of the community cloud that will benefit the participation of the community.

The operation of the cloud is distributed to the community’s resources, so it doesn’t require expensive machinery and the use of excessive power to function. These environmentally friendly techniques could be a drawing power for potential users who are sensitive to such matters.

FUTURE RESEARCH DIRECTIONS

Data sharing is a familiar practice to Web 2.0 users. However those who are looking for the next trend, have already seen the emerge of the Web 3.0 era which is known as semantic web (Allemang, 2011). How Government as a Service will be conformed to the world of Web 3.0?

Semantic web defines simasia (meaningful value) to sites’ content so that these can be reused predictably by modalities to satisfy human needs. A current keyword search using a web search engine will show results that include the keywords but not the simasia of the question of the user. For

instance a question such as “which is the fastest administrative act regarding the family status?” cannot be answered efficiently by the existing web intelligence; humans have to study and evaluate results, exclude choices, further elaborate their search criteria. Semantic web mining engines promise to extract simasia out of the question and compare it against simasia extracted from RDF structured content.

Ontologies are being developed to define meaning so that semantic mining is applicable. The semantic web mining technology results to greater satisfaction among the members of a cloud community that will trigger more requests for the public services. At the same time more satisfaction and less time needed for the issuing of public documents leads to more community members; the more the members the more the computing resources to support the user’s needs. Communities will start lending computing resources to each other and get ready to answer to such question as “lend me for k weeks the appropriate computing resources to efficiently provide the x documents to my members”. Part of the previous question refers to the computing resources that need to be calculated and offered to another community. We envision a semantic mining engine able to assist us in such questions. Perhaps this is a post semantic web evolution.

Semantic web research efforts share the following characteristics (Stumme, 2006):

1. A common syntax for machine language oriented phrases.
2. A common vocabulary.
3. A common logic language.
4. The usage of the language for exchange of proof.

Most researchers nowadays focus on the meaning of the web page’s data while building a search engine, a post semantic web will also is able to calculate the requested computing resources. E.g. A post semantic web engine will not only but be able to answer the question “which is the fastest

administrative act regarding the X process?” but will come up with results to the question “which is the fastest administrative act of the X process and what is the bandwidth needed for the issuing of the document in real time for forty users?” The development of a hybrid semantic engine that will be able to search for a specific public document or combine public data with the computing resources needed for their support, is the key to continue sharing information in an advanced web. The implementation of such a hybrid semantic engine requires a suitable ontology driven language.

Except the technical research direction, there are legal concerns as well about the use of the community cloud computing in the public sector (Winkler, 2011). The fact that the citizen’s data do not remain in a local system, but they are archived in the cloud, it raises a legal concern about who is the authorized user to access and edit this data. From the government’s side, it is important to ensure that will secure the privacy of the data by establishing an appropriate agreement between the government and the cloud computing vendor. The terms and conditions of the agreement should clarify the authority about the control of each party over the citizen’s data and have to ensure a specific service level from the vendor’s side. In terms of security, there should be defined a process about the vendor to notify the government in case of unauthorized access to the data by hackers or requests by foreign governments about access to the data (Snooks, 2012). In confidentiality level, the vendor should be aware of the confidentiality degree of each set of data in order to manage the data in the appropriate way.

CONCLUSION

Cloud governance is a concept that has to be adjusted in the new challenges of the Information Technology. The application of e-government in the autonomous communities through cloud

computing is a way to reduce the corruption and empower the bonds between the members of the community. On the other hand, the communities have to use the advantages of the cloud computing in an ethical way. The recent findings that ISIS was using a network of playstation devices for the communication between the terrorists, is a sad fact. I wonder if a Community cloud computing network would be a strong weapon for the terrorists. The question here should be how secure is a community cloud network and what is the mission of this type of network?

However, I believe that the issues of unauthorized access to secret information and the website hacking actions will be reduced if the Government as a Service model will be developed from the citizen’s side. As long as there is limited political motivation, the citizens will probably develop a more responsible and respectful behavior towards the public policy. The proposed approach increases sharing and participation within the users’ communities and guides them to respect governments and their politics.

REFERENCES

- Allemang, D., & Hendler, J. (2011). *Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OW*. Waltham, MA: Elsevier.
- Anderson, P. (2008). *Our community: changing attitudes, empowering communities*. Retrieved November 13, 2013, from <http://www.psawa.com/Deinstitutionalisation.html>
- Badger, T., Grance, T., Patt-Corner, R., & Voas, J. (2011). *Draft Cloud Computing Synopsis and Recommendations*. National Institute of Standards and Technology, U.S. Department of Commerce.
- Birman, K. (2012). *Guide to Reliable Distributed Systems: Building High-Assurance Applications and Cloud-Hosted Services*. Ithaca, NY: Springer. doi:10.1007/978-1-4471-2416-0

Briscoe, G., & Marinos, A. (2009). Lecture Notes in Computer Science: Vol. 5931. *Community cloud computing* (pp. 472–484). Berlin: Springer. doi:10.1007/978-3-642-10665-1_43

Krutz, R., & Vines, R. (2010). *Cloud Security: A Comprehensive Guide to Secure Cloud Computing*. Indianapolis, IN: Wiley Publishing, Inc.

Lee, C., Desai, N., & Brethorst, A. (2014) A Keystone-Based Virtual Organization Management System. *Cloud Computing Technology and Science (CloudCom), 2014 IEEE 6th International Conference* (pp 727-730). Singapore: IEEE.

Manion, M. (2004). *Corruption by Design: Building Clean Government in Mainland China and Hong Kong*. Cambridge, MA: Harvard University Press.

Ostermann, S., Iosup, A., Yigitbasi, N., Prodan, R., & Fahringer, T. (2009). A Performance Analysis of EC2 Cloud Computing Services for Scientific Computing. *Cloud Computing*, 4, 115–131.

Snooks, A. (2012). *Negotiating the cloud – legal issues in cloud computing agreements*. Retrieved March 6, 2014, from <http://www.finance.gov.au/cloud/>

Soder, N. (2008). *Community currency: an approach to economic sustainability in our local bioregion*. Berlin: Humboldt University.

Stumme, G., Hotho, A., & Berendt, B. (2006). Semantic Web Mining, State of the Art and Future Directions. *Journal of Web Semantics*, 4(2), 124–143. doi:10.1016/j.websem.2006.02.001

Wang, L., Ranjan, R., Chen, J., & Benatallah, B. (2011). *Cloud Computing: Methodology, Systems, and Applications*. New York: CRC Press.

Winkler, V. (2011). *Securing the Cloud: Cloud Computer Security Techniques and Tactics*. Waltham, MA: Elsevier.

ADDITIONAL READING

Agrawal, D., Candan, S., & Wen-Syan, L. (2011). *New Frontiers in Information and Software as Services: Service and Application Design Challenges in the Cloud*. New York: Springer. doi:10.1007/978-3-642-19294-4

Anttiroiko, A., & Mälkiä, M. (2007). *Encyclopedia of Digital Government*. Hershey, PA: IGI Global. doi:10.4018/978-1-59140-789-8

Box, R. (1998). *Citizen Governance: Leading American Communities Into the 21st Century*. Thousand Oaks: SAGE. doi:10.4135/9781483328249

Brown, W., Laird, R., Gee, C., & Mitra, T. (2009). *Soa Governance: Achieving And Sustaining Business And It Agility*. London: Pearson Education.

Cardoso, J., & Sheth, A. (2010). *Semantic Web Services, Processes and Applications*. New York: Springer.

Chaston, I. (2012). *Public Sector Reformation: Values-driven Solutions to Fiscal Constraint*. Hampshire: Palgrave Macmillan. doi:10.1057/9780230379350

Fox, A., & Patterson, D. (2013). *Engineering Software As a Service: An Agile Approach Using Cloud Computing*. San Francisco: Strawberry Canyon LLC.

Gallagher, S. (2013). *VMware Private Cloud Computing with vCloud Director*. Hoboken, NJ: John Wiley & Sons.

Gil-Garcia, J. (2012). *Enacting Electronic Government Success: An Integrative Study of Government-wide Websites, Organizational Capabilities, and Institutions*. New York: Springer. doi:10.1007/978-1-4614-2015-6

Great Britain, Parliament, House of Commons, Public Administration Select Committee. (2011) Government and IT - “a recipe for rip-offs”: time for a new approach, twelfth report of session 2010-11, Vol. 2: Oral and written evidence, London: The Stationery Office.

Greer, M. (2009). *Software As a Service Inflection Point: Using Cloud Computing to Achieve Business Agility*, Bloomington, Indiana: iUniverse.

Hatch, R. (2008). *SaaS Architecture, Adoption and Monetization of SaaS Projects Using Best Practice Service Strategy, Service Design, Service Transition, Service Operation and Continual Service Improvement Processes*, Lulu.com.

Kashyap, V., Bussler, C., & Moran, M. (2008). *Semantic Web: Semantics for Data and Services on the Web*. New York: Springer.

Katzan, H. (2011). The Education Value Of Cloud Computing. In *Contemporary. Issues in Educational Research*, 3(7), 37–42.

Lane, J. (2000). *The Public Sector: Concepts, Models and Approaches*. Thousand Oaks: SAGE.

McDonald, K. (2010). *Above the Clouds: Managing Risk in the World of Cloud Computing*, Ely: IT Governance Ltd.

Mergel, I. (2012). *Social Media in the Public Sector: A Guide to Participation, Collaboration and Transparency in The Networked World*. Hoboken, NJ: John Wiley & Sons.

Nordic Council of Ministers. (2012). *Nordic Public Sector Cloud Computing - a Discussion Paper*, Copenhagen: Norden.

Reddick, C. (2010). *Politics, Democracy, and E-government: Participation and Service Delivery*. Hershey, PA: IGI Global. doi:10.4018/978-1-61520-933-0

Roebuck, K. (2011). *SaaS - Software As a Service: High-Impact Emerging Technology - What You Need to Know: Definitions, Adoptions, Impact, Benefits, Maturity, Vendors*. Newstead, QLD: Emereo Pty Limited.

Scholes, K., & Johnson, G. (2001). *Exploring Public Sector Strategy*. London: Pearson Education.

Vitvar, T., Peristeras, V., & Tarabanis, K. (2010). *Semantic Technologies for E-Government*. New York: Springer. doi:10.1007/978-3-642-03507-4

Weerakkody, V., & Reddick, C. (2012). *Public Sector Transformation Through E-government: Experiences from Europe and North America*. London: Routledge. doi:10.4018/978-1-4666-1776-6

Williams, A., & Robinson, A. (2012). *Public Cloud Computing: Security and Privacy Guidelines*. New York: Nova Science Publishers, Incorporated.

Williams, B. (2012). *The Economics of Cloud Computing: An Overview For Decision Makers*. San Jose: Cisco Press.

KEY TERMS AND DEFINITIONS

Cloud Computing: A platform for delivering electronic services from any place and in any time.

Community Cloud: A community operated platform for delivering electronic services from any place and in any time.

Government as a Service: Virtual governance that service platform in which both software and data are located on the cloud.

Online Community: A community that operates through a public or private electronic network.

Semantic Web: A conversion of the World Wide Web that stores information in webpages which is readable from other computers and will be used for the understanding of the webpage's content.

Virtual Organization: An organization that operates through a public or private electronic network.

Web 2.0: The inclusion of social elements on the World Wide Web.

Clouds of Quantum Machines

Nilo Sylvio Serpa

Universidade Paulista, Brazil

INTRODUCTION

This work is a continuation of studies started at 2014. Since then, I improved some ideas about theoretical quantum machines interacting in cloud operation, as well as enhancements on the concept of quantum entanglement itself. Some recent works have been added to the original references, although the classic treatises remain in effect, given the slowness with which the subject progresses.

As we know, one of the great challenges in quantum computing is how to preserve quantum entanglement, since microphysical systems are extremely sensitive to external disturbances. It is one of my aims to show a way to minimize this problem. Moreover, a quantum orchestration metalanguage is introduced into a first schematic representation of the operation of a quantum cloud aiming to optimize further constructions of quantum algorithms. While it is essential that the reader becomes familiar with quantum mechanics, some parts of my first work are reproduced here in order to reduce the effort to understand the subject.

BACKGROUND: SERVICES AND CLOUDS IN A CONTEMPORARY APPROACH

Services are cybernetic replicas of human practices, being evoked by well-established environmental motivations. In turn, SOA is an architecture that integrates in a standard manner several service units, each of them sending their features as sets of tasks over the network. Only service interfaces are exposed to consumers as exported methods (Nakamura *et al.*, 2004). Therefore, when services

are requested, SOA seeks the best responses to those environmental motivations according to the internal logic of each service. In particular, this architecture is now strongly linked to the theme of “enterprise application integration” (EAI) in contexts where legacy applications already established are performed on different platforms.

The literature on SOA comprises several milestone contributions as the works of Nakamura *et al.* (2004), Erl (2005), Anderson & Ciruli (2006), Natis (2007), Sha (2007) and, markedly, Frenken *et al.* (2008) about device-level service deployment. On this latter subject, it is noteworthy that, in the process of architectural development, devices which access legacy applications are created and interact using a protocol defined by the system. In turn, the system returns the aggregated information from the various legacy applications, preferably without any additional code. The architectural development also takes care of the service interface, prescribing the information required to access the competences of that. It is worth remembering that the existence of interfaces and descriptions of accessibility is *sine qua non* for the implementation of SOA. More recent works show the state-of-art in services orchestration (MEF Forum, 2015; Lemos *et al.*, 2015).

In SOA projects, the so-called Enterprise Service Bus (ESB) is thought to be the main component of the infrastructure layer. It is the mediator between provider and service consumers, and its responsibility is to provide integration and interoperability between different systems. Embedded in this responsibility is also the mission of cleaning the databases by a service that tracks and recognizes all of the systems which shall be linked. Connectors are created in the databases feeding a new datawarehouse completely normal-

DOI: 10.4018/978-1-5225-2255-3.ch091

ized, such that any updates made on the original basis are automatically computed and reflected in the standardized repository.

As a logical consequence of the advent of the Internet and the concept of SOA, we can say that cloud computing is a cybernetic implementation by which all IT resources (hardware, software, networking, storage, etc.) are provided as services on-demand to consumers via Internet, remaining managed to ensure fast delivery, high availability, security and quality. In short, cloud computing is a model of computation by which those IT resources are randomly dispersed in the network, being offered as services paid as they are consumed. Although this subject promotes a lot of controversy about information security, everything suggests that the process of agglomeration of servers in clouds is irreversible.

Cloud computing and SOA have contributed significantly with one another and should remain so for a long time. In the words of David Linthicum, “SOA can be used as a key technology-enabling approach to leverage cloud computing” (Linthicum, 2009). Thus, the use of SOA can be galvanized by the cloud structure, since it allows on-demand delivery beyond the limitations imposed by the firewall constraints of the enterprise environment. A cloud computing system, whether formed by quantum machines and evolving to the point of hosting thousands or millions of servers, leads to a probabilistic approach of the states of each component as well as the states of information. As in quantum mechanics, where the physical position of a particle is random-dispersed until the intervention of an observer, the information is “nowhere”, and, at the same time, “everywhere” until an objective request is done. This probabilistic approach of the cyber states of the cloud and their components is consistent with a vector representation in Hilbert space, just as the latter is used in quantum mechanics. With respect to the theoretical foundations and not on the scale of physical phenomena, the above conceptual and formal analogy improves the ways to understand

the behavior of a “quantum cloud”, and, reversely, our understanding of quantum physics and its paradigmatic importance to modern informatics, a fact that certainly contributes to future advances in the field of cybernetics. In the next sections I will explain the representative basis to build quantum clouds, discussing the main outcomes expected from such a technology.

PROBABILITY AND SYMMETRY IN QUANTUM MECHANICS

The concept of physical probability was really born with the adventure of quantum mechanics, even though in the core of this discipline it has been treated systematically as the expression of the inexact knowledge.

The focal point was to interpret the so called wave function ψ — the amplitude of the wave itself — and Max Born was the prime to achieve it. As we know, ψ is solution of the famous wave equation for one particle with mass m , due to Schrödinger

$$-\frac{\hbar^2}{2m}\nabla^2\psi + V\psi = i\hbar\frac{\partial\psi}{\partial t}, \quad (1)$$

where V is the potential and $\hbar$ is the Planck constant divided by 2π . The functions ψ are in general complex. The connection of such quantities to the “real” world (or, which came to be the same, the acquirement of quantities called “observables”) is represented by means of operations such as $\psi_i^*\psi_i$, where ψ_i^* is the complex conjugate of ψ_i . It prays a fundamental postulate of quantum mechanics that $|\psi(r,t)|^2 = \psi_i^*\psi_i$ is the density of the probability $P(r,t)$ for a particle of mass m to be found at the point r , on time t . Therefore, the likelihood to locate the particle inside an infinitesimal volumn of space τ on that time t is

$$|\psi(r, t)|^2 d\tau. \quad (2)$$

Only in the case of one particle, the configuration space of the function ψ is isomorphic to the tridimensional space of positions. For two particles, for example, the wave function of the system ($\psi(r_1, r_2)$) is defined in a configuration space of six dimensions.

Since the summation of the probabilities referring to events that are mutually exclusive is 1, it follows

$$\int P(r, t) d\tau = \int |\psi(r, t)|^2 d\tau = 1. \quad (3)$$

Once ψ is not an observable quantity, there is a certain freedom of choice of its form. Besides, the solutions of linear equations, like Schrödinger's equation, may be multiplied by complex numbers, remaining solutions, so that expression (3) turns possible to choose a correct amplitude factor. The point of view of the physical interpretation sustains that the probability $P(r, t)$ is in fact the reflection of an objective property of the "particle", which is that the possible eigenvalues coexist as propensions in a reference class until a macroscopic intervention (a measurement) takes place. Such intervention changes drastically the original reference class. Let us take a system with states $|\psi\rangle$ and $|\psi'\rangle$ respectively before and after the experimental intervention. It's clear that the function ψ is somewhat conjectural here, but, for all theoretical purposes, is ever possible to think this function as a set of states reducible to a unique state (the reduction of the "wave packet"). We must consider the set $|\psi\rangle$ while not specified any function ψ by the apparatus of measurement, in such manner that we have two distinct instances of the reality, one before and other after the observation.

Quantum measurements are represented by a collection $\{\hat{\phi}_\lambda\}$ of operators that act upon the

phase space of the system under observation. The subindex λ labels the possible results of measurement. Let us suppose that the system is in the initial state $|\cdot\rangle$. The probability for a certain state λ after the measurement is,

$$P(\lambda) = \langle \cdot | \hat{\phi}_\lambda^\dagger \hat{\phi}_\lambda | \cdot \rangle, \quad (4)$$

where $\hat{\phi}_\lambda^\dagger$ is the conjugate transposed of $\hat{\phi}_\lambda$,

$$\langle \cdot | \cdot \rangle = 1$$

and

$$\hat{\phi}_\lambda^\dagger \hat{\phi}_\lambda = 1.$$

This is what mathematically signifies to be a normalized state in quantum mechanics. Now, let $|\mu_1\rangle, |\mu_2\rangle, |\mu_3\rangle, \dots, |\mu_q\rangle$ be an orthonormal base. Thus,

$$\hat{\phi}_\lambda = |\mu\rangle \langle \mu| \quad (5)$$

is a quantum measurement. The intervention of the apparatus modifies the state of the system to

$$\frac{\hat{\phi}_\lambda |\cdot\rangle}{(\langle \cdot | \hat{\phi}_\lambda^\dagger \hat{\phi}_\lambda | \cdot \rangle)^{1/2}} = |\cdot'\rangle; \quad (6)$$

$$\frac{|\mu\rangle \langle \mu | \cdot \rangle}{(\langle \cdot | \hat{\phi}_\lambda^\dagger \hat{\phi}_\lambda | \cdot \rangle)^{1/2}} = |\cdot'\rangle; \quad (7)$$

$$\frac{|\mu\rangle \langle \mu | \cdot \rangle}{(\langle \cdot | \langle \mu | \mu \rangle | \mu \rangle \langle \mu | \cdot \rangle)^{1/2}} = |\cdot'\rangle; \quad (8)$$

$$\frac{|\mu\rangle\langle\mu|\cdots\rangle}{(\langle\cdots|\mu\rangle\langle\mu|\cdots\rangle)^{1/2}} = |\cdots\rangle; \quad (9)$$

$$\frac{|\mu\rangle\langle\mu|\cdots\rangle}{|\langle\mu|\cdots\rangle|} = |\cdots\rangle. \quad (10)$$

THE VECTORIAL BACKBONE OF QUANTUM MECHANICS

The vector space of quantum mechanics is a Hilbert space, that is, an orthonormal vector space in which

1. The vector components are complex scalars;
2. The scalar product satisfies $\langle\psi|\psi\rangle > 0$ for $|\psi\rangle \neq 0$, otherwise $\langle\psi|\psi\rangle = 0$;
3. If a and b are complex scalars, then

$$\langle\chi|a\psi_1 + b\psi_2\rangle = a\langle\chi|\psi_1\rangle + b\langle\chi|\psi_2\rangle;$$

4. The space is complete in the norm $\|\psi\| = \sqrt{\langle\psi|\psi\rangle}$.

Finally, the implementation of symmetries in generalized quantum mechanical coordinates¹ may be represented by a unitary operator in the Hilbert space, so that,

$$\mathcal{U}^\dagger \mathcal{U} = 1, \quad [H, \mathcal{U}] = 0;$$

for the groundstate of the Hamiltonian

$$H = -\frac{\hbar^2}{2m} \nabla^2 + V,$$

$$\mathcal{U}|\psi_0\rangle = |\psi_0\rangle; \text{ (this is not so obvious!)}$$

$$H|\psi_0\rangle = E_0|\psi_0\rangle;$$

$$\mathcal{U}H|\psi_0\rangle = E_0(\mathcal{U}|\psi_0\rangle).$$

In fact, accordingly von Neumann theorem, a coordinate transformation that corresponds to a symmetry of the Hamiltonian let invariant the canonical commutation relations of the system and (here is the power of the theorem) may always be implemented by an unitary manner in the Hilbert space of the states. So,

$$\hat{q}'_i = \mathcal{U}(S)\hat{q}_i\mathcal{U}^\dagger(S) = S_{ij}\hat{q}_j;$$

$$\mathcal{U}(S) = e^{i\omega\hat{O}},$$

where $\hat{O}$ is an operator that defines a motion constant (thereby furnishing good quantum numbers for the states of the system) so that $\hat{O}^\dagger = \hat{O}$, and ω is the set of parameters defining the matrix S .

Of course, as an effect of the macroscopic intervention, μ shows some classic traces inherited from the apparatus. But quantum mechanics says nothing about the world out of the experiment. Also, it is important to clarify that it is not always possible to carry out a complete and decisive experiment in this area. For instance, with respect to gravity, an approach by quantum field theory would need 1) an understandable model of gravitation accordingly some quantization algorithm applied to general relativity, which seems little bearing, and 2) an experimental frame able to reproduce the physical conditions under which the hypothetical quantum nature of gravity may come about, such as in a black hole singularity. In fact, one reason to brush aside an experimental program in this way is the difficulty of formulating quantum theory in a cosmological context in which the observers must be part of the system. Although it appears out of the blue, we may suppose there is a real messenger of gravity and imagine a “metaframe” to render gravitation in a familiar figurative language with no *a priori* concerns whether the messenger and its supersym-

metric partner follow Bose or Fermi statistics beneath lab apparatus. This was my proposal: a supersymmetric meta-field theory on gravity (Serpa, 2015). So, I define meta-field theory as a theory that introduces a supersymmetric metaframe to describe fields as sets of particular transformations between two types of entities, the supersymmetric partners in focus.

As in the supersymmetric meta-field theory, it is possible to build a similar metaframe to describe cloud computing in its continuous process of increasing complexity. I will try, so much as possible, to refine the presentation of the formalism in order to avoid time lost with unclear notations and conventions.

THE QUANTUM BIT

The quantum bit, or qubit, is the quantum tile of information and differs from the classical bit by the fact that it is generally given in a superposition of two basic states, e. g., $|0\rangle$ and $|1\rangle$, so that the Dirac ket of the time-dependent state-function of a qubit is denoted by

$$|\Psi(t)\rangle = c_0(t)|0\rangle + c_1(t)|1\rangle,$$

where $c_0(t)$ and $c_1(t)$ are complex time functions. The binary assigned to the basic states are associated with discrete values assumed by physical degrees of freedom of elementary particles, such as the spin. The qubit state fulfills equation (1) in such manner that the Hamiltonian operator takes the form

$$H(t) = h_{00}(t)|0\rangle\langle 0| + h_{01}(t)|0\rangle\langle 1| + h_{10}(t)|1\rangle\langle 0| + h_{11}(t)|1\rangle\langle 1|.$$

Sometimes it is useful to rewrite Schrödinger's equation in matrix formalism as

$$i\hbar \begin{pmatrix} \dot{c}_0(t) \\ \dot{c}_1(t) \end{pmatrix} = \begin{pmatrix} h_{00}(t) & h_{01}(t) \\ h_{10}(t) & h_{11}(t) \end{pmatrix} \begin{pmatrix} c_0(t) \\ c_1(t) \end{pmatrix}.$$

Last, normalization condition applies

$$|c_0|^2 + |c_1|^2 = 1.$$

QUANTUM TELEPORTATION

Quantum teleportation is a very different conception of their science fiction counterparts. Since the nineties authors have studied the subject in theoretical and experimental approaches (Deutsch & Jozsa, 1992; Braunstein, 1996; Bouwmeester, 1997; Zhang *et al.*, 2002; Bowen, 2003).

The concept of quantum teleportation and the so-called quantum entanglement form the basis of cloud computing as conceived here. The latter is one of the biggest sources of confusion in science, since quantum entanglement became a paradox in quantum theory because of its conflict with causality. Two quantum objects are said entangled if they are linked in such manner that their behaviors are bonded never minding how much distant they are from one another. I will try to explain the central idea in the cloud context with a maximum of formal consistence on the previous sections.

We start with two state functions to be entangled at server A , $|\Psi_{\uparrow}\rangle_A$ and $|\Psi_{\downarrow}\rangle_A$. The entanglement is given by the instruction 0 (I_0) of the experiment

$$I_0 : |\Psi_{\uparrow}\rangle_A \# |\Psi_{\downarrow}\rangle_A \Rightarrow |\Psi_{(\uparrow\downarrow)(\uparrow\downarrow)}\rangle_A. \quad (11)$$

The two state functions are now non-causally correlated. After the entanglement, we apply by instruction 1 (I_1) a classical procedure P_1 to carry the second entangled state function on server B , that is,

$$I_1 : P_1 \left| \Psi_{(\uparrow\downarrow)(\uparrow\downarrow)} \right\rangle_A \stackrel{\circ}{=} \left| \Psi_{(\uparrow\downarrow)(\uparrow\downarrow)} \right\rangle_A (-) \left| \Psi_{(\uparrow\downarrow)} \right\rangle_A \equiv \left| \Psi_{(\uparrow\downarrow)} \right\rangle_B.$$

Now we perform a measurement $|M\rangle$ in server A on the combined state which we gain putting $\left| \Psi_{(\uparrow\downarrow)} \right\rangle_A$ in contact with the unknown state function $|\chi\rangle$. Having done this interaction, server A transmits to server B , through a classical channel reachable by procedure P_2 , a complete description of the results of the quantum measurement on $|\chi\rangle \left| \Psi_{(\uparrow\downarrow)} \right\rangle_A$ in order to enable server B to perform certain linear transformation $|\delta\rangle$ on $\left| \Psi_{(\uparrow\downarrow)} \right\rangle_B$; in fact, the measurement described annihilates $|\chi\rangle$, but the linear transformation $|\delta\rangle$ rebuilds the latter at server B from $\left| \Psi_{(\uparrow\downarrow)} \right\rangle_B$, so that by instruction 2 (I_2)

$$I_2 : P_2 |M\rangle |\chi\rangle \left| \Psi_{(\uparrow\downarrow)} \right\rangle_A \stackrel{\circ}{=} |\delta\rangle \left| \Psi_{(\uparrow\downarrow)} \right\rangle_B \equiv |\chi\rangle.$$

This is what we call teleportation of the state $|\chi\rangle$ from server A to server B . Quantum teleportation refers to the “blind” teleport of the state of a quantum system about which there was no information. The measurement does not provide any information from the state function $|\chi\rangle$. All of the quantum state information is passed by the non-causal link between the entangled states $\left| \Psi_{(\uparrow\downarrow)} \right\rangle_A$ and $\left| \Psi_{(\uparrow\downarrow)} \right\rangle_B$. The main consequence of the process is the annihilation of the initial quantum state at server A rebuilt at server B . It must be understood that it is the quantum states which are destroyed and recreated in the teleportation process, and not material components. Thereby, cloning is an impossible operation in quantum physics; we simply can generate an almost-perfect

replica of the original destroyed after teleportation. It is also important to remember that quantum information within a state function is available only as probabilities or, as we commonly say, expectation values.

ENTANGLEMENT: THE PICTURES ON THE POOL

It was pointed that quantum processing was born from “purely philosophically motivated questions” (Walther, 2006) on non-locality and completeness of quantum mechanics fomented mainly by Einstein from his collaborative work with Podolsky and Rosen in 1935. In fact, as once observed, it was Einstein whom restored in modern science the Cartesian metaphysical sense of philosophy, turning physics into a real theory of knowledge (Charon, 1967). This important note remembers to us that philosophy will always be present in the process of creation. It is precisely its absence that determines little creativity that prevails today in all fields. Thus, to understand what is entanglement it will be necessary a reflective process of reconstruction of the conceptual foundations of physics, which will lead to a comprehensive review of the applicability of the notion of causality.

The main controversies of quantum mechanics ever resided in the difficulty of the human mind to separate the physical fact from its perception or representation. Indeed we always work with our perceptions; we took from them the full potential of human development and survival offered, creating representations for all we observe. There was a time when I was a follower of a kind of fruitless and paralyzing materialism that insisted to reify the world. Later, influenced by some physicists adepts of the operationalism, I came also to sympathize with the dresser and foolish idea that the only thing that matters is the calculation and not the ultimate nature of things. Thanks to my growing interest in quantum computing, I could deepen those controversial discussions and reach my own conclusions about them. Of course, long before the

seventies there were eloquent speeches from the great thinkers of modern physics. Weizsäcker, for instance, in the Spanish version of 1974: “*El átomo no es inmediatamente perceptible para nuestros sentidos, y cualquier experimento lleva sólo una determinada propiedad del átomo al ámbito de una perceptibilidad mediata*”² (Weizsäcker, 1974). But that was still little; not just to observe a predicate and describe it by means of classical concepts. It was necessary a phenomenal texture made by the experimental apparatus from which one could then extract useful measurements (information). In this it would lie a deepening of the famous complementarity of Bohr: the ultimate hidden object and its accessible and inseparable image.

Inspired by those philosophical texts from the first half of the twentieth century and early second half, I could refine my ideas and reach an understanding which I consider acceptable, although limited by the nature of human thought. Now I believe that the understanding of the quantum entanglement, one of the most intriguing phenomena of the quantum world, rises, for happiness of the philosophers, in a reflection on the edge of a pool. One summer night, I sat in a chair right in front of a lighted lamp whose flickering light was reflected in the pool. The image of the lamp stretched like a rubber with the ripples of the water and sometimes came to double or even to quintuple depending on the swings of the water. Both, the lamp and its images in water, are real, belonging to the world of mater and perceptions. But imagine that we could not see the lamp, only their images reflected in the water. We would think that two objects born of a unique (duplicate picture) would be irrevocably united, although separated; any change in one of them would “cause” an instantaneous change in the other. With respect to the quantum world is passing up something similar. We have no direct access to the ultimate reality (as the hidden lamp), only to the images produced by our experiments. What we see are the “pictures in the pool” and these are as real as the object that produced them. Clearly, these images carry information from the ultimate object, which makes them tractable to

control. Instead of using the ultimate object we use them with all their informational potential. This potential is the base of the teleport process, since we teleport physical states, not matter in itself. In short, the quantum world is so light and sensible to our presence that it would be impossible to get direct benefits from their objects. All we can do is work with “pools”. As Weizsäcker said: “*Todo experimento es un acto material que es simultáneamente um acto de percepción*”³ (Weizsäcker, 1974).

To get a slight idea of quantum entanglement we consider the process of creation and destruction of a pair of quantum bits called “pair of Einstein, Podolsky and Rosen” (EPR pair). So, let us suppose a quantum bit in a zero-state,

$$|\Psi_1\rangle = 1|0\rangle + 0|1\rangle = |0\rangle.$$

Now, let us take the Hadamard matrix H_2

$$H_2 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

and

$$U_2 = \frac{1}{\sqrt{2}} H_2.$$

We make

$$\begin{aligned} |\Psi'_1\rangle &= U_2 |\Psi_1\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \\ &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{1}{\sqrt{2}} |0\rangle + \frac{1}{\sqrt{2}} |1\rangle. \end{aligned}$$

Also we take another quantum bit in zero-state

$$|\Psi_2\rangle = 1|0\rangle + 0|1\rangle = |0\rangle.$$

Performing a tensor product between $|\Psi'_1\rangle$ and $|\Psi_2\rangle$ we gain

$$|\Psi'_1\rangle \otimes |\Psi_2\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} = \frac{1}{\sqrt{2}} |00\rangle + 0 |01\rangle + \frac{1}{\sqrt{2}} |10\rangle + 0 |11\rangle.$$

It is convenient to define certain unitary transformation, called “control NOT-gate” (CNot). Using Pauli matrices

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

we may write

$$CNot = \frac{1 + \sigma_z}{2} \otimes 1 + \frac{1 - \sigma_z}{2} \otimes \sigma_x =$$

$$= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \text{ so that}$$

$$CNot |\Psi'_1 \Psi_2\rangle = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} =$$

$$= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} = \frac{1}{\sqrt{2}} |00\rangle + \frac{1}{\sqrt{2}} |11\rangle. \quad (12)$$

The point is that for entangled states, as expressed in the above result, decomposition does not hold, that is,

$$\left\{ \#(\varphi_1, \varphi_2) / \varphi_1 \otimes \varphi_2 = \frac{1}{\sqrt{2}} |00\rangle + \frac{1}{\sqrt{2}} |11\rangle \right\}.$$

The very strangeness of entanglement may be explained taking two distinct moments of experimental intervention. Just prior the *CNot* transformation we perform a measurement to obtain the probability of $|0\rangle$. The measurement operator is

$$M_0 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Before the *CNot* transformation the system is in state

$$|\Psi'_1 \Psi_2\rangle = \frac{1}{\sqrt{2}} |00\rangle + 0 |01\rangle + \frac{1}{\sqrt{2}} |10\rangle + 0 |11\rangle.$$

Therefore,

$$\begin{aligned} p(0) &= \langle \Psi'_1 \Psi_2 | M_0^\dagger M_0 | \Psi'_1 \Psi_2 \rangle = \\ &= \langle \Psi'_1 \Psi_2 | M_0 | \Psi'_1 \Psi_2 \rangle = \end{aligned}$$

$$= \frac{1}{\sqrt{2}} (1, 0, 1, 0) \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} =$$

$$= \left(\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}, 0 \right) \begin{pmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix} = 1.$$

After measurement we get for the state of the system

$$\frac{M_0 |\Psi_1 \Psi_2\rangle}{\sqrt{\langle \Psi_1 \Psi_2 | M_0^\dagger M_0 | \Psi_1 \Psi_2 \rangle}} = \frac{\begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \\ 0 \end{pmatrix}}{1} = |\Psi_1 \Psi_2\rangle. \quad (13)$$

We see that measurement had no influence on the first quantum bit that remains in a superposition of $|0\rangle$ and $|1\rangle$. This is not the case when we perform the same measurement just after the *CNot* application. Now we start from

$$\Psi_3 = \frac{1}{\sqrt{2}} |00\rangle + \frac{1}{\sqrt{2}} |11\rangle.$$

Let us compute the probability of $|0\rangle$ by means of

$$p(0) = \langle \Psi_3 | M_0^\dagger M_0 | \Psi_3 \rangle = \langle \Psi_3 | M_0 | \Psi_3 \rangle =$$

$$= \frac{1}{\sqrt{2}} (1, 0, 0, 1) \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} = \frac{1}{2}.$$

Thereby, probability of $|0\rangle$ was changed to $1/2$. After measurement, the state vector of the system took the form

$$\frac{M_0 |\Psi_3\rangle}{\sqrt{\langle \Psi_3 | M_0^\dagger M_0 | \Psi_3 \rangle}} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \times$$

$$\times \begin{pmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ 0 \\ \frac{1}{\sqrt{2}} \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} = |00\rangle.$$

This is somewhat astonishing. Measuring one quantum bit we modify the probabilities of the other quantum bits of the system. However, as strange as the phenomenon is, the role of science is only to describe what happens, leaving aside ontological speculations about the “why” of the things being as they are.

MAIN FOCUS

Operators in quantum mechanics commonly provide representations of experimental interventions. But this does not mean that reality is determined by the observer. In fact, the experiment is just a physical intervention rationally controlled similar to the natural physical interventions occurring at random in the interactions between systems. An-

other important consideration to be made is about the clear representational role that mathematical constructs which serve to quantum mechanics must have, for example, the construct “quantum gate”. Accordingly McMahon (2008),

A gate can be thought of as an abstraction that represents information processing. [...] In a quantum computer, information is also processed using gates, but in this case the “gates” are unitary operations. Since quantum gates are just unitary operators, we’ll often go back and forth between the words gate and operator --- so keep in mind they mean the same thing in this context. (McMahon, 2008).

Then, we may say that quantum gates are connectors that allow us to building quantum circuits. They act upon quantum bits; thus, generalizing the concept, we may think of them as formal representations of particular circumstances imposed by the environment or the observer, never mind- ing whether they are controlled or not. They are operators of certain type acting on quantum bits to do something, that is, to produce some specific quantum configuration.

THINKING ABOUT NON-UNITARY OPERATIONS EMBEDDED IN ENTANGLED STATES: DOES IT MAKE SENSE?

In its current formal representation, quantum computation deals only with quantum gate operations which are necessarily unitary, a fact that turns difficult or even impossible to solve central problems such as decoherence and feasibility of measurements in the middle of the computation. From my theoretical background, the restriction to unitary gates and pure quantum states seems very arbitrary. Quantum gates can perfectly represent general quantum operations, not exclusively unitary, providing more flexibility and facility to building algorithms. Unitary operations can,

at best, be elected to represent the evolution of a quantum system under observational control, but not necessarily to represent quantum systems free of human intervention. Many works were performed on quantum gates (Barenco et al., 1995; Raussendorf & Briegel, 2000; Wang et al., 2001), so that the reader can deepen his particular search as desired.

What I want to show is that, given two entangled quantum bits, it is not possible to know whether the entanglement arose from the interference of some non-unitary action on a given quantum bit in a pure initial state. This means that quantum bits separated by large distances may carry effects of primary out-of-measurement processes that originated them. So, could we embed quantum states with these non-unitary actions, that is, these out-of-measurement transformation rules? If so, how would be the protocol for that?

First of all, as already suggested, I assume that certain operators reflect environmental conditions that favor the creation of new qubits from a primary qubit. In my former works, I developed the math entity named “progenitor” as the Kronecker operator to obtain two entangled qubits. According to my considerations on Bohr’s philosophical thinking, there must be an abstract quantum physical description of the natural process that leads to entangled qubits interacting through a quantum channel. I assume that if what connects two entangled qubits far from one another is a non-classical channel, that is, no transaction between entangled qubits occurs in common spacetime, so what triggered such a connection is an imaginary “operation”. So, the transfer of information in a quantum system, based on the non-classic connection between qubits, is holding through an imaginary “operation” like

$$\hat{G} := \frac{1}{\sqrt{2}} \begin{pmatrix} \mathbf{i} & -\mathbf{i} \\ 0 & -\mathbf{i} \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & -\mathbf{i} \\ 0 & -\mathbf{i} \end{pmatrix},$$

where $\mathbf{i} = \sqrt{-1}$, and

$$\begin{pmatrix} \mathbf{i} & -\mathbf{i} \\ 0 & -\mathbf{i} \end{pmatrix} \begin{pmatrix} -\mathbf{i} & \mathbf{i} \\ 0 & \mathbf{i} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

The object

$$\frac{1}{\sqrt{2}} \begin{pmatrix} 0 & -\mathbf{i} \\ 0 & -\mathbf{i} \end{pmatrix}$$

is the progenitor, that is, an operator which acting on a qubit by a Kronecker product on the left gives one two-qubit system in a certain configuration such that, under a control gate, it outputs a pair of entangled qubits. The Kronecker product protocol (KP) comes

$$\frac{1}{\sqrt{2}} \begin{pmatrix} 0 & -\mathbf{i} \\ 0 & -\mathbf{i} \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 0 \\ -\mathbf{i} \\ 0 \\ 0 \\ 0 \\ -\mathbf{i} \\ 0 \end{pmatrix}. \quad (14)$$

Applying the correct imaginary gate, it follows

$$\begin{pmatrix} \mathbf{i} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \mathbf{i} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \mathbf{i} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{i} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathbf{i} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \mathbf{i} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{i} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{i} \end{pmatrix} \times \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 0 \\ -\mathbf{i} \\ 0 \\ 0 \\ 0 \\ -\mathbf{i} \\ 0 \end{pmatrix} = \frac{-\mathbf{i}}{\sqrt{2}} \begin{pmatrix} 0 \\ 0 \\ \mathbf{i} \\ 0 \\ 0 \\ 0 \\ \mathbf{i} \\ 0 \end{pmatrix} =$$

$$= \frac{1}{\sqrt{2}} |11\rangle + \frac{1}{\sqrt{2}} |00\rangle. \quad (15)$$

The reader must remember that tensor product warrants the permanence of the superposition principle, i.e., there is a way to have a little of the $\hat{G}$ -operator and a little of the original quantum state in the final state.

The advantage of this approach is the presumption of a genuine quantum channel through which imaginary transformations occur, even though we do not know the dynamic essence of the non-local phenomenon; so, the imaginary operations are the logical quantum channel paths (not physical paths in spacetime).

PRESERVING QUANTUM ENTANGLEMENT

The question is not so much of mathematical theorems but appropriate representations for the physical phenomena examined here. Thus, all my effort was directed to set representations that can be implemented in quantum algorithms to solve or minimize decoherence problems.

The quantum world is very difficult to understand because of the lack of realistic correspondence with Euclidian world. But, if we open our minds and enlarge common geometry to imagine the geometry of an object with no extension we shall be very close to the language we need to describe quantum mechanical facts in quantum computing. For the sake of simplicity, we could call “point” any indivisible structure, not uniquely the point of geometry. Thus, two entangled particles would constitute a “point” in which refers their interactions. However, this point was called “tapestry” in my quantum language. It may sound strange to call “tapestry” an entity with no extension, but this is intentional. A tapestry is understood from the onefold coverage provided by unique progenitor tensor product on one qubit. This is like to define the element of a wool rug, the minimal knot that begins a complete Persian tapestry. Thereby, one simple knot is in fact a tapestry of one element.

A fundamental issue about entanglement is that, no matter how far apart two entangled particles are from one another, what happens to one brings instantaneous response from the other. I sustain that if we use the particle image of matter, it is impossible to conceive the phenomenon of quantum entanglement, since what affects the particle can not propagate instantaneously to another particle. On the other hand, if we think about a continuous entity as a tapestry not made up by parts, then it is easy to see much more; to have no parts means to be indivisible in the space of configuration. In fact, two entangled particles constitute a physical *monad* as in the metaphysical Leibinizian sense. What is really missing is a rational form of expression suitable for such a phenomenon, both mathematically as literarily. We can well conceive the quantum *monad* as an imaginary tapestry of one element, cohesive, indivisible and intractable by common sense.

Our tapestry is an imaginary quantumfold, that is, a onefold that only exists in quantum descriptions of nature. From this quantumfold we may obtain an imaginary representation of entangled states by tensor operations applied on it. As the quantumfold is not geometrically thinkable because it is not composed by parts, it is covered, as I said, by unique progenitor tensor product to one qubit. To extract entangled states from this coverage, that is, to obtain real descriptions of states not separable from superposition principle, we logically need one imaginary gate, the only to transform (that is, to Wick-rotate) representations of imaginary objects into real ones. This gate builds a bridge between the two representations. So, let us take an important definition.

Definition 1: A tapestry $\mathbb{T}$ is a map that carries a pair progenitor-cum-qubit $(\xi^\dagger, \psi^\pm)$ on a column vector $\mathbb{I}$ with entries $(0, -\mathbf{1})$ by a Kronecker product $\langle\langle \dots, \dots \rangle\rangle$, so that

$$\mathbb{T} : \langle\langle \xi^\dagger, \psi^\pm \rangle\rangle \rightarrow \mathbb{I}_{(0, -\mathbf{1})}.$$

This approaching is derived from Serpa's proposal on Wick-rotations (Serpa, 2015). In some sense, tapestry is the generalized geometrical *locus* of qubit transformations that lead to entanglements.

Definition 2: A quantum channel is a manifold built by the connected sum of two or more tapestries such that the canonical representation of each tapestry is a word as

$$\Gamma_1 \Gamma_2 \dots \Gamma_g \Lambda_1 \Lambda_2 \dots \Lambda_h = 1,$$

meaning that the tapestry has *genus* g and h holes.

CONNECTED TAPESTRIES OVER SERVERS IN CLOUD

But not everything is perfect, since the phenomenon of decoherence — the loss of entanglement by environmental interferences — haunts quantum computing labs and brings puzzles to theorists. The problem of decoherence tantalizes scientists since long ago. However, it was demonstrated that the existence of quasiparticle excitations named non-Abelian anyons, neither classified as bosons nor fermions, is related to certain topological configurations that make immune to local decoherence the quantum information stored in such configurations (Sarma, *et al.*, 2006). Also it was reported recently what would be the first experimental demonstration of a loss resilient entanglement-based protocol, probing that, in some circumstances, it's possible to preserve properties acquired by qubits at first in entangled states (Zhang, 2013). These findings encourage theoretical investigations, including progenitor's protocol. According to the last, when a pair of qubits is produced by progenitor's action on a qubit, the imaginary tensor “operation” is somewhat “memorable” inside the generated pair in

tapestry. As we know, tensor product retains a little of the progenitor and a little of the original quantum state in the final state. To understand exactly where I want to go, let us consider the concept of measurement from the point of view of quantum mechanics.

Following von Neumann, we say that a consistent description of the measurement process in quantum physics must consider the interaction between the quantum system under observation and the quantum measurement apparatus (von Neumann, 1932). Thereby, a measurement is an intervention described by a unitary transformation that evolves the initial global state of the combined system. In this sense, the application of a certain controlled-gate builds a fact from what really happens. From the point of view of the model presented here, quantum entanglement precedes every physical measurement operation; two qubits are said entangled if they result of the transformation of a single qubit via progenitor, such that there must be at least one controlled-gate (a Wick-rotation matrix) capable of translating this entanglement as mathematically associated with an observable, albeit indirectly. The proposed protocol establishes the mathematical design of two entangled qubits from one qubit and one progenitor instead of two former qubits. Theoretically, from the notions of tapestry, imaginary quantum channel and progenitor it is possible to reduce the loss of the amount of entanglement. Now we take the Kronecker protocol (14), from which we have the tapestry representation of entangled states

$$KP|10\rangle = 0|01\rangle + 0|01\rangle + 0|11\rangle + 0|11\rangle + \\ + 0|00\rangle + 0|10\rangle - \frac{\hbar}{\sqrt{2}}|00\rangle - \frac{\hbar}{\sqrt{2}}|10\rangle.$$

Relating the terms of last expression with the formalism of surface topology, we can imagine an algorithm that converts this representation into a string (or word) such that in server *A* we read

$$|iI\rangle_A = aacbbdef$$

with

$$\begin{aligned} a &:= 0|01\rangle; \\ c &:= 0|11\rangle; \\ b &:= 0|00\rangle; \\ d &:= 0|10\rangle; \\ e &:= -\hbar/\sqrt{2}|00\rangle; \\ f &:= -\hbar/\sqrt{2}|10\rangle. \end{aligned}$$

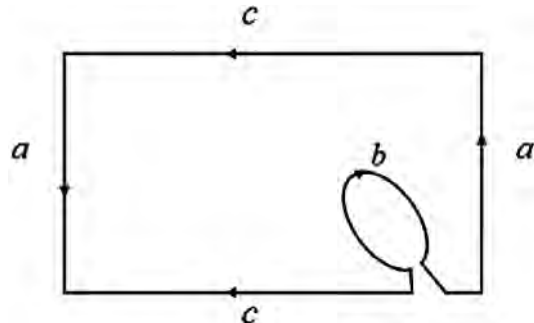
We can make $bdef \rightarrow b$, so that canonical representation gives

$$aacbb = 1,$$

which is a Klein bottle formed by two cross caps (aa, cc) and a hole (b) (see Figure 1).

Clearly, the topology shall depend on the initial number of entangled qubits. It is obvious that cyclic order is not important to identify the topology, but once determined this latter, the cyclic order operates as a part of the signature of the entanglement itself; we may fix the total additional information about the implicit manifold, including that signature, to be transmitted by a classical channel to server *B* into a complete topological information-state packet coupled to the qubit, so that we may ensure high efficiency and fidelity repairing entanglement. The number

Figure 1. Topological planar model of $KP|10\rangle$



of possible anagrams (signatures) from a word in which there is repetition is given by

$$P_n^{(q_1, q_2, q_3 \dots)} = \frac{n!}{q_1! \cdot q_2! \cdot q_3! \dots},$$

where $q_1, q_2, q_3 \dots$ are the numbers of times that repeated letters appear in the word. For the case of Klein bottle,

$$P_5^{(2,2)} = \frac{5!}{2!2!} = 30.$$

Thus, there are 30 possible signatures for the Klein bottle.

That information-state packet is the “memory” of entanglement and serves to preserve it from external perturbations. The connected sum (entanglement) $|i\mathcal{I}\rangle_{A\#B}$ represents the physical connection — the quantum channel — between the qubits and is written as

$$|i\mathcal{I}\rangle_{A\#B} = a_1 a_1 c_1 c_1 b_1 a_2 c_2 c_2 b_2.$$

Then, entanglement is a quantum phenomenon in which, through the quantum channel, the two tapestries in servers A and B are connected as the two *tori* in Figure 2. All we have to do is to transmit the topological information-state associated to the Kronecker protocol from server A to server B . If noise removes entanglement between two qubits, the reapplication of the control imaginary gate (the reconstruction of the quantum channel between the qubits) in server B through a quantum circuit will restore, in thesis, the entanglement from that memorized information shared by the two qubits.

$$a_1 b_1 a_1^- b_1^- a_2 b_2 a_2^- b_2^-.$$

Figure 2. Connected sum of two tori



ORCHESTRAL METALANGUAGE

To think about quantum cloud computing is a natural consequence when we investigate the potential of quantum machines. From here, it rises a plethora of interesting subjects, including security and privacy of computation by blind quantum computing based on the transmission of individual photonic qubits (Barz *et al.*, 2012). Thus, it is a great challenge to understand the complexity of services to be orchestrated in quantum clouds.

The task of identifying and matching services is far from obvious. Potential applications of quantum mechanics in everyday life always bump into a linguistic modeling problem, a fact which greatly complicates the accurate understanding of what we want to do and what we can really do. For this reason, the structuring of an orchestral metalanguage is decisive for the correct construction of the topology of services, evidencing scalability and giving descriptive accuracy and, in the same breath, great facility of implementation of changes. In addition, as well observed by Metodi and Chong, a quantum computer of practical value must be up to storing and orchestrating a system comprising tens of millions qubits (Metodi & Chong, 2006). Therefore, that orchestral metalanguage would be an interesting tool to represent such a complex dynamics.

With respect to the above introduced metalanguage and to make it clearer, it is important to consider the idea of orchestration as referring not

only to the service composition, but also to the topology that determines the order in which the services occur in a given process. We begin by defining a functional f as a unique service for state analysis. It informs a given composite service S about the best orchestration of elementary services to be performed depending on demand and the current state of the environment or on a collection of orchestrations identified on the same basic services to be initialized in parallel, depending on the overhead consumption. There are two ways of doing work one functional on a composite service as we shall see.

The lexicon of the metalanguage, that is, its main catalogue of elementary connectors and words, is given by what we call a primitive base as follows:

Now we look at the grammar of the metalanguage, that is, the rules to combine connectors and words in such manner that we may build meaningful statements (axioms, definitions, sentences, *etc.*):

Table 1. Primitive base

S	Composite service
s	Elementary service
f	Functional
Υ	Combined with (to compose services)
$\succ$	Call forward
$\prec$	Call backward
$\prec \succ$	Parallel running
$\wedge$	Defined as
$\langle f $	Right functional
$ f \rangle$	Left functional
O^1	Orchestration in one dimension

Definition 1: It is called weak coupling the orchestration O^1 of elementary services s_i to which there is mutual communication among the orchestration component services.

Axiom 1: For any set of elementary services s_i there is at least one weak coupling O^1 on s_i , such that O^1 on s_i is an application of the topology T^1 on s_i equivalent to the composite service S^1 :

$$\left\{ \forall \{s_i\} \exists O_{\{s_i\}}^1 / O_{\{s_i\}}^1 : T^1 \{s_i\} \Leftrightarrow S^1 \right\}.$$

Definition 2: It is called agglomerate the meeting O^k of k weak couplings on the same basic services, being $k = 1, 2, 3 \dots n$ the size of the agglomerate and $O^k \Leftrightarrow S^k$.

Axiom 2: For every set of basic services s_i , there is at least one agglomerate O^k on s_i , such that O^k on s_i is an application of the topology T^k on s_i equivalent to the composite service S^k :

$$\left\{ \forall \{s_i\} \exists O_{\{s_i\}}^k / O_{\{s_i\}}^k : T^k \{s_i\} \Leftrightarrow S^k \right\}.$$

Definition 3: It is called “left” functional action the initialization of the weak coupling O^1 which has the best performance among all identified weak couplings on the same basic services.

Definition 4: It is called “right” functional action the initialization of the agglomerate O^k .

Corollary: All weak coupling O^1 is an agglomerate of dimension $k = 1$.

It must be understood that both lexicon and grammar can be enlarged as the representational complexity advances. For instance, let us take an example of a functional action at “left” and at “right”,

$$\begin{aligned}
& |f\rangle S^1 \wedge \langle s_1 \succ \langle s_2 \succ s_3 \rangle \succ \langle s_4 \succ s_5 \succ s_6 \rangle \rangle; \\
& S^4 \langle f | \wedge \left\| \begin{aligned} & \langle s_1 \succ \langle s_2 \succ s_3 \rangle \succ \langle s_4 \succ s_5 \succ s_6 \rangle \rangle \\ & \langle s_1 \succ \langle s_2 \succ s_3 \rangle \succ \langle \langle s_4 \succ s_5 \rangle \succ s_6 \rangle \rangle \\ & \langle s_1 \succ \langle s_2 \succ s_3 \rangle \succ \langle s_4 \succ s_5 \succ s_6 \rangle \rangle \\ & \langle s_1 \succ \langle s_2 \succ s_3 \rangle \succ \langle s_4 \succ \langle s_5 \succ s_6 \rangle \rangle \end{aligned} \right\|, \\
& S^4 \langle f | \wedge \left\| \begin{aligned} & O^1 \\ & O^2 \\ & O^3 \\ & O^4 \end{aligned} \right\|.
\end{aligned} \tag{16}$$

Expression (16) is called parallelism matrix and its dimension k depends primarily on the number of elementary services and the number of available qubits. It is understood that service combinations, expressed between kets, must precede external calls. Large service chains with related functions can be represented in this way, documenting all the required topologies. Considering two servers, A and B , and taking the last column-matrix, quantum teleportation of this four-dimensional agglomerate from server A to server B , by the instruction 2 from Equation 6, would be given by

$$\begin{aligned}
I_2 : P_2 |M\rangle |O^1 O^2 O^3 O^4\rangle |\Psi_{(1)}\rangle_A &= > \\
\Rightarrow |\delta\rangle |\Psi_{(1)}\rangle_B &\equiv |O^1 O^2 O^3 O^4\rangle.
\end{aligned}$$

Operating under metacomputation, B will perform the same task as server A could do, however, with no needs to repeat the process of topological analysis done by A . In server B , all topological possibilities can be used simultaneously to perform parts of a computation (service).

Figure 3 (in appendix) outlines the SOA overlay intermediating clients and quantum machines in a certain hypothetical production environment. Servers A , B and C are supercomputers originally sharing the same quantum channel. At the right of the ESB a client requests certain complex service. By means of a protocol translation hardware (PT)

located at the left of the ESB, a query is addressed to the quantum machines B and C to know whether the requested service is already available. If the answer is negative, the request is forwarded to the quantum analyzer A . Until now, all we have done took place by means of classic channels. From now on, having defined the best orchestration, the quantum analyzer teleports the state matrix to both servers B and C ; the probabilistic parallel processing begins in B and C at the same time that state matrix is destroyed at A . This diagram was inspired in a more general scheme called one-to-many teleportation. Due to entanglement, probabilities in B and C interfere with one another. The requested service results from the instantaneous “collapse” (see Figure 4 in appendix) of the copies of the state matrix into new states at PT (mathematically, this is a change of probabilistic reference class). Finally, service is really available for the client through ESB. While in machines B and C the service stays divided in probabilities, only manifesting as an effective product in daily world after protocol translation. With continuous reductions, that is, with unbroken chain of very fast reductions, availability and quickness are theoretically warranted in a level never seen before. Lastly, as pointed out by Schmidt and colleagues, the virtualized infrastructure of the bus allows it to grow or shrink according to the workload which it is supporting (Schmidt, 2005).

The one-to-many teleportation has become well known since the end of the nineties (Murao *et al.*, 2000), now considered by Ghiu (2012). Let us first establish the initial state in server A related to a half-spin particle,

$$|\psi\rangle_A = \alpha|0\rangle + \beta|1\rangle.$$

We want to broadcast the information of this state to servers B and C , so that they share the final state

$$|\Psi\rangle_{BC} = \alpha|\phi_0\rangle + \beta|\phi_1\rangle.$$

The general representation of a quantum channel shared by the three servers is given from

$$|C\rangle_{ABC} = N(|0\rangle_A |\phi_0\rangle_{BC} + n|1\rangle_A |\phi_1\rangle_{BC}),$$

where

$$N = 1 / \sqrt{1 + n^2}.$$

If we take parameter $n = 1$ we get one-to-many teleportation. Now considering Bell-states, it follows the whole system state

$$|\psi\rangle|C\rangle = \frac{N}{\sqrt{2}} \left[|\Psi_a^+\rangle(\alpha|\phi_0\rangle + \beta|\phi_1\rangle) + |\Psi_a^-\rangle(\alpha|\phi_0\rangle - \beta|\phi_1\rangle) + |\Psi_b^+\rangle(\beta|\phi_0\rangle + \alpha|\phi_1\rangle) + |\Psi_b^-\rangle(-\beta|\phi_0\rangle + \alpha|\phi_1\rangle) \right].$$

Lastly, servers B and C have to make local appropriate transformations to get final state

$$|\psi\rangle = \frac{1}{\sqrt{|\alpha|^2 + n^2|\beta|^2}} (\alpha|\phi_0\rangle + n\beta|\phi_1\rangle) = \alpha|\phi_0\rangle + \beta|\phi_1\rangle.$$

Now we apply the metalanguage defined above. Each elementary service has a complete state function, so that, with Bell-basis, we may write for the whole system

$$S^4\langle f | \wedge \frac{N}{\sqrt{2}} \left\| \begin{array}{l} \langle \psi_1 \succ \langle \psi_2 \succ \psi_3 \rangle \succ \langle \psi_4 \succ \psi_5 \succ \psi_6 \rangle \rangle \\ \langle \psi_1 \succ \langle \psi_2 \succ \psi_3 \rangle \succ \langle \psi_4 \succ \psi_5 \rangle \succ \psi_6 \rangle \rangle \\ \langle \psi_1 \succ \langle \psi_2 \succ \psi_3 \rangle \succ \langle \psi_4 \succ \psi_5 \succ \psi_6 \rangle \rangle \\ \langle \psi_1 \succ \langle \psi_2 \succ \psi_3 \rangle \succ \langle \psi_4 \succ \langle \psi_5 \succ \psi_6 \rangle \rangle \rangle \end{array} \right\|, \quad (17)$$

So we can say that the functional applied to the right in the above expression corresponds to the initialization of the quantum channel between the functions ψ_i that make up the orchestration S^4 , so that, for each elementary service s_i we have

$$\psi_i = \left[|\Psi_{i,A}^+\rangle(\alpha|\phi_0\rangle + \beta|\phi_1\rangle) + |\Psi_{i,A}^-\rangle(\alpha|\phi_0\rangle - \beta|\phi_1\rangle) + |\Psi_{i,B}^+\rangle(\beta|\phi_0\rangle + \alpha|\phi_1\rangle) + |\Psi_{i,B}^-\rangle(-\beta|\phi_0\rangle + \alpha|\phi_1\rangle) \right].$$

DISCUSSION

According to *Definition 1* and *Axiom 1*, S^1 has flexible nature enough to incorporate virtually any weak coupling O^1 on s_i services. Based on the *Definition 2* and the *Axiom 2*, this flexibility extends to n dimensions according to *Definition 4*, which states that in practice an array of identified weak couplings is executable under demand and according to the availability of resources. *Definition 3* requires logistical criteria previously established in the architecture itself. It is worth remembering that both auxiliary services and application services fall within the formal framework described above.

The word “agglomerate” was used rather than “cluster” precisely to avoid confusion with the concept of “cluster of machines”. Based on quantum principles, Server A arrived at the best possible solutions in four dimensions. During the short period of processing, the computer repeated the test in hundreds of different ways to make sure that there was not a better selection of ways to perform the required task. Thus, given a cloud under a SOA overlay disposing services $s_1, s_2, s_3, s_4, s_5, s_6$, this agglomerate shall be analyzed in server B according to local environmental conditions. The advantage is that the teleported matrix already contains the best selections of orchestrations on the same services for a certain global task to be replied in a remote location.

Due to the uncommon nature of the qubit itself, in comparison with the classical bit, quantum computers are expected to prove in labs to be able to operate many times faster executing complex tasks of analysis and recombination. Nevertheless, as Nielsen and Chuang pointed out, “we do not understand what exactly it is that makes quantum

computers powerful, or on what class of problems they can be expected to outperform classical computers” (Nielsen & Chuang, 2000). In fact, our example of orchestration was a simple one, but in reality quantum servers shall deal with a high number of services and dimensions, a situation now difficult to govern by common computers. In addition, quantum principles applied in computation should help to solve the most challenging problem in computer science: the construction of learning machines. By making computers select and analyze teleported agglomerates (as server *B*) based on previous experiences (server *A*), there is hopeless to improve artificial intelligence in clouds for complex decisions related to global scenarios of production. I think that there shall be not by way of individual processing but by metacomputation that we shall obtain the best performance gains and cyber intelligence with quantum machines.

SPONTANEOUS ENTANGLEMENT

In fact, the creation of a pair of qubits from one qubit, as presented previously, is understandable as an outcome of the growth in the complexity of cybernetic autonomous devices of information interchanging and their links, but changes of complexity stay obscure; they require spontaneous entanglement. This is what keeps physicists separate from the world outside the laboratory.

Quantum processors need to operate at superconductivity regime in order to make superposition happen. A viable way to achieve this is using metal niobium and lowering the temperature of the apparatus to -272.98°C, close to absolute zero. This is a physical precondition to hold quantum phenomena. It happens that spontaneous entanglement is a response to complex stimuli from the environment; the more you induce the increasing required complexity of the system, the more you rise the chances of new entanglements. As the phenomenon of mutation useful for the survival of a species, or the emergence of new synapses in the brain, allowing connections between in-

tellective processes, it is not known precisely how occurs spontaneous quantum entanglement between qubits from the incitement of the process until the conflagration of the fact itself; it is an evolutionary interval that remains confined to a black box. Spontaneous entanglements are not observed, in the same way that it is not feasible a snapshot of the natural extinction of a species.

A METAFRAME FOR CLOUDS IN HILBERT SPACE

Constructions of type-cloud are more than sets of devices. Clearly, there is a succession of scales if we agree that to be a member of the larger system (cloud), the element (server) must be enrolled in some metric with tier below the tier of the metric of the first. Altaisky (2001) understood very well the problem of formal description of complex cybernetic systems, including systems of type-cloud. A state function to describe the members x (servers) of an object X (cloud) would be

$$\{\Psi(X), \Psi(X, x)\},$$

being $\Psi(X)$ the global state function and $\Psi(X, x)$ the state function of the elements. Since the $\Psi(X, x)$ are in $\Psi(X)$, it is not valid the commutation rule

$$[\Psi(X), \Psi(X, x)].$$

This means that, in principle, we can not apply operators on these functions such that we can measure both simultaneously; either we observe the overall behavior of the cloud, or the isolated behavior of one of its members. In theoretical terms, in a perfectly entangled quantum cloud, to measure the state of a server blurs the overall state of the cloud and vice versa. As the objects $\Psi(X, x)$ and $\Psi(X)$ inhabit different functional spaces,

their signatures or sets of coordinates associated with them assume a hierarchical network. Each level of the hierarchy is described by the structure

$$\mathcal{M} = \langle \ell, \beta^\ell, T^{\beta^\ell} \rangle$$

in which

$\ell \rightarrow$ the signature of scale,

$\beta^\ell \rightarrow$ the symmetry group at scale ℓ ,

$T^{\beta^\ell} \rightarrow$ some topology on β^ℓ (in other words, the coordinates at ℓ -th level).

In Hilbert space $\mathcal{H}$ of these hierarchical states,

$$\Psi_1, \Psi_2 \in \mathcal{H}; \alpha, \beta \in C \Rightarrow \alpha\Psi_1 + \beta\Psi_2 \in \mathcal{H},$$

where C is the set of the complex numbers, we would have for the general state, by definition,

$$\alpha\Psi(X) = \left\{ \alpha\psi_X^{\ell_1}(\tau^{\beta^{\ell_1}}), \right. \\ \left. \left\{ \alpha\psi_{X_{x_1}}^{\ell_2}(\tau_1^{\beta^{\ell_2}}), \dots, \alpha\psi_{X_{x_n}}^{\ell_2}(\tau_n^{\beta^{\ell_2}}) \right\}, \dots \right\},$$

remembering that by the re-ingoing nature of $\Psi(X)$ there is no commutation.

The structure of $\mathcal{M}$ is enough to preach cybernetic systems of type “cloud”, although here I have provided only a brief formalism. This structure includes complexities such as dynamic provisioning of computing resources, dynamic balancing of the workload and performance monitoring. The application of cloud computing is a reality on the Internet (Google and Yahoo). In 2008 the total of the clouds of the five largest Internet search companies amounted to around 2 million servers. The main advantage of this computational model is the significant reduction of

the time-to-market for on-demand e-business and Web 2.0 applications, that is, the gradual allocation of resources by necessity.

THE BEEHIVE EFFECT

Quantum communication uses the informational content of entangled systems in order to obtain an extra resource. Quantum entanglement is essential to reach the exponential speed-up anticipated by some quantum algorithms. From the theoretical point of view, all the information in a state of maximum entanglement is contained in the joint properties of the systems and not in individuals separately. The so-called “beehive effect” originates from a large scattering of entangled states distributed across multiple quantum servers via broadcast channels. Aiming to make an effective distribution of information by entangled states we would have to build a linked set of transmitters like in the most stable going experiments in which we start from the distribution of entangled photons through glass fibers, since they are installed underground and thus they are less vulnerable to external disturbances. Eventually, it will be necessary the use of satellite technology. Due to the spread of entangled states over arbitrary distances, the cloud assumes a global behavior from each stimulus locally introduced. The global response is resulting not only from the entanglement, but also from the state analysis promoted by auxiliary services of SOA architecture on the entangled states.

It is usual to expect that quantum computing comes to fruition in the next ten/twelve years, mainly solving problems about the transfer of large amounts of complex data by teleporting based on quantum entanglement. The evolution to an intelligent cloud of entangled quantum servers with the ability to send and receive large amounts of data analyzing and deciding what to do is a more distant win I suppose, but even so it would be risky to make estimates from the present stage of research.

FUTURE RESEARCH DIRECTIONS

In the future, clouds will be able to make global decisions supported by entangled states of information shared among all quantum servers, including via teleportation of entanglement itself. As well as, local decisions will be madden after the beehive analysis of the situation, a complex task which requires multiple entanglements and teleportations. Quantum cloud computing is still in its infancies, but it is very far from science fiction.

CONCLUSION

Physicists agree on that we are a long way off —decades, they suppose — from employing quantum features to build quantum hardware with practical applicability (Woo, 2013). The correlation quantumness between particles is in principle given by entanglement, and entanglement is very sensible to environmental perturbations, which turns impracticable to maintain superposition of states. Researchers are now working on new ideas to quantify the disagreement (the “discord”) between quantum and classical ways of calculating the same property as a manner to solve the problem of sensibility to the environment, but it is still not clear if discord really fulfills general computing quantumness (Tyler, 2013). As Professors Hadjiivanov and Todorov said,

Quantum mechanics, created during the first quarter of XX century is finding wide applications only after the invention of the transistor in 1948 and the development of the laser in the late 1950's. The true applications of the 'second quantum revolution' are yet to come. (Hadjiivanov & Todorov, 2015).

Currently, with the improvement of laser technology, there are several advanced experiments

applying quantum entanglement/teleportation over distances of about 89 mi. We expect that computers endowed of quantum microprocessors shall develop capabilities to perform what I call “self-entanglement”, transmitting packets of entangled states with embedded spinor-like gates to other computers in a cloud, interacting with high efficiency by teleportation of states. The “quantum cloud” will be many times faster to provide services than any conceived architecture now available. This is the beginning of an increasing intelligence, since entanglement and teleportation open doors to infinity of interactions from server to server.

ACKNOWLEDGMENT

I acknowledge Doctor José Abdalla Helayël-Neto, at Brazilian Center of Physics Research, by the great support to this work.

REFERENCES

- Altaisky, M. (2000). *What can biology bestow to quantum mechanics?* arXiv:quant-ph/0007023v1
- Anderson, R., & Ciruli, D. (2006). *Scaling SOA with distributed computing*. Retrieved June 12, 2011, from <http://www.ddj.com/architect/193104809>
- Barenco, A., Bennett, C. H., Cleve, R., DiVincenzo, D. P., Margolus, N., Shor, P., & Weinfurter, H. et al. (1995). Elementary gates for quantum computation. *Physical Review A*, 52(5), 3457–3467. doi:10.1103/PhysRevA.52.3457 PMID:9912645
- Barz, S., Kashefi, E., Broadbent, A., Fitzsimons, J., Zeilinger, A., & Walther, P. (2012). Demonstration of blind quantum computing. *Science*, 335(6066), 303–308. doi:10.1126/science.1214707 PMID:22267806

- Bowen, W., Treps, N., Buchler, B. C., Schnabel, R., Ralph, T. C., Bachor, H.-A., & Lam, P. K. et al. (2003). Experimental investigation of continuous variable quantum teleportation. *Physical Review A*, 67(3), 032302. doi:10.1103/PhysRevA.67.032302
- Bouwmeester, D., Pan, J.-W., Mattle, K., Eibl, M., Weinfurter, H., & Zeilinger, A. (1997). Experimental quantum teleportation. *Nature*, 390(6660), 575–579. doi:10.1038/37539
- Braunstein, S. (1996). Quantum teleportation without irreversible detection. *Proceedings of the Royal Society of London*, 53, 1900–1903. PMID:9913086
- Charon, J. (1967). *De la fisica al hombre*. Madrid: Ediciones Guadarrama.
- Deutsch, D., & Jozsa, R. (1992). Rapid solution of problems by quantum computation. *Proceedings of the Royal Society of London. Series A*, 439(1907), 553–558. doi:10.1098/rspa.1992.0167
- Erl, T. (2005). *Service-oriented architecture: Concepts, technology, and design*. Upper Saddle River, NJ: Prentice Hall PTR.
- Frenken, T., Spiess, P. & Anke, J. (2008). A flexible and extensible architecture for device-level service deployment. *LNCS*, 5377, 230–241.
- Ghiu, I. (2012). *Generalized telebroadcasting of entanglement*. Department of Physics, University of Bucharest.
- Hadjiivanov, L., & Todorov, I. (2015). Quantum Entanglement. *Bulg. J. Phys.*, 42, 128–142.
- Lemos, A., Daniel, F., & Benatallah, B. (2015). Web service composition: A survey of techniques and tools. *ACM Computing Surveys*, 48(3), 33. doi:10.1145/2831270
- Linthicum, D. (2009). *Cloud computing and SOA convergence in your enterprise*. New York: Addison-Wesley.
- McMahon, D. (2008). *Quantum computing explained*. John Wiley & Sons, Inc.
- Forum, M. E. F. (2015). *The third network: Life-cycle service orchestration vision*. Whitepaper, Retrieved May 26, 2016, from https://www.mef.net/Assets/White_Papers/MEF_Third_Network_LSO_Vision_FINAL.pdf
- Metodi, T., & Chong, F. (2006). *Quantum computing for computer architects*. Morgan & Claypool Publishers.
- Murao, M., Plenio, M., & Vedral, V. (2000). Quantum information distribution via entanglement. *Physical Review A*, 61(3), 032311. doi:10.1103/PhysRevA.61.032311
- Nakamura, M. (2004). Implementing integrated services of networked home appliances using service oriented architecture. In M. Aiello, M. Aoyama, F. Curbera, M. Papazoglou (Eds.), *Service-Oriented Computing - IC-SOC 2004: Second International Conference*. doi:10.1145/1035167.1035206
- Natis, Y. (2007). *Twelve common SOA mistakes and how to avoid them*. Retrieved May 10, 2013, from http://www.gartner.com/it/content/754400/754413/twelve_common_soa_mistakes.pdf
- Nielsen, M., & Chuang, I. (2000). *Quantum computation and quantum information*. Cambridge, UK: Cambridge University Press.
- Raussendorf, R., & Briegel, H. (1995). *Quantum computing via measurements only*. arXiv:quant-ph/0010033
- Sarma, S., Freedman, M., & Nayak, C. (2006). *Topological quantum computation*. *Physics Today*. American Institute of Physics.
- Serpa, N. (2015). *New lectures on supergravity*. Saarbrücken: Verlag/Lambert Academic Publishing.

Shah, M. (2007). *SOA reusability: Shrinking the lag between business and IT*. Retrieved June 12, 2011, from [http:// today.java.net/pub/a/ today/2007/ 07/24/soa-reusability-shrinking-lag-time.html](http://today.java.net/pub/a/today/2007/07/24/soa-reusability-shrinking-lag-time.html)

Schmidt, M., Hutchison, B., Lambros, P., & Phippen, R. (2005). The Enterprise Service Bus: Making service-oriented architecture real. *IBM Systems Journal*, 44(4), 4. doi:10.1147/sj.444.0781

Tyler, C. (Ed.). (2013). *Quantum discord: Theoretical advance toward practical quantum computing*. Los Alamos Science and Technology Magazine.

von Neumann, J. (1932). *Mathematische grundlagen der quantenmechanik*. Berlin: Springer Verlag.

Zhang, T. (2002). *Quantum teleportation of light beams*. arXiv: quantph/0207076.

Zhang, Z., Tengner, M., Zhong, T., Wong, F., & Shapiro, J. (2013). *Entanglement's benefit survives an entanglement-breaking channel*. arXiv: quantph/13035343

Walther, P., & Zeilinger, A. (2006). Quantum entanglement, purification, and linear-optics quantum gates with photonic qubits. In L. Accardi, M. Ohya, & N. Watanabe (Eds.), *Quantum probability and white noise analysis* (Vol. 19, pp. 360–369). Singapore: World Scientific Publishing Co.

Wang, X. (2001). Multi-bit gates for quantum computing. *Physical Review*, 86, 3907–3910. PMID:11329354

von Weizsäcker, C. (1974). *La imagen física del mundo*. Madrid: Biblioteca de Autores Cristianos.

Woo, M. (2013). How Caltech physicists are helping to bring us ever closer to our quantum future. *Engineering & Science*. Retrieved July 08, 2013, from http://eands.caltech.edu/articles/LXXVII/2013_Spring_What_Is_a_Quantum_Computer.pdf

KEY TERMS AND DEFINITIONS

Beehive Effect: The supposed global — and even intelligent — behavior of a cloud of servers acting under quantum principles.

Cloud Computing: A model of computation by which IT resources are randomly dispersed in the network, being offered as services.

Progenitor: The gate generator of a two-qubit system which under the action of a control gate creates a pair of entangled states.

Quantum Bit (or Qubit): The quantum tile of information that can assume both states 0 and 1 at the same time.

Quantum Entanglement: The matting of quantum states to which decomposition does not hold.

Quantum Machine: A computer whose general operation follows the laws of quantum mechanics.

Quantum Teleportation: The long-distance replication of a quantum state.

SOA (Service Oriented Architecture): A computational architecture for the provision of services as packages of specific tasks over the network.

ENDNOTES

¹ A symmetry in quantum mechanics is a discrete transformation or a group of continuous transformations that let invariant the Hamiltonian (or the Lagrangian) and the canonical commutation relations of the system.

² The atom is not immediately perceptible to our senses, and any experiment takes only a specific property of the atom to the ambit of a mediated sensibility.

³ Every experiment is a material act which is simultaneously an act of perception.

APPENDIX

Figure 3. Operating sketch of a cloud of quantum machines, showing simple quantum cloud architecture linked to an Enterprise Service Bus through the protocol translator hardware

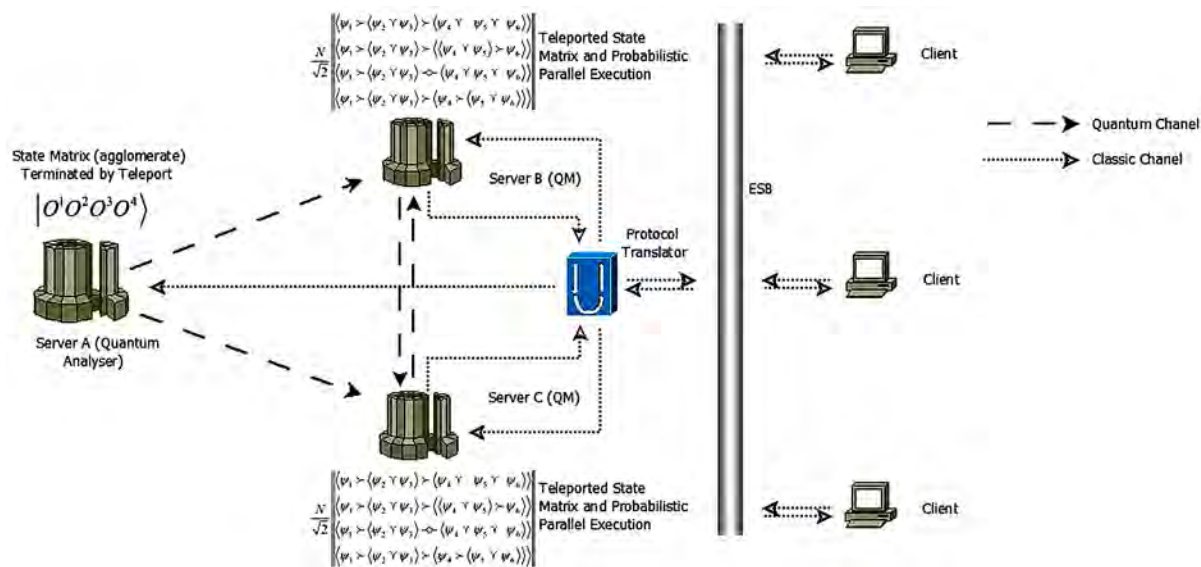
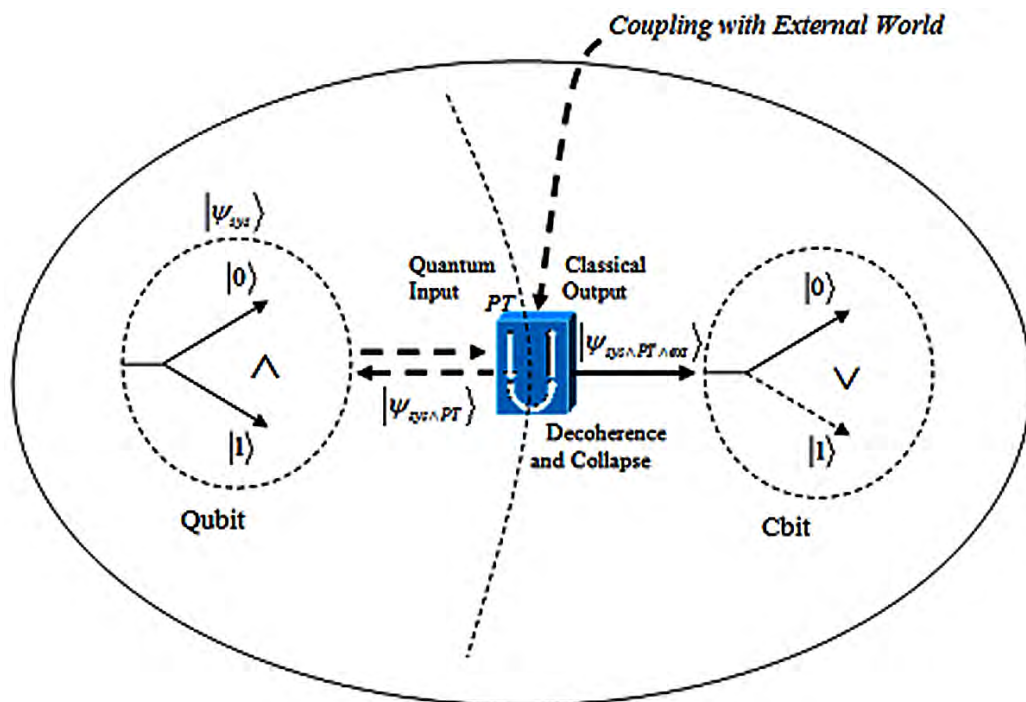


Figure 4. Detail of the interactions at protocol translator



Cyberinfrastructure, Cloud Computing, Science Gateways, Visualization, and Cyberinfrastructure Ease of Use

C**Craig A. Stewart***Indiana University, USA***Richard Knepper***Indiana University, USA***Matthew R. Link***Indiana University, USA***Marlon Pierce***Indiana University, USA***Eric Wernert***Indiana University, USA***Nancy Wilkins-Diehr***San Diego Supercomputer Center, USA*

INTRODUCTION

Computers accelerate our ability to achieve scientific breakthroughs. As technology evolves and new research needs come to light, the role for cyberinfrastructure as “knowledge” infrastructure continues to expand. In essence, cyberinfrastructure can be thought of as the integration of supercomputers, data resources, visualization, and people that extends the impact and utility of information technology. This article defines and discusses cyberinfrastructure and the related topics of science gateways and campus bridging, identifies future challenges in cyberinfrastructure, and discusses challenges and opportunities related to the evolution of cyberinfrastructure and cloud computing.

BACKGROUND

Today’s US national cyberinfrastructure ecosystem grew out from the National Science Foundation-funded supercomputer centers program of the 1980s (National Science Foundation, 2006). Four centers provided supercomputers and support for their use by the US research community. Researchers generally accessed one supercomputer at a time, sometimes logging into a front-end interface. At this time, the focus of the research computing community was centered on supercomputers – traditionally defined as computers that are among the fastest in existence. Over time there have been several different supercomputer architectures, but the key points were that supercomputers were monolithic systems that were among the fastest in the world. At present we can think of supercomputers as being a subset of the more general term high performance computer (HPC) – where HPC means that many computer processors work together, in concert, to solve large computational

DOI: 10.4018/978-1-5225-2255-3.ch092

challenges and where the computer processors communicate via very fast, networks internal to the HPC system. HPC focuses on computing problems where a high degree of communication is needed among the processors working together on a particular problem. HPC is a more general term than supercomputers because there are many HPC systems that are modest in total processing capacity relative to the fastest supercomputers in the world (cf. Top500.Org, 2016).

In the early days of supercomputing, using multiple supercomputers in concert was not possible. In the late 1980s, the National Research and Education Network initiative created several testbeds for distributed computing, including the CASA testbed which linked geographically distributed supercomputers to solve large-scale scientific challenges (U.S. Congress Office of Technology Assessment, 1993). A turning point in distributed high performance computing was the I-WAY project – a short-term demonstration of innovative science enabled by linking multiple supercomputers with high performance networks (Korab & Brown, 1995). It demonstrated the possibilities to advance science and engineering by linking supercomputers using high-speed networks.

In the late 1990s, the NASA Information Power Grid provided a production grid of multiple supercomputers connected by a high-speed network (Johnston, Gannon, & Nitzberg, 1999). Around this time began also the concept of high throughput computing (HTC) with a software system called Condor (Litzkow, Livny, & Mutka, 1988). HTC takes the approach of breaking a problem up into small pieces of work and distributing them to multiple CPUs over network connections that may be relatively slow. HTC best suits problems where relatively little communication is needed among the processors working together on a particular problem or simulation. Because HTC applications can operate relatively efficiently on processors with little communication among the processors, HTC applications have always fit naturally into a distributed computing environment (Thain,

Tannenbaum, & Livny, 2005). Today, a popular framework for distributed storage and processing of large data sets is Apache Hadoop (The Apache Software Foundation, 2006).

Over time, distributed computing evolved into ‘grids,’ with grids emerging as a commonly used term in the late 1990s. Typically, computational grids are the hardware and software infrastructure which provides access to the computational capabilities (Foster & Kesselman, 1998, 2004). *Middleware* is a key software component of cyberinfrastructure, enabling the disparate components of cyberinfrastructure to work together. In effect, middleware manages complex interactions between resources which allows for the development of new networked applications (National Science Foundation, 2004). Around the turn of the century, the US government funded two major grid projects – TeraGrid and the Open Science Grid. In 2001, the NSF funded an experimental computational, storage, and visualization resource called TeraGrid, which developed grid capabilities for supercomputer centers (National Science Foundation, 2006). The Open Science Grid (OSG) (Livny et al., 2006; Open Science Grid, 2015), first funded with that name in 2006, grew out of three projects that developed HTC grids for the purpose of analyzing data from physics experiments (Avery, 2007).

Tying geographically distributed computing systems together into grids to create a whole greater than the sum of its parts was widespread around the turn of the century. However, the term grid computing was becoming laden with sometimes competing definitions. In addition to computing and data grids, other terms such as collaboration, semantic, and peer-to-peer grids emerged, distinguished by the characteristics of the protocols and interactions between components (Fox, 2006). The potential for confusion and competing definitions of different types of grids led Dr. Ruzena Bajcsy, then NSF assistant director of the Computer and Information Science and Engineering Directorate, to use the term cyberinfrastructure when charging a new advisory group to offer advice to

the NSF – the “Blue Ribbon Advisory Panel on Cyberinfrastructure.” The term cyberinfrastructure had been used before, in a different sense, by Richard Clarke, then US National Coordinator for Security, Infrastructure Protection, and Counterterrorism (Clarke & Hunker, 1998). Bajcsy stated that she used the term cyberinfrastructure because she wished to “create a program ... that would involve the broader computer science/information technology community” (Bajcsy, 2013). The committee report goes on to state, “the newer term cyberinfrastructure refers to infrastructure based upon distributed computer, information and communication technology. If infrastructure is required for an industrial economy, then we could say that cyberinfrastructure is required for a knowledge economy” (Atkins et al., 2003).

Bajcsy’s successor at the NSF, Dr. Peter Freeman, stated that this report “led to the creation of a term for infrastructure that attempts to capture the integration of computing, communications, and information for the support of other activities (especially scientific in the case of NSF)” (Freeman, 2013). In 2007, Freeman wrote “cyberinfrastructure can have many definitions and, to some extent, the definition is in the eye of the beholder” (Freeman, 2007). To make it clearer for scientists outside of science and physics, Indiana University developed a definition identifying components and function:

Cyberinfrastructure consists of computing systems, data storage systems, advanced instruments and data repositories, visualization environments, and people, all linked together by software and high performance networks to improve research productivity and enable breakthroughs not otherwise possible. (Stewart et al., 2010)

The EDUCAUSE Campus Cyberinfrastructure Working Group and the Coalition for Academic Scientific Computation developed a definition which includes teaching and learning:

Cyberinfrastructure consists of computational systems, data and information management, advanced instruments, visualization environments, and people, all linked together by software and advanced networks to improve scholarly productivity and enable knowledge breakthroughs and discoveries not otherwise possible. (Dreher et al., 2009)

The characteristics distinguishing cyberinfrastructure from other IT terms and concepts is the inclusion of resources like instruments and sensor networks as well as people and a focus on knowledge breakthroughs. Cyberinfrastructure may be distinguished in particular from the more European term eScience on the basis of the explicit role of people in cyberinfrastructure. eScience is defined as “the large scale science that will increasingly be carried out through distributed global collaborations enabled by the Internet. Typically, a feature of such collaborative scientific enterprises is that they will require access to very large data collections, very large scale computing resources and high performance visualization back to the individual user scientists” (National e-Science Centre, 2010).

CYBERINFRASTRUCTURE TODAY

The broad use of cyberinfrastructure in science and engineering envisaged by Bajcsy is in ample evidence today. That cyberinfrastructure enables breakthroughs not otherwise possible is demonstrated by two Nobel prizes for work made possible by major cyberinfrastructure resources – the Open Science Grid and XSEDE.

The Open Science Grid is an international HTC resource. Many different organizations own the computers participating in the grid (Open Science Grid, 2015). OSG’s people part of cyberinfrastructure is organized through dozens of Virtual Organizations (VOs) that use the computational resources of the OSG, each supporting its own uses and users. Analysis of data from the Large

Hadron Collider (LHC) is the paradigmatic use case for HTC. LHC data can be broken down into large numbers of small data sets, each of which may be analyzed in isolation. The 2013 Nobel Prize for Physics was awarded to François Englert and Peter Higgs for the theoretical discovery of the particle now known as the Higgs Boson. The existence of the Higgs Boson was verified in experiments at the LHC, with the data analyses made possible by the OSG.

The largest HPC-oriented cyberinfrastructure in the use is the eXtreme Science and Engineering Discovery Environment (XSEDE) (Towns et al., 2014). XSEDE is a single, virtual system which is comprised of a collection of integrated and highly-advanced digital resources and constitutes the largest HPC resource funded by the US government (Towns et al., 2014; XSEDE, 2013, 2015). The 2013 Nobel Prize in Chemistry was awarded to Martin Karplus, Michael Levitt and Arieh Warshel, for the development of multiscale computer models of complex chemical systems. Karplus used resources of the TeraGrid, the predecessor of XSEDE, and Warshel uses resources of XSEDE (XSEDE, 2015).

Cyberinfrastructure may support a particular research domain or application. Cyberinfrastructure has also been widely adopted in the private sector, particularly in advanced engineering, medicine and pharmaceuticals, mining and oil exploration, finance, and manufacturing (Tabor Griffin Communications, 1998).

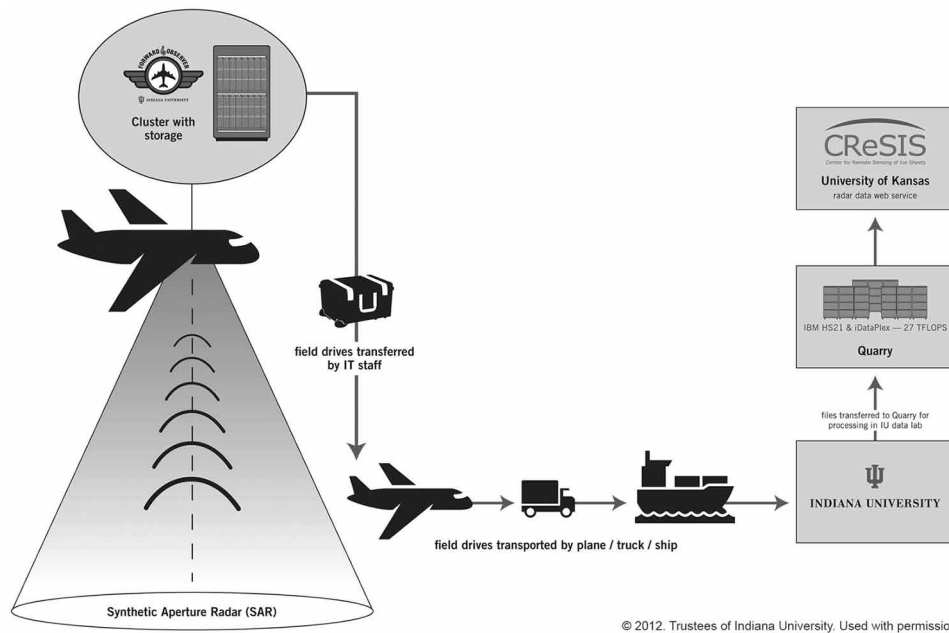
Cyberinfrastructure systems need not be massive to be important. A Specialized cyberinfrastructure supports NASA's *Operation IceBridge* in measuring polar ice sheets in Greenland and Antarctica. *Operation IceBridge* uses sophisticated synthetic aperture radar (SAR) systems to study polar ice and map the bedrock base in Greenland and Antarctica (Hayden, Fox, & Gogineni, 2007; Knepper, Link, & Standish, 2015). One of the characteristics of SAR is that one doesn't get an image out of SAR systems directly; a great deal of computation is required to generate an image. In-plane computation and data storage provide real-time analysis of multiple radar data sources (Figure 2). This cyberinfrastructure is highly specialized to deal with the rigors of fieldwork in Antarctica. The cyberinfrastructure designed to support *Operation IceBridge* enables real-time

Figure 1. The open science grid



Figure 2. NASA operation icebridge field radar data processing cyberinfrastructure

Source: Knepper, Standish, & Link 2015



reactions to data as it is being collected in an airplane over the Antarctic ice sheets – something not possible before this system was developed.

EVOLVING COMPONENTS OF CYBERINFRASTRUCTURE INCLUDING CLOUD COMPUTING

In his 2007 article, Freeman stated that the definition of cyberinfrastructure will evolve over time (Freeman, 2007). Cloud computing can thus be thought of as a particular approach to computing infrastructure and as a component of cyberinfrastructure which includes computational resources and data storage resources. According to the National Institute of Standards and Technology (NIST),

Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and

released with minimal management effort or service provider interaction. (Mell & Grance, 2011)

Key in cloud computing are on-demand self-service, broad network access, resource pooling, rapid elasticity, and measured service. It can be used as a solution to applications that are “hosted in the cloud” or integrated into cyberinfrastructure. Like a high performance computer, a cloud computing solution can be monolithic or integrated into a larger cyberinfrastructure facility. For example, the NSF recently funded a cloud system that will be integrated as part of XSEDE called Jetstream (Stewart et al., 2015). It can be used in isolation as a scientific cloud system or as part of a larger integrated cyberinfrastructure facility.

Data storage systems, advanced instruments and data repositories have also changed over time. Some recent changes in needs and data resources are described in the final report of the NSF Advisory Committee for Cyberinfrastructure Task Force on Data and Visualization (NSF Advisory Committee for Cyberinfrastructure Task Force on Data and Visualization, 2011) and the work

from Hey, Tansley, and Tolle (2009). The NSF has recently funded a new data-oriented storage and analytics facility called Wrangler to add new resources for big data to the US infrastructure funded by the NSF and supported by XSEDE.

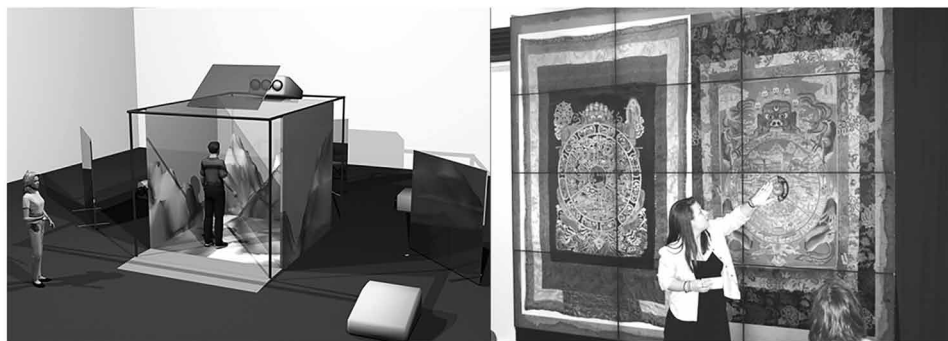
Middleware has evolved significantly since the early days of cyberinfrastructure. Globus, one of the most widely used families of software in the world, now includes authentication, secure access capabilities, and data and metadata management tools (Foster, 2005; Globus Online, 2013). Other middleware includes workflow systems that coordinate the use of cyberinfrastructure and automate complex analyses; examples include Apache Airavata (Maru et al., 2011), Kepler (Ludäscher et al., 2006), and Pegasus (Deelman et al., 2005).

Visualization systems — hardware (display, visualization, and interaction) and software (applications, libraries, middleware, and data format standards) — have evolved dramatically since the inception of the term cyberinfrastructure. Visualization was one of the earliest cyberinfrastructure components to promote distributed applications and high levels of interoperability, largely because of the network of homogeneous CAVE Automatic Virtual Environments (CAVEs) and smaller devices using similar software launched in the last half of the 1990s (NCSA, 2001). Us-

ers at multiple sites could synchronously interact with the same data sets and observe remote participants via virtual avatars while communicating over IP-based audio and video channels. CAVEs and similar devices introduced new capabilities for understanding complex 3D and 4D data from other cyberinfrastructure resources. However, cost and scarcity, limited their impact on day-to-day scientific investigation. The 2000s saw affordable graphics cards, projectors, and high-definition, stereoscopic displays. Consumer-level technologies spurred a range of innovative systems for stereoscopic and ultra-resolution visualization (Sherman, O’Leary, Whiting, Grover, & Wernert, 2010), democratizing advanced visualization systems and techniques. Figure 3 shows a CAVE diagram and an ultra-high resolution tiled wall assembled from commodity HD televisions. Shown in the figure at right, scholars at Indiana University’s Mathers Museum of World Cultures use an IQ-Wall to compare high-resolution images of textiles. This 3x4 wall is free-standing and was installed in a museum gallery in an afternoon. Such a display wall, which supports collaborative research, can now be created for a few tens of thousands of dollars, making them widely accessible in research environments.

Figure 3. At left is a CAVE, a room-scale visualization environment. At right is an ultra-high resolution tiled wall built in 2013 using commodity HDTV displays.

Source: © 2015, Trustees of Indiana University. Used with permission



FUTURE RESEARCH DIRECTIONS

Cloud Computing, Cyberinfrastructure, Exascale Computing, and the Economics of Computing

Cloud computing and more traditional HPCs (supercomputers) have complementary strengths and weaknesses. Cloud computing facilities may have internal networks of modest speed compared to supercomputers. On the other hand, cloud computing may be purchased in modest increments and are thus more accessible to a larger user community than supercomputers. Cloud computing is commonly used for “big data” applications, characterized by data volume, velocity, and variety (Laney, 2001). US President Obama’s recent executive order (Obama, 2015) sets a new agenda for the creation of exascale computing facilities (capable of 10^{15} mathematical operations per second) while calling for joint development of exascale and big data/cloud computing facilities. Fox and collaborators (Fox, Qui, Kamburugamuve, Jha, & Luckow, 2015) depict many of the commonalities between cloud computing and HPC and propose an alignment and set of commonalities between HPC and big data stacks that can form a foundation for the sort of joint development of both approaches called for by President Obama in his recent executive order.

Cloud computing and HPC need not be an “either/or” choice. There are business cases for selecting cloud computing or local servers (Brumec & VrčEk, 2013; Marston, Li, Bandyopadhyay, Zhang, & Ghalsasi, 2011). However, choosing between cloud and HPC can be very complicated as cloud resources may not be able to support science applications that require low-latency internal networks or large amounts of memory as prerequisites, even if the cost of cloud computing seems lower than HPC per CPU hour. A locally owned HPC or HTC system can be acquired as a one-time cost where the capacity of the system limits the usage over time but remains useable for

several years. In contrast, use of cloud computing may have a smaller cost for initial use but requires ongoing payments. This suggests tradeoffs in “locally owned” versus “cloud” that will suggest solutions strongly influenced by local conditions and financial systems at any given organization.

Cyberinfrastructure challenges include documenting return on investment and energy costs to operate at large scale. An analysis of return on investment in XSEDE is explored by Stewart and collaborators (Stewart et al., 2015). Energy costs and the economics of large-scale data centers help drive many activities into cloud computing. Security and data privacy are also concerns.

Science Gateways, Campus Bridging, and Cyberinfrastructure Ease of Use

For years, researchers accessed cyberinfrastructure exclusively through command-line interfaces. This sort of interface made it difficult to do long complex tasks and they were not particularly user friendly. Today, access to and the utility of cyberinfrastructure has been considerably expanded through the deployment of science gateways, use of cloud computing tools, and campus bridging. In particular, science gateways provide access to cyberinfrastructure to a broad set of users by employing graphical user interfaces and sophisticated tools for orchestrating computational workflows.

Science gateways make it possible to weave together a set of complicated tasks to achieve an overarching goal – like search for drug candidates or predict the path of a tornado. More formally, science gateways are defined as “a community-specific set of tools, applications, and data collections that are integrated together via a portal or a suite of applications” that can “support a variety of capabilities including workflows, visualization as well as resource discovery and job execution services” (Wilkins-Diehr, 2007). There are now dozens of science gateways in use or in development (Lawrence et al., 2015). For example, the CiPRES portal (Cyberinfrastructure for Phylogen-

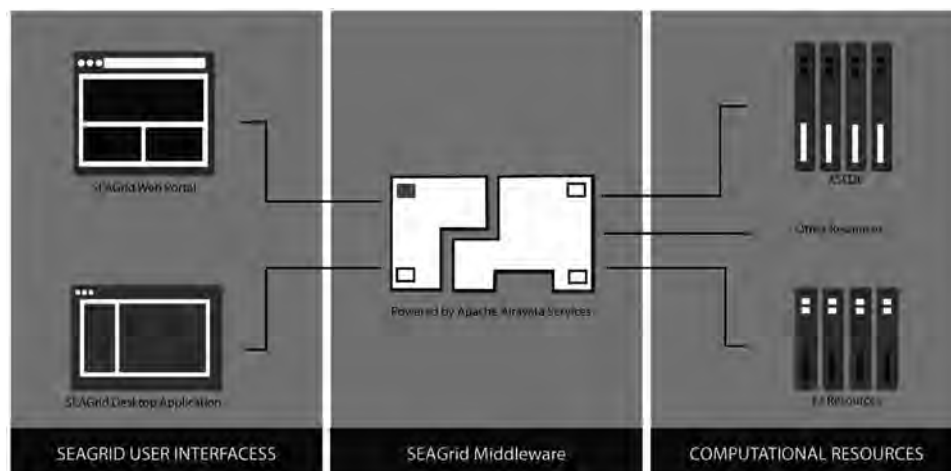
tic Research) enabled thousands of researchers – including research students – to do sophisticated analyses of evolutionary histories from genetic information (CIPRES, 2016). Another important science gateway is SEAGrid – the Science and Engineering Applications gateway (SEAGrid, 2015). SEAGrid is geared toward chemical and mechanical analyses, and can, for example, be used to search for potential new drug candidates. The major components of SEAGrid (see Figure 4) exemplify a multi-tiered approach commonly used in science gateways.

Science gateways have also had a profound impact on citizen science—the public contribution to scientific discoveries (OpenScientist, 2011). Zooniverse is a web-based front end to several science gateways supporting citizen science. Dozens of projects use citizen science in weather, archaeology, biology, and medicine, where thousands of people help analyze research data – particularly image data – that might otherwise go unanalyzed for months or years into the future (Zooniverse, 2015).

Campus bridging approaches a different set of cyberinfrastructure issues. One of the significant challenges in cyberinfrastructure is integrating across scales of resources. Campus bridging focuses on integrating local, often modest scale cyberinfrastructure facilities with regional, na-

tional, and even international cyberinfrastructure resources. The goal of campus bridging is to “bridge” from the campus to the approach enables small or resource-constrained groups to operate as if the higher-level resources were close at hand (Hallock, Knepper, & Stewart, 2015; NSF Advisory Committee for Cyberinfrastructure Task Force on Campus Bridging, 2011). Some campus bridging issues are solvable simply with funding; for instance, networking becomes ever cheaper and it becomes ever more feasible to have good network connectivity from campus to national resources. Recent efforts have focused on technical interoperability among campus computing clusters and national level resources like XSEDE. The technical aspects of such interoperability include adding software to local institutional cyberinfrastructure systems that match those used on nationally-shared resources and as a result enabling training and educational materials developed for nationally shared systems to be used in support of smaller, local resources. Networking in support of bridging from campus to national resources has also been furthered by a network concept called the Science DMZ, which explicitly creates a portion of network “specifically engineered for science applications and does not include support for general-purpose use. By separating the high-performance science network

Figure 4. Science gateways such as SEAGrid uses a multi-tiered gateway architecture



(the Science DMZ) from the general-purpose network, each can be optimized without interfering with the other” (ESnet, 2016).

CONCLUSION

Cyberinfrastructure has evolved from supercomputer centers into an integrated and distributed suite of powerful and flexible resources that integrate supercomputers, data resources, visualization, and people in ways that go beyond the capabilities of any of the individual components of cyberinfrastructure. It has led to new products, medical treatments, and improved business processes that improve quality of life. In the long run we believe that cloud computing, high performance computing, and high throughput computing will be seen not as alternatives but as complementary tools used flexibly in response to the particular science and engineering needs and particular local conditions of researchers and organizations making use of cyberinfrastructure.

In summary, the future offers tremendous opportunities for science and society to use cyberinfrastructure to enable new discoveries and improve the quality of life of people everywhere as new tools for visualization, science gateways, campus bridging, citizen science, and cloud computing evolve and deliver new capabilities to the public and the scientific and technical communities worldwide.

ACKNOWLEDGMENT

This material is based upon work supported by the National Science Foundation under grants 0504075, 0451237, 0723054, 1062432, 0116050, 0521433, 0503697, 1053575, and ACI-1445604, and support provided by the Indiana University Pervasive Technology Institute. Any opinions, findings and conclusions or recommendations expressed herein are those of the authors and do

not necessarily reflect the views of the supporting agencies. Robert Quick of IU created the OSG map in Figure 1. Editing by Greg Moore and Winona Snapp-Childs is gratefully acknowledged; any errors are the responsibility of the senior author.

REFERENCES

- Atkins, D. E., Droegemeier, K. K., Feldman, S. I., Garcia-Molina, H., Klein, M. L., Messerschmitt, D. G., ... Wright, M. H. (2003). *Revolutionizing Science and Engineering Through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure*. Retrieved from <http://www.nsf.gov/cise/sci/reports/atkins.pdf>
- Avery, P. (2007). Open science grid: Building and sustaining general cyberinfrastructure using a collaborative approach. *First Monday*, 12(6). doi:10.5210/fm.v12i6.1866
- Brumec, S., & Vrček, N. (2013). Cost effectiveness of commercial computing clouds. *Information Systems*, 38(4), 495–508. doi:10.1016/j.is.2012.11.002
- CIPRES. (2016). *CIPRES Cyberinfrastructure for Phylogenetic Research*. Retrieved June 30, 2016, from http://www.phylo.org/sub_sections/portal/
- Clarke, R., & Hunker, J. (1998). *Press Briefing by Richard Clarke, National Coordinator for Security, Infrastructure Protection, and Counter-Terrorism; and Jeffrey Hunker, Director of the Critical Infrastructure Assurance Office*. Retrieved from <http://www.fas.org/irp/news/1998/05/980522-wh3.htm>
- Deelman, E., Singh, G., Su, M.-H., Blythe, J., Gil, Y., Kesselman, C., ... Katz, D. S. (2005). Pegasus: A framework for mapping complex scientific workflows onto distributed systems. *Sci. Program.*, 13(3), 219–237. Retrieved from <http://dl.acm.org/citation.cfm?id=1239653>

- Dreher, P., Agarwala, V., Ahalt, S. C., Almes, G., Fratkin, S., Hauser, T., ... Stewart, C. A. (2009). Developing a Coherent Cyberinfrastructure from Local Campuses to National Facilities: Challenges and Strategies. *EDUCAUSE*. Retrieved from <http://www.educause.edu/Resources/DevelopingaCoherentCyberinfras/169441> or <http://hdl.handle.net/2022/5122>
- ESnet. (2016). *Why Science DMZ*. Retrieved June 30, 2016, from <https://fasterdata.es.net/science-dmz/motivation/>
- Foster, I. (2005). Globus Toolkit Version 4: Software for Service-Oriented Systems. In H. Jin, D. Reed, & W. Jiang (Eds.), *Network and Parallel Computing* (Vol. 3779, pp. 2–13). Springer Berlin Heidelberg. http://doi.org/doi:10.1007/11577188_2
- Foster, I., & Kesselman, C. (1998). *The Grid: Blueprint for a New Computing Infrastructure*. Morgan Kaufmann.
- Foster, I., & Kesselman, C. (2004). *The Grid 2, Second Edition: Blueprint for a New Computing Infrastructure*. Morgan Kauffman.
- Fox, G. (2006). Collaboration and Community Grids. In *International Symposium on Collaborative Technologies and Systems* (pp. 419–428). IEEE Computer Society. <http://doi.org/doi:10.1109/CTS.2006.24>
- Fox, G. C., Qui, J., Kamburugamuve, S., Jha, S., & Luckow, A. (2015). HPC-ABDS High Performance Computing Enhanced Apache Big Data Stack. *2015 15th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid)*, 1057–1066. <http://doi.org/doi:10.1109/CCGrid.2015.122>
- Freeman, P. A. (2007). Is Designing Cyberinfrastructure or, Even, Defining It Possible? *First Monday*, 12(6). doi:10.5210/fm.v12i6.1900
- Globus Online. (2013). *Globus Toolkit Homepage*. Retrieved from <http://www.globus.org/toolkit/>
- Hallock, B., Knepper, R., & Stewart, C. A. (2015). *Workshop Report: Campus Bridging: Reducing Obstacles on the Path to Big Answers, 2015. IEEE Cluster 2015*. Retrieved from <http://hdl.handle.net/2022/20538>
- Hayden, L., Fox, G., & Gogineni, P. (2007). Cyberinfrastructure for Remote Sensing of Ice Sheets. Madison, WI: TeraGrid. Retrieved from <http://cerser.ecsu.edu/citeam/teragrid07.pdf>
- Johnston, W. E., Gannon, D., & Nitzberg, B. (1999). Grids as production computing environments: the engineering aspects of NASA's Information Power Grid. *High Performance Distributed Computing, 1999. Proceedings. The Eighth International Symposium on*, 197–204. <http://doi.org/doi:10.1109/HPDC.1999.805298>
- Knepper, R., Link, M. R., & Standish, M. (2015). Big Data on Ice: The Forward Observer System for In-flight Synthetic Aperture Radar Processing. *Procedia Computer Science*, 51, 1504–1513. Retrieved from <http://hdl.handle.net/2022/20471>
- Korab, H., & Brown, M. D. (1995). Virtual Environments and Distributed Computing at SC'95: GII Testbed and HPC Challenge Applications on the I-WAY. In H. Korab & M. D. Brown (Eds.), *SUPERCOMPUTING '95*. New York, NY: ACM.
- Laney, D. (2001). 3D Data Management: Controlling Data Volume, Velocity, and Variety. *META Delta*. META Group. Retrieved from <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>
- Litzkow, M., Livny, M., & Mutka, M. (1988). Condor - A Hunter of Idle Workstations. *8th International Conference of Distributed Computing Systems*, 104–111.

- Livny, M., Avery, P., Pordes, R., Foster, I., & Lazzarini, A. (2006). *Sustaining and Extending the Open Science Grid: Science Innovation on a PetaScale Nationwide Facility*. National Science Foundation. Retrieved from <http://nsf.gov/award-search/showAward.do?AwardNumber=0621704>
- Ludäscher, B., Altintas, I., Berkley, C., Higgins, D., Jaeger, E., Jones, M., & Zhao, Y. et al. (2006). Scientific workflow management and the Kepler system. *Concurrency and Computation*, 18(10), 1039–1065. doi:10.1002/cpe.994
- Marru, S., Gunathilake, L., Herath, C., Tangchaisin, P., Pierce, M., & Mattmann, C., ... Weerawarana, S. (2011). Apache airavata: a framework for distributed applications and computational workflows. In *Proceedings of the 2011 ACM workshop on Gateway computing environments* (pp. 21–28). Seattle, WA: ACM. <http://doi.org/doi:10.1145/2110486.2110490>
- Marston, S., Li, Z., Bandyopadhyay, S., Zhang, J., & Ghalsasi, A. (2011). Cloud computing - The business perspective. *Decision Support Systems*, 51(1), 176–189. doi:10.1016/j.dss.2010.12.006
- Mell, P., & Grance, T. (2011). *The NIST Definition of Cloud Computing (SP 800-145). Recommendations of the National Institute of Standards and Technology*. Retrieved from <http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf>
- National Science Foundation. (2006). *Cyberinfrastructure: From Supercomputing to the TerraGrid*. Retrieved from http://www.nsf.gov/news/special_reports/cyber/fromsctotg.jsp
- NCSA. (2001). *CAVERN Users Society*. Retrieved November 30, 2015, from <http://cavernus.ncsa.illinois.edu/>
- NSF Advisory Committee for Cyberinfrastructure Task Force on Campus Bridging. (2011). *NSF Advisory Committee for Cyberinfrastructure Task Force on Campus Bridging Final Report*. Retrieved from http://www.nsf.gov/od/oci/taskforces/TaskForceReport_CampusBridging.pdf or <http://pti.iu.edu/campusbridging/>
- NSF Advisory Committee for Cyberinfrastructure Task Force on Data and Visualization. (2011). *NSF Advisory Committee for Cyberinfrastructure Task Force on Data and Visualization Final Report*. Retrieved from http://www.nsf.gov/od/oci/taskforces/TaskForceReport_Data.pdf
- Obama, B. (2015). *Executive Order -- Creating a National Strategic Computing Initiative*. Retrieved from <https://www.whitehouse.gov/the-press-office/2015/07/29/executive-order-creating-national-strategic-computing-initiative>
- Open Science Grid. (2015). *About the Open Science Grid*. Retrieved November 30, 2015, from <http://www.opensciencegrid.org/>
- Sherman, W. R., O'Leary, P., Whiting, E. T., Grover, S., & Wernert, E. A. (2010). IQ-station: A low cost portable immersive environment. *Lecture Notes in Computer Science*, 6454, 361–372. http://doi.org/doi:10.1007/978-3-642-17274-8_36
- Stewart, C. A., Cockerill, T. M., Foster, I., Hancock, D., Merchant, N., & Skidmore, E., ... Gaffney, N. (2015). Jetstream - A self-provisioned, scalable science and engineering cloud environment. In *Proceedings of the 2015 XSEDE Conference: Scientific Advancements Enabled by Enhanced Cyberinfrastructure*. <http://doi.org/doi:10.1145/2792745.2792774>
- Stewart, C. A., Roskies, R., Knepper, R., Moore, R. L., Whitt, J., & Cockerill, T. M. (2015). XSEDE Value Added, Cost Avoidance, and Return on Investment. In *Proceedings of the 2015 XSEDE Conference: Scientific Advancements Enabled by Enhanced Cyberinfrastructure* (pp. 23:1–23:8). New York, NY: ACM. <http://doi.org/doi:10.1145/2792745.2792768>
- Stewart, C. A., Simms, S., Plale, B., Link, M., Hancock, D., & Fox, G. (2010). What is Cyberinfrastructure? SIGUCCS 2010. doi:10.1145/1878335.1878347

Tabor Griffin Communications. (1998). *High-Performance Computing Contributions to Society*. Tabor Griffin Communications. Retrieved from <http://www.tgc.com/hpcbook>

Thain, D., Tannenbaum, T., & Livny, M. (2005). Distributed computing in practice: The Condor experience. *Concurrency and Computation*, 17(2-4), 323–356. <http://doi.org/doi:10.1002/cpe.938>

The Apache Software Foundation. (2006). *Hadoop Distributed File System*. Retrieved May 6, 2014, from <http://hadoop.apache.org/>

Top500.Org. (2016). *The List*. Retrieved June 30, 2016, from <https://www.top500.org/>

Towns, J., Cockerill, T., Dahan, M., Foster, I., Gaither, K., Grimshaw, A., ... Wilkins-Diehr, N. (2014). XSEDE: Accelerating Scientific Discovery. *Comput. Sci. Eng.*, 16, 62. doi:10.1109/MCSE.2014.80

U.S. Congress Office of Technology Assessment. (1993). *Advanced Network Technology--Background Paper*. Retrieved from <https://www.princeton.edu/~ota/disk1/1993/9304/9304.PDF>

XSEDE. (2013). *2013 Nobel Prize in Chemistry winners bring HPC to the lab*. Retrieved November 30, 2015, from <https://www.xsede.org/2013-nobel-prize-in-chemistry>

XSEDE. (2015). *Overview*. Retrieved November 30, 2015, from <https://www.xsede.org/overview>

KEY TERMS AND DEFINITIONS

Campus Bridging: The seamlessly integrated use of cyberinfrastructure operated with other local or remote cyberinfrastructure as if they were proximate to the user.

Citizen Science: The work of individuals or teams of amateur, non-professional, or volunteer scientists who conduct research, gather and analyze data, perform pattern recognition, and develop technology, often in support of professional scientists.

Cloud Computing: On-demand, affordable access to a distributed, shared pool of computing and storage resources, applications, and services usually via the Internet for a large number of users.

Cyberinfrastructure: Computational systems, data and information management, advanced instruments, visualization environments, and people, all linked together by software and advanced networks to improve scholarly productivity and enable knowledge breakthroughs and discoveries not otherwise possible.

eScience: Computationally intensive science carried out through distributed global collaborations enabled by the Internet, involving access to large data collections, very large scale computing resources and high performance visualization.

High Performance Computing: Many tightly integrated computer processors that run very large scale computations and data analyses quickly where communication among the many processors is required.

High Throughput Computing: A computing paradigm that focuses on the efficient execution of a large number of loosely-coupled tasks

Science Gateways: Community-developed tools, applications, and data integrated via a portal or a suite of applications, usually in a graphical user interface, and customized to the needs of specific communities.

Fault Tolerant Cloud Systems

C

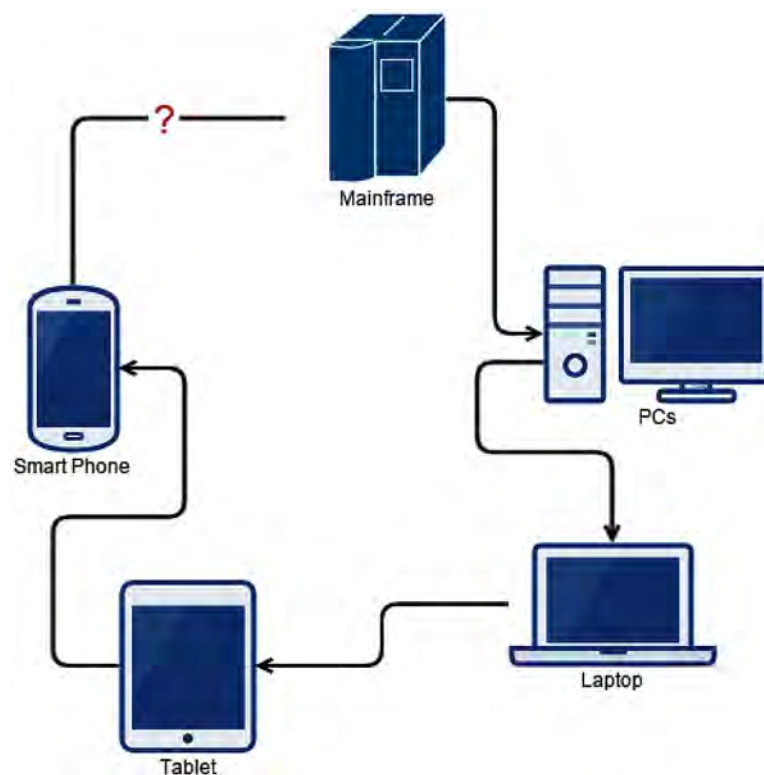
Sathish Kumar*VIT University, India***Balamurugan B***VIT University, India*

INTRODUCTION

Computing is a study of algorithms, automation, programming the information. Programming is a way of designing algorithms which are aimed at controlling, executing the computing devices. These devices have the basic features such as the amount of data they can store and process speed to perform in a reliable time. Traditionally in

1980's desktop personal computers (PCs) are used to support in creating, editing and manipulating documents. Further, these PCs are connected to the devices like a scanner to scan the documents, printer to take hard copies of the documents, etc. Later these devices are connected together to form a simple network. Since PCs has more of devices and it occupies more space the devices like laptop, tablet, mobile phone came into the context.

Figure 1. Sample computing paradigm shift



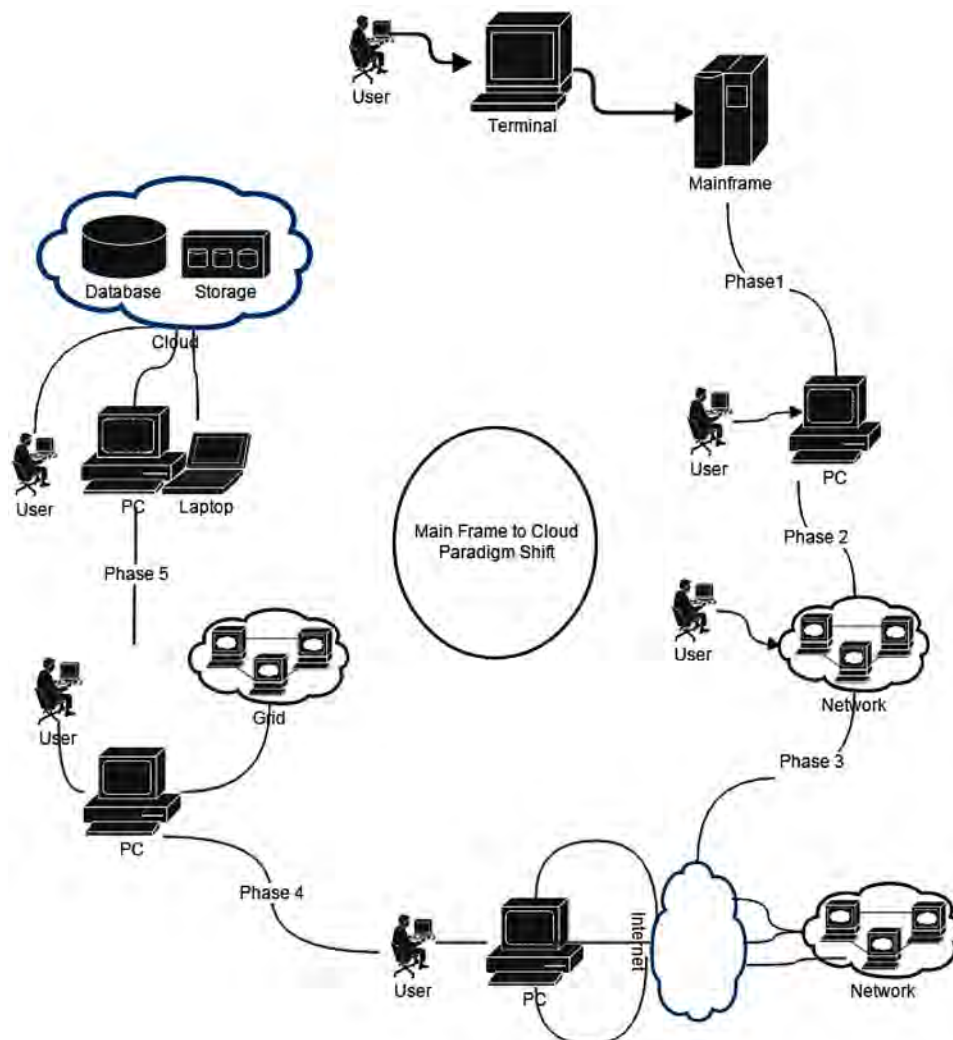
BACKGROUND

Computing Shift from Mainframe to Cloud

There are five distinct stages that cloud computing arrived. Initially one computer terminals like keyboard monitor to access the mainframes systems. In stage1, personal computers (PCs) were used to manipulate user requirements. In stage2, several PCs were connected to form a network called local

network and user can access the PCs from their own PCs. In stage3, several local networks were connected to a global network called the internet. From the internet, the users can remotely access the systems. In stage4, the grid computing came into the context where resources were shared distributedly. The user uses PCs to access the grid. In stage5, the user employs a computing technique called cloud computing that allows users to access the resources through the internet.

Figure 2. Mainframe to cloud shift



COMPUTING TECHNIQUES ERA

Cluster Computing

A cluster computing consists of several stand-alone computers which are a distributed loosely or tightly connected system and performs several tasks which are viewed as a single system. The features of cluster computing are reducing cost, power; it uses improved network technology, availability, and scalability.

Service-Oriented Architecture (SOA)

Service-oriented architecture is a loosely coupled distributed system that follows standard protocols to provide services over a network. The aim of service-oriented architecture is to divide the problems into separate distinguishable sections and emphasizes in a single software. The feature of the service-oriented architecture is it is independent of any product or technology.

Grid Computing

Grid computing is the technology that came into the contrast of electrical power grid, which allows the resource to be shared independently between power grids. Grid computing is a loosely coupled distributed system connected over a network. Grid computers are geographically distributed and heterogeneous in nature. The feature of grid computing is it improves scalability and performance of information system.

Utility Computing

Utility computing is a technology makes computing resource available as a service and provides a resource on demand and charges them depends on the usage. Utility computing used some form of virtualization like storage virtualization and hardware virtualization. The feature of utility computing is in reducing the cost of buying a resource in rental.

Cloud Computing

Cloud computing is an enabling technology that provides access to computer resource on-demand. Cloud computing follows model pay-as-you-go models. The important properties of cloud computing are on-demand self service, broad network access, resource pooling, elasticity, and measured service.

TYPICAL COMPUTING RESOURCE

The devices that are connected physically or virtually, internally or externally to the computer systems are called resources. Basically, computing resource can be classified into two categories namely physical resource and logical resource.

Physical Resource

Physical resources are the devices that are connected physically to the system. For example, a personal computer has several components connected together like, a monitor used to display the video outputs, a keyboard to give inputs, a secondary memory disk to store the inputs and outputs and a primary memory used for processing. Henceforth we can classify physical resource as follows, central processing unit, memory, storage, workstations, network elements, sensors.

Central Processing Unit (CPU)

Cpu is one of the most important units in the computer system where most of the processing were done. In the context to cloud computing utilization of cpu is considered. That is the amount of task processed by the cpu.

Memory

In early days, memory is managed statically. Since cloud computing is dynamic there is a need to fill dynamic memory allocation techniques for

cloud computing. Basically, memory is a block to store arrived data or task based on some policies. Memory can be allocated in three ways namely, schematic memory allocation technique: were memory allocation follows policies like FIFO, LRO. Static memory allocations technique: were memory block are divided into two parts and each part is partitioned using some static ratios. Dynamic memory allocation technique: were memory block are divided into two parts and each part is partitioned dynamically.

Storage

Cloud is models were resources are accessed remotely; the data's are stored in the remote database. These storage are maintained redundantly to ensure the client that they can access the data on demand. By this way, reliability is achieved in the storage of cloud and cloud maintains this storage as service.

Workstations

Workstations are the CPUs with high processing speed and have good internet connectivity so that the user can attain good services all the times. Usually, the workstation is automated without human interaction so that high performance is attained.

Network Elements

The components like routers, hubs, switches, bridges, etc are called network elements. To build an efficient data center to cloud switches like Top of Rack (TOR) switch, End of Rack (EOR) switch, and the virtual switch was used.

Logical Resources

Logical resources are used to control the physical resources so that we can use physical resource efficiently. Some of the logical resources are operating systems, network throughput, network loads, protocols, network delays etc.

Operating System

Operating systems help in using the available resource efficiently by managing hardware's, software's, files, devices, performance, and fault tolerance.

Network Throughput

The maximum number of bits per transferred per second is called throughput. Likewise, the maximum number of data bits transferred per second in network link is called network throughput. Higher the network throughput higher network efficiently.

Network Delay

A small difference in a second or even in millisecond made a major difference in reliability, scalability, and even in performance also. For example, a delay in virtual machine setup, virtual machine migration causes a delay in availability of virtual machines. So the minimal in delay can increase the system accuracy.

Network Protocols

The irregular communication between workloads leads to inability. So, some set of rules to be defined to overcome those inability. Protocols are kinematics of identifying the devices and connect them with each other. It also setup the rules for packing the data into a message and how to transfer these data's.

VIRTUALIZATION METHODOLOGY

Virtualization is a methodology used in cloud computing to determine an abstraction of software or hardware. Virtualization provides the resource that follows go-to technology. It helps in improving a cost effective performance by hiding the physical resource and communication between these resources and users.

Table 1. Evolution of virtualization

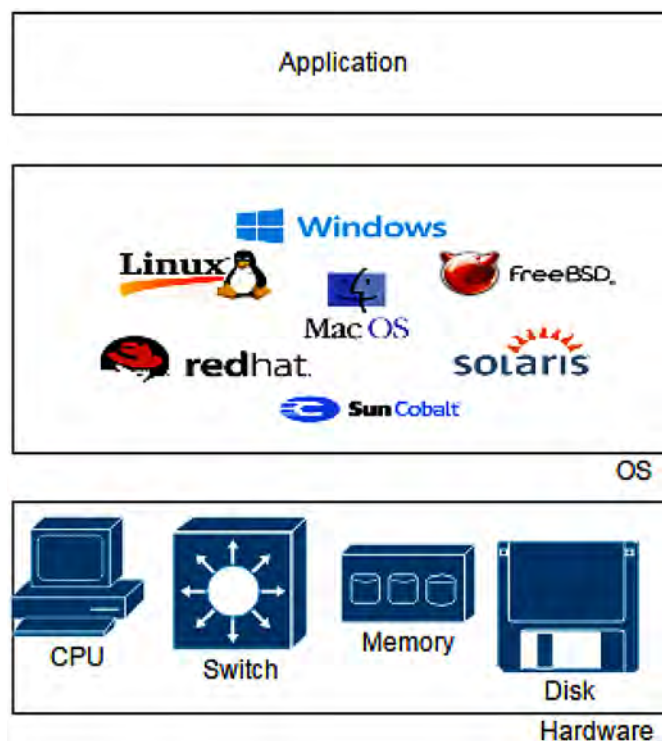
Technology	Period	Use
Virtual Memory	Late 1950	Developed for automatic page replacement for transistorized main frame computers.
Virtual Machine	In 1960	IBM developed a first virtual machine to run multiple operating systems on a single processor machine.
Virtual Machine with Time Sharing	Mid 1960 Mid 1970	Hardware virtualization is developed. The concept of virtual machine monitor is introduced to control this hardware virtualization. Every operating system accepted to use virtualization.
Hypervisor	In 1972	IBM developed first hypervisor VM-CP.

Evolution of Virtualization

The term virtual has introduced in late 1950, in transistorized mainframe system by creating a virtual memory for page replacement. In 1960 the first virtual machines are introduced by IBM. These virtual machines allow the user to run multiple operating systems on a single processor machine. In the mid of 1960 hardware virtualiza-

tion with time sharing is introduced. The concept of virtual machine monitor is developed during this time which is used to run a virtual machine in a protected environment. In the mid of 1970, every operating system accepted to use virtualization. In 1972 IBM has developed hypervisors called VM-CP which is used to control the virtual machines. VM-CP is the type I hypervisor which runs directly on hardware to control virtual machines.

Figure 3. Basic virtualization



Terminologies and Components of Virtualization

Host Machine

A host machine is a machine that includes the resources such as, CPU, memory, hard disk, network components that were used by virtual machines,

Virtual Machine (VM)

The virtual machine is a virtualized physical machine implemented in host machine that acts as real host machines. In the host machine, it is created as a single file or a single folder. It is controlled by a software called virtual software.

Host Operating System

Host operating system (OS) is the OS installed in the host system and manages resources such as memory, processing speed of the host machine. Finally, it assigns these resources to the virtual machines on demand.

Guest Operating System

Guest operating system (OS) is the OS installed on the virtualized setting. This guest OS may be installed in the client system, physical server etc. This guest OS is controlled by hypervisors.

Hypervisor

Hypervisor or virtual machine manager (VMM) is the software that runs on the virtual machine. Each OS in the host machine has its own memory, processing speed, allocation scheme etc. The responsibility for VMM is to control the allocation of resources and memory to the VMs. There are two groups of hypervisors namely: TYPE I hypervisor and TYPE II hypervisor. The TYPE I hypervisor so called native or bare-metal hypervisor runs directly on the hardware. It is responsible for al-

locating the resources between the host machine and VMs. It does not have a separate OS on it. TYPE II hypervisor so called hosted hypervisor has a separate OS called guest OS which is responsible for allocating the resources between the host machine and VMs.

TYPES OF VIRTUALIZATION

Basically there are two types of virtualization techniques followed. One is software-based virtualization and another one is hardware-assisted virtualization. In *software-based virtualization*: the VMM are maintained as software and the hardware interface is handed over to the guest OS. Now VMM has full control over Guest OS by writing dynamic binary translation code. It has two types of virtualization Full virtualization and para virtualization. In *hardware-assisted virtualization*: the VMM is directly installed on hardware so the dynamic binary translation is removed and the performance overhead during this dynamic binary translation is overcome.

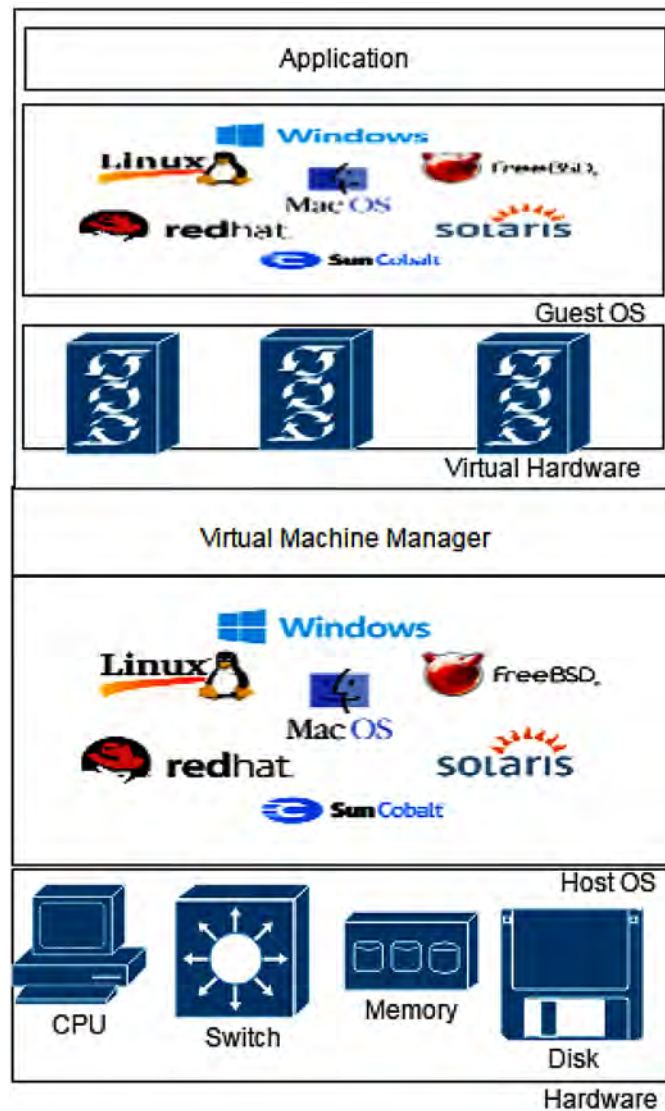
Emulation

Emulation is software created to replicate the hardware and process of a system. Guest OS need not be modified. The instructions from the guest OS should be interpreted by the host OS. This leads guest OS to slow down. So emulation suits for the systems where speed is not significant. The disadvantages of emulation are low performance and low density.

Full Virtualization/ Native Virtualization

Full virtualization is a hardware virtualization uses a hypervisor to translate the instruction to the host OS. This leads to a significant impact on system performance. Guest OS is not modified to make feel that it is running as host OS.

Figure 4. Full virtualization



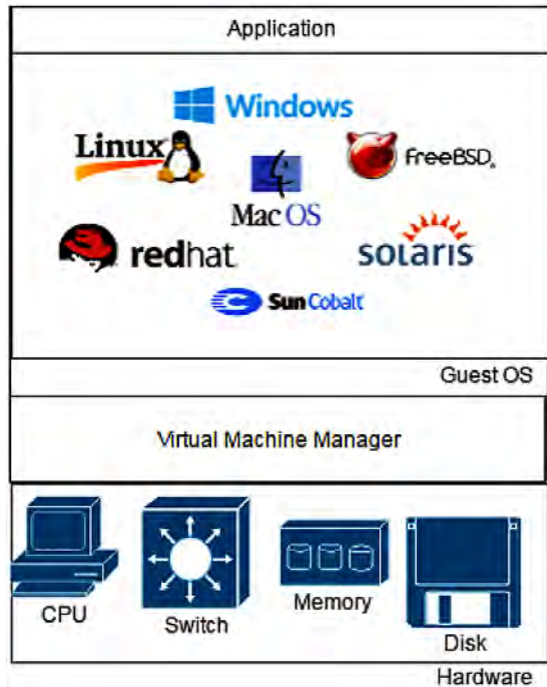
Para Virtualization

The guest OS is modified and it can be able to communicate directly with the hypervisor. This tends in the reduction of instruction translation time and operational cost. The OS available to the guest is limited that depends on the availability of the hypervisor.

Operating System Level Virtualization (OSLV)

OSLV is the method to provide multiple separate user-space instances for an operating system kernel. It allows the user-space application to run independently from other software.

Figure 5. Para virtualization



Resource Virtualization

The resources such as software resource (software, applications) and hardware resource (storage, network device etc.) are shared by the guest OS. Resource virtualization is classified into

- Storage Virtualization
- Server Virtualization
- Network Virtualization
- Desktop Virtualization
- Application Virtualization

Storage Virtualization

Storage virtualization is a technique that shares the multiple physical storage resources into a single storage and managed centrally. Storage virtualization generally implemented in file systems, tape systems, and storage area network.

Server Virtualization

Server virtualization is a technique that shares a single physical machine into multiple virtual machines (VM). Each VM have its own virtual memory, virtual interface, virtual cpu, and any one the functions such as mail, file, internet.

Network Virtualization

Network virtualization is a technique that changes the physical network into software called virtual network. The networking devices such as routers, switches, port, firewall etc. A virtual network will act same as a physical network. Network virtualization is the seamless step after storage and server virtualization. The network virtualization is grouped into two namely; External network virtualization: where multiple Local Area Networks were combined to form a Virtual Local Area Network (VLAN). Internal network virtualization: where hypervisors and containers combined to control the networks.

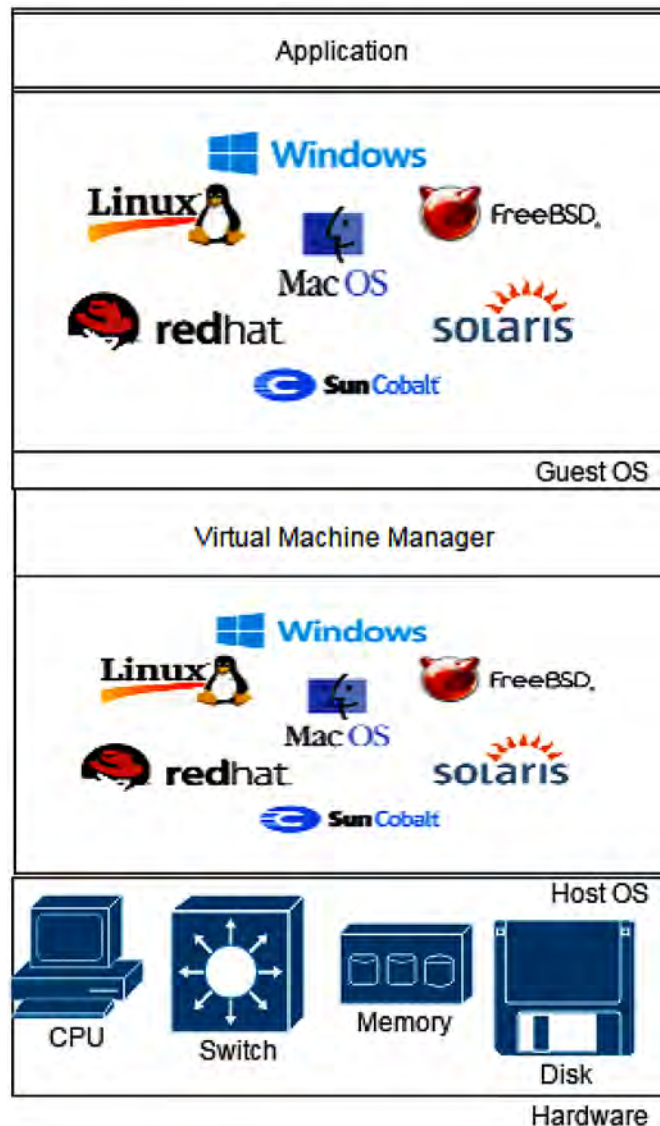
Desktop Virtualization

Desktop virtualization is a technique, on an existing desktop running OS it creates a separate OS. There are two man groups of desktop virtualization such as, Remote Desktop Virtualization also called as server-hosted virtualization where the operations are hosted on servers and accessed by the client over the network, Local desktop virtualization also called as client- hosted virtualization where local physical systems and hardware's are virtualized and monitored by the software called hypervisor.

Application Virtualization

Application virtualization is a technique that virtualizes particular application from the desktop

Figure 6. Operating system level virtualization



installed locally in the container. The container controls the interaction of the applications and components on how to communicate with other systems and components.

VIRTUALIZATION TOOLS

In the view of computing technology, a tool is a software use to process or to control a task or a request. Henceforth, virtualization tool is a software use to achieve virtualization for hardware

or any other resource in computer systems. The virtualization tools are broadly classified into the following modes: full virtualization, para virtualization, native virtualization, Operating system level virtualization and emulators. The availability of these tools can be commercial, open source and freely available. In full virtualization mode, many tools are available, most commonly used virtualization tool is based on VMware used to manage a virtual infrastructure. The other tools of VMware are VMware workstation and VMware Venter converter that runs on open source OS.

Table 2. Comparing features of virtualization

Type	Advantage	Disadvantage
Virtualizations	• Cost of hardware implementation and cost of energy is reduced. • Backup and recovery s faster and earlier. • More Efficient.	• Considering software license. • Experts in virtualization are less. • Maintaining the resource in virtualization environment with leasing plans is difficult.
Operating system Virtualization	• Reduce overhead. • Increase performance. • Increase compatibility.	• Have one base operating system and multiple guest operating systems. • If base OS fails the entire guest OS also fails.
Hypervisors	• Guest OS can be modified or unmodified. • In an unmodified OS if any one crashed it does not affect entire OS. • Supports wide range of hardware.	• In a modified OS if any one OS fails all others guest OS also fails. • If base OS is rebooted all other guest OS also rebooted reduces availability of VM's.

VMware server is freely available tool that runs on both windows and Linux platforms. Microsoft virtual PC is a commercially available tool that only works on Microsoft platforms. In para virtualization mode xen is an open source tool used for virtual machine migration.usermode Linux (UML), Vserver, Linux container(LXC) are the open source tools available for linux hardwares. LXC is a container based virtualization tool. In native virtualization mode, QEMU is an open source tool used for heterogeneous hardware's and used as an emulator. A virtual box is a commercial tool used to control remote desktop protocols. Openvz is an open source operating system level tool used to partition the resource efficiently and it is a container based tool. Bochs is an open source emulator tool used to debug guest OS.

FAULT TOLERANCE

For a failure or fault system fault tolerance is a system that should identify the fault and recovers the failed system back without any damage.

Fault: A fault is an underlying defect of a system that leads to an error.

Error: An error is a faulty system state, which May leads to failure.

Failure: A failure is a system error that affects its functionality of a system.

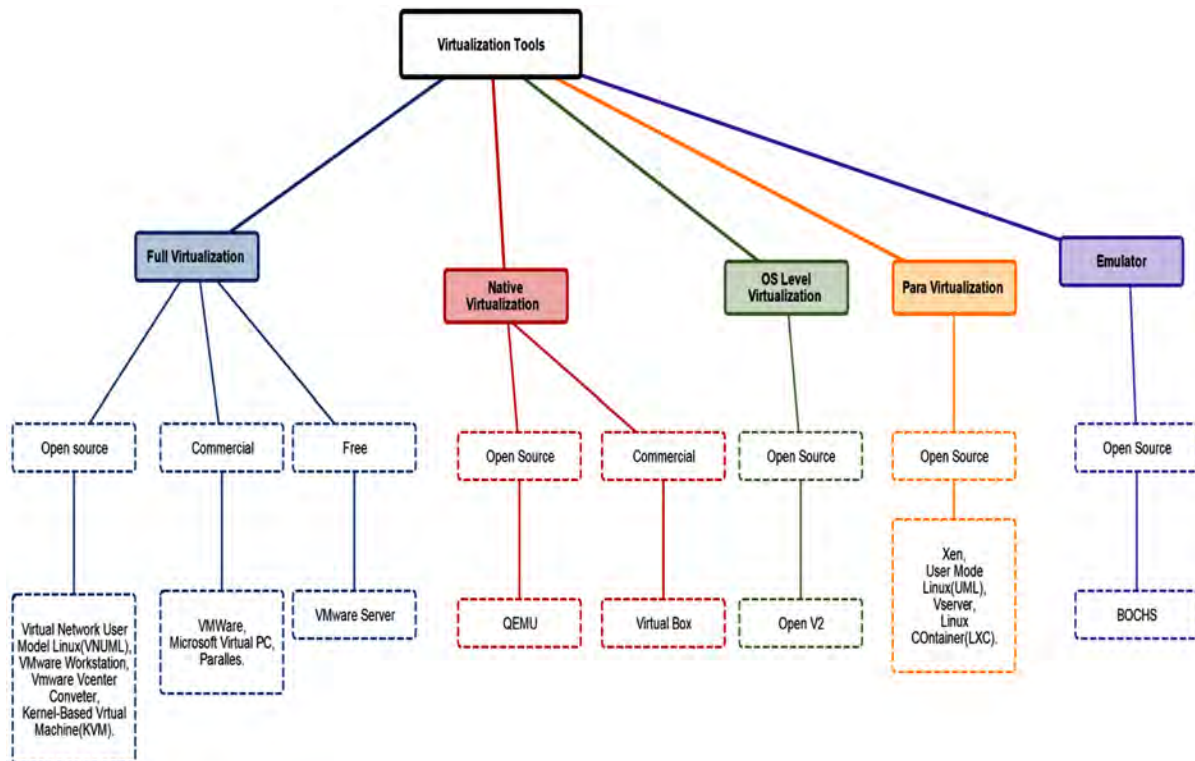
Fault Detection: Fault detection is a technique used to detect a fault or error of a system.

Fault Recovery: Fault recovery is a technique used to repair a fault or to repair a failed system after fault detection.

Table 3. Comparison of virtualization tools

Tool	Virtualization type	Host CPU	Guest CPU	Host OS	Guest OS	Live Migration	License
Xen	Para Virtualization	X86,X64	X86,X64	Linux, Unix	Windows, Linux, Unix	Yes	Open Source
KVM	Full Virtualization	X86,X64	X86,X64	Linux	Windows, Linux, Unix	Yes	Open Source
Virtual Box	Full Virtualization	X86,X64	X86,X64	Windows, Linux, Unix	Windows, Linux, Unix	Yes	Commercial
VMWare	Full Virtualization	X86,X64	X86,X64	Proprietary Unix	Windows, Linux, Unix	Yes	Open Source

Figure 7. Virtualization tools



FAULT TOLERANCE TECHNIQUES

There are numerous fault tolerance techniques followed to overcome faults. Fault tolerance can be classified based on redundancy, load-balancing, and policies.

Fault Tolerance Based on Redundancy

Redundancy is the method, where multiple copies are maintained so that availability of a resource is attained efficiently. Based on redundancy fault tolerance is classified as, hardware redundancy, software redundancy, time redundancy.

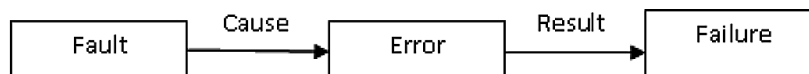
Hardware Redundant Fault Tolerant

Numerous redundant modules are created to perform the function the requirement. Later the outputs were ranked and if any fault occurred, errors were removed by this ranking.

Software Redundant Fault Tolerant

In software redundant fault tolerant two or more solutions of a problem are identified. In either way, two or more versions of software are also maintained so that failures will not occur for large inputs.

Figure 8. System failure



Time Redundant Fault Tolerant

To achieve an effective fault tolerant, time redundant fault tolerant uses a technique such as it execute the result repeatedly in different hardware and software and increase the accuracy. It also executes the result with increased time with the same hardware and software.

Fault Tolerance Based on Load Balancing

Load balancing is a technique used to balance the work load of a network in a peak traffic time. Based on load balancing it is categorized into hardware load balancing, network load balancing, and dispatcher software load balancing.

Hardware Load Balancing Fault Tolerant

Hardware based load balancing controls single host IP and sends to multiple hosts using network address translation. Backup load balancers were, used to handle the failure of single host IP.

Dispatcher Load Balancing Fault Tolerant

Dispatcher load balancing handle single point failure by transmitting dispatch function to another computer after failure. It maintains a separate server to control the entire incoming request.

Network Load Balancing Fault Tolerant

Network load balancing is distributed software used to create redundancy host and controls the entire request for an efficient load balancing.

Fault Tolerance Based on Policies

There are numerous fault tolerant techniques follows some kind of policies such as creating check points, migration of tasks, replicating the task

or workflow, resubmission of the task, retrying the failed job to recovers etc. There is two kinds of fault tolerance techniques categorized based on policies such as proactive fault tolerance and reactive fault tolerance.

Reactive Fault Tolerance Technique

When a failure occurs, reactive fault tolerance follows some kind of policies and helps to recover from failed state. The following or the policies were reactive fault tolerance follows,

Check Point/Restart

The process of saving the system states periodically when a system is in the failure-free state is called checkpoint. If a failure occurred during the execution the system is restarted from the last saved check point and it is roll backed or recovered to the fail free state.

Replication

Replication is the process used to store the most commonly used data to multiple locations and ensures the availability of data when needed. The data are accessed from nearest locations when a failure occurred.

Proactive Fault Tolerance Technique

Proactive fault tolerance techniques used to predict the fault in prior state and either replaces the components that are to fail or transfer the component that is predicted to be failed to failure free component. Some of the proactive fault tolerance policies are,

1. **Software Rejuvenation:** It is used to periodically reboot the system to ensure the system is in fail free state,
2. **Self-Healing:** It is a technique used to automatically handle the failure state and

Table 4. Tools for fault tolerant technique

Fault Tolerant Policy	Tool	Technique
Reactive	HAProxy, Hadoop, Amazon E2	Replication
Reactive	HAProxy	Job migration
Reactive	SGaurd	Checkpoint/restart
Reactive	Scientific Work Flow Systems	Task Resubmission
Reactive/Proactive	Assure	Retry, Self-Healing
Reactive	Adaptive Fault Tolerance (AFTRC)	Checkpoint/restart
Reactive	Low Latency Fault Tolerance (LLFT)	Replication
Reactive	FTWS	Replication/Resubmission
Reactive	FTM	Replication
Reactive/Proactive	CANDY	Preemptive migration
Reactive	Astrolabe	Checkpoint/restart
Reactive	VisPerf	Replication
Reactive	GEMS	Checkpoint/restart
Proactive	Cloudinit.d	Software Rejuvenation

3. **Preemptive Migration:** Used to transfer the failed component based on feedback from clients.

MTTR: Mean time to repair.

Availability: is defined as the availability of a system when it is needed without any fail and able to complete the opted task. Availability of a system is given by,

$$\text{Availability (A)} = \text{MTBF} / \text{MTBF} + \text{MTTR}.$$

Reliability at time 't' is given by the probability of time 'T' to failure after time 't'.

$$R(t) = P\{T > t\}.$$

DEPENDABILITY

Dependability used to define the reliability of the system in offering a service. Dependability generally studies the following metrics: reliability, availability, confidentiality, integrity, security.

Reliability: is defined as the probability of the system performs without any variations in the system for a specific period of time. Reliability is given by,

$$\text{Reliability function} = n(t) / N$$

= Failure free elements/number of elements at time zero.

Reliability is also given by,

$$\text{Reliability } R = \text{Availability} / \text{MTBF} + \text{MTTR}.$$

MTBF: Mean time between failures.

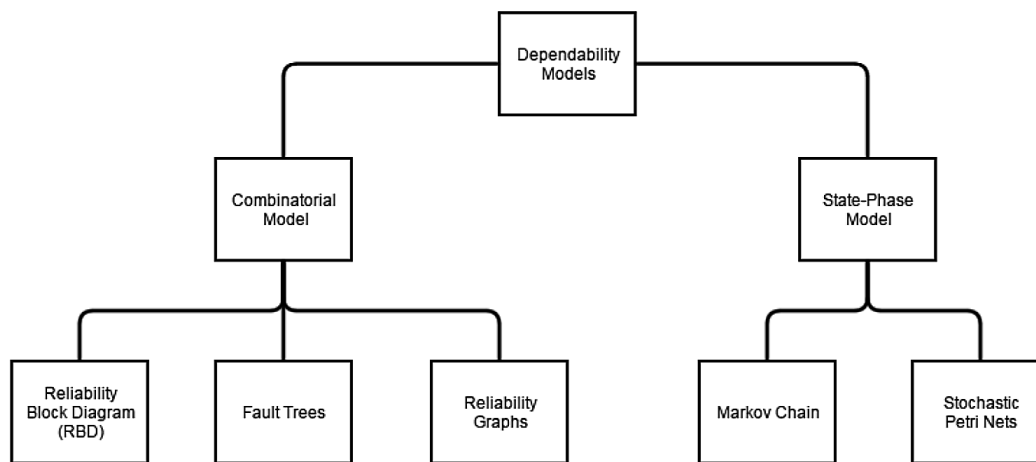
ESTIMATING DEPENDABILITY METRICS

In general to estimate dependability metrics two models are used. One is combinatorial model and another model is a state-phase model. *Combinatorial model:* used to study the system interaction that leads to system failure in terms of the structural relation between them. *State-phase model* considers the state and event of system dynamically. Following are the models used to represent both combinatorial and state-phase models,

Table 5. Comparison of fault tolerance models

Types	Fault Tolerant Properties				Fault Tolerant Techniques & Policies				
Fault Tolerance Model	Fault Tolerance Type	Reliability	Performance	Response Time	Self-Healing	Preemptive Migration	Checkpoint/Restart	Replication	Job Migration
LLFT	Reactive Fault Tolerance	High	High	Average	No	No	No	Yes	No
FTM	Reactive Fault Tolerance	High	Low	Average	No	No	Yes	Yes	Yes
FTMC	Reactive Fault Tolerance	High	Low	Average	No	No	Yes	Yes	No
AFRTC	Proactive Fault Tolerance	High	High	Average	No	No	Yes	Yes	Yes
FTWS	Reactive Fault Tolerance	Average	Average	Average	No	No	Yes	Yes	No
Map-Reduce	Proactive Fault Tolerance	High	High	Average	Yes	Yes	No	No	No

Figure 9. Dependability models



CONCLUSION

Formal definitions of computing, computing components, virtualization, fault, error, fault tolerance, fault tolerant model are discussed. Also tools for virtualization and fault tolerance techniques are discussed. A mathematical representation of fault tolerance is modeled.

In this chapter, fault tolerant techniques for the virtual machine were discussed. Looking at the performance metrics and features of application

the fault tolerant techniques discussed here helps in improving fault tolerance.

Future Trends

Considering the limitations and problems, following problems are identified.

1. The efficiency of fault tolerant technique in the real cloud is to be evaluated.

2. A fault tolerant technique for performance overhead in the large heterogeneous cloud is modeled.
3. A complete fault tolerant technique is developed and implemented as one the service in the cloud.
4. Based on the application type an optimum fault tolerant technique is to be developed.
5. If a number of component increases it increases the probability of failure. So an efficient fault tolerance technique is modeled.

In the context of data transmission if a large amount of data is transmitted for a long distance it increases the probability of failure. Here, an efficient fault tolerance technique is modeled.

REFERENCES

- Bala, A., & Chana, I. (2012). Fault tolerance-challenges, techniques and implementation in cloud computing. *IJCSI International Journal of Computer Science Issues*, 9(1), 1694–0814.
- Bressoud, T. C., & Schneider, F. B. (1996). Hypervisor-based fault tolerance. *ACM Transactions on Computer Systems*, 14(1), 80–107. doi:10.1145/225535.225538
- Chen, P. M., & Noble, B. D. (2001, May). When virtual is better than real [operating system relocation to virtual machines]. In *Hot Topics in Operating Systems, 2001. Proceedings of the Eighth Workshop on* (pp. 133-138). IEEE.
- Cheraghrou, M. N., Khadem-Zadeh, A., & Haghparast, M. (2016). A survey of fault tolerance architecture in cloud computing. *Journal of Network and Computer Applications*, 61, 81–92. doi:10.1016/j.jnca.2015.10.004
- Chouikhi, S., El Korbi, I., Ghamri-Doudane, Y., & Saidane, L. A. (2015). A survey on fault tolerance in small and large scale wireless sensor networks. *Computer Communications*, 69, 22–37. doi:10.1016/j.comcom.2015.05.007
- Dijkstra, E. W. (1974). Programming as a discipline of mathematical nature. *The American Mathematical Monthly*, 81(6), 608–612. doi:10.2307/2319209
- Egwutuoha, I. P., Levy, D., Selic, B., & Chen, S. (2013). A survey of fault tolerance mechanisms and checkpoint/restart implementations for high performance computing systems. *The Journal of Supercomputing*, 65(3), 1302–1326. doi:10.1007/s11227-013-0884-0
- Frohlich, D., Thomas, P., Hawley, M., & Hirade, K. (1997). Inaugural issue editorial: Future personal computing. *Personal Technologies*, 1(1), 1–5. doi:10.1007/BF01317881
- García-Valls, M., Cucinotta, T., & Lu, C. (2014). Challenges in real-time virtualization and predictable cloud computing. *Journal of Systems Architecture*, 60(9), 726–740. doi:10.1016/j.sysarc.2014.07.004
- Lira, V., Tavares, E., Fernandes, S., & Maciel, P. (2015). Dependable virtual network mapping. *Computing*, 97(5), 459–481. doi:10.1007/s00607-014-0431-8
- Manvi, S. S., & Shyam, G. K. (2014). Resource management for Infrastructure as a Service (IaaS) in cloud computing: A survey. *Journal of Network and Computer Applications*, 41, 424–440. doi:10.1016/j.jnca.2013.10.004
- Masood, A., Sharif, M., Yasmin, M., & Raza, M. (2015). Virtualization tools and techniques: Survey. *Nepal Journal of Science and Technology*, 15(2), 141–150. doi:10.3126/njst.v15i2.12131
- Padhy, R. P. (2012). Virtualization techniques & technologies: state-of-the-art. *Journal of Global Research in Computer Science*, 2(12), 29–43.
- Patra, P. K., Singh, H., & Singh, G. (2013). Fault tolerance techniques and comparative implementation in cloud computing. *International Journal of Computers and Applications*, 64(14).

Qi, Z., Xiang, C., Ma, R., Li, J., & Wei, D. (n.d.). *ForenVisor: A Tool for Acquiring and Preserving Reliable Data in Cloud Live Forensics*. Academic Press.

Sahoo, J., Mohapatra, S., & Lath, R. (2010, April). Virtualization: A survey on concepts, taxonomy and associated security issues. In *Computer and Network Technology (ICCNT), 2010 Second International Conference on* (pp. 222-226). IEEE. doi:10.1109/ICCNT.2010.49

Sandholm, T., & Lee, D. (2014). Notes on Cloud computing principles. *Journal of Cloud Computing*, 3(1), 1–10.

Singh, G., & Kinger, S. (2013, June). A survey on fault tolerance techniques and methods in cloud computing. *International Journal of Engineering Research and Technology*, 2(6).

Singh, H. (n.d.). *Cloud Computing Paradigms and Challenges*. Academic Press.

Stanković, R., Štula, M., & Maras, J. (2015). Evaluating fault tolerance approaches in multi-agent systems. *Autonomous Agents and Multi-Agent Systems*, 1–27.

Sun, D., Chang, G., Miao, C., & Wang, X. (2013). Analyzing, modeling and evaluating dynamic adaptive fault tolerance strategies in cloud computing environments. *The Journal of Supercomputing*, 66(1), 193–228. doi:10.1007/s11227-013-0898-7

Tedre, M. (2011). Computing as a science: A survey of competing viewpoints. *Minds and Machines*, 21(3), 361–387. doi:10.1007/s11023-011-9240-4

White, J., & Pilbeam, A. (2010). *A survey of virtualization technologies with performance testing*. arXiv preprint arXiv:1010.3233

ADDITIONAL READING

Benmessaoud, N., Tulloch, M., Williams, C., & Mudigonda, U. M. (2014). *Microsoft System Center Network Virtualization and Cloud Computing*. Pearson Education.

Jaramillo, D., Furht, B., & Agarwal, A. (2014). Mobile Virtualization Case Study. In *Virtualization Techniques for Mobile Systems* (pp. 27–36). Springer International Publishing. doi:10.1007/978-3-319-05741-5_4

Ruest, N., & Ruest, D. (2009). *Virtualization, A Beginner's Guide*. McGraw-Hill, Inc.

Xing, Y., & Zhan, Y. (2012). Virtualization and cloud computing. In *Future Wireless Networks and Information Systems* (pp. 305–312). Springer Berlin Heidelberg. doi:10.1007/978-3-642-27323-0_39

KEY TERMS AND DEFINITIONS

Cloud Customer: A cloud customer is who consumes the services provided by the cloud provider.

Cloud Models: Cloud is basically deployed in three models as public, private and hybrid.

Cloud Provider: Cloud providers are who provides the services to the customer who needs the cloud services.

Cloud Services: Cloud provides their resources as services to their customers. So the customers pay for the services and use them.

Virtual Machine Migration: The process of migrating, the virtual machine from one location to another is called virtual machine migration.

Work Load: The number of resources consumed and shared by the users for a period of time.

Fault Tolerant Data Management for Cloud Services



Wenbing Zhao

Cleveland State University, USA

INTRODUCTION

The pervasiveness of cloud services has significantly increased the dependability requirement of cloud systems. Most cloud services are implemented according to a three-tier architecture where the presentation, application logic execution, and data management are separately handled by each tier (Zhao, Moser, and Melliar-Smith, 2005). The middle-tier servers implement the application logic and they are designed to be either stateless or to only maintain session state. Hence, this separation of concerns has greatly increased the scalability of such systems because the middle-tier servers can be easily scaled out. On the other hand, this design makes the data management tier ever more important because the availability and integrity of the services hinges on the data management tier. The data must be made highly available and protected against various hardware faults and malicious attacks.

While it is relatively straightforward to ensure high availability for Web servers and application servers by simply running multiple copies according to the three-tier architecture, it is not so for a database management system, which has abundant state. The subject of highly available database systems has been studied for more than two decades and there exist many alternative solutions (Agrawal et al., 1997; Cecchet, Candea, & Ailamaki 2008; Drake et al., 2005; Garcia, Rodrigues, & Prego, 2011; Kemme, & Alonso, 2000; Patino-Martinez et al., 2005). In this article, we provide an overview of two most popular database high availability strategies, namely database replication and database clustering. The emphasis is given to those

that have been adopted and implemented by major database management systems (Banker 2011; Davies & Fisk, 2006; Vallath 2004).

BACKGROUND

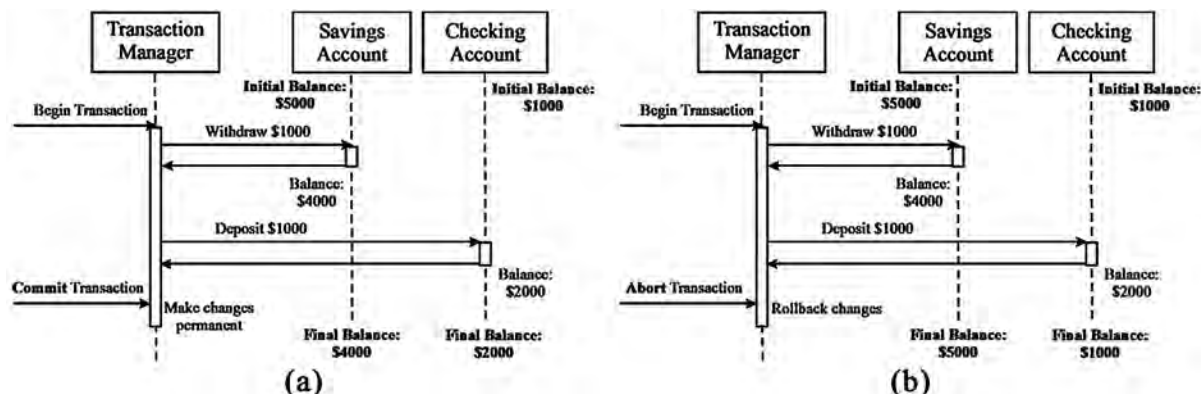
A database management system consists of a set of data and a number of processes that manage the data. These processes are often collectively referred to as database servers. The core programming model used in database management systems is called transaction processing. In this programming model, a group of read and write operations on the some data set are demarcated within a transaction. A transaction has the following ACID properties (Gray & Reuter, 1993):

- **Atomicity:** All operations on the data set agree on the same outcome. Either all the operations succeed (the transaction commits), or none of them are (the transaction aborts).
- **Consistency:** If the database is consistent at the beginning of a transaction, then the database remains consistent after the transaction commits.
- **Isolation:** A transaction does not read or overwrite a data item that has been accessed by another concurrent transaction.
- **Durability:** The update to the data set becomes permanent once the transaction is committed.

An example of an atomic transaction is shown in Figure 1. This transaction involves a debit

DOI: 10.4018/978-1-5225-2255-3.ch094

Figure 1. An example of atomic transactions



operation on a savings account and a credit operation on a checking account. If both operations are successful and the user decides to commit the transaction, the changes to both accounts are made permanent, as shown in Figure 1(a). On the other hand, if the user decides to abort the transaction, the changes to both accounts will be reversed so that the account balances of both accounts are restored to the values at the beginning of the transaction, as illustrated in Figure 1(b).

To support multiple concurrent users, a database management system uses sophisticated concurrency control algorithms to ensure the isolation of different transactions even if they access some shared data concurrently (Bernstein *et al.*, 1987). The strongest isolation can be achieved by imposing a serializable order on all conflicting read and write operations of a set of transactions so that the transactions appear to be executed sequentially. Two operations are said to be *conflicting* if both operations access the same data item and at least one of them is a write operation, and they belong to different transactions. Another popular isolation model is the snapshot isolation. Under the snapshot isolation model, a transaction performs its operations against a snapshot of the database taken at the start of the transaction. The transaction will be committed if the write operations do not conflict with any other transaction that has committed since the snapshot was taken. The snapshot

isolation model can provide better concurrent execution than the serializable isolation model.

A major challenge in database replication, the basic method to achieve high availability, is that it is not acceptable to reduce the concurrency levels. This is in sharp contrast to the replication requirement in some other field, which often assumes that the replicas are single-threaded and deterministic (Castro & Liskov, 2002).

To achieve high availability, a database system must try to maximize the time to operate correctly without a fault and minimize the time to recover from a fault. The transaction-processing model used in database management systems has some degree of fault tolerance in that a fault normally cannot corrupt the integrity of the database. If a fault occurs, all ongoing transactions will be aborted on recovery. However, the recovery time would be too long to satisfy the high availability requirement. To effectively minimize the recovery time, redundant hardware and software must be used. Many types of hardware fault can in fact be masked. For example, power failures can be masked by using redundant power supplies, and local communication system failures can be masked by using redundant network interface cards, cables and switches. Storage medium failures can be masked by using RAID (redundant array of inexpensive disks) or similar techniques.

To tolerate the failures of database servers, several server instances (instead of one) must be used so that if one fails, another instance can take over. The most common techniques are database replication and database clustering. These two techniques are not completely distinct from each other, however. Database replication is typically used to protect against total site failures. In database replication, two or more redundant database systems operate in different sites, ideally in different geographical regions, and communicate with each other using messages over a (possibly redundant) communication channel. Database clustering is used to provide high availability for a local site. There are two competing approaches in database clustering. One uses a shared-everything (also referred to as shared-disk) design, such as the Oracle Real Application Cluster (RAC) (Valath 2004). The other follows a shared-nothing strategy, such as the MySQL Cluster (Davies & Fisk, 2006) and most of DB2 shared database systems. To achieve maximum fault tolerance and hence high availability, one can combine database replication with database clustering.

DATABASE REPLICATION

Database replication means that there are two or more instances of database management systems, including server processes, data files and logs, running on different sites. Usually one of the replicas is designated as the primary and the rest of the replicas as backups. The primary accepts users' requests and propagates the changes to the database to the backups. In some systems, the backups are allowed to accept read-only queries. It is also possible to configure all replicas to handle users' requests directly. But doing so increases the complexity of concurrency control and the risk of more frequent transaction aborts.

Depending on how and when changes to the database are propagated across the replicas, there are two different database replication styles, often referred to as eager replication and lazy replica-

tion (Gray & Reuter, 1993). In eager replication, the changes (i.e., the redo log) are transferred to the backups synchronously before the commit of a transaction. In lazy replication, the changes are transferred asynchronously from the primary to the backups after the transactions have been committed. Because of the high communication cost, eager replication is rarely used to protect site failures where the primary and the backups are usually far apart. (Eager replication has been used in some shared-nothing database clusters.)

Eager Replication

To ensure strong replica consistency, the primary must propagate the changes to the backups within the boundary of a transaction. For this, a distributed commit protocol is needed to coordinate the commitment of each transaction across all replicas. The benefit for doing eager replication is that if the primary fails, a backup can take over instantly as soon as it detects the primary failure.

The most popular distributed commit protocol is the two-phase commit (2PC) protocol (Gray & Reuter, 1993). The 2PC protocol guarantees the atomicity of a transaction across all replicas in two phases. In the first phase, the primary (which serves as the coordinator for the protocol) sends a prepare request to all backups. If a backup can successfully log the changes, so that it can perform the update even in the presence of a fault, it responds with a "Yes" vote. If the primary collects "Yes" votes from all backups, it decides to commit the transaction. If it receives even a single "No" vote, or it times out a backup, the primary decides to abort the transaction. In the second phase, the primary notifies the backups of its decision. Each backup then either commits or aborts the transaction locally according to the primary's decision and sends an acknowledgment to the primary.

As can be seen, the 2PC protocol incurs significant communication overhead. There are also other problems such as the potential blocking if the primary fails after all backups have voted to

commit a transaction (Skeen, 1981). Consequently, there has been extensive research on alternative eager replication techniques, e.g., the epidemic protocols (Agrawal et al., 1997; Stanoi et al., 1998), and multicast-based approaches (Kemmer & Alonso, 2000; Patino-Martinez et al., 2005). However, they have not been adopted by any major commercial product due to their high overhead or complexities.

Lazy Replication

Most commercial database systems support lazy replication. In lazy replication, the primary commits a transaction immediately. The redo log, which reflects the changes made for the recently committed transactions, is transferred to backups asynchronously. Usually, the backup replicas lag behind the primary by a few transactions. This means that if the primary fails, the last several committed transactions might get lost.

Besides the primary/backup replication approach, some database management systems allow a multi-primary configuration where all replicas are allowed to accept update transactions. If this configuration is used with lazy replication, different replicas might make incompatible decisions, in which case, manual reconciliation is required.

DATABASE CLUSTERING

In recent several years, database clustering has evolved to be the most promising technique to achieve high availability as well as high scalability (Vallath 2004; Davies & Fisk, 2006). Database clustering, as the name suggests, uses a group of computers interconnected by a high speed network. In the cluster, multiple database server instances are deployed. If one instance fails, another instance takes over very quickly so high availability is ensured.

Database clustering not only brings high availability, but the scaling-out capability as well. Scaling-out means that the capacity of a database

management system can be dynamically increased by adding more inexpensive nodes while keeping the old equipment.

There are two alternative approaches in database clustering. In one approach, which is pioneered in Oracle RAC, adopts a shared-everything architecture. A number of other products choose to use the shared-nothing architecture. Both approaches have their challenges and advantages.

Shared Everything Cluster

In a shared-everything database cluster, all server instances share the same storage device, such as a storage area network. The cluster nodes typically connect to the shared storage device via a fiber channel switch or shared SCSI for fast disk I/O. The shared storage device must also have built-in redundancy such as mirrored disks to mask disk failures. To minimize disk I/O, all server instances share a common virtual cache space. The virtual cache space consists of local cache buffers owned by individual server instances. A number of background processes are used to maintain the consistency of the data blocks in the cache space. These processes are also responsible to synchronize the access to the cached data blocks because only one server instance is allowed to modify a data block at a time.

Each server instance has its own transaction logs stored in the shared disk. If a server instance fails, another server instance takes over by performing a roll-forward recovery using the redo log of the failed server instance. This is to ensure that the changes made by committed transactions are recorded in the database and not get lost. The recovery instance also rolls back the transactions that were active at the time of the failure and releases the locks on the resources used by those transactions.

The shared-everything design makes it unnecessary to repartition the data, and therefore, eases the tasks of cluster maintenance and management. However, this benefit does not come for free. The most prominent concern is the cost

of inter-node synchronization. Unless high-speed interconnect is used and the workload is properly distributed among the server instances, the inter-node synchronization might limit the scalability of the cluster. Also, the requirement for high-speed shared disk system also imposes higher financial cost than using conventional disks.

Shared Nothing Cluster

In a shared-nothing database cluster, each node runs one or more server instances and has its own memory space and stable storage. Essential to the shared-nothing approach, the data must be partitioned either manually or automatically by the database system across different nodes. Each partition must be replicated in two or more nodes to keep the desired redundancy level. Concurrency control and caching are carried out in each local node, and therefore, they are more efficient than those in shared-everything clusters. However, to ensure the consistency of replicated data and fast recovery, the two-phase commit protocol is often used to ensure atomic commitment of the transactions in the cluster. Comparing with the shared-everything approach, the cost of inter-node synchronization is essentially replaced by that of distributed commit.

The shared nothing approach faces the additional challenge of split-brain syndrome prevention (Birman, 2005). The split-brain syndrome may happen if the network partitions, and if each partition makes incompatible decisions on the outcome of transactions or their relative orders. To prevent this problem, typically only the main partition is allowed to survive. The minor partition must stop accepting new transaction and abort active transactions. Usually, the main partition is the one that consists of the majority of the replicas, or the one that contains a special node designated as the arbitration node (Davies & Fisk, 2006).

FUTURE TRENDS

Non-Traditional Database Systems

An interesting development in database systems is the popularity of NoSQL databases. NoSQL databases are designed to offer highly available storage services for large number of users by using weaker consistency models than those used in traditional relational database systems (Bartholomew 2010). Furthermore, such systems offer explicit user control regarding the placement of the replications. As such, NoSQL databases, such as Dynamo (DeCandia et al., 2007), MongoDB (Chodorow 2013), and Cassandra (Han et. al., 2011) are predominately used for cloud services (Bermbach et al., 2011).

In the following, we give a brief overview of the replica placement schemes and replica consistency levels offered by Cassandra. In Cassandra, three replica placement schemes are offered to users. In the default scheme (SimpleStrategy), nodes are organized into a logical ring and nodes that are next to each other on the ring are selected as replicas. In the NetworkTopologyStrategy scheme, users can choose the number of replicas from different datacenters. In the OldNetworkTopologyStrategy scheme, one replica is chosen from one datacenter, and the remaining replicas are chosen from another datacenter. Cassandra offers a variety of replica consistency levels to its users for both read and write operations. The strongest consistency level is ALL, which means the operation (read or write) would be applied to all replicas. For a read-ALL operation, it would fail if one or more replica are unavailable. A slightly weaker level is EACH_QUORUM, where the operation is applied to a quorum of replicas in every datacenter. The next weaker level is QUORUM, where the operation is applied to a quorum of replicas in any one of the datacenters. Another consistency level is

LOCAL_QUORUM, where a read/write operation would be applied to a quorum of replicas in the datacenter where the coordinator node resides. Furthermore, a user could choose to apply to one, two, or three replicas, as well as to use the SERIAL and LOCAL_SERIAL consistency levels to achieve linearizable consistency.

Beyond Crash Fault Tolerance

Existing database systems are designed to tolerate process crash fault and hardware fault. However, considering the increased pace of security breaches, future database management systems must be designed to be intrusion tolerant, i.e., they should provide high availability against a variety of security threats, such as the unauthorized deletion and alteration of database records, the disruption of distributed commit (may cause replica inconsistency), and the exposure of confidential information.

To make a database system intrusion tolerant, many fundamental protocols such as the 2PC protocol must be enhanced. There may also be a need to design special tamper-proof storage devices to protect data integrity (Strunk et al., 2000). Even though there have been intensive research in this area (Deswarte et al., 1991; Garcia, Rodrigues, & Prego, 2011; Mohan et al., 1983; Prez-Sorrosal et al., 2006; Zhang et al., 2012; Zhao, 2014), the results have rarely been incorporated into commercial products yet. The primary barrier is the high commutation and communication cost, the complexity, and the high degree of replication required to tolerate malicious fault.

CONCLUSION

Database systems are the corner stones of today's information systems. The availability of database systems largely determines the quality of service provided by the information systems. In this article, we provided a brief overview of the state of the art database replication and clustering techniques. For

many, a low-cost shared-nothing database cluster that uses conventional hardware might be a good starting point towards high availability. We envisage that future generation of database management systems will be intrusion tolerant, i.e., they are capable of continuous operation against not only hardware and process crash fault, but a variety of security threats as well.

REFERENCES

- Agrawal, D., El Abbadi, A., & Steinke, R. C. (1997). Epidemic algorithms in replicated databases. *Proceedings of the ACM Symposium on Principles of Database Systems*. Tucson, AZ: ACM Press.
- Banker, K. (2011). *MongoDB in action*. Manning Publications Co.
- Bartholomew, D. (2010). SQL vs. NoSQL. *Linux Journal*, 2010(195), 4.
- Bermbach, D., Klems, M., Tai, S., & Menzel, M. (2011, July). Metastorage: A federated cloud storage system to manage consistency-latency tradeoffs. In *Cloud Computing (CLOUD), 2011 IEEE International Conference on* (pp. 452-459). IEEE. doi:10.1109/CLOUD.2011.62
- Bernstein, P. A., Hadzilacos, V., & Goodman, N. (1987). *Concurrency Control and Recovery in Database Systems*. Reading, MA: Addison-Wesley.
- Birman, K. (2005). *Reliable Distributed Systems: Technologies, Web Services, and Applications*. Springer-Verlag.
- Cecchet, E., Candea, G., & Ailamaki, A. (2008, June). Middleware-based database replication: the gaps between theory and practice. In *Proceedings of the 2008 ACM SIGMOD international conference on Management of data* (pp. 739-752). ACM. doi:10.1145/1376616.1376691
- Chodorow, K. (2013). *MongoDB: the definitive guide*. O'Reilly.

- Davies, A. & Fisk, H. (2006). *MySQL Clustering*. MySQL Press.
- DeCandia, G., Hastorun, D., Jampani, M., Kaku-lapati, G., Lakshman, A., Pilchin, A.,... Vogels, W. (2007, October). Dynamo: Amazon's highly available key-value store. In *SOSP* (Vol. 7, pp. 205-220). Academic Press.
- Deswarte, Y., Blain, L., & Fabre, J. C. (1991). Intrusion tolerance in distributed computing systems. *Proceedings of the IEEE Symposium on Research in Security and Privacy*. Oakland, CA: IEEE Computer Society Press.
- Drake, S., Hu, W., McInnis, D. M., Sköld, M., Srivastava, A., Thalmann, L., & Wolski, A. (2005). Architecture of highly available databases. In *Service Availability* (pp. 1-16). Springer Berlin Heidelberg. doi:10.1007/978-3-540-30225-4_1
- Garcia, R., Rodrigues, R., & Preguiça, N. (2011, April). Efficient middleware for byzantine fault tolerant database replication. In *Proceedings of the sixth conference on Computer systems* (pp. 107-122). ACM. doi:10.1145/1966445.1966456
- Gray, J., & Reuter, A. (1993). *Transaction Processing: Concepts and Techniques*. San Mateo, CA: Morgan Kaufmann Publishers.
- Han, J., Haihong, E., Le, G., & Du, J. (2011, October). Survey on NoSQL database. In *Pervasive computing and applications (ICPCA), 2011 6th international conference on* (pp. 363-366). IEEE.
- Kemme, B., & Alonso, G. (2000). A new approach to developing and implementing eager database replication protocols. *ACM Transactions on Database Systems*, 25(3), 333-379. doi:10.1145/363951.363955
- Mohan, C., Strong, R., & Finkelstein, S. (1983). Method for distributed transaction commit and recovery using Byzantine agreement within clusters of processors. *Proceedings of the ACM Symposium on Principles of Distributed Computing*. Montreal, Quebec, Canada: ACM Press. doi:10.1145/800221.806712
- Patino-Martinez, M., Jimenez-Peris, R., Kemme, B., & Alonso, G. (2005). Middle-R: Consistent database replication at the middleware level. *ACM Transactions on Computer Systems*, 23(4), 375-423. doi:10.1145/1113574.1113576
- Prez-Sorrosal, F., Patino-Martinez, M., Jimenez-Peris, R., & Vuckovic, J. (2006). Highly available long running transactions and activities for J2EE applications. *IEEE International Conference on Distributed Computing Systems*. Lisboa, Portugal: IEEE Computer Society Press. doi:10.1109/ICDCS.2006.47
- Skeen, D. (1981). Nonblocking commit protocols. *Proceedings of the ACM International Conference on Management of Data*. Ann Arbor, MI: ACM Press.
- Stanoi, I., Agrawal, D., & El Abbadi, A. (1998). Using broadcast primitives in replicated databases. *Proceedings of the IEEE International Conference on Distributed Computing Systems*. Amsterdam, Netherlands: IEEE Computer Society Press.
- Strunk, D., Goodson, G., Scheinholtz, M., Soules, C., & Ganger, G. (2000). Self-securing storage: protecting data in compromised systems. *Proceedings of Symposium on Operating Systems Design and Implementation*. San Diego, CA: USENIX Association.
- Vallath, M. (2004). *Oracle real application clusters*. Elsevier.

Zhang, H., Chai, H., Zhao, W., Melliar-Smith, P. M., & Moser, L. E. (2012). Trustworthy coordination of Web services atomic transactions. *IEEE Transactions on Parallel and Distributed Systems*, 23(8), 1551–1565. doi:10.1109/TPDS.2011.292

Zhao, W. (2014). A Novel Approach to Building Intrusion Tolerant Systems. *International Journal of Performability Engineering*, 10(2), 123.

Zhao, W., Moser, L. E., & Melliar-Smith, P. M. (2005). Unification of transactions and replication in three-tier architectures based on CORBA. *IEEE Transactions on Dependable and Secure Computing*, 2(1), 20–33. doi:10.1109/TDSC.2005.14

ADDITIONAL READING

Abadi, D. (2012). Consistency tradeoffs in modern distributed database system design: CAP is only part of the story. *Computer*, 45(2), 37–42. doi:10.1109/MC.2012.33

Barber, R. J., Herbert, D. M., Mohan, C., Somani, A., Watts, S. J., & Zaharioudakis, M. (2001). U.S. Patent No. 6,173,292. Washington, DC: U.S. Patent and Trademark Office.

Birman, K. P., Freedman, D. A., Huang, Q., & Dowell, P. (2012). Overcoming cap with consistent soft-state replication. *Computer*, 45(2), 50–58. doi:10.1109/MC.2011.387

Blundell, C., Lewis, E. C., & Martin, M. (2005, June). Deconstructing transactional semantics: The subtleties of atomicity. In *Annual Workshop on Duplicating, Deconstructing, and Debunking (WDDD)* (pp. 48-55).

Borr, A. J. (1984, August). Robustness to Crash in a Distributed Database: A Non Shared-Memory Multi-Processor Approach. In *VLDB* (pp. 445-453).

Brewer, E. (2012). CAP twelve years later: How the. *Computer*, 45(2), 23–29. doi:10.1109/MC.2012.37

Castro, M., & Liskov, B. (2002). Practical Byzantine fault tolerance and proactive recovery. *ACM Transactions on Computer Systems*, 20(4), 398–461. doi:10.1145/571637.571640

Cattell, R. (2011). Scalable SQL and NoSQL data stores. *SIGMOD Record*, 39(4), 12–27. doi:10.1145/1978915.1978919

Chai, H., Zhang, H., Zhao, W., Melliar-Smith, P. M., & Moser, L. E. (2013). Toward trustworthy coordination of Web services business activities. *IEEE Transactions on Services Computing*, 6(2), 276–288. doi:10.1109/TSC.2011.57

Escriva, R., Wong, B., & Sirer, E. G. (2012). HyperDex: A distributed, searchable key-value store. *Computer Communication Review*, 42(4), 25–36. doi:10.1145/2377677.2377681

Gray, J., Helland, P., O’Neil, P., & Shasha, D. (1996, June). The dangers of replication and a solution. [J]. ACM.]. *SIGMOD Record*, 25(2), 173–182. doi:10.1145/235968.233330

Han, J., Song, M., & Song, J. (2011, May). A Novel Solution of Distributed Memory NoSQL Database for Cloud Computing. In *Computer and Information Science (ICIS), 2011 IEEE/ACIS 10th International Conference on* (pp. 351-355). IEEE. doi:10.1109/ICIS.2011.61

Lent, B., Swami, A., & Widom, J. (1997, April). Clustering association rules. In *Data Engineering, 1997. Proceedings. 13th International Conference on* (pp. 220-231). IEEE. doi:10.1109/ICDE.1997.581756

Levandowski, J. J., Lomet, D. B., Mokbel, M. F., & Zhao, K. (2011). Deuteronomy: Transaction Support for Cloud Data. In *CIDR* (pp. 123-133).

Lloyd, W., Freedman, M. J., Kaminsky, M., & Andersen, D. G. (2011, October). Don’t settle for eventual: scalable causal consistency for wide-area storage with COPS. In *Proceedings of the Twenty-Third ACM Symposium on Operating Systems Principles* (pp. 401-416). ACM. doi:10.1145/2043556.2043593

Malkhi, D., & Reiter, M. (1997). Byzantine quorum systems. *Proceedings of the ACM Symposium on Theory of Computing*, El Paso, Texas: ACM Press. 569 – 578.

Membrey, P., Plugge, E., & Hawkins, T. (2010). *The definitive guide to MongoDB: the noSQL database for cloud and desktop computing*. Apress.

Padhy, R. P., Patra, M. R., & Satapathy, S. C. (2011). RDBMS to NoSQL: Reviewing Some Next-Generation Non-Relational Databases. *International Journal of Advanced Engineering Science and Technologies*, 11(1), 15–30.

Ramakrishnan, R., & Gehrke, J. (2000). Database management systems. Osborne/McGraw-Hill.

Shapiro, M., Preguiça, N., Baquero, C., & Zawirski, M. (2011). Conflict-free replicated data types. In *Stabilization, Safety, and Security of Distributed Systems* (pp. 386–400). Springer Berlin Heidelberg. doi:10.1007/978-3-642-24550-3_29

Stonebraker, M. (2010). Errors in Database Systems, Eventual Consistency, and the CAP Theorem. *Communications of the ACM, BLOG@ACM*.

Thomson, A., Diamond, T., Weng, S. C., Ren, K., Shao, P., & Abadi, D. J. (2012, May). Calvin: fast distributed transactions for partitioned database systems. In *Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data* (pp. 1–12). ACM. doi:10.1145/2213836.2213838

Wada, H., Fekete, A., Zhao, L., Lee, K., & Liu, A. (2011, January). Data Consistency Properties and the Trade-offs in Commercial Cloud Storage: the Consumers' Perspective. In *CIDR* (Vol. 11, pp. 134–143).

Wiesmann, M., Pedone, F., Schiper, A., Kemme, B., & Alonso, G. (2000). Understanding replication in databases and distributed systems. In *Distributed Computing Systems, 2000. Proceedings. 20th International Conference on* (pp. 464–474). IEEE. doi:10.1109/ICDCS.2000.840959

Zhao, W. (2014). *Building dependable distributed systems*. John Wiley & Sons. doi:10.1002/9781118912744

Zhao, W. (2015). Optimistic Byzantine fault tolerance. *International Journal of Parallel, Emergent and Distributed Systems*, 1–14.

KEY TERMS AND DEFINITIONS

Database Cluster (Shared-Everything, Shared-Nothing): A database management system runs on a group of computers interconnected by a high speed network. In the cluster, multiple database server instances are deployed. If one instance fails, another instance takes over very quickly to ensure high availability. In the shared-everything design, all nodes can access a shared stable storage device. In the shared-nothing design, each node has its own cache buffer and stable storage.

Database Recovery (Roll-Backward, Roll-Forward): Recovery is needed when a database instance that has failed is restarted or a surviving database instance takes over a failed one. In roll-backward recovery, the active transactions at the time of failure are aborted and the resources allocated for those transactions are released. In roll-forward recovery, the updates recorded in the redo log are transferred to the database so that they are not lost.

Database Replication (Eager, Lazy): Multiple instances of a database management system are deployed in different computers (often located in different sites). Their state is synchronized closely to ensure replica consistency. In eager replication, the updates are propagated and applied to all replicas within the transaction boundary. In lazy replication, the changes are propagated from one replica to others asynchronously.

High Availability (HA): The capability of a system to operate with long uptime and to recover quickly if a failure occurs. Typically, a highly available system implies that its measured uptime is

five nine (99.999%) or better, which corresponds to 5.25 minutes of planned and unplanned down-time per year.

NoSQL Database: This is a database that adopts a data storage and retrieval mechanism that is different from that used in traditional relational databases. The actual mechanisms differ in different NoSQL databases, such as key-value stores, document store, and graph, etc.

Transaction: A transaction is a group of read/write operations on the some data set that succeed or fail atomically. More accurately, a transaction

has the atomicity, consistency, isolation and durability properties.

Two-Phase Commit (2PC) Protocol: This protocol ensures atomic commitment of a transaction that spans multiple nodes in two phases. During the first phase, the coordinator (often the primary replica) queries the prepare status of a transaction. If all participants agree to commit, the coordinator decides to commit. Otherwise, the transaction is aborted. The second phase is needed to propagate the decision to all participants.

From Information Systems Outsourcing to Cloud Computing

C

Mohammad Nabil Almunawar
Universiti Brunei Darussalam, Brunei

Hasan Jawwad Almunawar
P. T. Tegar Kupas Mediatama, Indonesia

INTRODUCTION

Outsourcing is a business term to describe a mechanism in which a company utilizes services provided by another company, normally through a contract, to fulfill some of its required business resources or functions. Outsourcing is commonly practiced by business organizations, as it is believed that it can cut costs and simplify management. For service providers, outsourcing gives them a long-term source of revenue.

Nowadays most business organizations outsource some part of their business operations. One of the most common is information systems (IS) outsourcing. This may range from computer maintenance, website development and maintenance, e-Business to the whole IS function (Dibbern, Goles, Hirschheim, & Jayatilaka, 2004). Actually, IS outsourcing is an old story which started as early as 1963 when Frito-Lay and Blue Cross & Blue Shield outsourced their data processing jobs to Electronic Data Systems (Lacity & Hirschheim, 1993). In fact, Eastman Kodak outsourced the whole of its IS functions to IBM, DEC and Businessland in 1989, 25 years ago (Gupta & Gupta, 1992). In the early stages of IS outsourcing, the issue being addressed in business organizations was whether they should outsource. Over time, the issue was no longer on whether to outsource or not to outsource, but how much to outsource (Lee, Huynh, Kwok, & Pi, 2003). This indicates that IS outsourcing has been adopted by many business organizations.

The advancement of Internet technology, especially the Web as well as high-speed and broadband access to the Internet, enabled a new computing model, “cloud” computing. The new model allows organizations to outsource some components or whole of their IS in the cloud that can be controlled and utilized from anywhere with a web browser. With this model organizations do not need to purchase hardware and expensive software licenses and surely they do not need to worry about software and hardware maintenance, which is normally a large portion of the total ownership costs of an IS to estimate the overall cost (direct and indirect) of an IS in a given time frame. Cloud computing vendors normally offer a pay-per-use method for their services, making cloud computing services like paying utilities. Perhaps cloud computing is the realization of McCarthy’s dream of utility computing, a package of computing resources that can be rented or subscribed just like other utilities (Garfinkel, 2011).

What makes the cloud computing system different from conventional computing systems? In conventional computing systems (mainframe, client-server or personal computer systems), most of the computing resources owned by an organization normally reside in the organization’s premises. The organization has to manage these resources to make sure they can be utilized to support the organization in attaining its goals. The organization incurs all costs in owning these resources, which may include investment, operation and maintenance costs. In contrast, an organization

does not need to own most of the computing resources in a cloud computing system. Instead, the organization utilizes computing resources offered by a provider and accesses the resources as needed. The organization only needs to own client devices (low cost terminals or thin clients) to utilize the computing resources through the Internet. Consequently, the organization does not need to bear the burden of all the costs mentioned previously. Of course, the organization needs to pay the provider for using the resources with a pay-per-use method of payment.

The numbers of providers offering various computing resources in the cloud are growing and some big players include IBM, Amazon.com, Google and Microsoft. These companies foresee a lucrative business in cloud computing as it offers a new business model that may attract many customers. There are three types of customers: small organizations, medium and large organizations, and consumers. However, there are some adoption issues that need to be addressed properly by providers (Kim, Kim, Lee, & Lee, 2009).

This chapter discusses concepts and applications of cloud computing. The history of the development as well as some related computing concepts such as grid computing will be highlighted. Advantages and disadvantages of cloud computing, including several issues, including adoption issues will be discussed. Future direction will be presented in the last part of this chapter. The next section will discuss the development of cloud computing, computing models and available services. Section 3 will focus more on core technology, business model and related issues, including some criticisms of cloud computing. Section 4 is the future direction and the last section (Section 5) is the conclusion.

BACKGROUND

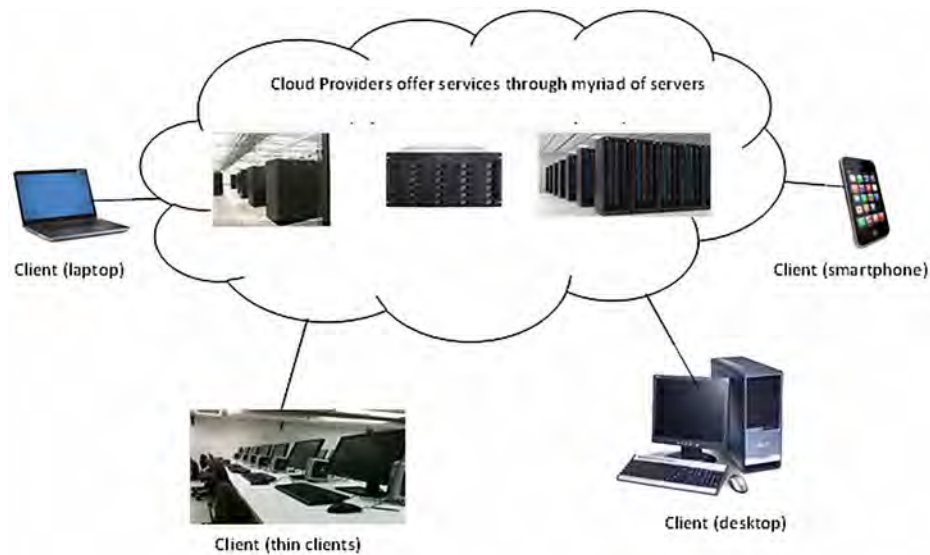
IS outsourcing has evolved to cloud computing where many business organizations see that it is a good option to outsource their computing re-

sources to cloud computing providers. There are many definitions of cloud computing. A study on definitions of cloud computing (Vaquero, Rodero-Merino, Caceres, & Lindner, 2009) found there are at least 20 definitions. This study summarizes three necessary components of cloud computing: a large pool of computing resources accessible through a computer network, dynamically and scalable resource allocations, and a pay-per-use method of payment. The National Institute of Standards and Technology (NIST) at the U.S. Department of Commerce provides a short definition for cloud computing: “Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g. networks, servers, storage, applications and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.” (Mell & Grance, 2011). Figure 1 illustrates a cloud computing models where clients’ machines access computing resources offered by cloud providers. The client machines can be desktops, laptops, smart phones, terminals or thin client machines that access the computing resources (normally computer servers) offered by cloud providers through the Internet or other networks.

There are many services that can be provided through cloud computing. In general, these services can be grouped into three service models: Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS). Organizations can outsource their IS through these three models. We will discuss these services with examples in the next section.

There are some pros and cons of cloud computing. However, many experts concur with McCarthy’s prediction: “Computing may someday be organized as a public utility just as the telephone system is a public utility” (Garfinkel, 2011). At the consumer level, people, especially mobile users access software and storage online as they are connected to the Internet most of the time. According to Pew Internet report on the future of cloud computing (Anderson & Rainie, 2010),

Figure 1. The cloud computing model



a solid majority of experts and stakeholders who participate in the survey on expected future of the Internet by 2020 agree that most people will access software applications and share information through the Internet using remote servers rather than local machines. Cloud computing will become dominant and will overtake desktops in the next decade.

We are witnessing smart phones overtake desktops and laptops at the consumer level. People use their smart phones to access the cloud, from sharing information through social networking sites (such as Facebook, LinkedIn, WhatsApp) to processing and sharing documents using Google Docs.

Information Systems Outsourcing and Cloud Computing

As highlighted in the introduction, IS outsourcing is not new. It started in the 60's and became a hot issue when Eastman Kodak outsourced the whole function of its IS. However, there were some successes and failures in implementing IS outsourcing and organizations need to carefully consider before outsourcing their IS to vendors (Lacity & Hirschheim, 1993).

Outsourcing is far more complex than is generally understood and IS outsourcing is even more complex. Many pros and cons have been identified and discussed in the literature (Gupta & Gupta, 1992; Palvia, 1995; Vitharana & Dharwadkar, 2007; Weidenbaum, 2005). Gonzalez (2009) studied the reasons and risks of IS outsourcing and confirmed them through surveys on large companies in Spain. With the increasing adoption of cloud computing, rapid changes in IS outsourcing is expected, as cloud computing is its latest trend (Dhar, 2012).

The development of cloud computing can be traced back to mainframe computing. In fact, virtualization—creating several virtual versions from one entity, the core technology in cloud computing, is not new as it is part of mainframe technology (Zhang, Cheng, & Boutaba, 2010). Mainframe computing relies of a powerful machine located in a specialized room. The machine is accessed via dumb terminals connected to it.

The two closely related computing models with cloud computing is grid and utility computing. Grid computing is a middleware consisting of interconnected heterogeneous computer systems in a high-speed network to solve computation-intensive problems. The idea was developed to

solve scientific problems that need intensive and high power computations in a low cost virtual supercomputer as compared to a high cost supercomputer (Buyya & Venugopal, 2005; Weinhardt et al., 2009). Utility computing is a computing model where the available computing resources can be rented based on a pay-per-use payment model.

Essentially cloud computing uses the client-server computing model where the servers belong to the providers. Users can access services offered by cloud computing servers remotely over a network (normally the Internet) through their client machines (thin client machines should be enough). To create powerful computing power it also uses the grid computing model connecting many servers using a computer network. However, unlike grid computing that connects computers from many organizations for a sharing purpose, servers in cloud computing normally belong to an organization (provider) for a commercial purpose.

Virtualization is a key enabler of cloud computing. Virtualization creates virtual versions of a real thing. Virtualization of a machine normally creates virtual versions of the machine, meaning several independent virtual machines for different purposes can be generated. In cloud computing, virtualization is used to generate virtual servers or virtual resources (such as storage) dynamically. Therefore, multiple virtual machines (possibly with different operating systems) can be generated and muted on-demand on a single machine, creating optimal flexibility (Buyya, Yeo, Venugopal, Broberg, & Brandic, 2009; Rosenthal et al., 2010; Zhang et al., 2010).

Most of the literature in cloud computing classified services offered into three clusters (Figure 2). Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS) (Badger, Grance, Patt-Corner, & Voas, 2012; Sultan, 2009; Zhang et al., 2010).

IaaS providers normally possess a myriad of servers in their data centers. For example, Amazon Elastic Computer Cloud (EC2) cloud has 454,400 Linux servers (Liu, 2012) in its data centers. With

virtualization technology, EC2 can create millions of virtual servers and one or more virtual servers can serve an organization. Besides virtual servers, related services such network-accessible storages and network infrastructure components are provided on-demand (Badger et al., 2012). PaaS providers offer the infrastructure as well as application development platform. All tools for development of applications are provided by PaaS providers, including automation in designing, deploying, testing and administering applications to simplify application development (Badger et al., 2012; Zhang et al., 2010).

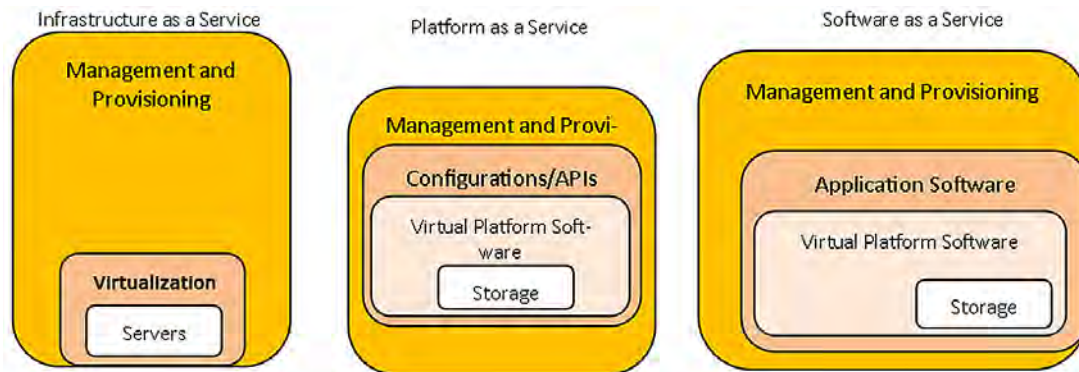
The most common type of cloud computing service is SaaS, accessible through the Web. Normally users access available applications on-demand using a web browser and pay the service based on the usage (per execution, per record processed, etc.). There are many SaaS vendors offering their services on the Web. Some examples are Google Apps from Google, iCloud from Apple and Adobe Creative Cloud from Adobe.

We interact with cloud applications every day such as Facebook, Twitter, YouTube, Wikis, and email (such as Live, Yahoo mail or Gmail). These applications are deployed for public or public cloud. There are four deployment models of cloud computing namely public cloud, private cloud, community cloud, and hybrid cloud (Mell & Grance, 2011; Subashini & Kavitha, 2011).

A public cloud as mentioned above is cloud infrastructure for general public use. The public has no control on the infrastructure, platform, or the applications. The cloud can be owned or managed by individuals, academic institutions, private or public organizations. A private cloud is cloud infrastructure exclusively owned and used by an organization to serve its intended users (such as workers or customers). The owner of the cloud (depending on the service model) has some control of the cloud. A community cloud is a cloud infrastructure for a particular community serving its members. It may be owned and managed by one or more entities in the community. A hybrid cloud is composed of two or three combinations

Figure 2. Cloud computing services

Source: Adapted from Yankee Group



of cloud deployment models, such as combination of private and public clouds, which serve different purposes but have same look and feel such as Google Apps.

BENEFITS, ISSUES, AND CHALLENGES OF OUTSOURCING IN THE CLOUD

The IS or IT outsourcing market is growing. According to Gartner, global IT outsourcing is predicted to reach \$ 288 billion in 2013, a slight increase (2.8 percent) from the previous year. Annual growth of the market is forecast to grow 5.4 percent. It was mentioned that cloud computing has taken away IS/IT outsourcing and caused downward pricing on IT services (Gartner, 2013; Overby, 2013). The growth of the public cloud computing market is much faster than IS/IT outsourcing in general. The global public cloud market is forecast to grow 18.5 percent in 2013 with total \$131 billion as compared to the previous year \$111 billion and IaaS segment market is predicted growing faster, 42.4 percent according to Gartner (2013).

According to 451 Research, the market size of enterprise cloud computing is expanding rapidly, much faster than the expansion of IT (451 Research, 2013). The annual growth of enterprise cloud computing is predicted to be 36

percent (Columbus, 2013b). Global spending in IaaS is predicted to reach \$16.5 billion in 2015. SaaS market will grow globally, from \$49 billion in 2015 to \$67 billion by 2018. Cloud applications have been dominating mobile data traffic globally, 81% in 2014 and reaching 90% in 2019 (Columbus, 2015).

The rapid growth of the cloud computing market is the result of growing adoption of cloud computing by the public as well as business organizations properly (Kim, Kim, Lee, & Lee, 2009). There are some obvious advantages of cloud computing, which influence its adoption. The following are some of these advantages (Armbrust et al., 2009; Buyya, Yeo, Venugopal, Broberg, & Brandic, 2009; Zissis & Lekkass, 2012).

Some Benefits of Cloud Computing

The first and important benefit of cloud computing is *simplicity and flexibility*. In a cloud computing environment, users are not required to own hardware other than terminals for connection, no software licenses needed hence users are free from the hassle of maintenance. Unlike the traditional computing model, users can scale-up and scale-down the usage as needed. Cloud computing providers normally upgrade their systems to improve services and the users should benefit from the ever up-to-date technology, both hardware and software.

The second benefit is *organizational agility*. Organizational agility is the ability of an organization to adapt quickly and effectively to the changing environment, such as rapid and effective response to customers' preferences while achieving strategic adaptation (adapting to trends and issues) in the medium term and long term shaping. As computing resources can be utilized in a very short time and portable applications or services can be developed quickly, cloud computing supports organizational agility. In addition, all services provided can be accessed any time globally with various client devices.

The third benefit is reliability and availability. Each cloud has a myriad of servers at different sites. Redundancy can be implemented to increase reliability and availability. The traditional computing systems rely on one or a few servers. Problems in these servers or even upgrading the servers may interrupt services. It is common that users of traditional computing systems receive messages from their system administrators that there is an interruption of services for a few hours because of maintenance. To ensure reliability and availability, conventional computing systems use redundant data centres. However, this is costly. A cloud computing service provider will definitely ensure high reliability and availability of its services as this is at the root of its credibility. In cloud computing systems, maintenance or server upgrades will not affect the services as they are addressed by other servers seamlessly.

The fourth benefit is *cost reduction*. Cost reduction on IT spending is highly expected as cloud computing can eliminate many conventional IT costs such as investment costs of hardware and software licenses as well as maintenance cost. As both investment costs and maintenance costs can be removed from IT spending, overall IT expenses in cloud computing is cheaper than conventional computing.

The fifth benefit is *scalable infrastructure*. In the cloud computing system, new hardware and new nodes can be added easily with limited modification to infrastructure setup and software

as the cloud architecture is designed to accommodate horizontal and vertical expansions. Also, the scale of the service can be expanded as requested. Hence, users can start with small and relatively inexpensive services and then expand the services as needed later.

The sixth benefit is *seamless upgrading and migration to new technology*. Technology, especially IT, keeps changing. It is hard to keep pace with the fast changing technology. However, it is important to keep updated with the latest technology as it helps improve computing performance. Updating to the latest technology is expensive and migration to a new system is not simple. Frequently, a migration requires a shutdown for a while, followed by a testing of the new system, which can create interruption of services. Upgrading and migration to new technology in cloud computing is seamless and the interruption of services is not necessary. Cloud computing providers normally keep their system up to date and the system update will have no implication for services.

Some Issues, Challenges, And Possible Outcomes

Although many advantages of cloud computing are discussed in the literature, there are many problems, issues and challenges as well such as issues on security, transparency and trust. We will discuss some issues, risks, and challenges of cloud computing shortly followed by possible solutions. Since the space is limited, we will only raise some important ones. We encourage the reader to find papers in the references or reading list for more related information.

Trust

The first and obvious issue is users (customers of cloud computing) lose control of their data. This can be a big hurdle for the adoption of cloud computing by both public and private organizations. With the conventional computing systems, all data and information are stored in the organizations'

server, policy of access on the data are established and enforced locally. Organizations make sure their information resources, especially data and information are protected from unauthorized access. Can organizations trust cloud providers to protect data on the same level of protection as if the data was stored locally? This is a very difficult question to deal with. Trust cannot be established easily, especially if there is no third party that can guarantee the security and privacy of data or information stored in the cloud. Everett (2009) discussed the issue of trust in cloud computing. One possible solution of this problem is to establish an assurance by an independent party through a certification or accreditation (such as ISO 27001 or SAS 70). An assurance from an independent third party can help establish trust between a provider and its customers.

Security

The second issue, which is highly related to the first issue, is security and privacy. This issue is also the issue of conventional computing, however, since users and their data are geographically separated and the users access the data through an open network such as the Internet, the issue highly affects the users' confidence in the cloud (Armbrust et al., 2010; Zhu, 2010). The security problems may happen in servers within the cloud, the client machines, and the network and each service model has its security issues. Subashini and Kavitha (2011) classify security issues on cloud computing into four categories: security related to third party resources, application security, data transmission security, and data storage security.

The security issues, especially on information security, can be systematically addressed using three components of information security: confidentiality, integrity, and availability (CIA). Confidentiality on data and information belonging to customers (users) is to protect the privacy of the data and information. Protecting privacy of data in conventional computing system is a technical challenge and the challenge is more complicated

in the cloud computing environment, which is distributed, and the customers are not aware of the location of data and access authorization. Sharing resources, although logically separated, if not taken care properly may pose another risk related to confidentiality of data (Badger et al., 2012).

There are some possible solutions on confidentiality of data. To keep the privacy of data, especially the sensitive ones, data must be stored in an encrypted form in the cloud storage. This may slow down data processing, as data need to be decrypted before processing. It is important to make sure whether the cloud provider has a strong encryption algorithm ready to be used. Otherwise, customers need to have their own encryption to protect their data. To avoid unintended loss of the logical separation of resource, customers may need to rest physical resource exclusively instead of rely on virtual machines allocated by the provider.

On the integrity of data or information (accuracy and consistency of stored data or information), cloud computing poses complex challenges. In a single standalone system, data integrity can be preserved by a database system through atomicity, consistency, isolation, and durability (ACID). The distributed nature of cloud computing makes data integrity a difficult issue (Abadi, 2009; Subashini & Kavitha, 2011) and cloud computing does not have assurance of data integrity. A new protocol that can solve data integrity on the cloud is needed. Kumar & Poornima (2012) proposes encoding data and data recovery that can ensure data integrity in the cloud.

Availability of data when needed is crucial. Just imagine what will happen when the data is needed for an urgent matter but the cloud is not available, neither applications nor data can be accessed. There are several possible causes of this problem; among others are attacks of denial of service (DOS), outage (including power failures, target of regulatory actions, or out of business), or simply storage problems in the cloud. Data redundancy and replication, including local replication of data can help to solve the problem.

There are some other issues on security. A survey on cloud computing security issues can be seen in (Subashini & Kavitha, 2011) and further discussion on security issues of cloud computing can be seen in (Jansen & Grance, 2011; Roberts & Al-Hamdani, 2011).

Lock-In and Interoperability

One of the business strategies of providers is to create switching cost to lock in customers so that they pay high prices to switch to another provider. Although interoperability among platforms may be supported by vendors, application program interfaces (APIs) for cloud-storage systems are proprietary and there is no standard for the APIs in the horizon (Armbrust et al., 2010; Marston, Li, Bandyopadhyay, Zhang, & Ghalsasi, 2011). According to Stallman (2008) customer lock-in on proprietary systems is a business trap since providers can change the price of the service at will. In the worst situation, the provider goes out of business, leaving customers wondering how to deal with their data stored in a proprietary format as the case of Coghead in 2009 (Marks, 2009). To avoid customer lock-in, standardization on APIs for storage systems is needed to allow interoperability among cloud providers.

Integration

It is highly likely that an organization will have multiple public clouds or hybrid clouds. Each cloud has its own data and APIs and the integration of applications or processes and data is needed. In addition, the existing legacy systems need to be integrated as well. Technology for such integration exists such as federated or heterogeneous database systems and enterprise application integration. These technologies can be adapted for the integration of clouds (Kim et al., 2009). There are many works to solve this issue; in fact some major vendors such as IBM and Oracle have offer their solutions, however, the complexity of the problem is not easy to solve as it may involve

other issues such security, scalability, management, and open platforms.

Performance

The distance between client machines and servers in the cloud and the network speed highly affect the performance of computation. The possibility of data transfer bottlenecks as the intensity of data processing and transfer as well as the number of users accessing the data increase may complicate the performance and costs as data transfer consumes the communication bandwidth (Armbrust et al., 2010; Kim et al., 2009). If the slow performance cannot be tolerated, the customers need to ask the provider to increase the bandwidth of data transfer. If we assume the provider can scale-up its infrastructure, the customer only incurs the additional cost for the scale-up. However, if the provider is unable to fulfill the demand as the scale-up may be beyond their expectations, things become complicated.

Disk access for data-intensive computation can be another source of performance problem as the I/O operation is much slower than CPUs and main memories (possibly shared by many virtual machines) and disks most likely shared by many users may cause unpredictable performance. Perhaps flash memory can be a solution, as the price is getting cheaper, consumes less energy and is much faster than disks (Armbrust et al., 2010).

FUTURE RESEARCH DIRECTIONS

Establishing trust between cloud providers and their customers is one of key success factors of cloud computing adoption. Trusted third parties can help increase the level of trust. Cloud computing governance model needs to be developed, as this can be used as a framework to strengthen the relationship a cloud provider has with its customers.

As the interest in the cloud computing is increasing and business in cloud computing is

growing, standardization of APIs and data format need to be developed to support interoperability and integration of the clouds. It can also remove vendor lock in impression. Perhaps XML data format is a good candidate for storing data in the clouds' storages as demonstrated by Knight (2009) in integrating Salesforce data.

The large penetration of small mobile devices (SMD) such tablet computers and smartphones globally has changed the way people interact and communicate. Users of these devices are connected to the Internet and they use cloud computing for many purposes, such as social networking, emailing, blogging, gaming, and storing information in the cloud storage. Mobile cloud computing is growing very fast most public clouds will be accessed by SMD. A recent survey in mobile cloud computing is discussed in (Fernando, Loke, & Rahayu, 2013). The integration of mobile and cloud computing is an interesting research area as it allows SMD to access powerful applications in the cloud.

Research on how education uses cloud computing so far and how cloud computing can support education will be very interesting, including mobile learning through the cloud and scientific research in the cloud. There are many opportunities created by cloud computing for education. Some cloud-computing vendors such as IBM and Google actively promote cloud computing as research tools and there are growing adoption of cloud computing by education institutions (Sultan, 2009).

Most of the APIs in cloud computing are proprietary and this raises concern of free/open source software communities. In fact, Richard Stallman (2008), the founder of Free Software Foundation, accused that cloud computing is another lock-in trap of proprietary systems with a different model. This accusation could be true as the APIs and data formats are proprietary as discussed above. Perhaps the open source community needs to create Open Cloud Computing fully supported by open source software using a grid computing model for the hardware connections.

CONCLUSION

Looking back at the development of IS technologies and methodology, it is apparent that the advent of cloud computing as it is today is a logical and inevitable advancement. We see that the ever-changing world of technology and globalization necessitates that businesses keep their information systems up to date and remotely accessible, and those who cannot keep up risk going out of business. Hence, the turn towards outsourcing their information systems to take advantage of a system that is scalable, never obsolete, secure and costs less.

The option to employ a cloud computing service, though, is not without its risks and questions. Will the data stored be secure from the likes of hackers and espionage? Can the providers be trusted with the privacy and sensitivity of the information? Can they keep their systems up to date? What if the provider itself goes out of business? Businesses must have full confidence in their providers before fully outsourcing their information systems.

Nowadays with smart phones becoming more and more powerful yet affordable, cloud computing will surely soon dominate, with desktops becoming less of a necessity. With the market for IS outsourcing in the hundreds of billions of dollars and expanding, more issues become apparent. Standardization must be established with regards to the data format and API's to prevent from customer lock-ins. This would in turn reduce the dependency of individuals and businesses on a single provider, reducing the risks associated with such a choice as outsourcing their IS.

The development of cloud computing is in full swing, and there is no question that as humans become more mobile and connected, the need for robust and secure cloud computing systems is becoming more important. Its vast effect on society is apparent in the popularity of web-based applications such as e-mail and social networking. New businesses want to setup their information systems to be web-based and accessible from thin

clients, and the developers of the web applications for these businesses want to develop their applications using computing resources in the cloud. For some, there is no choice but to employ a cloud computing solution. In fact, all must use a form of cloud computing in order to stay competitive in today's world, and this is the leading factor why it is predicted that the dominance of cloud computing will soon overtake desktop computers.

REFERENCES

- 451Research. (2013, August). *Market Monitor: Cloud Computing*. Retrieved on November 10, 2013, from <https://451research.com/market-monitor-cloud-computing>
- Abadi, D. J. (2009). Data Management in the Cloud: Limitations and Opportunities. *A Quarterly Bulletin of the Computer Society of the IEEE Technical Committee on Data Engineering*, 32(1), 3–12.
- Anderson, J. Q., & Rainie, L. (2010). *The future of cloud computing*. Pew Internet. Retrieved from <http://pewInternet.org/Reports/2010/The-future-of-cloud-computing.aspx>
- Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R. H., Konwinski, A., ... Zaharia, M. (2009). *Above the Clouds: A Berkeley View of Cloud Computing* (No. Technical Report No. UCB/EECS-2009-28). Electrical Engineering and Computer Sciences, University of California at Berkeley. Retrieved from <http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-28.html>
- Armbrust, M., Stoica, I., Zaharia, M., Fox, A., Griffith, R., Joseph, A. D., & Rabkin, A. et al. (2010). A view of cloud computing. *Communications of the ACM*, 53(4), 50. doi:10.1145/1721654.1721672
- Badger, L., Grance, T., Patt-Corner, R., & Voas, J. (2012). *Cloud Computing Synopsis and Recommendations* (No. Special Publication 800-146).
- Buyya, R., Yeo, C. S., Venugopal, S., Broberg, J., & Brandic, I. (2009). Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility. *Future Generation Computer Systems*, 25(6), 599–616. doi:10.1016/j.future.2008.12.001
- Columbus, L. (2013a, July 10). Why salesforce is winning the cloud platform war. *Forbes*. Retrieved November 9, 2013, from <http://www.forbes.com/sites/louiscolumbus/2013/10/07/why-salesforce-is-winning-the-cloud-platform-war/>
- Columbus, L. (2013b, September 4). Predicting Enterprise Cloud Computing Growth. *Forbes*. Retrieved November 10, 2013, from <http://www.forbes.com/sites/louiscolumbus/2013/09/04/predicting-enterprise-cloud-computing-growth/>
- Columbus, L. (2015, November 4). Roundup of cloud computing forecasts and market estimates q3 update, 2015. *Forbes*. Retrieved November 4, 2015, from <http://www.forbes.com/sites/louiscolumbus/2015/09/27/roundup-of-cloud-computing-forecasts-and-market-estimates-q3-update-2015/>
- Dhar, S. (2012). From outsourcing to Cloud computing: Evolution of IT services. *Management Research Review*, 35(8), 664–675. doi:10.1108/01409171211247677
- Dibbern, J., Goles, T., Hirschheim, R., & Jayatilaka, B. (2004). Information systems outsourcing: A survey and analysis of the literature. *SIGMIS Database*, 35(4), 6–102. doi:10.1145/1035233.1035236
- Everett, C. (2009). Cloud computing – A question of trust. *Computer Fraud & Security*, 6, 5–7. doi:10.1016/S1361-3723(09)70071-5

- Fernando, N., Loke, S. W., & Rahayu, W. (2013). Mobile cloud computing: A survey. *Future Generation Computer Systems*, 29(1), 84–106. doi:10.1016/j.future.2012.05.023
- Garfinkel, S. (2011). The Cloud Imperative. *MIT Technology Review*. Retrieved November 2, 2015, from <http://www.technologyreview.com/news/425623/the-cloud-imperative/>
- Gartner. (2013, July 17). *Gartner Says Worldwide IT Outsourcing Market to Reach \$288 Billion in 2013*. Retrieved November 10, 2013, from <http://www.gartner.com/newsroom/id/2550615>
- Gupta, U. G., & Gupta, A. (1992). Outsourcing the IS function: Is It Necessary for Your Organization? *Information Systems Management*, 9(3), 44–47. doi:10.1080/10580539208906881
- Jansen, W., & Grance, T. (2011). *Guidelines on Security and Privacy in Public Cloud Computing* (No. NIST SP 800-144). NIST - national Institute of Standards and Technology. Retrieved from <http://csrc.nist.gov/publications/nistpubs/800-144/SP800-144.pdf>
- Kajko-Mattsson, M., & Gustafsson, L. (2010). Cloud Outsourcing requires a proper handover process (position paper). In *2010 6th International Conference on Advanced Information Management and Service (IMS)* (pp. 142–146). Academic Press.
- Kim, W., Kim, S. D., Lee, E., & Lee, S. (2009). Adoption issues for cloud computing. In *iiWAS '09 Proceedings of the 11th International Conference on Information Integration and Web-based Applications & Services* (pp. 3–6). Kuala Lumpur: ACM.
- Knight, R. (2009, June 30). *The new role of XML in cloud data integration*. CT316. Retrieved November 13, 2013, from <http://www.ibm.com/developerworks/webservices/library/x-sftoeap/>
- Kumar, V. V., & Poornima, G. (2012). Ensuring Data Integrity in Cloud Computing - EICA272. *Journal of Computer Applications*, 5(EICA2012-4), 513–520.
- Lacity, M. C., & Hirschheim, R. (1993). The Information Systems Outsourcing Bandwagon. *MIT Sloan Management Review*. Retrieved on November 4, 2013, from <http://sloanreview.mit.edu/article/the-information-systems-outsourcing-bandwagon/>
- Lee, J.-N., Huynh, M. Q., Kwok, R. C.-W., & Pi, S.-M. (2003). IT Outsourcing evolution—: Past, present, and future. *Communications of the ACM*, 46(5), 84–89. doi:10.1145/769800.769807
- Liu, H. (2012, March). Amazon data center size. *Huan Liu's Blog*. Retrieved from <http://huanliu.wordpress.com/2012/03/13/amazon-data-center-size/>
- Marston, S., Li, Z., Bandyopadhyay, S., Zhang, J., & Ghalsasi, A. (2011). Cloud computing — The business perspective. *Decision Support Systems*, 51(1), 176–189. doi:10.1016/j.dss.2010.12.006
- Mell, P., & Grance, T. (2011). *The NIST Definition of Cloud Computing* (Special Publication No. 800-145). National Institute of Standards and Technology. Retrieved from <http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf>
- Overby, S. (2013, August 2). *Gartner Predicts Limited IT Outsourcing Growth and Increased Volatility*. CIO. Retrieved November 10, 2013, from http://www.cio.com/article/737472/Gartner_Predicts_Limited_IT_Outsourcing_Growth_and_Increased_Volatility
- Palvia, P. C. (1995). A dialectic view of information systems outsourcing: Pros and cons. *Information & Management*, 29(5), 265–275. doi:10.1016/0378-7206(95)00030-9

- Reyes Gonzalez, J. G. (2009). Information Systems Outsourcing Reasons and Risks: An Empirical Study. *International Journal of Humanities and Social Science*, 4(3), 181–192.
- Roberts, J. C. II, & Al-Hamdani, W. (2011). Who can you trust in the cloud?: a review of security issues within cloud computing. In *Proceedings of the 2011 Information Security Curriculum Development Conference* (pp. 15–19). New York, NY: ACM. doi:10.1145/2047456.2047458
- Stallman, R. (2008, September 29). Cloud computing is a trap, warns GNU founder Richard Stallman. *The Guardian*. Retrieved from <http://www.theguardian.com/technology/2008/sep/29/cloud.computing.richard.stallman>
- Subashini, S., & Kavitha, V. (2011). A survey on security issues in service delivery models of cloud computing. *Journal of Network and Computer Applications*, 34(1), 1–11. doi:10.1016/j.jnca.2010.07.006
- Sultan, N. (2009). Cloud computing for education: A new dawn? *International Journal of Information Management*, 30(2), 109–116. doi:10.1016/j.ijinfomgt.2009.09.004
- Vaquero, L. M., Roderio-Merino, L., Caceres, J., & Lindner, M. (2009). A break in the clouds: Towards a cloud definition. *SIGCOMM Comput. Commun. Rev.*, 39(1), 50–55. doi:10.1145/1496091.1496100
- Vitharana, P., & Dharwadkar, R. (2007). Information Systems Outsourcing: Linking Transaction Cost and Institutional Theories. *Communications of the Association for Information Systems*, 20(1). Retrieved from <http://aisel.aisnet.org/cais/vol20/iss1/23>
- Weidenbaum, M. (2005). Outsourcing: Pros and cons. *Business Horizons*, 48(4), 311–315. doi:10.1016/j.bushor.2004.11.001
- Zhang, Q., Cheng, L., & Boutaba, R. (2010). Cloud computing: State-of-the-art and research challenges. *Journal of Internet Services and Applications*, 1(1), 7–18. doi:10.1007/s13174-010-0007-6
- Zhu, J. (2010). Cloud Computing Technologies and Applications. In *Handbook of Cloud Computing* (pp. 21–45). Retrieved from http://content.schweitzer-online.de/static/content/catalog/newbooks/978144196/9781441965233/9781441965233_Excerpt_001.pdf
- Zissis, D., & Lekkas, D. (2012). Addressing cloud computing security issues. *Future Generation Computer Systems*, 28(3), 583–592. doi:10.1016/j.future.2010.12.006

ADDITIONAL READING

- Abadi, D. J. (2009). Data Management in the Cloud: Limitations and Opportunities. A *Quarterly Bulletin of the Computer Society of the IEEE Technical Committee on Data Engineering*, 32(1), 3–12.
- Anderson, J. Q., & Rainie, L. (2010). The future of cloud computing. PewInternet. Retrieved from <http://pewInternet.org/Reports/2010/The-future-of-cloud-computing.aspx>
- Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R. H., & Konwinski, A. ... Zaharia, M. (2009). Above the Clouds: A Berkeley View of Cloud Computing (No. Technical Report No. UCB/EECS-2009-28). Electrical Engineering and Computer Sciences, University of California at Berkeley. Retrieved from <http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-28.html>

- Armbrust, M., Stoica, I., Zaharia, M., Fox, A., Griffith, R., Joseph, A. D., & Rabkin, A. et al. (2010). A view of cloud computing. *Communications of the ACM*, 53(4), 50. doi:10.1145/1721654.1721672
- Badger, L., Grance, T., Patt-Corner, R., & Voas, J. (2012). Cloud Computing Synopsis and Recommendations (No. Special Publication 800-146).
- Buyya, R., Yeo, C. S., Venugopal, S., Broberg, J., & Brandic, I. (2009). Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility. *Future Generation Computer Systems*, 25(6), 599–616. doi:10.1016/j.future.2008.12.001
- Columbus, L. (2013, September 4). Predicting Enterprise Cloud Computing Growth. *Forbes*. Retrieved November 10, 2013, from <http://www.forbes.com/sites/louiscolumbus/2013/09/04/predicting-enterprise-cloud-computing-growth/>
- Dibbern, J., Goles, T., Hirschheim, R., & Jayatilaka, B. (2004). Information systems outsourcing: A survey and analysis of the literature. *SIGMIS Database*, 35(4), 6–102. doi:10.1145/1035233.1035236
- Everett, C. (2009). Cloud computing – A question of trust. *Computer Fraud & Security*, (6), 5–7. doi:10.1016/S1361-3723(09)70071-5
- Fernando, N., Loke, S. W., & Rahayu, W. (2013). Mobile cloud computing: A survey. *Future Generation Computer Systems*, 29(1), 84–106. doi:10.1016/j.future.2012.05.023
- Gartner. (2013, July 17). Gartner Says Worldwide IT Outsourcing Market to Reach \$288 Billion in 2013. Retrieved on November 10, 2013, from <http://www.gartner.com/newsroom/id/2550615>
- Gupta, U. G., & Gupta, A. (1992). Outsourcing the IS function: Is It Necessary for Your Organization? *Information Systems Management*, 9(3), 44–47. doi:10.1080/10580539208906881
- Jansen, W., & Grance, T. (2011). Guidelines on Security and Privacy in Public Cloud Computing (No. NIST SP 800-144). NIST - national Institute of Standards and Technology. Retrieved from <http://csrc.nist.gov/publications/nistpubs/800-144/SP800-144.pdf>
- Kim, W., Kim, S. D., Lee, E., & Lee, S. (2009). Adoption issues for cloud computing. In *iiWAS '09 Proceedings of the 11th International Conference on Information Integration and Web-based Applications & Services* (pp. 3–6). Kuala Lumpur: ACM.
- Knight, R. (2009, June 30). The new role of XML in cloud data integration. CT316. Retrieved November 13, 2013, from <http://www.ibm.com/developerworks/webservices/library/x-sftoeap/>
- Kumar, V. V., & Poornima, G. (2012). Ensuring Data Integrity in Cloud Computing - EICA272. *Journal of Computer Applications*, 5(EICA2012-4), 513–520.
- Lacity, M. C., & Hirschheim, R. (1993, October 15). The Information Systems Outsourcing Bandwagon. *MIT Sloan Management Review*. Retrieved November 4, 2013, from <http://sloan-review.mit.edu/article/the-information-systems-outsourcing-bandwagon/>
- Liu, H. (2012, March). Amazon data center size. Huan Liu's Blog. Retrieved from <http://huanliu.wordpress.com/2012/03/13/amazon-data-center-size/>
- Marston, S., Li, Z., Bandyopadhyay, S., Zhang, J., & Ghalsasi, A. (2011). Cloud computing - The business perspective. *Decision Support Systems*, 51(1), 176–189. doi:10.1016/j.dss.2010.12.006
- Mell, P., & Grance, T. (2011). The NIST Definition of Cloud Computing (Special Publication No. 800-145). United State: National Institute of Standards and Technology. Retrieved from <http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf>

- Overby, S. (2013, August 2). Gartner Predicts Limited IT Outsourcing Growth and Increased Volatility. CIO. Retrieved on November 10, 2013, from http://www.cio.com/article/737472/Gartner_Predicts_Limited_IT_Outsourcing_Growth_and_Increased_Volatility
- Palvia, P. C. (1995). A dialectic view of information systems outsourcing: Pros and cons. *Information & Management*, 29(5), 265–275. doi:10.1016/0378-7206(95)00030-9
451. Research. (2013, August). Market Monitor: Cloud Computing. Retrieved on November 10, 2013, from <https://451research.com/market-monitor-cloud-computing>
- Reyes Gonzalez, J. G. (2009). Information Systems Outsourcing Reasons and Risks: An Empirical Study. *International Journal of Humanities and Social Science*, 4(3), 181–192.
- Roberts, J. C. II, & Al-Hamdani, W. (2011). Who can you trust in the cloud?: a review of security issues within cloud computing. In *Proceedings of the 2011 Information Security Curriculum Development Conference* (pp. 15–19). New York: ACM. doi:10.1145/2047456.2047458
- Stallman, R. (2008, September 29). Cloud computing is a trap, warns GNU founder Richard Stallman. The Guardian. Retrieved from <http://www.theguardian.com/technology/2008/sep/29/cloud.computing.richard.stallman>
- Subashini, S., & Kavitha, V. (2011). A survey on security issues in service delivery models of cloud computing. *Journal of Network and Computer Applications*, 34(1), 1–11. doi:10.1016/j.jnca.2010.07.006
- Sultan, N. (2009). Cloud computing for education: A new dawn? *International Journal of Information Management*, 30(2), 109–116. doi:10.1016/j.ijinfomgt.2009.09.004
- Vaquero, L. M., Rodero-Merino, L., Caceres, J., & Lindner, M. (2009). A break in the clouds: Towards a cloud definition. *SIGCOMM Comput. Commun. Rev.*, 39(1), 50–55. doi:10.1145/1496091.1496100
- Vitharana, P., & Dharwadkar, R. (2007). Information Systems Outsourcing: Linking Transaction Cost and Institutional Theories. *Communications of the Association for Information Systems*, 20(1). Retrieved from <http://aisel.aisnet.org/cais/vol20/iss1/23>
- Weidenbaum, M. (2005). Outsourcing: Pros and cons. *Business Horizons*, 48(4), 311–315. doi:10.1016/j.bushor.2004.11.001
- Zhang, Q., Cheng, L., & Boutaba, R. (2010). Cloud computing: State-of-the-art and research challenges. *Journal of Internet Services and Applications*, 1(1), 7–18. doi:10.1007/s13174-010-0007-6
- Zhu, J. (2010). Cloud Computing Technologies and Applications. In *Handbook of Cloud Computing* (pp. 21–45). Retrieved from http://content.schweitzer-online.de/static/content/catalog/newbooks/9781441961441965233/9781441965233_Excerpt_001.pdf
- Zissis, D., & Lekkas, D. (2012). Addressing cloud computing security issues. *Future Generation Computer Systems*, 28(3), 583–592. doi:10.1016/j.future.2010.12.006

KEY TERMS AND DEFINITIONS

Cloud Computing: Is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g. networks, servers, storage, applications and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.

Grid Computing: Is a middleware consisting of interconnected heterogeneous computer systems in a high-speed network to solve computation-intensive problems.

IaaS: Is a type of cloud computing service that provide infrastructure, especially servers or virtual servers located in data centers to customers. The servers or virtual servers are normally accessed through a computer network such the Internet.

Outsourcing: Is a business term to describe a mechanism in which a company utilizes services provided by another company, normally through a contract, to fulfill some of its required business resources or functions.

PaaS: Is a type of cloud computing service that offer the infrastructure as well as application development platform. PaaS providers accessible

through a computer network or the Internet provide all tools for development of applications. These tools may include automation in designing, deploying, testing, and administering applications to simplify application development.

SaaS: Is a type of cloud computing service that offer various software accessible to customers through the Internet.

Virtualization: Is mechanism to create virtual version of a real thing. Virtualization of a machine normally creates virtual versions of the machine, meaning several independent virtual machines for different purposes can be generated. In cloud computing, virtualization is used to generate virtual servers or virtual resources (such as storage) dynamically.

Improved Checkpoint Using the Effective Management of I/O in a Cloud Environment

Bakhta Meroufel

University of Oran1, Algeria

Ghalem Belalem

University of Es Senia Oran1, Algeria

INTRODUCTION

The emergence of cloud computing has brought a new dimension to the world of information technology. Although cloud computing offers several advantages such as virtualization, cost reduction, multi-tenancy, etc., there are risks and failures associated with it (Yang *et al.*, 2014). A key challenge for research in cloud computing is to ensure the reliability of the system without reducing the overall system performance. Among of fault tolerance, there is the strategy of checkpointing. The major problem of checkpointing is the overhead caused by the storage time of checkpointing files in stable storage, this time is estimated at 70% of checkpointing process time caused by the storage (Ouyang *et al.*, 2009a; Cornwell & kongmunvatana, 2011a), Figure1 shows the main phases of the checkpointing process. This process is based on three phases: i) suspend communication between processes and ensure consistent state; ii) use the checkpointing library to create and store checkpoints; iii) re-connect processes and continue execution.

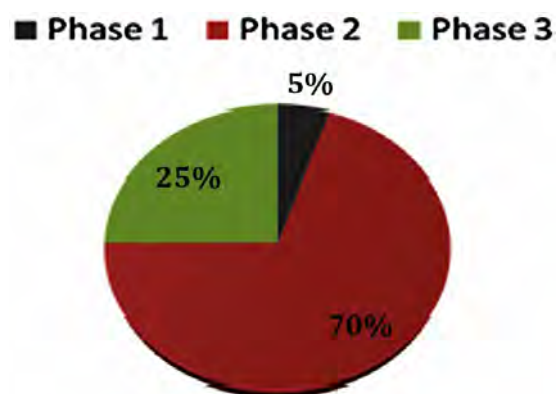
The aim of our work is to minimize the overhead of checkpointing by minimizing its storage time. To ensure this goal, we improve the I/O management and we propose a checkpointing strategy of three phases:

1. The construction of VRbIO topology (Virtual RbIO): RbIO proposed in (Lui *et al.*, 2010) is

a virtual hierarchical topology; it minimizes checkpointing time and I/O time at the same time. In our system, each VM has a reactive agent responsible of the local I/O management; at the end of this phase some of these reactive agents will be activated to manage the I/O of a group of VMs of the server. In this case, the I/O will be hierarchical.

2. Creating the checkpointing files using co-ordinated checkpointing protocol.
3. Ensuring a lightweight and fault-tolerant storage of these files by using Collective and Selective Data Sieving input/output (CSDS I/O), which is executed by only the active agents. CSDS is an improved ROMIO I/O strategy. However, this strategy has several problems and limitations (Fu *et al.*, 2011).

Figure 1. The time of the phases of the checkpointing process



DOI: 10.4018/978-1-5225-2255-3.ch096

Our algorithm with its three phases provides solutions for most issues raised by the use of classical checkpointing with ROMIO as an I/O strategy. The rest of the chapter is organized as follows: Section 2 presents the background in the field of aggregating I/O techniques with a comparative study. ROMIO and its features are illustrated in Section 3. Section 4 presents our contribution, each service of this contribution is described in details, and all the problems cited in previous section are solved in this section. Section 5 presents some experimental results, followed by a conclusion and future research directions.

BACKGROUND

An important reason for the limitations of I/O systems is that applications often send smaller queries disjoint. This access mode generates a first additional cost to the large number of applications running on various transmission channels, but more significantly increases the processing time of the latter (Sadiku *et al.*, 2014). To deal with this problem, several “aggregation” methods have been proposed we can distinguish two types of aggregations strategies: dependent and collective.

Independent I/O is a straightforward form of I/O and is widely used in parallel applications. This form of I/O can be called independently by an individual process or any subset of processes of a parallel application. The advantage of independent I/O is that users have the freedom to perform I/O for each individual process or any subset of the processes that open the file.

The buffering is an Independent I/O (Cornwell & Kongmunvattana, 2011a). In conventional strategies, the write operation transfers data from the buffer to the local disk from their reception. Buffering proposes that the buffer will be used for temporary storage of I/O. The write operation includes small blocks in a buffer (of limited size). Once the buffer is completely filled, it will be forwarded to the local disk.

The “List I/O” approach (Thakur *et al.*, 1999a) provides routines to indicate within a single call access number. A list of torque (offset, size) describes the distribution of data in memory and a similar list is used to perform matching on disk.

Data sieving (Thakur *et al.*, 1999a) is one of the techniques proposed to address this issue by aggregating small requests into large ones. Instead of accessing each small piece of data separately, data sieving accesses a large contiguous scope of data that includes the small pieces of data. The additional unrequested data are called holes. The size of holes compared to the requested data controls the efficiency of data sieving.

For many parallel applications, even though each process may access several non-contiguous portions of a file, the requests of multiple processes are often interleaved and may constitute a large contiguous portion of a file together (Chen *et al.*, 2010). In order to achieve better I/O performance, a group of processes may cooperate with each other in reading or writing data in a collective and efficient way, which is known as collective I/O.

The collective I/O is a general idea that exploits the correlations among accesses from multiple processes of a parallel application and optimizes its I/O accesses. The basic idea behind this technique is to coordinate I/O accesses from different processors. The processors exchange information regarding what data each of them needs to access. This information is used to derive an efficient I/O schedule.

The collective I/O can distinguished by the “physical place” where the operation group is performed: if aggregation is executed among processes, the most used method is the “Two-Phase I/O” approach; if aggregation is performed at the records, we are talking about system “Disk-Directed I/O” if the approach is finally realized within a server, the method is “server-directed I/O”.

“Two-Phase I/O”, this method (Del Rosario *et al.*, 1993), as its name suggests, consists of two main phases: after a consensus between the

Table 1. Comparative study between aggregation techniques

Criteria	Data Sieving	Coll. I/O	ListI/O	ViewI/O	Buff
Useless Data Transfer	Yes	No	Yes	No	No
Network Data Transfer	Read: 1 Write: 2	2	1	1	1
Responsible	All	All	All	All	Manager
Resource	Local buffer	Local Buffer	Piggy-backed Message	Local/ Coll Buffer	Local/Coll Buffer
File Distribution	Logic	Logic	Logic\Physic	Logic\Physic	Logic

processes involved, the first step is to retrieve the data, the second concerns the redistribution between each of the latter processes.

To improve the I/O strategies (independent and collective I/O), the “File View” (Isaila *et al.*, 2003) proposes to specify the access patterns and reacts according to each model. The access pattern is defined according to the knowledge of Logical/Physical distribution of files in client and memory level.

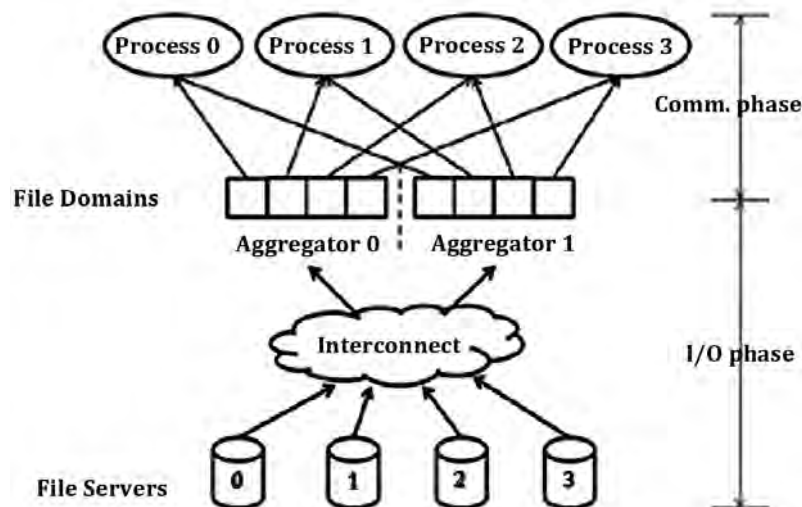
ROMIO

ROMIO is one of the most famous I/O implementation, it combine between *Two phases Collective I/O* and *Data Sieving* to create contiguous requests

in node and system levels (Thakur *et al.*, 1999a). ROMIO three phases is observed:

- Independent requests, each process run independent requests to access a contiguous region of data.
- Create contiguous blocks using “Data Sieving” in each process.
- Collective I/O: processes communicate between them to collect contiguous blocks of the same file for different processes.

Figure 2 presents an example of ROMIO I/O in four processes. After phase ‘Read’ the Data sieving is applied in each process to create contiguous blocks. ROMIO does not use a common buffer and the data are distributed over the buffers

Figure 2. Example ROMIO (Chen *et al.*, 2010)

of different processes (number of aggregators = number of processes) (Chen *et al.*, 2010). In this case, the process should contact each other to create more contiguous. The contiguous requests for the same file.

The checkpointing using the ROMIO I/O management generally suffers from several problems such as:

- **Problem 1:** Expensive communication phase. The number of messages sent during the communication phase increases with the number of processes and the number of files blocks to read or write.
- **Problem 2:** Limited buffer of nodes. ROMIO is completely distributed, so the local buffers of processes are insufficient to store contiguous blocks.
- **Problem 3:** Quantity of unnecessary data can be high. ROMIO creates contiguous blocks by using Data sieving. It is possible that the holes beyond the useful data, which makes reading/writing expensive.
- **Problem 4:** The storage time is still high. Although ROMIO uses Data sieving and collective I/O to create contiguous blocks and eliminate duplicate blocks, and since ROMIO is completely distributed, each process will be blocked to achieve I/O.
- **Problem 5:** Files storing is not fault tolerant. To ensure the correctness of checkpointing, we must ensure the storage atomicity. The checkpointing with ROMIO is responsible for I/O without taking into account the atomicity.

Like other IT services, cloud services are not immune to failures. In addition in such systems the failure rate increases with the size of the system itself. In this context, the application of operational safety is an element of primary importance, so a fault tolerance mechanism is necessary to ensure certain aspects such as availability. The checkpoint is one of tolerance mechanisms most used in distributed systems failures. Our contribution is

proposed an approach for the management of I/O to minimize the time of checkpointing by reducing its storage time. In this work, we propose an adaptive checkpointing approach that solves each of these problems using a virtual hierarchic topology and a lightweight I/O strategy, we called CSDS (Collective and Selective Data Sieving) I/O and which ensures parallelism and scalability making it adequate to cloud computing.

CONTRIBUTION

The checkpointing time is divided into two parts: the overhead and the latency. Overhead checkpointing is the time necessary to ensure the consistency and capturing the current state of the system (Create checkpointing files). It depends on the used checkpointing protocol (coordinated, uncoordinated, induced communication,...). We have proposed a fault tolerant checkpointing strategy that provides a hierarchical storage without losing the notion of I/O parallelism and also improves the I/O management by a smart strategy CSDS.

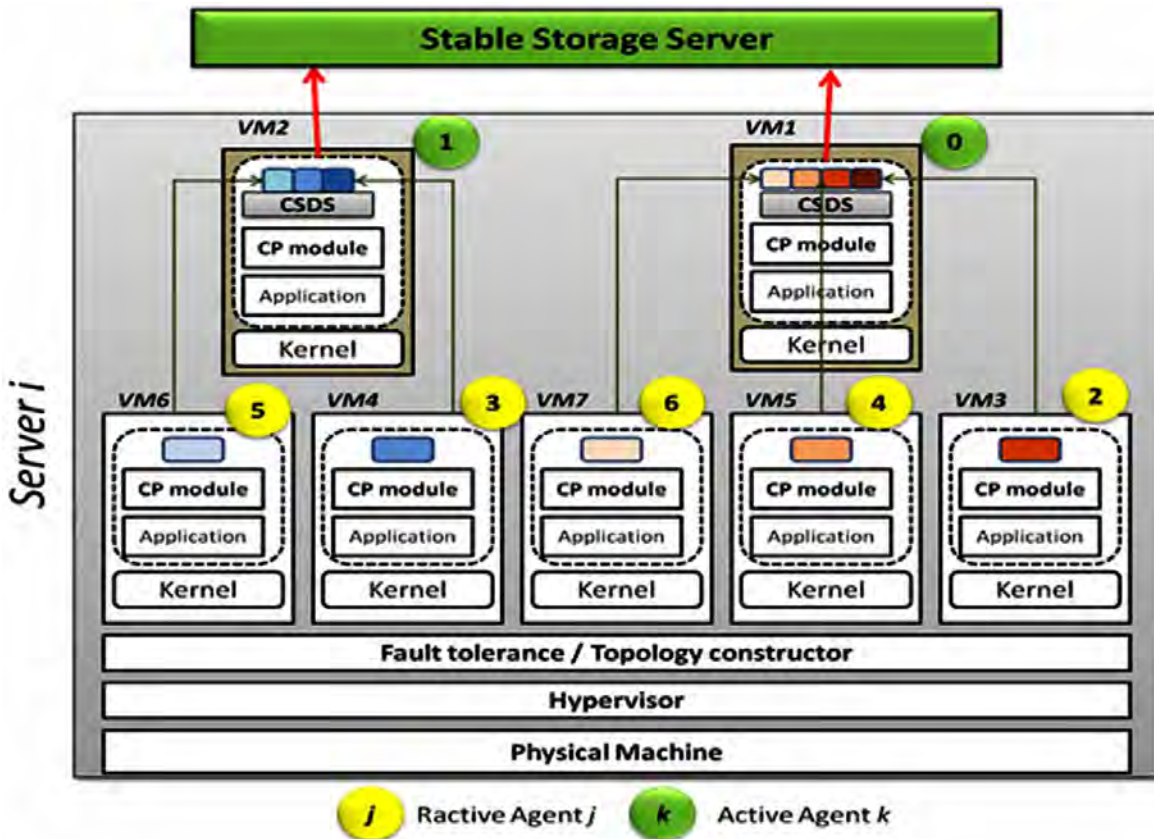
System Model

Figure 3 shows the different services in the server. Servers (Physical Machine) are assigned to one or more virtual machines (VMs) using a hypervisor. The hypervisor is the software that provides a virtualized hardware environment to support running multiple operating systems simultaneously using a single physical server.

The server contains a topology constructor service and fault tolerance service. The objectives of these services are respectively: creating VR-bIO topology and providing fault tolerance using checkpointing protocol. (controlling checkpointing interval, ensuring checkpointing atomicity, ensuring a consistent state,...).

Each VM can accept an application for special user when executing an application. The kernel in the virtual machine contains the communication modules and task management. We have extended

Figure 3. Server architecture



this architecture by implementing a checkpoints module (CP module) inside each VM. CP module creates checkpoints file that saves the actual VM state. A reactive agent inside the VM is responsible of storing the checkpointing file created by the CP module. The colored boxes in each VM represent the buffers where the checkpointing files will be stored temporarily. The server is structured in virtual topology named VRbIO where the reactive agents inside VMs can be activate to manage an intelligent I/O of their server using CSDS modules (Collective and Selective Data Sieving). So CSDS module is activated only in case of active agent. We also assume that each data center contains a permanent storage memory where the checkpointing files will be stored. Only active agents have the right to write into the memory.

Algorithm of Our Approach

The algorithm consists of three main phases to ensure fault tolerance by checkpointing: Create a virtual topology VRbIO, create checkpointing files using coordinated checkpointing protocol then run CSDS I/O to store files in the stable storage server.

VRbIO Topology

ROMIO is a strategy completely distributed. So, all the nodes use their buffers for temporarily storing files. Then they communicate with each other to create more contiguous blocks (Collective I/O) which increases the time of communications phase. To solve this problem (problem 1), we propose to use a hierarchical topology for managing I/O. According to the comparative study of different

topologies proposed in (Lui *et al.*, 2010), it is clear that RbIO is the best topology that ensures the scalability and the parallelism. In RbIO there are two types of nodes: Worker and Writers. The Writer is the node responsible for managing I/O of a set of Workers. The Worker is an ordinary node; it belongs to a single writer on its server. RbIO proposed in (Lui *et al.*, 2010) is physical. However, in our work we will create a virtual fault tolerant RbIO named VRbIO in each server (Host). In our server, we suppose that each VM has a reactive agent responsible of managing the local I/O. In order to simulate the comportment of Worker and Writers of the physical RbIO, our VRbIO activates some reactive agents of the server to play the role of Writers. The other inactivated agents (reactive agents) will be Workers. The existence of reactive agents in all VMs and the activation process allow more dynamicity and flexibility in the architecture. When an active agent fails, it is sufficient to activate one reactive agent in its group, without any intervention. The activation of an agent means the activation of the CSDS service. Only the active agent can access to the stable memory to store the checkpointing files.

To create the VRbIO, the topology constructor of the server i selects at first γ_i reactive agents to be activated. We suppose that γ_i is equal to the square root of the number of VMs created in that server i . The list of VMs in server i is sorted in descending order of speed and buffer size. Starting from the beginning of the sorted list, the topology constructor activates the first γ_i reactive agents (the rest agents of the list remain reactive). The reactive agents are assigned to their active agents in a round-robin fashion (See Formula (1)). We note $AA_{k,i}$ as the active agent k in server i . so

$AL_i = \{AA_{0,i}, AA_{1,i}, \dots, AA_{(\gamma-1),i}\}$ is the list of active agents in the server i . The list of reactive agents is $RL_i = \{RA_{\gamma,i}, RA_{(\gamma+1),i}, \dots, RA_{n,i}\}$ where $RA_{j,i}$ is the reactive agent j in server i and n is the number of VMs in the server. We also define $G(AA_{k,i})$ as the list of reactive agents associated to the active agent $AA_{k,i}$ (the Workers).

$$RA_{j,i} \in G(AA_{k,i}) \Leftrightarrow k = j \bmod \gamma_i \neq 0$$

We note also $VM(RA_{j,i})$ as the VM of reactive agent j (and $VM(AA_{k,i})$ as the VM of active agent k). The VRbIO improves the checkpointing, When checkpointing interval expires, the $VM(RA_{j,i} \setminus RA_{j,i} \in G(AA_{k,i}))$ creates a checkpoint file (temporarily stored in the local buffer) then sends this file to its active agent $AA_{k,i}$. Upon the reception of checkpoints files, the $AA_{k,i}$ uses CSDS to improve I/O and minimize time storage of these files (See next section).

The fact that each reactive agent knows exactly its active agent, the communication time is reduced dramatically (problem.1). The problem with the strategy RbIO and even in ROMIO is the buffer size which may be insufficient to store checkpoints files (problem.2). In our approach, the VM of active agent (Writer) must be able to store the files of its group of VMs with reactive agents (Workers) to optimize I/O using CSDS. To consider this problem, we used an idea proposed in BlobCR (Nicolae & Cappello, 2011) strategy. According to this strategy, after the construction of VRbIO, each $VM(RA_{j,i})$ gives a part of the buffer to the VM of its active agent $VM(AA_{k,i})$. The Algorithm 1 gives the detail of VRbIO construction.

In Figure 3 that represents the server i , the list of active agents is: $AL_i = \{AA_{0,i}, AA_{1,i}\}$ where

$$\{VM(AA_{0,i}) = VM_1, VM(AA_{1,i}) = VM_2\}.$$

The list of reactive agent is: $RL_i = \{RA_{2,i}, RA_{3,i}, RA_{4,i}, RA_{5,i}, RA_{6,i}\}$ where $\{VM(RA_{2,i}) = VM_3, VM(RA_{3,i}) = VM_4, VM(RA_{4,i}) = VM_5, VM(RA_{5,i}) = VM_6, VM(RA_{6,i}) = VM_7\}$

For the group affectation, we have $G(AA_{0,i}) = \{RA_{2,i}, RA_{4,i}, RA_{6,i}\}$ and $G(AA_{1,i}) = \{RA_{3,i}, RA_{5,i}\}$

Checkpointing Creation

In our work, we used coordinated checkpointing. Where all VMs synchronize their efforts to create a consistent state using control messages. In a consistent state, orphan and transit messages are

Algorithm 1. The VRbIO construction

```

// $\gamma_i$ : Number of Active agents in the server  $i$ . //  $ListVM_i$ : The list of VMs in the server  $i$ .
//  $AL_i$ : List of active agents in the server  $i$ . //  $RL_i$ : The list of Reactive agents in the server  $i$ .
// Initiation
For  $j = 1$  to  $ListVM_i.size$  do
// All the agents in the VM list are reactive in the beginning
 $Agent(ListVM_i(j)).State = Reactive$ ;
 $RL_i.add(Agent(ListVM_i(j)))$ 
End for
 $\gamma_i \leftarrow \sqrt{ListVM_i.size()};$ 
 $ListVM_i$  sorted in descending order of speed and buffer size
// the activation of reactive agents
For  $j = 1$  to  $\gamma_i$  do
 $Agent(ListVM_i(j)).State = Active$ ;
 $Agent(ListVM_i(j)).ID = i - 1$ ;
 $RL_i.add(Agent(ListVM_i(j)))$ ;
 $RL_i.remove(Agent(ListVM_i(j)))$ ;
End for
For  $j = 1$  to  $RL_i.size$  do
 $RL_i(j).ID = (\gamma_i - 1) + j$ ;
End for
// The affectation of reactive agents to their active //agents
For  $j = 1$  to  $RL_i.size \neq 0$  do
For  $k = 0$  to  $AL_i.size \neq 0$  do
If  $RL_i(j).ID \bmod \gamma_i = k$  Then
 $G(AL_i(k)).add(RL_i(j))$ 
End If
End for
End for

```

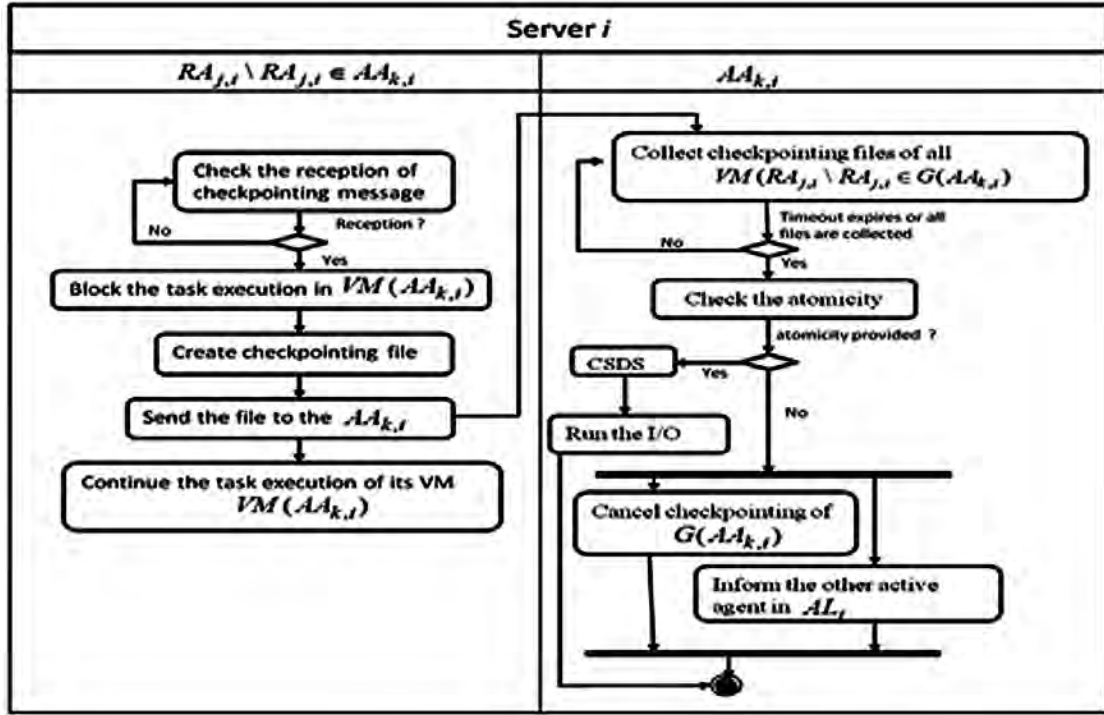
avoided by blocking the communication during the checkpointing phase and storing the susceptible transit messages (Kangarlou *et al.*, 2012). In this case, the system stores a single recovery line in the stable memory. The coordinated checkpointing is used in several studies (Arockiam & Geo Francis, 2012), (Kangarlou *et al.*, 2012).

Fault tolerance service and topology constructor are responsible for selecting the checkpointing interval $Time_{cp}$ adequate to system features. If $Time_{cp}$ expires, the service informs its VMs by sending a request for checkpointing. Upon receiving this message, each VM creates its own checkpointing file (stored in local buffer). VRbIO topology allows us to use the technique of Soft checkpointing. In conventional strategies (Hard checkpointing) VMs store their files directly in the stable memory which increases the storage time (Chen *et al.*, 2010).

In our strategy, each $VM(RA_{j,i} \setminus RA_{j,i} \in G(AA_{k,i}))$ sends its checkpoints file to $VM(AA_{k,i})$ and then continues immediately the execution. The $AA_{k,i}$ is responsible for storing the files of its reactive agent's VMs in the stable memory. This hierarchical checkpointing storage is called Soft checkpointing and it is a solution for the problem 4.

The major problem in the checkpointing is to ensure its atomicity, that is to say, we must ensure that all concerned VMs complete the creation of their checkpointing (create the file checkpointing and stored). In our approach, the active agent ensures partial checkpointing atomicity of the VMs of its group of reactive agent including its VM. If during the checkpointing, a $VM(RA_{j,i} \setminus RA_{j,i} \in G(AA_{k,i}))$ is incapable of creating its checkpoint (this is the case of a failed VM, blocked VM, busy VM, ...), its reactive agent $AA_{k,i}$ will be informed (either by a special message, fault

Figure 4. Checkpointing management



detection service or even a timeout). In this case, $AA_{k,i}$ cancels the checkpointing of VM in $G(AA_{k,i})$ by sending a message “Delete” to them (VMs that have not yet send their checkpointing file) and to the other active agents in AL_i of its server i . If a $AA_{j,s} \setminus s \neq i$ receives “Delete” from $AA_{k,i}$, it cancels the checkpointing of its $G(AA_{k,i})$. This partial atomicity technique solves the Problem 5 and makes the checkpointing itself fault tolerant. After the $AA_{k,i}$ collects checkpointing files of its $G(AA_{k,i})$, it uses the CSDS service to optimize the I/O (See Figure 4).

CSDS for Checkpointing Storage

After the $AA_{k,i}$ ensures checkpointing atomicity of its group, its CSDS service handles the I/O management to minimize the checkpointing latency. CSDS is similar to ROMIO. But in ROMIO, the Data Sieving and collective I/O will always be executed regardless of the size of useless data quantity caused by Data Sieving (Holes). In this case, the amount of unnecessary data can be large

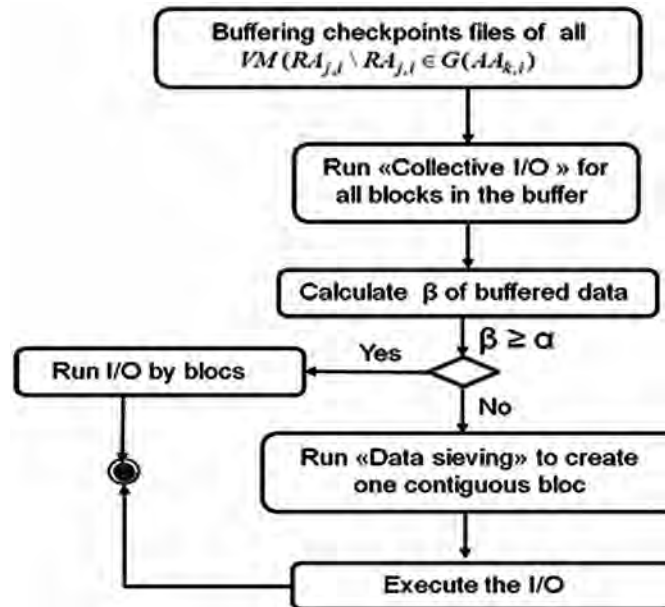
compared with the useful data, which increases the cost and the time of the I/O (problem 3). To resolve this issue, the CSDS of the active agents $AA_{k,i}$ executes Collective I/O for data requested by different $VM(RA_{j,i} \setminus RA_{j,i} \in G(AA_{k,i}))$, and then performs Data Sieving only if the size of useless data UD_j does not exceed a certain threshold α_j versus the total size of data to be written D_j by the active agent j . The threshold α is specified by SLA criteria (System Level Agreement).

The β_j parameter of active agent j represents the percentage of useless data written relative to the entries data (see Formula 2). The β_j will be compared to α_j to decide to perform a Data Sieving or not (simple buffering).

$$\beta_j = \frac{(100 \times UD_j)}{D_j} \quad (2)$$

This strategy eliminates the problem 3. The details of CSDS algorithm executed by an active agent $AA_{k,i}$ are illustrated in Figure 5.

Figure 5. CSDS Algorithm



EXPERIMENTAL RESULTS

Our approach noted CP_CSDS is destined to improve the checkpointing performances and it uses CSDS to manage the I/O, so we compare it with a classical coordinated checkpointing using ROMIO as I/O strategy (CP_ROMIO). We used several scenarios and parameters and we implement the

approaches in CloudSim simulator (Calheiros *et al.*, 2011).

In the first experiment, we measured the storage time of checkpoints files (checkpointing latency) in both I/O strategies using a server of 10 to 20 VMs and a checkpointing interval=10 minutes. The results are presented in Figure 6. CP_CSDS reduces the storage time of checkpoints files by a

Figure 6. Impact of number of checkpoints on storage time

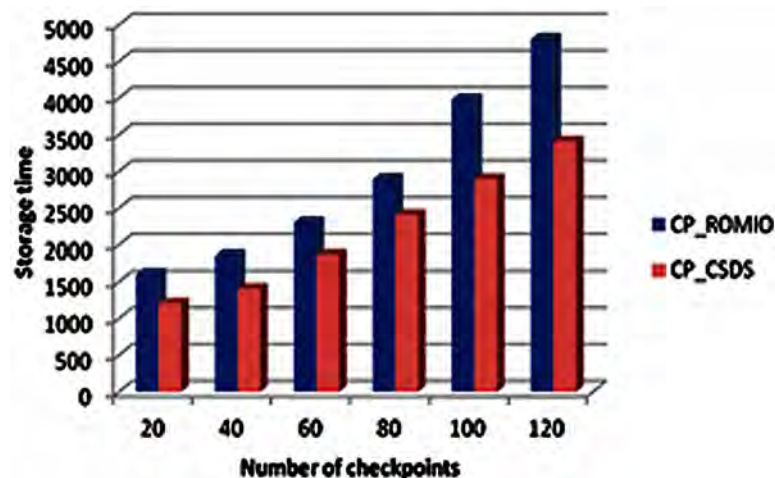


Figure 7. Impact of checkpoints sizes on storage time

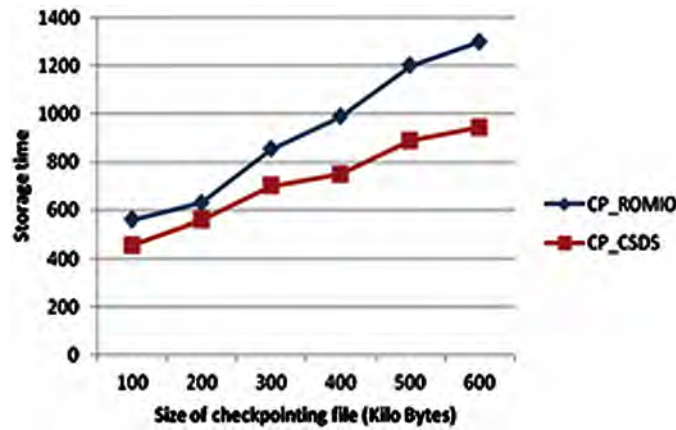
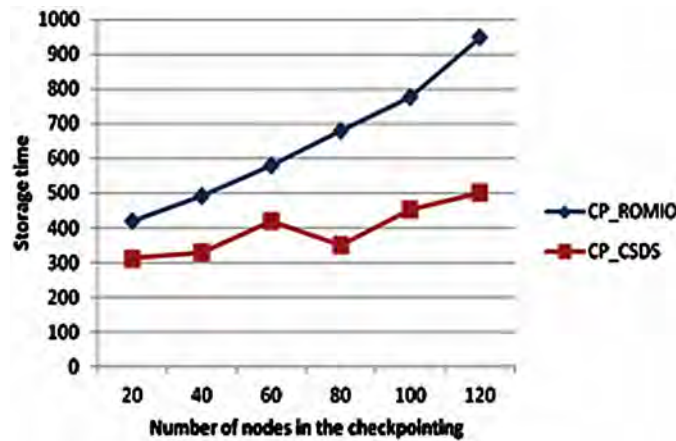


Figure 8. Impact of number of nodes involved in the checkpointing on storage time



most 30% compared to CP_ROMIO. CP_CSDS uses VRbIO to reduce the communication phase. This is due to the hierarchical checkpointing storage used by RbIO and the I/O management based on CSDS that reduces the communication phase.

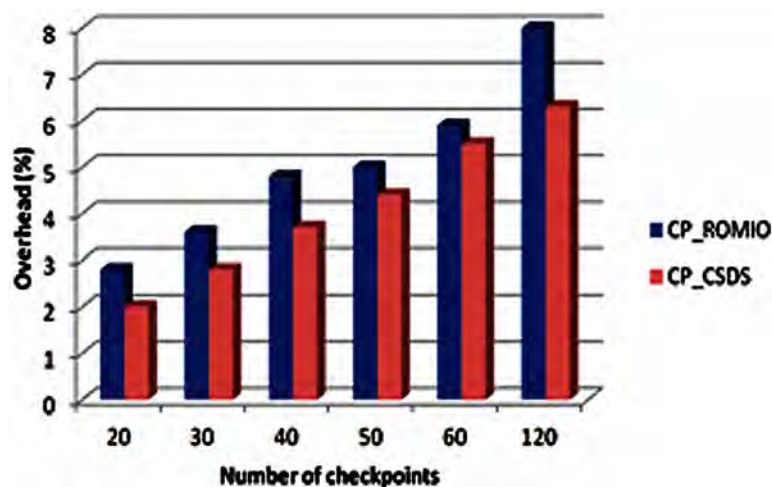
The second experiment is destined to measure also the storage time in case of different sizes of checkpointing files. According the results illustrated in Figure 7, the storage time increases if the size of files increases. However, our strategy CP_CSDS is better than CP_ROMIO (a gain of 20%) because the CSDS reduces the transfer of useless data during the I/O.

The purpose of the third experiment is to study the scalability of our strategy in terms of number of participating nodes in checkpointing

process. From the results shown in Figure 8, increasing the number of participating nodes in the checkpointing automatically increases the storage time in the classical approach of CP_ROMIO. In CP_ROMIO, all the nodes must communicate with each other to ensure the “Collective I/O” phase. In our approach CP_CSDS, the topology VRbIO provides a load balancing and scalability. Just the active agents communicate with each other to ensure the I/O which greatly reduces the storage time. The gain in term of storage time is $\approx 30\%$ compared to CP_ROMIO.

The last experiment calculates the overhead caused by the checkpointing using both approaches I/O (See Figure 9). It is clear that our approach CP_CSDS reduces significantly the

Figure 9. Impact of number of checkpoints on the Overhead



overhead of checkpointing ($\approx 18\%$) because in addition to reducing the storage time, it uses the soft checkpointing.

CONCLUSION

Checkpointing is very powerful fault tolerance technique in term of fault management and can be used for different objective such as migration and load balancing. The major problem of this technique is the storage time which can increase response time and minimizes the possibility to meet the deadlines specified by the user and the SLA rules. To improve the performance of checkpointing in the clouds, we have focused in this paper on minimizing the storage time of checkpointing files in stable storage server. Our approach involves three main steps: the construction of the virtual topology VRbIO, creating checkpointing

files using coordinated checkpointing protocol and finally save these files in the stable memory using a new strategy I/O that combines between the collective I/O and intelligent data sieving CSDS. In this paper we explained the various problems of using checkpointing with ROMIO as I/O strategy (CP_ROMIO). Our approach solves the majority of classical problems and makes the checkpointing appropriate for clouds environment. A brief comparison between our approach CP_CSDS and that of checkpointing with ROMEIO I/O (CP_ROMIO) is presented in Table2. In this table, the parameter n presents the number of nodes (VMs) in the server.

FUTURE RESEARCH DIRECTIONS

Several directions can be exploited, we can mention:

Table 2. CP_ROMIO vs CP_CSDS

	Topology	Number of Writer	Flexibility	Type of Checkpointing	Communication Phase	Buffer Problem	Transfert of Useless	Atomicity
CP_ROMIO	Distributed	n	No	Hard	Yes	Yes	High	No
CP_CSDS	Hierarchical (VRbIO)	$\sqrt{n}$	Yes	Soft	No	No	Reduced	Yes

- Improve strategies of VRbIO topology by making the assignment of workers (reactive agents) to their writers (active agents) dynamic depending on the load and the characteristics of the system;
- Improve CSDS by the choice of other critical criteria to swing between the transfer block and the data Sieving as: i) data transfer time; ii) energy consumption; iii) the cost.
- Extend CSDS in a predictive system by pre-loading for user checkpoint file playback which minimizes recovery time in case of failures.

REFERENCES

- Arockiam, L. & Geo Francis E. (2012). FTM-A Middle Layer Architecture for Fault Tolerance in Cloud Computing. *ICNICT*, (2), 12-16.
- Calheiros, R. N., Ranjan, R., Beloglazov, A., DeRose, C. A. F., & Buyya, R. (2011, January). CloudSim: A toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms. *Software, Practice & Experience*, 41(1), 23–50. doi:10.1002/spe.995
- Chen, Y., Sun, X.-H., Thakur, R., Song, H., & Jin, H. (2010). Improving Parallel I/O Performance with Data Layout Awareness, *Proceedings of the 2010 IEEE International Conference on Cluster Computing*, 302-311. doi:10.1109/CLUSTER.2010.35
- Cornwell, J. & Kongmunvattana, A. (2011). Efficient System-Level Remote Checkpointing Technique for BLCR. *Eighth International Conference on Information Technology: New Generations*, 1002-1007. doi:10.1109/ITNG.2011.172
- Del Rosario, J., Bordawekar, R., & Choudhary, A. (1993). Improved parallel I/O via a two-phase runtime access strategy. *ACM SIGARCH Computer Architecture News*, 21(5), 31-38.
- Fu, J., Min, M. S., Latham, R., & Carothers, C. D. (2011). Parallel I/O Performance for Application-Level Checkpointing on the Blue Gene/P System. *Workshop on Interfaces and Architectures for Scientific Data Storage (IASDS)*, 465-473. doi:10.1109/CLUSTER.2011.81
- Isaila, F., & Tichy, W. F. (2003). View I/O: improving the performance of non-contiguous I/O. *The Third IEEE International Conference on Cluster Computing*, 336-343.
- Kangarlou, A., Eugster, P., & Xu, D. (2012). VNsnap: Taking Snapshots of Virtual Networked Infrastructures in the Cloud. *IEEE Transactions on Services Computing*, 5(4), 484–496. doi:10.1109/TSC.2011.29
- Liu, N., Fu, J., Carothers, C. D., Sahni, O. K. E., Jansen, K. E., & Shephard, M. S. (2010). Massively Parallel I/O for Partitioned Solver Systems. *Parallel Processing Letters*, 20(4), 377–395. doi:10.1142/S0129626410000302
- Nicolae, B., & Cappello, F. (2011). BlobCR: Efficient checkpoint-restart for HPC applications on IaaS clouds using virtual disk image snapshots. *Int'l Conference on High Performance Computing Networking, Storage and Analysis (SC'2011)*, 1-34. doi:10.1145/2063384.2063429
- Ouyang, X., Gopalakrishnan, K., & Panda, D.-K. (2009a). Accelerating checkpoint operation by node-level write aggregation on multicore systems. *Proceeding of International Conference on Parallel Processing (ICPP'2009)*, 34-41. doi:10.1109/ICPP.2009.73
- Sadiku, M. N. O., Musa, S. M., & Momoh, O. D. (2014). Cloud computing: Opportunities and challenges. *Potentials IEEE Journal*, 33(1), 34–36. doi:10.1109/MPOT.2013.2279684

Thakur, R., Gropp, W., & Lusk, E. (1999a). Data Sieving and Collective I/O in ROMIO. *Proceedings of the Seventh Symposium on the Frontiers of Massively Parallel Computation*, 182–189. doi:10.1109/FMPC.1999.750599

Yang, B., Pang, X.-Q., Du, J.-Q., & Xie, D. (2014). Effective Error-Tolerant Keyword Search for Secure Cloud Computing. *Journal of Computer Science and Technology*, 29(1), 81–89. doi:10.1007/s11390-014-1413-1

KEY TERMS AND DEFINITIONS

Cloud Computing: Refers to applications and services offered over the Internet. These services are offered from data centers all over the world, which collectively are referred to as the “cloud”.

Checkpointing: A technique to add fault tolerance into computing systems. It basically consists

of saving a snapshot of the application’s state, so that it can restart from that point in case of failure.

Collective I/O: In many parallel applications, despite the fact that each process may need to access several noncontiguous portions of a file, the requests of different processes are often interleaved and may together span large contiguous portions of the file. Collective I/O procedure is used to improve significantly the I/O performance by merging the requests of different processes and servicing the merged request.

Data Sieving: To reduce the effect of high I/O latency in parallel applications, it is critical to make as few requests to the file system as possible. Data sieving is a technique that enables an implementation to make a few large, contiguous requests to the file system even if the user’s request consists of several small, noncontiguous accesses.

ROMIO: A portable MPI-IO implementation that works on many different machines and file systems.

Service Quality and Perceived Value of Cloud Computing–Based Service Encounters

C

Eges Egedigwe

Dallas County Community College, USA

INTRODUCTION

Cloud computing based technology is becoming increasingly popular as a way to deliver quality education to community colleges, universities and other organizations. At the same time, compared with other industries, colleges have been slow on implementing and sustaining cloud computing services on an institutional level because of budget constraints facing many large community colleges, in addition to other obstacles. Faced with this challenge, key stakeholders are increasingly realizing the need to focus on service quality as a measure to improve their competitive position in today's highly competitive environment. The purpose of this article is to present a study that examined the expectations and perceptions of instructors' usage of cloud computing based technology on overall quality of service (QoS). The article explores literature review that establishes the rationale and framework for this investigation, research methodology, data analysis, and results. A final section will include a summary of the findings, conclusions, and recommendations from the study.

BACKGROUND

Cloud Computing Environments

The explosive growth in computer usage by business, government, educational institutions, combined with global collaboration provided by the Internet, and competition has brought a considerable increase towards computer usage along

with the associated need to maximize the use of available resources while minimizing costs. One area of growing interest for meeting these needs is the use of cloud computing to centralize computing and information management functions for large, often geographically dispersed organizations. Users only need to pay for the services they actually use (Kim, Kim, Lee, & Lee, 2009). It offers potential benefits related to reductions of server/storage infrastructure and delivery of services (Leavitt, 2009). Some of the primary types of cloud computing services include infrastructure as a service, platform as a service, and software as a service (Leavitt, 2009; "National Institute," 2011). Leavitt (2009) also included a general group called services, which consist of storage, middleware, collaboration, and databases provided via the Internet. These technologies and services together comprise the majority of the types of computing services available from cloud computing, ranging from hardware and software services, to entire computing environments. Cloud computing offers potential benefits related to reductions of server/storage infrastructure and delivery of services (Kim et al., 2009; Robinson, 2009). Cloud computing can be highly beneficial in educational settings. Among the possible benefits is the enhanced usefulness of the existing technology (Erenben, 2009). With its emphasis on the delivery of low-cost or free applications anywhere on the Internet, cloud computing is a promising prospect for educational institutions faced with budget restrictions and mobile student population (Denton, 2012). This study builds on the SERVQUAL model, discussed next, to analyze

DOI: 10.4018/978-1-5225-2255-3.ch097

the significance of expectations and perceptions of instructors' usage of cloud computing technology in community colleges.

SERVICE QUALITY MEASUREMENT

According to Renganathan (2011), quality is a subjective concept that has no generally agreed definition for it. The word quality means different things to people according to the context. In general it is difficult to measure and quantify service quality. The main purpose of measuring service quality is to ensure whether service is provided as per the expectations of the customers. There are several well-known tools for measuring service quality or customer satisfaction. The most eminent instrument in attempting to systematize the service quality is "The gap model" of service or SERVQUAL developed by Parasuraman et al. (1985). This conceptual framework was developed initially to measure customer perception of service quality for the financial service sectors but later extended to sectors such as hospitality, telecommunications and healthcare. The SERVQUAL's model, which was developed by Parasuraman et al. (1988) used a survey to ask respondents for an indication of their expectations as well as their perceptions of service, and establishes the gap between the two. Other researchers, such as Cronin and Taylor (1992), held that only the perception of quality is important. The next section highlights how SERVQUAL has been used in universities to assess satisfaction not only with teaching and

learning, but with support services such as information technology (Smith et al., 2007).

MEASURING SERVICE QUALITY IN HIGHER EDUCATION

Frequently, higher education institutions seek to provide high quality services in all parts of their educational curricula and administrative processes. Therefore, the importance of service quality makes its measurement and its subsequent management an issue of utmost importance (Shekarchizadeh, Rasli & Hon-Tat, 2011).

The review of literature shows that some studies used the SERVQUAL model to measure service quality in higher education. Boulding, Kalra, Staelin, and Zeithaml (1993) used SERVQUAL model to study expectations and perceptions linked with the delivery of services in an educational environment. Their study used SERVQUAL to measure students' satisfaction with overall quality of service in a higher educational setting (Al-alak & Alnaser, 2012). Table 1 below shows hypotheses' testing results of their study. All but the sixth hypothesis was accepted. Hampton (1993) also used SERVQUAL model to measure college student satisfaction with professional service quality. In examining students' perceptions of service delivery, he applied the gap model (the disparity between expectations and experiences). These studies support the use of SERVQUAL model to measure instructors' usage of cloud computing technology

Table 1. Hypotheses' testing results

H		Result
1	There is a significant relationship between service quality dimensions and students satisfaction.	Accepted
2	There is a significant relationship between tangibles and students satisfaction.	Accepted
3	There is a significant relationship between reliability and students satisfaction.	Accepted
4	There is a significant relationship between empathy and students satisfaction.	Accepted
5	There is a significant relationship between assurance and students satisfaction.	Accepted
6	There is a significant relationship between responsiveness and students satisfaction.	Rejected

Source: Schwantz, 2012, p. 161, Table 4

Another study that supports the use of a modified version of SERVQUAL in academia to measure student satisfaction of the service they received was done by Schwantz (1996). Schwantz modified the usage of SERVQUAL instrument to make the comparison between traditional and non-traditional students' views of the quality of service in one higher educational institution. Students were asked to compare the quality of service (expected and received) of the support staff with that of faculty members. Based on factor analysis, the researcher identified the dimensions of the instrument where he used two dimensions instead of five, which are acknowledged by Parasuraman, Zeithaml and Berry (1990). The outcome of this study revealed no significant difference in the expectations or perceptions of traditional versus non-traditional students. There were no significant differences in students' expectations for support staff versus faculty. However, there was a significant difference in the students' perceptions of service quality from support staff versus service quality from faculty.

Other studies have borrowed some of the dimensions of SERVQUAL model to investigate the impact of a number of service quality attributes on satisfaction and loyalty in a higher education setting. Investigating the differences in student satisfaction and identifying dimensions of overall perceived quality, a study by Ong and Nankervis (2012) revealed that students with different academic performances perceived the impact of quality attributes on satisfaction differently compared with students with lesser performances. It was also shown that differences in overall satisfaction with educational experience were found among different lines of specializations. Drawing concepts from services marketing and assessment literature, Duque and Weeks (2010) developed a conceptual model to assess student learning outcome. It was found that student perceptions of educational quality had a noticeable impact on student satisfaction.

There has also been considerable research to re-examine the reliability and validity of SERVQUAL

(Asubonteng, McCleary & Swan, 1996; Brown, Churchill & Peter, 1993; Ladhari, 2008; Lam, 1997; Shahin, 2004). Ladhari (2008) suggested that industry-specific measures of service quality might be more appropriate than a single generic scale. He then encouraged researchers and scholars toward the development of an alternative industry-specific research instruments for measuring service quality. Lam (1997) found that the results are consistent with those reported in Babakus, Boller (1992), and Parasuraman et al. (1996), suggesting that both measures exhibit desirable levels of reliability and internal consistency. This view was echoed by Asubonteng et al., in their 1996 research: that SERVQUAL will predominate as a service quality measure.

Summarizing, although, SERVQUAL has been proven to be a reliable instrument for measuring expectations and perceptions of service quality (Parasuraman, Zeithaml & Berry, 1990; Parasuraman, Zeithaml & Berry, 1994) and most of studies have focused on students' satisfaction with overall quality of service and/or with professional service quality in high education. But the literature review reveals the lack of studies on instructors' satisfaction with cloud computing technology in educational settings. This study fulfils this need. The next section will describe the methodology used for the study followed by the reliability and validity of survey instrument and data analysis.

METHODOLOGY

The current study involves the use of a quantitative method to collect and analyze data received from the sample population regarding instructors' perception of the service quality provided by cloud computing based system in large community colleges in Texas. The use of quantitative methods ensure the current study is specific and narrow, whereby the researcher can uncover measurable, observable data on the variables. Quantitative research enables the collection of data from instruments with preset questions and responses, and

Table 2. Reliability analysis

CONSTRUCTS	Number of Items	Cronbach Alpha (α)
ETN (Service Quality Expectation on Tangibles)	6	0.869
ERL (Service Quality Expectation on Reliability)	4	0.949
ERS (Service Quality Expectation on Responsiveness)	5	0.974
EAS (Service Quality Expectation on Assurance)	6	0.909
EEM (Service Quality Expectation on Empathy)	5	0.968
PTN (Service Quality Perception on Tangible)	7	0.843
PRL (Service Quality Perception on Reliability)	4	0.955
PRS (Service Quality Perception on Responsiveness)	4	0.955
PAS (Service Quality Perception on Assurance)	5	0.982
PEM (Service Quality Perception on Empathy)	5	0.972

acquire data from a large population (Creswell, 2005). The study identified one primary research question that was used to guide this investigation, which is: Do the difference between instructors' perception and expectation significantly affect their perceived service quality of cloud computing based systems? The study collected and analyzed data based on the above research question from large community (or two year) colleges in Texas.

The target population for this study comprised of instructors or faculty members (referred to as 'participant') with sufficient experience using cloud computing technology in two year colleges in the State of Texas. Subjects were drawn mostly from faculty members' of three large community colleges (Dallas County Community College District (DCCCD), Houston College System (HCS), and Lone Star College System (LSC)) in Texas that provided Institutional Review Board approvals. The sample size is determined based on the size of the target population and the desired accuracy of the study. The target population is 11,395. A random sample of 470 email addresses of faculty were selected from the target population using the "Random Numbers Generator" feature of the SPSS statistical package. All the 470 instructors that were randomly selected received an online survey hosted by SurveyMonkey.com.

The model for this study leverages service quality (SERVQUAL) multi-item scale developed to

assess customer perceptions of service quality in service and retail businesses (Parasuraman et al., 1988). The scale breaks down the notion of service quality into five dimensions identified through this process and assessed using 22 item scale: Tangibles - physical facilities, equipment, staff appearance, etc.; Reliability - ability to perform service dependably and accurately; Responsiveness - willingness to help and respond to customer need; Assurance - ability of staff to inspire confidence and trust; and Empathy - the extent to which caring individualized service is given. Survey questions were designed based on the 22 questions of SERVQUAL. Some modifications to the wording were made to make them relevant to the cloud computing based environment.

RELIABILITY AND VALIDITY OF SURVEY INSTRUMENT

The reliability of each of the SERVQUAL's dimensions was assessed using Cronbach (1951)'s alpha as depicted in Table 2 - Reliability Analysis. The survey was also pre-tested for its reliability (Nunnally & Bernstein, 1994; Straub, 1989). Reliability in this context is the extent to which a measurement procedure is free from error. The estimation of alpha under the different number of items were examined using SPSS. The reliability coefficient

(Cronbach's alphas) for all the constructs ranges between 0.869 - 0.982. This supports reliability and face validity for the SERVQUAL scores for all the constructs.

The survey did not require personally identifying information. Anonymity was guaranteed by instructing participants to avoid placing their name, return address, or any identifying information on the survey. Strict controls over all data collected were maintained by not sharing the responses from any participants. Once the data was collected and downloaded, a Likert-scale type result spreadsheets/database of the survey instruments was generated. A codebook was built for this study describing each independent, dependent and other variables used in the data analysis. The responses to the variables were entered into the statistical applications software package - Mplus version 7.3 and IBM Statistical Package for the Social Sciences (SPSS) v22 – used for analysis.

DATA ANALYSIS AND RESULTS

According to Marshall and Rossman (1995), “Data analysis is the process of bringing order, structure, and meaning to the mass of collected data. It is a messy, ambiguous, time-consuming, creative, and fascinating process” (p.111). The survey responses from 301 participants were analyzed

using a mixture of statistical approaches in an effort to provide order, structure, and meaning to the survey data collected. Data was scanned for univariate and multivariate outliers, defined as values that are greater than 3.29 standard deviations from the mean (Stevens, 2009). Three participants were removed for being multivariate outliers. Another 46 participants were removed from the data collected for not completing major sections of the survey. A random sample of 470 potential participants was selected. From those, 301 participants (64%) took part in the study. Data analysis was conducted on 252 participants (54%) after removing sixteen (16%) percent of those responses that were incomplete or unusable.

The analysis of the data was reported using the research questions as a foundation. The analysis plan of hypothesis testing is shown in Table 3. The statistical data were analyzed using descriptive statistics (frequencies and percentages, mean, standard deviation, skewness, and kurtosis) and inferential statistics (correlation, shared covariance, structural equation modeling (SEM), and ANCOVAs). Descriptive research answers the questions of who, what, where, when, and how; however, it is not used to create a causal relationship (Gay, 1992). Table 4 presents the descriptive statistics for the expected, perceived, and differences scores.

Table 3. Analysis plan of hypothesis testing

Number	Hypothesis	Statistical Test
RQ1: Do the difference between instructors' perception and expectation significantly affect their perceived service quality of cloud computing based systems?		
H ₀₁	The difference between expectation and perception of reliability will not significantly load onto perceived service quality of cloud computing based systems - (DRL).	SEM & z-test
H ₀₂	The difference between expectation and perception of assurance will not significantly load onto perceived service quality of cloud computing based systems - (DAS).	SEM & z-test
H ₀₃	The difference between expectation and perception of tangibles will not significantly load onto perceived service quality of cloud computing based systems - (DTN).	SEM & z-test
H ₀₄	The difference between expectation and perception of empathy will not significantly load onto perceived service quality of cloud computing based systems - (DEM).	SEM & z-test
H ₀₅	The difference between expectation and perception of responsiveness will not significantly load onto perceived service quality of cloud computing based systems - (DRS).	SEM & z-test

Table 4. Descriptive statistics for expected, perceived, and difference scores

Score	Min	Max	M	SD	Skewness	Kurtosis
Expected (E)						
ETN	1.00	7.00	5.27	1.00	-1.03	1.87
ERL	1.00	7.00	5.83	1.37	-1.17	0.86
ERS	1.00	7.00	5.71	1.42	-0.98	0.17
EAS	1.00	7.00	5.71	1.27	-1.29	1.68
EEM	1.00	7.00	5.74	1.33	-1.11	1.09
Perceived (P)						
PTN	1.00	7.00	5.09	1.09	-0.42	0.37
PRL	1.00	7.00	5.69	1.35	-0.89	0.32
PRS	1.00	7.00	5.37	1.47	-0.60	-0.20
PAS	1.00	7.00	5.39	1.40	-0.84	0.51
PEM	1.00	7.00	5.32	1.42	-0.49	-0.23
Difference (E – P)						
DTN	-2.43	2.29	0.15	0.76	0.13	1.05
DRL	-2.25	3.00	0.08	0.81	0.43	1.93
DRS	-2.00	3.40	0.33	0.95	1.37	1.96
DAS	-2.33	3.17	0.33	0.82	1.18	2.07
DEM	-3.00	3.80	0.42	1.04	0.82	1.60

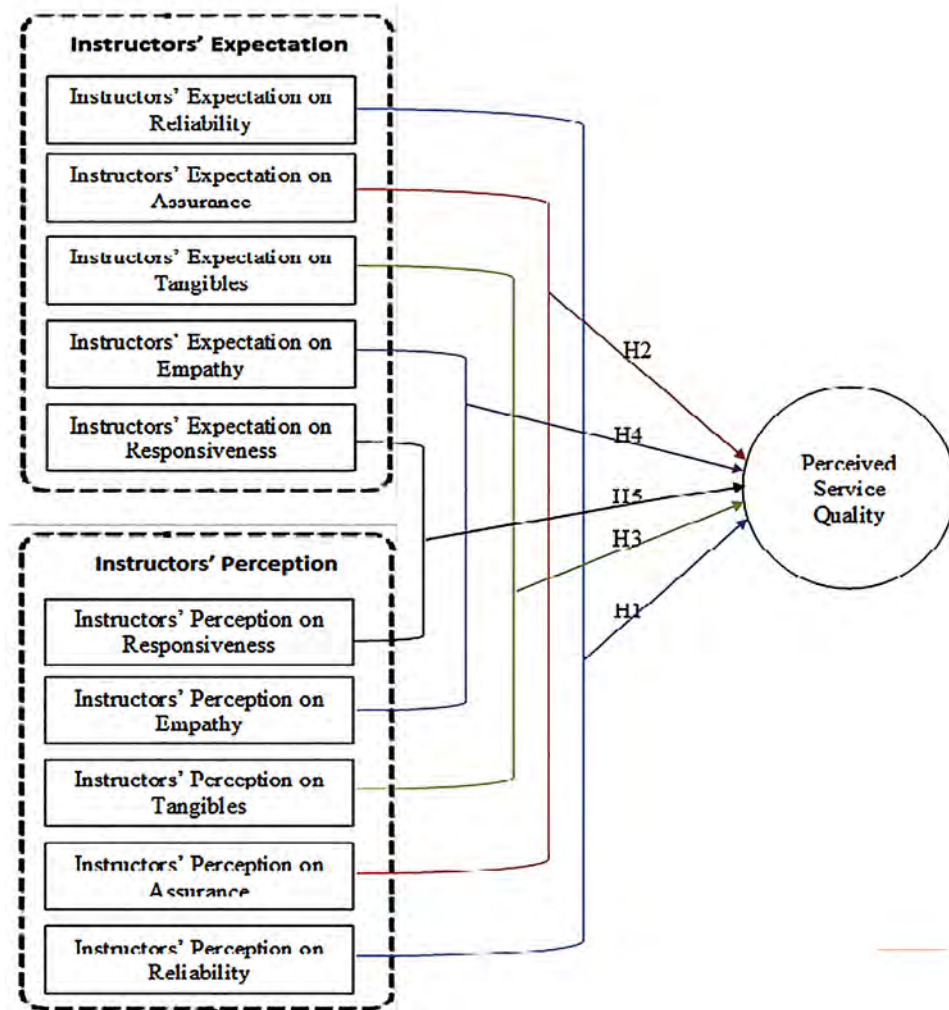
Prior to assessing the research question, the model fit of the empirical model (Figure 1) was examined through structural equation modeling (SEM) for goodness-of-fit. SEM was used to determine the model fit. Perceived quality of service (PSQ) is not a measured construct, and thus regression analysis is not possible. PSQ is a first order latent variable made up of the differences between expectation and perception of reliability, assurance, tangibles, empathy, and responsiveness.

To have a good model fit, the model should have a non-significant χ^2 statistic. However, since the χ^2 statistic can be unreliable for larger sample sizes, additional fit indices were also examined for to determine model fit (Kline, 2005). The comparative fit index (CFI) should be above 0.90. The root mean square error of approximation (RMSEA) should be below 0.10. Due to poor model fit, $\chi^2(5) = 32.36$, $p < .001$, CFI = .94, TLI = .88, RMSEA = .15, modification indices were examined to assess how the model can be improved empirically.

Modification indices provided ways to empirically improve the model. If the changes make theoretical sense, then the modification indices was tracked to improve the model to fulfill the requirements for a good model fit (Kline, 2011). Using SPSS, a statistical program, a correlation matrix was calculated to estimate the relationships among all the variables (i.e., between all possible pairs of variables) as shown in Table 5 - Correlation Matrix. The correlation for all paired variables are “statistically significant” at the 0.01 level (2-tailed).

In order to address research question 1 and hypotheses 1 – 5, structural equation model (SEM) (a confirmatory factor analysis (CFA)) was conducted to assess if the DTN, DRL, DRS, DAS, and DEM variables loaded onto the single perceived service quality (PSQ) latent construct. The data was entered into MPlus for analysis. The results provided a good model fit for the data, $\chi^2(4) = 15.01$, $p = .004$, CFI = .98, TLI

Figure 1. Conceptualization of model to be tested. Paths between expectation and perception represents the differences between these values



$=.94$, $RMSEA = .10$. Because the good model fit was found, the individual factor loadings for each of the variables was examined in order to address the research questions. The standardized estimates for the factor loadings were examined to determine the significance of each indicator. All indicator variables had a p value that was less than .001, thus showing significance. Because significance was found for each of the variables, null hypotheses 1 – 5 were all rejected in favor of the alternative hypotheses. Table 6 presents the results of the confirmatory factor analysis. Table 7 presents model fit information for the original

and modified models. Table 8 presents that results of hypotheses 1 – 5 tested. The next section will discuss future research directions.

FUTURE RESEARCH DIRECTIONS

The current study attempted to investigate instructors' perception regarding the service quality provided by cloud computing based system in large community (or two year) colleges in Texas as expected. Hence, it would be beneficial for future research to consider the following suggestions:

Table 5. Correlation matrix

Correlations											
	ETN	ERL	ERS	EAS	EEM	PTN	PRL	PRS	PAS	PEM	PSQ
ETN	1										
ERL	.734**	1									
ERS	.758**	.835**	1								
EAS	.742**	.770**	.852**	1							
EEM	.722**	.747**	.866**	.876**	1						
PTN	.609**	.531**	.607**	.635**	.656**	1					
PRL	.578**	.657**	.617**	.646**	.662**	.745**	1				
PRS	.604**	.583**	.679**	.654**	.731**	.825**	.792**	1			
PAS	.557**	.549**	.585**	.678**	.649**	.722**	.744**	.810**	1		
PEM	.541**	.528**	.580**	.605**	.684**	.766**	.719**	.870**	.854**	1	
PSQ	.442**	.409**	.414**	.474**	.472**	.587**	.528**	.618**	.640**	.656**	1

**. Correlation is significant at the 0.01 level (2-tailed).

Table 6. Parameter estimates for PSQ CFA model

Variable	Unstandardized Estimate	Standard Error	Standardized Estimate	p
DTN	1.00	-	.51	-
DRL	1.14	0.18	.54	<.001
DRS	2.31	0.30	.95	<.001
DAS	1.45	0.20	.68	<.001
DEM	1.90	0.25	.71	<.001
DEM with DAS	.15	0.04	.34	<.001

Note. Model: $\chi^2(4) = 15.01$, $p = .004$, CFI = .98, TLI = .94, RMSEA = .10.

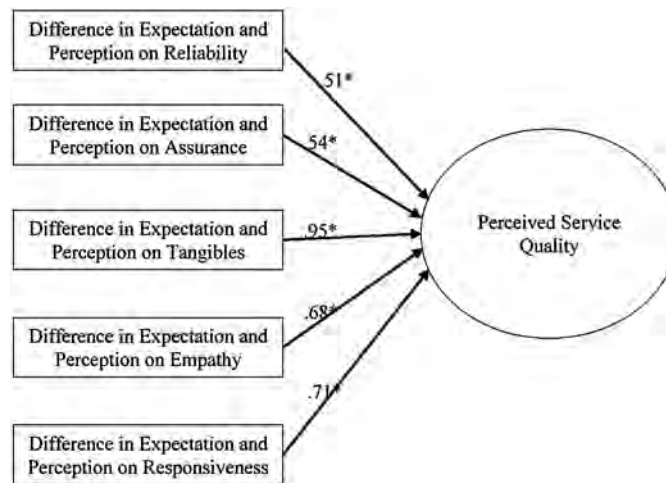
Table 7. Model fit statistics for original and modified models

Model	χ^2	df	p	CFI	TLI	RMSEA
Original	32.36	5	.001	.94	.88	.15
Modified	15.01	4	.004	.98	.94	.10

Table 8. Hypotheses' testing results

Number	Hypothesis	Result
RQ1: Do the difference between instructors' perception and expectation significantly affect their perceived service quality of cloud computing based systems?		
H ₀ 1	DRL	Rejected
H ₀ 2	DAS	Rejected
H ₀ 3	DTN	Rejected
H ₀ 4	DEM	Rejected
H ₀ 5	DRS	Rejected

Figure 2. SEM (CFA) results for PSQ



1. Further studies using the same methodology for the same population to examine the long-term implications of service quality improvement efforts.
2. Expansion of the study to include all two year colleges and not just large community colleges in the state to establish competitive benchmarks, track defections to other clouding computing providers caused by poor service delivery, and promotes a statewide service quality measurement and instructors satisfaction.
3. Evaluation can be made on the most common service quality measurement instruments in higher education. A comparative study will also be useful in this domain.
4. Additional exploratory, qualitative, and empirical research on the impact of instructor satisfaction vis-à-vis the wide variety of instructor demographic variables.
5. Further studies of the many types of service encounters, including service failures and recoveries, present in higher education.
6. An extension and testing of a model to measure internal customer satisfaction between service providers and institutional departments.

CONCLUSION

This study of instructor perceptions of service quality and perceived service quality (or satisfaction) of cloud computing technology in large community colleges in the State of Texas yielded support for the model tested, and expanded on previous service quality research in business and higher education. The findings of the study added inputs to the body of knowledge in higher education in Texas. Therefore, it is hoped that other colleges/universities will replicate the implications of this study.

It is quite evident that service quality has significant positive relationships with instructors' satisfaction based their continued use of cloud computing technology. Thus, instructors' satisfaction level can be enhanced by improving the service quality which confirms with the findings of other research work. This is critically important as community colleges have to compete with other higher academic institutions to serve the interests of the students. The results of the hypotheses tests are presented in Table 8. The results indicated that the differences between the expectation and perception on all five SERVQUAL dimensions load to the instructors' perceived service qual-

ity. Institutions should be held accountable for effectively meeting or exceeding instructors' expectations of the quality of services it provides.

REFERENCES

- Al-alak, B. A., & Alnaser, A. S. (2012). Assessing the relationship between higher education service quality dimensions and student satisfaction. *Australian Journal of Basic and Applied Sciences*, 6(1), 156–164.
- Asubonteng, P., McCleary, K. J., & Swan, J. E. (1996). SERVQUAL revisited: A critical review of service quality. *Journal of Services Marketing*, 10(6), 66–81. doi:10.1108/08876049610148602
- Babakus, E., & Boller, G. (1992). An empirical assessment of the SERVQUAL scale. *Journal of Business Research*, 24(3), 253–268. doi:10.1016/0148-2963(92)90022-4
- Babakus, E., & Mangold, W. G. (1992). Adapting the SERVQUAL scale to hospital services: An empirical investigation. *Health Services Research*, 26(6), 767–786. PMID:1737708
- Boulding, W., Kalra, A., Staelin, R., & Zeithaml, V. (1993). A dynamic process model of service quality: From expectations to behavioral intentions. *JMR, Journal of Marketing Research*, 30(1), 7–27. doi:10.2307/3172510
- Brown, T. J., Churchill, G. A., & Peter, J. P. (1993). *Improving the measurement of service quality* (Working Paper No. 92-94). Madison, WI: A.C. Nielsen Center for Marketing Research.
- Creswell, J. W. (2005). *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research*. Upper Saddle River, NJ: Pearson Education, Inc.
- Cronbach, L. J. (1951). Coefficient Alpha and the internal structure of tests. *Psychometrika*, 16(13), 297–334. doi:10.1007/BF02310555
- Cronin, J. J., & Taylor, S. A. (1992). SERVPERF versus SERVQUAL: Reconciling performance-based and perceptions-minus-expectations measurement of service quality. *Journal of Marketing*, 58(1), 125–131. doi:10.2307/1252256
- Denton, D. W. (2012). Enhancing instruction through constructivism, cooperative learning, and cloud computing. *TechTrends*, 56(4), 34–41. doi:10.1007/s11528-012-0585-1
- Duque, L. C., & Weeks, J. R. (2010). Towards a model and methodology for assessing student learning outcomes and satisfaction. *Quality Assurance in Education*, 18(2), 84–105. doi:10.1108/09684881011035321
- Erenben, C. (2009). Cloud computing: The economic imperative. *ESchool News*, 12(3), 13–19.
- Finn, D. W., & Lamb, C. W. (1991). An evaluation of the SERVQUAL scales in a retailing setting. *Advances in Consumer Research. Association for Consumer Research (U. S.)*, 18, 483–490.
- Garcia-Aracil, A. (2009). European graduates level of satisfaction with higher education. *Higher Education*, 57(1), 1–21. doi:10.1007/s10734-008-9121-9
- Gay, L. R. (1992). *Educational research: competencies for analysis and application* (4th ed.). New York: Macmillan Publishing Company.
- Hampton, G. (1993). Gap analysis of college student satisfaction as a measure of professional service quality. *Journal of Professional Services Marketing*, 9(1), 115–128. doi:10.1300/J090v09n01_10
- Kim, W., Kim, S. D., Lee, E., & Lee, S. (2009). *Adoption issues for cloud computing: ACM Proceedings of the 7th International Conference on Advances in Mobile Computing and Multimedia in Kuala Lumpur, Malaysia*. ACM.
- Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York: The Guilford Press.

- Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York: The Guilford Press.
- Ladhari, R. (2008). Alternative measure of service quality: A review. *Journal of Managing Service Quality*, 18(1), 65–86. doi:10.1108/09604520810842849
- Lam, S. S. K. (1997). SERVQUAL: A tool for measuring patients opinions of hospital service quality in Hong Kong. *Total Quality Management*, 8(4), 145–152. doi:10.1080/0954412979587
- Leavitt, N. (2009). Is cloud computing really ready for prime time. *Computer*, 42(1), 15–20. doi:10.1109/MC.2009.20
- Mashall, C., & Rossman, G. B. (1995). *Designing Qualitative Research*. Thousand Oaks, CA: Sage Publications.
- National Institute of Standards and Technology. (2011). *Cloud Computing*. Information Technology Laboratory. Retrieved from <http://www.nist.gov/itl/cloud/index.cfm>
- Nunnally, J. C., & Bernstein, I. (1994). *Psychometric theory*. New York, NY: McGraw-Hill.
- Ong, W. M., & Nankervis, A. (2012). Service quality in higher education: Students' perceptions in Australia and Malaysia. *Review of Integrative Business Economics Research*, 1(1). Retrieved from http://sibresearch.org/uploads/2/7/9/9/2799227/riber2012-223_277-290.pdf
- Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1985). A Conceptual Model of Service Quality and Its Implications for Future Research. *The Journal of Marketing*, 49(4), 41–50.
- Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1988). A multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing*, 64(1), 12–40.
- Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1990). *Delivering quality service: Balancing customer perceptions and expectations*. London: The Free Press.
- Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1991a). Understanding customer expectations of service. *Sloan Management Review*, 32(3), 39–48.
- Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1991b). Refinement and Reassessment of the SERVQUAL Scale. *Journal of Retailing*, 67(4), 420–450.
- Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1994). Reassessment of expectations as a comparison standard in measuring service quality: Implications for further research. *Journal of Marketing*, 58(1), 111–124. doi:10.2307/1252255
- Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1996). The Behavioral Consequences of Service Quality. *Journal of Marketing*, 60(2), 31–64. doi:10.2307/1251929
- Plank, R., & Chiagouris, L. (1997). Perceptions of quality of higher education: An exploratory study of high school guidance counsellors. *Journal of Marketing for Higher Education*, 8(1), 55–67. doi:10.1300/J050v08n01_05
- Renganathan, R. (2011). Service quality in hospitality services: Gap model and factor analysis. *European Journal of Soil Science*, 26(2), 159–175.
- Robinson, B. (2009). Gathering storm. *Federal Computer Week*, 23(1), 28–29.
- Schwartz, G. (1996). *Service quality in higher education: Expectations and perceptions of traditional and non-traditional students* (Doctoral Dissertation). Retrieved from repositories.tdl.org
- Shahin, A. (2004). *SERVQUAL and model of service quality gaps: A framework for determining and prioritizing critical factors in delivering quality services*. Retrieved from www.qmconf.com/docs/0077.pdf

Shekarchizadeh, A., Rasli, A., & Hon-Tat, H. (2011). SERVQUAL in Malaysian universities: Perspectives of international students. *Business Process Management Journal*, 17(1), 67–81. doi:10.1108/14637151111105580

Smith, G., Smith, A., & Clarke, A. (2007). Evaluating service quality in universities: A service department perspective. *Quality Assurance in Education*, 15(3), 334–351. doi:10.1108/09684880710773200

Stevens, J. P. (2009). *Applied multivariate statistics for the social sciences* (5th ed.). Mahwah, NJ: Routledge Academic.

Straub, D. W. (1989). Validating instruments in MIS research. *Management Information Systems Quarterly*, 13(2), 147–169. doi:10.2307/248922

Young, J. R. (2011). Colleges unite to drive down cost of cloud computing. *The Chronicle of Higher Education*, 58(9).

ADDITIONAL READING

Behrend, T. S., Johnson, E. C., London, J. E., & Wiebe, E. N. (2011). Cloud computing adoption and usage in community colleges. *Behaviour & Information Technology*, 30(2), 231–240. doi:10.1080/0144929X.2010.489118

Bowen, N. K., & Guo, S. (2012). *Structural equation modeling*. New York, NY: Oxford University Press.

Choy, J. Y., Lam, S. Y., & Lee, T. C. (2012). Service quality, customer satisfaction and behavioral intentions: Review of literature and conceptual model development. *International Journal of Academic Research*, 4(3), 11–15.

Freeze, R. D., & Raschke, R. L. (2007). An assessment of formative and reflective constructs in IS research. Conference: Proceedings of the Fifteenth European Conference on Information Systems, ECIS 2007, St. Gallen, Switzerland, 1481 – 1492.

Landrum, H., Prybutok, V., Zhang, X., & Peak, D. (2009). Measuring IS System Service Quality with SERVQUAL: Users' Perceptions of Relative Importance of the Five SERVPERF Dimensions. *International Journal of an Emerging Transdiscipline*, 12(3), 17–35.

Sultan, N. (2010). Cloud computing for education: A new dawn? *International Journal of Information Management*, 30(2), 101–182. doi:10.1016/j.ijinfomgt.2009.09.004

KEY TERMS AND DEFINITIONS

Cloud Computing Based Technology (or System): Includes information technology services and capabilities that are perceived as useful and accessible, which enable the users to provide beneficial services via technological interfaces.

Expectation (E): The performance anticipated or expected by the consumer. They form the reference point against which product or service performance is compared.

Perception (P): The customer's judgment about the service encounter – the actual service received.

Provider: An organization supplying cloud computing resources to outside users.

Service Quality or Perceived Service Quality (PSQ): Service quality is defined as the customer's perception of the level of success or failure in meeting expectations (P-E).

User: An organization or individual that uses cloud computing resources as a customer of a cloud computing provider.

Understanding Business Models on the Cloud



Arash Najmaei

Australian Catholic University, Australia

INTRODUCTION

The relationship between business models and technologies is an increasingly important issue. New technologies instigate new business models and new business models enable commercialization and success of new technologies. Building on this notion, this chapter synthesizes the literature on business models and cloud computing- a relatively new information technology- It shows that the relationship between business models and cloud-based systems create two distinctive technological trajectories: 1) cloud computing as a business model which can be seen as the development of cloud computing business models' (hereafter CCBM) and 2) cloud computing-enhanced or enabled business models which can be broadly thought of as traditional business modeling with improvements using cloud computing (hereafter BMCC). The former refers to a technological model or a system that commercializes value created solely by cloud-based systems in public and private sector. Whereas the latter refers to various ways that cloud-based systems can be and have been integrated into existing business models to enhance improve or enable creation and commercialization of new value. An explicit acknowledgement of these two trajectories and their underlying architecture are remarkably absent in the literature. Therefore, discussion outlined in this chapter is a timely contribution to the growing body of knowledge on how and why cloud-computing matters as a revolutionary technology.

The chapter proceeds as follows: first an overview of the concepts of business model within the framework of technology management is offered

in order to outline where the business modeling stands with respect to cloud computing as an emerging technology. Then cloud computing as an emerging technological paradigm will be discussed. Next, the architectural and structural ways through which business models and cloud computing systems are integrated will be illuminated in terms of two technological trajectories as briefly outlined above. Finally, the chapter concludes with a discussion on the implications of this conceptualization for theory and practice.

BACKGROUND

Technologies are developed to provide solutions to the problems humanity face. One of the most striking characteristics of our era is the exponential growth in the amount of information produced and the consequential need to process and store it. Information technologies refer to a family of technologies that solve such problems by enabling us to systematically collect, storage, clean, organize (mine), analysis, transform and present (visualization) information to satisfying our evolving needs.

Technologies do not emerge in vacuum. Their development is based on a purposeful orchestration of resources and application of knowledge. Business models play a fundamental role in this process. A business model defines the purpose of the technology and determines how resources are configured and orchestrated to bring the technology into the market (Chesbrough & Rosenbloom, 2002). When a technology is brought into the market it opens new avenues to develop new busi-

ness models (Teece, 2010). This interplay pushes the technological frontiers forward, makes old technologies obsolete and creates demand for new technologies (Najmaei, 2012, 2014b). Let's look at the information technologies as an illustration of this process. The traditional and still prevalent business model of ITs is based on investment in infrastructure. Simply said, if one needs to store and process data one has no choice but to buy or build a datacenter and a server to get these tasks done. This business model has worked well for years, and evolved into a global cluster of industries that offer advanced tools, systems and models all requiring heavy upfront investment and ongoing maintenance. In addition, technologies to process and store information are not easily available when we need them, where we need them.

Cloud computing (CC) is a technology developed based on a radically different business model aimed at solving these problems. It is a revolutionary approach to the provision of information technologies and radically changes the traditional business model of IT. CC offers a new way to make information technologies available to those individuals and organizations who demand it, when they need it, and where they need it by reducing the need for initial investment, hosting and ongoing maintenance (Hayes, 2008).

The idea behind CC is to bring the long-held dream of IT as a utility into life (Armbrust et al., 2009). CC is not only based on a new business model but also has a great potential to enable many new business models and radically change existing ones (Chang, Bacigalupo, Wills, & De Roure, 2010; DaSilva, Trkman, Desouza, & Lindic, 2013; Khanagha, Volberda, & Oshri, 2014; Lindgren & Taran, 2011; Weinhardt et al., 2009).

The constant interaction between existing and new technologies and business models epitomizes an evolutionary process which can be called technological business modeling (Najmaei, 2014a; Najmaei, Rhodes, & Lok, 2015). Drawing on this analogy, this chapter posits that the interplay between cloud computing and business models creates two distinctive technological trajectories

each with important theoretical and practical implications: 1) the business modeling with cloud computing and 2) the cloud computing business model. The insights from the literature on technological paradigms, clusters and trajectories will be used to develop this argument.

CLOUD COMPUTING, BUSINESS MODELS AND TECHNOLOGICAL PARADIGM

As noted, technologies are developed to solve problems. In doing so each technology follows a specific paradigm. A technological paradigm is a model or pattern of solution of selected technological problems (Dosi, 1982). Technologies related to each other in a given paradigm form technological clusters (Dosi, 1982). Technological clusters such as nuclear technologies, biotechnologies, semiconductor technologies and information technologies are, hence, a group of technologies sharing the same paradigm and addressing similar problems (Dosi, 1982). Technologies advance and evolve within a cluster along technological trajectories. A technological trajectory is the pattern of normal problem solving activity on the ground of a technological paradigm (Dosi, 1982). These trajectories represent evolutionary directions of technological changes within a paradigm until a paradigm shift takes place when paradigms collapse, change, merge or new paradigms emerge (Dosi 1982).

Technological trajectories and paradigms are closely related to business models (Teece, 2010). A business model is a logic or a set of assumptions that determines how well a technology fits in the market (Chesbrough & Rosenbloom, 2002). Technologies are brought to the markets when there is a demand for them. A market for a given technology is a place where users of the technology meet its sellers to exchange the rights to own and/or use the technology. This exchange can't happen without a business model in place. Therefore, a business model as a model is a sim-

plified representation of reality of the technology in a particular paradigm to be realized the market on a specific trajectory (Najmaei, 2014b). Take for example, online social networking technology. Business models of companies such as Facebook, Snapchat, WhatsApp and Twitter all represent a simplified reality of social networking on the web using firm-specific technologies which are all in one paradigm but evolving along different trajectories. The specific business models they use enable us to differentiate them from one another and at the same time relate them to each other. For a technology-developer or techno-preneur such as business model is essentially a mind-set or simply a blue print that describes how a technology creates value for its users and captures value for its sellers in an identifiable manner (Zott & Amit, 2013; Zott, Amit, & Massa, 2011). For a technology user such a business model is a means to place a business in a paradigm along a trajectory. Therefore, a business model is a bridge between suppliers of a given technology and its buyers. A well-functioning business model performs this functions by defining how a technology attracts buyers by showing them how it creates value and provides impetus for sellers by telling them how it is commercialized to capture its value. Both value creation and value capture are based on the assumptions of the paradigm within which the core technology of the business model is situated.

More specifically, creation of value refers to the way the business model uses the technology to offer new benefits, enhance execution of some hitherto difficult or complicated tasks, or to present an improved utility for users of the technology by simplifying jobs and solving problems at hand more efficiently (Bowman & Ambrosini, 2000). Creation of value is captured by the concept of 'use value' (Bowman & Ambrosini, 2000) which refers to the perceived usefulness of the technology by its users.

When value is created it should be captured as well. Capturing value refers to how a technology is commercialized in the market place in order to cover its costs and create profit by convinc-

ing users to pay for it (Bowman & Ambrosini, 2000). Value capture is tied to the profit formula associated with the technology and is simply represented by the concept of exchange value. That is the monetary value exchanged between seller and buyer for the use of the technology (Bowman & Ambrosini, 2000, 2007). As noted both value creation and capture are informed by the underlying assumptions of technological paradigm and what customers want, how they want it and how they can be enticed to pay for it (Teece, 2010). A technology can not address these assumptions on its own unless it uses a specific business model. Thus, the business model of the technology is an organizing framework which shows how a technology can achieve its full potential by creating and capturing value. To conclude, no technology can succeed without a business model and any given successful business model is based on a potent technology.

Constant changes in consumers' demands create new business models and new technologies. These changes take place within existing technological paradigms and progress along technological existing trajectories or create new ones. Many current technological paradigms and trajectories did not exist twenty years ago (e.g. share economy paradigm and technologies behind business model of firms such as air BnB). Drawing on this logic it can be argued that technological paradigms and trajectories are in fact business modelling trajectories and business modeling paradigms. Technologies that address related problems use similar business models that are based on or locked-in an existing dominant paradigm. Radical changes in technological capabilities or shifts in demands cause technologies and their business models to break out from of a dominant paradigm and develop an emerging one (Dolfsma & Leydesdorff, 2009). History of mobile phone technology and success of touch-screen smart phones is a vivid example of this phenomenon. The emergence and success of cloud computing is another manifestation of this phenomenon which creates two particular business modeling

trajectories. To understand this phenomenon, let's first, see what cloud computing technology is.

WHAT IS CLOUD COMPUTING?

The United States' National Institute of Standards and Technology (NIST) defines cloud computing as "....a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction." (Mell & Grance, 2011). It is also called 'on-demand computing', 'software as a service' (SaaS), or the 'Internet as platform' (Hayes, 2008). Regardless of its name, cloud computing as a technology subscribes to a new paradigm because it addresses computing problems differently. In a more technical term, the Cloud Computing paradigm creates a new approach to combine computing hardware and software. It involves "both the applications delivered as services over the Internet and the hardware and systems software in the data centers that provide those services" (Armbrust, Stoica, et al., 2010, p. 50). The services are referred to as Software as a Service (SaaS). The data center hardware and software is what is called a Cloud (Armbrust, Stoica, et al., 2010; Marston, Li, Bandyopadhyay, Zhang, & Ghalsasi, 2011b; Zhang, Cheng, & Boutaba, 2010). A Cloud can be made available in a pay-as-you-go manner to the general public which is then called a Public Cloud or be made as internal data centers of a business or other organization, not made available to the general public in the form of Private Cloud (Armbrust, Stoica, et al., 2010). Hence, the concept of cloud in its broadest terms is defined by a shift from locally installed programs to on-demand software which can be made available to various users on demand by being on the cloud. This shift which is fundamentally in the geography of technology is the defining feature of the cloud computing (Hayes, 2008). By virtue of this shift

in the logistics and geography of computing, the cloud computing disrupts the traditional IT paradigm which is essentially an internally developed, monitored and managed process. Consider for example the traditional office package as opposed to the cloud-based Google Docs: "when you create a spreadsheet with the Google Docs service, major components of the software reside on unseen computers, whereabouts unknown, possibly scattered across continents" (Hayes, 2008, p. 9). Thus, cloud computing represents emergence of a new paradigm in the science of providing information technologies.

HOW CLOUD COMPUTING CREATES VALUE

The value creating potential of cloud computing is simple. It opens numerous doors to create new 'use value' by delivering "all the functionality of existing information technology services (and in fact enable new functionalities that are hitherto infeasible) even as it dramatically reduces the upfront costs of computing that deter many organizations from deploying many cutting-edge IT services" (Marston et al., 2011b, p. 176). As such, Marston et al. (2011b, p. 178) argue that cloud computing as a new technology is suitable for:

- Businesses that use compute-intensive analytics,
- Businesses left behind IT revolution in less-developed countries,
- New businesses and expansion of established businesses by providing an almost immediate access to hardware resources, with no upfront capital investments for users, leading to a faster time to market in many businesses,
- Group of organizations with similar businesses because it becomes an adaptive infrastructure that can be shared by different end users, each of whom might use it in very different ways.

- Innovative companies because it lowers IT barriers to innovation,
- Growth-oriented companies by making it easier for businesses to scale their services. Since the computing resources are managed through software, they can be deployed very fast as new requirements arise. In fact, the goal of cloud computing is to scale resources up or down dynamically depending on client load with minimal service provider interaction.
- Developing new computing-intensive applications that were not possible before such as a) mobile interactive applications that are location-environment- and context-aware and that respond in real time to information provided by human users, nonhuman sensors (e.g. humidity and stress sensors within a shipping container) or even from independent information services (e.g. worldwide weather data); b) parallel batch processing, that allows users to take advantage of huge amounts of processing power to analyze terabytes of data for relatively small periods of time; c) business analytics that can use the vast amount of computer resources to understand customers, buying habits, supply chains and so on from voluminous amounts of data; and d) extensions of compute-intensive desktop applications that can offload the data crunching to the cloud leaving only the rendering of the processed data at the front-end, with the availability of network bandwidth reducing the latency involved.

Drawing on Marston et al. it can be argued that cloud computing changes the value creating logic of traditional IT business models in three ways. First, it challenges the prevalent assumptions about the accessibility and availability of IT. By making IT services available and accessible in new ways, perceived limitations in initial costing, investments, setup complications and speed as well as ease of delivery will be removed. Second,

cloud computing opens new horizons on making IT available and accessible in places where it was not possible or feasible before, making IT available to more users from a wider spectrum of demands. Finally, by offering IT on demand it changes the perceived usefulness of IT services as less-flexible and rigid offerings. All in all, cloud computing enables creation of new 'use value' by making IT-based offerings available to new people and new place in a more flexible and timely manner. That said, the next question is how cloud computing can capture new value created in these ways.

HOW VALUE OF CLOUD COMPUTING CAN BE CAPTURED

The value capturing potential of cloud computing lies in its ability to commercialize the dramatic reduction in the cost per unit of computing and the upfront costs of computing that deter businesses from using various IT services (Marston et al., 2011b). Cloud providers entice clients to join the cloud using this promise. They can then convert this revenue into profit because the clouds operate on extremely large-scale, commodity-computer data centers at low-cost location (Armbrust, et al., 2010). As Armbrust et al. (2010) argue, this factors "combined with statistical multiplexing to increase utilization compared to traditional data centers, meant that cloud computing could offer services below the costs of a medium-sized data center and yet still make a good profit" (p. 52). The profit formula is simple. Companies can access customized state-of-the-art computing services on a pay-as-you-go and/or subscription formats.

Let's have a look at three value-capturing scenarios. First, when demand for a service varies with time cloud computing pay by the hour for computing resources would lead to massive cost saving compared to traditional hosting (Armbrust, et al., 2010; Garrison, Wakefield, & Kim, 2015; Marston et al., 2011b). In the second scenario, demand is unknown in advance. In this case, subscribing to the cloud on an on-demand basis would give the

business a capacity to deal with operational complexity and uncertainty more efficiently (Armbrust et al., 2010). Finally, pay by the hour for computing enables analytics-intensive companies to finish computing faster when performing batch analytics. In all these scenarios and many similar ones, the value capturing potential of cloud computing is associated with its pay-as-you-go formula which companies treat as a conventional capital expense to operational expenses (Armbrust, Stoica, et al., 2010; Garrison et al., 2015; Hayes, 2008). All in all, reduced costs per unit of computing enables a more efficient capture of exchange values with higher margins. This would help cloud computing companies to entice more customers to pay for their IT services in an enhanced more flexible in pay-as- you-go and on-demand manner.

CLOUD COMPUTING AND BUSINESS MODELS: TWO TECHNOLOGICAL TRAJECTORIES

The foregoing sessions briefly discussed the value creation and capturing potential of cloud computing. As noted, a business model is a framework that combines these two. Business models make use of cloud computing by benefiting from the synergistic power created by the convergence of two forces. First is the “IT efficiency, whereby the power of modern computers is utilized more efficiently through highly scalable hardware and software resources” (Marston et al., 2011b, p. 177) and the second is the operational agility or strategic resiliency “whereby IT can be used as a competitive tool through rapid deployment, parallel batch processing, use of compute-intensive business analytics and mobile interactive applications that respond in real time to user requirements” (Marston et al., 2011b, p. 177). The notion of ‘business modeling’ is used to demonstrate how cloud computing and business models interact. Business modeling in its broadest meaning refers to the process of creating and constantly adjusting

business models to maintain their evolutionary fitness (Najmaei, 2014a; Najmaei et al., 2015). Understanding the business modeling side of cloud computing is just as important as understanding the technology side itself. Without a proper business modeling, cloud computing fails at fulfilling (i.e. creating use value) its potential and never succeeds at markets (i.e. capturing exchange value). If the cloud computing represents a paradigm shift in IT, then there would be two potential technological trajectories along which business models can be developed. More specifically, business modeling in the context of cloud computing technology takes two forms each representing a technological trajectory. These two are namely the business modeling with cloud computing (BMCC) and the cloud computing business models (CCBM). The next two sections explain these two trajectories.

BUSINESS MODELING WITH CLOUD COMPUTING (BMCC)

First, when cloud computing enables or enhances existing business models which use traditional IT, the process shall be called the business modeling with cloud computing or more simply cloud-computing enabled business modeling. This trajectory is growing rapidly on a worldwide scale as more businesses across industries are adopting the cloud computing and incorporating the cloud technology into their existing business models (Garrison et al., 2015; Marston et al., 2011b; McAfee, 2011; Zhang et al., 2010). Few examples include universities which adopt cloud-based online education systems, banks which use cloud to store and process transactions data, manufacturers which adopt cloud to manage their production across a diverse network of suppliers, retailers which use cloud to more quickly and efficiently manage their inventory and category management.

This trajectory is important because an enhanced speed, availability and efficiency of cloud computing enable existing business models to

become more agile and competitive. The challenge, however, is to transit from a traditional-IT enabled business models to cloud-enabled ones in a smooth and ideally least disruptive way. A study by KPMG in India suggests that privacy, security, cultural and structural resistance, execution risks and complexities and regulatory ambiguity are some of the most important challenges faced by executives when planning to execute this transformation (Parakala, Udhas, & Khanapurkar, 2011). Not every executive is well equipped and suited to lead this transformation: "...the true benefits of cloud computing will surprise you. But you'll need the right people to lead the transformation" (McAfee, 2011, p. 124). Although it is not an easy task and in many cases it has to be outsourced to cloud experts but given the current trend (Chang et al., 2010; Hashem et al., 2015; Marston, Li, Bandyopadhyay, Zhang, & Ghalsasi, 2011a; Prasanth, Bajpei, Shrivastava, & Mishra, 2015; Qian, Luo, Du, & Guo, 2009; Ranjan, Buyya, Nepal, & Georgakopoulos, 2015; Weinhardt et al., 2009; Zhang et al., 2010) it seems reasonable to argue that future belongs to cloud-enabled business models. Therefore, business modeling with cloud computing seems to be a worthy strategic direction to take. McAfee (2011) in support of this trend states: over time the economics of building and running a technology infrastructure will favour the cloud over on-premise computing" (p.131). Furthermore, business modeling with cloud computing not only permits many companies to concentrate on their core competencies while leaving the task of running IT infrastructure to the cloud service providers (Parakala et al., 2011) but also to develop new innovative competencies to target more customers in new markets. Through the development of virtual and "open" business processes, cloud computing enables its various stakeholders including customers, business partners, suppliers, etc. to connect and do new businesses more seamlessly (Parakala et al., 2011).

THE CLOUD COMPUTING BUSINESS MODEL (CCBM)

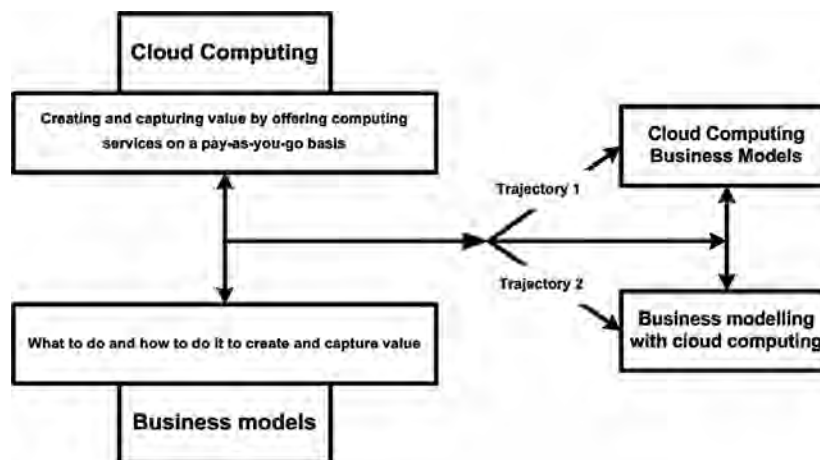
C

The first business modeling trajectory within the paradigm of cloud computing technology was about different ways that cloud computing could improve functionality of existing businesses and transform performativity of their business models. As interesting and potentially transformative as it sounds, this trajectory is only one side of the story. A look at the cluster of technologies formed around the cloud computing shows another trajectory that is just as potential to transform economies as the first one. This trajectory is not about cloud-enabled business modeling or transformation of other business models by adopting the cloud computing but is about formation of cloud-based business models. That is, developing business models that are entirely based on the cloud computing. These business models can be called 'cloud computing business models'. Some examples of such business models are Drop Box and Google Docs. They are independent entities whose existence depends entirely on the cloud computing. They did not exist before the invention of cloud computing. They create value by helping millions of people in both public and private sectors share, store and transfer files using the cloud technology. They also capture this value using various pay-as-you-go and subscription plans. In addition, some of the world's pioneering IT firms have begun to develop separate divisions, entities and spin-off companies with their own business models to tap into this market. Examples of such business models are the Google App Engine TM, AMAZON Web Service TM, and Microsoft Azure TM.

These business models are used by autonomous and independent entities which offer computing services over the internet to clients and have the ability to be joined to other related technologies including Grid Computing, Utility Computing, virtualization and Automatic Computing (Zhang et al., 2010).

The cloud computing business models are service-driven and can generally be grouped into

Figure 1. BMCC vs. CCBM



three different service business models which deliver Infrastructure as a Service (IaaS), Platform as a Service (PaaS), or Software as a Service (SaaS) (Zhang et al., 2010). These business models can be made available solely to the general public in the form of public clouds or to clients in the private sector in the form of private clouds to be used exclusively by a single organization or even in a hybrid form to both the general public and the private sector (Armbrust, Fox, et al., 2010; Prasanth et al., 2015; Zhang et al., 2010).

It is to be noted that, developing business models for cloud computing is different from but closely related to the process of business modeling with cloud computing. More specifically, the two trajectories converge on a technological path through which the cloud computing business models allow firms to do their business modeling with the help of cloud computing. The technological cluster shaping around these two trajectories enables this emerging paradigm to delineate its boundaries more clearly as a revolutionary rather than an evolutionary change in the IT paradigm which transforms both the provision of IT services (CCBM) and the way such services are adopted and used by organizations in both public and private sector (BMCC). The discussion outlined here has important implications for theory, practice and research which will be illustrated in the next section.

IMPLICATIONS AND FUTURE RESEARCH DIRECTIONS

Discussions of the link between the cloud computing technology and business models of companies which offer and/or adopt cloud computing are missing in the literature mainly because the boundaries between these two issues are somehow unclear. The primary contribution of this paper is to address this gap. It was argued that in the wake of cloud computing era the business landscape is changing by two business modeling forces: first the existing business models are transformed by the cloud computing. The cloud technology is making them more agile, adaptive and competitive. This trend is progressing along a technological trajectory which can be called business modeling with cloud computing (BMCC). Second, new business models are developed entirely based on the cloud computing breeding new entities and forming new industries and technological clusters. This trend is, however, progressing along a different trajectory which can be called the emergence of cloud computing business models. Figure 1 schematically illustrates this conceptualization.

Porter and Millar (1985) argue that IT changes business landscape in two interrelated ways: either by revolutionizing the established ways of doing business or by creating entirely new ways to do business. The two business modelling trajectories

within the cloud computing paradigm conform to this logic and point to an implication for the theory and practice of business in the era of cloud computing. The adoption of cloud computing seems essential especially in computing-intense industries, so is having a well-functioning business model. The challenge is to make sure that the cloud technology contributes to both value creation and value capturing aspects of the business model. A business model functions well only when its value creation and value capturing parts are well-aligned. Cloud computing enhances both creation and capture of value. The main challenge is to develop the internal alignment between these two (Low, Chen, & Wu, 2011; McAfee, 2011). This alignment is likely to disrupt the business modeling routines of a business but will pay off if properly planned and done under expert supervision. Note that, business modeling with cloud computing is path-dependent whereas the cloud computing business models are path creating. These two interact within the paradigm of cloud computing. Therefore, managers should keep abreast of cloud computing business models to successfully develop strategic scenarios when planning their business modeling with the cloud.

CONCLUSION

On a final thought, any new phenomenon is a fruitful field of research. Numerous unanswered questions exist with regard to the interplay between business models and cloud computing. First, little is known about how exactly small and large companies differ in their capacity to adopt cloud computing technologies. Similarly, little is known about how existing business models are transformed into cloud-enabled ones and how service companies differ from manufacturing ones in performing this process. Secondly, cloud computing business models have primary been developed secretly by large IT companies such as Microsoft and Amazon. Little investigation has been conducted to describe how these business

models have been planned, designed, executed and being managed. Addressing these questions requires exploratory, context-specific case studies which can enable rigorous theory-building and cumulative knowledge creation in both trajectories. This chapter is only a first step in this direction and should be regarded as a starting point for a lively debate on the theory and practice of business modeling in the context of cloud computing.

REFERENCES

- Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A., & Stoica, I. et al. (2010). A view of cloud computing. *Communications of the ACM*, 53(4), 50–58. doi:10.1145/1721654.1721672
- Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A., & Zaharia, M. (2009). *Above the Clouds: A Berkeley View of Cloud Computing*. UC Berkeley Reliable Adaptive Distributed Systems Laboratory.
- Armbrust, M., Stoica, I., Zaharia, M., Fox, A., Griffith, R., Joseph, A. D., & Rabkin, A. et al. (2010). A view of cloud computing. *Communications of the ACM*, 53(4), 50. doi:10.1145/1721654.1721672
- Bowman, C., & Ambrosini, V. (2000). Value Creation Versus Value Capture: Towards a Coherent Definition of Value in Strategy. *British Journal of Management*, 11(1), 1–15. doi:10.1111/1467-8551.00147
- Bowman, C., & Ambrosini, V. (2007). Firm value creation and levels of strategy. *Management Decision*, 45(3), 360–371. doi:10.1108/00251740710745007
- Chang, V., Bacigalupo, D., Wills, G., & De Roure, D. (2010). *A categorisation of cloud computing business models*. Paper presented at the 2010 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing. doi:10.1109/CCGRID.2010.132

- Chesbrough, H., & Rosenbloom, R. S. (2002). The role of the business model in capturing value from innovation: Evidence from Xerox Corporations technology spin-off companies. *Industrial and Corporate Change*, 11(3), 529–555. doi:10.1093/icc/11.3.529
- Chesbrough, H., & Rosenbloom, R. S. (2002). The role of the business model in capturing value from innovation: Evidence from Xerox Corporations technology spin-off companies. *Industrial and Corporate Change*, 11(3), 529–555. doi:10.1093/icc/11.3.529
- DaSilva, C. M., Trkman, P., Desouza, K., & Lindic, J. (2013). Disruptive technologies: A business model perspective on cloud computing. *Technology Analysis and Strategic Management*, 25(10), 1161–1174. doi:10.1080/09537325.2013.843661
- Dolfsma, W., & Leydesdorff, L. (2009). Lock-in and break-out from technological trajectories: Modeling and policy implications. *Technological Forecasting and Social Change*, 76(7), 932–941. doi:10.1016/j.techfore.2009.02.004
- Dosi, G. (1982). Technological Paradigms and Technological Trajectories. *Research Policy*, 11(2), 147–162. doi:10.1016/0048-7333(82)90016-6
- Garrison, G., Wakefield, R. L., & Kim, S. (2015). The effects of IT capabilities and delivery model on cloud computing success and firm performance for cloud supported processes and operations. *International Journal of Information Management*, 35(4), 377–393. doi:10.1016/j.ijinfomgt.2015.03.001
- Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Khan, S. U. (2015). The rise of big data on cloud computing: Review and open research issues. *Information Systems*, 47, 98–115. doi:10.1016/j.is.2014.07.006
- Hayes, B. (2008). Cloud computing. *Communications of the ACM*, 51(7), 9–11. doi:10.1145/1364782.1364786
- Khanagha, S., Volberda, H., & Oshri, I. (2014). Business model renewal and ambidexterity: Structural alteration and strategy formation process during transition to a Cloud business model. *R & D Management*, 44(3), 322–340. doi:10.1111/radm.12070
- Lindgren, P., & Taran, Y. (2011). Business Models and Business Model Innovation in a Secure and Distributed Cloud Clustering (DISC) Society. *Wireless Personal Communications*, 58(1), 159–167. doi:10.1007/s11277-011-0285-8
- Low, C., Chen, Y., & Wu, M. (2011). Understanding the determinants of cloud computing adoption. *Industrial Management & Data Systems*, 111(7), 1006–1023. doi:10.1108/02635571111161262
- Marston, S., Li, Z., Bandyopadhyay, S., Zhang, J., & Ghalsasi, A. (2011a). Cloud computing—The business perspective. *Decision Support Systems*, 51(1), 176–189. doi:10.1016/j.dss.2010.12.006
- Marston, S., Li, Z., Bandyopadhyay, S., Zhang, J., & Ghalsasi, A. (2011b). Cloud computing — The business perspective. *Decision Support Systems*, 51(1), 176–189. doi:10.1016/j.dss.2010.12.006
- McAfee, A. (2011). What Every CEO Needs to Know About The Cloud. *Harvard Business Review*, 89(11), 124–132. PMID:21370809
- Mell, P., & Grance, T. (2011). *The NIST Definition of Cloud Computing: Recommendations of the National Institute of Standards and Technology*. NIST Special Publication 800-145.
- Najmaei, A. (2012). Business Model: Strategic Logic of Disruptive Innovation. In R. Karim Mohammed (Ed.), *Strategic and Pragmatic E-Business: Implications for Future Business Practices* (pp. 100–131). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-1619-6.ch005
- Najmaei, A. (2014a). A Behavioral View of Business Modeling. In T. K. Das (Ed.), *Behavioral Strategy: Emerging Perspectives* (pp. 177–203). Charlotte, NC: Information Age Publishing.

- Najmaei, A. (2014b). Business Model Value Creation, Value Capture, and Information Technologies. In M. Khosrow-Pour (Ed.), *Encyclopedia of Information Science and Technology* (3rd ed.; pp. 549-557). Hershey, PA: IGI Global.
- Najmaei, A., Rhodes, J., & Lok, P. (2015). Practice-Based View of Business Modeling: Cognition and Knowledge in Action. In T. K. Das (Ed.), *The Practice of Behavioral Strategy* (pp. 77-104). Charlotte, NC: Information Age Publishing.
- Parakala, K., Udhas, P., & Khanapurkar, N. (2011). The Cloud: Changing the Business Ecosystem. KPMG India. Retrieved from kpmg.com/
- Porter, M. E., & Millar, V. E. (1985). How information gives you competitive advantage. *Harvard Business Review*, 63(4), 149-160.
- Prasanth, A., Bajpei, M., Shrivastava, V., & Mishra, R. G. (2015). Cloud Computing: A Survey of Associated Services. In *Cloud Computing: Reviews, Surveys, Tools, Techniques and Applications*. HCTL Open.
- Qian, L., Luo, Z., Du, Y., & Guo, L. (2009). Cloud Computing. *An Overview*, 5931, 626-631.
- Ranjan, R., Buyya, R., Nepal, S., & Georgakopoulos, D. (2015). A note on resource orchestration for cloud computing. *Concurrency and Computation*, 27(9), 2370-2372. doi:10.1002/cpe.3421
- Teece, D. J. (2010). Business Models, Business Strategy and Innovation. *Long Range Planning*, 43(2-3), 172-194. doi:10.1016/j.lrp.2009.07.003
- Weinhardt, C., Anandasivam, D.-I.-W. A., Blau, B., Borissov, D.-I. N., Meinel, D.-M. T., Michalk, D.-I.-W. W., & Stöber, J. (2009). Cloud computing—a classification, business models, and research directions. *Business & Information Systems Engineering*, 1(5), 391-399. doi:10.1007/s12599-009-0071-2
- Zhang, Q., Cheng, L., & Boutaba, R. (2010). Cloud computing: State-of-the-art and research challenges. *Journal of Internet Services and Applications*, 1(1), 7-18. doi:10.1007/s13174-010-0007-6
- Zott, C., & Amit, R. (2013). The business model: A theoretically anchored robust construct for strategic analysis. *Strategic Organization*, 11(4), 403-411. doi:10.1177/1476127013510466
- Zott, C., Amit, R., & Massa, L. (2011). The Business Model: Recent Developments and Future Research. *Journal of Management*, 37(4), 1019-1042. doi:10.1177/0149206311406265

ADDITIONAL READING

- Abolfazli, S. A. E. I. D., Sanaei, Z., Sanaei, M., Shojafar, M., & Gani, A. (2015). *Mobile cloud computing: The-state-of-the-art, challenges, and future research*. Encyclopedia of Cloud Computing. USA: Wiley.
- Chang, V., Kuo, Y. H., & Ramachandran, M. (2016). Cloud computing adoption framework: A security framework for business clouds. *Future Generation Computer Systems*, 57, 24-41. doi:10.1016/j.future.2015.09.031
- Chang, V., Walters, R. J., & Wills, G. B. (2016). Organisational sustainability modelling—An emerging service and analytics model for evaluating Cloud Computing adoption with two case studies. *International Journal of Information Management*, 36(1), 167-179. doi:10.1016/j.ijinfomgt.2015.09.001
- Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Khan, S. U. (2015). The rise of big data on cloud computing: Review and open research issues. *Information Systems*, 47, 98-115. doi:10.1016/j.is.2014.07.006

Pasupuleti, S. K., Ramalingam, S., & Buyya, R. (2016). An efficient and secure privacy-preserving approach for outsourced data of resource constrained mobile devices in cloud computing. *Journal of Network and Computer Applications*, 64, 12–22. doi:10.1016/j.jnca.2015.11.023

Singh, A., Mishra, N., Ali, S. I., Shukla, N., & Shankar, R. (2015). Cloud computing technology: Reducing carbon footprint in beef supply chain. *International Journal of Production Economics*, 164, 462–471. doi:10.1016/j.ijpe.2014.09.019

KEY TERMS AND DEFINITIONS

Business Model: The strategic logic that defines how a technology creates value and how its value can be captured.

Business Modeling: The process of continuously managing business models.

Business Modelling with Cloud Computing (BMCC): A business modeling process in which existing business models adopt cloud computing to improve their agility and functionality.

Cloud Computing Business Models (CCBM): A business modeling process in which new business models based entirely on the cloud computing (its value creation and capture potential) come into existence.

Cloud Computing: A new technological paradigm in IT that is based on computing services as utility and provides them on a pay-as-you-go basis to public and private sector businesses.

Technological Paradigms: A model or pattern of solution of selected technological problems.

Technological Trajectories: The pattern of normal problem solving activity on the ground of a technological paradigm.

Value Capture: The act of capturing profit from a technology by developing a profit formula that entices customers to pay the specified price to obtain the technology and converts the revenue into profit.

Value Creation: The act of creating a value proposition for a technology that clearly defines its benefits/value for customers and specifies how they can be perceived as useful and unique in the market place.

Understanding Cloud Computing in a Higher Education Context



Lucy Self

University of Sussex, UK

Petros Chamakiotis

University of Sussex, UK

INTRODUCTION

This chapter explores cloud computing (CC)—an internet-based type of computing which allows users to share resources which are provided to computers or other remote devices (Hasan, 2011)—within a higher education (HE) context. The study is premised on the view that university students increasingly rely on online resources, electronic media and computing applications for nearly all their work, which has been seen as improving the quality of university studies (e.g. Robinson & Schlegl, 2005). Given the limited number of studies in this area—reviewed in the next section—this chapter aims to advance understanding of CC in HE by adopting an innovative methodological approach involving suppliers, implementers, and end users of CC in a selected HE context. The study—which is qualitative and exploratory, and draws on interview data—informs the extant literature and offers recommendations for researchers, practitioners, and universities. In what follows, the authors discuss the extant literature and then move on to present the research study.

BACKGROUND

CC is a relatively new phenomenon in the technology industry, developed from traditional hosting, which is becoming pervasive in numerous sectors (Armbrust, 2010). Mell and Grance (2011) consider it “a model for enabling ubiquitous,

convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction” (p.2). It is not clear when the term CC was officially coined, because it has been used for many years in network diagrams to depict an area of uncertainty (Stevenson, 2009). It is believed the term was coined in 2006 when large companies such as Google and Amazon Web Services started naming their hosting services as CC (Regalado, 2011). CC has been viewed as a disruptive technology; as Krikos (2011) puts it, “*CC has all the markings of a disruptive technology—those that change the game as it’s currently played both by traditional software licensing companies and by private, on-premises datacenters*” (p.2). A disruptive technology often begins by only satisfying a niche segment of the market—for example CC being best suited to businesses ‘*born on CC*’—and later expanding to other sectors (Danneels, 2004).

From a technical perspective, the extant CC literature posits different CC service models:

- **Infrastructure as a Service (IaaS):** When the customer is able to provision processing, storage, networks, and other fundamental computing resources.
- **Software as a Service (SaaS):** Which allows the consumer to access the provider’s applications running on a **cloud solutions (CS)** but cannot manage the underlying infrastructure.

- **Platform as a Service (PaaS):** Where the customer can control the applications but not the infrastructure (e.g. Mell and Grance, 2011).

Further, CC types may also vary; from ‘*Private Cloud*’ (a high cost option but also the most secure, often preferred by banks and governments), through to ‘*Community*’ (shared by multiple organizations), ‘*Public*’ (allowing pay-per-use) and finally ‘*Hybrid*’ (a mix of the previous) (Chou, 2015).

However, this chapter is interested in CC in the HE sector in particular. Within this context, Sultan (2010) discusses the reasons that it adds value for universities: Firstly, the rapid amount of change in the technology industry puts a lot of pressure on universities to keep up with the ever changing software and hardware; this is not always easy for institutions as they are often under governmental pressure to reduce spending. Therefore, investing in CC prevents universities from having to purchase hardware, which is costly (Wheeler & Waggener, 2009). In turn, it also means that universities can reduce labor costs relative to the management of information technology (IT) systems. Moreover, it is difficult for universities to develop their own premise infrastructure to achieve the economies of scale that a CC provider can achieve, which also leads to cost savings (Katzan, 2010). By allowing a CC provider to maintain the infrastructure, the pressure is also taken off the IT staff at the institution (Ercan, 2010). Further, moving staff away from the low skill tasks can potentially allow them to be reallocated to more value adding activities (Low, Chen & Wu, 2011). Another important feature that CC can bring to universities is collaboration, by, for example, the sharing of resources, such as lecturers, learning content, and amongst students too. This could happen not just internally, but with other universities as well (Wheeler & Waggener, 2009).

In the UK, universities have begun to take CC seriously. For example, London South Bank University (2014, May 6) have recently adopted

CS, confirming a deal for £14 million for IBM’s Exceptional Student Experience. This has numerous benefits: it (a) incorporates analytics, mobile, social and IT security; (b) allows students to access university applications from mobile devices, and use social communities for file sharing; (c) encourages more interaction between students; and (d) gives access for teachers to assess and feedback online (Miller, 2008; Rao, Sasidhar & Kumar, 2012; Sasikala, 2011).

Katz (2008) sees CC in HE as more of a necessity. As he puts it, “*students will arrive on campus with their own IT architectures and service arrangements. These students... will have little use for or patience with college or university offerings that underperform or force them to lose precious connections to people and processes that they have accumulated since childhood*” (p.18). Therefore, whereas Sultan (2010) believes moving to CC is more of a logical step for universities, Katz (2008) argues they will have to out of necessity. Building on this view, Alabbadi (2011) proposes that universities should carefully assess budgets and challenges to choose CC infrastructure which is best suited to their current IT activities.

Recently, however, it has been argued that CC in the HE context does not come without challenges. For example, Culley and Panteli (2015) found that IT staff at universities may view CS as a threat for their jobs; thus, a repositioning of IT services in the HE sector might be necessary. As they put it,

[HE IT staff should] develop a new suit of capabilities, so that the IT department can function when most of the technology is delivered by the cloud [...] IT staff can act as the bridge understanding the business and the technology. (p.60)

Clearly, as it follows, CC is a new, prevalent and disruptive technology with unprecedented benefits for both universities and students and yet there is a clear lack of empirical evidence to enable understanding of CC within the HE context, since much of the extant literature is based around

the pre-implementation phase of CC in HE. Also, this literature is limited in, for example, the impact of CS on HE IT departments (Culley & Panteli, 2015). What about how this is experienced by others—e.g. developers and users—involved in CC in HE? What about its overall impact on HE? In this study, the authors take a more holistic approach by adopting a multi-perspective research approach and considering the perceptions of different types of individuals involved in CC within the HE context, which is described next.

THE RESEARCH STUDY

Methodological Approach

The study is qualitative due to the newness of this topic and involves a multi-perspective approach in terms of data collection in order to gain deep understanding of the phenomenon under investigation. Semi-structured interviews were conducted in February/March 2016 with three groups of participants:

1. **CS Suppliers:** Four employees of a global, Fortune-100 technology company, Epsilon (a pseudonym).
2. **CS Implementers:** Four employees of a UK university's IT department (Hudson University; a pseudonym).
3. **CS End Users:** Four students at Hudson University.

This multi-perspective approach is robust in that it offers the perceptions of both users and developers as well as other individuals involved in the implementation of CS within the HE context. A total of 12 in-depth, semi-structured interviews were conducted, involving open-ended questions which allowed for the participants' experiences and stories to emerge. The interviews lasted for about 40-50 minutes each and the interview data were analyzed following a manual, qualitative thematic analysis approach (Braun & Clarke, 2006).

Analysis was interpretive and inductive—guided largely by the participants' perceptions—and allowing for themes to emerge purely from the data, in line with the exploratory character of the study. In presenting the findings, the names of the participants as well as those of the selected organization (i.e. Epsilon) and university (i.e. Hudson) were anonymized using pseudonyms.

Research Findings

Presented first in this section are the research participants along with their main demo-biographical characteristics (Table 1).

Presented next as three stories are the findings from the qualitative analysis for each group.

Epsilon Staff

There was consensus among Epsilon participants that,

[CC] is the ability to have an underlying IT environment that supports innovation by providing flexibility, speed and agility to allow that innovation to meet the changing requirements of a client core business... (Bert)

Interestingly, Epsilon participants argued that CC constitutes a hidden technological force which serves as an enabler of efficient work that HE students largely benefit from; yet it remains unacknowledged:

I think [students] indirectly benefit from CS. CC is the enabler that gives [students] the agility, speed and flexibility, so something else that provides them benefit. Quite often what students don't realize, they think 'this is a fantastic mobile application or website,' what they don't think is, 'wow, I love the CC that this sits on.' (Bert)

Its disruptive character and potential value was also highlighted:

Table 1. Presentation of the research participants

	Pseudonym	Gender	Age Group	Tenure (in Years)	Subject of Studies	Area of Work*	Organizational Role
Epsilon Staff	Alex	Male	16-25	4	Business and Management	Professional Services	Cloud Technical Sales
	Bert	Male	36-45	18	Economics and Politics	IT Services	Head of Business Development of Cloud Division
	Charles	Male	46-55	5	Marketing	IT Services	Cloud Business Development Executive
	Derek	Male	46-55	30	Chemistry and Economics	Marketing, Sales and IT	Academic Initiative Leader for Europe
Hudson University IT Staff	Edward	Male	26-35	15	Psychology	Infrastructure Services	Head of Platform Development
	Fred	Male	36-45	15	Medieval Studies	Client Services	Senior Information Delivery Manager
	Graham	Male	36-45	20	General IT Training	Client Services	Software & Technical Support Officer
	Harry	Male	46-55	25	Engineering	Infrastructure Services	Head of Technology & Architecture
Hudson University Students	Ian	Male	16-25	4	Computer Science with Management	IT at a Cosmetics Company	HE Student
	Jack	Male	16-25	4	Business and Management	Technology Company	HE Student
	Karl	Male	16-25	4	Computer Science	Technology Company	HE Student
	Lewis	Male	16-25	4	Physics	N/A	HE Student

* For University of Hudson students, this section refers to the type of company within which they carried out their placement year (internship).

Having an alternative and compelling value proposition that competes with those sort of disruptive issues is important [for universities]. (Charles)

The participants here posited a shift in terms of the activities that will hereafter be undertaken by university IT departments:

IT staff [at universities] will spend less time managing PCs and physically locking stuff down. They will be spending more time on the network that enables people to have flexible access and they'll be spending more time on security. (Derek)

This led to the idea of physicality and the benefits afforded by CS in HE contexts. For example,

Every square foot they had would be better served serving their core business rather than housing a computing environment. (Bert)

The issue of security was also found to be related to the above:

A lot of universities have a lot of physical security, and that is fine apart from the fact that, as history has already proved to us, if people either can get past the physical security or your security risk is already inside your network – you've only got to look at people like Edward Snowden. (Derek)

The above quote that rather than have IT staff keep the datacenters secure, universities with CC could utilize those staff members more efficiently.

This way, CS could contribute to taking the focus of universities away from managing datacenters and putting it back on delivering education.

Moreover, Epsilon participants shared their views about the future impact of CS on HE. On the one hand, Charles argued that there is an expectation that more tailored and industry-specific CS will start being developed, for example bundled services aimed specifically at universities. Overall, the participants used the hotel industry and the case of Airbnb, Inc. in particular to stress the level of impact CC can have on a business model. The question that was raised was: Why would a university ‘own’ its students since CC can enable access to foreign universities and remove the issue of tuition?

In expressing his views about the future impact of CC on HE, Bert outlined two reasons:

One is the immediate value, speed and agility it brings and secondly the future of IT is served not just by the benefits that Cloud can bring today but the way in which Cloud will position data to drive cognitive analytics on top of it. The idea of universities using cognitive analytics in the future for better management could be used for things such as tracking students who are performing poorly or helping to pick modules for students based on their preferences.

Finally, Alex suggested the up and coming nature of CS in HE:

I think it's a snowball effect, once you see a number of universities leveraging Cloud services, you'll see the technology within universities increase even more.

Hudson University IT Staff

This group agreed that CC solutions enhance HE:

I think it is enhanced because there are things that you can do because of the scale of some CC

which would never feasibly be possible if you only ever have internal IT. (Harry)

The issue of resistance was posited among Hudson IT staff:

It's very difficult for an IT department who is used to having everything in its own sphere of control to have it outside of its sphere of control, even if that sphere of control is statistically likely to provide a higher up time than we do. (Edward)

However, IT Staff thought that moving workloads to the CC would not lead to redundancies; rather, it would encourage staff to retrain to become more useful in value adding areas:

Our plan was not to make anyone redundant but it's actually to redeploy them to more valuable work, moving up the value chain and getting people doing interesting valuable work because there's clearly value to complex projects. (Edward)

It was mentioned that these cultural fears of CC came from the early 2000s when there was a large shake up in technology jobs due to outsourcing.

Data protection and IT security seemed to be a concern. As a university, issues of intellectual property (IP) need to be kept very secure; however, the university is taking a risk-based model around CC as the benefits lost without CC would be too great for the university:

People in all areas have concerns about security, particularly in light of TalkTalk—there will always be, to an extent, the argument of we've done it ourselves for years and it's all been fine, how do we know it's still going to be fine if we move it to a Cloud service? (Harry)

The ability to supply teaching virtually in order to reduce the congestion on campus was found to be critical. With the increasing number of students, it's no longer becoming a viable option to purely work on an on-premise basis, therefore moving

to a CS are becoming more of a necessity rather than just a benefit; and this may lead to a potential change in the way in which teaching is delivered:

We need a certain number of students to be the kind of university that we have the aspiration to be and that means challenges for capacity on campus. It's expensive to invest in infrastructure but if we can mitigate that through delivering some classes remotely and reducing the need to be on campus then we're clearly in a good position [...] we need to think of new hybrid teaching models where some of the course content is delivered on-line or is delivered through webinars and virtual classroom learning. (Edward)

In terms of how CC businesses could target universities better, this came down to cost and licensing:

We are massively price aware, we aren't just choosing based on return-on-investment (ROI) we're choosing on opportunity cost on spend because we're constrained to the point if a product would give us a positive ROI, we might not be able to go ahead with it if there's something else we could put in if it'll give us more value. (Edward)

However—similarly to participants from Epsilon—Hudson IT staff commented on CC sustainability:

We have reached a sustainable model in which the university is continuing to invest which means [CC] is becoming more mature, more accessible and more usable and therefore of greater value to us as end users. (Harry)

Overall, it was found that adding value was more important than price; therefore, Hudson IT staff argued that bundling products together, e.g. Microsoft bundling Office licenses into their CC offering, could be of significant value.

Hudson University Students

When asked to define CS, students emphasized the possibilities for mobile work as well as for synchronous collaborative work that these afford:

Cloud to me is being able to be mobile whilst working. If I save a document to the Cloud it means that other people that have access to that Cloud can access it anywhere around the country and even globally. (Jack)

CC was found to make collaboration between students, and also between students and lecturers significantly easier:

It's easy off the Google Docs and you can email it off to the lecturer or they can review it and it allows the lecturer to then look on the file and add comments and review it. (Karl)

Further to the issue of collaboration, the opportunities for knowledge sharing that CC was seen as enhancing were thought of by the students as paramount. Students highlighted the inter-university aspect of this, arguing that CS may allow them to access materials from top universities at a global level, thus improving their learning experience significantly:

We already have [...] something called the '[Omega, a pseudonym] Physics Network' and they have a room that has cameras dotted all around it, a big projector and then there's ten or so in the [area of the UK] that can share lectures. (Lewis)

I definitely think it would. I would be able to use lectures from the best universities in the US and the best lecturers in Oxford and it would just be my normal course. (Graham)

DISCUSSION AND RECOMMENDATIONS

This section discusses (a) the themes that emerged from the study and their contribution to the extant literature; and (b) recommendations for cloud suppliers, universities and students. It has to be noted that interviewees spoke about different CC service models; for example, Epsilon staff shared their views from an IaaS perspective, whereas Hudson IT staff and students from SaaS and PaaS perspectives. Further, though all themes add equally to the literature, Figure 1 below shows the relationships between the different themes, i.e. the degree to which the different groups agree on certain themes.

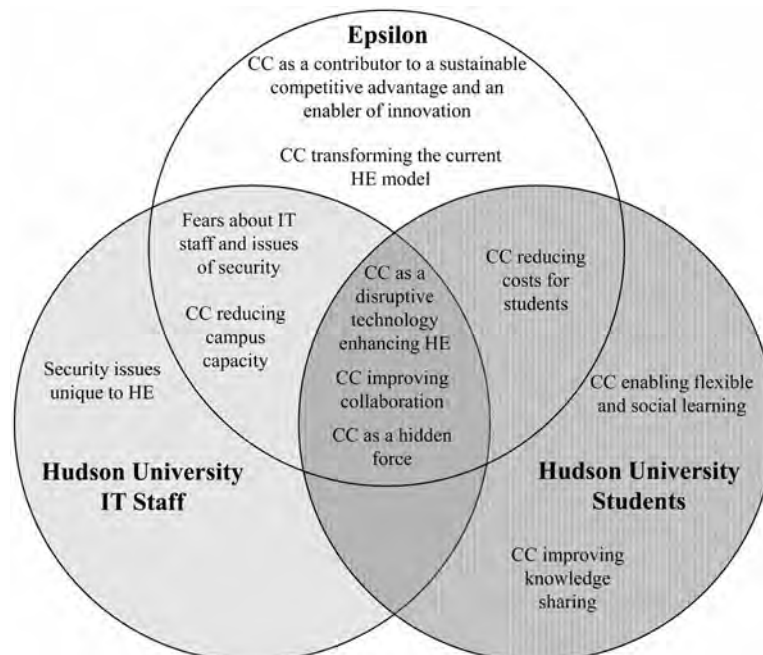
Emergent Themes

A dominant theme amongst the participants from Epsilon and Hudson IT staff was the idea of CC as a sustainable competitive advantage and its potential for value creation for organizations. It was purported that the more mature CC becomes,

the more value it adds to HE institutions. This expands the traditional literature on competitive advantage sustainability and cost reduction or differentiation as a driver of innovation (Porter, 1985), by explaining how CC can contribute to the development of a competitive advantage in HE. With a more granular look, CC capabilities in HE relate to the notion of temporary advantage (D'Aveni, Dagnino & Smith, 2010) suggesting that the volatility of today's markets only allows products to have an advantage for a limited time before another innovation emerges (Christensen, 1997).

Epsilon participants and Hudson IT staff both spoke about the fears around CC. These fears stem from cultures developed inside IT departments that believe outsourcing internal IT infrastructure to an external supplier would lead to jobs being lost. Some believed it was linked to the early 2000s when businesses saw a big shift to outsource their infrastructure. However, both groups stressed that this could be seen as an opportunity for current staff to develop different skills, corroborating and adding to recent literature speaking about employee

Figure 1. Relationships between the emergent themes from the three groups



resistance and also potential repositioning and new capabilities of universities' IT departments (Culley & Panteli, 2015; Low et al., 2011).

Security featured as another fear, especially after recent data leaks from big companies such as TalkTalk. The study stresses the added importance of this within the HE context due to IP issues characterizing universities' research activities. However, Epsilon and Hudson IT staff participants agreed that the benefits that HE institutions could obtain from a CS were greater than the limited chance of being hacked. IT staff also agreed that if someone really wanted to access their data, it would not be incredibly difficult for them to hack their on-premise infrastructure, thus taking a risk-based model towards CC is necessary. CC security is an area heavily covered in the literature, however, universities taking a risk-based model towards CC is not.

All three groups agreed that CC is a hidden force that will enhance education. Epsilon participants believed that this is down to the capabilities of CC which will allow higher levels of collaboration and will act as a base for cognitive analytics. IT staff claimed it would enhance education because it will allow greater scale than they could provide on-premise and students stressed they would benefit from it as it could potentially allow them to receive lectures over the internet from top institutions.

An idea put forward by Epsilon participants was that CC will completely alter the current model of HE Institutions by enabling new types of universities to set up online with much faster processes and cost far less for students. Or, similar to the cases of Airbnb, Inc. and Uber, lectures could be delivered virtually and tutors could be available locally. This may allow students to access the top lecturers globally without having to leave home. This is something that the IT staff agreed with and they added that if the university did not move quicker to tackle this pace of technology, they could seriously lose out. Students thought the idea of having access to top lecturers and cheaper tuition fees would be highly valuable. An

important issue suggested around this by Epsilon was that the universities that do not disrupt the current state now, will only be disrupted later.

An issue that follows from the aforementioned themes is the disruptive aspect of CC as a technological innovation (e.g. Krikos, 2011). These findings challenge deterministic approaches which see technology as a support activity within organizations, and instead highlight its potential to transform existing business models and bring value to organizations, as it is the case with CC in the HE context, which has been shown in the study presented here. Further to its theoretical significance, the study has recommendations for practitioners, which are outlined next.

Recommendations for Practitioners

Universities appreciate value creation and the simplicity of bundling multiple products and services into one license. It is important therefore for CS suppliers to consider this in order to ensure that the value added by their proposed CS has the potential to create sustainable value. If the current model for universities is disrupted by CS, the external suppliers will need to have a part in this process. Therefore, targeting universities effectively should take into consideration the educational boundaries that are being pushed. Further, a significant selling point that CC suppliers could consider relates to another concept mentioned by Epsilon with the idea of an '*Industry Cloud*'; a could offering based off of the needs of specific sectors, for example the HE sector.

It would overall be recommended that universities embrace CC for their own and their students' benefit. It is vital that any cultural perceptions within their IT departments be reassured in an effort to minimize uncertainty among IT staff. A repositioning of universities' IT departments and development of new capabilities for IT staff might prove fruitful. An acknowledgement that today's HE institution business model may be altered due to disruptive technology might also help. Willingness to accept this now will lead

to less disruption in the future; universities that do not technologically progress now might be disrupted by the universities that are willing to take a technical risk. It is finally recommended that university students embrace the opportunities afforded by CC by capitalizing on collaboration as well as accessing materials that exceed the boundaries of the traditional university.

FUTURE RESEARCH DIRECTIONS

The study presented here has limitations which give rise to directions for future research. First are methodological issues, for example, the issue of gender imbalance and limited statistical generalizability due to the relatively small sample. These can be overcome with larger studies and quantitative approaches as well as studies in other sectors, as CC is prevalent in numerous areas further to HE. From a theoretical perspective, researchers could build on the findings reported here to further explore the disruptive aspect of CC and its potential for the creation of a sustainable competitive advantage. Researchers could consider the relevance of the 'Business Model Canvas' (Osterwalder and Pigneur, 2010) examining such issues as value propositions, cost structures, and customer segments.

CONCLUSION

Overall, the exploratory study presented in this chapter was premised on the view that very little is known about CC in HE, despite its popularity and importance in this context. The study shed light on this issue by unpacking themes relating to CC in HE by considering the views of different groups of individuals (suppliers, implementers, end users). These themes add to our understanding in this area and—as discussed in the earlier sections—could be used to guide both future researchers, as well as practitioners in the field.

REFERENCES

- Alabbadi, M. M. (2011). Cloud computing for education and learning: Education and learning as a service (ELaaS). In *14th International Conference on Interactive Collaborative Learning (ICL)* (pp. 589-594). Piestany, Slovakia: IEEE. doi:10.1109/ICL.2011.6059655
- Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A., & Zaharia, M. et al. (2010). A view of cloud computing. *Communications of the ACM*, 53(4), 50–58. doi:10.1145/1721654.1721672
- Avram, M. (2014). Advantages and Challenges of Adopting Cloud Computing from an Enterprise Perspective. *Procedia Technology*, 12, 529–534. doi:10.1016/j.protcy.2013.12.525
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. doi:10.1191/1478088706qp063oa
- Chou, D. (2015). Cloud computing: A value creation model. *Computer Standards & Interfaces*, 38, 72–77. doi:10.1016/j.csi.2014.10.001
- Christensen, C. (1997). *The innovator's dilemma*. Boston, MA: Harvard Business School Press.
- Culley, G., & Panteli, N. (2015). Exploring the Impact of Cloud Computing on IT Departments. In *9th European Conference on IS Management and Evaluation: ECIME 2015* (p. 54-60). Bristol, UK: Academic Conferences and Publishing Limited.
- Danneels, E. (2004). Disruptive Technology Reconsidered: A Critique and Research Agenda. *Journal of Product Innovation Management*, 21(4), 246–258. doi:10.1111/j.0737-6782.2004.00076.x
- DAveni, R., Dagnino, G., & Smith, K. (2010). The age of temporary advantage. *Strategic Management Journal*, 31(13), 1371–1385. doi:10.1002/smj.897

- Ercan, T. (2010). Effective use of cloud computing in educational institutions. *Procedia: Social and Behavioral Sciences*, 2(2), 938–942. doi:10.1016/j.sbspro.2010.03.130
- Hassan, Q. (2011). Demystifying cloud computing. *The Journal of Defense Software Engineering*, (Jan/Feb), 16–21.
- Katz, R.N. (2008). The tower and the Cloud. *Higher Education in the Age of Cloud Computing*, 9.
- Katzan, H. Jr. (2010). The education value of cloud computing. *Contemporary Issues in Education Research*, 3(7), 37. doi:10.19030/cier.v3i7.219
- Krikos, A. (2011). Cloud Computing as a Disruptive Technology. *Cloudbook*, 2(2), 1–5.
- Low, C., Chen, Y., & Wu, M. (2011). Understanding the determinants of cloud computing adoption. *Industrial Management & Data Systems*, 111(7), 1006–1023. doi:10.1108/02635571111161262
- LSBU invests £14m to transform student experience through personalised educationLSBU invests £14m to transform student experience through personalised education. (2014, May 6). London South Bank University. Available at: <https://www.lsbu.ac.uk/about-us/news/lbsu-ibm-exceptional-student-experience>
- Marston, S., Li, Z., Bandyopadhyay, S., Ghalsasi, A., & Zhang, J. (2011). Cloud Computing: The Business Perspective. *SSRN Electronic Journal*, 51(1), 176–189.
- Mell, P., & Grance, T. (2011). *The NIST Definition of Cloud Computing*. Available at: <http://faculty.winthrop.edu/domanm/csci411/Handouts/NIST.pdf>
- Miller, M. (2008). *Cloud Computing: Web-Based Applications That Change the Way You Work and Collaborate Online*. Indianapolis, IN: Que.
- Osterwalder, A., & Pigneur, Y. (2010). *Business Model Generation*. Wiley.
- Porter, M. (1985). Technology and Competitive Advantage. *The Journal of Business Strategy*, 5(3), 60–78. doi:10.1108/eb039075
- Rao, N. M., Sasidhar, C., & Kumar, V. S. (2012). *Cloud computing through mobile-learning*. arXiv preprint arXiv:1204.1594
- Regalado, A. (2011). Who Coined ‘Cloud Computing’? *MIT Technology Review*. Available at: <http://www.technologyreview.com/news/425970/who-coined-Cloud-computing/>
- Robinson, A. M., & Schlegl, K. (2005). Student Use of the Internet for Research Projects: A Problem? Our Problem? What Can We Do About It? *Political Science and Politics*, 38(2), 311–315. doi:10.1017/S1049096505056544
- Sasikala, S., & Prema, S. (2011). *Massive centralized cloud computing (MCCc) exploration in higher education*. Academic Press.
- Schurr, A. (2008). *Keep an eye on cloud computing*. Network World. Available at: <http://www.networkworld.com/article/2281563/data-center/keep-an-eye-on-cloud-computing.html>
- Stevenson, D. (2009). *Why is it called “Cloud Computing”?* Available at: <http://it.toolbox.com/blogs/original-thinking/why-is-it-called-quotCloud-computingquot-30713>
- Sultan, N. (2010). Cloud computing for education: A new dawn? *International Journal of Information Management*, 30(2), 109–116. doi:10.1016/j.ijinfomgt.2009.09.004
- Wheeler, B., & Waggener, S. (2009). Above-campus services: Shaping the promise of cloud computing for higher education. *EDUCAUSE Review*, 44(6), 52–67.

ADDITIONAL READING

Culley, G., & Panteli, N. (2015). Exploring the Impact of Cloud Computing on IT Departments. In *9th European Conference on IS Management and Evaluation: ECIME 2015* (p. 54-60). Bristol, UK: Academic Conferences and publishing limited.

Ercan, T. (2010). Effective use of cloud computing in educational institutions. *Procedia: Social and Behavioral Sciences*, 2(2), 938–942. doi:10.1016/j.sbspro.2010.03.130

Katzan, H. Jr. (2010). The education value of cloud computing. *Contemporary Issues in Education Research*, 3(7), 37. doi:10.19030/cier.v3i7.219

Sultan, N. (2010). Cloud computing for education: A new dawn? *International Journal of Information Management*, 30(2), 109–116. doi:10.1016/j.ijinfomgt.2009.09.004

Wheeler, B., & Waggener, S. (2009). Above-campus services: Shaping the promise of cloud computing for higher education. *EDUCAUSE Review*, 44(6), 52–67.

KEY TERMS AND DEFINITIONS

C

Cloud Computing: A model enabling, convenient, on-demand network access to a shared pool of computing resources.

Cloud Solutions: The delivery of IT services, retrieved through Internet-based tools and applications rather than via a direct server connection.

Data Protection: Legal control over access to, and use of, data in computers.

Datacenter: A group of networked computer servers used for remote storage, processing, or distribution of large amounts of data.

Higher Education (HE): Tertiary education at a university leading to a recognized degree.

Information Technology (IT): The application of computers to store, retrieve, transmit and manipulate data.

IT Security: Protecting information from theft or damage, either via hardware or software and the information stored on them.

License: Agreement that stipulates the terms of use for an application and defines the rights of the software producer and of the end-user.

Virtually: Not physically existing, but made by software to appear to do so.

Vertical Integration Between Providers With Possible Cloud Migration

Aleksandra Kostic-Ljubisavljevic

University of Belgrade, Serbia

Branka Mikavica

University of Belgrade, Serbia

INTRODUCTION

The continuous growth of Internet traffic is significantly pushed by emerging high bandwidth demanding contents. All participants in content provisioning process including content providers, service providers, Content Delivery Networks (CDN) and customers are influenced by bandwidth requirements. Appropriate bandwidth demand estimation and consequently, network dimensioning are of great importance for addressing resource investment. However, peak bandwidth demand varies during the day. Provisioning of self-owned resources that satisfy peak bandwidth demand leads to network underutilization in the periods of low or normal traffic load. Hence, over-provisioning is cost ineffective. On the other hand, under-provisioning of network resources leads to rejection of customers' requests for service provisioning. Considering all aforementioned, vertically integrated providers in content provisioning process need to consider cloud migration in order to minimize costs and improve Quality of Service (QoS) and hence Quality of Experience (QoE) of their customers. Cloud providers maintain large-scale data centres to offer storage and computational resources in the form of Virtual Machines instances at a relatively low cost. Depending on the characteristics, cloud instances are available at different prices. In general, cloud providers offer three different pricing plans, such as reservation, on-demand and spot pricing. With the aim of obtaining

optimal integration charging strategy, concepts like Revenue Sharing, Cost Sharing, Wholesale Price etc. are applied frequently. Involvement of cloud providers into content provisioning process introduces additional complexity for choosing appropriate interconnection contract. The vertically integrated content provider's incentives for cloud migration can induce significant changes in integration contracts, and consequently in costs and requests' rejection rate.

In this chapter different methods of vertical integration charging among providers are analyzed and compared. With the aim of QoS and QoE improvement, providers can perform cloud migration. In that manner, cloud resources utilization concepts like reservation, on demand, and potential application of spot-pricing are going to be analyzed.

BACKGROUND

In order to provide content to the customers, vertical integration between providers is necessary. Term vertical integration refers to interconnection, i.e. physical and logical connecting, among providers operating at different network's levels. For instance, long-distance operator and local operator can be vertically integrated. Intrinsically, interconnection enables customers connected to one network to communicate with customers of the same or different network. It is a set of legal rules, technical and operational arrangements

which providers use to connect their equipment, networks and services. Vertical integration is economical, efficient and it enables achievement of economies of scale (Dai & Tang, 2009). Often applied contracts are Revenue Sharing, Cost Sharing and Wholesale Price. Instead of explicitly defined interconnection tariffs, operators often apply Revenue Sharing, which establishes fixed revenue share among providers. This type of contract is characterized with operational simplicity, and it can rebalance providers' returns when retail prices are distorted for any reason. Some of the greatest challenges that providers are dealing with are increasing profitability of the offered services, assuring higher charges for improved services and obtaining a fair share of the increased revenues. A fair Revenue Sharing contract based on the weighed proportional fairness criterion is proposed by He & Warland, (2006). They also show that non-cooperative strategies between providers may lead to unfair distribution of profit and may even discourage future upgrades to the network. Modelling of non-cooperative interaction between service providers and content provider as a Stackelberg game is performed by Wu, Kim, Hande, Chiang & Tsang (2011). These authors propose Revenue Sharing contract between service providers that jointly provide network connectivity between content provider and customers. They introduce profit division factor into the contract with the aim of social profit's maximization. Revenue Sharing contract between content provider and two service providers under network neutrality debate is observed by Coucheney, Maille, & Tuffin (2014). In this model, service providers enable direct connectivity to a fixed proportion of the content and compete in terms of price for customers. Relations between service providers are established using Revenue Sharing contract in order to maximize customers' welfare.

Revenue Sharing and Cost Sharing contracts are often formulated using Shapley value (Lee, Jang, Yi & Cho, 2013; Bogomolnaia, Holzman & Moulin, 2010). This mathematical method allocates resources according to contributions

of each party in a fair manner. Application of Shapley value in order to obtain a proper revenue distribution between content and service provider with elastic and inelastic demand is proposed by Ma, Chiu, Lui, Misra & Rubenstein (2008). The following formulations of Cost Sharing contract in backbone networks: volume-device, 95th percentile-device, 95th percentile-customer, customer peak-device, aggregate peak-device and Shapley-device, are analyzed and compared by Gyarmati, Sirivianos, & Laoutaris, (2012). The results show that Shapley-device contract better reflects the costs of the customers and improves fairness compared to other observed contracts. Key characteristic of Cost Sharing contract is enabling a compensation for the costs that provider incurs in carrying traffic generated by other providers.

Under Wholesale Price contract, service provider's payment to content provider for content provisioning depends on the established wholesale unit price and the traffic volume. The process of determining wholesale prices for mobile virtual networks operators and relations among vertically integrated incumbents when downstream entries are present under Wholesale Price contract are described by Song (2010). It appears that wholesale price can be the lower limit of the retail price, and determines the competitiveness of mobile virtual networks operators. Related regulatory issues have also been considered and it emphasizes the fact that regulators should have long-term perspective for the market.

In order to improve content provisioning process and assure satisfactory QoS and QoE, bandwidth demand estimation is essential. Content requires a certain minimal bandwidth which assures that customers' experience in content provisioning is satisfactory. Also, content has assigned content popularity factor, addressing the probability of access to certain content. In bandwidth demand estimation, parameter depicting content popularity is an important parameter and has to be included in network dimensioning process. Contents' differentiation according to its assigned popularity is of great importance for load

balancing, as well (Janaszka, Bursztynowski, & Dzida, 2012). Several observations on the suitable content popularity distribution can be found in the literature: Zipf (Zipf, 1932), Zipf-Mandelbrot (Tang, Fu, Cherkasova, & Vahdat, 2007), stretched exponential (Guo, Tan, Chen, Xiao, & Zhang, 2008), Zipf with exponential cut-off tail (Cha, Kwak, Rodriguez, Ahn, & Moon, 2007), etc.

In general, demand has stochastic characteristics. However, it appears that Internet traffic exhibits similar daily patterns and similar peak values every day. Demand attains the highest value in the evening hours, whereas the lowest demand occurs in the early morning hours. Provisioning capacity that satisfies total bandwidth demand is not economically efficient. Hence, content provider's resources would be idle most of the time. In order to avoid over-provisioning, servers' capacity might be constrained by parameter determining the portion of the peak bandwidth demand. The adequate bandwidth of content provider's resources is from 40% up to 60% of the peak bandwidth demand (Li et al., 2011). Intuitively, the less capacity provisioned, less resources are underutilized. However, this situation leads to greater rejection of customers' requests for content provisioning and hence, it must be taken in consideration. A promising solution for this challenge is cloud migration. Providers' incentives for partial cloud migration are analyzed by Li, Zhong, Liu, Li, & Xu (2011). Generic framework that facilitates cost-effective cloud migration of live streaming applications is provided by Wang, Liu, Chen, & Wang (2016). Cost-minimizing framework for migration of content distribution services into the cloud is presented by Qiu, Li, Wu, Li, & Lau, (2012). Costs for partial cloud migration and providers' profits obtained under different providers' integration contracts are analyzed by Mikavica & Kostic-Ljubisavljevic (2016).

Cloud resources are organized in the form of different types of Virtual Machine instances. Instances are available at different prices, depending on the applied cloud pricing plan. Two cloud

pricing plans that are widely implemented among cloud providers are: on-demand and reservation. On-demand pricing plan considers payment only for utilized resources, charging a fixed rate per billing cycle, usually on hourly basis. Cloud instances are available for indefinite time. However, cloud provider retains the right to terminate cloud instance. Unused time (less than an hour) whether end user or another provider terminated connection, charges as a full hour. If the instance is terminated by cloud provider, particular unused time is free of charge. Reservation cloud pricing plan enables reservation of cloud resources for a certain period of time, usually a year or three years. Beside upfront fee for resource reservation, cloud provider's customer pays discounted hourly resource usage price. Reserved resources are available at any time on request. Cloud pricing plans' classification is presented by Kansal, Singh, Kumar, & Kaushal, (2014). Lately, spot pricing has emerged as promising solution for enhancement of cloud provider's resources utilization and decreasing prices for cloud customers. Even with resources reservation and on-demand instances' utilization, some cloud resources remain in idle state. Thus, cloud provider offer its available resources, namely spot instances, in cloud market at lower price with ability to terminate instances when needed. Spot pricing presents dynamic pricing mechanism based on auctions (Toosi, Vanmechelen, Khodadadi, & Buyya, 2016) and reverse auctions (Roovers, Vanmechelen, & Broeckhove, 2012). Cloud customer creates a bid, representing maximum price willing to pay for specific instance. The access to spot instance is enabled as long as bid exceeds cloud provider's last computed market clearing price. Resource allocation and cloud pricing plans are closely related (Pal & Hui, 2013). Revenue maximization along with optimization of capacity allocation under different cloud pricing plans may be achieved by means of admission control (Toosi, Vanmechelen, Ramamohanarao, & Buyya, 2015).

VERTICAL INTEGRATION BETWEEN PROVIDERS

The comprehensive growth of Internet traffic is caused by emerging high bandwidth demanding contents such as video on demand, High Definition Television (HDTV), real time video, online gaming, file sharing and cloud computing. Participants in content provisioning process include content providers, service providers, Content Delivery Network (CDN) and customers. Generally, these participants may perform different combinations of functionalities depending on whether or to what degree an operator's business model implies vertical integration (BEREC Report BoR (12) 130, 2012).

Content providers' main functionality is content creation and aggregation. In addition, content providers establish their own hosting capabilities or deploy their own network infrastructure. Content providers encompass various players such as: platforms enabling transactions (e.g. Amazon, eBay, etc.), social platforms (e.g. Facebook), search engines (e.g. Google, Bing), live and on-demand radio and video services (e.g. broadcasters), entertainment services (e.g. Youtube, Dailymotion, etc.), application providers (e.g. Skype), video on demand (e.g. Netflix).

Generally, term service provider refers to operators selling broadband access and connectivity to the Internet at the retail level and at the wholesale level through transit and other forms of integration. These providers enable connectivity for different types of customers, whether it is end customer of content provider.

CDNs serve as aggregators of content on behalf of content providers. Their function is delivering content closer to the terminating network. CDN is a system of servers, placed at various locations at the network edges in order to enable efficient access from any customers' location. Customers' requests are redirected to the servers nearest to them. Different services provided by CDNs are targeted at the different types of content, such as video services, game services, software distribu-

tion updates, which content providers distribute. Each type of content has its own characteristics, such as data volume, upstream and/or downstream traffic, peak traffic, for which different CDNs can provide added value.

Term customer refers to both residential and business customers of a broadband access in their function of passively consuming content. Generally, customers form downstream traffic volume in consuming the content provided by the content providers. Overlapping between customers and content providers is possible. Players that predominantly act as customers may, in certain situations, act as content providers and provide content and services or applications, e.g. Youtube videos, Internet blogs, etc. Retail customers are private households whereas business customers may range from small to large business and industry customers.

If partial cloud migration is performed, content provider stores all original contents in self-owned servers and serves major part of the customers' requests for content provisioning. A certain part of the content provider's contents is migrated to the cloud resources provided by cloud provider. Cloud provider is a party which provides cloud services and performs activities necessary to ensure its delivery to the cloud customers as well cloud service maintenance. Term cloud service refers to one or more capabilities (application, infrastructure or platform capabilities type) offered via cloud computing invoked using a defined interface (ITU-T Recommendation Y.3500, 2014). In order to ensure satisfactory QoE, cloud provider delivers contents to the cloud CDN. Subsequently, cloud CDN delivers contents to the service provider using a global network. In the periods of low traffic load, content provider's resources can satisfy overall demand so data transfer through cloud CDN is not significant. However, in the periods of high traffic load, cloud CDN serves meaningful data volume. In order to assure the access to the most popular contents, content provider occupies cloud storage capacity according to appropriate cloud pricing plan. When demand exceeds capacity of the

content provider's resources, customer's request for content provisioning will be accomplished if required content is being stored into the cloud. Otherwise, customer's request for content provisioning is being rejected. Therefore, determination of the portion of the contents' replication on cloud has direct impact on requests' rejection, especially in the periods of high traffic load. In order to minimize risk of content unavailability, it is recommendable that cloud migrations occur in the period of low traffic load, so the content uploading to the cloud would not additionally load the system. Content replication during cloud migration assures that most popular contents are available even in the periods of high traffic load and it lowers number of rejected requests for content provisioning. The least rejection rate attains for the greatest content provider's capacity and the greatest portion of cloud migration.

Each provider aims to obtain the highest profit along with satisfactory QoS and QoE which, in long term, assures customers' loyalty. In general, providers' profits can be obtained by deducting corresponding costs from revenues. Service provider's revenue generated from the customers depends on the retail price and number of customers. Their costs largely depend on the selected integration contract. Prerequisite for selecting optimal vertical integration contract which improves content provisioning process is appropriate bandwidth demand estimation. In the context of partial cloud migration, vertical integration is being established among content and service provider. Relations among content and cloud provider are in accordance with the cloud pricing plans. Which cloud pricing plan will be chosen depends on the costs' prediction, demand, content provider's self-owned capacities and tolerance for requests' rejection rate. Vertical integration contracts that are often applied between providers are Revenue Sharing, Cost Sharing and Wholesale Price. Formulations of previously mentioned contracts used in this chapter are provided by Yoon, Yoo & Choi (2010).

In order to enable their customers the access to certain contents under Revenue Sharing contract, service provider shares its revenue with content provider. Hence, content provider's revenue under Revenue Sharing vertical integration contract depends on the service provider's revenue and on the parameter determining revenue share among them. The parameter depicting revenue share among providers is predefined constant value in the range $(0, 1)$. Greater revenue share assures greater revenue for content provider. The opposite state is valid for service providers.

Similarly to Revenue Sharing contract which defines how revenues of content and service provider are shared, Cost Sharing contract defines how costs in content provisioning process should be shared among providers. Extensive analysis of content and service provider's costs in the content provisioning process will be presented later. The parameter defining cost share is constant value in the range $(0, 1)$. Higher value of this parameter provides greater revenue for content provider. In addition, content provider's revenue under Cost Sharing contract depends also on profit margin which indicates a provider's profitability and takes values in the range $(0, 1)$.

The parameter profit margin coexists in the Wholesale Price contract with the same characteristics as in the Cost Sharing contract, as well. Hence, greater values for profit margin implicates greater content provider's revenue. Considering the fact that the same parameter coexists in Cost Sharing and Wholesale Price contracts, these contracts are easy to compare.

Content and service provider's costs are highly influenced with a selection of proper vertical integration contract. Service provider's costs can be divided into cost for customers' requests provisioning and vertical integration payment to content provider. Number of the customers' requests provisioned through cloud CDN depends on the portion of contents' replication on the cloud. Hence, partial cloud migration introduces additional costs. However, cloud migration has numerous short-term benefits including lower

requests' rejection rate, and long-term benefits which consist of customer's loyalty improvement. Analysis of content provider's costs indicates that enhancement of self-owned resources increases costs, which is primary motivation for cloud migration. The appropriate combination of content provider's self-owned resources and the amount of cloud migration should be selected in accordance with the realized profits depending on integration contracts between content and service provider. Content provider's costs in content provisioning with partial cloud migration can be divided into costs for serving part of requests for content provisioning from self-owned resources and costs for cloud assistance. Cloud instances are being occupied depending on the following factors: estimated future demand, necessary storage capacity for desired portion of the most popular contents on the cloud and offered cloud pricing plans. Hence, content provider's costs for renting cloud resources consist of: cost for reservation, if resource reservation is enabled; costs for resource utilizing, expressed as a fixed rate per billing cycle if on-demand or spot instances are being utilized and costs for data transfer from cloud CDN to the customers. Cost for content provisioning from cloud CDN to service provider's network is based on the transferred data volume.

Content provider's profits under Revenue Sharing vertical integration contract increases with revenue share enhancement. Greater content replication on cloud increase costs. Self-owned capacities enhancement increase costs, as well. Profit margin assures certain amount of content provider's revenue. In addition, cost share has considerable impact on profits. Greater profit margin provides greater profits. Likewise, greater cost share leads to greater profits. However, under Cost Sharing and Wholesale Price contracts, unlike Revenue Sharing, increasing of self-owned resources and content replication to cloud has minor impact on profits (Mikavica & Kostic-Ljubisavljevic, (2016). Considering service provider's profits under different vertical integration contracts and different levels of cloud migration

it is notable that higher values of revenue share, cost share or profit margin decrease profit.

Analyzing requests' rejection rate, costs and occupied resources under providers' vertical integration it is notable that cloud migration is preferable. However, content provider is imposed on additional costs. Considering that less popular contents are not likely to be requested, complete contents' replication on cloud would be cost ineffective. Hence, partial cloud migration which involves storing the most popular contents on the cloud is promising solution. In order to achieve the greatest cost savings, appropriate determination of content popularity and demand estimation is essential. In addition, cloud pricing plans in cloud migration may significantly affect content provider's costs, which in turn, may affect on proper selection of vertical integration contract with service provider. Hence, compromising solution is necessary.

FUTURE RESEARCH DIRECTIONS

Revenue Sharing vertical integration contract has proved to be preferable among Cost Sharing and Wholesale Price contracts, when partial cloud migration is performed (Mikavica & Kostic-Ljubisavljevic, 2016). However, only reservation cloud pricing plan has been observed. Hybrid pricing plan which involves partial resources' reservation and on-demand instances provisioning may decrease costs and reduce requests' for content provisioning rejection, concurrently. The impact of these pricing plans on selection appropriate vertical integration contract is subject of further research. Another promising solution for achieving goals of cloud provider's customers, i.e. providers and customers is spot pricing. However, there are some challenges related to spot pricing, including resources' availability and allocation, complexity for customer's understanding, optimal bidding in auctions etc. Migration to Superclouds allows relocation contents from terminating spot instance into the instance in another availability

zone, or even on another cloud provider. This can enhance content provisioning process performances. In addition to observation of different vertical integration contracts and their comparison, aforementioned issues will be addressed in future.

CONCLUSION

The aim of this chapter is to present complex issue of vertical integration between providers involved in content provisioning process. Undertakings that are concerned include content provider, service provider and cloud provider with corresponding CDN along with partial cloud migration. It has to be observed from different aspects. Primarily, appropriate bandwidth demand estimation is of crucial importance. Main reason for this estimation is to mitigate all disadvantages of under-provisioning or over-provisioning of resources. Requests' for content provisioning rejection rate is another relevant parameter since it significantly affects customers' experience in content provisioning process. Cost allocation and predicted profits are decisive factors in investment for partial cloud migration and selection appropriate vertical integration contract between providers. It appears that partial cloud migration can be cost-effective solution for providing contents with satisfactory QoS and QoE along with costs' minimization for maintenance of self-owned resources. Considering the fact that providers' main business goal is achieving the highest possible profits, selection of proper vertical integration contract is essential. In this chapter, Revenue Sharing, Cost Sharing and Wholesale Price contracts between content and service provider have been analyzed and compared. Introducing cloud provider in content provisioning process increases complexity in vertical integration issues, but it may offer (some) different optimal solutions in that process.

REFERENCES

- BEREC Report BoR (12) 130. (2012). *An assessment of IP interconnection in the context of Net Neutrality*.
- Bogomolnaia, A., Holzman, R., & Moulin, H. (2010). Sharing the Cost of a Capacity Network. *Mathematics of Operations Research*, 35(1), 173–192. doi:10.1287/moor.1090.0435
- Cha, M., Kwak, H., Rodriguez, P., Ahn, Y.-Y., & Moon, S. (2007). I Tube, You Tube, Everybody Tubes: Analyzing the World's Largest User Generated Content Video System. In *Proceedings of Internet Measurements Conference*, (vol. 7, pp. 1-14). San Diego, CA: ACM. doi:10.1145/1298306.1298309
- Coucheney, P., Maille, P., & Tuffin, B. (2014). Network Neutrality Debate and ISP Inter-relations: Traffic Exchange, Revenue Sharing, and Disconnection Threat. *Economics Research and Electronic Networking*, 15(3), 155–182. doi:10.1007/s11066-014-9090-3
- Dai, R., & Tang, S. (2009). Analysis on Vertical Integration of Telecom Operators. In *Proceedings of 2009 International Conference on Management of e-Commerce and e-Government*. Nanchang, China: IEEE. doi:10.1109/ICMeCG.2009.24
- Guo, L., Tan, E., Chen, S., Xiao, Z., & Zhang, X. (2008). The Stretched Exponential Distribution of Internet Media Access Patterns. In *Proceedings of Symposium on Principles of Distributed Computing*, (vol. 27, pp. 283-294). Toronto, Canada: ACM. doi:10.1145/1400751.1400789
- Gyarmati, L., Sirivianos, M., & Laoutaris, N. (2012). Sharing the Cost of Backbone Networks: Simplicity vs. Precision. In *Proceedings of NetEcon 2012 Workshop, Part of INFOCOM*, (pp. 171-176). Orlando, FL: IEEE. doi:10.1109/INFCOMW.2012.6193483

- He, L., & Warland, J. (2006). Pricing and Revenue Sharing Strategies for Internet Service Providers. *IEEE Journal on Selected Areas in Communications*, 24(5), 942–951. doi:10.1109/JSAC.2006.872875
- ITU-T Recommendation Y.3500. (2014). *Information technology – Cloud computing – Overview and vocabulary*. ITU.
- Janaszka, T., Bursztynowski, D., & Dzida, M. (2012). On Popularity-Based Load Balancing in Content Networks. In *Proceedings of International Teletraffic Congress*, (vol. 24, pp. 1-8). Krakow, Poland: IEEE.
- Kansal, S., Singh, G., Kumar, S., & Kaushal, S. (2014). Pricing Models in Cloud Computing. In *Proceedings of the International Conference on Information and Communication Technology for Competitive Strategies*. Udaipur, India: ACM.
- Li, H., Zhong, L., Liu, J., Li, B., & Xu, K. (2011). Cost-effective Partial Migration of VoD Services to Content Clouds. In *Proceedings of the International Conference on Cloud Computing* (vol. 4, pp. 203-210). Washington, DC: IEEE. doi:10.1109/CLOUD.2011.41
- Ma, R. T. B., Chiu, D., Lui, J. C. S., Misra, V., & Rubenstein, D. (2008). Interconnecting Eyeballs to Content: A Shapley Value Perspective on ISP Peering and Settlement. In *Proceedings of the 3rd International Workshop on Economics of Networked Systems* (pp. 61-66). Seattle, WA: ACM. doi:10.1145/1403027.1403041
- Mikavica, B., & Kostić-Ljubisavljević, A. (2016). Interconnection Contracts between Service and Content Provider with Partial Cloud Migration. *Elektronika ir Elektrotehnika*, 22 (6), 92-98.
- Pal, R., & Hui, P. (2013). Economic Models for Cloud Service Markets: Pricing and Capacity Planning. *Theoretical Computer Science*, 496, 113–124. doi:10.1016/j.tcs.2012.11.001
- Qiu, X., Li, H., Wu, C., Li, Z., & Lau, F. C. M. (2012). Cost-Minimizing Dynamic Migration of Content Distribution Services into Hybrid Clouds. In *Proceedings of the International Conference on Computer Communications: Mini-Conference*, (vol. 31, pp. 2571-2575). Orlando, FL: IEEE.
- Roovers, J., Vanmechelen, K., & Broeckhove, J. (2012). A Reverse Auction Market for Cloud Resources. In *Proceedings of the 9th International Conference Economics of Grids, Clouds, Systems, and Services*, (pp. 32-45). Berlin, Germany: Springer-Verlag Berlin. doi:10.1007/978-3-642-28675-9_3
- Song, J. (2010). Various wholesale price equilibria for mobile virtual network operators. *Telecommunications Policy*, 34(10), 633–648. doi:10.1016/j.telpol.2010.07.002
- Tang, W., Fu, Y., Cherkasova, L., & Vahdat, A. (2007). Modeling and Generating Realistic Streaming Media Server Workloads. *Computer Networks*, 51(1), 336–356. doi:10.1016/j.comnet.2006.05.003
- Toosi, A. N., Vanmechelen, K., Khodadadi, F., & Buyya, R. (2016). An Auction Mechanism for Cloud Spot Markets. *ACM Transactions on Autonomous and Adaptive Systems*, 11(1), 25–57. doi:10.1145/2843945
- Toosi, A. N., Vanmechelen, K., Ramamohanarao, K., & Buyya, R. (2015). Revenue Maximization with Optimal Capacity Control in Infrastructure as a Service Cloud Markets. *IEEE Transactions on Cloud Computing*, 3(3), 261–274. doi:10.1109/TCC.2014.2382119
- Wang, F., Liu, J., Chen, M., & Wang, H. (2016). Migration Towards Cloud-Assisted Live Media Streaming. *IEEE/ACM Transactions on Networking*, 21(1), 272–282. doi:10.1109/TNET.2014.2362541

Wu, Y., Kim, H., Hande, P. H., Chiang, M., & Tsang, D. H. K. (2011). Revenue Sharing among ISPs in Two-Sided Markets. In *Proceedings of International Conference on Computer Communications*, (vol. 30, pp. 596-600). Shanghai, China: IEEE. doi:10.1109/INFCOM.2011.5935234

Yoon, Y. S., Yoo, J., & Choi, M. (2010). Revenue Sharing is the Optimal Contractual Form for Emerging App Economy. In *Proceedings of International Conference on Information and Communication Technology Convergence*, (pp. 329-334). IEEE.

Zipf, J. K. (1932). *Selective Studies and the Principle of Relative Frequency in Language*. Cambridge, MA: Harvard University Press. doi:10.4159/harvard.9780674434929

ADDITIONAL READING

Al-Roomi, M., Al-Ebrahim, S., Buqrais, S., & Ahmad, I. (2013). Cloud Computing Pricing Models: A Survey. *International Journal of Grid and Distributed Computing*, 6(5), 93–106. doi:10.14257/ijgdc.2013.6.5.09

Bilo, V., Fanelli, A., Flammini, M., Melideo, G., & Moscardelli, L. (2010). Designing Fast Converging Cost Sharing Methods for Multicast Transmissions. *Theory of Computing Systems*, 47(2), 507–530. doi:10.1007/s00224-009-9207-5

Li, Z., Zhang, H., OBrien, L., Jiang, S., Zhou, Y., Kihl, M., & Ranjan, R. (2016). Spot Pricing in the Cloud ecosystem: A Comparative Investigation. *Journal of Systems and Software*, 114, 1–19. doi:10.1016/j.jss.2015.10.042

Moulin, H. (2013). Cost Sharing in Networks: Some Open Questions. *International Game Theory Review*, 15(2), 1–10. doi:10.1142/S021919891340001X

Pooyan, J., Ahmad, A., & Pahl, C. (2013). Cloud Migration Research: A Systematic Review. *IEEE Transactions on Cloud Computing*, 2(1), 142–157.

Wang, X., Wang, X., Che, H., Li, K., Huang, M., & Gao, C. (2015). An Intelligent Economic Approach for Dynamic Resource Allocation in Cloud Services. *IEEE Transactions on Cloud Computing*, 3(3), 257–289. doi:10.1109/TCC.2015.2415776

Wei, L., & Yang, X. (2015). Revenue Sharing Contract in a Cloud Computing Service Supply Chain with a Monopoly AIP and Multiple Competing ASPs under Asymmetric Information. In *Proceedings of International Conference on Service Systems and Service Management*, (vol. 12, pp. 1-6). Guangzhou, China. IEEE.

Xu, H., & Li, B. (2013). Dynamic Cloud Pricing for Revenue Maximization. *IEEE Transactions on Cloud Computing*, 1(2), 158–171. doi:10.1109/TCC.2013.15

Xu, Y., & Bisi, A. (2012). Wholesale-price contracts with postponed and fixed retail prices. *Operations Research Letters*, 40(4), 250–257. doi:10.1016/j.orl.2012.04.001

KEY TERMS AND DEFINITIONS

Cloud Migration: Storing a portion of the most popular content provider's contents into the cloud resources with the aim of enabling access to those contents whenever capacity of the self-owned servers is occupied.

Cost Sharing: Integration contract which allocates costs to all undertakings in a fair manner. Cost share among providers is defined in accordance with contribution in total costs of each provider in content provisioning process.

On-Demand: Pricing plan which enables cloud providers' customers to pay only for utilized resources per time unit, usually on the hourly basis. Cloud instances are retained for indefinite time.

Reservation: Cloud providers' customers pay an upfront reservation fee in order to reserve cloud resources for a specific period of time. In return, they receive a significant discount on the hourly resource usage price.

Revenue Sharing: Integration contract which defines modality of revenue shares between providers involved in all forms of provider's integration. It is characterized with operational simplicity and possibility of rebalancing the returns of the providers when retail prices are distorted.

Spot Pricing: An auction-like mechanism where cloud providers' customers submit bids, i.e. maximum price they are willing to pay for specific cloud instance. The access to cloud instances is enabled as long as submitted bid exceeds cloud provider's last computed market clearing price.

Vertical Integration: The term refers to inter-connection between all undertakings involved in content provisioning process, enabling physical and logical connecting with the aim of providing the access to contents.

Wholesale Price: Integration contract among providers based on the determination of the unit wholesale price for a given service. Demand and traffic volume are highly dependent on defined wholesale price.

C

Virtualization as the Enabling Technology of Cloud Computing

Mohamed Fazil Mohamed Firdhous
University of Moratuwa, Sri Lanka

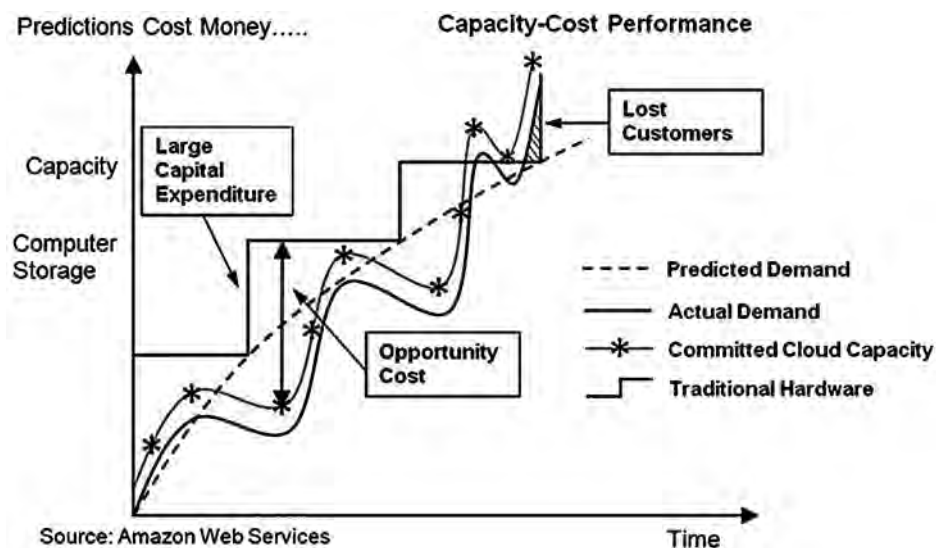
INTRODUCTION

Cloud computing has become the newest computing paradigm that has made a long time dream of making computing a utility a reality (Buyya, Yeo, Venugopal, Broberg, & Brandic, 2009). The initial attempt at providing computing as a utility was attempted by IBM through renting its mainframe computing resources to the Wall Street using remotely connected dumb terminals. Then the idea just died down until it again became popular in the early 2000s as cloud computing. Cloud computing makes it possible to purchase and use computing resources similar to utilities such as electricity, water, gas and telephony. The utility model allows the customers to consume any amount of resources and pay for only the resources that have actually been used. Prior to the arrival cloud computing,

computer systems were generally purchased outright. The systems thus purchased had fixed capacity irrespective of the actual demand. Figure 1 shows the capacity-utilization curve developed by Amazon Web Services depicting the usage of disk storage under cloud computing and traditional computing models (AWS, 2012).

From Figure 1, it can be seen that the traditional computing model shown by the step wise resource allocation is unable to follow the demand pattern. On the other hand, the cloud based resource provisioning closely follows the demand pattern in both short term as well as long term fluctuations. Thus cloud computing provides the right solution for any venture irrespective of their nature of business. On cloud computing, the committed computing resources and the money spent would have a very close and strong relationship with that

Figure 1. Capacity-utilization curve
 Source: AWS, 2012



DOI: 10.4018/978-1-5225-2255-3.ch101

of the real demand. This would make the money spent on these resources fully worth. Also cloud computing effectively transforms the money spent on the computing resources into an operational expenditure rather than a capital expenditure that has traditionally been. This helps businesses to utilize their capital fully on core business activities (Jaatmaa, 2010).

BACKGROUND

The technology that works behind to make cloud computing possible is hardware virtualization (Siddhisena, Warusawithana, & Mendis, 2011). In cloud computing, a single system is partitioned and hosted in parallel. These parallel systems can be spawned on demand and allocated to different users (Siddhisena et al., 2011). Once a user has completed his work, the virtual system can be removed and all the resources are released so that they can be allocated to another user in the future.

The virtualized computing infrastructure is created by installing a Virtual Machine Manager (VMM) on the physical hardware (Li, Yang, Kandula, & Zhang, 2010). The VMM provides the necessary isolation and security between the multiple virtual machines running in parallel on a single physical computer. Hosting individual fully functional systems on virtual machines maximizes the utilization of the physical systems as the physical system can be allocated to many customers. When many virtual machines are hosted on a single system simultaneously, its performance may start to degrade after a point due to the competition for resources between the hosted systems. So, in order to maintain the service quality, the maximum number of virtual machines hosted on a system must be limited.

Advantages and Disadvantages of Virtualization

Virtualization of computer hardware results in many benefits along with certain shortcomings.

Though there are a few disadvantages, the advantages for sure outweigh those disadvantages. Ameen and Hamo (2013) have carried out an extensive analysis on the advantages and disadvantages of hardware virtualization.

The most prominent advantages and disadvantages of hardware virtualization are listed below:

- Ability to have better server and application consolidation
- Improve security through sandboxing
- Multiple simultaneous execution environments
- Better use of existing hardware
- Reduction in IT infrastructure costs
- Reduced downtime or increased uptime and faster failure recovery
- Simplified system administration
- Ability to have gradual capacity expansion
- Support for legacy systems and applications
- Simplified system installation and deployment
- Better system and application testing business continuity and disaster recovery capabilities
- Improved business agility
- Better resource sharing and utilization
- Increased flexibility, availability and scalability
- Enhanced security, isolation and hardware independence
- Better load balancing
- Opportunity for single points of failure
- Reduced performance, if not managed properly
- Complicated management interface
- Increased networking complexity and debugging time

HISTORY OF VIRTUALIZATION TECHNOLOGY

The initial work on creating virtualized systems started way back in the early 1960s by organiza-

tions like General Electric, Bell Labs, IBM and MIT (Ali & Meghanathan, 2011). IBM was the first to bring the concept of Virtual Machines to the commercial environment. Virtual Machines as they were initially implemented on IBM's Mainframes are still in use today, though the use of mainframes has diminished with the development of distributed computing along with the capability of low end computer systems. On the lower end of the computing spectrum, namely the personal computer systems, the implementation of virtual systems started only in the 1990s. The virtualization support for the x86 class microprocessors were initially achieved through complex software implementations as the hardware support for such implementations were not available at time (Bugnion, Devine, Rosenblum, Sugerma, & Wang, 2012).

VMware Workstation was the first virtualization solution implemented for the 32 bit x86 based platforms. The virtualization solution implemented for the x86 platform was very complicated compared to that of the mainframe counterparts. The main reason behind the success of the mainframe virtualization was the monolithic single vendor implementation of different components of the computer system. In the mainframe arena, the processor, the computer system, the operating system, the virtual machine manager and to a great extent the main applications were all developed by a single vendor (Bugnion et al., 2012). On the other hand the x86 industry had been totally fragmented into multiple vendors implementing various components mostly with vendor specific designs with least uniformity.

SERVER VIRTUALIZATION TECHNIQUES

Researchers have proposed many virtualization technologies that can be used for hosting multiple systems on a single physical infrastructure. These technologies can be divided into the following groups; they are namely full virtualization, para-

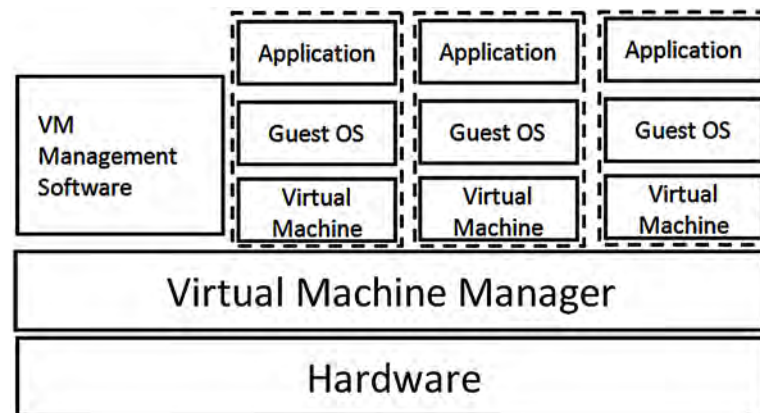
virtualization, hardware-assisted virtualization, operating system virtualization (Zhang, et al., 2010; Bazargan, Yeun, & Zemerly, 2012). All these types of virtualization present different independent virtual machines to the users. In addition to these virtualization techniques, there are other virtualization techniques called storage virtualization, memory virtualization, network virtualization and application virtualization (Gao, Zhang, & Zhou, 2012; Kumar, Reddy, & Rupa, 2012; Costa, Migliavacca, Pietzuch, & Wolf, 2012; Khanna, Beaty, Kar, & Kochut, 2006).

Full Virtualization

Full virtualization creates a complete emulation of the underlying hardware in multiple instances that are presented to the users (Marinescu & Kroger, 2007). Since the full virtualization creates an exact copy of the underlying hardware as virtual machine instances, it is commonly known as the native virtualization. One of the important features of the native virtualization is the complete reflection of the every aspect of the underlying hardware in the virtual machines. every virtual machine instance created by full virtualization will include the complete instruction set, registers, input/output operations, interrupts, memory access etc., of the hardware on which they are hosted. Hence the environment that has been emulated though the full virtualization is an exact copy of the hardware enabling the users to take full advantage of the hardware without requiring any additional knowledge of tweaking. Thus any operating system and application that are supported by the raw hardware can be installed and executed on the virtual machines.

Implementing full virtualization is not a straight forward task as many challenges including security and isolation of virtual machines need to be handled properly. One such challenge that requires the special attention is the execution of privileged instructions evoked by the virtual machines. Privileged operations such as input/output when requested by a virtual machine must

Figure 2. The layered architecture of a fully virtualized system



be intercepted and emulated carefully by the virtual machine manager. The operations invoked by a single virtual machine must be confined to that virtual machine without affecting the status of other virtual machines or the general status of the overall physical system. Only the hardware operations that keep their effects entirely within the virtual machine elements managed by the control programs such as memory locations or arithmetic registers must be allowed to be executed directly on the hardware. All the other operations must be captured and emulated. Figure 2 shows the typical layered architecture of a system installed with full virtualization. From this figure, it can be seen that many different types of operating systems can be installed and executed concurrently on a single hardware system without affecting any one of them. The virtual machine management software executed along with the virtual machine manager would manage the necessary allocation of resources and handle the security between different virtual machines.

Full virtualization has been commonly used for:

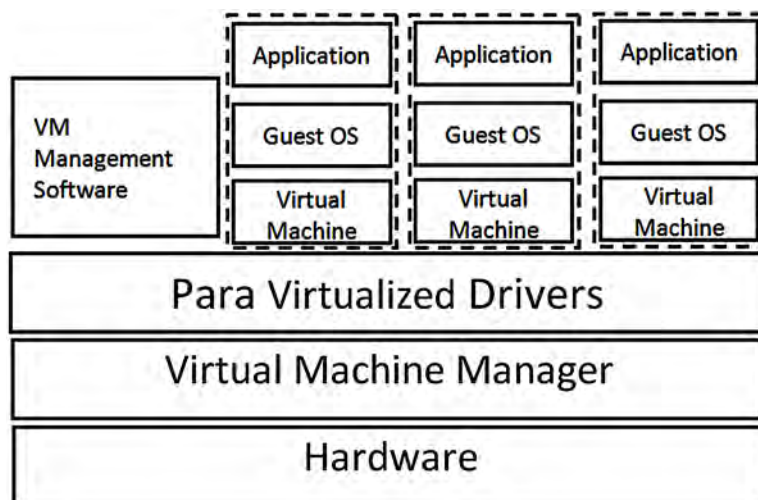
- Hosting multiple users on a single computer system;
- Isolating and protecting users from interference;
- Emulating hardware for improved reliability, security and productivity.

Paravirtualization

Paravirtualization is the presentation of a software interface for virtual machines that is slightly different from the underlying hardware (Marinescu & Kroger, 2007). Paravirtualization is also known as the operating system-assisted virtualization as it requires the modification of the guest operating system in order to achieve the successful execution of privileged operations (Ameen & Hamo, 2013). Paravirtualization has been achieved through replacing the non-virtualizable instructions with an alternative set of equivalents that can be easily virtualized. The modification has been intended to speed up the operations that are more difficult to execute in virtualized environments than in real environments. Paravirtualization creates special hooks that are specially defined to allow the guests and host to invoke and recognize those tasks that must be executed in a host domain rather than in the virtual domain. A successful paravirtualized platform makes the design and operation of the virtual machine manager simpler and improves the overall performance of the guest operating system.

The guest operating system to be installed on top of paravirtualized system must be specifically ported for the paraAPI. A conventional OS that has not been designed with special components necessary for running on paraAPI will not run on a paravirtualization VMM. Under para-virtualization, split drivers are used for handling I/O

Figure 3. Paravirtualization

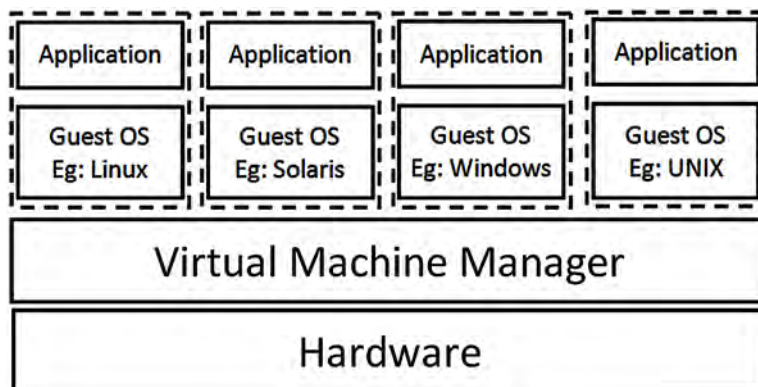


requests. The physical devices are accessed by another set of drivers known as back-end drivers that are part of a privileged virtual machine installed directly on the hardware (Barham et al., 2003). The front-end device driver within the guest operating system would pass the I/O requests from the operating system to the back-end driver in the privileged operating system. The back-end driver will then interpret the request and carry out the task on the hardware directly. Figure 3 shows a system that has been installed with the para-virtualization software. From Figure 3, it can be seen that paravirtualization is similar to full virtualization except for the modification of guest OSs that are running on top of virtual machines.

Hardware-Assisted Virtualization

Hardware-assisted virtualization does not use binary translation as in full virtualization nor hypercalls as in paravirtualization (Shen, Hamayun & Petrot, 2012). Hardware-assisted virtualization is been called by multiple names such as native virtualization, accelerated virtualization or hardware virtual machine. Hardware-assisted virtualization enables full virtualization with maximum efficiency with the aid of the hardware capabilities, especially the processors. Hardware-assisted virtualization provides for a higher level CPU instruction set communication as the hypervisor runs in a new advanced mode known as the root

Figure 4. Hardware-assisted virtualization



mode below the operating system kernel mode. In this virtualization form the critical and privileged instructions are enabled to automatically seize the hypervisor directly. Figure 4 illustrates the hardware-assisted virtualization architecture.

Hardware-assisted virtualization is commonly found in the server market due to its improved performance and security due to high virtual machine isolation. The VMM running on hardware controls and synchronizes the access of the guest operating systems to the hardware resources directly. The advantages of hardware-assisted virtualization include, reduction of maintenance overhead of para-virtualization as no changes to the guest operating system is needed and better performance compared to all other virtualization techniques. The downside of the hardware-assisted virtualization is the requirement of the explicit host CPU support that may not be available in all processors. Also a pure hardware-assisted virtualization approach suffers from high CPU overheads that limit the scalability and the efficiency of server consolidation due to many virtual machine traps involved.

Operating System Level Virtualization

Operating system level virtualization also called the Single Kernel Image (SKI) runs multiple

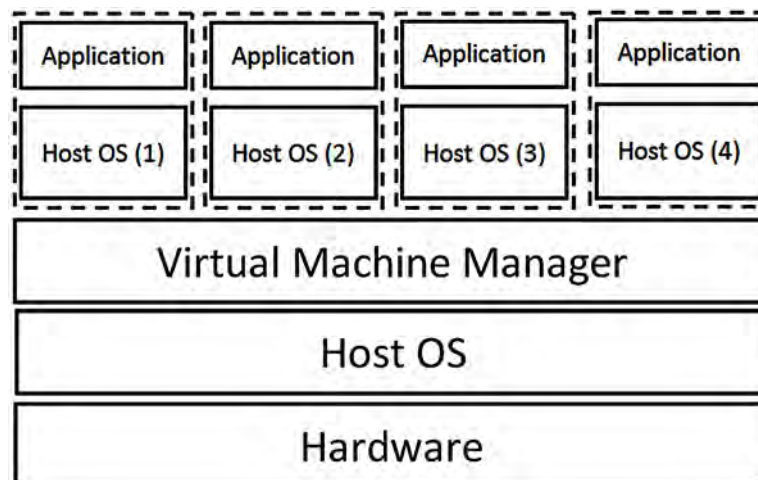
instances of the same host operating systems in parallel (Bazargan et al., 2012). Thus, it can be concluded that it is not the hardware that has been virtualized but the host OS is. All virtual machines running on the same hardware must execute the same virtualized operating system image. Figure 5 shows the architecture of the OS level virtualization. The virtualized operating system image is known as the virtualization layer.

The OS level virtualization reduces the administration effort needed for managing the system by allocating the resources both at creation of a virtual machine as well as at runtime. The OS layer virtualization is more efficient than other hardware level virtualizations but suffers from the shortcomings of reduced isolation between different virtual instances and lack of freedom for installing different OSs other than the host OS.

Storage Virtualization

Under storage virtualization, multiple disk drives are combined into a single logical drive and presented to the hosting system as such (Gao et al., 2012). Storage virtualization can be divided into two primary types. They are namely, file virtualization and block virtualization. Block virtualization separates logical storage from the physical storage that helps flexible access to the

Figure 5. Operating system-level virtualization



storage system without worrying about the details of the physical arrangement of the disks (Wooley, 2011). File virtualization operates at file level rather than block level or disk levels (Joukov et al., 2007). File virtualization makes the data accessed at file level independent from the physical location of the files. This helps the optimization of storage use through server consolidation and the migration of files from one system to another with minimum disruptions.

Storage virtualization may be carried out within a single computer, within a local area network or across multiple local area networks (Peng, Zhu & Luo, 2012). The storage that is directly attached to a server is known as Direct Attached Storage (DAS) (Yuan, 2011). In DAS the storage device is directly attached to the host computer by a cable. The internal disk drive of a computer or a directly attached tape drive is the simplest form of DAS. The access to the storage device is carried out through I/O requests that access devices directly. When a storage device along with a special integrated processor for handling data access requests is connected to the network directly, it is called Network Attached Storage (NAS) (Peng et al., 2012). The data access through a NAS device is carried out with the aid of specialized protocols. The NAS processor attached with the storage device translates will the file requests received from clients to device I/O requests. When the storage devices are housed in a special network and access to them is provided through a gateway, it is called a Storage Area Network (SAN) (Chishima, Yamaguchi & Oguchi, 2007). Similar to DAS, in SAN also data access is carried out through direct I/O requests.

Memory Virtualization

Similar to physical computers, even virtual machines require memory dedicated for them to keep the active programmes including the operating system, applications and other processes (Ali & Meghanathan, 2011). The total size of memory required is equal to its configured size and the

overhead required for virtualization. The configured memory is the amount of RAM including both physical and virtual memories made available to the guest operating system for hosting the applications and other processes (Kumar et al., 2012). For creating memory virtualization the VMM running host computer partitions the physical and virtual memory space of the host system. These portions of memory are isolated from each other in order to provide the required security for protecting them from interfering with each other.

For the virtual machines to work smoothly the virtual memory addresses used by the virtual machines must be translated to physical memory addresses. Address translation in virtual machines are carried out using several mapping techniques such as Extended Page Table (EPT), Shadow Tables (ST), Nested Page Table (NPT) etc., (Wang, Zang, Wang, Luo & Li, 2011). The conventional method of address translation is using shadow page tables. The VMM maintains a shadow page table for every virtual machine where the mapping of every virtual address to physical address is maintained. In addition to the shadow page table maintained by the VMM, every guest also maintains its own virtual to physical mapping table. The shadow mechanism carried out through shadow pages table suffers from the shortcoming of VM exits at times of page faults that require synchronization between two page tables. In order to avoid the VM exits, vendors like Intel and AMD have developed other techniques such as hardware assisted EPT and NPT for facilitating address translation. In hardware assisted techniques, the Memory Management Unit (MMU) manages a guest page table that maps the virtual addresses to physical addresses of the guests. In the hardware assisted techniques, a guest page fault is managed by the guest itself protecting the system from VM exits.

Network Virtualization

Network virtualization refers to the combination of network hardware and software into a single

virtual network (Bazargan, et al., 2012). When network virtualization has been implemented, it hides the details of the network implementation and provides a unified view that can be customized through a software interface to suit the user requirements.

Application Virtualization

Application level virtualization presents the user with a copy of components that are not shared with other processes. The components commonly protected this way include registry files, global objects, local variables and personalized environmental settings. The virtual environment prevents conflicts with other instances of the same application running in parallel on the same system. Java Virtual Machine (JVM) is an example of application level virtualization. The following sections briefly describe the different virtualization technologies.

QUALITY OF SERVICE ISSUES IN VIRTUALIZATION

Since virtualization is the underpinning technology that powers the revolutionary cloud computing paradigm, the quality of service of virtualization will decide the quality of cloud computing as well. It is a fact that a virtualized system cannot offer performances equal to that of a system running with native resources (Rao & Rao, 2015). Hence it is vital to ensure the quality of service in virtualization, if it is to provide a guaranteed service to customers while maintaining maximum efficiency and utilization of the resources (Calheiros, Ranjany & Buyya, 2011). Calheiros et al., (2011) further state that the achieving the required QoS targets agreed upon in the SLAs signed between the customers and service providers is often complicated due to many issues including the uncertain behavior of virtualized IT resources

and network elements. The leading hypervisors in the market such as Xen employ fair schedulers that manage the job allocation based on pointers or weights assigned by the administrators (Rao & Rao, 2015). The performance of the jobs thus scheduled is very sensitive to these scheduling parameters. The main downside of this scheme is the static nature of these pointers that cannot take the dynamic nature of the application requirements. It must be noted that the exact relationship between the domain parameters and performance measures such as response time or throughput are neither obvious nor static. Also, the resource provisioning by servers may encounter many unpredictable situations that obstruct the smooth delivery of application services at run times. The problem is further compounded in large public data centers by the dynamic and unpredictable workload behavior with spikes depending on the time of day, application popularity, customer characteristics and expectations etc. Hence, if the cloud systems need to meet the demands of the customers, the underlying technology that enables the cloud operations possible must be designed in such a manner that it can respond to the variations positively and promptly.

VIRTUALIZATION SECURITY

Despite the advantages the virtualization technologies provide the users, they have certain drawbacks too, if not handled properly. Security is one of the main issues that must be given special attention when setting up a virtualized system. Some of the most common security issues identified with virtualization include virtual machine escape, virtual machine hopping, malicious code injection, hyperjacking, virtual machine sprawl, side channel attacks, virtual machine poaching and unsecured virtual machine migration (Kedia, Nagpal & Singh, 2013; Kapadia & Gulati, 2014).

FUTURE RESEARCH DIRECTIONS

Virtualization area provides the researchers ample opportunity for further research into the foreseeable future. The following are some of the directions, which the future research can be directed to:

- **Virtualization Security:** Many guest systems are loaded on top of a single hardware platform sharing the resources. So, the security of the loaded systems and applications may be compromised through leaks and attacks. Hence further research into the system isolation and security is needed for achieving complete isolation between hosted systems while efficiently sharing the common resources.
- **Energy Optimization:** Datacenters have been identified as the entities to have large energy footprints. Largest amount of energy is consumed during the switching on and off phase of virtual machines. Hence the energy efficient virtual machine allocation provides an opportune area for future researchers.
- **Efficient Virtual Machine Allocation:** Cloud computing obtains its advantages through efficient management of hardware resources. Hence efficient allocation and reallocation of virtual machines on the fly is a must. Thus, this area too provides opportunities for the development of new and improved virtual machine allocation strategies, mechanisms and schemes.
- **Elastic Resource Management:** Effective elastic resource management requires the consideration of many areas including resource allocation, resource provisioning, resource mapping and resource adaptation. Hence extensive future research is needed in these areas to come up with effective and efficient elastic resource management schemes.
- **QoS in Virtualized Systems:** Real-time business applications generally have very

stringent QoS requirements. Implementing QoS in virtualized environments would determine the success of the entire cloud computing landscape. Hence, there are many avenues for future research in the area of virtualization service quality.

CONCLUSION

This chapter took an in depth look at virtualization from the points of their advantages, disadvantages, different types and implementation schemes. The chapter also looked at the history of virtualization in order to present a complete picture of this ground breaking technology starting from the IBM's mainframe based virtual machines to today's virtualized systems implemented on various hardware platforms and resources. This chapter also briefly highlights the quality of service and security issues that might affect the performance of virtualized systems.

REFERENCES

- Ali, I., & Meghanathan, N. (2011). Virtual machines and networks: Installation, performance, study, advantages and virtualization options. *International Journal of Network Security & Its Applications*, 3(1), 1–15. doi:10.5121/ijnsa.2011.3101
- Ameen, R. Y., & Hamo, A. Y. (2013). Survey of server virtualization. *International Journal of Computer Science and Information Security*, 11(3), 1–10.
- AWS. (2012). *Capacity vs Utilization Curve*. Retrieved May 15, 2013, from Amazon Web Services: <http://www.amazon.com/economics>
- Barham, P., Dragovic, B., Fraser, K., Hand, S., Harris, T., Ho, A., & Warfield, A. et al. (2003). Xen and the art of virtualization. *19th ACM Symposium on Operating System Principles*, (pp. 164-177). Bolton Landing, NY: ACM.

- Bazargan, F., Yeun, C. Y., & Zemerly, M. J. (2012). State-of-the-art of virtualization, its security threats and deployment models. *International Journal for Information Security Research*, 2(3/4), 335–343.
- Bugnion, E., Devine, S., Rosenblum, M., Sugerman, J., & Wang, E. Y. (2012). Bringing virtualization to the x86 architecture with the original VMware workstation. *ACM Transactions on Computer Systems*, 30(4), 1–51. doi:10.1145/2382553.2382554
- Buyya, R., Yeo, C. S., Venugopal, S., Broberg, J., & Brandic, I. (2009). Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility. *Journal of Future Generation Computer Systems*, 25(6), 599–616. doi:10.1016/j.future.2008.12.001
- Calheiros, R. N., Ranjan, R., & Buyya, R. (2011). Virtual machine provisioning based on analytical performance and QoS in cloud computing environments. *International Conference on Parallel Processing*, 295–304. doi:10.1109/ICPP.2011.17
- Costa, P., Migliavacca, M., Pietzuch, P., & Wolf, A. L. (2012). NaaS: Network-as-a-Service in the cloud. *2nd USENIX Workshop on Hot Topics in Management of Internet, Cloud, and Enterprise Networks and Services*, 1–6.
- Gao, Y., Zhang, Y., & Zhou, Y. (2012). Building a virtual machine-based network storage system for transparent computing. *International Conference on Computer Science and Service System*, 2341–2344. doi:10.1109/CSSS.2012.581
- Jaatmaa, J. (2010). *Financial aspects of cloud computing business models* (MSc Thesis). School of Economics, Aalto University, Finland.
- Kapadia, G. S., & Gulati, R. M. (2014). Security concern for virtualization in cloud computing. *International Journal of Computer Science Engineering and Information Technology Research*, 4(5), 107–116.
- Kedia, P., Nagpal, R., & Singh, T. P. (2013). A survey on virtualization service providers, security issues, tools and future trends. *International Journal of Computers and Applications*, 69(24), 36–42. doi:10.5120/12123-8491
- Khanna, G., Beaty, K., Kar, G., & Kochut, A. (2006). Application performance management in virtualized server environments. *10th IEEE/IFIP Network Operations and Management Symposium*, 373–381. doi:10.1109/NOMS.2006.1687567
- Kumar, P. P., Reddy, A. R., & Rupa, A. (2012). Snapshot based virtualization mechanism for cloud Computing. *International Journal of Computer Science Issues*, 9(5), 226–231.
- Marinescu, D., & Kroger, R. (2007). State of the art in autonomic computing and virtualization. Technical Report, Wiesbaden University of Applied Sciences.
- Rao, V. V., & Rao, M. V. (2015). A survey on performance metrics in server virtualization with cloud environment. *Journal of Cloud Computing*, 1–12.
- Shen, H., Hamayun, M. M., & Petrot, F. (2012). Native simulation of MPSoC using hardware-assisted virtualization. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 31(7), 1074–1087. doi:10.1109/TCAD.2012.2187526
- Siddhisena, B., Warusawithana, L., & Mendis, M. (2011). Next generation multi-tenant virtualization cloud computing platform. *13th International Conference on Advanced Communication Technology*, 405–410.
- Zhang, B., Wang, X., Lai, R., Yang, L., Luo, Y., & Li, X. (2010). A survey on I/O virtualization and optimization. *5th Annual ChinaGrid Conference*, 117–123.

ADDITIONAL READING

Arrasjid, J. Y., Lin, B., Veeramraju, R., Kaplan, S., Epping, D., & Haines, M. (2011). *Cloud computing with VMware vCloud Director*. Berkeley, CA: USENIX.

Buyya, R., Vecchiola, C., & Selvi, S. T. (2013). *Mastering cloud computing*. New Delhi: Tata McGraw Hill. Portnoy, M. (2102). *Virtualization essentials*. Indianapolis, IN: John Wiley & Sons.

Ruest, D., & Ruest, N. (2011). *Virtualization: A beginner's guide*. New York, NY: McGraw Hill.

Tulloch, M. (2010). *Understanding Microsoft virtualization solutions: From the desktop to the datacenter*. Redmond, WA: Microsoft Press.

Turnbull, J. (2014). *The docker book: Containerization is the new virtualization*. Retrieved from http://www.amazon.com/dp/B00LRROT14/ref=rdr_kindle_ext_tmb

Vaquero, L. M., Cáceres, J., & Hierro, J. J. (2012). *Open source cloud computing systems: Practices and paradigms* (pp. 1–378). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-0098-0

Wang, L., Ranjan, R., Chen, J., & Benatallah, B. (2011). *Cloud computing: Methodology, systems, and applications*. Boca Raton, FL: CRC Press.

Wolf, C., & Halter, E. M. (2005). *Virtualization: From the desktop to the enterprise*. New York City, NY: Apress.

Yang, C., & Huang, Q. (2013). *Spatial cloud computing: A practical approach*. Boca Raton, FL: CRC Press. doi:10.1201/b16106

KEY WORDS AND DEFINITIONS

Attack: It is an attempt to gain unauthorized access to, steal, expose, alter, disable or destroy computing resources by unauthorized personnel or through unauthorized means.

Cloud Computing: It is the newest paradigm in computing that delivers the resources including hardware, platform and application as services.

Elasticity: It is the capability of a computer system to adjust its capacity dynamically to cater to the demands of users.

Infrastructure as a Service: It is a service provided through cloud computing technology where users can access virtualized hardware and pay on a utility basis.

Multiple Tenancy: Refers to accommodating multiple users on a single hardware system and enabling them to access the underlying resources as if they are the sole users.

Quality of Service: Refers to the capability of a system to provide services to customers within the expected limits of user preference.

Security: Denotes the protective mechanisms including policies, guidelines and algorithms applied to computers, computer networks, users and data stored and transmitted over them.

Virtualization: It is the process of creating multiple logical hardware replicas based on an underlying infrastructure and assigning them to different users.

Category C

Communications Theory

Communication, Information, and Pragmatics

Adriana Braga

Pontifical Catholic University of Rio de Janeiro, Brazil

Robert K. Logan

University of Toronto, Canada

INTRODUCTION

The intention of this article is to clarify the relationship between communication and information by considering pragmatics. Although these terms are closely related, they are not the same. In order for communication to occur the information that is transmitted must be processed within the social context of the sender and the receiver or in other words through the use of pragmatics. That is, there is no communication between the sender and the receiver if the receiver does not understand the information sent by the sender. Information before it is interpreted is therefore nothing more than the signal. It only becomes communication, if it is properly interpreted by the receiver of the information. And it is only through the context or the pragmatics that the receiver can understand the intended meaning of the sender and therefore as a result communication can take place. Misinterpretation leads to miscommunication. As no interpretation is perfect as pragmatics between the sender and the receiver is never perfect the content of the communication depends on the user and the user's interpretation, which McLuhan formulated with his iconic one-liner *the user is the content*. In actuality the sender, the receiver and their understanding of each other are all part of the content.

BACKGROUND

The context in which the information is interpreted is the only way that the meaning that was intended by the sender can be understood by the receiver, but the meaning that the receiver attaches to the information sent by the sender will always vary to some degree (Chan, Walker & Gleaves, 2015; Gibbs et al, 2015; Gui & Argentin, 2011; Introna & Nissenbaum, 2000). Because the 'user is the content' all communication is miscommunication to a certain degree. Perfect communication is an ideal that all communicators strive to achieve through the art of rhetoric. To sum up what we have just posited: information is required for communication but does not necessarily result in communication and never results in perfect communication. The extra ingredient that is required to transform information into communication is context or pragmatics, which is never perfect. Recent methodological innovations try to account for the pragmatics on digital environments (Boyd and Crawford, 2012; Hine, 2005; Lee and Chen, 2015).

In this article we will first examine the meaning and significance of information, which will entail a critique of Shannon Information Theory. We will show it is really a theory of the transmission of signals. We describe how MacKay and Bateson with their respective formulations of "information

is the distinction that makes a difference (MacKay, 1969)” and “information is the difference that makes a difference (Bateson, 1973)” adds the element of meaning to the definition of information. We then examine the proposition of Kauffman, Logan et al. (2007) that organization is a form of information and that life entails the propagation of organization.

Assuming that it is not possible not to communicate, we emphasize the pragmatic dimension of communication. We argue that ‘information,’ ‘communication’ and ‘social interaction’ are inseparable elements of production of meaning, even if analytically they can be conceived as independent concepts. Thus, as in any communication there are three simultaneous dimensions operating as a system – syntactic, semantics and pragmatics, and it is also the case that ‘information,’ ‘communication’ and ‘social interaction’ are operating as a system. In this sense, speech acts owe their meaning to performances in the context of what information the sender sends, the interpretation of the receiver and the social context that exists between the sender and the receiver. We argue that, from the differential emphasis on the syntactic, semantic or pragmatic dimensions of communication, lies a major difference between models for the theory of communication: communication as transmission of information or communication as a relational activity.

Information: From Origins to Shannon’s Information Theory

The English word information according to the Oxford English Dictionary (OED) first appears in the written record in 1386 by Chaucer. It is derived from Latin through French by combining the word “inform” meaning “giving a form to the mind” with the ending “ation” denoting a noun of action. This earliest definition refers to information as an item of training or molding of the mind. Information is not an object but a process of forming or informing the mind.

The notion of information as something capable of storage in or the transfer to something inanimate and the notion of information as a mathematically defined quantity does not arise until the 20th century. The beginning of the modern theoretical study of information is attributed to Claude Shannon (1948), who is recognized as the father of information theory. He defined information as a message sent by a sender to a receiver. Shannon wanted to solve the problem of how to best encode information that a sender wished to transmit to a receiver. Shannon gave information a numerical or mathematical value based on probability defined in terms of the concept of information entropy more commonly known as Shannon entropy. Information is defined as the measure of the decrease of uncertainty for a receiver. The amount of Shannon information is inversely proportional to the probability of the occurrence of that information, where the information is coded in some symbolic form as a string of 0s and 1s or in terms of some alpha-numeric code.

MacKay’s Counter Revolution: Where Is the Meaning in Shannon Information?

According to Claude Shannon (1948) his definition of information is not connected to its meaning. However, as Shannon suggested, information in the form of a message often contains meaning but that meaning is not a necessary condition for defining information. So it is possible to have information without meaning, whatever that means.

Not all of the members of the information science community were happy with Shannon’s definition of information. Three years after Shannon proposed his definition of information Donald Mackay (1969) suggested that information should be defined as “the change in a receiver’s mind-set,” and thus with meaning. He defined information as “a distinction that makes a difference.” and not just the sender’s signal (Hayles, 1999). The notion of information independent of its meaning or context is like looking at a figure isolated from its ground.

As the ground changes so too does the meaning of the figure. This is a central idea of Marshall McLuhan's (1964) approach to communications known as media ecology.

[McLuhan] believed that to understand the meaning of a figure one must take into account the ground in which it operates and in which it is situated. The true meaning of any "figure," whether it is a person, a social movement, a technology, an institution, a communication event, a text, or a body of ideas, cannot be determined if one does not take into account the ground or environment in which that figure operates. The ground provides the context from which the full meaning or significance of a figure emerges. The concern with the figure/ground relationship is consistent with McLuhan's emphasis on interface and pattern rather than on a fixed point of view (Logan 2011, p.2).

The problem with MacKay's definition was that meaning could not be measured or quantified and as a result the Shannon definition won out and changed the development of information science. The theorizing that Shannon conducted through his combination of electrical engineering and mathematics came to be known as information theory. It is ironic that the OED cites the first use of the term "information theory" as that of MacKay's who used the term in a heading in an article he published in the March 1950 issue of the Philosophical Magazine. Gregory Bateson (1973) defined information as "the difference that makes a difference," which is now more often quoted than MacKay's formulation. Both MacKay's and Bateson's formulation contain the notion that it is the meaning of the information that makes the difference. Another one-line definition of information that incorporates the notion of its meaning is Fredkin's "The meaning of information is given by the processes that interpret it." This is a very insightful definition because it explicitly incorporates the notion that information depends on context.

If information is the distinction (MacKay) or the difference (Bateson) that makes a difference then if there is no distinction or no difference then there can be no information. This would mean chaos or random numbers contain no information because there is no difference or distinction in one part of the stream of numbers as opposed to another part of the stream because of a lack of organization. This is opposite to the conclusion of Shannon who claims that a stream of random numbers contains the maximum information. While it is true each element is different from the next and is a complete surprise it is also true that the overall pattern of chaos and randomness is the same and hence there is no distinction nor is there any difference in the stream of random numbers.

Feedforward and Cybernetics

I. A. Richards' area of research was rhetoric, which he considered to be more than just the art of persuasion. Richards was concerned with the accuracy of human communication. He considered the field of rhetoric to be about finding remedies for avoiding misunderstandings and hence improving communication as well as understanding how words work. He believed the notion of feedforward was an important tool for achieving these ends. Feedforward is basically a form of pragmatics where pragmatics is the use of context to assist meaning.

Richards first introduced the term feedforward in his address to the Macy conferees in 1951:

Perhaps this thing on which I want to put the spotlight will be considered to be included in some ingenious way under the word "feedback." But what I am going to stress stands in an obvious and superficial opposition to "feedback," and it will, in certain frames of thought, be given nearly, if not quite so much, importance, and sometimes more importance than feedback itself in certain connections. It is certainly as circular. You have no doubt fed forward enough to see that what I am going to talk about from now on is feedforward.

I am going to try to suggest its importance in describing how language works and, above all, in determining how languages may best be learned (Richards, 1952, p. 54).

The coining of the term by Richards was no doubt influenced by the term feedback used by cyberneticians and according to the OED first introduced into the English language in 1920. But as Richards pointed out feedforward stands in superficial opposition to feedback. Feedback is basically reactive whereas feedforward is proactive. Feedforward anticipates where one is headed and sets one's goals. Feedback allows one to see how close one gets to their goals. Richards who stressed the importance of providing the context of what one wanted to communicate might have coined the term feedforward to complement the term feedback used by cyberneticians precisely because the audience that he was addressing at the Macy Conference included the man who coined the term cybernetics, namely Norbert Wiener. The term feedforward as used by Richards suggested that in order to have one's communication understood it was necessary to literally feedforward the context of what one was planning to talk about.

Organization as Information

In an article entitled Propagating Organization: An Enquiry (POE) Kauffman, Logan, Este, Goebel, Hobill and Shmulevich (2007) argued that Shannon's (1948) classical definition of information as the measure of the decrease of uncertainty was not valid for a biotic system that propagates its organization. The core argument of POE was that Shannon information "does not apply to the evolution of the biosphere" because Darwinian preadaptations cannot be predicted and as a consequence "the ensemble of possibilities and their entropy cannot be calculated." Therefore a definition of information as reducing uncertainty does not make sense since no matter how much one learns from the information in a biotic system the uncertainty remains infinite because the number

of possibilities of what can evolve is infinitely non-denumerable. This contradicts Shannon who specified that the number of possible messages was finite.

Instead of Shannon information we defined a new form of information, which we called instructional or biotic information, not with Shannon, but with constraints or boundary conditions. The amount of information will be related to the diversity of constraints and the diversity of processes that they can partially cause to occur. By taking this step, we embed the concept of information in the ongoing processes of the biosphere, for they are causally relevant to that which happens in the unfolding of the biosphere.

We therefore conclude that constraints are information and... information is constraints... We use the term 'instructional information' because of the instructional function this information performs and we sometimes call it 'biotic information' because this is the domain it acts in, as opposed to human telecommunication or computer information systems where Shannon information operates (Kauffman, Logan, et. al., 2007).

A living organism is an open system, which von Bertalanffy (1968) "defined as a system in exchange of matter with its environment, presenting import and export, building-up and breaking-down of its material components." Instructional or biotic information may therefore be defined as *the organization of that exchange of energy and matter*. The fact that a biotic system is an open system can be used to argue against the association of instructional or biotic information with cybernetics because cybernetics focuses strictly on the flow of information and does not deal with the flow of energy and matter.

In POE it has been the associated biotic or instructional information with the organization that a biotic agent is able to propagate. This contradicts Shannon's definition of information and the notion that a random set or soup of organic chemicals has more Shannon information than a

structured and organized set of organic chemicals found in a living organism.

The biotic agent has more meaning than the soup, however. The living organism with more structure and more organization has less Shannon information. This is counterintuitive to a biologist's understanding of a living organism. We therefore conclude that the use of Shannon information to describe a biotic system would not be valid. Shannon information for a biotic system is simply a category error. A living organism has meaning because it is an autonomous agent acting on its own behalf. A random soup of organic chemicals has no meaning and no organization (Kauffman, Logan, et. al., 2007).

According to Shannon's definition of information a set of random numbers transmitted over a telephone line would have more information than the set of even numbers transmitted over the same line. Once 2, 4, 6, 8, 10, 12 was received the receiver would be able to correctly guess that the rest of the numbers to follow the sequence would be the set of even numbers. The random numbers have no organization but the even numbers are organized so the mystery of the relevance of Shannon information deepens, as one must counter-intuitively conclude that information and organization can be at cross-purposes in Shannon's scheme of things.

This argument completely contradicts the notion of information of a system biologist who would argue that a biological organism contains information. It is by virtue of this propagating organization that an organism is able to grow and replicate, as pointed out by Kauffman (2000) in *Investigations*. From the contradiction between Shannon and biotic information we already have a hint that there is possibly more than one type of information and that information is not an invariant like the speed of light in relativity theory, which is independent of its frame of reference. We also see that perhaps Shannon's definition of information might have limitations and might not represent a universal notion of information. After all Shannon formulated his concept of information as information entropy to solve a specific

problem namely increasing the efficiency or the signal to noise ratio in the transmission of signals over telecommunication lines.

The term information is generally regarded as some uniform quantity or quality, which is the same for all the domains and phenomena it describes. In other words, information is an invariant like the speed of light, the same in all frames of reference. But as Kauffman, Logan et al. (2007) have argued the notion of Shannon information and biotic information are quite different.

Is It Possible Not to Communicate?

Living beings (be they bacteria, plants, fish or humans) interact with their environments by reacting to stimuli, that is, by defining their immediate surroundings and adapting their conduct to them. That means interpretation of signs, processing of information and choosing a definite line of conduct. That is, *communicating* with their environment. In the same way that syntax, semantics and pragmatics are inseparable in communication phenomena, information, communication and interaction are inseparable in biotic systems, human culture included.

Action is innate in humans, and is not just a reaction. Meaning is foundational for the direct feeling and perception of "reality". Meaning is not inherent in the objects, but the ownership of the propositions, statements, and beliefs belongs to the receiver through their interpretations of the originating signals of the sender. The sense of presence in the world presupposes a real world through experience.

Meaning is not inherent to objects, but attributed by people in particular social situations. Additionally, the attribution of meaning to things never ceases. In the presence of another person, an object, a message or a social situation, an individual immediately tries to define the situation and attribute to it a meaning in the context of the immediate situation.

Assuming that it is not possible not to communicate, we emphasize the pragmatic dimension

of communication. We argue that ‘information,’ ‘communication’ and ‘social interaction’ are inseparable elements of production of meaning, even if analytically they can be conceived as independent concepts. Thus, as in any communicational process there are three simultaneous dimensions operating as a system – syntactic, semantics and pragmatics, and it is also the case that ‘information,’ ‘communication’ and ‘social interaction’ are operating as a system. In this sense, speech acts owe their meaning to performances in the context of what information the sender sends, the interpretation of the receiver and the social interaction that exists between the sender and the receiver.

The pragmatist motto of Peirce states that the production of meaning is oriented to action, and that the idea of what a thing ‘is’ lies on the somatry of the effects that can be conceived as possible from it.

Beside this two main branches of pragmatist philosophy, there is also the approach of Ludwig Wittgenstein (mainly in his *Philosophical Investigations*, 1953), in which the meaning of an expression lies on its practical uses; that of John Austin (1962) and John Searle (1965), in which ordinary language is seen as a resource for philosophical analysis, as well as those who see in pragmatism an adequate technical resource for the renewal of a transcendent philosophy of communication, such as Jurgen Habermas (1984) and Karl-Otto Apel.

According to Charles W. Morris, a semiotician and disciple of George Herbert Mead, the study of language has been traditionally divided between: a) a semantic approach, dealing with the relation of signs with the things they represent, that is, of signs with their meaning; and b) a syntactic approach, dealing with the relation of signs with other signs, such as the relation between words within a sentence, searching for the rules that ensure its meaning.

To Morris, both approaches do not solve neither the problem of meaning nor the problem of truth. Thus, a third approach – pragmatics – is needed to

deal with the relation of signs with its users, that is, of sentences with the people who speak them.

The main concepts of a pragmatist approach towards language are, thus, missing concepts in semantic and syntactic approaches.

Thus, under a pragmatist perspective, communication is inseparable from the social act in which it is embedded. Communication is the mediation that allows collective social action. G. H. Mead’s masterpiece *Mind, Self and Society* (1934) expresses in its title the central spots from which he understands human activity. ‘Mind,’ ‘Self’ and ‘Society’ are different elements of the same social process, his basic analytical units.

To Mead, the *act* (a complete unit of human conduct) derives from an impulse that produces perception, attribution of meaning, evaluation by participants and a final outcome: a process, which cannot be analyzed from just one of its parts. Mead’s notion of society presupposes conscious individuals, who actively interpret the world in a never-ending relational process.

FUTURE RESEARCH DIRECTIONS

The understanding of information as an inseparable part of the integral process of meaning production converges to some of the most urgent and imperative demands of our era. We live in a society of individuals, of advancing of human and political rights and a fast pace of technological development. The outcomes of the process of cultural integration to the newest technology of communication are yet unconcluded and unknown (Braga & Logan, 2014; Hargittai & Hinnant, 2008; Livingstone, 2004; Park, 2012; Van Deursen & Van Dijk, 2010, 2015; Van Dijk, 2005).

Thus, research on Human-Computer-Interaction (HCI) may have a significant development if it includes into its research agenda the sensibility to social contexts in which its phenomena take place.

Future research directions would include a re-evaluation of the use of Shannon Information

theory in the various disciplines in which it has been implemented to determine in what manner information and signal have been conflated. A signal no matter how accurately it has been transmitted and how accurately its information content has been calculated using Shannon's mathematical formula still has no meaning if the context or the pragmatics has not been also considered.

Another research trend to be explored regards the role of technologies of information and communication in social environments, where those technologies are essential to their organization, strategic development and their outcome.

CONCLUSION

What is communication? There are two alternative communication concepts that have coexisted since the term began to be used in the mid-nineteenth century: "communication as transmission" and "communication as ritual". Trying to answer the difficult question: what is communication?, Carey states that "communication is a symbolic process whereby reality is produced, maintained, repaired and transformed" (1989, p. 23) This definition challenges the notion (of Cartesian origin) that there is a dualism between 'fact' and 'discourse', between 'reality' and 'language.' 'reality' is a symbolic, social product. "The reality is constructed by communication through the use of symbolic forms." (Carey, 1989, p. 25)

All these distinctions, apart from their metaphorical subtleties, point out to the difference between a group of communication theorists who see communication as a matter of *transmission of information* (that is, syntax + semantics), and, on the other side, a group of theorists that regard communication as a matter of *social interaction* (syntax + semantics + pragmatics). In the first group, focusing on 'information', people (the "audience" or the "public") are seen as simple recipients receiving their messages, regardless of what they can actually do or think about it. The

second group of theorists, in order to account for the pragmatics of social situations of 'follow the actors', integrate cultural approaches with communication practices. (Braga, 2008)

Depending on the emphasis a theory places on transmission or interaction results a completely different picture of what is "communication", "information" or "reality". These approaches are in fact complimentary and would enhance each other's reach and analytic potential if integrated.

As we have stated before, 'information,' 'communication' and 'social interaction' are inextricable elements of the production of meaning. By denying the pragmatic dimension of communication (reducing communication phenomena to the 'content' – or information) makes it easier for a researcher to analyse (as the avoidance the pragmatics would certainly reduce complexity). However, this reduction will have a cost: as we have seen, when pragmatics is not considered, there is a loss of meaning, this reduction causes the phenomenon to 'disappear'.

The transmission and social perspectives are not contradictory, but complementary. To emphasize the relational perspective is to value the idea that communication is human action, made by people. Under ethnographic perspective, it is possible to observe values and principles of a system in place being contradicted or reversed, through social resilience and creative subversions of the industrial prescriptions. (Braga and Logan, 2014) This perspective stresses that communication problems are "community issues", which relate to the communities that they have created and in which we live.

Thus we conclude that the processes of 'transmission of information' and communication are not same. Communication requires that the information that is transmitted, which is only a signal after all, must be processed within the ground of pragmatics or social context. Put simply communication is contextualized information contextualized through pragmatics and social interactions.

REFERENCES

- Austin, J. (1962). *How to do things with words*. Oxford, UK: Oxford University Press.
- Bateson, G. (1973). *Steps to an Ecology of Mind*. St. Albans, UK: Paladin Frogmore.
- Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological and scholarly phenomenon. *Information Communication and Society*, 15(5), 662–679. doi:10.1080/1369118X.2012.678878
- Braga, A. (2008). *Personas Materno Eletrônicas* [Maternal Electronic Personae]. Porto Alegre, RS: Editora Sulina.
- Braga, A. A., & Logan, R. K. (2014). Social resilience in action. Social uses of cell phone in Brazil. In H. Rahman & R. Sousa (Eds.), *Information Systems and Technology for Organizational Agility, Intelligence, and Resilience* (pp. 297–315). Hershey, PA: Information Science Publishing. doi:10.4018/978-1-4666-5970-4.ch014
- Carey, J. (1989). *Communication as Culture*. Boston, MA: Unwin Hyman.
- Chan, N. N., Walker, C., & Gleaves, A. (2015). An exploration of students lived experiences of using smartphones in diverse learning contexts using a hermeneutic phenomenological approach. *Computers & Education*, 82, 96–106. doi:10.1016/j.compedu.2014.11.001
- Gibbs, M., Meese, J., Arnold, M., Nansen, B., & Carter, M. (2015). #Funeral and Instagram: Death, social media, and platform vernacular. *Information Communication and Society*, 18(3), 255–268. doi:10.1080/1369118X.2014.987152
- Gui, M., & Argentin, G. (2011). Digital skills of internet natives: Different forms of internet literacy in a random sample of northern Italian high school students. *New Media & Society*, 13(6), 963–980. doi:10.1177/1461444810389751
- Hargittai, E., & Hinnant, A. (2008). Digital inequality difference in young adults use of the internet. *Communication Research*, 35(5), 602–621. doi:10.1177/0093650208321782
- Hayles, K. (1999). *How We Became Posthuman*. Chicago, IL: University of Chicago Press. doi:10.7208/chicago/9780226321394.001.0001
- Hine, C. (Ed.). (2014). *Virtual Methods: issues on social research on the internet*. Oxford, UK: Berg.
- Introna, L. D., & Nissenbaum, H. (2000). Shaping the web: Why the politics of search engines matters. *The Information Society*, 16(3), 169–185. doi:10.1080/01972240050133634
- Kauffman, S. (2000). *Investigations*. Oxford, UK: Oxford University Press.
- Kauffman, S., Logan, R. K., Este, R., Goebel, R., Hobill, D., & Shmulevich, I. (2007). *Propagating organization: An enquiry. Biology and Philosophy*, 23(1), 27–45. doi:10.1007/s10539-007-9066-x
- Lee, M. R., & Chen, T. T. (2015). Digital creativity: Research themes and framework. *Computers in Human Behavior*, 42(1), 12–19. doi:10.1016/j.chb.2014.04.001
- Livingstone, S. (2004). Media literacy and the challenge of new information and communication technologies. *Communication Review*, 7(7), 3–14. doi:10.1080/10714420490280152
- Logan, R. (2011). Figure/Ground: Cracking the McLuhan code. *E-Compos*, 14(3), 1–13.
- MacKay, D. (1969). *Information, Mechanism and Meaning*. Cambridge, MA: MIT Press.
- McLuhan, M. (1964). *Understanding Media: Extensions of Man*. New York, NY: MacGraw Hill.
- Mead, G. H. (1974). *Mind, Self and Society: from the standpoint of a social behaviorist*. Chicago, IL: The University of Chicago Press.

Park, S. (2012). Dimensions of digital media literacy and the relationship to social exclusion. *Media International Australia*, 142(1), 87–100.

Richards, I. A. (1952). Communication Between Men: The Meaning of Language. In *Transactions of 8th Macy Conference - Cybernetics: Circular Causal and Feedback Mechanisms in Biological and Social System*. New York, NY: Josiah Macy, Jr. Foundation.

Shannon, C.E. (1948). A Mathematical Theory of Communication. *Bell Syst. Techn. J.*, 27, 379–423, 623–656.

Van Deursen, A., & Van Dijk, J. (2010). Internet skills and the digital divide. *New Media & Society*, 13(6), 893–911. doi:10.1177/1461444810386774

Van Deursen, A., & Van Dijk, J. (2015). Internet skills level increase, but gaps widen: A longitudinal cross-sectional analysis among the Dutch population. *Information Communication and Society*, 18(7), 782–797. doi:10.1080/1369118X.2014.994544

Van Dijk, J. (2005). *The deepening divide: inequality in the information society*. Thousand Oaks, CA: Sage.

Von Bertalanffy, L. (1968). *General Systems Theory: Foundations, Development, Applications*. New York, NY: George Braziller.

Wittgenstein, L. (2001). *Philosophical Investigations*. London: Blackwell Publishing.

ADDITIONAL READING

Becker, H. (2007). *Telling About Society*. Chicago, IL: University of Chicago Press.

Becker, H. (2008). *Art Worlds*. Los Angeles, CA: University of California Press.

Braga, A. (2005). *CMC, Identidades e Genero*. Covilhã PT: LabCom Books.

Button, G., Coulter, J., Lee, J. R. E., & Sharrock, W. (1995). *Computers, Minds and Conduct*. Cambridge, UK: Polity Press.

Goffman, E. (1967). *Interaction Ritual*. New York, NY: Anchor Books.

Goffman, E. (1986). *Frame Analysis*. New York, NY: Harper Row Press.

Luffi, P., Gilbert, N., & Frohlich, D. M. (Eds.). (1990). *Computer and Conversation*. London, UK: Academic Press.

Peirce, C. S. (1958). *Collected Papers of Charles Sanders Peirce*. Cambridge, MA: Harvard University Press.

Van Dijk, T. A. (2014). *Discourse and Knowledge: A sociocognitive approach*. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9781107775404

KEY TERMS AND DEFINITIONS

Communication: A symbolic process whereby the social reality is produced, understood, maintained and updated in a given situation. Communication is the mediation that allows collective social action. Communication is contextualized information through pragmatics.

Data: The pure and simple facts without any particular structure or organization, the basic atoms of information.

Information: Structured, organized data. It is composed by syntax and semantics, that is, with reference to a code and to a given meaning.

Knowledge: The ability to use information strategically in a social context to achieve one's objectives.

Naturalistic Perspective: Empiricist approach of the Social Sciences based on the premise of collecting data essentially from “natural” situations, those that happen despite of the presence or participation of the researcher.

Pragmatics: A discipline that stands between Philosophy and Linguistics, and tries to define to which degree the human sense of 'reality' is determined by language. The production of meaning is oriented to action, and that the idea of what a thing 'is' lies on the sum of the effects that can be conceived as possible from it. Pragmatics is the use of social context to assist meaning.

Social Interaction: Mutual action and/or influence among co-participants of the same social situation.

Wisdom: The capacity to choose objectives consistent with one's values and within a social context.

C

Data Journalism

Andreas A. Veglis

Aristotle University of Thessaloniki, Greece

Charalampos P. Bratsas

Aristotle University of Thessaloniki, Greece

INTRODUCTION

The introduction of ICTs (Information Communication Technologies) had a profound impact on every aspect of human activities. In the case of journalism, the utilization of ICTS has transformed the profession through the digitalization of the work process as well as the introduction of the internet along with its services (Veglis 2009). Today the journalist is expected to have the ability to firstly employ many tools and services in order to be instantly informed about breaking news as well as current events, and secondly, use a variety of tools and applications in order to prepare and disseminate news articles (Veglis & Bratsas, 2017). Many new types of journalism have emerged, among which, data journalism (Gray, Chambers, & Bounegru, 2012), which requires journalists to have special ICT skills.

In the recent years, data journalism has drawn significant attention in the academic literature as well as in the area of new developments in digital news production (Appelgren & Nygren, 2014; Fink & Anderson, 2015; Mair & Keeble, 2014). Data journalism is considered to be a new form of journalism. It has appeared gradually in the dawn of the new century. Many factors have contributed to the introduction of data journalism, but one of the most prominent is believed to be the availability of data in digital form (Veglis & Bratsas, 2017). Data Journalism is a journalistic specialty reflecting the increased role of the numerical data has in the production and distribution of information in the digital era. Data can be the source of data journalism, and/or it can be the tool

with which the story is told (Gray, Chambers, & Bounegru, 2012).

This chapter examines current trends and future perspectives of data journalism. The background section provides historic evolution and definitions of data journalism. Next, the stages of data journalism are presented in detail. Also, the relation between data journalism and open data is discussed due to the importance of the later in the development of data journalism. Finally, recommendations and future research direction are briefly discussed.

BACKGROUND

Evolution and Definition

Although the term data journalism started to attract attention at the end of the previous century, initial examples of data journalism appeared quite early. According to Simon Rogers the first example of data journalism was published at Guardian in 1821. It concerned the number of students who attended school and the costs per school in Manchester (Gray, Chambers, & Bounegru, 2012).

At the end of the 20th century, employing large data to write an article was difficult and required skills that went beyond the capabilities of the average journalist. That resulted in the phenomenon that some news organization in the United States and Great Britain were hiring programmers that worked on novel news products (Parasied & Dagiral, 2013). Traditionally, journalists used to rely on information provided by various sources

(governments, officials, research studies, etc.). Of course, there were some cases of investigative journalism where journalists were able to find resources to gather and analyze their own data and publish their results in articles (Veglis & Bratsas, 2017). But as a growing amount of data gradually became available online, and efficient tools with which anyone can analyze, visualize and publish large amounts of data appeared, things changed significantly (Sirkkunen, 2011).

The concept of data journalism is not new. It has been around since the beginning of the digitalization. Digital data has been utilized in news production since the late 60s in US newspapers (Parasied & Dagiral, 2012). Data journalism gradually emerged with the rapid introduction of ICTs and the availability of data in digital form. The term data journalism is synonymous with data-driven journalism while the older term, computer-assisted reporting has vanished since it was introduced at the early stages of computer history (Bradshaw, 2010). It is worth noting that in the case of data journalism there is an increased interaction between journalists and several other fields such as design, computer science and statistics (Thibodeaux, 2011; Veglis & Bratsas, 2017).

The term data journalism is attributed to Simon Rogers that first mentioned it in a post to the Guardian Insider Blog (Knight, 2015). It can be viewed as a process that begins with analyzing, and continues with filtering and visualizing data in a form that links to a narrative (Lorenz, 2010). It combines spreadsheets, graphics data analysis and the biggest news stories (Rogers, 2008). It is fundamentally the production of news graphics and includes elements of design and interactivity (Bradshaw, 2010; Lorenz, 2010; Rogers, 2008). Megan Knight (2015) describes data journalism as “a story whose primary source or “peg” is numeric (rather than anecdotal), or a story which contains a substantial element of data or visualization”.

Veglis and Bratsas (2017) proposed a definition in order to better address the power of visualization and interactivity that are significant factors in data journalism. They defined data journalism

as the process of extracting useful information from data, writing articles based on the information and embedding visualizations (interacting in some cases) in the articles that help readers understand the significance of the story or allow them to pinpoint data that relate to them.

Journalists' Skills

Today the journalist is expected to possess various ICT skills in order to cope successfully with the challenges in his everyday work (Veglis & Bratsas, 2017). Typical examples are writing news articles, constructing diagrams via spreadsheet applications, communication via e-mail, visualizing data with the help of various applications, publishing material on the WWW (Peebles 2011). Also, they often seek information on the web and by e-mail (Veglis, 2013). Veglis and Pomportsis (2012, 2014) organized the journalists' ICT skills into five categories, namely *Basic skills*, *Web publishing skills*, *Web 2.0 skills*, *Web casting skills* and *Data Journalism skills*. Veglis & Bratsas (2017) extended this categorization by adding a category that refers to *Web 3.0*. Next, the six categories are briefly presented:

- **Basic Skills:** The journalist has the ability to work efficiently with office automation suites (which include word processing, spreadsheet, presentation, database), and with the basic Internet services (WWW, email). Specifically, the journalist is expected to have basic typing and formatting skills, and he must be able to perform basic functions in a spreadsheet. He must have at least a general understanding of how to use data to support news stories and he must also be able to use relational database programs to cross-check those data files to find various information (Veglis & Pomportsis, 2014).
- **Web Publishing Skills:** In this case, basic knowledge of HTML is considered to be a necessary prerequisite, as well as the abil-

ity to use Content Managements Systems. The journalist should have an understanding of the basic concepts of HTML and cascading styling sheet (CSS) (Peebles, 2011). Except web publishing, the journalist must be able to work with widely used Desktop Publishing Applications (Veglis & Pomportsis, 2014).

- **Web 2.0 Skills:** Web 2.0 includes the utilization of various tools and services such as blogs, RSS, wikis, social bookmarking, and social networking (Facebook, Google+, Twitter, etc.). The journalist must be able to update the media organization's profile on social networks, and also to interact with the audience through various services, such as blogs, Twitter, Facebook, Google+, etc. (Veglis & Pomportsis, 2014). It is worth noting that the majority of the Web 2.0 tools and services can also be used by the journalists in order to receive updates on current events from various news organizations.
- **Webcasting Skills:** In this case journalists must have the necessary skills to create and publish podcasts and videocasts. The journalist must be able to record the audio of an interview, perform simple editing on the audio recording of that interview, and upload it to the WWW for the audience to have access to the podcast. In the case of videocasts, the journalist must be capable of making, at least, a short video story even if it is shot with a smartphone and must have the skills of using entry-level nonlinear video editing software, to move

scenes around, to create a basic news video (Fletcher, 2008). Podcasts are considered to be an effective way to use audio to broadcast news stories since they are easy to create and their size is relatively small. Videocasts are considered to be powerful tools to convey news, but they are more difficult to create in comparison with podcasts and their sizes are quite large (Veglis & Pomportsis, 2014).

- **Web 3.0 Skills:** Include basic knowledge and experience with Web 3.0 technologies. Journalists must be able to a) understand the basics of Web 3.0, b) use the limited Web 3.0 tools currently available (but expected to grow exponential in the near future) c) create articles that comply with open data rules and link to other open data (Veglis & Pomportsis, 2014).
- **Data Journalism Skills:** Journalists must be able to find datasets, clean and filter the data, put them in context, find the story in the data, visualize the results and integrate them into a news story (Veglis & Bratsas, 2017).

DATA JOURNALISM STAGES

Veglis and Bratsas (2017) organized the data journalism workflow in six stages, entitled: *Data Compilation*, *Data Cleaning*, *Data Understanding*, *Data Validation*, *Data Visualization* and *Article Writing*. The workflow model is depicted in figure 1. Next, we briefly describe each stage.

Figure 1. Data journalism workflow



- **Data Compilation:** A data journalism project begins in one of two ways: either the journalist has a question that needs data, or a dataset that needs questioning. The compilation of data can take one of the following forms: (i) data may be supplied directly by an organization (in some cases in the form of open data), (ii) data may be found with the help of advanced searching techniques, (iii) data may be compiled by scraping web pages, (iv) data may be collected by converting documents to other formats that can be analyzed, and (v) data may be collected by means of observation, surveys, online forms or crowdsourcing (Bradshaw, 2011). Data scraping is a process in which a software tool extracts data from a human-readable output that originates from other software. Also, there are free tools on the internet that allow users to convert documents to other forms that can be analyzed. For example, DocumentCloud is a web-based software platform created specifically for journalists to allow the searching, analyzing, annotation and publication of primary source documents used in reporting (Veglis & Bratsas, 2017).
- **Data Cleaning:** Also known as data scrubbing is the process of detecting and correcting corrupted or incorrect records from a dataset (Wu, 2013). This can be accomplished by removing human errors and converting the data into a format that is consistent with other data the journalist is using. Typical examples include empty entries, duplicate entries, the use of default values to save time or where no information was held, incorrect formatting (for example words instead of numbers), corrupted entries or entries with HTML code, multiple names for the same thing and missing data (Bradshaw, 2011). Cleaning data can be done in simple ways, for example using find and replace commands or filters in spreadsheets. There are also specialized tools, like Google's OpenRefine, which is a standalone open source desktop application for data cleanup and transformation to other formats. The application's interface resembles spreadsheet applications and it can open files of various spreadsheet formats (Veglis & Bratsas, 2017).
- **Data Understanding:** Datasets usually include various codes that represent categories, classifications or locations, and special terminology that it is not understood by journalists. Frequently further data is required in order for existing data to become meaningful. Overall journalists must be data-literate, meaning that they must be able to consume knowledge, produce coherently and think critically about data (Veglis & Bratsas, 2017). Additionally, journalists must possess statistical knowledge and also understand how to work with large datasets, how they were produced, how to connect various datasets and how to interpret them (Gray, Chambers, & Bounegru, 2012).
- **Data Validation:** This stage includes the process of cross-checking the original data and obtaining further data from sources in order to enrich the available information (Silverman, 2014; Veglis, 2013). It is worth noting that like any source, datasets cannot always be trusted since they come with their own histories, biases, and objectives. That means that journalists have to investigate issues like: who gathered it, when, and for what purpose, and how it was gathered (Bradshaw, 2011). This can be accomplished by investigating the history of the creation of the dataset, by finding references to the dataset or by using other sources of information that refer to the same subject (Silverman, 2014; Veglis & Bratsas, 2017).
- **Data Visualization:** Data visualization is a modern branch of descriptive statistics that involves the creation and study

of the visual representation of data. It is the graphical display of abstract information for data analysis and communication purposes (Cairo, 2012). Static data visualizations offer only pre-composed “views” of data. Interactive data visualization supports multiple static views in order to present a variety of perspectives on the same information. Important stories include “hidden” data and interactive data visualization is the appropriate way to discover, understand and present these stories. In interactive data visualization, there is a user input (a control of some aspect of the visual representation of information) and the changes made by the user must be incorporated into the visualization in a timely manner (Veglis, 2015; Veglis & Bratsas, 2017).

It is worth mentioning that infographics are also part of the static visualization. Infographics are graphic visual representations of data or knowledge, which are able to present complex information quickly and clearly (Smiciklas 2012). They are often used in newspapers, to show the weather, as well as maps, site plans, and graphs for statistical data (Veglis & Bratsas, 2017).

There are many types of data visualizations. Heer, Bostock, and Ogievetsky (2010) defined the types (and also their sub-categories) of data visualization: (i) Time-Series Data (Index Charts, Stacked Graphs, Small Multiples, Horizon Graphs), (ii) Statistical Distributions (Stem-and-Leaf Plots, Q-Q Plots, Scatter Plot Matrix -SPLOM, Parallel Coordinates), (iii) Maps (Flow Maps, Choropleth Maps, Graduated Symbol Maps, Cartograms), (iv) Hierarchies (Node-Link Diagrams, Adjacency Diagrams, Enclosure Diagrams) and (v) Networks (Force-Direct Layout, Arc Diagrams, Matrix Views).

Today there are a lot of available online tools that can be used for creating interactive data visualizations. All of them are either free or offer a free version (except a paid one that includes more

features). Some of the most widely used tools are CartoDB, Google Chart Tools, Google Fusion Tables, Tableau Public, Many Eyes, and Infogr.am. A complete list of the available tools can be found at <http://selection.datavisualization.ch>.

Article Writing: The last stage in a data journalism project includes the writing of the news article. Depending on the intended publication medium, the article may include special characteristics (for example external links other articles or related material, multimedia content, mashups, static or interactive visualizations) in order to fully take advantage of the medium’s potentials (Veglis & Bratsas, 2017). The amount of text that is included in the data journalism article along with the visualizations may vary considerably. Specifically, we can have the case where the visualization supplements the text (which is quite extended) as well as the case where the visualization is the center of the project and the text plays a supplemental role, explaining parts of the visualization.

FUTURE RESEARCH DIRECTIONS

Tim Berners-Lee believes that “data-driven journalism is the future” and urges reporters to hunt for stories in datasets (Arthur, 2010). We are convinced that the Semantic Web and Linked Open Data will play a significant role in the evolution of data journalism. In the future, journalists should be able to comprehend and utilize advanced technologies which will include more “intelligent web” (Bradshaw & Rohumaa, 2011). Semantic Web technologies, smart devices (smartphones etc.) and tools are continuously being transformed and upgraded. Data journalism ought to adopt these technologies. It is worth noting that Web 3.0 and open data can play an important role in the case of data journalism as they support the acquisition and validation of data which are the main source of data journalism.

Tim Berners-Lee, the inventor of the Web and Linked Data initiator developed a 5-star deployment scheme for Open Data, in order to encourage

Figure 2. Star rating system for the use of good linked open data for journalists



people – especially government data owners to produce well linked Open Data (2010). Veglis and Bratsas (2017) proposed an adaptation of the system for the journalists in order to act as a guide for them to produce articles in the Web 3.0 era. The star rating system is presented in figure 2 and includes the steps that a journalist can take in order to accomplish good Web 3.0 practices in journalism.

One other issue that needs to be addressed is the lack of online platforms that can support all stages of a data journalism project. Today there are several online tools that can be used in different stages of a data journalism project. Each tool has its own unique interface and the journalist must spend time in order to learn how to use it effectively. The majority of them do not offer any interconnectivity with other tools. Thus, the jour-

nalist spends a lot of time in order to transfer data from one platform to another in order to exploit the strong points of each tool. Data journalists can strongly benefit from the introduction of an online platform that will interface various online tools that can be used in data journalism projects.

**it is a string of characters used to identify a name of a resource over the WWW. The most common form of URI is the uniform resource locator (URL), frequently referred as a web address.*

CONCLUSION

This chapter discusses the issue of data journalism. Data journalism is considered to be a journalistic specialty that is expected to flourish in the world

of data that we are living today. Data journalism stages were presented and discussed in detail. These stages define the skills that a journalist ought to possess in order to develop data journalism projects.

The above findings can guide journalism educators in order to adapt their programs. This is already happening since the constant changes in market demands, as far as journalistic skills are concerned, has forced journalism educators to adjust their programs in order to better facilitate the needs of the industry (Wenger & Owens, 2012). Currently data journalism is taught in a limited number of journalism schools in Europe. Of course this is something worth investigating more thoroughly in the near future. A recent survey, concerning the use of data journalism in Greece, found a low penetration of data journalism practices in media organizations. Nevertheless the majority of the journalists appeared to agree with the importance of working with data and was interested in acquiring more knowledge and skills concerning data journalism (Veglis & Bratsas, 2017).

There is no doubt that the majority of the journalists today are not involved in data journalism projects, although they are willing to work on such projects in the future. Online journalists appear to be more involved and more motivated towards data journalism and since there is an ongoing shift to online journalism (Cokley, Edstrom, McBride, & Ranke, 2015), we have to expect many new developments in the area of data journalism.

REFERENCES

- Appelgren, E., & Nygren, G. (2014). Data Journalism in Sweden Introducing new methods and genres of journalism into old organizations. *Digital Journalism*, 2(3), 394–405. doi:10.1080/21670811.2014.884344
- Arthur, C. (2010, November 22). Analysing data is the future for journalists, says Tim Berners-Lee. *The Guardian*. Retrieved March 18, 2016, from <http://www.theguardian.com/media/2010/nov/22/data-analysis-tim-berners-lee>
- Berners-Lee, T. (2010). Linked Data- Design Issues. *World Wide Web Consortium (W3C)*. Retrieved March 18, 2016, from [http://www.w3.org/DesignIssues/ LinkedData.html](http://www.w3.org/DesignIssues/LinkedData.html)
- Bradshaw, P. (2010, October 1). How to be a data journalist. *The Guardian Data Journalism*. Retrieved March 18, 2016, from <http://www.Guardian.co.uk/news/datablog/2010/oct/01/data-journalism-how-to-guide>
- Bradshaw, P. (2011). The inverted pyramid of data journalism. *Online Journalism Blog*. Retrieved March 18, 2016, from <http://onlinejournalismblog.com/2011/07/07/the-inverted-pyramid-of-data-journalism>
- Bradshaw, P., & Rohumaa, L. (2011). *The Online Journalism Handbook: Skills to survive and thrive in the digital age*. Longman Publishing Group.
- Cairo, A. (2012). *The Functional Art: An introduction to information graphics and visualization*. New Riders.
- Cokley, J., Edstrom, M., McBride, J., & Ranke, A. (2011). Moving Away from 'Big Media': Students, Jobs and Long-tail Theory. *Australian Journalism Review*, 33(1), 107–118.
- Fink, K., & Anderson, C. W. (2015). Data Journalism in the United States. *Journalism Studies*, 16(4), 467–481. doi:10.1080/1461670X.2014.939852
- Fletcher K. (2008). *Journalism Skills Survey*. National Council for the Training of Journalists.
- Gray, J., Chambers, L., & Bounegru, L. (2012). *The Data Journalism Handbook*. Retrieved March 18, 2016, from <http://datajournalismhandbook.org>

- Heer, J., Bostock, M., & Ogievetsky, V. (2010). A tour through the visualization zoo. *Magazine Communications of the ACM*, 53(6), 59–67. doi:10.1145/1743546.1743567
- Knight, M. (2015). Data Journalism in the UK: A preliminary analysis of form and content. *Journal of Media Practice*, 16(1), 55–72. doi:10.1080/14682753.2015.1015801
- Lorenz, M. (2010). *Data driven journalism: What is there to learn?* Paper presented at IJ-7 Innovation Journalism Conference, Stanford, CA.
- Mair, J., & Keeble, R. L. (2014). *Data Journalism: Mapping the future*. Abramis Academic Publishing.
- Parasie, S., & Dagiral, E. (2013). Data-driven journalism and the public good: Computer-assisted-reporters and programmer-journalists in Chicago. *New Media & Society*, 15(6), 853–871. doi:10.1177/1461444812463345
- Peebles, J. (2011, March 9). Digital media skills every young journalist needs. *Net Worked*. Retrieved March 18, 2016, from <http://blogs.spjnetwork.org/tech/2011/03/09/digital-media-skills-every-young-journalist-needs>
- Rogers, S. (2008, December 18). Turning official figures into understandable graphics, at the press of a button. *Inside the Guardian Blog*. Retrieved March 18, 2016, from <http://www.theguardian.com/help/insideguardian/2008/dec/18/unemploymentdata>
- Silverman, C. (2014). *Verification Handbook*. European Journalism Centre. Retrieved March 18, 2016, from http://verificationhandbook.com/downloads/verification_handbook.pdf
- Sirkkunen, E. (2011). *Trends in Data Journalism*. Wp 3, Hyperlocal D.3.2.1.2.b. Retrieved March 18, 2016, from http://virtual.vtt.fi/virtual/next-media/Deliverables-2011/D3.2.1.2.B_Hyperlocal_Trends_In%20Data_Journalism.pdf
- Smiciklas, M. (2012). *The Power of Infographics: Using Pictures to Communicate and Connect with Your Audience*. Pearson Education Inc.
- Thibodeaux, T. (2011, October 6). 5 tips for getting started in data journalism. *Poynter: A Global Leader in Journalism*. Retrieved March 18, 2016, from <http://www.poynter.org/news/media-innovation/147734/5-tips-for-getting-started-in-data-journalism/>
- Veglis, A. (2009). Cross Media Communication in newspaper organizations. *4th Mediterranean Conference on Information Systems*, Athens, Greece.
- Veglis, A. (2013). Education of Journalists on ICTs: Issues and Opportunities. *Journal of Applied Journalism & Media Studies*, 2(2), 265–279. doi:10.1386/ajms.2.2.265_1
- Veglis, A. (2015). *Teaching Data Journalism in the School of Journalism & MC-Greece*. Paper presented at Ejta2015, Narrative Journalism and Storytelling on Digital Platforms (Teacher Conference), Milan, Italy.
- Veglis, A., & Bratsas, C. (2017). Journalists in the age of Data Journalism: The case of Greece. *Journal of Applied Journalism & Media Studies*, 6(2).
- Veglis, A., & Pomportsis, A. (2012). The e-citizen in the cyberspace – A journalism aspect. In *Proceedings of the 25th International Conference on Information Law and Ethics*. Kerkyra, Greece: Nomiki Bibliothiki.
- Veglis, A., & Pomportsis, A. (2014). Journalists in the age of ICTs: Work demands and Educational needs. *Journalism & Mass Communication Educator*, 69(1), 61–75. doi:10.1177/1077695813513766
- Wenger, D. H., & Owens, L. C. (2012). Help Wanted 2010: An Examination of New Media Skills Required by Top U.S. News Companies. *Journalism & Mass Communication Educator*, 67(1), 9–25. doi:10.1177/1077695811426826

Wu, S. (2013). A review on coarse warranty data and analysis. *Reliability Engineering and System*, 114, 1–11. doi:10.1016/j.ress.2012.12.021

ADDITIONAL READING

Allemang, D., & Hendler, J. (2011). *Semantic Web for the Working Ontologist* (2nd ed.). Morgan Kaufmann.

Auer, S. R., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., & Ives, Z. (2007). DBpedia: A Nucleus for a Web of Open Data. *The Semantic Web. Lecture Notes in Computer Science*, 4825, 722–735. doi:10.1007/978-3-540-76298-0_52

Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked Data - the story so far. *International Journal on Semantic Web and Information Systems*, 5(3), 1–22. doi:10.4018/jswis.2009081901

Bradshaw, P. (2011, September 22). Data journalism survey: a mixed picture. *Data Driven Journalism*. Retrieved March 18, 2016, from http://datadrivenjournalism.net/news_and_analysis/data_journalism_survey_analysis

Brickley, D., & Guha, R. V. (2014). RDF Shcema. Retrieved March 18, 2016, from <http://www.w3.org/TR/2014/REC-rdf-schema-20140225>

Bull, A. (2010). *Multimedia Journalism: A Practical Guide*. London, New York: Routledge.

Card, S., Mackinlay, J., & Shneiderman, B. (1999). *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann.

Cokley, J., Gilbert, L., Jovic, L., & Hanrick, P. (2016). Growth of Long Tail in Australian journalism supports new engaging approach to audiences. *Continuum (Perth)*, 30(1), 58–74. doi:10.1080/10304312.2015.1099152

Dietrich, D., Gray, J., McNamara, T., Poikola, A., Pollock, R., Tait, J., & Zijlstra, T. (2012). *The Open Data Handbook*. An Open Knowledge Foundation project, Retrieved March 18, 2016, from <http://opendatahandbook.org>

European Journalism Centre. (2010). Data-driven journalism: what is there to learn? Retrieved March 18, 2016, from http://mediapusher.eu/datadriven-journalism/pdf/ddj_paper_final.pdf

Friedman, V. (2007, August 2). Data Visualization: Modern Approaches. *Smashing Magazine*. Retrieved March 18, 2016, from <http://www.smashingmagazine.com/2007/08/02/data-visualization-modern-approaches>

Friedman, V. (2008, January 14). Data Visualization and Infographics. *Smashing Magazine*. Retrieved March 18, 2016, from <http://www.smashingmagazine.com/2008/01/14/monday-inspiration-data-visualization-and-infographics>

Friendly, M. (2007). A Brief History of Data Visualization. in C. Chen, W. Hardle and A. Unwin (eds.) *Handbook of Computational Statistics: Data Visualization*, III, Springer. Retrieved March 18, 2016, from <http://www.datavis.ca/papers/hbook.pdf>

McGuinness, D. L., & van Harmelen, F. (2014). Ontology Web Language OWL. Retrieved March 18, 2016, from <http://www.w3.org/TR/owl-features>

Murray, S. (2013). *Interactive Data Visualization for the Web*. O'Reilly.

Open Knowledge. (n.d.). Open Data Handbook. Retrieved March 18, 2016, from <http://opendatahandbook.org/>

Prud'hommeaux, E., & Seaborne, A. (2013). SPARQL Query Language for RDF. Retrieved March 18, 2016, from <http://www.w3.org/TR/rdf-sparql-query>

Shneiderman, B. (1996). The eyes have it: a task by data type taxonomy for information visualizations. In Proceedings at the *IEEE Symposium on Visual Languages*, 336 – 343. doi:10.1109/VL.1996.545307

Spyridou, L. P., Masiola, M., Veglis, A., Kalliris, G., & Dimoulas, C. (2013). Journalism in a State of Flux: Changing Journalistic Practices in the Greek Newsroom. *International Communication Gazette*, 75(1), 76–98. doi:10.1177/1748048512461763

Vatant, B., & Vandenbussche, P. Y. (2014). *Linked Open Data Vocabularies*, Retrieved March 18, 2016, from <http://lov.okfn.org/dataset/lov>

Ward, M., Grinstein, G., & Keim, D. (2010). *Interactive Data Visualization: Foundations, Techniques, and Applications*. A K Peters/CRC Press.

Ware, C. (2008). *Visual Thinking: for Design*. Morgan Kaufmann Series in Interactive Technologies.

Ware, C. (2012). *Information Visualization. Third Edition: Perception for Design.* Morgan Kaufmann.

KEY TERMS AND DEFINITIONS

Data Journalism: The process of extracting useful information from data, writing articles based on the information and embedding visualizations in the articles that help readers understand the significant of the story or allow them to pinpoint data that relate to them.

Dataset: A collection of data that contains individual data units organized in a specific way and accessed by a specific access method that is based on the data set organization.

Data Scraping: The process in which a software tool extracts data from human-readable output that originates from other software.

Data Cleaning or Data Scrubbing: The process of detecting and correcting corrupted or incorrect records from a dataset.

Data Validation: The process of cross-checking the original data and obtaining further data from sources in order to enrich the available information.

Data Visualization: The graphical display of abstract information for data analysis and communication purposes.

Open Data: Data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and share alike.

Investigating Diachronic Variation and Change in New Varieties of English

Rita Calabrese

University of Salerno, Italy

BACKGROUND

The Study of Indian English From a Language-Contact Perspective

The issue of contact instantiating processes of pidginization and creolization can be differentiated according to varieties that arise through contact with languages coming outside the area, especially through colonialism and varieties that arise through internal contact among languages already indigenous to the area (Schiffman 2010, p.741). From this perspective, the South-Asian region can be considered a paradigm example of the phenomenon known as “convergence area” (Weinreich 1958) referred to phenomena specifically occurring in language contact situations that lead to changes in all areas of grammar.

Some decades after Weinreich’s descriptive study, Charles Ferguson (1992) published a comprehensive essay on some features of “language use” that make South Asia an interesting subject of study as “sociolinguistic area”. His attention to “shared patterns of use” and not only of shared grammatical structures was an important step towards the understanding of general processes of language change in contact situations. Ferguson’s particular perspective, in fact, “looks for the relationship between diachronic language change and language development, phonology and syntax, social conventionalization and cognitive processing, and language universals and individual differences” (Huebner 1996, p.12). Consequently, in studies concerning the development of new varieties, a crucial issue is the extent to which

universals of language (Pinker 2003, p.23) and language contact exert their influence on shaping those language systems. Recent research has tested current hypotheses on the interrelationship between language universals and language variation and given rise to new challenging theories on contact varieties. Namely, the notion of “vernacular universals” (Chambers 2004) limits the supposed tendency towards the *absolute creativity* of these varieties relying on the identification of universally shared features across varieties of English around the world.

The Study of Indian English From a Comparative Perspective

To date, two important issues have not been exhaustively examined by experts in the field of variationist studies: the first deals with the characterization of Standard English at the time of colonization that was slightly different from today’s standard against which new varieties of English are usually investigated, the second concerns possible internal and deterministically governed developments occurring in both early / late Modern English and its new emerging varieties¹.

When considering the peculiar contact situation in the Asian subcontinent English has played a major role in influencing local South Asian languages, though it was not the first European language to have an impact on them since Portuguese was already attested in the area before the founding of the East India Company in the early 1600s. By the end of the 18th century, the knowledge of English had grown greatly and re-

placed Portuguese as the *lingua franca* of India (Nihalani 2005, vi). However, it is worth noting that the *General Report on the Census of India, 1891* still records a low average percentage (4.4%) of ‘those who know English’ (Baines 1893: 224) and “not anyone who learned English in India was taught directly by a native speaker of the language” (Nihalani 2005, vi). Moreover, the local British community is at that time supposed to be expanding to what Schneider (2007, p. 37) defines as “British plus: genuinely British no doubt, but seasoned with the additional flavor of the colonial experience which those who stayed ‘home’ do not share”. ‘Colonial lag’ is the expression used to refer to the consequent conservatism in colonial varieties as a potential factor in distinguishing them ‘from their home counterparts in all levels of language’ (Bauer 2002, p. 5). Overall, some features of a colonial dialect can be predicted from the form used by the majority of the settlers (Bauer, 2002, p. 11) who, in the case of India, came originally from the city of London (Salaja 2009, p. 95)². Anyway, ‘in the colonial situation, a lot of speakers of many different dialects come face to face, and in the short term the result is a period of diversity where everyone is accommodating to everyone else [...]. In most cases the form used by the majority will be the form that survives in the new mixed dialect’ (Bauer 2012, p. 8).

The Emergence of a New Variety

Apart from deviations and occasional ‘errors’, it is therefore possible to classify the features of IndEng into three groups (Meshtrie & Bhatt 2008, p. 47):

1. Dialect features of the superstrate which have survived despite the norms of instructed English
2. Features of early modern English surviving in a particular colony
3. ‘True’ innovations in IndEng with no equivalent forms in modern English.

Some of the possible effects of these factors on the characterization of the emerging variety can be seen in Table 1. Significantly, when comparing the number of innovations characterizing IndEng and late Modern English, the distance between them dramatically reduces at least for some of those aspects that are generally considered as Indian English-specific (Calabrese 2012, 2015).

In particular, the increase in Verb + particle combinations⁴ within the class of multi-word verbs, made up of a common verb, often one syllable combined with a preposition, has been considered as one of the most important characteristics of the modern English vocabulary (Baugh & Cable 2002, p. 345). As a matter of fact, the changing uses of prepositions clearly reflect the idiomatic changes⁵ in a language from one age to another (ibid., p. 248) following the developmental path from synthetic⁶ to analytical constructions. This steady process dating back to Old English led to a gradual structural shift “from a productive system of verbal prefixes to a new system of post-verbal particles [in which] phrasal verbs as well as prepositional verbs come to be the functional equivalents of the older prefixed verbs” (Brinton 1988, p. 185). The same tendency to favor phrasal verbs instead of simplex verbs to express aspectual meanings can frequently be observed in contact situations as well (Danchev 1992, p. 30). The resulting poly-verbal constructions represent a type of paraphrase that has been

Table 1. Linguistic features emerging in Late Modern English and Indian English

Linguistic feature	Late Modern English	Indian English
Increase in Verb+particle combinations	✓	✓
Extensive use of Verb+particle combinations as nouns	✓	✓
Tendency to eliminate 3 rd pers –s marking of <i>do</i> ³	✓	✓
Lexical expansion (borrowings, loanwords, neologisms)	✓	✓

described as an ‘achievement strategy’ that helps communication when the semantic components of a lexical gap in the interlanguage are rearranged linearly and made explicit through separate free morphemes (Danchev, 1992).

Earlier research (Sorace 2005; Calabrese 2008, 2010) on the interpretation of prepositional phrases (PPs) as adjuncts or arguments of verb phrases in Second Language Acquisition (SLA) had highlighted the difficulty by second language learners to automatically integrate syntactic knowledge and information from other domains of language (e.g. morphology and semantics). As a result, the most discriminative feature in determining the choice of a given preposition proved to rely on semantic factors directly depending on the class of the verb. The logical form associated with a given verb is supposed to enhance the interpretation and production of verbs with similar logical forms and sub-categorization frames (Calabrese 2010: 58). For example, the verb *provide* belongs to the same semantic domain as *give* so that the same sub-categorization frame and logical form are assigned to it in the process of SLA leading to deviant forms such as **provide to* attested in IndEng.

A CASE STUDY

The concepts outlined so far and the related methodologies of investigation have provided the rationale for the present preliminary study aiming to diachronically investigate the development of new verb-particle combinations in spoken or speech-like data. Following a corpus-driven approach to the analysis of data, a sample diachronic corpus of Indian English has been queried in order to investigate the structural properties of this new variety from a diachronic perspective. Indeed, the study is based on the assumption that looking backwards at past stages of language development will contribute to identify convergence phenomena over time and gradually reconstruct the shape of a new variety as a standardized language. In

particular, the main focus will be on changes in the lexico-grammar interface which shows more gradient properties typical of structural nativization (Schneider 2007).

Since a key aspect of corpus design for most studies concerning diachronic variation is including the range of linguistic variation that exists in a language in a given historical period and not the proportions of variation (Biber et al. 1998, p.247), a stratified approach has been adopted by sampling data from different genres and time periods to study the frequency and use of verb-particle constructions as a structural property of IndEng rather than an effect of its ‘deviant use’ of English.

METHOD

Materials

The corpus, on which the present study is based, combines data of the Diachronic Corpus of Indian English (henceforth DiCIE)⁷ covering the years from 1835 to 2010 and parallel selected sections of the Indian component of the International Corpus of English (ICE-IND) dating back to 1978. The corpus has been designed on the model of both ICE-IND (Greenbaum 1996) and the Corpus of Contemporary Indian English (CCIE, Balasubramanian 2009) so that the final version of the entire corpus will have a comparable, balanced configuration with respect to the model corpora considered in the research. The corpus contains printed editions of the Indian national newspaper *The Statesman* that were copied from microfilms of the 19th century issues and are part of the British Library Newspaper Collection.

For the purpose of this preliminary study, certain sections of ICE-IND including spoken data such as speeches, reported legal cross-examinations and written data such as letters⁸ to the editor (tagged as <S>, <LCE>, <LE> respectively) have been selected in order to be compared to similar samples from the DiCIE.

More specifically, the spoken data includes audio files of Gandhi (<GANDH-S>), the past Prime Minister Singh (<SINGH-S>) and the MP Lalu Prasad's (<LP-S>) speeches sampled for the years 1930-47 and 2010 respectively⁹.

Procedure

The sections of the corpus including <LCE> and <LE> were created by converting in electronic format the printed texts published in *The Statesman* between the years 1835-1951.¹⁰ The oral data derive from audio/video files available online along with their transcriptions convertible in .txt files to make them readable by concordancers and annotation tools. The collected data were automatically parsed by using the language analysis tools available at the VISL website (<http://beta.visl.sdu.dk/>). The parsers provided by the VISL interface are based on the theoretical framework of the *Constraint Grammar*, a methodological paradigm widely adopted in *Natural Language Processing* (NLP) which can provide both syntactic and semantic information on a given constituent structure by assigning tags of lemmatization, inflection, derivation, syntactic function, constituent dependency, valency, semantic classification. The system also marks the *dependency relations* between parts of speech (POS) with the symbol @ placed before (>) or after (<) the head and proves therefore to be particularly useful for investigations on lexico-grammatical and morpho-syntactic patterns in specific variety usage. Upper case tags describe word classes as well as morphological inflection (e.g. MV= main verb, PRP= preposition, N = noun, GN= genitive). Once annotated, the corpus can be therefore queried for specific tags using a concordancer to find out possible examples of convergence or divergence from the standard usage in the variety under investigation.

Verb-particle combinations were searched for in the corpus and then mapped onto VISL tags by observing the constituent structure of the Verb Phrases and the Prepositional Phrases along with their functional categories. In order to accurately

classify and estimate all MV+PRP occurrences in the annotated corpus, a specific syntactic setting was established in the queries and all examples of MVs occurring with PRPs as their right collocates could easily be extracted. The features examined in the study and the corresponding VISL tags are shown in Table 2. As an example, the following tags have been searched for in order to study the frequency of Verb-particle combinations (tagged as PIV) as well as verb + adjunct constructions (tagged as ADVL) in Indian English.

Examples (1) and (2) show different annotations for different functional interpretations of the PPs in the annotated corpus.

In example (1) the PERS pronoun [*him*] is annotated as an argument of the preposition (P) on the left (@P<, while the prepositional phrase [*upon him*] is annotated as an adjunct/free adverbial.

- (1) They[they] PERS 3P NOM @SUBJ> immediately[immediately] ADV @ADVL> **sprung**[spring] <mv> V PCP2 PAS @ ICL-N< **upon** [upon] **PRP** @<**ADVL** him [he] PERS MASC 3S ACC @P<, [,] PU @ PU (<LE>, 1835)

In example (2), the prepositional phrase is not interpreted as an adverbial or adjunct and the noun (*copies*) is annotated as prepositional object / valency.

- (2) you [you] PERS 2S/P NOM @SUBJ> can [can] <aux> V PR @FS-<**ADVL** be [be] <aux> V INF @ICL-AUX< **furnished** [furnish] <mv> V PCP2 PAS @ICL-AUX<

Table 2. Features examined in the study and corresponding VISL tags

POS	Functional Category	Definition
MV /PRP	@[<] ADVL [>]	adjunct [free] adverbial
MV /PRP	@P<	argument of preposition
MV /PRP	@[<] PIV [>]	prepositional object / valency

with[with]PRP @ <PIV copies [copy] <cjt-head> <sem> <cc-r> <idf> <nhead> N P
NOM @P<

The starting point of the analysis was a selection of features (in grey) from the list of grammatical traits identified in past literature (Meshtrie and Bhatt 2008; Sailaja 2009; Schneider 2004, 2007, 2011; Schneider & Zipp 2013) as characteristic of IndEng (Table 3). Once annotated, tags for each selected feature could automatically be extracted from the corpus with the application of the *ConcApp* concordancer and then manually mapped to the corresponding structural patterns selected for the study (Table 3).

In order to estimate the occurrence of those patterns in British English as well, the findings were compared to sample data taken from an interactive database, namely *Phrases in English* (PIE) which provides quantitative information on recurrent phrase structures in the *British National Corpus* (BNC). The queries were ordered according to specific criteria such as minimum frequency of the searched item (e.g. $n=1$), data chunk size

(e.g. $n=1000$ which corresponds to the average number of tokens examined in DiCIE) and word forms or POS tags to match or exclude from a given query within a specified range of *n-grams* (from 1 to 3).

Figure 1 shows examples from BNC through *Phrases in English* (<http://pie.usna.edu/>) which are significantly attested in IndEng sources.

Then, the forms not attested in the BNC were searched for in the *New Oxford English Dictionary* (OED) in order to get a further matching with a valuable lexicographic source including information about the history of each entry under study. This procedure was necessary to find out possible rare as well as dated occurrences attested in BE as well as in the DiCIE. This method revealed that forms in the DiCIE such as *be fathered upon/on*, *descanted on*, *lavished upon*, *removed to* not attested in the BNC were archaisms attested in the OED. The comparative analysis was therefore carried out to test two main hypotheses:

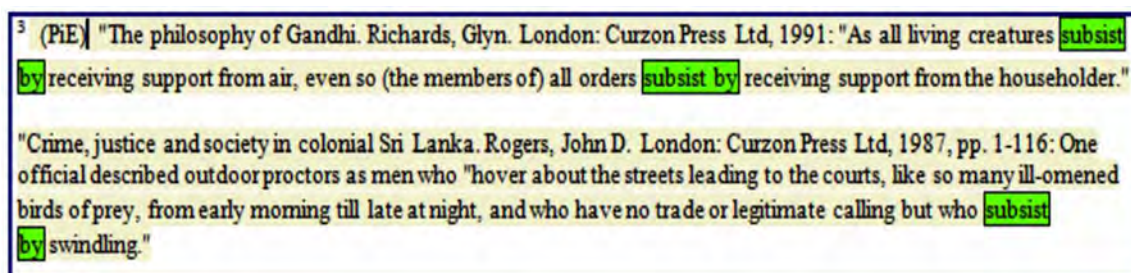
Hypothesis 1: Data dating back to different time periods show a number of shared traits of convergence toward a set of nativized/localized forms or linguistic habits.

Hypothesis 2: Data dating back to different time periods show clear signs of divergence with respect to: (a) the target norms (BE) and (b) the set of localized forms identified in previous literature.

Table 3. List of structural features (in grey) examined in the study

Grammatical Features	Type of VISL Tags Used
Verb Phrase: -Particle verbs -Changes in valency patterns	MV PRP functionally specified as: @ADVL=adverbial; @PIV=prepositional object
Preposition Phrase: -Prepositions	MV </>
Lexico-grammar - Collocations/Phraseology	(as dependency markers)

Figure 1. Examples from BNC through phrases in English



Results

Particle Verbs

In previous descriptions of the IndEng variety, particle verbs have been pointed out as the most typical and innovative area of lexicosyntactic Indianisms (Sedlatschek 2009, p. 149). To test whether any of the early attested combinations (in 1835) had followed a gradual path toward integration and stabilization in IndEng over the years, the comparative approach of using each sub-corpus compiled for each decade side-by-side was employed to describe the development and use of those combinations.

Table 4 gives the occurrences of the individual combinations across decades in the DiCIE as well as in the British National Corpus.

Since the interpretation and selection of prepositions as arguments and/or adjuncts is a semantic class phenomenon rather than an item-specific phenomenon (see ‘Background’), the forms attested in the corpus were then grouped by both semantic domains and following particles (Table 5).

Though the individual counts for each ‘deviant’ occurrence emerging from the analysis of the data are very low to reach statistical significance, their occurrence in the corpus is a ‘linguistic event’ that cannot be disregarded, ‘since the corpus linguist is interested in assessing and interpreting the results anyway’ (Schneider & Zipp 2013).

Discussion

The study of the data attested in the DiCIE has been carried out following two types of analy-

Table 4. Occurrences of verb-particle combinations across time periods

	BNC (2011)	1835	1894	1909- 1931	1940- 1947	1951	1978	2010	N=
<i>accept of</i>	tr.	2					+obj.		3
<i>acquire of</i>	tr.	✓							1
<i>approve of</i>	tr.	2					✓	✓	4
<i>be adapted into</i>	To					✓		✓	2
<i>chase to</i>	after, in	✓	✓	✓					3
<i>contain of</i> ¹¹	tr.	✓	✓			✓	✓		4
<i>convene on</i>	(for/no obj.)				✓				1
<i>excluded over</i>	From	✓							1
<i>be fathered upon</i>	On		✓						1
<i>fitted of</i>	with, in		✓						1
<i>justified in</i>	tr.	✓						✓	2
<i>lack in (of)</i> ¹²	tr..			✓					1
<i>lavished upon</i>				✓					1
<i>quibble over</i>	About				✓				1
<i>reflect to</i>	on, upon							✓	1
<i>removed to</i>	From	✓	✓	✓					3
<i>sprung upon</i>	up, from	✓							1
<i>subsist by</i>	(n=2) ¹³ on/in	✓	✓	✓					3
TOTAL									33

Table 5. Distribution of verb-particle constructions by prepositions and semantic domains

Preposition	Example	Semantic Domain	n=
<i>of</i>	<i>accept of</i>	Mental	1
	<i>acquire of</i>	activity	1
	<i>approve of</i>	Mental	4
	<i>contain(of)¹⁴</i>	Existence	4
	<i>fitted of</i>	Existence	1
<i>in/into</i>	<i>be adapted into</i>	Effort	2
	<i>justified in</i>	mental	2
	<i>lack in</i>	existence	2
	<i>transmitted in</i>	Activity	2
<i>to</i>	<i>chase to</i>	Activity	1
	<i>reflect to</i>	Existence	1
<i>on/upon</i>	<i>convene on</i>	Communicative	1
	<i>be fathered upon</i>	Existence	1
	<i>sprung upon</i>	Activity	1
	<i>lavished upon</i>	Activity	1
<i>over</i>	<i>excluded over</i>	Activity	1
	<i>quibble over</i>	Communicative	1
<i>for</i>	<i>contend for</i>	Communicative	1
<i>by</i>	<i>subsist by</i>	Existence	3

sis. Following a purely data-driven procedure, the annotated corpus was first queried for the structural patterns underlying both phrasal verbs and particle verbs through their morphological features and syntactic dependencies. Then, the patterns of occurrence of the grammatical features identified in the DiCIE were compared to those attested in BE by querying the BNC. The results of previous analyses of the corpus had already shown that particular lexico-grammatical patterns are attested in DiCIE but not in ICE-IND where some other non-standard English patterns can occur (e.g. SINGH-S *India's regard and respect in our ancient culture, My respects to all at home* vs BE *respect for*) (Calabrese 2012). In some other cases (ib.), two competing constructions could co-occur within the same corpus (e.g. ICE-IND *Health permitting I would like to attend to Platinum Jubilee* along with *It was a very good*

news that you attended a meeting of the Board). The current study has shown some changes in the valency patterns of some verbs (e.g. *accept of, acquire of, approve of, lack in*) resulting in the creation of new verb-particle constructions as well. From a diachronic perspective, therefore, a remarkable difference emerges between the data from 1835, 1909, 2010 sources and the BNC as far as the occurrence of verb-particle constructions is concerned. While the texts from the year 2010 contain examples of items such as *quibble over, return +obj.+ to*, no such constructions can be found in the samples from the year 1909.

FUTURE RESEARCH DIRECTIONS

The comparison of data from different time periods suggests that, along with regular standard usage of common particle-verb combinations the creation of new combinatory lexico-grammar patterns is an overall phenomenon occurring regularly and steadily over time, even though regularity in the occurrence of individual items is attested only for a couple of items as shown in Table 4. At this stage, there is no definitively explanatory information from the available materials to establish whether the innovative elements attested in the corpus reflect an individual or an entire community of speakers' linguistic habit. Regular occurrence of innovative features in other sources across time needs to be further investigated to provide evidence of their stabilization in the variety.

CONCLUSION

The analysis of the data has shown rare signs of convergence over time. As matter of fact, some evidence of 'divergence' represented by innovations in lexico-grammatical patterns has emerged probably due to an augmented transparency. Linguists generally tend to distinguish between intralinguistic and extralinguistic factors when discussing principles underlying change and

evolution. Conditions which are purely internal to language depend on speech production, perception, and processing, whereas extralinguistic conditions include factors determined by language use in historical and sociolinguistic settings. Some of them can be explained by speakers' simplification strategies aiming to increase the economy of speech production in language contact situations which leads to the omission of inflectional endings, copulas, articles or the use of redundant form / meaning pairs. Some other forms attested only in 1909 sources far from being considered as individual or community innovations have simply become obsolete (e.g. DiCIE-LET-1909- *I crave your indulgence*) in modern English usage (see OED). Moreover, two of the most effective factors influencing any variety of English are represented by spatial distance and degree of contact with the changes that take place in the language among native speakers which will result in substantial differences across varieties. Although these findings cannot be given a highly statistical backing, they appear to reflect a general attitude on the part of IndEng speakers using such items more frequently in the years 1835, 1894, 1909 and 2010 than in the other years examined.

REFERENCES

- Baines, J. A. (1893). *General Report on the Census of India, 1891*. London: Eyee and Spottiswoode.
- Balasubramanian, C. (2009). *Register Variation in Indian English*. Amsterdam: J. Benjamins Company. doi:10.1075/scl.37
- Bauer, L. (2002). *An Introduction to International Varieties of English*. Edinburgh, UK: Edinburgh University Press.
- Baugh, A. C., & Cable, T. (2002). *A History of the English Language* (5th ed.). London: Routledge.
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511804489
- Brinton, L. J. (1988). *The Development of English Aspectual Systems: Aspectualizers and Post-verbal Particles*. Cambridge, UK: Cambridge University Press.
- Calabrese, R. (2008). *Insights into the Lexicon-Syntax Interface in Italian Learners' English. A generative Framework for a Corpus-based Analysis*. Roma: Aracne.
- Calabrese, R. (2010). The interpretation of Prepositional Phrases as arguments and adjuncts in L2 acquisition. In V. Duško, & C. Krstev (Eds) *Proceedings of the 29th International Conference on Lexis and Grammar*, (pp. 51-61). Belgrade: Faculty of Mathematics, University of Belgrade.
- Calabrese, R., Chambers, J., & Leitner, G. (Eds.). (2015). *Variation and Change in Postcolonial Contexts*. Newcastle-upon-Tyne, UK: Cambridge Scholars Publishing.
- Calabrese, R. (2012). Standardizing variation in a multilingual context: insights into the past and present state of Indian English. In *Regenerating Community, Territory, Voices. Memory and Vision*, (pp. 43-52). Napoli: Liguori Editore.
- Chambers, J. K. (2004). Dynamic typology and vernacular universals. In B. Kortmann (Ed.), *Dialectology meets Typology* (pp. 127-145). Berlin: Mouton de Gruyter.
- Culpeper, J., & Kytö, M. (2010). *Early Modern English Dialogues. Spoken Interaction as Writing*. Cambridge, UK: Cambridge University Press.
- Danchev, A. (1992). The evidence for analytic and synthetic developments in English. In M. Rissanen, O. Ihalainen, & T. Nevalainen (Eds.), *History of Englishes: New Methods and Interpretations in Historical Linguistics* (pp. 25-41). Berlin: Mouton de Gruyter. doi:10.1515/9783110877007.25

- Givón, T. (2009). Multiple routes to clause union. The diachrony of complex verb phrases. In T. Givón & M. Shibatani (Eds.), *Syntactic Complexity. Diachrony, acquisition, neuro-cognition, evolution*. Amsterdam: Benjamin Publishing. doi:10.1075/tsl.85.04mul
- Greenbaum, S. (1996). *Comparing English Worldwide: The International Corpus of English*. Oxford, UK: Clarendon Press.
- Hickey, R. (2010). Contact and Language Shift. In R. Hickey (Ed.), *The Handbook of Language Contact*. Malden, MA: Wiley-Blackwell. doi:10.1002/9781444318159.ch7
- Huebner, T. (Ed.). (1996). *Sociolinguistic Perspectives: Papers on Language in Society, 1959-1994*. Oxford, UK: Oxford University Press.
- Kennedy, A. G. (1920). *The Modern English verb-adverb combination*. Palo Alto, CA: Stanford University Press.
- Law, N. N. (1915). *Promotion of Learning in India by Early European Settlers*. London: Longmans, Green and Co.
- Mahmood, S. (1895). *A History of English Education in India (1781-1893)*. MAO College.
- Mesthrie, R., & Bhatt, R. M. (2008). *World Englishes. The Study of New Linguistic Varieties*. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511791321
- Nihalani, P., Tongue, R. K., & Hosali, P. (2005). *Indian and British English. A Handbook of Usage and Pronunciation* (2nd ed.). Oxford, UK: Oxford University Press.
- Pinker, S. (2003). Language as an adaptation to the cognitive niche. In S. Kirby & M. Christiansen (Eds.), *Language Evolution: States of the Art* (pp. 16–37). New York: Oxford University Press. doi:10.1093/acprof:oso/9780199244843.003.0002
- Sailaja, P. (2009). *Indian English*. Edinburgh, UK: Edinburgh University Press.
- Schiffman, H. F. (2010). Language contact in South Asia. In R. Hickey (Ed.), *The Handbook of Language Contact* (pp. 738–756). Malden, MA: Wiley-Blackwell. doi:10.1002/9781444318159.ch36
- Schneider, E. W. (2004). How to trace structural nativization: Particle verbs in world Englishes. *World Englishes*, 2(2), 227–249. doi:10.1111/j.0883-2919.2004.00348.x
- Schneider, E. W. (2007). *Postcolonial English: Varieties Around the World*. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511618901
- Schneider, E. W. (2011). *English Around the World: An Introduction*. Cambridge, UK: Cambridge University Press.
- Schneider, G., & Zipp, L. (2013). Discovering new verb-preposition combinations in New Englishes. In *Corpus Linguistics and Variation in English: Focus on Non-Native Englishes*. Retrieved from www.helsinki.fi/varieng/journal/volumes/13
- Sedlatleck, A. (2009). *Contemporary Indian English. Variation and Change*. Amsterdam: John Benjamins Company.
- Sorace, A. (2005). Syntactic optionality at interfaces. In *Syntax and Variation: Reconciling the Biological and the Social*. Amsterdam: John Benjamins.
- Thim, S. (2012). *Phrasal Verbs. The English Verb-Particle Construction and its History*. Berlin: de Gruyter Mouton.
- Weinreich, U. (1958). On the compatibility of genetic relationship and convergent development. *Word*, 14(2-3), 374–379. doi:10.1080/00437956.1958.11659677

ADDITIONAL READING

Chambers, J. K., & Schilling, N. (Eds.). (2013). *The Handbook of Variation and Change* (2nd ed.). Oxford: Wiley/Blackwell. doi:10.1002/9781118335598

Deumert, A., & Vandenbussche, W. (2003). Research directions in the study of language standardization. In *Germanic Standardizations. Past to Present*. Deumert, A. & Vandenbussche, W. Amsterdam: Benjamin Publishing. doi:10.1075/impact.18.18deu

Filppula, M., Klemola, J., & Palausto, H. (Eds.). (2009). *Vernacular Universals and Language Contacts. Evidence from Varieties of English and Beyond*. London: Routledge.

Gupta, R. S. (2001). English in post-colonial India: an appraisal. In *Who's Centric Now? The Present State of Post-Colonial Englishes*. Moore, B. Oxford: Oxford University Press.

Subbarao, K. V. (2008). Typological characteristics of South Indian Languages. In B. B. Kachru, S. N. Sridhar, & Y. Kachru (Eds.), *Language in South Asia* (pp. 49–78). Cambridge University Press. doi:10.1017/CBO9780511619069.005

KEY TERMS AND DEFINITIONS

Colonial Lag: The expression used to refer to the consequent conservatism in colonial varieties as a potential factor in distinguishing them from their home counterparts in all levels of language.

Convergence: Here used to characterize phenomena specifically occurring in language contact situations which lead to changes in grammar.

Diachronic: Linguistics investigates the historical development of languages and their changes over time.

Synchronic: Linguistics describes a 'state' of a given language at one point in time.

Sociolinguistic Area: An area where shared patterns of use can be observed especially in contact situations.

Structural Nativization: Structural changes in a new variety combining features from both the substrate and the superstrate languages.

Vernacular Universals: Universally shared features across varieties of English around the world.

ENDNOTES

¹ Early Modern English covers the period when the first British settlers arrived in India; Late Modern English can be referred to the period when early attempts to introduce English in formal education were made in India (Mahmood 1895; Law 1915).

² See for example Bauer (2002, p. 9) who motivates rothicity in North America with the origin of most English-speaking settlers from rothic regions of Great Britain.

³ See Baugh & Cable 2002, 344.

⁴ A comprehensive and updated study of verb-particle constructions has been recently published by Thim (2012) who also provides an extensive review of past literature.

⁵ To effectively illustrate the process of development from the literal meaning to the figurative one, Kennedy (1920, p.15) compares, among the others, the different meanings of compound verbs such as *overlook*, *overtake*, *overturn*, *withdraw* and *withstand* with the corresponding verb-particle combinations *look over*, *take over*, *turn over*, *draw with* and *stand with*. The aspectual meaning of post-verbal particles such as *away*, *down*, *through*, *up* is interpreted as a figurative extension of the concrete meanings of the original adverbs (ibid.).

⁶ Synthetic constructions like pre-verbal incorporated adpositions can be found all over Germanic and Romance languages, in which prepositions have been incorporated

- as verb prefixes (Givón 2009, p. 109). The same process of incorporation can be said to be still going on in English in which prepositions are incorporated post-verbally in conformance with the current V(erb) O(bject) word order (ibid., p. 110).
- ⁷ The DiCIE consists of 833,990 tokens of spoken and written data collected in the year 2010 as part of a four-year project funded by the University of Salerno, Italy.
- ⁸ Letters signed with European names were excluded from the research.
- ⁹ Other variables like the geographical distribution and educational background of the speakers were not considered for the purpose of the present study.
- ¹⁰ The data conversion was carried out by using the Optical Character Recognition (OCR) system which turns scanned texts into digital. doc and.txt texts.
- ¹¹ This entry is also attested in Nihalani (2005).
- ¹² Also attested in Nihalani (2005).
- ¹³ The forms attested in the BNC come from IndEng sources.
- ¹⁴ This entry is also attested in Nihalani (2005).

Negotiating Local Norms in Online Communication



Jonathan R. White
Högskolan Dalarna, Sweden

INTRODUCTION

This entry deals with norms of language use in online communication. Norms are a controversial issue for language learners, in particular when it comes to English usage. There are many varieties of English around the world, but they have different statuses. It is still the case that, for many learners, British and American English have the highest status, and are the only really “proper” varieties to learn; but for others the colonial and cultural hegemony associated with them makes them irrelevant, and even inappropriate, in local contexts.

The Internet has got the reputation of being very open regarding freedom of speech (cf. the Wikileaks scandals), and for being a place where very informal, speech-like language is used. Research has shown, though, that language use online can be a way for social minorities to find a voice, and for those with lower confidence levels to communicate more. Also, it has been demonstrated that non-native speakers contribute more in discussions online, especially with native speaker interlocutors, compared to face-to-face communication. Discourse communities in general have also been argued to develop their own language practices, through the act of using language. The issue under investigation here is what norms do Internet novices adopt in online communication with native speakers of English?

Language production in online seminars by learners of English is analysed to see what norms they develop. The learners are studying on a net-based MA programme in English Linguistics. They were novice Internet users and had never

taken a net-based course before this programme. Mostly they were native speakers of Vietnamese, apart from two speakers of Bangla, both cultures which value respect for teachers. The specific norm analysed are reduced forms, where users reduce either the form or formality of linguistic expressions, for example, writing *info* instead of *information*, or *yeah* instead of *yes*. Given this background, and given that it is an education context, we might predict that their language production will be more oriented towards the norms of their native English-speaking teachers, and be more formal. In fact, this was not the case, and there was evidence that the learners clearly developed their own norms of language use. Thus, this constitutes strong support for the idea that online communities develop their own local norms through using language.

The next section presents the theoretical background to these issues.

BACKGROUND

English in the World

The status of English as a global language is a phenomenon of a modern globalised world. However, the power of English is controversial, a controversy which for many has its roots in colonialism and cultural hegemony. While English is the premier lingua franca for international communication, many see using its norms, and particularly those of British and American English, as betrayals of their cultural identity.

Kachru (1985) described the classic three concentric circles of English in the world, moving from the minority Inner Circle of native speaker norms, to the Outer Circle of the (mostly colonial) countries where English is the institutional language of government and education, etc., to the Expanding Circle of the rest of the world where English is a second or foreign language. Out of this World Englishes movement has come the strong belief that all varieties of English should have equal status, especially as targets for learners of English.

However, Expanding Circle varieties have had an equally controversial status, just as British and American English have. While some varieties have come to be viewed by the local population as acceptable varieties to learn and use, others have been derided as examples of “bad English”. However, even such well-entrenched varieties such as Singapore English are not immune to criticism. Rubdy (2001) noted that the Singapore government had launched a “good English” campaign to clean up the use of Singapore English. In countries like China, there is even stronger criticism of the local variety, demeaningly named “Chinglish”. In their surveys of attitudes of Chinese learners and teachers of English, Jin (2005) and Hu (2005) reported that speakers were very negative towards Chinglish, and for them, American English in particular was preferable. However, the situation among teachers of English is somewhat mixed. Hu (2005) noted that half of all teachers surveyed preferred to teach China English, and two thirds of them thought that it would become the standard for teaching English in China (the same has also been reported by He & Li, 2009 and Xie, 2014). The same debate is ongoing concerning Korean English or Konglish (cf. Rüdiger, 2014, for example).

An alternative view comes from the English as a Lingua Franca (ELF) movement. Given the fact that the Expanding Circle is much bigger than the rest, with roughly two-thirds of English speakers being non-native speakers (cf. the SIL Ethnologue,

www.ethnologue.org), learners should be taught explicitly how to communicate intelligibly with other learners. For instance, Jenkins (2002, 2007) has proposed a core of pronunciation features that are relevant for learners in lingua franca settings. However, this focus on communication involving non-native speakers only, plus the fact that there is still a prescribed set of norms that should be taught, has been criticised as being as extreme a position as one where native speaker varieties are seen as the only correct ones.

Canagarajah (2007: 94) refers to ELF as “a social process constantly reconstructed in sensitivity to environmental factors”; and Park & Wee (2012: 46) state ELF is “always shifting in form and situated within specific contexts”. Regarding native speakers’ place in such International Englishes, Berns (2008: 329) puts it that “native speakers have an important role, not as norm-setters, ..., but as partners with non-native speakers...”. Thus, we see it that all speakers need to adapt in cross-cultural communication settings, and it is in the interactions between speakers of English, both native and non-native, that norms of usage are negotiated. Indeed, we can even argue that the native/non-native dichotomy is obsolete, and that we should simply talk about “users of English”.

This leads us onto norms in online discourse communities.

Norms in Online Discourse Communities

As argued by Pennycook (2010) and Park and Wee (2012) for example, language is a product of social action, and embodies the social practices that brought it about. According to Pennycook’s (2010) ecological perspective, languages adapt to the environment they are used in. Language practices are negotiated and embodied in discourse communities, which are a “locally created social category” (Llamas, 2007: 581); and are “constituted by the language practices they engage in” (Pennycook, 2010: 124).

In the community of practice approach, discourse communities are seen as groups of individuals engaged in a particular task, with mutual engagement in that task, and with discourse practices to discuss that task (Lave & Wenger, 1991; Wenger, 1998). Communities set their own norms and practices internally, although there may be outside influence on them. As Canagarajah (2013: 7) puts it, "...grammatical norms ... evolve from local language practices sedimented over time" but crucially adds that "...they are always open to renegotiation and reconstruction as users engage with new communicative contexts". So, norms in communities are fluid and can be changed. Li (2010) has noted that the Internet is leading to changes in perceptions of norms in English, which is why we will focus on online communication in our analysis.

Communities of language learners are an area of special interest to us, as they very much depend on the norms they are introduced to through pedagogical materials. The choice of what norms to promote and admit in assessment is of vital importance for teachers and learners alike. We mentioned in the introduction that we will concentrate on one particular norm of online communication, reduced forms. White (2011, 2014a, 2014b, 2015b) has argued in previous work that the use of reduced forms is evidence that a community of practice is being built and sustained, as they are examples of discourse practices negotiated by that online community. Thus, our basic question here is what kinds of norms for reduced forms are negotiated in online learner communities, those of native speakers or the learners' own?

Much research in online discourse has identified reduction processes for the purpose of speeding up communication, among other functions (cf. Herring & Zelenkauskaitė, 2009). There are many different classifications of reduction processes, but we concentrate here on some very general categories. The following simplified system for reduced forms will be adopted for this analysis based on the common categories identified in the literature, as argued for in White (2015a: 76):

1. *Clippings (using pl for please, plus acronyms and initialisms).*
2. *Homophone respellings (using 4 for for).*
3. *Phonetic respellings (using yeah for yes, or evry for every).*
4. *Mixed categories of clipping plus homophone or phonetic respelling (using cos for because, or plz for please).*

The focus here is specifically on the reductions in orthography and morphology, although formality is an issue in phonetic respellings like *yeah* as well, as we see that these mark a particular type of discourse community. Other categories such as the orthographic expression of emotion and intonation, etc. are important features of CMC in textchat, but they are not of interest in this work.

Now, we present the data and participants in the study before we move on to the analysis.

NORMS IN ONLINE COMMUNITIES

Data and Participants

The participants in this study were users of English as a second language (30 in total), and were students on a net-based MA programme in English Linguistics run by a university in Sweden. They were not specially selected for a study, although they were told that the work they produced on the programme could be used by teachers for research purposes. Their consent was sought for the material to be used, informing them that they could say no at any time, and only those who gave their consent were included in the study. Two were Bangladeshi students studying in Sweden, and the rest were Vietnamese students based around Vietnam. The Vietnamese students were teachers of English at universities and colleges. All students required an average IELTS score of 7.0 to be admitted onto the programme. Their ages ranged from 23 to 45, and seven were men. An important reason why this material was used for a study was that the students were novice In-

ternet users, and so it is unlikely they will have had access to native speaker norms in English computer-mediated communication before. As mentioned in the introduction, this was also their first experience of net-based education. Thus, the way they developed their discourse community and language practices was of particular interest.

The course where the data was taken from was the first course on their programme, an introduction to core topics in English linguistics, such as phonetics, morphology, and different sociolinguistic topics run in the Autumn semester of 2007. Thus, this was their absolutely first experience of net-based education. The data comes from seminar discussions with three teachers who are users of American, British and Irish English, respectively. The discussion in the seminars took place through Skype textchat. It is important to remember that the classes were not focused on proficiency, which will affect the way language is used. All students have been made anonymous in the presentation of the data, and are referred to as, e.g. Student 15. Typographical errors have been preserved.

The data was analysed linguistically through manual analysis first of all, by identifying reduced forms and counting which students used which forms. These were then checked by using the free concordancer software, AntConc, available from http://www.antlab.sci.waseda.ac.jp/antconc_index.html.

Having presented the data and discussed issues to do with the methodology employed, we present our analysis.

How Were Norms Negotiated?

If we consider first how the students interact and negotiate norms, we see that they generally support one another when a new reduced form is used:

Extract 1

Student 28: *bb*, all you

Student 6: *take care*

Student 4: *Goodbye*

Student 13: *Bye, Teacher 1*

Student 11: *Bye. will we meet tomorrow, summer group?*

Student 6: *bye everyone*

Student 8: *bb*

[Phonology seminar, Spring and Summer groups]

We find Student 28 using the informal *bb* for *goodbye*, which is then picked up by Student 8. There are, though, many examples of students persisting in their own norms:

Extract 2

Student 27: *ex: asked me in full pronunciation is {aSk t mi}, but when we speak fast we only pronounce {as mi}*

Student 23: *yes*

Student 22: *Elision is a process that not pronouncing segment that might be present in the deliberately pronouncing of a word isolation*

Student 22: *Do you think so*

Student 25: *elision is typical of rapid, casual speech, the process of change in phoneme realisations produced by change in speed and casualness of speech*

Student 21: *can you give more examples*

Student 24: *I agree with Student 22*

Student 25: *eg: loss of weak vowels*

[Phonology seminar, Winter group]

Student 27 uses *ex* for *for example*, but Student 25 uses *eg* which is the more standard (Inner Circle) norm.

In these communities, leaders in groups have a high status. The cohorts on the programme were divided into four groups, each with a leader. This leader was responsible for keeping order in discussions, and for arranging and leading group discussions without teachers. We see that leaders particularly strongly affect the others when they use particular reduced forms. Consider the following extract:

Extract 3

Student 21: *I think so*

Student 21: *u r just going too fast*

Student 20: *I agree*

Student 14: *I think so*

Student 22: *the same to me*

Student 18: *u have to describe the position of each ending sounds and u know why*

[Phonology seminar, Autumn and Winter groups]

The leader of Winter group, Student 21, uses *u* as a reduction of *you*, and this is picked up by Student 18. However, it does happen that leaders are not followed, as the following shows:

Extract 4

Student 21: *qu3*, page 69

Student 25: *ing*

Teacher 1: *yes*

Student 25: *Q3*

[Morphology seminar, Winter group]

Here, we see Student 21 again using a reduced form, this time *qu* instead of *question*. In this case, though. Student 25 does not adopt this form, but uses *Q* instead. This non-adaptation, though, does not happen so often, and thus students are most often cooperative in adopting forms. It is also rarer, but we do also see non-leaders affecting other non-leaders:

Extract 5

Student 2: *EX pat -bat*

Student 12: *sorry I've lost which part are you discussing?*

Student 3: *how can we know that they are not overlap*

Teacher 1: *Are we all clear on how to establish if two sounds are allophones of one phoneme?*

Student 2: *I mean minimal pair*

Student 4: *complementary distribution, Student 12*

Student 9: *for ex in the words She and Shoe*

[Phonology seminar, Spring and Summer groups]

Student 2 uses *ex* for *for example*, which is then picked up by Student 9 (this form did not appear earlier in the discussion). Note that a similar form is supported in a separate group in Extract 2 above. Their groups are led by Students 1 and 7.

We have seen here, therefore, that students are generally cooperative when negotiating norms, but that individuals can persist in their own usage. High status individuals in the community in particular have power in this process. This is expected in dynamic communities where norms can be renegotiated, and it is also evidence that we have a community of practice which is in the process of negotiating its linguistic practices.

Next, we consider the role of the teachers in norm-setting.

Teachers and Norms

Given their status in Vietnamese and Bangladeshi cultures, we might expect the norms the English-speaking teachers follow would be adopted by the students. Indeed, we do find examples of this type:

Extract 6

Teacher 1: *Student 4, what exactly do you want ex of?*

Student 9: *very clear*

Student 8: *it's ok but too abstract*

Teacher 1: *Student 5? can you be specific?*

Student 8: *I mean too abstract*

Student 5: *ex of co-articulation effect*

[Phonology seminar, Spring and Summer groups]

This first extract has Teacher 1 using *ex* as a reduction of *example*, and this is adopted by Student 5. Thus, we see that native speakers do have some influence on norms in this community. However, these are rare examples, and mostly students do not adopt teachers' norms at all. The following is a representative example:

Extract 7

Teacher 1: *T.Hanh*, all sounds are allophones of some phoneme

[... three contributions missing...]

Student 18: Can 2 allophones have the same environment?

[... seven contributions missing...]

Teacher 1: Hanh, not if they are allophones of the same phoneme

Student 17: no, I mean

Student 18: Can I give an example of Vietnamese words?

Student 17: example of what?

Student 7: please go on *thuy hanh*

[Phonology seminar, Autumn and Winter groups]

Teacher 1 reduces the name of one student to *T.Hanh*, Student 18. However, Student 7 does not adopt this practice, though, but uses the full name, *Thuy Hanh*.

On the contrary, we find teachers adopting student practice, as in the following:

Extract 8

Student 15: We all have done *Q.3*

Student 17: Teacher 2, may be we should take another textbook.

Student 18: no problem with handout

Teacher 2: yes, i agree. we have one we use with other students that's much better

Student 20: that's good

Student 15: but with Research tasks

Student 14: yes

Student 15: difficult

Student 17: which one?

Teacher 2: so what did you get for *Q3*?

[Morphology seminar, Autumn group]

Teacher 2 uses the reduction *Q* for *question*, following Student 15's use.

Thus, the teachers can be seen to have limited influence on the norms negotiated by the community. Given the background of these students, namely that they were novice Internet users and

were new to net-based education, it is surprising that they still actively negotiated their own norms separately from those practiced by their teachers. We propose that it is an affordance of net-based education that they were able to do this. The community they formed were able to negotiate their own norms due to the distance created by the online environment.

Finally, we turn to how the teachers were addressed.

Addressing Teachers

When it comes to the status of teachers in this community, we see that they still do have high status. As mentioned in the introduction, teachers are held in high respect in Vietnamese and Bangladeshi society, and so should be addressed formally. We find this in the following common greetings:

Extract 9

Student 4: Thanks, *teacher*. and I am still in vague about bound morpheme, are functions only as parts of words are bound morphemes?

[Language and the media seminar, Autumn-Winter group]

Extract 10

Student 4: more examples on co-articulation effect and assimilation, *Prof Una*

[Phonology seminar, Spring and Summer groups]

Extract 11

Student 25: hOW ARE YOU, *Mrs Una*?

[Phonology seminar, Winter group]

We see a variety of formal address forms. There are plenty of reductions, though, especially of this author's name which tended to be reduced to single syllables like *John/Jon* or *Jo*, as we see in the following:

Extract 12

Student 13: *Jo online now.*

[Morphology seminar, Spring group]

As the teachers were told informally by the students, they would not be able to refer to their Vietnamese teachers in this way, and this social pressure is still apparent, in that we see students asking whether it is ok to use a particular reduced form to address a teacher:

Extract 13

Student 17: *Hello, Jon. Can we call you like that?*

[Morphology seminar, Autumn group]

Student 17 asks if it is ok to use *Jon* as an address form. In face-to-face interactions with the same teachers at a university in Hanoi, the students were much more reluctant to use similar shorter forms, and even to use the teachers' given names. Given that very different behaviour is evidenced in online discussions, we see that this is further evidence that it is an affordance of net-based education to create closer, less formal discourse communities, and to create new social environments for interaction. The same sorts of conclusions have been made for computer-mediated communication regarding the amount and quality of communication by non-native speakers in much literature (cf. Sauro, 2012; Jenks, 2014, among others).

Recommendations

We have given evidence that even novices to net-based education create their own norms through using language. These norms are negotiated by the discourse community itself, mostly not adopting the norms that their teachers practiced. This suggests that communities act independently in negotiating their own language practices. This can be argued to promote the conditions for community autonomy and will lead more readily to

learner autonomy. Therefore, we would like to propose that as teachers in academic discussion classes we should be more allowing of these practices. Such discourse environments have specific language practices, and other communities will adopt similar or completely different norms. It is up to the community itself to negotiate and adopt these practices. Through this, they strengthen the community and thereby facilitate the individual autonomy necessary for learning.

FUTURE RESEARCH DIRECTIONS

This research can be taken in a number of directions. Firstly, similar communities can be followed over a longer period of time to see if similar practices are negotiated, especially in an academic environment. Less formal communities may well negotiate in very different, and even more explicit, ways. In terms of pedagogy, an experiment can be set to compare a group that is allowed freedom to set their own language practices with one whose norms are restricted. How would the latter group react to these restrictions? A hypothesis that could be tested here is that the restricted group would prove to be less autonomous than the open group.

CONCLUSION

To conclude, we have seen that communities set their language practices independently of outside control. Groups of novices at net-based education were found to negotiate and adopt their own practices. This was argued to be strong evidence that a community had formed and was able to independently set its own practices. We hope we have convincingly demonstrated that in spite of cultural and linguistic influences, users of English can still practice and create their own language. To us, this is a sign of a strong community which allows its own norms to be negotiated and practiced.

REFERENCES

- Berns, M. (2008). World Englishes, English as a lingua franca, and intelligibility. *World Englishes*, 27(3/4), 327–334. doi:10.1111/j.1467-971X.2008.00571.x
- Canagarajah, S. (2013). *Translingual Practices: Global Englishes and Cosmopolitan Relations*. New York: Routledge.
- He, D., & Li, D. C. S. (2009). Language attitudes and linguistic features in the China English debate. *World Englishes*, 28(1), 70–89. doi:10.1111/j.1467-971X.2008.01570.x
- Herring, S. C., & Zelenkauskaitė, A. (2009). Symbolic capital in a virtual heterosexual market: Abbreviation and insertion in Italian iTV SMS. *Written Communication*, 26(1), 5–31. doi:10.1177/0741088308327911
- Hu, X. (2005). China English, at home and in the world. *English Today*, 21(3), 27–38. doi:10.1017/S026607840500307X
- Jenkins, J. (2002). A sociolinguistically based, empirically researched pronunciation syllabus for English as an International Language. *Applied Linguistics*, 23(1), 83–103. doi:10.1093/applin/23.1.83
- Jenkins, J. (2007). *English as a Lingua Franca: Attitude and Identity*. Oxford, UK: Oxford University Press.
- Jenks, C. J. (2014). *Social Interaction in Second Language Chat Rooms*. Edinburgh, UK: Edinburgh University Press.
- Jin, J. (2005). Which is better in China, a local or a native English-speaking teacher? *English Today*, 21(3), 39–46. doi:10.1017/S0266078405003081
- Kachru, B. B. (1985). Standards, codification and sociolinguistic realism: The English language in the outer circle. In *English in the World: Teaching and Learning the Language and Literatures*. Cambridge, UK: Cambridge University Press.
- Lave, J., & Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511815355
- Li, D. C. S. (2010). When does an unconventional form become an innovation? In *The Routledge Handbook on World Englishes*. Oxford, UK: Routledge.
- Llamas, C. (2007). A place between places: Language and identities in a border town. *Language in Society*, 36(4), 579–604. doi:10.1017/S0047404507070455
- Park, J. S.-Y., & Wee, L. (2012). *Markets of English: Linguistic Capital and Language Policy in a Globalizing World*. New York: Routledge.
- Pennycook, A. (2010). *Language as a Local Practice*. New York: Routledge.
- Rudby, R. (2001). Creative destruction: Singapore's Speak Good English movement. *World Englishes*, 20(3), 341–355. doi:10.1111/1467-971X.00219
- Rüdiger, S. (2014). The nativization of English in the Korean context: Uncharted territory for World Englishes. *English Today*, 30(4), 11–14. doi:10.1017/S0266078414000340
- Sauro, S. (2012). L2 performance in text-chat and spoken discourse. *System*, 40(3), 335–348. doi:10.1016/j.system.2012.08.001
- Wenger, E. (1998). *Communities of Practice: Learning, Meaning and Identity*. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511803932
- White, J. R. (2011). Reduced forms in chat language. In *Language for the Future: Papers from the ASLA Symposium in Falun*. Falun: Högskolan Dalarna.
- White, J. R. (2014a). Standardisation of reduced forms in English in an academic community of practice. *Pragmatics and Society*, 5(1), 105–127. doi:10.1075/ps.5.1.05whi

White, J. R. (2014b). The pragmatics of reduced forms in an Internet community of practice. In *Lodz Studies in Language* 33: Cognitive and Pragmatics Aspects of Speech Actions. Frankfurt am Main: Peter Lang. doi:10.3726/978-3-653-03213-0

White, J. R. (2015a). Processes and variations in language economisation. *Ampersand*, 2, 72–82. doi:10.1016/j.amper.2015.06.001

White, J. R. (2015b). Building and sustaining online communities of practice through language economy. *WorldCALL 2013*.

Xie, J. (2014). Challenges and opportunities for the pluricentric approach in ESL/EFL teaching. *English Today*, 30(2), 43–50. doi:10.1017/S0266078414000121

ADDITIONAL READING

Artes, B., Eslami, Z., & Wright, K. (2015). Incorporating World Englishes into undergraduate ESL education courses. *World Englishes*, 34(3), 485–501. doi:10.1111/weng.12149

Kim, H. Y. (2011). Promoting communities of practice among non-native speakers of English in online discussions. *Computer Assisted Language Learning*, 24(4), 353–370. doi:10.1080/09588221.2011.572899

Mollin, S. (2014). Exploring World Englishes: Language in a global context. *English World-Wide*, 35(3), 353–357. doi:10.1075/eww.35.3.05mol

Schreiber, B. R. (2015). “I am what I am”: Multilingual identity and digital translanguaging. *Language Learning & Technology*, 19(3), 69–87.

Sung, C. C. M. (2013). I would like to sound like Heidi Klum: What do non-native speakers say about who they want to sound like? *English Today*, 29(2), 17–21. doi:10.1017/S0266078413000102

Terhune, N. M. (2013). Learning to learn digitally: Getting students on the road to autonomy. *International Journal of Computer-Assisted Language Learning and Teaching*, 3(4), 9–24. doi:10.4018/ijcallt.2013100102

KEY TERMS AND DEFINITIONS

Discourse Community: A group of language users who communicate on a particular topic.

Inner Circle: The highest status varieties of English consisting of the “colonist” varieties.

Lingua Franca: A language variety used for communication between non-native speakers.

Net-Based Education: Education taking place primarily through online resources.

Norm: A linguistic form agreed upon by a discourse community as appropriate in a particular situation.

Reduced Form: Linguistic expressions reduced in form or formality.

Textchat: Quasi-synchronous use of language in written form.

Category C

Computer Engineering

Architecture of an Open-Source Real-Time Distributed Cyber Physical System



Stefano Scanzio
CNR-IEIT, Italy

INTRODUCTION

Cyber physical systems (CPSs) (Guan et al., 2016) are based on a number of nodes connected through a communication network, which can interact with the environment. Application contexts that foresee the use of CPSs are Industry 4.0, building automation and smart grids (Yu & Xue, 2016). In such systems, nodes are connected through a wired or a wireless network, but commonly a typical configuration relies on the combination of both. In many applications, such as those regarding manufacturing systems, determinism on data exchanges between nodes of the system is the most important requirement.

On the other hand, in recent years, open-source software and applications are approaching CPS market. Open-source provides a low cost and effective method to implement inexpensive, performing and bug free applications. The need of open-source components for many application contexts derives from their capability to be highly customizable. This characteristic impacts directly on performance to perform actuations or sensing physical quantities, in terms of determinism and latency. The use of open-source software for real embedded systems has been described in the book (Cibrario Bertolotti, & Hu, 2015). Other examples of use of open-source software will be provided in the next sections. In the PC world, the Linux operating system is an example of widely adopted open-source platform, which offered long-term support and development. These last features are really important for systems such as CPS, which are updated and replaced infrequently. The Linux

operating system is often chosen by researchers, users and developers. Moreover, the code of the operating system, and of its software suite, is well documented, allowing to expert users a really high degree of control of any behavioral aspect. From the network point of view, Ethernet is the “de facto” standard for wired networks. Regarding wireless, a number of technologies exist, but the most common, fast and easily interoperable with Ethernet is IEEE 802.11, also known as Wi-Fi. For cost reasons and ease of configuration and installation, these technologies are the best candidates for a CPS based on open-source components. It is worth pointing out that, while the majority of protocols for industrial automation rely on Ethernet, possible candidates in industrial CPS for wireless extensions are IEEE 802.11, Bluetooth or wireless sensor networks based on the IEEE 802.15.4 standard (Lu et al., 2016).

In this chapter, a possible implementation of a CPS based on the Linux operating system is presented. The proposed architecture exploits the RTAI or XENOMAI hard real-time schedulers to guarantee the required degree of determinism of nodes. It makes use of synchronization protocols, in both the wired and the wireless extensions, to provide all the nodes a common view of time. Transmission latency can be reduced in the wired network by using hard real-time protocol stacks (such as, e.g., RTnet) and channel access methods as the time division multiple access (TDMA). Regarding Wi-Fi, redundancy techniques based on the transmission of the very same data packet on two non-overlapping networks allow to reduce both the number of packets lost and latency. Fi-

nally, we will show how the proposed architecture simplifies the integration of popular industrial protocols (e.g., EtherCAT or Modbus TCP) within the communication system.

BACKGROUND

CPSs usually interact with the surrounding environment through a wired or a wireless network. Sensing applications used to measure physical values are characterized by few timing requirements. Usually, a network wide notion of the time is maintained through specific synchronization protocols. Common time allows network nodes to associate a time to measured values. It is the case of wireless sensor networks (WSN) (Zaman, Ragab, & Abdullah, 2012) or monitoring systems. A large class of CPS applications is control systems (CSs). In CSs, remote actuators have to be continuously managed by a control unit. In its control-loop, the CS cyclically makes use of data acquired by remote sensors to properly command actuators. The main requirements of CSs are the determinism of the communication network and its reliability. In practice, a network packet has to reach the receiving node within a predefined deadline. In hard real-time (HRT) systems deadlines cannot be exceeded, while in soft real-time (SRT) systems they can be exceeded with low probability, i.e., the deadline miss ration has to be bounded. For wired networks, a number of protocols, technologies and node architectures, typically originated from the industrial world, are able to ensure HRT constraints. Nowadays, only SRT is possible for wireless networks. This is mainly due to the non-exclusive use of the communication medium (i.e., the ether) by nodes involved in the communication. Nevertheless, a number of countermeasures appeared in scientific literature to increase determinism and reliability of wireless protocols, especially for Wi-Fi. Finally, in such type of networks, nodes have to send packets with the correct timing constraints, to ensure the determinism required by the application. To this

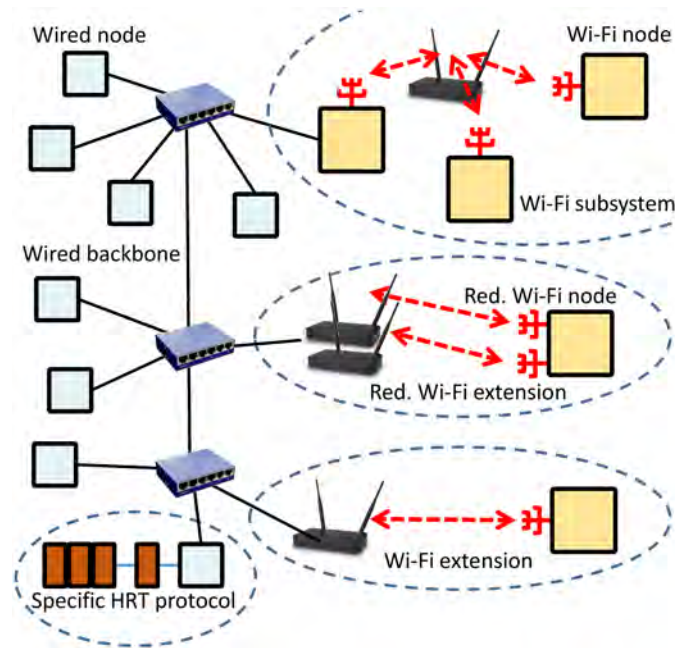
extent, RT properties of nodes have to be improved by using, e.g., a RT operating system. In the following, design guidelines and references will be provided for the implementation of a distributed wired/wireless CPS, with completely open-source components, RT capabilities and conventional hardware.

OPEN-SOURCE ARCHITECTURE FOR A REAL-TIME CPS

Architecture

A reference network architecture of a CPS is schematically presented in Figure 1. A *wired backbone* interconnects *wired nodes*. In the proposed open-source architecture, it consists in standard Ethernet cables and switches. Some *wireless extensions* are connected to the wired backbone. They allow node mobility and cable replacement for such applications in which cabling is hard or cables are prone to wear and tear (e.g., in robotic arms). Wi-Fi has been chosen as the reference wireless technology for wireless extensions because it is faster if compared with other WSNs technologies, it is completely interoperable at the data link layer with Ethernet, and its basic components (i.e., access points (APs) and Wi-Fi adapters) are available off-the-shelf at relatively low cost. APs are used as interconnection elements between the wired backbone and wireless extensions. Two types of Wi-Fi nodes can coexist in the same CPS: conventional *Wi-Fi nodes* and redundant *Red. Wi-Fi nodes* equipped with two or more wireless adapters. The latter type of node is used for SRT applications with demanding latency and reliability requirements. Some nodes can be equipped with both wired and wireless adapters. An example is the border node of a *Wi-Fi subsystem* (e.g., the interface node of an industrial machinery produced by a specific vendor) which is connected to the wired backbone to communicate with other components of the CPS and internally exploits wireless communication for the connec-

Figure 1. Reference architecture of a CPS



tion with moving elements. Finally, some specific network segments may require a high degree of determinism. In this case, *specific HRT network protocols* can ensure the expected performance. Both wired and wireless nodes of the CPS are personal computers (PCs) or embedded systems running the Linux operating system.

Wired Backbone

In the proposed architecture, wired backbone is based on the IEEE 802.3 Ethernet standard. Switches are used to interconnect wired nodes. As a matter of fact, the level of determinism in terms of latency of switched Ethernet is enough for the majority of the CPS application contexts. Nevertheless, switches introduce indeterminism on delivery time, because packets queued inside them experience a delay that depends on network traffic. To avoid this problem, a synchronization protocol can be used to provide a common notion of time to nodes. As a consequence, if network packets containing a command arrive to the receivers sufficiently in advance, synchronized actuations can be performed by different nodes,

even if the underlying communication network is not deterministic enough. Synchronization protocols will be analyzed in detail later. If the network packet containing the actuation command cannot be sent sufficiently in advance, because the actuation command is based on other sensors data that must be adequately new, network determinism has to be improved. A common and effective remedy is traffic prioritization. Most important traffic has to be tagged with the higher priority, as define in IEEE P802.1p standard, to increase Quality of Service (QoS). Such solution is adopted by the EtherNet/IP industrial protocol. Further advantages can be achieved by using specifically modified switches. For example, in TTEthernet high priority traffic is sent on specific time intervals. TTEthernet ensures that in such intervals other traffic cannot interfere with the high priority one. TTEthernet preserves the compatibility with standard Ethernet (i.e., traffic is routed with conventional Ethernet rules outside the reserved intervals). Unfortunately, as other HRT protocols, TTEthernet requires specific and expensive hardware. A software-based, less expensive and less deterministic solution, is TDMA. In TDMA,

synchronized nodes subdivide the communication network in time slots. Each node is then allowed to transmit only in its assigned slots, preventing possible interference with other network nodes. The main disadvantage of this technique is that all nodes have to implement TDMA.

Quite counter-intuitively, in relatively small networks, the major source of indeterminism is not the network, but is the node itself. A simplified schema of a Linux-based PC is reported in Figure 2.

In a PC, concurrently, a number of processes make use of the central processing unit (CPU). Between these processes, the control application (CA) is in charge to manage the transmission and reception of the relevant data of the CPS. The transmission of a packet (e.g., to perform an actuation on a remote node) by means of the CA is usually triggered by a timer or by the reception of another packet, which for instance contains a new sensed datum. In both cases, when the CPU is in use by others process, a context switch operation to substitute the running process with that of the CA is needed. The additional delay to save the context of the old process and to restore the CA context worsen the HRT capability of the node, if compared with the more favorable case

in which the CPU is idle when the CA process is triggered. To send the packet, the CA that runs in user-space makes use of the *system calls* provided by the operating system. System calls software is executed in kernel-space. The transition from user to kernel spaces requires another context switch operation, which leads to further indeterminism. Moreover, the use of shared resources by means of concurrent processes decreases determinism (e.g., the access to the Ethernet adapter or to the main memory is performed through the shared system bus). Additionally, the reaction time of the system to an interrupt depends on pending interrupts not yet managed by the CPU. A first remedy is the installation of specific Linux kernel patches with the aim to reduce the time intervals in which the operating system cannot be preempted by other ready processes, such as the CA. The most known is RT patch. Among its most important features, it minimizes the time in which the CPU is not sensible to interrupts and it allows the Linux kernel to assign a priority to each interrupt (i.e., a major priority can be assigned to Ethernet adapter and timing interrupts). Although RT patch allows to increase determinism, the resulting system can only be considered SRT. A more effective countermeasure relies on the use of HRT extensions, such as RTAI or XENOMAI. Normal processes and the operating system itself are scheduled by these extensions as non-real-time processes. HRT processes are not scheduled by the Linux kernel but by RTAI or XENOMAI. In this way, they do not suffer of the interference caused by other non-real-time processes. Unfortunately, to guarantee HRT performance, CA software cannot make use of system calls, and device drivers have to be programmed exploiting specific RTAI or XENOMAI libraries.

The most recent article of performance comparison between standard Linux kernel, RT patch and RTAI is (Cereia, Cibrario Bertolotti, & Scanzio, 2011). Some results of that work are reported in Table 1. It analyzes the jitter (i.e., the deviation from the desired actuation time) in a system without additional load and with an inten-

Figure 2. Simplified schema of a PC wired node

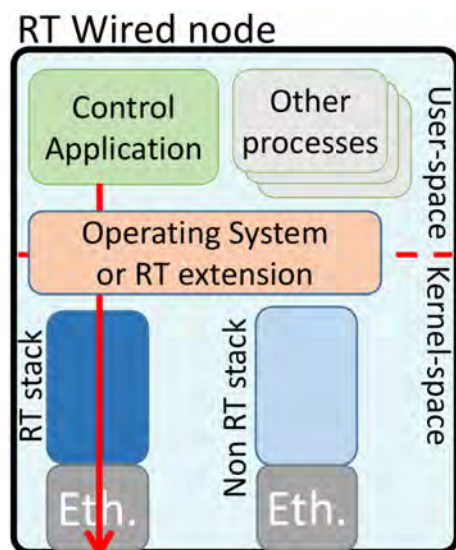


Table 1. Comparison between standard Linux kernel, RT patch and RTAI

	Jitter (μ s)			
	No Load		I/O Load	
	σ	Max	σ	Max
Std. Linux kernel	5.72	1434.0	5.75	1151.0
RT patch	0.12	1.8	0.63	10.4
RTAI	0.12	2.5	0.32	8.0

Data source: (Cereia, Cibrario Bertolotti, & Scanzio, 2011).

sive I/O load of 500 IRQs per second. In the latter configuration, results confirm the impossibility to use a standard Linux kernel in RT applications (maximum jitter is 1151 μ s), and the effectiveness of RT patch and RTAI (maximum jitter is 10.4 and 8.0 μ s, respectively).

When packets have to be sent timely, also driver and protocol stack implementations must have RT properties. A commonly used HRT protocol stack for wired adapters is RTnet, which ensures a bounded delay on packets delivery. Jitters of standard Linux and RTnet were analyzed in (Cena, Scanzio, Valenzano, & Zunino, 2011). As reported in Table 2, the maximum jitter of the standard driver in the I/O load condition is for many application contexts too high (1910.6 μ s). On the other side, with RTnet, jitter remains bounded to 38 μ s.

Wi-Fi Extensions

Wi-Fi extensions are really useful in many application contexts. Unfortunately, they cannot guarantee the determinism of cables. In particular, the number and type of wireless nodes in proximity of the CPS is usually out of the control of the application designer. Choosing the band and the transmission channel of the Wi-Fi system in a frequency not exploited by other wireless networks is the first best practice. Sometimes this solution is not possible, because in a specific environment users have not control over the surrounding networks. With the widespread diffusion of mobile devices, the elimination of these interfering nodes is becoming

Table 2. Comparison between standard and RTnet protocol stack

	Jitter (μ s)			
	No Load		I/O Load	
	σ	Max	σ	Max
Standard	11.2	72.3	35.2	1910.6
RTnet	6.0	38.0	4.0	33.6

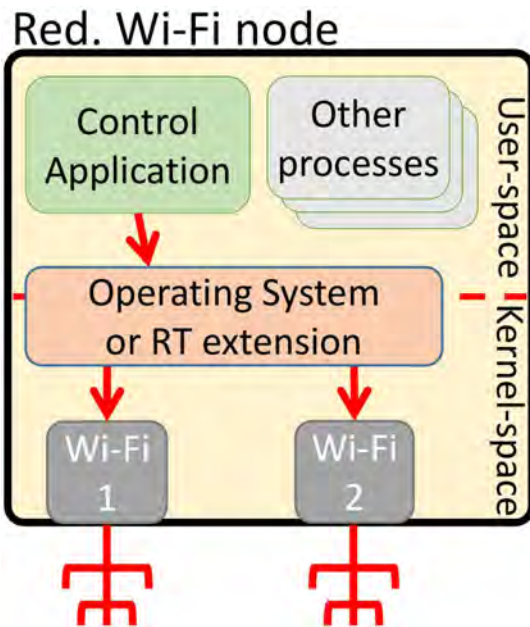
Data source: (Cena, Scanzio, Valenzano, & Zunino, 2011).

even more complicated. Moreover, electromagnetic disturbances are another source of packets loss. TDMA techniques have been extensively exploited in Wi-Fi to prevent the indeterminism of the CSMA/CA method used by nodes to manage channel access (Sevani, Raman, & Joshi, 2014). As for the wired case, synchronized nodes are allowed to transmit only in their assigned slots. Unfortunately, even with TDMA, sensitivity to channel disturbances is equal to normal Wi-Fi. In addition, to obtain a sufficient degree of determinism, these methods require that all the wireless nodes follow TDMA rules. Even if TDMA and the other listed techniques can noticeably improve channel quality, they cannot deal with protracted disturbances or heavy interferences.

An effective solution is seamless redundancy, which is based on frequency diversity. The architecture of a *Red. Wi-Fi node* implementing seamless redundancy and based on a PC is reported in Figure 3. The most significant difference with a typical Wi-Fi node is the presence of two (or more) Wi-Fi adapters.

With seamless redundancy, two copies of the same message are transmitted on distinct wireless channels. The receiver node takes care of discarding duplicates, and it forwards to the application layer only the first message copy arrived to destination. The probability that a disturbance or an interference affects all the channels involved in the parallel transmission is relatively small. Seamless redundancy for Wi-Fi was firstly proposed by (Rentschler & Laukemann, 2012) and it was based on the Parallel Redundancy Protocol (PRP), which is a protocol aimed at ensuring end-to-end

Figure 3. Simplified schema of a PC-based Red Wi-Fi node



redundancy for wired Ethernet networks. Its first adaptation to Wi-Fi relied on specific devices named RedBoxes that implement the standard PRP protocol over Ethernet. Wi-Fi equipment was connected to RedBoxes to analyze PRP performance in a wireless network. Seamless redundancy based on PRP is an end-to-end solution, i.e., between sender and receiver nodes. In Wi-Red technique (Cena, Scanzio, Valenzano, & Zunino, 2014), seamless redundancy was applied at the link level. For instance, in the communication path between two Wi-Fi nodes, the first link is between the node and the AP, while the second

is between the AP and the second node. With Wi-Red, further improvements are possible: by implementing specific techniques to avoid the sending of packets already arrived through other channels, or by exploiting heuristics for scheduling the traffic on the two interfaces.

Results published in scientific literature (Cena, Scanzio, & Valenzano, 2016) confirm that significant improvements can be achieved in real installations in terms of packet losses and transmission latencies. Table 3 compares performances of two single channels with those obtained with seamless redundancy. Results, obtained sending periodically (every 100 ms) unicast and multicast traffic, are in terms of average latency, maximum latency and percent of packet lost much better than those of the single channel. Even if, at present, a HRT driver implementation similar to RTnet for Wi-Fi adapters is not available, the use of HRT extensions such as RTAI or XENOMAI is recommend to improve determinism as much as possible.

Clock Synchronization

Synchronization makes possible the implementation of TDMA and offers some essential services at the application level. To permit the comparison between values sensed by different nodes, sensed data must be associated with a temporal information. In addition, synchronization enables different nodes to perform simultaneous actuations, addressing the indeterminism of the communication network by sending packets in advance. Synchronization protocols are based on

Table 3. Comparison between Wi-Fi transmissions on individual channels and on redundant channels

	Unicast			Multicast		
	μ	Max	Perc. Lost	μ	Max	Perc. Lost
	(ms)	(ms)	(%)	(ms)	(ms)	(%)
Channel A	0.92	100.77	0.0	1.57	106.30	0.51
Channel B	0.95	100.92	0.0	2.16	92.71	1.11
Channel A+B	0.71	8.93	0.0	1.12	44.35	0.02

Data source: (Cena, Scanzio, & Valenzano, 2016).

the exchange of specific sequences of packets, and on timestamps obtained on sending and/or reception of a subset of these packets. Timestamps are used by synchronization protocols for the regulation of nodes time by means of specific clock regulation algorithms (Mongelli & Scanzio, 2016). Their precision is directly related to timestamps accuracy. If hardware timestamping is not supported by the adapter, timestamps are typically acquired at the first instruction of the Interrupt Service Routine (ISR), executed to serve the interrupt raised by the network adapter, which follows the reception of a packet. Synchronization quality in the case of software timestamps is in the order of tens microseconds, while for hardware timestamps a typical accuracy is in the order of some nanoseconds. For wired networks, the “de facto” standard is the master-slave IEEE 1588 protocol (Cena, Cibrario Bertolotti, Scanzio, Valenzano, & Zunino, 2013). The IEEE 1588 protocol is based on a tree synchronization hierarchy, where a number of slave nodes synchronize with a master node. A slave node can act as a master for its sub-tree. For Wi-Fi, specific synchronization protocols, namely the *timing measurement* and *timing advertisement*, have been included in the 2012 version of IEEE 802.11 standard, but at the time of writing they are not implemented in most of the available off-the-shelf Wi-Fi adapters, and they have some disadvantages (Cena, Scanzio, Valenzano, & Zunino, 2015). The porting of IEEE 1588 to Wi-Fi has been analyzed in (Mahmood, Exel, & Sauter, 2014). The AP acts as a slave for the wired backbone and as a master for the Wi-Fi extension. This solution is really effective, and, above all, it makes possible the sharing of the same synchronization protocol to all the nodes of the CPS. Unfortunately, implementation requires modifications to the AP. The Reference Broadcast Infrastructure Synchronization (RBIS) protocol (Cena et al., 2015) does not require any modification to Wi-Fi equipment. As a consequence, it can be directly implemented in software on conventional hardware. RBIS requires a Wi-Fi node as

a master, i.e., the master is not the AP. This last requirement limits the possible network topologies. In fact, the master node, to synchronize with the wired backbone, has two network adapters: a wired interface connected to the backbone and a Wi-Fi interface connected to the AP. On the other side, RBIS enables the connection of Wi-Fi subsystems to the wired backbone, as represented in Figure 1. Synchronization error achieved by RBIS on a real implementation based on a conventional PC is less than 3.3 μ s (Cena et al., 2015).

Other RT Protocols

Usually, actuations and sensing have to be performed within a given deadline. To this extent, some network segments can rely on specific and more deterministic protocols, especially if the CPS is used for CSs with safety requirements. A number of HRT protocols exist, each with its own peculiarities. EtherCAT (Scanzio, 2012) is an example of really high deterministic protocol. In EtherCAT, master node can be implemented on a conventional PC. This feature and the availability of open-source software make it particularly suitable to be used in Linux-based CPS. Slave nodes, those performing sensing and actuations, are implemented in hardware to increase determinism. EtherCAT has its own synchronization protocol that guarantees a synchronization quality smaller than 50 ns (Cena, Cibrario Bertolotti, Scanzio, Valenzano, & Zunino, 2012), and which can be easily integrated with IEEE 1588 through specific EtherCAT slave devices. The most powerful open-source EtherCAT master software for Linux is EtherLab. A description of EtherLab and some guidelines for its use and configuration can be found in (Scanzio, 2012), while its performance was evaluated in (Cereia et al., 2011). An EtherCAT master implemented on a Linux-based conventional PC requires a very high determinism on sending Ethernet frames. For this reason, the use of HRT extensions as RTAI or XENOMAI is highly recommended.

SOLUTIONS AND RECOMMENDATIONS

A completely open-source architecture for CPS based on Linux, off-the-shelf network devices and standard PCs has been proposed. Determinism of nodes is achieved with HRT extensions (i.e., RTAI or XENOMAI) or with a SRT patch named RT patch. Moreover, specific HRT protocol stacks, such those provided by RTnet, can be used for wired connections. Depending on the application requirements, the proposed architecture has different levels of capability and scalability. For wired networks, it may rely on conventional Ethernet, but TDMA techniques can be used to increase determinism. Specific HRT protocols such EtherCAT can be exploited to further increase determinism. Regarding wireless extensions, if conventional Wi-Fi is not enough, TDMA is a first improvement, which can be further enhanced with techniques based on seamless redundancy. However, to the current state-of-the-art, Wi-Fi can be considered only for SRT systems, in which deadlines can be exceeded with low probability (Cena, Scanzio, & Valenzano, 2016). Finally, a synchronization service is transparently available in all nodes of the CPS. It is based on the IEEE 1588 standard for wired segments, and on recent researches for Wi-Fi extensions.

FUTURE RESEARCH DIRECTIONS

For many years, a constant research activity was targeted to improve determinism of wired networks. Its outcomes were a suite of HRT protocols that rely on Ethernet, and deterministic software and operating systems for nodes. Many of these software (and hardware) are commercial products, some are open-source, which are the most used in research. Nowadays, research activity on CPS wired networks is aimed to increase performance and to seamlessly integrate all network components (Bangemann, Riedl, Thron, & Diedrich, 2016). Regarding determinism, HRT extensions installed

on PCs have counter-intuitively worst performance than in the past, because today's CPUs are multi-core and their target is to maximize throughput instead of determinism. The evolution of the hardware requires the implementation of new algorithms and software. The use of big-data analysis and machine learning techniques is a way to improve the quality of some services, such as for synchronization protocols. Nodes virtualization allows to decrease configuration and management effort in large CPS, and it is a very promising research area, which opens many interesting investigation aspects regarding especially determinism. An even more promising research area concerns wireless. CPS has a really high demand of deterministic communication over the ether. Seamless redundancy is a new remedy, but not a definitive solution. Very large improvements can be achieved on existing techniques, and the opportunity to propose new effective methods is very high. A really heterogeneous set of wireless technologies is currently available in the market, with very orthogonal characteristics (e.g., IEEE 802.11, Bluetooth, IEEE 802.15.4, WirelessHART, LoRa, etc.). For each of these technologies, techniques to increase determinism and reliability have to be specifically implemented to obtain the maximum improvement (Lu et al., 2016). Finally, other services such localization need to be improved and included in wireless networks of the future, in order to better support node mobility.

CONCLUSION

The proposed architecture for the implementation of a CPS is based on open-source software, on standard hardware and on the Linux operating system. Other architectural solutions are possible, especially if commercial products are considered. The latter are typically easier to install and configure, but their integration in the overall architecture is more complex. The proposed open-source solution is very scalable and customizable,

especially because software components are easily modifiable and highly configurable (i.e., source code is open). The reference wireless technology for wireless extensions discussed in this chapter is Wi-Fi. It offers a number of advantages in terms of both integration with Ethernet networks and throughput. Nowadays, general purpose systems (e.g., PC equipped with Linux) and off-the-shelf devices (e.g., Ethernet and Wi-Fi adapters, or APs and switches) can be the basic components of high performing CPS. This is a great achievement. Moreover, future research activities are very promising and aimed to replace cable with wireless technology for an increasing number of applications.

REFERENCES

- Bangemann, T., Riedl, M., Thron, M., & Diedrich, C. (2016). Integration of Classical Components Into Industrial Cyber-Physical Systems. *Proceedings of the IEEE*, 104(5), 947–959. doi:10.1109/JPROC.2015.2510981
- Cena, G., Cibrario Bertolotti, I., Scanzio, S., Valenzano, A., & Zunino, C. (2012). Evaluation of EtherCAT Distributed Clock Performance. *IEEE Transactions on Industrial Informatics*, 8(1), 20–29. doi:10.1109/TII.2011.2172434
- Cena, G., Cibrario Bertolotti, I., Scanzio, S., Valenzano, A., & Zunino, C. (2013). Synchronize your watches: Part I: General-purpose solutions for distributed real-time control. *IEEE Industrial Electronics Magazine*, 7(1), 18–29. doi:10.1109/MIE.2012.2232354
- Cena, G., Scanzio, S., & Valenzano, A. (2016). Seamless Link-Level Redundancy to Improve Reliability of Industrial Wi-Fi Networks. *IEEE Transactions on Industrial Informatics*, 12(2), 608–620. doi:10.1109/TII.2016.2522768
- Cena, G., Scanzio, S., Valenzano, A., & Zunino, C. (2011). Evaluation of the real-time properties of open-source protocol stacks. In *IEEE 16th Conference on Emerging Technologies & Factory Automation (ETFA)* (pp. 1–8). doi:10.1109/ETFA.2011.6059003
- Cena, G., Scanzio, S., Valenzano, A., & Zunino, C. (2015). Implementation and Evaluation of the Reference Broadcast Infrastructure Synchronization Protocol. *IEEE Transactions on Industrial Informatics*, 11(3), 801–811. doi:10.1109/TII.2015.2396003
- Cereia, M., Cibrario Bertolotti, I., & Scanzio, S. (2011). Performance of a Real-Time EtherCAT Master Under Linux. *IEEE Transactions on Industrial Informatics*, 7(4), 679–687. doi:10.1109/TII.2011.2166777
- Cibrario Bertolotti, I., & Hu, T. (Eds.). (2015). *Embedded Software Development: The Open-Source Approach*. Boca Raton, FL: CRC Press.
- Guan, X., Yang, B., Chen, C., Dai, W., & Wang, Y. (2016). A comprehensive overview of cyber-physical systems: from perspective of feedback system. *IEEE/CAA Journal of Automatica Sinica*, 3(1), 1–14.
- Lu, C., Saifullah, A., Li, B., Sha, M., Gonzalez, H., Gunatilaka, D., & Chen, Y. et al. (2016). Real-Time Wireless Sensor-Actuator Networks for Industrial Cyber-Physical Systems. *Proceedings of the IEEE*, 104(5), 1013–1024. doi:10.1109/JPROC.2015.2497161
- Mahmood, A., Exel, R., & Sauter, T. (2014). Delay and Jitter Characterization for Software-Based Clock Synchronization Over WLAN Using PTP. *IEEE Transactions on Industrial Informatics*, 10(2), 1198–1206. doi:10.1109/TII.2014.2304413
- Mongelli, M., & Scanzio, S. (2016). A neural approach to synchronization in wireless networks with heterogeneous sources of noise. *Ad Hoc Networks*, 49, 1–16. doi:10.1016/j.adhoc.2016.06.002

- Rentschler, M., & Laukemann, P. (2012). Towards a reliable parallel redundant WLAN black channel. In *9th IEEE International Workshop on Factory Communication Systems (WFCS)* (pp. 255-264). Lemgo, Germany. doi:10.1109/WFCS.2012.6242573
- Scanzio, S. (2012). SoftPLC-Based Control: A Comparison between Commercial and Open-Source EtherCAT Technologies. In M. Khan & A. Ansari (Eds.), *Handbook of Research on Industrial Informatics and Manufacturing Intelligence: Innovations and Solutions* (pp. 440-463). Hershey, PA: Information Science Reference. doi:10.4018/978-1-4666-0294-6.ch019
- Sevani, V., Raman, B., & Joshi, P. (2014). Implementation-Based Evaluation of a Full-Fledged Multihop TDMA-MAC for WiFi Mesh Networks. *IEEE Transactions on Mobile Computing*, 13(2), 392-406. doi:10.1109/TMC.2012.251
- Yu, X., & Xue, Y. (2016). Smart Grids: A Cyber-Physical Systems Perspective. *Proceedings of the IEEE*, 104(5), 1058-1070. doi:10.1109/JPROC.2015.2503119
- Zaman, N., Ragab, K., & Abdullah, A. (2012). *Wireless Sensor Networks and Energy Efficiency: Protocols, Routing and Management*. Hershey, PA: IGI Global. doi:10.4018/978-1-4666-0101-7
- Cena, G., Scanzio, S., Vitturi, S., & Zunino, C. (2011). Industrial Ethernet and TCP/IP-Based Systems. In B. G. Liptak & H. Eren (Eds.), *Instrument Engineers' Handbook, Volume 3: Process Software and Digital Networks, Fourth Edition* (pp. 636-653). Boca Raton, FL: CRC Press.
- Cucinotta, T., Mancina, A., Anastasi, G. F., Lipari, G., Mangeruca, L., Checcozzo, R., & Rusina, F. (2009). A Real-Time Service-Oriented Architecture for Industrial Automation. *IEEE Transactions on Industrial Informatics*, 5(3), 267-277. doi:10.1109/TII.2009.2027013
- Djukic, P., & Mohapatra, P. (2012). Soft-TDMAC: A Software-Based 802.11 Overlay TDMA MAC with Microsecond Synchronization. *IEEE Transactions on Mobile Computing*, 11(3), 478-491. doi:10.1109/TMC.2011.56
- Mongelli, M., & Scanzio, S. (2014). Approximating Optimal Estimation of Time Offset Synchronization With Temperature Variations. *IEEE Transactions on Instrumentation and Measurement*, 63(12), 2872-2881. doi:10.1109/TIM.2014.2320400
- Perahia, E., & Stacey, R. (Eds.). (2013). *Next Generation Wireless LANS: 802.11n and 802.11ac* (2nd ed.). New York, NY: Cambridge University Press. doi:10.1017/CBO9781139061407
- Wu, Y. C., Chaudhari, Q., & Serpedin, E. (2011). Clock Synchronization of Wireless Sensor Networks. *IEEE Signal Processing Magazine*, 28(1), 124-138. doi:10.1109/MSP.2010.938757

ADDITIONAL READING

- Cena, G., Cibrario Bertolotti, I., Scanzio, S., Valenzano, A., & Zunino, C. (2013). Synchronize Your Watches: Part II: Special-Purpose Solutions for Distributed Real-Time Control. *IEEE Industrial Electronics Magazine*, 7(2), 27-39. doi:10.1109/MIE.2013.2248431
- Cena, G., Scanzio, S., Valenzano, A., & Zunino, C. (2014). Ethernet for Control Automation Technology. In R. Zurawski (Ed.), *Industrial Communication Technology Handbook* (2nd ed., pp. 18-1-18-27). Boca Raton, FL: CRC Press.

KEY TERMS AND DEFINITIONS

Cyber Physical System: A system that integrates computer and networking to control a physical process.

IEEE 802.11 (Wi-Fi): A standard for the implementation of high-throughput wireless local area networks.

IEEE 802.3 (Ethernet): A collection of standards for the definition of Ethernet. Ethernet is the most known wired network technology.

Real-Time Networks: Communication networks with demanding requirements in terms of latency and determinism.

Seamless Redundancy: Inclusion of duplicate components, not strictly necessary, to improve reliability.

Synchronization Protocols: Network protocols for the distribution of a common time reference to the nodes of a communication network.

Time Division Multiple Access (TDMA): A channel access method to manage a mutual exclusive access of transmitting nodes to a network.

Wi-Red: A seamless redundancy technique applied to Wi-Fi to improve determinism and reliability of IEEE 802.11 wireless networks.

C

Consistency Is Not Enough in Byzantine Fault Tolerance

Wenbing Zhao

Cleveland State University, USA

INTRODUCTION

In Byzantine fault tolerance (BFT), a core concern is to ensure the consistency of replicas despite malicious attacks from the clients and compromised replicas (Zhao, 2014). This is accomplished by totally ordering incoming requests and by rendering the replica's operations deterministic (Zhang et al., 2011). In the presence of application non-determinism, such as the access of local clocks, replicas are rendered deterministic by forcing all non-faulty replicas to use the same values either supplied by the primary or computed deterministically. While this approach works well for some applications, such as a replicated file system, doing so could lead to the compromise of the service integrity for applications that rely on the use of random numbers.

For example, consider an Internet application that relies on the use of session-ids for stateful interactions between the server and its clients. As pointed out in (Dorrendorf, Gutterman, & Pinkas, 2007), if the session-id of an active session can be predicted, the client's session with the server could be hijacked, which could lead to the leak of confidential information regarding the client, such as name, address, and the order history, or unauthorized orders (if the one-click option for placing orders is enabled). The session-id may be predicted by searching the limited entropy space if weak random bits are used in an application. For example, the authors of (Dorrendorf, Gutterman, & Pinkas, 2007) reverse-engineered a version of Tomcat (a popular Java Servlet Engine) and the related operations in a Window's based Java Virtual Machine. They could attack the system

by performing about 251 searches in finding an active session-id.

Therefore, it is critical not to weaken the strength of the random bits essential for the integrity of their operations when replicating these systems for Byzantine fault tolerance. For a sound coordination algorithm, it is essential to enable each replica to access its own entropy source and maintain its independence in such operations. However, this desire is in conflict with the basic requirement for state machine replication (Schneider, 1990), which mandates that the replicas must be deterministic or rendered deterministic to maintain strong replica consistency. The conflicting requirements for security and replication must be reconciled to avoid the dilemma of either favoring security over high availability by not performing state machine replication of the systems, or trading security for high availability by removing the randomness of the systems in order to perform state machine replication.

In this chapter, we present a novel replica coordination algorithm, referred to as the Collective-Determination BFT algorithm, or CD-BFT algorithm in short, towards the reconciliation of the conflicting requirements for security and for strongly consistent replication. The central idea behind this algorithm is that all random numbers to be used by the replicas are collectively determined, and furthermore, the determination is based on the contributions made by a quorum of replicas, at least one of which is correct.

In the CD-BFT algorithm, the replicas first reach a Byzantine agreement on the set of contributions from replicas, and then apply a deterministic algorithm, such as the bitwise exclusive-or

operation (Young & Yung, 2004), to compute the final random value. The freshness of the random numbers generated is application dependent. Our approach does not alter the freshness of the random numbers. If a pseudo-random number generator is used, it should be periodically reseeded from a good entropy source.

BACKGROUND

An arbitrary fault is often referred to as a Byzantine fault. The term was introduced in (Lamport, Shostak, & Pease, 1982) to highlight a specific faulty behavior that a Byzantine faulty process may disseminate conflicting information to other processes. For example, a compromised process might exhibit such Byzantine faulty behavior. Byzantine fault tolerance refers to the capability of tolerating Byzantine faults in a system. It can be achieved by introducing sufficient redundancy into the system and by using a sophisticated replica coordination algorithm that can cope with Byzantine faulty replicas and clients. A basic requirement for such an algorithm is to ensure that all server replicas agree on the total ordering of the requests received despite the existence of Byzantine faulty replicas and clients. Such an agreement is often referred to as Byzantine agreement (Lamport, Shostak, & Pease, 1982).

Recently, a number of efficient BFT algorithms (Castro & Liskov, 2002; Kotla et al., 2007; Yin et al., 2003) have been proposed. Our CD-BFT algorithm is derived from the PBFT algorithm and we use the same system model as that in (Castro & Liskov, 2002). The PBFT algorithm operates in an asynchronous distributed environment. The safety property of the algorithm, *i.e.*, all correct replicas agree on the total ordering of requests, is ensured without any assumption of synchrony. However, for the algorithm to make progress towards a Byzantine agreement (*i.e.*, liveness), certain synchrony is needed, for example, a reasonable assumption is that the message transmission and processing delay has an asymptotic upper bound. This bound

is dynamically explored in the algorithm in that each time a view change occurs, the timeout for the new view is doubled.

How to ensure strong replica consistency in the presence of replica non-determinism has been of research interest for a long time (Castro & Liskov, 2002; Castro, Rodrigues, & Liskov, 2003; Powell, 1991; Slember & Narasimhan, 2006), most of which are for fault tolerant systems using the benign fault model. However, while the importance of the use of good random numbers has long been recognized in building secure systems (Viega & McGraw, 2002), we have yet to see substantial research work on how to preserve the randomized operations necessary to ensure the system integrity in a fault tolerant system. For the type of systems where the use of random numbers is crucial to their service integrity, the benign fault model is obviously inadequate and the Byzantine fault model must be employed if fault tolerance is required.

Some form of replica non-determinism, in particular those related to timestamp operations, has been studied in the context Byzantine fault tolerant systems (Castro & Liskov, 2002; Castro, Rodrigues, & Liskov, 2003). However, as we will argue in the next section, the existing approach is vulnerable to the presence of colluding Byzantine faulty replicas and clients. We also studied the replica non-determinism issue with the emphasis of classification of non-determinism types and the systematic handling of various types of replica non-determinism (Zhang et al., 2011).

RECONCILIATION OF SECURITY AND REPLICATION REQUIREMENTS

Pitfalls in Controlling Replica Randomness

In this section, we analyze a few well-known approaches that one may attempt to use to ensure replica consistency in the presence of replica randomness. We show that they are not robust

against Byzantine faulty replicas and clients. We reiterate here the importance of using random bits for security-related operations. Any attempt to weaken or remove such randomness from the system for the purpose of replication may compromise the system integrity. Furthermore, for Byzantine fault tolerance, it is essential *not* to place undue trust to any single replica, such as the primary replica, because it is virtually impossible to verify in real-time if the random number proposed by a replica is truly random, and consequently, if that particular replica is compromised, the integrity of the entire replicated system will be lost.

For systems that use a pseudo-random number generator, the replica consistency can be easily achieved by using the same seed value to initialize the generator. One might attempt to use the sequence number assigned to the request as the seed. Even though this approach is perhaps the simplest way to render replicas deterministic (since no extra communication step is needed and no extra information is to be included in the control messages for total ordering of requests), it takes the randomness away from the system. Consequently, a Byzantine faulty client can easily guess the seed and predict the random numbers. A seemingly more robust approach is to use the timestamp as the seed to the pseudo-random number generator. As shown in (Viega & McGraw, 2002; Young & Yung, 2004), the use of timestamp does not offer more robustness to the system because it a Byzantine faulty client can easily guess the seed.

The only option remaining seems to be the use of a truly random number to seed a strong pseudo-random number generator (or to obtain random numbers entirely from a high entropy source). We note that the elegant mechanism described in (Castro & Liskov, 2002) cannot be used in this case because backups have no means to verify whether the number proposed by the primary is taken from a strong random number generator seeded periodically using a high-entropy source, or is generated according to a deterministic algorithm. If the latter is the case, the Byzantine faulty primary could continue colluding with Byzantine faulty clients

without being detected. A slight improvement over this scheme is to proactively rotate the primary. However, this improved scheme does not solve the problem: while the faulty replica serves as the primary, it can use a predictable seed and/or a weak random number generator without being detected, and hence, the system integrity cannot be guaranteed during this period.

Therefore, the most effective way in countering such threats seems to collectively determine the random numbers, based on the contributions from a quorum of replicas so that Byzantine faulty replicas cannot influence the final outcome.

The Collective-Determination BFT Algorithm

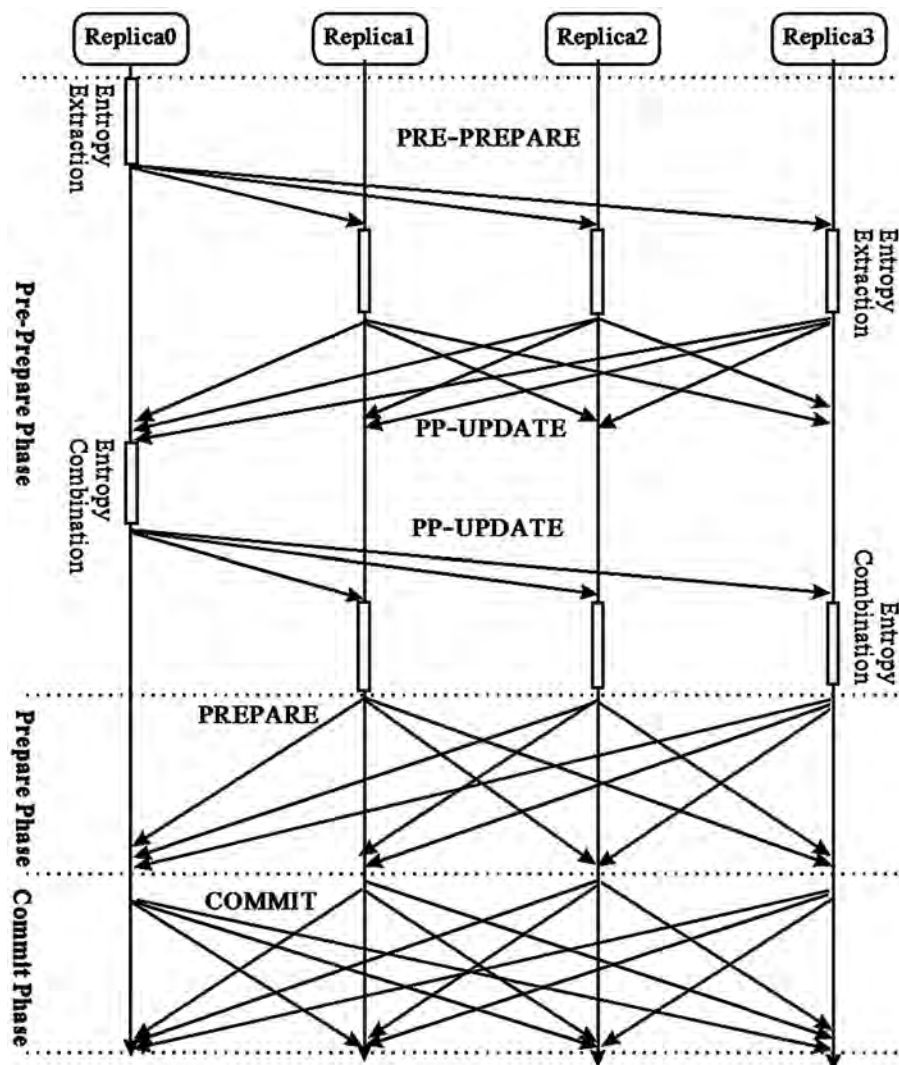
The normal operation of the CD-BFT algorithm is illustrated in Figure 1. The algorithm is executed by a set of $3f + 1$ replicas to tolerate up to f Byzantine faulty replicas. One of the replicas is designated as the primary while the remaining replicas are backups. Each replica is assigned a unique id i , where i varies from 0 to $3f$. For view v , the replica whose id i satisfies $i = v \bmod (3f + 1)$ would serve as the primary. The view starts from 0. For each view change, the view number is increased by one and a new primary is selected.

All messages exchanged among the replicas, and those between the replicas and the clients are protected by an authenticator (Castro & Liskov, 2002) (for multicast messages), or by a message authentication code (MAC) (for point-to-point communications). An authenticator is formed by a number of MACs, one for each target of the multicast. We assume that the replicas and the clients each has a public/private key pair, and the public keys are known to everyone. These keys are used to generate symmetric keys needed to produce/verify authenticators and MACs. To ensure freshness, the symmetric keys are periodically refreshed by the mechanism described in (Castro & Liskov, 2002). We assume that the adversaries have limited computing power so that they cannot break the security mechanisms described above.

A compromised replica may replace a high entropy source to which it uses to seed its random number generator with a deterministic algorithm, and convey such an algorithm and the random numbers collectively determined via a covert channel to its colluding clients. We assume that a faulty replica cannot explicitly piggyback such information with a reply through the normal communication channel, which could be ensured by using an application-level gateway, or a privacy firewall as described by Yin et al. (2003). However, we assume that the bandwidth of the covert channel is very limited such that a faulty replica

cannot transmit confidential state information to its colluding clients *in real time*, more specifically, given the size of the secret data L , the refresh period T , and the covert channel bandwidth R , we assume $T < LR$. For example, if the random number is of size 32 bits, and the covert channel bandwidth is 0.1 bit/second, the refresh period should be made less than 320 seconds. How to limit the bandwidth of covert channels is still under intense research (Gianvecchio & Wang, 2007; Shah, Molina, & Blaze, 2006), and it is out of the scope of this chapter.

Figure 1. Normal operation of the CD-BFT algorithm



When ordering a request, the primary determines the order of the request (*i.e.*, assigns the next sequence number to the request), and queries the application for the type of operation associated with the request. If the operation involves with a random number as input, the primary activates the mechanism for the CD-BFT algorithm. The primary then obtains its share of random number by extracting from its own entropy source, and piggybacks the share with the PRE-PREPARE message multicast to all backups. The pre-prepare message has the form $\langle \text{PRE-PREPARE}, v, n, d, R_p \rangle_{\alpha_p}$, where v is the view number, n is the sequence number assigned to the request, d is the digest of the request, R_p is the random number generated by the primary, and α_p is the authenticator for the message.

When a backup receives a PRE-PREPARE message, it performs the usual procedure such as the verification of the authenticator before it accepts the message. It also checks if the request will indeed trigger a randomized operation, to prevent a faulty primary from putting unnecessary loads on correct replicas (which might lead to a denial of service attack). If the PRE-PREPARE message is acceptable, the replica creates a PRE-PREPARE certificate for storing the relevant information, generates a share of random number from its entropy source, and multicasts to all replicas a PP-UPDATE message, in the form $\langle \text{PP-UPDATE}, v, n, i, R_i, d \rangle_{\alpha_i}$, where i is the sending replica identifier, R_i is the random number share from replica i .

When the primary has collected $2f$ PP-UPDATE messages, it combines the random numbers received according to a deterministic algorithm in the entropy combination step, as shown in Figure 1, and builds a PP-UPDATE message with slightly different content than those sent by backups. In the PP-UPDATE message sent by the primary, the R_i component is replaced by a set of $2f + 1$ tuples containing the random numbers contributed by replicas (including its own share), S . Each tuple

in the set S has the form $\langle R, i \rangle$. The replica R_i identifier is included in the tuple to ease the verification of the set at backups.

On receiving a PP-UPDATE message, a backup accepts the message and stores the message in its data structure provided that the message has a correct authenticator, it is in view v and it has accepted a PRE-PREPARE message to order the request with the digest d and sequence number n . A backup proceeds to the entropy combination step provided that it has accepted a PP-UPDATE message from the primary and $2f$ PP-UPDATE messages sent by the replicas referenced in the set S . The backup requests a retransmission from the primary for any missing PP-UPDATE message.

When it finishes the entropy combination, a backup multicasts a prepare message. The message has the form $\langle \text{PREPARE}, v, n, i, d' \rangle_{\alpha_i}$, where d' is the digest of the request concatenated by the combined random number. The replica then waits for $2f$ valid PREPARE messages from different replicas (possibly including the message sent or would have been sent by itself), after which, it multicasts a commit message to all replicas in the form $\langle \text{COMMIT}, v, n, i, d' \rangle_{\alpha_i}$. When a replica receives $2f+1$ valid COMMIT messages, it decides on the sequence number and the collectively determined random number. At the time of delivery to the application, both the request and the random number are passed to the application.

In Figure 1, the duration of the entropy extraction and combination steps have been intentionally exaggerated for clarify. In practice, the entropy combination can be achieved by applying a bit-wise exclusive-or operation on the set of random numbers collected, which is very fast. The cost of entropy extraction depends on the scheme used. Some schemes, such as the TrueRand method (Lacy et al., 1993), allow very fast entropy extraction. TrueRand works by gathering the underlying randomness from a computer by measuring the drift between the system clock and the interrupts-generation rate on the processor.

DISCUSSION

For clarity, when describing the CD-BFT algorithm, we have assumed that each remote method invocation involves only a single random number. In practice, an invocation might need to access a number of random numbers. Our algorithm can be trivially modified to accommodate this need. Let n be the number of random numbers needed for an operation. In the CD-BFT algorithm, during the pre-prepare phase, each replica obtains n random numbers from its entropy source and includes them in the PRE-PREPARE and PP-UPDATE messages. The entropy combination step would now combine n random numbers instead of 1. The remaining steps are identical to those in the original algorithm.

If the entropy consumption rate exceeds the entropy generation rate of the entropy source, which is the case for many practical systems, a pseudo-random number generator is normally used. If such systems are replicated for Byzantine fault tolerance, unless the pseudo-random number generator is frequently reseeded, it is not wise to collectively determine the seed to the random number generator so that the sequence of pseudo-random numbers can be generated deterministically without replica coordination. This is because if a compromised replica leaks the collectively-determined seed to some faulty clients (an adversary can take the time to leak the seed number through a very low-bandwidth covert channel, if the reseeding is not carried out frequently), they can predict all the future random numbers produced by the generator, which would compromise the integrity of the system (*i.e.*, the backward security will be lost). For better security, the replicas should instead seed their pseudo-random number generators independently and collectively determine *every* random number generated. It is also advised that each replica uses a different (strong) pseudo-random number generator for enhanced security. In fact, even if all correct replicas use the same relatively weak random number generator that can be broken by

searching 2^E space, our algorithm can expand the search space significantly to $2^{(f+1)E}$ because the collectively determined random number is based on contributions from at least $f + 1$ correct replicas.

FUTURE RESEARCH DIRECTIONS

It would be interesting to do a comparison study of our approach and the threshold cryptography (Desmedt, 1994; Rabin, 1998; Shoup, 2000). Threshold cryptography can be readily used to perform many secure operations across several replicas without ever exposing the security key, and hence, it is more robust against malicious attacks because it works even in the presence of high-bandwidth covert channels between a compromised replica and its colluding clients. However, threshold cryptograph is not without its limitations shown below. How to overcome these limitations would be very interesting research topics.

- Threshold cryptograph is extremely computationally expensive. We have experimented with a Java-based threshold cryptography framework on our test-bed. For a 1024 bit-long key, the key shares generation operation takes over 1 second, and the key shares combination operation takes over 2 seconds. This might be too expensive for most Web-based applications, which require soft real-time responses. It is worthwhile to investigate if a shorter key is sufficient depending on the security requirement of an application.
- To achieve proactive threshold cryptography, the key shares must be frequently refreshed, which involves not only expensive computations, but substantial message exchanges among the replicas as well. How this step impacts the overall system performance should be evaluated.
- Threshold cryptography assumes the availability of a trusted dealer to distribute the

initial shares. How to implement such a trusted dealer would be an interesting research challenge.

Nevertheless, threshold cryptography has been integrated into some Byzantine fault tolerance systems, such as in Byzantine fault tolerant distributed hash tables (Young et al., 2009), and Byzantine fault tolerant distributed storage (Cachin and Tessaro, 2005).

CONCLUSION

The use of good random numbers is crucial to the security of many mission-critical systems. However, when such systems are replicated for Byzantine fault tolerance, a serious issue arises, i.e., how do we preserve the integrity of the systems while ensuring strong replica consistency? Despite the fact that there exists a large body of work on how to render replicas deterministic under the benign fault model, the solutions regarding the random number control are often overly simplistic without regard to the security requirement, and hence, they are not suitable for practical Byzantine fault tolerance. In this chapter, we have presented a novel integrity-preserving replica coordination algorithm for Byzantine fault tolerant systems. The central idea behind our CD-BFT algorithm is that all random numbers to be used by the replicas are collectively determined, based on the contributions made by a quorum of replicas, at least $f+1$ of which are not faulty.

REFERENCES

Cachin, C., & Tessaro, S. (2006, June). Optimal resilience for erasure-coded Byzantine distributed storage. In *Proceedings of the International Conference on Dependable Systems and Networks*, (pp. 115-124). IEEE. doi:10.1109/DSN.2006.56

Castro, M., & Liskov, B. (2002). Practical Byzantine fault tolerance and proactive recovery. *ACM Transactions on Computer Systems*, 20(4), 398–461. doi:10.1145/571637.571640

Castro, M., Rodrigues, R., & Liskov, B. (2003). BASE: Using abstraction to improve fault tolerance. *ACM Transactions on Computer Systems*, 21(3), 236–269. doi:10.1145/859716.859718

Desmedt, Y. (1994). Threshold cryptography. *European Transactions on Telecommunications*, 5(4), 449–457. doi:10.1002/ett.4460050407

Dorrendorf, L., Gutterman, Z., & Pinkas, B. (2007). Cryptanalysis of the random number generator of the windows operating system. *Proceedings of the 14th ACM Conference on Computer and Communications Security*, 476–483.

Gianvecchio, S., & Wang, H. (2007). Detecting covert timing channels: an entropy-based approach. *Proceedings of the 14th ACM Conference on Computer and Communications Security*, 307–316. doi:10.1145/1315245.1315284

Gutterman, Z., & Malkhi, D. (2005). Hold your sessions: An attack on Java session-id generation. *Lecture Notes in Computer Science*, 3376, 44–57. doi:10.1007/978-3-540-30574-3_5

Kotla, R., Alvisi, L., Dahlin, M., Clement, A., & Wong, E. (2007). Zyzzyva: Speculative byzantine fault tolerance. *Proceedings of 21st ACM Symposium on Operating Systems Principles*, 45–58.

Lacy, J., Mitchell, D., & Schell, W. (1993). Cryptolib: Cryptography in software. *Proceedings of the 4th USENIX Security Symposium*, 1–17.

Lamport, L., Shostak, R., & Pease, M. (1982). The Byzantine generals problem. *ACM Transactions on Programming Languages and Systems*, 4(3), 382–401. doi:10.1145/357172.357176

Powell, D. (Ed.). (1991). *Delta-4: A Generic Architecture for Dependable Distributed Computing*. Springer-Verlag. doi:10.1007/978-3-642-84696-0

Rabin, T. (1998). A simplified approach to threshold and proactive RSA. *Proceedings of the 18th Annual International Cryptology Conference*, 89–104. doi:10.1007/BFb0055722

Schneider, F. (1990). Implementing fault-tolerant services using the state machine approach: A tutorial. *ACM Computing Surveys*, 22(4), 299–319. doi:10.1145/98163.98167

Shah, G., Molina, A., & Blaze, M. (2006). Keyboards and covert channels. *Proceedings of the 15th USENIX Security Symposium*, 59–75.

Shoup, V. (2000). Practical threshold signatures. *Lecture Notes in Computer Science*, 1097, 207–220. doi:10.1007/3-540-45539-6_15

Slember, J., & Narasimhan, P. (2006). Living with nondeterminism in replicated middleware applications. *Proceedings of the ACM/IFIP/USENIX 7th International Middleware Conference*, 81–100. doi:10.1007/11925071_5

Viega, J., & McGraw, G. (2002). *Building Secure Software*. Addison-Wesley.

Yin, J., Martin, J.-P., Venkataramani, A., Alvisi, L., & Dahlin, M. (2003). Separating agreement from execution for byzantine fault tolerant services. *Proceedings of the ACM Symposium on Operating Systems Principles*, 253–267. doi:10.1145/945445.945470

Young, A., & Yung, M. (2004). *Malicious Cryptography: Exposing Cryptovirology*. Wiley.

Young, M., Kate, A., Goldberg, I., & Karsten, M. (2010, June). Practical robust communication in dhds tolerating a byzantine adversary. In *Proceedings of the IEEE 30th International Conference on Distributed Computing Systems* (pp. 263–272). IEEE. doi:10.1109/ICDCS.2010.31

Zhang, H., Zhao, W., Moser, L. E., & Melliar-Smith, P. M. (2011). Design and Implementation of a Byzantine Fault Tolerance Framework for Non-Deterministic Applications. *IET Software*, 5(3), 342–356. doi:10.1049/iet-sen.2010.0013

Zhao, W. (2014). *Building dependable distributed systems*. John Wiley & Sons. doi:10.1002/9781118912744

ADDITIONAL READING

Adya, A., Bolosky, W. J., Castro, M., Cermak, G., Chaiken, R., Douceur, J. R.,... & Wattenhofer, R. P. (2002). FARSITE: Federated, available, and reliable storage for an incompletely trusted environment. *ACM SIGOPS Operating Systems Review*, 36(SI), 1–14.

Aiyer, A. S., Alvisi, L., Clement, A., Dahlin, M., Martin, J. P., & Porth, C. (2005). BAR fault tolerance for cooperative services. *Operating Systems Review*, 39(5), 45–58. doi:10.1145/1095809.1095816

Amir, Y., Coan, B., Kirsch, J., & Lane, J. (2008, June). Byzantine replication under attack. In *Dependable Systems and Networks With FTCS and DCC*, 2008. DSN 2008. IEEE International Conference on (pp. 197–206). IEEE. doi:10.1109/DSN.2008.4630088

Amir, Y., Danilov, C., Dolev, D., Kirsch, J., Lane, J., Nita-Rotaru, C.,... Zage, D. (2006, June). Scaling Byzantine Fault-Tolerant Replication to Wide Area Networks. In *Dependable Systems and Networks*, 2006. DSN 2006. International Conference on (pp. 105–114). IEEE. doi:10.1109/DSN.2006.63

Amir, Y., Danilov, C., Dolev, D., Kirsch, J., Lane, J., Nita-Rotaru, C., & Zage, D. et al. (2008). Steward: Scaling Byzantine Fault-Tolerant Replication to Wide Area Networks. *IEEE Transactions on Dependable and Secure Computing*, 80–93.

Bessani, A. N., da Silva Fraga, J., & Lung, L. C. (2006, April). BTS: A Byzantine fault-tolerant tuple space. In *Proceedings of the 2006 ACM symposium on Applied computing* (pp. 429–433). ACM. doi:10.1145/1141277.1141377

- Chai, H., Melliar-Smith, P. M., Moser, L. E., Zhang, H., & Zhao, W. (2013). Toward Trustworthy Coordination of Web Services Business Activities. *IEEE Transactions on Services Computing*, 6(2), 276–288. doi:10.1109/TSC.2011.57
- Chun, B. G., Maniatis, P., & Shenker, S. (2008, June). Diverse Replication for Single-Machine Byzantine-Fault Tolerance. In *USENIX Annual Technical Conference* (pp. 287-292).
- Clement, A., Wong, E. L., Alvisi, L., Dahlin, M., & Marchetti, M. (2009, April). Vol. 9, pp. 153–168). Making Byzantine Fault Tolerant Systems Tolerate Byzantine Faults. In NSDI.
- Cowling, J., Myers, D., Liskov, B., Rodrigues, R., & Shriram, L. (2006, November). HQ replication: A hybrid quorum protocol for Byzantine fault tolerance. In *Proceedings of the 7th symposium on Operating systems design and implementation* (pp. 177-190). USENIX Association.
- Distler, T., & Kapitza, R. (2011, April). Increasing performance in byzantine fault-tolerant systems with on-demand replica consistency. In *Proceedings of the sixth conference on Computer systems* (pp. 91-106). ACM. doi:10.1145/1966445.1966455
- Driscoll, K., Hall, B., Sivencrona, H., & Zumsteg, P. (2003). Byzantine fault tolerance, from theory to reality. In *Computer Safety, Reliability, and Security* (pp. 235-248). Springer Berlin Heidelberg. doi:10.1007/978-3-540-39878-3_19
- Garcia, R., Rodrigues, R., & Preguiça, N. (2011, April). Efficient middleware for byzantine fault tolerant database replication. In *Proceedings of the sixth conference on Computer systems* (pp. 107-122). ACM. doi:10.1145/1966445.1966456
- Hendricks, J., Ganger, G. R., & Reiter, M. K. (2007). Low-overhead byzantine fault-tolerant storage. *Operating Systems Review*, 41(6), 73–86. doi:10.1145/1323293.1294269
- Kapitza, R., Behl, J., Cachin, C., Distler, T., Kuhnle, S., Mohammadi, S. V., & Stengel, K. et al. (2012, April). CheapBFT: resource-efficient byzantine fault tolerance. In *Proceedings of the 7th ACM european conference on Computer Systems* (pp. 295-308). ACM.
- Kotla, R., & Dahlin, M. (2004, June). High throughput Byzantine fault tolerance. In *Dependable Systems and Networks, 2004 International Conference on* (pp. 575-584). IEEE. doi:10.1109/DSN.2004.1311928
- Veronese, G. S., Correia, M., Bessani, A. N., & Lung, L. C. (2009, September). Spin One's Wheels? Byzantine Fault Tolerance with a Spinning Primary. In *Reliable Distributed Systems, 2009. SRDS'09. 28th IEEE International Symposium on* (pp. 135-144). IEEE. doi:10.1109/SRDS.2009.36
- Yin, J., Martin, J. P., Venkataramani, A., Alvisi, L., & Dahlin, M. (2002, June). Byzantine fault-tolerant confidentiality. In *Proceedings of the International Workshop on Future Directions in Distributed Computing* (pp. 12-15).
- Zhang, H., Chai, H., Zhao, W., Melliar-Smith, P. M., & Moser, L. E. (2012). Trustworthy coordination of Web services atomic transactions. *IEEE Transactions on Parallel and Distributed Systems*, 23(8), 1551–1565. doi:10.1109/TPDS.2011.292
- Zhang, Y., Zheng, Z., & Lyu, M. R. (2011, July). BFTCloud: A byzantine fault tolerance framework for voluntary-resource cloud computing. In *Proceedings of the IEEE International Conference on Cloud Computing* (pp. 444-451). IEEE. doi:10.1109/CLOUD.2011.16
- Zhao, W. (2008, December). Integrity-preserving replica coordination for byzantine fault tolerant systems. In *Proceedings of the 14th IEEE International Conference on Parallel and Distributed Systems* (pp. 447-454). IEEE. doi:10.1109/ICPADS.2008.45

Zhao, W. (2009). Design and implementation of a Byzantine fault tolerance framework for Web services. *Journal of Systems and Software*, 82(6), 1004–1015. doi:10.1016/j.jss.2008.12.037

Zhao, W. (2015). Optimistic Byzantine fault tolerance. *International Journal of Parallel, Emergent and Distributed Systems*, 1-14 (preprint).

Zhao, W., & Zhang, H. (2009). Proactive service migration for long-running Byzantine fault-tolerant systems. *IET software*, 3(2), 154-164.

KEY TERMS AND DEFINITIONS

Entropy: It is a measure of uncertainty, or randomness.

Message Authentication Code (MAC): A MAC is produced by a keyed secure hash function on a message. It is used to ensure the integrity of the message such that if a message protected by a MAC is tampered, it can be detected by comparing the MAC included with in the message and the recomputed MAC.

Pseudo-Random Number Generator: It is an algorithm used to generate a sequence of numbers that approximate the properties of random numbers. The algorithm depends on one or

a small set of initial numbers, usually referred to as the seed to the generator.

Quorum: A quorum of a set consists of the minimum number of components to perform a predefined function. In Byzantine fault tolerance replication, $2f+1$ is needed to form a quorum in a set of $3f+1$ replicas.

Random Number: It is a number generated by some process that cannot be reproduced or predicted.

Security: The security of a system refers to its capability of protecting itself from harm, such as external attacks. More specifically, a secure system is one that guarantees confidentiality, integrity, and the availability of the system.

Strong Replica Consistency: The states of the replicas of a process should remain identical at the end of the processing of each request.

System Integrity: The integrity of a system refers to the capability of performing correctly according to the original specification of the system under various adversarial conditions.

Threshold Cryptography: Basic cryptographic operations such as encryption, decryption, signature generation, and verification are performed by a group of processes without reconstructing the shared secret.

Category C

Computer Simulation

Data Visualization Strategies for Computer Simulation in Bioelectromagnetics



Akram Gasmelseed
Qassim University, Saudi Arabia

Ali H. Alharbi
Qassim University, Saudi Arabia

INTRODUCTION

Bioelectromagnetics is a branch of science that examines how living organisms interact with electromagnetic fields. In a broader sense, it is a discipline that spans the fields of signal processing, electromagnetics, medical imaging, physiology, medical physics, anatomy, occupational and environmental health, behavioral, and computer sciences.

Due to its provision of interactive and flexible programming environment, this chapter describes a LabVIEW based data visualization system that has been implemented and used in bioelectromagnetics simulations (Gasmelseed, 2013). The graphical nature of the LabVIEW programming environment makes it very easy to create a sophisticated application in a minimum amount of time, especially when compared to designing graphical user interfaces in a low-level programming language.

BACKGROUND

Bioelectromagnetics

Bioelectromagnetics is the field that depicts the interaction of electromagnetic (EMW) waves with biological systems (Furse et al., 2009). The Common core area of study in bioelectromagnetics includes the study of the interactions between electromagnetic waves and human tissue. Recent

advances in computational techniques have made it easier to assess the EMW behavior within the exposed biological tissue model numerically.

Computational techniques are extremely powerful for solving bioelectromagnetic problems and have been successfully employed in the modelling of microwave applicators used in hyperthermia (Polk & Postow, 1995; Maruyama et al., 2006), calculating electromagnetic signal absorption in human body models (Reyhani & Ludwig, 2006), and for dosimetry of mobile handhelds (Lin, 2006).

Laboratory Virtual Instrument Engineering Workbench (LabVIEW)

LabVIEW (Laboratory Virtual Instrument Engineering Workbench) is a National Instruments program development application (<http://www.ni.com/labview/>) with computational capabilities (e.g. analysis, visualization) similar to those found in C, Fortran or Matlab development systems (Samsi et al., 2010). LabVIEW is different from those applications in one important respect: other programming systems use text-based languages to create lines of code, while LabVIEW uses a graphical programming language called 'G' to create programs in block diagram form (Kalkman, 1995; Olansen & Rosow, 2001). LabVIEW also has extensive libraries of functions and subroutines for most programming tasks such as data analysis, data presentation, and data storage. LabVIEW includes conventional program development tools, allowing the user to set breakpoints, animate pro-

DOI: 10.4018/978-1-5225-2255-3.ch108

gram execution to see how data passes through the program and single-step through the program to make debugging and program development easier. Graphical programming is based on the concept of data flow, in contrast to the sequential logic of most text-based programming languages. This means that execution of a block or graphical component is dependent on the flow of data, or, more specifically, a block executes when data are made available at all of the block's input and output data are sent to all other connected blocks.

LabVIEW is a multithreaded programming language, so that specific operations within a single application can be subdivided into individual threads, each of which can theoretically run in parallel. For example, the block diagrams in LabVIEW are parallel programs. The LabVIEW compiler automatically breaks up these parallel programs into multiple threads for the user and passes these threads to the operating system for assignment to multiple processing cores. In comparison, text-based languages are sequential, that is, the code is basically run line by line.

Moreover, researchers in the field of bioelectromagnetics prefer to integrate C subroutines into MATLAB or Simulink code in order to speed-up their execution (Laakso & Hirata, 2012; Quinto et al., 2011). Integrating C subroutines with MATLAB or with Simulink is a difficult task and demands knowledge of the MATLAB MEX interface or the Simulink S-function interface.

On the other hand, in LabVIEW, C subroutine can be integrated using the code interface nodes (CINs) (Johnson, 1998). LabVIEW will create the prototype function headers and linking references, easing the insertion of the required C subroutine.

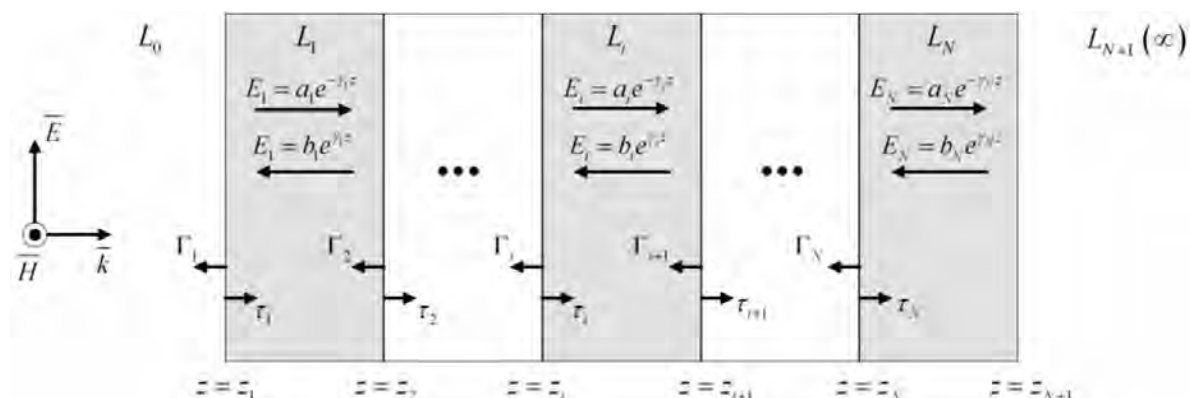
In this chapter, a LabVIEW development environment is designed to simulate those EM/biological tissue interactions. Two standard methods are presented to investigate the effect of electromagnetic radiation on the human tissue: the multilayered method, and the finite difference time domain (FDTD) method. The LabVIEW development environment provides a graphical user interface to the model that makes it easy to observe and modify the behavior of each element of the model, relatively easy to understand and they can be easily modified for the user's specific purpose.

METHODOLOGY

Multilayered Method

One of the standard methods to model EM signal absorption in human tissue is the multilayered method. The method is based on using simple planar models to solve the absorption problem. While these planar models may be simple, important properties of the absorption can be

Figure 1. General biological multilayered structure
Adapted from Akram & Jasmy, 2008.



illustrated (Akram & Andrew, 2006; Samaras & Christ, 2007).

Figure 1 shows the geometry of the problem. A plane wave travelling in the $+z$ -direction with the electric field $\vec{E}$ vector linearly polarized along the x -axis is incident upon biological multilayered structure.

The incident field is assumed to have harmonic time variation $e^{j\omega t}$. There are $N+2$ layers ($L_0, L_1, \dots, L_{N+1}$), and $N+1$ interfaces ($z_1, z_2, \dots, z_{N+1}$). The i^{th} layer indicated by L_i has permittivity ϵ_i [F/m], permeability μ_i [H/m], and conductivity σ_i [S/m]. Further, we denote the left-most layer (i.e. air) as L_0 , while we denote the right-most layer by L_{N+1} . Then, the electric ($\vec{E}_i$) and magnetic ($\vec{H}_i$) fields in the i^{th} layer are given by,

$$\vec{E}_i = (a_i e^{-j\gamma_i z} + b_i e^{j\gamma_i z}) \hat{x}$$

$$\vec{H}_i = \frac{(a_i e^{-j\gamma_i z} - b_i e^{j\gamma_i z})}{\eta_i} \hat{y}$$

where $i = 0, 1, \dots, N+1$. The constants a_i and b_i are the amplitudes of the incident and re-

flected waves, respectively, where the amplitude of the incident electric field is represented by a_0 , and is known in advance. γ_i and η_i denotes the propagation constant and the characteristic impedance at L_i , respectively.

The wave impedance is given by,

$$Z_i = \eta_i \frac{(1 + \Gamma_i e^{2j\gamma_i z_i})}{(1 - \Gamma_{i+1} e^{2j\gamma_{i+1} z_{i+1}})} \quad (5)$$

At the interface $z = z_i$, the reflection (Γ_i) coefficient is given by,

$$\Gamma_i = \frac{Z_i - \eta_{i-1}}{Z_i + \eta_{i-1}} e^{-2j\gamma_{i-1} z_i} \quad (6)$$

LabVIEW Implementation of Multilayered Code

To demonstrate our approach we have adapted an example that considers a multilayered biological structure (Karwowski, 2007) composed of three layers of tissues: skin, fat and muscle. In this simulation, the layered tissue model shown in Figure 2 is assumed to be irradiated by a 5.5 GHz plane-wave traveling in air, with the electric field component

Figure 2. Multilayered tissue model consisting of skin-fat-muscle

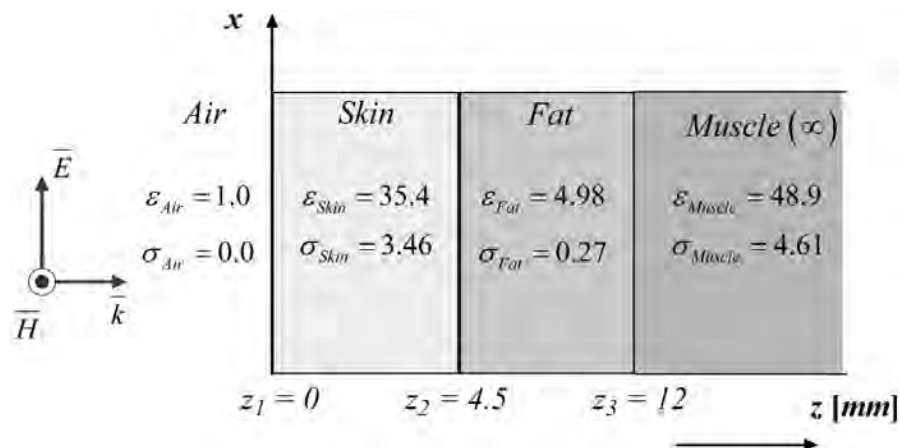
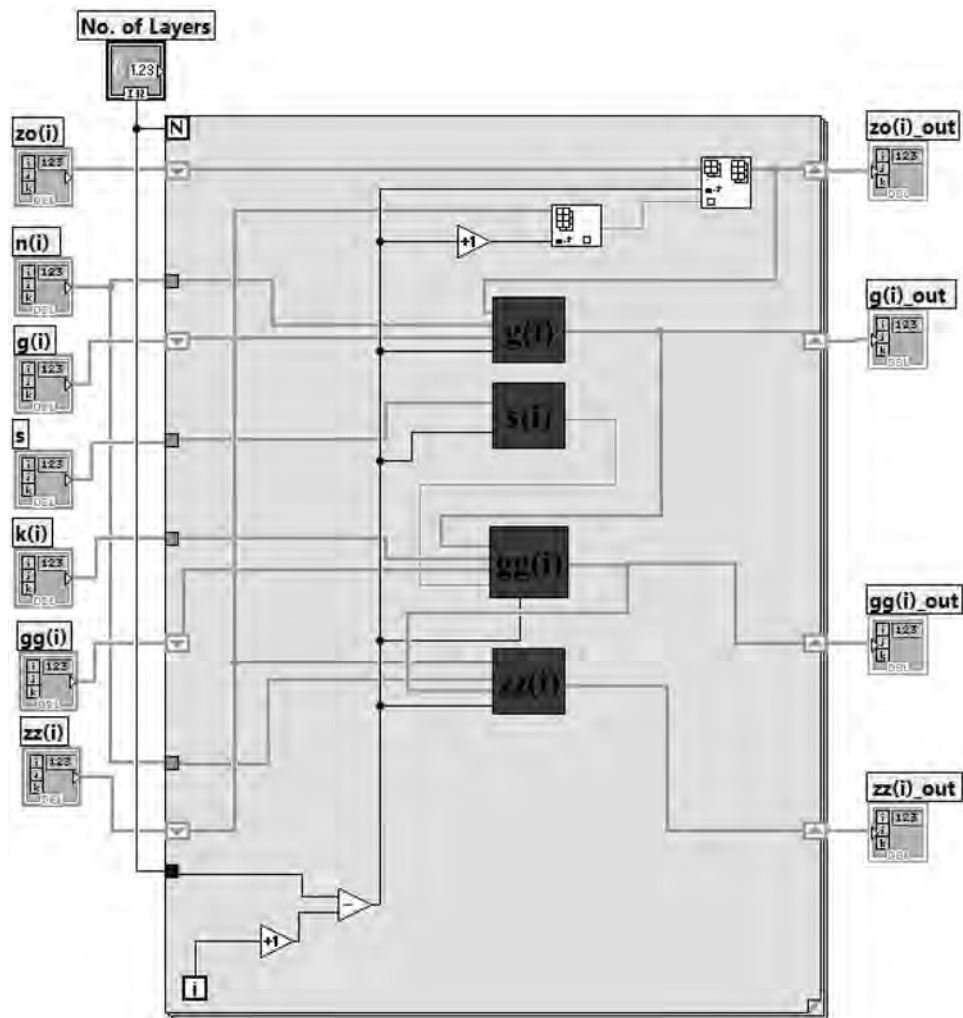


Figure 3. The block diagram that executes equations (5) and (6)



E_x having a unit magnitude, impinging normally at $z = 0$ on a tissue-air interface. LabVIEW's block diagram (shown in Figure 3) consists of a *for-loop*, *shift-registers*, and *array* tools.

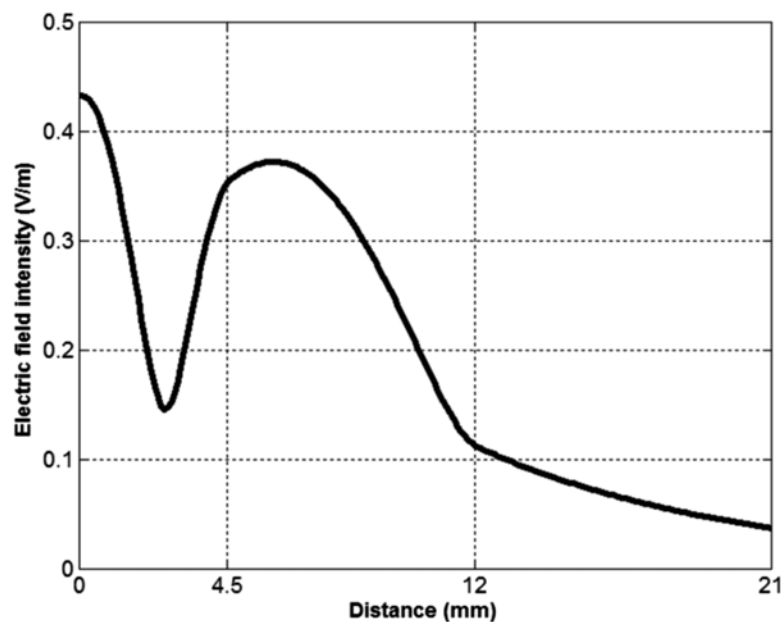
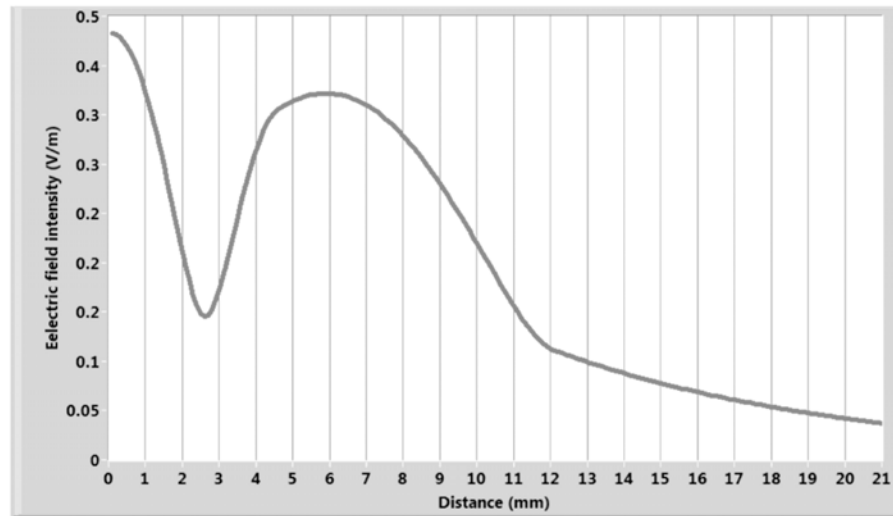
Figure 4 shows results of computed electric field strength (1) obtained from LabVIEW program and MATLAB program.

Finite Difference Time Domain (FDTD) Method

The finite-difference time-domain (FDTD) method has become widely used for electromagnetic computation since the initial work of Yee KS (Yee, 1966) due to its power in solving Max-

well's equations accurately and efficiently within a highly scattering medium in a discretized form. The method uses a leapfrog scheme on staggered Cartesian grids where the electric field (E) is offset spatially and temporally from the magnetic field (H). This approach solves the present fields throughout the computational domain in terms of the past fields.

In this section, the FDTD method for a basic scattering modeling is presented (Akram & Jasmy, 2007). We present a one-dimensional (1-D) formulation with simple outer radiation boundary condition (ORBC) to avoid reflections from the edges of the modeling grid.

Figure 4. Comparison of computed electric field strength (E_i) in both (a) LabVIEW and (b) MATLAB

LabVIEW Implementation of One Dimensional (1-D) Formulation

To give a description of the FDTD method, initially a 1-D formulation will be used. Maxwell's scalar equations for a non-permeable (Taflov & Hagness, 2005) lossy medium ($\mu = \mu_0$, $\sigma^* = 0$) are,

$$\sigma E_z + \epsilon \frac{\partial E_z}{\partial t} = \frac{\partial H_y}{\partial x}$$

$$\mu \frac{\partial H_y}{\partial t} = \frac{\partial E_z}{\partial x}$$

These equations represent a plane wave with electric field oriented in the z -direction, the magnetic field oriented in the y -direction, and traveling in the x -direction. Equations (8) and (9) yield the discretized update equations when approximated by finite differences,

$$E_z^{n+1}[i] = Ca[i] \cdot E_z^n[i] + Cb[i] \cdot \left(H_y^{n+\frac{1}{2}}\left[i + \frac{1}{2}\right] - H_y^{n+\frac{1}{2}}\left[i - \frac{1}{2}\right] \right)$$

$$H_y^{n+\frac{1}{2}}\left[i + \frac{1}{2}\right] = H_y^{n-\frac{1}{2}}\left[i + \frac{1}{2}\right] + \frac{\Delta t}{\mu \delta} \cdot \left(E_z^n[i+1] - E_z^n[i] \right)$$

where the terms $Ca[i]$ and $Cb[i]$ are given by,

$$Ca[i] = \frac{1 - \frac{\sigma[i] \Delta t}{2\varepsilon[i]}}{1 + \frac{\sigma[i] \Delta t}{2\varepsilon[i]}}$$

$$Cb[i] = \frac{\frac{\Delta t}{\varepsilon[i] \delta}}{1 + \frac{\sigma[i] \Delta t}{2\varepsilon[i]}}$$

Equations (10) and (11) are easily programmed for solution by iteration of each equation. To illustrate the algorithm, a sinusoidal wave (1-GHz) propagating in a nonpermeable lossy medium ($\varepsilon_r = 1.0$, $\sigma = 0.005$ S/ms modeled) is used. The grid resolution ($dx = 2.5$ cm) is chosen to provide 20 samples per wavelength. The Courant factor $S = c \cdot dt/dx$ is set to the stability limit $S = 1$ to ensure numerically stable update in Equations (10) and (11). The number of grid cells (IE) in the x -direction is 200, and the total number of time steps is 240. A MATLAB program segment that executes time-stepping of E_z in the FDTD space lattice can be written as,

Figure 5. The block diagram of E_z field calculation

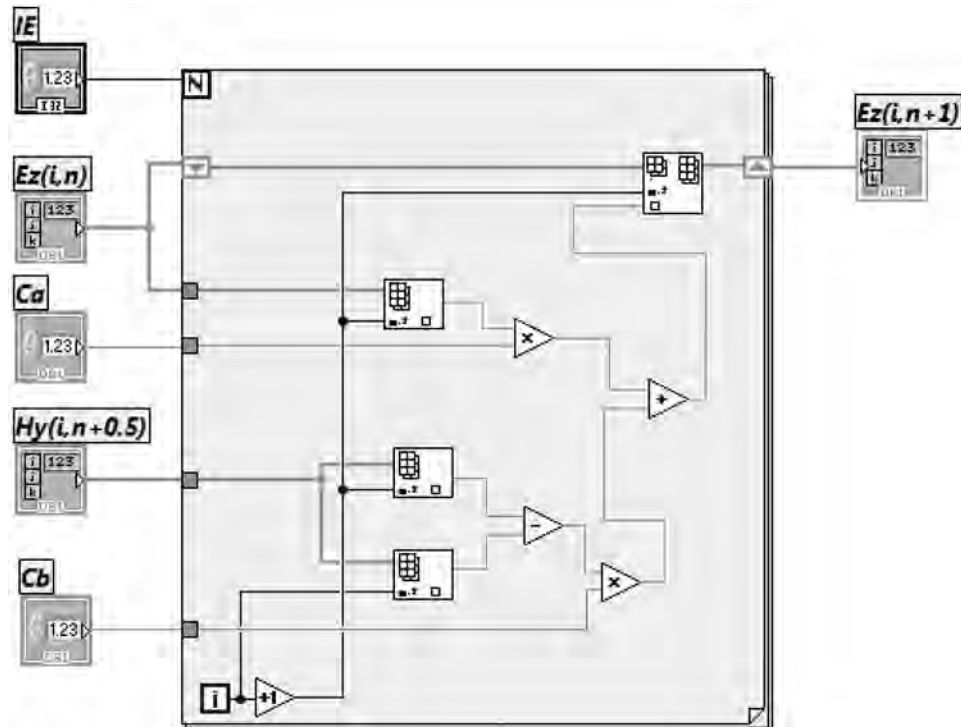
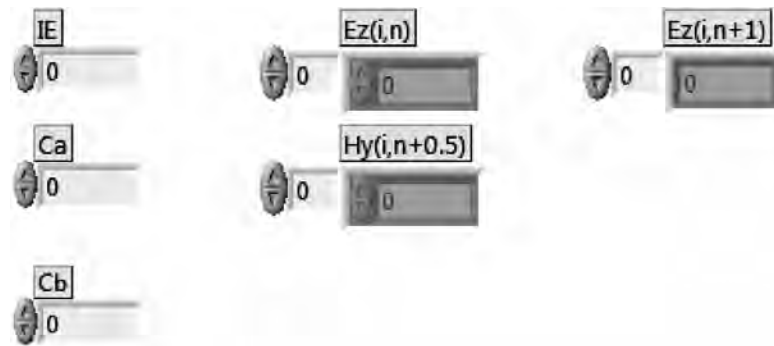


Figure 6. The front panel of E_z field calculation



$$E_z(2:IE) = Ca * E_z(2:IE) + Cb * (Hy(2:IE) - Hy(1:IE-1));$$

Following closely the LabVIEW instructions and using the FDTD model, the *block diagram* for E_z field calculation (10) (and shown in Figure 5) consists of a *for-loop*, *shift-registers* and *array* tools. The corresponding *front panel* for the E_z field calculation is shown in Figure 6 and Results of this programming construct are shown in Figure 7.

LabVIEW *Graphs* and *charts* are used to display data in a graphical form. *Graphs* and *charts* differ in the way they display and update data. *VIs* with graphs usually collect data in an *array* and then plot the data to the graph, which is similar to a spreadsheet that first stores the data and then generates a plot of it. In contrast, a *chart* appends new data points to those already in the display to create a history. On a *chart*, we can see the current reading or measurement in context with data previously acquired.

Recently, more sophisticated models were introduced and their aims is to provide a realistic human model. These anatomical-based voxelized inhomogeneous models are extracted from computed tomography (CT), magnetic resonance imaging (MRI) or X-ray scans (e.g. Visible Human (Spitzer & Ackerman, 1996) and Virtual Family (Christ et al., 2007)). In Matlab, creating a highly inhomogeneous model is very easy by providing to each array cell a material index and using indirect addressing of the updating coefficients.

In order to incorporate such models into LabVIEW applications, it is needed to use the IMAQ (image acquisition) Vision for LabVIEW (National Instruments Corp, 2007). IMAQ Vision for LabVIEW is a library of LabVIEW virtual instruments (VIs) that can be used to develop machine vision and scientific imaging applications.

FUTURE RESEARCH DIRECTIONS

The theory and simulation experiments in this chapter will lead to further understanding of electromagnetic interactions.

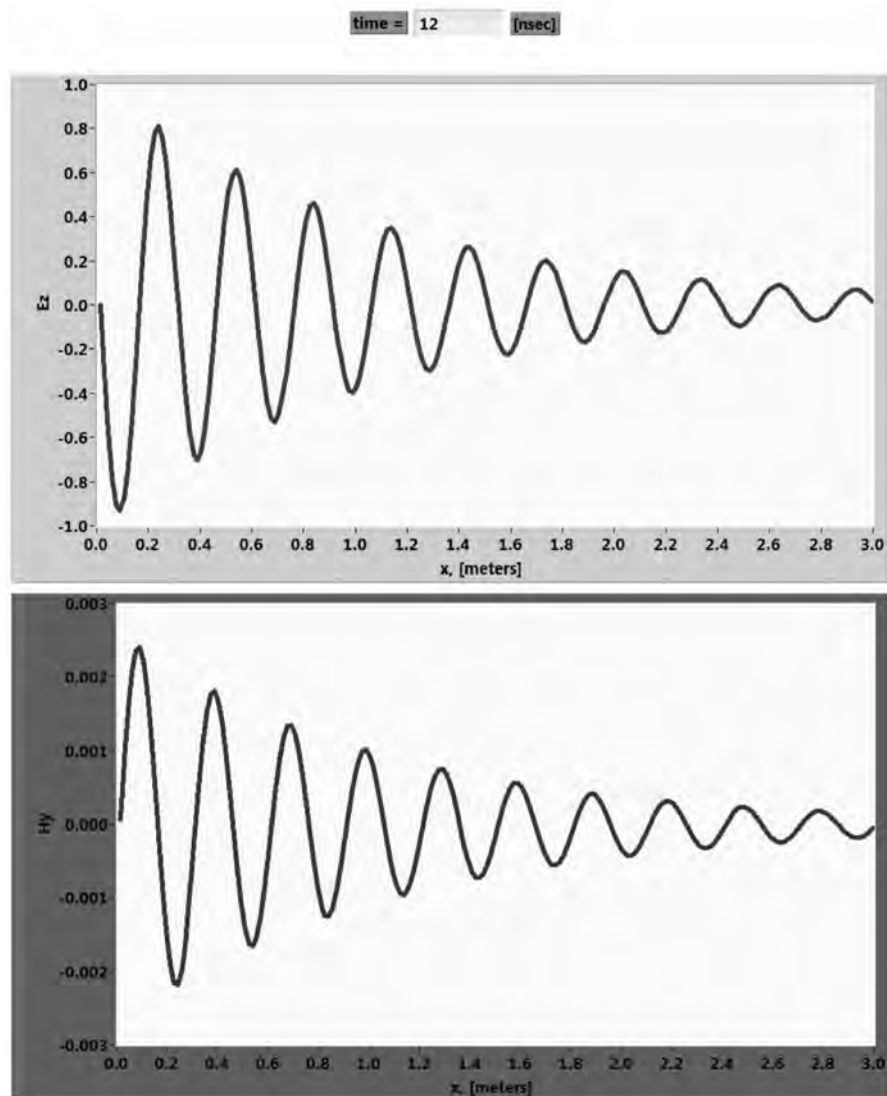
In future, research can be conducted to conveniently integrate DAQ cards in our software frame of LabVIEW. As a scientific issue, the interaction of electromagnetic energy with biological systems is hardly new, and such integration can be considered a novelty in scientific literature of research on electromagnetic interactions.

With DAQ integration and the availability of fast computers in modern times, we will be able to examine the electromagnetic interactions in real time.

CONCLUSION

Computer modelling will likely remain the best method to determine the electric fields inside body organs and tissues in many practical situations. There is a close relationship between the

Figure 7. The FDTD computed E_z and H_y at every time step



mathematical description used in the biological model and the computational method for solving the fields inside the model. Further, the availability of more powerful computers and refinement of computer codes will allow the determination of the fields inside the biological model in greater details and allow for more realistic models of EM emitting devices to be incorporated.

In this article, a LabVIEW development environment is discussed to simulate those EM-biological tissue interactions. Two standard methods are presented to investigate the effect of

electromagnetic radiation on the human tissue: the multilayered method, and the finite difference time domain (FDTD) method. The LabVIEW development environment provides a graphical user interface to the model that makes it easy to observe and modify the behavior of each element of the model, is relatively easy to understand and can be easily modified for the user's specific purpose.

The advantages of using the LabVIEW development environment are many. First, the graphical nature of the development environment makes it easy to create sophisticated applications in a

minimum amount of time, especially when compared to designing graphical user interfaces in a low-level programming language like C/C++. Moreover, the structure of LabVIEW makes it easy to add new components to the model when the need arises. The user interface allows the user to observe and gain insight into the behaviour of the model. Finally, VIs has both an interactive user interface and a source code equivalent and they accept parameters from higher-level VIs.

REFERENCES

- Akram, G., & Andrew, T. (2006). A multilayered model of human head irradiated by electromagnetic plane wave of 100 MHz–300 GHz. *International Journal of Scientific Research*, 16, 397–403.
- Akram, G., & Jasmy, Y. (2007). Numerical simulation of the FDTD method in LabVIEW. *IEEE Microwave Magazine*, 8(6), 90–99. doi:10.1109/MMM.2007.906920
- Akram, G., & Jasmy, Y. (2008). LabVIEW-based multilayered model for estimation of the absorbed energy inside biological tissues. *IEEE Antennas and Propagation Magazine*, 50(2), 152–158. doi:10.1109/MAP.2008.4562280
- Bitter, R., Mohiuddin, T., & Nawrocki, M. (2000). *LabVIEW: Advanced Programming Techniques*. Boca Raton, FL: CRC. doi:10.1201/9781420039351
- Christ, A., Kainz, W., Hahn, E., Honegger, K., Shen, J., Rascher, W., ... Kuster, N. (2007). *The virtual family project - development of anatomical whole-body models of two adults and two children*. Paper presented in at the 23rd annual Review of Progress in Applied Computational Electromagnetics.
- Furse, C., Durney, C. H., & Christensen, D. A. (2009). *Basic Introduction to Bioelectromagnetics*. Boca Raton, FL: CRC. doi:10.1201/9781420055436
- Gasmelseed, A. (2013). New development environment for modern bioelectromagnetics signal processing applications. *IEEE Microwave Magazine*, 14(5), 134–152. doi:10.1109/MMM.2013.2259404
- Johnson, G. W. (1998). *LabVIEW Power Programming*. McGraw-Hill Prof. Publishing.
- Kalkman, C. J. (1995). LabVIEW: A software system for data acquisition, data analysis, and instrument control. *Journal of Clinical Monitoring*, 11(1), 51–58. doi:10.1007/BF01627421 PMID:7745456
- Karwowski, A. (2004). Improving accuracy of FDTD simulations in layered biological tissues. *IEEE Microwave and Wireless Components Letters*, 14(4), 151–152. doi:10.1109/LMWC.2004.827103
- Laakso, I., & Hirata, A. (2012). Fast multigrid-based computation of the induced electric field for transcranial magnetic stimulation. *Physics in Medicine and Biology*, 57(23), 7753–7765. doi:10.1088/0031-9155/57/23/7753 PMID:23128377
- LabVIEW System Design Software. (n.d.). Retrieved from <http://www.ni.com/labview/>
- Lin, J. C. (2006). Update of IEEE radio frequency exposure guidelines. *IEEE Microwave Magazine*, 7(2), 24–28. doi:10.1109/MMW.2006.1634013
- Maruyama, T., Sawaya, Y., Sato, F., Matsuki, H., Aiba, S., Ito, Y., & Sato, T. (2006). Heat element of magnetic hyperthermia in mouse melanoma model. *IEEE Transactions on Magnetics*, 42(10), 3590–3592. doi:10.1109/TMAG.2006.879088
- National Instruments Corp. (2007). *IMAQ Vision Concept Manual*. The National Instruments Corp.
- Olansen, J. B., & Rosow, E. (2001). *Virtual Bio-Instrumentation, Biomedical, Clinical and Healthcare Applications in LabVIEW*. Prentice-Hall PTR.

Polk, C., & Postow, E. (1995). *Handbook of Biological Effects of Electromagnetic Fields*. Boca Raton, FL: CRC.

Quinto, M., Boutami, S., & Hazart, J. (2011). Fast computation of dipole radiation in stratified background using graphics processing unit. *Progress in Electromagnetics Research*, 20, 115–126. doi:10.2528/PIERM11052505

Reyhani, S. M., & Ludwig, S. A. (2006). An implanted spherical head model exposed to electromagnetic fields at a mobile communication frequency. *IEEE Transactions on Bio-Medical Engineering*, 53(10), 2092–2101. doi:10.1109/TBME.2006.881770 PMID:17019874

Samaras, T., Christ, A., Klingenbock, A., & Kuster, N. (2007). Worst case temperature rise in a one-dimensional tissue model exposed to radiofrequency radiation. *IEEE Transactions on Bio-Medical Engineering*, 54(3), 492–496. doi:10.1109/TBME.2006.890498 PMID:17355061

Samsi, S., Gadepally, V., & Krishnamurthy, A. (2010). MATLAB for signal processing on multiprocessors and multicores. *IEEE Signal Processing Magazine*, 27(2), 40–49. doi:10.1109/MSP.2009.935421

Spitzer, V., Ackerman, M. J., Scherzinger, A. L., & Whitlock, D. (1996). The visible human male: A technical report. *Journal of the American Medical Informatics Association*, 3(2), 118–130. doi:10.1136/jamia.1996.96236280 PMID:8653448

Taflove, A., & Hagness, S. C. (2005). *Computational Electrodynamics: The Finite Difference Time Domain Method*. Norwood, MA: Artech House, Inc.

Yee, K. S. (1966). Numerical solution of initial boundary value problems involving Maxwells equations in isotropic media. *IEEE Transactions on Antennas and Propagation*, 14(3), 302–307. doi:10.1109/TAP.1966.1138693

ADDITIONAL READING

Akram, G. (2011). New head models extracted from thermal infrared (IR) images for dosimetry computations. *Computer Methods in Biomechanics and Biomedical Engineering*, 14(7), 665–671. doi:10.1080/10255842.2011.563738 PMID:21480080

Akram, G. (2011). Electromagnetic energy absorption patterns in subjects with common visual disorders. *Electromagnetic Biology and Medicine*, 30(3), 136–145. doi:10.3109/15368378.2011.596248 PMID:21861692

Akram, G., & Jasmy, Y. (2007). The effect of aging-related variations of normal human head on the induced specific absorption rate (SAR): plane multilayered head model. Paper presented at the IEEE Asia-Pacific Conference on Applied Electromagnetics, Melaka. doi:10.1109/APACE.2007.4603868

Bitter, R., Mohiuddin, T., & Nawrocki, M. (2000). *LabVIEW: Advanced Programming Techniques*. Boca Raton, FL: CRC. doi:10.1201/9781420039351

Derat, B., & Bolomey, J. C. (2006). Various optimization problems of electromagnetic power absorption in homogeneous and heterogeneous phantoms. *IEEE Electromagnetic Compatibility*, 48(4), 641–647. doi:10.1109/TEMPC.2006.884619

Drossos, A., Santoma, V., & Kuster, N. (2000). The dependence of Electromagnetic energy absorption upon human head tissue composition in the frequency range of 300–3000 MHz. *IEEE Transactions on Microwave Theory and Techniques*, 48(11), 1988–1995. doi:10.1109/22.884187

Eker, S. S., & Yeldiren, B. (2001). *Multi-layered lossy finite length dielectric cylindrical model of man at oblique incidence*. Paper presented at the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Istanbul. doi:10.1109/IEMBS.2001.1019721

Sullivan, D. M. (1997). An unsplit step 3-D PML for use with the FDTD method. *IEEE Microwave and Guided Wave Letters*, 7(7), 184–186. doi:10.1109/75.594858

Sullivan, D. M. (2000). *Electromagnetic Simulation Using the FDTD Method*. NY: IEEE Press. doi:10.1109/9780470544518

Xie, Y., Guo, B., Li, J., & Stoica, P. (2006). Novel multistatic adaptive microwave imaging methods for early breast cancer detection. *EURASIP Journal on Advances in Signal Processing*.

Zombolas, C., & Sargent, A. (2006). *SAR compliance in Australia - requirements for mobile and portable transmitting devices*. Paper presented at *IEEE 17th IEEE International Symposium on Electromagnetic Compatibility*, Zurich. doi:10.1109/EMCZUR.2006.214880

KEY TERMS AND DEFINITIONS

Bioelectromagnetics: A branch of science that describes the interaction of electromagnetic (EMW) waves with living organisms.

Computational Techniques: The techniques that use computers to analyze and solve scientific and engineering problems.

Dosimetry: The measurement and assessment of the absorbed electromagnetic radiation in biological tissue.

Finite-Difference Time-Domain (FDTD): A mathematical technique used for finding an approximate solution to differential equations used in various areas of applied sciences.

LabVIEW: An acronym for “Laboratory Virtual Instrument Engineering Workbench” and is a development environment for a visual programming language from National Instruments (NIs) designed for every engineer and scientist to solve real world problems compared to text-based programming (e.g. Matlab, C).

Multilayered Method: A technique used in electromagnetic analysis to determine the absorbed power in a one-dimensional biological model which has been arranged into several distinct tissues layers with different intrinsic properties using transmission line analogy.

Outer Radiation Boundary Condition (ORBC): A condition that is needed to be satisfied at all the outside edges or surface of the computational space domain in which a set of differential equations is to be solved.

Ergonomic Design of a Driver Training Simulator for Rural India

Prabir Mukhopadhyay

Indian Institute of Information Technology Design and Manufacturing Jabalpur, India

Vipul Vinzuda

National Institute of Design Gandhinagar, India

INTRODUCTION

Hired cars (taxi) have become popular in the rural areas of Gujarat state in India for commuting, leading to an increased demand for taxi drivers. To improve the lifestyle of the rural population, the local government in Gujarat state of western India has taken an initiative. Their focus is to improve upon the transportation facilities for the people in these areas. The first step is to increase the number of hired cars (known as taxi locally) in different routes. This was a long standing demand of the people as buses and similar modes of public transportation are not popular in this part due to variety of reasons like high operational cost, absence of metallic roads etc. Thus there is an increase in demand of taxi drivers. To train the driver's new car driver training schools are opening up in these areas. This is serving two major purposes; first improving the connectivity between different places and second providing employment opportunity for the local youths. The youths are also keen on taking up driving as a mode of earning as it is lucrative when compared to agricultural work (which requires lot more physical labor) and many are supplementing their income from agricultural work by driving part-time. The local government has come up with schemes for providing soft loans to these youths for purchasing their own cars. The cars that ply in these areas are the standard cars manufactured indigenously in the country and the government rule does not

permit cars of any other manufacturers to ply in these routes.

As training and issue of licenses for these drivers necessitate the applicant to go through successive training on simulators followed by training on the roads, simulator training is becoming an important aspect of the driving schools. The simulators are mainly imported from the western world and are installed in the schools. Hence the novice drivers are trained on a simulator, which does not give them that confidence of driving on rural Indian roads, which are different from the western world. On repeated requests from the trainers, alumni of the driving schools, novice drivers, and other stake holders, ergonomic analyses of the existing simulators were done to reveal mismatches in the traffic conditions of the rural India and the western roads. There were various problems with the existing configuration related to comprehension (88%), readability (92%), navigation (89%), error messages (96%) and feedback menu (69%). Based on users feedback new simulators were designed.

BACKGROUND

Around 65% of the total population in India lives in the rural areas. Growth and prosperity of the country is largely dependent on these areas, with employment for the local youth becoming a challenge (Mukhopadhyay, 2006). It has been reported that (World Bank, 2000) 74% of the rural

population in India are still not integrated into the national economy due to lack of proper transportation facilities. Thus better transportation in rural India is a necessity as it gives the rural people an opportunity to travel for different purposes like health, education, social purposes and services (Ramaswamy, 1998). Thus focusing on rural local youths as taxi drivers.

It has been established that simulator training plays an important role in the training of a novice driver. For example it was found that while training car drivers on a simulator, the simulated environment was (Underwood et al., 2011) capable of increasing the hazard perception of drivers, thus indicating its criticality in driver training. There have been quite a few studies to indicate that the data from simulator training was very close to that of training in actual environment (Hallvig et al., 2013, Antonsen et al., 2014, Chan et al., 2010, Hallvig et al., 2013). For example in a comparative study (Hallvig et al., 2013) on sleepy driving on the real road and the simulator it was found that for both real and simulated driving the response to night driving appeared to be similar for subjective sleepiness and sleep physiology. In a study on the effect of landscape, different objects and landmarks on the driving behavior, it was found that the presence of any object on the road had a speed reducing effect which was in line with the simulator data. It was further established in the same study that different objects, land marks and landscapes affected the driver behavior in similar manner, both on the simulator and in the actual environment. Similarly another group (Antonsen et al., 2014) reported that open landscapes encouraged the drivers in actual environment to drive fast, and the drivers reported less stress compared to other conditions of the road. This was again very similar to the data obtained on the simulator. Another utility of the driving simulator is that it is useful (Chan et al., 2010) in evaluating certain driving attributes of the novice drivers like hazard anticipation, speed management, and attention maintenance. So these studies indicate that driving simulators are an essential tool in the training of

novice drivers. For the training to be effective, it's important that the simulated environment should be close to the actual driving environment of rural India which is different from the western world.

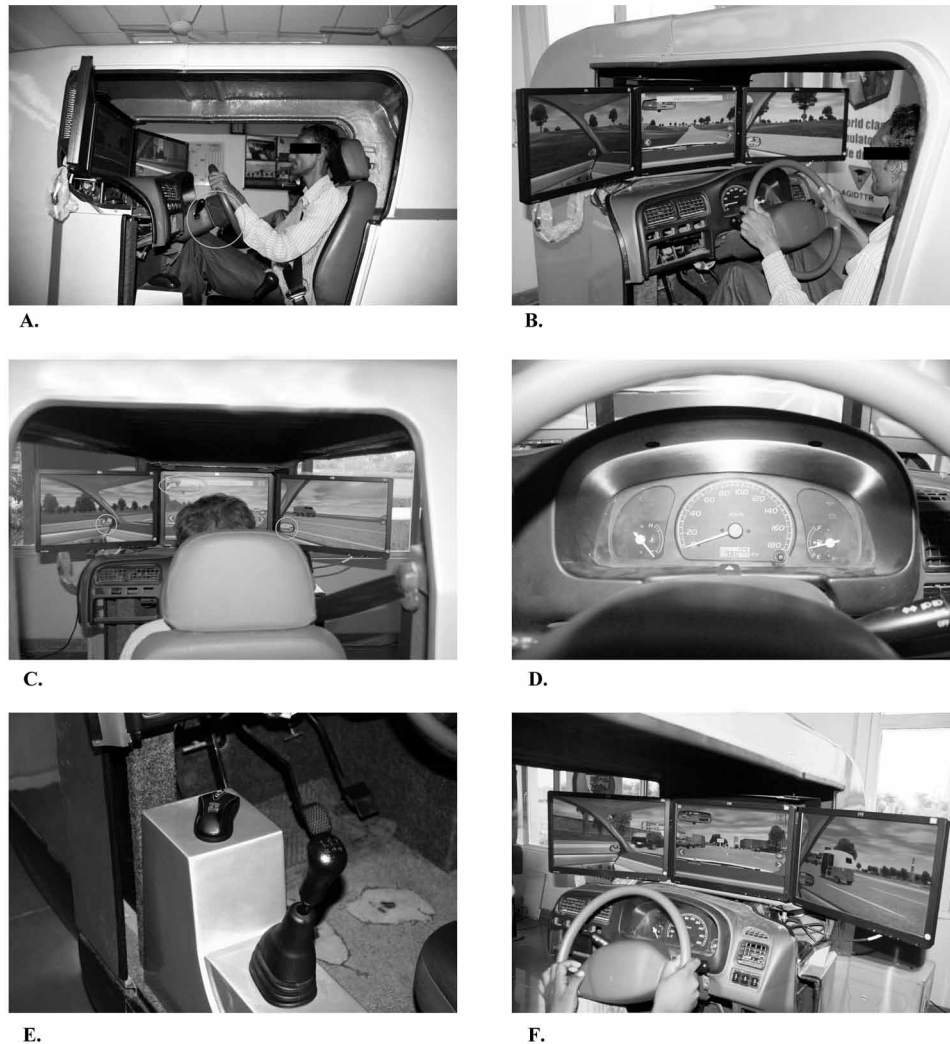
Research has indicated that there are hazards (Gurr et al., 1998) associated in transferring any technology without due considerations of the local culture, as this goes against ergonomic norms and leads to mismatch at the user interface which in turn leads to error. Safety and culture of a geographical region determines what specific problems pertaining to drivers training need to be addressed. So it is important to take feedback from the local end users about the exact features of the driving world which should be incorporated at the simulator interface, so that the novice drivers get a more realistic feel of the driving environment. This research was an attempt in that direction, to do an ergonomic analysis of the existing simulator interfaces and suggest suitable concept designs in tandem with the Indian traffic system in rural areas. Technology today is a traveller. It is conceived in one country, manufactured in a second country and implemented in a third country. Thus time has come for customization of the user interface which is the front end of any technology, in tandem with the target users of the target. This is not only pertinent in the Indian context but similar context where a foreign technology is often imposed as an implant in different context.

MAIN FOCUS OF THE ARTICLE

Issues, Controversies, Problems

Driving in rural India is a different experience than anywhere in the world. Majority of the roads are made of mud, traffic lanes are few, surprises on the roads like cows sitting, and kids playing are many. Apart from this there are issues of low illumination and waterlogged roads. Thus driving in rural India requires a different type of training and the same training simulator used to train prospective drivers in the west will not work in the

Figure 1. Existing simulator console



Indian rural scenario. The rural scenario is much more disorganized when compared to the western counterpart. The current research focused on the car simulator for the only brand of car which was currently used in this part of the country as public carriers. The emphasis was to train rural people and get them accustomed to the driving environment so that chances of mistakes and accidents are minimized when they train on real training cars. The simulator under study had two LCD screens (Figure 1A-B) which modelled the windshield of a car with a seating system along with gears and accelerators like that of a standard car. Participants were investigated while on the simulator so

that they are able to give actual feedback on the problems faced by them.

Study Design

Five different methodologies were applied. Direct observation and activity analysis along with still and videography recorded all the different activities of the participants at the existing simulator interface. Questionnaires were applied on the participants for the existing configuration. Scores were applied on the participants for the existing as well as the two concept designs. Subjective testing for usability of the interfaces for the existing

as well as the two concept designs was applied. Finally objective testing for performance, namely total time taken and the number of errors made for a set task was again applied on the participants for the existing and the two concept designs.

Participant Selection

Thirty-three right-handed male participants were selected for the study and had acquainted themselves with the simulator for the past fifteen days. The reason for fifteen days was the tenure of a course/training on a simulator; then they were trained on a training car. Only males were selected as in rural India females do not drive a taxi.

Qualitative Analysis

There were in total 28 observations, which were made by the research team. The entire team comprised of two members both trained in ergonomics. Each member was instructed to observe the activities of the driver from the beginning till the end. These were similar to other studies done to train drivers (Waard et al., 2008). The observational study started from the moment the driver boarded the simulator and the simulator started. All tasks performed by the driver, from the way he holds the handles to get inside the car, takes his seat, activates different controls, adjust the rear view mirrors were observed. There was special focus on the stages in which the participants became confused or made mistakes.

Questionnaire Study

A twenty questions survey questionnaire was administered to the participants to gain insight into the different types of problems encountered. It was ensured that other passengers and staff were not around, to ensure that people nearby did not bias the users. Some of the 20 questions were related to age, difficulty in accessing different controls, problem in reading and comprehending different

displays, difficulty in navigating under different adverse conditions of the road etc.

Interface Analysis

Different types of control and display elements on the simulator console were analyzed for usability issues such as placement, size and shape, graduations, color etc. to name a few. The ergonomic design issues for each of the interface elements were studied in detail. Rating scores were used to test the existing configuration and the concept designs. These scores were done on different sets of criteria based upon the feedback received from the trainers and taking lead from similar works reported in the literature (Stevenson et al., 2000). The first set of score criteria was “Comfortable to Use” or “Satisfactory”. The second set of score criteria was “Awkward” or “Very Uncomfortable”. The users were asked to rate the concept design on the basis of these two quantifiable questions. The percentage of total responses for different interface features was recorded. The third set of scores involved four levels in terms of perceived usability (Sade et al., 1998). These were “Good”, “Minor Problems”, “Major Problems”, “Failure”. “Good” indicated that participants did not have any problem at the interface. “Minor Problems” indicated that participants were occasionally confused which slowed down their performance. “Major Problems” indicated that participants encountered serious problems that slowed down their performance considerably. They were confused with the basic function of the concept design. “Failed,” indicated that participants were not able to operate. These scores indicated the total number of participants who agreed to these four levels in terms of perceived usability. A participant was allowed to give multiple responses.

Performance Analysis

Two dependent variables: total time taken and the number of errors made, were selected for objectively measuring the performance of the existing

Table 1. Demographics of the participants (n=33)

	Age (Years)	Experience (Months)
Mean	30.4	2.1
SD	7.9	0.7
Range	21-47	1-3

and the two concept designs. Participants were given a set task and were asked to perform it on the existing and the concept design interfaces. The total time taken and the number of errors made in performing the tasks were recorded.

Participants

The mean age of the participants (Table 1) was 30.4 years (± 7.9) with age ranging from 21 to 47 years old. The mean experience of on road driving was 2.1 months (± 0.7) with experience ranging from 1 to 3 months.

RESULTS AND DISCUSSIONS

The driving simulator had two major components. The physical component and the software user interface component. The later was imported from the west. The integration of the hardware and software again took place at the local vendors place. The physical component consisted of a seat without any doors and a dash-board (Figure 1B). There were three LCD screens in front where the simulation was displayed (Figure 1C). There was no haptic feedback from the gears, breaks or pedals of the car. In fact, the seat-belt which teaches the first safety aspects of driving was without any feedback when it was attached (Figure 1D). This was necessary as all the cars in this part of the country, which are used as taxi, have a feedback for seat belt on the dashboard. There was complete absence of any surrounding sounds to create the ambience of traffic or driving on a typical Indian road.

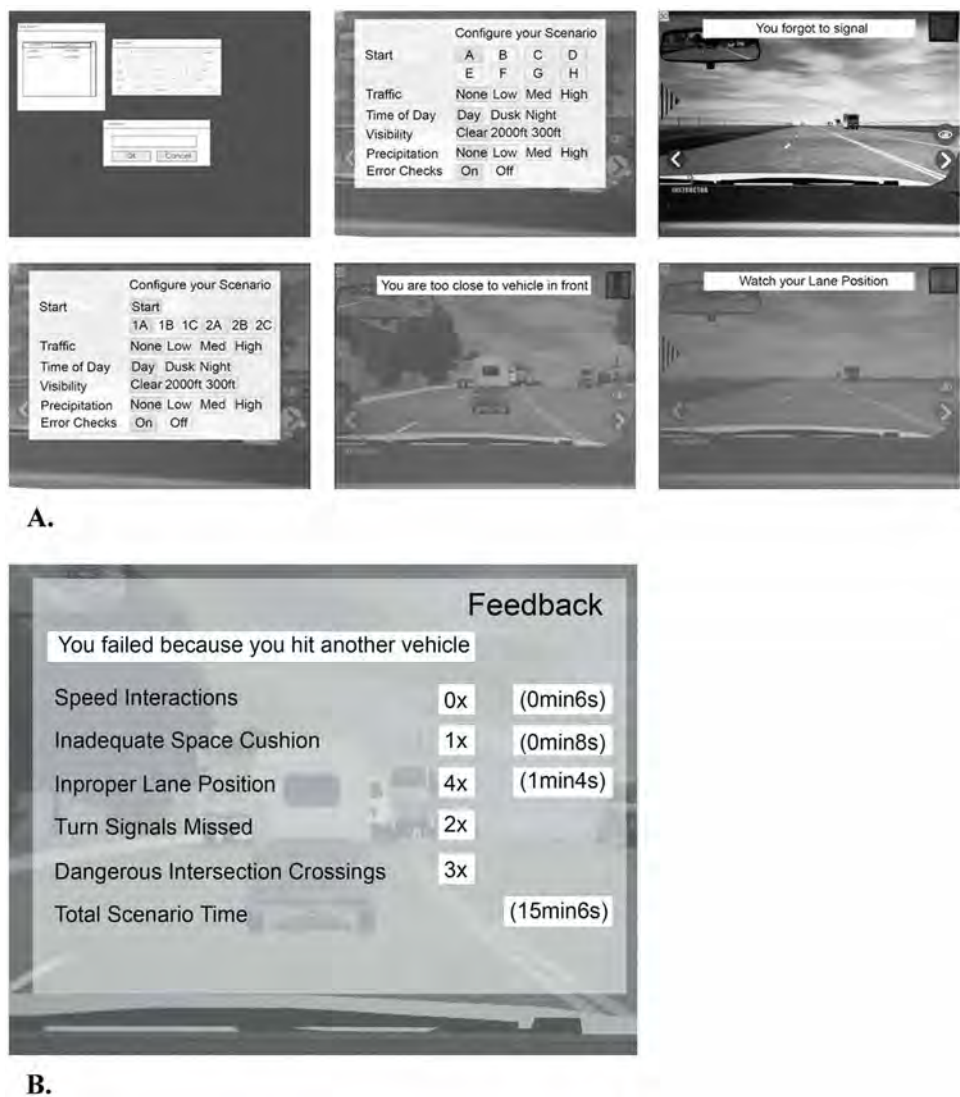
The interface of the simulator could be divided into two categories: the first one was the login and subsequent screens and the second one comprised of the main menu which was further made up of six sub- menus (Figure 2) in different permutations and combinations of the driving environment (Figure 3). After a successful login, users were able to go to the main menu, which contained six driving modes (Figure 3). The first one was day- time but with low illumination level. The second one was with precipitation of different types ranging from low to high. The third one was with hazards on the road. There were onscreen indicators with arrows and text messages indicating the position of the vehicle on the road as to whether it was in the appropriate lane and whether at a safe distance from other vehicles. The participant had to log in after their session was over to check their scores and then close the programme. The main problem with the interface was that it never mimicked the Indian road conditions and was representation of the western world.

The questionnaire study on the problems at the interface (Figure 4) revealed five major elements. 88% of the participants had problems in comprehending the information conveyed by the display in general. 92% of the participants were not able to read many of the pieces of information presented to them. There was difficulty in navigating through the menu structures reported by 89% of the participants. Error messages were not comprehensible (96%) and so was the feedback menu for 69% of the participants.

The overhead and rear view mirrors (virtual) were small in size, and placed outside the driver's visual cone thus rendering them virtually useless while driving the simulator (Figure 1C). The speedometer was not properly color coded for different speeds, thus it gave the driver no clue as to the safe driving speed (Figure 1D). The overall look and feel of this simulator interface was more like that of a video game and not at all similar to the Indian road conditions. On the roads few Indian- made vehicles were shown, but the roads were representation of the western world. The

Figure 2. Existing user interface of the simulator

C

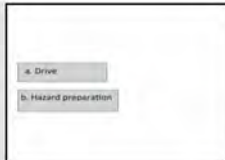
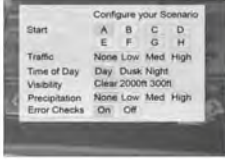
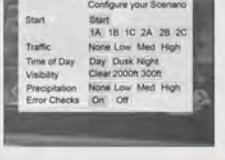
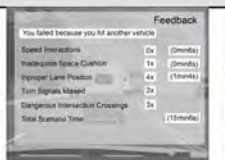


interface lacked “Indian” elements on the roads like potholes, grazing animals, moving traffic in wrong direction, people crossing, or vehicles changing lanes without warning. Another major drawback of the interface was the driver’s feedback system. There was no facility for taking a print-out of the feedback which could be referred to later for improving performance. In fact the manner in which the feedback was given to the users was not comprehensible to many because it received few satisfaction scores from the users. The users expected the mistakes in terms of percentage of the total navigation rather than in absolute num-

bers. A complete comparative analysis of all the features of the existing and the concept designs (1 and 2) are represented in Figure 3.

The main features of this concept (concept design 1) were that (Table 2) the mouse platform was removed to ensure free movement of the left hand when operating the gears. The instrument clusters were redesigned with a contrasting color (Figure 5). The main menu configuration was provided in the console of the dashboard itself (Figure 6). For the interface part (Figure 7) the selection menu was split up into discrete steps with images to facilitate the ease of selection.

Figure 3. Comparison between the existing and the two design concepts of the simulator interface

Description	Existing	Concept 1	Concept 2
Home page / Log in			
Driving mode			
Configuration Selection Menu 1			
Configuration Selection Menu 2			
Driving scenario 1			
Driving scenario 2			
Feedback menu			

The different menu options like mode and traffic, visibility and time of the day were combined and chunked together based on their relevance. The participants were guided by a blinking cursor for moving to the next option, before moving to the

next screen. When an option was selected it was highlighted by a bright green color. The “next” button was enabled only when all the options on the screen were selected.

Figure 4. Percentage of participants that mentioned a problem for different ergonomic issues at the existing interface

Note: As there were multiple responses the total does not add up to 100.

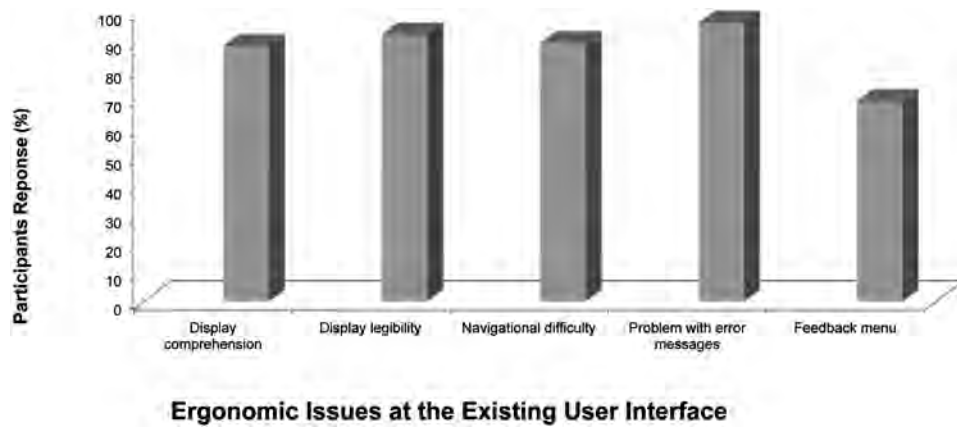


Figure 5. Comparison between the existing and the proposed design concept of the instrument panel



Table 2. Ergonomic design features of design concept 1 of the simulator console panel

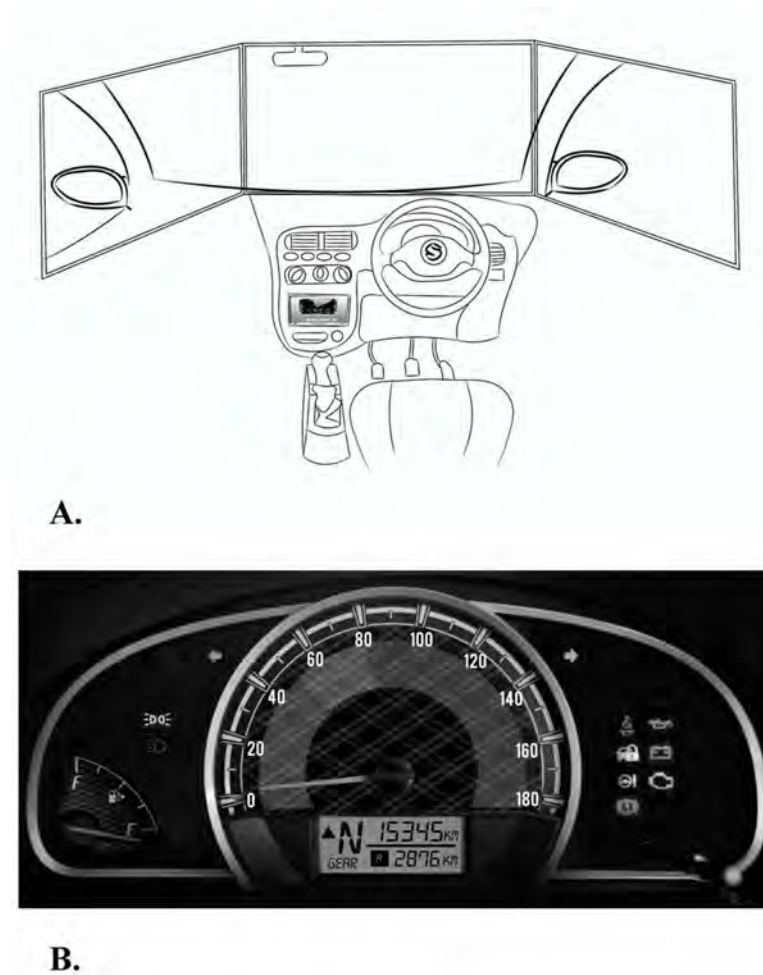
Features
• Using the existing cab arrangement with ergonomic updates • Bigger screen with better orientation • Mouse platform is removed • Main menu configuration is provided in the console of the dashboard • Redesigned instrument cluster • Cluster- includes gear shift and seat indicators and colour zones • Trip meter • Steering is provided with tilt and height adjustments
Salient Features
• Using existing cab- cost is minimal • Using the center console already available in cab • Bigger screens • Height adjustment steering will provide better viewing angle

SOLUTIONS AND RECOMMENDATIONS

After analyzing the problems two concepts were developed. The main features of the first concept (concept design 1) were that (Table 2) the mouse platform was removed to ensure free movement of the left hand when operating the gears. The main menu configuration was provided in the console of the dashboard itself. The instrument clusters were redesigned with a contrasting color (Figure 5 and Figure 6). For the interface part (Figure 7) the selection menu was split up into discrete steps with images to facilitate the ease of selection.

For the physical part the dashboard (concept design 2) was recreated (Figure 8 & 9) and Table 3. Two seats were provided like a real training car. A bonnet was fabricated to train the participants

Figure 6. Proposed layout of the console and the instrument cluster: design concept 1



on making a judgment on the road. This was essential as this category of taxis has a bonnet, which comes in the visual field of the drivers with shorter torso. For this concept (design 2), (Table 3) similar scenarios were developed. The screen here consisted of a timer to indicate how much time the user has spent on the simulator, along with a “pause” and a “home button”. Here also fifteen feet is considered a safe distance between two vehicles on the road. Unlike the previous concept design, there was no change in color to indicate whether the vehicle is coming in proximity to another vehicle on the road. Instead an icon at the center was provided based on the expected quadrant of the visual field, with text indicating whether one was at a safe distance, needs to keep a safe distance or is too close to a vehicle.

Table 3. Concept-2 ergonomic design features of the simulator console pane

Features
• Recreating the car dashboard • Two seats- instructor sits along with the driver in a closed cabin arrangement • Bonnet mock-up is provided to get a sense of stopping and starting • Single curved screen is used- avoiding distortion of imagery • Passenger side glove box modified to hold a keyboard and mouse • Steering provided with tilt and height adjustment
Salient Features
• Instructor can also sit inside the cab at all times • Provision of mouse and keyboard • Curved screen- better viewing angle • Printed feedback form- better assessment

Figure 7. Proposed interface of design concept 1

C



Figure 8. Proposed layout of design concept 2(exterior) and the layout of the console (below) with the design of the instrument cluster remaining the same as in design concept 1

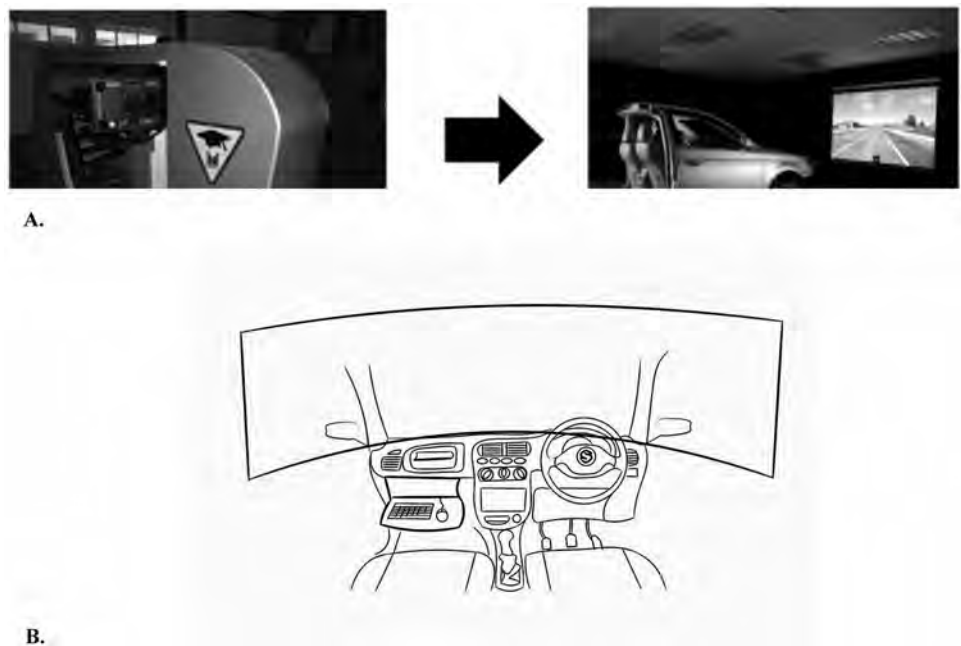
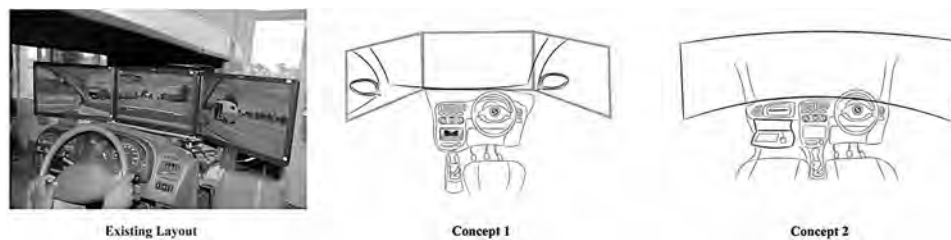


Figure 9. Comparison between the existing layout of the simulator console with the two new design concepts for the console layout



As the user study revealed the need for contextualizing the interface to Indian conditions, different scenarios were developed (Figure 9) for this purpose. Six different types of scenarios were framed, and on these scenarios (Figure 10) different types of vehicles were placed. The six major scenarios were city (representing any modern Indian city) with wide roads but haphazard placement of billboards and with traffic rules being violated by other vehicles. The second scenario was a typical non-metallic mud road as found in rural and semi-urban India. The third scenario depicted a traffic signal on an asphalt road popularly known as metallic road (this is where maximum violation of traffic rules happens, and maximum accidents take place). The fourth scenario depicted surprises such as stray animals on the road, which are very common features in the Indian roadway context. The fifth scenario depicted roundabouts, which in India are again an extremely vulnerable zone for road accidents including collision and overturning of other vehicles like three-wheelers. The sixth scenario depicted vehicles being driven on the wrong side of the road which is extremely important, as a good driver needs to be extra vigilant to prevent casualties on the road. A separate driver's feedback form was developed and the users could carry a printout of the same for their reference (Figure 11). The information (number of errors for a set task) was structured as per users feedback with the mistakes represented in percentage of the total travel time. Driving in rural India entails travelling through stretches of semi-urban areas as well, thus such areas were

deliberately included in the concept designs to make it more realistic.

The users rated different interface elements in the new designs for the two layouts in comparison to the existing layout (Figure 2). The scores for "Comfortable to Use" or "Satisfactory" for "simulator loading" were higher (Figure 12) for concept design 1(88%) and concept design 2(92%) compared to the existing configuration (03%). The scores for "password log in" were again higher for concept design 1(92%) and concept design 2(98%) compared to the existing one (01%). The scores for "main menu" for concept design 1(87%) and concept design 2 (91%) showed similar trend as well. For "driver menu" concept design 1(89%) and concept design 2(88%) again scored over the existing configuration (12%). "Hazard preparation" for concept design 1(91%) and concept design 2(94%) was higher than the existing one (4%). The other element like "on screen error messages" for concept design 1(89%) and concept design 2(90%) was again more than existing (3%). Similarly for "feedback menu" it was higher for concept design 1(92%) and concept design 2(94%) compared to the existing one (9%). To compare the scores between the existing interface and the newly developed concept designs 1 and 2, a One-Way Analysis of Variance (ANOVA) was applied on the scores for "Comfortable to Use" or "Satisfactory" for the existing configuration and concept designs 1 and 2. The scores were significantly different from one another ($p < 0.001$). To differentiate between the scores for different designs, Student-Newman-Keul's tests were performed on the existing configuration as well

Figure 10. Proposed interface of design concept 2



A.



B.

Figure 11. Comparison between the online feedback form and newly designed form in print format for drivers

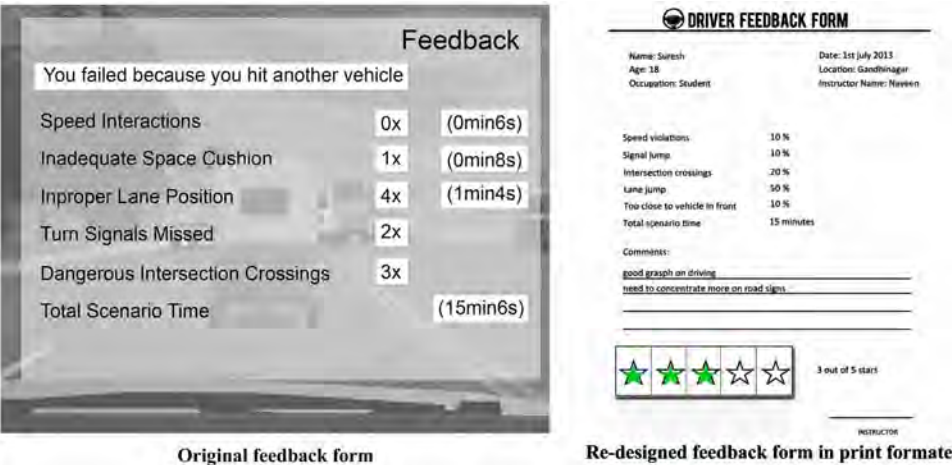
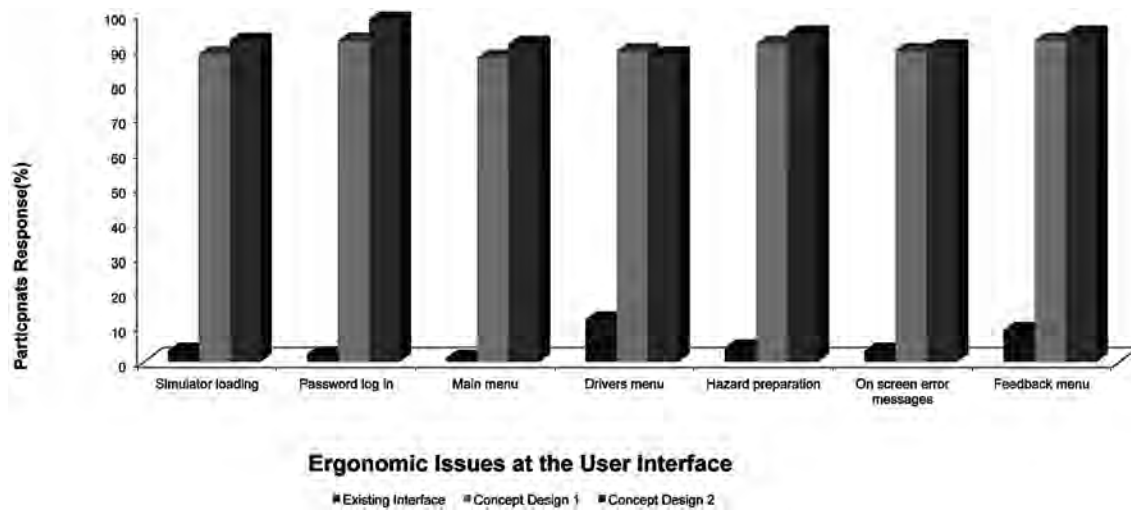


Figure 12. Participants (%) rating the existing and the concept design interfaces as “Comfortable to Use” or “Satisfactory”

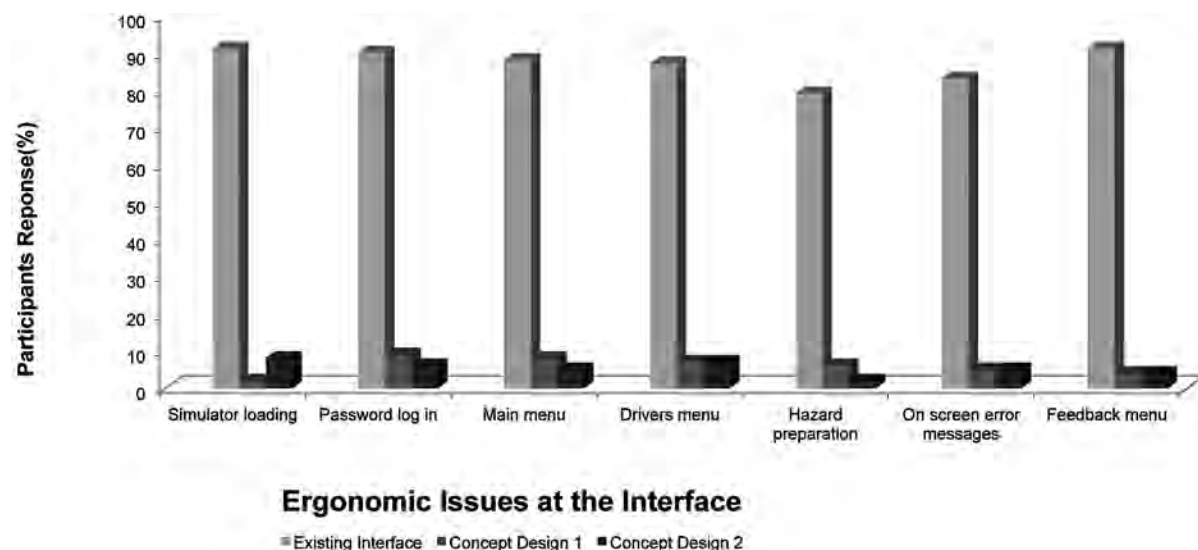


as concept designs 1 and 2. There was no difference in scores between concept designs 1 and 2, but for the existing configuration the scores were significantly ($p < 0.05$) lower than concept designs 1 and 2.

The scores for “Awkward” or “Very Uncomfortable” (Figure 13) for “simulator loading” was at 91% for the existing configuration, but decreased to 2% for concept design 1 and 8% for concept

design 2. For “password log in” the score for the existing configuration was at 90% whereas for concept design 1 it was 9% and for concept design 2 it was 6%. Along similar lines for “main menu” the score for “existing configuration” was at 88% which declined to 8% for concept design 1 and 5% for concept design 2. The scores for “driver menu” for the existing configuration was at 87% but declined to 7% for concept design 1 and 7% for

Figure 13. Participants (%) rating the existing and the concept design interfaces as “Awkward” or “Very Uncomfortable”



concept design 2. A similar trend was observed for “hazard preparation”. For the existing configuration, the scores were at 79% which declined to 6% for concept design 1 and 2% for concept design 2. For “on screen error messages” the scores for existing configuration were at 83% and for concept design 1 at 5% and concept design 2 at 5% again. The “feedback menu” followed similar trends. It was 91% for existing configuration and 4% for concept design 1 and 4% for concept design 2.

A One-Way Analysis of Variance (ANOVA) was applied on the scores for “Awkward” or “Very Uncomfortable” for the existing configuration and concept design designs 1 and 2. The scores were significantly different from one another ($p < 0.001$). To differentiate between the scores for different designs, Student-Newman-Keul’s tests were performed on the existing configuration as well as concept designs 1 and 2. There was no difference in scores between concept designs 1 and 2, but for the existing configuration the scores were significantly ($p < 0.05$) higher than concept designs 1 and 2.

For the existing (Table 4) configuration the scores for “Good” and “Minor Problems” were at 0. For “Major Problems” it was at 11 and for “Failure” it was at 5. For concept design 1 the scores for “Good” was at 12, “Minor Problems” and “Major Problem” at 2. For “Failure” it was at 0. Similarly for concept design 2 the score for “Good” was at 14, “Minor Problems” and “Major Problems” it was at 1. For “Failure” the score was at 0.

Table 4. Performance comparison with the existing and the design concept interfaces. The numbers below indicate the number of users who responded.

	Existing Interface	Concept 1	Concept 2
Good	0	12	14
Minor problems	0	2	1
Major problems	11	2	1
Failure	5	0	0

N.B.: As there were multiple responses the total does not add up to 33 in many cases, 33 being the total sample size. Zero indicates that subjects did not comment on that particular level of perceived usability.

A measure of performance was also conducted on the existing configuration and the two concept designs. Two dependent variables were selected, one was the total time taken (minutes) to perform a set task and the second was the number of errors made in performing the task. The mean time taken for the existing configuration (Table 5) was 77.9 minutes ($SD=9.1$). For concept design 1 it was 64.4($SD=15.6$) and for concept design 2 it was 75.6($SD=15.8$). A One-Way Analysis of Variance (ANOVA) was applied on the mean time for the existing and the two new concept designs. The mean performance time for the existing configuration, concept design 1 and concept design 2 were significantly different from one another ($p < 0.024$). A Post Hoc Students-Newman Keuls test revealed that the scores for the original configuration were significantly higher than concept designs 1 and 2

Table 5. Total time taken and number of errors made at the existing and the design concept interfaces for performing a set task

	Existing Interface	Concept 1	Concept 2
Time in Minutes			
Mean (SD)	77.9(9.1)	64.4(15.6)	75.6(15.8)
Range	66-91	33-99	40-94
Error (Number)			
Mean (SD)	19.4(6.4)	5.1(2.2)	8.1(1.6)
Range	11-40	2-9	6-11

(at $p < 0.05$) with the existing configuration taking the maximum time followed by concept design 2 and then concept design 1. The second measure of performance was the number of errors made. These errors were defined on the basis of the number of mistakes the participants made as discussed in the methods section. For the existing configuration (Table 9) the mean errors were at 19.4(SD=6.4). For concept design 1 it was at 5.1(SD=2.2) and for concept design 2 it was at 8.1(SD=1.6). A One-Way Analysis of Variance (ANOVA) was applied on the error rate for the existing and the two new concept designs. The mean error rate for the existing configuration, concept design 1 and 2 were significantly different from one another ($p < 0.001$). A Post Hoc Students-Newman Keuls test revealed that the error rates for the original design and the two concept designs 1 and 2 were all significantly different from one another (at $p < 0.05$), with the existing configuration yielding the highest number of errors followed by concept design 2, then concept design 1.

FUTURE RESEARCH DIRECTIONS

Simulator of similar nature could also be designed for training the novice drivers in other domains of public transportation in India like motor boat, tram, three wheelers etc. so that they are confident before they start driving the real vehicle. The present study could also be extended for screening of older drivers for these types of vehicles. It has been established that the existing screening system for older drivers (medical test, road test, cognitive screening) is unreliable (Casutt et al., 2014). Thus further work on this type of simulators could be directed in this direction so as to ensure that drivers who get old are capable of driving safely.

This research could be extended to other domains like medicine, aerospace, communication etc. where many a times the technology is

imported from a different country and is used by the target users in a different country without any customization of the user interface in tandem with the local users and the context. Such research into identifying mismatches at the user interfaces would be interesting to probe into in terms of the exact location, nature and quantum of the ergonomic or human factors problems.

CONCLUSION

The participants accepted the new concept designs of the car simulator in the Indian context in principle. There were some minor problems related to information design, which were auditory feedback (which needs further development as it was very soft), road signage (which needs refinement considering night time driving and driving during heavy rainfall) for which work is ongoing. The simulator and if it meets preference standards has to be now developed as a working concept design and then tested on participants once again to see how it functions.

The research carried out here depicts the importance of humanizing technology in tandem with the target users and the local context in which it is to be used. In this era of globalization, when technology is conceived, developed and finally used by the target users in different geographical locations of the world, it becomes extremely important to customize the technology for the local context in order to reap the maximum benefit out of it. The study on simulator in this paper was an example in that direction, pointing to the fact that each technology developed in remote locations needs customization for the target users.

ACKNOWLEDGMENT

Acknowledgement is owed to all those who volunteered for this study.

REFERENCES

- Antonsen, H., Ahlstrom, C., Mardh, S., Blomqvist, G., & Wiklund, M. (2014). Landscape heritage objects effect on driving: A combined driving simulator and questionnaire. *Accident; Analysis and Prevention*, 62(1), 168–177. doi:10.1016/j.aap.2013.09.021 PMID:24172083
- Antonson, H., Mardh, S., Wiklund, M., & Blomqvist, G. (2009). Effect of surrounding landscape on driving behavior: A driving simulator study. *Journal of Environmental Psychology*, 29(4), 493–502. doi:10.1016/j.jenvp.2009.03.005
- Casutt, G., Martin, M., Keller, M., & Jancke, L. (2014). The relation between performance in on-road driving, cognitive screening and driving simulator in older healthy drivers. *Transportation Research Part F: Traffic Psychology and Behaviour*, 22(1), 232–244. doi:10.1016/j.trf.2013.12.007
- Chan, E., Pradhan, A. K., Pollatsek, A., Knodler, M. A., & Fisher, D. L. (2010). Are driving simulators effective tools for evaluating novice drivers hazard anticipation, speed management and attention maintenance skills? *Transportation Research Part F: Traffic Psychology and Behaviour*, 13(5), 343–353. doi:10.1016/j.trf.2010.04.001 PMID:20729986
- Gurr, K., Straker, L., & Moore, P. (1998). Cultural hazards in the transfer of ergonomics technology. *International Journal of Industrial Ergonomics*, 22(3), 397–404. doi:10.1016/S0169-8141(97)00094-2
- Hallvig, D., Anund, A., Fors, C., Kecklund, G., Karlsson, J. G., Wahde, M., & Akerstedt, T. (2013). Sleepy driving on the real road and in the simulator - A comparison. *Accident; Analysis and Prevention*, 50(1), 44–50. doi:10.1016/j.aap.2012.09.033 PMID:23149323
- Kantowitz, B. H. (2011). Using driving simulators outside of North America. In J. K. Caird & J. D. Lee (Eds.), *Handbook of Driving Simulation for Engineering, Medicine and Psychology*. Boca Raton, FL: CRC Press/Taylor and Francis. doi:10.1201/b10836-4
- Mukhopadhyay, P. (2006). Global ergonomics: Looking beyond the horizon. *Ergonomics in Design*, 4(2), 4–6. doi:10.1177/106480460601400302
- Ramaswamy, N. S. (1998). Rural transport in India. *Current Science*, 75(1), 800–815.
- Sade, S., Neminen, M., & Riihio, S. (1998). Testing usability with 3D paper concept design-case hilton system. *Applied Ergonomics*, 29(1), 67–73. doi:10.1016/S0003-6870(97)00027-6 PMID:9769091
- Sleepy driving on the real road and in the simulator: a comparison. (n.d.). *Accident Analysis Preview*, 50(2), 44–50.
- Stevenson, M. G., Coleman, N., Long, A. F., & Williamson, A. M. (2000). Assessment, re-design and evaluation of changes to the drivers cab in a suburban electric train. *Applied Ergonomics*, 31(3), 499–566. doi:10.1016/S0003-6870(00)00014-4 PMID:11059463
- The World Bank. (2000). Rural road a lifeline of villages in India: Connecting hinterland to social services and markets. *Technical Representative. World Bank in India*, 28(1), 1–44.
- Underwood, G., Crundall, P., & Chapman, P. (2011). Driving simulator validation with hazard perception. *Transportation Research Part F: Traffic Psychology and Behaviour*, 14(6), 23–28. doi:10.1016/j.trf.2011.04.008
- Waard, D., Kruizinga, A., & Brookhuis, K. A. (2008). The consequences of an increase in heavy goods vehicle for passenger car drivers mental workload and behavior: A simulator study. *Accident; Analysis and Prevention*, 40(2), 818–828. doi:10.1016/j.aap.2007.09.029 PMID:18329437

ADDITIONAL READING

Bridger, R.S. (2009). Introduction to ergonomics. CRC Press, Taylor and Francis Group.

Garner, S. (1991). *Human factors*. Oxford University Press.

Woodson, W.E., Tillman, P., & Tillman, B. (1991). *Human factors design handbook*. Mc Graw-Hill Publishing Company.

KEY TERMS AND DEFINITIONS

Feedback: The output from the machine to the user which makes the user at ease with the machine he is operating at.

Interface: It is the connecting link between the user and the technology with which he is interacting.

Mental Model: It is the image in our brain of what we see in real life. This image persists even after a long time.

Participants: They are the people who participate in the experiments.

Scenario: This is the artificial creation of a particular situation in order to test the robustness of any design meant for the users.

Simulator: A device which gives similar feedback to the users like the original product so that novice users get an experience of how the actual product works.

Usability: It is the ease with which a product can be used by the users for fulfilling their specific needs.

3D Scanning and Simulation of a Hybrid Refrigerator Using Photovoltaic Energy

C

Edith Obregón Morales*Centro de Investigación y Desarrollo Tecnológico en Electroquímica, Mexico***José de Jesús Pérez Bueno***Centro de Investigación y Desarrollo Tecnológico en Electroquímica, Mexico***Juan Carlos Moctezuma Esparza***Universidad Politécnica Metropolitana de Hidalgo, Mexico***Diego Marroquín García***Universidad Tecnológica de San Juan del Río, Mexico***Arturo Trejo Pérez***Universidad Tecnológica de San Juan del Río, Mexico***Roberto Carlos Flores Romero***Universidad Tecnológica de San Juan del Río, Mexico***Juan Manuel Olivares Ramírez***Universidad Tecnológica de San Juan del Río, Mexico***Maria Luisa Mendoza López***Tecnológico Nacional de México, Instituto Tecnológico de Querétaro, Mexico***Juan Carlos Solís Ulloa***Instituto Tecnológico Superior de Cintalapa, Mexico***Yunny Meas Vong***Centro de Investigación y Desarrollo Tecnológico en Electroquímica, Mexico***Víctor Hugo Rodríguez Obregón***Centro de Investigación y Desarrollo Tecnológico en Electroquímica, Mexico*

INTRODUCTION

In this chapter, a methodology that starts with the measurement and recording of real prototype geometries up to simulations to evaluate parameters, improvements or performance under various conditions is proposed. Here it is presented a case study of a solar powered refrigerator with storage capacity for 50 kg of fruit. The refrigerator comprises two systems, vapor-compression and

Peltier. The methodology consisted in acquiring by a 3D laser scanner or Coordinate Measuring Machine (CMM) and in some small complex items using a 3D photogrammetry scanner (James, 2017; Ozdarici-Ok, 2015). These data were transferred first as a CAD or SolidWorks® geometry and subsequently transferred to domains geometry useful for ANSYS or COMSOL simulation software. These models with high-resolution brings the simulations closer to real prototypes. As a

DOI: 10.4018/978-1-5225-2255-3.ch110

source of direct information from the prototypes, thermal images obtained using a thermographic camera were taken. Also, wireless sensors were installed for temperature and humidity monitoring. The analyses of the energy efficiencies of both prototypes were performed. The idea of presenting this methodology is to demonstrate that it is possible to apply it in various prototypes regardless of sizes, geometries or whether they are or not solar powered.

The work seeks to develop a hybrid cooling system based on a system of traditional compression system and a Peltier cooling system. The construction or production of this system is focused on being able to assist the agribusiness sector that is away from the public electric network, directly in the crop areas, proposing that this system will be powered completely by photovoltaic modules and in turn, be able to innovate with the cooling systems adaptations.

The starting point was a commercial compression system, to build or make an approach of the required refrigerator with a capacity of 50 kilograms. The material used in the Peltier cooling system consisted of: four Peltier modules of 50 W, eight heat sinks and eight fans that assist with the distribution of heat within the refrigerator.

The following steps were the elaboration of the construction drawings, drilling and placing of the Peltier systems, in a distribution geometry designed in the SolidWorks® software. All the parts were put in place both physical and mechanical. Some of them were affected during the modification process. The geometry was acquired and exported for its analysis in a virtual simulation module of finite element (Work Bench ANSYS), to obtain a validation or improvement of the system with the efficiency of space and location, considering the fluid dynamics and thermodynamics.

A general scanning was conducted for a commercial refrigerator with a laser system (ZX 800 Corp) and, for the smaller parts, with a system of photogrammetry (PicoScan Pro). The scannings

were conducted using these devices for the development of a virtual 3D modeling. In this process, a polygonal mesh was applied to the virtual model initially constituted only of a point cloud (Bici, 2014), which is transformed into a file compatible with the simulation software.

Once that the process with the commercial refrigerator model reaches the simulation stage, the methodology is applicable to the construction and adaptation of the Peltier system using as a guide the electrical drawings previously made. To achieve a relationship between the Peltier system and the compression system uses a temperature sensor programmed to monitor the functioning of these systems on the basis of specific conditions known during the simulation process.

BACKGROUND

Design and Computer Aided Engineering

Design is an activity that is projected toward the solution of problems in the process of adaptation to the environment that surrounds it, for such purpose resources as the technology CAD/CAE are used. The most developed technique in Computer Aided Engineering (CAE) is the application of Finite Element Analysis (FEA) that, with the improvement of computers has become more accessible to all users. These techniques are used industrially from design to manufacturing getting optimize costs, quality, time, security, etc.

The implementation of the CAD software in engineering includes the development of synoptic charts, diverse kinds of diagrams, statistical graphics, standard representation of parts for its design and construction, tridimensional representation of dynamic models in multimedia, finite element analysis, applications in virtual reality, robotics, etc.

Figure 1. Puller and refrigerator motor



Table 1. Characteristics of the refrigerator

Refrigerant	TYPE R-134 a
Rated Voltage	115V
Rated Frequency	60 Hz
Current Rating	2.7 A

Cooling Powered by Photovoltaic Panels

At present, the cooling in the households represents more than 45% of the total electricity consumption, which is why on the basis of this challenge, solutions were found that represent a technological breakthrough where renewable energy acquired a very important role.

The challenges associated with this work is the distance that exists between the electrical distribution network and the crops of fruits. Therefore, the solution is the implementation of a photovoltaic system that meets the energy demand of the refrigerator. The refrigerator will be a field, for that reason, production losses are minimal during the fruit collect (as a model, guava was chosen).

This work deals with a refrigerator LINE HI SPEC (VR09) with the features shown in Table 1.

The cooling system is commercial, with a refrigerant *Type R-134 a* and a power of 310.5 W at 2.7A (Figure 1).

The solar panels for the power supply of the refrigerator had cells of 250 W at 24 V being monocrystalline panels with an area of 1.6 (Figure 2; Table 2).

Calculations and Batteries

To get the calculations, it is possible to start from the power demand that was required and the time in which it was engaged, multiplied by a factor of safety (Table 3).

There was a total of 2,916 W hr multiplied by the factor of safety: $2,916 \text{ W hr} (0.20) = 3,499.2 \text{ W hr}$. The batteries work at 12 V and the design implies several batteries in parallel to provide the required current.

Table 2. Features of the panel

Operation Voltage	24 V
Open Circuit Voltage	22 V
Current Rating	8 A

Figure 2. Solar panel 250W at 12V

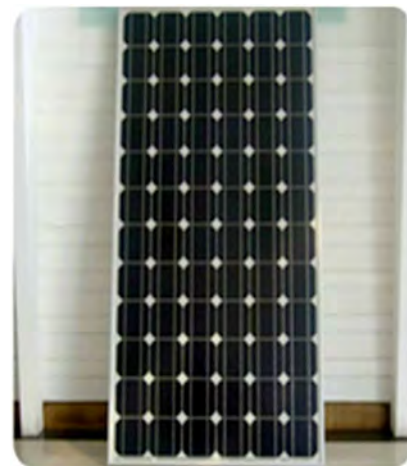


Table 3. Characteristics of the batteries

Article	Consumption (W)	Time of Use (hr)	Daily Consumption (W hr)
Refrigerator	324	9	2916

$$\text{Amperes - hour/day} = \frac{\text{Watts} - \text{hr} / \text{day}}{\text{Volts}(\text{under})} =$$

$$\frac{3499.2}{24} = 145.8 \text{ Amp-hr day}$$

This is multiplied by the number of days of autonomy and divided between the depth of discharge:

$$\text{Amperes - hour/day} = \frac{\text{Amperes} - \text{hours} / \text{day} * \text{No autonomy days}}{\text{depth of discharge}}$$

$$= \frac{145.8 \frac{\text{Amperes}}{\text{hours}} * 3}{0.4} = 1093.5 \text{ Amp-hrs/day}$$

To determine the number of batteries, a capacity is selected in the A-H manufacturer catalog. In this case 12v, 115 A-H:

$$\text{Not batteries} = \frac{1092.5 \text{ Amperes} - \text{hour} / \text{day}}{115 \text{ Amperes} - \text{hour} / \text{batteries}}$$

It is considered the installation in the state of Queretaro, México, latitude 20.6 °N and longitude 100.3 °W, and the worst conditions are with a solar peak at December of 4.4 hr/Day, as shown in Table 4.

Table 4. hr/day solar peak in México

State	City	Jan	Feb	Mar	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec	Max.	Min.	Avg.
Qro.	Queretaro	5	5.7	6.4	6.8	6.9	6.4	6	6.4	6.3	5.4	5	4.4	6.9	4.4	5.9

Peak current of the arrangement = (Amp-hr /Day average)

$$/ (\text{battery } \eta) / (\text{hr/Day peak solar}) = 145.8 \text{ Amp-hr /Day} / 0.8 / 4.2$$

$$= 40.5 \text{ A}$$

The number of modules in parallel is:

$$\text{Number of modules in parallel} = \frac{\text{peak current of the arrangement}}{\text{peak current module}} \text{ FS}$$

FS = safety factor that takes into account the aging and dirt. It is taken as 1.2

$$\text{No. of modules in parallel} = 1.2 * (40.5 \text{ A} / 8 \text{ (A)}) = 5 \text{ modules in parallel}$$

$$\text{No. of modules in series} = \frac{\text{system voltage}}{\text{nominal module voltage}} = \frac{24}{12} = 2$$

No. total modules = No. modules in series * No. modules in parallel

$$= 2 * 5 = 10 \text{ modules}$$

The controller, with specified rated current whose value is given.

Short circuit current of the system: circuit current module * number of modules in parallel * FS

$$\text{The short circuit current of the system} = (24 \text{ A}) (5) (1.25) = 150 \text{ A}$$

In the following list, it can be observed the actions carried out within the process of analysis of the structure to be able to reach the required data.

- Sizing and assessment of the prototype (refrigerator).
- Development and construction of the prototype in SolidWorks®.
- Assembly and installation of the prototype, taking into account Peltier cells.
- Scanning of the prototype through photogrammetry (Peak Scan Pro) for small parts.
- Scanning of the prototype using the laser system (ZX 800 Corp) for large parts.
- Development of reverse engineering (Bici, 2014; James, 2017; Ziadi, 2014; Lopez Herrejon, 2015; Thomopoulos, 2015; Ozdarici-Ok, 2014; Janik, 2014).
- Flow Simulation of the obtained solid.

In this section, you will be studying the hoped-for outcome and the way in which the project can be developed in a way that is efficient; it is important to have this point already that depart from here and are born all the ideas of optimization and improvement of processes.

The average of the prototype was obtained to perform the construction of the same, it is necessary that the measures were as accurate as possible since it is designed to be the closest to reality, to obtain the results with a minimum error rate.

Evaluation and Construction of the Prototype in SolidWorks®

In this part, this work presents a proposal, which is based on design portions of the refrigerator separately. One of the qualities of SolidWorks® is that when working separately (individual design of the parts), once drawn all the parts, they can be joined to form a single solid, which can be simulated without any problem.

Construction of the Refrigerator

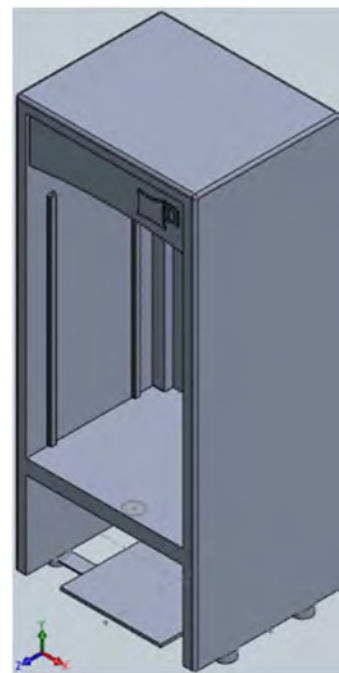
This part of the work began creating a sketch in SolidWorks® of the refrigerator rectangular geometry, depending on its measured dimensions, continuing with an extrusion to give depth to the figure, thus creating the solid. With extrusion cuts, it is possible to segment the solid and giving a better form to the structure of the refrigerator (Figure 3).

Construction of Heat Exchanger in SolidWorks®

This part began creating a sketch, where the original geometry was an aluminum foil and using a linear array (operation to make a geometry on a repetitive basis) to be able to carry out the set and with an operation of extrusion create the solid foils.

The following step was to build the copper pipe that runs through all the leaves, where the refrigerant passes to make the heat exchange, creating

Figure 3. Prototype of the refrigerator (solid)



the pipes on one side of the leaf trespassing all the foils, according to the measured dimensions.

Between each tube there is an outbound turn which represents an elbow, it turns and follows the flow of the entire copper line (Figure 4).

To create details such as edges inside of the refrigerator, new sketches in the face were used in that places where the work continues and thus be able to shape that surface in 2D.

With different design, it was possible to carry out the various details and thus have a more comprehensive and closer to reality model (Figure 4).

Construction of the Heat Sink in SolidWorks®

This part was worked through the creation of different levels, which according to the obtained heights, represented each of them.

In the foreground, the fan was built, starting with a sketch of the form, subsequently completely extruding the rectangle to make the solid form, following with slices for extrusion, cutting some spaces and detailing the corners with round turns.

Then, the work was focused with the foils, in the same way, first creating a single sheet with the particular geometry and with a linear array and continue with the repetition of the form and thus all the foils were created.

The drawing in Figure 5 is the base where the Peltier cell was placed and by which two copper pipes crossed the structure making contact with the base.

Construction of the Ventilator and Molding

After measure the dimensions, the stage began creating the simple trapezium-shaped figure. The cuts were represented in Figure 6. The details such as the moldings and the grille where the fan is located were also created in different sketches, to create the solid, which was in the form of foil.

Figure 4. Heat exchanger (solid)

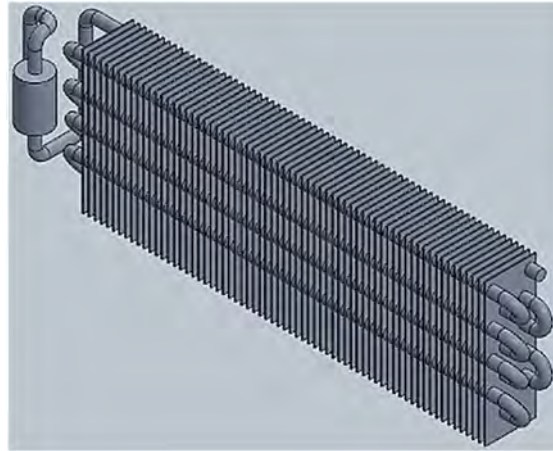


Figure 5. Heat sink (solid)

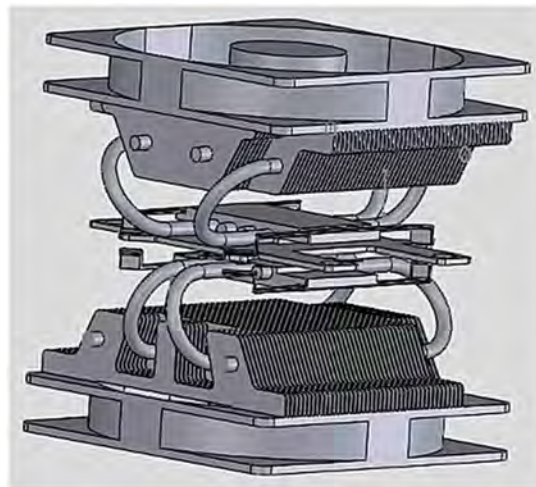
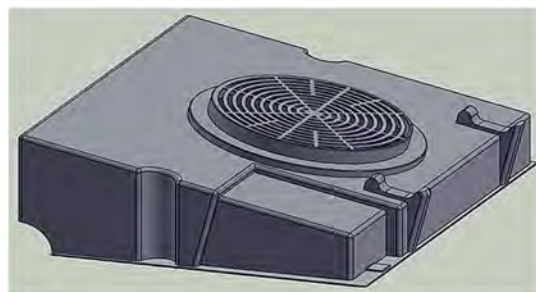


Figure 6. An extractor fan (solid)



The realization of this geometry was very critical because in this space the heat sinks and the Peltier cells going to be put. Therefore, each these parts was made with the greatest possible accuracy regarding to the real measured object.

The molding which is located below the fan was also developed with different sketches to make the moldings and the shape of the grid according to the measures.

It is worth mentioning that this is a detail to get a better view of the refrigerator, as this part was just to make it as similar as possible to reality (Figure 7).

Assembly of the Prototype Considering Peltier Cells

The following stage was for creating the assembly, adding all the created pieces according to their corresponding positions, which work by coincidence, distances, concentrics and thus be able to align all these figures.

Creating different levels according to the place where position relationship was needed, a sketch was created by adding constructive lines, which serve to punt in position each part.

With this, it can be seen an improvement in design making it more complex and detailed as seen in Figure 8, Figure 9 and Figure 10.

The construction plans were made for better positioning of each of the components (Figure 11).

Scanning of the Prototype through Photogrammetry (Peak Scan Pro) for Small Parts

Photogrammetry is a technology which is based on the scanning using triangulation; it is a technology that allows obtaining the desired object in a point cloud. The model used for this works, projects a pattern (which in this case was a small projector) on the desired object, the reflected light was picked up by a high-definition camera, leading to the reading of the shape in the form of points and transforming it into a point cloud.

Figure 7. Molding (solid)

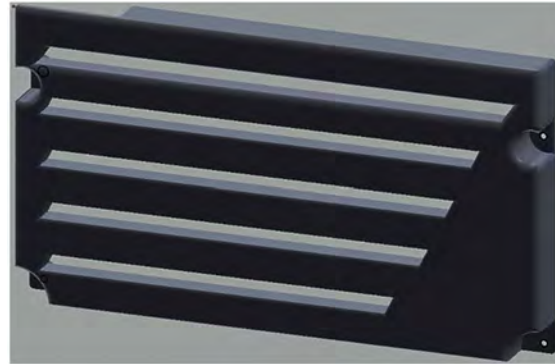


Figure 8. Refrigerator assembly (solid)



After capturing some photos, the figure was recorded but it had to be cleaned of any noise that might have occurred in the process of taking the photographs. Here, noise refers to all the spots or areas out of real surfaces that appear as displayed imperfections in the image.

This was one of the ways in which it was possible to scan small objects for the final prototype. The work was focused in obtaining a model very close to the real object to improve the simulation,

Figure 9. Location of heat sinks (solid)

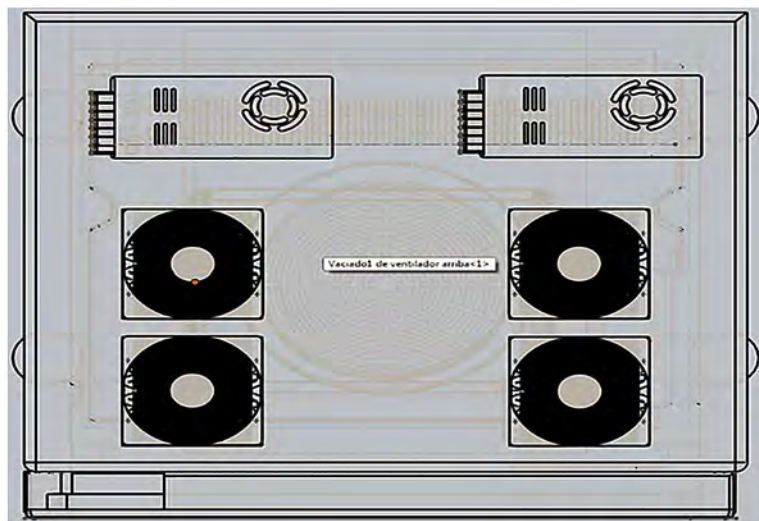
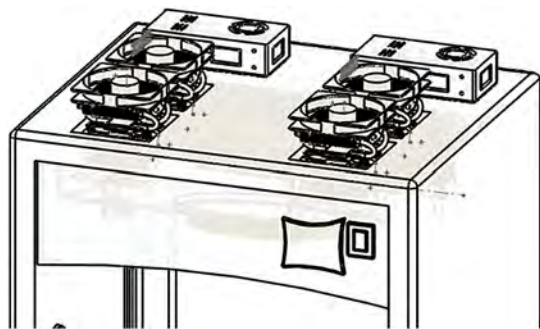


Figure 10. Heat sinks (solid)



which depends in part on how successful was the scan. The better was the scanning implied that it was very likely to have a simulation more related to reality.

Also, an objective of this work is the use in the crop field and improve the productive sector of agribusiness in those areas that are remote from the electrical grid (Thomopoulos, 2015). That is why, for the use of this equipment, it was necessary to conduct a scan of fruit, in this case the guava as a model fruit individually and in a group within a container, and to analyze the behavior or influence of this kind of fruits and their container in the refrigerator performance with the use of the Peltier cells.

The computer will scan in 3D varying the angle of view to be able to do the recording of the fruit (guava), and after the recording, conduct a test of simulation. It started by adjusting the work table where the recording took place, as the set should not suffer any movement during the scan process, to ensure a point cloud with enough quality. Two programs were used (Mephisto PEAK PRO and Mephisto Process), the first helps in capturing the photos according to the settings arranged before beginning the recording of the object and the second for post-processing the data.

It is necessary the system calibration, specifying the size of the step in degrees for rotating the object for taking the photos. Also, other parameters were specified such as the distance between the object, the camera and the projector, adjustments related the surfaces in case of being shiny or black, bodies with slight movements or completely static, among others.

The next step was taking the photos, as shown in Figure 12, Figure 13 and Figure 14. If the images of the object were very sharp, the process of reverse engineering could be very simple.

The Mephisto Process® program connects all the figures to form the figure as a point cloud, which was cleaned of all the noise (Figure 15).

Figure 11. Refrigerator level

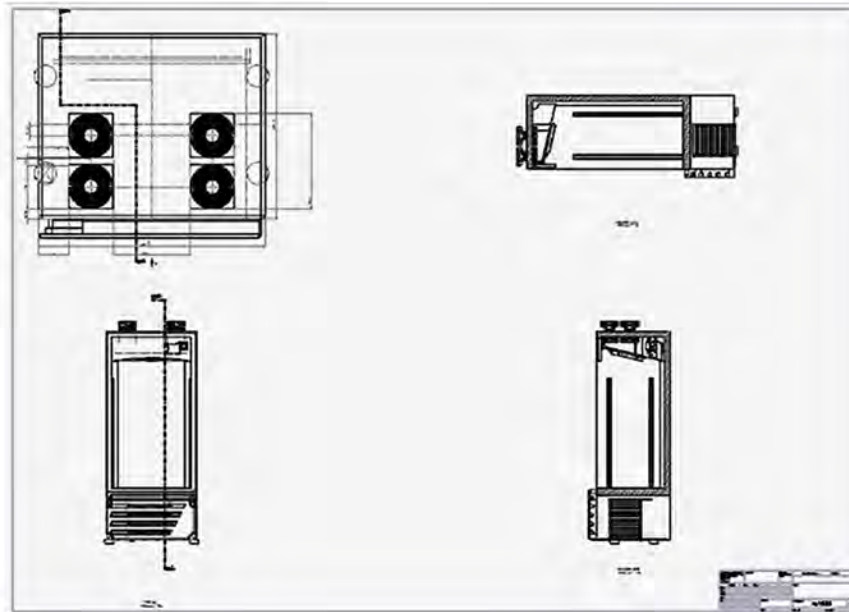


Figure 12. Peak scan PRO



Reverse Engineering

Geomagic Design X

In *Geomagic Studio*®, the scan is acquired with several options of format types. The one that resulted very useful was the file with format “.PLY” (Figure 16).

The file that contains a polygonal mesh (is a construction based on the point cloud by which lines were drawn a link between each) was opened

with the software, which will allow a real geometry refrigerator. Once the object was in the program, the next step was to proceed with the cleanliness of the polygonal mesh with the option “Mesh Doctor”. This option allowed to clean, refine and fill the defects that appeared in the scanning.

In Figure 17, each of the options that has the Mesh Doctor were used to clean the mesh properly.

Each of the spaces was cleaned and filled and now can be seen shaded as “islands” at Figure 18.

Figure 13. Scanning guava 1

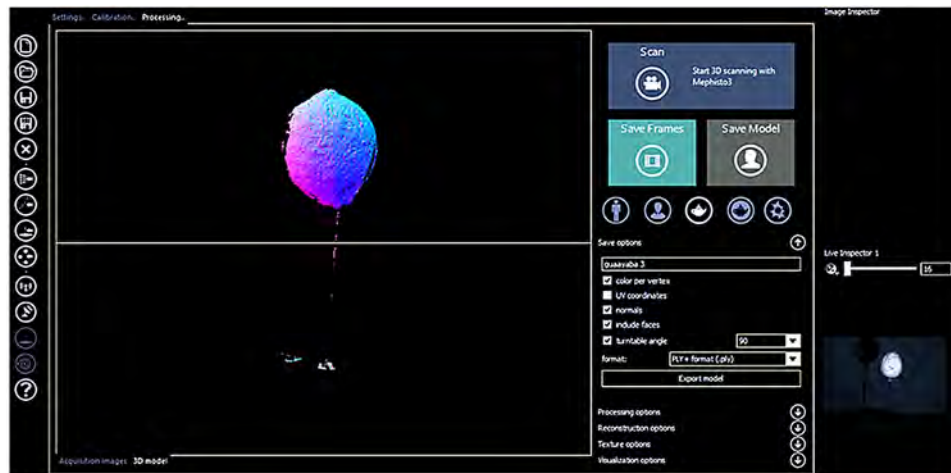


Figure 14. Scanning guava 2

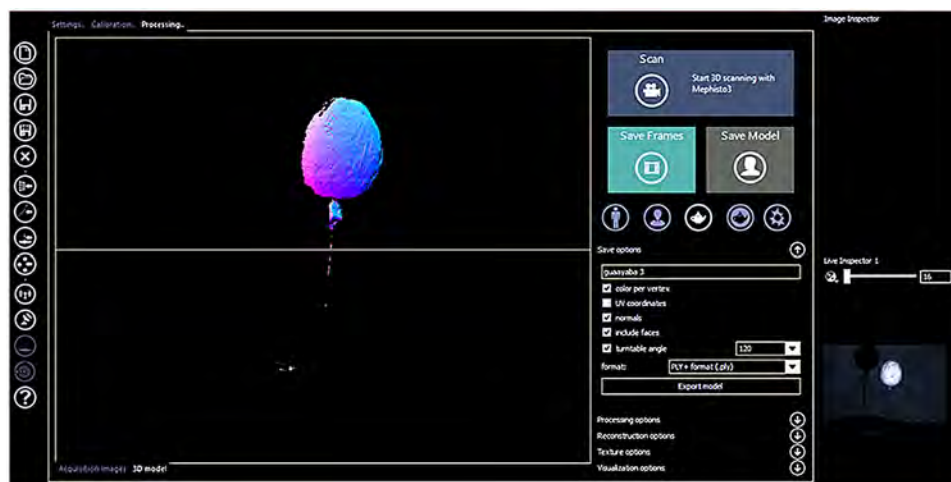


Figure 15. Point Cloud (guava)



Figure 16. Geomagic Studio 1

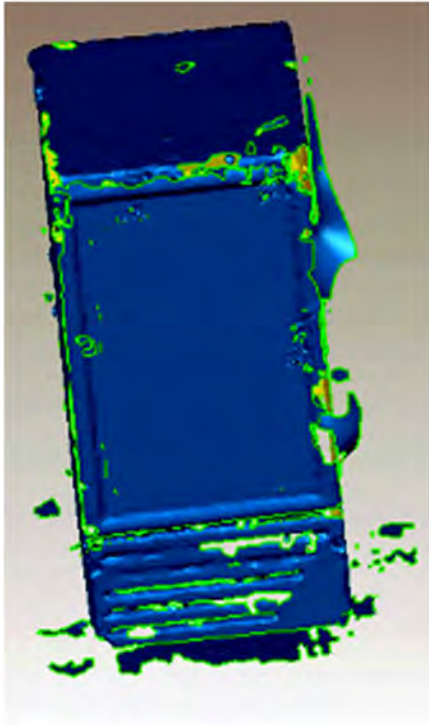


Figure 17. Geomagic Studio 2

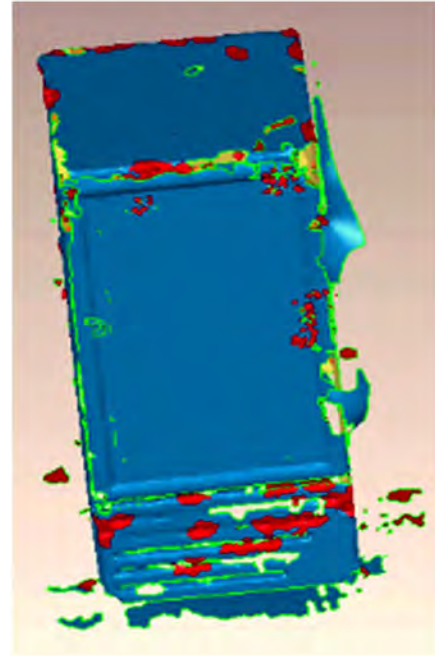
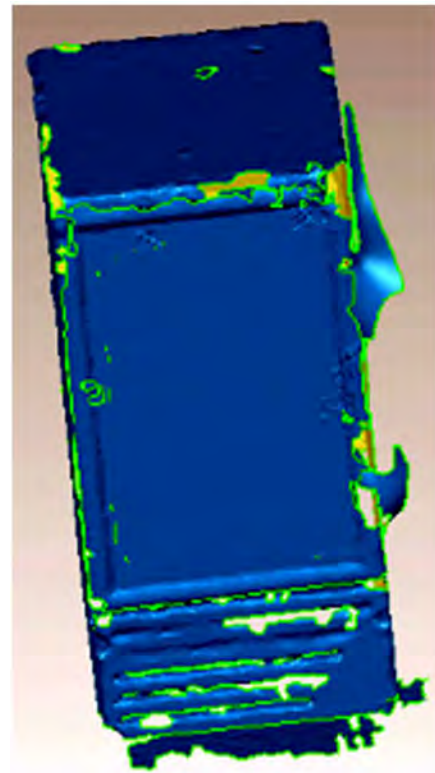


Figure 18. Geomagic Studio 3



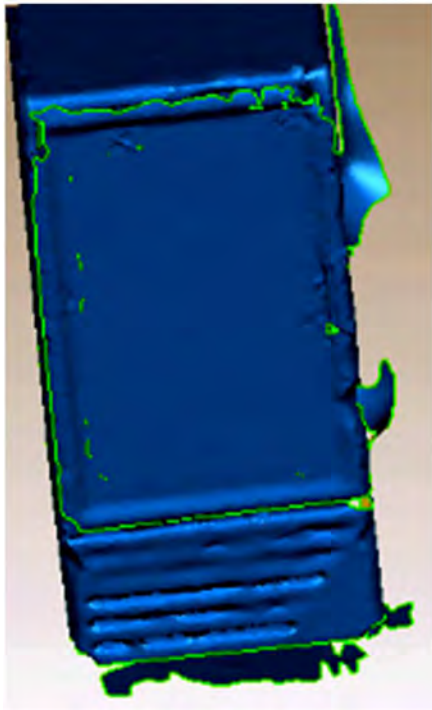
The mesh was manually cleaned and comes with a conversion from mesh to a point cloud. The point cloud was used to remove the noise generated by the scan (only the noise that was seen with the naked eye, separated from the main geometry). The noise was eliminated, and a polygonal mesh was performed with the tool “warp”. This completed the procedure to clean the polygonal mesh (Figure 20).

The next step for the reverse engineering was to make the change from a polygonal mesh to a solid body, usually an object modeled in a virtual space with a design software is covered by a polygonal mesh on the surface. The reverse engineering starts from a polygonal mesh to obtain the solid scan.

Geomagic Design X®

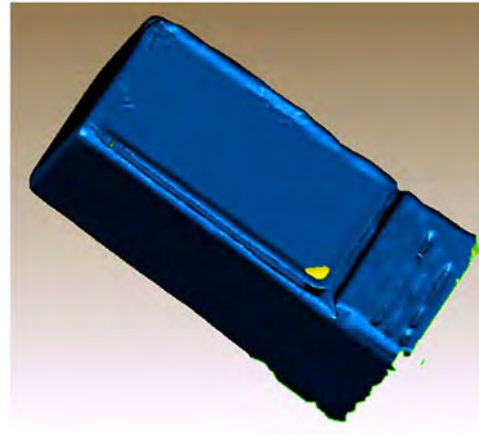
In this software the reverse engineering process was terminated. It is the final step, which gets the real geometry and transform it to a solid. It is important to clarify that the polygon mesh and its

Figure 19. Geomagic Studio 4



cleaning is only a shell of a real object and can not act as a functional object in a simulation. For use such representation of the object, it is necessary to solidify the piece. Nonetheless, this could be a complicated part of the reverse engineering.

Figure 20. Geomagic Studio 5®



The following step continues with the importation of the polygon mesh, keeping with the format “.PLY”, to manipulate it. Once having the mesh, the next step was the section of the interactive alignment, since this program generates a series of drawings X, Y, Z in a three dimensional space (Figure 21).

Once that there was an alignment of the object, there were two steps to be able to do a reverse engineering.

Figure 21. Geomagic design X I®

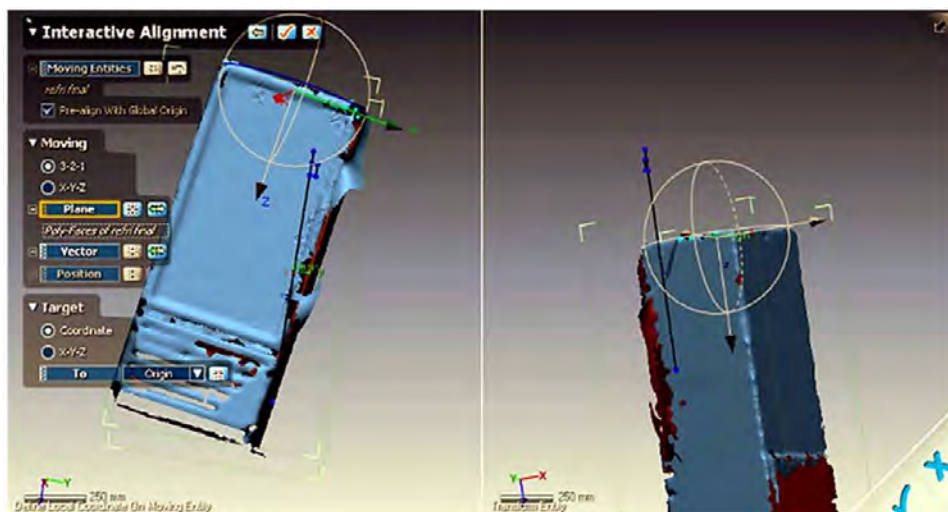
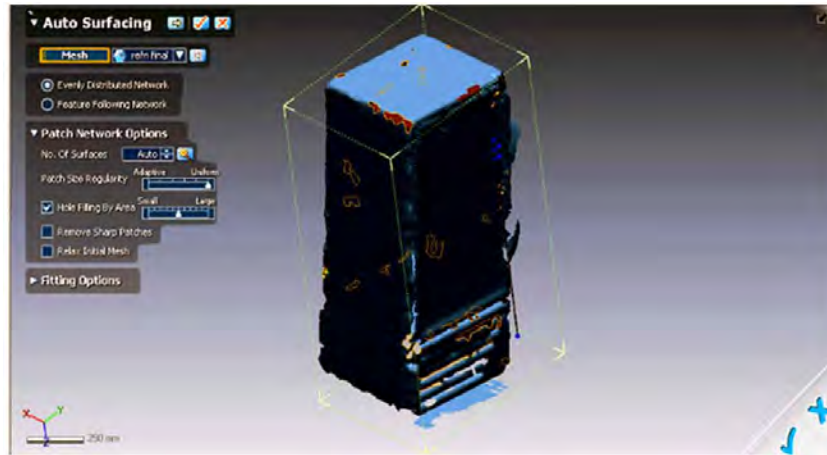


Figure 22. Geomagic design X2



Auto Surface

This tool in the program allows to adjust a surface and refill a body by a polygonal mesh. It creates the solid from a mesh and, in this way, it can be used in reverse engineering, since the tool does all the calculations and adjustments. The polygonal mesh must be complete and without errors, completely cleaned of any noise (Figure 22).

Figure 23 shows the final result.

Manual Mode

To make an reverse engineering, it is necessary to starts with the polygonal mesh of a geometry to make a solid body. The envelope generated during the process is completely hollow. To be able to generate the solid body, the whole process of a reverse way, in other words, instead of generating the drawing in an X-plane and then convert it to a three dimensional body, it begins with a drawing (polygonal mesh) and creates a plane on the surface or the base of the real object. Once having a plane, it is possible to remove the drawing. The polygonal mesh can be fragmented into the necessary drawings. Once fragmented, the basis of the object represented by a pink shading, it is possible to go to a real drawing in color blue, and the tool “Auto sketch” allows to do the stroke in blue by avoiding errors in the sizes.

Once done the dimensional drawing in blue, it generates a reference axis or a vector of extrusion (Figure 24).

Figure 25 shows the reference plane for the removal of the real dimensional drawing of the scanning.

Taking the real drawing and generating it on a reference plane with extrusion in its normal

Figure 23. Geomagic design X3 final result



Figure 24. Manual mode

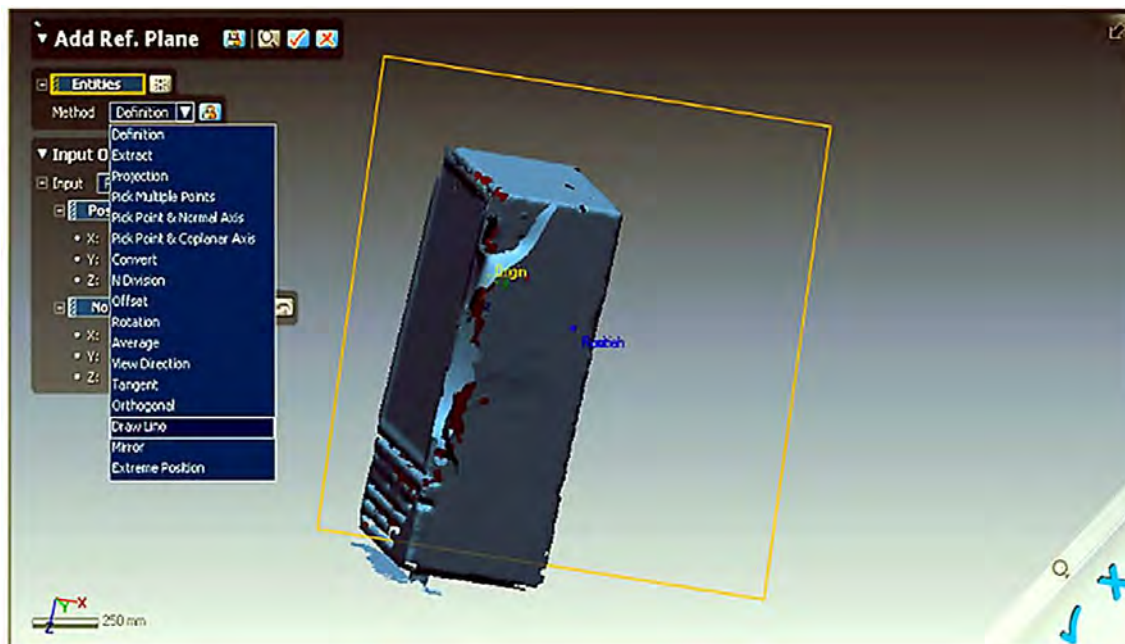
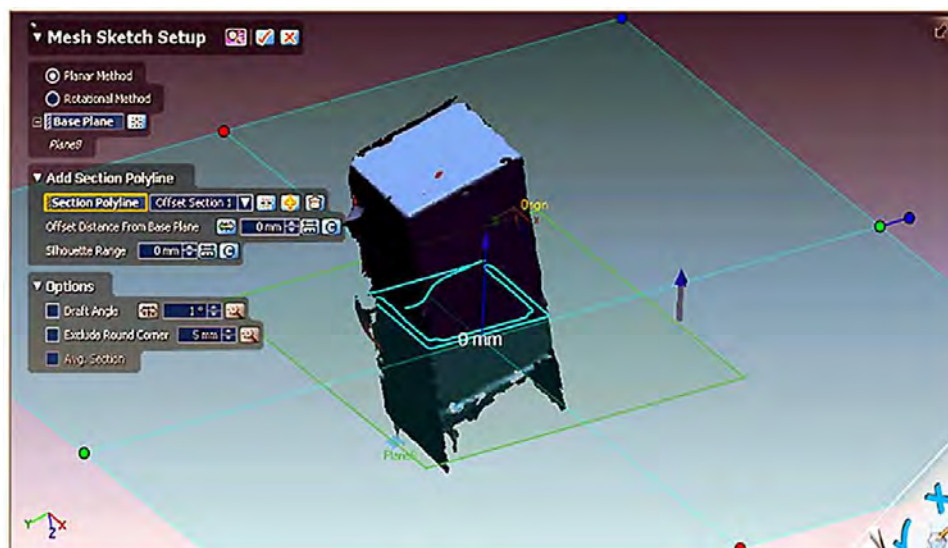


Figure 25. Reference Level, manual mode

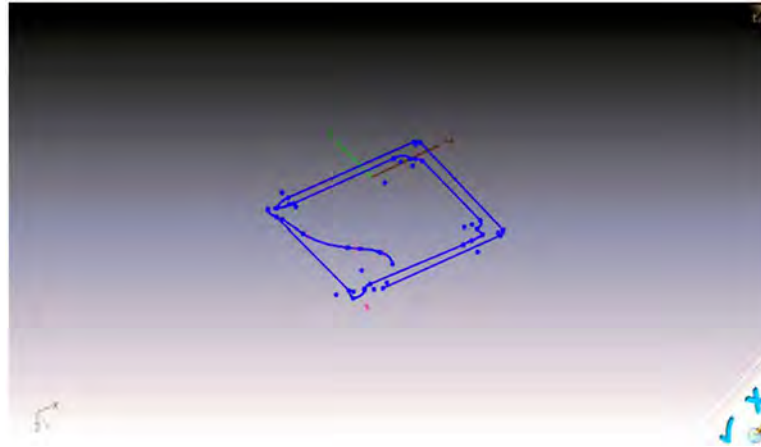


direction (direction perpendicular to the plane), Figure 26.

Once drawn the vector of extrusion, it was possible to follow the creation of the solid base. There was the option to extrude. The base object was extruded to the desired dimension and the program can do it easily.

The process can be repeated as often as necessary. If it is a regular figure without any complex shape, the work is faster, but if it is a more complex piece, the process can take much more time. This step was done for each part of the object, in the case of requiring or desiring them to be separate elements, because it is possible to enable the op-

Figure 26. Base Level, manual mode



tion to prevent the union of the extruded areas and keep them as separate bodies (Figure 27). Once having the piece created by the method of extrusion, several tools were used to replace a polygonal mesh on the surface and create a set that fully-matched to the export a real geometry of the desired object (Figure 28).

Simulation in ANSYS

ANSYS is a third generation software. It is divided into three main tools named modules: pre-processor (creation of geometry and mesh), processor and post-processor. Both the pre-processor as the post-processor are fitted with a graphical interface. This processor for the finite element solution of mechanical problems includes: analysis of static

Figure 27. Drawings, manual mode

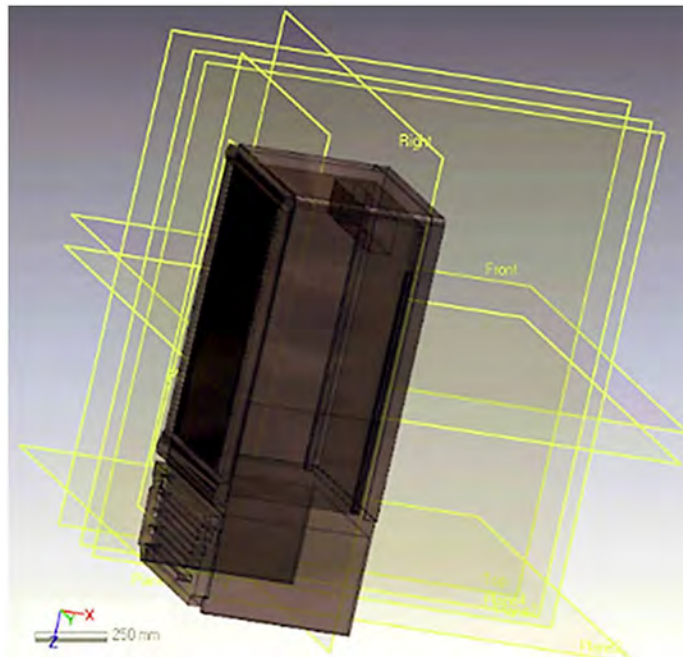
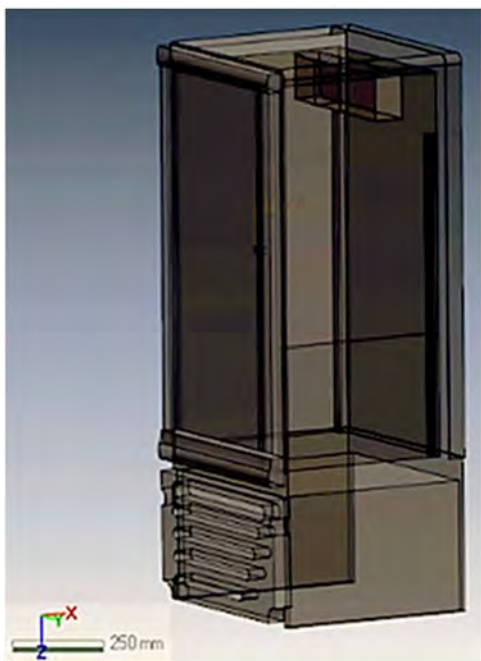


Figure 28. Final Geometry, manual mode



and dynamic structures (both for linear and not linear problems), analysis of heat transfer and fluid dynamics, as well as problems of acoustic, and electromagnetism. Normally, these tools are used simultaneously achieving mixed problems of structures along with problems of heat transfer as a whole. This software is also used in simulation software of civil and electrical engineering, physics and chemistry.

Once received the solid, it is transferred to ANSYS, where the behavior of the flows was seen according to the variables both, internal and external. It is worth mentioning that this simulation of flow was unique. In this simulation, the flows can be observed from the direction of the fans with air going up ↑ or the air going down ↓ (Figure 29).

Figure 30 shows a demonstrative flow simulation of the Peltier system with a configuration of flow going up ↑↑.

Figure 29. Simulation 1

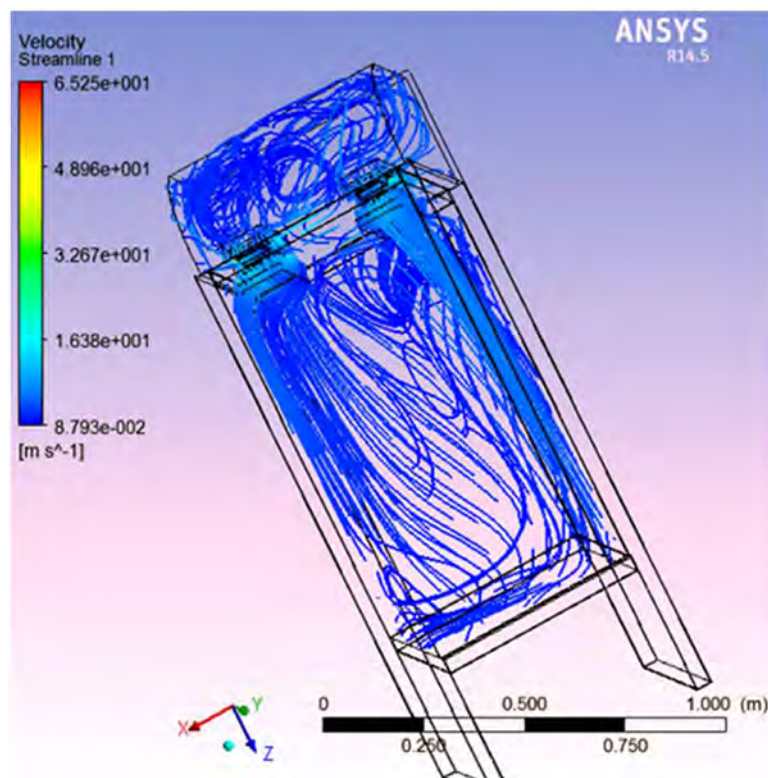


Figure 30. Simulation 2

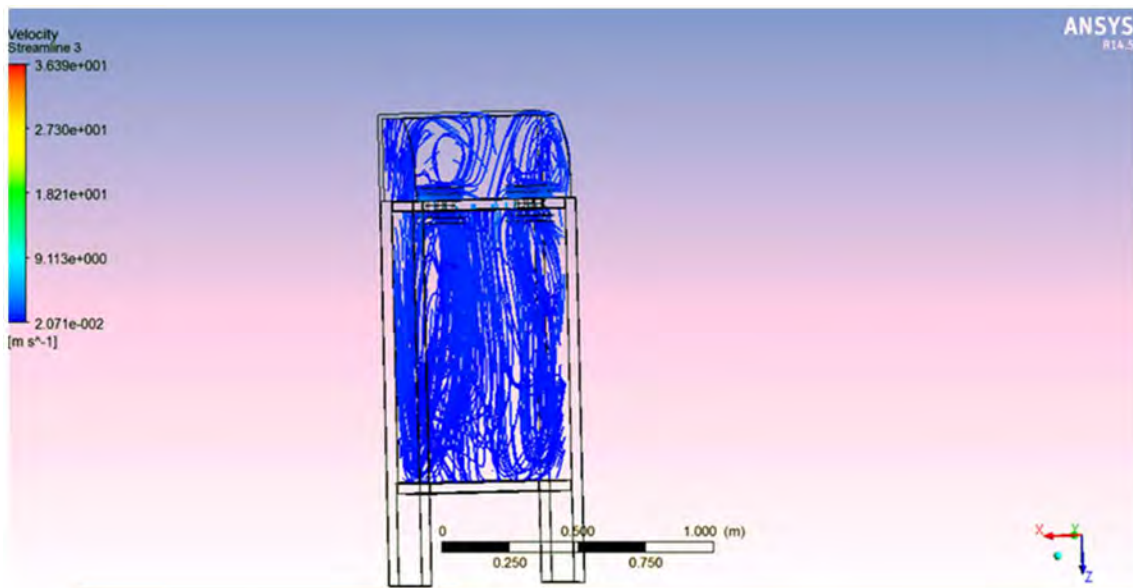
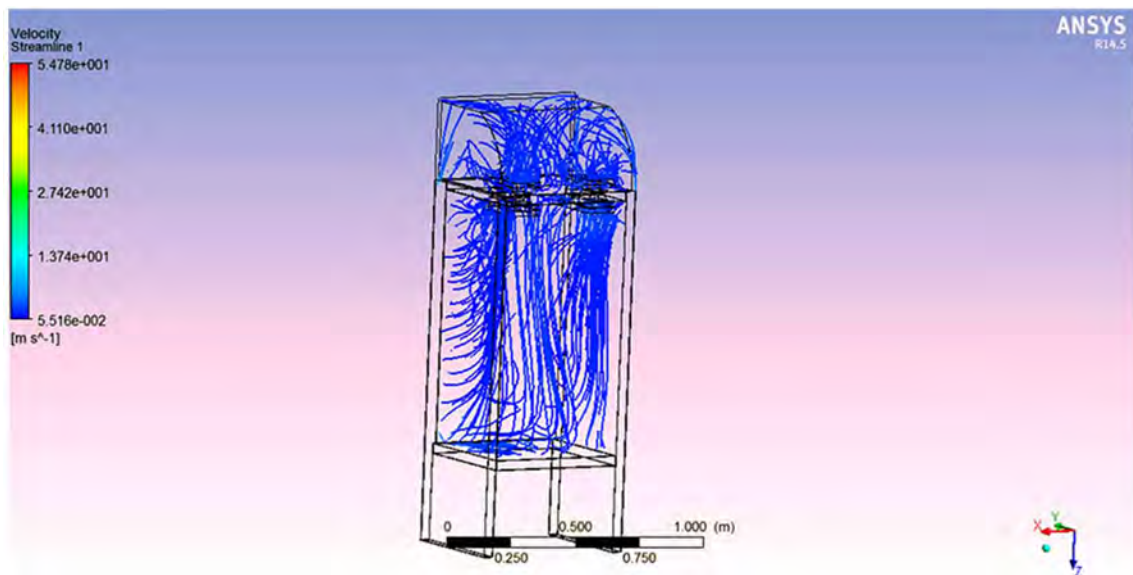


Figure 31 shows a demonstrative flow simulation of the Peltier system with a configuration of flow going down ↓↓.

FUTURE RESEARCH DIRECTIONS

- The developed prototype can be scaled but require new adjustments for changes in geometry or capabilities.

Figure 31. Simulation 3



- The technological development and studies were made for geometry, loads and specific characteristics of use in rural areas.
- The vapor-compression combined with Peltier cells was proposed for the particular application of *in situ* utilization in the place of harvest.

CONCLUSION

Nowadays the use of renewable energy sources is a challenge, the acceptance of the society still has some difficulty, and it is for this reason that there are implemented projects that demonstrate the potential of this technology in the industrial sectors.

The agroindustrial sector is one of the most important sectors worldwide both, in size and for its strategical activity associated with the food production. Nevertheless, the crops usually are rural areas far from the urban electric network. For this reason, this work proposes the implementation of an autonomous refrigerator system using photovoltaic modules. The results demonstrated that performance is optimal giving rise to a considerable improvement, increasing the possibility to be applied in the agroindustrial sector or in any sector that has or dependent on mobile cooling systems.

From the tests performed and the compilation of the simulations, the efficiency achieved in the hybrid system is very similar to the commercial system, with the difference of temperature fluctuations which are smaller, there is a lower power consumption based on the reaction time and temperature.

The Peltier cells, useful for cooling by contact, have the disadvantage of requiring more current. As a result, there were almost no savings in power consumption in comparison with the compression system. However, the advantages of Peltier cells are important, the fact of being more compact or

smaller is very useful when deployed on mobile applications or facilities that need to be transported.

It is recommended that in the scanning process of the device, getting a sound recording would facilitate the process of reverse engineering. This technology is easily manipulable and it can be deployed in a wide variety of cases, which facilitates and improves the processes and data in time and accuracy.

REFERENCES

- Bici, M., Campana, F., Petriaggi, S., & Tito, L. (2014). Study of a Point Cloud Segmentation with Part Type Recognition for Tolerance Inspection of Plastic Components via Reverse Engineering. *Computer-Aided Design and Applications*, 11(6), 640–648. doi:10.1080/16864360.2014.914382
- James, D. W., Belblidia, F., Eckermann, J. E., & Sienz, J. (2017). An innovative photogrammetry color segmentation based technique as an alternative approach to 3D scanning for reverse engineering design. *Computer-Aided Design and Applications*, 14(1), 1–16. doi:10.1080/16864360.2016.1199751
- Janik, W., Grabowik, C., & Ćwikła, G. (2014). The practical approach to freeform shape elements reverse engineering. *Applied Mechanics and Materials*, 657, 755–759. doi:10.4028/www.scientific.net/AMM.657.755
- Lopez Herrejon, R. E., Linsbauer, L., Galindo, J. A., Parejo, J. A., Benavides, D., Segura, S., & Egyed, A. (2015). An assessment of search-based techniques for reverse engineering feature models. *Journal of Systems and Software*, 103, 353–369. doi:10.1016/j.jss.2014.10.037
- Ozdarici-Ok, A., & Akyurek, Z. (2014). Object-based classification of multi-temporal images for agricultural crop mapping in Karacabey Plain, Turkey. *Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 40(7), 127-132.

Thomopoulos, R., Croitoru, M., & Tamani, N. (2015). Decision support for agri-food chains: A reverse engineering argumentation-based approach. *Ecological Informatics*, 26(P2), 182–191. doi:10.1016/j.ecoinf.2014.05.010

Ziadi, T., Henard, C., Papadakis, M., Ziane, M., & Le Traon, Y. (2014). Towards a language-independent approach for reverse-engineering of Software Product Lines. *Proceedings of the ACM Symposium on Applied Computing*, 1064–1071. doi:10.1145/2554850.2554874

ADDITIONAL READING

Ali, S., Durupt, A., Adragna, P. A., & Bosch Mauchand, M. (2014). A Reverse Engineering for Manufacturing approach. *Computer-Aided Design and Applications*, 11(6), 694–703. doi:10.1080/16864360.2014.914387

King, G., Pan, J., & Roberts, M. E. (2014). Reverse-engineering censorship in China: Randomized experimentation and participant observation. *Science*, 345(6199), 1251722. doi:10.1126/science.1251722 PMID:25146296

Kumar, V., Michikawa, T., & Suzuki, H. (2016). Improving medial surfaces for reverse engineering. *Computer-Aided Design and Applications*, 13(6), 786–791. doi:10.1080/16864360.2016.1168221

Li, F., Longstaff, A. P., Fletcher, S., & Myers, A. (2014). Rapid and accurate reverse engineering of geometry based on a multi-sensor system. *International Journal of Advanced Manufacturing Technology*, 74(1-4), 369–382. doi:10.1007/s00170-014-5997-y

Paulic, M., Irgolic, T., Balic, J., Cus, F., Cupar, A., Brajliah, T., & Drstvensek, I. (2014). Reverse engineering of parts with optical scanning and additive manufacturing. *Procedia Engineering*, 69, 795–803. doi:10.1016/j.proeng.2014.03.056

Wang, Y., Chang, K. H., & Staub, P. G. (2014). Formability Investigation of a Thin-wall Part of Double Curvature Using an Integrated Reverse Engineering Environment. *Computer-Aided Design and Applications*, 13(1), 63–76. doi:10.1080/16864360.2015.1059197

KEY TERMS AND DEFINITIONS

Extrude: It refers to the creation of an object with three dimensions from a two dimensional drawing.

Hybrid Refrigerator: A device with two or more different kinds of systems for cooling or chilling, such as: Peltier, Vapor - Compression, Ammoniac – water.

Modeling: The transference to a virtual environment of an object with a 3D visualization.

Noise in Photogrammetry Images: It refers to all the spots or areas out of real surfaces that appear as displayed imperfections in the image.

Peltier Cells: Semiconductor devices that by an electrical current can cool or chill one of their sides and heat the other.

Photogrammetry: The obtaining of reliable measurements from photographic images. Also, a scanning method comprising the projection of a pattern and photo capture in its 360° of an object to make a three-dimensional reconstruction.

Point Cloud: Points at a virtual space resulting from scans that represent an object surface but only reconstructing the location of benchmarks or targets.

Polygonal Net: Surface film produced by nodules or triangulated areas for the geometries recording.

Refrigeration: The cooling or chilling of food at domestic or industrial uses.

Simulation: The imitation of behavior or performance of a real phenomenon or object using a software that represents the object imposing the initial and boundary conditions.

Solar Energy: The electromagnetic radiation coming from the sun absorbed and transformed into electricity to be used by electronic devices in human activities.

Vector of Extrusion: It is an imaginary line which creates a three-dimensional object through a flat base, by means of the vector or imaginary line, the object will acquire the direction and distance in three dimensions.

Uniform Random Number Generation With Jumping Facilities

C

E. Jack Chen

BASF Corporation, USA

INTRODUCTION

As computer capacities and simulation technologies advance, simulation has become the method of choice for modeling and analysis. The fundamental advantage of simulation is that it can tolerate far less restrictive modeling assumptions, leading to an underlying model that is more reflective of reality and thus more valid, leading to better decisions (Lucas et al. 2015). Simulation studies are typically proceed by transforming in a more or less complicated way of a sequence of numbers between 0 and 1 produced by a pseudorandom generator into an observations of the measure of interest. A facility for generating sequences of pseudorandom numbers is a fundamental part of computer simulation systems. Furthermore, random number generators also play an important role in cryptography. A collection of random variables $x_1, x_2, \dots, x_n$ is a random sample if they are independent and identically distributed. True random numbers cannot be produced by a deterministic algorithm, and hence, random numbers generated by using a recursive equation are referred to as pseudorandom numbers. The deterministic nature of these techniques is important because it can be reproduced in computations. A facility for generating sequences of pseudorandom numbers is a fundamental part of computer simulation systems. Usually, in practice, such a facility produces a deterministic sequence of values, but externally these values should appear to be drawn independently from a uniform distribution between 0 and 1, i.e., they are independent and statistically indistinguishable from a truly random sequence. Furthermore, multiple independent streams of

random numbers are often required in simulation studies, for instance, to facilitate synchronization for variance-reduction purposes, and for making independent replications.

A random number generator (RNG) is an algorithm that starting from an initial seed (or seeds), produces a stream of numbers that behaves as if it were a random sample when analyzed using statistical tests. The RNG is closely related to the Deterministic Random Bit Generators (DRBGs). See L'Ecuyer (1990, 2013) and references therein for more information on RNGs. We describe a portable set of software utilities for uniform random-number generation. It provides for multiple generators (streams) running simultaneously, and each generator (stream) has its sequence of numbers partitioned into many long disjoint contiguous substreams, see L'Ecuyer et al. (2002). Simple procedure calls allow the user to make any generator “jump” ahead/back v steps (random numbers). Implementation issues are discussed. The basic underlying generator CMRG (Combined Multiple Recursive Generator) combines two multiple recursive random number generators with a period length of approximately 2^{191} ($\approx 3.1 \times 10^{57}$), good speed, and excellent theoretical properties, e.g., the lattice structure, see Kroese et al. (2011) for a list of desired properties.

BACKGROUND

There are a number of methods for generating the random numbers, of which the most popular are the congruential methods (mixed, multiplicative,

DOI: 10.4018/978-1-5225-2255-3.ch111

and additive). The (mixed) linear congruential generators (LCGs) are defined by

$$x_i = (ax_{i-1} + c) \text{MOD } m, \quad u_i = \frac{x_i}{m}, \quad x_0 \in \{1, \dots, m-1\} \quad i > 0.$$

where m (the modulus) is a positive integer (usually a very large primary number), a (the multiplier) $\in \{1, \dots, m-1\}$ and c (the increment) is a nonnegative integer. This mathematical notation signifies that x_i is the remainder of $(ax_{i-1} + c)$ divided by m . Hence, $x_i \in \{1, \dots, m-1\}$. Thus, random variable u_i is a uniform 0, 1 variable. Note that

$$x_{i+v} = \left(a^v x_i + \frac{c(a^v - 1)}{a - 1} \right) \text{MOD } m.$$

Hence, every x_i is completely determined by m , a , c , and x_0 . The sequence x_i repeats once it returns to a previously visited value. The period of a generator is the length of a generated steam before it begins to repeat. If $u_0 = u_p$ (where $p > 0$), then the length p is called the period. The longest possible period for a LCG is m , i.e., m represents the desired number of different values that could be generated for the random numbers. Hence, the modulus m is often taken as a large prime number close to the largest integer directly representable on the computer (i.e., equal or near $2^{31}-1$ for 32-bit computers). If $p=m$, we say that the generator has full period. The required conditions on how to choose m , a , and c so that the corresponding LCG will have full period are known, see Knuth (1997) or Law (2014).

LCGs are sensitive with respect to the parameters, especially the value of a . When $c > 0$, the LCGs are called mixed LCGs. When $c = 0$,

$$x_i = ax_{i-1} \text{MOD } m, \quad u_i = \frac{x_i}{m}, \quad x_0 \in \{1, \dots, m-1\} \quad i > 0.$$

These LCGs are called multiplicative LCGs. Most existing implementations of LCGs are multiplicative LCGs, because in general the value of c does not have a large impact of the quality of an LCG. Note that if $x_i = 0$, then all subsequent x_i are identically 0. Thus, the longest possible period for a multiplicative LCG is $m-1$. Furthermore,

$$x_{i+v} = a^v x_i \text{MOD } m$$

Most experts now recognize that small LCGs with moduli around 2^{31} or so should no longer be used as general-purpose random-number generators. Not only can one exhaust the period in a few minutes on a PC (personal computer), but more importantly the poor structure of the points can dramatically bias simulation results for sample sizes much smaller than the period length.

One way of extending the basic LCG is to combine two or more LCGs through summation. Another way of extending the basic LCG is to use a higher-order recursion. A multiple recursive random number generator (MRG), which goes from integer to integer according to the recursion

$$x_i = (a_1 x_{i-1} + \dots + a_k x_{i-k}) \text{MOD } m, \quad u_i = \frac{x_i}{m}$$

$$A \text{ seed } x_0, \dots, x_{k-2}, x_{k-1} \in \{1, \dots, m-1\}$$

where i , k , and m are positive integers, and $a_1, \dots, a_k \in \{0, 1, \dots, m-1\}$. To increase the efficiency and ease the implementation, the MRG algorithm usually set all but two a_i 's to 0. Furthermore, the nonzero a_i should be small. However, L'Ecuyer (2013) points that these conditions are generally in conflict with those required for having a good lattice structure and statistical robustness. See Law (2014) on the lattice structure of pseudorandom numbers. The longest possible period for a MRG is m^k-1 (L'Ecuyer, 1996). Other way of extending the basic LCG is the additive congruential RNG (ACRON). The ACRON sets $a=1$ and replace

c by some random number proceeding x_i in the sequence, for example, x_{i-1} (so that more than one seed is required to start calculating sequence). Wilkmaratna (2008) indicated that the ACRON is a special case of a multiple recursive generator.

Other classes of RNGs are available, e.g., twisted generalized feedback shift register generators (Matsumoto & Kurita, 1994; Matsumoto & Nishimura, 1998). The RNGs discussed so far are by computational methods. Another approach is by physical methods, e.g., Kanter et al. (2010). The output based on a physical method is usually unpredictable. Thus, this class of random bit generators is commonly known as non-deterministic random bit generators (NRBGs). L'Ecuyer (2010) points out that those physical devices are unnecessary and unpractical for Monte Carlo methods, because deterministic algorithmic methods are much more convenient than hardware devices. See L'Ecuyer (2013) and the numerical algorithms group (2013) for more information on other methods of generating random numbers.

MAIN FOCUS

We discuss the algorithm and implementation of L'Ecuyer's (1999) combined multiple recursive random number generator. A combination of generators can have a much longer period than any of its components. Furthermore, with well-chosen parameters, the CMRG has good structural properties and passes many statistical tests, see Kroese et al. (2011). A list of classical LCGs and their properties can be found on Karl Entacher's website: <http://random.mat.sbg.ac.at/results/karl/server/>.

Basic Generators

A newly employed generator is the combined multiple recursive random number generator, which goes from integer to integer according to the recursion

$$x_{j,n} = (a_{j,1}x_{j,n-1} + \dots + a_{j,k}x_{j,n-k}) \text{MOD } m_j;$$

for $j=1, \dots, J$

$$Z_n = \sum_{j=1}^J (\delta_j x_{j,n}) \text{MOD } m_1$$

$$u_n = \begin{cases} z_n / (m_1 + 1) & \text{if } z_n > 0 \\ m_1 / (m_1 + 1) & \text{if } z_n = 0 \end{cases}$$

where J is the number of MRGs used in the CMRG, j, n, k , and m_j are positive integers, and each $a_{j,k}$ belongs to $Z_m = \{0, 1, \dots, m-1\}$, δ_j are integers, and the greatest common divisor of δ_j and m_j is one for each j . To have fast algorithms, with a given j most $a_{j,k}$ should be 0. The transformation of u_n makes sure that u_n is never equal to 0 or 1 (otherwise, trouble may arise, e.g., when taking the logarithm to generate an exponential random variate). Assume that the m_j are distinct primes and that each recurrence j has a primitive characteristic polynomial, and thus period length $\rho_j = m_j^k - 1$ and the longest possible period for a CMRG is $\rho_1 \dots \rho_J / 2^{J-1}$ (L'Ecuyer, 2013).

The combined generator is a close approximation to a single MRG with a modulus equal to the product of the moduli of the components MRGs. Thus, the CMRG has the advantages associated with a larger modulus while permitting an implementation using smaller values. The CMRG combining multiple recursive sequences provides an efficient way of implementing random-number generators with long periods and good structural properties. If the parameters are well chosen, such generators are statistically far more robust than simple linear congruential generators that fit into a single computer word (L'Ecuyer, 1999).

The Need for Multiple Substreams

Many disjoint random number subsequences are often required in simulation studies, for instance, (1) to make independent replications and/or (2) to

associate distinct “streams” of random numbers with different sources of randomness in the system to facilitate synchronization for variance reduction. To produce such “streams,” different seeds (values of vector S_i) must be obtained far enough apart in the sequence to insure that the streams do not overlap. Selection of seeds should also consider statistical properties between streams, such as apparent independence.

In other words, given any X_i (with seed S_i) and positive integer v , there should be a quick way to compute X_{i+v} (with seed S_{i+v}) without generating all intermediate values. The availability of efficient jump-ahead methods is very useful because it permits one to partition the RNG sequence into long disjoint stream and substreams of random numbers. Most packages offer no facility for jumping ahead directly from X_i to X_{i+v} or to compute distant seeds efficiently. Many simulation languages offer a limited number of streams, all based on the same generator, but using fixed starting seeds set say 100,000 values apart. This provides relatively low flexibility. Suppose, for instance, that you want to perform independent pairs of replications with common random numbers across the configurations (i.e., between any two runs of the same pair) in order to compare two different configurations of a system.

To insure proper synchronization, you want every generator to start from the same seed in both runs of the same pair. However, in general, these two runs will make a different number of calls to a generator, and programming “tricks” should be used to skip a proper amount of random numbers to resynchronize the generators for the next pair without overlap in the random number streams. This requires extra programming effort and could be error prone. Good software tools should ease the programmer’s task in that respect. A simple procedure call should permit resetting a generator to previous seed or jumping ahead to a new seed for the next run. Of course, the sequence of “new seeds” (one per run) should be the same of both configurations for the system. Implementing such tools requires efficient “jumping ahead” facilities,

which in turn ask for efficient procedures to compute $(a \times s \text{ MOD } m)$, where a and s are positive integers.

Computing $(a \times s \text{ MOD } m)$

Consider a 32-bit computer on which all integers between -2^{32-1} and 2^{32-1} (exclusive) are well represented. We want to compute $(a \times s \text{ MOD } m)$, where a , s , and m are positive integers smaller than 2^{32-1} . Without loss of generality, we assume that $a < m$ and $s < m$ (if not, replace a and s by $a \text{ MOD } m$ and $s \text{ MOD } m$, respectively). In order to keep seeds with 32-bit precision, all operations can not produce any number greater than 2^{53} (the IEEE-754 standard, i.e., the floating-point numbers have at least 53 bits of precision for the mantissa). Therefore, special algorithm was needed to compute $(a \times s \text{ MOD } m)$, where a and s are less than 2^{32} (around 4.3×10^9).

The following algorithm was used. Let

$$a = a_1 \times 2^{17} + a_2$$

so

$$a \times s = a_1 \times s \times 2^{17} + a_2 \times s$$

Therefore,

$$a \times s \text{ MOD } m = \left((a_1 \times s \text{ MOD } m) \times 2^{17} + a_2 \times s \right) \text{ MOD } m$$

where $a_1 < 2^{15}$, so $a_1 \times s < 2^{47} (= 2^{15} \times 2^{32})$, and $a_2 < 2^{17}$, so $a_2 \times s < 2^{49} (= 2^{17} \times 2^{32})$. Because $z = (a_1 \times s \text{ MOD } m) < 2^{32}$, we have $z \times 2^{17} < 2^{49}$ so that all the intermediate terms in the above computations are less than 2^{53} . Therefore, all seed values will have exact accuracy.

Note that $(a \times s \text{ MOD } m)$, where a and s are less than 2^{32} , can be computed directly on computers that are capable of performing operations with 128 bit quadruple precision, i.e., numbers as large as 2^{113} (around 1.0×10^{34}).

Computing the Jumping Matrices

The initial state of substreams can be computed easily if jumping-ahead facilities are available for the individual MRG components; that is, if an efficient algorithm is available for computing the state of the MRG v steps ahead of the current one, for large values of v . L'Ecuyer (1990) explained one way of doing that, based on the fact that the MRG can be viewed as a LCG in matrix form, whose state is a k -dimensional vector and whose multiplier is a $k \times k$ matrix A . To jump ahead by v values, just multiply the current state by $(A^v \text{MOD } m)$. The matrix $(A^v \text{MOD } m)$ can be pre-computed in time $O(\log_2 v)$, using the divide-and-conquer algorithm (Knuth, 1997).

That is, for the MRG random number generator, X_{i+v} can be computed directly from X_i using

$$\begin{aligned} X_{i+v} &= (A^v X_i) \text{MOD } m \\ &= (A^v \text{MOD } m) X_i \text{MOD } m \end{aligned}$$

where

$$A = \begin{pmatrix} 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ a_k & a_{k-1} & \cdots & a_1 \end{pmatrix}$$

is an invertible $k \times k$ matrix and

$$X_i = \begin{pmatrix} X_i \\ X_{i+1} \\ \vdots \\ X_{i+k-1} \end{pmatrix}$$

is a $k \times 1$ vector. When A has this special structure, the first $k-1$ components of X_i are obtained by shifting the last $k-1$ components of X_{i-1} , and the last component of X_i is a linear combination of the components of X_{i-1} , according to the MRG

(L'Ecuyer, 1990). Kroese et al. (2011) point out that an MRG can be interpreted and implemented as a matrix multiplicative congruential generator, i.e.

$$X_i = (A X_{i-1}) \text{MOD } m, i = 1, 2, \dots$$

The divide-and-conquer algorithm (Knuth, 1997) was used to compute the jumping matrices A^v using the following recursion:

$$A^v \text{MOD } m = \begin{cases} A & \text{if } v = 1; \\ A(A^{v-1} \text{MOD } m) \text{MOD } m & \text{if } v > 1, v \text{ odd}; \\ (A^{v/2} \text{MOD } m)(A^{v/2} \text{MOD } m) \text{MOD } m & \text{if } v > 1, v \text{ even} \end{cases}$$

A Random Number Package

We now propose a portable set of utilities for random number generation. We consider a basic underlying generator of period p . Let s_0 be the basic seed (initial state) for this generator and $s_1, s_2, \dots$ be its sequence of successive states. Let T denote the transition function of the generator, that is, the operator $T: S \rightarrow S$ such that $T(s_i) = s_{i+1}$, and T^q its q -fold composition ($T^q(s_i) = s_{i+q}$). Starting from states $I_1 = s_0, I_2 = s_q = T^q(I_1), I_3 = s_{2q} = T^q(I_2), \dots, I_G = s_{(G-1)q} = T^q(I_{G-1})$, respectively. Each of these G sequences corresponds to a “stream”, with length q . Note that q needs to be a very large number, say no smaller than 10^9 . Hence, there will be G virtual RNGs. I_g is called the *initial seed* of stream g . At any moment during program execution, stream g is in some state C_g , say, in its subsequence number r , that is, such that $C_g = T^{r-1}(I_g)$. We call C_g the *current state* of stream g .

The product of $a_{ij}(m_j - 1)$ is less than 2^{53} . If this condition holds, then the integer $a_{ij}x_{ij}$ is always represented exactly in floating point on a 32-bit computer that supports the IEEE floating-point arithmetic standard, with at least 53 bits of precision for the mantissa. The generator can then be

implemented directly in floating-point arithmetic, which is typically faster than an integer arithmetic implementation. On the other hand, with this implementation, the state of the generator is represented over $64k$ J bits, as opposed to $32k$ J bits when the x_{ij} are represented as 32-bit integers.

The mrg32k3a of L'Ecuyer (1999) implemented a CMRG with 2 components ordered 3, whose coefficients satisfy above condition. The moduli and coefficients are $m_1 = 2^{32} - 209 = 4294967087$, $a_{11} = 0$, $a_{12} = 1403580$, $a_{13} = -810728$, $m_2 = 2^{32} - 22853 = 4294944443$, $a_{21} = 527612$, $a_{22} = 0$, $a_{23} = -1370589$. Its period length is $\rho = (m_1^3 - 1) / (m_2^3 - 1) / 2 \approx 2^{191} \approx 3.1 \times 10^{57}$. This implementation set $\delta_1 = -\delta_2 = 1$. The parameters have been chosen so that the period is long, a fast implementation is available, and the generator performs well with respect to the spectral test. The mrg32k3a has been implemented as one of the core random number generators in many simulation packages.

Before this procedure is called for the first time, one must initialize the global variables s_{10} , s_{11} , s_{12} to (exact) non-negative integer values less than m_1 and not all zero, and s_{20} , s_{21} , s_{22} to non-negative integers less than m_2 and not all zero. The vectors (s_{10}, s_{11}, s_{12}) and (s_{20}, s_{21}, s_{22}) are the initial values of $(x_{1,0}, x_{1,1}, x_{1,2})$ and $(x_{2,0}, x_{2,1}, x_{2,2})$, respectively. They constitute the *seed*.

Therefore, in this implementation

$$x_{1,n} = (1403580.0x_{1,n-1} - 1370589.0x_{2,n-3}) \text{ MOD } 4294967087$$

$$x_{2,n} = (527612.0x_{2,n-1} - 1370589.0x_{2,n-3}) \text{ MOD } 4294944443$$

$$Z_n = (x_{1,n} - x_{2,n}) \text{ MOD } 4294967087$$

$$u_n = \begin{cases} z_n / 4294967088 & \text{if } z_n > 0 \\ 4294967087 / 4294967088 & \text{if } z_n = 0 \end{cases}$$

The matrix A_1 for s_{1n} to jump ahead one step is

$$A_1 = \begin{pmatrix} 0.0 & 1.0 & 0.0 \\ 0.0 & 0.0 & 1.0 \\ -810728.0 & 1403580.0 & 0.0 \end{pmatrix}$$

and the matrix A_2 for s_{2n} to jump ahead one step is

$$A_2 = \begin{pmatrix} 0.0 & 1.0 & 0.0 \\ 0.0 & 0.0 & 1.0 \\ -1370589.0 & 0.0 & 527612.0 \end{pmatrix}$$

The List 1 is an implementation of the CMRG (L'Ecuyer's mrg32k3a) in C.

The mrg32k3a RNG software packages are available for download via url <http://www.iro.umontreal.ca/~lecuyer/myftp/streams00/>

Jumping Backward

The jump-back matrix also exists. This is true because each component has a primitive characteristic polynomial (a necessary condition). Given a vector s_n , one may jump back v steps to s_{n-v} . For example, the jump-back-one-step matrix B_j is $A_j^{\rho_j-1} \text{ MOD } m_j$, where ρ_j is the period of the j th MRG of the CMRG. As one might expect, $A_j^v \times B_j^v \text{ MOD } m_j$ turns out to be the identity matrix I , for $v = 1, 2, \dots, \rho_j$. This equation is consistent with our intuition that $A_j^{\rho_j-1} \text{ MOD } m_j = I$, and this follows from Fermat's theorem. The matrix B_j generates the same stream but in reverse order. As we can see from the new recursion below, the values of the parameters b_{ij} are much larger than the original ones, where b_{ij} are the parameters of the new CMRG. Furthermore, we no longer have $b_{ij}(m_j-1) < 2^{53}$ for all b_{ij} . Therefore, the original recursion provides a more efficient generator.

The matrix B_1 for s_{1n} to jump back one step (i.e., jump ahead ρ_1-1 step) is

List 1.

```

#define norm          2.328306549295728e-10          // 1.0/(m1+1)
#define norm2         2.328318825240738e-10          // 1.0/(m2+1)
#define m1            4294967087.0
#define m2            4294944443.0
// the initial seed
double               s[2][3] = {
{0.0,                0.0,                1.0},
{0.0,                0.0,                1.0}
};
double MRG32k3a()
{
long k;
double p;
p = 1403580.0 * s[0][1] - 810728.0 * s[0][0];
k = long (p / m1);      p -= k*m1; if (p < 0.0) p += m1;
s[0][0] = s[0][1]; s[0][1] = s[0][2]; s[0][2] = p;
p = 527612.0 * s[1][2] - 1370589.0 * s[1][0];
k = long (p / m2);      p -= k*m2; if (p < 0.0) p += m2;
s[1][0] = s[1][1]; s[1][1] = s[1][2]; s[1][2] = p;
if (s[0][2] <= s[1][2]) return ((s[0][2] - s[1][2] + m1) * norm);
else return ((s[0][2] - s[1][2]) * norm);
}

```

$$B_1 = \begin{pmatrix} 184888585.0 & 0.0 & 1945170933.0 \\ 1.0 & 0.0 & 0.0 \\ 0.0 & 1.0 & 0.0 \end{pmatrix}$$

$$x_{2,n} = (360363334.0x_{2,n+2} - 4225571728.0x_{2,n+3}) \text{ MOD } 4294944443$$

or

$$x_{2,n} = (360363334.0x_{2,n+2} - 69372715x_{2,n+3}) \text{ MOD } 4294944443$$

and the matrix B_2 for s_{2n} to jump back one step (i.e., jump ahead ρ_2-1 step) is

$$B_2 = \begin{pmatrix} 0.0 & 360363334.0 & 4225571728.0 \\ 1.0 & 0.0 & 1.0 \\ 0.0 & 1.0 & 0.0 \end{pmatrix}$$

The reverse stream follows the recursion

$$x_{1,n} = (184888585.0x_{1,n+1} - 1945170933.0x_{1,n+3}) \text{ MOD } 4294967087$$

Generating Nonuniform Random Variate

To simulate the underlying system under study, various nonuniform random variates are required. These random variates are used to simulate the observations from the desired distribution. Random variates from distributions other than the uniform [0,1] are generated by applying further transformations to the output value u_i of the uniform RNG. The simplest way of generating a random variate X with distribution function F

is using the inverse transformation. Let $U(a,b)$ denote the uniform distribution between a and b . Based on the fact that $y=F(x)\sim U(0,1)$, we generate a random variate $U\sim U(0,1)$ and return the value $X=F^{-1}(U)$ as a variate with distribution F . Here F^{-1} is the inverse distribution function (or quantile function) and is defined by

$$X = F^{-1}(U) = \inf \{x : F(x) \geq U\}.$$

Note that

$$\begin{aligned} P(X \leq x) &= P(F^{-1}(U) \leq x) \\ &= P(U \leq F(x)) = F(x) \end{aligned}$$

For more information of using inverse transform to generate random variates, see Law (2014). Other methods are available when F^{-1} is difficult or expensive to compute, see Kroese et al. (2011).

FUTURE TRENDS

Random number generators continue to be key components of any computer simulation, cryptography, and other areas where producing an unpredictable result is desirable. As computers become 64-bits based operation, RNGs based on 64-bits will be developed. Utilities for RNGs will be incorporated into many simulation packages and will be readily available for simulation practitioners.

CONCLUSION

We have discussed the backbone CMRG random-number generator, several utilities to enhance its practical use, and implementation issues. Our implementation eliminates the need to use a fixed number of pre-computed seeds, which would provide little flexibility. The proposed random-number-generation package provides jumping

facilities, has good speed, a long period length, and excellent theoretical/statistical properties. It provides for multiple generators (streams) running simultaneously, and each generator (stream) has its sequence of numbers partitioned into many long disjoint contiguous substreams. The RNG mrg32k3a has been studied intensively and implemented in several software packages.

REFERENCES

- Kanter, I., Aviad, Y., Reidler, I., Cohen, E., & Rosenbluh, M. (2010). An optical ultrafast random bit generator. *Nature Photonics*, 4(1), 58–61. doi:10.1038/nphoton.2009.235
- Knuth, D. E. (1997). *The art of computer programming*. 2 (3rd ed.). Reading, MA: Addison-Wesley.
- Kroese, D. P., Taimre, T., & Botev, Z. I. (2011). *Handbook of Monte Carlo Methods*. New York: John Wiley & Sons. doi:10.1002/9781118014967
- L'Ecuyer, P. (2010). Pseudorandom number generators. In R. Cont (Ed.), *Encyclopedia of Quantitative Finance*. doi:10.1002/9780470061602.eqf13003
- L'Ecuyer, P. (2013). *Random number generation*. Retrieved January 3, 2013, from <http://www.iro.umontreal.ca/~lecuyer/myftp/papers/handstat.pdf>
- Law, A. M. (2014). *Simulation modeling and analysis* (5th ed.). New York: McGraw-Hill.
- LEcuyer, P. (1990). Random numbers for simulation. *Communications of the ACM*, 33(10), 85–97. doi:10.1145/84537.84555
- LEcuyer, P. (1996). Combined multiple recursive generators. *Operations Research*, 44(5), 816–822. doi:10.1287/opre.44.5.816
- LEcuyer, P. (1999). Good parameter sets for combined multiple recursive random number generators. *Operations Research*, 47(1), 159–164. doi:10.1287/opre.47.1.159

LEcuyer, P., & Cote, S. (1991). Implementing a Random Number Package with Splitting Facilities. *ACM Transactions on Mathematical Software*, 17(1), 98–111. doi:10.1145/103147.103158

LEcuyer, P., Simard, R., Chen, E. J., & Kelton, W. D. (2002). An object-oriented random-number package with many long streams and substreams. *Operations Research*, 50(6), 1073–1074. doi:10.1287/opre.50.6.1073.358

Matsumoto, M., & Kurita, Y. (1994). Twisted GFSR generators II. *ACM Transactions on Modeling and Computer Simulation*, 4(3), 254–266. doi:10.1145/189443.189445

Matsumoto, M., & Nishimura, T. (1998). Mersenne twister: A 623-dimensionally equidistributed uniform pseudorandom number generator. *ACM Transactions on Modeling and Computer Simulation*, 8(1), 3–30. doi:10.1145/272991.272995

Sen, R. P. (2010). *Operations Research: Algorithms and Applications*. New Delhi: PHI Learning Private Limited.

The Numerical Algorithms Group. (n.d.). *G05 – Random number Generators*. Retrieved January 3, 2013, from http://www.nag.co.uk/numeric/fl/nagdoc_fl23/pdf/G05/g05intro.pdf

Wilkrmaratna, R. S. (2008). The additive congruential random number generator – A special case of a multiple recursive generator. *Journal of Computational and Applied Mathematics*, 216(2), 371–387. doi:10.1016/j.cam.2007.05.018

ADDITIONAL READING

Barker, E. & Kelsey, J. (2012). *Recommendation for random number generation using deterministic random bit generators*. NIST SP800-90A.

Brent, R. P. (2007). Some long-period random number generators using shifts and xors. *ANZIAM Journal*, 4, C188–C202. doi:10.21914/anziamj.v48i0.40

Gentle, J. E. (2003). *Random number generation and Monte Carlo methods* (2nd ed.). New York: Springer.

Glasserman, P. (2003). *Monte Carlo methods in financial engineering*. New York: Springer. doi:10.1007/978-0-387-21617-1

Goresky, M., & Klapper, A. (2003). Efficient multiply-with-carry random number generators with maximal period. *ACM Transactions on Modeling and Computer Simulation*, 13(4), 310–321. doi:10.1145/945511.945514

Haramoto, H., Matsumoto, M., Nishimura, T., Panneton, F., & LEcuyer, P. (2008). Efficient jump ahead for F2-linear random number generators. *INFORMS Journal on Computing*, 20(3), 385–390. doi:10.1287/ijoc.1070.0251

Hörmann, W., & Leydold, J. (2003). Continuous random variate generation by fast numerical inversion. *ACM Transactions on Modeling and Computer Simulation*, 13(4), 347–362. doi:10.1145/945511.945517

Hörmann, W., Leydold, J., & Derflinger, G. (2004). *Automatic nonuniform random variate generation*. Berlin: Springer-Verlag. doi:10.1007/978-3-662-05946-3

Lucas, T. W., Kelton, W. D., Sánchez, P. J., Sanchez, S. M., & Anderson, B. L. (2015). Changing the paradigm: Simulation, now a method of first resort. *Naval Research Logistics*, 62(4), 293–303. doi:10.1002/nav.21628

Panneton, F., & LEcuyer, P. (2010). Resolution-stationary random number generators. *Mathematics and Computers in Simulation*, 80(6), 1096–1103. doi:10.1016/j.matcom.2007.09.014

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (2007). *Numerical Recipes: The Art of Scientific Computing* (3rd ed.). New York: Cambridge University Press.

Wichmann, B. A., & Hill, I. D. (2006). Generating good pseudo-random numbers. *Computational Statistics & Data Analysis*, 51(3), 1614–1622. doi:10.1016/j.csda.2006.05.019

KEY TERMS AND DEFINITIONS

Common Random Numbers: A variance reduction technique that uses the same random number streams to simulation different systems.

Deterministic Algorithm: An algorithm that given the same input always products the same outputs.

Deterministic Sequence: A sequence that the value of any future elements can be determined when the value of the current element is known

Jump Ahead (Back) ν Steps: A utility to advance (go back) ν steps in the random number stream without going through the stream sequentially.

Period Length (of RNGs): The length of the random number sequence before it recycles itself.

Quantile Function (Inverse Distribution Function): The quantile function (given a probability p) returns the value at or below which 100 p percent of the population lies.

Random Number Generators (RNGs): Computational or physical devices designed to generate a sequence of numbers that appear to be random, as determined by statistical tests.

Variance Reduction: A technique to reduce the variance of the estimate, hence, increase the precision of the estimate.

Category C

Computer Vision and Image Processing

A Critical Overview of Image Segmentation Techniques Based on Transition Region

Yu-Jin Zhang

Tsinghua University, China

INTRODUCTION

Image segmentation is the key step from image processing to image analysis, and is an important technique of image engineering (Zhang, 2009). It refers to the general procedures and processes for dividing an image into regions with each has some particular characteristics and extracting the target region of interest. Since there is no general theory for image segmentation itself, yet, the image segmentation has always been in-depth study of attention. A large variations of theories and technologies have been applied to solve the image segmentation problem (Zhang, 2006). According to a study based on EI Compendex, the first segmentation algorithm is proposed in 1962, the number of literatures on image segmentation has reached 77,000 half-century late (Zhang, 2015b). According to the increasing tendency, it is estimated that this number should be around 100,000 now.

Image segmentation based on transition region is a special or distinctive type of techniques that are different from traditional boundary-based or region-based techniques. Its strategy is first detecting a transition region between object regions and background regions, then segmenting image into object region and background region based on the properties of transition region. The first technique using transition region is often referred to effective average gradient (EAG) that in fact is just the name of the core parameter of this technique. EAG has raised more than a quarter of a century ago (Zhang, 1989). Journal paper (Zhang, 1991) introduced this technique in detail,

and journal paper (Zhang, 1996) further extended this technique to more general situations. Since then, there are many subsequent related researches and applications, and a series of papers in the literature citing (Zhang, 1991) and/or (Zhang, 1996) are published worldwide (Zhang, 2015a).

Using Google Scholar, a number of papers citing (Zhang, 1991) and/or (Zhang, 1996) are searched and selected, a study on the statistics of these papers is conducted, an analysis of these papers is performed, and discussions on these results are provided.

BACKGROUND

The main contributions of the original journal papers (Zhang, 1991) and (Zhang, 1996) can be summarized into five points:

1. **They Clarified the Three Evidences for the Existing of Transition Region:** Firstly it can prove that, according to Shannon sampling theorem, the discrete edge obtained by sampling a continuous image has always a transition region of at least one pixel wide (Gerbrands, 1988). Secondly the transition region in the actual image can be directly observed. Finally, the transition region can be generated by means of smoothing property inherent in the nature of image capturing.
2. **They Analyzed the Rationality and Feasibility of Using the Transition Region for Segmentation:** The transition region is spatially located between the object and

the background, so the boundary separating the object and background should be in the transition region. Moreover, the gray-values of transition region are in the middle of the gray-values of object pixels and background pixels, so if the thresholding techniques is used for segmentation, the appropriate threshold values should be belong to the value set of the transition region pixels. In addition, selecting the threshold value from the transition region must be more reliable than selecting the threshold value from the full image.

3. **They Put Forward a Specific Method for Determining the Transition Region:**

Although the existing of transition regions is an objective reality, the effective detection method is required. To determine the transition region in actual images, a method consisting of clip transformation of images, computation of the EAG curves, determination of the optimal dynamic range, and thereby extracting the transition region is proposed. This method converts the problem of determining the transition region into an optimization problem, without any preset parameter to be adjusted in the determination. This fully automatic method has the iterative nature and uses the global information of images, so it is relatively stable and robust. Finally, according to the characteristics of the EAG curve, a fast realization way is designed to improve the computational efficiency.

4. **They Proved Three Properties of the Two Gray-Values Defining the Transition Region:**

Depending on the aforementioned computation method for the transition region, the dynamic range of transition region is defined by two gray-values (i.e., L_{low} and L_{high}) obtained from EAG curve. With the help of mathematical analysis, three properties of these two values are formally proved: 1) They always exist and only one for each transition region; 2) They can make both corresponding functions $EAG(L_{low})$ and

$EAG(L_{high})$ reaching the maximum values; 3) They satisfy $L_{low} < L_{high}$ in practical situations. In (Zhang, 1996), these three properties are also proved for images having more than one transition region.

5. **They Discussed How to Use the Transition Region Further to Segment Image:**

On the basis of the transition region extracted, according to the property that the gray-value range of transition region is between gray-values of object and background, a suitable threshold value can be determined with the help of gray-values of the transition region, for example, this threshold value can be the average value or the mode value of the gray-value ranges of the transition region.

SURVEY ON THE FOLLOW-UP LITERATURE

A large number of follow-up researches for (Zhang, 1991) and/or (Zhang, 1996) have been conducted, many new papers have been published in referring/citing them. With the help of Google Scholar, many of them (most of them are journal papers) can be found.

Citation-Statistics Over the Years

According to the search results from Google Scholar, so far the number of records citing (Zhang, 1991) is 87, the distribution of these records for each year is shown in Figure 1; and the number of records citing (Zhang, 1996) is 59, the distribution of these records for each year is shown in Figure 2.

The total number of records citing these two papers reaches 146. However, after a careful reading and checking of all these records, it is found that 7 of them are incorrectly picked up by Google Scholar (3 records for citing (Zhang, 1991), 4 records for citing (Zhang, 1996)). On the other side, there are 13 records citing both (Zhang, 1991) and (Zhang, 1996). After screening, there are totally 126 papers citing (Zhang, 1991)

Figure 1. The distribution of the 87 records citing (Zhang, 1991) for each year (from Google Scholar)

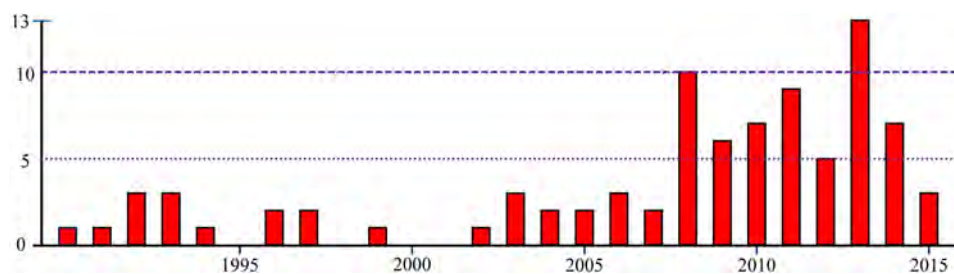
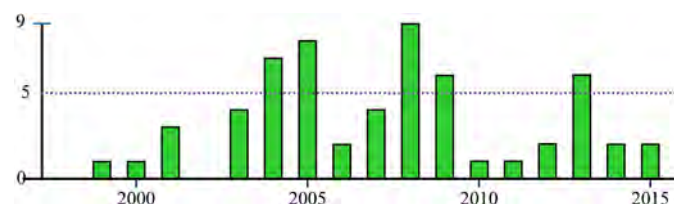


Figure 2. The distribution of the 59 records citing (Zhang, 1996) for each year (from Google Scholar)



and/or (Zhang, 1996), their distribution is shown in Figure 3. In Figure 3, the red (dark) columns denote the papers citing only (Zhang, 1991), the green (light) columns denote the papers citing only (Zhang, 1996), while the columns with pattern denote the papers citing both (Zhang, 1991) and (Zhang, 1996).

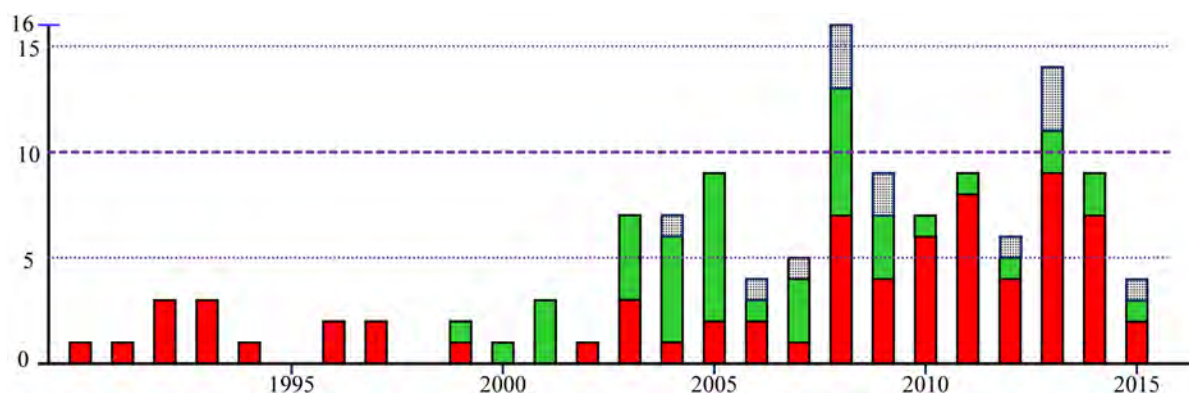
From Figure 3, it is seen that the majority of follow-up papers are published in the past 10 years. For example, there are 42 records in the 5 year's period from 2011 to 2015; there are 41 records in the 5 year's period from 2006 to 2010; while there are only 43 records for the rest 15 years

(before 2004). It seems the related researches for transition region-based image segmentation have an upward trend, though many years have passed since its publication.

Purpose-Statistics Over the Years

From the perspective of scientific researchers, the purposes for paper citations of academic and technical literatures can be attributed mostly to the following three groups.

Figure 3. The distribution of the 126 records citing both (Zhang, 1991) and (Zhang, 1996) for each year



1. **Introduction/Indication:** Citing the paper as the example of a kind of typical techniques, especially in the review articles or in the literature review section of a paper.
2. **Applications:** Using the concepts and/or algorithms presented in the cited paper for a practical work (implementation may be required, and some parameters may often be adjusted to suit specific application needs).
3. **Follow-Up Research:** With the analysis of the principles and/or deficiencies presented in the paper, put forward some new concepts and ways to improve the technical performance, or to modify/extend the technique to a wider range of application fields.

By analyzing the above 126 papers with emphasis on their main purposes, it is found that 29 of them can be classified into the first group, 44 of them can be classified into the second group, while 53 of them can be classified into the third group. It seems that this type of techniques not only be widely applied, but also be grasped by the attention of many researchers. The distributions of these three groups for each year are shown in Figure 4, with the red (dark) columns denoting the first group, the green (light) columns denoting the second group, and the columns with pattern denoting the third group.

Content-Statistics Over the Years

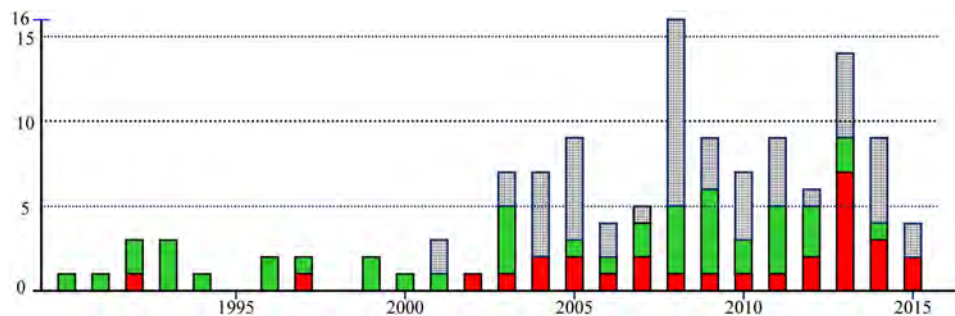
From the point of view of academic and technical research, the 53 papers belong to the third groups are worthy of further analysis.

Firstly, all these papers hold a positive attitude for the existence of the transition region in images, and some of them also provide further refinement or more examples. Most of them also recognized the rationality and feasibility of the transition region for image segmentation. They conducted a number of works for the determination of transition regions as well as for the segmentation of image based on transition regions.

The original methods mentioned in (Zhang, 1991) and (Zhang, 1996) for computing the transition region is relatively simple and intuitive, the follow-up researches started from different angles, and designed different methods for improving the performances, such as the segmentation accuracy and noise robustness. Among these methods, most of them can be described from the consideration of their understanding of the transition region.

To extract the transition region from images, it is necessary to separate the transition region from both object and background regions. This task can be regarded as a classification problem, and suitable features (representing the characteristics of different regions) for classification

Figure 4. The classification of the 126 records into 3 groups according to their citing purposes for each year



are required. One important characteristic of the transition region is the relatively significant gray-level variations in a comparatively small space or scale. It may be caused by two factors, one is the existence of large gray-level difference between adjacent pixels in transition region, and other is the presence of frequent changes of gray-level among neighboring pixels in transition region. Of course, the effects of these two factors may also be conjoined. In other words, the transition regions would be characterized by large variations and/or high frequency of pixel values.

For the 53 papers belong to the third group, most of them still determine the transition region based on these two factors, some papers use one of two factors and several papers combine two factors. There are also certain papers with other particular/specific approaches. So, according to the contents, these papers can be classified into 4 categories:

1. Consider the magnitude change of gray values.
2. Consider the frequency change of gray values.
3. Consider both magnitude and frequency changes in conjunction.
4. Other approaches based on different ideas and focuses.

The distributions of these four categories for each year are shown in Figure 5, with the red (dark) columns denoting the first category, the green (light) columns denoting the second category, the columns with pattern denoting the

third category, and the white columns denoting the fourth category. One thing can be seen from Figure 5 is that most papers in the third category are published recently.

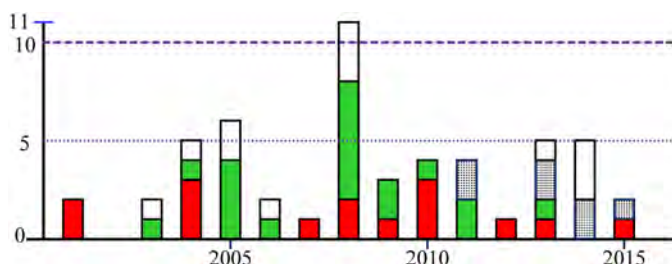
ANALYSIS AND DISCUSSION ON SOME ISSUES

In the follow-up study of the transition region-based image segmentation techniques, there have been some doubtful claims and statements. A number of them have already been announced or discussed in papers (Zhang, 1991) and (Zhang, 1996), but it seems that they have not been noticed or have not been properly understood. Below, some of them raised from the following four citing papers are analyzed and discussed, in order to guide the further works along correct directions.

Paper (Groenewald, 1993)

From the perspective of gradient computation, the paper (Groenewald, 1993) compared the effective average gradient (EAG) method (Zhang, 1991) with the average gradient (AG) method (Watanabe, 1974) and the more general histogram transformation / modification (HT/HM) method (Weszka, 1979). It deduced that EAG is a smoothed version of AG or HT/HM. Moreover, it also designed an example to point out there may be situations with $L_{low} > L_{high}$, and thus to challenge the third property of the two gray-values defining the transition region. However, these claims are questionable.

Figure 5. The classification of the 53 records into 4 categories of contents for each year



First, it must highlight that EAG is not a smoothed version of AG. According to the definition of EAG, where only the pixels with not zero (or very small) values of gradient are considered in the EAG computation, EAG is a selected statistics, and is different from AG. In addition, in order to compute $EAG(L_{low})$ and $EAG(L_{high})$ curves, special clip transformations (cut section will be set to cutting value) in low-end and high-end are introduced respectively, only the pixels retained after the clip transformation are considered in the EAG computation. Both selections are corresponding to non-linear transformations, and the result is a function of clip level (cutting value), so the relation between EAG and AG is absolutely not a simple smooth one. Furthermore, although both EAG and AG make use of the 2-D local contrast information in image, EAG also utilizes 2-D global information in computation. Different from HT/HM that uses each time only the overall information of pixels in one gray value, EAG utilizes each time the overall information of all the pixels with non-zero values of gradient, and thus is more robust to noise. This can also be seen directly from the comparison of Figure 4 and Figure 5 in (Zhang, 1991).

Second, it should point out that the paper (Groenewald, 1993) ignored the pixels with zero gradient value in the simplified approximation to gradient computation. This is equivalent to only take the values of low-end clip and high-end clip as the minimum and maximum values of images, respectively, and does not reflect the dynamic influence caused by the change of clip values. It is seen that, on the one hand the simplification in (Groenewald, 1993) is just a rough approximation to HT/HM; on the other hand the gradient computation in (Zhang, 1991) is not the traditional computation. Considering the common segmentation model that assumes an object on background, then the most of the pixels in the image are either in the internal of object or in the internal of background, the pixels on the border between object and background are only a small part. If one relays the limit of zero gradient as the

approximation did in (Groenewald, 1993), most of the pixels will be ignored in that way, such an approximation is obviously unreasonable.

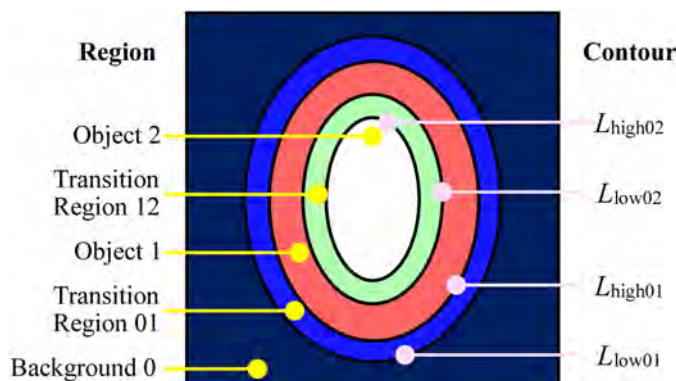
Finally, the paper (Groenewald, 1993) gives an example to illustrate that the situation of $L_{low} > L_{high}$ might happen, but in fact the example given is just a case that there are two non-overlap transition regions (the corresponding profiles of two transition regions are connected) in a gray value image. If extending the 1-D edge curve from the example of (Groenewald, 1993) whose gradient curve has two pics in 2-D plane, the situation would be as shown in Figure 6, where the first pic would correspond to the transition region 01 between background 0 and object 1, while the second pic would correspond to the transition region 12 between object 1 and object 2. In this case, the object 1 could be seen as the background of object 2. Since there are two transition regions, then the computation of L_{low} and L_{high} should be performed for each pair of them. The results would be L_{low01} and L_{high01} corresponding to transition region 01, and L_{low02} and L_{high02} corresponding to transition region 02. In (Zhang, 1996), it pointed out that the three properties of the two gray-values defining the (single) transition region still valid for the case of multiple transition regions, as shown in Figure 9 of (Zhang, 1996), in which $L_{low01} < L_{high01}$ for the first transition region and $L_{low02} < L_{high02}$ for the second transition region. (Groenewald, 1993) compared L_{low02} with L_{high01} , hence the abnormality appeared. In reality, L_{low02} and L_{high01} are belong to two different transition regions, there is no means to make them into comparison.

Paper (Zhang, 2004)

For calculating the gradient values needed to obtain EAG curves, the paper (Zhang, 2004) designed a procedure to accelerate the computation. The result obtained when using Sobel gradient operator is that the number of multiplications and additions can be reduced to $18/256 \approx 7\%$ of direct computation.

In reality, (Zhang, 1991) already introduced a fast procedure to perform similar computation

Figure 6. The illustration of two transition regions in the image



with the help of Fibonacci sequence, the amount of computation can be reduced to less than 5% of direct computation, and this reduction does not depend on the gradient operators used.

Paper (Liu, 2007)

The paper (Liu, 2007) made a review of transition region-based image segmentation methods proposed before 2007, and considered paper (Weszka, 1979) as the first one to introduce the transition region theory into image segmentation. It seems that this statement does not understand the meaning of transition region and does not distinguish these two methods clearly.

Literally, there has been no term “transition region” appeared in (Weszka, 1979), let alone “transition region theory”. In fact, (Weszka, 1979) discussed the histogram modification / histogram transformation (HM/HT) method, whose basic idea is to rise the difference between the peaks and valleys in the histogram of the original image by using other information than just pixel gray values, so as to increase peak-valley discrimination and to facilitate the selection of threshold from histogram. On the other side, EAG of (Zhang, 1991) and (Zhang, 1996) directly determines the transition region and then the threshold for segmentation, without any histogram computation. Obviously, these two methods are different both in basic ideas and in technical route. They should not be confused.

In (Zhang, 1991), the evaluation and comparison of EAG and HM/HT are already performed. HM/HT takes the edge value of each pixel to weight its role played in the selection of threshold, but does not distinguish the edge values caused by gray value variation among neighboring pixels with that caused by noise. In EAG method, the contribution of each pixel to the selection of threshold does not only depend on the gradient value of that pixel but also depend on the distribution of gray values in the entire image. This is equivalent to incorporate not only the edge values but also the global 2-D information (such as pixel distribution information reflected by clip transformation) into the detection of the transition region. Since EAG method combines more *a priori* knowledge than that of HM/HT method, it is expected to provide better result than that of HM/HT method. In addition, except the quantity of *a priori* knowledge, how to use the knowledge also influences the performance. Both EAG and HM/HT methods utilize the local contrast information (edge value information and gradient information), but the ways are different. HM/HT method uses the local contrast information for each gray level separately, while EAG method uses simultaneously the local contrast information of all pixels with non-zero gradient values. As more samples are used in statistics, the EAG method is more robust to noise influence than HM/HT method (Zhang, 1991).

Incidentally, (Liu, 2007) attributed the method of (Wang, 2005) to the class of mathematical

morphology methods, while in fact, there is no any mention of mathematical morphology in (Wang, 2005). Although (Wang, 2005) does some edge thinning operations by using gray values and gradient values in neighborhood, but does not employ any mathematical morphology procedure.

Paper (Hua, 2008)

The paper (Hua, 2008) used a supervised way to specify both gray values for limiting the dynamic range of the transition region (for the purpose of comparison with the original documents (Zhang, 1991) and (Zhang, 1996), these two values will also be referred to as L_{low} and L_{high}), and performed the following clip transformation (to remain the contributions come from pixels in the range of $[L_{\text{low}}, L_{\text{high}}]$ and to remove the contributions come from pixels outside this range):

$$f_{\text{TR}}(x, y) = \begin{cases} L_{\text{low}} & \text{if } f(x, y) < L_{\text{low}} \\ f(x, y) & \text{if } L_{\text{low}} < f(x, y) < L_{\text{high}} \\ L_{\text{high}} & \text{if } f(x, y) > L_{\text{high}} \end{cases}$$

This transformation has basically the same idea as that of the two clip transformations introduced in the original paper (Zhang, 1991). It can be seen as one formula combined two transformations. It has the same effect as the two transformations if initially fixing the first gray value while changing the second gray value, and then fixing the second gray value while changing the first gray value. Nevertheless, two gray values are fixed together in (Hua, 2008), so the potential benefits of (Zhang, 1991) for iterative computation of optimal values separately were lost.

FURTHER RESEARCH DIRECTIONS

As indicated in Figure 3, it seems the related researches are still ascendant. Below are some directions that would be worth for further research.

1. Why were there so many follow-up research works? One important reason is that the transition region has no unique and quantitative definition, yet. Note in (Zhang, 1991), the transition region is just defined by the computation results. Many researches have been largely carried out along this line of thought in these years. For example, local complexity or local entropy based methods need to determine a threshold value for detecting the transition region. Pixels with local complexity or local entropy values greater than (or less than) the threshold will be assigned to the transition region. Paper (Yan, 2005) attempts to define a general transition region, though it does relax the selection for features representing the properties of transition region, a threshold value is still required to detect the transition region. In other words, using different threshold values will result in different transition region. In addition, the fact that there is no unique and quantitative definition for transition region also causes no ground truth in the method evaluation. Moreover, no suitable test database makes the research results difficult to compare, with a number of uncertainties.
2. As mentioned above, for determining the transition region, most methods use information either on the magnitude variation of gray values or on the frequency variation of gray values. Both of them have their own characteristics, each of them has its own application areas. Combination would be a potential solution, but how to learn from each other needs more in-depth research works.
3. Till now, the related researches are basically concentrated in the direction along the third point indicated in background section. However, there would be some research spaces in the direction along the fifth point. After obtaining the transition region, various characteristics of this region can be used for image segmentation. Firstly, the gradation characteristics can be exploited. The range

of gray values is between that of object and background, so the mean or extreme values of transition region can be used to determine a segmentation threshold (Zhang, 1991). Secondly, the spatial characteristics can be exploited. Since the transition region is a strip-shaped region surrounding the boundary of the object, so it can be refined (e.g., using a medial axis transformation or performing a strike skeleton operation) to obtain the object contour inside it (Zhang, 2001). Further researches on other promising characteristics would be worthwhile.

4. Several follow-up researches classify the transition region based methods into two categories: direct methods (directly compute pixels belong to transition region) and indirect methods (first computing range values, then determining the transition region). With the help of common terms in image segmentation, the former can be regarded as the region-based methods, while the latter can be regarded as the boundary-based methods. It is generally believed that the direct methods obtain the transition region straightly, while the indirect methods have turned a corner. However, the ultimate goal of transition region determination is to segment images into object and background. If a threshold value is needed, for direct methods, determining the threshold should be performed after obtaining the transition region; while for indirect methods, the threshold could be determined directly after obtaining the range values of transition region. In case that the transition region is determined by L_{low} and L_{high} , the threshold can be directly computed from these two values, even the explicit extraction of the transition region becomes unnecessary. Here, it seems that the “direct” and “indirect” just swap each other. What lessons could be learnt from it?

CONCLUSION

A survey for researches and applications on image segmentation based on transition region determination (literature searched with the help of Google Scholar) in these years is conducted. These related papers are sorted first according to the publishing year, and then grouped according to their purposes and contents (with techniques used). Some questionable issues in these papers are pointed out and critically discussed, and several further research directions are indicated and analyzed. The statistics shows that the related researches catch a lot of attention, especially in recent years, so further researches, applications, and improvements will continue.

ACKNOWLEDGMENT

This work has been supported by the National Natural Science Foundation under Grants NNSF-61171118 and the Ministry of Education under Grants SRFDP-20110002110057.

REFERENCES

- Gerbrands, J. J. (1988). *Segmentation of noisy images*. Delft University.
- Groenewald, A. M., Barnard, E., & Botha, E. C. (1993). Related approaches to gradient-based thresholding. *Pattern Recognition Letters*, 14(7), 567–572. doi:10.1016/0167-8655(93)90107-O
- Hua, Q. M., Luo, S. H., Qiao, Y., & Qian, G. Y. (2008). Supervised grayscale thresholding based on transition regions. *Image and Vision Computing*, 26(12), 1677–1684. doi:10.1016/j.imavis.2008.05.003
- Liu, S. L., & Yang, J. Y. (2007). Survey on extraction methods of transition region. *Engineering and Science*, 9(9), 89–96.

Wang, H., Gao, J. B., & Huang, F. (2005). Extraction of transition region of image based on ramp width reduction. *Journal of Shenzhen University Science and Engineering*, 22(1), 7–11.

Watanabe, S. (1974). An automated apparatus for cancer prescreening: CYBEST. *CVGIP: Computer Graphics and Image Processing*, 3(4), 350–358.

Weszka, J. S., & Rosenfeld, A. (1979). Histogram modification for threshold selection. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1), 38–52. doi:10.1109/TSMC.1979.4310072

Yan, C. X. (2005). Two basic notes on image transition region extraction and segmentation method. *Computer Engineering and Design*, 26(10), 2589–2590.

Zhang, A. H. (2004). A novel fast algorithm of mono transition region determination with gray image. *Wuhan University Journal of Natural Sciences*, 9(3), 309–312. doi:10.1007/BF02907884

Zhang, Y.-J. (1989). *Development of a 3-D Computer Image Analysis System, Application to Megakaryocyte Quantitation*. Liège University.

Zhang, Y.-J. (1996). Transition region and image segmentation. *Tien Tzu Hsueh Pao*, 24(1), 12–17.

Zhang, Y.-J. (2001). *Image Segmentation*. Beijing: Science Press.

Zhang, Y.-J. (2006). *Advances in Image and Video Segmentation*. IIR Press. doi:10.4018/978-1-59140-753-9

Zhang, Y.-J. (2009). *Image Engineering: Processing*. Cengage Learning, Singapore: Analysis, and Understanding.

Zhang, Y.-J. (2015a). A survey on transition region-based techniques for image segmentation. *Journal of Computer-Aided Design and Computer Graphics*, 27(3), 379–387.

Zhang, Y.-J. (2015b). Half century for image segmentation. *Encyclopedia of Information Science and Technology*. 3rd Edition, 5906–5915 doi:10.4018/978-1-4666-5888-2.ch584

Zhang, Y.-J., & Gerbrands, J. J. (1991). Transition region determination based thresholding. *Pattern Recognition Letters*, 12(1), 13–23. doi:10.1016/0167-8655(91)90023-F

ADDITIONAL READING

Li, Z., Liu, G., Zhang, D., & Xu, Y. (2016). Robust single-object image segmentation based on salient transition region. *Pattern Recognition*, 52, 317–331. doi:10.1016/j.patcog.2015.10.009

Parida, P., & Bhoi, N. (2016). Transition region based single and multiple object segmentation of gray scale images. *Engineering Science and Technology, an International Journal*. 19(3):1206–1215

Peng, B., & Li, T. R. (2013). A probabilistic measure for quantitative evaluation of image segmentation. *IEEE Signal Processing Letters*, 20(7), 689–692. doi:10.1109/LSP.2013.2262938

Vojodi, H., Fakhari, A., & Eftekhari, M. et al.. (2013). A new evaluation measure for color image segmentation based on genetic programming approach. *Image and Vision Computing*, 31(11), 8770886. doi:10.1016/j.imavis.2013.08.002

Zhang, Y.-J. (2015c). A review of image segmentation evaluation in the 21st century. *Encyclopedia of Information Science and Technology*, 3rd Edition, Chapter 579 (5857–5867).

Zhang, Y.-J. (2015d). Statistics on image engineering literatures. *Encyclopedia of Information Science and Technology*, 3rd Edition, Chapter 595 (6030–6040). doi:10.4018/978-1-4666-5888-2.ch595

Zhang Y.-J. (2016). Image Engineering in China: 2015. *Journal of Image and graphics*, 21(5): 533–543.

KEY TERMS AND DEFINITIONS

Clip Transform: One operation performed on the gray values of pixels. Classic clip transform assigns the maximum or minimum gray values to the cutting parts. In transition region determination based image segmentation, a particular clip transform is designed, which assigns the clip values to the cutting parts to avoid the effects caused by too big contrast at cutting edge, and to reduce the influence of various interferences.

Effective Average Gradient (EAG): A special expression for or a particular parameter of the average gradient of images, in which only those pixels with non-zero gradient values are involved in the computation. Since the influence of zero gradient pixels is removed, so it is called “effective”. It is a selective statistics of images, and the basis of transition region determination.

Image: An entity that was captured by some visual systems in looking at the real world and that can be sensed to produce perception. It is a representation, likeness, or imitation of an object or thing, a vivid or graphic description, something introduced to represent something else.

Image Analysis (IA): One of three layers of image engineering, which is concerned with the extraction of information (by meaningful measurements with descriptive parameters) from an image (especially from interesting objects).

Image Engineering (IE): An integrated discipline/subject comprising the study of all the different branches of image and video techniques. As a general term for all image techniques, it could be considered as a broad subject encompassing mathematics, physics, biology, physiology, psychology, electrical engineering, computer science, automation, etc. Its advances are also closely related to the development of telecommunications, biomedical engineering, remote sensing, document processing, industrial applications, *etc.*

Image Segmentation: A process consists of subdividing an image into its constituent parts and extracting these parts of interest (objects) from the image.

Image Techniques: A collection of various branches of techniques for processing (such as acquiring, capturing, sensing, storing, enhancing, filtering, deblurring, inpainting, transforming, coding, transmitting, manipulating, *etc.*) analyzing (such as segmenting, representing, describing, featuring, measuring, classifying, recognizing), and understanding (such as modeling, registering, matching, reconstructing, training, learning, reasoning, interpreting, *etc.*) images.

Thresholding: Thresholding techniques are the most popularly used segmentation techniques. A set of suitable thresholds need to be first determined, and then the image can be segmented by comparing the pixel properties with these thresholds.

Development of Image Engineering in the Last 20 Years



Yu-Jin Zhang

Tsinghua University, China

INTRODUCTION

In recent years, images become more and more an important medium for human beings to observe the information of the real world around. Images can be obtained by using different observing and capturing systems in various forms and manners. The human visual system is just a typical example of an observation system. In its general sense, the word “image” could include all entities that can be visualized by human eyes, such as a still image or picture, a clip of video, as well as animations, cartoons, charts, drawings, graphics, paintings, even also texts and visual patterns, etc. In the early days, images were called pictures. Nowadays, with the progress of information science and society, the term “image” rather than “picture” is generally used because computers store numerical images of a picture or scene.

Image techniques are those techniques that have been invented, designed, implemented, developed, and utilized to treat various types of images for different and specified purposes (Zhang, 2009b). The researches and applications based on images are turned out to be a hot spot in the modern information society. They have attracted more and more attention in recent years with the fast advances of mathematic theories and physical principles, as well as the progress of computer and electronic devices, etc.

Image engineering (IE), an integrated discipline/subject comprising the study of all the different branches of image techniques, which has been formally proposed and defined around 20 years ago (Zhang 1996a; Zhang 1996c) to cover

the whole domain, has been evolved very quickly. Image engineering is now a very broad subject encompassing mathematics, physics, biology, physiology, psychology, electrical engineering, computer science, automation, etc. Its advances are also closely related to the development of telecommunications, biomedical engineering, remote sensing, surveying and mapping, document processing, industrial applications, etc.

Twenty years ago, a bibliography series on IE have been started. With a set of carefully selected journals and a thoroughly reading on the papers published, several hundreds of papers related to IE are chosen each year for further classification and statistical analysis. This work has been made already for consecutive 20 years (Zhang, 1996a; Zhang, 1996b; Zhang, 1997; Zhang, 1998; Zhang, 1999; Zhang, 2000; Zhang, 2001; Zhang, 2002; Zhang, 2003; Zhang, 2004; Zhang, 2005; Zhang, 2006; Zhang, 2007; Zhang, 2008a; Zhang, 2009a; Zhang, 2010; Zhang, 2011; Zhang, 2012; Zhang, 2013; Zhang, 2014a; Zhang, 2015). The last summary of this survey series can be found in (Zhang, 2014b).

The main purpose of this survey work is triple, that is, to capture the up-to-date development of IE, to make available a convenient means of literature searching facility for readers working in related areas, and to supply a useful reference for the editors of journals and potential authors of papers. This series summarize the selected papers for each year, analyzes the distributions of the selected papers from various sources and provides various statistics about the classified papers in each subject group. This paper will present an

overview of this survey series by showing the ideas behind and considerations on this work, as well as the comprehensive statistics obtained from this work. Some insights from it are also discussed.

BACKGROUND

For image engineering that is a new discipline, its scope is first described.

IE, from a perspective more oriented to techniques, could be referred to as the collection of three related and partially overlapped groups of image techniques, that is, Image Processing (IP) techniques (in its narrow sense), Image Analysis (IA) techniques and Image Understanding (IU) techniques. In a structural sense, IP, IA and IU build up three inter-connected layers of IE as shown in Figure 1. Each of them operates on different elements (IP's operand is pixel, IA's operand is object, and IU's operand is symbol) and works with altered semantic levels (from low at IP, via middle at IA, and to high at IU). The three layers follow a progression of increasing abstractness (left up arrow) and of decreasing compactness (right down arrow) from IP to IA to IU.

The techniques covered by IP primarily include the acquisition, representation, compression, enhancement, restoration and reconstruction of images. While IP is concerned with the manipulation of an image to produce another (improved) image, the techniques covered by IA are more concerned with the extraction of information from an image

(especially from the objects in it). Compared to IP that takes an image as input and outputs also images, IA takes also an image as input but outputs data extracted from input. Here, the data can be the measurement results associated with specific image properties or the representative symbols of certain object attributes. Based on IA, IU refers to a body of knowledge used in transforming the data into certain commonly understood descriptions, and for making subsequent decisions and actions according to the interpretation of the image contents.

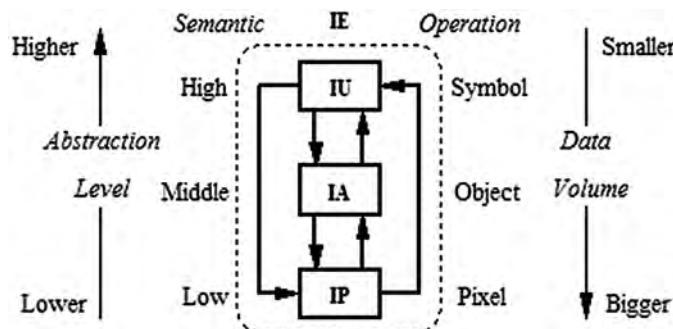
STATISTICS ON IMAGE ENGINEERING LITERATURES

After many years of development, IE (including IP, IA, and IU) has been greatly progressed. What is the current "picture" of IE? Answering this question is the foremost intention of this survey series. For such a purpose, the selection of suitable reference sources and the classification of selected references according to their contents are two important factors. Also for such a purpose, two statistics made by this survey are illustrated in the following.

Selection of Reference Source

As with any other emerging discipline, a large number of references related to IE have been published worldwide. They appear in books

Figure 1. Three layers of image engineering



(encyclopedia, handbook, monograph, and textbook), citation (secondary literature), conference proceedings, journals (periodicals), newspaper, patents, standard, theses (dissertations), yearbook, etc. The most popular ones are books, conference proceedings, and journals. Each of them has its own characteristics. Considering the fast publishing rate, the conference proceedings would be ranked first, followed by journals and then books. Considering the comprehensiveness, the books would be ranked first, followed by journals and then conference proceedings. Considering the quality, coverage, and regularity (which is important for a yearly survey), journals would be ranked higher than that of conference proceedings and books. Combining all these consideration, journals would be a better choice for such a survey series.

Classification Scheme

The classification scheme used in the bibliography series should reflect the contents of references.

A classification problem can be considered as a problem of partitioning a set into subsets. An appropriate classification of references into groups and/or sub-groups should satisfy the following four conditions:

1. Every reference must be in a group.
2. All groups together could include all references.
3. The references in the same group should have some common properties.
4. The references in different groups should have certain distinguishing properties.

Taking into consideration the above four conditions and the current status of IE development in the field, a complete and compact classification of the theories and techniques of IE is proposed (Zhang, 1996a; Zhang, 1996b), evolved and modified with the progress of domain (Zhang, 2001; Zhang, 2006). The current one is listed in Table 1. As a bibliography series, not only the three

Table 1. Classification scheme of literatures in image engineering

Group	Sub-Group
IP: Image Processing	P1: Image capturing (including camera models and calibration) and storage P2: Image reconstruction from projections or indirect sensing P3: Filtering, transformation, enhancement, restoration, inpainting, quality assessing, <i>etc.</i> P4: Image and/or video coding/decoding and international coding standards P5: Image digital watermarking, forensic, image information hiding, <i>etc.</i> P6: Image processing with multiple-resolutions (super-resolution, decomposition and interpolation, resolution conversion, <i>etc.</i>)
IA: Image Analysis	A1: Image segmentation, detection of edge, corner, interest points, A2: Representation, description, measurement of objects (bi-level image processing) A3: Analysis and feature measurement of color, shape, texture, position, structure, motion, <i>etc.</i> A4: (2-D) object extraction, tracking, discrimination, classification and recognition A5: Human face and organ (biometrics) detection, location, identification, categorization, <i>etc.</i>
IU: Image Understanding	U1: (Sequential, volumetric) image registration, matching and fusion U2: 3-D modeling, representation and real world/scene recovery U3: Image perception, interpretation and reasoning (semantic, expert system, machine learning) U4: Content-based image and video retrieval (in various levels, related annotation) U5: Spatial-temporal technology (high-dimensional motion analysis, 3-D gesture detection, tracking, manners judgment and behavior understanding, <i>etc.</i>)
TA: Technique Applications	T1: System and hardware, fast algorithm implementation T2: Telecommunication, television, web transmission, <i>etc.</i> T3: Documents (texts, digits, symbols) T4: Bio-medical imaging and applications T5: Remote sensing, radar, surveying and mapping T6: Other application areas
SS: Summary and Survey	S1: Cross category summary (combination of image processing / analysis / understanding)

layers of IE are distinguished, the technique application is also to be discriminated. In addition, the survey articles should belong to an impartial group. Consequently, this scheme classifies the literatures in image engineering firstly into five groups and further into 23 sub-groups. It is easy to verify that the above four conditions are fulfilled by this classification scheme.

Summary Over Years

The first statistic made from this survey series is a summary of the number of publications in the last twenty years, and also the classification of IE publications in each group, as shown in Table 2.

In Table 2, the total number of papers published in the selected journals (#T), the number of papers

selected for survey as they are related to IE (#S), and the selection ratio (SR), which equals to #S / #T, for each year have been provided. In addition, the paper numbers for five groups (and their percentages in the year) are also listed.

Some interesting points can be noted from Table 2:

1. This work has attended a quite large scale, with nearly 50500 papers involved and more than 12000 papers selected and classified (more than 50 papers per month).
2. IE is an (more and more) important topic for electronic engineering, computer science and automation. The average SR is more than 1/5 (it is more than 1/4 in several recent years),

Table 2. Summary over the last twenty years

Year	#T	#S	SR	IP	IA	IU	TA	Survey
1995	997	147	14.74	36(24.5%)	51(34.7%)	14(9.52%)	46(31.3%)	0
1996	1205	212	17.59	52(24.5%)	72(34.0%)	30(14.2%)	55(25.9%)	3(1.42%)
1997	1438	280	19.47	104(37.1%)	76(27.1%)	36(12.9%)	60(21.4%)	4(1.43%)
1998	1477	306	20.72	108(35.3%)	96(31.4%)	28(9.15%)	71(23.2%)	3(0.98%)
1999	2048	388	18.95	132(34.0%)	137(35.3%)	42(10.8%)	73(18.8%)	4(1.03%)
2000	2117	464	21.92	165(35.6%)	122(26.3%)	68(14.7%)	103(22.2%)	6(1.29%)
2001	2297	481	20.94	161(33.5%)	123(25.6%)	78(16.2%)	115(23.9%)	4(0.83%)
2002	2426	545	22.46	178(32.7%)	150(27.5%)	77(14.3%)	135(24.8%)	5(0.92%)
2003	2341	577	24.65	194(33.6%)	153(26.5%)	104(18.0%)	119(20.6%)	7(1.21%)
2004	2473	632	25.60	235(37.2%)	176(27.8%)	76(12.0%)	142(22.5%)	3(0.47%)
2005	2734	656	23.99	221(33.7%)	188(28.7%)	112(17.1%)	131(20.0%)	4(0.61%)
2006	3013	711	23.60	239(33.6%)	206(29.0%)	116(16.3%)	143(20.1%)	7(0.98%)
2007	3312	895	27.02	315(35.2%)	237(26.5%)	142(15.9%)	194(21.7%)	7(0.78%)
2008	3359	915	27.24	269(29.4%)	311(34.0%)	130(14.2%)	196(21.4%)	9(0.98%)
2009	3604	1008	27.97	312(31.0%)	335(33.2%)	139(13.8%)	214(21.2%)	8(0.79%)
2010	3251	782	24.05	239(30.6%)	257(32.9%)	136(17.4%)	146(18.7%)	4(0.51%)
2011	3214	797	24.80	245(30.7%)	270(33.9%)	118(14.8%)	161(20.2%)	3(0.38%)
2012	3083	792	25.69	249(31.4%)	272(34.3%)	111(14.0%)	151(19.1%)	9(1.14%)
2013	2986	716	23.98	209 (29.2%)	232 (32.4%)	124 (17.3%)	146 (20.4%)	5 (0.70%)
2014	3103	822	26.49	260 (31.6%)	261 (31.8%)	121 (17.7%)	175 (21.3%)	5 (0.61%)
Total	50478	12126	—	3923 (32.35)	3725 (30.72)	1802 (14.86)	2576 (21.24)	100 (0.83)
Average	2524	606	24.02	196	186	90	129	5

which is remarkable in considering the wide coverage of these journals.

3. IE publication evolves quite steadily. From Table 2, #S has been increased a lot these year. For example, its value in 2009 has attended near 7 times bigger compared to that in the starting year (while the #T in the same year is only 3.6 times bigger compared to that in the starting year). It is clear that the literature for IE evolves fast than other disciple/subjects covered in these journals.
4. The growing rates of publications for IP, IA, IU and TA are different. To make it clear, Figure 2 shows the numbers of publications for these four groups graphically. The four curves in Figure 2 run quite smoothly and have not intercrossed in last seven years (becoming stable), with the rank from top to bottom: IA, IP, TA and IU.
5. The number of publications for IP and IA constitute near 2/3 of the total number of IE publications. This shows that the current research focus of IE are on IA and IP. In contrast, research work on IU needs to be promoted.

Detailed Statistics on Sub-Groups

The second statistic is a detailed classification of IE publications in each sub-group and for each year. The results are listed in Table 3.

Many types of information could be obtained from Table 3; however, only three important observations are pointed out here:

1. From the number of publications in different sub-groups, the top five sub-groups are image segmentation and edge detection (A1), various typical processing techniques (P3), object extraction and recognition (A4), image registration, matching and fusion (U1), and image compression (P4). They are the most important research topics in all these years. However, their development tendencies are different. Since there is no general theory on image segmentation, the research in this sub-group always attracts many attentions, so the number of A1 is continuous rising. On the other side, P4 is continuous declining as researches become matured, just as predicted in (Zhang 2008b).

Figure 2. Number variation of the first four groups of IE literatures in selected publications for last twenty years

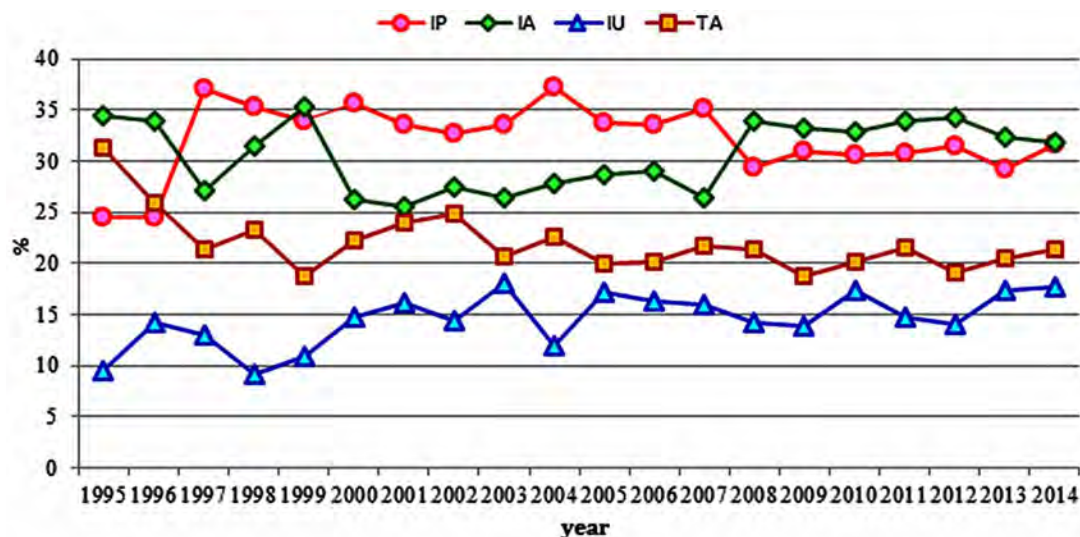
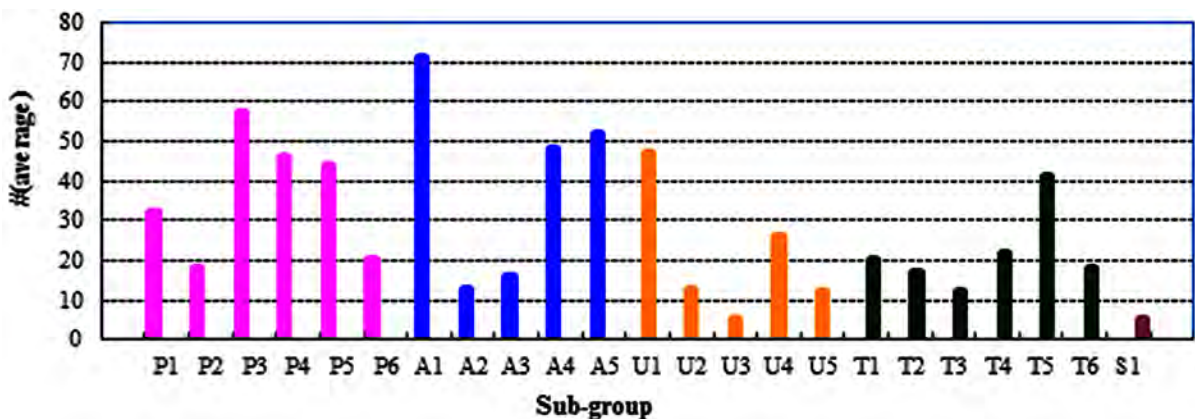


Table 3. Detailed classification of IE publications in sub-groups over the last twenty years

Year	P1	P2	P3	P4	P5	P6	A1	A2	A3	A4	A5	U1	U2	U3	U4	U5	T1	T2	T3	T4	T5	T6	S1
1995	3	9	20	4	—	—	24	9	8	10	—	5	8	1	—	—	9	13	6	13	5	—	—
1996	6	9	15	22	—	—	27	16	11	18	—	13	15	2	—	—	8	6	5	17	11	8	3
1997	11	13	15	65	—	—	31	16	5	24	—	14	19	3	—	—	14	9	14	7	6	10	4
1998	15	13	23	57	—	—	42	19	7	28	—	13	9	6	—	—	14	7	15	8	19	8	3
1999	19	10	29	74	—	—	62	24	19	32	—	26	13	3	—	—	21	9	10	18	10	5	4
2000	20	14	36	75	20	—	52	14	14	28	14	32	18	2	16	—	18	13	14	19	20	19	6
2001	27	7	30	66	31	—	49	18	13	28	15	36	15	6	21	—	22	12	15	18	27	21	4
2002	31	14	49	50	34	—	56	15	11	34	34	36	10	4	27	—	17	23	24	28	17	26	5
2003	22	16	47	60	49	—	63	7	16	29	38	56	14	3	31	—	19	15	16	17	25	27	7
2004	36	16	77	49	57	—	70	19	28	28	31	38	11	2	25	—	16	25	20	25	34	22	3
2005	33	15	58	52	48	15	83	10	15	48	32	54	12	3	27	16	26	31	10	13	33	18	4
2006	35	19	68	57	49	11	82	11	13	46	54	62	12	4	25	13	20	35	16	15	40	17	7
2007	48	38	79	54	71	25	90	9	21	59	58	72	20	7	36	7	30	32	14	28	63	27	7
2008	28	25	82	53	61	20	111	15	27	88	70	69	10	5	38	8	20	19	17	37	80	23	9
2009	53	25	102	46	60	26	102	4	26	96	107	83	11	4	22	19	29	24	13	45	84	19	8
2010	46	17	80	38	43	15	95	9	18	75	60	71	15	9	31	10	18	13	10	25	59	21	4
2011	30	29	102	32	30	22	107	4	15	85	59	72	8	7	28	3	24	19	4	19	71	24	3
2012	56	19	78	26	46	24	91	14	20	76	71	58	12	7	20	14	23	12	3	31	67	15	9
2013	36	26	69	29	35	14	91	5	17	56	63	63	16	7	20	18	14	10	4	31	72	15	5
2014	78	22	87	20	28	25	82	12	22	78	67	60	14	5	26	16	31	9	7	28	70	30	5
Summary	633	356	1146	929	662	197	1410	250	326	966	773	933	262	90	393	124	393	336	237	442	813	355	100

Figure 3. Numbers of publications per year for different sub-group



- Listing the top ranking in each group could also reveal the current research focus. In IP, P3 and P4 are ranked in the top. In IA, except A1, A4 has also obtained a high score. These two sub-groups are closely related in the sense of object detection with object model, the difference is that the former is more like an un-supervised task and the latter is more supervised. Some connections existed between the top one in IU, U1, and the top one in TA, T5. They are both related to remote sensing, the difference is that the former is more focused on theoretical aspects while the latter is more concentrated on application aspects
- Finally, most of the summary in Table 3 are made for all twenty years. However, some sub-groups have only been added to survey in recent years, along the techniques covered appearing with the evolution. Considering this factor, the average numbers of publications per year for different sub-group are supplied in a histogram form in Figure 3. Compared to Table 3, P5 and A5 provide another two promising recent research directions.

A COMPARISON OF FOUR PERIODS OF “FIVE-YEARS”

This survey series started since 1995, and has lasted already 20 years. These 20 years can be divided into four periods of “five-years”. The summary information of selecting image engineering literatures for these four periods of “five-years” are given in Table 4, from column 2 to column 5, respectively. Moreover, the total aggregate data and annual average data for these 20-years are also provided separately in column 6 and column 7. To compare and comprehend trends conveniently, the ratio of second “five-years” over first “five-years”, the ratio of third “five-years” over second “five-years”, and the ratio of fourth “five-years” over third “five-years”, are also computed as listed in column 8 to column 10, respectively. In which, the ratio bigger than 1 represents an increasing development, and the ratio smaller than 1 represents a decreasing development.

The discussions for each line of Table 4 are as follows:

- The total numbers of papers in the first three periods of “five-years” are increasing gradually, but a slight decrease is appearing in the last period of “five-years”. In recent years, the length of many papers are increased, the volume of journal issue is thus relatively reduced.

Table 4. Statistics of four “five-years” selections for publications of image engineering

	First “Five- Years”	Second “Five- Years”	Third “Five- Years”	Fourth “Five- Years”	Sum Over 20 Years	Average Per Year	Second / First	Third / Second	Fourth / Third
# of Total Publications	7165	11654	16022	15637	50478	2524	1.63	1.37	0.98
# of Selected Publications	1333	2699	4185	3909	12126	606	2.03	1.55	0.93
Selection Ratio (%)	18.60	23.16	26.12	25.00	24.02	24.02	1.24	1.13	0.96

- The numbers of selected papers in the first three periods of “five-years” are also increasing gradually, and similarly a slight decrease is appearing in the last period of “five-years”. Apparently, image engineering entered a relatively stable period of development.
- The selection ratios in the first three periods of “five-years” are equally increasing gradually, with a slight decreasing in the last period of “five-years”. The interpretation should be substantially the same as (2).

In order to be more intuitive and clear, Figure 4 to Figure 7 give the classification and comparison of the numbers of papers in 22 sub-groups (except E1) for these four periods of “five-years”. Figure 4 to Figure 7 correspond to groups P, A, U, T, respectively.

In Figure 4 to Figure 7, each bar represents the sum of the five-years, so the analysis for the variations and trends should be more reliable and credible. From the viewpoint of individual technologies, the following points should be noted:

From the viewpoint of individual technologies, the following points should be noted:

- Among the 17 sub-groups began from the first year (existed 20 years ago), four cases can be distinguished. There are six sub-groups with statistical data increasing every stage. They are P1, P3, A1, A4, U3, and T5. These research directions are still rising and expanding in depth and breadth. There are eight sub-groups with statistical data decreasing in one stage. They are P2, A3, U1, U2, T1, T2, T4 and T6. Do these fluctuations

Figure 4. Comparison of the numbers of publications of group P for different techniques in four periods of “five-year”

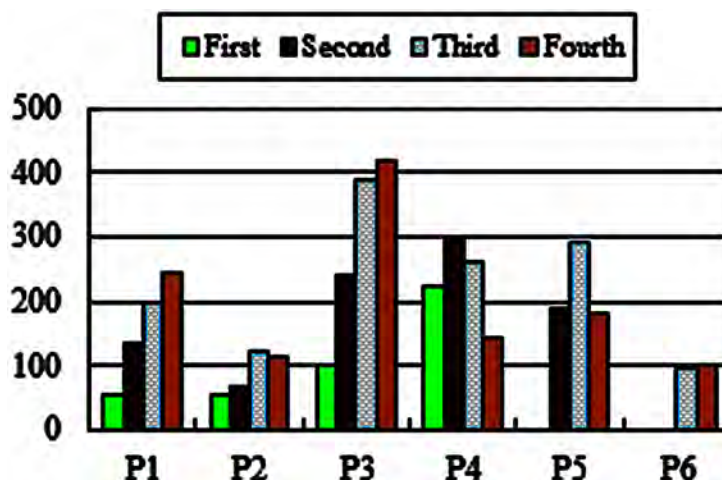


Figure 5. Comparison of the numbers of publications of group A for different techniques in four periods of “five-year”

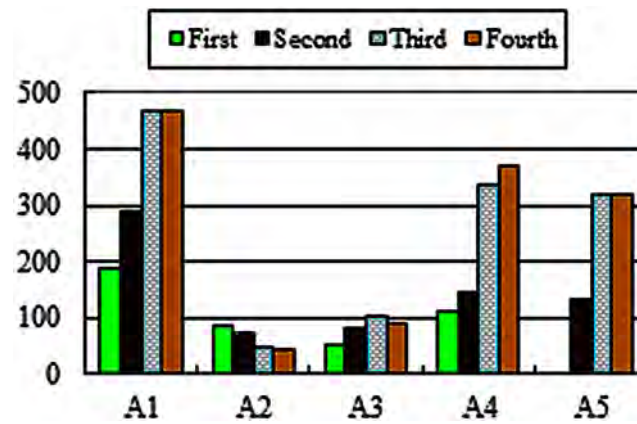


Figure 6. Comparison of the numbers of publications of group U for different techniques in four periods of “five-year”

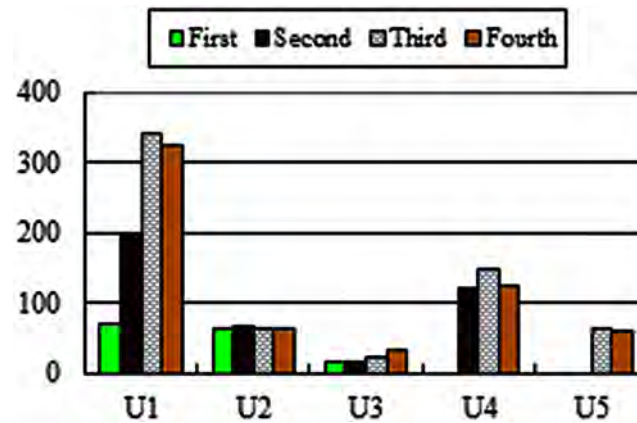
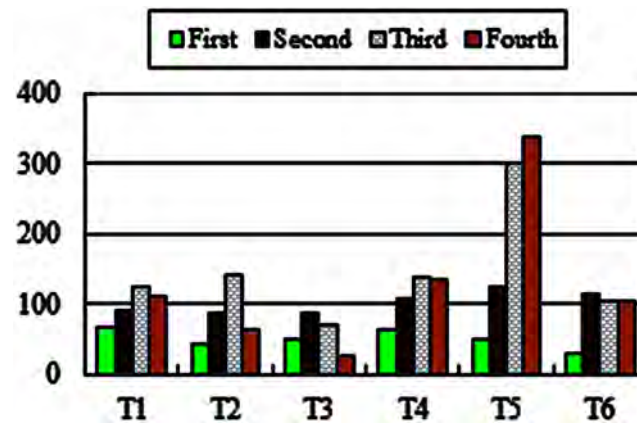


Figure 7. Comparison of the numbers of publications of group T for different techniques in four periods of “five-year”



indeed reflecting the rise and fall of the corresponding research directions? More time is needed to make judgment. There are two sub-groups with statistical data decreasing in two stages. They are P4 and T3. These research directions are relatively matured and widely applied, so the gradual reduction of new researches and results is expected. There is one sub-groups with statistical data decreasing in all three stages. This is A2. It may be disappear in the future.

2. Among the 3 sub-groups began from the second period of “five-years”, sub-group A5 maintained continuously increasing in two stages, and the absolute numbers have been kept at a high level. The other 2 sub-groups are P5 and U4. They are increasing in the second stage and decreasing in the third stage. Considering the coverage of research domain, sub-group A5 should maintain the vitality of a period of time; sub-group P5 may be leveling off; sub-group T4 should have some research space in theory and have to be pioneered in the applications.
3. Among the 2 sub-groups began from the third period of “five-years”, sub-group P6 is increased slightly and sub-group U5 is decreased slightly. As they are relatively new, related researches are still deepening, so it needs to be judged according to the development trend of the future situation.

FURTHER RESEARCH DIRECTIONS

It could be predicted that, from a research point of view, the IE research has a tendency that is going from low level to high level, that is, from image processing to image understanding, via image analysis. As shown in Figure 2, after a chaos in the beginning, Image Processing publications went to top in the early years, and now the Image Analysis overpasses it for several years. It

is expected that more research and application requirements would make Image Understanding goes up fast in near future.

The field of IE has changed enormously in recent years. Many techniques have been developed, exploited or applied only in this century. It is seen that techniques for IE being implemented and used on a scale few would have predicted a few years ago. It is also likely that these techniques will find many new applications in the future.

Viewing the prospective of IE, the work for survey on IE could also be pushed deeply, at least, in two ways. First, since this survey provides an up-to-date picture regarding IE and its research advance, so further research could be advanced and promoted in appropriate directions. Second, according to the principles and methods of bibliometrics, a systematic investigation of the factors of the articles indexed in the survey series could be made.

CONCLUDING REMARKS

This chapter shows an overview of a survey series on IE made in the last twenty years. The idea behind this survey, as well as a thorough summary of obtained statistics for this survey series are illustrated and discussed. All these provide much of useful information regarding the tendency of fast progresses of IE in China and worldwide. On view of the helpfulness of this series for the research community, the 21st and 22nd in this series have also been made (Zhang, 2016; Zhang 2017).

Such a work not only provides a convenient means for literature searching in IE, but also presents a detailed picture of hot research topics in the field. Moreover, it may be useful for publishers who want to quickly capture the general trends of development in IE, and for potential authors who wish to disseminate widely their research results in the associated communities.

ACKNOWLEDGMENT

This work has been supported by the National Natural Science Foundation under Grants NNSF-61171118 and the Ministry of Education under Grants SRFDP-20110002110057.

REFERENCES

- Zhang, Y. J. (1996a). Image engineering in China: 1995. *Journal of Image and Graphics*, 1(1), 78–83.
- Zhang, Y. J. (1996b). Image engineering in China: 1995 (supplement). *Journal of Image and Graphics*, 1(2), 170–174.
- Zhang, Y. J. (1996c). Image engineering and bibliography in China. Technical Digest of ISIST'96, 158–160.
- Zhang, Y. J. (1997). Image engineering in China: 1996. *Journal of Image and Graphics*, 2(5), 336–344.
- Zhang, Y. J. (1998). Image engineering in China: 1997. *Journal of Image and Graphics*, 3(5), 404–414.
- Zhang, Y. J. (1999). Image engineering in China: 1998. *Journal of Image and Graphics*, 4(5), 427–438.
- Zhang, Y. J. (2000). Image engineering in China: 1999. *Journal of Image and Graphics*, 5A(5), 359–373.
- Zhang, Y. J. (2001). Image engineering in China: 2000. *Journal of Image and Graphics*, 6A(5), 409–424.
- Zhang, Y. J. (2002). Image engineering in China: 2001. *Journal of Image and Graphics*, 7A(5), 417–433.
- Zhang, Y. J. (2003). Image engineering in China: 2002. *Journal of Image and Graphics*, 8A(5), 481–498.
- Zhang, Y. J. (2004). Image engineering in China: 2003. *Journal of Image and Graphics*, 9(5), 513–531.
- Zhang, Y. J. (2005). Image engineering in China: 2004. *Journal of Image and Graphics*, 10(5), 537–560.
- Zhang, Y. J. (2006). Image engineering in China: 2005. *Journal of Image and Graphics*, 11(5), 601–623.
- Zhang, Y. J. (2007). Image engineering in China: 2006. *Journal of Image and Graphics*, 12(5), 753–775.
- Zhang, Y. J. (2008a). Image engineering in China: 2007. *Journal of Image and Graphics*, 13(5), 825–852.
- Zhang, Y. J. (2008b). A study of image engineering. Encyclopedia of Information Science and Technology (2nd ed.). Academic Press.
- Zhang, Y. J. (2009a). Image engineering in China: 2008. *Journal of Image and Graphics*, 14(5), 809–837.
- Zhang, Y. J. (2009b). *Image Engineering: Processing*. Cengage Learning, Singapore: Analysis, and Understanding.
- Zhang, Y. J. (2010). Image engineering in China: 2009. *Journal of Image and Graphics*, 15(5), 689–772.
- Zhang, Y. J. (2011). Image engineering in China: 2010. *Journal of Image and Graphics*, 16(5), 693–702.
- Zhang, Y. J. (2012). Image engineering in China: 2011. *Journal of Image and Graphics*, 17(5), 603–612.
- Zhang, Y. J. (2013). Image engineering in China: 2012. *Journal of Image and Graphics*, 18(5), 483–492.
- Zhang, Y. J. (2014a). Image engineering in China: 2013. *Journal of Image and Graphics*, 19(5), 649–658.

Zhang, Y. J. (2014b). Statistics on Image Engineering Literatures. Encyclopedia of Information Science and Technology (3rd ed.). Academic Press.

Zhang, Y. J. (2015). Image engineering in China: 2014. *Journal of Image and Graphics*, 20(5), 585–598.

Zhang, Y. J. (2016). Image engineering in China: 2015. *Journal of Image and Graphics*, 21(5), 533–543.

Zhang, Y. J. (2017). Image engineering in China: 2016. *Journal of Image and Graphics*, 22(5), 525–535.

KEY TERMS AND DEFINITIONS

Bibliometrics: Bibliometrics is a branch of science. It is a type of research method used in library and information science. It utilizes quantitative analysis and statistics to describe patterns of publication within a given field or body of literature.

Image: An entity that was captured by some visual systems in looking at the real world and that can be sensed to produce perception. It is a representation, likeness, or imitation of an object or thing, a vivid or graphic description, something introduced to represent something else.

Image Analysis (IA): One of three layers of image engineering, which is concerned with the extraction of information (by meaningful measurements with descriptive parameters) from an image (especially from interesting objects).

Image Engineering (IE): An integrated discipline/subject comprising the study of all the different branches of image and video techniques. As a general term for all image techniques, it could be considered as a broad subject encompassing mathematics, physics, biology, physiology, psychology, electrical engineering, computer science, automation, etc. Its advances are also closely related to the development of telecommunications, biomedical engineering, remote sensing, document processing, industrial applications, etc.

Image Processing (IP): One of three layers of image engineering, which encompasses processes whose inputs and outputs are both images, with outputs are improved version of inputs.

Image Techniques: A collection of various branches of techniques for processing (such as acquiring, capturing, sensing, storing, enhancing, filtering, deblurring, inpainting, transforming, coding, transmitting, manipulating, etc.) analyzing (such as segmenting, representing, describing, featuring, measuring, classifying, recognizing), and understanding (such as modeling, registering, matching, reconstructing, training, learning, reasoning, interpreting, etc.) images.

Image Understanding (IU): One of three layers of image engineering, which transforms data extracted from images into certain commonly understood descriptions, and makes subsequent decisions and actions according to the interpretation of the images.

Particle Shape Analysis Using Digital Image Processing



Katia Tannous

University of Campinas, Brazil

Fillipe Souza Silva

University of Campinas, Brazil

INTRODUCTION

The fluidization technology has been closely related to traditional processes, such as petrochemical, pharmaceutical, polymer, food, and incineration (Kunii & Levenspiel, 1991). In these processes, the characterization of solid particles is vital to obtain higher efficiencies and lower process costs. Recently, many clean energy processes (e.g.: mechanical processes, pyrolysis, combustion, and gasification) involving biomass (non-woody and woody plants, and organic wastes) need fundamental parameters to understand energy conversion (physical, chemical, thermal and biological) in obtaining bio-fuels (gas, liquid or solid), electricity or heat. All of them require versatility in a wide range of chemical reactor configurations and designs (Cui & Grace, 2007; Dai & Grace, 2011; Tannous e Lourenço, 2015).

Fundamental parameters are necessary to understand the dynamic behavior of solids in these reactors because the particles have irregular non-spherical shapes. According to Guo, Chen, and Liu (2012), biomass particles are extremely irregular due to their high contents of cellulose, hemicellulose and lignin. Mandø and Rosendahl (2010), and also, Tannous, Lam, Sokhansanj, and Grace (2013) reviewed fundamental definitions for solid particles such as mean diameter and shape.

Different definitions have been presented in the literature to calculate shape factors of solid particles based on two-dimensional (2D) and three-dimensional (3D) analyses. Depending on

the shape, two-dimension representations may be insufficient to represent the true particle shapes (Zavala, 2012; Rodriguez, Edeskär, & Knutsson, 2013; Tannous, Lam, Sokhansanj, & Grace, 2013). The geometry of biomass particles is complex because it depends of the type of raw material and process applied changing its size and shape from the beginning to the end of the process (Dai & Grace, 2011). There is a need to develop appropriate geometrical definitions as well as algorithms to describe particles shapes (Rodriguez, Johansson, & Edeskär, 2012).

This chapter will discuss new software, Particles and Geometric Shapes Analyzer (APOGEO) aiming the determination of major and minor axes, aspect ratio and sphericity of solid particles (mainly biomass) for four different definitions by image processing without any manual work. This software was developed to reach users from academy and industry.

BACKGROUND

The influence of biomass particle shapes cannot be ignored in particle transportation, mixing and fluidization processes. Various particle shapes result in different particle surface areas, in which are heat and mass transfer processes (Guo et al., 2012). Rodriguez et al. (2013) have presented a review about different methods and techniques to determine the geometrical shape of the particles. The authors observed that there is no agreement

on the usage of the descriptors and is not clear which descriptor is the best. A large scale shape classification has been a problem. In addition, the authors considered that image analysis is a promising tool; it presents advantages like low time consumption or repeatability.

When non-spherical solid particles are observed through a microscope, various methods can be used for their sizing, resulting in terms of an equivalent spherical particle. Projected images in microscopes (optical, scanning and transmission) are in two-dimensional and depend on the orientation of the particles (Turbitt-Daoust, Alliet, Kaye, & Matchett, 2000). Particles in a stable orientation tend to have a maximum area causing microscopic measurements larger values than those presented by other methods, i.e., when smaller particle sizes are discarded.

This technique requires the analysis of a number of particles statistically significant, which has needed the use of automatic image analysis programs, conducted with the aid of computers and specific software. This process includes several steps: 1) acquisition and image scanning, 2) pre- and post-processing of the scanned image, 3) measurements (shape, size and count), 4) analysis and data presentation (Papini, 2003).

Currently, different equipments can be found on the market, but a few of them can be use for biomass particles, for example: Olympus BX51 from Olympus Corporation; Morphologi G3 and Sysmex FPA3000 from Malvern Instruments Ltd; Camsizer and Camsizer XT from Retsch Technology GmbH; CILAS granulometer and expert Shape from Compagnie Industrielle des Lasers; Nikon E200 and ImageJ software, DOMAS Digital Imaging System from Fiber and Particle Engineering Laboratory (University of Oulu), and Eyecon™ from Innopharma Labs. These equipments calculate, in general, the aspect ratio, the roundness, the convexity, the elongation, and the linearity of the materials. Remarking that, the quality of the results is dependent on the skill and experience of the operator, and also the correct calibration of the microscope.

In view of the impossibility of obtain such equipments for economic reasons, in 2011, Tannous and Silva (2012) have decided to seek alternatives to solve many research problems concerning biomass particles through of development of their own software to obtain the shape and sphericity of these kind of particles. This laboratory has experience in educational software development (Rimoli, Assis, & Tannous, 2006; Maranesi & Tannous, 2009; Tannous & Rocha, 2012; Tannous & Santos, 2012).

DEVELOPMENT OF APOGEO® SOFTWARE

General Information

APOGEO® (Particles and Geometric Shapes Analyzer) software was developed to reach professionals in different area of expertise (e.g: chemist; physicist; and chemical, mechanical and food engineers). It is based on image processing technique in order to transform pixels in computational objects. This software was published in the National Institute of Industrial Property-Brazil (n° 13085-6) by Tannous and Silva (2012).

The software can quantify the major and minor axes correlating two or three dimensions of particles (e.g.: woods, coconut fiber, sugarcane bagasse, and rice husk) to obtain its sphericity. The particles can be associated with different geometries such as: rectangular parallelepiped, cylinder, oblate and prolate spheroids, and irregular. The results are presented in histograms and tables, but also can be saved in a spreadsheet (Excel file, OpenOffice Calc file).

Minimum Hardware Requirements

APOGEO was developed on personal computer sufficiently supplied with memory and processing of operational systems. The hardware chosen was a processor of 800 MHz, 1 GB of RAM memory, monitor of 15", mouse, keyboard, and Java Runtime Environment (version 1.6.0 or higher).

Methodology Applied

Particle Selection and Recognition

The development of APOGEO software for mapping images requires that it be able to manage with clustering and pattern recognition. Furthermore, it has to manage with both filters to improve image quality, ensuring a valid background and also the impurities on the image plane. According to Starck, Murtagh, and Fadili (2010), the technique of sparse matrices is an efficient solution to solve these problems. This approach is widely used to identify the size and shape of minerals through images by microscopy.

Therefore, using this mathematical technique, it became possible to approach our research of pattern recognition as a problem of sparse matrix assembly. After that, the new task was to transform these sparse matrix computational objects in an easy access and of the manipulation. One solution was found using the recommendation by Cormen, Leiserson, Rivest, and Stein (2002), concerning on the isolation of particles, and saving them into data structures.

The results of this technique were satisfactory, since that became easy to manipulate the data. In terms of development, the software program was simplified allowing the reuse of the data for later processing such as filtering impurities, manually delete particles, and adjusting of the contrast level with the background image.

Image Processing

The result obtained from the image processing should be a group of particles expressed by a set of linked data structures (linked list). However, the transformation of an image into a set of particles is not an elementary task. Various approaches were used in order to avoid memory bottlenecks that would risk the program functionality. Moreover, complex libraries for image processing could not be applied here, because in some of them the portability requirement of the software would be

impaired. In other cases, they would have a considerable increase in the space in disk occupied.

The solution was to use a very simple technique (Cormen et al., 2002), the recurrence or recursion models. The application of this model was given by route taken in account the map of a particle. Once the user find a pixel of a potential particle, one must go through this whole neighborhood, without losing the referential. This approach is called backtracking (McAllister, 2009).

However, it presents a downside of consuming a lot of memory and processing, since the use of recursive functions in programming creates a memory overhead, leading to a stack overflow, which would risk the usability of software. Thus, the maximum image size would be 100x100 pixels, because there could be damaging to the operating system.

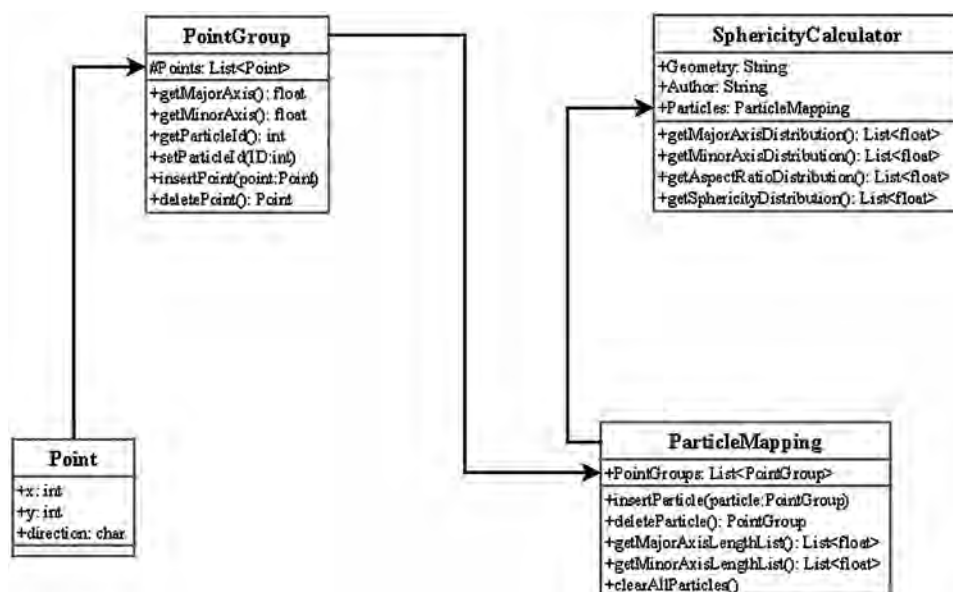
In order to enhance the software quality (APOGEO) and overcome its limitation, it was adopted a solution proposed by Abelson, Sussman, and Sussman (1996), which is the replacement of the recursive function model by a data structure called stack. Using this approach, the image processing time reduced more than 10 times, and reduced the file sizes with a maximum of 3200x3200 pixels.

Development Stages

Once defined the approaches for image processing and data storage, the software APOGEO was created considering different steps, such as: image processing, interaction with the end user, and statistical data analysis. Although they are quite distinct in terms of programming approaches, the use of a simplified model of classes by standard Singleton design enabled the interconnection between the three different demands. The use of this model (Gamma, Helm, Johnson, & Vlissides, 1994) led the project in a simplified pattern resulting in class diagram, as shown in Figure 1.

The initial stage of the program consisted in transforming the original image into a binary matrix. The determination of the values, zero (background) and one (potential solid particle),

Figure 1. Class diagram



was made through the comparison between the sum of the scales of each RGB pixel with a pre-defined value named contrast level. After that, with the binary matrix already established, its transformation in a particle group was defined through the flow diagram (Figure 2).

User Interface

The software interface construction was conducted using the Java Swing libraries and Java AWT, which are the libraries for building of Java Graphic User Interface (GUI). Through its simple and standard model becomes possible to construct similar functional interfaces on any operating system.

The APOGEO software was created using NetBeans software (version 6.9.1). It provides a framework for generation of Swing and AWT interfaces. Thus, the effort to generate production code for the graphical interface has been replaced by a dedicated work to create a more intuitive interface, which would provide a better user experience.

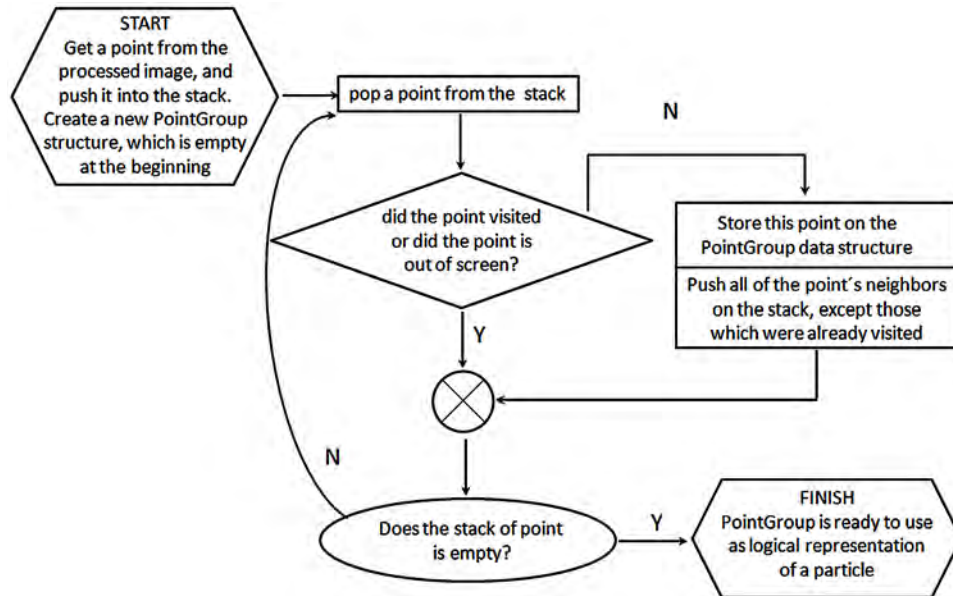
Once the graphical development tools were properly chosen, the final design consisted in exploring three scenarios:

1. A panel of image processing, consisting of the original and processed images, and tabs to change the contrast level, excluding impurities, and actual particle sizes;
2. A panel for selecting particles and equations, in which the user could actively interfere in the processes image by excluding some of those particles, and allowing the choice of the desired equations;
3. Statistical information, consisting of histograms, and other statistical data (mean, median, and standard deviation).

A problem inherent in the construction of these interfaces was the use of space to create a rich interface, simple and objective. Based on this, the software APOGEO has four main frames, side frames, and informational messages, in order to assist the user in different stages of the program.

Despite the reservations made to the AWT and Swing interfaces, they are still heavy in terms of processing, but even quite simple. In other words,

Figure 2. Program flow

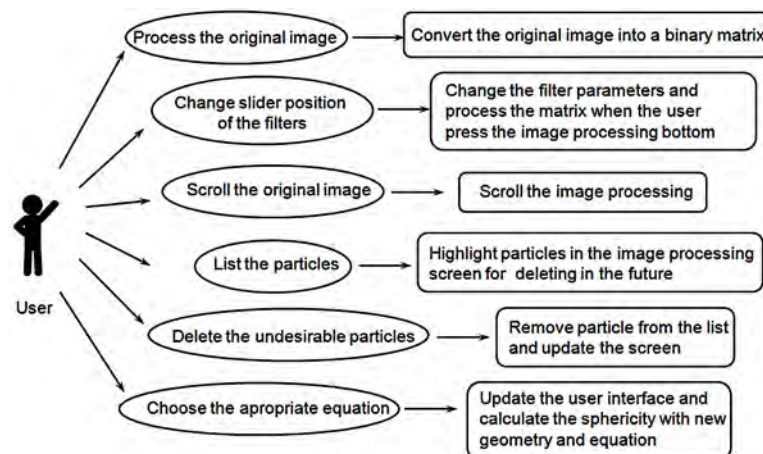


with the basic modules applied (Java Swing for Java AWT), if the user wants to create a more complex interface or even a graphical effects as transparency and fade, it becomes a difficult task since the framework does not provide a WSIWYG tool. The solution for using more complex libraries creates platform dependencies, which in turn hamper the portability of the software. Figure 3 shows the diagram which the user interacts with the APOGEO interface.

Testing and Reliability

According to Sommerville (2007), the software testing is as important as the software development, once it delivers to the end user one software with quality, avoiding usability, and coherence problems. Following the methodologies suggested by this author, the APOGEO software had two stages of quality assurance: unit and integration tests.

Figure 3. User interface design



Unit Tests

In order to test the program, the JUnit framework was used for some methods and classes. The Unit tests were applied for each empirical equation, in which every result was compared against the others of the adopted references. Moreover, for each routine of image processing there were several comparative tests. Although the computer measurements were executed in few seconds, the manual measurements take a long time to perform, once it had to be made using Technical Standardization. These standards involve several measurements with predefined instruments, such as pachymeter and micrometer.

Integration Tests

Once the unit tests have already concluded, the next step was to integrate both functional methods and the user interface in order to lead a final

product. In this step, some of the interface methods were tested, as well as all equations with different combinations.

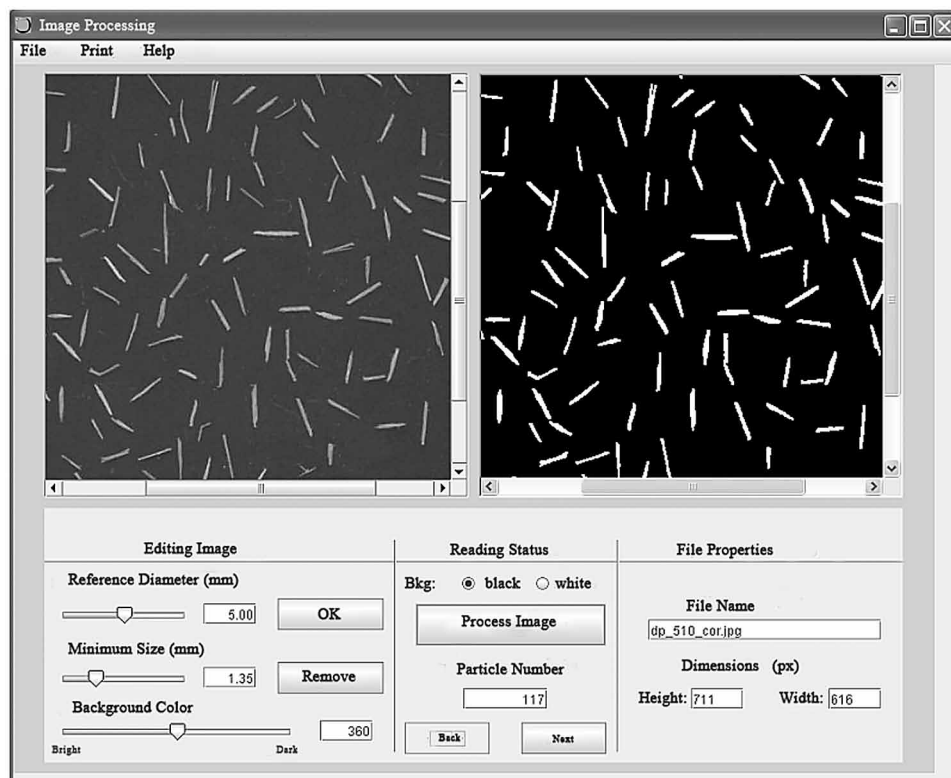
We also tested the program walkthrough in order to evaluate the help messages, errors treatment, histograms, and final results. We have also used it in the laboratory in order to evaluate the software performance and usability in research activities.

RESULTS AND DISCUSSION

Opening and Image Processing

The initial frame (Figure 4) shows the menu of image processing along with control buttons considering three parts: customize image, background and number of particles, and file properties. The buttons were placed in order to facilitate the user, in a just one run, to learn how to setup the entire

Figure 4. Program interface: Digital image capture and processing



environment before taking the measurements, without reading the manual over and over.

- The first step of the program is the opening image file. In order to improve the program functioning, the image should be in JPEG format and lower than 3200x3200 pixels. It is noteworthy that the larger the image, the longer will take its processing.
- The second step is the image processing. First of all, the user should choose the background image. If the particle is dark and the background is light in color, we adopt the choice WHITE, otherwise BLACK (default). Once the user has processed the image, it is possible to adjust the contrast level in order to obtain better results in comparison with the first image processing.
- The third step is to choose the particle size. In order to get the correct size, the user must have at least one reference object such as a coin. Because the user knows its dimensions, sets the maximum length of particle looking at the referenced object. By clicking in OK, the user confirms the operation.
- The fourth step is not required, but it is useful if the user wants to filter small particles or powder, that could influence the final results. Adjusting the minimum size, the user can remove powder or small particles from his sample.
- Taking into account the number of particles is very important because the particle images could be overlapping other particles compromising the final results. After following all these steps, click NEXT button.

If the procedure order is not followed, a warning window will appear indicating what details are missing.

Editing Image

The second frame (Figure 5) allows visualizing the reading process results together with the table data (box aside) to the processed image, removing undesired particles, choosing the particle geometry and equations for sphericity calculation, which are based on four references (authors) from the literature.

Waste Removal

The users could also delete the particle in this frame, in order to avoid particles on the screen. To remove them, just select it in the created list and click DELETE PARTICLE or press DELETE on the user keyboard. A warning will appear indicating which particle has been deleted. To check the result, just click on the underlying or adjacent element of the list.

The Choice of Particle Geometry and Equation

There are five different options of geometry in this software: cylindrical, oblate spheroid, prolate spheroid, rectangular parallelepiped and irregular. The geometries were associated with food and biomass particles (Figure 6) such as: (a) coconut fiber and sugarcane bagasse, (b) lentil, (c) rice husk, (d) Brazilian woods, (e) Hedgehog Brazil nut.

Four definitions are referenced in the software considering two (2D) and three dimensions (3D) of particles according to the traditional literature of fluidization technology: Wadell (1935), Riley (1941), Curray (1951) and Massarani and Peçanha (1986). The software measures two particle dimensions (A and B). To calculate the sphericity in three distinct dimensions of particles (e.g.: oblate and prolate spheroid, and rectangular parallelepiped), it is necessary to assign the value represented by the letter C (Figure 6). The user can choose two paths to introduce this value in the software: (1) measuring the third dimension of the sample by experimental work (e.g: 100 particles) (Farias,

Figure 5. Program interface: Editing image

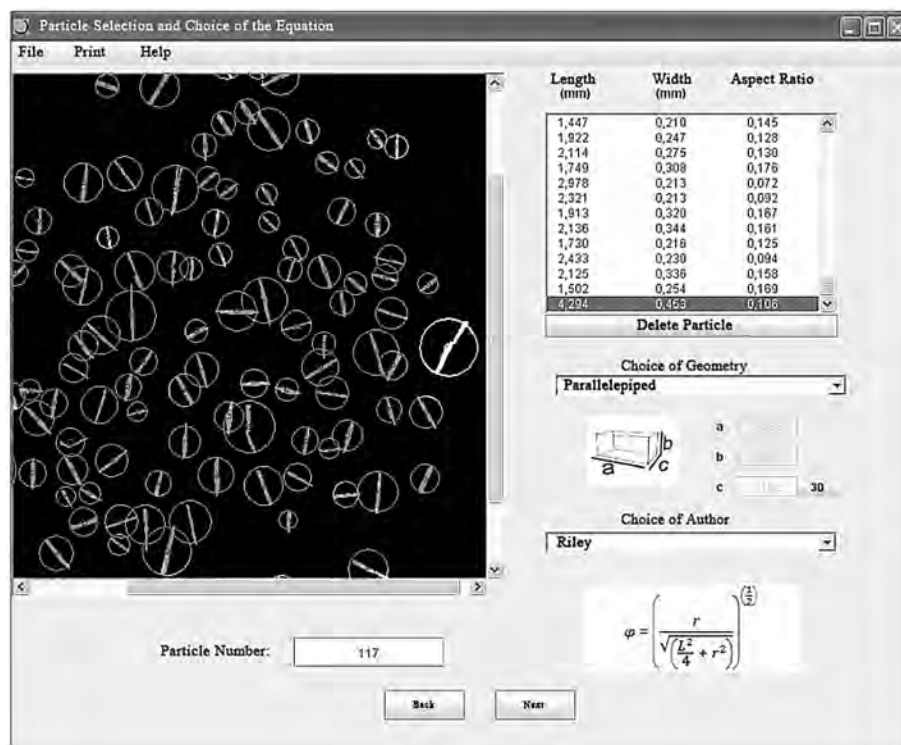
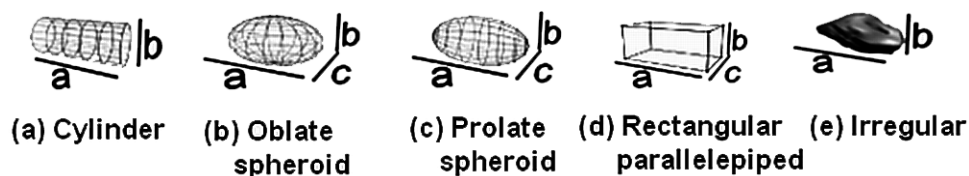


Figure 6. Particle geometry



2012; Lourenço, 2012) or (2) adopting the value suggested in the software, clicking on the button “3D” (choice of third dimension). After choosing the desired definition, just click on the NEXT button which displays a frame containing the data distribution in histograms.

Histograms

The third frame (Figure 7) provides a graphic display of histograms which allows a better perception over the distribution of measurements. The relative data distributions were concerning to the minor and major particles axes, aspect ratio

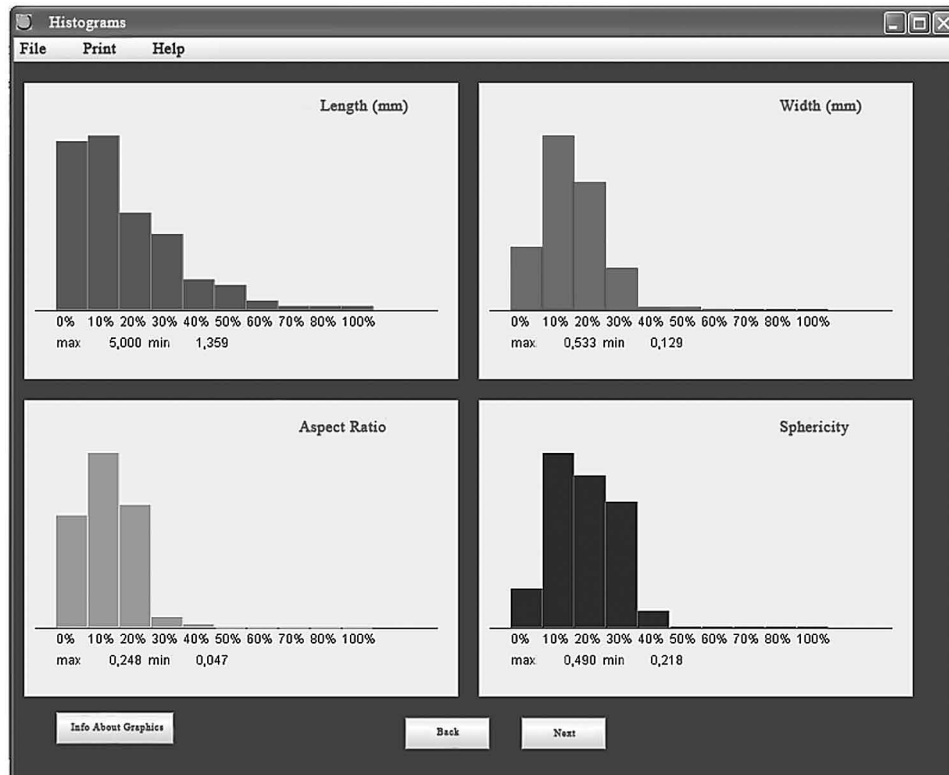
(L/D) and sphericity considering 2D or 3D. The histograms allowed for identifying which particle are out of expected results, and then, the users could delete these particles in order to get better results in their experimental analysis.

Besides that, a help menu was created to aid the users in case the misunderstanding of representations.

Final Results

The fourth and final frames (Figure 8) show the overall statistics for each measurement (major and minor axes, aspect ratio and sphericity) accord-

Figure 7. Program interface: Histograms



ing to their minimum, maximum, mean, median and standard deviation. Three significant digits were considered, due to the uncertainty of the conversion of pixel to millimeters. Furthermore, in this frame the user could SAVE THE DATA in a spreadsheet (Excel file, OpenOffice Calc file), where it could use them anytime and anywhere. Saving the data, the users should insert the file and folder names.

The navigation through these four frames is simple and it is not necessary to reprocess the image. The user can back and forward whenever one wants, because the information is stored in the memory during software execution. There is no database embedded on this software, but the results can be saved in a spreadsheet as Microsoft Excel or OpenOffice Calc.

The software APOGEO has a great advantage over possible competitors: a platform independent approach. In comparison with other similar software, such as ImageJ and some modules of

Matlab, its small space demanded for storage, language (Java) and portability are very attractive in favor for using.

Currently, our software has played a fundamental part in the conduct of research on biomass, because with automated measures implied by the program, the time for quantification and measurement of the particles decreased from hours to less than 1 minute. Without modifying its structure, it could contribute to the industry, automating systems and being able to choose particles or objects to any process. The disadvantage of the software is the lack of efficient algorithms for image processing, which would allow large files at high resolutions.

FUTURE RESEARCH DIRECTIONS

For future works, some new features will be implemented in the software, such as: increasing

Figure 8. Program interface: Final results

The screenshot shows a software window titled 'Final Results' with a menu bar containing 'File', 'Print', and 'Help'. The main area is divided into four quadrants, each displaying statistical data for a specific particle property. Each quadrant has a title and five rows of data, each with a label and a text input field containing a numerical value.

Length		Width	
Mean (mm)	2,226	Mean (mm)	0,264
Std. Deviation	0,698	Std. Deviation	0,066
Median (mm)	2,054	Median (mm)	0,258
Minimum (mm)	1,359	Minimum (mm)	0,129
Maximum (mm)	5,000	Maximum (mm)	0,533

Aspect Ratio		Sphericity	
Mean (mm)	0,126	Mean (mm)	0,349
Std. Deviation	0,040	Std. Deviation	0,055
Median (mm)	0,125	Median (mm)	0,352
Minimum (mm)	0,047	Minimum (mm)	0,218
Maximum (mm)	0,248	Maximum (mm)	0,490

Below the data tables, there is a button labeled 'Save in a Datasheet'. At the bottom of the window, there are two buttons: 'Back' and 'Next'.

the image size (higher than 3200 pixels) in order to read high-definition images; new filters can be applied in order to improve the image quality, and the software performance; choosing the particle that will be deleted by clicking over it, instead of navigating through each particle in order to find it; applying of the APOGEO software in a web service available for several platforms; and the software could be used as a pattern recognition software.

CONCLUSION

The software APOGEO was created to overcome the need for a automation tool of size and shape of solid particles (mainly biomass) measurements in a consistent and efficient manner. The computational methodology allowed the construction of a robust and simplified program. Furthermore, the alliance between the computational expertise and technical knowledge in particle analysis was es-

sential for determining their functions, equations, as well as its user interface. The validation of the results and the establishment of limits for its use were performed using a detailed analytical study. In a comparative analysis, APOGEO has proved to be very efficient compared to its competitors, but moreover, it also has the advantage of being portable to different operating systems, and have minimum requisites compatible to the most current devices. The field of applications and future improvements of the APOGEO is quite promising, leading to a wide range for its use, in academy or industry.

ACKNOWLEDGMENT

The authors would like to thank the contribution of students of Laboratory of Particle Technology and Multiphase Processes, School of Chemical Engineering at University of Campinas.

REFERENCES

- Abelson, H., Sussman, G. J., & Sussman, J. (1996). *Structure and interpretation of computer program*. New York, NY: Mc Graw-Hill Publisher.
- Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2002). *Algoritmos – Teoria e prática*. Rio de Janeiro, RJ: Campus Publishing Company.
- Cui, H., & Grace, J. R. (2007). Fluidization of biomass particles: A review of experimental multiphase flow aspects. *Chemical Engineering Science*, 62(1-2), 45–55. doi:10.1016/j.ces.2006.08.006
- Curry, J. K. (1951). *Analysis of sphericity and roundness of quartz grains* (Master's thesis). The Pennsylvania State University, University Park, PA.
- Dai, J., & Grace, J. R. (2011). Biomass granular screw feeding: An experimental investigation. *Biomass and Bioenergy*, 35(2), 942–955. doi:10.1016/j.biombioe.2010.11.026
- Farias, F. O. M. (2012). *Characterization of Brazilian biomasses for the purpose of energy generation* (Master's thesis). University of Campinas, Brazil.
- Gamma, E., Helm, R., Johnson, R., & Vlissides, J. (1994). *Design patterns: Elements of reusable object-oriented software*. Boston, MA: Addison Wesley Publishing Company.
- Guo, Q., Chen, X., & Liu, H. (2012). Experimental research on shape and size distribution of biomass particle. *Fuel*, 94, 551–555. doi:10.1016/j.fuel.2011.11.041
- Kunii, D., & Levenspiel, O. (1991). *Fluidization Engineering*. London, UK: Butterworth-Heinemann Publisher.
- Lourenço, J. (2012). *Study of fluid dynamic characteristics of mixtures of eucalyptus wood and the tucumã endocarp with inert aiming application in thermoconversion processes* (Master's thesis). University of Campinas, Brazil.
- Mandø, M., & Rosendahl, L. (2010). On the motion of non-spherical particles at high Reynolds number. *Powder Technology*, 202(1-3), 1–13. doi:10.1016/j.powtec.2010.05.001
- Maranesi, L. A. H., & Tannous, K. (2009). Real time interface for fluidized bed reactor simulator. In M. Khosrow-Pour (Ed.), *Encyclopedia of Information Science and Technology* (2nd ed.; pp. 3205–3212). Hershey, PA: Information Science Reference; doi:10.4018/978-1-60566-026-4.ch512
- Massarani, G., & Peçanha, R. P. (1986, October). Feature size and shape of particles. *Proceedings of the Fourth Meeting on Flow in Porous Media*.
- McAllister, W. (2009). *Data structures and algorithms using Java*. Sudbury, MA: Jones and Bartlett Publishers, LLC.
- Papini, C. J. (2003). *Comparative study of methods for determination of particle size* (Master's thesis). Nuclear Energy Research Institute, University of São Paulo, Brazil.
- Riley, N. A. (1941). Projection sphericity. *Journal of Sedimentary Research*, 11(2), 94–95.
- Rimoli, D., Assis, D. M., & Tannous, K. (2006, August). Simulator software applied to fluidized bed reactors – SEREA. *Proceedings of the Fifth International Conference for Conveying and Handling of Particulate Solids*.
- Rodriguez, J. M., Edeskär, T., & Knutsson, S. (2013). Particle shape quantities and measurement techniques—A Review. *Electronic Journal of Geotechnical Engineering*, 18. Retrieved March 12, 2014 from <http://www.ejge.com/2013/Abs2013.016.htm>
- Rodriguez, J. M., Johansson, J. M. A., & Edeskär, T. (2012, May). Particle shape determination by two-dimensional image analysis in geotechnical engineering. *Proceedings of the Nordic Conference on Soil Mechanics and Geotechnical NGM*, 207–218.

Sommerville, I. (2007). *Software engineering* (8th ed.). Boston, MA: Addison Wesley Publishing Company.

Starck, J. L., Murtagh, F., & Fadili, J. M. (2010). *Wavelets, curvelets, morphology diversity*. New York, NY: Cambridge University Press; doi:10.1017/CBO9780511730344

Tannous, K., Lam, P. S., Sokhansanj, S., & Grace, J. R. (2013). Physical properties for flow characterization of ground biomass from Douglas Fir wood. *Particulate Science and Technology*, 31(3), 291–300. doi:10.1080/02726351.2012.732676

Tannous, K., & Lourenço, J. B. (2015). Fluid dynamic and mixing characteristics of biomass particles in fluidized beds. In K. Tannous (Ed.), *Innovative Solutions in Fluid-Particle Systems and Renewable Energy Management* (pp. 54–91). Hershey, PA: Engineering Science Reference.. ch003 doi:10.4018/978-1-4666-8711-0.ch003

Tannous, K., & Rocha, A. B. (2013). Fluidized bed distributor N° 13591-0. Brazil: National Institute of Industrial Property, N° 2199.

Tannous, K., & Santos, G. E. (2012). Bernoulli equation and head loss in piping systems (BE-HELPS) N° 13310-1. Brazil: National Institute of Industrial Property, No 2187.

Tannous, K., & Silva, F. S. (2012). Particles and geometric shapes analyzer (APOGEO) N° 13085-6. Brazil: National Institute of Industrial Property, N° 2187.

Turbitt-Daoust, C., Alliet, D. F., Kaye, B. H., & Matchett, A. J. (2000). Orientation effects in the evaluation of populations of irregularly shaped particles by image analysis. *Particle & Particle Systems Characterization*, 17(4), 159–166. doi:10.1002/1521-4117(200012)17:4<159::AID-PPSC159>3.0.CO;2-P

Wadell, H. (1935). Volume, shape and roundness of quartz particles. *The Journal of Geology*, 43(3), 250–280. doi:10.1086/624298

Zavala, J. M. R. (2012). *Particle shape quantities and influence on geotechnical properties – A review* (Tech. Report). Luleå, Suécia: Luleå University of Technology. Retrieved March 12, 2014 from http://pure.ltu.se/portal/files/41279547/shape_report_november_2012_A4.pdf

ADDITIONAL READING

Abràmoff, M. D., Magalhães, P. J., & Ram, S. J. (2004). Image processing with ImageJ – *Biophotonics International*. Retrieved March 14, 2014 from <http://webeye.ophth.uiowa.edu/dept/biograph/abramoff/ImageJ.pdf>

Lezoray, O., & Grady, L. (Eds.). (2012). *Image processing and analysis with graphs: Theory and practice*. Boca Raton, FL: CRC Press.

Shenbagavadivu, S., & Devi, M. R. (2013). An investigation of noise removing techniques used in spatial domain image processing. *International Journal of Computer Science and Mobile Computing*, 2(7), 378–384.

Tannous, K. (2007). Project-based learning in chemical engineering education using distance education tools. In S. Mishra & R. C. Sharma (Eds.), *Cases on global e-learning practices successes and pitfalls* (pp. 202–217). Hershey, PA: Idea Group. doi:10.4018/978-1-59904-340-1.ch016

KEY TERMS AND DEFINITIONS

Data Structure: A particular way of storing information, allowing to a high level approach on the software implementation.

Fluidized Bed: Technology related to the fine and large solids transforming into a fluid-like state through contact with a gas or liquid.

Graphical Interface: The interface between a human and a computer through windows, symbols and images.

Image Processing: A computer technique which process a two dimensional picture.

Object-Oriented Programming: A programming paradigm that uses “objects” and their interactions to design applications and computer programs.

Particle Sphericity: The description of the overall shape of the particle irrespective of angularity of edges and corners.

Recurrence or Recursion: A function or method that replicates itself into entries until reach the base case.

C

The Understanding of Spatial–Temporal Behaviors

Yu-Jin Zhang

Tsinghua University, China

INTRODUCTION

The human visual process can be seen as a complex process ranging from feeling (feel the projection of 3-D world on the resulting 2-D image) to perception (capturing the content and meaning of 3-D world with 2-D images) (Kong, 2002). The ultimate goal of vision, from the narrow sense, is to make a meaningful interpretation of the world and the description of the scene; while broadly speaking, it includes also designing a behavior plan based on these explanations and descriptions, and in accordance with the surroundings and wishes of viewers. Computer vision is to realize the task to make computer refers to the human visual function, hoping to make a meaningful judgment on realistic goals and scenarios based on perceived images (Shapiro, 2001). This is just also the goal of understanding.

An important task in computer vision and image understanding is to analyze the scene through image operation on the image of scene in order to guide the action. To do this, one needs to locate the objects in the scene, and to determine how they change its position, attitude, speed and relationships in the space over time. In short, it is to grasp the action in time and space, to determine the purpose of the operation, and thus to understand the semantics of the information they passed. This is refereed as the understanding of spatial-temporal behaviors.

Behavior-based automatic target image/video understanding is a very challenging research issue. It includes the access of objective information (image/video acquisition sequence), the process

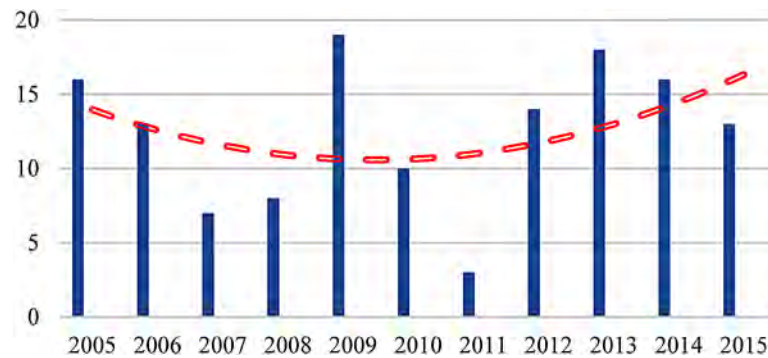
of relevant visual information, the analysis of contents (expression and description), as well as the interpretation on the basis of the image/video information, in order to achieve learning and recognition of behavior.

BACKGROUND

Research forces and results around such a topic are just appeared in recent years, some statistics can be seen from the survey on image engineering (Zhang, 2015c). The annual survey series of the yearly bibliographies on image engineering has started in 1995 and has been carried out for 21 years (Zhang, 2016). When the series enters its second decade (for the literature statistics of 2005), with the appearance of some new hot spots in the image engineering research and application, a new subcategories (C5): spatial-temporal technology (including 3-D motion analysis, gesture and posture detection, object tracking, behavior judgment and understanding) has been added into the image understanding category (C) (Zhang, 2006). The emphasis here is the comprehensive utilization of a variety of information possessed by the image/video in order to make the according interpretation for the dynamics of scene and objects inside.

In the past eleven years, the number of publications belong to the subcategory C5 in the annual survey series has attend a total of 153. There are five subcategories in category C, and the total number of publications belong to category C in these eleven years is 1352, so the subcategory C5

Figure 1. Some statistics of the numbers of publications for spatial-temporal technology



is still a small subcategory. Their distributions in each years are shown in the bars in Figure 1, in which a 3-order polynomial curve fitting to the number of publications of each year is drawn to show the change trends. Overall, this is still a relatively new field of research, so its development is not too stable, yet.

MAIN FOCUS OF THE ARTICLE

The definition, development, and stratification of spatial-temporal technology are first provided. Then, from low-level to high-level, the detection of points of interest, the forming of dynamic trajectory and activity path, the example techniques for action classification and recognition, as well as modeling for action and activity, are introduced consecutively. Several further research directions are discussed before some final concluding remarks are delivered.

Spatial-Temporal Technology

Spatial-temporal technology is oriented to the understanding of spatial and temporal behavior. Currently, the main target of the research on spatial-temporal technology is a moving person or object, and the object (particularly a human) changes. According to the expression level of abstraction and description, it can be divided into multiple levels from bottom to top.

1. **Action Primitives:** Atomic units to build a significant action, generally corresponds to a short period of motion information in the scene.
2. **Action:** A series of acts composed of a number of action primitives, a collection of ordered and meaningful models, generally represents a simple execution often carried out by one person, and often lasts only seconds in duration. The results of human actions often lead to changes in body posture and/or object position.
3. **Activity:** A series of actions performed by the host/initiator (mainly emphasize logic combinations). It is a relatively large-scale sports events, generally depends on the environment and people in interaction (Aggarwal, 2011). It is often an occurrence or a consequence of complex operations performed by more than one person, and often last for a relative long period of time.
4. **Event:** It refers to certain activities occurred in specific periods and specific spatial location. Typically one of the actions performed by a plurality of host / initiator (group activity). Detection of specific events often associated with abnormal activities, an overview on this can be seen from (Popoola, 2012).
5. **Behavior:** Subject/initiator mainly refers to a human or animal, emphasizing the body/originator dominated by ideological change action in a particular environment/context, the ongoing activities and describe events.

The following five table tennis pictures, for instance, provide some typical examples (Figure 2). Player's leg lifting, racket waving, etc. can be seen as typical action primitives in table tennis. Players complete a ball serving (including ball tossing, arm swing, wrist shaking, ball hitting, etc.) or a ball returning (including the pace moving, arm outstretching, wrist flipping, ball stroking, etc.) are typical types of action, but one player went to apron and pick a ball are often seen as an activity. The two players hit the ball back and forth in order to win points is a typical scene of activity. Competition between teams is generally viewed as an event, so as a wonderful shot, and the awarding after the match is also a typical event. Fist after winning of players can be seen as a self-motivated action, but more often it is seen as a behavior of the players. When players made good exchange, the applause of the audience, as well as shouting, cheering, etc., are also attributed to the behavior of the audience.

It should be noted that the concepts of the last three levels are often not strictly distinguished in many practical studies. For example, some activities are called events, this generally refers to some unusual activities (such as the disputes between two persons, the falling of an elderly person, etc.). Sometimes, the activities are referred to as behavior, it emphasis more on the meaning of activities and the particularity of activities (such as burglary behaviors or over the wall burglary theft). On the other side, they are often closely related. For instance, many human activities are directly related to their behavior (Khair, 2012). In the following discussion, unless special emphasis

would be made, the (generalized) activity will be used as the main representatives of three levels.

The research in spatial-temporal technology has been conducted from points (point of interest) to curves (trajectory of object = multiple points), to surfaces (allocation of activity = compound curves), and to volume (variation of behaviors = stack of surfaces).

Spatial-Temporal Points of Interest

Scene changes come from the motion of objects in the scene, especially accelerated motion. In video of scenes, the accelerated motion of partial structures corresponds accelerated motion of objects, they are at some particular positions in the image, with unconventional change values. The information on these positions (called points of interest) are helpful for understanding the scene.

First consider still images. The modeling of image $f(x, y)$ can be made with the help of linear scale space L , $L: \mathbf{R}^2 \rightarrow \mathbf{R}$:

$$L(x, y; \sigma_l^2) = g(x, y; \sigma_l^2) \otimes f(x, y) \quad (1)$$

where g is a Gaussian kernel with variance σ_l^2 :

$$g(x, y; \sigma_l^2) = \frac{1}{2\pi\sigma_l^2} \exp\left[-\frac{(x^2 + y^2)}{2\sigma_l^2}\right] \quad (2)$$

The idea behind the typical Harris detector of interesting points is to determine the spatial locations in $f(x, y)$ with significant change in both horizontal and vertical directions. For a

Figure 2. Some typical frames of images from a table tennis game



given observation scale σ_l^2 , these points can be computed with the help of the matrix of second order moment that is obtained by summation in a Gaussian window with variance σ_l^2 :

$$\begin{aligned} \mu(x, y; \sigma_i^2, \sigma_l^2) &= g(x, y; \sigma_i^2, \sigma_l^2) \otimes \\ &\left\{ \left[\nabla L(x, y; \sigma_l^2) \right] \left[\nabla L(x, y; \sigma_l^2) \right]^T \right\} \\ &= g(x, y; \sigma_i^2, \sigma_l^2) \otimes \begin{bmatrix} (L_x)^2 & L_x L_y \\ L_x L_y & (L_y)^2 \end{bmatrix} \end{aligned} \quad (3)$$

where L_x and L_y are the Gaussian differentials at local scale σ_l^2 . The eigenvalues λ_1 and λ_2 form a change descriptor of $f(x, y)$ along two directions, respectively. If both λ_1 and λ_2 have large values, then a point of interest exists. To detect such points, the positive maximum of corner point function can be computed:

$$H = \det(\mu) - k \text{trace}^2(\mu) = \lambda_1 \lambda_2 - k(\lambda_1 + \lambda_2)^2 \quad (4)$$

The above computation for points of interest in spatial space can be extended to spatial-temporal space for detection of spatial points of interest in a particular location on time. This tendency is started around 10 years ago (Laptev, 2005). Detection of such points is essential for extracting low level motion features, and no background modeling is required.

For modeling spatial-temporal image sequence, start from $L, L: \mathbf{R}^2 \rightarrow \mathbf{R}$, and convolve it with non-isotropic (space variance σ_l^2 and time variance τ_l^2 are not correlated) Gaussian kernel to form a linear scale space $L, L: \mathbf{R}^2 \times \mathbf{R} \rightarrow \mathbf{R}$.

$$L(x, y; \sigma_l^2, \tau_l^2) = g(x, y; \sigma_l^2, \tau_l^2) \otimes f(x, y) \quad (5)$$

where the Gaussian kernel for the separation of space and time is

$$g(x, y, t; \sigma_l^2, \tau_l^2) = \frac{1}{\sqrt{(2\pi)^3 \sigma_l^4 \tau_l^2}} \exp \left[-\frac{x^2 + y^2}{\sigma_l^2} - \frac{t^2}{2\tau_l^2} \right] \quad (6)$$

Similar to the spatial domain, a spatial-temporal matrix of second order moment is a 3×3 matrix containing spatial differentials of space and time that convolved with Gaussian function $g(x, y; \sigma_l^2, \tau_l^2)$:

$$\mu = g(x, y; \sigma_i^2, \tau_i^2) \otimes \begin{bmatrix} L_x^2 & L_x L_y & L_x L_t \\ L_x L_y & L_y^2 & L_y L_t \\ L_x L_t & L_y L_t & L_t^2 \end{bmatrix} \quad (7)$$

where the first order differentials are

$$\begin{aligned} L_x(x, y; \sigma_l^2, \sigma_l^2) &= \partial_x (g \otimes f) \\ L_y(x, y; \sigma_l^2, \sigma_l^2) &= \partial_y (g \otimes f) \\ L_t(x, y; \sigma_l^2, \sigma_l^2) &= \partial_t (g \otimes f) \end{aligned} \quad (8)$$

To detect a point of interest, the search for μ with significant eigenvalues λ_1 , λ_2 , and λ_3 is conducted. This can be accomplished by extending eq.(4) to spatial-temporal domain:

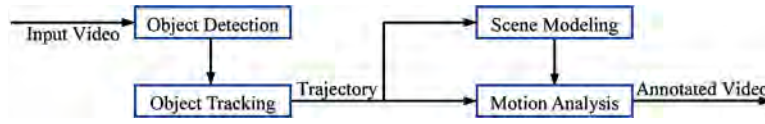
$$H = \det(\mu) - k \text{trace}^3(\mu) = \lambda_1 \lambda_2 \lambda_3 - k(\lambda_1 + \lambda_2 + \lambda_3)^3 \quad (9)$$

Dynamic Trajectory Learning and Analysis

Dynamic trajectory learning and analysis attempt to provide certainty for monitoring the state of the scene by understanding and characterization of the change of each target position and the moving results (Morris, 2008). A flowchart for dynamic trajectory learning and analysis in video is shown in Figure 3. First, the target is detected and tracked; then, the scene model is automatically constructed with the trajectory obtained; finally, the model is used to monitor the situation and provide labels for the activities.

In the scene modeling, the points of interest (POI) are first determined within an image area and considered as the location where some events happening, then in the learning step the activities

Figure 3. Flowchart for dynamic trajectory learning and analysis



are defined along an activity path (AP). The path is to characterize how the target moving between the points of interest. Such constructed models can be called POI/AP model.

The main tasks in the POI/AP Learning include:

1. Activity learning;
2. Adaption;
3. Feature selection.

Automatic Scene Modeling

Modeling scenarios by means of dynamic trajectory automatically include the following three tasks (Makris, 2005):

1. **Object Tracking:** It needs for each observable object to achieve identity maintenance in each frame. For example, pedestrians can be detected by camera installed on a moving car (Jia, 2007).
2. **Points of Interest Detecting:** The first task in image scene modeling is to figure out a region of interest. In topographic map for object tracking, these regions correspond to the nodes in the graph. Two types of nodes mostly considered are in/out regions and stop regions.

In/out regions are the locations of the objects to enter or leave the field or the locations of tracked objects appear or disappear. Stop regions come from scene landmark points, that is, the positions that an object tends to stay over a period. There are two different ways for locating stop regions. One is to find the region inside which the tracked points have moving speed lower than a pre-determined threshold; another is to find the region inside

which the tracked points keep in a limited distance range within a certain period.

3. **Active Path Finding:** To understand the behavior, the active path (AP) needs to be determined. This can be obtained by filtering out false alarms or tracking interrupted noise from training set with POI, keeping only the paths started after entering the region and ended before leaving the region. An activity of interest should be defined between the two end points (points of interest).

Active Path Learning

As the active path provides the information for object motion, an original trajectory can be represented by a sequence of dynamic measurements. For example, a common representation for trajectory is often a motion sequence:

$$Diam(B) = \max_{i,j} [D(b_i, b_j)] \quad b_i \in B, \quad b_j \in B \quad (10)$$

where the motion vector is

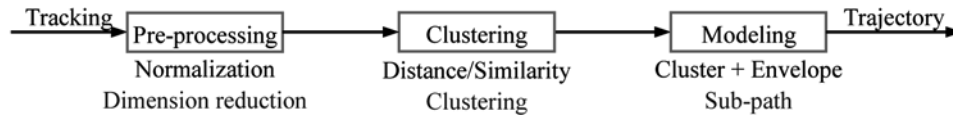
$$Diam(B) = \max_{i,j} [D(b_i, b_j)] \quad b_i \in B, \quad b_j \in B \quad (11)$$

It represents the dynamic parameters obtained from the object tracking at time t , i.e., position $[x, y]^T$, speed $[v_x, v_y]^T$, and acceleration $[a_x, a_y]^T$.

It is possible to learn AP, with unsupervised way, by using only the trajectory. The basic process is shown in Figure 4 that includes three steps.

1. **Pre-Processing:** It consists of two tasks, normalization to ensure that all the paths have

Figure 4. Steps for path learning



the same length; dimension reduction for mapping the paths to a new low-dimensional space to perform more robust clustering.

2. **Clustering:** In order to produce meaningful clustering, three considerations are the definition of a distance measure (corresponding to similarity), the updating policy for clusters, and the clustering validation.
3. **Modeling:** Two methods can be used, as shown in Figure 5. One considers the complete path by using the cluster centers and the envelopes (to indicate the range of path), as in Figure 5(a). Other one divides the whole path into a number of sub-paths (using tree structure), as in Figure 5(b).

Automatic Activity Analysis

Once the scene model is established, then the activities and behaviors of the objects can be analyzed. In addition to simply identify a particular behavior, all atypical events need to check. By observing a scene for long time period, the system can analyze a series of activities, and can further learn to find what the event of interest is.

Some typical activity analysis examples are:

1. **Virtual Fencing:** Any monitoring system has a monitoring range. By setting up early warning on the border, an equivalent virtual

fence is established, and the invasion will trigger system to start an analysis procedure.

2. **Speed Profiling:** Virtual fencing uses only position information, while tracking can also provide dynamic information for speed-based early warning.
3. **Path Classification:** Speed profiling uses only the current tracking data, while activity path (AP) can also provide historical information to predict and interpret the coming data.
4. **Abnormality Detection:** Since the activity path can indicate the characteristics of typical/normal activities, if the new path shows certain differences, an abnormality will be detected.
5. **Online Activity Analysis:** Enabling online analysis, identification, and evaluation of activities would be more powerful and useful than just describing the activity with the path. An online system needs to inference the current behavior based on incomplete data.
6. **Object Interactive Characterization:** Even higher level analysis is expected to explain the interaction among different objects, their motions may need a group of paths, and their activities may need several image sequences.

Figure 5. Two methods for path modeling



Action Classification and Recognition

Vision-based human action recognition is a process to use action labels for marking an image sequence (video). On the basis of obtaining the representation of the observed image sequence or video, this process can be turned into a classification problem.

Action Classification

Many techniques for action classification have been proposed (Poppe, 2010).

1. **Direct Classification:** It does not pay special attention to the time domain, even the video is used. The related methods put all observed information (from all frames) into a single expression, or identify and classify actions separately for each frame. As high-dimensional representation for image is required, a large number of computation is inevitable.
2. **Time Status Model:** Both generative models and descriptive models can be used. Generative models learn the joint distribution between observation and action, and model each action class. Typical generative models include hidden Markov model (HMM), maximum entropy Markov model (MEMM), etc. Descriptive models learn the conditional probability to observe the action class. They do not concern for the class model, but for the difference between classes. Conditional random fields (CRF) is an often used descriptive model.
3. **Action Detection:** It does not explicitly model the object representation in the image, nor the action in the image. It connects the observed sequence with the labeled video sequence to directly detect (defined) actions. The main methods for action representation and description can be classified into two groups: appearance-based and body model-based.

Action Recognition

Many methods for action recognition have been developed. They are depended on the purpose of the research and the application domain (Moeslund, 2006). For monitoring systems, the human activity and human interaction are considered. For scene interpretation, representation can be independent from the objects inducing the motion.

1. **Holistic Recognition:** It puts the overall emphasis on the whole body identification or on identifying each individual part of single person. Most techniques are based on human silhouette or outline, and they have less distinction among various parts of the body.
2. **Posture Modeling:** Recognition of human action is closely related to the body posture estimation. Body posture can be divided into action attitude and postural attitude (gesture). The former corresponds to the movement behavior at a certain moment, while the latter corresponds to the orientation of human body in 3-D space. The main methods for posture representation and description can be classified into three groups: appearance-based, body model-based, and 3-D reconstruction-based.
3. **Activity Reconstruction:** Action results in a change of posture. If each stationary body posture is defined as a state, then the building of a sequence of actions (activity) can be obtained by conducting a traversal through the corresponding attitudes of the state, with the help of state-space method (also known as probability Cyber Law). Based on such a sequence of actions (activity), body action and posture can be recovered.
4. **Inter-Activity:** Inter-activity is relative complex. Two main types can be distinguished: interaction between human and environment; interaction among different persons. In the former, the human is the initiator of the activity, such as taking a book up from

a table, driving a car on the road, etc. This can be referred as single (person) activity. In the latter, the human interacts each other. It often refers to the exchange activities or contact behavior of two (or multiple) persons. It can be seen as the combination of several single (atomic) activities.

5. **Group Activity:** Many quantitative changes may cause a qualitative change. Large number of objects involved in the activity could pose new problems and require new solutions. Semantic analysis of the polymerization, dissipation, differentiation, and combination of object groups becomes indispensable for capturing the tendency and situation of whole scene.
6. **Scene Interpretation:** Different from recognition of objects in scene, scene interpretation considers to comprehend the entire image rather than to verify a particular person or object. In practice, many methods do recognize activity by only observing the motion of objects in the images, without determining the identity of objects.

Modeling Activity and Behavior

A general action/activity recognition system should perform, from an image sequence to high-level interpretation, several tasks (Turaga, 2008):

1. Capturing input video or image sequence.
2. Extracting low-level image features.
3. Describing middle-level actions.
4. Interpreting image with high-level semantics.

General practical activity recognition system has hierarchical structure. In the low-level, there

are modules for foreground-background separation, for object detection and tracking. In the middle level, the main module is for action recognition. In the high-level, the most important module is inference engine, which understands activity with the aid of learning.

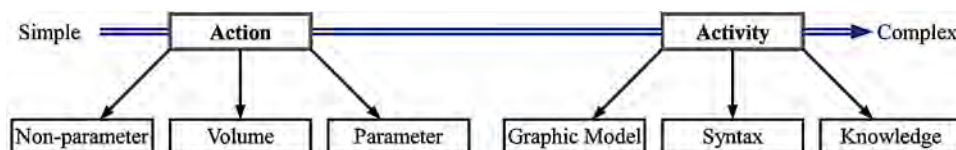
From an abstract point of view, the level of activity is higher than that of action. From the technology point of view, modeling and recognizing the action and activity often using different techniques. A categorization scheme is shown in Figure 6 (Turaga, 2008).

Action Recognition

The methods for action recognition can be divided into three groups: non-parameter modeling, volume modeling, (timing) parameter modeling.

1. **Non-Parameter Modeling:** It extracts, from each frame of video, a group of features, and matches these features with stored templates. Typical methods include: using 2-D template, using 3-D object model, manifold learning, etc.
2. **Volume Modeling:** It does not extracts features from each frame of video, while considers video as a collection of 3-D voxels, and extend the standard 2-D image features (such as scale space extreme values) to 3-D. Typical methods include: spatial-temporal filtering, using parts of 3-D space (for example, spatial-temporal points of interest), sub-volume matching, tensor-based methods, etc.
3. **Parameter Modeling:** It dynamically models the motion time, and estimates the specific parameters for a group of actions

Figure 6. Classification of approaches for action and activity recognition



from training sets. It is more suitable for complex actions cross over different time domain. Typical methods include: hidden Markov model (HMM), linear dynamic system, non-linear dynamic system, etc.

Activity Modeling and Recognition

Compared to action, activity not only has long duration but also often involve many initiators. For modeling complex scenario, it is necessary to perform high-level representation and reasoning for the intrinsic structure and semantics of behavior (Zheng, 2012). The methods for activity modeling and recognition can be divided into three groups: by means of graphic model, with the help of syntax, using knowledge.

1. **Graphic Model:** Belief network (such as Bayesian network) and Petri net are two popular networks used, which can be represented by graph.
2. **Syntax Methods:** Two key points for syntax methods are grammar and rule. The grammar uses a group of generative rules to describe the structure under considerations, so as to model body motion and multiple-person interaction. Stochastic context-free grammar has been used for modeling semantic of activity.
3. **Knowledge Methods:** Knowledge is closely related to logic. Logic rules are suitable for representing input knowledge and high-level reasoning results, so they can be used to describe general domain knowledge for activity explanation. Ontology can standardize the definition of activity, increase inter-operability among different systems, and compare the performance of various behavior assessments.

FURTHER RESEARCH DIRECTIONS

Spatial-temporal behavior understanding is a relatively new field of research, many future research works are required, and several potential directions are:

1. As can be seen from the above introduction and discussion, research in this field has gone from points to curves, from curves to surfaces, and from surfaces to volumes. What would be the next step?
2. Action and activity, both are the physic bases of behavior. They are often the most significant and also the most attractive parts of image and video. In this regard, the attention-based recognition techniques (Zheng 2015) should be beneficial.
3. With big data collected from all areas, from one side, a large number of situations requiring behavior understanding arise; from other side, using data mining techniques, it is possible to learn some effective information for accomplish such tasks. To this end, content based visual information retrieval would be called for behavior classification (Zhang 2015b).
4. Deep learning, more general, machine learning, technology has been proved being a powerful tool in many information processing task recently (Zhang 2015a; Zheng 2014; Zheng 2016). How to efficiently use this new technology for understanding of spatial-temporal behavior is expected.

CONCLUDING REMARKS

This chapter introduces a cutting-edge research field of computer vision and image understanding – the spatial-temporal behavior understanding. The main concepts, the focus of research, the typical technology, the fast development, etc. of this new field in recent years are overviewed. From the research advances and trends in the five

levels of spatial-temporal technology, the most of current research efforts seems going to the more challenging high levels. From the point of view in considering the scope of the study involved in the content, except for image technology itself, medicine, psychology, behavioral science, cognitive science, sociology and other disciplines will also play an active role in the future.

ACKNOWLEDGMENT

This work has been supported by the National Natural Science Foundation under Grants NNSF-61171118 and the Ministry of Education under Grants SRFDP-20110002110057.

REFERENCES

- Aggarwal, J. K., & Ryoo, M. S. (2011). Human activity analysis: A review. *ACM Computing Surveys*, 43(3), 1–43.
- Jia, H. X., & Zhang, Y. J. (2007). A survey of computer vision based pedestrian detection for driver assistance system. *Acta Automatica Sinica*, 33(1), 84–90.
- Khair, N. M., Yaacob, S., & Hariharan, M. (2012). A study of human emotional: Review. *Proc. 2012 International Conference on Biomedical Engineering*, 393–399.
- Kong, B. (2002). Comparison between human vision and computer vision. *Nature Magazine*, 24(1), 51–55.
- Laptev, I. (2005). On space-time interest points. *International Journal of Computer Vision*, 64(2/3), 107–123.
- Makris, D., & Ellis, T. (2005). Learning semantic scene models from observing activity in visual surveillance. *IEEE-SMC-B*, 35(3), 397–408. PMID:15971910
- Moeslund, T. B., & Hilton, A., & Krüger, V. (2006). A survey of advances in vision-based human motion capture and analysis. *Computer Vision and Image Understanding*, 104, 90–126.
- Morris, B. T., & Trivedi, M. (2008). A survey of vision-based trajectory learning and analysis for surveillance. *IEEE-CSVT*, 18(8), 1114–1127.
- Popoola, O. P., & Wang, K. (2012). Video-based abnormal human behavior recognition: A review. *IEEE-SMC-C*, 99(99), 1–14.
- Poppe, R. (2010). A survey on vision-based human action recognition. *Image and Vision Computing*, 28, 976–990.
- Shapiro, L., & Stockman, G. (2001). *Computer Vision*. Prentice Hall.
- Turaga, P., Chellappa, R., & Subrahmanian, V. S. (2008). Machine recognition of human activities: A survey. *IEEE-CSVT*, 18(11), 1473–1488.
- Zhang, Y.-J. (2006). Image engineering in China: 2005. *Journal of Image and Graphics*, 11(5), 601–623.
- Zhang, Y.-J. (2015a). Machine learning for image classification. *Encyclopedia of Information Science and Technology* (3rd ed.). Information Science Reference, Chapter 21 (215–226).
- Zhang, Y.-J. (2015b). Up-to-date summary of semantic-based visual information retrieval. *Encyclopedia of Information Science and Technology* (3rd ed.). Information Science Reference, Chapter 123 (1294–1303).
- Zhang, Y.-J. (2015c). Statistics on image engineering literatures. *Encyclopedia of Information Science and Technology* (3rd ed.). Information Science Reference, Chapter 595 (6030–6040).
- Zhang, Y.-J. (2016). Image engineering in China: 2015. *Journal of Image and Graphics*, 21(5), 533–543.

Zheng, Y., Chen, Q.-Q., & Zhang, Y.-J. (2014). Deep learning and its new progress in object and behavior recognition. *Journal of Image and Graphics*, 19(2), 175–184.

Zheng, Y., Zemel, R. S., Zhang, Y.-J., & Larochelle, H. (2015). A neural autoregressive approach to attention-based recognition. *International Journal of Computer Vision*, 113(1), 67–79.

Zheng, Y., Zhang, Y.-J., & Larochelle, H. (2016). A deep and autoregressive approach for topic modeling of multimodal data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(6), 1056–1069. PMID:26372202

Zheng, Y., Zhang, Y.-J., Li, X., & Liu, B. D. (2012). Action recognition in still images using a combination of human pose and context information. *Proceedings of the 19th International Conference on Image Processing*, 785–788.

KEY TERMS AND DEFINITIONS

Clustering: Clustering is also called unsupervised learning and is a powerful technique for pattern classification. It is a process to group, based on some defined criteria, two or more terms together to form a large collection. In the context of image segmentation, it is often considered as the multi-dimensional extension of the thresholding technique.

Computer Vision (CV): A field that includes methods for acquiring, processing, analyzing, and understanding images from the real world in order to produce numerical or symbolic information, e.g., in the forms of decisions. A theme in the development of this field has been to duplicate the abilities of human vision by electronically perceiving and understanding an image. Computer vision has also been described as the enterprise of automating and integrating a wide range of processes and representations for vision perception.

Deep Learning (DL): A new area of machine learning research. It uses a cascade of many layers of nonlinear processing units for feature extrac-

tion and transformation. Each successive layer uses the output from the previous layer as input. It also learns multiple levels of representations that correspond to different levels of abstraction. Deep learning algorithms are based on distributed representations. The composition of a layer of nonlinear processing units used in a deep learning algorithm depends on the problem to be solved.

Image Classification: Aims at associating different images with some semantic labels to represent the image contents abstractly. To achieve this goal, various machine learning and pattern recognition techniques could be used.

Image Engineering (IE): An integrated discipline/subject comprising the study of all the different branches of image and video techniques. As a general term for all image techniques, it could be considered as a broad subject encompassing mathematics, physics, biology, physiology, psychology, electrical engineering, computer science, automation, etc. Its advances are also closely related to the development of telecommunications, biomedical engineering, remote sensing, document processing, industrial applications, etc.

Image Techniques: A collection of various branches of techniques for processing (such as acquiring, capturing, sensing, storing, enhancing, filtering, de-blurring, in-painting, transforming, coding, transmitting, manipulating, etc.) analyzing (such as segmenting, representing, describing, featuring, measuring, classifying, recognizing), and understanding (such as modeling, registering, matching, reconstructing, training, learning, reasoning, interpreting, etc.) images.

Image Understanding (IU): One of three layers of image engineering, which transforms data extracted from images into certain commonly understood descriptions, and makes subsequent decisions and actions according to the interpretation of the images.

Machine Learning (ML): A powerful tool for pattern classification. It uses the theory of statistics in building mathematical models, and programs computers to optimize a performance criterion using example data or past experience.

Category C

Criminal Science and Forensics

Forensic Investigations in Cloud Computing

Diane Barrett

Bloomsburg University of Pennsylvania, USA

INTRODUCTION

Cloud computing environments add an inherent layer of complication to a digital forensic investigation. The content of this article explores current forensic acquisition processes, why current processes need to be modified for cloud investigations, and how new methods can help in an investigation. A section will be included that provides recommendations for more accurate evidence acquisition in investigations. A final section will include recommendations for additional areas of research in the area of investigating cloud computing environments and acquiring cloud computing based evidence.

BACKGROUND

Cloud Computing Environments

Cloud computing is encompassed in the capabilities of almost all existing technologies. The concept behind cloud computing is a production environment in which resources and software services do not function locally. Instead, the Internet or the internal network of an organization seamlessly connects numerous host machines running on a virtualized platform (Budriene & Zalieckaite, 2012).

Pallis (2010) provides a general layered architecture of cloud infrastructures as a basic model by classifying the architecture into three abstract layers using two models: deployment and service, along with a set of characteristics. The layers

from the bottom up are infrastructure, platform, and application. The infrastructure layer provides fundamental computing resources such as processing, storage, and networks. The platform layer delivers higher-level services and abstractions for integration of the ability to perform application functions in the environment. The application layer allows the capability for applications as a service (AaaS).

These three layers are further broken down into service models, deployment models, and attributes. The three well-recognized cloud service models are infrastructure as a service (IaaS), platform as a service (PaaS), and software as a service (SaaS). The four cloud deployment models are community, hybrid, public, and private. The attributes consist of measured and on-demand self-service, resource pooling, rapid elasticity, and broad network access. This is the exact layered architecture outlined by National Institute of Standards and Technology (NIST) in the final issuance of the cloud computing definition dated September 2011.

Environmental Variables

Complex and dynamic business environments such as cloud computing environments drive organizations of all sizes to respond rapidly to market changes and pursue creative resource saving solutions. In addition to being a technology solution, cloud computing is a new business model. Cloud computing environments offer unrestricted scalability and lower data-center setup costs by using multitenancy.

The multitenancy and virtualization characteristics of a cloud computing environment present difficult implementation demands in the areas of security and access control (Almutairi, Sarfraz, Basalamah, Aref, & Ghafoor, 2012). The unique security and access control challenges presented by the use of multitenancy and virtualization in cloud computing environments exist because many individual environments share the same set of hardware. The sharing of storage blocks can result in the accidental and unauthorized flow of information (Werner, 2011). The diversity of services offered in cloud computing environments requires variable levels of granularity when implementing access control mechanisms. The risk of resource exploitation by unauthorized users is significantly increased when there are insufficient or untrustworthy authorization mechanisms implemented in a cloud computing environment (Werner, 2011).

Cloud computing environments offer many organizational benefits by providing scalable but complex computing infrastructures. Every cloud deployment and service model instance is different. For example, one SaaS implementation can be completely different from the next. There are many newly emerging challenges associated with the use of cloud computing environments and existing issues are not yet addressed. Automated service provisioning, virtual machine migration, server consolidation, and the management of power and security are just beginning to garner research community attention.

Digital Evidence Seizure

Digital forensics focuses on the retrieval and analysis of data found on digital devices relative to some type of unauthorized or criminal activity (Garfinkel, 2010). Traditional digital forensics processes consist of crime scene evidence collection, evidence preservation, evidence analysis, and presentation of the analysis results (Greengard, 2012). Current traditional digital acquisition processes include maintaining chain of custody control of forensic evidence data. This chain of

custody control occurs in the evidence collection phase through the imaging of a system (Decker, Kruse, Long, & Kelley, 2011).

Cloud computing technology disrupts the initial step of evidence collection in conducting a digital forensic investigation. Traditionally, the seizure of forensic evidence occurs through a search warrant or other legal request. In a cloud computing environment, the emphasis is on the cloud provider contract as opposed to a search warrant or legal request because the data can be geographically disbursed and is under the control of the service provider (Dykstra & Sherman, 2011). Cloud contracts play a very important part in forensic practitioner's ability to conduct a sound forensic investigation (Svetcov, 2011). In some cases, the forensic investigator never gets to acquire the evidence. The designated employee at the service provider performs the evidence acquisition and supplies the results to the client, who then sends the information to the investigator (James, Shosha, & Gladyshev, 2013). This is contrary to current forensic practice where legal documents provide the case investigator the ability to acquire the evidence.

In a cloud computing environment, data are spread out amongst numerous servers and other components. When a crime happens, the location of recoverable data includes the client machine, the equipment of the Internet Service Provider (ISP), and the data backups of the cloud service provider (Shin, 2013). This makes acquiring evidence from a crime committed in a cloud computing environment much more difficult since the environment is not localized like it is in a traditional digital forensic evidence acquisition (Dykstra & Sherman, 2011). Traditional imaging cannot be performed because it is not possible to take down and create a forensic image of such a large environment (James et al., 2013).

The digital forensics discipline faces new challenges where cloud computing is concerned (Berman, Kesterson-Townes, Marshall, & Srivathsa, 2012). Traditional forensic evidence acquisition processes do not fit well into cloud

computing because of the way cloud computing works (Svetcov, 2011). Traditional digital forensic acquisition processes focus on individual computers and isolated environments, while cloud computing forensic acquisition processes include the intricacies of complex infrastructures including virtual servers, applications, and diverse operating platforms that may be located in foreign countries (James et al., 2013).

Common forensic procedure is to shut down the system in order to take a forensic image of the system. This procedure cannot apply to hosted cloud services due to disruption of service to a wide scope of users (Cho, Chin, & Chung, 2012). Cloud computing systems consist of multiple user environments, using a variety of services. Shutting down a cloud computing system disrupts services to all the user environments hosted on the system.

When the user or service provider shuts down the virtual machine because of inactivity, service providers reallocate the storage space originally allocated to the virtual machine to other virtual machines (Huth & Cebula, 2011). Upon deletion of the virtual machine and the overwriting of storage space, the evidence disappears (Svetcov, 2011). Although it is possible to recover forensic data from a running virtual machine, if the files associated with the virtual machine are not collected, valuable evidence is missed (Fiterman & Durick, 2010). Associated virtual machine files contain information such as configurations settings and memory content are often not collected as part of the acquisition process.

Two well-known and commonly used commercial cloud services are Amazon Elastic Compute Cloud (EC2) and Rackspace's cloud services. Hosting cloud services on geographically diverse servers using distributed file systems provides scalability and reliability (Cho et al., 2012). Current forensic methods do not exist for analyzing this type of cloud computing environment. The only way to recover evidence is through vendor provided tools that are designed for gathering metrics. Forensic analysis of distributed file systems is very difficult (Cho et al., 2012). Multi-national

web-based organizations such as Facebook and Yahoo use the Hadoop file system. Hadoop is a common distributed file system used in large data environments such as cloud computing. Due to the way the Hadoop file system works, it processes petabytes of data and erasure of forensic evidence could easily go without notice (Cho et al., 2012). The development of Hadoop occurred with fast processing of data in mind and there was very little consideration given to security or forensic analysis capabilities.

Current Cloud Acquisition Approaches

Without a clear understanding of how cloud computing works and the underlying infrastructure components, law enforcement agencies cannot properly investigate cloud computing environments (James et al., 2013). Some forensic practitioners have taken the view that the acquisition of evidence from the cloud is merely an extension of existing actions used to acquire data from a network because the cloud data is stored similar to network data (Adams, 2013). Contrary to this view, there are differences between cloud computing forensics and network forensics. The primary similarities and differences between cloud computing forensics and network forensics stem from the key areas of physical access, logical access, segregation of evidence, methodology, and legal considerations.

Ruan, Carthy, and Kechadi (2012) proposed that cloud forensics is a subsection of network forensics and a cross-discipline between cloud computing and digital forensics, adopting the same definitions of cloud computing and digital forensics provided by NIST. The methodology used originates at three distinct dimensions: legal, technical, and organizational. Choosing this methodology encompasses jurisdictional issues, the lack of international cooperation, and formation of foundational standards and sound policies. Ruan et al. (2012) developed a high-level conceptual model that analyzed the forensic im-

part of cloud computing environments based on deployment models, service offerings, cloud actor roles, segregation of duties, and forensic artifacts obtainable from the environment.

Adams (2013) proposed the advanced data acquisition model (ADAM). The model breaks down forensic processes into three main functions derived from analysis of previous models introduced between the years 2001 and 2004. The actions in the first stage occur upon notification of the scope of work to the investigator. The activities in this stage happen prior to arriving at the incident location. In the next stage, the investigator prepares for the data acquisition upon entry of the incident scene. The final stage consists of acquiring and handling the evidentiary data.

Adams (2013) used a distinctions, systems, relationships, and perspectives (DSRP) thinking patterns in creating ADAM, tying scientific thinking to the forensic processes. Initial testing of ADAM compared ADAM to the processes used in three previous investigations. The results produced a viable model because cases in Australia and New Zealand are seeing flowcharts introduced as a manner in which to help the jury and court understand scientific forensic evidence.

Shende (2010) approached the digital investigations of cloud computing environments as the mapping of computational and storage structures falling within the investigative scope. Emphasizing the cross boundaries of responsibility and access, service level agreements play a large part in the ability to investigate a cloud computing environment. The key to data access in cloud computing investigations lies in the dialogue between the organization and cloud service vendor in establishing incident response requirements, processes, and protocols when drafting a service level agreement (SLA) (Shende, 2010).

Lu, Lin, Liang, and Shen (2010) proposed using a secure source structure based on the bilinear pairing technique to deliver reliable digital forensic evidence in cloud computing environments. Using a cloud computing architecture comprised of a cloud service provider (SP), a trusted sys-

tem manager (SM), and an enterprise size user environment, the secure provenance SP method demonstrates confidentiality of information, authorized access, anonymous authentication, and ownership tracking. Extending this method for forensic use provides ownership information associated with every version of a document to the SM. The SM then uses an ownership-tracking algorithm to follow the evidence to a unique user identity. This proposed design encompasses a wide range of security features, making it ideal for use in the forensic investigation of cloud computing environments.

Witek and Daranyi (2012) took an approach similar to Lu et al. (2010), in that the work focused on the tracking of digital assets. Using a digital preservation (DP) policy methodology ensures that digital objects remain valid over an extended time period. Cloud computing infrastructure provides a platform ad-hoc computation along with distributed digital preservation. Witek and Daranyi (2012) proposed a middleware solution in the business services layer of the Sustaining Heritage Access through Multivalent Archiving (SHAMAN) core infrastructure. The business layer serves as the intermediary between the top presentation layer and the lower infrastructure layer. The proposed solution has promising future implications for digital preservation.

Current Theory Approaches

Bayesian theory and Dempster-Shafer theory are among the few theories that have any documented research on forensic application. There is widespread use of Dempster-Shafer theory in information security because it applies in many areas such as risk assessment, attack analysis, and forensic image analysis. In forensic analysis, combining Dempster-Shafer theory with other theories and methodologies allows a more accurate means to identify evidence. Current methods of application using Bayesian theory and Dempster-Shafer theory have limitations where cloud computing is concerned (Zhou & Mao, 2012). Bayesian and

Dempster-Shafer approaches use the mathematics of probability theory and numerical measures of uncertainty. These theories are applicable in intrusion detection systems, data mining techniques, and data privacy, but cloud computing environments have too many variables to postulate correct hypotheses for successful mathematical calculation (Chou, 2011). Additionally, the objective of the research is to provide a framework for human use. Bayesian and Dempster-Shafer theories are machine-based applicable theories that are not conducive to human application (Zhou & Mao, 2012).

Although there is proven application of Dempster-Shafer to forensic image analysis, this is a limited use application. Little is known about the application of current forensic evidence acquisition methods to cloud computing environments (Lallie & Pimlot, 2012).

APPLYING TRADITIONAL ACQUISITION METHODS TO CLOUD COMPUTING

Although there are theories about when traditional forensic acquisition methods are applicable in cloud forensic environments, when the methods can be modified, and when new methods are required, it is not currently known which traditional forensics methods apply to cloud computing environments, which traditional processes can be modified, and when new methods are required to be developed. Carlton (2007) used a grounded theory methodology to identify the tasks forensic examiners perform during forensic data acquisitions. Carlton's (2007) study identified distinct acquisition procedures for live or running systems that include performing a random access memory (RAM) dump, collecting volatile data, performing a live acquisition, and determining the programs currently running. Concluding tasks commonly used by examiners include preserving and sealing original acquisition media, placing original media in an evidence vault, performing an inspection to

verify the acquired image is readable, and creating a restore image of the original media for return to the owner (Carlton, 2007).

The acquisition tasks of Carlton's (2007) survey were reviewed for relevance and the 20 top ranked tasks were selected. A panel of international experts then reviewed the tasks for applicability to cloud computing environments. The findings indicated that about 50% of the current digital acquisition processes applied to cloud computing environments; the rest required modification or new process development. Post-acquisition processes were most suited for application in cloud computing environments. Following post-acquisition processes in order of applicability were live acquisition processes. The main limitation in this area was access to the cloud server in order to perform the processes. The results analysis suggested this category of processes had solutions for digital evidence acquisitions in cloud computing environments because the processes were modeled after already established network forensic processes.

Ten percent of the 20 pre-selected traditional forensic processes required the development of new processes for the forensic acquisition of digital evidence in cloud computing environments. The panel experts agreed that current acquisition processes in the category of dead drive acquisitions did not fit into cloud computing environments and many of the processes should no longer be included because they were not applicable or modifiable. From the results, the contingency framework shown in Table 1 was built based on the understanding of when current traditional forensic processes apply, identifying when process modification is acceptable, and when the development of new processes is required for forensic evidence acquisitions in cloud computing environments. The categories for forensic evidence acquisition activity are summarized as pre-acquisition, acquisition, imaging, and post-acquisition processes.

The digital forensic acquisition cloud contingency model combines the key contributions from the process design contingency model proposed

Table 1. Contingency framework

Process Category	Recommended Application	Contingency Variable(s)
Pre-acquisition processes	Modification	Fluidity of environment
Live acquisition processes	Application with limitations	Access, tools, scope
Dead acquisition processes	Partial application, modification or develop new	Cloud implementation, access
Imaging processes	Application, modification or develop new	Cloud implementation, access, scope
Post-acquisition processes	Application with limitations	Fluidity of environment, access

by Austin and Devin (2009) and the information security management contingency model demonstrated by Tassabehji (2005).

The model depicts the processes first because processes are the only constant. Once the digital forensic examiner choose the process to be performed, the cloud computing environment uncertainty is introduced which produces the contingency. Based on the contingency, the appropriate application methodology is executed. Figure 1 represents a diagrammatical illustration of the digital evidence forensic acquisition cloud contingency model.

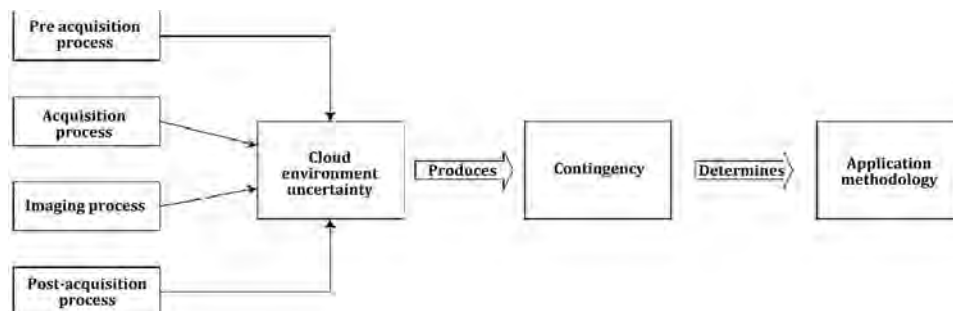
The contingency framework in Table 1 used the panel results as the foundation to create a digital forensic acquisition cloud contingency model, concentrating on how uncertainty and contingencies affect particular processes and guide a course of research that can support and enrich the model. When the world changes in a way that requires different functionality in a process or practice currently used, flexibility allows adjustment of the process or practice to reflect the changed world (Austin & Devin, 2009).

RECOMMENDATIONS

When forensic examiners lack knowledge about the environments from which evidence is being acquired, there is a high likelihood that critical information is missed during the acquisition process (Hale, 2013). Tools will gradually become outdated and computer forensic practitioners will no longer be able to rely on forensic analysis results, unless the forensic community formulates a vibrant strategy for developing methods that build upon each other (Garfinkel, 2010). Cloud forensics is a relatively new area of digital forensic practices with few industry professionals capable of providing required training (Ruan et al., 2012).

Currently, NIST and the Scientific Working Group on Digital Evidence (SWGDE) offer policy and guidance on a federal level. The digital evidence forensic acquisition cloud contingency model does not fit into the SWDGE model because the digital evidence forensic acquisition cloud contingency model has contingent variables requiring deviation from the proven reliable processes established in the SWDGE model (SWDGE,

Figure 1. A diagrammatical illustration of digital evidence forensic acquisition cloud contingency model



2014). When working groups create models on which policy is based, a broader more flexibly model and policy is required to reflect the fluid environment that is now part of every business and most digital forensic acquisitions.

The organizations and universities that build and deliver curriculum in digital forensic areas need to participate in the advancement of the digital forensics field by addressing cloud computing acquisitions. Guidance organizations must work together to provide solid direction in order to establish agreement on basic definitions and simplified models regarding cloud computing environments to help the industry move forward so practitioners accept and implement suggested guidance.

Academia and the digital forensic training community will need to create and encourage the development of new training programs so that practitioners may better respond to situations where the acquisition of cloud computing environments are required. Standards will also have to be set by legal departments where service level agreements are concerned and when other parties must be relied upon to provide evidence. In setting new standards for proper acquisition of forensic evidence in cloud computing environments, organizations must create legal requirement documentation encompassing both vendor contracts and the organization. After the legal requirements are set, digital forensic practitioners have the availability to set best practices, but those best practices will be contingent upon the existing conditions.

FUTURE RESEARCH DIRECTIONS

The opportunity for researchers to make innovative contributions and have a substantial impact on the cloud computing industry has only just begun. The findings in this panel study are a bridge to a very small body of literature, setting the stage for new and promising lines of inquiry. The results of the study produced a contingency framework

and digital evidence forensic acquisition cloud contingency model to help guide a course of research that can test the model.

Using contingency frameworks to address other research questions provides a fresh perspective on the application of digital forensic acquisition methods to cloud computing environments. Additional studies could firmly establish that the choices available for the application of digital forensic methods to cloud computing environments are ingrained in contingency frameworks. One of the significant contributions of this panel study is the identification of contingency factors such as available tools, access, availability, and acquisition scope as the underlying elements when choosing the application of digital forensic methods to cloud computing environments. These contingency factors are easily ported to other evidence acquisition methods for expanding research in this area.

The digital evidence forensic acquisition cloud contingency model suggests other important directions of research. Accepting a contingency perspective on how to choose digital process application in cloud computing environments can serve as a powerful theoretical lens both in interpreting the results of prior models and in shaping rigorous research models for future inquiry. Another direction for future research would be to examine the influence of multiple contingencies on the process application within individual cloud types.

CONCLUSION

As an industry, digital forensics is lacking the tools, published processes, and guidance for proper acquisition of digital evidence in cloud computing environments. Pre-acquisition processes are most suited for modification in cloud computing environments while post-acquisition processes are most suited for application in cloud computing environments. The digital acquisition processes that applied to cloud computing environments were modeled after already established

network forensic processes. The cloud computing environment uncertainty produces contingencies that affect selecting the appropriate acquisition process application.

The result of this panel study was a developed robust contingency framework for deciding when to use traditional forensic acquisition practices, when to use modified processes, and when it is necessary to develop new forensic acquisition processes through the evaluation of 20 conventionally recognized forensic acquisition processes by a panel of experts. The presented contingency framework used the study results as the foundation to create a digital forensic acquisition cloud contingency model, concentrating on how uncertainty of cloud computing environments and contingencies affect particular processes. The contingencies are easily ported to other evidence acquisition methods for expanding research in this area.

REFERENCES

- Adams, R. (2013). The emergence of cloud storage and the need for a new digital forensic process model. In K. Ruan (Ed.), *Cybercrime and cloud forensics: Applications for investigation processes* (pp. 79–104). doi:10.4018/978-1-4666-2662-1.ch004
- Almutairi, A., Sarfraz, M., Basalamah, S., Aref, W., & Ghafoor, A. (2012). A distributed access control architecture for cloud computing. *IEEE Software*, 29(2), 36–44. doi:10.1109/MS.2011.153
- Austin, R. D., & Devin, L. (2009). Weighing the benefits and costs of flexibility in making software: Toward a contingency theory of the determinants of development process design. *Information Systems Research*, 20(3), 462–479. doi:10.1287/isre.1090.0242
- Berman, S. J., Kesterson-Townes, L., Marshall, A., & Srivathsa, R. (2012). How cloud computing enables process and business model innovation. *Strategy and Leadership*, 40(4), 27–35. doi:10.1108/10878571211242920
- Budriene, D., & Zalieckaite, L. (2012). Cloud computing application in small and medium-sized enterprises. *Issues Of Business & Law*, 4(1), 99–130. doi:10.520/ibl.2012.11
- Carlton, G. H. (2007). A grounded theory approach to identifying and measuring forensic data acquisition tasks. *Journal of Digital Forensics, Security and Law*, 2(1), 35–56. Retrieved from <http://www.jdfsl.org/>
- Cho, C., Chin, S., & Chung, K. S. (2012, July). Cyber forensic for hadoop based cloud system. *International Journal of Security and Its Applications*, 6(3), 83–90. Retrieved from <http://www.sersc.org/journals/IJSIA/>
- Chou, T. S. (2011). Cyber security threats detection using ensemble architecture. *International Journal of Security and Its Applications*, 5(2), 11–15. Retrieved from <http://www.sersc.org/journals/IJSIA/>
- Decker, M., Kruse, W., Long, B., & Kelley, G. (2011). *Dispelling common myths of “live digital forensics”*. Retrieved from <http://www.dfcb.org/docs/LiveDigitalForensics-MythVersusReality.pdf>
- Dykstra, J., & Sherman, A. T. (2012, August). Acquiring forensic evidence from infrastructure-as-a-service cloud computing: Exploring and evaluating tools, trust, and techniques [Supplement]. *Digital Investigation*, 9, S90–S98. doi:10.1016/j.diin.2012.05.001
- Fiterman, E. M., & Durick, J. D. (2010). Ghost in the machine: Forensic evidence collection in the virtual environment. *Digital Forensics Magazine*, 2, 73–77. Retrieved from <http://www.digitalforensicsmagazine.com/>
- Garfinkel, S. L. (2010). Digital forensics research: The next 10 years. *Digital Investigation*, 7(Supplement), S64–S73. doi:10.1016/j.diin.2010.05.009
- Greengard, S. (2012, November). On the digital trail. *Communications of the ACM*, 55(11), 19–21. doi:10.1145/2366316.2366323

Hale, J. S. (2013, October). Amazon cloud drive forensic analysis. *Digital Investigation*, 10(3), 259–265. doi:10.1016/j.diin.2013.04.006

Huth, A., & Cebula, J. (2011). The basics of cloud computing. *United States Computer Emergency Readiness Team (US-CERT)*. Retrieved from http://www.us-cert.gov/reading_room/USCERT-CloudComputingHuthCebula.pdf

James, J. I., Shosha, A. F., & Gladyshev, P. (2013). Digital forensic investigation and cloud computing. In K. Ruan (Ed.), *Cybercrime and cloud forensics: Applications for investigation processes* (pp. 1–41)., doi:10.4018/978-1-4666-2662-1.ch001

Lallie, H., & Pimlott, L. (2012). Challenges in applying the ACPO principles to cloud forensic investigations. *Journal of Digital Forensics Security and Law*, 7(1), 71–86. Retrieved from <http://www.jdfsl.org/>

Lu, R., Lin, X., Liang, X., & Shen, X. S. (2010, April). Secure provenance: the essential of bread and butter of data forensics in cloud computing. *Proceedings of the 5th ACM Symposium on Information, Computer and Communications Security*, 282–292. doi:10.1145/1755688.1755723

Ruan, K., Carthy, J., & Kechadi, T. (2012, May). NIST cloud computing reference model and its forensic implications. *Journal of Digital Forensics, Security and Law, Conference Proceedings*. Retrieved from http://www.digitalforensics-conference.org/subscriptions/proceedings_2012.htm

Shende, J. (2010). Live forensics and the cloud. *DFI News*. Retrieved from <http://www.dfinews.com/article/live-forensics-and-cloud-part-1>

Shin, D. H. (2013). User centric cloud service model in public sectors: Policy implications of cloud services. *Government Information Quarterly*, 30(2), 194–203. doi:10.1016/j.giq.2012.06.012

Svecov, E. (2011). *An introduction to cloud forensics*. Retrieved from <http://blog.datacraft-asia.com/2011/01/an-introduction-to-cloud-forensics/>

Tassabehji, R. (2005). Principles for managing information security. *Encyclopedia of Multimedia Technology and Networking*, (pp. 842–848). doi:10.4018/978-1-59140-561-0.ch119

Werner, D. (2011). *Wake-up call*. Retrieved from: <http://www.defensenews.com/section/C4ISR/C4ISR-Journal C4ISR>

Wittek, P., & Darányi, S. (2012). Digital preservation in grids and clouds: A middleware approach. *Journal of Grid Computing*, 10(1), 133–149. doi:10.1007/s10723-012-9206-7

Zhou, X., & Mao, F. (2012, August). A semantics web service composition approach based on cloud computing. *Fourth International Conference on Computational and Information Sciences (ICCIS)*, 807–810. doi:10.1109/ICCIS.2012.43

ADDITIONAL READING

Almulla, S. A., Iraqi, Y., & Jones, A. (2014). State-of-the-art review of cloud forensics. *Journal of Digital Forensics. Security and Law*, 9(4), 7–28.

Dalins, J., Wilson, C., & Carman, M. (2015). Monte-Carlo Filesystem Search—A crawl strategy for digital forensics. *Digital Investigation*, 13, 58–71. doi:10.1016/j.diin.2015.04.002

Daryabar, F., Dehghantanha, A., & Udzir, N. I. (2013). A review on impacts of cloud computing on digital forensics. *International Journal of Cyber-Security and Digital Forensics*, 2(2), 77–94.

Desai, P., Solanki, M., Gadhwal, A., Shah, A., Patel, B. (2015, January) Challenges and Proposed Solutions for Cloud Forensic. *International Journal of engineering Research and Applications*, 1(5), 37–42

Farina, J., Scanlon, M., Le-Khac, N., & Kechadi, T. (2105, August). Overview of the Forensic Investigation of Cloud Services. *International Workshop on Cloud Security and Forensics (WCSF 2015)*.

Grispos, G., Storer, T., & Glisson, W. B. (2012). Calm Before the Storm: The Challenges of Cloud Computing in Digital Forensics. *International Journal of Digital Crime and Forensics*, 4(2), 28–48. doi:10.4018/jdcf.2012040103

Pătrașcu, A., & Patriciu, V. V. (2014). Digital Forensics in Cloud Computing. *Advances in Electrical and Computer Engineering*, 14(2).

Pichan, A., Lazarescu, M., & Soh, S. T. (2015). Cloud forensics: Technical challenges, solutions and comparative analysis. *Digital Investigation*, 13, 38–57. doi:10.1016/j.diin.2015.03.002

Shirkhedkar, D., & Patil, S. (2014). Analysis of Various Digital Forensic Techniques for Cloud Computing. *International Journal of Advanced Research in Computer Science*, 5(4), 104–107.

Zawoad, S., & Hasan, R. (2016, March). Trustworthy Digital Forensics in the Cloud. *Computer*, 49(3), 78–81. doi:10.1109/MC.2016.89

KEY TERMS AND DEFINITIONS

Cloud Computing: Cloud computing is an environment in which the Internet or the internal network of an organization seamlessly connects numerous host machines running on a virtualized platform to provide resources and software services.

Digital Forensics: Digital forensics is the division of forensic science that focuses on investigating and analyzing information artifacts located on digital devices involved in computer crime cases.

Infrastructure as a Service (IaaS): IaaS is an Internet delivered virtual environment in which the service provider supplies cloud infrastructure services that can include software, hardware, and networking equipment.

Multitenancy: Multitenancy is a software architecture in which one server hosts a multitude of different virtual environments for unrelated organizations, or tenants.

Platform as a Service (PaaS): PaaS is an Internet delivered virtual environment in which the service provider supplies cloud platform services that include the operating system and other platform-associated services.

Software as a Service (SaaS): SaaS is an Internet delivered virtual environment in which the service provider supplies commercial applications, eliminating associated purchasing and installation costs for the organization.

Service-Level Agreement (SLA): A SLA is a binding contract between service providers and purchasing organizations that specifies the service parameters and recourse if an unexpected business interruption occurs.

Virtual or Virtualized: Virtual or virtualized is a term used to explain the process of running an environment other than the native computing environment. fffffff

Internet–Facilitated Child Sexual Exploitation Crimes

Keith F. Durkin

Ohio Northern University, USA

Ronald L. DeLong

University of Dayton, USA

INTRODUCTION

The computer innovations of the late 20th Century had a transformative impact on society, revolutionizing many aspects of social life including education, commerce, and recreation. These technological developments have also impacted crime. Many types of criminal behavior have changed and new forms emerged in response to the advent of the Internet. Examples include hacking, cyberbullying, fraud, cyberstalking, and gambling. Internet crimes against children are among the most prominent manifestations of this phenomenon.

BACKGROUND

Internet crimes against are defined as those offenses that include an element of computer-facilitated sexual exploitation of minors (Alexy, Burgess, & Baker, 2005). First, the Internet is now the main mechanism for accessing and distributing child pornography. Second is the well-publicized problem of adults soliciting minors for sex. Third, the Internet is used to facilitate the commercial sexual exploitation of children (e.g., sex trafficking of juveniles, commercial distribution of child pornography). Fourth, the Internet serves as a social consolidation mechanism allowing adults with a sexual interest in children to network with each other. Additionally, the Internet plays a secondary role in the victimization of children

by allowing offenders who have a pre-existing relationship with a minor (e.g., family members or acquaintances) to use this medium to facilitate contact offending.

Offenses, Offenders, and Victims

The Internet is now the primary medium for accessing and distributing child pornography. These materials can now be instantaneously shared with, an anonymously accessed, by a global audience. For instance, an Interpol child abuse image data base contains more than a half a million images (Elliot & Beech, 2009). A recent study of the peer-to-peer file sharing network Gnutella found approximately 250,000 American computers shared 120,148 unique child pornography images during a one-year period (Wolak, Liberatore, & Levine, 2014). On the most fundamental level, these images represent a permanent record of the sexual exploitation of a minor (Lanning, 2010). The mere demand for this material fuels the further sexual abuse and exploitation of children (Seto, Hanson, & Babchishin, 2011). Concerns have been raised that this easy access to child pornography online might contribute to a new category of offender who succumbs to temptations that might have otherwise been held in check (Babchishin, Hanson, & VanZuylen, 2015).

Individuals who collect and/or traffic child pornography via the Internet are traditionally referred to as “traders” by law enforcement (Alexy et al., 2005). The National Juvenile Online Victimization

tion (NJOV) studies have gathered data from a national sample of law enforcement agencies regarding arrests for child pornography (Wolak, Finkelhor, & Mitchell, 2011; 2012a; 2012b). The results revealed that offenders in these cases are overwhelmingly single white men over the age of 25. Most of the offenders arrested on child pornography charges had material depicting prepubescent children, and a substantial portion had images of children less than 5 years of age. There appears to be an association between the possession of child pornography and engaging in contact sexual offenses against minors. An examination of 2009 arrest data suggests that one-third of suspects arrested for online child pornography offenses were actively molesting minors (Wolak et al., 2012a). Moreover, a meta-analysis of published studies of online (primarily child pornography) offenders found that approximately 1 in 8 of these men had a prior criminal record for contact sexual offenses (Seto et al., 2011). However, the same analysis found that about one-half of these individuals self-reported committing contact offenses. Similarly, Bourke and Hernandez (2009), who conducted a study of 155 offenders who were serving terms for online child pornography charges, found that although 26% of these men had official criminal records for sex crimes against minors, 85% of these men admitted to having committed at least one contact offense against a minor while in treatment.

A second category of Internet crimes against children involves adults using the Internet to solicit children for sexual purposes. These offenders have been called “travelers” by law enforcement (Alexy et al., 2005), and their online behavior has been referred to a “grooming”, “luring” or “predation” (Urbas, 2010). They engage in sexually orientated communications with minors, often with the intention of arranging offline encounters. They have been the subject of intense media scrutiny in the United States, and police stings aimed at catching these offenders were sensationally featured on the popular television show *Dateline NBC: To Catch a Predator*. Communication is frequently on-going, and a relationship between the of-

fender and target can develop over the course of several weeks or months. These solicitations are sometimes aggressive, and can contain explicit sexual content. Offenders may send pornography to their target, as well as exposing their genitals and masturbating on web cams (Marcum, 2007). They may also encourage minors to send them suggestive or explicit photos (Wolak et al., 2011). The most recent research suggests these offenders are increasingly turning to social networking sites to locate and groom potential victims (Balfe et al., 2015).

It is important to understand that the online behavior of these men frequently constitutes a crime. The sexually explicit communication with a minor may constitute the offense of enticement or importuning (DeLong, Durkin, & Hundersmarck, 2010). However, the media coverage of “Internet predators” tends to give an inaccurate picture of this type of offense. Based on research this evidence, this crime often fits the model of statutory rape, inasmuch as adult men try to develop a relationship with young teens and “seduce” them into an eventual encounter (Wolak, Finkelhor, Mitchell, & Ybarra, 2008). Data from the NJOV studies indicate that over 90% of the victims in these cases are between the ages of 13-17. Most of the victims are female adolescents. Violence or force is used less than 5% of these cases. A study of online solicitation in Europe produced very similar findings (Schulz, Bergen, Schuhmann, Hoyer, & Santtila, 2015). These offenders are normally not pedophiles, rather they are antisocial adult males who are willing to pursue sexual activities with minors who show signs of sexual development but who are below the age of consent (Seto, 2010).

Research has also focused on the risks factors associated with the online sexual solicitation of minors. Results of the second Youth Internet Safety Survey (YISS2), a national study of 11 to 17 year old adolescents, identified various factors that are associated with receiving aggressive online sexual solicitations (Mitchell, Ybarra, & Finkelhor, 2007). These included being female, visiting chat rooms, discussing sexual topics

online, and sharing personal information with individuals they met online. According to Wolak, Evans, Nguyen and Hines (2013), there is a higher risk of victimization for troubled youths, such as those with a history of abuse or poor parental relations, as well as boys who are gay or questioning. Many of the minors who are solicited online are not stereotypical innocents. These young people may be curious and rebellious adolescents who are actively seeking sexual information or contact (Lanning, 2010).

The Internet is playing a major role in the commercial sexual exploitation of juveniles (Mitchell, Jones, Finkelhor, & Wolak, 2011). With these offenses, children are exploited as commodities in the pursuit of profit. One example that is receiving a great deal of public attention is juvenile prostitution. Classified advertisement sites such as Craigslist can be used by the purveyors of child prostitution to place advertisements complete with a physical description of available children to potential consumers online (McCabe, 2008). In a study of Internet-related juvenile prostitution cases, Wells, Mitchell, and Ji (2012) discovered more than two-thirds of the victims were age 15 or younger, and in more than one-quarter of these cases the adult exploiter was a family member or acquaintance of the victim. The Internet is also used to promote sexual tourism. This phenomenon, which involves traveling with the intent to engage in sexual behavior, often, involves child victims. For a number of years, sex tours in Southeast Asia involving juvenile workers have been promoted via the Internet (Hughes, 2000). There have been reports of pedophiles trading information about potential child victims for pedophiles traveling abroad (Holt, Blevins, & Burkert, 2010).

Moreover, the Internet can serve a networking or social consolidation function for individuals who are sexually attracted to children. There is an entire subculture online of adults with a sexual interest in minors (Holt et al., 2010). This consists of a myriad of websites, computer forums, and

chat rooms. Adult-child sex advocacy groups such as NAMBLA (the North American Man/Boy Love Association) are remarkably active online. There is a vast array of literature, such as BoyWiki and The Boylove Manifesto, which supports the pedophile viewpoint (Durkin & Hundersmarck 2008). Although this networking is generally not criminal (Lanning, 2010), there are serious concerns that these networks may encourage criminal conduct. Research on this phenomenon has concluded these sites were criminogenic since they foster relationships among users and expose these users to rationalizations and justifications (e.g., this behavior does not harm children, children can freely consent to sex with an adult) conducive to sexual offending (Balfie et al., 2015; D'Ovidio, Mitnam, El-Burki, & Shumar, 2009; Durkin & Bryant, 1999; O'Halloran & Quayle, 2010). Furthermore, Schulz et al. (2015) found that among adults who solicit juveniles online, those individuals who participate in pedophilia websites and forums are much more likely to target minors age 13 and younger.

The Internet also plays a secondary (rather than primary) role in the sexual abuse of children. An offender can use computer technology to facilitate sex crimes against juveniles whom they already have a face-to-face relationship with (Wolak et al., 2012b). These offenders can be relatives, teachers, people in the ministry, coaches, or family friends. They can use chat or e-mail to seduce or "groom" the victim, as well as to arrange meetings where the abuse will occur. Data from an NJOV study reveals that nearly half of the victims in these cases are between the ages of 6 and 12 (Mitchell, Finkelhor, & Wolak, 2005). This is significant since the supra majority of victims in Internet initiated cases are older juveniles (between the ages of 13 and 17). Additionally, Internet pornography, including child pornography, can be used by sex offenders to lower the inhibitions of children they already have a face-to-face relationship with, thus facilitating their eventual victimization (Lanning, 2010).

Assessment and Classification of Internet Sex Offenders

Since this is a relatively new type of offending many different types of general sex offender assessments are currently utilized (e.g., pre-sentence, treatment planning and placement, classification, and risk assessment). Due to the unknown elements of associated with this type of offender; risk-level appears to be most requested type of assessment. Additional goals of this assessment should focus level of deviance, identification of pro-offending attitudes, accountability, level of denial and general psychosocial areas (see DeLong et al., 2010). Questions should be formulated in a manner to ascertain the degree, nature, and extent of the sexual and legal problems of the offender. Important in this process is the clinician's ability to address areas that may be uncomfortable for general practitioners such as sexual practices, fantasy, use and description of sexually explicit materials (including child pornography)(Quayle & Taylor, 2002). One should be aware of the offender's level of guardedness and defensiveness, as well as their possible high degree of shame and guilt. Asking questions in a gentle, nonjudgmental and non-confrontational manner is imperative to have an environment where the offender can discuss their offense dynamics and private areas of their life without the fear of judgment and retribution.

Quayle and Taylor (2002) developed a four factor framework that could be utilized in a semi-structured setting in order to address Internet specific issues during an assessment and to identify possible offense cycle dynamics. These four areas are: action, reflection, excitement, and arousal. In *action*, information obtained is from questions addressing length of time on the Internet, contact with others, emotional withdraw from significant others, degree of pleasure received from the activities, level of social isolation, images obtained or developed and sent. *Reflection* focuses on the offender's level of preoccupation with re-living past experiences; details of other online persons contacted; breaking promises about

quitting; and difficulties in concentrating or keeping off-line. *Excitement* addresses degree and type of risks in accessing materials and persons; if the materials were obtained while others were present (especially child pornography with children present); if the offender corresponds with others with a sexual interest in children; and attempts to make contact with children. *Arousal* involves the offender's level of masturbation associated with online activities; changes in sexual behaviors since beginning online activities; changes in the receipt of images and text messages particularly with children; material accessed; and if arousal happens with other non-child images.

Assessing the dynamics of the Internet offending process may help to formulate an explanation of online behaviors in order to understand the motive, intent, and goal(s) of the offender. This will assist the clinician in identifying factors pertinent to establishing treatment goals for the offender. Unfortunately, *the Diagnostic and Statistical Manual of Mental Health Disorders, Fifth Edition* does not identify a specific diagnosis or diagnostic criteria for Internet sex offenders. Therefore, the utilization of offense-specific typologies could appear a logical instrument for the understanding and classification of this population. There are a variety of typologies which attempt to identify a specific set of criteria based on the behavior demonstrated by the online behaviors and processes of the offender (see Seto, 2013 for a listing). These can permit a clinician or other evaluator to simplify and organize the complexities of the behaviors in order to understand similarities and differences in these offenses and offenders. However, it is imperative to note that the typology should be a tool for understanding and assimilating the behaviors, not a diagnosis.

Important in the development of the typologies is Cusson's (1993, in Taylor & Quayle, 2006) multi-stage explanation of the criminal process with Internet sex offenders which focuses on the tasks of search, pre-criminal situations, pre-criminal opportunity, and criminal tactics. This process identifies the following stages: 1) the of-

fender searchers for a need to find a pre-criminal situation, 2) once found the pre-criminal stage is introduced with the potential to commit an offense, 3) once the perceived payoff exceed the risk for the offender, the pre-criminal opportunity is established thereby allowing the offender to seek online communications, and 4) criminal tactics are utilized in order for the offender to demonstrate a sequence of events for the act to occur. Although offenses seemingly well thought out, the offender does not take into account that the actions are restrictive due to the physical limitation of the communicative processes and technologies (Taylor & Quayle, 2006). Therefore, an important consideration involved in whether or not an offender will travel to engage a physical meeting with an intended victim, is the method of communication and ease of transmission regarding messages, images and other items of interests for the offender (Elliot & Beech, 2009; DeLong, et al., 2010).

Treatment of Internet Sex Offenders

Andrews and Bonta (2010) promoted the concept and use of risk, need, and responsivity (RNR), with each area identifying a specific principle in which the offender would be treated. The “*risk principle*” addresses the intensity of services for the offenders should be in line with their level of recidivism risk. Thus, this assumes that the lower risk offender would need less restrictive services versus the high risk offender. The “*need principle*” focuses on the dynamic risk factors which are consistent to the criminogenic needs of the offender. This principle attests that dynamic risk factors which can be changed which are consistent with the offender’s recidivism risk. Finally, the “*responsivity principle*” identifies that when treatment is utilized, the information provided should meet the level and skill of the offender for maximum effect. Although many other factors are important (i.e., client-counselor relationship, genuine approach by the counselor, etc.), the impact by the utilization of the RNR ap-

proach has been found to be a much more effective treatment dynamic (Hanson, Bourgen, Helmus, & Hodgson, 2009).

Since the inception of online offending, treatment options have increased but are not absolute in the consistency of treatment. In fact, the need for separating contact and Internet-only offenders has been addressed as imperative since the treatment needs of each group differs. Seto (2013) identified several treatment programs that have increased in use and awareness. For instance, sexual addiction groups and cybersex addiction programs have increased in availability. A cybersex addictions model has been developed by the Internet Behavior Consulting group, which has a combination of their own workbook, relapse prevention, cognitive therapy, and a 12-step program approach (<http://www.internetbehavior.com>). Because of the research suggesting some Internet-only offenders are lower in risk (Seto, 2013), such treatment would not have to be lengthy or intensive but focused on the dynamics of RNR principals. It would be of importance to assess online behaviors which are central to this behavioral pattern: length of computer time with general viewing, legal adult pornographic materials, social networking sites, and visiting/interacting with other likeminded persons regarding their sexual interests. However, increased risk (online offending which progresses or used in concert to physical sex offending) would need more intensive treatment focusing on both dynamics of online and direct contact offending.

SOLUTIONS AND RECOMENDATIONS

Law enforcement agencies in the U.S. have taken a very active role in combating Internet crimes against children, and it has been estimated that 95% of their criminal cases involving these offenses result in guilty pleas or convictions (Wolak, Finkelhor, & Mitchell, 2009). In 2009, U.S. law enforcement agencies made an estimated 8,194 arrests for technologically-facilitated child sexual

exploitation offenses (Wolak et al., 2012b). The most common offenses are those related to child pornography, with approximately 5,000 arrests for child pornography possession in 2009 (Wolak et al., 2012a). Law agencies are becoming more aggressive in trying to locate child pornography offenders. This includes officers posing as fellow “traders” in proactive investigations, as well as tracking suspects through peer-to-peer file sharing services. This has resulted in a progressive increase in the number of child pornography arrests since 2000. On the other hand, arrests of offenders who solicit undercover police officers posing as minors have been decreasing in recent years (Wolak et al., 2012b). There are two possible explanations for this decline. First, resources used for proactive investigations have been shifted to child pornography offenders. Second, the extensive media coverage has raised public awareness about online solicitation of minors and has possibly deterred some adult men from engaging in this conduct.

One of the most popular theories in criminology, routine activities theory (Cohen & Felson, 1979), provides a useful framework to address strategies to prevent these offenses. According to this theory, crime is a function of everyday life, and it occurs when three factors simultaneously converge. The first factor is motivated offenders. These are individuals who are seeking a criminal opportunity. Victimization risks are considered a function of an individual’s exposure to offenders (Mustaine & Tewksbury 2000). The second factor is attractive targets. These are individual targets that are appealing to offenders --- whether it is a vulnerable individual or something worth stealing. The final factor is a lack of capable guardians—people who can discourage an offense from happening (Felson, 2001). Individuals can also exercise personal guardianship to prevent their own victimization. Proactive police investigations designed to catch individuals involved with child pornography or solicitation is one way to try and curb the presence of motivated offenders. Furthermore, prevention efforts should be targeted at children and adolescents to increase their levels

of personal guardianship. There is a need for developmentally appropriate prevention strategies that target high-risk behaviors (e.g., visiting chat rooms, corresponding with strangers, and discussing sexual matters online). Some examples include teaching children avoidance skills and educating adolescents about the problematic (and criminal) nature of sexual relationships with adults (Wolak et al., 2008).

CONCLUSION

As technology advanced, so did the opportunities for the exploitation of children. Internet crimes against children constitute a serious albeit misunderstood problem. Child pornography is now a worldwide problem with horrific consequences for victims. The adults who are involved with this material often represent a very serious threat to society. The online solicitation of minors is another contemporary problem. However, the majority of these cases are consistent with statutory rape. Proactive police investigations, particularly into child pornography, are an especially valuable asset in reducing Internet crimes against children. Moreover, realistic prevention strategies offer promise in further reducing online victimization. As technology continues to expand with handheld personal communication devices such as cellular phones, opportunities for sex offenses against minors will also expand. It is essential that society be aware of these possibilities, and be ready to act in an adaptable fashion to prevent the victimization of children and adolescents.

FUTURE RESEARCH DIRECTIONS

Since Internet crimes against children are a new phenomenon, research on this topic is in its infancy. There appear to be two especially compelling directions for future research. First, it will be important to know the relative risk of future contact offending for adults arrested for soliciting

minors online. Second, research should focus on determining which specific strategies are effective in controlling these offenses and preventing victimization of minors online.

REFERENCES

- Alexy, E. M., Burgess, A. W., & Baker, T. (2005). Internet offenders: Traders, travelers, and combination trader-travelers. *Journal of Interpersonal Violence, 20*(7), 804–812. doi:10.1177/0886260505276091 PMID:15914702
- Andrews, D. A., & Bonta, J. (2010). *The psychology of criminal conduct* (5th ed.). Cincinnati, OH: Anderson.
- Babchishin, K. M., Hanson, R. K., & VanZuylen, H. (2015). Online child pornography offenders are different: A meta-analysis of the characteristics of online and offline sex offenders against children. *Archives of Sexual Behavior, 44*(1), 45–66. doi:10.1007/s10508-014-0270-x PMID:24627189
- Balfe, M., Gallagher, B., Masson, H., Balfe, S., Brugha, R., & Hackett, S. (2015). Internet child sex offenders concerns about online security and their use of identity protection technologies: A review. *Child Abuse Review, 24*(6), 427–439. doi:10.1002/car.2308
- Bourke, M. L., & Hernandez, A. E. (2009). The Butner Study redux: A report of the incidence of hands-on child victimization by child pornography offenders. *Journal of Family Violence, 24*(3), 183–191. doi:10.1007/s10896-008-9219-y
- Cohen, L. E., & Felson, M. (1979). Social change and crime rate trends: A routine activity approach. *American Sociological Review, 44*(4), 588–608. doi:10.2307/2094589
- D'Ovidio, R., Mitnam, T., El-Burki, J., & Shumar, W. (2009). Adult-child sex advocacy websites as social learning environments. *International Journal of Cyber Criminology, 3*, 421–440.
- DeLong, R., Durkin, K. F., & Hundersmarck, S. (2010). An exploratory analysis of the cognitive distortions of a sample of men arrested in Internet sex stings. *Journal of Sexual Aggression, 16*(1), 59–70. doi:10.1080/13552600903428235
- Durkin, K. F., & Hundersmarck, S. F. (2008). Pedophiles and child molesters. In E. Goode & D. Angus Vail (Eds.), *Extreme Deviance* (pp. 144–150). Thousand Oaks, CA: Pine Forge Press.
- Elliott, I. A., & Beech, R. (2009). Understanding online pornography use: Applying sexual offense theory to Internet offenders. *Aggression and Violent Behavior, 14*(3), 180–193. doi:10.1016/j.avb.2009.03.002
- Felson, M. (2001). Routine activity theory: The theorist's perspective. In C.D. Bryant (Ed.), *Encyclopedia of Criminology and Deviant Behavior, Volume I: Historical, Conceptual, and Theoretical Issues* (pp. 338–339). Philadelphia: Brunner-Routledge.
- Hanson, R. K., Bourgon, G., Helmus, L., & Hodgson, S. (2009). The principals of effective correctional treatment also apply to sexual offenders: A meta-analysis. *Criminal Justice and Behavior, 36*(9), 865–891. doi:10.1177/0093854809338545
- Holt, T. J., Blevins, K. R., & Burkert, N. (2010). Considering the pedophile subculture online. *Sexual Abuse, 22*, 3–24. PMID:20133959
- Hughes, D. (2000). Welcome to the rape camp: Sexual exploitation and the Internet in Cambodia. *Journal of Sexual Aggression, 6*(1-2), 1–23. doi:10.1080/135526000008413308
- Lanning, K. V. (2010). *Child molesters: A behavioral analysis*. Washington, DC: National Center for Missing & Exploited Children.

- Marcum, C. (2007). Interpreting the intentions of Internet predators: An examination of online predatory behavior. *Journal of Child Sexual Abuse*, 16(4), 99–114. doi:10.1300/J070v16n04_06 PMID:18032248
- McCabe, K. A. (2008). The role of Internet service providers in cases of child pornography and child prostitution. *Social Science Computer Review*, 26(2), 247–251. doi:10.1177/0894439307301438
- Mitchell, K. J., Finkelhor, D., & Wolak, J. (2005). The Internet and family and acquaintance sexual abuse. *Child Maltreatment*, 10(1), 49–60. doi:10.1177/1077559504271917 PMID:15611326
- Mitchell, K. J., Jones, L. M., Finkelhor, D., & Wolak, J. (2011). Internet-facilitated commercial sexual exploitation of children: Findings from a nationally representative sample of law enforcement agencies in the United States. *Sexual Abuse*, 23, 43–71. PMID:20852011
- Mitchell, K. J., Ybarra, M., & Finkelhor, D. (2007). The relative importance of online victimization in understanding depression, delinquency, and substance abuse. *Child Maltreatment*, 12(4), 314–324. doi:10.1177/1077559507305996 PMID:17954938
- Mustaine, E. E., & Tewksbury, R. (2000). Comparing the lifestyles of victims, offenders, and victim-offenders: A routine activity theory assessment of similarities and differences for criminal incident participants. *Sociological Focus*, 33(3), 339–362. doi:10.1080/00380237.2000.10571174
- OHalloran, E., & Quayle, E. (2010). A content analysis of a boy love support forum: Revisiting Durkin and Bryant. *Journal of Sexual Aggression*, 16(1), 71–85. doi:10.1080/13552600903395319
- Quayle, E., & Taylor, M. (2002). Pedophiles, pornography, and the Internet: Assessment issues. *British Journal of Social Work*, 32(7), 863–875. doi:10.1093/bjsw/32.7.863
- Schulz, A., Bergen, E., Schuhmann, P., Hoyer, J., & Santtila, P. (2015). Online Sexual Solicitation of Minors How Often and between Whom Does It Occur? *Journal of Research in Crime and Delinquency*, 53(2), 165–188. doi:10.1177/0022427815599426
- Seto, M. (2010). Child pornography use and Internet solicitation in the diagnosis of pedophilia. *Archives of Sexual Behavior*, 39(3), 591–593. doi:10.1007/s10508-010-9603-6 PMID:20182786
- Seto, M. (2013). *Internet sex offenders*. Washington, DC: American Psychological Association. doi:10.1037/14191-000
- Seto, M., Hanson, K. R., & Babchishin, K. M. (2011). Contact sexual offending by men with online sexual offenses. *Sexual Abuse*, 23, 124–145. PMID:21173158
- Taylor, M., & Quayle, E. (2006). The Internet and abuse images of children: Search, pre-criminal situations and opportunity. In R. Wortley & S. Smallbone (Eds.), *Situational Prevention of Child Sexual Abuse* (pp. 169–195). New York: Criminal Justice Press.
- Urbas, G. (2010). Protecting children from online predators: The use of covert investigation techniques by law enforcement. *Journal of Contemporary Criminal Justice*, 26(4), 410–425. doi:10.1177/1043986210377103
- Wells, M., Mitchell, K. J., & Ji, K. (2012). Exploring the role of the internet in juvenile prostitution cases coming to the attention of law enforcement. *Journal of Child Sexual Abuse*, 21(3), 327–342. doi:10.1080/10538712.2012.669823 PMID:22574847
- Wolak, J., Evans, L., Nguyen, S., & Hines, D. A. (2013). Online predators: Myth versus reality. *New England Journal of Public Policy*, 25, 1–6.
- Wolak, J., Finkelhor, D., & Mitchell, K. (2009). Trends in Arrests of “Online Predators”. Durham, NC: Crimes Against Children Research Center.

Wolak, J., Finkelhor, D., & Mitchell, K. (2011). Child pornography possessors: Trends in offender and case characteristics. *Sexual Abuse, 23*, 22–42. PMID:21349830

Wolak, J., Finkelhor, D., Mitchell, K., & Ybarra, M. L. (2008). Online predators and their victims: Myths, realities, and implications for prevention and treatment. *The American Psychologist, 63*(2), 111–128. doi:10.1037/0003-066X.63.2.111 PMID:18284279

Wolak, J., Finkelhor, D., & Mitchell, K. J. (2012a). Trends in Arrests for Child Pornography Possession: The Third National Juvenile Online Victimization Study (NJOV-3). Durham, NC: Crimes against Children Research Center.

Wolak, J., Finkelhor, D., & Mitchell, K. J. (2012b). Trends in Law Enforcement Responses to Technology-Facilitated Child Sexual Exploitation Crimes: The Third National Juvenile Online Victimization Study (NJOV-3). Durham, NC: Crimes against Children Research Center.

Wolak, J., Liberatore, M., & Levine, B. N. (2014). Measuring a year of child pornography trafficking by US computers on a peer-to-peer network. *Child Abuse & Neglect, 38*(2), 347–356. doi:10.1016/j.chiabu.2013.10.018 PMID:24252746

ADDITIONAL READING

Bourke, M. L., & Hernandez, A. E. (2009). The Butner Study redux: A report of the incidence of hands-on child victimization by child pornography offenders. *Journal of Family Violence, 24*(3), 183–191. doi:10.1007/s10896-008-9219-y

DeLong, R., Durkin, K. F., & Hundersmarck, S. (2010). An exploratory analysis of the cognitive distortions of a sample of men arrested in Internet sex stings. *Journal of Sexual Aggression, 16*(1), 59–70. doi:10.1080/13552600903428235

Elliott, I. A., & Beech, R. (2009). Understanding online pornography use: Applying sexual offense theory to Internet offenders. *Aggression and Violent Behavior, 14*(3), 180–193. doi:10.1016/j.avb.2009.03.002

Mitchell, K. J., Finkelhor, D., & Wolak, J. (2005). The Internet and family and acquaintance sexual abuse. *Child Maltreatment, 10*(1), 49–60. doi:10.1177/1077559504271917 PMID:15611326

Mitchell, K. J., Jones, L. M., Finkelhor, D., & Wolak, J. (2011). Internet-facilitated commercial sexual exploitation of children: Findings from a nationally representative sample of law enforcement agencies in the United States. *Sexual Abuse, 23*, 43–71. PMID:20852011

OHalloran, E., & Quayle, E. (2010). A content analysis of a boy love support forum: Revisiting Durkin and Bryant. *Journal of Sexual Aggression, 16*(1), 71–85. doi:10.1080/13552600903395319

Quayle, E., & Taylor, M. (2002). Pedophiles, pornography, and the Internet: Assessment issues. *British Journal of Social Work, 32*(7), 863–875. doi:10.1093/bjsw/32.7.863

Wolak, J., Finkelhor, D., Mitchell, K., & Ybarra, M. L. (2008). Online predators and their victims: Myths, realities, and implications for prevention and treatment. *The American Psychologist, 63*(2), 111–128. doi:10.1037/0003-066X.63.2.111 PMID:18284279

Wolak, J., Finkelhor, D., & Mitchell, K. J. (2012). Trends in Law Enforcement Responses to Technology-Facilitated Child Sexual Exploitation Crimes: The Third National Juvenile Online Victimization Study (NJOV-3). Durham, NH: Crimes against Children Research Center.

Wolak, J., Liberatore, M., & Levine, B. N. (2014). Measuring a year of child pornography trafficking by US computers on a peer-to-peer network. *Child Abuse & Neglect, 38*(2), 347–356. doi:10.1016/j.chiabu.2013.10.018 PMID:24252746

KEY TERMS AND DEFINITIONS

Child Molester: Any adult who engages in any type of sexual contact with an individual legally defined as a child.

Internet-Facilitated Child Sexual Exploitation: Offenses in which Internet technology is utilized for purposes of sexually exploiting minors.

Pedophilia: A psychiatric disorder involving an adult who experiences a primary or exclusive sexual attraction to prepubescent children (13-years-old or younger).

Traders: Individuals who traffic and/or collect child pornography.

Travelers: Individuals who use the Internet attempting to solicit minors for sexual purposes.

C

Knowledge-Based Forensic Patterns and Engineering System

Vivek Tiwari

International Institute of Information Technology Naya Raipur, India

R. S. Thakur

Maulana Azad National Institute of Technology, India

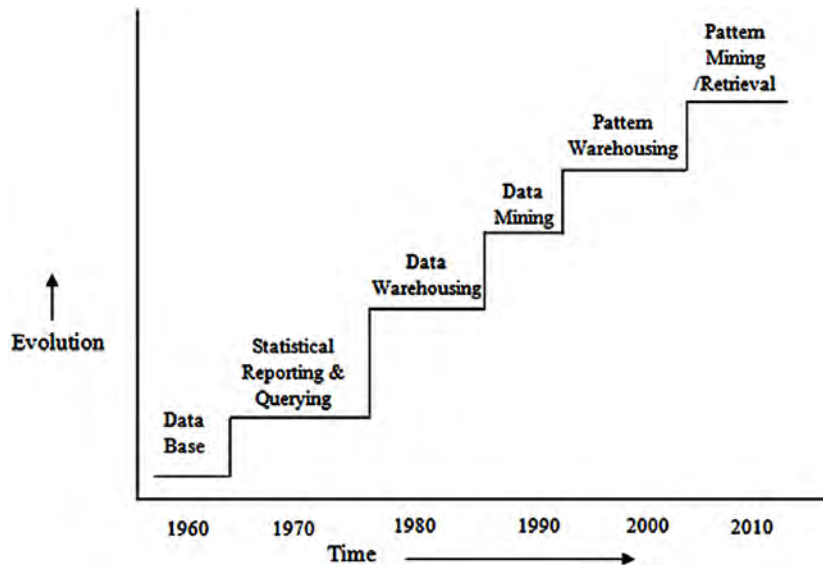
INTRODUCTION

The primary center behind this chapter is outline a devoted framework which can store forensic digital data patterns permanently. The issues, challenges, conceptual multilayer framework for developing forensic pattern warehouse for betterment of forensic prediction and forecasting has been discussed. A sequence of phases is involved during the design processes of forensic pattern warehouse and initially, conceptual modeling is one of the significant and the initial phase of forensic pattern warehouse design, because it constructs the solid framework for the next level of phases. In view of high semantic nature, there is a need to take a different approach for forensic pattern warehouse design to accommodate flexibility, isolation, extensibility and robustness. In this way, the context oriented forensic pattern warehouse design is one of the possible ways to manage patterns in a better way. The context of underlying pattern also helps to guide the queries to give more satisfactory result. Moreover, the logical modeling of the pattern warehouse should incorporate a mechanism to properly hierarchized the patterns and able to make sharp isolation among patterns. Furthermore, there is a need to provide additional information in logical modeling with patterns to improve query processing by revealing the sense of the underlying domain of source data.

BACKGROUND

In the mid of the 90s, organizations have started to recognize the strategic use of databases as a new discipline which was entirely different from theme of operational database (Tiwari V. at al 2010). Traditionally, operational database has been used to full fill mission-critical, day to day needs for online transaction processing (Inmon W.H., 2005). Organizations have a variety of computer based data processing system such as financial, making, feedback, attendance, and sales etc. which generate operational data (Agrawal R. & Srikant R. 1994). These kinds of data contain detailed, non-redundant and updated values. Those organizations that have recognized the power of information timely will have huge advantages over their competitors and it leads to design an effective data warehouse strategy (Kimball R. & Ross M. 2011). A data warehouse and data mining have given a platform to recognize the role of information behind successful business and much more. Data warehousing improves the productivity of an expert's decision making though consolidation, conversion, transformation, and integration of scattered data, and represents a trusted view of the enterprise. The data warehouse should not consider as a product rather than it is an environment. The data warehousing is a layered process to construct of information system that helps with

Figure 1. Evolution of database technology



organization in their decision making by giving historical data (Romero et al, 2010)). There are some following reasons that make data warehouse is very special:

- Prompt decision need to be made correctly by analyzing available data.
- Business workers are expert in their domain, not in the computer.
- The amount of data doubles in every one and half year, which create the problem in deep data analysis.
- The required infrastructure cost of development of a data warehouse is continuing to decline.
- Organizations have to take critical decisions based on the entire data rather than using rough estimates based on incomplete data.

In the recent evolution of database technology as depicted in Figure 1, patterns are being managed and maintained by Pattern Warehouse Management System (PWMS). Pattern warehouse is a brand-new concept and little emphasis has been given till date. A pattern warehouse is as attractive as data warehouse as the main repository of an

organization's historical pattern and is optimized for on-demand reporting and analysis (Mohammad R. et al (2009), Bartolini et al, (2003)). By nature, patterns are not persistent. There is a need to execute pattern generating methods when patterns are required. Pattern warehouse is a way to make the patterns persist by storing them permanently and it can be considered as a collection of persistently stored patterns.

In order for someone to be able to exploit these patterns on-demand, an efficient general PWMS is required for handling (storing/processing/retrieving) the patterns. The general means, the system must be able to handle all kinds of patterns. Patterns share some characteristics that make traditional DBMSs unable to represent and manage them. Patterns thus, can be regarded as knowledge units that effectively describe entire subsets of data (in this sense, they are compact) (Tiwari V. & Thakur R.S. 2015b). Patterns can be generated from different application domain so they are very heterogeneous in structures and often, heterogeneous patterns are needed to be managed together. The structure of patterns is required to be investigated for the design of a flexible pattern warehouse management system. Pattern management is an important issue in

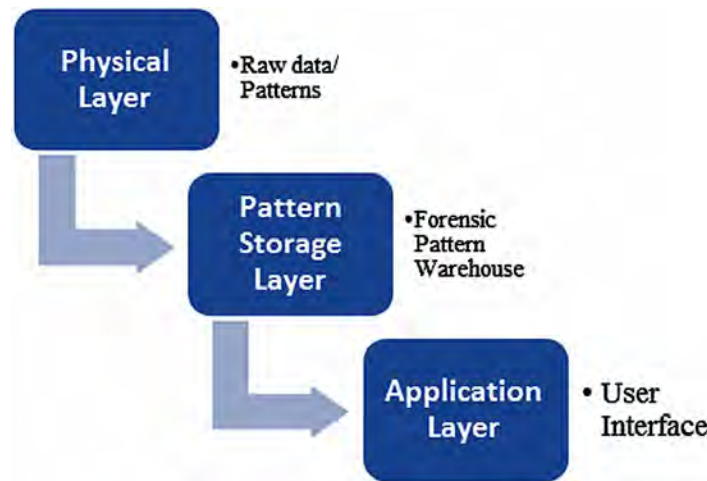
many different contexts and domains. The most important contexts in which pattern management is required are business intelligence and data mining. Business intelligence concerns a broad category of applications and technologies for gathering, storing, analyzing, and providing access to data to help enterprises in business decisions.

Moreover, many government and private forensic databases can help to both law enforcement investigators and the scientists who support their work in the lab. Forensic Pattern Warehouse (FPW) is a centralized forensic data repository that integrates forensic data from various transactional, legacy, applications and external sources. The Forensic Pattern warehouse provides an environment that is separate from the operational systems and is completely designed for decision-support, analytical-reporting, ad-hoc queries, and data mining. This isolation and optimization enables queries to be performed without any impact on the daily transactional and operational systems (Tiwari V. & Thakur R.S., 2015c). Benefits with a successful implementation of a forensic pattern warehouse include:

- Enhanced forensic intelligence.
- Increased processing of large and complex queries.
- Forensic intelligence from multiple sources.
- Instant access to forensic pattern (Save Time).
- Enhanced forensic pattern quality and consistency.
- Provide historical forensic pattern intelligence.
- Convert forensic data into actionable form.
- Increase scope of forensic data availability.
- Generates a high ROI (return on investment).
- Provide wide variety of forensic patterns.
- Provide various kinds of trends report.
- Decrease computational cost and increase productivity.

Decision making on forensic data is very crucial. Decisions that affect the strategy and operations of organizations will be based upon credible facts and will be backed up with evidence and actual data (Geradts Z. & Bijhold J, (2002), David L.W. & Andrew J. (2013)). Insights will be gained by using forensic pattern warehouse (FPW) through improved information access. A data warehouse is designed for storing large volumes of data and being able to rapidly query the data but forensic pattern warehouse is designed and constructed with a focus on speed of knowledge retrieval and analysis. Pattern warehouse does not concentrate for creation and modification of forensic data. In contrast, the forensic pattern warehouse is built for analysis and decision making. In conventional ways, forensic data are stored around transactional database and have limited accessibility (Geradts Z. & Bijhold J, (2001), Mikkonen S. & Astikainen T., (1994)). It is almost impossible to any organization to share their transactional databases. There are large amount of forensic data in country but scattered in organization and institute wise. They hardly exchange their forensic data for analysis purpose (Sibert R.W., (1994), Catherine et al. (2016)). The result of any analysis process is directly depends on amount of quality data. Decision making process lies on credibility of analysis. Analysis on single organization's forensic data cannot bring such credibility (Tobias K. et al, 2011). For many organizations, forensic information systems are comprised of multiple subsystems, physically separated and built on different Platforms and formats. Moreover, gathering data from multiple disparate data sources is a common need when conducting forensic data analysis for decision making (John H. et al, 2011). We need to gather forensic data from various small or big organizations time to time and need to put in central repository system (Werrett D.J., 1997). We perform integration of existing disparate data sources and make them accessible in one place. This repository system consisted of same grade (forensic) of data with different formats. These

Figure 2. Layered forensic pattern warehouse framework



technical issues can be covered in preprocessing phase. There is a separate forensic data integration system, known as ETL, within a forensic pattern warehouse environment (Tiwari V. & Thakur R.S. 2015a). This system consolidates forensic data from multiple source systems and transforms the data into a useful format. Further, user of forensic data will be able to query data directly with less information technology support. This allow to users to spend more time performing data analysis and less time gathering data.

MAIN FOCUS

A Multi-Layered Architecture

An elementary model is proposed to develop an overall integrated environment for uniformly representing (create, store and update) and querying various kinds of forensic patterns. A multi-layered architecture is introduced and considered one of the suitable solutions for forensic pattern warehouse management system. The architecture is made up of the number of interconnected layers which include Physical layer, Pattern storage layer and Application layer as depicted in Figure 2.

Physical Layer

Bottom of the architecture depicts the data store that contains the raw data. Raw data can be either managed by a DBMS (Database Management System) or can be stored in files, streams or any other external physical data source. External data sources are operational systems/flat files/ text files, etc., used to manage the data to support critical operations and relatively small number of well-defined business transactions (Golfarelli M. & Rizzi S., 2009).

Pattern Storage Layer

The middle layer contains patterns, which forms the core part of the pattern warehouse management system. The pattern stored in the pattern warehouse is uploaded by pattern interface. This layer can be further divided into Pattern-Tier and Context-Tier.

- **Pattern-Tier:** This is populated with patterns which are extracted through various techniques pushed here from the preceding layer.

- **Context-Tier:** Context gives the ability to organize patterns according the context and helps to give sense about what kind of knowledge they are carrying. It defines schema and procedure to organize patterns contextually. It holds built-in and user defined types for patterns and also it describes the syntax of the patterns.

Collectively, pattern tier and context tier form the pattern warehouse where patterns, trends, and knowledge reside. Pattern warehouse is the core of the pattern warehouse management system and it can think as a logical and compact representation of processed and mined data.

Application Layer

The top of the architecture consists application layer. This is the representation of other applications, machines, users, etc. The application layer of the pattern management architecture is the layer to the end-user deals with directly. This layer also includes the hardware and software involved in displaying and presentation of patterns.

Pattern Representation

The following section covers are various ways to define and declare the forensic patterns to keep the sematic of patterns and forensic data intact (Manolis T. & Panos V. (2003), Rizzi, S. et al (2003)).

Patterns are represented with triple (Pattern_Type, Pattern, Class).

Pattern_Type: A pattern type '*pt*' is a quintuple -

$$pt = (n, ss, ds, ms, f)$$

Where:

n is the name of pattern type.

ss (structure schema) is a definition of pattern space.

ds (source schema) define related raw data space.

ms (measure schema) quantify the quality

f is a function that describes the relationship between source space and pattern space.

Pattern: Let $pt = (n, ss, ds, ms, f)$ be a pattern type.

A pattern '*p*' instance of '*pt*' is a quintuple:

$$P = (pi, s, d, m, e)$$

Where:

pid- pattern identifier

s- is a value for type '*ss*'

d- dataset

m- is a value of type *ms*

e- region of the source space

Example of a pattern of association type is:

pid: 001

s: (head = { 'Thumb_Finger_Print' }, body = { 'Knife', 'Minor_Age' })

d: 'SELECT SETOF record FROM

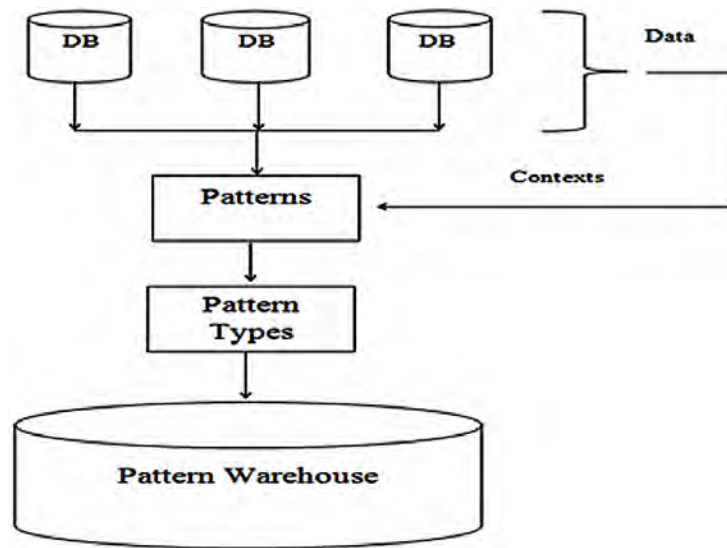
minor_age_prison_db GROUP BY recordID

m: (confidence = 0.75, support = 0.30)

e: { transaction: { 'Thumb_Finger_Print', 'Knife', 'Minor_Age' } }

In this chapter, the attention is given on the central issues: conceptual and logical schema design only. In the view, Figure 3 depicts very abstract the conceptual view of forensic pattern warehousing including data sources, patterns, type and pattern warehouse. There is not found context based conceptual and logical schema. The conceptual modeling of pattern warehouse is with clear goals and objectives, such as completeness (all kinds of patterns), summarizability (ability to compute aggregate or derived pattern), and knowledge Independence (every pattern can be answered using the pattern warehouse only) as constraints (Tiwari V. & Thakur R.S. 2015b).

Figure 3. An abstract conceptual view of pattern warehousing



FUTURE RESEARCH DIRECTIONS

The presented work covers conceptual the forensic pattern warehouse while discussion on physical modeling is missing. Future work includes the dynamics of changing, modifying and updating of forensic patterns in warehouse. We must address the pattern's storage technology and schema toward quantity, quality and heterogeneity. There are also need to explore about duplication, compression, optimization and pattern placement. Forensic pattern warehouse refreshment problem is very important to handle carefully.

CONCLUSION

An elementary model is proposed to develop an overall integrated environment for uniformly representing (create, store and update) and querying various kinds of forensic patterns. A multi-layered architecture is introduced and considered one of the suitable solutions for pattern warehouse management system. The architecture is made up of the number of interconnected layers which include Physical layer, Pattern storage layer and Application layer. In presented model, each layer

is proposed to perform concise and specific task. The model has enough flexibility and extensibility to use its own way. This allows the user to: 1) use standard pattern types or define new ones; 2) generate patterns and represented according to existing standards; 3) allow using new standards and mapping.

REFERENCES

- Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules. *Proc. 20th Int. Conf. Very Large Data Bases*, 1215, 487-499.
- Bartolini, L., Elisa, B., Barbara, C., Paolo, C., Matteo, G., Marco, P., & Rizzi, S. (2003). Patterns for Next-generation DAtabase systems: preliminary results of the PANDA project. *Proceedings of the Eleventh Italian Symposium on Advanced Database Systems*.
- Catherine, C., Aaron, K., & Robert, C. (2016). Development of a normalized extraction to further aid in fast, high-throughput processing of forensic DNA reference samples. *Forensic Science International: Genetics*, 25, 112-124.

- David, L. W., & Andrew, J. (2013). *Digital Forensics Processing and Procedures* (1st ed.). Elsevier.
- Geradts, Z., & Bijhold, J. (2001). New developments in forensic image processing and pattern recognition. *Science & Justice*, 41(3), 159–166.
- Geradts, Z., & Bijhold, J. (2002). Content Based Information Retrieval in Forensic Image Databases. *Journal of Forensic Sciences*, 47(2), 40–47.
- Golfarelli, M., & Rizzi, S. (2009). *Data warehouse design: Modern principles and methodologies*. McGraw-Hill, Inc.
- Inmon, W. H. (2005). *Building the data warehouse*. New York: John Wiley & Sons, Inc.
- John, H., David, J., & Lamb, M. (2011). A Framework for the Forensic Investigation of Unstructured Email Relationship data. *International Journal of Digital Crime and Forensics*, 3(3).
- Kimball, R., & Ross, M. (2011). *The data warehouse toolkit: the complete guide to dimensional modeling*. John Wiley & Sons.
- Manolis, T., & Panos, V. (2003). *Architecture for Pattern Base Management Systems*. Department of Electrical and Computer Engineering. National Technical University of Athens.
- Mikkonen, S., & Astikainen, T. (1994). Database classification system for shoe sole Patterns - identification of partial footwear impression found at a scene of crime. *Journal of Forensic Sciences*, 39(5), 1227–1236.
- Mohammad, R., Keivan, K., Reda, A., & Mick, J. (2009). Data modeling for effective data warehouse architecture and design. *International Journal of Information and Decision Sciences*, 1(3), 282–300.
- Rizzi, S., Bertino, E., Catania, B., & Terrovitis, M. (2003). Towards a logical model for patterns. *Proceedings. ER Conference*.
- Romero, Oscar, & Abelló. (2010). A framework for multidimensional design of data warehouses from ontologies. *Data & Knowledge Engineering*, 1138–1157.
- Sibert, R. W. (1994). DRUGFIRE revolutionizing forensic firearms identification and providing the foundation for a national firearms identification network. *USA Crime Lab Digest*, 21(4), 63–68.
- Tiwari, V., & Thakur, R. S. (2015a). *Improving Knowledge Availability of Forensic Intelligence through Forensic Pattern Warehouse (FPW)*. In *Encyclopedia of Information Science and Technology* (3rd ed., pp. 1326–1335). IGI-Global.
- Tiwari, V., & Thakur, R. S. (2015b). Contextual snowflake modelling for pattern warehouse logical design. *Sadhana*, 40(1), 15–33.
- Tiwari, V., & Thakur, R. S. (2015c). P²MS: a phase-wise pattern management system for pattern warehouse. *International Journal of Data Mining, Modelling and Management*, 7(4), 331–350.
- Tiwari, V., Tiwari, V., Gupta, S., & Tiwari, R. (2010). Association rule mining: A graph based approach for mining frequent itemsets. *2010 International Conference on Networking and Information Technology*, 309–313. doi:10.1109/ICNIT.2010.5508505
- Tobias, K., Claus, V., & Marcus, L. (2011). *Automated Forensic Fingerprint Analysis: A Novel Generic Process Model and Container Format*. In *Biometrics and ID Management* (Vol. 6583, pp. 262–273). Springer.

Werrett, D.J. (1997). The national DNA database. *Forensic Science International*, 88(1), 33–42. doi:10.1016/S0379-0738(97)00081-9

KEY TERMS AND DEFINITIONS

Business Intelligence: The process, technologies, and tools needed to turn data, patterns into information, information into knowledge, and knowledge into plans that help for forensic decisions.

Data Acquisition: It is a processes used to collect information to document or analyze some phenomenon.

Data Mining: It is the practice of analyzing, examining data from different perspectives and summarizing to generate new information.

Forensic Data Analysis (FDA): It is the process of using of controlled and documented analytical and investigative techniques to identify, collect, examine, and preserve digital information.

Heterogeneous Pattern: Pattern with different structure.

Pattern Warehouse Management System (PWMS): Pattern management system used to model, store, retrieve, and manipulate patterns in an efficient and effective way.

Pattern: It is a compact and rich in semantic representation of raw data.

Pattern-Warehouse: It is a collection of persistently stored patterns.

Uncovering Limitations of E01 Self-Verifying Files

Jan Krasniewicz

Birmingham City University, UK

Sharon A. Cox

Birmingham City University, UK

INTRODUCTION

Teaching good practice in computer forensics is important to understand the correct operation and limitations of computer forensic hardware and software. One task is to demonstrate the self-verification feature of evidence file formats such as the EnCase E01 file format that contains an image of acquired data. The E01 file contains the data plus extra data in the form of hash values and Cyclic Redundancy Check (CRC) values used by computer forensic software to check the data contained within the file has not been tampered with. Students are taught how to carry out this task and verify the file by making a change to the generated file and observing mismatches between hash values and Cyclic Redundancy Check (CRC) values generated when the data was copied and when the file is loaded into computer forensic software. Whilst creating teaching materials for students to carry out this task an anomaly was identified in one of the forensic file formats, the E01 format, commonly used by practitioners. The anomaly allows changes to be made to certain bytes within the file that are not detected by computer forensic software when verified by the associated hash and CRC values. This paper describes the anomaly in the file format, discussed the implications for relying on the self-verification feature of the E01 file format and concludes on methods to make any change to the file contents detectable.

Background

One of the first tasks before conducting a computer forensic analysis of data is to make a forensically sound copy of the data stored on, for example, a hard disk drive. This task forms the acquisition stage of an investigation. By “forensically sound” it is meant that the copying process does not alter the source data resulting in an exact copy of the data (Casey, 2007). This task involves making a bit-for-bit copy of the data and using a method that assists in determining the integrity of the resulting copy as part of the chain of custody.

It is important to be able to determine that the copy of data has not been changed before it is analysed. It is common practice and recommended by organisations such as the Association of Chief Police Officers (ACPO) and National Institute of Standards and Technology (NIST) to use a mathematical function to calculate a unique value for the data at the time of copying. Examples of mathematical functions used to check the integrity of data are Cyclic Redundancy Check (CRC) and cryptographic hash (Schneier, 1996). These functions are implemented in computer programs to compute a value from a computer file or entire contents of a storage device. The value is recorded so that whenever the digital evidence is analysed the value is recomputed and compared to the original value.

Computer programs have been developed to automate the copying process and calculate the integrity values for the acquired data. These values are stored within the resulting copy of the data. Storing the integrity values within the file allows the copy to be *self-verifying* when analysed with computer forensic software. When the copy is used by a computer forensic software application, such as Guidance Software's EnCase and AccessData's FTK, the application recalculates the unique value and then compares it with the value stored in the file. The program displays a warning message when the original and calculated values are different as this difference indicates the file has changed, the change could be as a result of corruption or it could be more sinister due to a deliberate change by an individual.

This paper considers the integrity values stored in the copy of the data, commonly known as the image file or digital evidence container file (Common Digital Evidence Storage Format Working Group, 2006). The paper describes mathematical functions used to calculate the integrity values and how the property of the function allows data to be validated. The paper then describes how a practical exercise to demonstrate self-verification features to students identified an anomaly where it was possible to change a byte within the file without the self-verification detecting that the copy had been changed. The paper explains why additional integrity values should be calculated based on the entire data, copy and integrity values combined, to further enhance confidence the copy has not been altered after it has been made.

Hash Functions in Computer Forensics

Hash functions are one approach to solving the problem in computer systems of being able to perform a fast lookup of data in a data store such as RAM or disk drive (Knuth, 1998). The hash function takes as input a characteristic or characteristics of data and computes a value based on the characteristic. This value is used to locate

the data within a data store, for example. When a program needs to find data within a set of data the hash value is calculated and the value used to examine a location in the data store for the presence of that data. As a result, the hash value offers almost direct access to the data.

In comparison algorithms such as linear search are slow as it involves examining each data location for the presence of a specific data value from the first to the last. The amount of time to find data in the data store using a linear search is proportional to the number of data items in the data store and as the number of items increases so does the time to search the data store. The advantage of the hash function is it computes a hash value in constant time irrespective of the location of the data in the data store. Consider the example of searching for an individual's personal details based on some identifying data within a data store comprising of 1,000,000 individuals. A linear approach would inspect every location until a match is found where the time take to find the data is proportional to the amount of data, more data results in a longer search, potentially. A hash function would be applied to the identifying data and a value calculated that is then used to access a location directly.

Ideally hash values calculated are unique for different data but for some applications this is not entirely a requirement (Knuth 1998). Computer forensics make use of hash functions but a specific type called cryptographic hash functions that are designed to produce unique values for data. This property of calculating a unique value allows the value to be considered as a digital signature or fingerprint of the data allowing for data to be found by its hash value or verified because its value is unique (Cowen, 2013).

Cryptographic hash functions are used for the validation of arbitrary messages (the data) between individuals (Schneier, 1996). A key property of cryptographic hash algorithms is that they permit the verification of arbitrary streams of data by producing a unique value for that message exactly. In the case of cryptographic hash functions like

MD5 (Rivest, 1992) and SHA1 (RFC3174), they are designed to produce a different value when there is a small change to the message such as the changing of one bit. This property of cryptographic hash functions allows an individual to check the integrity of the data before use. The MD5 function produces a 128bit value from a data stream of any length, for example an individual file or contents of an entire storage device, where this value is unique to the data theoretically. The uniqueness of MD5 is that there is a 1 in 2^{128} or 1 in 3.4028236692093846346337460743177x10³⁸ chance of two streams of data producing the same MD5 hash value even when the difference between the two streams is one bit and the stream of bits is many millions of bits. The SHA-1 function calculates a 160 bit value where the chances of two different data producing the same value is 2^{160} or 1 in 1.4615016373309029182036848327163x10⁴⁸. However there is the issue of collisions in cryptographic hash functions where a collision occurs when two different messages produce the same hash value (Sotirov et al., 2011). Collisions present a problem as a malicious data could be constructed to generate a hash value with the same value as valid data (sKyWIper Analysis Team, 2012). Thomson (2005) discusses the implications for collisions and birthday attacks to computer forensics and concludes that whilst there are problems with MD5 collisions they can be overcome by using alternative algorithms. There have been a number of legal challenges to Encase regarding the verification process as covered in (Encase, 2014) where Encase 6 was augmented the Secure Hash Algorithm (SHA1) to provide an additional means to verify the data contained in an E01 file.

Further properties of cryptographic hash functions are that (Schneier, 1996):

- It is easy to compute the hash value for any message.
- It is difficult to compute the message from the hash value.

- It is difficult to determine another message that produces the same hash value of the message.

This latter property gives the function the property of being collision resistant where it should not be easy to compute another message that results in the same hash value. If this should happen then it would be easy to substitute one message with another without being detected. Examples of cryptographic hash functions are Message Digest (MD), Secure Hash Algorithm (SHA) and RACE Integrity Primitives Evaluation Message Digest (RIPE-MD) (Schneier, 1996). These properties of cryptographic hash functions are significant to computer forensics as they allow data to be verified after acquisition by calculating and comparing hash values.

MD5 and SHA1 functions are implemented in computer programs that are used to make the copy of the data, examples of which are AccessData's FTK Imager and EnCase. There have been some concerns with using MD5 and SHA-1 (Wang, Yin and Yu, 2005) as attacks on the algorithms have demonstrated that it is possible to generate the same value for different data. However it is acceptable and standard practice to use MD5 and SHA-1 in a legal court (Guidance Software, 2009).

As well as using cryptographic hash functions to verify the overall integrity of the data, the CRC function is used to calculate values for blocks of acquired data and stored in the digital evidence file. The CRC value's properties are that it allows data to be checked for errors where CRC is used to check packets of data communicated over a network (Halsall, 1995). A CRC value is calculated for a block of data and transmitted over the network with the data. The receiver performs the CRC check only on the received data and compares with the received CRC value. If the CRC values are the same then this means no errors have been introduced into the data as a result of transmission. If the CRC values are different then this means there are errors in the data due to transmission and

the receiver would request retransmission of the data. CRC function has similar properties to the hash function but it does not generate a unique value for data only a value that allows the data to be verified. CRC is used in digital evidence files to validate smaller blocks, for example 64 sectors worth of data (Bunting, 2012), within the data so should there be a hash mismatch a program can identify a location within the file.

It is important to note CRC and cryptographic hash functions cannot verify the data acquired are correct; they only assist in determining acquisition produced an exact copy of the original data. As a result any existing corruption in the data or successful attempts to wipe the data cannot be identified with cryptographic functions or CRC.

Concentrating on cryptographic hash functions, the hash value generated at acquisition and then the hash value generated when the data are verified are called the acquisition and verification hash values, respectively (Bunting, 2012). When the acquisition and verification hash values are the same then the data in the file are understood to be consistent with the original data copied from the device. Issues about the acquisition process are outside the scope of this paper. This paper assumes that the process of acquisition is correct and that a valid evidence file has been produced. When the values are different this means there has been some alteration of the evidence file contents. For example, it may be the case that the device the evidence file is stored on has some hardware fault corrupting the file. It could be that person or persons unknown have tried to tamper with the file by using a hex editor to modify key bytes in the file. In both cases the CRC and hash values help in determining that corruption has occurred and

where this corruption affects the acquired data, for example corruption in the file corresponds to specific sectors. This self-verification property of the evidence file format allows it to be forensically sound.

Evidence File Formats

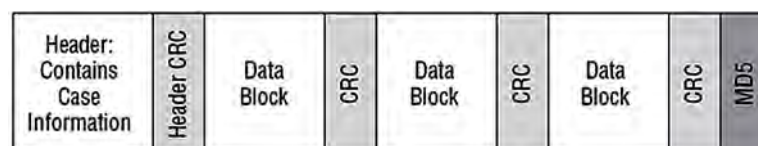
The inclusion of extra data to allow the digital evidence file to be validated automatically has resulted in a number of evidence file formats (Common Digital Evidence Storage Format Working Group, 2006). The primary objectives of evidence file formats are to be an accurate copy of the original data and allow for corruption and/or alteration of the contents to be detected. Unlike other file formats evidence file formats cannot be updated with new data; they are effectively sealed. Should an error in the creation of the evidence file occur for example, incorrectly selecting source devices and data were not copied, then, the entire process has to be repeated to create a new file.

Examples of evidence file formats are Access Data's Advanced Forensic Format (AFF), Guidance Software's E01 format and ASR Data's SMART format. These file formats all make a binary copy of data and use CRC and cryptographic hash values to validate the data. The evidence file format used for the exercise is the E01 evidence file format described below.

The E01 Evidence File Format

The E01 format is used by a number of computer forensic software such as Guidance Software's EnCase and AccessData's FTK. The logical structure of the E01 format is shown in Figure 1.

Figure 1. The logical structure of the E01 evidence file format (source: Bunting 2012)



The evidence file is generated at acquisition time by a computer program called an imager (Casey, 2009). The Header stores details of the acquisition such as a case name, evidence number, examiner name and brief notes. The Data Blocks in Figure 1 are regions containing the binary copies of the acquired data. A CRC value calculated for the Header and Data Block is stored in the file at creation. The purpose of the CRC values is for performing an integrity check on the Data Blocks and Header separately. As a software application loads the data from the file it recalculates the CRC value based on the data in the block and compares it to the stored CRC value identifying any corruption in the data block.

At the end of the file is the MD5 cryptographic hash value. This is calculated only on the acquired data and does not include the header and CRC values stored in the file. The E01 file format now includes a SHA-1 hash value (not shown in Figure 1) that is also calculated on only the acquired data and stored at the end of the file. The purpose of this additional hash is to provide a second way to check integrity using a different function to calculate hash and therefore having two values to compare against. The inclusion of the SHA-1 value helps with mitigating claims about the reliability of the MD5 algorithm to produce unique values for different data and discussed in section 2 of this paper. In combination, CRC values determine if there has been corruption in a data block and the cryptographic hash function values determine if there is corruption in the entire acquired data. This combination addresses the smaller (in terms of number of bits) CRC value where there is a greater possibility of corruption producing the correct CRC value and, hence, the corruption not being detected.

When an evidence file is loaded into computer forensic software the software begins recalculating CRC and cryptographic hashes based on the acquired data. Once the software completes recalculating the values it compares the CRC and hashes calculated with the CRC and hashes in the file. The software displays an error message

indicating a mismatch when the cryptographic hash values, acquisition and verification, do not match and an examiner can find out which data blocks are in error.

It is good computer forensic practice to check these values before conducting an analysis on the data as results based on an invalid data could be called into question. Students on a BSc Forensic Computing complete a practical exercise on verifying a digital evidence file and see the value self-validating evidence file formats by making a change to the file to produce a mismatch between acquisition and verification hashes. The exercise is described in the following section.

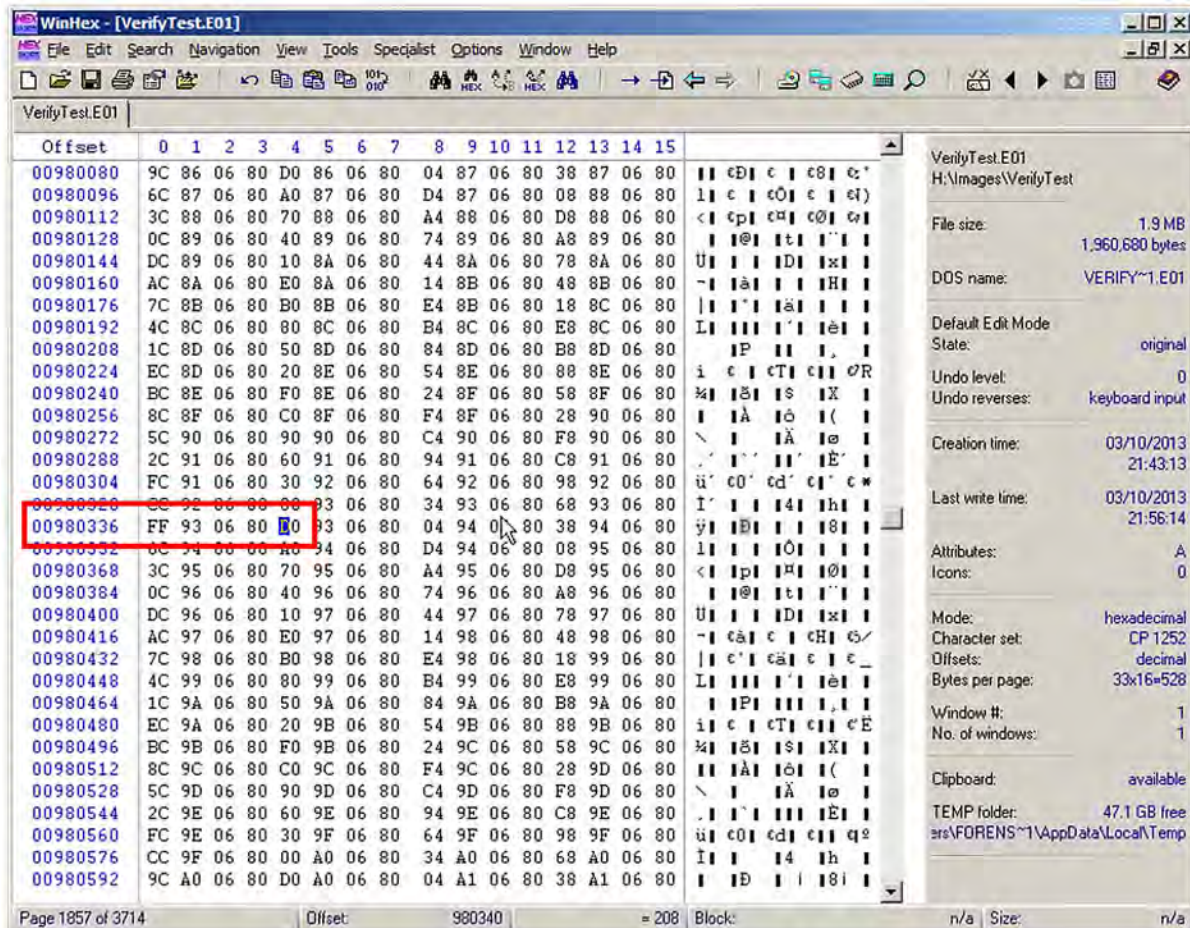
File Verification Exercise

One of the initial exercises students on a BSc Forensic Computing course do is to validate an evidence file and gain experience to determine whether a file is valid or not. The exercise comprises of two parts. Part one requires students to load the test evidence file into computer forensic software and wait for verification to complete. They then check the acquisition and verifications hash values displayed by the program are the same or what message the program should display when the verification has been completed without mismatch.

The second part involves students using a hex editor program to load and view the binary contents of the evidence file. They are instructed to change a byte at a location within the file from its current value to a different value. The file is saved and reloaded into computer forensic software for validation where students are instructed to check the acquisition and verification hash values are different as a result of the change. The hash values themselves are not of significance to the exercise but the fact they are different is significant as it indicates a problem in the integrity of the file.

The location of the byte to be changed in the second part of the exercise was determined by obtaining the logical size of the file and dividing by two to be the approximate mid-point. The total

Figure 2. Location of byte to be changed for the exercise showing the original value



size of the evidence file is 1,960,680 bytes and for simplicity the byte at 980340 was to be changed by students from its original value to 0xFF. The choice of the value 0xFF was arbitrary based on the assumption any change in the value whether it was a minor or major change should result in a verification error. A minor change would be to calculate the binary value for the byte value stored at the location and change one bit, for example the original value is D0 in the hexadecimal number system. The binary value for D0 is 11010000 where changing the right most bit that is the least significant in the number results in 11010001 or D1. A major change is to change many bits to result in a significantly different number such as reverse the binary value from 11010000 to 00101111. FF was chosen as this results in the binary pattern

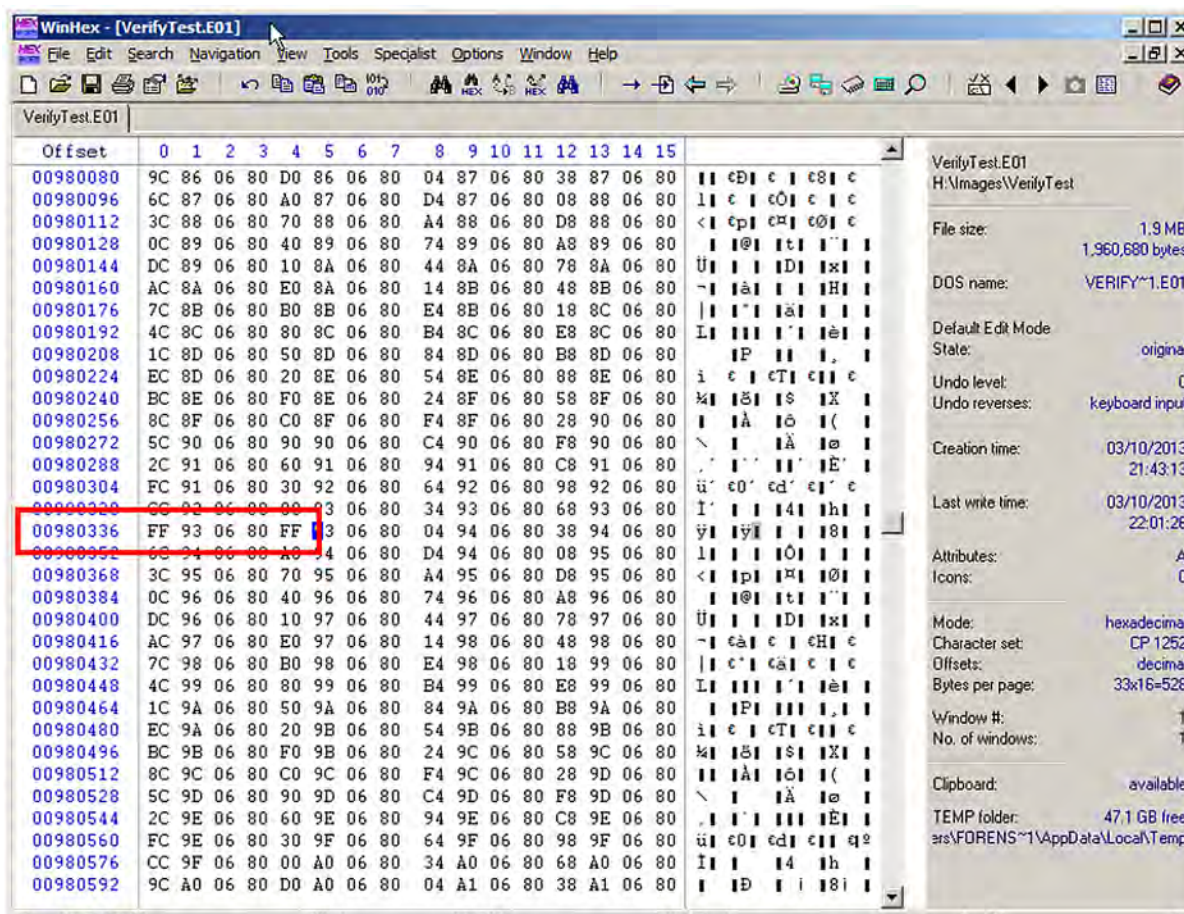
11111111 with 5 bits changed. Figure 2 shows the location and original value stored at that location.

The value at the location D0 was overwritten with the value FF as shown in Figure 3.

RESULTS

The exercise was conducted by the tutor in order to check that the exercise would produce the mismatch for the second part of the exercise. The file was saved and loaded into the EnCase computer forensic software where the expected result from the program was that there is a verification error due to the calculated hash values being different. However, the actual result was that the program completed the verification successfully with no

Figure 3. Location of byte after being changed to FF



verification errors indicated. The EnCase function to verify file integrity was invoked explicitly where the same result of no verification errors was received. Figure 4 shows the results of verifying the file's integrity.

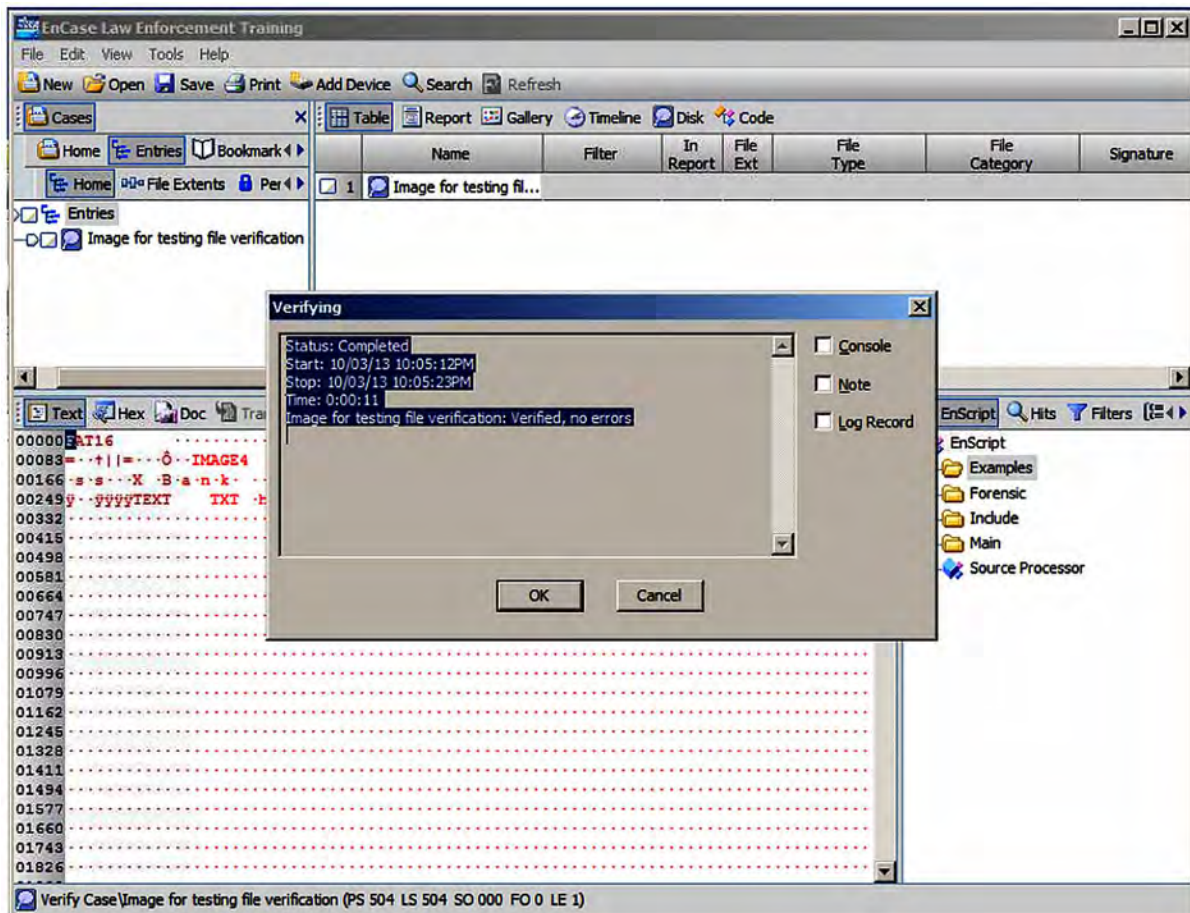
This result was not expected and the test was repeated by loading the evidence file into another computer forensic program AccessData's FTK Imager that performs a verification check using the same MD5 and SHA-1 cryptographic hash functions. The experiment was repeated to identify if the problem was specific to EnCase. On completing verification FTK Imager also displayed no difference in the computed verification hash values and they match. Figure 5 shows the results of verifying the file's integrity using the FTK Imager program. The conclusion based

on this result alone is that the values, CRC and hash, stored have not been sufficient to allow the program to detect a change in the evidence file itself. The CRC and hash values are computed solely for the acquired data not the evidence file per se. As a result changes to data outside the data blocks, CRC and hash value are susceptible to change without detection.

DISCUSSION OF RESULTS

The results in the previous section suggest that a byte can be changed in an E01 evidence file without resulting in an error being detected when the file is verified automatically by computer forensic software such as EnCase and FTK Imager. This

Figure 4. Modified file verification of modified showing no errors



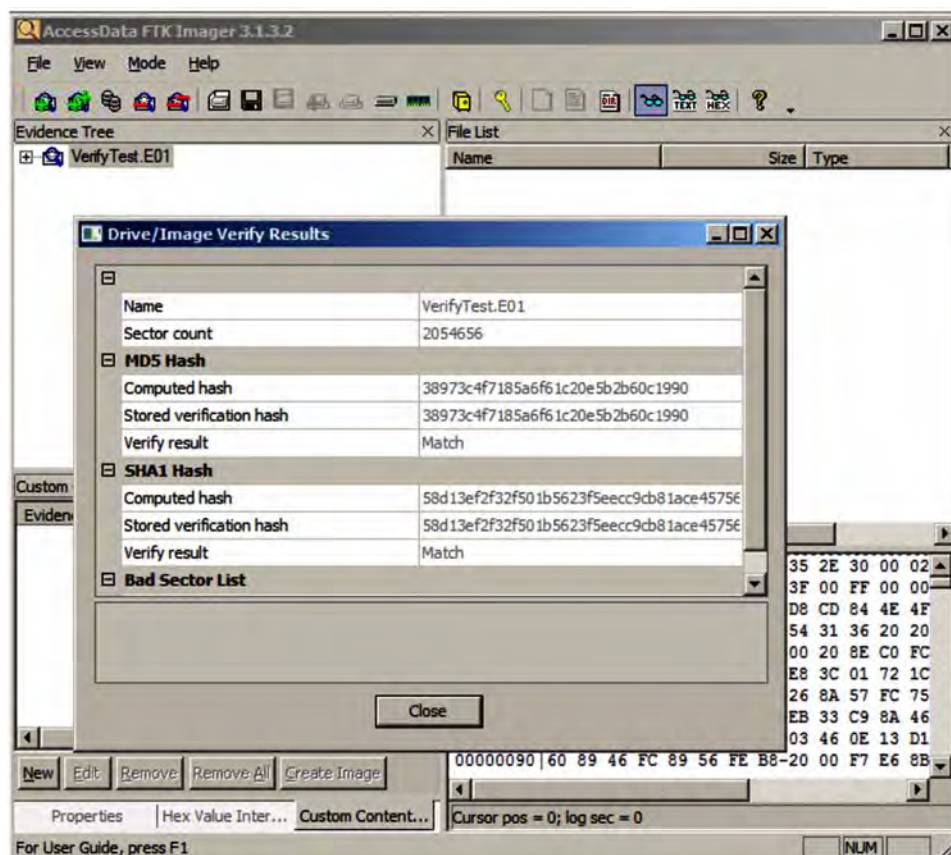
result was unexpected given the claims that E01 files are self-verifying and changes to the file are detected through the use of CRC and hash values embedded in the evidence file. As a result the self-verification hash values and CRC should only be considered as validating the acquired data and not the entire file. Based on the E01's specification a mismatch should occur when a byte has changed, as that change will either be in a data block and identified by the hash value and CRC as expected. Alternatively the change should be detected because it occurred in a CRC or hash value because these data are the only other data in the file. The change results in an invalid CRC or hash value in the E01 file and results in a verification error because the CRC and hash calculated from the data block will be different. It would appear

there is at least one location in the file that can be changed without being detected using the file's self-verification feature.

Confirming the integrity of the entire E01 file is achievable by calculating the hash value for the entire image file after it has been created. This would be consistent with computer forensic practices of using hashes to confirm data integrity (Kumar et al. 2012).

This result calls into question relying purely on the self-verification of the E01 evidence file format to verify the data. A similar problem occurred when a legal challenge to the E01 file format was made due to the use of MD5 hash function used to verify the data. The challenge was made based on academic research that demonstrated it was possible to generate the same hash value for

Figure 5. Verification results produced by the FTK Imager program on the modified file.



two different data (Boer and Bosselaers, 1994). The challenge was successfully countered and where an additional hash value was used in the file to further improve data integrity (Guidance Software, 2009).

Likewise the results presented in this paper show a problem in self-verification property of the E01 file format for the whole file but significance of the result to the E01 file format may not be major. Further work is needed to confirm this assertion that the change made to a file was at a location that was not in a data block, CRC or hash value.

A solution to relying on the internal hash and CRC values is to use hash function such as the MD5 and SHA-1 to calculate the hash value for the E01 file as a whole, header, data blocks, CRC values and hash values. The E01 file hash value would be calculated immediately after acquisition was

completed and a record made of the hash value. The hash value is recalculated whenever the E01 file is to be used in an analysis and the calculated hash value compared with the original hash. The result of the comparison would verify the file's integrity or indicate there has been some change to the file after acquisition. This check could be performed manually by using a program such as the one above to calculate hash values. Alternatively the E01 file can be augmented to include extra hash values for validating the file.

CONCLUSION

The primary conclusion of this paper is that is possible to make a change to a E01 evidence without that change being detected though the self-verification feature of the file format. This

result contradicts the accepted fact that E01 files are self-verifying and can detect the changes to the file. A more correct interpretation is the verification feature of E01 files detects changes that have occurred only within data blocks acquired. As a result confidence in the E01 file format to be forensically sound with regards to verifying the data in contains is maintained. However the same is not true for verifying the E01 file in its entirety.

The recommendation of this paper is that to verify the integrity of the entire E01 file an additional hash value using an accepted hash function such as MD5 or SHA-1 or both is calculated and recorded as part of good practice. This hash value is calculated at the same time the image file is created. Whenever the E01 file is to be used the hash value is recalculated on the entire image file and compared with the original hash. When the values are the same it can be taken that the E01 file has integrity and can be used. When the values are different then there is some doubt over the integrity of the file irrespective of the CRC and hash values embedded in the E01 file.

FUTURE RESEARCH DIRECTIONS

Further work is needed to identify whether the anomaly identified in an E01 file is specific to the test file or it can be consistently produced for other E01 files of differing sizes (in bytes) and for split E01 files. This further work will also involve changing bytes at other locations in an E01 file in order to ascertain whether the acquired data has been changed but the change not detected during self-verification as described previously. In addition other evidence file formats will be tested as well as using alternative programs such as hashmyfiles.

The secondary conclusion is the importance of testing computer forensic hardware and software and associated artefacts in computer forensic teaching. Testing helps students understand using

hardware and software in order to build confidence in what the hardware and software does and its correct operation. It also helps with emphasising the need for a rigorous testing in order to demonstrate procedures are correct to ensure evidence is admissible.

REFERENCES

- Bunting, S. (2012). *EnCase Computer Forensics -- The Official EnCE: EnCase Certified Examiner Study Guide*. Indianapolis, IN: John Wiley.
- Casey, E. (2007). What does forensically sound really mean? *Digital Investigation*, 4(2), 49–50. doi:10.1016/j.diin.2007.05.001
- Casey, E. (2010). *Handbook of Digital Forensics and Investigation*. Burlington, MA: Elsevier Academic Press.
- Common Digital Evidence Storage Format Working Group. (2006). *Survey of Disk Image Storage Formats* [PDF document]. Retrieved October 25, 2013 from <http://www.dfrws.org/CDESf/survey-dfrws-cdesf-diskimg-01.pdf>
- Cowen, D. (2013). *Computer Forensics InfoSec Pro Guide*. New York: McGraw Hill.
- Den Boer, B. (1993, May). A. *Collisions for the compression function of MD5*. Paper presented at Advances in Cryptology — EUROCRYPT '93 Workshop on the Theory and Application of Cryptographic Techniques, Lofthus, Norway.
- Eastlake, D. (2001). *US Secure Hash Algorithm 1 (SHA1)*. Retrieved October 25, 2013 from <https://www.ietf.org/rfc/rfc3174.txt>
- Encase Legal Journal. (2014). *Encase Legal Journal*. Retrieved September 22, 2016 from <https://www.guidancesoftware.com/docs/default-source/document-library/publication/encase-legal-journal---5th-edition.pdf?sfvrsn=12>

Guidance Software. (2009). *EnCase Legal Journal*. Retrieved October 29, 2010 (Requires registration) from <https://www.guidancesoftware.com/resources/Pages/doclib/Document-Library/EnCase-Legal-Journal.aspx>

Halsall, F. (1992). *Data Communications, Computer Networks and Open Systems* (3rd ed.). Wokingham, UK: Addison-Wesley.

Knuth, D. (1998). *The Art of Computer Programming: Vol. 3. Sorting and Searching* (2nd ed.). London: Addison-Wesley.

Kumar, K., Sofat, S., Jain, S. K., & Aggarwal, N. (2012). Significance of Hash Value Generation in Digital Forensic: A Case Study. *International Journal of Engineering Research and Development*, 2(5), 64–70.

Rivest, R. (1992). *The MD5 Message-Digest Algorithm*. Retrieved October 25, 2013 from <https://www.ietf.org/rfc/rfc1321.txt>

Schneier, B. (1996). *Applied Cryptography: Protocols, Algorithms and Source Code in C* (2nd ed.). Chichester, UK: John Wiley.

sKyWIper Analysis Team. (2012). *sKyWIper (a.k.a. Flame a.k.a. Flamer): A complex malware for targeted attacks*. Retrieved September 22, 2016 from <http://www.crysys.hu/skywiper/skywiper.pdf>

Sotirov, A., Stevens, M., Appelbaum, J., Lenstra, A., Molnar, D., Osvik, D. A., & de Weger, B. (2011). *MD5 considered harmful today*. Retrieved September 22, 2016 from <http://www.win.tue.nl/hashclash/rogue-ca/>

Wang, X., Yin, Y. L., & Yu, H. (2005, August). *Finding Collisions in the Full SHA-1*. In *Advances in Cryptology – CRYPTO 2005*, Paper presented at the 25th Annual International Cryptology Conference, Santa Barbara, CA.

ADDITIONAL READING

Boddington, R. (2016). *Practical Digital Forensics*. Birmingham, UK: Packt Publishing.

Encase image file format (n. d.). Retrieved from http://www.forensicswiki.org/wiki/Encase_image_file_format

Hayes, D. R. (2014). *A Practical Guide to Computer Forensics Investigations* (1st ed.). Indianapolis, IN: Pearson IT Certification.

Robinson, M. K. (2015). *Digital Forensics Workbook: Hands-on Activities in Digital Forensics*. Scotts Valley, CA: On-Demand Publishing LLC.

Sammes, A., & Jenkinson, B. (2010). *Forensic Computing: A Practitioner's Guide* (2nd ed.). London, UK: Springer.

Sammons, J. (2014). *The Primer for Getting Started in Digital Forensics* (2nd ed.). Burlington, MA: Syngress.

Sammons, J. (2014). *The Basics of Digital Forensics* (2nd ed.). Burlington, MA: Syngress.

KEY TERMS AND DEFINITIONS

Acquisition: The process by which data is acquired from a storage device and stored in an image file.

Bit-for-Bit Copy: A bit level copy of an arbitrary stream of data.

Cyclic Redundancy Check: An algorithm that computes a value for a data stream to use used for error detection and possible correction.

E01: A type of image file format that includes a bit-for-bit copy of source data plus hash values calculated from the original data.

Image File: A file containing acquired from a storage device.

MD5: A type of hashing algorithm that computes a fixed size value for an arbitrary data stream.

Verification: The process by which an image file is verified before use. Involves comparing hash values computed when the image file was made and when it is loaded into computer forensic software. Any discrepancy between the hash values suggests tampering or corrupting of the data within the image file.

Category C

Crisis Response and Management

Crisis Response and Management

Sergey V. Zykov

National Research University Higher School of Economics, Russia

INTRODUCTION

In the 1960s, the so-called “software crisis” triggered the advent of software engineering as a discipline. This term originated from the critical development complexity, which happened due to the rapid growth of computational power. At that time, the computing power of the machines became so overwhelming that a number of software development projects were over budget, late or unsuccessful. One well-known example was the first General Electric’s payroll system launched in 1954 at Louisville, Kentucky; this was late, over budget, and missing crucial features (Topi, & Tucker, 2014). Irrespective of human efforts, the complexity of the hardware and software systems was hard to cope with by means of the old methods and techniques. The challenge was so dramatic that in 1967 NATO arranged an invitation-only conference, where world leaders in IT research and practice searched for an efficient response. At the conference, the term “software crisis” was coined by F.Bauer and used by E.Dijkstra (Naur, & Randell, 1968).

Another term suggested at the conference by the same F. Bauer was software engineering. The idea was to apply the engineering methods of material production to the new domain of large-scale concurrent software systems in order to make the software projects more accurate and predictable. This software engineering approach was feasible, though the methods and practices used had to differ substantially from those used in the material production. Specifically, the experts examined bridges as the instances of complex material systems.

The attendees concluded that the distribution of time and cost by the lifecycle phases, especially for

the post-delivery maintenance was very different for software and material production. This is why the new software engineering discipline was in need of new methodologies, techniques and tools.

The focus of the software engineering discipline was the “serial” production of substantially large-scale, complex and high quality software systems. Concerning software complexity, at least two dimensions were identified; these were technical and management complexity (Booch, 2006). To measure software product complexity and quality, a set of attributes and metrics was suggested. The quality attributes included performance, reliability, security, fault tolerance, usability, strategic reusability and maintainability; their importance depended on the product size and scope (Lattanze, 2008). The complexity metrics included product size in terms of lines of code, function points, nesting levels, cyclomatic complexity and a number of more sophisticated ones (Debbarma, Debbarma, Chakma, & Jamatia, 2013). These metrics assisted in the divide-and-conquer strategy; later, they this general approach transformed into elaborate product estimation techniques and software development methodologies (Jensen, 2014).

Researchers argue whether the crisis in software engineering is over yet (Colburn, Hsieh, Kehrt, & Kimball, 2008) or it still exists (Buettner, Dai, Pongnumkul, & Prasad, 2015). This happens because of the fundamental differences in the lifecycles of software and material products. One critical difference between large-scale software and material production is the distribution of time and cost by the development lifecycle phases. Therewith, maintenance is the most time and cost consuming, it often exceeds 60% of the software project expenses (Schach, 2011). The other crucial difference is that

DOI: 10.4018/978-1-5225-2255-3.ch120

software production often depends dramatically upon human factors. These human factors relate to the management aspects of software complexity, whereas the technology factors relate to the technological aspects. Certain product categories are far more complex in terms of management than in terms of technology; however, the influence of the human factors on their development is largely underestimated. For such software product categories as enterprise information systems and defense management information systems, neglecting these human factors often results in project delays or even failures (Booch, 2006).

Therewith, the software crisis originates from a number of factors; these are human-related and technology-related factors. To manage this crisis, the authors suggest a set of software engineering methods, which systematically optimize the lifecycles for both types of these influencing factors. This lifecycle optimization strategy includes crisis-responsive methodologies, system-level architectural patterns, informing process frameworks, and a set of knowledge transfer principles (Zykov, 2009; Zykov, Shapkin, Kazantsev, & Roslovtshev, 2015; Zykov, 2015).

Software development usually involves customers, developers and their management; each of these parties has different preferences and expectations. Therewith, these parties often differ in their vision of the resulting product; typically, the customers focus on business value while the developers are concerned with technological aspects. Such a difference in focus often results in crises. Thus, the software crises often has a human factor-related root cause. To deal with these kind of crises, software engineers should enhance their skillset with managerial skills, such as teamwork, communications, negotiations, and risk management.

BACKGROUND

In the 1960s, the software production lifecycle was unstable as no systematic approach existed.

At that time, software development did not allow for adequate planning and management of a project timeline and budget. The software products were unique masterpieces; they used a build-and-fix approach as the core “methodology”. A systematic approach to product lifecycle was required in order to manage this crisis of development. This approach was to include certain technical and management aspects. The technical aspects should include justified architecture selection and high-level design. The management aspects should include teamwork and transparent communication between the client and the developer.

In this period, software development involved a number of parties with significant differences in goals and expectations. These were clients, developers and their management. Each side usually had a different understanding of the future product, as the clients were business oriented, while the developers focused on technology.

At present, this same lack of common understanding hampers software development; it is a possible source of a software crisis resulting from management complexity. To deal with this kind of crisis, software engineers enhance their technical abilities by a specific “soft” skillset. This includes collaborative teamwork, risk management, communications and negotiations. The “soft” skills assist software engineers in their management of the present-day software development crisis, which often results from human factors.

The following decades of the 1970s and 1980s changed software development from an art to a science, though it had not yet become a serial production technology. Techniques and methods appeared which are currently known as the Programmer’s Workbench (Dolotta, & Mashey, 1976), third-generation programming languages and the supporting Structured Analysis and Design Technique (Dahl, Dijkstra, & Hoare, 1972). Further, in the 1990s software development technologies became even more advanced. The new technologies which arose at that time were more process focused; a few examples of these include OOAD (Jacobson, Christerson,

Jonsson, & Overgaard, 1992) and the capability maturity models of CMM and CMM(I) (Paulk, Curtis, Chrissis, & Weber, 1993; Linden, 1994). Major software R&D centers appeared, such as the Software Engineering Institute (<http://www.sei.cmu.edu>). A number of mission-critical systems (i.e. essential to the organization survival) appeared. Relative value of software increased. Software estimating tools and methods appeared, which made software development more crisis-responsive and better manageable (Jensen, 2014). A few examples of these methods are: COCOMO and COCOMO II (Boehm, 1981; Boehm, Egyed, Clark, Horowitz, Westland, Madachy, & Selby, 1995), Price-S (Park, 1990), Sage (Jensen, 2007), and Seer (Jensen, 1980). However, the crisis in software engineering lasted longer and was deeper than that in material manufacturing. One possible reason for this is that human-related factors are much more complex in their nature and harder to manage (Jensen, 2014).

Absence of a universal software development methodology indicates that certain consequences of the crisis persist. To manage the crisis, life-cycle optimization is required. This optimization should systematically address crisis technological and management aspects by means of software engineering models, methods, tools, techniques and practices.

Any software architecture has multiple perspectives. These include process, data and system interfaces (Lattanze, 2008). Any single-perspective software architecting, or even a two-“dimensional” one, usually results in insufficient crisis agility. The present-day crisis manifests itself in high turbulence and fluctuating customer demand. Inefficient clients with excessive overheads often end up with dramatic profit losses. In crisis, the software architecture often requires extending beyond the traditional software engineering of “heavy” methodologies, such as Rational Unified Process (Kruchten, & Kroll, 2003) or Microsoft Solution Framework (Turner, 2006). Human factor-responsive software development methodologies, such as Agile (Agile Manifesto, 2001),

Scrum (Takeuchi, & Nonaka, 1986) or Extreme Programming (Beck, & Andres, 2004) can often better assist in crisis management.

TECHNOLOGY-BASED CRISIS RESPONSIVE SOFTWARE DEVELOPMENT

Issues related to software and hardware system development tend to become more essential in a crisis. A positive solution is to create uniform architectural patterns for such complex systems (Zykov, 2010).

Crisis-responsive software engineering uses general principles of architecture, such as components and connectors (Lattanze, 2008). Top-level architectural design is critical; it determines key concepts and interfaces for the subsystems of software products to be developed/optimized. At this level, high-level patterns are applicable, which describe the interfaces for the software system components. E. Gamma laid the foundations of the software design patterns (Gamma, Helm, Johnson, & Vlissides, 1998), which are applicable to large-scale software systems (Fowler, 2002; Freeman, Bates, Sierra, & Robson, 2004; Zykov, 2009).

However, each instance of the patterns is dependent upon both the problem domain and the nature of the client. That is why the authors suggest adding human factor-based patterns to the technological and architectural ones (Zykov, 2009). In crisis, both components may critically influence software development. Therefore, human factor issues often manifest themselves in undisciplined development and communication problems.

In a crisis, the technological factors that influence software development have three architectural perspectives: business process perspective, data flow perspective, and system interface perspective. Each of these three perspectives should be projected to the business management levels, which include everyday management, short-term planning and strategic decision-making. The

combination of these architectural perspectives and the business management levels results in the crisis agility matrix. As such, crisis agility should be understood, interpreted and measured in terms of mission-critical crisis quality attributes, such as portability, maintainability and adaptability. To design a crisis-agile software system, architects should plan for its decomposability to smaller subsystems and future reengineering (Zykov, Shapkin, Kazantsev, & Roslovtsev, 2015).

The crisis agility matrix allows the identification of technology and human factor-related mission critical dependencies for the software systemic properties. These dependencies suggest a set of constraints which guard the system development process from potential analysis and design errors. This set of constraints warns of potential architectural problems, challenges or risks represented in terms of contradictions between the business processes, data flows and system interfaces. The matrix allows for crisis responsiveness estimation in terms of process management, data integrity and interface quality (Zykov, Shapkin, Kazantsev, & Roslovtsev, 2015).

An enhancement for the agility matrix is a pattern-based crisis-responsive methodology of software development. This methodology extracts common enterprise software patterns of module level and applies them to a series of heterogeneous implementations. The approach extends the spiral lifecycle (Boehm, 1986) by formal models for data representation and management, and by development tools based on domain specific languages (Hohpe, & Woolf, 2004). Implementations include oil and gas resource planning, airline dispatching and nuclear power plant construction support (Zykov, 2010).

CRISIS-RESPONSIVE ARCHITECTURE FOR MISSION- CRITICAL SYSTEMS

Mission-critical software systems usually include subsystems of different levels. The higher levels

assist in decision-making and long-term planning; the lower ones support data acquisition hardware. Different levels serve the information requests of different groups of users, such as top and middle management, analysts and designers. Thus, these software levels and their interfaces are human factor dependent.

It is possible to extend the pattern-based approach to software systems development (Gamma, Helm, Johnson, & Vlissides, 1998; Fowler, 2002) for software system engineering; the focus is the pattern of high-level system architecture (Zykov, Shapkin, Kazantsev, & Roslovtsev, 2015).

A crisis-responsive system should have a scalable architecture so that the system supports easy decomposition and reengineering. Typically, a mission-critical system architecture has three perspectives. These three perspectives focus on processes, data and hardware (Lattanze, 2008). An architectural design, which does not support all three perspectives, is usually insufficient for crisis agility. The system lifecycle should consider the dependencies between these three perspectives; for each individual system, it is essential to address these dependencies at the early stages of analysis and design.

The crisis agility matrix includes the three perspectives for processes, data and system interfaces (Zykov, Shapkin, Kazantsev, & Roslovtsev, 2015). The first perspective represents dynamics; it assists in decomposing strategic goals into business processes. The second perspective contains the decomposed data objects for these business processes. The third perspective includes the hardware and software systems, which operate these data.

The system architecture framework incorporates the three perspectives (Zykov, Shapkin, Kazantsev, & Roslovtsev, 2015); it is based on high-level patterns (Fowler, 2002; Freeman, Bates, Sierra, & Robson, 2004; Hohpe, & Woolf, 2004). Each level communicates directly with the adjacent ones; it is the consumer for the lower level and the provider for the higher level.

The top, strategic level represents decision-making and business intelligence. This level analyses and aggregates data from the software for planning enterprise resources; it provides instant online access through a dashboard-like interface.

The interactive level represents the interaction of the developers with their clients and partners. This level propagates the enterprise knowledge to the previous level for strategic analysis and decision-making. An instance of this level is a customer relations management (CRM) system.

The next level represents the resource planning; it consolidates the data from the lower level(s) to form a forecast dynamics of the key production indexes, such as revenues, profits, overheads and expenditures. An example of this level is a resource planning system.

The accounting level contains relatively lower-level subsystems for accounting, warehousing, inventory management and the like.

The supervisory level contains the interfaces between software and hardware components. This level embraces the SCADA systems for field-based devices and sensors, which perform assembly line production and similar plant level operations.

The data level represents databases and data warehouses. This level includes DBMS with data mining plug-ins, middleware, and lower level content management tools.

The hardware level is the lowest; it includes hardware devices and human-machine interfaces. The hardware level aggregates analog data, digitizes and processes them for the upper levels (Zykov, 2015).

HUMAN FACTORS: BUILDING A CRISIS-RESPONSIVE COMMUNICATION

In a crisis, the so-called “human factor errors”, which result from critical uncertainties and anarchic, undisciplined lifecycle management, often dramatically influence the product quality and the project outcomes.

Communication as a “soft” skill and as a process is mission-critical for software development; miscommunications often result in crises. The crisis agility suggests a process framework based on informing science (Cohen, 1999) and Shannon’s communication model (Shannon, & Weaver, 1949). The basic framework includes a transmitting side (i.e. the client), a receiving side (i.e. the developer), and the environment (i.e. the set of software systems at their sites). The model for the basic process framework is an oscillating circuit with a capacitor and an inductive coupling, which uses a feedback loop to oscillate.

A crisis requires a more complex model. First, the communicating sides are informing systems by themselves (Gill, & Bhattacharjee, 2007). Second, the informing circuit requires a feedback controller to model resonant communication. Third, to model bidirectional feedback control the informing circuit requires positive and negative feedback loops.

Typically, the positive feedback increases the amplifier gain, whereas the negative one reduces it. Uncontrolled positive feedback may cause unwanted spontaneous oscillations; this models a situation when the resonance no longer promotes communication. Uncontrolled negative feedback decreases the gain and models the situation when the output value is too low for communication to happen. The enhanced informing circuit provides controlled resonance, which models reliable communication in the oscillating environment with crisis uncertainties. For controlled resonance, the model requires a dedicated couple of controlled positive and negative feedback loops for each of the communicating sides (Zykov, 2015).

To maintain controlled oscillation, the model requires positive and negative feedback compensators, which turn on when the signal input or output value is out of the preset range. These compensators assist in avoiding spontaneous resonance, which models crisis-related communication problems. The circuit requires thorough compensator testing to ensure that both kinds of feedback are within prescribed ranges. The posi-

tive feedback assists in a resonant communication; it also helps to compensate for the noise, which is a likely source of a miscommunication and, consequently, a crisis. For controlled communication, the enhanced circuit uses a calibrated bidirectional feedback circuit, which keeps the informing signal within the prescribed operating range (Zykov, 2015).

CRISIS-RESPONSIVE KNOWLEDGE TRANSFER PRACTICES

In a crisis, it appears that a set of key practices essentially influences knowledge transfer between the client and the developer. Some of these knowledge transfer practices are also known as the “seven principles”; they assist the “informed” communications between the client and the developer (Ambrose, Bridges, DiPietro, Lovett, & Norman, 2010) and significantly enhance the informing science framework (Cohen, 1999).

To promote knowledge transfer, the following practices are mission-critical: prior knowledge, knowledge organization, mastery, feedback, course climate and motivation. Concerning prior knowledge, the two communicating sides usually relate what they learn to what they have known previously; they interpret new knowledge in terms of their prior knowledge (National Research Council, 2000). The client’s side tends to relate prior knowledge to that which is new; however, in crisis this prior knowledge often is inappropriate or incomplete (Resnick, 1983; Alvermann, Smith, & Readence, 1985). Analogical reasoning tends to bridge the gap between the two kinds of knowledge; it helps focusing on deeper concept relationships rather than misleading surface similarities (Gentner, Loewenstein, & Thompson, 2003; Zykov, 2015). The clients with clearer goals often use better strategies for crisis-responsive communication (Barron, & Harackiewicz, 2001).

Mastery as a crucial human-related factor includes competence and consciousness (Sprague, & Stuart, 2000). In crisis software management,

junior developers and architects often underestimate application of their skills to multiple contexts; adaptive mentoring tends to build up their mastery (Cognition and Technology Group at Vanderbilt, 1994; Zykov, 2015).

The synergy of feedback and practice yields to well-balanced theoretical knowledge and a practical skillset. For resonant knowledge transfer, practice and feedback should focus on the same project objectives (Ambrose, Bridges, DiPietro, Lovett, & Norman, 2010). In crisis, the feedback should be prompt and frequent; even a short feedback, if immediate, frequent and individually mentored, promotes knowledge transfer (Hattie, & Timperley, 2007; Anderson, Corbett, Koedinger, & Pelletier, 1995).

Personal and team process improvement requires an efficient metacognitive cycle with iterative evaluation and adjustment (Pascarella, & Terenzini, 2005). In crisis, metacognitive skills are mission-critical; recommended software engineering techniques include self-reflection, team-based concept mapping and brainstorming (Novak, 1998; Zimmerman, 2001).

Other crisis-responsive practices are clear-cut ground rules, transparent policies, cognitive load minimization goal-directed feedback, and an open and friendly environment (Ambrose, Bridges, DiPietro, Lovett, & Norman, 2010; Gill, & Bhat-tacherjee, 2007; Zykov, 2015).

FUTURE RESEARCH DIRECTIONS

The authors found that the crisis in software development was driven by technological and human-related factors. The technological factors were largely driven by software system technical complexity; they can be managed by divide-and-conquer strategy. In a crisis, however, human factors often become mission-critical; this group of factors relates to software management complexity. To manage these human-related factors, agile development methods are usually efficient; these methods are typically team focused, flexible and

adaptive. The authors identified such agile methods and practices; however, their exact influence on particular human factors are intricate, hard to detect and to describe in detail. These intricate relationships between agile methodologies and the mission-critical human-related factors are a prospective research area. The authors also identified and prioritized a set of principles for successful information transfer between software product clients and developers. However, a number of dependencies between these principles is still a matter of future research.

CONCLUSION

The authors analyzed the nature and root causes of the software development crisis which started in the 1960s and, as some researchers argue, is still present. The crisis resulted from a number of factors, which included technological ones and human-related ones. The first group of factors originates from technical complexity, while the second one results from management complexity of the software products.

Software production involves multiple sides with clearly different expectations and goals. In a crisis, efficient managing of these human-related factors becomes mission-critical. Communicational misconceptions between the sides often result in a software product which may not meet the clients' expectations. To manage this software delivery crisis, the authors recommend an optimization strategy for the technical and human factors by means of software engineering models, methods, techniques, practices and tools. This strategy includes risk management models, agile methodologies, system-level architectural patterns, an informing process framework, and a set of knowledge transfer principles. This strategy, if applied systemically, should essentially help to optimize the software development process and to manage the crises of software production.

REFERENCES

- Agile Manifesto. (2001). Retrieved March 14, 2016, from <http://www.agilemanifesto.org/>
- Ambrose, S. A., Bridges, M. W., DiPietro, M., Lovett, M. C., & Norman, M. K. (2010). *How Learning Works. Seven Research-Based Principles for Smart Teaching*. John Wiley & Sons.
- Barron, K., & Harackiewicz, J. (2001). Achievement goals and optimal motivation: Testing multiple goal models. *Journal of Personality and Social Psychology*, 80(5), 706–722. doi:10.1037/0022-3514.80.5.706 PMID:11374744
- Beck, K., & Andres, C. (2004). *Extreme Programming Explained: Embrace Change* (2nd ed.). Addison-Wesley.
- Boehm, B. W. (1986). A Spiral Model of Software Development and Enhancement. *ACM SIGSOFT Software Engineering Notes*, 11(4). doi:10.1145/12944.12948
- Boehm, B. W., Egyed, A., Clark, B., Horowitz, E., Westland, C., Madachy, R., & Selby, R. (1995). Cost Models for Future Software Life Cycle Processes: COCOMO 2.0. *Annals of Software Engineering*, 1(1), 57–94. doi:10.1007/BF02249046
- Booch, G. (2006). *Software Architecture*. Retrieved March 14, 2016, from http://paginas.fe.up.pt/~aaguiar/as/AS-200506-Software_Architecture_by_Booch.pdf
- Buettner, M., Dai, P., Pongnumkul, S., & Prasad, R. (2015). *Software Engineering is indeed in a crisis*. Retrieved March 14, 2016, from https://www.researchgate.net/publication/265070905_Software_Engineering_is_indeed_in_a_crisis
- Cognition and Technology Group at Vanderbilt. (1994). From visual word problems to learning communities: Changing conceptions of cognitive research. In K. McGilly (Ed.), *Classroom lessons: Integrating cognitive theory and classroom practice* (pp. 157–200). Cambridge, MA: MIT Press/Bradford Books.

- Cohen, E. B. (1999). Reconceptualizing Information Systems as a Field of the Transdiscipline Informing Science: From Ugly Duckling to Swan. *Journal of Computing and Information Technology*, 7(3), 213–219.
- Colburn, A., Hsieh, J., Kehrt, M., & Kimball, A. (2008). *There is no software engineering crisis*. Retrieved March 14, 2016, from cs.washington.edu/courses/cse503/08wi/crisis-con.pdf
- Dahl, O. J., Dijkstra, E. W., & Hoare, C. A. R. (1972). *Structured Programming*. London: Academic Press.
- Debbarma, M. K., Debbarma, S. N., Chakma, K., & Jamatia, A. (2013). A Review and Analysis of Software Complexity Metrics in Structural Testing. *International Journal of Computer and Communication Engineering*, 2(2), 129–133. doi:10.7763/IJCE.2013.V2.154
- Dolotta, T. A., & Mashey, J. R. (1976). An introduction to the Programmer's Workbench. *Proceedings of the 2nd international conference on Software engineering*, 164–168.
- Fowler, M. (2002). *Patterns of Enterprise Application Architecture*. Addison-Wesley.
- Gamma, E., Helm, R., Johnson, R., & Vlissides, J. (1998). *Design Patterns CD: Elements of Reusable Object-Oriented Software*. Addison-Wesley.
- Gill, G., & Bhattacharjee, A. (2007). The Informing Sciences at a crossroads: The role of the Client. *Informing Science Journal*, 10, 17–39.
- Hohpe, G., & Woolf, B. (2004). *Enterprise Integration Patterns: Designing, Building, and Deploying Messaging Solutions*. Addison-Wesley.
- Jacobson, I., Christerson, M., Jonsson, P., & Overgaard, G. (1992). *Object-oriented software engineering: a use case driven approach*. Addison-Wesley.
- Jensen, R. W. (2014). *Improving Software Development Productivity: Effective Leadership and Quantitative Methods in Software Management*. NJ: Prentice Hall.
- Kruchten, P., & Kroll, P. (2003). *The Rational Unified Process Made Easy: A Practitioner's Guide to the RUP*. Addison-Wesley.
- Lattanze, A. (2008). *Architecting Software Intensive Systems: A Practitioner's Guide*. Auerbach. doi:10.1201/9781420045703
- National Research Council. (2000). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.
- Naur, P., & Randell, B. (1968). *Software Engineering: Report on a Conference sponsored by the NATO Science Committee*. Garmish.
- Park, R. E. (1990). The Central Equations of the Price Software Cost Model. Moorestown, NJ: G.E. Price Systems.
- Paulk, M. C., Curtis, B., Chrissis, M. B., & Weber, C. V. (1993). Capability maturity model, version 1.1. *IEEE Software*, 10(4), 18–27. doi:10.1109/52.219617
- Resnick, L. B. (1983). Mathematics and science learning: A New Conception. *Science*, 220(4596), 477–478. doi:10.1126/science.220.4596.477 PMID:17816206
- Schach, S. R. (2011). *Object-Oriented and Classical Software Engineering* (8th ed.). McGraw-Hill.
- Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana, IL: University of Illinois Press.
- Takeuchi, H., & Nonaka, I. (1986). The new product development game. *Harvard Business Review*, 64(1), 137–146.

Topi, H., & Tucker, A. (2014). *Computing Handbook: Information Systems and Information Technology* (3rd ed.). CRC Press.

Turner, M. S. V. (2006). *Microsoft Solutions Framework Essentials: Building Successful Technology Solutions*. Microsoft Press.

Zykov, S. (2009). Designing patterns to support heterogeneous enterprise systems lifecycle. In *Proceedings of the 5th Central and Eastern European Software Engineering Conference in Russia* (pp. 83-88). IEEE. doi:10.1109/CEE-SECR.2009.5501184

Zykov, S. (2015). Enterprise Applications as Anthropic-Oriented Systems: Patterns and Instances. In *Proceedings of the 9th KES Conference on Agent and Multi-Agent Systems: Technologies and Applications* (vol. 38, pp. 275-283). Springer. doi:10.1007/978-3-319-19728-9_23

Zykov, S. (2015). Human-Related Factors in Knowledge Transfer: A Case Study. In *Proceedings of 9th KES Conference on Agent and Multi-Agent Systems: Technologies and Applications* (vol. 38, pp. 263-274). Springer. doi:10.1007/978-3-319-19728-9_22

Zykov, S., Shapkin, P., Kazantsev, N., & Roslovtshev, V. (2015). Agile Enterprise Process and Data Engineering via Type-Theory Methods. *Proceedings of the 5th International Symposium ISKO-Maghreb*.

ADDITIONAL READING

Alvermann, D. E., Smith, L. C., & Readence, J. E. (1985). Prior knowledge activation and the comprehension of compatible and incompatible text. *Reading Research Quarterly*, 20(4), 420–436. doi:10.2307/747852

Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *Journal of the Learning Sciences*, 4(2), 167–207. doi:10.1207/s15327809jls0402_2

Boehm, B. W. (1981). *Software Engineering Economics*. Englewood Cliffs, NJ: Prentice Hall.

Freeman, E., Bates, B., Sierra, K., & Robson, E. (2004). *Head First Design Patterns*. O'Reilly.

Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. *Journal of Educational Psychology*, 95(2), 393–405. doi:10.1037/0022-0663.95.2.393

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. doi:10.3102/003465430298487

Jensen, R. W. (1980). A Macrolevel Software Development Cost Estimation Methodology. In *Proceedings of the Fourteenth Asilomar Conference on Circuits, Systems, and Computers*. Pacific Grove, CA.

Jensen, R. W. (2007). *Sage Users Guide*. Brigham City, UT: Software Engineering, Inc.

Linden, van der P. (1994). *Expert C Programming*. Englewood Cliffs, NJ: Prentice Hall.

Novak, J. (1998). *Learning, creating, and using knowledge: Concept maps as facilitative tools in schools and corporations*. Mahwah, NJ: Erlbaum.

Pascarella, E., & Terenzini, P. (2005). *How college affects students: A third decade of research*. San Francisco: Jossey-Bass.

Sprague, J., & Stuart, D. (2000). *The speaker's handbook*. Fort Worth, TX: Harcourt College Publishers.

Zimmerman, B. J. (2001). Theories of self-regulated learning and academic achievement: An overview and analysis. In B. J. Zimmerman & D. H. Schunk (Eds.), *Self-regulated learning and academic achievement* (pp. 1–3). Hillsdale, NJ: Erlbaum.

Zykov, S. (2010). Pattern Development Technology for Heterogeneous Enterprise Software Systems. *Journal of Communication and Computer*, 7(4), 56–61.

KEY TERMS AND DEFINITIONS

Agile: A software development methodology; an alternative to traditional project management where emphasis is placed on empowering people to collaborate and make team decisions in addition to continuous planning, continuous testing and continuous integration.

Agility: Ability to adapt to uncertainties and changes of environment.

CMM (Capability Maturity Model): A methodology used to develop and refine an organization's software development process.

CMM(I) (Capability Maturity Model Integration): A framework of best practices in managing, measuring and monitoring software development processes.

COCOMO (Constructive COst MOdel): An algorithmic method for evaluating and estimating the cost of software development.

COCOMO II (CONstructive COst Model II): A method for estimating the cost, effort, and schedule when planning a new software development activity.

Crisis: Misbalanced production and realization of a surplus value, the root cause of which is separation between the producers and the means of production.

CRM (Customer Relationship Management): A set of practices, strategies and technologies that companies use to manage and analyse

customer interactions and data throughout the customer lifecycle, with the goal of improving business relationships with customers.

Enterprise Agility Matrix: Matrix, the columns of which correspond to processes, data and systems, and the rows of which contain enterprise system levels. Used to detect and predict local crises of software production.

Human-Related Factor: A factor originating from human nature, which influences requirements elicitation, and, consequently, software development.

MSF (Microsoft Solution Framework): A software development methodology; a set of principles, models, disciplines, concepts, and guidelines for delivering information technology solutions from Microsoft.

OOAD (Object-Oriented Analysis and Design): An approach in the system analysis and design through the application of the object-oriented paradigm and concepts including visual modelling.

Oscillator Circuit: An electric circuit, which consists of capacitor and inductive coupling. Uses feedback for oscillation.

Price-S: A composite modelling system for estimating software cost.

Quality Attribute: A systemic property of a software product, which critically influences its quality.

RUP (Rational Unified Process): A software development methodology; a software development methodology from Rational. RUP organizes the development of software into four phases, each consisting of one or more executable iterations of the software at that stage of development: inception, elaboration, construction, transition.

SADT (Structured Analysis and Design Technique): A software engineering methodology for describing systems as a hierarchy of functions.

Sage: A model for estimating software development; includes project management and personnel characteristics.

SCADA (Supervisory Control and Data Acquisition): A category of software application for process control; gathers data in real time from remote locations in order to control equipment and conditions.

Scrum: An iterative and incremental agile software development methodology for managing product development.

Seer: A model for estimating IT project and service costs.

Software Design Pattern: A general reusable solution to a commonly occurring problem within a given context.

Software Engineering: A set of tasks, methods, tools and technologies used to design and implement complex, replicable and high-quality software systems, which include a database.

Third-Generation Programming Language: A high-level programming language such as FORTRAN, COBOL, BASIC, Pascal and C.

XP (eXtreme Programming): A software development methodology which is intended to improve software quality and responsiveness to changing customer requirements; a pragmatic approach that emphasizes business results and takes an incremental approach to building the product through continuous testing and revision.

Information Science and Technology in Crisis Response and Management



Randy Basham

University of Texas at Arlington, USA

INTRODUCTION

Emergent technological innovations in robotics and miniaturization of robotics, drone technologies, acoustical sensors, and others are revolutionizing the effectiveness of crisis response and management efforts, on smaller local events, and in combination, may be applied to larger disaster events such as major community or regional crises. These have also been supported by advances in tracking, communications, information dissemination, identity verification, and location technologies developed over that past decade since Hurricanes Katrina and Rita devastated the central populated region of the United States and Central North America. The combination of these technologies show promise of rapid intervention and rescue, within the most devastated areas affected by major environmental and man-made disasters as they are scaled up to address larger population supports. Location and allocation of resources to the individuals, families, and populations of need may occur nearly unimpeded by obstacles, debris, contaminants and human first responder high risk environments. As these technologies are systematized and develop some level of autonomy in technological problem solving, they are likely to improve overall survivability of populations at risk in the most disaster fragile of environments.

BACKGROUND

Early efforts to link crisis management and information technology were concerned with the protection and maintenance of data within private

business organizations. Businesses' have become dependent on technology to perform work, distribute products and improve productivity and efficiency. Loss of technology functions and critical information could cause irrevocable damage to a business. Therefore crisis management in the business sense has meant anticipatory planning for disruptions and protecting data and process critical for business success. The use of experts within a domain of business services coupled with domain decision makers and gathered into crises management teams; has been highly useful in considering critical scenarios of information or other critical processes disruption or loss. Constructing anticipatory action plans has proven to be a key part of adequate preparation for response, though these must be updated as situational contexts and key personnel change (Esbensen & Krisciunas, 2008).

Approximately a decade ago, in North America the major disasters of Hurricanes Katrina and Rita (Hurricanes: Science and Society, 2015) and there combined effects on a region of the south central United States were heavily televised, in part because of the urban location of the disaster in a modern and well known city. Major humanitarian service providers such as FEMA and the American Red Cross began to grapple with the scale of information needed and service provision possible for such large scale disasters and began seeking community and academic partnerships with corporate information technology providers, the defense industry, uniformed military service providers (especially those engaged in search and rescue efforts) and governmental agencies to improve crises response and management, for disasters of scale (Hurricane Katrina Disaster Relief, 2015).

DOI: 10.4018/978-1-5225-2255-3.ch121

Since this time several additional cataclysms have increased awareness for the need to look at international and global response systems involving information technology, relative to preparedness, systemization, planning, pre-positioning of resources and equipment, logistics and distribution capacity in the management of crisis and disaster management. Public attention has been galvanized by devastating natural disasters such as; earthquakes in Pakistan in 2005 and in Sichuan China in 2008 and again in Haiti in 2010, the Tohoku earthquake and tsunami striking Japan in 2011, Cyclone Nargis that struck Myanmar in 2008 and by recent recollections of earlier but recent large scale disasters prior to the North American hurricanes of 2005 such as; the Sumatra tsunami of 2004, and the Iran earthquake of 2003, and several earlier to these and many others since (Lists of Disasters, 2015).

CRISIS RESPONSE AND MANAGEMENT: TECHNOLOGY CONSIDERATIONS

Crises and disasters are sudden and unforeseen. Each is serious, disruptive, overwhelming and often exceeds capacity to restore order and normalcy in any expected time frame. However, crisis or disaster events may occur with some frequency and predictability, and standardized responses may be thoughtfully and consistently applied. Others may be anticipated to occur occasionally but irregularly, and require a level of creativity and innovation over a short period in response. There are of course, improbable and unexpected crises and disasters which occur infrequently, but of which there are no precedents or examples to follow, and for which we have little or no prior experience. These are often widespread and devastating occurrences, requiring longer term determination and resolve to manage and to overcome or restore some level of normalization. Technology and information based systems may be designed to address each, as these relate to specific probabilities for their occurrence,

and with attention to scalability of responses as needed for each type of crisis or disaster.

SWOT (Strengths, Weaknesses, Opportunities, and Threats) Analysis for Institutions

Though there are a number of possible and foreseeable crises that may affect any organization, information science technology is emerging as a critical crisis response consideration. Economic downturns, loss of funding, loss of physical and human assets, and employee litigation (Crisis Management, 2016) may all be tied to the consistency sustainability and privacy issues associated with information technology management protocols and continuity. A strategic planning method that may be quickly used by employees not fully familiar with an organization is to develop a simple SWOT plan of key and critical inquiries relative to information technology management or other strategic plan, crisis response initiative of the organization. Some possible lines of inquiry and planning for information technology management are given in Table 1 below, these should be strategically prioritized and determined both internally and externally:

Information Science Technology: Institutional Crisis Response and Management

Crisis response and management as related to institutions, or business, information science and technology requires crisis management teams are often designated in advance of a crises and focus crisis and recovery plans on avoiding a crisis, preparing and planning for crises, identifying threats and critical systems for protection, training key personnel for crisis events and recovery, communicating plans to organizational members, and reviewing and updating plans and disseminating crisis response guidelines. Phases of institutional Crises Stages (Fink, 1986) involving information technology have been identified as:

- **Prodromal Stage:** The first time those potential crisis symptoms occur. Early warnings may be missed, or may not occur creating unexpected crisis and causing decision makers to revert to a damage control process rather than preventing or limiting damage (Fink, 1986).
- **Acute Stage:** There is no preventative recovery, or turning back. The crisis begins and the eruption is noticeable. The crisis ends when the damage or disruption becomes unnoticeable (Fink, 1986).
- **Chronic Stage:** The recovery phase includes investigations, audits, explanations, self-analysis, self-doubt and healing. This phase is often used to analyze what went wrong and how to improve response in future (Fink, 1986).
- **Resolution Stage:** The crisis is resolved by managing the three prior phases of crisis. Alerts or prodromes are taken seriously in anticipation of avoiding or reducing the effect of any upcoming potential or foreseeable crises (Fink, 1986).

Most institutions with crisis management experience relative to information technology develop contingencies to suppress the destructive potency of a crises or critical event.

Key Theories Underlying Crisis Response and Disaster Response

At the institutional level crises are perceived as situational threats that may affect institutional reputation which may be moderated by application of communication theory, to minimize reputational harm (Coombs, 2007). Though larger scale disruptions external to the institution that may require disaster response are influenced by chaos theory and the perceived need to restore normalcy (Kiel, 1995). Similarities in each of these approaches to retain stability, or restore stability during periods of extreme stress to the institution or greater environment may be identified in

classic crisis intervention theory as conceived by Aguilera and Messick as early as 1970. Crisis Intervention Theory maintains, that pre-existing equilibrium may be disrupted and result in disequilibrium among human organisms at the onset of extreme or prolonged stress. There is a pressing need to restore equilibrium. If there are sufficient balancing factors present, then no crisis occurs, or is sustained. Though, if sufficient balancing factors are absent, then a crisis will occur, or be sustained, though a lesser or less adequate level of equilibrium may still be possible (Aguilera, Messick, & Farrell, 1970).

Developing Crisis and Disaster Prevention and Response Infrastructure

A number of organizations and corporations are actively engaged in developing the technological architecture for preventing both institutional and non-institutional crises and disasters that occur with some frequency, and require standardized responses. These are preventable crises and probable disasters, or those of which a known infrastructure may be developed to rapidly and efficiently respond, so as to reduce loss of control. The costs of development are justifiable, and incidence rate may be statistically tracked with some accuracy for prediction of occurrence. These include those events that may be anticipated, and allow sufficient time to prepare for. Preparation is based upon what is known and what may be known, of which, there is an ability to take preventative action, and that there is an ability to plan to act to maintain or regain control where possible. Recent information technology advances related to crisis or disaster response capacity may be seen in Table 2 below:

Individual Hazards Reduction: Disaster Response and Management

Disaster response and emergency management may be related to single institutions as causal (i.e. Bhopal India, Chemical Spill in the 1980's),

though the effects are often experienced community wide or regionally, whether manmade or natural. A single natural event may also be causal, but cause unexpected community and regional effects across institutions and technologies. These disrupt daily living and livelihoods, and cause loss of life, community structures, and normalcy, are widespread, and last for longer periods. Emergency Management Phases (U.S., FEMA, 2015) involving local or regional recovery, but those focuses upon individual or small group responses have been identified as:

- **Mitigation Phase:** Preventing future emergencies or minimizing their effects. Includes any activities that prevent an emergency, reduce the chance of an emergency happening, or reduce the damaging effects of unavoidable emergencies. Buying flood and fire insurance... (U.S., FEMA, 2015).
- **Preparedness Phase:** Preparing to handle an emergency. Includes plans or preparations made to save lives and to help response and rescue operations. Evacuation plans and stocking food and water are both examples of preparedness. Preparedness activities take place before an emergency occurs (U.S., FEMA, 2015).
- **Response Phase:** Responding safely to an emergency. Includes actions taken to save lives and prevent further property damage in an emergency situation. Response is putting preparedness plans into action. Seeking shelter from a tornado or turning off gas valves in an earthquake... (U.S., FEMA, 2015).
- **Recovery Phase:** Recovering from an emergency. Includes actions taken to return to a normal or an even safer situation following an emergency. Recovery includes getting financial assistance to help pay for the repairs. Recovery activities take place after an emergency (U.S., FEMA, 2015).

Most individuals, families, small groups, or some communities with disaster or crisis management training or experience develop contingencies to suppress the destructive effects of an emergency or disaster, including preparing for disruptions in information and media access, and energy, transportation, and communications disruptions. However, larger scale disasters may overwhelm these individual efforts.

Fragile and At Risk Systems and Environments

Within fragile, or at risk, business, manufacturing, or service systems, or within geographically defined areas such as a neighborhood, community, nation, or region, infrastructure to resist various forms of crisis or disaster effects, either natural or manmade is often inadequate for the risk profile of the institution or area (ReliefWeb International, 2015). Technology may be defined differentially, based upon the specific crisis prevention and response infrastructure needed for recovery. For example, regional risks may include inadequate roads, transportation, water supplies, food, or agricultural supports, weather reports, available media, energy access, shelter or insulation from the elements, access to money, or commercial goods needed for survival, medication, medical care, hygiene supplies, information, and communication, though this listing is not exhaustive. Each facility, institution, community, or geographic area or environment has a uniquely definable fragility, or risk profile. Planning, pre-positioning of equipment, services, or supplies and the systematic addressing of infrastructure limitations may attenuate the devastating effects of a crisis or disaster. Information technology is an essential part of and critical requirement in the evaluation of fragility and the provision of effective decision supports to responders and disaster response organizations. Recent device technology advances related to crisis or disaster response capacity may be seen in Table 3, below:

Humanitarian Action in Crises and Disaster Response

Whether the crisis is of smaller scale or more broadly distributed and classified as a disaster, local people such as: organizational volunteers, employees, or available citizens, with or without adequate supportive equipment or training, are most likely first responders. Approximately 90% of lives saved in a life threatening disaster are saved by local people (Calhoun, 2015). These first responders are themselves vulnerable to the same threats and losses as those they are attempting to rescue. These responders often have limited access to adequate and timely information, communication capacity, and needed health care services and medical expertise, or medications and medical supplies. In smaller crises, lack of information, communication and available but needed equipment and expertise are common. Smaller crises however, if unattended may increase risks for the development of full blown disasters if unresolved, within a risk reduction time frame (Author, 2006).

Post Crisis or Disaster: Resilience Engineering

The quote attributed to George Santayana, *Those who cannot remember the past are condemned to repeat It* (Internet Encyclopedia of Philosophy, 2015) with a variant frequently used by Sir Winston Churchill *Those who fail to learn from history are doomed to repeat It* (National Churchill Museum Blog, 2015) provide in their historical contexts sufficient example and justification to improve conditions following crises and disasters that could not be reasonably predicted to assure that they are less likely in the future.

Engineers and decision makers are increasing reliance on the principles and practices of resilience engineering for strengthening infrastructures and to reduce fragility and risk, due to potential damages from crisis or disaster.

Resilience engineering (Resilience Engineering Association, 2015) is an emerging and non-conventional approach to risk management, especially in engineering that examines methods to improve the ability at all levels of organizations to create processes that are robust yet flexible, to monitor and revise risk models, and to use resources proactively in the face of disruptions or ongoing production and economic pressures. Relative to information science and technology, the concept of information technology, communication technology, or computing technology failures, by robust design, are not due to the breakdown or malfunctioning of normal system functions, but are more likely to represent the limitations of human operators to anticipate adaptations necessary to changing conditions in the real world.

Issues, Controversies, Problems

Improving the saving of life and the limitation of longer term adverse effects during disaster is in part due to recent advances in strategic hardware devices such as remote telemetry, surveillance, multipurpose vehicles, including amphibious vehicles, and aerial unmanned surveillance and package delivery vehicles (drones). The capacity to gather situational awareness information remotely through acoustic sensors and bio monitors, ground and object penetrating scanning devices have greatly improved critical information gathering for crisis and disaster management. Additionally, mechanical or robotic search devices, and the interconnectivity of these between multiple information systems and operators, either human or computing, are also reducing barriers to rescue and recovery efforts.

SOLUTIONS AND RECOMMENDATIONS

Often technologies that have intelligence gathering capacity, logistical capacity, risk evaluation

capacity and autonomous decision making ability are used for defensive and military purposes. However, these same technologies show great promise of being able to ameliorate the damage and destruction created by both natural and manmade disasters, as well as crisis response at the institutional level. Deploying these technologies and a wide scale and imaginative engineering focused upon improving the hardiness or resilience of civil, municipal, transportation, communication, information science and related structures will also mitigate the effects of crisis and disasters and improve overall response, and recovery.

FUTURE RESEARCH DIRECTIONS

Improved information science driven mapping and location technologies, including, for persons at risk or endangered, in a crises or disaster, automated deployment of assets, at a threshold level of early warning, improved logistics in planning and prepositioning and distribution of needed resources, are all critically needed areas for additional research. Rapidly self-correcting decision making software or applications for critical information systems restoration (financial, energy, information, etc.) would reduce legacy losses of vital knowledge needed to continue organizational, institutional, community or regional functions.

Advanced light weight composite material with protective and insulating properties that may be used as wearable information science interface platforms as clothing and with embedded bio sensors and others focused upon external environmental or risk conditions need further additional inquiry and would likely enhance the effectiveness of response personnel. Miniaturization, stable energy source, and autonomous functioning and sensing capacities for robots, drones and others would rapidly reduce risks in locating those in need and reducing risks. Networked visual (camera systems) through damages areas either urban or rural would also contribute to crises response.

CONCLUSION

Crisis response and management have improved steadily with access to improved situational awareness and information distribution. Satellite systems have improved weather prediction, and advanced warning alerts for some naturally occurring disasters. Emerging sensor technologies may improve earthquake and tidal wave or tsunami predication as previously difficult to predict natural events. Global positioning satellite systems have improved mapping and location technologies to enhance disaster location and improve logistical responses. Satellites have also improved knowledge of slowing developing disasters such as droughts, or water level and availability changes. These have assisted in providing additional supports to affected regions over time. The transmission of vital information over the internet and in combination with satellite technology, media broadcasting globally, and the advent of cloud storage or off site storage of critical information and redundant information archives have limited data and information loss and improved consistency and availability of access to both decision makers and affected populations during an emergency. Indeed the emphases on crisis response or disaster response and management knowledge for corporations, other institutions, helping organizations, the military services and governments have led to widespread and more focused levels of preparation and likelihood of limiting damages during a crisis, information based, or physical.

REFERENCES

- Aguilera, D. C., Messick, J. M., & Farrell, M. S. (1970). *Crisis Intervention: Theory and Methodology*. St. Louis, MO: The C.V. Mosby Company.
- Calhoun, L. (2015). *COPE: Community Outreach Promoting Emergency Preparedness: Map your neighborhood*. Retrieved October 14, 2015, from <http://www.cope-preparedness.org/html/Map%20Your%20Neighborhood-1.pdf>

- Coombs, W. (2007). Protecting Organization Reputations During a Crisis: The Development and Application of Situational Crisis Communication Theory. *Corporate Reputation Review*, 10(3), 163–176. doi:10.1057/palgrave.crr.1550049
- DHS. (2010). *Department of Homeland Security Risk Lexicon*. Retrieved October 14, 2015, from <http://www.dhs.gov/xlibrary/assets/dhs-risk-lexicon-2010.pdf>
- Earth, G. (2015). *Google Earth: Let's Go Explore*. Retrieved October 1, 2015, from <https://www.google.com/earth/>
- Esbensen, L. H., & Krisciunas, T. (2008). *Crisis management & information technology (Unpublished master thesis)*. INFM02 VT2008 (15hp). Retrieved September 26, 2015, from lup.lub.lu.se/record/1337842/file/1646737.pdf
- FEMA. (2015). *Animals in disasters: The four phases of emergency management. Module A, Unit 3*. Retrieved October 1, 2015, from http://www.training.fema.gov/emiweb/downloads/is10_unit3.doc
- Fink, S. (1986). *Crisis management: Planning for the inevitable, an authors guild backprint.com edition*. iUniverse Inc.
- Google Crisis Response. (2015). *Google Crisis Response (Google Crisis Maps, etc.)*. Retrieved October 12, 2015, from <https://www.google.org/crisisresponse/resources.html>
- Google Person Finder. (2015). *A Google Org. Project*. Retrieved October 6, 2015 from <https://google.org/personfinder/global/home.html>
- Hurricane Katrina Disaster Relief. (2015). *Wikipedia; The Free Encyclopedia*. Retrieved October 2, 2015, from https://en.wikipedia.org/wiki/Hurricane_Katrina_disaster_relief
- Hurricanes: Science and Society. (2015). *Hurricanes Katrina and Rita*. Retrieved October 14, 2015, from <http://www.hurricanesociety.org/history/storms/2000s/rita/>
- Internet Encyclopedia of Philosophy. (2015). *A peer reviewed academic resource. George Santayana (1863-1952)*. Retrieved October 5, 2015, from <http://www.iep.utm.edu/santayan/>
- Kiel, D. (1995). Chaos Theory and Disaster Response management: Lessons for Managing Periods of Extreme Instability. In *What Disaster Response Management Can Learn From Chaos Theory, Conference Proceedings*. Retrieved July 30, 2016 from; http://www.library.ca.gov/crb/96/05/over_12.html
- Lists of Disasters. (2015). *Wikipedia; The Free Encyclopedia*. Retrieved October 4, 2015, from https://en.wikipedia.org/wiki/Lists_of_disasters
- Lynxx Technologies. (2015). *Lynxx Technologies; The new standard*. Retrieved October 3, 2015, from; <http://lynxxtechnologies.ca/video/amphibious-design/>
- Management, C. (2016). Retrieved, July 23, 2016, from <http://www.strengtheningnonprofits.org/resources/e-learning/online/managingcrisis/default.aspx?chp=0>
- National Center for Missing and Exploited Children. (2015). *Unaccompanied Minors Registry*. Retrieved October 11, 2015, from <http://www.missingkids.com/DisasterResponse>
- National Churchill Museum Blog. (2015). *Designated by Congress as America's National Churchill Museum, Those who fail to learn from history*. November 12, 2012. Retrieved October 12, 2015, from <https://www.nationalchurchillmuseum.org/blog/churchill-quote-history/>
- ReliefWeb International. (2015). *International Alert, Policy Brief May 2015: Compounding risk: Disasters, fragility and conflict*. Retrieved October 11, 2015, from http://reliefweb.int/sites/reliefweb.int/files/resources/ClimateChange_DisastersFragilityConflict_EN_2015.pdf

Resilience Engineering Association. (2015). *The official home of resilience engineering*. Retrieved October 11, 2015, from <http://www.resilience-engineering-association.org/>

Search and Rescue Optimal Planning System. (2015). *Wikipedia; The Free Encyclopedia*. Retrieved October 11, 2015, from https://en.wikipedia.org/wiki/Search_and_Rescue_Optimal_Planning_System

Shelterbox. (2015). Retrieved September 29, 2015, from <http://www.shelterbox.org/>

ADDITIONAL READING

Adrot, A., & Robey, D. (2008). Information technology, improvisation and crisis response: Review of literature and proposal for theory. *AMCIS 2008 Proceedings*, 397.

Alfonso, G.-H., & Suzanne, S. (2008). Crisis Communications Management on the Web: How Internet-Based Technologies are Changing the Way Public Relations Professionals Handle Business Crises. *Journal of Contingencies and Crisis Management*, 16(3), 143–153. doi:10.1111/j.1468-5973.2008.00543.x

Apvrille, L., Tanzi, T., & Dugelay, J. L. (2014, August). Autonomous drones for assisting rescue services within the context of natural disasters. In *General Assembly and Scientific Symposium (URSIGASS), 2014 XXXIth URSI* (pp. 1-4). IEEE. doi:10.1109/URSIGASS.2014.6929384

Boin, A., & McConnell, A. (2007). Preparing for critical infrastructure breakdowns: The limits of crisis management and the need for resilience. *Journal of Contingencies and Crisis Management*, 15(1), 50–59. doi:10.1111/j.1468-5973.2007.00504.x

Comfort, L. K. (1993). Integrating Information Technology into International Crisis Management and Policy. *Journal of Contingencies and Crisis Management*, 1(1), 15–26. doi:10.1111/j.1468-5973.1993.tb00003.x

Fan, L. (2012). Shelter strategies, humanitarian praxis and critical urban theory in post-crisis reconstruction. *Disasters*, 36(s1), S64–S86. doi:10.1111/j.1467-7717.2012.01288.x PMID:22687157

Franke, J., Charoy, F., & Ulmer, C. A Model for Temporal Coordination of Disaster Response Activities. *7th International Conference on Information Systems for Crisis Response and Management (ISCRAM'2010)*, May 2010, Seattle, WA, United States. 2010, <<http://www.iscram.org>>. <inria-00505194>

Hitt, M. A., Ireland, R. D., & Hoskisson, R. E. (2012). *Strategic management cases: competitiveness and globalization*. Cengage Learning.

Rao, R. R., Eisenberg, J., & Schmitt, T. (Eds.). (2007). *Improving disaster management: the role of IT in mitigation, preparedness, response, and recovery*. National Academies Press.

KEY TERMS AND DEFINITIONS

Crisis Management: The application of strategies designed to help an organization, or community, deal with a sudden and significant negative event.

Crisis Response: The methods used to offer short term immediate help to individuals who have experienced a crisis, disaster or catastrophic event.

Device Technology- Crisis/Disaster Response: Mobile devices and technology enhanced mechanical devices with computing and problem solving capabilities, networked communication capacity, remote sensing and transport abilities to provide, alerts, response assistance and rescue assistance.

Fragility – Fragile, At Risk, Environments: Easily broken, shattered, or damaged; delicate; brittle; frail: a fragile ceramic container; a very fragile alliance, also applies to communities, regions or nations with limited capacity or infrastructure to recover from calamity.

Information Science: The study of processes for storing, organizing, retrieving or recovering information, especially scientific or technical information.

Information Technology- Crisis/Disaster Response: The study or use of systems (especially computers and telecommunications) for storing,

retrieving, and sending information; routinely or in special event circumstances, or condition.

Manmade Disaster: Anthropogenic hazards or human-made hazards can result in the form of a human-made disaster. In this case, anthropogenic means threats having an element of human intent, negligence, or error; or involving a failure of a human-made system. This is as opposed to natural hazards that cause natural disasters.

Natural Disaster: A natural event such as a flood, earthquake, or hurricane that causes great damage or loss of life.

APPENDIX 1

Table 1. SWOT Analysis for IT Management and Crisis Response (Situation being analyzed: Information Technology Loss or Disruption)

Strengths What is the company doing well, relative to information technology management? What information technology resources does the company provide that are important to continuity, sustainability and crisis response? Is the company making best use of available information technology?	Weaknesses What resources or services could the company improve, relative to information technology management? What areas does the organizational competition have a technological edge relative to information technology management? What adverse information technology should the organization actually avoid?
Opportunities How will information technology change? What are the interesting trends in technology that could be exploited? What needs in the information technology market are not being addressed by others?	Threats What is the competition doing well, relative to information technology? What information security threats should the business be most concerned about? What new innovations are others bringing, that would affect information technology management?

APPENDIX 2

C

Table 2. Information science and technology: crisis or disaster information technology advance

Information Technology Advance	Crisis/ Disaster Response Capacity
Satellite and transmission technology (including radio, video, ham radio, etc.)	Satellite conveyed information has become the primary conveyance system for information, communication, and geographic location data. Though satellites were originally designed to serve military and entertainment roles, over time they have become a tool for humanitarian and educational transfer of knowledge.
Computing capacity	Communications, information dissemination, data archiving, automation of routine tasks such as sending or updating alerts, automated calling or text messages
Data archiving capacity	Cloud based archiving of data, large hard drives, large servers, portable memory devices, data backup services and devices, etc.
Graphical user interface	Windows and Macintosh operating systems and proprietary corporate and other derivations, including a mouse or similar device, to select and display information rapidly
Global positioning data, global imaging data	GPS location, Google Earth and similar photo mapping applications (Google Earth, 2015).
Environmental data server, geographic information systems	Mapping, weather, and available distinguishing features, services available, terrain and available transportation routes (Google Crisis Response, 2015).
Mobile technology, personal and configurable, connectivity to information systems	Cellular phones, paging systems, text message, and email or automating phone message delivery upon detection of any predetermined activating event; capacity to archive directories of contacts and key services locally; capacity to access the internet to search for information or help; capacity to download customized application on rendering first aid, locating transportation, etc.
Crisis informatics	Continuous situational updating and risk parameter information either queries or automatically updated and displayed, or communicated
Crowdsourcing of data.	Matching or linking applications, that match available supplies to need within an area, match volunteers to disaster area, match perspective labor with available work, etc. (Google Person Finder, 2015).
Identity verification	Online financial and membership application that confirm, or authenticate identity
Tracking technologies	RFID chips, embedded in credit cards, or ID cards and GPS applications located in mobile phones, vehicles, or with assigned personnel or emergency broadcasting device
Post disaster reunification of children with families	(The National Center of Missing and Exploited Children; Unaccompanied Minors Register, 2015), (USA)
Simulators	(Search and Rescue Optimal Planning System, 2015). US Coast Guard (SAROPS) is made up of the Graphical User Interface (GUI), the Environmental Data Server (EDS) and the Simulator (SIM) which can predict: 1. probability of containment, 2. probability of detection, and 3. probability of success
Social networking	Provide an online location for family members, friends or contacts to check in and locate each other, update news, or share alerts: Facebook, Twitter, and similar networks

APPENDIX 3

Table 3. Information science and technology: interdependent crisis or disaster technology devices

Device Technology Advance	Crisis/ Disaster Response Capacity
Remote and portable sensors and transmitters; acoustic, biological, chemical, infrared, microwave, motion, thermal, radiation, visual, and others	Depending upon energy supply, sensors may be connected to computing and tele communications networks to provide continuous or situational monitoring and updating broadcasted to any configured receiver at distance.
Smart clothing, insulating and protective fabrics, sensor and transmission devices, exoskeletons, etc.	Clothing may be configured as shelter, protective device against injury, as a technology platform for sensing and communication; with designs currently available; exoskeletons permit the wearer to reduce limitations of some disabilities and increase body strength or endurance
Biological and health sensors; on person, in location	Exposure to specific contaminants, Wi-Fi enabled health monitoring software (cardio, pulse, EKG, EEG, kinesiology, etc.).
Linked camera systems, municipal and other	Traffic cams, sport event cams, accessible, security and weather cams, etc.
Water purification and transport devices and distribution	Solar powered portable mobile water purification systems, mobile dispatch software for water transport supplies, logistics tracking and volume
Food preservation and long term storage devices and distribution	Pulsed electric and magnetic field treatment to inactivate bacteria and limit quality loss in transport and storage, packaging and oxygen removal, mobile dispatch software for food transport supplies, logistics tracking and volume
Energy storage and production and distribution	Long term marine batteries that hold charge up to 10 years, mobile fuel powered generators, portable solar panel, mobile power plants, etc.
Portable and temporary shelter	(ShelterBox, 2015), an international disaster relief agency along with Rotary International, provides inexpensive relief shelters. The organization has tracking systems to monitor weather systems, the likely scale of hurricanes and cyclones and earthquake alert system. Other temporary shelter companies provide similar.
Advanced extraction, evacuation and hazard clearing tools	Devices or mechanical appendages often networked and monitored which are attached to multipurpose transportation devices and have specialized functions. Example; chain saws to clear hazards or obstacles, mobile battering ram, bomb detection, etc.
Transportations systems (unmanned) and human responder interface, or independent function	(Lynxx, 2015) Multipurpose vehicles, amphibious vehicles, drone technologies, etc.
Robotics and human responder interface, search, burrowing, medical, etc.	Larger automated and frequently mobile mechanical devices with some computing and communication capacity and other specialized functions or capacities such as: traversing difficult and hazardous terrain and gathering risk and other information, transporting supplies, digging or repetitive manualized labor tasks, and performing human supervised medical procedures at a distance
Miniaturization of robotics, and human responder interface, or independent function (disaster response micro robots)	Remote sensing devices, including miniature cameras attached to miniaturized robots to perform search or information gathering functions in high risk environments

Category C

Curriculum Development and Instructional Design

Addressing Digital Competencies, Curriculum Development, and Instructional Design in Science Teacher Education

Isha DeCoito

Western University, Canada

INTRODUCTION

New realities of the 21st century demand individuals with different competencies than those considered appropriate for success in the past. Consequently, education must change. Milton (2015) maintains that surface changes in education will not equip students for the 21st century and that change is needed at the core of educational practice. A shift must occur from the traditional view of educational practice to a transformative view. Moreover this shift must aim to incorporate technologies in schooling in a manner that digress from disciplinary experts' determinations of what and how students should learn – a classic perspective which has resulted in challenges for educators as they continue to search for strategies to effectively address the development of skills reminiscent of the preferred learning styles of today's students.

With each passing year, technology becomes a more predominant part of educational culture (Bolstad & Gilbert, 2006; Cox, 2008). Simply introducing technological tools and infrastructure into schools will not trigger beneficial and meaningful educational change. Moreover, technology cannot be effective in the classroom without teachers who are knowledgeable about both the technology itself and its implementation to meet educational goals, that is, teachers who are technologically literate. Thus, it can be said that, while technology use in the classroom is increasing, improving learning through the application of these literacies should remain the

goal. Changes are inevitable if technology is to make a difference in curriculum design and address the needs of 21st century learners. The impact of technology and the changing face of curriculum, as well as the accompanying changes in the roles of teachers can no longer be ignored; roles must be reconceived in order to engage learners in many decisions about their learning (Bennett, 2002; Bolstad, Gilbert, McDowall, Bull, Boyd, & Hipkins, 2012). Achieving changes associated with the integration of technology in the overall learning environment will require efficient teacher training in teacher education programs (Brush & Bannon, 1998). This begs the question: What kinds of modeling and scaffolding should educators or designers provide to help learners engage in this process?

Teacher professional development (PD) is absolutely essential if technology is to be used effectively; PD should entail initial preparation/training – pre-service, in-service, and ongoing pedagogical and technical support for teachers as they address their daily challenges and responsibilities. Training and on-going inquiry-based approaches imply that support should go beyond teaching skills in technology use and focus on the effective pedagogical use of the technology to support teaching and learning goals (DeCoito & Richardson, 2016).

In this article the author reports on a mixed-methods study with a focus on science teacher education. Specifically, the study addresses the development of secondary science teacher candidates' (TCs') digital competencies as they

explored the integration of digital literacies in a science methods course, and its potential to enhance teaching and learning in science, including curriculum and instructional design. The author maintains that in order to develop the necessary skills and application practices of technology integration, and enhance technological literacy, TCs must be presented with appropriate experiences in teacher induction programs.

BACKGROUND

Teaching and Learning in the 21st Century

The preparation of young people for lifelong learning in a 21st century knowledge-based information society has become an increasingly important objective of educational systems worldwide (Dagiené, 2011). Multi-literate, creative and innovative individuals are seen as “drivers of the 21st century and the prerequisites to economic success, social progress and personal empowerment” (Canadians for 21st Century Learning and Innovation, 2012). A primary challenge for education is to transform student’s learning processes to engage student interest in gaining 21st century skills and knowledge. Lemke (2004) reported a link between 21st century skills and academic achievement, making the case for incorporating teaching activities, including digital technologies that nurture these skills (see further elaborations, National Research Council, 2011). Despite the fact that these skills have been around for a very long time, currently the challenge is, according to Bolstad et al. (2012), “how to achieve a shift that creates a more coherent educational ecology that can support what is known about good learning *and* that can accommodate new knowledge about learning and, importantly, new *purposes* for learning in a changing world”. As such, there is a need for students to experience opportunities that will result in enhancing 21st century learning skills.

The acquisition of 21st century skills in teaching and learning will require a shift in what we teach, how we teach it, the tools we use and how we educate, train, nurture, and retain our teachers. It is undeniable that we cannot change how our students learn unless our teachers are equipped to teach in new ways (Murnane & Levy, 2004). Thus, the scope and components of the research underlying this article were informed by the conceptual framework of a constructivist model of learning applied to teacher education – acknowledging that for TCs to develop their ability to teach, they must be provided opportunities to actively construct their understandings of pedagogical content knowledge (Chai, Koh, & Tsai, 2010; Graham, 2011) and integrate new understandings with prior knowledge (Hollingsworth, 1989; Shulman, 1986) under a technological, pedagogical, content knowledge (TPACK) framework (Mishra & Kohler, 2006). The constructivist view of education suggests that the theory of individual learning is flawed and furthermore, is not conducive to multi-literacy pedagogy in multimodal environments. The constructivist view is that students learn best, not by assimilating what they are told, but rather through a knowledge-construction process (Bereiter, 2002). In order for individuals to learn how to construct knowledge, it is necessary that the process be modeled and supported in the surrounding community; that is, a learning community. In a learning community the goal is advancing the collective knowledge while supporting the growth of individual knowledge (Scardamalia & Bereiter, 1994). The defining quality of a learning community is that there is a culture of learning, in which everyone is involved in a collective effort of understanding. This was in fact nourished, supported, and modeled throughout the science methods course as teachers engaged with digital literacies on various levels.

Digital Literacies

Education today is faced with the challenge of adapting to an environment where literacies are

becoming ever more important –how knowledge is represented, as well as the mode and media chosen, is a crucial aspect of knowledge construction, making the form of representation integral to meaning and learning more generally (Cope & Kalantzis, 2000; Martin & Grudziecki, 2006). It follows, then, that to better understand learning and teaching in the multimodal environment of the contemporary classroom, it is essential to explore the ways in which representations in all modes feature in the classroom (Jewitt, 2008).

The term ‘digital literacy’ is contested, and understood in different ways by different people and is often referred to as “digital competency”, meaning a set of knowledge elements, abilities, dispositions, and conducts that enable individuals to know how ICTs work, what they are for, and how they can be used to attain specific objectives (Castells, 2009; Grant, 2010). Digital literacy has also been defined as an umbrella framework for a number of complex and integrated sub-disciplines comprised of skill, knowledge, ethics and creative outputs in the digital network environment (Calvani, Cartelli, Fini, & Ranieri, 2008). The shift in our society’s growing reliance on technology mandates that education emphasize technological/digital literacy including: digital competence (skills, concepts, approaches, attitudes), digital usage (professional/discipline application), and digital transformation (innovation/creativity) (Belshaw, 2011). According to Spires et al. (2012), when considering digital literacy one can organize the related cognitive and social processes into three categories: i) locating and consuming digital content; ii) creating digital content; and iii) communicating digital content.

ICT are reshaping the ways in which students learn and prefer to learn, including fluency in multiple media and simulation-based virtual settings and co-designing learning experiences that are personalized to individual needs and preferences (Annetta, 2008). While educators and others have listed and described 21st century skills, teachers continue to search for strategies to effectively address the development of these

skills in the preferred learning styles of today’s students (Bolstad et al., 2012; Greenhow, Robelia & Hughes, 2009). Emerging learning styles signal that teachers adapt their teaching styles, and according to Ng (2010), the new styles should reflect multi-literacies pedagogy which focuses on multimodal ways of conveying understanding. Improving digital literacy underpins not only a nation’s capacity to provide individuals and groups with equity of access to social opportunity; it is a necessity for participation in the digital economy (DeCoito & Richardson, 2016).

Digital Competencies and Science Education

Digital literacy is important for science learning as it i) assists students to learn more effectively with the range of ICT-enabled affordances that have the capacity to motivate and enable better understanding of science concepts, and ii) lessens the working memory’s cognitive load, while learning science that is ICT-based (Ng, 2011). The affordances offered by ICT benefit science learning and potential benefits of integrating digital literacy in science education have been documented, including increasing student motivation (Osborne & Collins, 2000), expanding the pedagogical resources available to science teachers (Al-Alwani, 2005), promoting cognitive development, and highlighting relevance via relating science to students’ real-life experiences (Webb, 2005). For example, DeCoito (2014) explored how digital scientific timelines can be used to teach about the nature of science, and more recently, the potential of digital games to enhance scientific and technological literacy while teaching STEM concepts (DeCoito & Richardson, 2016).

The challenge for educators is to utilize these aforementioned literacies in ways that facilitate the highest level of learning outcomes. Educators face a number of challenges when deciding to integrate digital literacies in the classroom. Among these are choices about which one(s) to use, how effective the choices are in reinforcing learning, and the

role of these digital literacies in shifting from an instruction paradigm, which is teacher focused, to a learning paradigm, which is student focused (Cox, 2008). Dede (2014) maintains that the goal of using technology should be to empower teachers to make better use of instructional strategies that provide opportunities for self-directed learning, which foster academic engagement, self-efficacy, and tenacity by requiring students to define and pursue specific interests. Furthermore, it should personalize learning, which ensures that students receive instruction and supports that are tailored to their needs and responsive to their interests.

Research studies (Ertmer et al., 2012; Sancar-Tokmak, Sürmeli, & Özgelen, 2014; Sasseville, 2004; Zhang & Martinovic, 2008) revealed that TCs' levels of self-efficacy (Skaalvik & Skaalvik, 2008) and attitudes toward ICT determined whether or not, and how, they used ICT in their classrooms. For example, Ertmer et al. (2012) found that teachers' attitudes towards technology play an important role in their integration efforts. In general, TCs believe that they should become more ICT literate because it may enhance teaching and learning in all classrooms; however, they believe that they do not possess the skills to integrate ICT in their future practices (Gomes, 2005). Dilworth et al. (2012) demonstrated that to use technology effectively, a deep understanding of content and related pedagogical knowledge is essential.

TCs provided various reasons for their attitudes toward the use of ICT in teaching and learning, including lack of adequate teacher training in effective use of ICT in the classroom (Albirini, 2006); fear that time taken for technological training infringes on time for learning content; and concern that an emphasis on technology may distract students and teachers from teaching and learning (Dipetta & Woloshyn, 2009). Despite the affordances offered by ICT, the development of digital literacy may require a focus on technology for initial periods of time. As expertise develops with the technology, learning becomes more learner-centred and over time, the focus shifts to the engagement of learners with science content (Ng, 2011).

THE STUDY

The Course

The science methods course was offered at the senior level and catered to TCs whose speciality was biology, chemistry, or physics. The implementation of technology was promoted widely in this science methods course to expose TCs to multi-literacy pedagogy and develop their technological literacy, as well as engage them in creating digital science content, and promote creativity. The instructor used technology extensively and invited guests to conduct workshops on technology enhanced learning, including biotechnology and the effective use of digital games. The course assignments spanned over two semesters and incorporated the integration of digital literacies (information literacy, computer literacy, media literacy, communication literacy, visual literacy, and technology literacy) on a variety of levels as TCs developed a) digital interactive case studies exploring socioscientific issues; b) scientific timelines using, e.g., Prezi, Movie Maker, Tiki Toki, etc.; c) concept presentations in science (e.g., including digital games, simulations, virtual laboratories, etc.); and d) science resource websites focusing on multimedia interactive activities.

Methodology

A mixed-methods design (Mills, Durepos, & Wiebe, 2010) was utilized to address the development of TCs' digital competencies over a 2-year period. Participants included 39 TCs enrolled in a senior science methods course in a Faculty of Education at a Canadian university. Data sources included questionnaires (views of, experiences with, and attitudes towards technology; and development of course assignments); TCs reflections on the course assignments; audio-taped, semi-structured interviews about TCs' views of the role of digital literacies in science education; and student work. Quantitative data were analyzed using statistical analysis (Microsoft Excel) and de-

scriptive statistics. Qualitative data were analyzed through an interpretational analysis framework executed through the process of thematic coding and constant comparative method (Stake, 2000) to address the goals of the study.

Findings

Teaching and Learning Through Digital Literacies in Science

There was overwhelming agreement amongst the participants for incorporating digital literacies in all levels of science teaching. Reasons cited by TCs included i) engaging students, ii) fostering learner choice and learner control; iii) highlighting relevance between science and technology, iv) personalizing learning, and v) addressing 21st century trends, as illustrated below:

They [students] can do everything they were able to do before, but now they have another door to open and go through ... I taught viruses and the kids were not excited. I had them do Facebook profiles of viruses [like we did in class]... suddenly they were on the Internet and became engaged, interested. I think it's just from the standpoint that they are doing something that isn't usual ... difference is a good thing ... they appreciate the difference in learning and it is very helpful because students can find what they need to learn. This is something the education system is lacking ... by high school, students need to start exacting control over their learning.

So much of this world that we are in today is all about technology, and so, if students aren't learning about technology, they are not really keeping up, so as teachers we have to adapt our teaching techniques to what's going on in the world. If we're going more towards the technology based world, we have to prepare our students to be successful in that world ... students will appreciate it and they will learn different things. If we can teach our students about technology and how to use it,

that would be beneficial for them for whatever they do in their lives.

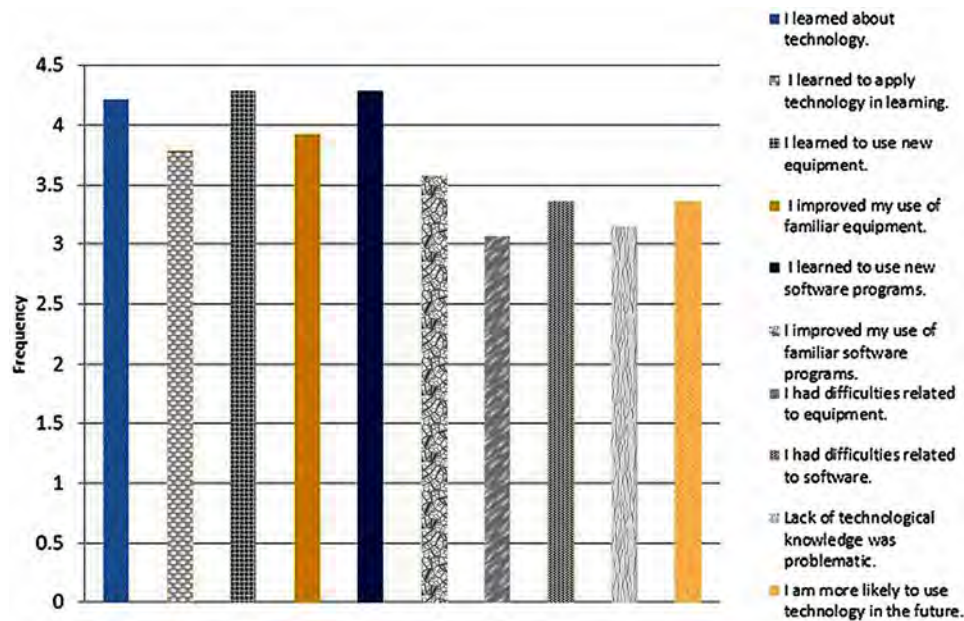
These findings are supported by Bolstad et al., (2012) and Cox (2008) who maintain that the teacher's role must be reconceived when implementing digital literacies in order to engage learners, their interests, experiences, and knowledge in many decisions about their learning.

Enhancing Teacher Candidates' Technological Literacy and Self-Efficacy

Assignments in the science methods course incorporated digital literacies on a variety of levels. In terms of developing technological knowledge and skills, participants reported that the course assignments unquestionably impacted their technological literacy and self-efficacy (Skaalvik & Skaalvik, 2008) in a positive manner. Figure 1 illustrates TCs' responses to survey statements related to developing their technological literacy. It is evident that TCs felt that they learned about technology, how to use new equipment, and how to use new software programs or improve their usage of familiar programs. Some TCs experienced challenges, however, this did not dissuade them from seeking out peer and instructor support to become familiar with the equipment and programs and how to use them effectively in order to complete their assignments. One noteworthy finding is the fact that TCs felt that they learned to apply technology in teaching and learning. Furthermore, there was a positive trend towards using technology in their future practices. TCs' survey responses demonstrated heightened technological literacy in terms of digital competence (skills), digital usage (professional/discipline application), and digital transformation (innovation/creativity) (Belshaw, 2011).

TCs also reported affective outcomes including enhanced motivation, engagement, confidence, interest, and pride in their experiences with digital literacies in the science methods course. They felt confident in their own ability to plan, organize, and

Figure 1. Teacher candidates' responses to survey statements exploring technological literacy



carry out activities incorporating digital literacies to attain teaching and learning goals in science, as reported by TCs:

It just allowed me to see what the world of digital technology is actually like; it's complicated, but it's also not as intimidating as you'd think. I didn't know that much was available and out there and from the standpoint of understanding ... like actually being able to see what makes a good game for learning, and what makes a good digital application, not just for engagement, but for learning as well, and just seeing, understanding in myself that it's not difficult to recognize that sort of a thing, and I think just from having to go through so many concepts and trying to find resources and just even building a website, I think I can quickly find things better for my students ... now I know what I'm looking for, and I know what's available.

I learned a lot about different digital resources that I could use in my class, and I used a lot of them during my last practicum. So after being able to do my timeline and learning about Prezi, I was

always implementing it in the classroom, ... it was really useful as I learnt a lot of different software, a lot of digital media, and I became more literate. I feel very confident. I used it a lot. I was always showing them videos, playing games ...

Curriculum Development and Instructional Design

The inclusion of digital literacies offered TCs: 1) comprehensive and flexible resources in terms of design and accessibility; 2) organized, professional and aesthetically pleasing outcomes; 3) innovative means to differentiate and personalize instruction; and 4) a repertoire of digital teaching/learning tools. One TC commented,

It is very engaging, and kids nowadays are always on the computer anyway, it is a way to engage them, hook them in and get them ready. The technologies we explored in the science course were really exciting and I think the kids would really like that ... having them use Prezi, when they're presenting will really get them involved and interested. You always have those artsy kids that find

science very boring, and I think that would be a really good way to have them use their creativity to learn about science.

TCs also felt that digital literacies provide the affordance to personalize learning, as summed up by a participant:

A lot of the students I found were visual learners, so I think it benefits them most. They definitely get to see it, hear about the concepts, and if they have any questions they can see it again, and so basically, it's through visual learning that they benefit. Also, it's not just memorizing what is written in a textbook. They understand the concepts and they know how to explain them in their own words.

The sharing of multiple, varied representations of concepts helped TCs grasp complex material by showing them alternative forms of the same underlying idea (Rose & Gravel, 2012).

Challenges/Constraints to Implementing Digital Literacies

TCs cited a number of potential challenges to implementing digital literacies in the classroom, including the fact that teachers need to be trained in order to effectively utilize digital literacies and achieve success in terms of developing learning communities in the classroom. Several participants felt that digital literacies should be introduced in the lower grades, thus allowing students to become “more comfortable and confident when using these technologies” as students who are not “tech savvy” might be resistant to learning through “new” technologies. Another challenge is the variation in choice of digital literacies available; some teachers felt that “their choice of representation” did not always coincide with what was being promoted in their host school. Finally, access to computers was cited by participants as the greatest obstacle to implementing digital literacies.

FUTURE RESEARCH DIRECTIONS

Despite the fact that using digital technologies as part of the educational environment fits into the philosophy of active learning and constructivism, it poses a tremendous challenge for teaching and learning in different contexts, specifically those that may not have access or experience a “digital divide”. Consequently, a digital divide decreases an individual’s access to social opportunity. Thus, research that promote deeper understandings of where knowledge gaps exist in the research base around the impact of digital technology on teaching and learning is warranted. Findings, including gaps and barriers, will make significant contributions to our understanding of socially informed digital literacy and can inform educators, researchers, and policy makers as to benefits and challenges of implementing digital and technology based projects in different contexts.

CONCLUSION

The research study conducted with TCs reveal that integrating digital literacies in science teacher education can potentially impact TCs’ digital competence, digital usage, and digital transformation (Belshaw, 2011). Findings indicate TCs experienced enhanced technological literacy in terms of learning about technology, software programs, and equipment. TCs also reported affective outcomes including enhanced motivation, heightened self-efficacy, increased interest, confidence and pride in their experiences with digital literacies in the science methods course. Digital usage was evident in their course work and some TCs reported that they had implemented various technologies in their practicum. In terms of teaching and learning, digital usage was at the forefront as it introduced personalized learning pedagogy (Dede, 2014); promoted engagement; fostered learner choice and control; promoted 21st century learning skills; and highlighted the relevance and application of digital literacies in teaching and learning science.

Digital transformation, including innovation and creativity, was prevalent in their course work. The inclusion of digital literacies offered TCs i) the opportunity to design and create comprehensive and flexible curricular resources; ii) organized, professional and aesthetically pleasing outcomes; iii) innovative means to differentiate instruction; and iv) ample opportunities for knowledge construction resulting in a repertoire of digital teaching/learning tools such as digital timelines and multi-media interactive website resources.

In this study, TCs embraced a constructivist view and engaged in a knowledge construction process as they developed a learning community (Scardamalia & Bereiter, 1994); one that promoted, modeled, and supported digital literacies. Through the course assignments and practicum experiences, TCs were provided opportunities to locate and consume digital content, create digital content, and communicate digital content (Spires et al., 2012). Despite the successes, a number of challenges were identified in terms of implementing digital literacies in the classroom, including poor teacher preparation, a finding corroborated by Dipetta and Woloshyn (2009) who maintain that without appropriate training, educators will continue to use technology as a supplement to instruction rather than a tool to facilitate instruction. In addition, some TCs felt that there was disconnect between pedagogy related to digital literacies that they promoted and implemented during their practicum and digital competencies of host teachers, a challenge raised by Jewitt (2008) around knowledge construction thus affecting the development of learning communities. Finally, TCs felt that access to computers is the greatest barrier, a finding consistent with research (OECD, 2013), which maintain that despite the ubiquity to computers in the home, student access to computers at school is still limited.

By the end of the study, participants possessed a good understanding of digital literacies and ways to effectively incorporate them in their future teaching practices. There was evidence that TCs utilized digital literacies learnt in the

course during their practicum. Results indicate that the explicit integration of digital literacies created and engaged learning communities, while improving technological and scientific literacies in a purposeful manner. Results may contribute to the information available to educators and curriculum consultants about developing learning communities through the integration of digital literacies in teacher education through activities (e.g., digital timelines, websites, etc.); products of personal creativity that vary according to students' unique learning styles. Finally, the provision and use of digital literacies may assist TCs in helping students develop 21st century skills, enhance teachers' and students' scientific and technological literacy, and improve attitudes toward teaching and learning science in technology enriched environments.

REFERENCES

- Al-Alwani, A. (2005). *Barriers to Integrating Information Technology in Saudi Arabia Science Education* (Unpublished doctoral dissertation). University of Kansas.
- Albirini, A. A. (2006). Teachers attitudes toward information and communication technologies: The case of Syrian EFL teachers. *Journal of Computers and Education*, 47(4), 373–398. doi:10.1016/j.compedu.2004.10.013
- Annetta, L. A. (2008). Video games in education: Why they should be used and how they are being used. *Theory into Practice*, 47(3), 229–239. doi:10.1080/00405840802153940
- Belshaw, D. (2011). *The Essential Elements of Digital Literacies*. Retrieved from <http://doug-belshaw.com/ebooks/digilit/>
- Bennett, F. (2002). The future of computer technology in K-12 education. *Phi Delta Kappan*, 83(8), 621–626. doi:10.1177/003172170208300811
- Bereiter, C. (2002). *Education and mind in the Knowledge Age*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

- Bolstad, R., & Gilbert, J. (2006). *Creating digital age learners through school ICT projects: What can the Tech Angels project teach us? Discussion paper prepared for the Ministry of Education*. Wellington: Ministry of Education.
- Bolstad, T., Gilbert, J., McDowall, S., Bull, A., Boyd, S., & Hipkins, R. (2012). *Supporting future-oriented learning and teaching. A New Zealand perspective*. Ministry of Education.
- Brush, T., & Bannon, S. (1998). Characteristics of technology leaders: A survey of school administration in the United States. *International Studies in Educational Administration*, 26(2), 47–56.
- Calvani, A., Cartelli, A., Fini, A., & Ranieri, M. (2008). Models and Instruments for Assessing Digital Competence at School. *Journal of e-Learning and Knowledge Society*, 4(3), 183–193.
- Canadians for 21st Century Learning and Innovation (C21 Canada). (2012). *Shifting Minds: A 21st Century Vision of Public Education for Canada*. Author.
- Castells, M. (2009). *Communication Power*. Oxford, UK: Oxford University Press.
- Chai, C. S., Koh, J. H. L., & Tsai, C.-C. (2010). Facilitating pre-service teachers' development of technological, pedagogical, and content knowledge (TPACK). *Journal of Educational Technology & Society*, 13(4), 63–73.
- Cope, B., & Kalantzis, M. (2000). Introduction: Multiliteracies. In B. Cope & M. Kalantzis (Eds.), *Multiliteracies: Literacy learning and the design of social futures*. London: Routledge.
- Cox, T. (2008). Learning Styles and Students' Attitudes toward the Use of Technology in Higher and Adult Education Classes. *Institute for Learning Styles Journal*, 1, 1–13.
- Dagiene, V. (2011). The development of ICT competency in pre-service teacher education. *International Journal of Digital Literacy and Digital Competence*, 2(2), 1–10. doi:10.4018/jdlldc.2011040101
- DeCoito, I. (2014). Teaching about the nature of science through digital scientific timelines. In P. D. Morrell & K. Popejoy (Eds.), *A few of our favorite things: Teaching ideas for K-12 science methods instructors*. Sense Publishers.
- DeCoito, I., & Richardson, T. (2016). Focusing on Integrated STEM Concepts in a Digital Game. In M. Urban & D. Falvo (Eds.), *Improving K-12 STEM Education* (pp. 1–23). IGI Global.
- Dede, C. (2014). *The Role of Digital Technologies in Deeper Learning. Students at the Center: Deeper Learning Research Series*. Boston, MA: Jobs for the Future.
- Dilworth, P., Donaldson, A., George, M., Knezek, D., Searson, M., Starkweather, K., & Robinson, S. et al. (2012). Preparing teachers for tomorrow's technologies. *TeachTrends*, 56(4), 11–14. doi:10.1007/s11528-012-0581-5
- DiPetta, T., & Woloshyn, V. E. (2009). Exploring teacher-candidates' experiences, beliefs and to technology as an instructional learning tool following instruction in a “technology rich” classroom. *Journal of Educational Technology*, 6(1), 54–64.
- Ertmer, P., Ottenbreit-Leftwich, A., Sadik, O., Sendurur, E., & Sendurur, P. (2012). Teacher beliefs and technology integration practices: A critical relationship. *Computers & Education*, 59(2), 423–435. doi:10.1016/j.compedu.2012.02.001
- Gomes, C. (2005). Integration of ICT in Science Teaching: A Study Performed in Azores, Portugal. *Recent Research Development in Learning Technologies*, 47(4), 47–61.
- Graham, C. R. (2011). Theoretical considerations for understanding technological pedagogical content knowledge (TPACK). *Computers & Education*, 57(3), 1953–1960. doi:10.1016/j.compedu.2011.04.010
- Grant, L. (2010). *Connecting digital literacy between home and school*. Bristol: Futurelab.

- Greenhow, C., Robelia, E., & Hughes, J. (2009). Research on Learning and Teaching with Web 2.0: Bridging Conversations. *Educational Researcher*, 38(4), 280–283. doi:10.3102/0013189X09336675
- Hollingsworth, S. (1989). Prior beliefs and cognitive change in learning to teach. *American Educational Research Journal*, 26(2), 160–189. doi:10.3102/00028312026002160
- Jewitt, C. (2008). Multimodality and Literacy in School Classrooms. *Review of Research in Education*, 32(241), 241–267. doi:10.3102/0091732X07310586
- Lemke, J. (2004). *Why study digital gameworlds? Notes toward a basic research agenda for learning technologies*. Retrieved from <http://www-personal.umich.edu/~jaylemke/games.htm>
- Levy, F., & Murnane, R. J. (2004). *The New Division of Labor*. New York: Russell Sage. doi:10.1515/9781400845927
- Martin, A., & Grudziecki, J. (2006). DigEuLit: Concepts and Tools for Digital Literacy Development. University of Glasgow. Retrieved from <http://ics.heacademy.ac.uk/italics/vol5iss4/martin-grudziecki.pdf>
- Mills, J. M., Durepos, G., & Wiebe, E. (2010). *Encyclopedia of Case Study Research*. Thousand Oaks, CA: Sage Publications Inc. doi:10.4135/9781412957397
- Milton, P. (2015). *Shifting Minds 3.0: Redefining the Learning Landscape in Canada*. C21 Canada.
- Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, 108(6), 1017–1054. doi:10.1111/j.1467-9620.2006.00684.x
- National Research Council (NRC). (2011). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. Washington, DC: The National Academies Press.
- Ng, W. (2010). Empowering students to be scientifically literate through digital literacy. In S. Rodrigues (Ed.), *Multiple literacy and science education: ICTs in formal and informal learning environments* (pp. 11–31). Hershey, PA: Global Publishing. doi:10.4018/978-1-61520-690-2.ch002
- Ng, W. (2011). Why digital literacy is important for science teaching and learning. *Teaching Science*, 57(4), 26–32.
- OECD. (2013). *Trends Shaping Education 2013*. OECD Publishing. doi:10.1787/trends_edu-2013-en
- Osborne, J. F., & Collins, S. (2000). *Pupils' and parents' views of the school science curriculum*. London: King's College London.
- Rose, D. H., & Gravel, J. W. (2012). *Curricular Opportunities in the Digital Age: The Students at the Center Series*. Boston, MA: Jobs for the Future.
- Sancar Tokmak, H., Sürmeli, H., & Ozgelen, S. (2014). Pre-service science teachers' perceptions of their TPACK development after creating digital stories. *International Journal of Environmental and Science Education*, 9(3), 247–264.
- Sasseville, B. (2004). Integrating Information and Communication Technology in classroom: A comparative discourse analysis. *Canadian Journal of Learning and Technology*, 30(2), 5–27.
- Scardamalia, M., & Bereiter, C. (1994). Computer support for knowledge-building communities. *Journal of the Learning Sciences*, 3(3), 265–283. doi:10.1207/s15327809jls0303_3
- Shulman, L. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–15. doi:10.3102/0013189X015002004
- Skaalvik, E. M., & Skaalvik, S. (2008). Teacher self-efficacy: conceptual analysis and relations with teacher burnout and perceived school context. In R. Craven, H. W. Marsh, & D. McInerney (Eds.), *Self-processes, learning, and enabling human potential* (pp. 223–247). Information Age Publishing.

Spires, H. A., Wiebe, E., Young, C. A., Hollebrands, K., & Lee, J. K. (2012). Toward a new learning ecology: Professional development for teachers in 1:1 learning environments. *Contemporary Issues in Technology & Teacher Education*, 12(2), 232–254.

Stake, R. (2000). Case Studies. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of Qualitative research* (pp. 435–454). Thousand Oaks, CA: Sage.

Webb, M. (2005). Affordances of ICT in science learning: Implications for an integrated pedagogy. *International Journal of Science Education*, 27(6), 705–735. doi:10.1080/09500690500038520

Zhang, Z., & Martinovic, D. (2008). ICT in teacher education: Examining needs, expectations and attitudes. *The Canadian Journal of Learning and Technology*, 34(2), 149–166.

ADDITIONAL READING

Clark, D., Nelson, B., Sengupta, P., & D'Angelo, C. (2009). *Rethinking science learning through digital games and simulations: Genres, examples, and evidence*. Washington, DC: National Research Council.

de Jong, T. (2006). Technological advances in inquiry learning. *Science*, 312(5773), 532–533. doi:10.1126/science.1127750 PMID:16645080

DeCoito, I. (2012). Digital Games in Science Education: Developing Students' 21st Century Learning Skills. In Z. Karadag and Y. Devocioglu-Kaymakci (Eds.), *Proceedings of the International Dynamic, Explorative, and Active Learning (IDEAL) Conference* (pp. 159–172). Turkey: Bayburt University. ISBN: 978-605-61893-4-0.

Dede, C. (2009). Immersive interfaces for engagement and learning. *Science*, 323(5910), 66–69. doi:10.1126/science.1167311 PMID:19119219

Dostál, J. (2009). Educational software and computer games – tools of modern education. *Journal of Technology and Information Education*, (1)1.

Hayes, E. R., & King, E. M. (2009). Not just a dollhouse: What *Sims2* can teach us about women IT learning. *On The Horizon*, 17(1), 60–69. doi:10.1108/10748120910936153

Hennessy, S., Wishart, J., Whitelock, D., Deane, R., Brawn, R., Velle, L., & Winterbottom, M. et al. (2007). Pedagogical approaches for technology-integrated science teaching. *Computers & Education*, 48(1), 137–152. doi:10.1016/j.compedu.2006.02.004

Koç, M. (2005). Implications of Learning Theories for Effective Technology Integration and Pre-service Teacher Training: A Critical Literature Review. *Journal of Turkish Science Education*, 1(2), 1–18.

Lee, Y. F., Guo, Y., & Ho, H. J. (2008). Explore Effective Use of Computer Simulations for Physics Education. *Journal of Computers in Mathematics and Science Teaching*, 27(4), 443–466.

Martin, A. (2005). DigEuLit – a European Framework for Digital Literacy: a Progress Report. *Journal of eLiteracy*, 2, 130–136.

Olsson, L., & Edman-Ståhlbrand, E. (2008). Digital literacy as a challenge for Teacher Education. In M. Kendall & B. Samways (Eds.), *Learning to Live in the Knowledge Society* (pp. 11–18). Boston: Springer. doi:10.1007/978-0-387-09729-9_2

Prain, V., & Waldrip, B. (2006). An exploratory study of teachers and students use of multimodal representation of concepts in primary science. *International Journal of Science Education*, 28(15), 1843–1866. doi:10.1080/09500690600718294

Rutten, N., Joolingen, W. R., & Veen, J. T. (2012). The learning effects of computer simulations in science education. *Computers & Education*, 58(1), 136–153. doi:10.1016/j.compedu.2011.07.017

Smillie-Adjarkwa, C. (2005). *Is the Internet A Useful Resource For Indigenous Women Living In Remote Communities In Canada, Australia and New Zealand To Access Health Resources?* Retrieved from <http://www.yorku.ca/anthna/DigitalDivide.pdf>

Wegener, M., McIntyre, T. J., McGrath, D., Savage, C. M., & Williamson, M. (2012). Developing a virtual physics world. *Australasian Journal of Educational Technology*, 28(3), 504–521. doi:10.14742/ajet.847

Winsberg, E. (2010). *Science in the age of computer simulation*. The University of Chicago Press. doi:10.7208/chicago/9780226902050.001.0001

Yeo, S., Loss, R., Zadnik, M., Harrison, A., & Treagust, D. (2004). What do students really learn from interactive multimedia? A physics case study. *American Journal of Physics*, 72(10), 1351–1358. doi:10.1119/1.1748074

KEY TERMS AND DEFINITIONS

C

Digital Divide: The gap between those with ready access to tools of ICT, and to the knowledge that they provide access to, and those without such access or skills.

Digital Literacy: A combination of digital competence (skills, concepts, approaches, attitudes), digital usage (professional/discipline application) and digital transformation (innovation/creativity).

Prezi: A cloud based presentation software that allows individuals to collaborate and edit presentations with multiple users. It allows students to construct and present their knowledge in different learning styles (videos, audio, etc.).

Socioscientific Issues: (SSI): Open-ended problems which are controversial social issues related to science and highlights the application of moral and scientific reasoning to real world situations.

Tiki Toki: Web-based software that can be used to create interactive timelines that can be shared on the Internet.

Designing Engaging Instruction for the Adult Learners

Karen Swanson

Mercer University, USA

Geri Collins

Mercer University, USA

INTRODUCTION

Instruction in higher education continues to follow a traditional model of instructor delivery and content focus. In a study by Nunn (1996), he found that less than six percent of class time involves student participation. There was no evidence of any instance in which students spoke for more than 23% of the total class time. A teacher-centered instructional method has received significant rebuke in the last decade, the stereotypical image is one of instructors droning on to a captive audience of undergraduates taking feverish notes for the sole purpose of reciting the information on a subsequent test. Hoyt and Perera (2000) surveyed faculty as to which type of instruction approach they incorporated into their practice. Forty-five percent identify some combination which used lecture as a primary approach. So the question then becomes what are the best practices for an engaging course design that benefits both faculty and adult learners?

Instructional design for the adult learner is a growing field of study in higher education. Engaging instruction for adult learners will be defined in this chapter in two ways: designing courses using the significant learning taxonomy, and a paradigm shift to support faculty to involve student participation.

Adult education is defined as “activities intentionally designed for the purpose of bringing about learning among those whose age, social

roles, or self-perceptions, define them as adults” (Merriam & Brockett, 2014, p. 8). The adult learner can also include those who are parents, working full time or older and more experienced than the traditional college student. Adults are motivated by learning that has a practical application and personal significance. In addition, the growing abundance of information makes covering a discipline almost impossible. So the alternative or anecdote to teaching a volume of information is to design the learning for significance in a complex and rich way. Here again the need to address student motivation becomes an issue to examine.

BACKGROUND

When planning instruction for adult learners it is important to examine their motivation for being in the classroom. They have chosen to pursue a higher education degree and the faculty’s challenge is to plan and execute engaging, relevant learning experiences. The traditional model of instruction in higher education, such as lecture and testing, while possibly engaging, is limited. If the goal is to be able to explicitly assess how students are making the connections among theory, content, and personal experience there needs to be a master course design. Fink (2007) explains that “our current instructional procedures are not working very well. Students are not learning even basic general knowledge, they are not developing higher-level

cognitive skills, and they are not retaining their knowledge very well. In fact there is no significant difference between student who take courses and students who do not” (Fink, 2013, p. 4).

Learner-Centered Instruction

Teaching adult learners should be focused on the learning process of the students. Teacher-directed instruction, such as lecture and rote memorization can be transformed to be more engaging through activities such as discussion and application.

Huba and Freed (2000) contrast teacher and learner-centered instruction as a model or paradigm. Their research shifts the focus from what faculty teach (content) to how students learn. They acknowledge that a teacher-centered methodology is not ineffective, but a shift from solely lecturing to a teaching style that incorporates more student interaction and a demonstration of learning beyond rote memorization which supports long-term retention. Huba and Freed (2000) characterize a teacher-centered paradigm in several ways:

1. Content is primarily delivered by the instructor and students are solely learners.
2. Content is not contextualized but rather students passively receive the information.
3. Assessment is the responsibility of the instructor, requires only rote memorization and is summative in nature.
4. Learning is the responsibility of the individual and courses can be constructed in a competitive nature (p. 5).

These characteristics may be more evident in large, lecture style courses where individual interaction is more cumbersome. However, the challenge is for faculty to begin to consider a possible paradigm shift towards learning centered instruction. Fink (2013) states, “...a major change already taking place in American higher education.....is a paradigm shift in which institutions are thinking less about providing instruction (the teaching paradigm) and more about producing

learning (the learning paradigm)” (p. 20). The following sections will explore this paradigm shift, how it impacts instructional design and how to engage students in this new perspective of learning.

Three Models: Transmission, Generative, and Transformative

As faculty, we design courses with a myriad of assumptions about teaching and learning. Those assumptions influence the decisions we make about how course time is used, assignments are structured and assessments are designed. Therefore, a theoretical background may provide a framework to place assumptions into a context for examination. Some assumptions that may be present include the student’s lack of prior content knowledge; more content is equivalent to increased rigor and time constraints (it’s faster to lecture). It may be assumed that teacher directed instruction requires less planning over time and is the format most appropriate in higher education. The objective of this section is to provide three general models to examine how moving towards learning centered instruction can produce rich educational experiences for the students.

Wink (2000) identifies three perspective models on teaching: the transmission model, the generative model and the transformative model. The transmission model has instructors filling students with information through primarily lecture and the students passively receiving information. The instructor controls the scope and sequence of information and evaluation is typically an examination which requires students to repeat information in the form they were given it (p. 121).

Researchers have come to know that students must construct knowledge from information which defines the generative model of teaching. This model provides a space for students to engage in the learning using groups to build on their own knowledge. Learning is no longer passive. The role of the instructor is to structure the course activities to guide students toward learning outcomes (p. 122).

The transformative model is a learner-centered classroom where students actively engage in determining the curriculum, activities and assessment of their understanding. The focus becomes application of knowledge to solving real-life issues (Wink, 2000, pp. 122-123). The examination of these models shows the paradigm shift toward learning centered instruction and challenges higher education instructors to find ways to engage students in the whole learning process.

A balanced classroom ebbs and flows between faculty delivery of information delivery and the students' application or demonstration of understanding. Palmer (2007) writes that there is a need to —rebalance the scales between an overemphasis on teaching technique and the value of the emotion of learning (p. 64). Palmer remarks that it behooves instructors to balance the rhetoric, to not hold too tightly to either possible polar position, but to use the best of each to enhance student learning. Palmer presents pedagogical attributes to creating an open learning environment.

1. The space should be bounded and open. Bounded refers to using a question, text or body of data to focus the class. The information should be very clear and compelling to attract student attention. Space is achieved through providing room for surprises, tangents and possibility to arrive at a destination not previously determined by the instructor.
2. The space should be hospitable and —charged|. Is this class a safe place to express my ideas? Is it hospitable? The charged element comes in the process of delving into compelling topics in meaningful ways where students can't marginalize or avoid the depths of new understandings.
3. The space should invite the voice of the individual and the voice of the group. Students should be expected to express their ideas regardless of how polished the presentation. True learning comes when students place their individual ideas in the company of

others. It is then that they are molded and stretched by the voice of the group.

4. The space should honor the little stories of the students and the big stories of the disciplines and tradition. This refers to the internal and external discussion between the individual ideas and the tradition of the discipline. If the individual is too present the discipline is lost, if the discipline is too present the process of internalization is compromised.
5. The space should support solitude and surround it with the resources of community. As with Baxter Magolda and Kings (2004) work on self-authorship in supporting students to become independent learning. Class time must incorporate time for students to consider material individually and also grapple with content interpersonally with their peers to make meaning.
6. The space should welcome both silence and speech. Silence can make instructors nervous. But silence is also a sign of reflection that is critical in the learning process. Instructors can specifically incorporate reflection points during or after class. (pp. 77- 79)

The image that emerges from Palmer's work shows instruction that spurs critical thinking as students become engaged in the process of the delivery of content. Again focusing on the learning aspect of instruction which includes both questioning and discussion topics. Although the instructor maybe in the front of the class leading this discourse, learning is still at the center of the lesson. Course design done well can create dynamic learning as well as promote the creation of independent learners.

Self-Authorship

The structure of a course can purposefully address each of these domains to expand the independence of students. The prior knowledge students bring to a course can be incorporated into the presenta-

tion of new information from a discipline. Done purposefully, the instructor can facilitate the acquisition and situation of new knowledge. They can also minimize the resistance by students to consider new perspectives regarding the subject matter. Individual and group assignments can be made for both during and outside of class time to create opportunities for students to grapple with new content. One theoretical framework to assist instructors to visualize how students learn and develop is Baxter Magolda and King's (2004) self-authorship model. They describe the process of cultivating sophisticated learners.

Educators try to make learning to fly as safe as possible, yet a certain amount of crashing is to be expected. Balancing the degree of risk during college with the necessity of expertise for adult life is a contrast learning curve in which educators sometimes crash and sometimes soar. (p. xvii)

Keegan (1994) defines self-authorship “as internally coordinating beliefs, values and interpersonal loyalties rather than depending on external values beliefs and interpersonal loyalties” (in Baxter Magolda & King, 2004, p. xviii). Briefly, they identify three domains in which students’ process information. The first is an epistemological foundation in which knowing is contextual built on a belief system that continuously evaluates, interprets and judges new information based on their prior knowledge. The epistemological domain wrestles with new information and the delivery person. Does the information and delivery blend with their current moral and intellectual conceptions or does it create a conflict?

The second domain, intrapersonal foundation, is how people view themselves and create an identity based on what they value and experience. This is the domain of individual reflection on learning which supports students making new connections. Examples of supporting coursework would be journaling, creating graphic organizers of their understanding of connections, or reading the text using marginalia to create dialogue with the author.

The third is the interpersonal domain where students place their identity or knowledge in context with other students and instructors (Baxter Magolda & King, 2004, pp. 8-9). The interpersonal domain provides a learning centered instructional model where information or ideas are presented by students working with others to make sense of that information as a group. For example, study groups or group projects would require students to determine how their ideas fit with others and promote a creation of new understandings that would not have been possible individually. Therefore the bridge between supporting adult learners to develop self-authoring behaviors is embedded in creating significant learning experiences.

The bridge between tradition and innovation may be found in learning outcomes that promote self-authoring. Fink (2013) has created a course design structure called significant learning experiences. Significant learning experiences are defined as teaching that results in “a learning experience resulting in something that is truly significant in terms of the students’ lives” (p. 6). The goal is to focus on process and outcomes that result in the following four characteristics:

Process

1. **Engaged:** Students are engaged in their learning.
2. **High Energy:** Class has a high energy level.

Results, Impact, Outcomes

3. **Significant and Lasting Change:** Course results in significant changes in the students, changes that continue after the course is over and even after the students graduate.
4. **Value in Life:** What the students learn has a high potential for being of value in their lives after the course is over, by enhancing their individual lives, preparing them to participate in multiple communities, or preparing them for the world of work. (Fink, 2013, p. 8)

The transition from the characteristics of significant learning experiences is to evaluate or section a syllabus or course design into three types of learning. The first is the *delivery* of content to provide a foundation for subsequent higher order activities. Content can be obtained through readings, lecture, or online modules to name a few. These activities can incorporate both teacher-directed and learner-centered methods. The second section is to outline the *process* elements. These would include presentations, designing prototypes, community experiences, and other similar authentic experiences. The final section is for students to *evaluate and reflect* on their learning experience and for faculty to assess their work (Fink, 2007). The scaffolding of student learning and course construction is vital to engaged student participation and readiness for class.

Fink (2013) uses a Taxonomy of Significant Learning to address areas instructors may find useful as they design instruction that motivates learning:

1. Foundational Knowledge provides the basic understanding of broad subjects that is necessary to create knowledge for learning in multiple content areas. For example, an understanding of historical events such as the Revolutionary War, Civil War, Civil Rights, etc.
2. Application allows other kinds of learning to become useful for example in the use of knowledge to perform a task. Application is evidenced in critical thinking and problem-solving.
3. Integration is making connections between and among concepts or ideas.
4. Human Dimension informs students about the human significance of what they are learning. This is similar to Baxter Magolda and King's (2004) self-authors model that empowers the student to alter their internal conversation and what they know about themselves.
5. Caring. When students care about something, they then have the energy they need for learning more about it and making it a part of their lives. Without energy for learning, nothing significant happens.
6. Learning how to learn is when student understand the strategies of how to learn independently such as reading strategies and writing patterns. It enables students to continue learning in the future and to do so with greater effectiveness (pp. 39-40).

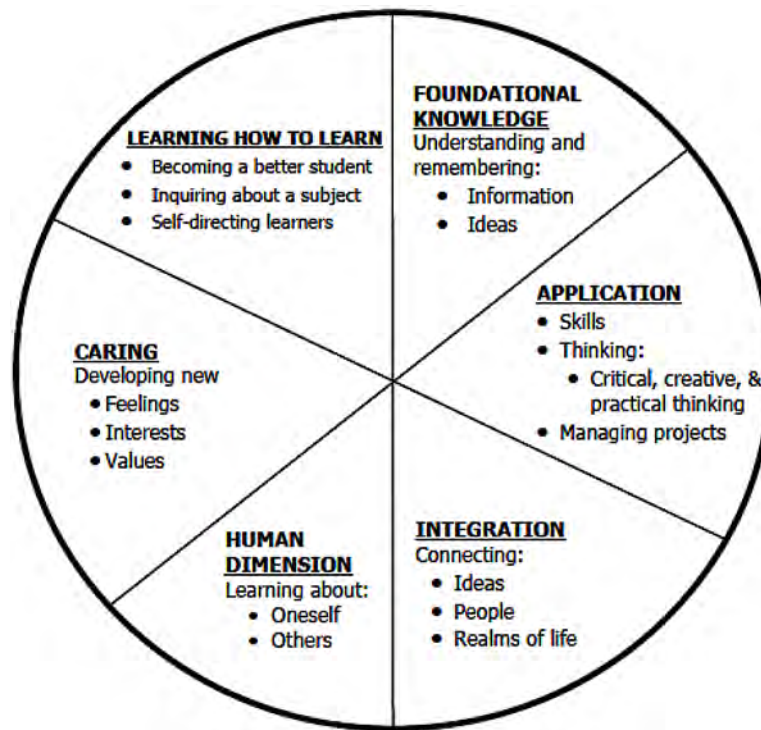
These areas become the foundation for learning centered instruction and the underlying theme is motivating students in the classrooms.

Instructional Design

Good teachers possess a capacity for connectedness. They are able to weave a complex web of connections among themselves, their subjects, and their students so that students can learn to weave a world for themselves. The methods used may vary widely: discussion, questioning lectures, Socratic dialogues, laboratory experiments, collaborative problem solving, and creative chaos. The connections made by good teachers are held not in their methods but in their hearts “meaning *heart* in the ancient sense, as the place where intellect and emotion and spirit and will converge in the human self” (Palmer, 2007, p.11).

In a powerful learning experience, students will be engaged in their own learning, there will be “high energy level associated with it, and the whole process will have important outcomes or results (Fink, 2013, p. 8). The idea of the faculty/student relationship is important to explore. It is important to connect with students in order to have them want to engage in the process of learning. The connection must be professional and intellectual. It is important to always relate the knowledge without being the end all on any given subject. Allow students to disagree with what is stated and then discuss the differences that are in question. This gives students the ability to connect

Figure 1. Fink's significant learning experience taxonomy (2013, p. 35)



in a positive way and learn how to view different perspectives. If students are invited to take the risk of speaking their mind, teachers must be able to guide the discussion without being intimidated by the student's perspective. This two-way street gives the attitude of professionalism and intellectualism to the classroom environment. Students learn how to use their voice in a professional way and learn from the acceptance of the teacher. "Both artistry and technical skills are required as you develop presentation techniques that are well fitted to your unique instruction circumstances" (Savage, Savage & Armstrong, 2006, p. 242).

Once more, a level of rapport must be developed between the teacher and the students. Classroom with this kind of instruction become active learning environments. Content can either be dry and boring or exciting and interesting. Humor is a good way to engage students and can lead to understanding if connections are made even if they are funny. "Enthusiasm is contagious. If you express a high level of interest and sense of importance about

your topic, students often become spellbound, anxious to find out what so interesting" (Moore, 2009, p.150). The whole environment of the classroom revolves around how well the teacher can ignite interest in the subject and help the students to become a part. Adding points of interest that capture the meaning of a concept gives student a way to remember what needs to be learned. The demeanor of the instructor during lecture is important as well. The speaker should smile, make eye contact with student when they ask questions and move around the room to better include all students (Sullivan & McIntosh, 1996, p.7). Moving about the classroom also adds a physical connection that keeps students aware that something is actually happening in the classroom.

Student Engagement

The shift for higher education faculty is to plan with the end in mind. Faculty redesign their delivery to engage students with them in the process

of learning rigorous content. It brings the words being spoken to life and relays the importance of what students need to know. A well-prepared class must move beyond the required reading, lecture and test model. The questioning and discussion regarding the reading can be ways for students to give their perspective. Student reflection can move the lecture to more depth as analysis is used to present the information. As lectures evolve, students begin to understand the broader aspects of what they need to understand about any given subject (Brookfield, 2005). It is also important to acknowledge the students contribution to the lecture. Moore (2009) refers to this as reinforcement, "...rather than give reinforcement early in the questioning sequence, you should allow as many students as possible to respond to the question then reinforce them for their contribution" (p.163).

Two methods to create accountability for students to come to class prepared to discuss readings include student created reading cards and reading panels. The reading cards are 3x5 cards which students prepare notes using the assigned readings for week. The cards are turned in the first five minutes of class, usually left at the door in a basket. The instructor puts a special stamp on those turned in on time (late comers loose out) and students can use stamped cards on the exams. Another strategy is to create a reading panel. Student names are put into a hat and five names are drawn each class. Those five students are responsible to be prepared to lead a discussion of the reading without the assistance of the instructor. Each week names are drawn, all the names are available each week, so that once a student participates they are not out of circulation hence stop reading.

The classroom should be a two way street, this means that students also have a responsibility for their learning by coming to class prepared. Faculty "must focus on changing the role of the student from passive observe to active participant. In this approach, the responsibility for meeting learning

objectives is shared by the instructor and each student" (Sullivan & McIntosh, 1996, p.5). Student preparation includes reading information and involves in-depth analysis of what is read which increases the student's ability to express their perspective in class. Students should not hesitate to raise questions; the interruption forces the instructor to engage with the students. If students take the challenge to insert themselves into the lecture with questions or comments they encourage others to join the discussion. Teachers must be open to let students become involved, but it is as important for students to be prepared as much as it is for the teacher.

Questioning

Questioning is one of the best ways to engage students. However, questions only engage the students if they lead to more discussion. Open-ended questions provide students the opportunity to think and give them ownership of the discussion. The use of focus questions can inform instructors what students know and create interest in the topic. Questions which "may be factual, empirical, productive, or evaluative are used to direct student attention" (Moore, 2009, p.158). Questioning is a skill that teachers need to acquire. Part of the technique is using wait time which is a purposeful delay to fill the quiet space created when students are forming a response. Instructor pacing generally allows for a one second wait time; however when wait time is increased to three to five seconds Moore (2009) found:

1. Student response time increased.
2. Failure to respond tends to decrease. Student asked more questions.
3. Unsolicited responses increased.
4. Student confidence increased (p.162).

Wait time allows student time to think and again, engages them in the lesson. The question needs to invoke the thinking process and lead

to more in depth discussion of the subject. One questioning technique that is highly used is the Socratic Method which Paul and Elder (2007) describe as:

disciplined questioning that can be used to pursue thought in many directions and for many purposes...The key to distinguishing it from other types of questioning is that Socratic questioning is systematic, disciplined and deep and usually focuses on foundational concepts, principles, theories, issues, or problems. (p. 36)

The objective is for discovery on the part of the students to understand the concept being examined, the idea being explored, the process being described or the content being explained. The faculty and the students become involved in a dialogue which allows the students to see different perspectives as they begin to form their own ideas based on a broader understanding of the elements being presented.

Lecture Recitation

Lecture recitation is another method that can be effective and also efficient. Moore (2009) points out the use of a lecture formed around questioning adds another dimension to the instruction. The time spent giving needed information should always be as engaging as any hands-on activity. Lecture recitation helps to invite students into a conversation. The faculty becomes the director of the concepts being discussed and leads the flow of information towards the desired content. Discovery is the purpose for students as they formulate their answers to open ended thought provoking questions. The whole class can participate as the instructor directs questions that pique the interest of each student. All answers must be accepted and used to continue the discussion which adds depth to the content being studied. Students are equally responsible to come to class prepared with readings and homework complete to best participate with their peers.

FUTURE RESEARCH DIRECTIONS

C

As previously stated, "...a major change already taking place in American higher education.....is a paradigm shift in which institutions are thinking less about providing instruction (the teaching paradigm) and more about producing learning (the learning paradigm)" (Fink, 2013, p. 20). The long-term outcomes for higher education appears to be shifting towards graduates who are critical thinkers, collaborative, flexible and creative. While content is essential for a well-rounded and informed person, it is the higher order thinker who will be successful to the future workplace.

This is a challenge for faculty who love their content and love being the expert. There will need to be a new way to matriculate new faculty to begin thinking more about learning. There must be professional development for current faculty to realize that the transformation of their syllabi, teaching and assessment is a process towards a new paradigm, rather than throwing the baby out with the bath water. Faculty enjoy seeing student reactions when they grasp a concept or a skill, the future of teaching and learning in higher education will be to design for those significant learning experiences from the beginning.

CONCLUSION

The need to examine how to provide engaging instruction for adult learners is the rationale behind this study. The investigation revealed adult learners' desire to obtain more in depth instruction which requires a paradigm shift towards learning centered pedagogy. The basic themes presented in the new paradigm addresses how to motivate the adult learners, how to design engaging institutions to enhance motivation and how to focus on learning centered instruction which invites students to participate in the learning process there by gaining ownership of the learning. This study gave suggestions for actual classrooms instruction as well as ways for instructors to entourage different ways

of learning. It was the intent to provide theoretical background for the paradigm shift and also way in which to implement different strategies to create learning centered instruction.

REFERENCES

- Baxter Magolda, M., & King, P. (2004). *Self-authorship as the common goal of the 21st-century education. In Learning partnerships. Theory and models of practice to educate for self-authorship.* Sterling, VA: Stylus.
- Brookfield, S. D., & Preskill, S. (2005). *Discussion as a way of teaching: Tools and techniques for democratic classrooms* (2nd ed.). San Francisco, CA: Jossey Bass.
- Fink, D. (2007). The power of course design to increase student engagement and learning. *Peer Review*, 9(1), 13–17.
- Fink, D. (2013). *Creating significant learning experience: An integrated approach to designing college courses.* San Francisco, CA: Jossey-Bass.
- Hoyt, D., & Perera, S. (2000). *Validity of the IDEA student ratings of instruction system: An update, IDEA research report #2.* Manhattan, KS: Kansas State University, Center for Faculty Evaluation and Development.
- Huba, M., & Freed, J. (2000). *Learner-centered assessment on college campus. Shifting the focus from teaching to learning.* Boston, MA: Allyn and Bacon.
- Kegan, R. (1994). *In over our heads. The mental complexity of modern life.* Cambridge, MA: Harvard Press.
- Merriam, S. B., & Bierema, L. L. (2014). *Adult learning. Linking theory and practice.* San Francisco, CA: Jossey Bass.
- Moore, K. (2009). *Strategies from theory to practice.* Thousand Oaks, CA: Sage.
- Nunn, C. E. (1996). Discussion in the college classroom: Triangulating observational survey results. *The Journal of Higher Education*, 67(3), 243–266. doi:10.2307/2943844
- Palmer, P. (2007). *The courage to teach. Exploring the inner landscape of a teacher's life.* San Francisco: Wiley & Sons.
- Paul, R., & Elder, L. (2007). Critical thinking: The art of Socratic questioning. *Journal of Developmental Education*, 31(1), 36–37.
- Savage, T., Savage, M., & Armstrong, D. (2006). *Teaching in the secondary school* (6th ed.). Upper Saddle River, NJ: Pearson/Merrill Prentice Hall.
- Sullivan, R. & McIntosh, N. (1996, December). *Delivering effective lectures.* Paper #5, JHPIECO Corporation: Johns Hopkins University.
- Wink, J. (2000). *Critical pedagogy. Notes from the real world.* New York: Addison Wesley Longman.

KEY TERMS AND DEFINITIONS

Adult Education: “Activities intentionally designed for the purpose of bringing about learning among those whose age, social roles, or self-perceptions, define them as adults.

Self-Authorship: Refers to a phase of development within the lifelong process of self-evolution.

Significant Learning Experiences: Defined as teaching that results in—a learning experience resulting in something that is truly significant in terms of the students’ lives.

Educational Ontology Development

C**Galip Kaya***Havelsan Inc., Turkey***Arif Altun***Hacettepe University, Turkey*

INTRODUCTION

Amount of data on the web is constantly growing day by day and it is a new skill to be developed for users to reach the information for their needs, tasks and/or goals among vast amount of information. Likewise, in educational domain, it is a big challenge for teachers and learners to find educational resources on the web and to be sure that these resources meet their requirements. Semantic web technologies, including ontology use, offer a solution for those problems to be solved within educational domain by using domain ontologies.

Ontology provides a framework to determine a domain of interest and to build a common understanding in this domain between different applications by representing concepts and relationships that exist between these concepts. Ontology use in education environments can be explained in three main groups: content access and/or retrieval, content creation, and personalization. Use of ontologies also helps content creators to design and develop online courses; provide smart searches and content suggestions; and, design personalized learning environments for learners. All these advantages come with a price, though: It is not always possible for educators to find the best ontology for their needs in their Learning Management System (LMS). Consequently, users generally have to create their own ontologies from scratch.

In developing an ontology, there is no single and/or standard method to follow that can be ap-

plied to all domains since they are widely differ across domains as in law, medicine, information technologies, education, military etc. Since the ontology creation is a complex process, it might not be always easy or even possible to create an ontology for a selected domain and use it across domains.

To address these concerns, in this paper, first, a review of state-of-the art literature regarding the ontologies in educational domain will be reviewed. Second, the challenges and difficulties in ontology development process for educational domain will be addressed and explained in detail. Finally, design suggestions to the difficulties expressed in the literature and further opportunities in ontology design and development will be presented taking the existing ontology evaluation frameworks into account.

BACKGROUND

Approaches to Ontology Development for Educational Purposes

There are several classification methodologies when developing ontologies in education, among which are domain-task, task, and domain ontologies (Breuker, 1999; Devedzic, 2004; Allert et al., 2006; Al-Yahya et al., 2014). Based on their purposes, we can categorize educational ontologies in three groups:

- Ontologies for content access (Aroyo et al., 2002; Mitrovic & Devedzic, 2004; Lemnitzer et al., 2008; Lama et al., 2012; Solomou et al., 2015),
- Ontologies for content creation (Simon et al., 2004; Neto & Gauthier, 2006; Boyce & Pahl, 2007; Oprea, 2011; Manganello et al., 2013; Çelik et al., 2014)
- Ontologies for personalization (Henze et al., 2004; Karampiperis & Sampson, 2005; Fok, 2006; Vargas-Vera & Lytras, 2008; Chen et al., 2011; Clemente et al., 2011)

In the following section, each of these groups will briefly be described.

Ontologies for Content Access

Content access is usually related to providing an access to the existing content within the curriculum in a domain. The main assumption is to list all the standards/goals in a specific course curriculum and map the ontology to these standards accordingly. In their ontology development, for example, Al-Yahya et al. (2014) have designed a curriculum ontology named CURONTO. In their research, they used CURONTO to manage and assess the existing curriculum to provide feedback to the teaching program. They also claimed that, their ontology model is used to establish relations between goals, learning process and the curriculum so that the related stakeholders could benefit from querying the ontology to monitor to what extend the curriculum is aligned with their existing goals and learning processes.

Some researchers took the curriculum model a step further by integrating a user model within the ontology. Clemente et al. (2011, 2014), for example, created a student model by using the ontology as a knowledge representation model. In their ontology, student's cognitive process is taken into account so that with use of ontology, a link is created between student's cognitive process, knowledge and teaching goals. The overall system included three ontologies working in line with

each other. First one, student profile ontology, is used to keep personal information; second, student state ontology is used to track learning process of student; and, learning objectives ontology is used to assess learning outcomes.

Ontologies for Content Creation

Çelik et al. (2014) created an ontology based agent-based Educational Activity Discovery System which is designed to propose educational methodologies for parents or educators about pervasive developmental disorder. They used OWL to build knowledge base and created an ontology of pervasive developmental disorder with concepts and relations of this disorder. By the help of ontology, the system allowed educators to create semantic searches and to support their teaching processes.

Karkar et al. (2014) proposed an educational system for children with intellectual challenges, by creating multimedia tutorials for children dynamically. Researchers used several techniques to create tutorials, such as natural language processing, text extraction, ontology usage and online content retrieval. Researchers went further to claim that their system could dynamically customize learning content by taking learners' preferences into account. Ontology of the system is built from scratch and SPARQL queries are used to list and propose multimedia content.

Marzano and Notti (2015) used ontologies for educational assessment, to assess the system and learning in particular. They created a learning environment in their previous study named EduOntoWiki. This environment is aimed to create an ontology based repository for learning objects. Users of this environment can create content by using ontological structures. EduOntoWiki is used as a test environment for educational assessment in this research. They finally stated that with this approach of educational assessment, it is possible to use different methodologies and models in order to assess systems automatically in an efficient and effective way.

Solomou et al. (2015) have proposed a specialized ontology based version of IEEE LOM standard to use in any e-learning environment to improve discovery and retrieval rates of the educational materials. They called this profile as educational metadata profile (EMP). They further stated that, EMP could effectively describe educational and technical parts of a learning object. By use of this ontology as a basis for EMP, semantic relationships between learning materials can be defined. EMP is used in a distance course on object oriented programming by incorporating learning objects when delivering the online course from a distance. Finally, the researchers concluded that with EMP, learning objects could successfully be retrieved even if the selected concept is not directly stated in the knowledge domain.

Ontologies for Personalization

Kaya and Altun (2011a) proposed an ontology based learner model for e-learning systems. Researchers stated that, although there are standards of consortiums like IEEE (Personal and Private Information, PAPI) or IMS (Learner Information Package, LIP) that provide learner models, these models are very complex to use in real applications. Therefore, researchers have proposed new learner/user models for educational and/or other adaptive systems. This user model included Demographic information (learner related demographic information); current learner status (learner's current knowledge about the domain); expectations (learning goals which learners are expected to acquire, will be provided by the curriculum ontology); individual attributes (individual related data such as cognitive attributes, learning styles, presentation types etc. will be kept); performance (completed courses, resolved tests, progress status, achieved gains and other results which emerge in the interaction of system and learner) and context attributes (technical data such as connection speed, operating system of learner etc.)

Cakula and Sedleniece (2013) proposed an ontology based personalized e-learning model that

uses basic elements of personalization: Student personality, knowledge level, course content and technologies. They explained adaptation of knowledge management for e-learning, learning object as used for basic building blocks of e-learning and metadata used for learning object access. Ontology and semantic web use is also explained in this research, and with this approach, researchers propose that ontologies support semantic search and it is possible to make smart queries from multiple repositories by creating connections to learning objects with ontologies.

CHALLENGES AND DIFFICULTIES

Despite building educational ontologies has its own challenges and difficulties, education domain shares same issues with ontology development in general ontological engineering area. First of all, most of the ontologies are created using OWL DL ontology language and for most of the researchers who are new to ontology creation, it is hard to understand formal description logic (DL) language in OWL DL. DL is a union of formal knowledge representation (KR) languages based on first-order logic and is based on a formal, logic-based semantics (Baader and Nutt, 2003). First-order logic can express structures of a theory in a finite domain (which can be identified as problem domain). Users also shall have a knowledge of open world reasoning to understand DL. Because it is necessary to understand and implement all rules of DL to implement an ontology in OWL DL, ontology developers usually suffer in realizing all these rules in their ontology.

Another ontology creation issue is related to reusing existing ontologies. Despite there are several methodologies exist describing ontology generation from scratch (Uschold & Gruninger, 1996; Fernández-López, 1997; Noy & McGuinness, 2001), there is no single formal methodology that describes how to reuse existing ontologies or how to incorporate other existing ontologies into a new one. Although some researchers encourage to

reuse existing ontologies (i.e., Borst, 1997; Pinto & Martins, 2001; Park et al., 2011; Kalfoglou & Schorlemmer, 2003), lack of an existing and working model makes it difficult to incorporate the existing one into new domains. Therefore, starting from scratch becomes a preferred method, which results in more time and effort. More importantly, when aligning (mapping, merging, integration) ontologies across domains, besides using the same (similar at least) relationships inside domains, it is an issue to use existing relationships that connect and combine different domains (Noy & Musen, 2000; Pinto & Martins, 2001; Kalfoglou & Schorlemmer, 2003; Choi et al., 2006). Also, aligned ontologies shall answer the queries consistently. Thus, after aligning different consistent ontologies, the new created ontology shall remain as consistent in their structure.

There are also several pitfalls in ontology development (Poveda et al., 2009b) such as, using synonyms as classes, defining relationships as inverse even if they are not, recursive definitions, polysemy, using “is” relationships, defining classes entailing other classes, defining relationships out of current development domain, incomplete information. For example “is-a” or “part-of” relation describes only the inclusion of a relation between concepts. Using just these relations eventually turns the ontology into a taxonomy with a hierarchical structure. Thus, power of ontologies to represent the domain relationships cannot be realized thoroughly (Mansur & Yusof, 2013).

In educational domain, several concepts are combined to stand for a new concept. So it is a challenging task to define these concepts and relations across domains. It is also harder to infer meaning from new combined concepts. For example, in math, a function could be represented in multiple forms; whereas, a function is usually considered as a relationship in liberal arts. Once developing the ontology for educational domain, educators are challenged to represent such concepts whether they are concepts or relationships.

Another problem is about using automated ontology building technologies by using textual materials. In this process, all the concepts, restrictions, and relations are extracted from a given text. Since the natural language processing and text extraction methodologies are not perfect yet, all the remaining post processing work need to be completed by ontology engineers. Since this creates a very challenging and time consuming job, it is generally neglected and development process starts with various limitations.

DESIGN SUGGESTIONS

As the ground zero of ontology development is using OWL DL, the first suggestion would be related to the clarification of OWL usage. Rector et al. (2004) has once listed the most common errors and problems about OWL DL in detail and have also suggested solutions about these issues. Solutions of using OWL starts with using the disjoint axioms. Ontology developers need to be familiar with reasoning and formal logic before starting development with OWL DL, otherwise, it will be harder to design an ontology leading to a faulty logic. Furthermore, domain and range constraints should also be carefully defined so that inconsistencies become minimized.

Design patterns consists of shared guidelines to help minimize the design problems. Design patterns are commonly used in software engineering and is also used in ontological engineering to solve common problems. Svatek (2004), Gangemi (2005), De Cea et al. (2008), Gangemi & Presutti (2009), Blomqvist (2009) explained ontology design patterns (ODP) in detail. ODP help users solve ontological design problems and help in modelling ontologies (Poveda et al., 2009a). An example ontology design patterns is partition design pattern. In this pattern, a concept is considered as partition. Another approach is to use the worst practices (Poveda et al., 2009a). Worst practices in ontology design process are listed as using recursive relations, lazy elements,

Table 1. Differences of educational ontology development methodology

	Noy and McGuinness (2001)	Kaya and Altun (2009)
Step 1	Determine the domain and scope of the ontology	Determine the domain and scope of the ontology due to subject matter of educational domain
Step 2	Consider reusing existing ontologies	Consider reusing existing ontologies
Step 3	Enumerate important terms in the ontology	Enumerate important terms in the ontology, considering educational domain and relation of these terms to other terms in other domains
Step 4	Define the classes and the class hierarchy	Define the classes and the class hierarchy due to subject matter of educational domain
Step 5	Define the properties of classes (slots)	Define the properties of classes, considering educational domain and relation of these classes to other educational domains
Step 6	Define the facets of the slots	Define the facets of the slots
Step 7	Create instances	Create instances

undefined inverse relations, and defining classes instead of individuals, etc.

Authors of this paper have reviewed basic ontology development methodologies (Uschold & Gruninger, 1996; Fernández-López, 1997; Noy & McGuinness, 2001), and implemented Noy and McGuinness (2001)'s methodology to create an ontology for educational environments (Kaya & Altun, 2009). As Noy and McGuinness' model for ontology development is mostly suitable for educational ontology development, authors updated some steps of this methodology to make it a more suitable development model and proposed an updated ontology development methodology as one of the results of paper. Differences are given in following table:

One of the main problems in educational ontology development is related to the concept representation. Since concepts are highly dependent on the subject matter, it is not quite easy to apply a standard methodology for extracting concepts from a subject domain. Likewise it is an issue to clarify hierarchies of classes and relations (e.g. slots and facets in ontology) between concepts. Hierarchy of educational concepts often requires multiple inheritances which make inferencing instable within the ontology. So, determining hierarchy of classes in educational domain requires a common consensus across content experts.

Noy and Musen (2000) suggested to use mediators in ontology when reuse and/or integration is preferred. Mediators (Wiederhold, 1992) can answer questions about an information source. Thus, each information source shall have an interface between itself and the mediator. Within this framework, a common data model could then be created by incorporating the information sources and required targets to attach to the data model. As a result, matching rules will be defined so that the sources could be translated into the target; in other words, queries could be executed over these rules. To sum up, it can be concluded that educational environments needs the following fundamental ontology types:

- Domain ontologies, contain information about current domain, holds information about concepts, relationships, restrictions etc. about whole domain.
- Curriculum ontologies, contains information about whole curriculum. This ontology shall reuse several domain ontologies to build an entire curriculum e.g. for a lesson.
- User/learner ontologies, contains information about user, user's development, current information, progress of user, completed or studied materials of user.

All these ontologies will contain concepts, relationships and restrictions within their own domains. It is hard to build ontologies from scratch that covers all the limitations into account. Sometimes ontologies can contain or include other ontologies, and as stated above, there is no standard procedure for users to merge ontologies and use them in cooperation.

Finally, ontology evaluation is another challenge to deal with. Groza et al. (2014) presented a specification of Ontology Building Competition 2013, by providing a set of metrics to be used to evaluate ontologies. The metrics covered five dimensions: Structural, semantic, design patterns, worst practices and capability to answer to competency questions. In structural evaluation, concept and attribute richness are measured. Concept richness is ratio of non-empty classes to total number of instances. Attribute richness measures quantity of information that is held by an individual. Modularization and connectedness measure the reuse of information in an ontology such as the reuse of other ontologies in current ontology.

In semantic evaluation, it is assumed that it is better if an ontology has more expressive logic. In this evaluation consistency and coherence metrics are emphasized. Domain coverage, on the other hand, is evaluated by answering predefined competency questions and predefined terms.

Another ontology analysis based on design patterns is developed by Poveda et al. (2009a). Researchers explained patterns and anti-patterns by illustrating how anti-patterns could be used as an ontology analysis method. Researchers created a decision tree to classify patterns, anti-patterns and worst practices. When deciding which analysis method is the most effective one, first, it is checked whether a design pattern is used when creating the ontology. Secondly, similar step is taken to check whether design fits into anti-pattern. Finally, it is checked whether any of the worst practices in ontology design is observed in the existing ontology. Some of the worst practices in ontology design is listed and finally worst practices when

creating the relationships in ontology are provided to researchers.

Ontologies can also be compared by using precision and recall parameters which are widely used in information retrieval literature with F scores (harmonic mean of these parameters). These parameters are used in evaluation of ontologies (e.g., Khan et al., 2004; Nagypal, 2005; Shehata et al., 2007, Lemnitzer et. al., 2008) and discussed in detail in Kaya and Altun (2011b). Precision is the rate of retrieved document count to total document count, recall is the rate of retrieved relevant document count to total relevant document count either retrieved or cannot be retrieved. F score is harmonic mean of precision and recall values. This methodology can be applied to concepts addressed by ontologies, and number of concepts in ontologies can be used instead of document numbers. As a result of these calculations, precision, recall parameters and F scores can guide developers to decide which ontology could be preferred.

CONCLUSION

In this paper, literature of educational ontology use is reviewed and some of the basic problems for educational ontology development are addressed. Based on educational ontology literature, ontology development process and ontology evaluation techniques have been discussed and synthesized. Furthermore, some suggestions were provoked based on the reported issues and experienced problems and challenges.

FUTURE RESEARCH DIRECTIONS

Use of ontology development design patterns can be used in educational ontology development process, as described in “Design Suggestions” section. But as these blue prints are generally created for ontology development for all domains, educational ontology developers need to be more

specific when developing and applying design patterns and solutions in their future research.

As the ontology development is a very complex process to start from scratch, automated or semi-automated development methodologies are required. By the help natural language processing and data mining methodologies from computer science literature, ontology development process could be elaborated and hopefully more educational ontology-based applications will be utilized.

In evaluating the existing ontologies, researchers tend to apply measures to observe how content is accessible in accurate and fastest way. When ontologies are introduced for educational purposes, on the other hand, educators are also very interested in whether the content created an opportunity for learners to better grasp the provided content. To design an experimental research to observe this effectiveness is also a challenging task. Therefore, more research is needed to measure the effectiveness of existing ontologies across or within a domain.

REFERENCES

- Al-Yahya, M., George, R., & Alfaries, A. (2015). Ontologies in E-learning: Review of the literature. *Int. J. Softw. Eng. Appl*, 9(2), 67–84.
- Allert, H., Markannen, H., & Richter, C. (2006, October). Rethinking the use of ontologies in learning. In *Proceedings of the 2nd International Workshop on Learner-Oriented Knowledge Management and KM-Oriented Learning (LOKMOL 06), in conjunction with the First European Conference on Technology-Enhanced Learning (ECTEL 06)* (pp. 115-125).
- Aroyo, L., Dicheva, D., & Cristea, A. (2002). Ontological Support for Web Courseware Authoring. In S. Cerri, G. Gouardères, & F. Paragauçu (Eds.), *Intelligent Tutoring Systems* (Vol. 2363, pp. 270–280). Springer Berlin Heidelberg. doi:10.1007/3-540-47987-2_31
- Baader, F., & Nutt, W. (2003, January). Basic description logics. In *Description logic handbook* (pp. 43-95). Academic Press.
- Blomqvist, E., Gangemi, A., & Presutti, V. (2009, September). Experiments on pattern-based ontology design. In *Proceedings of the fifth international conference on Knowledge capture* (pp. 41-48). ACM. doi:10.1145/1597735.1597743
- Borst, W. N. (1997). *Construction of engineering ontologies for knowledge sharing and reuse*. Universiteit Twente.
- Boyce, S., & Pahl, C. (2007). Developing Domain Ontologies for Course Content. *Journal of Educational Technology & Society*, 10(3).
- Breuker, J., Muntjewerff, A., & Bredewej, B. (1999, July). Ontological modelling for designing educational systems. *Proceedings of the AI-ED 99 Workshop on Ontologies for Educational Systems*.
- Cakula, S., & Sedleniece, M. (2013). Development of a personalized E-learning model using methods of Ontology. *Procedia Computer Science*, 26, 113–120. doi:10.1016/j.procs.2013.12.011
- Çelik, D., Elçi, A., Inan, N., Elverici, E., & Elci, A. (2014). Educational Activity Discovery System for Individuals with Pervasive Developmental Disorders through Web Ontology Language (OWL). *Egitim ve Bilim*, 39(171).
- Chen, B., Lee, C.-Y., & Tsai, I. (2011). Ontology-Based E-Learning System for Personalized Learning. *International Proceedings of Economics Development & Research*, 18.
- Choi, N., Song, I. Y., & Han, H. (2006). A survey on ontology mapping. *SIGMOD Record*, 35(3), 34–41. doi:10.1145/1168092.1168097
- Clemente, J., Ramírez, J., & De Antonio, A. (2011). A proposal for student modeling based on ontologies and diagnosis rules. *Expert Systems with Applications*, 38(7), 8066–8078. doi:10.1016/j.eswa.2010.12.146

- Clemente, J., Ramírez, J., & De Antonio, A. (2014). Applying a student modeling with non-monotonic diagnosis to Intelligent Virtual Environment for Training/Instruction. *Expert Systems with Applications*, 41(2), 508–520. doi:10.1016/j.eswa.2013.07.077
- De Cea, G. A., Gómez-Pérez, A., Montiel-Ponsoda, E., & Suárez-Figueroa, M. C. (2008). Natural language-based approach for helping in the reuse of ontology design patterns. In *Knowledge Engineering: Practice and Patterns* (pp. 32–47). Springer Berlin Heidelberg.
- Devedzic, V. (2004). Education and the semantic web. *International Journal of Artificial Intelligence in Education*, 14(2), 165–191.
- Fernández-López, M., Gómez-Pérez, A., & Juristo, N. (1997). *Methontology: from ontological art towards ontological engineering*. Academic Press.
- Fok, A. W. (2006). Peonto-integration of multiple ontologies for personalized learning. In *Proceedings of the 5th IASTED International Conference on Web-based Education*. ACTA Press.
- Gangemi, A. (2005). Ontology design patterns for semantic web content. In *The Semantic Web—ISWC 2005* (pp. 262–276). Springer Berlin Heidelberg. doi:10.1007/11574620_21
- Gangemi, A., & Presutti, V. (2009). Ontology design patterns. In *Handbook on ontologies* (pp. 221–243). Springer Berlin Heidelberg. doi:10.1007/978-3-540-92673-3_10
- Groza, A., Bernadette, V., & Vacca, M. (2014, October). A learning environment for building and evaluating ontologies: Case study of 2013 Ontology Building Competition. In *The International Scientific Conference eLearning and Software for Education* (Vol. 4, p. 157). “Carol I” National Defence University.
- Henze, N., Dolog, P., & Nejd, W. (2004). Reasoning and Ontologies for Personalized E-Learning in the Semantic Web. *Journal of Educational Technology & Society*, 7(4), 82–97.
- Kalfoglou, Y., & Schorlemmer, M. (2003). Ontology mapping: The state of the art. *The Knowledge Engineering Review*, 18(01), 1–31. doi:10.1017/S0269888903000651
- Karampiperis, P., & Sampson, D. (2005). Adaptive learning resources sequencing in educational hypermedia systems. *Journal of Educational Technology & Society*, 8(4), 128–147.
- Karkar, A. G., Saleh, M., Saad, S., Ja’am, A., & Mohamad, J. (2014, April). An arabic ontology-based learning system for children with intellectual challenges. In *Global Engineering Education Conference (EDUCON)*, (pp. 670–675). IEEE. doi:10.1109/EDUCON.2014.6826165
- Kaya, G., & Altun, A. (2009). E-Öğrenme Ortamları İçin Özlü Sözlük Ontolojisi ve Görselleştirme Yazılımı Geliştirilmesi. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 36(36).
- Kaya, G., & Altun, A. (2011a). A learner model for learning object based personalized learning environments. In *Metadata and Semantic Research* (pp. 349–355). Springer Berlin Heidelberg. doi:10.1007/978-3-642-24731-6_35
- Kaya, G., & Altun, A. (2011b). Öğrenme nesneleri için ontoloji tabanlı gezinim aracının değerlendirilmesi. Paper presented at the 11th international educational technology conference (IETC 2011), Istanbul, Turkey.
- Khan, L., McLeod, D., & Hovy, E. (2004). Retrieval effectiveness of an ontology-based model for information selection. *The VLDB Journal—The International Journal on Very Large Data Bases*, 13(1), 71–85.

- Lama, M., Vidal, J. C., Otero-García, E., Bugarín, A., & Barro, S. (2012). Semantic Linking of Learning Object Repositories to DBpedia. *Journal of Educational Technology & Society*, 15(4).
- Lemnitzer, L., Mossel, E., Simov, K., Osenova, P., & Monachesi, P. (2008). In M. Iskander (Ed.), *Using a Domain-Ontology and Semantic Search in an E-Learning Environment Innovative Techniques in Instruction Technology, E-learning, E-assessment, and Education*. Springer Netherlands. doi:10.1007/978-1-4020-8739-4_49
- Manganello, F., Falsetti, C., Spalazzi, L., & Leo, T. (2013). PKS: An Ontology-based Learning Construct for Lifelong Learners. *Journal of Educational Technology & Society*, 16(1).
- Mansur, A. B. F., & Yusof, N. (2013). Social learning network analysis model to identify learning patterns using ontology clustering techniques and meaningful learning. *Computers & Education*, 63, 73–86. doi:10.1016/j.compedu.2012.11.011
- Marzano, A., & Notti, A. (2015). Eduonto: An ontology for Educational Assessment. *Journal Of E-Learning And Knowledge Society*, 11(1).
- Mitrovic, A., & Devedzic, V. (2004). A model of multitutor ontology-based learning environments. *International Journal of Continuing Engineering Education and Lifelong Learning*, 14(3), 229–245. doi:10.1504/IJCEELL.2004.004971
- Nagypal, G. (2005, October). Improving information retrieval effectiveness by using domain knowledge stored in ontologies. In *On the Move to Meaningful Internet Systems 2005: OTM 2005 Workshops* (pp. 780–789). Springer Berlin Heidelberg. doi:10.1007/11575863_98
- Neto, W. C. B., & Gauthier, F. A. (2006, June). Sharing and reusing information on web-based learning. *Workshop on Applications of Semantic Web Technologies for E-Learning*.
- Noy, N., & McGuinness, D. L. (2001). *Ontology development 101*. Knowledge Systems Laboratory, Stanford University.
- Noy, N. F., & Musen, M. A. (2000, August). Algorithm and tool for automated ontology merging and alignment. In *Proceedings of the 17th National Conference on Artificial Intelligence (AAAI-00)*. Available as SMI technical report SMI-2000-0831.
- Oprea, M. (2011). An Educational Ontology for Teaching University Courses. *roceedings of the 6th International Conference on Virtual Learning-ICVL*.
- Park, J., Oh, S., & Ahn, J. (2011). Ontology selection ranking model for knowledge reuse. *Expert Systems with Applications*, 38(5), 5133–5144. doi:10.1016/j.eswa.2010.10.002
- Pinto, H. S., & Martins, J. P. (2001, October). A methodology for ontology integration. In *Proceedings of the 1st international conference on Knowledge capture* (pp. 131–138). ACM.
- Poveda, M., & Suárez-Figueroa, M. C., & Gómez-Pérez, A. (2009a, October). Ontology analysis based on ontology design patterns. In *Workshop on Ontology Patterns* (p. 155).
- Poveda, M., Suárez-Figueroa, M. C., & Gómez-Pérez, A. (2009b). Common pitfalls in ontology development. In *Current Topics in Artificial Intelligence* (pp. 91–100). Springer Berlin Heidelberg.
- Rector, A., Drummond, N., Horridge, M., Rogers, J., Knublauch, H., Stevens, R., & Wroe, C. et al. (2004). OWL pizzas: Practical experience of teaching OWL-DL: Common errors & common patterns. In *Engineering Knowledge in the Age of the Semantic Web* (pp. 63–81). Springer Berlin Heidelberg. doi:10.1007/978-3-540-30202-5_5
- Shehata, S., Karray, F., & Kamel, M. (2007, August). A concept-based model for enhancing text categorization. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 629–637). ACM. doi:10.1145/1281192.1281260

Simon, B., Dolog, P., Mikl, Z., Olmedilla, D., & Sintek, M. (2004). Conceptualising Smart Spaces for Learning. *Journal of Interactive Media in Education*, 4.

Solomou, G., Pierrakeas, C., & Kameas, A. (2015). Characterization of Educational Resources in e-Learning Systems Using an Educational Metadata Profile. *Journal of Educational Technology & Society*, 18(4), 246–260.

Svatek, V. (2004, April). Design patterns for semantic web ontologies: Motivation and discussion. *Proceedings of the 7th Conference on Business Information Systems*.

Uschold, M., & Gruninger, M. (1996). Ontologies: Principles, methods and applications. *The Knowledge Engineering Review*, 11(2), 93–136. doi:10.1017/S0269888900007797

Vargas-Vera, M., & Lytras, M. (2008). *Personalized learning using ontologies and semantic web technologies. In Emerging Technologies and Information Systems for the Knowledge Society* (pp. 177–186). Springer. doi:10.1007/978-3-540-87781-3_20

Wiederhold, G. (1992). Mediators in the architecture of future information systems. *Computer*, 25(3), 38–49. doi:10.1109/2.121508

ADDITIONAL READING

Al-Yahya, M., George, R., & Alfaries, A. (2015). Ontologies in E-learning: Review of the literature. *Int. J. Softw. Eng. Appl*, 9(2), 67–84.

Gangemi, A., & Presutti, V. (2009). Ontology design patterns. In *Handbook on ontologies* (pp. 221–243). Springer Berlin Heidelberg. doi:10.1007/978-3-540-92673-3_10

Guarino, N., Oberle, D., & Staab, S. (2009). What is an ontology? In *Handbook on ontologies* (pp. 1–17). Springer Berlin Heidelberg. doi:10.1007/978-3-540-92673-3_0

Noy, N., & McGuinness, D. L. (2001). *Ontology development 101*. Knowledge Systems Laboratory, Stanford University.

Poveda, M., Suárez-Figueroa, M. C., & Gómez-Pérez, A. (2009b). Common pitfalls in ontology development. In *Current Topics in Artificial Intelligence* (pp. 91–100). Springer Berlin Heidelberg.

Uschold, M., & Gruninger, M. (1996). Ontologies: Principles, methods and applications. *The Knowledge Engineering Review*, 11(2), 93–136. doi:10.1017/S0269888900007797

KEY TERMS AND DEFINITIONS

Design Pattern: Design patterns consists of shared guidelines helping design problems. Design patterns commonly used in software engineering and is also used in ontological engineering to solve common problems.

Ontology: Ontology provides a framework to determine a domain of interest and to build a common understanding in this domain between different applications by representing concepts and relationships that exist between these concepts.

Personalization: Personalization is adapting learning experience to different learners due to the analysis of knowledge, skills and learning preferences of individuals.

Semantic Web: Semantic web defines data and relations of the data significantly so that a common knowledge base will be provided among applications and more efficient search, integration and reuse operations will be provided. Semantic web provides an environment which human and machines can communicate.

Factors Contributing to the Effectiveness of Online Students and Instructors

C

Michelle Kilburn

Southeast Missouri State University, USA

Martha Henckell

Southeast Missouri State University, USA

David Starrett

Columbia College, USA

INTRODUCTION

As technological advances become mainstream in higher education, web-based learning continues to gain focus and momentum. The number of online courses in K-12, technical, professional and liberal arts education has increased significantly (Allen & Seman, 2013; Picciano, 2001; Setzer & Lewis, 2005). Multimedia technology has changed the landscape of distance education (Zirkle, Norris, Winegardner & Frustaci, 2006). Gray (2013) purports barriers to online learning are being addressed and students have access to more educational opportunities than ever before.

Online learning programs often tout more interactive, individualized and independent learning (Chen, Czerwinski & Macredie, 2000; Inan, Yukselturk & Grant, 2009; Park & Lee, 2003). However, a major challenge with web-based learning is identifying the idiosyncrasies between learning online and the traditional learning format (Inan, et. al., 2009; Martinez, 2003; Rovai, 2003). Another challenge is identifying the very different profile of online students (Dutton, Dutton & Perry, 2002; Sikora & Carroll, 2003).

Identifying the positive attributes of students and instructors in the online environment will contribute to the understanding of how we can enhance the learning experience for the student and the teaching experience for the instructor. This article will assist students and instructors

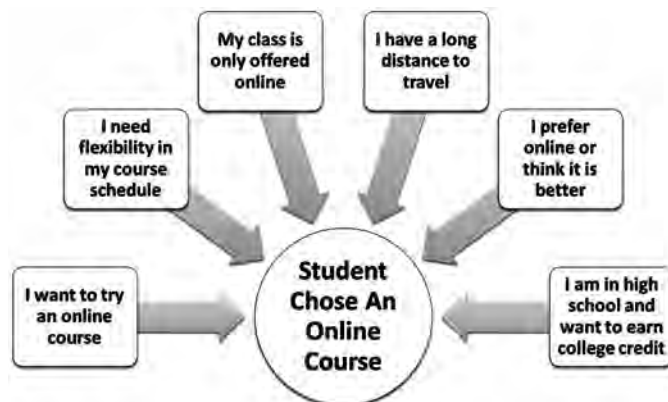
in understanding the differences that may be experienced in the online environment versus the face-to-face environment and provide the opportunity to consider whether online learning and/or teaching is a “good fit” for them. Understanding why students and/or instructors might choose the online environment will also assist administrators in developing successful, quality online programs that enrich the experiences for both students and instructors.

BACKGROUND

In 1981, the first online classes were developed at the School of Management and Strategic Studies at Western Behavior Sciences Institute in La Jolla, California. An evaluation of the program, and the discussions that took place, revealed that the quality of the online course was higher than the information collected in the traditional classroom setting (Feenberg, 1999).

Since that time, a number of studies have compared the effectiveness of online instruction to traditional lecture formats. Findings have admittedly been mixed (Rivera, & McAlister, 2001; Ungerleider & Burns, 2004; Zhang, 2005). Even though a majority of the studies find no difference in student performance and student satisfaction, regardless of the delivery format (Lim, Kim, Chen & Ryder, 2008; McFarland & Hamilton,

Figure 1. Motivations to take online courses



2006), there is concern regarding online student retention. Knowledge of student characteristics and how they can possibly affect online course success can provide the opportunity for faculty to intervene before grades are affected or dropout rates increase (Cochran, et al., 2014).

Kilburn (2005) developed the following conceptual map regarding student motivations to take an online course at a particular University in the Midwest. (see Figure 1)

CHARACTERISTICS OF SUCCESSFUL STUDENTS AND INSTRUCTORS

In the upcoming section, an examination of student and instructor characteristics and how each of those different roles contributes to the quality of an online course will help provide insight into the foundational underpinnings of web-based learning.

Student Characteristics

It is estimated that five out of six students taking an online course are employed and would not be able to attend traditional classes (Thomas, 2001). Literature suggests that the growth in online courses is based on attracting new students rather than “stealing” from students enrolled in

current on-campus programs (Mangan, 2001, Thomas, 2001). Kauffman (2014) points out the convenience and flexibility that online education affords students with work and family demands. Undergraduate online students are commonly older, married, or have dependents (Dotterweich & Rochelle, 2012). Regardless of gender or age, students who need the flexibility of online or distance education classes in order to obtain a degree may find online to be a good option for them (Dotterweich & Rochelle, 2012).

Kilburn (2005) found that positive characteristics identified in studies stress the importance of an active versus passive student role in an online course. Some researchers have attempted to identify student abilities that suggest whether a student will complete an online course, or be less satisfied with an online course, in comparison to the traditional classroom setting. In Kirmizi’s study (2015), student satisfaction was used as the dependent variable and sub-dimensions of a learner’s readiness: computer self-efficacy, self-directed learning, learner control, motivation, and online self-efficacy, as the independent variables. Kirmizi identified self-directed learning as the most valuable predictor of success, with learner control and motivation as next in importance. Shea and Bidjerano (2010) also suggested that motivation, self-efficacy, and self-regulation are the central components of learner presence in an online course. Studies have consistently reported

students need to develop self-regulatory abilities and take control of their learning (Dabbagh & Kitsantas, 2004; Picciano, 2001; Rovai, 2003; Saba, 2000).

One attraction to online learning is the presumed capacity to increase access and equity to learners by removing some of the barriers to participation (Harasim, 1990). Sullivan (2001) found that students appear to find the online environment more welcoming for quiet and shy students than the traditional classroom. Varela, et. al., (2012) found that low gregariousness and achievement orientation were able to predict learning. In another particular study, using the Myers-Briggs Type Inventory, almost half of the online students surveyed were introverts, sensors, or judges (Mupinga, Nora & Yaw, 2006). Online students also may not be as comfortable working in groups and prefer the more individualized environment that online courses can provide (Haigh, 2007).

Sullivan (2001) asserts that nontraditional students, particularly female, adult learners with children or familial responsibilities, value online courses. Research has suggested that the asynchronous nature of the online environment might encourage a more reflective type of interaction that changes the dynamics of classroom discussion in a way that female students find rewarding (Selfe, 1999). Arbaugh (2000) found that males communicate online in a more competitive manner, while females viewed Internet-based communication as a medium to develop collaborations in online learning. It should be noted that some studies reported differences between genders (Wladis, et al., 2015; Dotterweich & Rochelle, 2012; Yang, et al., 2011) while others did not find significant differences (Lim, 2001). Regardless of gender, students often benefit from online courses because of the ability to balance work and family, among other convenience factors (Al-Asfour & Bryant, 2011). While Wladis et al. (2015) discovered both men and women performed worse in online science, technology, engineering, and mathematics courses, they did not find a significant difference in regards to the online medium and ethnicity. At

some institutions the push to offer more online courses is promoted in order to reach populations who have not had access to higher education because of social or geographical barriers (Enoch & Soker, 2006).

The age of a student in relationship to course completion rates has been an area of focus in current research. Some research suggests that students over 30 and under 50 years old are more likely to choose and finish a distance course than traditional students age 18 to 29 (Cao, et. al., 2011; Sikora & Carroll, 2003). Other studies have generated mixed results. Xenos, et al., (2002) found a positive correlation between age and dropout rates. Jones et al. (2004) suggested that as students get older, more dropouts are likely to occur. In contrast, Cochran et al. (2014) found freshman to have the highest rate for online course withdrawal, with a steady decrease as class rank increases.

Students should possess the ability to navigate the Internet and deal effectively with computer software and hardware difficulties. Technological difficulties and “feeling lost in cyberspace” were cited as negatives by students (El Mansour & Mupinga, 2007). Motteram and Forrester (2005) suggest the first hurdle students struggle with in the online environment is becoming familiar with the computer (i.e., how to access, enter and navigate sites).

In the online environment, successful students shift from a more passive role in the exchange of knowledge to a more active role. Several studies purport that the following are the most influential factors affecting a student’s active participation in online learning: (1) prior knowledge of online learning, (2) knowledge of a given subject area, (2) strategies to cope with information overload, (3) personality traits, (4) instructor facilitation, and (5) appropriate feedback (Vonderwell & Zachariah, 2005). Koohang’s research suggests active learning has three essential stages – the underpinning state, the ownership state and the engaging stage (Koohang, 2012 & Koohang, et. al., 2013).

Instructor Characteristics

The relationship between student-teacher interactions and learning outcomes has been well documented in the traditional classroom and distant education (Powers & Rossman, 1985). Of particular importance in traditional classrooms is teacher “immediacy tendencies.” Immediacy alludes to the psychological distance between student and instructor (Weiner & Mehrabian, 1968). Research suggests that a teacher’s verbal and non-verbal immediacy behaviors can lessen the perceived distance between themselves and their students. By lessening the perception of disconnect between the instructor and student in a course, instructors can help facilitate (directly or indirectly) effective learning. Pollard, et. al., (2013) found a significant association between teacher presence and learner motivation in the online environment.

With the importance of interactions established as a crucial component of learning, one might assume it would be equally important online. Certain researchers have suggested that asynchronous media are less capable of representing the social presence of participants (Short, Williams & Christie, 1976). Researchers with experience teaching online contest this view, arguing that rather than being impersonal, computer-mediated communication often seems to be hyper-personal (Walther, 1994). According to Kilburn (2005), research indicates that participants in online courses communicate, argue, and create social presence by projecting immediacy behaviors (LaRose & Whitten, 2000; Garrison & Anderson, 2003; Rourke, Anderson, Garrison & Archer, 2001;). Kizilcec, Bailenson and Gomez (2015) highlighted the benefits of social and other nonverbal cues from the instructor during online video lectures that included the instructor’s “talking head”. This reportedly helped learner’s focus and feel more connected.

Interactions among students seem to clearly matter in online discussions. The development of social presence and the perceived interaction with

others is one of the cornerstones for the development of online communities (Rourke, Anderson, Garrison & Archer, 2001). Lu and Jeng (2006) discovered that instructors who strived to serve as both facilitator and co-participant were helpful in “enhancing knowledge construction,” particularly in discussion forums (p. 196). Students also prefer courses and assignments that included opportunities to apply content and not just reference the textbook (Deggs, Grover, Kacirek, 2010).

Jiang and Ting (2000) report correlations between perceived learning in online courses and the specificity of the instructors’ discussion, instructions and the percent of course grades based on discussion responses. Recent research has found that online graduate students place great importance on the availability of, access to, and feedback from faculty (Fahy, Spencer, & Halinski, 2008). These studies exemplify the magnitude of importance that is placed on the instructor to assure the quality of an online course.

Researchers have begun to look at the changing roles of teachers in online classrooms (Guasch, Alvarez & Espasa, 2010). Coppola, Hiltz and Rotter, (2001) assert that in any environment teachers have three roles – cognitive, affective, and managerial. They found that with online courses the cognitive role often shifts to one of deeper complexity. The affective role requires instructors to find new tools to express emotion and the managerial role requires greater attention to detail, more structure, and additional student monitoring. Easton (2003) reported that prior to the start of class, the instructor role was more like an instructional designer and subject matter expert; however, once the class began, the instructor’s activities took on more of an interactive, facilitator role. Darabi, Sikorski & Harvey (2006) describe the various instructors’ roles as: managerial, social and technical. Mupina, et. al., (2006) reported the top four needs of online students were: technical help, flexible and understanding instructors, early information about the course, and sample assignments. Faculty must also communicate regularly with students through consistent feedback, responses to post-

ings, and responses to concerns (Vesely, Bloom & Sherlock, 2007).

Paul and Cochran (2013) suggest four components are “necessary and important as single entities for delivering and utilizing online education, the larger risks and rewards for online education occur where these components intersect” (p. 50). The four components are institution, student, faculty and technology.

Janicki & Liegle (2001) evaluated the work of a wide range of instructional design professionals and developed a list of ten concepts believed to support effective design of web-based instruction. The findings included: (1) instructors acting as facilitators, (2) usage of a variety of presentation styles, (3) multiple exercises, (4) hands-on problems, (5) learner control of pacing, (6) frequent testing, (7) clear feedback, (8) consistent layout, (9) clear navigation, and (10) available help screens. With the number of online programs and course offerings increasing, faculty roles and the nature of teaching is changing. More faculty and support staff are required to meet the demands of online learning at many institutions (Bennett & Lockyer, 2004; Wiesenbergs & Stacey, 2008). Online teachers “are required to possess a diverse set of competencies and their extent of utilization relies on the context or role they are required to perform and also the kind of resources and support available” (Bawane & Spector, 2009, p. 387). Britt (2015) suggests that allowing students to use real world knowledge is vital for a valuable online classroom experience. Faculty must ensure they are up-to-date with the latest methods to enhance online education.

Interaction has been acknowledged as an important component of learning in conventional and distance education (Moore, 1993). Studies have argued for the importance of instructors’ social or interpersonal feedback when attempting to improve learning achievement in online courses (Jung, 2000). Dennen, Darabie & Smith (2007) have suggested effective online instructors adopt the following practices: (1) maintaining frequency of contact, (2) having a presence in class discussion, and (3) making expectations clear to

learners. Similarly, Hunter (2011) suggests that interaction is a critical expectation that influences student satisfaction. Live chat session, instant messaging options, discussion board participation, course announcements, grading feedback and timely response to e-mails are just a few ways that increased accessibility can be accomplished.

Instructors must balance academic interaction related to the subject matter with more personal student interaction that will cultivate a sense of relationship and community (Moller, 1998). Dennen (2007) purports that “instructors establish a persona via both presence (amount of instructor posts) and position (interaction relative to those in the student role)” (p. 95). Some online learners find it challenging to stay on task. Consequently, they fail to complete coursework. Prompt e-mail responses may help students remain on task and on schedule (Nandi, Hamilton & Harland, 2012). Baran, Correia, and Thompson (2011) suggest that because online students are expected to take control of their learning and become actively involved in the process, facilitation of online learning plays a greater role.

FUTURE RESEARCH DIRECTIONS

As online learning continues to grow in the higher education arena, more research needs to be done on the changing roles of students and instructors in the online environment. Time is no longer spent in the classroom passively listening to a lecture; students are required to play a more active role in their learning. Instructors often struggle to adapt their teaching approaches to develop active interaction and engagement in their online courses.

Rethinking quality assurance, how to measure student success and what constitutes quality interaction will be a crucial step in moving instructors and administrators from the traditional teaching and classroom management frame of reference to the “virtual campus.” Encouraging instructors and students to “think outside the box” and experiment with ways to enhance interaction will help shape the future of online learning.

CONCLUSION

Students will need to adjust to a more self-directed, student-centered approach to learning, while instructors will need to move toward a facilitating model of instruction. Understanding how to assist both students and instructors to adjust to more nontraditional roles in academe is critical. Instructional designers in higher education have the crucial role of helping outstanding face-to-face instructors translate their expertise, knowledge and personality into outstanding online courses. This discussion has reinforced the argument that developing an online course is not a matter of simply putting lecture notes and presentations online. Instructors must assume both an affective and managerial role, going so far as predicting which students may warrant intervention. The instructional designer must help them stock a new “tool box” to effectively transfer knowledge, communicate with their students and develop a sense of community in the online environment. Administrators may also need to rethink how to evaluate instructors in the promotion and tenure process in order to acknowledge instructors who excel in the online environment and make that crucial connection with their students.

Understanding the premises and concepts in this discussion will not only assist administrators and instructors in developing quality online courses and programs, but also provide an understanding of how online learning has resulted in an evolution of student and instructor roles. Successful students take on a more active and participative role and successful instructors take on a more facilitative and supportive role.

REFERENCES

- Al-Asfour, A., & Bryant, C. (2011). Perceptions of Lakota Native American students taking online business courses at Oglala Lakota College (OLC). *American Journal of Business Education*, 4(10), 43–50.
- Allen, E., & Seaman, J. (2013). *Changing course: Ten years of tracking online education in the United States*. Babson Survey Research Group and Qualtrics Research Group. Retrieved from: <http://www.onlinelearningsurvey.com/reports/changingcourse.pdf>
- Arbaugh, J. D. (2000). An exploratory study of the effects of gender on student learning and class participation in an Internet based MBA course. *Management Learning*, 31(4), 533–549. doi:10.1177/1350507600314006
- Baran, E., Correia, A., & Thompson, A. (2011). Transforming online teaching practice. Critical analysis of the literature on the roles and competencies of online teachers. *Distance Education*, 32, 421–439. doi:10.1080/01587919.2011.610293
- Bawane, J., & Spector, J. (2009). Prioritization of online instructor roles: Implications for competency-based teacher education programs. *Distance Education*, 30(3), 383–397. doi:10.1080/01587910903236536
- Bennett, S., & Lockyer, L. (2004). Becoming an online teacher: Adapting to a changed environment for teaching and learning in higher education. *Educational Media International*, 41(3), 231–248. doi:10.1080/09523980410001680842
- Berge, Z. L., & Collins, M. P. (1995). Computer-mediated scholarly discussion groups. *Computers & Education*, 24(3), 183–189.
- Boaz, M., Elliott, B., Foshee, D., Hardy, D., Jarmon, C., & Olcott, D. (1999). *Teaching at a distance: A handbook for instructors*. Mission Viejo, CA: Archipelago Productions.
- Britt, M. (2015). How to better engage online students with online strategies. *College Student Journal*, 49(3), 399–404.
- Cao, Y., & Sakchutchawan, S. (2011). Online vs. traditional MBA: An empirical study of students’ characteristics, course satisfaction, and overall success. *The Journal of Human Resource Management and Adult Learning*, 7(2), 1–12.

- Carr, S. (2000). As distance education comes of age, the challenge is keeping the students. *The Chronicle of Higher Education*, 46(23), A39–A41.
- Chen, C., Czerwinski, M., & Macredie, R. (2000). Individual differences in virtual environments: Introduction and overview. *Journal of the American Society for Information Science American Society for Information Science*, 51(6), 499–507. doi:10.1002/(SICI)1097-4571(2000)51:6<499::AID-ASI2>3.0.CO;2-K
- Cochran, J. D., Campbell, S. M., Baker, H. M., & Leeds, E. M. (2014). The role of student characteristics in predicting retention in online courses. *Research in Higher Education*, (55): 27–48.
- Coppola, N. W., Hiltz, S. R., & Rotter, N. (2001). *Becoming a virtual professor: Pedagogical roles and ALN*. Piscataway, NJ: Institute of Electrical and Electronics Engineers Press.
- Dabbagh, N., & Bannan-Ritland, B. (2005). *Online learning: Concept, strategies, and applications*. Upper Saddle River, NJ: Pearson Education.
- Darabi, A., Sikorski, E., & Harvey, R. (2006). Validated competencies for distance teaching. *Distance Education*, 27(1), 105–122. doi:10.1080/01587910600654809
- Deggs, D., Grover, K., & Kacirek, K. (2010). Expectations of adult graduate students in an online degree program. *College Student Journal*, 44(3), 690–698.
- Dennen, V. P. (2007). Presence and positioning as components of online instructor persona. *Journal of Research on Technology in Education*, 40(1), 95–108. doi:10.1080/15391523.2007.10782499
- Dennen, V. P., Darabi, A. A., & Smith, L. J. (2007). Instructor-Learner Interaction in Online Courses: The Relative Perceived Importance of Particular Instructor Actions on Performance and Satisfaction. *Distance Education*, 28(1), 65–79.
- Dotterweich, D. P., & Rochelle, C. F. (2012). Online, instructional television and traditional delivery: Student characteristics and success factors. *American Journal of Business Education*, 5(2), 129–137.
- Dutton, J., Dutton, M., & Perry, J. (2002). How do online students differ from lecture students? *Journal of Asynchronous Learning Networks*, 6(1), 1–20.
- Easton, S. S. (2003). Clarifying the instructor's role in online distance learning. *Communication Education*, 52(2), 87–105. doi:10.1080/03634520302470
- El Mansour, B., & Mupinga, D. M. (2007). Students' positive and negative experiences in hybrid and online classes. *College Student Journal*, 41(1), 242–248.
- Enoch, Y., & Soker, Z. (2006). Age, gender, ethnicity and the digital divide: University students' use of web-based instruction. *Open Learning*, 21, 99–110. doi:10.1080/02680510600713045
- Fahy, P. J., Spencer, B., & Halinski, T. (2008). The self-reported impact of graduate program completion on the careers and plans of graduates. *The Quarterly Review of Distance Education*, 9(1), 51–71.
- Feenberg, A. (1999). No frills in the virtual classroom. *Academe*, 85(5), 26–31. doi:10.2307/40251765
- Garrison, D. R., & Anderson, T. (2003). *2st century: A framework for research and practice*. London: Routledge Falmer.
- Gray, D. (2013). Barriers to online post-secondary education crumbles: Enrollment in face-to-face courses decline as enrollment in online courses increases. *Contemporary Issues in Education Research*, 6(3), 345–348.

- Guasch, T., Alvarez, I., & Espassa, A. (2010). University teacher competencies in a virtual teaching/learning environment: Analysis of a teacher training experience. *Teaching and Teacher Education*, 26(2), 199–206. doi:10.1016/j.tate.2009.02.018
- Haigh, M. (2007). Divided by a common degree program? Profiling online and face-to-face information science students. *Education for Information*, 25, 93–110.
- Harasim, L. (1990). *Online Education: Perspectives on a new environment*. New York: Praeger.
- Harasim, L. M. (1995). *Learning networks: A field guide to teaching and learning online*. MIT Press.
- Hunter, D. (2011). Who holds the pen? Strategies to student satisfaction scores in online learning environments. *The Business Review*, 18(2), 75–81.
- Inan, F. A., Yukelturk, E., & Grant, M. M. (2009). Profiling potential dropout students by individual characteristics in an online certificate program. *International Journal of Instructional Media*, 36(2), 163–176.
- Janicki, T., & Liegle, J. O. (2001). Development and evaluation of a framework for creating web-based learning modules: A pedagogical and systems approach. *Journal of Asynchronous Learning Networks*, 5, 189–206.
- Jiang, M., & Ting, E. (2000). A study of factors influencing students' perceived learning in a web-based course environment. *International Journal of Educational Telecommunications*, 6, 317–338.
- Jones, P., Packham, G., Miller, C., & Jones, A. (2004). Roles of the students and teachers in distance education. *Turkish Online Journal of Education*, 5(4).
- Jung, I. S. (2000). Technology innovations and the development of distance education. *Open Learning*, 15(3), 217–231. doi:10.1080/713688402
- Jung, I. S., & Lee, D. S. (1993). Improving interactivity in distance education through computer-mediated communication. *Journal of Distance Education*, 7(4), 127–226.
- Kauffman, H. (2014). A review of predictive factors of student success in and satisfaction with online learning. *Communication Sciences and Disorders*, (23), 1–13.
- Kilburn, M. (2005). *A descriptive analysis of online learning at a Midwest university* (Doctoral dissertation). University of Missouri, Columbia, MO.
- Kirmizi, O. (2015). The influence of learner readiness on student satisfaction and academic achievement in an online program at higher education. *The Turkish Online Journal of Educational Technology*, 1(14), 133–142.
- Kizilcec, R., Bailenson, J., & Gomez, C. (2015). The instructor's face in video instruction: Evidence from two large-scale field studies. *Journal of Educational Psychology*, 107(3), 724–739.
- Koohang, A. (2012). Active learning in e-learning: Advancing a systemic model. *Issues in Information Systems*, 13(1), 68–76.
- Koohang, A., & Paliszkievicz, J. (2013). Knowledge construction in e-learning: An empirical validation of an active learning model. *Journal of Computer Information Systems*, 53(3), 109–114.
- LaRose, R., & Whitten, P. (2000). Re-thinking instructional immediacy for web courses: A social cognitive exploration. *Communication Education*, 49, 320–338. doi:10.1080/03634520009379221
- Lim, C. K. (2001). Computer self-efficacy, academic self-concept, and other predictors of satisfaction and future participation of adult distance learners. *American Journal of Distance Education*, 15(2), 41–51. doi:10.1080/08923640109527083

- Lim, J., Kim, M., Chen, S. S., & Ryder, C. E. (2008). An empirical investigation of student achievement and satisfaction in different learning environments. *Journal of Instructional Psychology*, 35(2), 113–119.
- Lu, L. L., & Jeng, I. (2006). Knowledge construction in in-service teacher online discourse: Impacts of instructor roles and facilitative strategies. *Journal of Research on Technology in Education*, 39(2), 183–202. doi:10.1080/15391523.2006.10782479
- Mangan, K. S. (2001). Expectations evaporate for online MBA programs. *The Chronicle of Higher Education*, 48(6), A31-A32.
- Martinez, M. (2003). High attrition rates in e-learning: Challenges, predictors and solutions. *The ELearning Developers' Journal*. Retrieved on October 13, 2013, from <http://www.elearning-guild.com/pdf/2/071403MGT-L.pdf>
- McFarland, D., & Hamilton, D. (2006). Factors affecting student performance and satisfaction: Online versus traditional course delivery. *Journal of Computer Information Systems*, 46(2), 25–32.
- Moller, L. (1998). Designing communities of learners for asynchronous distance education. *Educational Technology Research and Development*, 46(4), 115–122. doi:10.1007/BF02299678
- Moore, M. G. (1993). *Distance education: New perspective*. London: Croom Helm.
- Motteram, G., & Forrester, G. (2005). Becoming an online distance learner: What can be learned from students' experiences of induction to distance programmes? *Distance Education*, 26(3), 281–298. doi:10.1080/01587910500291330
- Mupinga, D. M., Nora, R. T., & Yaw, D.C., (2006). The learning styles, expectations and needs of online students. *College Teaching*, 54(1), 185-189. Retrieved on October 13, 2013 from Ebscohost.com
- Nandi, D., Hamilton, M., & Harland, J. (2012). Evaluating the quality of interaction in asynchronous discussion forums in fully online courses. *Distance Education*, 33(1), 5–30. doi:10.1080/01587919.2012.667957
- Park, O., & Lee, J. (2003). Adaptive instructional system. In D. H. Jonassen (Ed.), *Handbook of research for educational communications and technology* (pp. 403–437). Mahwah, NJ: Lawrence Erlbaum Associates.
- Paul, J. A., & Cochran, J. D. (2013). Key interactions for online programs between faculty, students, technologies and educational institutions: A holistic framework. *The Quarterly Review of Distance Education*, 14(1), 49–62.
- Picciano, A. G. (2001). *Distance learning: Making connections across virtual space and time*. Upper Saddle River, NJ: Prentice-Hall.
- Pollard, H., Blevins, R., Connor, M., & McGovern, L. (2013). An examination of the relationship between teaching presence, social presence, learner motivation, and self-reported learning among online MBA students. *Journal of American Academy of Business*, 18(2), 23–30.
- Powers, S., & Rossman, M. (1985). Student satisfaction with graduate education: dimensionality and assessment in college education. *Psychology: A Quarterly Journal of Human Behavior*, 22(2), 46–49.
- Reeves, T. C., & Reeves, P. M. (1997). The effective dimensions of interactive learning on the WWW. In B. H. Khan, (Ed.), *Web-based Instruction*. Englewood Cliffs, NJ: Academic Press.
- Rivera, J. C., & McAlister, M. K. (2001). A comparison of student outcomes and satisfaction between traditional and web-based course offerings. *Proceedings of the 2001 Information Resources Management Association International Conference*, 770-772.

- Rourke, L., Anderson, T., Garrison, D. R., & Archer, W. (2001). Assessing social presence in asynchronous text-based computer conferencing. *Journal of Distance Education, 14*(3), 51–71.
- Rovai, A. P. (2003). In search of higher persistence rates in distance education online programs. *The Internet and Higher Education, 6*(1), 1–16. doi:10.1016/S1096-7516(02)00158-6
- Saba, F. (2000). Student attritions: How to keep your online learner focused. *Distance Education Report, 4*(14), 1–2.
- Selfe, C. (1999). Technology and literacy: A story about the perils of not paying attention. *College Composition and Communication, 50*, 411–436. doi:10.2307/358859
- Setzer, J. C., & Lewis, L. (2005). *A profile of participation in distance education: 1999-2000*. Washington, DC: National Center for Educational Statistics.
- Shea, P., & Bidjerano, T. (2010). Learning presence: Towards a theory of self-efficacy, self-regulation, and the development of communities of inquiry in online and blended learning environments. *Computers & Education, 55*, 1721–1731. doi:10.1016/j.compedu.2010.07.017
- Short, J., Williams, E., & Christie, B. (1976). *The social psychology of telecommunications*. Toronto: Wiley.
- Sikora, A. C., & Carroll, C. D. (2003). *A profile of participation in distance education; 1999--2000*. Washington, DC: National Center for Educational Statistics. doi:10.1037/e610562011-009
- Sullivan, P. (2001). Gender differences and the online classroom: Male and female college students evaluate their experiences. *Community College Journal of Research and Practice, 25*, 805–818. doi:10.1080/106689201753235930
- Thomas, K. Q. (2001). Local colleges providing online learning programs. *Rochester Business Journal, 16*(43), n.p..
- Ungerleider, C., & Burns, T. (2003). *A systematic review of the effectiveness and efficiency of networked ITC in education: A state of the art report to the Council of Ministers Canada and Industry Canada*. Ottawa: Industry Canada.
- Varela, O. E., Cater, J. J., & Michael, N. (2012). Online learning in management education: An empirical role of personality traits. *Journal of Computing in Higher Education, 24*, 209–225. doi:10.1007/s12528-012-9059-x
- Vesely, P., Bloom, L., & Sherlock, J. (2007). Key elements of building online community: Comparing faculty and student perceptions. *MERLOT Journal of Online Learning and Teaching, 3*(3). Retrieved on October 13, 2013 from <http://jolt.merlot.org/vol3no3/vesely.htm>.
- Vonderwell, S., & Zachariah, S. (2005). Factors that influence participation in online learning. *Journal of Research on Technology in Education, 38*(2), 213–230. doi:10.1080/15391523.2005.10782457
- Walther, J. (1994). Interpersonal effects in computer mediated interaction. *Communication Research, 21*, 460–487. doi:10.1177/009365094021004002
- Weiner, M., & Mehrabian, A. (1968). *Language within language: Immediacy, a channel in verbal communication*. New York: Appleton-Century-Crofts.
- Wiesenberg, F., & Stacey, E. (2008). Teaching philosophy: Moving from face-to-face to online classrooms. *Canadian Journal of University Continuing Education, 34*(1), 63–69.

Wladis, C., Conway, K. M., & Hachey, A. C. (2015). The online STEM Classroom—Who succeeds? *Community College Review*, 2(43), 142–164.

Xenos, M., Pierrakeas, C., & Pintelas, P. (2002). A survey on student dropout rates and dropout causes concerning the students in the course of informatics of the Hellenic Open University. *Computers & Education*, 39, 361–377. doi:10.1016/S0360-1315(02)00072-6

Yang, Y., Cho, Y., Mathew, S., & Worth, S. (2011). College student effort expenditure in online versus face-to-face courses. *Journal of Advanced Academics*, 4(22), 619–638.

Zhang, D. (2005). Interactive multimedia-based e-learning: A study of effectiveness. *American Journal of Distance Education*, 19(3), 149–162. doi:10.1207/s15389286ajde1903_3

Zirkle, C., Norris, C., Winegardner, A., & Frustaci, E. (2006). Distance education programming barriers in business education teacher preparation programs in the United States. *Career and Technical Education Research*, 31(2), 101–118. doi:10.5328/CTER31.2.101

ADDITIONAL READING

Allen, I. E., & Seman, J. (2004). *Entering the mainstream: The quality and extent of online education in the United States, 2003 and 2004*. Neham, MA: The Sloan Consortium.

Brusilovsky, P. (1999). Adaptive and intelligent technologies for web-based education. *Kunstliche Intelligenz*, 4, 19–25.

Hawisher, G.E. & Pemberton, M. A. (1997). Writing across the curriculum encounters asynchronous learning networks or WAC meets up with ALN. *Journal of Asynchronous*.

Holt, L. (1993). Learning that transcends time and place. *Educational Technology*, 79(4), 50–56.

Keller, J. M. (1999). Motivation in cyber learning environments. *International Journal of Educational Technology*, 1(1), 7–30.

Kerka, S. (1996). *Distance Learning, the Internet, and the world wide web*. Columbus, OH: Office of Educational Research and Improvement.

Khan, B. H. (1997). *Web-based instruction*. Englewood Cliffs, NJ: Educational Technology Publications.

Klesius, J., Homan, S., & Thompson, T. (1997). Distance education compared to traditional instruction: The students' view. *International Journal of Instructional Media*, 24(3), 207–220.

Lim, C. I. (1999). Integrated approach in designing interactive WBI. *Korea Journal of Educational Technology*, 15(1), 3–24.

Moore, M. G., & Kearsley, G. (1996). *Distance education: A systems view*. Belmont, CA: Wadsworth Publishing Company.

Ryan, S. (2001). Is online learning right for you? *American Agent and Broker*, 73(6), 54–58.

Thompson, M. M., & Chute, A. G. (1998). A vision for distance education: Networked learning environments. *Open Learning*, 2(13), 4–11. doi:10.1080/0268051980130203

Weisband, S. P. (1992). Group discussion and first advocacy effects in computer-mediated and face-to-face decision making groups. *Organizational Behavior and Human Decision Processes*, 53(3), 352–380. doi:10.1016/0749-5978(92)90070-N

Willis, B. (1993). *Distance education: Strategies & tools*. Englewood Cliffs, NJ: Educational Technology Publications.

KEY TERMS AND DEFINITIONS

Asynchronous Learning: Electronic communication in which the student and teacher interact via e-mail and listservs, but do not do so

by being on the Internet at the same time (Berge & Collins, 1995).

Distance Learning: Learning that occurs when the instructor and students are separated by physical distance and technology is used to bridge the instructional gap (Boaz, Elliott, Foshee, Hardy, Jarmon, & Olcott, 1999).

Online Learning/Course: A context for learning in which students interact using technology and do not meet in a physical classroom with the instructor.

Synchronous Learning: Adjective used to describe an operation performed at the same time as another event (Boaz, et al., 1999).

Web-Based Instruction: A media-rich online environment allowing people to interact with others asynchronously or synchronously in collaborative and distributed environments (Harasim, 1995), to gain access to remote multi-media databases for active and resource-based learning (Jung & Lee, 1993), and to manage self-paced individual learning in a flexible way (Reeves & Reeves, 1997).

Increasing Student Engagement and Participation Through Course Methodology

C

T. Ray Ruffin*University of Phoenix, USA & Grand Canyon University, USA & Ashford University, USA & North Carolina Wesleyan College, USA***Donna Patterson Hawkins***University of Phoenix, USA***D. Israel Lee***Southern Illinois University, USA & University of Phoenix, USA*

INTRODUCTION

Educators have been criticized for limited resources in building capacity for continued growth through reflective learning (Demirbulak, 2012). A large number of educational institutions concentrate on acquiring formal knowledge of the subject matter excessively, with theory being overstated and the practical realities of the classroom not being effectively addressed. This is a formidable challenge that educators face and has led to research in trying to develop lifelong learning by being reflective. Reflective education is observing one's own educational practices and thinking about what works and evaluating the approach to see if it is effective or not (Demirbulak, 2012).

Reflection bridges the chasm between theory and practice enabling educators to apply their knowledge for effective learning that is vital for success. This chasm could be the cause of educators avoiding uncertain, possibly rewarding, activities and ignoring cognitive activities (Edwards & Protheroe, 2003). Educators tend to prefer the more traditional teaching methods that are teacher-centered and express anxiety with the perception of student-centered learning (Ozgun-Koca & Sen, 2006). Concurrently, other factors (e.g., poor classroom management, limited knowledge of correlated subjects, insecurity) limit the opportunity for a more rich educational experi-

ence (Demirbulak, 2012). Even with these factors considered, it is the educator's responsibility to design courses that utilize reflective educational practices that are effective (Demirbulak, 2012). Therefore, dynamic educational approaches are necessary and should be integrated in a well-balanced manner. Practices that are learner-focused increase the probability of success by increasing student engagement in the classroom.

According to Günüç and Kuzu (2014a), the increase use of technology is facilitating greater learner participation and is a boon to the academic environment. Günüç and Kuzu posit that it is essential to take technology into consideration in the classroom due to the near universal acceptance of technology into every aspects of life. Digital natives, individuals surnamed because they were born in an era immersed in the technology, see this technology (e.g., iPads, Surface Pros, Smart Phones...etc.) as normal. This particular demographic of learner differ from other generations in this respect (Günüç & Kuzu, 2014a). For the adult learner (historically referred to as non-traditional learner), likes the convenience the technology provides (e.g., access to education and coursework virtually 24 hours a day, from any location). However, all adult learners are not the same, do not learn the same, and desire a more dynamic educational approach. Despite this perspective, a singular approach to the learning process tends

DOI: 10.4018/978-1-5225-2255-3.ch126

to dominate educational practices—lecture. This myopic approach is based on principles, theories, and teaching methodologies that have existed for centuries that do not take into account the diversity and cultural differences of learners. Concomitantly, educators and educational systems are responding to the needs of learners with a combination of educational methodological approaches—pedagogy and andragogy.

Pedagogy uses prescribed subject matter with little room for deviation, it is supported by external motivation factors such as grades to promote engagement and learners are dependent upon the instructor to determine how much and how well they learn (Educational Technology and Mobile Learning, 2015). Andragogy, uses technological tools to enhance learning for adults, encourages self-directedness, and incorporates personal experiences into the learning process (Educational Technology and Mobile Learning, 2015). This chapter investigates the aforementioned methodologies found to be effective for adult learners. The chapter will consist of the following sections: background; issues controversies and problems; solutions and recommendations; future research directions; and the conclusion. The information contained herein will enhance the field of education and course methodology and influence modern educational institutions and society in general.

BACKGROUND

Foundational theorist such Immanuel Kant explored education in, *On Pedagogy* (Über Pädagogik). Here, the image of Greek pedagogues walking alongside their charges, or sitting with them in classrooms, represents the quintessential picture of exemplary education. Eisner (1979, 1985, 1994) argued that the ability to reflect, imagine, and respond involves developing “the ideas, the awareness, the skills, and the mind to create work that is well balanced, adeptly accomplished, and inventive, regardless of the domain in which an individual works ” (this needs a page reference).

Schön’s (1983) work on reflective practice and his critique of the sort of “technical rationality” crudely employed within more “scientific” approaches to practice has been influential (where and/or how?).

This approach to education and thinking is consistent with pedagogues and their ability to reflect, make judgements, and respond (Smith and Smith, 2008). Alegria (2014) found that critical pedagogy literature emphasized the following themes:

- Teaching that focuses on the students ‘culture, socioeconomic status, familial connections, and identity (who are they, identity/personal growth).
- Teaching that emphasizes academic skills and knowledge of the student as well as his or her development of critical thinking (academic/cognitive).
- Teaching that develops students’ critical understanding of society, power, the inequality embedded in activities to make the student ware of their role in society, their understanding of inequality (who is in power and why and their own personal power to change their status or role).

Andragogy was first coined in Europe of 1833 by Alexander Kapp. However, many credit Malcom Knowles with the patriarchal role of the evolution and sculpting of andragogy in the United States (“Culture and process of adult learning”, 2013). Malcolm Knowles, an American practitioner and theorist of adult education, defined andragogy as “the art and science of helping adults learn”. Knowles was convinced that adults and children learned differently. This provided the basis for a distinctive field of investigation. Knowles identified the six principles of adult learning as:

- Adults are internally motivated and self-directed.
- Adults bring life experiences and knowledge to learning experiences.

Table 1. Pedagogy vs. andragogy

	Pedagogical	Anagogical
The Learner	The learner is dependent upon the instructor for all learning. The teacher assumes full responsibility for what is being taught and how it is learned. The teacher evaluates learning.	The learner is self-directed. The learner is responsible for their own learning. Self-evaluation is characteristic of this approach.
Role of the Learner's Experience	The learner comes to the activity with little experience that could be tapped as a resource for learning. The experience of the teacher is most influential.	The learner brings a greater volume and quality of experience. Adults are a rich resource for one another. Different experiences assure diversity in groups of adults. Experience becomes the source of self-identity.
Readiness to Learn	Students are told what they have to learn in order to advance to the next level of mastery.	Any change is likely to trigger a readiness to learn. The need to know in order to perform more effectively in some aspect of one's life is important. Ability to assess gaps between where one is now and where one wants and needs to be.
Orientation to Learning	Learning is a process of acquiring prescribed subject matter. Contents units are sequenced according to the logic of the subject matter.	Learners want to perform task, solve problem, and live in a more satisfying way. Learning must have relevance to real-life tasks. Learning is organized around life and work situations rather than subject matter units.
Motivation for Learning	Primarily motivated by external pressures, competition for grades, and the consequences of failure.	Internal motivators, self-esteem, recognition, better equality of life, self-confidence, and self-actualization.

Retrieved from: Educational Technology and Mobile Learning (2015)

- Adults are goal oriented.
- Adults are relevancy oriented.
- Adults are practical.
- Adult learners like to be respected.

Eminent theorists, such as Paulo Freire, talked about pedagogy in relation to working with adults, there are others such as Knowles (1970) who argue that it cannot escape its roots is bound up with practice with children. Table 1 illuminates the differences between pedagogy and andragogy.

Student engagement symbolizes the time and effort adult learners dedicate to activities that are empirically connected to desired outcomes of educational institutions to induce students to participate in these activities (Kuh, 2009): Table 2 is the foundations of engagement historical growth.

Today's students referred to as learners of the 21st century, learners of the new millennium, generation Y, Internet generation, and technological natives are defined by Prensky (2001) as "digital natives" as they are born into technology use and

technology culture. Many have stated for digital natives, technology is much more than a tool because it is considered a way of life and because technological environments are regarded as an ordinary natural environment.

Table 2. The foundations of engagement historical growth as taken from Kuh (2009)

Name	Authors and Time Frames
Time on Task	(Tyler, 1930s),
Quality of Effort	(Pace, 1982)
Student Involvement	(Astin, 1984).
Social and Academic Integration	(Tinto, 1993).
Good practices in Undergraduate Education	(Chickering & Gamson, 1987)
College Impact	(Pascarella, 1985)
Student Engagement	(Kuh, 2009)

INCREASING STUDENT ENGAGEMENT AND PARTICIPATION THROUGH COURSE METHODOLOGY ISSUES, CONTROVERSIES, PROBLEMS

When educators discuss learning, they use the word pedagogy. At its root, pedagogy refers to “leading children”, this concept can be misleading due to the effectiveness of the approach on undergraduate students. The apparent reference to the science of leading children seems contrary to preparing a learner for an adult world that is evolving technologically (Batson, 2011). There is definite need to understand how adults learn and design the undergraduate experience accordingly. A primary function of critical pedagogy is to critique, expose, and challenge the manner in which education impacts the political and cultural life of students. Educators need to recognize how colleges and universities bond knowledge and power, and how they can work to influence or impede the formation of critically thinking and socially active individuals (Darder, 1995).

Adult learners’ attentiveness, perceptions and willingness are fairly important for academic achievement and attendance. Their engagement helps them benefit efficiently from courses. Krause and Coates (2008) related student engagement to high quality in learning outcomes. Both the learner and the institution, the two sharers of education, are responsible for increasing the level of student engagement (Trowler, 2010). Adult learner engagement is a sociological and psychological concept (Kahu, 2013). Gunuc and Kuzu (2014b) defined student engagement as the quality and quantity of learners’ psychological, cognitive, emotional and behavioral reactions to the learning process in or out of class academic and social activities to achieve successful learning outcomes. The traditional perceptions of education claims to be impartial and apolitical, pedagogy interprets all educational theory as closely linked to ideologies shaped by power, politics, history, and culture (Batson, 2011). Given this view, edu-

cation functions as an environment of continuing struggle over what will be accepted as legitimate knowledge and culture. Pedagogy must earnestly address the concept of cultural politics by legitimizing and challenging cultural experiences that encompass the histories and social realities that in turn comprise the forms and boundaries that give meaning to student lives (Kelly, 2014).

The body of research about learning and pedagogy is evolving. Educators are still of the opinion of not knowing even the rudimentary theoretical terminology to use, to better comprehend the changes and make knowledgeable decisions about changes (Batson, 2011). Clearly, we have moved from knowledge stability and entered a time of extraordinary change, with Web 2.0, a term coined in 2004, is a description of the new Web architecture (Batson, 2011). It is historic marker between the era of contented stability and the era of disconcerting change. Many educators have accordingly turned to learning and away from teaching. Most educators are unaware of the degree of change necessary or the degree to which change will continue over the coming years. Educator in higher education have been searching around the edges of learning research for decades, and have dealt daily with the issues of learning (Batson, 2011). One research thread that seemed to lead to a rich stratum of ideas about learning started with a Google search of the term “situated cognition”. Situated cognition, and related research threads, seems a useful concept for beginning to understand the tendencies of information technology for teaching and learning (Batson, 2011).

Andragogy has produced many scholarly efforts with puzzling empirical investigations. Debates abound over an accurate and consistent definition for “andragogy”. Educator’s definitions range from “adult leader” to “leader of adults” to “learning of adults” which has caused skepticism (“Culture and process of adult learning”, 2013). Over the years, dubious empirical research has existed testing the reliability and validity of the theory of andragogy. Unlike pedagogy, which is known for teacher centered learning; andragogy

has an elasticity of meaning. Perhaps there are essential actions researchers can examine to help validate andragogy's long-standing theory. Rachal (2002) postulates seven criteria to help put the theory of andragogy into perspective:

- **Voluntary Participation:** The andragogy researcher should examine or design learning situations in which the adult learner wants to participate for her own personal fulfillment or some other internal motivator.
- **Adult Status:** Future andragogy studies should avoid college settings if the various groups being compared are partly comprised of traditional college students.
- **Collaboratively-Determined Objectives:** The andragogy researcher should examine or design learning situations in which, the adult learner plays a significant or even primary role in the determination of the learning objectives.
- **Performance-Based Assessment of Achievement:** When the purpose of the analogical learning experience is primarily proficiency or competence in a content area, the andragogy researcher must examine achievement.
- **Measuring Satisfaction:** Many adult education activities do not have as their objective the mastery of some content or acquisition of a skill, but rather the inherent pleasure or satisfaction of participating in a learning activity.
- **Appropriate Adult Learning Environment:** Future andragogy studies should make every attempt to insure that both the physical and the psychological environments are as congruent as possible with Knowlesian guidelines for adult learning settings.
- **Technical Issues:** Ideally, random assignment of participants should occur, but the realities of adult education research are such that in situ groups are the norm and should be considered acceptable (p. 219-224).

to disengage students from learning and create conflict between adult learners and the educational institution (Pedró, 2006).

SOLUTIONS AND RECOMMENDATIONS

In recent years, educators have drawn attention to such problems as low level of student engagement, dropout at early age, poor student behavior, and low level of academic achievement (Harris, 2008). Unfortunately, inflexible, intolerant, or arrogant educators who were successful using each of the approaches, purposefully or inadvertently, created a learning situation where pedagogy and andragogy were siloed; and not considered for the combination possible in a blend of the approaches. However, Strohschen and Elazier (2011) and Peterson and Ray (2013) suggest there is an educational approach that combines the advantages of pedagogy and andragogy to benefit the learner by focusing intensely on the needs of the learner. Educators and adult learners come from two generations, which have two different technological cultures, coming to conflict in a classroom environment. Some adult learners believe that the Internet has an instructional potential than educators (Kolikant, 2009). On the other hand, educators believe students are non-attentive in classes and students are reluctant to learn. Prensky (2001) suggests technology may be the bridge needed to connect the two generational differences. One where educators understand the language of digital natives and use methodological approaches that are considerate of the adult learner, as educational institutions become a central part of adult learners' lives (Kolikant, 2009).

Educators having zeal to educate across the human spectrum must include an understanding of how to maintain educational effectiveness. Scholarly proliferation of educational philosophies, theories, strategies, and approaches produced by the introduction of prefixes attached to agogos, with accompanying academic support, has created learning situations across the anthropological spec-

trum. Pedagogical processes positions learners as having limited knowledge and no life experiences as they relate to the subject matter for the acquisition of knowledge. Andragogical processes positions learners as knowledgeable, responsible, and self-directed. This perspective assumes that adult learners will be more receptive to the learning process and will experience a higher level of engagement than adult learners who are educated using the aforementioned pedagogical approach. By connecting disciplines, in part or entirely, the transdisciplinary educator demonstrates the utility of the information contained in the course in a myriad of life situations that are useful inside and outside of the classroom. This aspect of the educator's learning approach is consistent with the andragogical approach and contains characteristics of pedagogical methodology.

McCaslin and Scott (2012) promoted that metagogy is an independent-interdependent approach to the learning process with the educator functioning as the "potentiator" of the learning process. McCaslin and Scott (2012) suggest that the educator creates the possibility for the acquisition of knowledge, with the adult learner ultimately responsible for learning. In the metagogical approach, the educator switches between pedagogy and andragogy to create a learning environment that compensates for the adult learner's low level of knowledge in an area and promote self-directedness in areas where the adult learner is proficient. In this environment, the transdisciplinary educator is the catalyst for the learning process and depends on the adult learner being intrinsically driven to explore the richness of the learning experience. According to Blaschke, (2012)) heutagogy is defined as the study of

self-determined learning, as a progression from earlier educational methodologies to a point where there is an interdependent learning relationship between the educator and adult learner. The complex, adaptive, and self-determined nature of the adult learner, the environment, and the situation creates a learning situation that can be exciting (Canter, 2012). The heutagogical approach uses intrinsically motivated principles with an understanding that the extrinsic impetus comes from the completing the task at hand Blaschke, (2012).

Pedagogical, andragogical, metagogical, and heutagogical educational approaches to the learning process are individually appropriate and effective for facilitating the engagement of the adult learner. When using these processes to educate adult learners, there are active and passive period of interactions that occur between the adult learner and educator in the course of the acquisition and transfer of information that will constitute learning. Table 3 is an example of the interactions, dependent upon the level of learning; the adult learner will be directed or determined in the education process

In order to select the proper educational approach to facilitate learning for the adult learner, a proper evaluation of the adult learner's level of preparedness has to be identified. Amy Herman, author of *The Art of Perception*, advocates the "Four A's" to evaluate most education and training situations:

1. Assess.
2. Analyze.
3. Articulate.
4. Adapt (Herman, 2015).

Table 3. An example of the interactions taken from DiLullo, C. McGee, P. Kriebel, R. M. (2011)

Type	Adult Learning Orientation	Drivers of Education
Pedagogy	Passive Receiver	Educator Directed
Andragogy	Active Receiver-Passive Transmitter	Educator-Learner Directed
Metagogy	Active Transmitter-Active Receiver	Educator-Learner Determined
Heutagogy	Active Transmitter	Learner Determined

During the assessment phase, the focus is to determine the knowledge, skill, and aptitude of the adult learner. This can be accomplished using any number of procedures (e.g., formative assessment, questionnaire, general conversation, classroom interaction) via any number of methods of delivery (Canter, 2012). Essential to this process is suspending judgment of the information until the analysis phase; therefore, all information is considered. Herman (2015) analytical approach to the educational process is relative to the educator for two reasons:

1. To unpack assumptions.
2. To select an educational approach most beneficial for the adult learner.

In the analysis phase, the information received is critically examined to determine what information is necessary and what information should be discarded. The essential information is segregated and extracted from the non-essential information and weighed against the need(s) of the adult learner to ensure a reasonable degree of certainty that the approach considered is commensurate with needs of the adult learner.

FUTURE RESEARCH DIRECTIONS

In today's global economy, it is necessary for workers and adult learners to have critical thinking skills, creative problems solving abilities, and self-direction to initiate change for improvement in education, workplace practices and life. Self-directedness is essential to workers and learners in the 21st century due to the rapidly proliferation of technology and the changing nature of the workplace. According to Guglielmino (2014), it is nearly impossible for the educational and curriculum designers to remain current on what exactly is required of the adult learners in the classroom without the self-determined nature inherent in the learning processes. Holford and Van der Veen (2003) reported that in European

countries (i.e., Spain, Finland, and Belgium) the desire is to have secondary educators instill adult learners with the ability to think and reason for themselves in order to become a part of growing responsible citizenry. According to Freidman (2005) with more and more companies relying on a "flat" hierarchical arrangement in the workplace, it is imperative that adult learners and workers inculcate a self-directed philosophy if a company is to remain viable in the 21st century.

The information contained in this report supports need for lifelong and advanced learning by workers and adult learners in the 21st century. Livingstone and Rakov (2013) suggest professional and industry workers in Canada depend on certification, recertification, and continuous learning to maintain proficiency and employment. The 21st century adult learners and professionals in growing and emerging fields bring prior knowledge, life and work experience, and expert knowledge to the learning environment that are integral components of advance learning processes (i.e., metagogy and heutagogy). Concurrently, Formosa (2014) writes of the coming of the "third age" learners. These are people are near retirement, retired, or past the age of employment, however they engage in formal and informal education. They actively seek the engagement, companionship, and continued mental growth in educational environments. Through the metagogical and heutagogical approaches broached in the classroom, adult learners have the opportunity to provide critical inquiry and expect resolution and/or inclusion to assuage the concerns on the adult learners and provide effective instruction in the classroom.

CONCLUSION

Clearly, we have moved from knowledge stability and entered a time of extraordinary change, with Web 2.0, a term coined in 2004. It is a historic marker between the era of contented stability and the era of disconcerting change. Many educators

have accordingly turned to learning and away from teaching. Most educators are unaware of the degree of change necessary or the degree to which change will continue over the coming years. In recent years, Educators have drawn attention to such problems as low level of student engagement, dropout at early age, poor student behavior, and low level of academic achievement. Scholarly proliferation of educational philosophies, theories, strategies, and approaches produced by the introduction of prefixes attached to agogos, with accompanying academic support, has created learning situations across the anthrological spectrum. There is a critical need for educational approaches that are adult learner-focused and utilitarian across cultures. Implementing education where adult learners are involved with and have input into the content and execution of the course is appealing in both reason and utility. Through the metagogical and heutagogical approaches broached in the classroom, adult learners have the opportunity to provide critical inquiry and expect resolution and/or inclusion to assuage the concerns on the adult learners and provide effective instruction in the classroom.

REFERENCES

- Alegria, A. (2014). Supporting English Language Learners in the Science Classroom through Critical Pedagogy. *International Journal of Science and Mathematics Education*, 12(1), 99–121. doi:10.1007/s10763-013-9399-7
- Batson, T. (2011). *The Problem of 'Pedagogy' in a Web 2.0 Era*. Retrieved from: <http://campustechnology.com/articles/2011/06/15/the-problem-of-pedagogy-in-a-web-2.0-era.aspx>
- Blaschke, L. M. (2012). Heutagogy and lifelong learning: A review of heutagogical practice and self-determined learning. *The International Review of Research in Open and Distance Learning*. Retrieved from <http://www.irrodl.org/index.php/irrodl/article/view/1076/2087>
- Canter, M. (2012). *E-Heutagogy for life-long learning*. Retrieved from; doi:10.1016/j.protcy.2012.02.025
- Darder, A. (1995). Buscando America: The Contributions of Critical Latino Educators to the Academic Development and Empowerment of Latino Students in the U.S. In *Multicultural Education, Critical Pedagogy and the Politics of Difference*. New York: SUNY Press.
- Demirbulak, D. (2012). Fostering Propensities for Reflective Practice at Undergraduate Methodology and Practicum Courses. *Eurasian Journal of Educational Research*, 48, 39–58.
- DiLullo, C., McGee, P., & Kriebel, R. M. (2011). *Demystifying the Millennial student: A reassessment in measures of character and engagement in professional education*. *Anatomical sciences education*. doi:10.1002/ase.240
- Educational Technology and Mobile Learning. (2015). *Awesome Chart on Pedagogy VDS Andragogy*. Retrieved from: <http://www.educatorstechnology.com/2013/05/awesome-chart-on-pedagogy-vs-andragogy.html>
- Edwards, A., & Protheroe, L. (2003). Learning to see in classrooms: What are student teachers learning about teaching and learning while learning to teach in schools. *British Educational Research Journal*, 29(2), 227–240. doi:10.1080/0141192032000060957
- Eisner, E. W. (1979, 1985, 1994). *The educational imagination: on the design and evaluation of school programs*. New York: Macmillan.
- Formosa, M. (2014). *Four decades of Universities of the Third Age: Past, present, future*. Retrieved from <http://search.proquest.com.contentproxy.phoenix.edu/docview/1465054379?pq-origsite=summon&accountid=35812>

- Guglielmino, L. (2014). *Guglielmino: Self-Directed Learning-What About the Adults?* Retrieved from <http://www.p21.org/news-events/p21blog/1475-guglielmino-self-directed-learning-what-about-the-adults>
- Gunuc, S., & Kuzu, A. (2014a). Student engagement scale: Development, reliability and validity. *Assessment & Evaluation in Higher Education*. doi:10.1080/02602938.2014.938019
- Günüç, S., & Kuzu, A. (2014b). Factors Influencing Student Engagement and the Role of Technology in Student Engagement in Higher Education: Campus-Class-Technology Theory. *Turkish Online Journal of Qualitative Inquiry*, 5(4), 86–113. doi:10.17569/tojqi.44261
- Harris, R. (2008). A Phenomenographic Investigation of Teacher Conceptions of Student Engagement in Learning. *Australian Educational Researcher*, 35(1), 57–79. doi:10.1007/BF03216875
- Hase, S., & Kenyon, C. (2007). Heutagogy: A child of complexity theory. *Complicity: An International Journal of Complexity and Education*. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.397.5180&rep=rep1&type=pdf>
- Herman, A. (2015). *Seeing what matters*. Boston, MA: Houghton Mifflin Harcourt.
- Holdford, J., & Van der Veen, R. (2003). *EU research of social sciences and humanities: Life-long learning, governance, and active citizenship in Europe*. Retrieved from http://ec.europa.eu/research/social-sciences/pdf/etgace_en.pdf
- Kahu, E. R. (2013). Framing Student Engagement in Higher Education. *Studies in Higher Education*, 38(5), 758–773. doi:10.1080/03075079.2011.598505
- Kelly, T.J. (2014). *Students as investigators, teachers as researchers: documenting a critical history pedagogy and its impact on diverse learners in a tenth-grade world history classroom* (Doctoral Dissertation). James Madison University. Retrieved from <http://www.jamesmadison.gov/pdf/Madison%20Fellows/TimothyKelly.pdf>
- Kolikant, B. (2009). Digital Students in a Book-Oriented School: Students' Perceptions of School and the Usability of Digital Technology in Schools. *Journal of Educational Technology & Society*, 12(2), 131–143.
- Krause, K., & Coates, H. (2008). Students Engagement in First-Year University. *Assessment & Evaluation in Higher Education*, 33(5), 493–505. doi:10.1080/02602930701698892
- Kuh, G. (2009). What Student Affairs Professionals Need to Know About Student Engagement. *Journal of College Student Development*, 50(6), 683–706. doi:10.1353/csd.0.0099
- Livingstone, D. W., & Raykov, M. (2013). *Adult learning trends in Canada: Basic findings of the WALL 1998, 2004, and 2010 surveys*. Center for the Study of Education and Work. Retrieved from <http://wall.oise.utoronto.ca/Adult-Learning-Trends-in-Canada-2013.pdf>
- McCaslin, M., & Scott, K. W. (2012). *Building communities of potential through teaching and learning: the nature of metagogy*. Unpublished manuscript. Retrieved from http://h12.cgpublisher.com/proposals/234/index_html
- Ozgun-Koca, S. A., & Sen, A. I. (2006). The belief and perceptions of pre-service teachers enrolled in a subject area dominant teacher education program about effective education. *Teaching and Teacher Education*, 22(7), 946–960. doi:10.1016/j.tate.2006.04.036
- Peterson, C. M., & Ray, C. M. (2013). Andragogy and metagogy: The evolution of neologisms. *Journal of Adult Education*, 42, 80–85.

- Prensky, M. (2001). Digital natives, digital immigrants. *On the Horizon*, 9(5), 1–5. doi:10.1108/10748120110424816
- Rachal, J. R. (2002). Andragogics detectives: A critique of the present and a proposal for the future. *Adult Education Quarterly*, 52(3), 210–227. doi:10.1177/0741713602052003004
- Schön, D. (1983). *The Reflective Practitioner: How professionals think in action*. London: Temple Smith.
- Smith, H., & Smith, M. (2008). *The art of helping others: Being around, being there, being wise*. London: Jessica Kingsley Publishers.
- Strohschen, G., & Elazier, K. (2011). *The blended shore concept of international/intercultural adult education program and delivery: A disclosure and process*. Retrieved from <http://ahea.org/files/pro2009strohschen.pdf>
- Trowler, V. (2010). Student Engagement Literature Review. York, UK: Higher Education Academy. Available online [http://www.heacademy.ac.uk/assets/York/documents/ourwork/studentengagement/StudentEngagementLiteratureReview.pdf](http://www.heacademy.ac.uk/assets/York/documents/ourwork/studentengagement/St%20udentEngagementLiteratureReview.pdf)
- Hase, S., & Kenyon, C. (2000). From andragogy to heutagogy. *Ultibase Articles*, 5(3), 1–10.
- Kagan, D. M. (1992). Professional growth among preservice and beginning teachers. *Review of Educational Research*, 62(2), 129–169. doi:10.3102/00346543062002129
- McCaslin, M. L., & Flora, J. (2013). *Living, learning, and leading within the integral space: Energizing integral leadership through experiential learning*. Integral Leadership Review.
- Merriam, S. B. (2001). Andragogy and self-directed learning: Pillars of adult learning theory. *New Directions for Adult and Continuing Education*, 2001(89), 3–14. doi:10.1002/ace.3
- Nason, D. S. (2011). Crisis Drift: A Meta-gological Glue for Learning & Teaching Design. *Design Principles & Practice. International Journal (Toronto, Ont.)*, 5(6).
- Nix, J. V., Lion, R. W., Michalak, M., & Christensen, A. (2015). Individualized, Purposeful, and Persistent: Successful Transitions and Retention of Students at Risk. *Journal of Student Affairs Research and Practice*, 52(1), 104–118. doi:10.1080/19496591.2015.995576

ADDITIONAL READING

- Ashton, J., & Elliott, R. (2007). Juggling the balls—study, work, family and play: Student perspectives on flexible and blended heutagogy. *European Early Childhood Education Research Journal*, 15(2), 167–181. doi:10.1080/13502930701321378
- Astin, A. W. (1984). Student involvement: A developmental theory for higher education. *Journal of College Student Personnel*, 25(4), 297–308.
- Dunleavy, J., & Milton, P. (2009). *What did you do in school today? Exploring the Concept of Student Engagement and its Implications for Teaching and Learning in Canada*. Toronto: Canadian Education Association.

KEY TERMS AND DEFINITIONS

Agogos: From the Greek *agein* “to lead” or “leader of.”

Anthropological Spectrum: The process of human learning across a continuum of educational approaches.

Heutagogical Approaches: Educational approaches where the self-directedness of the learning process allows the learner the flexibility to adjust the learning situation to meet the changing nature of the educational or social phenomena.

Metagogical Approaches: Educational approaches that combine the advantages of pedagogy and andragogy to increase the effectiveness of the educational process.

Potentiator: A person that creates learning opportunities.

Siloed: “Silos” are organizations where each department concentrates on their own objectives. The departments are not working for the good or the whole of the organization. An extreme example of this negative impact of what “silos” can do in an organization is when one department sabotages another. For the purpose of this “soloed” is sepa-

rating pedagogy and andragogy and not using the methodology together.

Third Age Learners: People who are near retirement, retired, and/or past the age of employment but engage in formal and informal education.

Transdisciplinary: The framework for allowing members of an educational team to contribute knowledge and skills, collaborate with other members, and collectively create knowledge.

C

Instructional Real World Community Engagement

Caroline M. Crawford

University of Houston – Clear Lake, USA

INTRODUCTION

The imperative needs towards the integration of real world community engagement throughout the instructional process is inherent within the instructional design shifts that are occurring. The Digital Age has introduced the ability to more fully engage the larger community realm within the instructional process, engaging real world professionals, personal relationship engagements and subject matter experts towards supporting the learner's conceptual understandings of the information, cognitive load and self-regulatory concerns of the learner, as well as a recognition that there is a real world connection between learned information and real world implementation of information.

Instructional efforts are important towards developing a knowledge base of subject matter understanding, as well as working with the information in new and different ways so as to further develop a level of knowledge prowess and engage with the information. Yet towards reaching higher order thinking skills (Aini, Nor & Razak, 2015; Anderson & Krathwohl, 2001; Bloom, 1956, 1984; Bloom, Englhart, Furst, Hill & Krathwohl, 1956; Bloom & Krathwohl, 1956; Krathwohl, Bloom & Masia, 1964; Mishra & Kotecha, 2012, 2016; Ong, Hart & Chen, 2016; Zohar & Barzilai, 2015), it is necessary and appropriate to implement the information in project-focused endeavors. Although traditionally these have been class-based projects or controlled projects for a distinct clientele, the engagement of the learner's understanding within a larger real world community engagement impacts

the learner's levels of personal and professional motivation, enhanced sense of self-efficacy, as well as self-regulatory efforts that reflect not only course-based efforts but also posing within real world community engagement that may include familial, personal and professional communities of engagement. The impact of real world community engagement upon the instructional environment reflects strongly upon learner motivation, external-to-course knowledge checks and analytical feedback, as well as enhancing the positive self-efficacy of the learner internal to and external to the course environment, with the ultimate impact of the instructional experience extending far beyond the requisite course session designations.

BACKGROUND

Discussions surrounding the important elements that impact the instructional environment are worthy of consideration, as the knowledge base within which this real world community engagement occurs is vitally important. The importance and impact of the real world community engagement within the instructional realm is supported by an understanding of cognitive taxonomies, conceptual frameworks of understanding, social discourse, active instructional engagement and implicit cognitive vulnerability that undergirds the holistic understanding of the distinctive importance associated with instructional real world community engagement.

Cognitive Taxonomies

Bloom's Taxonomy of the Cognitive Domain (Aini, Nor & Razak, 2015; Anderson & Krathwohl, 2001; Bloom, 1956, 1984; Bloom, Englhart, Furst, Hill & Krathwohl, 1956; Bloom & Krathwohl, 1956; Krathwohl, Bloom & Masia, 1964; Mishra & Kotecha, 2012, 2016; Ong, Hart & Chen, 2016; Zohar & Barzilai, 2015) engages in a thought process wherein the instructional process must begin at the knowledge level of learner engagement with the subject matter and then slowly progress through each subsequent level of knowledge engagement and informational understanding (i.e., beginning at the knowledge level, progressing through comprehension, application, analysis, synthesis and evaluation) until the learner has the opportunity to achieve the highest level of the higher order thinking skill capability, specified as the evaluation level of informational engagement. Of interest is Anderson and Krathwohl's (2001) redesigned Taxonomy of the Cognitive Domain for the Digital Age, focusing upon the learner's engagement with the understanding as designated by the lowest level of knowledge engagement (i.e., remember) and then progressing towards the highest level of learner's cognitive engagement with the information (i.e., from remember, and progressing through understand, apply, analyze, evaluate, and create as the highest level of informational engagement and thought process).

The importance of Bloom's Taxonomy, including Anderson and Krathwohl's revised taxonomy, considers the viable considerations revolving around exactly how the learner has the opportunity to work with and to more fully understand the subject matter information through progressively more engaged and analytical aspects of engagement. Towards considering the learner's engagement with this information, it is important to consider the real world implications and intimations, not only how this information may be successfully implemented within the real world environment but also how this new subject matter information begins to fit and frame within a

learner's prior subject matter understandings that also include a learner's conceptual framework of understanding.

Conceptual Frameworks of Understanding

Vygotsky's theoretical delineation of a learner's conceptual framework of understanding (Barwell, 2015; Esteban-Guitart, 2015; Metraux, 2015; Montealegre, 2016; Vygotsky, 1933/1966, 1935, 1981; Wang, 2015) critically frames and supports the understanding that the learner brings forward prior knowledge and understanding of information into any new learning experience. The importance of the learning effort is not only towards understanding the subject matter, but also towards appropriately integrating the new information within prior knowledge and concepts, also referred to as conceptual frameworks of understanding. By integrating the social discourse and knowledge checking efforts, inclusive of real world community engagement, the conceptual frameworks of understanding are more strongly engaged within the learner's informational attainment.

Social Discourse

The importance of social discourse was embraced by Wittgenstein (1961) due to his research that reflected a focused interest upon the importance of social engagement and word choice to describe and reflect the learner's understanding of the subject matter (Burr, 2015; Gorski, 2016; Gupta, 2016; Inwood, 2015; Shotter, 2015). Although subject-specific terminology is vital towards understanding the subject matter and respective field of informational engagement, the social aspects as regards communicating about the subject matter and word choices that more holistically reflect the learner's attainment of subject matter information within a bounded engagement level reflects a learned understanding and effort. With consideration towards real world community engagement, social discourse with individuals within

the real world community is vitally important for the learner, not only towards implementing the terminology appropriately but also towards understanding the social discourse that is appropriate and necessary towards more fully engaging with the subject matter and developing a more fully active understanding as regards how the subject matter more directly impacts the real world environment, no matter whether within the professional community of expertise but also within the social and global community of understanding that is beyond the bounds of the “four walls of the classroom”. This active social discourse within a real world community is supported by the concept of active instructional engagement.

Active Instructional Engagement

Instructional efforts and understandings have shifted over the decades, as new and differentiated learning theories have impacted the instructional process. As well, the introduction of technology within the global community has directly impacted the instructional process through not only instructional technology integration into the physical learning environment, but the introduction of disruptive instructional environments such as blended learning, online learning, mobile learning and flipped classroom conceptions have truly unbound the traditional understanding of the instructional environments. However, actively engaging with the subject matter is understood as being integrally important, reflecting a concept termed active instructional engagement. Within active instructional engagement, tactics that have been highlighted include: Tiered Learning; Responsive Design; Immersive Scenarios; Mixed Media Approach; Micro-learning; and, Gamification.

A tiered learning approach breaks the learning process into conceptually relevant areas of understanding. The initial approach is an audio or a video-based visually engaging understanding of the subject matter; one may consider this to be a focused overview of the subject matter to be

learned, so that the learner can understand the progressive steps towards achieving an understanding of the subject matter. A next step in this approach is textually-framed information that is offered as a detailed understanding and representation of the information to be learned. The third component of the tiered learning approach is the ability to work with the new information, to engage and apply the information within a practice-style effort. The importance is the content, but instead of everything focusing upon the content it is a progressive, cognitively conceptual approach that slowly presents information to the learner and then expects the application of the learned information in new and different ways, such as through higher order thinking skills realms (Aini, Nor & Razak, 2015; Anderson & Krathwohl, 2001; Bloom, 1956, 1984; Bloom, Englhart, Furst, Hill & Krathwohl, 1956; Bloom & Krathwohl, 1956; Krathwohl, Bloom & Masia, 1964; Mishra & Kotecha, 2012, 2016; Ong, Hart & Chen, 2016; Zohar & Barzilai, 2015).

Responsive design is an instructional approach that is embraced within the digital age approach to computer mediated learning efforts. This is due to the ability of the technology-interface environment to slowly progress in a breadth and depth of understanding and engagement with the subject matter, normally increasing access to the information based upon the decision of the learner. As well, the learner also chooses when to engage with the information within a more expansive environmental community.

Immersive scenarios support the learner’s progressive engagement with the subject matter that is normally a process-oriented approach or may engage in an understanding of soft skills subject matter, with the primary approach being that multiple branching choice structures systematize one’s subject-based understandings of learner choices within the immersive scenario offering a progressively responsive learning path towards a designated outcome. As such, instead of supporting the learner’s understanding of information in a behaviorally laden process-oriented step-by-step

approach, the immersive scenarios more fully engage the learner in the progressive decision making process and resulting outcomes of decisions made by the learner. Such immersive scenarios are more focused upon higher order thinking skill choices such as analysis, evaluation, synthesis and creativity, as a learner's social engagement and conceptual frameworks of understanding support not only the understanding of learned subject matter but also the external, real world implications associated with choices that impact the immersive scenario outcomes.

A mixed media approach suggests an active instructional engagement, wherein an overarching theme is implemented throughout the instructional environment, but a differentiated approach is brought forward to support and reflect that theme for the learner. What this suggests is that graphics, audio, video, simulations, animations and innumerable modeling efforts reflect the learning efforts thematic support that is meant to more fully support the learner's conceptual framework of understanding and implicit memory attainment as regards the subject matter under study. As suggested within a mixed media approach's focus upon overarching themes, a semiotic approach is inherent within this framework. A semiotic approach is the study and understanding of meaning-making within a humanist perspective; the study and understanding of meaningful communication and the processes understood by signs and symbols is inherent within this semiotic approach (Chandler, 2014; Chang, 2007; Semiotic Society of America, n.d.; Streeter, 2012a, 2012b).

Micro-learning is the concept of learning modules, within cognitively aware and relevant short modules of subject matter understanding. This concept is also supported by the tiered learning approach, wherein a progressive understanding of the subject matter must be realized so as to ensure that the learner's cognitive conceptualization is supported in a progressively engaged and cognitively sensitive understanding revolving around higher order thinking skills understanding.

Finally, the gamification of information is an effort that has been embraced by instructional technologists. Even as the recognition of gamification has heightened and then waned over the years, the progression from cardboard-based floor games towards cognitively structured pseudo-simulation environments that have become significantly easier to develop within recent years, has again risen the concept of gamification within the distributed learning environments (electronic learning, or elearning) and mobile learning environments. Much as with the conception of Massive Open Online Courses (MOOCs) the short-term behavioral rewards system concept of coins, badges, medals and short term rewards structures that are conceptualized as meaningful for different learner personalities have also been impacting gamification efforts.

Active instructional engagement is really framed as different ways through which to implement instructional efforts, towards reaching different types and personalities of learners. Within bounds, each of these instructional efforts is relevant and worthy of consideration. Towards considering the learner within this effort, the subject matter information may be more fully tested and understood within a real world community environment.

REAL WORLD COMMUNITY ENGAGEMENT

The ultimate effort associated with the learning process, including subject matter information attainment and understanding, is towards not only learning for the love of learning nor towards a more well-rounded citizenry, but in fact the ability to understand the subject matter within the real world community environment is an ultimate understanding and desire to successfully reach this level of subject matter understanding and attainment by each learner engaged in the instructional process. The ability to implement new information within a

grounded real world environment is important, not only towards more fully understanding the learned information but also towards more fully recognizing the worth and impact of the subject matter within a real world community engagement. For this reason, discussions revolving around learner-focused considerations of real world community engagement, knowledge checks occurring within real world community engagement, and types of real world community engagement are offered.

Learner-Focused Considerations

Real world community engagement is vitally important within the learning process for several different reasons. First is the concept of personal and professional motivation, wherein a recognition that learners may engage differently with the learning process and deeply engaging with the subject matter when motivational considerations (Maslow, 1977, 1986, 1997), including Vroom's expectancy theory (1964; Chen, Ellis, Suresh, 2016; Holdford & Lovelace-Elmore, 2001; Hung, Zhuang, & Lin, 2015; Lazaroiu, 2015; Parsons & Geoff, 1978; Purvis, Zagenczyk, & McCray, 2015; Shweiki, Martin, Beekley, Jenoff, Koenig, Kaulback, Lindenbaum, Patel, Rosen, Weinstein, Zubair, & Cohen, 2015; Whittington, 2015) are imbedded. The learner's motivational levels are also inherently important towards a learner's sense of self-efficacy (Bandura, 1977, 1986, 1997), wherein the learner's ability to recognize their own abilities and potentials towards success support the learner's efforts in working with the information, especially in a real world community environment that embraces the opportunity to understand the subject matter within daily usage in new and different manners of understanding. Also of interest is Bandura's (1991) discussion of the impact regarding social cognitive theory upon self-regulation. Conceptually, Bandura suggests that the learner self-influences their success through the ability of a person to self-monitor one's efforts and associated impacts, the ability of a person to judge their behaviors within the bounds of the

environment and associated standards of expectation, as well as the learner's ability to understand the affective influence, or emotional self-reaction within different situations. Primarily, the learner's ability to understand their own actions and reactions within any style of knowledge attainment and information expansion environments is viable and significantly important towards the learner's ability to successfully engage and more fully embrace opportunities towards working with the subject matter information in new and real world community efforts of engagement.

Knowledge Checks

The real world community engagement is a perfect opportunity for a learner to check their understanding of the knowledge under study, as well as more fully engage with the subject matter information in new and different ways that will more fully develop an understanding and "fill in the cognitive blanks" for the learner. This holistic understanding of the subject matter is vitally important, towards progressing out of lower order thinking skills and engaging in higher order thinking skill efforts. Much as what was described in the tiered learning effort, the progressive efforts associated with framing a tiered approach that begins with a controlled understanding of the information, such as in-course knowledge attainment efforts, then shifts towards working with the information in practice and application efforts, with a shift towards external-to-course real world understandings of the subject matter. This real world community engagement offers knowledge checks and analytical feedback efforts that offer the opportunity to critically analyze and correct any conceptual framework of understanding misalignments that are more readily addressed within a real world environment. Engaging the learners within reflective understandings of the subject matter information within the real world community engagement opportunities supports a learner's ability to critically reflect upon why the shift from classroom knowledge attainment and

skills application and practical application within a controlled environment might be differentiated from a real world community engagement with the subject matter is considered vitally important; the analysis and re-learning understandings more fully engage the learner in their own cognitive framing of understanding, as well as a learner's recognition associated with their own self-efficacy (Bandura, 1977, 1986, 1997; Chen, Ellis, Suresh, 2016; Holdford & Lovelace-Elmore, 2001; Hung, Zhuang, & Lin, 2015; Lazaroiu, 2015; Parsons & Geoff, 1978; Purvis, Zagenczyk, & McCray, 2015; Shweiki, Martin, Beekley, Jenoff, Koenig, Kaulback, Lindenbaum, Patel, Rosen, Weinstein, Zubair, & Cohen, 2015; Vroom, 1964; Whittington, 2015), meaning that the learner recognizes the ability to learn and re-learn information within differentiated contexts while still recognizing the learner's ability to successfully overcome and achieve quality levels of knowledge acquisition, subject matter information implementation and successfully reframing conceptual frameworks of understanding.

Types of Real World Community Engagement

Within an instructional environment, there are innumerable ways through which real world community engagement can be successfully addressed and implemented. Much as within a tiered approach towards learning, a tiered approach towards instructional real world community engagement is realized within an instructionally impactful environment. Within an instructional environment, real world community engagement may fall within three distinct constructs: Academic Communities of Engagement; Personal Communities of Engagement; and, Professional Communities of Engagement. Within a tiered community approach, the primary community may be considered the academic community of engagement, the secondary community may be considered the personal communities of engagement, and the tertiary (although by no means lesser

in importance) community may be considered the professional communities of engagement.

Academic Communities of Engagement

Academic communities of engagement revolve around the conception of co-learners and collegial instructors who are all engaged within an academic effort. This focus is towards not only fulfilling their own academic impetus towards learning but also within a collegial effort that engages within a group or communal ability to work with subject matter in new and different ways.

First may be considered the current course colleagues, who are currently engaged in the same course in which the subject matter is under focused consideration. Each course is focused upon its own subject matter focus, as reflected through instructional goals and instructional objectives; for this reason, current course colleagues are all focused upon successfully attaining specific instructional goals and objectives within the designated course environment.

From a differentiated approach, prior course collegial support systems are in place, based upon collegial learner relationships that have been developed and maintained even after a designated course's instructional experience concludes. These learner colleagues may follow different paths towards their own ultimate academic success, but their bounded approach towards academic pursuits and falling within a specific community environment continues to frame prior course collegial support systems as inherently viable support systems. Not only do the prior course collegial support system members understand the culture and associated quality efforts towards attaining academic success, but prior experiences offer a binding motivational support system within which members find collegial and relational support through recognized successes, adventures, and difficulties.

Current and prior instructors also fall within academic communities of engagement, as mentor-apprentice and professional relationships within

the academic community and academic culture naturally arise. The ability of the learners to develop professional relationships with current and prior instructors is integrally important, not only towards the learner's motivational levels and sense of self-efficacy, but also towards engaging with the subject matter in new and different ways; this is true no matter whether the learner engages with the subject matter within a current instructor's course and desires a deeper analytical approach towards understanding the subject matter, or if the learner desires to work with prior instructors towards engaging in a deeper analytical approach and perhaps a differentiated discussion revolving around the subject matter taught within more current coursework so as to analyze and engage with the subject matter in new and different manners that may not be comfortable nor acknowledged by the current instructor.

The academic community of engagement is a viable community environment due to its primary initial level of importance within a learning process as framed through a traditional view of the learning process.

Personal Communities of Engagement

Personal communities of engagement revolve around the conception of social communities of influence who may not have similar understandings of the subject matter under study within the academic community realm. Such a community supports the learner's self-efficacy and self-regulatory efforts that directly revolve around focused efforts and a desire towards successful engagement in a progressive manner. Within the personal community of engagement, one may specify at least three groups of impact: Friends; Family; and, Social Connections.

Friends are an integrally important real world community of impact upon the learner. The learner has a developed relationship with friends, who have a direct impact upon the learner's cognitive stresses and cognitive load concerns, suggesting that friends hold a prestigious level of support and

impact within a learner's ability to learn, engage and reflect upon the subject matter learned within their prior academic community structure. Within a personal community construct, the learner must feel a sense of accomplishment and quality of attainment, to retain information as well as to sustain their level of cognitive engagement in the learning process.

Family have deep-seated roots that are emotional and bonding in nature, and which also mirror the learner's family-specific labels of success or perhaps detrimental underpinnings. The learner's strong sense of support and self-efficacy may begin within the bounds of the family structure. As a family structure, this personal community of engagement bounds the perceptions of the learner by other's opinions of her/himself as a successful learner, as an intelligent person, as well as engaging in motivational behaviors towards the learner's own successes. The support of a family structure is vitally important towards supporting the learner's underlying self-esteem, self-efficacy and motivational efforts. From a more negative point of view, the learner must reflect upon and cognitively engage in a holistic understanding of their familial community of engagement, so as to recognize and potentially rectify any potential concerns that may result in the learner's blocking of successful attainment of their instructional successes that also include working with the subject matter within their personal communities of engagement.

Personal communities of engagement also include social connections. The types of personal social connections that impact a person on a daily basis directly impact the learner's ability to engage in intellectual and holistically emotive and supporting levels of real world community engagement, that may more fully support the learner's ability to recognize their own skill sets and cognitive strengths that result in social successes and analytical empathetic understandings revolving around different viewpoints associated with subject matter from a cultural and social level of community engagement.

The personal community of engagement is a viable community environment due to its secondary level of importance within a learning process as framed through a cognitive constructivist view of the learning process. The personal, cultural and social impact upon a learner's understanding of the subject matter directly impacts the learner's understanding of the subject matter within a real world community engagement scenario.

Professional Communities of Engagement

Professional communities of engagement revolve around the conception of professional communities of influence who may have differentiated backgrounds of subject matter expertise and understanding, yet the ability to engage with the subject matter within a mentor-apprentice relational understanding strengthens the learner's ability to cognitively structure an understanding of the subject matter beyond the realms of academic learning and personal community influence, towards a real world professional community engagement more fully reflects opportunities towards higher order thinking skills efforts and informational attainment. Within the professional community of engagement, one may specify at least three groups of impact: Current Work Colleagues; Prior Work Collegial Support System; and, Professional Organization Colleagues.

Current work colleagues may be considered the first of three levels of professional communities of engagement. Within this community structure, the learner has the opportunity to fully implement the subject matter information with colleagues who may be equally informed as pertains to the depth and breadth of information that the learner brings forward within a professional community environment. Conceptual frameworks of understanding are integrally important within this community structure, as the ability to offer professionally ap-

propriate checks and balances towards supporting the learner's engagement with the subject matter information is vitally important. From a negative viewpoint, unknowledgeable, uninformed or uncomfortably critical colleagues may irreparably damage the learner without the additional mirroring of acceptability by others from within each of the other real world communities of engagement.

Prior work collegial support systems are integrally important towards a learner's successful integration and implementation of the subject matter within the professional environment. This is not only due to the ability of prior work colleagues to dissuade the learner's beliefs and labels as bound by uninformed or ignorant current work colleague critical actions, but also towards developing differentiated understandings about the learned subject matter information in new and different ways, that are beyond the potential bounds of a current work site effort and into a more holistic understanding of the professional communities of engagement. This reflects not only upon the importance and impact of the subject matter information, but may also further engage and reflect upon the learner's conceptual framework of understanding in an extension of engagement and perception beyond prior bounds of comprehension.

The final of the three professional communities of engagement is defined as the professional organization colleagues. These subject matter professionals offer differentiated levels of expertise with the subject matter information with which the learner is working, laboring and toiling. These professional organization colleagues offer an inherent interest in expanding their own professional understandings as revolves around the subject matter, as well as learning new information. This dynamic style of relational real world community engagement is not only a mentor-apprentice relationship but may shift between a more collegial view of subject matter engagement, questioning and analytical creativity.

SOLUTIONS AND RECOMMENDATIONS

Instructional real world community engagement is an integrally important aspect within any instructional process. Within a tiered learning approach, not only must the subject matter be offered in a carefully framed presentation and cognitively discrete understanding of the learner, but the real world community engagement of the learner must also be understood. This real world community engagement of the learner does offer a direct instructional impact upon the learner's cognitive understanding of the subject matter, conceptual framework of understanding, as well as motivation, self-efficacy and fully engaged understanding of the subject matter within a holistic understanding of the subject matter's real world implementation.

FUTURE RESEARCH DIRECTIONS

More fully framing the developing understandings revolving around the real world community engagement's impact upon the instructional process must be more fully realized within academic pursuits. Not only is the real world engagement with the subject matter inherently vital towards the learner's understanding of the information, but recognizing and reflecting upon the differentiated real world community understandings of the subject matter may develop the learners into more reflective practitioners and professional citizenry.

CONCLUSION

Instructional real world community engagement is vitally integral and important towards structuring a knowledge base of subject matter understanding, while also recognizing the working knowledge impact upon the learners. Through an understanding of the theoretical undergirdings as well as a recognition of the real world community engage-

ment components more fully frame the impact of real world community engagement upon the instructional process.

REFERENCES

- Aini, Z., Nor, N. H., & Razak, K. A. (2015). Generating higher-order thinking skills in Islamic education through Information Communication Technology (ICT). *Turkish Online Journal of Educational Technology*, 2015, 215–221.
- Anderson, L.W., & Krathwohl. (Eds.). (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84(2), 191–215. doi:10.1037/0033-295X.84.2.191 PMID:847061
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Bandura, A. (1991). Social cognitive theory of self-regulation. *Organizational Behavior and Human Decision Processes*, 50(2), 248–287. doi:10.1016/0749-5978(91)90022-L
- Bandura, A. (1997). *Self-Efficacy: The exercise of control*. New York, NY: W. H. Freeman.
- Barwell, R. (2015). Formal and informal mathematical discourses: Bakhtin and Vygotsky, dialogue and dialectic. *Educational Studies in Mathematics*, 1–15.
- Bloom, B. S. (Ed.). (1956). *Taxonomy of educational objectives: The classification of educational goals: Handbook I, cognitive domain*. New York: Longmans, Green.
- Bloom, B. S. (1984). *Taxonomy of educational objectives*. Boston, MA: Allyn and Bacon.

- Bloom, B. S., Englhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (Eds.). (1956). *Taxonomy of educational objectives, the classification of educational goals, Handbook I: cognitive domain*. New York: Longmans.
- Bloom, B. S., & Krathwohl, D. R. (1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals, by a committee of college and university examiners. In Handbook I: Cognitive Domain*. New York: Longman, Green.
- Burr, V. (2015). *Social constructionism*. Routledge. doi:10.1016/B978-0-08-097086-8.24049-X
- Chandler, D. (2014). *Semiotics for Beginners*. Retrieved from <http://visual-memory.co.uk/daniel/Documents/S4B/sem01.html>
- Chang, H.-L. L. (2007). *Semiotics*. University of Chicago. Retrieved from <http://csmt.uchicago.edu/glossary2004/semiotics.htm>
- Chen, L., Ellis, S. C., & Suresh, N. (2016). A supplier development adoption framework using expectancy theory. *International Journal of Operations & Production Management*, 36(5), 592–615. doi:10.1108/IJOPM-09-2013-0413
- Esteban-Guitart, M. (2015). LS Vygotsky and education by Moll, LC: (2014). New York, NY: Routledge, 173 pp. *Journal of Language, Identity, and Education*, 14(4), 295–297. doi:10.1080/15348458.2015.1070601
- Gorski, P. S. (2016). The Matter of Emergence: Material Artifacts and Social Structure. *Qualitative Sociology*, 39(2), 211–215. doi:10.1007/s11133-016-9328-5
- Gupta, A. B. (2016). M. Foucault on Knowledge and Some Parallels With Wittgenstein: A Comparison. *Global Journal for Research Analysis*, 5(3), 108–109. Retrieved from <http://worldwidejournals.in/ojs/index.php/gjra/article/viewFile/528/527>
- Holdford, A. D. A., & Lovelace-Elmore, B. (2001). Applying the principles of human motivation to pharmaceutical education. *Journal of Pharmacy Teaching*, 8(4), 1–18. doi:10.1300/J060v08n04_01
- Hung, C. C., Zhuang, W. L., & Lin, C. H. (2015). The Relationships among Buyers' Perceived Risk, Exhibitors' Brand Equity, Purchase Postponement and Switching Intention-From the Perspectives of Perceived Risk Theory and Expectancy Theory. *Information Management and Business Review*, 7(4), 74.
- Inwood, M. (2015). Wittgenstein and Heidegger. *The Philosophical Quarterly*, 00(00), pqv113. doi:10.1093/pq/pqv113
- Krathwohl, D. R., Bloom, B. S., & Masia, B. B. (1964). *Taxonomy of educational objectives: The classification of educational goals. Handbook 11: The affective domain*. New York: David McKay.
- Lăzăroiu, G. (2015). Work Motivation and Organizational Behavior. *Contemporary Readings in Law and Social Justice*, (2), 66–75.
- Maslow, A. H. (1943). A Theory of Human Motivation. *Psychological Review*, 50(4), 370–396. doi:10.1037/h0054346
- Maslow, A. H. (1954). *Motivation and personality*. New York: Harper and Row.
- Maslow, A. H. (1962). *Towards a psychology of being*. Princeton, NJ: D. Van Nostrand Company. doi:10.1037/10793-000
- Maslow, A. H. (1964/1970a). *Religions, values, and peak experiences*. New York: Penguin.
- Maslow, A. H. (1968). *Toward a Psychology of Being*. New York: D. Van Nostrand Company.
- Maslow, A. H. (1970b). *Motivation and personality*. New York: Harper & Row.

- Métraux, A. (2015). Lev Vygotsky as seen by someone who acted as a go-between between eastern and western Europe. *History of the Human Sciences*, 28(2), 154–172. doi:10.1177/0952695115575410
- Mishra, R., & Kotecha, K. (2012). Teaching Communication Skills Student Engagement Through Blogging. *International Journal of Humanities and Social Science*, 2(2), 4.
- Mishra, R., & Kotecha, K. (2016). Are we there Yet! Inclusion of Higher Order Thinking Skills (HOTs) in Assessment. *Journal of Engineering Education Transformations*. doi: 10.16920/jeet/2016/v0i0/85686
- Montealegre, R. (2016). Controversias Piaget-Vygotski en psicología del desarrollo/Piaget-Vygotsky controversies in developmental psychology. *Acta Colombiana de Psicología*, 19(1).
- Ong, K. K. A., Hart, C. E., & Chen, P. K. (2016). Promoting Higher-Order Thinking Through Teacher Questioning: A Case Study of a Singapore Science Classroom. *New Waves*, 19(1), 1.
- Parsons, J. E., & Goff, S. B. (1978). Achievement & motivation: Dual modalities. *Journal of Educational Psychology*, 13(1), 93–96. doi:10.1080/00461527809529199
- Purvis, R. L., Zagenczyk, T. J., & McCray, G. E. (2015). Whats in it for me? Using expectancy theory and climate to explain stakeholder participation, its direction and intensity. *International Journal of Project Management*, 33(1), 3–14. doi:10.1016/j.ijproman.2014.03.003
- Semiotic Society of America. (n.d.). *Semiotic Society of America*. Retrieved from <http://semioticsocietyofamerica.org/>
- Shotter, J. (2015). Undisciplining social science: Wittgenstein and the art of creating situated practices of social inquiry. *Journal for the Theory of Social Behaviour*. Retrieved from https://www.researchgate.net/profile/John_Shotter2/publication/271226696_Undisciplining_Social_Science_Wittgenstein_and_the_Art_of_Creating_Situated_Practices_of_Social_Inquiry/links/5523cf810cf2c815e0732846.pdf
- Shweiki, E., Martin, N. D., Beekley, A. C., Jenoff, J. S., Koenig, G. J., Kaulback, K. R., ... Cohen, M. J. (2015). Applying Expectancy Theory to residency training: proposing opportunities to understand resident motivation and enhance residency training. *Advances in Medical Education and Practice*, 6, 339.
- Streeter, T. (2012a). Semiotic Terminology. *Semiotics and Advertising*. Retrieved from http://www.uvm.edu/~tstreete/semiotics_and_ads/terminology.html
- Streeter, T. (2012b). *The Semiotics of Media*. Retrieved from http://www.uvm.edu/~tstreete/semiotics_and_ads/contents.html
- Vroom, V. H. (1964). *Work and motivation*. New York: Wiley.
- Vygotsky, L. S. (1933/1966). Play and its role in the mental development of the child. *Social Psychology*, 12(6), 62–76.
- Vygotsky, L. S. (1935). *Mental development of children during education*. Moscow: Uchpedzig.
- Vygotsky, L. S. (1981). The genesis of higher mental functions. In J. V. Wertsch (Ed.), *The concept of activity in Soviet psychology*. Armonk, NY: Sharpe.

Wang, R. (2015). LS Vygotsky and education. *British Journal of Educational Studies*, 63(1), 112–114. doi:10.1080/00071005.2015.1006957

Whittington, K. D. (2015). Does Motivation Predict Persistence and Academic Success? *Open Journal of Nursing*, 5(01), 10–16. doi:10.4236/ojn.2015.51002

Wittgenstein, L. (1961). *Tractatus logico-philosophicus* (D. F. Pars & B. F. McGuinness, Trans.). New York: Humanities Press.

Zohar, A., & Barzilai, S. (2015). Metacognition and teaching higher order thinking (HOT) in science education: Students' thinking, teachers' knowledge, and instructional practices. In R. Wegerif, L. Li, & J. Kaufman (Eds.), *Routledge international handbook of research on teaching thinking* (pp. 229–242). Oxon, UK: Routledge.

ADDITIONAL READING

Boggs, G. L. (2015). Listening to 21st century literacies: Prehistory of writing in an academic discipline. *Linguistics and Education*, 29, 15–31. doi:10.1016/j.linged.2014.10.001

Dondlinger, M. J., & McLeod, J. K. (2015). Solving real world problems with alternate reality gaming: Student experiences in the global village playground capstone course design. *Interdisciplinary Journal of Problem-Based Learning*, 9(2), 3. doi:10.7771/1541-5015.1488

Frank, A. (2015). Reaching the summit: explorations in meaningful learning through community engagement: International Higher Education Teaching and Learning (HETL)–Scholarship of Teaching and Engagement (SoTE) Conference, Utah Valley University, Orem UT, 20–22 January 2015. *The Town Planning Review*, 86(6), 725–732. doi:10.3828/tpr.2015.41

Garcia, N. A., & Longo, N. V. (2015). Community voices: Integrating local and international partnerships through storytelling. *Partnerships: A Journal of Service-Learning and Civic Engagement*, 6(2), 1–18.

Hervani, A. A., Helms, M. M., Rutti, R. M., LaBonte, J., & Sarkarat, S. (2015). Service learning projects in online courses: Delivery strategies. *Journal of Learning in Higher Education*, 11(1).

Newman, J. L., Dantzler, J., & Coleman, A. N. (2015). Science in action: How middle school students are changing their world through STEM service-learning projects. *Theory into Practice*, 54(1), 47–54. doi:10.1080/00405841.2015.977661

Savery, J. R. (2015). Overview of problem-based learning: Definitions and distinctions. *Essential readings in problem-based learning: Exploring and extending the legacy of Howard S. Barrows*, 5–15.

Smith, W., Butcher, E., Litvin, S. W., & Frash, R. (2015). Incorporating an Instructional Scaffolding Approach into the Classroom: Teaching for Authentic Learning in Hospitality and Tourism Education. *Journal of Teaching in Travel & Tourism*, 15(3), 264–277. doi:10.1080/15313220.2015.1059306

Swan, K. (2016). The Community of Inquiry Framework, Blended Learning, and the i2Flex Classroom Model. In *Revolutionizing K-12 Blended Learning through the i2Flex Classroom Model* (pp. 12–25). IGI Global.

Toven-Lindsey, B., Rhoads, R. A., & Lozano, J. B. (2015). Virtually unlimited classrooms: Pedagogical practices in massive open online courses. *The Internet and Higher Education*, 24, 1–12. doi:10.1016/j.iheduc.2014.07.001

KEY TERMS AND DEFINITIONS

Academic Community: This type of learner-realized real world community engagement falls within three realms of influence, specifically current course colleagues, prior course collegial support system, and current and prior instructors.

Community: This group of people are influential members of a cultural or social environment that may be geographic, emotional or impactful in some important manner of influence. Within this discussion, the community are a group of persons who impact a learner's sense of self as well as sense of subject matter understanding.

Personal Community: This type of learner-realized real world community engagement falls within three realms of influence, specifically friends, family, and social connections.

Professional Community: This type of learner-realized real world community engagement falls

within three realms of influence, specifically current work colleagues, prior work collegial support system, and professional organization colleagues.

Self-Efficacy: This term refers to a learner's ability to identify and distinguish one's own potential towards success. Within this discussion, the term refers to the learner's ability to identify and distinguish one's own potential towards learning success.

Self-Regulation: This is the learner's ability to understand one's own affective actions and reactions, recognize and control one's own environmental conditions, as well as observe one's own conduct and reactive performance.

Tiered Learning: Tiered learning is an instructional design approach. Within the concept of an instructional real world community, the academic, personal and professional communities are tiered engagements.

Learner Engagement in Blended Learning



Kristian J. Spring

Brigham Young University, USA

Charles R. Graham

Brigham Young University, USA

Tarah B. Ikahihifo

Brigham Young University, USA

BACKGROUND

Definitions of Blended Learning

Use of the term *blended learning* remains relatively new in higher education, K-12, and corporate settings. While this is the most commonly used label, the construct is sometimes described with the terms *mixed mode* and *hybrid learning* (Moskal, Dziuban, & Hartman 2013; Picciano, 2014b). Due to the flexible nature of blended learning, the debate continues over a precise definition of the term (Picciano, 2014b). While some consider this ambiguity a weakness that prohibits blended learning from use as a discriminating label (Oliver and Trigwell, 2005), others submit that a more narrow definition would impede “great potentials of the concept” (Alammary, Sherad, & Carbone, p. 443, 2015).

The most widely accepted basic position is that effective BL environments are a combination of F2F learning with technology-mediated instruction (Graham, 2006, 2013). Many individuals and institutions build upon this broad definition include caveats about seat time (Mayadas & Picciano, 2007), and the quality of the blend (Garrison & Kanuka, 2004) or quantity of instruction placed online (Allen & Seaman, 2007). Most current definitions of BL focus on the physical dimensions of the blend (e.g., online and face-to-face). However, future definitions may emphasize more

of the psychological/pedagogical dimensions of the blend (Graham, Henrie, & Gibbons, 2014).

Across contexts and institutions, varying ideas exist of what constitutes a BL environment (Porter, Graham, Spring & Welch, 2014). This distinction is most noticeable between postsecondary and K-12 sectors. Although BL at both levels is similar in many ways, it must be adapted to fit the K-12 setting (Staker & Horn, 2014). Horn & Staker’s (2015) three-part definition of BL focuses on the element of student control over their own learning experience, learning in a supervised brick-and-mortar location away from home, and the importance of an integrated learning experience. The integration aspect focuses on the coherence between the F2F and online components to deliver cohesive instruction for the learner about a given topic (Horn & Staker, 2015). An effective implementation of blended learning is well-coordinated with each component supporting the other.

Despite disagreement on an exact definition, many institutions are adapting BL to suit their specific needs. In each case, institutional context plays an important role in the construction of an operational definition and strategy (Moskal, Dziuban, & Hartman, 2013). The loose definition is “plastic enough to adapt to local needs and constraints of the several parties employing them, yet robust enough to maintain a common identity across sites” and thus allows the creation and implementation of customized institutional blends (Star & Griesemer, 1989, p. 393)

Reasons for Blending

A blended approach offers many advantages for both instructors and students. Stein & Graham (2014), Moloney et. al (2011), and Poon (2013) found that these benefits closely align with the Sloan-C Five Pillars of Quality:

1. Improved learning outcomes (including potential for learning communities and collaboration/active learning)
2. Cost reduction and effective use of resources
3. Access and flexibility
4. Student satisfaction
5. Faculty satisfaction

Reduced seat time, flexibility in time and space, and maintaining F2F interaction are some of the advantages of a blended approach (Moskal, Dziuban, & Hartman, 2013). Many educators choose to adopt a blended approach to avoid sacrificing benefits of one method for benefits of the other (e.g. convenience of an asynchronous distributed environment without eliminating the benefit of human contact in the F2F environment). While BL can provide the “best of both worlds” (Bonk & Graham, 2012; Moskal, Dziuban, & Hartman, 2013), if not designed with thoughtful consideration of the combined methods, content, and intended learners’ needs, BL environments can also mix the least effective elements of F2F and online instruction. Thus clearly articulated models are needed to help guide practice and research (Graham, Henrie & Gibbons, 2014; Halverson, Graham, Spring, Drysdale & Henrie, 2014).

LEARNER ENGAGEMENT IN BLENDED LEARNING

Learner engagement has been heavily researched (Azevedo, 2015) and termed the “holy grail” of learning (Sinatra, Heddy & Lombardy, 2015). Engagement is strongly connected to a variety of benefits (Carini, Kuh & Klein, 2006; Mountford-

Zimdars, Sabri, Moore, Sanders, Jones & Higham, 2015). Some positive student outcomes linked to learner engagement include student performance (Casuso-Holgado et al., 2013; Kuh et al., 2008), persistence and resilience (Kuh et al., 2008), and psychological growth (Harper & Quaye, 2009).

Defining Learner Engagement

The educational field has yet to agree upon a standard definition of engagement (Sinatra, Heddy & Lombardy, 2015). Possible competing and/or overlapping definitions span integrating learning strategies and motivations (Richardson & Newby, 2006), a combination of motivation and cognition (Järvelä, Veermanis & Leinonen, 2008), and “emotionally positive” and “cognitively focused” (Skinner and Pitzer, 2012). We suggest that engagement may be conceptualized as the “quality and quantity” of energy exerted (Henrie, Bodily, Manwaring, & Graham, 2015).

The student experience is different in an on-line setting than in a traditional one, and blended engagement presents unique challenge and opportunity for engagement. While the research has not clearly identified the underlying causes for improved outcomes in blended contexts (Means et al., 2013) Blended learning has been identified a means for increasing student engagement (Graham & Robison, 2007; Northey, Bucic, Chylinski, Govind, 2015). More research is needed on the types of blends that produce “deep and meaningful learning” (Bernard, Borokhovski, Schmid, Tamim & Abrami, 2014, p. 116).

Models of Learner Engagement

There are several models that researchers have used to conceptualize engagement. One prominent model characterizes engagement as part of three domains: behavioral, emotional, and cognitive (Fredricks, Blumenfeld, & Paris, 2004). Behavioral engagement is connected to a student’s actions and participation, emotional engagement is related to a student’s feelings about the learning

experience, and cognitive engagement focuses on investment and exerting effort to understand material. This model looks at engagement as a meta-construct which combines multiple aspects of the student experience (Fredricks, Blumenfeld, & Paris, 2004).

Skinner & Pitzer (2012) extended behavioral, emotional, and cognitive engagement to contextual levels with specific outcomes at each. The most expansive is engagement with prosocial *institutions*, like churches and clubs, which produces generally positive development. Second engagement is more limited within a *school* context, including classes and extracurricular activities, which encourage students to remain in school and graduate. Next comes engagement with a *class*, including the instructor, material, and other students, which bolsters student achievement. Finally students can engage in specific *learning activities* which develop academic assets like learning, coping and resilience (Skinner & Pitzer, 2012). This model is useful for identifying the levels at which engagement is measured and conceptualized (Henrie, Halverson & Graham, 2015).

Researchers continue to develop and test new models for measuring engagement in order to track engagement in technology-mediated and blended contexts. Lam et al. (2014) developed a scale to measure cognitive, behavioral, and affective engagement internationally and tested it with over

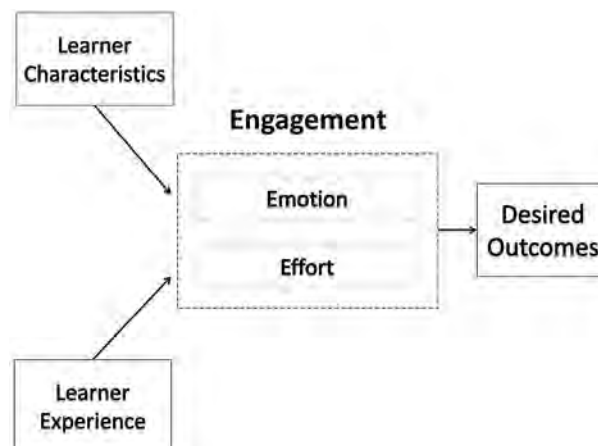
3,000 students in 12 countries; they found that the scale is valid and can be applied further to student engagement elsewhere in the world. Henrie, Halverson & Graham (2015) identified and analyzed existing methods of measuring engagement, in order to suggest potential improvements in engagement measurement. Henrie, Bodily, Manwaring & Graham (2015) used longitudinal data to study engagement in a blended course and discovered that clear instruction and relevant activities had a greater effect on student engagement than the instructional medium (Figure 1).

Measuring Learning Engagement

Measuring Engagement is a complicated but important task. Gaging learner engagement is difficult because it is challenging to define the construct, to select the granularity at which to measure the engagement, and to avoid disrupting learning while collecting data in real-time (Sinatra, Heddy & Lombardy, 2015).

Granularity, the “size” of engagement, must be addressed. Researchers should carefully examine the theory, definition and grain size of possible measures while planning a study of engagement. Engagement ranges from the micro level (i.e. individual students for short periods of time) to macro level (i.e. many students for sustained periods of time) and measurement methods should correspond

Figure 1. Model of learner engagement



to the engagement level to be measured (Sinatra, Heddy & Lombardy, 2015). Methods for collecting data to measure engagement span the classic self-report (Greene, 2015) to the highly technical eye-tracking (Miller, 2015). Each measurement method has its place, dependent upon fit for theoretical underpinnings and research goals (Fredricks & McColskey, 2012), though some methods are only possible in a blended or online context.

Self-reporting and similar types of data collection might interrupt engagement. (Baker et. al, 2012). Signaling students to respond right away after an event, several times during a learning experience can provide a more complete understanding of their engagement (Järvelä, Veermans, & Leinonen, 2008). However, the more interruptions a student experiences, the more invasive the disruptions may be. Computer-generated information in an online or blended context can be harnessed to allow researchers to collect data in real time without such disruptions (D'Mello & Graesser, 2012; Graesser, D'Mello, & Strain, 2014; Henrie, Halverson & Graham, 2015). It may be possible to infer student engagement from a student's interaction with an online interface with no to minimal disruption to the student's learning experience (Gobert, Baker & Wixton, 2015).

The possibilities for real-time, actionable measures are greatly enhanced by technology. Face and gesture recognition programs allow researchers to gauge someone's emotions based on their facial and upper body expressions (Baltrušaitis et al., 2011). This type of data has been applied to educational contexts (D'Mello & Kory, 2015). Data automatically collected by a learning management system (LMS) can also help determine a student's engagement level. Information about the number of times a student participates in a discussion, sends mail, or completes an assessment can help an instructor understand his students engagement levels (Macfayden & Dawson, 2010; Vivekananthamoorthy & Naganathan, 2015). Harnessing an LMS to monitor student engagement can also help a professor provide timely, meaningful feedback to struggling students.

Learner Engagement in Online and Blended Learning

In a traditional learning environment, students interact face-to-face with instructors in a classroom setting. A typical approach in this context, especially at the postsecondary level, is instruction provided through lecture (Smith & Cardaciotto, 2011). This approach is particularly efficient when dealing with large class sizes, however, a major concern is that of student engagement (Blood & Neel, 2008). It is easier for students to become disengaged in a larger class as there is less personalized instruction and fewer opportunities to actively participate. Generally, students have a more passive role in a traditional setting with instruction being more "educator-focused" (Northey, Bucic, Chylinski & Govind, 2015, p. 171). When opportunities to participation do arise, students indicate their willingness to do so by raising their hand (Barr, 2013). Although some students feel comfortable participating in this manner, others may find this response method intimidating, lessening their desire to contribute to the classroom discussion (Barr, 2013). Problem-based learning and student response systems have been used as methods to increase student engagement in a traditional classroom setting. Studies show these approaches can increase engagement and participation in the traditional context, allowing students to have a more active role (Barr, 2013; Blood & Neel, 2008).

Online learning environments allow students the flexibility and autonomy to control their own learning experience (McBrien, Jones, & Cheng, 2009; Robinson & Hullinger, 2008). As the internet has increased in accessibility so has the number of students opting for online courses (Allen & Seaman, 2008). Robinson & Hullinger (2008) found that higher levels of engagement were reported by online students compared to on-campus students for each of their benchmarks used to measure engagement. The four benchmarks were level of academic challenge, student-faculty interaction, active and collaborative learning,

and enriching educational experience. While the F2F element is not present in online instruction, students are expected to work together, fostering collaboration among participants (Chen, Lambert, & Guidry, 2010). The use of synchronous conference technologies provide opportunities for social interaction in a typically asynchronous environment (McBrien, Jones, & Cheng, 2009). According to Chen, Lambert, & Guidry (2010), students who actively participate in Web-based learning are more likely to make use of practices such as reflective learning, higher order thinking, and integrative learning.

Blended environments allow instructors to harness the capacities of both traditional and online engagement strategies, as well as unique combinations of the two. Complementing face-to-face activities with out-of-class, asynchronous learning opportunities allows educators to create a community-focused space for personalized learning experiences for students (Northey, Bucic, Chylinski, & Govind, 2015). Both environments support each other and offer opportunities for learners to be co-creators of their knowledge. Northey, Bucic, Chylinski & Govind (2015) found that adding an asynchronous component to a course, in the form of social media, improved student engagement and final grades. Their study also found that the positive interactions students experienced through the online portion of a blended course had a positive spillover effect in the F2F setting.

Even within the context of blended learning environments, the method in which the F2F component is delivered can affect the level of engagement. Delialioğlu (2012) found that students reported higher levels of engagement in a problem-based learning environment as opposed to the traditional lecture-based method of delivery. Regardless of the difference between these methods, students were highly satisfied with the blended learning configuration of the course (Delialioğlu, 2012). Although technology plays an integral role in the lives of students, learning how to effectively

leverage this tool in meaningful and engaging ways academically is a task for instructors and students alike (Dahlstrom & Bichsel, 2014).

Interaction is one way to understand how engagement and technology are connected. Moore (1989) divided interaction in distance education into three domains: learner-content, learner-instructor, and learner-learner. In a meta-analysis of experimental literature on interaction in distance education Bernard et al. (2009) found all three types of interaction to positively influence student learning, possibly due to increased cognitive engagement. Interaction should be mindfully included in courses and activities designed to promote engagement (Meyer, 2014). These types of interaction in online and blended learning can also be understood as human-technology interaction (i.e. learner-content interaction) and human-human interaction (i.e. learner-instructor and learner-learner interaction).

Strong learner-content interaction can improve learner outcomes. Zimmerman (2012) found that learners who spent more time interacting with content earned higher grades in a course. Creating meaningful interactions with content, rather than simply providing it, also improves student outcomes (Dunlap, Sobel & Sands, 2007). Hillman, Willis, & Gunawardena (1994) extended interaction to a fourth domain, learner-interface interaction and argued that specific strategies should be applied to this type of learning experience. We would conceptualize both learner-content and learner-interface interaction as human-technology interaction in an online or blended context. There are obvious differences between a static book, or even an e-book or eLearning course and a the dynamic possibilities provided by modern technology in terms of interaction as well as data collection. Automated engagement detection programs, which track a student's engagement individually, in real time, have success rates that are comparable to those of human raters (Whitehill, Serpell, Lin, Foster & Movellan, 2014). Interventions are even being tested to respond to

students apparent emotions with appropriate facial expression from a computer generated model (Jayareka & Rajamohamed, 2015).

Students may also engage differently when they are experiencing learner-human interaction as opposed to learner-content or learning-technology interaction. A plethora of technologies are available to foster human-human interaction in online and blended contexts both synchronously and asynchronously. Blogs, wikis, video conferencing, live presentation tools, emails, chats, and many other emerging technologies are versatile ways to include human-human interaction at a distance (Beldarrain, 2006). While there was variation between types of interactions, Borokhovski, Tamim, Bernard & Abrami et al. (2012) found that contextual interactions which provided an environment for student interaction but were not purposefully created to promote collaboration were less effective than designed interactions which were more beneficial to student achievement. Learner-instructor interaction is also an important dimension of human-human interaction. Shackelford & Maxwell (2012) found instructor modeling, encouragement and other facets of learner-instructor interaction cultivated a greater sense of community for students.

FUTURE RESEARCH DIRECTIONS IN ENGAGEMENT IN ONLINE AND BLENDED CONTEXTS

In 2006 Bonk, Kim, and Zeng (p. 560) made some predictions about the future of blended learning:

1. Increased use of mobile devices in BL
2. Greater use of visualization tools and hands-on learning
3. Increased learner input in designing their own learning programs
4. Increased connectedness, community, and collaboration
5. Increased authenticity and on-demand learning
6. Stronger ties between work and learning
7. More flexible calendaring system
8. More inclusion of BL course designations in program descriptions
9. Increased movement in instructor roles toward mentoring, coaching, and counseling

Some of these predictions are manifest in the succeeding literature as developing trends. Models such as the self (à la carte) blend, buffet blend, or field blend give students more input in designing their own learning. Additionally, formats like the flex model, rotation model, the emporium and fully online blends are beginning to shift the role of the instructor. The literature has mentioned increased connectedness, community, and collaboration (see Poon, 2013). However, for some of the other predictions it may be too early to know whether they truly become trends.

While the future of BL environments is hard to predict, the use of BL in K-12, higher education, and corporate contexts will likely continue to increase and may become the norm as predicted. Future learning environments may be predominantly BL, and the adjective *blended* may no longer be needed. Focus in higher education may be on whether blending is happening or not, and universities may be more concerned with the quality of the blend and seek to understand how faculty can be trained and supported to teach in BL environments. Some related trends might develop:

1. Increasing use of learning analytics to help improve learning (online and blended models allow for easier data collection; see Picciano, 2014b). A better understanding of student engagement and more advanced mechanisms for measuring, responding to, and supporting student engagement in real-time.
2. Increasing personalization of learning (see Horn et al., 2011), especially in response to individual student engagement data.
3. Movement of instructor roles toward mentoring, coaching, and counseling as learner-content or human-technology interaction

can be made more engaging for information delivery and human-human interaction can focus on the personal aspects of learner.

Corporate and military contexts are most likely to continue to push the technological envelope, exploring more expensive technologies, although simulations may be used more commonly in K-12 and higher education classrooms.

CONCLUSION

Because access to and understanding of technology are spreading, we see unprecedented availability of instructional resources and efficient communication. However online and F2F learning environments have unique benefits. Because of varied strengths, the present move toward blending online and F2F instruction has valuable potential that warrants further consideration and research. Blended learning has much potential for improving both student engagement and the ability for researchers and instructors to measure and develop that engagement. Researchers must continue to work toward a consistent definition and validated models of learner engagement. We feel that in this endeavor researchers and practitioners will be able to enhance blending strategies to accomplish the objectives for which BL efforts are initiated.

REFERENCES

- Alammary, A., Sheard, J., & Carbone, A. (2014). Blended learning in higher education: Three different design approaches. *Australasian Journal of Educational Technology*, 30(4), 440–454. doi:10.14742/ajet.693
- Allen, I. E., & Seaman, J. (2007). *Online nation: Five years of growth in online learning*. Sloan Consortium report.
- Allen, I. E., & Seaman, J. (2008). Staying the course. *Online Education in the United States*, 2008. *Sloam Consortium*, 22. doi:10.1242/jeb.049304
- Azevedo, R. (2015). Defining and Measuring Engagement and Learning in Science: Conceptual, Theoretical, Methodological, and Analytical Issues. *Educational Psychologist*, 50(1), 84–94. doi:10.1080/00461520.2015.1004069
- Baker, R. S. J., Gowda, S. M., Wixon, M., Kalka, J., Wagner, A. Z., & Aleven, V., ... Rossi, L. (2012). Towards sensor-free affect detection in cognitive tutor algebra. In K. Yacef, O. Zaïane, H. Hershkovitz, M. Yudelson, & J. Stamper (Eds.), *Proceedings of the 5th International Conference on Educational Data Mining* (pp. 126–133).
- Baltrušaitis, T., McDuff, D., Banda, N., Mahmoud, M., El Kaliouby, R., Robinson, P., & Picard, R. (2011, March). Real-time inference of mental states from facial expressions and upper body gestures. In *Automatic Face & Gesture Recognition and Workshops (FG 2011)*, 2011 *IEEE International Conference on* (pp. 909–914). IEEE. doi:10.1109/FG.2011.5771372
- Barr, M. L. (2013). Encouraging College Student Active Engagement in Learning: The Influence of Response Methods. *Innovative Higher Education*, 1–13. doi:10.1007/s10755-013-9276-x
- Beldarrain, Y. (2006). Distance Education Trends: Integrating new technologies to foster student interaction and collaboration. *Distance Education*, 27(2), 139–153. doi:10.1080/01587910600789498
- Bernard, R. M., Abrami, P. C., Borokhovski, E., Wade, C. A., Tamim, R. M., Surkes, M. A., & Bethel, E. C. (2009). A meta-analysis of three types of interaction treatments in distance education. *Review of Educational Research*, 79(3), 1243–1289. doi:10.3102/0034654309333844

- Bernard, R. M., Borokhovski, E., Schmid, R. F., Tamim, R. M., & Abrami, P. C. (2014). A meta-analysis of blended learning and technology use in higher education: From the general to the applied. *Journal of Computing in Higher Education*, 26(1), 87–122. doi:10.1007/s12528-013-9077-3
- Blood, E., & Neel, R. (2008). Using Student Response Systems in Lecture-Based Instruction: Does It Change Student Engagement and Learning? *Journal of Technology and Teacher Education*, 16(3), 375–383. Retrieved from <http://www.eric.ed.gov/ERICWebPortal/detail?accno=EJ797352>
- Bonk, C. J., & Graham, C. R. (2012). *The handbook of blended learning: Global perspectives, local designs*. San Francisco, CA: Pfeiffer.
- Bonk, C. J., Kim, K.-J., & Zeng, T. (2006). Future directions of blended learning in higher education and workplace learning settings. In C. J. Bonk & C. R. Graham (Eds.), *The handbook of blended learning: Global perspectives, local designs* (pp. 550–567). San Francisco, CA: Pfeiffer Publishing.
- Borokhovski, E., Tamim, R., Bernard, R. M., Abrami, P. C., & Sokolovskaya, A. (2012). Are contextual and designed student–student interaction treatments equally effective in distance education? *Distance Education*, 33(3), 311–329. doi:10.1080/01587919.2012.723162
- Carini, R. M., Kuh, G. D., & Klein, S. P. (2006). Student Engagement and Student Learning: Testing the Linkages. *Research in Higher Education*, 47(1), 1–32. doi:10.1007/s11162-005-8150-9
- Casuso-Holgado, M. J., Cuesta-Vargas, A. I., Moreno-Morales, N., Labajos-Manzanares, M. T., Barón-López, F. J., & Vega-Cuesta, M. (2013). The association between academic engagement and achievement in health sciences students. *BMC Medical Education*, 13(33), 1–7. doi:10.1186/1472-6920-13-33 PMID:23446005
- Chen, P.-S. D., Lambert, A. D., & Guidry, K. R. (2010). Engaging online learners: The impact of Web-based learning technology on college student engagement. *Computers & Education*, 54(4), 1222–1232. doi:10.1016/j.compedu.2009.11.008
- D ’Mello, S. K., & Kory, J. (2015). A Review and Meta-Analysis of Multimodal Affect Detection Systems. *ACM Computing Surveys*, 47(43). doi:10.1145/2682899
- Dahlstrom, E., & Bichsel, J. (2014). *Study of Undergraduate Students and Information Technology, 2014*, 50. Retrieved from <http://www.educause.edu/library/resources/ecar-study-undergraduate-students-and-information-technology-2013>
- Delialioğlu, Ö. (2012). Student engagement in blended learning environments with lecture-based and problem-based instructional approaches. *Journal of Educational Technology & Society*, 15(3), 310–322. doi:10.2307/jeductechsoci.15.3.310
- DMello, S., & Graesser, A. (2012). AutoTutor and affective autotutor. *ACM Transactions on Interactive Intelligent Systems*, 2(4), 1–39. doi:10.1145/2395123.2395128
- Dunlap, J., Sobel, D., & Sands, D. (2007). Designing for deep and meaningful student-to-content interactions. *TechTrends*, 51(4), 20–31. doi:10.1007/s11528-007-0052-6
- Finn, A. (2002). Trends in e-learning. *Learning Circuits*, 3(11). Retrieved from <http://www.learningcircuits.org/2002/nov2002/finn.htm>
- Fredricks, J. A., Blumenfeld, P. C., & Paris, H. (2004). School Engagement: Potential of the Concept, State of the Evidence. *Review of Educational Research*, 74(1), 59–109. doi:10.3102/00346543074001059

- Fredricks, J. A., & McColskey, W. (2012). The measurement of student engagement: A comparative analysis of various methods and student self-report instruments. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of Research on Student Engagement* (pp. 763–782). Boston, MA: Springer US; doi:10.1007/978-1-4614-2018-7_37
- Garrison, D. R., & Kanuka, H. (2004). Blended learning: Uncovering its transformative potential in higher education. *The Internet and Higher Education*, 7(2), 95–105. doi:10.1016/j.iheeduc.2004.02.001
- Gobert, J. D., Baker, R. S., & Wixon, M. B. (2015). Operationalizing and Detecting Disengagement Within Online Science Microworlds. *Educational Psychologist*, 50(1), 43–57. doi:10.1080/00461520.2014.999919
- Graham, C. R. (2013). Emerging practice and research in blended learning. In M. G. Moore (Ed.), *Handbook of distance education* (3rd ed.; pp. 333–350). New York, NY: Routledge. doi:10.4324/9780203803738.ch21
- Graham, C. R., Henrie, C. R., & Gibbons, A. S. (2014). Developing models and theory for blended learning research. In A. G. Picciano, C. D. Dziuban, & C. R. Graham (Eds.), *Blended learning: Research perspectives* (2nd ed.; pp. 13–33). New York, NY: Routledge.
- Graham, C. R., & Robison, R. (2007). Realizing the transformational potential of blended learning: Comparing cases of transforming blends and enhancing blends in higher education. In A. G. Picciano & C. D. Dziuban (Eds.), *Blended Learning: Research Perspectives* (pp. 83–110). Needham, MA: The Sloan Consortium.
- Graham, C. R. (2006). Blended learning systems: Definition, current trends, and future directions. In C. J. Bonk & C. R. Graham (Eds.), *Handbook of blended learning: Global perspectives, local designs* (pp. 3–21). San Francisco, CA: Pfeiffer Publishing.
- Greene, B. A. (2015). Measuring Cognitive Engagement With Self-Report Scales: Reflections From Over 20 Years of Research. *Educational Psychologist*, 50(1), 14–30. doi:10.1080/00461520.2014.989230
- Gunuc, S., & Kuzu, A. (2015). Confirmation of Campus-Class-Technology Model in student engagement: A path analysis. *Computers in Human Behavior*, 48, 114–125. doi:10.1016/j.chb.2015.01.041
- Halverson, L. R., Graham, C. R., Spring, K. J., Drysdale, J. S., & Henrie, C. R. (2014). A thematic analysis of the most highly cited scholarship in the first decade of blended learning research. *The Internet and Higher Education*, 20, 20–34. doi:10.1016/j.iheeduc.2013.09.004
- Harper, S. R., & Quaye, S. J. (2009). Beyond sameness, with engagement and outcomes for all. In *Student engagement in higher education: Theoretical perspectives and practical approaches for diverse populations* (pp. 1–15). Retrieved from <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Beyond+Sameness,+with+Engagement+and+Outcomes+for+All#0>
- Henrie, C. R., Bodily, R., Manwaring, K. C., & Graham, C. R. (2015). Exploring intensive longitudinal measures of student engagement in blended learning. *International Review of Research in Open and Distributed Learning*, 16(3), 131–155. doi:10.19173/irrodl.v16i3.2015
- Henrie, C. R., Halverson, L. R., & Graham, C. R. (2015). Measuring Student Engagement in Technology-Mediated Learning: A Review. *Computers & Education*, 90, 36–53. doi:10.1016/j.compedu.2015.09.005
- Hillman, D. C., Willis, D. J., & Gunawardena, C. N. (1994). Learner-interface interaction in distance education: An extension of contemporary models and strategies for practitioners. *American Journal of Distance Education*, 8(2), 30–42. doi:10.1080/08923649409526853

- Horn, M. B., & Staker, H. (2015). *Blended: Using disruptive innovation to improve schools*. San Francisco, CA: Jossey-Bass.
- Horn, M. B., Staker, H., Hernandez, A., Hassel, B., & Ableidinger, J. (2011). *The rise of K-12 blended learning: Profiles of emerging models*. Innosight Institute. Retrieved from <http://www.innosightinstitute.org/innosight/wp-content/uploads/2011/01/The-Rise-of-K-12-Blended-Learning.pdf>
- Hoxie, A.-M., Stillman, J., & Chesal, K. (2014). Blended learning in New York City: The iLearn-NYC Program. In A. G. Picciano, C. D. Dziuban, & C. R. Graham (Eds.), *Blended learning: Research perspectives* (Vol. 2, pp. 304–324). New York, NY: Routledge.
- Järvelä, S., Veermans, M., & Leinonen, P. (2008). Investigating student engagement in computer-supported inquiry: A process-oriented analysis. *Social Psychology of Education*, 11(3), 299–322. doi:10.1007/s11218-007-9047-6
- Jayareka, K. S., & Rajamohamed, S. (2015). 3D Human Computer-Intelligent Interaction by Facial Expression analysis using Facial Action Coding System-FACS. *International Journal of Scientific & Engineering Research*, 6(4), 56–63.
- Kuh, G. D., Cruce, T. M., Shoup, R., Kinzie, J., Gonyea, R. M., & Gonyea, M. (2008). Unmasking the Effects of Student Engagement on First-Year College Grades and Persistence. *The Journal of Higher Education*, 79(5), 540–563. doi:10.1353/jhe.0.0019
- Lam, S., Jimerson, S., Wong, B. P. H., Kikas, E., Shin, H., Veiga, F. H., & Zollneritsch, J. et al. (2014). Understanding and measuring student engagement in school: The results of an international study from 12 countries. *School Psychology Quarterly : The Official Journal of the Division of School Psychology, American Psychological Association*, 29(2), 213–232. doi:10.1037/spq0000057 PMID:24933218
- Macfadyen, L. P., & Dawson, S. (2010). Mining LMS data to develop an early warning system for educators: A proof of concept. *Computers & Education*, 54(2), 588–599. doi:10.1016/j.compedu.2009.09.008
- Mayadas, A. F., & Picciano, A. G. (2007). Blended learning and localness: The means and the end. *Journal of Asynchronous Learning Networks*, 11(1), 3–7.
- McBrien, J. L., & Jones, P. (2009). Virtual Spaces : Employing a Synchronous Online Classroom to Facilitate Student Engagement in Online Learning. *International Review of Research in Open and Distance Learning*, 10(3), 1–17. doi:10.19173/irrodl.v10i3.605
- Meyer, K. A. (2014). Student engagement in online learning: What works and why. *ASHE Higher Education Report*, 40(6), 1–114. doi:10.1002/aehe.20018
- Miller, B. W. (2015). Using reading times and eye-movements to measure cognitive engagement. *Educational Psychologist*, 50(March), 31–42. doi:10.1080/00461520.2015.1004068
- Moloney, J. F., Hickey, C. P., Bergin, A. L., Boccia, J., Polley, K., & Riley, J. E. (2011). Characteristics of successful local blended programs in the context of the Sloan-C pillars. *Journal of Asynchronous Learning Networks*, 14(1), 71–89. Retrieved from <http://files.eric.ed.gov/fulltext/EJ909843.pdf>
- Moore, M. G. (1989). Editorial: Three types of interaction. *American Journal of Distance Education*, 3(2), 1–7. doi:10.1080/08923648909526659
- Moskal, P., Dziuban, C., & Hartman, J. (2013). Blended learning: A dangerous idea? *The Internet and Higher Education*, 18, 15–23. doi:10.1016/j.iheduc.2012.12.001
- Mountford-Zimdars, A., Sabri, D., Moore, J., Sanders, J., Jones, S., & Higham, L. (2015). *Causes of differences in student outcomes*. Academic Press.

- Northey, G., Bucic, T., Chylinski, M., & Govind, R. (2015). Increasing Student Engagement Using Asynchronous Learning. *Journal of Marketing Education*, 37(3), 171–180. doi:10.1177/0273475315589814
- Oliver, M., & Trigwell, K. (2005). Can blended learning be redeemed? *E-learning*, 2(1), 17–26. doi:10.2304/elea.2005.2.1.2
- Picciano, A. (2014a). Big Data and Learning Analytics in Blended Learning Environments: Benefits and Concerns. *International Journal of Interactive Multimedia and Artificial Intelligence*, 2(7), 35. doi:10.9781/ijimai.2014.275
- Picciano, A. (2014b). Introduction. In A. G. Picciano, C. D. Dziuban, & C. R. Graham (Eds.), *Blended learning: Research perspectives* (Vol. 2, pp. 1–9). New York, NY: Routledge.
- Poon, J. (2013). Blended learning: An institutional approach for enhancing students' learning experiences. *Journal of Online Learning and Teaching*, 9(2), 271–288. Retrieved from http://jolt.merlot.org/vol9no2/poon_0613.pdf
- Porter, W. W., Graham, C. R., Spring, K. A., & Welch, K. R. (2014). Blended learning in higher education: Institutional adoption and implementation. *Computers & Education*, 75, 185–195. doi:10.1016/j.compedu.2014.02.011
- Richardson, J. C., & Newby, T. (2006). The role of students cognitive engagement in online learning. *American Journal of Distance Education*, 20(1), 23–37. doi:10.1207/s15389286ajde2001_3
- Robinson, C. C., & Hullinger, H. (2008). New benchmarks in higher education : Student engagement in online learning. *Journal of Education for Business*, 84(2), 101–109. doi:10.3200/JOEB.84.2.101-109
- Shackelford, J. L., & Maxwell, M. (2012). Contribution of Learner – Instructor Interaction to Sense of Community in Graduate Online Education. *MERLOT Journal of Online Learning and Teaching*, 8(4), 248–260.
- Sinatra, G. M., Heddy, B. C., & Lombardi, D. (2015). The challenges of defining and measuring student engagement. *Educational Psychologist*, 50(1), 1–13. doi:10.1080/00461520.2014.1002924
- Skinner, E. A., & Pitzer, J. R. (2012). Developmental dynamics of student engagement, coping, and everyday resilience. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of Research on Student Engagement* (pp. 21–44). Boston, MA: Springer US; doi:10.1007/978-1-4614-2018-7_2
- Smith, C. V., & Cardaciotto, L. (2011). Is active learning like broccoli? Student perceptions of active learning in large lecture classes. *Journal of the Scholarship of Teaching & Learning*, 11(1), 53–61. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,url,cookie,uid&db=ehh&AN=58601456&site=ehost-live>
- Staker, H., & Horn, M. B. (2014). Blended learning in the K-12 education sector. In A. G. Picciano, C. D. Dziuban, & C. R. Graham (Eds.), *Blended learning: Research perspectives* (Vol. 2, pp. 287–303). New York, NY: Routledge.
- Star, S. L., & Griesemer, J. R. (1989). Institutional ecology, translations and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 190739. *Social Studies of Science*, 19(3), 387–420. doi:10.1177/030631289019003001
- Stein, J., & Graham, C. R. (2014). *Essentials for blended learning: A standards-based guide*. New York, NY: Routledge.
- Vivekananthamoorthy, N., & Naganathan, E. R. (2015). Engagement And Interaction In E-Learning Using Web. *International Journal of Applied Engineering Research*, 10(5), 12253–12276.
- Whitehill, J., Serpell, Z., Lin, Y. C., Foster, A., & Movellan, J. R. (2014). The faces of engagement: Automatic recognition of student engagement from facial expressions. *IEEE Transactions on Affective Computing*, 5(1), 86–98. doi:10.1109/TAFFC.2014.2316163

Young, J. R. (2002, March 22). “Hybrid” teaching seeks to end the divide between traditional and online instruction. *The Chronicle of Higher Education*, A33.

Zimmerman, T. D. (2012). Exploring learner to content interaction as a success factor in online courses. *International Review of Research in Open and Distance Learning*, 13(4), 152–165. doi:10.19173/irrodl.v13i4.1302

ADDITIONAL READING

Garrison, D. R., & Vaughan, N. D. (2008). *Blended learning in higher education: Framework, principles, and guidelines*. San Francisco, CA: Jossey-Bass.

Graham, C. R. (2006). Blended learning systems: Definition, current trends, and future directions. In C. J. Bonk & C. R. Graham (Eds.), *Handbook of blended learning: Global perspectives, local designs* (pp. 3–21). San Francisco, CA: Pfeiffer Publishing.

Graham, C. R. (2013). Emerging practice and research in blended learning. In M. G. Moore (Ed.), *Handbook of distance education* (3rd ed., pp. 333–350). New York, NY: Routledge. doi:10.4324/9780203803738.ch21

Henrie, C. R., Halverson, L. R., & Graham, C. R. (2015). Measuring Student Engagement in Technology-Mediated Learning: A Review. *Computers & Education*, 90, 36–53. doi:10.1016/j.compedu.2015.09.005

Means, B., Toyama, Y., Murphy, R., & Baki, M. (2013). The Effectiveness of Online and Blended Learning: A Meta-Analysis of the Empirical Literature.pdf. *Teachers College Record*, 115(3). Retrieved from <http://www.tcrecord.org/library/content.asp?contentid=16882>

Means, B., Toyama, Y., Murphy, R., Bakia, M., & Jones, K. (2009). Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning studies (p. 94). Retrieved from <http://www2.ed.gov/rschstat/eval/tech/evidence-based-practices/finalreport.pdf>

Picciano, A. G., & Dziuban, C. (Eds.). (2007). *Blended Learning Research Perspectives*. The Sloan Consortium.

KEY TERMS AND DEFINITIONS

Blended Learning Environment: A learning environment that combines face-to-face and computer-mediated instruction.

Cognitive Engagement: The effort and cognitive energy that students devote to mastering the educational objects.

Distributed Learning Environment: A learning environment in which participants are not co-located, but use computer-based technologies to access instruction and communicate with others.

Emotional Engagement: Students’ emotional responses (e.g. attitudes, relationships) to other people and content in the learning environment.

Flipped Classroom: An instructional approach where initial instruction in course curricula occurs prior to the face-to-face meeting and application activities occur in class so that the instructor is present to guide and provide feedback.

Hybrid Course: Another term for blended course—typically a course that replaces some F2F instructional time with computer-mediated activities.

Learning Analytics: The collection and analysis of data regarding learning performance and activity in order to understand the learning process and help individual students succeed.

Measuring Text Readability Using Reading Level

C

James C. Brewer

Texas Tech University, USA

INTRODUCTION

Reading Level is a calculation by formula which assigns a number to a text in order to represent the grade level at which a reader should be able to comprehend written English-language material. A variety of such formulas are in use, with the Flesch-Kincaid Grade Level as the best known and most widely used in the United States. The computation is supported by a variety of Web tools and by some of the most popular word processing tools. The formula is based on work done by Rudolf Franz Flesch (1911-1986) to develop the Flesch Reading Ease formula—as distinct from the Flesch-Kincaid Grade Level—which assigns a number between zero and one hundred twenty (120) to a text when expressing its reading difficulty. Zero means practically unreadable and numbers up to one hundred twenty (120) mean easier to read. (Theoretically, numbers less than zero are possible for extreme cases.) Examples are *Reader's Digest* at 65 and the *Harvard Law Review* in the low 30s. Note that for a journal, individual articles will be at different reading levels, and for this reason a journal will have a range or average value to describe its reading level. Following the work of Flesch, J. Peter Kincaid (1942-) used the Flesch Reading Ease formula as a basis for the computation today called the Flesch-Kincaid Grade Level. When comparing several different texts, the Flesch Reading Ease formula can be used as a measure of which of the texts is more difficult. The Flesch-Kincaid Grade Level allows a grade number to be assigned to reading material as suitable for that grade level.

Publishers of children's literature and school texts are often users of technologies to assign reading level to their published materials, but a much wider range of users can be found: producers of technical manuals which need to reach a range of educational backgrounds, those who create medical consent forms, and state and federal government agencies that produce documents covering insurance benefits or driver education manuals and tests, for example. Within the classroom or online education setting, reading level can also provide benefits to alert presenters and instructors about the complexity of the material that is introduced.

BACKGROUND

In mid-nineteenth century America schoolrooms were generally not divided into grade levels. Over time grade levels were added and methods were developed to measure the grade level of texts. As the need for graded material grew, there was an extensive push for more scientific methods to measure the grade level of specific texts used in the classroom. The result was the first readability formulas coming in to use in the 1920s. (Wolf, 2013) Much of the early work on the Flesch-Kincaid Grade Level formula was done by Rudolf Flesch in the 1940s (Flesch, 1979). Flesch had been conducting reading studies, observing readers and how they approach long words, and examining punctuation and sentence length. He was an active proponent of “plain English” and was known through many books, in particular *Why Johnny Can't Read: And What You Can Do About*

It (1955). He was a critic of the “look-say” method of teaching reading popular in the 1950s, and he advocated a method for teaching reading that became known as the “look and guess” method and started a revival of phonics which taught learners to sound out words using rules (Blumenfeld, 2015). The Flesch-Kincaid Readability Test and Reading Ease Test came about from the initial work of Rudolf Flesch and subsequent refinements by J. Peter Kincaid. These changes were added by Kincaid while performing work for the United States Navy. Noting that the Flesch-Kincaid grade level was developed for adults, J. Peter Kincaid pointed out:

Among other things we can reasonably measure: the number of commonly understood words, sentence complexity, the number of abstract ideas, and the use of personal pronouns. Beyond these factors, it takes the expertise of the writer and editor to judge organization of the text and whether or not the text conveys the proper information. (McClure, 1987)

In a 1987 interview J. Peter Kincaid stated:

I derived the formula [Kincaid Readability Formula] by testing a large sample of Navy technical personnel on their understanding of Navy technical passages. Next, I determined the personnel’s reading ability level with the Gates-MacGinitie Reading Test. I used a procedure devised for an Army study to calculate the actual equation. (McClure, 1987)

In the intervening decades after the original publication of the Flesch formula, many other formulas were published. It is the Flesch-Kincaid version which has remained the most popular formula over time.

The Flesch Reading Ease formula is given as

$$206.835 - 1.015 \left(\frac{\text{total words}}{\text{total sentences}} \right) - 84.6 \left(\frac{\text{total syllables}}{\text{total words}} \right)$$

(Flesch Reading Ease Readability Score, n.d.), which produces a number between 0 and 120. The higher the number, the greater ease in reading the text. The Flesch-Kincaid Grade Level is given as

$$0.39 \left(\frac{\text{total words}}{\text{total sentences}} \right) + 11.8 \left(\frac{\text{total syllables}}{\text{total words}} \right) - 15.59$$

(McClure, 1987), which then represents the grade level of the text (such as 6th grade level or 17th grade level, for example). Pennsylvania was the first state to require that auto insurance policies are written at the 9th grade level. (McDonnell, 2014)

A 1975 U.S. Navy report discusses in detail the sample size and the types of material which were presented for testing. The Grade Level formula was devised to assist the Navy Department in developing technical manuals so that the information is understandable to personnel who were being trained. The formula helped address the need to accommodate recruits over a wide range of reading skills. Goals included simplifying language through avoidance of synonyms, use of simpler words, and writing with a consistent sentence structure. (Kincaid, J.P., Fishburne, R.P., Rogers, R.L., and Chissom, B.S., 1975)

Nowadays, organizations that make use of Consent Forms (such as university research projects with test subjects) may find that the Institutional Review Board (IRB) will impose a Reading Level of 6th grade. This is often true for health organizations as well when asking patients to sign a consent form. The seemingly low Reading Level grade comes about in the following considerations. Participants called upon to sign a consent may be asked to do so in a short span of time, in an environment which is unfamiliar to them. By imposing a low reading level on the form, it is more likely that the language will be easier for the participant or patient to follow under a stressful condition and they will be much clearer about what is being asked of them. From the perspective of the author of a consent form, achieving the desired precision of language could prove difficult.

HOW TO USE THE FORMULA

Should writers check their own work for reading level? As the preceding discussion shows, the reading level numbers can help focus on problems with a text. The number alone may not have high significance by itself, but it does assume importance when the goals of the writer are to be considered. Who is the intended audience and what was the expected reading level? If these do not match up, it is a quick trigger to spend time editing.

How does the author determine the reading level of his work? Some word processing programs, such as Microsoft Word in its various releases, can be configured to provide the reading level of texts entered. Or texts can be cut and pasted into these programs from other sources and thus the reading level can be checked. Additionally, there are a number of free online websites which provide an online calculation of the reading level, generally by cutting and pasting text into a text window. A recent comparison of Microsoft Word and two online websites using the same texts showed reading level scores ranging over as much as one grade level or more. While this amount is not greatly significant, it does show that the computation of the score is implementation dependent. It would be wise to double-check any values used on multiple sources for the sake of consistency. Websites which accept PDF formats of a document will often show different Reading Level value scores of over a grade level when the PDF is converted to a Microsoft Word document and checked against those tools or websites which support other formats. This again is another indication that reading level checks available to authors are often implementation dependent.

In examining the Reading Level computation, it is noted that 'total syllables' is a component. Depending on how this term is computed, the resulting formula value as calculated may change. Electronic dictionaries used by word processing software or online services need to be up-to-date for a precise calculation. New terms, especially new scientific terms and foreign word borrow-

ings, pose particular concerns, since the rules of syllabification would be different. The actual mix of such terms becomes important if they are an overriding characteristic of the text.

There are several other observations from the formulas which should be noted as well. Both the Flesch Reading Ease and the Flesch-Kincaid Grade Level depend on total sentences and total words. This implies that the score may be more accurate when working with longer texts. Actual numbers on short sections of large texts will often fall within a grade level of each other as when compared to using the full text. Reading Level Grade should be viewed as a guideline rather than an exact indication on the scale of complexity in view of the variances described so far.

Both formulas also show that reading level can be improved (grade level reduced) by breaking longer sentences into shorter structures. To the extent that it is possible without loss of meaning, a more readable text will generally involve using shorter sentences. Consider, for example, the case of a translation from a foreign language: the translator should be aware that the original sentence length should not be preserved in most cases for an English-speaking audience, when content and not poetic elements are concerned. Consider, for example, the case of German scientific text which allows lengthy sentences of complex structure due primarily to the nature of German word construction and German grammar. Such structure should be reworked when the translated German content is presented to an English-speaking audience.

A writer can further determine which sections of his text are most in need of reworking by cutting and pasting selected sections into a word processing program that does the reading level checks. In this way, it is easier to isolate specifically where to make the changes leading towards simplification.

In summary, users of the computations should pay close attention to the following:

- Applying the formulas to longer texts will be more accurate. For example, on a short consent form, one long or complex sen-

tence alone will cause the reading level value to be high, compared to when that longer sentence is broken into shorter sentences.

- Sentence length plays an important role in the calculation. When considering a verbal presentation to an audience, not only is the overall complexity of the text reduced, but listeners will be given a better chance to follow the discussion that they are hearing for the first time.
- Note that neither the Flesch Reading Ease Formula nor the Flesch-Kincaid Grade Level takes into consideration the overall editorial objective of conveying ideas clearly and without undue complexity. These are issues which should be addressed by editors prior to making a determination about reading level of a text using the computations. The formulas should be applied on documents and writings after the editing phase has taken place and clarity of the overall presentation has been worked out.

Users of the formulas should understand that Reading Level calculations need to be evaluated in terms of intrinsic factors of the text which is being analyzed. Technical documents in medicine or chemistry may borrow from Latin and Greek roots and produce a type of English resulting from the construction of neologisms that is not easily measured with evaluative tools. Such situations would suggest that Flesch-Kincaid formulas will not apply well, since these do not take into account the significance of these terms from other sources.

Many writers who have learned a foreign language know that there are often terms which are not easily translatable into English. The converse is also true: many English words do not have direct equivalents in other languages, a fact which poses unique problems to readers from specific cultures. Writers will of course have no idea of the native languages of their potential readers and what pitfalls to avoid in their word choices. This issue looms larger when the context is sci-

entific papers, which are the principal means by which new findings are communicated to broad audiences. Writers of scientific papers can often have limited training in composition beyond their secondary education. Since English is most often the primary discourse for science, speakers whose native language is not English will be writing in a language choice which can pose significant obstacles. While spelling, punctuation and grammar will be corrected by automated tools and careful editors, the conceptual design of the article may reflect the non-English word choices and syntax of the writer's native skills. Scientific integrity does not always correlate with the ability to write simply. Writing is a skill that must be acquired through instruction and practice.

Manual computation of the formulas is ill-advised, especially for longer texts, since this involves counting syllables, words and sentences. A wide range of tools are available online, many of which are freely available (Readability Formulas, n.d.). Free tools cover a range of different formulas and scores, such as the Gunning fog index, the SMOG Index, the Coleman-Liau Index, Automated Readability Index and Linsear Write Formula. Recall that Rudolf Flesch mentions in the 1950s alone two dozen competing formulas existed for the purpose of computing reading level, each with their own specific approaches but offering similar information. A brief digest of these formulas is given in Table 1. Table of Common Reading Level Tools, which follows below.

Many users of the Flesch-Kincaid Grade Level or Flesch Reading Ease Formula will want to rely on Microsoft Word for these computations, which can be configured as part of the word count computations that are available in the tool. Microsoft Word could be used for text created within the tool itself or for texts cut and pasted in the tool. Conversions exist so that Microsoft Word can import PDF files. Details to enable this computation should be determined from the Help file for Microsoft Word, since specific settings may change depending on the software release.

Table 1. Table of common reading level tools

Tool or Formula	Description
Flesch Reading Ease formula	A number from 0 to 120 to evaluate the relative ease of reading a given text. Can be used to compare texts and determine which are more easily read. Lower numbers reflect greater reading difficulty, and higher numbers reflect greater reading ease.
Flesch-Kincaid Grade Level	A decimal number that translates roughly as the grade school, high school or college level grade of text readability.
Gunning fog index	Developed by Robert Gunning to measure number of school years needed to read a given text.
SMOG grade	SMOG stands for the “Simple Measure of Gobbledygook” and attempts to measure the number of years of education needed to understand a text. It was widely adapted in health care.
Coleman-Liau Index	The index is another alternative to the Flesch-Kincaid Grade Level and assigns a grade level number relating to the United States usage of grade levels. The index is intended to be one that can be applied manually from a text.
Automated Readability Index	The index is an alternative to the Flesch-Kincaid Grade Level Formula, and is used for English language materials. The index attempts to supply a grade level number relating to the United States usage of grade levels. This index makes use of characters per word instead of syllables in the word.
Linsear Write calculation	The Linsear Write calculation was developed by the United States Air Force for determining the readability of technical manuals. It relies on manual calculation and takes note of words that have three or more syllables.

Reading Level is an active area of research, particularly in professional areas such as medicine which need to bridge the gap between research findings and helping the public to understand these results. (Neuhoff, 2016) Moreover, writers can benefit from the number of well-crafted websites which provide reading level tools to measure the complexity and accessibility of their writing, whether fiction or non-fiction. (Readability Score, n.d.) Researchers are addressing issues such as achieving useful, readable publications and working with readership. Surprisingly, lower reading level can correlate positively with commercial popularity of publications.

CHALLENGES ON THE INTERNET

English has become the de facto standard on the Internet. This fact alone simply informs us that many users who are not native English speakers are reading from and producing Web information. Depending on the countries of origin of the readers and writers, access to resources for reading and writing in English can vary greatly. Even where resources for language learning and language reference are available on the Internet,

questions about bandwidth, firewalls and paywalls, and device capabilities start to play a role. Online tools are now available to check the reading level of a website. (Checking the Reading Level of Your Website, n.d.)

Reading Grade Level formulas are generally not discussed outside the English-language context. Several considerations can help to explain this phenomenon.

- The English language has among the highest word counts of all languages. Depending on estimates used and the types of words included, totals run from several hundred thousand ordinary words to over one million for complete dictionaries with foreign borrowings and all scientific terms and chemical names. (Global Language Monitor, 2014)
- English grows rapidly over time due the force of the language as the principal language of science with the resulting inclusion of new terminologies to accommodate the advancement of sciences. According to a recent estimate, a new word comes into the English language every 98 minutes. (Global Language Monitor, 2014)

- English is widely taught to students of all levels of income, background and ethnic origins, due in part to the United States educational policies of inclusion, perhaps among the widest number of groups, cultures and backgrounds.

Thus, English presents special concerns in its ability to allow widely different kinds of texts with a proliferation of related words offering fine shadings of meaning. The fact English is a growing language that borrows freely from all areas of the world makes it challenging to define proficiency. In order to successfully reach a diverse audience, measures such as Reading Grade Level computation can help to a greater degree by bringing standards to the evaluation of texts for use in a range of educational purposes. Many other language groups have relatively restrictive boards or organizations which monitor new material coming into the language and reject newly constructed foreign terminologies in favor of native word formations or native word roots.

CONSIDERATIONS FOR NON-NATIVE ENGLISH USERS

Today the abundance of translation tools (browsers and applications) reduces the need for the reader to rely on English solely and may in fact over time weaken English language skills due to lack of practice. When it comes time to write in English, the non-native writer would be facing a steeper climb as the result of nonuse of his language skills. In effect translation tools reduce the importance of mastering English unless the user needs to write in English. This scenario further supports the idea of English as a base global language, regardless of the reader's native tongue, since he can easily get to an equivalent in his mother tongue from English, whatever the native language is. The translation tools further the idea that English can retain its prominence in terms of language representation over the long range, carrying the

idea into further generations of Web users. It may also be of significance that in U.S. higher education foreign language study has declined through the weakening of language requirements and the decline in faculty positions. The implication here is that those writing in English-speaking countries become less aware to the complexities in writing for a foreign language audience.

A faculty member could check planned readings from the Web before these become assignments. Rather than omitting specific readings, more preparation could be given to students in the class when reading level scores show very high values for texts that will be assigned reading. It is easy to select sections from a text and examine reading level by cutting and pasting sections into a tool which does the Flesch-Kincaid calculation. While native speakers of English may not experience difficulty with many texts--even when the reading levels are high--many classrooms today have a mix of students from a variety of ethnic backgrounds and reading abilities. Preparing focus questions for review that single out the difficult sections of the reading can prove valuable in bringing student understanding forward.

When faculty members are developing an article or a presentation, working with the Flesch-Kincaid calculation could prove beneficial. A verbal presentation will benefit from a lower reading level value, even when publication is intended. When listeners are seated for a presentation, they can be greatly helped if the material has been screened for unnecessary complexity. But it should never be assumed that Flesch-Kincaid will detect all problem areas. Areas that can be focused on when delivering a written talk verbally include the following: long sentences, complex sentences, and word length. It is important to stress that oversimplification in this area is not intended, but it is a way of calling attention to areas which can create challenges for listeners. Naturally, careful inflection of words, insertion of pauses in materials presented, and slowing down during difficult passages can be especially beneficial and more productive than rewriting.

Due to the need to have published conference proceedings, presentations are frequently given as material read before an audience. The speaker's familiarity with the content may further result in obscuring the content since the challenges of a first listen will be lost to the speaker. Awareness of these issues is the best insulation to preventing them from happening.

A number of other methods can be used in conjunction with Flesch-Kincaid to assist with producing documents that are to be read before a group. Many researchers have showed that when writers make use of multiple senses (e.g. hearing, sight) during writing, the result can be significant. It is therefore reasonable to bring sound into editing a text. A number of editing tools will read back in spoken form text that is being written. This allows the writer to bring in his sense of hearing along with the writing skill in a way that can produce a more acceptable result. When writing, using a voice recorder can help in a number of ways by taking advantage of playback. Spoken material will allow additional skills to come into use, and the writer will be able to both hear and record his intentions. A number of tools exist to take the recorded voice file and convert to text in a highly accurate manner. Or the writer can use one of the computer tools that directly convert spoken material into computer text. Another method that is of great value is to be able to listen to existing text from a computer file spoken by a variety of voices, male and female and with various voice styles or accents.

FUTURE RESEARCH DIRECTIONS

The topic of reading level continues to reemerge in very active discussion among educators due to a variety of issues. Generations of teachers with distinct differences in how reading should be presented have come and gone. New forms of media have come about from the growth in technologies. The entire public is called upon to view more information, often in electronic format, with

little guidance on how to evaluate what is being read. Questions about literacy come to forefront as standardized tests show areas of deficiency even though methods abound to address such concerns. As schools change their structures and reinvent themselves, where do the traditional skills that had emphasis in the past now fit in a new place?

Current research in reading now includes bringing in new technologies such as fMRI, which can identify the areas of the brain that are involved in specific activities relating to reading. In fact, it is possible to make use of fMRI in many cases to pinpoint those students who are currently experiencing or will encounter reading difficulties. (Dehaene, 2009) As a greater understanding of the biological aspects of reading is gained, it will be possible to redesign reading programs to address specific brain functions and more precisely measure results.

CONCLUSION

Reading Level is an important tool for publishers for a variety of audiences, from students learning new material to those who are studying training manuals to become proficient in a new skill. The Flesch-Kincaid Grade Level can be helpful to authors in creating their texts by providing an objective measure of the complexity of the material. The resulting computation can indicate to the author if the material is suitably targeted for the intended audience. When Internet publication is planned, this will also help with reaching non-English audiences who may experience difficulty with more complex texts as determined by the Grade Level computation. Faculty members can also make use of the Flesch-Kincaid Grade Level as a measure of complexity when making reading assignments. In the case where the material is especially difficult, the techniques described here can point to the need for further follow-up classroom discussion or giving additional information prior to reading assignment to alert students to the more challenging material. (Readability Score, n.d.)

REFERENCES

- Blumenfeld, S. (2015). *The Legacy of Rudolf Flesch*. Retrieved July 19, 2016, from <http://www.thenewamerican.com/reviews/opinion/item/19903-the-legacy-of-rudolf-flesch>
- Checking the Reading Level of Your Website. (n.d.). Retrieved July 19, 2016, from <http://etc.usf.edu/techease/4all/web-accessibility/checking-the-reading-level-of-your-website/>
- Dehaene, S. (2009). *Reading in the Brain: the new science of how we read*. New York: Penguin.
- Flesch, R. (1979). *How to Write Plain English. Chapter 2: Let's Start with the Formula*. Retrieved July 19, 2016, from <http://pages.stern.nyu.edu/~wstarbuc/Writing/Flesch.htm>
- Flesch Reading Ease Readability Score. (n.d.). Retrieved July 19, 2016 from <http://www.rfp-templates.com/readability-scores/flesch-reading-ease>
- Global Language Monitor. (2014). *Number of Words in the English Language: 1,025,109.8*. Retrieved July 19, 2016 from <http://www.languagemonitor.com/number-of-words/no-of-words/>
- Kincaid, J. P., Fishburne, R. P., Rogers, R. L., & Chissom, B. S. (1975). Derivation of New Readability Formulas (Automated Readability Index, Fog Count, and Flesch Reading Ease Formula) for Navy Enlisted Personnel, Research Branch Report 8-75, Chief of Naval Technical Training, Naval Air Station, Memphis, Millington, TN. Retrieved July 19, 2016 from <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6449109>
- McClure, G. M. (1987). Readability Formulas: Useful or Useless? *IEEE Transactions on Professional Communication*, 30(1), 12–15. doi:10.1109/TPC.1987.6449109
- McDonnell, P. J. (2014). Using hard words (or, if you prefer, complex verbiage). *Ophthalmology Times*. Retrieved July 19, 2016, from <http://ophthalmologytimes.modernmedicine.com/ophthalmologytimes/content/tags/flesch-kincaid-readability-grade-level/using-hard-words-or-if-you-pr?page=full>
- Neuhoff, E. (2016). *Flesch-Kincaid Reading Grade Level Re-examined: Creating a Uniform Method for Calculating Readability on a Certification Exam*. Retrieved November 25, 2016, from <http://opensiuc.lib.siu.edu/cgi/viewcontent.cgi?article=1189&context=ojwed>
- Readability Formulas. (n.d.). Retrieved July 19, 2016, from <http://readabilityformulas.com/>
- Readabilty Score. (n.d.). Retrieved November 25, 2016 from <https://readability-score.com/>
- Snow, S. (2015). *This Surprising Reading Level Analysis Will Change the Way You Write*. Retrieved November 25, 2016, from: <https://contently.com/strategist/2015/01/28/this-surprising-reading-level-analysis-will-change-the-way-you-write/>
- Training, Naval Air Station, Memphis, Millington, TN. (n.d.). Retrieved July 19, 2016 from <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6449109>
- Wolf, R. A. (2013). *The Flesch Readability Formula: Still alive or still life?* Retrieved July 19, 2016, from <http://essay.utwente.nl/64204/>

ADDITIONAL READING

- Larson, E., Foe, G., & Lally, R. (2015). Reading Level and Length of Written Research Consent Forms. *Clinical and Translational Science*, 8(4), 355–356. <http://onlinelibrary.wiley.com/doi/10.1111/cts.12253/abstract;jsessionid=8AB7AACD4DC4291DBB55B7E254457B79.f04t03> Retrieved November 27, 2016 doi:10.1111/cts.12253 PMID:25580939

Lyon, G. R. (1998). Why Reading is Not a Natural Process. *Educational Leadership*, 55(6), 14–18. <http://www.ascd.org/publications/educational-leadership/mar98/vol55/num06/Why-Reading-Is-Not-a-Natural-Process.aspx> Retrieved November 27, 2016

Lyon, G. R., & Chhabra, V. (2004). The Science of Reading Research. *Educational Leadership*, 61(6), 12–17. <http://www.ascd.org/publications/educational-leadership/mar04/vol61/num06/The-Science-of-Reading-Research.aspx> Retrieved November 27, 2016

Moats, L. C. (1999). *Teaching reading is rocket science*. Washington, DC: American Federation of Teachers. Retrieved November 27, 2016, from http://www.aft.org/sites/default/files/reading_rocketscience_2004.pdf

Pondisco, R. (2014). *Leveled reading: The making of a literacy myth*. Retrieved November 27, 2016, from <https://edexcellence.net/articles/leveled-reading-the-making-of-a-literacy-myth>

KEY TERMS AND DEFINITIONS

Automated Readability Index: The index is an alternative to the Flesch-Kincaid Grade Level Formula, and is used for English language materials. The index attempts to supply a grade level number relating to the United States usage of grade levels. This index makes use of characters per word instead of syllables in the word.

Coleman-Liau Index: The index is another alternative to the Flesch-Kincaid Grade Level and assigns a grade level number relating to the United

States usage of grade levels. The index is intended to be one that can be applied mechanically from text. It relies on the average number of letters per 100 words and the average number of sentences per 100 words.

Flesch Reading Ease Formula: The Reading Ease Formula was developed by Rudolf Flesch to provide a number between zero and 120 to represent the difficulty of a text. Theoretically, there is no lower bound to the number assigned. Lower numbers represent greater difficulty and higher numbers represent greater ease of readability. For example, *Reader's Digest* is around 65 and the *Harvard Law Review* is around 30.

Flesch Kincaid Grade Level Formula: The Flesch-Kincaid Grade Level Formula was originally developed by Rudolf Flesch and further revised by J. Peter Kincaid while working on evaluating the reading level of technical manuals for the United States Navy.

Gunning Fog Index: Robert Gunning developed the Gunning fog index in the 1950s to measure the number of years of school that would be required to read a given text. For example, a fog index value of 17 would require a college graduate.

Linsear Write Calculation: The Linsear Write calculation was developed by the United States Air Force for determining the readability of their technical manuals. It relies on manual calculation and note of words that have three or more syllables.

SMOG Grade: SMOG stands for the “Simple Measure of Gobbledygook” and attempts to measure the number of years of education needed to understand a text. It was widely adapted in health care.

Multimodal Literacy

Maryann Tatum Tobin

University of Miami, USA

INTRODUCTION

Reading in today's technology-rich world has become an increasingly complex maze of symbols, systems, and patterns. Current research in the field of literacy proposes an updated, more robust definition of reading that extends beyond traditional printed text to include reading-to-learn as a multimodal activity incorporating emerging technologies into daily literacy instruction. Heath & Street (2008) define multimodal literacy as the ability to navigate the "combined representations of written, oral, visual, and gestural systems of communication." However, the implication on educators is such that additional modifications to pedagogy must take into account the ever-changing nature of media. Generational terms such as *digital natives* (those born after the year 1980s) (Prensky, 2006) and *millennials* (those reaching young adulthood sometime around the year 2000) hardly encapsulate the huge difference between growing up with an offline home/micro-computer in your home to the vast array of knowledge available by spending your childhood with a smart phone in your hand. Digital natives process information in a vastly different way than previous generations, however, "native-ness" with technology and the ability to use it to learn in efficient ways are not always synonymous. Facility with social media and multitasking does not equal deep understanding or a natural instinct on how to program or create media for and about learning. Curriculum and schooling has begun to change to accommodate this world of hypertext, distance, and simulated learning, but many scholars find that change has not come fast enough to keep up with the pace required by the exponential growth of technology. For example, in the field of literacy

education and the teaching of reading, the effects of technology on the classroom environment, and most-specifically how the truncated and fast-paced world of online, shallow reading has affected the reading and learning stamina of children saturated in media (Carr, 2010). These critical factors are what will dictate the knowledge capacities and learning modalities of today's children and future generations.

BACKGROUND

Since the latter part of the 20th century, the traditional notion of literacy has been redefined. Just as the invention of the printing press moved society into the typographic culture, now we are in what Provenzo (1986) termed the "post-typographic culture," where the definition of a literate person now includes affinity with the computer, media, and visual literacies. In the post-typographic world, microcomputers revolutionized commerce and news media, but education is still late to the game. There are pockets of classroom use, but by comparison to how, for example, marketing, was transformed by social media and online metrics, education has largely seen isolated use of technology. However, within the last decade, tablet and portable device-driven learning has begun to see a more comprehensive change, and print books are taking a "back seat" to multimedia in the school life of students. Print-based texts contain elements of written language, images, and design elements such as typography and graphic design, but media-rich texts involve moving images, sound effects, and increasingly robust digitally-rendered elements that must be "read" by students (Sarafini, 2014).

DOI: 10.4018/978-1-5225-2255-3.ch130

Many students are immersed in media-centered environments that are different from classrooms of the past. This new technology comes with several acquired understandings for young readers, as they must learn to maneuver their way through both print text and a series of icons, images, and symbols. Literacy instruction, therefore, the “becoming literate” has moved from an ability to navigate and decode a printed text for the purposes of gaining new knowledge to the ability to access, manipulate, and apply information from a variety of print and non-print courses (Gee, 2007). These new literacies require a pedagogical revolution that incorporates processing a variety of modalities simultaneously in the classroom. Engagement with robust new literacies helps prepare students to not just be receptacles of information, but to think critically, observe and evaluate like scientists and historians (Abilock, 2001). To foster inquiry and project-based learning, schools are developing links between content area teachers and the central texts (both print and digital) through *disciplinary literacy*, that is the confluence of content knowledge, personal experiences, and the ability to read, write, listen and speak academically (Wineburg & Resiman, 2015).

Engaging students in open-ended, authentic, problem-based tasks using technology has become the key to this kind of transformative learning (Wiggins & McTighe, 1998). Technology has been, for many disciplinary literacy scholars, a blessing in that a myriad of open access resources are now available to learners. However, this blessing has also created a curse. Most of the text available online are less dense and more truncated than ever before. Imagine the difference of reading the New York Times cover to cover versus getting the news in pop-up updates from your CNN app. The fast-paced world of online reading is the demand of the day. Although we may long for the times when deep reading was valued, the machine of automation has driven society forward and the field of literacy education is no different. To paraphrase Marshall McLuhan, we cannot drive the information super-highway by looking in the

rear-view mirror (1964). However, literacy educators today fight a constant battle to combat the foundational pangs of “shallow-reading,” including poor reading stamina, shorter attention spans, and the inability of students to show motivation to read anything that appears time consuming or lacking in value (Carr, 2010).

Digital Natives?

In the literature, *digital natives* and *millennials* are often used interchangeably, although there has been some considerable debate about whether or not the distinction still holds relevance. Both terms have come to refer to the same social group: those born after 1980 that have become working adults around the turn of the 21st century. The modern educational system is currently populated by both *digital natives* and *digital immigrants* (those who have learned to use the technology as adults). In fact, many of today’s teachers are also by definition *digital natives*, although many argue that the level of native-ness varies greatly between those born in the 1980s and those born today. Having a computer with dial-up Internet in your home as a child is quite different than having a smart phone in your hand at all times. Classroom teachers must navigate the distance between their own technological knowledge and skill and the technological fluency of their students, and that savvy with technology has been pared down to a fundamental difference between how digital natives and digital immigrants think.

McLuhan commented that “we shape our tools and thereafter, our tools shape us” and part of that shaping includes reshaping how students acquire, retain, and exchange information. Take communication, for instance. To a digital immigrant, long-distance telephone communication was something that was once expensive and limited, almost a novelty used only by those who could afford it. Letter writing and post-cards were the medium of choice when communicating with those distant cousins. But, to a digital native, real-time connections across the world via text messaging

services and apps are commonplace and in many cases, free. In fact, in an effort to abbreviate communication even more so, truncated texts, slang, and emoticons provide for casual multimodal connection so that natives can efficiently and thriftily communicate information in fewer characters or exclusively through images.

Digital natives act as both content consumers and content creators. Termed “Dot Nets” by Pew’s Research Center because of their facility with navigating the complex geography of the Internet, millennials are comfortable using a diverse range of media to

multitask, and they “demand interactivity as they construct knowledge” (Bauerlein, 2009, pg. 73). This issue of multitasking, though, has mixed challenges for literacy educators. During homework, for example, students often deal with multiple inputs from friends, family, and school all simultaneously. This presents a few important issues for educators (and parents) to consider:

- Mental downtime from media to encourage the development of other parts of the brain.
- Sleep disruption due to too much digital stimulation.
- Other distractions produced by digital devices (such as driving distractions).
- Addiction behaviors towards digital devices (Steyer, 2012).

Splitting focus is nearly impossible, even for the mature brain. Humans are unable to effectively process two streams of informational stimuli at once, and while hypertext has been credited with increasing informational access, reading motivation, fast problem-solving and data-juggling, it is often also related to limited attention span, skimming, cognitive overload, and diminished long-term memory (Steyer, 2012). Much of homework multitasking involves double screens and browsers open to television and video services, like Hulu and Netflix, and the market certainly has not missed this opportunity. In fact, in 2014, Nielsen’s rating service began measure YouTube view and found

that it reaches more 18-35 year olds than any cable network (YouTube Statistics, 2016).

This culture shift has called for many technology activists, like Marc Prensky and Jonathan Fanton, to declare that perhaps the richest learning environment we could offer young people is no longer the traditional classroom. *Flipped learning* allows for students to take advantage the web’s wealth of content independently and then come to the classroom for teacher-facilitated cooperative inquiry. Fine for the student who reads well and learns well independently, but quite a challenge for the fifth grader who still struggles with basic literacy skills like decoding and phonics. Naysayers of flipped learning model critique the reliance on low-income students’ home access to efficient connectivity, those who have to share limited computer time with family, or those students who simply do not have the time and focus to watch a direct instruction lesson online. However, the flipped model does show benefits in extending teacher time during the day and allowing for more time for students, especially those with exceptionalities, to have more one-on-one time with the teacher (Weisen, 2014).

A challenge of such increased screen time is that screen intelligence doesn’t always transfer to print experiences (Bauerlein, 2009). While the Internet’s saturation of videos and music does train the brain to be adept at processing multiple stimuli simultaneously, what it does not promote is understanding for deep meaning. Media-savvy minds are open and flexible to what grabs their attention, but the “digital mind” turns away from the one-dimensional printed text. In fact, many of the new standards for college and career readiness purported by educational systems urge the use of technology to “master the kinds of higher-order thinking and decision-making skills employers seek today” (Federation of American Scientists, 2006).

So how do we reconcile the fact that today’s college graduates have spent less than 5,000 hours of their lives reading and over 20,000 hours on the computer, watching TV and playing video

games (Prensky, 2011)? Are we simply leveling up to meet the promise of the new world, just as 15th-century civilians did with the arrival of Gutenberg's printing press, or are we fulfilling a Kubrick-like philosophy of succumbing to the individuality-numbing of artificial intelligence (Carr, 2011)? Fatalists would believe the latter. But, educators, who have been confronted since the dawn of modern schooling with the "latest and greatest" new fads have learned to adapt instruction to meet the ever-changing needs of young minds. As an example, Gee (2007) presents several principles of learning in his study of the self-directed learning that takes place as students engage in playing video games, including critical learning, the use of semiotics, the development of a metalanguage for and about learning, and the understanding that text is an "embodied experience." As learners negotiate multimodal texts, initial, mostly verbal understandings, give way to larger, cross-modal understanding that involve the "reading" of images, sounds, renderings, actions, and consequences.

In the case of reading, which is the first window onto the world presented to any young mind, preliminary exposure to texts in the modern toddler is often digital in nature (mom's iPhone or dad's e-reader). But it is still exposure to print nevertheless. In fact, the draw of technology is often a gateway to new and deeper exposure to text (Tobin, 2012). Reading teachers and educators who tackle reading in the content areas have come to adopt a new and deeper paradigm for creating literate learners.

Multimodal Literacy Learning

The term "literacy" and the meanings to which it refers have changed in recent years. People who could not read and write printed letters were often considered to be illiterate, as early definitions of literacy encompassed decoding, mastering the alphabetic principle, effectively communicating with texts, learning grammar and sentence structure, comprehension, and becoming an expressive,

persuasive, or informational writer (Moje, 2003). However, the advent of technology in schools extended the notion of literacy to "the practice of navigating many different symbol systems and discourse communities to make meaning from and with written text" (Moje & Sutherland, 2003, p. 152). People who can only read and write alphabetic letters are viewed as lacking the appropriate literacy skills vital to the digital age, which is characterized by images, sound, videos, and other texts on various screens. Thus, the term *multimodal literacy* has emerged, representing the combined representations of written, oral, visual, and gestural systems of communication provided by technology or other media (Heath & Street, 2008; Kress, 2000).

The main difference between reading in a print-only world and reading online is that text in a world of hypermedia is rife with choices. The market is such that we are inundated with images and distractors that encourage us to click away from what we are reading or doing online. The pitfalls are many, not the least of which is to develop in young readers the core competencies of research: targeted reading, making informed judgements, and culling information from a variety of sources, which is termed *informational literacy* (Gilster, P., 2000). Digital immigrants, therefore, were taught to read in a print world but have learned to adapt their reading focus and stamina to the online world. This dichotomy is illustrated by Knobel and Lankshear (2007) who developed a two-threaded definition of "little 'l' literacy and big 'L' literacy." Little 'l' literacy is the traditional, print-based, foundational view of literacy: the actual processes of reading and writing. In contrast, Big 'L' literacies are connected to the patterns and discourse that integrate ways of communicating through gestures, actions, spoken and written words, images, and sounds. This kaleidoscope of literacies is still conceptualized in relations to text, but now *text* is defined as any instance of communication using any semiotic resource (Kress, 2003). Research has found that the inclusion of new technologies, even with

early readers, provides for new resources for both students and teachers, and the potentialities to increase the development of little 'l' literacy skills.

(Leu, Kinzer, Coiro, & Cammack, 2004; Lankshear & Knobel, 2006; Ranker, 2008). One of the fastest growing big 'L' literacy is *media literacy*, which is defined in relation to a person's attitude toward, evaluation of, and experience in reading texts and designing hypertexts made possible by technology (Hobbs, 2007). A key tenet of multimodal literacy learning is that digital natives draw from a prolonged access to social networks and informational materials, computer editing programs, and constantly changing communication technologies. So, being literate in today's media-rich world includes recognizing and negotiating the social positionings inherent in the exchange of both prolonged and immediate interactions with a variety of multimodal texts (Alvermann & Hagood, 2000).

Shallow Reading

While the average child today may spend exponentially more time in front of a screen than their predecessors, their facility with technology is often limited to "surface technologies" like texting and social media. This parallels what Nicholas Carr calls "shallow reading" – where skimming and scanning are valued and contemplation and reflection are lost (Carr, 2010). What we need is a way to bring deeper and more meaningful interactions with technology into the classroom so that students are exposed to more complex digital tools like HTML codes, programming, and spreadsheets, and we need to train educators to bring deeper intellectual discussions to social media platforms for student consumption, so that the surface technology becomes an outlet for rich discussion.

Careers in the sciences, particularly medicine, still rely heavily on linear thinking and development, as opposed to the creativity and entrepreneurial thinking that may await a career at Google. So there is still a need to focus on little 'l' literacy

learning methods to some extent in classrooms, otherwise reading stamina and focus is lost in the fragmented world of online reading (Carr, 2010). The world cannot be "de-technologized," so educators must learn to teach linear processes in a non-linear world.

FUTURE RESEARCH DIRECTIONS

In order to train literacy educators to be successful in preparing young minds for whatever demands are placed on them regardless of career path, a healthy balance of collaborative learning and "do-it-on-your-own" learning must be presented. We must be prepared to plan literacy across language, modes, media, ideologies, and cultures (Hull & Nelson, 2009). Old paradigms have been hard to change, but most school systems have accepted the call to implement best practices for 21st century learning, including shifting the instructional focus from the teacher to the student (Tapscott, 2009).

Broadcast Learning vs. Interactive Learning

This is the main shift in education that needs extensive continual study, especially given the high-stakes testing mentalities that the latest research is working hard to combat. The focus on learning needs to not be about what students know when they graduate but on their capacity for lifelong learning and their metacognitive awareness of how they learn and collaborate best. When presented, for example, with a multimodal literacy lesson, which modalities do they prefer to use and which modalities help them learn more efficiently. Each modality comes with its own semiotic metalanguage for and about the activity. In other words, each learning tool comes with its own set of rules for interactions. If a reader finds that listening to music helps them focus, then the teacher needs to encourage that. If stopping for questions or supplementing a text with video improves comprehension, then the

teacher needs to incorporate than into his/her lesson planning. This is the malleable nature of a truly multimodal classroom. Broadcast learning, which is very teacher-centered with a “sage on stage” is passé. Interactive learning should rule the day. Classrooms should be learner-centered, collaborative, and based in inquiry and discovery (Tapscott, 2009).

Micro-Level Literacy

With all the focus on multimodal literacy and its affect on changing the way we view reading and thinking as a society, we must not lose sight of the foundational, little ‘I’ literacy practices that promote linear learning. Micro-level literacy promote basic skills that transition across subject areas and are transferable resources available to students throughout their academic lives (Ivanic, 2010). Previous micro-literacy studies focused on transferring skills in the narrow framework of class-to-class – in other words, making sure that what is learned in a reading class is able to be used independently to understand a text in history or science classes. Teaching students to scale their micro-level skills to the macro-level in the digital world provides a framework for refining communication and inquiry within the global community (Brown & Slagter van Tryon, 2010). Understanding the micro aspects of something instantaneous like a tweet and how it can impact networking on a macro-level is a key 21st century literacy skill that needs further study in the context of the digital world.

The Global Marketplace

This is an unparalleled time in literacy education, where new technologies have provided opportunities to create educational paradigms that cross cultural and linguistic boundaries (Zapata & Roach, 2011). The ideology of multimodality is one that promotes social semiotics and transcends the boundaries placed on us by global borders. We must study how to best prepare young consumers

of knowledge to be aware of the hypermobility that being literate provides and the positions of success granted to those who attain literacy. We must also study the effects of those who are not advantaged in greater detail, as the digital divide can continue to stratify those without access on broader, more global levels.

CONCLUSION

We read everyday. We read text, facial expressions, actions, signs, patterns, and symbols. Most of what we do as literate members of society is automatic, as we have reached a level of fluency, learning to read with print and adapting to the digital world. For digital immigrants, while seemingly disadvantaged by the digital native-ness of the young, the benefit of learning in a micro-literacy world and adapting to a macro-literacy world must not be diminished. It is this facility that allows for the true understanding of how to develop curriculum for a multimodal world. Through multimodal literacy, teachers and students, collaboratively, can explore the changing nature of reading using different semiotic resources across a variety of modes. Adapting instruction and schooling to this new reality will develop more robust learners prepared for the challenges of the rest of this century and beyond.

REFERENCES

- Abilock, D. (2001). Using technology to enhance student inquiry. In J. L. LeBaron & C. Collier (Eds.), *Technology in its place*. San Francisco, CA: Jossey-Bass.
- Alvermann, D., & Hagood, M. (2000). Critical media literacy: Research, theory, and practice in New Times. *The Journal of Educational Research*, 93(3), 193–205. doi:10.1080/00220670009598707

- Bauerlein, M. (2009). *The dumbest generation: How the digital age stupefies young Americans and jeopardizes our future or, don't trust anyone under 30*. New York, NY: Penguin.
- Brown, A., & Slagter van Tryon, P. (2010). Twenty-first century literacy: A matter of scale from micro to mega. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 83(6), 235–238. doi:10.1080/00098655.2010.484438
- Carr, N. (2010). *The shallows: What the internet is doing to our brains*. New York, NY: W. W. Norton & Company.
- Carr, N. (2011). Is Google making us stupid? In M. Bauerlein (Ed.), *The digital divide: Arguments for and against Facebook, Google, texting, and the age of social networking*. New York, NY: Penguin.
- Federation of American Scientists. (2006). *Summit on Educational Games: Harnessing the Power of Video Games for Learning*. Author.
- Gee, J. (2007). *What video games have to teach us about learning and literacy* (2nd ed.). New York, NY: Palgrave Macmillan.
- Gilster, P. (2000). Digital literacy. In R. Pea (Ed.), *The Jossey-Bass reader on technology and learning*. San Francisco, CA: Jossey-Bass.
- Heath, S., & Street, B. (2008). On ethnography: Approaches to language and literacy research: Vol. 4. *Approaches to language and literacy research: An NCRL research in language and literacy series*. New York: Teachers College Press.
- Hobbs, R. (2007). *Reading the media: Media literacy in high school English*. New York: Teachers College Press.
- Hull, G., & Nelson, M. (2010). Literacy, media, and morality: Making the case for an aesthetic turn. In M. Baynham & M. Prinsloo (Eds.), *The future of literacy studies* (pp. 199–227). New York, NY: Palgrave Macmillan.
- Ivanic, R. (2010). Bringing literacy studies into research on learning across the curriculum. In M. Baynham & M. Prinsloo (Eds.), *The future of literacy studies* (pp. 100–122). New York, NY: Palgrave Macmillan.
- Knobel, M., & Lankshear, C. (2007). Online memes, affinities, and cultural production. In M. Knobel & C. Lankshear (Eds.), *A new literacies sampler* (pp. 199–227). New York: Peter Lang.
- Kress, G. (2003). *Literacy in the new media age*. London: Routledge. doi:10.4324/9780203164754
- Lankshear, C., & Knobel, M. (2006). *New literacies: Everyday practices and classroom learning* (2nd ed.). Maidenhead, NY: Open University Press.
- Leu, D., Kinzer, C., Coiro, J., & Cammack, D. (2004). Toward a theory of new literacies emerging from the Internet and other information and communication technologies. In R. Ruddell & N. Unrau (Eds.), *Theoretical models and processes of reading* (5th ed., pp. 1570–1613). Newark, DE: International Reading Association.
- Levy, P. (1997). *Collective intelligence: Mankind's emerging world in cyberspace*. Cambridge, MA: Perseus Books.
- McLuhan, M. (1964). *Understanding media*. Cambridge, MA: MIT Press.
- Moje, E., & Sutherland, L. (2003). The future of literacy teacher education. *English Education*, 35(2), 149–164.
- Prensky, M. (2006). *Don't bother me mom, I'm learning*. St. Paul, MN: Paragon House.
- Prensky, M. (2011). Digital natives, digital immigrants. In M. Bauerlein (Ed.), *The digital divide: Arguments for and against Facebook, Google, texting, and the age of social networking*. New York, NY: Penguin.

Provenzo, E. (1986). *Beyond the Gutenberg galaxy: Microcomputers and the emergence of the post-typographic culture*. New York, NY: Teachers College Press.

Ranker, J. (2008). Composing across media in digital video productions. *Written Communication*, 25, 196–234. doi:10.1177/0741088307313021

Sarafini, F. (2014). *Reading the visual: An introduction to teaching multimodal literacy*. New York, NY: Teachers College Press.

Shaw, D., Becker, H. J., Bransford, J. D., Davidson, J., Hawkins, J., Malcom, S., & Lehoczky, S. et al. (1998). Report to the President on the Use of Technology to Strengthen K-12 Education in the United States: Findings Related To Research and Evaluation. *Journal of Science Education and Technology*, 7(2), 115–126. doi:10.1023/A:1022539822660

Steyer, J. (2012). *Talking back to Facebook*. New York, NY: Scribner.

Tapscott, D. (2000). Growing up digital: The rise of the net generation. In R. Pea (Ed.), *The Jossey-Bass reader on technology and learning*. San Francisco, CA: Jossey-Bass.

Tapscott, D. (2009). *Grown up digital: How the net generation is changing your world*. New York: McGraw-Hill.

Tobin, M. (2012). Digital storytelling: Reinventing literature circles. *Voices from the Middle*, 20(2), 35–43.

Weisen, N. (2014). Flipping the classroom for students with learning disabilities. *Scientific Learning*. Retrieved from <http://www.scilearn.com/blog/-/284>

Wiggins, G., & McTighe, J. (1998). *Understanding by design*. Alexandria, VA: Association for Supervision and Curriculum Development.

Wineburg, S., & Resiman, A. (2015). Disciplinary literacy in history: A toolkit for digital citizenship. *Journal of Adolescent & Adult Literacy*, 58(8), 636–639. doi:10.1002/jaal.410

YouTube Statistics. (2016). Retrieved from <https://www.youtube.com/yt/press/statistics.html>

Zapata, A., & Roach, A. (2011). *Movement in literacy: New directions in multilingual, multicultural, multinational, and multimodal literacy studies*. The National Council of Teachers of English.

ADDITIONAL READING

Carr, N. (2015). *The glass cage: How our computers are changing us*. New York, NY: W. W. Norton & Company.

Crockett, L., Jukes, J., & Churches, A. (2011). *Literacy is not enough: 21st century fluencies for the digital age*. Thousand Oaks, CA: Corwin Press.

Issacson, W. (2014). *The innovators: How a group of hackers, geniuses, and geeks created the digital revolution*. New York, NY: Simon & Shuster.

Jacobs, H. (2014). *Mastering Digital Literacy*. Bloomington, IN: Solution Tree.

Palfrey, J., & Gasser, U. (2008). *Born digital: Understanding the first generation of digital natives*. New York, NY: Basic Books.

Perkins, D. (2014). *Future wise: Educating our children for a changing world*. San Francisco, CA: Jossey-Bass.

Prensky, M. (2012). *Brain gain: Technology and the quest for digital wisdom*. New York, NY: St. Martin's Press.

Trilling, B., & Fadel, C. (2012). *21st Century Skills: Learning for life in our times*. San Francisco, CA: Jossey-Bass.

Turkle, S. (2012). *Alone together: Why we expect more from technology and less from each other*. New York, NY: Basic Books.

KEY TERMS AND DEFINITIONS

Big “L” Literacy: Literacy connected to the patterns and discourse that integrate ways of communicating through gestures, actions, spoken and written words, images, and sounds (Knobel & Lankshear, 2007).

Digital Immigrants: An individual born before the widespread growth of digital technologies.

Digital Natives: Being “born digital,” “native speakers” of the digital language of computers, videos, video games, social media and other sites on the internet (Prensky, 2001).

Disciplinary Literacy: The confluence of content knowledge, personal experiences, and the ability to read, write, listen and speak academically (Wineburg & Resiman, 2015).

Flipped Learning: An educational model in which content is delivered to the student independently through videos and web content, followed by classroom times for teacher-facilitated cooperative inquiry.

Hypertext: Text displayed on a computer or other digital devices that has embedded hyperlinks,

images, or media that can be accessed to enhance the primary text.

Informational Literacy: The ability to access and cull information from a variety of print and non-print sources.

Little “l” Literacy: The traditional, print-based, foundational view of literacy: the actual processes of reading and writing (Knobel & Lankshear, 2007).

Media Literacy: The experience of reading texts and designing hypertexts made possible by technology (Hobbs, 2007).

Millenials: A demographic cohort representing anyone born between 1980-2000.

Multimodal Literacy: The design of discourse (for example, language, gesture, images) co-deployed across various modalities (for example, visual, aural, somatic) as well as their interaction and integration in constructing a coherent text.

Shallow Reading: The practice of reading quickly, in truncated forms (e.g. texts, images, memes), without critical analysis.

Text: Any instance of communication using any semiotic resource (Kress, 2003).

An Open Learning Format for Lifelong Learners' Empowerment

C

Sabrina Leone

Università Politecnica delle Marche, Italy

INTRODUCTION

Open learning and lifelong learning (LLL) are not new concepts in research literature (Brügelmann, 1975; Coffey, 1988; Kent, 1987). Anyhow, the current fervid debate is on the need for disruptive education at the light of the fundamental technological and societal changes that have occurred in recent decades (Stevens & Kirst, 2015). The whole life course is being reworked. Open, lifelong and ubiquitous learning, cloud computing and smart city frameworks are the pillars of the change that is replacing the traditional education and work models and transforming the way crowds of people learn, communicate, collaborate, teamwork, produce value and growth for the entities of which they are part (Leone & Biancofiore, 2015).

Education systems are urged to meet the growing demand for personalised learning paths, and individuals are driven to acquire awareness, tools and competences towards personal learning goals.

In particular, emerging practices with social semantic computing technologies and research findings signal the need for more personal, social and participatory approaches that foster learners in co-producing learning resources, rather in managing the learning process as a whole, and in pursuing personal life goals and needs. Indeed, learning occurs increasingly for the most part outside the traditional formal situations, especially for adult lifelong learners (European Commission/EACEA/ Eurydice, 2015).

In this regard, over the last years highlighting on the shift from formal to informal eLearning through knowledge management and sharing has been placed. Growing attention has been paid to

Personal Learning Environments (PLEs) as an effective framework for lifelong learners, and to the need of integration of formal and informal learning. Formal teaching spaces are defined within Learning Management Systems (LMSs). Spaces agreed in informal learning communities, instead, are to be used by social software. They are aimed to build networks of virtual identities and to define PLEs of dynamic contents, based on continuous accesses, validations, dialogic exchanges. ELearning 2.0 has mediated the shift from formal to informal eLearning, from LMSs, which are organisation-centred spaces that overlook individual differences and potential, to PLEs as emerging learner-centred spaces.

A wider and wider range of open source and free learning applications on the web are offering lifelong learners powerful tools to construct and characterise their own PLEs.

Technologically speaking, this change of perspective manifests in a learning web where information, activities and relations are distributed across sites and web-based learning applications, of which social networks have become a key factor. Consequently, knowledge management becomes an issue, and personalisation requires the support of adaptive mechanisms and of semantics applied to social components.

This work presents the updated version of *SSW4LL* (*Social Semantic Web for Lifelong Learners*), an adaptive, modular, flexible and integrated learning format which has been devised to support the characterisation of adult lifelong learners' PLEs by implicit and explicit tools of personalisation, in a learner-centred framework (Leone, 2013; 2014; 2015). The *SSW4LL* system,

the technological architecture, is presented as a whole made up of components of formal and informal learning environments: Moodle 2.9 integrated with an adaptive mechanism (conditional activities) and some tools of Social Semantic Web (Semantic MediaWiki 2.3, Diigo and Google+), respectively. The *SSW4LL* format was successfully validated during the course *SSW4LL 2011*.

BACKGROUND: PERSONALISATION IN A LIFELONG LEARNING VISION

Over time a great amount of theories and definitions of personalisation of learning have been produced. The expressions *personalised learning*, *personalising learning*, *individualised learning* have been coined to characterise and support with different emphasis the basic common view that each learner should be able to choose a tailored learning path, in order to meet personal needs, interests and abilities (Bentley & Miller, 2004); to promote both independence and dependability (Downes, 2007); to enhance social skills and sense of responsibility toward others (Pankhurst, 1922; Keller, 1968); to improve creative, intellectual, social and moral growth and develop personality fully (OECD, 2006; Rogers, 1983).

As the activator of economical, cultural and social growth of the knowledge society, the individual should be seen as an active, responsible and self-motivated learner, a co-author of the script that determines how education is delivered (Knowles, 1970; Leadbeater, 2004), often with extensive use of technology in the process; briefly, a lifelong learner (Leone, 2010; Martinez-Pons, 2002; Zimmerman, 2002).

The emphasis in relation to personalising education is that learning is inclusive, lifelong and therefore reaching beyond the traditional confines of schools (OECD, 2006).

Personalisation is thus a strategy aimed at designing, implementing and supporting learner-centred institutional practices (Maharey, 2007), and at drawing on wider resources for learning

beyond formal education, by creating new flexibilities to meet new demands (Bentley, 2005).

In this regard, Leadbeater's model of surface and deep personalisation (Leadbeater, 2004), where the student steadily progresses from consumer to producer behaviour, calls attention to a great deal of crossover with PLEs.

Since its appearance, the concept of PLE has posed numerous educational theories and implications (Chatti et al., 2010; Downes, 2010; Leo, Manganello & Chen, 2010; Motschnig-Pitrik & Mallich, 2004; Pettenati, 2010; Wild, Mödritscher & Sigurdarson, 2008). In this research, a PLE is a concept rather than specific software, a group of techniques and a variety of tools to gather information, explore and develop relationships between pieces of information (Leone & Guazzaroni, 2010). A PLE helps to view the subject as a landscape as well as individual pieces of information; to create a personal repository of materials and relationships clustered around a unifying topic or concept; to document, reflect, communicate, collaborate. Information and knowledge reside in digital and non-digital sources. A PLE, at the same time, develops and is fed by autonomy, pragmatics, relevance, building on prior knowledge, goal-directed approach (Leone, 2009).

Certainly, the concept of PLE has challenged the existing education systems and institutions, since new forms of learning are based on trying things and action, rather than on more abstract knowledge. Thus the dichotomy LMS vs PLE has been transformed into models of integration (Giovannella, 2008; Leo Manganello & Chen, 2010) of formal and informal learning.

Anyhow, in relation to personalisation of learning, LMSs, the formal component of the integrated learning environment, are weak (Graf, 2007), and the management of an effective PLE requires information overload and diversity, distribution, trustworthiness and evaluation of the collection of learning resources to be tackled as critical elements.

To this end, adaptive mechanisms and Social Semantic Web can offer suitable tools of

respectively implicit and explicit personalisation (Braynov, 2004) of learning. Implicit learner's profiling is automatically carried out by the LMS through an adaptive plug-in by tracking and monitoring user's behaviour in order to identify learning patterns. Explicit profiling, instead, involves the learner's active participation by Social Semantic Web tools, thereby allowing the user to express directly needs and modalities and thus to control and share his/her learning path.

Many pedagogical and psychological theories sustain that learners have diverse learning modes (Felder & Silverman, 1988; Gardner, 1983; Kolb, 1984) and that students showing a strong preference for a specific learning style could have difficulties if the teacher's approach is dissonant (Felder & Soloman, 1997). On this basis, the evaluation of the several models for the detection of learners' learning styles and of adaptive educational systems (Oppermann & Simm, 1994) that could be integrated in a LMS is crucial (Graf, 2007; Leone, 2013).

Some researchers (Graf, Kinshuk & Ives, 2010; Limongelli, Sciarrone & Vaste, 2011) have successfully implemented adaptive plug-ins in Moodle 1.9, where adaptivity operates on the basis of the detection of students' learning styles by the Felder-Silverman learning styles model (FSLSM) (Felder & Silverman, 1988) as the most acknowledged model in this kind of applications. Explorative studies have been conducted on the use of artificial neural networks to increase the accuracy level of learning style identification (Bernard, Chang, Popescu, & Graf, 2015). Anyhow, no similar research experience seems to have been carried out with the latest versions of Moodle to date; besides, none of the adaptive plug-ins that have been analysed for this study has been adapted for Moodle 2.9 yet, nor it seems that learning formats that exploit the Moodle 2.9 conditional activities as an adaptive mechanism have been devised.

Recently several research projects funded by the European Union (EU) have been carried out

to provide networked and personalised learning environments for lifelong learners, but a few have investigated new applications that exploit Social Semantic Web to enhance LLL, like *Learning Technologies for Lifelong Learning* (LTfLL) (www.ltfill-project.org) and *Responsive Open Learning Environments* (ROLE) (Kroop, Mikroyannidis & Wolpers, 2015).

THE SSW4LL FORMAT

The *SSW4LL* format offers a learner-centred framework to back the characterisation of adult lifelong learners' PLEs through implicit and explicit tools of personalisation. The format is devised for adult lifelong learners in general, rather than for one specific cluster within, and for the expansion of all knowledge domains. Besides, the *SSW4LL* format supports mobile learning, but ubiquitous learning features (Leone & Leo, 2011) could be implemented as well, as an extension.

The synergy of formal and informal learning arises from the smooth integration of the different technological components, the light e-moderation of the learning environment by a facilitator, the support of a technical e-tutor and the continuous enrichment of the initial learning resources (formal environment) by social software and Social Semantic Web tools (informal environment).

The *SSW4LL* system, the technological architecture, is made up of Moodle 2.9 integrated with an adaptive mechanism (conditional activities), as the formal learning environment component, and of Semantic MediaWiki 2.3, Diigo and Google+, as Social Semantic Web tools and informal learning environment elements.

As a whole, the *SSW4LL* format offers an adaptive (Oppermann & Simm, 1994) modular, flexible and integrated architecture, compatible with future Moodle releases.

The format has been successfully implemented and validated by the course *SSW4LL2011* (Leone, 2013; 2014). Further experiences are in progress.

Needs Analysis, Paradigm, and Learning Strategies

In reference to the European framework of key competences for LLL (European Parliament and Council of the European Union, 2006) and to the results of the EU survey on adult education (AES) (European Commission, 2011), the *SSW4LL* format can contribute to: (1) learning to learn; (2) form active citizens, that is individuals who are engaged in the development of the multiple dimensions of citizenship, beyond knowledge towards the enhancement of competences and attitudes by experiencing active participation in different contexts; (3) support and deliver personalised and flexible learning; (4) facilitate learner-centred technology-enhanced learning; (5) promote inclusion; (6) improve digital and social skills.

The *SSW4LL* format is created to personalise learning in terms of self-organisation by adult lifelong learners working with the support of the facilitator, the adaptive mechanism, the technical eTutor and the peers (Leadbeater, 2004). In this format, personalisation aims to valorise learners' full potential and to empower individuals through knowledge sharing and co-construction. Learners are active co-designers of the learning experience (Maharey 2007). As a result, *learning work* turns into a *learning adventure* (Leo, Manganello & Chen, 2010), that is a learner-centred, holistic experience which involves a complex, continual, chaotic and co-creative process.

The theoretical background of *SSW4LL* draws on andragogy (Knowles, 1970) and socio-constructivism (Vygotsky, 1986; Varisco, 2002) theories. Learning is negotiated between the facilitator, the learners and the learning material, and expertise can be placed beyond the formal setting through learners' PLEs. The *SSW4LL* format can be developed through the following strategies:

- Brainstorming (in Moodle forums, and in Google+ by posts and hangouts);
- Problem solving;

- Collaborative and cooperative learning (in Moodle forums, Diigo and Semantic MediaWiki);
- Webquest (which includes cooperative work and problem solving);
- Reflection activities;
- Learning by doing;
- Self-learning, with the support of an adaptive mechanism in the formal learning environment (Moodle).

Technological Architecture: The SSW4LL System

The *SSW4LL* system is made up of integrated components of formal and informal learning environments.

Specifically, the design of the framework has been developed considering (Giovannella, 2008):

- The web used as an environment, where various tools and contents can be aggregated for the construction of a PLE;
- Open source, open content, open society and, therefore, adoption of open “machine-readable” standards;
- The learner's capability of managing his/her learning processes and of setting up his/her e-portfolio as aggregator of personal knowledge and competences (Lubesky, 2006).

The formal learning environment consists in Moodle 2.9, in which adaptive learning is enabled by its conditional activities. The elements of the informal learning environment are Semantic MediaWiki 2.3, Diigo and Google+, that allow adult lifelong learners for a qualitative different bottom-up approach to learning

By fully exploiting Moodle 2.9 adaptation features, that in the *SSW4LL* system are based on the detection of learners' learning styles (Felder & Silverman, 1988; Felder & Soloman, 1997), this LMS can deploy a personalised scaffolded learning

environment for self-regulated learners. Moreover, social software can be smoothly integrated in the architecture by Diigo widgets, by the GoogleLogin extension that allows Semantic MediaWiki users to login with their Google account, and by Moodle plug-in for Google+.

Formal Learning Environment: Moodle 2.9

All LMSs provide a wide range of features that teachers can exploit to create and deliver online courses. For this reason, they are commonly used by educational institutions to successfully offer technology-enhanced learning. All the same, LMSs normally do not consider learners' individual differences and supply very little or, in most cases, no adaptivity (Graf, 2007).

Additionally, the knowledge society and the LLL vision have required a large variety of skills to be developed in lifelong learners: the ability to locate and evaluate information effectively and efficiently; facility with making meaning by aligning new information with prior knowledge; and an ability to synthesise, critically analyse, and create new information within the context of larger social practices (Lin, 2011).

On the basis of these preliminary remarks, two relevant studies (Graf, 2007; Lin, 2011) and the author's prior direct experience have informed the choice of Moodle as the most suitable component for the formal learning environment in the *SSW4LL* system.

Furthermore, Moodle allows the implementation of different degrees of learning personalisation between the following extremes: (a) teachers can opt to roughly guide students, by scaffolding, and let them find their own way; (b) teachers can opt to give them a detailed map with checkpoints they must reach along the way, that is a more structured approach. Further, the determination of learners' learning styles is possible by creating a suitable quiz.

In the *SSW4LL* format, the abovementioned option (a) and the FSLSM (Felder & Silver-

man, 1988) are adopted to meet learners' needs and profiles, in accordance with the theoretical background of the format. Furthermore, several researchers concur on the validity of the FSLSM as the most appropriate to be used in adaptive learning systems (Graf, Kinshuk & Ives, 2010; Limongelli, Sciarrone & Vaste, 2011), as the *SSW4LL* system is.

Conditional activities aim to allow the educational designer to scaffold learners in accessing learning resources on the basis of one or more conditions. In the back end of the resources and activities (e.g., a forum) that the designer previously wishes to add to the course a *restrict availability* area is available for setting (e.g., on a date, on a grade, on the completion of an activity, etc).

The second element to make the learning environment adaptive is the *activity completion tracking* facility, which is the ability for users to mark tasks as "done" (by a tick) or to have tasks checked automatically by the system if the teacher chooses this option in the back end.

In the *SSW4LL* format a suitable and thorough combination of these conditions allows to produce different learning sequences, in which learning objects are proposed according to the students' learning profiles that are initially detected through an entry quiz.

Informal Learning Environment: Semantic MediaWiki, Diigo, and Google+

In the *SSW4LL* system the elements of the informal learning environment are Semantic MediaWiki, Diigo and Google+.

They have been selected among several alternative solutions on the basis of effectiveness and efficiency in relation to *SSW4LL* target learners' goals. Evaluation has been carried out on the following criteria, attributing a value on a 5-point Likert scale: social semantic features, effectiveness as tools for characterising adult lifelong learners' PLEs, novel features, possible integration with Moodle 2.9, easy-to-use interface and mobile

learning features. The categories of tools that needed to be considered for the *SSW4LL* system were aggregators, tools of semantic annotation, social bookmarking and recommended search, and social networks.

Organisation

Technical Competences Required

The implementation and management of the *SSW4LL* format requires the following professional profiles:

- A learning technologist, who has extensive knowledge of the use of Moodle, Semantic MediaWiki, Diigo and Google+, and of all the other technologies that could support learning and teaching within the *SSW4LL* format. He/she deals with the implementation, updating and troubleshooting of the different technological components of the format;
- An ICT technician, who has knowledge of a range of ICT hardware and applications commonly used, and has good problem-solving and organisational skills. He/she provides the facilitator and learners with technical support, guidance and maintenance in order to use all software/hardware correctly during the learning path;
- A (or more) teacher-facilitator, who is familiar with the design, implementation and management of Moodle courses, including the settings of conditional activities;
- Learners, who have basic digital competences and use of Web 2.0 tools (forums, wikis, social bookmarking and social networks).

Devices

The implementation of the *SSW4LL* format in a distributed learning environment requires an Internet connection and one of the following sets of equipment, according to the students' location:

- A workstation, a webcam and a headphone set if the students are in a fixed location;
- Smartphones or tablets if the students are in a mobile learning environment;
- Laptops, smartphones, ultra-mobile PCs or tablets together with the use of sensor network nodes, contact-less smart cards, RFID (Radio Frequency Identification) and QR (Quick Response) codes, if the students are in a ubiquitous learning environment (as an extension).

Evaluation and Assessment

The *SSW4LL* format is monitored through an entry and a final survey, and entry and formative self-assessment tests to evaluate its effectiveness in terms of participants' expectations and satisfaction, and achievement of learning goals, respectively. A final cooperative essay or project work provides summative assessment.

Recommendations for an Optimal Implementation

The format *SSW4LL* is very flexible for both the teacher-facilitator and the learners. Anyhow, essential elements for the success of the experience are: (1) thorough organization and management of the necessary hardware and software; (2) an adequate familiarisation of the learners with the learning environment (technology, tools and learning approach); (3) a light e-moderation by the facilitator, in order to provide a modulation of self-regulated and shared learning on the basis of the students' silent and/or expressed requests, with the aim of supporting participants' high motivation.

SWOT Analysis

Strengths of the *SSW4LL* format are scaffolded self-regulated learning, personalised and flexible learning, novel tools for the characterisation of adult lifelong learners' PLEs. Weaknesses are the initial technological issues that can origin

demotivation and disengagement. Opportunities are various: growing availability of open software and learning materials; increasing individuals' awareness of the importance of adopting a LLL vision. Vice-versa, possible threats are a teacher-centred approach, that could void the aims and flexibility of the format, and an insufficient Internet connection.

FUTURE RESEARCH DIRECTIONS

This research could open ways for advanced learning systems, which are able to meet learners' needs and characteristics, merge assets of formal and informal learning environments, and provide learners with dynamic personalisation of their PLEs.

In the future, improvements of the *SSW4LL* system include enabling the integration of additional social semantic tools to tackle differently knowledge management, syndicating resources and trustworthiness, that are actual research issues related to the enhancement of dynamic PLEs. Moreover, in-depth observation will be conducted in higher education on how the learning outcomes improve by transferring responsibility for the choice and configuration of the learning environment from the teacher to the learner by social semantic tools.

CONCLUSION

At the light of the fundamental technological and societal changes that have occurred in recent decades, current policies and research debate are focused on the development of disruptive education solutions on the basis of open, lifelong and ubiquitous learning frameworks. This change of vision is manifesting in a growing awareness that more personal, social and participatory approaches to learning are necessary to meet the lifelong learners' need to become co-producers

of their learning paths and to control the whole learning process.

The shift from formal to informal learning, from LMSs as organisation-centred spaces to PLEs as learner-centred spaces, can be turned into an integration of these two in a flexible and integrated learning environment.

This perspective manifests in a learning web where information is distributed across sites. As a result, knowledge management becomes an issue, and personalisation requires the support of adaptation and of semantics applied to social components (i.e., the Social Semantic Web).

The *SSW4LL* learning format has been devised to support the characterisation of adult lifelong learners' PLEs. The format offers an adaptive, modular, flexible and integrated architecture; the *SSW4LL* system is made up of components of formal (Moodle 2.9 with conditional activities as a suitable mechanism of learning adaptation) and informal learning environments (Semantic MediaWiki 2.3, Diigo and Google+). The format is compatible with future Moodle releases and easy to use for teachers-facilitators.

The influence of the informal learning components of the *SSW4LL* system is strong: where social software gives users freedom to choose their own processes and supports the collaboration of adult lifelong learners *anytime, anywhere*, Semantic Web technology gives the possibility to structure information for easy retrieval, reuse, and exchange between different systems and tools.

The format is conceived to empower adult lifelong learners by facilitating the acquisition of some of the skills necessary for the 21st century.

REFERENCES

Bentley, T. (2005). *Personalised learning in the UK: from vision to strategy? Personalised learning: high expectations symposium*. Sydney, NSW: Department of Education and Training.

- Bentley, T., & Miller, R. (2004). *Personalised learning: creating the ingredients for system and society-wide change*. IARTV Occasional Paper No. 87. Melbourne: Incorporated Association of Registered Teachers of Victoria.
- Bernard, J., Chang, T. W., Popescu, E., & Graf, S. (2015). Using Artificial Neural Networks to Identify Learning Styles. *Proceedings of the International Conference on Artificial Intelligence in Education (AIED 2015)*, 541-544. doi:10.1007/978-3-319-19773-9_57
- Braynov, S. (2004). Personalization and Customization Technologies. *The Internet Encyclopedia*. Retrieved 4 Nov. 2015 from <http://onlinelibrary.wiley.com/doi/10.1002/047148296X.tie141/full>
- Brügelmann, H. (1975). Open curricula—A paradox? *Cambridge Journal of Education*, Vol. 1, No. 5. *Lent Term*, 1975, 12–20.
- Chatti, M. A., Agustiawan, M. R., Jarke, M., & Specht, M. (2010). Toward a Personal Learning Environment Framework. *International Journal of Virtual and Personal Learning Environments*, 1(4), 66–85. doi:10.4018/jvple.2010100105
- Coffey, J. (1988). Guest Editorial: The Opening Learning Movement. *Innovations in Education & Training International*, 25(3), 195–196. doi:10.1080/1355800880250301
- Downes, S. (2007). Learning Networks in Practice. In *Emerging technologies for learning*. Retrieved 4 Nov. 2015 from http://www.downes.ca/files/Learning_Networks_In_Practice.pdf
- European Commission/EACEA/Eurydice. (2015). *Adult Education and Training in Europe: Widening Access to Learning Opportunities*. Eurydice Report. Luxembourg: Publications Office of the European Union.
- European Parliament, Council of the European Union. (2006). Recommendation of the European Parliament and of the Council of 18 December 2006 on key competences for lifelong learning. *Official Journal of the European Union*, L394, 30.12.2006, p. 10-18.
- Felder, R. M., & Silverman, L. K. (1988). Learning and teaching styles in engineering education. *English Education*, 78, 674–681.
- Felder, R. M., & Soloman, B. A. (1997). *Index of Learning Styles Questionnaire*. Retrieved 4 Nov. 2015 from <http://www.engr.ncsu.edu/learning-styles/ilsweb.html>
- Giovannella, C. (2008). Learning 2.0? *Atti del V congresso Sie-I*. Macerata, Italia: EUM.
- Graf, S. (2007). *Adaptivity in Learning Management Systems focusing on Learning Styles* (PhD thesis). Vienna University of Technology.
- Graf, S., Kinshuk, & Ives, C. (2010). A flexible mechanism for providing adaptivity based on learning styles in Learning Management Systems. *Proceeding of the IEEE International Conference on Advanced Learning Technologies (ICALT 2010)*, (pp. 30-34). Sousse: IEEE Computer Society. doi:10.1109/ICALT.2010.16
- Keller, F. S. (1968). Good-bye, teacher..... *Journal of Applied Behavior Analysis*, 1(1), 79–89. doi:10.1901/jaba.1968.1-79 PMID:16795164
- Kent, J. (1987). *Principles of Open Learning*. Stafford, UK: Witan Books.
- Knowles, M. S. (1970). *The Modern Practice of Adult Education: Andragogy versus Pedagogy*. New York: Associated Press.
- Kolb, D. A. (1984). *Experiential Learning: experience as the source of learning and development*. Prentice-Hall.

- Kroop, S., Mikroyannidis, A., & Wolpers, M. (2015). *Responsive Open Learning Environments*. New York: Springer. doi:10.1007/978-3-319-02399-1
- Leadbeater, C. (2004). *Personalisation through participation: a new script for public services*. Demos.
- Leo, T., Manganello, F., & Chen, N.-S. (2010). From The Learning Work To The Learning Adventure. *Proceedings of EDEN 2010 Annual Conference*.
- Leone, S. (2009). PLE: A Brick in the Construction of a Lifelong Learning Society. In J. O'Donoghue (Ed.), *Technology Supported Environment for Personalised Learning: Methods and Case Studies*. IGI-Publisher.
- Leone, S. (2010). F2F learning vs eLearning: the lifelong learner's point of view. *Proceedings of INTED 2010*.
- Leone, S. (2013). *Characterisation of a Personal Learning Environment as a lifelong learning tool*. New York: Springer Science & Business Media. doi:10.1007/978-1-4614-6274-3
- Leone, S. (2014). Social Semantic Web for Lifelong Learners - SSW4LL: a learning format. In M. Khosrow-Pour (Ed.), *Encyclopaedia of Information Science and Technology* (3rd ed.). Hershey, PA: IGI Global.
- Leone, S. (2015). *Exploring the Social Semantic Web for Lifelong Learners*. Retrieved 4 Nov. 2015 from <http://www.igi-global.com/video.aspx?ref=exploring-social-semantic-web-lifelong&titleid=137628>
- Leone, S., & Biancofiore, G. (2015). Leveraging teamwork by Google+ in a lifelong learning perspective. *Knowledge Management & E-Learning: An International Journal*, 7(2).
- Leone, S., & Guazzaroni, G. (2010). Pedagogical sustainability of interoperable formal and informal learning environments. In F. Lazarinis, S. Green, & E. Pearson (Eds.), *Developing and utilizing e-learning applications*. IGI-Publisher. doi:10.4018/978-1-61692-791-2.ch001
- Leone, S., & Leo, T. (2011). The synergy of paper-based and digital material for ubiquitous foreign language learners. *Knowledge Management & E-Learning: An International Journal*, 3(2).
- Limongelli, C., Sciarrone, F., & Vaste, G. (2011). Personalized e-learning in Moodle: the Moodle_LS System. *Journal of e-Learning and Knowledge Society*, 7(1), pp. 49-58.
- Lin, T.-J. (2011). Review of Moodle 2.0. *Language Learning & Technology*, 15(2), 27-33.
- Lubeskey, R. (2006). The present and future of Personal Learning Environments (PLE). *Deliberations*. Retrieved 4 Nov. 2015 from <http://www.deliberations.com.au/2006/12/present-and-future-of-personal-learning.html>
- Maharey, S. (2007). *Organising for personalising learning*. Ministry of Education media release, 13 March. New Zealand Government.
- Martinez-Pons, M. (2002). A social cognitive view of parental influences on student academic self-regulation. *Theory into Practice*, 41(2), 126-131. doi:10.1207/s15430421tip4102_9
- Motschnig-Pitrik, R., & Mallich, K. (2004). Effects of Person-Centered Attitudes on Professional and Social Competence in a Blended Learning Paradigm. *Journal of Educational Technology & Society*, 7(4), 176-192.
- Oppermann, R., & Simm, H. (1994). Adaptability: user-initiated individualization. In R. Oppermann (Ed.), *Adaptive User Support: Ergonomic Design of Manually and Automatically Adaptable Software* (pp. 14-64). Hillsdale, NJ: Lawrence Erlbaum Associates.

Organisation for Economic Co-operation and Development. (2006). *Schooling for tomorrow: personalising education*. Paris: Centre for Educational Research and Innovation, OECD.

Parkhurst, H. (1922). *Education On The Dalton Plan*. New York: E. P. Dutton & Company.

Pettenati, M. C. (2010). Proposal of a Personal Knowledge Environment. *Knowbie*. Retrieved August 8, 2011 from <http://knowbie.wordpress.com/tag/personal-knowledge-environment/>

Rogers, C. R. (1983). *Freedom to Learn for the 80's*. Columbus, OH: Charles E. Merrill Publishing.

Stevens, M. L., & Kirst, M. W. (Eds.). (2015). *Remaking College: The Changing Ecology of Higher Education*. Palo Alto, CA: Stanford University Press.

Varisco, B. M. (2002). *Costruttivismo socio-culturale. Genesi filosofiche, sviluppi psicopedagogici, applicazioni didattiche*. Roma, Italia: Carocci.

Vygotsky, L. S. (1986). *Thought and Language*. Cambridge, MA: MIT Press.

Wild, F., Mödritscher, F. & Sigurdarson, S. (2008). Designing for Change: Mash-Up Personal Learning Environments. *eLearning Papers*, (9).

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into Practice*, 41(2), 64–72. doi:10.1207/s15430421tip4102_2

ADDITIONAL READING

Alberici, A., Catarsi, C., Colapietro, V., & Liodice, I. (2007). *Adulti e Università. Sfide ed innovazioni nella formazione universitaria e continua*. Milano, Italia: Franco Angeli.

Attwell, G. (2008). The Social Impact of Personal Learning Environments in Connected Minds. In S. Wheeler (Ed.), *Emerging Cultures: Cybercultures in Online Learning*. Charlotte, NC: Information Age Publishing.

Barr, R., & Tagg, J. (1995). From Teaching to Learning: A New Paradigm for Undergraduate Education. *Change Magazine*, 2(12), 8–12.

Berners-Lee, T., Handler, J., & Lassila, O. (2001). The Semantic Web. *Scientific American*. Retrieved 4 Nov. 2015 http://www-sop.inria.fr/acacia/cours/essi2006/Scientific%20American_%20Feature%20Article_%20The%20Semantic%20Web_%20May%202001.pdf

Bratsas, C., Kapsas, G., Konstantinidis, S., Koutsouridis, G., & Bamidis, P. D. (2009). A semantic wiki within Moodle for Greek medical education. *Proceedings of the 22nd IEEE International Symposium on Computer-Based Medical Systems*, pp.1-6. IEEE Computer Society Press. doi:10.1109/CBMS.2009.5255417

Breslin, J., Passant, A., & Decker, S. (2009). *The Social Semantic Web*. Springer-Verlag Berlin Heidelberg. doi:10.1007/978-3-642-01172-6

Brusilovsky, P., Kobsa, A., & Nejdl, W. (Eds.). (2007). *The Adaptive Web: Methods and Strategies of Web Personalization*. Berlin, Heidelberg: LNCS Springer. doi:10.1007/978-3-540-72079-9

Brusilovsky, P., & Peylo, C. (2003). Adaptive and Intelligent Web-Based Educational Systems. *International Journal of Artificial Intelligence in Education*, 13, 156–169.

Cedefop (2011). *Lifelong guidance across Europe: reviewing policy progress and future prospects*. Luxembourg: Publications Office of the European Union.

- Cormier, D. (2010). Five points about PLEs and PLNs. *Dave Cormier* (Blog post). Retrieved 4 Nov. 2015 from <http://davecormier.com/edblog/2010/09/12/5-points-about-ples-plns-for-plenk10/>
- EAEA. (2015). *European Agenda for Adult Learning*. Retrieved 4 Nov. 2015 from <http://www.eaea.org/en/policy-advocacy/european-agenda-for-adult-learning.html>
- EDUCAUSE. (2009). 7 Things you should know about PLE's. Retrieved 4 Nov. 2015 from <http://net.educause.edu/ir/library/pdf/ELI7049.pdf>
- European Commission. (2000). *E-learning – Designing Tomorrow's Education*. Brussels: European Commission.
- European Commission. (2001). *Making a European area of lifelong learning a reality*. Brussels: European Commission.
- European Commission. (2002). *European Report on Quality Indicators of Lifelong Learning*. Brussels: European Commission.
- European Commission. (2010). *Key competences for adult learning professionals*. Brussels: European Commission.
- Graf, S., & Kinshuk (2014). Adaptive technologies. In Spector, J.M. Merrill, M.D. Elen, J. Bishop, M.J. (Eds.), *Handbook of Research on Educational Communications and Technology*, pages 771–779.
- Gruber, T. (2008). Collective knowledge intelligence: Where the Social Web meets the Semantic Web. [Amsterdam, The Netherlands: Elsevier Science Publishers B. V.]. *Web Semantics: Science, Services, and Agents on the World Wide Web*, 6(1), 4–13. doi:10.1016/j.websem.2007.11.011
- Herzig, D. M., & Ell, B. (2010). Semantic MediaWiki in Operation: Experiences with Building a Semantic Portal. *Proceedings of the 9th International Semantic Web Conference (ISWC-10)*, p.114-128. Springer-Verlag Berlin, Heidelberg. doi:10.1007/978-3-642-17749-1_8
- Leadbeater, C. (2009). *We-Think: Mass Innovation, Not Mass Production*. London, UK: Profile Books.
- Leone, S. (2008). Lifelong learning and tertiary education in Italy. *The International Journal of Interdisciplinary Social Sciences*, 3(2), 7–14. doi:10.18848/1833-1882/CGP/v03i02/52529
- Leone, S. (Ed.). (2014). *Synergic integration of formal and informal eLearning environments*. Hershey, PA: IGI Global.
- Leone, S., Guazzaroni, G., Carletti, L., & Leo, T. (2010). The increasing need of validation of non-formal and informal learning. The case study of the community of practice Webm.org. *Proceedings of CELDA 2010, Timisoara (Romania), 15-17 October 2010*.
- Posea, V., & Trausan-Matu, S. (2010). Bringing the Social Semantic Web to the Personal Learning Environment. *Proceedings of the 10th IEEE International Conference on Advanced Learning Technologies (ICALT)*. doi:10.1109/ICALT.2010.48
- Trentin, G. (2005). From “formal” to “informal” e-Learning through knowledge management and sharing. *Journal of e-Learning and Knowledge Society*, 1(2), 209-217.
- UNESCO Institute for Education. (1999). *Glossary of Adult Learning in Europe*. Paris, France: UNESCO.

KEY TERMS AND DEFINITIONS

Adaptivity: In adaptive learning systems, it provides increased user's efficiency, effectiveness and satisfaction by greater correspondence between learner's needs and goals and the characteristics of the system.

Adult Education: All forms of non-vocational adult learning, whether of a formal, non-formal or informal nature.

Formal Learning: Hierarchically structured, chronologically graded educational system running from primary through to tertiary institutions.

Informal Learning: Unstructured learning that allows persons to acquire attitudes, values, skills and knowledge from daily experience, within the individual's environment (i.e., family, friends, peer groups, etc.).

Learning Management System: Organisation-centred learning space where online interaction takes place between learners and teachers.

Lifelong Learners: Self-regulated learners, characterized as demonstrating perseverance, initiative, and adaptive abilities. Self-regulation relates to an ability to recognize a need for further learning as well as to be proactive in gaining access to and accomplishing learning.

Lifelong Learning: A holistic vision of learning in different contexts (formal, non-formal and informal) and throughout life, based on the evolution of provider-driven education toward

personalised learning and aiming at improving knowledge, skills and competencies within a personal, civic, social and/or employment-related outlook.

Personal Learning Environment: A concept, a group of techniques and a variety of tools to (1) gather information, explore and develop relationships between pieces of information; (2) facilitate the access to and the aggregation, the configuration and the management of the individual's learning experiences. A PLE is an activity-based learning environment, user-managed and learner-centred; it is an open system, interconnected with other PLEs and with other external services.

Social Semantic Web: As a result of the merging of Social Web, the web of social relations and collective intelligence, and the Semantic web, the web of data, it consists in explicit and semantically rich knowledge representations created by social interactions on the Web.

Reflection as a Process From Theory to Practice



Sonia Bharwani

Indian School of Management and Entrepreneurship, India

Durgamohan Musunuri

Bhavan's Usha and Lakshmi Mittal Institute of Management, India

INTRODUCTION

Global boundaries, markets, customer demands, technologies, products, processes – all seem to be in a constant state of flux in today's dynamic business environment. There is considerable debate about the efficacy of the pedagogy used in developing and training management students to take on the mantle of being effective managers in this challenging and ever-changing business milieu (Boyatzis, Cowen & Kolb, 1995; Mintzberg, 2004; Raelin, 2009). Several researchers have critiqued that management education has become chain-bound by narrow vocationalism where functional and technical competencies are overemphasized at the cost of experiential learning which is an integral ingredient for the holistic development of a business manager (Reed & Anthony, 1992; Albert & Grzeda, 2015). It has been observed that management graduates often display analytical detachment and methodological elegance to the detriment of insight that comes from hands-on experience (Hayes & Abernathy, 1980; Inamdar & Roldan, 2013).

Researchers in the field of pedagogic practices like Hill (2003) and Raelin (2009) argued that learning by doing i.e. experiential learning where “concrete experience is the basis for observation and reflection” (Kolb, 1984, p.21) is really what managers need, to be effectual in today's dynamic business arena. There is a growing clamour for intersection of theory and practice in the field of management education and a shift in focus from

a teacher-centric approach to a student-centric, learning-focused approach. It has become imperative to address this issue by engaging in a deeper approach to learning. This can be done by laying emphasis on higher order thinking skills such as analysis, synthesis and evaluation which allow students to achieve enhanced levels of cognitive processing through transformative learning.

Thus, students of business management need to hone their cognitive abilities, learn from experiences and develop skills to question their mental models and underlying assumptions. This would enable them to develop new models and theories which are more relevant and *au courant* in terms of their current practice and thus, undergo transformative learning. Transformative learning is a deep-rooted, structural shift in basic foundations of thought, feelings, and actions (Mezirow, 1998). The pedagogical concepts of double loop learning, experiential learning and metacognition are important building blocks for attaining these transformative learning goals.

Double loop learning involves correction of errors by questioning the framework of learning systems and accordingly changing the governing values which underlie actions (Argyris, 2002). Experiential learning is based on the notion that experiences contribute to understanding and learning (Kolb, 1984) and metacognition (Flavell, 1979) involves thinking about one's own self and the attendant assumptions and mental images that underlie the thinking. A key aspect of transformative learning involves engaging in critical

reflection of one's experiences and the underlying assumptions which leads to a change in one's meaning schemes and perspectives (Mezirow, 1998). Thus, at the core of all of the above mentioned pedagogic concepts is the practice of reflection.

BACKGROUND

The Significance of Reflection

According Dewey (1933), an early pioneer of this precept, reflection can be defined as "active, persistent and careful consideration of any belief or supposed form of knowledge in the light of the grounds that support it and the further conclusion to which it tends" (p.9). Reflection as a process is meant to facilitate self-awareness in context of practice and encourage students to think critically. Further, reflection facilitates self-actualization and encourages the development of new knowledge and theory. In a dynamic environment where behaviour has to change on a continuous basis to adapt to the challenges of the situation at hand, reflection is "fodder for self-regulation" (Scott, 2010, p.432).

Daudelin (1996) defined reflection as "the process of stepping back from an experience to ponder, carefully and persistently, its meaning to the self through the development of inferences." Reflection is directly connected to learning since it aids students in drawing inferences from events and experiences which form the basis of future actions. Page & Meerabeau (2000) elaborated on this idea when they advocated for reflection because "various taken-for-granted elements of every day practice can be unearthed, made visible and considered for professional scrutiny."

Reflection, as such, refers to the practice of examining and analysing an experience, in the form of thought, feeling, or action, and ascribing meaning to it to glean its cognitive essence while it is happening or subsequently. Mezirow (1998) considered critical reflection as key instrument

by which one could question the validity of one's world views. He suggested three types of reflection based on their roles in transforming meaning schemes and perspectives: content reflection, process reflection, and premise reflection. The role of reflection in the students' learning cycle in transitioning from theory to practice and back to theory cannot be over-emphasized. Several researchers and academics have through their expositions clarified the meaning and utility of reflection in going beyond the surface approach to learning. Equally interesting is how practitioners i.e. teachers and teacher educators use the process of reflection to foster deep learning or life-long learning and improve teaching (Fry, Ketteridge & Marshall, 2009).

OBJECTIVES OF THE STUDY

Over the years, the education system in India has been following the surface approach to learning (Manikutty, 2010) which is typified by lecture-based classroom teaching, rote learning and an intention to merely complete the task at hand, with limited emphasis on critical thinking, active learning and collaborative learning. Cross-cultural studies comparing learning styles of students across countries show that in India, reflective practices and structured reflection are not widely prevalent concepts in mainstream professional education (Jaju, Kwak and Zinkhan, 2002) which could be attributed to the country's collectivist culture with a high degree of power distance and authority acceptance (Manikutty, Anuradha & Hansen, 2007). Hence, there is almost no research about the use and the efficacy of reflection as a learning tool in the context of Indian management education. This research paper endeavours to bridge this research gap by mapping the Indian management students' perception of reflection and the value they place on reflection as an instrument of learning and continuous development in their professional and personal pursuits.

FOCUS GROUP STUDY

One way of gaining powerful research insights within social science fields such as education and healthcare are focus group interviews which are usually grounded in a qualitative paradigm. Focus groups are “designed to obtain perceptions on a defined area of interest in a permissive, non-threatening environment” (Krueger, 1994, p.6) by encouraging open conversation and debate among the research participants to gather an in-depth understanding of their reality in terms of their knowledge and experiences.

This methodology provides an emic perspective or an insider’s view and uses inductive processes to analyse and interpret participants’ opinions and views. Focus group interviews were chosen over individual interviews keeping in mind that “the dynamic interaction afforded by focus groups enables the eliciting of a diversity of views from participants and the immediate clarification of issues that affect this diversity” (Hollis, Openshaw & Goble, 2002, p.4). The study, thus, used a qualitative approach by conducting focus group interviews to present the student participants with a forum to air opinions. It aimed at understanding students’ perceptions and interpretations of use reflection and reflective practices as a learning pedagogy without subjecting them to the pressure of arriving at a consensus.

Participants

Participants for the focus group study, which was conducted in May 2013, were chosen from among the undergraduate students of a hospitality and business management institute situated in Gurgaon, India. The institute had been conducting the undergraduate degree programmes of a well-established, U.K.-based university in India for more than a decade. The course curriculum of the university actively incorporated and used reflection and reflective practices in the pedagogy adopted.

The selection of the research participants was made keeping in mind the students’ exposure to and familiarity with the concepts and the practice of reflection as part of their course curriculum. Three focus group interviews were conducted in the same week. It was decided to restrict the number of participants in each focus group to a maximum of 7 students. Three different cohorts were chosen from each of the two undergraduate programmes, B.Sc. in Hospitality Management and B.B.A. (Hons.) in Business Management programme for the three focus group interviews. Focus Groups A and B comprised of students in the Final Years of the Hospitality Management and Business Management programmes, respectively. Focus Group C was selected from the Second Year students of the Business Management programme.

For Focus Group A, seven students were selected using convenience sampling from volunteers from the Final Year Cohort of the Hospitality Management programme. The selection was done on the basis of computer-generated random numbers. The group comprised of 3 female students and 4 male students. In a similar manner for Focus Group B, six students were selected from the Final Year Cohort of the Business Management programme and another six students were selected from the Year 2 Cohort of the Business Management programme for Focus Group C. Both the latter groups comprised of 4 female students and 2 male students. These ratios of male and female participants were representative of the gender profiles of the respective cohorts.

Procedure

The focus group studies were conducted within broad parameters outlined by Krueger (1994) with the help of question routes which were developed using protocols and strategies enunciated by him. Each focus group was of an average duration of 45 minutes. A series of introductory, transitional, key, probing and closing questions, as outlined in Table 1 in Appendix 1, were framed and phrased

in a conversational format to gather richer details about the participants' experiences with reflective practices. Dichotomous questions and questions which required students to make non-contextual judgments were avoided (Hollis et al., 2002).

At the outset of the focus group, the participants were apprised of the purpose of the interaction which was to understand students' perceptions and viewpoints regarding reflection and reflective practices. Basic rules for the interaction were defined and it was emphasized that there was no pressure for consensus. For each of the focus group session one of the researchers (SB) played the role of the moderator to facilitate group interaction. The second researcher (DM) played the role of an independent observer and took field notes of significant comments while observing and analysing the group dynamics. The entire proceedings of each of the focus groups were video-recorded and then transcribed. Immediately after each session, the moderator and observer conducted a debriefing session in order to facilitate later analysis.

DATA ANALYSIS AND FINDINGS

The transcripts of the video recordings were minutely examined and coded to align common concepts and significant themes together. This facilitated thematic analysis of the findings from the three focus group interviews. The observer's notes were used concurrently with the transcripts to identify the emerging categories in the theme of reflection from the proceedings of the focus groups. These were as follows:

- Perception of Reflection
- Practice of Reflection
- Experience of Reflection
- Benefits of Reflection
- Factors Enhancing and Inhibiting Reflection

Perception of Reflection

Students across all three cohorts viewed reflection as a tool for self-analysis and evaluation, learning from their experiences and course-correcting in the future in case of deviant outcomes. Words such as 'introspection', 'contemplation', 'flashback', 'self-criticism', 'self-evaluation', 'self-analysis', 'self-monitoring' and 'questioning one's assumptions' were used by the students while enunciating the meaning of reflection. A common theme which emerged was that reflection could be perceived as a process of revisiting one's own experiences to understand them better and 'learn from one's own mistakes and those of others' so that future experiences and decision can be informed. Another important element discussed in Groups A and B which comprised of Final Year students from both disciplines, included making connections between existing knowledge and newly acquired knowledge to integrate and enhance understanding for better guiding future action.

There was high degree of congruence between the students' perception of reflection and the definitions and descriptions of reflection found in academic literature. According to Rodgers (2002) reflection is a 'meaning-making' process. It is a practice which helps convert raw experience into intelligible theory grounded in concrete experience and supported by existing theoretical concepts. It forms the foundation of intelligent actions which are deliberated upon (rather than impulsive). Intelligent actions find their origins in information garnered from experience and are driven by one's final goals. But undergoing an 'experience' is not the only sufficient condition for reflection since experience is not primarily cognitive (Dewey, 1944). What is equally important is thoughtful consideration of the experience to accord meaning to it and give it value. Thus, it is essential to develop "the ability to perceive and then weave meaning among the threads of experience" (Rodgers, 2002, p.847- 848).

Practice of Reflection

In discussing the practice of reflection, students identified different ways in which they indulged in reflective practices. In the academic context writing reflective statements, maintaining reflective diaries and journals and keeping portfolios were identified as the most commonly used methods of structured reflection. They adeptly identified modules in their curriculum where reflection was an actively used pedagogy. Besides writing reflective expositions, ‘thinking’, ‘contemplating’, ‘voicing one’s thought’, and ‘sharing one’s views’ were also described as ways of reflecting used by students in the academic context.

Students across all three focus groups used the terms ‘mulling over’, ‘ruminating’ and ‘musing’ while describing reflecting in a personal, non-academic domain. These terms were used to describe reflection in isolation of others, in an unstructured manner when students were simply ‘revisiting’ their personal experiences for the purpose of introspection and self-analysis. However, some students of Focus Group A stated that while they did reflect in isolation, they believed that ‘sharing reflections’ with friends and mentors and using others as ‘sounding boards’ may enhance the quality of the ‘take-aways’ or learnings from the reflection. Another interesting revelation by this group was the increasing use of social media like Facebook (status updates) and Twitter as well as personal blogs for reflection by voicing their thoughts and airing their opinions on-line and also revisiting their ‘older posts’ to analyse their earlier thought processes.

Le Cornu’s (2009) model of reflection which advocated that individuals reconstruct and reflect on assimilated knowledge by thinking, speaking and writing, corroborates with the students’ views. While thinking about an experience is the first step in reflection, by itself it may not prove to be adequately sufficient for the task. External articulation in the form of speech or writing may be required “in order for individuals to consolidate and concretize their thinking” (Le Cornu, 2009,

p.282). Further, Rodgers (2002) also contended that reflection is a collaborative process and needs to happen in interaction with others. Reflection can be done in solitude, but reflection in community facilitates a broader understanding of an experience in the learner. Use of social networking sites (SNS) for reflective learning enhances student engagement by providing them with a platform for individual expression as well as collaborative connection (Park & Son, 2011). The findings from the focus group studies supported these premises as discussed above.

Experiences with Reflection

While discussing their experiences with reflection Focus Group A described their initiation to structured reflection in the course of their Hotel Operations module which required them to train in practical skills for restaurant, kitchen and front office operations. The students were required to maintain a reflective diary about their practical experiences on the shop floor and analyse their performance on a regular basis. They discussed the importance of maintaining a written record as an aid to ‘reflecting-on-action’ (thinking retrospectively on practice) since they had limited experience and skills in ‘reflecting-in-practice’ (thinking that shapes practice during practice) (Schön, 1987).

Both Focus Group A and B were appreciative of the role of the reflective portfolio that they were required to maintain during their year-long industry internship, in enhancing the learning take-aways from their practical experiences. Focus Group C, being academically junior, had still not been exposed to subject modules which used reflective practices to the extent that their seniors had and hence were unable to value-add to the discussion. All the focus groups concurred that for the uninitiated, maintaining a reflective log can be ‘a tedious activity’ and ‘a chore.’ However, they conceded that the process becomes less tiresome and even a positively viewed activity with increased familiarity and repetition.

In response to the query of whether the students had the opportunity to reflect in a structured manner or indulged in practices which could be construed as reflection prior to undertaking their undergraduate studies, the majority of the students across the focus groups answered in the negative. A couple of exceptions were students who had done their schooling in Australia and Canada and thus had engaged in reflective practices even at the school level. On probing further, some of the students who had done their schooling within India recalled instances where they wrote a 'summer vacation diary' or had a teacher-led in-class reflective discussion or were required to write essays which had an element of self-analysis and meaning-making.

The students' perception regarding the practice of reflection as discussed above correlates with Rodgers' (2002) contention that reflection is a distinct, rigorous and disciplined way of thinking. According to her, reflection is more than haphazard 'mulling over' something. As a process it finds its roots in scientific inquiry and is made up of structured and precise steps which include observation and description of experience, analysis of experience in terms of generation of explanations and development of theories, followed by intelligent action and active experimentation.

Benefits of Reflection

In process of discussing why students reflect, the students' perceptions of the benefits of reflection were identified. 'Self-awareness', 'knowing one's strength and weaknesses', 'betterment through experience', 'mental clarity', 'reassurance from peers' and 'validation of one's thought processes and actions by others' were some of the key themes identified.

According to the focus group findings, the meta-knowledge gained through the process of reflection made the students more confident and self-aware and contributed to their emotional intelligence quotient. It also brought about a 'change in attitudes' and inculcated 'independent thinking'. To quote a participant from Focus Group A, 'I

understood the mistakes I did, wrote them down to see how I could rectify and work around them and be a better leader. The process enhanced my learning curve. You don't make the same fundamental mistakes the next time over.' Students undertook unstructured reflection to analyse and understand their experiences especially 'when things had gone wrong'. Thus, the role of reflection in helping a learner move from the state of disequilibrium or disturbed perplexity to a harmonious state of understanding was highlighted during the focus group discussions. Literature also highlights the important outcome of this process of exploration and internal scrutiny as transformative learning through changing of one's perspective as new information and experiences are encountered (Peltier, Hay & Drago, 2005).

Factors Enhancing and Inhibiting Reflection

The focus group discussions also focused on the factors that facilitated and enhanced reflective processes and practices among the participants and inhibiting factors for reflection. The students identified collaborative environment, personal disposition and time as the key factors that bolstered reflective practices. 'You reflect when you want to do better in life' and 'the college environment, faculty members, family and friends encourage one to reflect' were some of the opinions voiced by Focus Group A. Their contention is supported by Bulpitt & Martin (2005) who summed up the conditions necessary for reflection as "support, time and space and a collaborative environment" (p.208). Further, Le Cornu (2009), too, highlighted the role of a "respectful, receptive ear" in encouraging dialogue and fostering reflection.

While identifying the constraining factors, Focus Group C was vocal in enunciating that 'you need to courage to do it' since deep critical self-reflection can be uncomfortable process requiring the reflector 'face the honest truth'. Lack of time and opportunities for structured reflection along with poorly developed reflective skills were also

identified as impediments to reflection by Focus Group B. The students' perception in this context were on similar lines with Boud & Walker's (1993) research which has identified internal barriers in terms of preparedness and intent and external barriers such as support systems and inadequate preparedness which impact upon the ability of the learners to engage in reflective practices.

Table 2 in Appendix 2 summarises the findings of the focus group study based on the various themes related to reflection.

DISCUSSION

The aim of the study is to throw light on the practice of reflection as understood by students pursuing mainstream professional education courses in India. Reflection brings unconscious and implicit thinking and understanding to an explicit level. It refers to the cognitive ability to examine the dissonance of personal experiences in the context of one's prior thoughts and assumptions. It is a meaning-making process which entails a quest for alternative explanations and interpretations in order to reconcile ambiguity and inconsistency between one's experiences, actions and assumptions, and thus, fosters transformative learning.

Research showed that the students from the final year cohorts of both the curriculum streams who had more experience in practicing reflection had a better grasp of the concept and practice of reflection and were more comfortable in engaging with reflection for their personal as well as professional growth. Focus Group C which was selected from the Year 2 cohort which had comparatively lower exposure to structured reflective practices had limited understanding and found reflection to be a 'difficult' and 'not so comfortable' practice.

Another important finding which emerged was the students perceived that reflection was more a collaborative process rather than an independent, individual process. Rodgers (2002), too, advocated the need for and the benefits of collaborative reflection as under:

- It leads to the affirmation of the value of one's own experience which otherwise may have been summarily dismissed as 'unimportant'.
- It helps in seeing matters in a new light when others offer their perspective and help in broadening the realm of understanding.
- It supports the process of inquiry (which might be difficult to sustain alone) by making the learner responsible and accountable to a group.

This highlights the importance of role of the teacher in encouraging and developing the students' reflective skills and supporting them by providing a collaborative environment to practice reflection. In addition, there is an increasing use of social media such Facebook and Twitter as well as personal online blogs by students to externally articulate reflections.

The focus groups perceived that personal dispositional factors played an important role in encouraging reflective behaviour. Reflection requires an attitude that values one's own personal and intellectual growth and that of others. Along with cognitive discipline, reflection also requires emotional discipline on part of the individual to objectively accept and analyse what evidence indicates. To gain the maximum from reflective practices in terms of personal and intellectual growth, reflection should be ideally guided by a basic attitude of single-mindedness to engage in the experience in an undistracted manner. Further, open-mindedness is also essential to entertain many interpretations of the experience. This allows one to gain a broad and varied understanding of one's response and the actions that flow from it, rather than cling doggedly to preconceived notions. Lastly, one must be grounded in reality and question the real-life implications of the understanding gathered through reflection. This requires acknowledging that a shift in understanding of an experience may result in a paradigm shift in outlook.

LIMITATIONS AND FUTURE RESEARCH DIRECTIONS

The sample for the research was a non-probability one, since the participants for the focus groups were chosen from among volunteers and volunteers by their very nature preclude randomness. Since both the researchers were familiar with the participants by virtue of being faculty members at the institute, the focus groups were formed by selecting from among the volunteers on the basis of randomly generated computer-assigned numbers to eliminate bias in selection.

Another limitation that stems from the sample relates to its size which was limited to a maximum of 7 participants in each focus group. While the sample size is small in absolute terms, in relation to the average cohort size of 30 students at the institute it was considered a fair representative of the internal population who practice reflection. However, the study being a focus group study, by its very nature limits external validity and the generalisability of its results to a larger population. However, in context of the research objectives, the sample size was considered to be adequate for the purpose of the research which was preliminary in nature. Further, in order to maintain consistency, similar questioning routes were used in all focus groups with the same researchers playing the roles of moderator and observer respectively.

The preliminary findings discussed above contribute to understanding the Indian students' perspective on the utility of reflection in going beyond the surface approach to learning and transcending from theory to practice. However, further quantitative research which would help generalise these preliminary findings to a broader population and make a case for introducing and actively using this learning tool in management education in the Indian context, is required.

CONCLUSION

If we recognize and accept that to be effectual in today's dynamic business environment students

will have to deal with complex problem solving on a regular basis, it is then critical that they develop skills to engage in reflective practice. Reflection will enable student practitioners to distil meaning from their experiences and apply the essence of the learning to the establishment or evolution of processes in the complex organisational context, thus enabling them to synergise theory and practice.

Since reflective thinking is not necessarily an innate trait, the teacher or educational instructor plays a key role in creating a learning environment in which reflective practices are facilitated. Such an environment encourages open discussion, allows for questioning of established paradigms and voicing of divergent views and enables learning to transcend beyond the surface approach. Thus, for reflective learning to flourish, the teacher should be regarded by the learners as a mentor and a coach who promotes learning by engendering a non-threatening learning environment by encouraging questions and open discussions. Clearly, the role of the teacher in nurturing reflection and reflective practices among students cannot be undermined.

A deeper understanding of the meaning, utility and the criteria for inculcating reflective practices is one of the key ingredients for academic success and plays a vital role in elevating the educational experience of learners. Using active pedagogy which allows for observing and reflecting on experiences creates learning that involves "challenging assumptions, seeking alternatives, identifying areas for improvement... (and) shows active and conscious engagement, characteristics commonly associated with a deep approach to learning." (Peltier et al., 2005).

To sum up, it can be held that reflection plays a crucial role in creating meaning out of experiences in an iterative forward-moving crescendo of learning, alternating between practice and theory. Further, the criticality of the role of an instructor in fostering high order reflective and transformative learning cannot be over-emphasised.

REFERENCES

- Albert, S., & Grzeda, M. (2015). Reflection in Strategic Management Education. *Journal of Management Education*, 39(5), 650–669. doi:10.1177/1052562914564872
- Argyris, C. (2002). Double-loop learning, teaching, and research. *Academy of Management Learning & Education*, 1(2), 206–218. doi:10.5465/AMLE.2002.8509400
- Boud, D., & Walker, D. (1993). Barriers to reflection on experience. In D. Boud, R. Cohen, & D. Walker (Eds.), *Using experience for learning* (pp. 73–86). Buckingham, UK: Open University Press.
- Boyatzis, R., Cowen, S., & Kolb, D. (1995). *Innovation in professional education: Steps on a journey from teaching to learning*. San Francisco, CA: Jossey-Bass.
- Bulpitt, H., & Martin, P. J. (2005). Learning about reflection from the student. *Active Learning in Higher Education*, 6(3), 207–217. doi:10.1177/1469787405057751
- Daudelin, M. (1996). Learning from experience through reflection. In J. Osland, D. Kolb, & I. Rubin (Eds.), *The organizational behavior reader* (pp. 67–74). Upper Saddle River, NJ: Prentice Hall.
- Dewey, J. (1933). *How we think: A restatement of the relation of reflective thinking to the education process*. Boston, MA: D.C. Heath.
- Dewey, J. (1944). *Democracy and education*. New York, NY: Free Press.
- Flavell, J. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *The American Psychologist*, 34(10), 906–911. doi:10.1037/0003-066X.34.10.906
- Fry, H., Ketteridge, S., & Marshall, S. (2009). *A handbook for teaching and learning in higher education: Enhancing academic practice* (3rd ed.). New York, NY: Routledge.
- Hayes, R., & Abernathy, W. (1980). Managing our way to economic decline. *Harvard Business Review*, 58(4), 67–77.
- Hill, L. (2003). *Becoming a manager: How new managers master the challenges of leadership*. Boston, MA: Harvard Business School Press.
- Hollis, V., Openshaw, S., & Goble, R. (2002). Conducting focus groups: Purpose and practicalities. *British Journal of Occupational Therapy*, 65(1), 2–8. doi:10.1177/030802260206500102
- Inamdar, S., & Roldan, M. (2013). The MBA capstone course: Building theoretical, practical, applied, and reflective skills. *Journal of Management Education*, 37(6), 747–770. doi:10.1177/1052562912474895
- Jaju, A., Kwak, H., & Zinkhan, G. (2002). Learning styles of undergraduate business students: A cross cultural comparison between US, India and Korea. *Marketing Education Review*, 12(2), 49–60. doi:10.1080/10528008.2002.11488787
- Kolb, D. (1984). *Experiential learning as sources of learning and development*. Englewood Cliffs, NJ: Prentice Hall.
- Krueger, R. (1994). *Focus Groups: A Practical Guide for Applied Research* (2nd ed.). Thousand Oaks, CA: Sage Publication.
- Le Cornu, A. (2009). Meaning, internalization and externalization: Toward a fuller understanding of the process of reflection and its role in the construction of self. *Adult Education Quarterly*, 59(4), 279–297. doi:10.1177/0741713609331478
- Manikutty, S. (2010). The Effects of Nationality on the Approaches to Learning and Studying. In *Cross Cultural Approaches to Learning and Studying*. New Delhi: Macmillan.
- Manikutty, S., Anuradha, N., & Hansen, K. (2007). Does culture influence learning styles in higher education? *International Journal of Learning and Change*, 2(1), 70–87. doi:10.1504/IJLC.2007.014896

Mezirow, J. (1998). On critical reflection. *Adult Learning Quarterly*, 48(3), 185–198. doi:10.1177/074171369804800305

Mintzberg, H. (2004). *Managers not MBAs: A hard look at the soft practice of managing and management development*. San Francisco, CA: Barrett-Koehler.

Page, S., & Meerabeau, L. (2000). Achieving change through reflective practice: Closing the loop. *Nurse Education Today*, 20(5), 365–372. doi:10.1054/nedt.2000.0430 PMID:10895118

Park, J., & Son, J. (2011). Expression and connection: The integration of reflective learning into social network sites. *Journal of Online Learning and Teaching*, 7(1), 171–178.

Peltier, J., Hay, A., & Drago, W. (2005). The reflective learning continuum: Reflecting on reflection. *Journal of Marketing Education*, 27(3), 250–263. doi:10.1177/0273475305279657

Raelin, J. (2009). The practice turn-away: Forty years of spoon-feeding in management education. *Management Learning*, 40(4), 37–52. doi:10.1177/1350507609335850

Reed, M., & Anthony, P. (1992). Professionalizing management and managing professionalization. *Journal of Management Studies*, 29, 591–613. doi:10.1111/j.1467-6486.1992.tb00680.x

Rodgers, C. (2002). Defining reflection: Another look at John Dewey and reflective thinking. *Teachers College Record*, 104(4), 842–866. doi:10.1111/1467-9620.00181

Schön, D. (1987). *Educating the reflective practitioner*. San Francisco, CA: Jossey-Bass.

Scott, S. G. (2010). Enhancing reflection skills through learning portfolios: An empirical test. *Journal of Management Education*, 34(3), 430–457. doi:10.1177/1052562909351144

ADDITIONAL READING

Brookfield, S. (1995). *Becoming a critically reflective teacher*. San Francisco, CA: Jossey-Bass.

Grey, C., & French, R. (1996). *Rethinking management education: An introduction*. London, England: Sage.

Halpern, D. F. (2014). *Thought and Knowledge: An Introduction to Critical Thinking* (5th ed.). New York, NY: Psychology Press.

Mezirow, J. (2000). *Learning as transformation: Critical perspectives on a theory in progress*. San Francisco: Jossey-Bass.

Porter, L., & McKibbin, L. (1988). *Business education and development: Drift or thrust into the 21st century*. New York, NY: McGraw-Hill.

KEY TERMS AND DEFINITIONS

Deep Learning: Learning based on critical examination and understanding of new facts, theories and concepts by linking them with existing cognitive structures and knowledge.

Experiential Learning: A process through which knowledge, skills and attitudes are developed through direct experiences that are supported by reflection, critical analysis and synthesis.

Metacognition: Self-monitoring one's own thought processes and the attendant assumptions that underlie the thinking.

Reflection: The practice of examining and analysing an experience to glean its cognitive essence in order to foster deep or life-long learning.

Surface Learning: Learning based on fact memorisation and lower level cognitive abilities.

Transformative Learning: Learning that brings about a change of perspective in the learner.

APPENDIX 1

C

Table 1. Question route for focus group interviews

Opening address Statement of purpose of the focus group session and the mechanics of the process. Emphasis on 'no pressure for consensus' to put participants at ease. Students were also assured of the confidentiality of their responses. Introductory question How do you interpret and understand the terms 'reflection' and 'reflective practice'? Key questions Why do you reflect? What factors encourage you to reflect? What are the impediments to the process of reflection? Transitional questions Before joining this institute and following the U.K. University curriculum have you ever reflected in a structured manner or indulged in practices which could be construed as reflection? Can you describe any of your experiences with reflection? Key questions In what ways can you reflect? Do you reflect in isolation or share your reflections with others? Probing questions Do you reflect in the academic context alone? Closing questions Any other thoughts about 'reflection'? Is there anything else that you feel may have been overlooked?

APPENDIX 2

Table 2. Thematic analysis of the findings of the focus group study

Themes of the Study	Findings
Perception of Reflection	• Tool for introspection, self-analysis and evaluation of experiences • Meaning-making process which serves as a guide to future action
Practice of Reflection	• Reflection can be practiced in isolation but works more effectively as a collaborative process • External articulation helps consolidate and concretize reflective thoughts • Social media such as Facebook, Twitter and on-line blogs have emerged as popular tools for external articulation and collaborative reflection
Experience with Reflection	• Increased familiarity and exposure to reflective practices improves perceived utility of reflection and reduces associated tedium. • Limited use of reflection as a pedagogical tool in mainstream education in India
Benefits of Reflection	• Helps change attitude and perceptions based on analysis of experiences • Enhances learning curve and leads to personal and intellectual growth • Increases self-awareness, confidence and emotional intelligence
Factors Enhancing and Inhibiting Reflection	• Enhancing Factors: Personal disposition, open-mindedness, collaborative environment and time • Inhibiting Factors: Lack of time, mental unpreparedness and lack of support systems conducive to reflective practices

Relationship Among Intelligence, Achievement Motivation, Type of School, and Academic Performance of Kenyan Urban Primary School Pupils

Jessina Mukomunene Muthee
Kenyatta University, Kenya

Catherine G. Murungi
Kenyatta University, Kenya

INTRODUCTION

In Kenya and many other countries examinations have been accepted as an important part of the education system. Examinations have also been used as basis for judging students ability and means of selection for educational advancement and employment.

Background

Every year thousands of Kenyan pupils sit for the Kenya certificate of primary education (K.C.P.E.) examinations. This is an examination done at the conclusion of eight years of learning in primary school. In spite of learners being exposed to the same environment, uniform program of classroom instruction and some may even possess similar IQ'S, discrepancies in academic performance have been observed to arise every year. Private schools albeit the high cost ones have attained high scores in Kenya certificate of primary education and contributed to nearly 80% of form- one students in national schools.

Extensive literature survey on academic performance of primary school showed that very few studies addressed the relationship between psycho-social factors and academic performance in Kenyan context. Most of the available literature reviewed comprise of studies done in western

countries and small proportion from India. Studies in Kenyan setting are fewer in comparison to efforts made abroad. The present study attempted to explore the nature and degree of relationship between academic performance and selected psychosocial variables such as intelligence, type of school and achievement motivation.

Cognitive abilities are widely regarded as a key component of intelligence which is a concept that is difficult to define (Hucken burry and Hucken Burry, 1997). It has been said that there are probably as many definitions of intelligence as there are experts who study it. Simply put however intelligence is the biological substrate of mental ability, the brains neuroanatomy and physiology, the manifestation of intelligence, the level of performance on psychometric tests of cognitive ability (Gardner, 1993).

Numerous researches have been done in western countries on the relationship between intelligence and academic performance (Vygotsky, 1978; Brody 1997; Ceci & Williams 1997; Ediseth, 2002, Parker et al; 2004) and these researches seem to agree that there exists a relationship between intelligence and academic performance and that higher educational achievement is predictive of higher intellectual outcomes.

Intelligence as measured by the Raven's standard progressive matrices has been found to be the best predictor of student's grade point average

(GPA) in all grades (Laidra et al; 2007). Deary et al; (2007). Also found a strong and positive relationship between intelligence and academic achievement. This study examined between psychometric intelligence at age 11 years and education achievement in 25 academic subjects at age 16. The trait and a latent trait of education achievement were 0.81. General intelligence contributed to success on all 25 academic subjects.

Direct relationship between intelligence and academic performance has also been reported. (Gagne & Stpere, 2002; Kossowska, 1999; Smith & Dobbs, 1999). In addition some researcher's view intelligence and academic achievement as identical constructs. Others assert that intelligence is causally related to academic achievement of success in academic work (Laidra, Pull mann & Allik, 2007).

Academic achievement motivation has been defined as learners need or drive towards the achievement of success in academic work (Amalaha, 1975). Motivation affects people's lives every day. People are motivated to engage in academic tasks if they value the task and if they expect to succeed at a task (Eccless, 1993, and Eccless & Wigfield, 1992).

It has been observed that many factors promote or hinder the success of students in educational institutions. Among them are achievement motivation and satisfaction with school experiences (Lampart, 1993). A great deal of research has also revealed that students of high academic motivation are more likely to have increased levels of academic achievement and lower school dropout rates (Blank, 1997).

Other studies have also reported positive relationship between motivation and academic performance (Johnson, 1996, Sandra, 2002; Brouard and Garrison 2004; and Skaalvik, 2004, 2006).

Literature evaluating the effects of school type on academic performance has grown rapidly over the past decade. For example the impact of Catholic schools (Private) on academic performance has received considerable attention in USA.

Good school facilities which are mostly found in private schools are observed to support educational enterprises. Research evidence abound (Cash, 1993; Earthman and Lemaster, 1996; Lemaster, 1997; Lackney, 1999; Cotton, 2001; and Schneider, 2002) that clean air, good light, small quiet comfortable and safe environment are important for good academic performance. Schools that have poor conditions because of factors such as overcrowding, differed maintenance or poor designs signal poor operating practices and hence poor academic performance. Such schools are mostly identified among public sector.

High school type appears to be strongly associated with educational outcomes (Checci et al; 1999). Boero et al; (2001) also confirms the importance of high school type for academic performance using his study sample of Italian college graduates that showed that the final graduation mark dropped significantly when one compared general (public) and private schools graduates. Private schools seemed to do better.

Evans and Schwab (1995) report a study on the effects of Catholic schools on high school completion and college enrolment probability. They also highlighted the endogeneity issues that can arise from self selection of students into Catholic schools and used instrumental variables to identify the effects of Catholic schools attendance on measures of academic success. They concluded that catholic schools raise subsequent educational outcomes. The Catholic schools in this case represented private schools.

Although many studies are in favour of private schools doing better in academic performance as compared to public schools, other studies also seem to differ with the idea. For example, Bertola & Cheddi (2002), Studied a sample of University students from the University of Millan. Their study revealed that those coming from general (public) schools scored better than otherwise comparable students on a range of performance indicators. They also considered the differences in academic performance between public and private schools students, finding that public schools were associ-

ated to better performance followed by religious private schools and lay private schools.

Hypothesis

The study was designed to test the tenability of the following Hypotheses:

1. There is a significant relationship among intelligence, achievement motivation, school type and academic performance.
2. Pupils of private and public schools differ significantly in (a) academic performance, (b) intelligence and (c) Achievement motivation.

Method

Sample: The subjects of study comprised 200 (101 boys and 99 girls) standard eight pupils from Nairobi city primary schools. The stratified random sampling technique was used to have private and public school representation. Their age ranged from 12 and 14.5years.

Measures

1. **Raven's Standard Progressive Matrices (SPM: Raven, 1939):** This was used as a screening test of cognitive ability. The test is designed to measure a person's ability to form perceptual relations and to reason by analogy independent of language and formal schooling and may be used with person ranging from six years old to adulthood. The candidate is provided with multiple choices to identify the missing segment required to complete a larger pattern in each test item. In this study, the test was administered in groups under the supervision of trained research assistants and the role scores were converted to percentile scores for analysis. The test was selected as it eliminates the bias of language. Total scores indicate an individual's intellectual capacity regardless of his/her level of education.

2. **Measure of Academic Performance:** The overall marks obtained by pupils in the class examination of various subjects were collected from school records. The scores were then converted into percentages and were used as indices for academic performance.
3. **Achievement Motivation Inventory (Muthee & Thomas, 2009):** This is a 32 item scale intended to assess the achievement motivation levels among students. Cronbach alpha computed for the scale was found to be 0.749 indicating that the scale has satisfactory internal consistency reliability. Validity for the scale is claimed on the basis of the fact that it is modeled after well known inventories meant for measuring achievement motivation. The scoring of the scale is done in such a manner that high scores indicate high levels of achievement motivation.

Statistical Techniques

The major statistical techniques used for analysis of data were t- tests and multiple regressions. The results obtained using these techniques are detailed below. Mean scores on different study variables obtained by students belonging to private and public schools were analyzed using t – testes. The results obtained are presented in Table 1.

In Table 1, the results of t – tests indicate that the students of private schools differ significantly from students of public schools in terms of academic achievement, and intelligence but not in terms of achievement motivation.

An examination of details of scores obtained by the sub groups indicate that private school students obtained greater means scores in academic achievement (71.22) and intelligence (mean percentile =76.57) and when compared to scores obtained by public school students (60.26% and percentile of 60.24 respectively).

The findings in the present study are in agreement with those of Boyle (1962) who examined the effects of differences in quality high schools

Table 1. Differences based on type of school

Variables	Type of School	N	Mean	Std. Deviation	t-values	Sig.
Academic achievement	Private	100	71.22	11.68836	5.884	.000
	Public	100	60.26	14.50226		
intelligence	Private	100	76.57	19.54665	4.867	.000
	Public	100	60.24	27.25607		
Achievement motivation	Private	100	125.18	13.67988	-1.108	.269
	Public	100	123.01	14.01694		

in Western Canada. His findings revealed that students who attend good quality schools has higher educational aspirations and expectations and hence perform better.

One explanation that may be offered for the present finding is that those who are studying in better schools are already a select group who can be expected to be above average in many respects including intellectual ability. In such a case, it may be not surprising that their achievement is also superior to those studying in schools with less facility.

From Table 2 it can be seen that among the variables considered for the study, intelligence was the first and the most important variable to enter the regression equation. The multiple R of intelligence is 0.593 and the R^2 is 0.352. This means that intelligence alone accounts for 35.2 percent of variance in academic performance of the students. The beta value associated with intelligence is 0.445. The positive sign of beta value indicates that there is a positive correlation between intelligence and academic performance.

Achievement motivation was the second variable which entered into the regression equation for prediction of academic achievement. With the addition of this variable, the multiple R improved to 0.655 and R. square became 0.429. The change in R square was 0.076. This meant that achievement motivation accounted for an additional 7.6 percent of variance in academic achievement. Intelligence and achievement motivation in combination could predict 42.8 percent of variance in academic achievement. The beta value for the achievement motivation was 0.282 which indicated positive relation between achievement motivation and academic achievement.

Table 3 describes the third and final variable which entered into regression equation for prediction of academic achievement was type of school. the beta value for school type was 0.208 which indicated a positive relation between type of school and academic achievement. With the additional of this variable the multiple R became 0.683. The change in R squares was 0.038. This showed that school type together could account for 46.6%variance in academic performance.

Table 2. Relative contribution of the Independent variables to the prediction

Predictors	Unstandardized Coefficients		Standardized Coefficients			Multiple Correlations		
	B	Std. Error	Beta	t	Sig.	R	R ²	R ² Change
Constant	3.259	7.174		.454	.650			
Intelligence	.253	.034	.445	7.503	.000	.593	.352	.352
Achievement motivation	.293	.058	.282	5.042	.000	.655	.429	.076
Type of school	5.962	1.649	.208	3.615	.000	.683	.466	.038

Table 3. List of variables which got excluded from the final regression equation

Variable	Beta in	t	Sig
Gender of the child	-.90°	1.679	.095
No. of brothers	-.013°	-.236	.813
No. of sisters	-.003°	-.058	.953
Living with	-.048°	-.875	.383
Education of the father	.081°	1.176	.241
Education of the mother	-.037°	.519	.605
Monthly income	.116°	1.564	.119
Occupational status of father	.088°	1.474	.142
Occupational status of mother	-.036°	-.615	.540
Socio economic status T.score	.016°	.205	.838
Class climate	-.036°	-.641	.522
Home environment	-.004°	-.078	.938

The multiple correlation of 0.683 which amounts to an explained variance of 46.6% shall be considered sizeable, in view of the fact that only 3 variables are involved in the equation. Among these three variables the most important one is intelligence which alone accounts for 35.2% of the explained variance. The finding is understandable and agrees with the common sense notion that the major part of one's academic achievement is determined by one's intellectual potential. The results obtained are also in consonance with the work of several earlier researchers (e.g. Vygotsky, 1978, Brody, 1997, Ceci & Williams, 1997; Cronbach, 1949).

Of greater interest and significance is the finding that achievement motivation, which is a non-cognitive variable, has a role in academic achievement that is over and above the role played by one's cognitive potentiality. This implies that significant improvement can be made in achievement of a student by optimizing his or her motivational level. Findings similar to the present one have been reported by Johnson 1996; Broussard and Garrison, (2004).

With regard to the variable type of school, the results obtained in the present analysis imply that being in a private school is a predictor of higher academic achievement when compared to be-

ing in a public school. This finding is of special significance in the context of Kenyan society. Where private schools in the private sector provide better facilities for students and are well staffed when compared to schools in the public sector. Summarizing these findings, multiple regression analysis reveal that better academic achievement among standard eight pupils in Nairobi, Kenya is predicted by higher levels of IQ, higher academic motivation and being placed in private schools for studying.

FUTURE RESEARCH DIRECTIONS

A similar study may be conducted on predictors of academic performance and motivation in primary schools of rural areas to cross check whether the conclusions arrived at in the present study which was conducted on urban setting, could be generalised.

The present study concentrated on correlates of academic achievement and motivation among adolescents in primary schools. There is need to examine the factors which influence academic performance and motivation in higher education among post adolescent learners.

Personality variables should be also taken up besides school variables and home environment variables in order to get more comprehensive picture of the variables affecting academic achievement and motivation.

CONCLUSION

The association found between various psychosocial variables viz., achievement motivation, type of school level of intelligence with academic achievement, revealed in the present study have lent some additional support to the findings of earlier studies.

The study helped to generate empirical support to the contention that both cognitive factors (intelligence) and non cognitive factors e.g., motivational and environment are equally important in determining academic performance of students.

The importance of achievement motivation as a predictor of academic achievement, revealed in the present study, as well as the identification of the significant predictors of achievement motivation has important implications for both educators and parents. Motivation being a non cognitive trait, amenable to modification through training, one may attempt to improve the motivational level of learners in order to bring positive results in academic performance.

From the finding the impact of type of school on academic achievement, implies that the school atmosphere, determined by various factors like facilities available, quality of teaching, evaluation, accountability and discipline maintained, etc., has some relevance in academic achievement of students. It is hoped that these findings shall be given serious attention by school administrators and governmental authorities and initiate positive changes in the administrative and academic set up of schools.

REFERENCES

- Amalaha, B. M. (1975). Academic achievement Motivation of Ibo fifth formers. *Abstract Int.*, 36(1), 123-A.
- Batnagar. (1991). Children's psychological development as related to home Environment. *Indian Journal of Applied Psychology*, 28, 1.
- Bertola, G., & Checci, D. (2002). *Sorting and private Education in Italy*. Centre for Economic Policy Research discussion paper No: 3198.
- Black, M. W. (1997). *Principles of Behavior: An introduction to behavior theory*. New York, NY: Applicator.
- Boero, G., Mcknigght, A., Naylor, R. A., & Smith, J. (2001). Graduates and graduates labor market in the UK and Italy. *Central Economic Journal*, 11(1), 20–22.
- Brody, N. (1997). Intelligence, schooling and society. *American Psychologists*, 52, 1046–1050.
- Broussard, S. C., & Garrison, M. E. (2004). The relationship between classroom motivation and Academic achievement in elementary school aged children. *Family and Consumer Sciences Research Journal*, 33(2), 106–120. doi:10.1177/1077727X04269573
- Cash, C. (1993). *A study of the relationship between school building condition and student achievement behavior* (Unpublished doctoral dissertation). Virginia Polytechnic Institute and State University.
- Ceci, S. J., & William, S. W. M. (1997). Schooling, intelligence and income. *The American Psychologist*, 52(10), 1058–1063. doi:10.1037/0003-066X.52.10.1051

- Checchi, D., Ichino, A., & Rustichini, A. (1999). More equal but less mobile? Intergenerational mobility and inequality in Italy and in the US. *Journal of Public Economics*, 74(33), 351–393. doi:10.1016/S0047-2727(99)00040-7
- Cotton, K. (2001). *New small learning communities: Findings from recent research*. Portland, OR: Northwest Regional Education Library. Retrieved 02/11/2010 from: <http://www.nwret.org/scpd/sirs/nslc/pdf>
- Cronbach, L. J. (1949). *Essentials of psychological testing*. New York: Harpern.
- Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and Educational achievement. *Intelligence*, 35(1), 13–21. doi:10.1016/j.intell.2006.02.001
- Earthman, G. I., & Lemaster, L. (1996). *Review of research on the relationship between schools buildings. Students achievement and student behavior*. Paper presented at the annual meeting of the council of educational facility planners international, Tarpon Springs, FL.
- Eccless, J. S. (1993). Expectancies value and Academic behavior. In J. T. Spence (Ed.), *Achievement and Achievement motives* (pp. 75–146). San Francisco, CA: Freeman.
- Ediseth, A. (2002). The relationship between intelligence approaches to learning and academic achievement. *Scandinavian Journal of Educational Research*, 46(2), 219–230. doi:10.1080/00313830220142218
- Evans, W. N., & Schwab, R. M. (1995). Finishing High School and starting college: Do Schools make a difference? *The Quarterly Journal of Economics*, 110(4), 941–974. doi:10.2307/2946645
- Gagne, F., & Stpere, F. (2002). When IQ is controlled, does motivation still predict achievement? *Intelligences*, 30(1), 71–100. doi:10.1016/S0160-2896(01)00068-X
- Gardner. (1993). *Multiple Intelligences: The Theory in practice*. Basic Books.
- Huckenbury, D. H., & Huckenbury, S. E. (1997). *Psychology*. London: Worth.
- Johnson, J.O. (1996). *Child Psychology*. Calabar: Nigeria Wusen Press Limited.
- Kossowska, S (1999). Learning styles: Differential effects of self control and deep level information processing on academic achievement. *Personality Psychology in Europe*, (7), 263 –281.
- Lackney, J.A. (1999). *Assessing school facilities for learning/Assessing the impact of the Physical environment on the educational process*. Mississippi State.
- Laidra, K., Pullmann, H., & Allik, J. (2007). Personality and intelligence as predictors of Academic Achievement: A cross sectional study from elementary to secondary school. *Personality and Individual Differences*, 42(3), 441–451. doi:10.1016/j.paid.2006.08.001
- Lampart. (1993). *A theory of goal setting and task performance*. Academic Press.
- Lemasters, L. K. (1997). *A synthesis of studies pertaining to facilities, student achievement and student behavior* (Unpublished doctoral dissertation). Virginia Polytechnic and State University.
- Muthee, J. M., & Thomas, I. (2009). *Achievement Motivation Inventory*. Trivandrum: Department of Psychology, University of Kerala, Kariavattom.
- Parker, J. D. A., Creque, R. E. Sr, Barnhart, D. L., Harris, J. I., Majeski, S. A., Wood, L. M., & Hogan, M. J. et al. (2004). Academic achievement in high school: Does emotional intelligence matter? *Personality and Individual Differences*, 37(7), 1321–1330. doi:10.1016/j.paid.2004.01.002
- Raven, J., Raven, J.C., & Court, J.H. (1938). *Raven Manual: Section I Standard progressive Matrices*. Oxford, UK: Oxford Psychologists Press Ltd.
- Sandra, A. (2002). Predictors of University Academic performance in Colombia. *The Journal of Educational Research*, 35, 411–417.

Schneider, M. (2002). *Do school facilities affect academic outcomes?* Washington, DC: National Clearinghouse for Educational Facilities. Accessed from <http://www.edfacilities.org/pubs/outcomes.pdf>

Skaalvik, E. M., & Skaalvik, S. (2006). Self concept and self efficacy in mathematics: Relation with mathematics motivation and achievement. *Proceedings of the international conference on learning science*. available at: <http://www.find.articles.com>

Skaalvik, S., & Skaalvik, E. M. (2004) *Gender differences in Math's and verbal self concept, performance exceptions and motivation sex role*. Available at: <http://www.find.articles.com>

Vygotskky, J. L. (1978). Learning and intelligence. *Personality and Individual Differences*, 27, 715–735.

Wigfield, A., & Eccless, J. S. (1992). The development of Achievement task values: A theoretical analysis. *Developmental Review*, 12(3), 265–310. doi:10.1016/0273-2297(92)90011-P

ADDITIONAL READING

Dweck, C. (2000). *Self theories, their role in motivation, personality and development*. Washington, DC: Psychologist Press.

KEY TERMS AND DEFINITIONS

Academic Performance: As used in the contest of this chapter, the expression academic performance refers to the percentage of marks obtained by the pupils on each of the five compulsory subjects after two terms of study in standard eight class. The overall percentage on five subjects

was calculated from this information. Information forms containing columns to enter total marks obtained by the students and the maximum marks on each of the five compulsory subjects were used for the purpose. This information was provided by the school authorities.

Intelligence: The term ‘intelligence’ as used in the context of the present study refers to cognitive abilities. This construct was operationalised using standard progressive matrices, which was developed by Raven’s (1938) to test people’s ability for observation and clear thinking. It measures a person’s ability to form perceptual relations and to reason by analogy. The test can be administered to any individual in the age range of 6 years and above irrespective of language and schooling. The total score indicates an individual’s intellectual capacity regardless of his or her level of education.

Motivation: The expression ‘motivation’ in the present study refers to the persons aroused desire for participation in the learning process and perform at high level of competence. This construct was operationalised using an achievement motivation inventory constructed by the present investigators (Muthee, murungi and Thomas, 2009) The tool measured four different aspects related to academic motivation, viz., ⁽¹⁾ motivation for achievement ⁽²⁾ inner resources of the learner ⁽³⁾ interpersonal strengths in learners and ⁽⁴⁾ work habits.

Type of School: In Kenya there are two types of schools. The government sponsored schools (public schools) and the private owned schools (private schools). In this study both public and private primary schools participated in the study. Primary school level: this is the elementary level school, it houses pupils of ages ranging from seven to sixteen years. In this study pupil’s in the age range of eleven years to fourteen under half years participated in the study. These learners are expected to be in standard eight class in Kenyan primary schools.

Screencasts and Learning Styles

Rui Alberto Jesus

CESPU, Instituto de Investigação e Formação Avançada em Ciências e Tecnologias da Saúde, Portugal

INTRODUCTION

Many people have heard about learning styles. Besides the criticism about the instruments that detect and classify these preferred learning modalities, learning styles appear to explain something that is obvious: people learn in different ways.

In broader terms, people have different personalities, different ways of processing information, and different sensory preferred modes (among others factors), that influence how each person relates to a learning environment.

In this chapter, the emphasis is on the different sensory modalities. Individuals perceive stimuli from the outside through the five senses, but each individual may have one or two of these senses more accurate than others. As such, he tends to acquire information more through that sense. This preference for a sensory modality has been investigated to explain the success or failure students have to assimilate certain learning content (e.g.: more orally or more in written form).

On the other hand, professors can use several didactic materials to deliver instruction to their students (particularly in eLearning). Text, images and diagrams, audio, video, simulations are all valid means to deliver pedagogical information. But which of these means suites best a particular learning style?

This chapter discusses the contribution of screencasts as one possible solution to that problem. A screencast is a digital recording of computer screen output, including mouse movements and clicks. Also known as a video screen capture, screencasts can include audio narration to explain the process that is being documented by the screencast. If well

planned and recorded, screencasts can include text, images, diagrams, audio, video and simulations, thus aiming to reach several learning modalities, including the preferred one of a particular student.

BACKGROUND

What Are Learning / Cognitive Styles?

One of the pioneers of the term ‘cognitive styles’ was Gordon Allport (1937), which defined them as the usual or typical way of an individual processing information. In other words how he perceives, thinks and remembers information, and how he uses it to solve problems. Since then, there have been many researchers who have dedicated themselves to study this concept, with the consequent identification of different cognitive and learning styles. For example, the work of Messick (1976) identified 19 different dimensions of cognitive styles (field dependence versus field independence, global versus analytic, inductive versus deductive, visualizer versus verbalizer, etc.), some of which are referred in the Additional Readings section.

Before moving on, it should be clarified that in this chapter, like in most of the area’s texts, the terms ‘cognitive styles’ and ‘learning styles’ are used to describe the same concept, although the first one is more used in the context of academic research, while the latter one is more related to their practical applications. Moreover, the term ‘cognitive styles’ is more connoted with a bipolar characteristic (e.g.: a student is either inductive

or deductive), while the term ‘learning styles’ does not require the existence of two poles (e.g.: one student may be visual and kinesthetic at the same time).

The VARK Model of Learning Styles

During the 1980s, and in informal conversations with college students, Neil Fleming realized that many of them attributed their learning difficulties, to the way learning content was presented. Some students said they had more difficulties with content presented orally; others, with written texts; some more, with ideas that were presented in graphical form; and others with subjects that were presented without any connection to practical applications. This finding led the author to focus on the sensory modalities as a dimension of learning styles with some prominence in relation to the other dimensions (Fleming & Mills, 1992).

In addition, the author found some basis for this assumption in his research in the neuro-linguistic programming area, which years before, had already identified three different sensory modalities – aural, visual, and kinesthetic – described below:

- *Aural* individuals learn best by listening, for example, to a lecture given by a teacher.
- *Visual* individuals learn best by viewing, for example, a video, an image, etc.
- *Kinesthetic* individuals learn best when, for example, attend a lecture but write what they hear or perform something practical.

Thus, more than studying the various dimensions of learning styles, which seemed to have little practical application for students (Fleming & Mills, 1992), Fleming was interested primarily in the sensory preferences of students, giving rise to what he designated the VARK model (which is an acronym for Visual, Aural, Read / write & Kinesthetic).

As noted by the acronym, Fleming added a second visual modality to the neuro-linguistic programming model referred above. It is the

read / write modality, because, according to the author, there are some students who prefer text content, i.e., written words basically, while other students prefer other symbolic forms to represent information such as maps, diagrams and charts. Although both preferences are visual, seldom they are present in the same person (Fleming & Baume, 2006).

In summary, the VARK model proposes the following sensory modalities to enhance learning (Fleming, 1995):

- **Visual:** Learning centered on viewing images, graphs and diagrams, as well as in color and formatting variations of documents (e.g.: highlighted boxes); good perception of symbolic information.
- **Aural:** Learning centered on hearing, i.e., using the ears to receive the commonest way to exchange information in society, which is speech.
- **Read / Write:** Learning through texts, that is, through the written word; it is called “read / write” because this type of students use reading and writing as the first choice to receive information.
- **Kinesthetic:** Multisensory and practical oriented learning, i.e., this type of students like to experience learning through all senses, and although they like to learn by doing, they can also learn concepts and more abstract materials, provided they are accompanied by concrete examples or real life scenarios (in a phrase, they like learning through action).

Each individual has one or more of these sensory modalities by which he prefers to learn, and there is a questionnaire with sixteen questions, which helps students to diagnose those preferences (VARK Learn Limited, 2016).

One of the advantages of this learning styles model is that, besides allowing to know the sensory preferences of an individual, it also provides study tips for more effective learning. This is because

the authors emphasize that the way information is presented to students, does not have to be the same in that they record it for later study. For example, if a teacher gives priority to oral presentations in his classes, a visual student should make diagrams and drawings in his notebook, which better reflect the verbal information he just heard (Fleming, 1995).

This VARK model priority – to promote the adaptation of students to the formats in which teachers present their contents – seemed more feasible to the authors, than to promote that teachers diversify their teaching methodologies (Fleming & Mills, 1992). This is because, knowing how to make this adjustment, students gain tools to get the best out of any course, and not only those in which the teacher diversifies his teaching materials to meet several sensory preferences. On the other hand, it is known that teachers tend to present their contents according to their own sensory preference (Fleming & Baume, 2006).

Nevertheless, the model also motivates teachers to reach as much sensory modalities as possible, and gives tips on how to design classes and teaching materials with that goal in mind. In fact, the VARK model dissemination among the teaching community made them more aware for the need to go beyond the most frequent presentation, which is the unimodal or, at most, bimodal (even with eLearning, that in many cases, only converts speech into text, i.e. the aural modality by the read / write modality) (Fleming & Baume, 2006).

In terms of the VARK questionnaire's statistical validity – i.e., can that instrument detect and classify correctly, the preferred learning modalities of students? – although there are potential problems related to item wording and the scale's scoring algorithm, there is evidence of VARK's validity for measuring learning preferences (Leite, Svinicki, & Shi, 2010). The potential problems are due to the fact that respondents can select more than an option, in each of the questionnaire's sixteen questions. Each question presents a scenario – e.g.: *"You have a problem with your heart."* – and four options for the respondent to choose his preferred sensory modality(ies) – e.g.: *"You would prefer*

that the doctor: (i) used a plastic model to show what was wrong; (ii) gave you something to read to explain what was wrong; (iii) described what was wrong; (iv) showed you a diagram of what was wrong."

Despite the limitations of the VARK model, that does not decrease its usefulness to illustrate that different students prefer to perceive teaching materials in different formats, which may help to explain the relations between screencasts and learning styles (the aim of this chapter).

Another common confusion is to consider VARK's sensory preferences as skills or strengths of individuals. However, to like something does not necessarily mean that a person is good at it, i.e. the VARK model helps to know how a student likes to communicate, but says nothing about the quality of that communication (Fleming & Baume, 2006).

Screencasts

A screencast is a digital recording of computer screen output, often accompanied by audio narration to explain the process being described (Peterson, 2007). As the PrintScreen key allows to capture the image of a computer screen, a screencast is a film of what can be seen in this screen, over a given period of time.

Hence there is the need for a specific software in order to record screencasts (a screencasting software), which besides recording what is happening on the screen, it also allows the edition of the screencast's sequences (e.g.: cutting a part where the speaker coughed), and the production of the final result in a universal file format (e.g.: Adobe Flash) (EDUCAUSE Learning Initiative, 2006).

However, unlike face-to-face classes that can be recorded with a traditional video camera, or even an audio recorder, classes that deal with computer-based content require a higher quality recording, in order to allow an effective viewing (and hearing) of the computer screen content. Normally, with regular video cameras one cannot achieve that quality.

Screencasts are typically indicated to show tasks that are performed with a computer (such as how to operate with a particular software), but not only...

Within a course, the demonstrative tasks are the first candidates to be recorded in screencasts (e.g.: “How to search a scientific database?”). However, also the lectures can be made available in screencasts, not just to be accessible to students who miss face-to-face classes, but also because these classes can be more effective this way than face-to-face (at least for some students). Why? Basically for these three reasons (Jesus & Moreira, 2009):

- Unlike a classroom lecture, students can view the screencasts when they want to, especially in the times of the day that enhance their ability to concentrate / retain the subject matter (e.g.: there are students who prefer to study in the evening, others in the morning, etc.).
- In addition, there are students who feel that the classroom environment contains a certain ‘noise’ – e.g.: the whispering from behind, the side colleague who asks a question – that hinder the full message reception transmitted by the teacher. When viewing the screencast of the class, the student can eliminate this ‘noise’.
- Furthermore, the student can stop the screencast when he wants to and continue viewing it later; or else, if he gets confused, he can rewind the screencast, and repeat the visualization of the most difficult parts to understand them better. The same goes even weeks or months later – the student can always go back to the screencast to refresh his memory in relation to any topic of interest.

Of course, these reasons apply to screencasts of all kinds of classes, but with particular relevance to lectures, because they can address contents of a higher level of abstraction.

Another major advantage of screencasts is that the recording task does not require a class taking place face-to-face, in order to be recorded “live”. The recording phase can (and should) be conducted during the planning of classes, prior to the beginning of the school year. Thus, screencasts of all the course’s subjects can be delivered to students in the first day of the semester. Among other things, this means that the most interested students can also view the screencast of the next class before it takes place. Those who do, will take an additional advantage of screencasts – use them to prepare face-to-face classes, taking advantage of the time with the professor, mostly to elucidate questions and doubts that did not get perfectly understood by simply viewing the screencast. This is the major teaching strategy advocated by flipped learning supporters (Faculty Focus, 2015).

SCREENCASTS AND LEARNING STYLES: A CASE STUDY

In order to discover if there is a relation between screencasts and sensory preferences, a study was designed with a group of nursing students. What kind of students, i.e., with what kind of sensory preferences, would value the most, learning through screencasts? Are screencasts more valued by multimodal students than by unimodal ones? These are some of the questions that this section tries to answer. It starts by identifying the research methodology, followed by the major results found and its discussion.

Methods

The author of this chapter is the professor of a statistics course whose participants are third year students from the Instituto Politécnico de Saúde do Norte (IPSN, a Portuguese private higher education school; URL: <http://www.cespu.pt/en/>). The course aims to prepare nursing students to conduct research projects in their field. With a 2-hour class per week (face-to-face), a page

in a Moodle platform, and screencasts of all the course's subjects to support the course during and between classes, the unit of Introduction to Research takes place in a fully equipped classroom with personal computers and broadband Internet connection.

A transversal study was conducted, at the beginning of 2015/2016 academic year, where 57 undergraduate nursing students (convenience sample) filled out the VARK questionnaire (version 7.1) (VARK Learn Limited, 2016), and another survey with sociodemographic questions, as well as questions related to the usefulness and satisfaction with screencasts, Moodle resources and face-to-face classes.

Besides those two questionnaires, data collection methods included the students' ratings in the course's assessment moments, and the Moodle's log files information, besides the records of class attendance. All this information was stored and utilized for analysis after removing all individual identification data in order to protect the research subjects' privacy. The participants were informed of that practice prior to the beginning of the study, and signed a consent form approving that method.

The collected data was analyzed with descriptive and inferential statistics methods (using SPSS version 23). Correlation between VARK scores and screencasts satisfaction, as well as mean differences tests (e.g.: Mann-Whitney U test) were utilized to answer the research questions, always with a significance level of 0.05.

Results and Discussion

The following topics summarize the sample characterization:

- 57 students with more females than males (3/4 versus 1/4);
- The vast majority (86%) were traditional students (coming directly from high school to higher education), with a span of 20 – 22 years of age;

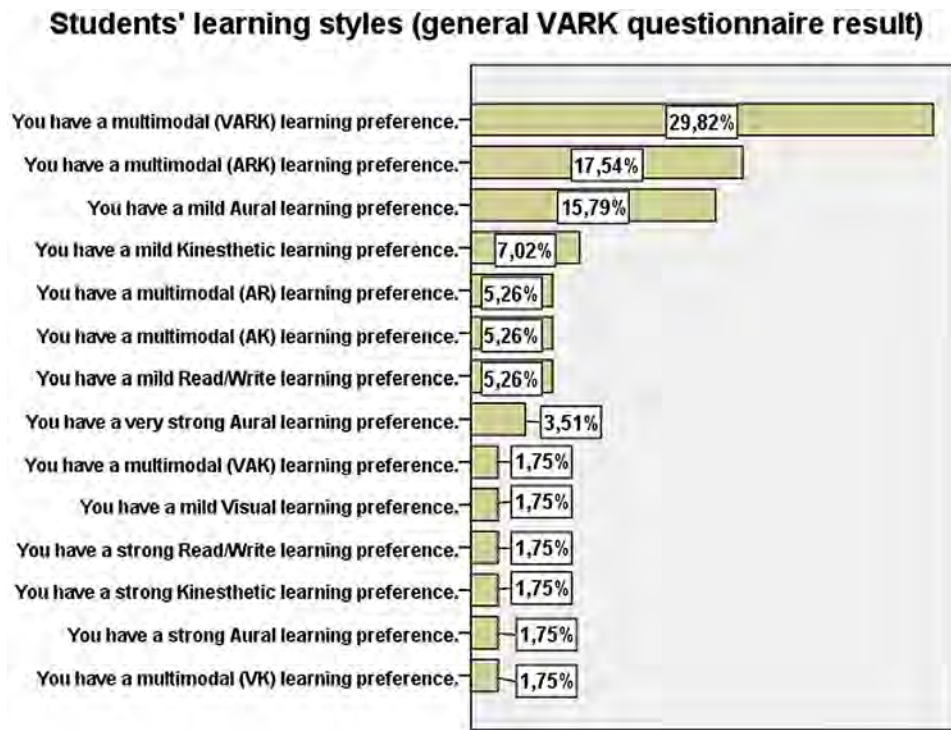
- Almost all students (90%) have chosen the bachelor's degree in Nursing as their first choice in higher education (hence, they were motivated);
- For almost all students (93%) it was the first time they watched screencasts;
- For three quarters of students it was the first time they accessed a Moodle platform;
- The students ended the curricular unit with good grades (median of 14.1 ± 2.4 in a scale of zero to 20), and only 4 students failed to approve.

VARK Modalities and Scores

Figure 1 shows that the majority of these students were multimodal (61.4%), as opposed to unimodal (38.6%). This result is in line with the general population (VARK Learn Limited, 2015), as well as with other samples of medical students (Prithishkumar & Michael, 2014) (Samarakoon, Tharanga, Rodrigo, & Rajapakse, 2013) (D'Amore, James, & Mitchel, 2012). In the multimodal subgroup, quad modal (VARK) was the predominant style, with 29.8% of all students. In the unimodal subgroup, the aural modality (mild, strong and very strong) was the predominant one, with 21.1% of all students.

As for the frequency of V, A, R and K scores, calculated from the questionnaire, remember that each individual score varies from zero to 16 (the number of items in the questionnaire). But as there were students who chose more than one option in each question (i.e., who chose more than one sensory preference), it is best to convert the scores into the scale of zero to 100%. Thus, if a student only chose one sensory preference in each one of the 16 questions, each score was divided by 16 to obtain the percentage. But if a student chose more than one sensory preference in the survey questions, each score was divided by the total number of options that the student has selected (> 16). Figure 2 illustrates the average frequency of each V, A, R and K scores, in percentage, for all respondents. Thus, in the scale of zero to

Figure 1. Students' learning styles (general VARK questionnaire result)



100%, predominated the aural modality (average score of $31.67\% \pm 10.57\%$), then the read / write modality (average score of $26.86\% \pm 8.6\%$), followed by the kinesthetic modality (average score of $25.33\% \pm 10.08\%$), and finally, the visual modality (average score of $16.13\% \pm 8.52\%$). In other studies with medical students around the world, summarized in a recent literature review (Khanal, Shah, & Koirala, 2014), the predominant sensory modalities were the kinesthetic, followed by the aural. The kinesthetic being justified by the practical nature of health degrees (e.g.: clinical internships); and the aural being justified by the fact that students are used to lectures throughout their entire school path.

Students' Satisfaction with Screencasts

Figure 3 shows that more than three quarters of these students thought that oral explanations were the most useful component in the screencasts. In

other words, they valued the most the teacher's audio narration that explained the process being described by the screencasts. This result is consistent with the predominant aural modality found above.

Followed by close and selected by 71.9% of all students, the examples and real scenarios were the second most valued component of screencasts. This component is more addressed to kinesthetic students, whose score was also well represented in this sample, as seen above.

Diagrams and illustrations (more addressed to visual students), and written text (more addressed to read / write students) were the least valued components of screencasts, referred by less than half of these students. And if the visual modality had the smallest average score in this sample (16.13%), it is strange that the written text was so poorly evaluated by students, because in terms of average scores, the RW modality was the second most predominant in this group of students. Maybe the media rich context of screencasts, which

Figure 2. Students' average scores in the VARK questionnaire

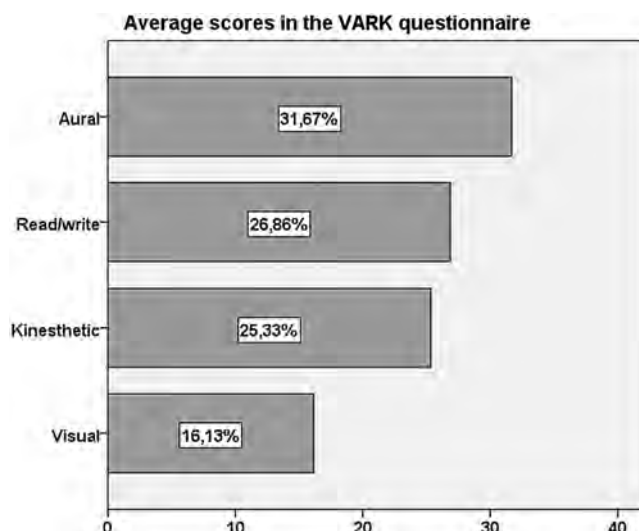
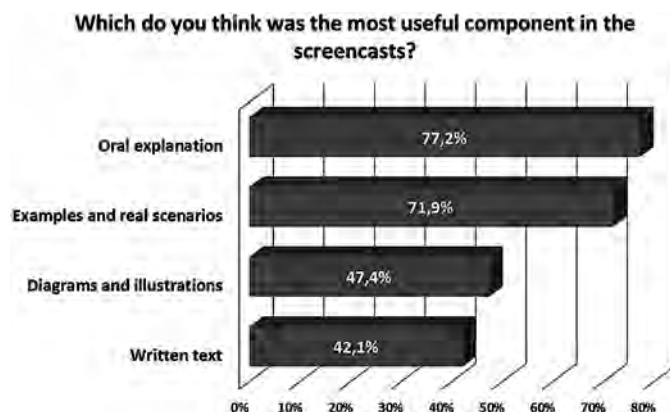


Figure 3. Most useful component in the screencasts, for students



can include images, diagrams, audio, video and simulations, has relegated the text component to an inferior rank, in these students' opinion.

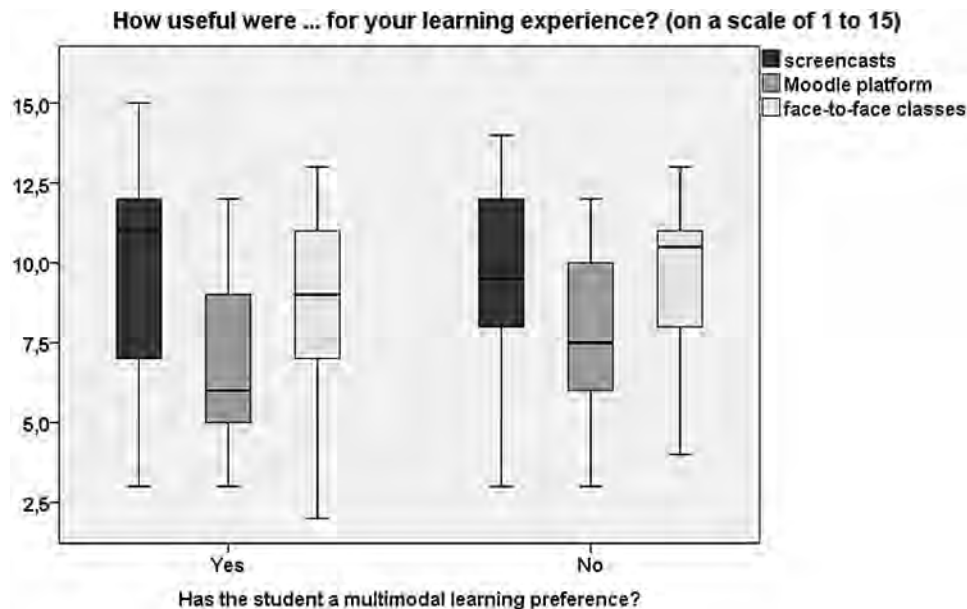
Relations Between Screencasts and Learning Styles

Based on several questions that were present in the second questionnaire (such as: "*How useful were screencasts / Moodle / classes for your learning experience?*", "*Describe how did you prepared yourself for the final exam*", and "*In the future, when you need to apply the skills and knowledge that you have learned in this unit, what kind of*

resources will you use?"), a total utility score was computed for each one of the three major didactic materials of this course: screencasts, Moodle and face-to-face classes.

Figure 4 illustrates that multimodal students found screencasts to be more useful than face-to-face classes and Moodle resources (screencasts' median = 11; classes' median = 9; Moodle's median = 6), and those differences were statistically significant ($p = 0.006$; effect size = 17.5%). On the contrary, unimodal students found face-to-face classes to be more useful than screencasts and Moodle resources (classes' median = 10.5; screencasts' median = 9.5; Moodle's median =

Figure 4. Usefulness of screencasts, Moodle and face-to-face classes, for the students learning experience



7.5), although those differences were not statistically significant ($p = 0.074$). This shows that screencasts are more appealing to multimodal students, and face-to-face classes are more appealing to unimodal students.

Comparing each component between multimodal and unimodal students, only in the screencasts component there is a higher usefulness perception of multimodal students, compared with unimodal ones (multimodal median = 11 versus unimodal median = 9.5). Both in Moodle resources and in face-to-face classes, unimodal students have perceived more value than multimodal ones (Moodle unimodal median = 7.5 versus Moodle multimodal median = 6; classes' unimodal median = 10.5 versus classes' multimodal median = 9). However, none of these between groups differences were found to be statistically significant.

Following there are other findings that try to answer to the question: "What kind of students, i.e., with what kind of sensory preferences, would value the most, learning through screencasts?":

- Students who found that the most useful component were screencasts, had an aural score lower than those who found that the

most useful component were face-to-face classes (classes' median aural score = 33.3 versus screencasts' median aural score = 26.1; $p = 0.011$; effect size = 11.7%);

- As the students' visual score increases, so does the enjoyment for learning through screencasts ($\rho = 0.34$, $p = 0.009$), and the amount of time invested on watching them ($\rho = 0.47$, $p < 0.001$);
- As the students' visual score increases, the enjoyment for learning through classes decreases ($\rho = -0.33$, $p = 0.012$).

This reveals that, in this sample, visual students were the ones who valued the most, learning through screencasts; and aural students were the ones who valued the most, learning through face-to-face classes. But caution interpreting the previous sentence because the majority of these students were multimodal (61.4%), as opposed to unimodal (38.6%). So, the majority of these students had multiple sensory preferences, and cannot be considered just 'visual students', nor just 'aural students'.

FUTURE RESEARCH DIRECTIONS

Taking into account that, in this experiment, students were only recipients of screencasts, it would be interesting to see how they would behave, if they were creators of screencasts. From this idea, one could design multiple experiences. For example, what would be the effects on their learning, if students were asked to design screencasts on certain thematic units, in order to share them with the rest of the class? Or, what if instead of writing a paper, students could present the results of their research project in screencast format? What kind of students (e.g., multimodal vs. unimodal) would be the best screencasts' creators?

Another suggestion is to test the feasibility of screencasts in eLearning assessments. As one knows, in eLearning settings is very difficult to warrant that the person doing the evaluation is the student himself. What if it was mandatory for students, to record their assessment test in a screencast (with voice narration), and send it to the teacher at the end of it? How would students react to this new form of surveillance? Would they have technical difficulties to record and submit screencasts? Would this new variant of evaluation – evaluate how you get to the end result, and not just the final outcome – would be beneficial to the students' learning process? These questions are pertinent to validate this proposed model of evaluation in eLearning.

CONCLUSION

Learning styles are not mutually exclusive categories but preferences that may be mild, moderate or strong. This explains the wide variation among learners with the same learning styles, when prior knowledge, experience, and skill level are factored into the learning style equation.

Can a teacher design instruction that addresses all these individual differences? Of course not and neither is that the point. But if teachers diversify their teaching methodologies, in order to achieve

a good balance, for instance with respect to the sensory channels that they appeal, the chances of different kinds of students feel comfortable with learning the teacher's message will increase, thus promoting a better learning experience for all.

Screencasts seem to be a good solution for this diversity, because they can include text, images, diagrams, audio, video and simulations, thus aiming to reach several learning modalities, including the preferred one of a particular student. In favor of this solution is the fact that, in a sample of undergraduate students examined by the author, screencasts were found to be more appealing to the multimodal ones (the majority of the sample and also of society in general (VARK Learn Limited, 2015)). Inversely, unimodal students (which were strongly composed by aural learners) found face-to-face classes more appealing to their learning, probably because they are lectures to a great extent. Prithishkumar & Michael (2014) also advocate the use of screencasts as a way to diversify the teaching strategies that reach several learning styles.

There is one last enduring message to be taken from the debate about learning styles. Addressing the learning needs of students is way more complicated than it seems. The ideal balance among learning style categories depends on the subject, level, and learning objectives of the course and the backgrounds and skills of students. That is the problem teachers should be working to solve but without expecting a single right answer.

REFERENCES

- Allport, G. W. (1937). *Personality: A psychological interpretation*. New York: Henry Holt and Co.
- D'Amore, A., James, S., & Mitchel, E. K. (2012, July). Learning styles of first-year undergraduate nursing and midwifery students: A cross-sectional survey utilising the Kolb Learning Style Inventory. *Nurse Education Today*, 32(5), 506–515. doi:10.1016/j.nedt.2011.08.001 PMID:21889237

EDUCAUSE Learning Initiative. (2006, March 15). *7 Things You Should Know About Screencasting*. Retrieved March 21, 2016, from <https://library.educause.edu/resources/2006/3/7-things-you-should-know-about-screencasting>

Faculty Focus. (2015). *Flipped Classroom Trends: A Survey of College Faculty*. Madison, WI: Magna Publications. Retrieved March 21, 2016, from <http://www.facultyfocus.com/free-reports/flipped-classroom-trends-a-survey-of-college-faculty/>

Fleming, N. (1995). I'm different; not dumb. Modes of presentation (VARK) in the tertiary classroom. In *Proceedings of the 1995 Annual Conference of the Higher Education and Research Development Society of Australasia (HERDSA) (vol. 18, pp. 308-313)*. Rockhampton, Australia: HERDSA.

Fleming, N., & Baume, D. (2006). Learning Styles Again: VARKing up the right tree! *Educational Developments*, 7(4), 4-7.

Fleming, N., & Mills, C. (1992). Not Another Inventory, Rather a Catalyst for Reflection. *To Improve the Academy*, 11, 137-155.

Jesus, R., & Moreira, F. (2009). Students Prefer Screencasts: The New Face of Early Days Distance Education. In *Proceedings of the IADIS International Conference: e-Learning 2009* (pp. 155-162). Algarve: International Association for Development of the Information Society.

Khanal, L., Shah, S., & Koirala, S. (2014). Exploration Of Preferred Learning Styles In Medical Education Using VARK Modal. *Russian Open Medical Journal*, 3(3), 1-8. doi:10.15275/ru-somj.2014.0305

Leite, W. L., Svinicki, M., & Shi, Y. (2010). Attempted Validation of the Scores of the VARK: Learning Styles Inventory with Multitrait-Multimethod Confirmatory Factor Analysis Models. *Educational and Psychological Measurement*, 70(2), 323-339. doi:10.1177/0013164409344507

Messick, S. (1976). *Individuality in Learning*. San Francisco: Jossey-Bass.

Peterson, E. (2007). Incorporating Screencasts in Online Teaching. *International Review of Research in Open and Distance Learning*, 8(3), 1-4. doi:10.19173/irrodl.v8i3.495

Prithishkumar, I. J., & Michael, S. A. (2014). Understanding your student: Using the VARK model. *Journal of Postgraduate Medicine*, 60(2), 183-186. doi:10.4103/0022-3859.132337 PMID:24823519

Samarakoon, L., Tharanga, F., Rodrigo, C., & Rajapakse, S. (2013). Learning styles and approaches to learning among medical undergraduates and postgraduates. *BMC Medical Education*, 13(1), 1-6. doi:10.1186/1472-6920-13-42 PMID:23521845

VARK Learn Limited. (2015). *Research & Statistics*. Retrieved March 21, 2016, from <http://vark-learn.com/introduction-to-vark/research-statistics/>

VARK Learn Limited. (2016). *The VARK Questionnaire*. Retrieved March 21, 2016, from <http://vark-learn.com/the-vark-questionnaire/>

ADDITIONAL READING

Brown-Sica, M., Sobel, K., & Pan, D. (2009). Learning for all: Teaching students, faculty, and staff with screencasting. *Public Services Quarterly*, 5(2), 81-97. doi:10.1080/15228950902805282

Entwistle, N. J. (1981). *Styles of Learning and Teaching*. London: Wiley & Sons.

Gazzaniga, M. S. (1998). The Split Brain Revisited. *Scientific American*, 279(1), 50-56. doi:10.1038/scientificamerican0798-50 PMID:9648298

Goldstein, K. M., & Blackman, S. (1978). *Cognitive Styles: Five Approaches and Relevant Research*. New York: Wiley & Sons.

Jesus, R., & Moreira, F. (2010). Best Practices of Blended-Learning with Screencasts. In *Proceedings of the 2010 International Conference on Frontiers in Education: Computer Science and Computer Engineering* (pp. 310-315). Las Vegas: World Congress in Computer Science, Computer Engineering, and Applied Computing.

Jung, C. G. (1923). *Psychological types*. Princeton, NJ: Princeton University Press.

Morris, C., & Chikwa, G. (2014). Screencasts: How effective are they and how do students engage with them? *Active Learning in Higher Education*, 15(1), 25–37. doi:10.1177/1469787413514654

Witkin, H. A., & Goodenough, D. R. (1981). *Cognitive Styles: Essence and Origins*. New York: International University Press.

KEY TERMS AND DEFINITIONS

eLearning: Contraction of electronic learning. Interactive training, available through the Internet or other means of electronic communication,

making the learning process independent of time and place.

Learning Style: Usual way of an individual processing information. How he/she perceives, thinks and remembers information, and how he/she uses it to solve problems.

Moodle: Open source learning management system used by all types of schools and training centers, to add web technology to their courses.

Multimodal Student: A student that has two, three or even the four VARK Model dimensions equally developed. He/she prefers to use several sensory channels to acquire information.

Screencast: Digital recording of computer screen output, including audio voiceover.

Sensory Modalities: Senses through which humans perceive stimuli from the outside world.

Unimodal Student: A student that has one of the four VARK Model dimensions more developed than the other three. He/she prefers to use a single sensory channel to acquire information.

VARK Model: Model of learning styles based in the sensory preferences of students. Its 4 dimensions (which form the model's acronym) are: Visual, Aural, Read / write and Kinesthetic.

Self-Awareness and Motivation Contrasting ESL and NEET Using the SAVE System

C

Laura Vettraino*Learning Community, Italy***Valentina Castello***CIOFS FP, Italy***Marco Guspini***educommunity – Educational Community, Italy***Eleonora Guglielman***Learning Community, Italy*

INTRODUCTION

Youth unemployment has been increasing over time in the last 10 years, although with very different rates, in all Countries. The last economic crisis has intensified the phenomenon of people being not in education, employment or training (NEET), extending it to areas where it was most content. According to OECD (2016), about 15% of youth aged 15 to 29 (corresponding to roughly 40 million young people) are NEET. However, crisis represents only a contingent factor: effective policies and intervention should address their structural drivers, mainly related to the quality of the educational, VET and job market systems and dynamics, as well as their dynamic alignment (Eurydice and Cedefop, 2014; Global Opportunity Report, 2016). It is well documented that reduced employment opportunities and increased likelihood of unemployment, poverty and social exclusion is strictly linked to early leaving from education and training (OECD 2012; European Commission, 2011).

The strong correlation between education and employment, links the NEET and early school leaving (ESL) phenomena in a vicious circle (Eurostat 2016) having negative impacts on social

inclusion and development, also in a medium-long term perspective. Young people with low level of education are three times more likely to become NEET compared to the others (Eurofound report 2012). The core issue of this relationship is not merely related to a lack of “professionalization”; it also calls for the progressive lack of expectations and motivations and in turn, in very low (or absent) active youth behaviours. The consequent vicious circle between ESL and NEET brings serious social and economical impacts, also in a long term perspective.

Although the both are featured by very diverse sub-groups and driving factors, three main common features and intervention directions emerge as critical: the improvement of the guidance, training and orienting systems in terms of learning and orienting personalization; the stronger focus on the active engagement and motivational aspects; the employment of a multi-dimensional and multi-actor approach, addressing contemporarily system, trainers, targets (European Commission, 2015; Eurydice and Cedefop report, 2014).

Referring to this scenario, this paper provides a review of the main effective approaches to contrast ESL and NEET phenomena; presents the methodological framework and the technology enhanced

DOI: 10.4018/978-1-5225-2255-3.ch135

guidance and learning system, integrating the main effective methodological approaches in an embedded Instructional Design System, guiding both the learners and trainers to the curriculum development and building of digital portfolio. The System, aimed at supporting trainers to contrast ESL and NEET phenomena, particularly in Vocational Education and Training (VET), is designed within a European project, named SAVE - Self Awareness, evaluation and motivation system Enhancing learning and integration and contrast ESL and NEET - 2014-1-IT01-KA202-002472 carried on by an European partnership (IT, ES, CZ, UK) under Program ERASMUS +. Then the article presents the result of the SAVE system piloting, engaging over 60 trainers and about 250 students in Europe, especially in Italy, UK and Spain; last discusses the further implications, challenges and effort required to prevent ESL and NEET.

BACKGROUND: MAIN DRIVERS AND POLICIES TO REDUCE THE PHENOMENON OF ESL AND NEETS

The heterogeneity of the NEET phenomenon is largely agreed, in terms of driving factors. Starting from the relation with ESL and, generally, with the educational level, other common causes emerging as critical refer to family and other social background factors and to school/work transition, Labour Market access, guidance system (Italia Lavoro, 2011; Eurofund, 2012; EU Commission, 2015 a). One of the main issues is the “lack of ownership” existing in the current setting, with many NEETs, particularly for 16/17 year olds, having little or no contact with institutions. This creates a serious risk of disconnection and by the time they enter the “formal” system, the damage may already be done.

A combination of personal, social, economic, educational and family-related factors, combined with certain features of education and training systems may aggravate the educational disadvantage, generating additional barriers for learners

who are struggling and delay their educational pathways. Creating an outcome related intervention with a focus on prevention and development of skills, rather than recovery, therefore appears essential. It appears even more essential when addressing the impacts of this “disconnection” on the motivational dimension, playing a central role, confirmed by the Eurofound report (2012) that distinguishes the NEETs population in subgroups in terms of engagement degree, from “disengaged” to “opportunity seekers”; and the last EU Report on dropout in Europe (2015 a), stressing the importance of the active engagement of youth and learning personalization as key factors. Researches show (Eurydice and Cedefop 2014) that systems characterized by grade retention, early insufficient support for learners, lack of quality vocational education and training (VET) [...] *The evidence suggests that individualising instruction and aid is likely to be more effective and resource efficient measure than making students repeat a year of study. A pedagogical practice that meets the needs of all students requires a flexible curriculum. [...] the evaluation of students' performance through the different combination of continuous formative assessment and external assessment is also crucial in this respects.* (Overcoming failure at school OECD 1998 p. 57).

According to the European DG Education and Culture (EU Commission, 2015), Policies to reduce ESL should be embedded in a whole school approach based on a stronger cooperation with a wide range of stakeholders (e.g. social services, outreach care workers, psychologists, guidance specialists, local authorities, business, unions, etc.); and in an overall inclusive learner-centred vision of education. That means to implement:

- A stimulating and conducive learning climate;
- Engaging curricula and effective teaching approaches;
- The learners' well-being;
- The learners' voice and participation in school life;

- Career education and guidance;
- Extra-curricular activities;
- Early detection mechanisms;
- A systemic support framework;
- Specific support for non-native speakers.

Furthermore, researches show that, in order to prevent ESL and disengagement, the curriculum should be linked to clear learning goals, and be connected to the next education (or professional) level. As much as possible, classroom activities related to the curriculum should develop student knowledge of real world problems (Dumont, Istance and Benavides, OECD 2010; OECD 2012). Students differ in ability, conceptions of learning, learning styles and strategies, interest, motivation, self-efficacy beliefs and emotion, as well as in their linguistic, cultural and social backgrounds. Students bring to the classroom different prior knowledge that substantially influences their learning process, and there is a constant and complex interaction between capacity and experience that shapes learning. That is why a set of guiding principles of learning environment are highlighted:

- **Learner-Centred:** Dependent on learner, not as an alternative to the critical role of teachers;
- **Personalised:** Sensitive to individual and group differences in background, prior knowledge, motivation and abilities, and offers tailored and detailed feedback.
- **Inclusive:** Able to valorise individual and group differences, including the weakest learners, in an educational agenda that excludes no-one.
- **Social:** Taking place in group settings, in a community that is itself the learning environment, inspired to collaborative strategy.

Summarising, NEETs population differentiations need to be addressed by policies, interventions and programs; at the same time, emerges a common and central role of motivational aspects.

This has represented (within a wider compared scenario carried on) the conceptual and operative

framework for the design and implementation of the SAVE methodological framework and technology enhanced system.

C

MAIN FOCUS OF THE ARTICLE: METHODOLOGICAL AND TECHNOLOGICAL FRAMEWORK TO CONTRAST AND PREVENT ESL AND NEET

The design of the SAVE methodological framework and digital system, takes move from the above mentioned background, addressing:

- Engagement and motivation of students by implementing effective self-awareness and evaluation tools and inclusive learning 2.0 solutions;
- Empowerment of the guidance role of trainers, by defining new pedagogical models, also technology enhanced, based on: self-evaluation; learning personalization and digital portfolio;
- Enhancement of the School and initial VET responsiveness, by leveraging on the competencies' recognition within the digital portfolio frame. The SAVE portfolio is a result of a self-reflection process driven in a motivational scale, helping to overcome the obstacles for children at risk of becoming NEET.

SAVE research confirmed the relevance of a twofold approach employing a multifactorial approach and valorising contemporarily: the active engagement of targets; the system capability in implementing learning and guidance personalised processes.

Furthermore, educators and employers agree that there is a growing gap between the Soft Skills that companies expect from their employees and those that many job candidates currently possess (OECD, 2015; CEDEFOP 2012; CEDEFOP 2008). With an increasingly competitive Job Market, it is imperative that job seekers are equipped

with a range of skills and competences that move beyond the technical knowledge and ability to perform specific tasks. These skills include emotional and social attributes allowing young people to demonstrate their motivation, self-confidence, ability to solve problems, initiative and communication/teamwork skills. In many cases young people are not aware of the soft skills they have, and even if they are aware, have difficulty in articulating them.

That is why a key goal of the SAVE project is to ensure that young people at risk of ESL and NEETs are provided with every opportunity to become, and to remain, actively engaged in learning in order to prevent exclusion from school and/or work. To this aim a range of interventions have been targeted at both:

- **Young People at Risk of Exclusion:** Improving their motivation and active engagement in learning and work by supporting their personal development, particularly in relation to their soft skills, self-esteem and by enhancing their employability;
- **Trainers:** Designing approaches, methods and resources supporting guidance and learning personalization and enhancing youth engagement and personal development.

Methodological Framework

Mistrust, the loss of self confidence, no perspective or vision about the future, no motivation, no interests, no expectations, no taking on responsibility, are the common feelings and attitudes of a NEET or ESL. That is why a learner-centered solution is necessary, involving all the dimension of the learner (Guspini 2009; Hoz 1982, 1997; Järvelä 2006, Meirieu 1992, 2004; Przesmycki 1991; Vettraino, Guglielma, Guspini; 2011), restarting from what the person is able to and from his/her potential. The SAVE methodological framework is based on strategies compliant with these requirements, aimed at the co-design of the

learning program and process (Przesmycki 1994), through a self-evaluation process (Guspini 2003; Varisco 2004; Davis 2001; De Bartolomeis 1974), of increasing complexity leading to the identification of learning challenges more than learning objectives, supported by digital portfolio's evidences (Cumming, Maxwell 1999; Wiggins 1993; Johnson, Rose 1997, Ajello, Belardi 2007).

The SAVE System integrates the main effective recalled methodological Learner-centred, Personalised approaches and Inclusive and Social learning environment, into an embedded Instructional Design System, guiding both the learners and trainers to the curriculum development. The system strategically combines from a methodological and technological point of view: motivational elements, evaluation strategies and digital portfolio resources to move along a self-realization scale where competences became visible.

Technological Framework

The SAVE system is supported by a customized Moodle platform. It embeds an instructional design model inspired to the ADDIE model (Analyse, Design, Develop, Implement, Evaluate), encouraging an active role of the students who go through the analysis phase, design develop and implementation, evaluation of the educational curriculum together with the guide of the trainer. So that another relevant implied technology is the students' mind (Hickman, 2000; Dewey 1910). Nevertheless, the digital support, the web services and the digital system guide the students and the trainers through the process of instructional design and curriculum development (Morrison, Gary, 2010), virtualizing this process, making it visible (Lévy, 1995) so that it can be retraced, appraised, improved, at any time. The SAVE portfolio mirrors this process, making visible all the accomplished tasks and progresses in real time. The SAVE System guides the climbing of the self-realization scale:

Step 1: I do not want

Step 2: I cannot

Step 3: I have not competences

Step 4: I will think about it

Step 5: I can

Step 6: I want

Step 7: Nothing can stop me

Step 8: I did it

Step 9: I did it better than I thought

A specific set of exercises corresponds at each step, presented by the following sections:

- **My Skills:** Contains resources and tools guiding the learner, from step 1 to 3, to acquire awareness about what s/he is able to. The learner collects all the information necessary to move to the next steps. (Analysis step of the Instructional design).
- **My Personal Plan:** Contains resources and tools guiding the learner, from step 4 to 6, to co-design with the coach a personal plan of development, oriented to the choice of an internship experience. (Design step of the Instructional design).
- **My Work Experience:** Provides the coach and the learner, from step 7 to 8, with tools and resources for collecting evidences of the work experience during the internship, to improve self-awareness of this experiences and give evidence to the related competences acquired and personal process; (Development/Implementation steps of the Instructional design).
- **My Portfolio:** Collects in real time the tangible results of the self-awareness rising process. It mirrors and live updates, step by step, the information recorded by the learner during each phase, leading to the step 9 (Evaluation step of the Instructional design).

The process is guided but it is not conditioned. The trainer provides suggestions and alerts; each page of the system shows information and instructions about steps progression, but the learner can move from a section to another. SAVE system structure integrates all the tools, frames and

supporting resources within a strategic, technology enhanced, learning personalization strategy. That is the characterizing element of SAVE: to be integrated in terms of actors, tools, strategies. Basically the trainers guide the learners in the effective use of the tools available in the System to improve their personal plan of development, to plan and realize an internship experience, to build and refine their portfolio. The trainers see exactly the same functions available for the learners, but having for each area of the system some guidelines on how to use each area, tool, resource. Moreover, trainers have control over the list of their learners and the correspondent process, in order to monitor and intervene when necessary with alerts, reminds, cockades.

In the section “My skills”, questionnaires and interviews, focused on what the students are able to, are the main support of the process. At the starting point questions offer closed options and descriptions: the students are asked to mirror with and choose the description that better fits with themselves. At the very beginning, indeed, students do not want, do not care, and do not willing (or are not able) to give deep personal descriptions. Time by time the interviews, encouraging and training self-reflection process, propose items of increasing openness and complexity. They move from closed questions with maximum three options, through a Self Awareness and Evaluation Diagnostic tool, to interviews based on biographical approach (Mezirow 2003; Demetrio, 2002).

For each questionnaire the coach can reinforce the self-awareness process by engaging the learner in simple tasks related to the information provided during the exercises. For example, if the learner states that s/he is able to cook a cake the trainer invites the learner to realize a cake and upload evidences of the artefact, reinforcing the accomplishment of the assigned task with virtual grades. The direct connection with a tangible result, with a real experience, more than with theories or contents, is a decisive variable in order to contrast the typical NEETs or ESLs attitudes and behaviours.

The Self Awareness and Evaluation Diagnostic (and its associated resources) is a core tool sup-

porting the learner to move along a self realization scale and enhancing trainers' supporting role. The questionnaire, focused on soft skills analysis, is designed to get young people to decide whether they agree or disagree to 48 statements about how they typically behave and like to work. Based on the range of potential answers to this questionnaire the Algorithm model for mixed variable programming was devised in order to produce individual reports. The questionnaire and its associated algorithms are loaded into the web-based SAVE Platform enabling learners to complete the questionnaire anywhere and at any time. It is loosely based on Marston's 4 Quadrant theory of Behaviour (4QB) (Marston, 1999). Designed to help young people become more 'self-aware', it uses the principles of personality profiling in order to analyse and feedback on the primary behavioural styles of the learners who have completed the assessment. The tool draws from a questionnaire and algorithms that have been particularly developed for use with young people who are at risk of becoming ESLs or NEETs.

The System seeks to identify an individual's primarily working style – how they prefer to work and interact – and to make young people more self-aware of these attributes. The outputs of the System suggest behavioural styles or preferences as well as strategies to make the most of strengths and cope with aspects of work, working relationships and learning that may not come easily to the learner. The outputs are designed to be positive and motivational. The benefits of this are twofold:

- As a result of taking the test and receiving professional feedback, the learners will have both greater self-awareness and increased confidence;
- For the trainers, the reports provide a useful insight into a learner's working strengths, style and potential thus providing an accurate and informed starting point for a Soft Skills development programme.

In order to use the Style Reports effectively, the trainer needs to recognise the primary behaviours

of each of the 4 styles identified by the diagnostic and understand how to work constructively with these when providing one to one feedback with a learner. To this extent, a Trainers Guide and associated resources includes: a description of the key characteristics of each of the 4 styles identified; guidance on how to explain the different styles to learners so that they become more Self Aware; how to use reports in order to draw up Individual Development Action Plans; a presentation for trainers to use with learners that explains the behavioural characteristics of the 4 styles, including their approach to learning and taking on new challenges; exercises and activities designed to support soft skill development.

The Self Awareness tool is integral part of the process leading to the portfolio development.

In the section "My personal plan", calendar functions, questionnaires and interviews are aimed at the co-design of a formative pact helping students to focus on their learning challenges and guiding to the choice of an internship experience.

In the section "My work experience", questionnaires and interviews guide and support learners to take notes and evidences on a virtual diary of their work experience, focusing on what they learnt and are able to.

The portfolio development is applied and integrated in this process, strictly connected with practical and real life, learning experience, or internship experience. All the proposed activities are finalized to support the choice of the apprenticeship experience or of a decision relevant for the learner. In SAVE the portfolio is then not a mere question of choosing format, tools or resources to fill in a folder. It is the arrival point of a progressive and recursive process of self awareness rising, aimed at moving from a condition of no motivation and mistrust toward a condition of gratification and motivation. At the very beginning the portfolio is merely a mirror of the learner profile. Step by step, the trainer supports the learner to improve the portfolio giving it an aspect more attained to the job perspective than to a mirror supporting the self-reflection. The trainer can suggest to re-phrase some sentences in a professional way.

The available qualification repertoires (i.e. European Qualification Framework) can be used to describe the acquired competences at the different level of expertise and mastery. SAVE portfolio starts as a self-reflection informal “mirror” and, step by step, ends in a more competence and job oriented working area that can be also valorised within external tool or networks (i.e. LinkedIn), or other social recruiting network and online job recruiting services.

EVIDENCES FROM THE SAVE SYSTEM APPLICATION

About 250 students at a risk of ESL, plus sixty teachers (engaged in Vocational Education and Training (VET), experienced the SAVE System, from June to October 2016. Validation activity with trainers, supporting an iterative process and the system on-going fine tuning has been carried out the year before.

Trainers and students, employing a multilevel and multidimensional assessment frame, have been asked to answer questions related to usability (and “utility”) main dimensions on all the areas, tools and resources of the system; trainers were also asked to observe the behaviour of the students using a common checklist. Indicators and descriptors of the checklist focused on behaviours of students showing: interest, concentration, completion of tasks. These behaviours indeed show antonyms attitudes compared with the usual approaches of NEETs or ESL. Furthermore, trainers were asked to add their interpretations and comments. A qualitative analysis of the collected remarks offered the following highlights.

The relationship between trainers and students experienced in the use of the SAVE System was mainly of one trainers to ten learners, from a minimum of one to three, and a maximum of one trainer to thirty learners. More than half of the trainers reported positive reactions of students showing interest, concentration, completion of tasks, together with an effective virtuous circle,

of increasing self confidence and awareness of areas of development by the students. trainers confirmed the effectiveness of the interview approach, focused on strengths, that gradually moves from multiple choice closed options, to opened biographic approach, as a powerful training of the self reflection attitude.

In order to enlarge, on a both qualitative and quantitative dimension, the feedbacks collected, the validation activities have been organised with different setting (from supporting the one-to one or small group guidance colloquium to the plenary class collective utilization and analysis of the system) and different target subgroups (in terms of age/class grade/curriculum).

A systematic observation would have required a longer elapse of time in a necessary longitudinal approach, for a more precise quantitative analysis. Nevertheless, the collected remarks, by a considerable sample of people, show a positive trend related to the system effectiveness. Almost the totality of answers related to the usability were positive, nevertheless students remarked that the graphic was not enough appealing and that they would have liked to have the system accessible and usable also by their mobile phone. Additionally, trainers and student highlighted the impact of the “reflective action” promoted by SAVE on the guidance and learning process and, more generally, on the personal development also in terms of “more aware” choices, improvement of the personal (and emotional) behaviour, enhancement of active behaviour. Finally, the first evidences from some focused validation action, also allowed to ground the valorisation of the improved awareness and transparency of competences (vertical but also behavioural) acquired, made visible by the digital portfolio.

FURTHER RESEARCH DIRECTIONS

The SAVE System, which encourages and motivates self awareness and self-evaluation is a valuable tool in helping young people to activate

a positive, aware and effective personal development (and employability) processes, supported by trainers within a more personalised guidance and learning systems. Coupled with the enhancement of active and personalised processes (at the centre of ESL&NEET prevention), the reflective and (self)evaluated processes promote the development (and awareness of the development) of soft skills, recognized as the core competences required to youths from companies and thus representing and essential elements of employability in addition to appropriate vocational.

However, these inputs represent only one piece, although central, of the puzzle. The question is, can schools, VET and colleges help provide young people with access to resources, on top of their educational attainment, supports more effective progression towards sustained employment? And if so, how can these organisations successfully provide resource to those young people whose personal circumstances make it hardest to access the experience, knowledge and contacts which might allow them to enter the labour market on a more even footing?

Work-based learning is a fundamental aspect of vocational training. It is directly linked to the mission of VET to help learners acquire knowledge, skills and competences which are essential in working life. The System, which supports the enhancement of the School and VET responsiveness, by leveraging on the competencies' recognition within the digital portfolio frame is a valuable tool in supporting trainers and employers within the management of the dual system application. To this extent, two evidences, among the others, can be quoted: some SAVE tools and resources will be valorised within the shared toolkit of the European Alliance for Apprenticeship; a dedicated toolkit for companies and job intermediates has been realised.

Creating opportunities for high-quality work-based learning as well as preventing and contrasting ESL and NEET are at the centre of the political and programmatic debates. The urgency of effective addressing both paths is referred to the high youth unemployment rate but also to the

vicious circle among ESL & NEET and its negative impacts: at the individual level, on personal development and social inclusion; at a system level, on the development of competitive, inclusive and innovative society.

CONCLUSION

There has been a broad range of interventions aimed in various different ways at tackling the issue of youth unemployment, ESL and NEETs. The most striking finding is how poor evaluations of these interventions are when it comes to attempting to judge successful programs. More needs to be done in this setting to ensure that we are able to identify which interventions are most effective in terms of outcomes and cost-effectiveness.

Consequently, SAVE open challenges and recommendations address both the policy and system level, the VET and job guidance system and the professional levels. Different considerations have been developed in each; the common frame is their integration. The most relevant effectiveness dimension is in fact the design and implementation of integrated policies and interventions, impacting contemporarily on youth motivation and evaluation and on system capacity in defining and managing personalized and engaging learning and guidance systems. Within this scenario and employing this approach, SAVE project has been designed and implemented. Actually the SAVE System is available by www.saveproject.it.

REFERENCES

- Ajello, A. M., & Belardi, C. (2007). *Valutare le competenze informali. Il portfolio digitale*. Roma: Carocci Faber.
- Cedefop. (2008). Skill shortages. In *Modernising vocational education and training: Fourth report on vocational education and training research in Europe: Background report*. Luxembourg: Publications Office.

- Cedefop. (2012). *Skill mismatch. The role of the enterprise*. Luxembourg: Publications Office of the European Union.
- Cumming, J. J., & Maxwell, G. S. (1999). Contextualizing Authentic Assessment. *Assessment in Education: Principles, Policy & Practice*, 6(2).
- Davis, A. (2001). *I limiti della valutazione educativa*. Roma: Anicia.
- De Bartolomeis, F. (1974). *Valutazione e orientamento*. Torino: Loescher.
- Demetrio, D. (2002). *Il metodo autobiografico*. Milano: Guerini e Associati.
- Dewey, J. (1910). *How we think*. Boston: D.C. Heath & Co. doi:10.1037/10903-000
- DNV GL AS. (2016). *Global Opportunity Report*. Oslo: Høvik.
- Dumont, H., Istance, D., & Benavides, F. (2010). *The Nature of Learning: Using research to inspire practice*. Paris: OECD.
- Eurofound. (2012). *NEETs – Young people not in employment, education or training: Characteristics, costs and policy responses in Europe*. Luxembourg: Publications Office of the European Union.
- European Commission. (2011). *Reducing early school leaving - Accompanying document to the Proposal for a Council Recommendation on policies to reduce early school leaving*. Author.
- European Commission. (2015). Schools policy A whole School approach to tackling early school leaving, DG Education and Culture. *Education + Training*, 2020.
- European Commission. (2015 a). *Report on Dropout and Completion in Higher Education in Europe*. Luxembourg: Publications Office of the European Union.
- European Commission EACEA. (2014). *Tackling Early Leaving from Education and Training in Europe: Strategies, Policies and Measures. Eurydice and Cedefop Report*. Luxembourg: Publications Office of the European Union.
- Eurostat. (2016). *Early leavers from education and training*. Author.
- Eurydice and Cedefop Report. (2014). *Tackling early leaving from education and training in Europe – Strategies, policies and measures*. Author.
- Guspini, M. (2003). @ dell'autovalutazione. In Learning Audit. Autovalutazione per l'istruzione e la formazione nell'era della conoscenza. Roma: Anicia.
- Guspini, M. (2009). Riflessioni attuali sulla personalizzazione degli apprendimenti. In P. Di Rienzo (Ed.), *Formazione degli Adulti e Università*. Roma: Anicia.
- Hickman, L. A. (2000). *La tecnologia pragmatica di Johnn Dewey*. Roma: Armando.
- Hoz Garcia, V. (1982). *L' educazione personalizzata. Individualizzazione e socializzazione nell'insegnamento*. Firenze: Le Monnier.
- Hoz Garzia, V. (1997). *Dal fine agli obiettivi dell'educazione personalizzata*. Palermo: Palumbo.
- Järvelä, S. (2006). L'apprentissage personnalisé? Comment stimuler la capacité d'apprendre, nouveaux éclairages. In *Personnaliser l'enseignement*. Paris: OCDE. doi:10.1787/9789264036628-3-fr
- Johnson, N. J., & Rose, L. M. (1997). *Portfolio*. Lancaster, PA: Technomic.
- Lévy, P. (1995). *Qu'est-ce que le virtuel?* Paris: La Découverte.
- Marston, W. M. (1999). *Emotions of Normal People*. Taylor & Francis Ltd.

- Meirieu P. (1991), *Individualisation, différenciation, personnalisation: De l'exploration d'un champ sémantique aux paradoxes de la formation*. Association des enseignants et chercheurs en sciences de l'éducation.
- Meirieu, P. (2004). *L'école, mode d'emploi: Des "méthodes actives" à la pédagogie différenciée*. Issy-les-Moulineaux: ESF.
- Mezirow, J. (2003). *Apprendimento e trasformazione. Il significato dell'esperienza e il valore della riflessione nell'apprendimento degli adulti*. Milano: Cortina.
- Morrison, G. R. (2010). *Designing Effective Instruction* (6th ed.). John Wiley & Sons.
- OECD. (2012). *Equity and Quality in Education*. OECD.
- OECD. (2015a). *Skills Outlook 2015: Youth, Skills and Employability. Trends in improving young people's education and skills*. OECD.
- OECD. (2015b). *Education at a Glance. Interim Report*. OECD.
- OECD. (2016). *Society at a Glance*. OECD.
- Przesmycki, H. (1991). *Pédagogie différenciée*. Broché, Editions HACHETTE Education.
- Przesmycki, H. (1994). *La pédagogie de contract*. Paris: Hachette Livre.
- Varisco, B. M. (2004). *Portfolio. Valutare gli apprendimenti e le competenze*. Roma: Carocci.
- Vettraino, L., Guglielman, E., & Guspini, M. (2011). *European Model of Personalization for Adult and Elderly Learners*. Roma: Anicia.
- Vygotskij, L. (1990). *Pensiero e linguaggio*. Roma, Bari: Laterza.
- Wiggins, G. P. (1993). *Assessing student performance*. San Francisco: Jossey Bass.

KEY TERMS AND DEFINITIONS

Learning Personalization: A learner centred approach aimed at involving all the dimensions of the learner (personal, emotional, vocational, etc.) encouraging the expression and development of personal attitudes and talents, in order to valorise the potential that each person can reach and express.

NEET: It is an acronym standing for “Not in Education, Employment or Training”, that in accordance with the International literature identifies people aged 15 to 29 demotivated to re-enter the Educational or Labour Market System.

Self Evaluation: A recursive process of self reflection, aimed at increasing awareness of strength points and areas of further development.

Category C

Customer Relationship Management

Analysis of Two Phases Queue With Vacations and Breakdowns Under T-Policy

Khalid Alnowibet

King Saud University, Saudi Arabia

Lotfi Tadj

Fairleigh Dickinson University, Canada

INTRODUCTION

Waiting in line is an experience that everyone practices, almost, on a daily basis. The waiting takes different forms and settings. Nowadays, the competition of customer satisfaction and low cost becomes very intense to the point that any customer that waits too long in line is potentially a lost customer to another compactor that provides better service or better waiting environment.

Waiting lines are basic in structure to the external (customer-facing) and the internal business processes. Queueing structures generally include staffing, scheduling and inventory levels. For this reason, businesses often utilize queueing theory as a competitive advantage. Although queueing is undesired for anyone, it is the cornerstone of efficiency and organization for many companies. The idea is simple: At any given moment, there can be more people or cases needing service, help or attention than an organization can handle. Queues help workers and managers track, prioritize and ensure the delivery of services and transactions.

The theory of waiting lines provides insight and identifies management options for improving customer service. A wide variety of queueing models have been developed and successfully exploited for very complex service situations. This chapter describes one such queueing model. For a comprehensive classification of various control policies applied in queueing systems, see the survey by Tadj and Choudhury (2005).

The service system considered in this chapter is characterized by an unreliable server. Random breakdowns occur on the server and the repair may not be immediate. It is assumed that at the end of a given service, the server may either take a vacation or start serving the next customer. When the queue is empty, the server again takes vacations and scans the queue periodically, every T units of time, to check if some customers have arrived while he was away. The third assumption is that the actual service of any arrival takes place in two consecutive phases. Both service phases are independent of each other.

BACKGROUND

The unreliability of the server is one of the main features of the queueing system studied in this chapter. Queueing systems prone to failure are commonly encountered in the real world. The server breakdown was first analyzed by White and Christie (1958). Since then, queueing systems with unreliable servers have been extensively studied by many researchers; see Tadj et al. (2012) for a comprehensive survey on the subject.

The next feature of the system of interest in this chapter is the Bernoulli vacation schedule. The classical vacation scheme with Bernoulli service discipline was introduced and developed by Keilson and Servi (1986). Various aspects of Bernoulli vacation models have been discussed

by a number of authors; see the survey of Ke et al. (2010).

The other important feature considered in our service system is the server T -policy. The M/G/1 queue with a T -policy was first studied by Heyman (1977). Many variants of the T -policy discipline model have been considered in the literature since then. There is no recent survey on the subject of T -policy discipline; however, some details of the latest contributions are listed here. Wang and Ke (2002) consider a single non-reliable server in the ordinary M/G/1 queueing system operating under the N -policy, the T -policy and the $\text{Min}(N, T)$ -policy. They show that the optimal N -policy and the optimal $\text{Min}(N, T)$ -policy are always superior to the optimal T -policy. Tadj (2003) studies an M/G/1 quorum queueing system under T -policy. The quorum or q -policy means that the server does not start service unless a specified number of customers are in the queue, and service is always rendered to groups of a fixed size. Ke (2005) studies an M/G/1 queueing system with an unreliable server, startup, and the following modified T -policy: After all the customers are served in the queue exhaustively, the server deactivates and takes at most J vacations of constant time length T repeatedly until at least one customer is found waiting in the queue upon returning from a vacation. If no customers arrive by the end of the J^{th} vacation, the server remains dormant in the system until at least one customer arrives. This model is generalized by Ke (2008) to the case of compound Poisson arrival process. More complex scenarios for the server are considered by Ke (2006). Kim and Moon (2006) study an M/G/1 queueing system where the server can take a vacation time T after the system becomes empty. At that time, the server is switched off with probability p and takes a vacation or remains on serving the arriving customers with probability $(1-p)$. Wang et al. (2009a) investigate the T -policy M/G/1 queue with server breakdowns and startup times. The server is turned on after a fixed length of time T repeatedly until at least one customer is present in the waiting line. The

same model is studied by Wang et al. (2009b) who use the maximum entropy approach to solve for the steady-state probabilities. Zhang et al. (2011) clarify the concept of regeneration cycle used in evaluating the average operating cost of the M/G/1 queue with T -policy. Two ways of defining the regeneration cycle are compared and advantages and disadvantages of each are pointed out.

Randomized control policies, with random control of the server at the beginning of the service when at least one customer appears, have also been combined recently with the T -policy. Yang et al. (2008) study the randomized T -policy in an unreliable M/G/1 queueing system with second optional service to the customers and server startup. Ke and Chu (2008) compare the randomized T -policy and the randomized N -policy for an M/G/1 queueing system with second optional service. Ke and Chu (2009) deal with a variation of the randomized T -policy of Ke and Chu (2008). Wang et al. (2010) compare the randomized T -policy and the randomized N -policy for an M/G/1 queueing system with second optional service and server startup. They show that the optimal randomized N -policy outperforms the optimal randomized T -policy. Kuo et al. (2015) obtain the optimal N - T threshold for a two-phase service M/G/1 system with vacation and random failure of service. In their system, they consider the case where the server starts service if queue length reaches threshold value N , or when the waiting time of the leading customers reaches time T . Also, Wua et al. (2014) study an M/G/1 queue with vacation under N -policy and a single phase of service. Where server is subject to breakdowns and the repair facility may fail during the repair period.

Recently there have been several contributions considering queueing systems of M/G/1 type in which the server may provide a second phase of service. The case where both phases of services are exponentially distributed is the so called Coxian distribution C_2 . Bertsimas and Papaconstantinou (1988) consider such distribution to design a multi-server queue with application in a transportation system. Madan (2000) studies an M/G/1 queue

where the server provides a first phase of regular service to all the arriving customers, whereas only some of them receive a second phase of optional service. The first regular service time follows a general distribution but the second optional service is assumed to be exponentially distributed. Some examples of queueing situations where such service mechanism can arise are also given. Medhi (2002) generalizes the model by considering the second optional service time is also governed by a general distribution. Recently, Choudhury (2003) investigates this queueing model in more depth and generalizes the results for a batch arrival queueing system with more than one optional service in (2003). Choudhury and Paul (2005) extend the two phase service model considering feedback. Krishnakumar et al (2002) consider an M/G/1 retrial queue with additional phase of service. The motivation for this type of model comes from some computer and communication networks where messages are processed in two stages by a single server. Further, Choudhury et al (2007) are able to develop the steady state distribution of the two phase service system with bulk arrivals and scheduled vacation.

MAIN FOCUS OF THE CHAPTER

The primary purpose of waiting lines theory is to obtain the performance measures of the service system. We thus obtain probability generating functions of the system state and the main performance measures such as the mean number of customers in the system, and the mean lengths of the idle period, busy period, and busy cycle. These measures are then used to design an optimal management policy of the system, by balancing various types of costs. We illustrate the results obtained by presenting a numerical example and perform some sensitivity analysis to show the effect of the system parameters on the optimal policy. The method of Zhang et al. (2011) is applied to obtain the system performance measures. These

will be used to find the vacation length that will optimize the system performance.

Model Description and Notation

The queueing system considered in this chapter is characterized by a single server and a Poisson arrival process with positive rate λ . It is assumed that the waiting room can accommodate an infinite number of customers. The server provides the service on a first-in, first-out discipline (FIFO). In addition, the actual service process takes place in two consecutive phases, a first phase of service (FPS) followed by a second phase of service (SPS). Let $i = 1$ correspond to FPS and $i = 2$ correspond to SPS. Let B_i denote the duration of phase i of service. Then, B_i is distributed according to the cumulative distribution function (CDF) denoted by $B_i(t)$ with Laplace-Stieltjes (LST) denoted by $B_i^*(\theta)$ and moments $b_i^{(j)}, j \geq 1$ and $i = 1, 2$.

During the actual service, the server may break down in any of the phases or in both. It is assumed that breakdowns occur independently in each phase according to a Poisson process with positive rate α_i ($i = 1, 2$). When a service interruption occurs, repair is delayed so that a delay period precedes the repair period. Let D_i denote the duration of the delay time during phase ($i = 1, 2$). Then, D_i is distributed according to the CDF given by $D_i(t)$ with LST given by $D_i^*(\theta)$ and moments $d_i^{(j)}, j \geq 1$ and $i = 1, 2$.

As soon as the server of phase i ($i = 1, 2$) is repaired, it immediately returns to provide his service until the queue becomes empty. Service time is the cumulative duration of actual service times, delays, and repair times. Thus, let R_i denote the duration of the repair time during phase i ($i = 1, 2$) of service. Then, R_i is distributed according to the CDF $R_i(t)$ with LST $R_i^*(\theta)$ and moments $r_i^{(j)}, j \geq 1$ and $i = 1, 2$.

Now, let H_i denote the duration of the modified service time of phase $(i = 1, 2)$. Then, H_i is distributed according to the CDF denoted by $H_i(t)$ with LST given by $H_i^*(\theta)$ and moments $h_i^{(j)}, j \geq 1$. Thus,

$$H_i^*(\theta) = B_i^* \left(\theta + \alpha_i \left(1 - D_i^*(\theta) R_i^*(\theta) \right) \right) \quad (1)$$

From Equation (1), the first two moments are defined as

$$\begin{aligned} h_i^{(1)} &= b_i^{(1)} \left[1 + \alpha_i \left(d_i^{(1)} + r_i^{(1)} \right) \right], \\ h_i^{(2)} &= b_i^{(2)} \left[1 + \alpha_i \left(d_i^{(1)} + r_i^{(1)} \right) \right]^2 + \alpha_i b_i^{(1)} \left(d_i^{(2)} + 2d_i^{(1)} r_i^{(1)} + r_i^{(2)} \right) \end{aligned} \quad (2)$$

Note that, in order to derive relation (1), the random variables B_i, D_i and $R_i (i=1, 2)$ are assumed to be independent of each other. This relation also allows determining the moments of the generalized service times in terms of the system parameters. Therefore, let H denote the duration of the generalized service time of a customer. Then $H = H_1 + H_2$ is distributed according to the CDF $H(t)$ with LST $H^*(\theta)$ and moments $h^{(j)}, j \geq 1$, so that

$$H^*(\theta) = H_1^*(\theta) H_2^*(\theta), \quad (3)$$

and

$$\begin{aligned} h^{(1)} &= h_1^{(1)} + h_2^{(1)}, \\ h^{(2)} &= h_1^{(2)} + 2h_1^{(1)} h_2^{(1)} + h_2^{(2)}. \end{aligned} \quad (4)$$

It is also assumed in this system that the server implements the Bernoulli vacation schedule, so that at the end of a service, the server may take a vacation of length T with some positive probability p , or may remain working with the complementary probability $q = 1 - p$. Let G denote the duration of the required service time

of a customer, which includes the generalized service time and a possible vacation. Then, G is distributed according to the CDF $G(t)$ with LST $G^*(\theta)$, and moments $g^{(j)}, j \geq 1$. Since

$$G = \begin{cases} H + T, & \text{with probability } p, \\ H, & \text{with probability } q, \end{cases}$$

we have

$$G^*(\theta) = (q + p e^{-\theta T}) H^*(\theta), \quad (5)$$

and

$$\begin{aligned} g^{(1)} &= h^{(1)} + pT, \\ g^{(2)} &= h^{(2)} + 2h^{(1)} pT + pT^2. \end{aligned} \quad (6)$$

The final main characteristics of the system is that the server takes a vacation if a customer departure leaves the system empty, in accordance with the T -policy. The server scans the queue T time units after the end of the last busy period (the beginning of the idle period). When the server finds at least one customer waiting in the queue, a busy period starts until the system becomes empty. Otherwise, if the server finds no customers waiting in the queue, another vacation starts, and a new scan is made after T units of time, and so on.

Additional Queue Size Distribution Due to the Idle Period

Suppose that after $\nu - 1$ vacations, the server finds no customer in the queue and after ν vacations, he finds Y_ν customers in the queue. Let $q(z) = e^{-\lambda T(1-z)}$ denote the probability generating function (PGF) of a Poisson random variable. Then, the probability mass functions of Y_ν and ν can be easily derived either by direct probability arguments or using fluctuation theory. Since

$P(Y = 0) = q(0)$ and $P(Y > 0) = 1 - q(0)$, we have

$$P(\nu = j) = [1 - q(0)]^{j-1} q(0), j \geq 1. \quad (7)$$

$$E[z^\nu] = \frac{z[1 - q(0)]}{1 - zq(0)}. \quad (8)$$

$$E[\nu] = \frac{1}{1 - e^{-\lambda T}}. \quad (9)$$

To obtain the distribution of Y_ν , let $k_j = e^{-\lambda T} \frac{(\lambda T)^j}{j!}$ denote the probability that j units arrive during a constant length of vacation time T . So,

$$P(Y_\nu = j) = \frac{k_j}{1 - q(0)}, j \geq 1. \quad (10)$$

$$E[z^{Y_\nu}] = \frac{q(z) - q(0)}{1 - q(0)}. \quad (11)$$

$$E[Y_\nu] = \lambda T E[\nu]. \quad (12)$$

The mean idle period is obtained, when the length of the idle period I , by utilizing Little's Law in expression (12), we get the expected length of the idle period as

$$I^{(1)} = \frac{\left. \frac{dE[z^{Y_\nu}]}{dz} \right|_{z=1}}{\lambda} = \frac{T}{1 - e^{-\lambda T}}. \quad (13)$$

Applying the discrete renewal theory the results (see Takagi (1990)), we can obtain the PGF of the number of arrivals during the residual life of an idle period as:

$$\psi(z) = \frac{1 - E[z^{Y_\nu}]}{\left. \frac{dE[z^{Y_\nu}]}{dz} \right|_{z=1} (1 - z)} = \frac{1 - e^{-\lambda T(1-z)}}{\lambda T(1 - z)}, \quad (14)$$

which is the PGF of the additional queue size distribution due the idle period for the T-policy model.

Queue Size Distribution at a Departure Epoch

Now, let $Q(t)$ denote the number of customers in the system at an arbitrary instant of time t and let Q_n be the number of customers in the system at the completion epoch T_n of the n^{th} service, that is, $Q_n = Q(T_n^+)$. Also, let V_{n+1} represent the number of customer arrivals during the $(n+1)^{\text{st}}$ required service. Then, Q_n satisfies the following recursion relation

$$Q_{n+1} = \begin{cases} Y_\nu + V_{n+1} - 1, & Q_n = 0, \\ Q_n + V_{n+1} - 1, & Q_n \geq 1. \end{cases} \quad (15)$$

Therefore, the process $\{Q_n; n = 0, 1, \dots\}$ is a Markov chain. Then, the i^{th} row of the transition probability matrix (TPM) A of the Markov chain $\{Q_n; n = 0, 1, \dots\}$ has PGF:

$$A_i(z) = \begin{cases} z^{-1} E[z^{Y_\nu}] E[z^{V_{n+1}}], & i = 0, \\ z^{i-1} G^*(\lambda - \lambda z), & i \geq 1, \end{cases} \quad (16)$$

where

$$E[z^{V_{n+1}}] = G^*(\lambda - \lambda z).$$

For the ergodicity of $\{Q_n\}$, Abolnikov and Dukhovny (1991) have shown that a Markov chain, whose TPM is a Δ -matrix, is ergodic if and only if

$$\left. \frac{dG^*(\lambda - \lambda z)}{dz} \right|_{z=0} < 1. \quad (17)$$

Since the TPM A is a Δ -matrix, it satisfies that:

$$\rho = \lambda g^{(1)} < 1. \quad (18)$$

Therefore, in the rest of the chapter, it is assumed that condition (18) is satisfied so that the Markov chain $\{Q_n; n = 0, 1, \dots\}$ is ergodic and a steady-state distribution for the number of customers in the system exists.

Now, denote by $P(z) = \sum_{i=0}^{\infty} p_i z^i$ the probability generating function of Q_n and by $p_i = \lim_{n \rightarrow \infty} P(Q_n = i)$ the steady-state probability of the number of customers present in the system.

Since $P(z) = \sum_{i=0}^{\infty} A_i(z) p_i$ then we have

$$P(z) = \frac{G^*(\lambda - \lambda z)(E[z^{Y_\nu}] - 1)p_0}{z - G^*(\lambda - \lambda z)}. \quad (19)$$

The summability-to-one condition $P(1) = 1$ yields

$$p_0 = \frac{1 - \rho}{E[Y_\nu]}, \quad (20)$$

where $E[Y_\nu]$ is given by (12). Combining equations (19) and (20), the steady-state probabilities for number of customers present in the system is defined for all n .

Queue Size Distribution at an Arbitrary Point in Time

The process $\{Q(t); t \geq 0\}$ is a semi-regenerative process relative to the sequence $\{T_n; n = 0, 1, \dots\}$ of service completion epochs. It is ergodic if and only if $\rho < 1$. Denote by $\pi(z) = \sum_{i=0}^{\infty} \pi_i z^i$ the probability generating function PGF in the steady-state of $Q(t)$ and by $\pi_i = \lim_{t \rightarrow \infty} P(Q(t) = i)$ the steady-state distribution. It can be shown using the main convergence theorem for semi-regenerative processes, see e.g., Çınlar (1975), that this PGF $\pi(z)$ is given by

$$\pi(z) = P(z). \quad (21)$$

Idle Period Distribution

Let I and $I^*(\theta)$ be the idle period random variable and its LST, respectively. Then, from the theory of fluctuating processes, it can be shown that

$$I^*(\theta) = \frac{e^{-\theta T}(1 - e^{-\lambda T})}{1 - e^{-(\theta + \lambda)T}}. \quad (22)$$

Delayed Busy Period Distribution

Let B and $B^*(\theta)$ be the busy period random variable and its LST, respectively. Then, from expression (11), we have

$$B^*(\theta) = \frac{e^{-\lambda T[1-\varpi^*(\theta)]} - e^{-\lambda T}}{1 - e^{-\lambda T}}, \quad (23)$$

where $\varpi^*(\theta)$ is the LST of the busy period of an ordinary unreliable M/G/1 queue under Bernoulli vacation schedule satisfying Takác's functional equation

$$\varpi^*(\theta) = (q + pe^{-\theta T}) \prod_{i=1}^2 H_i^*(\theta + \lambda - \lambda \varpi^*(\theta)). \quad (24)$$

System Performance Measures

One of our objectives of this analysis is to define the key important performance measures of the system in steady-state based on our findings. In this section we introduce the mathematical expressions for selected performance measures.

1. **Expected Number of Customers in the System:** This measure is found by:

$$L = \rho + \frac{\lambda T}{2} + \frac{\lambda^2 g_2}{2(1-\rho)}. \quad (25)$$

2. **Mean Idle Period:** The average time that a server remains idle is the time until the total workload accumulates to exceed the value T . Thus, the mean idle period, denoted by $I^{(1)}$ is equal to the mean first excess time. Then,

$$I^{(1)} = \frac{T}{1 - e^{-\lambda T}}. \quad (26)$$

3. **Mean Delayed Busy Period:** The busy period is the interval of time that keeps the server busy and this goes on until the first subsequent instant when the system becomes empty. Using equation (23), the mean delayed busy period is found to be

$$B^{(1)} = \frac{\lambda(h^{(1)} + pT)}{1 - \lambda h^{(1)}} \frac{T}{1 - e^{-\lambda T}}. \quad (27)$$

4. **Mean Busy Cycle:** The busy cycle is the difference between the starting times of two consecutive idle periods, $C^{(1)} = I^{(1)} + B^{(1)}$. So, the mean busy cycle is obtained by:

$$C^{(1)} = \frac{1 + \lambda p T}{1 - \lambda h^{(1)}} \frac{T}{1 - e^{-\lambda T}}. \quad (28)$$

5. **Probabilities of Idle and Busy Server:** Since the server takes vacations, the server may remain idle even with customers present in the system. Thus, the probability of idle server, denoted as π_0 , is given by:

$$P_I = \frac{I^{(1)}}{C^{(1)}} = 1 - \frac{\lambda(h^{(1)} + pT)}{1 - \lambda h^{(1)}}, \quad (29)$$

$$P_B = \frac{B^{(1)}}{C^{(1)}} = \frac{\lambda(h^{(1)} + pT)}{1 - \lambda h^{(1)}}.$$

6. **System Intensity:** The system intensity is defined by $\mathcal{I} = \sum P\beta$ where $P = (p_0, p_1, \dots)$ and $\beta = (\beta_0, \beta_1, \dots)$ with $\beta = E[T_{n+1} - T_n | Q_n = i]$. Since

$$\beta_i = \begin{cases} TE[\nu] + g_1, & i = 0, \\ g_1, & i \geq 1, \end{cases}$$

then we have

$$\mathcal{I} = \frac{\lambda T p_0}{1 - e^{-\lambda T}} + \lambda g_1. \quad (30)$$

Now using the expression found for p_0 in (20), we see that the system intensity is equal to the server load 1, as in the M/G/1 queueing system.

The quantity $P\beta$ is called the mean service cycle, and is equal to $1/\lambda$ both in the M/G/1 and in our model. This quantity needs to be finite and is essential when applying the main convergence theorem for semi-regenerative processes cited above.

Optimal Management Policy

Having obtained the required performance measures, we now are ready to develop a management policy for the system. We are interested in obtaining the optimal value of the length T of the vacation period. Denote by $TC(T)$ the total expected cost per unit of time where T is the decision variable. Then,

$$TC(T) = c_h L + c_o \frac{B^{(1)}}{C^{(1)}} + c_a \frac{I^{(1)}}{C^{(1)}} + c_s \frac{1}{C^{(1)}}, \quad (31)$$

where c_h is the holding cost per unit for each customer present in the system, c_o is the cost per unit time for keeping the server on and in operations, c_a is the startup cost per unit time for the preparatory work of the server before starting service, and c_s is the setup cost per busy cycle. We illustrate in the next section by some numerical results. Using the previous expressions, we have

$$TC(T) = c_h \left[\rho + \frac{\lambda T}{2} + \frac{\lambda^2 g_2}{2(1-\rho)} \right] + c_o \frac{\lambda(h^{(1)} + pT)}{1 - \lambda h^{(1)}} + c_a \left[1 - \frac{\lambda(h^{(1)} + pT)}{1 - \lambda h^{(1)}} \right] + c_s \frac{1 - \lambda h^{(1)}}{1 + \lambda pT} \frac{1 - e^{-\lambda T}}{T}. \quad (32)$$

The optimal value T^* is found by solving

$$\frac{dTC(T)}{dT} = 0. \quad (33)$$

The nonlinear expression (33) cannot be solved analytically. However, it can be solved numerically using some mathematical software such as MATLAB. Also, we don't compute the second derivative to show that the cost function is convex, although extensive numerical experiments suggest that it is.

Numerical Example

To illustrate the results obtained, assume the service times, the repair times, and the delay times all follow the exponential distribution. Therefore, the LST of the service time, the delay time and the repair time for each phase are given by $B_i^*(s)$, $D_i^*(s)$ and $R_i^*(s)$, ($i = 1, 2$), respectively which are defined as:

$$B_i^*(s) = \frac{1}{1 + sb_i^{(1)}}, D_i^*(s) = \frac{1}{1 + sd_i^{(1)}}, R_i^*(s) = \frac{1}{1 + sr_i^{(1)}}.$$

From expression (1) defined earlier, we get the LST for the modified service time given by:

$$H_i^*(\theta) = \left[1 + \left(\theta + \alpha_i - \alpha_i \left(\frac{1}{1 + \theta d_i^{(1)}} \right) \left(\frac{1}{1 + \theta r_i^{(1)}} \right) \right) b_i^{(1)} \right]^{-1}$$

It is assumed in this example that the arrival rate is $\lambda = 0.4$, the Bernoulli schedule probability is $p = 0.1$, the first phase breakdown rate $\alpha_1 = 0.01$ and the second phase breakdown rate $\alpha_2 = 0.02$. The moment's values of for each service phase are shown in Table 1. The monetary parameters were taken as follows:

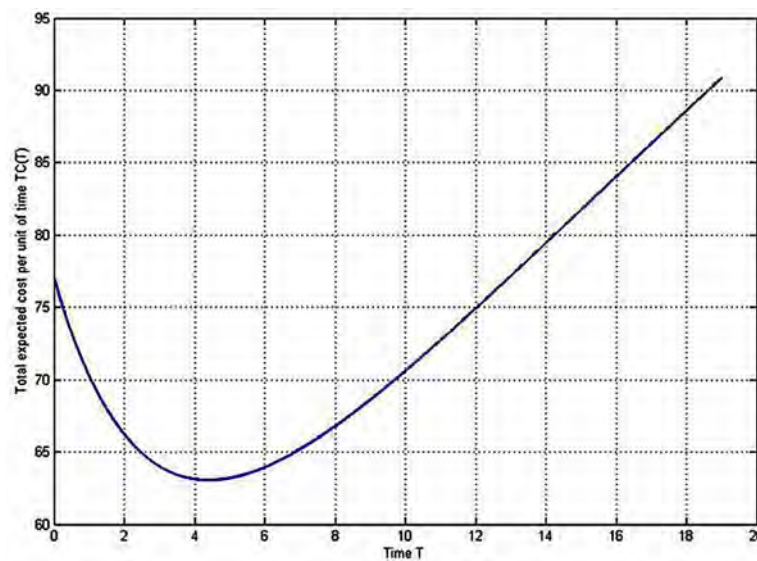
$$c_h = 5, c_o = 20, c_a = 100, c_s = 200.$$

Figure 1 shows the variations of the expected total cost per unit of time. Clearly, the total cost function $TC(T)$ is a convex function of T . This means that $TC(T)$ has a global unique optimal

Table 1. Nonmonetary data

FPS Service Parameters	Value	FPS Service Parameters	Value
FPS Mean service time	$b_1^{(1)} = 0.05$	SPS Mean service time	$b_2^{(1)} = b_1^{(1)} / 2$
FPS Mean delay time	$d_1^{(1)} = b_1^{(1)} / 20$	SPS Mean delay time	$d_2^{(1)} = b_2^{(1)} / 20$
FPS Mean repair time	$r_1^{(1)} = b_1^{(1)} / 10$	SPS Mean repair time	$r_2^{(1)} = b_2^{(1)} / 10$

Figure 1. Variations of total expected cost per unit of time



value of T that minimizes the cost function. In this case, the optimal length of a vacation period is $T^* = 5.39$, which corresponds to an optimal cost per unit of time $TC(T^*) = 63.0476$.

Sensitivity Analysis

A sensitivity analysis is conducted to assess the effect of the main system parameters on the optimal policy. The first sensitivity analysis is conducted by varying the arrival rate λ from 0.125 arrival per unit of time to 0.375 and fixing all other values of the model. The second analysis is done by varying the values of the Bernoulli schedule probability p from 0.01 to 0.1. Table 2 shows that the optimal cost per unit of time decreases as p increases.

The analysis shows that the optimal value of T increases as the arrival rate increases. The optimal value of T converges to some value as the arrival rate takes high values. Details of the analysis are listed in Table 2 and Figure 2. However, the value of the optimal cost is a concave function with respect to the increase of the arrival rate (see Figure 2). In fact, the main purpose of applying control policies is to choose the best value of T that reduces the optimal cost. In the absence of the control policy it is expected that the total cost function strictly increases as the arrival rate increases.

With respect to the Bernoulli schedule probability, as the probability of taking a vacation increases the optimal T decreases. Similar to the effect of increasing the arrival rate, the optimal

Table 2. Effect of arrival rate on optimal cost and optimal policy

λ	0.125	0.150	0.175	0.200	0.225	0.250	0.275	0.300	0.325	0.350	0.375
$TC(T^*)$	26.61	31.52	35.85	39.73	43.28	46.58	49.69	52.63	55.43	58.11	60.65
T^*	1.00	2.04	3.28	4.06	4.57	4.89	5.11	5.24	5.33	5.37	5.39

Figure 2. Effect of arrival rate on optimal policy and optimal cost

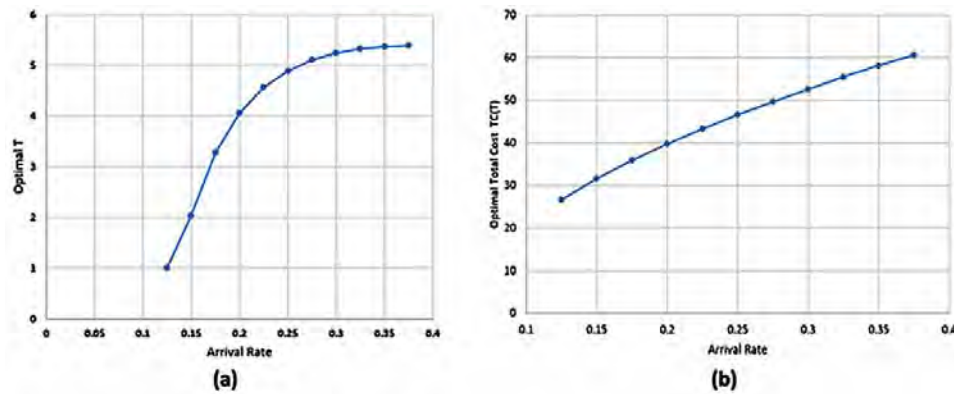
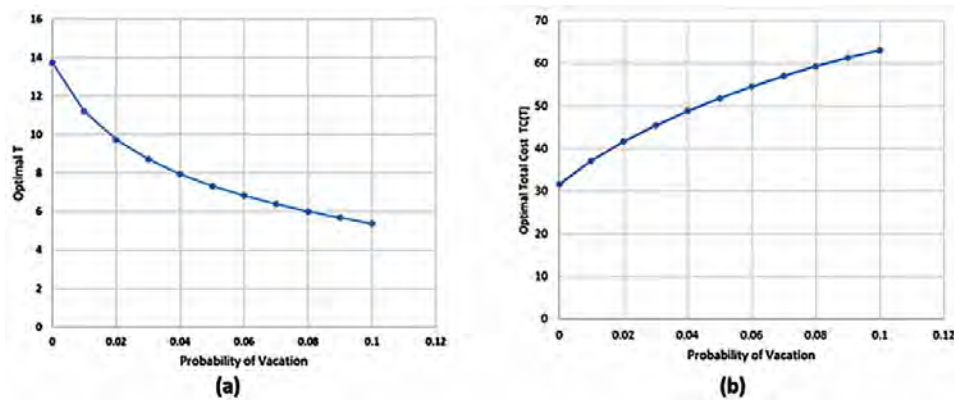


Table 3. Effect of Bernoulli schedule probability on optimal cost and optimal policy

p	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
$TC(T^*)$	31.5553	37.1595	41.6272	45.4265	48.772	51.7796	54.5146	57.0101	59.2755	61.2988	63.0476
T^*	13.74	11.22	9.74	8.72	7.95	7.33	6.82	6.39	6.01	5.68	5.39

Figure 3. Effect of Bernoulli schedule probability on optimal policy and optimal cost



policy value T seems to converge to some optimal value as the probability of vacation increases for higher values. (See Figure 3).

Figure 4 shows that the optimal total cost decreases linearly as the average service time for the first phase increases. The last analysis listed

in Table 5 represents the effect of changing breakdown rates for both service phases on optimal cost and optimal policy. Although, there is a relatively wide range of change in the breakdown rates the change in the total cost remains very limited. In addition, the optimal policy T^* remains

Table 4. Effect of average first phase service time on optimal cost and optimal policy

$b_1^{(1)}$	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
$TC(T^*)$	61.0713	61.5654	62.0594	62.5535	63.0476	63.5416	64.0355	64.5292	65.0228	65.5162
T^*	5.42	5.41	5.40	5.39	5.39	5.38	5.37	5.36	5.35	5.34

Figure 4. Effect of average first phase service time on optimal policy and optimal cost

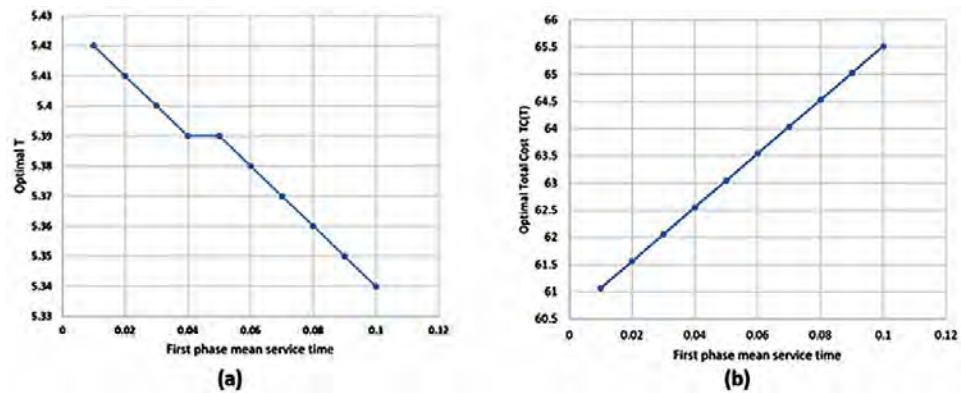


Table 5. Effect of breakdown rates on optimal cost and optimal policy

		α_2 The Breakdown Rate of Second Phase Service								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
α_1 The Breakdown Rate of First Phase Service	0.1	63.049	63.049	63.050	63.050	63.050	63.051	63.051	63.051	63.051
	0.2	63.050	63.051	63.051	63.051	63.051	63.052	63.052	63.052	63.053
	0.3	63.051	63.052	63.052	63.052	63.053	63.053	63.053	63.054	63.054
	0.4	63.053	63.053	63.053	63.054	63.054	63.054	63.055	63.055	63.055
	0.5	63.054	63.054	63.055	63.055	63.055	63.055	63.056	63.056	63.056
	0.6	63.055	63.055	63.056	63.056	63.056	63.057	63.057	63.057	63.058
	0.7	63.056	63.057	63.057	63.057	63.058	63.058	63.058	63.059	63.059
	0.8	63.058	63.058	63.058	63.059	63.059	63.059	63.059	63.060	63.060
	0.9	63.059	63.059	63.059	63.060	63.060	63.060	63.061	63.061	63.061

the same for all cases ($T^* = 5.39$). This means that the effect of breakdown rate of either phase is very minimal on the performance of the entire system.

FUTURE RESEARCH DIRECTIONS

The theory of queues, as a part of business analytics, uses past business performance to drive business planning. From past data, system parameters such as arrival rate and service rate can be estimated. Queueing models are then applied to obtain performance measures, thus gaining insight into the current state of the business. As shown above, these performance measures can then be used to devise an optimal management policy for the firm. It is through the use of analytics that the firm gains a competitive edge. A large number of queueing models is available. However, the development of more models is imperative since they provide the analyst with a powerful tool for designing and evaluating the performance of service systems. For example, the model developed in this chapter can be generalized to the case of bulk arrival, batch service, or both.

CONCLUSION

We have considered in this chapter a queueing system under T-policy, that is, where the server checks the queue length periodically, every T units of time, during the idle period. Customers require two consecutive phases of service, and the server may break down while providing either phase of service. A Markov chain approach is used to obtain the steady state system size probabilities and different performance measures. The optimal value of the threshold level is obtained analytically. Similar analyses can be applied to more complex systems where either the arrival process or the service process is bulk. Combinations of the T-policy with either D-policy or N-policy are also possible.

REFERENCES

- Abolnikov, L., & Dukhovny, A. (1991). Markov chains with transition delta matrix: Ergodicity conditions, invariant probability measures and applications. *Journal of Applied Mathematics and Stochastic Analysis*, 4(4), 333–356. doi:10.1155/S1048953391000254
- Bertsimas, D., & Papaconstantinou, X. (1988). On the steady state solution of the $M/C_2(a,b)/S$ queueing system. *Transportation Science*, 22(2), 125–138. doi:10.1287/trsc.22.2.125
- Choudhury, G. (2003). A batch arrival queueing system with an additional service channel. *International Journal of Information and Management Sciences*, 14(2), 17–30.
- Choudhury, G. (2003). Some aspects of an $M/G/1$ queueing system with optional second service. *Top (Madrid)*, 11(1), 141–150. doi:10.1007/BF02578955
- Choudhury, G., & Paul, M. (2005). A two phase queueing system with Bernoulli feedback. *Information and Management Sciences*, 16(1), 35–52.
- Choudhury, G., Tadj, L., & Paul, M. (2007). Steady state analysis of an $M^x/G/1$ queue with two phase service and Bernoulli vacation schedule under multiple vacation policy. *Applied Mathematical Modelling*, 31(6), 1079–1091. doi:10.1016/j.apm.2006.03.032
- Çınlar, E. (1975). *Introduction to stochastic processes*. Englewood Cliffs, NJ: Prentice-Hall.
- Heyman, D. (1977). The T-policy for the $M/G/1$ queue. *Management Science*, 23(7), 775–778. doi:10.1287/mnsc.23.7.775
- Ke, J.-C. (2005). Modified T vacation policy for an $M/G/1$ queueing system with an unreliable server and startup. *Mathematical and Computer Modelling*, 41(11-12), 1267–1277. doi:10.1016/j.mcm.2004.08.009

- Ke, J.-C. (2006). Optimal NT policies for M/G/1 system with a startup and unreliable server. *Computers & Industrial Engineering*, 50(3), 248–262. doi:10.1016/j.cie.2006.04.004
- Ke, J.-C. (2008). Two thresholds of a batch arrival queueing system under modified T vacation policy with startup and closedown. *Mathematical Methods in the Applied Sciences*, 31(2), 229–247. doi:10.1002/mma.905
- Ke, J.-C., & Chu, Y.-K. (2008). Optimization on bicriterion policies for M/G/1 system with second optional service. *Journal of Zhejiang University: Science A*, 9(10), 1437–1445. doi:10.1631/jzus. A0820103
- Ke, J.-C., & Chu, Y.-K. (2009). Notes of M/G/1 system under the (p, T) policy with second optional service. *Central European Journal of Operations Research*, 17(4), 425–431. doi:10.1007/s10100-009-0101-7
- Ke, J.-C., Wu, C.-H., & Zhang, Z. G. (2010). Recent developments in vacation queueing models: A short survey. *International Journal of Operations Research*, 7(4), 3–8.
- Keilson, J., & Servi, L. D. (1986). Oscillating random walk models for GI/G/1 vacation system with Bernoulli schedules. *Journal of Applied Probability*, 23(03), 790–802. doi:10.1017/S0021900200111933
- Kim, D.-J., & Moon, S.-A. (2006). Randomized control of T-policy for an M/G/1 system. *Computers & Industrial Engineering*, 51(4), 684–692. doi:10.1016/j.cie.2006.05.005
- Krishna Kumar, B., Vijaykumar, A., & Arivudainambi, D. (2002). An M/G/1 retrial queueing system with two phase service and preemptive resume. *Annals of Operations Research*, 113(1/4), 61–79. doi:10.1023/A:1020901710087
- Kuo, C., Ke, J.-C., & Choudhury, G. (2015). Optimal NT Policies for a Two-Phase Service M/G/1 System with Bernoulli Vacation Schedule. *Quality Technology & Quantitative Management*, 12(3), 343–353. doi:10.1080/16843703.2015.11673385
- Madan, K. C. (2000). An M/G/1 queue with second optional service. *Queueing Systems*, 34(1/4), 37–46. doi:10.1023/A:1019144716929
- Medhi, J. (2002). A single server Poisson input queue with a second optional channel. *Queueing Systems*, 42(3), 239–242. doi:10.1023/A:1020519830116
- Tadj, L. (2003). On a quorum queueing system under T-policy. *The Journal of the Operational Research Society*, 54(5), 466–471. doi:10.1057/palgrave.jors.2601536
- Tadj, L., & Choudhury, G. (2005). Optimal design and control of queues. *Top (Madrid)*, 13(1), 359–414. doi:10.1007/BF02579061
- Tadj, L., Choudhury, G., & Rekaş, K. (2012). A two-phase quorum queueing system with Bernoulli vacation schedule, setup, and N-policy for an unreliable server with delaying repair. *International Journal of Services and Operations Management*, 12(2), 139 - 164.
- Takagi, H. (1990). *Queueing analysis: A foundation of performance evaluation* (Vol. 1). Amsterdam: Elsevier.
- Tian, N., & Zhang, Z. G. (2006). *Vacation queueing models - Theory and applications*. New York, NY: Springer Science.
- Wang, K.-H., & Ke, J. C. (2002). Control policies of an M/G/1 queueing system with a removable and non-reliable server. *International Transactions in Operational Research*, 9(2), 195–212. doi:10.1111/1475-3995.00350

Wang, K.-H., Yang, D.-Y., & Pearn, W. L. (2010). Comparison of two randomized policy M/G/1 queues with second optional service, server breakdown and startup. *Journal of Computational and Applied Mathematics*, 234(3), 812–824. doi:10.1016/j.cam.2010.01.045

Wang, T.-Y., Wang, K.-H., & Pearn, W. L. (2009a). Optimization of the T policy M/G/1 queue with server breakdowns and general startup times. *Journal of Computational and Applied Mathematics*, 228(1), 270–278. doi:10.1016/j.cam.2008.09.021

Wang, T.-Y., Wang, K.-H., & Pearn, W. L. (2009b). Maximum entropy analysis to the T Policy M/G/1 queue with server breakdowns and startup times. *International Journal of Information and Management Sciences*, 20, 395 - 414.

White, H., & Christie, L. G. (1958). Queueing with preemptive priority or breakdown. *Operations Research*, 6(1), 79–96. doi:10.1287/opre.6.1.79

Wua, W., Tanga, Y., & Yub, M. (2014). Analysis of an M/G/1 Queue with Multiple Vacations, N-policy, Unreliable Service Station and Repair Facility Failures. *International Journal of Supply and Operations Management*, 1(1), 1–19.

Yang, D.-Y., Wang, K.-H., Ke, J.-C., & Pearn, W. L. (2008). Optimal randomized control policy of an unreliable server system with second optional service and startup. *Engineering Computations: International Journal for Computer-Aided Engineering and Software*, 25(8), 783 - 800.

Zhang, Z. G., Tadj, L., & Bounkhel, M. (2011). Cost evaluation in M/G/1 queue with T-policy revisited: Technical Note. *European Journal of Operational Research*, 214(3), 814–817. doi:10.1016/j.ejor.2011.05.043

KEY TERMS AND DEFINITIONS

Bernoulli Vacation Schedule: Service discipline specifying that, at the end of a service, the server may either take a vacation or serve the next customer.

Optimal Management Policy: Refers to the selection of a best element from some set of available alternatives.

Queue: A line of waiting persons, jobs, or objects.

Queueing System: Consists of three major components: (1) the source population and the way customers arrive to the system, (2) the serving systems, and (3) how customers exit the system.

Server Vacation: Term used when the server is not serving the customers and is performing some secondary job.

T-Policy: A service discipline specifying that, when the system is empty, the server takes vacations of fixed length T, as long as no customers arrive to the system.

Unreliable Server: A server that can breakdown while providing service, and become unavailable to customers.

Customer Lifetime Value

Kijpokin Kasemsap

Suan Sunandha Rajabhat University, Thailand

INTRODUCTION

Customer lifetime value (CLV) is the present value of all future profits obtained from customers over the life of their relationship with a company (Benoit & van den Poel, 2009). Companies have been increasing their focus on establishing and maintaining the good customer relations during the customers' life in the company (Haenlein, Kaplan, & Beeser, 2007), and thus, researchers and practitioners have realized the importance of the lifetime value of customers. CLV is the value of a customer's entire lifetime with the company, forecasts in the literature usually focus on a short-time period (Donkers, Verhoef, & de Jong, 2007).

CLV has been studied under the name of customer value, customer equity, and customer profitability (Kahreh, Tive, Babania, & Hesani, 2014). CLV, as an important metric in customer relationship management (CRM), has attracted the widespread attention over the last decade (Chen & Fan, 2013). In relationship marketing, CLV is the discounted profit streams of a customer across the entire customer life cycle (Ma, Li, & Chen, 2008). Companies can manage customers with different age groups, gender, and self-construal characteristics with different strategies to maximize benefits at different levels of customer loyalty to improve CLV (Qi, Qu, Zhou, & Li, 2015).

This article aims to bridge the gap in the literature on the thorough literature consolidation of CLV. The extensive literatures of CLV provide a contribution to practitioners and researchers by indicating the important perspectives on CLV in order to maximize the impact of CLV in global marketing.

BACKGROUND

Berger et al. (2006) indicated that companies need to consider revenues and costs based on cash flow of profits in measuring CLV. CLV is the sum of the revenues gained from company's customers over the lifetime of transactions after the deduction of the total cost of attracting, selling, and servicing customers, taking into account the time value of money (Hwang, Jung, & Suh, 2004). Companies can be more profitable if they identify the most profitable customers and invest disproportionate marketing resources in them (Malthouse & Blattberg, 2005). A profitable customer is one who can create profits, increase revenues, and assist in reducing losses (Kotler & Armstrong, 1996), and the difference between the revenues and costs generated by a customer is CLV (Dwyer, 1989).

CLV is presented to evaluate customers in terms of recency, frequency, and monetary (RFM) variables (Cao, Yu, & Zhang, 2015), toward to describe the value of a client through time in terms of profitability (Moro, Cortez, & Rita, 2015). Measuring RFM is an important method to evaluate CLV (Liu & Shih, 2005). Bult and Wansbeek (1995) explained the RFM terms (i.e., recency, frequency, and monetary perspectives). R (Recency) is the period since the last purchase, and a lower value corresponds to a higher probability of the customers making a repeat purchase. F (Frequency) is the number of purchases made within a certain period. Higher frequency indicates higher loyalty. M (Monetary) is the amount of money spent during a certain period. A higher value indicates that the company should focus more on that customer (Bult & Wansbeek, 1995).

IMPORTANT PERSPECTIVES ON CUSTOMER LIFETIME VALUE

This section emphasizes the prospect of CLV and the importance of CLV in global marketing.

Prospect of Customer Lifetime Value

Customer lifetime value (CLV) is viewed as the present value of the future cash flows associated with a customer (Pfeifer, Haskins, & Conroy, 2005). The main goal of CLV is to specify the importance level of each customer for a company (Hiziroglu & Sengul, 2012). CLV considers a customer's value to the company based on predicted future costs and transactions (Kumar, 2010). Gladly et al. (2009) defined CLV as the discounted value of future marginal earnings, based on the customer's activity. Knowing the CLV of individual customers enables the decision maker to improve the customer segmentation and marketing resource allocation efforts (Kim & Lee, 2007) and this perspective will lead to the higher retention rates and profits for the company (Hawkes, 2000).

CLV models can estimate the value of a customer over the entire customer's lifetime (Kahreh et al., 2014). The application of CLV in marketing, used for segmenting customers or formulating strategies, has been found in literature (Cheng, Chiu, Cheng, & Wu, 2012). Kim et al. (2006) proposed a CLV computation model considering the past profit contribution, potential benefit, and defection probability of a customer. Kim et al. (2006) covered a framework for analyzing customer value, and segmenting customers based on their values, and applied their approach to formulating marketing strategies for a wireless communication service company. Shih and Liu (2008) proposed applying a collaborative filtering technique through CLV and customer demand to develop a product recommendation system.

Regarding CLV, a monetary value of a future relationship with a customer is a fundamental concept for the long-term management of the

customer base and the planning of marketing activities (Karvanen, Rantanen, & Luoma, 2014). Kahreh et al. (2014) indicated that data inputs commonly used when making CLV calculations include seven major factors (i.e., acquisition cost, churn rate, discount rate, retention cost, time period, periodic revenue, and profit margin). Acquisition cost is the amount of money a marketing department has to spend, on average, to acquire a single new customer. Acquisition cost is positively associated with customer retention, future profits, and current market values (Livne, Simpson, & Talmor, 2011). Customer retention is a major driver of CLV and is a key performance metric in marketing management (Becker, Spann, & Schulze, 2015).

Churn rate is the percentage of customers who end their relationship with a company in a given time period (Kahreh et al., 2014). Customer churn (or customer attrition) is key for measuring success in the logistics industry (Chen, Hu, & Hsieh, 2015). CLV model considers the past contribution, potential value, and churn probability at the same time (Kim, Kim, & Sohn, 2009). Discount rate is the cost of capital utilized to discount future revenue from a customer. Retention cost is the amount of money a company has to spend in a given time period to retain an existing customer. Time period is the unit of time into which a customer relationship is divided for analysis. Periodic revenue is the amount of revenue collected from a customer in the time period. Profit margin is the profit as a percentage of revenue (Kahreh et al., 2014).

Customer loyalty is a driver of CLV (Qi, Zhou, Chen, & Qu, 2012). The aim of customer loyalty programs in modern organizations is to identify and encourage those customers exhibiting a high CLV (Steinhoff & Palmatier, 2016). Regarding CLV, the differential unit cost of marginal effects, ceiling rate, efficiency, and allocation of spending on acquisition and retention to achieve market share growth can maximize customer equity (Tsao, 2013). Value creation through customer-to-customer exchange occurs when the perceived

benefits of a company's offering are increased as a result of customers' interaction with one another (Gruen, Osmonbekov, & Czaplewski, 2007).

Importance of Customer Lifetime Value in Global Marketing

The concept of CLV receives much attention from marketing practitioners and academicians (Blattberg, Malthouse, & Neslin, 2009). CRM has been widely adopted in modern business (Fang, Shyng, Lee, & Tzeng, 2012). CLV has been an important metric for CRM to evaluate the marketing decisions over the last decade (Borle, Singh, & Jain, 2008). As CRM has become popular, CLV has emphasized a longer relationship with the customers (Kim et al., 2009). Chan et al. (2010) utilized the CLV to predict the company's long-term return on investment. CLV models are employed in the field of marketing to evaluate the lifetime value of customers in conventional businesses (Etzion, Fisher, & Wasserkrug, 2005).

Applying CLV to guide marketing decisions significantly encourages companies to recognize the major differences among customers and begin to create business value (Pfeifer & Ovchinnikov, 2011). The combination of the CLVs of the company's current and future customers (net of acquisition costs) is a measure of the value of the company (Rust, Zeithaml, & Lemon, 2000). In the context of customer development, CLV can be used as the object to select the promotions for the cross-selling and up-selling (Khan, 2009). Mulhern (1999) utilized the concept of present value and future value to measure CLV from the financial perspective. Nadeem (2006) investigated the models of marketing strategy for the long-term and short-term CLVs in an e-business context.

CLV plays an essential role in designing the marketing programs and allocating the marketing resources (Benoit & van den Poel, 2009). Liu et al. (2007) applied different variables to CLV in the service context and defined several relationships to customer value. An important managerial task is to take an array of past retention numbers

for a given group of customers and project them into the future to make more accurate predictions about customer tenure and CLV (Fader & Hardie, 2007). In the context of customer retention, the individual-level CLV gives a sense of how much is being lost due to churn (Rosset, Neumann, Eick, & Vatnik, 2003), and the decrease of CLV is utilized as a criterion to define churners in noncontractual settings (Glady et al., 2009). Examples of non-contractual context are retail purchases where the company does not know whether a customer has defected or not (Singh, Borle, & Jain, 2009).

FUTURE RESEARCH DIRECTIONS

The classification of the extensive literature in the domains of CLV will provide the potential opportunities for future research. CRM is the managerial philosophy according to which the company's goal can be achieved through the identification and satisfaction of the customer's requirement (Kasemsap, 2015a). To be the successful global retailers, the ability to apply and adjust the retail marketing mix elements and retail brand strategy concerning specific target-market conditions and a wider global marketing strategy, is functionally important in retail settings (Kasemsap, 2015b).

Marketing strategies include all basic and long-term activities in the field of marketing, thus dealing with the analysis of the strategic situation of an organization and the formulation, evaluation and selection of market-oriented strategies and effectively contribute to the goals of the organization and its marketing objectives (Kasemsap, 2015c). The capability of social media in building brand in the global marketplace is practically important in modern advertising (Kasemsap, 2015d). Developing a good relationship with the target market is essential for brand management in global marketing (Kasemsap, 2016). An examination of linkages among CLV, CRM, brand management, marketing strategies, and social media platforms in global marketing would seem to be viable for future research efforts.

Consumer attitude toward products can be changed by highlighting the new functions of the products, by changing the beliefs the consumer has regarding the products, and by getting the consumer more involved in the products (Kasemsap, 2017a). Web mining techniques can be applied with the effective analysis of the clearly understood business needs and requirements (Kasemsap, 2017b). Business intelligence is the computer-based techniques used in analyzing business data, such as sales revenue by products or associated costs and incomes (Kasemsap, 2017c). In the customer-related supply chain processes, information sharing can help in formulating customer experience strategies, increase customer service effectiveness, and improve supply chain operations (Kasemsap, 2017d). Considering the relationships among CLV, consumer attitude, web mining, business intelligence, and information sharing in global marketing would be beneficial for future research directions.

CONCLUSION

This article highlighted the prospect of CLV and the importance of CLV in global marketing. CLV is the total of the financial profit, calculated from the existing period to the future. CLV develops the optimal strategies for customer engagement, promotes the understanding of potential value of a customer, and enables the workforce to effectively improve customer relationships. Companies utilizing CLV can practically plan for availability of the service or product by predicting the future demand by customers. CLV can be a good indicator for operational efficiency and customer satisfaction. CLV can be managed to implement the initiatives that extend the life span of a customer through brand loyalty, increase the frequency of purchases, and develop the directed marketing campaigns to find more customers with similar behaviors.

CLV is an important metric for determining how much money a company wants to spend on acquiring new customers and how much repeat

business a company can expect from certain consumers. CLV can be a crucial perspective for costs to be associated with the promotions and communications to attract the new customers and retain the existing customers. CLV can help individuals estimate a customer's monetary worth to a business after factoring in the value of the relationship with a customer over time. CLV can determine the amount that will be required to retain customer segment. Promoting CLV has the potential to enhance marketing performance and reach strategic goals in global marketing.

REFERENCES

- Becker, J. U., Spann, M., & Schulze, T. (2015). Implications of minimum contract durations on customer retention. *Marketing Letters*, 26(4), 579–592. doi:10.1007/s11002-014-9293-2
- Benoit, D. F., & van den Poel, D. (2009). Benefits of quantile regression for the analysis of customer lifetime value in a contractual setting: An application in financial services. *Expert Systems with Applications*, 36(7), 10475–10484. doi:10.1016/j.eswa.2009.01.031
- Berger, P. D., Eechambadi, N., George, M., Lehmann, D. R., Rizley, R., & Venkatesan, R. (2006). From customer lifetime value to shareholder value: Theory, empirical evidence, and issues for future research. *Journal of Service Research*, 9(2), 156–167. doi:10.1177/1094670506293569
- Blattberg, R. C., Malthouse, E. C., & Neslin, S. A. (2009). Customer lifetime value: Empirical generalizations and some conceptual questions. *Journal of Interactive Marketing*, 23(2), 157–168. doi:10.1016/j.intmar.2009.02.005
- Borle, S., Singh, S. S., & Jain, D. C. (2008). Customer lifetime value measurement. *Management Science*, 54(1), 100–112. doi:10.1287/mnsc.1070.0746

- Bult, J. R., & Wansbeek, T. J. (1995). Optimal selection for direct mail. *Marketing Science*, 14(4), 378–394. doi:10.1287/mksc.14.4.378
- Cao, J., Yu, X., & Zhang, Z. (2015). Integrating OWA and data mining for analyzing customers churn in e-commerce. *Journal of Systems Science and Complexity*, 28(2), 381–392. doi:10.1007/s11424-015-3268-0
- Chan, S. L., Ip, W. H., & Cho, V. (2010). A model for predicting customer value from perspectives of product attractiveness and marketing strategy. *Expert Systems with Applications*, 37(2), 1207–1215. doi:10.1016/j.eswa.2009.06.030
- Chen, K., Hu, Y. H., & Hsieh, Y. C. (2015). Predicting customer churn from valuable B2B customers in the logistics industry: A case study. *Information Systems and e-Business Management*, 13(3), 475–494.
- Chen, Z. Y., & Fan, Z. P. (2013). Dynamic customer lifetime value prediction using longitudinal data: An improved multiple kernel SVR approach. *Knowledge-Based Systems*, 43, 123–134. doi:10.1016/j.knosys.2013.01.022
- Cheng, C. J., Chiu, S. W., Cheng, C. B., & Wu, J. Y. (2012). Customer lifetime value prediction by a Markov chain based data mining model: Application to an auto repair and maintenance company in Taiwan. *Scientia Iranica, Transactions E. Industrial Engineering (American Institute of Industrial Engineers)*, 19(3), 849–855.
- Donkers, B., Verhoef, P. C., & de Jong, M. G. (2007). Modeling CLV: A test of competing models in the insurance industry. *Quantitative Marketing and Economics*, 5(2), 163–190. doi:10.1007/s11129-006-9016-y
- Dwyer, F. R. (1989). Customer lifetime profitability to support marketing decision making. *Journal of Direct Marketing*, 3(4), 8–15. doi:10.1002/dir.4000030404
- Etzion, O., Fisher, A., & Wasserkrug, S. (2005). E-CLV: A modeling approach for customer lifetime evaluation in e-commerce domains, with an application and case study for online auction. *Information Systems Frontiers*, 7(4), 421–434. doi:10.1007/s10796-005-4812-6
- Fader, P. S., & Hardie, B. G. S. (2007). How to project customer retention. *Journal of Interactive Marketing*, 21(1), 76–90. doi:10.1002/dir.20074
- Fang, S. K., Shyng, J. Y., Lee, W. S., & Tzeng, G. H. (2012). Exploring the preference of customers between financial companies and agents based on TCA. *Knowledge-Based Systems*, 27(1), 137–151. doi:10.1016/j.knosys.2011.09.003
- Gladly, N., Baesens, B., & Croux, C. (2009). Modeling churn using customer lifetime value. *European Journal of Operational Research*, 197(1), 402–411. doi:10.1016/j.ejor.2008.06.027
- Gruen, T. W., Osmonbekov, T., & Czaplewski, A. J. (2007). Customer-to-customer exchange: Its MOA antecedents and its impact on value creation and loyalty. *Journal of the Academy of Marketing Science*, 35(4), 537–549. doi:10.1007/s11747-006-0012-2
- Haenlein, M., Kaplan, A., & Beeser, A. (2007). A model to determine customer lifetime value in a retail banking context. *European Management Journal*, 25(3), 221–234. doi:10.1016/j.emj.2007.01.004
- Hawkes, V. A. (2000). The heart of the matter: The challenge of customer lifetime value. *CRM Forum Resources*, 13, 2–10.
- Hiziroglu, A., & Sengul, S. (2012). Investigating two customer lifetime value models from segmentation perspective. *Procedia: Social and Behavioral Sciences*, 62, 766–774. doi:10.1016/j.sbspro.2012.09.129

- Hwang, H., Jung, T., & Suh, E. (2004). An LTV model and customer segmentation based on customer value: A case study on the wireless telecommunication industry. *Expert Systems with Applications*, 26(2), 181–188. doi:10.1016/S0957-4174(03)00133-7
- Kahreh, M. S., Tive, M., Babania, A., & Hesani, M. (2014). Analyzing the applications of customer lifetime value (CLV) based on benefit segmentation for the banking sector. *Procedia: Social and Behavioral Sciences*, 109, 590–594. doi:10.1016/j.sbspro.2013.12.511
- Karvanen, J., Rantanen, A., & Luoma, L. (2014). Survey data and Bayesian analysis: A cost-efficient way to estimate customer equity. *QME: Quantitative Marketing and Economics*, 12(3), 305–329.
- Kasemsap, K. (2015a). The role of customer relationship management in the global business environments. In T. Tsiakis (Ed.), *Trends and innovations in marketing information systems* (pp. 130–156). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-8459-1.ch007
- Kasemsap, K. (2015b). The roles of corporate marketing strategies and brand management in the global retail industry. In E. Pantano (Ed.), *Successful technological integration for competitive advantage in retail settings* (pp. 310–339). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-8297-9.ch014
- Kasemsap, K. (2015c). The role of marketing strategies in modern business and society. In A. Jean-Vasile, I. Andreea, & T. Adrian (Eds.), *Green economic structures in modern business and society* (pp. 274–300). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-8219-1.ch015
- Kasemsap, K. (2015d). The role of social media in international advertising. In N. Taşkıran & R. Yılmaz (Eds.), *Handbook of research on effective advertising strategies in the social media age* (pp. 171–196). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-8125-5.ch010
- Kasemsap, K. (2016). The roles of social media marketing and brand management in global marketing. In W. Ozuem & G. Bowen (Eds.), *Competitive social media marketing strategies* (pp. 173–200). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-9776-8.ch009
- Kasemsap, K. (2017a). Mastering consumer attitude and sustainable consumption in the digital age. In N. Suki (Ed.), *Handbook of research on leveraging consumer psychology for effective customer engagement* (pp. 16–41). Hershey, PA: IGI Global. doi:10.4018/978-1-5225-0746-8.ch002
- Kasemsap, K. (2017b). Mastering web mining and information retrieval in the digital age. In A. Kumar (Ed.), *Web usage mining techniques and applications across industries* (pp. 1–28). Hershey, PA: IGI Global. doi:10.4018/978-1-5225-0613-3.ch001
- Kasemsap, K. (2017c). Mastering business process management and business intelligence in global business. In M. Tavana, K. Szabat, & K. Puranam (Eds.), *Organizational productivity and performance measurements using predictive modeling and analytics* (pp. 192–212). Hershey, PA: IGI Global. doi:10.4018/978-1-5225-0654-6.ch010
- Kasemsap, K. (2017d). Advocating information system, information integration, and information sharing in global supply chain. In G. Jamil, A. Soares, & C. Pessoa (Eds.), *Handbook of research on information management for effective logistics and supply chains* (pp. 107–130). Hershey, PA: IGI Global. doi:10.4018/978-1-5225-0973-8.ch006
- Khan, R., Lewis, M., & Singh, V. (2009). Dynamic customer management and the value of one-to-one marketing. *Marketing Science*, 28(6), 1063–1079. doi:10.1287/mksc.1090.0497
- Kim, E., & Lee, B. (2007). An economic analysis of customer selection and leverage strategies in a market where network externalities exist. *Decision Support Systems*, 44(1), 124–134. doi:10.1016/j.dss.2007.03.006

- Kim, G., Kim, A., & Sohn, S. Y. (2009). Conjoint analysis for luxury brand outlet malls in Korea with consideration of customer lifetime value. *Expert Systems with Applications*, 36(1), 922–932. doi:10.1016/j.eswa.2007.10.004
- Kim, S. Y., Jung, T. S., Suh, E. H., & Hwang, H. S. (2006). Customer segmentation and strategy development based on customer life time value: A case study. *Expert Systems with Applications*, 31(1), 101–107. doi:10.1016/j.eswa.2005.09.004
- Kotler, P., & Armstrong, G. (1996). *Principle of marketing*. Englewood Cliff, NJ: Prentice Hall.
- Kumar, V. (2010). A customer lifetime value-based approach to marketing in the multichannel, multimedia retailing environment. *Journal of Interactive Marketing*, 24(2), 71–85. doi:10.1016/j.intmar.2010.02.008
- Liu, B. S. C., Petruzzi, N. C., & Sudharshan, D. (2007). A service effort allocation model customer lifetime value in service marketing. *Journal of Services Marketing*, 21(1), 24–35. doi:10.1108/08876040710726275
- Liu, D. R., & Shih, Y. Y. (2005). Hybrid approaches to product recommendation based on customer lifetime value and purchase preferences. *Journal of Systems and Software*, 77(2), 181–191. doi:10.1016/j.jss.2004.08.031
- Livne, G., Simpson, A., & Talmor, E. (2011). Do customer acquisition cost, retention and usage matter to firm performance and valuation? *Journal of Business Finance & Accounting*, 38(3/4), 334–363. doi:10.1111/j.1468-5957.2010.02229.x
- Ma, M., Li, Z., & Chen, J. (2008). Phase-type distribution of customer relationship with Markovian response and marketing expenditure decision on the customer lifetime value. *European Journal of Operational Research*, 187(1), 313–326. doi:10.1016/j.ejor.2007.03.018
- Malthouse, E. C., & Blattberg, R. C. (2005). Can we predict customer lifetime value? *Journal of Interactive Marketing*, 19(1), 2–16. doi:10.1002/dir.20027
- Moro, S., Cortez, P., & Rita, P. (2015). Using customer lifetime value and neural networks to improve the prediction of bank deposit subscription in telemarketing campaigns. *Neural Computing & Applications*, 26(1), 131–139. doi:10.1007/s00521-014-1703-0
- Mulhern, F. J. (1999). Customer profitability analysis: Measurement, concentration, and research directions. *Journal of Interactive Marketing*, 13(1), 25–40. doi:10.1002/(SICI)1520-6653(199924)13:1<25::AID-DIR3>3.0.CO;2-L
- Nadeem, M. M. (2006). How e-business leadership results in customer satisfaction, and customer lifetime value. *Business Review (Federal Reserve Bank of Philadelphia)*, 6(1), 218–224.
- Pfeifer, P. E., Haskins, M. E., & Conroy, R. M. (2005). Customer lifetime value, customer profitability and the treatment of acquisition spending. *Journal of Managerial Issues*, 17(1), 11–25.
- Pfeifer, P. E., & Ovchinnikov, A. (2011). A note on willingness to spend and customer lifetime value for firms with limited capacity. *Journal of Interactive Marketing*, 25(3), 178–189. doi:10.1016/j.intmar.2011.02.003
- Qi, J. Y., Qu, Q. X., Zhou, Y. P., & Li, L. (2015). The impact of users' characteristics on customer lifetime value raising: Evidence from mobile data service in China. *Information Technology & Management*, 16(4), 273–290. doi:10.1007/s10799-014-0200-6
- Qi, J. Y., Zhou, Y. P., Chen, W. J., & Qu, Q. X. (2012). Are customer satisfaction and customer loyalty drivers of customer lifetime value in mobile data services: A comparative cross-country study. *Information Technology & Management*, 13(4), 281–296. doi:10.1007/s10799-012-0132-y

Rosset, S., Neumann, E., Eick, U., & Vatnik, N. (2003). Customer lifetime value models for decision support. *Data Mining and Knowledge Discovery*, 7(3), 312–339. doi:10.1023/A:1024036305874

Rust, R. T., Zeithaml, V. A., & Lemon, K. N. (2000). *Driving customer equity: How customer lifetime value is reshaping corporate strategy*. New York, NY: The Free Press.

Shih, Y. Y., & Liu, D. R. (2008). Product recommendation approaches: Collaborative filtering via customer lifetime value and customer demands. *Expert Systems with Applications*, 35(1/2), 350–360. doi:10.1016/j.eswa.2007.07.055

Singh, S. S., Borle, S., & Jain, D. C. (2009). A generalized framework for estimating customer lifetime value when customer lifetimes are not observed. *QME: Quantitative Marketing and Economics*, 7(2), 181–205.

Steinboff, L., & Palmatier, R. W. (2016). Understanding loyalty program effectiveness: Managing target and bystander effects. *Journal of the Academy of Marketing Science*, 44(1), 88–107. doi:10.1007/s11747-014-0405-6

Tsao, H. Y. (2013). Budget allocation for customer acquisition and retention while balancing market share growth and customer equity. *Marketing Letters*, 24(1), 1–11. doi:10.1007/s11002-012-9217-y

ADDITIONAL READING

Aarikka-Stenroos, L., & Jaakkola, E. (2012). Value co-creation in knowledge intensive business services: A dyadic perspective on the joint problem solving process. *Industrial Marketing Management*, 41(1), 15–26. doi:10.1016/j.indmarman.2011.11.008

Arslanagic-Kalajdzic, M., & Zabkar, V. (2015). The external effect of marketing accountability in business relationships: Exploring the role of customer perceived value. *Industrial Marketing Management*, 46, 83–97. doi:10.1016/j.indmarman.2015.03.002

Benedek, G., Lublóy, Á., & Vastag, G. (2014). The importance of social embeddedness: Churn models at mobile providers. *Decision Sciences*, 45(1), 175–201. doi:10.1111/deci.12057

Cermak, P. (2015). Customer profitability analysis and customer life time value models: Portfolio analysis. *Procedia Economics and Finance*, 25, 14–25. doi:10.1016/S2212-5671(15)00708-X

Chandler, G. N., Broberg, J. C., & Allison, T. H. (2014). Customer value propositions in declining industries: Differences between industry representative and high-growth firms. *Strategic Entrepreneurship Journal*, 8(3), 234–253. doi:10.1002/sej.1181

Chen, S., Ho, T. K., Vyas, A., Cao, J., & Spiess, J. (2014). A multi-layer dynamic model for customer experience analytics. *Bell Labs Technical Journal*, 18(4), 19–32. doi:10.1002/bltj.21643

Choo, H. J., Moon, H., Kim, H., & Yoon, N. (2012). Luxury customer value. *Journal of Fashion Marketing and Management*, 16(1), 81–101. doi:10.1108/13612021211203041

Clempner, J. B., & Poznyak, A. S. (2014). Simple computing of the customer lifetime value: A fixed local-optimal policy approach. *Journal of Systems Science and Systems Engineering*, 23(4), 439–459. doi:10.1007/s11518-014-5260-y

Enz, M. G., & Lambert, D. M. (2012). Using cross-functional, cross-firm teams to co-create value: The role of financial measures. *Industrial Marketing Management*, 41(3), 495–507. doi:10.1016/j.indmarman.2011.06.041

- Eskafi, M., Hosseini, S., & Yazd, A. M. (2013). The value of telecom subscribers and customer relationship management. *Business Process Management Journal*, 19(4), 737–748. doi:10.1108/BPMJ-Feb-2012-0016
- Ethiraj, S. K., Ramasubbu, N., & Krishnan, M. S. (2012). Does complexity deter customer-focus? *Strategic Management Journal*, 33(2), 137–161. doi:10.1002/smj.947
- Gray, G. T., & Wert-Gray, S. (2012). Customer retention in sports organization marketing: Examining the impact of team identification and satisfaction with team performance. *International Journal of Consumer Studies*, 36(3), 275–281. doi:10.1111/j.1470-6431.2011.00999.x
- Heidemann, J., Kaiser, M., Klier, M., & Probst, F. (2012). Customer lifetime value-based sales force control in the financial services industry: An incentive-compatible remuneration model. *Journal of Management Control*, 23(1), 27–51. doi:10.1007/s00187-012-0152-0
- Hibbert, S. A., Piacentini, M. G., & Hogg, M. K. (2012). Service recovery following dysfunctional consumer participation. *Journal of Consumer Behaviour*, 11(4), 329–338. doi:10.1002/cb.1391
- Lambert, D. M., & Enz, M. G. (2012). Managing and measuring value co-creation in business-to-business relationships. *Journal of Marketing Management*, 28(13/14), 37–41.
- Maicas, J. P., & Sese, F. J. (2015). Customer-base management in network industries: The moderating role of network size and market growth. *European Management Review*, 12(4), 209–220. doi:10.1111/emre.12048
- Matsuno, K., Zhu, Z., & Rice, M. P. (2014). Innovation process and outcomes for large Japanese firms: Roles of entrepreneurial proclivity and customer equity. *Journal of Product Innovation Management*, 31(5), 1106–1124. doi:10.1111/jpim.12147
- Mittal, V., Frennea, C., & Westbrook, R. A. (2014). Managing customer value in business-to-business markets. *Marketing Review St. Gallen*, 31(3), 46–54. doi:10.1365/s11621-014-0362-y
- Moon, H., Miller, D. R., & Kim, S. H. (2013). Product design innovation and customer value: Cross-cultural research in the United States and Korea. *Journal of Product Innovation Management*, 30(1), 31–43. doi:10.1111/j.1540-5885.2012.00984.x
- Morgan, N. A., & Rego, L. L. (2012). The value of different customer satisfaction and loyalty metrics in predicting business performance. *Marketing Science*, 25(5), 426–439. doi:10.1287/mksc.1050.0180
- Mzoughia, M. B., & Limam, M. (2015). An improved customer lifetime value model based on Markov chain. *Applied Stochastic Models in Business and Industry*, 31(4), 528–535. doi:10.1002/asmb.2053
- Nikkhahan, B., & Tarokh, M. J. (2013). Citizen lifetime value in e-government. In R. Eid (Ed.), *Managing customer trust, satisfaction, and loyalty through information communication technologies* (pp. 202–218). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-3631-6.ch012
- Prior, D. D. (2013). Supplier representative activities and customer perceived value in complex industrial solutions. *Industrial Marketing Management*, 42(8), 1192–1201. doi:10.1016/j.indmarman.2013.03.015
- Rathi, T., & Ravi, V. (2014). Customer lifetime value measurement using machine learning techniques. In J. Wang (Ed.), *Encyclopedia of business analytics and optimization* (pp. 565–573). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-5202-6.ch051

Sousa, R., Amorim, M., Rabinovich, E., & Sodero, A. C. (2015). Customer use of virtual channels in multichannel services: Does type of activity matter? *Decision Sciences*, 46(3), 623–657. doi:10.1111/deci.12142

Trocchia, P. J., & Luckett, M. G. (2013). Transitory bias as a source of customer dissatisfaction: An exploratory investigation. *Journal of Consumer Behaviour*, 12(1), 32–41. doi:10.1002/cb.1403

Wouters, M., & Kirchberger, M. A. (2015). Customer value propositions as interorganizational management accounting to support customer collaboration. *Industrial Marketing Management*, 46, 54–67. doi:10.1016/j.indmarman.2015.01.005

KEY TERMS AND DEFINITIONS

Customer: Individual that buys products or services.

Customer Equity: The value of potential future revenue generated by a company's customers in a lifetime.

Customer Lifetime Value: The total gain expressed on a net present value basis that a business expects from having an enduring commercial relationship with a customer over time.

Customer Loyalty: The likelihood of previous customers to continue to buy from a specific organization.

Customer Retention: The assessment of the product or service quality provided by a business that measures how loyal its customers are.

Customer Satisfaction: The degree of satisfaction provided by the products or services of a company.

Marketing: The social and managerial processes by which individuals or companies obtain what they need or want through creating, offering, and exchanging products with each other.

Revenue: The income that a government or company regularly receives.

Facilitating Customer Relationship Management in Modern Business

Kijpokin Kasemsap

Suan Sunandha Rajabhat University, Thailand

INTRODUCTION

In a highly competitive market, companies need to maintain positive relationship with their customers (Hassan, Nawaz, Lashari, & Zafar, 2015). Customer relationship management (CRM) encompasses the broad perspectives of activities starting with the segmentation of customers in the database and continuing with acquiring new customers and retaining existing customers (Matis & Ilies, 2014). CRM systems are a group of information systems that enable organizations to contact customers and collect, store, and analyze customer data to provide a comprehensive view of their customers (Khodakarami & Chan, 2014). Customer data and information technology (IT) organize the groundwork upon which any successful CRM strategy is established (Tsai, 2011).

CRM becomes one of the most important business strategies in the digital age, thus involving organizational capability of managing business interactions with customers in an effective manner (Kasemsap, 2015a). CRM refers to a firm's activities for establishing and maintaining relationships with its customers (Kalaighnam & Varadarajan, 2012). The merger of CRM systems with social media technology has given way to a new concept of CRM that incorporates a more collaborative and network-focused approach to managing customer relationships (Trainor, Andzulis, Rapp, & Agnihotri, 2014).

CRM is a critical issue in services management (Cheng & Yang, 2013). CRM systems are essential for companies seeking growth and profits in the global marketplace (Heidemann, Klier, Landherr, & Zimmermann, 2013). Keramati et al. (2013)

indicated that the nature of CRM is multidimensional, and it is one of the most complicated innovations in modern organizations. Customer value, customer satisfaction, and brand loyalty have mediated positive effect on CRM performance (Kasemsap, 2014a). This article aims to bridge the gap in the literature on the thorough literature consolidation of CRM. The extensive literature of CRM provides a contribution to practitioners and researchers by describing the multifaceted applications of CRM in order to maximize the business impact of CRM in modern business.

BACKGROUND

Effective CRM has emerged as a strategic imperative for companies in every business (Benmoussa, 2005). CRM has been recognized as a set of methodologies and organizational processes to attract and retain customers through their increased customer satisfaction and customer loyalty (Cousment & van den Poel, 2008). CRM is used to define the process of creation and maintain relationship with customers (Hassan et al., 2015). CRM is defined as the adoption of IT to develop new customers and retain old customers so as to keep the long-term customer relationship, which aims to improve customer relationship and can help increase customer loyalty, customer retention, and customer profitability (Hennig-Thurau, Gwinner, & Gremler, 2002).

The rapid growth of CRM systems raises the opportunity within many firms to effectively utilize the customer data over time to secure competitive advantage in modern business (Eichorn, 2004).

CRM data is among the most important and comprehensive information available to executives in many organizations (Stein, Smith, & Lancioni, 2013). With the increasing importance of CRM in every industry domain, CRM classification practitioners demand a standardized framework with the streamlined data mining processes capable of delivering the satisfactory result for CRM data with all data mining challenges (Tu & Yang, 2013). CRM leverages information, technology, and people to create and deliver value to the targeted customer at a profit (Salim & Keramati, 2014).

CRM relational information process includes information reciprocity, information capture, information integration, information access, and information use, and expected the process to enable the facilitation of CRM performance (Jayachandran, Sharma, Kaufman, & Raman, 2005). CRM relational information processing is a process in which firms effectively engage in the systematic registration, integration, and analysis of customer information (Hillebrand, Nijholt, & Nijssen, 2011). The key objective of the CRM relational information process is a focus on initiating, maintaining, and retaining the long-term customer relationships (Becker, Greve, & Albers, 2009).

CHARACTERISTICS OF CUSTOMER RELATIONSHIP MANAGEMENT

This article describes the overview of CRM; CRM and technological utilization; and the facilitation of CRM in modern business.

Overview of Customer Relationship Management

Customer relationship management (CRM) has been widely recognized as an important business approach to build the long-term relationships with specific customers (Coltman, Devinney, & Midgley, 2011). CRM is recognized as a manage-

rial tool for guiding sales teams and supporting engineers and other specialists in developing sales prospects, creating appropriate business proposals, dealing with customer objections, and providing post-sales customer support (Agrawal, 2003). Chen and Popovich (2003) defined CRM as a cross-functional, customer-driven, and technology-integrated business process management strategy that maximizes relationships. Kumar and Reinartz (2006) considered CRM as the strategic approach of identifying the firm's customers that can be profitably served.

CRM applications are considered as the closed-loop systems focusing on a set of priorities and time-critical functions at the customer level (Ku, 2010). CRM includes four major dimensions (i.e., customer identification, customer attraction, customer retention, and customer lifetime value) toward developing the sustainable relationships (Ngai, Xiu, & Chau, 2009). Current models of customer lifetime value consider the discounted value of profits that a customer generates over an expected lifetime of relationship with the firm (Singh, Murthi, & Steffes, 2013). Providing free trials for the particular products is an essential way to give feedbacks to the customers toward promoting CRM (Wei, Lee, Chen, & Wu, 2013). Mohammad et al. (2013) stated that promoting CRM improves organizational performance in the tourism-related hotel industry.

CRM not only enhances the flow of information within an organization, makes it easier to store customer data, and increases information accessibility (Hillebrand et al., 2011), but also leads to the greater insights into customer needs, customer behavior, and customer expectations by developing and maintaining the interactive relationships (Spiller, Vlastic, & Yetton, 2007). The development of long-term relationship with customers and turning them into the loyal customers are the important aspects of successful marketing because customers are the crucial assets in modern business (Wang, 2013).

Customer Relationship Management and Technological Utilization

A large proportion of firms that adopt CRM technology find it challenging to integrate CRM technology into their core marketing processes and utilize CRM to improve their performance (Saini, Grewal, & Johnson, 2010). In the context of CRM, there are various metrics used to define firm performance, including financial performance (Noble, Sinha, & Kumar, 2002), the elements of the balanced scorecard (Chang, 2007), and those that are either process-related perspective (Kim & Kim, 2009) or sales-related perspective (Bodenberg, 2001), as well as those that are related to customer satisfaction and economic performance (Keramati, Mehrabi, & Mojir, 2010). CRM-compatible organizational alignment and CRM technology positively moderate the relationships between CRM processes and economic performance (Reinartz, Karfft, & Hoyer, 2004).

CRM recognizes the new capabilities enabled by the technological and social shifts concerning social media applications (Trainor, 2012). The CRM-performance link is fully mediated by differentiation and cost leadership (Reimann, Schilke, & Thomas, 2010). CRM can contribute to the advancement of the web mining research (Tuzhilin, 2012). IT is recognized as an important component of CRM (Chang, Park, & Chaib, 2010). The practices of IT, knowledge management, and project management efficiently support each other in order to achieve planned goals in modern business (Kasemsap, 2015b). CRM technology is viewed as the sales or marketing-centric tool that increases the employee's effectiveness (Crittenden, Peterson, & Albaum, 2010).

Social media (e.g., Facebook and Twitter) and Web 2.0 technologies can help promote CRM in modern business. Social media allows access to data and video information to a broad range of potential consumers (Kasemsap, 2015c). Social media technologies need to be integrated with CRM processes to establish the firm-level capabil-

ity that influences the organizational performance (Trainor et al., 2014). Social media technologies are integrated throughout the organization to facilitate the capability development and to moderate the capability-performance relationship (Nevo & Wade, 2010). Social media enables the creation of knowledge value chain to customize information and delivery for a technological business growth (Kasemsap, 2014b). Social media encourages individuals to be visible to others and establish connections with other people in modern organizations (Kasemsap, 2014c).

Facilitation of Customer Relationship Management in Modern Business

Interaction orientation and CRM readiness are recognized as the moderators on the relationship between CRM relational information processes and customer-based performance (Chang, Wong, & Fang, 2015). Interaction orientation supports the creation of a trustworthy relationship with customers and leads to an increase in customer loyalty (Homburg & Muller, 2011). Interaction orientation reflects a firm's ability to interact with its customers and to take advantage of information obtained from them through the successive interactions in order to achieve the profitable customer relationships (Ramani & Kumar, 2008). Interaction orientation is proposed to assist in the accomplishment of CRM relational information processes and promotes the customer-based performance (Chang et al., 2015).

Bohling et al. (2006) stated that successful implementation of CRM by firms depends on their strategic fit on CRM, marketing strategy, and cooperation in modern business. CRM implementation can be favorable if there is the effective cross-functional incorporation of people, operation, and marketing expertise that is promoted by the process of technology, application, and information (Richard, Thirkell, & Huff, 2007). It is worth emphasizing the role that employees can play in the achievement of organizational goals to

adopt CRM initiatives as part of its operational activities, such as call centers (Abdullateef, Muktar, Yusoff, & Ahmad, 2014). The successful CRM is depended on internal IT service capability (Yang, 2012). CRM failures have occurred through a lack of strategic focus (Frow & Payne, 2009).

CRM systems include three categories: operational systems (used for automation and increased efficiency of CRM processes), analytical systems (utilized for the analysis of customer data and knowledge), and collaborative systems (utilized to integrate the communication channels and the customer interaction touch points) (Geib, Kolbe, & Brenner, 2006). CRM systems are the information systems that enable organizations to contact with customers, provide services for them, store customer information, and analyze that information to provide a comprehensive view of the customers (Karakostas, Kardaras, & Papathanassiou, 2005). CRM systems help organizations acquire and generate customer knowledge (Khodakarami & Chan, 2014). Customer knowledge is the critical asset toward gaining the valuable activity for organizations (Winer, 2001). Knowledge for customers is required to support the customer's relations and satisfy the customer's knowledge needs (Smith & McKeen, 2005).

FUTURE RESEARCH DIRECTIONS

The classification of the extensive literature in the domains of CRM will provide the potential opportunities for future research. Regarding the increasing advances in technology, IT and knowledge management effectively improve the strategic tools for providing the direct link between customers and tourism organizations, thus encouraging the communication channels in global tourism (Kasemsap, 2016). Understanding the importance of good customer service is essential for a profitable business in creating new customers, keeping loyal customers, and developing referrals for future customers. Customer experience is the

practice of designing and reacting to customer interactions for the purpose of meeting customer expectations, thus increasing customer satisfaction and customer loyalty.

Customer orientation is the important approach to sales and customer-relations in which executives and marketing professionals focus on helping customers to meet their long-term needs and wants. Customer satisfaction is the significant outcome of good marketing practice. Customer satisfaction is valuable from both a customer goodwill perspective and an organization's financial perspective. The study of consumer behavior helps marketers recognize and forecast the purchase behavior of the consumers while they are purchasing a product, toward increasing customer satisfaction. The relationships among CRM, IT, knowledge management, customer service, customer experience, customer orientation, customer satisfaction, and consumer behavior will be the important topics for future research directions.

CONCLUSION

The overview of CRM; CRM and technological utilization; and the facilitation of CRM in modern business are highlighted in this article. Efficiently dealing with all customers and providing them what they actually need increases the customer satisfaction. This increases the chance of getting more business which ultimately enhances turnover and profit. CRM is the important strategy to learn more about customer needs and customer behaviors in order to develop the stronger relationships with them. CRM is the most efficient approach in maintaining and creating the relationships with customers in modern business. CRM helps business gain the insight into the behavior of their customers and modify their business operations to ensure that customers are served in the best possible way.

Modern organizations should have an effective CRM system to cope up with all the business

needs. CRM helps business recognize the value of its customers and to capitalize on the improved customer relations. CRM system is utilized to deal with the existing customers and is also useful in acquiring new customers. CRM involves a historical view and analysis of all the acquired or to be acquired customers. CRM can predict customer requirements and can increase business growth. Installing a CRM system can definitely improve the business solutions and can help in challenging the new ways of marketing and business in an efficient manner. Facilitating CRM has the potential to enhance organizational performance and gain sustainable competitive advantage in modern business.

REFERENCES

- Abdullateef, A. O., Muktar, S. S. M., Yusoff, R. Z., & Ahmad, I. S. B. (2014). Effects of customer relationship management strategy on call centre's employee intention to quit: Evidence from Malaysia call centers. *Procedia: Social and Behavioral Sciences*, 130, 305–315. doi:10.1016/j.sbspro.2014.04.036
- Agrawal, M. L. (2003). Customer relationship management (CRM) & corporate renaissance. *Journal of Service Research*, 3(2), 149–171.
- Becker, J. U., Greve, G., & Albers, S. (2009). The impact of technological and organizational implementation of CRM on customer acquisition, maintenance, and retention. *International Journal of Research in Marketing*, 26(3), 207–215. doi:10.1016/j.ijresmar.2009.03.006
- Benmoussa, C. (2005). Supporting salespersons' CRM efforts through location-based mobile support systems. *Journal of Systems Science and Systems Engineering*, 14(1), 97–114. doi:10.1007/s11518-006-0184-9
- Bodenberg, T. M. (2001). *Customer relationship management: New ways of keeping the customer satisfied*. New York, NY: The Conference Board.
- Bohling, T., Bowman, D., LaValle, S., Mittal, V., Narayandas, D., Ramani, G., & Varadarajan, R. (2006). CRM implementation effectiveness issues and insights. *Journal of Service Research*, 9(2), 184–194. doi:10.1177/1094670506293573
- Chang, H. H. (2007). Critical factors and benefits in the implementation of customer relationship management. *Total Quality Management & Business Excellence*, 18(5), 483–508. doi:10.1080/14783360701239941
- Chang, H. H., Wong, K. H., & Fang, P. W. (2015). The effects of customer relationship management relational information processes on customer-based performance. *Decision Support Systems*, 66, 146–159. doi:10.1016/j.dss.2014.06.010
- Chang, W., Park, J. E., & Chaiy, S. (2010). How does CRM technology transform into organizational performance? A mediating role of marketing capability. *Journal of Business Research*, 63(8), 849–855. doi:10.1016/j.jbusres.2009.07.003
- Chen, I. J., & Popovich, K. (2003). Understanding customer relationship management (CRM): People, process and technology. *Business Process Management Journal*, 9(5), 672–688. doi:10.1108/14637150310496758
- Cheng, L. Y., & Yang, C. W. (2013). Conceptual analysis and implementation of an integrated CRM system for service providers. *Service Business*, 7(2), 307–328. doi:10.1007/s11628-012-0160-z
- Coltman, T., Devinney, T. M., & Midgley, D. F. (2011). Customer relationship management and firm performance. *Journal of Information Technology*, 26(3), 205–219. doi:10.1057/jit.2010.39

- Coussement, K., & van den Poel, D. (2008). Integrating the voice of customers through call center emails into a decision support system for churn prediction. *Information & Management*, 45(3), 164–174. doi:10.1016/j.im.2008.01.005
- Crittenden, V. L., Peterson, R. A., & Albaumb, G. (2010). Technology and business-to-consumer selling: Contemplating research and practice. *Journal of Personal Selling & Sales Management*, 30(2), 103–110. doi:10.2753/PSS0885-3134300201
- Eichorn, F. L. (2004). Internal customer relationship management (IntCRM) a framework for achieving customer relationship management from the inside out. *Problems and Perspectives in Management*, 1(1), 154–177.
- Frow, P. E., & Payne, A. F. (2009). Customer relationship management: A strategic perspective. *Journal of Business Market Management*, 3(1), 7–27. doi:10.1007/s12087-008-0035-8
- Geib, M., Kolbe, L. M., & Brenner, W. (2006). CRM collaboration in financial services networks: A multi-case analysis. *Journal of Enterprise Information Management*, 19(6), 591–607. doi:10.1108/17410390610708481
- Hassan, R. S., Nawaz, A., Lashari, M. N., & Zafar, F. (2015). Effect of customer relationship management on customer satisfaction. *Procedia Economics and Finance*, 23, 563–567. doi:10.1016/S2212-5671(15)00513-4
- Heidemann, J., Klier, M., Landherr, A., & Zimmermann, S. (2013). The optimal level of CRM IT investments. *Electronic Markets*, 23(1), 73–84. doi:10.1007/s12525-012-0096-0
- Hennig-Thurau, T., Gwinner, K. P., & Gremler, D. D. (2002). Understanding relationship marketing outcomes: An integration of relational benefits and relationship quality. *Journal of Service Research*, 4(3), 230–247. doi:10.1177/1094670502004003006
- Hillebrand, B., Nijholt, J. J., & Nijssen, E. J. (2011). Exploring CRM effectiveness: An institutional theory perspective. *Journal of the Academy of Marketing Science*, 39(4), 592–608. doi:10.1007/s11747-011-0248-3
- Homburg, C., Muller, M., & Klarmann, M. (2011). When does salespeople's customer orientation lead to customer loyalty? The differential effects of relational and functional customer orientation. *Journal of the Academy of Marketing Science*, 39(6), 795–812. doi:10.1007/s11747-010-0220-7
- Jayachandran, S., Sharma, S., Kaufman, P., & Raman, P. (2005). The role of relational information processes and technology use in customer relationship management. *Journal of Marketing*, 69(4), 177–192. doi:10.1509/jmkg.2005.69.4.177
- Kalaighnam, K., & Varadarajan, R. (2012). Offshore outsourcing of customer relationship management: Conceptual model and propositions. *Journal of the Academy of Marketing Science*, 40(2), 347–363. doi:10.1007/s11747-011-0291-0
- Karakostas, B., Kardaras, D., & Papathanassiou, E. (2005). The state of CRM adoption by the financial services in the UK: An empirical investigation. *Information & Management*, 42(6), 853–863. doi:10.1016/j.im.2004.08.006
- Kasemsap, K. (2014a). The role of brand loyalty on CRM performance: An innovative framework for smart manufacturing. In Z. Luo (Ed.), *Smart manufacturing innovation and transformation: Interconnection and intelligence* (pp. 252–284). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-5836-3.ch010
- Kasemsap, K. (2014b). The role of social media in the knowledge-based organizations. In I. Lee (Ed.), *Integrating social media into business practice, applications, management, and models* (pp. 254–275). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-6182-0.ch013

- Kasemsap, K. (2014c). The role of social networking in global business environments. In P. Smith & T. Cockburn (Eds.), *Impact of emerging digital technologies on leadership in global business* (pp. 183–201). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-6134-9.ch010
- Kasemsap, K. (2015a). The role of customer relationship management in the global business environments. In T. Tsiakis (Ed.), *Trends and innovations in marketing information systems* (pp. 130–156). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-8459-1.ch007
- Kasemsap, K. (2015b). The roles of information technology and knowledge management in project management metrics. In G. Jamil, S. Lopes, A. Malheiro da Silva, & F. Ribeiro (Eds.), *Handbook of research on effective project management through the integration of knowledge and innovation* (pp. 332–361). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-7536-0.ch018
- Kasemsap, K. (2015c). The role of social media in international advertising. In N. Taşkıran & R. Yılmaz (Eds.), *Handbook of research on effective advertising strategies in the social media age* (pp. 171–196). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-8125-5.ch010
- Kasemsap, K. (2016). The roles of information technology and knowledge management in global tourism. In A. Nedelea, M. Korstanje, & B. George (Eds.), *Strategic tools and methods for promoting hospitality and tourism services* (pp. 109–138). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-9761-4.ch006
- Keramati, A., Mehrabi, H., & Mojir, N. (2010). A process-oriented perspective on customer relationship management and organizational performance: An empirical investigation. *Industrial Marketing Management*, 39(7), 1170–1185. doi:10.1016/j.indmarman.2010.02.001
- Keramati, A., Nazari-Shirkouhi, S., Moshki, H., Afshari-Mofrad, M., & Maleki-Berneti, E. (2013). A novel methodology for evaluating the risk of CRM projects in fuzzy environment. *Neural Computing & Applications*, 23(1), 29–53. doi:10.1007/s00521-012-1216-7
- Khodakarami, F., & Chan, Y. E. (2014). Exploring the role of customer relationship management (CRM) systems in customer knowledge creation. *Information & Management*, 51(1), 27–42. doi:10.1016/j.im.2013.09.001
- Kim, H., & Kim, Y. (2009). A CRM performance measurement framework: Its development process and application. *Industrial Marketing Management*, 38(4), 477–489. doi:10.1016/j.indmarman.2008.04.008
- Ku, E. C. S. (2010). The impact of customer relationship management through implementation of information systems. *Problems and Perspectives in Management*, 1(1), 154–177.
- Kumar, C., & Reinartz, W. (2006). *Customer relationship management: A database approach*. New York, NY: John Wiley & Sons.
- Matis, C., & Ilies, L. (2014). Customer relationship management in the insurance industry. *Procedia Economics and Finance*, 15, 1138–1145. doi:10.1016/S2212-5671(14)00568-1
- Mohammad, A. A., Rashid, B. B., & Tahir, S. B. (2013). Assessing the influence of customer relationship management (CRM) dimensions on organization performance: An empirical study in the hotel industry. *Journal of Hospitality and Tourism Technology*, 4(3), 228–247. doi:10.1108/JHTT-01-2013-0002
- Nevo, S., & Wade, M. R. (2010). The formation and value of IT-enabled resources: Antecedents and consequences of synergistic relationships. *Management Information Systems Quarterly*, 34(1), 163–183.

- Ngai, E. W. T., Xiu, L., & Chau, D. C. K. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Systems with Applications*, 36(2), 2592–2602. doi:10.1016/j.eswa.2008.02.021
- Noble, C. H., Sinha, R. K., & Kumar, A. (2002). Market orientation and alternative strategic orientations. *Journal of Marketing*, 66(4), 25–39. doi:10.1509/jmkg.66.4.25.18513
- Ramani, G., & Kumar, V. (2008). Interaction orientation and firm performance. *Journal of Marketing*, 72(1), 27–45. doi:10.1509/jmkg.72.1.27
- Reimann, M., Schilke, O., & Thomas, J. S. (2010). Customer relationship management and firm performance: The mediating role of business strategy. *Journal of the Academy of Marketing Science*, 38(3), 326–346. doi:10.1007/s11747-009-0164-y
- Reinartz, W. J., Karfft, M., & Hoyer, W. D. (2004). The customer relationship management process: Its measurement and impact on performance. *JMR, Journal of Marketing Research*, 41(3), 293–305. doi:10.1509/jmkr.41.3.293.35991
- Richard, J. E., Thirkell, P. C., & Huff, S. L. (2007). An examination of customer relationship management (CRM) technology adoption and its impact on business-to-business customer relationships. *Total Quality Management*, 18(8), 927–945. doi:10.1080/14783360701350961
- Saini, A., Grewal, R., & Johnson, J. L. (2010). Putting market-facing technology to work: Organizational drivers of CRM performance. *Marketing Letters*, 21(4), 365–383. doi:10.1007/s11002-009-9096-z
- Salim, F. S., & Keramati, A. (2014). How to make Iranian banks customer oriented with use of customer relationship management concepts? Methodology and comprehensive architecture. *Systemic Practice and Action Research*, 27(4), 377–395. doi:10.1007/s11213-013-9284-x
- Singh, S., Murthi, B. P. S., & Steffes, E. (2013). Developing a measure of risk adjusted revenue (RAR) in credit cards market: Implications for customer relationship management. *European Journal of Operational Research*, 224(2), 425–434. doi:10.1016/j.ejor.2012.08.007
- Smith, H. A., & McKeen, J. D. (2005). Developments in practice XVIII – customer knowledge management: Adding value for our customers. *Communications of the Association for Information Systems*, 16, 744–755.
- Spiller, J., Vlasic, A., & Yetton, P. (2007). Post-adoption behavior of users of Internet service providers. *Information & Management*, 44(6), 302–311. doi:10.1016/j.im.2007.01.003
- Stein, A. D., Smith, M. F., & Lancioni, R. A. (2013). The development and diffusion of customer relationship management (CRM) intelligence in business-to-business environments. *Industrial Marketing Management*, 42(6), 855–861. doi:10.1016/j.indmarman.2013.06.004
- Trainor, K. J. (2012). Relating social media technologies to performance: A capabilities-based perspective. *Journal of Personal Selling & Sales Management*, 32(3), 317–331. doi:10.2753/PSS0885-3134320303
- Trainor, K. J., Andzulis, J., Rapp, A., & Agnihotri, R. (2014). Social media technology usage and customer relationship performance: A capabilities-based examination of social CRM. *Journal of Business Research*, 67(6), 1201–1208. doi:10.1016/j.jbusres.2013.05.002
- Tsai, H. H. (2011). Research trends analysis by comparing data mining and customer relationship management through bibliometric methodology. *Scientometrics*, 87(3), 425–450. doi:10.1007/s11192-011-0353-6

- Tu, Y., & Yang, Z. (2013). An enhanced customer relationship management classification framework with partial focus feature reduction. *Expert Systems with Applications*, 40(6), 2137–2146. doi:10.1016/j.eswa.2012.10.028
- Tuzhilin, A. (2012). Customer relationship management and Web mining: The next frontier. *Data Mining and Knowledge Discovery*, 24(3), 584–612. doi:10.1007/s10618-012-0256-z
- Wang, M. (2013). An evaluation of customer relationship management in hospital-based and privately run nursing homes in Taiwan. *Total Quality Management & Business Excellence*, 24(9/10), 1004–1021. doi:10.1080/14783363.2011.637783
- Wei, J. T., Lee, M. C., Chen, H. K., & Wu, H. H. (2013). Customer relationship management in the hairdressing industry: An application of data mining techniques. *Expert Systems with Applications*, 40(18), 7513–7518. doi:10.1016/j.eswa.2013.07.053
- Winer, R. S. (2001). A framework for customer relationship management. *California Management Review*, 43(4), 89–105. doi:10.2307/41166102
- Yang, Y. (2012). Service capabilities and customer relationship management: An investigation of the banks in Taiwan. *Service Industries Journal*, 32(6), 937–960. doi:10.1080/02642069.2010.545394
- Ata, U. Z., & Toker, A. (2012). The effect of customer relationship management adoption in business-to-business markets. *Journal of Business and Industrial Marketing*, 27(6), 497–507. doi:10.1108/08858621211251497
- Awasthi, P., & Sangle, P. S. (2012). Adoption of CRM technology in multichannel environment: A review (20062010). *Business Process Management Journal*, 18(3), 445–471. doi:10.1108/14637151211232641
- Babakus, E., & Yavas, U. (2012). Customer orientation as a buffer against job burnout. *Service Industries Journal*, 32(1), 5–16. doi:10.1080/02642069.2010.545396
- Chen, Y., Li, P., & Evans, K. R. (2012). Effects of interaction and entrepreneurial orientation on organizational performance: Insights into market driven and market driving. *Industrial Marketing Management*, 41(6), 1019–1034. doi:10.1016/j.indmarman.2012.01.017
- Choo, H. J., Moon, H., Kim, H., & Yoon, N. (2012). Luxury customer value. *Journal of Fashion Marketing and Management*, 16(1), 81–101. doi:10.1108/13612021211203041
- Ernst, H., Hoyer, W. D., Krafft, M., & Krieger, K. (2011). Customer relationship management and company performance: The mediating role of new product performance. *Journal of the Academy of Marketing Science*, 39(2), 290–306. doi:10.1007/s11747-010-0194-5

ADDITIONAL READING

- Alam, A., Arshad, M. U., & Shabbir, S. A. (2012). Brand credibility, customer loyalty and the role of religious orientation. *Asia Pacific Journal of Marketing and Logistics*, 24(4), 583–598. doi:10.1108/13555851211259034
- Eskafi, M., Hosseini, S., & Yazd, A. M. (2013). The value of telecom subscribers and customer relationship management. *Business Process Management Journal*, 19(4), 737–748. doi:10.1108/BPMJ-Feb-2012-0016
- Guo, X., Hao, A. W., & Shang, X. (2011). Consumer perceptions of brand functions: An empirical study in China. *Journal of Consumer Marketing*, 28(4), 169–279. doi:10.1108/07363761111143169

- Han, Y., Yang, F. Z., & Wu, G. (2013). Applying data mining to the customer relationship management in retail trade. *Social Media Retrieval and Mining: Communications in Computer and Information Science*, 387, 116–124. doi:10.1007/978-3-642-41629-3_10
- Jamali, R., Moshabaki, A., Aramoon, H., & Alimohammadi, A. (2013). Customer relationship management in electronic environment. *The Electronic Library*, 31(1), 119–130. doi:10.1108/02640471311299173
- Komunda, M., & Osarenkhoe, A. (2012). Remedy or cure for service failure? Effects of service recovery on customer satisfaction and loyalty. *Business Process Management Journal*, 18(1), 82–103. doi:10.1108/14637151211215028
- Lukkari, P. (2011). Merger: Institutional interplay with customer relationship management. *Management Research Review*, 34(1), 17–33. doi:10.1108/014091711111096450
- Mainela, T., & Ulkuniemi, P. (2013). Personal interaction and customer relationship management in project business. *Journal of Business and Industrial Marketing*, 28(2), 103–110. doi:10.1108/08858621311295245
- Mittal, V., Frennea, C., & Westbrook, R. A. (2014). Managing customer value in business-to-business markets. *Marketing Review St. Gallen*, 31(3), 46–54. doi:10.1365/s11621-014-0362-y
- Mohebbi, N., Hoseini, M. A., & Esfidani, M. R. (2012). Identification & prioritization of the affecting factors on CRM implementation in edible oil industry. *Journal of Basic and Applied Scientific Research*, 2(5), 4993–5001.
- Moreno, G. A., & Melendez, P. A. (2011). Analyzing the impact of knowledge management on CRM success: The mediating effects of organizational factors. *International Journal of Information Management*, 31(5), 437–444. doi:10.1016/j.ijinfomgt.2011.01.002
- Munnukka, F., & Farvi, P. (2012). The price-category effect and the formation of customer value of high-tech products. *Journal of Consumer Marketing*, 29(4), 293–301. doi:10.1108/07363761211237362
- Nguyen, T. H., & Waring, T. S. (2013). The adoption of customer relationship management (CRM) technology in SMEs: An empirical study. *Journal of Small Business and Enterprise Development*, 20(4), 824–848. doi:10.1108/JSBED-01-2012-0013
- Padmavathy, C., Balaji, M. S., & Sivakumar, V. J. (2012). Measuring effectiveness of customer relationship management in Indian retail banks. *International Journal of Bank Marketing*, 30(4), 246–266. doi:10.1108/02652321211236888
- Pamsari, M. B., Dehban, M., & Lulemani, H. K. (2013). Assessment of the key success factors of customer relationship management. *Universal Journal of Management and Social Sciences*, 3(4), 23–29.
- Persson, A. (2013). Profitable customer management: Reducing costs by influencing customer behavior. *European Journal of Marketing*, 47(5/6), 857–876. doi:10.1108/03090561311306912
- Saarijarvi, H., Karjaluo, H., & Kuusela, H. (2013). Extending customer relationship management: From empowering firms to empowering customers. *Journal of Systems and Information Technology*, 15(2), 140–158. doi:10.1108/13287261311328877
- Sen, A., & Sinha, A. P. (2011). IT alignment strategies for customer relationship management. *Decision Support Systems*, 51(3), 609–619. doi:10.1016/j.dss.2010.12.014
- Wang, Y., & Feng, H. (2012). Customer relationship management capabilities: Measurement, antecedents and consequences. *Management Decision*, 50(1), 115–129. doi:10.1108/00251741211194903

Zahay, D., Peltier, J., & Krishen, A. S. (2012). Building the foundation for customer data quality in CRM systems for financial services firms. *Journal of Database Marketing & Customer Strategy Management*, 19(1), 5–16. doi:10.1057/dbm.2012.6

KEY TERMS AND DEFINITIONS

Customer: Individual that buys products or services.

Customer Loyalty: The likelihood of previous customers to continue to buy from a specific organization.

Customer Relationship Management: The management philosophy according to which the company's goal can be best achieved through the

identification and satisfaction of the customers' stated and unstated requirement.

Customer Retention: The assessment of the product or service quality provided by a business that measures how loyal its customers are.

Customer Satisfaction: The degree of satisfaction provided by the products or services of a company.

Information Technology: The set of tools, processes, and associated equipment employed to collect, process, and present the information.

Marketing: The social and managerial processes by which individuals or companies obtain what they need or want through creating, offering, and exchanging products with each other.

Marketing Strategy: The company's strategy that combines all of its marketing goals into practical results.

Implementing a Customer Relationship Management (CRM) System



Dimitra Skoumpopoulou
Northumbria University, UK

Benjamin Franklin
Northumbria University, UK

INTRODUCTION

A report published by Gartner shows that the worldwide CRM market grew 13.7% from \$18B in 2012 to \$20.4B in 2013. They anticipate that CRM usage will increase to \$36.5B worldwide by 2017, which is a significant increase from the \$20.6B forecasted for this year. However, research also suggests that 80% of CRM implementations fail (Rowley, 2002; Muhmin, 2012; Nguyen & Mutum, 2012). Although there has been extensive research into CRM implementation and why it often fails, studies predominantly concentrate on the technical performance of the implementation (Payne & Frow, 2005; Xu & Walton, 2005; Cox 2013), or data migration issues involved (Manjunath, Hegadi & Archana, 2012; Canter, 2013). Therefore, the purpose of this study is to research CRM implementation and the challenges involved in such complex implementations.

This paper will start by exploring the benefits of CRM systems and why companies decide to implement them. We will review the literature on the challenges that companies face during implementation and then we will examine a real-life implementation case study. Our conclusions will provide academics and practitioners with a clearer understanding of the CRM implementation and will summarise the main aspects that need to be taken into consideration during such complex implementations.

BACKGROUND

Clemons (2000), defines CRM from a customer's viewpoint and suggest that in essence a tiny proportion of a company's customers will generate the bulk of its profits, "identifying, collecting and keeping these clients is the very essence of customer relationship management" (Clemons 2000:25). However, according to Krauss (2002) a CRM implementation and strategy can only be deemed successful only if it can actually improve customer loyalty. One of the main reasons for failure is the reluctance to put CRM and the customer at the heart of the company. Girishankar (2000) proposes that for CRM systems to become successful companies must take a holistic approach to their implementation.

Research claims that corporate structure, employee involvement and management support are the key factors for the successful implementation of a CRM system (Galbreath & Rogers, 1999; Lindgreen, 2004; Pinto & Rouhiainen, 2001). Companies need to be connected and open to communication while management must be de-centralized and encourage subordinates to innovate since often employees are the ones that can steer the vision of a CRM system. A good project leader will set out a clear mission on what the CRM system will achieve, gain management support, schedule a clear implementation plan, communicate frequently with the end user, recruit

a successful well trained team, train employees successfully and gain end-user acceptance once the implementation is finished.

Siddiqi, Akhgar, and Wise (2011) agree that a clear report must be produced and analysed before a formula for success is produced because success depends highly on how the company views and aims to use the system in the first place. Organizations must choose between a customer focused system that allows a personalized service, or a company focused system that is mainly used for analysis and forecasting. It is likely that using a “soft” approach to the system, targeting specific customer in tailored ways will bring the best financial gains, and therefore employee motivation and support. Furthermore, Kotorov (2003) suggests companies must view CRM as a strategy rather than a technology solution.

Payne and Frow (2005) and Parvatiyar and Sheth (2001) believe that the lack of expertise and communication from IT departments is the most important factor in failure. Payne and Frow (2005) state that due to marketing and IT departments having no previous link, the lack of communication can lead to migration problems. This is a continuation of Parvatiyar and Sheth (2001) model, who suggest that in an enthusiasm to implement CRM, companies can hastily pass the implementation on to the in-house IT department. In order for a CRM implementation to be successful, employees must be involved in the design since the ineffective communication between IT and other departments leads to the implementation of a technically biased system, rather than one with wider benefits (Lindgreen, 2004; Pinto and Rouhiainen, 2001).

Xu and Walton (2005) state the reason for most failures is the lack of technology advancements and companies not pushing for greater analytical power from the systems. They believe that by being comfortable with what the providers currently offer, they are not harnessing CRM to its full potential. Cox (2013) agrees with this, believing that the technology on many systems is not being used to its full potential. However, he suggests that

this is due to the technology selection process, rather than the lack of potential technology. If a company selects an inferior product that does not benefit the employees then there will be issues.

Once the correct solution for the company has been picked, migrating data and information from previous systems is a large part of the implementation. Kostojohn, Paulen and Johnson (2011) are one of the few who studied data migration. They argue that correct migration can add tremendous value. They state that data migration does not cleanse data, and merely drags everything across from the old systems. In order for any implementation to be successful, data must be cleansed and organised. Similarly, Epicor (2005) emphasises the importance of thoroughly cleansing any data and an effective way to achieve this, is by employing data-administrators or specialists to help with the removal of duplicates.

Alongside the need for clean data migration, Kostojohn, Paulen and Johnson (2011) explore the need for adequate training. They state that in order for CRM to succeed, employees must be encouraged to use the system. This is a continuation of Reinartz, Krafft and Hoyer’s (2004) research that encourages relevant, well-planned and effective training that rewards employees for engaging in CRM activities. Becker, Greve & Albers (2009) also linked training directly to customer retention rates. Their research is one of the few that directly links components of the implementation plan to the outcome and success of a CRM implementation. They state that ineffective training will have a direct impact on the success of the system. Similarly, Zablah, Bellenger and Johnson’s (2004) research proposes that training can be used as a way to push companies’ strategies onto its employees. In order for implementations to prosper, training must be used so that all employees know how the company wants to use the system, its core principles and how it will benefit the customer and themselves.

There are limited case studies of CRM implementation that use literature and models to define success. Bull (2003) studied the implementation of a CRM system using the company ELMS as a

case study while Piskar and Faganel (2009) used a real-life case study of Sanolobar that offered direct insights into many of the problems faced during a CRM implementation.

The first issue with Sanolobar was that the IT managers had a clear vision of what they wanted the CRM system to achieve; namely to meet global competition and respond faster to market changes. This however was not explained clearly to their employees. Also, the company's manager saw little need for change while the employees had no previous knowledge of CRM systems so explaining the benefits to them was difficult. Piskar and Faganel (2009) state that the project manager plays a very important role in a CRM implementation. In this case the manager had no prior experience of a CRM system which led to a number of misjudgments and implementation errors.

Similarly, in the case study reported by Bull (2003) the manager lacked experience to produce a necessary evaluation of what type of CRM system the company needed. The lack of a strategic trained manager, led to the IT department working on an ineffective system which overstretched them and essentially led to the company not being able to use the CRM system in the desired way. Bull (2003) also explained that ineffective training had a knock on effect when striving to get employees on board with the CRM strategy. This created a fear of the CRM system within employees who now saw it as a cost cutting exercise. It was lack of communication and clear training that caused these issues.

MAIN FOCUS OF THE ARTICLE

Qualitative research has often been criticized for not being guided by theory in its development. However Mertz and Anfara (2006) disagree with this, suggesting that combining both the theory and the researcher's experiences can only benefit the research.

The knowledge gained through qualitative investigations is more informative, richer and offers enhanced understandings while the holistic nature of the research means that any question or answer can be explored and then analysed. Tewksbury (2009) states that qualitative methods gather how people understand, experience and operate within a certain parameter.

In order to gain maximum credible qualitative data the authors decided that conducting interviews was the best method. There are three types of interviews: unstructured, structured and semi-structured. Bloom and Crabtree (2006) suggest that no interview can be fully unstructured, but inevitably arise from the participants' reactions and environment. Structured interviews are particularly suitable where it is necessary to capture precise data, and correspond more to quantitative research. Although they tend to be quick to undertake they place a rigidity that is not beneficial. Our research has no use for pre-established categories; and as such structured interviews did not fit the research. We therefore decided to use semi-structured interviews.

Having established a semi-structured method, the authors were keen to collect data that was not biased and therefore elected to use the literature review to formulate the approach. In doing this, the authors split the question structure into sections:

- **Technology:** The authors used Payne & Frow (2005), Xu & Walton (2005), Parvatiyar & Sheth (2001) and Kostojohn, Johnson and Paulen's (2011) research to develop questions. The main line of inquiry was to establish if the product was the correct one for the company and offered everything that the participants needed.
- **Leadership/Communication/Structure:** The questions in this section used the research of Galbreath and Rogers (1999), Lindgreen (2004) and Pinto and Rouhiainen (2001) as basis. They aimed to understand the importance of effective leadership and

how a company's structure plays a part in the implementation.

- **Taking an Holistic Approach and Personal Views:** This section concentrated on the motivation of the employees and how the company uses the system based on the research of Kotorov (2003) and Girishankar (2000). This part aimed to explore the employees' perceptions about the system and how it was used.
- **Training and How This Affected Motivation:** Throughout the literature, the importance of training was prominent (Reinartz et al. 2004; Becker, Greve & Albers, 2009). The questions within this section aimed to explore the training offered; how it was tailored to individuals and the effect it had on motivation to use the new system.
- **General:** The last three questions aimed to sum up the interview, working chronologically on the most important parts of the implementation from an employees' perspective.

We interviewed 10 people from a technology company that for anonymity purposes we will be referring as Company X. Table 1 shows the interviewees' codes and job titles. Once the interviews had been completed and recorded the

Table 1. Interviewees' details

Code Name	Job Title
Participant 1	Operation Support
Participant 2	Software Support
Participant 3	Software Support
Participant 4	Technical Support
Participant 5	Operation Support
Participant 6	Software Support
Participant 7	Software Support
Participant 8	Operations & Technical lead
Participant 9	Technical Support
Participant 10	Software Support

authors transcribed them. The authors had already developed an understanding of the main issues with CRM and therefore transcribing the data allowed clear and thorough analysis.

King (2012) states that the choice of method of data analysis needs to be guided by the methodological position of the piece of research, in such King suggests that template analysis encourages the researcher to be explicit in their analysis as every conclusion can be traced back to the research. Template analysis, in essence gives the researcher a way of ordering the data into set codes. The authors used the initial five sections discussed above (Technology, Leadership/Communication/Structure, Taking an Holistic Approach & Personal Views, Training and How This Affected Motivation, General) in order to prepare the data. Once an initial analysis had been conducted the researchers were able to update the codes and analyse further. Crabtree and Millar (1999) suggest that by using a second set of codes it decreases the risk of missing valuable data, and allows the researcher to look beyond the initial codes.

SOLUTIONS AND RECOMMENDATIONS

Company X is a technology company and in 2013 they undertook the decision to change the company's CRM system from System A to System B (for ethical reasons we will be using code names for the systems). The UK was used as a tester region and System B was implemented as a pilot in this region. Similarly to many other companies Company X does not use System B as a sole CRM system which means that they also use other management and forecasting tools.

Success of Implementation

Often it is challenging to understand the intangible benefits that CRM systems offer, thus it is very difficult to measure their success. In order to discover how effective Company X has been,

the authors will use the employee's reaction to the question how successful has the implementation been? Participant 1 believes "it has gone very well" Participant 2 agrees, stating that "everything ran quite smoothly". However, Participant 3 and Participant 4 are more negative suggesting that "it wasn't successful in the slightest, too slow and buggy" and "not successful at all" respectively. It is interesting that the rest of the participants suggest a change in opinion overtime, Participant 5 suggests there is "room for improvement", Participant 6 that "I wouldn't say it was unsuccessful, it certainly wasn't without fault" and Participant 7 that "Overtime it is becoming more successful". These varied answers suggest that there were ways in which Company X could have improved the roll-out and the following analysis will decipher the most important parts.

Leadership

Throughout the literature a strong emphasis has been placed on having a strong leader to guide the implementation. Pinto & Rouhiainen (2001) state that a good project lead will ensure that communication is frequent with the end user. It is interesting that when probed on who was leading the implementation only Participant 4 & Participant 8 had an idea of who was the leader whilst the rest were "not aware" of any leadership. System B lacked a clear focal point from the outset, and from the employee's point of view, this instantly placed strain on the implementation.

Selecting the Solution

Xu and Walton (2005) state that if a company selects an inferior product that does not benefit the employees then this could potentially create a lot of issues jeopardizing the successful introduction of the system. This is evident in our case with participants often unhappy with the features offered or the technical performance of the system. In every interview there were references to changing certain aspects of the CRM system to enable

it to be more effective for employees. Over half suggest the need for "offline capabilities", whilst Participant 3 and Participant 7 talk negatively about the "lack of personalised features that make the system unique to each user". However, the lack of functionality is not as serious as the usability and performance issues. Participant 6 describes the system as "excruciatingly slow", with simple tasks taking "up to an hour longer per day" and Participant 2 explains hang-time of "up to 15 seconds loading per page". It is the usability of the system that is hindering the successfulness of the implementation and Participant 9 suggests that the system's inconsistency means that they "never actually want to look at it".

Employee Involvement

One of the main reasons for the technical issues is the lack of involvement from employees who use the system daily. Parvatiyar and Sheth (2001) state that although technical workers may build a system that is technically sound, it may not be effective for front-line employees. Participant 2 recognised the lack of involvement as a "big failure". Overall the front-line employees feel that if they had been included more in the design the implementation would have been more successful.

Furthermore, six months into the project the leadership team decided to take some of this feedback on board and set up an Ambassador program. In essence it was a selective group of end users who were given extra training and guidance on the system. The process would work two ways; firstly the developers would gain valuable insight and feedback, and secondly to create a taskforce of skilled users who act as advocates. Employees were shown "functions that were coming, however these were not available so it was difficult to get enthusiastic, felt that queries were being listened to but there was no improvement". It is interesting that when asked what were the main issues of the implementation Participant 5 stated that "I would want more input from the people on the ground, people who use it every day" and Participant 4

stated that “I would want a seller task force, who are the people who are going to be using it, and what do they need in the system. I would ensure that they were fully trained and became advocates.” It is worrying that although that the new system was set up to “spread the word” (Participant 8) six out of the ten participants were “not aware” that it existed. This would suggest a lack of communication throughout the project.

Communication

Galbreath and Rogers (1999) suggest the need for the company to be open to communication and Bull (2003) blamed the fear of the CRM system in their study on a lack of communication. When asked whether there was communication, all but one participant recognised that emails were sent regularly; the issue came with the quality of what was being sent. Participant 3 stated that they “felt that these were put in technical language rather than in a way that we could use” and Participant 7 felt that “answers were difficult to find”. Participant 4 described the dangers of giving communication that employees cannot relate to: “there were so many updates and bug fixes that you lost interest and didn’t take them seriously. Motivation was affected and it was used as an excuse not to use it (System B)”.

Managerial Support

Kotorov (2003) states that in order for an implementation to be successful it needs to have support from the highest echelon in the company. When asked what they believed to be the main issues, Participant 9 stated “we need to be on the same page and we are not at the moment, from senior leadership right down to the bottom users”. Ultimately, sending out communication that is untruthful holds back the improvement of the product. Employees become unengaged and demotivated, feedback becomes muted, and essentially the system is used for what the developers have developed, rather than what suits the

employees. Participant 7 identified that they “do not think” there is any encouragement in using the new system and Participant 5 reported that “as long as it is clean, not fussed about System B”. This would suggest that at lower-manager level the system is not being accepted, and other in house systems are being used instead.

Heart of the Selling Process

Girishankar (2000) proposes that for it to become successful the company must place CRM at the heart of everything it does. In Company X’s case this would mean removing other in house reporting systems to leave one customer-facing tool. This is widely recognised by the interviewees as a major issue with the implementation. Within every interview each participant states the need for the system to become more “collaborated”. Participant 4 said “Company X has several other systems that marketing and sellers can use which means that we don’t get the main benefits out of a CRM”. Participant 2 suggests that, “coming from an organization where everything was inter-connected and there was one system that records all, to a place where you have 3 or 4 systems seems odd. I think there should be integration”. This affects motivation to use it, but more importantly affects benefits that the customer can receive. Ultimately Company X has created a solution that misses the fundamental point of a modern CRM system.

Company X’s Structure

In order for a CRM implementation to be successful the structure of Company X would need to be “open to change”. The participants had varied responses to whether change was needed, and ultimately how the new system compared to the old system A. Although Participant 4 believes that “System A was an outdated first generation and needed updating”, on the other hand, Participant 6 didn’t “know why we moved from System A”. This again links to the communication and leadership. If Company X had allowed every user to

know the reasons for change then the transition may have been smoother. In order to achieve success Galbreath & Rogers (1999) argue that a company's structure must encourage subordinates to innovate. In our case study it is evident that it is not the employees that need encouragement to innovate, but perhaps the company that needs to improve.

Training

The literature presents a strong need for the training on any new system to be extremely important and it appears that Company X have had real issues in this area. When asked to comment on the effectiveness of the training all but one of the participants talked negatively. Participant 7 stated "It was a waste of time, so I taught myself", Participant 9 "I didn't think we were trained enough, I wanted to use it but ended up being frustrated", Participant 2 suggested "the reason I did not know what I was doing is because the training was not great". Most importantly, Participant 6 linked the poor training directly to motivation "It's hard to get motivated about something you don't fully understand". Kostojohn, Paulen & Johnson (2011) and Reinartz et al. (2004) recommend that one way of improving training is to offer tailored guidance specific to employee's roles. Participant 3 recognised the lack of individuality stating that the training was just "several people on a webinar looking at an automated screen, no personalisation". Company X will need to improve this moving forward and use training as a way of engaging employees. They could use this to explain the need to change from System A and create a clear and consistent message to its employees.

Data Migration

A major part of implementations that often gets less limelight is the migration of data from the old system. It appears again that Company X had varying success with this. Participant 2, Participant 7 and Participant 1 suggest there were "no issues",

Participant 3 recognised there were "teething problems" and Participant 10 suggested that there were "initial issues that were all fixed". Participant 4 and Participant 6 however were affected massively by ineffective migration. Participant 6 argued, "data did not transfer across and deleted entries. I had to re-add opportunities" this added "a huge increase in time and frustration". Participant 4 stated "there were too many problems, when it first came across I stopped using the system due to the data being all over the place". This suggests extreme failure in parts of the migration with employees using this as an excuse not to use the system.

FUTURE RESEARCH DIRECTIONS

It is evident that although this research brings together various researches, the study of CRM is still relatively uncommon. CRM systems are continuing to grow and as the technological age advances, these systems will play a bigger role. Our case study is unique due to the size and scale of the project; which has brought up problems that perhaps would not be faced in smaller organizations. The authors would suggest applying the ten factors for success identified in this paper in other companies in order to further investigate the more important issues involved during CRM implementation. It will be useful and very interesting to conduct further case study analysis in companies of various size and various industries. The authors believe that CRM systems will be a vital part of the vast majority of businesses in a few years and therefore more research is needed to further understand these complex systems. It is disappointing that in a world where technology is so impressive; much of the power of CRM is being lost.

CONCLUSION

Through reviewing the literature and by contacting a primary case study, it is evident that there

are many variables that may influence the success or otherwise of CRM implementation. Listed and explained below are the ten key points that the authors believe must be in place:

- **The Need for Change:** The Company needs to have a genuine reason for implementing a system, which will benefit both the customer and the employees. This must be explained clearly to the people who will be affected.
- **A Beneficial Corporate Structure:** The structure of the company must be open to this change, and any communication that comes with it. It must have the resources in place, and it is beneficial if it encourages innovation.
- **Communication:** There must be clear and truthful communication throughout. This needs to be put in a language that can be understood by the whole company, allowing all levels of the organisation to establish a shared understanding.
- **Leadership:** A strong leader must be in place with accountability for the implementation. The leader will be experienced with CRM systems, know how to get the most out of them, motivate employees around them and create a solid technical team.
- **Selecting the Correct Solution:** The solution must meet the company's requirements, and add genuine value to what is already there. It must be tested thoroughly with a correct sample size, and must be able to work independently before any parallel running is removed.
- **Using as a Modern CRM System:** The solution must be placed at the heart of the company; it is to be used as the focal point for all customer interaction, adding benefit rather than analysis. The system must be inter-departmental, encompassing the whole organisation.
- **Employee Involvement:** Front line employees and end users must be consulted from the start and throughout. Feedback from the system users will be invaluable in order to know the way the system is used and what is really needed.
- **Top-Level Support:** The company's leaders must offer support and resources to the IT and development team. Their message must be one of resilience; this will filter down through the rest of the company. Timescales must be set to allow time for effective implementation.
- **Data Cleansing:** It is important to ensure before data is migrated to any new system that it is cleansed and data-administrators put in place. This will improve efficiency and subsequent employee support.
- **Training:** Effective training is a key aspect to any successful implementation. Training should focus on the usage and purpose of the new system in a companywide context. It is essential that training takes place on fully operating systems and is focused on the roles of the individuals participating.

It is important to obtain feedback from employees and where possible customers to identify further improvements to the system, taking it to its full potential. The benefits that CRM systems offer mean that customers should gain an improved service. By not harnessing constant improvement, success rates will fall. This research provides practitioners a general overview to the problems faced, and potential solutions to many issues.

We created a strategy or model that would allow practitioners a collaborated view to ensure future success in CRM implementations. The authors brought together the points found in the literature review, as well as in our analysis and findings in order to create ten factors that can help organisations to realise the full potentials of CRM systems. This is not available in published articles and in such, places a genuine need and importance for our study.

REFERENCES

- Ahearne, M., Rapp, A., Mariadoss, B. J., & Ganesan, S. (2012). Challenges of CRM Implementation in Business-to-Business Markets: A Contingency Perspective. *Journal of Personal Selling & Sales Management*, 32(1), 117–130. doi:10.2753/PSS0885-3134320110
- Anshari, M., & Almunawar, M. N. (2011). Evaluating CRM Implementation in Healthcare Organization. *Proceedings of 2011 International Conference on Economics and Business Information*.
- Becker, J., Greve, G., & Albers, S. (2009). The Impact of Technological and Organizational Implementation of CRM on Customer Acquisition, Maintenance, and Retention. *International Journal of Research in Marketing*, 26(3), 207–215. doi:10.1016/j.ijresmar.2009.03.006
- Bloom, B., & Crabtree, B. (2006). The Qualitative Research Interview. *Medical Education*, 40(4), 314–321. doi:10.1111/j.1365-2929.2006.02418.x PMID:16573666
- Bull, C. (2003). Strategic issues in customer relationship management (CRM) implementation. Manchester Metropolitan University.
- Canter, J. (2013). *Successful CRM Hinges On Data Quality*. Retrieved 27, 2014, from Destination-CRM: <http://www.destinationcrm.com/Articles/Web-Exclusives/Viewpoints/Successful-CRM-Hinges-On-Data-Quality-41571.aspx>
- Chen, I., & Popovich, K. (2011). *Understanding customer relationship management (CRM): People, process and technology*. MCB UP Ltd.
- Clemons, E. (2000). *Gathering the Nectar, Understanding CRM*. Retrieved 27, 2014, from Financial Times: <http://citeseer.uark.edu:8080/citeseerx/showciting;jsessionid=D34C23EF616DAFC5D3B3698F96F870C3?cid=822690>
- Cox, A. (2013). *Framework for successful CRM implementation*. Epiphany Information.
- Crabtree, B., & Millar, W. (1999). *Doing Qualitative Research*. London: Sage Publications.
- Epicor. (2005). *Strategies for a Successful CRM Implementation*. Epicor.
- Galbreath, J., & Rogers, T. (1999). *Customer relationship leadership: a leadership and motivation model for the twenty-first century business*. The TQM Magazine. MCB UP Ltd.
- Girishankar, S. (2000). Companies Want CRM tools to manage business relationships. *Information Week*, 17(April), 65.
- King, N. (2012). *Template Analysis*. Retrieved March 17, 2014, from University of Huddersfield: http://hhs.hud.ac.uk/w2/research/template_analysis/index.htm
- Kostojohn, S., Johnson, M., & Paulen, B. (2011). *CRM Fundamentals* (Vol. 1). New York: Apress. doi:10.1007/978-1-4302-3591-0
- Kotorov, R. (2003). Customer relationship management: Strategic lessons and future directions. *Business Process Management Journal*, 9(5), 566–571. doi:10.1108/14637150310496686
- Krauss, M. (2002). *At many firms, technology obscures CRM*. Retrieved 27, 2014, from Marion: <http://marionpartners.com/publications/marketingnews/mkn03182002.htm>
- Lindgreen, A. (2004). The design, implementation and monitoring of a CRM programme: A case study. *Marketing Intelligence & Planning*, 22(2), 160–186. doi:10.1108/02634500410525841
- Manjunath, T., Hegadi, R., & Archana, R. A. (2012). A study on Sampling techniques for data testing. *International Journal of Computer Science and Communication*, 3(1), 13–16.
- Mertz, N., & Anfara, V. (2006). *Theoretical Frameworks in Qualitative Research* (Vol. 1). SAGE Publications.

- Muhmin, A. G. A. (2012). CRM technology use and implementation benefits in an emerging market. *Journal of Database Marketing & Customer Strategy Management*, 19(2), 82–97. doi:10.1057/dbm.2012.13
- Nguyen, B., & Mutum, D. S. (2012). A review of customer relationship management: Successes, advances, pitfalls and futures. *Business Process Management Journal*, 18(3), 400–419. doi:10.1108/14637151211232614
- Parvatiyar, A., & Sheth, J. N. (2001). *Customer Relationship Management: Emerging Practice, Process, and Discipline*. *Journal of Economic and Social Research*.
- Payne, A., & Frow, P. (2005). *A Strategic Framework for Customer Relationship Management*. American Marketing Association. *Journal of Marketing*, 69(4), 167–176. doi:10.1509/jmkg.2005.69.4.167
- Pinto, J., & Rouhiainen, P. (2001). *Building Customer-Based Project Organizations*. New York: John Wiley & Sons.
- Piskar, F., & Faganel, A. (2009). *A Successful CRM Implementation Project in a Service Company: Case Study*. Sanolabor.
- Reichheld, F. (2001). *The Loyalty Effect: The Hidden Force Behind Growth, Profits, and Lasting Value*. Harvard Business Press.
- Reinartz, W., Krafft, M., & Hoyer, W. (2004). The Customer Relationship Management Process: Its Measurement and Impact on Performance. *JMR, Journal of Marketing Research*, 41(3), 293–305. doi:10.1509/jmkr.41.3.293.35991
- Rowley, J. (2002). Eight questions for customer knowledge management in e-business. *Journal of Knowledge Management*, 6(5), 500–511. doi:10.1108/13673270210450441
- Siddiqi, J., Akhgar, B., & Wise, T. (2011). *A framework for the implementation of a Customer Relationship Management strategy in retail sector*. Sheffield University.
- Tewskbury, R. (2009). Qualitative versus Quantitative Methods: Understanding Why Qualitative Methods are Superior for Criminology and Criminal Justice. *Journal of Theoretical and Philosophical Criminology*, 1.
- Xu, M., & Walton, J. (2005). *Gaining customer knowledge through analytical CRM*. Emerald Group Publishing Limited.
- Zablah, A., Bellenger, D., & Johnston, W. (2004). Customer Relationship Management (CRM) Implementation Gaps. *Journal of Personal Selling & Sales Management*, 24, 279–295.

ADDITIONAL READING

- Balaram, A., & Adhikari, B. (2010). Managing Customer Relationships in Service Organizations. *Administration and Management Review*, 21(2), 65–78.
- Crockett, B., & Reed, K. (2001). *Three Approaches to Customer-Centric Understanding*. Retrieved 27, 2014, from Revenue Performance: <http://mthink.com/article/three-approaches-customer-centric-understanding/>
- Davenport, T. H., Harris, J. G., & Kohli, A. K. (2001). How do they know their customers so well? *MIT Sloan Management Review*, 42(2), 63–73.
- Dyché, J. (2001). *The CRM Handbook: A Business Guide to Customer Relationship Management* (1 ed.). Addison-Wesley Professional.
- Finnegan, D., & Currie, W. (2010). A multi-layered approach to CRM implementation: An integration perspective. *European Management Journal*, 28(2), 153–167. doi:10.1016/j.emj.2009.04.010

Henneberg, S. (2003). *CRM Implementation: Hard Choices and Soft Choices*. Bath University. Henneberg.

Kim, S.-Y., Jung, T.-S., Suh, E.-H., & Hwang, H.-S. (2006). Customer segmentation and strategy development based on customer lifetime value: A case study. *Expert Systems with Applications*, 31(1), 101–107. doi:10.1016/j.eswa.2005.09.004

Kumar, V., & Reinartz, W. (2012). *Customer Relationship Management: Concept, Strategy, and Tools* (Vol. 2). New York: Springer Heidelberg. doi:10.1007/978-3-642-20110-3

Light, B. (2001). *A Review of the Issues Associate with Customer Relationship Management Systems*. University of Salford.

Osarenkhoe, A., & Bennan, A.-E. (2007). An exploratory study of implementation of customer relationship management strategy. *Business Process Management Journal*, 13(1), 139–164. doi:10.1108/14637150710721177

Peppard. (2000). Customer Relationship Management (CRM) in Financial Services. *European Management Journal*, 18 (3), 312–327

Sin, L., Tse, A., & Yim, F. (2005). CRM: Conceptualization and scale development. *European Journal of Marketing*, 39(11), 1264–1290. doi:10.1108/03090560510623253

Unit, E. I. (2007). *Improving customer relationships: an integrated approach*. SAP. London: SAP.

KEY TERMS AND DEFINITIONS

Case Study: An up-close, in-depth, and detailed examination of an organization.

Communication: The means of sending or receiving information.

Customer Relationship Management (CRM): Identifying, collecting and keeping clients.

End Users: The people who actually use a particular system.

Implementation: The process of putting a plan into action.

Leadership: Leading a group of people or an organization.

Training: The action of learning a new skill.

The Intelligence of E-CRM Applications and Approaches on Online Shopping Industry

Bashar Shahir Ahmed

University Abdelmalek Essaadi (LEROSA Laborator), Morocco

Fadi Amroush

Universidad de Salamanca, Spain

Mohammed Ben Maati

University Abdelmalek Essaadi, Morocco

INTRODUCTION

Many organizations are implementing electronic Customer Relationship Management systems in an effort to secure and maintain competitive advantage. Developing and enhancing long-term relationships with customers requires using quality data about the customers. Unfortunately, many e-CRM systems have failed to achieve their objective because of the issues associated with data integrity (Chen & Ling, 2008). The integrity of the data used to make decisions regarding customers in an organization affects the quality of the decisions made.

Protecting data integrity is challenging particularly in big organizations. The challenges associated with data integrity are increased by the complexity of the current business environment. Furthermore, marketing strategies and programs usually change rapidly in organizations (Fan, W. et-al, 2011). Therefore, to get the most out of relationship marketing, the customer data has to conform to the standards in a given industry. In addition, the data should be time stamped to make its validity clear. In addition, the data should be complete and accurate by having all the required values correctly entered. Finally, it is important for the customer data to be consistent (Chen &

Ling, 2008). This means that the data values should be aligned across all systems in an organization.

BACKGROUND

Various studies have been done before by a number of scholars as far as the quality and accuracy of data in e-CRM systems are concerned. Research on optimising the value of e-CRM application in E-commerce aims at determining the responsive prospect of customers to E-commerce applications (Ahmed, Maati & Al Mohajir, 2015a). The study revealed that customers often become frustrated if the e-CRM system is redundant and incapable of meeting their needs and expectations. A revelation from the study is that quality of data goes far deeper to include ergonomic relationships (Ahmed, Maati & Al Mohajir, 2015a, pg. 150). As such, any design or approach to improve data quality must measure and ascertain that the implementation of an e-CRM system meets the ergonomic conditions. With the study aiming to determine whether potential customers can respond to the e-CRM application, the researchers established that culmination of the application should deal with all customer-related issues. Such issues include customer services facilitation, sales and

marketing, and field support (Ahmed, Maati & Al Mohajir, 2015a, pg. 152).

In addition to data quality issues, companies embracing e-CRM systems face difficulties in developing effective technological infrastructure considering the limited time and resources (Ahmed, Maati & Al Mohajir, 2014, pg. 214). A recent study revealed that 66 percent of loaded orders are discarded before checkout, and only 5 percent of customers who visit organisations' online stores become customers (Ahmed, Maati & Al Mohajir, 2014, pg. 214). The aim of the research was to show that E-CRM is not limited to internet data but includes other devices such as phones, set-top boxes, pagers and so forth.

In another research by same authors, Ahmed, Maati, and Al Mohajir on how to improve the E-CRM intelligence by using CRM data analytic tools such as OLAP, data mining and web analytics, it was established that CRM analytics tools not only contribute to exceptionally productive customer relationship as regards sales and service delivery but in the development of adverts, planning, and the analysis of customer data as well (Ahmed, Maati & Al Mohajir, 2015b, pg. 7). As such, it is evident that real experience for an online client is dependent on an intelligent, concise, and convenient application.

In another study, authors targeted to a larger extent the problems resulting from formatting errors, data inconsistencies and lexical errors. A good example is the token based and attribute selection algorithms (Z. Yuhang, 2010). The algorithms allow for eliminating duplicate records data in databases. The algorithms' implementation is in such a way that users pick on the source and the preferred mode of repair. This functionality is realized courtesy of the application of distance function on the attributes that are unclear. In these algorithms, the repair time and progress is visible (Z. Yuhang, 2010). Yuhang suggested a new technique for cleaning using the frequency of words appearance. In clustering texts, C- cluster-

ing algorithms were used. It was a very effective technique in the reducing of tendencies of doing a mistake whenever a word that is rare is to be filtered (Nguyen & Mutum, 2012). With the help of the open source tool called Febrl, new algorithms and techniques for record linkage can be developed. It possess the ability to allow for varied formats of files and has a number of methods for encoding.

Problem Statement

CRM systems benefits organisations in a number of ways. The systems enhance coordination of internal processes and organizations' responsiveness to customers' needs. Data integrity is however an issue of major concern as far as e-CRM systems are concerned. In the current business environment, data is available from a lot of sources (Chen & Ling, 2008), in large quantities and huge complexities. According to a recent research by Bluewolf Consulting, data quality is a major issue of concern in e-CRM systems. According to the survey conducted by Bluewolf Consulting, 38 percent of the respondents were not confident in the accuracy of their instance's data.

A different research revealed that data quality problems cost an organization around \$ 100 per dirty record. The \$ 100 per dirty record includes costs such as the cost of printing and mailing to bad address, the losing disgruntled customer, sales conflict over the same lead and incorrect marketing segmentation and personalization. It is also important to point out that bad data is one of the three reasons why CRM projects fail. This is because bad data results in misleading, incomplete and confusing information. Data quality issues are therefore important when implementing e-CRM systems and can lead to limited acceptance or outright failure of the systems if measures are not put in place to enhance data integrity (Vitt, 2014). In this paper, proposed framework and algorithm for enhancing data quality in e-CRM systems is presented.

Literature Review

The goal of integrating business intelligence systems with E-CRM tools and techniques is very valuable to organizations today. It is due to the fact that this integration allows organizations to gather information on the customers, by applying sophisticated tools on the available data. This information can be accumulated from various media channels or points of contact that can be used as a communication node. These communication nodes are usually inclusive of the organization's contact number, websites and several other online forums that enable live chatting, direct mails, advertising materials and social media. Hence, the business intelligence systems are used to enhance the applications of the CRM systems by making use of datasets, which contains the customer's background and information such as; customer's personal information, purchase records, buying preferences and other concerns, to aid in decision making (Batenburg & Versendaal, 2007; Botha, Bothma, & Geldenhuys, 2008).

Therefore, the audacity of the E-CRM applications to acquire a robust system that assimilates such massive customer data, while also delivering the interactive tools to make the business process much more efficient and automated requires the integration of business intelligence systems. Hence, in an aftermath several E-CRM systems integrated with the business intelligence systems have been designed to accommodate the organizational needs (Busaidi, 2013; Chen & Lin, Enhancing the relationship benefit to develop relationship value, 2009).

Since the conceptual framework of integrating E-CRM and business intelligence systems, it has become the most robust technology related topic. In accord with the economists, CRM has not been a new notion, in fact with the current development and advancements in its application by integrating information technology CRM has extended more practically. The use of E-CRM is now done mainly for the marketing purposes to infuse the goal of taming the long term sustain-

ability of the customers by wandering away from the product-centric marketing services is a key driver that is profitable for the organizations. However, the use of business intelligence systems can make E-CRM more effective by tailoring the organizations decision on a particular product or service, thus boosting the customer portfolio (Chen & Ling, 2008).

With the increased competition circumventing the global industry, the need for E-CRM systems is getting a lot of cool. According to many organizations, the E-CRM systems must be designed to refine the organizational operations from top to bottom mainly focusing on the organizational endeavors incorporated in all the departments. It is noteworthy that the integration of business intelligence systems and E-CRM have aided in the manufacturing process, product testing along with its assembling, purchasing, and billing. Thus, an E-CRM system has become more effective and is beneficial for almost every organizational department, such as; human resources, marketing, sales and engineering (Chua, 2011; Fjermestad & Romano, 2006).

However, the integrated system have increased the complexity of E-CRM applications as the market competition has been escalated and organizations searches for sophisticated techniques that includes deep data mining of the customer data that is retrieved from all the touch points of the customer to accomplish their target. The accumulated data allows organizations to create and enable the holistic view of their customer profile (Anderson & Kerr, 2002). Moreover, the additional benefit to the organization is that it is now proficient at determining its target customers and therefore, predict the trends of their future purchases, which allows them to alter their marketing strategies accordingly (Horwitz, 2014; Jones & Ranchhod, 2007).

The integrated business intelligence and E-CRM systems have created an opportunity for the effective approach that seamlessly interlinks the sales, marketing, customer service, support and other operations concerned with customers.

Hence, the integrated system is centralized to keep important and profitable customers remain dedicated to the organization with reduced cost, increased quality and high revenues. Thus, the modern customer relationship management concept was shaped and motivated by the notions of quality management. In a lot of organizations today, E-CRM is considered as a technological solution that forms up an individual database and empowers automation in sales and marketing strategies and operations of the organization in order to improve their focused efforts (Raab, Ajami, V., & Goddard, 2008).

From customer's point of view, the E-CRM systems offer customization and ease of implementing transactions irrespective of the type of channel used for the organizational and customer interaction. A lot of businesses have realized the standing worth of business intelligence systems and E-CRM with its potential to assist the organization's goal and endure a striking competitive edge in the global competitive market. Due to various, yet frequent alterations and variations in the global environment and competition, organizations cannot compete constructively with inconsequential advantages and tricks that can easily be replicated by the competing firms (Gosney & Boehm, 2000). Hence, they need something new and innovative to deal with. The integration and implementation of business intelligence systems and E-CRM is an opportunity to increase the advantages for the organizations and enable them to focus more on the development of actual relationships between

them and their customers. Successful organizations always deliver and cater for their customers' needs. In general terms, the integration of business Intelligence systems and E-CRM tends to provide a futuristic approach for the organizations by improvising their decision making capabilities (Peppers & Rogers, 2011).

Methodology

The E-CRM systems when integrated with the business intelligence systems provides ubiquitous features that have the ability to deliver sources of customer credentials at much lower costs than the traditional network technologies. In an organization, the E-CRM systems can be used to hoard, store, preserve, and distribute customer knowledge on a particular product or service. However, it is noteworthy that the effective administration of the information plays a very important role in using E-CRM, as it can be used for product modification, service and innovation, while keeping under consideration the views of customers (Lamb, Hair, & McDaniel, 2011).

To improve the business process it is very much essential to first improve the quality of the accessible data as depicted in Figure 1. However, to improve the data quality a proposed framework and algorithm should be implemented. The following section discusses the framework and algorithm that can improve the data quality for E-CRM systems.

Figure 1. The flow chart to improve decision making skills of an organization

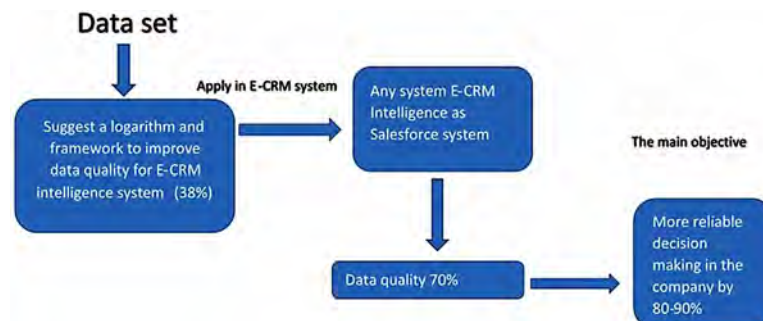


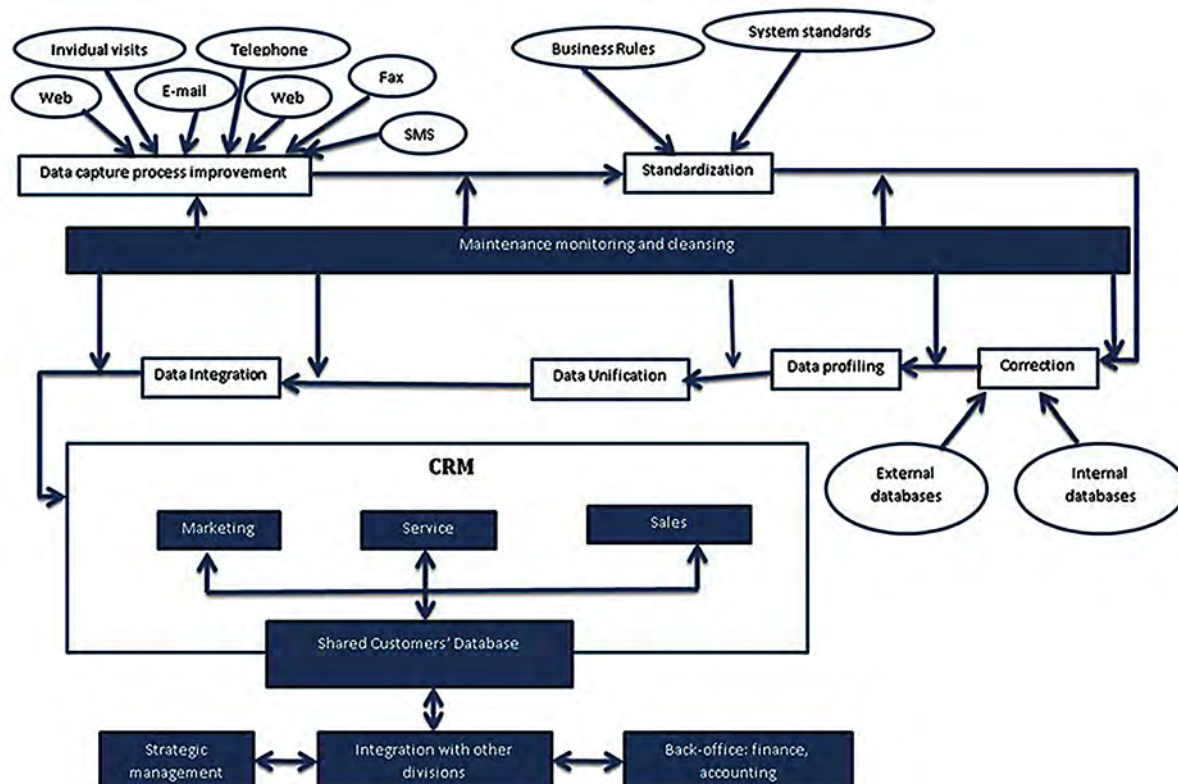
Figure 2 shows the proposed data quality framework that can be used to enhance the quality of data for e-CRM systems. First, the system has mechanisms in place to eliminate user errors. In alleviating user errors, the proposed framework has the capacity to flag suspicious or questionable records by comparing the various data values. Customers may at times participate in submitting wrong and misleading data either knowingly or otherwise. To flag such data as suspicious, the proposed framework incorporates data mining tools that can compare and analyze data values with the aim of detecting abnormal data values. It is difficult to recover data integrity once it is compromised. As such, the data quality framework incorporates systematic safeguards that counteract and prevent dirty data from entering e-CRM systems.

The second step in the data quality framework is the data capture process improvement. An

organization can capture its customers' data at many points of contact such as SMS, E-mail and telephone conversation. It is important for the organization in question to ensure that the process of capturing its customers' information is efficient and error free. This can be achieved by identifying and profiling various channels that customers use to interact with the organization (Chen & Ling, 2008). Each channel is then treated with a customized solution that enhances data quality in the channel. Data capture process improvement should also entail standardization of data at each channel to ensure consistency and eliminate data duplication (Batenburg & Versendaal, 2007).

The third step in the data quality framework is data standardization. Standardization involves creating structure in unstructured or semi structured data to ensure that the collected customer data conforms to business rules and system standards (Chen & Lin, 2009). The fourth phase

Figure 2. Data quality framework



in the framework is data correction. The data correction phase entails correcting customer data to ensure that it conforms to its definition, value domain and range. It is also in this phase that the missing value algorithm is defined (Powell & Childerhouse, 2010).

The fifth phase in the data quality framework is data profiling. An organization should know the source and various attributes associated with its customer data. This will enable the organization in question to select the most effective cleansing algorithms and functions that can enhance the quality of the data. The sixth phase in the data quality framework is data unification (Nguyen & Mutum, 2012). In this phase, the new customer data is matched with the existing data in order to establish a unique identity across an organization. The process entails identifying similar data within and across data sources and appending new data or providing missing data to enhance the quality and value of the data (Nguyen & Mutum, 2012).

The seventh phase in the data quality framework is data integration. The main challenge in the implementation of e-CRM systems is that wide range of separate CRM applications and transactional systems that hinder an integrated view of customer data. As established by Fjermestad and Romano, there is an inadequate integration of operational e-CRM systems (Fjermestad & Romano, 2006). As such, system managers have to cope continually with a vast number of e-CRM applications. It complicates the attainment of high sales management because consumer attendants are unable to find the required information about customers. The study also demonstrated that there is an inadequate integration of various data sources, thus making each product to have its unique customer database (Fjermestad & Romano, 2006).

In most organizations, customer data are usually stored in marketing databases. It is however common to find customers interacting with various departments in an organization such as technical support and shipping departments on different occasions (Chen & Ling, 2008). The resulting

interaction may therefore end up in different databases across the organization. It is thus vital to avail all the pieces of customer data to the CRM system by integrating the various databases in the organization. This will ensure a total view of customer behavior in the organization. This is because integrating systems within an organization ensures that updates in one system automatically update other related systems.

After intergration, the customer data can then be used in e-CRM systems. It is however important to design the systems in a manner that allows for data augmentation. This entails incorporating additional external or internal data that is not directly related to the customer data in question in order to gain insight. The customer data can be combined with data from third parties in order to enhance understand of customers' behaviors (Chua, 2011). This can enable an organization to easily segment its customers in order to meet the needs of each segment cost effectively.

The final step in the data quality framework is the maintenance monitoring and cleansing. Maintenance monitoring and cleansing is important in ensuring data integrity. This is because achieving data integrity is not a one-shot effort. There is need for an ongoing vigilance. It is therefore important for an organization to put in place policies, processes and tools that will enable it to monitor and solve data quality issues in its e-CRM system.

Data Cleansing Algorithm

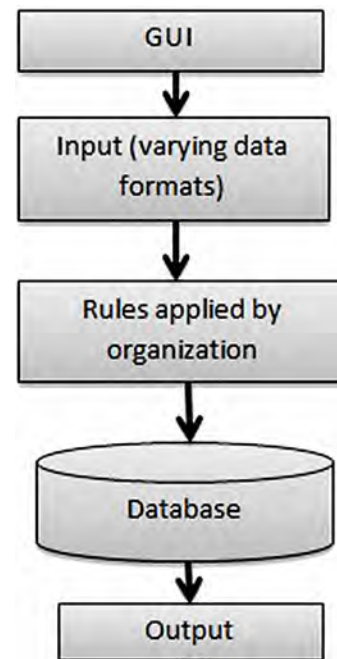
The organizational data on the website is usually not correct and outdated (Chen & Ling, 2008). This is usually due to duplicate record and misleading information. The front end of the system is designed using Java or any other web language, such as; PHP. The proposed backend system is designed using SQL for the database. The integration between the front end and the backend system completes the algorithm. Besides the storage, retrieval, and modification, data quality issues

also may arise in the implementation process. This is due to the lack of software technology, infrastructure, customer-centric strategy, and commitment from a corporation.

Generally, CRM software can enable business transactions, facilitate data mining and warehousing, create operational-centered databases, and provide marketing campaign supporting tools (Chen & Ling, 2008). However, different firms provide CRM software with different package specifications. Issues regarding the quality of data arise when a corporate, say an e-commerce site selects CRM software that do not meet their business specifications. Qualities issues also are common in the case of a defective infrastructure. The seamless E-CRM performance requires a well configured corporate infrastructure. Poor quality of data is prevalent in cases of meager network setup, storage capacity, data backup, and backlogged web servers.

However, with this proposed approach, the varying date formats in the fields of type- date are exploited. The date is used as an illustrative example, but this could be applied to all data input types with predefined predicates. Simply, this proposed approach is not limited to one single data type but the rest too even the higher in the precedence order of SQL server such as XML, datetime2, or date time offset (Vitt, n.d). Such is true as long as the algorithm for the define rules using Java programming model as graphical user interface (GUI) and SQL at the backend. If a user or operator merges two expressions having unique data types rules defining data type precedence specifies that lower precedence data type to be converted to higher precedence data type (Vitt, n.d). In regard to the illustrative example, problems faced include the realization of a normalized standard date format, disguised date values, varying time zones, differing abbreviations and different forms of separators between the date parameters. The flowchart in Figure 3 gives the flow of the suggested data cleansing algorithm.

Figure 3. The proposed model



Explanation of the Proposed Algorithm

In the proposed algorithm the user or the customer will be asked to fill in the details regarding the product or service quality. Moreover, the GUI shall be designed in such a way that it provides the clientele's personal information, such as the purchase history, the preferences and the comments of the customer on the product or services as this informatics are essential for CRM improvisation. The data inputted must not have a limitation of data formats, i.e., the proposed system must allow all data formats to be evaluated and used accordingly. The rules must be applied by an organization on the dataset to extract the useful information. These rules may include application data mining tools as well, such as: R-Programming, Weka and Rapid Miner. Additionally, these rules set by an organization will define that what sort information is needed by an organization to improve its processes. However, this is the most crucial part

of the system where rules are being made and set by the organization itself. The set rules will improve the data quality, if the right information is extracted from the data then and only then the quality of data can become better.

The extracted data is then stored in a database where series of functions can be applied on it inclusive of the sorting of data, data mining and other sophisticated tools to get the desired results. The applied tools and techniques will provide information on the basis of majority hits, i.e. it will provide with what majority feels about the product or the services. Hence, the proposed model can easily help in improving the data quality which could further improve the business process and decisive skills of an organization. The algorithm proposed by Wing Ning Li, Johnson Zhang and Roopa Bheemavaram sorts the data to make it more efficient. The sorting algorithm determines if the two inputted records are related in any way or not with the any other records and thus partitions the records based on its relation (Li, Zhang,

& Bheemavaram, 2006). As the data depicted in Figure 4 is sorted out using the algorithm and the result is illustrated in Figure 5.

Similar, Algorithms have been introduced to improve the data quality and can be efficiently used by organizations to increase viability of their ECRM processes.

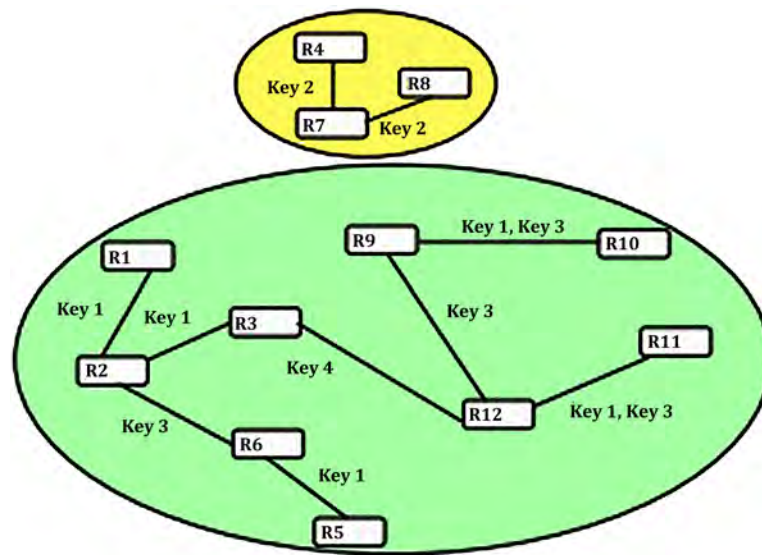
RESULT

The proposed model will enhance the data quality as the accurate information is extracted from the database and is analyzed to enhance the customer relationship with the organization. The tagged data will be accessible to the organization for decision making and improvise their product or service quality. Moreover, the E-CRM when integrated with business intelligence also provides additional features that can be benefited from. These features of the proposed system will include following benefits;

Figure 4. Different fields of data indicated by record no. and set no.

RecordNo.	Key 1	Key 2	Key 3	Key 4	Set Number
R1	C1K1V4	C1K2V10	C1K3U30	C1K4U31	1
R2	C1K1V4	C1K2U2	C1K3U3	C1K4V2	2
R3	C1K1V4	C1K2V5	C1K3U8	C1K4V5	3
R4	C2K1U18	C2K2V3	C2K3V3	C2K4V3	4
R5	C1K1V5	C1K2V7	C1K3U21	C1K4V7	5
R6	C1K1V5	C1K2V4	C1K3U3	C1K4V4	6
R7	C2K1V9	C2K2V3	C2K3U5	C2K4V4	7
R8	C2K1V16	C2K2V3	C2K3U6	C2K4V5	8
R9	C1K1V7	C1K2U13	C1K3U14	C1K4U15	9
R10	C1K1V7	C1K2V42	C1K3U14	C1K4V34	10
R11	C1K1V23	C1K2V6	C1K3U14	C1K4V6	11
R12	C1K1V23	C1K2V1	C1K3U14	C1K4V5	12

Figure 5. The data records sorted out in a database



Automation in Marketing Processes

The integration of business intelligence and E-CRM systems implement automated capabilities to easily manage the order of repetitive tasks and enhance the marketing exertions accessible to the customers from the different stages of the production lifecycle. For instance, if a sales prospect is considered, the system will inevitably direct the marketing content via mail or social media in case to spin the sales, building a mature customer for the organization (Cunningham, 2002).

Automation in Contact Services Operations

This tool in the integrated system of business intelligence and E-CRM is designed to lessen the monotonous operations of a contact center agent's job. The computerized solution will include a recorded audio that helps in customer problem solving queries and distribution of information. Numerous tools are assimilated with the agent's desktop tools in order to manage the customer requests and bring down the call intervals, while simplifying the whole service process (Brink, 2009).

FUTURE RESEARCH DIRECTION

There already exists a couple of e-CRM technologies aimed at improving service delivery and interaction experiences between the companies and customers. For these to be achieved, both the companies and the customer personnel must have a fair interaction with the e-CRM system. The main focus of this paper was to realize a better interaction experience between the companies and the e-CRM system. The proposed framework and algorithm are aimed at enhancing data quality for e-CRM systems.

Another area that this paper adds knowledge to is on data integration and duplication. These two areas are new and continue to evolve, and hence less regard is given towards their implication. Only a few studies on the issues are given address perhaps because of the faster rate upon which the integration is invoked. This is because, the more data is brought together; any e-CRM is likely to face a range of data issues regarding quality that require attention.

CONCLUSION

For businesses that have adopted e-Business, it is a vital business function to enhance the quality of their services. This can be achieved by using high quality data in their e-CRM systems. Data quality improvement allows for consistent data types with the help of an implementable framework and algorithm that can go a long way in ensuring data quality in e-CRM systems.

REFERENCES

- Ahmed, B., Ben Maâti, M. L., & Al Mohajir, B. (2014). The Intelligence of E-CRM Applications and Approaches on Online Shopping Industry. *International Journal of Innovation and Scientific Research*, 12(1), 213–216.
- Ahmed, B., Maati, M. B., & Al Mohajir, B. (2015a). The Effectiveness Of The Intelligence E-CRM Application In Enhancing Positive Customer Relationships. *Journal International Journal of Business Intelligence Research*, 6(1), 1–10. doi:10.4018/IJBIR.2015010101
- Ahmed, B. S., Maati, M. L. B., & Al Mohajir, B. (2015b). Improve Intelligence of E-CRM Applications and Customer Behavior in Online Shopping. *International Journal of Business Intelligence Research*, 6(1), 1–10. doi:10.4018/IJBIR.2015010101
- Anderson, K., & Kerr, C. (2002). *Customer relationship management*. New York: McGraw-Hill.
- Bashar, , & BenMati,, & Al Mohajir. (2014). The Intelligence of E-CRM Applications and Approaches on Online Shopping Industry. *International Journal of Innovation and Scientific Research*, 12(1), 213–216.
- Batenburg, R., & Versendaal, J. (2007). Business/IT-alignment for customer relationship management: Framework and case studies. *IJECRM*, 1(3), 258. doi:10.1504/IJECRM.2007.017795
- Botha, J., Bothma, C. H., & Geldenhuys, P. (2008). *Managing e-commerce in business*. Juta and Company Ltd.
- Busaidi, K. (2013). Aligning customer knowledge management tools with business strategy. *IJECRM*, 7(2), 117. doi:10.1504/IJECRM.2013.056510
- Chen, T., & Lin, Y. (2009). Enhancing the relationship benefit to develop relationship value. *IJECRM*, 3(2), 167. doi:10.1504/IJECRM.2009.025286
- Chen, T., & Ling, Y. (2008). Customer relationship management resources and practices and customer voluntary behaviour: The evidence of passengers at Taiwans international airport. *IJECRM*, 2(3), 276. doi:10.1504/IJECRM.2008.020412
- Chua, A. (2011). How Web 2.0 supports customer relationship management in Amazon. *IJECRM*, 5(3/4), 288. doi:10.1504/IJECRM.2011.044693
- Cunningham, M. (2002). *Customer relationship management*. Oxford, UK: Capstone Pub.
- de Andrade, T. L., & de Souza, R. C. G. (2011). Optimization of algorithm to identification of duplicate tuples through phonetic based on multithreading. *12th International Conference on Parallel and Distributed Computing, Applications and Technologies*. doi:10.1109/PDCAT.2011.58
- Fan, W., Geerts, F., Li, J., & Xiong, M. (2011). Discovering Conditional Functional Dependencies. *IEEE Transactions on Knowledge and Data Engineering*, 23(5), 683–698. doi:10.1109/TKDE.2010.154
- Fjermestad, J., & Romano, N. (2006). *Electronic customer relationship management*. Armonk, NY: M.E. Sharpe.

Gosney, J., & Boehm, T. (2000). *Customer relationship management essentials*. Roseville, CA: Prima Tech.

Horwitz, L. (2014). *Study says CRM boosts CEM, but data challenges still prevail*. *SearchCRM*. Retrieved 28 October 2015, from <http://searchcrm.techtarget.com/news/2240232157/Recent-study-says-CRM-boosts-CEM-but-data-challenges-still-prevail>

Jones, S., & Ranchhod, A. (2007). Marketing strategies through customer attention: Beyond technology-enabled Customer Relationship Management. *IJECRM*, 1(3), 279. doi:10.1504/IJECRM.2007.017797

Muther, A. (2002). *Customer relationship management*. Berlin: Springer. doi:10.1007/978-3-642-56222-8

Nguyen, B., & Mutum, D. (2012). Customer relationship management: Advances, dark sides, exploitation and unfairness. *IJECRM*, 6(1), 1. doi:10.1504/IJECRM.2012.046467

Powell, M., & Childerhouse, P. (2010). Retail value stream management. *IJECRM*, 4(3), 209. doi:10.1504/IJECRM.2010.035963

Raab, G. (2008). *Customer relationship management*. Aldershot, UK: Gower.

Shah, M. (2008). Fuzzy logic: A realistic tool for management of customer relation. *IJECRM*, 2(2), 158. doi:10.1504/IJECRM.2008.019930

Vitt, E. (2014). *Data Quality Solutions*. *Technet.microsoft.com*. Retrieved 9 May 2016, from <https://technet.microsoft.com/en-us/library/aa964137%28v=sql.90%29.aspx>

Yuhang, Z. (2010). Research on Data Cleaning in Text Clustering. *IEEE Conference at International Forum on Information Technology and Applications*. doi:10.1109/IFITA.2010.73

KEY TERMS AND DEFINITIONS

Business: Business environment which operates in Intelligence of E-CRM.

Data Quality: The level of quality of data from E-CRM application (this is our subject).

E-CRM Systems: Electronic Customer Relationship Management system encompasses all the CRM functions with the use of the net and organizations environment.

Intelligence: One of the features that have E-CRM systems and their process.

Organization: Place where operator Intelligence of E-CRM systems such as institutions and companies etc..

Process: All about related of business generated from Intelligence of E-CRM application.

Relationship Management: Strategy in which a continuous level of engagement is maintained between an organization and Intelligence of E-CRM application.

Optimizing Cloud Computing Costs of Services for Consumers

C

Eli Weintraub*Afeka Tel Aviv College of Engineering, Israel***Yuval Cohen***Afeka Tel Aviv College of Engineering, Israel*

INTRODUCTION

Cloud Computing (CC) typically deals with organizations using computing services, communication and web applications. The National Institute of Standards and Technology (NIST) defines CC as a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (for example networks, servers, storage, applications and services) that can be rapidly provisioned and released, with minimal management effort or service-provider interaction. Cloud computing targets four main groups of organizational customers: private, public, community and hybrid. Customers choose their CC provider which gives them maximum value in minimal costs. This paper examines users' motivations in choosing their pricing model; certain customers are looking for best value, others looking for least cost while others look for a combination of both reasons.

This research reviews the main motivations and obstacles to adopting the cloud technology by companies, and develops a cost model for optimizing the consumer costs. Providers currently offer software services as bundles consisting of services which include the software, platform and infrastructure services. Providers also offer platform services bundled with infrastructure services. This bundling policy is likely to change in the long run since it contradicts economic free market rules and competition conditions, causing an unfair pricing model and locking-in consumers

to specific service providers. A famous example is the Israel telecom revolution where the introduction of competition in 1995 led to an extremely high subscriber growth rate, one of the highest in the world. As of 2014, penetration stands at approximately 125% (The Israel Ministry of Communication, 2016). Vendor lock-in is a major barrier to the adoption of cloud computing, due to the lack of standardization (Opara-Martins, Sahandi, & Tian, 2016). This research assumes that in the future market forces will push providers to act in a free competitive market, in which consumers are free to switch their services among providers. The proposed model is aimed at the potential customer who wishes to find the optimal combination of service providers which minimizes his costs. The objective of this paper is proposing possible strategies for implementation of the model in organizations, optimizing consumers' costs.

BACKGROUND

Comparing CC pricing models is a complicated task due to variance among providers' services and structure of tariff tables. Researchers found that cost saving is the strongest incentive for organizations considering CC adoption (Yung-Ming & Chia-Ling, 2012). CC services are usually sorted to three groups: SaaS (Software as a service), PaaS (Platform as a Service) and IaaS (Infrastructure as a Service), each service belongs to a specified group, and is offered for specific prices.

DOI: 10.4018/978-1-5225-2255-3.ch141

There exist two main pricing models. Pay-per-use is the most used model, in which the consumer is charged a fee for a used unit in a specified duration. The unit used may be a certain computing unit of hardware, software or application. Fixed-price model, in which the user is charged for using a service unit for a fixed price, usually in periods of month or year. In the fix-price model consumers may consume an unlimited amount of unit resources, although in some contracts consumption is limited to a maximal amount which consumers do not intend to reach. In the fixed-price model consumers might be charged for resources they have not actually consumed. Al-Roomi, Al-Ebrahim, Buqrais, & Ahmad, (2013) surveyed pricing models, and classified them to three groups: fixed – in which the customer is charged the same amount all the time, dynamic – in which prices changes dynamically according to purchased volumes and market-dependent in which prices change according to market conditions. Lai (2005) claims that market competition powers using pay per use pricing model could bring efficient allocations of computing facilities. Weinhardt et al., (2009) illustrate that current trends in CC show an ambition to base pricing models on dynamic pay-per-use pricing models. In certain cases consumers prefer to pay a fixed price, ignoring pay-per-use model advantages which fit their exact consumption and might minimize their costs (Anandasivam & Premm, 2009; Pueschel, Anandasivam, Buschek, & Neumann, 2009). Wu & Banker, (2010) found that some providers offer pay-per-use pricing and leave some consumer surplus to the customers in order to be more attractive. Researchers explored cloud provider pricing models using cluster analysis and found common business models; first cluster includes niche providers who use fixed pricing, second cluster includes mass players using pay-per-use pricing models (Labes, Erek, & Zarnekow, 2013). A possible explanation of using fixed prices is lock-in situations prevalent among niche players' products. Lilienthal (2013) who compared costing

schemes offers a decision model which calculates financial trade-off with respect to the workloads.

Several researchers studied anomalies in consumer decisions. Lambrecht & Skiera (2006) identified fixed-prices biases in which consumers prefer a fixed price model although they would pay less on a pay-per-use tariff for reasons of budget confidence, and cases of consumers prefer a pay-per-use model paying more for operational flexibility. Koehler, Anandasivam, Dan, & Weinhardt (2010) also found that the insurance effect has significant influence on the fixed-price bias while the pay-per-use bias is influenced by the flexibility effects.

Several researchers state that providers use to offer free of charge services using lock-in strategies (Koehler et al., 2010). Researchers found differences between private and organizational consumers (Weinhardt et al., 2009). Most cloud services which are focused on private consumers are free of charge while organizational consumers are usually charged, and only some add-on services are free of charge. PaaS providers often offer their development tools for free. Walterbusch, Martens, & Teuteberg (2013) raise the awareness of indirect and hidden costs in cloud computing pricing models. They found that some providers try to attract customers by a low price per storage while charging hidden costs for data transfer. Chen, Han, Cao, Jiang, & Chen (2013) state that customers face difficulties in evaluating prices of cloud services, those difficulties are one of the main reasons preventing customers from adopting cloud services.

This paper proposes a pricing model which optimize customers' costs in a future cloud computing free-competition market, thus eliminating the above discussed market in-efficiencies.

MAIN FOCUS OF THE ARTICLE

Research literature in the CC pricing models domain names three competition barriers' features:

bundling of services, lack of transparency, and varying tariff structures.

Bundling of Services

Examining providers' pricing models as published in their websites (Amazon, 2016; Google, 2016; Microsoft, 2016) illustrates the phenomena of bundling services, where each provider offers only a couple of services while other providers offer their preferred services. Weintraub & Cohen (2015) introduce new definitions of two kinds of bundling: first is horizontal bundling, second is vertical bundling. In horizontal bundling a provider offers several services, all belong to one layer. For example Amazon EC2 offers several bundles each one is composed of the following components: CPU, ECU, memory, instance storage, and operating system. In such bundling situations consumers may not use their own operating system. In vertical bundling a provider offers services which belong to lower layers, in addition to the main needed service. For example Amazon offers SaaS services, in which the consumer is asked to choose the configuration of infrastructure he wants the software application to run. A consumer looking for SaaS services may not use a PaaS service such as his own operating system or an operating system he bought from another cheaper service provider.

According to Weintraub & Cohen (2015) consumers should be able to choose another IaaS provider instead of being forced to use the infrastructure services of the SaaS or PaaS main provider. Providers use to bundle services in ways that customers are unable to know the real prices of each service. Such a situation contradicts economic competition principles, causing an unfair pricing model when examining customers' optimal alternatives.

Lack of Transparency

Bundling masks the prices of services, in both situations: vertical and horizontal bundling. The

provider offers a tariff for the whole bundle without breaking it to its components' services, in a package deal. In such situations customers do not know the price of a specific service which is part of the bundle. Researchers found that public cloud customers receive no insight into the underlying IT infrastructure and have restrictive administrative rights (Walterbusch, Martens, & Teuteberg, 2013). El Kihal, Schlereth, & Skiera (2012) state that transparency of service costs in cloud computing is a key factor to popular wide usage by organizations. Blau, Neumann, Weinhardt, & Michalk (2008) propose a pricing model which includes incentives to providers who are willing to present the pricing components of their services and also the configuration of the technological implementation such as the assets consumed for each service. Improving transparency will be feasible by breaking bundles to component services so that a consumer may choose each service by comparison to competing providers' prices.

Varying Tariff Structures

A consumer wishing to compare a service offered by several providers may find it difficult to perform, sometimes impossible at all. The large number of cloud providers' services based on varying pricing schemes has led to complexities in cloud service selection (Rehman, Hussain, & Hussain, 2011). According to El Kihal, Schlereth & Skiera (2012) pricing models providers use different tariff structures; some providers such as Google charge separately for each service, and providers like Amazon and Microsoft offer predefined bundles of services. This situation is due to reasons such as services having different functionalities, computing resources having different technological characteristics, differences in service levels, differences in contract duration and differences in discounts offered for certain volumes. The variability of charges between current SPs does not give sufficient common ground for a simple comparison. The issue of tariff structure variability is intensified due to the usage of two

different pricing models: fix-rate pricing and the pay per use pricing which enable customers incorporate risk analysis and cash-flow considerations.

This paper proposes a pricing model in which the customer is free to choose service providers according to his own pricing preferences, composing the bundled services by his own. In the proposed model tariff tables are normalized to hourly time units. The model simulates situations in which consumers may install software or hardware components on multi providers' sites, optimizing consumers' total CC expenses.

SOLUTIONS AND RECOMMENDATIONS

Solutions

Cloud computing architecture is described in literature as consisting of three layers: IaaS, PaaS and SaaS, each layer performs certain functions and also supporting functions requested by upper layers. Weinhardt et al. (2009) introduced a framework of CC architecture, composed of three layers of functions; The *Infrastructure layer* – focuses on providing technologies as basic hardware components for software services. *Platform layer* - includes services which are using cloud infrastructures needed for their functioning. *Application layer* - consists of the human interfaces used by the organizations' end-users. Applications are running on cloud assets, making use of platform and infrastructure layers, managing underlying infrastructures for the offered service. For example a SP suggesting a SaaS product is also bundling into the product the PaaS and IaaS layers.

Using current cloud computing technologies, a certain provider may run an application using another provider's infrastructure, but in practice both providers are parts of the same organization (Zhang, Cheng, & Bautaba, 2010). Bundling practices used by service providers force customers pay for components they may buy cheaper from other providers. Paraiso, Haderer, Merle, Rouvoy,

& Seinturier (2012) state that in the future developers will plan their cloud applications enabling migration of services among multiple clouds. Velte, Elsenpeter, & Velte (2009) state that applications belonging to different layers will be run on separate geographical locations.

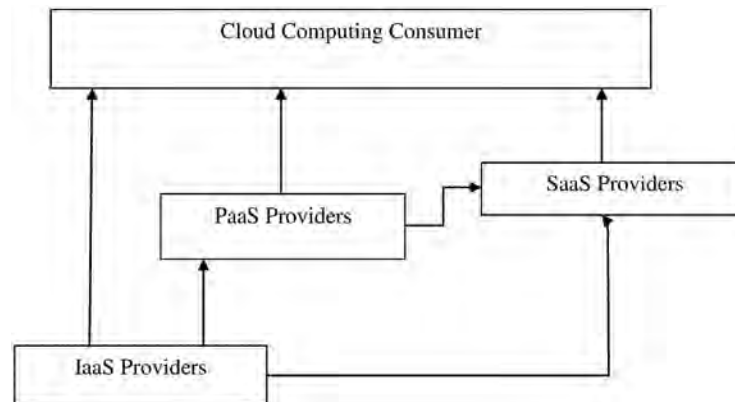
The business model proposed in this article enables implementing functionalities of a service provider interfacing the underlying layers by other service providers. This functionality puts two requirements on cloud architecture. First, the architecture should use open standards which enables interfacing between many components of many providers. Second, the architectures' building blocks should be loosely coupled. Implementation of those two functionalities enables connectivity among vertical and horizontal services, thus eliminating the bundling phenomena. Figure 1 describes the new suggested Cloud Computing Competition-enabled Architecture. Arrows describe services supplied by underlying layers. Rectangles describe computing services. A consumer looks for services of each one of three levels. Each provider may buy certain services from other providers. For example a consumer looks for SaaS services, he finds the cheapest SaaS provider, who also buys certain PaaS and IaaS services from other providers.

The inability to compare SP offerings is a competitive market failure which hinders not only SP competitiveness, but also the decisions of the customers. In an efficient market, customers should be able to easily compare all their options, and choose the best. Vendors will have to adapt their pricing models to standard schemes in order to improve their competitiveness. In the following, the proposed model suggests two pricing models in which service providers are competing for offering the best service to customers' requirements.

The Case Study Example

The simplest basis for comparing pricing schemes is using a ratio of Cents/Hour of usage. A more advanced pricing basis may include an addition

Figure 1. Cloud computing competition-enabled architecture



to the usage payments, periodical fixed payments for various services. To enable simple comparison of SP's Each SP base its offer on its in-house offering with complimentary services bought from its business partners. Also, when a SP buys another underlying service from another SP, s/he must pay not only for the service, but also for the administrative work and the interfaces between suppliers involved in the purchase.

This section describes a theoretical example of three SPs and their tariffs, and a consumer who chooses certain SP's based on his foreseen consumption. Following, an introduction of three pricing models which will make use of the data described in this section. Table 1 summarizes the customer estimated monthly requirements of hypothetic cloud services. All requirements are either in hours, volume or other specified unit.

The customer contacts three candidate SPs to get a price quote. The published tariffs of these SPs appear in Table 2. Each SP provides all the services using his/her tariff.

So, based on current practices bundled services per SPs would give:

- SP1 price per monthly usage: \$ 4038.40
- SP2 price per monthly usage: \$ 4220.80
- SP3 price per monthly usage: \$ 4964.00

Therefore, SP1 is the least cost provider (\$4038.40). However, the suggested model enables

Table 1. Customer estimated monthly requirements

Service Name	Requirement in Units	Units of Service
SaaS		
Data transfer	4,800	1 TB
Email services	4,800	1000 messages
Cloud search	1,200	1 search
Documents Mgt.	400	1 hour
ERP	4,800	1 hour
PaaS		
Operating system	1,200	1 hour
Memory	2,000	1 hour, 1 GB
Instance storage	2,000	1 hour, 1 GB
Developer support	480	1 hour
IaaS		
Relational Database services	1760	1 hour
Storage standard vol.	480	1 TB, 1 hour
Backup (GB)	2,000	1 GB, 1 hour

the consumer to set an efficient market price that further minimizes his/her expenses. This could take several forms depending on the main supplier of choice:

Hierarchical Pricing Model

Since fitting SaaS services to the customer is more sensitive to customer requirements this

Table 2. Tariffs of three service providers (Cents/Hour); example prices, assuming basic usage

Service Name	SP1 Tariff	Sp2 Tariff	SP3 Tariff
SaaS			
Data transfer	2	4	2
Email services	4	6	2
Cloud search	6	8	6
Documents Mgt.	4	2	2
ERP	100	94	120
PaaS			
Operating system	26	36	14
Memory	40	50	70
Instance storage	24	20	30
Developer support	6	8	4
IaaS			
Relational Database services	18	12	38
Storage standard vol.	34	30	50
Backup	40	56	40

model assumes that each SP maximizes its SaaS capabilities and look for purchasing the best combination of platform and infrastructure services that best complements its own offerings in these levels. Since SPs seek simple management of sub-contracted services, only one SP could be chosen for complementing the platform or the infrastructure level. The Platform SPs can also purchase infrastructure services. Also, when a SP buys another underlying service he must pay two fees: a fixed monthly sum of money for initiating, controlling and maintaining the connection to the other SP, and a fee for each executed transaction.

While SaaS is the highest level in the hierarchy, the computations start from the lowest level (IaaS) and progress through PaaS to the decision taken by the SP based on their SaaS and possibly sub-contracted PaaS and/or IaaS. In this example the IaaS total monthly costs (in \$) per SP are as calculated in Table 3, Table 4 and Table 5.

Table 3. IaaS monthly prices per SP for the example

Service	SP1 \$/Month	SP2 \$/Month	SP3 \$/Month
IaaS			
Relational DB services	158.4	105.6	334.4
Storage standard vol.	81.6	72	120
Backup	400	560	400
Total	640	737.6	854.4

Table 4. PaaS total costs (in \$) per SP for the example

Services	SP1 \$/Month	SP2 \$/Month	SP3 \$/Month
PaaS			
Operating system	156	216	84
Memory	400	500	700
Instance storage	240	200	300
Developer support	14.4	19.2	9.6
Total	810.4	935.2	1093.6

Table 5. SaaS monthly price per SP for the example

Service	SP1 \$/Month	SP2 \$/Month	SP3 \$/Month
SaaS			
Data transfer	48	96	48
Email service	96	144	48
Cloud search	36	48	36
Document Mgt.	8	4	4
ERP	2400	2256	2880
Total	2588	2548	3016

Thus SP1 is the IaaS provider of choice for the requirements of this customer.

Again SP1 is also the PaaS provider of choice for the requirements of this customer.

Finally, the SaaS total costs (in \$) per SP per month are as follows:

Here SP2 is the SaaS provider of choice for this customer.

The customer in this model would choose at each level the provider of choice for the requirements. Ignoring the fixed monthly sum of money for initiating, controlling and maintaining the connection between SP2 and SP1, SP2 is chosen for SaaS (\$2548, Table 5) and SP1 for PaaS (\$810.4, Table 4) and IaaS (\$640, Table 3).

The total cost per month would be the sum of minimum: $640+810.40+2548 = \$ 3998.60$

Assuming there is a fix monthly fee of \$60 for initiating, controlling and maintaining the connection between two different SPs in each two consecutive levels. The monthly fee of 60\$ is chosen as an example only, for model illustration, without limiting the generality of the model since comparing investment alternatives usually involve fix and variable costs, two components which the model includes. An additional \$ 60 per month would be charged for the connection SP2(SaaS) - SP1(PaaS), and no charge for the same SP1 between PaaS and IaaS. In that case, the total cost would be: $\$ 3998.60+60 = \$ 4058.40$.

Let F be the fixed monthly fee for initiating, controlling and maintaining the connection between SP2 and SP1.

If $F < 2588-2548$ the above policy would remain optimal with monthly cost of $F+3998.60$.

If $F > 2588-2548$ than Choosing SP1 to supply the three levels of service would yield:

$$640+810.4+2588=\$4038.40$$

Of course supplier selection decisions require sensitivity analysis (finding the impact of small changes in requirements) but the example here is just for illustrating the required computations.

The Simple Pricing Model

This model relaxes the assumptions about hierarchy and the need for simple management and control over sub-contracted services. Thus, in this model each SP offers the bundle of services that is composed of the minimal tariffs. The cost of sub-

contracting management and control is assumed to be a fixed sum per service per period. Thus, each SP supplies his/her own services if their tariffs are minimal. Otherwise they sub-contract other suppliers (as shown in Table 6). For example SP1 offers the services included in SaaS PaaS and IaaS (Table 1 column SP1 tariff).

The minimal tariff of each item yields total monthly price for the customer's requirements of: \$ 3663.2. In addition, the customer must contact the other two SPs to establish the purchases and track the transactions. Assuming a monthly cost per SP per service of \$ 60.00 for the administrative work of ordering, tracking and payment management yields:

Main SP1: $3663.2+8*60 = 4143.20$

Main SP2: $3663.2+7*60 = 4083.20$

Main SP3: $3663.2+5*60 = 3963.20$

Thus, Main SP3 is chosen with monthly expenses of: 3963.20, compared with the minimal cost SP, this is annual savings of \$ 902.40).

RECOMMENDATIONS

Three preconditions are required for effective comparison of pricing models. Service providers have to offer standard functionalities of their products, standardize tariff table structures and show all cost components thus having transparent tariffs. Providers should build their services according to open standards thus enabling connectivity among different services offered by suppliers.

Consumers should enhance the usage of current technologies which offers several applications in which they can compare products prices, for example ShopSavvy and ShopAdviser. Such applications will enable CC services comparisons easily. Consumer organizations should switch their CC technological architecture to use the new offered pricing models; starting with the hierarchical model and when possible moving

Table 6. Minimal cost supplier for each service

Service Name	SaaS Minimal Tariff	SP1 Sub Contract (Bald)	SP2 Sub Contract (Bald)	SP3 Sub Contract (Bald)
SaaS				
Data transfer	2	SP1	SP3	SP3
Email services	2	SP3	SP3	SP3
Cloud search	6	SP1	SP3	SP3
Documents Mgt.	2	SP3	SP2	SP3
ERP	94	SP2	SP2	SP2
Service Name	PaaS Min. Tariff	SP1 Sub Contract	SP2 Sub Contract	SP3 Sub Contract
PaaS				
Operating system	14	SP3	SP3	SP3
Memory (GB)	40	SP1	SP1	SP1
Instance storage GB	20	SP2	SP2	SP2
Developer support	4	SP3	SP3	SP3
Service Name	IaaS Min Tariff	SP1 Sub Contract	SP2 Sub Contract	SP3 Sub Contract
IaaS				
Relational Database services	12	SP2	SP2	SP2
Storage TB standard volume.	30	SP2	SP2	SP2
Backup (GB)	40	SP1	SP1	SP3

gradually to the second simple pricing model thus achieving optimal costs.

Governments and regulatory agencies should build technological infrastructures and economical ecosystems, to enhance completion among providers and forbid providers using monopoly and lock-in strategies.

FUTURE RESEARCH DIRECTIONS

Further research is needed for incorporating additional consumers' considerations such as preference of a fix-price model on a pay-per-use model due to risk aversion, and providers' impacts on consumers' biases in buying decisions. Further research is needed for better understanding the providers' considerations in the cloud computing market.

CONCLUSION

This paper proposes two cost minimization models for cloud computing consumers, (while keeping the published tariffs). Choosing one of the models depends on its similarity to the actual needs and conditions in the CC market. The first model is hierarchical; one supplier is chosen for each of the three layers (SaaS, PaaS, IaaS). The hierarchical model is easy to implement and also reduces customers' costs compared to the current situation. This model produces only a limited amount of services unbundling and only limited transparency of prices. The second model called the simple pricing model, enables high transparency and unbundling of services and further cost reduction. Implementing this model is more difficult since the control of various providers and services is more complicated. The paper also gives recommendations to all parties involved in the CC market: private and organizational consumers, service providers and government agencies.

REFERENCES

- Al-Roomi, M., Al-Ebrahim, S., Buqrais, S., & Ahmad, I. (2013). Cloud Computing Pricing Models: A Survey. *International Journal of Grid and Distributed Computing*, 6(5), 93–106. doi:10.14257/ijgdc.2013.6.5.09
- Amazon Web Services. (2016). *How AWS Pricing Works*. Retrieved June 10, 2016 from https://d36cz9buwru1tt.cloudfront.net/AWS_Pricing_Overview.pdf
- Anandasivam, A., & Premm, M. (2009). Bid Price Control and Dynamic Pricing in Clouds. *ECIS 2009 Proceeding*. Retrieved from <http://aisel.aisnet.org/ecis2009/284>
- Blau, B., Neumann, D., Weinhardt, C., & Michalk, W. (2008). Provisioning of service mashup topologies. *Proceedings of the 16th European conference on information systems*.
- Chen, Z., Han, F., Cao, J., Jiang, X., & Chen, S. (2013). Cloud Computing-Based Forensic Analysis for Collaborative Network Security Management System. *Tsinghua Science and Technology*, 18(1), 40–50. doi:10.1109/TST.2013.6449406
- El Kihal, S., Schlereth, C., & Skiera, B. (2012). *Price comparison for Infrastructure-as-a-Service*. In European Conference on Information Systems. Retrieved from <http://aisel.aisnet.org/ecis2012/161>
- Google Cloud Platform. (2016). *Product Pricing Details*. Retrieved June 10, 2016 from https://cloud.google.com/?gclid=Cj0KEQjwmPKeBRCj4bOro6nBitABEiQABa2FJJiC0E3d61rVTIdBElecEgzYWDTaf_EyAckJ_nOyKH4aAs-nH8P8HAQ6
- Koehler, P., Anandasivam, A., Dan, M., & Weinhardt, C. (2010). Customer heterogeneity and tariff biases in cloud computing. *Proceedings of the Thirty First International Conference on Information Systems*.
- Labes, S., Ereke, K., & Zarnekow, R. (2013). Common Patterns of Cloud Business Models. *Proceedings of the Nineteenth Americas Conference on Information Systems*.
- Lai, K. (2005). Markets are dead long live markets. *ACM Exchanges*, 5(4), 1–10. doi:10.1145/1120717.1120719
- Lambrecht, A., & Skiera, B. (2006). Paying Too Much and Being Happy about It: Existence, Causes and Consequences of Tariff-Choice Biases. *JMR, Journal of Marketing Research*, 43(2), 212–223. doi:10.1509/jmkr.43.2.212
- Lilienthal, M. (2013). A Decision Support Model for Cloud Bursting. *Business & Information Systems Engineering*, 5(2), 71–81.
- Microsoft Cloud Services. (2016). *Cloud Services Pricing*. Retrieved June 10 2016, from <https://azure.microsoft.com/en-us/pricing/details/cloud-services/>
- Opara-Martins, J., & Sahandi, R., & Tian, F. (2016). Critical analysis of vendor lock-in and its impact on cloud computing migration: a business perspective. *Journal of Cloud Computing: Advances Systems and Applications*, 5(4), 1–18.
- Paraiso, F., Haderer, N., Merle, P., Rouvoy, R., & Seinturier, L. (2012). A Federated Multi-Cloud PaaS Infrastructure. *IEEE Fifth International Conference on Cloud Computing*, 392–399.
- Rehman, U. Z., Hussain, F. K., & Hussain, O. K. (2011). Towards Multi-Criteria Cloud Service Selection. *Fifth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing*, 44–48. doi:10.1109/IMIS.2011.99
- The Israel Ministry of Communications. (2016). *Telecommunications in Israel 2015 – From Monopoly to competition*. Retrieved August 24, 2016, from <http://www.moc.gov.il>
- Velte, A., Elsenpeter, R., & Velte, T. J. (2009). *Cloud Computing: A practical approach*. Tata McGraw-Hill Education Pvt. Ltd.

- Walterbusch, M., Martens, B., & Teuteberg, F. (2013). Evaluating cloud computing services from a total cost of ownership perspective. *Management Research Review*, 36(6), 613–638. doi:10.1108/01409171311325769
- Weinhardt, C., Anandasivam, A., Blau, B., Borisso, N., Meinel, T., Michalk, W., & Stöber, J. (2009). Cloud Computing – A Classification, Business Models, and Research Directions. *Business & Information Systems Engineering*, 1(5), 391–399. doi:10.1007/s12599-009-0071-2
- Weintraub, E., & Cohen, Y. (2015). Cost Optimization of Cloud Computing Services in a Networked Environment. *International Journal of Advanced Computer Science and Applications*, 6(4), 148–157. doi:10.14569/IJACSA.2015.060420
- Wu, S., & Banker, R. (2010). Best Pricing Strategy for Information Services. *Journal of the Association for Information Systems*, 11(6), 339–366.
- Yung-Ming, L., & Chia-Ling, C. (2012). Analyzing the Pricing Models for Outsourcing Computing Services. Proceedings of PACIS 2012.
- Zhang, Q., Cheng, L., & Bautaba, R. (2010). Cloud computing: State-of-the-art and Research challenges. *Journal of Internet Services and Applications*, 1(1), 7–18. doi:10.1007/s13174-010-0007-6
- Ouedraogo, M., Mignon, S., Cholez, H., Furnell, S., & Dubois, E. (2015). *Security transparency: the next frontier for security research in the cloud*. Journal of Cloud Computing: Advances. *Systems and Applications*, 4, 12.
- Zhang, Q., Chen, H., Shen, Y., Ma, S., & Lu, H. (2016). Optimization of virtual resource management for cloud applications to cope with traffic burst. *Future Generation Computer Systems*, 58, 42–55. doi:10.1016/j.future.2015.12.011
- Zheng, S., Zhu, G., Zhang, J., & Feng, W. (2015). Towards an adaptive human-centric computing resource management framework based on resource prediction and multi-objective genetic algorithm. *Multimedia Tools and Applications*, 1–18.
- Zhu, Y., Hu, H., Ahn, G., & Yau, S. (2012). Efficient audit service outsourcing for data integrity in clouds. *Journal of Systems and Software*, 85(5), 1083–1095. doi:10.1016/j.jss.2011.12.024

KEY TERMS AND DEFINITIONS

Bundling: Product bundling is offering a package of several products for sale as one combined product. Bundling can be regarded as an unfair use of market power because it limits consumers' choices.

Cloud Computing: Cloud computing is the delivery of on-demand computing services over the internet on a pay-for-use basis, with minimal consumer setup efforts.

Competition: Companies compete each other over the same group of customers. Competition causes commercial firms to develop new services, giving consumers better services. The greater selection causes lower prices for the services, compared to a market with limited competition.

Infrastructure as a Service (IaaS): A cloud computing model in which the provider offers remote hardware and basic software services such as operating systems, database management systems and utilities.

ADDITIONAL READING

- Faniyi, F., & Bahsoon, R. (2015). A systematic review of service level management in the cloud. *ACM Computing Surveys*, 48(3), 1–27. doi:10.1145/2843890
- Hummaida, A. R., & Norman, W. (2016). Paton & Sakellariou, R. (2016). *Adaptation in cloud resource configuration: a survey*. Journal of Cloud Computing: Advances. *Systems and Applications*, 5, 7.

Platform as a Service (PaaS): A cloud computing model in which the provider offers software components such as compilers and debuggers which are used for application development.

Software as a Service (SaaS): A cloud computing model in which the provider offers software applications aimed for remote usage by

IT professionals and business users. SaaS applications includes applications such as CRM and ERP.

Transparency: Transparency of a cloud computing service is the ability of the system to provide visibility of the structure and behavior of a process. Transparency leads to consumers' trust in the supplied service.

C

Taxonomy for “Homo Consumens” in a 3.0 Era

Carlos Ballesteros

Universidad Pontificia Comillas, Spain

INTRODUCTION

Consumption is a basic human activity that implies a wide range of consequences: economic, psychological, environmental and social ones. Although it is true that consumption is one of the principal and most necessary engines for modern economies, there are important ecological and social costs in its practice. As human beings we are co-responsible for our consumption habits. We can no longer think that to consume is a free act that only implies ourselves and what can we afford. Through our lifestyles and consumption patterns we transform the world. Consumption can be seen in many different ways: from a liberal perspective nobody has the right to influence and be critical about your consumption but yourself: you have the freedom and the right to buy what you can/want/desire; from a reformist point of view, as a society, a consumer society is the best, the ideal society we can have, but it has some drawbacks that we can reform or correct; The sustainable side states that it is possible to move towards a cleaner (and a fairer) world with certain behaviours; Last but not least there is even a political viewpoint where consumption can be seen as a way to change the world. There could be also a radical position (also seen as destructive), or also an ascetic one, that pleads for a new relationship between the Planet and Humankind in a back-to-the-roots path.

It was Erich Fromm, in 1965, who was the first to call the human species *Homo consumens*, a term that evolved from *Homo sapiens* to denote how we devote all our efforts, beyond working from 9:00 to 5:00, to consume compulsively as a means to

sublimate our frustrations. In addition, Saad (2007) and Lipovetsky (2006) have used the term *Homo consumericus* to refer to the Darwinian principles of evolution and consumer behavior. Lipovetsky has noted that modern times have brought about the rise of a third type of *Homo consumericus*, who is unpredictable and insatiable.

Homo Consumens is the man whose main goal is not primarily to own things, but to consume more and more, and thus to compensate for his inner vacuity, passivity, loneliness and anxiety...Homo Consumens is under the illusion of happiness, while unconsciously suffers from his boredom and passivity. The more power you have over machines, the more powerless it becomes as a human being; the more you consume, the more you become a slave to the ever increasing needs. (Fromm, 1965, p. 214)

Since then, most of the literature on symbolic consumption and consumerism has recreated this concept to deepen, if not openly criticize, a society that alienates the person, relegating it to a mere facet, often irresponsible and compulsive, and unconsciously, a buyer of increasingly sophisticated products and services. For example, Robert Bocoock (1995) states that consumption is a social, psychological, and cultural process that affects the way in which individuals establish and maintain an awareness of what they are or would be; Luis Enrique Alonso (2005) includes consuming as a form of material appropriation of the social surplus, understood as the production, circulation, and use of signs, circumscribing all practices of each

social position; Naomi Klein (2002), aligned with her “No Logo,” dared to raise a frontal attack on a logoized world; and Zygmunt Bauman (2007) points to how today’s world, with its globalization and the Web, cannot be understood without simultaneously considering, to a greater or lesser extent, consumers being consumers and consumer’s objects themselves: “individuals become simultaneously the promoters of commodities and the commodities that they promote” (p. 6). Even Adela Cortina (2002) raised the need to work toward a more conscious and critical consumption, without necessarily revolutionizing or destroying the system but reforming from within those aspects that do not work properly. Requena (2009) raised the possible existence of “consumers of nows,” the so-called transumers who would collect and accumulate “nows” in intense and prolonged fashion. Effectively, September 11, 2001, and its subsequent consequences, among which is a strong feeling of vulnerability (the collapse of the twin towers, representing the heart of capitalism, violated the consumer society, which could happen anywhere) caused us to believe that no future could be imagined, resulting in a feeling of *carpe diem*: Consume “nows,” because we do not know if there will be “afters.”

Thus, “we are all consumers,” as Kennedy said in his famous 1962 speech: “We are the largest market economic group, affecting and affected by almost every public and private economic decision.” In this developed society, it is true that we do not understand/cannot imagine our identity(ies) without that idea of being a buyers’ being. We are consumers, recognize ourselves as such, and cannot understand our society (relationships, institutions, symbols, and myths) without this essence. The concept of “mass man” (a term coined by Ortega in 1929) refers to the human being as characterized by a large increase in the consumption of goods and services of all kinds.

However, we know that there are not “consumers,” targeted as such and predictable in behavior throughout their existence, but consumption

habits, moments, and behaviors. The financial crisis, globalization, and social networks have changed the way people indulge in the satisfaction of their needs. Traditionally, consumers were labeled according to their behaviors and habits and classified in a particular segment according to sex, age, habitat, and life moment. Today, these classifications are geared more toward consumption moments in which each person behaves according to the circumstances: sometimes rational, sometimes emotional, sometimes price based, sometimes hedonistic. For example, you can be a manager by day, a parent on weekends, and a player at night. Thus, segmenting consumers is increasingly difficult, especially when they are now better informed, have become more demanding, and are sovereign in their decisions. It seems that Belk and Pollay (1985;889) were premonitory when they said, “We are building our identity in a way that had never before tasted”.

BACKGROUND

Homo Sapiens and Homo Consumens Ludens

Perhaps the first subclassification in this new taxonomy is to differentiate between rational and emotional consumers. The SmartShopper is logical and rational and tries to maximize a personal return on investment when buying food, household items, and even clothes, while the consumer is looking for emotions and experiences and seeking to enjoy every purchase. It is, however, that old distinction between “go shopping” and “do the shopping.”

It seems logical and even obvious to think that when a buyer chooses to purchase a good or service that he or she should always opt for that object that will provide greater value. The satisfaction of a need understood as that which fills the gap between a state that is presented as an ideal and a current state that keeps us unhappy is not only

the starting point of any action in marketing but also at the bottom of all advertising messages. Indeed, any slogan or advertising creation can be translated as “I’m the best solution to your problem.” Thus, *Homo sapiens* as consumers seek this solution among the functional characteristics of objects, while *Homo ludens* would be set on the components of what is called “extended product” (Kotler & Armstrong, 1999). Products compete not so much with their technical characteristics, more efficient after-sales service, or more useful or attractive packaging but for those experiences and feelings that the product causes the consumer: peace, pride, sympathy for the brand, and so on. According to some authors, in this world of mature capitalism where everybody has achieved an acceptable (even affluent) materialistic status, what is important is not to own but to feel (Davis, 2003). Physical needs have long ceased to be part of our priorities, because now what we seek is mental happiness. Thus, products are chosen according to how they can increase that spiritual happiness. Miller (2001) raised the idea that when acquiring certain objects we act simply as social primates trying to impress the opposite sex through the reactions of rejection or attraction based on what we wear or show to others. The best example of this experiential consumption is transmitted by the famous advertisement of a prestigious German car brand in which the driver’s hand is outside the vehicle’s window, playing with the wind in time to pleasant music, while the ad asks, “Do you like to drive?” Moreover, buying products with certain social characteristics, apart from the intrinsic value of the product, provide a pleasant feeling of contributing to a better world. People who do not have time or are not willing to provide another form of solidarity (volunteering, etc.) can contribute to solving the world’s problems from their everyday shopping baskets.

From Latin to English: Consumer and Prosumer

When Alvin Toffler coined the term *prosumer* in 1980, he was possibly thinking of those who produce for their own consumption. Thus, as in the early stages of the history of mankind, only a small part of the population could have the privilege of not hunting, growing, harvesting, or meeting their needs by their own means. In the postmodern era, there are people who choose to grow their own food or make their own clothes, which is likely due not so much to a need to cover basic requirements (a correct analysis of the effectiveness/efficiency of cultivating your own garden in purely economic terms would not hold) but in that need of *Homo ludens* alluded to in the preceding paragraph to experience something. Growing your own tomatoes, even a single plant in a pot on your terrace, provides a sense of pride and peace and cures stress. The do-it-yourself trend, sponsored by a financial crisis that has left us with less money to spend, has made us aware of the exaggerated and pernicious era of “throwaway” consumerism in our pursuit of shared leisure, family time, and booming new housing markets. “If Toffler is right with the rise of active prosumers’, marketing managers face a challenging future, if not frustrating” (Kotler, 1986, p. 510).

FUTURE RESEARCH DIRECTIONS: SOVEREIGN CONSUMERS

This brief taxonomy of consumers in a 3.0 era may conclude posing a need to empower and return sovereignty to the consumer. In economics, the consumer is always treated with respect and affection for it is the *raison d’être* of the market, the reason why that occurs and the object of desire of brands competing for your will, your faithfulness,

and your pocket. When looking at an economic dictionary for a definition of consumer sovereignty, such entries as “characteristic of a free market system where consumers orient production” and “Idea according to which consumers ultimately decide what should be produced (or not) by the act of choosing what will buy it (and what not)” often appear. Ultimately, it is the idea of consumer empowerment that becomes an undisputed market operator. The current development model based on economic growth and personal achievement towards individuals’ enrichment today is no longer a valid pattern. The idea that the world would be a better place if each would improve individually has been proved erroneous, because the achievement of somebody’s welfare (the less), according to this model, is at the expense of others (the more). The paradigm that happiness would come for possession of consumer goods and objects is taking us to a society of accumulation, which addresses us waste, pollution and injustice. The problem of being rich is the obligation of being happy every day.

Consumer, allegedly subject of rights and duties, can (although sometimes don’t want to) exercise them. In Europe theoretically, any citizen has the right to buy only what you want to actually buy. In practice this is not true. Consumer Rights are largely unknown and distant to the citizen and are written entirely with the individual consumer in mind. Protecting their safety, their health and their economic interests; promoting information and education of consumers and users with full freedom to choose (but don’t forget to choose), etc. Regarding duties is a simpler thing: the unique duty for a consumer seems to go shopping and pay. Rarely is referred the duty of every consumer to be informed about production conditions (social and environmental) of what he/she is buying, neither the responsibility to reduce consumption

However, this feature of a sovereign omnipotent that with its preferences guides the economy is not entirely true or tenable. In a competitive market and based on unbridled world consumption, the

trick is to convince consumers that they are free to choose whatever they want, whenever they want it. Like the absolute monarchs in the enlightened despotism of the 18th century who used their authority to introduce reforms in the political and social structures of their countries, we seem to be currently attending to an illustrated consumerism: Everything for the consumer but the consumer.

Precisely, this primacy of the individual consumer as market owner and master of this new concept of consumer sovereignty is counterpoised. If food sovereignty is the right of people to control their agricultural policies, to decide what to grow, to produce locally respecting the territory, and to control the natural resources (water, seeds, soil), then consumer sovereignty should be understood as the right of individuals to decide collectively and responsibly what, why, and where they want to consume. Consumer sovereignty, also known as “Consumocracy”, could be defined then as “a system by which consumers can exert moral or other authority on enterprises through a more enlightened selection of consumer products. It is therefore a voluntary regulation system within which corporate behavior is in part subordinated to consumer demands obeying both logics of individualism and voluntary solidarity”¹.

The market mechanism should serve as a form of political participation in which consumers move to rationality and utilitarianism as criteria fundamental to global transformation that puts people’s behaviors, the planet, and consumer relations in the center of the decision. Political consumerism is an expression used by Micheletti (2003, 2001), Andersen and Tobiasen, (2001) Barnett et al, (2005) Follesdal, (2006), Ballesteros (2007) that should be considered as the fundamentals of a new Theory of Revolutionary Consumption based in the well know French Revolution principles of Liberté, Egalité, Fraternité. As several authors have observed, for some individuals consumer behaviour is perceived as a means of making a public statement of personal values and beliefs, and/or as a means of influencing policy-making.

Authors have named this movement ethical consumerism (Shaw and Shiu, 2003; Shaw et al, 2006), consumer activism (Barnett et al, 2005; Smith, 2005) or political consumerism.

Micheletti's (2003) definition of political consumerism “actions by people who make choices among producers and products with the goal of changing objectionable institutional or market practices” (p. 2) assumes that political action can only occur through purchasing. This idea is reinforced when she explains the three types of political action. “[It] can take the form of exit, voice and loyalty, which equals boycotts, demands on producers, and smart shopping” (p. 15). Other authors have defined political consumerism in the same way, taking for granted that consumers can only express their political choices through purchasing (see Barnett et al, 2005; Follesdal, 2006; Shaw et al, 2006).

Consumer activism should therefore be considered as a new social movement (Hollenbeck and Zinkham, 2006, Kozinets and Handelman, 2004), and it can be done a distinction between two types of responsible consumers: reformers and revolutionaries, in the sense of the classifications of new social movements made by Held and McGrew (2002), Kiely (2000) and Starr, (2000). The difference between these two types is better understood by analyzing their targets and the form of power they are exercising. Let us focus first on the targets. The main difference between reformist consumers and revolutionary consumers lies in the targets they choose to attack. Reformists seek changes within the market, without contesting the current paradigm. Reformists choose the best performers and punish the worst ones: reformists boycott unethical brands and reinforce the best-in-class manufacturers. Therefore, their targets are specific examples of best (e.g. fair trade, organic food, best-in-class performer) and worst ways of production (e.g. unethical companies). However, for revolutionary consumers framing a decision as a choice between brand A and brand B is limiting the decision to the lesser of two evils. Revolutionary consumers, on the other hand, think first of

their needs, assess whether they are peremptory or not, and then decide whether they can satisfy them without engaging in an economic transaction. They believe that many purchasing decisions are not individual, rational or sovereign. On the contrary, individuals' purchasing decisions are heavily influenced by “contextual social forces” (e.g. corporate messages) and “structural forces” that make it necessary to increase or maintain consumption levels (Conca et al, 2001). Their targets are these nodes of power shaping consumption choices: they attack the structural roots of the social, economic, and political problems.

CONCLUSION: CONSUMER 3.0

The new consumer's conceptualization envisaged in this chapter is clearly aligned with the information science and technology society. The unlimited access to information that allows a more and better (empowered) consumer is crucial. The “prosumption” activity in social media is also a phenomenon broadly discussed in the literature as social networks in turn contribute to this need to create and produce as well as to consume.. A large number of Web 2.0 users are simultaneously content producers and consumers. YouTube channels, Facebook pages, Pinterest boards, Instagram photos.

From other perspective, social networking is clearly a consumer's activism tool. Worldwide consumers' unions are connected and develop campaigns through the internet. The firms' control by the consumers is permitted and enhanced through the exchange of knowledge that the information technologies allow. Apps for cellular phones which permits on-line comparison of prices and qualities while at the supermarket; web comparators that allows a smart-shopping behavior; the possibility to directly write to the company that market a product with an immediate response from the brand managers, etcetera, are good examples of how the new technologies contributes to this consumer empowerment.

REFERENCES

- Alonso, L. E. (2005). *La era del consumo*. Madrid, Spain: Siglo XXI.
- Andersen, J. G., & Tobiasen, M. (2001). *Political Consumption and Political Consumers. Globalization and Political Consumption in Everyday Life*. Aarhus, Denmark: Institut for Statskundskab.
- Ballesteros, C. (2007). *Tu compra es tu voto*. Madrid, Spain: HOAC.
- Barnett, C., Clarke, N., Cloke, P., & And Malpass, A. (2005). The political ethics of consumerism. *Consumer Policy Review*, 15(2), 45–51.
- Bauman, Z. (2007). *Consuming life*. Cambridge, UK: Polity Press.
- Belk, R., & Pollay, R. W. (1985). Images of ourselves: The good life in twentieth century advertising. *The Journal of Consumer Research*, 11(4), 887–897. doi:10.1086/209024
- Bocock, R. (1995). *El consumo*. Madrid, Spain: Talasa.
- Conca, K., Princen, T., & Maniates, M. F. (2001). Confronting consumption. *Global Environmental Politics*, 1(3), 1–10. doi:10.1162/152638001316881377
- Cortina, A. (2002). *Por una ética del consumo*. Madrid, Spain: Taurus.
- Davis, M. (2003). *Por qué deseamos lo que deseamos?* Barcelona, Spain: Urano.
- Follesdal, A. (2006). Political consumerism as chance and challenge. In M. Micheletti, A. Follesdal, & D. Stolle (Eds.), *Politics, products, and markets. Exploring political consumerism past and present*. Transaction Publishers.
- Fromm, E. (Ed.). (1965). *Socialist humanism: An international symposium*. New York, NY: Doubleday.
- Held, D., & McGrew, A. (2002). *Globalization/Anti-globalization*. Cambridge, UK: Polity Press.
- Hollenbeck, C. R., & Zinkham, G. M. (2006). Consumer activism on the Internet: The role of anti-brand communities. *Advances in Consumer Research. Association for Consumer Research (U. S.)*, 33, 479–485.
- Kennedy, J. F. (1962). Special message to the Congress on protecting the consumer interest. *The American Presidency Project*. Retrieved from <http://www.presidency.ucsb.edu/ws/?pid=9108>
- Klein, N. (2002). *No logo*. Madrid, Spain: Paidós Iberica.
- Kotler, P. (1986). The prosumer movement: A new challenge for Marketers. In R. J. Lutz (Ed.), *Advances in consumer research* (Vol. 13, pp. 510–513). Provo, UT: Association for Consumer Research.
- Kotler, P., & Armstrong, G. (1999). *Principles of marketing* (8th ed.). Upper Saddle River, NJ: Prentice Hall.
- Kozinets, R. V., & Handelman, J. M. (2004). Adversaries on consumption: Consumer movements, activism, and ideology. *The Journal of Consumer Research*, 31(4), 691–704. doi:10.1086/425104
- Kyeli, R. (2002). Actually existing globalization, deglobalization and the political economy of anticapitalist protest. *Historical Materialism*, 10(1), 93–121. doi:10.1163/156920602760231677
- Lipovetsky, G. (2006). *Le bonheur paradoxal: Essai sur la société d'hyperconsommation*. Paris, France: Gallimard.
- Micheletti, M. (2001) Why Political Consumerism? A Theoretical Context for Understanding the Phenomenon. *International Seminar on Political Consumerism*. City University.

Micheletti, M. (2003). *Political virtue and shopping. Individuals, consumerism, and collective action*. New York: Palgrave MacMillan. doi:10.1057/9781403973764

Miller, D. (2001). The poverty of morality. *Journal of Consumer Culture*, 1(2), 225–243. doi:10.1177/146954050100100210

Ortega, J. (1929). *La rebelion de las masas*. Madrid, Spain: Espasa.

Requena, G. (2009, December 16). Mega-tendencias: Qué nos depara el 2010. *Marketing News*. Retrieved from <http://www.marketingnews.es/varios/opinion/1044478028705/mega-tendencias-depara-2010.1.html>

Saad, G. (2007). *The evolutionary bases of consumption*. Mahwah, NJ: Lawrence Erlbaum.

Shaw, D., Newholm, T., & Dickinson, R. (2006). Consumption as voting: An exploration of consumer empowerment. *European Journal of Marketing*, 40(9/10), 1049–1067. doi:10.1108/03090560610681005

Shaw, D., & Shiu, E. (2003). Ethics in consumer choice: A multivariate modeling approach. *European Journal of Marketing*, 37(10), 1485–1498. doi:10.1108/03090560310487202

Starr, A. (2000). *Naming the enemy: anti-corporate movements confront globalization*. London, UK: Zed Books.

Toffler, A. (1980). *La tercera ola*. Barcelona, Spain: Plaza & Janes.

ADDITIONAL READING

Dabholkar, P. A. (1996, February). Consumer evaluations of new technology-based self-service options: An investigation of alternative models of service quality International. *Journal of Research in Marketing*, 13(1), 29–51. doi:10.1016/0167-8116(95)00027-5

Hernandez, M. A. (2012). *Marketing digital: Marketing móvil, SEO y analítica web*. Madrid, Spain: Anaya.

Hoyer, W. D., Chandy, R., Dorotic, M., Krafft, M., & Singh, S. S. (2010, August). Consumer Cocreation in New Product Development. *Journal of Service Research*, 13(3), 283–296. doi:10.1177/1094670510375604

KEY TERMS AND DEFINITIONS

Consumer Sovereignty: Characteristic of a free market system where consumers orient production.

Crowdsomers: Consumers who make use of the crowd platforms, both to take advantage of group purchasing and to participate in financing cultural projects or creating solidarity with minimum contributions (but operating in massive number).

Knowcosters: A movement initiated in Spain and posed by the need to consume as one thinks, knowing and being aware, in addition to the visible product costs (reflected in the retail price) of all indirect costs, which are not seen but paid by us, including planet welfare state cost.

Localsomers: Consumers rooted in their local consumption.

Presumers/Preconsumers: People involved in the process of creating products and services. By bringing an accelerated version to market, product forms often make companies decide to launch beta versions, which are improved and complemented with market use.

Trysumers: Consumers wanting to try new things.

Virgin Consumers: Consumers who increasingly meet with products that they do not know how to use, processes with which they are unfamiliar, and brands that they do not know and yet they are not afraid and decide to experiment, test, and comment on their experiences via social networks.

Wikisumers: Consumers or collaboratives exercise their right to consumption in nontraditional ways by sharing, exchanging, loaning, renting, and giving, all redefined through modern technology and communities.

ENDNOTE

- ¹ <http://consumarchy.blogspot.com/2006/03/what-is-consumarchy.html>

Category C

Cyber and Network Security

Cyber Security Protection for Online Gaming Applications



Wenbing Zhao

Cleveland State University, USA

INTRODUCTION

By online gaming applications, we mean both distributed applications that enable large number of users to play multiplayer games and those that enable people to gamble online because both types of applications could have huge financial stakes and the security and dependability challenges for both are rather similar. On the one hand, such systems must ensure continuous high availability so that users around the globe could play the games 24 by 7. This requires that the game servers be replicated to provide non-stop services. On the other hand, state-machine replication requires that the replicas be deterministic or rendered deterministic. This requirement does not work well with gaming applications because random numbers are essential to their operation. For example, in a card game, the random numbers are used to shuffle the cards. If the random numbers used are not robust, the hands in the card game may become predictable, which could damage the integrity of the game and may lead to financial losses to the game provider and/or honest game players. The nature of this type of applications poses a particular challenge to ensure cyber security because it is difficult to ensure high availability while preserving the integrity of the operation of these applications (Arkin et al., 1999; Viega & McGraw, 2002; Young & Yung, 2004; Zhao, 2007; Zhao, 2008; Zhang et al., 2011).

Byzantine fault tolerance (Castro & Liskov, 2002) is a well-known technique to achieve cyber security (Zhao, 2014). The technique aims to tolerate various malicious attacks to online systems by employing state machine replication (Schneider,

1990). However, as we mentioned earlier, Byzantine fault tolerance cannot be used as it is because it is not equipped with built-in solution to resolve the conflict of replication determinism requirement and the intrinsic randomness of the server operation. In this article, we elaborate how we address this dilemma using an online poker game application as a running example. In this application, a pseudo-random number generator (PRNG) is used to generate the pseudo-random numbers for shuffling the cards. We present two alternative strategies to cope with the intrinsic application nondeterminism. One depends on a Byzantine consensus algorithm and the other depends on a threshold signature scheme. Furthermore, we thoroughly discuss the strength and weaknesses of these two schemes.

BACKGROUND

In this section, we provide a brief introduction of PRNG, the entropy concept, and the methods to collect and enhance entropy.

A PRNG is a computer algorithm used to produce a sequence of pseudo-random numbers. It must be initialized by a seed number and can be reseeded prior to each run. The numbers produced by a PRNG are not truly random because computer programs are in fact deterministic machines. Given the same seed, a PRNG will generate the same sequence of numbers. Consequently, if an adversary knows the seed to a PRNG, then he/she can generate and predict the entire stream of random numbers (Young & Yung, 2004). Therefore, to make the random numbers unpredictable,

it is important that the seeds to the PRNG cannot be guessed or estimated. Ideally, a highly random number that is unpredictable and infeasible to be computed is required to seed the PRNG in order to produce a sequence of random numbers.

The activity of collecting truly random numbers is referred to as “collecting entropy” by cryptographers (Young & Yung, 2004). Entropy is a measure of the degree of randomness in a piece of data. As an example, consider using the outcome of coin flipping as 1 bit of entropy. If the coin-toss is perfectly fair, then the bit should have an equal chance of being a 0 or a 1. In such a case, we have a perfect 1 bit of entropy. If the coin-toss is slightly biased toward either head or tail, then we have something less than 1 bit of entropy. Entropy is what we really want when we talk about generating numbers that cannot be guessed. In general, it is often difficult to figure out how much entropy we have, and it is usually difficult to generate a lot of it in a short amount of time.

It is a common practice to seed a PRNG with the current timestamp. Unfortunately, this is not a sound approach to preserve the integrity of the system, as described by Arkin et al (1999) in the context of how a Texas Hold'em Poker online game can be attacked. They show that with the knowledge of the first few cards, they can estimate the seed to the PRNG and subsequently predict all the remaining cards.

TECHNIQUES FOR ENHANCING THE TRUSTWORTHINESS

In this section, we describe two possible strategies for enhancing the trustworthiness of online gaming applications. One depends on a Byzantine consensus algorithm and the other depends on a threshold signature algorithm. Both algorithms ensure that all replicas adopt the same value to seed their PRNGs, while each replica is taking entropy from its respective entropy source.

Byzantine Fault Tolerance

A Byzantine fault (Lamport, Shostak, & Pease, 1982) is a fault that might bring a service down, or compromise the integrity of a service. A Byzantine faulty replica may use all kinds of methods to disrupt the normal operation of a system. In particular, it might propagate conflicting information to other replicas. To tolerate f Byzantine faulty replicas in an asynchronous environment, we need to have at least $3f+1$ number of replicas (Castro & Liskov, 2002). An asynchronous environment is one that has no bound on processing times, communication delays, and clock skews. Internet applications are often modeled as asynchronous systems. Usually, one replica is designated as the primary and the remaining ones as backups.

Any robust Byzantine fault tolerance (BFT) algorithm can be modified to cope with the use of random numbers. In the following, we describe a solution based on the well-known Practical BFT (PBFT in short) algorithm developed by Castro and Liskov (2002). The PBFT algorithm has three communication phases in normal operation. During the first phase, the pre-prepare phase, upon receiving a request from the client, the primary assigns a sequence number and the current view number to the message and multicasts a Pre-Prepare message to all backups. In the second phase, referred to as the prepare phase, a backup broadcasts a Prepare message to the rest of replicas after it accepts the Pre-Prepare message. Each non-faulty replica enters into the commit phase, i.e., the third phase, only if it receives $2f$ Prepare messages (from other replicas) that have the same view number and sequence number as the Pre-Prepare message, then it broadcasts the Commit message to all replicas including the primary. A replica commits the corresponding request after it receives $2f$ matching commit messages from other replicas. To prevent a faulty primary from intentionally delaying a message, the client starts a timer after it sends out a request and waits for $f+1$ consistent responses from different replicas. Due to the assumption that at most f replicas can

be faulty, at least one response must have come from a non-faulty replica. If the timer expires, the client broadcasts its request to all replicas. Each backup replica also maintains a timer for similar purposes. On expiration of their timers, the backups initiate a view change and a new primary is selected. In the PBFT algorithm, digital signature or an authenticator is employed to ensure the integrity of the messages exchanged.

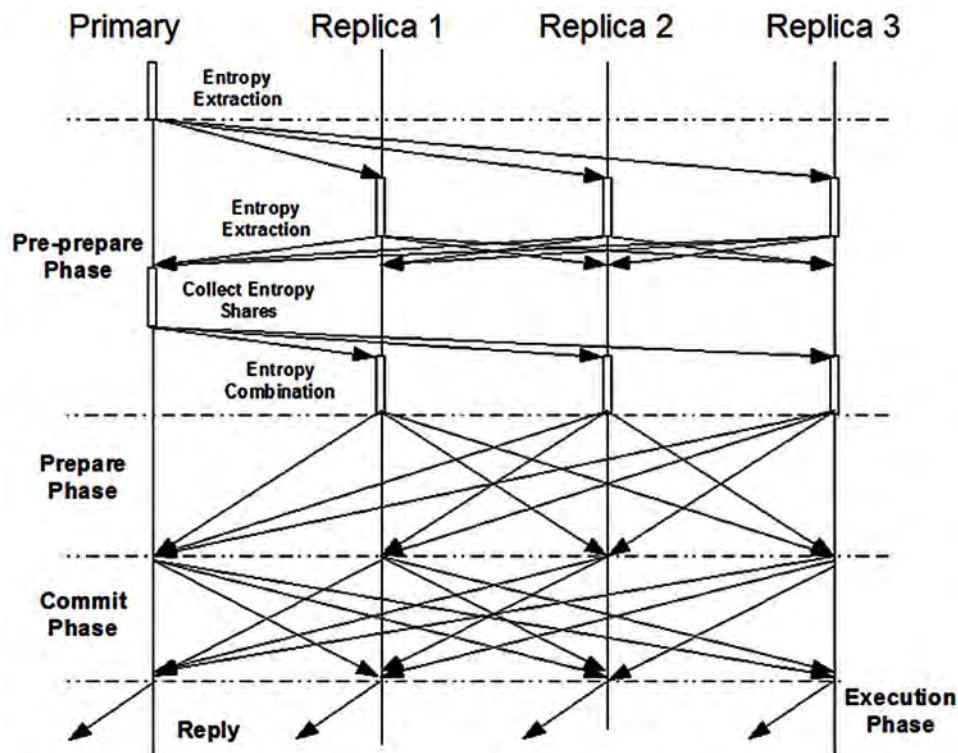
The above PBFT algorithm is modified in the following way to cope with the replica nondeterminism caused by the use of random numbers. The modified algorithm also consists of three phases, as shown in Figure 1. In the beginning of the first phase, the primary invokes ENTROPY-EXTRACTION operation to extract its entropy and append the entropy to the Pre-Prepare message. It then multicasts the Pre-Prepare message to the backups. Each replica records the primary's entropy from the Pre-Prepare message in its log and then invokes the ENTROPY-EXTRACTION operation to obtain its own share of entropy as well.

Each backup then multicasts a Pre-Prepare-Update message, including its share of entropy extracted. When the primary collects $2f$ Pre-Prepare-Update messages from the backups, it constructs a Pre-Prepare-Update message, including the digest of the $2f+1$ entropy shares ($2f$ received, plus its own), together with the corresponding contributor's identity, and multicasts the message.

Upon receiving the Pre-Prepare-Update message from the primary, each replica invokes the ENTROPY-COMBINATION operation to combine the entropy from the $2f+1$ shares. The outcome of the ENTROPY-COMBINATION operation ensures a highly random number, due to the contributions from the non-faulty replicas. The combined number is provable secure and will be used to seed the PRNG if the BFT algorithm terminates.

The second and third phases are similar to the corresponding phases of the PBFT algorithm, except each replica will append the digest of the entropy set determined in the first phase to the

Figure 1. The adapted BFT algorithm to handle entropy extraction and combination



Prepare and Commit messages. These two phases are necessary to ensure all nonfaulty replicas to agree on the same message total ordering and the entropy value despite the presence of Byzantine faulty replicas.

We now highlight the details of the ENTROPY-EXTRACTION and the ENTROPY-COMBINATION operations.

Entropy-Extraction

The ENTROPY-EXTRACTION operation is based on software-based entropy collection. There are a number of techniques that can be used to extract the entropy, most of which are based on the timing of internal activities in a computer (Young & Yung, 2004). A well-known technique is called TrueRand, developed by Don Mitchell and Matt Blaze. The idea behind TrueRand is to gather the underlying randomness from idle CPUs by measuring the drift between system clock and the generation of interrupts on the processor. Other frequently used techniques include recording network traffic as it comes into the server, timing how long it takes to seek the disk, and capturing kernel state information that changes often.

Entropy-Combination

The ENTROPY-COMBINATION operation combines the $2f+1$ entropies the replica collected using the exclusive-or (XOR) operator (Young & Yung, 2004). This operation has several benefits.

First, it combines a number of weak sources of entropy to form an effective strong entropy source. Consider two entropy sources from coin flipping, and the case which source 1 results in head and source 2 results in tail, or source 1 results in tail and source 2 results in head, and the probability for source 1 to result in head is $10/16$ and that for source 2 is $12/16$ (i.e., both are biased). If we combine these two sources, the probability of getting a head is $7/16$. This shows that the coin flipping resulting from XOR-ing the bits from the two sources is the same or better than the best

flip in either of the two sources. Furthermore, the more sources we use, the higher entropy we get.

Second, the ENTROPY-COMBINATION operation eliminates any negative impact from malicious replicas. Among the $2f+1$ entropy shares collected, there can be up to f of them coming from faulty replicas. The XOR operation guarantees that if at least one share coming from a good entropy source, the combined entropy is at least as good as that entropy share. This requirement is met because there are at least $f+1$ shares contributed by non-faulty replicas. Any low-entropy or predictable shares generated by faulty replicas are virtually ignored.

Third, the ENTROPY-COMBINATION operation results in a single high entropy share used by all non-faulty replicas, which ensures the consistency of the replicas when they are involved with intrinsically nondeterministic operations.

Threshold Signature

The other strategy to ensure consistent Byzantine fault tolerance replication for nondeterministic operations is to employ threshold cryptography (Deswarte, Blain, & Fabre, 1991; Desmedt & Frankel, 1990). Threshold cryptography is a good way to distribute trust among a group of players to protect either information or computation (Zhou, Schneider, & Renesse, 2002).

A well-known secret sharing scheme (Shamir, 1979) is the (k, n) threshold digital signature scheme. In the (k, n) secret sharing scheme, a secret is divided into n sets and distributed to the same number of players. The secret can be reconstructed if k out of n number of players could combine their shares. However, fewer than k players cannot collude to forge the secret. The (k, n) threshold digital signature scheme allows a set of servers to collectively generate a digital signature in a way similar to reconstructing a secret using the (k, n) secret sharing scheme.

In the (k, n) threshold digital signature scheme, a private key is divided into n shares, each owned by a player. To produce a valid threshold digital

signature, at least k players must pool their shares together (i.e., fewer than k players would not be able to generate a valid threshold digital signature). Each player uses its private key share to generate a partial signature on a message and these partial signatures can be combined into a threshold signature on the message. The threshold signature can be verified using the public key corresponding to the divided private key.

The RSA Shoup scheme (Shoup, 2000) is one of the practical threshold digital schemes. In this scheme, a dealer generates a key pair and divides the private key into n shares at first. Each key share has a key verifier. Then the dealer distributes the message to be signed and the key shares to n players. Each player uses its key share to generate a partial signature on the message. Furthermore, each player sends the signed message with the verifier to a trusted server, which verifies the signature shares and combines the partial signatures into a threshold signature verifiable by the public key.

In the following, we show how to integrate the threshold digital signature scheme with Byzantine fault tolerance for online gaming applications. The adapted PBFT algorithm consists of three phases (under normal operation) for Byzantine agreement and an additional phase run at the beginning for key shares distribution. The Byzantine agreement algorithm works similar to the PBFT algorithm except the third phase, where each replica generates a partial signature (using its key share) to sign the client's message and piggybacks the partial signature to the Commit message. Each replica combines the partial signatures into a threshold signature. The signature is then mapped into a number to seed the PRNG.

Despite the elegance of the threshold signature, the algorithm, however, might not be practical in the Internet environment. First of all, it depends on a trusted dealer at the beginning to generate a key pair, divide the private key into several key shares and it must also be responsible for distributing the key shares to all replicas. If the dealer is compromised, the entire system would become vulnerable to adversaries. Furthermore,

the threshold signature is computationally expensive, especially when generating the threshold signature. For example, for a 1024-bit threshold signature it usually takes 73.9ms on a PC equipped with a single 1.0GHz Pentium III CPU and 1.5 GB RAM (Rhea et al., 2003). Furthermore, the validity on the use of the threshold signature as the seed to the PRNG remains to be proved secure.

FUTURE TRENDS

Not only for online gaming applications, PRNGs are widely used in nearly every field in computer and networking area. In particular, for cryptography, the access to truly random numbers is extremely important. Even though there is moderate success in implementing PRNG, it remains to be vulnerable under cryptanalytic attacks and the attacks against its internal state. Furthermore, it is easy to see that even if the PRNG is robust against many potential threats, once the seed is discovered the numbers generated by the PRNG are no longer unpredictable. In light of this observation, more efforts should be engaged in how to gather and evaluate entropy in a secure and dependable manner. The research described in this article can be regarded as the first step towards this direction.

There are many open issues to be resolved before we can confidently apply these techniques in practice. The most interesting research issue is how to maintain replica consistency. Common Byzantine fault tolerance techniques require deterministic execution of replicas despite the fact that all practical applications contain some degree of nondeterminism (Zhao, 2007), for example, clock values, CPU speed, multithreading, etc. Note that these types of nondeterminism are not considered as good entropy sources accordingly to the cryptography standard, but the presence of these types of nondeterminisms nevertheless poses a big threat to maintaining replica consistency. We recently proposed a Byzantine fault tolerance framework that is able to handle various

nondeterministic applications in a systematic and efficient manner (Zhang et al., 2011).

So far we have focused on addressing the issues for the client-server based online gaming applications. Peer-to-peer (P2P) based online games have become popular in recent years. The P2P design brings additional challenges to enhancing trustworthiness of such applications because each client maintains substantial private state. There are new attacks that may be launched from clients to the systems. Various schemes have been designed to enforce the integrity of such applications (Jha, 2007; Wierzbicki, 2006).

Finally, we only addressed the high availability and the associated integrity issues for online gaming applications. Additional research is needed to address other aspects of risks faced by such applications. For example, a player could cheat by changing the rendering routines so that he/she could see scenes that are not allowed, thereby gaining an unfair advantage over honest players. These types of threats cannot be addressed by resorting to replication.

CONCLUSION

In this article, we pointed out the threats to online gaming applications and presented two strategies that can be used to build secure and dependable online gaming applications. These strategies not only seek the solution for gathering entropy to seed the PRNG used in such applications, but also intend to eliminate malicious intrusions to protect the seed and to maintain replica consistency. By applying these techniques, the online gaming applications can ensure its service integrity (both the service providers and the innocent players are protected) and guarantee high availability despite the presence of Byzantine faults. Finally, we outlined some open research issues in this field.

REFERENCES

- Arkin, B., Hill, F., Marks, S., Scjmod, M., & Walls, T. (1999). *How we learned to cheat at on-line poker: A study in software security*. Available at http://www.developer.com/java/other/article.php/10936_616221
- Castro, M., & Liskov, B. (2002). Practical Byzantine fault tolerance and proactive recovery. *ACM Transactions on Computer Systems*, 20(4), 398–461. doi:10.1145/571637.571640
- Desmedt, Y., & Frankel, Y. (1990). Threshold cryptosystems. *Lecture Notes in Computer Science*, 435, 307–315. doi:10.1007/0-387-34805-0_28
- Deswarte, Y., Blain, L., & Fabre, J. (1991). Intrusion tolerance in distributed computing systems. *Proceedings of the IEEE Symposium on Research in Security and Privacy*, 110–121.
- Jha, S., Katzenbeisser, S., Schallhart, C., Veith, H., & Chenney, S. (2007, May). Enforcing semantic integrity on untrusted clients in networked virtual environments. In *Proceedings of the IEEE Symposium on Security and Privacy*, (pp. 179–186). IEEE.
- Lamport, L., Shostak, R., & Pease, M. (1982). The Byzantine generals problem. *ACM Transactions on Programming Languages and Systems*, 4(3), 382–401. doi:10.1145/357172.357176
- Luby, M. (1996). *Pseudorandomness and Cryptographic Applications*. Princeton University Press.
- Rhea, S., Eaton, P., Geels, D., Weatherspoon, H., Zhao, B., & Kubiatowicz, J. (2003). Pond: the OceanStore Prototype. *Proceedings of the 2nd USENIX Conference on File and Storage Technology*, 1–14.
- Rivest, R., Shamir, A., & Adleman, M. (1978). A method for obtaining digital signatures and public key cryptosystems. *Communications of the ACM*, 21(2), 120–126. doi:10.1145/359340.359342

Schneider, F. (1990). Implementing fault-tolerant services using the state machine approach: A tutorial. *ACM Computing Surveys*, 22(4), 299–319. doi:10.1145/98163.98167

Shamir, A. (1979). How to share a secret. *Communications of the ACM*, 22(11), 612–613. doi:10.1145/359168.359176

Shoup, V. (2000). Practical threshold signature. *Lecture Notes in Computer Science*, 1807, 207–223. doi:10.1007/3-540-45539-6_15

Viega, J., & McGraw, G. (2002). *Building Secure Software: How to Avoid Security Problems the Right Way*. Addison-Wesley.

Wierzbicki, A. (2006). Trust enforcement in peer-to-peer massive multi-player online games. *Lecture Notes in Computer Science*: vol. 1180. On the Move to Meaningful Internet Systems 2006: CoopIS, DOA, GADA, and ODBASE (pp. 1163–1180). Springer Berlin Heidelberg. doi:10.1007/11914952_7

Young, A., & Yung, M. (2004). *Malicious Cryptography: Exposing Cryptovirology*. Indianapolis, IN: Wiley Publishing.

Zhang, H., Zhao, W., Moser, L. E., & Melliar-Smith, P. M. (2011). Design and implementation of a Byzantine fault tolerance framework for non-deterministic applications. *Software, IET*, 5(3), 342–356. doi:10.1049/iet-sen.2010.0013

Zhao, W. (2007). Byzantine Fault Tolerance for Nondeterministic Applications. *Proceedings of the 3rd IEEE International Symposium on Dependable, Autonomic and Secure Computing*, 108–115. doi:10.1109/DASC.2007.11

Zhao, W. (2008, December). Integrity-preserving replica coordination for byzantine fault tolerant systems. In *Proceedings of the 14th IEEE International Conference on Parallel and Distributed Systems* (pp. 447–454). IEEE. doi:10.1109/ICPADS.2008.45

Zhao, W. (2014). *Building dependable distributed systems*. John Wiley & Sons. doi:10.1002/9781118912744

Zhou, L., Schneider, F., & Renesse, R. (2002). COCA: A secure distributed online certification authority. *ACM Transactions on Computer Systems*, 20(4), 329–368. doi:10.1145/571637.571638

ADDITIONAL READING

Atighetchi, M., Pal, P., Webber, F., Schantz, R., Jones, C., & Loyall, J. (2004). Adaptive cyber defense for survival and intrusion tolerance. *IEEE Internet Computing*, 8(6), 25–33. doi:10.1109/MIC.2004.54

Baughman, N. E., & Levine, B. N. (2001). Cheat-proof payout for centralized and distributed online games. In *Proceedings of the Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies*. (Vol. 1, pp. 104–113). IEEE. doi:10.1109/INFCOM.2001.916692

Bessani, A. N., Sousa, P., Correia, M., Neves, N. F., & Verissimo, P. (2007). Intrusion-tolerant protection for critical infrastructures.

Bessani, A. N., Sousa, P., Correia, M., Neves, N. F., & Verissimo, P. (2008). The CRUTIAL way of critical infrastructure protection. *Security & Privacy, IEEE*, 6(6), 44–51. doi:10.1109/MSP.2008.158

Chen, T., Hsiao, T.-C., & Chen, T.-L. (2004). An efficient threshold group signature scheme. *Proceedings of the IEEE Region 10 Conference*, Volume B (pp. 21–24). Chiang Mai, Thailand.

Deswarte, Y., & Powell, D. (2006). Internet security: An intrusion-tolerance approach. *Proceedings of the IEEE*, 94(2), 432–441. doi:10.1109/JPROC.2005.862320

- Feng, D., & Xiang, J. (2005, July). Experiences on intrusion tolerance distributed systems. In *Proceedings of the 29th Annual International Computer Software and Applications Conference* (Vol. 1, pp. 270-271). IEEE.
- Garcia, M., Bessani, A., Gashi, I., Neves, N., & Obelheiro, R. (2011, June). OS diversity for intrusion tolerance: Myth or reality? In *Proceedings of the IEEE/IFIP 41st International Conference on Dependable Systems & Networks* (pp. 383-394). IEEE.
- Hampel, T., Bopp, T., & Hinn, R. (2006, October). A peer-to-peer architecture for massive multiplayer online games. In *Proceedings of 5th ACM SIGCOMM workshop on Network and system support for games* (p. 48). ACM. doi:10.1145/1230040.1230058
- Hoglund, G., & McGraw, G. (2007). *Exploiting online games: cheating massively distributed systems*. Addison-Wesley Professional.
- Hsiao, T. Y., & Yuan, S. M. (2005). Practical middleware for massively multiplayer online games. *IEEE Internet Computing*, 9(5), 47–54. doi:10.1109/MIC.2005.106
- Iimura, T., Hazeyama, H., & Kadobayashi, Y. (2004, August). Zoned federation of game servers: a peer-to-peer approach to scalable multi-player online games. In *Proceedings of 3rd ACM SIGCOMM workshop on Network and system support for games* (pp. 116-120). ACM. doi:10.1145/1016540.1016549
- Kim, J., Bentley, P. J., Aickelin, U., Greensmith, J., Tedesco, G., & Twycross, J. (2007). Immune system approaches to intrusion detection—a review. *Natural Computing*, 6(4), 413–466. doi:10.1007/s11047-006-9026-4
- Li, S., Chen, C., & Li, L. (2007, November). Using group interaction of players to prevent in-game cheat in network games. In *Proceedings of the First International Symposium on Data, Privacy, and E-Commerce*, (pp. 47-49). IEEE. doi:10.1109/ISDPE.2007.112
- Li, S., Chen, C., & Li, L. (2008, January). Prevent In-Game Cheat in Network Games. In *Proceedings of the First International Workshop on Knowledge Discovery and Data Mining*, (pp. 513-518). IEEE. doi:10.1109/WKDD.2008.95
- Loose, T. C. (2014). U.S. Patent No. 8,632,405. Washington, DC: U.S. Patent and Trademark Office.
- Mettouris, C., Maratou, V., Vuckovic, D., Papadopoulos, G. A., & Xenos, M. (2015). Information Security Awareness through a Virtual World: An end-user requirements analysis. In *Proceedings of the 5th International Conference on Information Society and Technology* (pp. 273-278).
- Nae, V., Prodan, R., & Fahringer, T. (2010, January). Monitoring and fault tolerance for real-time online interactive applications. In *Proceedings of the Euro-Par 2009—Parallel Processing Workshops* (pp. 255-265). Springer Berlin Heidelberg. doi:10.1007/978-3-642-14122-5_30
- Nguyen, B. T., Wolf, B. D., & Underdahl, B. (2014). U.S. Patent No. 8,708,791. Washington, DC: U.S. Patent and Trademark Office.
- Yan, J. (2003, December). Security design in online games. In *Proceedings of the 19th Annual Computer Security Applications Conference*, (pp. 286-295). IEEE.
- Yan, J., & Randell, B. (2005, October). A systematic classification of cheating in online games. In *Proceedings of 4th ACM SIGCOMM workshop on Network and system support for games* (pp. 1-9). ACM. doi:10.1145/1103599.1103606
- Yan, J., & Randell, B. (2005). Security in computer games: From pong to online poker. University of Newcastle upon Tyne. *Computer Science*.
- Yan, J., & Randell, B. (2009). An investigation of cheating in online games. *Security & Privacy, IEEE*, 7(3), 37–44. doi:10.1109/MSP.2009.60

Zhao, W. (2008, May). Towards practical intrusion tolerant systems: a blueprint. In Proceedings of the 4th Annual Workshop on Cyber security and information intelligence research: developing strategies to meet the cyber security and information intelligence challenges ahead (p. 19). ACM. doi:10.1145/1413140.1413162

KEY TERMS AND DEFINITIONS

Byzantine Fault Tolerance: A replication-based technique used to ensure high availability of an application subject to Byzantine fault.

Byzantine Fault: It is used to model arbitrary fault. A Byzantine faulty process might send conflicting information to other processes to prevent them from reaching an agreement.

Dependable System: A dependable system is one that is trustworthy to its users. It requires that the system to be highly available (to legitimate users) while ensuring high degree of service integrity.

Digital Signature: A digital signature aims to serve as the same purposes as a real-world signature. A sound digital signature ensures that the sender of the digital signature can be authenticated, the sender cannot later repudiate that she has sent the signed message, and a receiver cannot forge a digital signature (without being detected).

Entropy Combination: The operation that combines a number of entropy shares into one. The combination is usually achieved by using the

exclusive-or (XOR) operator. Entropy combination is an effective defense against adversaries that substitute a random value by a predictable one. The combined entropy is often of higher quality than each individual share.

Entropy Extraction: The operation that extracts entropy from a random variable (referred to as the entropy source). Entropy can be extracted using both software and hardware based methods.

Entropy: A metric used to evaluate and describe the amount of randomness associated with a random variable.

Pseudorandom Number Generator (PRNG): A PRNG is a computer algorithm used to produce a sequence of pseudo-random numbers. It must be initialized by a seed number and can be re-seeded prior to each run. The numbers produced by a PRNG are not truly random. Given the same seed, a PRNG will generate the same sequence of numbers.

Threshold Digital Signature: In the (k, n) threshold digital signature scheme, a private key is divided into n shares, each owned by a player. A valid threshold digital signature can be produced if k players combine their shares. However, no valid signature can be generated by fewer than k players. Each player uses its private key share to generate a partial signature on a message and these partial signatures can be combined into a threshold signature on the message. The threshold signature can be verified using the public key corresponding to the divided private key.

Piracy and Intellectual Property Theft in the Internet Era

Shun-Yung Kevin Wang

University of South Florida – St. Petersburg, USA

Jeremy J McDaniel

Principal Financial Group, USA

INTRODUCTION

Internet has quickly become an essential part of contemporary society across country borders for its capacity to offer a wide array of functions, ranging from information distribution, communications, financial and business management, to entertainments. Also, the Internet has evidenced itself as a unique medium with the fastest speed of diffusion in human history. With hundreds of thousand miles of optical fiber that connect servers and mega-storing devices together globally, several terabytes of digital information, as huge as those stored in the U.S. Congress Library, can be easily transferred from one end of the world to the other within minutes (Britz, 2013). In conjunction with widely available Wi-Fi and telecommunication (e.g., 3G, 4G, LTE) in many areas of the world, it is never this easy for an average user to transmit valuable information in digital format via mobile devices.

The information technology advances with incremental innovation, but business is the instrument that facilitates the widespread of the technology. The mechanism of business determines when to release certain technology, and the nature of business makes it user friendly for the purpose of obtaining a larger market share and a higher level of profit (Felson and Clarke, 1997). While legitimate opportunities are created in the process, some offenders may take advantage. Like many innovations that have a tendency to crime (Merton, 1968), the growing capacity of Internet probably is too good to be true, as it has created

new forms of intellectual property (IP). Before further discussing IP and elaborating the victimization of piracy, background of some theoretical frameworks of crime is necessary.

BACKGROUND

Basic Elements of Crime and Socio-Technical Gap

In their theory of crime, Cohen and Felson (1979) point out three elements of a crime incident: a suitable target, a motivated offender, and the absence of capable guardians. A suitable target is something valuable to potential offenders, and the target must be easy enough to be removed. Although crime rate is the highest among young males, motivated offenders can be anybody in the population, if an adequate opportunity is present. The guardians against crime do not necessarily refer to law enforcement. Instead, the owner of the targeted property, friends and neighbors of the property owners serve better roles of capable guardians that discourage potential offenders. In the scenario of burglary, potential perpetrators probably would less likely to choose houses that the owners are present or their friends/neighbors pay attention to. In the business settings, for another example, an office suite's receptionists who watch people entering the office can serve as the role of guardian. In sum, for a crime to occur, the above three elements have to emerge.

There is little doubt that industry has incentives to make their products lighter, more portable, more convenient, and more added functions and values, but this tendency naturally leads to some unwanted consequences of the products, such as suitable targets to theft. However, the social system (e.g., laws, justice agencies) usually simply reacts to the consequences of technological advancements pushed by industry and business. That is, technology proactively runs at the front, and the social system passively chases behind and (hopefully) fixes problems and challenges. In the era of Internet, the discrepancy between fast-growing Internet and information technology and the slow-reacting social system in the virtual space has created a cybergap in which crimes emerge (Huang and Wang, 2009). Explicitly, many more new digital IP are valuable targets with little to no meaningful guardians that trigger motivation of potential offenders in the cyberspace. The following section provides a description of IP theft and piracy. The discussion of IP and piracy in the present article is focused within the arena of those using digital technology, with an intention to compare and contrast several major incidents.

IP, IP Theft, and Piracy

The discussion of IP traditionally revolves copyrights, patents, trademarks, and trade secrets. Piracy has been generally defined as “the unauthorized use or reproduction of another’s work,” and it encompasses any individual or corporation that utilizes intellectual property in a digital form without the authorization of the originator (Business Software Alliance, n.d.; Filby, 2007). The nature of such behaviors is perceived as illegitimate, with some noticeable variation across different levels of civilization and cultures. For example, in some Asian societies with long histories, scholarly works are traditionally viewed as public goods contributing to the advancement of the entire society, and the scholars are informally “rewarded” with socially-recognized reputations and their social status. On the other hand, in the

United States and many European countries, where the right of tangible or intangible personal property are better defined and protected by laws, such kind of theft has been criminalized. Generally speaking, intellectual properties are well respected, formally and informally, in civilized societies.

Properties can be generally divided into tangible and intangible items, and the age-old theft usually involves tangible goods that perpetrators have to physically move away and turn into financial gains. IP theft is different from stealing of physical property in many ways: IP theft implies depriving people of their ideas, inventions, or creative expressions, and thereby this type of asset is intangible. Nevertheless, it is not saying that there is no overlapping between tangible and intangible properties, as IP also requires some kind of medium to load on, to store, or to distribute. For instance, the physical piracy of music – the production and/or distribution of illegally made copies of sound recordings without the consent of the rights proprietor – needs cassettes, discs, USB, hard drives, or other storing media. Within the past two decades, the significant improvement of personal computing devices equipped with large storage capacity inflamed the popularity of digital IP. In addition, in conjunction with growing Internet users, the expanding capacity of broadband and wireless technology increases the movement of digitalization. Collectively, the advancement of information technology has dramatically increased the amount of IP in digital format.

Based on the contents, there are two broad groups of highlighted IP: entertainment (music¹, movies, games, TV programming, etc.) and instrumental software. This typology is adopted for its exhaustive categories that ease the discussion, although ‘Internet piracy’ and ‘digital piracy’ are the terms used more often in varied news media and public reports. Internet piracy is a somehow broadly used term which generally means that the Internet is employed to distribute unauthorized creative content amongst users², and this term is used to generalize any use of creative content on the Internet that infringe on copyright laws

(Higgins, 2011). It had been a common myth to some that customers who purchase a legal copy of a music CD or USB believe that they own the music because they bought it in the format of optical disc or portable storage. In fact, the ownership of the recorded music, as a form of IP protected by the copyright laws, belongs to the writers. Consequently, if the purchasers massively multiply the disc/USB or upload the content to computers and share with others via the Internet, they would be committing so-called 'Internet piracy.' In a similar vein, piracy offenders may use the Internet for advertising, offering, acquiring, or distributing of other copyright protected contents.

Naturally, IP theft and piracy are business-threatening issues to the corporations of the ownership, and they can be problems to individuals as well. Intuitively, the pirated copies, which may appear to be legitimate ones, hurt the profit of the producing industry. Counterfeits sound recordings are produced without the required permission of the proprietor and then packaged to bear a resemblance to the original. Another noticeable type of piracy, bootlegs, consist of recording live or broadcast concerts without the consent of the proprietor that are replicated and re-sold. Similar violation may include other forms of performance and artworks. These actions are taken to mislead customers into believing that they are buying an original (and legal) version and supporting the creations.

The other substantial IP subject to today's piracy is instrumental software. Software piracy is commissioned by unlawfully multiplying or distributing of copyrighted software with the intention to gain financially. Legally purchased software (only) grants buyers the licenses of usage. The consensus, as it is reflected in the laws in many developed countries, is that buyers pay for user licenses, not the ownership of the IP of software. Based on the purpose and volume of duplication, two categories of software piracy are recognized: end user piracy and reseller piracy. End user piracy is the installation of the software onto computers or terminals exceeds the user license

allows. Reseller piracy occurs when an individual or an organization consciously produces multiple copies of the software for the purpose of selling for profits. Sometimes, pirate copies even come with counterfeit certificates with an obvious intent to deceive purchasers. In sum, copying, downloading, installing, sharing, or selling multiple copies of copyrighted software are behaviors subject to software piracy.

Corporations have adopted a good number of strategies to fight against software piracy. The most straight-forward method is to establish a call-free reporting hotline by the corporations. The operating system giant Microsoft, for example, utilized its anti-piracy hotline to collect information and sued 8 Toronto-area computer resellers suspected of unlawfully dispensing unlicensed software. The lawsuit was the single largest piracy sweep operation in a North American city of the time (National Post, 1999). A related strategy, similar to criminal case reporting system or campaign (e.g., "see something, say something"), is collaboration among stakeholders and offer meaningful incentives to encourage reporting. Business Software Alliance (BSA) promoted a program named "know it, report it, reward it" that provides qualified individuals who report software piracy a cash reward of up to \$1 million. This program led to a lawsuit that BSA filed against a Nevada based engineering company that duplicated and distributed unauthorized copies of Autodesk software used for drafting (PR Newswire, 2009). Another developed approach is working with direct-related industries or holding them legally accountable, such as Internet platform providers. In 2008, the Software and Information Industry Association (SIIA) filed nine separate lawsuits on behalf of members of Adobe Systems and Symantec alleging that the defendants deliberately sold illicit copies of software on eBay (The Daily Record of Baltimore, MD., 2008). This largest round of lawsuits, since eBay initiated its auction website anti-piracy program, signaled the seriousness of the issue and also sent out a clear message to similar industries. Still another approach is called digital

rights management (DRM) that places control over digital products. The methods of managing the digital products can be as simple as entering serial numbers to activate installed software, and the methods can be as sophisticated as requiring authentication with an online server and/or with a smart-card (Kumari, Khan, & Li, 2016). This approach has been adopted to ensure authorized access, distribution, and consumption of the products/services by software and entertainment companies. The implications of this approach can cover the entire lifecycle of the digital products.

Amplified by the global economy, the matter of IP theft and piracy can be much more complicated and worse when the majority of IP thefts initiated outside of the property owner's country territory where the right of IP are usually better protected by laws. Traditionally, stealing of ideas emerged in the developing countries where the criminal and civil laws are permissive and the enforcement of law is more likely to be compromised for economic reasons. However, Sweden has ironically been a piracy haven because of the well-known file-sharing website Pirate Bay is hosted within its territory. Although the Swedish court system convicted four major players of the website and punished them with prison sentences and a substantial amount of fines in 2009, the website is astoundingly running (Pfanner, 2009)³. Another extreme example is the entire counterfeit Apple store, in which the staffs even believe that they are Apple employees, found in Kunming, China (BBC, 2011), and such kind of "fake" Apple store did not seem to exist solely (Chang, 2015). It appears that some well-off Chinese appreciate real Apple products by paying double amount of the money, but they do not seem to care the sources of IP products. At the same time, surveys show that the overall rate of IP infringement in mainland China is close to the global average (Broadhurst, Bouhours, & Bouhours, 2013).

Proactive approaches to combating piracy have been spread both domestically and internationally. For example, in 2006, the United States government created an organization called Strategy

Targeting Organized Piracy (STOP) in an effort to deal with piracy. Following suit was the European Union that has created a different strategy based off of the STOP ideology. The European Union began working with some developing countries through effective implementation and enforcement of existing laws of IP. This proactive approach by both the United States and the European Union has in turn influenced the Nigerian government, for example, to take similar actions. The Nigerian government worked in partnership with the private sector and other stakeholders to set up the Strategic Action Against Piracy (STRAP), and the Nigerian Copyright Commission (NCC) worked with Microsoft to combat piracy (Africa News, 2006).

VICTIMIZATION OF PIRACY AND IP THEFT AND PROTECTIVE ACTIONS

The victimization of piracy and IP theft usually involves multiple parties in different context. The advancing information technology makes it easy for pirating websites to sell illegally duplicated IP, or even portray themselves as a legitimate business to deceive inexperienced online consumers⁴. In another scenario, many people who acquire pirated software over the Internet never actually receive the software or entertainment items they paid for. In that sense, those who consciously purchase pirated IP online potentially play dual roles: law violators and victims.

There used to be a clear-cut distinction between the victim(s) and the offender(s) when a criminal act takes place. We have little to no doubt that a street thief who snatches a purse is the obvious criminal offender, and the victim is the individual whose purse was stolen. As this series of behavior occurs in the physical world, it could be witnessed by others or video recorded by devices that leave evidence for criminal investigations. However, the Internet has changed the rule, and thereby somehow changes the distinction between victims and offenders. With the advancement of information

technologies, the physicality of traditional crimes has become intertwined with the creation of new types of criminal acts in the cyberspace, which are much less visible. Internet also substantially lowers the threshold of committing crimes, especially true in the violation of IP – clicking the mouse or touching on the screen and downloading a pirated copy of software, film or music from a remote server (wherever it is physically located in the world) can be a violation of criminal codes! Furthermore, the sense of borderless of many cybercrimes complicates the issue, as multiple countries, with a general disagreement about the seriousness of IP violations, are typically involved (Altbach, 1988). Collectively, it has become more challenging to define a victim in an intellectual theft incident. This has also created burdens on policy makers and police when defining and enforcing new criminal and civil statutes intended to protect truly innocent victims.

In the short history of digital piracy, the creation of Napster and its impact on music industry is one of the important milestones. In 1999, Shawn Fanning, a college student at the University of Boston, created an online music peer-to-peer file sharing service called Napster. This ground-shaking service allowed some music enthusiasts around the world to share music with each other without purchasing music CDs or tapes from physical stores. From many aspects, Napster could be a successful business, as it rapidly became the largest platform of music sharing of the time. Rather, Napster has often been blamed from the economic perspective – its existence leads to a major revenue loss of recording industry. Largely because of the erosion of industry's profit, the Recording Industry Association of America (RIAA) sued Napster on behalf of the major record labels and proprietors (musicians) that are the victims of music piracy (Richtel, 2001). RIAA won the law suit, however, the winning lawsuit that shutdown Napster did not lead to the intended consequences – regain billions of dollars in potential revenue. In contrast, the result in Napster's disappearance created a number of unregulated peer-to-peer file sharing

networks such as BearShare, Grokster, Gnutella, KaZaa, Morpheus, Pirate Bay, and others (Dodge, 2005). Even in a report, IFPI (2010) claims that the revenues from recorded music reduced by 50% in a decade after the creation of Napster. In a sense, the RIAA probably ironically retain the financial damages to its party by shutting down Napster instead of working with this emerged online platform, which once had 38 million users, to rectify their strategic flaws of business. The law case symbolizes the significance of copyrighted content protection, but the music industry is still somehow suffering from this lawsuit. The advancement of related information technology has made it difficult for corporations to protect themselves from being victimized, if the fundamental changes that can be made by new technologies are not carefully appraised. It is hard to deny that “as long as Philips makes and sells CD burners, and as long as Sony makes and sells MiniDisc players/recorders, they're directly profiting from the very technology that they claim is hurting them” (Rodman & Vanderdonckt, 2006: 253). In other words, the increase of digital piracy of music and the decline of music industry revenues are largely the consequences of information technology advancement.

Piracy of music has existed for decades, but digital piracy of music significantly hit the industry because of its slower, if not unwise, response to the technologies, in addition to passive response from the social system (Huang & Wang, 2009). An interesting fact is that the successful business model of selling music online was accomplished by a computer company, not the music industry – Apple introduced iTunes online music store after Napster lawsuit, and it remained its leading place in the market when this manuscript was prepared in 2016.

In addition to corporations being impacted by the emerged cyber-platforms, individual users are subjects to corporations' legal actions. In 2003, the music recording industry sued over 200 Americans for sharing songs on peer-to-peer file sharing networks, and the message cannot be

clearer when a minor living under poverty was on the list. Moreover, the recording industry has filed, settled, or threatened legal actions against at least thirty thousand individuals in order to protect its potential profits. It probably is socially justified to consider these individual offenders in that they illegally shared intellectual property, regardless the media that are used. However, the proportionality of penalties (e.g., charges and retributions that these individuals have to pay) and damages might be debatable. To some, this case makes sued individuals become scapegoat when the balance and rules are greatly impacted by the introduced new applied technology. For instance, Jammie Thomas of Brainerd, MN was accused of unlawfully sharing over 1,500 songs, and the RIAA lawsuit sought damages under a federal law that allowed a fine of \$750 to \$30,000 for each copyright violation, which totaled a judgment of over \$1.2 million (Freed, 2007). To others, “it’s stupid (and reckless) to antagonize one of your principal sources of long-term income by publicly attacking them as criminals” (Rodman & Vanderdonckt, 2006: 259).

Similarly, the widespread of BitTorrent and illegal download of movie contents had significantly impacted revenues of film makers. Another important milestone of combating piracy is focusing on sites that offer free massive online storage space, which were misused by some to store and share contents of large files, such as movies. In a recent study that assesses the economic impacts of supply-side intervention of illegal downloading indicates that the shutdown of Megaupload⁵ chronologically leads to a significant increase in digital sales and rentals. Countries with higher pre-shutdown usage of Megaupload experience larger increase of digital sales than their counterparts with lower pre-shutdown usage do (Danaher & Smith, 2013).

IP theft and piracy will undoubtedly continue to occur, but the approaches of solutions can be different. In the case of film and music piracy, it will be up to stakeholders, including policy

makers, legal experts, and the industry to come out a balanced approach. More efforts should be taken placed in innovative ways to better educate consumers about sharing movies and music and to reduce costly legal actions. A more recent case of a Vietnam-based social media website “Zing,” which includes an infringing deeplinking music portal, shows that collaborations, instead of lawsuits, may have been identified as a win-win solution between right holders and online platforms (Executive Office of the President of the United States, 2012).

FUTURE RESEARCH DIRECTIONS

From the preceding sections, the trend of working with continuously-advancing information technology and online platforms instead of fighting against it appears to be a more effective approach in IP protection with more beneficial results to stakeholders. Although the Internet is borderless, the undeniable “territory” in cyberspace is largely structured by users and administrations’ nationality, culture, ideology, and language, which is somehow related to their geographical locations. Future research can systematically compare cases in the area of IP protection strategies and evaluate their outcomes across different geographical regions and industries. Instead of narrowing down to the benefits of certain stakeholders like music industry, the outcomes should include social and economic indicators that capture the impacts on the development of legal system and competitive advantage of the economy. Future research probably should investigate the impacts of cultures, including pop cultures, on how people perceive and support IP rights. Given some contradicting phenomenon between international survey findings and news reports, using dimensions like the socioeconomic development, level of civilization, and law protections seem to overly simplify global citizens by their nationality.

CONCLUSION

The main challenge of piracy and IP violations that today's societies face is the discrepancy between the fast-pace information technology, which is believed by some to lead advancement of human societies, and the slow reactions of laws, which is supposed to guide and protect citizens. The rapid progression of technology, combined with weak laws, has allowed criminals to exploit the international community and run illegal "business" of piracy. As a result, the entertainment and software industries, in particular, have been victimized of extensive pirating schemes, which have caused substantial revenue and job loss worldwide. Information technology has greatly expanded the avenues that copyright-protected works can be duplicated and distributed, while lowering the costs. After the mass investment of optical fiber in late 1990s, the Internet has facilitated the transfer of vast amounts of information among large numbers of people at little cost to law violators. But, the cost of piracy, ranging from financial, economic, safety, to health, is tremendous to the IP owners and the entire society, as the pirates require no investment in research, development, and warranties, and consumers are risked by failing counterfeiting parts and products.

The uniqueness of Internet's fastest diffusion, along with the reactive nature of social system, has created a good number of challenging social issues in the socio-technical gap. From a criminological perspective like routine activities, this information super highway does provide very fast transportation of attractive targets with some to not-meaningful guardians in the cyberspace to motivated (and somewhat skilled) offenders. The Internet and related technology might have also blurred the clear line of ownership of productive assets, as well as the products and services originate from them. These clear lines are central to the foundation of capitalism; the social system cannot continue to function without a set of understandable rules that define ownership of the properties and legitimate business models (Thurow, 2000).

Rather than making progress, the system would fall apart starting from the erosion of the incentive structure that encourages the continuity of research and development, which leads to further loss of trust in business and confidence in the economy (van Dijk and Terlouw, 1996). Indeed, the Internet probably has brought another wave of significant challenges to the issues of intellectual property and piracy, since Xerox introduced its first commercial photocopy machine in 1959.

The advancement of information technology has also intentionally increased digitalized IP and "migrate" those in earlier generation formats to the cyberspace. The Internet search giant Google, for example, has implemented many ambitious projects to massively digitalize social artifacts. There is no doubt that information is so affordable to so many people in the human history because of the Internet, and Google, so far, has been the leading private provider that effectively organizes and prioritizes the online contents to the majority users of the developed countries. While some may argue against Google's project of "displaying" instead of "copying" the IP in the cyberspace, another wave of concern of IP and piracy may involve the conflicts between the profit-making nature of business and the belief of public-belonging of human heritage (Vaidhyanathan, 2011).

REFERENCES

- Africa News. (2006, December 6). *Nigeria: Tackling the problem of piracy in Nigeria*. Retrieved from www.lexisnexis.com/hottopics/lnacademic
- Altbach, P. G. (1988). Economic progress brings copyright to Asia. *Far Eastern Economic Review*, 139(9), 62–63.
- BBC. (2011). *Fake Apple store found in China*. Retrieved July 11, 2016, from <http://www.bbc.co.uk/news/technology-14258135>
- Britz, M. (2013). *Computer Forensics and Cyber Crimes: An Introduction* (3rd ed.). Upper Saddle River, NJ: Pearson Education Inc.

- Broadhurst, R., Bouhours, B., & Bouhours, T. (2013). Business and the risk of crime in China. *The British Journal of Criminology*, 53(2), 276–296. doi:10.1093/bjc/azs059
- Business Software Alliance. (n.d.). *What is software piracy*. Retrieved July 10, 2016, from <http://www.bsa.org/country/Anti-Piracy/What-is-Software-Piracy.aspx>
- Chang, C. (2015). *The Most Phenomenal Fake Apple Store in China*. Retrieved July 20, 2016, from <http://micgadget.com/14648/the-most-phenomenal-fake-apple-store-in-china/>
- Cohen, L. E., & Felson, M. (1979). Social change and crime rate trends: Routine activity approach. *American Sociological Review*, 44(4), 588–608. doi:10.2307/2094589
- Danaher, B., & Smith, M. (2013). *Gone in 60 Seconds: The Impact of the Megaupload Shutdown on Movie Sales*. Available at SSRN # 2229349.
- Dodge, D. (2005). *Perspective: Napster's learning curve*. Retrieved February 17, 2014, from http://news.cnet.com/Napsters-learning-curve/2010-1027_3-5901873.html
- Executive Office of the President of the United States. (2012, December 13). *Out-of-Cycle Review of Notorious Markets*. Author.
- Felson, M., & Clarke, R. V. (1997). *Business and Crime Prevention*. Monsey, NY: Criminal Justice Press.
- Filby, M. (2007). Confusing the captain with the cabin boy: the dangers posed to reform of cyber piracy regulation by the misrepresented interface between society, policy makers and the entertainment industries. In *Proceedings of the British & Irish law, education and technology association conference* (pp. 1-47). Stanford, CA: Creative Commons.
- Freed, J. (2007). *Brainerd woman first to go to trial for music sharing: Jammie Thomas at risk for a judgment of millions of dollars*. Retrieved February 17, 2014, from http://brainerddispatch.com/stories/100207/new_20071002025.shtml
- Higgins, G. E. (2011). *Digital Piracy: An Integrated Theoretical Approach*. Durham, NC: Carolina Academic Press.
- Huang, W., & Wang, S. K. (2009). Emerging Cybercrime Variants in the Socio Technical Space. In B. Whitworth, & A. de Moor (Eds.), *Handbook of Research on Socio-Technical Design and Social Networking Systems*. Hershey, PA: IGI.
- IFPI. (2010, July). *IFPI Response to Commission Green Paper on Creative and Cultural Industries*. IFPI.
- Kumari, S., Khan, M. K., & Li, X. (2016). A more secure digital rights management authentication scheme based on smart card. *Multimedia Tools and Applications*, 75(2), 1135–1158. doi:10.1007/s11042-014-2361-z
- Merton, R. K. (1968). *Social Theory and Social Structure*. Glencoe, IL: The Free Press.
- National Post. (1999, August 13). *Eight computer resellers face lawsuits after latest Microsoft anti-piracy sweep: Computers allegedly loaded with unlicensed software*. Retrieved from www.lexisnexis.com/hottopics/lnacademic
- Newswire, P. R. (2009, March 31). *BSA settles lawsuit against Nevada-based engineering company for \$205,000*. Retrieved from www.lexisnexis.com/hottopics/lnacademic
- Pfanner, E. (2009, April 18). Four convicted in Sweden in Internet piracy case. *The New York Times*, p. 2.
- Richtel, M. (2001, February 13). The Napster decision: The overview; appellate judges back limitations on copying music. *The New York Times*, p. 1.

The Daily Record of Baltimore. (2008). *News Summary - 3/21*. Retrieved from www.lexisnexis.com/hottopics/lnacademic

Thurow, L. C. (2000). The product of a knowledge-based economy. *The Annals of the American Academy of Political and Social Science*, 570(1), 19–31. doi:10.1177/0002716200570001002

Vaidhyanathan, S. (2011). *The Googolization of Everything: (And Why We Should Worry)*. Berkeley, CA: University of California Press.

van Dijk, J., & Terlouw, G. (1996). An international perspective of the business community as victims of fraud and crime. *Security Journal*, 7, 157–167. doi:10.1016/0955-1662(96)00168-3

ADDITIONAL READING

Allum, F., & Siebert, R. (Eds.). (2003). *Organized crime: The challenge to democracy*. London: Routledge.

Beresford, A., Desilets, C., Haantz, S., Kane, J., & Wall, A. (2005). Intellectual property and white-collar crime: Report of issues, trends, and problems for future research. *Trends in Organized Crime*, 8(4), 62–78. doi:10.1007/s12117-005-1014-z

Bhattacharjee, R., Gopal, R. D., Lerwachara, K., & Marsden, J. R. (2006). Impact of legal threats on online music sharing activity: An analysis of music industry legal actions. *The Journal of Law & Economics*, 49(1), 91–114. doi:10.1086/501085

Branstetter, L. G., Fisman, R., & Foley, C. F. (2006). Do stronger intellectual property rights increase international technology transfer? Empirical evidence from U.S. firm-level panel data. *The Quarterly Journal of Economics*, 121(1), 321–349. doi:10.1093/qje/121.1.321

Broadhurst, R. G. (2005). International Cooperation in Cyber-crime Research. In *Proceedings 11th UN Congress on Crime Prevention and Criminal Justice, Workshop 6: Measures to Combat Computer Related Crime*, 1–12, Bangkok, Thailand.

Chen, Y., & Puttitanum, T. (2005). Intellectual property rights and innovation in developing countries. *Journal of Development Economics*, 78(2), 474–493. doi:10.1016/j.jdevec.2004.11.005

Choate, P. (2005). *Hot Properties: The Stealing of Ideas in an Age of Globalization*. New York, NY: Knopf: Distributed by Random House.

Drahos, P. (2002). Developing countries and international intellectual property standard-setting. *The Journal of World Intellectual Property*, 5(5), 765–789. doi:10.1111/j.1747-1796.2002.tb00181.x

Eaton, J., & Kortum, S. (1996). Trade in ideas: Patenting and productivity in the OECD. *Journal of International Economics*, 40(3–4), 251–278. doi:10.1016/0022-1996(95)01407-1

Friedman, T. L. (2005). *The world is flat: A brief history of the twenty-first century*. New York: Farrar, Straus, and Giroux.

Grossman, G. M., & Lai, E. L. C. (2005). International protection of intellectual property. *The American Economic Review*, 95(5), 1635–1653.

International Federation of the Phonographic Industry (IFPI). (2003). *The record industry commercial piracy report, 2003*. London: IFPI.

Keller, W. (2004). International technology diffusion. *Journal of Economic Literature*, 42(3), 752–782. doi:10.1257/0022051042177685

Lai, E. L. C. (1998). International intellectual property rights protection and the rate of product innovation. *Journal of Development Economics*, 55(1), 133–153. doi:10.1016/S0304-3878(97)00059-X

Maher, M. K., & Thompson, J. M. (2002). Intellectual property crimes. *The American Criminal Law Review*, 39(2), 763–816.

Motivans, M. (2004). *Intellectual Property Theft, 2002*. Bureau of Justice Statistics, U.S. Department of Justice. doi:10.1037/e306472005-001

Nhan, J. (2009). “It’s like printing money”: Piracy on the Internet. In F. Schmalleger & M. Pittaro (Eds.), *Crimes of the Internet* (pp. 356–383). Upper Saddle River, NJ: Pearson Education Inc.

Piquero, N. L., & Piquero, A. R. (2006). Democracy and intellectual property: Examining trajectories of software piracy. *The Annals of the American Academy of Political and Social Science*, 605(1), 104–127. doi:10.1177/0002716206287015

Ponte, L. M. (2009). The warez scene: Digital piracy in the online world. In F. Schmalleger & M. Pittaro (Eds.), *Crimes of the Internet* (pp. 384–407). Upper Saddle River, NJ: Pearson Education Inc.

Rodman, G. B., & Vanderdonckt, C. (2006). Music for nothing or, I want my MP3. *Cultural Studies*, 20(2–3), 245–261. doi:10.1080/09502380500495734

Ryan, M. P. (1998). *Knowledge Diplomacy: Global Competition and the Politics of Intellectual Property*. Washington, D.C.: Brookings Institution Press.

Singer, P. W., & Friedman, A. (2013). *Cybersecurity and Cyberwar: What Everyone Needs to Know*. Oxford: Oxford University Press.

Wall, D. S. (2007). *Cybercrime: The Transformation of Technology in the Networked Age*. Cambridge: Polity Press.

World Intellectual Property Organization (WIPO). (2001). *WIPO intellectual property handbook: Policy, law, and use*. WIPO Publication no. 489(E). Geneva, Switzerland: WIPO.

KEY TERMS AND DEFINITIONS

End User Piracy: The installation of the software onto computers or terminals exceeds the user license allows.

Intellectual Property (IP): Copyrights, patents, trademarks, and trade secrets. Copyrights protect literary or artistic creation, ranging from books, articles, scripts, and audio and video recordings. Patents protect registered and determined inventions or discoveries that exclude others from making, using, offering for sale, or selling, and the examples include the composition of a new medication. Trademarks protect a companies’ name, products’ name, logos, slogans, or package designs. Trade secrets secure formulas, production procedures, and even business lists, and Coca Cola’s secret recipe is an example.

Internet Piracy: Piracy that is conducted or facilitated by the Internet where unauthorized creative contents are distributed.

IP Theft: Stealing law-protected intellectual property that reflects creators’ ideas, inventions, or creative expressions, in tangible or intangible format.

Piracy: The unauthorized use or reproduction of works that belong to another individual(s) or entity for the purpose of financial gain.

Reseller Piracy: When an individual or organization consciously produces multiple copies of the software with the intent of selling for profits.

Software Piracy: Unauthorized duplicating or distributing of copyrighted software with the intention to gain financially.

ENDNOTES

- ¹ According to the International Federation of the Phonographic Industry, there are four kinds of illegal activities that fall under the music piracy umbrella: physical music piracy, counterfeit, bootlegs, and Internet piracy.

² Like many other oft-used terms describing Internet-related crimes, such a term that includes the medium/instrument (Internet) carrying IP in its name usually distracts the discussions and thereby confuses people.

³ One of the four founders was later arrested for the conviction in 2012 in Cambodia.

⁴ According to the Business Software Alliance, a software industry trade association, there

are roughly 840,000 Internet sites partaking in the illegal selling of software as the real thing.

⁵ Megaupload at one point was the 13th most visit site online, according to Alexa.com. It was accounted for 4% of worldwide Internet traffic.

Secure Group Key Sharing Protocols and Cloud System



Vaishali Ravindra Thakare
VIT University, India

John Singh K
VIT University, India

INTRODUCTION

Secure group communication is an important research issue in the field of cryptography and network security, because group applications like online chatting programs, video conferencing, distributed database, online games etc. is expanding rapidly. Group key agreement protocols allow that all the members agree on the same group key, for secure group communication, and the basic security criteria must be hold. Many group key agreement protocols have been established for secure group communication.

Since the group generation processes takes many modular exponentiations and long time in generation of group key. For achieving higher security, group key protocol should be dynamic, means it should change for each new join or leave member, so that new member have not any knowledge about prior information. Therefore group key management protocol focusing on the group key generation efficiently. The authors have identified the research gaps in the existing protocols and these are communication, computation overhead while generating and sharing digital envelopes and security issues while sharing group key with encryption algorithms. These research problems in existing framework motivate authors to focus on security and efficiency of the system. Many practical systems have been proposed (Liu et al., 2014a, 2015; Pan et al., 2011; Sanchez-Artigas,

2013; Li et al., 2015) of which the most familiar one is the TGDH key distribution system. After analyzing the demand for sharing data with multiple users in groups by reducing computational complexity and for achieving productive benefits, an efficient solution is proposed in this paper.

Modular exponentiation is very expensive in computation of group key. The number of exponentiations for membership depends on group size as when the group size increased the number of exponents will also increase. Tree Based Group Diffie Hellman (TGDH) (Kim et. al., 2004) uses the concept of Diffie-Hellman key exchange with logical tree structure to achieve efficiency. The efficiency of TGDH is $O(\log 2n)$, where n is the group size. However, some extra overhead occurred in maintaining a perfect key tree balance. Skinny tree has lower communication overhead, but it increases computation. Burmester–Desmedt (BD) distributes and minimizes computation by using more messages broadcast. All these protocols using similar security properties including group key independence. From the broad study it is found that, tree-based CGKA (Contributory Group Key Agreement) methods are more efficient since they reduce the complexity from $O(n)$ to $O(\log n)$ while computing the new group key, where n is the group size. Consequently, this unit considers only the existing tree-based CGKA protocols.

BACKGROUND

Group key agreement protocols allow that all the members agree on the same group key, for secure group communication, and the basic security criteria must be hold. In 1994 Mike Burmester and Yvo Desmedt Proposed A Secure and Efficient Conference Key Distribution System (BD Protocol), In 2000 Group Diffie Hellman (GDH) was proposed by Steiner et.al., Skinny Tree (STR) Wong et al. 2000, ID-AGKA (Identity based authenticated group key agreement protocol) by K C Reddy and Divya Nalla in 2002, Kim et al. proposed TGDH (Tree Based Group Diffie Hellman) in 2004, In 2006 CCEGK was proposed by Szheng, Moreover in 2009 QGDH (Queue Based Group Diffie Hellman) by Hong S.

After understanding the real time issues in real time groupware applications like voice & video conferences, distributed computation over the insecure network.

(Zheng et al., 2007) Proposed a two round key agreement protocol for dynamic peer group (DPG). The protocol is proven secure against passive attack by using indistinguishable method. Moreover, both perfect forward secrecy (PFS) and key independence (KI) were achieved. Author's proposed protocol greatly reduces the computation complexity of each member, definite identification and time stamp are added to the protocol to effectively avoid replay attacks and it satisfies PFS, dynamic and it provides a session key for wireless group members due to which its messages are transmitted through broadcasting. Meanwhile, authors proved correctness, tolerance for passive attacks, secure against active adversaries in the random oracle model as the security and efficiency analysis of this protocol.

(Liu et al., 2013) Proposed a secure multi-owner data sharing scheme (Mona) for dynamic groups in cloud applications. The Mona aims to realize that a user can securely share the data with others via the un-trusted cloud servers, and efficiently support dynamic group interactions. In this scheme, a new user can directly decrypt data files

without pre-contacting with data owners, and user revocation is achieved by revocation list without updating the secret keys of the remaining users.

(Xue & Hong., 2013) proposed a novel secure group sharing framework for public cloud and it can take the effective advantage of cloud help by taking care that no sensitive data should be exposed to cloud provider and an attacker. It combines proxy signature, enhanced TGDH-based binary tree, proxy re-encryption as a protocol, using which the authors have achieved the objective. In this scheme, authors used TGDH with binary tree to negotiate and update the group key pairs with the help of cloud servers

(Jaiswal & Tripathi., 2015) Proposed an alternative approach to group key agreement, i.e., a novel queue-based group key agreement protocol, which uses the concepts of elliptic curve cryptography to reduce unnecessary delays, considers member diversity with filtering out low performance members in group key generation processes. After analyzing many prior group key agreement protocols like TGDH, STR, BD, and QBDH, they provide better security. They take more computational overheads. So, authors have used elliptic curve cryptographic technique that removes exponentiation to reduce computational overheads, and hence the results are better than that of the other group key agreement protocols.

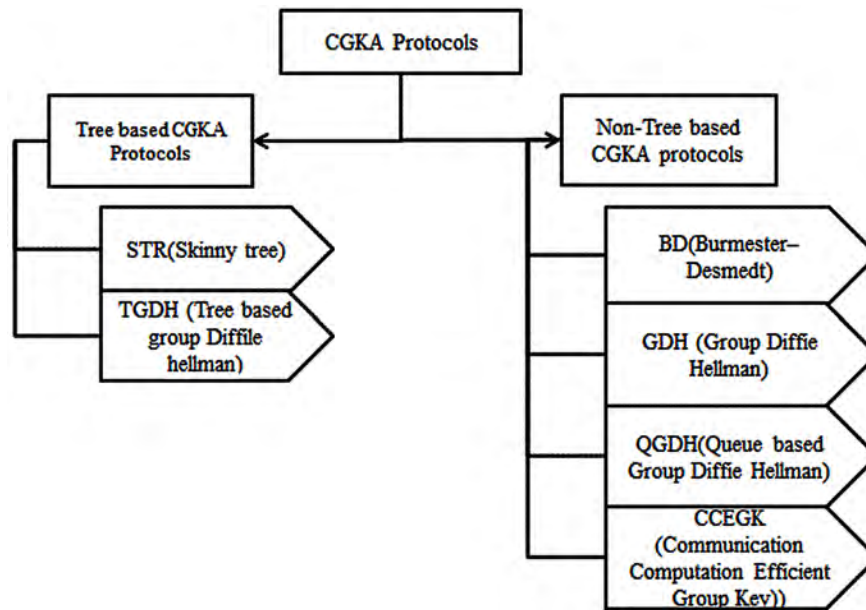
Figure 1 shows the tree structure of CGKA, the management of secure communication among groups of participants requires a set of secure and efficient operations some protocols are better in communication cost some are in computation of secure group key while some are having security issues. These all protocols fall under CGKA (Contributory Group Key Agreement) scheme. CGKA is further again Divided into two sub categories:

Tree Based-CGKA

STR (Skinny Tree)

The STR protocol is modified to provide dynamic group operations. The protocol has a relatively

Figure 1. Tree structure of CGKA (contributory group key agreement) protocols



low communication overhead and is well suited for adding new group members. Robustness is easily provided. However, it is relatively difficult for member exclusion ($O(n)$ modular exponentiation). The GK = $g^{K_n} g^{K_{n-1}} \dots g^{K_2} g^{K_1}$. The STR is communication efficient and it is more secure against attacks specific to group communication

TGDH (Tree Based Group Diffie Hellman)

The first tree based CGKA protocol called TGDH was proposed in which key tree and DH protocol are combined to generate the group key. After comparison of TGDH with STR and GDH protocol. The TGDH is comparatively efficient in join events, and it is best in leave events. The TGDH is more efficient in both computation and communication than GDH. The 'Skinny Tree' (STR) employs the key tree concept used in TGDH. TGDH provides the most efficient group key agreement protocol. The protocol has low communication overhead and low computation overhead ($O(\log n)$ modular exponentiation). In addition,

TGDH provides robustness. The GK = $g^{g^{K_1} K_2 g^{K_3} K_4 \dots g^{K_n} - 3 K_n - 2 g^{K_n} - 1 K_n}}$

In every operation of TGDH (e.g., Join, merge, leave, and partition), every message uses the broadcast scheme, so authors implement the initialization operation in TGDH using a broadcast scheme. TGDH key management protocol for dynamic secure group data sharing In ternary tree with TGDH all group members are initially arranged as the leaves of ternary tree and three party TGDH generates the final group key. This generation of group key is done in two stages:

1. Up-flow stage,
2. Down-flow stage.

Consider two prime numbers p and q such that $q|p-1$ and size of p and q are large enough so that solving the discrete logarithm problem in G is infeasible computational, where G is a subgroup with order q of a finite field Z_p^* [2] g is the generator. Consider, Alice, Bob and Charlie are three users that generates a shared secret key between them.

Up-flow:

- Alice constructs a set $\{g^x \bmod p\}$ and send it to Bob.
- Bob constructs a set $\{g^{xy} \bmod p, g^y \bmod p, g^x \bmod p \text{ (received from Alice)}\}$. Bob sends this set to Charlie- Similarly, Charlie also construct a set $\{g^{xyz} \bmod p, g^{yz} \bmod p, g^{xz} \bmod p\}$.
- Here, $g^{xyz} \bmod p$ is a cardinal value and treated as a group key K . Where, x , y and z are random numbers assumed by Alice, Bob and Charlie respectively.

Down-flow:

- Charlie broadcast the remaining set element ($g^{yz} \bmod p, g^{xz} \bmod p$) to other members in a group i.e. Alice and Bob.
- Now, every members assumes their random numbers for computing the same group key.
- (Alice $\Rightarrow K: (g^{yz})^x \bmod p$ and Bob $\Rightarrow K: (g^{xz})^y \bmod p$).

Non Tree Based-CGKA

BD (Burmester-Desmedt)

The BD protocol (Burmester and Desmedt, 1994) supports dynamic group operations. The protocol has a relatively low computational overhead due to two modular exponentiations. However, it needs more message exchanges to generate GK. The $GK = g^{K1K2+K2K3+...+Kn-1Kn}$.

GDH (Group Diffie-Hellman)

It provides high security assurance (Cui et. al., 2014). However, it has a relatively high computation cost ($O(n)$ modular exponentiation) and is relatively hard to provide robustness. The $GK = g^{K1K2K3K4...Kn-1Kn}$

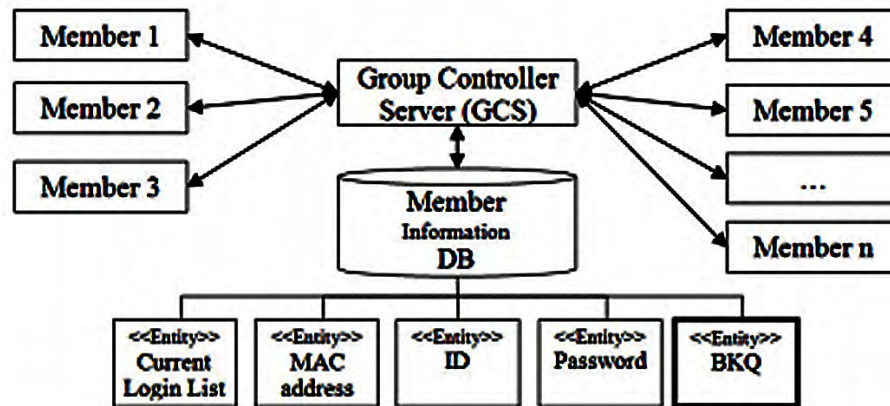
QGDH (Queue Based Group Diffie Hellman)

Figure 2 shows Queue-based Group Diffie-Hellman Entity Model (Hong S., 2009, Jaiswal & Tripathi, 2015). In Figure 2, a Group Controller Server (GCS) has a member information DB containing a current login member list, MAC addresses, IDs, Passwords, and Blind Key Queues (BKQ). If each member logs in to the GCS, then the GCS will validate his ID, password, and MAC (Media Access Control) address by checking his member information DB. After approving member's identification, all members start to generate a group key by sending his blind key to the GCS who collects all blind keys and store them into the Blind Key Queues (BKQ) in order of arrival. Then the GCS informs participants who will join the next level of the group key generation.

CCEGK (Communication-Computation Efficient Group Key Algorithm)

The CCEGK was developed for large and dynamic groups that combine two existing protocols such as EGK and TGDH (Lee et. al., 2003). It outperforms EGK, TGDH and STR protocols. CCEGK designed to provide both efficient communication and computation, addressing performance, security and authentication issues of CCEGK. An analytical comparison of all algorithms revealed eight similar methods: add, remove, merge, split, mass add, mass remove, initialize, and key refresh. Comparing the cost in terms of communication and computation, we found CCEGK to be more efficient across the board. With the advent of new arenas such as wireless ad-hoc and low powered distributed computing and communication devices, designers of group key encryption algorithms can no longer ignore communication in favor of computation or vice versa. In some environments the power cost of communication may be sufficiently high to warrant low cost communication protocols, whereas in other environments computation cost may be

Figure 2. Queue-based group diffie-hellman entity model



the dominant feature. CCEGK is a group key management algorithm based upon two preceding group key management algorithms, EGK and TGDH. By extending this previous work, CCEGK considerably improves both communication and computation costs of their related operations.

CCEGK, as do EGK and STR, always joins at the root of the tree, resulting in potentially far fewer sequential exponentiations. TGDH will join at the root only if the tree is a full tree, in an attempt to keep the tree more balanced. STR, on the other hand always has a skinny tree, where every internal node has one child that is a leaf.

CCEGK, as with EGK, merges the two roots of the trees, resulting in lower cost but a potentially more unbalanced tree. TGDH merges the shallower tree to the deeper one, or at the root, whichever creates the shallowest final tree. STR adds one tree to the bottom of the other.

SOLUTIONS AND COMPARITIVE ANALYSIS OF SECURE GROUP KEY SHARING PROTOCOLS

Group Key Management Operations

Throughout this chapter authors addressed several group key management operations used by

group key systems. The operations are described as following:

- **Initialization Operation:** This is the initial creation of the group key and organization of the key management infrastructure.
- **Join:** This operation brings a new member into the existing group.
- **Leave:** This operation is used to remove a member from the group.
- **Partition:** This operation, different from mass leave, occurs when a single group is divided into two or more component groups.
- **Merge:** This operation, as opposed to mass join, is used when another group is combined with the existing group to become a new group.
- **Key Refresh:** To prevent the secret key from becoming stale, it should be changed. Moreover, to prevent an adversary from breaking in, users should refresh the original key and generate a new secret key periodically.
- **Number of Rounds:** This is a generic time unit used to compare the number of steps taken in different operations. The algorithms often require synchronization between rounds; therefore, this number

Table 1. Communication and computation cost analysis of tree based-CGKA protocols

Tree Based CGKA Protocols		Communication Cost		Computation Cost
		Rounds	Messages	Exponentiation
STR	Join	2	3	4
	Leave	1	1	$\frac{3n}{2} + 2$
	Partition	1	1	$\frac{3n}{2} + 2$
	Merge	2	k+1	3m+1
TGDH	Join	2	3	$\frac{3n}{2}$
	Leave	1	1	$\frac{3n}{2}$
	Partition	$\text{Min}(\log_2 p, h)$	2h	3n
	Merge	$\text{Log}_2 k+1$	2k	$\frac{3n}{2}$

becomes important when taking synchronization time into account.

- **Number of Messages:** This is the sum of all unicast messages and broadcast messages. users use this number to determine the total time of communication in an underlying broadcast network.
- **Number of Exponentiations:** During an operation there will be a series of computationally expensive cryptographic operations (such as modular exponentiation used in the DH algorithm). The algorithms in the literature often require the results of one cryptographic operation prior to the execution of another. This metric represents the worst case scenario, the longest sequence of dependencies of these costly cryptographic operations in the operation.

The authors of this article have analyzed many existing group key agreement protocols like TGDH, STR, BD, QBDH and CCEGK.

They provide better security, but some take more computational overheads, some take more communication overhead. So, Following Table 1 shows communication and computation cost analysis of Tree Based-CGKA Protocols and Table 2 shows communication and computation cost analysis of Non Tree Based-CGKA Protocols.

FUTURE RESEARCH DIRECTIONS

Recently, group data sharing (GDS) has attracted significant attention for its wide applications in many different fields. A GDS can be seen as a special dynamic and distributed group, so the secure communication is essential in it. Surely, the most common method is to encrypt messages with a group key only shared by the included nodes, so that those outside the group cannot decode the encrypted messages. Thus, the protocol to achieve the group shared key is crucial, which users often name the *key agreement protocol*.

Table 2. Communication and computation cost analysis of non tree based-CGKA protocols

Non Tree Based CGKA Protocols		Communication Cost		Computation Cost
		Rounds	Messages	Exponentiation
BD	Join	2	$2n+2$	3
	Leave	2	$2n-2$	3
	Partition	2	$2n-2p$	3
	Merge	2	$2n+2m$	3
GDH	Join	4	$n+3$	$n+3$
	Leave	1	1	$n-1$
	Partition	1	1	$n-p$
	Merge	$m+3$	$n+2m+1$	$n+2m+1$
QDGH	Join	2	$2n-2$	$\frac{3(\log_2 n)}{2}$
	Leave	1	$2n-2$	$\frac{3(\log_2 n)}{2}$
	Partition	1	$2n-2$	$3(\log_2 n)$
	Merge	2	$2n-2$	$\frac{3(\log_2 n)}{2}$
CCEGK	Join	1	2	1
	Leave	1	1	$3h-3$
	Merge	1	N	$n-1$

Over the years, numerous excellent key agreement protocols for dynamic peer group have been proposed, such as the GDH, TGDH and extended STR protocols. However, not all of them are communication efficient when applied to cloud networks. All of these have better security built-ins. Now a day in every fields cloud computing is used, health care, IT industries, small and medium enterprises (SMEs). So, The scope of these protocols is everywhere widely growing because of the requirements of the changing world towards cloud computing. There will be various kinds of requirements of healthcare organizations, as healthcare organizations are the primary priority in every country, Healthcare clouds must be as

secure as possible. According to the organizations requirements healthcare cloud groups will be formed for the secure communication between other organizations in healthcare and these protocols will take care about security and confidential data, communication and computation overheads.

CONCLUSION

Data sharing in cloud environment by forming groups is considered as a common approach. But secure data sharing in groups is still a challenging issue for the researchers. There are many frameworks available to share data in groups in

the previous researches listed in above discussion of this chapter, after analyzing the existing frameworks drawbacks; it is found the tree based CGKA protocols are more efficient than others. From Table 2, TGDH is better protocol for secure group data sharing and it can be used in healthcare cloud applications. From Table 3, CCEGK protocol is better but it is non-tree CGKA protocol but its communication and computation overhead cost is lesser than the BD, GDH and QGDH protocols.

REFERENCES

- Burmester, M., & Desmedt, Y. (1994, May). A secure and efficient conference key distribution system. In *Workshop on the Theory and Application of Cryptographic Techniques* (pp. 275-286). Springer Berlin Heidelberg.
- Cui, Y., Peng, Z., Song, W., Li, X., Cheng, F., & Ding, L. (2014, September). A Time-Based Group Key Management Algorithm Based on Proxy Re-encryption for Cloud Storage. In *Asia-Pacific Web Conference* (pp. 117-128). Springer International Publishing. doi:10.1007/978-3-319-11116-2_11
- Hong, S. (2009). Queue-based Group Key Agreement Protocol. *International Journal of Network Security*, 9(2), 135–142.
- Jaiswal, P., Kumar, A., & Tripathi, S. (2015). Design of Queue-Based Group Key Agreement Protocol Using Elliptic Curve Cryptography. In *Information Systems Design and Intelligent Applications* (pp. 167–176). Springer India. doi:10.1007/978-81-322-2250-7_17
- Kim, Y., Perrig, A., & Tsudik, G. (2004). Tree-based group key agreement. *ACM Transactions on Information and System Security*, 7(1), 60–96. doi:10.1145/984334.984337
- Lee, S., Kim, Y., Kim, K., & Ryu, D. H. (2003, October). An efficient tree-based group key agreement using bilinear map. In *International Conference on Applied Cryptography and Network Security* (pp. 357-371). Springer Berlin Heidelberg. doi:10.1007/978-3-540-45203-4_28
- Liu, X., Zhang, Y., Wang, B., & Yan, J. (2013). Mona: Secure multi-owner data sharing for dynamic groups in the cloud. *IEEE Transactions on Parallel and Distributed Systems*, 24(6), 1182–1191. doi:10.1109/TPDS.2012.331
- Renugadevi, N., & Mala, C. (2014). Tree based group key agreement—a survey for cognitive radio mobile ad hoc networks. In *Advanced Computing, Networking and Informatics* (vol. 2, pp. 85-94). Springer International Publishing. doi:10.1007/978-3-319-07350-7_10
- Shi, H., He, M., & Qin, Z. (2006, December). Authenticated and communication efficient group key agreement for clustered ad hoc networks. In *International Conference on Cryptology and Network Security* (pp. 73-89). Springer Berlin Heidelberg. doi:10.1007/11935070_5
- Steiner, M., Tsudik, G., & Waidner, M. (2000). Key agreement in dynamic peer groups. *IEEE Transactions on Parallel and Distributed Systems*, 11(8), 769–780. doi:10.1109/71.877936
- Wong, C. K., Gouda, M., & Lam, S. S. (2000). Secure group communications using key graphs. *IEEE/ACM Transactions on Networking*, 8(1), 16–30. doi:10.1109/90.836475
- Xue, K., & Hong, P. (2014). A dynamic secure group sharing framework in public cloud computing. *IEEE Transactions on Cloud Computing*, 2(4), 459–470. doi:10.1109/TCC.2014.2366152
- Zheng, S., Manz, D., & Alves-Foss, J. (2007). A communication–computation efficient group key algorithm for large and dynamic groups. *Computer Networks*, 51(1), 69–93. doi:10.1016/j.comnet.2006.03.008

Zheng, S., Wang, S., & Zhang, G. (2007). A dynamic, secure, and efficient group key agreement protocol. *Frontiers of Electrical and Electronic Engineering in China*, 2(2), 182–185. doi:10.1007/s11460-007-0034-7

Zou, X., Ramamurthy, B., & Magliveras, S. S. (2005). Dynamic Conferencing Schemes. *Secure Group Communications over Data Networks*, 91–103.

Zou, X., Ramamurthy, B., & Magliveras, S. S. (2005). Tree Based Key Management Schemes. *Secure Group Communications over Data Networks*, 49–89.

ADDITIONAL READING

Asokan, N., & Ginzboorg, P. (2000). Key agreement in ad hoc networks. *Computer Communications*, 23(17), 1627–1637. doi:10.1016/S0140-3664(00)00249-8

Kim, Y., Perrig, A., & Tsudik, G. (2004). Group key agreement efficient in communication. *IEEE Transactions on Computers*, 53(7), 905–921. doi:10.1109/TC.2004.31

Kumar, A., & Tripathi, S. (2014). Ternary Tree based Group Key Agreement Protocol Over Elliptic Curve for Dynamic Group. *International Journal of Computers and Applications*, 86(7).

Roman, A., & Szykuła, M. (2015). Forward and backward synchronizing algorithms. *Expert Systems with Applications*, 42(24), 9512–9527. doi:10.1016/j.eswa.2015.07.071

Samanthula, B. K., Elmehdwi, Y., Howser, G., & Madria, S. (2015). A secure data sharing and query processing framework via federation of cloud computing. *Information Systems*, 48, 196–212. doi:10.1016/j.is.2013.08.004

Sanchez-Artigas, M. (2013). Toward efficient data access privacy in the cloud. *IEEE Communications Magazine*, 51(11), 39–45. doi:10.1109/MCOM.2013.6658650

Tripathi, S., & Biswas, G. P. (2009, January). Design of efficient ternary-tree based group key agreement protocol for dynamic groups. In 2009 First International Communication Systems and Networks and Workshops (pp. 1–6). IEEE. doi:10.1109/COMSNETS.2009.4808834

Wang, L. L., Chen, K. F., Mao, X. P., & Wang, Y. T. (2014). Efficient and provably-secure certificateless proxy re-encryption scheme for secure cloud data sharing. *Journal of Shanghai Jiaotong University (Science)*, 19(4), 398–405. doi:10.1007/s12204-014-1514-6

KEY TERMS AND DEFINITIONS

Cloud Computing System: Cloud computing is a model for delivering information technology services in which resources are retrieved from the internet through web-based tools and applications rather than a direct connection to a server.

Computational Complexity: A mathematical characterization of the difficulty of a computing group key which describes the resources required by a computing machine to compute the group key. The mathematical study of such characterizations is called computational complexity theory and is important in many branches of theoretical computer science, especially cryptography.

Proxy Re-Encryption: Proxy re-encryption schemes are cryptosystems which allow third parties (proxies) to alter a ciphertext which has been encrypted for one party, so that it may be decrypted by another.

Proxy Signature: Proxy signature, which allows an original signer to delegate his/her signing right to another party (or proxy signer), is very useful in many applications.

Security of Internet–, Intranet–, and Computer–Based Examinations in Terms of Technical, Authentication, and Environmental, Where Are We?

Babak Sokouti

Tabriz University of Medical Sciences, Iran

Massoud Sokouti

Mashhad University of Medical Sciences, Iran

INTRODUCTION

Over the past decades, the use of Internet and its various applications in several environments such as academic communities and industries has been dramatically increased. Fundamental aspects of these applications are commonly related to sending and receiving small or large amounts of data through the local or global network communications. Transferring data over the insecure tunnel of Internet requires applying security services when especially the sensitive data are involved. The main security services are confidentiality, integrity, and availability namely called the CIA triad. The recent advancements in computer networking lead many technologies to be solely or partly upgraded their traditional environments into e-environments and hence, dependent on electronic-based objects such as e-mail and digitalized records and images. The security of shared or archived data or information personally or in organizations are vital and important due to the development of information science in modern society with everyday dramatically increase. Because of its rapid changing nature driven mostly by technology such as smart phones and development of new virtual communities, there are several demands for generating information systems in which security properties play vital roles. By considering these developments, the

e-learning parts of academic communities (i.e., online education) at universities have been affected in order to keep up with the knowledge innovation and they also provide some of their degrees and courses in a distance-based structure by taking the advantage of web-based infrastructures such as LMS (Learning Management System) and MOODLE (Modular Object- Oriented Dynamic Learning Environment). And, this makes the research areas open for modern education era to conduct paperless examinations by providing more security and efficiency. Although, this technological transition in the educational section is to some extent valuable and time effective, however, in most cases the evaluations and examinations of examinee are carried out and monitored in a supervised manner. Precisely speaking, for taking an exam in a traditional examination environment generally examinees, proctors, professors, pen and exam papers and a secure or isolated examining hall are included. For implementing such exams in a distance- or electronic-based model, many security issues whether they are technical or environmental are raised and essential to be resolved. In this regard, an examinee may take the exam alone at home; the professor makes the exam questions and sends them to a web server; the proctors' role can be performed supervised or unsupervised;

the sufficient security measurements should be in place of e-exam's environments.

As of now, providing the most of security services for unsupervised e-exams is still in its infancy and needs much of attention to be taken in to consideration, however, most of technical and communication-based security issues have been addressed in abundance.

In this chapter, the literature researches carried out on potential security aspects of conducting a robust secure e-exam are discussed, then their pros and cons with respect to their provided security services are evaluated and reviewed, and finally a hybrid security model for satisfying most of security properties will be proposed as a secure e-exam model for all conditions along with negotiating future directions.

BACKGROUND

For managing and conducting any types of e-exam systems, literature researchers have performed diverse studies considering special security aspects of them whether they are supervised or unsupervised including technical (e.g., networking, question generation, servers, clients), examinee authentication and identification (e.g., passwords, tokens, and biometrics), environmental monitoring (e.g., webcams, microphones) for preventing possible cheatings during the examination period. Nowadays, for universities providing distance learning courses and degrees are getting epidemically widespread and educationally of much of interest over the world. To survey the current position of state of art related to online exams considering the security aspects, the SCOPUS database is searched and reviewed with the keywords “online” AND “exam” AND “security” in which 19 out of 49 were filtered out based on their relevant contents (i.e., by using title and abstract) for this study. However, one out of 19 was proposed as a web tool for teachers and did not discuss the security aspects of the online examinations (Castillo-Ramoran, 2008).

In order to prevent various types of cheatings, some of the supervised approaches are carried out by employing video and audio devices such as webcams and microphones, biometric recognition tools, manually controlling the port and IP of the connected network devices most of which are relevant case studies for their classrooms in the local area networks (LANs) (Bari, Sullivan, & Blair, 2004; Jung & Yeom, 2009; Savulescu, Polkowski, & Alexandru, 2015). However, these environments are monitored manually via instructor-student networking communications.

From technical point of view and assuming that the only security problems can be carried out by outside attackers, Ming-Ming and Yan have proposed a simple security environment by considering only the data encryption by using triple DES algorithm and secure communication channel (i.e., virtual private network (VPN)) for secure connection of students and stopping the outside attackers (Ming-Ming & Yan, 2013). Also recently, a simple secure computer based examination was also carried out using the B/S structure (Singh & Tiwari, 2016). Another two approaches were conducted as frameworks in which the security countermeasures and the access control to the database servers (e.g., SQL server) were considered (Darong & Huimin, 2010; Haiyan, GHong, Lijun, & Jie, 2014).

For physically distance monitoring of a classroom or a room in which online exams were conducted, an environmental based secure framework was proposed by Stocco and colleagues by incorporating sets of logical and physical sensors (Stocco, Otsuka, & Beder, 2012). Employing the logical architecture of object oriented business application which has five layers including interface layer, interface control business layer, data access, and data storage and management will also bring some advantages (Basar & Haji-zada, 2014). These pros are summarized in scalability, flexibility, and security (Basar & Haji-zada, 2014). So, only the network access manages to stop all types of attacks applied on local and global networks. In other study, the authors have considered four

phases (i.e., registration, examination, marking, and notification) for providing security properties (Giustolisi, Lenzini, & Bella, 2013). In these phases, eligibility, identification, authorization, authentication, authorship, integrity, anonymous marking, mark integrity and privacy are of great importance. Another type of supervised framework is the study which incorporates both teacher and manager to monitor and manage the students during the online exams by considering security properties such as authenticity, privacy, secrecy, receipt of exam answer and copy detection (Castella-Roca, Herrera-Joancomarti, & CDorca-Josa, 2006). During the technological improvements, tablets with their built-in cameras took researchers attentions to develop secure apps for their courses (Gramoll, 2015).

The remaining literature studies have discussed the security properties of the online based examinations in terms of cryptographic algorithms in order to authenticate exam candidates (Karim & Shukur, 2015). Karim and Shukur have demonstrated a review on the list of various authentication methods derived from different databases including Scopus, IEEE Xplore, Springer link, Science Direct, ACM, and Google Scholar (Karim & Shukur, 2015). Some of the samples of the review performed by Karim and Shukur took the advantages of uni or multi authentication methods using biometric properties along with/without username and password each of which have their own strengths and weaknesses. The biometric features include palm print (Al-Saleem & Ullah, 2014), keystroke dynamic, voice recognition, stylometry, face recognition, iris recognition, mouse movements, and hand writing (Karim & Shukur, 2015). Besides these, use of challenge questions (Ullah, Xiao, Barker, & Lilley, 2014) and IP addresses with timestamps (Gao, 2012) are other methods to mention a few.

A typical multimodal biometric technology embedded for authenticating the candidates was proposed to enhance the security of the users' authentication (Asha & Chellappan, 2008). This dynamic based authentication incorporates the

physiological (fingerprint technology) and behavioral (mouse dynamics) properties of the user by means of a mouse device with built-in fingerprint scanner (Asha & Chellappan, 2008). An interactive and secure e-examination unit (ISEEU) was proposed by Sabbah and colleagues where its one-on-one supervised based framework makes it good for continuous authentication using webcams during the video calls in order to reduce the potential user cheatings (Sabbah, Saroit, & Kotb, 2011) by considering that for having a secure environment three user security properties including identity, authentication, and presence should be checked continuously. Moreover, checking the presence via tracking the timestamp having been spent in front of the computer have also discussed elsewhere (Agulla, Rifon, Castro, & Mateo, 2008). It has also been shown that using the keystroke recognition on long-text input under different conditions has less accuracies while different devices in terms of laptop and desktop pc are used which is not applicable for real world conditions (Villani et al., 2006).

Although, these available research papers have proposed, assessed and emphasized on different aspects of security issues related to electronic-based or online exams, however, there are no robust models or protocols to satisfy all security aspects.

MAIN FOCUS OF THE ARTICLE

Issues, Controversies, Problems

Based on the abovementioned literature articles published in the theme of online-exams in terms of security issues, it can be deduced that all of them are not proposed as a whole response for providing and considering all aspects of security. Some of them were only technical based and required information security protocols and standards for a network engineer to meet the essential security measures. On the other hand, several algorithms and approaches were proposed in order to authenticate the user by incorporating several aspects of

authentication whether they were used in a uni or multi factor authentication using properties for “What you are” (i.e., it can satisfied by providing biometric features), “What you have” (i.e., it can be met by a token), and “What you know” (i.e., username and password or multiple question challenges). However, more terms also need to be considered such as “Where you are” and “What you do” that can be checked by monitoring the target environment using video calls, surveillance cams and web cams. Most of recent studies do only satisfy one, two or three out of abovementioned terms which are not a complete answer for having a secure online-examination environment.

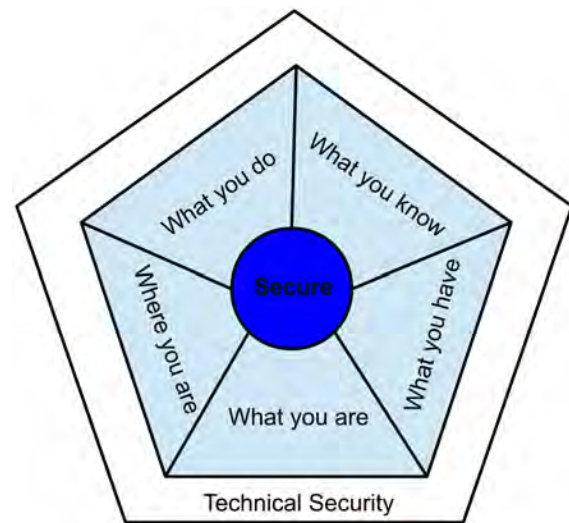
These three categories including technical, authentication, and monitoring of the environment are studied several times but not all together in the literature by its special category. However, they have still their own threats and issues in terms of security and still need more researches on proposing new cryptographic algorithms for identification and authentication in different situations.

SOLUTIONS AND RECOMMENDATIONS

By considering the strengths and weaknesses of the approaches studied in the themes of three categories, none of the categories alone by themselves can be enough for meeting the required security properties that have been discussed before. So, for having a comprehensive security model which includes all five security services shown in Figure 1, it is required to merge all three categories to one and took the advantages of the resulted model in all online exams whether they are taken in a local or global network. By incorporating all of three categories with various combinations, one may conclude that the current model will satisfy all the security properties required for stopping the users’ cheating in any types of local positions or internet based conditions.

In other words, by implementing the unsupervised online exam environments one must be

Figure 1. Required security services for a secure online examination



assured of that all of the security services listed in Figure 1 are in place. For example, providing technical security is solely depending on the local or global administrators’ opinion in order to stop any types of insider (i.e., student, manager, and teacher) and outsider attacks, denial of services (DOS), distributed denial of service (DDOS), and replay attacks by setting up required security countermeasures such as securing the network communication channels (VPNs), data encryption, tracking timestamps, and etc. For the sake of user authentication depending on the situation, username and passwords, tokens, and multi biometric (e.g., fingerprint, mouse dynamics) features are required as one multi authentication package. And finally, the security needs required for an unsupervised situation, several environmental monitoring devices should be employed such as behavioral and movement sensors, webcams, and environmental microphones.

FUTURE RESEARCH DIRECTIONS

Nowadays, the shadow of security can be seen spreading its wings all over the technology ranging from the intelligent tablets to professional

laptops and new car products. The improvements covering all aspects of an object which may include the hardware and technology, software and applications, and cryptographic algorithms and their assessments (i.e., encryption, decryption, digital signatures, key generations) (Bahar, Sokouti, & Sokouti, 2010; M. Sokouti, Sokouti, & Pashazadeh, 2009; Massoud Sokouti, Sokouti, Pashazadeh, Feizi-Derakhshi, & Haghipour, 2013; M. Sokouti, Sokouti, Pashazadeh, & Khanli, 2010; M. Sokouti, Zakerolhosseini, & Sokouti, 2014) will have extensive effects on the performance of online exams in terms of security.

Moreover, in the future studies it is essential for the authors to provide a web site or database server of their implementations as it can be seen widely in the bioinformatics and systems biology realm (B. Sokouti, Church, Morris, & Dastmalchi, 2015; B. Sokouti, Rezvan, & Dastmalchi, 2015, 2016; B. Sokouti, Rezvan, Yachdav, & Dastmalchi, 2014). This will help the researchers assessing their new conducted model or algorithms with previously published approaches.

The last but not the least, it is sometimes recommended to assess and find out the shortcomings of the existing methodologies in order to improve them rather than implement an unknown method with high risks of security issues.

CONCLUSION

In this chapter, security aspects of online exams in different situations were discussed. Three categories of the research interest including technical, authentication methodologies, and environmental monitoring were also covered in terms of both supervised and unsupervised. Moreover, based on the shortcomings and weaknesses in all three categories, a new security model can be carried out considering various combinations of these categories to reach a gold standard and a secure framework for online exams without any types of cheatings.

REFERENCES

- Agulla, E. G., Rifon, L. A., Castro, J. L. A., & Mateo, C. G. (2008). *Is my studnet at the other side? applying biometric web authenntication to e-learning environments*. Paper presented at the Eighth IEEE International Conference on Advanced Learning Technologies. doi:10.1109/ICALT.2008.184
- Al-Saleem, S. M., & Ullah, H. (2014). Security considerations and recommendations in computer-based t esting. *TheScientificWorldJournal*, 2014, 1–7. doi:10.1155/2014/562787 PMID:25254250
- Asha, S., & Chellappan, C. (2008). *Authentication of E-learners using multimodal biometric technology*. Paper presented at the International Symposium on Biometrics and Security Technologies (IS-BAST 2008). doi:10.1109/ISBAST.2008.4547640
- Bahar, H. B., Sokouti, M., & Sokouti, B. (2010). A first study of improving transposition cryptosystem. *Journal of Discrete Mathematical Sciences and Cryptography*, 13(1), 1–9. doi:10.1080/09720529.2010.10698272
- Bari, J., Sullivan, R., & Blair, J. (2004). *Security method in distance-learning*. Paper presented at the 34th ASEE/IEEE Frontiers in Education Conference, Savannah, GA.
- Basar, E., & Haji-zada, T. (2014). *Object oriented business archtechtue on online-exam and assignment system*. Paper presented at the 25th Annual Conference EAEEIE.
- Castella-Roca, J., & Herrera-Joancomarti, J., & CDorca-Josa, A. (2006). *A secure E-exam management system*. Paper presented at the Proceedings of the 1st International Conference on Availability, Reliability and Security (ARES'06). doi:10.1109/ARES.2006.14

- Castillo-Ramoran, V. (2008). *Web tool for teachers: Information systems on categorized teacher materials and online examination*. In *Innovative Techniques in Instruction Technology, E-learning, E-assessment, and Education* (pp. 570–575). Springer Science-Business Media.
- Darong, H., & Huimin, H. (2010). *Realiation and research of online exam system based on S2SH framework*. Paper presented at the International Conference on Web Information Systems and Mining.
- Gao, Q. (2012). Biometric Authentication to Prevent e-Cheating. *International Journal of Instructional Technology & Distance Learning*, 9(2), 3–13.
- Giustolisi, R., Lenzini, G., & Bella, G. (2013). *What security for electornic exams?* Paper presented at the International Conference on Risks and Security of Intrnet and Systems (CRiSIS' 13).
- Gramoll, K. C. (2015). *Development and implementation of a tablet-based exam app for engineering course*. Paper presented at the 122nd ASEE Annual Conference & Exposition, Seattle, WA. doi:10.18260/p.23850
- Hai-yan, L. V., GHong, L. V., Lijun, Z., & Jie, Z. (2014). *Research and design of the common curriculum online examination system that used in military academies*. Paper presented at the 2nd International Conference on Information Technology and Electronic Commerce (ICITEC 2014), Dalian, China. doi:10.1109/ICITEC.2014.7105585
- Jung, I. Y., & Yeom, H. Y. (2009). Enhanced security for online exams using group cryptography. *IEEE Transactions on Education*, 52(3), 340–349. doi:10.1109/TE.2008.928909
- Karim, N. A., & Shukur, Z. (2015). Review of user authentication methods in online examination. *Asian Journal of Information Technology*, 14(5), 166–175.
- Ming-Ming, F., & Yan, M. (2013). The online examination system of distance education. *Applied Mechanics and Matrials*, 411-414, 2901–2905. doi:10.4028/www.scientific.net/AMM.411-414.2901
- Sabbah, Y., Saroit, I., & Kotb, A. (2011). *An interactive and secure e-examination unit (ISEEU)*. Paper presented at the 10th Roedunet International Conference (RoEduNet2011). doi:10.1109/RoEduNet.2011.5993713
- Savulescu, C., Polkowski, Z., & Alexandru, A. I. (2015). *The online and computer aided assessment*. Paper presented at the ECAI 2015-International Conference on Electronics, Computers, and Artificial Intelligence, Bucharest, Romania. doi:10.1109/ECAI.2015.7301226
- Singh, S. K., & Tiwari, A. K. (2016). Design and implementation of secure computer based examination system based on B/S structure. *International Journal of Applied Engineering Research*, 11(1), 312–318.
- Sokouti, B., Church, W. B., Morris, M. B., & Dastmalchi, S. (2015). Structural and Computational Approaches in Drug Design for G Protein-Coupled Receptors. In M. Khosrow-Pour (Ed.), *Encyclopedia of Information Science and Technology* (pp. 479–489). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-5888-2.ch046
- Sokouti, B., Rezvan, F., & Dastmalchi, S. (2015). Applying random forest and subtractive fuzzy c-means clustering techniques for the development of a novel G protein-coupled receptor discrimination method using pseudo amino acid compositions. *Molecular BioSystems*, 11(8), 2364–2372. doi:10.1039/C5MB00192G PMID:26108102
- Sokouti, B., Rezvan, F., & Dastmalchi, S. (2017). GPCRTOP v.1.0: One-step web server for both predicting helical transmembrane segments and identifying G protein-coupled receptors. *Current Bioinformatics*, 12(1), 80-84. doi:10.2174/1574893611666160901122236

Sokouti, B., Rezvan, F., Yachdav, G., & Dastmalchi, S. (2014). GPCRTOP: A novel G protein-coupled receptor topology prediction method based on hidden Markov model approach using Viterbi algorithm. *Current Bioinformatics*, 9(4), 442–451. doi:10.2174/1574893609666140516010018

Sokouti, M., Sokouti, B., & Pashazadeh, S. (2009). An Approach in Improving Transposition Cipher System. *Indian Journal of Science and Technology*, 2(8), 9–15.

Sokouti, M., Sokouti, B., Pashazadeh, S., Feizi-Derakhshi, M.-R., & Haghipour, S. (2013). Genetic-based random key generator (GRKG): A new method for generating more-random keys for one-time pad cryptosystem. *Neural Computing & Applications*, 22(7), 1667–1675. doi:10.1007/s00521-011-0799-8

Sokouti, M., Sokouti, B., Pashazadeh, S., & Khanli, L. M. (2010). FPGA Implementation of Improved Version of Vigenere Cipher. *Indian Journal of Science and Technology*, 3(4), 459–462.

Sokouti, M., Zakerolhosseini, A., & Sokouti, B. (2014). Improvements over GGH Using Commutative and Non-Commutative Algebra. In M. Khosrow-Pour (Ed.), *Encyclopedia of Information Science and Technology* (3rd ed.; pp. 3404–3418). IGI-Global.

Stocco, J. A., Otsuka, J. L., & Beder, D. M. (2012). *Logical and physical sensor-based online assessment security support*. Paper presented at the 12th IEEE International Conference on Advanced Learning Technologies.

Ullah, A., Xiao, H., Barker, T., & Lilley, M. (2014). Evaluating security and usability of profile based challenge questions authentication in online examinations. *Journal of Internet Services and Applications*, 5(2), 1–16.

Villani, M., Tappert, C., Ngo, G., Simone, J., Fort, H. S., & Cha, S.-H. (2006). *Keystroke bimetric recognition studies on long-text input under ideal and application-oriented conditions*. Paper presented at the IEEE Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06). doi:10.1109/CVPRW.2006.115

ADDITIONAL READING

Baumgart, D. C., Wende, I., & Grittner, U. (2016). Tablet computer-based multimedia enhanced medical training improves performance in gastroenterology and endoscopy board style exam compared with traditional medical education. *Gut*, 65(3), 535–536. doi:10.1136/gutjnl-2015-309895 PMID:26123026

Blevins, A., & Besaw, M. E. (2011). Reflecting on quiz placement in online tutorials: Creating a more active learning experience. *Medical Reference Services Quarterly*, 30(3), 316–324. doi:10.1080/02763869.2011.590427 PMID:21800988

Emery, M. (2010). With no shelf exam available, SAEM online clerkship testing tool best option. *Academic Emergency Medicine*, 17(9), 1028, author reply 1029. doi:10.1111/j.1553-2712.2010.00855.x PMID:20836790

Gibson, S. I. (2015). Promoting an active form of learning out-of-class via answering online study questions leads to higher than expected exam scores in General Biology. *PeerJ*, 3, e1322. doi:10.7717/peerj.1322 PMID:26500828

Glatz, K., Savic, S., Glatz, D., Francz, G., Barascud, A., Grilli, B., & Bubendorf, L. et al. (2006). An online quiz uncovers limitations of morphology in equivocal lung cytology. *Cancer*, 108(6), 480–487. doi:10.1002/cncr.22286 PMID:17091509

Gray, J., & Jackson, C. (2004). The development of an online quiz for drug calculations. *Nursing Times*, 100(4), 40–41. PMID:14974263

Johnson, B. C., & Kiviniemi, M. T. (2009). The Effect of Online Chapter Quizzes on Exam Performance in an Undergraduate Social Psychology Course. *Teaching of Psychology*, 36(1), 33–37. doi:10.1080/00986280802528972 PMID:20046908

Khan, R., Krupinski, E., Graham, J. A., Benodin, L., & Lewis, P. (2012). Assessing first year radiology resident competence pre-call: Development and implementation of a computer-based exam before and after the 12 month training requirement. *Academic Radiology*, 19(6), 752–758. doi:10.1016/j.acra.2011.12.019 PMID:22297203

Metz, A. M. (2008). The effect of access time on online quiz performance in large biology lecture courses. *Biochemistry and Molecular Biology Education*, 36(3), 196–202. doi:10.1002/bmb.20184 PMID:21591191

KEY TERMS AND DEFINITIONS

C

Authentication: A process for knowing if someone who he is declared to be.

Availability: A guarantee for an information resource being accessed all the time by authorized users. CIA Triad: There most required ones for security services (i.e., Confidentiality, Integrity, and Availability).

Confidentiality: A security service equivalent to privacy which can be achieved by encrypting the message using a cryptographic algorithm.

Identification: It is related to the identity of a person who claims about it.

Integrity: An assurance for information not being altered or changed during the communication.

Online Exam: A type of exam which can be taken overseas or locally with or without manual supervision.

A Three-Vector Approach to Blind Spots in Cybersecurity

Mika Westerlund

Carleton University, Canada

Dan Craigen

Carleton University, Canada

Tony Bailetti

Carleton University, Canada

Uruemu Agwae

Carleton University, Canada

INTRODUCTION

With the increased use of network technologies (Clements & Kirham, 2010), cybercrime is on the rise. PricewaterhouseCoopers estimates that 120,000 cyberattacks occur daily (PwC, 2016). There is a need for cybersecurity throughout society. Cybersecurity is defined as “the organization and collection of resources, processes, and structures used to protect cyberspace and cyberspace-enabled systems from occurrences that misalign de jure from de facto property rights” (Craigen, Diakun-Thibault, & Purse, 2014). It is also the measure of preparedness including recovery, protection, and triage against the losses caused by cyberattacks (Maughan, 2010).

Cybersecurity is key for protecting valuable assets such as intellectual property, virtual currencies, and industrial control systems (Kritzinger & Solms, 2010; Smith & Rupp, 2002). However, cyberattacks are often successful due to “blind spots,” which refer to various biases and preconceived information that affects organizational and human decision making (Heuer, 1999; Pronin, Lin, & Ross, 2002), leading to unawareness of malicious activity (Boehm & Turner, 2005). It

is important to understand how to mitigate all blind spots, particularly those that can lead to massive economic losses (Flowers, Zeadally, & Murray, 2013).

The objective of this chapter is to investigate eight cyberattack cases (“attack scenarios”) from the viewpoint of “the core vectors” which include economic, technological and psychological perspectives to blind spots. While previous research has viewed core vectors in isolation from each other (Baker, 2014; Garfinkel, 2012; Singer & Friedman, 2014), this chapter focuses on how to mitigate blind spots in cybersecurity by using a holistic three-vector approach. The holistic view to cybersecurity has been suggested by many authors (Emami-Tabatabaie, Amoui, & Tahvildari, 2013; Hua & Bapna 2013; Hughes & Cybenko, 2013).

Section one provides an overview of cyberattacks and blind spots that enable attacks. Section two discusses core vectors conceptualized as psychological, economic, and technical perspectives to blind spots. Sections three and four discuss research methods and eight scenario cases. Section five presents a summary table of the cases included in the sample. Finally, sections six and

seven discuss future research avenues and implications to practice.

BACKGROUND

Han and Dongre (2014) list political, economic, and socio-cultural motives as primary motives for cyberattacks, and emphasize that attackers can be organizational insiders or outsiders. Political motives include cyber terrorism against foreign nations or multinationals (Hua & Bapna, 2013) and ethically fighting for justice and human rights (Gandhi et al., 2011). Other motives may be plain entertainment. Regardless, there is a propensity for harm when cyberattacks occur. Understanding what enables these attacks enables mitigation, and will contribute to the theory on blind spots in cybersecurity (Chen, Huang, Xu, & Lai, 2015; Nathan & Petrosino, 2003).

Blind spots are dangerous because they are about biases and preconceptions (Pronin et al., 2002). Humans tend to interpret new information so that prior conclusions remain intact. A 2012 survey by the National Cyber Security Alliance (NCSA) and Symantec revealed that 83% of small U.S. companies did not have a formal plan for keeping their business cyber-secure although over 70% responded that a safe and trusted Internet is critical to their day-to-day operations. A total of 76% thought that their company was safe from cyber-security breaches.

Given that blind spots are inevitable, there is a need to develop more efficient means to mitigate them. Thus, it is critical to understand how (i) the business, (ii) the psychological, and (iii) the technological perspectives might help organizations and individuals to recognize and avoid blind spots. This core vector thinking is supported by the cybersecurity assessment factors by Gavins and Hemenway (2010) and the categorization of attacker motivations by Han and Dongre (2014). Combining vectors enables a comprehensive analysis of past cyberattack scenarios in order to mitigate blind spots.

PERSPECTIVES TO CYBERSECURITY

C

Cybersecurity should be examined through psychological, economic, and technological vectors. Wiederhold (2014) argues that psychology in human nature is the weakest link in cyberspace. Although technology is sophisticated, humans still fall victim to social engineering (Bauer & van Eeten, 2009). Economic factors are important as computers are infested with malware to enrich attackers, and victims ponder associated costs (Hua & Bapna, 2013).

Psychological Vector

Human factors affect cybersecurity the most (Baker, 2014), and when cyber threats are not understood or are ignored, adverse consequences accrue. What people perceive of risks affects how they behave (West, 2008). The 2012 NCSA/McAfee survey showed that 64% of Americans feel their smartphone is safe from cyberattacks even though they had not installed any security software.

Sukamol and Markus (2008) argue that many individuals make security decisions from a simplistic understanding of risk. In organizations, risk has focused on technology, but firms are now expanding their technology-centered perspective to include people and processes (PwC, 2016). Heuer (1999) notes that it takes more information and more unambiguous information to recognize an unexpected phenomenon than an expected one. Moreover, Cebula and Young (2010) argue that many people use the same combination of user ID and password for different information systems because of human memory limitations thereby creating a huge risk to cybersecurity.

The psychological vector is associated to human thinking and cognitive limitations affecting decision making. Humans are especially prone to blind spots when it comes to complex systems like law, politics, or cyberspace. A typical user may not be able to identify an email phishing attack or

a cloned website; whatever captures their minds as close enough is good enough. Thus, “unknown unknowns” associated with complex systems make an intellectual blind spot (Brotherton, 2015), and human understanding of risk plays a big role in fostering secure actions.

Economic Vector

Different industries are exposed to different cyber threats. The banking and finance sector is most concerned about financial fraud, and 65% of the frauds are cyberspace related (Chakrabarty, 2012). The information and communication technology sector is most concerned about illicit access to email and other electronic communications, because they offer access to trade secrets and business-sensitive information potentially harming the whole value chain (PwC, 2016). The insurance sector is primarily concerned about financial losses due to sensitive information being exposed unintentionally. Attackers clearly have financial motives and the victim normally suffers economic losses. Globally, the financial loss from cybersecurity incidents is increasing by 34% annually (PwC, 2016).

Unfortunately, individuals and organizations decide on cybersecurity investments under economic constraints such as limited budgets (Heuer, 1999). Although the increased engagement of the Board of Directors has impacted a boost in security spending (PwC, 2016), the financial cost attributed to cybersecurity investment is seen as huge by most businesses. Major issues among businesses is the inability to devise a cybersecurity investment strategy and the organizational inability to invest in the necessary personnel and processes to guard against attacks. Security risks can be mitigated through understanding one’s own vulnerabilities and the motivations of cybercriminals.

Supply chain partners are also dangerous. More than two-thirds of companies in the automotive industry attribute security incidents to partners such as resellers that have trusted access to their internal networks and data (PwC, 2016). However,

security checks for such partners are normally not practiced. The 2012 NCSA/Symantec survey shows that two-thirds of small companies are not protecting nor letting the partners know of their cybersecurity practices. Companies that have not suffered an attack may not know where to prioritize their cybersecurity investments, while companies that have suffered an attack are better handling security gaps and investing wisely.

Technological Vector

It is often assumed that technically inclined individuals have the knowledge to solve cybersecurity issues. However, the responsibility of handling security issues is not purely an IT department concern, but is everyone’s responsibility (Kritzing & von Solms, 2010). This raises the question of whether technology is a solution. West (2008) argues that sophisticated technology does not decrease system vulnerabilities, and that the more complex the technology the more vulnerabilities. Although organizations and individuals may use technology to avert attacks, they cannot guarantee immunity.

Many network security solutions, such as firewalls, are automated background processes (Garfinkel, 2012). However, cybersecurity requires users or operators to act appropriately; the lack of network security understanding may result in abuse of privacy and security (Singer & Friedman, 2014).

Simpler designs may help in developing technology with fewer security vulnerabilities. A system with many features and functionality is more likely to be compromised (Saltzer & Schroeder, 1975). It is important to check access and objects continually, enforce automatic security updates, and provide multiple levels of security so that if any bit is compromised, there is a safe alternative. Encryption makes it more difficult for data to be compromised, and technical alert systems can be used to prevent or mitigate unwanted occurrences (Shapiro & Cohen, 2007). However, technological

risks will never be totally eliminated as attackers shift to targeting vulnerable mobile technology.

METHODS

This study analyzes eight scenarios of successful cyberattacks. These scenarios occurred between 2007 and 2013, and the information used was collected from various online news websites. Possible vulnerabilities of each cyberattack is elaborated from the perspective of psychological, technological, and business vectors. Blind spot theory recognizes that people tend to focus on their perceptions of relevance, yet there are blind spots in regard to what people consider irrelevant or “unknown unknowns” (Taleb, 2012). These vectors provide an understanding of the enablers of blind spots; had these vulnerabilities been identified and addressed before a cyberattack, threats could have been mitigated and attacks avoided.

The analysis makes use of cybersecurity operational risks by Cebula and Young (2010), according to whom actions of people can be deliberate (fraud, sabotage, theft) or accidental (mistakes, errors), or a lack of action. Increased risk can also result from technology failures (systems design, specifications, and complexity; software configuration, settings, and practices; hardware capacity and obsolescence). Moreover, failed internal processes (problems in the internal business process flow, documentation, and alerting), process controls (status monitoring), and supporting processes (staffing, training, funding) affect the ability to implement, manage, and sustain cybersecurity. Finally, external events (dependencies on service providers, economic conditions) may result in risks.

FINDINGS

This sections summarizes eight cybersecurity incidents; the findings from the perspective of the three core vectors are emphasized with brackets.

- C
U. S. Banks: In September 2012, there was a series of Distributed Denial of Service (DDoS) attacks on major U.S. financial institutions, including JPMorgan and Citigroup. Virtual private web servers were used to flood the online banking system with junk email, attacking vulnerable web applications. The high volume of Internet traffic made the websites unavailable and resulted in client disruption. Iran was suspected to be the attacker to retaliate for the economic sanctions from the U.S. and its Arab allies. The banks assumed their servers were secure but discovered otherwise when they were attacked (psychology), although the technical functionalities of the hosted servers worked according to design (technology). The attack failed to affect the core of the business (business).
- Mat Honan:** In August 2012, Mat Honan’s entire digital life was destroyed by a hacktivist group who called Applecare customer tech support to obtain a temporary password for Honan’s me.com email using his billing address and last four digits of the credit card for identification. They found Honan’s billing address by searching his web domain information and received the last four digits by calling Amazon.com and requesting to add an additional credit card to Honan’s existing account. Although hackers failed to pass security questions, they were able to reset Honan’s AppleID password. His iPhone was disconnected from the Internet and iCloud login disabled, his twitter and Gmail accounts were deleted, and all his Apple devices were remotely erased. Honan had daisy-chained his Google and iCloud accounts using the same email prefix and did not have two-factor authentication for convenience reasons (psychology). The ‘Find my Mac’ application, a supposedly helpful feature, turned into an attacking tool used to wipe out Honan’s details (technology). Apple

violated their internal security processes and Amazon.com made access to sensitive information too easy for the sake of business and customer convenience (business).

- **Dalai Lama:** In 2008, the Office of His Holiness the Dalai Lama (OHHDL) sent out a confidential email to an unnamed country who was later warned by the Chinese government not to host the Dalai Lama. Investigations revealed infiltrations into the OHHDL's system. There were several unauthorized logins into email accounts from IP addresses belonging to Chinese and Hong Kong providers. The hackers composed emails that appeared to come from the OHHDL, causing recipients to visit malicious websites. Using phishing, the hackers were able to install rootkits on computers in several embassies, foreign ministries, and organizations and download sensitive data. Obviously, human thinking assumed it was safe to open these emails and click the links (psychology). OHHDL's web service was not encrypted and the vulnerabilities in office software allowed malicious content to be downloaded (technology). Organizations affected by the attack focused on their core processes and did not evaluate the potential loss of data or information due to their unsafe practices (business).
- **Koobface:** The Koobface worm has terrorized users of social media since 2008. Infection allowed hackers to access users' sensitive banking information and passwords. They aimed at users with invitations to watch a funny or sexy video. Those who clicked the link got a message to update their computer's Flash software, which began the download of the malware. Victims' computers were drafted into a network of infected computers, and were sent advertisements of fake antivirus software. Also, their web searches were hijacked and the clicks delivered to unscrupulous marketers. The group made money from people

who bought the bogus software and from unsuspecting advertisers. Koobface checks the user's Internet cookies to identify the social network the user has access to. Thus, they were able to send messages to the networking site for the user and there was no need to create a fake account on the social networking site (technology). They also capitalized on their victims' unwise click choices and the trust between the user and their social network contacts to increase the rate of malicious infection (psychology). More users, more traffic, more revenue is the pattern of thought of social media sites, and security that would prevent this kind of abuse comes later (business).

- **Cutting Sword of Justice:** Aramco is Saudi Arabia's largest oil and gas firm with affiliates all over the world. In 2012, Aramco's computer network was infested with a replicating virus Shamoon that affected its business process and caused the loss of data. Shamoon infected over 30,000 computers and caused a major disruption to the firm's business. It took almost two weeks to recover from the effect of the attack. The virus activates at a particular time, displays an image of a burning American flag and then sends reports of information to the attacker; moreover, files can be deleted from infected computers. The effects of the Shamoon can cause a firm to lose its Intellectual Property. Both an anti-oppression group and the Cutting Sword of Justice group took responsibility for the attack. The virus only affected computers used for Aramco's enterprise activities, not those used for oil production (technology). Thus, an in-house security issue may be involved (psychology). Surprisingly, there was no backup done for the drilling and production data during the period of the attack (technology). Later, Aramco invested heavily on cybersecurity and fast recovery from similar situations (business).

- Target:** The U.S. retailing company Target was attacked in late 2013. The attack was enabled by a malware planted in their point-of-sale terminals that allowed hackers to steal customers' information from the magnetic stripe of millions of debit and credit cards. The hackers accessed Target's network through a third party that had access to the network because they were monitoring the energy and temperature in stores for fluctuations in temperature that might affect the shopping experience (psychology). The hackers accessed the third party network using an email malware. Target's information system was not upgraded to work with the EMV chip technology that would enable them to read card information in an encrypted format (technology). Target had worked with Visa for over ten years to introduce smart cards, which would limit fraud by encrypting sensitive data contained on the card. However, they discarded the project after spending tens of millions because it was cheaper to issue and accept traditional magnetic-stripe cards than credit cards with silicon chips (business).
- Spamhaus:** In 2013, Spamhaus, an anti-phishing group, added Cyberbunker to its public black list, suspecting Cyberbunker to be an avenue of spam. Cyberbunker was a commercial Internet Service Provider with apparently undisclosed and shady activities. After Spamhaus blacklisted Cyberbunker, they experienced a severe DDoS against their email and web servers making it impossible for anyone to access their website. Soon thereafter, the spokesman of Cyberbunker, who was arrested, said "Spamhaus had too much control in determining and in deciding what spam is and what isn't". He operated within a mobile computing van from a well-equipped bunker and the attacks were generated by swarms of computers known as botnets (technology). After his arrest, there was an anonymous demand for his release, adding that if he were not released, there would be a larger attack. Since Cyberbunker felt hurt by Spamhaus, they decided to hurt and attack Spamhaus in return (psychology). Spamhaus was doing their job by listing those hosts who they felt were a spamming threat. Cyberbunker, on the other hand, was also pursuing their business interests by allowing spamming services to be hosted on their servers (business).
- New York Times (NYT):** In 2013, hackers were able to access every employee's password. The attack was achieved through spear-phishing where the hackers emailed malicious attachments or links to employees. Once an employee clicked on the email or link containing the malicious code, the hackers were able to install a remote access tool which aided them to extract data, passwords, images, keystrokes, documents, computer webcam and microphone recordings. The hackers focused on reporters who were preparing investigative reports on how the Chinese Prime Minister's family had gained their billion-dollar fortune from questionable business dealings. The hackers, assumed to be Chinese, wanted to determine the writers' sources of information for these reports. The attackers installed malicious software for a period of three months, during which time the firewall and antivirus program only discovered one instance of the presence of malicious software and had it quarantined. Hackers no longer focus their efforts in attempting to break firewalls, rather they capitalize on the social lapses of an unassuming human who will click on suspicious links and files received by email (psychology). Hashed passwords were made available for employees to make it difficult to steal their

passwords, but rainbow tables used by hackers nullify the security (technology). The NYT relied on a security provider to combat malicious threats while focusing on their own core business; however, no provider can mitigate all hacker exploits of human vulnerabilities (business).

A THREE-VECTOR APPROACH

The scenario analysis used three perspectives to blind spots in cybersecurity: (i) psychological (limitations of the human mind), (ii) technological (critical flaws and the need for secure

technology), and (iii) business (commercial logic behind attacks and security investments). Table 1 provides a combined three-vector approach to attack scenarios in order to illustrate blind spots as enablers of cyberattacks.

The resulting eight blind spots explain how human cognitive limitations; technological exploitation of vulnerable information systems; and economic motives of attackers lead to cyberattacks towards under-resourced. The eight blind spots are identified in Table 1. The blind spots were derived from a multi-vector analysis, providing a more holistic view compared to traditional approaches to cybersecurity risks.

Table 1. Blind spots and their mechanisms identified from the eight scenario cases

Example	Blind Spot	Mechanism of Enabler
U.S. Banks	Internal focus	Emphasis on digital technology that protects core operations and systems and secures internal processes (prevents hacking aimed at fraud and theft) allows attackers to harm non-core activities by targeting the outside interface (DDoS to prevent customers from entering a bank website)
Mat Honan	Convenience first	Allowing the customer to turn off critical security features and ignoring corporate security procedures in favor of customer convenience allows attackers to exploit holes, especially when the customer links social media accounts and login credentials for convenience
Dalai Lama	Habitual aspect	Office solutions have security vulnerabilities, and the attacker can exploit the victims' habitual aspect for unauthorized access to their systems
Koobface	Hedonistic enjoyment	The attacker is able to lure victims to click malicious files while they search for funny and entertaining content; installed malware is configured to spread through the victim's online social networks arising from a human need for social interaction and social enjoyment
Cutting Sword of Justice	Insider sabotage	Sabotage is driven by political or economic motives, and may require an insider who is able to contaminate the victim's systems and hinder their ability to recover from the attack; sabotage may be based on deterrence and creates harm even when aimed at the victim's non-core activities
Target	Wrong priorities	Wrong priorities for the sake of business and smooth operations, such as customer convenience, costs savings and functionality before technical security, as well as the unnecessary technical connection between core systems and supportive systems expose organizations to cyberattacks
Spamhaus	Egotistical behavior	Egotistical behavior of actors may spark cyber warfare, where both parties feel that their actions are justified; one side's under-evaluation of the other side's narcissistic motivation (i.e. revenge) and subsequent actions exposes them to large-scale and systematic full-frontal cyberattacks
NY Times	Technology effect	Overreliance on technology is dangerous if organizations and people believe a single technology prevents all kinds of cyber threats; however, such one-stop technology does not exist, and even the most complex security solutions are useless when "the keys to the lock" are easily accessible

FUTURE RESEARCH DIRECTIONS

Many individuals and organizations guard themselves against cyberattacks; yet some are affected by the lack of implementing sufficient safe practices. In lieu of this, there are newly devised and sophisticated avenues of attacks that are highly dynamic. It would take substantial effort to provide guidance to respond to all avenues. Thus, this chapter addresses ‘what perspectives can be used to identify blind spots and how the perspectives can be combined to help us mitigate blind spots?’

The examined scenarios showed that attacks could be averted or mitigated, if the victims focused attention on the basic principles and limitations of their actions. Future research could analyze additional scenarios to better contribute to the theory of blind spots and to spark the building of a systematic tool that applies the results to real businesses. The tool could be enriched by a more in-depth “what happened and what did not happen” analysis of the scenarios. An evident limitation to this type of research is data availability; many attacks are not publicized because of victims’ reluctance to discuss their specific scenarios.

CONCLUSION

The outcome of this study is that the theory of blind spots in the cybersecurity context can be advanced by making use of a combination of perspectives (“a three-vector approach”). Blind spots exist because there is no focus or there are unknowns, and because people tend to focus on what they think is relevant. These blind spots, illustrated by findings from the study, allow hostile parties to execute cyberattacks. Thus, the responsibility of handling security issues in organizations is not solely for the IT department; rather it is everyone’s responsibility. Organizations and individuals who handle valuable assets in cyberspace will find the present study useful because they will be better able to spot cybersecurity gaps and avoid social and economic costs from successful cyberattacks.

Suggested ways of improving cybersecurity include creating defenses against immediate threats by enhancing mutual situational awareness of threats and vulnerabilities. Moreover, organizations and individuals should develop the ability to swiftly respond to vulnerabilities and prevent network intrusions. In addition, expanding cyber education, research and development efforts, and strategies will likely mitigate malicious activities in cyberspace. A number of organizations, however, are not willing to increase awareness or disclose vulnerabilities because they believe it will create a negative impression for investors and organizational image. In order to enhance cybersecurity return on investment, businesses should provide funding for shorter terms rather than the typical 3-5 years business planning.

REFERENCES

- Baker, E. W. (2014). A model for the impact of cybersecurity infrastructure on economic development in emerging economies: Evaluating the contrasting cases of India and Pakistan. *Information Technology for Development*, 20(2), 122–139. doi:10.1080/02681102.2013.832131
- Bauer, J. M., & van Eeten, M. J. (2009). Cybersecurity: Stakeholder incentives, externalities, and policy options. *Telecommunications Policy*, 33(10), 706–719. doi:10.1016/j.telpol.2009.09.001
- Boehm, B., & Turner, R. (2005). Management challenges to implementing agile processes in traditional development organizations. *Software, IEEE*, 22(5), 30–39. doi:10.1109/MS.2005.129
- Brotherton, R. (2015). *Suspicious minds: Why we believe conspiracy*. Bloomsbury.
- Cebula, J. J., & Young, L. R. (2010). *A taxonomy of operational cyber security risks*. Technical Note CMU/SEI-2010-TN-028. Retrieved from Carnegie Mellon University website: <http://www.sei.cmu.edu/reports/10tn028.pdf>

- Chakrabarty, K. C. (2012). Indian Banking Sector: Towards the next orbit. In Academic Foundation (Ed.), *Economic developments in India: Analysis, reports, policy documents*. New Delhi: AF.
- Chen, Y.-Z., Huang, Z.-G., Xu, S., & Lai, Y.-C. (2015). Spatiotemporal patterns and predictability of cyberattacks. *PLoS ONE*, 10(5), e0124472. doi:10.1371/journal.pone.0124472 PMID:25992837
- Clements, S., & Kirkham, H. (2010). Cybersecurity considerations for the smart grid. In *Proceedings of the IEEE Power and Energy Society General Meeting*. Minneapolis, MN: PES. doi:10.1109/PES.2010.5589829
- Craigien, D., Diakun-Thibault, N., & Purse, R. (2014). Defining cybersecurity. *Technology Innovation Management Review*, 4(10), 13–21.
- Emami-Tabla, M., Amoui, M., & Tahvildari, L. (2013, August). On the road to holistic decision making in adaptive security. *Technology Innovation Management Review*, 59–64.
- Flowers, A., Zeadally, S., & Murray, A. (2013). Cybersecurity and US legislative efforts to address cybercrime. *Journal of Homeland Security and Emergency Management*, 10(1), 29–55. doi:10.1515/jhsem-2012-0007
- Gandhi, R., Sharma, A., Mahoney, W., Sousan, W., Zhu, Q., & Laplante, P. (2011). Dimensions of cyber-attacks: Cultural, social, economic, and political. *Technology and Society Magazine, IEEE*, 30(1), 28–38. doi:10.1109/MTS.2011.940293
- Garfinkel, S. L. (2012). The cybersecurity risk. *Communications of the ACM*, 55(6), 29–32. doi:10.1145/2184319.2184330
- Gavins, W., & Hemenway, J. (2010). Cybersecurity: A joint terminal engineering office perspective. In *Proceedings of the IEEE Military Communications Conference*. San Jose, CA: MILCOM. doi:10.1109/MILCOM.2010.5679588
- Han, C., & Dongre, R. (2014). What motivates cyber-attackers? *Technology Innovation Management Review*, 4(10), 40–42.
- Heuer, R. J. (1999). *Psychology of intelligence analysis*. Washington, DC: CIA.
- Hua, J., & Bapna, S. (2013). The economic impact of cyber terrorism. *The Journal of Strategic Information Systems*, 22(2), 175–186. doi:10.1016/j.jsis.2012.10.004
- Hughes, J., & Cybenko, G. (2013, August). Quantitative metrics and risk assessment: The three tenets model of cybersecurity. *Technology Innovation Management Review*, 15–24.
- Kritzinger, E., & von Solms, S. H. (2010). Cyber security for home users: A new way of protection through awareness enforcement. *Computers & Security*, 29(8), 840–847. doi:10.1016/j.cose.2010.08.001
- Maughan, D. (2010). The need for a national cybersecurity research and development agenda. *Communications of the ACM*, 53(2), 29–31. doi:10.1145/1646353.1646365
- Nathan, M. J., & Petrosino, A. (2003). Expert blind spot among preservice teachers. *American Educational Research Journal*, 40(4), 905–928. doi:10.3102/00028312040004905
- Pronin, E., Lin, D. Y., & Ross, L. (2002). The bias blind spot: Perceptions of bias in self-versus others. *Personality and Social Psychology Bulletin*, 28(3), 369–381. doi:10.1177/0146167202286008
- PwC. (2016). *The global state of information security survey 2016*. Industry Report. Retrieved from PricewaterhouseCoopers website: www.pwc.com/gsis
- Saltzer, J., & Schroeder, M. (1975). The protection of information in computer systems. *Proceedings of the IEEE*, 63(9), 1278–1308. doi:10.1109/PROC.1975.9939

Shapiro, J. N., & Cohen, D. K. (2007). Color bind: Lessons from the failed homeland security advisory system. *International Security*, 32(2), 121–154. doi:10.1162/isec.2007.32.2.121

Singer, P. W., & Friedman, A. (2014). *Cybersecurity and cyberwar: What everyone needs to know?* Oxford University Press.

Smith, A. D., & Rupp, W. T. (2002). Issues in cybersecurity; understanding the potential risks associated with hackers/crackers. *Information Management & Computer Security*, 10(4), 178–183. doi:10.1108/09685220210436976

Sukamol, S., & Markus, J. (2008). Using cartoons to teach internet security. *Cryptologia*, 32(2), 137–154. doi:10.1080/01611190701743724

Taleb, N. N. (2012). *Antifragile: Things that gain from disorder*. Random House.

West, R. (2008). The psychology of security. *Communications of the ACM*, 51(4), 34–40. doi:10.1145/1330311.1330320

Wiederhold, B. K. (2014). The role of psychology in enhancing cybersecurity. *Cyberpsychology, Behavior, and Social Networking*, 17(3), 131–132. doi:10.1089/cyber.2014.1502 PMID:24592869

ADDITIONAL READING

Davis, A. (2015). Building cyber-resilience into supply chains. *Technology Innovation Management Review*, 5(4), 19–27.

Kadivar, M. (2014). Cyber-attack attributes. *Technology Innovation Management Review*, 4(11), 22–27.

Muegge, S., & Craigen, D. (2015). A design science approach to constructing critical infrastructure and communicating cybersecurity risks. *Technology Innovation Management Review*, 5(6), 6–16.

Uma, M., & Padmavathi, G. (2013). A survey on various cyberattacks and their classification. *International Journal of Network Security*, 15(5), 390–396.

KEY TERMS AND DEFINITIONS

Attack Scenario: A cybersecurity incident where an attacker was able to gain illegitimate access to a victim's sensitive information and/or information systems.

Blind Spots: Various biases and preconceived information that affect organizational and human decision making; enabled by unknowns or false focus on what is relevant.

Core Vectors: Economic, technological and psychological perspectives to blind spots; used in combination to obtain a better understanding of what enables cyberattacks.

Cyberattack: A deliberate exploitation of computer systems, technology-dependent enterprises and networks.

Cybersecurity: The organization and collection of resources, processes, and structures used to protect cyberspace from occurrences that misalign de jure from de facto property rights.

Category C

Cyber Crime, Cyber Bullying, and Digital Terrorism

Cyber Bullying

C

Jo Ann Oravec

University of Wisconsin – Whitewater, USA

INTRODUCTION

“Cyberbullying” comprises a wide spectrum of behaviors that have negative and often devastating impacts upon their targets (or “victims”). This article is intended to analyze research trends on cyberbullying as well as related concerns involving online harassment, online reputational damage, and cyberstalking. Its focuses are as follows: (1) analyze the creative and innovative conceptual work and research that have emerged on the technological as well as social and ethical aspects of these issues; (2) present insights as to how cyberbullying and reputational damage can best be mitigated, given the technological capabilities and emerging know-how of technical specialists, educators, and organizational consultants; and (3) discuss why cyberbullying is of continuing importance to a broad business, government, non-profit, and educational audience. The article also includes reflections about the moral and personal dimensions of cyberbullying.

Bullying incidents are intricate and frustrating phenomena from whatever contexts they emerge, face-to-face or online. Cyberbullying often involves words and pictures that are considered as protected speech under various national and local laws, often providing cover for those who are attempting to abuse or unsettle a victim (Fraser, Bond-Fraser, Buyting, Korotkov, & Noonan, 2013; Oravec, 2012). The prevalence of cyberbullying is difficult to determine, given the privacy with which many cases dealing with juvenile offences (as well as offences in workplace settings) are handled. In their “Scoping Review on Studies of Cyberbullying Prevalence Among Adolescents,” Brochado, Soares, and Fraga (2016) found the following:

Most of the studies tend to assess cybervictimization experiences. However, even considering the same perspective, the same country, and the same recall period, a high variability in the estimates was observed. As a main conclusion, the way in which the prevalence of cyberbullying is estimated is influenced by methodological research options.

Added to the difficulties in studying cyberbullying is the observation that some individuals who are victimized may not display signs of damage or even choose to respond to the bully. Other individuals may be extremely harmed by comparable words and pictures; some may choose to fight back while others become depressed or even suicidal. Generally, cyberbullying consists of repetitive behavior that has a particular focus on a victim; the bully’s attacks can be shielded from video from others who could possibly intervene. As related in Bonanno and Hymel (2013), “Cyberbullying also takes place on a virtual playground that makes it possible to victimize a peer within the sanctity of one’s own home, at any time of the day or night, in complete anonymity, and with maximal exposure and hence potential embarrassment for the intended target” (p. 646). Kamali (2015) adds an angle relating to the growing assortment of networked devices: “the perpetrator can employ varied means (e.g., cell-phones, texts, blogs, Internet, social media, etc.)” in conducting the bullying (p. 43). Anonymity can give some protection to bullies in shielding them from observation in whole or part of their bullying (Barlett, Gentile, & Chew, 2016). Cyberbullies can also combine anonymous interactions with personally-identifiable ones to make it appear that more than one individual is involved in the attacks, potentially intensifying the negative social impacts of the bullying.

DOI: 10.4018/978-1-5225-2255-3.ch148

Equipping individuals in workplace, community, and educational contexts to be aware of cyberbullying issues may enable them to become more sensitive and empathetic as well as more effective as the front line of defense against these phenomena. It will also help them to mitigate bully-related problems in their organizational and community roles and provide an “early warning system” for the new forms of bullying various technological changes may engender (such as drones and virtual reality). Emerging research efforts may also enhance understanding of cyberbullying as well as empower citizens and organizational participants in their efforts to mitigate it.

BACKGROUND

Public discourse on bullying and mobbing has increased dramatically in the past decade: in the *Journal of Psychohistory*, Dervin (2010) labelled 2010 as “The Year of the Bully” because of the number of shocking incidents involving young people, many of which incorporated some social media component. The classic novel *Lord of the Flies* (Goulding, 1960) has often been mentioned in discourse on youth-related cyberbullying; many of the young people characterized by Goulding had forms of autonomy and power within their environments but did not have the moral guidance to understand fully the consequences of their actions. Cyberbullying attracted public attention in part because of cases such as that of Rutgers University student Tyler Clementi who committed suicide after having some of his sexual activities broadcast via webcam on the Internet (Byers, 2013; Oravec, 2012). “Cyberstalking” has been distinguished as a phenomenon related to and often overlapping with cyberbullying (Adam, 2002; Kamali, 2015); it is more often associated with the conduct of adults than cyberbullying, the latter which has been more often associated with children and adolescents (Carter, 2015). Research from Milosevic (2015) shows that in US mainstream media “overall debate on cyberbullying is narrow, focused on incidents that resulted in suicides, and

subsequent blaming of individuals involved. Such framing can have implications for audience’s support of punitive policies, inability to comprehend complexity of the issue and moral panic around children’s use of technology.” (p. 492).

Computer networking is adding new and complex dimensions to bullying, providing the venue for complex cases of online reputational damage and privacy invasion via social media (Oravec, 2013; Van Royen, Poels, Daelemans, & Vandebosch, 2015). Victims, bullies, and onlookers can often interact in debilitating and confusing patterns in cyberbullying incidents, given their relatively-recent emergence in society and lack of common understandings concerning them. As new forms of cyberbullying emerge, individuals who work online and engage in networked educational and social activity can encounter harassment, belittling, damaging or doctored materials that refer to them, and related harms. Technologies such as digital video, text messaging, and social media can be combined in ways that make victims perceive that they have nowhere to hide and no real legal or social recourse. Victims often lose the ability to contribute their full efforts to their jobs, schoolwork, and social interactions, and sometimes even choose to end their lives.

Cyberbullying can affect people of all gender affiliations; it can also affect individuals of various religious heritages as well as racial and ethnic backgrounds (Cappadocia, Craig, & Pepler, 2013; Stanbrook, 2014). Individuals with disabilities can be targeted as well as become perpetrators themselves (Kowalski *et al.*, 2016). Certain patterns can be found in the way cyberbullying and online reputational damage emerge in the context of various groups and settings, which can help in detecting bullying and mitigating its consequences. For example, some varieties of cyberbullying infuse forms of hate speech (as described in Adam, 2002, Byers, 2013, and Oravec, 2000), integrating themes involving race, gender, religion, or other perceived affiliation; some of this speech, however debilitating to the victim and bystanders, may be protected by the First Amendment (Conn, 2015).

MAIN FOCUS OF THE ARTICLE

Issues, Controversies, Problems

As individuals engage in increased levels of online activity, the prospects that they will find their online interactions and reputations negatively impacted by others who have somehow targeted them increase as well. Many recent political debates have been affected by cyberbullying, often preventing individuals from engaging fully in citizenship and public forums. For example, Schwartz (2016) describes a number of circumstances in which women's blogs with some political content have been placed under attack with repetitive harassment and abuse. Gender-related and ethnic insults have been common in the online attacks perpetrated in social media as well as email, listservs, and texting venues (Slonje, Smith, & FriséN, 2013); attacks involving sexting (sex-themed technologically-mediated communications) have proliferated as well (Durham, 2016; Oravec, 2012b). In many of these incidents, cyberbullies are enabled to remain anonymous or disguise their identities in ways that only a few of their followers would be able to decipher, increasing the difficulties for law enforcement or administrators to mitigate the situations.

For its victims, cyberbullying has been shown to be linked with subsequent stress and even suicidal ideation (Kowalski et al., 2014). The impact of cyberbullying is often intensified by the ubiquitous nature of computer networking: "For the victim, cyber bullying is pervasive and persistent; once a cyberattack is launched, it is difficult, if not impossible, to eliminate... As a result, the impact of cyber victimization can be (or can be perceived to be) far more negative than that of traditional forms of bullying, and victims may have a much more difficult time coping" (Bonanno & Hymel, p. 686). Cyberbullying has also been shown to affect academic and workplace motiva-

tion, making its control and mitigation a practical as well as moral concern for institutions (Oravec, 2012a; Young-Jones, 2015). Research by Coyne *et al.* (2016) explores "the relationship between experiencing workplace cyberbullying, employee mental strain and job satisfaction," formulating a "disempowerment" approach to understanding its negative effects on the workplace. Kamali (2015) calls for increases in college-directed activity in this arena: "Cyberbullying research on college campuses lacks a unified definition of the concept. Although the states mandated most school districts to develop and enact some sort of policy, the law is silent on the college level cyberbullying" (p. 43).

The study of "serial cyberbullying" (cyberbullying behaviors that are repeated over time by one individual or group) is providing clues as to how bullying operates and who perpetrates it. From an organizational systems perspective, serial bullying is "a cycle that generates when a target is singled out, bullied, and driven from the workplace and regenerates when another target is singled out, bullied, driven from the workplace, and so on" (Lutgen-Sandvik & Tracy, 2012, p. 16). Serial bullies, simply put, can become increasingly proficient in what they do, and their negative impacts on organizations and individuals can be compounded over time as reinforcing network structures grow around them. Repetitive cyberbullying conduct on the part of particular individuals could signal deepseated psychological conditions that can be mitigated through mental health treatment. There are dangers in labeling specific individuals as "serial bullies," however; if these individuals are blocked from subsequent employment and educational opportunities (perhaps unfairly) their life chances can be diminished and opportunities to rehabilitate reduced. However, learning about serial bullying patterns and watching for them can empower human resource specialists and educators to contain repetitive bullying behaviors before they injure more people.

SOLUTIONS AND RECOMMENDATIONS

Cyberbullying prevention and response often involves assistance with the technologies incorporated in cyberbullying for those with low levels of computer literacy. For example, victims often need guidance in documenting cyberbullying and online reputational damage so that their cases can be examined and cyberbullying mitigated. They may need updating on aspects of social media reputation protection and personal privacy guardianship. Unfortunately, many of the organizations that support online platforms (as well as obtain profits from their use) are slow to respond to cyberbullying complaints; a combined initiative among many organizational and online units needs to be maintained for cyberbullying to be mitigated (Stanbrook, 2014). It may certainly be difficult in many workplace contexts for administrators to admit that their organizations have issues with cyberbullying and seek assistance. New research findings about cyberbullying along with various social and artistic initiatives are beginning to generate insights for assistance in helping victims and concerned onlookers. For example, YouTube videos, compelling posters and artwork, and celebrity appearances along with Twitter emojis are being used in public service campaigns to signal that bystanders are becoming aware and involved and cyberbullying is not acceptable behavior (Hayman & Coleman, 2016). Doane, Kelley, and Pearson (2016) describe their research showing that “a brief cyberbullying video is capable of improving, at one-month follow-up, cyberbullying knowledge, cyberbullying perpetration behavior, and TRA constructs known to predict cyberbullying perpetration” (p. 136).

Many well-meaning individuals apparently have an impulse to disseminate widely information about particular cyberbullying cases in public venues. However, publicity about cyberbullying can backfire, despite the intentions of reducing negative outcomes (Pieschl, Kuhlmann, & Porsch, 2015). Publicity can sometimes put the victim’s

own personal history and personal characteristics in a problematic light as people wonder why he or she was targeted. Such publicity may have bad repercussions whether or not the victim’s name was directly mentioned as rumors spread about who was involved.

Surveillance by well-meaning organizational or governmental authorities can play roles in mitigating some aspects of bullying, identifying and documenting potential bullying situations in hopes of containing their damage (Van Royen, *et al.*, 2015). However, with surveillance measures, privacy invasions can emerge as well as the real potential that whatever materials are collected will be misused or taken out-of-context. In turn, Mendola (2015) presents the case that such monitoring has the potential to displace the roles of household controls in a “Big Brother as Parent” scenario. Understanding how surveillance and monitoring can be used to document and frame bullying so as to mitigate its effects will require extensive experimentation in organizational contexts as well as the sharing of best practices.

FUTURE RESEARCH DIRECTIONS

The needs for continuing research on cyberbullying and for the proactive approaches and insights it will engender are considerable. As forms of social media expand, cyberbullying is developing new modes and inflicting new kinds of reputational and emotional devastation. Many participants, onlookers, and victims in cyberbullying and online reputational damage are young adults who may not yet understand the full gravity of harassment, misrepresentation, and related behaviors (Carter, 2015; Oravec, 2013). They are being faced with a difficult and confusing assortment of challenges as they establish their social lives as well as take on moral leadership roles in these contexts. It can be difficult to make a decision as to whether to join with a cyberbully in an online attack (and participate in “mobbing,” defined in the glossary) or have a positive and effective

role as an active intervener or victim supporter. Bullies can be perceived as powerful people, and joining with them may seem at first as a way to enlarge one's own presence and become more powerful in a workplace or educational context (Bussey, Fitzpatrick, & Raman, 2015). Some of the creative tactics against cyberbullying that have been examined by researchers include forms of humor levied by the victims themselves. In a study of Turkish adolescents, Sari (2016) found that "as aggressive humor decreases [on the part of victims], likelihood of cyberbullying perpetration increases. These results indicate that maladaptive humor styles (aggressive humor and self-defeating humor) may successfully address cyberbullying behavior" (p. 555). Research on perceptions of cyberbullying, and how positive associations of them can be eroded (perhaps through humor), may elicit insights as to how to deal with these phenomena.

International comparisons and cross-cultural investigations of cyberbullying will reveal a great deal not just about bullying as a phenomenon but about cultural variations in social psychology (Bayraktar, 2015). For example, in some pioneering longitudinal research on adolescents in Canada, Cappadocia, Craig, and Pepler (2013) found that "Risk factors associated with cyberbullying included higher levels of antisocial behaviors and fewer prosocial peer influences. Risk factors associated with cybervictimization included being in the transition year for high school, as well as higher levels of traditional victimization and depression." (p. 171). Scales such as the Revised Cyberbullying Inventory (RCBI) can be used to elicit results that can provide some sort baselines for what is happening with cyberbullying phenomena over time and across cultures. Using the RCBI in a study of British adolescents, Brewer and Kerslake (2015) found the following:

Empathy was a significant individual predictor of cyberbullying perpetration, such that as empathy decreases, likelihood of cyberbullying perpetration increases. These findings indicate

that self-esteem and empathy oriented interventions may successfully address cyberbullying behavior. (p. 255)

C

Cultural and psychological insights on cyberbullying may be productively combined with cybercounseling approaches to design workplace and educational interventions (Mishna, Bogo, & Sawyer, 2015) which can be implemented and evaluated in real-life contexts.

CONCLUSION

Decades ago, many of the future projections for the uses of computer networks seemed to present bright and optimistic portraits (Natale & Balbi, 2014; Oravec, 1996); computing technology innovations would be used primarily to inform, share, and befriend others. However, many dystopian scenarios are emerging worldwide of individuals who are afraid to participate in online interaction because of the negative impact it can have on their lives, scenarios that are echoed in the real-life experiences of many students and employees (Adam, 2002; Byers, 2013). Cyberbullying research has indeed documented specific cases of cyberbullying and elicited trends, but the deeper and more difficult moral questions remain of why so many individuals use computer technology in ways that disempower and wound people. These phenomena are apparently taking on new forms as computing technologies expand their reach societally (in terms of their national and international permeation) and personally (in terms of how we live our everyday lives).

In conclusion, managers, administrators, educators, and concerned citizens need to consider how to deal with generations of individuals (from young people to older adults) who have suffered from painful and reputation-damaging personal attacks online. Workplaces have also been sites for extended and sophisticated cyberbullying efforts, which have served to decrease levels of organizational morale as well as productivity. Many

students have become acquainted with cyberbullying in school contexts, in some cases learning how to be participants and in others how it feels to be a victim; educational institutions may be able to foster more positive and healthy behaviors as well. Inclusive excellence cannot be obtained if individuals live in fear. People cannot be expected to learn effectively or produce their best work if they regularly need to deal with cyberbullies or are themselves inappropriately or unfairly being construed in that light. Continuing research and public discourse on these matters is necessary for society to benefit fully from computer technology and social media.

REFERENCES

- Adam, A. (2002). Cyberstalking and Internet pornography: Gender and the gaze. *Ethics and Information Technology*, 4(2), 133–142. doi:10.1023/A:1019967504762
- Barlett, C. P., Gentile, D. A., & Chew, C. (2016). Predicting cyberbullying from anonymity. *Psychology of Popular Media Culture*, 5(2), 171–180. doi:10.1037/ppm0000055
- Bayraktar, F. (2015). *Cross-national and cross-cultural variances in cyberbullying*. In *The Wiley Handbook of Psychology, Technology and Society* (pp. 158–175). New York: Wiley.
- Bonanno, R. A., & Hymel, S. (2013). Cyber bullying and internalizing difficulties: Above and beyond the impact of traditional forms of bullying. *Journal of Youth and Adolescence*, 42(5), 685–697. doi:10.1007/s10964-013-9937-1 PMID:23512485
- Brewer, G., & Kerslake, J. (2015). Cyberbullying, self-esteem, empathy and loneliness. *Computers in Human Behavior*, 48, 255–260. doi:10.1016/j.chb.2015.01.073
- Brochado, S., Soares, S., & Fraga, S. (2016). A scoping review on studies of cyberbullying prevalence among adolescents. *Trauma, Violence & Abuse*. doi:10.1177/1524838016641668 PMID:27053102
- Bussey, K., Fitzpatrick, S., & Raman, A. (2015). The role of moral disengagement and self-efficacy in cyberbullying. *Journal of School Violence*, 14(1), 30–46. doi:10.1080/15388220.2014.954045
- Byers, D. (2013). Do they see nothing wrong with this?: Bullying, bystander complicity, and the role of homophobic bias in the Tyler Clementi case. *Families in Society*, 94(4), 251–258. doi:10.1606/1044-3894.4325
- Cappadocia, M. C., Craig, W. M., & Pepler, D. (2013). Cyberbullying prevalence, stability, and risk factors during adolescence. *Canadian Journal of School Psychology*, 28(2), 171–192.
- Carter, S. (2015). *The hostile environment: Students who bully in school*. New York: Lexington Books.
- Conn, K. (2015). From Student Armbands to Cyberbullying: The First Amendment in Public Schools. *Legal Frontiers in Education: Complex Law Issues for Leaders, Policymakers and Policy Implementers* (Advances in Educational Administration, Volume 24). New York: Emerald Group Publishing Limited, 35–58.
- Coyne, I., Farley, S., Axtell, C., Sprigg, C., Best, L., & Kwok, O. (2016). Understanding the relationship between experiencing workplace cyberbullying, employee mental strain and job satisfaction: A dysempowerment approach. *International Journal of Human Resource Management*, 1–28. doi:10.1080/09585192.2015.1116454
- Dervin, D. (2011). 2010: The Year of the Bully. *The Journal of Psychohistory*, 38(4), 337–345.

- Doane, A. N., Kelley, M. L., & Pearson, M. R. (2016). Reducing cyberbullying: A theory of reasoned action-based video prevention program for college students. *Aggressive Behavior*, 42(2), 136–146. doi:10.1002/ab.21610 PMID:26349445
- Durham, M. G. (2016). *Technosex: Precarious corporealities, mediated sexualities, and the ethics of embodied technics*. New York: Springer. doi:10.1007/978-3-319-28142-1
- Fraser, I., Bond-Fraser, L., Buyting, M., Korotkov, D., & Noonan, S. (2013). Cyber bullying and the law: Are we doing enough. *The American Association of Behavioral and Social Science Journal*, 17(1), 26–39.
- Goulding, W. (1960). *Lord of the flies*. London: Faber & Faber.
- Hayman, S., & Coleman, J. (2016). *Parents and digital technology: How to raise the connected generation*. New York: Routledge.
- Kamali, A. (2015). Assessing cyber-bullying in higher education. *Information Systems Education Journal*, 13(6), 43.
- Kowalski, R. M., Giumetti, G. W., Schroeder, A. N., & Lattanner, M. R. (2014). Bullying in the digital age: A critical review and meta-analysis of cyberbullying research among youth. *Psychological Bulletin*, 140(4), 1073–1137. doi:10.1037/a0035618 PMID:24512111
- Kowalski, R. M., Morgan, C. A., Drake-Lavelle, K., & Allison, B. (2016). Cyberbullying among college students with disabilities. *Computers in Human Behavior*, 57, 416–427. doi:10.1016/j.chb.2015.12.044
- Leymann, H. (1990). Mobbing and psychological terror at workplaces. *Violence and Victims*, 5, 119–126. PMID:2278952
- Lutgen-Sandvik, P., & Tracy, S. J. (2012). Answering five key questions about workplace bullying how communication scholarship provides thought leadership for transforming abuse at work. *Management Communication Quarterly*, 26(1), 3–47. doi:10.1177/0893318911414400
- Mendola, C. E. (2015). Big Brother as parent: Using surveillance to patrol students' Internet speech. *British Columbia Journal of Law & Social Justice*, 35, 153.
- Milosevic, T. (2015). Cyberbullying in US mainstream media. *Journal of Children and Media*, 9(4), 492–509.
- Mishna, F., Bogo, M., & Sawyer, J. L. (2015). Cyber counseling: Illuminating benefits and challenges. *Clinical Social Work Journal*, 43(2), 169–178. doi:10.1007/s10615-013-0470-1
- Natale, S., & Balbi, G. (2014). Media and the imaginary in history. *Media History*, 20(2), 203–218. doi:10.1080/13688804.2014.898904
- Oravec, J. A. (1996). *Virtual individuals, virtual groups*. New York: Cambridge University Press. doi:10.1017/CBO9780511574986
- Oravec, J. A. (2000). Countering violent and hate-related materials on the Internet: Strategies for classrooms and communities. *Teacher Educator*, 35(3), 34–45. doi:10.1080/08878730009555233
- Oravec, J. A. (2012a). Bullying and mobbing in academe: Challenges for distance education and social media applications. *Journal of Academic Administration in Higher Education*, 49, 46–58.
- Oravec, J. A. (2012b). *The ethics of sexting: Issues involving consent and the production of intimate content*. In *Digital Ethics: Research and Practice* (pp. 129–145). New York: Peter Lang.

Oravec, J. A. (2013). Gaming Google: Some ethical issues involving online reputation management. *Journal of Business Ethics Education*, 10, 61–81. doi:10.5840/jbee2013104

Pieschl, S., Kuhlmann, C., & Porsch, T. (2015). Beware of publicity! Perceived distress of negative cyber incidents and implications for defining cyberbullying. *Journal of School Violence*, 14(1), 111–132. doi:10.1080/15388220.2014.971363

Sari, S. V. (2016). Was it just a joke? Cyberbullying perpetrations and their styles of humor. *Computers in Human Behavior*, 54, 555–559. doi:10.1016/j.chb.2015.08.053

Schwartz, A. (2016). Critical blogging: Constructing Femmescares online. *Ada: A Journal of Gender, New Media, and Technology*, (9).

Slonje, R., Smith, P. K., & Frisé, N. A. (2013). The nature of cyberbullying, and strategies for prevention. *Computers in Human Behavior*, 29(1), 26–32. doi:10.1016/j.chb.2012.05.024

Stanbrook, M. B. (2014). Stopping cyberbullying requires a combined societal effort. *Canadian Medical Association Journal*, 186(7), 483–483. doi:10.1503/cmaj.140299 PMID:24664656

Van Royen, K., Poels, K., Daelemans, W., & Vandebosch, H. (2015). Automatic monitoring of cyberbullying on social networking sites: From technological feasibility to desirability. *Telematics and Informatics*, 32(1), 89–97. doi:10.1016/j.tele.2014.04.002

Young-Jones, A., Fursa, S. F., Byrket, J. B., & Sly, J. S. (2015). Bullying affects more than feelings: The long-term implications of victimization on academic motivation in higher education. *Social Psychology of Education*, 18(1), 185–200. doi:10.1007/s11218-014-9287-1

ADDITIONAL READING

Jaffe, E. M. (2016). Swatting: The new cyberbullying frontier after *Elonis v United States*. *Drake Law Review*, 64, 455–553.

Maiuro, R. (Ed.). (2015). *Perspectives on bullying: Research on childhood, workplace, and cyberbullying*. New York: Springer Publishing Company.

Starcevic, V., & Aboujaoude, E. (2015). Cyberchondria, cyberbullying, cybersuicide, cybersex: New psychopathologies for the 21st century? *World Psychiatry: Official Journal of the World Psychiatric Association (WPA)*, 14(1), 97–100. doi:10.1002/wps.20195 PMID:25655165

KEY TERMS AND DEFINITIONS

Bystanders: Individuals who have access to information about current or potential bullying or harassment incidents, whether or not they choose to respond. In online contexts, information that is received by bystanders may be fragmented and highly confusing.

Cyberstalking: Repetitive and personally invasive surveillance and monitoring of a target or victim, often with unsettling forms of unwanted contact. Cyberstalking is often associated with adult behaviors more than those of children or young adults (Kamali, 2015).

Mobbing: Group member cooperation and collaboration in bullying. From the victim's perspective, mobbing compounds the negative power of abusive behavior with the sense of exclusion from the group, producing a form of "psychological terror." From the mobbers' perspectives, mobbing may make the participants feel less guilty about their behaviors since their peers and comrades are also engaging (Leymann, 1990).

Online Reputational Damage: Injury to the online profiles and compiled reputational information about an individual. This injury can be conducted through such means as doctoring photographs and manipulating online information in a negative way (Oravec, 2013).

Serial Cyberbullying: Individuals who engage in bullying who have chosen more than one target or victim, either simultaneously or sequentially. Serial bullying has been widely documented

in non-computer assisted forms of bullying as well as online varieties. It has also been observed among children and young people (Carter, 2015).

Victim: The intended target of the online abuse. The term “victim” can it itself have negative implications for those for whom it is applied, with connotations of “opportunistic victimhood” and assumptions that the target somehow is benefiting from the abuse or did something to attract it intentionally.

C

Cyberbullying Among Malaysian Children Based on Research Evidence

Sarina Yusuf

Universiti Putra Malaysia, Malaysia

Md. Salleh Hj. Hassan

Universiti Putra Malaysia, Malaysia

Adamkolo Mohammed Mohammed Ibrahim

Universiti Putra Malaysia, Malaysia & University of Maiduguri, Nigeria

INTRODUCTION

The Internet, a technology that cuts across geographical borders and overcomes the challenges of time is considered the most influential technological development in the 21st Century (Dholakia & Kshetri 2004). It has been described as the modern Pandora's Box that has opened a new cyberspace of threats to unwary users (Ktoridou, Eteokleous & Zachariadou 2012). Given the broad use and the rapid sophistication of the cyber technology over the years, the Internet is gradually replacing conventional (traditional) media of communication and becoming not only a new means communication but also a new means of interaction and socialization for millions of users nowadays (Ktoridou et al. 2012).

Rapid innovative advancements in the cyber technology and cyber-related technologies have exceptionally augmented the capacity and categories of users over the decades. According to Dholakia and Kshetri (2004), the Internet took only three years to reach 50 million users around the world compared to the 13 years taken by the radio and 38 years by the television. Approximately 2.9 billion people all over the world have access to the Internet, representing around 41% of the world population, and this number continues to increase (Kenda 2014).

The cyberspace is rapidly becoming a medium for people to meet their everyday needs like searching for information, communicating and entertainment (University of Southern California, [USC] Annenberg School Center for the Digital Future [ASCDF], 2008). The cyberspace has certainly provided enormous benefits and advantages to the people (Lenhart, Madden & Hitlin 2005). A meta-analysis study by Guan and Subrahmanyam (2009) has found that the cyber technology is a useful tool for promoting youth cognitive, social and physical development. In addition, scholars have emphasised the great utility of the Internet in successfully raising awareness regarding health issues, enhancing social relationships, maintaining community links among the young people and numerous other benefits (Barak & Sadovsky 2008; Subrahmanyam & Greenfield 2008; Flicker, Maley, Ridgley, Biscope, Lambardo & Skinner 2008).

Despite the positive effects of the cyberspace on users, the technology exerts some negative influences as well (e.g., cyberbullying), especially on children and the youth (Livingstone, 2012). Nowadays, Malaysian children and adolescents are increasingly using cyber technology at younger ages, with diverse devices and technologies (Balakrishnan 2015). In recent years however, there has been growing concern by the public, schools and parents about children and youth involvement

in online risk behaviors such as cyberbullying and harassment (Finkelhor, Mitchell & Wolak 2000; Wolak, Mitchell & Finkelhor 2006). Many Malaysian youth and children have been exposed on daily basis to numerous Internet risks and harms (Balakrishnan 2015).

Given the ease of access and use of Internet technologies, even children learn how to use it; some children even go to the extent of protecting their online privacy from the prowling eyes of their parents (Yusuf, Osman, Hassan & Teimoury 2014). With just a click on the keyboard/keypad, children can access almost anything, including any kind of video, article and image they desire, including what they might stumble across accidentally (Balakrishnan 2015; Ktoridou et al. 2012), which often leads to incidences of cyberbullying, where children are being bullied by others while online (Yusuf et al. 2014).

This suggests that cyberbullying has been posing growing concern, especially for the online safety of the youth and children. Furthermore, there have been significant and rapid changes in children and youth's online activities in the last one decade in the country (Balakrishnan 2015; Yusuf et al. 2014). Evidently, there has been remarkable increase in the use of smartphones by the youngsters coupled with their migration from social activity to social networking sites across societies in the world (Jones, Mitchell & Finkelhor 2012) and in Malaysia (Abu Bakar 2013; Balakrishnan 2015).

Abu Bakar (2013) has empirically supported the issue that cyberbullying is a silent epidemic in the Malaysian society. The scholar indicated that online harassment and sexually based cyberbullying are the most common forms of cyberbullying in the country. Over the recent years, many incidents of cyberbullying had occurred on the Internet, which led to cases such as homicide, sexual assault, humiliation and assault. Cyberbullying incidents are increasing every in the country and becoming a potent threat to children, parents

and other stakeholders (Neging, Musa, R. & Abdul Wahab 2013).

Norton Family Report (NOFR) (2010) indicates that the rising popularity of the new media and the increasing amount of time children spend online pose great challenges to parents and or guardians who want to protect their children from the threats and harms of the cyberspace. Hence, this study was prompted by the urge to understand the prevalence of cyberbullying in Malaysia by determining the level of the phenomenon among children. One major area that this chapter has contributed to is the development of a scale suitable for measuring cyberbullying in Malaysian society. Given that previous research focuses on qualitative approach such as face-to-face and telephone interviews, there has been a dearth of quantitative scale to measure cyberbullying, particularly on sexually based cyberbullying and online harassment dimensions. Therefore, this study adopted a quantitative approach.

BACKGROUND

Cyberbullying in Malaysia

According to an Interpol statistics, online harassment is growing rampantly all around the world, and Malaysia is no exception. It has been reported that one in every five children online becomes the target of cyber predators and 30% of female children have been sexually harassed in a chat room (Azizan 2012).

The cyber security unit of Malaysia have conducted a survey on the effect of cyber threats on children and adolescents and found that about 60% of the cases reported to them were about cyberbullying on social networking sites such as Facebook and MySpace (Nik Anis, Abdul Rahim, & Lim 2012). Despite the limited number of literature on cyberbullying among Malaysian youngsters, a study conducted by Balakrishnan (2015) found

that cyberbullying does occur among Malaysian youth. However, Balakrishnan (2015) suggests that occurrence of the phenomenon among the respondents of the study (youth aged 17 to 30 years old) was not as prevalent as it is among children and teenagers, and that social networking sites are the most likely domains where cyberbullying incidences take place.

In addition, there are many none-empirical evidences of cyberbullying occurrences in Malaysia. In 2010, not less than 60 cases of cyberbullying were reported to the cyber security unit of Malaysia in 2007 (McKenna 2010). Surprisingly however, the number of reported cases rose to 5,181 in 2009 and 11,930 in 2010 (Cyber Security Malaysia Report [CSMR] 2011). Furthermore, various media reports in Malaysia have suggested that cyberbullying is prevalent among children and the youth. However, much of it could have been under-reported because most people are unaware that it has become a serious issue (Eek 2009).

According to Roberts and Samani (2013), the McAfee and ABA (Anti Bullying Alliance) surveys have found that cyberbullying is one of the critical online phenomena that children and youth face nowadays. Nik Anis et al. (2012) also reported that besides cyberbullying, the main online threats to Malaysian children while online are cyber-grooming, identity theft and child pornography.

The findings of a national survey conducted by Malaysian Cyber Security in 2011 indicates that more than 90% of children aged 5 to 18 in Malaysia may have been exposed to cyber threat. Findings of NOFR (2010) survey indicate that due to their excessive use of the Internet, 87% of 553 Malaysian children surveyed indicated having been bullied by cyber predators. Furthermore, the results of a national survey of school-going children in Malaysia shows that almost 30% of the children indicated that they have been bullied while online. However, only 13% of them admitted being bullied almost every day on the Internet (CSMR 2014, 2013).

LITERATURE REVIEW

Children in Digital Age

The literature stresses that social networking sites like Facebook and Twitter, search engines such as Google, MSN (Microsoft Network), Yahoo and WIKIs are an established part of Internet users' daily life events (Eteokleous & Pavlou 2010). Adolescents mainly under the age of 18 are the highest users of this digital realm. These adolescents are commonly referred to as Digital Learners (Murugesan 2009), Net Generation (Myers, McCaw & Hemphill 2011), Millennial Teens or Generation Y (Negin et al. 2013) and Digital Natives (Prensky 2001) since they were born and growing up in the Information Technology (IT) age (Ktoridou et. al 2012). This group of users is the main consumer of the Internet since the first official outbreaks of the Internet in the mid-1990s (Ward 2011).

In contrast, children that were born or growing up during the time of Internet technological evolution are classified as the digital immigrants (Prensky 2001). These users are using the cyber technology and associated technologies based on their needs for survival by adjusting to the environments as part of their daily lives (Ktoridou et al. 2012). Since new generations are the frequent users of the Internet, it follows that they are more aware of the use, the tools and the features of the Internet than the older generations (Davidson & Martellozzo 2008). Consistent with the facts, a research conducted in 21 European countries has showed that 75% of children in European countries are using the Internet and the rate continues to grow everyday (Livingstone & Haddon 2009).

Adolescents and youths are the most fervent users of the Internet, who have embraced cyber technology and digital media as a lifestyle for communicating, making connections, interacting and social networking (Ktoridou et al. 2012). The researchers further indicated that the Internet has the ability to communicate with people without

limits regardless of the languages, cultural and geographical diversity, simply by using the social network (e.g. Facebook and Twitter), chat rooms (e.g. Skype and Yahoo Messengers) and various other forums. Through all the convenient tools for socializing, people from all over the world with various background and cultures can connect easily and quickly. According to Eteokleous (2011), this technology enabled children to understand, to discover new things, and to express their feelings without boundaries and limitations. In addition, a survey by the Malaysian Communication and Multimedia Commission (MCMC 2014) reports that almost 65% of Malaysian population has access to the Internet out of which 17% are children and adolescents aged 7 to 19 years old.

A recent research conducted in the United Kingdom (UK) reported that one of the main predictors of risks and harm among children as young as nine years old is indulging in risky online behaviors (Livingstone & Gorzig 2014). Furthermore, nearly 11% and 13% of adolescents within the age range of 9 to 16 years old in the UK have already seen an online pornography and seen or received hateful messages respectively (Livingstone 2012).

A study found that 95% of young adults aged 18 to 29 have easy access almost anywhere to the Internet via high-speed broadband and smartphones (Pew Internet and American Life Project [PI&ALP] 2013; Zickhur & Smith 2013). According to them, this scenario has led us to the disturbing reality of the Internet norms these days that not all of the access to the Internet is safe, particularly for adolescents and children.

A number of research surveys have suggested that majority of children aged between 12 and 14 years use the Internet intensively, spending about 15 to 20 hours per week communicating, passing pastime and entertaining themselves (Livingstone, Bober & Helsper 2005; USC ASCDF 2008, 2005; Van den Eijnden, Spijkerman, Vermulst, Van Rooij & Engels 2010; World Internet Project [WIP] 2013; Zamaria & Fletcher 2007).

The Concept of Bullying

According to Carter (2011), resolving conflicts amongst adolescence is a normal part of their maturation process, and usually will be disclosed in bullying behaviors, with bullies threatening other students who may be perceived as vulnerable or weak. This kind of behavior is identified as either traditional bullying or cyberbullying. Since bullying behaviors occurred during the process of growing up, Banks (1997) has identified that bullying might increase during elementary school, will be at uttermost frequency during middle school years and will be decreasing in high school.

In order to comprehend the phenomenon called cyberbullying, it is important to understand the concepts of traditional bullying better. Olweus (2002) formulated a systematic terminology of bullying. The researcher defines victimization or bullying as a situation when a student is exposed repeatedly and over time to negative actions on the part of one or more students. The researcher further explains that negative action is when someone intentionally inflicts, or attempts to inflict, injury or discomfort upon another (Olweus 2002). Smith, Mahdavi, Carvalho, Fisher and Russell (2008) comprehensively but concisely defined bullying as the act of verbal, physical, relational and indirectly bullying often referred to in the scholarly literature, media and Internet as traditional bullying.

This early concept of bullying has been used for decades by researchers to identify the bullying behaviors among children and adolescents. Nansel, Overpeck, Pilla, Simons-Morton and Scheidt (2001) reinforced the term bullying with an additional important characteristic, the imbalance of power that exists in which the bully or bullies are considered more powerful by attacking a victim who is less powerful. Based on this consideration, experts of bullying behaviors have shared same views that bullying is a major concern in the society as it leads to negative outcomes for the victims' physical and psychological well-being

(Noordahl, Poole, Stanton, Walden & Beran 2008; Smokowski & Kopasz 2005). Many victims of bullying can vividly recall being harassed during childhood, especially victims who cannot properly defend himself or herself because of the difference in strength and size (Hinduja & Patchin 2009). According to a dissertation by Carter (2011), bullying problems often go undetected and unreported because many people view bullying as a normal part of life especially in middle and high school, as most people believe that it is normal for children to fight and they have to learn how to protect themselves the hard way.

The Concept of Cyberbullying

Cyberbullying has been defined as the act of willfully using electronic and/or cyber technologies to deliberately and repeatedly harass or threaten someone by sending or posting cruel text and/or graphic messages (Abu Bakar 2013; Berson, Berson & Ferron 2002; Finkelhor, Mitchell & Wolak 2001, 2000; Hinduja & Patchin 2009). Smith et al. (2008) defined cyberbullying as an aggressive, intentional act carried out by a group or individuals, using electronic forms of contact, repeatedly and over time against a victim who cannot easily defend him or herself.

Researchers argued that cyberbullying could occur anywhere and at any time, this empowers bullies much more than traditional bullying does because more targets could be harassed through cyberbullying with less effort on the part of the bully (Garinger 2008). This indicates that the time taken by a cyberbully to harass someone is now limited to the speed of a click of the mouse and the spread of threats is no longer limited to the cyberbully and cybervictim (Mitchell 2011). Meanwhile, surveys have reported that in Malaysia, majority of children spend as many as 19 hours per week on average using the Internet (NOFR 2010). According to the report, the rising popularity of the new media and the increasing amount of time children spend online pose great challenges

to parents and or guardians who want to protect their children from the threats and harms of the cyberspace. A new form of bullying commonly referred to as cyberbullying has become a new growing problem on the Internet in nowadays (Carter 2011). Smith et al. (2008) defined cyberbullying as an aggressive, intentional act carried out by a group or individual, using electronic forms of contact, repeatedly and over time against a victim who cannot easily defend him or herself.

Generally, cyberbullying is not that much different from traditional bullying, even though it has changed overtime through the advancement of technology. Bullies nowadays have moved from the school grounds to behind-the-cellular-phone or computer screen (Hinduja & Patchin 2010). According to the researchers, the availability and use of technology by the young people is on the rise and so is the ability to become cyberbullies or cyber predators.

Similarly, Sharif (2008) indicate that rapid advancement of Internet and cellular technologies has opened up a new and infinite space that young people can explore with fewer restrictions, as the digital media offers convenient opportunities to humiliate, bully, or harass someone online. A new variation of bullying has transformed from the physical to the virtual, it becomes insidious when cyberbullying grows into a form of psychological cruelty (Sharif & Hoff 2007) and in some cases will end up with the loss of lives (Abu Bakar 2013).

Types of Cyberbullying

Cyberbullying could take place every hour, every day, throughout the week (24/7) from anywhere (Willard 2007). According to the researcher, cyberbullying can take place through the information technology that (IT) students access every day, including cell phones, text messages, e-mail, instant messaging, social networks, pictures and video clips. Garinger (2008) suggests that since cyberbullying can occur anywhere and at any time, this empowers bullies to a greater degree

compared to traditional bullying. More targets can be harassed through cyberbullying with less effort on the part of the bully (Garinger 2008; Willard 2007). This implies that the time a cyberbully uses to perpetrate an attack is now limited to the speed of a mouse click and the spread of threatening attack is no longer limited to the bully and target (Mitchell 2011). There are eight distinct forms of cyberbullying, namely flaming, harassment, denigration, impersonation, outing and trickery, exclusion, cyberstalking and cyberthreats (Willard 2007). Table 1 contains eight distinct forms of online harassment, which Willard (2007) has outlined.

Trolling, happy slapping and sexually based bullying are other forms of cyberbullying (Ybarra & Mitchell 2004). Online harassment is defined as “threats or other offensive behaviors targeted directly at youth through new technology channels (e.g., Internet, text messaging) or posted online about youth for others to see” (Jones et al. 2012, p. 54). According to Willard (2007) online harassment is the act of aggressive behaviors, including insulting someone online, making nasty or threatening remarks through e-mail or instant messages, spreading unwanted images or video clips of someone and creating offensive pictures or images and posting them online in order to humiliate somebody.

Sexually based cyberbullying is also a common form of cyberbullying (Mitchell & Jones 2011) and it involves sexual solicitations and unwanted exposure to sexual materials such as pornographic pictures, pornographic video clips and sexually suggesting gestures posted online or done with the aim of arousing sexual feelings in the targeted persons or making them feel sexually abused. According to Abu Bakar (2013), online harassment and sexually based cyberbullying are the two most prevalent forms of cyberbullying in Malaysia.

According to Mitchell and Jones (2011), sexual solicitations and exposure to unwanted sexual materials are a wide range of negative cyber experiences affecting children. The researchers defined sexual based cyberbullying as sexual

Table 1. Categories of cyberbullying

Type of Bullying	Description
Flaming	• Short lived argument between protagonists • Extended, heated argument leading to threats of violence • Questionable credibility of threats
Harassment	• Repeated ongoing assault • Usually one sided • Can have multiple protagonists harassing a target • Some protagonists may not even know the target • Could be criminal especially if involving hate or bias • Target is direct recipient of material
Denigration	• Harmful, untrue, or cruel speech about a target • Posted online and/or circulated via email, texting, instant messenger • Target is not direct recipient • Included public postings and sending of digital images (which may have been digitally altered) • May include defamation or invasion of privacy
Impersonation	• Falsely identifying as the target • Posting or sending material that shows the target in a bad light or interferes with target's relationships and friendships • Often a means of the protagonist getting the target in trouble with authorities
Outing and Trickery	• Posting or otherwise circulating images and other personal communications that are embarrassing to the target • Target is not direct recipient of the attack • Images are often sexually suggestive and verge on sexual harassment
Exclusion	• Deliberate exclusion of the target from communications to which he or she was previously privy to • Often occurs when the protagonist convinces multiple people to defriend [unfriend] the target
Cyberstalking	• Including the repeated harassment threats of harm • Can be intimidating, offensive, or involve extortion • Protagonist often lulls target into sense of false security before slowly escalating harassment

Source: Willard (2007).

solicitation, which involves an urge to participate in a sexual activity willingly or unwillingly, made by someone online. Sexually based cyberbullying also involves exposure to sexual materials such

as pictures of naked people or people having sex while surfing the web, doing online search, opening e-mails or instant messages, or opening links in e-mails or instant messages (Abu Bakar 2013; Mitchell & Jones 2011).

Patterns of Cyberbullying

Furthermore, through commonly used channels such as blogs, chat rooms, file-sharing applications, social networking sites and mobile phone messages, children and young people can be threatened, excluded from activities or humiliated by having misleading messages or photos about them posted online (Eek 2009). According to an online survey, these aggressions among youngsters are mostly channeled through social networking sites; 54% of 10,000 participants aged 12 to 17 have reported having experienced cyberbullying on Facebook (Roberts & Samani 2013).

Impact of Cyberbullying

Over the past decade, cyberbullying has become more dangerous and more virulent form of bullying to the extent that insidious forms of victimization now lurk in online community resulting from cyberbullying (Raskauskas & Stoltz 2007). Cyberbullying is one of the biggest threats confronted by young people and it has become a common phenomenon among youth nowadays (Abu Bakar 2013). The effect of cyberbullying is more traumatic than traditional bullying since victims can be bullied 24 hours and 7 days a week from anywhere as long as there is a working Internet connection (Willard 2007). Besides the unlimited accessibility to the victims, cyberbullying also causes psychological abuse to the victims including teasing, intimidation and exclusion in ways that do not exist in traditional bullying (Raskauskas & Stoltz 2007).

Cyberbullies usually seek to exert dominance and power over the victim and inflict psychological torture such as the feeling of low self-esteem, loneliness, poor mental health, peer rejection,

depression, isolation and hopelessness in their victims (Hinduja & Patchin 2010; Pranjic & Bajraktarevic 2010). According to the Center for Disease Control and Prevention [CDC&P] (2010), these negative feelings could even influence children to consider committing a suicide.

Between 2003 and 2012, some 34 cases involving cyberbullicides (committing suicide because of being cyberbullied) were reported, while in 2013 alone, seven cases were reported (Abu Bakar 2013). In 2007, suicide was found to be the third leading cause of death among younger citizens of USA aged 15 to 24 years old (American Foundation for Suicide Prevention [AFSP] 2007). Generally, however, research directly linking cyberbullying to suicide is quite limited (Brunstein-Klomek, Sourander & Gould 2010; Pranjic & Bajraktarevic 2010).

Material and Methods

Cyberbullying dimensions were constructed based on the conceptual framework of this study and a review of extant literature. As mentioned earlier, there are many forms of cyberbullying. However, this study focuses on sexually based cyberbullying and online harassment. For the measurement of sexually based cyberbullying dimension, a scale developed by Mitchell and Jones (2011) and Livingstone et al. (2011) was adopted. While, for the measurement of online harassment dimension (which is related to acts of aggressive behavior online), a scale developed by Livingstone et al. (2011) was adopted with modifications. The original survey scale consists of four items on risk and harm towards child's safety (i.e. bullying, pornography, sexting and meeting strangers online).

This study designed a scale consisting of an inventory of 10 items measured using five-point Likert scale was developed. Two constructs (online harassment and sexually based cyberbullying) were involved, each measured by five items. However, to ensure that the scale is valid in Malaysian context and to help the respondents

(children) understand the questions better, several modifications were made. Respondents were asked to answer the questions based on a 5-point Likert scale: “1 never, 2 seldom, 3 sometimes, 4 often and 5 very often. As part of the modifications, some words/phrases in the original scale that sound sensitive in Malaysian context were changed based on the advised given by Ministry of Education in the country. For example, having sex was replaced with obscene acts and pictures of naked persons was replaced with inappropriate materials in the modified scale. Table 2 shows the complete original and modified items in the scale.

Given that Malay language is the primary national language and English language is treated as a secondary national language, to ensure that the respondents understood the questions very well, the survey instrument was designed in both languages.

Stratified random sampling perspective was adopted to select 378 pupils aged 9 to 16 years old from eight primary schools in Selangor State.

Only children that had experienced cyberbullying in the past 12 months were selected. Paired, coded survey forms were administered to the young participants in order to avoid non-response bias respondents (see Singer 2006).

The respondents were collected into small groups during class hours and were administered with the questionnaire, which they completed under the supervision of their researchers. Either of the children’s parents was also selected to participate as a respondent. A few trained enumerators were employed who read out the questionnaire aloud for the younger respondents aged 9 to 11 years old. Although the average time to complete the questionnaire was about 15 minutes, it took the younger respondents about 30 minutes or more to complete. The questionnaires were retrieved on the same day after the pupils had finished completing them; 61 forms from the children and 49 forms from their parents were recovered.

Factor analysis was used, and the questionnaire was reduced to 378 after data cleansing. Construct

Table 2. Original and adapted cyberbullying scales

Original Items (Qualitative Approach)	Adapted Items (Quantitative Approach)	Dimension and Source
Has anyone ever sent you nude or nearly nude photos or videos of kids who were under the age of 18 that others took?	I have been sent inappropriate messages on the Internet	Sexually-based Cyberbullying (YISS 1, 2 and 3) by Mitchell and Jones (2011)
Did you ever open a message or a link that showed actual people/naked people/people having sex that you did not want?	I have seen inappropriate materials posted where other people could see it on the Internet.	
Did you ever find yourself in a website that showed pictures of naked people/having sex when you did not want to be in it?	I have seen other people perform obscene acts.	
Did anyone on the Internet ask you for sexual information? Like, what your body looks like or sexual things you have done.	I have been asked on the Internet for a photo or video showing my private parts.	
Did anyone on the Internet ever try to get you to talk online about sex when you did not want?	I have been asked to talk about nasty acts with someone on the Internet.	
Nasty or hurtful messages were sent to me.	I have received nasty or hurtful messages.	Online Harassment (EUKidsonline) by Livingstone et al. (2011)
Nasty or hurtful messages about me were passed around or posted where others could see.	Nasty or hurtful messages about me were posted or passed around online where others could see.	
Other nasty or hurtful things on the Internet.	I have received other nasty or hurtful things on the Internet.	
I was threatened on the Internet.	No changes made	
I was left out or excluded from a group or activity on the Internet.	No changes made	

reliability test was then run. The Cronbach alpha coefficient value (α) for the construct for pre-test was $\alpha = .830$ ($n = 63$)¹ while for actual data collection was $\alpha = .75$ ($n = 375$). This indicates that the scale was very reliable (see Salkind 2008). Furthermore, a committee of experts reviewed the cyberbullying items and validated it. Approval from Malaysian ministry of education and the Selangor State department of education was also obtained. Similarly, consent was obtained from authors whose works were adopted in this study.

RESULTS

Respondent's Demographic Profile: Children

The respondents' demographic information shows female respondents were almost twice as many as male respondents. That is, the disparity between the proportion of the male (35.2%) and female (64.8%) pupils was very wide. The respondents were divided into two age groups for data analysis convenience. The first group was the younger group, comprising 155 pupils aged between 9 and 11 years old. The second group was the teenage group, comprising 232 respondents aged between 13 and 16 years old, with exception of 15 year-old age group whose members were sitting a compulsory examination. Generally, the respondents were aged between 9 and 16 years old ($M=12.51$), with more than two-thirds of them (60.1%) aged between 13 and 14 years old. Many (57.6%) of them lived in urban areas (see Table 3).

Respondents' Demographic Profile: Parents

Male parents are slightly fewer than the female parents were, fathers, 44.8% and mothers, 55.2%. Many (56.8%) of them were aged between 42 and 53 years old. Quite many (40.5%) of them were aged 30 to 41 years old, with only few of them

Table 3. Descriptive statistic of respondents' demographic profile (n=375)

Variables	Frequency	Percentage
Gender		
Male	132	35.2
Female	243	64.8
Age		
9	48	12.8
10	49	13.1
11	54	14.4
13	83	22.1
14	90	24.0
16	51	13.6
Ethnicity		
Malay	275	73.3
Chinese	49	13.1
Indian	46	12.3
Others	5	1.3
Religion		
Islam	277	73.9
Buddha	41	10.9
Hindu	42	11.2
Christian	12	3.2
Others	3	.8
Living area		
Urban	216	57.6
Rural	159	42.4

(2.4% and .3%) were aged between 54 to 65 years old and 18 to 29 years old respectively ($M=43$, $SD=5.43$) (see Table 4).

Majority of the parents were Malay (73%) by ethnic identity and Muslims (74%) by religious identity. The parents' educational qualification data indicates that 26.7% of them were Bachelor degree holders, 25.3% held SPM², 20.5% held Diploma and less than 10%. Only very few (1.3%, $n=5$) of them had never been to any school. The monthly disposable income data shows that majority very many (60.8%) of the parents earned RM3,000³ or less.

Table 4. Descriptive statistic of parents' profile (n=375)

Variables	Frequency	Percentage
Gender		
Male	168	44.8
Female	207	55.2
Age		
18 – 29	1	.3
30 – 41	152	40.5
42 – 53	213	56.8
54 – 65	9	2.4
Mean = 42.89		
SD = 5.43		
Ethnicity		
Malay	273	72.8
Chinese	48	12.8
Indian	47	12.5
Others	7	1.9
Religion		
Islam	277	73.9
Buddha	42	11.2
Hindu	41	10.9
Christian	13	3.5
Others	2	.5
Education Level		
Never been to school	5	1.3
Standard 6	11	2.9
SRP/PMR	26	6.9
SPM	95	25.3
STPM	26	6.9
Diploma	77	20.5
Bachelor's	100	26.7
Master's	33	8.8
PhD	2	.5
Monthly Income (RM)		
< 3,000	228	60.8
3,001 – 5,000	76	20.3
>5,001	71	18.9

Prevalence of Cyberbullying Incidences Among the Children

C

This study discovers that the gaps between the children's responses on 'never experienced', 'sometimes experienced' and 'often experienced' were wide. Unpredictably, nearly all (94.1%) of the respondents indicated never having experienced any online threat ($M=1.11$, $SD=.22$). Extremely few of them experienced any cyberbullying as indicated by the overall mean value ($M=1.15$, $SD=.33$) (see Table 5).

Level of Cyberbullying Incidences Among the Children

Overview

Level of cyberbullying refers to the degree of prevalence of the occurrence of cyberbullying incidences. The level is measured using 'Low', 'Moderate' and 'High' scales (see Sharma, Mukherjee, Kumar & Dillon 2005). The scores of the level are derived based on the cut-off mark method, by subtracting the minimum scores from the maximum scores of the responses and are then divided into three levels (see Sa'ari, Wong & Roslan 2005). For example, the overall mean score of cyberbullying is $M=1.15$, $SD=.33$. The level of cyberbullying (see Table 6) is determined marking off this mean score in the level scale, which was designed based on the method mentioned above.

The overall mean value of cyberbullying scale is $M=1.15$, $SD=.33$ (refer to Table 5). Based on the results in Table 6, the level of cyberbullying among Malaysian children is moderate.

DISCUSSION

The results reveal that extremely few (3.9%) of the respondents have ever been threatened online. In addition, only quite a few of them experienced any

Table 5. Percentage of frequency, mean and standard deviation of cyberbullying (n=375)

Items	Likert Scale (% of Frequency)					Mean	SD
	1	2	3	4	5		
I have seen an inappropriate materials posted where other people could see it on the Internet.	79.2	17.7	1.1	1.0	1.0	1.24	.44
I have received nasty or hurtful messages.	82.9	13.7	2.0	1.1	.3	1.22	.40
I have been sent an inappropriate message on the Internet.	84.0	13.5	1.0	1.0	.5	1.21	.40
I have seen other people perform obscene acts.	84.5	12.4	2.0	.5	.5	1.21	.38
I have received other nasty or hurtful things on the Internet.	87.7	10.5	1.5	.5	.3	1.14	.34
I was left out or excluded from a group or activity on the Internet.	89.9	6.2	2.1	1.0	.8	1.12	.34
I have received nasty or hurtful messages about me were passed around or posted where others could see.	90.9	7.3	1.2	.3	.2	1.11	.31
I was threatened on the Internet.	94.1	5.2	.5	.1	.1	1.11	.22
I have been asked on the Internet for a photo or video showing my private parts.	95.8	3.7	.3	.1	.1	1.05	.22
I have been asked to talk about nasty acts with someone on the Internet.	96.0	4.0	-	-	-	1.04	.20
Overall score						1.15	.33

Note: Likert scale: 1 never, 2 seldom, 3 sometimes, 4 often and 5 very often; M: Mean, SD: Standard Deviation

Table 6. Level of cyberbullying (n=375)

Level	Percentage	Overall Mean Value of the Scale	
		Mean	SD
Low (1.00 – 1.04)	10.0	1.15	.33
Moderate (1.05 – 1.23)	80.0		
High (1.24 – 5.00)	10.0		

Note: SD: Standard Deviation.

form of sexually based cyberbullying as follows. Only a few (20.8%) of them had seen inappropriate materials at least once posted online, quite few (16.0%) of them had ever seen inappropriate materials posted online and only extremely few (14.5%) of them had seen people commit obscene acts online.

Similarly, quite few of the children have ever been harassed online as indicated by these percentages. Only 11.1% of them have ever suffered the humiliation of being expelled from an online group; and others have ever ridiculed only 10.1% of them through posting their pictures online where others could see them. Only 4.2% of them have ever been asked by someone to share the picture or video of their private parts online; and only 3.0%

of them have ever been cornered by someone to talk about nasty things online. The results generally suggest that majority of the children have only mildly (moderately) experienced at least one form of cyberbullying or another at some point in time within the past one year. However, majority of them never experienced or got involved in any form of cyberbullying or its related incidence with any serious degree.

This outcome further suggests that either many of the children have been involved in cyberbullying incidences (as cybervictims or cyberbullies) more than once in the past one year but have been unaware of what it constitutes, or (fortunately) have never been exposed to any cyberbullying act. This finding is consistent with the findings of a

survey conducted by the Malaysian cyber security in collaboration with a mobile telecommunication company in 2013 (CSMR 2013).

Therefore, putting the frequency distributions into perspective, this chapter suggests that mild cyberbullying incidences actually occur among Malaysian children but that majority of them have been unaware of what it constitutes. The children may have been passing it for a normal part of online social interaction, especially if the cyberbullying incidences occur on social network sites (see Balakrishnan 2015; Lenhart, Madden, Smith, Purcell, Zickhur & Rainie 2011).

Online harassment has been the most prevalent type of cyberbullying incidence in the country. Wolak et al. (2006) supports this finding empirically. Similarly, Bernama (2012) supported this finding. Although mildly, incidences of online harassment are increasingly becoming prevalent nowadays because children and youth have unprecedented access to information and communication technologies such as smartphones, tablets and Internet connection (Abu Bakar 2013; Hinduja & Patchin 2009).

FUTURE RESEARCH DIRECTIONS

This chapter identifies the absence of a standard cyberbullying measurement scale suitable for Malaysian context as a challenge and limitation. The scale adopted in this study, which were used in the United States (see Mitchell & Jones 2011) and various European countries (see Livingstone et al. 2012) have been suitably reliable in those contexts, and have produced predictable results. However, the findings of the current study have rather been unpredicted.

This study is only limited to quantitative approach. Hence, this chapter recommends that future research should focus on qualitative approach in Malaysian context. Qualitative data can be used to conduct an in-depth observation of the causal relationships between cyberbullying other social factors such as demographics or parent-child

relationships. Therefore, future research should attempt to explore the relationships between cyberbullying and demographic characteristics such as gender and age, as well as parental attachments. It is expected that this approach can generate new and reliable cyberbullying scale for Malaysian society. Qualitative methodologies studies such as in-depth interviews, focus group discussion, content analysis and audio recordings can provide valuable insights and facilitate a clearer interpretation of factors affecting cyberbullying among Malaysian children (see Festl & Quandt 2013).

This study focuses on cyber victim (being a victim of cyberbullying) dimension only because most Malaysian children and adolescents that have got involved in cyberbullying acts have been victims rather than bullies (Balakrishnan 2015). However, to get a comprehensive understanding of cyberbullying, future research should investigate the effect of cyberbullying on other parties involved in the behavior, namely the bully, the bully-victim and the by-stander (audience) (see Camacho, Hassanein & Head 2014).

CONCLUSION

This study has discovered that generally, the level of cyberbullying among Malaysian children is moderate. Specifically, however, incidences of online harassment occur more often than those of sexually based cyberbullying do. The findings also reveal that majority of the children have experienced a cyberbullying incidence at least once in a year, which gives the hints that with time, cyberbullying could become a menace to the innocent Malaysian child online (see Balakrishnan 2015). Nonetheless, the threat posed by cyberbullying to the children has been moderate, which implies that the online safety of majority of Malaysian children is guaranteed at present. Therefore, this study concludes that mild cyberbullying incidences actually do occur among Malaysian children, however, most of the children have been unaware of what it portends.

However, parents are advised to control their children's access to and usage of technologies as well as monitor their behavior. Parents should report any cyberbullying incidence case to the Police. The Malaysian cyber security should maintain the annual cyberbullying survey. Children should be pedagogically instructed about online risks and safety. Therefore, the Ministry of Education is advised to introduce subjects on cyber safety and threats in elementary and high schools. Lastly, the Malaysian Communication and Multimedia Commission (MCMC) should organize seminars on online safety and threats periodically to enlighten parents, adolescents and young adults about the dangers lurking in the online milieu and teach them ways of avoiding or managing involvement in cyberbullying.

REFERENCES

- Abu Bakar, H. S. (2013). Investigating the emergence themes of Cyberbullying phenomenon: A grounded theory approach. In *Educational Media, 2013 IEEE 63rd Annual Conference International Council for Educational Media (ICEM)* (pp. 1-14). Nanyang Technological University.
- AFSP. (2007). *Facts and figures: National Statistics*. Retrieved on February 4, 2014 from http://Ofea9f-b064-4092-bl-135c-3a70delfda.afsp.org/index.cfin?fuseaction_home.viewvpageandpageicl_0S
- Armsden, G. C., & Greenberg, M. T. (1987). The inventory of parent and peer attachment: Individual differences and their relationship to psychological well-being in adolescence. *Journal of Youth and Adolescence*, 16(5), 427-454. doi:10.1007/BF02202939 PMID:24277469
- Azizan, H. (2012). Do you know whom your kids are talking to? *The Star Online*. Retrieved on January 28, 2014 from: <http://www.thestar.com.my/News/Nation/2012/04/29/Do-you-know-who-your-kids-are-talking-to/>
- Balakrishnan, V. (2015). Cyberbullying among young adults in Malaysia: The roles of gender, age and Internet frequency. *Computers in Human Behavior*, 46, 149-157. doi:10.1016/j.chb.2015.01.021
- Banks, R. (1997). Bullying in schools. *ERIC Digest*. Retrieved on July 23, 2014 from: <http://npin.org/library/pre1998/n00416/n00416.html>
- Barak, A., & Sadovsky, Y. (2008). Internet use and personal empowerment of hearing-impaired adolescents. *Computers in Human Behavior*, 24(5), 1802-1815. doi:10.1016/j.chb.2008.02.007
- Bernama. (2012). *Early intervention on cyber security to safeguard young*. Cyber Security Malaysia. Retrieved on August 7, 2014 from http://www.cybersecurity.my/bahasa/knowledge_bank/news/2012/main/detail/2225/index.html
- Berne, S., Frisén, A., Schultze-krumholz, A., Scheithauer, H., Naruskov, K., Luik, P., & Zukauskienė, R. et al. (2013). Aggression and violent behavior cyberbullying assessment instruments: A systematic review. *Aggression and Violent Behavior*, 18(2), 320-334. doi:10.1016/j.avb.2012.11.022
- Berson, I. R., Berson, M. J., & Ferron, J. M. (2002). Emerging risks of violence in the digital age: Lessons for educators from an online study of adolescent girls in the United States. *Journal of School Violence*, 1(2), 51-72. doi:10.1300/J202v01n02_04
- Brunstein-Klomek, A., Sourander, A., & Gould, M. (2010). The association of suicide and bullying in childhood to young adulthood: A review of cross-sectional and longitudinal research findings. *Canadian Journal of Psychiatry*, 55(5), 282-288. PMID:20482954
- Camacho, S., Hassanein, K., & Head, M. (2014). *Understanding the factors that influence the perceived severity of cyber-bullying*. Paper presented at Human-Computer Interaction (HCI) International, Crete, Greece. doi:10.1007/978-3-319-07293-7_13

- Carter, J. M. (2011). *Examining the relationship among physical and psychological health, parent and peer attachment and cyberbullying in adolescents in urban and suburban environments* (PhD Thesis). Wayne State University.
- CDC&P. (2010). *Suicide: Risk and protective factors*. Retrieved on March 17, 2014 from <http://ViolencePrevention/youthviolence/risk-protectivefactors.html>
- CSMR. (2011). Children and teenagers exposed to negative cyber culture. *Bernama News*. Retrieved on January 28, 2014 from http://www.cybersecurity.my/en/knowledge_bank/news/2011/main/detail/2087/index.html
- CSMR. (2014, 2013). *A National Survey Report 2013. Safety net: Growing awareness among Malaysian schoolchildren on staying safe online*. Cyber Security of Malaysia. Retrieved on December 4, 2014 from https://digi.cybersafe.my/files/article/DiGi_Survey_Booklet_COMPLETE.pdf
- Davidson, C. J., & Martellozzo, E. (2008). Protecting vulnerable young people in cyberspace from sexual abuse: Raising awareness and responding globally. *Police Practice and Research*, 9(41), 277–289. doi:10.1080/15614260802349965
- Dholakia, N., Dholakia, R. R., & Kshetri, N. (2004). Global diffusion of the Internet. In H. Bidgoli (Ed.), *The Internet Encyclopedia* (2nd ed.; pp. 38-51). John Wiley & Sons. doi:10.1002/047148296X.tie072
- Eek, T. S. (2009). *A big sister helps students cope with bullying*. UNICEF. Retrieved October 4, 2009, from http://www.unicef.org/infobycountry/malaysia_51549.html
- Eteokleous, N., & Pavlou, V. (2010). Digital natives and technology literate students: Do teachers follow? *Proceedings of the Cyprus Scientific Association of Information and Communication Technologies in Education*, 113-124.
- Festl, R., & Quandt, T. (2013). Social relations and cyberbullying: The influence of individual and structural attributes on victimization and perpetration via the Internet. *Human Communication Research*, 39(1), 101–106. doi:10.1111/j.1468-2958.2012.01442.x
- Finkelhor, D., Mitchell, K., & Wolak, J. (2001). *Highlights of the youth Internet safety survey*. Juvenile Justice Fact Sheet-FS200104. Washington, DC: US Government Printing Office. Retrieved November 20, 2007 from <http://www.unh.edu/ccrc/pdf/jvq/CV46.pdf>
- Finkelhor, D., Mitchell, K. J., & Wolak, J. (2000). *On-line victimization: A report on the nations' youth*. Alexandria, VA: National Centre for Missing and Exploited Children.
- Flicker, S., Maley, O., Ridgley, A., Biscope, S., Lambardo, C., & Skinner, H. A. (2008). Using technology and participatory action research to engage youth in health promotion. *Action Research*, 6(3), 285–303. doi:10.1177/1476750307083711
- Garinger, H. M. (2008). Cyber pox: A look at female adolescent cyber bullying. *Michigan Journal of Counseling*, 35, 24–32.
- Guan, S. A., & Subrahmanyam, K. (2009). Youth Internet use: Risks and opportunities. *Current Opinion in Psychiatry*, 22(4), 351–356. doi:10.1097/YCO.0b013e32832bd7e0 PMID:19387347
- Hinduja, S., & Patchin, J. W. (2007). Offline consequences of online victimization. *Journal of School Violence*, 6(3), 89–112. doi:10.1300/J202v06n03_06
- Hinduja, S., & Patchin, J. W. (2009). *Bullying beyond the schoolyard: Preventing and responding to cyberbullying*. Thousand Oaks, CA: Corwin Press.
- Hinduja, S., & Patchin, J. W. (2010). Bullying, cyberbullying and suicide. *Archives of Suicide Research*, 14(3), 206–221. doi:10.1080/13811118.2010.494133 PMID:20658375

- Jones, L. M., Mitchell, K. J., & Finkelhor, D. (2012). Online harassment in context: Trends from three youth Internet safety surveys (2000, 2005, 2010). *Psychology of Violence*, 3(1), 53–69. doi:10.1037/a0030309
- Kenda, M. (2014). *Internet Society Global Internet Report 2014: Open and sustainable access for all*. Global Internet Report. Retrieved July 2, 2014, from <http://www.Internetsociety.org/map/global-Internet-report/>
- Kraft, E. (2006). *Cyberbullying: A worldwide trend of misusing technology to harass others*. WIT Transactions on Information and Communication Technologies. Retrieved on June 19, 2014 from <http://www.witpress.com/elibrary/wit-transactions-on-information-and-communication-technologies/36/16302>
- Ktoridou, D., Eteokleous, N., & Zachariadou, A. (2012). Internet uses and threats for students: Investigating parents' awareness. In T. Bastiaens & M. Ebner (Eds.), *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2011* (pp. 179-188). Chesapeake, VA: Association for the Advancement of Computing in Education (AACE). Retrieved January 25, 2015 from <http://www.editlib.org/p/37862/>
- Lenhart, A., Madden, M., & Hitlin, P. (2005). *Teens and technology. Youth are leading the transition to a fully wired and mobile nation. 2005*. Pew Internet and American Life Project. Retrieved from: http://www.pewInternet.org/pdfs/PIP_Teens_Tech_July2005web.pdf
- Lenhart, A., Madden, M., Smith, A., Purcell, K., Zickhur, A., & Rainie, L. (2011). *Teens, kindness and the cruelty on social network sites*. Pew Internet and American Life Project.
- Livingstone, S. (2012). Critical reflections on the benefits of ICT in education. *Oxford Review of Education*, 38(1), 9–24. doi:10.1080/03054985.2011.577938
- Livingstone, S., Bober, M., & Helsper, E. (2005). *Inequalities and the digital divide in children and young people's Internet use: Findings from the UK Children Go Online project*. London, UK: The London School of Economics and Political Science.
- Livingstone, S., & Gorzig, A. (2014). When adolescents receive sexual messages on the Internet: Explaining experiences of risk and harm. *Computers in Human Behavior*, 33, 8–15. doi:10.1016/j.chb.2013.12.021
- Livingstone, S., & Haddon, L. (2009). Risky experiences for children online: Charting European research on children and the Internet. *Children & Society*, 22(4), 314–323. doi:10.1111/j.1099-0860.2008.00157.x
- Livingstone, S., Haddon, L., & Görzig, A. (2012). *Children, risk and safety the Internet: Research and policy challenges in comparative perspective*. Bristol: Policy Press. doi:10.1332/policypress/9781847428837.001.0001
- Livingstone, S., Haddon, L., Gorzig, A., & Olafsson, A. (2011). *Risks and safety on the Internet: The perspective of European children: Full findings*. London, UK: EU Kids Online.
- McKenna, P. (2010). Rise of cyberbullying. *Reader's Digest Asia*. Retrieved on April 3, 2014 from: http://origin-www.rdasia.com/rise_of_cyber_bullying/
- MCMC. (2014). *Internet users survey 2014*. Retrieved on November 13, 2014 from <http://www.skmm.gov.my/Resources/Statistics/Internet-users-survey/Hand-Phone-Users-Survey-2014>
- Mitchell, K. J., & Jones, L. M. (2011). *Youth Internet Safety Study (YISS): Methodology Report*. Durham, NH: Crimes against Children Research Center University of New Hampshire. Retrieved on May 2, 2014 from: [http://www.unh.edu/ccrc/pdf/YISS%20Methods%20Report%20formatted%20\(final\).pdf](http://www.unh.edu/ccrc/pdf/YISS%20Methods%20Report%20formatted%20(final).pdf)

- Mitchell, M. S. (2011). *Cyberbullying and academic achievement: research into the rates of incidence, knowledge of consequences, and behavioral patterns of cyberbullying* (PhD Thesis). University of Connecticut.
- Murugesan, S. (2009). *Social issues and web 2.0: A closer look at culture in e-learning. Handbook of Research on Web 2.0, 3.0, and X.0: Technologies, Business, and Social Applications, Information Science*. New York, NY: Prentice Hall.
- Myers, J. J., McCaw, D. S., & Hemphill, L. S. (2011). *Responding to cyber bullying. An action tool for school leader*. Corwin Press, a Sage Company. doi:10.4135/9781483350516
- Nansel, T., Overpeck, M., Pilla, R., Simons-Morton, B., & Scheidt, P. (2001). Bullying behaviors among U.S. youth: Prevalence and association with psychosocial adjustment. *Journal of the American Medical Association*, 285(16), 2094–2100. doi:10.1001/jama.285.16.2094 PMID:11311098
- Neging, P., Musa, R., & Abdul Wahab, R. (2013). The determinants and outcomes of pathological Internet use (PIU) among urban millennial teens: A theoretical framework. *World Academy of Science, Engineering and Technology*, 74, 2-21.
- Nik Anis, N. M., Abdul Rahim, R., & Lim, Y. (2012). Najib: Cyberbullying a serious threat to kids. *The Star Online*. Retrieved on January 28, 2014 from <http://www.thestar.com.my/News/Nation/2012/10/10/Najib-Cyber-bullying-a-serious-threat-to-kids/>
- NOFR. (2010). *Global insights into family life online*. Retrieved from http://us.norton.com/content/en/us/home_homeoffice/media/pdf/nofr/Norton_Family-ReportUSA_June9.pdf
- Nordahl, J. K., Poole, A., Stanton, L., Walden, L. M., & Beran, T. N. (2008). A review of school-based bullying interventions. *Democracy & Education*, 8(1), 16–20.
- Olweus, D. (2002). A profile of bullying at school. *Educational Leadership*, 60, 12–17.
- Online Bullying Is a Top Concern Among Youth. (2012, June 25). *Microsoft global youth online behavior survey*. Microsoft News Center. Retrieved from <http://www.microsoft.com/en-us/news/press/2012/jun12/06-25youthbehaviorpr.aspx>
- PI&ALP. (2013). *The Demographic of Media Social Users-2012*. Washington, DC: Pew Internet and American Life Project. Retrieved from <http://pewInternet.org/Reports/2013/Social-media-users.aspx>
- Pranjic, N., & Bajraktarevic, A. (2010). Depression and suicide Ideation among secondary school adolescents involved in school bullying. *Primary Health Care Research and Development*, 1(1), 349–362. doi:10.1017/S1463423610000307
- Prensky, M. (2001). Digital natives, digital immigrants part 1. *On the Horizon*, 9(5), 1–6. doi:10.1108/10748120110424816
- Raskauskas, J., & Stoltz, A. D. (2007). Involvement in traditional and electronic bullying among adolescents. *Developmental Psychology*, 43(3), 564–575. doi:10.1037/0012-1649.43.3.564 PMID:17484571
- Roberts, L., & Samani, R. (2013). *Digital deception: The online behavior of teens*. McAfee: An Intel Company. Retrieved on May 5, 2014 from <http://oxfordhandbooks.com/view/10.1093/oxfordhb/9780199561803.001.0001/oxfordhb-9780199561803-e-019>
- Sa'ari, J. R., Wong, S. I., & Roslan, S. (2005). Attitudes and perceived information technology competency among teachers. *Malaysian Online Journal of Educational Technology*, 2(3), 70–77.
- Shariff, S. (2008). *Cyberbullying: Issues and solutions for the school, the classroom and the home*. New York: Routledge.
- Shariff, S., & Hoff, D. (2007). Cyber bullying: Clarifying legal boundaries for school supervision in cyberspace. *International Journal of Cyber Criminology*, 1(1), 76–118.

- Sharma, S., Mukherjee, S., Kumar, A., & Dillon, W. R. (2005). A simulation study to investigate the use of cut-off values for assessing model fit in covariance structure models. *Journal of Business Research*, 58(7), 935–943. doi:10.1016/j.jbusres.2003.10.007
- Smith, P. K., Mahdavi, J., Carvalho, M., Fisher, S., Russell, S., & Tippett, N. (2008). Cyberbullying: Its nature and impact in secondary school students. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 49(4), 376–385. doi:10.1111/j.1469-7610.2007.01846.x PMID:18363945
- Smokowski, P. R., & Kopasz, K. H. (2005). Bullying in school: An overview of types, effects, family characteristics and intervention strategies. *Children & Schools*, 27(2), 101–110. doi:10.1093/cs/27.2.101
- Subrahmanyam, K. & Greenfield, P. (2008). Online communication and adolescent relationships. *The Future of Children*, 8(1), 119–146.
- Tokunaga, R. S. (2010). Following you home from school: A critical review and synthesis of research on cyberbullying victimization. *Computers in Human Behavior*, 26(3), 277–287. doi:10.1016/j.chb.2009.11.014
- USC ASCDF. (2005). *The digital future report 2005: Surveying the digital future: Year five*. Los Angeles, CA: University of Southern California.
- USC ASCDF. (2008). *2008 Digital future report*. Los Angeles, CA: University of Southern California.
- Van den Eijnden, R., Spijkerman, R., Vermulst, A., Rooij, T., & Engels, R. (2010). Compulsive Internet use among adolescents: Bidirectional parent-child relationships. *Journal of Abnormal Child Psychology*, 38(1), 77–89. doi:10.1007/s10802-009-9347-8 PMID:19728076
- Vignoli, E., & Mallet, P. (2004). Validation of a brief measure of adolescents parent attachment based on Armsden and Greenbergs three-dimension model. *Revue Europeene de Psychologie Appliquee*, 54(4), 251–260. doi:10.1016/j.erap.2004.04.003
- Vivolo-Kantor, A. M., Martell, B. N., Holland, K. M., & Westby, R. (2014). A systematic review and content analysis of bullying and cyberbullying measurement strategies. *Aggression and Violent Behavior*, 19(4), 424–434. doi:10.1016/j.avb.2014.06.008 PMID:26752229
- Ward, W. (2011). Technology correspondent. *BBC News*. Retrieved 24 January 2014 from <http://www.how-the-web-went-worldwide>
- Willard, N. (2007). *Cyberbullying and cyber-threats: Responding to the challenge of online social aggression, threats and distress*. Research Press.
- WIP. (2013). International Report (5th ed.). Prentice Hall.
- Wolak, J., Mitchell, K., & Finkelhor, D. (2006). Online victimization of youth: 5 years later. National Center for Missing and Exploited Children, 7, 6–25.
- Yusuf, S., Osman, M. N., Hassan, M. S. H., & Teimoury, M. (2014). Parents influence on childrens online usage. *Procedia: Social and Behavioral Sciences*, 155, 81–86. doi:10.1016/j.sbspro.2014.10.260
- Zamaria, C., & Fletcher, F. (2007). *Canada online: Yeartwo highlights, 2007*. Canada Internet Project.
- Zickhur, A., & Smith, A. (2013). Digital differences. Washington, DC: Pew Internet and American Life Project. Retrieved from <http://www.pewInternet.org/2012/04/13/digital-differences/>

ADDITIONAL READING

Abd Jalil, S., Abd Jalil, H., & Abdul Latiff, A. (2010). Social media and our youths today: Exploring the impact of social media on Malaysian youths. International Conference on Communications and Media (ICCM), 18 - 20 June 2010, Bayview Hotel, Malacca, Malaysia.

Abdullah, A. D. A., & Chan, C. M. (2016, September). Social media use among teenagers in Brunei Darussalam. In *Conference on e-Business, e-Services and e-Society* (pp. 195-205). Berlin, Germany: Springer International Publishing. doi:10.1007/978-3-319-45234-0_18

Abu Bakar, H. S. (2015). The emergence themes of cyberbullying among adolescences. *International Journal of Adolescence and Youth*, 20(4), 393–406. doi:10.1080/02673843.2014.992027

Ang, R. P. (2016). Cyberbullying: Its Prevention and Intervention Strategies. In *Child Safety, Welfare and Well-being* (pp. 25–38). New Delhi, India: Springer India. doi:10.1007/978-81-322-2425-9_3

Ang, R. P., Tan, K. A., & Mansor, A. T. (2010). Normative beliefs about aggression as a mediator of narcissistic exploitativeness and cyberbullying. *Journal of Interpersonal Violence*, 26(13), 2619–2634. doi:10.1177/0886260510388286 PMID:21156699

Choi, L. J., Syawal, N. B. A. R. M., & Narawi, B. (2016). Bullying Cases among students: School principals' self-efficacy. *Academy of Social Science Journal*, 1(3), 38–53.

DAntona, R., Kevorkian, M., & Russom, A. (2010). Sexting, texting, cyberbullying and keeping youth safe online. *Journal of Social Sciences*, 6(4), 523–528. doi:10.3844/jssp.2010.523.528

Daud, A., Omar, S. Z., Hassan, M. S., Bolong, J., & Teimouri, M. (2014). Parental mediation of children's positive use of the Internet. *Life Science Journal*, 11(8), 360–3691.

Mustaffa, N., Ibrahim, F., Mahmud, W. A. W., Ahmad, F., Kee, C. P., & Mahbob, M. H. (2011). Diffusion of innovations: The adoption of Facebook among youth in Malaysia. *The Public Sector Innovation Journal*, 16(3), 1–15.

KEY TERMS AND DEFINITIONS

Adolescents: This term refers to growing up children, aged between 13 and 17 years old.

Children: This is an umbrella term that refers to pupils (boys and girls), usually aged between six and 16 years old.

Cyber Predator: This term refers to online bullies, i.e., individuals who bully cyber victims.

Cyber Victim: This term refers to a boy or a girl who has been affected by cyberbullying.

Cyberbullying: This refers to intentional inflicting of psychological torture and hurling of abusive words online (in text, graphics or audio-visual formats) by a person (adult or minor) targeted at a minor or a group of minors using the Internet aimed at scaring them.

Cyberspace: This term refers to the online environment where cyberbullying occurs, i.e., the Internet.

Internet: This term refers to the global, wireless networked digital technology service that connects one computer device to another and facilitates sharing of information and effective communication between users smoothly and promptly.

Online: This term is often used as a synonym of Internet. It also refers to an activity performed on Internet-enabled platform or application such as the World Wide Web (www).

Online Risk Behavior: This refers to a particular social behavior that portends threats exhibited by persons (adult or children) online targeted at other online users, usually minors.

Online Safety: This term refers to web or online condition that guarantees a bullying-free, risk-free and threat-free surfing (use) of the Internet, especially by children.

Teenagers: This term refers to growing up children between the ages of 13 and 19 years old.

Youth: This term refers to certain age-brackets of young adults, usually between the ages of 17 and 30 years old.

ENDNOTES

¹ n: A sample size

² SPM: Sijil Pelajaran Malaysia (Malaysia Education Certificate)

³ RM: Ringgit Malaysia (the Malaysian unit of currency)

The Nature, Extent, Causes, and Consequences of Cyberbullying

C

Michelle F. Wright

Masaryk University, Czech Republic

INTRODUCTION

Children and adolescents spend a great deal of time using and interacting through electronic technologies, including cell phones, gaming consoles, and the Internet. Some of their engagement with electronic technologies involves many benefits, such as the ability to engage in quick communication with just about anyone, including friends and family, and having access to a multitude of rich information. Despite the many opportunities afforded by electronic technologies, many children and adolescents are exposed to risks. One risk associated with electronic technology usage among adolescents and children is cyberbullying. Cyberbullying occurs through electronic technologies, including gaming consoles, email, instant messaging, chatrooms, social media, and text messages via mobile phones. The literature in this chapter draws on research from various disciplines, including communication, computer science, education, media studies, psychology, social work, and sociology. Furthermore, the literature involves a variety of different research designs, including cross-sectional and longitudinal methodologies as well as qualitative and quantitative designs. The chapter is organized into the following eight sections:

1. The first section provides a background of the nature of cyberbullying by focusing on defining it and describing the technologies used to target others and the features of anonymity as applied to cyberbullying.
2. The second section describes the extent of cyberbullying by focusing on the prevalence rates of children's and adolescents' involvement in cyberbullying.
3. The third section describes the individual characteristics and risks associated with children's and adolescents' involvement in cyberbullying.
4. The fourth section details the role of parents and families in children's and adolescents' cyberbullying perpetration and victimization.
5. Similar to the fourth section, the fifth section explains the role of peers and school in children's and adolescents' cyberbullying involvement.
6. The purpose of the sixth section is to review literature on the psychological, behavioral, and academic consequences associated with cyberbullying involvement among children and adolescents.
7. The seventh section discusses future research directions.
8. The final section provides concluding remarks on cyberbullying.

BACKGROUND

As deliberately embarrassing or intimidating, cyberbullying involves the usage of modern electronic technologies to harm others using hostile and repetitive behaviors (Wolak, Mitchell, & Finkelhor, 2007; Ybarra, West, & Leaf, 2007). Cyberbullying is described as an extension of traditional face-to-face bullying, and it also includes elements of an imbalance of power between the bully and the victim as well as the incorporation

DOI: 10.4018/978-1-5225-2255-3.ch150

of a technological component (Olweus, 1999). These behaviors are repetitive, deliberate, and intentionally carried out by bullies with malicious intent. Similar to traditional face-to-face bullying, cyberbullying can also include behaviors with a face-to-face equivalent, such as spreading a rumor about a victim, harassment, physical threats, social exclusion, humiliation, gossiping about a victim to get others not to like the victim, and/or verbal insults. There are also physical forms of cyberbullying, like in traditional face-to-face bullying, which can include hacking. It can include making anonymous phone calls, theft of identity information by pretending to be someone else, distributing explicit videos via various websites, and harassment using instant messenger, social networking websites, and text messages through mobile phones (Wolak et al., 2007; Ybarra & Mitchell, 2004). Other forms of cyberbullying involve happy slapping and flaming (Smith et al., 2008). Furthermore, cyberbullying can involve using various electronic technologies, instant messaging tools and social networking websites.

Researchers have attempted to understand why children and adolescents engage in cyberbullying. One proposal is that new electronic technologies allow cyberbullies to hide their identities, furthering the power differential between the cyberbully and the cybervictim (Wright, 2013; Ybarra et al., 2007). Many children and adolescents who engage in cyberbullying choose to remain anonymous while perpetrating cyberbullying (Dehue, Bolman, Vollink, & Pouwelse, 2012; Wright, 2014a; Ybarra & Mitchell, 2004). Another proposal relates to new electronic technologies' ability to allow the cyberbully to perpetrate frequent, repeated, and prolonged harassment (Wright, 2014a; Ybarra & Mitchell, 2004). Wright (2013; 2014a) found that when individuals felt more confidence in their ability to remain anonymous while using electronic technologies they engaged in more cyberbullying, especially when they could perpetrate cyberbullying anonymously.

THE EXTENT OF CYBERBULLYING

When an awareness of cyberbullying was gained by researchers, there were many studies conducted to understand how frequently children and adolescents were involved in these behaviors. In 2007, Wolak and colleagues (2007) conducted one of the earliest studies on cyberbullying in the United States. They found that 50% of children and adolescents in their sample were victimized by cyberbullying. Lower prevalence rates have been found in other studies conducted on cyberbullying involvement in the United States. For instance, Kowalski and Limber (2007) found that 11% of their 3,767 sample from middle school (aged 11-14) were victimized by cyberbullying, 4% had cyberbullied others, and 7% were classified as cyberbullies-cybervictims. Similar estimates were found in a study of cyberbullying by Patchin and Hinduja (2006). Among their sample, 29% of participants reported that they were cybervictims, while 47% reported to have witnessed cyberbullying at least once. More recent research revealed that 4.9% of children and adolescents in their sample ($N = 4,441$, 6-12th grade) were cybervictims within the past 30 days (Hinduja & Patchin, 2013).

Increasing evidence indicates that cyberbullying is not localized to one country. Instead, research has revealed that cyberbullying is a global concern. In one study, Cappadocia et al. (2013) found that out of their 1,972 sample of Canadian 10th graders, 2.1% perpetrated cyberbullying, 1.9% were cybervictims, and 0.6% were cyberbullies-cybervictims. Rates of cyberbullying involvement have also been examined in Australia as well. In this research, Campbell and colleagues (2012) found that 4.5% of their sample ($N = 3,112$; grades 6-12th) were victims of cyberbullying. Research has also documented cyberbullying involvement among European children and adolescents. Using a large sample of Swedish adolescents ($N = 22,544$; ages 15-18), Laftman et al. (2013)

found that 5% of their sample were classified as cybervictims, 4% as cyberbullies, and 2% as cyberbullies-cybervictims. Rates were similar for cyber victimization in Ireland and Belgium such that 6% of Irish adolescents ($N = 876$; ages 12–17; Corcoran, Connolly, & O'Moore, 2012) and 6.3% of Belgian adolescents ($N = 1,042$, M age = 15.47; Heirman & Walrave, 2012) were classified as cybervictims. Prevalence rates of cyber victimization were much higher among Italian adolescents. For instance, Brighi and colleagues (2012) found that 12.5% of the 2,326 Italian adolescents in their sample were cybervictims.

Although slower to develop, some research has focused on cyberbullying involvement rates among children and adolescents in Asia. In one study, Huang and Chou (2010) found that 63.4% of their sample ($N = 545$) witnessed cyberbullying, 34.9% were cybervictims, and 20.4% were cyberbullies. Jang and colleagues (2014) examined rates of cyberbullying involvement among Korean adolescents, with findings revealing that 43% of adolescents in their sample ($N = 3,238$) were cybervictims or perpetrators of cyberbullying. Among a sample of Chinese adolescents ($N = 1,438$), Zhou et al. (2013) found that 34.8% were cyberbullies and 56.9% were cybervictims. Rates were similar among a sample of 1,676 13 through 17 year olds in Singapore such that 59.4% were cybervictims through Facebook and 56.9% were cyberbullies through Facebook (Kwan & Skoric, 2013).

Comparisons of prevalence rates are difficult as the researchers relied on different samples and measurement of cyberbullying. Regardless of the shortcomings associated with studies on prevalence rates of cyberbullying involvement, it is clear that cyberbullying is a concern for children and adolescents around the world. Such a consideration highlights the importance of studying cyberbullying by attempting to understand the risk factors related to children's and adolescents' involvement in cyberbullying as victims, perpetrators, and bystanders.

CHARACTERISTICS AND RISK FACTORS ASSOCIATED WITH CYBERBULLYING INVOLVEMENT

C

After much of the research on the prevalence rates of cyberbullying involvement, many researchers directed their attention to understanding the characteristics and risk factors related to children's and adolescents' involvement in cyberbullying. Most of these earlier investigations focused on the role of demographic variables in cyberbullying perpetration and victimization. Age has been examined as a factor related to cyberbullying among children and adolescents. One study revealed that younger adolescents were at a greater risk of cyber victimization, while older adolescents were most frequently classified as cyberbullies (Ayas & Horzum, 2012). On the other hand, Wade and Beran (2011) found that 9th graders in their sample experienced the greatest risk of cyberbullying involvement when compared to adolescents in middle school. Such conflicting findings from studies focused on age indicated that this variable might be an inconsistent predictor of cyberbullying involvement. Gender is also another inconsistent predictor of cyberbullying involvement among children and adolescents. In particular, some researchers (Ybarra et al., 2007) have found that boys self-reported more cyberbullying perpetration when compared to girls, with some research revealing that girls experienced more cyber victimization in comparison to boys (Hinduja & Patchin, 2013; Kowalski & Limber, 2007). In contrast, some researchers (e.g., Dehue et al., 2012) have found that girls were more often the perpetrators of cyberbullying in comparison to boys, while boys were much more likely to be victims of cyberbullying (e.g., Huang & Chou, 2010). On the other hand, some researchers have found no gender differences in children's and adolescents' cyberbullying involvement (e.g., Wright & Li, 2013).

Some research has been conducted on other characteristics and risk factors associated with

cyberbullying involvement among children and adolescents. One risk factor for cyberbullying involvement is internet usage. In this research, cybervictims and cyberbullies typically report higher rates of chatting, emailing, blogging, gaming, and sending instant messages when compared to uninvolved children and adolescents (Smith et al., 2008; Ybarra & Mitchell, 2004). To explain the association between greater internet usage and cyberbullying involvement, Ybarra and colleagues (2007) proposed that cybervictims and cyberbullies are more likely to disclose personal information online, which worsens their exposure to online risks, such as cyberbullying. Thus, frequent internet usage increases children's and adolescents' willingness to disclose private information through electronic technologies. Higher normative beliefs concerning both traditional face-to-face bullying and cyberbullying were also related to cyberbullying perpetration among adolescents (Burton, Florell, & Wygant, 2013). Such findings indicate that cyberbullies view bullying behaviors as normal, which contributes to their perpetration of these behaviors. Defined as children's and adolescents' beliefs that bullying is unacceptable and that defending victims is valued, provictim attitudes are also related to cyberbullying perpetration. In particular, Elledge and colleagues (2013) found that holding lower levels of provictim attitudes were associated with greater cyberbullying perpetration. Other research has revealed that low self-control, less empathy, and moral disengagement are also contributors to cyberbullying perpetration (Jang et al., 2014; Robson & Witenberg, 2013).

Few studies have focused on the longitudinal risk factors related to cyberbullying involvement. Utilizing a longitudinal design, Fanti and colleagues' (2012) study revealed that media violence exposure, along with callous and unemotional traits, predicted cyberbullying perpetration one year later. Media violence exposure was also a risk factor associated with cyber victimization. Current research has indicated that perceived stress from parents, peers, and academics also increases

cyberbullying perpetration one year later among adolescents (Wright, in press). Considering each of the studies examined in this section, it is clear that there are a variety of individual characteristics and risk factors which make children and adolescents vulnerable to cyberbullying involvement. Other research has focused on the role of other individuals in children's and adolescents' involvement in cyberbullying. Parents and families have also been examined as having a role in children's and adolescents' cyberbullying perpetration and victimization.

THE ROLE OF PARENTS AND FAMILIES

Parental monitoring and parenting styles have been examined in relation to children's and adolescents' involvement in cyberbullying. Mason (2008) found that 30% of adolescents in his sample utilized the Internet for three hours or more daily, with 50% of these adolescents reporting that their parents monitored their activities. Even when parents do monitor their children's online activities, they are usually not effective because some parents lack technological skills and many do not follow-up on the strategies that they have implemented. In addition, many parents in one sample reported that they were not certain of what their children did online and they were less certain on how to discuss the topic of online activities with their children. Because parents are not able to engage in an active and ongoing dialogue with their children about online activities, and considering that experiencing online risks is almost inevitable for children, it is not likely that some children have the skills necessary to effectively deal with problematic situations that arise online.

Another line of research focuses on the associations between family characteristics and children's and adolescents' cyberbullying involvement. In this research, Ybarra and Mitchell (2004) revealed that lower caregiver monitoring and poor emotional bonds with caregivers were risk factors

associated with cyberbullying. They did not find evidence that family income, parental education, and caregiver status (single or married) were related to cyberbullying perpetration or cyber victimization. On the other hand, Aricak and colleagues' (2012) study revealed that parental unemployed increased children's and adolescents' risk of being involved in cyberbullying. Neglectful parenting is also a risk factor associated with cyberbullying perpetration and cyber victimization. In particular, Dehue and colleagues (2012) found that victims and perpetrators of cyberbullying were more likely to report that their parents engaged in neglectful parenting styles when compared to uninvolved children and adolescents. Similarly, their study also indicated that authoritarian parenting style increased children's and adolescents' risk of cyber victimization. More attention should be given to the role of parents and families in children's and adolescents' cyberbullying involvement.

THE ROLE OF SCHOOLS AND PEERS

Once children enter formal schooling, they spend most of their time in such settings until they graduate from high school or college. Consequently, some researchers have directed their attention to the linkages of peers and schools to children and adolescents' bullying involvement. Although schools' role in cyberbullying involvement is a bit complicated, some research has focused on how schools, particularly teachers and school climate, impact children's and adolescents' cyberbullying involvement. In one study, many teachers perceived the impact of overt forms of traditional face-to-face bullying as much worse than covert forms of bullying, like relational aggression and some forms of cyberbullying (Sahin, 2010). Cyberbullying and schools' involvement in these incidences is further complicated because many situations involving cyberbullying occur outside of school through electronic technologies. This makes it difficult for teachers to know about such

incidences and for them to intervene or provide support to the victim. Additionally, many teachers have not been properly trained on how to deal with and recognize cyberbully behaviors as well as the risk factors associated with these behaviors, furthering hindering schools' ability to deal with these behaviors. Cassidy and colleagues (2012a) found that some of the Canadian teachers in their sample were unfamiliar with some of the newer electronic technologies. Their lack of knowledge made it difficult for them to deal with cyberbullying, even if a victim came to them for help.

A negative school climate and lower school commitment increase cyberbullying perpetration at schools (Ybarra & Mitchell, 2004). Another problem inhibiting schools' ability to deal with cyberbullying is that there are few implemented policies and programs, even when teachers were concerned with these behaviors. It is important that teachers are concerned about cyberbullying and understand how to prevent it. Teachers' motivations for learning about cyberbullying and their self-efficacy for believing in their ability to deal with these behaviors decreased from elementary school to middle school.

Another feature of the school environment are children's and adolescents' peers. Festl and colleagues (2013) examined the role of peers in adolescents' cyberbullying involvement. Their findings revealed that one of the best predictors of cyberbullying involvement were classrooms in which the highest incidences of cyberbullying occurred. In addition, adolescents who believed that their friends perpetrated cyberbullying were more likely to perpetrate these behaviors as well (Hinduja & Patchin, 2013). Some research has focused on the quality of peer relationships to understand children's and adolescents' involvement in cyberbullying. In particular, Burton and colleagues (2013) found that higher levels of peer attachment were related negatively to cyberbullying perpetration and cyber victimization. In addition, being rejected by one's peers was also related to cyberbullying, even after controlling for face-to-face victimization (Wright & Li, 2013).

Research has also revealed that cyberbullying perpetration might be used to boost adolescents' social status among their peer group. For example, Wright's (2014b) study suggested that higher levels of perceived popularity, a reputational type of popularity usually linked to traditional face-to-face bullying perpetration, increased adolescents' cyberbullying perpetration six months later.

PSYCHOLOGICAL, BEHAVIORAL, AND ACADEMIC DIFFICULTIES

Researchers, educators, and parents are particularly concerned about the psychological, behavioral, and academic difficulties associated with adolescents experiencing and/or perpetrating cyberbullying. Perpetrators and victims of cyberbullying reported internalizing difficulties, such as depression and anxiety, and externalizing difficulties, including drug and alcohol use as well as violence (Patchin & Hinduja, 2006; Ybarra et al., 2007). Cyberbullying involvement is also linked to problems in school. In one study, Wright (2015) examined the longitudinal associations between school functioning and cyberbullying involvement. She found that cyberbullying perpetration and cyber victimization were each associated with poor academic performance, absenteeism, and school behavioral problems (e.g., classroom disruptions) one year later.

The bulk of the research on the psychological, academic, and behavioral difficulties associated with children's and adolescents' cyberbullying involvement do not control for traditional face-to-face bullying involvement. Controlling for these variables is important as there are strong correlations between cyberbullying involvement and traditional face-to-face bullying involvement. To this end, Bonanno and colleagues (2013) examined cyberbullying perpetration and cyber victimization in relation to depressive symptoms and suicidal ideation. After including face-to-face bullying and victimization in the same model, they

still found associations between these variables. However, Brighi and colleagues (2012) did not find a relationship between cyber victimization and loneliness or self-esteem, after accounting for traditional face-to-face bullying victimization.

FURTHER RESEARCH DIRECTIONS

There are a variety of gaps in the cyberbullying literature, which need to be investigated through future research. One recommendation is that follow-up research needs to focus on understanding more about the concept of anonymity as applied to cyberbullying. Anonymity is not only a component of cyberbullying, but anonymous forms of traditional face-to-face bullying can occur as well. For example, a nasty note can be slipped into a peer's locker anonymously. Thus, research should be conducted to understand differences in children's and adolescents' conceptualizations of anonymous acts in the cyber context and the offline world. There are few longitudinal studies conducted on cyberbullying involvement. In future research, longitudinal designs should be utilized to examine both traditional face-to-face bullying and cyberbullying involvement from childhood into late adolescence or even emerging adulthood. Investigations focused on the developmental trajectory of bullying involvement in both social contexts are important as much of the research on cyberbullying does not take into account previous levels of traditional face-to-face bullying, although these variables are strongly correlated (Wright & Li, 2013; Ybarra et al., 2007).

CONCLUSION

This chapter reviewed the nature, extent, causes, and consequences associated with cyberbullying. Researchers are beginning to understand that attacks which occur in cyberspace might be just as powerful as face-to-face attacks. Our communities

need to recognize this threat and devote time to dealing with it, as it affects everyone. Cyberbullying will probably not disappear anytime soon, due to our increasing reliance of electronic technologies. Therefore, we all must stand united in our efforts to reduce or prevent cyberbullying.

REFERENCES

- Aricak, T., Siyahhan, S., Uzunhasanoglu, A., Saribeyoglu, S., Ciplak, S., Yilmaz, N., & Memmedov, C. (2008). Cyberbullying among Turkish adolescents. *Cyberpsychology & Behavior*, 11(3), 253–261. doi:10.1089/cpb.2007.0016 PMID:18537493
- Ayas, T., & Horzum, M. B. (2010). *Cyberbully / victim scale development study*. Retrieved from: <http://www.akademikbakis.org>
- Bonanno, R. A., & Hymel, S. (2013). Cyber bullying and internalizing difficulties: Above and beyond the impact of traditional forms of bullying. *Journal of Youth and Adolescence*, 42(5), 685–697. doi:10.1007/s10964-013-9937-1 PMID:23512485
- Brighi, A., Guarini, A., Melotti, G., Galli, S., & Genta, M. L. (2012). Predictors of victimisation across direct bullying, indirect bullying and cyberbullying. *Emotional & Behavioural Difficulties*, 17(3-4), 375–388. doi:10.1080/13632752.2012.704684
- Burton, K. A., Florell, D., & Wygant, D. B. (2013). The role of peer attachment and normative beliefs about aggression on traditional bullying and cyberbullying. *Psychology in the Schools*, 50(2), 103–114. doi:10.1002/pits.21663
- Campbell, M., Spears, B., Slee, P. H., Butler, D., & Kift, S. (2012). Victims perceptions of traditional and cyberbullying, and the psychosocial correlates of their victimisation. *Emotional & Behavioural Difficulties*, 17(3-4), 389–401. doi:10.1080/13632752.2012.704316
- Cappadocia, M. C., Craig, W. M., & Pepler, D. (2013). Cyberbullying: Prevalence, stability and risk factors during adolescence. *Canadian Journal of School Psychology*, 28, 171–192.
- Cassidy, W., Brown, K., & Jackson, M. (2012a). Making kind cool: Parents suggestions for preventing cyber bullying and fostering cyber kindness. *Journal of Educational Computing Research*, 46(4), 415–436. doi:10.2190/EC.46.4.f
- Corcoran, L., Connolly, I., & O'Moore, M. (2012). Cyberbullying in Irish schools: An investigation of personality and self-concept. *The Irish Journal of Psychology*, 33(4), 153–165. doi:10.1080/03033910.2012.677995
- Dehue, F., Bolman, C., Vollink, T., & Pouwelse, M. (2012). Cyberbullying and traditional bullying in relation to adolescents' perceptions of parenting. *Journal of Cyber Therapy and Rehabilitation*, 5, 25–34.
- Elledge, L. C., Williford, A., Boulton, A. J., De-Paolis, K. J., Little, T. D., & Salmivalli, C. (2013). Individual and contextual predictors of cyberbullying: The influence of childrens provictim attitudes and teachers ability to intervene. *Journal of Youth and Adolescence*, 42(5), 698–710. doi:10.1007/s10964-013-9920-x PMID:23371005
- Fanti, K. A., Demetriou, A. G., & Hawa, V. V. (2012). A longitudinal study of cyberbullying: Examining risk and protective factors. *European Journal of Developmental Psychology*, 8(2), 168–181. doi:10.1080/17405629.2011.643169
- Festl, R., Schwarkow, M., & Quandt, T. (2013). Peer influence, internet use and cyberbullying: A comparison of different context effects among German adolescents. *Journal of Children and Media*, 7(4), 446–462. doi:10.1080/17482798.2013.781514
- Heirman, W., & Walrave, M. (2012). Predicting adolescent perpetration in cyberbullying: An application of the theory of planned behavior. *Psicothema*, 24, 614–620. PMID:23079360

- Hinduja, S., & Patchin, J. W. (2013). Social influences on cyberbullying behaviors among middle and high school students. *Journal of Youth and Adolescence*, 42(5), 711–722. doi:10.1007/s10964-012-9902-4 PMID:23296318
- Huang, Y., & Chou, C. (2010). An analysis of multiple factors of cyberbullying among junior high school students in Taiwan. *Computers in Human Behavior*, 26(6), 1581–1590. doi:10.1016/j.chb.2010.06.005
- Jang, H., Song, J., & Kim, R. (2014). Does the offline bully-victimization influence cyberbullying behavior among youths? Application of general strain theory. *Computers in Human Behavior*, 31, 85–93. doi:10.1016/j.chb.2013.10.007
- Kowalski, R. M., & Limber, S. P. (2007). Electronic bullying among middle school students. *The Journal of Adolescent Health*, 41(6), 22–30. doi:10.1016/j.jadohealth.2007.08.017 PMID:18047942
- Laftman, S. B., Modin, B., & Ostberg, V. (2013). Cyberbullying and subjective health: A large-scale study of students in Stockholm, Sweden. *Children and Youth Services Review*, 35(1), 112–119. doi:10.1016/j.chilyouth.2012.10.020
- Mason, K. (2008). Cyberbullying: A preliminary assessment for school personnel. *Psychology in the Schools*, 45(4), 323–348. doi:10.1002/pits.20301
- Olweus, D. (1999). Sweden. In K. Smith, Y. Morita, J. Junger-Tas, D. Olweus, R. Catalano, & P. Slee (Eds.), *The nature of school bullying: A cross-national perspective* (pp. 7–27). New York, NY: Routledge.
- Patchin, J. W., & Hinduja, S. (2006). Bullies move beyond the schoolyard: A preliminary look at cyberbullying. *Youth Violence and Juvenile Justice*, 4(2), 148–169. doi:10.1177/1541204006286288
- Robson, C., & Witenberg, R. T. (2013). The influence of moral disengagement, morally based self-esteem, age, and gender on traditional bullying and cyberbullying. *Journal of School Violence*, 12(2), 211–231. doi:10.1080/15388220.2012.762921
- Sahin, M. (2010). Teachers perceptions of bullying in high schools: A Turkish study. *Social Behavior and Personality*, 38(1), 127–142. doi:10.2224/sbp.2010.38.1.127
- Smith, P. K., Mahdavi, J., Carvalho, M., Fisher, S., Russell, S., & Tippett, N. (2008). Cyberbullying: Its nature and impact in secondary school pupils. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 49(4), 376–385. doi:10.1111/j.1469-7610.2007.01846.x PMID:18363945
- Wade, A., & Beran, T. (2011). Cyberbullying: The new era of bullying. *Canadian Journal of School Psychology*, 26(1), 44–61. doi:10.1177/0829573510396318
- Wolak, J., Mitchell, K., & Finkelhor, D. (2007). Does online harassment constitute bullying? An exploration of online harassment by known peers and online-only contacts. *The Journal of Adolescent Health*, 41(6), 51–58. doi:10.1016/j.jadohealth.2007.08.019 PMID:18047945
- Wright, M. F. (2013). The relationship between young adults beliefs about anonymity and subsequent cyber aggression. *Cyberpsychology, Behavior, and Social Networking*, 16(12), 858–862. doi:10.1089/cyber.2013.0009 PMID:23849002
- Wright, M. F. (2014a). Predictors of anonymous cyber aggression: The role of adolescents beliefs about anonymity, aggression, and the permanency of digital content. *Cyberpsychology, Behavior, and Social Networking*, 17(7), 431–438. doi:10.1089/cyber.2013.0457 PMID:24724731

Wright, M. F. (2014b). Longitudinal investigation of the associations between adolescents popularity and cyber social behaviors. *Journal of School Violence, 13*(3), 291–314. doi:10.1080/15388220.2013.849201

Wright, M. F. (2015). Adolescents cyber aggression perpetration and cyber victimization: The longitudinal associations with school functioning. *Social Psychology of Education, 18*(4), 653–666. doi:10.1007/s11218-015-9318-6

Wright, M. F. (in press). Cyber victimization and perceived stress: Linkages to late adolescents' cyber aggression and psychological functioning. *Youth & Society*.

Wright, M. F., & Li, Y. (2013). The association between cyber victimization and subsequent cyber aggression: The moderating effect of peer rejection. *Journal of Youth and Adolescence, 42*(5), 662–674. doi:10.1007/s10964-012-9903-3 PMID:23299177

Ybarra, M. L., Diener-West, M., & Leaf, P. (2007). Examining the overlap in internet harassment and school bullying: Implications for school intervention. *The Journal of Adolescent Health, 1*(6), 42–50. doi:10.1016/j.jadohealth.2007.09.004 PMID:18047944

Ybarra, M. L., & Mitchell, K. J. (2004). Online aggressor/targets, aggressors, and targets: A comparison of associated youth characteristics. *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 45*(7), 1308–1316. doi:10.1111/j.1469-7610.2004.00328.x PMID:15335350

Zhou, Z., Tang, H., Tian, Y., Wei, H., Zhang, F., & Morrison, C. M. (2013). Cyberbullying and its risk factors among Chinese high school students. *School Psychology International, 34*(6), 630–647. doi:10.1177/0143034313479692

ADDITIONAL READING

Antoniadou, N., & Kokkinos, C. M. (2015). A review of research on cyber-bullying in Greece. *International Journal of Adolescence and Youth, 20*(2), 185–201. doi:10.1080/02673843.2013.778207

Barlett, C., & Coyne, S. M. (2014). A meta-analysis of sex differences in cyber-bullying behaviour: The moderating role of age. *Aggressive Behavior, 40*(5), 474–488. doi:10.1002/ab.21555 PMID:25098968

Bauman, S. (2011). *Cyberbullying: What counsellors need to know*. Alexandria, VA: American Counseling Association.

Bauman, S., Cross, D., & Walker, J. (2013). *Principles of cyberbullying research: Definitions, measures, and methodology*. New York, NY: Routledge.

Boyd, D. (2014). *It's complicated: The social lives of networked teens*. New Haven, CT: Yale University Press.

Cassidy, W., Faucher, C., & Jackson, M. (2013). Cyberbullying among youth: A comprehensive review of current international research and its implications and application to policy and practice. *School Psychology International, 34*(6), 575–612. doi:10.1177/0143034313479697

Chisholm, J. F. (2014). Review of the status of cyberbullying and cyberbullying prevention. *Journal of Information Systems Education, 25*(1), 1–14.

Cowie, H. (2013). Cyberbullying and its impact on young peoples emotional health and well-being. *The Psychiatrist, 37*(5), 167–170. doi:10.1192/pb.bp.112.040840

- Ey, L.-A., Taddeo, C., & Spears, B. (2015). Cyberbullying and primary-school aged children: The psychological literature and the challenge for sociology. *Societies*, 5(2), 492–514. doi:10.3390/soc5020492
- Gorzig, A., & Frumkin, L. A. (2013). Cyberbullying experiences on-the-go: When social media can become distressing. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, 7(1), article 4. Retrieved from: <http://www.cyberpsychology.eu/view.php?cisloclanku=2013022801&article=4>
- Hamm, M. P., Newton, A. S., Chisholm, A., Shulhan, J., Milne, A., Sundar, P., & Hartling, L. et al. (2015). Prevalence and effect of cyberbullying on children and young people: A scoping review of social media studies. *JAMA Pediatrics*, 169(8), 770–777. doi:10.1001/jamapediatrics.2015.0944 PMID:26098362
- Hinduja, S. K., & Patchin, J. W. (2014). *Bullying beyond the schoolyard: Preventing and responding to cyberbullying*. Thousand Oaks, CA: Corwin.
- Kiriakidis, S. P., & Kavoura, A. (2010). Cyberbullying: A review of the literature on harassment through the Internet and other electronic means. *Family & Community Health*, 33(2), 82–93. doi:10.1097/FCH.0b013e3181d593e4 PMID:20216351
- Kowalski, R. M., Guimetti, G. W., Schroeder, A. N., & Lattanner, M. R. (2014). Bullying in the digital age: A critical review and meta-analysis of cyberbullying research among youth. *Psychological Bulletin*, 140(4), 107–137. doi:10.1037/a0035618 PMID:24512111
- Kowalski, R. M., Limber, S. P., & Agatston, P. W. (2012). *Cyberbullying: Bullying in the digital age*. Indianapolis, IN: Wiley Publishing.
- Li, Q. (2015). When cyberbullying and bullying meet gaming: A systematic review of the literature. *Jahrbuch für Psychologie und Psychotherapie*, 5, 1–11.
- Li, Q., Cross, D., & Smith, P. K. (2012). *Cyberbullying in the global playground: Research from international perspectives*. West Sussex, UK: Blackwell Publishing. doi:10.1002/9781119954484
- Modecki, K. L., Minchin, J., Harbaugh, A. G., Guerra, N. G., & Runions, K. C. (2014). Bullying prevalence across contexts: A meta-analysis measuring cyber and traditional bullying. *The Journal of Adolescent Health*, 55(5), 602–611. doi:10.1016/j.jadohealth.2014.06.007 PMID:25168105
- Navarro, R., Yubero, S., & Larranaga, E. (2016). *Cyberbullying across the globe: Gender, family, and mental health*. Switzerland: Springer International Publishing. doi:10.1007/978-3-319-25552-1
- Nixon, C. L. (2014). Current perspectives: The impact of cyberbullying on adolescent health. *Adolescent Health, Medicine and Therapeutics*, 5, 143–158. PMID:25177157
- Notar, C. E., Padgett, S., & Roden, J. (2013). Cyberbullying: A review of the literature. *Universal Journal of Educational Research*, 1(1), 1–9.
- Patchin, J. W., & Hinduja, S. (2013). *Cyberbullying prevention and response: Expert perspectives*. New York, NY: Routledge.
- Perren, S., Corcoran, L., Cowie, H., Dehue, F., Garcia, D., Mc Guckin, C., & Vollink, T. et al. (2012). Tackling cyberbullying: Review of empirical evidence regarding successful responses by students, parents, and schools. *International Journal of Conflict and Violence*, 6(2), 283–293.
- Schneider, S. K., O'Donnell, L., Stueve, A., & Coutler, R. W. S. (2012). Cyberbullying, school bullying, and psychological distress: A regional census of high school students. *American Journal of Public Health*, 102(1), 171–177. doi:10.2105/AJPH.2011.300308 PMID:22095343

Shariff, S. (2015). *Sexting and cyberbullying: Defining the line for digitally empowered kids*. New York, NY: Cambridge University Press. doi:10.1017/CBO9781139095891

Smith, P. K., & Steffgen, G. (2013). *Cyberbullying through the new media: Findings from an international network*. New York, NY: Psychology Press.

Van Geel, M., Vedder, P., & Tanilon, J. (2014). Relationship between peer victimization, cyberbullying, and suicide in children and adolescents. *JAMA Pediatrics*, 168(5), 435–442. doi:10.1001/jamapediatrics.2013.4143 PMID:24615300

Wright, M. F. (2016). *A social-ecological approach to cyberbullying*. Hauppauge, NY: NOVA Science Publishers.

KEY TERMS AND DEFINITIONS

Cyberbullying: When young people harass, humiliate, embarrass, intimidate, and/or threaten another young person via information and communication technologies.

Cyber Victimization: When young people experience harassment, humiliation, embar-

rassment, intimidation, and/or are threaten via information and communication technologies by another young person.

Face-to-Face Bullying: When young people harass, humiliate, embarrass, intimidate, and/or threaten another young person offline.

Face-to-Face Victimization: When young people experience harassment, humiliation, embarrassment, intimidation, and/or are threaten offline by another young person.

Information and Communication Technologies: This board term encompasses communication device or applications, such as computers and cellular phones.

Parental Involvement: Involves the participation of parents in their children's activities.

Parental Mediation of Technology Use: The strategies parents use to control, supervise, or interpret information and communication technology content for their children.

Peer Attachment: The feelings and beliefs that one's peers will be there when needed.

School Climate: The quality and character of school life, which includes norms, values, interpersonal relationships, social interactions, and organizational processes and structures.

Category D

Data Analysis and Statistics

Advanced Recommender Systems

D**Young Park**

Bradley University, USA

INTRODUCTION

It is increasingly difficult to find the right information on the Web in the age of explosive information overload. Recommender systems provide users with personalized suggestions of goods, services, or information and thus help them find the most relevant and interesting goods, services, or information for them. Over the last two decades since the first major recommender systems emerged in the mid-1990s (Konstan et al., 1997; Resnick & Varian, 1997), numerous recommender systems have been developed and used in various application domains including e-commerce, education, and engineering (Aggarwal, 2016; Jannach, Zanker, Felfernig, & Friedrich, 2011; Manouselis, Drachsler, Verbert, & Santos, 2014; Ricci, Rokach, & Shapira, 2015; Robillard, Maalej, Walker, & Zimmermann, 2014). Recommender systems have also proven very useful in various application domains.

A basic personalized recommender system suggests a list of items that seem to be most relevant for a given single target user without considering the context that the user is in by using users' ratings of items on a single overall criterion where both users and items are in a single domain (Jannach et al., 2011). The basic recommender system can be extended in several ways. There are four major extensions, i.e., suggesting items for a group of target users rather than a single user (group recommendations), suggesting items by considering a specific context of the target user (context-aware recommendations), suggesting items using ratings on multiple criteria rather than a single overall criterion (multi-criteria recommendations), and suggesting items by using users and items in

multiple domains rather than a single domain (multiple-domain recommendations).

In this chapter, we present a brief and systematic overview of four major advanced recommender systems — *group* recommender systems, *context-aware* recommender systems, *multi-criteria* recommender systems, and *cross-domain* recommender systems. We characterize and compare them within a unifying model as extensions of the basic recommender systems. Future research topics and directions in the area of advanced personalized recommendations are discussed.

BACKGROUND

Recommender Systems (RS)

A *recommender system (RS)* is a software-intensive system that provides a given target users with personalized recommendations of items such as goods, services, or information to guide the user to find the most relevant items (Aggarwal, 2016; Jannach et al., 2011; Ricci et al., 2015). The personalized recommendations are made by using profiles of the target user and other users with respect to items.

A typical recommender system generally uses the following three types of data — data about the users (U), data about the items such as goods, services, or information (I), and data about the relevance (such as rating, evaluation, purchase, or interest) information between the users and the items (R) where

- U contains a set of all existing users.
- I contains a set of all existing items.

- R contains relevance ratings of items to users that are represented as a matrix that maps a user-item pair into a relevance value and are constructed over time either explicitly, inferred implicitly, or both.

Given a target user T , a recommender system recommends a list of new items that are likely to be most relevant to the target user by using U , I , and R . The personalized recommendation process generally uses two phases: the prediction phase and the suggestion phase. The prediction phase predicts the unknown relevance values of new items by the target user based on the *similarity* between users or between items. There are a number of similarity metrics including *Pearson correlation* and *Cosine similarity* to represent the degree of similarity between users or items (Jannach et al., 2011). The suggestion phase generates a list of top n items with the highest predicted relevance values. Various metrics such as *mean absolute error* (MAE), *mean square error* (MSE), *root mean squared error* (RMSE), *precision* and *recall* are used for evaluating the accuracy of the recommended items (Herlocker, Konstan, Terveen, & Riedl, 2004; Jannach et al., 2011).

There are a number of different approaches for generating personalized recommendations. Three major approaches for personalized recommendations are *collaborative recommender*, *content-based recommender* and *hybrid recommender*. For recommending relatively complex items there is an approach called *knowledge-based recommender* that is based on deep domain *knowledge* about users and items.

Collaborative filtering-based recommendation is based on *user collaborations* and is the most widely used and proven method of providing recommendations (Ekstrand, Riedl, & Konstan, 2011; Schafer, Frankowski, Herlocker, & Sen, 2007). There are two types of collaborative filtering: *neighborhood-based collaborative filtering* such as user-based collaborative recommender and item-based collaborative recommender (Linden, Smith, & York, 2003) and *model-based collab-*

orative filtering (Koren, Bell, & Volinsky, 2009). The collaborative filtering-based recommenders, however, have the relevance feedback sparsity problem and the new item problem. Content-based recommendation is based on the *content* of items (Lops, Gemmis, & Semeraro, 2011; Pazzani & Billsus, 2007). The content-based recommender recommends a list of items with similar content to the items that the target user gave good ratings. But, content-based recommendation has limitations such as the item overspecialization problem and the new user problem. Hybrid recommendation recommends items via *hybridization*, i.e., by combining the content-based recommendation and the collaborative filtering-based recommendation together (Burke, 2007).

Basic Recommendation

We will use a model of basic personalized recommendation to characterize advanced recommendations as extensions of the basic model. We define the *basic recommender system* (BRS) as a collaborative recommender system that recommends a list of items:

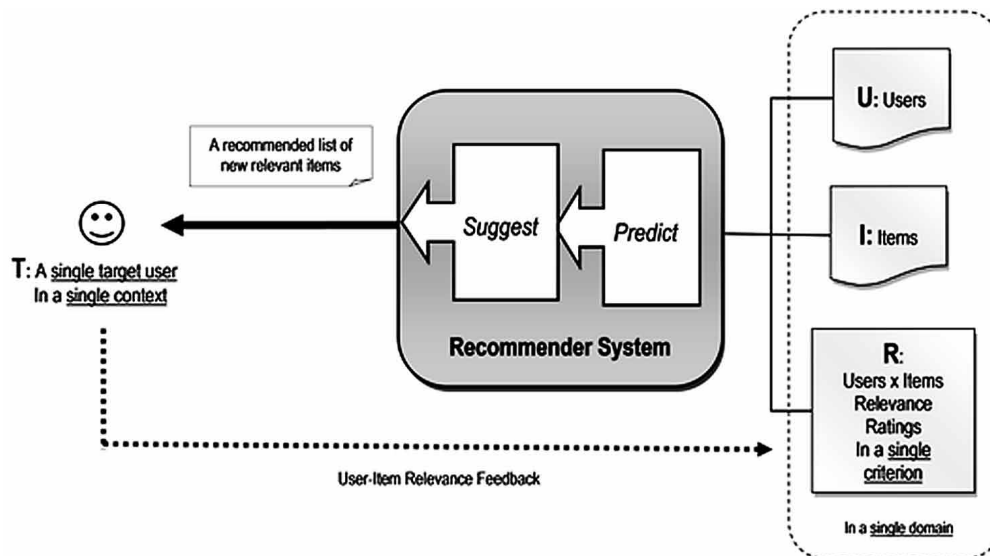
- For a single target user,
- In a single context,
- By using users' ratings of items on a single overall relevance criterion,
- where both users and items are in a single domain.

Figure 1 depicts the overall architecture of our basic personalized recommender system that is represented as *single target user*, *single context*, *single relevance criterion* & *single domain*.

ADVANCED RECOMMENDATIONS

There are four major advanced recommendations, i.e., group recommendation, context-aware recommendation, multi-criteria recommendation, and cross-domain recommendation. We comparatively

Figure 1. Basic personalized recommendation – Single target user, single context, single relevance criterion & single domain



characterize these major advanced recommendations as extensions of the basic personalized recommendation model shown in Figure 1 in a unifying framework.

Group Recommender Systems (GRS)

Basic recommender systems recommend a list of items most relevant to a single individual user. However, there are many situations that call for recommendations of relevant items for a group of users instead of a single user, for example, restaurant recommendation for a user and the user's friends (Berkovsky & Freyne, 2010; Jameson & Smyth, 2007; Kompan & Bielikova, 2014;

Masthoff, 2011). The recommendations are shared by the group of users.

A *group recommender system* (GRS) is a recommender system that provides a group of users with a shared list of items that are most relevant to the users in the group.

Group recommendation as an extension of the basic single-user recommendation is illustrated in Figure 2. Groups can be real or virtual. There can be different types of groups, such as established groups, occasional groups, random groups, and automatically identified groups.

We generate group recommendations using the basic single-user recommender system. There are three general approaches for the group recom-

Figure 2. Group recommendation – Group of target users

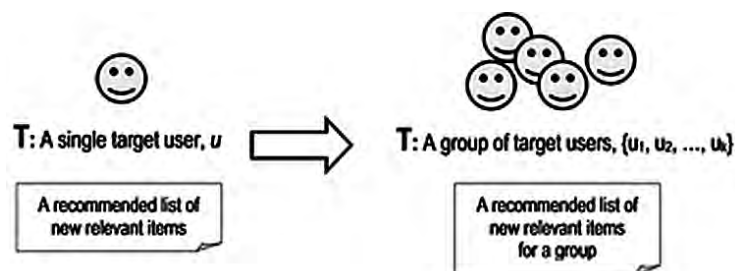
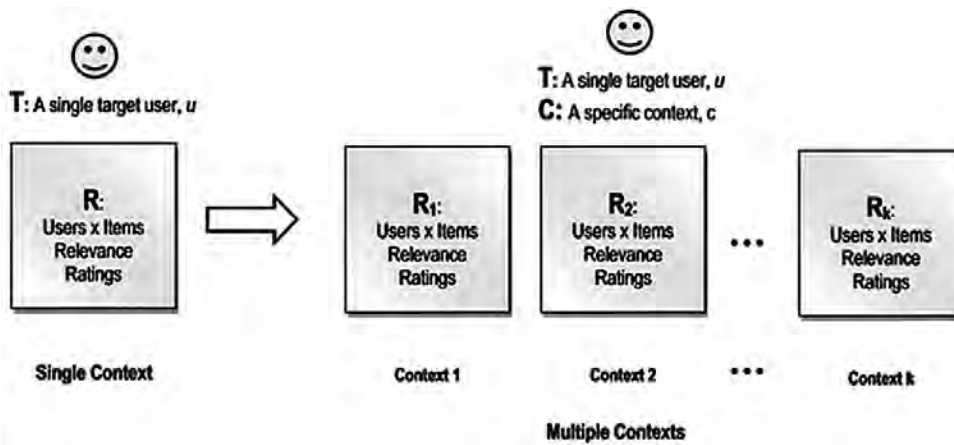


Figure 3. Context-aware recommendation – Multiple contexts



mendation process (Berkovsky & Freyne, 2010; Jameson & Smyth, 2007; Kompan & Bielikova, 2014; Masthoff, 2011):

Approach 1: Aggregate the users $u_1, u_2, \dots, u_k$ in the group into a single aggregated group user $u_{\text{aggregated}}$ and apply the single-user recommender system to the aggregated group user $u_{\text{aggregated}}$ to produce the group recommendation of items.

Approach 2: Apply the single-user recommender system to each user u_i for $i = 1, 2, \dots, k$ in the group and aggregate k single-user recommendations of items into the group recommendation of items.

Approach 3: Represent the whole group of target users as a single user u_{group} and apply the single-user recommendation to the user u_{group} to produce the group recommendation of items.

Various aggregation strategies can be used for user and recommendation aggregation, including additive strategy, multiplicative strategy, plurality strategy, fairness strategy, average strategy, least misery strategy, and most pleasure strategy. Most aggregation strategies seem to show similar results.

Group recommendation can be used for the cold start problem with new users. Many group recommenders have been proposed in various

application domains, including movies, songs, books, television programs, news, restaurants, recipes, tours, and learning.

Context-Aware Recommender Systems (CARS)

A basic recommender system recommends a list of items most relevant to a target user without considering any additional specific contextual information about the user. However, in many situations, recommendations of more relevant items need additional contextual information about the target user (Adomavicius, Mobasher, Ricci & Tuzhilin, 2011; Adomavicius & Tuzhilin, 2011; Liu, Ma, Chen, & Xiong, 2013; Panniello & Gorgoglione, 2011; Symeonidis, Ntempos, & Manolopoulos, 2014). For example, more enhanced recommendation of music can be possible by considering contexts like the location, time, and mood of a user. This additional contextual information can provide better recommendations.

A *context-aware recommender system* (CARS) is a recommender system that provides a target user within a specific context with a list of items that are most relevant to that user in that context.

Figure 3 shows context-aware rating recommendation as an extension of the basic single-context rating recommendation.

Contextual information in recommender systems is the context in which recommendations are provided to a target user, such as time, location, surroundings, companions, occasions, and so on. Time is a useful context and used in time-aware recommendations (Campos, Díez, & Cantador, 2014). Location is an important contextual information in mobile environments for location-aware recommendations (del Carmen, Ilarri, Trillo-Lado, & Hermoso, 2015; Levandoski, Sarwat, Eldawy, & Mokbel, 2012).

We can incorporate contextual information in the basic single-context recommender system in order to generate context-aware recommendations in the following three ways (Adomavicius et al., 2011; Adomavicius & Tuzhilin, 2011; Panniello & Gorgoglione, 2011):

Approach 1: Select the user-item ratings R_c for a given specific context c from the all-context user-item ratings R_1 to R_k (called *contextual pre-processing*), and then apply basic single-context rating recommendation using the user-item ratings R_c in the context c for a given target user u to produce the contextual recommendation of items.

Approach 2: Apply basic single-context rating recommendation using the all-context user-item ratings R_1 to R_k to produce the recommendation of items for all contexts for a given target user u , and then select the contextual recommendation of items for a given context c from the recommendation of items for all contexts (called *contextual post-processing*).

Approach 3: Apply either a similarity-based or a model-based multidimensional rating recommendation for a given target user u and a given context c to produce the contextual recommendation of items (called *contextual modeling*).

These three approaches can be combined to improve recommendation accuracy. Various context-aware recommenders have been proposed

in several mobile application domains, such as restaurant, travel, tour, learning, and shared economy.

Multi-Criteria Recommender Systems (MCRS)

Basic recommender systems recommend a list of most relevant items by using the relevance ratings of items provided by users in a single criterion of overall relevance. However, there are many situations in which the recommendation process requires more information than just single overall relevance ratings, i.e., relevance ratings of items in many aspects (Adomavicius & Kwon, 2007; Adomavicius, Manouselis, & Kwon, 2011; Lakiotaki, Matsatsinis, & Tsoukias, 2011; Manouselis & Costopoulou, 2007). For example, hotels can be rated on many dimensions such as the room, the service, and the neighborhood, as well as overall satisfaction. This additional information can enhance recommendation quality.

A *multi-criteria recommender system* (MCRS) is a recommender system that provides a target user with a list of items that are most relevant to that user by using the relevance ratings of items in multiple criteria that are provided by the users.

Multi-criteria rating recommendation as an extension of the basic single-criterion rating recommendation is shown in Figure 4. Here, r_i is the rating in overall criterion c_i and r_i for $i = 2, 3, \dots$, k is the rating in each criterion c_i for $i = 2, 3, \dots$, k .

Figure 4. Multi-criteria recommendation – Multiple relevance criteria

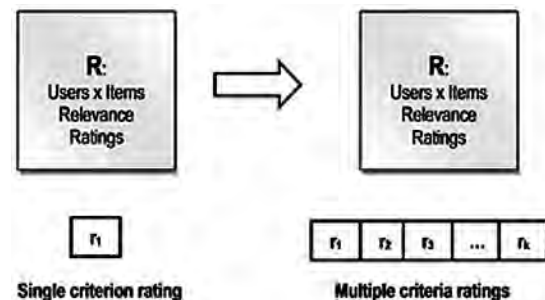
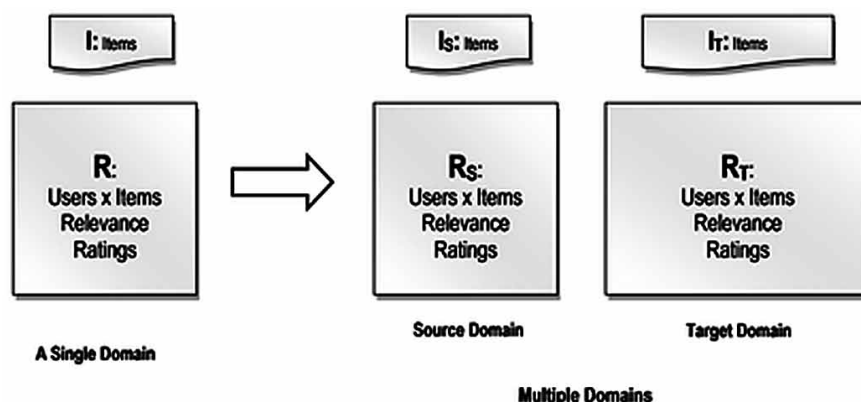


Figure 5. Cross-domain recommendation – Multiple domains



To generate multi-criteria recommendations, we incorporate multi-criteria rating information during the rating prediction phase and the item suggestion phase of the basic single-criterion recommender system (Adomavicius & Kwon, 2007; Adomavicius et al., 2011; Lakiotaki et al., 2011).

In the neighborhood-based rating prediction phase, we use the following two approaches to incorporate the multi-criteria ratings:

Approach 1: Compute for each criterion c_i for $i = 1, 2, 3, \dots, k$ the similarity between two users using some similarity measure in the single-criterion rating recommender system. Then, compute the overall similarity between two users by aggregating these k similarities in individual criteria using some aggregation strategy such as average similarity, worst-case similarity, or aggregated (weighted sum of individual similarities) similarity.

Approach 2: Compute the overall similarity between two users using some multidimensional distance similarity measure such as Manhattan distance, Euclidean distance, or maximal value distance.

In the model-based rating prediction phase, approaches for incorporating the multi-criteria ratings include the aggregation function approach, the probabilistic modeling approach, and the multilinear singular value decomposition approach.

In the aggregation function approach, the overall rating is viewed as an aggregate of other criteria ratings and the relationship is modeled as an aggregate function of other criteria ratings.

The multi-criteria ratings can also be used in the item suggestion phase if the overall rating criterion is not available. A number of multi-criteria recommenders have been proposed in various application domains including movies, music, books, restaurants, hotels, tours, and learning.

Cross-Domain Recommender Systems (CDRS)

A basic recommender system suggests a list of most relevant items in a single domain (such as movies or books) and is thus a domain-specific recommender system. In many cases, the same users provide relevant ratings of different types of items, like movies and books, in different domains (Cantador, Fernández Tobías, Berkovsky, & Cremonesi, 2015; Dong, 2012; Fernández-Tobías, Cantador, Kaminskas, & Ricci, 2012; Li, 2011). For domains that share some resources like users, items and ratings, knowledge from one domain (the source domain) can be exploited in and transferred to better recommendations in another domain (the target domain).

A *cross-domain recommender system* (CDRS) is a recommender system that provides a target user with a list of items in the target domain that

are most relevant to the target user by exploiting knowledge from the source domain that shares resources with the target domain.

Figure 5. illustrates cross-domain rating recommendation as an extension of the basic single-domain rating recommendation. Domains can be defined at different levels, such as item attribute level, item type level, item level, and system level. Resource sharing between the source and target domains can be both user and item overlap, user overlap only, item overlap only, and no overlap.

In general, there are two approaches to incorporating knowledge from the source domain to the target domain for cross-domain recommendations (Cantador et al., 2015; Fernández-Tobías et al.2012):

Approach 1: Aggregate knowledge from the source domain(s) to produce recommendations of items in the target domain (called *knowledge aggregation*).

Approach 2: Link and transfer knowledge from the source domain(s) to produce recommendations of items in the target domain (called *knowledge linkage and transfer*).

There are various techniques for knowledge aggregation across domains:

- Merge user ratings and apply the single-domain recommender system to the merged ratings to produce the target domain recommendation.
- Build user modeling from the source domain for use in applying the single-domain recommender system to the target domain to produce the target domain recommendation.
- Apply the single-domain recommender system to the source domain and the target domain and combine the two resulting recommendations to produce the target domain recommendation.

There are several techniques for knowledge linkage and knowledge transfer across domains:

- Link items in the source and target domains using some common knowledge.
- Share common latent features between the source and target domains.
- Transfer rating patterns between the source and target domains.

Cross-domain rating recommendation can address the cold start problem and the sparsity problem, and thus enhance recommendation accuracy. Various cross-domain recommenders have been proposed in several application domains including movies, television shows, music, books, and restaurants.

FUTURE RESEARCH DIRECTIONS

Advanced recommender systems such as group recommendation, context-aware recommendation, multi-criteria recommendation, and cross-domain recommendation are all relatively new research areas presenting many underexplored and open research topics.

Possible future research topics in the area of group recommender systems may include subareas such as modeling users and groups, aggregation strategies, incorporating personality and social aspects, explaining group recommendations, privacy and security issues in group recommendations, experimental evaluation, new application domains, and the connection with multi-criteria recommendations.

In the area of context-aware recommendation, interesting future topics may include modeling contexts, strategies for context incorporation, explaining context-aware recommendations, context suggestions, evaluation on practical applications, new application domains, and the connection with mobile recommendations.

Topics like building criteria for items, selecting the best set of criteria, collecting real-world datasets for experimental evaluation, new application domains, and the connection with group recommendations are possible research subareas in the area of multi-criteria recommender systems.

Incorporating knowledge from the source domain, real-world cross-domain datasets for experimental evaluation, new application domains, and the connection with context-aware recommendations are interesting future research topics in the area of cross-domain recommender systems.

Most current research has focused on each advanced recommendation separately, but combining different advanced recommendations opens new and exciting possibilities for the future.

CONCLUSION

We have presented a systematic overview of four major advanced recommender systems. We have comparatively characterized group recommender systems, context-aware recommender systems, multi-criteria recommender systems, and cross-domain recommender systems within a unifying model as extensions of basic recommender systems. Future research directions in the area of advanced personalized recommendation have been discussed. We believe that advanced recommender technologies will continue to advance.

REFERENCES

- Adomavicius, G., & Kwon, Y. (2007). New recommendation techniques for multicriteria rating systems. *IEEE Intelligent Systems*, 22(3), 48–55. doi:10.1109/MIS.2007.58
- Adomavicius, G., Manouselis, N., & Kwon, Y. (2011). Multi-criteria recommender systems. In *Recommender systems handbook* (pp. 769–803). New York, NY: Springer. doi:10.1007/978-0-387-85820-3_24
- Adomavicius, G., Mobasher, B., Ricci, F., & Tuzhilin, A. (2011). Context-Aware Recommender Systems. *AI Magazine*, 32(3), 67–80.
- Adomavicius, G., & Tuzhilin, A. (2011). Context-aware recommender systems. In *Recommender systems handbook* (pp. 217–253). New York, NY: Springer. doi:10.1007/978-0-387-85820-3_7
- Aggarwal, C. (2016). *Recommender Systems – The Textbook*. Berlin: Springer. doi:10.1007/978-3-319-29659-3
- Berkovsky, S., & Freyne, J. (2010, September). Group-based recipe recommendations: analysis of data aggregation strategies. In *Proceedings of the fourth ACM conference on Recommender systems* (pp. 111–118). ACM. doi:10.1145/1864708.1864732
- Burke, R. (2007). Hybrid web recommender systems. In *The Adaptive Web* (pp. 377–408). Berlin: Springer. doi:10.1007/978-3-540-72079-9_12
- Campos, P. G., Díez, F., & Cantador, I. (2014). Time-aware recommender systems: A comprehensive survey and analysis of existing evaluation protocols. *User Modeling and User-Adapted Interaction*, 24(1-2), 67–119. doi:10.1007/s11257-012-9136-x
- Cantador, I., Fernández-Tobías, I., Berkovsky, S., & Cremonesi, P. (2015). Cross-domain recommender systems. In *Recommender systems handbook* (pp. 919–959). New York, NY: Springer. doi:10.1007/978-1-4899-7637-6_27
- del Carmen, M., Ilarri, S., Trillo-Lado, R., & Hermoso, R. (2015). Location-Aware Recommendation Systems: Where We Are and Where We Recommend to Go. In *Proceedings of the Workshop on Location-Aware Recommendations* (pp. 1–8). ACM.
- Dong, G. (2012). Cross domain similarity mining: Research issues and potential applications including supporting research by analogy. *ACM SIGKDD Explorations Newsletter*, 14(1), 43–47. doi:10.1145/2408736.2408744

- Ekstrand, M. D., Riedl, J. T., & Konstan, J. A. (2011). Collaborative filtering recommender systems. *Foundations and Trends in Human-Computer Interaction*, 4(2), 81–173. doi:10.1561/11000000009
- Fernández-Tobías, I., Cantador, I., Kaminskas, M., & Ricci, F. (2012, June). Cross-domain recommender systems: A survey of the state of the art. *Spanish Conference on Information Retrieval*.
- Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. (2004). Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems*, 22(1), 5–53. doi:10.1145/963770.963772
- Jameson, A., & Smyth, B. (2007). Recommendation to groups. In *The adaptive web* (pp. 596–627). Berlin: Springer. doi:10.1007/978-3-540-72079-9_20
- Jannach, D., Zanker, M., Felfernig, A., & Friedrich, G. (Eds.). (2011). *Recommender Systems: An Introduction*. Cambridge University Press.
- Kompan, M., & Bielikova, M. (2014). Group recommendations: Survey and perspectives. *Computing and Informatics*, 33(2), 446–476.
- Konstan, J. A., Miller, B. N., Maltz, D., Herlocker, J. L., Gordon, L. R., & Riedl, J. (1997). GroupLens: Applying collaborative filtering to usenet news. *Communications of the ACM*, 40(3), 77–87. doi:10.1145/245108.245126
- Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix Factorization Techniques for Recommender Systems. *Computer*, 42(8), 30–37. doi:10.1109/MC.2009.263
- Lakiotaki, K., Matsatsinis, N. F., & Tsoukias, A. (2011). Multicriteria user modeling in recommender systems. *IEEE Intelligent Systems*, 26(2), 64–76. doi:10.1109/MIS.2011.33
- Levandowski, J. J., Sarwat, M., Eldawy, A., & Mokbel, M. F. (2012). Lars: A location-aware recommender system. In *2012 IEEE 28th International Conference on Data Engineering* (pp. 450–461). IEEE. doi:10.1109/ICDE.2012.54
- Li, B. (2011, November). Cross-domain collaborative filtering: A brief survey. In *Tools with Artificial Intelligence (ICTAI), 2011 23rd IEEE International Conference on* (pp. 1085–1086). doi:10.1109/ICTAI.2011.184
- Linden, G., Smith, B., & York, J. (2003). Amazon.com Recommendations: Item-to-Item Collaborative Filtering. *Internet Computing*, 7(1), 76–80. doi:10.1109/MIC.2003.1167344
- Liu, Q., Ma, H., Chen, E., & Xiong, H. (2013). A Survey of Context-Aware Mobile Recommendations. *International Journal of Information Technology & Decision Making*, 12(01), 139–172. doi:10.1142/S0219622013500077
- Lops, P., Gemmis, M., & Semeraro, G. (2011). Content-based Recommender Systems: State of the Art and Trends. In *Recommender Systems Handbook* (pp. 73–105). Berlin: Springer. doi:10.1007/978-0-387-85820-3_3
- Manouselis, N., & Costopoulou, C. (2007). Analysis and classification of multi-criteria recommender systems. *World Wide Web (Bussum)*, 10(4), 415–441. doi:10.1007/s11280-007-0019-8
- Manouselis, N., Drachsler, H., Verbert, K., & Santos, O. (Eds.). (2014). *Recommender Systems for Technology Enhanced Learning: Research Trends and Applications*. Berlin: Springer. doi:10.1007/978-1-4939-0530-0
- Masthoff, J. (2011). Group recommender systems: Combining individual models. In *Recommender systems handbook* (pp. 677–702). New York, NY: Springer. doi:10.1007/978-0-387-85820-3_21

Panniello, U., & Gorgoglione, M. (2011). Context-Aware Recommender Systems: A Comparison of Three Approaches. In DART@ AI* IA.

Pazzani, M. J., & Billsus, D. (2007). Content-based recommendation systems. In *The Adaptive Web* (pp. 325–341). Berlin: Springer. doi:10.1007/978-3-540-72079-9_10

Resnick, P., & Varian, H. R. (1997). Recommender systems. *Communications of the ACM*, 40(3), 56–58. doi:10.1145/245108.245121

Ricci, F., Rokach, L., & Shapira, B. (Eds.). (2015). *Recommender Systems Handbook* (2nd ed.). Berlin: Springer. doi:10.1007/978-1-4899-7637-6

Robillard, M., Maalej, W., Walker, R., & Zimmermann, T. (Eds.). (2014). *Recommendation Systems in Software Engineering*. Berlin: Springer. doi:10.1007/978-3-642-45135-5

Schafer, J. B., Frankowski, D., Herlocker, J., & Sen, S. (2007). Collaborative filtering recommender systems. In *The Adaptive Web* (pp. 291–324). Berlin: Springer. doi:10.1007/978-3-540-72079-9_9

Symeonidis, P., Ntempos, D., & Manolopoulos, Y. (2014). *Recommender Systems for Location-based Social Networks*. New York, NY: Springer. doi:10.1007/978-1-4939-0286-6

ADDITIONAL READING

Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6), 734–749. doi:10.1109/TKDE.2005.99

Aggarwal, C. (2016). *Recommender Systems – The Textbook*. Berlin, Heidelberg: Springer. doi:10.1007/978-3-319-29659-3

Felfernig, A., Jeran, M., Ninaus, G., Reinfrank, F., & Reiterer, S. (2013). Toward the Next Generation of Recommender Systems: Applications and Research Challenges. *Multimedia Services in Intelligent Environments - Advances in Recommender Systems*. 24, 81–98.

Jannach, D., Zanker, M., Felfernig, A., & Friedrich, G. (Eds.). (2011). *Recommender Systems: An Introduction*. Cambridge University Press.

Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix Factorization Techniques for Recommender Systems. *Computer*, 42(8), 30–37. doi:10.1109/MC.2009.263

Manouselis, N., Drachsler, H., Verbert, K., & Santos, O. (Eds.). (2014). *Recommender Systems for Technology Enhanced Learning: Research Trends and Applications*. Berlin, Heidelberg: Springer. doi:10.1007/978-1-4939-0530-0

Ricci, F., Rokach, L., & Shapira, B. P. (Ed.). (2015). *Recommender Systems Handbook*, 2nd edition. Berlin/Heidelberg: Springer.

Robillard, M., Maalej, W., Walker, R., & Zimmermann, T. (Eds.). (2014). *Recommendation Systems in Software Engineering*. Berlin, Heidelberg: Springer. doi:10.1007/978-3-642-45135-5

Symeonidis, P., Ntempos, D., & Manolopoulos, Y. (2014). *Recommender Systems for Location-based Social Networks*. New York, NY: Springer. doi:10.1007/978-1-4939-0286-6

KEY TERMS AND DEFINITIONS

Advanced Recommendation: Personalized recommendation as an extension of basic recommendation such as group recommender systems, context-aware recommender systems, multi-criteria recommender systems, and cross-domain recommender systems.

Basic Personalized Recommendation: Personalized recommendation of items for a single target user in a single context by using users' ratings of items on a single overall relevance criterion where both users and items are in a single domain.

Collaborative Recommender System: A recommender system that recommends items through user collaborations and are the most widely used and proven method of providing personalized recommendations. There are three major types: user-to-user collaborative filtering based on user-to-user similarity, item-to-item collaborative filtering based on item-to-item similarity, and latent factor model-based collaborative filtering based on user-item matrix factorization.

Context-Aware Recommender System: A recommender system that provides a target user within a specific context with a list of items that are most relevant to the target user in the specific context.

Cross-Domain Recommender System: A recommender system that provides a target user

with a list of items in the target domain that are most relevant to the target user by exploiting knowledge from the source domain that shares resources with the target domain.

Group Recommender System: A recommender system that provides a group of users as a whole with a shared list of items that are most relevant to the users in the group.

Multi-Criteria Recommender System: A recommender system that provides a target user with a list of items that are most relevant to the target user by using the relevance ratings of items in multiple criteria that are provided by the users.

Recommender System: A software system that provides a single target user within a single context with personalized recommendations of items such as goods, services or information to guide the target user to find most relevant items using ratings on a single relevance criterion (i.e., overall) and where both users and items are in a single domain.

Centrality Analysis of the United States Network Graph

Natarajan Meghanathan
Jackson State University, USA

INTRODUCTION

Network Science is one of the emerging fields of Data Science to analyze real-world networks from a graph theory point of view. Several real-world networks have been successfully modeled as undirected and directed graphs to study the intrinsic structural properties of the networks as well as the topological importance of nodes in these networks. The real-world networks that have been subjected to complex network analysis typically fall under one of these categories: social networks (Ghali et. al., 2012), transportation networks (Cheung & Gunes, 2012), biological networks (Ma & Gao, 2012), citation networks (Zhao & Strotmann, 2015), co-authorship networks (Ding, 2011) and etc. One category of real-world networks for which sufficient attention has not yet been given are the regional networks featuring the states within a country. In this chapter, we present a comprehensive analysis of a network graph of the states within a country with respect to the four commonly used centrality metrics in complex network analysis (Newman, 2010): degree, eigenvector, betweenness and closeness centralities.

We opine the chapter to serve as a model for anyone interested in analyzing a connected graph of the states within a country from a Network Science perspective. The approaches presented in this chapter could be useful to determine the states (and their cities) that are the most central and/or influential within a country. For example, the ranking of the vertices based on the shortest path centrality metrics (closeness and betweenness) could be useful to choose the states (and their cities) that could serve as hubs for transportation

networks (like road and airline networks). We could identify the states that are most the central states as well as identify the states that could form a connected backbone and geographically well-connected to the rest of the states within a country and use this information to design the road/rail transportation networks. The degree centrality and eigenvector centrality metrics as well as the network-level metrics like minimum connected dominating set and maximal clique size could be useful to identify fewer number of venues (with several adjacent states to draw people) for political campaigns/meetings that would cover the entire country.

We choose the United States (US) as the country for analysis and build a connected network graph of the contiguous states (48 states and the District of Columbia, DC) of the US: each state and DC is a node (vertex) and there exists a link (edge) between two vertices if the two corresponding states/DC share a common border. Though some prior studies have been conducted on transportation networks (Cheung & Gunes, 2012) and food flow networks (Lin et. al., 2014) in the United States, to the best of our knowledge, there has been no prior study of centrality analysis on the graph of the contiguous US states solely based on their geographical locations. Table 1 lists the contiguous states and DC in alphabetical order, their two character codes and the IDs used to refer to them in the chapter. The rest of the chapter is organized as follows: Section 2 introduces the centrality metrics and presents the results of the analysis on the states graph for each of them. Section 3 discusses related work. Section 4 concludes the chapter by summarizing the results

Table 1. List of contiguous states (including DC) of the US in alphabetical order

ID	State/District Name	Code	ID	State Name	Code	ID	State Name	Code
1	Alabama	AL	18	Maine	ME	34	Ohio	OH
2	Arizona	AZ	19	Maryland	MD	35	Oklahoma	OK
3	Arkansas	AR	20	Massachusetts	MA	36	Oregon	OR
4	California	CA	21	Michigan	MI	37	Pennsylvania	PA
5	Colorado	CO	22	Minnesota	MN	38	Rhode Island	RI
6	Connecticut	CT	23	Mississippi	MS	39	South Carolina	SC
7	Delaware	DE	24	Missouri	MO	40	South Dakota	SD
8	District of Columbia	DC	25	Montana	MT	41	Tennessee	TN
9	Florida	FL	26	Nebraska	NE	42	Texas	TX
10	Georgia	GA	27	Nevada	NV	43	Utah	UT
11	Idaho	ID	28	New Hampshire	NH	44	Vermont	VT
12	Illinois	IL	29	New Jersey	NJ	45	Virginia	VA
13	Indiana	IN	30	New Mexico	NM	46	Washington	WA
14	Iowa	IA	31	New York	NY	47	West Virginia	WV
15	Kansas	KS	32	North Carolina	NC	48	Wisconsin	WI
16	Kentucky	KY	33	North Dakota	ND	49	Wyoming	WY
17	Louisiana	LA						

of Section 2. For the rest of the chapter, the terms ‘network’ and ‘graph’, ‘node’ and ‘vertex’, ‘link’ and ‘edge’ are used interchangeably. They mean the same. The layout for the US States Network graph presented in Figure 1 is drawn using the Fruchterman Reingold layout algorithm (Fruchterman and Reingold, 1991), available in Gephi (Cherven, 2013).

BACKGROUND

Centrality Metrics

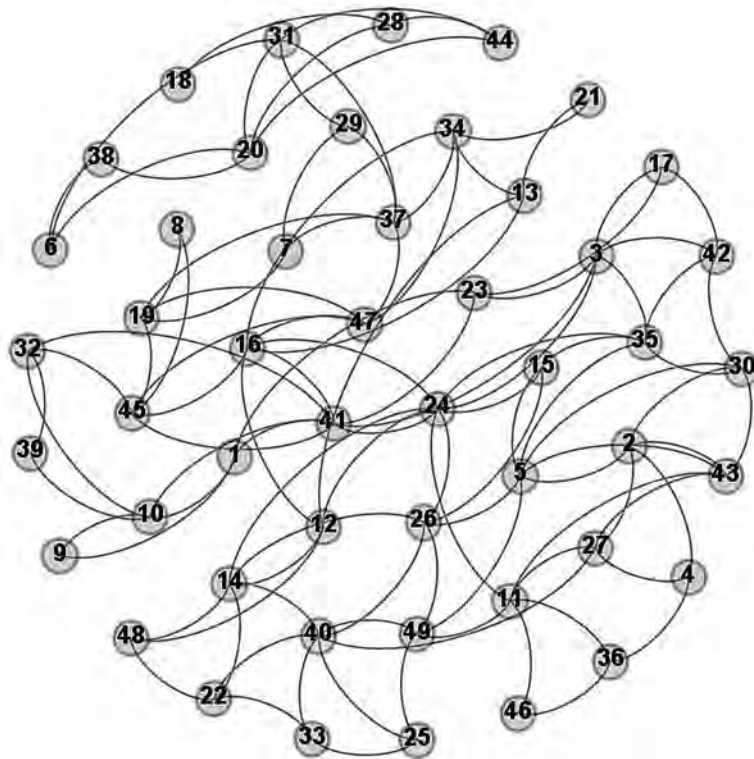
In this section, we introduce the centrality metrics for which we will run their respective algorithms on the US States Network graph and present the results (including their distribution and ranking of the vertices). The four centrality metrics analyzed are: degree centrality, eigenvector centrality, closeness centrality and betweenness centrality.

Since the US States graph is an undirected graph, the adjacency matrix of the graph is symmetric and there is only one value per vertex for each centrality metric.

Degree Centrality

The degree centrality (DegC) of a vertex is the number of edges incident on it. Table 2 presents the degree centrality of the vertices and the corresponding rank (in the decreasing order of their values) in the US States network graph; vertices with identical values for DegC have the same rank. The state of Missouri has the largest degree centrality value of 9, followed by the state of Tennessee with the second largest degree centrality value of 8. The state of Maine has the smallest degree centrality value of 1 (as New Hampshire is its only adjacent state). There are no ties among vertices for the largest and second largest values of degree centrality as well as for the vertex with

Figure 1. Fruchterman Reingold layout of the US network states graph



the smallest degree centrality. However, as we can notice (from Table 2): for other values of degree centrality, there are several instances of ties among vertices (we assign the same rank for all such vertices with identical values for degree centrality).

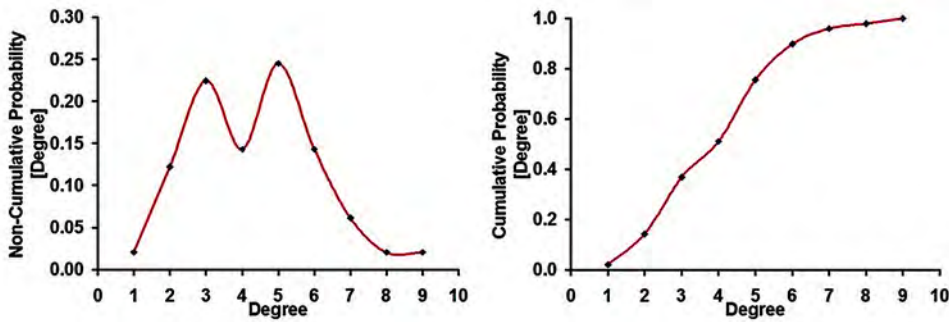
Figure 2 illustrates the cumulative and non-cumulative degree distributions of the vertices in the states graph. The cumulative distribution curve indicates that more than 85% of the vertices have degree of 6 or lower (the degree values are mostly ± 1 away from either of the two peaks), even though the largest degree observed is 9. The non-cumulative distribution curve illustrates that the degree distribution is bi-modal with a Poisson pattern (Balakrishnan & Nevzorov, 2003) with the peaks observed at degree values of 3 and 5. To corroborate this assertion, we observe the average degree of the vertices to be 4.37 (roughly close to the average of the two peak degree values of 3 and 5), with a standard deviation and Kurtosis

of 1.72 and 2.75 respectively (all indicating that the degree distribution is close to being a normal/Poisson distribution; Balakrishnan & Nevzorov, 2003). Kurtosis is a measure of the “tailedness” of a probability distribution (Balanda & MacGillivray, 1988). For normal distribution, the expected value for the Kurtosis is 3; distributions with Kurtosis values above or below 3 are said to be fat-tailed and thin-tailed respectively (Balanda & MacGillivray, 1988). Further, the spectral radius ratio for node degree (defined as the ratio of the principal eigenvalue of the adjacency matrix of the graph and the average node degree; Meghanathan, 2014) is observed to be 1.24. The spectral radius ratio for node degree is a measure of the variation in node degree of the vertices; the minimum possible value for this metric is 1.0 and farther the value of the spectral radius ratio for node degree from 1, the larger is the variation in node degree. Though the spectral radius ratio for node degree value of 1.24 is not much farther from 1, the value is not

Table 2. Ranking of the vertices in the US states network graph based on degree centrality

Rank	ID	DegC	Rank	ID	DegC	Rank	ID	DegC	Rank	ID	DegC
1	24	9	5	2	5	6	23	4	7	48	3
2	41	8	5	30	5	6	42	4	7	28	3
3	5	7	5	27	5	6	15	4	7	44	3
3	16	7	5	43	5	6	32	4	7	33	3
3	40	7	5	49	5	6	13	4	7	25	3
4	3	6	5	20	5	6	22	4	8	9	2
4	35	6	5	31	5	7	4	3	8	38	2
4	26	6	5	19	5	7	17	3	8	8	2
4	37	6	5	12	5	7	36	3	8	39	2
4	45	6	5	34	5	7	6	3	8	46	2
4	11	6	5	47	5	7	7	3	8	21	2
4	14	6	6	1	4	7	29	3	9	18	1
5	10	5									

Figure 2. Bi-modal poisson distribution (non-cumulative and cumulative) for the degree centrality metric of the US states graph



as close to 1 (Meghanathan, 2014) as observed for the US Football network (Girvan & Newman, 2002): a network with a unimodal Poisson degree distribution of the vertices; the value of 1.24 could be attributed to the variation due to the bi-modal degree distribution of the vertices.

Eigenvector Centrality

The eigenvector centrality (EVC) of a vertex is a measure of the degree of the vertex as well as the degree of its neighbors. A vertex is likely to have a higher EVC if it has a larger degree and its neighbor(s) also have a larger degree. The

EVC for a vertex corresponds to the entry for the vertex in the principal eigenvector of the adjacency matrix of the graph (Chung, 2006). The EVC of the vertices is computed by implementing the power-iteration algorithm [(Lay et al., 2015). Table 3 lists the EVC values (rounded to four decimal places) of the vertices in the US States graph. We observe a tie between the states of Vermont and Connecticut (values of 0.0050 each). The rest of the 47 vertices have a unique ranking. The US States graph is another example to illustrate that (unlike the degree centrality metric) the eigenvector centrality metric is more likely to return unique values for the vertices in

Table 3. Ranking of the vertices in the US states network graph based on Eigenvector Centrality

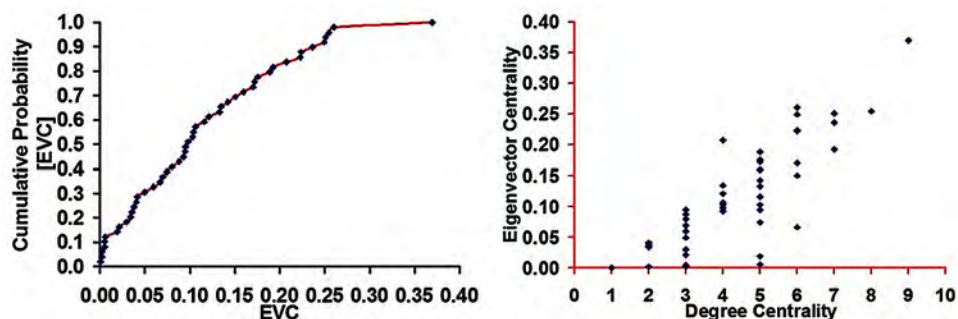
Rank	ID	EVC	Rank	ID	EVC	Rank	ID	EVC	Rank	ID	EVC
1	24	0.3697	14	11	0.1703	26	48	0.0951	38	21	0.0380
2	26	0.2605	15	43	0.1595	27	10	0.0946	39	39	0.0354
3	41	0.2546	16	45	0.1506	28	1	0.0927	40	9	0.0344
4	5	0.2509	17	2	0.1423	29	17	0.0880	41	7	0.0298
5	35	0.2494	18	42	0.1346	30	25	0.0803	42	29	0.0212
6	16	0.2362	19	27	0.1330	31	19	0.0742	43	31	0.0191
7	3	0.2235	20	23	0.1210	32	33	0.0696	44	20	0.0061
8	14	0.2227	21	47	0.1158	33	37	0.0666	45	44	0.0050
9	15	0.2076	22	22	0.1065	34	4	0.0597	45	6	0.0050
10	40	0.1924	23	13	0.1039	35	36	0.0496	46	28	0.0021
11	12	0.1887	24	34	0.1029	36	8	0.0413	47	38	0.0020
12	49	0.1752	25	32	0.0983	37	46	0.0404	48	18	0.0004
13	30	0.1720									

real-world network graphs (Meghanathan, 2015). The state of Missouri has the largest EVC value (0.3697), distantly followed by the states of Nebraska (0.2605) and Tennessee (0.2546); the state of Maine has the lowest EVC value (0.0004), far away from the immediately larger value of 0.0020 for the state of Rhode Island. The largest values for the state of Missouri with respect to both the degree centrality and eigenvector centrality metrics indicate that the state of Missouri not only has the largest degree, its neighboring states also have a relatively larger degree.

Figure 3 illustrates the cumulative probability distribution of the EVC values of the vertices in the US States graph. Except the state of Missouri

(that has an EVC value much larger than the rest of the vertices), each of the other states have an EVC value that is closer to one or two other states. Thus, the distribution of the EVC values of the vertices follows a Poisson distribution. Figure 3b presents a comparison of the degree centrality and eigenvector centrality values of the vertices and the Spearman's rank-based correlation coefficient (Daniel, 2000) between the two degree-based centrality metrics is 0.80. We observe that vertices with the same degree centrality have a wide range of values for the eigenvector centrality. Though vertices with larger degree centrality appear to be more likely to have a larger eigenvector centrality, there are several vertices

Figure 3. Eigenvector Centrality of the vertices in the us states graph



for which the eigenvector centrality is relatively lower (compared to the EVC of vertices that have a relatively lower degree centrality) even if they have a higher degree centrality. For example, the state of Massachusetts (degree centrality - 5) has a much lower EVC (0.0061) than the state of Georgia (degree centrality - 3; EVC - 0.0946): primarily, attributed to the relatively higher degree of the neighbors for the state of Georgia.

Betweenness Centrality

The Betweenness Centrality (BWC) of a vertex is a measure of its presence among the shortest paths between any two vertices in the graph (Newman, 2010). The BWC of a vertex v_i is the sum of the fractions of the shortest paths between any two vertices (v_j and v_k ; $i \neq j \neq k$) that go through v_i . Quantitatively (Newman, 2010),

$$BWC(i) = \sum_{j \neq k \neq i} \frac{sp_{jk}(i)}{sp_{jk}}, \text{ where } sp_{jk} \text{ is the total}$$

number of shortest paths between vertices v_j and v_k ; $sp_{jk}(i)$ is the number of such shortest paths between v_j and v_k that go through vertex v_i . We determine the BWC of the vertices in the US States graph by implementing the Brandes' algorithm

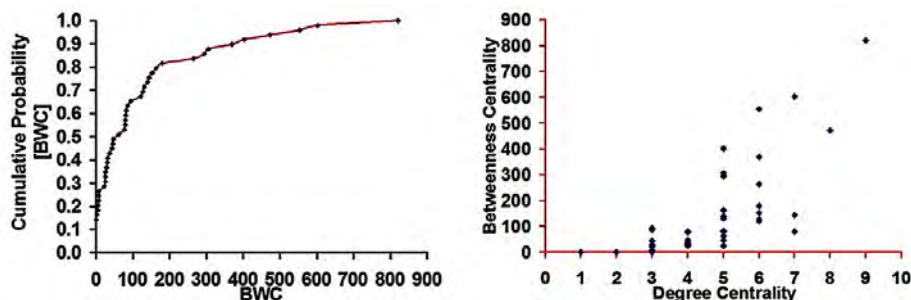
(Brandes, 2001). Table 4 ranks the vertices in the US States graph based on BWC; the state of Missouri has the largest BWC value (821.4), followed by the states of Kentucky (602.4) and Pennsylvania (554.4). Seven vertices (DC, Florida, Maine, Michigan, Rhode Island, South Carolina and Washington) have a BWC value of 0.0 - indicating that these vertices do not lie on the shortest path between any two vertices in the graph. In addition to the above tie, when rounded to the first decimal value for the non-zero values of BWC, we notice that there are ties between Delaware and New Jersey (with a lower BWC of 7.0 each). A total of 42 distinct rank values could be assigned for the vertices.

Figure 4 illustrates the cumulative probability distribution of the BWC of the vertices. We notice that about 81% of the vertices have BWC values less than 180; while the largest BWC value observed is 821.4. Thus, the BWC metric exhibits a Power-law style distribution (Balakrishnan & Nevzorov, 2003) for the vertices in the US States graph. From Figure 4b, we also notice that though vertices with a higher degree are more likely to have a higher BWC and the Spearman's rank-based correlation coefficient is 0.87; we do observe instances wherein the BWC values could vary

Table 4. Ranking of the vertices in the US states network graph based on Betweenness Centrality

Rank	ID	BWC	Rank	ID	BWC	Rank	ID	BWC	Rank	ID	BWC
1	24	821.4	14	20	140.0	26	1	47.0	38	7	7.0
2	16	602.4	15	19	131.0	27	30	45.5	38	29	7.0
3	37	554.4	16	26	127.2	28	6	43.0	39	17	4.7
4	41	472.7	17	35	121.4	29	22	36.7	40	25	4.5
5	31	403.0	18	28	94.0	30	23	31.3	41	33	1.7
6	11	368.9	19	44	86.0	31	36	30.4	42	8	0.0
7	34	305.6	20	27	81.8	32	42	28.1	42	9	0.0
8	47	294.1	21	5	80.4	33	2	25.4	42	18	0.0
9	45	264.4	22	10	79.3	34	43	25.2	42	21	0.0
10	14	179.8	23	13	78.9	35	15	25.1	42	38	0.0
11	12	163.2	24	32	76.9	36	48	21.9	42	39	0.0
12	3	152.6	25	49	61.7	37	4	7.5	42	46	0.0
13	40	144.3									

Figure 4. Betweenness Centrality of the vertices in the US states graph



appreciably even among vertices with the same degree centrality.

Closeness Centrality

The Closeness Centrality (CIC) of a vertex is a measure of the number of hops on the shortest paths from the vertex to every other vertex in the graph (Newman, 2010). The CIC of a vertex is the inverse of the sum of the lengths (hops) of the shortest paths from the vertex to the rest of the vertices in the graph (determined using the Breadth First Search algorithm; Cormen et. al., 2009). We observe the state of Missouri to have the largest CIC, followed by the states of Kentucky and Tennessee. The state of Maine has the lowest

CIC indicating that the sum of the length of the shortest paths from this state to the rest of the states is the largest. From Table 5, a total of 40 unique values could be observed for the CIC of the vertices (there are several instances where two or three states have the same values for the CIC). Figure 5a captures the cumulative distribution of the CIC; we observe the values to be uniformly distributed albeit within a smaller range (unlike the EVC and BWC) resulting in a relatively steep curve. Figure 5b captures the relationship between degree centrality and closeness centrality; vertices with a larger degree are likely to have a larger closeness centrality and the Spearman's rank-based correlation coefficient value is 0.75.

Table 5. Ranking of the vertices in the US states network graph based on Closeness Centrality

Rank	ID	CIC	Rank	ID	CIC	Rank	ID	CIC	Rank	ID	CIC
1	24	0.00787	10	15	0.00625	20	43	0.00543	29	2	0.00474
2	16	0.00763	11	32	0.00602	21	8	0.00538	30	29	0.00469
3	41	0.00730	12	19	0.00578	21	30	0.00538	31	25	0.00459
4	12	0.00680	13	23	0.00575	21	42	0.00538	32	4	0.00455
5	45	0.00667	13	37	0.00575	22	21	0.00521	33	9	0.00450
5	47	0.00667	14	1	0.00562	22	22	0.00521	34	33	0.00446
6	34	0.00654	14	10	0.00562	23	17	0.00518	35	20	0.00405
7	3	0.00649	15	5	0.00559	24	36	0.00508	36	44	0.00402
7	11	0.00649	16	40	0.00556	25	46	0.00503	37	6	0.00400
7	26	0.00649	17	48	0.00552	26	7	0.00483	38	28	0.00344
8	14	0.00645	18	49	0.00549	27	31	0.00481	39	38	0.00341
9	13	0.00633	19	27	0.00546	28	39	0.00476	40	18	0.00296
9	35	0.00633									

Figure 5. Closeness Centrality of the vertices in the US states graph

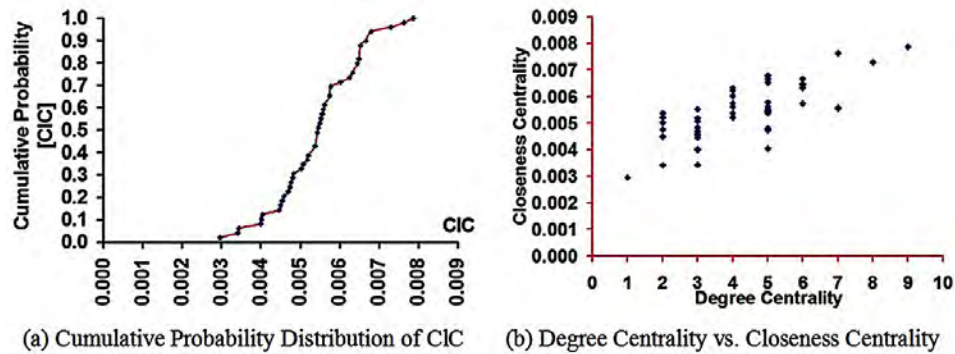
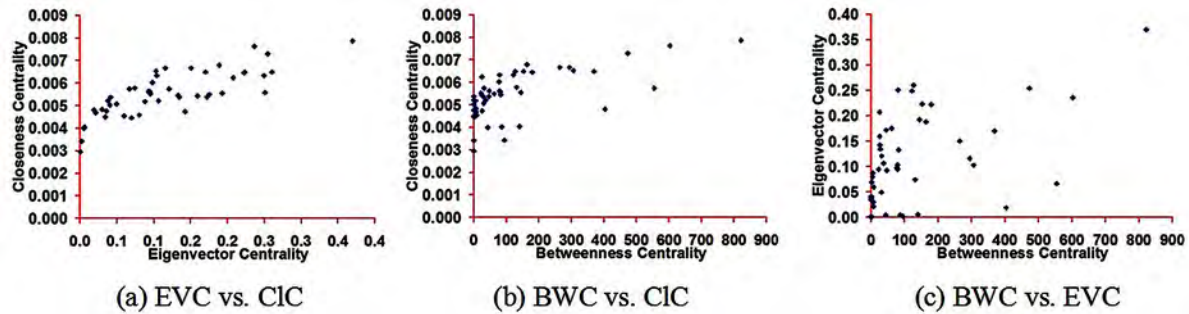


Figure 6. Relationship between Eigenvector, Closeness and Betweenness Centrality metrics for the US states graph



Ranking of Vertices based on the Centrality Metrics

Figures 3, 4 and 5 respectively capture the relationship between degree centrality and each of the other three centrality metrics (EVC, BWC and CIC). Figures 6a, 6b and 6c capture the relationship between EVC, BWC and CIC. We present the values for the Spearman's rank-based correlation coefficient between any two centrality metrics in Table 6. The larger the value of the correlation coefficient for any two metrics, the more the similarity in the ranking of the vertices with respect to the two metrics in consideration. We observe the ranking between DegC and BWC to have the highest correlation coefficient (0.87), whereas the ranking between EVC and BWC has the lowest cor-

relation coefficient (0.52). The lower correlation coefficient values for EVC-BWC and BWC-CIC metrics indicate the rankings of the vertices with respect to each of the two combinations are quite different from each other. Nevertheless, the state of Missouri has the largest value for all the four centrality metrics.

Table 6. Spearman's rank-based correlation coefficient among the Centrality metrics

	EVC	BWC	CIC
DegC	0.80	0.87	0.75
EVC		0.52	0.80
BWC			0.68

FUTURE RESEARCH DIRECTIONS

Very few works have been conducted on network graphs related to the US. We review these works below: E. A. Fogarty et al (2008) conducted a network analysis-based study on the hurricanes that made landfalls in the US from 1851 to 2008. A set of 23 non-overlapping regions (nodes) of the US that were affected with at least one hurricane were identified; two nodes were linked with an edge if at least one hurricane impacted the regions corresponding to both of them. The centrality metrics were extensively used in this study to estimate the connectivity of the regions. One of the interesting conclusions from this study was that regions (like Louisiana) with a high occurrence rate of hurricanes had a low connectivity with the rest of the regions; on the other hand, regions with high connectivity (like Virginia) had a low occurrence rate.

Lin et al (2014) conducted a network analysis of food flows within the US and had the following results: The distributions for the degree centrality and betweenness centrality were observed to be normal and Weibull (Balakrishnan & Nevzorov, 2003) in nature. A power-law relationship (Balakrishnan & Nevzorov, 2003) existed between the degree centrality and betweenness centrality metrics, indicating a vulnerability to the disturbance of key nodes. A. Lyte et al (2015) conducted a citation network-based analysis of the different sections that fall under the 52 titles of United States Code; each section is a node and there exists a directed edge from one section to another section if the former cites the latter. The betweenness and eigenvector centrality metrics were used in this study to identify major pathways of references from one section to another. The modularity-based Louvain community detection algorithm (Blondel et. al., 2008) was used to identify communities of sections that had similarities with respect to concepts and codes. It was observed that though sections under two or more related titles formed a single community, most of the communities

detected were a collection of sections under a particular title.

Cheung and Gunes (2012) conducted a complex network analysis study of the US air transportation network as of 2011 and compared it with the networks that existed in 1991 and 2001. Their study revealed no major changes in the features (like centrality and connectivity of the airports) of the air transportation networks that evolved with time (with increase in the number of airports and flight connections). A critical finding from the study was that the US air transportation network of 2011 has been identified to be more vulnerable to airport closures than it was in the past. The degree distribution of the 2011 US air transportation network only follows a partial Power-law (i.e., the distribution exhibited Power-law only after a degree value > 1), unlike the world-wide air transportation network that follows Power-law starting from degree value of 1 (Guimera et. al., 2005).

CONCLUSION

Our high-level contribution in this chapter is to illustrate centrality analysis of a connected graph of the states within a country. We implemented the algorithms to compute the centrality metrics and ran them on the US States network graph. The state of Missouri is the top-ranked node with respect to all the commonly studied centrality metrics such as degree, betweenness, closeness and eigenvector centralities. The states of Tennessee and Kentucky are in the top 2-3 rankings for at least two of the four centrality metrics. The state of Maine has the lowest ranking with respect to all the four centrality metrics, attributed to its location in the northeast corner of the country and with only one neighboring state in the US. The degree distribution appears to mimic a bi-modal Poisson distribution, while the betweenness centrality (BWC) exhibits a Power-law style distribution. The eigenvector and closeness centralities also exhibit

a Poisson-style distribution. The spectral radius ratio for node degree is 1.24 (moderately high for a Poisson network, vindicating the bi-modal degree distribution of the vertices). The approach taken in this chapter could be used to identify the most central states within a country and use this information to design the road/rail transportation networks. For countries with a reasonably larger area and an appreciable number of states, each state (except those in the corners of the country) typically shares border with a similar number of states. Hence, we anticipate the distribution of values for the node-level centrality metrics to be about the same for several other countries too. We thus opine the chapter to serve as a model for anyone interested in analyzing a connected graph of the states within a country from the perspective of node centrality.

REFERENCES

- Balakrishnan, N., & Nevzorov, V. B. (2003). *A Primer on Statistical Distributions* (1st ed.). Hoboken, NJ: Wiley-Interscience. doi:10.1002/0471722227
- Balanda, K. P., & MacGillivray, H. L. (1988). Kurtosis: A Critical Review. *The American Statistician*, 42(2), 111–119.
- Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast Unfolding of Communities in Large Networks. *Journal of Statistical Mechanics*, P10008, 1–11.
- Brandes, U. (2001). A Faster Algorithm for Betweenness Centrality. *The Journal of Mathematical Sociology*, 25(2), 163–177.
- Chung, L. L. F. (2006). *Complex Graphs and Networks*. American Mathematical Society.
- Cherven, K. (2013). *Network Graph Analysis and Visualization with Gephi*. Packt Publishing.
- Cheung, D. P., & Gunes, M. H. (2012). A Complex Network Analysis of the United States Air Transportation, *Proceedings of the International Conference on Advances in Social Network Analysis and Mining* (pp. 699–701), Istanbul, Turkey: IEEE. doi:10.1109/ASONAM.2012.116
- Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2009). *Introduction to Algorithms* (3rd ed.). Cambridge, MA: MIT Press.
- Daniel, W. W. (2000). *Applied Nonparametric Statistics* (2nd ed.). Boston, MA: Cengage Learning.
- Ding, Y. (2011). Scientific Collaboration and Endorsement: Network Analysis of Coauthorship and Citation Networks. *Journal of Informetrics*, 5(1), 187–203. doi:10.1016/j.joi.2010.10.008 PMID:21344057
- Fogarty, E. A., Elsner, F. B., Jagger, T. H., & Tsonis, A. A. (2008). Network Analysis of U. S. Hurricanes. *Springer Hurricanes and Climate Change*, 153–167.
- Fruchterman, T. M. J., & Reingold, E. M. (1991). Graph Drawing by Force-Directed Placement. *Software, Practice & Experience*, 21(11), 1129–1164. doi:10.1002/spe.4380211102
- Ghali, N., Panda, M., Hassanien, A. E., Abraham, A., & Snasel, V. (2012). Social Network Analysis: Tools, Measures and Visualization, *Computational. Social Networks*, 3–23.
- Girvan, M., & Newman, M. E. J. (2002). Community Structure in Social and Biological Networks. *Journal of the National Academy of Sciences of the United States of America*, 99(12), 7821–7826. doi:10.1073/pnas.122653799 PMID:12060727
- Guimera, R., Mossa, S., Turtschi, A., & Amaral, L. A. N. (2005). The World-Wide Air Transportation Network: Anomalous Centrality, Community Structure, and Cities Global Roles. *Journal of the National Academy of Sciences*, 102(22), 7794–7799. doi:10.1073/pnas.0407994102 PMID:15911778

Lay, D. C., Lay, S. R., & McDonald, J. J. (2015). *Linear Algebra and its Applications* (5th ed.). Pearson Publishers.

Lin, X., Dang, Q., & Konar, M. (2014). A Network Analysis of Food Flows within the United States of America. *Environmental Science & Technology*, 48(10), 5439–5447.

Lyte, A., Slater, D., & Michel, S. (2015). *Network Measures of the United States Code. Technical Report*. McLean, VA: MITRE.

Ma, X., & Gao, L. (2012). Biological Network Analysis: Insights into Structure and Functions. *Briefings in Functional Genomics*, 11(6), 434–442. doi:10.1093/bfgp/els045 PMID:23184677

Meghanathan, N. (2014). Spectral Radius as a Measure of Variation in Node Degree for Complex Network Graphs, *Proceedings of the 7th International Conference on u- and e- Service, Science and Technology*, 30-33. doi:10.1109/UNESST.2014.8

Meghanathan, N. (2015). Exploiting the Discriminating Power of the Eigenvector Centrality Measure to Detect Graph Isomorphism. *International Journal in Foundations of Computer Science and Technology*, 5(6), 1–13.

Newman, M. (2010). *Networks: An Introduction*. Oxford University Press.

Zhao, D., & Strotmann, A. (2015). *Analysis and Visualization of Citation Networks* (1st ed.). San Rafael, CA: Morgan & Claypool Publishers.

KEY TERMS AND DEFINITIONS

Betweenness Centrality: A measure of the number of shortest paths that go through the node between any two vertices in the network.

Centrality: A quantitative measure of the importance of the node on the basis of the topological structure of the network.

Closeness Centrality: A measure of the number of hops of the shortest paths from the node to every other node in the network.

Correlation Coefficient: A measure of the statistical relationship between two or more datasets.

Degree Centrality: A measure of the number of neighbor nodes for the node.

Eigenvector Centrality: A measure of the number of neighbors of the node as well as the degree centrality of the neighbor nodes.

Context-Aware Approach for Restaurant Recommender Systems

D**Haoxian Feng***University of Ottawa, Canada***Thomas Tran***University of Ottawa, Canada*

INTRODUCTION

A recommender system provides online consumers with product information and recommendations through online e-business website to in order to enhance customers' shopping intention and improve the customers' shopping experience. In the specific area of restaurant recommendations, recommender systems aim to find the best restaurant that meets a user's need such as food type, environment type and so on.

The Chicago Restaurants Union has been using a FindMe type (Burke, Hammond, & Yound, 1997) recommender system called Entrée, which is used to recommend restaurants to the customers. This recommender system makes use of a typical recommendation algorithm based on knowledge (Sarwar, Karypis, Konstan, & Riedl, 2001, pp.285-295), which only uses the online operations of users as a basis, and is therefore lack of using historical data of users. This paper uses a context-aware approach to consider users' historical data for recommender systems that provide suggestions for restaurants.

The rest of this paper is organized as follows: Section 2 provides background information restaurant recommender systems. Section 3 presents the recommender algorithms used in our experiments. Section 4 describes the Entrée dataset; present data scrubbing process and the user behavior mining on this dataset; report the experimental results. Section 5 discusses the value of the results, their potential applications, and challenges and weak-

nesses. Section 6 concludes the paper and proposes some future research directions.

BACKGROUND AND RELATED WORK

Upsurge of Recommender Systems

In the 90s of the last century, the field of electronic commerce began to flourish and personalized recommender system also rose its first wave. Some large e-commerce sites such as Amazon launched their own personalized recommender system. Literature reported that 35 percent of Amazon's incremental sales came from their recommender system (Figiel, Epstein, McDonald, 1998, pp.20-25). At the same time, a sensational paper published by Amazon in 2000, entitled "Item-based collaborative filtering recommendation algorithms" (Sarwar, et al., 2001), has become one of the most famous literature in this area. Ever since then, personalized recommendation technology has become an irresistible trend.

Restaurant Recommender System: The Entrée System

As mentioned above, in the era of the Internet economy, recommender systems have been widely used in various business areas, which also include food industry. Burke (1999) stated the restaurant recommender system, Entrée, "makes its recom-

DOI: 10.4018/978-1-5225-2255-3.ch153

mentations by finding restaurants in a new city similar to restaurants the user knows and likes” (p.69).

Entrée is a recommender system using FindMe approach to assist users browsing which allows users to navigate by stating their preferences with respect to a given restaurant, thereby refining their search criteria.

When using this system, firstly a user needs to submit an “entry point” which is a known restaurant. (Or, the system will give a default “entry point” for new users.) Then the system offers a range of criteria, and according to the criteria chosen by the user, it will display a few restaurants that meet the users’ need. At the same time, the user would set their preferences until an acceptable restaurant is recommended.

As shown in Figure 1, the Entrée system finds a restaurant called “BrauhoF” that is similar to a restaurant called “Biergasthof” that the user used to choose. Then, if the user wants to find one that is similar to ” BrauhoF ” but slightly cheaper, she can click the “Less \$\$” button. Figure 2 shows the result of such action. In the same way, the user can choose cheaper, nicer or more traditional restaurants by clicking the corresponding buttons. The system will refresh the recommendation results based on the user’s operation, and finally help her to find the restaurant she desires.

The recommendation technology that Entrée use is knowledge-based similarity search. There are two basic search modes: one is based on the similarity retrieval, and the other is based on users’ evaluation result. When finding some similar

Figure 1. Entrée: A FindMe type system (Jannach, Zanker, Felfernig, & Friedrich, G. 2010)



Figure 2. Entrée: A FindMe type system (Jannach et al., 2010)



restaurants, the user selects in a given item directory, and requests for other similar items.

However, this traditional recommender system only uses the online operations of the users as a basis with just a few options like cheaper, quieter, etc. So, it is far from meeting the real needs of users. The main reason for this problem is the insufficient use of historical data of the users. Robin Burk and Hammond and Yound (1997) proposed the concept of hybrid recommendation models based on the original knowledge-based recommendation and designed an improved the Entrée system. He firstly mapped the navigation information of the Entrée system to a series of implicit ratings, and user-based collaborative filtering algorithm (Wang, De Vries, & Reinders, 2006) was then used to get the prediction score. It proved to achieve better results.

Related Work

Collaborative Filtering

A common approach for recommender systems is collaborative filtering (Breese, Heckerman, & Kadie, 1998). In general, it recommends the content that the users may be interested in according to the preference of the group sharing the same hobbies. Collaborative filtering analyzes the responses of users' preferences for goods and the main way to respond is rating or social filtering (Shardanand & Maes, 1995).

There are two very popular methods to compute the similarity measure $\text{sim}(u, u')$ between users in collaborative recommender systems. One is a correlation-based approach (Jannach et al., 2010):

$$sim(x, y) = \frac{\sum_{s \in S_{xy}} (r_{x,s} - \bar{r}_x)(r_{y,s} - \bar{r}_y)}{\sum_{s \in S_{xy}} (r_{x,s} - \bar{r}_x)^2 \sum_{s \in S_{xy}} (r_{y,s} - \bar{r}_y)^2} \quad (2.1)$$

where the $r_{x,s}$ and $r_{y,s}$ are the ratings of the items given by the user x and y , S_{xy} is the set of all items that rated both by the user x and y .

Another one is a cosine-based approach (Jan-nach et al., 2010):

$$sim(\vec{x}, \vec{y}) = \frac{\vec{x} \cdot \vec{y}}{|\vec{x}| * |\vec{y}|} \quad (2.2)$$

where the $\cdot$ means the dot product of vectors, $|\vec{x}|$ and $|\vec{y}|$ are the Euclidian length of the vectors.

There are also some specific calculations. For example, Oku and Nakajima and Miyazaki and Uemura (2006) proposed a context-aware SVM methods by using SVM to calculate the similarity between restaurants based on the context parameters of users' and characters of restaurants.

Sarwar et al. (2001) showed that item-based collaborative filtering can outperform the original user-based collaborative filtering.

Context-Aware Recommender Systems

This section discusses recommender systems based on context-aware (Foltz, P. W. 1990), which may consider the time, location or other factors when making recommendations. The traditional recommender system only considers two dimensions – users and items:

$$R : \text{User} * \text{Item} \rightarrow \text{Rating} \quad (2.3)$$

But context-aware recommender systems would take context information into account:

$$R : \text{User} * \text{Item} * \text{Context} \rightarrow \text{Rating} \quad (2.4)$$

There are 3 main methods in context-aware recommender systems:

1. Contextual pre-filtering, which selects only the information and data related to context for building recommender systems;
2. Contextual post-filtering, which filters the results of recommender system and retains only the results that contain contextual information; and
3. Contextual modeling, which uses contextual information directly in the recommendation algorithm.

The figure of the above methods is shown as Figure 3.

Contextual pre-filtering uses context information to filter when choosing the data for recommender system by using the current context-aware information to make a match.

Musto and Semeraro and Lops and de Gemmis (2013) extended the eVSM (enhanced vector space model) which is exactly a pre-filtering type context-aware recommendation.

Contextual post-filtering considers the contextual information only when the recommended item list is generated, then it will consider context information to filter out irrelevant items and adjust the order of items in the list. This process is illustrated in Figure 4.

Contextual post-filtering can be divided into two types, namely heuristic-based and model-based contextual post-filtering. Heuristic-based approach filters and changes the order according to some common characteristics of a user in the certain scenario. And model-based approach builds a model to estimate the probability that a user selects a certain item in a certain context.

Despite the fact that contextual pre-processing and contextual post-processing can be realized by the two-dimensional recommendation function, the contextual model can be extended

Figure 3. Paradigms for incorporating context in recommender systems: (a) Contextual pre-filtering; (b) contextual post-filtering; (c) contextual modeling (Abbar, Bouzeghoub, & Lopez, 2009)

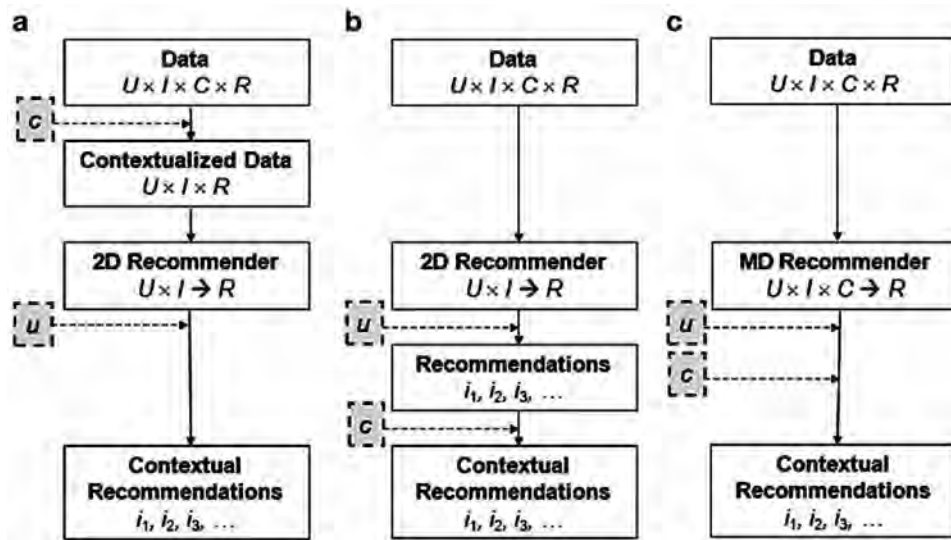
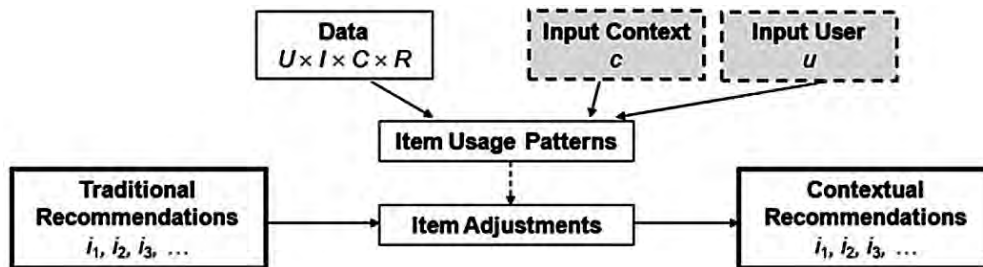


Figure 4. Contextual post-filtering approach: recommendation list adjustment (Abbar et al., 2009)



to higher dimensional recommendation model. Lee and Teng (2007) proposed collaborative recommender with multi-criteria decision analysis with a dataset from Zagat. It considers multiple aspects of the restaurant include food, decor, service and price. Zhang et al. (2016) proposed a model that uses both implicit (e.g. demographic and geographic information) and explicit (e.g. ratings and reviews) preferences of users to make restaurant recommendations. For example, a lot of restaurant recommendation websites, such as Yelp, Zagat and Dianping, are using Location-Based Services to both website and mobile users.

PROPOSED APPROACH

Approach Overview

This paper builds a recommender system for which the underlying algorithm is the bias-based SVD algorithm, which is a matrix factorization algorithm enhanced with baseline predictors and stochastic gradient descent process. In the next section, the authors will compare the performance of the bias-based SVD algorithm with the item-based collaborative algorithm.

In the coming subsections, the authors are going to provide detailed descriptions of the item-based

collaborative filtering algorithm, as well as the algorithms involved in bias-based SVD method.

Item-Based Collaborative Filtering

The K-nearest neighbor algorithm (Horton & Nakai, 1997) is usually used in item-based collaborative filtering algorithm. It computes the similarity between the two items through the nearest adjacency matrix and then produces the predicted value for the user u using the weighted average approach. In addition, mechanisms like Pearson correlation, vector cosine similarity as (2.2), etc. can be added in this algorithm to improve the prediction result.

The advantage of this method is that it is easy to interpret the results, which is one of the most important features of the recommender system. But its shortcomings are also obvious: First, it needs user's score. Secondly, if the data is sparse, its performance will become poor.

This paper uses the Pearson correlation and obtains the prediction using similarity as a weight to calculate the weighted summation.

SVD- Singular Value Decomposition

Matrix Factorization

The mathematical foundation of the matrix factorization algorithm (Ma, Yang, Lyu, & King, 2008) is the transformation of matrices. They have $A = pEq = pq$, where A is a user-item rating matrix, E is a diagonal matrix, p can be viewed as the feature matrix of users, and q as the feature matrix of items (Koren, Bell, & Volinsky, 2009).

Readers can imagine that each line of the matrix represents a user and each column represents an item, and the numbers in cells represent the ratings of each item from each user. For example:

$$\begin{matrix} 5 & 2 & 4 & 4 & 3 & u_{11} & u_{12} & i_{11} & i_{12} & i_{13} \\ 3 & 1 & 2 & 4 & 1 & u_{21} & u_{22} & i_{21} & i_{22} & i_{23} \\ 2 & & 3 & 1 & 4 & u_{31} & u_{32} & i_{31} & i_{32} & i_{33} \end{matrix} \quad (2.5)$$

This is how matrix factorization works to deal with missing values in the matrix. The blank one is the missing value. Our goal is to find those two matrices on the right side by using existing ratings. More intuitively, researchers can also regard the rating matrix as the dot product of p and q as:

$$\widehat{r_{ui}} = p_u^T q_i \quad (2.6)$$

where the p_u^T and q_i are the same as p and q .

Baseline Predictors

Matrix factorization is using the hidden class to connect the users and items. But in some cases, the rating system has some inherent attributes of users and items while items or users also have some inherent attributes between each other. In order to settle this shortage, Koren (2009) proposed a baseline predictor to deal with this problem.

They set μ to be the mean value of all the ratings of the training set. They use b_i and b_u to represent the preference of items and users and $\widehat{r_{ui}}$ represent the predict rating of user u for item i . In particular, b_u represents the rating tendency of user that some users prefer to rate higher while some do not. b_i represents the rated pattern of the particular item in the dataset.

$$\widehat{r_{ui}} = \mu + b_i + b_u \quad (2.7)$$

These two parameters can be calculated as an optimization problem such as the least square procedure:

$$\min_{b^*} \sum_{(u,i) \in \mathbb{R}} (r_{ui} - \mu - b_i - b_u)^2 + \lambda_1 (\sum_u b_u^2 + \sum_i b_i^2) \quad (2.8)$$

where b^* represents all biases including user and item, r_{ui} represent the observed rating of user u for item i and $R(i)$ or $R(u)$ represent the whole set of item or user. The first term, the regulariza-

tion parameter, λ_1 , avoids overfitting by penalizing the magnitudes of the parameters (Koren 2009).

Bias-Based SVD

In this paper, the authors use SVD with baseline predictor, which is the combination of matrix factorization and baseline predictors. The authors call it bias-based SVD.

As an evolution of equation 2.6 and the matrix factorization methods, the rating model can be improved as (Koren et al., 2009):

$$\hat{r}_{ui} = \mu + b_i + b_u + p_u^T q_i \quad (2.9)$$

In this case, researchers use this loss function:

$$\min_{b^+, q^+, p^+} \sum_{(u,i) \in \mathbb{R}} (r_{ui} - \mu - b_i - b_u - p_u^T q_i)^2 + \lambda_2 (\sum_u b_u^2 + \sum_i b_i^2 + \|q_i\|^2 + \|p_u\|^2) \quad (2.10)$$

$\lambda_2 (\sum_u b_u^2 + \sum_i b_i^2 + \|q_i\|^2 + \|p_u\|^2)$ is the penalty term.

And researchers use stochastic gradient descent to optimize equations 2.11 - 2.14:

$$b_i \leftarrow b_i + \gamma(e_{ui} - \lambda_2 * b_i) \quad (2.11)$$

$$b_u \leftarrow b_u + \gamma(e_{ui} - \lambda_2 * b_u) \quad (2.12)$$

$$q_i \leftarrow q_i + \gamma(e_{ui} * p_u - \lambda_2 * q_i) \quad (2.13)$$

$$p_u \leftarrow p_u + \gamma(e_{ui} * q_i - \lambda_2 * p_u) \quad (2.14)$$

where those four parameters need to be pushed forward along the steepest descent direction. γ is learning rate which needs to be chosen according to the experiment. The stochastic gradient descent algorithm is the basic optimization algorithm in optimization theory. It finds the steepest descent direction by the partial derivative of the parameter.

Then the authors have the error formula as 2.15:

$$e_{ui} = r_{ui} - \hat{r}_{ui} = r_{ui} - p_u^T q_i = r_{ui} - \sum_{k=1}^k p_{uk} q_{ki} \quad (2.15)$$

EXPERIMENT AND EVALUATION

Dataset

In this paper, the authors use the Entrée dataset. The researchers collected the data of user behaviors from September 1996 to April 1999 and organized them according to year quarter (Q). Specifically, in Q3 1996 and Q2 1999 each contains only one month. This dataset tagged 9 users' behavior records (UC Irvine Machine Learning Repository, 2000) from L to T, representing the following information respectively [Burke et al., 1997; Breese et al., 1998]:

{ 'Browse', 'Entry Point', 'End Point', 'Nicer', 'More Traditional,' 'More Creative', 'Cheaper', 'Livelier', 'Quieter', 'Change Cuisine' }.

The original dataset looks like the website log. It documents every click of the users. Each line in the dataset represents a session of a user's interaction with the system. The (tab-separated) fields are as follows:

Date, IP, Entry point, Rated restaurant1,..., Rated restaurantN, Endpoint

The dataset has a total of 50672 lines like the line above.

User Behavior Mining

To better understand the model, the researchers consider the semantic information as some specific ratings (Yu, Nakamura, Jang, Kajita, S., & Mase, 2007; Foltz, 1990) include both explicit evaluation (Carroll & Swain, 1993) and implicit evaluation (Joachims, Granka, Pan, Hembrooke, & Gay, 2005).

Explicit Behavior

1. The content that users are interested in:
 - a. Here, Entry Point is entered by the user, then the system will recommend some restaurants similar to the Entry Point restaurant, according to the characteristics of this restaurant. It is very likely that the user will choose their most ideal restaurant as the standard to find similar candidates. Therefore, for this kind of restaurant, the authors use a function $X(u, r) = 1$ to express the strongest interest rate.
2. The content that users are not interested in:
 - a. For evaluations such as 'Nicer', 'More Traditional', 'More Creative', 'Cheaper', 'Livelier', 'Quieter', 'Change Cuisine', which states that the users do not like the restaurant, and the restaurant will not be their ideal candidate at least for a long period. Therefore, authors score them as $X(u, r) = -1$, which represents a strong discomfort.

Table 1. Behavior-Score mapping

Behaviors	Scores
Browse	-0.5
Change Cuisine	-1
Cheaper	-1
Nicer	-1
More Creative	-1
More Traditional	-1
Livelier	-1
Quieter	-1
Entry Point	1
End Point	0.8

Implicit Behavior

For Exit Point, the user would confirm the end of the search on this "exit" (for some users that directly log out without confirming the Exit Point, the system will not record this value). If there is an Exit Point, the authors consider that the user is very likely to choose this restaurant. But this is not for sure because they may just follow the ending process. Therefore, for such implicit behavior, the authors can only give $X(u, r) = 0.8$, assuming that the user is very likely to choose this kind of restaurant.

For another behavior, 'Browse', while users do not give the restaurant a negative evaluation directly, they do not seem to choose this kind of restaurant eventually. Therefore, authors give $X(u, r) = -0.5$.

The authors summarize these points in Table 1.

After finishing the behavior-score mapping process authors have Table 2.

Table 2. Original dataset

User	Restaurants	Positive Ratings	Negative Ratings
22396	1149	50583	189291

Data Scrubbing and Pre-processing

The data are recorded over 20,000 users' information, many of which are too sparse, for example, some users browsed very quickly and departed directly in a few minutes. Therefore, the authors need to preprocess the data properly:

1. The number of evaluation restaurants of our users must be more than 15 so that the authors can have a less sparse rating matrix and make sure every user is in the rating matrix have enough information for recommender systems.
2. Users may use the system many times in a short period. So, the authors consider the same user in different quarters as different users.
3. Users may rate the same restaurant many times. The authors documented every time of it and calculated the sum of all the ratings of each user. The authors set the rating to be 1 if the calculated sum is larger than 1 and set the rating to be -1 if the calculated sum is smaller than -1. As such, the authors make use of every single rating of the users to make a better prediction.
4. In order to fit in the algorithm, the authors move the range of the ratings (-1,1) to (1,3) by adding 2 points to each rating.
5. For restaurants, researchers require at least three users have been evaluated it. Because some users have been removed in step 1, the number of restaurants will have to be reduced.

After the preprocessing, the information of our dataset is shown in Table 3.

Table 3. Dataset after pre-processing

User	Restaurants	Positive Ratings (2,3)	Negative Ratings (1,2)
4089	672	15975	64245

Results and Evaluation

Evaluation Metrics

MAE, standing for Mean Absolute Error, is a measurement to measure how close the predictions are to the real value.

$$MAE = \frac{\sum_{u,i \in T} |r_{ui} - \hat{r}_{ui}|}{|T|}$$

RMSE, Root-Mean-Square Error, is a measurement to the square root of the ratio of the difference between observed value and the true value and the observation times.

$$RMSE = \sqrt{\frac{\sum_{u,i \in T} |r_{ui} - \hat{r}_{ui}|^2}{|T|}}$$

In the above two formulas, u represent users, i represent items, r_{ui} represents the rating of an item i from the user u , and $\hat{r}_{ui}$ represents the predicted rating calculated by the recommender system.

They are both accuracy measurements and scale-dependent which cannot be used to make comparisons between series that are on different scales (Hyndman & Athanasopoulos, 2014).

Mean absolute error weights all individual differences equally while root mean squared error gives relatively high weight to large errors and punishes large errors more severely.

Table 4. Errors of various iteration times

Iterations	1	5	10	15	20	25
RMSE	0.578135	0.56587	0.560847	0.557987	0.556211	0.555176
MAE	0.431502	0.422098	0.41853	0.416308	0.415167	0.414898
Iterations	30	35	40	45	50	
RMSE	0.554555	0.554171	0.553933	0.553794	0.553711	
MAE	0.413728	0.413546	0.413543	0.413324	0.413243	

Process and Results

The authors run the dataset after pre-processing on both the item-based collaborative filtering algorithm and the bias-based SVD algorithm, and separate the training set and the test set in ratio 8:1.

In order to find an appropriate iteration time for the bias-based SVD algorithm, the authors have done some pre-training (see Table 4).

In Figure 5, the x-axis represents the iteration times, and the y-axis represents the test set's RMSE and MAE value. The RMSE and MAE values have a significant drop at iteration times $k = 5$ then starts to decay much more gently. And both of them almost convergence when k is larger than 20. Hence, the authors chose 20 as the iteration times for the next SVD experiments.

The authors train the item-based and SVD model for each quarter and the whole data set.

Table 5. RMSE and MAE value of the whole test dataset

	Item-Base	Bias-Based SVD
RMSE	0.615462	0.587139
MAE	0.461834	0.438032

Finally, the authors obtain the RMSE and MAE values as follows:

It can be seen in Table 5 that the bias-based SVD algorithm outperforms the item-based collaborative filtering one in both evaluation criteria: 4.6% in RMSE and 5.2% in MAE. These are the results that the authors hoped for. Note that the dataset has a lot of missing values, which significantly affect the correlation calculation. Although the data scrubbing process set a threshold of 15 ratings per user, there are still a lot of missing

Figure 5. Error-Steps values line chart

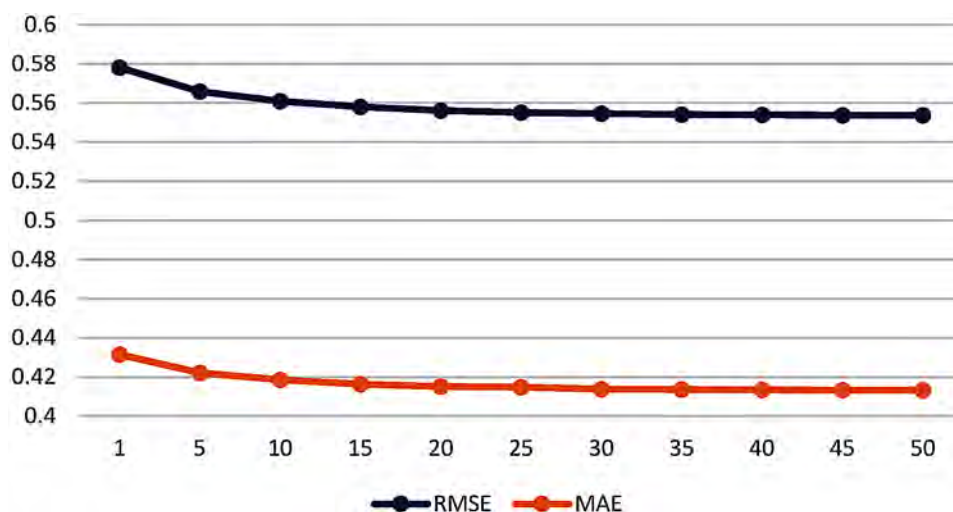
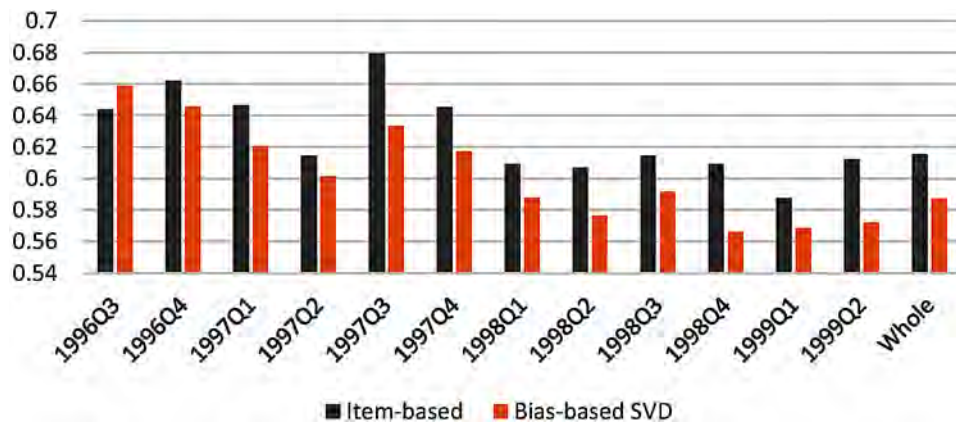


Figure 6. RMSE value for two different models



values in the dataset. But the bias-based SVD method can fulfill all those missing values and help with the prediction.

Also, the authors compare each quarter of year's performance from Figures 6 and 7.

As shown in Figure 6 and Figure 7, the RMSE values show that bias-based SVD outperforms the item-based collaborative filtering method in every quarter, except only for the third quarter in 1996. This shows that bias-based SVD algorithm has stronger capabilities to deal with missing values.

Also, both methods show the relatively high quality of prediction. Thus, using the transformation from the user behaviors to ratings, researchers

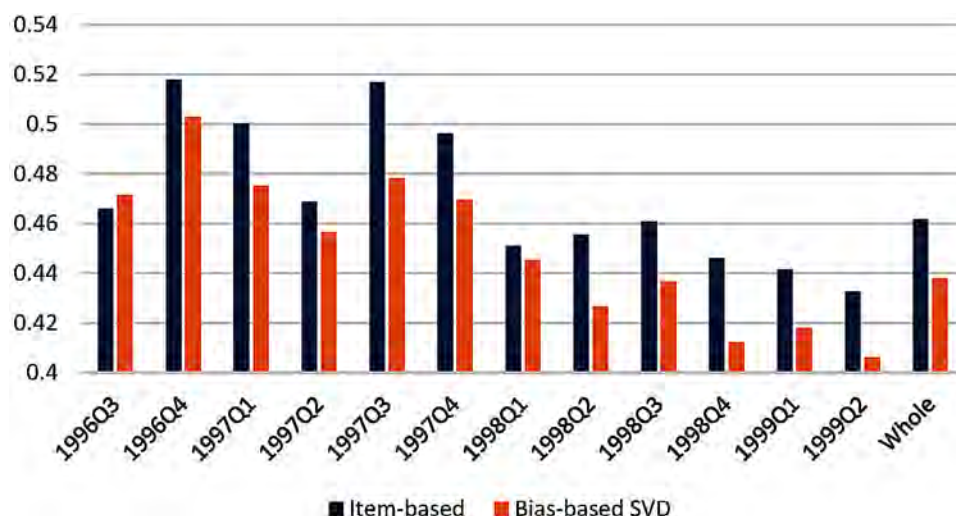
can empower those classical recommendation algorithms to achieve good results.

DISCUSSION

Bias-based SVD methods had a great success in Netflix Prize on the movie or video recommendation systems. And it now turns out to have very good performance in this restaurant dataset.

It is encouraging to learn that the bias-based SVD method achieves very good performance on the Entrée dataset after all those data preprocessing steps. Actually, this kind of transformation from

Figure 7. MAE value for two different models



user behaviors to ratings can be used in different domains. For example, in the music recommendation domain, developers can use a tag system to allow users to tag the music they heard. Those tags can form the context of the music dataset to which the process described in this paper can be re-applied. Also, this method can be used in web browsing record-based recommendation systems.

But since the Entrée system is a knowledge-based system originally, the transformation from user behaviors to ratings is based on the assumption by the researchers, which may not reflect user personal preferences precisely. This is because there are some opposite behaviors among all the behaviors, such as ‘nicer’ with ‘cheaper’ etc., which can weaken the relation or the opposition between the behaviors.

CONCLUSION

The internet is changing the way of people’s thinking, their habits of life and the relationship between them, and is building unprecedented business and social models. Our information is growing in geometric exponential level, but our ability to obtain information remains the same as our ancestors’. The information explosion has made our life comfortable but also created heavy work for us to find the things we want. The emergence of recommender systems meets our need in this new era.

This paper uses a bias-based SVD method for restaurant recommendations and compares it with the traditional item-based collaborative filtering algorithm, both operating on the Entrée dataset after some pre-processing steps. The authors consider different user behaviors and transform them into ratings, which are then fed as input into the recommender systems. The results show that the bias-based SVD method outperforms the item-based collaborative filtering algorithm although both methods attain relatively high prediction quality thanks to the behavior-score mapping and pre-processing.

FUTURE RESEARCH DIRECTIONS

As mentioned in the Discussion section, a shortcoming of our approach is in the context-aware process. In the future, the authors can take the opposite user behaviors into account. The authors find that “Nicer”/ “Cheaper”, “Livelier”/ “Quieter”, “more Traditional”/ “more Creative” are opposite to each other. Considering those behaviors in pairs may lead to a better result. The authors can also use the characteristic of restaurants to help build a better dataset to represent the preferences of users. In addition, the authors can use Bayesian Personalized Ranking (Rendle, Freudenthaler, Gantner, & Schmidt-Thieme, 2009) to improve the process of implicit information. Also, the authors plan to look into alternative methods for better evaluating the value of our proposed approach in the context of related works.

REFERENCES

- Abbar, S., Bouzeghoub, M., & Lopez, S. (2009, August). Context-aware recommender systems: A service-oriented approach. In *VLDB PersDB workshop* (pp. 1-6). VLDB.
- Breese, J. S., Heckerman, D., & Kadie, C. (1998, July). Empirical analysis of predictive algorithms for collaborative filtering. In *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence* (pp. 43-52). Morgan Kaufmann Publishers Inc.
- Burke, R. (1999, July). Integrating knowledge-based and collaborative-filtering recommender systems. In *Proceedings of the Workshop on AI and Electronic Commerce* (pp. 69-72). Academic Press.
- Burke, R. D., Hammond, K. J., & Yound, B. C. (1997). The FindMe approach to assisted browsing. *IEEE Expert*, 12(4), 32–40. doi:10.1109/64.608186

- Carroll, S., & Swain, M. (1993). Explicit and implicit negative feedback. *Studies in Second Language Acquisition*, 15(03), 357–386. doi:10.1017/S0272263100012158
- Figiel, G. S., Epstein, C., McDonald, W. M., Amazon-Leece, J., Figiel, L., Saldivia, A., & Glover, S. (1998). The use of rapid-rate transcranial magnetic stimulation (rTMS) in refractory depressed patients. *The Journal of Neuropsychiatry and Clinical Neurosciences*, 10(1), 20–25. doi:10.1176/jnp.10.1.20 PMID:9547462
- Foltz, P. W. (1990, March). Using latent semantic indexing for information filtering. *ACM SIGGROUP Bulletin*, 11(2-3), 40–47. doi:10.1145/91478.91486
- Horton, P., & Nakai, K. (1997, June). Better Prediction of Protein Cellular Localization Sites with the it k Nearest Neighbors Classifier. In ISMB (Vol. 5, pp. 147-152). ISMB.
- Hyndman, R. J., & Athanasopoulos, G. (2014). *Forecasting: principles and practice*. OTexts.
- Irvine Machine Learning Repository, U. C. (2000). *Entrée Chicago recommendation data dataset* [Data file]. Available from <https://archive.ics.uci.edu/ml/datasets/Entree+Chicago+Recommendation+Data>
- Jannach, D., Zanker, M., Felfernig, A., & Friedrich, G. (2010). *Recommender systems: an introduction*. Cambridge University Press. doi:10.1017/CBO9780511763113
- Joachims, T., Granka, L., Pan, B., Hembrooke, H., & Gay, G. (2005, August). Accurately interpreting clickthrough data as implicit feedback. In *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 154-161). ACM. doi:10.1145/1076034.1076063
- Koren, Y. (2009). The bellkor solution to the netflix grand prize. *Netflix Prize Documentation*, 81, 1-10.
- Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer*, 42(8), 30–37. doi:10.1109/MC.2009.263
- Lee, H. H., & Teng, W. G. (2007, August). Incorporating multi-criteria ratings in recommendation systems. In *2007 IEEE International Conference on Information Reuse and Integration* (pp. 273-278). IEEE. doi:10.1109/IRI.2007.4296633
- Ma, H., Yang, H., Lyu, M. R., & King, I. (2008, October). Sorec: social recommendation using probabilistic matrix factorization. In *Proceedings of the 17th ACM conference on Information and knowledge management* (pp. 931-940). ACM. doi:10.1145/1458082.1458205
- Musto, C., Semeraro, G., Lops, P., & de Gemmis, M. (2013, August). Contextual evsm: A content-based context-aware recommendation framework based on distributional semantics. In *International Conference on Electronic Commerce and Web Technologies* (pp. 125-136). Springer Berlin Heidelberg. doi:10.1007/978-3-642-39878-0_12
- Oku, K., Nakajima, S., Miyazaki, J., & Uemura, S. (2006, May). Context-aware SVM for context-dependent information recommendation. In *Proceedings of the 7th international Conference on Mobile Data Management* (p. 109). IEEE Computer Society.
- Rendle, S., Freudenthaler, C., Gantner, Z., & Schmidt-Thieme, L. (2009, June). BPR: Bayesian personalized ranking from implicit feedback. In *Proceedings of the twenty-fifth conference on uncertainty in artificial intelligence* (pp. 452-461). AUAI Press.
- Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001, April). Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th international conference on World Wide Web* (pp. 285-295). ACM.

Shardanand, U., & Maes, P. (1995, May). Social information filtering: algorithms for automating “word of mouth”. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 210-217). ACM Press/Addison-Wesley Publishing Co. doi:10.1145/223904.223931

Wang, J., De Vries, A. P., & Reinders, M. J. (2006, August). Unifying user-based and item-based collaborative filtering approaches by similarity fusion. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 501-508). ACM. doi:10.1145/1148170.1148257

Yu, Z., Nakamura, Y., Jang, S., Kajita, S., & Mase, K. (2007, July). Ontology-based semantic recommendation for context-aware e-learning. In *International Conference on Ubiquitous Intelligence and Computing* (pp. 898-907). Springer Berlin Heidelberg. doi:10.1007/978-3-540-73549-6_88

Zhang, F., Yuan, N. J., Zheng, K., Lian, D., Xie, X., & Rui, Y. (2016, April). Exploiting Dining Preference for Restaurant Recommendation. In *Proceedings of the 25th International Conference on World Wide Web* (pp. 725-735). International World Wide Web Conferences Steering Committee.

Baltrunas, L., & Ricci, F. (2014). Experimental evaluation of context-dependent collaborative filtering using item splitting. *User Modeling and User-Adapted Interaction*, 24(1-2), 7-34. doi:10.1007/s11257-012-9137-9

Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted Interaction*, 12(4), 331-370. doi:10.1023/A:1021240730564

Chen, L., & Pu, P. (2012). Critiquing-based recommenders: Survey and emerging trends. *User Modeling and User-Adapted Interaction*, 22(1-2), 125-150. doi:10.1007/s11257-011-9108-6

De Lathauwer, L., De Moor, B., & Vandewalle, J. (2000). A multilinear singular value decomposition. *SIAM Journal on Matrix Analysis and Applications*, 21(4), 1253-1278. doi:10.1137/S0895479896305696

McCarthy, J. F. (2002, April). Pocket restaurant-finder: A situated recommender system for groups. In *Workshop on Mobile Ad-Hoc Communication at the 2002 ACM Conference on Human Factors in Computer Systems*.

Thompson, C. A., Goker, M. H., & Langley, P. (2004). A personalized system for conversational recommendations. *Journal of Artificial Intelligence Research*, 21, 393-428.

ADDITIONAL READING

Adomavicius, G., Sankaranarayanan, R., Sen, S., & Tuzhilin, A. (2005). Incorporating contextual information in recommender systems using a multidimensional approach. *ACM Transactions on Information Systems*, 23(1), 103-145. doi:10.1145/1055709.1055714

Badeau, R., & Boyer, R. (2008). Fast multilinear singular value decomposition for structured tensors. *SIAM Journal on Matrix Analysis and Applications*, 30(3), 1008-1021. doi:10.1137/060655936

KEY TERMS AND DEFINITIONS

Baseline Predictors: A method that considers the user bias, the item bias and the global average of ratings in the dataset to provide better recommendation results.

Behavior-Score Mapping: A transformation process from the behavior to the ratings under some rules.

Context-Aware Recommender System: A recommender system that considers not only the ratings but also includes other information such as time, emotion and so on.

Gradient Descent: A fundamental optimization algorithm that is used to find the steepest descent direction and the local minimum.

Item-Based Collaborative Filtering: A classical and most widely used recommendation algorithm that uses the similarities between items.

Matrix Factorization: A way to decompose a matrix to the product of several matrices.

Singular Value Decomposition (SVD): One of the matrix factorization methods to decompose a matrix to the product of two orthogonal matrices and one diagonal matrix.

D

Data-Centric Benchmarking

Jérôme Darmont

Université de Lyon, Lyon 2, ERIC EA3083, France

INTRODUCTION

In data management, both system designers and users casually resort to performance evaluation. On one hand, designers need to test architectural features and hypotheses regarding the actual (vs. theoretical) behavior of a system, especially in terms of response and scalability. Performance tuning also necessitates accurate performance evaluation. On the other hand, users are also keen on comparing the efficiency of different technologies before selecting a software solution. Thence, performance measurement tools are of premium importance in the data management domain.

Performance evaluation by experimentation on a real system is generally referred to as benchmarking. It consists in performing a series of tests on a given system to estimate its performance in a given setting. Typically, a data-centric benchmark is constituted of two main elements: a data model (conceptual schema and extension) and a workload model (set of read and write operations) to apply on this dataset, with respect to a predefined protocol. Both models may be parameterized. Most benchmarks also include a set of simple or composite performance metrics such as response time, throughput, number of input/output operations, disk or memory usage, etc.

The Transaction Processing Performance Council (TPC), a non-profit organization founded in 1988, plays a preponderant role in data-centric benchmarking. Its mission is to issue standard benchmarks, to verify their correct application by the industry, and to publish performance test results. TPC members include all the major industrial actors from the database field.

The aim of this chapter is to present an overview of the major past and present state-of-the-art

data-centric benchmarks. Our review includes the TPC standard benchmarks, but also alternative or more specialized benchmarks. We survey benchmarks from three families: transaction benchmarks aimed at On-Line Transaction Processing (OLTP), decision-support benchmarks aimed at On-Line Analysis Processing (OLAP) and big data benchmarks. Eventually, we discuss the issues, tradeoffs and future trends in data-centric benchmarking.

BACKGROUND

Transaction Processing Benchmarks

The first TPC benchmark for relational, transactional databases, TPC-C (TPC, 2010), has been in use since 1992. TPC-C features a complex business database (a classical customer-order-product-supplier model with nine types of tables bearing various structures and sizes) and a workload of diversely complex transactions that are executed concurrently. The performance metric in TPC-C is transaction throughput. As all TPC benchmarks, TPC-C's only parameter is a scale factor SF that determines data size. TPC-C was complemented in 2007 by TPC-E (TPC, 2015a), which simulates a brokerage firm with the aim of being representative of more modern OLTP systems. In its principles and features, TPC-E is otherwise very similar to TPC-C.

There are few alternatives to TPC-C and TPC-E for relational applications. Yet, some benchmarks fit niches where there is no standard benchmark. For instance, OO7 (Carey et al., 1993) and OCB (Darmont & Schneider, 2000) are object-oriented database benchmarks modeling engineering applications, e.g., computer-aided design or software

DOI: 10.4018/978-1-5225-2255-3.ch154

engineering. However, their complexity makes both these benchmarks hard to understand and implement. Moreover, with objects in databases being more commonly managed in object-relational systems nowadays, object-relational benchmarks such as BUCKY (Carey et al., 1997) and BORD (Lee et al., 2000) now seem more relevant. Such benchmarks focus on queries implying object identifiers, inheritance, joins, class and object references, multivalued attributes, query un-nesting, object methods, and abstract data types. However, typical object navigation is considered already addressed by object-oriented benchmarks and is not taken into account. Moreover, object-relational database benchmarks have not evolved since the early 2000's, whereas object-relational database systems have.

Similarly, XML benchmarks aim at comparing the various XML storage and querying solutions developed since the late nineties. From the early so-called XML application benchmarks that implement a mixed XML database that is either data-oriented (structured data) or document-oriented (in general, random texts built from a dictionary), XBench (Yao et al., 2004) stands out. XBench is indeed the only benchmark proposing a true mixed dataset (i.e., data *and* document-oriented) and helping evaluate all the functionalities offered by XQuery. FlexBench (Vranec & Mlýnková, 2009) also tests a large set of data characteristics and proposes query templates that allow modeling multiple types of applications. Finally, Schmidt et al. (2009) and Zhang et al. (2011) propose benchmarks that are specifically tailored for testing logical XML model-based systems, namely native XML and XML-relational database management systems, respectively.

Decision-Support Benchmarks

TPC-H (TPC, 2014a) has long been the only standard decision-support benchmark. It exploits a classical product-order-supplier database schema, as well as a workload that is constituted of twenty-two SQL-92, parameterized, decision-support

queries and two refreshing functions that insert tuples into and delete tuples from the database. Query parameters are randomly instantiated following a uniform law. Three primary metrics describe performance in terms of power, throughput, and a combination of power and throughput.

However, TPC-H's database schema is not a star-like multidimensional schema that is typical in data warehouses. Furthermore, its workload does not include any true OLAP query. TPC-DS (TPC, 2015b) now fills in this gap. Its schema represents the decision-support functions of a retailer under the form of a constellation schema with several fact tables and shared dimensions. TPC-DS' workload is constituted of four classes of queries: reporting queries, ad-hoc decision-support queries, interactive OLAP queries, and extraction queries. SQL-99 query templates help randomly generate a set of about five hundred queries, following non-uniform distributions. TPC-DS features one primary throughput metric that takes both query execution and data warehouse maintenance into account.

Given the primordial importance of data integration in many data-centric (including data warehousing) scenarios, TPC-H was recently complemented by TPC-DI (TPC, 2014b). TPC-DI focuses on Extract, Load and Transform (ETL) processes. Data are first generated in a staging area as if they were extracted from a virtual retail brokerage firm's operational databases. Then, data are transformed through, e.g., type conversions, attribute splits or merges, and error checks. Finally, data are loaded into a warehouse constituted of five fact tables and eight dimension tables. There are two load phases: an initial, so-called historical load, and then incremental updates. Transformations are different in these two phases. TPC-DI's main metric is a combination of throughputs from the historical load and two incremental updates.

There are, again, few decision-support benchmarks out of the TPC, but with TPC-DS having had an eight-year long development, alternative data warehouse benchmarks were proposed. Published by the OLAP council, a now inactive

organization founded by OLAP vendors, APB-1 (OLAP Council, 1998) was the first of them and actually predates TPC-DS. APB-1 has been intensively used in the late nineties. However, APB-1 is very simple and rapidly proved limited to evaluate the specificities of various activities and functions. Thus, more elaborate alternatives were proposed, such as DWEB (Darmont et al., 2007), which can be parameterized to generate various ad-hoc synthetic data warehouses and workloads that include typical OLAP queries, and SSB (O'Neil et al., 2009), which is based on TPC-H's database remodeled as a star schema and features a query workload that provides both functional and selectivity coverage.

It is also worth noting that TPC-DS is a complex benchmark. Thence, simpler benchmarks are still in use, especially for testing OLAP scenarios in the cloud. For instance, TPC-H was used to benchmark Hadoop and Pig (Moussa, 2012) and SSB for testing the efficiency of view materialization in the cloud (Perriot et al., 2014). Niche benchmarks also rely a lot on TPC-H. XWeB (Mahboubi & Darmont, 2010) proposes a unified reference model for XML warehouses and its associate XQuery decision-support workload. RTDW-bench (Jedrzejczak et al., 2012) is designed for testing the ability of a real-time data warehouse to handle a transaction stream without delay, given an arrival rate. Eventually, Bär and Golab (2012) propose a benchmark for stream data warehouses that measures the freshness of materialized views.

Finally, a couple of benchmarks are even more specific (and unrelated from TPC-H), e.g., Spadawan (Lopes Siqueira et al., 2010), which allows performance evaluation of specific, complex operations in spatial data warehouses, and BenchDW (Triplet & Butler, 2013), which targets biological data warehouses and particularly focuses on performance metrics, with twenty-two different metrics such as documentation quality, accuracy and response time.

Big Data Benchmarks

In the timely trend of big data analytics, benchmarking needs are as high as ever to compare alternative systems, including the many NoSQL database systems. In this context of data variety, volume and velocity, adaptability and scalability are premium features that must be evaluated. The TPC promotes two benchmarks for big data. TPC-VMS (TPC, 2013) is actually a benchmarking environment for virtualized databases that allows running TPC-C, TPC-E, TPC-H or TPC-DS on three virtual machines (VMs). Its metric is the minimum value of the selected benchmark's metric on all VMs. TPCx-HS (TPC, 2015c) focuses on Hadoop and MapReduce-based applications. It models a simple application (with no data model directly available) and features, in addition to classical metrics, availability and energy metrics.

Before the TPC could issue its big data benchmarks, and still in parallel, there are many industrial and academic initiatives. MalStone (Open Cloud Consortium, 2009) is a benchmark for assessing data intensive parallel processing. It features MalGen, a synthetic data generator that produces large datasets generated probabilistically following specified distributions. In the same line, HiBench (Huang et al., 2010) is a set of Hadoop programs, ranging from data sorting to clustering, aimed at measuring metrics such as response time, HDFS bandwidth consumption and data access patterns. SWIM (Chen et al., 2013) also measures the performance of Hadoop/MapReduce systems. SWIM contains suites of workloads of thousands of jobs, with complex data, arrival, and computation patterns, and therefore provides workload-specific optimizations. Finally, HcBench (Saletore et al., 2013) models real Hadoop usages in a datacenter. HcBench features various job types and data sizes.

By contrast, YCSB (Cooper et al., 2010) is a framework that focuses on data, and more specifically on performance evaluation of key-

value stores. YCSB defines several metrics and workloads to measure system behavior in different situations, or the same system when using different configurations. OLTP-Bench (Curino et al., 2012) is the first true benchmarking framework designed for cloud transactional database systems as a service. OLTP-Bench actually features a set of existing micro-benchmarks (i.e., designed to test one very specific aspect of performance, e.g., ResourceStresser), popular benchmarks (e.g., TPC-C) and real-world applications (e.g., Wikipedia).

Regarding big data analytics, PRIMEBALL (Ferrarons et al., 2013) aims at providing a real-life context to cloud data warehouse benchmarking. Its authors provide the specifications of a fictitious news site hosted in the cloud that is to be managed by the framework under analysis, together with several objective use cases and measures for evaluating system performance. The Big Data Benchmark (Amplab, 2014) goes one step further by actually implementing existing analytical workload models by Pavlo et al. (2009). Its only metric is the response time of such relational queries as scans, aggregations and joins. It can be used for both MapReduce-based systems (such as Shark and Hive) and classical parallel database systems. BigBench (Rabl et al., 2015) is a so-called a specification-based benchmark that is independent from technology. BigBench relies a lot on TPC-DS, borrowing its data model and part of its workload model. The remainder of the workload is adapted from big data use cases issued by the McKinsey Global Institute. Finally, yet other benchmarks, i.e., CloudSuite (Yasin et al., 2014) and DCBench (Jia et al., 2013), feature machine learning and data mining-oriented workload models that mostly run on Hadoop and exploit the Mahout library.

Eventually, BigDataBench (Wang et al., 2014) aims at providing the widest possible scope of big data models and workloads. It includes nineteen benchmarks representing a large variety of data models, workload models and application scenarios from search engines, social networks,

e-commerce, multimedia analytics and bioinformatics. Workload models cover OLTP, “cloud OLTP” and OLAP. As BigBench, BigDataBench also allows alternative implementations, e.g., using MapReduce or Spark.

ISSUES AND TRADEOFFS IN DATA PROCESSING BENCHMARKS

Gray (1993) defines four primary criteria to specify a “good” benchmark.

1. **Relevance:** The benchmark must deal with aspects of performance that appeal to the largest number of potential users.
2. **Portability:** The benchmark must be reusable to test the performances of different DBMSs.
3. **Simplicity:** The benchmark must be feasible and must not require too many resources.
4. **Scalability:** The benchmark must adapt to small or large computer architectures.

In their majority, existing benchmarks aim at comparing the performances of different systems in given experimental conditions. This helps vendors position their products relatively to their competitors’, and users achieve strategic and costly software choices based on objective information. These benchmarks invariably present fixed data and workload models. Gray’s scalability factor is achieved through a reduced number of parameters that mainly allows varying database size in predetermined proportions. All TPC benchmarks notably feature a single scale factor parameter.

This solution is simple (still according to Gray’s criteria), but the relevance of such benchmarks is inevitably reduced to the test cases that are explicitly modeled. For instance, the typical customer-order-product-supplier data model from TPC benchmarks is unsuitable to many application domains. This leads benchmark users to design more or less elaborate variants of standard tools, when they feel these are not generic enough to

fulfill particular needs. Such users are generally not confronted to software choices, but to architectural choices or performance optimization tradeoffs within a given system or family of systems. In this context, it is essential to multiply experiments and test cases, and a monolithic benchmark is of reduced relevance.

To enhance the relevance of benchmarks aimed at system designers, three solutions are possible. The first one is to design an ad-hoc benchmark for a particular application, e.g., RTW-bench, Spadawan and BenchDW, for real-time, spatial and biological data warehouses, respectively. However, the benchmark's application span is necessarily quite narrow. One alternative is to resort to benchmark generators, also called tunable or generic benchmarks, such as OCB, DWEB or FlexBench, which help generate various data or workload models, and thus allow experiments to run in various conditions. The other alternative, which is preferred in recent benchmarks such as YSCB, OLTP-Bench and BigDataBench, is to offer a unifying framework that includes a comprehensive set of state-of-the-art benchmarks. However, the two latter approaches are mechanically detrimental to simplicity, which is a primordial criterion. It is thus necessary to devise benchmark suites that do not sacrifice simplicity too much.

FUTURE RESEARCH DIRECTIONS

The previous section showed that classical transaction and decision-oriented benchmarks are well established. However, big data benchmarking, which predominantly uses cloud technologies, faces a new paradigm and must measure new features. Thus, in addition to Gray's (1993) criteria for building a good benchmark, Folkerts et al. (2012) propose that the quality criteria that are commonly accepted by the benchmarking community must be revisited.

Although the cloud inherits from a long legacy of distributed systems, important issues are unique

to the cloud. For instance, the concept of elasticity applied to data management may translate in the ability to bring in new data sources dynamically to meet emerging needs (Pedersen, 2010). Thus, cloud benchmark data models should be dynamic. Moreover, the three or four Vs (if we include veracity) of big data are unequally addressed in current benchmarks, which mostly focus on scaling (volume) and, to a lesser extent, on variety with multi-benchmark suites such as BigDataBench. Among recent benchmarks, only HcBench features inter-job arrival rates that can simulate data streams. Yet, it is a quite low-level benchmark. Thus, RTDW-Bench and Bär and Golab's (2012) stream warehouse benchmark could be welcome additions in multi-benchmark suites.

With respect to veracity, specific security issues appeared in the new framework of the cloud, e.g., cloud provider or subcontractor espionage, cost-effective defense of availability or uncontrolled mashups (Chow et al., 2009). Such features are important to assess, for security is one of the top concern of cloud users and would-be users. Data consistency is also a concern, e.g., PRIMEBALL's metrics do not only target transaction performance and storage costs, but also data consistency. Bermbach et al. (2013) further advocate for a standard comprehensive benchmark for quantifying the consistency guarantees of eventually consistent storage systems. Moreover, for web application-based benchmarks, data quality assessment is also of premium importance. To the best of our knowledge, this intricate task has not been included yet in any big data benchmark.

Finally, the economic model of the cloud is fundamentally new. Instead of a costly initial investment, pay-as-you-go models allow users to pay a small amount per use, e.g., of a dataset, in return for a one-time advantage (Pedersen, 2010). Thus, cost is also a key criterion when benchmarking cloud/big data solutions. TPC benchmarks typically feature a cost metric, but it is presumably too high-level for fine-grained cost analyses.

CONCLUSION

Benchmarking is a small field, but it is nonetheless essential to data-centric research and industry. It serves both engineering and research purposes, when designing systems or validating solutions; as well as marketing purposes, when monitoring competition and comparing commercial products.

We subdivide benchmarks in three classes. First, standard, general-purpose benchmarks such as the TPC's do an excellent job in evaluating the global performance of systems. They are well suited to software selection by users and marketing battles by vendors, who try to demonstrate the superiority of their product at one moment in time. However, their relevance drops for some particular applications that exploit data or workload models that are radically different from those they implement. Ad-hoc benchmarks are a solution. They either are adaptations of general-purpose benchmarks, or specifically designed benchmarks. Designing myriads of narrow-band benchmarks is not time-efficient, though, and trust in yet another new benchmark might prove limited in the community. Hence, the last alternative is to use generic or multi-benchmarks that feature a common framework for generating various experimental possibilities. The drawback of this approach is that benchmark complexity must be mastered. In conclusion, before starting a benchmarking experiment, users' needs must be carefully assessed so that the right benchmark or benchmark class is selected, and test results are meaningful.

It is nonetheless clear that the TPC plays a primordial role in the data benchmarking community, not only by issuing standards, but also by structuring and leading the community, e.g., by organizing the annual Technology Conference on Performance Evaluation and Benchmarking (TPCTC). This event does not only promote the TPC's activity, but also greatly encourages industrial and academic advances in the field of performance evaluation and benchmarking, whether they are related to the TPC or TPC benchmarks or not.

REFERENCES

- Amplab. (2014). *Big Data Benchmark*. Retrieved from <https://amplab.cs.berkeley.edu/benchmark/>
- Bär, A., & Golab, L. (2012). Towards benchmarking stream data warehouses. In *Proceedings of the 15th ACM International Workshop on Data Warehousing and OLAP (DOLAP 2012)* (pp. 105-112). Maui, HI: ACM.
- Bermbach, D., Zhao, L., & Sakr, S. (2013). Towards Comprehensive Measurement of Consistency Guarantees for Cloud-Hosted Data Storage Services. In *Proceedings of the 5th Technology Conference on Performance Evaluation and Benchmarking (TPCTC 2013)* (Vol. 8391, pp. 32-47). Riva del Garda, Italy: Springer.
- Carey, M. J., DeWitt, D. J., & Naughton, J. F. (1993). The OO7 benchmark. In *Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD 1993)* (pp. 12-21). Washington, DC: ACM. doi:10.1145/170035.170041
- Carey, M. J., Dewitt, D. J., & Naughton, J. F. (1997). The BUCKY Object-Relational Benchmark. In *Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD 1997)* (pp. 135-146). Tucson, AZ: ACM. doi:10.1145/253260.253283
- Chen, Y., Alspaugh, S., Ganapathi, A., Griffith, R., & Katz, R. (2013). *The Statistical Workload Injector for MapReduce (SWIM)*. Retrieved from <https://github.com/SWIMProjectUCB/SWIM/wiki>
- Chow, R., Golle, P., Jakobsson, M., Shi, E., Stadon, J., Masuoka, R., & Molina, J. (2009). Controlling Data in the Cloud: Outsourcing Computation without Outsourcing Control. In *Proceedings of the 1st ACM Cloud Computing Security Workshop* (pp. 85-90). Chicago: ACM.

- Cooper, B. F., Silberstein, A., Tam, E., Ramakrishnan, R., & Sears, S. (2010). Benchmarking cloud serving systems with YCSB. In *Proceedings of the 1st ACM symposium on Cloud Computing (SoCC 2010)* (pp. 143-154). Indianapolis, IN: ACM. doi:10.1145/1807128.1807152
- Curino, C., Difallah, D. E., Pavlo, A., & Cudré-Mauroux, P. (2012). Benchmarking OLTP/ Web Databases in the Cloud: the OLTP-Bench Framework. *Proceedings of the 4th International Workshop on Cloud Data Management (CloudDB 2012)*, 17-20. doi:10.1145/2390021.2390025
- Darmont, J., Bentayeb, F., & Boussaid, O. (2007). Benchmarking Data Warehouses. *International Journal of Business Intelligence and Data Mining*, 2(1), 79–104. doi:10.1504/IJBIDM.2007.012947
- Darmont, J., & Schneider, M. (2000). Benchmarking OODBs with a Generic Tool. *Journal of Database Management*, 11(3), 16–27. doi:10.4018/jdm.2000070102
- Ferrarons, J., Adhana, M., Colmenares, C., Pietrowska, S., Bentayeb, F., & Darmont, J. (2013). PRIMEBALL: a Parallel Processing Framework Benchmark for Big Data Applications in the Cloud. In *Proceedings of the 5th Technology Conference on Performance Evaluation and Benchmarking (TPCTC 2013)* (Vol. 8391, pp. 109-124). Riva del Garda, Italy: Springer.
- Folkerts, E., Alexandrov, A., Sachs, K., Iosup, A., Markl, V., & Tosun, C. (2012). Benchmarking in the Cloud: What It Should, Can, and Cannot Be. *Selected Topics in Performance Evaluation and Benchmarking: 4th TPC Technology Conference, TPCTC 2012, Istanbul, Turkey, August 27, 2012, Revised Selected Papers* (Vol. 7755, pp. 173-188). Istanbul, Turkey: Springer.
- Gray, J. (Ed.). (1993). *The Benchmark Handbook for Database and Transaction Processing Systems* (2nd ed.). San Francisco: Morgan Kaufmann.
- Huang, S., Huang, J., Dai, J., Xie, T., & Huang, B. (2010). The HiBench benchmark suite: Characterization of the MapReduce-based data analysis. *Workshops Proceedings of the 26th International Conference on Data Engineering (ICDE 2010)*, 41-51.
- Jedrzejcak, J., Koszlajda, T., & Wrembel, R. (2012). RTDW-bench: Benchmark for Testing Refreshing Performance of Real-Time Data Warehouse. In *Proceedings of the 23rd International Conference on Database and Expert Systems Applications (DEXA 2012)* (Vol. 7446, pp. 199-206). Vienna, Austria: Springer, LNCS. doi:10.1007/978-3-642-32597-7_18
- Jia, Z., Zhan, J., Wang, L., & Zhang, L. (2013). *DCBench: A benchmark suite for data center*. Presented at the 19th IEEE International Symposium on High Performance Computer Architecture (HPCA 2013) (Tutorial), Shenzhen, China.
- Lee, S., Kim, S., & Kim, W. (2000). The BORD Benchmark for Object-Relational Databases. In *Proceedings of the 11th International Conference on Database and Expert Systems Applications (DEXA 2000)* (Vol. 1873, pp. 6-20). London, UK: Springer, LNCS.
- Lopes Siqueira, T. L., & Rodrigues Ciferri, R. (2010). Benchmarking Spatial Data Warehouses. In *Proceedings of the 12th International Conference on Data Warehousing and Knowledge Discovery* (Vol. 6263, pp. 40-51). Bilbao, Spain: Springer, LNCS. doi:10.1007/978-3-642-15105-7_4
- Mahboubi, H., & Darmont, J. (2010). XWeB: The XML Warehouse Benchmark. In *Proceedings of the 2nd Technology Conference on Performance Evaluation and Benchmarking (TPCTC 2010)* (Vol. 6417, pp. 185-203). Singapore: Springer, LNCS.

- Moussa, R. (2012). TPC-H Benchmark Analytics Scenarios and Performances on Hadoop Data Clouds. In R. Benlamri (Ed.), *Networked Digital Technologies. Communications in Computer and Information Science* (Vol. 293, pp. 220–234). Springer. doi:10.1007/978-3-642-30507-8_20
- O’Neil, P., O’Neil, E., Chen, X., & Revilak, S. (2009). The Star Schema Benchmark and Augmented Fact Table Indexing. In *Proceedings of the 1st Technology Conference on Performance Evaluation and Benchmarking (TPCTC 2009)* (Vol. 5895, pp. 237). Lyon, France: Springer. doi:10.1007/978-3-642-10424-4_17
- OLAP Council. (1998). *APB-1 OLAP Benchmark Release II*. Retrieved from <http://www.olapcouncil.org/research/bmarkly.htm>
- Open Cloud Consortium. (2009). *Generate synthetic site-entity log data for testing and benchmarking applications requiring large data sets*. Retrieved from <http://code.google.com/p/malgen/>
- Pavlo, A., Paulson, E., Rasin, A., Abadi, D. J., DeWitt, D. J., Madden, S., & Stonebraker, M. (2009). A comparison of approaches to large-scale data analysis. *Proceedings of the of the ACM SIGMOD International Conference on Management of Data (SIGMOD 2009)*, 165–178. doi:10.1145/1559845.1559865
- Pedersen, T. B. (2010). Research challenges for cloud intelligence. In *Proceedings of the 2010 EDBT/ICDT Workshops*. Lausanne, Switzerland: ACM International Conference Proceeding Series.
- Perriot, R., Pfeifer, J., D’Orazio, L., Bachelet, B., Bimonte, S., & Darmont, J. (n.d.). Cost Models for Selecting Materialized Views in Public Clouds. *International Journal of Data Warehousing and Mining*, 10(4), 1–25.
- Rabl, T., Frank, M., Danisch, M., Jacobsen, H. A., & Gowda, B. (2015). The Vision of BigBench 2.0. *Proceedings of the 4th Workshop on Data analytics in the Cloud (DanaC 2015)*, 3:1–3:4.
- Saletore, V. A., Krishnan, K., Viswanathan, V., & Tolentino, M. E. (2013). HcBench: Methodology, Development, and Full-System Characterization of a Customer Usage Representative Big Data/Hadoop Benchmark. In *Proceedings of the 3rd and 4th Workshops on Big Data Benchmarking (WBDB 2013) – Revised selected papers* (Vol. 8585, pp. 73–93). San Jose, CA: Springer. doi:10.1109/IISWC.2013.6704672
- Schmidt, K., Bächle, S., & Härder, T. (2009). Benchmarking Performance-Critical Components in a Native XML Database System. In *Proceedings of the 1st International Workshop on Benchmarking of XML and Semantic Web Applications (BenchmarX 2009)* (Vol. 5667, pp. 64–78). Brisbane, Australia: Springer. doi:10.1007/978-3-642-04205-8_7
- TPC. (2010). *TPC Benchmark C Standard Specification Revision 5.11*. Transaction Processing Performance Council. Retrieved from <http://www.tpc.org>
- TPC. (2013). *TPC Virtual Measurement Virtual System Standard Specification Version 1.2.0*. Transaction Processing Performance Council. Retrieved from <http://www.tpc.org>
- TPC. (2014a). *TPC Benchmark H Standard Specification Revision 2.17.1*. Transaction Processing Performance Council. Retrieved from <http://www.tpc.org>
- TPC. (2014b). *TPC Benchmark DI Standard Specification Version 1.1.0*. Transaction Processing Performance Council. Retrieved from <http://www.tpc.org>
- TPC. (2015a). *TPC Benchmark E Standard Specification Version 1.14.0*. Transaction Processing Performance Council. Retrieved from <http://www.tpc.org>
- TPC. (2015b). *TPC Benchmark DS Standard Specification Version 1.4.0*. Transaction Processing Performance Council. Retrieved from <http://www.tpc.org>

TPC. (2015c). *TPC Express Benchmark HS Standard Specification Version 1.3.0*. Transaction Processing Performance Council. Retrieved from <http://www.tpc.org>

Triplet, T., & Butler, G. (2013). BenchDW: a generic framework for biological data warehouse benchmarking. In *Proceedings of the 28th ACM Symposium on Applied Computing (SAC 2013)* (pp. 1328-1334). Coimbra, Portugal: ACM. doi:10.1145/2480362.2480612

Vranec, M., & Mlýnková, I. (2009). FlexBench: A Flexible XML Query Benchmark. In *Proceedings of the 14th International Conference on Database Systems for Advanced Applications (DASFAA 2009)* (Vol. 5463, pp. 421-435). Brisbane, Australia: Springer.

Wang, L., Zhan, J., Luo, C., Zhu, Y., Yang, Q., He, Y., & Qiu, B. et al. (2014). BigDataBench: A big data benchmark suite from internet services. *Proceedings of the IEEE 20th International Symposium on High Performance Computer Architecture (HPCA 2014)*, 488-499. doi:10.1109/HPCA.2014.6835958

Yao, B. B., Özsu, T., & Khandelwal, N. (2004). XBench Benchmark and Performance Testing of XML DBMSs. *Proceedings of the 20th International Conference on Data Engineering (ICDE 2004)*, 621-633. doi:10.1109/ICDE.2004.1320032

Yasin, A., Ben-Asher, Y., & Mendelson, A. (2014). Deep-dive analysis of the data analytics workload in CloudSuite. *Proceedings of the 2014 IEEE International Symposium on Workload Characterization (IISWC 2014)*, 202-211. doi:10.1109/IISWC.2014.6983059

Zhang, X., Liu, K., Zou, L., Du, X., & Wang, S. (2011). Renda-RX: A Benchmark for Evaluating XML-Relational Database System. In *Proceedings of the 12th International Conference on Web-Age Information Management (WAIM 2011)* (Vol. 6897, pp. 578-589). Wuhan, China: Springer. doi:10.1007/978-3-642-23535-1_49

ADDITIONAL READING

Alexandrov, A., Brücke, C., & Markl, V. (2013). Issues in big data testing and benchmarking. Presented at the 6th *International Workshop on Testing Database Systems (DBTest 2013)*. New York, USA. doi:10.1145/2479440.2482677

Armstrong, T. G., Ponnekanti, V., Borthakur, D., & Callaghan, M. (2013). LinkBench: a database benchmark based on the Facebook social graph. In *Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD 2013)* (pp. 1185-1196). New York, USA. doi:10.1145/2463676.2465296

Barata, M., Bernardino, J., & Furtado, P. (2014). Survey on Big Data and Decision Support Benchmarks. In *Proceedings of the 25th International Conference on Database and Expert Systems Applications (DEXA 2014)* (Vol. 8645, p. 174-182). Munich, Germany: Springer, LNCS. doi:10.1007/978-3-319-10085-2_15

Baru, C. K., Bhandarkar, M. A., Othayoth Nambiar, R., Poess, M., & Rabl, T. (2012). Setting the Direction for Big Data Benchmark Standards. In *Proceedings of the 4th Technology Conference on Performance Evaluation and Benchmarking (TPCTC 2012)* (Vol. 7755, pp. 197-208). Istanbul, Turkey: Springer, LNCS.

Han, R., Jia, Z., Gao, W., Tian, X., & Wang, L. (2015). Benchmarking Big Data Systems: State-of-the-Art and Future Directions. Retrieved from <http://arxiv.org/abs/1506.01494>

Huppler, K. (2009). The Art of Building a Good Benchmark. In *Proceedings of the 1st Technology Conference on Performance Evaluation and Benchmarking (TPCTC 2009)* (Vol. 5895, pp. 18-30). Lyon, France: Springer, LNCS. doi:10.1007/978-3-642-10424-4_3

- Iosup, A. (2013). IaaS cloud benchmarking: approaches, challenges, and experience. In *Proceedings of the 2013 international workshop on Hot topics in cloud services (HotTop-iCS 2013)* (pp. 1-2). Prague, Czech Republic. doi:10.1145/2462307.2462309
- Jia, Z., Zhou, R., Zhu, C., Wang, L., Gao, W., Shi, Y.,... Zhang, L. (2013). The Implications of Diverse Applications and Scalable Data Sets in Benchmarking Big Data Systems. Retrieved from <http://arxiv.org/abs/1307.7943>
- Kestelyn, J. (2014). Big Data Benchmarks: Toward Real-Life Use Cases. Retrieved from <http://blog.cloudera.com/blog/2014/08/big-data-benchmarks-toward-real-life-use-cases/>
- Kuhlenkamp, J., Klems, M., & Röss, O. (2014). Benchmarking Scalability and Elasticity of Distributed Database Systems. In *Proceedings of the VLDB Endowment (VLDB 2014)* (Vol. 7, No. 12, pp. 1219-1230). doi:10.14778/2732977.2732995
- Nambiar, R., & Poess, M. (Eds.). (2015). Performance Characterization and Benchmarking. Traditional to Big Data (Vol. 8904): Springer, LNCS. doi:10.1007/978-3-319-15350-6
- Nambiar, R., Poess, M., Masland, A., Taheri, H. R., Emmerton, M., Carman, F., & Majdalany, M. (2013). TPC Benchmark Roadmap 2012. *Selected Topics in Performance Evaluation and Benchmarking*, 7755, 1–20. doi:10.1007/978-3-642-36727-4_1
- Qin, X., & Zhou, X. (2013). A Survey on Benchmarks for Big Data and Some More Considerations. In *Proceedings of the 14th International Conference on Intelligent Data Engineering and Automated Learning (IDEAL 2013)* (Vol. 8206, pp. 619-627). Hefei, China: Springer, LNCS. doi:10.1007/978-3-642-41278-3_75
- Rabl, T., Jacobsen, H. A., Nambiar, R., Poess, M., Bhandarkar, M. A., & Baru, C. K. (Eds.). (2013). *Proceedings of the 3rd and 4th Workshops on Big Data Benchmarking (WBDB 2013) – Revised selected papers* (Vol. 8585). Xi'an, China – San Jose, USA: Springer, LNCS.
- T. Rabl, M. Poess, C. K. Baru, & H. A. Jacobsen (Eds.). (2012). *Proceedings of the 1st and 2nd Workshops on Big Data Benchmarking (WBDB 2012) – Revised selected papers* (Vol. 8163). San Jose, USA – Pune, India: Springer, LNCS.
- T. Rabl, K. Sachs, M. Poess, C. K. Baru, & H. A. Jacobsen (Eds.). (2014). *Proceedings of the 5th Workshop on Big Data Benchmarking (WBDB 2014) – Revised selected papers* (Vol. 8991). Postdam, Germany: Springer, LNCS.
- Rivero, C. R., Schultz, A., Bizer, C., & Ruiz, D. (2012). Benchmarking the Performance of Linked Data Translation Systems. Presented at the *Workshop on Linked Data on the Web (LDOW 2012)*. Lyon, France.
- Surdu, S., Gripay, Y., Scuturici, V. M., & Petit, J. M. (2013). P-Bench: Benchmarking in Data-Centric Pervasive Application Development. In *Transactions on Large-Scale Data- and Knowledge-Centered Systems XI* (vol. 8290, pp. 51-75). Springer, LNCS.
- Waage, T., & Wiese, L. (2015). Benchmarking Encrypted Data Storage in HBase and Cassandra with YCSB. In *Foundations and Practice of Security* (Vol. 8930, pp. 311-325). Springer, LNCS. doi:10.1007/978-3-319-17040-4_20
- Wang, H., Li, J., Zhang, Z., & Zhou, Y. (2014). Benchmarking Replication and Consistency Strategies in Cloud Serving Databases: HBase and Cassandra. In *Big Data Benchmarks, Performance Optimization, and Emerging Hardware* (Vol. 8807, pp. 71-82). Springer, LNCS.

KEY TERMS AND DEFINITIONS

Benchmark: A standard program that runs on different systems to provide an accurate measure of their performance.

Cloud Benchmarking: Use of cloud services in the respective (distributed) systems under test (Folkerts et al., 2012).

Database Benchmark: A benchmark specifically aimed at evaluating the performance of Database Management Systems (DBMSs) or DBMS components.

Data Model: In a data-centric benchmark, a database schema and a protocol for instantiating

this schema, *i.e.*, generating synthetic data or reusing real-life data.

Performance Metrics: Simple or composite metrics aimed at expressing the performance of a system.

Synthetic Benchmark: A benchmark in which the workload model is artificially generated, as opposed to a real-life workload.

Workload Model: In a data-centric benchmark, a set of predefined read and write operations or operation templates to apply on the benchmark's database, following a predefined protocol.

Data Linkage Discovery Applications

D

Richard S. Segall

Arkansas State University, USA

Shen Lu

University of South Florida, USA

INTRODUCTION

This chapter discusses the topic of linkage discovery for data and their applications. Since the publication of Segall and Lu (2014) which pertained to linkage discovery with glossaries many other investigators have performed research on applications of linkage discovery for big data under new conditions and criteria. This chapter enhances Segall & Lu (2014) by not only presenting a more coherent tabular summary of work of others presented in Segall & Lu (2014) and also again presented within this chapter, but also by including additional references that pertain to applications of linkage discovery not requiring a glossary framework.

BACKGROUND

This section discusses some of the terminology of text mining that is also used in linkage discovery that is web-based form of knowledge discovery. Latent Semantic Analysis (LSA) can be used to discover knowledge from text with a general mathematical learning method without knowing prior linguistic or perceptual similarity knowledge.

Latent Semantic Analysis (LSA) is a Natural Language Processing (NLP) technique that is based on similarity of words but not grammatical or syntactical structure and extracts knowledge through the similarity of individual words. The motivation of LSA in terms of psychology is that people learn knowledge only from similarity of

individual words taken as units, not with knowledge of their syntactical or grammatical function.

Experimental result for linkage discovery for glossaries was shown in Lu et al. (2011) and Lu et al. (2012) and by other investigators that, by combining glossaries with the text, we can extract more meaningful words from the text and then link similar sections together.

Latent Semantic Analysis (LSA) can provide the meanings of the terms based on the context. However, one article cannot include all of the domain knowledge and the definition extracted from the context where the term appears in that article is not accurate. But, in glossaries, all of the terms are defined clearly. In Lu et al. (2011) and Lu et al. (2012), we manually put the definitions of the terms in glossaries to those words in an article and use those definitions to improve the accuracy of the background knowledge we can extract from the context. In this way, we can define meaningful words and use them to decide the theme of the corresponding sections.

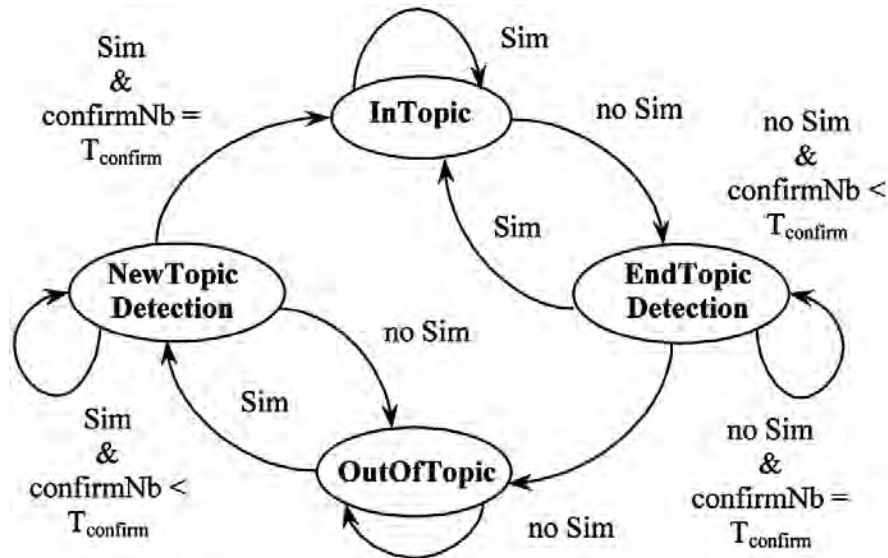
Ferret (2002) presented a method, called TOPICOLL, for using collocations for topic segmentation and link detection. Figure 1 below illustrates the automation of the algorithm of Ferret (2002) for detecting topic shifts.

TOPICOLL Algorithm [Source: Ferret (2002)]

Parameters:

1. State: {NewTopicDetection, InTopic, EndTopicDetection, OutOfTopic}
2. Sim: {True, False} // If

Figure 1. Automation for topic shift detection
[Ferret (2002)]



the context of the focus window and the context of the current segment is similar, Sim is True; otherwise, Sim is False.

```

3.      ConfirmNb: integer // the
        number of successive positions.
4.      Tconfirm: constant // the
        threshold of successive positions.
Input: Document
Begin:
Foreach FocusWindow is Document
{
1.      if(State == NewTopicDetection && Sim == False && ConfirmNb <
ConfirmNb){
2.          State = OutOfTopic;
3.          ConfirmNb = 0;
}
4.      if (state == NewTopicDetection && Sim = True && ConfirmNb <
Tconfirm){
5.          State = NewTopicDetection;
6.          ConfirmNb++;
}

```

```

7.          if(State == NewTopicDetection && Sim == True && ConfirmNb
≥Tconfirm){
8.              State = InTopic;
9.              ConfirmNb = 0;
}
10.         if(State == InTopic && Sim == True){
11.             State = InTopic;
}
12.         if(State == InTopic && Sim == False){
13.             State = EndTopicDetection;
14.             ConfirmNb = 0;
}
15.         if(State == EndTopicDetection && Sim == False && ConfirmNb
< Tconfirm){
16.             State = EndTopicDetection;
17.             ConfirmNb++;
}
18.         if(State == EndTopicDetection && Sim == True && ConfirmNb <
Tconfirm){

```

```

19.          State = InTopic;
20.          ConfirmNb = 0;
21.          if(State == EndTopicDe-
22.          tection && Sim == False && ConfirmNb
23.          ≥ Tconfirm){
24.              State = OutOfTop-
25.              ic;
26.              ConfirmNb = 0;
27.          }
28.          if(State == OutOfTopic &&
29.          Sim == False){
30.              State = OutOfTop-
31.              ic;
32.          }
33.          if(State == OutOfTopic &&
34.          Sim == True){
35.              State = NewTopic-
36.              Detection;
37.              ConfirmNb = 0;
38.          }
39.          }
40.      End;

```

As indicated in Hearst (1995), the field of information retrieval has traditionally focused on textbases consisting of titles and abstracts. Hearst (1995) developed an algorithm called TextTiling that partitions expository texts into multi-paragraph segments that reflect their subtopic structure. The name TextTiling was used because the segments are adjacent and non-overlapping.

MAIN FOCUS OF THE CHAPTER

The main focus of this chapter is to summarize research contributions by other investigators in areas related to data linkage discovery. As this is a research area that was initiated in the 1990's it is necessary to discuss some of these and later milestone contributions that lead to the current state of research in linkage discovery.

The below Table 1 provides a new tabular summary of some of the work that was both discussed in Segall & Lu (2014) and is also contained in this chapter but reorganized by data linkage

Table 1. Categorization of data linkage techniques

Data Linkage Technique	Data Linkage Technique	Representative Articles	Other Properties or Characteristics
Algorithms	① Latent Semantic Analysis (LSA) ② TOPICOLL ③ TextTiling ④ Maximum Entropy (ME) ⑤ Latent Dirichlet Analysis (LDA) ⑥ Field Association ⑦ Latent Semantic Indexing (LSI)	① Foltz (1996) ② (Ferret(2002) ③ Hearst (1997, 1995, 1993) ④ Chua et al. (2004) ⑤ Liu et al (2009) ⑥ Sharif et al. (2007) ⑦ Zekikovcitz and Hirsh (2002)	① text-based research ② using collocations ③ segmentation by multi-paragraph subtopic passages ④ for large broadcast news video archives ⑤ interactive topic based ⑥ dictionary applications ⑦ nearest neighbor text classification
Text decomposition/segmentation	① Passage retrieval ② Text segmentation ③ Field association terms	① Salton et al (1993) ② Ponte & Croft (1997) ③ Atlam et al. (2003)	① full text information systems ② by topic ③ documents similarity measurements
Semantic similarity estimation	① ontology-based ② emergent semantics ③ word sense disambiguation ④ corpus-based	① Sanchez and Batet (2011) ② Markines et al. (2009) ③ Patwardhan et al (2003) ④ Mihalcea and Corley (2006)	① biomedical domain ② social tagging ③ semantic relatedness measures ④ knowledge-based measures
Frameworks	① Digital libraries ② Silk link discovery framework ③ semantic link discovery	① Fox and Sornil (1999) ② Volz et al. (2009) ③ Hassanzadeh et al. (2009)	① information retrieval ② for web of data ③ relational data

techniques. Table 1 categorizes the data linkage techniques into techniques of algorithms, text decomposition/segmentation, semantic similarity estimation, and frameworks.

Algorithms

According to Foltz (1996), Latent Semantic Analysis (LSA) is a statistical model of word usage that permits comparisons of semantic similarity of textual information. Foltz (1996) summarized how three experiments that illustrates how Latent Semantic Analysis (LSA) can be used in text-based research. Perez et al. (2005) performed a study for automatic assessment of students' free-text answers using a combination of Latent Semantic Analysis with an algorithm previously used by others for evaluating machine translation systems.

Chua et al. (2004) discussed techniques, experience and trends in story boundary detection in large broadcast news video archives that entails segmentation of news videos. The techniques employed by Chua et al. (2004) include sophisticated machine learning models based on Maximum Entropy (ME) approaches, and concluded that it is advantageous to employ rigorous machine learning techniques with full multi-modality features to achieve good segmentation performance.

Liu et al. (2009) built an interactive, visual text analysis tool that aids users in analyzing a large collection of text. Liu et al. (2009) used a Latent Dirichlet Analysis (LDA) model that effectively conveys complex text summarization results. The work of Liu et al. (2009) also describes a set of rich interaction tools that allows users to work with a created visual text summary to further interpret the summarization results in context and examine the text collection from multiple contributions.

Reynar (1994) presented an outline for a new automatic method of finding topic boundaries and performed experiments on a series of concatenated documents using an optimization algorithm. The experiments performed by Reynar (1994) were on the same data set that consisted of 150 concatenations of articles containing a total of 660 articles

averaging 24.5 sentences in length. The difference between the two experiments in Reynar (1994) was that in the first experiment, boundaries were placed only at the ends of the sentences while in the second experiment; they were only placed at paragraph boundaries. Reynar (1994) showed that high recall rates can be obtained in the first experiment and high precision can be obtained in the second.

Sharif et al. (2007) discussed how using field associations (FA) that are a discriminating set of discriminating terms that can specify document fields can yield improvement of building field association term dictionary. Sharif et al. (2007) used field association with a dictionary composed of 38,372 articles and yielded about 24% more relevant more relevant FA terms that are appended to the earlier FA term dictionary and about 32% irrelevant FA terms are deleted.

Hearst and Plaunt (1993) discussed the merits of subtopic structuring for full-length document access by describing a new way of specifying queries on full-length documents. Hearst and Plaunt (1993) described an experiment in which making use of the recognition of how local structure achieves better results on a typical information retrieval task than does a standard information retrieval measure. Hearst (1997) implemented an algorithm called TextTiling that segmented text into multi-paragraph subtopic passages.

Wang & Si (2008) proposed two probabilistic models to formally model the evidence of a set of top ranked passages in a document, and performed an extensive set of experiments to show the effectiveness of each model and better results in comparison with a state-of-the-art document retrieval algorithm and a language model approach for passage-based retrieval.

Hersh et al. (1995) indicated that all of the methods currently used to evaluate information retrieval systems have limitations in their ability to measure how well users are able to acquire information. Hersch et al. (1995) utilized a new approach for measures of information retrieval evaluation by accessing information obtained based on the user's

ability to answer questions from a short-answer test. Experiments were performed by Hersch et al. (1995) with senior medical students using a ten-question test and then have them search one of two information retrieval systems on the five questions for which they were least certain of their answer. Hearst (1995) presented a visualization paradigm called TileBars that demonstrates the usefulness of explicit term distribution information in Boolean-type queries.

Zelikovcitz & Hirsh (2002) integrated background knowledge into nearest-neighbor text classification using Latent Semantic Indexing accessing document similarities in training and test documents.

Text Decomposition/Segmentation

Salton et al. (1996) performed automatic text decomposition using text segments and text themes. Salton et al. (1996) presented two main decomposition strategies that include a chronological decomposition into “text segments” and semantic decomposition into “text themes”. Choi et al. (2009) discussed lower and upper bounds for linkage discovery using graphical methods.

Ponte and Croft (1997) investigated text segmentation by topic and presented experimental results for an effective method that makes use of a query expansion technique to find common features for the topic segments. According to Ponte and Croft (1997), the applications of the method discussed include topic tracking of broadcast speech data and topic identification in full-text databases.

Atlam et al. (2003) outlined a new text manipulation system named FA-Sim that is useful for retrieving information in large heterogeneous texts and for recognizing content similarity in text excerpts. FA-Sim measures texts similarity by using specific field association (FA) terms instead of comparing all text information. Experiments performed by Atlam et al. (2003) showed precision and recall improved by 39% and 37% respectively over these two traditional methods.

Salton et al. (1993) discussed new approaches for implementing selection passage retrieval systems, and identifying text passages responsive to particular user needs, and used an automated encyclopedia search system to evaluate the usefulness of the proposed methods.

Semantic Similarity Estimation

Pivovarov and Elhadad (2012) presented a hybrid knowledge-based and data-driven approach to identifying semantically similar concepts in biomedical texts. Pivovarov & Elhadad (2012) proposed a comprehensive method that computes a similarity score for a concept pair by combining data-driven and ontology-driven knowledge. Pivovarov & Elhadad (2012) demonstrated their method on a corpus of clinical notes of patients with chronic kidney disease, and their method was able to prune out concepts that are similarly related from those that are semantically similar with accuracy of 92%.

Sanchez and Batet (2011) discussed semantic similarity estimation in the biomedical domain using an ontology-based information-theoretic perspective. According to Sanchez & Batet (2011) semantic similarity estimation is an important component of analyzing natural language resources like clinical records. Sanchez & Batet (2011) found that an information-theoretic redefinition of well-known semantic measures and similarity coefficients, and an intrinsic estimation of information content (IC) result with accuracy.

Pedersen et al. (2007) discussed how six-existing domain-independent measures originally based on WordNet, an English lexical database of concepts and relations can be adopted for semantic similarity and relatedness in the biomedical domain. Pedersen et al. (2007) concluded that there is a role both for more flexible measures of relatedness based on information derived from the body of documents, as well as for measures that rely on existing ontological structures.

Corley and Mihalcea (2005) introduced a method that combines word-to-word similarity

metrics into a text-to-text metric, and showed that this method outperforms the traditional text similarity metrics based on lexical matching. Sukkarieh et al. (2003) used auto-marking as a means of using computational linguistics to score short free text responses.

Markines et al. (2009) evaluated similarity measures for emergent semantics of social tagging. Experiments with resource and tag similarity performed by Markines et al. (2009) have pointed to folksonomy-based mutual information measures as the best at extracting semantic associations from social annotation data. Patwardhan et al. (2003) used measures of semantic relatedness for word sense disambiguation and conducting experiments using a lexical sample data and found that a semantic distance measure and floss overlap resulted in the highest accuracy. Mihalcea & Corley (2006) presented a method for measuring the semantic similarity of texts, using corpus-based and knowledge-based measures of similarity, and showed that the semantic similarity method outperformed methods based on simple lexical matching, resulting in up to 13% error rate reduction with respect to the traditional vector-based similarity metric.

Resnik (1999) presented a measure of semantic similarity in an “is a” taxonomy based on the notion of shared information content. Experimental evaluation by Resnik (1999) against a benchmark set of human similarity judgments demonstrates that the measure performs better than the traditional edge counting approach for application of ambiguity in natural language.

Frameworks

Hassanzadeh et al. (2009) presented a framework for online discovery of semantic links from relational data that is based on declarative specification of linkage requirements by the user that allows matching data items in many real-world scenarios. Hassanzadeh et al. (2009) claim that their framework for semantic link discovery over relational data lets data publishers to easily find

and publish high quality links to other data sources, and therefore could significantly enhance the value of the data in the next generation of web.

Kobayashi and Takeda (2001) discussed the development of new techniques targeted to resolve some of the problems associated with Web-based information retrieval, and speculate on future trends. Fox & Sornil (1999) reviewed and organized into a simple framework some of the developmental efforts related to digital libraries that have been in the information retrieval area, and highlighted needs of the future. Volz et al. (2009) created a link discovery framework called Silk for the Web of Data that accesses data sources over a protocol without the need to replicate datasets locally.

Zamir and Etzioni (1999) introduced Grouper a dynamic clustering interface to Web search engine results. Grouper developed and tested by Zamir & Etzioni (1999) dynamically groups the search results into clusters labeled by phrases extracted from document snippets. Milne et al. (2007) created a new search engine called Koru that offers effective domain independent knowledge-based information retrieval that is powered by Wikipedia, and conducted a detailed user study and found that Koru and its underlying knowledge base offers significant advantages over traditional keyword search. Marchionini (2006) explained that research tools for exploratory search success involve the creation of new interfaces that move the process beyond predictable fact retrieval.

Keim (2002) proposed and exemplified using examples of a classification of information visualization and visual data mining techniques that is based on the data type to be visualized, the visualization technique and the interaction of distortion technique.

FUTURE RESEARCH DIRECTIONS

In future, we will work on summarizing the additional references provided below in the Additional Readings section that were found upon the

writing of this chapter. Each of these Additional Readings will also need to be categorized with the possibility of additional categories.

CONCLUSION

The chapter shows that considerable work has been done by investigators on the topic of data linkage discovery for many applications using techniques that include algorithms, text decomposition/segmentation, semantic similarity estimation, and frameworks.

Linkage discovery has found many new methods for efficient and effective connecting of documents using extensions of data and text mining and is a field that is continuing to be investigated by researchers around the world.

REFERENCES

- Atlam, E.-S., Fuketa, M., Morita, K., & Aoe, J.-I. (2003). Documents similarity measurements using field association terms. *Information Processing & Management*, 39(6), 809–824. doi:10.1016/S0306-4573(03)00019-0
- Choi, S.-S., Jung, K. M., & Moon, B-R (2009). Lower and upper bounds for linkage discovery. *IEEE Transaction on Evolutionary Computation*, 13(2), 201-216.
- Chua, T.-S., Chang, S.-F., Chaisorn, L., & Hsu, W. (2004). Story boundary detection in large broadcast news video archives – Techniques, experience and trends. *Proceedings of the 12th Annual ACM International Conference on Multimedia (MM'04)*, 656-659. doi:10.1145/1027527.1027679
- Corley, C., & Mihalcea, R. (2005). Measuring the semantic similarity of texts. *Proceedings of the Association for Computational Linguistics (ACL) Workshop on Empirical Modeling of Semantic Equivalence and Entailment*, 13-18. doi:10.3115/1631862.1631865
- Ferret, O. (2002). Using collocations for topic segmentation and link detection, *Proceedings of 19th International Conference of Computational Linguistics (COLING 2002)*, 1, 1-7. doi:10.3115/1072228.1072261
- Foltz, P. W. (1996). Latent Semantic Analysis for text-based research. *Behavior Research Methods, Instruments, & Computers*, 28(2), 197–202. doi:10.3758/BF03204765
- Fox, E. A., & Sornil, O. (1999). Digital Libraries. In *Modern Information Retrieval*. Addison-Wesley Publishers.
- Hassanzadeh, O., Lim, L., Kementsietsidis, A., & Wang, M. (2009). A declarative framework for semantic link discovery over relational data. *WWW2009: Proceedings of the 18th International World Wide Web Conference*, 1101-1102. Retrieved from <http://www2009.org/proceedings/pdf/p1101.pdf>
- Hearst, M.A. (1993). *TextTiling: A quantitative approach to discourse segmentation*. Technical Report, Computer Science Division, University of California-Berkeley, Berkeley, CA.
- Hearst, M. A. (1995). TileBars: Visualization of term distribution information in full text information access. *CHI'95 Proceedings of the Special Interest Group on Human Computer Interface (SIGCHI) Conference on Human Factors in Computing Systems*, 59-66. doi:10.1145/223904.223912
- Hearst, M. A. (1997). TextTiling: Segmenting text into multi-paragraph subtopic passages. *Journal of Computational Linguistics*, 23(1), 33-64.
- Hearst, M. A., & Plaunt, C. (1993). Subtopic structuring for full-length document access. *SIGIR '93: Proceedings of the 16th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 59-68. doi:10.1145/160688.160695

- Hersh, W. R., Molnar, A., Elliot, D. L., Hickman, D. H., & Wolf, S. L. (1995). Towards new measures of information retrieval evaluation, *SIGIR'95: Proceedings of the ACM Special Interest Group on Information Retrieval*, 164-170. doi:10.1145/215206.215355
- Keim, D. A. (2002). Information visualization and visual data mining. *IEEE Transactions on Visualization and Computer Graphics*, 7(1), 100-107.
- Kobayashi, M., & Takeda, K. (2000). Information retrieval on the web. *ACM Computing Surveys*, 32(2), 144-173. doi:10.1145/358923.358934
- Liu, S., Zhou, M. X., Pan, S., Qian, W., Cai, W., & Lian, X. (2009). Interactive topic-based visual text summarization and analysis. *CIKM'09 Proceedings of the Conference on Information and Knowledge Management*, 543-552.
- Lu, S., Segall, R. S., & Belford, R. E. (2011). Linkage discovery with symposium proceedings. Proceedings of the 6th BioNanoTox and Applications International Research Conference. Available at http://softchallenge.net/documents/LU_SEGALL_BELFORD_bionanotox2011.pdf
- Lu, S., Segall, R. S., & Belford, R. E. (2012). Linkage discovery by combining latent semantic analysis with glossaries from bioscience domain. Poster presentation at the Arkansas Center for Plant Powered Production (P3) Annual Symposium, Jonesboro, AR.
- Marchionini, G. (2006). Exploratory Search: From finding to understanding. *Communication of the ACM*, 49(4), 41-46.
- Markines, B., Cattuto, C., Menczer, F., Benz, D., Hotho, A., & Stumme, G. (2009). Evaluating similarity measures for emergent semantics of social tagging. *WWW2009: Proceedings of the 18th International World Wide Web Conference*, 641-650. Retrieved from <http://www2009.org/proceedings/pdf/p641.pdf>
- Mihalcea, R., Corley, C., & Strapparava, C. (2006). Corpus-based and Knowledge-based measures of text semantic similarity. *Proceedings of the American Association for Artificial Intelligence (AAAI 2006)*, 775-780.
- Milne, D., Witten, I. H., & Nichols, D. M. (2007). A knowledge-based search engine powered by Wikipedia. *Proceedings of the ACM Conference on Information and Knowledge Management (CIKM'2007)*, 445-454. doi:10.1145/1321440.1321504
- Patwardham, S., Banerjee, S., & Pedersen, T. (2003). Using measures of semantic relatedness for word sense disambiguation. *Proceedings of the Fourth International Conference on Intelligent Text Processing and Computational Linguistics*, 241-257.
- Pedersen, T., Palhomov, S. V. S., Patwardham, S., & Chute, C. G. (2007). Measures of semantic similarity and relatedness in the biomedical domain. *Journal of Biomedical Informatics*, 40(3), 288-299. doi:10.1016/j.jbi.2006.06.004 PMID:16875881
- Perez, D., Gliozzo, A., Strapparava, C., Alfonsecc, E., Rodriguez, P., & Magnini, B. (2005). Automatic assessment of students free-text answer underpinned by the combination of a BLEU-inspired algorithm and latent semantic analysis. *Proceedings of the Eighteenth International Florida Artificial Intelligence Research Society Conference, FLAIRS 2005 - Recent Advances in Artificial Intelligence*, 358-363.
- Pivovarov, R., & Elhadad, N. (2012). A hybrid knowledge-based and data-driven approach to identifying semantically similar concepts. *Journal of Biomedical Informatics*, 45(3), 471-481. doi:10.1016/j.jbi.2012.01.002 PMID:22289420
- Ponte, J. M., & Croft, W. B. (1997). Text segmentation by topic. *Proceedings of the First European Conference on Research and Advanced Technology for Digital Libraries*, 120-129.

- Resnik, P. (1999). Semantic similarity in a taxonomy: An information-based measure and its application to problems of ambiguity in natural language. *Journal of Artificial Intelligence Research*, 11, 95–130.
- Reynar, J. C. (1994). An automated method of finding topic boundaries. *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics*, 331–333.
- Salton, G., Allan, J., & Buckley, C. (1993). Approaches to passage retrieval in full text information systems. *SIGIR '93: Proceedings of the 16th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 49–58. doi:10.1145/160688.160693
- Sanchez, D., & Batet, M. (2011). Semantic similarity estimation in the biomedical domain: An ontology-based information-theoretic perspective. *Journal of Biomedical Informatics*, 44(5), 749–759. doi:10.1016/j.jbi.2011.03.013 PMID:21463704
- Segall, R. S., & Lu, S. (2014). *Linkage Discovery with Glossaries*. In *Encyclopedia of Business Analytics & Optimization* (pp. 1411–1421). doi:10.4018/978-1-4666-5202-6.ch128
- Shariff, U. M., Ghada, E., Atlma, E., Fuketa, M., Morita, K., & Aoe, J.-I. (2007). Improvement of building field association term dictionary using passage retrieval. *Information Processing & Management*, 43(6), 1793–1807. doi:10.1016/j.ipm.2006.12.006
- Sukkarieh, J. Z., Pulman, S. G., & Raikes, N. (2003). Auto-marking: Using computational linguistics to score short, free text responses. *Proceedings of the 29th Annual Conference of the International Association for Educational Assessment (IAEA)*. Retrieved from <http://www.bibsonomy.org/bibtex/2262f8f1acf4f7a6bc9fb271f809219ee/ifland>
- Volz, J., Bizer, C., Gaedke, M., & Kobilarov, G. (2009). Silk-A link discovery framework for the web of data. *Proceedings of the Linked Data on the Web Workshop (LDOW2009)*. Retrieved from http://ceur-ws.org/Vol-538/ldow2009_paper13.pdf
- Wang, M., & Si, L. (2008). Discriminative probabilistic models for passage based retrieval. *SIGIR'08: Proceedings of the ACM Special Interest Group on Information Retrieval*, 419–426. doi:10.1145/1390334.1390407
- Wikipedia. (2016a). *Glossary*. Retrieved from <http://en.wikipedia.org/wiki/Glossaries>
- Wikipedia. (2016b). *Semantic analysis (linguistics)*. Retrieved from [http://en.wikipedia.org/wiki/Semantic_analysis_\(linguistics\)](http://en.wikipedia.org/wiki/Semantic_analysis_(linguistics))
- Zamar, O., & Etzioni, O. (1999). Grouper: a dynamic clustering interface to Web search results. *Proceedings of the 8th international conference on World Wide Web, Computer Networks: The International Journal of Computer and Telecommunications Networking*, 31, 1361–1374. doi:10.1016/S1389-1286(99)00054-7
- Zelikovitza, S., & Hirsh, H. (2002). Integrating background knowledge into nearest-neighbor text classification. *ECCBR'02 Proceedings of the 6th European Conference on Advances in Case-Based Reasoning*, 1–5. doi:10.1007/3-540-46119-1_1

ADDITIONAL READING

- Australian Institute of Health and Welfare. (2016). Data linkage publications. <http://www.aihw.gov.au/publications/data-linkage/>
- Biswas, S. Papachristou, Irwin, M.E. & Lin, S. (2003). Linkage analysis of the simulated data- evaluations and comparisons of methods. *BMC Genetics*, v.4 Supplement 1: S70. DOI: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1866509/10.1186/1471-2156-4-S1-S70>

- Blakely, T., & Salmond, C. (2002). Probabilistic record linkage and a method to calculate the positive predictive value. *International Journal of Epidemiology*, 31(6), 1246–1252. doi:10.1093/ije/31.6.1246 PMID:12540730
- Christen, P. (2006). Data Linkage Techniques: Past, Present and Future. Department of Computer Science, The Australian National University, <http://datamining.anu.edu.au/talks/2006/iapa-act-aug2006-slides.pdf>
- Epstein, P. E., Lin, X., & Boehnke, M. (2003). A tobit variance-component method for linkage analysis of censored trait data. *American Journal of Human Genetics*, 72(3), 611–620. doi:10.1086/367924 PMID:12587095
- Ferranate, A. (2009). The use of data-linkage methods in criminal justice research: A commentary on progress, problems and future possibilities. *Current Issues in Criminal Justice*, 20(3), 378–392.
- Fitrianah, D., Zen, R. A., Hidayanto, A. N., & Arymurthy, A. M. (2014). A fast method for multiple data source record linkage using hash map. *Proceedings of International Conference on Computer Engineering and Mathematical Sciences*. Langkavri, Malaysia, v.3
- Galloway, J., & Simoff, S. J. (2006). Network data mining methods and techniques for discovering deep linkage between attributes. APCCM '06: *Proceedings of the 3rd Asia-Pacific conference on Conceptual modelling*, pp. 21-32.
- Hannu, T., Toivonen, T., Onkamo, P., Vasko, K., Ollikainen, V., Sevon, P., & Kere, J. et al. (2000). Data Mining Applied to Linkage Disequilibrium Mapping. *American Journal of Human Genetics*, 67(1), 133–145. doi:10.1086/302954 PMID:10848493
- Herzog, T. N., Fritz, J., & Winkler, W. E. (2007). *Data Quality and Record Linkage Techniques*. Springer; <http://bmcmcdinformdecismak.biomedcentral.com/articles/10.1186/1472-6947-13-64>
- Hu, W., Yang, R., & Qu, Y. (2014). Automatically generating data linkages using class-based discriminative properties. *Data & Knowledge Engineering*, 91, 34–51. doi:10.1016/j.datak.2014.03.001
- Kirby, R. S., & Mason, C. A. (2015). CDC-CMS Data Linkage Training: Useful resources on record linkage, January. http://academyhealth.org/datalinkageproject/Record%20Linkage_Bibliography_2015.pdf
- Kobashi, H., Take, R., & Kawata, S. (2008). A distributed linkage method for a large amount of event data, *Proceedings of IEEE Fourth International Conference on eScience*, Indianapolis, IN, pp. 756-761. doi:10.1109/eScience.2008.173
- Kobashi, H., Take, R., & Kawata, S. (2008). A scalable linkage method for large scale event-data processing, *Fourth International Conference on Networked Computing and Advanced Information Management*, v.1, pp. 98-103. doi:10.1109/NCM.2008.10
- Lee, S., Okubo, Y., & Okada, Y. (2015). Performance evaluation of linkage pattern mining method. *Proceedings of the International Conference on Electronics and Software Science*, Takamatsu, Japan, pp. 27-34.
- Martínez, S., Valls, A., & Sánchez, D. (2011). An ontology-based record linkage method for textual microdata, *Frontiers in Artificial Intelligence and Applications*, Volume 232: Artificial Intelligence Research and Development, *Proceedings of the 14th International Conference of the Catalan Association for Artificial Intelligence*
- Ong, T. C., Mannino, M. V., Schilling, L. M., & Kahn, M. G. (2014). Improving record linkage performance in the presence of missing linkage data. *Journal of Biomedical Informatics*, 52, 43–54. doi:10.1016/j.jbi.2014.01.016 PMID:24524889
- Onkamo, P., & Toivonen, H. (2006). A survey of data mining methods for linkage disequilibrium mapping. *Human Genomics*, 2(336). PMID:16595078

Queensland Health. (2016). Queensland Data Linkage Framework, January, <https://www.health.qld.gov.au/hsu/pdf/other/qlddatalinkframework.pdf>

Randall, S. M., Ferrante, A. M., Boyd, J. H., Bauer, J. K., & Semmens, J. B. (2014). Privacy-preserving record linkage on large real world datasets. *Journal of Biomedical Informatics*, 50, 205–212. doi:10.1016/j.jbi.2013.12.003 PMID:24333482

Randall, S. M., Ferrante, A. M., Boyd, J. H., & Semmens, J. B. (2013). The effect of data cleaning on record linkage quality. *BMC Medical Informatics and Decision Making*, 13(64). doi:10.1186/1472-6947-13-64 PMID:23739011

Sakshaung, J. W., Wolter, S., & Kreuter, F. (2015). Obtaining record linkage consent: results from a wording experiment in Germany. *Survey Insights: Methods from the Field*. Retrieved from <http://surveysinsights.org?p=7288>

Schmidlin, K., Clough-Gorr, K. M., & Spoerri, A. (2015). Privacy preserving probabilistic record linkage (P3RL): A novel method for linkage existing health-related data and maintaining participant confidentiality. *BMC Medical Research Methodology*, 15(46). PMID:26024886

Sharp, S. (n.d.) Deterministic and probabilistic record linkage, National Records of Scotland, Sydney School of Public Health (2016). <http://www.gov.scot/resource/0039/00390031.pdf>

Tan, K. M., Flack, F. S., Bear, N. L., & Allen, J. A. (2015). An evaluation of a data linkage training workshop for research ethics committees. *BMC Medical Ethics*, 16(13). PMID:25888741

University of Sydney. (2016). Introductory analysis of linked data (PUBH5215) *Professional Development Course University of Sydney, Australia*. <http://sydney.edu.au/courses/uos/PUBH5215>

Winkler, W. E. (2005). Methods for record linkage and Bayesian networks. *Proceedings of Joint Statistical Meeting – Section on Survey Research Methods*, pp. 3743–3748. Also appeared as Research Report Series (Statistics #2002-05) Statistical Research Division, U.S. Bureau of the Census, Washington D.C. <https://www.census.gov/srd/papers/pdf/rrs2002-05.pdf>

KEY TERMS AND DEFINITIONS

Glossary: An alphabetical list of terms in a particular domain of knowledge with the definitions for those terms (Wikipedia, 2016).

Information Retrieval: The activity of obtaining information resources relevant to an information need from a collection of information resources. Searches can be based on metadata or on full-text indexing.

Information-Theoretic: Based upon information theory such as subfields of information security and language processing.

Latent Semantic Analysis: Statistical model of word usage that permits comparisons of semantic similarity between pieces of textual information (Foltz, 1996).

Linkage Discovery: Applications include discovery of linkage between different sections in electron publications.

Ontology-Based: Borrowing a word from traditional philosophy

Record Linkage: About how to find matching records or duplicates among entities and sections within or across files.

Semantic Analysis: In linguistics is the process of relating syntactic structures, from the levels of phrases, clauses, sentences and paragraphs to the level of the writing as a whole, to their language-independent meanings (Wikipedia, 2016).

Exploratory Data Analysis on Breast Cancer Prognosis

Mohammad Mehdi Owrang O.

American University, USA

Yasmine M. Kanaan

Howard University, USA

Robert L. Copeland Jr.

Howard University, USA

Melvin Gaskins

Howard University Hospital, USA

Robert L. DeWitty Jr.

Providence Hospital, USA

INTRODUCTION

Breast cancer is the most common female cancer in the US, the second most common cause of cancer death in women (“American Cancer Society”, 2016), and the main cause of death in women ages 40 to 59 (Siegel et al., 2012).

In 2016, it is estimated that 249,260 new cases of breast cancer will be diagnosed and estimated 40,890 breast cancer deaths; and an invasive breast cancer will be diagnosed in about 246,660 women and 2,600 men. An additional 61,000 new cases of in situ breast cancer will be diagnosed in women (“American Cancer Society”, 2016). The lifetime probability of developing breast cancer is one in six overall (one in eight for invasive disease) (“American Cancer Society”, 2013; Siegel et al., 2012).

Worldwide, breast cancer is the most frequently diagnosed cancer among women in 140 of 184 countries, according to the World Cancer Research Fund International (“BCRF”, n.d.). Since 2008, breast cancer incidence has increased by more than 20 percent and mortality has increased by 14 percent (“BCRF”, n.d.). Nearly 1.7 million new breast cancer cases were diagnosed in the last report of 2012.

Medical prognosis is an evaluative component of medicine that encompasses the science of estimating the complication and recurrence of disease and predictive survival of patients (Ohno-Machado, 2001). Medical prognosis plays an increasing role in health care outcome. Reliable prognostic models that are based on survival analysis statistics and techniques have been applied to a variety of domains with varying degrees of success (i.e., APACHE IV (Zimmerman, 2006)).

Breast cancer prognosis is a vital element of providing effective treatment for patients. It has become increasingly important that clinicians are provided with accurate prognostic information on which to base therapeutic decision as the range of options for the treatment of patients with breast cancer widens. A large number of factors, including tumor grade, tumor size, and lymph node status including other aspects may influence or correlate with prognosis for breast cancer patients.

Breast cancer prediction survivability has mainly been studied through clinical approaches, based on pathological factors such as tumor grade, tumor size and number of the positive lymph nodes, estrogen (ER), progesterone (PR), and human epidermal growth factor receptor 2 (Her2) receptors,

DOI: 10.4018/978-1-5225-2255-3.ch156

etc. Most studies are carried out in an effort to find factors that clarify the large unexplained variation in prognosis of the breast cancer patients. There is still uncertainty about the importance of most prognostic factors. There are other non-clinical factors such as age, ethnicity, obesity, and marital status that may have prognostic impact but are not used routinely in clinical practice.

This chapter is a survey of the significance of the non-clinical prognostic factors (i.e., age, ethnicity) in finding the prognosis for breast cancer patients. The National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) Public-Use Data (years 1977-2013; 421,056 records) ("SEER," n.d.) and the Howard University Cancer Center Tumor Registry (1995-2013; 1599 records) data were analysed. NPI (Nottingham Prognostic Index (Galea, 1992)) is a prognostic tool that enables grouping of patients based on calculated prognosis (i.e., excellent, good, moderate, poor) and the impact of each non-clinical prognostic factor on these subgroups. In addition, survival analysis tools such as Cox proportional hazards and Kaplan-Meier survival curve ("Cox proportional-hazards regression", 2013) were used in analyzing the data.

Our data analysis suggested that age, ethnicity, and marital status have some influence on the survivability rate of breast cancer patients. Whether such influence is significant enough in relationship with the clinical factors such as tumor size and grade remains to be further studied.

BACKGROUND

Several studies have been carried out on the survivability prediction of breast cancer using Naïve Bayes and Classification Trees, Artificial Neural Networks and statistical techniques of regression (Delen et al., 2005; Gupta et al., 2011). Delen and colleagues (Delen et al., 2005) have used data mining algorithms Artificial Neural Networks, decision trees, and logistic regression to develop

the breast cancer prediction models. The results indicated that the decision tree (C5) is the best predictor with 93.6% accuracy on the sample, artificial neural networks came second with 91.2% accuracy and the logistic regression models came to be the worst of the three with 89.2% accuracy.

Investigators (Owring & Hosseinkhah, 2013; Owring, 2014) have used the association rules data mining technique and data mining tool XLMiner ("XLMiner", n.d.) to investigate the significance of the ethnic factor on the patient's prognosis. The results, the association rules, suggested that the ethnicity had some significance in the survivability rate of the patient.

Existing reports suggest that young breast cancer patients have poorer outcomes compared to their older counterparts, which are in part attributed to later stage disease, more aggressive tumors, and less favorable receptor status (Anders, 2009; Bharat, 2009; Saghie et al., 2006). It appears that ethnicity may further influence breast cancer prognosis (Biffi et al., 2001; Liu et al., 2013). The large SEER Program and National Cancer Data Base analyses indicated that African-American women have more advanced disease at presentation and a higher death rate compared with non-Hispanic Caucasians. Investigators (Biffi et al., 2001) studied the influence of ethnicity on the prognosis of young breast cancer patients (≤ 35 years). Their study shows that Latino patients suffer more aggressive disease and poorer prognosis than other young women.

Most existing studies reflect single institution experiences that may not be representative of the whole population. Additionally, existing reports did not attempt to compare patients based on age and ethnicity relative to patients' prognostic group (i.e., excellent or good). This motivated us to examine the prognosis for breast cancer patients with respect to age, ethnicity, and marital status with a new approach based on patients' prognostic groups (i.e., excellent, moderate, good, poor).

BREAST CANCER PROGNOSIS

The prognosis of a patient diagnosed with cancer is often viewed as the chance that the disease will be treated successfully and that the patient will recover. Accurate prediction of survival is vital for planning effective breast cancer treatment.

To determine how appropriate a medical intervention is for a patient, physicians are required to estimate benefits and risks of each treatment based on their predicted prognosis. Oncologist most often use several factors to determine prognosis. The most common items are age, ethnicity, tumor stage, tumor size, nodal involvement, tumor grading, lymphovascular involvement, hormone receptor status, HER-2 receptor status and tumor differentiation (Bradley, 2007; Soerjomataram et al., 2008). They also may use a 23-gene panel (oncotype DX) to determine risk assessment for some early stage cancers. An exact percentage determination in most cases is not necessary as adjuvant therapy is given for most breast cancers. Prognostic determinants are very valuable in determining the necessity for aggressive therapy in early-stage breast cancers.

Breast Cancer Prognostic Factors

Although scientists do not know the exact cause of most breast cancers, they do know some of the risk factors that increase the likelihood of a woman developing breast cancer.

A prognostic factor may be defined as a measurable variable that correlates with the natural history of the disease. The prognostic factors used in the prediction of survival of breast cancer patient can be separated into two categories: chronological (based on the amount of time present, i.e., stage of cancer), or biological (based on the potential behavior of the tumor, i.e., histological grade) (Bradley, 2007; Soerjomataram et al., 2008).

Lymph node status, tumor size, histological grade are among the prognostic factors in use today (Bradley, 2007; Soerjomataram et al., 2008). Lymph node status is a time-dependent factor and

is directly related to survival. One of the most significant prognostic factors in breast cancer is the presence or absence of axillary lymph node involvement, which is usually assessed at the time of surgery using sentinel lymph node biopsy or axillary dissection (Bradley, 2007).

Survival is inversely related to the size of the tumor. The probability of long term survival is better with smaller tumor than with larger tumor (Soerjomataram et al., 2008). Tumor size has long been recognized as an independent prognostic factor and as a predictor of axillary node status, with larger tumors being associated with a poor prognosis and an increased likelihood of nodal metastasis.

Histological grade is being identified as being highly correlated with long term survival. Patients with a grade 1 tumor have a much better chance of surviving than patients with grade 3 tumor ("Breast Cancer," 2012). Delen (Delen et al, 2005) conducted sensitivity analysis on artificial neural networks model in order to gain insight into the relative contribution of the independent variables to predict survivability. The sensitivity results indicated that the prognosis factor "Grade" is by far the most important predictor, which is consistent with the previous research, followed by "Stage of Cancer", "Radiation", and "Number of Primaries". Why these prognostic factors are more important predictors than the other is a question that can only be answered by medical professional and further clinical and analytical studies.

Other factors include the patient's age, estrogen-receptor and progesterone-receptor levels in the tumor tissue, biological and genetic properties of the cancer cells (these properties, which are sometimes called biomarkers, can be determined by specific lab and imaging tests).

Breast Cancer Prognostic Tools

There have been efforts in building breast cancer risk assessment and prognostic tools. For example, National Cancer Institute (NCI) Breast Cancer Risk Assessment Tool ("National Cancer

Institute,” n.d.) uses a woman’s risk factors (i.e., age, menstruating at early age, older age at first birth or never having given birth, family history of breast cancer, etc.) to estimate her risk for breast cancer during the next 5 years and up to age 90. Multiparametric tools have been developed for breast cancer prognosis. These tools are briefly overviewed next.

Adjuvant! Online

Adjuvant! Online (“Adjuvant Online,” n.d.) provides the assessment of probability of survival and benefit from specific therapies (i.e., adjuvant) in patients with early breast cancer. The model takes into account age, menopausal status, comorbidities, ER status, tumor grade, tumor size, and number of involved axillary nodes.

PREDICT

PREDICT (“PREDICT,” n.d.), an online prognostication and treatment benefit tool designed to help clinicians and patients make informed decisions about treatment following breast cancer surgery. The survival estimates, presented both with and without adjuvant therapy (hormone therapy, chemotherapy), are provided for 5 and 10 years following surgery. The model uses the factors including patient age, tumor size, tumor grade, number of positive nodes, ER status, HER2 status, Ki67 status and mode of detection. Survival estimates, with and without adjuvant therapy, are presented in visual and text formats.

Nottingham Prognostic Index (NPI)

The Nottingham Prognostic Index (NPI) (Galea et al., 1992) is another prognostic tool used, widely in Europe, to determine prognosis following surgery for breast cancer. Its value is calculated using three pathological criteria: the size of the tumor; the number of involved lymph nodes; and the grade of the tumor.

The index is calculated using the formula: $NPI = [0.2 \times S] + N + G$, Where:

S is the size of the tumor in centimeters,
N is the number of lymph nodes involved: 0 =1, 1-3 = 2, >3 = 3,
G is the grade of tumor: Grade I =1, Grade II =2, Grade III =3

The interpretation is as follows:

Score 5-Year Survival Prognosis

≥2.0 to ≤ 2.4 93% Excellent Prognosis

> 2.4 to ≤ 3.4 85% Good Prognosis

> 3.4 to ≤ 5.4 70% Moderate Prognosis

> 5.4 50% Poor Prognosis

SIGNIFICANCE OF NON-CLINICAL PROGNOSTIC FACTORS ON PROGNOSIS

There have been some attempts to show that there are non-clinical factors including ethnicity, genetic, diabetics and obesity that may have prognostic impact on prognosis (Lan, 2014; Owrang, 2014). The results were found to have varying effects on breast cancer survival rates in different studies.

In the following, we describe the application of prognostic tool and statistical analysis in understanding the significance of the breast cancer prognostic factors. The two approaches can complement each other in providing a better estimate of the influence that a factor might have on the overall patient’s prognosis.

Significance of the Breast Cancer Prognostic Factors Based on Prognostic Tools

We have performed several analyses using NPI tool in order to show the significance of the non-clinical prognostic factors within different prognostic groups (i.e., excellent prognosis). The Adjuvant! Online and PREDICT have not been considered for the analysis as they are primarily used to calculate an individual patient’s prognosis.

Table 1. Variables considered in the analysis

Categorical Variable Name	Distinct Values
Behavior Code	4
Ethnicity	19
Marital Status at Diagnosis	6
Extension of Tumor	23
Radiation	9
Lymph Node Involvement	10
Grade	5
Diagnostic Confirmation	8
Stage of Cancer	5
Cause of Death	2
Primary Site	10
Continuous Variables	Range
Tumor Size (cm)	0-10
RX Summ -Surgery Primary Site	0-99
Number of Primaries	1-8
Number of Positive Nodes	0-50
Age at Diagnosis	17 – 102
ER status	2
PR status	2
HER2	2
Survival Time Recode Total Months	0 – 83

SEER's breast cancer dataset for the years 1977-2013 (421,056 records) has been analysed. Table 1 shows the attributes that are considered in the analysis.

Breast Cancer Prognosis Based on Ethnicity

Table 2 shows the calculated prognosis using NPI tool for different ethnic groups. In Table 2, the results suggest that there are differences in the survivability rates among ethnic groups. For example, in comparison of white and black patients, the calculated value for excellent prognosis is 19.31% and 11.8% respectively. To see if this disparity in the prognosis among ethnic groups is statistically significant, we have performed hypothesis testing (two sample proportion with summary) using StatCrunch statistical package ("StatCrunch," n.d.). For all categories of the prognosis (i.e., excellent prognosis), the hypothesis testing for the ethnic groups resulted in a P-Value <0.0001 in every case. The P-value indicates that we should reject the hypothesis that there are no differences within ethnic groups and that the differences between groups are statistically significant. Japanese patients have better survival rates than the other groups.

Table 2. Calculated prognosis using NPI prognostic tool for ethnicity

Ethnicity	Prognosis				
	Excellent	Good	Moderate	Poor	Total
Japanese	861	1161	1379	423	3824
%	22.52	30.36	36.06	11.06	
White	58044	80600	117634	44310	300588
%	19.31	26.81	39.13	14.74	
Black	4263	6882	16595	8387	36127
%	11.80	19.05	45.94	23.22	
Chinese	882	1323	2063	724	4992
%	17.67	26.50	41.33	14.50	
Others	2135	3558	6057	2427	14177
%	15.06	25.10	42.72	17.12	

Breast Cancer Prognosis Based on Marital Status

Similar analysis has been performed to study the significance of the marital status on breast cancer prognosis using the NPI tool (Table 3). The results suggest that married patients have better prognosis than single (never married, widowed, divorced) patients.

Breast Cancer Prognosis Based on Age

We have analysed Howard University breast cancer dataset using NPI prognostic tool. We considered patients of age (45 or less) as young and (45 plus) as old.

Table 4 shows the NPI prognostic categories (excellent, good, moderate, poor) for the younger and older black breast cancer patients.

Table 4 shows comparison between black younger and older patients such that values for “poor prognosis” is 37.74% and 25.97% respectively. The younger patients have less chance of

survival than the older patients. We have tested the differences using a two proportion sampling with summary hypothesis testing. A P-value of <0.0029 has been obtained, signifying that the difference is statistically significant. On other prognostic categories (i.e., excellent, good, and moderate), there is a difference between the younger and older patients; however, the difference is not statistically significant ($P\text{-value} \geq 0.05$). The findings from Table 4 can be summarized as prognosis tends to be poor in younger black women than in older black women and the difference is statistically significant, only on the “poor prognosis” category. However, such differences in other prognostic categories (i.e., moderate) are not statistically significant. This result, off course, can be hindered by small sample sizes.

To further investigate the notion that age (younger ≤ 45 and older > 45) has correlation with worse prognosis, we have used NPI tool on SEER breast cancer data set (1998-2013) with more than 400,000 records. Table 5 shows the results for the different prognostic categories for younger and older black and white patients. Simi-

Table 3. Calculated prognosis using NPI prognostic tool for marital status

Marital Status	Prognosis				
	Excellent	Good	Moderate	Poor	Total
Married	39007	54622	84040	30980	208649
%	18.70	26.20	40.30	14.80	
Single	25076	36196	56100	23828	141200
%	17.70	25.60	39.70	16.90	
P-Value	<0.0001	0.0003	0.0012	<0.0001	

Table 4. NPI prognosis categories based on ethnicity and age for Howard University Breast Cancer Dataset

Black/ Age	Prognosis								
	Excellent	%	Good	%	Moderate	%	Poor	%	Total
≤ 45	4	2.52	19	11.95	76	47.80	60	37.74	159
> 45	34	4.91	127	18.3	352	50.79	180	25.97	693
P-Value	0.19		0.05		0.50		0.0029		

lar to Howard University breast cancer dataset, data analysis shows that black younger patients (30.40%) compared to older patients with (21.49%) have “poor prognosis”. That is, the younger black patients have lower survival chance than the older black patients. We have tested the difference using a two proportion sampling with summary hypothesis testing. A P-value of <0.0001 confirms that the difference is statistically significant. With respect to “moderate prognosis”, the younger black patients have higher percentage than the older ones. The difference is statistically significant. That is, the younger are doing better than the older patients. Considering the “excellent” and “good” prognoses, the older black patients have higher percentages than the younger patients. In both cases, a P-Value of <0.0001 confirms that the differences are statistically significant. Therefore, the younger patients are faring worse than the older patients in these prognostic categories.

We should point out that in Table 4 (prognostic categories based on Howard University Data Set), the differences between the young and old black patients in the prognostic categories of “excellent”, “good”, and “moderate” were not statistically significant.

Considering young white patients in Table 5, similar to young black patients, they also have higher percentage of “poor Prognosis” (22.05%) than older white patients (13.55%). The difference is of course statistically significant. Regarding the other prognostic categories of excellent, good, and moderate, the differences between the young and

old white patients were also statistically significant (P-Value <0.0001). Overall, older white patients are doing better than the younger patients in every prognostic categories.

Breast Cancer Prognosis Based on Age and Ethnicity

Considering Table 5, on the “poor prognosis” category, the younger black patients have higher percentage (30.40%) than the younger white patients (22.65%). The difference between the two ethnic groups is statistically significant. This result suggests that the younger black patients have higher rate of poor prognosis than the white counterpart. Likewise, for the “moderate prognosis”, the difference was statistically significant. Similar results have also been observed for older black and white patients for the “poor prognosis” and “moderate prognosis” categories.

On the “excellent prognosis” and “good prognosis”, young black patients have lower percentage compared to young white patients. The difference between the two ethnic groups is statistically significant. This shows that younger black patients are doing poorer than younger white patients. Similar results have been observed regarding the older black and white patients.

Overall, young patients had worse prognosis when compared with the older group for the “poor prognosis” category, in both Howard University and SEER datasets.

Table 5. NPI prognosis categories based on ethnicity and age for NIH SEER Breast Cancer Data set

Ethnicity /Age	Prognosis								
	Excellent	%	Good	%	Moderate	%	Poor	%	Total
Black									
≤45	472	6.74	901	12.87	3499	49.99	2128	30.40	7000
>45	3791	13.02	5981	20.53	13096	44.96	6259	21.49	29127
P-Value	<0.0001		<0.0001		<0.0001		<0.0001		
White									
≤45	4460	11.34	7488	19.04	18471	46.97	8906	22.65	39325
>45	53584	20.51	73112	27.98	99163	37.96	35404	13.55	261263
P-Value	<0.0001		<0.0001		<0.0001		<0.0001		

Significance of the Breast Cancer Prognostic Factors Based on Statistical Analyses

Several statistical methods can be identified to rank prognostic factors based on their influence on the outcome of a disease. These schemes include Multivariate Data Analysis (Peat, 2005), Sensitivity Analysis (Peat, 2005), and Survival Analysis (Peat, 2005; “Cox proportional-hazards regression”, 2013). Cox proportional hazards model (“Cox proportional-hazards regression”, 2013) has been used extensively on a modelling approach to the analysis of survival data. The purpose of the model is to simultaneously explore the effects of several variables on survival.

We have used the Cox model to estimate the hazard ratio associated with each breast cancer prognostic factor. The hazard function is the probability that an individual will experience an event (for example, death) within a small time interval, given that the individual has survived up to the beginning of the interval. It can therefore be interpreted as the risk of dying at time t . Interpreting the Cox model involves examining the coefficients for each explanatory variable. A positive regression coefficient (the B value) for an explanatory variable means that the hazard is higher and thus the prognosis is worse. Conversely, a negative regression coefficient implies a decreased hazard and increased survival times. Exp (B) is the ratio of hazard rates that are one unit apart on the predictor. As an example, if the predictor age has a coefficient of 0.03, then the hazard rate increases by 0.03 (3%) with each unit increase in age.

Table 6 shows the result of running the Cox regression hazard ratio for the SEER breast cancer using the IBM SPSS Statistical software for PC. It shows that except for the ethnic factor, other factors have higher hazard and thus the prognosis is worse. A negative regression coefficient for ethnicity implies a decreased hazard and increased survival time.

Table 6. Cox Regression Hazard Ratio for the SEER breast cancer dataset 1977-2013

	B	Sig.	Exp(B)
Tumor_Grade	.185	.000	1.203
Lymph_Nodes_positive	.016	.000	1.016
New_Ethnicity	-.053	.000	.948
New_Marital_Status	.099	.000	1.104
New_Age_Group	.239	.000	1.270
New_ER_status	.369	.000	1.447
New_PR_status	.477	.000	1.612
New_Tumor_Size	.429	.000	1.536
New_HER2	.022	.038	1.001

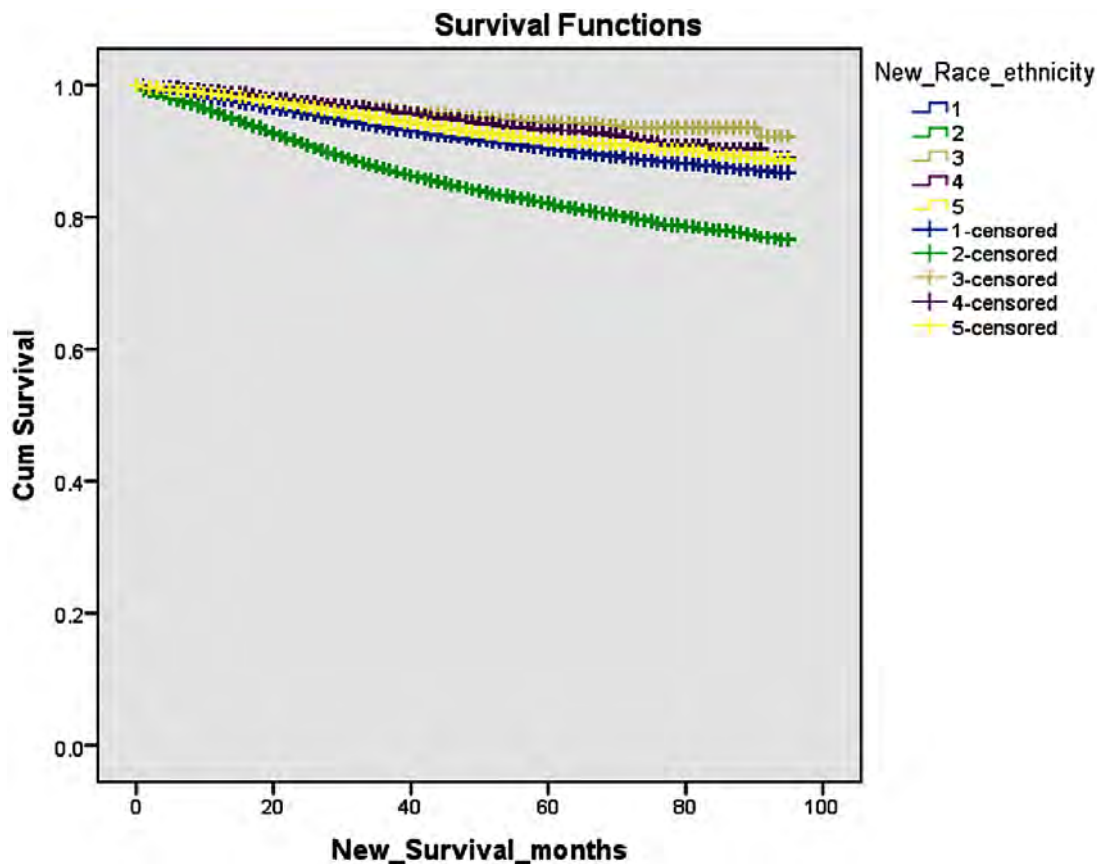
While Cox model compares covariates on overall hazard ratio, Kaplan-Meier survival curve (“Cox proportional-hazards regression”, 2013) compares the groups within a factor. It can be used to assess the statistical significance of any observed differences in a single covariate. Figure 1 shows the Kaplan-Meier survival curve for the ethnicity factor. It is the test of equality of surviving distribution for the different levels of the ethnicity factor. The graph shows that Japanese patients, first line from the top, have the best survival rate, followed by Chinese, Others, White, and Black Patients (2nd, 3rd, 4th, and 5th lines from the top of the graph, respectively).

Similar result has been obtained for the Marital Status factor based on Kaplan-Meier survival curve which shows that married patients have higher survival rate than single (windowed, divorced, never married) patients.

FUTURE RESEARCH DIRECTIONS

Current medical literature suggests that factors such as eating habit, work/occupational environment, genetics and family history, obesity could influence the survivability rate. More experiments should be conducted based on these factors to determine their impact on the prognosis. In addition, we must study the process in which existing

Figure 1. Kaplan-Meier survival curve for ethnicity prognostic factor
(1=White, 2=Black, 3=Japanese, 4=Chinese, 5=Others)



breast cancer prognostic tools (i.e., NPI) can be extended in order to accommodate the non-clinical prognostic factors in finding the survivability rate of breast cancer patients more accurately. Factors can be ranked based on their significance on the overall survivability (using sensitivity analysis, feature/factor selection, principle component analysis etc.) and weighted according to their ranking. Finally, such extended prognostic tool/system need to be evaluated for its usefulness in a clinical practice environment.

CONCLUSION

Accurate prediction of survival is an essential part of the decision making process following surgery for early breast cancer which allows clinicians to determine which patients will benefit from a treatment (i.e., Adjuvant therapy).

Breast cancer prognosis is generally understood through clinical approaches based on prognostic factors such as tumor grade, tumor size, lymph node status, ER, PR, Her2 receptors among others.

The aim of this chapter was to better understand the influence that non-clinical factors such as age, ethnicity, and marital status might have on breast cancer prognosis. Data analysis have been performed using the NPI prognostic tool, COX proportional hazard ratio and Kaplan-Meier survival curve to see if non-clinical factors have any impact on the overall survivability rate of the breast cancer patients.

Overall, data analysis indicates that breast cancer patients have different survivability rate considering their age, ethnicity, and marital status and such disparities in the final prognosis are statistically significant in most cases.

REFERENCES

- Adjuvant!OnLine. (n.d.). Retrieved from www.adjuvantonline.com/breast.jsp
- American Cancer Society. (2013). *How is Breast Cancer Staged*. Retrieved from <http://www.cancer.org/cancer/breastcancer/detailedguide/breast-cancer-staging>
- American Cancer Society. (2016). *Breast Cancer Facts & Figures*. Retrieved from <http://www.cancer.org/>
- Anders, C. K., Johnson, R., Litton, J., Phillips, M., & Bleyer, A. (2009). Breast cancer before age 40 years. *Seminars in Oncology*, 36(3), 237–249. doi:10.1053/j.seminoncol.2009.03.001 PMID:19460581
- Bharat, A., Aft, R. L., Gao, F., & Margenthaler, J. A. (2009). Patient and tumor characteristics associated with increased mortality in young women (< or =40 years) with breast cancer. *Journal of Surgical Oncology*, 100(3), 248–251. doi:10.1002/jso.21268 PMID:19330813
- Biffi, W. L., Myers, A., Franciose, R. J., Gonzalez, R. J., & Darnella, D. (2001). Is breast cancer in young Latinas a different disease? *American Journal of Surgery*, 182(6), 596–600. doi:10.1016/S0002-9610(01)00789-9 PMID:11839323
- Bradley, K. T. (2007). *Prognostic and Predictive Factors in Breast Cancer*. Retrieved from <http://www.cap.org>
- Breast Cancer Q & A/Facts and Statistics. (2012). Retrieved from <http://www.komen.org/bei/bhealth/QA/q-and-a.asp>
- Breast Cancer Research Foundation (BCRF). (n.d.). Received from <http://www.bcrfcure.org/breast-cancer-statistics-resources>
- Breast Cancer Risk Assessment Tool (Gail Model). (n.d.) Retrieved from <http://www5.komen.org/BreastCancer/GailAssessmentModel.htm>
- Cox proportional-hazards regression. (2013). *BMJ (Clinical Research Ed.)*, 347, f4919.
- Crowley, J., Hoering, A., & Ankerst, D. (Eds.). (2006) *Handbook of Statistics in Clinical Oncology* (2nd ed.). CRC Press, Taylor & Francis Group.
- Delen, D., Walker, G., & Kadam, A. (2005). Predicting breast cancer survivability: A comparison of three data mining methods. *Artificial Intelligence in Medicine*, 34(2), 113–127. doi:10.1016/j.artmed.2004.07.002 PMID:15894176
- Galea, M. H., Blamey, R. W., Elstone, C. E., & Ellis, I. O. (1992). The Nottingham Index in Primary breast cancer. *Breast Cancer Research and Treatment*, 22(3), 207–219. doi:10.1007/BF01840834 PMID:1391987
- Gupta, S., Kumar, D., & Sharma, A. (2011). Data Mining Classification Techniques Applied For Breast Cancer Diagnosis And Prognosis. *Indian Journal Of Computer Science And Engineering*, 2(2).
- Lan, N. H., Laohasiriwong, W., & Stewart, J. F. (n.d.). *Survival probability and prognostic factors for breast cancer patients in Vietnam*. Retrieved from <http://www.globalhealthaction.net/index.php/gha/article/view/18860>
- (NCI) National Cancer Institute. (n.d.). *Breast Cancer Risk assessment Tool*. Retrieved from <http://www.cancer.gov/bcrisktool/>
- Ohno-Machado, L. (2001). Modeling medical prognosis: Survival analysis techniques. *Journal of Biomedical Informatics*, 34(6), 428–439. doi:10.1006/jbin.2002.1038 PMID:12198763

Owring, O., & Hosseinkhah, F. (2013). Association Rules Mining for Breast Cancer Survivability Prediction. *Proceedings of the 28th International Conference on Computers and Their Applications (CATA-2013)*, 159-165.

Owring, O., & Mehdi, M. (2014). The Significance of the Race Factor in Breast Cancer Prognosis. *Proceedings of the 2014 International Conference on Data Mining*, 55-61.

Peat, J., & Barton, B. (2014). *Medical Statistics: A Guide to Data Analysis and Critical Appraisal* (2nd ed.). BMJ Books.

PREDICT. (n.d.). Retrieved from <http://www.predict.nhs.uk.technical.shtml>

SEER. (n.d.). *Surveillance, Epidemiology, and End Results*. Retrieved from www.seer.cancer.gov

Siegel, R., Naishadham, D., & Jamal, A. (2012). Cancer Statistics. *CA: A Cancer Journal for Clinicians*, 62(1), 10–29. doi:10.3322/caac.20138 PMID:22237781

Soerjomataram, I., Louwman, W. J., Ribot, J. G., Roukema, J. A., & Coebergh, J. W. W. (2008). An overview of prognostic factors for long-term survivors of breast cancer. *Breast Cancer Research and Treatment*, 107(3), 309–330. doi:10.1007/s10549-007-9556-1 PMID:17377838

StatCrunch. (n.d.). Retrieved from <http://www.statcrunch.com/>

XLMiner On Line. (n.d.). *User Manual*. Retrieved from <http://www.solver.com/xlminer-data-mining>

Zimmerman, J. E., Kramer, A. A., McNair, D. S., & Malila, F. M. (2006, May). Acute Physiology and Chronic Health Evaluation (APACHE) IV: Hospital mortality assessment for today's critically ill patients. *Critical Care Medicine*, 34(5), 1297–1310. doi:10.1097/01.CCM.0000215112.84523.F0 PMID:16540951

ADDITIONAL READING

El Saghir, N. S., Seoud, M., Khalil, M. K., Charafeddine, M., Salem, Z. K., Geara, F. B., & Shamseddine, A. I. (2006). Effects of young age at presentation on survival in breast cancer. *BMC Cancer*, 6(1), 194. doi:10.1186/1471-2407-6-194 PMID:16857060

Endo, A., Shibata, Hiroshi, & Tanaka, Hiroshi. (2008). Comparisons of Seven Algorithms to Predict Breast Cancer Survival. *Biomedical Soft Computing and Human Sciences*, 13(2), 11–16.

Hill, T., & Lewicki, P. (2007). *STATISTICS: Methods and Applications*. Tulsa, OK: StatSoft.

SEER. (n.d.) Retrieved from http://seer.cancer.gov/csr/1975_2010/

Weka, (n.d.) Retrieved from <http://www.cs.waikato.ac.nz/ml/weka/>

What are Hazard Ratios.(n.d.) Retrieved from http://www.medicine.ox.ac.uk/bandolier/painres/download/whatis/what_are_haz_ratios.pdf

What is Cox Model.(n.d.) Retrieved from http://www.medicine.ox.ac.uk/bandolier/painres/download/whatis/cox_model.pdf

What is Sensitivity Analysis.(n.d.) Retrieved from http://www.medicine.ox.ac.uk/bandolier/painres/download/whatis/What_is_sens_analy.pdf

KEY TERMS AND DEFINITIONS

Breast Cancer: A malignant (cancerous) growth that begins in the tissues of the breast in an uncontrolled way.

Data Analysis: The process of systematically applying statistical and/or logical techniques to describe and evaluate data.

Prognosis: The prediction of the survivability from a disease.

Prognostic Factor: A measurable variable that is used in the prediction of survival of breast cancer.

Prognostic Tool: A software tool that uses risk factors (i.e., age, ethnicity, genetic) to estimate breast cancer survivability rate.

Survival Analysis: A field in medical prognosis that deals with the application of various methods to estimate the survival of a particular patient suffering from a disease.

Survivability Rate: The probability of surviving and recovering from a disease.

D

In-Memory Analytics

Jorge Manjarrez Sánchez

Instituto Tecnológico Superior de Jerez, Mexico

INTRODUCTION

Processing speed and data set volume are a challenge for traditional analytics systems. Consider when do you want to make a recommendation for an online customer: right there while she is still browsing your catalogue or next time (if) she returns back? When do you want to detect a credit card is fraudulent: before a transaction is committed or after a client complain some days later? Score or decide when a credit card transaction is fraudulent or an incoming email is spam, is relatively easy from a computationally point of view provided the appropriate algorithms are implemented correctly. But in real-life systems where hundreds of thousands of such transactions arrive simultaneously and the correct decision must be taken instantaneously or in near real-time, speed is crucial. Even though only for the second example question an incorrect false negative, i.e., the no detection of a fraud, can have severe negative consequences, these examples highlight two main challenges in data analytics: the need to process ever larger data volumes and the need of obtaining fast or real time results. These are the concerns of In-memory analytics.

In what follows, we provide a definition and some concepts necessary to understand the problem and solution approach of in-memory analytics. Then we present a review of some important technological proposals to deal with two main challenges in data analytics: speed and scalability. The presentation is made from a data management perspective, independent of their nature (structured or unstructured), because an efficient data management strategy provides the infrastructure to handle and enable fast access to voluminous data sets for their analytical processing.

DOI: 10.4018/978-1-5225-2255-3.ch157

BACKGROUND

Analytics is the processing of data for facts and information discovery. Embedded within data there are facts about some events, perhaps several and a variety of them. These facts are knowledge helpful for taking decisions, recommendations systems, fraud detection, sentiment analysis in social networks and customer identification and many other applications. In a broad sense analytics can provide an answer for questions such as what happened? Why did it happen? What will happen? What should I do? And they correspond to the task of descriptive, diagnostic, predictive and prescriptive analytics respectively. The answer to these questions are obtained by the application of techniques from statistics, machine learning and computer science in the processing of data.

Analytics is the processing of data for making sense of all of it. Data sets are very large and to process them one has to cope with two orthogonal concerns: accuracy and speed of results. Accuracy needs to process as much data as possible, not just random samples. Not using complete datasets leads to loss of information (Wang, Callan, & Zheng, 2015) and obtained responses are approximations with certain level of error. Accuracy of the statistical model and henceforth that of predictions are proportional to samples size and quality. The bigger the better, but even that we can have today larger data sets, their processing takes considerable time. It consumes more processing time and disk input/output (IO) operations because data must be swapped back and forth from hard drives to RAM and CPU caches. But due to speed differences in data access and transfer between these devices, there is a processing bottleneck. One approach to avoid IO processing bottlenecks is to use as

much as possible RAM memory, hence the name In-Memory Analytics, where data is loaded and kept into RAM for their fast consumption by the processing algorithms.

FAST ANALYTICS

In this section we address the speed challenge. When the speed of results is a factor to consider for efficient analytics, and taking into consideration that it must be used as much data as possible for the abovementioned reasons, then one must reduce or eliminate processing bottlenecks. They can be the IO latency and CPU usage. CPU is good at computing mathematical operations and multicore CPU's enhance their performance. IO depends on the type of disk and RAM. Solid state disks are faster than hard disk drives, but random access memory is faster. A typical hard drive has a latency of 5 ms while RAM has only 100 ns, it is 50000 times faster. But also the faster the more expensive. There is one more rapid type of memory: cache memory, which is a small and expensive memory within the CPU die. Main memory or In-memory refers to the processing of data performed in RAM.

The data management infrastructure provides the processing capabilities according to the workload type. The typical workloads in a database, can be Online Analytical Processing (OLAP) or Online Transaction Processing (OLTP). The difference between both are their usage patterns, size of result sets and processing speed. Traditionally OLTP deals with faster small transactions and OLAP seeks for answer precision on very large data sets, but nowadays speed is an important factor to provide for both of them, hence the use of in-memory analytics, the usage of RAM fast data access and to enhance analytics performance. But the issue is that real data sets are usually larger than the memory affordable and available in a system. Also note that not all available memory can be used for data storage, enough space must be allocated to handle temporary objects and

intermediate results of computations. To address this concern, one approach is to use compression. For instance, by using compression and column-based storage it is possible to reduce storage space because only required attributes are retrieved from disk. Additionally, compression leverages CPU ability to handle numbers instead of text which combined with columnar storage allows an application to keep large quantities of data in memory. A drawback is that this approach is read optimized, insert or update operations are not easy. Some other approaches to make efficient use of memory while computing intermediate results are discussed in (Duan, Li, Tang, Xiao, & Li, 2013) within the context of Spark, an open source in-memory analytics platform. Their proposal makes automatic management of the memory space used by intermediate results susceptible of being used in further computations instead of letting the full responsibility to the programmer.

The idea of in-memory databases is not new (Lehman & Carey, 1986), it has been studied since more than two decades ago, but nowadays it is possible for real enterprise applications to process larger amounts of data because of RAM price drops. In addition to high prices, there was a technical limitation of the operating systems for the addressing of larger memory beyond some GB. Today these limitations have gone and prices started to decrease and now it is common for a single server to have 1 TB of RAM and a cluster of computers to sum up several Terabytes. The scalability issue to cope with big data sets is the topic of the next section.

In-memory analytics appliances from major vendors integrate an optimized computer with an in-memory database. The main database vendors have developed their in-memory proposal: Microsoft Hekaton (Diaconu et al., 2013) is an engine integrated within SQL Server, the memory optimized part of the database is marked as memory optimization. Code of stored procedures can also be compiled to enhance CPU usage. Oracle TimesTen (Lahiri & Folkman, 2013) is a separate product that can also be used as a fast data cache

for the mainstream Oracle database. SAP HANA (Plattner, 2014b) is one of the first proposals for columnar in-memory databases. Recent HANA improvements can handle successfully OLAP and OLTP workloads. Evidently all these systems provide mechanisms such as logging to achieve durability, transaction processing with full ACID (Atomicity, Consistency, Isolation, Durability) support, and some ad hoc enhancements. For instance, TimesTen can be used almost directly on a grid of computers.

Even if some analytical algorithms are implemented within the database (in-database) to have direct access to data and hence minimize data movement, data must be loaded into RAM to allow computation. The advantage of in-database analytics is that programmed algorithms can make use of some database structures that accelerates data access in addition to security and fault tolerance. In (Ordonez, 2013) it is presented a survey of some statistical and machine learning techniques that have been implemented inside a DBMS. The same article elucidates some approaches to program new analytic algorithms as User Defined Functions when it is not possible to modify the source code. We argue that a combination of in-database with in-memory obtains a better solution, indeed the same article mentions that programmed procedures in C or a similar language allows to push some computations to main memory.

But in spite of drop in memory prices or elimination of technological barriers not all required data can fit into memory and nor all database management systems or their processes are optimized to work into memory. Hence the need for a scalable proposal based on the combination of in-memory processing with parallel and distribute settings.

SCALABLE ANALYTICS

Now we turn to the data volume challenge. Suppose you have enough memory in a single server to handle small to medium size data sets. It arises

the situation when memory or processing power of a single machine is not enough. The efficient design of a scalable analytics infrastructure have been studied in (Zhang & Ré, 2014) and (Otto & Weber, 2013) among others. Zhang proposes a framework called DIMMWITTED to test the behavior of Support Vector Machines, logistic regression, neural networks and Gibbs sampling. The aim is to analyze the access methods (row or column), model replication and data replication to find a balance in the design space of these considerations where the efficiency of the system is maximized. The results are a guide to the design and implementation of analytical systems with a 100 times speed up in comparison with their baseline. In summary, the experiments show that there is no difference between row or columnar data access. Model replication is better when it is performed per node (the alternative are per core and per machine). Finally, the data replication, considering it is immutable (which in practice it is not often the case unless it is read only or a snapshot of a database) they simply state that avoiding data skew is a sufficient condition for achieving good performance. This can easily be attained by simple round robin partitioning of data among available processing nodes of the cluster.

Among the parallel architectures, shared-nothing is a choice because of its capability to be built from commodity machines and easy scalability. Data is partitioned across several computing nodes which process subsets of the data and then coordinate to consolidate final results. In-memory computing is less constrained. It scales far beyond a single server by using distributed in-memory computing examples are Hazelcast, memcached and Apache Spark ("Apache Spark," n.d.), but even on them memory must be used efficiently (Duan et al., 2013).

Statistical or machine learning algorithms must not only take advantage of local or single server's main memory, but to transparently scale to be able to handle big databases by the efficient use available memory in cluster or cloud settings. Although there are proposals that already exploits

elastic computing power, such as RAMCloud (J. Ousterhout et al., 2015) which proposes an in-memory distributed DRAM cache with a key value data model with low latency and fail tolerance capabilities. Low latency is in the order of microseconds and is achieved by using high speed network interconnections and guarantees 1-2 seconds availability when recovery from a failure. As usual in distributed systems there is a master which coordinates replication and helps to recover from failures.

In (Xin et al., 2013) it is proposed an scalable analytical engine based on a distributed memory abstraction. The engine is based on the MapReduce (Dean & Ghemawat, 2008) framework but performs 100 times faster than Hadoop among other improvements. For our discussion, it is important to note the in-memory abstraction, that handles all the available physical memory in the cluster to process analytical queries and reconstruct data lost in failures. Why it is important? Because RAM is a volatile memory and it is its main weakness. More on this in a section below.

Several research efforts have been made into leveraging the MapReduce framework as a starting point to develop scalable machine learning algorithms. The advantage of MapReduce is that it frees the developer from tasks allocation, communication issues and failure recovery. Also scalability of a parallel task is handled by the framework which automatically uses the available processing nodes in the distributed machine. Even if this programming model has several disadvantages (Doulkeridis & Nørnvåg, 2013) it has been used successfully in several real analytical projects. Among its enhancements there are main-memory implementations. One concern is its expensive initialization phase, but if you have to process real big data sets in offline mode MapReduce or one of its implementations such as Hadoop is a good option. Because the savings in the processing of a big data set compensates its expensive starting. A

survey of the weakness and limitations along with the proposed solutions are discussed in (Doulkeridis & Nørnvåg, 2013)

Probably one of the most notable in-memory open source implementations of MapReduce for analytics is Spark (Zaharia et al., 2012) a follow up of the project Shark (Engle et al., 2012; Xin et al., 2013), which provides tens fold speedup over Hadoop. Main Memory MapReduce, M3R (Shinnar, Cunningham, Herta, & Saraswat, 2012) aims at providing a fast MapReduce by eliminating communication overhead of the JobTracker and the heartbeat and promoting reuse of already in-memory structures.

Several machine learning algorithms used for descriptive analytics using unsupervised learning techniques such as the famous k-means (Jain, 2010) has also been implemented in MapReduce. It is unsupervised because it does not use predefined labels, it simply starts and discovers clusters of data based on their similarity. Its uses are varied and innumerable. For instance, it can be applied to discover frauds or lawbreakers or something out of normal by running the clustering process over the data points. After finishing the clustering process, the remaining data points (those that do not belong to any cluster) are called outliers. If each data point in a cluster represents a typical or acceptable behavior, then the outliers are the misbehaved ones. K-means is a computationally expensive algorithm that exemplifies that many analytical algorithms perform several scans over the data set imposing a strong overload in the IO system and thus is a viable candidate for in-memory implementation. Also to efficiently handle large data sets a parallel framework like MapReduce is often used. In (Bahmani, Moseley, Vattani, Kumar, & Vassilvitskii, 2012) a parallel version of k-means that perform an efficient initialization algorithm is proposed. Initialization is the finding of the centroids for each cluster, their choice and their number determines the quality of the clusters obtained.

FUTURE RESEARCH DIRECTIONS

Main memory is faster but their volatile nature makes it more sensitive to failures and data can be lost. This can be alleviated by implementing some failure recovery protocols, such as one based on checkpoints that saves database state at some intervals and in the presence of failures allows for recovery to the most recent saved state. Some recent research also proposes strategies for non-volatile persistence (Narayanan & Hodson, 2012).

New technologies arise to improve performance of computer applications. CPU speeds and efficiency constantly improves as well as that of RAM. A new type of storage aware non-volatile memory is being developed, it has the persistency capabilities of hard drives but preserving the speed of volatile RAM. Some prototypes have been developed to test their efficiency proving promissory results (Bailey, Hornyack, Ceze, Gribble, & Levy, 2013; S. Chen & Jin, 2015). These prototypes are based on memristors and phase-change memory which needs some improvement because of their write operations' latency, although they are still faster than solid state disks. Another drawback is that they are inconsistent with today block-oriented file systems. A recent work (Kimura, 2015) obtains positive experiments for the scalability of an ad hoc in-memory database for 240 cores and newly-emerging non-volatile memory (NVRAM).

If you want to optimize memory usage at a finer level, then consider the case of efficient CPU cache utilization. Cache is very small but faster than RAM. It is from there where the processor takes data and instructions to execute. If they are not available the processor stalls, simply meaning it does nothing. So when a programmer wants to exploit all the processor power, she must avoid cache misses and maximize cache hits. In (Tözün, Gold, & Ailamaki, 2013) experiments show that most misses come from the storage manager, specifically the indexing method in the processing of OLTP workloads.

Another research topic is related with the engineering of the software. Leveraging in-

memory processing with well-crafted off-the-shelf algorithms so data scientists or data analysts can concentrate on the design of experiments and interpretation of results without worrying about engineering high-performance algorithms.

Some other concerns are presented in (Tan et al., 2015). We do not discuss them here, because they relate to a general approach to data management and are not specific to analytics. However, as a further lecture, it can provide some clues about the issues to be solved and related to parallelism, concurrency and communication. Finally, one aspect we have not considered is the freshness of data: in order to have valid analytical results they should not be computed on stale data. So algorithms for in-memory updates or synchronization with the current data produced or stored on disk must be implemented.

CONCLUSION

Most of the statistical and machine learning algorithms required to perform analytical processing are iterative and pass over the data multiple times (take for instance the k-means algorithm). Handling the data on disk imposes a processing bottleneck and reduces system efficiency. In-memory analytics is a powerful approach to deliver fast and reliable data insights. Drop in prices and technological advances of RAM makes it affordable to have systems with larger memories capable of storing larger data sets.

IO operations are still an issue even if some new results seems to contradict this believe (K. Ousterhout, Rasti, Ratnasamy, Shenker, & Chun, 2015), but the fact is that even in that paper, it is stated that CPU is a remarkable issue because of the data compression used to save memory space. Hence, in reality the IO bottleneck is not solved, instead it is shifted to another vital resource: the CPU. Thus in the design of high performance analytics systems, one must make a tradeoff between sacrificing CPU at the cost of having a little extra space in memory by using some data compression

method or engineering efficient in-memory aware algorithms or perhaps spent some extra bucks into have a larger available memory. Memory savings are also obtained when serialization is used: data set size reduce up to three or more times but again, the corresponding deserialization mechanism required to process data consumes CPU and we are just shifting costs from one side to another.

REFERENCES

- Apache Spark. (n.d.). Retrieved from spark.apache.org
- Bahmani, B., Moseley, B., Vattani, A., Kumar, R., & Vassilvitskii, S. (2012). Scalable KMeans. In *VLDB End. PVLDB*.
- Bailey, K. A., Hornyack, P., Ceze, L., Gribble, S. D., & Levy, H. M. (2013). Exploring storage class memory with key value stores. In *Proceedings of the 1st Workshop on Interactions of NVM/FLASH with Operating Systems and Workloads - INFLOW '13* (pp. 1–8). New York: ACM Press. <http://doi.org/doi:10.1145/2527792.2527799>
- Chen, S., & Jin, Q. (2015). Persistent B + -trees in non-volatile main memory. *Proceedings of the VLDB Endowment*, 8(7), 786–797. doi:10.14778/2752939.2752947
- Dean, J., & Ghemawat, S. (2008). MapReduce: Simplified data processing on large clusters. *Communications of the ACM*, 51(1), 107–113. doi:10.1145/1327452.1327492
- Diaconu, C., Freedman, C., Ismert, E., Larson, P., Mittal, P., Stonecipher, R., ... Zwillig, M. (2013). Hekaton: SQL Server's Memory -Optimized OLTP Engine. SIGMOD.
- Doulkeridis, C., & Nørsvåg, K. (2013). A survey of large-scale analytical query processing in MapReduce. *The VLDB Journal*, 23(3), 355–380. doi:10.1007/s00778-013-0319-9
- Duan, M., Li, K., Tang, Z., Xiao, G., & Li, K. (2013). Selection and Replacement Algorithms for Memory Performance Improvement in Spark. *Concurrency and Computation*, 1–16. doi:10.1002/cpe
- Engle, C., Lupher, A., Xin, R., Zaharia, M., Franklin, M. J., Shenker, S., & Stoica, I. (2012). Shark: fast data analysis using coarsegrained distributed memory. *ACM SIGMOD International Conference on Management of Data*, 689–692. doi:10.1145/2213836.2213934
- Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8), 651–666. doi:10.1016/j.patrec.2009.09.011
- Kimura, H. (2015). FOEDUS: OLTP Engine for a Thousand Cores and NVRAM. In *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data - SIGMOD '15* (pp. 691–706). New York: ACM Press. <http://doi.org/doi:10.1145/2723372.2746480>
- Lahiri, T., & Folkman, S. (2013). Oracle TimesTen : An In-Memory Database for Enterprise Applications. *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering*, 36(2), 6–13.
- Lehman, T., & Carey, M. (1986). A Study of Index Structures for Main Memory Database Management Systems. *Proceedings of the VLDB Endowment*, 294–303.
- Narayanan, D., & Hodson, O. (2012). Whole-system persistence. In *Proceedings of the seventeenth international conference on Architectural Support for Programming Languages and Operating Systems - ASPLOS '12* (Vol. 47, p. 401). New York: ACM Press. <http://doi.org/doi:10.1145/2150976.2151018>
- Ordonez, C. (2013). Can We Analyze Big Data inside a DBMS?. DOLAP. doi:10.1145/2513190.2513198

- Otto, G., & Weber, T. (2013). Data Management Adaptive In-Memory Analytics with Teradata: Extending Your Existing Data Warehouse to Enable SAS ® High-Performance Analytics SAS Global Forum 2013 Data Management. *SAS Global Forum 2013*, 1–10.
- Ousterhout, J., Gopalan, A., Gupta, A., Kejriwal, A., Lee, C., Montazeri, B., & Stutsman, R. et al. (2015). The RAMCloud Storage System. *ACM Transactions on Computer Systems*, 33(3), 1–55. doi:10.1145/2806887
- Ousterhout, K., Rasti, R., Ratnasamy, S., Shenker, S., & Chun, B.-G. (2015). *Making sense of performance in data analytics frameworks*. Retrieved from <http://dl.acm.org/citation.cfm?id=2789770.2789791>
- Plattner, H. (2014b). The impact of columnar in-memory databases on enterprise systems. *Proceedings of the VLDB Endowment*, 7(13), 1722–1729. doi:10.14778/2733004.2733074
- Shinnar, A., Cunningham, D., Herta, B., & Saraswat, V. A. (2012). M3R: Increased performance for in-memory Hadoop jobs. *VLDB End.*, 5(12), 1736–1747. doi:10.14778/2367502.2367513
- Tan, K.-L., Cai, Q., Ooi, B. C., Wong, W.-F., Yao, C., & Zhang, H. (2015). In-memory Databases – Challenges and Opportunities From Software and Hardware Perspectives. *SIGMOD Record*, 44(2), 35–40. doi:10.1145/2814710.2814717
- Tözü, P., Gold, B., & Ailamaki, A. (2013). OLTP in wonderland. In *Proceedings of the Ninth International Workshop on Data Management on New Hardware - DaMoN '13* (p. 1). New York: ACM Press. <http://doi.org/doi:10.1145/2485278.2485286>
- Wang, Y., Callan, J., & Zheng, B. (2015). Should We Use the Sample ? Analyzing Datasets Sampled from Twitter ' s Stream API. *ACM Transactions on the Web*, 9(3).
- Xin, R. S., Rosen, J., Zaharia, M., Franklin, M. J., Shenker, S., Stoica, I., & Berkeley, U. C. (2013). Shark : SQL and Rich Analytics at Scale. *SIGMOD*, 13–24.
- Zaharia, M., Chowdhury, M., Das, T., Dave, A., Ma, J., M., M., ... Stoica, I. (2012). Fast and interactive analytics over Hadoop data with Spark. *USENIX Login*, 37(4), 45–51.
- Zhang, C., & Ré, C. (2014). DimmWitted: A Study of Main-Memory Statistical Analytics. *Proceedings of the VLDB Endowment*. Retrieved from <http://www.vldb.org/pvldb/vol7/p1283-zhang.pdf>

ADDITIONAL READING

- Alagiannis, I., Idreos, S., & Ailamaki, A. (2014, June). H2O: a hands-free adaptive store. In *Proceedings of the 2014 ACM SIGMOD international conference on Management of data* (pp. 1103–1114). ACM. doi:10.1145/2588555.2610502
- Chandramouli, B., Goldstein, J., Barnett, M., DeLine, R., Fisher, D., Platt, J. C., & Wernsing, J. et al. (2014). Trill: A high-performance incremental query processor for diverse analytics. *Proceedings of the VLDB Endowment*, 8(4), 401–412. doi:10.14778/2735496.2735503
- Leis, V., Kemper, A., & Neumann, T. (2014, March). Exploiting hardware transactional memory in main-memory databases. In *2014 IEEE 30th International Conference on Data Engineering* (pp. 580–591). IEEE doi:10.1109/ICDE.2014.6816683
- Mukherjee, N., Chavan, S., Colgan, M., Das, D., Gleeson, M., Hase, S., & Lahiri, T. et al. (2015). Distributed architecture of Oracle database in-memory. *Proceedings of the VLDB Endowment*, 8(12), 1630–1641. doi:10.14778/2824032.2824061

Mukherjee, N., Chavan, S., Colgan, M., Gleeson, M., He, X., Holloway, A.,... Macnaughton, N. (2016, May). Fault-tolerant real-time analytics with distributed Oracle Database In-memory. In 2016 IEEE 32nd International Conference on Data Engineering (ICDE) (pp. 1298-1309). IEEE. doi:10.1109/ICDE.2016.7498333

Plattner, H. (2014a). *A Course in In-Memory Data Management - The Inner Mechanics of In-Memory Databases* (2nd ed.). Springer-Verlag; doi:10.1007/978-3-642-55270-0

KEY TERMS AND DEFINITIONS

Data Analytics: The processing of data for information and facts discovery.

Descriptive Analytics: The gathering and processing of data to determine and understand

the state of things in an enterprise, gives an answer to what happened?

In-Memory Computing: Computer processing performed by loading all data in RAM in order to speed up data access.

Machine Learning: The science of developing techniques to give the computer inference and deduction capabilities to achieve diverse processing tasks autonomously.

Predictive Analytics: The analytics of data for forecasting outcomes, provides an answer to what will happen, based on estimates issued from probabilities and statistical analyses.

Prescriptive Analytics: After knowing the current state and the likelihood of the outcomes, prescriptive analytics concern is to provide guidance or advise on what an enterprise should do.

Scalability: It is the capability of a system to handle bigger workloads without compromising performance.

Innovative Formalism for Biological Data Analysis

Calin Ciufudean

Stefan cel Mare University, Romania

INTRODUCTION

Modern medical devices involves information technology (IT) based on electronic structures for data and signals sensing and gathering, data and signals transmission as well as data and signals processing in order to assist and help the medical staff to diagnose, cure and to monitors the evolution of patients. By focusing on biological signals processing we may notice that numerical processing of information delivered by sensors has a significant importance for a fair and optimum design and manufacture of modern medical devices. We consider for this approach fuzzy set as a formalism of analysis of biological signals processing and we propose to be accomplished this goal by developing fuzzy operators for filtering the noise of biological signals measurement. We exemplify this approach on neurological measurements performed with an Electro-Encephalograph (EEG).

We mention that automatic diagnosis based on fuzzy approach becomes more and more used in artificial intelligence framework of proposing new methods and tools for developing automatic medical diagnosis. Fuzzy techniques of analysis and diagnosis have a significant advantage compared to classical techniques, respectively the ability to deal with typical uncertainty and estimation that characterizes any physical system, including the human body.

This chapter is focused on reasoning with fuzzy logic for designing new type of filters capable to reject noises which overload the biological signals taken by invasive ad/or non-invasive medical tools. We will exemplify our approach on EEG measurements. This goal will be accomplished by using a

new structure and inference mechanism of fuzzy operators which will implement a new design of filters against impulsive and uniform distributed noise. This approach is inspired by the class 1-D (signals) and 2-D (images) of fuzzy filters proposed in (Zadeh, 1988; Russo & Ramponi, 1994; Russo, 1994). Building our approach we were guided by Professor Lofthi Zadeh's statement (Zadeh, 1988): "Fuzzy logic is a precise conceptual system of reasoning, deduction and computation in which the objects of discourse and analysis are, or are allowed to be, associated with imperfect information. Imperfect information is information which in one or more respects is imprecise, uncertain, incomplete, unreliable, vague or partially true". We assume that the reader is familiarized with fuzzy logic systems, or we address the reader to (Kumar, s.a., 2001; Adlassnig, 2007; Chin, 1991; Fathi-Torbaghan & Meyer, 1994; Nguyen & Kreinovich, 2001; Bounds, 1999). We notice that literature treat intensively the subject of applying fuzzy logic to biological data acquisition and processing for medical diagnosis (John & Innocent, 2005; Perner, 2002); therefore we try to add a piece in this puzzle hoping to speed up the process of having a clear picture. Automatic processing of sensory information is an actual research area, and fuzzy techniques are well framed in this content, and represents a useful and widely applicable technique with long run perspective. This chapter is focused on implementing rule-based fuzzy systems for processing, e.g. filtering, measured biologic data in order to determine the availability of the most important ones for the final diagnosis. We appreciate that our approach will framework the

computer systems for medical diagnosis. An example will emphasize our approach.

This chapter is organized as follows: section 2 briefly discuss the chapter's position on the topic of biological data processing and analysis, section 3 presents our fuzzy filter approach, section 4 deals with fuzzy inference mechanism for filtering electric signals, section 5 illustrates our approach by an example and section 6 concludes the present work and gives a few directions for future research in the area.

BACKGROUND

As modern world evolves, so does the technique, economy, the life style, and unfortunately, the diseases; therefore, nowadays mainly due to stress, pollution and rush life style we confront with a new medical problem: the diagnosis of multiple diseases and the diagnosis of mutant diseases. Therefore, nowadays mainly due to stress, low infrastructure facilities and rush life style we confront with a new problem: the diagnosis of multiple informational news and data releases of mutant informational transmission infrastructure. Failure diagnosis in complex information systems, such as medical ones, is a critical task due to respect the safe development of these systems and patient care. In an optimistic scenario, doctors are confronted with patients who suffer from one disease and the diagnosis tries to find the cause and the most adequate treatment (Cacioppo & Tassinari, 1990; Lyons, Budynek, & Akamatsu, 1999). This approach give birth to classic diagnosis, e.g. to ordered tables or lists well known in medicine as well as in medical informatics that make a correspondence between symptoms and disease (Ekman, Levenson, & Friesen, 1983; Ekman, Friesen, & Ellsworth, 1972; Posner, Russell, & Peterson, 2005). In our view this is only the starting point for developing a new way for applying computational modeling techniques in medical diagnosis procedure. This is explained by the fact that while in the above mentioned rule-based

systems deterministic models are considered (e.g. there are facts determined by certain factors), our approach deals with dependent probabilities of medical symptoms, signs and tests, and therefore we model the diagnosis procedure with artificial intelligence formalisms, such as fuzzy sets.

We assume that diagnosis of patients implies some risks and also some possible human errors or lack of an adequate logistics such as materials, equipment, medical devices, etc. (Pearl, 1988; Perner, 2006, 2008).

Basically, the doctors' expertise based on clinical symptoms, physical examination and laboratory results are the foundation for a correct diagnosis and treatment procedure for the patients. In order to optimize this complex process and to help performing correct computational modeling diagnosis our paper focuses on fuzzy built-in diagnosis procedure (Picard, Vyzas, & Healy, 2001; Groshan, 2012). An advantage of our model is that the use of large fuzzy models is not required. Another advantage is that it allows performing sensitivity analysis of an entire flexible system, as well as of its components. The novelty of this approach is that it incorporates the availability of the human factor, as fuzzy approach usually does, and in order to deal with this issue we built a fuzzy filter capable to reject both human and machine errors in medical data transmission and processing, e.g. capable to reject errors of patient diagnosis.

We consider for this approach fuzzy set as a formalism of analysis of biological signals processing and we propose to be accomplished this goal by developing fuzzy operators for filtering the noise of biological signals measurement. These fillers are capable to reject impulsive noise as well as the uniform noise in a very large proportion. The design of such filters is simple, starting by choosing the number of neighbors used for the correction, and obtaining an, easy to implement inference approach (Gadoras & Mihailev, 2009; Sharifalundian, 2009).

We exemplify this approach on neurological measurements performed with an Electro-Encephalograph (EEG) device.

PROPOSED FUZZY FILTER

In this section we present the design of fuzzy, systems devoted to the filtering of noisy measurement data. For this purpose, we shall focus on an interesting, and unfortunately not so much implemented, developments of a family of rule-based fuzzy operators, called FIRE (Fuzzy Inference Ruled by Else-Action) operators (Russo & Ramponi, 1994). The basic principle upon which these operators are based has been originally proposed in (Russo, 1994); now, FIRE operators represent a rapidly growing family of operators for 1-D and 2-D signal processing, as demonstrated by the number of applications in which these operators have become competitive with the classical techniques.

In the next section it will be presented the design of a new class of FIRE operators, e.g. Basic Support of Fuzzy Filters (BSFF) operators as we entitled them, whose filtering behavior can be adapted in order to address uniformly impulsive and distributed noise. Such a filter is based on the following features:

- An optimal choice of fuzzy sets;
- A new interference mechanism.

Let $d(n)$ be a digitized input signal in the range $(0, D-1]$. Let $d_0 = d(n)$ be the sample to be processed at time n , and let $M = \{d_i\}$ be the set of K neighbouring samples which belong to a window centred on d_0 , respectively $M = \{d_1, d_2, \dots, d_K\} = \{d(n-K/2), \dots, d(n-1), d(n+1), \dots, d(n+K/2)\}$, without d_0 .

The input signals of the filter are defined as the amplitude differences as follows:

$$a_{in} = d_i - d_0 \quad (1)$$

Where $i \in (1, K)$, and $a_{in} \in [-M+1, M-1]$

The output parameter a_{adj} represents the adjusting term added to sample d_0 for obtaining the new sample d_{01} :

$$d_{01} = d_0 + a_{adj} \quad (2)$$

Fuzzy sets for the input and output variables are defined in the domain of the amplitude differences. For this purpose, we use triangular-shaped fuzzy sets defined by two parameters (see Figure 1); where R and b represent the center and the half-width of the fuzzy set, respectively. We adopt, in particular, three fuzzy sets labelled *positive strict* (PS), *zero* (Z), and *negative strict* (NS), (John & Innocent, 2005).

The typical rule base of the fuzzy filter includes two symmetrical sub-rule bases and one ELSE-rule, thus identifying three different consequent actions.

Basically, the operation of the fuzzy rule base can be described by the following statements:

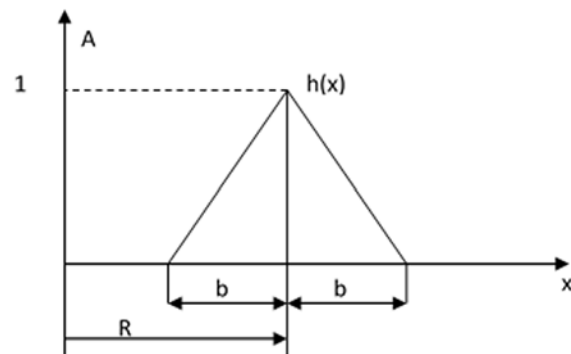
IF the value of a sample is higher than those of its neighbors. THEN decrease its amplitude;

IF the value of a sample is lower than those of its neighbors. THEN increase its amplitude;

ELSE do not change it.

More formally, the rule base of the fuzzy filter is expressed as follows:

Figure 1. Triangular fuzzy set used for building BSFF



IF (a_{in1}, A_{11}) AND... AND (a_{inK}, A_{1K}) THEN (a_{adj}, PS)

IF (a_{in1}, A_{21}) AND... AND (a_{inK}, A_{2K}) THEN (a_{adj}, PS)

IF (a_{in1}, A_{n1}) AND... AND (a_{inK}, A_{nK}) THEN (a_{adj}, PS) (3)

IF (a_{in1}, A_{11}) AND... AND (a_{inK}, A_{1K}) THEN (a_{adj}, NS)

IF (a_{in1}, A_{21}) AND... AND (a_{inK}, A_{2K}) THEN (a_{adj}, NS)

IF (a_{in1}, A_{n1}) AND... AND (a_{inK}, A_{nK}) THEN (a_{adj}, NS)

ELSE (a_{adj}, Z) .

Where A_{ij} is the fuzzy set (PS or NS) associated to the input variable number j in the rule number i . Rules having PS as the consequent set constitute the *first* sub-rule base; rules having NS as the consequent set constitute the *second* sub-rule base.

Fuzzy rules are designed in order to suppress noise while preserving the signal (Ndiaye, 2009); therefore the group of antecedent clauses of each rule takes care of a particular class of neighboring samples. Since the sub-rule bases in relations (3) are symmetrical, we can completely identify the structure of a particular operator by specifying only the group of fuzzy sets A_{ij} ($i = 1, \dots, n$, and $j = 1, \dots, K$). For this purpose, let A be the $(n \times K)$ matrix where elements a_{ij} are defined as follows: $a_{ij} = 0$ if $A_{ij} = PS$, and $a_{ij} = 1$ if $A_{ij} = NS$.

Using this matrix, we shall compactly define the rule base of any BSFF operator.

FUZZY INFERENCE MECHANISM

In order to describe the particular inference mechanism of the filter, we denote by C_1 and C_2

the degrees of activation of the first and the second sub-rule base, respectively. These degrees are evaluated by means of the following relationships (Gullapalli, 1990; Clouse, 1996; Guerrin, 1991), see Figure 1:

$$C_{1,2} = \max\{\min[m_{A_{ij}}(a_{inj})], \text{ with } i = 1, \dots, n, \text{ and } j = 1, \dots, K\} \quad (4)$$

Let C_0 denote the degree of activation of the ELSE-rule, where this degree is evaluated by using the following relation:

$$C_0 = 1 - (C_1 + C_2) \quad (5)$$

It should be observed that the ELSE-rule is reinforced by a weak activation of the other sub-rule bases. By adopting correlation-product inference (Guerrin, 1995), the output a_{adj} can be evaluated by using the following relationship, see Figure 1, and Figure 2:

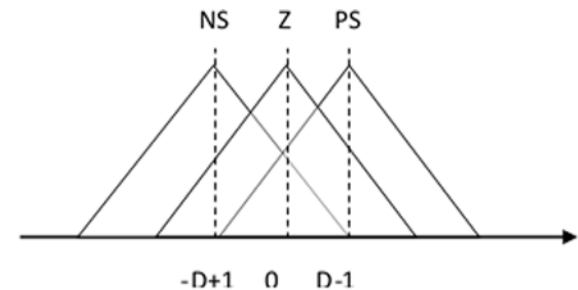
$$a_{adj} = (R_{PS} \cdot b_{PS} \cdot C_1 + R_Z \cdot b_Z \cdot C_0 + R_{NS} \cdot b_{NS} \cdot C_2) / (b_{PS} \cdot C_1 + b_Z \cdot C_0 + b_{NS} \cdot C_2) \quad (6)$$

From relation (1) and considering fuzzy triangular-shaped represented in Figure 1, and the fuzzy sets defined in Figure 2, we have:

$$R_{PS} = -R_{NS} = M-1; R_Z = 0 \quad (7)$$

$$b_{PS} = b_{NS} = b_Z = 2(M-1) \quad (8)$$

Figure 2. Fuzzy sets of BSFF



Thus, relation (6) becomes

$$a_{adj} = (M-1) \cdot (C_1 - C_2) \quad (9)$$

We mention that the fuzzy set parameters given by relations (7) and (8) represents the *optimal* parameters as this parameters allow a noise pulse to be rejected once it has been detected.

ILLUSTRATIVE EXAMPLE

Let V be the set of two neighboring samples defined by $V = \{d_{01}, d_{02} = \{d(n-1), d(n+1)\}$, where $K = 2$. The input variables a_{in1} and a_{in2} are consequently defined by relation (1). We can define a new fuzzy capable to suppress impulsive noise by using six THEN-rules.

$$\begin{aligned} &\text{IF } (a_{in1}, PS) \text{ AND } (a_{in2}, PS) \text{ THEN } (a_{adj}, PS) \\ &\text{IF } (a_{in1}, PS) \text{ AND } (a_{in2}, NS) \text{ THEN } (a_{adj}, PS) \\ &\text{IF } (a_{in1}, NS) \text{ AND } (a_{in2}, PS) \text{ THEN } (a_{adj}, PS) \end{aligned} \quad (10)$$

$$\text{IF } (a_{in1}, NS) \text{ AND } (a_{in2}, NS) \text{ THEN } (a_{adj}, NS)$$

$$\text{IF } (a_{in1}, NS) \text{ AND } (a_{in2}, PS) \text{ THEN } (a_{adj}, NS)$$

$$\text{IF } (a_{in1}, PS) \text{ AND } (a_{in2}, NS) \text{ THEN } (a_{adj}, NS)$$

$$\text{ELSE } (a_{adj}, Z).$$

The corresponding matrix A_{adj} is represented below:

$$A_{adj} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} \quad (11)$$

For exemplifying our approach we addressed the curtesy of the Medical Center Areni Suceava, Romania, which allow us to assist some EEG signals recording. By medical ethic reasons we erased from the following figures the patient name; Figure 3 displays the EEG measured from patient X, (Gadoras & Mihailkov, 2009; Sharifalundian, 2009; Perner, 2008).

Figure 3. EEG signals measured from patient X

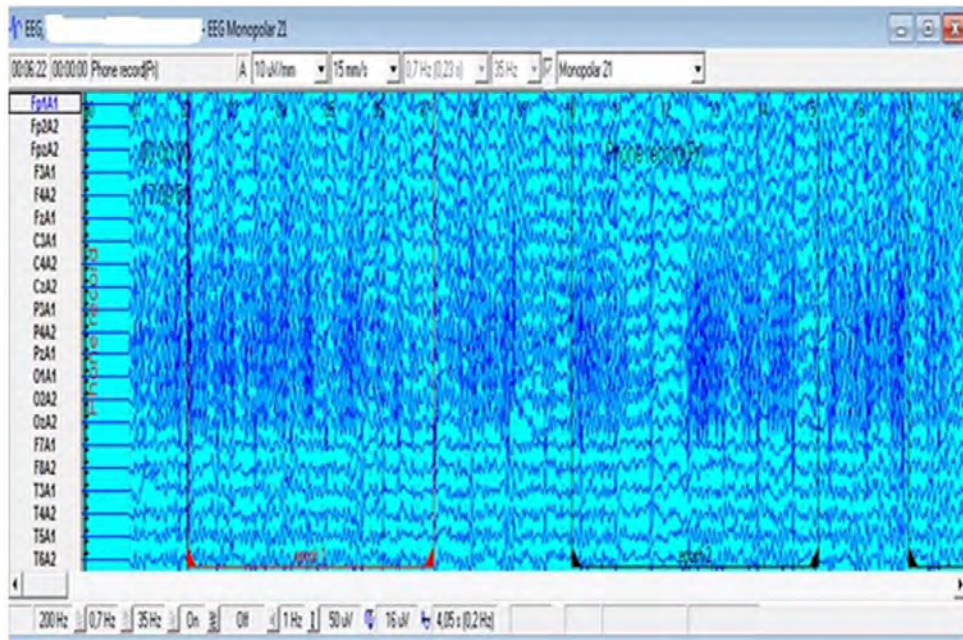
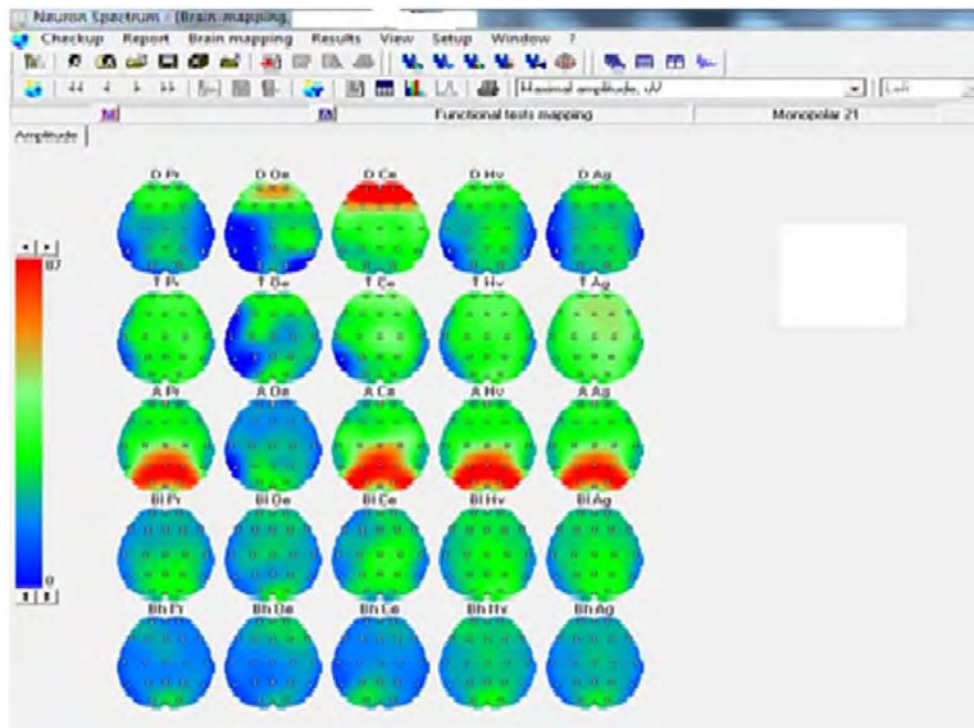


Figure 4. The active area of patient's brain measured by EEG signals



The chart showing the active area of EEG signals measured during the medical investigation time, and processed with a classic tool used by neurologist doctors, e.g. Neuro-Soft is shown in Figure 4. Light colors show the most active zone of the patient's brain.

In order to show the performance of the proposed operator, we have adopted a test waveform of EEG signals measured in an inadequate relational data transmission between the EEG and the computer used for register the EEG's waves and to process them according to our approach.

Basically, we used a non-insulated cable for data transmission which eventually results in intolerably fluctuating transmission path impedances which corresponds to adding impulsive noise (Perner, 2006, 2008).

We processed the biological EEG signals first with a classic tool used by neurologists, e.g. Neuro-Soft, and then we exported them in ASCII format in a Matlab environment in order to apply our approach (Figure 5).

In Figure 6 we display one of the 21st signals measured by EEG complex shown, i.e. signal Fp1A1 in Figure 3, which has been first processed on a time window in Matlab environment; here the signal is severely distorted as we mentioned before. Applying relations (10) and (11), e.g. by applying our fuzzy operator for the above mentioned signal we obtain the signal shown in Figure 7; as it can be seen, the noise has been severely diminished by almost 60%.

We repeat the procedure for all 21 EEG signals displayed in Figure 3 in order to filter the EEG signals, and finally we obtain the right power spectrum of each component of the filtered signals in Figure 8, i.e. the power spectrum of patient X's EEG.

CONCLUSION

Automatic processing of sensory information is an actual research area, and fuzzy techniques

Figure 5. Exporting EEG signals from Neuro Soft to Matlab

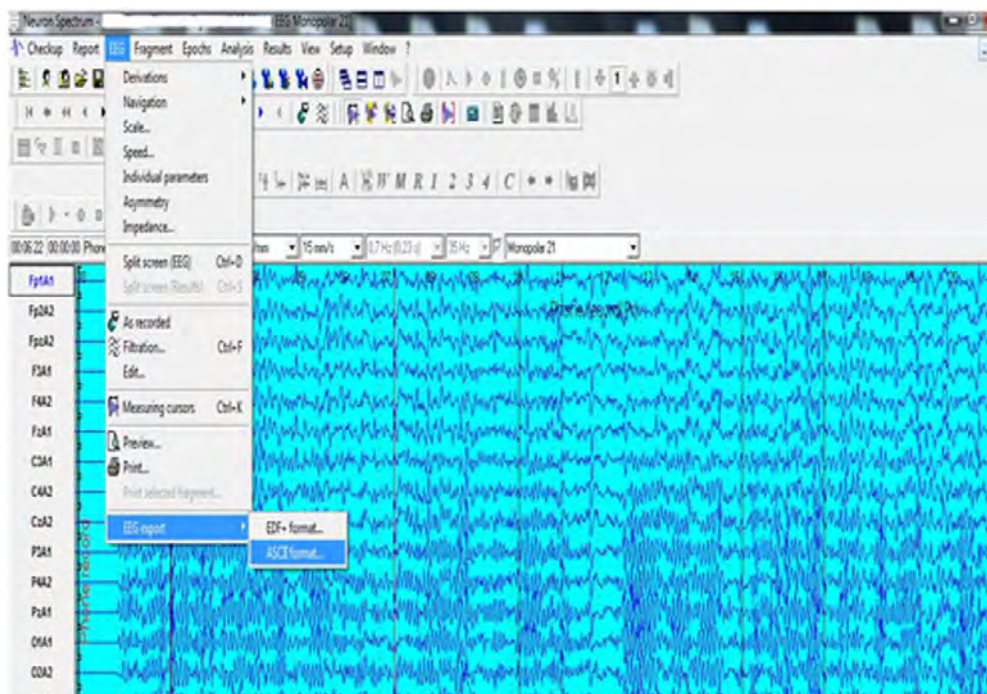


Figure 6. Signal Fp1A1 in processed in a Matlab window and severely distorted

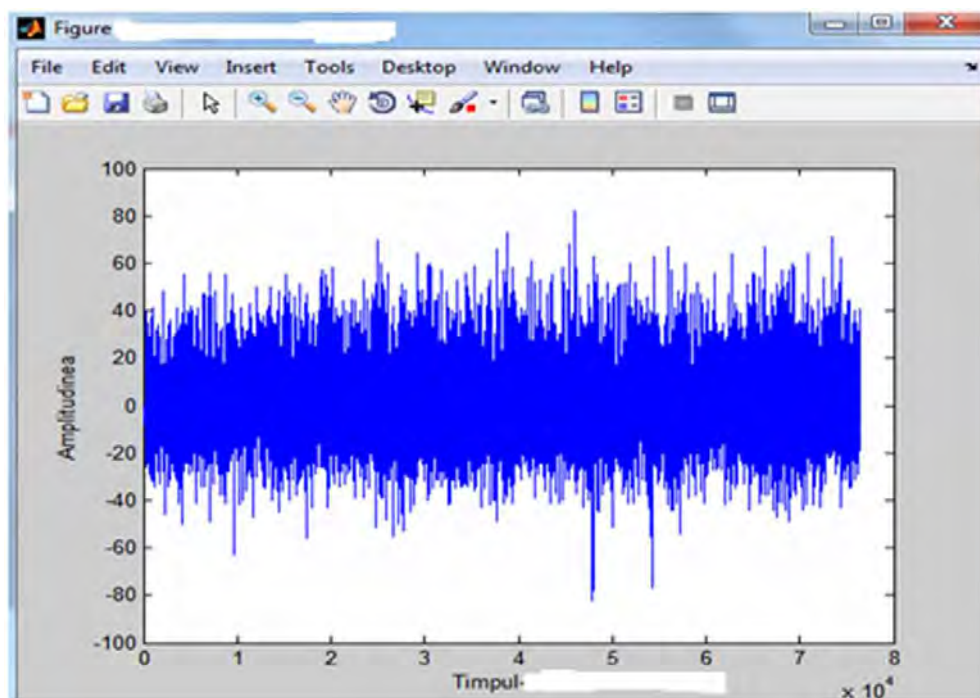


Figure 7. Signal Fp1A1 filtered according to our fuzzy approach

D

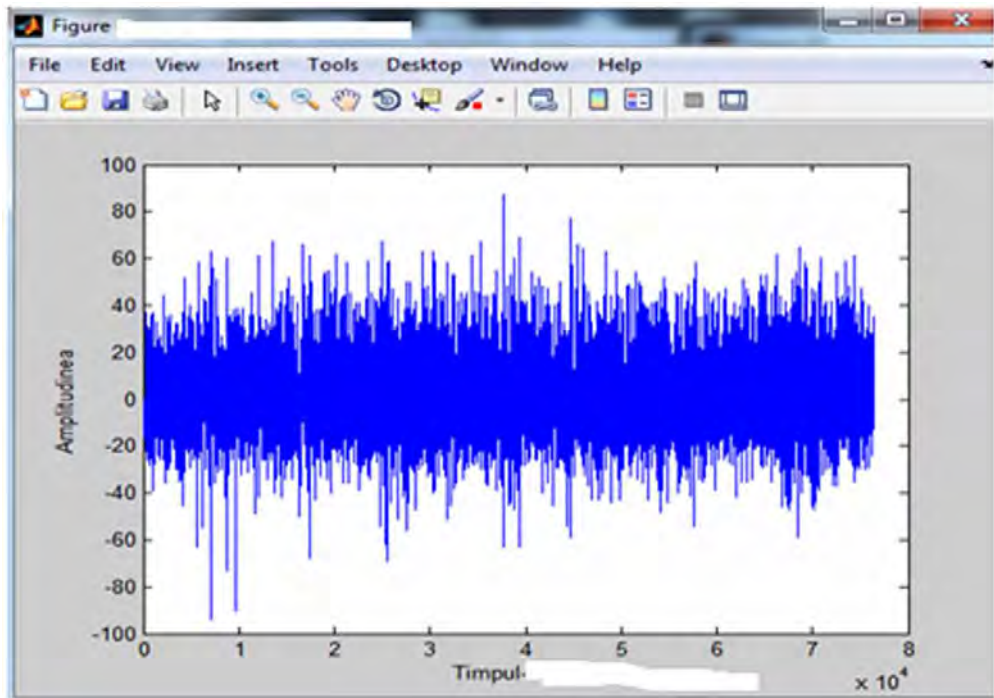
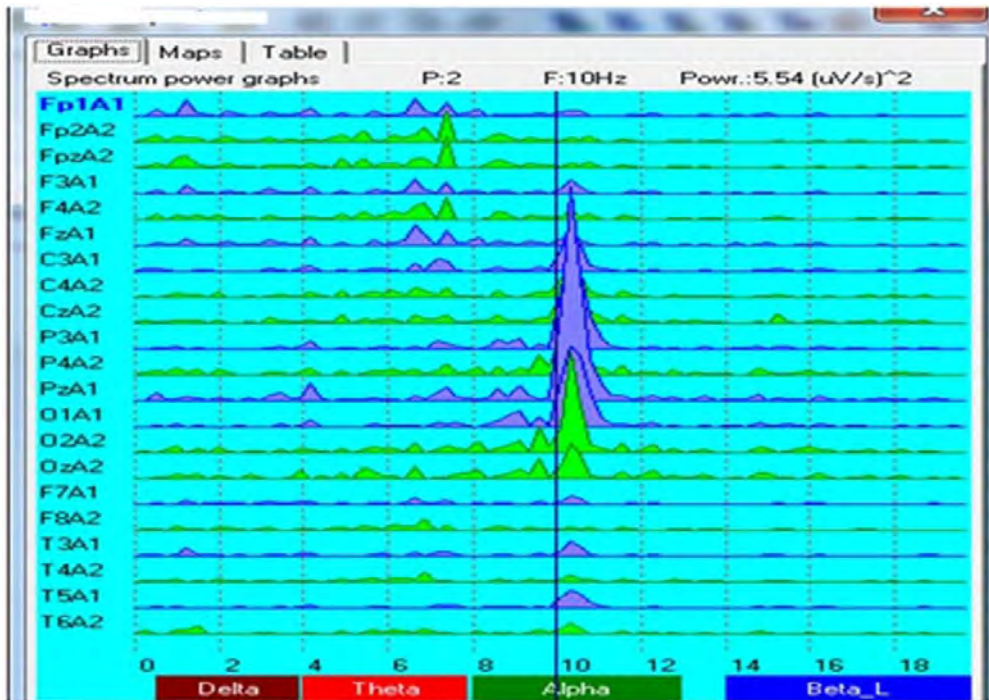


Figure 8. Power spectrum of the patient X's EEG



are well framed in this content, and represents a useful and widely applicable technique with long run perspective.

This chapter is focused on implementing rule-based fuzzy systems for processing, e.g. filtering measured biologic data. Our work resumes the researches of (Russo, 1994) focused on the filtering of noisy signals, and we propose an approach for filtering electrical signals encountered in biological measurement here entitled Basic Support of Fuzzy Filter (BSFF), as a new class of FIRE (Fuzzy Inference Ruled by Else-Action) operators. These fillers are capable to reject impulsive noise as well as the uniform noise in a very large proportion. The design of such filters is simple, starting by choosing the number of neighbors used for the correction, and obtaining an, easy to implement inference approach (Gadoras & Mihailkov, 2009; Sharifalundian, 2009).

One drawback of our method is that it necessitates more experiments in order to find out more appropriate certainty factors for a large range of diseases. However, as far as we experienced our approach under physicians' supervision at the Medical Center Areni Suceava, it proves to be a realistic way to improve the medical interaction physician (or nurse) – patient, as it allows us to modify in real time certainty factors.

FUTURE RESEARCH DIRECTIONS

An interesting extension, to be considered in future work, would consist of introducing a feedback mechanism that promotes the dynamic updating of the certainty factors considered as fuzzy weights (associated with different medical rules) according to the current patient status.

REFERENCES

- Adlassnig, K. P. (2007). Fuzzy Set Theory in Medical Diagnosis, *Systems, Man and Cybernetics. IEEE Transactions on*, 16(2), 260–275.
- Bounds, D. G., Lloyd, P. J., & Mathew, B. G. (1999). A comparison of neural network and other pattern recognition approaches to the diagnosis of low back disorders. *Neural Networks*, 3(5), 583–591. doi:10.1016/0893-6080(90)90008-9
- Cacioppo, J. T., & Tassinari, L. G. (1990). Inferring psychological significance from physiological signals. *The American Psychologist*, 45(1), 16–28. doi:10.1037/0003-066X.45.1.16 PMID:2297166
- Chin, T. L., & Lee, C. S. G. (1991). Neural-network-based fuzzy logic control and decision system, *Computers. IEEE Transactions on*, 40(12), 1320–1336.
- Clouse, J. A. (1996). *On integrating apprentice learning and reinforcement learning* (Unpublished doctoral dissertation). University of Massachusetts, Amherst, MA.
- Ekman, P., Friesen, W. V., & Ellsworth, P. (1972). *Emotion in the human face: guide-lines for research and an integration of findings*. New York: Pergamon Press.
- Ekman, P., Levenson, R. W., & Friesen, W. V. (1983). Autonomic nervous system activity distinguishes among emotions. *Science*, 221(4616), 1208–1210. doi:10.1126/science.6612338 PMID:6612338
- Fathi-Torbaghan, M., & Meyer, D. (1994). MEDUSA: A fuzzy expert system for medical diagnosis of acute abdominal pain. *Methods of Information in Medicine*, 33(5), 522–529. PMID:7869951
- Gadoras, I., & Mihailkov, L. (2009). An interpretable fuzzy rule-based classification methodology for medical diagnosis. *Artificial Intelligence in Medicine*, 47(1), 35–49. PMID:19112012
- Groshan, F. (2012). *Mental and emotional problems seen in a dual diagnosis*. Retrieved from <http://www.articlesbase.com/health-articles/mental-and-emotional-problems-seen-in-a-dual-diagnosis-1385569.html>

- Guerrin, F. (1991). Qualitative reasoning about an ecological process: Interpretative on hydro ecology. *Ecological Modelling*, 59(7), 165–201. doi:10.1016/0304-3800(91)90177-3
- Guerrin, F. (1995). Dualistic algebra for qualitative analysis, Qualitative reasoning workshop. Academic Press.
- Gullapalli, V. (1990). A stochastic reinforcement learning algorithm for learning real-valued function. *Neural Networks*, 3(6), 671–692. doi:10.1016/0893-6080(90)90056-Q
- John, R. I., & Innocent, P. R. (2005). Modeling uncertainty in clinical diagnosis using fuzzy logic. *Systems, Man, and Cybernetics, Part B: Cybernetics. IEEE Transactions on*, 35(6), 1340–1350.
- Kumar, S. (2001). An application of intuitionistic fuzzy sets in medical diagnosis. *Fuzzy Sets and Systems*, 117(2), 209–213. doi:10.1016/S0165-0114(98)00235-8
- Lyons, M. J., Budynek, J., & Akamatsu, S. (1999). Automatic classification of single facial images. *IEEE Tans. On Pattern Analysis and Machine Intelligence*, 21(12), 112–123.
- Ndiaye, A., Valle, G. D., & Roussel, P. (2009). Qualitative modelling of a multi-step process: The case of French bread making. *Expert Systems with Applications*, 36(2), 1020–1038. doi:10.1016/j.eswa.2007.11.006
- Nguyen, H. P., & Kreinovich, V. (2001). Fuzzy logic and its applications in medicine. *International Journal of Medical Informatics*, 62(2–3), 165–173. PMID:11470619
- Pearl, J. (1988). *Probabilistic reasoning in intelligent systems*. Morgan Kaufmann.
- Perner, P. (2002). *Data Mining on Multimedia Data*, LNCS 2558. Springer-Verlag Berlin Heidelberg New York.
- Perner, P. (2006). Case-base maintenance by conceptual clustering of graphs, *Engineering Applications of Artificial Intelligence*, 19(4), 381–395.
- Perner, P. (2008). Prototype-Based Classification. *Applied Intelligence*, 28(3), 238–246. doi:10.1007/s10489-007-0064-0
- Perner, P., & Jänichen, S. (2009). Learning of shape models from exemplars of biological objects in images. *Trans. On Mass-Data Analysis of Images and Signals*, 1(2), 198–209.
- Picard, R. W., Vyzas, E., & Healy, J. (2001). Toward machine emotional intelligence: analysis of affective physiological state. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 23(10), 1175–1191.
- Posner, J., Russell, J. A., & Peterson, B. S. (2005). The circumflex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology*, (17): 715–734. PMID:16262989
- Russo, F. (1996). Fuzzy Systems in Instrumentation: Fuzzy Signal Processing. *IEEE Transactions on Instrumentation and Measurement*, 45(2), 683–689. doi:10.1109/19.492809
- Russo, F., & Ramponi, G. (1994). A Totally Fuzzy Approach to Multi-sensor Instrumentation Processing Techniques. *Proc, IEEE Instrum. Measurem. Tech. Conf. IMTC'94*, 1, 121–132.
- Sharifalundian, E. (2009). Compression of vital bio-signals of patient by enhanced set partitioning in hierarchical trees algorithm. *World Congress on Medical Physics and Medical Bioengineering*, 1, 223–234.
- Zadeh, L. A. (1988). Fuzzy Logic. *Computer*, 1(4), 83–93. doi:10.1109/2.53

ADDITIONAL READING

Abbod, M. F., von Keyserlingk, D. G., Linkens, D. A., & Mahfouf, M. (2001). Survey of utilisation of fuzzy technology in Medicine and Healthcare. *Fuzzy Sets and Systems*, 120(2), 331–349. doi:10.1016/S0165-0114(99)00148-7

Belacel, N., & Boulassel, M. R. (2004). Multicriteria fuzzy classification procedure PROCFTN: Methodology and medical application. *Fuzzy Sets and Systems*, 141(2), 203–217. doi:10.1016/S0165-0114(03)00022-8

De, S. K., Biswas, R., & Roy, A. R. (2001). An application of intuitionistic fuzzy sets in medical diagnosis. *Fuzzy Sets and Systems*, 117(2), 209–213. doi:10.1016/S0165-0114(98)00235-8

Maji, P. K., Biswas, R., & Roy, A. R. (2001). Fuzzy soft set. *J. Fuzzy Math.*, 9(3), 677–692.

Massad, E., Ortega, N. R. S., Struchiner, C. J., & Burattini, M. N. (2003). Fuzzy epidemics. *Artificial Intelligence in Medicine*, 29(3), 241–259. doi:10.1016/S0933-3657(02)00070-2 PMID:14656489

Meenakshi, A. R., & Kaliraja, M. (2011). An application of interval valued fuzzy matrices in medical diagnosis. *Int. J. Math. Anal.*, 5(36), 1791–1802.

Neog, T. J., & Sut, D. K. (2011). Theory of fuzzy soft sets from a new perspective. *Int. J. Latest Trends Comput.*, 2(3), 439–450.

Yıldırım Çelik, Y., & Yamak, S. (2013). Fuzzy soft set theory applied to medical diagnosis using fuzzy arithmetic operations. *Journal of Inequalities and Applications*, 82, 1602–1701.

KEY TERMS AND DEFINITIONS

Automatic Processes: A category of cognitive processing.

Availability: The degree to which a system, subsystem is in a specified operable and committable state at the start of a mission, when the mission is called for at an unknown, *i.e.* a random, time.

Data Analysis: A process of inspecting, adapting, and modeling data with the goal of discovering useful information and decision making.

Diagnosis: Identification of the nature and cause of anything.

Electroencephalography: (EEG): A monitoring medical procedure to record electrical activity of the brain.

Fuzzy Logic: A mathematic analysis formalism which deal with concepts that cannot be expressed as the “true” or “false” but rather as “partially true” similar to human operators thinking.

Inference Mechanism: A tool from artificial intelligence analysis formalisms which addresses the search and control techniques.

Neurologist: A medical specialist in the nervous system.

Learning From Imbalanced Data

D**Lincy Mathews***M. S. Ramaiah Institute of Technology, India***Seetha Hari***Vellore Institute of Technology, India*

INTRODUCTION

Pattern Identification on various domains have become one of the most researched fields. Accuracy of all traditional and standard classifiers is highly proportional to the completeness or quality of the training data. Completeness is bound by various parameters such as noise, highly representative samples of the real world population, availability of training data, dimensionality etc.

Another very pressing and domineering issue identified in real world data sets is that the data is well-dominated by typical occurring examples but with only a few rare or unusual occurrences. This distribution among classes make the real world data inherently imbalanced in many domains like medicine, finance, marketing, web, fault detection, anomaly detection etc.

This chapter aims to highlight the existence of imbalance in all real world data and the need to focus on the inherent characteristics present in imbalanced data that can degrade the performance of classifiers. It provides an overview of the existing effective methods and solutions implemented towards the significant problems of imbalanced data for improvement in the performance of standard classifiers. Efficient metrics for evaluating the performance of imbalanced learning models followed by future directions for research is been highlighted.

BACKGROUND

The field of data mining has identified learning from data that suffer from imbalance distribution

as one of the top problems of today (Yang & Wu, 2006). However, a widely accepted issue is that the traditional learning algorithms assume a balanced distribution among the classes. It does not address nor recognize the presence of imbalance in the data. Data imbalance is evident when the number of instances representing the class of concern is much lesser than other classes.

To cite an example, the 1999 KDD Cup data set (UCI machine repository) is considered. The information collected by a simulated LAN environment consists of normal traffic with a relatively small number of intrusion attempts. The original data set consists of 23 classes, of which one of the classes belonged to normal traffic. When the data set was grouped down to a total of 2 classes, 'normal' and 'attack', the KDD data had 972,781 minority 'attack' class examples and 3,925,650 majority 'normal' class examples, which is approximately 80.14% majority examples. The training data thus will have only very few samples from the 'attack' class, due to which the classifier will be biased to the normal cases. This under representation of the interested class is evident in many applications such as intrusion detection, pollution detection, fault monitoring, biomedical, bioinformatics and remote sensing.

The under-represented class and well-represented class are known as the positive class (denoted by +1) and negative class (denoted by -1) respectively. As the class of interest indicates a positive case and is rarer by nature, it represents the minority data. The research community addresses the other well-defined classes as majority

class. The ratio between the instances of majority versus minority is termed as imbalance ratio.

The skew distribution present in the training data, leads to the bias by most classifiers. Studies have however shown that the base classifiers perform well when presented with balanced data than with imbalanced data (Weiss & Provost, 2001). This justifies the need for learning models that can address the challenges posed by imbalanced data.

CHARACTERISTICS OF IMBALANCED DATA

The imbalance ratio between the majority and minority instances need not necessarily affect the performance of classifiers if the degree of imbalance is moderate (Chen & Wasikowski, 2008). The inherent characteristics within minority data however; can cause degrade in performance by the learning models. Two basic categorization of minority instances exist; Safe and unsafe minority instances. Safe minority instances are instances where the misclassification is minimal by the base learners. These instances exist much away from the borderline of majority instances. Unsafe minority instances are so called because the misclassifications occur highly with these kinds of minority instances.

The causes of unsafe instances in imbalanced data sets are contributed by four significant occurrences. They are the presence of small disjuncts, lack of density in the sample space, noisy data and data shift. Addressing these issues alone can sometimes bring a positive effect on the accuracy of the classifier without having to address the imbalance factor (Japkowicz N, 2003).

Small disjuncts exist when there is a small cluster of similar instances amidst cluster of majority or minority instances. However, in case of imbalanced datasets, the presence of sub concepts (disjuncts) in the majority class will be rare as they are represented well. The occurrence of small disjuncts is frequent in minority class. The presence of small sub concepts can undermine the

performance of the classifier. As the data space; is very limited, this can lead to generalization and over fitting by the learning models. Weiss (2003) had identified the presence of disjuncts in imbalanced datasets as noise or sub concepts. Synthetic (artificial) samples were generated for these small disjuncts to overcome the lack of representation.

Lack of density is the existence of small sample size. Lack of samples coupled with high dimensional data becomes an even more difficult issue in imbalanced data. This creates rules that are too specific and can lead to over fitting. In domains such as face recognition, gene expression analysis etc; feature reduction method such as the Principal Component Analysis were used to address the high dimensional problem with respect to sample size (Yang et al., 2008). One sided or two sided feature selection technique addressed this problem (Alibeigi, 2012).

Presence of noise or outliers creates a negative impact on the performance of the classifier. As outliers occur much away from minority instances, it is most likely treated as noise. However, it might represent the rarest of cases that should be recognized and classified. There are cases of overlapping instances; presence of similar instances in majority and minority instances. These samples seemingly are the hardest to classify. The authors (Garcia et al., 2008) proposed two different frameworks for the k-Nearest Neighbor (k-NN) classifier. They focused on the ratio between the imbalance in the overlapping region versus the overall imbalance ratio. Khoshgoftaar et al. (2011) analyzed the relation between noise and imbalanced data. Bagging and boosting techniques were applied after implementing noise reduction techniques.

The data shift problem is realized when the training data distributions differs from the distribution of the test data (Alaiz & Japkowicz, 2008). The data shift phenomenon is mostly due to sample selection process. Stratified cross validation is used to measure performance thus avoiding sudden drift in performance. The other cause could be due to a very high degree of imbalance. Presence of data shift can highly affect the singular classification

error due to the low number of instances it has (Moreno et al., 2010). Preprocessing techniques can address the data shift problem.

TECHNIQUES FOR LEARNING FROM IMBALANCED DATA

The solutions proposed are categorized into data driven techniques, cost sensitive techniques and algorithmic solutions.

Data Driven Techniques

The aim is to change the distribution of data. Some mechanisms are applied to alter the imbalanced distribution of the data to a balanced distribution. One of the effective and efficient methods under data driven techniques is the sampling method.

Random sampling involves sampling of data to alter the skew distribution of existing data. Random under sampling involves random elimination of majority instances. This technique however, might eliminate informative samples that can affect the classifier performance while reducing the size of majority class. Random over sampling, appends instances to the minority class by sampling from the minority data. This naturally leads to over fitting and cause computational burden on the algorithm.

To overcome the deficiency of random under sampling, Informed under sampling selects majority samples that are redundant or acted as noise in classifying minority instances. Yen et al. (2009) proposed a cluster based approach; by clustering the majority instances from where the representative samples were then selected from each cluster as instances. Yua et al. (2013) selected significant and informative majority instances as per the selection frequency. This process however proved to be time consuming than the simple sampling procedure.

Synthetic Sampling, an over sampling technique was one of the powerful technique that proved to be a success in many applications. This

technique initially was adopted to address the high dimensionality problem. Synthetic Minority Oversampling Technique (SMOTE) (Chawla, 2002) generated instances by creating samples between nearest neighbor minority samples using the distance measure. The synthetic sample is defined as a point on the line joining a minority instance and its nearest neighbor; i.e., where, r is a random number between 0 and 1, x_i is the selected minority instance and, x_{nn} is nearest neighbor to. Synthetic instances generated closer to the majority class could be noise instances for the majority samples and can lead to overlapping.

Seetha et al. (2012) developed a novel multiple kernel learning approach to generate a synthetic training set larger than the original training set. This addressed the underperformance of Support Vector Machine (SVM) when the training sample had a very high dimension data. Viswanath et al. (2006) proposed a novel synthetic pattern generation by performing the join of sub patterns. The curse of dimensionality that affected the performance of Nearest Neighbor classifier was reduced significantly. Synthetic patterns to address the imbalance problem using the techniques mentioned in paper are currently being studied by the authors of the chapter.

Adaptive sampling techniques were proposed to overcome the overlapping limitation. Synthetic samples are generated based on the category of minority instances. Bartosz et al. (2014), proposed weighted one class classifier and adjusted weights depending on the type of minority instances - safe, borderline, rare and outlier. Random oversampling and SMOTE were used to improve the distribution of the data for increase in performance by decision tree, neural networks and SVM (Kerdprasop and Kerdprasop, 2012).

Cluster Based Clustering (CBC) approach deals with within-class imbalance problem. Yong (2012) proposed an approach where synthetic samples are produced using the genetic approach with the k- means classifier. Yen & Lee (2009) formed clusters of training data (majority class). The representative samples from the clusters were

selected as majority samples. Results showed that the cluster based under sampling improved prediction accuracy. Vasu & Ravi (2011) combined k means clustering on the negative class with k-reverse nearest Neighbor (kRNN) to detect outliers. The superiority of the method were effectively proved by classifiers such as SVM, Genetic programming and decision tree.

Another technique used is the Sampling with boosting method. Thanathamath et al. (2013) combined boundary data generation with boosting procedures for efficient classification by the Adaboost Classifier. Using the Hausdorff distance, the class boundary data were identified and eliminated. The training data was augmented by bootstrapping and synthesizing the new boundary area. Myoung et al. (2015) used geometric mean based boosting algorithm (GMBoost) to resolve data imbalance problem. The proposed method showed better performance than adaboost and cost sensitive boosting.

Features or dimensionality of the data plays an important role for the efficient and unbiased performance of classifiers. One Sided selection identifies features relevant to the class of interest. Two-sided selection identifies relevant features pertaining to both the classes. Feature selection metrics such the Chi square, Information Gain etc handles attributes if binary and nominal. Other metrics used for continuous data are Pearson Correlation Coefficient, Feature Assessment by sliding threshold etc. The probability density estimation of features for small disjuncts but with high dimensional data sets were used as the features selection criteria in Density Based Feature Selection (DBFS) (Alibeigi et al., 2012). DBFS in combination with correlation between features were adopted. K-S tree based on Kolmogorov-Smirnov proposed by Gong et al. (2012) selected relevant features and removed redundant variables. Oversampling and under sampling the data, is then applied to balance the distribution of data.

Cost- Sensitive Based Learning

One of the efficient approaches attempted by the researchers to reduce the bias of traditional learners is the sampling technique based on balance ratio. In Cost Sensitive approach, however, the costs of misclassified instances are considered. When there is a misclassification, the cost of misclassifying a minority instance as majority instance is usually greater than the cost of misclassifying a majority instance. Studies have shown that cost sensitive learning methods are superior to sampling technique in certain domains (Liu & Zhou, 2006) such as fraud, medical diagnosis etc.

Three approaches are adopted under the umbrella of cost sensitive learning. The first method is to assign the cost to the data space based on hard to classify samples, density based, borderline samples etc. Thus, based on the cost, the best instances from the training samples are inducted for next iteration. The second technique is to modify the existing classifier algorithms by incorporating the cost sensitive functions or characteristics. The third class involves developing cost sensitive classifiers by combining ensemble algorithms with the traditional algorithms.

Y Sun et al. (2007) proposed three algorithms: AdaC1, AdaC2 and AdaC3. Each algorithm assigns a greater misclassification cost so that the classifier receives instances that are more relevant. Uyar et al (2006) used oversampling and under-sampling with adjusting of decision threshold. This technique achieved optimum true positive and true negative.

Cost adjustments are applied to the decision threshold, split criteria at each node and on different pruning schemes considering the decision tree classifier. Maloof (2003) applied the different decision threshold criteria with respect to the misclassification cost. ROC evaluation metric was used to arrive at the most dominant point on the Curve. The plot was done to analyze the per-

formance when the decision threshold is moved against the maximum cost earned by misclassification of positive class and negative class. In a standard decision tree, splitting criterion is based on minimum error at each node. Pruning was applied to decision trees to remove leaves with class probabilities less than a threshold. Japkowicz & Stephen (2002) showed pruning the tree modeled from imbalanced data demonstrated less performance when compared to pruning of such trees.

On implementing cost sensitive approaches in Support Vector Machine (SVM); Zhao et al. (2011) worked on weighted maximum margin criterion to optimize the kernel. The weighted margin between the minority and majority class is been maximized in this method. Hwang et al. (2011) embedded weight parameters in the SVM Lagrangian formulation to speed up training and performance of Lagrangian Support Vector machine (WLSVM). By transforming a multiclass imbalance problem into symmetrical two-class problem, Modular Neural Network (MNN) based on Divide and Conquer principle was proposed by Zhao (2009). It was less complex and showed better performance when compared to other Neural Network techniques.

Ensemble Based Methods

Ensemble methods are combinations of individually trained classifiers whose decisions are summed to arrive at the class of a new object. The performance is proportional to the accuracy and diversity of the base classifier. Efficiency is been improved by forming different learners based on different training data. The combination techniques usually are boosting and bagging.

SMOTEBoost (Chawla et al., 2003) algorithm produces synthetic instances on each boosting iteration. Thus, each ensemble gets a different training set that are well defined and separated among the classes. The distribution among classes is improved by adding new minority instances produced by SMOTE. Barandela et al. (2003) proposed ensemble classifier where the training

data is a balanced set of minority and majority instances. The number of majority instances under sampled without replacement is equivalent to the number of minority instances. Hence, every training set will contain the same minority instances but different majority instances.

Gue & Viktor (2004) proposed DataBoost-IM that generated synthetic samples based on the “hard to classify” samples. These instances formed the focus for the next classifier in the boosting procedure. Synthetic samples were used as against random sampling to avoid overfitting. Lia et al. (2012) developed a genetic algorithm that integrated over sampling and under sampling with an ensemble of SVM classifier. Synthetic instances were generated by using random distance to nearest neighbor by which the nearest neighbor support vector machines (NNSSVM) was generated. Boosting algorithm was used to assemble the NNSSVM. Zhang (2012) introduced a dynamic classifier ensemble method (DCEID). This model outperformed other ensemble strategies such as the weighted random forests and is based on the adapting ensemble which selected sample from the Dynamic Classifier selection and the dynamic ensemble selection.

Table 1 highlights the pros and cons of the three proposed methods for learning from imbalanced data.

EVALUATION METRICS

The commonly and traditionally used measure is the Classification accuracy (CA). For observing the bias of imbalanced data, the evaluation metrics for three different data under different learners are shown in Table 2. The numbers of positive instances versus negative instances are indicated against the domain.

As we see, the classification accuracy shows an above average performance against all classifiers for the all data given. This measure however, fails to identify the misclassification rate in their respective class. Gue & Viktor (2004) discusses

Table 1. Advantages and disadvantages of learning techniques

Method	Advantages	Limitations (Disadvantages)
Sampling Techniques	• Implementation is easy. • Data is affected but not the Classifier model	• Oversampling - o Increase in memory space - o Cause over fitting - o Time Consuming - o Quality of synthetic and existing data should be similar - o Computational Burden • Under Sampling - o Removal of informative and significant samples
Cost Sensitive Method	Minimize the bias of the classifier towards majority samples by reducing the cost of misclassification	• Error costs are often unknown and decided based on domain • Can lead to over fitting
Ensemble Methods	• More resistant to noise • Multiple classifiers provide better prediction than single classifier.	• Time consuming • Over fitting can occur

Table 2. Performance measures of various classifiers

Data-Domain with Ratio of Imbalance (+: -)	Classifier	Classification Accuracy	Specificity	Sensitivity	Precision
Liver Patient (167:416)	Nearest Neighbor	74.97	80.00	34.41	41.37
	Random Forest Tree	69.64	85.10	31.14	45.61
	Support vector machine	61.23	74.60	56.46	41.80
	Neural Networks	63.15	63.15	23.52	69.56
Satellite (626:5809)	Nearest Neighbor	91.62	97.06	67.99	71.76
	Random Forest Tree	94.33	98.66	54.15	81.29
	Support Vector machine	94.34	96.04	78.59	68.35
	Neural Networks	93.58	94.92	78.48	57.94
Blood Transfusion (178:570)	Nearest Neighbor	73.65	78.86	34.14	46.97
	Random Forest Tree	74.20	87.72	30.90	44.00
	Support vector machine	68.71	76.32	44.74	22.31
	Neural Networks	78.57	79.43	60.00	12.00

the ineffectiveness of for imbalanced data. The above examples show that base classifiers are sensitive to the distribution of data. In response to the ineffectiveness of the CA, other metrics, insensitive to the distribution of data used by the research communities are namely

, where TP-True Positives, FP-False Positive

, where FN-False Negative

Recall is known as sensitivity/TPRate for a binary-class data.. Further looking into the response values of sensitivity and specificity, as shown in Table 2, specificity values indicates a better prediction of majority samples. However, sensitivity indicates the degree of misclassification of positive instances. The number of patients been classified as affected by liver diseases is only 34% by the nearest neighbor classifier. The remaining were misclassified as non liver disease patients which is a huge error rate. Sensitivity of 60% in-

icates that 40% of patients were misclassified as patients who did not receive the blood transfusion. The table thus shows the necessity of specificity, sensitivity and precision. All the parameters are estimated through a contingency matrix.

, where β is the non-negative coefficient that adjusts the weight age between recall and precision. The value generally assigned is 1. F measure measures the effectiveness of retrieval. Precision is sensitive to data distribution but recall is not. Precision does not indicate the number of positive samples correctly classified.

Higher value of G-mean indicates maximization of accuracy of both classes and at the same time balancing accuracy of the two classes. However, these metrics does not address the performance of different classifiers over the same distribution.

For comparison and evaluating different classifiers, Receiver Operating Characteristics (ROC) curve is used. The ROC graph are plotted against TPrate and FPrate (Fawcett, 2006). This provides the visualization between tradeoffs and costs of classifiers with respect to the data distribution. For classifiers that produce discrete class outputs, the ROC is just a point on the graph. Any point on the left top part of the curve shows a good performance as TPrate should approach to 1 and FPrate to zero. For classifiers that output continuous numeric values, then a threshold produces a series of points.

ROC curves have its limitations. The ROC graph cannot indicate performance of classifiers if the number of false positives increases. There will not be significant change in the FPrate since it represents FP only minimally as it is a combination of FP and TN. Hence, the Precision Recall curve (Davis & Goadrich, 2006) is used. This curve is advantageous to measure the classifiers performance on highly skewed data as it indicates the change in False Positive. The values should approach to right top corner of the PR graph representing high value of precision and recall.

FUTURE CHALLENGES

Works highlighted in the chapter focused on learning from imbalanced data. However, several issues and challenges are yet to be investigated. One of the issues is to find to what degree the imbalance ratio should be reduced. Secondly, no unified framework exists. Thirdly, no specific approach has been singled out for imbalanced data in general. There is a need for standardized curve based evaluation metric to handle imbalance data. Algorithms that can handle incremental learning from imbalanced data streams need focus. Applying semi supervised learning to imbalanced data is also a pressing issue that ought to be addressed. Lastly, identifying unlabelled data using minimal training data also proves to be a challenge.

CONCLUSION

This chapter highlighted a very challenging and predominant issue in the field of machine learning: Learning from imbalanced data. The insights of the chapter aimed to cover the characteristics of imbalanced data that causes learner models to be biased. It discusses the state of the art solutions aimed at improving the classifier performance. It covers several efficient assessment techniques to analyze the performance over a range of situations. The opportunities and challenges available in the area of imbalanced data mining were highlighted for future research and development.

REFERENCES

Alaiz-Rodriguez, R., & Japkowicz, N. (2008). Assessing the impact of changing environments on classifier performance. *Proceedings of the 21st Canadian Conference on Advances in Artificial Intelligence (CCAI'08)*, 13–24 doi:10.1007/978-3-540-68825-9_2

- Alibeigi, M., Hashemi, S., & Hamzeh, A. (2012). DBFS: An effective Density Based Feature Selection scheme for small sample size and high dimensional imbalanced data sets. *Data & Knowledge Engineering*, 81–82, 67–103. doi:10.1016/j.datak.2012.08.001
- Barandela, R., Sánchez, J. S., García, V., & Rangel, E. (2003). Strategies for learning in class imbalance problems. *Pattern Recognition*, 36(3), 849–851. doi:10.1016/S0031-3203(02)00257-1
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-Sampling Technique. *Journal of Artificial Intelligence Research*, 16, 321–357.
- Chawla, N. V., Lazarevic, A., Hall, L. O., & Bowyer, K. W. (2003). SMOTEBoost: Improving Prediction of the Minority Class in Boosting. *Proc. Seventh European Conf. Principles and Practice of Knowledge Discovery in Databases*, 107–119 doi:10.1007/978-3-540-39804-2_12
- Chen, X., & Wasikowski, M. (2009). Fast: A roc-based feature selection metric for small samples and imbalanced data classification problems. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 124–132
- Davis, J., & Goadrich, M. (2006). The Relationship between Precision- Recall and ROC Curves. *Proc. Int’l Conf. Machine Learning*, 233–240 doi:10.1145/1143844.1143874
- Fawcett, T. (2006). An Introduction to ROC Analysis. *Pattern Recognition Letters*, 27(8), 861–874. doi:10.1016/j.patrec.2005.10.010
- Garcia, V., Mollineda, R., & Sanchez, J. S. (2008). On the k-NN performance in a challenging scenario of imbalance and overlapping. *Pattern Analysis & Applications*, 11(3-4), 269–280. doi:10.1007/s10044-007-0087-5
- Geu, H., & Viktor, H. (2004). Learning from Imbalanced Data Sets with Boosting and Generation: The DataBoost-IM Approach. *SIGKDD Explorations*, 6, 30–39. doi:10.1145/1007730.1007736
- Gong, R., & Huang, S. H. (2012). A Kolmogorov–Smirnov statistic based segmentation approach to learning from imbalanced datasets: With application in property refinance prediction. *Expert Systems with Applications*, 39(6), 6192–6200. doi:10.1016/j.eswa.2011.12.011
- Hwang, J. P., Park, S., & Kim, E. (2011). A new weighted approach to imbalanced data classification problem via support vector machine with quadratic cost function. *Expert Systems with Applications*, 38(7), 8580–8585. doi:10.1016/j.eswa.2011.01.061
- Japkowicz, N. (2003). Class Imbalances: Are We Focusing on the Right Issue? *Proc. Int’l Conf. Machine Learning, Workshop Learning from Imbalanced Data Sets II*.
- Japkowicz, N., & Stephen, S. (2002). The Class Imbalance Problem: A Systematic Study. *Intelligent Data Analysis*, 6(5), 429–449.
- Kerdprasop, N., & Kerdprasop, K. (2012). On the Generation of Accurate Predictive Model from Highly Imbalanced Data with Heuristics and replication Technologies. *International Journal of Bio-Science and Bio-Technology*, 4, 49–64.
- Khoshgoftaar, T. M., Van, H., & Napolitano, J. A. (2011). Comparing boosting and bagging techniques with noisy and imbalanced data. *IEEE Transactions on Systems, Man, and Cybernetics. Part A, Systems and Humans*, 41(3), 552–568. doi:10.1109/TSMCA.2010.2084081
- Lia, H., & Suna, J. (2012). Forecasting business failure: The use of nearest-neighbour support vectors and correcting imbalanced samples e-Evidence from the Chinese hotel industry. *Tourism Management*, 33(3), 622–634. doi:10.1016/j.tourman.2011.07.004

- Liu, X. Y., & Zhou, Z. H. (2006). Training Cost-Sensitive Neural Networks with Methods Addressing the Class Imbalance Problem. *IEEE Transactions on Knowledge and Data Engineering*, 18(1), 63–77. doi:10.1109/TKDE.2006.17
- Maloof, M. A. (2003). Learning When Data Sets Are Imbalanced and When Costs Are Unequal and Unknown. *Proc. Int'l Conf. Machine Learning, Workshop Learning from Imbalanced Data Sets II*
- Moreno, T. J. G., & Herrera, F. (2010). A preliminary study on overlapping and data fracture in imbalanced domains by means of genetic programming-based feature extraction. *Proceedings of the 10th International Conference on Intelligent Systems Design and Applications (ISDA'10)*, 501–506.
- Myoung-Jong, K., Dae-Ki, K., & Hong, B. K. (2015). Geometric mean based boosting algorithm with over-sampling to resolve data imbalance problem for bankruptcy prediction. *Expert Systems with Applications*, 42(3), 1074–1082. doi:10.1016/j.eswa.2014.08.025
- Seetha, H., Sarvanan, R., & Narasimha, M.N. (2012). Pattern Synthesis Using Multiple Kernel Learning For Efficient SVM Classification. *Cybernetics and Information Technologies*, 12(4), 77–94.
- Sun, Y., Kamel, M. S., Wong, A. K. C., & Wang, Y. (2007). Cost-Sensitive Boosting for Classification of Imbalanced Data. *Pattern Recognition*, 40(12), 3358–3378. doi:10.1016/j.patcog.2007.04.009
- Thanathamath, P., & Lursinsap, C. (2013). Handling imbalanced data sets with synthetic boundary data generation using bootstrap re-sampling and AdaBoost techniques. *Pattern Recognition Letters*, 34(12), 1339–1347. doi:10.1016/j.patrec.2013.04.019
- Uyar, A., Bener, H. N., Ciracy, & Bahceci, M. (2006) Handling the Imbalance Problem of IVF Implantation Prediction. *IAENG International Journal of Computer Science*.
- Vasu, M., & Ravi, V. (2011). A hybrid under-sampling approach for mining unbalanced datasets: Applications to banking and insurance. *Int J. Data Mining Modelling and Management*, 3(1), 75–105. doi:10.1504/IJDMMM.2011.038812
- Viswanath, P., Narasimha, M. N., & Shalabh, B. (2006). Partition based pattern synthesis technique with efficient algorithms for nearest neighbor classification. *Science Direct. Pattern Recognition Letters*, 27(14), 1714–1724. doi:10.1016/j.patrec.2006.04.015
- Weiss, G. M. (2003). *The effect of small disjuncts and class distribution on decision tree learning* (PhD Thesis). Rutgers University
- Weiss, G. M., & Provost, F. (2001). *The Effect of Class Distribution on Classifier Learning: An Empirical Study. Technical Report MLTR- 43*. Dept. of Computer Science, Rutgers University.
- Yang, Q., & Wu, X. (2006). 10 challenging problems in data mining research. *International Journal of Information Technology & Decision Making*, 5(4), 597–604. doi:10.1142/S0219622006002258
- Yang, W. H., Dai, D. Q., & Yan, H. (2008). Feature Extraction Uncorrelated Discriminant Analysis for High-Dimensional Data. *IEEE Transactions on Knowledge and Data Engineering*, 20(5), 601–614. doi:10.1109/TKDE.2007.190720
- Yen, S., & Lee, Y. (2009). Cluster-based under-sampling approaches for imbalanced data distributions. *Expert Systems with Applications*, 36(3), 5718–5727. doi:10.1016/j.eswa.2008.06.108

Yong, Y. (2012). The Research of Imbalanced Data Set of Sample Sampling Method Based on K-Means Cluster and Genetic Algorithm. *International Conference on Future Electrical Power and Energy Systems. Energy Procedia*, 17, 164–170. doi:10.1016/j.egypro.2012.02.078

Yua, H., Nib, J., & Zhaoc, J. (2013). ACOSampling: An ant colony optimization-based undersampling method for classifying imbalanced DNA microarray data. *Neurocomputing*, 101, 309–318. doi:10.1016/j.neucom.2012.08.018

Zhang, Y., Zhang, D., Mi, G., Ma, D., Li, G., Guo, Y., & Zhu, M. et al. (2012). Using ensemble methods to deal with imbalanced data in predicting protein–protein interactions. *Computational Biology and Chemistry*, 36, 36–41. doi:10.1016/j.compbiolchem.2011.12.003 PMID:22286086

Zhao, Z., Zhong, P., & Zhao, Y. (2011). Learning SVM with weighted maximum margin criterion for classification of imbalanced data. *Mathematical and Computer Modelling*, 54(3-4), 1093–1099. doi:10.1016/j.mcm.2010.11.040

Zhao, Z. Q. (2009). A novel modular neural network for imbalanced classification problems. *Pattern Recognition Letters*, 30(9), 783–788. doi:10.1016/j.patrec.2008.06.002

ADDITIONAL READING

Estabrooks, A., Jo, T., & Japkowicz, N. (2004). A multiple resampling method for learning from imbalanced data sets. *Computational Intelligence*, 20(1), 18–36. doi:10.1111/j.0824-7935.2004.t01-1-00228.x

Fawcett, T. (2003). ROC Graphs: Notes and Practical Considerations for Data Mining Researchers, Technical Report HPL-2003-4, HP Labs

Kubat, M., & Matwin, S. (1997). Addressing the curse of imbalanced training sets: one sided selection. *Proc. 14th Intl. Conf. on Machine Learning*, 179–186.

Laurikkala, J. (2001). Improving Identification of Difficult Small Classes by Balancing Class Distribution. *Proc. Conf. AI in Medicine in Europe: Artificial Intelligence Medicine*, 63-66 doi:10.1007/3-540-48229-6_9

Liu, Y., Lohb, H. T., & Sunc, A. (2009). Imbalanced text classification: A term weighting approach. *Expert Systems with Applications*, 36(1), 690–701. doi:10.1016/j.eswa.2007.10.042

KEY TERMS AND DEFINITIONS

Bagging: It is the process of training multiple models on different samples (data splits) and averaging their predictions.

Bootstrapping: A statistical method generates artificial samples from the original training samples.

Classification: It is a process by which an unseen sample is assigned a class label by a model trained on data of known class labels.

Cost Matrix: A classification cost matrix is a matrix, where the element of value is the misclassification cost of guessing a case belongs to class X, when it actually belongs to class Y.

Overfitting: A modeling error that occurs when a function is too closely fit to a limited set of data point is known as Overfitting.

Sampling: It is a statistical analysis technique used to select, manipulate and analyze a representative set of data points in order to identify patterns in the larger data set.

Synthetic Data: It is defined as artificial generated data that is produced by some method.

Neighborhood Rough-Sets–Based Spatial Data Analytics

D

Sharmila Banu K
VIT University, India

B. K. Tripathy
VIT University, India

INTRODUCTION

With the huge amount of data generated every day, governments, corporates and scientists look at mining useful information from them. Considering the storage and organization costs of these data, the useful trade-off is to discover useful patterns in them. Transactional, telecommunication, spatial, satellite, remote sensing, medical informatics etc., are some of the domains dealing with *Big Data*. Discovering patterns and inferring predictions are vital to the well-being of man-kind at large. Analysis of spatial data usually includes construction of information system, dimensionality reduction, decision rule extraction based on a computational model and error analysis. This chapter focuses on the construction decision systems, proposes a hybrid method by substantiating the advantages and challenges involved with spatial data.

Spatial Data upholds an important perspective that it can be analysed only with respect to specific spatial reference (or a geographic area). Spatial data include spatial attributes like temperature, rainfall, humidity, slope, land cover etc., These features will be with respect to a spatial reference and encompass spatial auto correlation. Spatial auto correlation is the property of spatial data where the values of a feature will tend to be similar in a neighborhood and vary with increasing distance from the neighborhood. Figure 1 illustrates a sample region identified with latitudes and longitudes. A number of spatial features can be measured and some of them are shown in Table 1.

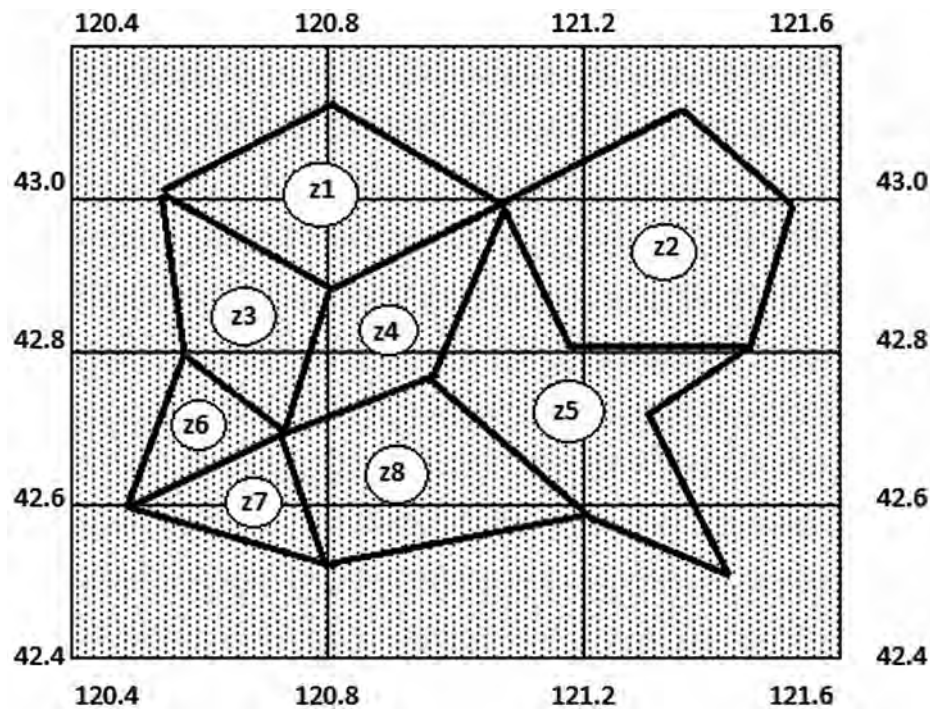
BACKGROUND

Mining spatial data is useful in fields like weather forecasting, natural calamity prediction, crime management, transmission and spread of infectious diseases and others. This calls for expertise in these areas and the nature of spatial data. For example, representing topology in spatial data modelling is inherent to dealing with uncertainties. And, Rough Sets have been used to deal with uncertainty in spatial data mining. Pawlak's (1982) Rough Set Theory (RST) has been used to model spatial regions with unclear boundaries. Beaubouef and Petry (1994) have demonstrated the use of rough sets have been used to query crisp data in relational databases. The Region Connection Calculus (RCC) proposed by Randell & Cohn (1992) and Egg-Yolk models by Cohn and Gotts (1996) have been blended with the ap-

Table 1. Sample Attribute Data of the region in Figure 1

Zones	Rainfall	Temperature	Slope
z1	21	18	1.6
z2	21	19	1.8
z3	22	19	2.1
z4	26	22	2.4
z5	16	28	1.4
z6	17	26	1.3
z7	19	27	1.7
z8	20	27	1.9

Figure 1. A sample geographic region



proximation concepts of RST to identify vague region boundaries. Rough sets have been used by Bai et. al (2010) to identify villages with birth defects, Ahlqvist (2005) for spatial classification and analysis, Leung et. al. (2007) for discovering classification rules in remote sensor data, Øhrn A (1999) for disease diagnosis and outcome prediction and Thangavel and Pethalakshmi (2006) for dimensionality reduction.

Spatial Data Analysis presents a myriad of challenges in modelling due to the uncertainty intrinsic to it. Spatial autocorrelation is yet another feature which presents the reference with core and periphery. And clearly RST can handle this with its approximation concepts.

Spatial Data: Challenges

RST uses equivalence relation to partition the universe which needs a discrete value space. Spatial Data involves continuous values and Jensen & Shen (2004) state that there may be loss of information upon discretization of continuous information. Lin (2000) proposed Neighborhood

Rough Sets and Dubois & Prade (2009), Fuzzy Rough Sets to solve this situation. Further, the granule of information supported in RST is based on partitions which yield single layer granulation structures. Neighborhood Rough Sets associate each element of universe with a family of Neighborhood granules which leads to multi-layered granulation.

Dimensionality reduction is also accomplished by rough sets based on indiscernibility. A subset of attributes that preserve the indiscernibility and henceforth induce the same partitioning of the universe are identified and called reducts. When applied on a spatial decision system, Bai & Ge (2009) say it is possible that some spatially relevant attributes may be removed while determining the reducts. It is comprehensive to treat spatial and non-spatial attributes separately. While most of the spatial attributes like temperature, rainfall, humidity, altitude etc., are continuous, using Neighborhood Rough Sets to identify knowledge granules will be accomplishing without the loss of information due to discretization.

This chapter will study the construction of a Spatial Decision System involving continuous and auto correlated spatial attributes for spatial data based on Neighborhood Rough Sets. It contains sections on Spatial Data Mining, Rough Sets on Spatial Data Mining; Rough Set based Decision Systems for Spatial Data Analytics, Neighborhood Granulation and Neighborhood Rough Sets. These sections discuss the current approaches, applications and prospects of the fields. Use of neighborhood based systems for spatial data is augmented with sections on coverings and spatial correlation. This is followed by conclusion and future directions of work.

DECISION SYSTEMS FOR SPATIAL DATA MINING

Spatial Data Mining probably dates back to the time ever since scientists started studying spaces and people. Environmentalists, epidemiologists, urban/highway planning authorities, demographers, market researchers, weather forecasters, retailers, telecommunication/automobile engineers, the defence and others have profound use in spatial data.

Scientists have been using Geographical Information Systems (GIS) and exclusive spatial databases to store, organize and retrieve spatial data. The rate at which the spatial data is being generated is colossal and this calls for studying patterns in them. Colocation pattern mining,

spatio-temporal mining, co-occurrence mining, identifying habitats – vegetation – dense areas, predicting bird nesting scenarios, crime analysis, spatial outlier significance, disease transmission modelling, identifying distribution of disease clusters, choosing a retail spot to outstand competition, analysing ocean data, land cover and associated prediction modelling are some of the mining problems extensively studied with respect to spatial data. This has led to informed decision making by the stakeholders, hence has led to the interest in exploring this area expecting only with more accuracy.

A sample decision system for spatial data is shown in Table 2. Komorowski et. al (1999) defined a *decision system* to consist of a finite set of objects constituting the universe U , a finite set of conditional attributes A , a set of decision attributes D and a value set V_a for A from the universe. The conditional attribute set in the decision system shown are $s1, s2, s3, s4$ and $s5$. The above sample has one decision attribute d . Constructing decision systems and applying soft computing techniques in spatial data mining has become inevitable. This is due to the accommodating nature of soft computing techniques like Rough Set Theory by Pawlak (1982), Fuzzy Sets by Zadeh (1965), their variants, enhancements and hybrid versions where they deal with inconsistent and uncertain data. These extensions of classical set theory with their unique properties have helped in resolving impacts due to data pitfalls. Besides, they also blend with other extensions like Neighborhood

Table 2. A sample decision system

object_id	s1	s2	s3	s4	s5	d
1	a	high	1.6	25-30	region1	yes
2	a	moderate	1.8	25-30	region2	yes
3	a	high	1.9	20-25	region1	no
4	b	massive	1.2	30-35	region3	yes
5	b	low	2.1	25-30	region1	no
6	c	massive	2.4	30-35	region2	no
7	c	low	1.5	20-25	region2	no

theories to overcome drawbacks and result in interesting findings.

The nature of data represented in decision systems should also be duly considered. Spatial data is entitled to unique features like high dimensionality, huge quantity, auto co-relation of attributes, involves a spatial reference and will include categorical or numerical attributes with discrete or continuous values. This chapter will include discussion on decision systems based on rough sets, fuzzy sets, rough fuzzy sets and Neighborhood systems.

ROUGH SETS IN SPATIAL DATA MINING

Pawlak proposed Rough Set theory (RST) in 1982 as an extension of classical set theory. RST expresses ambiguous sets using lower and upper approximation concepts. It created an approximate reasoning from the available data. This appealing nature of RST found its use in pattern recognition and mining problems. Researchers have used RST extensively for prediction modelling and dimensionality reduction. Øhrn (1999) enlists the use of RST in medicine, for predicting success of surgeries, myocardial infarctions, post-operative long & short term survival of patients, predicting diseases and length of treatment period. Bai et. al (2010) employ Rough Set Theory to identify distribution of neural tube defects in new born children. Dimensionality reduction using *reduct* of rough sets and identifying spatial factors involved in this defect were the highlights of the work. The work proved to be one of the eye openers to potential research in ecology, epidemiology and medicine.

Spatial data represents real world entities in spatial reference. Topological representations and relationships calls for handling without ambiguity. Association rule mining in SDM makes use of exclusive spatial topological relationships like *adjacent to*, *overlapping*, *intersection* etc., and such relationships between uncertain boundaries have to be addressed. Beaubouef and Petry

(2002) has demonstrated the use of fuzzy set approach for mining association rules with spatial factors involving topological uncertainty. They have suggested imbibing RST into data modelling will help in handling indeterminate nature of data. *RCC* and *Egg-Yolk* methods are applied to study geographies with unclear boundaries. Though these methods intuitively use approximation, it is not similar to the partitioning in RST based on equivalence classes. Further, egg-yolk method is reported to have performance issues when continuous space is involved. Beaubouef (1994) proposed rough set version of *RCC* and *Egg-yolk* and recorded better observations. The paper discusses modelling boundaries of Spatial Regions using RST where each grid of spatial data is considered as an equivalence class. Low resolution means coarser boundary region and vice-versa is also true. Modelling of topological features has been highlighted. Out of the eight relations that *RCC* encompasses, five are applicable to Lower Approximation that should be a subset of Upper Approximation. However, spatial auto correlation is not addressed in this paper.

Wang (2004) demonstrated rough interpretation and introduced the idea of rough topological space and relationships between the attributes. The work has suggested handling uncertain data without coercing them to crisp data. The work recalls various approaches like evidence theory, cloud theory, error band, epsilon band and others are used to handle some parts of uncertainty in data. As part of rough spatial interpretation, this work explores the probabilistic rough set defined by a membership function $\mu_x(x)$ where its value ranges from 0 to 1. A rough spatial entity is characterised by a pair of upper and lower approximation. Rough vector space to define a point, line or area is used and a raster space for shapes and forms. This helps in translating real world coordinates to computer screen coordinates. The approach also helps in image resolution features. Further, rough topology for relationships between spatial entities is also defined as rough relation between crisp-crisp, rough-crisp and rough-rough

entities. This *rough spatial interpretation* is used by Murgante et al.(2007) to define periurban fringe. The term *periurban fringe* is defined with respect to the inclusion (proximity to cities, population distribution etc.,) and exclusion (erosion, landslide area etc.,) of innate organic rules. He has compared the analysis from Map Algebra and Spatial Rough Sets and recorded that the latter has delivered improved interpretation.

Rough Sets Based Decision Systems for Spatial Data Analytics

Mining spatial rules has been the objective in a majority of work done with respect to spatial data. A decision system with a set of objects, conditional and decision attributes is constructed. It is defined as $(U, CA \cup DA)$ where U is the universe of non-empty objects, CA are the conditional attributes and DA the decision attributes. Construction of the decision requires understanding of the nature of objects we deal with. The attributes may be of crisp or fuzzy nature. There may be indiscernible objects in the universe. If we base the decision system on rough sets, equivalence classes based on equivalence relation are found and an associated indiscernibility is defined. The objects that belong in an indiscernibility relation are similar to each other with respect to the attributes based on which equivalence relation is defined. For any subset $X \subseteq U$ and any $F \subseteq CA$, is an associated equivalence relation $IND_s(F) = \{(x, x') \mid \forall a \in F, a(x) = a(x')\}$, called *F-indiscernibility* relation. The equivalence class of this is specified as $[x]_F$. Lower and upper approximations of X with respect to F are defined as $\underline{F}X$ and $\overline{F}X$ where $\underline{F}X = \{x \mid [x]_F \subseteq X\}$ and $\overline{F}X = \{x \mid [x]_F \cap X \neq \emptyset\}$. The difference of these two sets is the boundary set. A set is rough if the boundary region is not empty.

Also, the number of attributes involved may be very large. The attributes that preserve indiscernibility relation are retained. This is achieved by

identifying *reduct* and *core* in RST. Thangavel and Pethalakshmi (2006) proposed Reduct algorithm for feature selection, also other variants, genetic algorithm based reducts etc., are available to identify reducts. This ensures that classificatory performance of reduced set of attributes is same as the original set. After identifying the reducts, decision rules are extracted.

Usually, continuous attributes have to be discretized in the rough set based decision system to establish indiscernibility. But in the case spatial data, continuous values may have to be preserved to avoid loss due to generality.

NEIGHBOURHOOD ROUGH SETS

Neighborhood system was introduced for the study of Fechet (V) spaces, which are convex vector spaces used for representing multi-dimensional features. Lin (1988) later used it to study relationships between objects. Yao (1999) generated the idea of knowledge granulation structures from it. For a given object x in the universe U , a Neighborhood is associated with it called $n(x)$ which is a *subset* of U . If $n(x)$ includes x , it is a reflexive Neighborhood. A Neighborhood system $NS(x)$ of x is defined as a non-empty family of Neighborhoods of x and is denoted as $NS(x) = \{(x)_R\}$ from U/R where R is a reflexive relation. Approximations are defined by constructing a covering of the universe with all Neighborhoods in $NS(x)$. Neighborhoods are defined using nearest neighbour methods or by distance from central point to boundary. Using the latter method, for a given set of continuous attributes in space, a Neighborhood is defined for every object in U , as $\delta(x) = \{y \mid x \in U, d(x, y) \leq \delta\}$ where $\delta > 0$ and $\delta(x)$ is δ Neighborhood information granule of x . Neighborhood approximation is applied to identify the positive and boundary region. The set of objects in the positive region are identified into decision classes without ambiguity.

Non-spatial attributes can be reduced to core attributes using rough set approximation concepts. The minimal *reduct* induced can be intersected with the attributes determined using Neighborhood Rough Sets from which the rules can be inferred. Treating the spatial and non-spatial attributes separately and integrating them later will uphold the inherent spatial features critical to the semantics of rules.

Suitability of Neighborhood Rough Sets for Spatial Data

Spatial Data: Continuous

Spatial data involves attributes like temperature, rainfall, slope, humidity and generally are continuous and involve real number space. Discretising them may lead to loss of information.

Continuous attributes are modelled based on the domain expertise, transformation using discretization to fit decision systems or using Neighborhood relations. Use of domain expertise may not be a common approach and inevitably requires ways to ensure authentication. This may be an overhead in most of the cases. Discretization of continuous values is recorded to have loss of generality in some cases. In the context of Spatial Decision System, the continuous values of attributes are to be preserved. Neighborhood systems use similarity metric to relate data and reduce the chances of information loss.

Lin.T. Y (1989) used neighborhood system or topology to formalize approximation in relational databases. The work further states that neighborhood systems can be used for approximation in numerical analysis for which a neighborhood of tolerance has to be identified even before the approximation commences.

Neighborhood structure can be used for continuous attributes represented in the decision system. Nearest neighbour and distance measure are used in Neighborhood systems of which the latter can be used for handling continuous spatial data.

Spatial Auto Correlation

Spatial correlation is the feature where the values for an attribute will tend to be similar or vary in the order of small gradients in a geographic area. Spatial features like temperature, rainfall, humidity etc., are some of the examples. It will be significant to the decision rules generated if this correlation is upheld while partitioning. Neighborhood Rough Sets is a favourable candidate which does not discretise the attribute values to induce partitioning.

Neighborhood Based Coverings

The classical rough set theory induces the partitioning of universe based on equivalence relations. The approximations (lower and upper) generated by partitions will be vary according to partitions whereas in a covering based scenario, different coverings may induce the same approximations. Yao (1999) proposed neighborhood systems and augmented that it is covering based. It generates closely related covers which will be conducive for continuous values. Although, there is redundancy to be considered in the case of covering, as removal of a data item will leave the partition incomplete (because the partitions are based on equivalence relations). This may not happen in the case of Neighborhood based coverings.

NEIGHBORHOOD GRANULATION

Definition 1: Given an N dimension real number space U , two objects $x_i, x_j \in U$, the distance metric δ between the two data objects, then following properties are satisfied in the neighborhood model.

$$\delta(x_i, x_j) \geq 0, \text{ iff } i = j, \delta(x_i, x_j) = 0 \quad (1)$$

$$\delta(x_i, x_j) = \delta(x_j, x_i) \quad (2)$$

$$\delta(x_i, x_j) + \delta(x_j, x_k) \geq \delta(x_i, x_k) \quad (3)$$

Then we called (U, δ) is a real number space. The above three properties are reflexivity, symmetry and non-transitive properties of neighborhood relation.

Definition 2: Given a finite set of objects $U = \{x_1, x_2, x_3, \dots, x_n\}$ in real number space, for every object x_i in U , then the δ -neighborhood definition is as follows:

$$\delta(x_i) = \{x | x \in U, d(x, x_i) \leq \delta\} \quad (4)$$

where $\delta > 0$, $\delta(x_i)$ is δ neighborhood information granulation from x_i and called as x_i neighborhood granulation.

Neighborhood relations are a kind of similarity relations, which satisfy reflexivity and symmetry properties. The data objects are drawn together for similarity in terms of distances and the samples in the same neighborhood granule are close to each other. Considering two x_i, x_j objects in a M-dimensional space with attribute set $A = \{A_1, A_2, \dots, A_M\}$, $f(x, A_i)$ represents the value of object x in attribute A_i , a Minkowsky distance defined by

$$\delta_p(x_i, x_j) = \left(\sum_{i=1}^M |f(x_i, A_i) - f(x_j, A_i)|^p \right)^{1/p} \quad (5)$$

$\delta_p(x_i, x_j)$ is Manhattan distance if $p=1$, Euclidean distance if $p=2$ and Chebychev distance if $p=\infty$. Distance functions are discussed by Wilson & Martinez (1997) in detail. $\delta_A(x_i)$ is

the neighborhood of data object x_i and its size is based on the threshold δ .

Spatial attributes exhibit spatial auto correlation, where the values for an attribute tend to be similar in a geographic neighborhood. This is measured using Moran's Index in a Geographic Information System (GIS). This index can only compute autocorrelation of numerical data and not that of categorical data. Using neighbourhood granulation to calculate autocorrelation addresses this issue. Appropriate distance measure can be used to calculate neighbourhoods or spatial autocorrelation of numerical, categorical and mixed data.

Considering Table 1, we can calculate the neighborhood of objects in an attribute, for e.g. slope. We decide on a neighborhood threshold 0.2.

$$\delta(z1) = \{z1, z2, z5, z7\}, \delta(z2) = \{z1, z2, z7, z8\}, \\ \delta(z3) = \{z3, z8\}, \delta(z4) = \{z4\}$$

$$\delta(z5) = \{z1, z5, z6\}, \delta(z6) = \{z5, z6\}, \\ \delta(z7) = \{z1, z2, z7, z8\}, \delta(z8) = \{z2, z3, z8\}$$

If we need to calculate the neighborhood across two attributes which are numerical, *Euclidean* distance can be used. For attributes of categorical or mixed, suitable distance measures can be used as discussed in Wilson & Martinez (1997).

Neighborhood Approximation

Considering a set of objects $U = \{x_1, x_2, \dots, x_n\}$, a neighborhood relation R , then the neighborhood approximation space $S = \{U, R\}$ and for any $X \subseteq U$, lower and upper approximations are defined as follows:

$$\left\{ \begin{aligned} \underline{R}X &= \{x_i \in U | \delta(x_i) \subseteq X, x_i \in U\} \\ \overline{R}X &= \{x_i \in U | \delta(x_i) \cap X, x_i \in U\} \end{aligned} \right\}$$

The positive, negative and boundary region are as follows:

$$\begin{cases} POS(X) = \underline{R}X \\ NEG(X) = \sim \bar{R}X \\ BND(X) = \bar{R}X - \underline{R}X \end{cases}$$

A data object will fall in positive or boundary region. Also, roughness of subset X in the approximation space is inferred from the size of boundary region. The set of objects in positive region can be classified with certainty while the objects in boundary region will have indeterminate classification.

CONCLUSION

Rough sets and its variants have been used in mining decision rules in a number of applications. Use of rough sets in spatial data mining and its relevance has been discussed in the chapter. Considering Neighborhood Rough Sets based decision system for spatial data mining will help in avoiding the loss of information. The decision attributes in spatial decision system are considered to be crisp. Bai & Ge (2010) proposed Rough Fuzzy approach to spatial decision attributes. This was to accommodate a range of memberships for decision attributes with which real world scenario can be better represented. Using Neighborhood Rough Sets helps in preserving the continuous nature of spatial data and the inherent spatial autocorrelation without which these real world features might be overlooked in the decision system.

FUTURE WORK

More work can be carried out to study the potential of Rough Fuzzy and Rough Intuitionistic Fuzzy

System with Neighborhood Systems in constructing decision systems and evaluation of rules generated. NRST based Approximation strategies can be tested on various domains and compared with RST approximations. Variants involving tolerance relations can be studied. Identifying reducts for Neighborhood Rough Sets can be studied further. Rule generation and learning for NRST can be analysed and validated for various domains.

REFERENCES

- Ahlqvist, O. (2005). Using uncertain conceptual spaces to translate between land cover categories. *International Journal of Geographical Information Science*, 19(7), 831–857. doi:10.1080/13658810500106729
- Bai, H., & Ge, Y. (2009). Handling Spatial-correlated Attribute Values in a Rough Set. *Proc. of Intl. Conf. on Computational Science and Its Applications*. doi:10.1007/978-3-642-02454-2_33
- Bai, H., & Ge, Y. (2010). *Extracting Spatial Decision Rules Using Rough Fuzzy Sets*. Accuracy 2010 Symposium, Leicester, UK.
- Bai, H., Ge, Y., Wang, J., & Liao, Y. (2010). Using Rough Set Theory to identify villages affected by birth defects: The example of Heshun, Shanxi, China. *International Journal of Geographical Information Science*, 24(4), 559–576. doi:10.1080/13658810902960079
- Beaubouef, T., & Petry, F. (1994). Rough Querying of Crisp Data in Relational Databases. *Third International Workshop on Rough Sets and Soft Computing (RSSC 94)*.
- Beaubouef, T., & Petry, F. (2002). A Rough Set Foundation for Spatial Data Mining Involving Vague Regions. *Proc. of IEEE International Conference on Fuzzy Systems*, 761 – 772. doi:10.1109/FUZZ.2002.1005090

- Cohn, A., & Gotts, N. (1996). *The 'Egg-Yolk' Representation of Regions with Indeterminate Boundaries*. In *Geographic Objects with Indeterminate Boundaries*. GISDATA II, European Science Foundation.
- Dubois, D., & Prade, H. (1990). Rough fuzzy sets and fuzzy rough sets. *International Journal of General Systems*, 17(2), 191–209. doi:10.1080/03081079008935107
- Jensen, R., & Shen, R. (2004). Semantics-preserving dimensionality reduction: Rough and fuzzy-rough based approaches. *IEEE Transactions on Knowledge and Data Engineering*, 16(12), 1457–1471. doi:10.1109/TKDE.2004.96
- Komorowski, J., Pawlak, Z., Polkowski, L., & Skowron, A. (1999). Rough Sets: A Tutorial. *Rough-Fuzzy Hybridization: A New Trend in Decision Making*, 3–98.
- Leung, Y., Fung, T., Mi, J.-S., & Wu, W.-Z. (2007). A rough set approach to the discovery of classification rules in spatial data. *International Journal of Geographical Information Science*, 21(9), 1033–1058. doi:10.1080/13658810601169915
- Lin, T. Y. (1988). Neighborhood Systems and Approximation in Relational Databases and Knowledge Bases. *Proceedings of 4th International Symposium on Methodologies of Intelligent Systems*.
- Lin, T. Y. (2000). Data Mining and Machine Oriented Modeling: A Granular Computing Approach. *Applied Intelligence*, 13(2), 331–336. doi:10.1023/A:1008384328214
- Murgante, B., Las Casas, G., & Sansone, A. (2007). *A spatial rough set for locating the periurban fringe*. SAGEO.
- Øhrn, A. (1999). *Discernibility and rough sets in medicine: tools and applications* (PhD Thesis). Norwegian University of Science & Technology.
- Pawlak, Z. (1982). Rough Sets. *International Journal of Man-Machine Studies*, 21(2), 127–134. doi:10.1016/S0020-7373(84)80062-0
- Randell, D., Cui, Z., & Cohn, A. (1992). A spatial logic based on regions and connection. *Proceedings of 3rd International Conference on Knowledge Representation and Reasoning*, 165–176.
- Thangavel, K., & Pethalakshmi, A. (2006). Feature selection for medical database using rough system. *International Journal of Artificial Intelligence and Machine Learning*, 6(1), 11–17.
- Wang, S., Yuan, H., Chen, G., Li, D., & Shi, W. (2004). Rough Spatial Interpretation. *Lecture Notes in Artificial Intelligence*, 3066, 435–444. doi:10.1007/978-3-540-25929-9_52
- Wilson, D. R., & Martinez, R. (1997). Improved heterogeneous distance functions. *Journal of Artificial Intelligence Research*, 6, 1–34.
- Yao, Y. Y. (1999). Granular Computing Using Neighborhood Systems. *Advances in Soft Computing: Engineering Design and Manufacturing*. Springer-Verlag. doi:10.1007/978-1-4471-0819-1_40
- Zadeh, L. A. (1965). Fuzzy Sets. *Information and Control*, 8(3), 338–353. doi:10.1016/S0019-9958(65)90241-X

ADDITIONAL READING

- Du, Y., Hu, Q., Zhu, P., & Ma, P. (2011). Rule learning for classification based on neighborhood covering reduction. *Information Sciences*, 118(24), 5457–5467. doi:10.1016/j.ins.2011.07.038
- Hu, Q., Yu, D., Liu, J., & Wu, C. (2008). Neighborhood rough set based heterogeneous feature subset selection. *Information Science*, 178(18), 3577–3594. doi:10.1016/j.ins.2008.05.024

Hu, Q., Yu, D., & Xie, Z. (2008). Neighborhood classifiers. *Expert Systems with Applications*, 34(2), 866–876. doi:10.1016/j.eswa.2006.10.043

Pawlak, Z. (1997). Rough set approach to knowledge-based decision support. *European Journal of Operational Research*, 99(1), 48–57. doi:10.1016/S0377-2217(96)00382-7

Slowinski, R., & Vanderpooten, D. (2000). A generalized definition of rough approximations based on similarity. *IEEE Transactions on Knowledge and Data Engineering*, 12(2), 331–336. doi:10.1109/69.842271

Yao, Y., & Zhao, Y. (2008). Attribute reduction in decision theoretic rough set models. *Information Sciences*, 178(17), 3356–3373. doi:10.1016/j.ins.2008.05.010

Yong, L., Wenliang, H., Yunliang, J., & Zhiyong, Z. (2014). Quick attribute reduct algorithm for neighborhood rough set model. *Information Sciences*, 271, 65–81. doi:10.1016/j.ins.2014.02.093

Zhang, X., & Xu, W. (2013). Methods for Lower Approximation Reduction in Inconsistent Decision Table Based on Tolerance Relation. *Applications of Mathematics*, 4, 144–148.

KEY TERMS AND DEFINITIONS

Categorical Data: Data grouped to form a category when numerical representation is not adequate.

Distance Measure: Measures to calculate distance between numerical, categorical or mixed data.

Granulation: It is using groups or clusters of data objects formed on the basis of similarity in an incomplete information system.

Neighborhood Systems: They provide granulation structure for each element of a universe.

Rough Sets: An extension of classical set theory, which provides an approximation of crisp set in terms of lower and upper approximation.

Spatial Auto Correlation: The values of a spatial attribute tend to be close to each other and vary gradually from core to periphery of a geographic region.

Spatial Data Mining: Mining patterns or rules from spatial data with respect to a geographic reference.

The Ontology of Randomness

D

Jeremy Horne

The International Institute of Informatics and Systemics, USA

INTRODUCTION

“Random” commonly is associated with determinism, order, prophecy, and the future. Starkly put by one philosopher, “Randomness is unpredictability” (Eagle, 2005), quoting mainstream logician Suppes in saying “Phenomena that we *cannot predict* must be judged *random*” (Suppes, 1984, p. 32). People rely upon uncertainty (encryption and defeating bias). Tables of random numbers are based on the supposition that humans subconsciously create patterns and cannot generate randomness. Gambling casinos are about the “luck of the draw”. Human survival depends on remembering patterns of past events in order to adapt to the changing environment, as well as on randomness in science.

Our focus is not only on considering whether randomness exists, i.e., ontology, but the reason for it. Turmoil in the field of experts is expressed by “... there is no such thing as a random number - there are only methods to produce random numbers...” (Von Neumann, 1951; Fiorini and De Giacomo, 2014, p. 5). How can one create something that cannot exist? Yet, at the most basic building block level of existence, Planck scale, there is inherent uncertainty (perhaps immanent in the macro world) with which persons have difficulty accepting.

A new view of randomness will be presented here, based on the often neglected philosophy of dialectics, apprehension of something in terms of what it is not, a process understanding, rather than one of identifying objects.

This short space allows only a small sampling of approaches to the ontology of randomness for the reader to explore in depth through cited references. If one can begin to think beyond the

mechanics (as in random number generators) then this encyclopedia entry will have accomplished its purpose.

BACKGROUND

The Ontology of Randomness in this *Encyclopedia of Information Science and Technology*, focuses more on the thinking underpinning science, more specifically, whether randomness even exists, i.e., its ontology (Feibleman, 1951). Tests for randomness appear to assume what is trying to be shown, i.e., there is indeed randomness, begging the question of whether there is innate structure, or order, in the universe.

Because of very limited space, extended tutorials and discussions about ontology (the nature of existence), epistemology (how we know), the problems of induction (Hume, 1888; Mill, 1843; Russell, 1919; Ramsey, 1929; Keynes, 1921, pp. 305-314; p. 24 et seq.), stochastic analysis (series of random variables), and the problems of representation (Plato’s cave allegory as the philosophical foundations of statistics) have been omitted. Also omitted is a discussion of the role of randomness in logical scientific exploration (Popper, 1934; hypothetico-deductive, 2015; Copi, 1979; Rosser, 1953; Mendelson, 1997; Whewell, 1847; Feyerabend, 1975), as well as discussions of Abraham de Moivre (bell curve), Pierre-Simon Laplace (calculus of probabilities), and martingales (Birnbaum and Lukas, 1980). There are many other conversations about the differences between probability, chance (Keynes, 1921; Eagle, 2010), and randomness that would enrich a more complete treatment of the subject. This says nothing of the hundreds of mainstream

DOI: 10.4018/978-1-5225-2255-3.ch161

works of probability theorists and their views on randomness. Instead, given here is somewhat analogous to a brief literature search, with a focus on summarizing several main views of what people think randomness is and considering the implication of its existence status. If such can get the reader to think beyond technology, focusing more on the “why”, then this chapter will have accomplished its goal.

A Historical Perspective

Prediction for an early hunter meant being able to survive any challenge in the wild, that ability being taught to others. One thinks of shamans, soothsayers, and prophets being among the more esteemed members of various cultures.

Randomness came in the form of early games (Wolfram, 2016), a symbolized way of acting out in anticipation of meeting unpredictable events. As societies became more organized, those more skilled in correctly anticipating events led the military. Athletic competitions and often games emerged as simulations of military combat (History of Sport, 2016). In modern times, sophisticated modeling and simulation techniques describe, analyze, and plan for unexpected situations (MSBOK, 2016).

Current Views of Randomness

For the general public, “random” and its derivatives are defined variously as:

- “...proceeding, made, or occurring without definite aim, reason, or pattern”. (Random - Dictionary – Random House, 2016).
- “...the lack of pattern or predictability in events”. (Randomness - Wiki, 2016).
- “...without definite aim, direction, rule, or method; lacking a definite plan, purpose, or pattern”. (Random - Dictionary – Webster, 2016).

- “... Made, done, happening, or chosen without method or conscious” (Random - Dictionary - Oxford American, 2016).

More formally but non-exhaustive are the main areas of:

- Relative frequency concept of probability;
- Information complexity/compressibility of information – Chaitin-Kolmogorov-Smirnov-Shannon;
- Predictability and probability;
- Energy – Boltzmann-Clausius – entropy;
- Observation about completeness and certainty – Heisenberg, Gödel.

All of these overlap but have in common the inability of discerning future specificity, i.e., uncertainty. Following is a brief sketch of these views.

Relative Frequency Concept of Probability

Von Mises probability centers on “relative frequency”, the ratio of appearances of an attribute in an event to the total number of events (as in how a color of a flower varies in a seed type, the number of dots shown with each roll of dice, or the number of deaths in an age group) (Von Mises, 1957, p. 11). As more events are observed the ratio becomes more refined, or more stabilized (.5,.51,.519,.5196, etc.)(Ibid., pp. 14-15). A “collective” is a “sequence of uniform events or processes which differ by certain observable attributes, i.e., a set of events from which a certain characteristic is to be observed” (Ibid., p. 12). Probability is the likelihood of “...encountering a certain attribute in a given collective” (Ibid.). Further, it is the “...limiting value of the relative frequency in a true collective which satisfies the condition of randomness” (Ibid., p. 24).

A “true” or “collective appropriate for the application of probability” must indeed have

limiting values that “remain the same in all partial sequences which may be selected from the original one in an arbitrary way” (Ibid., pp. 24-25). So, if all the odd numbered elements are selected or ones designated by a prime number, the selection of this partial sequence being before the relative frequency is known, this “place selection” (deciding “whether an element should or should not be included” (Ibid.) regardless of the element’s attribute) will exhibit that relative frequency. That is, putting a member of the original sequence in the “selected partial sequence” is to be done apart from observing the value of the relative frequency. All sequences, then, should exhibit the same relative frequency, the observation being done after all the partial sequences have been identified. Randomness, then, is place selection on this basis.

Information Complexity/ Compressibility

If we cannot say what has generated a string of characters (method) that string is random, a view held by Kolmogorov (Kolmogorov complexity) and Chaitin (information compressibility), often combined as “Kolmogorov–Chaitin complexity, algorithmic entropy, or program-size complexity” (Kolmogorov complexity, 2016).

Kolmogorov (1998) argues that von Mises’ frequency view cannot be applied as probability theory to actual situations where there are to be a certain number of trials. While basically sound it also is not amendable to a rigorous pure mathematical analysis; there needing to be a way of measuring the algorithm’s complexity. Formally stated:

$$T = (t_1, t_2, \dots, t_n)$$

of N zeros and ones: $t_k = 0$ or 1 .

Such a table will be called random, if, while choosing the subset A of sufficiently large size from $1, N$ by different methods there is stability in the frequency

There is stability in the frequency

$$\pi(A) = 1/n \sum t_k$$

$$k \in A \quad (\text{Kolmogorov, 1998, p. 388})$$

Chaitin (1975, 1998), very much like Kolmogorov, says, “something is random if it is algorithmically incompressible or irreducible. More precisely, a member of a set of objects is random if it has the highest complexity that is possible within this set.” (Chaitin, Randomness, 1975, p. 71).

For example,

010101010101010101

01101100110111100010

The first can be stated concisely, “write ‘01’ 10 times”. The second does not have any method to generate a pattern of digits, hence no expression shorter than the 20 characters of 0 and 1 (not compressible) (Chaitin, Algorithmic, p. 16).

A common criticism is that neither version serves as statistical test for randomness (Eagle, 2005, pp. 11-12).

Predictability and Probability

Probability is not the same as randomness, although it relies on it by describing individuals selected without bias from an “infinite” population (Ramsey, p. 34-35, 39). In essence, if one cannot predict the nature of the next event (done by statistics), then that event is more random. Statistics increases predictability and lessens randomness, assuming that the future resembles the past (Whewell, 1847; Hume; 1888; Keynes, 1921, p. 305-314). A deviation of the expected values in a probability analysis would be considered random. If this observation doesn’t help in making the prediction, the event is said to be “random” (Eagle, 2005):

Entropy

Entropy has numerous definitions but forms a basis of randomness. If everything is uniform, there is no way of distinguishing something. There is no information or ability to say specifically what will happen next. Two main currents of entropy are energy dispersion and lack of information.

A system starts in an ordered state (arrangement according to a scheme) and increasing entropy allows many more possibilities of arrangements, i.e., less ordering. In a closed system, energy is equally dispersed when it attains maximum entropy (Clausius, 1870; Thompson, 1852). Ultimately, inside the final stage of the universe energy will be equally dispersed; there will be no differentiation of objects but only uniformity, an equal probability of occurrence, or randomness (Davies, 1984).

Shannon (1948, p. 393) said that compared to what is predicted in a message, what actually gets through determines entropy. That is, what would be the expected message length a person would need to be to transmit the measured value of a random variable?

Note that randomness is not the same as the more technical “chaotic”, as in chaos theory (Gleick, 1987).

Completeness and Uncertainty

Perhaps the most significant of all aspects of randomness in terms of predictability and certainty was formulated by Heisenberg, i.e.,

We simply cannot know the present in principle in all its parameters. Therefore all perception is a selection from a totality of possibilities and a limitation of what is possible in the future. ... [O]ne could be tempted to conclude that behind the observed, statistical world a “real” world is hidden, in which the law of causality is applicable. We want to state explicitly that we believe such speculations to be both fruitless and pointless. ... Because all experiments are subject to the laws

of quantum mechanics, it follows that quantum mechanics once and for all establishes the invalidity of the law of causality.” (Heisenberg, 1927)

This statement stands on its own, a source of much modern day angst.

Does Randomness Exist?

A major tenet of the ages of Reason and Enlightenment was to banish anything unpredictable as flying in the face of God’s design (Paley, 1881), hence atheistic. For example, Bernoulli said, “Everything that exists or originates under the sun, – the past, the present, or the future, – always has in itself and objectively the highest extent of certainty” (Bernoulli, 1713). In 1718 Abraham De Moivre (Central Limit Theorem – Gaussian curve, or binomial distribution), said one could account for all phenomenon in terms of how likely it would be for it to appear, i.e., degree of certainty, or probability. Like many other natural philosophers of the day, he did not think chance existed; everything was ordered (De Moivre, 1756, p. 253). Einstein said about uncertainty (randomness) that “God does not play dice” (Born, 1926, p. 90-91). Russell was unwilling to give up strict determinism, saying “Where determinism fails, science fails” (Russell, 1996, p.67).

“...[T]rue randomness’ is a mathematical impossibility” (Fiorini, 2014, p. 2). Calude (2015, p. 4) says, “There is no true randomness, irrespective of the method used to produce it”. Then, “Anyone who considers arithmetical methods of producing random digits is, of course, in a state of sin. ... There is no such thing as a random number - there are only methods to produce random numbers, and a strict arithmetic procedure of course is not such a method (Von Neumann, 1951, p. 36)”. Eagle says, “No mathematical definition of random sequence can adequately capture physical randomness” (Eagle, p. 14). Suffice it to say, such statements require one to look closely at terms often used in conjunction with “random”, such as “correlation”, “probability”, “chance”, and uncertainty.

Living With Randomness

Random numbers tables help avoid human bias in selecting samples. Other applications include cryptography, gaming, polling, and scientific experimentation, among many others (Applications of Randomness, 2015). People rely on randomness for uncertainty (bias-free selection). Random numbers sequences may be “true” (unpredictable by anyone and natural, as in radioactive decay), pseudo (unpredictable to one but predictable, to another), and quasi – uniformly distributed sequence of elements (RNG – types, 2015).

Eagle demands for an analysis of randomness:

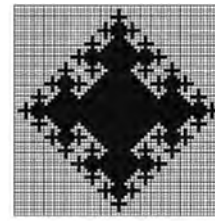
- Unbiased sampling, as in scientific experimentation and polling;
- Determining if single events are random or not, i.e., “finite randomness”.
- Explaining the behavior of systems, be it purposeful or not. Randomness needs empirical validation (confirmed) or invalidation (disconfirmed);
- Randomness needs to be compatible with determinism, as in processes like population genetics or chaotic dynamics. (Eagle, 2005, pp. 7-8)

Numerous tests for randomness include statistics and complexity analysis (as in Kolmogorov-Smirnov). If a sequence of elements (characters, images, and so forth) passes one or more of these tests, then it is considered “random”, given just the test. Note that these tests are applied to what has been generated. (Tests for Randomness, 2015)

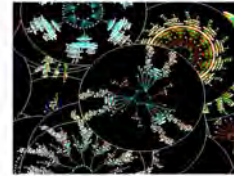
ISSUES

Emergence illustrates a form of induction, where the whole is greater than the sum of the parts. The founder of the journal *Emergence* says “... emergence is appealed to when the configuration of the components of a complex system offers more explanatory insight into the dynamics of the

Table 1. Patterns from randomly concatenated binary formula



(Wolfram, NKS, 2002, p. 170)



(Wuensche, 2015)

system than do explanations based on the parts alone. (Goldstein, 2015, p. 53) Wolfram (2002, p. 170) and Wuensche (2016) show how reputed randomly concatenated individual binary formula produce emerging patterns (here, fractals), as in:

Related to emergence, literature abounds on there being order from a “potential” or chaos (Kauffman; 1988; Gleick, 1994).

A pattern does not necessary mean what we think it represents (e.g., likeness of a person’s face in the clouds), or pareidolia (Pareidolia, 2015). Conversely, regularity appears in some cases to be a function of duration, an issue raised by the tests for randomness. The tests are performed on what has been generated, but long-term there might be regularity (or not - “pseudo order”). The study of statistics is based on the future resembling the past, one of Hume’s problems of induction.

Everything is apprehended in terms of what it is not, randomness being no exception. (dialectics). Reputable scientific institutions do not seem to hesitate in accepting that “...randomness is the opposite of order, the absence of pattern. “(MIT, 2016), but what is order?

Binary (propositional) logic is the foundation of the machine language used by computers but also represents the innate structure of the universe (Horne, 1997, 2012, 2015). At Planck Scale, no solidness exists, or at least anything that would remain in the same place (Heisenberg, 1983), but here are the fundamental and binary building blocks of the universe represented by two variables (Misner, 1973, p. 1235 et. seq.; Piaget, 1958;

Digital Physics, 2015; Wheeler, 1990, p. 5), the universe and a background into which something flicks in and out of existence (Hawking, 2015). This is where order meets randomness, one in terms of the other. In the quantum-cosmological arena, our universe stemmed from a singularity, uniformity. Here, the singularity dialectically existed either in terms of a “nothing” or another universe, or world.

Deductive systems may represent order, as in propositional logic, but ultimately, as Gödel (1931) indicated, there is incompleteness. One may ask simply what was used to construct the systems. Once someone points to an axiom, proposition, rule, etc. it may be asked what gave rise to these, ad infinitum. Conventions are only stopgaps (Horne, 1997).

Radioactive decay, Brownian movement, and “white noise” are random but these could follow some regularity which simply is unknown to us. In logic, there is the fallacy of *argumentum ignoratum*, or the fallacy from ignorance (Fallacy of Ignorance, 2016). Because one has not encountered an entity (a pattern), it does not mean that the entity does not exist.

Randomness inextricably involves time, at least what we think of time, assuming that it exists at all. If modern physicists are correct, time may be an illusion (Aaron, 2013). Einstein said that time is relative, each person having their own time. It is conceivable that from a distant part of the universe, one may be able to see our future, meaning that may have been determined, hence predictable in some way. Like time, then, predictability would be relative.

Finally, from a second order cybernetics [recursive] (Heylighen and Joslyn, 2001) perspective, one may ask whether this chapter is compressible; indeed it just may be random (☺).

FUTURE RESEARCH DIRECTIONS

Work should proceed in these areas. Randomness is tied to determinism, and determinism is about

free will. If everything is predetermined one may disregard making ethical choices. Social cohesion may rely on acting *as though* there were free will. We may act as though there were randomness, similar to that of doing a logic proof in drawing an assumption line and displaying the consequences.

There always is a persuasive idea that we are someone’s idea, analogous to a computer program (Bostrom, 2003). The question then arises about randomness is that world, even though we are predetermined.

Astrophysicists have determined that the universe is tending towards disorder, a heat death, all energy being equally dispersed, or a state of maximum entropy (Davies, 1984). Randomness, it seems will have its final say, winning out over order. However, this may not be the final word, if Hawking and others are correct about there being other universes and the possibility that ours in its end state may become a part of another dimension. Coupled with this is the recent speculation about black holes as portals into other worlds with other dimensions, i.e., as Hawking (2016) states, “Things can get out of a black hole, both to the outside, and possibly, to another universe.”

CONCLUSION

By no means has this chapter given complete treatment of randomness. Too, ours has been a Western orientation, not accounting adequately for East Asian views (Hinduism and Buddhism, for example). A case has been made for the opposite of random, order.

Investigators rely on generating a series of unexpected events that are given designations, such as numbers or letters, to identify what will be used to select members of a sample, scramble messages, or in simulating events in a model (among various uses). The ontology of randomness is reduced to discussing whether uncertainty is innate in the universe. Can what gives rise to Heisenberg-type uncertainties, the core of “natural randomness” ultimately be predicted? For now,

the answer appears to be “no”, but so were time and space thought to be absolute prior to Einstein. Physics (as expressed by its laws) (Feynman, 2001), like reputedly deductive systems is incomplete (Gödel, 1931), or something is missing or uncertain (Heisenberg, 1983). The uncertainty that we observe can produce emergence, or order that cannot be accounted for by the pieces that to us are assembled randomly.

Randomness has two aspects, with regularity, or uniformity, pursuing us in two domains, one of object and one of process. As an object, one thinks of Boltzmann, Clausian entropy, energy being dispersed evenly, with no distinction between one space-time coordinate and the next. The universe will die a heat death, in this way. Or we can think of the integrity of information, both by content (as in Kolmogorov-Chaitin compression) or integrity of transmission (Shannon, 1948). The second aspect is the regularity of process; the method of arriving at a conclusion is uniform, the future resembles the past, and the way of understanding these is the same: logic, unbiased and equally distributed sampling, scientific methods, replication, and so forth.

There is a parallel process within regards to knowledge. Knowledge acquisition comes through synthesis and analysis, the former being the assemblage of pieces to make a whole, the latter the subdivision of the whole into pieces, a process formalized by Descartes, who said, “... we cannot conceive body unless as divisible” (Descartes, 1641). At some point a subdivided world of knowledge may have such a small granularity of detail that one idea is no more significant than the next, a common complaint being an “information glut”. Too, knowledge is analogous to energy entropy, where once emitted cannot be “un-emitted”. Then, there is uniformity as a process that allows the experimental scientific method (Whewell, 1841). $E=MC^2$ says energy is known because of “matter” (material) and vice versa. Analogously, there is the duality of uniformity – process in terms of material and vice versa. Ultimately the ontology

of randomness (in all its aspects of entropy, uncertainty, and lack of prediction) is existence in terms of order, the dialectics of something being in terms of what it is not.

REFERENCES

- Aaron, J. (2013, October). Entangled toy universe shows time may be an illusion. *New Scientist*, 25. Retrieved from <https://www.newscientist.com/article/dn24473-entangled-toy-universe-shows-time-may-be-an-illusion/>
- Applications of Randomness. (2016). Retrieved 14 February 2016 from https://en.wikipedia.org/wiki/Applications_of_randomness
- Bernoulli. (2005). *The Art of Conjecturing* (O. Sheynin, Trans.). Retrieved 14 February 2016 from www.sheynin.de/download/bernoulli.pdf
- Born, M. (1926). Retrieved 14 February 2016 from https://archive.org/stream/TheBornEinsteinLetters/Born-TheBornEinsteinLetters_djvu.txt
- Bostrom, N. (2003). Are you living in a simulation? *The Philosophical Quarterly*, 53(211), 243–255. doi:10.1111/1467-9213.00309
- Calude, C.S. (2016). *Quantum randomness: From practice to theory and back*. CDTMS Research Report Series. CDMTCS-442 September 2013/ Revised November 2015 Centre for Discrete Mathematics and Theoretical Computer Science.
- Chaitin, G. (1975). Randomness and mathematical proof. *Scientific American*, 232(5), 47–52. doi:10.1038/scientificamerican0575-47
- Chaitin, G. (2001). *Exploring Randomness*. London: Springer-Verlag. Retrieved 14 February 2016 from <https://www.cs.auckland.ac.nz/~chaitin/ait/index.html>
- Chaitin, G. (2003). *Algorithmic Information Theory*. Yorktown Heights, NY: IBM.

- Clausius, R., & Hirst, T. A. (1867). *The Mechanical Theory of Heat: With Its Application to the Steam Engine and to the Physical Properties of Bodies*. London: J. Van Voorst.
- Copi, I. (1979). *Symbolic logic*. New York: Macmillan.
- Davies, P. (1984). *God and the New Physics*. New York: Simon and Schuster.
- Descartes, R. (1984). Meditations on first philosophy. In *The Philosophical Writings of René Descartes* (vol. 2). Cambridge, UK: Cambridge University Press.
- Digital Physics. (2016). Retrieved 14 February 2016 from https://en.wikipedia.org/wiki/Digital_physics
- Eagle, A. (2005). Randomness is unpredictability. *The British Journal for the Philosophy of Science*, 56(4), 749–790. doi:10.1093/bjps/axi138
- Fallacy of Ignorance. (2016). Retrieved 14 February 2016 from https://en.wikipedia.org/wiki/Argument_from_ignorance
- Feibleman, J. K. (1951). *Ontology*. Baltimore, MD: The Johns Hop-kins Press.
- Feyerabend, P. K. (1993). *Against method*. London: Verso.
- Feynman, R. P. (2001). *The Character of Physical Law*. Boston: MIT Press.
- Fiorini, R. A. (2014). The entropy conundrum: A solution proposal. *1st International Electronic Conference on Entropy and Its Applications*. Retrieved 14 February 2016 from www.sciforum.net/conference/ecea-1
- Fiorini, R. A., & De Giacomo, P. (2014). An open logic approach to EPM. *1st International Electronic Conference on Entropy and Its Applications*. Retrieved 14 February 2016 from www.sciforum.net/conference/ecea-1
- Gleick, J. (1987). *Chaos*. New York: Penguin Books.
- Gödel, K. (1962). On Formally Undecidable Propositions of Principia Mathematica and Related Systems. New York: Basic Books.
- Goldstein, J. (1999). Emergence. *Complexity and Organization*, 1(1), 49-72. Retrieved 14 February 2016 from http://www.anecdote.com/pdfs/papers/EmergenceAsAConsutructIssue1_1_3.pdf
- Hall, P., & Hyde, C. C. (1980). *Martingale Limit Theory and its Application*. New York: Academic Press.
- Hawking. (2016). Retrieved 14 February 2016 from <http://www.hawking.org.uk/into-a-black-hole.html>
- Heisenberg, W. (1983). The actual content of quantum theoretical kinematics and mechanics. *Zeit-schrift fur Physik*, 43(3-4), 172-198. Retrieved 14 February 2016 from https://ia600500.us.archive.org/10/items/nasa_tech-doc_19840008978/19840008978.pdf
- Heylighen, F., & Joslyn, C. (2001). Cybernetics and second-order cybernetics. In *Encyclopedia of Physical Science & Technology* (3rd ed.). New York: Academic Press. Retrieved 14 February 2016 from <http://pespmc1.vub.ac.be/papers/cybernetics-epst.pdf>
- History of Sport. (2016). Retrieved 14 February 2016 from https://en.wikipedia.org/wiki/History_of_sport
- Horne, J. (1997). Logic as the language of innate order in consciousness. *Informatica*, 22(4), 675–682.
- Horne, J. (2012). A new three dimensional bivalent hypercube. *Neuroquantology*, 10(1), 20–30. doi:10.14704/nq.2012.10.1.361

- Horne. (2016). Research and Applications in Global Supercomputing. *Supercomputers: A philosophy of their language*. Hershey, PA: IGI Global. DOI: 10.4018/978-1-4666-7461-5.ch017
- Hugenii, C. (1714). *Libellus De Ratiociniis In Ludo Aleae, or, The Value of All Chances in Games of Fortune; Cards, Dice, Wagers, Lotteries, &c. Mathematically Demonstrated*. London: S. Keimer. Retrieved 14 February 2016 from <http://www.stat.ucla.edu/history/huygens.pdf>
- Hume, D. (1888). *Hume's Treatise of Human Nature*. Oxford, UK: Clarendon Press. Retrieved 14 February 2016 from <http://plato.stanford.edu/entries/induction-problem/>
- Hypothetico-deductive. (2016). Retrieved 14 February 2016 from https://en.wikipedia.org/wiki/Hypothetico-deductive_model
- Kauffman, S. (1993). *The Origins of Order*. New York: Oxford University Press.
- Keynes, J. M. (1921). *A Treatise on Probability*. London: Macmillan.
- Kolmogorov, A. N. (1998). On tables of random numbers. *Theoretical Computer Science*, 207(2), 387–395. doi:10.1016/S0304-3975(98)00075-9
- Mendelson, E. (1997). *Introduction to Mathematical Logic* (4th ed.). London: Chapman & Hall.
- Mill, J. S. (1843). *System of Logic, Ratiocinative and Inductive: Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation* (Vol. 1). London: John W. Parker.
- Misner, C. W., Thorne, K. S., & Wheeler, J. A. (1973). *Gravitation*. New York: W.H. Freeman and Company.
- MIT. (2016). *First evidence that quantum processes generate truly random numbers*. MIT Technology Review.
- Moivre, D. (1756). *The Doctrine of Chances, or a Method of Calculating Probabilities in Play*. London: A. Millar. Retrieved 14 February 2016 from <http://www.ime.usp.br/~walterfm/cursos/mac5796/DoctrineOfChances.pdf>
- MSBOK. (2016). *Modeling and Simulation Body of Knowledge*. Retrieved 14 February 2016 from http://www.msco.mil/documents/_25_MSBOOK-20101022DistA.pdf
- Paley, W. (1881). *Natural Theology*. New York: American Tract Society. Retrieved 14 February 2016 from <https://ia600409.us.archive.org/1/items/naturaltheology00pale/naturaltheology-00pale.pdf>
- Pareidolia. (2016). Retrieved 14 February 2016 from <https://en.wikipedia.org/wiki/Pareidolia>
- Piaget, J. (1958). *Logic and Psychology*. New York: Basic Books, Inc.
- Popper, K. R. (1959). *The Logic of Scientific Discovery*. New York: Basic Books. Retrieved 14 February 2016 from www.oxforddictionaries.com/us/definition/american_english/random
- Random - Dictionary – Webster. (2016). Retrieved from www.merriam-webster.com/dictionary/random
- Randomized Communication Complexity. (2016). Retrieved 14 February 2016 from https://en.wikipedia.org/wiki/Communication_complexity#Randomized_communication_complexity
- RNG types. (2016). Retrieved 14 February 2016 from <http://www.mathworks.com/help/stats/generating-quasi-random-numbers.html> <https://www.random.org/randomness/>
- Rosser, J. B. (1953). *Logic for Mathematicians*. New York: Chelsea Publishing Company.

Russell, B. (1912). *Problems of Philosophy*. New York: Henry Holt and Company. Retrieved 14 February 2016 from <https://ia801409.us.archive.org/4/items/problemsofphilo00russuoft/problem-sofphilo00russuoft.pdf>

Russell, B. (1919). *Introduction to Mathematical Philosophy*. London: George Allen & Unwin, Ltd.

Russell, B. (1996). *A Fresh Look at Empiricism. Determinism and Physics*. London: Routlage.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technology Journal*, (27), 379–423. Retrieved 14 February 2016 from [https://en.wikipedia.org/wiki/Entropy_\(information_theory\)](https://en.wikipedia.org/wiki/Entropy_(information_theory))

Suppes, P. (1984). *Probabilistic Metaphysics*. Oxford, UK: Blackwell.

Tests for Randomness. (2016). Retrieved 14 February 2016 from <http://faculty.rhodes.edu/wetzel/random/mainbody.html#3tests>

Thomson, W. (1852). (Lord Kelvin) (1852). On a universal tendency in nature to the dissipation of mechanical energy. *Proceedings of the Royal Society of Edinburgh*, 19(April). Retrieved from www3.nd.edu/~powers/ame.20231/kelvin1852.pdf

Turing, A. M. (1936). On computable numbers, with an application to the Entscheidungs problem. *Proceedings of the London Mathematical Society*, 42, 230–265.

Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 59(236).

Von Mises, R. (1957). *Probability, Statistics, and Truth*. New York: Macmillan. Retrieved 14 February 2016 from http://155.97.32.9/~mhaber/Documents/Course%20Readings/vonMises_ProbStatsTruth.pdf

Von Neumann, J. (1951). *Various Techniques Used in Connection with Random Digits, Monte Carlo Method*. Retrieved 14 February 2016 from <https://dornsifecms.usc.edu/assets/sites/520/docs/VonNeumann-ams12p36-38.pdf>

Wheeler, J. A. (1990). Information, physics, quantum: The search for links. In W. Zurek (Ed.), *Complexity, Entropy, and the Physics of Information*. Redwood City, CA: Addison-Wesley.

Whewell, W. (1847). *The Philosophy of the Inductive Sciences*. London: John W. Parker.

Wolfram, S. (2002). *A New Kind of Science*. Champaign, IL: Wolfram Media. Retrieved 14 February 2016 from <http://www.wolframscience.com/nksonline/toc.html>

Wolfram, S. (2016). Retrieved 14 February 2016 from <https://www.wolframscience.com/reference/notes/968b>

Wuensche, A. (1993). Retrieved 14 February 2016 from <http://www.ddlab.com/>

ADDITIONAL READING

Cardano, D. L. A. L. (1657) Retrieved 14 February 2016 from http://www.filosofia.unimi.it/cardano/testi/operaomnia/vol_1_s_10.pdf

Chaitin, G. (2016). Home Page. Retrieved 14 February 2016 from <https://ufrij.academia.edu/GregoryChaitin>

Eagle, A. (2012) Chance versus randomness. *Stanford Encyclopedia of Philosophy*. First published Wed Aug 18, 2010; substantive revision Thu Feb 9, 2012 Retrieved 14 February 2016 from <http://plato.stanford.edu/entries/chance-randomness/>

Grinstead, C. M., & Snell, J. L. (2016). *Introduction to Probability*. Retrieved 14 February 2016 from https://www.dartmouth.edu/~chance/teaching_aids/books_articles/probability_book/amsbook.mac.pdf

Kolmogorov, A. N. (1956). *Foundations on the Theory of Probability*. 2nd ed. Nathan Morrison (Trans.). New York: Chelsae.

Krzysztof Maslanka, K., & Cislo, J. (2013). *Order from Randomness*. Retrieved 14 February 2016 from <http://arxiv.org/pdf/0706.3785v1.pdf>

Martin-Löf, P. (1966). The definition of random sequences. *Information and Control* 9 (6), 602–619. doi:. Retrieved 14 February 2016 from [http://statweb.stanford.edu/~cgates/PERSI/Courses/Phil166-266/Martin-Lof.pdf10.1016/s0019-9958\(66\)80018-9](http://statweb.stanford.edu/~cgates/PERSI/Courses/Phil166-266/Martin-Lof.pdf10.1016/s0019-9958(66)80018-9)

Random Organization. (2016). Retrieved 14 February 2016 from www.random.org

Thorn, C. D. (2003). Randomness and entropy: An introduction. Sans institute reading room site. Retrieved 14 February 2016 from <https://www.sans.org/reading-room/whitepapers/vpns/randomness-entropy-introduction-874>

Venn, J. (1888). *The Logic of Chance: An essay on the foundations and province of the theory of probability, with especial reference to its logical bearings and its applications to moral and social science and to statistics* (3rd ed., rev. and enl.), (pp. 96-118). London, Great Britain: Macmillan and Co, xxix, 508 pp. doi:10.1037/14079-005

KEY TERMS AND DEFINITIONS



Cartesianism: Repeated subdivision; reductionism.

Chance: Unexpected happening without regard to any other phenomenon or explanation.

Complexity: Amount and detail of intricacy, often expressed in the most reduced or abbreviated form.

Compressibility: Ability to produce the same essence of content in a reduced or compacted manner.

Deduction: In logic, if the premises are true, the conclusion is guaranteed. Closed systems.

Epistemology: A study of what exists.

Entropy: Either a dispersal of energy (Clausius) or loss of information integrity (Shannon).

Induction: In logic, the degree of probability that a conclusion will follow from the premises. Open systems.

Ontology: The study of what exists.

Probability: An inductive form of analysis by assessing the degree of certainty of future events, obtained by observing and analyzing an unbiased population sample.

Random: Uncertain; unpredictable; not having any method or algorithm to produce a phenomenon.

Scientific Method: Extrapolating from the past to project to the future.

Statistics: A form of computable induction, deciding the degree of certainty of future events.

Order Statistics and Applications

E. Jack Chen
BASF Corporation, USA

INTRODUCTION

Order statistics refer to the collection of sample observations sorted in ascending order and are among the most fundamental tools in non-parametric statistics and inference. Statistical inference established based on order statistics assumes nothing stronger than continuity of the cumulative distribution function of the population and is simple and broadly applicable. Important special cases of order statistics are the minimum and maximum value of samples, the sample median and other sample quantiles. Order statistics have been widely studied and applied to many real-world issues. It is often of interest to estimate the reliability of the component/system from the observed lifetime data. In many applications we want to estimate the probability that a future observation will exceed a given high level during some specified epoch. For example, a machine may break down if a certain temperature is reached. For more applications of order statistics, see (Arnold et al., 2008; David & Nagaraja, 2003; DasGupta, 2011).

It is the purpose of this chapter to present some of the more important results in the sampling theory of order statistics and of functions of order statistics and their applications, such as tests of independence, tests of goodness of fit, hypothesis tests of equivalence of means, ranking and selection, and quantile estimation. Order-statistics techniques continue to be key components of statistical analysis. For example, quantile estimates are used to build empirical distributions, tests of goodness of fit are used to check the validity of the fitted input distribution, hypothesis tests of equivalence of means, and ranking and selection procedures are used to compare the performance of multiple systems.

DOI: 10.4018/978-1-5225-2255-3.ch162

BACKGROUND

Let $X_i, i = 1, 2, \dots, n$, denote a sequence of mutually independent random samples from a common distribution of the continuous type having a probability density function (pdf) f and a cumulative distribution function (cdf) F . Let $X_{[u]}$ be the u^{th} smallest of these X_i such that $X_{[1]} \leq X_{[2]} \leq \dots \leq X_{[n]}$. Then $X_{[u]}, u = 1, 2, \dots, n$, is called the u^{th} order statistic of the random sample $X_i, i = 1, 2, \dots, n$. Note that even though the samples $X_1, X_2, \dots, X_n$ are independently and identically distributed (iid), the order statistics $X_{[1]}, X_{[2]}, \dots, X_{[n]}$ are not independent because of the order restriction. The difference $R = X_{[n]} - X_{[1]}$ is called the sample range. It is a measure of the dispersion in the sample and should reflect the dispersion in the population.

Suppose $U_1, U_2 \sim U(0, 1)$, where “ $\sim$ ” denotes “is distributed as” and “ $U(a, b)$ ” denotes a uniform distribution with range $[a, b]$. We are interested in the distribution of $U_{[2]} = \max(U_1, U_2)$, which can be viewed as the second order statistic. The cdf of $U_{[2]}$ is

$$\begin{aligned} F(x) &= P(\max(U_1, U_2) \leq x) \\ &= P(U_1 \leq x, U_2 \leq x) \\ &= P(U_1 \leq x)P(U_2 \leq x) \\ &= x^2 \end{aligned}$$

Furthermore, the cdf of $U_{[1]} = \min(U_1, U_2)$, which can be viewed as the first order statistic, is

$$\begin{aligned}
 F(x) &= 1 - P(\min(U_1, U_2) > x) \\
 &= 1 - P(U_1 > x)P(U_2 > x) \\
 &= 1 - (1 - x)^2
 \end{aligned}$$

Consider a device contains a series of two identical components and the lifetime of the component $L_c \sim U(0,1)$. If any of these two components fail, the device fails, such as chains. Then, the lifetime of the device $L_d \sim \min(L_{c1}, L_{c2})$. On the other hand, if these two components are configured in parallel and the device fails only when both components fail, such as a strand contains a bundle of threads. Then the lifetime of the device $L_d \sim \max(L_{c1}, L_{c2})$. These techniques can be used in reliability analysis to estimate the reliability of the component/system from the observed lifetime data.

When there are more than two samples, the distributions of sample minimum and sample maximum can be obtained via the same technique. However, the u^{th} ($u \neq 1$ and $u \neq n$) order statistic needs to be obtained via other techniques. Similarly, joint distributions of order statistics and conditional distributions of order statistics require more complicated derivation.

MAIN FOCUS

We discuss distributions of order statistics, joint and conditional distributions of order statistics, and empirical distribution functions and show how order statistics techniques are used in statistical analysis. We then present several procedures that are derived based on inference of order statistics, e.g., tests of independence, range statistics and hypothesis tests of equivalence of means, ranking and selection procedures, and quantile estimation.

Distributions of Order Statistics

Hogg et al. (2012) show that the distribution of the u^{th} order statistic of n samples of X (i.e., $X_{[u]}$) is

$$g_{u:n}(x) = \beta(F(x); u, n - u + 1) f(x),$$

where

$$\beta(x; a, b) = \frac{(a + b - 1)!}{(a - 1)!(b - 1)!} x^{a-1} (1 - x)^{b-1}.$$

That is,

$$g_{u:n}(x) = n! \frac{[F(x)]^{u-1} [1 - F(x)]^{n-u}}{(u-1)!(n-u)!} f(x).$$

Furthermore, the cdf

$$G_{u:n}(y) = \int_{-\infty}^y g_{u:n}(x) dx.$$

In particular, the first order statistic, or the sample minimum, $X_{[1]}$ has the pdf

$$g_{1:n}(x) = n[1 - F(x)]^{n-1} f(x)$$

and the cdf

$$G_{1:n}(x) = 1 - [1 - F(x)]^n.$$

The n^{th} order statistic, or the sample maximum, $X_{[n]}$ has the pdf

$$g_{n:n}(x) = n[F(x)]^{n-1} f(x)$$

and the cdf

$$G_{n:n}(x) = [F(x)]^n.$$

In the cast that f is the uniform $[0,1]$ distribution,

$$g_{u:n}(x) = \beta(x; u, n - u + 1).$$

That is, let $U = x \sim U(0, 1)$, then the random variate $y = \beta(U; u, n - u + 1)$ is the u^{th} order statistic of n random variables with uniform $[0, 1]$ as the parent distribution. From this we can deduct that

$$E[X_{[u]}] = \frac{u}{n+1},$$

and

$$\text{Var}[X_{[u]}] = \frac{u(n-u+1)}{(n+1)^2(n+2)}.$$

This result can be used in simulation to generate the u^{th} order statistic of any variates with a strictly increasing cdf. A straightforward way of simulating order statistics is to generate a pseudo-random sample from the distribution F and then sort the sample in ascending order. This general method requires sorting and can be avoided by using the process below. Based on the fact that $y = F(x) \sim U(0, 1)$, we generate a random variate $y = \beta(U; u, n - u + 1)$ and return the value $x = F^{-1}(y)$ as the u^{th} order statistic with parent distribution F . Here F^{-1} is the inverse distribution function (or quantile function) and is defined by

$$F^{-1}(y) = \inf \{x : F(x) \geq y\}.$$

In the case that F is uniform $[0, 1]$, $y = F^{-1}(y)$. For more information of using inverse transform to generate random variates, see (Law, 2014).

For any $0 < p < 1$, the $100p^{\text{th}}$ percentile (or p quantile) is the value such that about $100p\%$

of the observations are less than or equal to this value and $100(1-p)\%$ of the observations are larger. For example, the median is the 0.5 quantile, i.e., half of the observations are less than or equal to the median and one-half are greater. The median is a measure of location that is considered an alternative to the mean. Sample median (quantile) is a natural estimator of the population median (quantile) of the distribution and can be obtained from order statistics. The order statistics p quantile estimator is the u^{th} ($u = np$) order statistic, i.e., $X_{[u]}$, where z is the smallest integer that is greater than or equal to the real number z . Hence, np of the observations are less than or equal to $X_{[u]}$. One advantage of the sample median over the sample mean is that it is less sensitive to extreme observations. Note that if $U_1, U_2, \dots, U_n$ are iid $U(0, 1)$ random variables and $X_1, X_2, \dots, X_n$ are iid random variables with common distribution F such that $X_i = F^{-1}(U_i)$, then $X_{[i]} = F^{-1}(U_{[i]})$.

A properly selected set of quantiles can reveal all the essential distributional features of random variables being analyzed. Furthermore, in many situations, quantiles provide more useful information than the mean and variance. For example, to ensure the inventory is able to satisfy customers' demands 95% of the time, the stock should be the 0.95 quantile of the distribution of the demand; which can be estimated by the 0.95 sample quantile (i.e., the $0.95n^{\text{th}}$ order statistic of n observations).

Joint and Conditional Distributions of Order Statistics

The joint density function of $(X_{[u]}, X_{[v]})$, $1 \leq u < v \leq n$ is denoted by $g_{u,v:n}(x, y)$. It can be shown that

$$g_{u,v;n}(x,y) = n! \frac{[F(x)]^{u-1} [1-F(y)]^{n-v}}{(u-1)! (n-v)!} \frac{[F(y)-F(x)]^{v-u-1}}{(v-u-1)!} f(x)f(y).$$

In particular, the minimum and the maximum, $(X_{[1]}, X_{[n]})$, have the joint density

$$g_{1,n;n}(x,y) = n(n-1)[F(y)-F(x)]^{n-2} f(x)f(y).$$

From the joint distribution of two order statistics we can find the distribution of various other statistics, e.g., the sample range $R = X_{[n]} - X_{[1]}$. The pdf and cdf of the sample range R are, respectively,

$$g_R(R) = n(n-1) \int_{-\infty}^{\infty} \left[\frac{F(x+R)}{-F(x)} \right]^{n-2} f(x)f(x+R)dx$$

and

$$G_R(R) = n \int_{-\infty}^{\infty} \left[\frac{F(x+R)}{-F(x)} \right]^{n-1} f(x)dx.$$

The range statistics allow us to develop procedure to test the hypothesis of equivalence of means. Furthermore, the joint pdf of any number of the order statistics can be constructed, in particular, the joint pdf of all of the order statistics is

$$g_{1,2,\dots,n;n}(X_{[1]}, X_{[2]}, \dots, X_{[n]}) = n! f(X_{[1]})f(X_{[2]}) \dots f(X_{[n]}).$$

In the case that f is the uniform $[0,1]$ distribution,

$$g_{u,v;n}(x,y) = n! \frac{x^{u-1} [1-y]^{n-v} [y-x]^{v-u-1}}{(u-1)! (n-v)! (v-u-1)!}$$

and

$$\text{Cov}[X_{[u]}, X_{[v]}] = \frac{u(n-v+1)}{(n+1)^2(n+2)}.$$

As expected, $\text{Cov}[X_{[u]}, X_{[v]}]$ decreases as the difference between $v-u$ increases, given u or v . Moreover, $\text{Cov}[X_{[1]}, X_{[n]}] \rightarrow 0$, as $n \rightarrow \infty$. That is, in large samples the minimum and the maximum are in general approximately independent. Furthermore, the pdf and cdf of the sample range, respectively, are

$$g_R(R) = n(n-1)R^{n-2} \int_0^{1-R} dx = n(n-1)R^{n-2}(1-R) = \beta(R; n-1, 2),$$

and

$$G_R(R) = n(1-R)R^{n-1} + R^n.$$

From this we can deduce that

$$E[R] = \frac{n-1}{n+1}$$

and

$$\text{Var}[R] = \frac{2(n-1)}{(n+1)^2(n+2)}.$$

That is, the sample range R approaches the true range and variance of the sample range R becomes smaller as the sample size n becomes larger.

For $1 \leq u < v \leq n$, the conditional distribution of $X_{[v]}$ given $X_{[u]} = x_a$ is the same as the unconditional distribution of the $(v - u)^{\text{th}}$ order statistic in a sample of size $n - u$ from a new distribution, namely the original F truncated at the left at x_a . In notation,

$$\begin{aligned} g_{X_{[v]}|X_{[u]}=x_a}(x) &= \frac{(n-u)!}{(v-u-1)!(n-v)!} \left[\frac{F(x) - F(x_a)}{1 - F(x_a)} \right]^{v-u-1} \\ &\quad \left[\frac{1 - F(x)}{1 - F(x_a)} \right]^{n-v} \frac{f(x)}{1 - F(x_a)}, x > x_a. \end{aligned}$$

Similarly, for $1 \leq u < v \leq n$, the conditional distribution of $X_{[u]}$ given $X_{[v]} = x_b$ is the same as the unconditional distribution of the u^{th} order statistic in a sample of size $v - 1$ from a new distribution, namely the original F truncated at the right at x_b . In notation,

$$\begin{aligned} g_{X_{[u]}|X_{[v]}=x_b}(x) &= \frac{(v-1)!}{(u-1)!(v-1-u)!} \left[\frac{F(x)}{F(x_b)} \right]^{u-1} \\ &\quad \left[1 - \frac{F(x)}{F(x_b)} \right]^{v-1-u} \frac{f(x)}{F(x_b)}, x < x_b. \end{aligned}$$

For more information of conditional distribution of order statistics, see (DasGupta, 2011). These formula can be used to estimate the distribution of sample range R ; without storing and sorting the entire sequence.

Order Statistics from Correlated Normal Random Variables

Let $\Phi(x)$ and $\phi(x)$ be, respectively, the cdf and the pdf of a standardized normal random variable. In the case that $\{X_i\}$ for $i = 1, 2, \dots, n$, are standardized normal random variables with correlation matrix $\{\rho_{ij} = \rho \geq 0, \text{ for } i \neq j\}$, it can be shown (Gupta et al., 1973) that

$$\begin{aligned} F_u(h; \rho) &= \Pr[X_{[u]} \leq h] \\ &= \int_{-\infty}^{\infty} I_{q(x)}(u, n - u + 1) d\phi(x), \end{aligned}$$

where $q(x) = \Phi\left(\frac{x\sqrt{\rho} + h}{\sqrt{1 - \rho}}\right)$, and

$I_{q(x)}(a, b)$ ($a, b > 0$) is the usual incomplete beta function. When $u = n$ the equation is reduced to

$$\begin{aligned} F_n(h; \rho) &= \Pr[X_{[n]} \leq h] \\ &= \int_{-\infty}^{\infty} \Phi^n\left(\frac{x\sqrt{\rho} + h}{\sqrt{1 - \rho}}\right) \phi(x) dx. \end{aligned}$$

By the symmetry of the normal distribution, $\Pr[X_{[1]} \geq -h]$ is also equal to $F_n(h; \rho)$. When the correlation matrix has the structure $\rho_{ij} = c_i c_j$ ($i \neq j$), where $-1 \leq c_i \leq 1$ for $i = 1, 2, \dots, n$, then X_i can be represented as

$$X_i = c_i Z_0 + \sqrt{1 - c_i^2} Z_i, \text{ for } i = 1, 2, \dots, n,$$

where $Z_i, i = 0, 1, \dots, n$, are independent standard normal random variables. If $\rho_{ij} = \rho \geq 0$, then

the structural assumption that $\rho_{ij} = c_i c_j$ is satisfied by taking $c_i = \sqrt{\rho}$ for all i .

Normal distributions' pdf cannot be integrated in closed form, and hence tables of the cdf or computer programs are necessary to compute probabilities and quantiles for normal distributions. The complexity increases when variables are correlated. Gupta et al. (1973) point out that in several cases of applications, e.g., ranking and selection, the percentage points (quantile) of the distribution of $X_{[n]}$ are needed and provide tables of percentage points for several selected values of n and positive ρ . See Chen (2014) for more information.

Empirical Distribution Functions

In many cases, we want to use the observed data themselves to specify a distribution, called an empirical distribution, to approximate the true distribution. Consider an iid sequence $\{X_i\}$ for $i = 1, 2, \dots, n$, with distribution function F . One of the simplest and most important functions of order statistics is the sample cumulative distribution function F_n , which can be constructed by placing a mass $1/n$ at each observation X_i . Hence, F_n may be represented as

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I(X_i \leq x), -\infty < x < \infty.$$

Here $I(\cdot)$ is the indicator function. That is, the empirical cdf $F_n(x)$ for all real x , represents the proportion of sample values that do not exceed x . Namely,

$$F_n(x) = 0, x < x_{[1]},$$

$$F_n(x) = \frac{i}{n}, x_{[i]} \leq x < x_{[i+1]},$$

$$F_n(x) = 1, x_{[n]} \leq x.$$

For each fixed x , the strong law of large number implies that $F_n(x) \rightarrow F(x)$ asymptotically. Hence, we can get a good idea of what the true F is by studying F_n . Furthermore, natural estimators (obtained via empirical cdf) of estimating any characteristic properties of the distribution are appropriate and perform well.

In many cases, we are interested whether samples are drawn from the same distributions. It is very important in simulation to ensure that distributions of input parameters represent the underlying distributions correctly. The Q-Q (Quantile-Quantile) plot is a graphical evaluation of goodness of fit, it is the graphical representation of the points $(F^{-1}(p), x_{[np]})$, where population quantiles are recorded along the horizontal axis and the sample quantiles on the vertical axis. If the samples are in fact from F , we expect the Q-Q plot to be close to a straight line. If not, the plot may show nonlinearity at the upper or lower ends, which may be an indication of the presents of outliers. If the nonlinearity shows up at other points as well, one could question the validity of the assumption that the parent cdf is F . On the other hand, a goodness-of-fit test is a statistical hypothesis test that is used to assess formally whether the observations $X_1, X_2, \dots, X_n$ are independent samples from a particular distribution with distribution function F . The goodness-of-fit tests include Chi-square test, Kolmogorov-Smirnov test, and Anderson-Darling test (Law, 2014).

Furthermore, the density function $f = F'$. Hence, we can use the derivative of F_n to estimate f . We consider the central finite difference

$$f_n(x) = \frac{F_n(x+w) - F_n(x-w)}{2w}.$$

Here $2w$ is the width of bins (or intervals), which are disjointed categories. A graphical representation of $f_n(x)$ is called a histogram. See (Silverman, 1986) for more information on density estimation.

Test of Independence

The usual method of confidence interval construction from classical statistics, which assumes iid observations, is not directly applicable in many simulation procedures because simulation output data are generally correlated. Hence, it is important to determine whether the output data appear to be independent before applying random sampling theory. If the output data sequence appears to be dependent, special techniques need to be used to mitigate the effect of dependence, e.g., constructing confidence interval with batch means (Law, 2014). In dealing with this problem various tests of independence based on order statistics have been proposed, e.g., the runs-up and runs-down tests (Knuth, 1997). Briefly, a run up (run down) is a monotonically increasing (decreasing) subsequence. The tests compare the proportion of different run lengths to what would be expected under the independent null hypothesis in order to check whether the sequence appears to be independent.

The runs tests of independence require the underlying distributions of the random variables to be continuous. To adapt for discrete random variables, we will increase the run length with probability $1/(\tau+1)$ when two successive observations are equal, where τ is the current run length. Note that if there are $\tau+1$ observations, the possibility of the $(\tau+1)^{\text{th}}$ observation being the largest is $1/(\tau+1)$. If the distribution is continuous, the probability of a subsequence having run length τ is $\tau/(\tau+1)!$, under the null hypothesis that the sequence consists of iid random variables. We then use the Chi-square test on the observed frequency of each run length to

determine whether the sequence appears to be independent.

Lag l observations of sequence $X_i, i = 1, 2, \dots, n$, refers to the sequence $X_{(i-1)l+1}, i = 1, 2, \dots, n/l$. For example, lag 2 observations is the sequence $X_1, X_3, \dots, X_{(n/2-1)2+1}$.

Chen & Kelton (2010) use tests of independence to determine simulation run length. The procedure determines the lag l so that observations that at least $l-1$ observations apart will be essentially independent. The required simulation run length is then determined based on l . For example, if the required sample size for an iid sequence is n , then the required sample size for a correlated sequence is nl .

Indifference-Zone Selection

Discrete-event simulation has been widely used to compare alternative system designs or operating policies. When evaluating k alternative system designs, we select one or more system as the best and control the probability that the selected systems really are the best. Let CS denote the event of "correct selection." In a stochastic simulation, a CS can never be guaranteed with certainty. The probability of CS, denoted by $P(\text{CS})$, is a random variable depending on sample sizes and other uncontrollable factors. Moreover, in practice, if the difference between μ_{i_1} and μ_{i_2} is very small, we might not care if we mistakenly choose system i_2 , whose expected response is μ_{i_2} . The practically significant difference d^* , (a positive real number) between a desired and a satisfied system is called the indifference zone in statistical literature, and it represents the smallest difference which we care about. Therefore, we want a procedure that avoids making a large number of replications or batches to resolve differences less than d^* . That means we want $P(\text{CS}) \geq P^*$ provided that $\mu_{i_2} - \mu_{i_1} \geq d^*$, where the minimal CS probabil-

ity P^* and the “indifference” amount d^* are both specified by the users.

The indifference-zone selection procedure of Dudewicz and Dalal (1975) to select the best of k systems proceeds as follows.

1. Simulate the initial samples for all systems.
Compute the first-stage sample means with the equation and the sample variances with the equation

$$S_i^2(n_0) = \frac{1}{n_0 - 1} \sum_{j=1}^{n_0} \left(X_{ij} - \bar{X}_i^{(1)}(n_0) \right)^2$$

for $i = 1, 2, \dots, k$.

2. Compute the required sample sizes with the equation

$$N_i = \max \left(n_0 + 1, \left(h_1 S_i(n_0) / d^* \right)^2 \right), \text{ for } i = 1, 2, \dots, k.$$

Here h_1 (a positive real number that depends on k , P^* , and n_0) is a critical constant that will be described later.

3. Simulate additional $N_i - n_0$ samples, for $i = 1, 2, \dots, k$.
4. Compute the second-stage sample means with the equation

$$\bar{X}_i^{(2)}(N_i - n_0) = \frac{1}{N_i - n_0} \sum_{j=n_0+1}^{N_i} X_{ij},$$

for $i = 1, 2, \dots, k$.

5. Compute the weighted sample means with the equation

$$\widetilde{X}_i(N_i) = W_{i1} \bar{X}_i^{(1)}(n_0) + W_{i2} \bar{X}_i^{(2)}(N_i - n_0),$$

for $i = 1, 2, \dots, k$

and select the system with the smallest $\widetilde{X}_i(N_i)$.
The weights can be derived from the equations

$$W_{i1} = \frac{n_0}{N_i} \left[1 + \sqrt{1 - \frac{N_i}{n_0} \left(1 - \frac{(N_i - N_0)(d^*)^2}{h_1^2 S_i^2(n_0)} \right)} \right]$$

and $W_{i2} = 1 - W_{i1}$, for $i = 1, 2, \dots, k$. The expression for W_{i1} was chosen to guarantee that the variable $\left(\widetilde{X}_i(N_i) - \mu_i \right) / (d^* / h_1)$ would have a t distribution with $n_0 - 1$ df; see Dudewicz and Dalal (1975).

Let T_i for $i = 1, 2, \dots, k$, be independent t -distributed random variables with $n_0 - 1$ df and let F denote the cdf of the t distribution with $n_0 - 1$ df. Let $\tilde{X}_{c_u} = \min_{2 \leq i \leq k} \tilde{X}_{i1}$, under the null hypothesis that $\mu_{i1} + d^* = \mu_{i2} = \dots = \mu_{ik}$, we can write

$$\begin{aligned} P(CS) &= P \left[\tilde{X}_{i1} < \tilde{X}_{c_u} \right] \\ &= P \left[\frac{\tilde{X}_{i1} - \mu_{i1}}{d^* / h_1} < \frac{\tilde{X}_{c_u} - \mu_{c_u}}{d^* / h_1} + \frac{\mu_{c_u} - \mu_{i1}}{d^* / h_1} \right] \\ &= P \left[T_{i1} < T_{c_u} + h_1 \right] \\ &= \int_{-\infty}^{\infty} F(t + h_1) g_{1:k-1}(t) dt. \end{aligned}$$

We equate the right-hand side to P^* and solve for the critical constant h_1 . Note that T_{c_u} is the first order statistic of $k - 1$ random variate T .

Generalized Subset Selection

Mahamunulu (1967) considers a generalized version of the selection problem. The goal is to select a subset of size m contains at least c of the v best of k systems, where $\max(I, m + v + I - k) \leq c \leq \min(m, v)$ and $\max(m, v) \leq k - 1$. If $m = v = 1$,

then the problem is to choose the best system. When $m > v = 1$, we are interested in choosing a subset of size m contains the best. If $m = v > 1$, we are interested in choosing the m best systems. We now assess the P(CS) of this subset selection problem. The LFC (Least Favorable Configuration) of this problem is $\mu_{i_1} = \mu_{i_2} = \dots = \mu_{i_v}$ and $\mu_{i_v} + d^* = \mu_{i_{v+1}} = \dots = \mu_{i_k}$. Indifference-zone selection procedures determine the number of additional replications based on the LFC because the minimal P(CS) occurs in this setting. We derive the P(CS) of subset selection with order statistics. Let $\tilde{X}_c$ be the c^{th} smallest weighted sample mean from system i_l for $l = 1, 2, \dots, v$ and let μ_c be its unknown true mean. Let $\tilde{X}_u$ ($u = m - c + 1$) be the smallest weighted sample mean from system i_l for $l = v + 1, v + 2, \dots, k$ and let μ_u be its unknown true mean. Then correct selection occurs if and only if $\tilde{X}_c < \tilde{X}_u$. We would like to point out that systems c and u are unknown. However, $\tilde{X}_c \sim g_{c:v}(t)$ and $\tilde{X}_u \sim g_{u:k-v}(t)$. The P(CS) of subset selection can be written as

$$\begin{aligned} P(CS) &= P\left[\tilde{X}_c < \tilde{X}_u\right] \\ &= P\left[T_c < T_u + \frac{\mu_u - \mu_c}{d^* / h}\right] \\ &\geq P\left[T_c < T_u + h\right]. \end{aligned}$$

Here h is a critical constant. The inequality follows because $\mu_u - \mu_c \geq \mu_{i_{v+1}} - \mu_{i_v} \geq d^*$. Furthermore, if $d_{cu} = \mu_u - \mu_c$ is used instead of d^* in the above equations, then we obtain strict equality. For more information on subset selection, see (Chen, 2009). Note that the value h is determined such that $P\left[T_c < T_u + h\right] = P^*$. Let $\gamma = T_c - T_u$ then $P\left[\gamma < h\right] = P^*$. That is, under the LFC the value of the critical constant h is the P^* quantile of the distribution of γ . We can use this property to estimate the value of h for the problem at hand.

Hypothesis Tests of Equivalence of Means

This section investigates performing the hypothesis testing of equivalence of mean of multiple systems. The most useful comparison of the system means is done by comparing each system with the unknown best system or best systems within the group. Consider sample means of certain observations drawn at random from k systems. The sample mean of system i with N_i samples

$$\bar{X}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} X_{ij}.$$

Let $\bar{X}_j$, $j = 1, 2, \dots, k$ be another set of sample means that are independent of $\bar{X}_i$ and let

$$\bar{X}_o = \frac{1}{k} \sum_{j=1}^k \bar{X}_j \text{ and } S^2 = \frac{1}{k-1} \sum_{j=1}^k (\bar{X}_j - \bar{X}_o)^2.$$

The range of sample means $W = \bar{X}_{[k]} - \bar{X}_{[1]}$ and the ratio, $q = W/S$, of the sample range to an independent root-mean-square estimate, S , of the population standard deviation, σ , is called a studentized range. Tukey's range test to find which means are significantly different from one another is based on the distribution of q (Snedecor and Cochran, 1989). Note that Tukey's range test assumes equal variances and uses equal sample sizes for each system.

Chen (2011) considers the equivalence test with critical region $R = \left\{ \bar{X}_{[k]} - \bar{X}_{[1]} \leq \delta \right\}$, where $\delta > 0$ is a user-specified upper bound of the sample range for the hypothesis test. Note that in our cases variances may be unequal among systems and the studentized range can no longer be computed in a straightforward manner.

Let $\tilde{X}_i$ (i.e., $\tilde{X}_i(N_i)$) be the weighted sample means and let

$$T_i = \frac{\tilde{X}_i - \mu_i}{\delta / h_3}.$$

Here h_3 is a critical constant. Note that the weight needs to be computed with d replaced by δ and the required sample sizes N_i need to be computed by equation (1), which will be described later. Moreover, T_i 's are independent t -distributed random variables with $n_0 - 1$ degrees of freedom (df). Let F denote the cdf of the t distribution with $n_0 - 1$ df. Let μ_{i_l} be the l^{th} smallest of the μ_i 's, so that $\mu_{i_1} \leq \mu_{i_2} \leq \dots \leq \mu_{i_k}$ and let $\tilde{X}_{c_l}$ denote the l^{th} smallest weighted sample mean such that $\tilde{X}_{c_1} \leq \tilde{X}_{c_2} \leq \dots \leq \tilde{X}_{c_k}$. Under the null hypothesis that $\mu_{i_1} = \mu_{i_2} = \dots = \mu_{i_k}$, we can write the probability of correct decision

$$\begin{aligned} P(\text{CD}) &= P\left[\tilde{X}_{c_k} - \tilde{X}_{c_1} \leq \delta\right] \\ &= P\left[\frac{\tilde{X}_{c_k} - \mu_{c_k}}{\delta / h_3} - \frac{\tilde{X}_{c_1} - \mu_{c_1}}{\delta / h_3} \leq \frac{\delta}{\delta / h_3}\right] \\ &= P\left[T_{c_k} - T_{c_1} \leq h_3\right] \\ &= G_R(h_3). \end{aligned}$$

We then equate the right hand side to P^* and solve for h_3 . Numerical approximation has been used to solve the integration. Nevertheless, other approach can be used to solve h_3 , which will be discussed later.

The steps of testing the null hypothesis that all means are equal are as follows.

1. Simulate the initial samples for all systems. Compute the first-stage sample means with the equation and the sample variances with the equation

$$S_i^2(n_0) = \frac{1}{n_0 - 1} \sum_{j=1}^{n_0} \left(X_{ij} - \bar{X}_i^{(1)}(n_0)\right)^2$$

for $i = 1, 2, \dots, k$.

2. Compute the required sample sizes with the formula

$$N_i = \max\left\{n_0 + 1, \left(h_3 S_i(n_0) / \delta\right)^2\right\}, \text{ for } i = 1, 2, \dots, k. \quad (1)$$

3. Simulate additional $N_i - n_0$ samples.
4. Compute the second-stage sample means with the formula

$$\bar{X}_i^{(2)}(N_i - n_0) = \frac{1}{N_i - n_0} \sum_{j=n_0+1}^{N_i} X_{ij}, \text{ for } i = 1, 2, \dots, k.$$

5. Compute the weighted sample means $\tilde{X}_i$ for $i = 1, 2, \dots, k$. Reject the null hypothesis when

$$\tilde{X}_{c_k} - \tilde{X}_{c_1} > \delta;$$

otherwise do not reject the null hypothesis.

Estimating the Critical Constant

In this section, we present an approach to estimate the critical constant for selection procedures. Let variables T_{i_l} for $l = 1, 2, \dots, k$ be independently and identically t -distributed with $n_0 - 1$ df. The selection procedure of Dudewicz and Dalal (1975) is derived from the equation

$$\begin{aligned} P(CS) &\geq P\left[T_{i_1} < T_{i_l} + h_1, l = 2, 3, \dots, k\right] \\ &= \int_{-\infty}^{\infty} F^{k-1}(t + h_1) f(t) dt, \end{aligned}$$

where f and F are the pdf and cdf of the t distribution with $n_0 - 1$ df, respectively. The equality holds exactly under the LFC. We equal the right-

hand side to P to solve for h_1 . Recall that $T_{c_u} = \min_{2 \leq l \leq k} T_{i_l}$, hence, $P[T_{i_1} - T_{c_u} < h_1] = P^*$. That is, the critical constant h_1 is the P^* quantile of the variable $T_{i_1} - T_{c_u}$. Similarly, h_3 is the P^* quantile of the variable $T_{c_k} - T_{c_1}$ and h is the P^* quantile of the variable $T_c - T_u$.

Chen & Kelton (2008) develop a quantile-estimation procedure that controls the precision of quantile estimates by ensuring that the p quantile estimator $\hat{x}_p$ satisfies the following:

$$P[x_p \in \hat{x}_{p \pm \varepsilon}] \geq 1 - 2\alpha_1,$$

or equivalently $P[|F(\hat{x}_p) - p| \leq \varepsilon] \geq 1 - 2\alpha_1.$ (2)

Using this precision requirement (i.e., (2)), the required sample size np for a fixed-sample-size procedure of estimating the p quantile of an iid sequence is the minimum np that satisfies

$$np \geq \frac{z_{1-\alpha_1}^2 p(1-p)}{\varepsilon^2},$$

where $z_{1-\alpha_1}$ is the $1 - \alpha_1$ quantile of the standard normal distribution, ε is the maximum proportional half-width of the confidence interval, and $1 - 2\alpha_1$ is the confidence level. For example, if the data are independent and one wants to have 95% confidence that the coverage of the 0.9 quantile estimator has no more than $\varepsilon = 0.0001$ deviation from the true but unknown quantile, the required sample size is $np \geq 34574400$ (i.e., $1.960^2 0.9(1 - 0.9)/0.0001^2$). Consequently, we will have 97.5% confidence that the quantile estimate will cover at least $p - 0.0001$ (for $p \geq 0.9$), with sample size 34574400.

Quantile estimates obtained through order statistics directly (i.e., sample quantiles) required storing and sorting the entire sequence, which is computationally intensive. To avoid storing and sorting the entire sequence, a modified version of

the histogram- approximation procedure of Chen & Kelton (2008) was used to estimate quantiles.

FUTURE TRENDS

Order statistics continue to be key components of any data analysis. Algorithms developed based on order statistics will be incorporated into many statistical packages and will be readily available for practitioners.

CONCLUSION

Order statistics have proved useful for both theoretical and computational purposes. Important special cases of order statistics are the minimum and maximum value of samples, and the sample median and other sample quantiles. We have investigated several properties of order statistics and discussed how order statistics are applied in statistical analysis, e.g., generate the distributions of order statistics without sorting, build empirical distributions, and validate empirical distributions with goodness-of-fit tests. We also presented several procedures that are derived based on the inference of order statistics, e.g., hypothesis tests of equivalence of means, ranking and selection, and quantile estimation. Properties of order statistics allow us to derive many algorithms in a concise manner. These procedures are easy to understand and simple to implement.

REFERENCES

- Arnold, B. C., Balakrishnan, N., & Nagaraja, H. N. (2008). *A first course in order statistics*. SIAM-Society for Industrial and Applied Mathematics. doi:10.1137/1.9780898719062
- Chen, E. J. (2009). Subset selection procedures. *Journal of Simulation*, 3(4), 202–210. doi:10.1057/jos.2009.1

Chen, E. J. (2011). The range statistic and procedures to select the best systems. *Journal of Simulation*, 5(4), 292–300. doi:10.1057/jos.2011.6

Chen, E. J. (2014). Selection and order statistics from correlated normal random variables. *Discrete Event Dynamic Systems*, 24(4), 659–668. doi:10.1007/s10626-013-0180-4

Chen, E. J., & Kelton, W. D. (2008). Estimating steady-state distribution via simulation-generated histograms. *Computers & Operations Research*, 35(4), 1003–1016. doi:10.1016/j.cor.2006.05.015

Chen, E. J., & Kelton, W. D. (2010). Confidence-interval estimation using quasi-independent Sequences. *IIE Transactions*, 42(1), 83–93. doi:10.1080/07408170903232266

DasGupta, A. (2011). *Probability for statistics and machine learning*. New York: Springer. doi:10.1007/978-1-4419-9634-3

David, H. A., & Nagaraja, H. N. (2003). *Order statistics* (3rd ed.). New York: Wiley. doi:10.1002/0471722162

Dudewicz, E. J., & Dalal, S. R. (1975). Allocation of observations in ranking and selection with unequal variances. *Sankhya*, B37, 28–78.

Gupta, S. S., Nagel, K., & Panchapakesan, S. (1973). On the order statistics from equally correlated normal random variables. *Biometrika*, 60(2), 403–413. doi:10.1093/biomet/60.2.403

Hogg, R. V., McKean, J., & Craig, A. T. (2012). *Introduction to mathematical statistics* (7th ed.). Pearson Education.

Knuth, D. E. (1997). *The art of computer programming* (3rd ed.). Reading, Mass.: Addison-Wesley.

Law, A. M. (2014). *Simulation modeling and analysis* (5th ed.). New York: McGraw-Hill.

Mahamunulu, D. M. (1967). Some fixed-sample ranking and selection problems. *Annals of Mathematical Statistics*, 38(4), 1079–1091. doi:10.1214/aoms/1177698778

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. New York: Chapman and Hall. doi:10.1007/978-1-4899-3324-9

Snedecor, G. W., & Cochran, W. G. (1989). *Statistical methods* (8th ed.). Iowa State University Press.

ADDITIONAL READING

Ahsanullah, M., Nevzorov, V. B., & Shakil, M. (2013). *An Introduction to Order Statistics*. Atlantis Press. doi:10.2991/978-94-91216-83-1

Billingsley, P. (2012). *Probability and measure*. Anniversary ed. New York: Wiley.

Chen, C. H., & Lee, L. H. (2010). *Stochastic simulation optimization: An optimal computing budget allocation*. World Scientific Publishing. doi:10.1142/7437

Chen, E. J. (2016). *Crafts of Simulation Programming*. World Scientific Publishing. doi:10.1142/9899

Chen, E. J., & Lee, L. H. (2009). A multi-objective selection procedure of determining a pareto set. *Computers & Operations Research*, 36(6), 1872–1879. doi:10.1016/j.cor.2008.06.003

Nagaraja, H. N., Sen, P. K., & Morrison, D. F. (2012). *Statistical theory and applications: Papers in honor of Herbert A. David*. New York: Springer.

Shi, Y., Xiong, J., Zolotov, V., & Visweswariah, C. (2013). Order Statistics for Correlated Random Variables and Its Application to At-Speed Testing. *ACM Transactions on Design Automation of Electronic Systems*, 18(3), 42. doi:10.1145/2491477.2491486

KEY TERMS AND DEFINITIONS

Critical Constant: In ranking and selection, the critical constant is a required parameter for computing the required sample sizes and is the P^* quantile of the difference of two specific random values.

Empirical Distribution Function: The cumulative distribution function (cdf) associated with the empirical observations of the sample. The empirical distribution function estimates the true underlying cdf of the sample.

Histogram: A graphical representation showing the distribution of data.

Indifference Zone: We don't distinguish values that are deviated less than a significant amount (the indifference amount), i.e., they are within in the indifference zone.

Order Statistics: The collection of sample observations sorted in ascending order.

Q-Q Plot: A probability plot for comparing two probability distributions by plotting their quantiles against each other.

Quantile Function (Inverse Distribution Function): The quantile function (given a probability p) returns the value at or below which $100p$ percent of the population lies.

Range Statistics: The statistics of the difference between the maximum and minimum observations. Range statistics give an estimate of the spread of the data.

Tests of Independence: Statistical null hypothesis tests to determine whether sequences of observations appear to be independent.

Recommender Technologies and Emerging Applications

D**Young Park**

Bradley University, USA

INTRODUCTION

A *recommender system* (or *recommendation system*) is a software application that provides users with personalized recommendations of goods, services or other potentially relevant and interesting information, and thus helps users find useful items in the information overload (Aggarwal, 2016; Resnick & Varian, 1997; Ricci, Rokach, & Shapira, 2015). The field of recommender systems is highly interdisciplinary and based on various technologies. Though relatively new, recommender technologies have made significant progress for the last decade.

A variety of recommender systems have been developed and used mainly in e-commerce application domains, including Amazon.com, BarnesandNoble.com, Netflix.com, mystrands.com, and Yahoo.com (Konstan et al., 1997; Sarwar, Karypis, Konstan, & Riedl, 2000; Schafer, Konstan, & Riedl, 2001). Over the last decade, recommender systems have proven very useful in increasing sales and retaining consumers, and are considered an effective personalization tool in the e-commerce environment (Adomavicius & Tuzhilin, 2005; Goy, Ardissono, & Petrone, 2007; Jannach, Zanker, Felfernig, & Friedrich, 2011; Ricci et al., 2015; Sarwar et al., 2000; Schafer et al., 2001;). One illustration is the famous Netflix competition (2006-2009), which offered a one million dollar prize in exchange for an algorithm to enhance the recommendation accuracy (i.e., movie rating prediction) of its recommender systems (Bell, Koren, & Volinsky, 2010).

This chapter presents an overview of the field of recommender technologies and their emerging

application domains. We characterize current major recommender system approaches in a unifying model and describe emerging applications of recommender technologies beyond traditional e-commerce. We conclude with emerging and future trends and topics, as well as additional readings in the area of recommender technologies and applications.

BACKGROUND

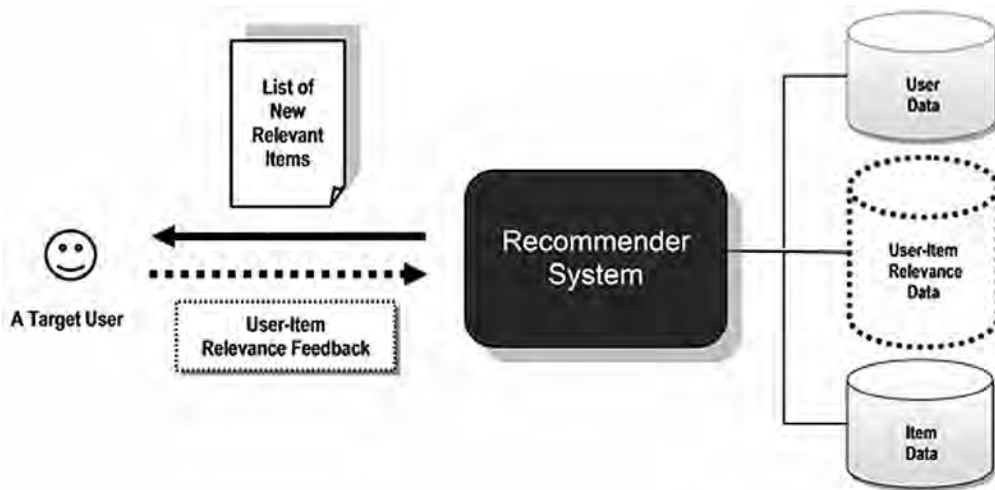
Since the first major recommender systems emerged in the mid-1990s (Resnick & Varian, 1997), a large number of recommender systems have been developed and used in a wide range of e-commerce environments and improved by continuing research.

A typical recommender system provides users with *personalized* recommendations of items such as goods, services or information to guide users to find items that are relevant to them. Recommendations are based on past and present profiles of users with respect to items. The *personalized recommendation problem* can be described as follows:

Given a target user, produce personalized recommendations of items relevant to the target user.

To solve this recommendation problem, a recommender system generally uses three types of data—data about the users (*U_data*), data about the items such as goods, services or information (*I_data*), and data about the relevance (such as rating, evaluation, purchase, or interest) relation between the users and the items (*R_data*):

Figure 1. A model for recommender systems



- U_data contains a set of all users and some optional additional information about all users.
- I_data contains a set of all items and some optional additional information about all items.
- R_data contains relevance information (such as rating, evaluation, purchase, or interest) of the item to the user. Such information can be viewed and represented as a partial function (or matrix) that maps a pair of user and item into a relevance value.

As shown in Figure 1, a recommender system can be modeled as follows:

Given a target user, a recommender system recommends a list of new items that are most relevant to the target user by using U_data , I_data , and R_data .

The user data is constructed for all existing users and the item data is constructed for all existing items. The user-item relevance data is constructed over time and can be obtained either explicitly from user participation (*explicit relevance feedback*), inferred implicitly from user behavior (*implicit relevance feedback*), or both.

In order to find items that are relevant to the target user, recommender systems are premised on the *similarity* between users or between items. A number of similarity metrics are used to represent the degree of similarity between users or items. *Pearson Correlation* and *Cosine Similarity* are among most widely used metrics in recommender systems.

The recommendation process typically consists of two phases – the prediction phase and the suggestion phase. First, the recommended list of most relevant and new items for the target user is determined by predicting the unknown relevance values of new items by the target user. The predicted relevance value of a new item is estimated from the relevance values of the item from some of the most similar users (i.e., nearest user neighbors) or the relevance values of some of the most similar items (i.e., nearest item neighbors) using some prediction algorithm such as the weighted sum. Second, the recommender system typically recommends a list of top n items with the highest predicted relevance values.

The quality of recommender systems is evaluated in many ways (Herlocker, Konstan, Terveen, & Riedl, 2004). The accuracy of the recommended items is the most frequently evaluated measure. Popular accuracy metrics include *mean absolute error* (MAE, or the average of absolute differences

between real feedback values and predicted feedback values), *mean square error* (MSE), *root mean squared error* (RMSE), *precision* (the number of relevant items recommended divided by the total number of recommended items), and *recall* (the number of relevant items recommended divided by the total number of existing relevant items).

MAJOR RECOMMENDER TECHNOLOGIES

Various recommendation technologies have been used to develop more effective and efficient recommender systems in a range of application domains (Adomavicius & Tuzhilin, 2005). Each approach has strengths and weaknesses. Major recommender system approaches can be classified as follows: *content-based recommender*, *collaborative recommender* and *hybrid recommender*. Other approaches include knowledge-based recommender and demographic recommender.

The major recommender approaches (Figure 2) can be characterized within a single unifying model (Figure 1). Each approach uses three kinds of data — U_data , I_data and R_data . However, each approach differs in

- The information contained in the data, and
- How the three kinds of data are used to produce the most relevant items for a given target user.

Content-Based Recommender

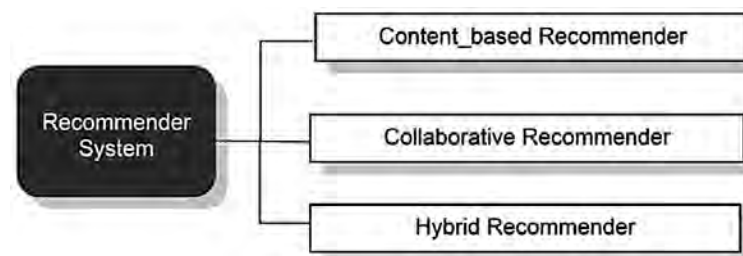
The content-based recommendation approach, which is based on the *content* of items, underlies a number of recommender systems (Pazzani & Billsus, 2007; Lops, Gemmis, & Semeraro, 2011). This system basically recommends a list of items with similar content to the items that were given good relevance feedback by the target user. The pure content-based recommender system uses

- U_data = A set of all users.
- I_data = A set of all items and a partial function that maps an item-feature pair (i.e., item, feature) into a content description value. (This function contains information about items and features describing the content of the items.)
- R_data = A partial function that maps a user-item pair into a relevance value.

Similarity between items with respect to content description features can be measured using various similarity metrics. Given a target user, the recommendation process of the content-based approach has three main steps:

- Step 1:** Find items for which the target user provided good relevance feedback using R_data .
- Step 2:** Select items with content similar to that of the items found in Step 1 using I_data .
- Step 3:** Recommend a list of items from the items selected in Step 2 that are most relevant and new to the target user.

Figure 2. Major recommender approaches



Limitations of Content-Based Recommender

The effectiveness of content-based recommender systems depends on how well the items are described by features, and thus is often limited. These systems tend to recommend items that are similar to those already recommended (called *the item overspecialization problem*), and cannot recommend items for new users until they provide enough relevance feedback (called *the new user problem*). However, content-based recommender systems can recommend new items.

Collaborative Recommender

The collaborative recommendation approach recommends items through *user collaborations*. Numerous recommender systems using the collaborative recommendation approach have been proposed, and this approach is the most widely used and proven method of providing recommendations (Herlocker et al., 2004; Schafer, Frankowski, Herlocker, & Sen, 2007). There are two types of *neighborhood-based* collaborative recommender systems (Sarwar, Karypis, Konstan, & Riedl, 2001; Linden, Smith, & York, 2003):

- *User-to-user collaborative recommender*
- *Item-to-item collaborative recommender*

User-to-User Collaborative Recommender

The user-to-user collaborative filtering approach is based on user-to-user similarity. It recommends a list of items given relevance feedback by other users similar to that provided by the target user. The user-to-user collaborative recommender system uses

- U_data = A set of all users.
- I_data = A set of all items.
- R_data = A partial function that maps a user-item pair into a relevance feedback value.

Similarity between two users is based on relevance feedback for all items provided by both users. Different similarity metrics can be used to describe user similarity. Given a target user, the recommendation process of the user-to-user collaborative recommender system has three main steps:

- Step 1:** Find consumers who are similar to the target consumer using R_data .
- Step 2:** Select items given good feedback by the similar consumers found in Step 1 using R_data .
- Step 3:** Recommend a list of items from the items selected in Step 2 that are most relevant and new to the target user.

Item-to-Item Collaborative Recommender

The item-to-item collaborative filtering approach is based on item-to-item similarity. It recommends a list of items that are similar to the items that were given good relevance feedback by the target user. The item-to-item collaborative recommender system uses

- U_data = A set of all users.
- I_data = A set of all items & a partial function that maps an item-item pair (i.e., item and item) into a similarity value. (This function contains information about similarity between items.)
- R_data = A partial function that maps a user-item pair into a relevance feedback value.

Similarity between two items is based on relevance feedback received from all users concerning both items. Note, however, that the content-based approach uses similarity between items with respect to content description features. The similarity between items with respect to users is measured using different similarity metrics. The partial function that maps an item-item pair to a similarity value can be constructed using

R_data in the offline stage in advance. Given a target user, the recommendation process of the item-to-item collaborative recommender system has three main steps.

Step 1: Find items given good feedback by the target user using R_data .

Step 2: Select items that are similar to the items found in Step 1 using I_data .

Step 3: Recommend a list of items from the items selected in Step 2 that are most relevant and new to the target user.

Comparison of Collaborative Recommender Approaches

While user-to-user collaborative filtering is based on similarity between users, item-to-item collaborative filtering is based on similarity between items. Item-to-item collaborative filtering is more algorithmically efficient and has shown better recommendation accuracy than user-to-user collaborative filtering.

Item-to-item collaborative filtering is an example of the *model-based* method. The model-based method builds a model of the user-item relevance feedback data offline then uses the model to generate items for recommendation. Amazon's product recommender system is an item-to-item collaborative recommender (Sarwar et al., 2001). In contrast, user-to-user collaborative filtering and the content-based method are examples of the *memory-based* method. The memory-based method memorizes and then uses the entire user-item relevance feedback data to find items for recommendation. The model-based method shows better scalability than the memory-based method..

Latent Factor Model-Based Collaborative Recommender

The latent factor model-based collaborative filtering approach is a model-based approach based on

the concept of matrix factorization (decomposition) of relevancy feedback data (R_data). This approach tends to show high recommendation accuracy in various applications, and a number of matrix factorization techniques are widely used in collaborative recommenders (Koren, Bell, & Volinsky, 2009). Models based on matrix factorization were used in the movie recommendation algorithm by the team who won the Netflix prize, BellKor's Pragmatic Chaos.

Limitations of Collaborative Recommender

The recommendation quality of collaborative filtering-based approaches depends on user feedback on items. But the available feedback is usually smaller compared with all possible feedback (called *the relevance feedback sparsity problem*). Collaborative filtering-based approaches also suffer from the new user problem and cannot recommend new items until they have enough feedback from users (called *the new item problem*).

Hybrid Recommender

In order to overcome the limitations of each approach and improve accuracy through synergy, the hybrid recommendation approach recommends items by combining two or more approaches, including the two major approaches (content-based and collaborative filtering-based) and the knowledge-based approach (Burke, 2007). Netflix's movie recommender system is a hybrid recommender that combines the content-based approach and the collaborative approach.

The combination (hybridization) can be done in many different ways - monolithic hybridization, parallel hybridization and pipelined hybridization. Many recommender systems adopt the hybrid approach and tend to result in better recommendations.

TRADITIONAL E-COMMERCE RECOMMENDER APPLICATIONS

With the explosive growth goods, services and information available on the Web under the e-commerce umbrella (in either commercial or non-commercial contexts), it has become increasingly difficult for consumers to find what they are searching for in a timely manner. By providing consumers with personalized recommendations, recommender systems can help consumers find relevant and interesting products in the age of information overload.

The traditional application domain of recommender technologies is e-commerce (Schafer et al., 2001; Lu, Wu, Mao, Wang, & Zhang, 2015; Ricci et al., 2015). In the e-commerce environment, the users are consumers and the recommendable items are products including a wide variety of goods, services or information. E-commerce recommender systems provide consumers with personalized recommendations of relevant products, including movies, videos, news, songs, books, images, blogs, games, restaurants, television programs, people, documents, jokes, places, hotels, travel destinations, accommodations, real estates, websites, and tags.

Personalized recommendations can be viewed as part of the personalization activity of e-commerce applications. Recommender technologies have become and will continue to be an important tool for e-commerce personalization.

EMERGING RECOMMENDER APPLICATIONS

Personalized recommender technologies have increasingly been applied to various other domains beyond e-commerce. Notable emerging application domains, discussed below, include

- **Education:** Learning, teaching and advising.

- **Engineering:** Software engineering, requirements engineering and data/knowledge engineering.

Recommender Applications in Education

One notable emerging application domain is that of education, namely, learning (including e-learning), teaching and academic advising. In the educational context, the users are students, teachers or academic advisors, and the recommendable items are products such as educational goods, services or information. Educational recommender systems can guide students, teachers or advisors by suggesting personalized recommendations of relevant educational goods, services or information during the learning, teaching or academic advising process.

A significant number of educational recommender systems, involving courses, learning materials, topics and grades, have been proposed and deployed in this application domain (Manouselis, Drachsler, Verbert, & Duval, 2013; Manouselis, Drachsler, Verbert, & Santos, 2014).

Recommender Applications in Engineering

Another notable emerging application domain is the technical domain of engineering, including software engineering, requirements engineering and data/knowledge engineering. Engineering processes involve many complex and knowledge-intensive activities. In the technical engineering environment, the users are engineers and the recommendable items are products such as engineering goods, services or information. Engineering recommender systems can assist engineers with personalized recommendations of relevant engineering goods, services or information during the complex engineering process.

Various software engineering recommender systems have been presented for products such as process, codes, issue reports, tools, commands and

operations (Robillard, Walker, & Zimmermann, 2010; Robillard, Maalej, Walker, & Zimmermann, 2014).

In the area of requirements engineering, various recommender systems have been proposed for products including stakeholders, features and requirements (Mobasher & Cleland-Huang, 2011; Felfernig et al., 2013).

An interesting approach using small data with contexts and heuristics called “recommendation in-the-small” was recently introduced and used in software engineering recommendations. (Inozemtseva, Holmes, & Walker, 2014).

A data/knowledge engineering recommender system with relevant constraints as the recommendable knowledge engineering product has been proposed (Felfernig et al., 2013; Felfernig, Jeran, Ninaus, Reinfrank, & Reiterer, 2013).

Recommender Applications in Other Domains

Healthcare is also a promising field for recommender systems. In the healthcare setting, the users are patients and healthcare professionals, and the recommendable items are products such as healthcare-related goods, services or information. Health recommender systems could help patients and healthcare professionals with personalized recommendations of relevant healthcare-related goods, services or information. Health and nutrition recommender systems have been discussed (Wiesner & Pfeifer, 2014; Mika, 2011).

Other proposed emerging application domains include persuasive systems and smart homes (Felfernig et al., 2013).

Recommender technologies will continue to be applied successfully in a variety of novel and interesting application domains.

FUTURE RESEARCH DIRECTIONS

Researchers have proposed a number of extensions to current recommender technologies and outlined

future challenges (Adomavicius & Tuzhilin, 2005; Felfernig, Friedrich, & Schmidt-Thieme, 2007; Felfernig et al., 2013).

Important future trends and research topics include:

- Group recommendations,
- Context-aware recommendations,
- Ubiquitous, mobile and location-aware recommendations,
- Multi-criteria recommendations,
- Cross-domain recommendations,
- Explanations of recommendations,
- Novelty, diversity and serendipity in recommendations,
- Privacy, security, and trust in recommendations,
- New recommendation application domains.

Leading researchers and companies in the field of recommender technologies, and their recent progress, can be found at conferences and in journals. The major conferences are the premier series of Association for Computing Machinery (ACM) Conference on Recommender Systems (RecSys), SIGIR Informational Retrieval, SIGKDD Knowledge Discovery and Data Mining, International Conference on World Wide Web (WWW), and International Conference on Intelligent User Interfaces (IUI). The major journals include ACM Transaction on Information Systems (TOIS), IEEE Intelligent Systems, Expert Systems with Applications (ESWA) and Decision Support Systems (DSS).

Future research will continue to focus on building more effective and efficient recommender systems for a wider range of application domains.

CONCLUSION

We have presented a brief overview of the field of recommender technologies and their emerging application domains. This chapter explains the current major recommender system approaches

within a unifying model, discusses emerging applications of recommender systems beyond traditional e-commerce, and outlines emerging trends and future research topics, along with additional readings in the area of recommender technologies. We believe that personalized recommender technologies will continue to advance and be applied in a variety of traditional and emerging application domains to assist users in the age of information overload.

REFERENCES

- Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6), 734–749. doi:10.1109/TKDE.2005.99
- Aggarwal, C. (2016). *Recommender Systems – The Textbook*. Berlin: Springer. doi:10.1007/978-3-319-29659-3
- Bell, R., Koren, Y., & Volinsky, C. (2010). All Together Now: A perspective on the Netflix prize. *Chance*, 23(1), 24–29. doi:10.1080/09332480.2010.10739787
- Burke, R. (2007). Hybrid web recommender systems. In *The Adaptive Web* (pp. 377–408). Berlin: Springer. doi:10.1007/978-3-540-72079-9_12
- Felfernig, A., Friedrich, G., & Schmidt-Thieme, L. (2007). Guest Editors Introduction: Recommender Systems. *IEEE Intelligent Systems*, 22(3), 18–21. doi:10.1109/MIS.2007.52
- Felfernig, A., Jeran, M., Ninaus, G., Reinfrank, F., & Reiterer, S. (2013). Toward the Next Generation of Recommender Systems: Applications and Research Challenges. *Multimedia Services in Intelligent Environments - Advances in Recommender Systems*, 24, 81–98.
- Felfernig, A., Ninaus, G., Grabner, H., Reinfrank, F., Weninger, D., Pagano, D., & Maalej, W. (2013). An Overview of Recommender Systems in Requirements Engineering. In *Managing Requirements Knowledge* (pp. 315–332). Berlin: Springer. doi:10.1007/978-3-642-34419-0_14
- Felfernig, A., Reiterer, M., Stettinger, M., Reinfrank, F., Jeran, M., Weninger, D., & Ninaus, G. (2013). Recommender Systems for Configuration Knowledge Engineering. In *International Configuration Workshop* (pp. 51–54). Berlin: Springer.
- Goy, A., Ardissono, L., & Petrone, G. (2007). Personalization in e-commerce applications. In *The Adaptive Web* (pp. 485–520). Berlin: Springer. doi:10.1007/978-3-540-72079-9_16
- Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. (2004). Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems*, 22(1), 5–53. doi:10.1145/963770.963772
- Inozemtseva, L., Holmes, R., & Walker, R. (2014). Recommendation Systems in-the-small. In *Recommendation Systems in Software Engineering* (pp. 77–92). Berlin: Springer. doi:10.1007/978-3-642-45135-5_4
- Jannach, D., Zanker, M., Felfernig, A., & Friedrich, G. (Eds.). (2011). *Recommender Systems: An Introduction*. Cambridge University Press.
- Konstan, J. A., Miller, B. N., Maltz, D., Herlocker, J. L., Gordon, L. R., & Riedl, J. (1997). Groups: Applying collaborative filtering to usenet news. *Communications of the ACM*, 40(3), 77–87. doi:10.1145/245108.245126
- Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix Factorization Techniques for Recommender Systems. *Computer*, 42(8), 30–37. doi:10.1109/MC.2009.263

- Linden, G., Smith, B., & York, J. (2003). Amazon.com Recommendations: Item-to-Item Collaborative Filtering. *Internet Computing*, 7(1), 76–80. doi:10.1109/MIC.2003.1167344
- Lops, P., Gemmis, M., & Semeraro, G. (2011). Content-based Recommender Systems: State of the Art and Trends. In *Recommender Systems Handbook* (pp. 73–105). Berlin: Springer. doi:10.1007/978-0-387-85820-3_3
- Lu, J., Wu, D., Mao, M., Wang, W., & Zhang, G. (2015). Recommender System application developments: A survey. *Decision Support Systems*, 74, 12–32. doi:10.1016/j.dss.2015.03.008
- Manouselis, N., Drachsler, H., Verbert, K., & Duval, E. (Eds.). (2013). *Recommender Systems for Learning*. (Springer Briefs in Electrical and Computer Engineering). Berlin: Springer. doi:10.1007/978-1-4614-4361-2
- Manouselis, N., Drachsler, H., Verbert, K., & Santos, O. (Eds.). (2014). *Recommender Systems for Technology Enhanced Learning: Research Trends and Applications*. Berlin: Springer. doi:10.1007/978-1-4939-0530-0
- Mika, S. (2011). Challenges for Nutrition Recommender Systems. *Proceedings of Workshop on Context Aware Intel. Assistance*, 25–33.
- Mobasher, B., & Cleland-Huang, J. (2011). Recommender Systems in Requirements Engineering. *AI Magazine*, 32(3), 81–89.
- Pazzani, M. J., & Billsus, D. (2007). Content-based recommendation systems. In *The Adaptive Web* (pp. 325–341). Berlin: Springer. doi:10.1007/978-3-540-72079-9_10
- Resnick, P., & Varian, H. R. (1997). Recommender systems. *Communications of the ACM*, 40(3), 56–58. doi:10.1145/245108.245121
- Ricci, F., Rokach, L., & Shapira, B. (Eds.). (2015). *Recommender Systems Handbook* (2nd ed.). Berlin: Springer. doi:10.1007/978-1-4899-7637-6
- Robillard, M., Maalej, W., Walker, R., & Zimmermann, T. (Eds.). (2014). *Recommendation Systems in Software Engineering*. Berlin: Springer. doi:10.1007/978-3-642-45135-5
- Robillard, M., Walker, R., & Zimmermann, T. (2010). Recommendation systems for software engineering. *IEEE Software*, 27(4), 80–86. doi:10.1109/MS.2009.161
- Sarwar, B. M., Karypis, G., Konstan, J., & Riedl, J. (2001). Item-based collaborative filtering recommendation algorithms. In *Proceedings of WWW10 Conference* (pp. 285–295). Hong Kong: ACM Press.
- Sarwar, B. M., Karypis, G., Konstan, J. A., & Riedl, J. (2000). Analysis of recommendation algorithms for e-commerce. In *ACM Conference on Electronic Commerce* (pp. 158–167). ACM Press.
- Schafer, J. B., Frankowski, D., Herlocker, J., & Sen, S. (2007). Collaborative filtering recommender systems. In *The Adaptive Web* (pp. 291–324). Berlin: Springer. doi:10.1007/978-3-540-72079-9_9
- Schafer, J. B., Konstan, J. A., & Riedl, J. (2001). E-commerce recommendation applications. *Data Mining and Knowledge Discovery*, 5(1/2), 115–153. doi:10.1023/A:1009804230409
- Wiesner, M., & Pfeifer, D. (2014). Health Recommender Systems: Concepts, Requirements, Technical Basics and Challenges. *International Journal of Environmental Research and Public Health*, 11(3), 2580–2607. doi:10.3390/ijerph110302580 PMID:24595212

ADDITIONAL READING

Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6), 734–749. doi:10.1109/TKDE.2005.99

Aggarwal, C. (2016). *Recommender Systems – The Textbook*. Berlin, Heidelberg: Springer. doi:10.1007/978-3-319-29659-3

Felfernig, A., Jeran, M., Ninaus, G., Reinfrank, F., & Reiterer, S. (2013). Toward the Next Generation of Recommender Systems: Applications and Research Challenges. *Multimedia Services in Intelligent Environments - Advances in Recommender Systems*. 24, 81-98.

Jannach, D., Zanker, M., Felfernig, A., & Friedrich, G. (Eds.). (2011). *Recommender Systems: An Introduction*. Cambridge University Press.

Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix Factorization Techniques for Recommender Systems. *Computer*, 42(8), 30–37. doi:10.1109/MC.2009.263

Linden, G., Smith, B., & York, J. (2003). Amazon.com Recommendations: Item-to-Item Collaborative Filtering. *Internet Computing*, 7(1), 76–80. doi:10.1109/MIC.2003.1167344

Manouselis, N., Drachsler, H., Verbert, K., & Duval, E. (Eds.). (2013). *Recommender Systems for Learning. (Springer Briefs in Electrical and Computer Engineering)*. Berlin, Heidelberg: Springer. doi:10.1007/978-1-4614-4361-2

Manouselis, N., Drachsler, H., Verbert, K., & Santos, O. (Eds.). (2014). *Recommender Systems for Technology Enhanced Learning: Research Trends and Applications*. Berlin, Heidelberg: Springer. doi:10.1007/978-1-4939-0530-0

Ricci, F., Rokach, L., & Shapira, B. P. (Ed.). (2015). *Recommender Systems Handbook*, 2nd edition. Berlin/Heidelberg: Springer.

Robillard, M., Maalej, W., Walker, R., & Zimmermann, T. (Eds.). (2014). *Recommendation Systems in Software Engineering*. Berlin, Heidelberg: Springer. doi:10.1007/978-3-642-45135-5

KEY TERMS AND DEFINITIONS

Collaborative Recommender Systems: Recommender systems that recommend items through user collaborations and are the most widely used and proven method of providing recommendations. There are three types: user-to-user collaborative filtering based on user-to-user similarity, item-to-item collaborative filtering based on item-to-item similarity, and latent factor-based collaborative filtering based on user-item matrix factorization.

Content-Based Recommender Systems: Recommender systems that are based on content of items and recommend a list of items with similar content to that of the items that were given good feedback by the target user.

Electronic Commerce: Activities of a traditional commerce process by using information and communication technologies through the Internet and intranets.

Hybrid Recommender Systems: Recommender systems that recommends items by combining two or more methods together, including the content-based method, the collaborative filtering-based method, the demographic method and the knowledge-based method.

Item-to-Item Collaborative Filtering: Collaborative filtering method that is based on similar items and recommends a list of items that are similar to the items that were given good relevance feedback by the target user.

Latent Factor Model-Based Collaborative Filtering: Collaborative filtering method that is based on latent factor models such as matrix factorizations and recommends a list of items that other user gave relevance feedback similar to that provided by the target user.

Personalized Recommendation Problem: Given a target user, produce personalized recommendations of items such as goods, services or information for the target user.

Recommender Systems (Recommendation Systems): Systems that provide users with personalized recommendations of items such as goods, services or information and thus help users find relevant items in the information overload.

User-to-User Collaborative Filtering: Collaborative filtering method that is based on similar users and recommends a list of items that other user gave relevance feedback similar to that provided by the target user.

D

Use of Data Analytics for Program Impact Evaluation and Enhancement of Faculty/Staff Development

Samuel Olugbenga King
Auburn University, USA

INTRODUCTION

One of the major developments in the 21st century academic development landscape is the pressure to meet “institutional concerns for quality control and accountability” (Land, 2011, p. 175). The now pervasive accountability culture, including a requirement for the use of “much more sophisticated evaluation mechanisms than have been the case in the past” (Brew & Peseta, 2008, p. 84) is typically driven by external accountability pressures. But there are also internal pressures. For example, Brew and Peseta (2008) describe how a change in management of an academic development unit can catalyze fresh demands for a systematic evaluation of the unit’s programs and effectiveness (pp. 83-84). *How might an academic development center that is faced with both external and internal accountability pressures demonstrate the impact and contributions of its programs to the broader institution’s strategic plan?*

To address this challenge, this paper will focus on the use of data analytics to satisfy the twin external and internal accountability demands (*summative evaluation*), while contributing to faculty and staff development in the process (*formative evaluation*). The goal is to meet the “increasing ‘troublesome’ expectations for academic development specialists to demonstrate the impact of their work” (Taylor, 2011, p. 95) through the use of data analytics, which enables systematic evaluation. In the next section, data analytics will be introduced within the context of the broader higher education institutional research ecosystem.

BACKGROUND: INSTITUTIONAL RESEARCH AND DATA ANALYTICS

Historically, the field of institutional research in higher education is predominantly focused on the strategic use of information and data sets to implement, review and enhance the academic mission of the university (Buller, 2012). The types of data that have been collected include system-generated behavioral data, e.g., human resource systems; survey data; transactional data, e.g., learning management systems; and frozen data, e.g., admission head counts (Bichsel, 2012, p. 16). But with the technology revolution, higher education is increasingly adopting and leveraging a range of data systems to support institutional capacity and meet strategic goals. The range of data systems that have been deployed to support higher educational units, such as student enrollment, information technology, budgeting and finance, human resources, student success, research administration, and facilities, include enterprise resource management (ERP) systems or business management/intelligence software, academic enterprise systems (e.g., LMS), customer relationship management (CRM) systems, and personalized learning environments, including assessment software (Norris & Baer, 2013, p. 9). An overview of the data analytics ecosystem in higher education is presented in Figure 1.

Moreover, 69% of higher education institutions surveyed in the USA on the use of data analytics indicated that analytics was either a “major institutional priority” or “major priority for some

Figure 1. Use of data analytics in higher education, with list of providers
Adapted from Norris & Baer, 2013, p. 19.

ERP/BI	• Oracle/PeopleSoft/Hyperion • Ellucian (Datatel/SunGard) • Campus Management	• Jenzabar • SAP/Business Objects • Workday (Big Data offering)	• Top School • Destiny Solutions • Adobe
LMS	• Blackboard/iStrategy • Moodle rooms • Desire2Learn	• Pearson/eCollege • LoudCloud • Sakai	• Instructure/Canvas • Campus Cruiser • Appendra
Visualization/CRM Dashboard/Analytics Consulting/Generalized Advising	• IBM (SPSS/Cognos) • Microsoft • eThorty	• Nuventive • eVisions • Pentaho	• iDashboard • Tableau • QlikTech
Advising/CRM → Learning Relationship Management	• Starfish Retention Solutions • Civitas	• EBI/MAPWorks • RapidInsight	• Hobsons • Respondus • Salesforce.com • Talisma • ConnectEDU • Campus Lab
Personalized Learning Environments			• WebStudy • Knewton • Cengage • Turning Technologies • Epsilen • SoftChalk • Ucompass • eXact Learning Solutions • GoingOn • SMART Technologies

departments but not entire institution” (Bichsel, 2012, p. 8). But this is not just an American phenomenon. For example, ERP systems (e.g., Oracle and PeopleSoft) and LMSs (e.g., Blackboard, Moodle and Sakai) are pervasive at institutions in Africa, Asia and Europe (e.g., Cochrane, Black, Lee, Narayan, & Verswijvelen, 2013; Nkurunziza, 2013). This pervasiveness of analytics systems is linked with their perceived benefits for student enrollment and academic success, optimization of the use of institutional resources, and demonstration of higher education’s effectiveness and efficiency (Bichsel, 2012, p. 11). As Petersen (2012) commented, the “current higher education landscape is replete with demands for improving accountability, increasing efficiency, and controlling costs. At the same time, information technologies make it easier to collect and analyze

information to measure outcomes or to assist in decision making” (p. 44).

It is therefore clear from the ubiquity of data analytics in higher education that institutions have adopted them as a systematic tool to measure, support and enhance performance. But this question may arise, *Are there any specific reasons why analytics would be strategically beneficial to academic development?* In the next section, I will highlight the role of data analytics in academic development.

THE ROLE OF DATA ANALYTICS IN ACADEMIC DEVELOPMENT

The field of academic development is evolving, but one of the requirements for further advancement

in the field is the need for the adoptions of tools and concepts that transcend local institutional contexts. For example, Land (2011) advocates the adoption of “new tools of analysis and some conceptual lenses to bring our educational practices into view somewhat differently, so that our understanding of what might be happening can unfold through a variety of perspectives and offer different ways forward” (p. 176). A review of the academic development literature (Gibbs, 2013; Gordon, 2012; Levinson-Rose & Menges, 1981; Ouellett, 2010; Steinert et al., 2006; Stes, Min-Leliveld, Gijbels, & Van Petegem, 2010; Taylor, 2011; Amundsen & Wilson, 2012) indicates that there are three other relevant requirements for further progress in the field:

1. The adoption of conceptual frameworks to help augment the transferability of study findings to similar contexts, e.g., transversal measures – Gordon (2012, p. 177);
2. The need for sophisticated, systematic evaluation; and
3. The need for detailed descriptions of practice to enhance transferability.

The position of this paper is that if data analytics is situated within a coherent theoretical framework and detailed descriptions of its applications are provided, then it is a compelling and systematic tool that can provide summative and formative evaluations of programs’ impact and effectiveness. Data analytics provides sophisticated evaluation mechanisms for academic development because of the following attributes:

- Ability to receive, incorporate and process large sets of data, including in real time;
- Use of advanced algorithms to detect patterns in data for analysis;
- Visualization of datasets (e.g., dashboards, infographics, widgets);
- Provision of big-picture, as well as fine-grained/granular data views;

- Makes feasible longitudinal evaluation, as they serve as data repositories.

The primary focus of this paper is the use of the data analytics system, Google Analytics, to provide summative and formative evaluation of a Center’s programs in response to external and internal accountability demands. But there will also be a secondary focus on the use of data analytics to evaluate the impact or uptake of academic development videos. The next section will highlight the use of Google Analytics within a cohesive framework to provide summative and formative evaluations of the impact of an academic development Center’s programs to meet both external and internal performance demands.

USE OF GOOGLE ANALYTICS AS A KNOWLEDGE PERFORMANCE INDICATOR (KPI)

Although there are different views of data analytics, in this paper, analytics is defined as the “process of data assessment and analysis that enables us to measure, improve, and compare the performance of individuals, programs, departments, institutions or enterprises, groups of organizations and/or entire industries” (van Barneveld, Arnold, & Campbell, 2012, p. 3). Data analytics in essence provides a set of objective key performance indicators (KPIs). This section will detail how Google Analytics, which focuses on the analysis of web site traffic data, is being used to evaluate the performance of a center for teaching and learning at a research university (see Matteson, 2013 for a practical guide to opening a free Google Analytics account).

Google Analytics Application Context: The Biggio Center

The Biggio Center for the Enhancement of Teaching and Learning engages, supports, and empowers

Auburn University's academic community in each phase of the scholarly teaching and transformative learning process. Some of the programs include new faculty and graduate student development, distance learning, testing services, professional development workshops and seminars, instructional technology, and active learning spaces and design. Moreover, the Center's programs are intentionally aligned with the Auburn University's mission and strategic plan. For example, one of the alignment goals is the provision of access to development resources and support throughout the academic career of Auburn faculty and staff. Many of the Center's programs are accessed through its web site, which is often the first point of contact for patrons. So the web site is integral to the operational effectiveness and impact of the Biggio Center.

The Biggio Center faces two accountability measures. The external accountability demand reflects the new accountability culture in higher education. The Center is required to produce annual reports that show how its programs are contributing to the wider university's strategic plan. Brew and Peseta (2008) highlighted how annual reports can be used to "establish the impact of academic work" (p. 83). In addition, the Center is under new management that is interested in reviewing past performance in order to consolidate and modify practices, so as to assure continued enhancement and progress (Gray & Radloff, 2008). So an evaluation of its web site traffic is invaluable to satisfying both internal and external requirements for impact evaluation and improvement. The results of the Google Analytics evaluation process are therefore designed to satisfy both the external demand, which is summative in thrust, and the internal requirements for improvement, which is quintessentially formative.

Moreover, the academic development literature furnishes several notable examples or potential applications of data analytics. For example, the authors of the study reported in Cochrane et al. (2013) used information from data analytics (#web

clicks per student) on the low learner engagement on the e-learning courses hosted on Blackboard at "New Zealand's largest polytechnic" (p. 276) to change the institution's e-learning strategy by re-orienting the courses from a teaching-centered to a learner-centered pedagogy through the implementation of communities of practice (pp. 276-277). The analytics indicated that "fewer than 40 of the over 1000 courses on Blackboard in 2004 evidenced interactivity beyond discussion boards" (p. 277).

Howard (2012) highlighted the value of "YouTube Universities" (p. 168), e.g., The MIT YouTube Channel, and implicitly linked the corresponding YouTube video analytics data as a measure of diffusion and impact. Similarly, Rhode and Krishnamurthi (2013) describe how they used tracking data analytics of the database system designed for faculty consultations to improve faculty development consultation services. On a broader scale, impact factors and citation counts are used as measures of journal prestige and faculty productivity respectively.

THE ANALYTICS PROCESS

The cohesive framework that has been adopted to guide the use of Google Analytics in the impact evaluation of the Biggio Center's web site is the *analytics process* (Figure 3), which entails five components:

1. Starting with a strategic question;
2. Finding or collecting the appropriate data to answer that question;
3. Analyzing the data with an eye toward prediction and insight;
4. Representing or presenting the findings in ways that are both understandable and actionable, and
5. Feeding back into the process of addressing strategic questions and creating new ones.

Figure 2. The Biggio Center web site home page

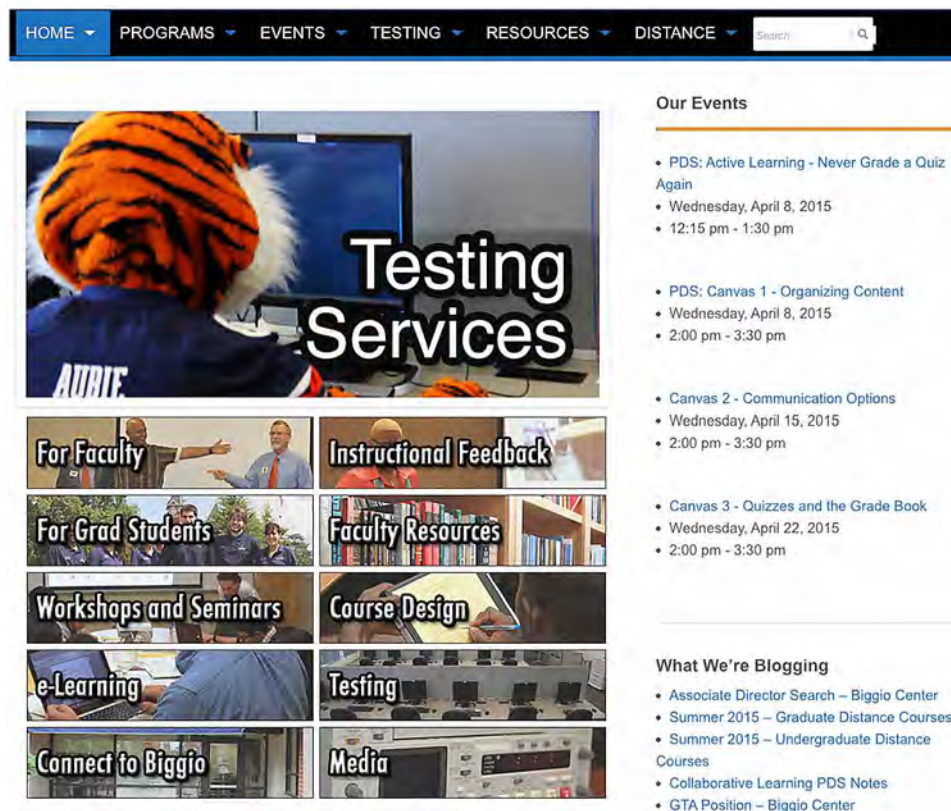


Figure 3. A pictorial overview of the analytics process
Adapted from Norris & Baer, 2013, p. 53.

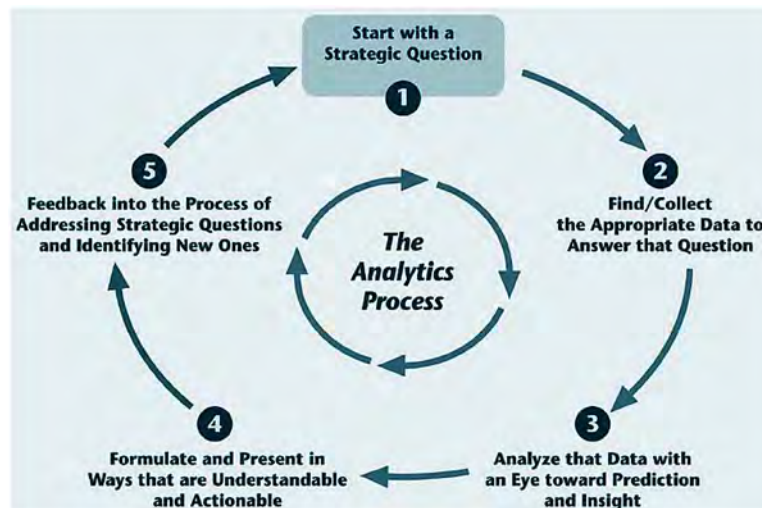
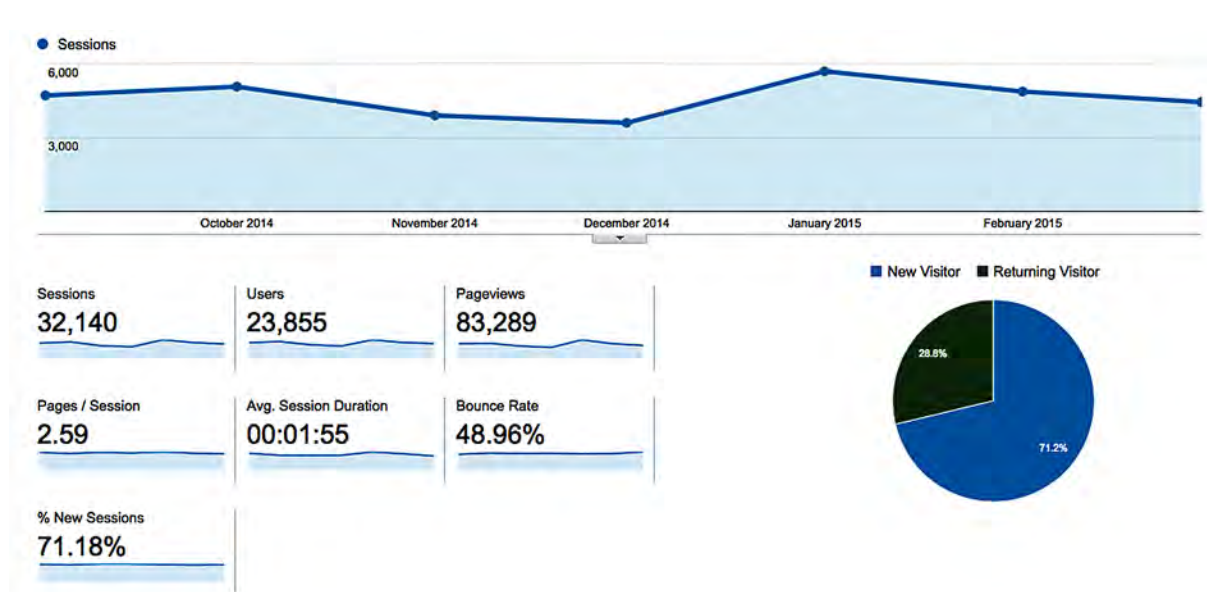


Figure 4. Composite data about the web traffic and users of the Biggio Center Web site



The strategic questions that were adopted and that Google Analytics has been used to investigate are:

1. What can be learned from the Biggio Center web site traffic about the uptake or usage (i.e., impact) of the Center's programs and services?
2. How may the findings be used to reinforce, modify, or enhance the Center's programs and services?

The next section will detail how these questions have been answered through Google Analytics.

Overview of the Uptake of the Biggio Center's Programs

The *line graph* in Figure 4 shows there was a total of 32,140 sessions or visits to the Biggio Center site by 23,855 visitors between September 1, 2014 and March 25, 2015. The peak period, with about 6,000 visits, was in January, 2015. This means that 121 people visited the web site per day in the period under review. Meanwhile, the number of total Auburn University employees who may be

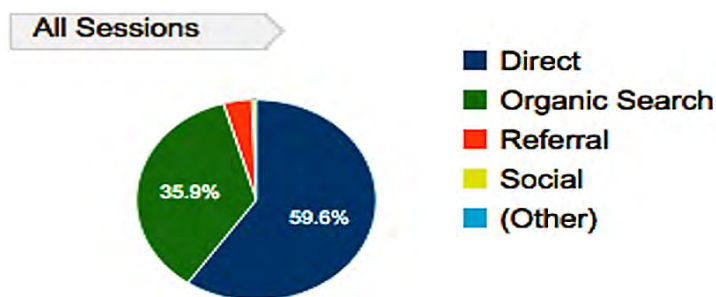
realistically expected to benefit from the Center's programs is about 4000.

The *pie chart* shows that nearly 28.9% of the total number of sessions were by *returning visitors*. This is perhaps an indication that the Biggio Center has a significant footprint in the Auburn University community and that patrons have high satisfaction with the level of programs offered. Further, the fact that seven of out 10 visits (71.2%) to the Biggio Center site is by *new visitors* indicates that the Center's programs are perhaps attractive to new clients.

The *areas of improvement*, based on this composite dataset overview, include reducing the bounce rate (49%), the number of people who leave the web site without any further interactions, and perhaps increasing the average session duration beyond 1:55 minutes. The peak period in January 2015 coincides with a major center initiative (a large professional development event), while the dip in December 2014 is understandable in a higher education context – there is a lull in programming, and a winding down of the academic calendar for the year.

One way to leverage the peak data is to create longitudinal anchors to the audience for the

Figure 5. How users found or were directed to the Biggio Center site



large event in January. For example, Gibbs (2013) describes how the “most important consequences of training of newly appointed teachers is not that they teach better in the classroom the next week, but that some come back to the educational development unit for help a decade later when they eventually become head of department” (p. 8). So the challenge is to create meaningful programming that would catalyze the ‘coming back’ of program participants both in the short term, and over the duration of their careers.

Mode of Site Navigation

The data in Figure 5 indicates that nearly six out of the 10 people (59.6%) who accessed the Biggio Center site did so *directly*. This includes typing the url directly into the browser, and also accessing the site through links in emails and relevant Biggio Center program applications. Similarly, at least one out of three people (35.9%) found the Biggio Center site through organic search, i.e., the use of search engines such as Google and Bing, either to navigate directly to the Biggio Center site, or by searching for similar services and programs.

Another possible area for improvement would be to include the program schedules or events calendar prominently on the Center’s home page, so that visitors would not have to navigate a couple of web pages before accessing the program schedule.

The *areas of improvement*, based on this channel dataset, includes correlating the peaks and troughs in web traffic (Figure 4) with how users navigated to the web site, as depicted in Figure 5. For example, since most users find the site directly, including the Center’s url in informational and program literature, and incorporating QR codes would facilitate easier navigation to the site. This is buttressed by the fact that the direct search traffic is positively correlated with the overall web traffic to the Center site (Figure 6). Social media contributes less than 1%, so it is not considered important as a source of traffic.

Use of Mobile Devices

The Biggio Center recently optimized its web site to ensure mobile accessibility, so an overview of which devices are used to access the web site is relevant. The data in Figure 7 indicates that about

Figure 6. Contributions from Direct and Organic Search to the total web traffic

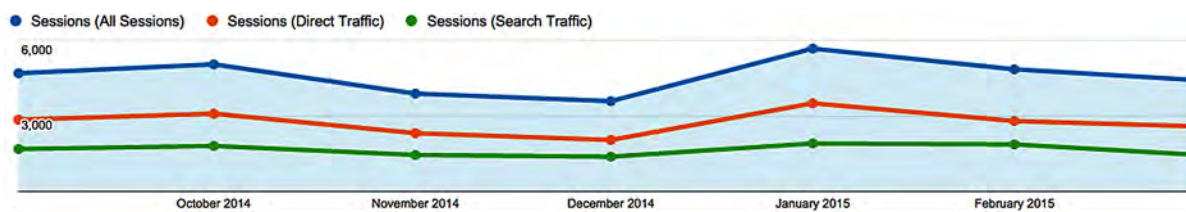
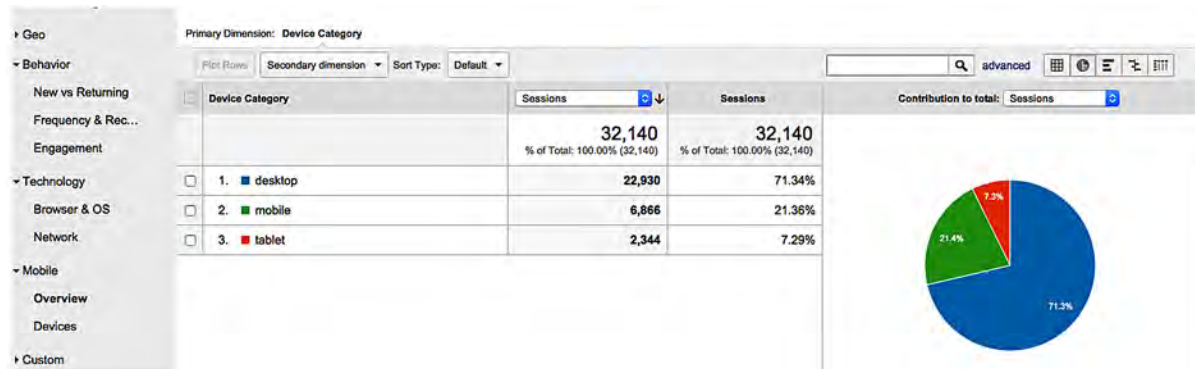


Figure 7. Percentage of visits to the Biggio Center web site via mobile devices



29% of visits to the site were through mobile devices (mobile phones and tablets).

Similarly, the data in Figure 8 shows that the top mode of navigating to the Biggio Center site is through the use of a desktop via direct navigation. In contrast, visitors who use a mobile phone to access the site are slightly more likely to do so via organic search (11.2%) than directly (9.6%). Further, tablet users clearly prefer navigating to the site directly (5.2%) rather than through organic search (1.9%).

Meanwhile, the data in Figure 9 shows the top 20 web pages that are visited on the Biggio Center web site. The most visited Center programs are the Distance Learning pages, including the Dietary Managers Program, and Testing Services.

Another perspective on visitor behavior on the Biggio Center site is to look at persistence – if visitors do not find what they really need on the landing page, for example, they might keep browsing till they get to the right web page. For example, the data in Figure 10 presents information on the third web pages (“2nd interaction”) that visitors navigate to after beginning their browsing from the “starting pages”. Again, the most visited web pages are the distance learning pages (13,200 sessions or about 42% of all sessions). An interesting data point for the 2nd interaction dataset is the fact that the Center’s Spring 2015 program schedule (“/ACADEMIC/spring2015”) seems to be valued by some of the visitors.

Figure 8. Correlation of device use with navigational mode patterns

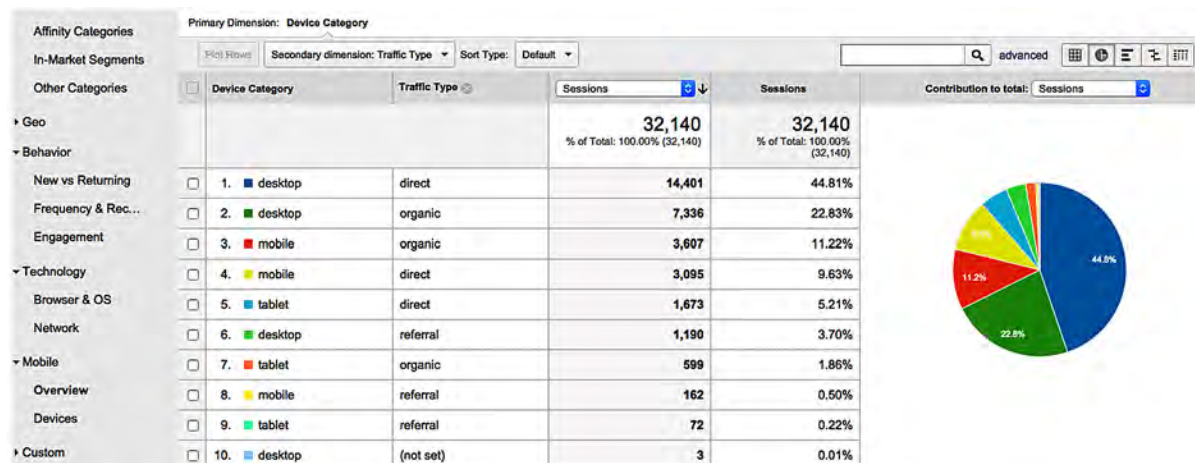


Figure 9. The top 20 most visited web pages on the Biggio Center web site (the type of device, mobile or desktop, used to access the site is also linked)

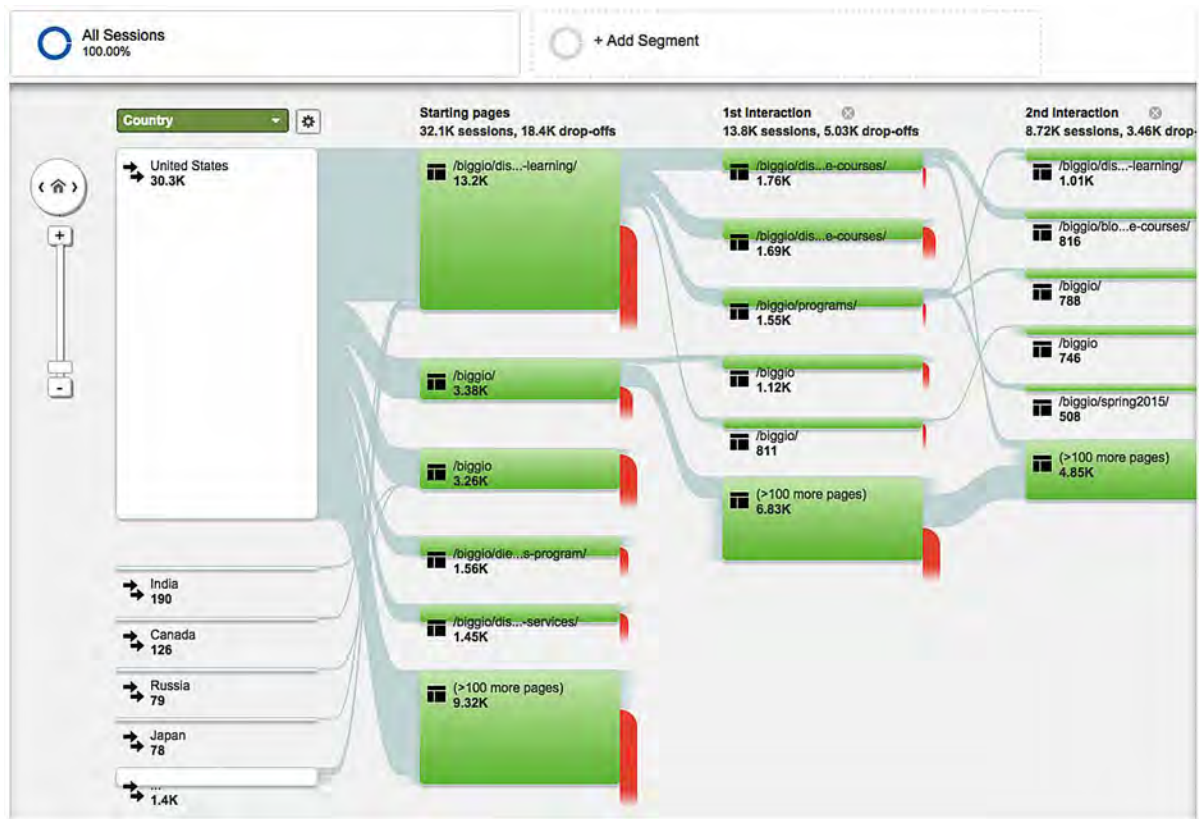
Device Category		Landing Page	Sessions	Sessions
			32,140 % of Total: 100.00% (32,140)	32,140 % of Total: 100.00% (32,140)
1.	desktop	/biggio/distance-learning/	9,151	28.47%
2.	desktop	/biggio/	2,872	8.94%
3.	mobile	/biggio/distance-learning/	2,746	8.54%
4.	tablet	/biggio/distance-learning/	1,278	3.98%
5.	desktop	/biggio/dietary-managers-program/	1,032	3.21%
6.	desktop	/biggio/distance-learning/auburn-distance-courses/undergraduate-distance-courses/	834	2.59%
7.	desktop	/biggio/distance-learning/testing-services/	751	2.34%
8.	mobile	/biggio/distance-learning/testing-services/	666	2.07%
9.	desktop	/biggio/distance-learning/auburn-distance-courses/graduate-distance-courses/	440	1.37%
10.	desktop	/biggio/oct/	435	1.35%
11.	desktop	/biggio/programs/breeder/	418	1.30%
12.	mobile	/biggio/dietary-managers-program/	403	1.25%
13.	desktop	/biggio/biggio-staff/	323	1.00%
14.	desktop	/biggio/distance-learning/notfound.php	320	1.00%
15.	mobile	/biggio/distance-learning/testing-services/register-for-an-exam/	317	0.99%
16.	desktop	/biggio/test-proctoring/	307	0.96%
17.	mobile	/biggio/	291	0.91%
18.	desktop	/biggio/summer-institute/	290	0.90%
19.	desktop	/biggio/ipad-cohort-program/	265	0.82%
20.	mobile	/biggio/distance-learning/auburn-distance-courses/undergraduate-distance-courses/	260	0.81%

The *areas of improvement*, based on the datasets presented in Figures 7-10, include brainstorming on whether it might be feasible to increase the social media traffic to the site. For example, the videos on the Biggio Center YouTube site have been viewed 6567 times (i.e., between February 2010 and April 2015). In contrast, there were only three visits from the YouTube traffic to the main Biggio Center web site. So there is no correlation between levels of activity on the Center's YouTube site and the traffic from the YouTube site to the Center's main web site. Further, there is scope to perhaps considerably enhance the electronic registration process for the Testing Services program, which currently includes a significant offline registration process. It would also be beneficial to post updates for the Distance Learning courses

and Testing Services on the home page to provide easier navigation aids for visitors.

In summary, the data clearly shows that there is significant web traffic on the Biggio Center's web site, which is an indication of update and usage levels. It is also clear that the most valued resources appear to be the Distance Learning courses, Testing Services, and to a lesser extent, program schedules.

It would be important to benchmark this dataset (September 2014 – March 2015) with future longitudinal datasets to more precisely evaluate the uptake of the Biggio Center's programs and services by the Auburn University community and other relevant stakeholders.

Figure 10. Mapping user behavior from the initial “starting pages” to the third page (“2nd interaction”)

ADDRESSING DISADVANTAGES OF USING DATA ANALYTICS

It is important to address the potential disadvantages of using Google Analytics or a similar data analytics system (KPIs) for summative or formative program evaluation. Some of the potential pitfalls associated with the use of data analytics include cost, data misuse, inaccurate analysis or findings, the relevance of the data analytics approach, and privacy (Bichsel, 2012; Oblinger, 2012; Stiles, 2012; van Barneveld et al., 2012).

These pitfalls have been mitigated in this study through the adoption of Google Analytics, which is free, with its use situated within the analytics process, to constructively align data collection and analysis with the larger external and internal program impact evaluation requirements. Data triangulation has also been adopted to corroborate the findings from Google Analytics. For example,

the peaks and troughs for the line graph in Figure 4 were corroborated with the information about the Center’s program schedule for December 2014 and January 2015. The data analysis has also focused on a seven-month period, which provides better data aggregation.

To assure user privacy, no individual or user identification has been collected or monitored through the Google Analytics process. The analysis has instead focused on the investigation of broad trends, which necessarily involves the masking of user identification data. So the manner of adoption of Google Analytics for this study could be construed as low “reputational risk” (Figure 11), as it is geared towards the protection of user identity.

Finally, the appropriateness of the use of Google Analytics for this study could be filtered through the five dimensions or benchmarks of good ass

Figure 11. Data sensitivity classification

Adapted from Stiles, 2012, p. 24.

	Most Critical <i>Highest level of sensitivity</i>	Critical <i>Moderate level of sensitivity</i>	Least Critical <i>Very low, but still requiring some protection</i>
Legal requirements	Protection of data is required by law (e.g., HIPPA and FERPA data elements and other PII protected by law)	The institution has a contractual obligation to protect the data (e.g., bibliographic citation data, bulk licensed software)	
Reputational risk	High	Medium	Low
Other institutional risks	Information that provides access to resources, physical or virtual	Smaller subsets of Most Critical data from a school, large part of a school, or department	
Data examples	Medical Student Prospective student Personnel Donor or prospect Financial Contracts Physical plant detail Credit card numbers Certain management information	Information resources with access to Most Critical data Research detail or results that are not Most Critical Library transactions (e.g., catalog, circulation, acquisitions) Financial transactions that do not include Most Critical data (e.g., telephone billing) Very small subsets of Most Critical data	Campus maps Personal directory data (e.g., contact information) E-mail Institutionally published data

essment rubric (Suskie, 2006). The use of the free Google Analytics software has been driven by the adoption of *strategic questions and goals*, which have provided *reasonably accurate and truthful results*, and which have in turn provided *valuable* summative and formative evaluation to feed into the iterative process of program,

reinforcement, modification and improvement (Figure 12).

Hence the analysis of the Google Analytics findings has been guided by the need to combine “analytics with sound reasoning” (Oblinger, 2012, p. 12).

Figure 12. The five dimensions of good assessment

Adapted from Suskie, 2006.

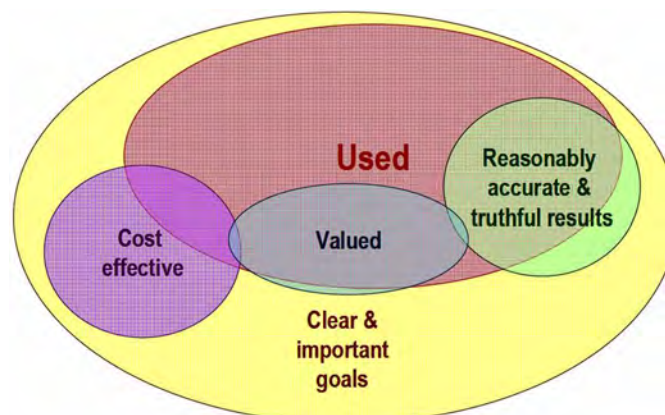


Table 1. A comparison of the digital video analytics for selected academic development centers in North America

Digital Video Analytics	Auburn University, USA	University of Southern Alabama, USA	University of Michigan, USA	University of Waterloo, Canada	Trent University, Canada
#Videos	76	59	129	37	11
#Views	6,567	678	13,464	738	421
#Subscribers	19	-	67	NA	1
Host Site	YouTube	Kaltura	YouTube	YouTube	YouTube
Date of Site Creation	February 22, 2010	September, 2014	October 13, 2011	-	-

Video analytics retrieved on April 3, 2015.

ANALYTICS (KPIs) FOR ACADEMIC DEVELOPMENT VIDEOS

Many academic development centers increasingly produce and integrate videos into their programs and services, which are typically hosted on sites, such as YouTube, Vimeo, or Kaltura. So one way to assess the analytics on the Biggio Center academic development digital videos is to benchmark the data with the digital videos produced by other academic development centers. The examples provided (Table 1) are therefore further data points about measuring comparative impact (van Barneveld et al., 2012), as they help provide a big picture view regarding the use and uptake of academic development videos in the United States and Canada.

This big picture view can also be extended to other geographical regions in order to provide international benchmarking of academic development videos. For example, digital video analytics for academic development centers in the United Kingdom, Singapore and Australia is provided in Table 2.

FUTURE RESEARCH DIRECTIONS

Future research will focus on comparative studies of analytics applications (e.g., Tableau and iDashboard), as well as an longitudinal evaluation of the impact of the use of data analytics on the programming efficacy of academic development centers.

Table 2. A comparison of the digital video analytics for selected academic development centers in the UK, Singapore, and Australia

Digital Video Analytics	London School of Economics (LSE), UK	Manchester Metropolitan University, UK	National University of Singapore (NUS)	University of Melbourne, Australia
#Videos	14	174	33	4
#Views	242	19, 885	459	574
#Subscribers	-	54	-	-
Host Site	YouTube	YouTube	YouTube	YouTube
Date of Site Creation	-	September 9, 2011	-	-

Video analytics retrieved on April 7, 2015.

CONCLUSION

This chapter has focused on how academic development centers may respond to the twin demands of internal and external pressures for demonstrations of impact and effectiveness, especially within a higher education context that is increasingly defined by an accountability culture. The focus has been on the use of data analytics based on its unique affordances to provide cost-effective, systematic, reliable and useful data to inform both summative and formative evaluation. The data analytics process has been conducted by situating the inquiry within a cohesive framework of strategic questioning and analysis.

The data analytics process has been applied to evaluate the impact of the programs of an academic development center (i.e., the Biggio Center) at a large research university, with the findings being filtered to enhance further development at the center. The findings from the center have also been benchmarked with data from similar centers in North America to provide a big picture overview of the emerging field. Similar comparisons have been provided for institutions in Europe, Australasia and Africa. Moreover, the data analytics process reported in this study has been designed to enrich the academic development field regarding the need for systematic evaluation of impact, detailed descriptions of practice to enhance generalizability, and the adoption of new tools of analysis to advance understanding in the field. Further, the potential pitfalls about using data analytics, including privacy, have been addressed.

It is astute to conclude that the revamped accountability culture in higher education and the associated troublesome expectation regarding “performance measurement is not going to go away”. Therefore one way to respond to the new pressures is by embracing a data-driven culture (Bichsel, 2012) that prioritizes the use of knowledge performance indicators, such as Google Analytics and YouTube Video analytics, as a way

to both demonstrate impact and enhance further development. This approach would guarantee a blending of higher education “accountability and assurance agendas more explicitly into our narrative” (Stefani, 2013, p. 296), so as to meet and exceed the troublesome expectations in stride.

ACKNOWLEDGMENT

The author wishes to acknowledge his gratitude to Dr. Diane Boyd for the helpful critique provided on the initial draft of this manuscript.

REFERENCES

- Amundsen, C., & Wilson, M. (2012). Are we asking the right questions? A conceptual review of the educational development literature in higher education. *Review of Educational Research*, 82(1), 90–126. doi:10.3102/0034654312438409
- Bichsel, J. (2012). *Analytics in higher education: Benefits, barriers, progress, and recommendations (Research Report)*. Louisville, CO: Educause Center for Applied Research.
- Brew, A., & Peseta, T. (2008). The precarious existence of the academic development unit. *The International Journal for Academic Development*, 13(2), 83–85. doi:10.1080/13601440802076467
- Buller, J. L. (2012). *Best practices in faculty evaluation: A practical guide for academic leaders*. San Francisco: Jossey-Bass.
- Cochrane, T., Black, B., Lee, M., Narayan, N., & Verswijvelen, M. (2013). Rethinking e-learning support strategies. *The International Journal for Academic Development*, 18(3), 276–293. doi:10.1080/1360144X.2012.733884
- Gibbs, G. (2013). Reflections on the changing nature of educational development. *The International Journal for Academic Development*, 18(1), 4–14. doi:10.1080/1360144X.2013.751691

- Gordon, G. (2012). It is time to strengthen the conceptual focus of SoTL. *The International Journal for Academic Development*, 17(2), 177–180. doi:10.1080/1360144X.2012.668750
- Gray, K., & Radloff, A. (2008). The idea of impact and its implications for academic development work. *The International Journal for Academic Development*, 13(2), 97–106. doi:10.1080/13601440802076525
- Howard, C. R. (2012). Technology in higher education. In W. Buskist & V. A. Benassi (Eds.), *Effective college and university teaching: Strategies and tactics for the new professoriate* (pp. 163–172). Washington, DC: Sage. doi:10.4135/9781452244006.n18
- Land, R. (2011). There could be trouble ahead: Using threshold concepts as a tool of analysis. *The International Journal for Academic Development*, 16(2), 175–178. doi:10.1080/1360144X.2011.568747
- Levinson-Rose, J., & Menges, R. J. (1981). Improving college teaching: A critical review of research. *Review of Educational Research*, 51(3), 403–434. doi:10.3102/00346543051003403
- Matteson, S. (2013). *Five things you should know about Google Analytics*. Retrieved August 25, 2015, from <http://www.techrepublic.com/blog/google-in-the-enterprise/five-things-you-should-know-about-google-analytics/>
- Nkurunziza, J. (2013). Can Moodle be used as a tool to educate the masses in Africa? In S. Retalis & M. de Raadt (Eds.), *Proceedings of the 2nd Moodle Research Conference* (pp. 65–73). Sousse, Tunisia: Moodle Research Conference.
- Norris, D. M., & Baer, L. L. (2013). *Building organizational capacity for analytics*. Boulder, CO: Educause.
- Oblinger, D. G. (2012). Let's talk analytics. *EDUCAUSE Review*, 10, 11–13.
- Ouellett, M. L. (2010). Overview of faculty development: History and choices. In *A guide to faculty development* (2nd ed.; pp. 3–20). San-Francisco: Jossey-Bass.
- Parr, C. (2013). *Africa's mobile phone e-learning transformation*. Times Higher Education.
- Petersen, R. J. (2012). Policy dimensions of analytics in higher education. *EDUCAUSE Review*, 44–49.
- Purcell, K. (2013). *Online video 2013*. Washington, DC: Pew Research Center.
- Rhode, J., & Krishnamurthi, M. (2013). A consultations tracking database system for improving faculty development consultation services. *To Improve The Academy*, 32, 199–216.
- Stefani, L. (2013). Performance measurement for academic development: Risk or opportunity. *The International Journal for Academic Development*, 18(3), 294–296. doi:10.1080/1360144X.2013.818542
- Steinert, Y., Mann, K., Centeno, A., Dolmans, D., Spencer, J., Gelula, M., & Prideaux, D. (2006). A systematic review of faculty development initiatives designed to improve teaching effectiveness in medical education. *Medical Teacher*, 28(6), 497–526. doi:10.1080/01421590600902976 PMID:17074699
- Stes, A., Min-Leliveld, M., Gijbels, D., & Van Petegem, P. (2010). The impact of instructional development in higher education: The state-of-the-art of the research. *Educational Research Review*, 5(1), 25–49. doi:10.1016/j.edurev.2009.07.001
- Stiles, R. J. (2012). *Understanding and managing the risks of analytics in higher education: A guide*. Boulder, CO: Educause.
- Suskie, L. (2006). *Principles of good assessment*. Retrieved August 25, 2015, from https://www.pratt.edu/uploads/principles_of_good_assessment_suskie_2006.pdf

Taylor, L. (2011). Shining the light on inquiry: Diverse goals, conceptual lenses, and methods in academic development scholarship. *The International Journal for Academic Development*, 16(2), 93–95. doi:10.1080/1360144X.2011.568278

van Barneveld, A., Arnold, K. E., & Campbell, J. P. (2012). *Analytics in higher education*. Boulder, CO: Educause.

ADDITIONAL READING

Daniel, B. (2015). Big data and analytics in higher education: Opportunities and challenges. *British Journal of Educational Technology*, 46(5), 904–920. doi:10.1111/bjet.12230

Johnson, L., Adams, B. S., Estrada, V., & Freeman, A. (2014). *NMChorizon report: 2014 higher education edition*. Austin, TX: The New Media Consortium.

KEY TERMS AND DEFINITIONS

Academic Development Center: The Centers at universities and colleges whose primary purpose is to support the professional development of faculty, staff and graduate students.

Accountability Culture: The significant push for quality measures and/or data metrics as evidence of a higher education unit's effectiveness.

Analytics Process: The pedagogical alignment of the use of data analytics software to address specific (identified) objectives, such as faculty development or student learning goals.

Data Analytics: The use of specialized software to analyze large datasets to provide actionable feedback or information.

Formative Evaluation: Assessment that is intrinsically used to inform, develop or shape both process(es) and product(s).

Knowledge Performance Indicators: Specialized programs or data metrics that provide quantifiable measures or indicators of performance.

Summative Evaluation: Assessment with primary focus on final outcomes or product(s).

Category D

Data Mining and Databases

Corporate Disclosure Measurement

Md. Salah Uddin Rajib

Jahangirnagar University, Bangladesh

Md. Qutub Uddin Sajib

China University of Geosciences (CUG), China

INTRODUCTION

Researchers have mentioned that about 90% of the entire world's data has been created in the past 2 years (e.g., Kolb & Kolb, 2013). Most of these data are made in electronic format. Corporate disclosures (CDs) are available now in electronic formats through different documents also. As the organizations are disclosing the disclosure annually or in the interim period, the numbers of volumes are increasing gradually. Organizations disclose disclosure through various media, e.g., annual reports, websites, supplementary of financial statements and others.

A number of researchers have investigated the corporate disclosure (CD) scenario in the last decades as it is assumed to have an impact on the capital market (Healy & Palepu, 2001). CD can be categorized into two broad groups – Mandatory disclosures and Voluntary disclosures. Mandatory disclosures are obligatory by the concern organization of a country (e.g., Securities and Exchange Commission). In contrast, voluntary disclosures are made by the firms voluntarily considering the interest of stakeholders. A number of topics on CD have been investigated e.g., financial disclosure (e.g., Malone et al., 1993), human resource capital (e.g., Bontis, 2003), environmental disclosure (e.g., Gamble et al., 1996), strategic disclosure (e.g., Wagenhofer, 1990) and others.

Very often, researchers have used content analysis to explore the scenario of CD practice (e.g., Bontis, 2003; Khan et al., 2011). In content

analysis they normally count the frequency of words to explore the scenario. Researchers normally select a list of related terms or keywords to the concern topic and search the terms in the annual reports or other related supplementary where disclosure has been presented by the organizations (e.g., Bontis, 2003).

Researchers have urged for more reliable measurement of CD practice (e.g., Rahman & Post, 2011). Consistent with that, this chapter proposes the use of text mining (TM) approach for measuring the CD. Currently, different areas of researches are applying the approaches and techniques of TM. Researcher of medical science, politics, business and others have used the techniques of TM to extract the information (Moohebat et al., 2015). The purpose of TM is to process unstructured textual information. In large text collection, TM works as a tool for knowledge discovery (Gomez et al., 2002). TM uses several techniques to extract the information, e.g., text classification, text clustering, text summarization, sentiment analysis and others. This chapter investigates the usability and advantages of TM approaches in the measurement of CD. Therefore, the aim of this paper is fourfold. First, it reviews the background of CDs and their measurement and background of TM. Second, it discusses the TM approaches, including general model of TM and methodology. Third, the usability of TM in measuring the CD has been investigated. And finally, it discusses the prospects of research in the area of using the TM in CD measurement.

BACKGROUND

As mentioned earlier, CD can be categorized into two broad groups - mandatory and voluntary disclosure. The demand for CD primarily arises from the perspective of regulation. Therefore, Securities and Exchange Commission (SEC) played an important role in the CD, made by the organizations. History states that because of various crises in US e.g., war, unemployment, racial conflict, environmental degradation and others, authority tempts to examine the fundamental principles underlying the body of corporate law for policy making (Schoenbaum, 1972). Sarbanes-Oxley (SOX) requires more and better information also. Consisting with the situation The Securities Exchange Act of 1934 imposes disclosure requirements on corporations in the US (Schoenbaum, 1972). With time, Generally Accepted Accounting Principles (GAAP) imposes some disclosures on corporations. Voluntary disclosure drives by various factors, e.g., manager-specific influences (e.g., Bamber et al., 2010), corporate governance (e.g., Hermalin & Weisbach, 2012), external pressure groups (e.g., Islam & Deegan, 2010) and others. Therefore, voluntary disclosures are associated with various areas and it is unstructured in nature. Researchers have tempted to find out the relationship of CD with the organization performances (Healy & Palepu, 2001). Therefore, measurement of CD emerges as a research area for the researchers.

This chapter conducts studies on the usability of text mining (TM) for CD measurement. TM is a discipline of natural language processing. The approach of TM has evolved from the simple word processing at the end of the 1990s to now (Antonio & Ferneda, 2008). TM derives much of its inspiration and direction from seminal research on data mining (Feldman & Sanger, 2007). The increased numbers of electronic documents have triggered the use of TM to extract the knowledge from textual data.

TEXT MINING: GENERAL MODEL AND METHODS



The primary objective of TM is to retrieve information from unstructured text and to present the extracted knowledge to the users in a concise form (Ananiadou & McNaught, 2006). It has been mentioned by the researchers that TM can add significant importance and economic value where the data are in electronic form and are in centralized warehouses (e.g., Weiss et al., 1999).

Text Mining (TM)

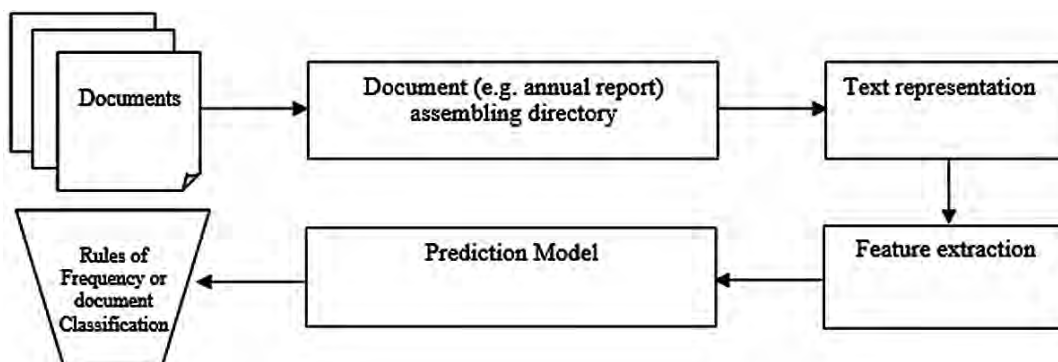
TM is defined as a tool for extracting the information from electronic documents by using natural language processing computer program. Broadly, TM can be defined as a knowledge-intensive process in which a user interacts with a document collection over time by using a suite of analysis tools (Feldman & Sanger, 2007). Like the data mining, TM extracts useful information from the sources through various approaches. It can collect data both from formalized database records and from unstructured textual data.

General Model of Text Mining

In a TM technique, computer programs scan text in a document and apply model consisting with the objective. The individual documents are used as raw cases for TM. The basic approach of TM can be shown by the following figure -

Figure 1 shows that the documents assemble directory drive from the documents that are raw cases for the TM techniques. After representing the text in the computerized program, the features of the text will be extracted. After extracting the feature, various pre-specified models can be applied. The frequency of the terms can be counted easily or document can be classified considering various aspects.

Figure 1. A general model of automated text mining
Adapted from Weiss et al., 1999.



Methods of Text Mining

There are various methods for TM and TM related tasks. The developments of data mining and computer program are enriching the TM regarding the methods and approaches (e.g., Billheimer et al., 2003). Considering the objective of this chapter, several common methods of TM have been discussed below. It seems that methods like Term identification, Document Clustering, Document Summarization and Automated Document classification will be helpful for measuring the CD practice.

Term Identification

Term identification is known as keyword identification also. Keywords are widely used to define queries with information retrieval systems as they are easy to define, revise, remember, and share (Rose et al., 2010). Generally the keywords or the keyterms represent the essence of a document. Terms or the keyterms are used by the researchers from the very beginning of measuring the CD practice.

The use of the keywords or the terms is visible in numerous areas. Therefore, Automatic term recognition (ATR) has received concentration of the researchers. ATR can be defined as a technique which automatically identifies and extracts terms from the text. For ATR, there are various methods arises from the different approaches

like rules based, co-occurrence-based, statistical and machine learning techniques (Ananiadou & McNaught, 2006).

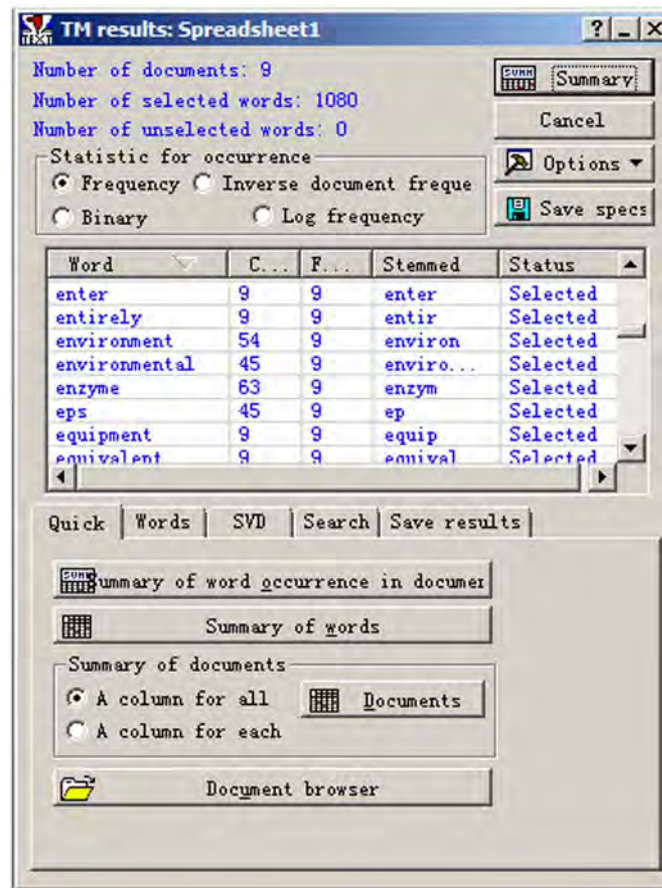
The ATR method provides a number of important features that are beneficial for various aspects. For example, ATR can recognize the ‘multi-world terms’ or ‘nested terms’ efficiently (Frantzi et al., 2000). By handling the primarily extracted terms from a document efficiently, the performance of ATR can be improved more. Acronyms, inflections, orthographic variants can be handled easily on primary extracted terms of TM through screening. It is natural that in a text document, one term will not present the same realization through the entire document. TM can handle these issues efficiently through various ways.

As an example of TM, texts have been extracted from the annual reports (for 2013) of the nine largest chemical companies (BASF, Dow Chemical, Bayer, Saudi Basic Industries, LyondellBasell Industries, E. I. Du Pont de Nemours, Air Liquide, Syngenta, Praxair) of the world. The screenshot (STATISTICA, version 8 has been used) has been presented in Figure 2.

Document Clustering

Document clustering or text clustering involves with the clustering or making groups of textual documents regarding various objects. It is used to classify the documents. The TM technique can perform the document clustering consider-

Figure 2. Text extraction from the annual reports of 9 largest chemical firms in the world



ing the intended use. A group of documents can be clustered as topic clusters, corresponding to a topic shared by all the documents included in the cluster (Thomas et al., 2011). In TM approach, there are a number of ways to cluster the documents (Aggarwal & Zhai, 2012). The documents can be clustered regarding a single term i.e., the documents can be clustered weighing the existence of a term frequency. Bag-of-words approach is used to cluster the document also. The bag-of-words approach uses the words of the documents and their frequencies of occurrence within a document in order to identify the central topic(s).

Automatic Document Classification

Automatic Document Classification involves with the automatically categorizing the digital documents into pre-defined class. Automatic document

classification identifies the underlying patterns and distinguishing features within documents that make them part of a defined group or class and uses this information to assign each new document to a known class (Thomas et al., 2011). There are many approaches for automatic document classification (Feldman & Sanger, 2007) e.g., neural networks, Bayesian logistic regression, probabilistic classifiers, regression methods and others. For executing automatic document classification, four main issues have to be ensured (Feldman & Sanger, 2007). First, the categories that will be used to classify the instances must be confirmed. Second, a training set for each of the categories has to be ensured. Third, features that represent each of the instances have to be confirmed. Finally, one should take a decision on the algorithm that will be used for the categorization. By applying the

automatic document classification one can easily classify a new document regarding the developed automatic classifier.

Document Summarization

Radev et al. (2002) have defined summary as a text that is produced from one or more texts, that conveys important information in the original text(s), and that is no longer than half of the original text(s) and usually significantly less than that. It derives from single or multiple documents and will be presented in a useful, and in short form. Document summarization arises from the assumptions, such as– a human reader breaks each sentence into a set of information fragments to which the sentence is referring; information fragments are mutually independent; and an information fragment has an important score (Radev & McKeown, 1999). There are various techniques of document summarization e.g., extraction, fusion, abstraction and compression.

Extraction is the procedure of identifying the important sections of the text and producing them verbatim. Abstraction aims to produce important material in a new way. Fusion combines the extracted parts coherently. And compression aims to throw out unimportant sections of the text (e.g., Radev & McKeown, 2002). Document summarization excludes redundant information scattered across the document(s) and can order the extracted sentences according to chronological relatedness of topic or precedence (Barzilay et al., 2002).

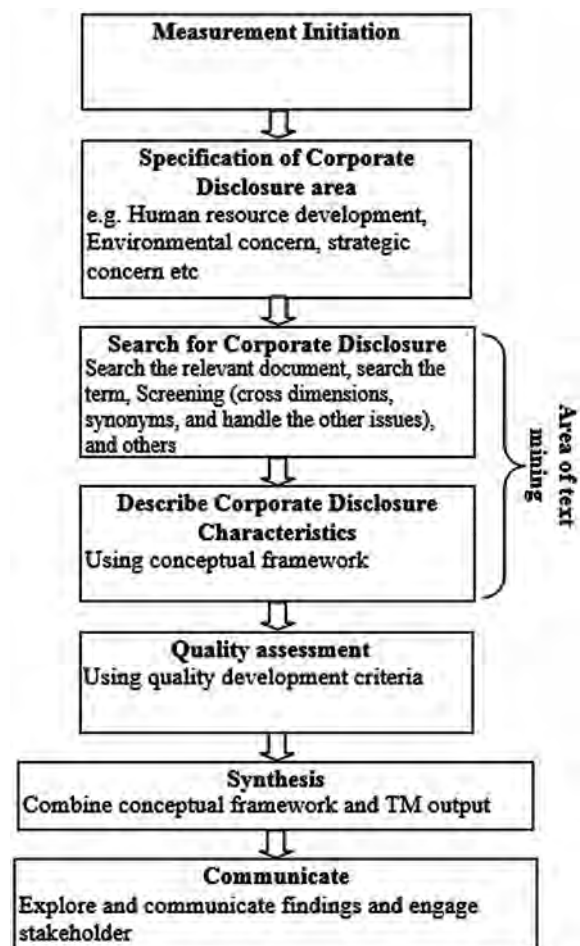
DISCUSSION ON TEXT MINING IN CORPORATE DISCLOSURE MEASUREMENT

In different areas, TM has been suggested in different implementation path. For example, Thomas et al. (2011) show the way of using TM for systematic reviews of literature. The previous research in the area of CD measurement indicates that

the measurement has been executed on the basis of simple content analysis where the frequency of keywords has been counted as the scale (e.g., Bontis, 2003; Khan et al., 2011). This approach can present misleading result very often, as it does not consider the position of words in the sentence and the meaning of the sentence. After discussing the basic approaches and methodologies, it seems plausible that TM can be used for CD measurement and it can produce more reliable results. Figure 3 has been presented to show the area where TM can be used in CD measurement.

Figure 3. Corporate disclosure measurement and area of text mining

Adapted from Thomas et al., 2011.



Text Mining Technique for Corporate Disclosure Measurement

In this section, different features of TM have been investigated to check the usability of TM for CD measurement. The discussion has been executed within four main methods of TM. 1) Term identification, in which keyterms from documents are identified automatically; 2) Document clustering, in which similar documents relating to the CD can be grouped together; 3) Automatic document classification, which can classify the documents automatically regarding the CD in terms of keywords and 4) Document summarization, which can ensure the quality of CD measurement by extracting the salient informative components of documents. Discussion has been arranged below considering the stage of application of TM.

Identifying the Relevant Document

To measure the CD practice of an organization, it is plausible to consider all the documents that are published by an organization. A number of previous researches have used annual reports only to measure the practice of CD (e.g. Gamble et al., 1996; Bontis, 2003). But, the organization publishes a number of electronic documents in a year, e.g., annual report, brochure, interim reports, and different memorandum. Moreover, publication of information through the website is a common practice nowadays.

Searching all the documents that are published by the organization is time consuming under the manual approach. Under this situation, TM can play an important role. TM offers a number of advantages to measure the CD. Firstly, researchers will be able to investigate a wider context in a shorter time. Secondly, this alternative approach will enrich the research methodology and will contribute to the development of reliable measurement of CD. It will contribute to the urge of demanding transparency and reliability for the measurement of CD (e.g., Rahman & Post, 2011).

Identification of Terms

One of the challenges in CD measurement is to select the search term. Previous research on CD measurement shows that over the years, researchers have developed indices to measure relevant area's CD (Marston, 1991).

TM can offer a broader scope for the identification of terms. Apart from the index terms, researchers can investigate the possible terms for inclusion in the measurement of CD by using the TM. The range of search terms will be much wider that will increase the reliability of CD measurement.

Identification of keyterms has two main characteristics. Firstly, it expands the scope of identifying the terms. Secondly, the cross-disciplinary terms will not be recognized efficiently. These features are effective for the CD measurement also. The automatic term identification of TM expands the scope for term identification to measure the CD. But, the cross-disciplinary terms will not be recognized efficiently. For example, voluntary disclosure which is unstructured in nature may include 'environmental concern' whereas some documents may recognize it by 'ecological concern'. Therefore, further processing is needed for efficient measurement of CD after automatic term identification.

Screening the Term and Screening Prioritization

The limitation of automatic term identification that has been mentioned in the previous section (e.g., cross-disciplinary term) demands screening the term.

To screen the term manually, very often two people are employed (Thomas et al., 2011). This labor-intensive task consumes considerable amount of cost and time. To reduce the time and cost in screening, automatic classification has been suggested (e.g., Thomas et al., 2011). The approach of automatic classification is applicable

for the measurement of CD also. At first, automatic classification can tell whether the documents are going to disclose financial issues, ecological issues or human resources developmental issues. Reviewing the documents after classification can reduce both the cost and time and ensure the reliability at the same time. Mandatory corporate disclosure, e.g., legal issues can be identified easily by screening.

Basically, researchers have suggested two approaches in TM that can help for screening. One is automatic document classification and another is the use of term recognition technique under screening prioritization.

When it is not possible to use automatic classification, the term recognition technique under screening prioritization may be a workable alternative (Thomas et al., 2011). For the use of term recognition, relatively little ‘training’ of the system is required. The term can be identified and weighted considering the degree of relevance with the specific CD. Then, a certain percentage of keywords of the lists can be checked manually for erroneous exclusions.

Screening can provide archival benefit by identifying the most relevant documents. This retrieval features will enhance the reliability of CDs measurement. Situation of multi-words can be handled efficiently by screening also. But, screening has limitations. There is a chance that researcher will be failed to quantify the number of terms or documents that have been missed in the screening (Thomas et al., 2011).

Applying Eligibility Criteria

Applying eligibility criteria is another approach for TM that has been suggested by the researchers for the information extraction from documents (e.g., Thomas et al., 2011). Under this approach, researchers identify a set of relevant titles and abstracts manually in order to train a classifier. Then the developed training set is used for the remaining relevant studies. This approach is fruitful when the numbers of documents are large and

increasing frequently. This approach is applicable for measuring the CD and can offer significant benefit as the documents for CDs are increasing rapidly nowadays.

Applying Pre-Existing Keywords Automatically

It has been found that the researchers have developed indices over the years to measure the CD (Marston, 1991). Therefore, the developed indices can be used easily by TM to measure the CD. Under this situation, TM can add value to the measurement of CD than general search as TM can help for mapping the plural words into their singular form (e.g., human resources and human resource).

Pattern Identification for Documents

TM can identify the pattern of documents weighing the trained data set or from the data set from bag of words approach. Pattern identification for documents method can identify groups of similar documents very quickly. Pattern identification for documents is named as document clustering also. Pattern identification for documents can provide valuable information in measuring the CD. As organizations publish CDs in various documents, pattern identification can be an effective tool for CDs measurement.

Summarizing the Documents

There are two types of document summarization. They are– 1) involves with the extracting important information from documents and 2) formulate a new document by paraphrasing the content of source document(s) (Thomas et al., 2011). Document summarization is applicable for the measurement of corporate disclosure also. Although advanced language generation technique is required for the second method, the first method can be used for the CD measurement easily. It can present the concise form of a document or

multi-documents by redundancy reduction and by ensuring the chronological order. For CD measurement, it can concise the relevant part of specific disclosures and can help the researchers to explain the nature of disclosures for different documents and organizations.

Utility of Text Mining in Corporate Disclosure Measurement

Most of the disclosures made by organizations in various sources are unstructured in nature. Under this situation, text mining can offer great advantages to the researchers for CD practice measurement.

Most of the research papers on the corporate disclosure measurement use different terms or keywords to measure the corporate disclosure. Use of terms or keywords is a good practice for the established theories and for structured texts. However, term based methods suffer from the problems of polysemy and synonymy (Zhong et al., 2012). The situation of 'Polysemy' rises when a word has multiple meanings. And situation of 'synonymy' rises when multiple words have the same meanings.

The probability of the problem of polysemy and synonymy is high for measuring the CD by keywords as the disclosures (especially voluntary) are unstructured in nature. These problems can be partially solved by the TM approaches by extracting the universal words of a document. By using the different TM approaches researchers can hold the sentence flavor also.

Therefore, theoretically it seems that the text mining can add value to the measurement of corporate disclosure which ultimately will impact positively on good business practice.

Summary of Methodology of Text Mining for Corporate Disclosure

The summary of technique of TM that has been discussed for the CD measurement is presented in Table 1. The development of Table 1 is in-

fluenced by Thomas et al. (2011). One notable thing that has been mentioned in the discussion is that some of the techniques of TM can be used individually where some can be used as associate techniques of others. For example, screening techniques can be used after identification of the terms from documents to ensure the appropriateness of measurement.

The efficient use of these methodologies depends on the researchers also. It has been mentioned by the researchers that ethical issues are inherent in the research (Rajib & Mou, 2014).

FUTURE RESEARCH DIRECTIONS

The use of TM in the measurement of CD is relatively new. Therefore, this area has miles to go. New innovations move through a series of stages from learning about the potential of a given innovation, through being persuaded to consider its potential, deciding whether to try it out, trying it out, and deciding whether or not to continue to use the innovation (Rogers, 1983). The proposed measurement of CD by TM is not an exception of this mentioned innovation. This paper develops the concepts only. It discusses the theoretical ground for the use of TM in the measurement of CD. Theoretically, it seems that TM can add value to the measurement of CD in great extent. In future, usability of TM in the measurement of CD can be tested considering various areas e.g., financial disclosure, human resources development, environment concern and others. The results of TM approaches can be compared with the previous research on CD measurement which is based on the simple content analysis.

The techniques of TM are still in the stage of development and evaluation. The methods are still relatively unknown and unused in the area of CD measurement. Therefore, use of TM in the measurement of CD can be investigated from various viewpoints in future.

Table 1. Summary of text mining for corporate disclosure measurement with advantages and disadvantages

Stage of Measurement	Text Mining Method	Advantages	Disadvantages/Limitations
Identification of search terms or keywords	Term recognition	Enriches the search terms and strategy.	More consciousness is required.
Screening the term	Term recognition	Provides the advantages of adjusting the cross-disciplinary terms, multi-terms and others. Reduces the necessity of manual 'double screening' by human beings.	Consciousness is required.
Applying eligibility Criteria	Automatic classification	Reduces the time to select the search term and for final search.	Further development is needed in this technique (e.g., Thomas et al. 2011).
Applying pre-existing keywords automatically	Automatic extraction and Automatic classification	Indices are available in literature of CD measurement practice. By using the previous indices, one can easily extract the frequency of the terms in document(s). By using the previous indices, one can easily classify the documents also.	Using the indices only will reduce the scope of inserting alternative words. Ultimately, it will reduce the reliability of measurement.
Pattern Identification of Documents	Automatic classification	Reduces the time to understand the nature of documents regarding the targeted CD measurement.	Depends on content of documents. It cannot explain critical concepts.
Clustering	Automatic document clustering	It gives an overview of the collection of documents to the researchers quickly.	It cannot recognize complex concepts that need explanation.
Summarization	Summarizing the documents	It can provide a quick summary for a document or for collection of documents from salient informative sentences.	It's not a good technique for the document(s) which content is heterogeneous.

CONCLUSION

Researches in the area of CD measurement are increasing in these days. The documents in which CD is searched are increasing in volume. Searching the CD in general approach is cumbersome. Moreover, the unstructured (especially voluntary disclosure) disclosure makes the measurement difficult.

Probably, the positive news under this scenario for the researchers is that most of the documents are available in electronic format nowadays. TM can be a helpful tool for the researchers to measure the CD in this emerged situation. It can reduce time and cost and can enhance the reliability. By the help of TM, researchers can identify the key-terms automatically as well as search the indexed terms in documents in the shortest time. It can categorize the documents from a large number of documents and helps to present the summary

of documents automatically. As the technology of TM is developing till now, it is expected that further development of text mining will increase the reliability of CD measurement.

REFERENCES

- Aggarwal, C. C., & Zhai, C. X. (2012). A survey of text clustering algorithms. In C. Charu (Ed.), *Mining text data* (pp. 77–128). New York: Springer US. doi:10.1007/978-1-4614-3223-4_4
- Ananiadou, S., & McNaught, J. (Eds.). (2006). *Text mining for biology and biomedicine*. Boston: Artech House.
- Antonio do Prado, H., & Ferneda, E. (Eds.). (2008). *Emerging technologies of text mining: Techniques and Applications*. New York: IGI Global. doi:10.4018/978-1-59904-373-9

- Bamber, L. S., Jiang, J., & Wang, I. Y. (2010). Whats my style? The influence of top managers on voluntary corporate financial disclosure. *Accounting Review*, 85(4), 1131–1162. doi:10.2308/accr.2010.85.4.1131
- Barzilay, R., Elhadad, N., & McKeown, K. (2002). Inferring strategies for sentence ordering in multi document news summarization. *Journal of Artificial Intelligence Research*, 17, 35–55.
- Billheimer. (2003). Methods and system for text mining using multidimensional subspaces. *Google Patents*. Retrieved on 1st March, 2015 from www.google.com/patents/US6611825
- Bontis, N. (2003). Intellectual capital disclosure in Canadian corporations. *Journal of Human Resource Costing & Accounting*, 7(1), 9–20. doi:10.1108/eb029076
- Frantzi, K., Ananiadou, S., & Mima, H. (2000). Automatic recognition of multi-word terms: The C value/NC-value method. *International Journal on Digital Libraries*, 3(2), 115–130. doi:10.1007/s007999900023
- Gamble, G. O., Hsu, K., Jackson, C., & Tollerson, C. D. (1996). Environment disclosure in annual reports: An international perspective. *The International Journal of Accounting*, 31(3), 293–331. doi:10.1016/S0020-7063(96)90022-9
- Gomez, M. M., Gelbukh, A., & Lopez, A. L. (2002). Text mining at detail level using conceptual graphs. In D. Corbett, G. Angelova, & U. Priss (Eds.), *Conceptual structure: Integration and interfaces* (Vol. 2393, pp. 122–136). Berlin: Springer. doi:10.1007/3-540-45483-7_10
- Healy, P. M., & Palepu, K. G. (2001). Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure analysis. *Journal of Accounting and Economics*, 31(1-3), 405–440. doi:10.1016/S0165-4101(01)00018-0
- Hermalin, B., & Weisbach, M. S. (2012). Information disclosure and corporate governance. *The Journal of Finance*, 67(1), 195–233. doi:10.1111/j.1540-6261.2011.01710.x
- Islam, M. A., & Deegan, C. (2010). Media pressures and corporate disclosure of social responsibility performance information: A study of two global clothing and sports retail companies. *Accounting and Business Review*, 40(2), 131–148. doi:10.1080/00014788.2010.9663388
- Khan, H. U. Z., & Islam, M. A. (2011). Corporate sustainability reporting of major commercial banks in line with GRI: Bangladesh evidence. *Social Responsibility Journal*, 7(3), 347–362. doi:10.1108/17471111111154509
- Kolb, J., & Kolb, J. (2013). *The big data revolution*. Plainfield, IL: Applied Data Labs Inc.
- Malone, D., Fries, C., & Jones, T. (1993). An empirical investigation of the extent of corporate financial disclosure in the oil and gas industry. *Journal of Accounting, Auditing & Finance*, 8(3), 249–273.
- Marston, C. L., & Shrives, P. J. (1991). The use of disclosure indices in accounting research: A review article. *The British Accounting Review*, 23(3), 195–210. doi:10.1016/0890-8389(91)90080-L
- Moohebat, M., Raj, R. G., Kareem, S. B. A., & Thorleuchter, D. (2015). Identifying ISI-Indexed articles by their lexical usage: A text analysis approach. *Journal of the Association for Information Science and Technology*, 66(3), 501–511. doi:10.1002/asi.23194
- Radev, D. R., Hovy, E., & McKeown, K. (2002). Introduction to the special issue on summarization. *Conceptual Linguistics*, 28(4), 399–408. doi:10.1162/089120102762671927
- Radev, D. R., & McKeown, K. (1999). Generating natural language summaries from multiple online sources. *Computational Linguistics*, 24(3), 469–500.

Rahman, N., & Post, C. (2011). Measurement issues in environmental corporate social responsibility (ECSR): Toward a transparent, reliable, and construct valid instrument. *Journal of Business Ethics*, 105(3), 307–319. doi:10.1007/s10551-011-0967-x

Rajib, M. S., & Mou, N. Z. (2014). Ethical Issues of Qualitative Research. In J. Wang (Ed.), *Encyclopedia of Business Analytics and Optimization* (pp. 852–863). Hershey, PA: Business Science Reference; doi:10.4018/978-1-4666-5202-6.ch080

Rogers, E. M. (1983). *Diffusion of innovations*. New York: Free Press.

Rose, S., Enger, D., Cramer, N., & Cowley, W. (2010). Automatic keyword extraction from individual documents. In *Text Mining: Applications and Theory*. New York: John Wiley & Sons Ltd. doi:10.1002/9780470689646.ch1

Schoenbaum, T. J. (1972). The relationship between corporate disclosure and corporate responsibility. *Fordham Law Review*, 40(3), 565–594.

Thomas, J., McNaught, J., & Ananiadou, S. (2011). Applications of text mining within systematic reviews. *Research Synthesis Methods*, 2(1), 1–14. doi:10.1002/jrsm.27 PMID:26061596

Wagenhofer, A. (1990). Voluntary disclosure with a strategic opponent. *Journal of Accounting and Economics*, 12(4), 341–363. doi:10.1016/0165-4101(90)90020-5

Weiss, S. M., Apte, C., Damerau, F. J., Johnson, D. E., Oles, F. J., Goetz, T., & Hampp, T. (1999). Maximizing text-mining performance, Intelligent Information Retrieval. *IEEE Intelligent Systems*, 14(July/August), 63–69. doi:10.1109/5254.784086

Zhong, N., Li, Y., & Wu, S. T. (2012). Effective pattern discovery for text mining. *IEEE Transactions on Knowledge and Data Engineering*, 24(1), 30–44. doi:10.1109/TKDE.2010.211

ADDITIONAL READING

Feldman, R., & Sanger, J. (2007). *Text mining handbook: Advanced approaches in analyzing unstructured data*. New York: Cambridge University Press.

KEY TERMS AND DEFINITIONS

Bag-of-Words Model: In this approach a text is considered as the bag of its words. The frequency or occurrence of each word is used as a feature in Bag-of-Words model to classify the documents.

Corporate Disclosure: Financial and non-financial information of organizations those are disclose to the public.

Mapping: Describing the characteristics of the studies that have been explored by the researchers.

Nested Terms: “Nested terms” is a concept to present the joint generalization of words.

Stemming: The term stemming refers to the reduction of words to their roots so that different grammatical forms or declinations of verbs are identified and indexed as the same word.

Training Set: A training set is a set of words that is used to discover potentially predictive model in text mining.

Data Mining and Knowledge Discovery in Databases

D

Ana Azevedo

Polytechnic Institute of Porto, Portugal

INTRODUCTION

The term knowledge discovery in databases or KDD, for short, was coined in 1989 to refer to the broad process of finding knowledge in data, and to emphasize the “high-level” application of particular Data Mining (DM) methods (Fayyad, Piatetski-Shapiro, & Smyth, 1996). Fayyad considers DM as one of the phases of the KDD process. The DM phase concerns, mainly, the means by which the patterns are extracted and enumerated from data. Nowadays, the two terms are, usually, indistinctly used.

Efforts are being developed in order to create standards and rules in the field of DM with great relevance being given to the subject of inductive databases (De Raedt, 2003) (Imielinski & Mannila, 1996). Within the context of inductive databases a great relevance is given to the so called DM languages.

This chapter presents a comprehensive introduction and summary of the main basic topics and bibliography in the area of DM, nowadays. Thus, the main contribution of this chapter is that it can be considered as a good starting point for newcomers in the area.

The remaining of this article is organized as follows. Firstly, DM and the KDD process are introduced. Following, the main DM tasks, methods/ algorithms, and models/patterns are organized and succinctly explained. SEMMA and CRISP-DM methodologies are next introduced and compared with KDD. A brief explanation of standards for DM is then presented. The article concludes with possible future research directions and conclusion.

BACKGROUND

In recent years, we have witnessed the growth and consolidation of the DM area. Since the first Workshop, IJCAI-89 Workshop on Knowledge Discovery in Databases, which took place at Detroit in 1989 and that led, in 1995, to the nowadays main annual conference in the area, ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, the number of publications and conferences dedicated to the area presents a significant growth. These conferences as well as several seminal papers, helped in the consolidation of the area. Since then, the evolution has been overwhelming, and DM can be considered as a consolidated research area (Azevedo, 2015).

DATA MINING AND THE KNOWLEDGE DISCOVERY IN DATABASES PROCESS

“The KDD process, as presented in (Fayyad, Piatetski-Shapiro, & Smyth, 1996), is the process of using DM methods to extract what is considered knowledge according to the specification of measures and thresholds, using a database along with any required preprocessing, sub sampling, and transformation of the database. There are five stages considered, namely, selection, preprocessing, transformation, data mining, and interpretation/evaluation as presented in Figure 1:

- **Selection:** This stage consists on creating a target data set, or on focusing in a subset

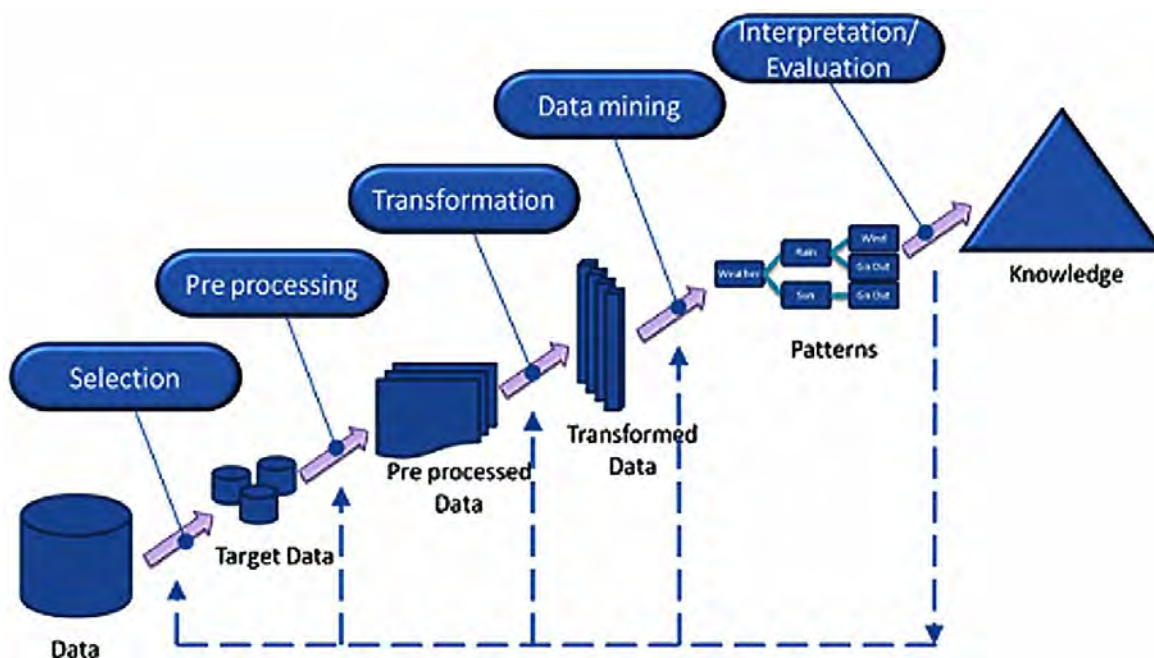
- of variables or data samples, on which discovery is to be performed;
- **Preprocessing:** This stage consists on the target data cleaning and preprocessing in order to obtain consistent data;
- **Transformation:** This stage consists on the transformation of the data using dimensionality reduction or transformation methods;
- **Data Mining:** This stage consists on the searching for patterns of interest in a particular representational form, depending on the DM objective (usually, prediction);
- **Interpretation/Evaluation:** This stage consists on the interpretation and evaluation of the mined patterns.” (Azevedo & Santos, 2008, p. 183)

The KDD process is preceded by the development of an understanding of the application domain, the relevant prior knowledge, and the goals of the end-user. It must be continued by knowledge

consolidation, incorporating this knowledge into the system. The KDD process is interactive and iterative, involving numerous steps with many decisions being made by the user (Brachman, J., & Anand, 1996).

As of the foundations of KDD and DM, several applications were developed in many diversified fields. The growth of the attention paid to the area emerged from the rising of big databases in an increasing and differentiated number of organizations. Nevertheless, there is the risk of wasting all the value and wealthy of information contained in these databases, unless the adequate techniques are used to extract useful knowledge (Chen, Han, & Yu, 1996) (Fayyad U. M., 1997) (Simoudis, 1996). The application of DM techniques with success can be found in a wide and diversified range of applications, for instance, bioinformatics, ecology and sustainability, finance, industry, marketing, scientific research, telecommunications, and other applications (Azevedo & Santos, 2011).

Figure 1. The KDD process



DATA MINING TASKS, METHODS/ALGORITHMS, AND MODELS/PATTERNS

Prediction and description were identified as the two “high-level” primary goals of DM (Fayyad, Piatetski-Shapiro, & Smyth, 1996). “Prediction involves using some variables or fields in the database to predict unknown or future values of other variables of interest. Description focuses on finding human-interpretable patterns on finding the data.” (Fayyad, Piatetski-Shapiro, & Smyth, 1996, p. 12)

To achieve these goals some DM tasks were used and its description can be found in the literature. Some of the most common DM tasks are classification, prediction, clustering, association, and summarization:

- Classification consists in finding a function that associates an instance of the independent variables to a specific pre-defined value of the target variable, named as class. The target variable should be of nominal type;
- Prediction consists in finding a function that associates an instance of the independent variables to some numerical value of a real-valued target variable, in order to predict future unknown values for that target variable;
- Clustering allows the identification of homogeneous groups containing several elements which have high similarity with all the other elements of the same group, and that have low similarity to all the elements of the other groups;
- Association consists in finding a model that describes significant dependencies between variables, that is to say, identifying facts that can be directly or indirectly associated;
- Summarization uses methods to discover a compact description for data, in order to find a better description of the data, thus improving its understanding.

In the literature, a significant number of DM methods/algorithms have been developed to accomplish each task, and different kinds of DM models/patterns can be obtained. Classification methods include decision trees, classification rules, neural networks, support vector machines, Bayesian data analysis, Bayesian networks, and k-nearest neighbor. Prediction methods include linear regression, nonlinear regression, neural networks, decision trees, and k-nearest neighbor. Clustering methods include partitioning, hierarchical and model-based methods. Association is accomplished with association rules. Summarization methods include EDA (Exploratory Data Analysis) and OLAP (On-Line Analytical Processing).

An outline of this DM tasks, methods/algorithms, models/patterns, and guidance is presented in Table 1.

There are different forms of evaluating models’ interestingness in each case, such as cross-validation, bootstrapping, bagging and boosting, estimating confidence intervals, or ROC curves. There are also a large variety of alternatives to provide guidance, including accuracy and error measures. We will not discuss in more detail each one of these issues, since we consider it is outside the scope of this text. Several textbooks can be found that cover these topics in more detail, e.g. (Han & Kamber, 2012) (Hand, Mannila, & Smyth, 2001) (Larose, 2014) (Myatt, 2014) (Witten & Frank, 2011) (Ye, 2003).

The emergence of more complex types of data led to the development of new methods and models to cope with the new task of mining complex data. As examples, we can point out text mining (do Prado & Ferneda, 2008), web mining (content, structure, and usage) (Markov & Larose, 2007), spatial data mining (Nlenanya, 2009), graph mining (Zhang, Hu, Xia, Zhou, & Achananuparp, 2008), mining time-series data (Liabotis, Theodoulidis, & Saraaee, 2006), among others. In (Kumar, 2011) some trends and new domains are explored.

Table 1. Outline of DM tasks, methods/algorithms, models/patterns, and guidance

DM Tasks	Methods/Algorithms	Models/Patterns	Guidance
Classification	Decision trees (ID3, C4.5, C5, CART, ...)	Tree	Information Gain/Gini Index
	Classification rules (PRISM, RIPPER, ...)	Rules	Accuracy and Coverage
	Neural Networks (multilayer perceptron, back propagation, kohonen networks)	Neural Network	Error
	Support Vector Machines	Maximum margin hyperplane	Error
	Bayesian data analysis	Probability distribution	Conditional Probability
	Bayesian networks (K2, TAN)	Directed acyclic graph	Conditional Probability
	k-nearest neighbor	Pattern space	Distance function
Prediction	Linear regression	Linear function	Error
	Nonlinear regression	Nonlinear function	Error
	Neural Networks (multilayer perceptron, back propagation, kohonen networks)	Neural Network	Error
	Decision trees (ID3, C4.5, CART, ...)	Tree	Information Gain Gini Index
	k-nearest neighbor	Pattern space	Distance function
Clustering	Partitioning (k-means, k-medoids, ...)	Diagram	Measure of dissimilarity
	Hierarchical (BIRCH, ROCK, Chameleon)	Diagram	Measure of dissimilarity
	Model-based (EM, Kohonen networks)	Diagram	Measure of dissimilarity
Association	Association rules (Apriori, GRI, FP-growth)	Rules	Support and Confidence
Summarization	EDA	Tables, Charts	-----
	OLAP	OLAP cubes	-----

KDD, SEMMA, AND CRISP-DM

The acronym SEMMA stands for Sample, Explore, Modify, Model, Assess, and refers to the process of conducting a DM project. The SAS Institute considers a cycle with 5 stages for the process, which are, sample, explore, modify, model, and assess. SEMMA offers an easy to understand process, allowing an organized and adequate development and maintenance of DM projects. It thus confers a structure for his conception, creation and evolution, helping to present solutions to business problems as well as to find the DM business goals (Santos & Azevedo, 2005).

CRISP-DM stands for CROss-Industry Standard Process for Data Mining. It consists on a cycle that comprises six phases, which are business understanding, data understanding, data preparation, modeling, evaluation, and deployment. CRISP-DM is extremely complete and documented. All its stages are duly organized, structured and defined, allowing a project to be easily understood or revised (Chapman, et al., 2000).

A comparative study of KDD, SEMMA, and CRISP-DM is made in (Azevedo & Santos, 2008). The summary of the correspondence between the three is presented in Table 2.

Table 2. Summary of the correspondences between KDD, SEMMA and CRISP-DM

KDD	SEMMA	CRISP-DM
Pre KDD	-----	Business understanding
Selection	Sample	Data Understanding
Pre processing	Explore	
Transformation	Modify	Data preparation
Data mining	Model	Modeling
Interpretation/ Evaluation	Assessment	Evaluation
Post KDD	-----	Deployment

Azevedo & Santos, 2008, p. 185.

STANDARDS FOR DATA MINING

Some efforts are being made seeking the establishment of standards in the DM area, both by academics, and by people in the industry field.

Some of the efforts in the industrial field concern the definition of processes/methodologies that can guide the implementation of DM applications, for instance, SEMMA and CRISP-DM. Other efforts in the industrial field focus on the development of software suites for implementing some selected DM algorithms. Over the past few years several DM suites have been developed (KD-Nuggets, 2015). These suites deliver user-friendly environments that allow users to apply DM freely and easily. There are other efforts being made that intend to develop standards that will allow model representation to be platform independent. These models mainly seek portability among models obtained in different tools, and some of them are included in most of the Business Intelligence (BI) tools in the market (Azevedo & Santos, 2011).

Above all the academic efforts towards a theory for DM and KDD follow closely the theory developed by Codd for the Relational Model. A promising research line is that of Inductive Databases, as presented by (De Raedt, 2003), (Imielinski & Mannila, 1996), and (Dzeroski, 2007). According to the Inductive Databases framework, data and DM models are stored on the same database and

can be queried at the same level. The inductive query language is a fundamental issue to consider in the research. Based on this framework, some theoretical research and prototypes have been developed, as well as some research about DM Languages, such as QMBE (Azevedo & Santos, 2012) (Azevedo & Santos, 2013), DMQL (Han, Fu, Wang, Koperski, & Zaiane, 1996), MINERULE (Meo, Psaila, & Ceri, 1998), MSQL (Imielinski & Virmani, 1999), SINBAD (Kramer, et al., 2006), SPQL (Bonchi, Giannotti, Lucchesse, Orlando, & Trasarti, 2007), KDDML (Romei, Ruggieri, & Turini, 2006), XDM (Meo & Psaila, 2006), and RDM (De Raedt, 2002). Some of these efforts contribute to the integration of DM in BI systems. The dissemination of DM tools is increasing in BI as well as the acknowledgment of the relevance of its usage in enterprise BI systems.

FUTURE RESEARCH DIRECTIONS

Despite it can be considered as an established field, many open research questions still remains and many appear daily in the DM field. Examining the primary conferences and journals in the DM field, it can be concluded that the main issues for research are related to improving data preparation for DM, to developing better algorithms and methods for specific problems and applications, and to measuring the utility of the obtained models. But the range of the problems and applications rises as time goes by.

One research line, focus on the use of DM in the realm of unstructured data, such as text, the web, audio, videos, or images, among other. Also, the so called “big data” gains momentum as time goes by. Those topics are increasing in importance, and the development of more research in the area is very important. These new types of “data”, present research challenges along all the KDD process, from the data preparation to the interpretation of the obtained data mining models.

Another research line concerns the definition of standards for DM. This is a very dynamic line

of research. The main goal from the academic point of view consists in defining a DM system similar to the Database Management Systems (DBMS) defined by Codd for the relational model for databases. The DBMS allow the user to easily access and manipulate data stored in databases and a similar system is desired for DM in order to allow the user to also access and manipulate de DM models. Some promising efforts were done, but there are some important drawbacks that need to be overcome. Just to present one example, the types of models that the languages, referred in the end of the previous section, can support are almost limited to rules.

This research line is strongly connected to a novel research line: DM integration with BI systems. This is a promising research line whose main goal is to allow final business users to really access and be able to manipulate DM models in order to completely extract their inner value.

CONCLUSION

This article presents a simple but comprehensive examination of the main issues related to DM. It presents the roots of the field. It also presents a very useful summarization of the main DM tasks and the respective methods/algorithms and models/patterns. The parallelism of KDD, SEMMA, and CRISP-DM is also very useful to the field. A usually not approached theme in DM texts is hereby included: standards for DM. This is a very important topic that, in my opinion, will represent the main core of research in the DM field hereafter.

Despite sometimes the DM field is considered as an established field with little remaining for research, this article presents several lines of research containing a good deal of open research issues waiting to be explored.

REFERENCES

- Azevedo, A. (2015). Data Mining and Business Intelligence: A Comparative, Historical Perspective. In A. Azevedo & M. Santos (Eds.), *Integration of Data Mining in Business Intelligence Systems* (pp. 1–11). Hershey, PA: Business Science Reference. doi:10.4018/978-1-4666-6477-7.ch001
- Azevedo, A., & Santos, M. F. (2008). KDD, SEMMA, and CRISP-DM: a Parallel Overview. In H. Weghorn & A.P. Abraham (Eds.), *Proceedings of the IADIS European Conference on Data Mining 2008* (pp. 182–185). Amsterdam: IADIS Press.
- Azevedo, A., & Santos, M. F. (2011). A Perspective on Data Mining Integration with Business Intelligence. In A. Kumar (Ed.), *Knowledge Discovery Practices and Emerging Applications of Data Mining: Trends and New Domains* (pp. 109–129). Hershey, PA: IGI Publishing. doi:10.4018/978-1-60960-067-9.ch006
- Azevedo, A., & Santos, M. F. (2012). Closing the Gap between Data Mining and Business Users of Business Intelligence Systems: A Design Science Approach. *International Journal of Business Intelligence Research*, 3(4), 14–53. doi:10.4018/jbir.2012100102
- Azevedo, A., & Santos, M. F. (2013). Extending QMBE Language with Clustering. *International Journal of Decision Support System Technology*, 5(4), 59–77. doi:10.4018/ijdsst.2013100104
- Bonchi, F., Giannotti, F., Lucchesse, C., Orlando, S. P., & Trasarti, R. (2007). On Interactive Pattern Mining from Relational Databases. *Lecture Notes on Computer Science*: vol. 4747. Knowledge Discovery in Inductive Databases - 5th International Workshop, KDID 2006 (pp. 42–62). Berlin: Springer-Verlag. doi:10.1007/978-3-540-75549-4_4

- Brachman, J. R., & Anand, T. (1996). The Process of Knowledge Discovery in Databases. In *Advances in Knowledge Discovery and Data Mining* (pp. 37–57). Menlo Park, CA: AAAI Press/The MIT Press.
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). *CRISP-DM 1.0 - Step-by-step Data Mining Guide. Technical Report: 1.0*. CRISP-DM Consortium.
- Chen, M., Han, J., & Yu, P. S. (1996). Data Mining: An Overview from a Database Perspective. *IEEE Transactions on Knowledge and Data Engineering*, 8(6), 866–883. doi:10.1109/69.553155
- De Raedt, L. (2002). Data Mining as Constraint Logic Programming. In *Lecture Notes on Artificial Intelligence: Vol. 2408. Computational Logic: Logic Programming and Beyond - Essays in Honour of Robert A. Kowalski - Part II* (pp. 526–547). Berlin: Springer-Verlag. doi:10.1007/3-540-45632-5_20
- De Raedt, L. (2003). A perspective on Inductive Databases. *SIGKDD Explorations*, 4(2), 6–77.
- do Prado, H. A., & Ferneda, E. (2008). *Emerging Technologies of Text Mining: Techniques and Applications*. Hershey, PA: IGI Publishing. doi:10.4018/978-1-59904-373-9
- Dzeroski, S. (2007). Towards a general framework for Data Mining. In *Lecture Notes in Computer Science: Vol. 4747. Knowledge Discovery in Inductive Databases - 5th International Workshop, KDID 2006* (pp. 259–300). Berlin: Springer-Verlag. doi:10.1007/978-3-540-75549-4_16
- Fayyad, U. M. (1997). Data Mining and Knowledge Discovery: Making Sense out of Data. *IEEE Expert*, 11(5), 20–25. doi:10.1109/64.539013
- Fayyad, U. M., Piatetski-Shapiro, G., & Smyth, P. (1996). From data mining to Knowledge Discovery: an overview. In *Advances in Knowledge Discovery and Data Mining* (pp. 1–34). Menlo Park, CA: AAAI Press/The MIT Press.
- Han, J., Fu, Y., Wang, W., Koperski, K., & Zaiane, O. (1996). DMQL: A Data Mining Language for Relational Databases. *Proceedings of the SIGMOD'96 Workshop on Research Issues on Data Mining and Knowledge Discovery*, 27–34.
- Han, J., & Kamber, M. (2012). *Data Mining: Concepts and Techniques* (3rd ed.). San Francisco, CA: Morgan Kaufman Publisher. doi:10.1007/978-1-4419-1428-6_3752
- Hand, D., Mannila, H., & Smyth, P. (2001). *Principles of Data Mining*. Cambridge, MA: The MIT Press.
- Imielinski, T., & Mannila, H. (1996). A Database Perspective on Knowledge discovery. *Communications of the ACM*, 39(11), 58–64. doi:10.1145/240455.240472
- Imielinski, T., & Virmani, A. (1999). MSQL: A Query Language for Database Mining. *Data Mining and Knowledge Discovery*, 3(4), 373–408. doi:10.1023/A:1009816913055
- KDNuggets. (2015). *Software Suites for Data Mining, Analytics, and Knowledge Discovery*. Obtido em November 2015, 10 from <http://www.kdnuggets.com/software/suites.html>
- Kramer, S., Aufschild, V., Hapfelmeier, A., Jarasch, A., Kessler, K., Reckow, S.,..... (2006). Inductive Databases in the Relational Model: the Data as the Bridge. In *Lecture Notes on Computer Science: Vol. 3933. Knowledge Discovery in Inductive Databases - 4th International Workshop - KDID2005* (pp. 124–138). Berlin: Springer-Verlag. doi:10.1007/11733492_8
- Kumar, A. (2011). *Knowledge Discovery Practices and Emerging Applications of Data Mining: Trends and New Domains*. Hershey, PA: IGI Publishing. doi:10.4018/978-1-60960-067-9
- Larose, D. T., & Larose, C. D. (2005). *Discovering Knowledge in Data: an Introduction to Data Mining* (2nd ed.). Hoboken, NJ: John Wiley & Sons.

Liabotis, I., Theodoulidis, B., & Saraee, M. (2006). Improving Similarity Search in Time Series Using Wavelets. *International Journal of Data Warehousing and Mining*, 2(2), 55–81. doi:10.4018/jdwm.2006040103

Markov, Z., & Larose, D. T. (2007). *Data Mining the Web: Uncovering Patterns in Web Content, Structure, and Usage*. Hoboken, NJ: Wiley-Interscience. doi:10.1002/0470108096

Meo, R., & Psaila, G. (2006). An XML-Based Database for Knowledge Discovery. In *Lecture Notes in Computer Science: Vol. 4254. Current Trends in Database Technology - EDTB 2006 Workshops* (pp. 814–828). Berlin: Springer-Verlag.

Meo, R., Psaila, G., & Ceri, S. (1998). An extension to SQL for mining Association Rules. *Data Mining and Knowledge Discovery*, 2(2), 195–224.

Myatt, G. J., & Johnson, P. W. (2007). *Making Sense of Data - a Practical Guide to Exploratory Data Analysis and Data Mining*. Hoboken, NJ: John Wiley & Sons.

Nlenanya, I. (2009). Building an Environmental GIS Knowledge Infrastructure. In H. Rahman (Ed.), *Data Mining Applications for Empowering Knowledge Societies* (pp. 262–279). Hershey, PA: IGI Publishing. doi:10.4018/978-1-59904-657-0.ch014

Romei, A., Ruggieri, S., & Turini, F. (2006). KDDMLA Middleware language and System for Knowledge Discovery in Databases. *Data & Knowledge Engineering*, 57(2), 179–200. doi:10.1016/j.datak.2005.04.007

Santos, M. F., & Azevedo, C. S. (2005). *Data Mining - Descoberta de Conhecimento em Bases de Dados*. Lisboa, Portugal: FCA - Editora de Informática.

Simoudis, E. (1996). Reality Check for Data Mining. *IEEE Expert*, 11(5), 26–33. doi:10.1109/64.539014

Witten, I. H., & Frank, E. (2011). *Data Mining - Practical Machine Learning Tools and Techniques* (3rd ed.). San Francisco, CA: Morgan Kaufmann Publishers.

Ye, N. (2003). *The Handbook of Data Mining*. Mahwah, NJ: Lawrence Erlbaum Associates.

Zhang, X., Hu, X., Xia, J., Zhou, X., & Achananuparp, P. (2008). A Graph-Based Biomedical Literature Clustering Approach Utilizing Terms Global and Local Importance Information. *International Journal of Data Warehousing and Mining*, 4(4), 84–101. doi:10.4018/jdwm.2008100105

ADDITIONAL READING

Botta, M., Boulicaut, J., Masson, C., & Meo, R. (2004). Query Languages Supporting Descriptive Rule Mining: A comparative Study. In Meo, R.; Lanzi, P. L. & Klemettinen, M. (Eds.), *Lecture Notes on Artificial Intelligence: Vol. 2682. Database Support for Data Mining Applications - Discovering Knowledge with Inductive Queries* (pp. 24–51). Berlin, Heidelberg: Springer-Verlag. doi:10.1007/978-3-540-44497-8_2

Calders, T., Goethals, B., & Prado, A. (2006). Integrating Pattern Mining in Relational Databases. In Fürnkranz, J.; Scheffer, T. & Spiliopoulou, M. (Eds.), *Lecture Notes on Artificial Intelligence: Vol. 4213. Knowledge Discovery in Databases - 10th European Conference on Principles and Practice of Knowledge Discovery in Databases - PKDD2006* (pp. 454–461). Berlin, Heidelberg: Springer-Verlag. doi:10.1007/11871637_43

Calders, T., Lakshmanan, L. V. S., Ng, R. T., & Paredaens, J. (2006). Expressive Power of an Algebra for Data Mining. *ACM Transactions on Database Systems*, 31(4), 1169–1214. doi:10.1145/1189769.1189770

- Catania, B., Maddalena, A., Mazza, M., Bertino, E., & Rizzi, S. (2004). A Framework for Data Mining Pattern Management. In Boulicaut, J.; Esposito, F. & Giannotti, F. (Eds.), *Lecture Notes on Artificial Intelligence: Vol. 3202. Knowledge Discovery in Databases - 8th European Conference on Principles and Practice of Knowledge Discovery in Databases - PKDD2004* (pp.87-98). Berlin, Heidelberg: Springer-Verlag. doi:10.1007/978-3-540-30116-5_11
- Data Mining Group. (2009). Predictive Model Markup Language (PMML). Data Mining Group Portal, 2009. Retrieved August, 1st 2009, from <http://www.dmg.org/>
- Fayyad, U. M., Piatetski-Shapiro, G., Smyth, P., & Uthurusamy, R. (1996). *Advances in knowledge discovery and data mining*. Menlo Park, California: AAAI Press/The MIT Press.
- Fromont, É., Blockeel, H., & Struyf, J. (2007). Integrating Decision Tree Learning into Inductive Databases. In Dzeroski, S. & Struyf, J. (Eds.), *Lecture Notes in Computer Science: Vol. 4747. Knowledge Discovery in Inductive Databases - 5th International Workshop, KDID 2006* (pp. 81-96). Berlin, Heidelberg: Springer-Verlag. doi:10.1007/978-3-540-75549-4_6
- Fung, G., & Stoeckel, J. (2007). SVM feature selection for classification of SPECT images of Alzheimers disease using spatial information. *Knowledge and Information Systems*, 11(2), 243–258. doi:10.1007/s10115-006-0043-5
- Gago, P., Fernandes, C., Pinto, F., & Santos, M. F. (2009). INTCare: On-line Knowledge Discovery in the Intensive Care Unit. INES'09 Proceedings of the IEEE 13th international conference on Intelligent Engineering Systems, 143-148.
- Gao, J., Cao, Y., Qi, Y., & Hu, J. (2005). Building Innovative Representations of DNA Sequences to Facilitate Gene Finding. *IEEE Intelligent Systems*, 20(6), 34–39. doi:10.1109/MIS.2005.100
- Gokhale, M., & Aslandogan, Y. A. (2003). A Visualization Oriented Data Mining Tool for Biomedical Images. In Smari, W. W. & Memon, A. M. (Eds.), *IEEE International Conference on Information Reuse and Integration, 2003, IRI 2003*, 219-226. Las Vegas, NV: IEEE Press. doi:10.1109/IRI.2003.1251417
- Gurbaxani, B., & Mallick, P. (2005). Finding Protein Domain Boundaries: An Automated, Non-Homology-Based Method. *IEEE Intelligent Systems*, 20(6), 26–33. doi:10.1109/MIS.2005.106
- Hsu, C., Chung, H., & Huang, H. (2004). Mining Skewed and Sparse Transaction Data for Personalized Shopping Recommendation. *Machine Learning*, 57(1), 35–59. doi:10.1023/B:MACH.0000035471.28235.6d
- Hu, X. (2005). A Data Mining Approach for Retailing Bank Customer Attrition Analysis. *Applied Intelligence*, 22(1), 47–60. doi:10.1023/B:APIN.0000047383.53680.b6
- Jamil, H. M. (2004). Declarative Data Mining Using SQL3. In Meo, R.; Lanzi, P. & Klemettinen, M. (Eds.), *Lecture Notes on Artificial Intelligence: Vol. 2682. Database Support for Data Mining Applications - Discovering Knowledge with Inductive Queries* (pp. 52-75). Berlin, Heidelberg: Springer-Verlag. doi:10.1007/978-3-540-44497-8_3
- John, G. H., Miller, P., & Kerber, R. (1996). Stock Selection Using Rule Induction. *IEEE Expert*, 11(5), 52–58. doi:10.1109/64.539017
- KDNuggets. (2011). Data Mining Community's Resource. Retrieved from <http://www.kdnuggets.com/>
- Kriegel, H., Borgwardt, K. M., Kröger, P., Pryakhin, A., Schubert, M., & Zimek, A. (2007). Future Trends in Data Mining. *Data Mining and Knowledge Discovery*, 15(1), 87–97. doi:10.1007/s10618-007-0067-9

- Lin, Y., Tsai, K., Shiang, W., Kuo, T.-C., & Tsai, C.-H. (2009). Research on using ANP to establish a performance assessment model for business intelligence systems. *Expert Systems with Applications*, 36(2), 4135–4146. doi:10.1016/j.eswa.2008.03.004
- Linoff, G. S. (2008). Survival Data Mining Using Relational Databases. *Business Intelligence Journal*, 13(3), 20–30.
- Linoff, G. S., & Berry, M. J. (2011). *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management*. Boulevard, IN: Wiley Publishing.
- Malerba, D., Appice, A., & Vacca, N. (2002). SDMOQL: An OQL-based Data Mining Query Language for Map Interpretation Tasks. *Proceedings of the EDBT 2002 Workshop on 'Database Technologies for Data Mining*, 3-18.
- Mannila, H. (2000). Theoretical Frameworks for Data Mining. *SIGKDD Explorations*, 1(2), 30–32. doi:10.1145/846183.846191
- March, S. T., & Hevner, A. R. (2007). Integrated decision support systems: A data warehousing perspective. *Decision Support Systems*, 43(3), 1031–1043. doi:10.1016/j.dss.2005.05.029
- Martin, J., Gibrat, J., & Rodolphe, F. (2005). Choosing the Optimal Hidden Markov Model for Secondary-Structure Prediction. *IEEE Intelligent Systems*, 20(6), 19–25. doi:10.1109/MIS.2005.102
- Mielikäinen, T. (2004). Inductive Databases as Ranking. In Kambayashi, Y.; Mohania, M. & Wöb, W. (Eds.), *Lecture Notes on Computer Science: Vol. 3181. Data Warehousing and knowledge Discovery - 6th International conference DaWak2004* (pp. 149-158). Berlin, Heidelberg: Springer-Verlag. doi:10.1007/978-3-540-30076-2_15
- Nijssen, S., & De Raedt, L. (2007). IQL: A Proposal for an Inductive Query Language. In Dzeroski, S. & Struyf, J. (Eds.), *Lecture Notes in Computer Science: Vol. 4747. Knowledge Discovery in Inductive Databases - 5th International Workshop, KDID 2006* (pp. 189-209). Berlin, Heidelberg: Springer-Verlag. doi:10.1007/978-3-540-75549-4_12
- North, M. A. (2012). Data Mining for the Masses. Retrieved from <https://sites.google.com/site/dataminingforthemasses/>
- Object Management Group. (2008). Knowledge Discovery Model (KDM). KDM Portal, 2006-2008. Retrieved from <http://kdmanalytics.com/kdm/index.php>
- Pal, N. R., & Jain, L. (2005). *Advanced Techniques in Data Mining and Knowledge Discovery*. London, UK: Springer-Verlag. doi:10.1007/1-84628-183-0
- Pan, J., Yang, Q., Yang, Y., Li, L., Li, F. T., & Li, G. W. (2007). Cost-Sensitive-Data Preprocessing for Mining Customer Relationship Management Databases. *IEEE Intelligent Systems*, 22(1), 46–51. doi:10.1109/MIS.2007.7
- Piatetsky-Shapiro, G. (2007). Data Mining and Knowledge Discovery 1996 to 2005: Overcoming the Hype and Moving from university to business and analytics. *Data Mining and Knowledge Discovery*, 15(1), 99–105. doi:10.1007/s10618-006-0058-2
- Pinto, F., Gago, P., & Santos, M. F. (2006). Data Mining as a New Paradigm for Business Intelligence in Database Marketing Projects. In Manolopoulos, Y., Filipe, J., Constantopoulos, P. & Cordeiro, J. (Eds.), *ICEIS 2006 - Proceedings of the Eighth International Conference on Enterprise Information Systems, Databases and Information Systems Integration*, 144-149. Paphos, Cyprus: INSTICC.

- Quintela, H., Santos, M. F., & Cortez, P. (2007). Real-Time Intelligent Decision Support System for Bridges Structures Behavior Prediction. In J. Neves, M. F. Santos, & J. Machado (Eds.), LNAI: Vol. 4874. *Proceedings of the 13th Portuguese Conference on Artificial Intelligence, EPIA 2007* (pp. 124-132). Berlin Heidelberg, Germany: Springer-Verlag. doi:10.1007/978-3-540-77002-2_11
- Rahman, H. (2009). *Data Mining Applications for Empowering Knowledge Societies*. Hershey, New York: IGI Publishing. doi:10.4018/978-1-59904-657-0
- Rantza, R. (2004). Frequent Itemset Discovery with SQL Using Universal Quantification. In Meo, R.; Lanzi, P. & Klemettinen, M. (Eds.), *Lecture Notes on Artificial Intelligence: Vol. 2682. Database Support for Data Mining Applications - Discovering Knowledge with Inductive Queries* (pp. 194-213). Berlin, Heidelberg: Springer-Verlag. doi:10.1007/978-3-540-44497-8_10
- Santos, M., Pereira, J., & Silva, Á. (2005). A Cluster Framework for Data Mining Models: an applications to intensive medicine. In Chen, C., Filipe, J., Seruca, I. & Cordeiro, J. (Eds.), *Proceedings of the 7th International Conference on Enterprise Information Systems, ICEIS 2005*, 163-168. Miami, USA: INSTICC.
- Santos, M. F., Cortez, P., Pereira, J., & Quintela, H. (2006). Corporate Bankruptcy Prediction Using Data Mining Techniques. In Zanasi, A.; Brebbia, C. A. & Ebecken, N. F. F. (Eds.), *WIT Transactions on Information and Communication Technologies: Vol. 37. Data Mining VII: Data, Text and Web Mining and their Business Applications* (pp. 349-357). Southampton, UK: WIT Press. doi:10.2495/DATA060351
- Santos, M. F., Cortez, P., Quintela, H., Neves, J., Vicente, H., & Arteiro, J. (2005). Ecological Mining - A Case Study on Dam Water Quality. In Zanasi, A.; Brebbia, C. A. & Ebecken, N. F. F. (Eds.), *WIT Transactions on Information and Communication Technologies: Vol. 35. Data Mining VI: Data mining, Text Mining and their Business Applications* (pp. 481-489). Southampton, UK: WIT Press.
- Sarawagi, S., Thomas, S., & Agrawal, R. (2000). Integrating Association Rule Mining with Relational Database Systems: Alternatives and Implications. *Data Mining and Knowledge Discovery*, 4(2-3), 89-125. doi:10.1023/A:1009887712954
- Tang, Z., & MacLennan, J. (2005). *Data Mining with SQL Server 2005*. Indianapolis, IN: Wiley Publishing.
- Taniar, D. (2008). *Data Mining and Knowledge Discovery Technologies*. Hershey, New York: IGI Publishing. doi:10.4018/978-1-59904-960-1
- Torgo, L. (2011). *Data Mining with R: Learning with Case Studies*. Boca Raton, FL: Taylor & Francis Group.
- Vercellis, C. (2009). *Business Intelligence: Data Mining and Optimization for Decision Making*. West Sussex, UK: John Wiley & Sons. doi:10.1002/9780470753866
- Wang, H., & Wang, S. (2008). A Knowledge Management Approach to Data Mining Process for Business Intelligence. *Industrial Management & Data Systems*, 108(5), 622-634. doi:10.1108/02635570810876750
- Wu, C., Yu, L., & Jang, F. (2005). Using Semantic Dependencies to Mine Depressive Symptoms from Consultation Records. *IEEE Intelligent Systems*, 20(6), 50-59. doi:10.1109/MIS.2005.115

KEY TERMS AND DEFINITIONS

Data Mining Algorithms: A sequence of steps that allows obtaining a data mining model.

Data Mining Language: Allows users to directly manipulate data and models at the same level.

Data Mining Models: Are obtained from the application of data mining methods/algorithms. There simple models, such as rules or trees, and more complex models, such as neural networks.

Data Mining Task: What is pretended to achieve when applying Data Mining, for instance classify.

Data Mining: One of the phases of the KDD process and concerns, mainly, to the means by which the patterns/models are extracted and enumerated from data. Many times is identified with the complete KDD process.

Inductive Database: Data and data mining models are stored at the same level on the same database and can be queried.

Standard: That obeys to settled parameters.

Data Mining and the KDD Process

D**Ana Funes***Universidad Nacional de San Luis, Argentina***Aristides Dasso***Universidad Nacional de San Luis, Argentina*

INTRODUCTION

The constant search for diseases causes, the improvement of automatic diagnoses methods, financial data analytics and market tendencies, among others, are only some of the innumerable applications where analysis and discovery of new patterns have fuelled the research and development of new methods, all related to machine intelligence, knowledge extraction from what is now being called ‘big data’, Knowledge Discovery in Databases or KDD, and Data Mining.

The development of these fields has benefited from the existence of large volumes of data proceeding from the most diverse sources and domains, e.g. entrepreneurial historical data bases, medical data bases, biological data bases, astronomical data, etc. KDD process and methods of Data Mining allows for the discovery of knowledge in data that is hidden to humans, particularly when data volumes are large or even extremely large, presenting the knowledge extracted under different ways: rules, equations, decision trees, etc., and helping to answer questions such as *what are the groups from a population of individuals with common characteristics?*, *is this client reliable?*, *is this e-mail spam?*, etc.

Answers to these questions, as well as to many others, are different from the traditional answers obtained from queries in On Line Transactions Processing (OLTP), where the information is not hidden neither is discovered, but it is presented summarized in an agreed format or report. They also differ from information proceeding from Online Analytical Processing (OLAP), which can

be presented in different perspectives or aggregated in different ways and not just summarized as in OLTP, and that can even escalate to the use of big data, where OLTP fails. However, both of these methods are not capable of discover new knowledge neither producing new patterns and rules as is the case with the KDD process.

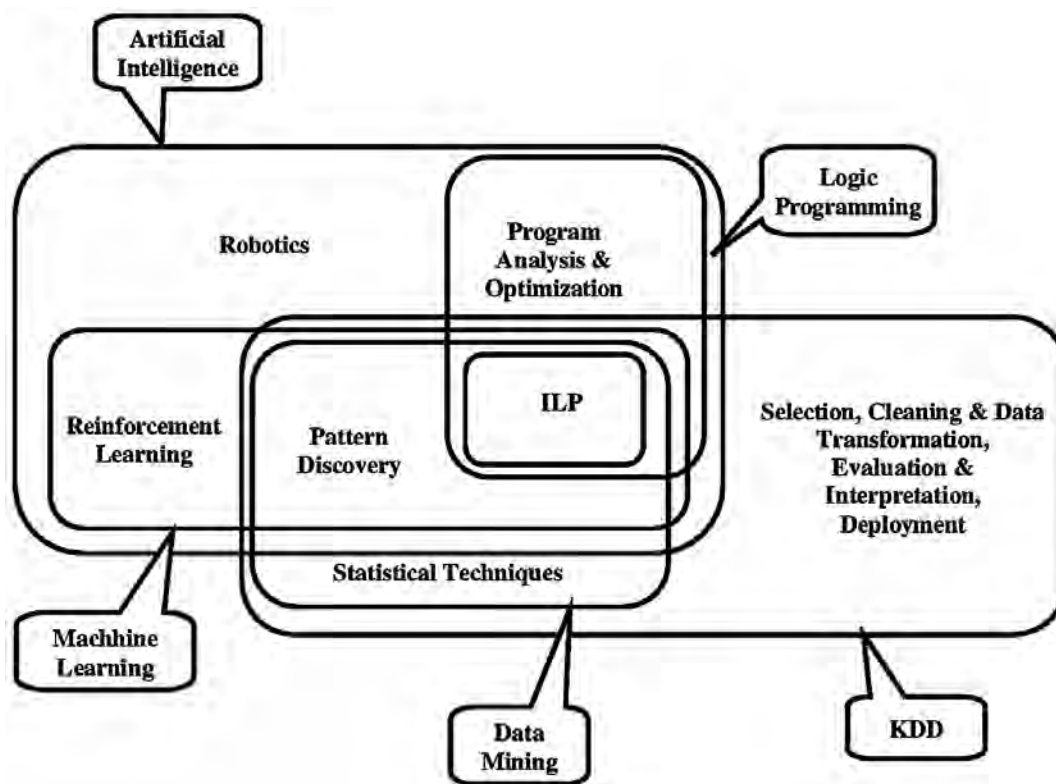
In this chapter, an overview of the KDD process and all its stages is given, including Data Selection, Cleaning, etc., with especial attention to the phase of Data Mining, its tasks and methods, as well as its relation to other areas such as Machine Learning, Inductive Logic Programming (ILP), Statistics, etc. A discussion of a possible classification of Data Mining methods is also given as well as an overview of future challenges in the field.

BACKGROUND

There exists some confusion in the use of the terms of *Knowledge Discovery in Databases* or KDD and *Data Mining*. Frequently these terms are interchanged, using Data Mining as synonym of KDD. Although they are strongly related, it is important to clarify the differences between them.

Several definitions of Data Mining can be found in the literature. Witten and Frank (2000) refers to Data Mining as the process of extraction of previously-unknown, useful and understandable knowledge from big volumes of data, which can be in different formats and come from different sources. In a much more short way, Hernández-Orallo, Ferri and Ramírez-Quintana (2004) define Data Mining as the process of converting data

Figure 1. Data mining and its relationships with other fields



into knowledge. Sometimes Data Mining is also referred by many other names including *knowledge extraction*, *information discovery*, *information harvesting*, *data archeology*, and *data pattern processing* (Fayyad et al, 1996a).

The notion of Data Mining is not new. Since the 60s, other terms as *Data Fishing* or *Data Dredging* have been used by statisticians to refer to the idea of finding correlations in data without a previous hypothesis as underlying causality. However, it is not until the late 80s that Data Mining became a discipline of Computer Science and scientific community adopted the term. In fact, as Witten and Frank (2005) point out, *the first book on data mining appeared in 1991 (Piatetsky-Shapiro and Frawley, 1991)—a collection of papers presented at a workshop on knowledge discovery in databases in the late 1980s.*

Data Mining is a branch of Artificial Intelligence, closely related to Machine Learning, where Machine Learning provides the technical

basis for Data Mining (Witten and Frank, 2005). Data Mining deals with *inductive learning* in a practical and not theoretical way (as Machine Learning does), making use of tools provided by Machine Learning. It applies Machine Learning techniques as well as other statistical and algebraic techniques to find structural patterns hidden in data, with the main objective of describing data or making predictions from them.

Artificial Intelligence comprises not only Machine Learning but other disciplines such as *Robotics*, *Logic Programming* and *Inductive Logic Programming* (ILP), which is a field of Machine Learning and Data Mining. Figure 1 shows a schematic view of the relationships between Data Mining, Machine Learning, Artificial Intelligence and other associated disciplines.

As it can be seen in Figure 1, Data Mining adopts techniques and methods not only from Machine Learning —area with which is closely related— but also from Statistics.

According to Tom Mitchell (1997), Machine Learning provides techniques; some of which are used by Data Mining for the development of software that is able to improve with experience. However, a valid question is *how a computer can improve with experience?* One answer to this question could be *by learning*. But, then, a new question arises: *can a computer learn?* Technically, when talking about *learning*, the only objective way of answer to this question is by using some kind of metric that allows measuring the performance when achieving a given task, and then observing how much better the same task has been solved after learning. For example, a set of data with information about bank customers, who have received a loan from the bank and have been classified as reliable/no reliable according with their loan payment history, could be used for learning the concept *reliable*. Later, an indicator of how much this concept has been learnt by the computer could be taken by measuring its capacity for correctly classifying new customers.

Machine Learning focus on the study of inductive learning; this inductive process consists in searching patterns hidden in a given evidence; once a model has been learnt, it can be used in the future to make predictions over new evidence or to describe new instances.

Although computers cannot already learn as human beings do, there is an important number of algorithms that work properly to solve certain tasks, e.g. image recognition, sound recognition, clustering, classification, categorization, among others.

Many authors, e.g. Hernández et al. (2004), Fayyad et al. (1996a), place Data Mining as part of a larger process called Knowledge Discovery in Databases (usually referred by its acronym KDD). This process involves a number of sub processes such as data selection and integration, preprocessing, Data Mining, and evaluation and interpretation of results. Therefore, in this context, Data Mining is viewed as a sub-process within the overall KDD process, which is concerned with the discovery of knowledge hidden in databases.

According to Klossgen and Zytkow (1996), KDD applies the knowledge discovery process to databases. Perhaps the most accepted KDD definition is Fayyad et al.'s (1996b), who define KDD as *a non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data*, giving a broader and more general application domain for the knowledge extraction process.

DATA MINING: A PHASE IN KDD PROCESS

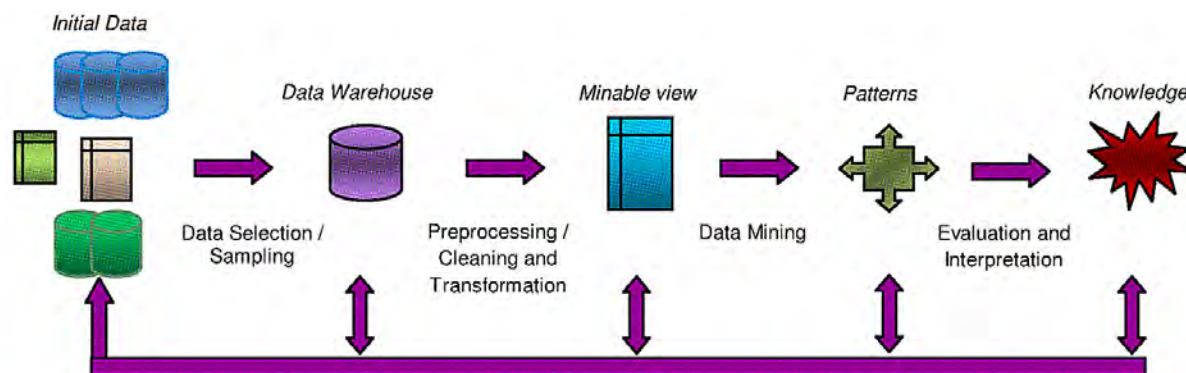
The KDD process is an iterative process that consists in the selection, cleaning and transformation of data coming not only from traditional databases but also from other heterogeneous sources, such as plain text, data warehouses, images, sound, etc., aimed to apply data mining algorithms to them in order to discover valid, novel, potentially useful, and understandable hidden patterns. Once patterns are obtained, they are interpreted and evaluated, going back if necessary to previous phases for a new iteration.

Figure 2. shows a graphical representation of the KDD process. Some authors decompose it in more phases; nevertheless, its activities are basically the same.

During the first phase (*Data Selection/Sampling*) of the KDD process, activities that have to do with data collection and integration take place. Frequently, data are collected from the most diverse sources (organizations, departments, databases, etc.); hence they do not share the same format and structure. In these cases, an activity of data integration is required. This has given place to the concept of *data warehouse*, a technology used by organizations to collect data from different sources for the purpose of analysis and decision making (e.g. evolution forecast, organizational analysis, and strategies' design).

Unlike databases, mainly used by applications based in traditional processing of data, known as OLTP (On-Line Transaction Processing), data

Figure 2. The KDD process



warehouses are operated by OLAP (On-Line Analytical Processing) tools. Data warehouses have a different data organization from databases, adopting a multidimensional data model.

Like OLAP, Data Mining techniques can make use of data from data warehouses, but as Hernández-Orallo et al. (2004) express, there is an important difference between OLAP and Data Mining: an OLAP tool allows to aggregate detailed information getting reports and views in real time, besides which, it permits to verify patterns and hypothesis suggested by the user, i.e. it is essentially a *deductive* process. On the contrary, Data Mining is an *inductive* process that uses data to infer patterns. Nevertheless, let us note that OLAP tools can also be used in the first stages of the KDD process for a better understanding of data.

A second phase of *Preprocessing/Cleaning and Transformation* is sometimes needed. During this stage, a *minable view* of the previously integrated data is created. A minable view is a subset of the integrated data, which will be the input to the data mining algorithms to be applied. A minable view eliminates unnecessary data, which are irrelevant for the data mining task. To that effect, activities such as creation of new attributes, selection of relevant attributes, attribute transformation and reduction in the number of observations to be considered take place. Sometimes, missing values and *outliers* (anomalous values) are also eliminated, especially when the algorithm used is sensible to their presence.

Once the minable view is ready, the *Data Mining* phase can proceed. In this phase, decisions about the method or technique (e.g. decision trees) to adopt for solving a task (e.g. classification) must be taken. In this phase the algorithms to be applied to the minable view (e.g. CART, C4.5) must be chosen. In the next section an overview of Data Mining's tasks, methods and algorithms is given.

The last phase of the KDD process, showed in the picture as *Evaluation and Interpretation* of patterns, is where the model learnt from data mining is evaluated, after which a new iteration begins starting from one of the previous stages whenever it is necessary.

Evaluation is a very important process since it involves determining which of the models among several alternatives is going to be kept for solving the problem at hand. Therefore, techniques to evaluate how well the different methods work to solve the problem and to compare them are mandatory.

Model evaluation techniques and measures depend on the kind of task to solve (descriptive/predictive). For example, in predictive tasks, such as classification, evaluation techniques split the evidence in sets for training and evaluation. This is done to guarantee that the model evaluation measure obtained from the test set is an independent measure, avoiding too optimistic estimations. For example, a typical evaluation measure for classification models is *error rate/accuracy*, defined as the proportion of incorrectly/correctly classified instances over the total number of instances.

The most simplistic evaluation technique for predictive task is *simple validation*, where the evidence is split in two sets, one for testing (*test set*) and the other for training (*training set*). However, there exist most sophisticated techniques such as *cross validation*, *n-fold cross validation*, useful when the evidence is scarce, or *bootstrapping* (Efron and Tibshirani, 1993) (Kohavi, 1995a).

In descriptive tasks, such as clustering and association rules, different measures are used. For example, *coverage* (or *support*) and *accuracy* (or *confidence*) are used for association rules.

LEARNING TASKS, METHODS AND ALGORITHMS IN DATA MINING

Data Mining Tasks

In Data Mining, as in many other disciplines, there exist families of problems. Each of these families, in Data Mining, is called *learning task* and it can be solved by means of diverse *methods* and *algorithms*. *Classification*, *regression*, *clustering*, *correlations* and *association rules* can be mentioned among the most important data mining tasks.

As a result of each task, according to the problem it solves, a prediction or a description is obtained, giving place to *predictive* and *descriptive tasks*. Predictive tasks, as its name indicates, estimate unknown or future values, while descriptive tasks search hidden patterns in data as a way of explaining or describing them.

One of the most common predictive tasks is *classification*, while correlations, association rules, functional dependencies and clustering are descriptive tasks.

A *classification task* consists in labeling each new instance with a tag (the *class*), from a set of possible classes. Classification predicts the class of an instance from the values of its attributes. For example, a model used to classify new messages as spam/no spam that previously have learned the concept of spam from old messages labeled as spam/no spam.

There are variations to classification like *categorization*, where an instance can be labeled with more than one class. Other variants to classification are *soft classification* and *probabilistic classification*. A soft classifier not only assigns a class but also an estimation of the degree of certainty. Probabilistic classification is a generalization of soft classification; here, for each instance, the classifier predicts a probability distribution for the set of possible classes, i.e. for each instance it predicts the probability of belonging to a class. Let us note that, when the number of classes is two, a probabilistic classifier is equivalent to a soft classifier.

Another descriptive task is *preference learning*, where a ranking is learnt. The learning process is carried out not from a set of instances as in classification but from a set of sequences of instances, where the order in a sequence represents the preference order of the instances in the sequence. For example, given a set of candidates for a job position, to get a preference order of the candidates based on their attribute's values and previous preference rankings from which to learn the new preference order.

Regression is also another important predictive task. In this case, a real value is predicted, i.e., a real function is learnt. An example of a regression task is the productivity estimation for a company based on past productivity values. When a real value is predicted for the future, this task is referred as *estimation*, and when the real value predicted is between two known values it is referred as *interpolation*.

On the other hand, *Correlations* are a descriptive task. They are useful to determine whether two numeric variables are correlated, i.e. whether they have a similar behavior. When there is a positive correlation, both variables increase and decrease at the same time; if the correlation is negative one decreases when the other increases. A correlation not only can be positive or negative but it also can be strong or weak or nonexistent. This task is used in data mining to determine redundant attributes or dependencies among them.

Association rules is also a descriptive task that, as correlations, finds existing relationships among categorical (non numerical) attributes. Given evidence from a set in $A_1 \times A_2 \times \dots \times A_n$, where A_i is the domain of an attribute, an association rule is defined as an implication, usually in the form

“if $A_i = a \wedge A_j = b \wedge \dots \wedge A_k = g$ then $A_r = u \wedge A_s = v \wedge \dots \wedge A_t = w$ ”

For example, from an analysis of purchase baskets the following rule could be inferred:

“if *baby_food* = YES then *diapers* = YES”

Functional dependencies are a variation of association rules, where only one attribute is determined from a subset of the remaining attributes. For example, a functional dependency could determine that a person is a homeowner from his/her age, income and marital status.

Finally, the descriptive task of *clustering* consists in partitioning a set of instances in natural groups called clusters. Originally, the central concept for clustering was the use of a numeric measure of similarity between the instances, creating clusters where the similarity among the members of a cluster is maximized and the similarity between the members of different clusters is minimized. A different approach to clustering is *conceptual clustering* proposed by Michalski (1980), which is not based in the notion of similarity but in the concept of *conceptual cohesion* (Michalski and Stepp, 1983a). In conceptual clustering, two individuals belong to the same cluster not because they are similar (according to a similarity measure or distance) but because they are part of a same concept or, in other words, they are described by a same concept. These two different views of clustering have given place to different approaches to this task: *similarity based clustering* and *conceptual clustering*.

The main application of clustering is in guessing that new instances that have been considered to be members of a cluster have similar behavior

or characteristics that the members of that cluster. For example, an on-line shop could make use of clustering to create groups of customers according to their purchase preferences, and then when a customer buys an article the system offers other articles based on the preferences of others customers in the same cluster.

Data Mining Methods and Algorithms

There is an important number of *learning methods* that support the previously described data mining tasks. For example, a classification task can be solved using different methods, e.g. neural networks or decision trees. At the same time, for each particular method different algorithms supporting it can be found. For example, there are several decision trees algorithms, such as ID3 (Quinlan, 1986) or its extension C4.5 (Quinlan, 1993). Many of them can be used to solve different tasks, e.g. *k*-nearest neighbor algorithm can be used for classification, regression and also for clustering.

In general, data mining methods can be placed into two large categories: *supervised* and *unsupervised* methods.

In supervised methods, models are learnt from examples, where there is a *teacher* or *supervisor* (expert) who defines the classes and provides examples for each class. These examples form a training set, which is used to build a model that allows predicting the class of new examples. Some supervised methods learn a function, i.e. the class are disjoint, while others learn a relationship, where classes can be overlapped. The first corresponds to the task of classification, while the second to categorization.

In unsupervised methods, learning is by observation and discovery. Here, the examples used for learning are not labeled with a class as in supervised methods but the methods must observe the examples and recognize patterns by themselves.

Both supervised and unsupervised methods adhere to one of two inductive learning approaches: symbolic learning or similarity-based learning. In

the next section a description of both approaches and an overview of their main methods are given.

Similarity-Based Methods

Similarity-base learning methods share the concept of *similarity* and other associated concepts like the notion of *distance* and neighborhood or proximity. Some kind of mechanism is needed to compare two objects and decide if they are similar. One way is by means of the use of a *similarity function*. A similarity function is a real function that given two objects returns a real value representing how similar these two objects are; it returns higher values while more similar the objects are. For example, a similarity function frequently used in Machine Learning is the *cosine similarity* measure between two vectors of real numbers; it is often used to compare documents.

The concept of *distance* or *metric* is associated to the notion of similarity, being the distance between two objects inversely proportional to the similarity between them. Even though a distance is a particular case of a similarity function, distances have certain properties that can be advantageously used by learning techniques.

Similarity-based methods support the idea that similar objects should behave in similar way. For example, given a similarity measure, a new

instance could be classified as belonging to the class that has the most similar instance; or in a clustering task, it could be placed it into the cluster that has the instances more similar to it.

A brief description of some well-known similarity-based supervised and unsupervised methods follows.

Supervised Similarity-Based Methods

Supervised Similarity-based methods learn from a set of labeled instances employing a similarity function or a distance to make their predictions. Examples of them are *k-Nearest Neighbors*, *Fisher's discriminant*, *Support Vector Machines*, *Learning Vector Quantization*, among others.

- **k-NN (k-Nearest Neighbors):** It is very simple and probably the best-known method. The idea behind *k-NN* is that the *k* nearest neighbors' majority class is assigned to each new instance.

The main problem is determining the optimum value for *k* (see Figure 3). Usually, $k = \sqrt{n}$ where *n* is the number of examples, is an option with a theoretical base.

Figure 3. Difference between the nearest neighbor classifications in function of the chosen value of *k*

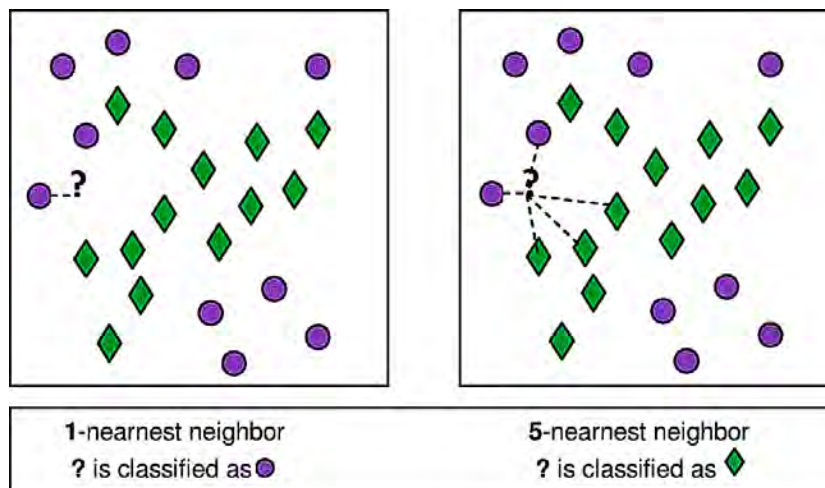
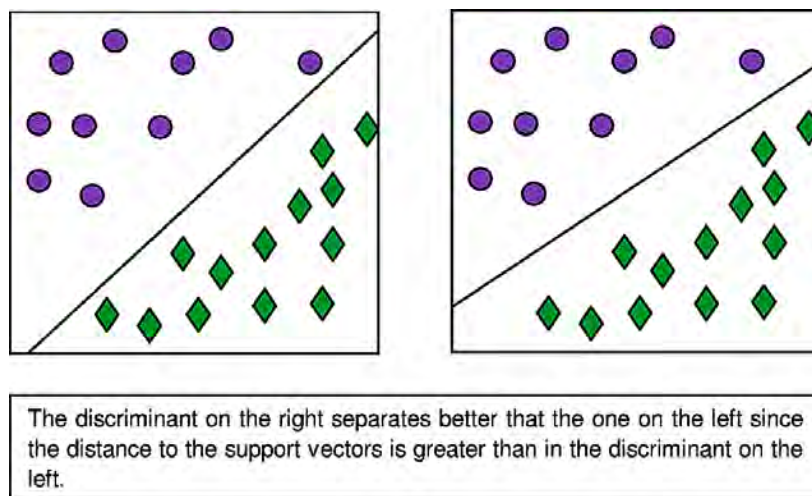


Figure 4. Linear discriminants



Any data representation can be used provided that a suitable similarity function is found.

- **Support Vector Machines (SVM):** This method searches a linear discriminant that maximizes the distance to the examples in the border of a class.

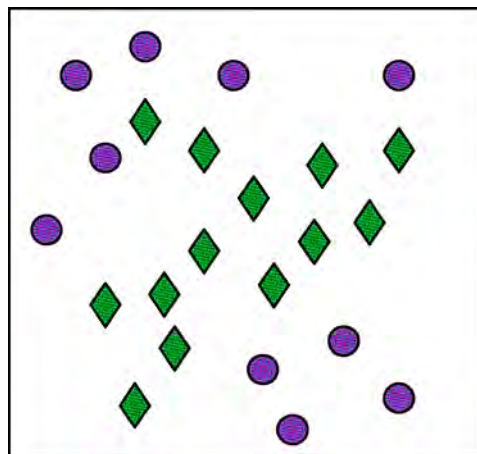
The classifier simply computes the hyperplane that perfectly separates the classes and that maximize the distance from the examples in the class border to the frontier (see Figure 4).

SVM are very efficient, even for hundred of dimensions because the linear discriminant has only to see a few points, called the *support vectors*, avoiding all the points that are not close to the frontier.

When data cannot be lineally discriminated (see Figure 5), a *kernel* function can be applied, augmenting the number of dimensions in such a way that the data are separable.

A SVM model is a representation of the examples as points in a space, transformed in such a way that the examples are clearly separated. When a new example must be classified, this is transformed by the kernel function into the same space and classified according to the side of the hyperplane where the new example falls.

Figure 5. Set of examples not lineally separable



- **Learning Vector Quantization (LVQ):** The learned model is a collection of *prototypes*¹, where a new example is classified according to its proximity to those prototypes.

Given a training set with k classes, the algorithm randomly choose k prototypes, whose positions change until a given threshold is reached. This is an iterative process, where, in first place, an example from the training set is chosen and is labeled with the class of the closer prototype. Next, the class' prototype is recomputed. Depend-

ing on the prototype's and example's labels, the prototype position will change, making closer the new prototype to the example if both labels match or the opposite if they do not match. The process stops when the change in the prototypes positions is lesser than a give threshold or there are no more training examples.

Unsupervised Similarity-Based Methods

In unsupervised similarity-based learning, the evidence is presented as a set of unlabeled elements and the objective is to find a descriptive model from the evidence, in this case using a similarity function.

Learning of descriptive models can be carried out in different ways and by multiple tasks, e.g. *clustering*, *correlations* and *factorizations*, *association rules* and *functional dependencies*.

A possible classification of clustering methods is splitting them into *hierarchical clustering* and *partitioning clustering*. A well-known example of partitioning clustering is *k-means* algorithm; examples of hierarchical clustering are *agglomerative* and *divisive hierarchical clustering*.

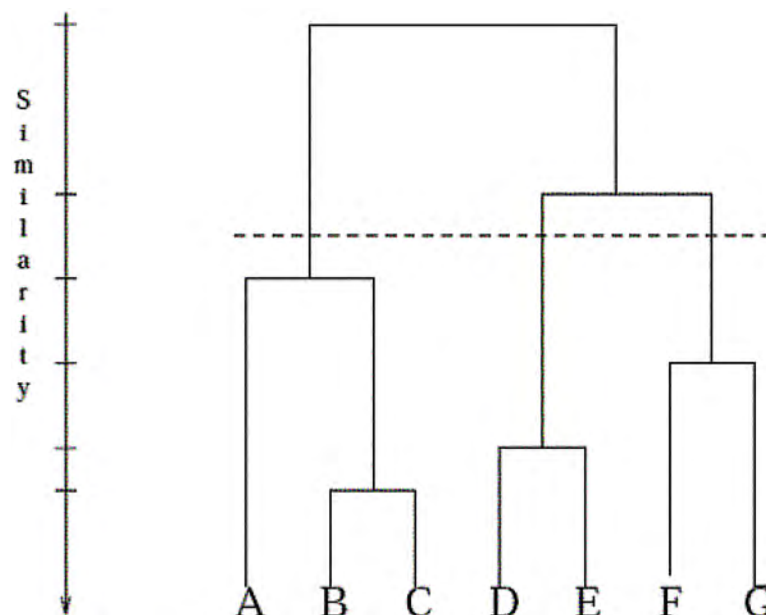
Below a brief description of similarity-based clustering is given.

- Hierarchical Clustering Algorithms:** These algorithms produce nested series of partitions of the evidence using as criterion for splitting or merging the clusters a similarity function (generally a distance). A hierarchical algorithm produces as result a *dendrogram*, which represents the nested clusters. It also shows the similarity levels where a clustering changes to a new clustering (see Figure 6).

A dendrogram can be decomposed at different levels obtaining different data partitions. For example, in Figure 6, the dotted line cuts the hierarchy at the level corresponding to the clustering formed by the groups {A, B, C}, {D, E} and {F, G}.

- Partitioning Clustering Algorithms:** Unlike hierarchical clustering algorithms, which create a clustering hierarchy, this kind of algorithms produces only one partition of the evidence.

Figure 6. Example of dendrogram produced by a hierarchical clustering algorithm



One of the most known partitioning algorithms is *k-means*. It begins the process working with an initial random partition formed by k clusters and proceeds reassigning the examples to other clusters driven by the distance between the example and the clusters' *centroids*¹. Once all the examples have been reassigned, the clusters' centroids are recomputed. This process continues until a convergence criterion is reached (e.g. the examples cannot be reassigned or the quadratic error has reached a threshold after a given number of iterations). Although *k-means* is a popular algorithm because his implementation is simple and his complexity is $O(n)$, where n is the number of examples, it has the drawback of being sensible to the selection of the initial partition.

Model-Based Symbolic Methods

In similarity-based learning, the discovered models are determined by the similarity between the elements of the evidence. This fact makes the model hard to understand by the user. This is a weakness of similarity-based methods since in many cases to have a model that is comprehensible by a human is a must; for example, a learned model should provide an explanation or description of its recommendations when refusing a loan to a customer, that goes beyond saying that the customer is classified as unreliable.

In this section some supervised and unsupervised learning methods that produce comprehensible models from the evidence are described. This family of methods or techniques is known as *Symbolic Learning*.

Supervised Model-Based Symbolic Methods

Possibly, the most well-known supervised symbolic method is *decision trees*, where a predictive model in the shape of a tree structure is learnt from the evidence. Decision tree branches represent attributes conjunctions and its leaves correspond

Table 1. Evidence used to construct the decision tree shown in Figure 7

Outlook	Temperature	Humidity	Windy	Play
sunny	hot	high	FALSE	no
sunny	hot	high	TRUE	no
overcast	hot	high	FALSE	yes
rainy	mild	high	FALSE	yes
rainy	cool	normal	FALSE	yes
rainy	cool	normal	TRUE	no
overcast	cool	normal	TRUE	yes
sunny	mild	high	FALSE	no
sunny	cool	normal	FALSE	yes
rainy	mild	normal	FALSE	yes
sunny	mild	normal	TRUE	yes
overcast	mild	high	TRUE	yes
overcast	hot	normal	FALSE	yes
rainy	mild	high	TRUE	no

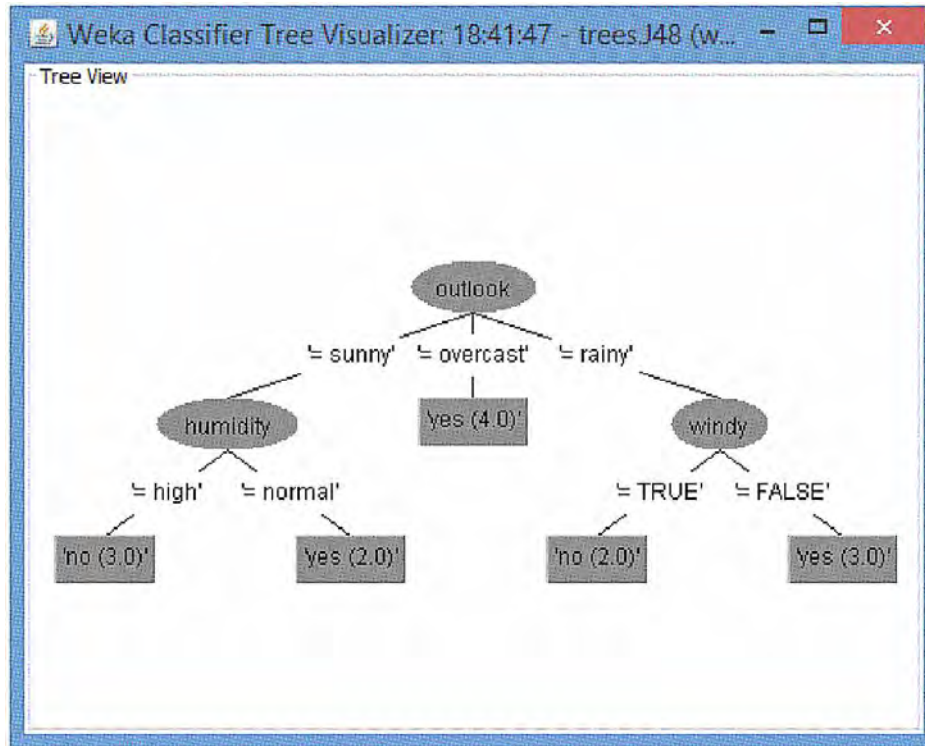
—in the case of *classification trees*—to a set of labels or classes, or to real values in the case of *regression trees*.

Figure 7 shows an example taken from (Witten and Frank, 2005) of a decision tree that has been induced from the data given in Table 1. This decision tree allows determine given a new instance whether to play a tennis match. The tree shows in its intermediate nodes the attributes outlook, humidity, and windy with their corresponding values on its arcs and the values yes/no of the class (play), on the leaves.

It can be noticed that although the attribute temperature appears in the data, this is not relevant to the decision (to play tennis or not), therefore temperature is not a node of the tree.

Whenever a new instance must be classified, the tree is traversed from the root according to the attribute values in the successive nodes until a leave is reached and the instance is classified with the value in the leave. For example, an instance with the attribute values outlook = sunny, temperature = cool, humidity = normal and windy = TRUE, will be classified with play = yes.

Figure 7. Decision tree to determine whether to play tennis



Unsupervised Symbolic Methods Based on the Model

In this section, two representative unsupervised symbolic methods, namely *Conceptual Clustering* (Michalski, 1979, 1980), (Michalski and Stepp, 1983a) and *Formal Concept Analysis* (FCA) (Ganter et al., 1999), which allow learning symbolic models from unlabeled examples, are described.

- Conceptual Clustering:** Classical distance-based unsupervised learning methods do not allow building conceptual descriptions of the discovered clusters. These methods leave to the user the problem of interpreting clusters. This is a limitation of distance-based methods since in some cases is useful not only to determine the groups but also to produce conceptual descriptions of them.

Michalski introduced, in 1980, the idea of grouping objects in categories that are described by concepts. The technique of conceptual clustering is composed of two main tasks: *structuring*, where the clusters are formed from a collection of objects and *characterization*, where concepts for each cluster of the structure are determined.

As it is explained in (Guerra-Sandón et al., 2012), the problem addressed by conceptual algorithms is of great practical usefulness. They are not limited to produce a list of objects that are in the same cluster and hence share similar characteristics or properties, but the intention of conceptual clustering is to provide additional information by giving the properties that the elements in each cluster satisfy.

Michalski was the precursor of these ideas and nowadays there are several algorithms for conceptual clustering. A conceptual clustering classification is given in (Guerra-Sandón et al., 2012); the authors classify conceptual clustering

algorithms into incremental and non incremental. Non incremental algorithms are those where all the training data are present at the same time, then the clustering is built from them; incremental algorithms process the examples one at the time, so each new example compels modifying the model obtained up to that moment. Incremental algorithms have been developed for those cases where the evidence is not completely given, i.e. is dynamic. Among the incremental algorithms, UNIMEM (Lebowitz, 1987), COBWEB (Fisher, 1987) and Galois (Carpineto and Romano, 1996) can be cited. Among the non incremental are CLUSTER/PAF (Michalski, 1979), CLUSTER/2 (Michalski and Stepp, 1983b), CLUSTER/S (Stepp and Michalski, 1986), WITT (Hanson and Bauer, 1989), ITERATE (Biswas et al., 1998), Conceptual K-Means (Ralambondrainy, 1995), LC Conceptual (Martínez-Trinidad, J. F. and Sánchez-Díaz, G., 2001) and HDCC (Funes et al., 2008), which is an hybrid between both paradigms (conceptual and distance-based).

Formal Concept Analysis. FCA allows identifying conceptual structures from data. These structures correspond to lattices of *formal concepts*, where each formal concept is a pair (*set of properties*, *set of examples*) and where the *set of properties* in the formal concept describes completely its *set of examples*. These lattices of formal concepts consist of a partial order relation defined as generalization relation between formal concepts. The main objective of FCA is offering compact descriptions of groups of data rather than a data clustering.

CONCLUSION

Our capacity for producing and storing data has notably increased during the last decades. This increment in the volume of data has gone up with our interest in exploring and analyzing

them. It is here that Data Mining has appeared as an emerging technology. Nowadays more and more enterprises, governmental, educational and all kind of organizations use Data Mining techniques for exploring, analyzing, understanding and applying the discovered knowledge extracted from big volumes of data. At the same time, an important number of both commercial and free tools have been developed; some of them oriented to particular domains, such as web mining and text mining tools, and others for general purpose mining.

FUTURE RESEARCH DIRECTIONS

Although data mining has achieved a big success and has given solution to many problems, there still exist some open research problems. In Qiang Yang and Xindong Wu (2006) ten challenging problems that give place to several research niches have been identified by consulting some of the most active researchers in data mining and machine learning. We enunciate them below as examples of possible fields of research:

- Developing a unifying theory of data mining.
- Scaling up for high dimensional data and high speed data streams.
- Mining sequence data and time series data.
- Mining complex knowledge from complex data.
- Data mining in a network setting.
- Distributed data mining and mining multi-agent data.
- Data mining for biological and environmental problems.
- Data Mining process-related problems.
- Security, privacy and data integrity.
- Dealing with non-static, unbalanced and cost-sensitive data.

REFERENCES

- Biswas, G., Weinberg, J. B., & Fisher, D. H. (1998). ITERATE: A Conceptual Clustering Algorithm for Data Mining. *IEEE Transactions on Systems, Man, and Cybernetics*, 28(2).
- Carpineto, C., & Romano, G. (1996). A lattice conceptual clustering system and its application to browsing retrieval. *Machine Learning*, 24(2), 95–122. doi:10.1007/BF00058654
- Efron, B., & Tibshirani, R. (1993). *An introduction to the bootstrap*. London: Chapman and Hall. doi:10.1007/978-1-4899-4541-9
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996a). The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM*, 39(11), 27–34. doi:10.1145/240455.240464
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996b). From data mining to knowledge discovery: an overview. In U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, & R. Uthurusamy (Eds.), *Advances in Knowledge Discovery and Data Mining*. AAAI Press.
- Fisher, D. H. (1987). Knowledge Acquisition Via Incremental Conceptual Clustering. *Machine Learning*, 2(2), 139–172. doi:10.1007/BF00114265
- Funes, A., Ferri, C., Hernández-Orallo, J., & Ramírez-Quintana, M. J. (2008). *Hierarchical Distance-based Conceptual Clustering*. In *Machine Learning and Knowledge Discovery in Databases, LNAI 5211/2008* (pp. 349–364). Heidelberg, Germany: Springer. doi:10.1007/978-3-540-87479-9_41
- Ganter, B., Stumme, G., & Wille, R. (1999). *Formal Concept Analysis, Mathematical Foundations*. Berlin: Springer. doi:10.1007/978-3-642-59830-2
- Guerra-Sandón, A., Vega-Pons, S., & Ruiz-Shulcloper, J. (2012). *Algoritmos de Agrupamientos Conceptuales: un estado del arte*. Reporte Técnico CENATAV, RNPS N°2142, ISSN 2072-6287, Junio 2012.
- Hanson, S. J., & Bauer, M. (1989). Conceptual Clustering, Categorization, and Polymorphy. *Machine Learning*, 3(4), 343–372. doi:10.1007/BF00116838
- Hernández-Orallo, J., Ramírez-Quintana, M. J., & Ferri, C. (2004). *Introducción a la Minería de Datos*. Madrid: Pearson Education S.A.
- Klosgen, W., & Zytkow, J. (1996). Knowledge discovery in databases terminology. In U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, & R. Uthurusamy (Eds.), *Advances in Knowledge Discovery and Data Mining*. Menlo Park, CA: AAAI Press/MIT Press.
- Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*. San Francisco: Morgan Kaufmann.
- Lebowitz, M. (1987). Experiments with Incremental Concept Formation: UNIMEM. *Machine Learning*, 2(2), 103–138. doi:10.1007/BF00114264
- Martínez-Trinidad, J. F., & Sánchez-Díaz, G. (2001). LC: A Conceptual Clustering Algorithm. *Proceedings of the Second International Workshop on Machine Learning and Data Mining in Pattern Recognition 2001*, (pp. 117–127). Springer-Verlag. doi:10.1007/3-540-44596-X_10
- Michalski, R. S. (1979). Conceptual Clustering: A Theoretical Foundation and a Method for Partitioning Data into Conjunctive Concepts. In Textes des exposés du Séminaire organisé par l'Institut de Recherche d'Informatique et d'Automatique. Academic Press.

Michalski, R. S. (1980). Knowledge acquisition through conceptual clustering. *Policy Analysis and Information Systems*, 4, 219–244.

Michalski, R. S., & Stepp, R. E. (1983a). Learning from Observation: Conceptual Clustering. In J. G. Carbonell, R. S. Michalski, & T. M. Mitchell (Eds.), *Machine Learning: An Artificial Intelligence Approach* (pp. 331–363). TIOGA Publishing Co.

Michalski, R. S., & Stepp, R. E. (1983b). Automated Construction of Classifications: Conceptual Clustering versus Numerical Taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-5(4), 396–410. doi:10.1109/TPAMI.1983.4767409 PMID:21869124

Mitchell, T. M. (1997). *Machine Learning*. New York, NY: McGraw-Hill.

Piatetsky-Shapiro, G., & Frawley, W. J. (Eds.). (1991). *Knowledge discovery in databases*. Menlo Park, CA: AAAI Press/MIT Press.

Quinlan, J. R. (1986, March). Induction of Decision Trees. *Machine Learning*, 1(1), 81–106. doi:10.1007/BF00116251

Quinlan, J. R. (1993). *C4.5: Programs for Machine Learning*. San Francisco, CA: Morgan Kaufmann Publishers.

Ralambondrainy, H. (1995). A conceptual version of the K-means algorithm. *Pattern Recognition Letters*, 16(11), 1147–1157. doi:10.1016/0167-8655(95)00075-R

Stepp, R. E., & Michalski, R. S. (1986). *Conceptual Clustering: Inventing Goal-Oriented Classifications of Structured Objects*. In R. S. Michalski & T. M. Mitchell (Eds.), *Machine Learning: An Artificial Intelligence Approach* (pp. 471–498). Morgan Kaufmann.

Witten, I. H., & Frank, E. (2000). *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. San Francisco, CA: Morgan Kauffman Publishers.

Witten, I. H., & Frank, E. (2005). *Data Mining Practical Machine Learning Tools and Techniques*. San Francisco, CA: Morgan Kauffman Publishers.

Yang, Q., & Wu, X. (2006). 10 Challenging Problems in Data Mining Research. *International Journal of Information Technology & Decision Making*, 5(4), 597–604. doi:10.1142/S0219622006002258

KEY TERMS AND DEFINITIONS

Classification: Inductive task where a predictive model is learnt from objects labeled with a class and whereby it is possible to predict the class of new objects.

Clustering: Inductive task where a set of unlabeled objects is partitioned into groups (clusters) and where objects in a same cluster have similar characteristics, maximizing the similarity intra cluster and minimizing the similarity inter cluster.

Data Mining: The process of extraction of implicit, previously unknown, and potentially useful knowledge from data. It uses Machine Learning, statistical and visualization techniques to discover and present knowledge in a form that is easily comprehensible to humans. It is a phase in a bigger process: the Knowledge Discovery in Databases (KDD) process.

Inductive Learning: Induction is the inference of information from data and inductive learning is a model building process where the data are analyzed to find hidden patterns.

KDD Process: The KDD process is an iterative process that consists in the selection, cleaning and transformation of data coming not only from databases but also from other heterogeneous sources, such as plain text, data warehouses, images, sound, etc., aimed to apply to them data mining algorithms in order to discover valid, novel, potentially useful, and understandable hidden patterns.

Supervised Learning: Learning process of a predictive model from a set of objects, where a supervisor define classes and supply objects of each class. Once the model has been formulated it can be used to predict the class(es) of new objects.

Unsupervised Learning: Learning process of a descriptive model (patterns) by observation and discovery from a set of unlabeled objects.

ENDNOTE

- ¹ A set *prototype* is the set element that is in the center of the elements according to its metric space; i.e. it minimizes the sum of the distances to the other elements in the set. When this element does not belong to the set, it is called centroid.

D

Data Mining to Identify Project Management Strategies in Learning Environments

Ana González-Marcos

Universidad de La Rioja, Spain

Joaquín Ordieres-Meré

Universidad Politécnica de Madrid, Spain

Fernando Alba-Elías

Universidad de La Rioja, Spain

INTRODUCTION

One of the most important organizational developments in recent years has been the significant growth of project work in different economic sectors and industries (Winter, Smith, Morris, & Cicmil, 2006). Thus, projects have become a key strategic working form. Further, it has been shown that all industries can benefit from project-based working (OPSR, 2003).

No longer just a sub-discipline of engineering, the management of projects -including program management and portfolio management- is now the dominant model in many organizations for strategy implementation, business transformation, continuous improvement and new product development (Winter et al., 2006). However, there is growing recognition that different types of projects require different approaches to their management (Müller & Turner, 2007). Furthermore, the increasing globalization of projects and project management adds intercultural challenges for project managers (Müller & Turner, 2004).

There is no doubt that management's configuration of projects affects the project's evolution. There are also factors like virtual teamwork and team building processes that are relevant to that evolution. Since effectiveness in managing projects depends on these factors, the authors

conducted this research to determine whether project performance varies according to project management and other factors. Thus, with a view to complementing other research to link project management to project success (Din, Abd-Hamid, & Bryde, 2011; Mir & Pinnington, 2014), this work considers factors such as virtual team configuration, team composition, knowledge competence, policy and strategy, project life monitoring and the level of detail implemented in managing projects that are undertaken in the learning process.

The data for this research was provided by an educational framework that was specifically designed to facilitate the learning experience of project management engineering students. The main purpose of this learning experience was to highlight how to move from simply learning content by rote to understanding, discussing and sharing (Alba-Elías, González-Marcos, & Ordieres-Meré, 2014). In this case, practitioners learned and applied by means of an experimental learning approach, a defined project management methodology that enables them to manage projects better.

Data mining and data analytics were used in this work to identify and understand the relationships between project performance and the analyzed factors. Data mining is widely applied in the educational area to predict students' performance

(Romero & Ventura, 2010; Shahiri, Husain, & Rashid, 2015). In the present research, we are much more interested in the project effectiveness that unexperienced project teams achieve, depending on specific factors. Thus, this study concentrates on using data mining for discovering patterns in project success, i.e., the performance of students as a team, instead of the performance of individual students.

The conclusions of our study can help higher education course designers, as well as teachers and students, by making clear the influence of smarter strategies in the learning process. In fact, the same benefits will help practitioners too, as they can improve their continuous learning procedures and adjust their own project management policies and strategies. Thus, the proposed research can be used as a specific decision tool of benefit in organizing projects according to specific parameters.

BACKGROUND

Project management is the application of knowledge, skills, tools and techniques to project activities to meet project requirements (PMI, 2008). The literature concerning project management includes many studies of critical success factors as can be seen in Wi & Jung (2010) or Alias, Zawawi, Yusof & Aris (2014). It is often agreed that project performance must achieve its objective and be aligned with criteria that the project stakeholders establish (Barclay, 2008). The usual metrics that are considered are cost, schedule and quality (ur Rehman Toor & Ogunlana, 2010). Indeed, the currently available standards are basically using these criteria in a process-oriented approach throughout the project's entire life cycle. Despite not being included in the three main strands of project management, sustainability has become a very important qualitative and quantitative step, particularly in the project's environmental aspects (Marcelino-Sádaba, González-Jaen, & Pérez-Ezcurdia, 2015). Thus, although sustain-

ability and project management traditionally have been tackled separately, the role of environmental management in projects is gaining more and more attention as an increasing number of publications in this field highlight (Marcelino-Sádaba et al., 2015; Økland, 2015).

Project performance has been of great interest to researchers and some authors have defined conceptual frameworks (Hewagamage & Hewagamage, 2011), as well as specific assessment models (González-Marcos, Alba-Elías, & Ordieres-Meré, 2016; Rehman, Usmani, & Al-Ahmari, 2012). Also, different studies have been conducted to investigate the impact of various factors on project performance. For instance, some studies for the Information and Communication Technology (ICT) projects, indicate that both users' knowledge of Information System and developers' knowledge of application domains have a significant impact on project performance (Byrd & Turner, 2001). In addition, insufficient personal knowledge about information system resources is a major reason for the failure of ICT projects (Tesch, Sobol, Klein, & Jiang, 2009).

However, despite the importance of technical factors, organizational and behavioral factors are considered in the literature to be the most relevant (Qureshi, Warraich, & Hijazi, 2009). Thus, all authors in the field recognize the relevance of continuous improvement and learning factors on project performance (Ojiako, Ashleigh, Chipulu, & Maguire, 2011). This applies not only to individuals, but also to the corporation. Therefore, elements such as lessons learned and project management maturity models are considered.

In this work an experiment that permits different layers of management (program and project management) to be reproduced was conducted. Thus, it was possible to study how managerial configuration of project management of virtual teams relates to project success. The main contribution of this work is the identification of quantitative relevance of described factors, as well as considering virtual teams as an additional driving factor.

DESCRIPTION OF THE ENVIRONMENT

Context

This experiment was designed to be a simulated environment in which challenge begins with a hypothetical company. It involves a new facility that the project owner intends to build in Spain to provide goods to the Spanish market. The owner is considering hiring a company to provide detailed engineering for its new plant. His owner expects to obtain the technical documentation that is necessary to authorize and build the facility. This includes both documents –specs, drawings, environmental impact assessment, etc.- and digital content like computer models and systems.

The requested companies that work as project teams in the program consist of students from three different universities - Universidad Politécnica de Madrid (UPM), Universidad de La Rioja (UR) and Universidad de León (ULE). Thus, different engineering students from different universities, different degrees and different levels of education were involved.

The experiment was conducted from September 2011 until May 2012. During this period all of the programs included different projects with different goals and durations. All of them were managed throughout with specifically designed information system that provided continuous monitoring of the progress during the period. This was the first time in all cases that practitioners were exposed to this methodology or project management theory and principles. Consequently, because of their previous knowledge, the participants can be expected to have been free of any bias.

The project enabled participants to assume various roles, ranging from project engineer (team members) to project and program manager (see Figure 1):

- Practitioners who were engaged as program managers (PGM) maintained their positions throughout the experiment. Thus,

they kept the scope for the whole program and they were responsible for the projects' evolution. Program managers were asked to produce an outline of the business case for projects that they wanted to manage in order to keep or increase the program's value. Close monitoring of project deliverables is expected, as the latter must be used as inputs for other projects or components of the final solution to be delivered to the customer.

- Project managers (PM) were responsible for defining and negotiating with the program managers the scope of the project, the deliverables to be produced by the project, etc. Project managers had the authority necessary for day-to-day management of the project.
- Team members (TM) were charged with the responsibility for engineering task development.
- Both program managers and project managers were able to negotiate and subcontract parts of its program or project to specialized companies. These companies consisted of one project manager (PMsc) and eight to ten team members (TMsc).

Teachers provided technical support in their feedback to different participants, although primarily to the program managers who assumed the role of owners for program outcomes. Thus, the instructor's team assumed the role of representatives for the aforementioned corporation. As an extension of their roles, teachers also worked as auditors to ensure that the basic technical rules for the experiment (e.g., procedures) were observed.

The methodology began with a proposal for five logical phases. These were similar to the five-phase project lifecycle that have been defined by the Project Management Institute (PMI) standard (PMI, 2008) for management of projects. They are Set Up, Planning, Execution, Control and Close. Theoretical classes emphasized the relevance of the methodology as a key aspect in managing

Figure 1. General structure of the experiment

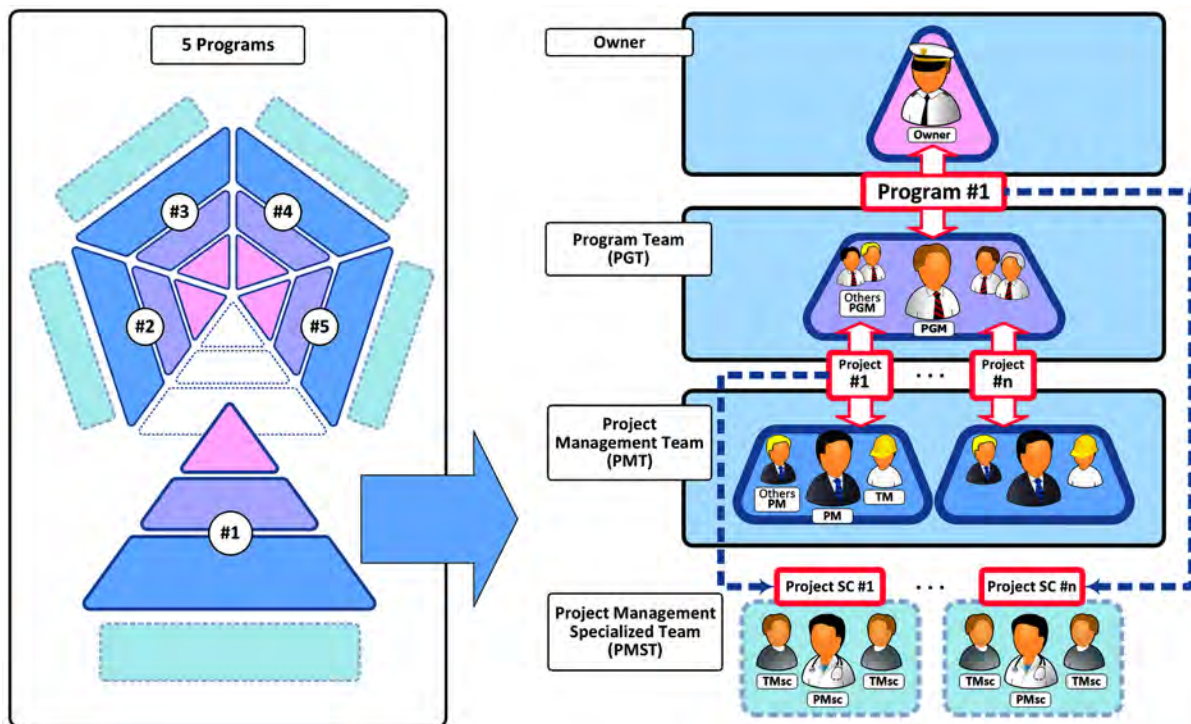
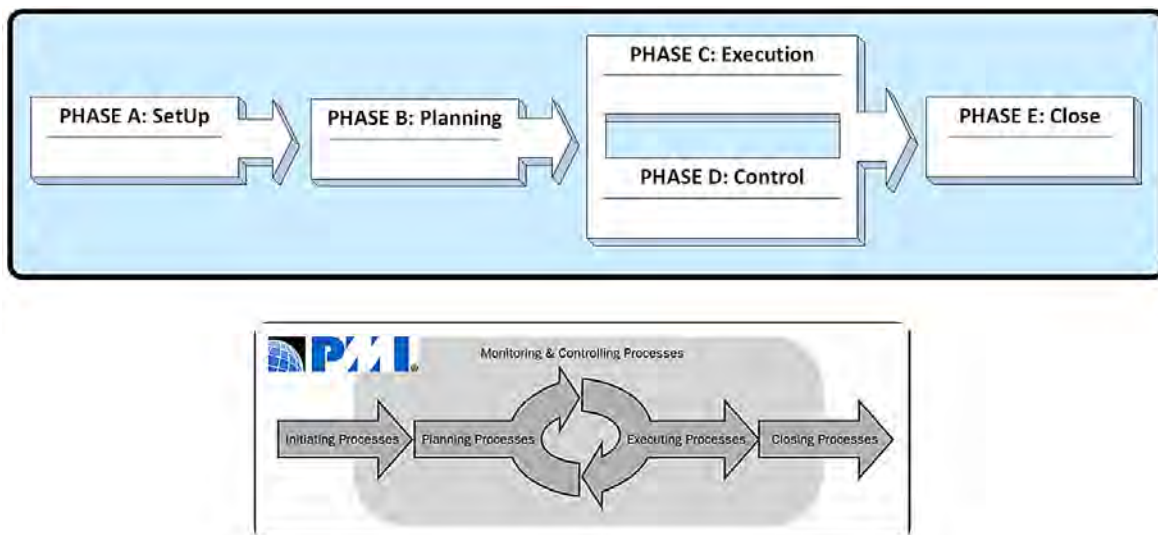


Figure 2. Designed program and project phases (top) based on the five-phase project lifecycle defined by PMI standard (bottom)
PMI, 2008.



properly the project, and also in reviewing those actions and their significance in practical life. Furthermore, written procedures were provided to participants allowing them to review due steps at any time.

Experimental Design

In designing the experimental, it was decided to operate five independent programs that addressed the same goal, but different criteria for the selection of team members in both gender and prior education. Projects were defined within those programs according to the business case and the limited number of resources that were assigned to each program. At the beginning of the experiment, program managers were informed that ten to fifteen projects that had an average fifteen members per project would be sufficient in initially deciding how to split the scope for the program.

The level of desired scope definition and control was determined by program managers. The amount of detail in the project definition was negotiated between project managers and program managers (their customers) as a matter of their responsibility. Then, the project managers selected their own management level of detail that they would provide to answer their customers' requirements and to be able to manage the projects. This behavior is more or less the same as what we can find in real situations. No additional managerial orientations or constraints were provided at the beginning of the experiment. Instead, it was decided to restrict feedback to mistakes that became apparent in the actions that were undertaken. Thus, it was possible to identify specific statistically significant evidence of structural behavior due to because of specific design factors. By structural behavior, the authors meant parameters related to project management. These include intensity of project control derived from evidence, level of detail in planning the work, individualization of work at task level or degree of knowledge shared among project team members, among others. Relationships between this structural behavior and

classical project monitoring parameters (time, effort and quality) was also studied.

The main design factors that were considered at the program level were:

- **Previous Work Experience:** Programs 1 and 2 were assigned program managers who had previous functional working experience, although none had experience in projects or project management.
- **Gender Factor:** The participants in programs 1 and 4 were women and the participants in program 2 were men. The other programs contained both men and women.
- **Cultural Factors:** Programs 3 and 4 included a mixture of members of various nationalities. The members of other groups belonged to the same country or had a similar cultural behavior. In this work, term culture is used as a term for nations, since our mental programming is at least partly a product of our national background.

Information System

The experiment used an information system that collects real-time information about low level actions made during the project on the Project Management Office software. Also, it enables the creation of periodic reports to better identify mistakes or inappropriate behaviors. Finally, more than three hundred different variables were collected during the project's life. This amount of information provided a global view of the program itself, as well as a view at the project level.

The authors have selected some open-source tools that are oriented to web project management applications, as well as other tools that provide database capabilities. They also have developed their own tools to produce automated auditing reports in order to have continuous monitoring by team member, etc. The system uses most of the information gathered for automatically providing feedback to students. More details of the designed web-based support system can be found

in González-Marcos, Alba- Elías, & Ordieres-Meré (2013).

In this way, teachers can monitor more objectively and efficiently, as well as evaluate practitioners continuously throughout the whole experiment. Eventually, at the end of the experiment, a final assessment, which is similar to a post-mortem analysis, is prepared that considers all of the activity reports that were recorded.

RESEARCH DEVELOPMENT

As previously indicated, this experiment was created to study potential sensitivity of project configuration factors, such as origin of project members (local/virtual), gender and national cultural considerations and structural behavior of project performance.

In this study, project effectiveness was measured as the average quality over the time required in producing its goals in comparison to the initial estimate for that project. Since projects are intended to produce some deliverables at an expected quality by using a defined set of resources in a period of time, the defined measure was considered as a coherent metric of effectiveness.

Policies and strategies decided at project management level can also influence project performance. Since the information system collects detailed information about the process itself, the following factors are analyzed: workpackage level at the Work Breakdown Structure (WBS), level of detail in planning the work (number of tasks), average number of persons assigned per task (socialization) and details of control exercised.

In summary, this work attempted to determine the intensity of those different factors against effectiveness in project management. Furthermore, it looked for sufficient statistical evidence on additional factors, such as gender, team virtualization and previous experience.

More specifically, we have considered the following factors which are closely related to the ones previously identified:

- **Project Effectiveness (P_EFF):** Is the ratio of the performance obtained to what was planned. Effective performance is defined as the sum of perceived quality – on a Likert scale – of all deliverables produced over the amount of effort spent to produce them.
- **Detailed Project Scope (DPS):** Has been defined as the number of workpackages in the project.
- **Detailed Work Plan (DW):** Has been defined as the number of planned tasks in the project. Both summary tasks and normal tasks are counted.
- **Collaborative Dimension of Work (CW):** Has been defined as the average number of members per defined project task.
- **Detail of Control Performed (DC):** Has been defined as the number of entries at project's blog, as it accounts for messages exchanged between the team asking for specific requests or actions related to their project.
- **Percentage of Virtualized Team (PVT):** Has been defined as the lower fraction of team members per University over the number of project team members.
- **Percentage of Multicultural Management (PMM):** Lower fraction of project management members that have a different nationality over the number of project management members.
- **Gender Management (GM):** Can vary between '1','2','3' as if the project management team can consist of only females, only males or a mixture of the sexes.
- **Owner's Cultural Influence (OCI):** Because the owner of a project is the program, it is possible that national culture may influence the projects in the program. It has been defined as the lower fraction of program managers with different nationality.
- **Owner's Gender Influence (OGI):** It is possible that gender approach to manage-

Table 1. Main factors observed after the programs ended

Program Name	Gender	Work Experience	Multicultural Team	Detailed Monitoring	Quality
1	1: F	Yes	0%	4	4
2	2: M	Yes	0%	5	3
3	3: X	No	50%	3	3
4	1: F	No	50%	2	2
5	3: X	No	0%	2	3

ment affects the development of projects in the program. OGI has been defined as the program's Gender Management.

- **Owner's Level of Control (OLC):** A rank between 1 (the lowest level of control) and 5 (the highest level of control), according to the strategy that has been adopted by the program managers.

Table 1 illustrates the main differences between programs. These include such aspects as members' gender, previous work experience, team's multiculturalism, detailed monitoring and final quality. The detailed monitoring values were obtained by a formal assessment of the WBS. A five point Likert-type scale that ranged from 1 to 5, in which 1 indicated very poor and 5 indicated very good, was used for the detailed monitoring and quality parameters.

In order to simplify the analysis, some ranges have been defined by parameter, like the Likert scale, to represent different situations as being very low/very poor, low/poor, high/good or very high/very good. This facilitates a better understanding of underlying dependences.

RESULTS AND DISCUSSION

After conducting the experiment, some basic information about its evolution is presented. First, the number of finally scheduled projects in each program was not the same. Table 2 illustrates how programs established different numbers of projects, according to differing ideas of scope control that was to be imposed on the projects. Since the total number of resources was almost constant, the effect was to produce lighter project teams. The number of project failures experienced during the development was also highlighted. They differ statistically by program, which suggests some intrinsic factors.

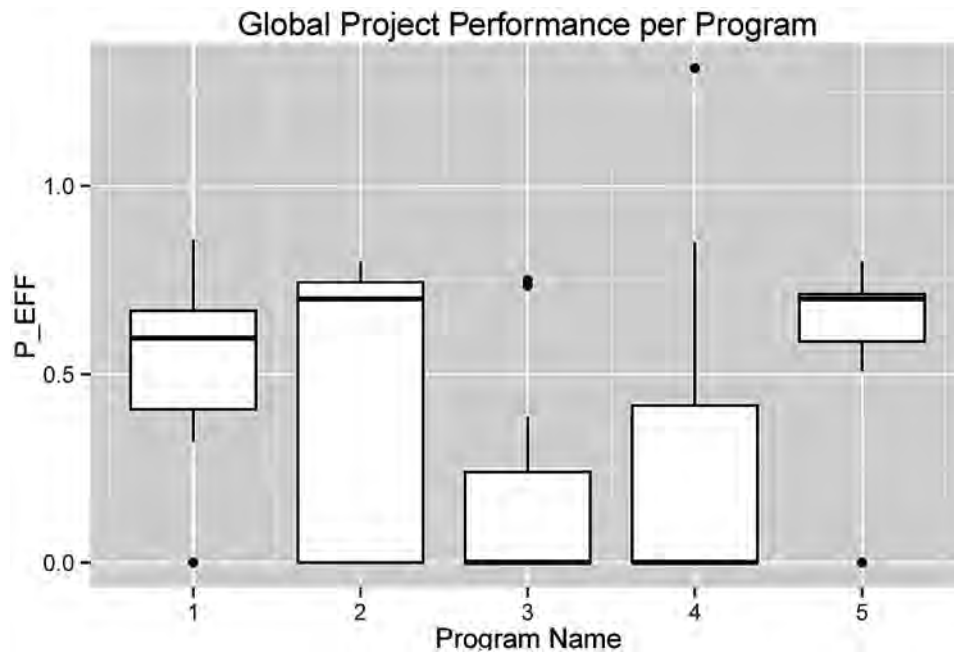
As a matter of general understanding of program relevance, the project effectiveness factor (P_EFF) by program was studied (Figure 3). To this end, it makes sense to discuss the outlier that was observed in program 4, where the effectiveness was higher than one. The reason for this is that the project team accomplished the work with good quality in much less times than initially estimated.

In order to assess the program's impact, formal equivalence tests have been conducted. In

Table 2. Factors identified at program level

Factor	Prg 1	Prg 2	Prg 3	Prg 4	Prg 5
Number of projects	14	18	11	14	11
Number of failed projects	1	5	8	9	1
Averaged number of PM	3.2	3.26	3.25	3.1	3.5
Averaged number of TM	11.5	9.16	11.18	9.6	11.27
Averaged virtualization	26%	27.2%	16.5%	26.6%	42%

Figure 3. General performance of projects by program



equivalence studies, if the evidence does not sufficiently favor equivalence, nonequivalence cannot be ruled out. Thus, one-way ANOVA shows that null hypothesis can be rejected. This means that there are differences between programs according to the F-test (see Table 3).

By using the well-known Tukey's "Honest Significant Difference" method, program pairs 1–3, 1–4, 2–3, 2–4, 2–5, 3–5 and 4–5 exhibited strong differences that enabled us to emphasize how relevant the management style of a program had become for the projects that belonged to it.

Another relevant aspect to discuss is which of the different configuration factors can explain project effectiveness and why this could happen. To examine this, the authors established a five point Likert scale of project effectiveness. In this case, 1 indicated very low performance and 5 indicated

very high performance. Then, a classification tree was constructed (see Figure 4).

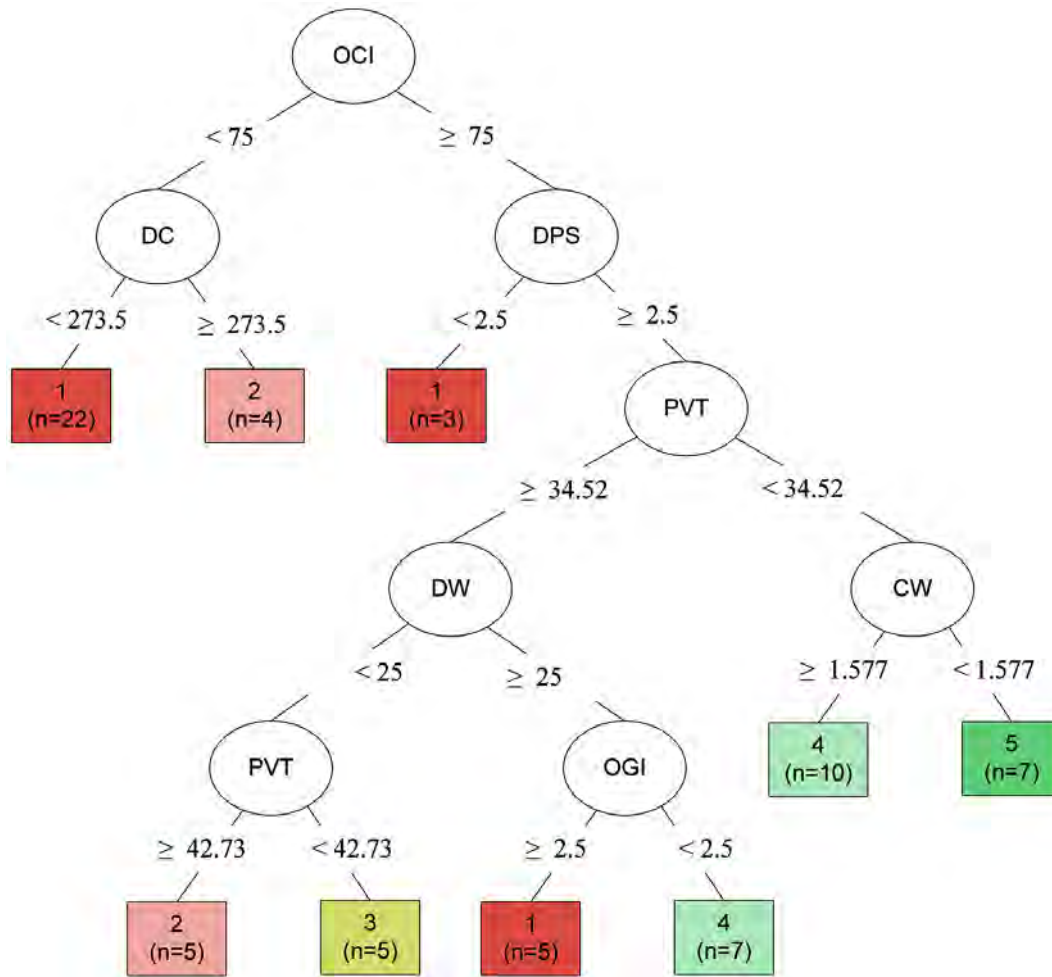
It is quite surprising how cultural factors (*OCI*) are becoming recognized as driving factors for program managers. It seems that program managers who belong to the same country or have similar cultural behavior provide better support. The underlying reason for this is not fully clear and additional research will be required. However, it appears to be related to how different program managers do their work according to their own standards and cultural patterns. These differences are perceived in the projects and deliver confusing messages about motivation and engagement.

In addition, it is clear that teams that do not pay enough attention to actively monitoring the project's progress ($DC < 274$ entries) evolve into

Table 3. Results of the ANOVA test of program relevance

	Df	Sum Sq	Mean Sq	F Value	Pr(>F)
Prg	4	1.64	0.41	4.14	0.0049
Residuals	63	6.26	0.10		

Figure 4. Relationship between project effectiveness and configuration factors



very low performing project teams. Similarly, a low level of scope definition ($DPS < 3$ workpackages) is recognized to be a driving force for a very low performing projects. The explanation is that a low scope detail leads to bigger project components for which defining the work to do and estimating the time required and cost is more difficult.

It is also worthwhile mentioning the impact of virtualized teams (PVT) on project effectiveness. The main reason is that teams' virtualization means more diverse skills (different behavior, ways of thinking, backgrounds, etc.) and, with careful management, teams become more productive.

For less virtualized teams (PVT) and mixed program management teams (OGI), it is important to mention how giving additional effort to managing the project gets recognition ($DW > 25$ tasks imply high performance).

SOLUTIONS AND RECOMMENDATIONS

The experiments were performed using a convenient sample of projects with a repeatability factor of five in order to consider diversity in them. They were designed to be developed without any

previous work experience at project level and considering it as a factor for programs. It was possible to experience slight variations if previous experiments become accepted at project level and this need to be investigated in the future.

On the other hand, despite the automatic system to collect evidence used in this experiment, it is always recommended that there be specific post-mortem analyses. Thus, it will be possible to identify the main reasons for specific behaviors and to discuss the relevance of introducing additional sensors to identify them for future projects.

FUTURE RESEARCH DIRECTIONS

Since individual skills could impact the team's morale and, therefore, the project effectiveness, future work will focus on measuring the evolution of individual competences during the experiment. In such work, the International Project Management Association (IPMA) competence model (IPMA, 2006) will be used as the reference for competence consideration.

Once individual competences have been measured, the authors intend to assess their impact on project effectiveness by using data mining.

CONCLUSION

The authors have designed and implemented an integrated platform that enables the monitoring of project activities on an individual basis. This infrastructure was used to run an experiment that was focused on analyzing the impact of various factors on project effectiveness. Thus, several guidelines can be derived that can help practitioners and course designers to improve the outcomes.

First, the significance of establishing program managers' teams based on similar cultural views becomes clear, as they will become the sponsors of projects being developed in those programs.

Also, it would be worthwhile to promote any kind of tool that encourages the sharing of infor-

mation about common mistakes to avoid and good managerial practices. It was made clear during this experiment that practitioners effectively exchange such kind of applied knowledge. Thus, it becomes easily disseminated throughout the organization. This must be accomplished by the work of highly motivated auditors, who play an important role in the early identification of mistakes and can make them visible for subsequent correction.

It also can be important to get teams virtualized as this allows participants to enrich their learning experience. Though exposure to different ways of thinking and the need to adopt additional management attitudes, they could increase the projects' effectiveness.

It is also relevant to emphasize the improvement is obtained when effective collaborative teams are organized. This is a project management strategy that must be emphasized to participants, not only from a theoretical point of view, but also on the basis of empirical evidence of the improvements achieved.

ACKNOWLEDGMENT

The authors wish to recognize the support of the Spanish "Ministerio de Educación" through its research funding program, as part of this work was supported by grants "EA-2010-0001" and "EDU2012-31080". The authors also appreciate the support of the "Vicerrectorado de Profesorado, Planificación e Innovación Docente" of the University of La Rioja through the "Dirección Académica de Formación e Innovación Docente" with its grant, APIDUR 2015.

REFERENCES

Alba-Elías, F., González-Marcos, A., & Ordieres-Meré, J. (2014). An active project management framework for professional skills development. *International Journal of Engineering Education*, 30(5), 1242–1253.

- Alias, Z., Zawawi, E. M. A., Yusof, K., & Aris, N. M. (2014). Determining Critical Success Factors of Project Management Practice: A Conceptual Framework. *Procedia: Social and Behavioral Sciences*, 153(16), 61–69. doi:10.1016/j.sbspro.2014.10.041
- Barclay, C. (2008). Towards an integrated measurement of is project performance: The project performance scorecard. *Information Systems Frontiers*, 10(3), 331–345. doi:10.1007/s10796-008-9083-6
- Byrd, T. A., & Turner, D. E. (2001). An exploratory analysis of the value of the skills of it personnel: Their relationship to its infrastructure and competitive advantage. *Decision Sciences*, 32(1), 21–54. doi:10.1111/j.1540-5915.2001.tb00952.x
- Din, S., Abd-Hamid, Z., & Bryde, D. J. (2011). ISO9000 certification and construction project performance: The Malaysian experience. *International Journal of Project Management*, 29(8), 1044–1056. doi:10.1016/j.ijproman.2010.11.001
- González-Marcos, A., Alba-Elías, F., & Ordieres-Meré, J. (2013). A web-based support system for providing effective monitoring, feedback and evaluation in project management education. In *Proceedings of the 5th International Conference on Computer Supported Education (CSEDU 2013)*. Aachen, Germany: SCITEPRESS – Science and Technology Publications.
- González-Marcos, A., Alba-Elías, F., & Ordieres-Meré, J. (2016). An analytical method for measuring competence in project management. *British Journal of Educational Technology*, 47(6), 1324–1339. doi:10.1111/bjet.12364
- Hewagamage, C., & Hewagamage, K. (2011). Redesigned framework and approach for IT project management. *International Journal of Software Engineering and its Applications*, 5(3), 89-106.
- IPMA. (2006). *ICB: IPMA Competence Baseline. Version 3.0*. Nijkerk, The Netherlands: International Project Management Association.
- Marcelino-Sádaba, S., González-Jaen, L. F., & Pérez-Ezcurdia, A. (2015). Using Project management as a way to sustainability. From a comprehensive review to a framework definition. *Journal of Cleaner Production*, 99, 1–16. doi:10.1016/j.jclepro.2015.03.020
- Mir, F. A., & Pinnington, A. H. (2014). Exploring the value of project management: Linking project management performance and project success. *International Journal of Project Management*, 32(2), 202–217. doi:10.1016/j.ijproman.2013.05.012
- Müller, R., & Turner, J. R. (2005). Cultural differences in project owner – Manager Communication. In D. Slevin, D. Cleland, & J. Pinto (Eds.), *Innovations: Project management research 2004* (pp. 404-418), Newtown Square, PA: Project Management Institute.
- Müller, R., & Turner, J. R. (2007). The influence of project managers on project success criteria and project success by type of project. *European Management Journal*, 25(4), 298–309. doi:10.1016/j.emj.2007.06.003
- Ojiako, U., Ashleigh, M., Chipulu, M., & McGuire, S. (2011). Learning and teaching challenges in project management. *International Journal of Project Management*, 29(3), 268–278. doi:10.1016/j.ijproman.2010.03.008
- Økland, A. (2015). Gap analysis for incorporating sustainability in project management. *Procedia Computer Science*, 64, 103–109. doi:10.1016/j.procs.2015.08.469
- OPSR. (2003). *Improving Programme and Project Delivery*. Office of Public Services Reform.

PMI. (2008). *A Guide to the Project Management Body of Knowledge (PMBOK Guide)* (4th ed.). Newtown Square, PA: Project Management Institute.

Qureshi, T. M., Warraich, A. S., & Hijazi, S. T. (2009). Significance of project management performance assessment (PMPA) model. *International Journal of Project Management*, 27(4), 378–388. doi:10.1016/j.ijproman.2008.05.001

Rehman, A. U., Usmani, Y. S., & Al-Ahmari, A. (2012). A study to assess significance of project management performance: Assessment model in applied projects. *International Journal of Applied Systemic Studies*, 4(3), 140–149. doi:10.1504/IJASS.2012.051129

Romero, C., & Ventura, S. (2010). Educational Data Mining: A Review of the State of the Art. *Transactions on Systems, Man, and Cybernetics - Part C: Applications and Reviews*, 40(6), 601–618. doi:10.1109/TSMCC.2010.2053532

Shahiri, A. M., Husain, W., & Rashid, N. A. (2015). A Review on Predicting Students Performance Using Data Mining Techniques. *Procedia Computer Science*, 72, 414–422. doi:10.1016/j.procs.2015.12.157

Tesch, D., Sobol, M. G., Klein, G., & Jiang, J. J. (2009). User and developer common knowledge: Effect on the success of information system development projects. *International Journal of Project Management*, 27(7), 657–664. doi:10.1016/j.ijproman.2009.01.002

ur Rehman Toor, S., & Ogunlana, S.O. (2010). Beyond the “iron triangle”: Stakeholder perception of key performance indicators (KPIs) for large-scale public sector development projects. *International Journal of Project Management*, 28(3), 228–236.

Wi, H., & Jung, M. (2010). Modeling and analysis of project performance factors in an extended project-oriented virtual organization (EProVO). *Expert Systems with Applications*, 37(2), 1143–1151. doi:10.1016/j.eswa.2009.06.051

Winter, M., Smith, C., Morris, P., & Cicmil, S. (2006). Directions for future research in project management: The main findings of a UK government-funded research network. *International Journal of Project Management*, 24(8), 638–649. doi:10.1016/j.ijproman.2006.08.009

ADDITIONAL READING

Anderson, M., & Gavan, C. (2016). *Developing Effective Educational Experiences through Learning Analytics*. Hershey, PA: IGI Global. doi:10.4018/978-1-4666-9983-0

Nooriafshar, M. (2013). The effectiveness of a lifecycle approach to learning project management. *The European Journal of Social & Behavioural Sciences*, 4(1), 836–847. doi:10.15405/FutureAcademy/ejsbs(2301-2218).2012.4.15

Peña-Ayala, A. (Ed.). (2014). *Educational Data Mining. Applications and Trends*. Switzerland: Springer. doi:10.1007/978-3-319-02738-8

Research, B. M. G. (2014). *Factors in project success*. Birmingham, UK: The Association for Project Management.

Silvius, A., & Schipper, R. (2014). Sustainability in Project Management Competencies: Analyzing the Competence Gap of Project Managers. *Journal of Human Resource and Sustainability Studies*, 2(02), 40–58. doi:10.4236/jhrss.2014.22005

KEY TERMS AND DEFINITIONS

Culture: The collective programming of the mind that distinguishes the members of one group or category of people from others.

Educational Data Mining: Techniques, tools, and research designed for automatically extracting meaning from large repositories of data generated by or related to people’s learning activities in educational settings.

Program: A group of related projects and other activities aligned with an organizational strategy or major goals, which are managed in a coordinated way to obtain benefits not available from managing them individually.

Project: A temporary endeavor consisting of coordinated and controlled activities with start and finish dates, undertaken to achieve an objective, i.e., to create a unique product, service, or result.

Scope: The sum of the products, services, and results with the specified features and functions to be provided as a project.

Virtual Team: A set of individuals who work across time, space and organizational boundaries and support the project manager in performing the work of the project to achieve its objectives.

Work Breakdown Structure (WBS): A hierarchical decomposition of the total scope of work to be carried out by the project team to accomplish the project objectives and create the required deliverables.

Workpackage: Unit of work or job defined at the lowest level of the work breakdown structure for which cost and duration can be estimated and managed.

Database Techniques for New Hardware

D**Xiongpai Qin***Renmin University of China, China***Yueguo Chen***Renmin University of China, China*

INTRODUCTION

In the last decade, computer hardware progresses with leaps and bounds. The advancements of hardware include: widely application of multi-core CPUs, using of GPUs in data intensive tasks, bigger and bigger main memory capacity, maturity and production use of non-volatile memory etc.

Database systems immediately benefit from faster CPU/GPU and bigger memory, and run faster. However, there are some pitfalls. For example, database systems running on multi-core processors may suffer from cache conflicts when the number of concurrently executing DB processes increases. To fully exploit advantages of new hardware to improve the performance of database systems, database software should be more or less revised (Ailamaki, 2004).

This chapter introduces some efforts of database research community in this aspect. And the following section gives a brief introduction to some new hardware technologies that database systems could utilized.

BACKGROUND

Multi-Core CPU

Improving the performance of CPU through increasing its clock frequency becomes more and more difficult. Researchers and engineers bring forth multi-core technology. In a typical multi-core CPU, 2, 4, 8 or more cores are integrated on one chip. Putting several cores on one die allows for

higher communication speeds between the cores, which will benefit many computing tasks.

The cores have their own private caches (Level 1 Cache, or L1 Cache), and share some larger but slower caches (L2 Cache). They access the shared main memory for parallel data processing. Database systems are basically multi-threaded, and they can benefit from multi-core CPUs without any modification. However, to fully utilize the cores to boost database performance, there is much work to do.

GPU for General Tasks

GPU is traditionally used to accelerate the specific task of graphic rendering. GPU vendors have integrated many computing units in a single die, and optimized the bandwidth to process large volume of graphic data. The highly parallelism of GPUs is exploited to speed up data intensive tasks as well, and GPU becomes GPGPU (General Purpose GPU).

Since GPU is designed primarily for graphic processing tasks instead of general tasks, the architecture of a GPU is rather different from CPU. It has its own unique thread hierarchy and memory hierarchy, which should be taken into account when using GPU for data processing tasks.

Bigger Memory

The price of memory is going down, now people can install more memory in a single server. It is rather common for a single server to possess a memory capacity as large as hundreds of giga-

bytes, or even up to terabytes. For moderate-size applications, it is possible to load the whole dataset into memory for fast access.

Non-Uniform Memory Access

Non-Uniform Memory Access (NUMA) machine is becoming more and more common. The NUMA architecture consists of a small number of processors, each having its own memory and I/O channels. Each group of CPUs is called a ‘node’. Memory that is local to a node is called local or near memory, while memory outside of a node is called foreign or far memory (Golding, 2010). Accessing foreign memory is much slower than accessing local memory. NUMA architecture requires changes of memory management.

Non-Volatile Memory, Solid State Disks, and Changes of the Memory Hierarchy

Non-volatile memory (NVRAM) is a type of computer memory that can keep the information even after the power is turned off. There are several flavors of NVRAM, including flash memory, and phase change memory (PCM) etc.

Flash memory devices use two different logical technologies, i.e. NOR and NAND, to map data. NOR flash reads and writes data in specific memory locations. It provides high-speed random access and it can retrieve a single byte. NAND flash reads and writes sequentially at high speed in small blocks called pages. In general, reading data from flash memory is much faster than writing data into it.

Phase-change memory (PCM) is a form of computer RAM (Random Access Memory). It stores data by altering the state of the matter from which the device is made. PCM is byte addressable. It incurs very low power consumption. And it has the potential to provide inexpensive, high speed, and high volume non-volatile storage.

A solid state disk (SSD) uses integrated circuit as memory to store data permanently. SSDs have

no moving parts, they are more resistant to physical shocks, run silently, have lower access time and latency, and consume much less power. SSD is usually made of flash memory. With the price of flash memory continues to decline, the price of SSDs is also declining over time.

With advent of non-volatile memory and SSDs, traditional memory hierarchy has been enriched with new intermediate levels. Recent and hot data could be kept in high speed memory levels, which is near to CPU, and historical and cold data could be migrated to low speed memory or disks.

MAIN FOCUS OF THE ARTICLE

Database researchers have devised many new techniques to utilize characteristics of above introduced hardware to improve performance of database systems. We classify these techniques into several categories, i.e. storage, index, query processing, transaction processing & concurrency control, and recovery.

Storage and Compression and Migration

Data Layouts

Traditional database systems are designed around disks. Data is primarily stored in disks, and it is loaded into memory buffer for later process. Now the memory capacity is so big, and it could be used as primary store of data. The whole dataset or most data can be kept in memory for faster access. Outdated or unused data is occasionally evicted to disks to make room for data that will be accessed soon (Zhang, Chen, Ooi, Wong, Wu, & Xia, 2015). The design principle is called “anti-caching”. In these scenarios, the optimization focus has shifted to how to efficiently exchange data between main memory and CPU cache for faster processing.

For a relational table, data could be stored in three styles of layouts: row-wise, columnar, and

PAX. It is straightforward to store rows of a table row by row in a data page. The layout is called N-ary Storage Model (NSM) (Gray, & Reuter, 1992). Traditional database systems for OLTP (online transaction processing) applications usually adopt a row-wise storage layout, since transactions on these systems access only a small number of rows. However, NSM model exhibits low cache utilization on multi-core platforms (Ailamaki, DeWitt, Hill, & Skounakis, 2001).

Columnar data layout, on the other hand, partitions a table into columns, and each column is stored contiguously. Columnar databases have shown to perform more than an order of magnitude better than traditional row-wise database systems on analytical workloads, in which most queries scan large volume of data, and touch only several columns (Abadi, Madden, & Hachem, 2008). Various compression techniques can be applied to reduce the space consumption. Columnar stores are actually successor of decomposed storage model (DSM) (Abadi, Boncz, & Harizopoulos, 2009).

Ailamaki et al. (Ailamaki, DeWitt, Hill, & Skounakis, 2001) have proposed a new data organization model called PAX (Partition Attributes Across). In PAX, a table is firstly horizontally partitioned into pages. In each page, all values of each attribute are grouped together. The PAX data layout greatly improves cache performance. According to their experimental results, PAX save at least 75% of stall time due to data cache accesses. Range select queries and updates on memory resident table execute 17-25% faster, and TPC-H queries execute 11-48% faster compared to NSM storage layout. PAX is highly scalable since the data is horizontally partitioned.

Row formats are ideal for OLTP workloads, which typically access only a small number of rows. Column formats are suited for analytics queries, which typically examine a small number of columns from a large number of rows. While PAX is suited for hybrid applications with OLTP and OLAP workloads.

Some researchers tried to combine read-optimized and write-optimized storage models

together to support hybrid workloads (OLTP and OLAP). In the work of (Kruger, Grund, Tinnefeld, Plattner, Zeier, & Faerber, 2010), a write-optimized buffer is used to maintain the delta data (recently inserted/updated data), in conjunction with the read-optimized main store, to support hybrid workloads.

Kissinger et.al. proposed an NUMA-aware in-memory storage engine named ERIS (Kissinger, Kiefer, Schlegel, Habich, Molka, & Lehner, 2014). ERIS uses an adaptive partitioning method to exploit the topology of the underlying NUMA platform. It reduces NUMA-related issues significantly. Experiments conducted on a system of 64 multiprocessors, 512 cores, and 8 TBs main memory and another smaller server demonstrated its superior performance over a NUMA-unaware index for different workloads.

Data Compression

Various compression techniques such as dictionary based encoding, bit vector encoding, run length encoding etc. (Krueger, Wust, Linkhorst, & Plattner, 2012) could be applied to memory resident data to reduce space consumption.

Data Migration in Memory Hierarchy

In many applications, data accesses have skewed patterns, where some records are hot (frequently accessed) but many others are cold (infrequently or never accessed). It will be economical to keep hot data in main memory and migrate cold data to low speed memories such as SSD.

In 2015, Sun proposed the *Sluice Gate Theory* (Sun, 2015) for data transferring among different levels of the memory hierarchy. In Sun's opinion, data transfer in a memory hierarchy is staged, and different stages have different capacities. Data is bumped and delayed at stage change (gate), and not all data goes to the next step. Data transferring in a memory hierarchy is something more like water transferring in sluice other than flowing in a river. The model has the potential to greatly improve

performance of data intensive computing when hardware and software are properly co-designed under the principle.

Eldawy et.al. (Eldawy, Levandoski, Larson, 2014) proposed the Siberia architecture, which migrates cold data to secondary storage and provides an interface to users to read and modify both hot and cold data, while minimizing number of accesses to cold storage. Records can be migrated between hot and cold stores while DBMS is still online. In main-memory optimized databases, appropriate cold data access rates incur an acceptable 7-14% throughput loss.

Index

When retrieving data from database, we can avoid full table scans when there are some indexes to located relevant data. New indexing techniques can be roughly classified into three categories, cache conscious indexes, cache oblivious indexes, and other index schemes such as those designed for non-volatile memory.

Cache conscious indexes are designed to reduce cache misses, based on parameters of memory hierarchy such as the number of memory levels, the block size of each level, and the relative R/W speed of each level. In contrast, cache oblivious indexes are designed to be independent of the parameters of memory hierarchy to achieve acceptable performance no matter what hardware platforms the database system is running on.

Cache Conscious Indexes

Rao and Ross proposed a cache conscious B+-tree called “Cache Sensitive B+-trees” (Rao, & Ross, 2000). The goal is to make B+-trees as cache conscious as CSS-trees (Cache Sensitive Search trees) without increasing the update cost too much. The tree stores all the child nodes of any given node contiguously, and keeps only the address of the first child in each node. The rest of the children can be located by adding an offset to that address. Since only one child pointer is stored

explicitly, the utilization of a cache line is higher. CSB+-trees support incremental updates in some way similar to B+-trees.

Dong and Xu presented a cache conscious R-tree variant (Dong, & Xu, 2007). Firstly, Quantized Bounding Spheres (QBSs) are introduced to approximate data points or bounding spheres. Then the advantages of QBSs and the Quantized Bounding Rectangles (QBRs) are incorporated into the cache conscious R tree (CSR+-tree). A new distance computation scheme is proposed to avoid the decompressing of QBSs or QBRs, which leads to significant run time reduction. The experiment results show that CSR+-tree consistently outperforms other representative indexes.

Faust et. al. (Faust, Schwalb, & Kruger, 2013) presented Paged Index, a sparse index tailored towards dictionary-encoded columns. By indexing the occurrence of values on a block level, the search scope for scan operations can be reduced drastically. They achieved up to two orders of magnitude of performance improvement for the column scan operations, paying only a little overhead for index maintenance and storage.

Binna et. al. (Binna, Pacher, Meindl, & Specht, 2014) proposed a cache conscious index structure DCB-tree for storing of short fixed size keys. They combined a hierarchical cache aligned node layout with delta encoding and pointer compression. Compared to other main-memory index structures, DCB-tree reduces the amount of memory required by 80% in the best case and by 30% in the worst case, while providing equal or better performance for large real world datasets.

Cache Oblivious Indexes

Bender et. al. proposed a cache oblivious B-tree (Bender, Demaine, & Farach-Colton, 2000). The first data structure permits optimal scans and searches from any node in the tree. The insertion and deletion costs for this data structure match the bound of $B = O((\log N)^2 (\log \log N)^4)$. The second data structure uses indirection and performs scans optimally. For this structure, insertions and dele-

tions match the bound of $B = O(\log N \log \log N)$. The third data structure also uses indirection, and insertions and deletions always match the B-tree bound.

Arge et. al. proposed a cache oblivious R-tree (Arge, Berg, & Haverkort, 2009). They created a cache oblivious data structure to store a set S , of N axis-aligned rectangles in the plane. The structure allows all rectangles in S , which intersect a query rectangle or a point, to be found efficiently. For any constant ϵ , a rectangle query can be answered using $O(\sqrt{T/B})$ memory transfers, and a point query can be answered using $O((N/B)^\epsilon)$ memory transfers, where T denotes the number of reported rectangles, and B denotes the block size of memory being transferred between any two levels of the memory hierarchy.

Lindstrom and Rajan (2014) proposed a general framework for generating cache oblivious layouts for binary search trees through recursively partitioning a tree into contiguous subtrees and prescribing an ordering amongst the subtrees. They also derived a new locality measure, i.e. the *Weighted Edge Product* to approximate the probability of cache misses at multiple levels. The *MIN WEP* layout for a complete binary search tree optimized by the *Weighted Edge Product* outperforms previous cache-oblivious layouts by around 20%.

Other Index Techniques

To speed up ad-hoc queries, Petraki et. al. (Petraki, Idreos, & Manegold, 2015) proposed a holistic indexing scheme. It makes use of available CPU resources and continuously adapts the index to workload changes, without human intervention. Runtime adaption does not affect the running queries negatively. In the meanwhile, the refinement of the physical design based on the changing workloads brings substantial performance improvements.

PCM's long write latency and high write energy bring challenges for data management over it. Li et. al. (Li, Jin, Yang, Wu, & Yue, 2016) propose

an improvement over the B+-tree for PCM. For write-intensive leaf nodes, they use an overflow-node technique to reduce PCM writes. For read-intensive ones, they adjust the tree structure to remove overflow nodes to improve the read performance. The proposed B+-tree outperforms the traditional B+-tree and the overflow B+-tree.

Query Processing

Query Processing with Multi-Core CPUs

Optimization of Database Operators

Kim et al. investigated join implementations on multi-core CPUs (Kim, Sedlar, Chhugani, Kaldewey, Nguyen, Blas, Lee, Satish, & Dubey, 2009), including hash join and sort-merge join. They offered optimized implementations, which have exploited large on-die cache, thread-level parallelism, data-level parallelism, and high memory bandwidth. The hash join performs consistently over a wide range of input data sizes from 64K to 128M tuples, and it is not affected by data skewness. The throughput of more than 100M tuples per second is 17X faster than the best reported performance on CPUs, and 8X faster than that reported for GPUs. On the other hand, the optimized sort-based implementation achieves a throughput of 47M to 80M tuples per second.

Albutiu et al. proposed a suite of new massively parallel sort-merge join (MPSM) algorithms (Albutiu, Kemper, & Neumann, 2012), which are based on partial partition-based sorting. The algorithms work on the independently created runs in parallel. This way the MPSM algorithms are NUMA-affine, since all the sorting is carried out on local memory partitions. Experiments on a 32-core machine with one terabytes of main memory prove the competitive performance of MPSM on large main memory databases with billions of tuples. The algorithms outperform the Vectorwise query engine by a factor of four.

Lang et. al. (Lang, Leis, Albutiu, Neumann, Kemper, 2013) identified the major performance

problems on NUMA architectures, and they developed a NUMA aware hash join algorithm. The algorithm distributes data carefully in order to provide balanced bandwidth on the inter-partition links without architecture-specific knowledge. Experimental results show that the hash join implementation outperforms previous hash joins by a factor of more than two, resulting in an unprecedented throughput of 3/4 billion join argument tuples per second.

Broneske et. al. (Broneske, Breß, &Saake, 2014) studied the performance of different variants of the scan operator. They investigated the impacts of four code optimizations (No Branching, Loop Unrolling Vectorization, and Parallelization) and their combination on the scan operator. They found that the performance of the scan operator for different processors gets harder to predict when several code optimizations are combined, which necessitate further investigation.

Zeuch and Freytag (Zeuch, & Freytag, 2015) extensively studied performance characteristic of selection operations on modern Intel CPUs. They concluded that a cost model that takes branch mispredictions as well as cache accesses with different weights into account can give more accurate cost estimations. Such a model could be generalized to other database operations such as projections, joins, or aggregations.

It is important to decide the degree of parallelism (DOP) of an operator in database query execution plans. Gawade et. al. (Gawade, & Kersten, 2016) proposed a feedback based parallelization scheme, which identifies the most expensive operator from previous execution. Their method generates better plans and improves the utilization of multi-cores.

Execution of Queries

Xi et. al. (Xi, Mishima, Yokota, 2014) proposed CARIC-DA, a middleware for achieving higher performance in DBMSs on multi-core processors by reducing cache misses with a new cache-conscious dispatcher for concurrent queries. CARIC-DA logically range-partitions the data

set into multiple subsets. This enables different cores to access different subsets by ensuring that different DB processes are pinned to different cores, and by dispatching queries to DB processes according to the data partitioning information. In this way, CARIC-DA is expected to achieve better performance via a higher cache hit rate for each core's private cache.

Interactive applications, and periodic jobs to generate daily reports have different response time requirements. Resource management based on priorities has been proposed for disk resident databases, which typically rely on multiplexing threads on a number of processing units. However, the strategy is unfavorable for main memory database using multi-core CPUs. A large number of context switches disrupt cache conscious algorithms and have a negative impact on intra-query parallelism. Wust et. al. (Wust, Grund, Plattner, 2013) proposed a resource management framework based on a task-based query execution that avoids the cost of thread multiplexing.

Furthermore, they (Wust, Grund, Hoewelmeyer, Schwalb, & Plattner, 2014) proposed a task-based query execution model with priority-based scheduling to effectively prioritize of transactional and analytical queries in a single in-memory database instance on multi-core CPUs. They found that a larger task size for long running operators negatively affects the response time of short running queries. They therefore limit the maximum task size to control the mutual performance impact of different query classes.

Query Processing with GPUs

Govindaraju and his colleagues presented a new algorithm - GPU TeraSort (Govindaraju, Gray, Kumar, & Manocha, 2006) to sort billion-record databases using GPUs. The algorithm uses the data parallelism and task parallelism of the GPU to perform memory-intensive and compute-intensive tasks, and uses the CPU to perform I/O and resource management. On their hardware, they achieved a record of sorting 60GB for a penny.

Fang et al. explored the design variants in query co-processing using GPUs (Fang, He, Lu, Yang, Govindaraju, Luo, & Sander, 2007). They have devised a GPU Sort algorithm which is based on bitonic sorting network, a GPU-CSS tree which is based on *Cache Sensitive Search Tree*, and a GPU-based *Min Max Join*. They presented promising results of their methods. Cederman and Tsigas proposed a modified quick sort algorithm for GPU (Cederman, & Tsigas, 2009). The new quick sort algorithm outperforms radix sort and bitonic sort. GPU based quick sort is a viable method for sorting large quantities of data.

He et al. designed a set of data-parallel primitives such as scan, scatter and split, and used these primitives to implement various joins (He, Yang, Fang, Lu, Govindaraju, Luo, & Sander, 2008) including indexed/non-indexed nested-loop, sort-merge and hash joins. The GPU-based algorithms are 2-20X faster than their CPU-based counterparts.

Kaldewey et al. revisited offloading large-scale relational join operations to GPUs (Kaldewey, Lohman, Muller, & Volk, 2012). GPU-based join algorithms achieved 3-8X performance improvement over the fastest reported implementation on multi-core CPUs. In the typical scenario of hash join, i.e., when the probe table is much larger than the build table, the GPU hash join implementation could achieve 96% utilization of the bandwidth (6.3GB/s) of PCI-E generation 2. They are also trying to handle larger data sets and process more complex queries on GPUs.

Karnagel, Muller and Lohman (Karnagel, Muller, & Lohman, 2015) study offloading grouping and aggregation operators, which are dominant operations in analytics queries after joins, to GPUs. Based on a detailed performance analysis of grouping and aggregation on the GPU, they provide a parametric model for a wide range of workloads with a heuristic optimizer to automatically pick the best implementation and performance parameters at execution time.

Transaction Processing and Concurrency Control

Concurrency control is critical for correct execution of simultaneously running transactions in database systems. It is based on locking or multi versioning. According to (Harizopoulos, Abadi, Madden, & Stonebraker, 2008), locking and latching account for 30% of instruction counts in transaction processing. When multi-core CPUs are widely used, and main memory is in abundance, designing an efficient concurrency control scheme for database systems is an important work to fully release the power of multi-core CPUs. The existing works are classified into two categories, lock-based methods and multi versioning based methods.

Lock-Based Concurrency Control

Jones et al. compared two low overhead concurrency control schemes (Jones, Abadi, & Madden, 2010). The first is a light-weight locking scheme, and the second is an even lighter-weight type of speculative concurrency control. The second scheme avoids the overhead of tracking reads and writes, but sometimes conducts some work that must be roll backed eventually. Speculative concurrency control can improve throughput by up to a factor of two using a modified TPC-C benchmark compared to other methods.

Ren et al. proposed *Very Lightweight Locking* (VLL) (Ren, Thomson, & Abadi, 2012), an alternative pessimistic concurrency control scheme for main-memory database system. VLL co-locates lock information (two simple semaphores) with the raw data, and forces all locks to be acquired at once by any transaction. A selective contention analysis (SCA) protocol helps systems implementing VLL to achieve high transactional throughput under high contention workloads. Experiment results on a database running on a multi-core single machine, and a distributed database running on a shared nothing cluster, show that VLL dramatically reduces locking overhead and increase transactional throughput.

Jung et. al. presented a scalable lock manager (Jung, Han, Fekete, Heiser, & Yeom 2013) for multi-core platforms. Based on the analysis on MySQL, it is identified that latch contention within the lock manager is the bottleneck that is responsible for the performance drops when the load increases. Adoption of barrier synchronization greatly reduces the usage of latch. Locks are pre-allocated in bulk, and the lock manager only needs to perform simple list manipulations during the acquiring and releasing phases of a transaction. De-allocation of the lock data structures is also done in bulk, which enables faster lock releasing. Their implementation can scale better than standard lock manager of MySQL.

Multi Versioning Based Concurrency Control

Larson et al. introduced two efficient concurrency control methods specifically designed for main-memory databases (Larson, Blanas, Diaconu, Freedman, Patel, & Zwilling, 2011). Both use multi versioning to isolate read-only transactions from updates, but differ in how atomicity is ensured, one is optimistic and the other is pessimistic. The authors also implemented a main-memory optimized version of single-version locking for comparison. Experimental results show that although single-version locking works well when transactions are short and contention is low, the performance quickly degrades under more demanding situations. Multi versioning schemes incur higher overhead, but are less sensitive to hotspots and long-running transactions.

Hyper database (Kemper, & Neumann, 2011) uses copy-on-update to preserve consistent snapshot to support OLAP queries in an OLTP main memory database. Hyper yields transaction rates as high as 100,000 per second and very fast OLAP query response times on a single system, executing hybrid workloads in parallel (using TPC-C and TPC-H benchmarks).

When the number of CPU cores increases, previous implementation of serializable snapshot

isolation does not scale well due to latch contention on internal data structures. Han et. al. proposed to detect the required read/write conflict conditions at runtime without the need to latch the transaction manager's internal structures (Han, Park, Jung, Fekete, Rohm, & Yeom, 2014). Their method scales well on an Intel 4-way 32 core machine. It outperforms their original implementation up to 20X, and maintaining high throughput when serving 500 clients.

Systems implementing snapshot isolation (SI) usually lack the support of full serializability. Neumann et. al. proposed a method to add serializability to existing snapshot isolation implementations (Neumann, Muhlbauer, & Kemper, 2015). The method is based on the adaptation of precision locking. It verifies that the writes of recently committed transactions do not intersect with the read predicate space of a committing transaction, by updating data in-place and storing versions as before-image deltas in undo buffers. Their MVCC achieved fast processing of point accessing and read-heavy transactions.

Kim et. al. proposed ERMIA (Kim, Wang, Johnson, & Pandis, 2016), a memory-optimized database system built from scratch to handle mixed workloads. They adopt snapshot isolation concurrency control instead of lightweight optimistic concurrency control (OCC) to coordinate heterogeneous transactions, and provides serializability when needed. ERMIA achieves comparable or superior performance and near linear scalability for a variety of workloads, compared to a lightweight OCC-based system.

Recovery

When a power-off takes place, the contents in the main memory database (except NVRAM) will disappear. Recovery system is a critical component for main memory databases. The goal of the recovery system is to recover the whole database as soon as possible after failures, without blocking of normal transaction execution too much when doing bookkeeping.

Recovery system of disk resident databases is mature and efficient. Usually it includes three components i.e. logging, checkpointing, and recovery. Researchers don't build recovery systems for main memory databases from scratch, but they borrow ideas from disk resident databases. Levy and Silberschatz presented a page-based incremental restart algorithm (Levy, & Silberschatz, 1992), which enables the resumption of transaction processing as soon as possible after the system restarts. Data pages are recovered individually and according to the access demands of the post-crash transactions. Lin and Dunham proposed a segmented fuzzy checkpoint (SFC) scheme (Lin, & Dunham, 1996), which uses a moving window to locate the redo-point more efficiently than traditional fuzzy checkpointing approaches. SFC can significantly reduce the log processing time during restart, only slightly degrading the system performance.

Lee, Kim, and Cha introduced a parallel recovery method (Lee, Kim, & Cha, 2001), which is based on a differential logging technique. Differential logging has two nice properties of commutative and associative. Logging records can be combined in any order and written to disks in parallel, which speeds up the execution of transactions. During recovery, the applying of log records to the data pages does not need to comply with the serialized order, thus different pages can be recovered in parallel, which accelerates the recovery process.

In modern high-throughput transaction processing systems, as transaction throughputs get higher, fine-grained, record-oriented ARIES-style physiological logging starts to take a substantial fraction of the overall transaction execution time, it is no longer the optimal way to recover a database system. Malviya et. al. (Malviya, Weisberg, Madden, Stonebraker, 2014) proposed a light weight command logging technique which only records the commands that were executed. The database is recovered by loading of a transactional consistent checkpoint and replaying the commands in the log. Command logging yield significantly

higher throughput at run-time. It offers 1.5X higher throughput than a main memory optimized implementation of ARIES-style physiological logging with TPC-C benchmark.

High speed byte addressable non-volatile memory allows people to rethink the durability mechanism of main memory databases. Schwalb et. al. (Schwalb, Kumar, Dreseler, Anusha, Faust, Hohl, Berning, Makkar, Plattner, & Deshmukh, 2016) proposed a new database Hyrise-NV, which updates data and index transactionally on NVRAM using multi version data structures. Hyrise-NV can instantly recover the database independently of its size, allowing the recovery of a table with 10 million rows in less than 100 ms.

Database Industry Efforts

Database management systems such as SQL Server was designed and implemented in the days when main memory was small, data lived in disks, machines had a single slow processor, and OLTP was the only workload. Nowadays the situation is changing. People have large memories, plenty of cores, and transactional and analytic workloads are equally important. Observing the potential of providing higher query processing performance by adapting database implementation to new hardware, database vendors and startups try to revise their current flagship product under new hardware environments or invent new databases.

Microsoft adapts SQL server to today's hardware by adding new features of column store indexes and in-memory tables (Larson, Hanson, & Zwilling, 2015).

IBM provided in memory BLU Acceleration for DB2 to improve its performance on mixed OLTP and BI analytics workloads (Barber, Lohman, Raman, Sidle, Lightstone, & Schiefer, 2015). Data is encoded and packed densely into bit-aligned vectors and kept in system cache, which can exploit SIMD (single instruction multiple data) instructions in query processing. BLU assigns independent data structures to each thread, incurring low synchronization cost, and achieves

scalable multi-core parallelism. BLU improves the performance on complex analytic queries by 10 to 50 times, on UPDATE and DELETE up to 112 times, compared to row organized tables.

Oracle provides a dual format in memory database (Lahiri, Chavan, Colgan, Das, Ganesh, Gleeson, ..., & Zait, 2015). The database allows data to be maintained in row-wise and columnar formats simultaneously and maintain transactional consistency between them. It employs a variety of optimizations including various levels of compression, SIMD vector processing, and in-memory storage indexes. In-memory column format achieves per CPU core scan speeds exceeding many billions of rows per second, which enables further query optimizations such as in memory aggregation. Costly joins and aggregations can be converted into a series of filtered scans. There is no need to create indexes. OLTP DML operations achieve a substantial speedup by eliminating index maintenance cost.

HP Lab envisions that next generation of servers will have thousands of cores, large DRAM and huge NVRAM, now no existing database is designed to running on such hardware. They developed a new database system FOEDUS (Kimura, 2015). It scales up in-memory database technologies, and allows transactions to manipulate data stored in both DRAM and NVRAM efficiently. The throughput of FOEDUS on a large NUMA machine with 240 physical cores is multiple orders of magnitude higher than H-Store with anti-caching using the TPC-C benchmark.

Besides the above mentioned systems, several other database systems deserve attention. Vectorwise is the commercialization of open sourced analytic main memory database of Monet. VoltDB is the commercialization of open sourced main memory database of H-Store, and it is aimed at OLTP workloads. SAP's HANA is an in memory database for high performance processing of hybrid workloads (Faerber, Dees, Weidner, Bauerle, & Lehner, 2015). MemSQL (Shamgunov, 2014) is a distributed in memory database, also supporting real time processing of OLTP and OLAP workloads.

FUTURE RESEARCH DIRECTIONS

Hardware-conscious database design is still a blossoming research direction when new hardware keeps emerging. Innovations on low level of storage and index, to up level of query processing, which fully utilize characteristics of new hardware, will improve the performance of database. To achieve higher performance, some researchers even accelerate database operations using specific Field Programming Gate Array (FPGA) (Mueller, & Teubner, 2009).

Another research direction is keeping database system in an independent progress track. Some researchers design general data models and query processing algorithms that are adaptive to various hardware, with the expectation that database can achieve consistently high performance on different hardware.

CONCLUSION

There are several trends in hardware development in recent years. Multi-core CPUs provide massive parallelism. GPUs provide even more parallelism and they are also used in data intensive tasks. With the price going down, people now install very large memory in a single machine, allowing them to keep more data in memory. Various non-volatile memory techniques become mature. High performance non-volatile memory and solid state disks introduce more intermediate levels in the memory hierarchy.

All these hardware advancements require that database systems are revised to fully leverage the power of hardware for higher performance. Some efforts of database research community are presented in this chapter. The techniques include data layouts, index schemes, query processing techniques, transaction processing & concurrency control methods, and recovery techniques, which accelerate data exchange between different levels of memory hierarchy and leverage parallelism of multi-core CPUs and GPUs.

ACKNOWLEDGMENT

This work is funded by NSF of China under Grants No. 61170013.

REFERENCES

- Abadi, D. J., Boncz, P. A., & Harizopoulos, S. (2009). *Column-oriented database systems*. VLDB 2009 Tutorial. Retrieved May 13, 2016, from <http://www.vldb.org/pvldb/2/vldb09-tutorial6.pdf>
- Abadi, D. J., Madden, S. R., & Hachem, N. (2008). Column stores vs. row stores: How different are they really? In *Proceedings of SIGMOD* (pp. 967-980). Vancouver, BC, Canada: ACM Press. doi:10.1145/1376616.1376712
- Ailamaki, A. (2004). Database architectures for new hardware. In *Proceedings of VLDB* (pp. 1241-1241). Toronto, Canada: Morgan Kaufmann.
- Ailamaki, A., DeWitt, D. J., Hill, M. D., & Skounakis, M. (2001). Weaving relations for cache performance. In *Proceedings of VLDB* (pp. 69-180). Roma, Italy: Morgan Kaufmann.
- Albutiu, M. C., Kemper, A., & Neumann, T. (2012). Massively parallel sort merge joins in main memory multi-core database systems. *PVLDB*, 5(10), 1064-1075.
- Arge, L., Berg, M. D., & Haverkort, H. (2009). Cache-oblivious R-trees. *Algorithmica*, 53(1), 50-68. doi:10.1007/s00453-007-9007-8
- Barber, R., Lohman, G. M., Raman, V., Sidle, R., Lightstone, S., & Schiefer, B. (2015). In-memory BLU acceleration in IBM's DB2 and dashDB: Optimized for modern workloads and hardware architectures. In *Proceedings of ICDE* (pp. 1246-1252). Seoul, South Korea: IEEE Computer Society. doi:10.1109/ICDE.2015.7113372
- Bender, M. A., Demaine, E. D., & Farach-Colton, M. (2000). Cache-oblivious B-trees. In *Proceedings of FOCS 2000* (pp. 399-409). Redondo Beach, CA: IEEE Computer Society.
- Binna, R., Pacher, D., Meindl, T., & Specht, G. (2014). The DCB-tree: A space-efficient delta coded cache conscious B-tree. In *Proceedings of IMDM@VLDB* (pp.30-41). Hangzhou, China: Springer.
- Broneske, D., Breß, S., & Saake, G. (2014). Database scan variants on modern CPUs: A performance study. In *Proceedings of IMDM@VLDB* (pp. 97-111). Hangzhou, China: Springer.
- Cederman, D., & Tsigas, P. (2009). *GPU-quick-sort: A practical quicksort algorithm for graphics processors*. Retrieved May 13, 2016, from <http://www.cse.chalmers.se/research/group/dcs/TechReports/gpuqsort.pdf>
- Dong, J. F., & Xu, X. H. (2007). CSR+-tree: Cache-conscious indexing for high-dimensional similarity search. In *Proceedings of SSDBM* (pp. 14-23). Banff, Canada: IEEE Computer Society. doi:10.1109/SSDBM.2007.9
- Eldawy, A., Levandoski, J. J., & Larson, P. A. (2014). Trekking through Siberia: Managing cold data in a memory-optimized database. *PVLDB*, 7(11), 931-942.
- Faerber, F., Dees, J., Weidner, M., Bauerle, S., & Lehner, W. (2015). Towards a web-scale data management ecosystem demonstrated by SAP HANA. In *Proceedings of ICDE* (pp. 1259-1267). Seoul, South Korea: IEEE Computer Society. doi:10.1109/ICDE.2015.7113374
- Fang, R., He, B. S., Lu, M., Yang, K., Govindaraju, N. K., Luo, Q., & Sander, P. V. (2007). GPUQP: Query co-processing using graphics processors. In *Proceedings of SIGMOD* (pp. 1061-1063). Beijing, China: ACM Press.

- Faust, M., Schwalb, D., & Kruger, J. (2013). Fast column scans: Paged indices for in-memory column stores. In *Proceedings of IMDM@VLDB 2013* (pp. 15-27). Riva del Garda, Italy: Springer.
- Gawade, M., & Kersten, M. L. (2016). Adaptive query parallelization in multi-core column stores. In *Proceedings of EDBT* (pp. 353-364). Bordeaux, France: OpenProceedings.org.
- Golding, B. (2010). *Introduction to NUMA*. Retrieved May 13, 2016, from <https://blogs.msdn.microsoft.com/msdnmagazine/2010/05/06/introduction-to-numa/>
- Govindaraju, N. K., Gray, J., Kumar, R., & Manocha, D. (2006). GPU TeraSort: High performance graphics coprocessor sorting for large database management. In *Proceedings of SIGMOD* (pp. 325-336). Chicago, IL: ACM Press. doi:10.1145/1142473.1142511
- Gray, J., & Reuter, A. (1992). *Transaction processing: Concepts and techniques*. Burlington, MA: Morgan Kaufmann.
- Han, H., Park, S., Jung, H., Fekete, A., Rohm, U., & Yeom, H. Y. (2014). Scalable serializable snapshot isolation for multicore systems. In *Proceedings of ICDE* (pp. 700-711). Chicago, IL: IEEE Computer Society. doi:10.1109/ICDE.2014.6816693
- Harizopoulos, S., Abadi, D. J., Madden, S., & Stonebraker, M. (2008). OLTP through the looking glass, and what we found there. In *Proceedings of SIGMOD* (pp. 981-992). Vancouver, BC, Canada: ACM Press. doi:10.1145/1376616.1376713
- He, B. S., Yang, K., Fang, R., Lu, M., Govindaraju, N. K., Luo, Q., & Sander, P. V. (2008). Relational joins on graphics processors. In *Proceedings of SIGMOD* (pp. 511-524). Vancouver, BC, Canada: ACM Press. doi:10.1145/1376616.1376670
- Jones, E. P. C., Abadi, D. J., & Madden, S. (2010). Low overhead concurrency control for partitioned main memory databases. In *Proceedings of SIGMOD* (pp. 603-614). Indianapolis, IN: ACM Press. doi:10.1145/1807167.1807233
- Jung, H., Han, H., Fekete, A. D., Heiser, G., & Yeom, H. Y. (2013). A scalable lock manager for multicores. In *Proceedings of SIGMOD* (pp. 73-84). New York, NY: ACM Press.
- Kaldewey, T., Lohman, G. M., Muller, R., & Volk, P. B. (2012). GPU join processing revisited. In *Proceedings of DaMoN @SIGMOD* (pp. 55-62). Scottsdale, AZ: ACM Press. doi:10.1145/2236584.2236592
- Karnagel, T., Muller, R., & Lohman, G. M. (2015). Optimizing GPU-accelerated group-by and aggregation. In *Proceedings of ADMS@VLDB 2015* (pp. 13-24). Kohala Coast, HI: Springer.
- Kemper, A., & Neumann, T. (2011). HyPer: A hybrid OLTP&OLAP main memory database system based on virtual memory snapshots. In *Proceedings of ICDE* (pp. 195-206). Hannover, Germany: IEEE Computer Society. doi:10.1109/ICDE.2011.5767867
- Kim, C., Sedlar, E., Chhugani, J., Kaldewey, T., Nguyen, A. D., Blas, A. D., & Dubey, P. et al. (2009). Sort vs. hash revisited: Fast join implementation on modern multi-core CPUs. *PVLDB*, 2(2), 1378–1389.
- Kim, K., Wang, T. Z., Johnson, R., & Pandis, I. (2016). *ERMIA: Fast memory-optimized database system for heterogeneous workloads*. Accepted by SIGMOD. doi:10.1145/2882903.2882905
- Kimura, H. (2015). FOEDUS: OLTP engine for a thousand cores and NVRAM. In *Proceedings of SIGMOD* (pp. 691-706). Melbourne, Victoria, Australia: ACM Press. doi:10.1145/2723372.2746480

- Kissinger, T., Kiefer, T., Schlegel, B., Habich, D., Molka, D., & Lehner, W. (2014). ERIS: A NUMA-aware in-memory storage engine for analytical workload. In *Proceedings of ADMS@VLDB 2014* (pp. 74-85). Hangzhou, China: Springer.
- Krueger, J., Wust, J., Linkhorst, M., & Plattner, H. (2012). Leveraging compression in in-memory databases. In *Proceedings of DBKDA* (pp. 147-153). Saint Gilles, Reunion Island: Xpert Publishing Services (XPS).
- Kruger, J., Grund, M., Tinnefeld, C., Plattner, H., Zeier, A., & Faerber, F. (2010). Optimizing write performance for read optimized databases. In *Proceedings of DASFAA* (pp. 291-305). Tsukuba, Japan: Springer. doi:10.1007/978-3-642-12098-5_23
- Lahiri, T., Chavan, S., Colgan, M., Das, D., Ganesh, A., Gleeson, M., & Zait, M. et al. (2015). Oracle database in-memory: A dual format in-memory database. In *Proceedings of ICDE* (pp. 1253-1258). Seoul, South Korea: IEEE Computer Society. doi:10.1109/ICDE.2015.7113373
- Lang, H., Leis, V., Albutiu, M. C., Neumann, T., & Kemper, A. (2013). Massively parallel NUMA-aware hash joins. In *Proceedings of IMDM@VLDB 2013* (pp. 1-12). Riva del Garda, Italy: Springer.
- Larson, P. A., Blanas, S., Diaconu, C., Freedman, C., Patel, J. M., & Zwilling, M. (2011). High-performance concurrency control mechanisms for main-memory databases. *PVLDB*, 5(4), 298-309.
- Larson, P. A., Hanson, E. N., & Zwilling, M. (2015). Evolving the architecture of SQL Server for modern hardware trends. In *Proceedings of ICDE* (pp. 1239-1245). Seoul, South Korea: IEEE Computer Society. doi:10.1109/ICDE.2015.7113371
- Lee, J., Kim, K., & Cha, S. K. (2001). Differential logging: A commutative and associative logging scheme for highly parallel main memory database. In *Proceedings of ICDE* (pp. 173-182). Heidelberg, Germany: IEEE Computer Society.
- Levy, E., & Silberschatz, A. (1992). Incremental recovery in main memory database systems. *IEEE Transactions on Knowledge and Data Engineering*, 4(6), 529-540. doi:10.1109/69.180604
- Li, L., Jin, P. Q., Yang, C. C., Wu, Z. L., & Yue, L. H. (2016). Optimizing B+-tree for PCM-based hybrid memory. In *Proceedings of EDBT* (pp. 662-663). Bordeaux, France: OpenProceedings.org.
- Lin, J. L., & Dunham, M. H. (1996). Segmented fuzzy checkpointing for main memory databases. In *Proceedings of the ACM symposium on Applied Computing* (pp. 158-165). Philadelphia, PA: ACM Press. doi:10.1145/331119.331168
- Lindstrom, P., & Rajan, D. (2014). Optimal hierarchical layouts for cache-oblivious search trees. In *Proceedings of ICDE* (pp. 616-627). Chicago, IL: IEEE Computer Society. doi:10.1109/ICDE.2014.6816686
- Malviya, N., Weisberg, A., Madden, S., & Stonebraker, M. (2014). Rethinking main memory OLTP recovery. In *Proceedings of ICDE 2014* (pp. 604-615). Chicago, IL: IEEE Computer Society.
- Mueller, R., & Teubner, J. (2009). FPGA: What's in it for a database? *SIGMOD 2009 Tutorial*.
- Neumann, T., Muhlbauer, T., & Kemper, A. (2015). Fast serializable multi-version concurrency control for main-memory database systems. In *Proceedings of SIGMOD* (pp. 677-689). Melbourne, Victoria, Australia: ACM Press. doi:10.1145/2723372.2749436
- Petraki, E., Idreos, S., & Manegold, S. (2015). Holistic indexing in main-memory column-stores. In *Proceedings of SIGMOD* (pp. 1153-1166). Melbourne, Victoria, Australia: ACM Press.
- Rao, J., & Ross, K. A. (2000). Making B+-trees cache conscious in main memory. *SIGMOD Record*, 9(2), 475-486. doi:10.1145/335191.335449
- Ren, K., Thomson, A., & Abadi, D. J. (2012). Lightweight locking for main memory database systems. *PVLDB*, 6(2), 145-156.

Schwalb, D., Kumar, G., Dreseler, M., Anusha, S., Faust, M., Hohl, A., & Deshmukh, P. et al. (2016). Hyrise-NV: Instant recovery for in-memory databases using non-volatile memory. In *Proceedings of DASFAA* (pp.267-282). Dallas, TX: Springer doi:10.1007/978-3-319-32049-6_17

Shamgunov, N. (2014). *The MemSQL in-memory database system*. IMDM@VLDB Keynote Talk. Retrieved May,15,2016, from <http://imdm.ws/2014/keynotes.html>

Sun, X. H. (2015). *The Sluice Gate Theory: Have we found a solution for memory wall?* Retrieved May,15,2016, from <http://www.cs.iit.edu/~scs/research/t-reports.html>

Wust, J., Grund, M., Hoewelmeyer, K., Schwalb, D., & Plattner, H. (2014). Concurrent execution of mixed enterprise workloads on in-memory databases. In *Proceedings of DASFAA* (pp.126-140). Bali, Indonesia: Springer. doi:10.1007/978-3-319-05810-8_9

Wust, J., Grund, M., & Plattner, H. (2013). Dynamic query prioritization for in-memory databases. In *Proceedings of IMDM@VLDB 2013* (pp. 56-68). Riva del Garda, Italy: Springer.

Xi, F., Mishima, T., & Yokota, H. (2014). CARIC-DA: Core affinity with a range index for cache-conscious data access in a multicore environment. In *Proceedings of DASFAA* (pp. 282-296). Bali, Indonesia: Springer. doi:10.1007/978-3-319-05810-8_19

Zeuch, S., & Freytag, J. C. (2015). Selection on modern CPUs. In *Proceeding of IMDM@VLDB 2015* (pp.5:1-5:8). Kohala Coast, HI: ACM Press. doi:10.1145/2803140.2803145

Zhang, H., Chen, G., Ooi, B. C., Wong, W. F., Wu, S. S., & Xia, Y. B. (2015). “Anti-Caching”-based elastic memory management for big data. In *Proceedings of ICDE* (pp. 1268-1279). Seoul, South Korea: IEEE Computer Society.

ADDITIONAL READING

Akhter, S., & Roberts, J. (2010). *Multi-core programming*. Nehru Palace. Delhi, India: BPB Publications.

Arulraj, J., Pavlo, A., & Dulloor, S. (2015). Let’s talk about storage & recovery methods for non-volatile memory database systems. In *Proceedings of SIGMOD* (pp. 707-722). Melbourne, Victoria, Australia: ACM Press. doi:10.1145/2723372.2749441

Cha, S. K. (2014). *Enterprise in-memory database: A revolution beyond crossing the chasm*. IMDM@VLDB Keynote Talk, Retrieved May,15,2016, from <http://imdm.ws/2014/keynotes.html>

DeBrabant, J., Arulraj, J., Pavlo, A., Stonebraker, M., Zdonik, S. B., & Dulloor, S. (2014). *A prodigymenon on OLTP database systems for non-volatile memory*. In *Proceedings of ADMS@VLDB 2014* (pp. 57-63). Hangzhou, China: Springer.

Feng, Z. Q., Lo, E., Kao, B., & Xu, W. J. (2015). *ByteSlice: Pushing the envelop of main memory data processing with a new storage layout*. In *Proceedings of SIGMOD* (pp. 31-46). Melbourne, Victoria, Australia: ACM Press. doi:10.1145/2723372.2747642

Kang, Q. Q., Jin, C. Q., Zhang, Z., & Zhou, A. Y. (2014). *MemTest: A novel benchmark for in-memory database*. In *proceedings of BPOE@ASPLOS/VLDB 2014* (pp. 34-46). Hangzhou, China: Springer.

Kemper, A., & Neumann, T. (2014). *Main-memory database systems*. In *Proceedings of ICDE* (pp. 1310-1310). Chicago, IL, USA: IEEE Computer Society Press.

Plattner, H., & Zeier, A. (2012). *In-Memory data management: Technology and applications* (2nd ed.). Berlin, Germany: Springer. doi:10.1007/978-3-642-29575-1

Viglas, S. D. (2015). *Data management in non-volatile memory*. In *Proceedings of SIGMOD* (pp. 1707-1711). Melbourne, Victoria, Australia: ACM Press.

Wilt, N. (2013). *CUDA handbook: A comprehensive guide to GPU programming*. Boston, Massachusetts: Addison-Wesley Professional.

KEY TERMS AND DEFINITIONS

Cache Oblivious: Database data/index layouts and processing techniques those are independent of parameters of memory hierarchy.

Cache Sensitive (Cache Conscious): Database data/index layouts and processing techniques those are adaptive to the parameters of memory hierarchy.

DSM: Decomposition Storage Model, vertically partitions an n-attribute relation table into columns, each of which is fetched only when queries need it.

GPU/GPGPU: General Purpose Graphics Processing Unit, is not only used for graphic

processing, but also for other tasks such as data processing.

Multi-Core CPU: Several cores are manufactured on a single die for higher performance by parallelism.

MVCC: Multi versioning concurrency control coordinates execution of concurrent transactions by maintaining multiple versions of the same data to avoid mutual interference among these transactions.

NSM: N-ary Storage Model, stores records contiguously starting from the beginning of each disk page, and uses an offset (slot) table at the end of the page to position the beginning of each record (row). NSM has poor cache performance because it loads the cache with unnecessary attributes.

NVRAM: Non-volatile memory is some type of computer memory that can keep information after the power is turned off.

PAX: In Partition Attributes Across, a table is firstly horizontally partitioned into pages, in each page all values of each attribute are grouped together, which greatly improves cache performance.

Ensemble Clustering Data Mining and Databases

Slawomir T. Wierzchon
Polish Academy of Sciences, Poland

INTRODUCTION

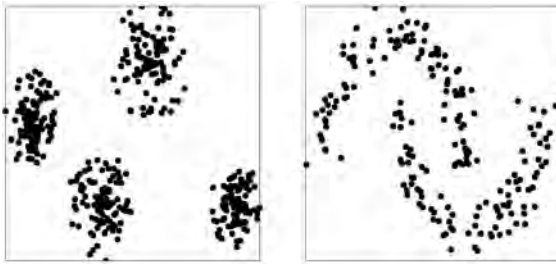
Clustering is an exploratory activity relying upon dividing a given collection X of objects, or entities, into a set of categories, called groups or clusters, in such a way that any two objects placed in the same group have more in common than any two objects assigned to different groups. Consensus clustering has been proposed to overcome some drawbacks of individual clustering algorithms. Usually we assume that the clusters are disjoint subsets of X such that the objects belonging to a single cluster are sufficiently similar to each other (i.e. the clusters should be homogeneous), while objects from different clusters should be sufficiently diversified (i.e. clusters should be well separated). Splitting given collection into disjoint clusters is termed hard clustering. Otherwise we say about soft clustering, i.e. – depending on the formalism used – probabilistic or fuzzy clustering.

The most popular clustering algorithm is the k -means algorithm producing hard partitions – consult (Jain, 2010) for historical background and deeper discussion of its current improvements and variations. Soft version of the algorithm, called fuzzy c -means, was proposed by Bezdek (1981). This author used letter c to name the number of clusters, hence the name of the algorithm. Both the algorithms minimize the squared-error criteria. They are computationally efficient and do not require the user to specify many parameters. However, there are three main disadvantages of both the algorithms. First, they require that the entities must be represented as points in n -dimensional Euclidean space. To alleviate this assumption Hathaway, Davenport and Bezdek (1989) introduced

relational version of fuzzy c -means algorithm: instead of the distance between the points representing the objects, a similarity measure between all pair of objects was used. In case of hard partitions the k -medoids algorithm was proposed: here a dissimilarity measure between pairs of objects replaces the distance measure – see e.g. Section 14.3.10 in (Hastie, Tibshirani, & Friedman, 2009). The second disadvantage results from the way in which objects are assigned to the clusters. Namely, in case of hard k -means each object is located to the cluster with nearest centroid (empirical mean of the cluster). Thus resulting clusters are spherical (more precisely, they are Voronoi regions). A similar assignment rule is used in the fuzzy c -means algorithm. Third disadvantage is that, the clusters should be of approximately similar cardinality and of similar shape. In case of unbalanced clusters erroneous results are frequently obtained. Similarly, if one cluster is located within a ball of small radius and the second – within an ellipsoid with one axis much greater than others, we can obtain erroneous results. Examples of “easy” and “difficult” data are depicted in Figure 1. Left panel presents compact, well separated, convex and linearly separated clusters, while non-convex clusters that are not linearly separated are shown in right panel.

To avoid these disadvantages, the ideas of ensemble methods used by machine learning community to improve results of classification methods, have been adapted to the requirements of clustering. In general, the ensemble methods use multiple models to obtain better predictive performance than could be obtained from any of the constituent models. A nice overview of these

Figure 1. Examples of data that are “easy” (left panel) and “difficult” (right panel) to the *k*-means algorithm



methods used in machine learning can be found e.g. in (Zhou, 2012). When transposed to the field of unsupervised learning (i.e. clustering) this idea translates to collecting multiple partitions of the same data. By combining these partitions, it is possible to obtain a good data partition even when the clusters are not compact and/or not well separated (Jain, 2010). Consensus clustering seems to be especially recommendable to analyze huge datasets. As noted in (Hore, Hall, & Goldgof, 2009): “The advantage of these approaches is that they provide a final partition of data that is comparable to the best existing approaches, yet scale to extremely large datasets. They can be 100,000 times faster while using much less memory.”

Irregular, of complex shape and structure, clusters is only one aspect of the problem. The other is strictly pragmatic. In some applications we are simply “knowledge consumers”, i.e. we use a knowledge created by others. In the context of clustering such knowledge is represented by a set of partitions, and consensus clustering is used to integrate these partitions into consistent form. Strehl and Ghosh (2003) propose the term “Knowledge reuse” to label such an activity. Further, the knowledge acquired in such a way may be prepared using different points of view, different needs or different criteria, and it may be generated by large number of sources. Thus these authors distinguish between feature distributed clustering (FDC) and object distributed clustering (ODC). In first case it is assumed that all the data are available, but each time they are clustered using only a subset

of features, or attributes, characterizing each piece of data. In the second case a fixed set of attributes is used but the collection of data vastly exceeds the size of a typical single memory. So, different partitions are obtained by using only pieces of the whole collection. Again consensus clustering allows consolidate these different clusterings into consistent partition (Hore, Hall, & Goldgof, 2009). A nice illustration of the FDC principle is e.g. the study by Helsen and Green (1991) who applied cluster analysis to define market segments for a new computer system. The dispersions in the opinions collected from 319 users resulted in different partitions. To make final judgments these authors used a Monte-Carlo based simulation method which can be classified as a precursor of consensus clustering.

To summarize, consensus clustering (called also ensembles clustering, or clustering aggregation) is a general purpose method that can be used to improve both the robustness and the stability of partition of large multidimensional datasets. As observed by Howard Firestone (2012): “The advantages of Cluster Ensemble include:

- Combining groupings from alternate and dissimilar sets of variables (e.g., demographics, lifestyle batteries, desired benefits or needs, etc.).
- Including a variety of clustering techniques when building the ensemble.
- Incorporating legacy clusters that are based on internal data.
- Uncovering better, more robust cluster solutions that are less sensitive to sample variations and outliers.
- Being able to find solutions that would not have been uncovered using a single approach.”

In this article we briefly review different approaches to the task of consensus clustering. After careful formulation of the problem we briefly characterize its components, that is: (a) the methods of obtaining different clustering, (b) the

methods of representing acquired knowledge, and (c) the methods of aggregating this knowledge into a consistent and manageable form. Some remarks on available software are provided.

A comprehensive introduction to the subject can be found e.g. in Ghosh and Acharya (2011), Goder and Filkov (2008), Vega-Pons and Ruiz-Shulcloper (2011), or in Section 8.3 of Kuncheva (2004). Gionis, Mannila, and Tsaparas (2007) provide a review of different applications of consensus clustering.

BACKGROUND

The problem of consensus clustering is formulated as follows, see Fred and Jain (2005), Filkov and Skiena (2004), or Meyer and Wessell (2012) for details. Suppose $X = \{x_1, \dots, x_m\}$ is a set of m objects. Let $\mathbf{P} = \{\mathbf{P}^1, \dots, \mathbf{P}^K\}$ be a family of K partitions obtained e.g. by using K different clustering algorithms. We call $\mathbf{P}$ the clustering ensemble. Each partition $\mathbf{P}^i$ splits the dataset into k_i clusters, i.e. $\mathbf{P}^i = \{C_1^i, \dots, C_{k_i}^i\}$, and k_i is the number of clusters in i -th partition. The problem is to find an agreed or compromised partition, $\mathbf{P}^*$, using the information available in the ensemble $\mathbf{P}$. Such a partition should satisfy at least two properties:

1. Consistency with the clustering ensemble $\mathbf{P}$.
2. Robustness to small variations in $\mathbf{P}$.

The first property states that the partition $\mathbf{P}^*$ should be as close as possible to all the partitions $\mathbf{P}^i$ from $\mathbf{P}$. In general, it is an NP-complete optimization problem known as median partition, consult (Cristofor & Simovici, 2002) for details. Some heuristic approaches to such a problem were proposed in the papers quoted at the beginning of this section. The second property requires that the number of clusters and the cluster membership in $\mathbf{P}^*$ should be resistant, or invariant, to small

perturbations in $\mathbf{P}$. Additionally, Fred and Jain (2005) introduce third requirement: The partition $\mathbf{P}^*$ should be consistent with ground truth information (true cluster labels of patterns), if available. Other variants of the compromise partition could be obtained by introducing extra requirements on $\mathbf{P}^*$, e.g. a fixed number of clusters in $\mathbf{P}^*$.

Figure 2 illustrates the problem. Original data, depicted in top left panel, present three nested circles (i.e. non-convex clusters): outer circle consists of 200 points, middle circle – of 100 points, and inner circle consists of 50 points. Since the data are not linearly separated, the k -means algorithm cannot identify proper groups. The panels (b) – (d) depict exemplary partitions obtained from k -means algorithm with different values of k . In this (simplified) case the family $\mathbf{P}$ consists of three partitions: $\mathbf{P}^1$ splits the data set into two subsets (Figure 2b), $\mathbf{P}^2$ – into three subsets (Figure 2c), and $\mathbf{P}^3$ – into four subsets (Figure 2d). The objects assigned to different clusters are depicted by different markers painted in different color. The problem relies upon “smart” integration of these partial pieces of evidence to obtain proper clustering into three groups.

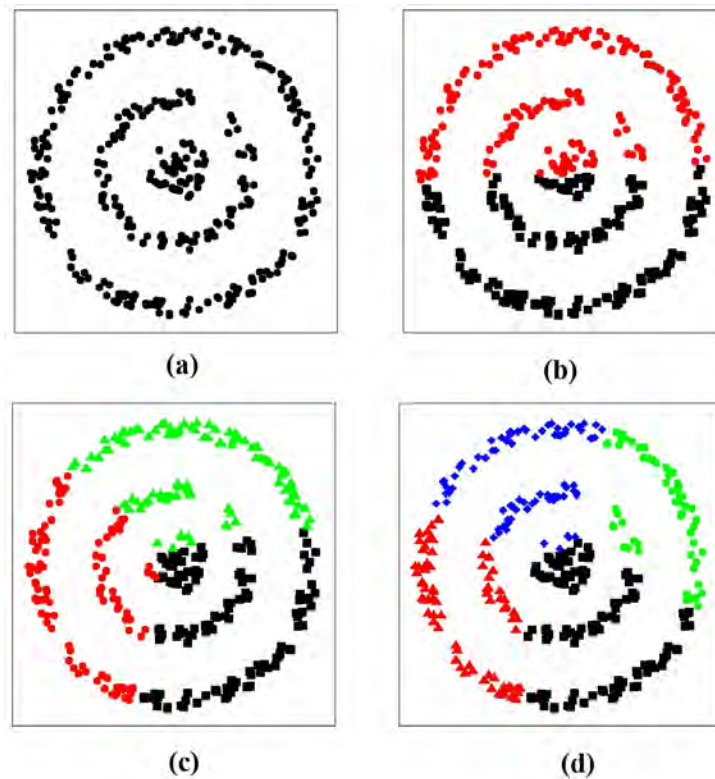
FINDING A CONSENSUS CLUSTERING

Solving the problem of consensus clustering requires answering at least three questions: (i) how to obtain different partitions $\mathbf{P}^1, \dots, \mathbf{P}^K$, (ii) how to represent the family of such partitions into manageable form, and (iii) how to extract the partition $\mathbf{P}^*$ from $\mathbf{P}$? Below we review possible answers to these questions.

Obtaining a Set of Partitions

The motivations for using consensus clustering, described in the “Introduction,” imply also the ways of obtaining a family of partitions. First, the natural sources of different partitions are different views on data. This is typical in applications like

Figure 2. The problem of consensus clustering: (a) original data, (b) – (d) partitions obtained by using the k -means algorithm with $k = 2, 3$, and 4



e.g. market segmentation, where customers may be segmented according to their needs, attitudes, demographic profiles, etc. – see again (Firestone, 2012) and (Helsen & Green, 1991). Bioinformatics (Goder & Filkov, 2008), (Grothjær *et al.*, 2006), Internet security (Zhuang *et al.*, 2012), or text clustering (Marcacini, Domingues, & Rezende, 2013) are another fields where different views should be aggregated to extract complicated clusters from highly dimensional data.

A general recipe to obtain different views is multiple application of any single algorithm with different initializations or by varying its (hyper)-parameters. Averaging over such formulated ensemble enables to extract quite complicated clusters, (Fred & Jain, 2005). Greene *et al.* (2004) lists a number of ways, in which such multiple runs can be realized. Focusing on k -means algorithm they distinguish between:

1. **Plain Method:** Use randomly initialized centroids in each run of the algorithm.
2. **Random- k Method:** Use randomly selected values of the number of clusters, k , from a user-specified interval. Such a method was also mentioned by Fred and Jain (2005). In their experiments these authors used k in the range $\{2, \dots, k_0+10\}$, where k_0 is the true number of clusters in a given dataset.
3. Random- $k+$ method is inspired by the observation that a collection of clusterings generated at a much higher resolution than the value of k_0 can provide better results. Thus Greene *et al.* (2004) used k in the range $\{2, \dots, k_0+30\}$.

Another source of variability is provided by using different clustering methods. Greene *et al.* (2004) used in their experiments standard k -means,

k -medoids and a fast “weak clustering” algorithm. In this last, highly simplified algorithm, k centroids are chosen at random and the remaining entities from the set X are assigned to the cluster with the nearest centroid. Such an allocation is performed only once with no further improvements. Other choices of clustering algorithm are possible; however their simplicity is the main criterion when selected.

Furthermore, variants of other methods developed for supervised learning can be used. Review of such method can be found in Kuncheva (2004). Greene *et al.* (2004) tested four of them:

4. **Bagging:** The subsets of the original data are generated by randomly selecting instances with replacement. Each subset is clustered separately.
5. **Random Subspacing:** Relies upon generating partitions using a randomly selected subset of the original attributes or dimensions.
6. **Random Projection:** Another effective ensemble creation method proposed in Fern and Brodley (2003), and used e.g. in Topchy, Jain, and Punch (2003). It produces a set of dissimilar clusterings by randomly projecting the data onto a subspace of dimension n' randomly selected from the set $\{1, \dots, n\}$.
7. **Heterogeneous Ensembles:** Here each partition is generated using a different clustering algorithm.

Minaei-Bidgoli *et al.* (2014) compare the efficacy of both bootstrap and subsampling (sampling without replacement) techniques in conjunction with several aggregating algorithms (discussed later). They observed that small subsamples generally suffice to represent the structure of the entire data set in the framework of clustering ensembles. Subsamples of small size can reduce computational cost for tasks with distributed sources of data. A unified framework for producing multiple partitions is presented in (Topchy *et al.*, 2005).

Representing Clustering Ensemble

Perhaps the simplest method, used e.g. in (Fred & Jain, 2005), is to form a symmetric co-association matrix A of size $m \times m$ with the entries a_{ij} denoting how many times the elements i and j were clustered together. If necessary, each entry can be divided by K , the number of clusterings, so a_{ij} 's are numbers belonging to the interval $[0, 1]$. Furthermore, we can assign weights, w_r , expressing e.g. a credibility, or simply, importance of each partition, $\mathbf{P}^r$. Then the a_{ij} 's are weighted sums of the form

$$a_{ij} = \frac{1}{K} \sum_{r=1}^K w_r \delta(C_i^r, C_j^r) \quad (1)$$

where C_i^r denotes the cluster to which i -th object has been assigned in r -th partition and $\delta(a, b) = 1$ if $a=b$, and $\delta(a, b) = 0$ otherwise. The main disadvantage of such an approach is the necessity of computing $m(m-1)/2$ elements of the (symmetric) matrix A .

That is why we rather compare partitions by using some measure $q(\mathbf{P}^i, \mathbf{P}^j)$ of agreement between the pairs $\mathbf{P}^i$ and $\mathbf{P}^j$ of partitions – consult Kuncheva (2004), Meila (2007), or Vinh, Epps, & Bailey (2010) for a review of different indices. Generally we distinguish among three classes of measures: based on pair counting, on set-matching, and information theoretic measures. Many of the indices belonging to the first class, like Rand index or its modifications, are popular in information retrieval, (Manning, Raghavan, & Schütze, 2009).

To give an example, denote n_{11} the number of pairs of objects from X , which were assigned to the same cluster in the partition $\mathbf{P}^i$ and to the same cluster in $\mathbf{P}^j$. Let n_{00} be the number of pairs of objects from X that are in different clusters in the partition $\mathbf{P}^i$ and are also in different clusters in $\mathbf{P}^j$. Both n_{00} and n_{11} are agreement quantities as in both partitions the pair of objects have been found to be either similar enough so as to be placed in the same cluster, or dissimilar enough so as to be placed in different clusters. We can define also

the two disagreement quantities n_{01} (meaning the number of pairs of objects that are in different cluster in $\mathbf{P}^i$ and in the same cluster in $\mathbf{P}^j$), and n_{10} (defined symmetrically). Then the Rand index, $R(\mathbf{P}^i, \mathbf{P}^j)$, is defined as follows:

$$R(\mathbf{P}^i, \mathbf{P}^j) = \frac{2(n_{00} + n_{11})}{m(m-1)}. \quad (2)$$

$R(\mathbf{P}^i, \mathbf{P}^j) = 1$ if the partitions agree completely, but does not have a constant value for the case when both partitions are drawn at random. To correct this disadvantage so-called adjusted Rand index was introduced, but its use also causes some problems, (Meila, 2007). Goder and Filkov (2008) used another distance defined as:

$$D_{GF}(\mathbf{P}^i, \mathbf{P}^j) = n_{01} + n_{10} \quad (3)$$

arguing that it is computable in linear time (with respect to the number of elements m), and that it can be updated very fast following one element moves in set partitions (this last property is important when the local search heuristics are used to optimize the distance).

Set-matching based measures, are based on finding matches between clusters in the two partitions. A popular measure is the classification error rate used e.g. in supervised learning. Several other indices are discussed in Meila (2007). Unfortunately, all these indices suffer from two major problems: (i) the number of clusters in the two partitions may be different, and (ii) even when the numbers of clusters are the same, the unmatched part of each matched cluster pair is still put outside consideration.

Lastly, the information theoretic measures are distinguished by their strong mathematical foundation, and ability to detect non-linear similarities. This class of measures has been popularized through the works of Strehl and Ghosh (2003) and Meila (2007), among others. Given two partitions, $\mathbf{P}^i, \mathbf{P}^j$, define n_{rs} to be the number of objects that are common to clusters $C_r^i \in \mathbf{P}^i$, and $C_s^j \in \mathbf{P}^j$. The cardinality of r -th cluster in the partition

$\mathbf{P}^i$ is denoted $n_{r\cdot} = \sum_s n_{rs}$, the cardinality of s -th cluster in the partition $\mathbf{P}^j$ is denoted $n_{\cdot s} = \sum_r n_{rs}$. These cardinalities can be easily converted into the estimates of probabilities p_{rs} , $p_{r\cdot}$, and $p_{\cdot s}$ by dividing corresponding values by the cardinality m . Then we can define the entropy of each partition:

$$H(\mathbf{P}^i) = \sum_{r=1}^{k_i} \frac{n_{r\cdot}}{m} \log \frac{n_{r\cdot}}{m},$$

$$H(\mathbf{P}^j) = \sum_{s=1}^{k_j} \frac{n_{\cdot s}}{m} \log \frac{n_{\cdot s}}{m} \quad (4)$$

and the mutual information between the two partitions

$$I(\mathbf{P}^i, \mathbf{P}^j) = \sum_{r=1}^{k_i} \sum_{s=1}^{k_j} p_{rs} \log \frac{p_{rs}}{p_{r\cdot} \cdot p_{\cdot s}} = \sum_{r=1}^{k_i} \sum_{s=1}^{k_j} \frac{n_{rs}}{m} \log \frac{m \cdot n_{rs}}{n_{r\cdot} \cdot n_{\cdot s}} \quad (5)$$

Meila (2007) gives a nice interpretation of this quantity. As there is no upper bound for the index I , Strehl and Ghosh (2003) proposed its normalized version

$$NMI(\mathbf{P}^i, \mathbf{P}^j) = \frac{I(\mathbf{P}^i, \mathbf{P}^j)}{\sqrt{H(\mathbf{P}^i) \cdot H(\mathbf{P}^j)}}. \quad (6)$$

This is a number from $[0,1]$, and such that $NMI(\mathbf{P}, \mathbf{P}) = 1$.

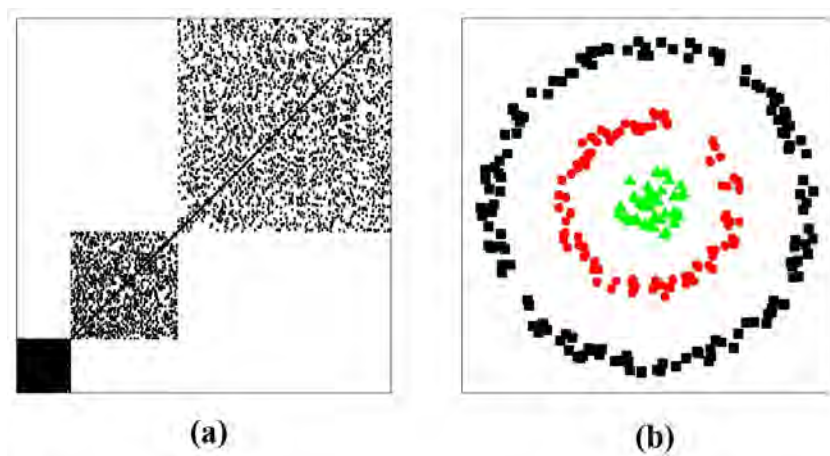
Meila (2007) introduced so-called variation of information

$$VI(\mathbf{P}^i, \mathbf{P}^j) = H(\mathbf{P}^i) + H(\mathbf{P}^j) - 2I(\mathbf{P}^i, \mathbf{P}^j), \quad (7)$$

which is a metric. Wu *et al.* (2009) suggested to use its normalized variant $NVI(\mathbf{P}^i, \mathbf{P}^j) = VI(\mathbf{P}^i, \mathbf{P}^j) / [H(\mathbf{P}^i) + H(\mathbf{P}^j)]$.

Now, let $q(\mathbf{P}, \mathbf{P}^i)$ be one of such defined measures of similarity, for instance, let $q(\mathbf{P}, \mathbf{P}^i) = NVI(\mathbf{P}, \mathbf{P}^i)$. We can introduce a combined quality function

Figure 3. Method of Meyer and Wessell (2012) applied to the data from Figure 2: (a) visualization of the double stochastic matrix P , (b) final partition obtained from this matrix



$$q(\mathbf{P}, \mathbf{P}) = [w_1 q(\mathbf{P}, \mathbf{P}^1) + \dots + w_K q(\mathbf{P}, \mathbf{P}^K)]/K, \quad (8)$$

and define the target partition $\mathbf{P}^*$ as such a partition which optimizes the index (8).

Consensus Function

By consensus function we mean a mapping of a given set of partitions $\mathbf{P}$ to a consensus representative $\mathbf{P}^*$. Again the simplest (conceptually) way to obtain such a representative is to use the co-association matrix A and treat it as a similarity matrix. Such a matrix can be thus clustered in a number of ways. Particularly, hierarchical clustering methods with single-link, complete-link, and average-link criteria could be used (Fred & Jain, 2005; Kuncheva, 2004). Strehl and Ghosh (2003) proposed three other graph-based approaches. Their adaptation to fuzzy consensus clustering is given in Punera and Ghosh (2008). Recently Meyer and Wessell (2012) introduced a method inspired by Simon-Ando theory of variable aggregation. This method reduces to finding best approximation of the co-association matrix A by a doubly stochastic matrix P , i.e. a matrix which rows and columns sum to 1. Extracting block-diagonal structure of this matrix reveals cluster structure of a given dataset. Figure 3 illustrates

the effect of such a procedure when applied to the data from Figure 2a.

Another group of methods uses the quality function (8) and searches for the optimal partition. Kuncheva (2004) provides a concise review of such approaches. Goder and Filkov (2008) search for the partition that is in minimal distance to all the partitions from $\mathbf{P}$ (i.e. they use the D_{GF} distance). These authors review a number of heuristics solving such stated task. What important, they have implemented these heuristics (the code is publicly available) and compared their performance in terms of efficacy and efficiency on many real and artificial datasets. They conclude that the clustering-based methods are generally faster, while the local-search heuristics result in a better performance. Thus, they propose an efficient hybrid algorithm, a greedy local search method starting from an initial consensus provided by the average link clustering. Cristofor and Simovici (2002) proposed to use genetic algorithm to find target partition.

FUTURE RESEARCH DIRECTIONS

Consensus clustering is a method proposed for improving robustness, stability and accuracy of clustering real data. In addition to many works

on this subject, a number of publicly software is available. For instance Goder and Filkov (2008) have written a C++ library for computing set partitions. MATLAB implementations are provided by Yu, Wong, and Wang (2007), Grotkjær *et al.* (2006), and Strehl (2003). Hornik (2005) offers elaborated implementation in R language, and Simpson, Armstrong, and Jarman (2010) describe an R package for analyzing microarray data. Lastly, Marcacini, Domingues, and Rezende (2013) provide a Java package for text mining.

Due to the unsupervised nature of the problem it is not possible to suggest the best consensus clustering method. But choosing a proper method would be easier if there was the theoretical foundation of combining multiple clustering results. Although a step toward this direction was suggested by Masson and Denoeux (2011), a solution to this problem is still in its early stages. Apart theoretical foundation, improving computational complexity in aggregating partial results seems to be an important investigation in future. Harnessing Nature-based methods to solve this task seems to be an intriguing direction – see e.g. Ghaemi *et al.* (2011) for a recent proposition. Another direction is initiated by Gullo, Domeniconi, & Tagarelli (2013). These authors suggest to use so-called projective clustering to aggregate different data partitions. On the other hand, Yousefnezhad and Zhang, D. (2015) proposed a novel framework for clustering problems, by exploiting some concepts from the areas of community detection and graph based clustering. Other advanced topics, like ensemble methods for semi-supervised learning, are discussed in Chapter 8 of (Zhou, 2012).

CONCLUSION

A solution returned by a single clustering algorithm is inevitably biased due to the peculiarities of such an algorithm. Consensus clustering was proposed to overcome this drawback. In addition to improving robustness and stability results it is perfectly fitted to integrate solutions from mul-

tle distributed sources of data. Furthermore this approach can also be used in multiobjective clustering as a compromise between individual partitions obtained by using conflicting objective functions.

REFERENCES

- Bezdek, J. C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*. Norwell, MA: Kluwer Academic Publishers; doi:10.1007/978-1-4757-0450-1
- Cristofor, D., & Simovici, D. (2002). Finding median partitions using information-theoretical-based genetic algorithms. *Journal of Universal Computer Science*, 8(2), 153–172.
- Fern, X. Z., & Brodley, C. E. (2003). Random projection for high dimensional data clustering: A cluster ensemble approach. In *Proceedings of the 20-th International Conference on Machine Learning, ICML 2003*, (Vol. 3, pp. 186–193). AAAI Press.
- Filkov, V., & Skiena, S. (2004). Integrating microarray data by consensus clustering. *International Journal of Artificial Intelligence Tools*, 13(04), 863–880. doi:10.1142/S0218213004001867
- Firestone, H. (2012). *Which segmentation solution is best? Cluster Ensembles! Ideas, Opinions, and Perspective from RTi Research*. Retrieved June 17, 2016 from <http://rtiresearch.com/blog/which-segmentation-solution-is-best-cluster-ensembles/?Itemid=73>
- Fred, A. L., & Jain, A. K. (2005). Combining multiple clusterings using evidence accumulation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(6), 835–850. doi:10.1109/TPAMI.2005.113 PMID:15943417
- Ghosh, J., & Acharya, A. (2011). Cluster ensembles. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1(4), 305–315. doi:10.1002/widm.32

- Gionis, A., Mannila, H., & Tsaparas, P. (2007). Clustering aggregation. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 1(1), article no. 4. doi:10.1145/1217299.1217303
- Goder, A., & Filkov, V. (2008). Consensus Clustering Algorithms: Comparison and Refinement. In *Proceedings of the Workshop on Algorithm Engineering and Experiments*. Retrieved June 17, 2016 from <http://web.cs.ucdavis.edu/~filkov/papers/consensuseval.pdf>
- Greene, D., Tsymbal, A., Bolshakova, N., & Cunningham, P. (2004, June). Ensemble clustering in medical diagnostics. In *Computer-Based Medical Systems, 2004. CBMS 2004. Proceedings. 17th IEEE Symposium on* (pp. 576-581). IEEE. doi:10.1109/CBMS.2004.1311777
- Grotkjær, T., Winther, O., Regenberg, B., Nielsen, J., & Hansen, L. K. (2006). Robust multi-scale clustering of DNA microarray data with the consensus algorithm. *Bioinformatics (Oxford, England)*, 22(1), 58–67. PubMed doi:10.1093/bioinformatics/bti746
- Gullo, F., Domeniconi, C., & Tagarelli, A. (2013). Projective clustering ensembles. *Data Mining and Knowledge Discovery*, 26(3), 452–511. doi:10.1007/s10618-012-0266-x
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). New York, NY: Springer-Verlag; doi:10.1007/978-0-387-84858-7
- Hathaway, R. J., Davenport, J. W., & Bezdek, J. C. (1989). Relational duals of the c-means algorithms. *Pattern Recognition*, 22(2), 205–212. doi:10.1016/0031-3203(89)90066-6
- Helsen, K., & Green, P. E. (1991). A computational study of replicated clustering with an application to market segmentation. *Decision Sciences*, 22(5), 1124–1141. doi:10.1111/j.1540-5915.1991.tb01910.x
- Hore, P., Hall, L. O., & Goldgof, D. B. (2009). A scalable framework for cluster ensembles. *Pattern Recognition*, 42(5), 676–688. PubMed doi:10.1016/j.patcog.2008.09.027
- Hornik, K. (2005). A CLUE for CLUster Ensembles. *Journal of Statistical Software*, 14(12), 1–25. doi:10.18637/jss.v014.i12
- Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8), 651–666. doi:10.1016/j.patrec.2009.09.011
- Kuncheva, L. I. (2004). *Combining Pattern Classifiers: Methods and Algorithms*. Hoboken, NJ: John Wiley & Sons; doi:10.1002/0471660264
- Manning, Ch. D., Raghavan, P., & Schütze, H. (2009). *An Introduction to Information Retrieval*. Cambridge, UK: Cambridge University Press.
- Marcacini, R. M., Domingues, M. A., & Rezende, S. O. (2013, May). Improving consensus clustering of texts using interactive feature selection. In *Proceedings of the 22nd international conference on World Wide Web companion* (pp. 237-238). doi:10.1145/2487788.2487910
- Masson, M. H., & Denoeux, C. (2011). Ensemble clustering in the belief functions framework. *International Journal of Approximate Reasoning*, 52(1), 92–109. doi:10.1016/j.ijar.2010.04.007
- Meila, M. (2007). Comparing clusterings – an information based distance. *Journal of Multivariate Analysis*, 98(5), 873–895. doi:10.1016/j.jmva.2006.11.013
- Meyer, C. D., & Wessell, Ch. D. (2012). Stochastic data clustering. *SIAM Journal on Matrix Analysis and Applications*, 33(4), 1214–1236. doi:10.1137/100804395
- Minaei-Bidgoli, B., Topchy, A. P., & Punch, W. F. (2014). Effects of resampling method and adaptation on clustering ensemble efficacy. *Artificial Intelligence Review*, 41(1), 27–48. doi:10.1007/s10462-011-9295-x

- Monti, S., Tamayo, P., Mesirov, J., & Golub, T. (2003). Consensus clustering: A resampling-based method for class discovery and visualization of gene expression microarray data. *Machine Learning*, 52(1/2), 91–118. doi:10.1023/A:1023949509487
- Punera, K., & Ghosh, J. (2008). Consensus-based ensembles of soft clusterings. *Applied Artificial Intelligence: An International Journal*, 22(7-8), 780–810. doi:10.1080/08839510802170546
- Simpson, T. I., Armstrong, J. D., & Jarman, A. P. (2010). Merged consensus clustering to assess and improve class discovery with microarray data. *BMC Bioinformatics*, 11(1), 590. doi:10.1186/1471-2105-11-590
- Strehl, A. (2003). Cluster Analysis and Cluster Ensemble Software. Retrieved June 17, 2016 from <http://strehl.com/soft.html>
- Topchy, A., Jain, A., & Punch, W. (2005). Clustering ensembles: Models of consensus and weak partitions. *Pattern Analysis and Machine Intelligence*, 27(12), 1866–1881. doi:10.1109/TPAMI.2005.237
- Topchy, A., Jain, A. K., & Punch, W. (2003, November). Combining multiple weak clusterings. In *Third IEEE International Conference on Data Mining. ICDM 2003* (pp. 331-338). IEEE.
- Vega-Pons, S., & Ruiz-Shulcloper, J. (2011). A survey of clustering ensemble algorithms. *International Journal of Pattern Recognition and Artificial Intelligence*, 25(03), 337–372. doi:10.1142/S0218001411008683
- Vinh, N. X., Epps, J., & Bailey, J. (2010). Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance. *Journal of Machine Learning Research*, 11(4), 2837–2854.
- Wu, J., Chen, J., Xiong, H., & Xie, M. (2009). External validation measures for k-means clustering: A data distribution perspective. *Expert Systems with Applications*, 36(2), 6050–6061. doi:10.1016/j.eswa.2008.06.093
- Yousefnezhad, M., & Zhang, D. (2015) Weighted spectral cluster ensemble. *IEEE International Conference on Data Mining* (pp. 549-558), IEEE. doi:10.1109/ICDM.2015.145
- Yu, Z., Wong, H.-S., & Wang, W. (2007). Graph-based consensus clustering for class discovery from gene expression data. *Bioinformatics (Oxford, England)*, 23(21), 2888–2896. doi:10.1093/bioinformatics/btm463
- Zhou, Z.-H. (2012). *Ensemble Methods: Foundations and Algorithms*. Chapman & Hall/CRC.
- Zhuang, W., Ye, Y., Chen, Y., & Li, T. (2012). Ensemble clustering for Internet security applications. *IEEE Transactions on Systems, Man and Cybernetics. Part C, Applications and Reviews*, 42(6), 1784–1796. doi:10.1109/TSMCC.2012.2222025

ADDITIONAL READING

- Abraham, A., Grosan, C., & Ramos, V. (2006). *Swarm Intelligence in Data Mining*. Berlin, Heidelberg: Springer; doi:10.1007/978-3-540-34956-3
- Azimi, J., Cull, P., & Fern, X. (2009). Clustering ensembles using ants algorithm. In J. Mira & J. M. Ferrández (Eds.), *J.-R. A. Sanchez, F. Paz, & J. Toledo (rds.) Methods and Models in Artificial and Natural Computation. A Homage to Professor Mira's Scientific Legacy, LNCS 5601* (pp. 295–304). Berlin, Heidelberg: Springer; doi:10.1007/978-3-642-02264-7_31

- Ben-David, S.,.... (2009) Clustering: Science or Art? Towards Principled Approaches. A NIPS 2009 Workshop, Dec. 11, 2009. The Hilton Whistler Resort & Spa. Retrieved June 17, 2016 from <http://stanford.edu/~rezab/nips2009workshop/>
- Brunet, J. P., Tamaro, P., Golub, T. R., & Mesirov, J. P. (2004). Metagenes and molecular pattern discovery using matrix factorization. [PubMed doi:10.1073/pnas.0308531101]. *Proceedings of the National Academy of Sciences of the United States of America*, 101(12), 4164–4169.
- Castro, P. D., Coelho, G. P., Caetano, M. F., & Von Zuben, F. J. (2005). Designing ensembles of fuzzy classification systems: An immune-inspired approach. In Jacob, C., Pilat, M., Bentley, P., Timmis, J. (Eds.) *Proceedings of 4-th International Conference on Artificial Immune Systems, ICARIS 2005, LNCS 3627* (pp. 469–482). Berlin, Heidelberg: Springer. doi:10.1007/11536444_36
- Coelho, A. L., Fernandes, E., & Faceli, K. (2010). Inducing multi-objective clustering ensembles with genetic programming. *Neurocomputing*, 74(1-3), 494–498. doi:10.1016/j.neucom.2010.09.014
- Domeniconi, C. & Al-Razgan, M. (2009). Weighted cluster ensembles: Methods and analysis. *ACM Trans. on Knowledge Discovery from Data*, 2(5), 17, 1-17.
- Duda, R. E., Hart, P. E., & Stork, D. G. (2000). *Pattern Classification* (2nd ed.). Hoboken, NJ: John Wiley & Sons.
- Faceli, K., de Carvalho, A. C., & de Souto, M. C. (2007). Multi-objective clustering ensemble. *International Journal of Hybrid Intelligent Systems*, 4(3), 145–156. Biburl: <http://dblp.uni-trier.de/rec/bib/journals/ijhis/FaceliCS07>
- Ghosh, J., Strehl, A., & Merugu, S. (2002). A consensus framework for integrating distributed clusterings under limited knowledge sharing. In *Proceedings of the NSF Workshop on Next Generation Data Mining* (pp. 99-108). Retrieved November 14, 2015 from <http://www.csee.umbc.edu/~kolari1/Mining/ngdm/ghosh.pdf>
- Hennig, H. (2015) What are the true clusters? *Pattern Recognition Letters*, 64, 53–62, doi: 10.1016/j.patrec.2015.04.009
- Xu, X., & Deng, S. (2005). A cluster ensemble method for clustering categorical data. *Information Fusion*, 6(2), 143–151. doi: 10.1016/j.patrec.2015.04.009
- Jiao, L. (2010). Soft spectral clustering ensemble applied to image segmentation. *Frontiers of Computer Science in China*, 5(1), 66-78. doi: 10.1016/j.patrec.2015.04.009
- Lancichinetti, A., & Fortunato, S. (2012). Consensus clustering in complex networks. *Scientific Reports*, 2, 336. doi:10.1038/srep00336 PMID:22468223
- Liu, R., Zhang, W., Jiao, L., & Liu, F. (2010, July). A multiobjective immune clustering ensemble technique applied to unsupervised SAR image segmentation. In *Proceedings of the ACM International Conference on Image and Video Retrieval* (pp. 158-165). ACM. doi:10.1145/1816041.1816067
- Monti, S., Tamayo, P., Mesirov, J., & Golub, T. (2003). Consensus clustering: A resampling-based method for class discovery and visualization of gene expression microarray data. *Machine Learning*, 52(1/2), 91–118. doi:10.1023/A:1023949509487
- Topchy, A., Jain, A. K., & Punch, W. (2004). A mixture model of clustering ensembles. In *Proceedings of the SIAM International Conference on Data Mining* (pp. 379-390). doi:10.1137/1.9781611972740.35

Yang, Y., & Kamel, M. (2003, April). Clustering ensemble using swarm intelligence. In *Proceedings of the 2003 IEEE Swarm Intelligence Symposium, SIS'03*. (pp. 65-71). IEEE. doi:10.1109/SIS.2003.1202249

Yearwood, J., Webb, D., Ma, L., Vamplew, P., Ofoghi, B., & Kelarev, A. (2009, December). Applying clustering and ensemble clustering approaches to phishing profiling. In *Proceedings of the Eighth Australasian Data Mining Conference* (Vol., 101 pp. 25-34). Australian Computer Society, Inc.

Zhang, X., Jiao, L., Liu, F., Bo, L., & Gong, M. (2008). Spectral clustering ensemble applied to SAR image segmentation. *IEEE Transactions on Geoscience and Remote Sensing*, 46(7), 2126–2136. doi:10.1109/TGRS.2008.918647

KEY TERMS AND DEFINITIONS

Cluster Analysis: It is an instance of unsupervised learning. In practice cluster analysis reduces to partitioning a given data set into an appropriate number of meaningful groups. However, the very nature of clustering is weakly understood. Consult e.g. the papers collected by Ben-David, *et al.* (2009), or Henning (2015).

Entropy: In information theory this notion, introduced by Claude Shannon, is used to express unpredictability of information content. For instance, if a data set containing n items was divided into k groups each comprising n_i items, then the entropy of such a partition is $H = p_1 \log(p_1) + \dots + p_k \log(p_k)$, where $p_i = n_i/n$. In case of two alternative partitions, the mutual information is a measure of the mutual dependence between these partitions.

NP-Hard Problems: A notion used in computational complexity theory. It denotes a class of problems that are, informally, “at least as hard as the hardest problems in NP”. The shorthand NP means “non-deterministic polynomial-time.”

Supervised Learning: The problem of supervised learning is formulated as follows: Let $X = \{x_1, \dots, x_m\}$ be a set of m objects or entities. Suppose these objects come from k classes and the membership to each object to exactly one class is known in advance. Then such a set is termed a training set T . It has the form $T = \{(x_1, l_1) \dots, (x_m, l_m)\}$ where l_i denotes the label of a class to which i -th object belongs. The aim of supervised learning is to find a mapping $f: X \rightarrow L$, such that $f(x_i) = l_i$, for all $i = 1, \dots, m$. The function f is said to be decision function or decision rule. If the labels l_i are not known we say about unsupervised learning. Its task is to discover a structure in the set X .

Unsupervised Learning: Contrary to the supervised case we have only the data set $X = \{x_1, \dots, x_m\}$ of m objects or entities. The aim of unsupervised learning is to find inner structure in this set, or equivalently, to find some regularities in this set. Clustering is a possible tool used to extract such information, but other approaches (e.g. self organizing maps) may be used as well. If the labels l_i are known for a (usually small) subset of X we are faced with semi-supervised learning.

Voronoi Regions: Suppose that the objects from the set X are represented as points in n -dimensional Euclidean space $\mathbf{R}^n$, and stands for the centroid (empirical mean) of i -th cluster C_i . The Voronoi region associated with this centroid is the set of all points in $\mathbf{R}^n$ whose distance to is not greater than their distance to the other centroids.

Graph-Based Concept Discovery

Alev Mutlu

Kocaeli University, Turkey

Pinar Karagoz

Middle East Technical University, Turkey

Yusuf Kavurucu

Turkish Naval Research Center Command, Turkey

INTRODUCTION

Multi-relational data mining (MRDM) (Džeroski, 2003) is concerned with discovering patterns hidden in data that is stored in relational format, i.e. as a database. One of the most commonly addressed tasks in MRDM is concept discovery (Džeroski, 2003), where the problem is inducing logical definitions of a specific relation, called *target relation*, in terms of other relations called *background knowledge*. Given a database consisting of kinship relations such that *mother* is the target relation; *father* and *wife* and are among background knowledge, a typical concept discovery system would induce a concept rule like *mother(A,B):-father(C,B), wife(A,C)*.

Concept discovery can be considered as a predictive learning task where target relation instances are generally represented as ground facts, background knowledge is represented either intensional or extensional; and concept descriptors are usually in the form of Horn clauses. Logic-based, more specifically Inductive Logic Programming (Muggleton, 1991) (ILP)-based, and graph-based approaches are two competing counterparts in concept discovery. ILP-based approaches (Quinlan, 1990; Muggleton 1995) benefit from the powerful data representation framework provided by first order logic and the easily interpretable concept descriptors they discover. Such systems suffer from long running times mainly due to evaluation of the large search space they

build, and are vulnerable to miss certain concept descriptors due to the local minima problem (Richards & Mooney, 1992). Graph-based approaches (Gonzalez, Holder, & Cook, 2002; Yoshida & Motoda, 1995) also provide powerful mechanisms to represent relational data and have heavily studied algorithms that can be modified to find concepts in graph data. Graph-based concept discovery systems can be classified as substructure-based approaches and pathfinding-based approaches. In the former, the assumption is that concepts should appear as frequent substructures in a graph, hence computationally expensive algorithms like subgraph isomorphism need to be employed. The assumption behind the systems that fall into the second category is that concepts should appear as finite length paths that connect certain vertices, hence such system need to employ path-profiling algorithms.

In this chapter we aim to introduce concept discovery problem and discuss several issues of graph-based concept discovery in depth. To this aim, in this chapter, we provide fundamentals of concept discovery in general, and in detail discuss state-of-the-art graph-based concept discovery methods by means of (i) data representation, (ii) search method, and (iii) concept evaluation mechanism.

The rest of the article is organized as follows. In the second section, an introduction to concept discovery and definitions of fundamental terms in concept discovery are provided; the concept

discovery problem is formally expressed. In the third section, classification of graph-based concept discovery approaches is provided; knowledge representation, search methods, and evaluation methods are introduced. In this section a running example is provided for a state-of-the-art pathfinding-based system presented in (Abay, Mutlu, & Karagoz, 2015a). In the fourth section, possible research directions for graph-based concept discovery are presented.

BACKGROUND

Multi-relational data mining is concerned with inducing patterns hidden in data stored in relational model, i.e. in a relational database. Traditional data mining algorithms are designed to work on flat files and cannot directly be work on such data. Two main directions in mining relational data are (i) converting the relational data into single table and then employing propositional learners (Kramer, Lavrač, & Flach, 2001), and (ii) developing new algorithms that can directly work on relational data (Džeroski, 2003). The main disadvantage of the first approach is the possible loss of information

during propositionalization process. Most of the algorithms that belong to the second direction have their roots in ILP. ILP is concerned with inducing general theories from examples. It was first proposed for learning binary classification rules, but now can model more complex data mining problems.

One of the most commonly addressed tasks in MRDM is concept discovery. Concept is a set of frequent patterns embedded in the features of the concept instances and its relations to other objects (Toprak, Senkul, Kavurucu, & Toroslu, 2007). *Completeness* and *consistency* are two measures to test quality of concept descriptors (Muggleton & De Raedt, 1994). A concept descriptor is called complete if it explains all of the positive target instances and consistent if it explains none of the negative target instances. As real world data is usually noisy, completeness and consistency are hard to meet in their pure definitions, hence they are extended to explain as many positive target instances as possible and as few negative target instances as possible, respectively (Pham & Afify, 2006). Concept discovery systems generally follow iterative covering approach to induce concept descriptors. They start with an initial hypothesis

Algorithm 1. Generic concept discovery algorithm

```

Input: E: target instances, B: background knowledge
Output: H: Complete and consistent concept descriptors
1: while E  $\neq \emptyset$  or H'  $\neq \emptyset$  do
2:     H' =  $\emptyset$ 
3:     Start with an initial hypothesis set H'
4:     for all h  $\in$  H' do
5:         refine(h)
6:         evaluate(h, B)
7:         if good(h) then
8:             cover(E, h)
9:             H' = H'  $\cup$  h
10:        end if
11:    end for
12:    H = H  $\cup$  H'
13: end while
14: return H

```

and refine it until complete and consistent concept descriptors are obtained. Once some concept descriptors are discovered, target instances that are deducible by the concept descriptor are removed from the data set – this step is called covering, and the process restarts with the remaining target instances. The algorithm terminates when all the target instances are explained; or the number of uncovered target instances drops below a certain threshold; or when no more concept descriptors can be discovered to model the remaining target instances. Algorithm 1 presents the generic concept discovery process.

ILP-based and graph-based approaches are two competitors in concept discovery. ILP-based approaches have long dominated the research in concept discovery and have been successfully applied in several domains such as engineering (Dolsak & Muggleton, 1992), bioinformatics (Turcotte, Muggleton, & Sternberg, 2001) and environmental sciences (Džeroski, Jacobs, Molina, Moure, Muggleton, & Van Laer, 1998). ILP-based concept discovery systems are classified as bottom-up and top-down. In bottom-up systems, an initial hypothesis, called *bottom clause*, that describes a single positive target instance is constructed and is iteratively generalized to cover as many positive target instances as possible and as few negative target instances as possible. In top-down systems, an overly general initial concept descriptor that models all the target instances is constructed and is iteratively specialized to cover as few negative target examples as possible. Hybrid systems that combine properties of both approaches are also studied in the literature.

Another research direction in concept discovery is based on graphs. Among others, a common problem faced by ILP-based concept discovery systems is the so-called locale plateau problem, where refinement by one literal at a time operators are insufficient to improve rule quality. Richards and Mooney (Richard & Mooney, 1992) proposed a graph pathfinding-based approach, which refines the concept descriptors with multiple literals at a time to solve the local plateau problem. Thence-

forth, several graph-based concept discovery systems have been proposed with successful applications. In literature there exist several studies that compare performance of ILP-based and graph-based concept discovery systems and argue that graph-based systems are competitive to ILP-based ones and might be of more expressive power (Gonzalez, Holder, & Cook, 2000).

In graph-based concept discovery systems relational data is represented as a graph and graph algorithms are utilized to mine the concept descriptors. Graph-based concept discovery systems are classified as substructure-based and pathfinding-based approaches. In substructure-based approaches, it is assumed that concepts should appear as frequent substructures in a graph; hence such substructures are sought. In pathfinding-based approaches, it is assumed that concept descriptors should appear as fixed length paths that originate from target instances and go through similar nodes. There also exist hybrid systems that combine properties of substructure-based and pathfinding-based approaches. Substructure discovery is known to be NP (Yan & Han, 2002) and path finding approaches require extensive path profiling mechanism.

GRAPH-BASED CONCEPT DISCOVERY

Input to graph-based concept discovery systems is a graph representation of the relational data. Graphs are one of the fundamental data structures and data representation frameworks of computer science. A graph, G , consists of non-empty set of *vertices*, V , and a set of *edges*, E . Edges connect two-element subsets of nodes which are called *endpoints* of the edges. G is called directed graph if endpoints are ordered and undirected otherwise. A graph is *labeled* if its vertices or edges or both are assigned specific values.

Graph-based concept discovery systems primarily differ from each other by means of the structure they consider to represent concepts. Other

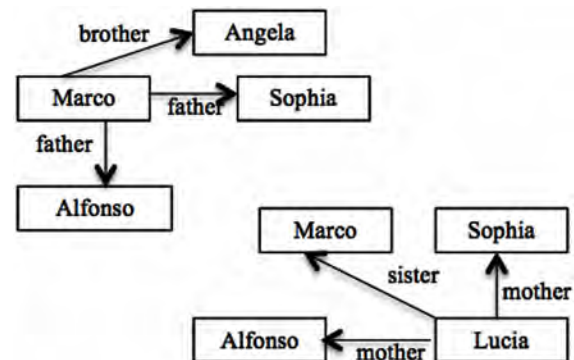
Table 1. Sample kinship dataset

Target Instances	Background Knowledge	
husband(Marco, Lucia).	father(Marco, Alfonso). father(Marco, Sophia). mother(Lucia, Alfonso). mother(Lucia, Sophia).	brother(Marco, Angela). sister(Lucia, Emilio).

than this, graph-based concept discovery systems also differ from each other by means of data representation, search technique, noise handling mechanism, and concept descriptor evaluation method. In the rest of this section we introduce pioneering works and state-of-the-art studies and discuss how they handle the above listed issues.

One of the earliest works that utilized graphs for concept discovery is Relational Pathfinding (Richard & Mooney, 1992). It is a pathfinding-based approach, where data is represented as an directed labeled graph where vertices represent relation arguments and edges connect arguments whose values form a fact, either a target instance or a background fact, e.g. representation of Table 1 is as in Figure 1. Although it is not clearly stated in the study, Relational Pathfinding allows target instances to be n-ary relations while limiting the background data with unary and binary relations. The system picks random target instance at random, and creates as many vertices as its arguments and marks them as unexpanded. Iteratively, each yet-unexpanded vertex is expanded by adding vertices that represent arguments of a relation where that particular

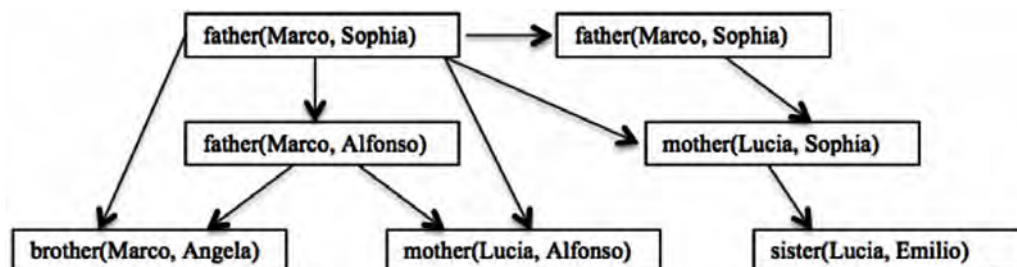
Figure 1. Most common two data representation formats: attribute-based representation



vertex takes part. Once a path that connects any two arguments of the target instance is established, the path is output as a concept descriptor.

Relational Paths Based Learning (RPBL) (Gao, Zhang, & Huang, 2009a) is pathfinding-based approach where data is labeled directed graph. Vertices represent facts from background data and edges connect those vertices that share an argument of the same type and value, e.g. representation of Table 1 is as in Figure 2. To discover concept descriptors, RBPL picks a number of target instances and constraints the source, s , and target, t , nodes based on them. If t is a target instance of form $t(A, B)$, then s is $T(A, *)$ and t is $p(?, B)$ where p is a relation from the background data. To find paths, RBPL employs breath first search. Once a vertex that matches p is discovered, the path is pinned as a concept descriptor. Similar to Relational Pathfinding algorithm, constants are replaced with variables to obtain the final concept descriptors.

Figure 2. Most common two data representation formats: fact-based representation



In (Gao, Zhang, & Huang, 2009b), RPBL is extended to handle domain theories and discover recursive concept descriptors. Domain theories are intensional definitions of relations. If such definitions are provided, graph is extended by adding new edges that connect vertices that hold for the domain theories. To discover recursive concept descriptors, target instances are also added to the graph. It follows the same search method with RPBL to discover concept descriptors.

Mode Directed Path Finding (MDPF) (Ong, de Castro Dutra, Page, & Costa, 2005) is an application of pathfinding-based concept discovery on bottom clauses. It differs from Relational Pathfinding and RPBL not only by means of working on a bottom clause, but also by employing hypergraph as a representation framework. Each literal in the body of the bottom clause is represented as a vertex in the hypergraph. Edges are created based on the input/output mode declarations among literal arguments. The head literal, i.e. target relation to be learnt, is represented with n distinct vertices, n being the arity of the head literal, which are sources of paths. Initially each path is of length 0 and is associated with a list, L_p , and contains the value of the vertex. Concept discovery method of MDPF is similar to Relational Pathfinding method. Vertices are expanded in breadth-first manner with their related facts, and at end of the each expansion the newly reachable values are added to the list. To find the concept descriptors, Mode Directed Path Finding compares each list against every other list. If any two lists intersect they are merged to form the final concept descriptor.

Another study that works on bottom clauses and makes use of mode declarations is Head Output Connected (HOC) Predicate learning system (Santos, Tamaddoni-Nezhad, & Muggleton, 2009). HOC class of clause is a special case of clauses induced by general pathfinding-based concept discovery systems. A clause $h \leftarrow b_1, b_2, \dots, b_n$ is said to be HOC if and only if (i) input variables for each body literal are either subset of the head input variables or are found in a previous body literal, and (ii) all head output variables are instantiated in

the body. Similar to Mode Directed Path Finding approach, HOC system builds a directed labeled graph from body literals of the bottom clause. It connects the vertices that contain input variables to those that have those input variables as output variables. It employs breadth-first search to find concept descriptors.

Some recent studies extended pathfinding-based approaches to graph databases. A recent study by Abay et al (Abay et al., 2015a) uses Neo4j graph database to mine relational data. Similar to Relational Pathfinding approach, vertices represent literal arguments and edges connect those that form a fact. It follows association rule mining techniques to guide the search and prune the intermediate concept descriptors. Best concept descriptor is selected based on F1 metric. The proposed approach is limited to work on binary relations that do not contain numeric values. The proposed method can infer concept descriptors of HOC class. This work also demonstrates applicability of No-SQL systems on concept discovery. The method is extended in (Abay, Mutlu, & Karagoz, 2015b) to handle n-ary relations and numeric attributes.

SUBDUE (Holder, Cook, & Djoko, 1994) is one of the pioneering works in the substructure-based concept discovery problem. It inputs a labeled, directed multi-graph where arguments are represented as vertices and edges connect arguments that form a valid fact and are labeled after the relation name. Initially, SUBDUE assumes each vertex as a valid substructure. Iteratively, substructures are expanded by one edge in all possible ways. At the end of each expansion, the substructures are evaluated and pruned based on their Minimum Description Length (Barron, Rissanen, & Yu, 1998) value. SUBDUE employs inexact graph match approach to find similar substructures; hence can handle noisy data. SUBDUE does not discover definitions of a specific target relation, but definitions of any concept that is hidden in the data.

SUBDUECL (Gonzalez, Holder, & Cook, 2002) is an extended version of SUBDUE to

discover concept descriptors for a specific target relation. Basically it employs the same data representation framework and core functions of SUBDUE. It differs from SUBDUE in learning and concept descriptor evaluation mechanisms. To evaluate the substructures, it employs set covering function instead of compression. Quality of a substructure is calculated with respect to the number of positive target instances that it does not explain and the negative target instances it explains. It employs constrained beam search and considers only as many substructures as the size of the beam at a time.

Concept Learning From Inference Patterns (CLIP) (Yoshida & Motoda, 1995) is a substructure-based concept discovery system, which utilizes colored, directed graph for data representation. In such a representation, colors encode attributes of a fact. CLIP assumes that each connected pair of vertices is a possible concept descriptor. CLIP employs beam search to find patterns. At each iteration, it finds similar pairs of vertices and replaces them with a new vertex and edges are rearranged, i.e. edges that originate or arrive at the vertices now point to the newly added vertex. CLIP decides the similarity of substructures based on their graph identities. Final concept descriptors are determined based on their compression degree of the original graph.

Graph-based Induction (GBI) (Yoshida, Motoda, & Indurkha, 1994) is another substructure-based concept discovery system. It employs basics of CLIP and discusses what types of problems can be represented as colored graphs. Matsuda, Motoda, Yoshida, & Washio (2002) further extends CLIP to handle self-loops and new pattern evaluation techniques.

In literature there are systems that combine properties of both approaches. One such study is by Mutlu & Karagoz (2013). It is similar to substructure-based approaches as it groups similar vertices and represents them as a single vertex. It is similar to pathfinding-based approaches, as it looks for paths in the compressed graph. It assumes that group of facts are similar if they are

related in the same way, i.e. have same value at the same argument place. To find paths, it follows iterative deepening search. The proposed algorithm restricts facts to be binary relations and can infer HOC type of concept descriptors. Igde, Kavurucu, & Mutlu (2015) extend this work to handle n-ary relations. Both studies evaluate the concept descriptors based on their coverage. Different than all other studies described above, these two systems utilize graphs to guide the search rather than mining.

Although graph-based for concept discovery systems are proposed to solve certain limitations of ILP-based systems, they pose certain problems. Both substructure-based and pathfinding-based approaches are computationally expensive. When compared to each other, SUBDUE distinguishes from other works as it can discover definitions of different relations at a time. CLIP is more tolerable to noise, as it considers substructures similar even they have some distortions. HOC system and MDPF build their search spaces based on a bottom clause, if the bottom clause is built for an improper target instance, the systems may fail to find concept descriptors of high quality. This problem also holds for Relational Pathfinding and RPBL. The hybrid approaches discussed above overcome this problem by considering all the target instances at the same time.

RUNNING EXAMPLE FOR GRAPH-BASED CONCEPT DISCOVERY

In this section we provide a working example for graph-based concept discovery. This is the demonstration of study by (Abay et al., 2015a). The study uses Neo4j database engine to represent relational data and induce concept descriptors. It uses No-SQL technology to retrieve data. The running example is based on the following kinship data set where *elti* is a Turkish kinship. Two women are called *elti* if they are married to two male brothers. In the proposed method, data is represented as directed, labeled graph. Vertices

Table 2. A sample kinship data set

Target Instances	Background Knowledge
elti(ayse, cemile). elti (ayten, ayse). elti (nalan, bedriye). elti (bedriye, nalan).	husband(altan, ayse). wife (cemile, osman). wife(ayse, altan). brother(altan, osman). husband(osman, cemile). wife(nalan, sadullah). husband(ismail, ayten).

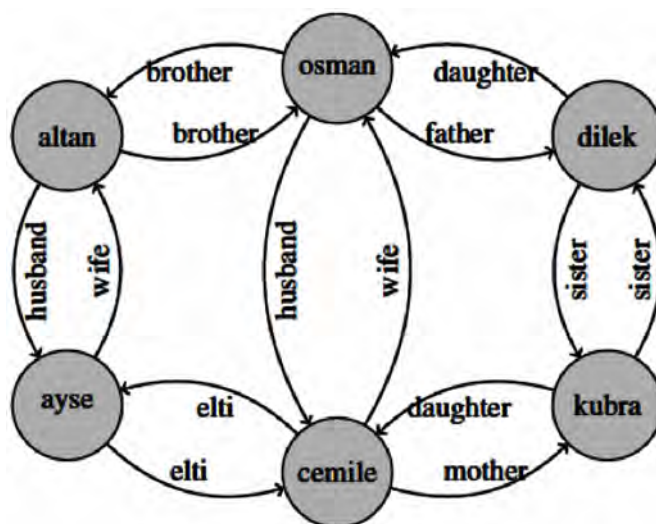
represent relation arguments and edges connect nodes that form a relation. Direction of edges implies the semantics of relations, i.e. osman is connected to cemile with an edge labeled husband, which indicates osman is husband of cemile. Graph representation of Table 2 is given in Figure 3.

To induce concept descriptors, the system makes use of Neo4j's path traversal methods and extracts all paths of length less than maximum rule length that originate from an argument of the target relation. Such paths include P1: elti(ayse,cemile); P2: elti(ayse,cemile), wife(ayse,altan), husband(altan,ayse), wife(ayse,altan); P3: elti(ayse,cemile), wife(ayse,altan), brother(osman,altan), husband(osman,cemile). Next, it eliminates all paths of length 1 as such paths are the target instances, i.e. P1, and paths that contain loops, i.e. P2, as revisited nodes do

not add extra information. Following the pruning step, the system replaces constants with variables. As an example, P3 becomes elti(A,B), wife(A,C), brother(C,D), husband(D,B). Finally support and confidence values of all such generalized paths are calculated and those that fall below user-defined thresholds are pruned. To evaluate support and confidence of concept descriptors they are transformed into SQL queries. Support of a concept descriptor is the number of target instances it deduces, and confidence is the number of instances, either from the target relation or background knowledge, it deduces. Generalized version of P3 maps to the following support and confidence queries:

- **Support:** *SELECT COUNT(DISTINCT CONCAT(elti.arg1, elti.arg2)) FROM elti, wife, brother, husband WHERE elti.arg1 = wife.arg1 AND wife.arg2 = brother.arg1 AND brother.arg2 = husband.arg1 and elti.arg2 = husband.arg2*
- **Confidence:** *SELECT COUNT(DISTINCT CONCAT(p1.arg, p2.arg)) FROM person as p1, person as p2, wife, brother, husband WHERE p1.arg = elti.arg1 AND p2.arg = husband.arg2 AND wife.arg2 = brother.arg1 AND brother.arg2 = husband.arg1*

Figure 3. Graph representation of sample kinship data set
Abay et al., 2015a.



Authors report in their study that graph construction takes 1 second for a data set with 224 instances and 2 seconds for a data set with 744 instances. When compared to ILP-based concept discovery system called CRIS, their study discovers concept descriptors with higher accuracy for some data sets. The running time of the system is also reported as feasible. These findings in coverage and running time comply with the findings of (Gonzales et al, 2000).

FUTURE RESEARCH DIRECTIONS

To improve running time of graph-based concept discovery systems there are several options. Parallelization poses several alternatives to improve running time of such systems. Graphs constructed to represent relational data may be very large. For example for PTE data set is represented by a graph that contains 22268 vertices and 22863 edges (Cai, Jonyer, & Paprzycki, 2005). Parallelization in graph-based concept discovery may be achieved in several manners. The first option is to distribute the data among several workers, each building its graph, and all searching for definitions of a specific target instances in their own graph. Another option is to distribute the entire data among the workers, each worker building the whole graph, and searching definitions of several target instances in parallel. Although there are several studies in literature that address such problems, there are still open problems such as reasonable load balancing, and communication cost.

Heuristics such as one proposed in (Mutlu, Karagoz, & Kavurucu, 2013) may be investigated to further improve efficiency of graph-based concept discovery systems where the order of the target instances effects the number of concept descriptors discovered and the running time of the system. Selecting the most promising target

instance to start the search may greatly improve the outcome of the concept discovery systems.

Introduction of fuzziness in such systems is also another research direction. As such systems prune possible concept descriptors based on metrics such as support, confidence and compression rate, pruning based on crisp thresholds may lead to loss of valuable concept descriptors.

Although some arguments of relations that take part in a concept descriptor are free variables, in the-state-of-the-art systems, relations take part in the concept descriptors in their full definitions. Systems may be modified to induce concept descriptors based on relation arguments rather than relations. By this way, concept definitions will not contain irrelevant information.

Functionalities of graph databases may be integrated into graph-based concept discovery systems. Such systems provide mechanism for exact and optional match of pattern and data. By such functionalities noisy data may better be analyzed.

CONCLUSION

In this chapter, we introduce concept discovery problem and compare several graph-based approaches by means of data representation, search method and concept evaluation technique. Although the problem has been researched for almost three decades now and its theoretical foundations are well established, it still has certain problems. To overcome the scalability issues, methods based on parallelization, heuristic development may further be investigated. To handle noisy data, fuzzy set oriented approaches may be investigated. In addition, new technologies such as graph databases deserve more attention for their possible applicability in concept discovery. Another possible research direction may include investigation of different models for concept descriptor representation.

REFERENCES

- Abay, N. C., Mutlu, A., & Karagoz, P. (2015a, July). A path-finding based method for concept discovery in graphs. In *Information, Intelligence, Systems and Applications (IISA), 2015 6th International Conference on* (pp. 1-6). IEEE. doi:10.1109/IISA.2015.7388092
- Abay, N. C., Mutlu, A., & Karagoz, P. (2015b, September). A Graph-Based Concept Discovery Method for n-Ary Relations. In *Big Data Analytics and Knowledge Discovery* (pp. 391-402). Springer International Publishing. doi:10.1007/978-3-319-22729-0_30
- Barron, A., Rissanen, J., & Yu, B. (1998). The minimum description length principle in coding and modeling. *Information Theory. IEEE Transactions on*, 44(6), 2743-2760.
- Blockeel, H., Witsenburg, T., & Kok, J. N. (2007, January). *Graphs, Hypergraphs, and Inductive Logic Programming*. MLG.
- Cai, M., Jonyer, I., & Paprzycki, M. (2005, September). Improving parallelism in structural data mining. In *Parallel Processing and Applied Mathematics* (pp. 455-462). Springer Berlin Heidelberg.
- Dolsak, B., & Muggleton, S. (1992). The application of Inductive Logic Programming to finite element mesh design. In *Inductive logic programming*.
- Džeroski, S., Jacobs, N., Molina, M., Moure, C., Muggleton, S., & Van Laer, W. (1998, July). Detecting traffic problems with ILP. In *International Conference on Inductive Logic Programming* (pp. 281-290). Springer Berlin Heidelberg. doi:10.1007/BFb0027332
- Gao, Z., Zhang, Z., & Huang, Z. (2009a, March). Learning relations by path finding and simultaneous covering. In *Computer Science and Information Engineering, 2009 WRI World Congress on* (Vol. 5, pp. 539-543). doi:10.1109/CSIE.2009.100
- Gao, Z., Zhang, Z., & Huang, Z. (2009b, April). Extensions to the relational paths based learning approach rpbl. In *2009 First Asian Conference on Intelligent Information and Database Systems* (pp. 214-219). doi:10.1109/ACIIDS.2009.40
- Gonzalez, J. A., Holder, L. B., & Cook, D. J. (2000). Graph based concept learning. *AAAI/IAAI*, 1072.
- Gonzalez, J. A., Holder, L. B., & Cook, D. J. (2002). *Graph-based relational concept learning*. Retrieved from: <http://ailab.wsu.edu/subdue/papers/GonzalezAAAI00.pdf>
- Holder, L. B., Cook, D. J., & Djoko, S. (1994, July). Substructure Discovery in the SUBDUE System. In *KDD workshop* (pp. 169-180).
- Iğde, M., Kavurucu, Y., & Mutlu, A. (2015). Graph Representation of Relational Database for Concept Discovery. *Procedia: Social and Behavioral Sciences*, 195, 1981-1989. doi:10.1016/j.sbspro.2015.06.212
- Kramer, S., Lavrač, N., & Flach, P. (2001). Propositionalization approaches to relational data mining. In *Relational data mining* (pp. 262-291). Springer Berlin Heidelberg. doi:10.1007/978-3-662-04599-2_11
- Matsuda, T., Motoda, H., Yoshida, T., & Washio, T. (2002, August). Knowledge discovery from structured data by beam-wise graph-based induction. In *PRICAI 2002: Trends in Artificial Intelligence* (pp. 255-264). Springer Berlin Heidelberg. doi:10.1007/3-540-45683-X_29
- Muggleton, S. (1991). Inductive logic programming. *New Generation Computing*, 8(4), 295-318.
- Muggleton, S. (1995). Inverse entailment and Progol. *New Generation Computing*, 13(3-4), 245-286.
- Muggleton, S., & De Raedt, L. (1994). Inductive logic programming: Theory and methods. *The Journal of Logic Programming*, 19, 629-679. doi:10.1016/0743-1066(94)90035-3

Mutlu, A., & Karagoz, P. (2013, August). A hybrid graph-based method for concept rule discovery. In *Data Warehousing and Knowledge Discovery* (pp. 327–338). Springer Berlin Heidelberg. doi:10.1007/978-3-642-40131-2_29

Mutlu, A., Karagoz, P., & Kavurucu, Y. (2013, September). A Counting-Based Heuristic for ILP-Based Concept Discovery Systems. In *Hybrid Artificial Intelligent Systems* (pp. 171–180). Springer Berlin Heidelberg. doi:10.1007/978-3-642-40846-5_18

Ong, I. M., de Castro Dutra, I., Page, D., & Costa, V. S. (2005, October). Mode directed path finding. In *Machine Learning: ECML 2005* (pp. 673–681). Springer Berlin Heidelberg. doi:10.1007/11564096_68

Pham, D. T., & Afify, A. A. (2006). SRI: A scalable rule induction algorithm. *Proceedings of the Institution of Mechanical Engineers. Part C, Journal of Mechanical Engineering Science*, 220(4), 537–552. doi:10.1243/09544062C18304

Quinlan, J. R. (1990). Learning logical definitions from relations. *Machine Learning*, 5(3), 239–266. doi:10.1007/BF00117105

Richards, L. R., & Mooney, J. M. (1992). Learning Relations by Pathfinding. *Proceedings. 10th National Conference on Artificial Intelligence*.

Santos, J. C., Tamaddoni-Nezhad, A., & Muggleton, S. (2009, October). An ILP system for learning head output connected predicates. In *Progress in Artificial Intelligence* (pp. 150–159). Springer Berlin Heidelberg. doi:10.1007/978-3-642-04686-5_13

Toprak, S. D., Senkul, P., Kavurucu, Y., & Toroslu, I. H. (2007, April). A new ILP-based concept discovery method for business intelligence. In *Data Engineering Workshop, 2007 IEEE 23rd International Conference on* (pp. 962–969). IEEE. doi:10.1109/ICDEW.2007.4401092

Turcotte, M., Muggleton, S. H., & Sternberg, M. J. (2001). Generating protein three-dimensional fold signatures using inductive logic programming. *Computers & Chemistry*, 26(1), 57–64. doi:10.1016/S0097-8485(01)00100-0 PMID:11765853

Yan, X., & Han, J. (2002). gspan: Graph-based substructure pattern mining. In *Data Mining, 2002. ICDM 2003. Proceedings. 2002 IEEE International Conference on* (pp. 721–724). IEEE.

Yoshida, K., Motoda, H., & Indurkha, N. (1994). Graph-based induction as a unified learning framework. *Applied Intelligence*, 4(3), 297–316. doi:10.1007/BF00872095

Yoshida, K. I., & Motoda, H. (1995). CLIP: Concept learning from inference patterns. *Artificial Intelligence*, 75(1), 63–92. doi:10.1016/0004-3702(94)00066-A

KEY TERMS AND DEFINITIONS

Breadth-First Search: Breadth-first search is a graph traversal method where firstly neighbor nodes are visited rather than sibling nodes.

Concept Discovery: Given a set of fact, concept discovery is the problem of learning definitions of a fact in terms of other facts.

Covering: Learning is an iterative process and at each rules possible with high accuracy and low coverage are induced. Removing learnt instances from the training set is called covering.

Depth-First Search: Depth first search is a graph traversal method where a branch is traversed as deep as possible without visiting any neighbor nodes.

Inductive Logic Programming: Given a set of observations, Inductive Logic Programming is concerned inducing general patterns that explain the observations.

Healthcare Data Analysis in the Internet of Things Era

George Tzanis

Aristotle University of Thessaloniki, Greece

INTRODUCTION

Health care advances are taking place rapidly all over the world. The newest medical discoveries and higher quality of medical practice have guided to serious improvements in human health and well-being. Life expectancy has been considerably increasing, leading to an aging population with different needs and desires. Moreover, world population has been rising exponentially, challenging the effectiveness of healthcare systems. New tests, procedures, and medicines have made medical care increasingly expensive, forcing the restraint of costs even while overall effectiveness improves (Vasanth & Sbert, 2012). Nowadays, prevention and early detection of disease signs as well as quality of life are becoming the points of focus more than ever before. Furthermore, personalized medicine is deemed particularly important and promises better disease treatment for each individual. Patients are becoming better-informed, want to know more about their health and their treatment, and demand to be involved in decision-making regarding their health.

All the aforementioned trends and changes have revealed the necessity to adopt a different model of health care. The modern healthcare delivery system should be more patient-centric, rather than career-centric, in order to deal effectively with the new challenges. Such a healthcare system should be organized in a layered structure (Pang, 2013) as shown in Figure 1. For example, the lower layer should comprise the single person and the higher one the hospital. Between these two terminal layers one can find the home and community layers. The lowest layer (person) is characterized by the

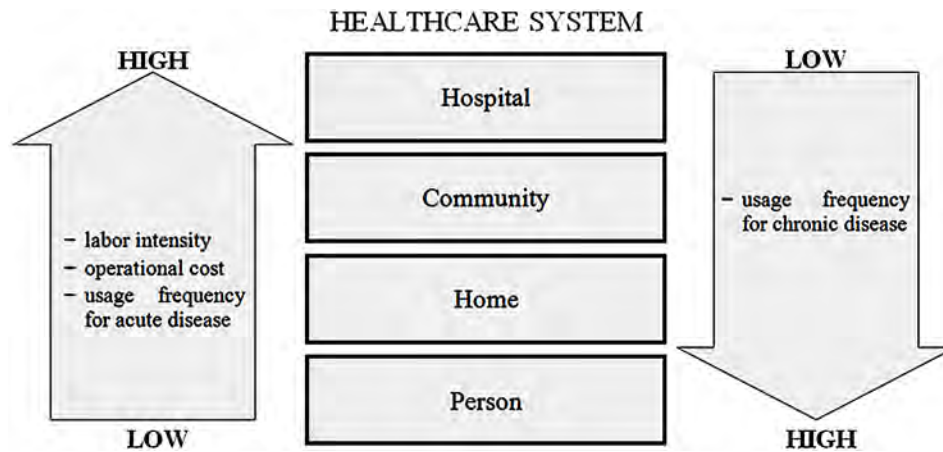
lowest labor intensity and operational cost, as well as the highest frequency of usage for chronic disease and lowest frequency of usage for acute disease. In contrast, the highest layer (hospital) has the highest labor intensity and operational cost, as well as the highest frequency of usage for acute disease and lowest for chronic disease.

The delivery model of health care is now standing at the starting point of its way to transform from the traditional hospital-centric to the modern home-centric/person-centric approach. The Internet of Things with its pervasive, personalized, and ubiquitous character can guide health care through this way. Many relevant concepts have been introduced to describe the future healthcare model powered by emerging information and communication technologies. Various terms, such as connected health, health Internet of Things (Health-IoT), pervasive healthcare (pHelath), ubiquitous healthcare (uHealth), mobile healthcare (mHealth), electrical healthcare (eHealth), telehealth, telemedicine, etc., have been utilized in order to express the plethora of these new concepts (Pang, 2013; Pawar et al., 2012; Rose et al., 2015). All these new concepts include, more or less, the fundamental aspects of the home-centric/person-centric healthcare model.

This article presents the concept of the Internet of Things as applied in health care discussing the benefits and opportunities as well as the posed obstacles and challenges. The article focuses on the basic strategies for analyzing healthcare data generated by the Internet of Things enabled devices and describes the current status, the major challenges, and the future trends of this revolutionary field.

DOI: 10.4018/978-1-5225-2255-3.ch172

Figure 1. The layered structure of a modern healthcare system



BACKGROUND

According to the definition provided by the telecommunication standardization sector of the International Telecommunication Union (ITU-T Recommendations, 2012), the Internet of Things (IoT) is “a global infrastructure for the information society, enabling advanced services by interconnecting (physical and virtual) things based on existing and evolving interoperable information and communication technologies”. With regard to the IoT, the term “thing” is defined as “an object of the physical world (physical thing) or the information world (virtual thing), which is capable of being identified and integrated into communication networks”. Exploiting various capabilities, concerning identification, data capture, processing, communication, etc., the IoT uses any “thing” to offer any service to any application domain. At the same time IoT has to abide by the requirements for security and privacy (ITU-T Recommendations, 2012).

Undoubtedly the IoT is the future of technology, with societal implications as well, that can make everyday life more efficient. The IoT can be recognized as the second digital revolution (The Government Office for Science, 2014), given that the first one was the World Wide Web, which connected people around the world establishing a global community. The IoT is built on the basis

of the World Wide Web extending its benefits and opportunities, but poses new big challenges. It is expected to greatly integrate leading technologies, such as technologies related to advanced machine-to-machine communication, autonomic networking, data mining and decision-making, security and privacy protection, as well as cloud computing, with technologies for advanced sensing and actuation. As shown in Figure 2, the IoT adds the dimension “any thing communication” to information and communication technologies (ICTs), which already provide “any time” and “any place” communication.

Among the plethora of application domains that can benefit by the utilization of the IoT, health care is one of the most important. Figure 3 presents the most popular IoT applications according to the study of IoT Analytics for Q3/2016. The popularity ranking of the IoT applications was based on 640 actual enterprise IoT projects not including any consumer IoT projects such as wearable devices, smart home or hobby projects.

Vilamovska et al. (2009) have classified IoT applications, in particular RFID applications, in healthcare across two axes:

- **IoT Enabling Functions:** These are four key and mutually exclusive functions concerning assets, staff and patients.
 - Tracking.

- Identification and authentication.
- Automatic data collection and transfer.
- Sensing.
- **Healthcare Applications:** These are the four most promising areas of applications
 - Patient safety and quality of care.
 - Pharmaceutical application (not including supply chain and counterfeit drug issues).

Figure 2. The new dimension (any thing communication) introduced with the Internet of Things
Source: ITU-T Recommendations, 2012.

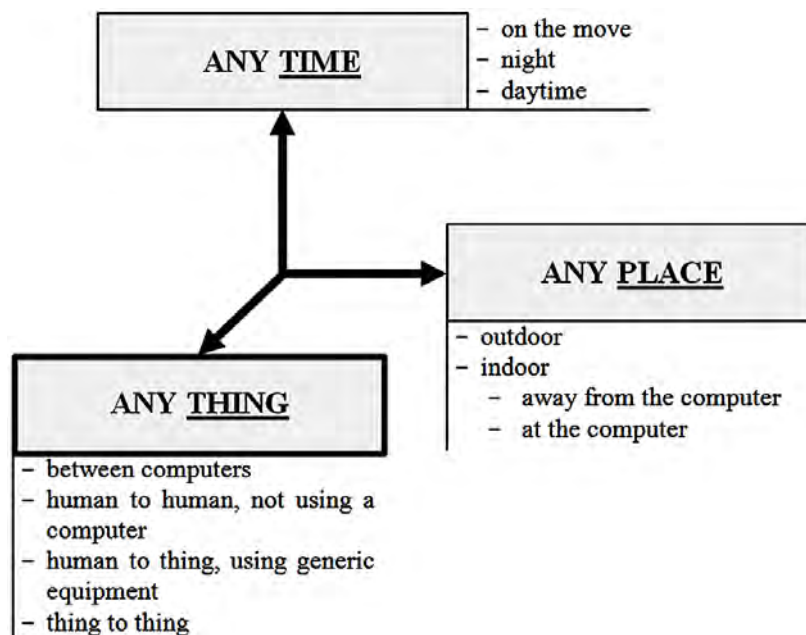
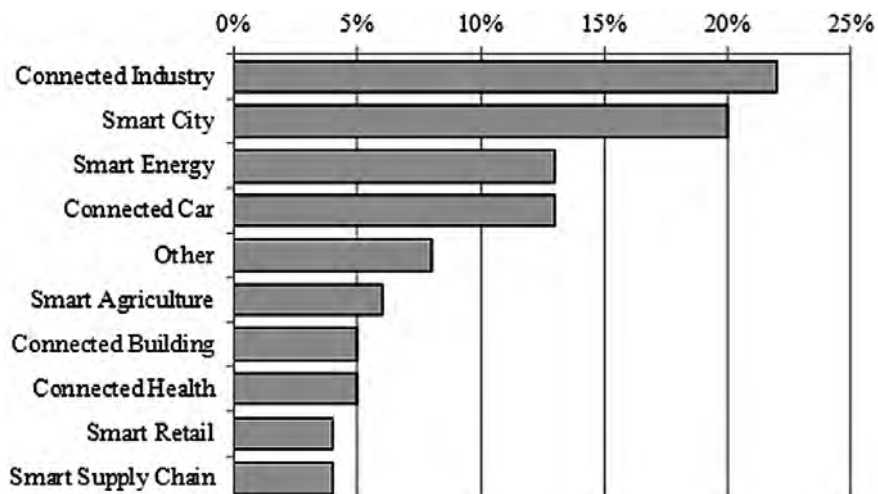


Figure 3. The most popular Internet of Things applications; valid for Q3/2016
Source: IoT Analytics, 2016.



- Management of devices, supplies and biological material.
- Patient and healthcare provider support and management.

Examples of healthcare applications for each enabling function are presented as a four-by-four matrix in Table 1.

HEALTHCARE DATA ANALYSIS

Healthcare data analysis and analytics can support the basic goals of a healthcare system. The objectives that can be addressed through healthcare data analysis have been classified into three major axes by Cortada, et al. (2012) as presented in the following lines:

- **Improvement of Clinical Effectiveness and Patient Satisfaction:**
 - Improvement of clinical quality of care.

- Improvement of patient safety and reduction of medical errors.
- Improvement of wellness, prevention and disease management.
- Better understanding of physician profiles and clinical performance.
- Improvement of patient satisfaction, acquisition and retention.

- **Improvement of Operational Effectiveness:**

- Reduction of costs and increase in efficiency.
- Optimization of network management.
- Improvement of pay for performance and accountability.
- Increase in operating speed and adaptability.

- **Improvement of Financial and Administrative Performance:**

- Increase in revenue and return on investment (ROI).
- Improvement of utilization.
- Optimization of supply chain and human capital management.

Table 1. Preliminary list of the most promising areas for RFID deployment in healthcare

	Patient Safety/Quality of Care	Pharmaceutical Application	Management of Devices, Supplies and Biological Material	Patient and Healthcare Personnel Support/Management
Tracking	Tracking vulnerable patients (dementia) Tracking patient location to account for patient time during treatment	Easy recall of products	Prevention of leftins during surgical procedures Equipment tracking (ensure hygiene, compliance, regular maintenance, fast spotting when needed)	Personnel tracking to improve workflow and reduce patient waiting times in emergency room
Identification and authentication	Mother–newborn accurate matching	Auto ID-enabled medication administration system	Maintenance of real-time clinical information associated with patient within a hospital	Patient identification to reduce incidents harmful to patients (wrong drug, dose, time, procedure)
Automatic data collection and transfer	Use of portable devices for care coordination.	Tracking pharmaceutical inventories	Inventory management for better use of staff time and faster care delivery	IoT-supported automated care, pathways, procedures audit, and management
Sensing	Patient vital signs to trigger alerts for medical personnel, and remote monitoring at patient's home	Patient compliance with prescribed medication treatment (in and out-patient)	Blood bags equipped with temperature sensors in hospital to ensure cold chain and efficacy	Personnel and asset tracking to ensure infection control (nosocomial infections)

Source: Vilamovska et al., 2009.

- Improvement of risk management and regulatory compliance.
- Reduction of fraud and abuse.

Data Mining

Data mining emerged in order to cope with the challenges that traditional data analysis techniques were facing up when dealing with large amounts of data. Moreover, these data often have a lot of peculiarities (e.g. missing values, noise, etc.). More specifically, data mining is the main step in the process of Knowledge Discovery in Databases or KDD Process (Figure 4). Knowledge Discovery in Databases is defined as the non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data (Fayyad et al., 1996). However, the term “data mining” is very often used to describe the whole KDD Process. Although the core of the process is the data mining step, where a data mining algorithm is applied in order to extract the patterns from data, the pre-processing and post-processing phases are very important too and contribute sensibly to the quality of the extracted knowledge. The pre-processing phase usually includes the selection of an appropriate portion of data, the cleansing of the selected data, as well as the transformation of data in more appropriate representations. The

post-processing phase deals with the management of the produced patterns and models and focuses on the evaluation and interpretation of data mining results.

Data mining, in practice, has the following two “high-level” primary goals (Fayyad et al., 1996):

- **Prediction:** Involves the use of some fields (variables) in a database to predict unknown or future values of other variables of interest.
- **Description:** Focuses on finding human-interpretable patterns describing the data.

Prediction and description are not equivalently important for every data mining application. In the context of Knowledge Discovery in Databases, description tends to be more important than prediction. In contrast, machine learning and pattern recognition applications, usually favor prediction as the primary goal. Prediction and description are achieved by using various data mining tasks, the most common of which are presented in Table 2 (Dunham, 2002). Depending on the nature of the data and the desired knowledge there is a large variety of algorithms for each task. A thorough review of the application of data mining in health care can be found in (Koh & Tan, 2005; Yoo et al., 2012).

Figure 4. The KDD process
Source: Revised from Tzani (2012).

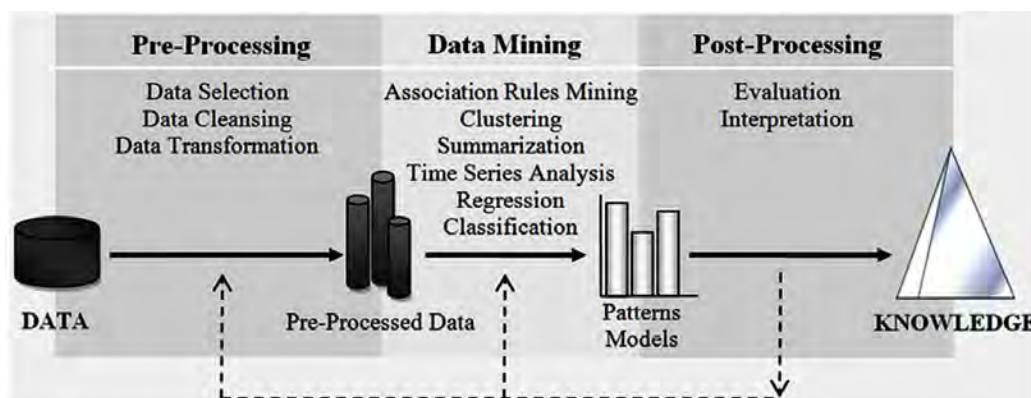


Table 2. Most common data mining tasks

Tasks for Description	Tasks for Prediction
• Association Rules Mining: Extracts rules expressing relationships among variables of data. Similar task is <i>sequence analysis</i> where relationships are based on time. • Clustering: Groups data, so that data in the same group (cluster) are more similar to each other than to those in other groups. • Summarization: Maps data into subsets with associated simple descriptions. Also called <i>characterization</i> or <i>generalization</i>.	• Classification: Maps data into predefined classes, using a model that is constructed on data with known class mappings. • Regression: Maps data into a real valued prediction variable, using a model constructed on data with known mappings. • Time Series Analysis: Examines the sequence of events where an event is determined by one or more of the preceding events.

Big Data Challenges

Although data mining provides the means to deal efficiently with large amounts of, usually, peculiar data, healthcare data analysis in the Internet of Things era is far from an uncomplicated task. The term “big data” is very popular nowadays and is used to describe the explosively extended collections of data, which carry significantly more peculiarities. The popularity of the “big data” concept is signified by the fact that a lot of challenges regarding the management and analysis of big data have been raising, whereas the benefits and opportunities ensuing from the exploitation of these data seem to be manifold and very attractive. The conventional data analysis tools, including machine learning and data mining, are not efficient anymore and have to be scaled in order to deal with the obstacles posed by the big data paradigm.

The basic peculiarities that figure the challenging character of big data are described by the “5-V’s”. Laney (2001) introduced the term “3-V’s” in big data community, namely volume, variety, and velocity. However, since then there have been proposed even more “V’s”, but two of them, variability and value, are the most popular (Fan & Bifet, 2012). These “5-V’s” are described in the following lines (Fan & Bifet, 2012; Fayyad, 2012):

- **Volume:** The volume of the data has been growing rapidly, more than ever before, providing challenges in loading, processing, and transferring.

- **Velocity:** The velocity of data streams that arrive continuously and need to be analyzed pose challenging real-time constraints.
- **Variety:** There are many different types of data (e.g. text, sensor data, images, audio, video, graph), various degrees of structure in data (structured, semi-structured, unstructured), and often different types and structures of data are mixed.
- **Variability:** The structure of the data as well as the way users are willing to interpret these data usually changes with time providing extra data and knowledge management challenges.
- **Value:** The value refers to the quality of the extracted knowledge that may give the ability of making better decisions and answering more questions.

SOLUTIONS AND RECOMMENDATIONS

Health care, like many other areas, follows the big data paradigm presented above. Healthcare data are becoming available at rapid rates and as the exploitation of the IoT infrastructure grows, data management and data analysis becomes even more challenging. Subsequently, individual hardware units are not any more adequate to satisfy the computational requirements for managing and analyzing these data. New algorithmic approaches with increased scalability and efficiency and infrastructures that make use of the computational power of multiple hardware units, such as distributed and cloud computing, are deemed necessary.

The prevalent trend in dealing with the big data storm is the concept of cloud computing. According to NIST definition (Mell & Grance 2011), “cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction”.

Cloud services can be grouped in three main groups (Mell & Grance 2011):

- **Software as a Service (SaaS):** The capability provided to the consumer is to use the provider’s applications running on a cloud infrastructure. The applications are accessible from various client devices through either a thin client interface, such as a web browser or a program interface.
- **Platform as a Service (PaaS):** The capability provided to the consumer is to deploy onto the cloud infrastructure consumer-created or acquired applications created using programming languages, libraries, services, and tools supported by the provider.
- **Infrastructure as a Service (IaaS):** The capability provided to the consumer is to provision processing, storage, networks, and other fundamental computing resources where the consumer is able to deploy and run arbitrary software, which can include operating systems and applications.

There are various cloud deployment models that have their own benefits and drawbacks. The selection of a proper model depends on the specific application domain. These deployment models are (Mell & Grance 2011; Botta et. al, 2016):

- **Private Cloud:** Provisioned for exclusive use by a single organization and may be owned, managed, and operated by the organization, a third party, or a combination of them.

- **Community Cloud:** Provisioned for exclusive use by a specific community of consumers from organizations that have shared concerns and may be owned, managed, and operated by some of the organizations in the community, a third party, or a combination of them.
- **Public Cloud:** Provisioned for open use by the general public and may be owned, managed, and operated by a business, academic, government organization, or a combination of them.
- **Hybrid Cloud:** Composed of two or more distinct cloud infrastructures (private, community, or public) that remain unique entities, but are bound together by standardized or proprietary technology that enables data and application portability.
- **Virtual Private Cloud:** Exploits virtual private network (VPN) technologies for allowing organizations to setup required network settings (e.g. topology, security, etc.).

Cloud computing and the IoT are two novel and promising concepts of technology that have been evolving independently. Although they are different, they have complementary characteristics as shown in Table 3. Due to this complementarity many researchers have been thinking about the possible benefits of the fusion of these two technologies proposing approaches that integrate

Table 3. Complementary characteristics of cloud computing and IoT

	Internet of Things	Cloud Computing
Displacement	Pervasive	Centralized
Reachability	Limited	Ubiquitous
Components	Real world “things”	Virtual resources
Computational capabilities	Limited	Virtually unlimited
Role of the Internet	Point of convergence	Means for delivering services
Big data	Source	Means to manage

Source: Botta et. al, 2016.

cloud computing and the IoT (Cloud-IoT) in various application domains (Alhakbaniet al., 2014; Botta et. al, 2016; Diaz et al., 2016; Gomes et al., 2014). The benefits are mutual for these two technologies. The IoT can exploit the virtually unlimited capabilities and resources of cloud computing to compensate its technological constraints and cloud computing can extend its scope to deal with real world “things” in a more distributed and dynamic manner, delivering novel services in a large number of real life cases.

The adoption of the innovative infrastructures of Cloud computing and IoT in health care has lead to a paradigm shift in healthcare data management and analysis. The traditional standalone, local, offline data analysis model has been replaced by a more dynamic, global and regularly online data management and analysis paradigm. Cloud-IoT is able to simplify healthcare processes, enhance the quality of medical services by enabling cooperation among interconnected entities. For example Ambient Assisted Living (AAL) aims at easing the daily lives of people with disabilities and chronic diseases (Botta et. al, 2016). Cloud-IoT may supply many innovative services, such as patients’ vital data collection through sensor networks, data transfer to medical centers’ Cloud infrastructures, management, sharing and ubiquitous access of Electronic Healthcare Records (EHR), etc (Botta et. al, 2016, Kuo, 2011). Moreover, Cloud allows better confrontation with common challenges of healthcare domain such as security, privacy, and reliability, by enhancing medical data security as well as service availability and redundancy. Cloud enables the execution of secure healthcare services, overcoming the issue of limited computational capacity and power. It provides the means for a flexible and scalable storage and processing in order to efficiently perform online and offline analyses of data.

FUTURE RESEARCH DIRECTIONS

The introduction of the Cloud-IoT paradigm in the healthcare domain can offer numerous benefits and opportunities that will considerably improve

the quality of healthcare services in the era of big data. This is a recently emerged research area and is expected to receive the focus of the future research efforts. Although the aims for storage and management of large amounts of data in the IoT infrastructure can be provided by the integration of the IoT and cloud computing, there are still open issues concerning verification, normalization, filtering and analysis of the IoT data (Diaz et al., 2016). The lack of open standards, the large diversity of technologies involved in the IoT, as well as the huge amounts of data require novel techniques to improve and optimize such integration problems.

From a broader perspective, the basic challenges that concern Cloud-IoT healthcare applications and should be addressed in the future are presented below (Atzori et al., 2010; Botta et. al, 2016; Rose et al., 2015):

- **Privacy and Security:** These challenges may concern the lack of trust in the cloud service providers, the lack of knowledge about the physical location of data, the exposure of the distributed system to possible attacks and vulnerabilities, etc.
- **Legal and Social Issues:** The legal challenges are important depending on the specific application and different international laws. Moreover, social issues become extremely important, since humans are becoming the source of sensor data.
- **Large Scale:** This relates to applications which include a very large number of “things”, distributed across very wide areas.
- **Reliability:** This arises with mission-critical applications.
- **Performance:** This concerns applications having specific performance and Quality of Service requirements at multiple levels (communication, computation, storage).
- **Heterogeneity:** This refers to heterogeneity of devices, platforms, operating systems, services, etc.

CONCLUSION

Undoubtedly the IoT is the future of technology, which can provide manifold benefits to health care. However, the posed challenges are also great. Concerning the analysis of healthcare data, various tools have been introduced to deal efficiently with the large volumes as well as the various peculiarities of data (e.g. missing values, noise, etc.). The most popular representative of these modern tools is data mining, or the KDD process, strictly speaking. Although the KDD process has provided a lot of solutions, these techniques have to be scaled in order to deal with the new challenges posed by the big data paradigm. Cloud computing is the modern infrastructure that can provide the means to efficiently manage big data. Both, cloud computing and the IoT are very promising concepts of technology and their complementary characteristics assure that their integration, Cloud-IoT, will provide a great potential of applications. The introduction of the Cloud-IoT paradigm in the healthcare domain can offer manifold benefits and opportunities that will considerably improve the quality of health care.

REFERENCES

- Alhakbani, N., Hassan, M. M., Hossain, M. A., & Alnuem, M. (2014). A framework of adaptive interaction support in Cloud-based Internet of Things (IoT) environment, In *Proceedings of the 7th International Conference* (vol. 8729, pp. 136–146). Calabria, Italy, Springer. doi:10.1007/978-3-319-11692-1_12
- Atzori, L., Iera, A., & Morabito, G. (2010). The Internet of Things: A survey. *Computer Networks*, 54(15), 2787–2805. doi:10.1016/j.comnet.2010.05.010
- Botta, A., De Donato, W., Persico, V., & Pescapé, A. (2016). Integration of Cloud computing and Internet of Things: A survey. *Future Generation Computer Systems*, 56, 684–700. doi:10.1016/j.future.2015.09.021
- Cortada, J.W., Gordon, D., & Lenihan, B. (2012). *The value of analytics in healthcare: from insights to outcomes*. IBM Global Business Services, Life Sciences and Healthcare, Executive Report.
- Diaz, M., Martin, C., & Rubio, B. (2016). State-of-the-art, challenges, and open issues in the integration of Internet of things and cloud computing. *Journal of Network and Computer Applications*, 67, 99–117. doi:10.1016/j.jnca.2016.01.010
- Dunham, M. H. (2002). *Data mining: Introductory and advanced topics*. Upper Saddle River, NJ: Prentice Hall.
- Fan, W., & Bifet, A. (2012). Mining Big Data: Current Status, and Forecast to the Future. *SIGKDD Explorations*, 14(2), 1–5. doi:10.1145/2481244.2481246
- Fayyad, U. M. (2012). Big Data Analytics: Applications and Opportunities in On-line Predictive Modeling. Keynote Talk. BigMine: BigData Mining Workshop KDD-2012, Beijing, China.
- Fayyad, U. M., Piatetsky-Shapiro, G., & Smyth, P. (1996). From Data Mining to Knowledge Discovery: An Overview. In *Advances in Knowledge Discovery and Data Mining*. AAAI Press/The MIT Press.
- Gomes, M. M., Da Rosa Righi, R., & Da Costa, C. A. (2014). Future directions for providing better IoT infrastructure. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication* (pp. 51–54). doi:10.1145/2638728.2638752
- IoT Analytics. (2016). *The top 10 IoT application areas – based on real IoT projects*. Retrieved September 29, 2016, from <https://iot-analytics.com/top-10-iot-project-application-areas-q3-2016>
- ITU-T Recommendations. (2012). *Overview of the Internet of things*. Recommendation ITU-T Y.2060. Retrieved September 29, 2016, from <http://handle.itu.int/11.1002/1000/11559>

Koh, H. C., & Tan, G. (2005). Data mining applications in healthcare. *Journal of Healthcare Information Management*, 19(2), 64–72. PMID:15869215

Kuo, A. M.-H. (2011). Opportunities and challenges of Cloud computing to improve health care services. *Journal of Medical Internet Research*, 13(3), e67. doi:10.2196/jmir.1867 PMID:21937354

Laney, D. (2001). *3-D Data Management: Controlling Data Volume, Velocity and Variety*. META Group Research Note.

Mell, P. & Grance, T. (2011). *The NIST Definition of Cloud Computing*. Recommendations of the National Institute of Standards and Technology. NIST, U.S. Department of Commerce. Retrieved September 29, 2016, from 10.6028/NIST.SP.800-145

Mohammed, E. A., Far, B. H., & Naugler, C. (2014). Applications of the MapReduce programming framework to clinical big data analysis: Current landscape and future trends. *BioData Mining*, 7(1), 22. doi:10.1186/1756-0381-7-22 PMID:25383096

Pang, Z. (2013). *Technologies and Architectures of the Internet-of-Things (IoT) for Health and Well-being* (Unpublished doctoral dissertation). Royal Institute of Technology, Stockholm, Sweden.

Pawar, P., Jones, V., Beijnum, B. F. V., & Hermens, H. (2012). A framework for the comparison of mobile patient monitoring systems. *Journal of Biomedical Informatics*, 45(3), 544–556. doi:10.1016/j.jbi.2012.02.007 PMID:22406009

Rose, K., Eldridge, S., & Chapin, L. (2015). *The Internet of Things: An Overview - Understanding the Issues and Challenges of a More Connected World*. Internet Society White Paper.

The Government Office for Science. (2014). *The Internet of Things: making the most of the Second Digital Revolution*. A report by the UK Government Chief Scientific Adviser. Retrieved September 29, 2016, from <https://www.gov.uk/government/publications/internet-of-things-blackett-review>

Tzanis, G. (2014). Biological and Medical Big Data Mining. *International Journal of Knowledge Discovery in Bioinformatics*, 4(1), 42–56. doi:10.4018/ijkdb.2014010104

Vasanth, K., & Sbert, J. (2012). *Creating solutions for health through technology innovation*. Texas Instruments White Paper. Retrieved September 29, 2016, from <http://www.ti.com/lit/wp/sszy006/sszy006.pdf>

Vilamovska, A.-M., Hatzianreou, E., Schindler, H. R., Van Oranje-Nassau, C., De Vries, H., & Krapels, J. (2009). *Study on the requirements and options for RFID application in healthcare - Identifying areas for Radio Frequency Identification deployment in healthcare delivery: A review of relevant literature*. RAND Europe.

Yoo, I., Alafaireet, P., Marinov, M., Pena-Hernandez, K., Gopidi, R., Chang, J. F., & Hua, L. (2012). Data mining in healthcare and biomedicine: A survey of the literature. *Journal of Medical Systems*, 36(4), 2431–2448. doi:10.1007/s10916-011-9710-5 PMID:21537851

ADDITIONAL READING

Bui, N., & Zorzi, M. (2011). Health care applications: a solution based on the internet of things. In *Proceedings of the 4th International Symposium on Applied Sciences in Biomedical and Communication Technologies (ISABEL '11)*. ACM, New York, NY, USA, Article 131. doi:10.1145/2093698.2093829

Chen, F., Deng, P., Wan, J., Zhang, D., Vasilakos, A. V., & Rong, X. (2015). Data Mining for the Internet of Things: Literature Review and Challenges. *International Journal of Distributed Sensor Networks*, 11(8), 431047.

Cloud Standards Customer Council. (2012). *Impact of Cloud Computing on Healthcare*, White Paper. Retrieved September 29, 2016, from <http://www.cloud-council.org/deliverables/CSCC-Impact-of-Cloud-Computing-on-Healthcare.pdf>

Hou, J.-L. and Yeh, K.-H. (2016). Novel authentication schemes for IoT based healthcare systems. *International Journal of Distributed Sensor Networks*, 11(11).

IoT Analytics. (2015). *IoT platforms - The central backbone for the IoT*, White Paper. Retrieved September 29, 2016, from <https://iot-analytics.com/product/iot-platforms-white-paper>

Khanna, A., & Misra, P. (2014). *The Internet of Things for Medical Devices - Prospects, Challenges and the Way Forward*. TATA Consultancy Services White Paper. Retrieved September 29, 2016, from http://www.tcs.com/resources/white_papers/Pages/Internet-of-Things-Medical-Devices.aspx

LeSueur, D. (2014). 5 Reasons Healthcare Data Is Unique and Difficult to Measure. Health Catalyst White Paper. Retrieved September 29, 2016, from <https://www.healthcatalyst.com/5-reasons-healthcare-data-is-difficult-to-measure>

Lueth, K. L. (2015). *IoT basics: Getting started with the Internet of Things*, IoT Analytics White Paper. Retrieved September 29, 2016, from <https://iot-analytics.com/product/whitepaper-iot-basics-getting-started-with-the-internet-of-things>

Riazul Islam, S. M., Kwak, D., Kabir, H., Hosain, M., & Kwak, K.-S. (2015). The Internet of Things for Health Care: A Comprehensive Survey. *IEEE Access*, 3, 678–708. doi:10.1109/ACCESS.2015.2437951

KEY TERMS AND DEFINITIONS

Ambient Intelligence: The discipline aims to bring intelligence to everyday environments making them sensitive and responsive to the presence of people in a non-intrusive way and with minimal explicit interaction.

Artificial Intelligence: The area of computer science that studies the construction of computer programs that can imitate human cognitive abilities and present behaviors indistinguishable from human behavior.

Internet of Medical Things (IoMT): A global infrastructure consisting of the collection of medical devices and applications that are interconnected through information and communication technologies.

Machine Learning: The field of study that deals with the construction of algorithms that can learn from data and make predictions on data.

Radio Frequency Identification (RFID): The technology that incorporates the use of electromagnetic fields to uniquely identify and track a tagged object.

Ubiquitous Computing: The discipline aims to bring computing anywhere and anytime in any object. It is also called pervasive computing.

Wi-Fi: The technology that allows electronic devices to connect to a wireless local area network (WLAN) and Internet using radio waves.

Wireless Sensor Network (WSN): The network of spatially distributed autonomous sensors which monitor physical or environmental conditions and cooperatively pass their data through the network to a main location.

N-Tuple Algebra as a Unifying System to Process Data and Knowledge

D

Boris Alexandrovich Kulik

Institute of Problems in Mechanical Engineering RAS, Russia

Alexander Yakovlevich Fridman

Institute for Informatics and Mathematical Modelling, Kola Science Centre of RAS, Russia

BACKGROUND

Theory of relations is mostly used in databases modeled by means of relational algebra (Codd, 1971). However, many diverse and (at first sight) different mathematical objects, namely graphs, semantic networks, expert rules, predicates, logical formulas, etc., can be expressed as relations. Relational algebra does not provide sufficient techniques to deal with such models and structures.

Modern intelligence systems comprise two types of dissimilar objects, namely databases (DBs) and knowledge bases (KBs). Their structures are quite distinct since they exploit fundamentally different theoretical approaches. Nowadays, data structures (numbers, vectors, graphs, networks, texts, etc.) correspond to algebraic methods (Wirth, 1976). As for KBs, their basic models (predicates, frames, semantic networks, rules and so on) are designed on the basis of declarative approach (Russel & Norvig, 2003). This incompatibility results in significantly different programming systems and structures for DBs and KBs, and consequently, in big difficulties to integrate a DB and a KB within one software system.

Conventional attempts to do so relate to the declarative approach mostly. The model of deductive databases (Ullman, 1989; Ceri, Gotlob, & Tanca, 1990) gives an example of such attempts. It is grounded on a modification of Prolog language designed within the declarative approach.

Conversely, some shortcomings inhere in this approach. One of the major problems is in the

necessity to reduce many tasks of logical analysis to satisfiability checks for a certain logical formula, this check being able to return only two possible answers (“yes” or “no”). Such a reduction is not simple. Moreover, it is unrealizable in cases when one needs to not only receive a binary answer but also to estimate the value of some parameters in the formal system or to assess the structure and/or number of objects that satisfy the given conditions.

To model and analyze modified reasoning (that is reasoning with hypotheses and abductive conclusions), many authors propose usage of non-classical logics (Russel & Norvig, 2003). However, such logics often have no interpretation or their interpretations do not correspond to semantics of the objects to be modeled. In the authors’ opinion, there are some reasons to suppose that problems in modeling modified reasoning within classical logic result from features of the declarative approach rather than from drawbacks of classical logic.

Algebraic approach to logical modeling can be originated from Aristotle’s syllogistics and Boole’s logical matrices, in more recent times only Arkady Zakrevskij (for instance, see (Zakrevskij & Gavrilov, 1969)) investigated set-theoretical approach to logic and proposed a programming language LYaPAS (Logical Language for Representation of Synthesis Algorithms) oriented toward programming of synthesis algorithms for finite-state and discrete devices. However, this language does not provide any means for logical inference.

To overcome the mentioned and other problems, the mathematical model of relational algebra was generalized by the authors to a more universal system for processing n -ary relations called n -tuple algebra (Kulik, 1995, 2007; Kulik, Fridman, & Zuenko, 2013, 2015).

N-TUPLE ALGEBRA: BASICS AND FEATURES

N -tuple algebra is a mathematical system to deal with arbitrary n -ary relations. In NTA, such relations can be expressed as four types of structures called *NTA objects*. Every NTA object is immersed into a certain space of *attributes*. Names of NTA objects contain an identifier followed by a sequence of attributes names in square brackets; these attributes determine the *relation diagram* in which the NTA object is defined. For example, $R[XYZ]$ denotes an NTA object defined within the space of attributes

$X \times Y \times Z$ that is $R[XYZ] \subseteq X \times Y \times Z$.

From the mathematical point of view, NTA is an *algebraic system* $A = \langle A, O, R \rangle$, where *carrier* A is an arbitrary totality of relations expressed as NTA objects; O is a set of *operations* comprising *algebra of sets' operations* (union, intersection, complement) and *operations on attributes* (see below); R is a set of verified *relations* between NTA objects, namely equality and inclusion. This algebraic system is proved to be complete (i.e. all operations and checks of relations are feasible for any totalities of NTA objects).

NTA objects defined on the same relation diagram are called *homotypic* ones. NTA objects contain n -tuples of sets and provide a condensed representation of n -ary relations. In NTA, *elementary n -tuples* correspond to n -tuples of elements in ordinary relations. For instance, the record $T[XYZ] = (a, b, c)$ means that T is an elementary n -tuple where $a \in X, b \in Y, c \in Z$.

NTA objects look like matrices having subsets of corresponding attributes (*components*) rather than elements as their cells. The totality of NTA components includes two *dummy components*. They are: the *complete component* $*$ that equals to the whole domain (i.e. the range of values) of the corresponding attribute in the relation diagram and the *empty component* $\emptyset$ that equals to the empty set and models the complement of the component $*$.

NTA admits 4 types of NTA objects, namely C - n -tuples and C -systems, D - n -tuples and D -systems.

A C - n -tuple is an n -tuple of components defined in a certain relation diagram, denoted within square brackets and interpreted as the set of elementary n -tuple contained in the Cartesian product of its components. For example, the right part of the formula $R[XYZ] = [A * C]$ expresses a C - n -tuple that can be transformed into an ordinary n -ary relation by calculating the Cartesian product $A \times * \times C$ where $A \subseteq X, * = Y$ is a complete component placed to the position of the attribute Y , and $C \subseteq Z$.

A C -system is a set of homotypic C - n -tuples that are denoted as a matrix in square brackets. As an example, the C -system

$$Q[XYZ] = \begin{bmatrix} A_1 & B_1 & * \\ * & B_2 & C_2 \end{bmatrix}$$

can be reduced to a set of elementary n -tuples calculated by formula

$$Q[XYZ] = (A_1 \times B_1 \times Z) \cup (X \times B_2 \times C_2).$$

A D - n -tuple expresses a structure that can be interpreted, in particular, as a complement of a certain C - n -tuple. Suppose, the C - n -tuple

$$R[XYZ] = [A \ B \ C]$$

is given. Its complement equals to the difference of C - n -tuples

$$[X \ Y \ Z] \setminus [A \ B \ C] = (X \times Y \times Z) \setminus (A \times B \times C).$$

The formula of difference for two Cartesian products is well known, for the given example it equals:

$$\begin{aligned} (X \times Y \times Z) \setminus (A \times B \times C) &= \\ ((X \setminus A) \times Y \times Z) &= \\ \cup (X \times (Y \setminus B) \times Z) &= \\ \cup (X \times Y \times (Z \setminus C)) &= \end{aligned}$$

In NTA terms, the right part of this formula can be denoted as the C -system

$$\overline{R} [XYZ] = \begin{bmatrix} \overline{A} & * & * \\ * & \overline{B} & * \\ * & * & \overline{C} \end{bmatrix}$$

where $\overline{A}$, $\overline{B}$, $\overline{C}$ mean complements of the components A , B , C to the domains of corresponding attributes.

The square C -systems like $\overline{R} [XYZ]$ where all nondiagonal cells contain complete components are called *diagonal C -systems*. Without any loss of information, they can be written as an n -tuple of the diagonal components. To distinguish them from C - n -tuples in NTA, we use reversed square brackets instead of regular ones. So, for the given example

$$\overline{R} [XYZ] =] \overline{A} \ \overline{B} \ \overline{C} [.$$

A D -system is a structure that consists of a set of homotypic D - n -tuples and equals to their intersection. The list of D - n -tuples in a D -system is limited with reversed square brackets.

NTA laws are based on the laws of algebra of sets and properties of the Cartesian product. Let us introduce main NTA relationships.

Suppose two homotypic C - n -tuples

$$P = [P_1 P_2 \dots P_n] \text{ and } Q = [Q_1 Q_2 \dots Q_n]$$

be given. Then:

1. $P \subseteq Q$, if and only if $P_i \subseteq Q_i$ for all $i = 1, 2, \dots, n$;
2. $P \cap Q = [P_1 \cap Q_1 \ P_2 \cap Q_2 \ \dots \ P_n \cap Q_n]$. Moreover, if for any i $P_i \cap Q_i = \emptyset$, then $P \cap Q = \emptyset$;
3. $P \cup Q = \begin{bmatrix} P_1 & P_2 & \dots & P_n \\ Q_1 & Q_2 & \dots & Q_n \end{bmatrix} \subseteq [P_1 \cup Q_1 \ P_2 \cup Q_2 \ \dots \ P_n \cup Q_n]$;
4. Equality $P \cup Q = [P_1 \cup Q_1 \ P_2 \cup Q_2 \ \dots \ P_n \cup Q_n]$ is possible in two cases only:
 - a. $P \subseteq Q$ or $Q \subseteq P$;
 - b. $P_i = Q_i$ for all corresponding pairs of components except one pair;
5. $\overline{P} =] \overline{P}_1 \ \overline{P}_2 \ \dots \ \overline{P}_n [.$

Numerous examples of representing different data and knowledge structures as NTA objects are given in our references listed below. Due to size limits of this paper, we give only one example of how to apply NTA structures for logical analysis. This example uses variables that contain more than two values, which makes it possible to analyze the logic models that go beyond the propositional calculus. Suppose we want to find the object, which is in one of three rooms. At the same time, you may not enter one of the rooms (there is a danger), and one of the rooms is empty. There are two following clues:

1. The first room is not dangerous, the second is not empty;
2. The third room is not dangerous, the second is not empty.

We also know that one of the clues is true, the other one is false.

Let us denote A the statement “the room contains the desired object”, B – “the room is empty” and C – “the room is dangerous”. Attributes names will be equal to the names of the rooms R_1, R_2, R_3 . Each of these rooms corresponds to one of situations from the permitted situations set $\{A, B, C\}$.

The given conditions can be expressed by C - n -tuples as follows.

$$T_1[R_1R_2R_3] = [\{A, B\} \{A, C\} *]; 2) T_2[R_1R_2R_3] = [* \{A, C\} \{A, B\}].$$

To solve the problem, it is necessary to consider two possible cases: 1) The first clue is true, and the second is false; 2) The first clue is false, and the second is true.

Let us consider the first case. If T_2 is false, its negation

$$\overline{T_2} =]\emptyset \{B\} \{C\}[$$

is true. To find a solution, one needs to find the intersection of the true conditions:

$$T_1 \cap \overline{T_2} = [\{A, B\} \{A, C\} *] \cap]\emptyset \{B\} \{C\}[=$$

$$[\{A, B\} \{A, C\} *] \cap \begin{bmatrix} * & \{B\} & * \\ * & * & \{C\} \end{bmatrix} = [\{A, B\} \{A, C\} \{C\}].$$

The resulting C - n -tuple contains 4 possible placement options (one can find this by calculating the Cartesian product of $\{A, B\} \times \{A, C\} \times \{C\}$). However, according to the conditions of the problem, the correct placement must have different situations in different rooms, i.e. the only placement (B, A, C) is correct. It means that the desired object is located in the second room. Similar results can be obtained when testing the second case.

The above listed relationships provide proving all the rest NTA laws, which assisted in composing algorithms of NTA operations and checks of equality and inclusion for any homotypic NTA objects.

To implement NTA operations and checks for any (possibly non-homotypic) NTA objects, it is necessary to define generalized operations and relations based on some operations upon attributes.

NTA: GENERALIZED OPERATIONS AND OPERATIONS UPON ATTRIBUTES

Let us introduce operations on attributes first.

- *Renaming of attributes* is used when it is necessary to substitute variables, particularly, in algorithms for calculating transitive closure of a graph.
- *Transposition of attributes* is an operation that swaps columns in an NTA object's matrix and respectively changes the order of attributes in the relation diagram.
- *Elimination of an attribute* ($-Attr$) is done in the following way: a column is removed from an NTA object, and the corresponding attribute is removed from the relation diagram.
- *Addition of a dummy attribute* ($+Attr$) is done when the added attribute is not yet present in the relation diagram of an NTA object. This operation adds the name of a new attribute into the relation diagram and simultaneously adds a new column with dummy components into the corresponding place; dummy components “*” are added into C - n -tuples and C -systems, and dummy components “ $\emptyset$ ” are added into D - n -tuples and D -systems.

By defining the operation $+Attr$, we have eliminated the restriction in the theory of relations stating that algebra-of-sets laws are only applicable to the relations defined upon the same Cartesian product.

The operation $+Attr$ is often used to reduce some different-type NTA objects to the same relation diagram by adding the missed attributes. Then one can perform all necessary operations and checks by means of standard NTA algorithms.

Considering this, we have introduced *generalized operations* ($\cap_G, \cup_G$). They differ from

corresponding operations of algebra of sets by the only feature: NTA objects are reduced to the same relation diagram before execution of these operations, which are semantically equal to logical connectives: disjunction and conjunction correspondingly. *Generalized relations* ($\subseteq_G$ and $=_G$) are defined similarly.

Algebra of n -ary relations with the mentioned generalized operations and relations is proved to be isomorphic to the ordinary algebra of sets.

NTA: CORRESPONDENCE WITH PROPOSITIONAL AND PREDICATE CALCULI

A C - n -tuple corresponds to *conjunction* of unary predicates with different variables. For example, in predicate calculus the C - n -tuple

$$P[XYZ] = [P_1 P_2 P_3]$$

where

$$P_1 \subseteq X; P_2 \subseteq Y; P_3 \subseteq Z$$

models the logical formula

$$H = P_1(x) \wedge P_2(y) \wedge P_3(z)$$

where X , Y and Z are ranges of values for the variables x , y and z .

A D - n -tuple corresponds to *disjunction* of unary predicates with different variables. For instance, the D - n -tuple

$$Q[XYZ] = [Q_1 \quad Q_2 \quad Q_3]$$

corresponds to the logical formula

$$F_2(x, y, z) = Q_1(x) \vee Q_2(y) \vee Q_3(z).$$

C -systems are convenient for representing *disjunctive normal forms* (DNFs) of unary predicates,

and D -systems model *conjunctive normal forms* (CNFs) of unary predicates.

An elementary n -tuple that is a part of a non-empty NTA object corresponds to a *satisfying substitution* in a logical formula.

An empty NTA object corresponds to an *identically false formula*.

An NTA object $P[XY...Z]$ that equals to $X \times Y \times \dots \times Z$ corresponds to a *valid formula*, or a *tautology*.

A non-empty NTA object models a *satisfiable formula*.

Now let us consider quantifiers in NTA.

Adding a dummy attribute to a C - n -tuple or a C -system corresponds to the *rule of generalization* in predicate calculus. For example, if an NTA object

$$G[XZ] = \begin{bmatrix} \{a, c\} & * \\ \{a, c, d\} & \{b, c\} \end{bmatrix}$$

corresponds to a formula $F(x, z)$ of this calculus, by adding a dummy attribute Y into this NTA object we get an NTA object

$$G_1[XYZ] = +Y(G[XZ]) = \begin{bmatrix} \{a, c\} & * & * \\ \{a, c, d\} & * & \{b, c\} \end{bmatrix}$$

which models the formula $\forall y F(x, z)$.

Elimination of an attribute from a C -system (D -system) is proved to be equivalent to the universal (existential) quantification on this attribute. Thus, if an NTA object $R[XYZ]$ models a formula $F(x, y, z)$, then $-Y(R[XYZ])$ corresponds to $\exists y F(x, y, z)$, if $R[XYZ]$ is a C -system, and to $\forall y F(x, y, z)$, if $R[XYZ]$ is a D -system.

To obtain NTA representations for formulas of propositional calculus containing logical variables $x_1, x_2, \dots, x_n$, it is enough to assign every variable x_i with the attribute $X_i = \{0, 1\}$. Then the dummy component $*$ corresponds to the set $\{0, 1\}$, literal x_i corresponds to $\{1\}$, and its nega-

tion $\neg x_i$ corresponds to $\{0\}$. For example, the formula

$$F(x, y, z) = (x \vee \neg z) \wedge (\neg y \vee z)$$

corresponds to the D -system

$$R[XYZ] = \begin{bmatrix} \{1\} & \emptyset & \{0\} \\ \emptyset & \{1\} & \{0\} \end{bmatrix}.$$

The negation of this formula

$$\neg F(x, y, z) = (\neg x \wedge z) \vee (y \wedge \neg z)$$

is modeled by the C -system

$$\overline{R}[XYZ] = \begin{bmatrix} \{0\} & * & \{1\} \\ * & \{0\} & \{1\} \end{bmatrix}.$$

It is easy to see that relations between C -objects (C - n -tuples and C -systems) and D -objects (D - n -tuples and D -systems) are in accordance with de Morgan's laws of duality. Due to this fact, they are called *alternative classes*.

All NTA objects model scopes of truth for their corresponding logical formulas. In NTA, attribute domains can be any arbitrary sets that are not necessarily equal to each other. Thus, NTA structures model formulas of *many-sorted predicate calculus*.

This paper does not consider matters of complexity in detail; they are described in (Kulik, 1995). It was proved that *computational complexity of algorithms on NTA structures fully corresponds to computational complexity of algorithms solving problems on logical structures*. However, the matrix properties of NTA objects allowed us to find some special cases that are solvable in polynomial time only (Kulik, 1995). Moreover, such structure is good to reduce computational complexity of NTA operations by their effective parallelization (Kulik & Fridman, 2014).

LOGICAL INFERENCE IN NTA

Theoretical principles of NTA provide solving the following problems of deductive analysis:

1. Problem of correctness check for an alleged consequence B from the given premises A_i .
2. *Problem of deriving possible consequences from the given premises A_i* considering certain semantical constraints, for instance, presence of given variables or their combinations in a consequence, deriving a consequence with the minimal number of significant variables, etc.

Unlike other logical systems that solve such problems by using inference rules with hardly optimized order of their implementation, NTA solves these problems by means of certain standard algorithms. To do so, we transform classical logic' formulas expressing premises and consequences into NTA objects and subject these objects to generalized operations and checks of equality ($=_G$) or inclusion ($\subseteq_G$). The transition to the algebraic representation becomes clearer, if you consider that NTA objects model scopes of truth for logical formulas. Then the correctness proof for an alleged consequence B from the premises A_i requires for calculating generalized intersections and checking the following generalized inclusion:

$$(A_1 \cap_G \dots \cap_G A_n) \subseteq_G B \quad (1)$$

Considering (1), the problem to derive possible consequences $\{B_j\}$ from the given premises A_i can be solved in NTA by using the relation that must be true for every B_j : $A \subseteq B_j$, where $A = A_1 \cap_G \dots \cap_G A_n$. Thus, any correct consequence from premises A_i can be modeled by an NTA object representing a superset of A . To find such objects, NTA proposes several techniques. One of them is *calculating projections of an NTA object*. This operation is simple if A is a C -system. Then the projection can be found by operation $-Attr$. It is easy to prove that any projection of an NTA

object models its superset. This technique allows checking the possibility to obtain consequences containing certain variables.

The relationship (1) allowed for an NTA definition of the *minimal consequence* from axioms $A_1, A_2, \dots, A_n$ that equals to the NTA object

$$A = A_1 \cap_G A_2 \cap_G \dots \cap_G A_n.$$

This consequence is the minimal one as it is proved to satisfy the two following statements:

1. Any strict subset of A is not derivable from the axioms $A_1, A_2, \dots, A_n$;
2. If S is an arbitrary NTA object from the proper universe, the NTA object $A \cup S$ is derivable from the axioms $A_1, A_2, \dots, A_n$.

In some kinds of reasoning models, a consequence B is considered a suitable solution of an inference task if this consequence contains fewer variables than the minimal consequence A (such a consequence can be found by using the above-introduced Technique 1). This decrease in number of variables in the formula $B = A \cup S$ results from transformation of some variables in the formula A into dummy ones after adding S to it.

NTA: MODELS OF PLAUSIBLE REASONING

Some difficulties in applying propositional and predicate calculi to analysis of plausible reasoning arose from the fact that mathematical logic recognizes inconsistency of a reasoning system (a theory) if and only if both a corollary and its negation follow from the same premises. Conversely, commonsense reasoning recognize inference of contrary corollaries (for instance, both statements “All A are B ” and “No A are B ”) as a definite criterion of inconsistency for a knowledge system though they are not formally contradictory.

In order to eliminate this and other discrepancies between formal logic and natural deduction, we propose a concept of “*collision*” in our logical analysis system (Kulik, 1995).

Unlike a logical contradiction, collisions can have opposite interpretations in different cases. For example, within one system the equality $R = \emptyset$ means the absence of an object that is necessary for existence of the system, and in another system (where the object R is not necessary) this equality specifies a status of this object. The first case requires for changing the premises while the second case provides a new useful knowledge.

Collisions mostly occur during defeasible reasoning when new knowledge or a hypothesis is included into the logical system. Collisions indicate violations of some formal rules or restrictions providing consistency of this system.

By using collisions, NTA allows for a formal definition of hypotheses. Let us suppose that a system of premises expressed as NTA objects $A_1, A_2, \dots, A_n$ is given and the NTA object $A = A_1 \cap_G A_2 \cap_G \dots \cap_G A_n$ is calculated.

Then a certain formula H is called a *hypothesis* under the following conditions:

1. $A \subseteq_G H$ is false. Otherwise, H is a consequence according to (1).
2. The hypothesis is correct, if the object $H \cap_G A$ contains no collisions.

The first condition can be expressed in other way: $A \setminus_G H \neq \emptyset$. The second condition assumes that the hypothesis is a premise or an axiom.

Usually, forming and checking of hypotheses accompany other analysis methods for defeasible reasoning. Below, we will describe usage of hypotheses in searching for abductive conclusions.

Abduction is forming of an explanatory hypothesis when one knows some of premises and an estimated consequence confirmed with facts or reasonable arguments, but a formal check does not infer it from the given premises. For example, abduction is used during diagnostics.

Let us now formally define abduction.

If B is an estimated consequence from the premises

$$A = A_1 \cap_G A_2 \cap_G \dots \cap_G A_n$$

and the statement $A \subseteq_G B$ is known to be false, then a formula H is an *admissible abductive conclusion* when the two following conditions are met:

1. H is a hypothesis (i.e. $A \subseteq_G H$ is false) and $H \cap_G A$ is not empty;
2. $(H \cap_G A) \subseteq_G B$, that is adding H into the system of premises results in deducibility of the estimated consequence B .

Considering the above definitions, we propose a search algorithm for abductive conclusions.

Examples of using this algorithm in (Kulik, Zuenko, & Fridman, 2015).

MEASURES IN NTA

NTA structures consist of sets of n -tuples being Cartesian products in essence. A measure (a probabilistic measure in particular) can be easily calculated for a Cartesian product, that is why NTA structures can be readily immersed into a space with a measure. For instance, if measures of a C - n -tuple' components are known, the measure of the C - n -tuple itself equals to the product of these measures. It is not that easy to calculate a measure for a C -system since intersections of some its elements can be not empty. That is why we have developed algorithms to transform any NTA structure to an orthogonal C -system that has no overlapping pairs of C - n -tuples (Kulik, 2007). Then its measure equals to the sum of the measures for the C - n -tuples comprising it.

Let us introduce two possible ways to use metric properties of NTA objects.

The first way proposes the number of elementary n -tuples as a measure based on the following regularity. Let a C - n -tuple be given in a space of attributes $P = [P_1 \ P_2 \ \dots \ P_n]$ with a finite number of their values. Then the power $|P_i|$ of every component P_i is a finite number as well, and the power (that is the number of elementary n -tuples) of the C - n -tuple P equals to $|P| = |P_1| \cdot |P_2| \cdot \dots \cdot |P_n|$. By using our orthogonalization algorithm, it is possible to count the power of every NTA object having finite sets as its components.

The second way to use metric properties of NTA objects is suitable for implementation of the logic-probabilistic analysis (LPA) for systems modeled by logical functions (Kulik, 2007).

This approach corresponds to the *direct problem* of LPA, when the probability of a complex event is calculated for given probabilities of elementary events. In the *inverse problem*, it is necessary to calculate probabilities of elementary events based on the data about probabilities of certain complex events. After this, one can calculate the probabilities of other complex events. Problems solved in probabilistic logic are of this type (Kulik, 2007; Nilsson, 1986).

Let us analyze an example from (Nilsson, 1986).

Assume that a totality of the given events is defined by formulas A and $A \supset B$ of propositional calculus, besides, $p(A) = p_1$ and $p(A \supset B) = p_2$. It is necessary to estimate the probability $p(B)$ for the event B .

We will solve this problem by means of NTA. Here are two logical variables A and B , which model elementary events as well. Let us suppose probabilities of these events equal to $p(A)$ and $p(B)$ correspondingly. The statement of this problem gives $p(A) = p_1$. Now we will express the given formulas in NTA structures within the universe

$$\{0, 1\}^2: A = [\{1\} *];$$

$$B = [* \{1\}];$$

$$A \supset B = \overline{A} \vee B =]\{0\} \{1\}[= \begin{bmatrix} \{0\} & * \\ \{1\} & \{1\} \end{bmatrix}$$

(here the D - n -tuple modeling the formula $A \supset B$ is transformed into an orthogonal C -system).

Then the probabilistic formulas for the events A and $A \supset B$ are as follows:

$$P(A) = p_1; P(A \supset B) = (1 - p(A)) + p(A)p(B) = p_2.$$

So, we have got a system of two equations:

$$p(A) = p_1;$$

$$(1 - p(A)) + p(A)p(B) = p_2$$

that allows to easily conclude

$$p(B) = \frac{p_1 + p_2 - 1}{p_1}.$$

This is the precise result for this problem while N. Nilsson (1986) gives an interval estimate for this result, namely

$$p_2 + p_1 - 1 \leq p(B) \leq p_2.$$

FUTURE RESEARCH DIRECTIONS

In the authors' opinion regarding NTA development, future research ought to include the following directions: context-oriented database (knowledge base) management systems; additional means of immersing NTA structures into measure spaces and assigning NTA objects with quantitative features; modeling dynamic intelligent systems (Vinogradov, Zhilyakova, & Osipov, 2002) within the situational approach; applications NTA methods to smart electromechanical systems (SEMS) (Kulik & Fridman, 2016).

CONCLUSION

This paper introduces a new mathematical instrument, n -tuple algebra (NTA), implementing the algebraic approach to the general theory of n -ary relations. Unlike relational algebra and theory of binary relations, NTA uses Cartesian product of sets rather than elementary n -tuples as a basic structure and is based on some generalized operations and relations. They provide logical processing of relations with different relation diagrams; these methods have no analogies in convenient theories. NTA belongs to the class of Boolean algebras. It is proved to be isomorphic to algebra of sets and supports unified implementation of various data/knowledge structures typical for artificial intelligence and many techniques of their semantic and logical analyses. Matrix properties of NTA objects allow to decrease complexity of intellectual procedures for some new structural and statistical classes of CNFs with polynomially identifiable satisfiability as well as by efficient parallelization of logical inference algorithms. NTA structures can be readily immersed into spaces with a measure. In particular, some new solutions were found for the direct and inverse problems of the logic-probabilistic analysis.

ACKNOWLEDGMENT

The authors would like to thank the Russian Foundation for Basic Researches (grants 14-07-00256, 14-07-00257, 15-07-04760, 15-07-02757, 16-29-04424, and 16-29-12901) for partial funding of this research.

REFERENCES

Ceri, S., Gottlob, G., & Tanca, L. (1990). *Logic programming and databases*. Berlin: Springer-Verlag. doi:10.1007/978-3-642-83952-8

Codd, E. F. (1971). A Database Sublanguage Founded on the Relational Calculus. *Proceedings of 1971 ACM-SIGFIDET Workshop on Data Description, Access and Control*. doi:10.1145/1734714.1734718

Kulik, B., Fridman, A., & Zuenko, A. (2013). Logical Inference and Defeasible Reasoning in N-tuple Algebra. In X. Naidenova & D. Ignatov (Eds.), *Diagnostic Test Approaches to Machine Learning and Commonsense Reasoning Systems* (pp. 102–128). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-1900-5.ch005

Kulik, B. A. (1995). A Logic Programming System Based on Cortege Algebra. *Computer and Systems Sciences International*, 33(2), 159–170.

Kulik, B. A. (1995). New Classes of Conjunctive Normal Forms with a Polynomially Recognizable Property of Satisfiability. *Automation and Remote Control*, 56(2), 245–255.

Kulik, B. A. (2007). N-Tuple Algebra-Based Probabilistic Logic. *Systems Analysis and Operations Research*, 46(1), 111–120.

Kulik, B. A., & Fridman, A. Ya. (2014). Data and Knowledge Parallel Processing by Means of N-Tuple Algebra. *Journal of Engineering and Applied Sciences (Asian Research Publishing Network)*, 9(8), 1354–1357.

Kulik, B. A., Zuenko, A. A., & Friedman, A. Ya. (2015). Deductive and Defeasible Reasoning on the Basis of a Unified Algebraic Approach. *Scientific and Technical Information Processing*, 42(6), 8–16. doi:10.3103/S0147688215060076

Nilsson, N. J. (1986). Probabilistic Logic. *Artificial Intelligence*, 28(1), 71–87. doi:10.1016/0004-3702(86)90031-7

Russel, S., & Norvig, P. (2003). *Artificial Intelligence: A Modern Approach* (2nd ed.). Prentice Hall.

Ullman, J. D. (1989). *Principles of Database and Knowledge-Base Systems*. Rockville, MD: Computer Science Press.

Vinogradov, A. N., Zhilyakova, L. Yu., & Osipov, G. S. (2002). Dynamic Intelligent Systems: I. Knowledge Representation and Basic Algorithms. *Computer and Systems Sciences International*, 41(6), 953–960.

Wirth, N. (1976). *Algorithms + Data Structures = Programs*. Prentice-Hall.

Zakrevskij, A. D., & Gavrilov, M. A. (1969). *Lyapas: A Programming Language for Logic and Coding Algorithms*. Academic Press.

ADDITIONAL READING

Kulik, B., Fridman, A., & Zuenko, A. (2012). Algebraic Approach to Logical Inference Implementation. *Computing and Informatics, Slovakia*, 31(6), 1295–1328.

Kulik, B. A., & Fridman, A. Ya. (2016). Logical Analysis of Data and Knowledge with Uncertainties in SEMS. In A. E. Gorodetskiy (Ed.), *Smart Electromechanical Systems* (pp. 45–59). Switzerland: Springer International Publishing. doi:10.1007/978-3-319-27547-5_5

Vinogradov, A. N., Zhilyakova, L. Yu., & Osipov, G. S. (2002). Dynamic Intelligent Systems: I. Knowledge Representation and Basic Algorithms. *Computer and Systems Sciences International*, 41(6), 953–960.

KEY TERMS AND DEFINITIONS

C-n-Tuple: An n -tuple of components defined in a certain relation diagram; domain of each component is a subset of the domain of the corresponding attribute.

Collisions: Situations occurring during defeasible reasoning when some new knowledge (hypothesis) is inputted. Such situations can be recognized as violations of some formally expressed rules and/or limitations which control consistency and meaning content of a logical system.

C-System: A set of homotypic C - n -tuples that are denoted as a matrix in square brackets and equal to the intersection of these C - n -tuples. The rows of this matrix are C - n -tuples.

D - n -Tuple: An n -tuple of components enclosed in reversed square brackets and equal to a diagonal C -system with the same diagonal components.

D -System: A set of homotypic D - n -tuples equal to the intersection of these D - n -tuples.

Generalized Operations and Relations: They differ from similar operations and relations of algebra of sets by the only feature: NTA objects (operands) are reduced to the same relation diagram before executing these operations or checking the relations.

N-Tuple Algebra: An algebraic system whose support is an arbitrary set of n -ary relations expressed by specific structures, namely, C - n -tuple, C -system, D - n -tuple, and D -system. These structures provide a compact expression for sets of elementary n -tuples.

D

A Proposed Framework for Incorporating Big-Data Technology in National Crisis Management Center

Magdy M. Kabeil

Al-Yamamah University, Saudi Arabia

Ahmad M. Kabil

University of Wisconsin – Whitewater, USA

INTRODUCTION

Along with the major development in many aspects of life, the current millennium has brought for several countries significant crises of different kinds and various degrees (Farazmand, 2014). The increasing jeopardizing to crises calls for a transformational change in countries' capability to manage crises. Crises may be natural or manmade. It may be on organizational or national levels.

The main measure of merit in crisis management is the ability to allocate all relevant capabilities to be the most effective on the right points at the right times over all crisis phases (Kabil & Kabeil, 2014 & 2011). On such measure, the national crisis management currently faces two critical challenges. The first is the high pace of crisis scenario relative to the corresponding management decision cycles. The second critical challenge is the large scale of volume, variety, and velocity of pertinent data.

However, if we record a crisis scenario on a video recorder and replay it in slow motion, we will see the crisis features very similar to a regular problem's features. On the other side, if a National Crisis Management Center (NCMC) can provide the decision makers with the capabilities that allow them to feel, recognize, abstract, comprehend, analyze, and decide faster, then the crisis management processes will be as manageable as the regular problem management processes.

The Big-Data Technology (BDT) as an accurate representation of real life could be used in both ways, to understand the real life and to control it. Fast cycles of understanding and controlling real life situations are the key functions of national crisis management.

This chapter presents a framework for incorporating BDT in traditional conceptual designs of NCMC. The updated conceptual design is validated using the Analytical Hierarchy Process (AHP) and the Quality Function Deployment (QFD) technique. The conceptual design provides a system that is responsive to decision makers' need in developing their own crisis management centers. The initial stages and steps required for implementing the conceptual design in a country are given.

BACKGROUND

The quality of decision, such as any other engineering product, is built from the very beginning all through the decision processing cycle (Davern et al. 2008). The generic decision cycle in crisis management context starts with data gathering that is processed further to higher levels of information, knowledge, intelligence, wisdom, and decision. Decisions are implemented through a Command, Control, Communication, Computer, Intelligence, Surveillance, and Reconnaissance

(C4ISR) System to move a real-world-situation to be more suitable for the next decision or action (Kabil & Kabeil 2014). The model has been updated in response to the evolution of BDT as depicted in Figure 1.

There are several definitions of the term “Big-Data.” Perhaps the most popular definition is based upon the IBM’s differentiation that is based on 3 attributes of the 3 V words, Volume, Variety, and Velocity (IBM, 2016; Sagioglu & Sinanc 2013). Leverage the traditional “Data” concepts and technologies to “Big-Data” ones affects all higher levels of the model.

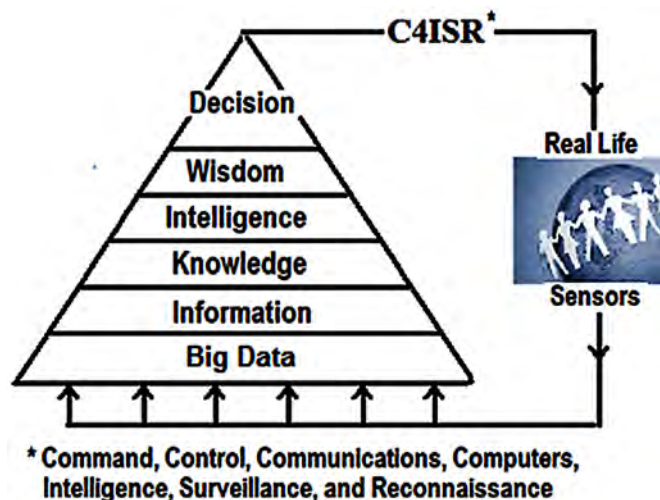
One of the most important implications of using BDT in crisis management is crisis crowdsourcing. Developing more new technologies capable of collecting, communicating and disseminating individuals information on macro scales has led to more forms of crowdsourcing in crisis management (Kahl et al., 2012). Recent research efforts on crisis crowdsourcing address new issues around using information and communication technologies to engage and coordinate with the wider public during crisis management. The Crisis Crowdsourcing Framework defined by Liu (2014) uses the same traditional information-gathering six dimensions of why, who, what, when, where, and how to identify the interaction mechanisms

that should be considered by designers of crisis crowdsourcing systems.

On the “Where” dimension, the geographically oriented Big-Data could be provided using Location-Based Services (LBS). The LBS could be used to locate living entities or objects across telecommunication networks. Several technologies are utilized by LBS for knowing where an information device is geographically located. Examples of technologies used for LBS are the Global Positioning Satellites (GPS), Assisted Global Positioning Satellites (AGPS), Cell Identification (Cell ID), Broadband Satellite Network (BSN), Radio Frequency Identification (RFID), and Near-Field Communication (NFC).

Ma (2011) proposes a real-time LBS that can enable hospitals treating victims during a disaster relief scenario. The author analyzes how a network of hospitals can use such LBS to handle medical aspects of disaster relief. He suggests a framework for following available hospitals capability, allocating victims among hospitals, tracking patients who are in-transit from one hospital to another, and developing strong collaborative network of personnel across all areas of the hospitals service. The emerging technology of smart watches with embedded sensors to recognize objects and forearm gestures expands the use of such appli-

Figure 1. From data to decision



cations in full integration with RFID and NFC technologies.

Remote sensing is a valuable source of spatial information in the event of a natural disaster, and its utility has been proven on many occasions around the world (Kumar et al., 2014). Joyce et al. (2009) address the use of remote sensing as a rapid-response data source and present a number of data types and image processing techniques along with airborne and satellite technologies used to map and monitor a crisis theater.

Big-Data is used for finding insights in new types of content to answer questions that were previously considered beyond the reach (LaValle, et.al, 2011). Che et al. (2013) consider Big-Data as greatly expanded assets to human and they argue that all what we need then is to develop the right tools for overcoming the scalability challenge of Big-Data and developing Big-Data Analytics.

Baboo and Kumar (2013) present an overview of the next generation data warehouse architecture that is based upon Big-Data for better business insights. The proposed architecture consists of three tiers, which are the extracting of Big-Data from multiple sources, storing and cleansing, and the analytical reporting. Within such architecture, a typical Big-Data processing framework relies on cluster computers with a high-performance computing platforms. The data processing tasks are being deployed by running parallel programming tools on a large number of computing nodes. A single data processing task is split into many small tasks each of which is running on one or multiple computing nodes (Wu et al., 2014). Che et al. (2013) review several attempts that have been made on exploiting massive parallel processing architectures for Big-Data Analytics.

The Big-Data Analytics helps analyzing online data about all relevant individuals and objects rather than measuring data from a sample of population. Measuring data from a sample usually carries two types of statistical errors known as “Type I” and “Type II” (Stevens, 2012). In crisis management context, “Type I” error is to neglect an individual or a group in population who should

be considered. And, “Type II” error is to consider an individual or a group who is not relevant to the case. Additional shortage of measuring data from a sample is the lack of concurrency and timeliness due to the period of conducting a sampling process.

The Big-Data paradigm provides a vast variety of new kinds of priors and estimation techniques to inform all sorts of decisions. The impact on the economy has been referred to as “the new oil” (Kolb & Kolb, 2013: p. 10). Its impact on the social sciences can be compared with the impact of the invention of the telescope for astronomy and the invention of the microscope for biology (Nature Editorial, 2008). The presented study can help investigating its impact on crisis management on the national level.

INCORPORATING BIG-DATA TECHNOLOGIES IN NCMC

Kabil and Kabeil (2014) suggest a generic conceptual design of National Crisis Management Center (NCMC) that covers initial requirements and basic functions for managing crises on the national level of a country. The conceptual design consists of four functional units, which are the Intelligence Unit, the Design & Operations Unit, the Choice Unit, and the Integration & Display Unit.

The Intelligence Unit includes 7 modules, which are Monitoring Module, Database Module, Planned Communication Module, Ad-Hoc Communication Module, Reconnaissance Module, Assumptions & Forecasting Module, and Validation & Reasoning Module. The Design & Operations Unit includes 6 modules, which are the Analysis & Knowledge Management Module, Modeling & Simulation Module, Operations Module, Social & Media Module, Political Module, and Government Departments Module. Both the Choice Unit and the Integration & Display Unit had not been decomposed into modules at this time.

The generic conceptual design of NCMC depicted above has been updated to include emerging BDTs. Keeping the same basic structure of the

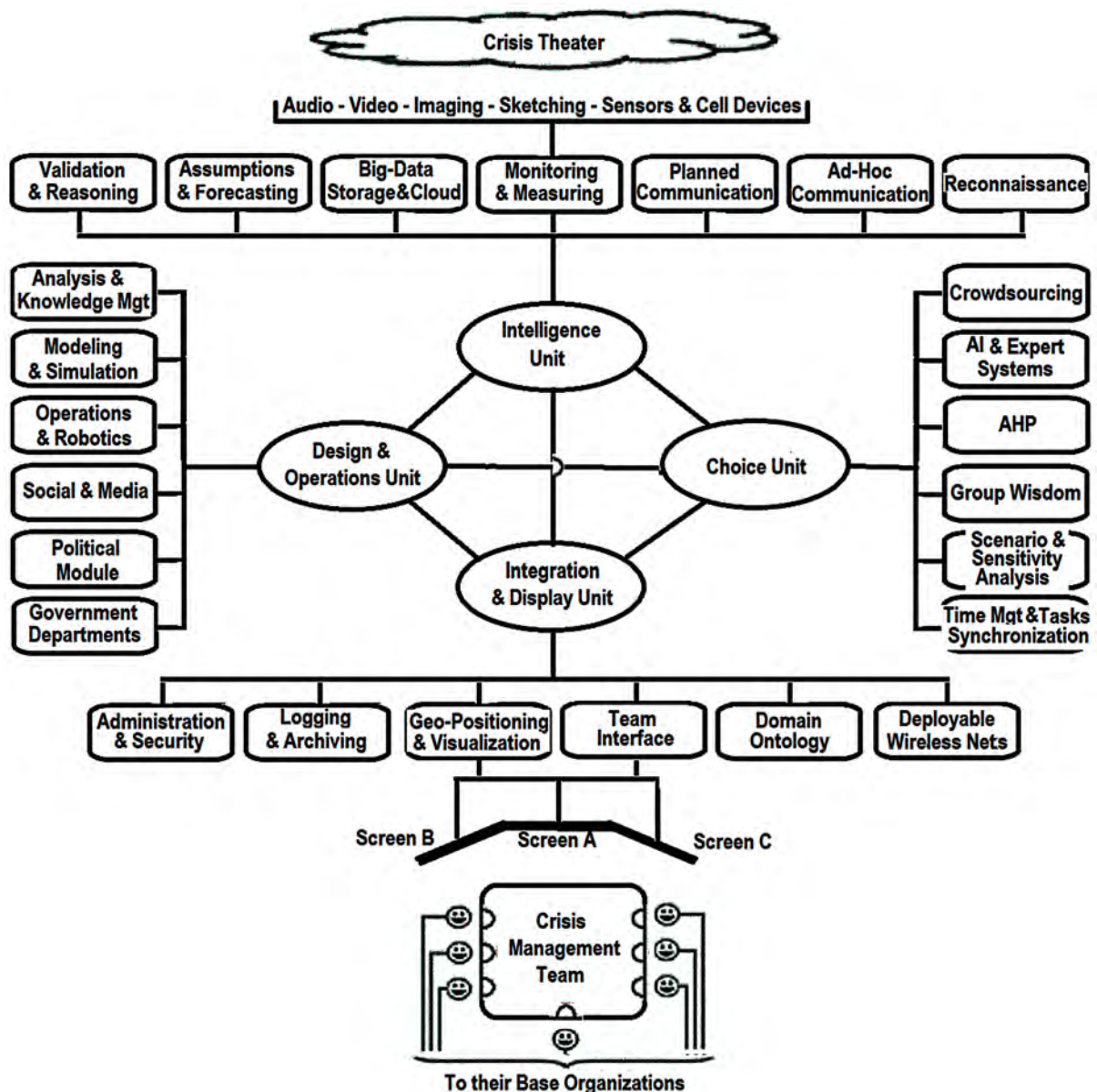
center, the functions of existing modules have been updated and several new modules have been added. The updating of the NCMC's conceptual design is illustrated in Figure 2.

The Intelligence Unit has been updated by adding a new element of Sensors & Cell Devices to the Monitoring & Measuring Module and by replacing the Database Module with the Big-Data Storage & Cloud Module. The functions of the remaining four modules of the Intelligence Unit

have been updated under the same names. Also, the functions of the six modules of the Design & Operations Unit have been updated under the same names. The Choice Unit has been decomposed into six modules and the Integration & Display Unit has been also decomposed into six modules.

The updated unites and modules of the NCMC's conceptual design are defined in the following items.

Figure 2. Updated conceptual design of NCMC



NCMC-1: Intelligence Unit

The information relevant to the decision situation is gathered and maintained by the intelligence unit all through the crisis life cycle. The unit includes seven updated modules as defined in the following items:

NCMC-1.1: Monitoring & Measuring Module, which supports the production, dissemination, and display of crisis theater data. It consists of five nodes that can create, receive, edit, transmit, and store video clips as well as images, graphics, voice, text data, and measuring signals.

NCMC-1.2: Big-Data Storage & Cloud Module, which includes an integrated collection of NCMC's core data corresponding to the needs of expected crises.

NCMC-1.3: Planned Communication Module, which is responsible for building and managing a network of distributed data sources belonging to several organizations and government departments.

NCMC-1.4: Ad-Hoc Communication Module, which is responsible for developing an index of data sources of different specialties and deploying an ad-hoc network of communication with them.

NCMC-1.5: Reconnaissance Module, which is responsible for conducting physical information- capturing activities of strategic or operational significance.

NCMC-1.6: Assumptions & Forecasting Module, which is responsible for filling gaps of data in the decision structure and conducting forecasting activities including meteorological forecasts, prediction of decision consequences, and early warning capabilities.

NCMC-1.7: Validation & Reasoning Module, which is responsible for checking the validity of data collected through the Intelligence Unit's modules and conducting data reasoning that allows crisis management team to

rely on information as facts, even though they have not specifically verified it in a physical manner.

NCMC-2: Design and Operations Unit

The main responsibility of the Design & Operations Unit is to develop enough alternatives for decision choices. The unit includes six updated modules as defined in the following items:

NCMC-2.1: Analysis & Knowledge Management Module, which is responsible for dealing with both explicit and tacit knowledge. The module is in full coordination with the Big-Data Storage & Cloud Module (NCMC-1.2) in the Intelligence Unit for Big-Data mining and extracting knowledge about behavior of individuals, groups, concepts, and objects.

NCMC-2.2: Modeling & Simulation Module, which is responsible for developing and maintaining the model-base of the NCMC.

NCMC-2.3: Operations & Robotics Module, which is responsible for conducting military actions in crisis context. Diverse military operations in crises context are a part of one or more of Direct Action, Psychological Operation, Civil Affairs, Information Operations and Hacking, and Robots Operations. Despite the fact that each one of these missions uses its own DSS-C4ISR system, interfacing these systems to the crisis decision structure is the duty of the Operations & Robotics Unit in the NCMC.

NCMC-2.4: Social & Media Module, which is responsible for communicating with the community and media at large.

NCMC-2.5: Political Module, which is responsible for handling the foreign affairs and international relations issue in crisis context. The module includes BDT in multilingual capability of communication in continuous accumulated processes.

NCMC-2.6: Government Departments Module, which is responsible for communicating and coordinating with other ministries and administrations according to their specialties in crisis management.

NCMC-3: Choice Unit

The Choice Unit is responsible for developing a preferences structure of group decision makers in crisis context (Limayem et al., 2006). The updated structure of the Choice Unit includes six modules as defined in the following items:

NCMC-3.1: Crowdsourcing Module, which is responsible for conducting public participation in crisis management. Crowdsourcing is an emerging, large-scale arena for Internet-mediated interaction that has implications for both informal and formal responses (Liu, 2014). With a focus on persistent citizen communications, the module follows their spatial arrangements, and how the emerging information serve different crisis management functions (Palen & Liu 2007).

NCMC-3.2: Group Wisdom Module, which is responsible for incorporating and integrating the wisdom of decision makers in the choosing process (Adkins et al. 2002).

NCMC-3.3: AHP Module, which is used for organizing the complexity of decision structure, determining the internal and total priorities of alternatives, forecasting about next move of crisis actors, allocating resources, selecting the most relevant cases to retrieve for investigation, integrating qualitative factors with quantitative factors in decision structure, and evaluating the ethical responsibilities in inverse decision support cases (Kabil 2012).

NCMC-3.4: Scenario & Sensitivity Analysis Module, which is responsible for studying consequences of decisions in each scenario, for checking the sensitivity of a decision to any changes may arise in decision parameters, and for checking the validity of data

received from the Assumptions & Forecasting Module of the Intelligence Unit.

NCMC-3.5: Artificial Intelligence & Expert Systems Module, which is responsible for improving the effectiveness and efficiency of decision-making processes using several Artificial Intelligence & Expert Systems methods (Khalil et al. 2012).

NCMC-3.6: Time Management & Tasks Synchronization Module, which is responsible for monitoring, controlling and adjusting the pace of decision cycle over crisis events.

NCMC-4: Integration and Display Unit

The Integration & Display Unit is responsible for integrating all unites and modules of the NCMC and facilitating its team to access transparently the internal and external components of the system (Huang et al., 2002). The updated structure of the Integration & Display Unit includes six modules as defined in the following items:

NCMC-4.1: Team Interface Module, which is responsible for creating and providing interface that minimizes the time of team comprehension. The module is able to combine several displays in a single dynamic presentation and to supply the displays being generated with various interactive facilities such as zoom, pan, rotate, focus, query, mark, filter, or follow a hyperlink.

NCMC-4.2: Domain Ontology Module, which is responsible for integrating the different ontologies into a new one and for overcoming the diversity of information sources as well as the diversity of information users.

NCMC-4.3: Geo-Positioning & Visualization Module, which is responsible for effective preparation, organization, and representation of Big-Data on intelligent maps suitable for anticipated recipients within the time constraints (Andrienko & Andrienko, 2007).

NCMC-4.4: Deployable Wireless Networks Module, which is responsible for administrating a rapidly deployable broadband wireless communications system for crisis management and emergency response.

NCMC-4.5: Logging & Archiving Module, which is responsible for logging and archiving data that are typically performed on a cyclic basis.

NCMC-4.6: Administration & Security Module, which is responsible for managing a role-based access control and security system. An integrated security mechanism is used to optimize security protocols and mechanisms of various heterogeneous wireless networks in NCMC.

CHECKING THE VALIDITY OF CONCEPTUAL DESIGN

The exercise suggested in this study for testing validity of the updated conceptual design of the NCMC is to recall the W3C's definition of critical steps in crisis response ontologies and to define the requirements for decision support throughout the crisis phases. Based upon the definition of such requirements, a model of QFD is developed to

test the capability of the NCMC to provide these types of support (Kabeil, 2009).

The World Wide Web Consortium (W3C, 2016) represents the noticeable effort in the way of building crisis response ontologies and getting the benefits of semantic web services. According to the W3C's definition, crisis response management should consider 14 critical steps that are included in the first 14 requirements in Table 1. To support decision making and planning, additional knowledge about the actors and types of resources and infrastructure should be considered. So, in addition to the W3C's definition, the table includes more requirements that are common for crisis management scenarios.

The crisis management requirements from the NCMC to handle the crisis are listed with their relative priorities using the AHP. The table includes all requirements in 25 rows along with the Weight and Relative Weight Percent.

The evaluation of the conceptual design modules is conducted using the QFD technique (Kabeil, 2010). The set of crisis requirements and their priorities are used as inputs to the QFD model. The core of QFD is called the House of Quality "HoQ" (Sankaran et.al., 2008) as depicted in Figure 3.

Figure 3. House of quality

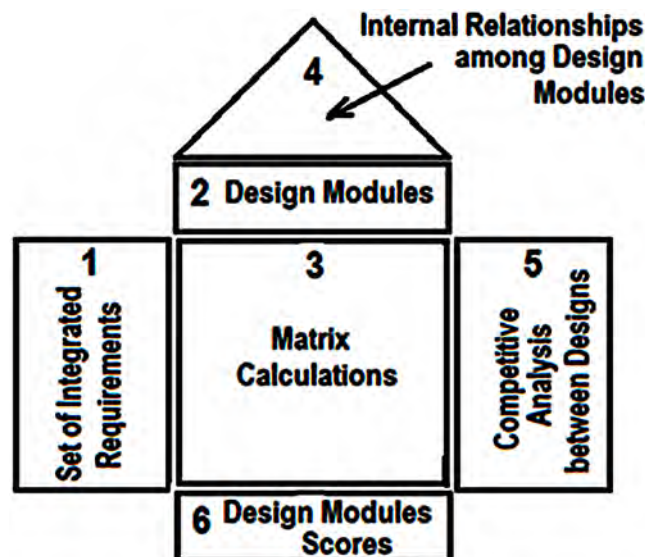


Table 1. Priorities of crisis management requirements

RN	% Rel. Weight	Weight (1-5)	Requirements
1	4.2%	5	Once crisis is widely anticipated, sharing of data describing response and resource characteristics are needed.
2	4.2%	5	As the crisis unfolds gathering of data on its scope and emerging effects.
3	4.2%	5	As the response begins, gathering of data on its outages and missing links and matching with relief capacity.
4	4.2%	5	As the response by first responders is overwhelmed, sharing relief requests to prioritize relieving the first responders who are most overloaded or tired.
5	4.2%	5	As the relief unfolds gathering and integrating data from all responders to build a common baseline map of the situation and facilitate probes and first attempts at proactive data gathering.
6	4.2%	5	Characterizing problem states as chaotic (no baseline and no reliable map), complex (changing too fast to identify causes, requires probes) or manageable.
7	4.2%	5	Rapidly deploying compatible information and communication systems to local authorities and institutions capable of dealing with the manageable situations.
8	4.2%	5	Calling for expert review of action proposals to limit/contain chaotic situations, and mass peer review of probes that better define complex ones, with intent to limit the unanticipated side effects of management decisions.
9	4.2%	5	Comparing predicted to measured effects of interventions within 48-72 hours.
10	4.2%	5	Identifying situations which are not improving and calling for more options or more resources.
11	4.2%	5	Helping experienced response teams to move on to the more complex situation by facilitating rapid handoff and just-in-time training of those less experienced.
12	4.2%	5	Guiding recovery and reconstruction efforts by identifying those outages or problems that most inhibit the resilience networks and outside relief efforts.
13	4.2%	5	Guiding resilience efforts by identifying which prevention and anticipation options (e.g. evacuation) could have prevented the most morbidity or loss of life-sustaining infrastructure.
14	4.2%	5	Passing off all data gathered in the disaster to the appropriate authority after the crisis passes, updating databases of vulnerable persons and places.
15	3.3%	4	Locating an individual, facility, vehicle, or object on a map.
16	3.3%	4	Searching for an individual on the map based on her/his profile.
17	4.2%	5	Receiving an emergency request from an individual with her/his location.
18	4.2%	5	Viewing the location of emergency requests on the map.
19	4.2%	5	Disseminating notifications and alerts to each concerned individual.
20	3.3%	4	Storing, maintaining and updating each individual profile and a list of her/his relatives and friends with their current location.
21	2.5%	3	Keeping historical Big-Data of the entire case for analysis and research.
22	4.2%	5	Insuring availability of the system
23	4.2%	5	Insuring security of the system
24	4.2%	5	System user being able to setup and manage the system.
25	4.2%	5	System admin being able to setup and manage the system.
		120	Total Scores

The HoQ matrix correlates requirements “what” to design modules “how” using symbols of “strongly related,” “moderately related,” “weakly related,” or “not related.” The internal relationships between design modules are identified in the upper part of the HoQ matrix. The row score of design

modules is the sum of the product between the relative weights of the integrated requirements and the correlations between requirements and modules as partially depicted in Figure 4.

The output of the HoQ matrix is summarized in Table 2. The relative weights (%) of the NCMC

Figure 4. HoQ of the NCMC

Table 2. Relative weight percent of the NCMC modules

NCMC-4.6	4.1
NCMC-4.5	3.3
NCMC-4.4	2.5
NCMC-4.3	9.5
NCMC-4.2	3.6
NCMC-4.1	5.1
NCMC-3.6	4.0
NCMC-3.5	0.9
NCMC-3.4	3.9
NCMC-3.3	4.2
NCMC-3.2	2.6
NCMC-3.1	6.6
NCMC-2.6	2.6
NCMC-2.5	0.5
NCMC-2.4	2.5
NCMC-2.3	2.1
NCMC-2.2	4.6
NCMC-2.1	3.3
NCMC-1.7	2.2
NCMC-1.6	2.6
NCMC-1.5	5.0
NCMC-1.4	3.2
NCMC-1.3	1.9
NCMC-1.2	11.0
NCMC-1.1	8.2

Cloud module (NCMC-1.2), the Geo-Positioning & Visualization module (NCMC-4.3), and the Monitoring & Measuring Module (NCMC-1.1).

IMPLEMENTATION AND FURTHER RESEARCH DIRECTIONS

Basic modules of NCMIS are developed as a conceptual design, which provides a general foundation for the detailed design and implementation of an application. The conceptual design provides a system that is flexible and responsive to decision makers' need in developing their own crisis management centers. The initial stages of the implementation of NCMC in a country consist of the establishment of facilities and staffs to carry out the basic functions of the center. Once an initial allocation of resources has been made and the functions of the system have been established in an operating condition, the major task in the development of the system involves integrating the various and widespread strategic information activities already existing in the country to the center.

The steps required for implementing the conceptual design of NCMC in a country are:

1. Defining the expected crises in the country and the probability of occurrence of each using historical approach, Delphi method, among other forecasting techniques.
2. Determining the weight of each crisis using the expected monetary value as the probability of occurrence times the value of crisis' consequences.
3. Determining the relative weight of each crisis using the normalization method.
4. Defining the list of requirements for each expected crisis.
5. Based on pairwise comparisons, a composite weight and relative priority are calculated for each item of requirements of each crisis using the AHP.

6. Determining the set of integrated requirements as the union of all requirements of all expected crisis of the country along with the total relative weight of each item in the set using the AHP as depicted in Figure 5.
7. Using the set of Integrated Requirements and their total relative weights as inputs to the HoQ matrix. The outputs of the matrix are the relative weights (%) of the NCMC modules that each component contributes in the implemented design.
8. Establishing other functions of the system in an operating condition.

The future research will be focusing on testing the proposed model in different contexts of country infrastructure and with different scenarios of crisis.

CONCLUSION

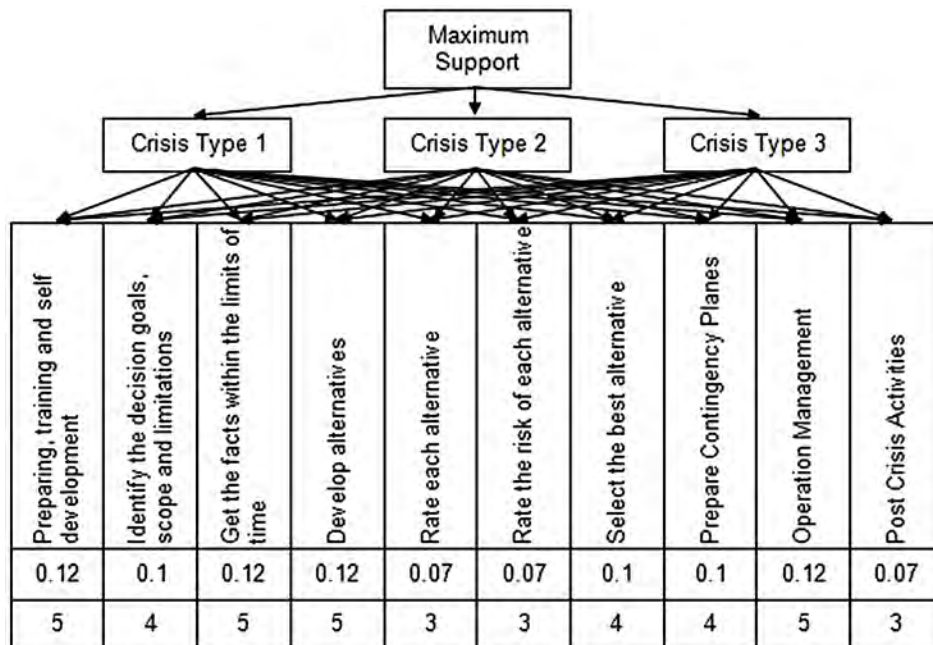
The national crisis management currently faces two critical challenges. The first is the high pace of crisis scenario relative to the corresponding

management decision cycles. The second critical challenge is the large scale of volume, variety, and velocity of pertinent data. Incorporating emerging BDT in the design of NCMC is the proposed solution in this study.

Basic modules of updated NCMC are presented as a conceptual design, which provides a general foundation for the detailed design and implementation of an application. The validation of the proposed conceptual design is conducted on a testing-case based upon the W3C's definition of crisis response ontology. The relative percentage of the total score that each module contributes in the conceptual design is evaluated using the QFD technique. The results show that each requirement of national crisis is supported by a set of integrated modules in a balanced way. The modules of the highest contribution in the conceptual design are the modules most related to BDT.

The conceptual design provides a system that is flexible and responsive to decision makers' need in developing their own crisis management centers. The initial stages of the implementation of NCMC in a country and main steps required

Figure 5. Example of integrated requirements with the relative weight of each



for implementing the conceptual design are given. By combining the conceptual design with a test of internal validity and steps of implementation, the study has provided useful information to build upon.

REFERENCES

- Adkins, M., Burgoon, M., & Nunamaker, J. Jr. (2002). Using Group Support Systems for Strategic Planning with the United States Air Force. *Decision Support Systems*, 34(3), 315–337. doi:10.1016/S0167-9236(02)00124-0
- Andrienko, N., & Andrienko, G. (2007). *A Framework for Decision-Centred Visualisation in Civil Crisis Management*. Springer Berlin Heidelberg. doi:10.1007/978-3-540-36728-4_33
- Baboo, S., & Kumar, P. R. (2013). Next generation data warehouse design with big data for big analytics and better insights. *Global Journal of Computer Science and Technology*, 13(7), 19–23.
- Che, D., Safran, M., & Peng, Z. (2013). From big data to big data mining: challenges, issues, and opportunities. In *Database Systems for Advanced Applications* (pp. 1–15). New York, NY: Springer. doi:10.1007/978-3-642-40270-8_1
- Davern, M., Mantena, R., & Stohr, E. (2008). Diagnosing Decision Quality. *Decision Support Systems*, 45(1), 123–139. doi:10.1016/j.dss.2007.12.012
- Editorial, N. (2008). Community cleverness required. *Nature*, 455(7209), 1–1. doi:10.1038/455001a PMID:18769385
- Farazmand, A. (2014). *Crisis and emergency management: Theory and practice*. Boca Raton, FL: CRC Press. doi:10.1201/b16715
- Huang, W., Wei, K., Watson, R., & Tan, B. (2002). Supporting Virtual Team-Building with a GSS: An Empirical Investigation. *Decision Support Systems*, 34(4), 359–367. doi:10.1016/S0167-9236(02)00009-X
- IBM. (2016). *IBM big data and information management*. Accessed January 5, 2016 from <http://www-01.ibm.com/software/data/bigdata/>
- Joyce, K. E., Belliss, S. E., Samsonov, S. V., McNeill, S. J., & Glassey, P. J. (2009). A review of the status of satellite remote sensing and image processing techniques for mapping natural hazards and disasters. *Progress in Physical Geography*, 33(2), 183–207. doi:10.1177/0309133309339563
- Kabeil, M. (2009). A Proposed Framework for Developing a National Crisis Management Information System. *International Journal of Information Systems for Crisis Response and Management*, 1(4), 50–74. doi:10.4018/jiscrm.2009071004
- Kabeil, M. (2010). An AHP-QFD Approach to Developing DSS for Crisis Management. *The International Journal of Management and Decision Making*, 11(1), 55–68. doi:10.1504/IJMDM.2010.033643
- Kabil, A. (2012). Evaluating Ethical Responsibility in Inverse Decision Support. *Advances in Decision Sciences*, 2012, 873710. doi:10.1155/2012/873710
- Kabil, A., & Kabeil, M. (2011). Initial Requirements of National Crisis Decision Support System. In *Crisis Response and Management and Emerging Information Systems: Critical Applications* (pp. 262–286). Hershey PA: IGI Global. doi:10.4018/978-1-60960-609-1.ch017
- Kabil, A., & Kabeil, M. (2014). Initial Requirements of National Crisis Decision Support System. In *Crisis Management: Concepts, Methodologies, Tools and Applications*. Hershey PA: IGI Global. doi:10.4018/978-1-4666-4707-7.ch049
- Kahl, A., McConnell, C., & Tsuma, W. (2012). Crowdsourcing as a tool in conflict prevention. *Conflict Trends*, (1), 27–34.

- Khalil, K. M., Abdel-Aziz, M., Nazmy, T. T., & Salem, A. B. M. (2012). Artificial Immune Systems Metaphor for Agent Based Modeling of Crisis Response Operations. In *Handbook on Decision Making* (pp. 417–428). New York, NY: Springer. doi:10.1007/978-3-642-25755-1_20
- Kolb, J., & Kolb, J. (2013). *The Big data Revolution*. Create-Space Independent Publishing Platform.
- Kumar, S., Morstatter, F., & Liu, H. (2014). *Twitter Data Analytics*. New York, NY: Springer. doi:10.1007/978-1-4614-9372-3
- LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big-Data, analytics and the path from insights to value. *MIT Sloan Management Review*, 52(2), 21–32.
- Limayem, M., Banerjee, P., & Ma, L. (2006). Impact of GDSS: Opening the Black Box. *Decision Support Systems*, 42(2), 945–957. doi:10.1016/j.dss.2005.08.004
- Liu, S. B. (2014). Crisis Crowdsourcing Framework: Designing Strategic Configurations of Crowdsourcing for the Emergency Management Domain. *Computer Supported Cooperative Work*, 23(4-6), 389–443. doi:10.1007/s10606-014-9204-3
- Ma, W. (2011). Real-time location systems for hospital emergency response. *IT Professional*, 13(2), 37–43. doi:10.1109/MITP.2011.31
- Palen, L., & Liu, S. B. (2007, April). Citizen communications in crisis: anticipating a future of ICT-supported public participation. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 727–736). doi:10.1145/1240624.1240736
- Sagiroglu, S., & Sinanc, D. (2013). Big-Data: A review. *International Conference on Collaboration Technologies and Systems (CTS)*, 42–47. doi:10.1109/CTS.2013.6567202
- Sankaran, R., Senthil, V., Devadasan, S., & Pramod, V. (2008). Design and Development of Innovative Quality Function Deployment Model. *International Journal of Business Innovation and Research*, 2(2), 203–222. doi:10.1504/IJBIR.2008.016653
- Stevens, J. P. (2012). *Applied multivariate statistics for the social sciences*. New York: Routledge.
- W3C. (2016). *Disaster Management*. Accessed January 5, 2016 from <http://esw.w3.org/topic/DisasterManagement>
- Wu, X., Zhu, X., Wu, G. Q., & Ding, W. (2014). Data mining with Big-Data. *Knowledge and Data Engineering. IEEE Transactions on*, 26(1), 97–107.

ADDITIONAL READING

- Estellés-Arolas, E., & González-Ladrón-de-Guevara, F. (2012). Towards an integrated crowdsourcing definition. *Journal of Information science*, 38(2), 189–200., 195.
- Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Khan, S. U. (2015). The rise of big data on cloud computing: Review and open research issues. *Information Systems*, 47, 98–115. doi:10.1016/j.is.2014.07.006
- Kabeil, M., & Kabil, A. (2015). Big Data in National Crisis Decision Support Systems. Chapter 5 in Van Vactor, J. (editor) *Crisis Management: A Leadership Perspective*. Nova Science Publishers.
- Li, Y., & Lu, Y. (2015). Analysis on Urban Crisis Conduction Under the Background of Big Data. In *Low-carbon City and New-type Urbanization* (pp. 317–328). Springer Berlin Heidelberg.
- Wybo, J. L., Fogelman-Soulié, F., Gouttas, C., Freyssinet, E., & Lions, P. (2015). Impact of social media in security and crisis management: A review. *International Journal of Emergency Management*, 11(2), 105–128. doi:10.1504/IJEM.2015.071045

KEY TERMS AND DEFINITIONS

Analytical Hierarchy Process: A technique for pairwise comparisons using judgements of experts to derive priority scales.

Big-Data Analytics: The systematic use of Big-Data and related business insights developed through applied analytical disciplines to drive fact-based decision-making.

Big-Data Technology: A high-resolution representation of real life in such a way that all movements, activities, thoughts, conversations, relations, feelings could be tracked, recorded and processed almost online.

Crisis: A time of danger when important decisions must be developed, communicated and implemented.

Crowdsourcing: The process of obtaining knowledge support by soliciting contributions from a large group of people, especially from an online community.

Location-Based Service: A technology used to locate living entities or objects across telecommunication networks.

National Crisis: A situation or time at which a nation faces intense difficulty, uncertainty, danger or serious threat to people and national systems and organizations and a need for non-routine rules and procedures emerge accompanied with urgency.

Near-Field Communication: A form of short-range wireless communication where the antenna used is much smaller than the wavelength of the carrier signal.

Quality Function Deployment: A technique correlates the design modules to common national crisis management requirements and translates them into specific plans to produce products to meet those needs.

Radio Frequency Identification: A generic term for technologies that use radio waves to automatically identify entities, whatever inanimate, movable, or live.

D

Quality Evaluation for Evolving Conceptual Database Design

Elvira Immacolata Locuratolo
CNR ISTI, Italy

INTRODUCTION

Research at boundary between *information systems* and *software development* is important for today's technology; (Buragga & Zaman, 2013). Before creating an information system, the business and organizational domain in which the information system is used must be represented. This representation, called *conceptual model*, is a critical point for the success of the information system development.

Model-driven engineering is an approach to software development that separates the specification of the system functionalities from its implementation on a particular platform. Models are the primary artifacts of software engineering. Two main categories of *mappings* between models are distinguished:

- Vertical mapping, relating system models at different abstraction levels;
- Horizontal mapping, relating models at the same abstraction level.

The requirements for engineering information modeling enclose complex objects, data exchange, data sharing, web based applications, imprecision uncertainty, and knowledge management. Techniques, such as web and artificial Intelligence, have been introduced in industrial applications.

Databases and Database systems play a main role in supporting data management and in implementing engineering information. Conceptual database design is defined by approaches to *mapping* from conceptual to logical database models (Elmasri & Navathe, 2015). Algorithms of

conceptual database graph partitioning (Locuratolo & Rabitti, 1998; Spagnolo, 2000), originally introduced as algorithms of mapping from semantic data models to object database systems, have been also exploited as techniques for constructive conceptual database design (Locuratolo, 2013). *Information systems and software development* is a dynamic, interesting research area, which provides the finest and fastest way to solve typical problems.

Innovative evaluation of quality is introduced to design a database conceptual model which can result from a vertical mapping or a horizontal mapping (Locuratolo, 2014a). In the former case, the cost of specifications/proofs is given through the cardinality of variables and constants; in the latter case, the conceptual model is evaluated in terms of hidden classes. These represent the *saving* that is obtained for what has not been explicitly specified/proven. The hidden classes can also be exploited to test other quality desiderata/dimensions, such as *easiness of use* and *flexibility*. The numeric saving of consistency can be determined from the corresponding conceptual evaluations. The vice-versa is not possible.

In this article, a new approach to conceptual database design, called *evolving conceptual database design*, is introduced. The approach integrates a *structure for the preservation of classes/concepts* within the conceptual database design: starting from a database conceptual graph, an ontology defined at the boundary between concept theory and computer science, called *structure for the preservation of database classes/concepts*, is designed (Locuratolo, 2015b); the leaves of this structure are mapped to classes of a logical database model. A

constructive approach of logical database design is considered to handle some dynamic aspects of the conceptual database design and the quality of this approach is evaluated.

BACKGROUND

Model-driven engineering is an approach to software development that separates the specification of the system functionality from its implementation on a particular platform. A *model-driven architecture* makes models the primary artefacts of software engineering. The development consists in the definition of a platform-independent model (PIM) of a system and in the application of parameterized transformations to this PIM in order to obtain one or more specific platforms. Another important specification is *architecture-driven modernization*, which is the process of understanding and evolving existing software. It involves modifications, reuse and enterprise-application integrations. An approach to modernization is a way to revitalize existing applications and systems. It concerns a wide variety of models and mappings between models, allowing integration and transformation of those models. Over the last few years, *model-driven architecture* has been integrated with *enterprise information systems*, and other technologies, such as *semantic web*, *semantic web services*, and *ontology* techniques.

Graphs of classes representing *conceptual/ logical database models* are based on concepts which are common in *enterprise information systems*, *model-driven engineering*, and *ontology*. New engineering requirements and current technologies favour the *conceptual database design evolution*. In the following, definitions useful to the purposes of the chapter are introduced: first a description of the *conceptual database design* is provided; then an extension of some definitions is given.

Conceptual Database Design

D

In this section, definitions useful to introduce the concept of conceptual database design are given:

- **[Database Conceptual Class]:** Couple (v, V) , where v is a label denoting the class objects and connoting the corresponding concept, and V are the attributes associated with the objects.
- **[Is-a Relationship]:** Conceptual class (u, U) is in *is-a relationship* with conceptual class (v, V) iff
 - The objects of class (u, U) are enclosed into the objects of class (v, V) ;
 - The attributes of class (u, U) are inherited from class (v, V) ;
 - Class (u, U) can have specific attributes
- **[Database Conceptual Graph]:** Oriented acyclic graph G_c of database classes in *is-a* relationship.
- **[Revised Partitioning]:** Step-sequence of decompositions resulting in the *finest partition* of G_c . At each step of decomposition, the following properties hold:
 - **Root Partitioning:** The root objects of a conceptual graph G_c are partitioned into the root objects of the conceptual graphs obtained after the decomposition.
 - **Root Labeling:** The root labels of the graphs obtained from decomposition represent the partition. These labels are obtained combining the root label of the decomposed graph with the labels of the root directed descendants.
 - **Root Structuring:** The root labels can be decomposed into two parts separated by the “-“sign: that on the left of this sign consists of label intersections, whereas that on the right consists of label unions. One of the two parts can be empty.

- **Consistency:** The following implicit information is specified through the root labels: only the attributes of class $\langle u \rangle$ are associated with a node labeled by $\langle u - v \rangle$, whereas the attributes of all the classes $\langle v_1 \rangle \dots \langle v_n \rangle$ with the node labelled by $\langle v_1 \cap \dots \cap v_n \rangle$.
- **Object/Logical Graph:** The obtained disjoint classes can be recomposed into an object/logical graph G_0 .
- **[Database Logical/Object Graph]:** Oriented acyclic graph G_0 of database classes in $is-a_0$ relationship. The $is-a_0$ differs from the $is-a$ relationship since the objects of class $\langle u, U \rangle$ and the objects of class $\langle v, V \rangle$ are disjoint.

Properties

In a *database conceptual graph*, each object instance can belong to any class of the graph thus enhancing *flexibility*.

The *revised partitioning* gives the properties characterizing the partitioning algorithms, which are algorithms of mapping from conceptual to logical database models. The objects of class $\langle u, U \rangle$ are denoted by $\langle u \rangle$. Exploiting the [Revised partitioning] properties, attributes of both, the classes of the decomposed database conceptual graphs or the disjoint classes of a logical/object database graph, can be associated to the classes.

In a *database logical graph*, each object instance can belong to one and only one class of the graph, thus enhancing *efficiency*. In the following, extended notions are given for a *constructive approach of conceptual database design* (Locuratolo, 2013).

Constructive Conceptual Database Graphs

This section is based on the extensions of previous definitions:

- **[Extended Database Conceptual Class]:** Triple (v, V, Op) that enlarges the couple (v, V) of the *database conceptual class* with the set Op of *basic operations* to add/remove objects or modify the class attributes.

In (v, V, Op) , both *attributes* and *basic operations* are defined by *functions*. Specifically, the basic operations are predicate transformers preserving the conceptual class constraints, i.e., *functions from predicates to predicates*. The weakest precondition semantics for a basic operation op is the following:

$$[Op(v, V)(par_list)]R \Leftrightarrow (conceptual\ class-constraints(v, V) \Rightarrow conceptual\ class-constraints(v, V)^*) \wedge R^*$$

where R is a predicate on the variables of class $\langle v, V \rangle$, *conceptual class-constraint* is the predicate which formalizes the conceptual class definition, and the star predicates are the mentioned predicates after the operation. Taking into consideration this observation, also the extended conceptual class (v, V, Op) can be denoted by $\langle v \rangle$.

- **[Extended Is-a Relationship]:** The extended conceptual class $\langle u \rangle$ is in an *extended is-a relationship*, denoted by $is-a'$ with class $\langle v \rangle$, iff
 - The objects of the extended class $\langle u \rangle$ are enclosed into the objects of the extended class $\langle v \rangle$.
 - The attributes and the elementary operations of the extended class $\langle u \rangle$ are inherited from The extended class $\langle v \rangle$ the extended.
 - class $\langle u \rangle$ can have specific attributes and elementary operations.
- **[Extended Database Conceptual Graph]:** Oriented acyclic graph G'_C of *extended classes* in *extended is-a* relationships, denoted by $is-a'$.

Further steps of *extended database conceptual graph* based on set-theory extensions can be proposed.

Observation

In a database conceptual graph, classes can be in *multiple inheritance*, i.e., classes with more *is-a* links to other classes. However, as this situation can be approached through the graph decomposition in disjoint trees, in the following, the considered graphs will be limited to graphs without classes in multiple inheritance.

Innovative Evaluation of Quality

The techniques proposed for the evaluations of models can be *theoretical*, such as ontological evaluations, *empirical*, such as experiments and cases studies, or *mixed*, such as theoretical approaches with empirical experience (Krogstie, Halpin & Keng, 2005). These techniques, which are mainly based on *metric analysis*, *cognitive analysis* or *statistical approaches*, result in numerical values, qualitative or statistical evaluations of quality. A *conceptual “measure” of quality*, given in terms of *hidden classes*, is proposed in (Locuratolo, 2014a). In order to reach this achievement, two different kinds of mapping are considered: the former kind is a vertical mapping which starts from the specification of a database application supported by a formal model, and results in a model based on semantic data models (Cardenas & McLeod, 1990). The latter is a horizontal mapping which starts from a database conceptual graph and achieves the resulting model of the previous approach. The use of *database notations* and of *is-a relationships* allows the achievement of the following conflicting quality dimensions/desiderata: *easiness of use and flexibility*. *Consistency* is a further quality dimension/desiderata to be proved.

The vertical mapping is exploited as a means for:

- Justifying the achievement of the *quality dimensions/desiderata*.
- Giving a *quantitative/numerical evaluation* of the *consistency costs*.

The *quantitative evaluation* is a number that expresses the *consistency cost* of what has been specified/proven in terms of variable and constant cardinality. This number gives no indication about the opportunity to specify or to prove at a *good cost*.

The horizontal mapping design is exploited as a means for

- Justifying the achievement of the *quality dimensions/desiderata*;
- Giving a *qualitative/conceptual evaluation* of the *consistency costs*.

The *qualitative/conceptual evaluation* is a set of hidden classes/extended classes that express the *saving* for what has been *implicitly specified/proven*. This “*measure*” gives indication about the *model goodness*, since it is related to the hidden classes and to their semantic richness. Other dimensions of quality, such as *easiness of use* and *flexibility*, can be tested on the hidden classes. The saving of the consistency costs can be determined *numerically*, starting from the corresponding qualitative evaluation. The vice-versa is not possible.

MAIN FOCUS OF THE ARTICLE

The main focus of this article is concerned with the proposal of a new approach of conceptual database design, called *evolving conceptual database design*, which introduces the *structure for the preservation of database classes/concepts* within the conceptual database design. The quality of the proposed approach is evaluated. The *structure for the preservation of database classes/concepts* is designed starting from a database conceptual graph. The leaves of this structure are mapped to a *logical/object database graph*. Horizontal steps

of constructive logical database design extend the model. The computational costs required to design the *structure for the preservation of database classes/concepts*, as well as the qualitative/conceptual costs of the logical models resulting from the constructive design, are discussed. In the following, the solution is given to enclose the *structure for the preservation of database classes/concepts within the conceptual database design*.

Solution and Recommendation

Given a conceptual graph G_c , the following *is-a*^c relationship is introduced to design a *generalization hierarchy of classes/concepts* (Locuratolo, 2015a; Locuratolo, 2016a).

Concept $[v]$ *is-a*^c concept $[u]$ if class $\langle u \rangle$ *is-a* class $\langle v \rangle$ in G_c .

A *structure for the preservation of database classes/concepts* (Locuratolo, 2015) is an ontology, defined at the boundary between *concept theory* and *computer science* that encloses:

- All and only the *classes/concepts* related to the conceptual graph G_c ;
- All and only the *is-a*^c relationships among *classes/concepts*;
- All and only the incompatible concepts/disjoint classes, which are leaves of the structure.

The structure for the preservation of database classes/concepts has been designed to achieve the *class/concept completeness* property, i.e., the integration of h generalization hierarchies of classes/concepts resulting from *partitioning algorithms of maximum steps* applied to the conceptual graph G_c , with h is the number of the direct sons of G_c root (Locuratolo, 2014b). A *mapping* is then established between the leaves of the *structure for the preservation of database classes/concepts* and the classes of a *database logical/object graph* G_o . The *revised partitioning* properties are

exploited to associate attributes with the classes during the mapping approach from the leaves of the structure to the logical/object database graph G_o . A constructive approach to database logical/object database design is proposed to extend the logical model G_o .

In the following, the definition of extended database logical/object graph is given:

- **[Extended Database Logical/Object Graph]:** Graph of database logical/object classes G'_o in *extended is-a'* relationship, denoted by *is-a'*_o.

Class $\langle u \rangle$ *is-a'*_o class $\langle v \rangle$ iff the objects of class $\langle u \rangle$ are disjoint from the objects of class $\langle v \rangle$ and class $\langle u \rangle$ inherits both *attributes* and *basic operations* from class $\langle v \rangle$. Class $\langle u \rangle$ may have *specific attributes/basic operations*.

Steps of further *extended logical database design* based on set-theory extensions can be proposed to enrich further the database logical graph G'_o .

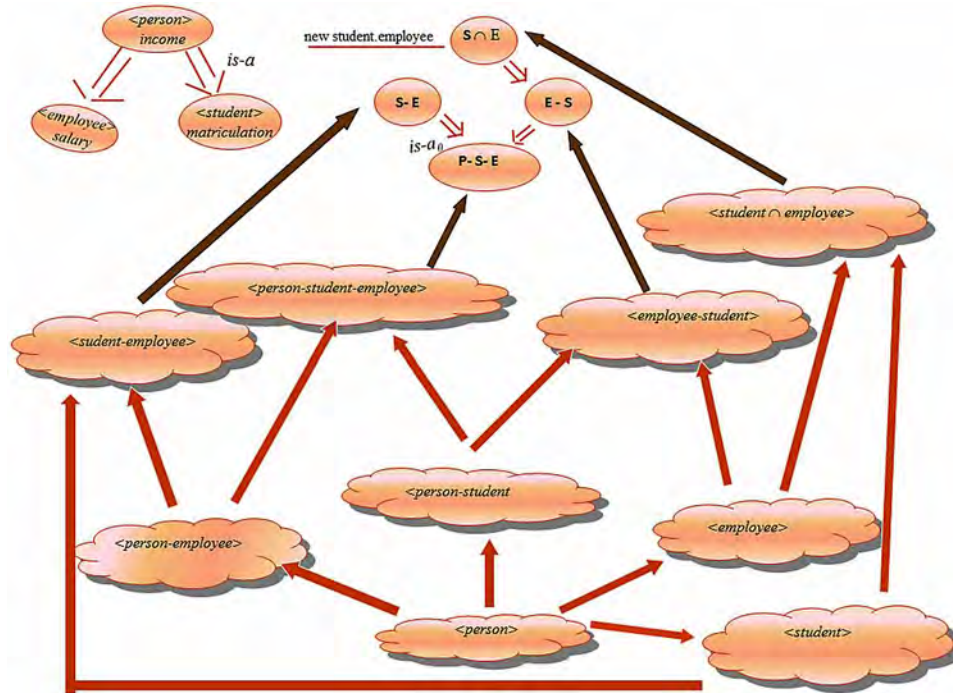
The approach is described in Figure 1: starting from the conceptual graph at the left side of the figure, the *structure for the preservation of database classes/concepts* is constructed and the leaves are mapped to a logical database graph. A new student employee is introduced in the appropriate class of the logical schema. In Figure 1, this operation is underlined.

Quality Evaluation

The quality evaluations of the *evolving conceptual database design* approach are based on the following costs:

1. Cost required to design the *structure for the preservation of database classes/concepts*;
2. Cost to map the leaves of the *structure for the preservation of database classes/concepts* to a database logical/object graph G_o ;
3. Cost of the *constructive logical database design*.

Figure 1. Approach description



In the following each of these costs is discussed:

Computational Cost

The cost required to design the *structure for the preservation of database classes/concepts* is the number of *classes/concepts* belonging to the structure. The algorithm resulting in the structure for the preservation of database classes/concepts is recursively defined starting from four initial cases. The formula introduced to evaluate the costs of the four initial cases is a function of the parameters n and h ; n is the number of the G_c classes, h is the number of the direct sons of the root. This cost, denoted by $C_1(h, n)$, gives the following values for the initial cases:

$$C_1(0, 1) = 1; C_1(1, 2) = 3; C_1(1, 3) = 5; C_1(2, 3) = 9$$

In the general case, this cost is *low* if G_c is a *single path tree*; it is *high* if G_c is a tree with only direct sons of the root, i.e. in the worst case,

and can be *acceptable* for applications with tree belonging to *intermediate cases*. An intermediate case can be transformed in the worst case, generating always necessary empty classes, which are not interesting in computer science but that can be interesting in concept theory.

The cost evaluation in the worst case is a function of the only parameter n . Specifically, it results $C_1(n) = 19/2 n!$ This cost represents an upper limitation to the complexity of the intermediate cases. As an example, $C_1(3) = 3 \cdot 19 = 57$ in the worst case, i.e. for a conceptual graph with three direct sons of the root and $C_1(2, 3) = 23$ for a conceptual graph with two direct sons of the root and three global sons. In the case in which the two direct descendents of the root have empty intersection, this cost is further reduced to 13.

Mapping Cost

The cost to map the structure for the preservation of database classes/concepts to the database logical graph G_0 is equal to the number of leaves of the

structure for the preservation of database classes/concepts. This cost, denoted by $C_2(n)$, is related to the conceptual tree topology, and its value belongs to the following interval: $n+1 \leq C_2(n) \leq 2^n$. The cost $C_2(n)$ is also the mapping cost of an algorithm of database graph partitioning initially introduced for the database conceptual design (Locuratolo & Rabitti, 1998). No approach of classes/concept preservation is part of this algorithm. In *evolving conceptual database design*, the *structure for the preservation of database classes/concepts* is part of the conceptual design. Although the mapping cost can be high in the worst case, its value can be really reduced for empty intersections among nodes, especially, in the case of applications with all the direct descendents of the tree root having empty intersections.

In (Locuratolo & Rabitti, 1998), both a table showing the mapping costs for tree with n nodes ranging in an interval from 1 to 16, where each node has at most p descendents ranging from 2 to 20 and a corresponding table showing the mapping cost for the same trees with disjoint direct descendents of the roots are provided. As a further observation, let us evidence that in real applications, not all the class/concepts of the *structure for the preservation of database classes/concepts* need to be mapped to the logical model, but only those where the corresponding logical class has at least one object. This observation can be useful exploited to handle the dynamic aspects of the *logical schema evolution* (Zicari, 1992).

Constructive Logical Database Design Cost

In order to evaluate the *costs of the constructive logical database design*, the following definition of specialized *logical class* is introduced:

- **[Specialized Logical Class]:** If (v, V) and (u, U) are logical classes, then class (u, U_e) , called *specialized logical class*, is defined as follows:

- class $\langle u \rangle$ and class $\langle v \rangle$ are disjoint.
- U_e is an enriched set of the attributes of class $\langle u \rangle$.
- Each attributes of class $\langle v \rangle$ is inherited by class $\langle u \rangle$.
- Each attribute of U_e belongs to class $\langle u \rangle$ or to class $\langle v \rangle$.

The *qualitative measure of quality* for a database logical graph G_o is defined subtracting from each specialized class (u, U_e) class (u, U) . Formally,

$$[u, (U_e - U)] = (u, U_e) - (u, U),$$

for each G_o class $\langle u \rangle$

This operation gives as a result, a class with the same label of the original class $\langle u \rangle$, and with attributes determined by set difference between the attributes of the specialized class, and those of the corresponding class. The “*measure*” of *quality* of the database logical graph represents *the saving* of implicit specifications. This saving can be determined *numerically* as a function of the specified variables and constants.

- **Property:** The “*measure*” of quality of an extended database logical graph G_o'' is the extension of the G_o' “*measure*” of quality.

This is a very interesting property: when the logical graph is extended, in order to evaluate the new saving, it is not required to begin *from scratch*. The new saving can be determined extending the previous saving. In order to be adherent with the other mentioned costs, in this paper, the *quality of the constructive, logical database design* is evaluated numerically. Finally, the following observation completes the section:

Observation

The *constructive logical database design* can be useful exploited to handle the dynamic aspects concerning the insertion/remotion of objects in the logical classes.

DISCUSSION

Semantic and object database models (Cardenas & McLeod, 1990) have similar mechanisms of abstraction; however, while semantic data models have never been implemented efficiently (Nixon & Mylopoulos, 1990), a remarkable level of efficiency has been achieved by object systems such as O2, or Object Store (Aloia, Barneva, Rabitti, 1992). On the contrary, while semantic data models are adequate for conceptual design, object database systems can display serious shortcomings in their ability to describe the dynamic nature of the real world entities.

The first algorithm of graph partitioning algorithm was proposed to achieve the two conflicting desiderata of *flexibility* and *efficiency*. Equivalence relations between conceptual and object database schemas ensure *correctness*. (Locuratolo & Rabitti, 1998). The results of this paper were used: within the refinement phase of the formal database design methodology ASSO (Locuratolo & Matthews, 1999; Locuratolo, 2005); to permit the database design to choose between two different kinds of target object models with different degrees of implementation efficiency, *Single Storage Object Data Model*, to retrieve an object with a single access; *Multi Storage Object Data Models*, to retrieve an objects with multiple storage access.

Algorithms of *graph partitioning*, exploited in conceptual database design, were reused to define a methodology of *database concept preservation* (Locuratolo & Palomäki, 2013; Locuratolo & Palomäki, 2015; Locuratolo, 2016b). Concept theory distinguishes *extensionality* from *intensionality*. In extensionality, a collection of objects is determined by its elements, whereas in intensionality, the interest is focused on concepts, properties, etc. It is possible to go from concepts to objects, not vice-versa. Although a formal background is available in concept theory, only recently, an algorithm, *the concept construction algorithm* was introduced (Locuratolo & Palomäki, 2013). The world of concepts is large and yet unexplored; on

the contrary, the world of database classes is well known. A relationship holds between an algorithm of graph partitioning, and the *concept construction algorithm*. The former results in disjoint classes that can be composed in a logical database graph; the latter in a concept structure, called *Ontology for database preservation*. It encloses:

- All and only the concepts related to an initial concept structure;
- All and only the logical relationships among the determined concepts;
- Results in leaves that can be mapped to a logical database graph.

Ontology for database preservation, which is formalized in concept theory, represents a concept structure difficult to be implemented, since it cannot enjoy the useful properties of the *is-a relationship*, exploited in modeling database conceptual graphs. The *structure for the preservation of database class/concepts* overcomes these problems since formalized in set-theory, and based on the *is-a^c* relationships.

FUTURE RESEARCH DIRECTIONS

The formal database design methodology ASSO has been defined to combine *easy* modification of the schema with *efficient* implementation, whilst ensuring specification *consistency* and design *correctness*. ASSO has been designed for *quality* achievements (Castelli & Locuratolo, 1995; Locuratolo & Matthews, 1999; Locuratolo, 2005). Object oriented methodologies, however, have always been very weak when dealing with data design; an example of methodology which overcomes this problem is IDEA (Ceri & Fraternali, 1997). In this methodology, data design comes first and application design follows. For the generalization hierarchies of IDEA, no transformation is required in the passage from analysis to design. The objects of a super class are partitioned into the objects of the subclasses.

This differs from the specialization hierarchy of the ASSO conceptual model, where the objects of the subclasses can overlap and a partitioning algorithm is exploited to map the conceptual model to the logical model. The methodology consists of two phases: conceptual design and refinements. The former focuses on the construction of the conceptual schema, a high level description of the database structure and behaviour. This schema is supported by a semantic data model extended with operations, modelled through predicate transformers. First order formulas permit to prove the conceptual schema consistency. Schema transformations resulting from appropriate extensions of the partitioning algorithms preserve the model semantics ensuring that the object schema is a correct implementation of the conceptual schema. The costs of the refinement approach of ASSO can be compared with the costs proposed in this chapter. From this evaluation, the opportunity to apply the *evolving conceptual database design* approach to the ASSO methodology results to be one of the possible future research direction.

CONCLUSION

This chapter proposes advancements of previously achieved research results: a new approach of conceptual database design, called *evolving conceptual database design* is proposed and the *quality* of this approach is evaluated. Specifically, the *structure for the preservation of database classes/concepts*, a concept structure defined at boundary between concept theory and computer science, is introduced within the *conceptual database design*: the *concept structure* is constructed starting from a database conceptual graph. The leaves of the concept structure are mapped to a *logical/object database graph*. Horizontal steps of constructive logical database design extend the model. The computational costs required to design the *structure for the preservation of database classes/concepts*, as well as the qualitative/conceptual costs of the logical models resulting from the constructive design, are discussed.

ACKNOWLEDGMENT

The author thanks Prof. Bruno Codenotti for the useful discussions during the preparation of this chapter; Dr. Paola Federica Canonico and Dr. Andrea Vannozzi for the technical support in the realization of the Figure.

REFERENCES

- Aloia, N., Barneva, S., & Rabitti, F. (1992). Supporting Physical Independence in Object Databases. *Database Technology*, 4(4), 265–286.
- Buragga, K. A., & Zaman, N. (Eds.). (2013). *Software Development Techniques for Constructive Information Systems Design*. Hershey, PA: Information Science Publishing. doi:10.4018/978-1-4666-3679-8
- Cardenas, A. F., & McLeod, D. (1990). *Research Foundations in Object-Oriented and Semantic Database Systems*. Englewood Cliffs, NJ: Prentice Hall.
- Elmasri, R., & Navathe, S. B. (2015). *Fundamentals of Database Systems*. London, UK: Addison Wesley.
- Krogstie, J., Halpin, T., & Keng, S. (Eds.). (2005). *Information Modeling Methods and Methodologies*. Hershey, PA: Idea Group Publishing. doi:10.4018/978-1-59140-375-3
- Locuratolo, E. I. (2005). Model transformations in designing the ASSO methodology. In P. van Bommel (Ed.), *Transformations of knowledge, Information and Data: Theory and Applications* (pp. 283–302). Hershey, PA: Information Science Publishing. doi:10.4018/978-1-59140-527-6.ch013

Locuratolo, E. I. (2013). A Constructive Approach for Conceptual Database Design. In K. A. Buragga & N. Zaman (Eds.), *Software development techniques for constructive information system design* (pp. 38–56). Hershey, PA: Information Science Publishing. doi:10.4018/978-1-4666-3679-8.ch002

Locuratolo, E. I. (2014a). Designing Conceptual Database Models for Innovative Evaluations of Quality. In B. Thalheim, H. Jaakkola, & Y. Kiyiki (Eds.), *Information Modelling and Knowledge Bases XXVI*. Amsterdam: IOS Press.

Locuratolo, E. I. (2015a). *Evolving Database Conceptual Graph Partitioning*. Cnr.isti. TR003.

Locuratolo, E. I. (2015b). *Preservation of Database Concepts: From Data Mining to Concept Theory*. TR-ISTI-029. In *Interdisciplinary Digital Preservation tools and technology*. IGI-Global.

Locuratolo, E. I. (2016a). Evolving Database Conceptual Graph Partitioning. In T. Welzer, M. Holbl, B. Thalheim, H. Jaakkola, & Y. Kiyiki (Eds.), *Information Modelling and Knowledge Bases XXVII*. Amsterdam: IOS Press.

Locuratolo, E. I. (2016b). Preservation of Database Concepts. In *Research in action*. Pisa: ISTI.

Locuratolo, E. I., & Palomäki, J. (2014b). Relating concept theory to computer science. In B. Thalheim, H. Jaakkola, & Y. Kiyiki (Eds.), *Information Modelling and Knowledge Bases XXVI*. Amsterdam: IOS Press.

Locuratolo, E. I., & Rabitti, F. (1998). Conceptual Classes and System Classes in Object Databases. *Acta Informatica*, 35(3), 181–210. doi:10.1007/s002360050118

Nixon, B., & Mylopoulos, J. (1990). Integration Issue in Implementing Semantic Data Models. In F. Bancilhon & P. Bueman (Eds.), *Advances in Database Programming Languages* (pp. 187–217). New York: ACM Press.

Spagnolo, A. M. (2000). *Incrementare la Qualità in Ambito Basi di Dati. Tesi*. Pisa, Italy: Pisa University.

Zicari, R. (1992). A framework for Schema Updates in an Object-Oriented Database System. In F. Bancilhon et al. (Eds.), *Building an Object-Oriented Database Systems. The story of O₂* (pp. 146–182). Burlington, MA: Morgan Kaufmann Publishers.

ADDITIONAL READING

Batini, C., & Scannapieco, M. (2006). *Data Quality Concepts, Methodologies and Techniques*. Berlin, Heidelberg: Springer.

Castelli, D., & Locuratolo, E. I. (1995). ASSO: a Formal Database Design Methodology. In H. Kangassalo, H. Jaakkola, S. Ohsuga, & B. Waangler (Eds.), *Information Modelling and Knowledge Bases VI* (pp. 145–158). Amsterdam, Berlin, Oxford, Tokyo, Washington: IOS Press.

Gladney, H. M. (Ed.). (2007). *Preserving digital Information*. Springer Verlag. Berlin: Heidelberg, New York.

Locuratolo, E. I., & Matthews, B. (1999). On the relationships between ASSO and B. In H. Jaakkola, H. Kangassalo, & E. Kawaguchi (Eds.), *Information Modelling and Knowledge Bases X* (pp. 235–253). Amsterdam, Berlin, Oxford, Tokyo, Washington: IOS Press.

Locuratolo, E. I., & Palomäki, J. (2013). Ontology for database preservation. In M. N. Ahmad, R. M. Colomb, & M. S. Abdullah (Eds.), *Ontology-based Applications for Enterprise Systems and Knowledge Management* (pp. 141–157). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-1993-7.ch008

Locuratolo, E. I., & Palomäki, J. (2015). Perspective for Database Preservation. In: Encyclopedia of Information Science and Technology-Third ed. Vol. III pp. 1855-1866. Mehdi Khosrow-Pour (ed.) Hershey, PA, US: IGI GLOBAL, 2015. doi:10.4018/978-1-4666-5888-2.ch179

KEY TERMS AND DEFINITIONS

Database Conceptual/Logical Class: Couple (v, V) , where the former component is a label denoting the class objects, and connoting the corresponding concept and the latter component is defined by attributes associated with the objects.

is-a Relationship: Conceptual class (u, U) is in *is-a relationship* with conceptual class (v, V) iff: the objects of class (u, U) are enclosed into the objects of class (v, V) ; the attributes of class (u, U) are inherited from class (v, V) ; class (u, U) can have specific attributes.

is-a^c Relationship: Concept $[v]$ is-a^c concept $[u]$ iff class (u, U) is-a class (v, V) in the conceptual graph G_c .

is-a_o Relationship: Relationship differing from the *is-a relationship*, since the objects of class (u, U) and the objects of class (v, V) are disjoint.

Qualitative/Quantitative “Measure” of Quality: Class defined “subtracting” from the specialized logical class (u, U_e) class (u, U) . Formally $(u, (U_e - U)) = (u, U_e) - (u, U)$. This class expresses the *saving* of implicit specification. From the qualitative “measure” of quality, the quantitative “measure” can be determined. The vice-versa is not possible.

Specialized Logical Class: Class (u, U_e) determined from class (u, U) is-a_o (v, V) : each attributes of class (v, V) is inherited by class (u, U_e) and each attribute of U_e belongs to class (u, U) or to class (v, V) .

Structure for the Preservation of Database Classes/Concepts: Ontology defined at the boundary between *concept theory* and *computer science* enclosing: all and only the *classes/concepts* related to a given conceptual graph G_c ; all and only the *is-a^c* relationships among *classes/concepts*; all and only the incompatible concepts/disjoint classes, which are leaves of the structure.

Query Languages for Graph Databases

D**Kornelije Rabuzin***University of Zagreb, Croatia*

INTRODUCTION

One can find more or less complicated definitions, but one could say that a database is an organized collection of data. It is assumed that a database is stored as files on a computer. Actually, any list of items on a paper could be called a collection of data as well, but today it is assumed that databases are stored digitally.

In order to create a text file, some text processor should be used (for example, Word, Wordpad, Notepad, etc.). In the same way, in order to create a database, some Database Management System (DBMS) should be used. Many systems are available, including Oracle, DB2, SQL Server, PostgreSQL, MySQL, Microsoft Access, Neo4j, MongoDB, etc. The majority of DBMSs are relational in nature. “Relational” means that the main structure that is used to store data is a relation, or as users usually call it, a table. One table contains zero or more rows. The idea of storing data into *relations* (tables) is quite old and relational databases have been used for over 40 years. Relations store data in a compact and organized way and redundancy and anomalies that are caused by redundancy are usually avoided. The relational model was introduced by dr. E. F. T. Codd. in the late 1960s and early 1970s.

In order to use a database management system, one has to be familiar with Structured Query Language (SQL). SQL is a standardized language that is supported by all major Relational DBMS software vendors (it does not belong to any vendor in particular). SQL contains many statements to work with data. The CREATE statement is used to create different objects in the database, as well as the database itself. The INSERT, UPDATE and

DELETE statements are used for data manipulation purposes. In order to retrieve data from one or more tables, SELECT statement is used. Many other, complex statements are available as well, but for our purposes, we can skip the details. Each DBMS has some (graphical) client program (one or more) that can be used to write queries (in SQL) in order to work with the database (for example, phpMyAdmin, Workbench, Navicat, pgAdmin, SQL Server Management Studio, etc.). More on SQL can be found in (Rabuzin, 2011, 2014).

Relational databases are mature technology; many DBMSs are available, SQL is standardized, etc. But in recent years things have changed. Extremely large amounts of unstructured and/or semi structured data come from different sources and relational databases can not cope with such large amounts of data in a satisfactory manner. Furthermore, certain queries require too much time to execute and sometimes such behavior is not acceptable. Because of this new types of databases are being invented (and used) on a daily basis, especially NoSQL technologies, which are becoming more and more important.

The word NoSQL had different meanings in the past (for example, No to SQL), but today people usually mean “Not Only SQL”. When one talks about NoSQL databases, several different types can be distinguished (some authors believe that XML databases should be included as well, as well as some other types, but it doesn’t make much difference):

- Document oriented databases,
- Column oriented databases,
- Key Value databases,
- Graph databases.

Document oriented databases, like MongoDB, do not use tables to store data. Instead, they use collections that store documents. Since one document contains all the data that are relevant, foreign keys are not used (at least not in a form that one is used to in relational databases). Document databases can use references to link between documents, but usually all the referenced data could/should be included within the document. So basically, collections would correspond to tables and documents would correspond to rows. It is important to have in mind that the document schema is flexible and each document within the collection can have different schema.

Column oriented databases use tables as well, but data is organized and stored differently. Unlike relational databases that store together values that belong to a single row (usually several rows are stored within a single unit of storage), column oriented databases store together values (for different rows) that belong to the same column. This is easy to justify. Namely, when one poses a query, in most cases one doesn't need all the columns in the table. Instead, only a few columns are usually selected. In data warehouses dimension tables can have very large number of columns. So queries that are posed on column oriented databases should be faster because less data has to be read from the hard disk drive since column data is stored together (and not values that belong to a single row).

Key Value (KV) databases store values for defined keys. Values can be simple as well as complex. Key value databases may look trivial, but sometimes they can be very useful.

For now it is important to have in mind that graph databases represent an interesting type as well, and because of that they will be discussed later. Namely, the main idea of this article is to present languages for graph databases so they have to be explored more thoroughly.

Although relational databases are the most commonly used, NoSQL databases have great potential. However, one has to have in mind that other database types (<http://db-engines.com/en/>) do exist (object oriented databases, content store,

event store, etc.), but they are not important for this research. So many different database types are available and they can be used for different purposes; one can store company data, images, sound, biometric data, etc. But the type that one plans to use depends on one's needs because all the database types mentioned above are more or less appropriate for different scenarios. For large amounts of interconnected data graph databases could represent the best choice whereas for large amounts of structured data relational databases may still represent the best choice. The idea behind NoSQL databases is not to suppress relational databases; it is just a question of the user's needs. So, relational databases now have an alternative, but which type to use depends on user's needs.

The main idea of this article is to explore several languages for graph databases. One of the languages that is in use is Cypher Query Language (CQL). However, it has some limitations as it will be shown later on. Because of that a new graphical language is proposed (implemented) that should be easier to learn and to use (at least for end-users). To resolve other limitations (recursion, views) that CQL has, it is shown how a logic programming language (Datalog) could be used as well.

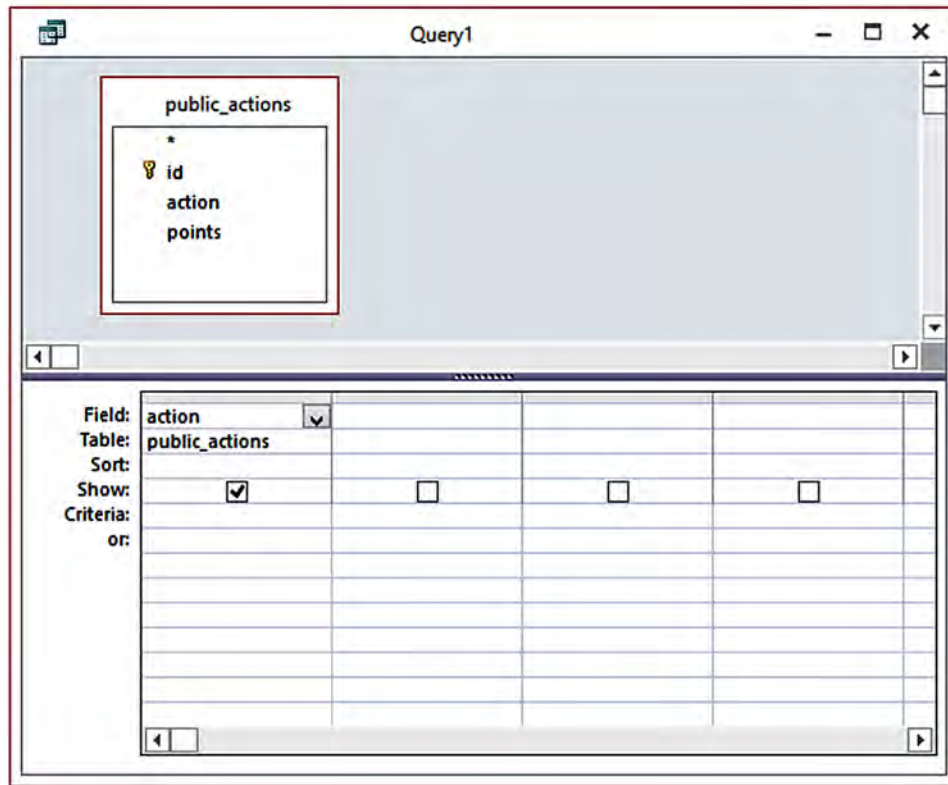
So, the article is organized as follows. First, the background is described. Then CQL is explained. In the next two sections a new language for graph databases is proposed and implemented and then it is shown how a logic programming language could be used to write queries that are not supported in CQL. In the end future research is given and the conclusion is presented. Finally, references are listed.

BACKGROUND

Problems With SQL

When SQL was introduced, it was supposed to be simple. However, it turned out to be quite complex and end-users had problems with more complex queries. Although the idea was that SQL had to be simple so that end users could write the queries

Figure 1. Query by example in Microsoft Access



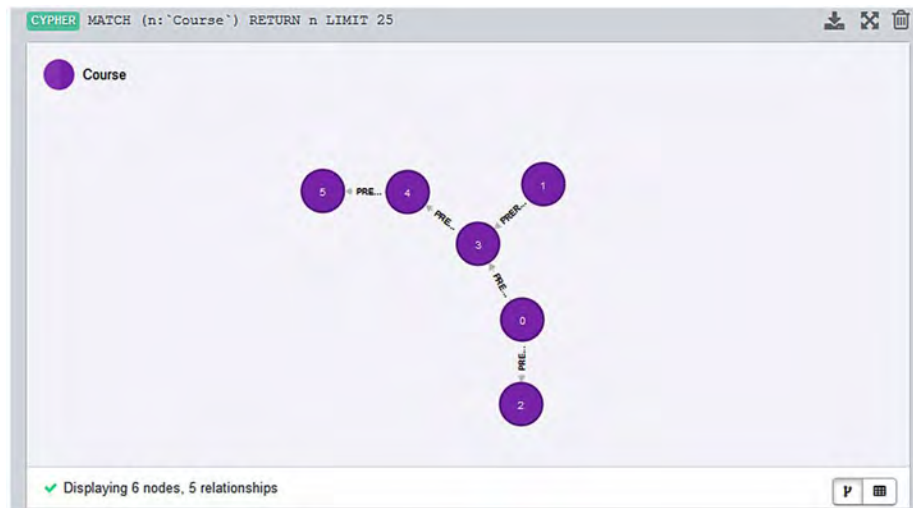
on their own, this was not the case. Namely, over the years several versions of the SQL standard became available (the version from the year 2008 had almost 5000 pages). Because of that even professionals can have problems with queries, and one query that could definitely cause problems to end-users is shown below (end-users are usually not capable of writing such queries):

```
SELECT TOP 100 calls.crm, calls.parent_id, calls.phone_c, izvor.broj_poziva
FROM calls INNER JOIN (SELECT crm, parent_id, phone_c, COUNT(*) broj_poziva FROM calls c GROUP BY crm, parent_id, phone_c) AS izvor ON
(calls.crm=izvor.crm AND calls.parent_id=izvor.parent_id AND calls.phone_c=izvor.phone_c) ORDER BY calls.crm, calls.parent_id, calls.phone_c, broj_poziva DESC
```

Since SQL queries can sometimes get complex, some other languages for relational databases were invented as well (in order to make the process of writing queries easier). One of them is called Query By Example (QBE) and it can be used in Microsoft Access (Figure 1). QBE is a graphical alternative to SQL. It should make it possible for end users to write queries successfully, although they do not have to know every SQL clause in order to write queries manually. This is one of the reasons why Microsoft Access is so popular.

In QBE one selects the attributes that one wants and one can enter certain criteria, specify the sort order and basically, this is enough to produce a query (Figure 1). This QBE idea is very important and it is going to be applied later on. Namely, the plan is to show that queries against graph databases could be built graphically as well. But first one should look at graph databases more thoroughly.

Figure 2. Graph database in Neo4j



Graph Databases

Graph theory is not new. However, such databases could be considered new. One small example of a graph database is given below (Figure 2).

Graph databases use vertices and edges to store information. The number of vertices and edges can be very large (measured in billions). The graph paradigm is a natural and very convenient way for some problem domains. For example, social network analysis is a good example. In order to represent a person and all the friends that the person has, and all the friends of friends (of the person), etc., a graph database is an excellent choice. For example, cities that are connected with roads, train stations, bus stations, etc., these are all scenarios that one could model by means of graph databases. Because of that graph databases are becoming more and more important, and at this point of time they are the database type that gains the most interest. However, for large amounts of structured (company) data relational databases would still represent the best choice. So each database type can be used for some problem domains, but it all depends on one's needs.

In (Robinson, Webber and Eifrem, 2013) one can find an interesting example. The example is about a database whose purpose was to store

information about people and their friends. It is important to say that relational database was used to store the database. Then SQL was used to build queries that should find friends of a person (the first level), then friends of friends were searched (the second level), etc. In the research authors also measured the time that was needed to get results to the queries. It is obvious that the “first level query” was executed almost immediately. The “second level query” took more time, and the “fifth level query” was even not executed in a real time. Since graph databases use edges to connect nodes, such queries are executed in a much faster way because in order to get a result to the query, it is enough to read the database (reading in graph databases is fast and there are no joins included as in SQL, because nodes are directly connected by means of edges).

There are many different graph databases available, including Neo4j, Titan, Giraph, FlockDB, AllegroGraph, ArangoDB, etc. All these systems use certain languages like CQL, Sparql, Gremlin, etc. All these systems have different characteristics (some are distributed, some support different query languages, some can work with extremely large volumes of data, etc.). In this article Neo4j has been used for implementation purposes. Neo4j is the most popular system for

graph databases (<http://neo4j.com/>). It has many users, many downloads, many organizations are implementing their databases in Neo4j, etc. Since Neo4j supports CQL, CQL has been selected to implement a new version of the language that is graphically oriented and that should behave as QBE does to SQL.

To conclude, in the past few years many NoSQL databases have emerged, including graph databases. NoSQL databases have certain advantages and they can be used in certain domains as an alternative to relational databases. In order to use graph databases, one needs to be familiar with specific languages like Cypher Query Language (CQL) or Gremlin. However, some statements in CQL can be considered too complex for end users (it is shown later on). Because of that the main idea of this article is to introduce and implement a new language for graph databases, called Visual Cypher, and to explore whether one of the existing, logic programming languages (Datalog), could be used on graph databases. The first language (Visual Cypher) is new and it has been implemented to pose queries on graph databases visually for the purpose of this article. Since CQL does not support recursion, views, etc., Datalog is used to show how to use recursion and views on a graph database. More information on NoSQL and graph databases could be found in (Buerli, 2012), (Redmond & Wilson, 2012) or (Robinson et al., 2013).

CQL

CQL is a query language for graph databases. CQL has many statements and some of them resemble those in SQL. The CREATE statement is used to create nodes and relationships (DELETE is used to delete them). MATCH is used to specify the graph pattern that should be matched and it is mostly used to get data from the database. WHERE is used to specify conditions and RETURN determines what should be returned. SET is used, for example, to set properties on nodes and relationships. One can

also use operators (comparison, mathematical, Boolean, etc). At the same time CQL uses some clauses that are used in SQL, like ORDER BY, etc.

The CREATE statement that is used to create the database that one can see in the Figure 2 is given below (this statements creates courses and relationships between them):

```
CREATE (p1:Course {name: "Mathematics",
ects: 7}),
(p2:Course {name: "Informatics"}),
(p3:Course {name: "Programming I",
ects: 7}),
(p4:Course {name: "Databases I"}),
(p5:Course {name: "Databases II",
ects: 6}),
(p6:Course {name: "Data warehouses
I", ects: 5}),
(p1)-[:PREREQ]->(p3),
(p1)-[:PREREQ]->(p4),
(p2)-[:PREREQ]->(p4),
(p4)-[:PREREQ]->(p5),
(p5)-[:PREREQ]->(p6)
```

The first part of the CREATE statement creates nodes. For example, p1:Course {name: "Mathematics", ects: 7} creates a node whose label is Course, and the rest of the statement is used to create relationships between nodes. For example, (p1)-[:PREREQ]->(p3) creates a relationship (called PREREQ) between two nodes that are identified by p1 and p3.

In order to list the database content, MATCH and RETURN clauses should be used. This is an equivalent to the SELECT statement in SQL. The following query searches for courses that are connected, i.e., PREREQ relationship exists between them:

```
MATCH (n:Course) -[:PREREQ]->
(m:Course)
RETURN n.name, m.name
n.name                m.name
Mathematics            Programming I
Mathematics            Databases I
```

Informatics Databases I
 Databases I Databases II
 Databases II Data warehouses
 I

One could also use WHERE, ORDER BY and LIMIT statements, but they are used in a pretty much the same way as in SQL. The problem with CQL is that it does not support recursion or views as it will be shown later on. This is why it is shown how to use two other languages to pose queries below. The first language is in fact a new visual language, Visual Cypher. The other is not new (Datalog), but it should enable deduction on graph databases (it will be used to implement recursion and views).

Some papers that are relevant for graph databases and graph query languages are (Cheng et al., 2008), (He & Singh, 2008), (Holzschuher & Peinl, 2013) and (Wood, 2009). However, none of the mentioned papers does include or mention a new language in the way it is proposed in the article. The ideas presented in this article have their roots in the project that explored deductive rules in data warehouses (Rabuzin, 2014a).

SOLUTIONS AND RECOMMENDATIONS

A New Language for Graph Databases: Visual Cypher (VC)

The first problem with CQL is that statements have to be created (entered) manually and sometimes this can cause problems for users. One can also say that another database language has to be learned (again). In order to pose queries on graph databases easily, a graphical interface has been developed. QBE was already mentioned earlier as a graphical alternative to SQL. In this section VC is presented and its purpose is to pose queries on graph databases graphically in the same way as QBE is used for relational databases. To the best of author's knowledge, this hasn't been done before.

Table 1. Neo4j database structure

Label Name	Property Name
Node Labels	
User	Firstname, Lastname
Book	Title, YearPublished
Author	Firstname, Lastname
Genre	Name
Relationship Types	
BORROWED	Connects_User_to_Book
WROTE	Connects_Author_to_Book
PART_OF	Connects_Book_to_Genre

In order to demonstrate how to use the language, here is graph database structure that is used in examples below (nodes and relationships, i.e., vertices and edges are specified):

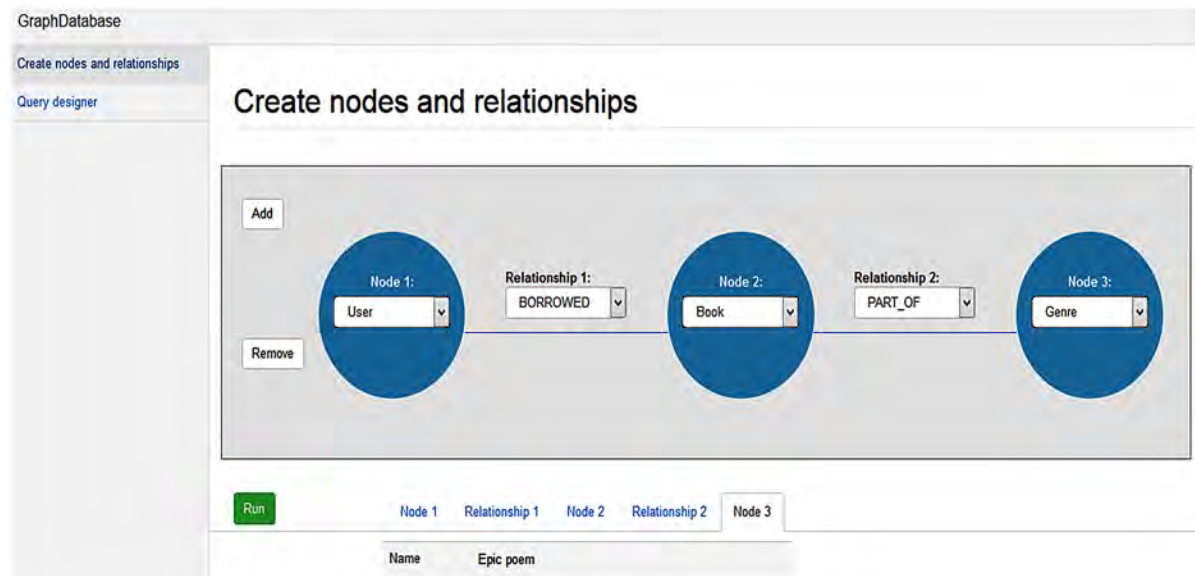
The language is implemented by means of Neo4jClient library for.NET. The interface can be used to work with the developed database, but without the Neo4j console (this console is used to write CQL statements in Neo4j). For now one can create nodes (one or more) as well as relationships and one can also pose queries against the database.

The first example is used to add nodes and relationships (Figure 3). One can add a node, two nodes and a relationship, etc. Since such scenarios are simple, one can see in Figure 3 how to add three nodes and two relationships:

The query behind the Figure 3 is given below (this is the query that the tool creates based on the items that have been selected on the interface, and the query is then executed in Neo4j):

```
Neo4jDb.Instance.Client.Cy-
pher.Match("(u:User), (b:Book),
(g:Genre)")
.Where((Book b) => b.Title==book.
Title)
.AndWhere((Book b) =>
b.YearPublished==book.YearPublished)
.AndWhere((User u) =>
u.Firstname==user.Firstname)
.AndWhere((User u) =>
```

Figure 3. Adding three nodes and two relationships



```
u.Lastname==user.Lastname)
.AndWhere((Genre g) => g.Name==genre.
Name)
.CreateUnique("(u)-[:BORROWED
{DateBorrowed:'' + dateBorrowed +
''}]>(b)-[:PART_OF]>(g)")
.ExecuteWithoutResults();
```

If one would like to add users and relationships manually, by means of CQL, here is exactly what one should write in CQL (users should be added before, in separate queries):

```
MATCH (u:User), (b:Book), (g:Genre)
WHERE b.Title='Judita' AND
b.YearPublished='1521' AND
u.Firstname='Ivan' AND
u.Lastname='Horvat' AND g.Name='Epic
poem'
CREATE UNIQUE (u)-[:BORROWED
{DateBorrowed:'13.1.2015.'}]>(b)-
[:PART_OF]>(g)
```

It is obvious that the developed prototype (interface) represents a much easier way for adding three nodes and two relationships than the CQL

query (the statement is generated automatically while the CQL query has to be entered manually).

Figure 4 shows how the interface for querying data (for a single node) looks like:

Figure 5 shows how to extract data from several nodes:

The method that has to be executed to extract the data is given below (it is used to retrieve data from the database):

```
IEnumerable<Object> result = Neo4jDb.
Instance.Client.Cypher
.Match("(user:User)-[:BORROWED]-
>(book:Book)<-[:WROTE]-
(author:Author)")
.Where((User user) => user.
Firstname==criteria1Prop1Node1)
.OrWhere((User user) => user.
Firstname==criteria2Prop1Node1)
.Return((user, book, r, author)
=> new { User=user.As<User>(),
Book=book.As<Book>(), R=r.
As<BORROWED>(), Author=author.
As<Author>() })
.OrderBy("user.Firstname")
.Limit(Convert.ToInt32(limit))
.Results.ToList<Object>();
```


Figure 4. Extracting data: single node

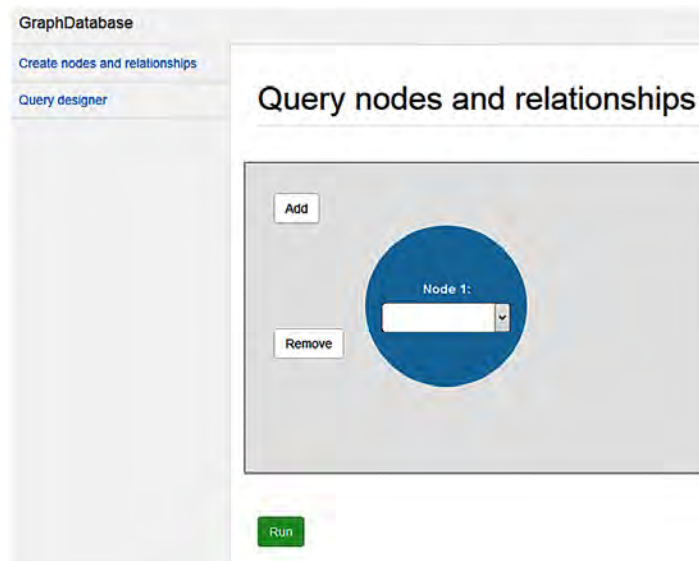
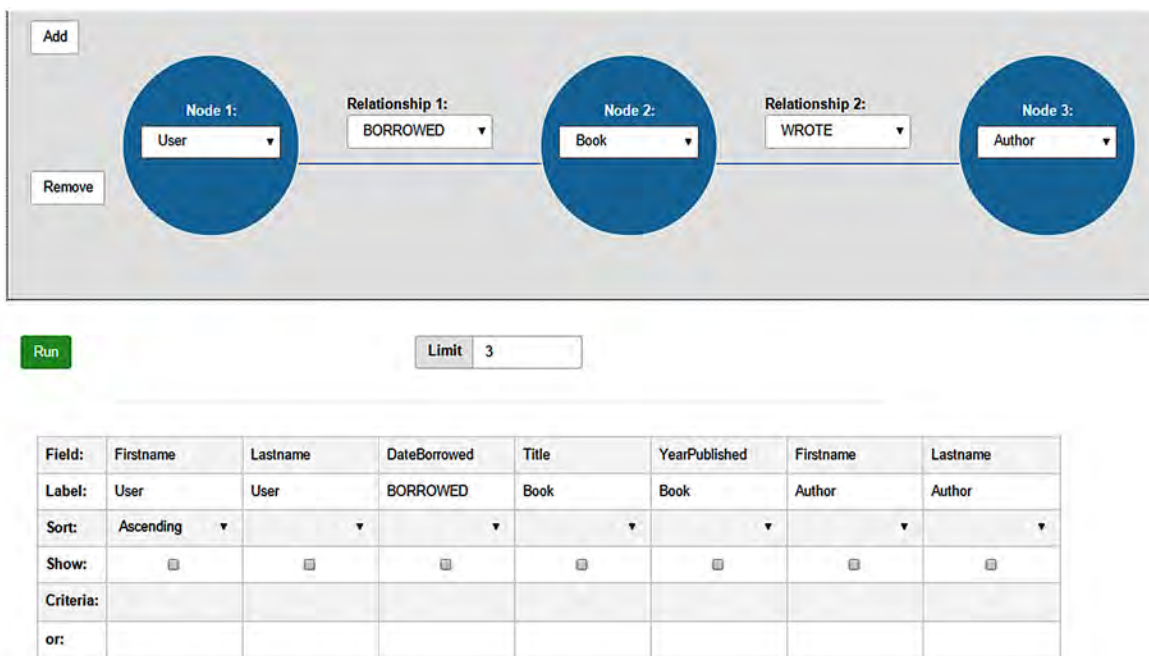


Figure 5. Extracting data: multiple nodes



To conclude, the developed interface can be used to add nodes and relationships as well as to extract data from the database. Other, more simple and complex examples could have been added as well, but the principle would have been the same. The main advantage of the approach is that users do not have to write CQL queries manually.

Instead, the visual interface makes the process of posing queries much easier. However, some other things should be implemented as well, like how to create other types of objects (for now one can create vertices and edges and one can retrieve data from the database), etc.

Datalog for Graph Databases

In this section Datalog is discussed in order to determine whether Datalog could be used to express queries on graph databases. Datalog is an old logic programming language and it is not explained here because people that work with databases are familiar with facts, rules and goals. However, the question is why to use Datalog on graph databases at all.

First of all, CQL does not support recursive queries. When SQL was introduced, it did not support recursive queries either. But some ideas from logic programming turned out to be very useful (recursive queries, for example). So today one can write recursive queries in SQL. CQL is probably on the same path and it is reasonable to expect that it will support recursive queries (some day), but for now it does not.

Second of all, views are not supported in CQL. In Datalog this is easy to achieve. If one had a rule, one could just call it and execute it as a query. In such a way the problem of views (that are not supported in CQL) could be resolved as well.

Furthermore, there are some other things that logic programming languages do support (hypothetical queries, for example), but for now the idea is to show how Datalog could be used to resolve two CQL problems that are mentioned above. In order to resolve the problems Datalog Educational Systems (DES) is used for testing purposes (in this section).

The graph database that was introduced to explain graph databases is used as an example. In order to find courses and their prerequisites in CQL, it is enough to read the graph database. It was noted earlier that MATCH and RETURN clauses (in CQL) should be used to extract data from the database:

```
MATCH (n:Course)-[:PREREQ]-
>(m:Course)
RETURN n.name, m.name
```

To find courses and their prerequisites on the second level, things get a little bit more complex, i.e., one has to add three nodes and two relationships to get the result:

```
MATCH (n:Course)-[:PREREQ]-
>(m:Course)-[:PREREQ]->(o:Course)
RETURN n.name, o.name
```

In order to find courses and their prerequisites (the first two levels), UNION should be used (UNION combines two queries in order to merge their results):

```
MATCH (n:Course)-[:PREREQ]-
>(m:Course)
RETURN n.name AS c, m.name AS p
UNION
MATCH (n:Course)-[:PREREQ]-
>(m:Course)-[:PREREQ]->(o:Course)
RETURN n.name, o.name
```

The same analogy could be applied further on (third level, etc.). But if one was not familiar with the total number of levels, such approach would not represent a good choice as queries would become more and more complex with each level. Imagine the query that finds courses and their prerequisites if one assumed that the total number of levels would be five.

In order to be able to use Datalog, the graph database should be transformed into a set of facts. But one thing is interesting here. Some graph database systems store graph databases in a native (graph) form, whilst some use relational or other types of databases for storage purposes. So, the thing that is crucial here is that relational representation of a graph database exists, in certain cases. This is the answer to the question and the reason why one could use Datalog on graph databases as well. But for the purpose of this example, one could transform nodes and relationships (from the graph database above) into a simple set of facts (only some courses are transformed to make this example small):

```
course('databases I', null).
course('databases II', 6).
course(mathematics, 7).
course(informatics, null).
```

The last course does not have the number of points, so NULL is used. NULL is used when one has to store data, but information is not known or it may be not applicable. The relationships between nodes could be transformed as follows:

```
prereq(informatics, 'databases I').
prereq(mathematics, 'databases I').
prereq('databases I', 'databases II').
prereq('databases II', 'data warehouses I').
```

After the facts are specified, the following rule could be defined:

```
comes_before(X,Y):- prereq(X,Y).
comes_before(X,Y):- prereq(X,Z),
comes_before(Z,Y).
```

This rule should find all the courses and their prerequisites (it is not required to know the number of levels in advance). This should be much easier than the example above (UNION in CQL).

And finally, a view is defined. When called, it lists all the courses that have 7 points. Such courses require special attention (the more points the course has, the more difficult it is):

```
attention(X):-course(X,Y), Y=7.
```

So now one could try to write queries and see their results. For a start, comes_before (X,Y) rule was invoked (Figure 6):

One can see that it is much easier to find courses and their prerequisites in Datalog, then it was in CQL.

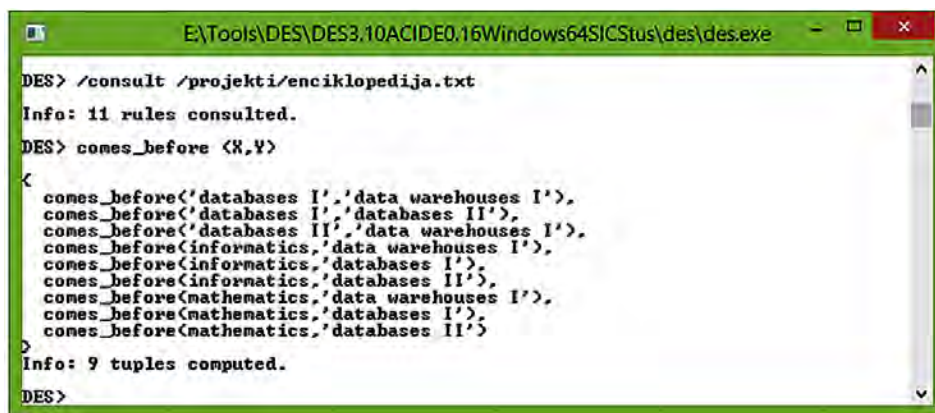
The next query finds the courses that come before the course 'Databases I':

```
DES> prereq(X,'databases I')
{
  prereq(informatics,'databases I'),
  prereq(mathematics,'databases I')
}
Info: 2 tuples computed.
```

In CQL one should write the following query:

```
MATCH (x:Course)-[:PREREQ]->(Course
{name:'Databases I'})
RETURN x.name
In order to find courses that require special attention, one should use:
```

Figure 6. Datalog in action



```
DES> attention(X)
{
  attention(mathematics)
}
Info: 1 tuple computed.
```

In CQL one should write:

```
MATCH (x:Course) WHERE x.ects = 7 RE-
TURN x.name
```

One can see that Datalog queries are less complicated than their equivalents in CQL. This has certain advantages. One of them is that Datalog has been used for a longer period of time and people are familiar with facts and rules. Furthermore, some things that cannot be expressed in CQL can be expressed in Datalog. In this section Datalog was not explained in a high detail, but for more information one could look at (Colomb, 2005) or (Piattini & Diaz, 2000).

FUTURE RESEARCH DIRECTIONS

Three languages for graph databases have been analyzed. CQL is a language that is used in Neo4j and both the language and the system are quite popular. However, people that work with databases do have to learn another new language to use graph databases. Since the number of graph databases as well as the number of languages is very large, the main idea of this research was to invent another, graphically oriented language that resembles to SQL alternatives (like QBE). This new language was introduced (Visual Cypher) and the interface to use it was implemented. As a consequence, queries on graph databases can be posed graphically, in a similar manner as QBE can be used to pose queries on relational databases (graphically).

Currently the prototype is functioning very well, although additional functionality certainly should be implemented. For example, currently one can define/create nodes and relationships, and one can query the nodes and relationships, but one cannot create other objects in the database.

Finally, since CQL and VC do not support recursion and views, it was shown how Datalog could be used to perform recursive queries as well as how it can be used to define views. Datalog is a well-known language that has been used for a very long period of time and people are usually familiar with rules, facts and goals (queries).

In the future one could extend the VC (as described above) and one could certainly explore other graph databases and their languages in the context of this research. For example, Visual Gremlin could be implemented, etc.

CONCLUSION

In this article three languages for graph databases have been discussed. The first language is called Cypher Query Language (CQL) and it can be used in some systems, such as Neo4j. However, it has some limitations; queries have to be specified manually, views are not supported, recursion is not supported, etc. Because of this the second language is proposed and implemented, called Visual Cypher. It is a new, graphically oriented language that makes it possible to add nodes and relationships into the graph database, as well as to extract data from the database. However, some more advanced features should be added as well. The third language (Datalog) is not new and it has been used for many years. However, since Datalog supports recursion and views, it has been shown how to use recursion and views on a relational representation of the graph database.

REFERENCES

- Buerli, M. (2012). *The Current State of Graph Databases*. Technical Report, Department of Computer Science, Cal Poly San Luis Obispo. Retrieved February 4, 2015, from http://www.cs.utexas.edu/~cannata/dbms/Class%20Notes/09%20Graph_Databases_Survey.pdf
- Cheng, J., Ke, Y., & Ng, W. (2008). Efficient Query Processing on Graph Databases. *ACM Transactions on Database Systems*, V, 1–44.
- Colomb, R. M. (2005). *Deductive databases and their applications*. London, UK: Taylor & Francis.
- Date, C. (2004). *An Introduction to Database Systems*. Boston: Addison-Wesley.
- Date, C. (2012). *Database Design and Relational Theory: Normal Forms and All That Jazz*. Sebastopol, CA: O'Reilly Media.
- He, H., & Singh, A. K. (2008). Graphs-at-a-time: Query Language and Access Methods for Graph Databases. *SIGMOD*, 08, 405–418. doi:10.1145/1376616.1376660
- Holzschuher, F., & Peinl, R. (2013). Performance of Graph Query Languages. *Proceedings of the Joint EDBT/ICDT 2013 Workshops*, 195–204. doi:10.1145/2457317.2457351
- Piattini, M., & Diaz, O. (2000). *Advanced Database Technology and Design*. Boston: Artech House.
- Rabuzin, K. (2011). *Uvodu SQL*. Čakovec, Croatia: Zrinski d. d.
- Rabuzin, K. (2014). *SQL – napredne teme*. Čakovec, Croatia: Zrinski d. d.
- Rabuzin, K. (2014a). Deductive data warehouses. *International Journal of Data Warehousing and Mining*, 10(1), 16–31. doi:10.4018/ijdw.2014010102
- Redmond, E., & Wilson, J. R. (2012). *Seven databases in seven weeks*. Dallas, TX: Pragmatic Programmers.
- Robinson, I., Webber, J., & Eifrem, E. (2013). *Graph Databases*. Sebastopol, CA: O'Reilly Media.
- Wood, P. T. (2009). Query Languages for Graph Databases. *Third Alberto Mendelzon International Workshop on Foundations of Data Management*. Retrieved February 4, 2015, from <http://www.dcs.bbk.ac.uk/~ptw/tutorial.pdf>

KEY TERMS AND DEFINITIONS

Cypher Query Language (CQL): In order to use graph databases CQL could be used, but other languages do exist as well.

Datalog: A logic programming language that is important for deductive databases.

Graph Databases: A database type that uses vertices and edges to store information.

NoSQL: New database types that are used as an alternative to the relational data model.

Query By Example (QBE): Visual query language supported in Microsoft Access.

Structured Query Language (SQL): A standardized language for relational databases.

Visual Cypher: A new, proposed graphical language for graph databases.

Schema Evolution in Conventional and Emerging Databases



Zouhaier Brahmia

University of Sfax, Tunisia

Fabio Grandi

University of Bologna, Italy

Barbara Oliboni

University of Verona, Italy

Rafik Bouaziz

University of Sfax, Tunisia

INTRODUCTION

In information systems, not only do data change over time, but also database schemata evolve frequently (Sjøberg, 1993). Changing database schema, which is a common but often troublesome task in database administration, occurs for many reasons such as changes in user requirements, compliance to new regulations, addition of new functionalities, or correction of deficiencies in the current schema. Usually, the database administrator changes a database schema through a schema definition language (e.g., SQL DDL).

Two fundamental aspects are involved by schema change: (i) *semantics of change*, which deals with the effects of the change on the schema itself, in order to maintain schema consistency after schema changes, and (ii) *change propagation*, which deals with the effects of the change on the underlying data, in order to guarantee data consistency with the modified schema.

For most database applications, changing the schema of the database without loss of existing data is a significant challenge: it is usually a time-consuming and error-prone task which must be done carefully. In the literature (Jensen et al., 1998), *schema evolution* has been defined as the modality for the management of schema

changes which relieves database programmers and administrators from this burden, by automatically recovering extant data and possibly adapting them to the new schema. During the last two decades, a lot of theoretical work has been done on schema evolution in conventional databases, within the relational, object-oriented, and XML settings, as described in (Brahmia et al., 2015). In recent years, the schema evolution problem has been addressed also in emerging databases (e.g., temporal, genome, embedded, mobile, and NoSQL databases), as a consequence of the growing use of these new databases by contemporary applications (e.g., digital libraries, mobile applications). However, schema evolution support remains so far almost neglected at the practical level: existing commercial systems (i.e., DBMSs like Oracle and DB2, and schema management tools) only provide a quite limited support. Thus, currently, each database administrator uses ad hoc techniques to manage the evolution of a database schema.

The main goal of this chapter is (i) to present the recent research proposals, not already covered in (Brahmia et al., 2015), that deal with schema evolution in traditional and emerging databases, and (ii) to discuss the recent advances on schema evolution support in mainstream DBMSs. The rest of this chapter is organized as follows. The next

section gives some background on our subject. In “Recent Research on Schema Evolution”, we present an update on recent research proposals concerning schema evolution. “DBMS Support for Schema Evolution” surveys the support of schema evolution in the state of the art of the latest database technology. Finally, future work directions and conclusion are provided.

BACKGROUND

In this section, we illustrate with a simple example the functioning of schema evolution, contrasting it with the lowest level of schema change support that can be embedded in a database, that is the modality of schema modification (Jensen et al., 1998). Assume that we have a relational database that contains only an AUTHOR relation with the attributes ID (primary key), NAME, PHONE, and COUNTRY. The first state of this database is shown in Figure 1.

The catalogues store information on the schema S1 of the AUTHOR relation. The table AUTHOR contains two tuples for two authors. Then consider the following schema changes:

```
ALTER TABLE AUTHOR
DROP COLUMN PHONE;
```

Figure 1.

S1

AUTHOR(ID, NAME, PHONE, COUNTRY)

ID	NAME	PHONE	COUNTRY
1	Aicha	11223344	Tunisia
2	Cristiana	55667788	Italy

Figure 2.

S2

AUTHOR(ID, NAME, COUNTRY, EMAIL)

ID	NAME	COUNTRY	EMAIL
(empty)			

```
ALTER TABLE AUTHOR
ADD COLUMN EMAIL CHAR(30);
```

The schema modification technique allows users to effect changes to the database schema, but neither previous schema nor its underlying data are preserved: the old schema is replaced by the new schema, which is initially empty as data populating the old schema are discarded. The effects in our example would be as shown in Figure 2.

The database designer or administrator must restore information concerning authors Aicha and Cristiana in database state S2 by explicitly inserting them as new tuples, through the following SQL statements (executed within the same transaction containing the schema changes or later):

```
INSERT INTO AUTHOR
VALUES(1, 'Aicha', 'Tunisia', 'aicha@
author.tn');
INSERT INTO AUTHOR
VALUES(2, 'Cristiana', 'Italy',
'cristiana@author.it');
```

Hence, this solution leads to loss of information and obsolescence of applications developed according to the original schema. Although this technique may seem unsuitable, since it is straightforward to implement, it is actually the most widely

Figure 3.

S2			
AUTHOR(<u>ID</u>, NAME, COUNTRY, EMAIL)			

ID	NAME	COUNTRY	EMAIL
1	Aicha	Tunisia	Null
2	Cristiana	Italy	Null

adopted technique for managing schema changes in existing information systems.

With schema evolution, after a schema change the old schema is replaced by the new one and old extant data, which are still conforming to the new schema, are automatically recovered. In our example, the effects would be as shown in Figure 3.

The new email information could then be introduced through the following SQL statements replacing Nulls with correct values:

```
UPDATE AUTHOR
SET EMAIL='aicha@author.tn'
WHERE ID = 1;
UPDATE AUTHOR
SET EMAIL='cristiana@author.it'
WHERE ID = 2;
```

The final state of the database is as shown in Figure 4.

Schema evolution can anyway lead to a partial loss of information. For instance, dropping columns means losing the related data. The change of the type of a column could lead to loss of data of this column if the new type is not compatible with the previous type. Furthermore, schema evolution technique does not guarantee continued functioning of existing applications, since programs that

use dropped columns or columns having changed types become obsolete after a schema change.

RECENT RESEARCH ON SCHEMA EVOLUTION

In recent years, the problems raised by schema evolution in conventional databases (i.e., relational and XML databases) have been further studied, with respect to what we surveyed in (Brahmia et al., 2015). Lately, researchers of the database community have also investigated schema evolution in emerging databases (e.g., mobile, genome, and NoSQL databases). In the two following subsections, we present the main recent contributions concerning schema evolution in conventional and emerging databases, respectively.

Schema Evolution in Conventional Databases

In this subsection, we further classify the research proposals dealing with database schema evolution into two subsections, according to the type of the setting, that is (object-) relational or XML, in which the research work has been done.

Figure 4.

S2			
AUTHOR(<u>ID</u>, NAME, COUNTRY, EMAIL)			

ID	NAME	COUNTRY	EMAIL
1	Aicha	Tunisia	aicha@author.tn
2	Cristiana	Italy	cristiana@author.it

Schema Evolution in (Object-) Relational Databases

In (Desanti et al., 2013), the authors propose an approach for schema evolution, based on a redundancy-free algorithm which removes possible redundancies in a set of schema change operations (i.e., a set of SQL DDL statements) specified on a relational database schema. A schema change operation is a “create”, “drop” or “modify” operation. The proposed approach (i) generates a delta version (i.e., a set of SQL DDL commands) with the difference between two schemas provided by a development team member, and (ii) merges a set of delta versions in one redundancy-free delta, by identifying and removing their redundancies, and executes such a delta, when the database administrator decides to generate a new version of the database schema.

Cleve et al. (2015) propose a systematic approach to understand how developers deal with database schema evolution in practice. It is based on an automatic four-step process in order to derive the global historical schema of an evolving relational database: (i) SQL code extraction, (ii) schema extraction, (iii) schema comparison, and (iv) visualization and exploitation. The produced schema contains all schema components that belonged (in past versions) or belong (in the current version) to the schema of the studied database, annotated with historical metadata (e.g., commit date, committer id).

Recently, Herrmann et al. (2015) propose CoDEL, a relationally complete database evolution language, which is built on the PRISM++ language (Curino et al., 2008a; Curino et al., 2010). CoDEL is intended to be a declarative language that consistently allows the specification of changes to the database schema and/or changes to stored data. Unlike PRISM++, CoDEL supports all possible schema change operations on tables. Relationally complete means that for any (relational) SQL script, which includes a sequence of SQL data definition and data manipulation statements, a semantically equivalent sequence

of CoDEL operations (i.e., SMO (Curino et al., 2008a) operations) exists. For each CoDEL operation, the authors formalize its operational semantics and provide an SQL-like syntax.

Schema Evolution in XML Databases

In (Klímek et al., 2012), the authors propose eXolutio, a tool which uses the principles of “Model Driven Architecture” (MDA) (Miller and Mukerji 2003) to design and evolve an XML Schema. It allows a designer to specify a “Platform Independent Model” (PIM) schema and multiple “Platform Specific Model” (PSM) schemata, each representing a part of a PIM schema, with a mapping from PSM schemata to the PIM one. eXolutio allows also the designer to evolve the set of specified schemata consistently, thanks to mappings between the two levels and propagation of schema changes made in one level to all the affected schemata.

As for updating XML documents, recently Cavalieri et al. (2013) proposed an algorithm for synthesising the changes between two XML documents as “Pending Update Lists” (PULs), that is unordered lists of update operations. If changes are represented as XQuery Update Facility (W3C, 2011) expressions, the evaluation of each expression results in a single PUL. Two phases are identified in the evaluation of an XQuery Update Facility expression: (i) production, in which the expression is evaluated producing a PUL, and (ii) application, in which the update operations in the produced PUL are applied.

In (Nösinger et al., 2013), the CoDEX approach of Klettke (2007) has been extended with “Evolution Language for XML-Schema” (ELaX), a domain-specific language that proposes to handle modifications on an XML Schema and to express such modifications formally. ELaX allows (i) specifying complex modifications by combining the add, delete, and update operations, and (ii) logging modifications on schema, which is an essential step for the adaptations of XML documents associated to XML Schema. Since an

XML Schema designer using the ELaX language could store a sequence of ELaX operations in a log for a long time period and execute it when required, Nösinger et al. (2014) propose a rule-based optimization algorithm “Rule-based Optimizer for ELaX” (ROfEL), which allows reducing the number of the logged operations before submitting them to be executed by (i) merging some of them, (ii) deleting unnecessary or redundant operations, and/or (iii) correcting, if possible, or removing invalid operations. Notice that the ROfEL algorithm applies a set of heuristic rules in order to identify unnecessary and redundant operations.

Amavi et al. (2014) propose a set of tools for conservative XML schema evolution and compatible XML document adaptation. They propose (i) an algorithm, which generates mappings that allow transforming an original schema into a conservative extension of it (and vice versa), via composition and inversion, and (ii) a method for translating each XML document which is valid to an original schema into a document which is valid to the extended schema.

Schema Evolution in Emerging Databases

Since contemporary applications, like multimedia, geographical information systems, digital libraries, and mobile applications, require new database models with new query languages for managing new types of data, new types of databases have emerged in the last few years. Among these emerging database paradigms, we could cite as examples multimedia, spatial, temporal, stream, biological/genome, embedded, mobile, grid, Big Data, and NoSQL/NewSQL databases. In the following paragraphs, we present contributions dealing with schema evolution in stream, embedded, and NoSQL databases.

Terwilliger et al. (2010) deal with schema evolution support in data stream management systems. They (i) propose to introduce a new element into data streams, called an accent, which announces evolutions to operators in a standing query, and

(ii) define a set of six common stream query operators (Select, Project, Union, Join, Window, and Aggregate) that can process data and schema evolution accents, considering three evolution primitives: “Add Attribute, Drop Attribute, and Alter Data”. So, the proposed framework allows data stream systems to support schema evolution without interrupting the evaluation of standing queries (i.e., queries respond automatically to evolution), whenever the evolution do not affect the semantics of a query.

Wu and Neamtiu (2011) propose an approach for automatically (i) extracting embedded database schemata from regular applications (e.g., written in C or C++), and (ii) detecting changes of these schemata as applications evolve. In fact, they present a tool, called “Schema extraCtion and eVolution analysis for embedded Databases” (SCVD), which shows the feasibility of their approach in the following way: given a set of releases of an application, SCVD automatically retrieves the source code for all these releases, extracts the database schemata embedded in each version of the application code, compares the extracted schema versions, and produces the schema evolution results organized in a useful manner.

In (Scherzinger et al., 2013), the authors aim at providing a generic interface for managing schema evolution in NoSQL data stores (Cattell, 2010). To that end, they propose (i) a declarative NoSQL schema evolution language (i.e., a set of basic common practical schema change operations that deal with most of the typical cases which are discussed in developer forums), and (ii) a generic and abstract NoSQL database programming language for accessing NoSQL data stores. The second language distinguishes the current state of the data store from the state of the objects available in the application space, and implements the proposed set of schema evolution operations. Through such an implementation, the authors show that the schema evolution process could be handled not only eagerly (i.e., executing the specified schema evolution operations directly on all involved entities), but also lazily (i.e., executing

the specified schema evolution operations on the concerned entities the next time they are fetched into the application space).

In (Scherzinger et al., 2015a), the authors propose Cleager, a web-based system prototype that allows eager schema evolution in schema-less NoSQL document stores (e.g., MongoDB, Google Cloud Datastore) by supporting the declarative NoSQL schema evolution language presented in (Scherzinger et al., 2013). Cleager (i) provides a web-based console for specifying eager schema modification operations, using the Cleager language (Scherzinger et al., 2013), and (ii) translates the specified operations and executes them as MapReduce jobs in the Google Cloud Platform.

In (Scherzinger et al., 2015b), the authors present ControVol, an Eclipse plugin which helps developers in resolving problems related to schema evolution during the process of development of Big Data applications on top of schema-less NoSQL data stores. Indeed, at development time, ControVol (i) detects changes (e.g., addition, renaming, or deletion of attributes) to the schema, in the application code, which are not compatible with legacy data; (ii) reports warnings to developers; (iii) proposes automatic and quick fixes to resolve the warnings. Consequently, ControVol guarantees, in the production environment, that not only legacy data could be compatible with new structures utilized in the new application code version, but also the new application version could continue running on legacy data.

Recently, Saur et al. (2015) propose “Key-Value store evolution” (KVolve), an approach and a tool for evolving a key-value NoSQL database without downtime. KVolve is implemented on top of a Redis (Redis Labs, 2015) NoSQL database which stores JSON (JSON, 2015) objects. It provides a domain-specific language to developers, which allows them writing changes to data formats (i.e., JSON objects) and specifying how legacy data are transformed to a new data format version. Such changes are applied lazily to the NoSQL database, in order to avoid downtime and database locking; objects that are stored according to a pre-

vious format are migrated on-the-fly to the new format only when accessed by applications. The developed tool supports changes to the fields of a JSON object (addition of new fields, and deletion or modification of existing fields) and also to the names of the keys.

DBMS SUPPORT FOR SCHEMA EVOLUTION

As reported in detail in (Brahmia et al., 2015), several survey works considered and comparatively analysed schema evolution support in traditional databases (Türker, 2000; Colazzo et al., 2010; Hartung et al., 2011; Curino et al., 2013). Such studies involved recent releases of all major commercial (object-)relational and XML DBMSs (including Oracle, Microsoft SQL Server, IBM DB2, MySQL, PostgreSQL, Sybase, Ingres, and Tamino), and change management tools (including Altova, IDERA, Embarcadero, RedGate, DTM DB, SwisSQL, and Liquibase).

Overall, these works show that the existing conventional DBMSs (i) support schema evolution partially, by providing only a limited set of schema changes, and (ii) do not use a common language to evolve schema. Besides, they show also that the support of XML schema updates (i) is either absent in some commercial DBMSs (like Microsoft SQL Server) and XML tools (like OXygen’s XML Editor or Stylus Studio) or (ii) limited in some other commercial DBMSs (like Oracle 11g, Tamino, or DB2 v.9) and XML tools (like Altova’s XMLSpy and DiffDog).

As for emerging DBMSs, Scherzinger et al. (2013) study, among other things, schema evolution support in state-of-the-art NoSQL data stores (Redis (Redis Labs, 2015) and Riak (Basho Technologies, 2015) as key-value stores, MongoDB (MongoDB, 2015) and Couchbase (Brown, 2013) as document stores, and BigTable (Chang et al., 2008), HBase (Apache, 2015a), App Engine Datastore (Google, 2015) and Cassandra (Apache, 2015b) as extensible record stores) and

present the results that follows. First, most NoSQL database systems do not provide a schema definition language (exception for some systems like Cassandra which supports a “CREATE TABLE” statement). Second, all existing NoSQL systems do not provide support for migrating legacy entities to a new schema specification and delegate such a task to the application logic layer. Third, in all available NoSQL data store systems, there is a lack of a schema evolution interface that helps in the maintenance of entities’ structures; currently, all schema changes are programmed by developers. Fourth, only Cassandra supports altering schema explicitly through its “ALTER TABLE” statement. So, it is not considered as a schema-less or schema-flexible NoSQL system. Fifth, NULL values, which result from some schema evolution operations (like adding a new field or column), are managed differently by NoSQL systems.

Moreover, in (Störl et al., 2015), the authors survey the state-of-the-art of Object-NoSQL Mappers, which are currently used by application developers to deal with schema evolution problems in schema-less NoSQL data stores, since available NoSQL systems do not provide any support for evolving schema of these data stores. Notice here that the lack of such a support could be due to the nature of these databases themselves which are designed to be schema-less or schema flexible.

FUTURE RESEARCH DIRECTIONS

There are several open research issues for schema evolution. Some of them are as follows:

1. Schema evolution is poorly supported in commercial DBMSs and schema management tools, as shown in the previous section. Thus, this aspect needs further efforts in order to provide more efficient practical solutions for database designers and administrators. Moreover, schema evolution support is also limited in the standard SQL language. Since relational databases continue to be the most

used ones around the world and since schema changes are unavoidable in information systems, SQL should be extended in order to provide more consistent help in managing evolving database schemata (e.g., new schema change operators, new features for graceful application of change propagation, etc.).

2. Nowadays, many enterprises use Web and distributed databases and, thus, schema changes have to be effected more frequently, since user requirements, laws and technologies rapidly change in this context. Therefore, it will be very interesting and useful to study schema evolution in Web and distributed environments. Among the first works in this area, we find (Curino et al., 2008b), which proposes a workbench for schema evolution in Web information system (after studying the Wikipedia database evolution during 4.5 years), and (Rae et al., 2013), which describes techniques used by Google’s F1 database management system (Shute et al., 2013) to support schema evolution in a globally distributed environment.
3. The database provenance is very useful in many modern information systems (like big science projects), in order to explain the lineage of the current data. This explanation often requires “flashing back” to the original schema. Thus, supporting database provenance under schema evolution could be an interesting issue. A first work in this area is (Gao & Zaniolo, 2012), which presents the main features of the Archived Metadata and Provenance Manager (AM&PM), which is the first system designed to support the provenance of both data and meta-data in the presence of schema evolution.

CONCLUSION

In this chapter, we studied recent advances on schema evolution in traditional and emerging

databases, and explored its support in commercial database systems and development tools.

Although a lot of work has been done at theoretical level and efficient solutions have been proposed by researchers, there is no efficient practical support for schema evolution at the state-of-the-art of database technology. Manual efforts are still required to successfully manage the schema evolution process: the database designer or administrator has to manually convert data defined under the previous schema in order to be valid with respect to the new schema, and the application developer(s) must carefully revise existing applications to adapt to the new schema.

REFERENCES

- Amavi, J., Chabin, J., Ferrari, M. H., & Réty, P. (2014). A ToolBox for Conservative XML Schema Evolution and Document Adaptation. In *Proceedings of the 25th International Conference on Database and Expert Systems Applications (DEXA 2014)*, (Part I, pp. 299-307). doi:10.1007/978-3-319-10073-9_24
- Apache. (2015a). *Apache HBase*. Retrieved from <http://hbase.apache.org/>
- Apache. (2015b). *Apache Cassandra*. Retrieved from <http://cassandra.apache.org/>
- Basho Technologies. (2015). *Riak*. Retrieved from <http://docs.basho.com/riak/latest/>
- Brahmia, Z., Grandi, F., Oliboni, B., & Bouaziz, R. (2015). Schema Evolution. In M. Khosrow-Pour (Ed.), *Encyclopedia of Information Science and Technology* (3rd ed.; pp. 7641-7650). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-5888-2.ch753
- Brown, M. (2013). *Developing with Couchbase Server*. O'Reilly Media, Inc.
- Cattell, R. (2010). Scalable SQL and NoSQL Data Stores. *SIGMOD Record*, 39(4), 12-27. doi:10.1145/1978915.1978919
- Cavalieri, F., Solimando, A., & Guerri, G. (2013). Synthetising Changes in XML Documents as PULs. *Proceedings of the VLDB Endowment*, 6(13), 1630-1641. doi:10.14778/2536258.2536273
- Chang, F., Dean, J., Ghemawat, S., Hsieh, W. C., Wallach, D. A., Burrows, M., Chandra, T., Fikes, A., & Gruber, R. E. (2008). Bigtable: A Distributed Storage System for Structured Data. *ACM Transactions on Computer Systems*, 26(2), Article 4.
- Cleve, A., Gobert, M., Meurice, L., Maes, J., & Weber, J. (2015). Understanding database schema evolution: A case study. *Science of Computer Programming*, 97(Part 1), 113-121. doi:10.1016/j.scico.2013.11.025
- Colazzo, D., Guerrini, G., Mesiti, M., Oliboni, B., & Waller, E. (2010). Document and Schema XML Updates. In C. Li & T. W. Ling (Eds.), *Advanced Applications and Structures in XML Processing: Label Stream, Semantics Utilization and Data Query Technologies* (pp. 361-384). Hershey, PA: Information Science Reference. doi:10.4018/978-1-61520-727-5.ch016
- Curino, C., Moon, H. J., Deutsch, A., & Zaniolo, C. (2013). Automating the database schema evolution process. *The VLDB Journal*, 22(1), 73-98. doi:10.1007/s00778-012-0302-x
- Curino, C. A., Moon, H. J., Deutsch, A., & Zaniolo, C. (2010). Update Rewriting and Integrity Constraint Maintenance in a Schema Evolution Support System: PRISM++. *Proceedings of the VLDB Endowment*, 4(2), 117-128. doi:10.14778/1921071.1921078
- Curino, C. A., Moon, H. J., Tanca, L., & Zaniolo, C. (2008b). Schema Evolution in Wikipedia: toward a Web Information System Benchmark. In *Proceedings of the 10th International Conference on Enterprise Information Systems (ICEIS 2008)*, (Volume DISI, pp. 323-332).

- Curino, C. A., Moon, H. J., & Zaniolo, C. (2008a). Graceful database schema evolution: the PRISM workbench. In *Proceedings of the 34th International Conference on Very Large Data Bases (VLDB 2008)*, (pp. 761-772). doi:10.14778/1453856.1453939
- Desanti, C., Bernardelli, M., Fuckner, M., & Stasiu, R. K. (2013). A Redundancy-Free Approach to Schema Evolution. In *Proceedings of the 28th Brazilian Symposium on Databases (SBBD 2013)*. Retrieved from <http://sbbd2013.cin.ufpe.br/Proceedings/artigos/pdfs/sbbd_shp_07.pdf>
- Gao, S., & Zaniolo, C. (2012). Supporting Database Provenance under Schema Evolution. In *Proceedings of the 5th International ER Workshop on Evolution and Change in Data Management and on Non Conventional Data Access (ECDM – NoCoDa 2012)*, (pp. 67-77).
- Google. (2015). *App Engine Datastore*. Retrieved from <<https://cloud.google.com/appengine/docs/java/datastore/>>
- Hartung, M., Terwilliger, J., & Rahm, E. (2011). Recent advances in schema and ontology evolution. In Z. Bellahsene, A. Bonifati, & E. Rahm (Eds.), *Schema Matching and Mapping* (pp. 149–190). Berlin, Germany: Springer-Verlag. doi:10.1007/978-3-642-16518-4_6
- Herrmann, K., Voigt, H., Behrend, A., & Lehner, W. (2015). CoDEL - A Relationally Complete Language for Database Evolution. In *Proceedings of the 19th East European Conference on Advances in Databases and Information Systems (ADBIS 2015)*, (pp. 63-76). doi:10.1007/978-3-319-23135-8_5
- JSON. (2015). *Introducing JSON*. Retrieved from <http://www.json.org/>
- Klettke, M. (2007). Conceptual XML Schema Evolution - the CoDEX Approach for Design and Redesign. In *Workshop Proceedings of Datenbanksysteme in Business, Technologie und Web (BTW 2007)*, (pp. 53-63).
- Klímek, J., Malý, J., Mlýnková, I., & Nečaský, M. (2012). eXolutio: Tool for XML Schema and Data Management. In *Proceedings of the 12th Annual International Workshop on Databases, TExtS, Specifications and Objects (DATESO 2012)*, (pp. 69-80).
- Miller, J., & Mukerji, J. (2003). *MDA Guide Version 1.0.1*. Object Management Group.
- Mongo, D. B. (2015). *MongoDB*. Retrieved from <https://www.mongodb.org/>
- Nösinger, T., Klettke, M., & Heuer, A. (2013). XML Schema Transformations - The ELAX Approach. In *Proceedings of the 24th International Conference Database and Expert Systems Applications (DEXA 2013)*, (Part I, pp. 293-302).
- Nösinger, T., Klettke, M., & Heuer, A. (2014). Optimization of Sequences of XML Schema Modifications - The ROFEL Approach. In *Proceedings of the 26th GI-Workshop on Foundations of Databases (Grundlagen von Datenbanken 2014)*, (pp. 11-16).
- Rae, I., Rollins, E., Shute, J., Sodhi, S., & Vingralek, R. (2013). Online, Asynchronous Schema Change in F1. *Proceedings of the VLDB Endowment*, 6(11), 1045–1056. doi:10.14778/2536222.2536230
- Redis Labs. (2015). *Redis*. Retrieved from <http://redis.io/>
- Saur, K., Dumitras, T., & Hicks, M. (2015). *Evolving NoSQL Databases Without Downtime*. Submitted for publication (CoRR abs/1506.08800). Retrieved from <<http://arxiv.org/pdf/1506.08800v2.pdf>>
- Scherzinger, S., Cerqueus, T., & Cunha de Almeida, E. (2015b). ControVol: A Framework for Controlled Schema Evolution in NoSQL Application Development. In *Proceedings of the 31st IEEE International Conference on Data Engineering (ICDE 2015)*, (pp. 1464-1467). doi:10.1109/ICDE.2015.7113402

Scherzinger, S., Klettke, M., & Störl, U. (2013). Managing Schema Evolution in NoSQL Data Stores. In *Proceedings of the 14th International Symposium on Database Programming Languages (DBPL 2013)*. Retrieved from <http://arxiv.org/pdf/1308.0514v1.pdf>

Scherzinger, S., Klettke, M., & Störl, U. (2015a). Cleager: Eager Schema Evolution in NoSQL Document Stores. In *Proceedings of the 16th Conference on Database Systems for Business, Technology, and Web (BTW 2015)*.

Shute, J., Vingralek, R., Samwel, B., Handy, B., Whipkey, C., Rollins, E., & Apte, H. et al. (2013). F1: A Distributed SQL Database That Scales. *Proceedings of the VLDB Endowment*, 6(11), 1068–1079. doi:10.14778/2536222.2536232

Sjøberg, D. (1993). Quantifying Schema Evolution. *Information and Software Technology*, 35(1), 35–44. doi:10.1016/0950-5849(93)90027-Z

Störl, U., Hauf, T., Klettke, M., & Scherzinger, S. (2015). Schemaless NoSQL Data Stores - Object-NoSQL Mappers to the Rescue? In *Proceedings of the 16th Conference on Database Systems for Business, Technology, and Web (BTW 2015)*.

Terwilliger, J. F., Fernández-Moctezuma, R., Delcambre, L. M. L., & Maier, D. (2010). Support for schema evolution in data stream management systems. *Journal of Universal Computer Science*, 16(20), 3073–3101.

Türker, C. (2000). Schema Evolution in SQL-99 and Commercial (Object-)Relational DBMS. In H. Balsters, B. de Brock, & S. Conrad, (Eds.), *Database Schema Evolution and Meta-Modeling—Proceedings of the 9th International Workshop on Foundations of Models and Languages for Data and Objects (FoMLaDO/DEMM 2000)*.

W3C. (2011). *XQuery Update Facility 1.0 - W3C Recommendation*. Retrieved from <<http://www.w3.org/TR/2011/REC-xquery-update-10-20110317/>>

Wu, S., & Neamtiu, I. (2011). Schema evolution analysis for embedded databases. In *Proceedings of the 3rd ICDE Workshop on Hot Topics in Software Upgrades (HotSWUp 2011)*, (pp. 151-156).

ADDITIONAL READING

Deufemia, V., Polese, G., & Vacca, M. (2009). Evolutionary Database: State of the Art and Issues. In V. E. Ferragline, J. H. Doorn, & L. C. Rivero (Eds.), *Handbook of Research on Innovations in Database Technologies and Applications: Current and Future Trends* (pp. 102–109). Hershey, PA: Information Science Reference. doi:10.4018/978-1-60566-242-8.ch012

Eder, J., & Wiggisser, K. (2009). Data Warehouse Maintenance, Evolution and Versioning. In L. Liu & M. T. Özsu (Eds.), *Encyclopedia of Database Systems* (pp. 664–669). Springer, US.

Ram, S., & Shankaranarayanan, G. (2003). *Research issues in database schema evolution: the road not taken. Technical Report #2003-15*. Tucson, AZ, USA: The University of Arizona.

Roddick, J. F. (2009). Schema Evolution. In L. Liu & M. T. Özsu (Eds.), *Encyclopedia of Database Systems* (pp. 2479–2481). Springer, US.

KEY TERMS AND DEFINITIONS

Database Conversion/Migration: Process consisting in adapting data valid under a given schema version to another schema version.

Database Schema: The formal structure of the database, described by means of a data definition language.

Schema Change Propagation: The effects of a schema change at instance level, involving suitable conversions necessary to adapt extant data to the new schema.

Schema Change Semantics: The effects of a schema change at schema level.

Schema Changes: Operations performed on a (populated) database schema, in order to adapt it to new application requirements.

Schema Evolution: A schema change modality that avoids the loss of extant data. No support is required for previous schemata.

Temporal Database: A database with built-in support for managing time-varying data.

D

Schema Versioning in Conventional and Emerging Databases

Zouhaier Brahmia

University of Sfax, Tunisia

Fabio Grandi

University of Bologna, Italy

Barbara Oliboni

University of Verona, Italy

Rafik Bouaziz

University of Sfax, Tunisia

INTRODUCTION

Persistent information and data-centric systems use databases to store data. The contents of a database must adhere to a formal structure that is fixed in advance, and is called the schema of the database (Date, 2003). In those systems, not only data changes are obvious tasks done almost every day but also schema changes are unavoidable, in order to reflect a change in the real world or in the user's requirements, to correct mistakes in the initial design, to migrate to a new platform or to allow the expansion of the application scope over time. Two main problems have to be considered when dealing with any schema change: *semantics of change* (i.e., the effects of the change on the schema itself) and *change propagation* (i.e., the effects of the change on the underlying data). Resolving the former guarantees schema consistency, while resolving the latter guarantees consistency of data with respect to the changed schema.

In the literature, *schema evolution* and *schema versioning* (Roddick, 1995; Jensen et al., 1998) are the two techniques that were proposed to support schema changes in a DBMS, without loss of extant data and with continued support of legacy applications. After applying schema changes,

schema evolution keeps only the current schema version and retains the data which are adapted to such a schema. On the other hand, each time schema changes are applied, schema versioning creates a new schema version, while preserving old schema versions and their corresponding data. With schema versioning, data access through any schema version is supported, which avoids applications developed with past schemata to become obsolete.

Schema versioning has been widely investigated, both in the context of traditional and temporal database research. Several models, languages and approaches, dealing with schema versioning, have been proposed during the two last decades in the relational, object-oriented, and XML databases. However, to the best of our knowledge, limited support of schema changes and no support of schema versioning is provided by commercial database management systems (DBMS). Therefore, diligent database designers and administrators have to work hard to solve the problem of evolving a database schema in an ad hoc manner. Besides, with the growing use of emerging databases (e.g., multimedia, temporal, biological, and NoSQL databases), research work has also recently done on the problems of schema versioning in such settings.

Figure 1.

S1			
AUTHOR(<u>ID</u>, NAME, PHONE, COUNTRY)			
ID	NAME	PHONE	COUNTRY
1	Aicha	11223344	Tunisia
2	Cristiana	55667788	Italy

The main goal of this chapter is (i) to present the recent research proposals, not already covered in (Brahmia et al., 2015), that deal with schema versioning, and (ii) to discuss the recent advances on schema versioning support in mainstream DBMSs. In particular, the next section gives some basic definitions related to the considered subject. In “Recent Research in Schema Versioning”, we present an update on recent research proposals on schema versioning. “DBMS Support for Schema Versioning” surveys the support of schema versioning in the state of the art of the latest database technology. Finally, future work directions and conclusion are provided.

BACKGROUND

The schema versioning technique allows changes to the database schema with continued support of previous schemata and their corresponding data, which are retained without any change. The newly created schema version is (usually) used to accommodate new data insertions, modifications and deletions. This technique neither leads to loss of information nor to obsolescence of existing applications, as they can still work with old schema versions. Further issues related to schema versioning support have been discussed in (Brahmia et al., 2015).

In this section, we illustrate the functioning of schema versioning with a simple example. Let us assume that we have a relational database that contains only an AUTHOR relation with the attributes ID (primary key), NAME, PHONE, and COUNTRY. The first state of this database is shown in Figure 1.

The catalogues store information on the schema S1 of the AUTHOR relation. The table AUTHOR contains two tuples for two authors. Then consider the following schema changes:

```
ALTER TABLE AUTHOR
DROP COLUMN PHONE;
ALTER TABLE AUTHOR
ADD COLUMN EMAIL CHAR(30);
```

In this case, the information related to unchanged attributes (i.e., ID, NAME and COUNTRY) are automatically recovered after executing schema changes. The effects of such schema changes are shown in Figure 2.

In particular, we obtain two schema versions, S1 and S2. The new email information could then be introduced through the following SQL statements:

```
UPDATE AUTHOR
SET EMAIL='aicha@author.tn'
WHERE ID = 1;
UPDATE AUTHOR
SET EMAIL='cristiana@author.it'
WHERE ID = 2;
```

Notice that here, without indicating the schema version under which emails should be introduced, the system is able to automatically determine that these new values have to be inserted in the table corresponding to the second version S2 of the AUTHOR relation.

Moreover, since schema versioning supports all schema versions and their underlying data, it (potentially) allows accessing these data, both retrospectively and prospectively, through user

are working, and the physical facet, i.e., the set of class definitions associated to the schema version, which is generated by the DBMS.

Lee et al. (2006) propose a schema version model, titled the extended “Rich Base Schema” (RiBS) model, for restructuring complex object hierarchies. A complex object hierarchy is a set of classes linked between them with reference relationships (which are based on object identifiers), such that one class of them is considered as the root class of the hierarchy and each one of the other classes as a part class of it. In this model, an attribute in a class could be either a leaf attribute, whose domain is a built-in type (e.g., integer, string), or a non-leaf attribute, whose domain is a user-defined class. The authors of the extended RiBS model focus on three aspects: schema evolution operations, query processing, and schema version merging.

Piccioni et al. (2013) propose an IDE-based approach for managing class versioning. More precisely, the authors present a new formal model for class attribute changes, four empirical studies to evaluate the proposed model, a measure of class evolution robustness, and a prototype tool for schema versioning. As for the proposed formal model, the authors deal with schema change operations, class transformations, and object transformers. As for empirical studies, the authors evaluate their model by analyzing four sets of classes written in Java and Eiffel languages while focusing on persistence-related code changes. As for the measure of class evolution robustness, the authors propose “P-Evolution-Robustness” (PER) to measure the robustness of an object-oriented software system with regard to class evolution. PER is defined as a ratio calculated on the basis of transformation functions between class versions. It is applied to all the classes in the “java.util” package. As for the prototype tool, the authors introduce the ESCHER system which implements the proposed model and shows its feasibility. It is integrated into the EiffelStudio IDE, but can also be used as an independent library.

Schema Versioning in XML Databases

D

Baqasah et al. (2013) propose XS-Diff, an algorithm for detecting changes between two XML Schema (or XSD) versions using relational databases. Indeed, XS-Diff starts by parsing and storing the XSD trees corresponding to the two XML Schema versions in relational tables. After that, it applies some specific SQL queries in order to match the identical nodes and store them in temporary tables. Finally, XS-Diff uses the information stored on these tables to generate the set of changes between the two versions and store them in delta tables.

Baqasah et al. (2014) propose an approach for merging XML Schema versions (i.e., XSD files), through the combination of three-way and operational-transformation approaches. Indeed, the authors (i) define a set of rules allowing an automatic three-way XML Schema version merging and conflict resolution, (ii) use transformation functions, in the merging algorithm, for correcting the paths of XSD nodes to be merged, and (iii) develop a merge tool prototype that shows the feasibility and the correctness of the proposed approach.

Malý et al. (2011) extend the framework presented in (Nečaský and Mlýnková 2009), for conceptually modeling XML data based on “Model-Driven Architecture” (MDA) (Miller and Mukerji 2003), to support schema versioning. More precisely, the authors deal with schema change propagation and propose an approach for the revalidation of existing XML documents against any newly created XML schema version. In fact, when a new XML schema version is added, a new version of each XML document, which is valid with respect to the previous schema version, must be correctly generated according to that new schema version. The proposed approach is based on an algorithm which compares the current XML schema version and the new one and returns a revalidation script in XSLT (W3C, 2007). Such a script is applied to the set of XML documents that

are valid with respect to the previous (logically) “modified” schema version in order to produce a set of new versions of the same XML documents, which are valid with respect to the new schema version.

Currim et al. (2009) provide an in-depth and long study on all aspects related to temporal instance versioning and temporal schema versioning in the τ XSchema framework, while presenting all details of the τ XSchema language (including the schema of an annotation document, the schema of a temporal schema document and the schema of a temporal document). Currim et al. (2012) extend such work by focusing on extending conventional XML Schema integrity constraints with temporal semantics, in the τ XSchema framework (Snodgrass et al., 2008) supporting schema versioning. They start by presenting the four types of conventional constraints that are provided by the XML Schema language: datatype, cardinality, identity, and referential integrity constraints. After that, they propose three classes of temporal constraints: (i) sequenced constraints, which are applicable at each point in time, (ii) current constraints, which represent a special kind of sequenced constraints since they are applicable at the current time, and (iii) non-sequenced constraints, which are applicable over a time interval (i.e., a part or the whole of the applicability bound of the constraint) rather than at a point in time. Finally, the authors describe the implementation of their approach (for managing temporal integrity constraints in τ XSchema) through a tool, named τ XMLLINT, which supports all introduced temporal constraints.

Brahmia et al. (2014a) propose a general approach for schema versioning in τ XSchema-based multitemporal XML repositories and deal with all kinds of schema changes (i.e., changes to conventional schema, changes to physical and logical annotations, and updates to the temporal schema) in an integrated environment. In this paper, the authors use their previous work on annotation document versioning (Brahmia et al., 2011) and conventional schema versioning (Brahmia et al., 2012) to provide a complete

figure of their approach for schema versioning in the τ XSchema framework and show, through a running example, how τ XSchema schema could be temporally versioned. The authors also study updates to temporal schema and propose a set of four primitives necessary for updating any temporal schema when its two components or one of them evolves (i.e., conventional schema and/or annotation document).

In order to help designers of τ XSchema schema, and to provide more-user friendly schema change operations that could be easily integrated in tools for XML schema versioning, Brahmia et al. (2014b) propose three sets of high-level operations each one of which is for the maintenance of a distinct component of the τ XSchema schema: forty nine operations acting on conventional schema, fifty-five operations acting on annotation documents, and four operations acting on temporal schema. Notice that a high-level schema change operation is defined as a valid and optimized sequence of schema change primitives (or low-level schema change operations), which expresses a frequent schema evolution need and specifies a complex schema change in a more compact way (Guerrini et al., 2005).

Schema Versioning in Emerging Databases

In the following paragraphs, we present contributions dealing with schema versioning in NoSQL databases.

Castelltort and Laurent (2013) present an approach for versioning graph-oriented NoSQL databases (Robinson et al., 2013). It is based on the following concepts: (i) Data graph: the current version of the graph (also called the operational graph database); (ii) Transaction: a sequence of change operations performed on the data graph in order to update it, while respecting the atomicity property; (iii) Revision: a new version of a graph created after the execution of a transaction; (iv) Version graph: a graph that includes all versions of a data graph (also called the historical graph

database). In this approach, graph versioning is managed as follows: (i) each node (relationship, respectively) in the data graph is represented by a node (relationship, respectively) in the version graph; (ii) to each element (node or relationship) in the data graph, also called trace element, a linked list of revision elements is associated; (iii) each revision element is characterized by (a) all properties of the corresponding element (in the data graph), (b) some useful meta-information (e.g., creation date, creator), and (c) a set for ingoing relationships and a set for outgoing relationships; (iv) each transaction produces a new version graph revision.

Felber et al. (2014) deal with the design and implementation space for providing versioning support on top of a distributed key-value store. They define an API for versioned data access supporting multiple writers and show that a plain key-value store does not offer the necessary synchronization power for implementing this API. They leverage the support for listeners at the key-value store level and propose a general construction for implementing arbitrary types of data structures for storing and querying versioned data. The produced versioning system, ALEPH, is implemented on top of Infinispan, which is an industrial-grade open-source key-value store. The authors evaluate ALEPH under a real workload from Wikipedia access traces.

Wen et al. (2014) introduce Llama, a distributed multi-versioned Model data management architecture to manage the schemata and attributes of structured data in NoSQL databases. In fact, Llama provides solutions for three problems which occur in Supervisory Control and Data Acquisition (SCADA) systems (Carcano et al., 2011) when the Model evolve (notice here that the Model is a set of data to describe the physical topology of monitored devices, including data definitions, device description, device association expression ...): (i) little capacity on schema management of structured data in NoSQL stores, (ii) Model has many versions but changes little, which lead to much unwanted redundancy, and (iii) SQL supported on NoSQL

could lead to a decline in access performance. The authors apply incremental version maintenance and time aggregation compression strategies to reduce the average storage overhead. They claim that Llama reaches the current hot spot version Model with low latency, with it cached in memory and write penetration implemented. Moreover, the authors show, through experimental results, that the average storage overhead is reduced to 20% in Llama, and the read speed of hot spot is more than 10 times faster.

Ruiz et al. (2015) propose a reverse engineering approach to infer the implicit schema of a NoSQL database, while taking into account all the versions of the entities and their relationships. These inferred schemas are named “Versioned Schemas”. Furthermore, the authors also show the usefulness of the inferred versioned schemata through two applications: schema visualization, and automated generation of data validators. Finally, the authors claim that their approach has been validated for three NoSQL DBMSs: MongoDB, CouchDB, and HBase.

DBMS SUPPORT FOR SCHEMA VERSIONING

As far as schema versioning support in conventional DBMSs is concerned, there are no updates that must be made to what we surveyed in (Brahmia et al., 2015). In practice, the only systems that partially provide schema versioning support for XML databases are SQL Server (limited to a subset of XML schema modifications) and IBM DB2 (limited to changes ensuring backward compatibility).

As far as schema versioning support in emerging DBMSs is concerned, Felber et al. (2014) compared versioning support in a set of most used key-value NoSQL systems (i.e., Cassandra, Riak, Couchbase, CouchDB, MongoDB, HBase, Dynamo, and HyperTable). They showed that, although all of them offer support for data versioning and some (i.e., MongoDB with multi-version

concurrency control, HBase, and HyperTable) also allow to maintain complete histories of temporal data versions, none of them provide schema versioning support.

Furthermore, Störl et al. (2015) survey the state-of-the-art of Object-NoSQL Mappers, which are currently used by application developers to deal with schema evolution problems in schema-less NoSQL data stores, since available NoSQL systems do not provide any support for evolving schema of these data stores (notice here that the lack of such a support could be due to the nature of these databases themselves, which are designed to be schema-less or schema flexible). The authors claim that although Object-NoSQL Mappers could manage different versions of objects, these tools do not provide an explicit support for managing schema versions.

FUTURE RESEARCH DIRECTIONS

The open research issues for schema versioning include:

1. Explicit schema versioning support is absent in all commercial systems (i.e., DBMSs and schema management tools), as shown in the previous section. Thus, this aspect needs further efforts in order to provide more efficient practical solutions for database designers and administrators.
2. Schema versioning support is also absent in the standard SQL language. Hence, SQL cannot be used to exploit the potentialities of a relational database supporting schema versioning. Assuming the proliferation of different versions of schema and associated data, suitable extensions to SQL should be defined and implemented in order to allow applications to access data defined under multiple schemas.
3. In a system where different schema versions coexist, the database designer or administrator could need to perform the following tasks:
 - a. To compare some schema versions in order to fulfill some requirements; therefore, comparing two different schema versions to identify the differences could also be an interesting research issue (available research results on schema integration could be adapted to the purpose).
 - b. To migrate data from one schema version to another one dynamically, even when a database continues to accept new data; thus, focusing on tools that could facilitate such a migration would be an interesting research topic which has rarely been addressed.
4. The support of schema versioning could lead to performance degradation due to the proliferation of versions. Thus, further interesting and challenging research topics are (i) efficient techniques to organize data storage spaces and (ii) advanced index structures to optimize data access under schema versioning.
5. Nowadays, many enterprises use Web and distributed databases and, thus, schema changes have to be effected more frequently, since user requirements, laws and technologies change rapidly. Therefore, it will be very interesting and useful to study schema versioning in Web and distributed environments. Among the first works in this area, we find (Rae et al., 2013), which describes techniques used by Google's F1 DBMS (Shute et al., 2013) to support schema evolution in a globally distributed environment.
6. The database provenance is very useful in many modern information systems (like big science projects), in order to explain the lineage of the current data. This explanation often requires "flashing back" to the original schema and, thus, supporting database provenance under schema versioning could be an interesting issue. Although provenance has been studied under schema evolution (Gao & Zaniolo, 2012), to the best of our knowledge, currently there is no work on data provenance under schema versioning.

CONCLUSION

In this chapter, we studied recent advances on schema versioning in traditional and emerging databases, and explored its support in commercial database systems.

Although a lot of work has been done at theoretical level and solutions have been proposed by researchers, there is no efficient practical support for schema versioning at the state-of-the-art of database technology. In order to successfully manage the schema versioning process, the database designer or administrator has to manually convert data defined under the previous schema in order to be valid with respect to the new schema, and the application developer(s) must carefully revise existing applications to support the new schema.

REFERENCES

- Baqasah, A., Pardede, E., Holubová, I., & Rahayu, W. (2013). On Change Detection of XML Schemas. In *Proceedings of the IEEE TrustCom/ISPA/IUCC 2013*, (pp. 974-982).
- Baqasah, A., Pardede, E., & Rahayu, J. W. (2014). A New Approach for Meaningful XML Schema Merging. In *Proceedings of the 16th International Conference on Information Integration and Web-based Applications & Services (iiWAS 2014)*, (pp. 430-439). doi:10.1145/2684200.2684302
- Brahmia, Z., Bouaziz, R., Grandi, F., & Oliboni, B. (2011). Schema Versioning in τ XSchema-Based Multitemporal XML Repositories. In *Proceedings of the 5th IEEE International Conference on Research Challenges in Information Science (RCIS 2011)*, (pp. 1-12).
- Brahmia, Z., Grandi, F., Oliboni, B., & Bouaziz, R. (2012). Versioning of Conventional Schema in the τ XSchema Framework. In *Proceedings of the 8th International Conference on Signal Image Technology & Internet Systems (SITIS'2012)*, (pp. 510-518).
- Brahmia, Z., Grandi, F., Oliboni, B., & Bouaziz, R. (2014a). Schema Change Operations for Full Support of Schema Versioning in the τ XSchema Framework. *International Journal of Information Technology and Web Engineering*, 9(2), 20-46. doi:10.4018/ijitwe.2014040102
- Brahmia, Z., Grandi, F., Oliboni, B., & Bouaziz, R. (2014b). High-level Operations for Creation and Maintenance of Temporal and Conventional Schema in the τ XSchema Framework. In *Proceedings of the 21st International Symposium on Temporal Representation and Reasoning (TIME'2014)*, (pp. 101-110). doi:10.1109/TIME.2014.14
- Brahmia, Z., Grandi, F., Oliboni, B., & Bouaziz, R. (2015). Schema Versioning. In M. Khosrow-Pour (Ed.), *Encyclopedia of Information Science and Technology* (3rd ed.; pp. 7651-7661). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-5888-2.ch754
- Carcano, A., Coletta, A., Guglielmi, M., Masera, M., Fovino, I. N., & Trombetta, A. (2011). A Multidimensional Critical State Analysis for Detecting Intrusions in SCADA Systems. *IEEE Transactions on Industrial Informatics*, 7(2), 179-186. doi:10.1109/TII.2010.2099234
- Castelltort, A., & Laurent, A. (2013). Representing history in graph-oriented NoSQL databases: A versioning system. In *Proceedings of the 8th International Conference on Digital Information Management (ICDIM 2013)*, (pp. 228-234). doi:10.1109/ICDIM.2013.6694022
- Currim, F., Currim, S., Dyreson, C. E., Joshi, S., Snodgrass, R. T., Thomas, S. W., & Roeder, E. (2009). *τ XSchema: Support for Data- and Schema-Versioned XML Documents*. Technical Report TR-91, TimeCenter. Retrieved from <http://timecenter.cs.aau.dk/TimeCenterPublications/TR-91.pdf>

- Currim, F., Currim, S., Dyreson, C. E., Snodgrass, R. T., Thomas, S. W., & Zhang, R. (2012). Adding Temporal Constraints to XML Schema. *IEEE Transactions on Knowledge and Data Engineering*, 24(8), 1361–1377. doi:10.1109/TKDE.2011.74
- Date, C. J. (2003). *An Introduction to Database Systems* (8th ed.). Addison-Wesley Longman, Inc.
- Felber, P., Pasin, M., Rivière, E., Schiavoni, V., Sutra, P., Coelho, F., & Vilaça, R. et al. (2014). On the Support of Versioning in Distributed Key-Value Stores. In *Proceedings of the 33rd IEEE International Symposium on Reliable Distributed Systems (SRDS 2014)*, (pp. 95-104). doi:10.1109/SRDS.2014.35
- Gao, S., & Zaniolo, C. (2012). Supporting Database Provenance under Schema Evolution. In *Proceedings of the 5th International ER Workshop on Evolution and Change in Data Management and on Non Conventional Data Access (ECDM – NoCoDa 2012)*, (pp. 67-77).
- Guerrini, G., Mesiti, M., & Rossi, D. (2005). Impact of XML Schema Evolution on Valid Documents. In *Proceedings of the 7th ACM International Workshop on Web Information and Data Management (WIDM 2005)*, (pp. 39-44). doi:10.1145/1097047.1097056
- Lee, S.-W., Ahn, J.-H., & Kim, H.-J. (2006). A schema version model for complex objects in object-oriented databases. *Journal of Systems Architecture*, 52(10), 563–577. doi:10.1016/j.sysarc.2006.04.001
- Liu, X., Nelson, D., Stobart, S., & Stirk, S. (2005). Managing Schema Versions in Object-Oriented Databases. In *Proceedings of the 9th East European Conference on Advances in Databases and Information Systems (ADBIS 2005)*, (pp. 97-108). doi:10.1007/11547686_8
- Malý, J., Mlýnková, I., & Nečaský, M. (2011). XML Data Transformations as Schema Evolves. In *Proceedings of the 15th International Conference on Advances in Databases and Information Systems (ADBIS 2011)*, (pp. 375-388). doi:10.1007/978-3-642-23737-9_27
- Miller, J., & Mukerji, J. (2003). *MDA Guide Version 1.0.1*. Object Management Group.
- Nečaský, M., & Mlýnková, I. (2009). On different perspectives of xml schema evolution. In *Proceedings of the 4th International Workshop on Flexible Database and Information System Technology (FlexDBIST 2009)*, (pp. 422-426).
- Piccioni, M., Oriol, M., & Meyer, B. (2013). Class Schema Evolution for Persistent Object-Oriented Software: Model, Empirical Study, and Automated Support. *IEEE Transactions on Software Engineering*, 39(2), 184–196. doi:10.1109/TSE.2011.123
- Rae, I., Rollins, E., Shute, J., Sodhi, S., & Vingralek, R. (2013). Online, Asynchronous Schema Change in F1. *Proceedings of the VLDB Endowment*, 6(11), 1045–1056. doi:10.14778/2536222.2536230
- Robinson, I., Webber, J., & Eifrem, E. (2013). *Graph Databases*. O'Reilly Media, Inc.
- Roddick, J. F. (1995). A survey of schema versioning issues for database systems. *Information and Software Technology*, 37(7), 383–393. doi:10.1016/0950-5849(95)91494-K
- Ruiz, D. S., Morales, S. F., & Molina, J. G. (2015). Inferring Versioned Schemas from NoSQL Databases and Its Applications. In *Proceedings of the 34th International Conference on Conceptual Modeling (ER 2015)*, (pp. 467-480).
- Shute, J., Vingralek, R., Samwel, B., Handy, B., Whipkey, C., Rollins, E., & Apte, H. et al. (2013). F1: A Distributed SQL Database That Scales. *Proceedings of the VLDB Endowment*, 6(11), 1068–1079. doi:10.14778/2536222.2536232

Snodgrass, R. T., Dyreson, C. E., Currim, F., Currim, S., & Joshi, S. (2008). Validating Quicksand: Schema Versioning in τ XSchema. *Data & Knowledge Engineering*, 65(2), 223–242. doi:10.1016/j.datak.2007.09.003

Störl, U., Hauf, T., Klettke, M., & Scherzinger, S. (2015). Schemaless NoSQL Data Stores - Object-NoSQL Mappers to the Rescue? In *Proceedings of the 16th Conference on Database Systems for Business, Technology, and Web (BTW 2015)*.

W3C. (2007). *XSL Transformations (XSLT) Version 2.0 - W3C Recommendation*. Retrieved from <<http://www.w3.org/TR/2007/REC-xslt20-20070123/>>

Wen, J., Dai, J., Han, J., Di, F., & Yan, H. (2014). Llama: Distributed Multi-versioned Model Data Management for SCADA. In *Proceedings of the 17th IEEE International Conference on Computational Science and Engineering (CSE 2014)*, (pp. 1968–1973).

ADDITIONAL READING

Bounif, H. (2009). Versioning Approach for Database Evolution. In V. E. Ferraggine, J. H. Doorn, & L. C. Rivero (Eds.), *Handbook of Research on Innovations in Database Technologies and Applications: Current and Future Trends* (pp. 91–101). Hershey, PA: Information Science Reference. doi:10.4018/978-1-60566-242-8.ch011

Eder, J., & Wiggisser, K. (2009). Data Warehouse Maintenance, Evolution and Versioning. In L. Liu & M. T. Özsu (Eds.), *Encyclopedia of Database Systems* (pp. 664–669). Springer, US.

Ram, S., & Shankaranarayanan, G. (2003). *Research issues in database schema evolution: the road not taken. Technical Report #2003-15*. Tucson, AZ, USA: The University of Arizona.

Roddick, J. F. (2009). Schema Versioning. In L. Liu & M. T. Özsu (Eds.), *Encyclopedia of Database Systems* (pp. 2499–2502). Springer, US.

KEY TERMS AND DEFINITIONS

Database Conversion/Migration: Process consisting in adapting data valid under a given schema version to another schema version.

Database Schema: The formal structure of the database, described by means of a data definition language.

Schema Change Propagation: The effects of a schema change at instance level, involving suitable conversions necessary to adapt extant data to the new schema.

Schema Change Semantics: The effects of a schema change at schema level.

Schema Changes: Operations performed on a (populated) database schema, in order to adapt it to new application requirements.

Schema Evolution: A schema change modality that avoids the loss of extant data. No support is required for previous schemata.

Schema Version: A complete database schema as created by the initial design or by the application of schema changes to an existing schema.

Schema Versioning: A schema change modality that provides for the maintenance of previous schema versions, such that extant data are accessible through any of the supported schema versions.

Temporal Database: A database with built-in support for managing time-varying data.

Temporal Schema Versioning: Schema versioning technique, in which different schema versions are indexed by timestamps.

Twitter Data Mining for Situational Awareness

Marco Vernier

University of Udine, Italy

Manuela Farinosi

University of Udine, Italy

Gian Luca Foresti

University of Udine, Italy

INTRODUCTION

The most recent catastrophic events, from the 2010 Haiti earthquake to the devastating 2013 Colorado floods, have shown a strong adoption of social media platforms by ordinary people. The data and meta-data produced by users during and after extra-ordinary situations could have enormous potentialities if integrated with traditional systems for emergency management and used for hyperlocal situational awareness (Foresti et al., 2015a). The great majority of current literature is focused on Twitter for several reasons, in particular, the architectures and practices of the platform itself. In the following section, an overview of social media use in extra-ordinary contexts will be presented, with a specific focus on the Twitter social media platform. Following this, in a section dedicated to the Twitter systems classification for situational awareness, a classification of the existing systems based on the analysis of Twitter data will be provided on the basis of the following three categories: 1) semantic systems; 2) meta-data systems; and 3) smart self-learning systems. Finally, in the future research trends section, an innovative and smart solution will be proposed for future development.

BACKGROUND

Social Media Use in Extra-Ordinary Contexts

Social media platforms are built from the beginning to be used socially, and oriented around collaboration and sharing. These potentialities are emphasized in extra-ordinary contexts, when ordinary people adopt these tools to provide or search for first-hand and real-time information regarding a certain event (i.e. an earthquake, flood, etc.) (Lindsay, 2011; Taylor et al., 2012). The most recent catastrophic events, from the 2010 Haiti earthquake to the devastating 2013 Colorado floods, in fact, have shown that these platforms have been strongly used both during and after disasters (Figure 1), allowing a real-time dissemination of information to the wider public, an effective situational awareness, and an up-to-date picture of what is happening on the ground (i.e. Farinosi & Micalizzi, 2013; White et al., 2014).

According to Kotsiopoulos (2014), in extra-ordinary situations, social media enable citizens to play at least three roles: 1) first responders/volunteers; 2) citizen journalists/reporters; and 3) social activists. Oftentimes citizens on the scene experience the event first-hand and are able to

provide updates more quickly than disaster response organizations and traditional news media (Sweetser & Metzgar, 2007; Procopio & Procopio, 2007; Farinosi & Micalizzi, 2013).

Given the increasing availability of data and meta-data produced and/or distributed on these online platforms, it is pivotal to understand how they should be used and integrated with traditional systems for situational awareness, supporting in this way the work of Civil Protection, Red Cross, Fire Department, and other agencies.

A Focus on Twitter

The great majority of existing research is focused on the social media platform Twitter. Twitter is a popular microblogging network that allow users to share up to 140 character text messages called tweets. Twitter was founded in 2006 by Jack Dorsey, Evan Williams, Biz Stone, and Noah Glass. Today it is one of the most used social platforms with about 250 million active users and about 500 million tweets shared per day (source: <https://dev.twitter.com/>). The platform is mainly based on the exchange of messages between networks of contacts. Each user can create his/her own custom networks, following other users' feeds. In turn, each user can be followed by other users called followers (Marwick and Boyd, 2011). Besides the possibility to share a short text content, Twitter also offers the opportunity to share visual content, such as videos and images, and URLs to web sites (Java et al., 2007; Huberman et al., 2008; Zhao & Rosson, 2009).

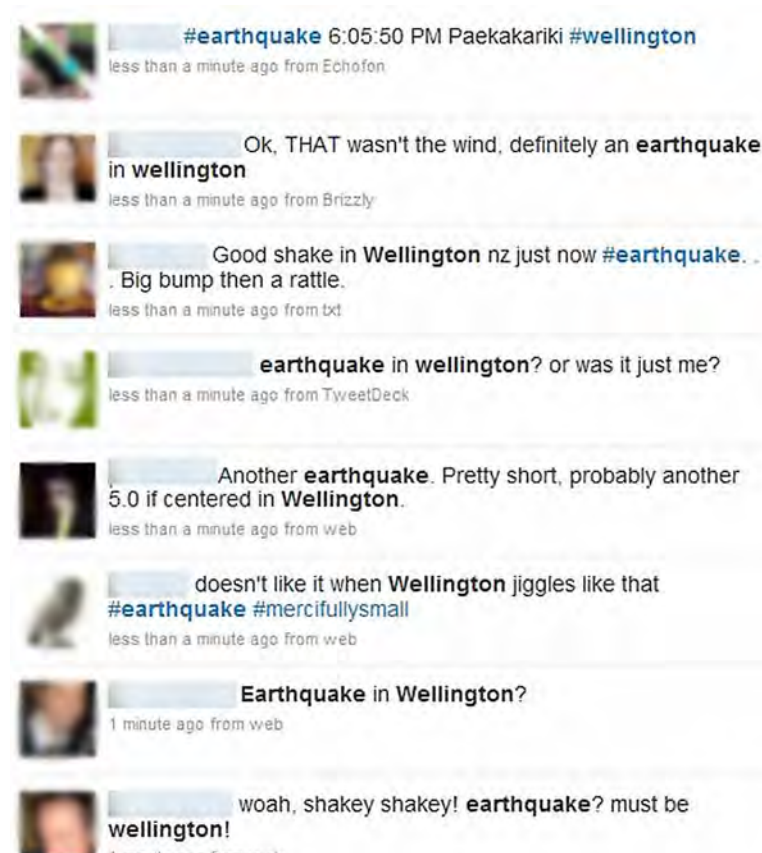
The reason why the majority of research is focused on Twitter is due to many factors. First of all, given the instantaneous nature of communication on Twitter, the platform is particularly suitable for real-time communication. Furthermore, the architecture and some specific features of Twitter seem to facilitate the widespread dissemination of information. Among the most important features offered by the Twitter platform is the possibility to share content using hashtags. This is an annotation format represented by the “#” symbol, used to

indicate with a single word (or a combination of words) the core meaning of a tweet. Conversations centered on specific hashtags promote focused discussions, even among people who are not in direct contact with each other.

In addition, the choice to analyze Twitter is also motivated by the prevailing public nature of the great majority of the accounts (only a small percentage of accounts are private), a feature that distinguishes this platform to other social networking sites (Bruns & Burgess, 2013). On the one hand, this peculiarity promotes public conversations, even among users that were not previously in contact with one another (Twitter in fact offers the possibility to interact with other users, to share content with them, and to reply or mention someone in your own tweet simply using the symbol “@” followed by the username of the person whom you want to tag). On the other hand, it makes it easier to conduct analysis that aims to reconstruct the spread of communication flows within the platform. Finally, yet importantly, this characteristic makes the use of tweets for research purposes less critical from an ethical point of view.

For example, the analysis of the 2011 floods in Queensland provides a detailed mapping of the general dynamics of Twitter use during an emergency and offers useful general indications (Bruns, 2012). Their findings highlight that the space-time variables represent a crucial element to obtain relevant data and to improve situational awareness during disasters. For instance, the physical distance of the Twitter users from the site of the catastrophe can reflect different types of needs to be met and a different perception of danger. In addition, given that a social platform like Twitter is structurally connected to forms of activation just in time, the time variable plays a fundamental role. Previous research demonstrates that immediately after the event there is a greater presence of forms of instinctive response, while tweets containing links to official news sources tend to arrive later (Acar & Muraki, 2011; Farinosi & Micalizzi, 2013).

Figure 1. An example of tweets posted on Twitter during the 2013 Wellington earthquake



DATA MINING ON TWITTER

Tweet data extraction operations involve the use of suitable API libraries, which provide all the tools and services to retrieve user tweets and their correlated information.

When a user posts a tweet on Twitter, besides the text content of 140 characters, he/she can share other useful information such as videos, images or GPS coordinates. Considering a given emergency event, this information could be useful to accurately locate the area where the event is occurring and to better coordinate rescue operations.

In the state-of-the-art there are several APIs (Application Programming Interfaces) which provide useful tools to collect and analyze tweets shared by users. In general, it is possible to dis-

tinguish official Twitter libraries and unofficial libraries created for specific programming languages (i.e. Neo4J source: <http://neo4j.com/>). Usually the unofficial libraries use the core libraries of the Twitter APIs to retrieve data from the Twitter server but are customized for specific purposes.

Twitter APIs are the official programming libraries proposed by Twitter (source: <https://dev.twitter.com/>). These APIs allow different type of operations to be performed such as the extraction of tweets from Twitter streams or creation of custom applications for websites. They are freely available although they can be used with some limitations (see “Twitter APIs limitations” section).

Twitter APIs can be generally classified into four groups: 1) Twitter for Websites, 2) Search APIs, 3) REST APIs, and 4) Streaming APIs.

Twitter APIs Limitations

Twitter APIs are freely available for the creation of custom applications with the goal of analyzing Twitter stream data and extracting useful information.

Nevertheless, APIs have certain limitations. First, there is a restriction regarding the number of requests that can be sent to Twitter servers. For example, if you want to extract data of a specific user (e.g. information about the profile of the user, his/her tweets etc.) or simply to collect tweets containing a specific keyword, it is only possible to send 180 requests every 15 minutes. If the maximum limit is reached, then the system automatically discards the following requests until the time window of 15 minutes has elapsed.

Moreover, there are other limitations regarding the use of Twitter APIs. For example, considering Streaming APIs, only a random sample corresponding to 1% of the total tweets is retrieved from Twitter stream data. This means that some relevant information could be missed during the harvesting process.

Furthermore, the REST APIs have some limitations. They do not provide access to tweets up to about a week old. Moreover, only tweets posted from public Twitter profiles can be retrieved, collected, and analyzed.

TWITTER SYSTEM CLASSIFICATION FOR SITUATIONAL AWARENESS

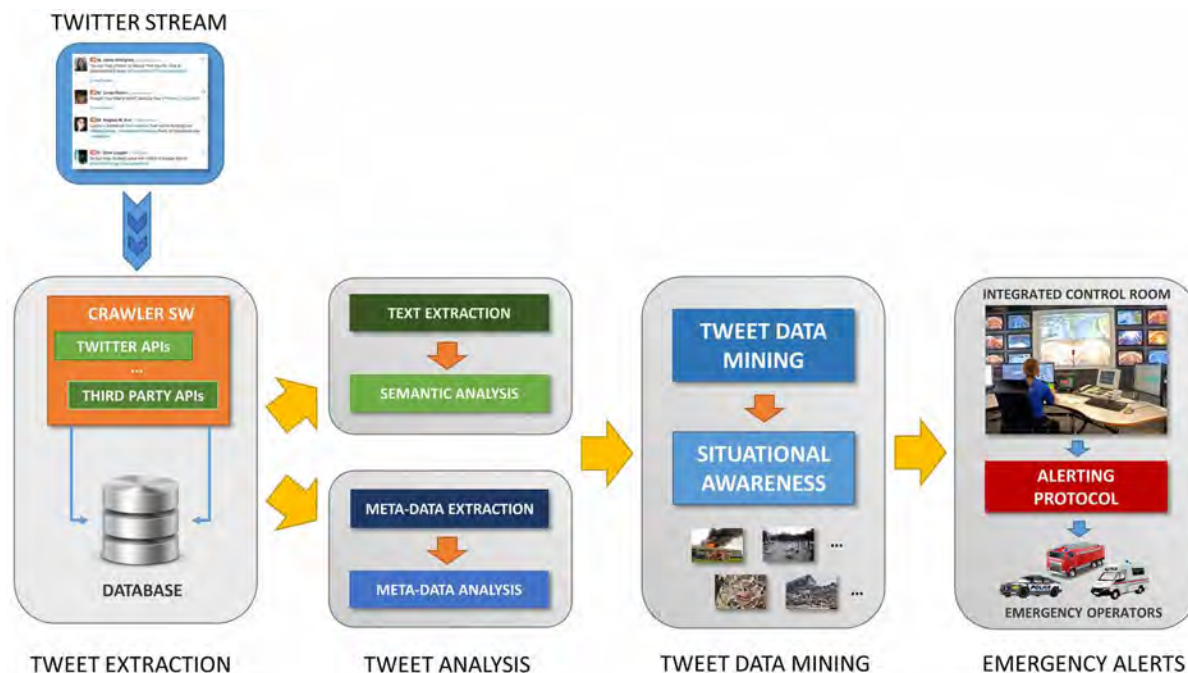
In the literature, there are many significant studies exploiting the use of Twitter to locate and analyze in real-time a given event for situational awareness. In particular, these systems are able to extract and save tweets from Twitter streams and analyze them with the goal to collect useful information. All these systems are very similar to each other, although they can be classified into three main groups on the basis of their characteristics:

1. Semantic systems, based on a textual analysis of the content;
2. Meta-data systems, based on the extraction of meta-data information;
3. Smart self-learning systems, able to identify in real-time a trending topic and to automatically search other secondary hashtags connected to the initial one.

Regarding the semantic systems focusing on situational awareness (Figure 2), this kind of application is characterized by the possibility to analyze the textual content of the tweets written and shared online by Twitter users. In several cases, the tweets posted by users could be very useful to identify relevant topics of discussion and, in doing so, to detect specific contextual information (i.e. collapsed buildings or numbers of wounded, dead, etc.). The first example of semantic systems is identified by Corvey et al. (2012) who are able to extract linguistic and behavioral information from tweet text to aid in the task of information integration. Through linguistic annotation, in the form of Named Entity Tagging and behavioral annotations, it captures tweets and then analyses and classifies their content, contributing to situational awareness. Another interesting example of this kind of system is that developed by Schulz et al. (2012), which uses crowdsourcing and Linked Open Data to enhance, classify, and filter the information shared by people on social media (i.e. Twitter or Facebook). This results in a structured dataset that contributes to the identification of a certain event.

A third example can be identified in the system implemented by Zielinski et al. (2012), which is able to analyze multilingual Twitter feeds for emergency events. Exploring the number of tweets posted online before and after ten earthquakes in the Mediterranean area, it detects a disaster by observing a rapid increase in Twitter, focusing in particular on posts written in the native language of the area hit by the seism.

Figure 2. An example of a system to extract data from Twitter stream during an emergency event



It is worth mentioning also the system developed by Klein et al. (2012), a Twitter crawler that analyzes the stream through a combined social and content-aware analysis approach. It conducts a grammatical analysis in order to classify the textual content of tweets in nouns, verbs, adjectives, etc. and groups single words in meaningful units, extracting in this way knowledge from texts through named entity recognition algorithms. By querying the Twitter Streaming API with ad hoc keywords focused on the specific disaster and the location, it is able to detect emergencies in real-time.

Another example of a tool that can be employed for situational awareness, even if originally conceived for a broader purpose, is Eddi, an interactive topic-based browser for Twitter feeds. It allows for the organization of tweets on the basis of specific topics (e.g. a crisis event or political elections, etc.) making their visualization more feasible. It provides different features: 1) a tag cloud overview for showing major topics, each of which is scaled proportionally to the number of tweets assigned to it; 2) a topic detail view, available by clicking

on each tag, provides a view of all tweets about a specific topic; 3) navigation list, that contains a complete list of the feed's topics sorted by popularity; 4) topic dashboard for displaying an overview of the most interesting topics that the user might like; and 5) a timeline that focuses on the temporal factor, stressing on spikes and trends over time. Among the possible contexts of use, Eddi could be adopted during and after a natural disaster to identify trending topics of discussion and relevant content of tweets.

Regarding the meta-data systems (Figure 2), these kinds of applications are able to use and aggregate different types of data. As for the semantic systems, they are able to extract and analyze the textual content of tweets related to a certain event, however, they also take into account other types of data directly produced by the social platform itself, such as time and date of the tweet, GPS coordinates, URLs shared, and so on. Twitcident (Abel et al., 2012), is the first example of a meta-data system. This is a framework and web-based system that is automatically able to filter, search, and analyze tweets about real-world incidents,

crises or natural hazards. Adopting semantic filtering strategies including tweet classification, named entity recognition, and linkage to related external online resources, it monitors emergency broadcasting services and automatically collects and filters tweets whenever an incident occurs.

The core of Twitcident is composed of several modules: 1) Incident detection module, which is able to identify incidents employing the P2000 communication network, an ad hoc broadcasting service used in The Netherlands by Police and other emergency operators. Whenever an incident is detected, the collected messages are semantically analyzed by the 2) Semantic enrichment module. This identifies the tweets relevant to a given incident and performs a real-time analysis. Twitcident also collects additional meta-data about the publisher of a tweet, his/her profile picture, number of followers, his/her location when publishing the tweet etc. All these data are useful to assess the trustworthiness of a tweet and improve the reliability of the whole system.

Another system for supporting emergency planning and risk assessment activities that also takes into account meta-data was developed by De Longueville (2009). This project analyzed the activity of Twitter users during a fire that broke out in Marseille in July 2009 through the Twitter's API and a search of the keywords "incendie" and "Marseille". They classified the content on the basis of the users and identified three major roles: citizens, media, and aggregators. This last category did not generate primary content but distributed existing sources (i.e. news produced by citizen journalists) through abbreviated URLs to blogs, news portals, and picture websites such as Flickr and Twitpic. It is worth noting also TEDAS (Rui Li et al., 2012), a system which uses keywords and GPS coordinates to discover spatial and temporal patterns of events, and the region of major influence. The high coverage of Twitter users, along with the rich information associated to tweets, allows TEDAS monitor events in a fast and accurate way. It is able to recognize new events, rank them according to their importance, and generate

a spatial-temporal pattern for any event. The tool takes a query, which contains keywords, a temporal period, and a location and extrapolates relevant tweets. Furthermore, it identifies and classifies tweets based on some relevant keywords, giving more relevance to those posted by an authority.

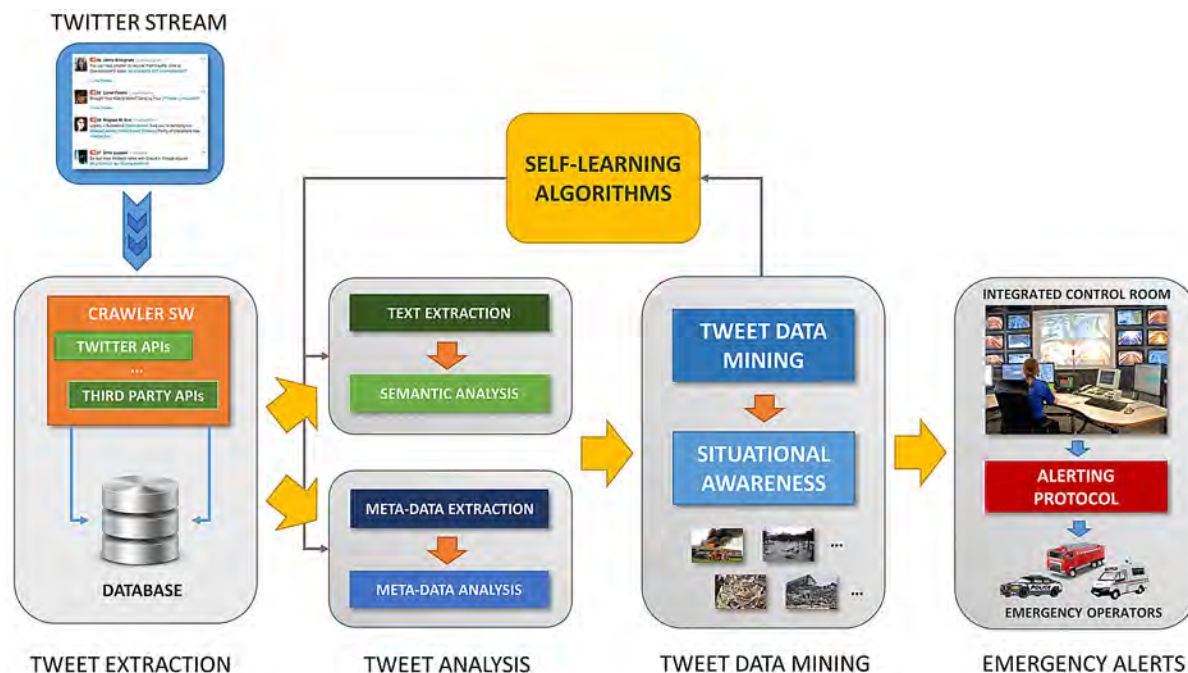
Another interesting system has been developed by Sakaki et al. (2010). Through the analysis of tweets and keywords (e.g. "earthquake" or "shake"), it is able to automatically detect target events and send notifications promptly to those who are registered. Moreover, the system applies Kalman and particle filtering in order to detect accurately the location of the event.

It is worth mentioning also a tool created to enhance the situational awareness in emergency events which uses Twitter Search APIs to find tweets through a query of case-insensitive terms and then analyses textual data, GPS-coordinates, and the registered location of the tweets, and visualizes their content by the E-Data Viewer (EDV) (Vieweg et al., 2010). A more specific system for disaster response is that developed to help relief workers during natural disasters (Ashktorab et al., 2014). It adopts different methods (i.e. sLDA, SVM) to classify in real-time the material shared from areas close to the disaster location and extracts the most relevant phrases about structural and infrastructural damage.

A similar system has been developed by Walther and Kaiser (2013). This is able to detect geo-spatial real-world events in real-time. It focuses on clusters of Tweets sent in a short time span and stores them for at least 48 hours in an open-source database (MangoDB). Then - through a Machine Learning component, which uses 41 features that address several of their aspects - the material is analyzed and filtered in order to eliminate tweets generated from bots or false positives.

For what concerns the last category, it refers to smart self-learning systems (Figure 3), and includes those applications able to identify in real-time a location affected by a certain event combining characteristics used by semantic and meta-data systems.

Figure 3. An example of smart self-learning system



However, in contrast to the previous categories, they are also able to self-learn from the content of tweets analyzed, identifying the most adopted words and hashtags and automatically looking for other secondary hashtags connected to the first ones. A first example of these systems is CrisisLex (Olteanu et al., 2014). The aim of this system is to improve the query to the Twitter servers in order to harvest the most relevant tweets during a specific crisis. As a first stage, the system creates a lexicon of crisis-related terms that frequently appear in relevant messages posted during different types of past crisis situations. The lexicon is then used by the system to automatically identify new terms that describe a given crisis. Moreover, it allows the automatization of the choice of keywords contained in the tweets. Data collection is optimized using keywords or geo-location searches. CrisisLex constitutes a first attempt to develop a smart self-learning system. Further research is needed to develop more accurate systems able to identify specific areas where the event is occurring, extracting only the data related to that event in that area.

In the next section, a possible innovative smart solution will be illustrated.

FUTURE RESEARCH TRENDS

As presented in the previous section, a lot of systems that collect and analyze in real-time data from Twitter to extract useful information of a given event already exist in the literature. They are mainly based on the semantic analysis of the text content of the tweets or on the analysis of meta-data linked to the tweets, such as the URLs shared by Twitter users or GPS coordinates. Even if this kind of system is able to analyze Twitter streams in real-time, future research will aim to develop more intelligent architecture integrating different types of data shared on Twitter by ordinary people during an event (Foresti et al., 2015b). Such a system should be able to extract data that are as accurate as possible both as regards the semantic analysis and the GPS location. When a certain event happens, the system could automatically detect the geographic area where it is occurring

with the goal to extract only the relevant information related to the event itself. For this purpose, state-of-the-art clustering algorithms could be used to identify and correctly classify the different events. The system should be able to consider several parameters such as the number of tweets shared by users in a certain lapse of time, and to compare them to the density of the population of the specific area, automatically recognizing if a certain event is happening or not. Another important feature which could be investigated in future research is automatic trend hashtags identification with the aim to improve situational awareness. The hashtag is an annotation format used on Twitter to indicate the topic of a tweet through one or more specific keywords. Nevertheless, when a crisis situation occurs, the hashtags adopted by Twitter users could be manifold, each one with a specific meaning. Therefore, the system should be able to classify hashtags considering their importance and use them to optimize future research on a given event. A solution could be the use of an automatic algorithm which gets as input all the hashtags adopted by the users to spread the event, and classifying them by keeping account of only the most relevant ones. Then, the identified hashtags could be used as input for future research with the main goal of optimizing them and also training the system.

Moreover, another feature that could be implemented to enhance rescue operations is the analysis of images posted by users on Twitter. In this way, future systems might be able to take into account not only textual data or meta-data, but also visual data uploaded by people and distributed on social media platforms. This visual information represents a useful resource, especially when it is linked to GPS coordinates, given that it means the possibility to detect in real time a hyperlocal situation and obtain a detailed picture of a certain location. The system should be able to automatically compare the visual content with other kinds of images (i.e. pictures from Google Earth), providing emergency operators real time

remote monitoring of the situation and highlighting critical points that require action.

D

CONCLUSION

In this contribution, the most important systems based on Twitter data mining have been classified into three different groups: 1) semantic systems, which are based on a textual analysis of the content; 2) meta-data systems, based on the extraction of meta-data information; and finally 3) smart self-learning systems, able to identify in real-time a trending topic and automatically look for other secondary hashtags connected to the initial one. The analyzed literature shows that there are a large number of systems able to detect and analyze different kind of crisis events. Nevertheless, nowadays there are only a few systems based on an effective real-time identification of an event and it would be really useful to improve the algorithms able to self-learn from previous events. For this purpose, in the section “future research trends”, potential new applications to enhance smart self-learning systems have been proposed. In the future such systems exploiting grassroots data generated and spread by citizens on the scene of disasters will gain a strategic role in emergency management and provide strong support to institutional operators.

REFERENCES

- Abel, F., Hauff, C., Houben, G., Tao, K., & Stronkman, R. (2012). Twitcident: Fighting Fire with Information from Social Web Streams. *Proceedings of IW3C2. WWW 2012 Companion*.
- Acar, A., & Muraki, Y. (2011). Twitter for crisis communication: Lessons learned from Japans tsunami disaster. *Int. J. Web Based Communities*, 7(3), 392–402. doi:10.1504/IJWBC.2011.041206

- Ashktorab, Z., Brown, C., Nandi, M., & Culotta, A. (2014). Tweedr: Mining Twitter to Inform Disaster Response. *Proceedings of the 11th International ISCRAM Conference*.
- Bruns, A., & Burgess, J., (2013). *Crisis Communication in Natural Disasters: The Queensland Floods and Christchurch Earthquakes, 2013*. Academic Press.
- Bruns, A., Burgess, J., Crawford, K., & Shaw, F. (2012). #qldfloods and @ QPSMedia: Crisis Communication on Twitter in the 2011 South East Queensland Floods. ARC Centre of Excellence for Creative Industries and Innovation, Brisbane.
- Corvey, W. J., Verma, S., Vieweg, S., Palmer, M., & Martin, J. H. (2012). Foundations of a Multilayer Annotation Framework for Twitter Communications during Crisis Events. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012)*.
- De Longueville, B., Smith, R. S., & Luraschi, G. (2009). "OMG, from here, I can see the flames!": a use case of mining location based social networks to acquire spatio-temporal data on forest fires. In *Proceedings of the 2009 International Workshop on Location Based Social Networks*, (pp. 73-80). doi:10.1145/1629890.1629907
- Farinosi, M., & Micalizzi, A. (2013). *NetQuake. Media Digitali e Disastri Naturali*. Milano: FrancoAngeli.
- Foresti, G. L., Farinosi, M., & Vernier, M. (2015a). Advanced emergency response management in smart environments. In M. Khosrow-Pour (Ed.), *Encyclopedia of Information Science and Technology* (3rd ed.; pp. 1337–1349). Hershey, PA: Information Science Reference. doi:10.4018/978-1-4666-5888-2.ch127
- Foresti, G. L., Farinosi, M., & Vernier, M. (2015b). Situational aware services in smart environments: Socio-mobile and sensor data fusion for emergency response to disasters. *Journal of Ambient Intelligence and Humanized Computing*, 6(2), 239–257. doi:10.1007/s12652-014-0227-x
- Huberman, B.A., Romero, D.M., & Wu, F. (2008). *Social Networks that Matter: Twitter Under the Microscope*. Academic Press.
- Java, A., Song, X., Finin, T., & Tseng, B. (2007). Why we twitter: understanding microblogging usage and communities. *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*, (pp. 56-65). doi:10.1145/1348549.1348556
- Klein, B., Laiseca, X., Casado-Mansilla, D., López-de-Ipiña, D., & Prada Nespral, A. (2012). Detection and Extracting of Emergency Knowledge from Twitter Streams. In J. Bravo, D. López-de-Ipiña, & F. Moya (Eds.), *Lecture Notes in Computer Science: Vol. 7656. Ubiquitous Computing and Ambient Intelligence* (pp. 462–469). doi:10.1007/978-3-642-35377-2_64
- Kotsiopoulos, I. (2014). Social Media in Crisis Management: Role, Potential and Risk. In *2014 IEEE/ACM 7th International Conference on Utility and Cloud Computing*, (pp. 681-686).
- Li, R., & Lei, K. H., Khadiwala, R., & Chang, K.C.-C. (2012). TEDAS: A Twitter-based Event Detection and Analysis System. In *Proceedings of the 2012 IEEE 28th International Conference on Data Engineering (ICDE)*. doi:10.1109/ICDE.2012.125
- Lindsay, B.R. (2011). *Social Media and Disasters: Current Uses, Future Options, and Policy Considerations*. Congressional Research Service 7-5700 R41987.
- Marwick, A., & boyd., (2011). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society*, 13(1), 114–133. doi:10.1177/1461444810365313
- Olteanu, A., Castillo, C., Diaz, F., & Vieweg, S. (2014). CrisisLex: A Lexicon for Collecting and Filtering Microblogged Communications in Crises. *Proceedings of ICWSM*.

- Procopio, C. H., & Procopio, S. T. (2007). Do You Know What It Means to Miss New Orleans? Internet Communication, Geographic Community, and Social Capital in Crisis. *Journal of Applied Communication Research*, 35(1), 67–87. doi:10.1080/00909880601065722
- Sakaki, T., Okazaki, M., & Matsuo, Y. (2010). Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors. In *Proceeding WWW '10 Proceedings of the 19th international conference on World wide web*. doi:10.1145/1772690.1772777
- Schulz, A., Paulheim, H., & Probst, F. (2012). Crisis Information Management in the Web 3.0 Age. In *Proceedings of the 9th International Conference on Information Systems for Crisis Response and Management (ISCRAM)*.
- Sweetser, K. D., & Metzgar, E. (2007). Communicating during crisis: Use of blogs as a relationship management tool. *Public Relations Review*, 3(3), 340–342. doi:10.1016/j.pubrev.2007.05.016
- Taylor, M., Wells, G., Howell, G., & Raphael, B. (2012). The role of social media as psychological first aid as a support to community resilience building. *Australian Journal of Emergency Management*, 27(1).
- Vieweg, S., Hughes, A. L., Starbird, K., & Palen, L. (2010). Microblogging During Two Natural Hazards Events: What Twitter May Contribute to Situational Awareness. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. doi:10.1145/1753326.1753486
- Walther, M., & Kaiser, M. (2013). Geo-spatial Event Detection in the Twitter Stream. In P. Serdyukov, P. Braslavski, S. O. Kuznetsov, J. Kamps, S. Rüger, E. Agichtein, & E. Yilmaz et al. (Eds.), *Lecture Notes in Computer Science: Vol. 7814. Advances in Information Retrieval* (pp. 356–367). doi:10.1007/978-3-642-36973-5_30
- White, J., Palen, L., & Anderson, K. (2014). Digital Mobilization in Disaster Response: The Work & Self-Organization of On-Line Pet Advocates in Response to Hurricane Sandy. In *Proceedings of the ACM 2014 Conference on Computer Supported Cooperative Work*. doi:10.1145/2531602.2531633
- Zhao, D., & Rosson, M. (2009). How and why people Twitter: the role that micro-blogging plays in informal communication at work. *Proceeding GROUP '09 Proceedings of the ACM 2009 international conference on Supporting group work*. doi:10.1145/1531674.1531710
- Zielinski, A., Bügel, U., Middleton, L., Middleton, S. E., Tokarchuk, L., Watson, K., & Chaves, F. (2012). Multilingual Analysis of Twitter News in Support of Mass Emergency Events. In *Proceedings of the 9th International Conference on Information Systems for Crisis Response and Management (ISCRAM)*.

ADDITIONAL READING

- Brunner, D., Lemoine, G., Thoorens, F. X., & Bruzzone, L. (2009). Distributed Geospatial Data Processing Functionality to Support Collaborative and Rapid Emergency Response. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2(1), 33–46. doi:10.1109/JSTARS.2009.2015770
- Chun, Y., Hwang, H. S., & Kim, C. S. (2014). Development of a Disaster Information Extraction System based on Social Network Services. *International Journal of Multimedia and Ubiquitous Engineering*, 9(1), 255–264. doi:10.14257/ijmue.2014.9.1.24
- Earle, P., Guy, M., Buckmaster, R., Ostrum, C., Horvath, S., & Vaughan, A. (2010). OMG Earthquake! Can Twitter Improve Earthquake Response? *Seismological Research Letters*, 81(2), 246–251. doi:10.1785/gssrl.81.2.246

KEY TERMS AND DEFINITIONS

Data Mining: A computational process of discovering patterns in large data sets. The main goal is to extract knowledge from a data set and transform it into an understandable structure for further use (i.e. machine learning or predictive analytics).

Hashtag: Annotation format represented by the “#” symbol, used to indicate with a single word (or a combination of words) the core meaning of a tweet.

Meta-Data: Data providing information about a certain item’s content. For example, a tweet may include meta-data specifying GPS coordinates, or information about when it was sent, by who, and so on, which can be analyzed to extract meaningful information.

Situation Awareness: Pre-requisite state of knowledge and cognition of events useful for making decisions in situations involving uncertainty and crisis.

Tweet: Message posted on Twitter which may contain up to 140 characters of text, photos, videos, links, and so on.

Twitter APIs: A set of software libraries used by developers to extract data from Twitter streams. Twitter provides different types of software libraries (e.g. REST APIs, Streaming APIs, etc.), each one with its own specific function.

Web Crawler: Software program able to automatically analyze the content of web pages in order to recognize useful information.

Category D

Decision Support Systems

Cognitive Process Elements of People Decision-Making

Thais Spiegel

Rio de Janeiro State University, Brazil

INTRODUCTION

With computers and the interest in information processing, a new image of man begins to emerge. In the 1950s, studies of Broadbent (1954), continuing the Cherry (1953) model, culminate in the model of human thought processes. A model that began with the information received by the senses, but concentrated on a new and important feature: the individual has a limited capacity for receiving and storing information (Gardner, 2003).

These results were related to the work of George Miller. Miller (1956) in “The magical number seven, plus or minus two: some limits on our capacity for processing information”, when referring to classification and coding, indicates that there are limitations to the ability to process sensory signals on the order of about seven. At the same time, Bruner led the “Cognition Project” at Harvard. Faced with the observations of human performance in tasks of formation and acquisition of concepts, Bruner, Goodnow and Austin (1956) suggest that individuals suffered a state of “cognitive strain” and tried to reduce it through strategies of simplification.

In the study of decision-making, the classical view of behavioral appropriateness or rationality was also challenged by psychological reasons. One major example was Simon’s theory (1957) of “bounded rationality”, in which it is proposed that cognitive limitations lead decision-makers to construct simplified models for dealing with the world. In the same study, Simon (1957) suggests several cognitive strategies, the average, the sum and subtraction, to explain the behavior of different economic agents. The best known example is the

“satisfaction” one, which explains the behavior of consumers looking for a “good enough” option in an uncertain environment, where the search for alternatives is costly. To satisfy is a heuristic decision which involves choosing the first alternative that meets its minimum requirements. To satisfy is simple in terms of cognitive operators; therefore it makes smaller demands to the scarce mental resources. However, it can lead to sub-optimal behavior, given that finding an acceptable option, the search and evaluation of other alternatives, possibly better ones, ceases.

When asked by Gigerenzer about why bounded rationality is not the same as irrationality, Simon responded with an analogy. “Bounded rationality is like scissors: the mind is a blade and the structure of the environment is the other blade. To understand the behavior, we have to look at both and in how they fit.” (Simon, 1990:7 *apud* Gigerenzer, 2004:397). In other words, in order to assess the cognitive strategies as rational or irrational, one must also consider the environment because a strategy is rational or irrational only with respect to a physical or social environment (Simon, 1990). Thus, models of bounded rationality bring the reality and consider how human with little time and knowledge behave. This term, coined by Simon (1955) is associated with three distinct programs: the study of optimization under constraints, the study of cognitive illusions and the study of fast and frugal heuristics.

In the search for descriptions that adhere more closely to the human decision-making process, this text briefly contextualizes the human perspective in the study of decision-making. It then proceeds, on the basis of a review of the literature

on cognition during decision-making, to propose the formulation of a model that identifies the roles of aspects of cognition, and their inter-relations, during the decision-making process.

BACKGROUND

Some models of decision-making involved in descriptive approaches, approximating to how decisions are really taken. In the heuristics and biases program, the focus is on the individual's process to reach a conclusion, i.e. in the judgment that leads to the decision. It can be said that the central input is to indicate the existence of particular types of processing, covering certain areas in greater or lesser extent, and that in their function there is a set of trends in decisions made. It is relevant to point out that people do not decide rationally. It refers to a set of evidences of bounded rationality and systematic deviation from the optimal model, the expected decisions. However, it fails to explain why the heuristics manifest, how they operate in cognitive terms.

The program of fast and frugal heuristics is particularly interesting as it locates individuals in the context, and considers the rooted empirical knowledge that they have. With the ecological rationality reinforces the kinds of decision-making process before which the heuristic, of specific domain, change. This direction to the context is done without losing the object of analysis - the decision-maker. It goes on, however, without explaining how the cognitive process works. Explanations relating the types of environments with cognitive functions are not formulated, for example.

The dual processing models advance in attempts to explain the processing operation "within" the individual's head. They contribute, fundamentally, to mark the existence of non-rational processes, guided by intuition. It is also important to strengthen processes under which people have no control and awareness. It features a large contribution, which goes toward closer

cognitive processes. It remains, however, without explanation as to the correlation between the actual functioning of processes and the types of problems that are presented to individuals.

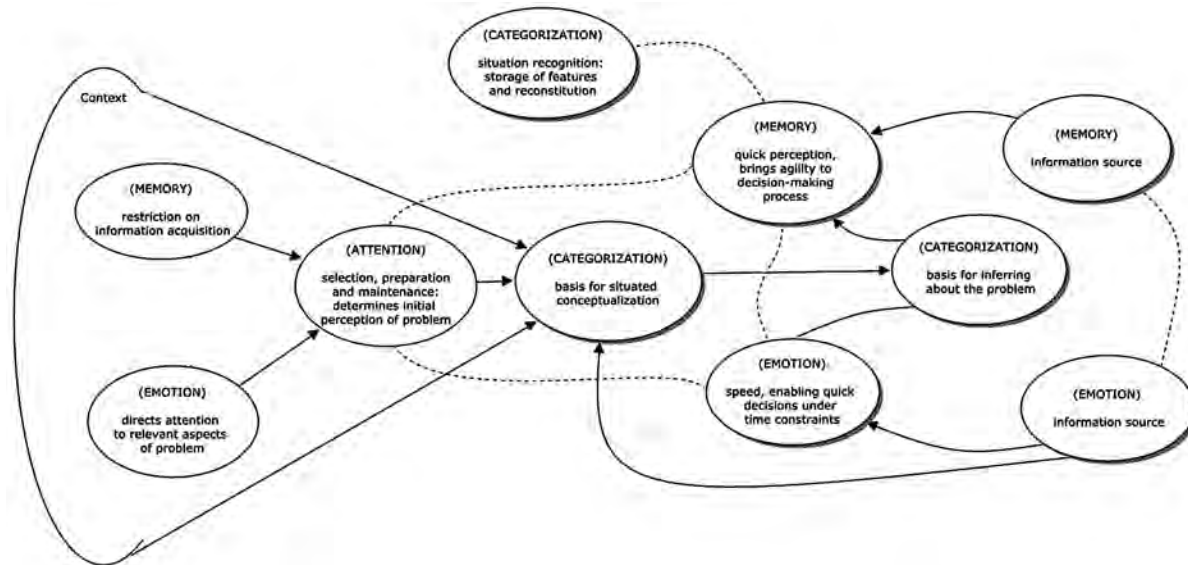
With cognitive models, the advance continues, recognizing structures formed by individuals that make up their understanding, assessment, and access or not to certain information and previous experiences. Particular aspects of each individual are emphasized as attention focuses on the formation of "schemes", which carry the history of the decision-maker.

The ramifications of this study lead to sense-making, through which one has contact with the attribution of meaning, and reinforces the unconscious processes. The concept of the production of retrospective explanations is of great contribution to the Decision Theory. The existence of the framing phenomenon commands researchers' attention to variations that cause information to be better or worse perceived and evaluated. The existence of a coding process in information processing that shapes the perception of the value of information is evidenced in this theoretical framework.

However, advances in the understanding of the cognitive process have been partly incorporated into organizational applications. Although cognition figures among the earliest focuses for research (Thagard et al., 2007), it is still a subject with many open questions in all the fields where it is considered. Very recently (as compared with other fields) organization-related inquiries entered this group of interested parties (Spiegel, 2013).

In order to understand decision-making "completely", and improve it, the underlying decision-making processes and the variables that affect the process must be examined (Roberts, 2002). Ola Svenson writes: "Human decision making cannot be understood simply by studying final decisions. The perceptual, emotional, and *cognitive* process which ultimately lead to the choice of a decision alternative must also be studied if we want to gain an adequate understanding of human decision making." (Svenson, 1979 apud Roberts, 2002, p.6)

Figure 1. Relationship among the roles of elements of cognition
 Source: Spiegel (2014).



On this question, Simon (1985, p.303) argues that: “Nothing is more fundamental in setting our research agenda and informing our research methods than our view of the nature of the human beings whose behavior we are studying”. Even when the focus is the outcome of the decision-making, ultimately it is the process that leads to it and allows it to be defined. While previous research was concerned with decision-making inputs and outputs, attention is now shifting to probe inside the “black box”, to examine the real processes involved in generating results from inputs (Roberts, 2002).

In that context, the hypothesis of this study is that the understanding of decision-making agents’ cognitive processes gives the basis necessary for understanding their decisions better. Once they are understood, that is believed to give access to the foundation underpinning description of human decisions (Spiegel, 2014).

MAIN FOCUS

Given the importance of opening the “black box”, this article presents the results of a review of the

related literature. It offers a structured presentation of the present state of our understanding of how cognition functions, and of its roles in relation to decision-making. From the literature, five elements - attention, consciousness, categorization, memory and emotion - were identified as representative aspects of cognition for understanding the decision-making process. In the search and model conception process, it was decided to work with consciousness as an element cross cutting the others. Accordingly, the state of (un)consciousness during the manifestation of the other aspects will be addressed when discussing the roles involved. Figure 1 summarizes the roles identified in the review presented in Spiegel (2014).

Table 1 summarizes an enhanced of the roles identified in the review presented in Spiegel (2014).

Below, a brief summary is given of the model constructed in the light of the roles set out in Spiegel (2014). Attention acts to selecting stimuli from the external and internal environments, and is limited by memory capacity, which acts to restrict information acquisition. Decision makers’ perceptions are shaped by their storage capacity and guided by emotional mechanisms “activated”

Table 1. Summary of roles played by elements of cognition

Cognition Role		Relevant Sources
Attention	Attention as selection, preparation and maintenance: determines initial perception of the problem	LaBerge (1999); Lachter et al. (2004); Posner & Rothbart (2007); Weber & Johnson (2009); Nejad (2014).
Categorization	Categorization as the basis for recognizing the situation: storage of the characteristics and the reconstitution	Kahneman & Tversky (2000); Miller & Cohen (2001); Cree & McRae (2003). Barsalou (2003, 2005); Cohen & Lefebvre (2005); Damásio (2005); Thagard et al. (2007).
	Categorization as the basis for situating perceptions	Smith et al. (2004); Barsalou (2003, 2005); Kourtzi (2010); McNamee, Rangel, O'Doherty (2013).
	Categorization as the basis for inferring about the problem	Barsalou (2005); McNamee, Rangel, O'Doherty (2013).
Memory	Memory as information source	Guenther (2002); Roberts (2002); Gonzalez (2013); Gutzwiller & Clegg (2013).
	Memory as constraint on information acquisition	Kareev et al. (1997); Rakow et al. (2010); Gutzwiller & Clegg (2013).
	Memory as bases for rapid perception, making for an agile decision-making process	Gupta et al. (2009); Gonzalez (2013); Gutzwiller & Clegg (2013).
Emotion	Emotion as information source	Mellers (2000); Slovic et al. (2002); Schwarz (2002); Pfister & Böhm (2008); Sun & Mathews (2012).
	Emotion in steering attention to important aspects of the problem	Berthoz (2006); Peters (2006); Pfister & Böhm (2008); Weber & Johnson (2009), Lerner et al. (2015).
	Emotion with a role in speed, permitting quick decisions when time is short	Bechara et al. (1997); Damásio (2005); Thagard et al. (2007), Brosch et al. (2013).

Source: Enhanced Spiegel (2014).

in the presence of the stimuli. Attention mechanisms bring not only accuracy, but also speed, to perception, preparing the system for information processing. In a manner integrated with the categorization process, in some situations, processing is accompanied by the maintenance of attention to certain stimuli.

In addition to this perceived information, which is discussed in the traditional decision-making literature, this present study emphasizes the importance of the decision maker's memory and emotions as conscious or subconscious sources of information. Emotional experiences at the various different stages of the decision-making process distinguished by Loewenstein & Lerner (2003) play a valuable role for decision makers, leaving them ready to react with pleasure or displeasure to the courses to be taken. It must be stressed that the experiences they are subjected to previously are critical here.

Returning to the internal or external stimuli, consciously or not, categorization plays a promi-

nent role by confronting them with patterns, enabling decision makers to recognize situations, which entails dimensions of storage, but also mental reconstitution, through mechanisms for simulating the experienced situation. The way the situation is presented activates distinct mechanisms with specific modes with regard to the role of categorization as situated perception, unlike interpretations by single mechanism acting in a decontextualized fashion.

The stored situations and subsequent simulations create a basis in categorization for inferences about problems. They thus speed the decision-making process. Memory and emotion as an information source are important influences in bringing greater agility to decision making. As information sources, these mechanisms, by establishing emotional relationships and reinforcement cycles in access to stored information, automate some steps of processing.

FUTURE RESEARCH DIRECTIONS

Future initiatives within the role of cognition in decision making include the answers that are observed whenever people face decisions from the experience and the description. The fact of the information that is relevant for the decision be necessarily built over the time, makes the decisions from the experience to be a good way to explore the sequential information processing (Barron & Leider, 2010). In addition, the “wealth” of the paradigm of decision from the experience with its focus on the acquisition and integration of information previous to the choice as well as the act of the choice, is available to more complete investigations in the individual differences in the cognitive processes (Rakow & Newell, 2010).

It became evident in the behavioral decision literature that these different approaches for the experimental study of decision-making, differ not only in the nature of the information provided but also in the way the acquired information is used to guide decisions (Hertwig *et al.*, 2004). In spite of this growing behavioral evidence, neurobiological studies of decision-making, haven’t been able to solve the way the differences of presentation may potentially influence the neural representations underlying these variables (Jessup & O’Doherty, 2010). Results indicate the decision variables generated based on the specific provision of descriptive information are not processed neurally the same way the decision variables acquired based on learning by trial and error. It’s important notice, that the brain systems identified in the two different conditions are not likely to be exclusively involved in one or other kind of decision-making (Barron & Leider, 2010). For instance, it was identified in other researches that the prefrontal cortex ventromedial represents decision values also under conditions that involves the providing of descriptive information. While MRI studies have found there is correlation of the activity in the ventral striatum with the prediction errors, a key sign to play a role in the learning by trial and error, and particularly in the associations between stimulus-result. A similar history is valid for risk

signs in the anterior insula and anterior cingulated (Chib *et al.*, 2009).

CONCLUSION

The decision-making study has as a target the understanding of the human ability of process multiple alternatives and choose an action course. Given the importance of decision makers in decision making, as examined in this chapter, these final remarks are intended to stress this aspect and to attest to the relevance of the contribution of the model formulated to yield descriptions closer to the realities of decision. The problem’s perception and/or the responsibility attribution by the resolution the decision-maker’s responsibility. The decision-maker is the individual or group on who lays the responsibility of dealing with the presented problem and, therefore, some fundamental factors linked to them change the way of how the decision occurs. The number of decision-makers, as well as the quality and the personal preference, defines the solution of the decision-making process.

Decisions are not equal and neither have the same difficulty degree, but they are all dependent on the mental model of their decision-makers. The agents may be evaluated by their individual knowledge about the problem, the resolution ability of each one of them, their desires and interests and the ethical and moral standards that guide them. Recognize the variety among the decision-makers is also necessary, since the performance will be determined by the adjustment of the tasks’ demands with the capacity, knowledge and the willingness to use effort of the decision-maker.

This chapter contributes to the body of knowledge on the subject by identifying and describing four elements of cognition that are representative for the human *decision-making* process. These elements interact, which also conditions the way the *decision-making* process occurs. Figure 1 shows the relationship among the “roles” of *attention*, *categorization*, *memory* and *emotion* over the course of *decision-making*.

REFERENCES

- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (pp. 89-195). Academic Press. doi:10.1016/S0079-7421(08)60422-3
- Barron, G., & Leider, S. (2010). The role of experience in the Gamblers Fallacy. *Journal of Behavioral Decision Making*, 23(1), 117-129. doi:10.1002/bdm.676
- Barsalou, L. W. (2003). Situated simulation in the human conceptual system. *Language and Cognitive Processes*, 18(5-6), 513-562. doi:10.1080/01690960344000026
- Barsalou, L. W. (2005). Situated Conceptualization. In H. Cohen & C. Lefebvre (Eds.), *Handbook of categorization in cognitive science* (pp. 619-650). Elsevier. doi:10.1016/B978-008044612-7/50083-4
- Bechara, A., Damásio, H., Damásio, A. R., & Lee, G. P. (1999). Different contributions of the human amygdala and ventromedial pre-frontal cortex to decision-making. *The Journal of Neuroscience*, 19, 5473-5481. PMID:10377356
- Bechara, A., Damásio, H., Tranel, D., & Damásio, A. R. (1997). Deciding advantageously before knowing the advantageous strategy. *Science*, 275(5304), 1293-1295. doi:10.1126/science.275.5304.1293 PMID:9036851
- Berthoz, A. (2006). *Emotion and Reason: The Cognitive Neuroscience of Decision Making*. Oxford University Press.
- Broadbent, D. (1954). *Perception and Communication*. London: Pergamon Press.
- Brosch, T., Scherer, K. R., Grandjean, D., & Sander, D. (2013). The impact of emotion on perception, attention, memory, and decision making. *Swiss Medical Intelligence*, 143, w13786. PMID:23740562
- Bruner, J., Goodnow, J., & Austin, A. (1956). *A Study of Thinking*. New York: Wiley.
- Cherry, E. C. (1953). Some experiments on the recognition of speech with one and two ears. *The Journal of the Acoustical Society of America*, 25(5), 975-979. doi:10.1121/1.1907229
- Chib, V. S., Rangel, A., Shimojo, S., & O'Doherty, J. P. (2009). Evidence for a common representation of decision values for dissimilar goods in human ventromedial prefrontal cortex. *The Journal of Neuroscience*, 29(39), 12315-12320. doi:10.1523/JNEUROSCI.2575-09.2009 PMID:19793990
- Cohen, H., & Lefebvre, C. (2005). *Handbook of Categorization in Cognitive Science*. Elsevier Science.
- Cree, G. S., & Mcrae, K. (2003). Analyzing the factors underlying the structure and computation of the meaning. *Journal of Experimental Psychology. General*, 132(2), 163-201. doi:10.1037/0096-3445.132.2.163 PMID:12825636
- Damásio, A. (2005). *Descartes' Error: Emotion, Reason, and the Human Brain*. Penguin.
- Gardner, H. (2003). *A nova ciência da mente: uma história da revolução cognitiva*. São Paulo: Editora da Universidade de São Paulo.
- Gigerenzer, G. (2004). Striking a Blow for Sanity in Theories of Rationality. In M. Augier & J. G. March (Eds.), *Models of a Man: Essays in Memory of Herbert A. Simon* (pp. 389-410). The MIT Press.
- Gonzalez, C. (2013). The boundaries of instance-based learning theory for explaining decisions from experience. In *Progress in Brain Research*. Elsevier. doi:10.1016/B978-0-444-62604-2.00005-8
- Guenther, R. K. (2002). Memory. In D. J. Levitin (Ed.), *Foundations of Cognitive Psychology: Core Readings* (pp. 311-359). The MIT Press.

- Gupta, R., Duff, M., Denburg, N., Cohen, N. J., Bechara, A., & Tranel, D. (2009). Declarative memory is critical for sustained advantageous complex decision-making. *Neuropsychologia*, 47(7), 1686–1693. doi:10.1016/j.neuropsychologia.2009.02.007 PMID:19397863
- Gutzwiller, R. S., & Clegg, B. A. (2013). The Role of Working Memory in Levels of Situation Awareness. *Journal of Cognitive Engineering and Decision Making*, 7(2), 141–154. doi:10.1177/1555343412451749
- Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science*, 15(8), 534–539. doi:10.1111/j.0956-7976.2004.00715.x PMID:15270998
- Hodgkinson, G. P., & Starbuck, W. H. (2008). *The Oxford Handbook of Organizational Decision Making*. Oxford University Press. doi:10.1093/oxfordhb/9780199290468.001.0001
- Jessup, R. K., & ODoherty, J. P. (2010). Decision Neuroscience: Choices of Description and of Experience. *Current Biology*, 20(20), 881–883. doi:10.1016/j.cub.2010.09.017 PMID:20971429
- Kahneman, D., & Tversky, A. (2000). *Choices, Values and Frames*. Cambridge University Press.
- Kareev, Y., Lieberman, I., & Lev, M. (1997). Through a narrow window: Sample size and the perception of correlation. *Journal of Experimental Psychology. General*, 126(3), 278–287. doi:10.1037/0096-3445.126.3.278
- Keren, G. (1996). Perspectives of Behavioral Decision Making: Some Critical Notes. *Organizational Behavior and Human Decision Processes*, 65(3), 169–178. doi:10.1006/obhd.1996.0016
- Kourtzi, Z. (2010). Visual learning for perceptual and categorical decisions in the human brain. *Vision Research*, 50(4), 433–440. doi:10.1016/j.visres.2009.09.025 PMID:19818361
- LaBerge, D. (1999). Attention. In B. M. Bly & D. E. Rumelhart (Eds.), *Cognitive Science* (pp. 43-97). Academic Press.
- Lachter, J., Forster, K. I., & Ruthruff, E. (2004). Forty-five years after Broadbent 1958: Still no identification without attention. *Psychological Review*, 111(4), 880–913. doi:10.1037/0033-295X.111.4.880 PMID:15482066
- Lerner, J.S., Li, Y., Valdesolo, P., & Kassam, K. (2015). Emotion and decision making. *Annual Review of Psychology*, 66(33), 1–33.
- Loewenstein, G., & Lerner, J. S. (2003). The role of affect in decision making. In R. Davidson, K. Scherer & H. Goldsmith (Eds.), *Handbook of affective science* (pp. 619-642). Oxford University Press.
- McNamee, D., Rangel, A., & ODoherty, J. (2013). Evidence for category-dependent and category-independent goal-value codes in human ventromedial prefrontal cortex. *Nature Neuroscience*, 16(4), 479–485. doi:10.1038/nn.3337 PMID:23416449
- Mellers, B. A. (2000). Choice and the relative pleasure of consequences. *Psychological Bulletin*, 126(6), 910–924. doi:10.1037/0033-2909.126.6.910 PMID:11107882
- Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, 24(1), 167–202. doi:10.1146/annurev.neuro.24.1.167 PMID:11283309
- Miller, G. (1956). The magical number seven plus or minus two: Some limits of our capacity for the processing of information. *Psychological Review*, 63(2), 81–97. doi:10.1037/h0043158 PMID:13310704
- Miller, S. J., & Wilson, D. C. (2006). Perspectives on Organizational Decision-Making. In S. Clegg, C. Hardy, T. Lawrence & W. R. Nord (Eds.), *The SAGE Handbook of Organization Studies* (pp. 469-484). Sage Publications Ltd. doi:10.4135/9781848608030.n16

- Myachykov, A., & Tomlin, R. S. (2015). Attention in the organization and development of Grammar. In *Handbook of Cognitive Linguistics* (pp. 31-53). MARC Record: HSK.
- Nejad, M. H. (2014). *Attention to Competition: The role of managerial cognition in shaping the response to competitive actions* (Thesis). The University of Ontario.
- Peters, E. (2006). The functions of affect in the construction of preferences. In S. Lichtenstein & P. Slovic (Eds.), *The construction of preference* (pp. 454-463). Cambridge University Press. doi:10.1017/CBO9780511618031.025
- Pfister, H. R., & Böhm, G. (2008). The multiplicity of emotions: A framework of emotional functions in decision making. *Judgment and Decision Making*, 3(1), 5-17.
- Posner, M. I., & Rothbart, M. K. (2007). Research on attention networks as a model for the integration of psychological science. *Annual Review of Psychology*, 58(1), 1-23. doi:10.1146/annurev.psych.58.110405.085516 PMID:17029565
- Rakow, T., & Newell, B. (2010). Degrees of uncertainty: An overview and framework for research on experience-based choice. *Journal of Behavioral Decision Making*, 23(1), 1-14. doi:10.1002/bdm.681
- Rakow, T., Newell, B., & Zougkou, K. (2010). The role of working memory in information acquisition and decision making: Lessons from the binary prediction task. *Quarterly Journal of Experimental Psychology*, 63(7), 1335-1360. doi:10.1080/17470210903357945 PMID:20013514
- Roberts, F. D. (2002). *The effects of decision aid recommendations on users' cognitive processes, memories, and judgments*. The University of Tennessee.
- Schwarz, N. (2002). Emotion, cognition, and decision making. *Cognition and Emotion*, 14(4), 433-440. doi:10.1080/026999300402745
- Shafir, E. (1999). Decision Making. In F. Keil & R. Wilson (Eds.), *The MIT Encyclopedia of the Cognitive Sciences* (pp. 220-223). MIT Press.
- Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, 69(1), 99-118. doi:10.2307/1884852
- Simon, H. A. (1957). *Models of Man*. Nova York. John Wiley.
- Simon, H. A. (1985). Human nature in politics. *The American Political Science Review*, 79(02), 293-304. doi:10.2307/1956650
- Simon, H. A. (1990). Invariants of human behavior. *Annual Review of Psychology*, 41(1), 1-19. doi:10.1146/annurev.ps.41.020190.000245 PMID:18331187
- Slovic, P., Finucane, M. L., Peters, E., & MacGregor, D. G. (2002). The affect heuristic. In T. Gilovich, D. Griffin & D. Kahneman (Eds.), *Intuitive judgment: Heuristics and biases* (pp. 397-420). Cambridge University Press. doi:10.1017/CBO9780511808098.025
- Smith, M. L., Gosselin, F., & Schyns, P. G. (2004). Receptive fields for flexible face categorizations. *Psychological Science*, 15(11), 753-761. doi:10.1111/j.0956-7976.2004.00752.x PMID:15482447
- Spiegel, T. (2013). *Cognitive sciences contributions to operations management: analysis of experience impact in operations manager*. Federal University of Rio de Janeiro.
- Spiegel, T. (2014). An Overview of Cognition Roles in Decision-Making. In J. Wang (Ed.), *Encyclopedia of Business Analytics and Optimization* (pp. 74-84). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-5202-6.ch008
- Sun, R., & Mathews, R. (2012). Implicit cognition, emotion, and meta-cognitive control. *Mind & Society*, 11(1), 107-119. doi:10.1007/s11299-012-0101-5

Thagard, P., Gabbay, D. M., & Woods, J. (2007). *Philosophy of Psychology and Cognitive Science*. North Holland.

Weber, E. U., & Johnson, E. J. (2009). Mindful Judgment and Decision Making. *Annual Review of Psychology*, 60(1), 53–85. doi:10.1146/annurev.psych.60.110707.163633 PMID:18798706

KEY TERMS AND DEFINITIONS

Attention: Attention refers to the ability of the individual to respond predominantly stimuli that are significant to the detriment of others.

Categorization: Categorization is the mental operation by which the brain classifies objects and events, comes the ability to group sensory events into meaningful categories.

Cognition: The concept of cognition comes to all capture processes of external stimuli through sensory resources and processing, reduction, storage, retrieval and use of these stimuli.

Consciousness: Consciousness is associated with the ability of human beings to get information itself, as well as get other people and objects.

Decision: The decision-making is understood as the deliberate choice of a course of action with the intention of producing a desired result.

Emotion: Emotions are mental operations are accompanied by a previous experience, capable of guiding the behavior and conduct physiological adaptations necessary.

Memory: The memory comprises a number of biological strategies and anatomical substrates; involves a complex mechanism that covers the storage and retrieval of experiences thus is closely associated with learning.

Comprehensible Explanation of Predictive Models

D

Marko Robnik-Šikonja
University of Ljubljana, Slovenia

INTRODUCTION

In many areas where machine learning methods are applied the practitioners and users of produced prediction models are interested in comprehensible explanation of their predictions. Unfortunately, the best performing predictive models do not offer an intrinsic introspection into their decision processes or provide explanations of their prediction. This is true for Support Vector Machines (SVM), Artificial Neural Networks (ANN), and all ensemble methods (for example, boosting, random forests, bagging, stacking and multiple adaptive regression splines). Approaches that do offer an intrinsic introspection such as decision trees or decision rules do not perform so well or are not applicable in many cases (Meyer et al., 2003). The areas where models' transparency is of crucial importance are for example most of business and marketing applications where the executives are just as interested in the comprehension of the decision process, explanation of the existing and new customers' needs and expectations in a given business case, as in the classification accuracy of the prediction model. The same is true for many areas of business intelligence, finance, marketing, insurance, medicine, science, policy making, and strategic planning where knowledge discovery dominates prediction accuracy.

To alleviate this problem two types of solutions have been proposed. The first type is based on internal working of each particular learning algorithm and exploits its learning process to gain insight into the presumptions, biases and reasoning leading to final decisions. A well-known example of such an approach are random forests for which several

visualizations exist mostly exploiting the fact that during bootstrap sampling some of the instances are not selected for learning and can serve as an internal validation set. With the help of this set important features can be identified and similarity between objects can be measured. The second type of explanation approaches are general and can be applied to any predictive model. Examples of this approach are methods EXPLAIN (Robnik-Šikonja & Kononenko, 2008) and IME (Štrumbelj et al., 2009). These two methods are based on efficient implementation of input perturbations. They can explain models' decision process for each individual predicted instance as well as the model as a whole. As both methods are efficient, offer comprehensible explanations, can be visualized, and are readily available in R package ExplainPrediction (Robnik-Šikonja, 2015) they are the focus of this article. Other general explanation methods are discussed in the background Section.

The objective of the article is to explain how EXPLAIN and IME explanation methods work and to show their practical utility in several real world scenarios. The first aim is achieved through explanation of their working principle and graphical explanation of models' decisions on a well-known data set. Two types of explanations are demonstrated, predictions of new unlabeled cases and the functioning of the model as a whole. This allows inspection, comparison, and visualization of otherwise opaque models. The practical utility of the methodology is demonstrated with short description of several applications: in medicine (Štrumbelj et al. 2010), macro economy (Pregeljc et al., 2012) and business consultancy (Bohanec et al., 2015).

DOI: 10.4018/978-1-5225-2255-3.ch181

BACKGROUND

In a typical data science problem setting, users are concerned with both prediction accuracy and the interpretability of the prediction model. Complex models have potentially higher accuracy but are more difficult to interpret. This can be alleviated either by sacrificing some prediction accuracy for a more transparent model or by using an explanation method that improves the interpretability of the model. Explaining predictions is straightforward for symbolic models such as decision trees, decision rules, and inductive logic programming, where the models give an overall transparent knowledge in a symbolic form. Therefore, to obtain the explanations of predictions, one simply has to read the rules in the corresponding model. Whether such an explanation is comprehensive in the case of large trees and rule sets is questionable.

For non-symbolic models there are no such straightforward explanations. A lot of effort has been invested into increasing the interpretability of complex models. A taxonomy of explanation methods and a review of neural network explanation approaches is given by Jacobsson (2005). For Support Vector Machines an interesting approach is proposed by Hamel (2006). Many approaches exploit the essential property of additive classifiers to provide more comprehensible explanations and visualizations, e.g., (Jakulin et al., 2005) and (Poulin et al. 2006).

Visualization of decision boundaries is an important aspect of model transparency. Barbosa et al. (2016) present a technique to visualize how the kernel embeds data into a high-dimensional feature space. With their Kelp method they visualize how kernel choice affects neighborhood structure and SVM decision boundaries. Schultz et al. (2015) propose a general framework for visualization of classifiers via dimensionality reduction. Goldstein et al. (2015) propose another useful visualization tool for classifiers that can produce individual conditional expectation plots, graphing the functional relationship between the predicted response and the feature for individual instance.

Some explanations methods (including the ones presented here) are general in a sense that they can be used with any type of classification model (Lemaire et al. 2008; Robnik-Šikonja and Kononenko 2008; Štrumbelj et al., 2010). This enables their application with almost any prediction model and allows users to analyze and compare outputs of different analytical techniques. Lemaire et al., (2008) applied their method to a customer relationship management system in telecommunications industry. The method which successfully deals with high-dimensional text data is presented in (Martens and Provost, 2011). Its idea is based on general explanation methods presented here and offers explanation in the form of a set of words which would change the predicted class of a given document. Bosnić et al. (2014) adapt the general explanation methodology to data stream scenario and show the evolution of attribute contributions through time. This is used to explain the concept drift in their incremental model.

Many explanation methods are related to statistical sensitivity analysis and uncertainty analysis (Saltelli et al., 2000). In that methodology sensitivity of models is analysed with respect to models' input. A related approach, called inverse classification (Aggarwal et al. 2010) tries to determine the minimum required change to a data point in order to reclassify it as a member of a different class. A SVM model based approach is proposed by (Barbella et al., 2009).

Another sensitivity analysis-based approach explains contributions of individual features to a particular classification by observing (partial) derivatives of the classifiers prediction function at the point of interest (Baehrens et al. 2010). A limitation of this approach is that the classification function has to be first-order differentiable. For classifiers not satisfying this criterion (for example, decision trees) the original classifier is first fitted with a Parzen window-based classifier that mimics the original one and then the explanation method is applied to this fitted classifier. The method was shown to be practically useful with kernel based classification method to predict molecular features (Hansen et al., 2011).

MAIN FOCUS

General explanation methods can be applied to any classification model which makes them a useful tool both for interpreting models (and their predictions) and comparing different types of models. By modification of feature values of interest what-if analysis is also supported. Such methods cannot exploit any model-specific properties and are limited to perturbing the inputs of the model and observing changes in the model's output (Robnik-Šikonja and Kononenko 2008; Lemaire et al. 2008, Štrumbelj et al., 2010).

True causal relationships between the dependent and independent variables are typically hidden except in artificial domains where all the relations as well as the probability distributions are known in advance. Therefore only explanations of prediction process for a particular model is of practical importance. The prediction accuracy and the correctness of explanation for a given model may be orthogonal: the correctness of the explanation is independent of the correctness of the prediction. However, empirical observations show that better models (with higher prediction accuracy) enable better explanations (Štrumbelj et al., 2009). Two types of explanations are discussed:

- *Instance explanation* explains predictions with the given model of a single instance and provides impact of input feature values on the prediction.
- *Model explanation* aggregates instance explanations over many (training) instances, which provides top-level explanations of features and their values. This aggregation over many instances enables identification of different roles attributes may play in the classifications of instances. To avoid loss of information due to summarization, in the presented visualization the evidence for and against each class is collected separately. In this way one can, for example, see that a particular value of an attribute supports specific class but not in every context.

Two general explanation methods are presented which provide the two types of explanations for prediction models. The key idea is that the contribution of a particular input value (or set of values) can be captured by "hiding" the input value (set of values) and observing how the output of the model changes. As such, the key component of general explanation methods is the expected conditional prediction - the prediction where only a subset of the input variables is known. Let Q be a subset of the set S of all input variables $Q \subseteq S = \{X_1, \dots, X_n\}$. Let $p_Q(y_k|x)$ be the expected prediction for x , conditional to knowing only the input variables represented in Q : $p_Q(y_k|x) = E(p(y_k|x) | X_i = x_{(i)}, \forall X_i \in Q)$.

Therefore, $p_S(y_k|x) = p(y_k|x)$. In practical settings the classification function of the model is not known - one can only access its prediction for any vector of input values. Therefore, exact computation of this equation is not possible and sampling-based approximations are used.

EXPLAIN: One-Variable-at-a-Time Approach

The method EXPLAIN computes the influence of a feature value by observing its impact on the model's output. The EXPLAIN assumes that the larger the changes in the output, the more important role the feature value plays in the model. The shortcomings of this approach is that it takes into account only a single feature at time, therefore it cannot detect certain higher order dependencies (in particular disjunctions) and redundancies in the model.

The EXPLAIN assumes that the characterization of the i -th input variable's importance for the prediction of the instance x is the difference between the model's prediction for that instance and the model's prediction if the value of the i -th variable was not known, namely: $p(y_k|x) - p_{S-\{i\}}(y_k|x)$. If this difference is large then the i -th variable is important. If it is small then the variable is less important. The sign of the difference reveals whether the value contributes towards or

against class value y_k . This approach was extended in (Robnik-Šikonja and Kononenko, 2008) to use log-odds ratios (or weight of evidence) instead of the difference in predicted class probabilities.

To demonstrate behavior of the method an example of an explanation is given. Let a binary domain contains three important (A_1 , A_2 , and A_3) and one irrelevant attribute (A_4), so the set of attributes is $S = \{1, 2, 3, 4\}$. Let us assume that the learned model correctly expresses the class value as the parity (xor) relation of three attributes $C = A_1 \oplus A_2 \oplus A_3$. The correct model would classify an instance $x = (A_1 = 1, A_2 = 0, A_3 = 1, A_4 = 1)$ to class $C = 0$, and assigns it probability $p(C = 0|x) = 1$. When explaining classification for this particular instance $p(C = 0|x)$, method EXPLAIN simulates the lack of knowledge of a single attribute at a time, so one has to estimate $p_{S-\{1\}}(C = 0|x)$, $p_{S-\{2\}}(C = 0|x)$, $p_{S-\{3\}}(C = 0|x)$, and $p_{S-\{4\}}(C = 0|x)$. Without the knowledge about the values of each of the attributes A_1 , A_2 , and A_3 , the model cannot correctly determine the class value, so the correct estimates of class probabilities are $p_{S-\{1\}}(C = 0|x) = p_{S-\{2\}}(C = 0|x) = p_{S-\{3\}}(C = 0|x) = 0.5$. The differences of probabilities $p(y_k|x) - p_{S-\{i\}}(y_k|x)$ therefore equal 0.5 for each of the three important attributes, which indicate that these attributes have positive impact on classification to class 0 for the particular instance x . The irrelevant attribute A_4 does not influence the classification, so the classification probability remain unchanged $p_{S-\{4\}}(C = 0|x) = 1$. The difference of probabilities $p_S(C = 0|x) - p_{S-\{4\}}(C = 0|x) = 0$ so the explanation of the irrelevant attribute's impact is zero.

IME: All-Subsets Approach

The one-variable-at-a-time approach is simple and computationally less-intensive but has some disadvantages. The main disadvantage is that disjunctive concepts or redundancies between input variables will result in unintuitive contributions for variables (Štrumbelj et al., 2009). A solution was proposed in (Štrumbelj and Kononenko, 2010), where all subsets of values are observed.

Such procedure demands 2^a steps and results in the exponential time complexity. However, the contribution of the variable corresponds to the Shapley value for the coalitional game of a players. This allows an efficient approximation based on sampling.

SOLUTIONS AND RECOMMENDATIONS

The working and practical utility of the one-variable-at-a-time contributions and their visualization are illustrated on the well-known Titanic data set. The task is to classify the survival of a passenger in the disaster of the Titanic ship. The three input variables report the passengers' status during travel (first, second, third class, or crew), age (adult or child), and gender. Note the similarity of the problem with many business decision problems, such as churn prediction, mail response, insurance fraud, etc. As an example PRBF classifier is used which is a special case of RBF network. Such models usually provide good prediction accuracy but are incomprehensible. Left-hand side of Fig. 1 shows an example of an *instance explanation* for the prediction of instance x (a first class adult male passenger). The text at the top includes the data set name, the model name, the class value in question, the model's prediction for this class value, and the actual correct class value for the instance. The input variables' names are shown on the left-hand side and their values for the particular instance are on the right-hand side. The thick dark shaded bars indicate the contributions of the instance's values for each corresponding input variable towards (or against) the class value "survived=yes". The thinner and lighter bars above indicate average contributions of these values across all instances. For the given instance one can observe that "sex = male" speaks strongly against survival and "status=first class" speaks in favor of survival, while being an adult has little influence. Thinner average bars above them reveal that being male is on average less

dangerous and being in the first class is even more beneficial than in the selected case. Note that the same visualization can be used even if some other classification method is applied.

A more general view of the model is provided by averaging the explanations for the training data and their visualization in a summary form, which shows average importance of each input variable and its values. An example of such a *model explanation* for Titanic data set is presented in the right-hand side of Fig. 1. On the left-hand side the input variables and their values are shown. For each value the average negative and the average positive contributions across all instances is displayed. Note that negative and positive contributions would cancel each other out if summed together, so it is important to keep them separate. The lighter bars shown are equivalent to the lighter bars in the instance explanation on Fig. 1a. For each input variable the average positive and negative contribution across all values and instances is shown (darker bars). The visualization reveals that the crew status and gender have approximately the same effect while the age is less important in the PRBF model. Travelling in the first or second class has an exclusively positive contribution towards the survival, being a child or female has greater positive than negative contribution, travelling third class or being a male has a predominately negative contribution, while being a member of the crew or an adult has a minor negative contribution.

Real World Applications

To show a broad spectrum of possible practical applications of the described methodology, three applications from different fields are shortly described.

Breast Cancer Recurrence

In a real-life breast cancer recurrence prediction Štrumbelj et al., (2010) illustrate the usefulness of the visualizations and the advantage of using

the general explanation method. Several machine learning algorithms were evaluated. Predictions were enhanced with instance explanations using the IME method. Visual inspection and evaluation showed that oncologists found the explanations useful and agreed with the computed contributions of features. Examples of explanations provided by the non-transparent PRBF model, which achieves 73% classification accuracy, are presented in Figures 2. If oncologists chose to replace their existing prediction model with a better performing one, the form of visualizations would remain the same and would not affect the user-experience. This is an advantage of using a general explanation method.

Enterprise Performance

Pregeljc et al., (2012) used traditional modeling approaches together with data mining to gain insight into the connections between the quality of organization in enterprises and the enterprises' performance. The data set contained economic results and indicators of organizational quality for 72 Slovenian enterprises' in the time span of four years. The hypothesis that a causal relationship exists between the latter and the former was confirmed based on the analysis consisting of the use of several classification algorithms to study these relationships and the evaluation of the prediction accuracy of the companies' economic results. The best performing models were complex and difficult to interpret, especially for non-technical users. Therefore, the general explanation methods presented here were employed to explain the effect of individual features on the model's predictions. The results showed that data mining models are successful at modeling the dependency relationship. The explanation of the influence of the input features on the predicted economic results provided insights with a meaningful economic interpretation. The interesting economic relationships and successful predictions come mostly from complex models such as random forests and ANN. Without proper explanation and visualization, these models are often neglected in favor of

Figure 1. An instance explanation (on the left-hand side) and a model explanation (on the right hand side) for the PRBF model classifying the Titanic data set. The tiny bars in the instance explanation represent the average positive and average negative contributions of the values and are equal to the corresponding value-bars in the model explanation (note the difference in scale).

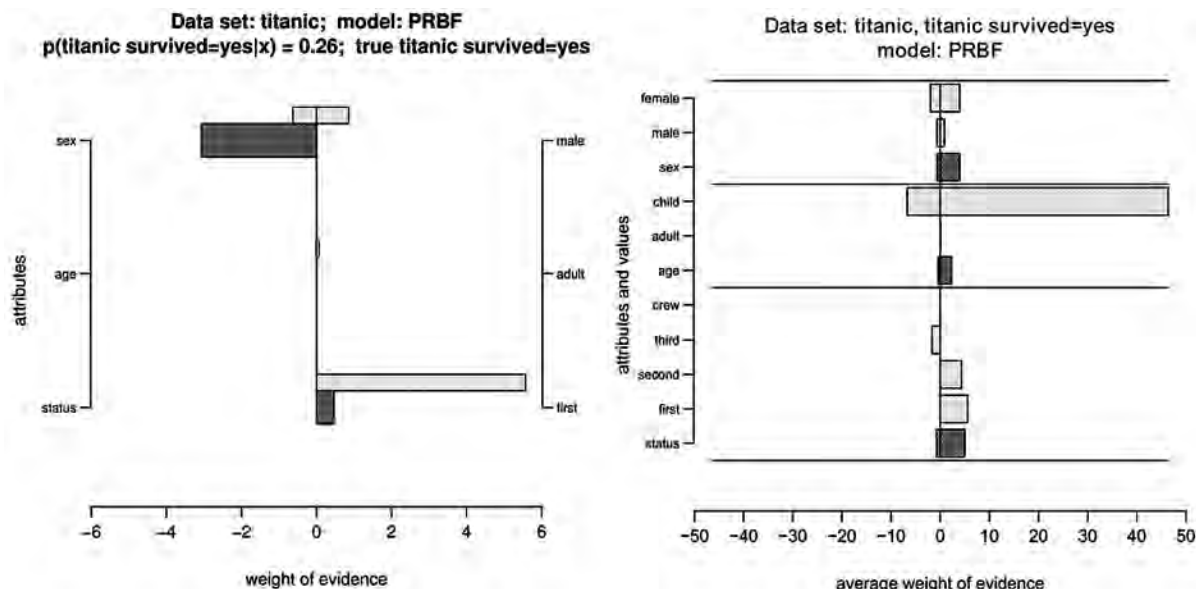
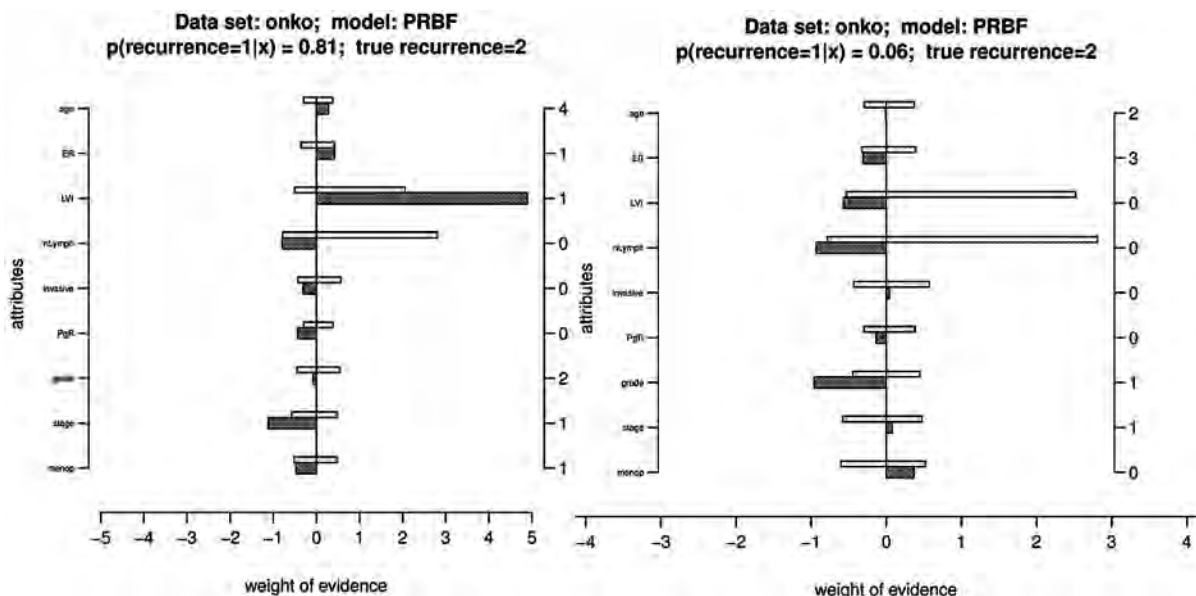


Figure 2. For the patient represented by the figure on the left-hand side the PRBF predicts a high likelihood of cancer recurrence (81%). While this prediction turned out to be incorrect as the cancer did not recur, the oncologist would have made the same prediction and agree that chance of recurrence is high due to lympho-vascular invasion (LVI = 1), while a low number of positive lymph nodes (nLymph = 0) and early stage of cancer (stage = 1) speak against recurrence. The patient on the right hand side is similar, but did not have lympho-vascular invasion. Therefore, the most relevant input variables speak against recurrence and the model correctly predicts a low chance of recurrence (6%).



weaker, but more transparent models. Experts from the economic-organizational field, which reviewed and interpreted the results of the study, agreed that such an explanation and visualization is useful and facilitates comparative analysis across different types of prediction models.

Forecasting Sales

Bohanec et al. (2015) constructed a real-world data set describing the sales history of a medium-sized company providing software solutions to clients in international B2B markets. The data set construction was motivated by possible improvements in understanding factors impacting the outcome of the sales process and improvements of the sales performance for new clients. Approximately 15 new (open) sales opportunities per month were analyzed with the help of an external consultant. Machine learning prediction models were used in knowledge discovery process to gain knowledge about the properties of decision process using general explanation methods EXPLAIN and IME. The analysis included explanations of individual decisions as well as the whole model. For new (open) cases decision makers were supported with the explanations to assess various scenarios with explanatory what-if analysis.

One can conclude that the explanation methodology presented is a useful tool for many different problems. It is especially important for problems where prediction performance has to be supplemented with model's transparency or knowledge discovery.

FUTURE RESEARCH DIRECTIONS

The explanation methodology presented has recently been implemented as R package ExplainPrediction (Robnik-Šikonja, 2015) making it freely available to data science community. The simplicity and elegance of the idea coupled with efficient implementation and visualization of instance and model based explanations will al-

low the application of the approach in new areas. Broader practical use shall spur additional research into explanation mechanisms and improvements in the visual design of explanations.

CONCLUSION

Two general methods for explanation of prediction models were presented. The methods allow explanation of individual decisions as well as the prediction model as a whole. Both of them can be efficiently computed and visualized. The explanations reveal how the individual input variables influence the outcome of otherwise completely opaque models, thus making it transparent and comprehensible. The general methods allow users to compare different types of models or replace an existing model without having to replace the explanation method. The explanation methods EXPLAIN and IME exhibit the following properties:

- **Instance Dependency:** Different instances are predicted differently, so the explanations will also be different.
- **Class Dependency:** Explanations for different classes are different, different attributes may have different influence on different classes (for two-class problems, the effect is complementary).
- **Model Dependency:** The decision process is taking place inside the model, so if the model is wrong for a given problem, explanation will reflect that and will be correct for the model, therefore wrong for the problem.
- **Capability to Detect Strong Conditional Dependencies:** If the model captures strong conditional dependency the explanations will also reflect that.
- EXPLAIN method is unable to detect and correctly evaluate the utility of attributes' values in instances where the *change in more than one attribute* value at once is needed to affect the predicted value. IME

method samples the space of feature interactions and therefore avoids this problem.

- **Visualization Ability:** The generated explanations can be graphically presented in terms of the positive/negative effect each attribute and its values have on the classification of the given instance.

REFERENCES

- Aggarwal, C. C., Chen, C., & Han, J. W. (2010). The inverse classification problem. *Journal of Computer Science and Technology*, 25(3), 458–468. doi:10.1007/s11390-010-9337-x
- Baehrens, D., Schroeter, T., Harmeling, S., Kawanabe, M., Hansen, K., & Müller, K. (2010). How to explain individual classification decisions. *Journal of Machine Learning Research*, 11, 1803–1831.
- Barbella, D., Benzaid, S., Christensen, J. M., Jackson, B., Qin, X. V., & Musicant, D. R. (2009). Understanding support vector machine classifications via a recommender system-like approach. In R. Stahlbock, S. F. Crone, & S. Lessmann (Eds.), *International Conference on Data Mining*. CSREA Press.
- Barbosa, A., Paulovich, F. V., Paiva, A., Goldenstein, S., Petronetto, F., & Nonato, L. G. (2016). Visualizing and Interacting with Kernelized Data. *IEEE Transactions on Visualization and Computer Graphics*, 22(3), 1314–1325. doi:10.1109/TVCG.2015.2464797 PMID:26829242
- Bohanec, M., Kljajić-Borštnar, M., & Robnik-Šikonja, M. (2015). Integration of machine learning insights into organizational learning: A case of B2B sales forecasting. In R. Bons, J. Versendaal, A. Pucihar, & M. Kljajić-Borštnar (Eds.), *Proceedings of the WellBeing conference*. Moderna organizacija.
- Bosnić, Z., Demšar, J., Kešpret, G., Rodrigues, P. P., Gama, J., & Kononenko, I. (2014). Enhancing data stream predictions with reliability estimators and explanation. *Engineering Applications of Artificial Intelligence*, 34, 178–192. doi:10.1016/j.engappai.2014.06.001
- Goldstein, A., Kapelner, A., Bleich, J., & Pitkin, E. (2015). Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation. *Journal of Computational and Graphical Statistics*, 24(1), 44–65. doi:10.1080/10618600.2014.907095
- Hamel, L. (2006). Visualization of support vector machines with unsupervised learning. *Proceedings of 2006 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology*. doi:10.1109/CIBCB.2006.330984
- Hansen, K., Baehrens, D., Schroeter, T., Rupp, M., & Müller, K.-R. (2011). Visual Interpretation of Kernel-Based Prediction Models. *Molecular Informatics*, 30(9), 817–826. doi:10.1002/minf.201100059 PMID:27467414
- Jacobsson, H. (2005). Rule extraction from recurrent neural networks: A taxonomy and review. *Neural Computation*, 17(6), 1223–1263. doi:10.1162/0899766053630350
- Jakulin, A., Možina, M., Demšar, J., Bratko, I., & Zupan, B. (2005). Nomograms for visualizing support vector machines. In R. Grossman, R. Bayardo, & K. P. Bennett (Eds.), *Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM. doi:10.1145/1081870.1081886
- Lemaire, V., Feraud, R., & Voisine, N. (2008). Contact personalization using a score understanding method. *Proceedings of International Joint Conference on Neural Networks (IJCNN)*. doi:10.1109/IJCNN.2008.4633863

Martens, D., & Provost, F. (2011). *Explaining Documents' Classifications*. Center for Digital Economy Research, New York University, Stern School of Business, working paper CeDER-11-01.

Meyer, D., Leisch, F., & Hornik, K. (2003). The support vector machine under test. *Neurocomputing*, 55(1), 169–186. doi:10.1016/S0925-2312(03)00431-4

Poulin, B., Eisner, R., Szafron, D., Lu, P., Greiner, R., Wishart, D. S., & Anvik, J. et al. (2006). Visual explanation of evidence with additive classifiers. In *Proceedings of AAAI'06*. AAAI Press.

Pregeljc, M., Štrumbelj, E., Mihelčič, M., & Kononenko, I. (2012). Learning and explaining the impact of enterprises' organizational quality on their economic results. In *Intelligent data analysis for real-life applications: Theory and practice*. IGI Global.

Robnik-Šikonja, M. (2015). *ExplainPrediction: Explanation of Predictions for Classification and Regression*. Retrieved from <http://cran.r-project.org/package=ExplainPrediction>

Robnik-Šikonja, M., & Kononenko, I. (2008). Explaining classifications for individual instances. *IEEE Transactions on Knowledge and Data Engineering*, 20(5), 589–600. doi:10.1109/TKDE.2007.190734

Saltelli, A., Chan, K., & Scott, E. M. (2000). *Sensitivity analysis*. Chichester, UK: Wiley.

Štrumbelj, E., Bosnić, Z., Zakotnik, B., Grašić-Kuhar, C., & Kononenko, I. (2010). Explanation and reliability of breast cancer recurrence predictions. *Knowledge and Information Systems*, 24(2), 305–324. doi:10.1007/s10115-009-0244-9

Štrumbelj, E., & Kononenko, I. (2010). An efficient explanation of individual classifications using game theory. *Journal of Machine Learning Research*, 11, 1–18.

Štrumbelj, E., Kononenko, I., & Robnik-Šikonja, M. (2009). Explaining instance classifications with interactions of subsets of feature values. *Data & Knowledge Engineering*, 68(10), 886–904. doi:10.1016/j.datak.2009.01.004

ADDITIONAL READING

Bosnić, Z., Vračar, P., Radović, M. D., Devedžić, G., Filipović, N. D., & Kononenko, I. (2012). Mining data from hemodynamic simulations for generating prediction and explanation models. *IEEE Transactions on Information Technology in Biomedicine*, 16(2), 248–254. doi:10.1109/TITB.2011.2164546 PMID:21846607

de Fortuny, E. J., & Martens, D. (2015). Active learning-based pedagogical rule extraction. *IEEE transactions on neural networks and learning systems*, 26(11), 2664–2677.

Gisbrecht, A., Schulz, A., & Hammer, B. (2015). Discriminative Dimensionality Reduction for the Visualization of Classifiers. In *Pattern Recognition Applications and Methods* (pp. 39–56). Springer International Publishing. doi:10.1007/978-3-319-12610-4_3

Kukar, M., Kononenko, I., & Grošelj, C. (2011). Modern parameterization and explanation techniques in diagnostic decision support system: A case study in diagnostics of coronary artery disease. *Artificial Intelligence in Medicine*, 52(2), 77–90. doi:10.1016/j.artmed.2011.04.009 PMID:21646000

Robnik-Šikonja, M. (2016). Data Generators for Learning Systems Based on RBF Networks. *IEEE transactions on neural networks and learning systems*, 27(5), 926–938.

Robnik-Šikonja, M., Kononenko, I., & Štrumbelj, E. (2012). Quality of classification explanations with PRBF. *Neurocomputing*, 96, 37–46. doi:10.1016/j.neucom.2011.10.038

Schulz, A., Gisbrecht, A., & Hammer, B. (2015). Using discriminative dimensionality reduction to visualize classifiers. *Neural Processing Letters*, 42(1), 27–54. doi:10.1007/s11063-014-9394-1

Starič, A., Demšar, J., & Zupan, B. (2015). Concurrent software architectures for exploratory data analysis. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 5(4), 165–180.

Štrumbelj, E., & Kononenko, I. (2011). A general method for visualizing and explaining black-box regression models. In *International Conference on Adaptive and Natural Computing Algorithms*, pp. 21–30. doi:10.1007/978-3-642-20267-4_3

Van Belle, V., & Van Calster, B. (2015). Visualizing risk prediction models. *PLoS ONE*, 10(7), e0132614. doi:10.1371/journal.pone.0132614 PMID:26176945

KEY TERMS AND DEFINITIONS

Attribute Evaluation: A data mining procedure which estimates the utility of attributes for given task (usually prediction). Attribute evaluation is used in many data mining tasks, for example in feature subset selection, feature weighting, feature ranking, feature construction, decision and regression tree building, data discretization, visualization, and comprehension.

Domain Level Explanation: Tries to find the true causal relationship between the dependent and independent variables. Typically this level is

unreachable except for artificial domains where all the relations as well as the probability distributions are known in advance.

Expressive Power of Explanation: Describes the language of extracted knowledge: propositional logic (that is, if-then rules), nonconventional logic (for example, fuzzy logic), first-order logic, and finite state machines (deterministic, nondeterministic, stochastic).

Feature Contribution: A value assigned to a feature (or its value) that is proportional to the feature's share in the model's prediction for an instance.

Instance Level Explanation: Explanation of prediction of a single instance with a given model, that is, at the model level.

Marginal Effect of a Feature: A difference between model's prediction with the feature and without it, holding all other features intact.

Model Level Explanation: Aims is to make transparent the prediction process of a particular model. Empirical observations show that better models (with higher prediction accuracy) enable better explanation at the domain level.

Portability of Explanation: Describes how well the explanation technique covers the set of available models.

Quality of the Explanation: Can be judged by several criteria: accuracy (generalization ability), fidelity (how well the explanation reflects behavior of the model), consistency (similarity of behavior for different models trained on the same task), and comprehensibility (readability and size of the extracted knowledge).

The Concept of the Shapley Value and the Cost Allocation Between Cooperating Participants

D

Alexander Kolker
GE Healthcare, USA

INTRODUCTION

Game theory is a branch of applied mathematics that studies strategic situations in which participants (players) act in order to maximize their returns (payoffs). As such, game theory provides models of rational behavior (decision-making) for strategic interactions.

Many types of problems that involve decision strategies for cooperating or non-cooperating participants present a fruitful ground for the application of mathematical game theory (Dowd, 2004; Cachon & Netessine, 2004).

In particular, cost allocation problems arise in many situations in which participants work together, such as healthcare providers who have to coordinate patient care in order to reduce the cost and improve the quality of care. It is demonstrated that a natural framework for developing a methodology for cost allocation problems could be based on game theoretical concepts (Tijs and Driessen, 1986; Roth, 1988; Young, 1994; Moulin, 2003). Several concepts for determining the 'fair' cost allocation have been proposed but only a few of these concepts have been used in practice: the nucleolus and the Shapley value.

In this chapter, these two concepts are illustrated side by side. The focus is on examples of practical application of the Shapley value, specifically in healthcare settings. The following two cases are considered in details: (i) the general application of the Shapley value methodology for cost allocation between cooperating providers of care applied to the bundled payment model mandated recently by the Center for Medicare Services' (CMS), and (ii) an

important particular case, in which each participant uses only a portion of the largest participant's asset (the so-called airport game).

BACKGROUND

By pooling resources and cooperating the participants usually reduce the total joint costs and realize savings. The question arises is how the reduced costs or the realized savings should be fairly allocated between them.

There could be different definitions of fair allocation. Some of them are:

- **Equitable Allocation:** Gives everyone the same satisfaction level, i.e. the proportion each player receives by their own valuation is the same for all of them. This is a difficult aim as players might not be truthful if asked their valuation.
- **Proportional Allocation:** Guarantees that each player gets his share. For instance, if three people divide up an asset then each gets at least a third by their own valuation.
- **Envy-Free Allocation:** Everyone prefers his own share to the others. No one is jealous of anyone else. No one would trade his share with anyone else's.
- **An Efficient or Pareto Optimal Allocation:** Ensures that no other allocation would make someone better off without making someone else worse off. The term efficiency comes from the economics idea of the efficient market.

DOI: 10.4018/978-1-5225-2255-3.ch182

- **Merit-Based Allocation:** The more one brings to the coalition, the more one gets out of the division of the accumulated gains.

A concept of fairness is rather subjective. It depends on the participants' socio-economic views and other factors. The fairness schemes described in the next sections form a basis of the two most popular cost allocation approaches: the nucleolus (Tijs and Driessen, 1986; Saad, 2009) and the Shapley value (Roth, 1988; Young, 1994).

MAIN FOCUS

The Nucleolus Concept

The nucleolus can be defined as an equilibrium that finds the 'center of gravity' of the so-called core. The core is defined as a set of inequalities that meet the requirement that no participant or a group of participants pays more than their stand-alone cost. The fairness criteria used by the nucleolus is minimizing the maximum "unhappiness" of a coalition. "Unhappiness" (or "excess") of a coalition is defined as the difference between what the members of the coalition could get by themselves and what they actually get if they accept the allocations suggested by the nucleolus.

More formally, an n -player game is defined by the set $N = \{1, 2, \dots, n\}$ and a function $v(\cdot)$, which for any subset gives a number $v(S)$ called the value of S . The characteristic value of the coalition S , denoted by $v(S)$, is the payoff that all players in the coalition S could jointly obtain. Let x_i be a payoff for player $i = 1, 2, \dots, n$. The nucleolus solution is defined as $x = (x_1, x_2, \dots, x_n)$ such that the excess ("unhappiness") $e_S(x) = v(S) - \sum_{i \in S} x_i$ of any possible coalition S cannot be lowered without increasing any other greater excess. With this definition, the nucleolus is a solution that makes the largest "unhappiness" of the coalitions as small as possible.

There is no general closed-form formula for the nucleolus calculation, except for the recently

developed analytic solution for a particular three-player case (Leng and Parlar, 2010). In general, the nucleolus has to be computed numerically in an iterative manner by solving a series of linear programming (LP) problems, or by solving a very large-scale LP problem. More specifically, the linear programming problem formulation is (Saad et al, 2009):

$$Z \rightarrow \min$$

subject to:

$$Z \sum_{i \in S} x_i \geq v(S) \quad (1)$$

$$\sum_{i \in N} x_i = v(N) \quad (2)$$

The advantage of the nucleolus is that it always exists and that it is unique for all non-empty cores. Therefore, some researchers have used this concept to analyze business and management problems. As an early application of the nucleolus concept, Barton (1992) suggested the nucleolus solution as the mechanism to allocate joint costs among entities who share a common resource. At the same time, due to the complexity of the calculations for large coalitions, the nucleolus has not been extensively used to solve the various allocation-related problems.

Another problem with the nucleolus is that it does not exhibit the monotonicity property (Tijs and Driessen, 1986). Cost allocation concepts that do not exhibit monotonicity could result in having some members paying less if the total cost increases or having paid more if the total cost decreases. An example of a water supply project is available in which cost overrun would actually benefit some participants if the nucleolus method is used for allocating costs: the higher total project cost results in lowering contributions of some participants (Young et al, 1982). Thus, if regulatory agencies do not have the means of monitoring

actual demand and costs, the use of the nucleolus that does not exhibit monotonicity could result in an abusive practice by some participants.

The Shapley Value Concept

The focus of this chapter is the Shapley value because it exhibits some attractive properties (including monotonicity) and it is easier to compute than the nucleolus. Therefore, it is more widely used for practical applications (Young, 1994).

The Shapley value aims at the ‘fair’ allocation of collective costs or gained savings between the collaborating participants based on the relative (marginal) contribution of each participant to their cooperative activities. Basically, the more one brings to the coalition, the more one gets out of the division of the accumulated gains. This is an example of meritocracy fairness.

It should also be noted that an approach was developed in which player’s payoff depends on the worth of the coalitions to which one belongs and not necessarily on one’s marginal contributions, the so-called membership values (Kleinberg and Weiss, 2013). This means that one would get some share regardless of whether or not one makes a marginal contribution to the welfare of the coalition. It was demonstrated that the set of non-marginal membership values includes those that embody widely held notions of fairness, such as partial “benefit equalization”, individual rationality and “greater rewards follow from greater contributions”, where one’s contributions are not measured marginally.

Marginal Contribution

Each participant is supposed to bring some value to the group when he joins it. Participant’s added value or the marginal contribution is defined as surplus that is produced when this participant joins the group.

For example, four participants form a group (coalition). Without one member, the other 3 members produce each \$75 worth of a product.

If this member works alone then he can produce \$80 worth of the product. If he joins the group, then each group member will be able to produce \$85 worth of the product. The one’s added value to the group is defined as $\$85 \times 4 - \$75 \times 3 - \$80 \times 1 = \35 , i.e. $\$85 - \$80 = \$5$ for the newly joined member and \$10 for each of three others. This is also called the ‘marginal contribution’ of the participant. It is a measure of what this group member’s presence is worth, above the minimum that he would require for his service. Notice that if one cannot produce anything working alone (\$0) then the added value becomes greater, e.g. $\$85 \times 4 - \$75 \times 3 = \$115$, i.e. \$85 for the newly joined member and \$10 for each of three others. Thus, joining the group provides more benefit if one cannot produce anything alone.

The Shapley Value Definition

A group of k cooperating members is called coalition, s . All participating members form the grand coalition S that consists of all n participants, $k \leq n$. Each non-empty coalition has a value $V(s)$, which represents the cost (or the value of this coalition). The Shapley value provides a ‘fair’ allocation in the sense that all members are compensated proportionally to their merit, i.e. proportionally to their marginal contributions, $V(s) - V(s-k)$. These contributions are then averaged over all possible different combinations in which the coalition can be formed.

The Shapley value, Sh_k , for each participant k is calculated as

$$Sh_k = \sum_{S \subset n} \frac{(s-1)!(n-s)!}{n!} [V(s) - V(s-k)],$$

where s is the number of participants in coalition S ; summation is performed over all possible coalitions, which participant k joins; $(s-1)!$ is the number of arrangements for participants before joining s ; $(n-s)!$ is the number of arrangements for participants after joining s , and $n!$ is the total number of all possible coalition combinations.

Thus, the Shapley value is computed by calculating the average marginal contribution that participant k brings to a coalition (group) s if this participant joins any coalition, and all coalitions for this participant k are formed in random order (there is no way of taking advantage of a more favorable coalition order).

Participants have incentives to stay voluntarily in the coalition if three conditions are satisfied: (i) the participant's coalition cost is lower than its stand-alone cost (or the saving is higher). This condition is called individual rationality; (ii) any sub-groups' costs are lower than that of combined stand-alone costs of the sub-groups' participants (or the saving is higher). This condition is called sub-group rationality; and (iii) the total costs (savings) must be completely distributed among all cooperating participants, i.e. no participant's share of the costs can be reduced without increasing another participant's share. This condition is called total distribution rationality. If these conditions are satisfied, then the so-called core of cost sharing is not empty. Sometimes, though, the core can be empty. This means that there is no unique cost allocation that satisfies all participants' conditions. Hence, there is no incentive to voluntarily form the coalition and stay in it. The Shapley value can always be calculated even if the core is empty. If the core is not empty, then all participants have an incentive to voluntarily stay in the coalition and get their 'fair' cost (savings) allocation.

SOLUTIONS AND RECOMMENDATIONS

Example 1: Three-Member Group

For a long time, policymakers have been increasingly frustrated with fee-for-service payment system in healthcare settings. Fee-for-service rewards volumes and encourages silos and fragmentation of care. Several provisions of the Affordable Care Act (ACA) healthcare legislation in the United States

seek to shift provider payments to value-based approaches that encourage quality improvement and cost reduction. Payment bundling promoted by the Center of Medicare Services' (CMS) is one such approach. In this payment model a single (bundled) payment for the episode of care is paid to all providers of services through a contracting organization. The contracting organization is responsible for allocating the payments among all providers (hospitals, physicians, skilled nursing facilities, rehabilitation units, etc.). CMS offers no methodology for sharing the savings or the extra costs between providers. The providers themselves are supposed to use a methodology of their choice for 'fair' distributing the payments among them. The Shapley value methodology illustrated in this and in the following examples could serve as a good choice.

The benefits of increased alignment (cooperation) between healthcare providers have been well established. The bundled payment offers a number of opportunities for hospitals to accrue savings through reductions in actual hospital costs (operating expenses), decreases in length of stay (LOS), reduction or avoidance of readmissions, and management post-acute care (Pearce and Harris, 2010).

Suppose for simplicity that the contracting organization includes three main providers: hospital (H), physician group (PG), and skilled nursing facility (SNF).

Let's consider, as an example, the representative costs of care for some medical condition related to an acute care episode for individual providers presented, for example, by Pearce and Harris (2010, exhibit 1) and indicated on Figure 1, line 1. Let's further assume that the 2-member and the 3-member coalitions have achieved the percent (%) cost reduction vs. stand-alone members' costs presented in lines 3 and 6 on Figure 1, respectively.

The Shapley value costs for participants 1, 2 and 3 (the group size $n=3$) are calculated as:

Figure 1. Stand-alone costs and the Shapley costs allocation for group size $n=3$. All cores are non-empty.

#	Participant (member)->	Hospital {1}	Physician group {2}	SNF {3}	Total
1	Median Medicare cost for stand-alone providers per episode-of-care	\$30,000	\$6,000	\$4,000	\$40,000
2	2-members coalitions->	{12}	{13}	{23}	
3	Cost reduction for 2-members coalition vs. stand-alone member cost, %	3.0%	2.0%	1.5%	
4	Cost for 2-member coalitions	\$34,920	\$33,320	\$9,850	
5	3-member coalition->	{123}			
6	Cost reduction for 3-member coalition vs. stand-alone member cost, %	3.5%			
7	Cost for 3-member coalitions	\$38,600			
8	Shapley value for Hospital: provider {1}	Shapley value for Physician group: provider {2}	Shapley value for SNF: provider {3}	Total	
9	\$29,290	\$5,555	\$3,755	\$38,600	
10	Saving {1}	Saving {2}	Saving {3}		
11	\$710	\$445	\$245	\$1,400	
12	Saving allocation % {1}	Saving allocation % {2}	Saving allocation % {3}		
13	2.4%	7.4%	6.1%		
14	Coalition->	{12}	{13}	{23}	
15	Condition 1 for non-empty core: Shapley costs is less than stand-alone members' costs	OK	OK	OK	
16	Coalition->	{12}	{13}	{23}	
17	Condition 2 for non-empty core: two-member coalition costs are less than the combined members' costs	OK	OK	OK	

$$Sh_1 = w_1 * C_1 + w_2 * (C_{12} - C_2) + w_3 * (C_{13} - C_3) + w_4 * (C_{123} - C_{23})$$

$$Sh_2 = w_1 * C_2 + w_2 * (C_{12} - C_1) + w_3 * (C_{23} - C_3) + w_4 * (C_{123} - C_{13})$$

$$Sh_3 = w_1 * C_3 + w_2 * (C_{13} - C_1) + w_3 * (C_{23} - C_2) + w_4 * (C_{123} - C_{12})$$

where C_1 , C_2 , and C_3 are the costs for stand-alone providers, respectively; C_{12} , C_{13} , and C_{23} are

the costs for 2-member coalitions, respectively; and C_{123} is the cost for the 3-member coalition.

For $s=1$ the weight $w_1 = (1-1)!(3-1)!/3! = 1/3$; for $s=2$ the weight $w_2 = (2-1)!(3-2)!/3! = 1/6$; and for $s=3$ the weight $w_3 = (3-1)!*(3-3)!/3! = 1/3$.

These formulas were built in the Excel spreadsheet. Results are presented on Figure 1.

Lines 15 and 17 include checking conditions for voluntary staying in the coalition, i.e. the conditions that participant's coalition costs are lower than its stand-alone costs (or the savings is

The Concept of the Shapley Value and the Cost Allocation Between Cooperating Participants

higher)-individual rationality condition, and (ii) any sub-groups' costs are lower than that for the corresponding combined stand-alone costs of the sub-groups' participants (or the saving is higher)-sub-group rationality condition. 'OK' means that the corresponding condition is satisfied and that

the core is not empty. Thus, the participants have incentives to stay voluntarily in the coalition.

However, if the projected combined cost reduction (%) for three participants is a little bit lower, say, 3% rather than 3.5%, then the sub-group rationality condition for coalition {12} is violated. This is indicated on Figure 2, line 17, as 'Core is

Figure 2. Stand-alone costs and the Shapley costs allocation for group size $n=3$. Some cores are empty.

#	Participant (member)->	Hospital {1}	Physician group {2}	SNF {3}	Total
1	Median Medicare cost for stand-alone providers per episode-of-care	\$30,000	\$6,000	\$4,000	\$40,000
2	2-members coalitions->	{12}	{13}	{23}	
3	Cost reduction for 2-members coalition vs. stand-alone member cost, %	3.0%	2.0%	1.5%	
4	Cost for 2-member coalitions	\$34,920	\$33,320	\$9,850	
5	3-member coalition->	{123}			
6	Cost reduction for 3-member coalition vs. stand-alone member cost, %	3.0%			
7	Cost for 3-member coalitions	\$38,800			
8	Shapley value for Hospital: provider {1}	Shapley value for Physician group: provider {2}	Shapley value for SNF: provider {3}	Total	
9	\$29,357	\$5,622	\$3,822	\$38,800	
10	Saving {1}	Saving {2}	Saving {3}		
11	\$643	\$378	\$178	\$1,200	
12	Saving allocation % {1}	Saving allocation % {2}	Saving allocation % {3}		
13	2.1%	6.3%	4.5%		
14	Coalition->	{12}	{13}	{23}	
15	Condition 1 for non-empty core: Shapley costs is less than stand-alone members' costs	OK	OK	OK	
16	Coalition->	{12}	{13}	{23}	
17	Condition 2 for non-empty core: two-member coalition costs are less than the combined members' costs	Core is Empty	OK	OK	

Empty'. Indeed, in this case, the Shapley value for provider 1 is \$29,357, and for provider 2 is \$5,622, respectively (compare to the values on Figure 1). Thus, their sum (\$34,979) exceeds the projected combined cost for stand-alone providers 1 and 2 (\$34,920). This makes the core empty. There is no financial incentive for providers 1 and 2 to stay voluntarily in the coalition. The coalition is likely to break down.

Using the developed spreadsheet, it is easy to play various input data scenarios to analyze which coalition is likely to voluntarily stay, and what would be the corresponding costs and savings allocations for the coalitions' members (providers).

Example 2: Four-Member Group

CMS comprehensive care for joint replacement (CJR) model went into effect on April 1, 2016 (Federal Register, 2015). Under this model, acute care hospitals in certain selected 67 metropolitan areas will receive retrospective bundled payments for episodes of care for lower extremity joint replacement. The providers are supposed to share the savings or bear the extra costs vs. the CMS baseline bundled cost.

According to Becker's Hospital Spine Review (Dyrda, 2016), the average Medicare cost breakdown per CJR episode includes 8 categories: inpatient hospital stay \$13,193; skilled nursing facility \$5,034; inpatient rehabilitation facility \$1,568; home health agency \$2,123; physicians \$1,675; hospital readmissions \$1,155; outpatient services \$604; and durable medical equipment \$122. To illustrate the Shapley cost allocation method let's combine for simplicity some of these costs to produce a 4-member participants group: inpatient hospital stay, physicians, hospital readmission (H), \$16,023; inpatient rehabilitation (IPR), \$1,568; skilled nursing facility (SNF), \$5,034; and home health agency, durable equipment, outpatient services (HHA), \$2,849. The total average baseline cost for the CJR episode of care is \$25,474.

As indicated on Figure 3, four providers can form 6 paired coalitions (line 2), 4 triple coalitions (line 5), and 1 grand coalition (line 8). Suppose that the 2-members' coalitions, the 3-members' coalitions, and the grand coalition have achieved the percent (%) cost reduction vs. stand-alone members' costs presented in lines 3, 6 and 9 on Figure 3, respectively.

The Shapley value for each provider contains 8 marginal values terms: 1 single, 3 double, 3 triple and 1 with grand coalition. These terms are summed up with coefficients $(1-1)!(4-1)!=6$, $(2-1)!(4-2)!=2$, $(3-1)!(4-3)!=2$, and $(4-1)!(4-4)!=6$, respectively, and the sum is then divided by $n!=4!=24$. These formulas were built in an Excel spreadsheet.

The Shapley values for all four providers, their savings and the percent of savings allocations are indicated on Figure 3, lines 11-16. The cores for all coalitions are not empty. All four providers have financial incentive to stay in coalitions because their stand-alone costs would be higher. The Shapley values allocate higher percentage savings to providers 2 (IPR) and 4 (HHA) (line 16) than simple cost-proportional allocations. This will likely encourage their cooperation with the hospital/physicians and SNF in providing coordinated patient care rather than position themselves as stand-alone providers.

At the same time, if the overall cost reduction for the contracting organization is only a little bit lower, say 4%, rather than 5%, then coalitions {13} and {134} will break-down, as indicated on Figure 4, lines 18 and 20. (To save space, only lines 8-20 of Figure 4 are included). Their corresponding cores are empty because the sum of the Shapley values for them will exceed the combined costs of the corresponding 2- and 3-member coalitions. Generally, the lower the percentage of cost reduction the higher the chances of breaking-down the voluntary coalition of providers. As in the previous example, it is easy to play various input data scenarios to analyze which coalition is likely to voluntarily stay, and what would be the corresponding costs and savings allocations for the coalitions' members (providers).

Figure 3. Stand-alone costs and the Shapley costs allocation for group size $n=4$. All cores are not empty.

#	Participant (member)->	Hospital (1)	IPR (2)	SNF (3)	HHA (4)		Total
1	Median cost for stand-alone providers per episode-of-care	\$16,023	\$1,568	\$5,034	\$2,849		\$25,474
2	2-members coalitions->	{12}	{13}	{14}	{23}	{24}	{34}
3	Cost reduction for 2-members coalition vs. stand-alone member cost	3%	4%	1%	2%	3%	2%
4	Cost for 2-member coalitions	\$17,063	\$20,215	\$18,683	\$6,503	\$4,284	\$7,725
5	3-member coalitions->	{123}	{124}	{134}	{234}		
6	Cost reduction for 3-member coalition vs. stand-alone member cost	3%	3%	4%	2%		
7	Cost for 3-member coalitions	\$22,059	\$19,827	\$22,950	\$9,262		
8	Grand coalition->	{1234}					
9	Cost reduction for grand coalition	5.0%					
10	Cost for grand coalition (over non-cooperating stand-alone providers)	\$24,200					
11		Shapley value for provider 1, H{1}	Shapley value for provider 2, IPR{2}	Shapley value for provider 3, SNF{3}	Shapley value for provider 4, HHA{4}		
12		\$15,476	\$1,410	\$4,705	\$2,608		
13		Saving {1}	Saving {2}	Saving {3}	Saving {4}	Total savings	
14		\$547	\$158	\$329	\$241	\$1,274	
15		Saving allocation % {1}	Saving allocation % {2}	Saving allocation % {3}	Saving allocation % {4}		
16		3.4%	10.0%	6.5%	8.5%		
17	Coalition ->	{12}	{23}	{34}	{13}	{14}	{24}
18	Condition 2 for non-empty core: 2-member coalition costs are less than the combined members' costs	OK	OK	OK	OK	OK	OK
19	Coalition ->	{123}	{124}	{134}	{234}		
20	Condition 3 for non-empty core: 3-member coalition costs are less than combined members' costs	OK	OK	OK	OK		

Simplified Shapley Value for Special Cases

The Shapley value formula is difficult to use when the coalition size increases above 4. The reason is that the number of permutations increases exponentially with the number of participants, and the costs data are needed for all k -member coalitions

($k=2, 3, 4, \dots, n$). These multiple cost data are difficult to get, or they are not known. Therefore, some special cases are important in which simplified or approximate expressions of the Shapley value are available (Reinhardt and Dada, 2005). In particular, there are widespread situations in which participants need to use only a portion of the available total capacity. The total capacity is

Figure 4. The Shapley costs allocation for group size $n=4$. Some cores are empty.

8	Grand coalition->	{1234}					
9	Cost reduction for grand coalition	4.0%					
10	Cost for grand coalition (over non-cooperating stand-alone providers)	\$24,455					
11		Shapley value for provider 1, H{1}	Shapley value for provider 2, IPR{2}	Shapley value for provider 3, SNE{3}	Shapley value for provider 4, HHA{4}		
12		\$15,540	\$1,474	\$4,769	\$2,672		
13		Saving {1}	Saving {2}	Saving {3}	Saving {4}	Total savings	
14		\$483	\$94	\$265	\$177	\$1,019	
15		Saving allocation % {1}	Saving allocation % {2}	Saving allocation % {3}	Saving allocation % {4}		
16		3.0%	6.0%	5.3%	6.2%		
17	Coalition ->	{12}	{23}	{34}	{13}	{14}	{24}
18	Condition 2 for non-empty core: 2-member coalition costs are less than the combined members' costs	OK	OK	OK	Core is Empty	OK	OK
19	Coalition ->	{123}	{124}	{134}	{234}		
20	Condition 3 for non-empty core: 3-member coalition costs are less than combined members' costs	OK	OK	Core is Empty	OK		

required to serve only the biggest participant. The users of the portions of the total capacity want to pay only their corresponding share of the biggest user's cost. One can think of toll roads, computer networks, or shared healthcare facilities.

One such practically important case is historically called an "airport game" developed originally for setting landing fees for different types of aircrafts (small and big) that share the same runway. The runway length (and the cost) is defined essentially by the biggest type aircraft that requires the full runway length while smaller type aircrafts can use only a portion of the total runway length. Littlechild and Owen (1973) first demonstrated that an empirical cost allocation rule for landing fees is the Shapley value in this particular case.

Let the cost of all stand-alone players be ordered from low to high, i.e.

$$C_1 < C_2 < C_3 < \dots < C_m,$$

where m is the number of types of players.

The number of players within each cost type are

$$n_1, n_2, n_3, \dots, n_m.$$

The total number of all players is $N = n_1 + n_2 + n_3 + \dots + n_m$.

Then, the Shapley value

for player 1 is $Sh_1 = C_1/N$

for player 2 is $Sh_2 = Sh_1 + (C_2 - C_1)/(N - n_1)$

for player 3 is $Sh_3 = Sh_2 + (C_3 - C_2)/(N - n_1 - n_2)$

for player 4 is $Sh_4 = Sh_3 + (C_4 - C_3)/(N - n_1 - n_2 - n_3)$

for player m is $Sh_m = Sh_{m-1} + (C_m - C_{m-1})/(N - \sum_{i=1}^{m-1} n_i)$

Table 1. Annual start-up and operating costs for new CDI

	Hospital 1	Hospital 2	Hospital 3	Hospital 4
Annual patient volumes	1000	950	800	1800
Estimated Annual Start-Up Costs	\$5M	\$4M	\$2M	\$9M

Thus, the pattern is that the Shapley value for each player is equal to that for the previous one plus the corresponding price differential divided by the total number of players but the sum of all previous smaller ones.

If only one player is present in each cost category, i.e. if $n_1=n_2=n_3=\dots=1$ then:

$$Sh_1 = C_1/N$$

$$Sh_2 = Sh_1 + (C_2 - C_1)/(N-1)$$

$$Sh_3 = Sh_2 + (C_3 - C_2)/(N-2)$$

$$Sh_4 = Sh_3 + (C_4 - C_3)/(N-3)$$

$$Sh_m = Sh_{m-1} + (C_m - C_{m-1})/(N-m+1)$$

These formulas can easily be built in an Excel spreadsheet to automate calculations.

Example 3

Four hospitals consider offering imaging service at a new Center for Diagnostic Imaging (CDI). Annual start-up and operating costs, as well as the projected number of patients that require service, are in Table 1.

The hospitals have agreed to use a portion of one largest \$9M facility that will serve their annual patient volumes.

How much will each hospital have to contribute to the largest facility?

What will be the minimal ‘fair’ charge per patient to recoup the costs over one year?

Step 1: Sort the costs in ascending order, i.e.

Hospital 3: $C_3 = \$2M$; 800 patients

Hospital 2: $C_2 = \$4M$; 950 patients

Hospital 1: $C_1 = \$5M$; 1000 patients

Hospital 4: $C_4 = \$9M$; 1800 patients

Step 2: Calculate the sum of all patient volumes:
 $N = 800 + 950 + 1000 + 1800 = 4550$.

Step 3: The Shapley value (minimal charge per patient):

Hospital 3:

$$Sh_3 = C_3/N = \$2M/4550 = \$439.56;$$

Total: $\$439.56 \times 800 = \$0.352M$

Hospital 2: $Sh_2 = Sh_3 + (C_2 - C_3)/(N - n_3) = \$439.56 + (\$4M - \$2M)/(4550 - 800) = \$972.89;$

Total: $\$972.89 \times 950 = \$0.924M$

Hospital 1: $Sh_1 = Sh_2 + (C_1 - C_2)/(N - n_3 - n_2) = \$972.89 + (\$5M - \$4M)/(4550 - 800 - 950) = \$1,330.04;$

Total: $\$1,330.04 \times 1000 = \$1.33M$

Hospital 4: $Sh_4 = Sh_1 + (C_4 - C_1)/(N - n_1 - n_2 - n_3) = \$1,330.04 + (\$9M - \$5M)/(4550 - 800 - 950 - 1000) = \$3,552.26;$

Total: $\$3,552.26 \times 1800 = \$6.394M$

Total cost is $\$0.352M + \$0.924M + \$1.33M + \$6.394M = \$9M$, i.e. the cost of the largest facility.

Example 4

Four dental groups consider opening dental clinics. The number of dentists and the total start-up costs are given in Table 2 (SNDP Manual, 2011).

If the dental groups use the corresponding portions of one largest 12-chairs clinic what will

Table 2. Dental clinic comparison chart: Capacity and costs for fixed clinics, and clinics using portable equipment

Size	3 Chairs	6 Chairs	9 Chairs	12 Chairs
Number of dentists	1	2	3	4
Total estimated Start-Up Costs	\$624,452	\$1,016,217	\$1,268,330	\$1,549,538

be the cost allocation for each dentist and the dental groups?

Total number of all dentists is $N=10$.

$$Sh_1 = \$624,452/10 = \$62,445 \text{ for dentist } 1.$$

$$Sh_2 = \$62,445 + (\$1,016,217 - \$624,452)/(10-1) = \$105,974 \text{ for dentist } 2.$$

$$\text{Total for dental group 2: } \$105,974 * 2 = \$211,948.$$

$$Sh_3 = \$105,974 + (\$1,268,330 - \$1,016,217)/(10-1-2) = \$141,990 \text{ for dentist } 3.$$

$$\text{Total for dental group 3: } \$141,990 * 3 = \$425,970$$

$$Sh_4 = \$141,990 + (\$1,549,538 - \$1,268,330)/(10-1-2-3) = \$212,292 \text{ for dentist } 4.$$

$$\text{Total for dental group 4: } \$212,292 * 4 = \$849,168.$$

Total cost for all dental groups is \$1,549,538, i.e. the cost of the largest facility.

Let's now consider only four dentists' groups without taking into account the number of dentists within each group. This means that we have only 1 member in each cost category; thus, $N=4$. In this case the Shapley values for each dental group are:

$$Sh_1 = C_1/N = \$624,452/4 = \$156,113$$

$$Sh_2 = Sh_1 + (C_2 - C_1)/(N-1) = \$156,113 + (\$1,016,217 - \$624,452)/(4-1) = \$286,701$$

$$Sh_3 = Sh_2 + (C_3 - C_2)/(N-2) = \$286,701 + (\$1,268,330 - \$1,016,217)/(4-2) = \$412,757$$

$$Sh_4 = Sh_3 + (C_4 - C_3)/(N-3) = \$412,757 + (\$1,549,538 - \$1,268,330)/(4-3) = \$693,965$$

Total for 4 dental groups is \$1,549,538, i.e. the cost of the largest facility.

FUTURE RESEARCH DIRECTIONS

As it was previously mentioned, the Shapley value formula is difficult to use when the group size n is greater than 4 because the number of double, triple, etc. permutations, P , increases exponentially with the group size as, $P = 2^n - (n+1)$. For example, for a modest group size $n=10$, more than 1000 sub-groups and their marginal costs should be considered. This is impractical. Although some special approximations were developed, such as an airport game, more research is needed for further developing other Shapley value approximations for large group sizes. Specifically, an approximation should be developed that does not necessarily require evaluating every possible coalition using, for example, the constraint generation technique.

Another research direction is further developing the concept of coalitional value, i.e. the possibility that some players may be more likely to act together than others. Despite the fact that coalition structures arise naturally in cost allocation problems, their applications have not been explored sufficiently (Costa and Garcia_Jurado, 2013). Taking into account the excellent properties of the Shapley value, it is natural to study how the Shapley values perform for games with a coalitional structure.

CONCLUSION

Cost allocation problem often arises in many business situations that benefit from the effect of cooperating partners. Examples are bundled payment to healthcare providers; pooling resources where savings from cooperation must be distributed ‘fairly’ between the participants, such as distribution of surgical costs in the presence of patient queues (Gonzalez and Herrero, 2004); joint hospitals’ imaging or surgical centers, etc.

A natural framework to study cost allocation problems is based on game theory approach. About a dozens of alternate concepts have been proposed to determine the ‘fair’ costs or savings allocation but only a few of them have been used in practice: the nucleolus and the Shapley value. It should be realized that there is no one best ‘fair’ cost allocation method. The choice of a method depends on the participants’ ideas about ‘fairness’, their economic and political views, and other subjective factors. Also, the ease of calculation and interpretation plays a significant role for the practical use.

The Shapley value’s axiomatic basis has some shortcomings (Tijs & Driessen, 1986; Young, 1994). Nonetheless, the attractive features of this concept make it valuable for various applications, especially in healthcare settings where providers’ cooperation and team based care is currently a long-term strategy for improving quality of care and cost reduction.

REFERENCES

- Barton, T.L. (1992). A unique solution for the nucleolus in accounting allocations. *Decision Sciences*, 23(2), 365–375. doi:10.1111/j.1540-5915.1992.tb00394.x
- Cachon, G., & Netessine, S. (2004). Game theory in supply chain analysis. In Z. J. Shen (Ed.), *Handbook of quantitative supply chain analysis* (pp. 13–65). Kluwer Academic Publisher. doi:10.1007/978-1-4020-7953-5_2
- Costa, J., & Garcia-Jurado, I. (2013). Coalitional values and cost allocation problems. *International Game Theory Review*, 15(2), 1–8. doi:10.1142/S0219198913400021
- Dowd, S. (2004). Applied game theory for the hospital manager. Three case studies. *The Health Care Manager*, 23(2), 156–161. doi:10.1097/00126450-200404000-00009 PMID:15192996
- Dyrda, L. (2016). CJR bundles to pay \$25k per episode: 8 statistics on cost breakdown. *Becker’s Spine Review*. Retrieved from <http://www.beckersspine.com/orthopedic-spine-practices-improving-profits/item/30556-cjr-bundles-to-pay-25k-per-episode-8-statistics-on-cost-breakdown.html>
- Federal Register. (2015). *Medicare Program; Comprehensive Care for Joint Replacement Payment Model for Acute Care Hospitals Furnishing Lower Extremity Joint Replacement Services*. Retrieved from <https://federalregister.gov/a/2015-29438>
- Gonzalez, P., & Herrero, C. (2004). Optimal sharing of surgical costs in the presence of queues. *Mathematical Methods of Operations Research*, 59, 435–446. doi:10.1007/s001860400350
- Klenberg, N., & Weiss, J. (2013). On membership and marginal values. *International Journal of Game Theory*, 42(2), 357–373. doi:10.1007/s00182-013-0367-9
- Leng, M., & Parlar, M. (2010). Analytic solution for the nucleolus of a three-player cooperative game. *Naval Research Logistics*, 57(7), 667–672. doi:10.1002/nav.20429

Littlechild, S., & Owen, G. (1973). A simple expression for the Shapley value in a special case. *Management Science*, 20(3), 370–372. doi:10.1287/mnsc.20.3.370

Moulin, H. (2003). *Fair division and collective welfare*. Cambridge, MA: The MIT Press.

Pearce, J., & Harris, J. (2010). The Medicare bundled payment pilot program: Participation considerations. *Healthcare Financial Management (HFM) Magazine*. Retrieved from <http://www.hfma.org/Templates/Print.aspx?id=22682>

Reinhardt, G., & Dada, M. (2005). Allocating the gains from resource pooling with the Shapley value. *The Journal of the Operational Research Society*, 56(8), 997–1000. doi:10.1057/palgrave.jors.2601929

Roth, A. (1988). *The Shapley value. Essays in honor of Lloyd Shapley*. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511528446

Saad, W., Han, Z., Debbah, M., Hjørungnes, A., & Basar, T. (2009). Coalitional game theory for communication networks: A tutorial. *IEEE Signal Processing Magazine*.

SNDC Manual. (2011). Facility design and staffing. Retrieved from <http://www.dentalclinicmanual.com/>

Tijs, S. H., & Driessen, T. S. (1986). Game theory and cost allocation. *Management Science*, 32(8), 1015–1028. doi:10.1287/mnsc.32.8.1015

Young, H. P. (1994). Cost allocations. In R. Aumann & S. Hart (Eds.), *Handbook of game theory with economic application* (Vol. 2, pp. 1193–1230). Elsevier Science B.V. doi:10.1016/S1574-0005(05)80066-9

Young, H. P., Okada, N., & Hashimoto, T. (1982). Cost allocation in water resources development. *Water Resources Research*, 18(3), 463–475. doi:10.1029/WR018i003p00463

KEY TERMS AND DEFINITIONS

Bundled Payment: A payment model to healthcare providers in which one single payment is disbursed to cover an episode of patient care through a contracting organization. The contracting organization is responsible for allocating the payments among all providers.

Coalition: A group of k cooperating partners.

Cost Allocation: A problem that arises in many business situations that benefit from the effect of economy of scale or cooperating partners.

Core: A set of inequalities that meet the requirement that no participant or a group of participants pays more than their stand-alone cost.

Empty Core: A lack of unique cost allocation that satisfies all participants. If the core is empty, then unsatisfied participants have incentive to leave the cost sharing coalition.

Game Theory: A branch of applied mathematics that studies strategic situations in which participants (players) act rationally in order to maximize their returns (payoffs).

Marginal Contribution: A value of the group with the player as a member minus the value of the group without the player minus the value created by the player working alone.

Nucleolus: A game theory concept defined as minimizing the maximum “unhappiness” of a coalition. “Unhappiness” (or “excess”) of a coalition is defined as the difference between what the members of the coalition could get by themselves and what they are actually getting if they accept the allocations suggested by the nucleolus.

Shapley Value: A game theory concept aimed at the ‘fair’ allocation of the collective costs or profits (savings) between several collaborative participants. It is based on allocating the costs to the cooperating participants proportionally to the marginal contributions of each participant that is averaged over all possible combinations in which participants can cooperate.

Decision Filed Theory

Lan Shao

University of Oulu, Finland

Jouni Markkula

University of Oulu, Finland

INTRODUCTION

Development advanced ICT based systems, applications and services for improving our daily life require understanding and modeling of how people make decisions. The human decision making has been of interest, and theories been developed, in different fields of science, including economics, psychology, sociology and computer science. There exist a number of well-known and widely applied decision making theories, such as causal decision theory, evidential decision theory and game theory. In ICT field, formal decision making models are used, for example, in developing decision support systems, artificial intelligence solutions and intelligent services. Growing number of application of these systems can be found in such areas as electronic commerce, intelligent traffic and health.

Decision Filed Theory (DFT) is one of the younger human decision making theories that has significant potential for practical applications in real-world decision making situations. It employs a dynamic decision making model, compared to static models that most of the traditional decision theories are relying on, and builds on psychology, whereas the earlier models are commonly based on economics and rational utility theories. These characteristics make it potentially useful for present day application needs. Along with development of technology, the DFT has been studied for solving various kinds of decision making problems in different contexts. Several

successful empirical studies have shown that DFT theory is able to explain human's dynamic behavior in many areas and it can be applied as a basis for ICT system design.

By a glance to the existing research, we can see that the DFT has been adapted to different application contexts. However, for having stronger evidence for its validity to practical empirical cases and understanding its trends and current state, the existing scientific literature needs to be reviewed and analyzed more systematically and in a structured way. This article presents the results of our study, where we analyze and synthesized the existing research knowledge on DFT with the means of Systematic Literature Review (SLR). The purpose is to give an evidence based view to the development of Decision Field Theory and its applications, which offers an understanding of its validity for practical applications and future potential.

BACKGROUND

Decision Field Theory, initially proposed by Busemeyer and Townsend (1993), is a dynamic cognitive approach to describe and explain human decision making under risk and uncertainty. It provides a formal theory and a mathematical model of human decision behavior, which takes into account two of the basic phenomena of human decision making that are not covered by most of the other theories: variability of preference of choices

and systematic relation between preferences and deliberation time (Busemeyer & Townsend, 1993). Before the theory has been applied to solve multiple options problems of decision making, it has experienced several stages that are the foundation for present theory. The early deterministic subject expected utility (SEU) theory is restricted to the choices between two options (Savage, 1954). This theory can be applied to predict the trend of preferences on options, but it can't explain the dynamic change of preference. It was followed by the Sequential SEU Theory (DeGroot, 1970), which take into account the sequential order of consecutive decisions. The main problem of the Sequential SEU theory is that the knowledge and experience in the past has no influence on current decision making, which goes against with real human deliberative process. The Random Walk SEU Theory (Busemeyer, 1985) solves this problem by setting an initial preference state. The deterministic decision making theories originate from the basic proposal that the action is either true or false between two conflict actions, which is referred to binary preference relation (Fishburn, 1988). In subsequent, it was used for two or more attributes decision making under dynamic environment (Diederich, 1997). After that, DFT was gradually mature and widely used.

DFT is a method that provides insights into relationship between preference evolution and dynamic decision making. The following original and classic mathematical equation is used to explain the dynamic evolution of preferences during the deliberation time.

$$P(t+h) = SP(t) + V(t+h)$$

In this equation, P is the Preference vector, S is the Growth-decay matrix and V the Valence vector, which are explained below. The equation can be decomposed further as

$$p_i(t+h) = s_{ii} \cdot p_i(t) + \sum_{k \neq i} s_{ik} \cdot p_k(t) + v_i(t+h)$$

In general, DFT consider time as a dynamic factor which lead to frequent change in human's preference among various options. So for the time series $\{t_0, t_1, t_2, \dots, t_j, \dots, t_n\}$, the preferences on options fluctuate upon specific time point and forms preferences series $\{p_0, p_1, p_2, \dots, p_j, \dots, p_n\}$. In the equation, the time points t and $t+h$ are two adjacent time moments, and parameter h , referred as time unit, denotes the distance between two moments of decision making. From microscopic side, decision making is a continuous process when the moment is sequential. So, the length of h approaches to zero in the limit (Abad, 2014). While from the macroscopic side, decision making is intervallic. Thus, time unit h is equal to an arbitrary value.

Preference Vector P

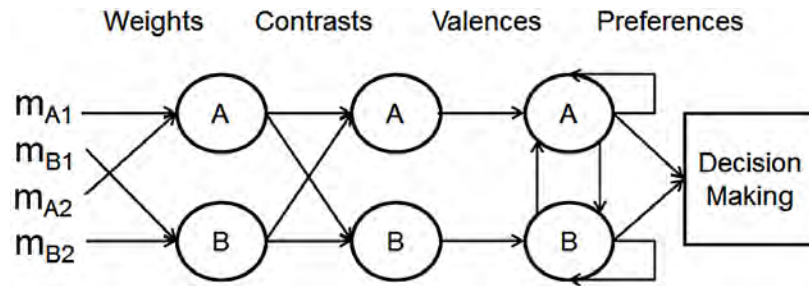
The row vector $P(t)' = [P_1(t), P_2(t), \dots, P_n(t)]$ represents the preference state at time t for the n options.

Growth-Decay Matrix S

The decision makers' perception of decision case is not only determined by current information, but it also would be affect by their previous knowledge or similar experience (Van Horne, 1967). Consequently, the preferences value in current moment is influenced positively or negatively by the decision makers' memory. The matrix S , referred as feedback vector, is formed as $S = (I - h\Gamma)$ that is assumed to be symmetric ($r_{ij} = r_{ji}$ for all i, j). The diagonal elements of s_{ii} provide memory for previous states of the system. The off-diagonal values of s_{ik} allow for competitive interactions among competing alternatives. The researchers (Busemeyer et al., 2008; Qin et al., 2013) explain the implication of s_{ii} value in four following conditions.

- $0 < s_{ii} < 1$ indicates that memory of decision maker has positive feedback on current decision problem.

Figure 1. Mechanism of DFT elements
(adapted from Qin et al. (2013))



- $s_{ii} > 1$ suggests that memory of decision maker has negative feedback on current decision problem.
- $s_{ii} = 0$ means that memory of decision maker has no influence on current decision problem.
- $s_{ii} = 1$ expresses that memory of decision maker has perfect impact, not decay or growth, on current problem.

Valence Vector V

The row vector $V(t)' = [v_1(t), v_2(t), \dots, v_n(t)]$ represents the valence input at time t for the n options. The valence vector $V(t)$ is consists of three matrices.

$$V(t) = CMW(t)$$

At time t , the anticipated value of an option on an attribute is compared with the anticipated values of other options on the same attribute. These comparisons produce a valence for each option, denoted $v_i(t)$, for the valence of option i at time t . The matrix C is called contrast matrix, the elements of which are valued as $c_{ii} = 1$ and $c_{ij} = -1/(n-1)$ for $i \neq j$ where n is the number of option. Matrix C is defined to compare all the options between each other and the structure of it guarantees the sum of $V(t)$ components always equals to 0. M is called attribute matrix that is an $n \times m$ matrix (n is the number of options and m is the number of attribute). The attribute matrix M

represents decision makers' evaluations for each option on each attribute under certain conditions. $W(t)$, called weight matrix, is a $m \times 1$ matrix (m is the number of attribute) that indicates the weights of attention people allocates to each attribute of M . Simply put, the number of each weight value indicates how important the attribute is in decision. The decision of people is changing moment by moment, which is caused by the variability of $W(t)$.

The Figure 1 illustrates the mechanism how each of the element of the model act in DFT. From this model you can see the deliberation process of people and the reasons causing dynamic decision making.

The essence of dynamic decision making is the changes on preferences value on options. The evolution process of preference state is the result of dynamic valence that is consists of three elements: contrast matrix C , attribute matrix M and weight matrix W . In original DFT model, the value of contrast matrix and attribute matrix is constant while the value of weight matrix is changeable.

DFT RESEARCH TRENDS AND APPLICATIONS

DFT was introduced in 1993 as a response to the limitations of static deterministic models of human decision making. Since then it has been utilized in many application domains and also validated with many different cases. In the field of supply chain management, a method of dynamic decision

making is structured in order to realize strategic vendor selection or switching based on the principle of hierarchical planning (Sucky, 2007). The results showed that the proposed method is able to explain behavior of vendor selection, and then decision maker could choose the optimal solution of vendor management among alternative offers. In the stock market, DFT is used to analyze and avoid risks based on the experiences (Cohen et al., 2008). In this study, the authors set up a model representing preferences under uncertain environment where the current risks are dependent on experience. Vescoukis et al. (2011) have proposed an architecture based on dynamic decision making in disaster management for plan constructing, retrieving, and exchanging. This architecture not only predicts human decision making in disaster management, but also provides recommendation and evaluation for solution of disaster management. Moreover, DFT is also adopted in many other contexts, such as transportation, agriculture and military (Cheung et al., 1999; Pearson et al., 2010; Du et al., 2010).

In general, the research, development and the extent of the DFT applications in time and with respect of various application domains is not well known. Aggregated knowledge of its successful validation in various empirical cases is not available. In order to address this gap of knowledge, to we have carried out a Systematic Literature Review (SLR) for synthesize the existing knowledge. Our literature review questions were the following.

RQ1: What has been the focus and trends in DFT related research?

RQ2: What are the practical application contexts where DFT has been applied?

By aggregating evidence of DFT and its practical application from large amount of studies, the development, current situation and future directions of DFT research can be assessed.

Systematic Literature Review Process

The SLR conducted in this study was done following the principles of the guidelines proposed by Kitchenham (2004). The phases of the process are described below.

Literature Databases

In this literature review, the following four literature databases were used. These literature databases can be expected to cover comprehensively the scientific publications on the topic.

- **Scopus:** Covers over 21.000 peer-reviewed titles from more than 5,000 international publishers and over 5.5 million conference papers.
- **ScienceDirect:** Includes publications for scientific, technical and medical research.
- **IEEE Xplore:** Includes 121 IEEE transactions, journals, magazines and over 400 conference proceedings published since 1988 plus select content back to 1950, and all current 900 IEEE standards.
- **ACM:** Includes all computer science literature originating to Association for Computer Machinery.

Search String

As the focuses of the study was to review the applications of DFT, the major search terms were chosen to be “Decision Field Theory” and “Application”. The synonyms, abbreviations and alternative spellings were identified. After several trials searches with different combination of terms, final search string was specified as:

(decision field theory OR dynamic decision making)
AND
(application OR model OR system OR framework OR software)

Selection Criteria

Based on the characteristics of review questions, the inclusion and exclusion criteria were pre-defined as follows.

Inclusion criteria (IC):

- IC1:** The paper should be conference paper, or journal paper, or book chapter, or survey report and written in English; the paper should be related to the topics specified by the search string.
- IC2:** The articles propose context in which the decision field theory or dynamic decision making approach is applied. Meanwhile, corresponding application or model (or system or framework or software) is developed.
- IC3:** The paper detail the generation process of application or model (or system or framework or software).

Exclusion criteria (EC):

- EC1:** The duplicated articles caused by combining of the different literature databases should be removed; the papers irrelevant to the topics specified by the search string should be rejected.
- EC2:** The context addressed is related to automatic decision making, for example, robotic studying, military command management, multi-core chip research and so forth.
- EC3:** The article only introduces basic and general conception or history of decision making.

Screening Process

Then, paper selection and screening was conducted rigorously in three steps, complying with the specified inclusion and exclusion criteria. The searching of the databases identified altogether 790 papers, which were screened according the following steps.

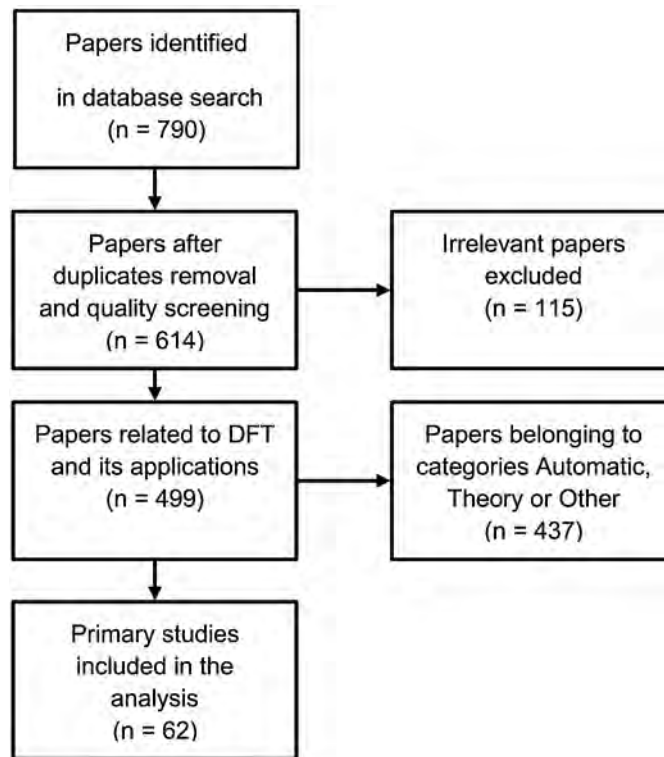
In the Step 1, the duplicates of the papers were removed and the inclusion and exclusion criteria

IC1 and EC1 were applied. During the process of Step 1, the title, keywords and abstract was also read, and the papers not related to the topic were rejected. Additionally, the papers were labeled according the application types by the keywords: Automatic (short for “automatic decision making”), Support (short for “decision support for human decision making”), Other (related to subject indirectly or the application type could not be identified exactly). After reading keywords, abstract and title, the papers labeled in support, automatic and others were accept while the irrelevant papers were reject. Consequently, the accepted papers were passed to Step 2 for further selection and the rejected papers were excluded.

In the Step 2, the inclusion and exclusion criteria IC2 and EC2 were applied to the papers remaining of the previous step. In this step, the papers were further classified according the research type to Application (short for “application of DFT”), Theory (short for “theory of DFT”) and Other (related to subject indirectly or the research type could not be identified exactly). The irrelevant papers were rejected. The accepted papers were then passed to the following step for further screening. As the focus of the study was DFT application to decision support, the Step 3 was carried out only to the papers belonging in the classes of application type Support and research type Application. The papers with a focus on automatic decision making or theory were excluded from the next step.

In the step 3, the IC3 and EC3 criteria were used for reviewing the remaining papers that belonged to the classes of application type Support and research type Application. The requirements in this step were that the paper should clearly present the application of the DFT in detail, rather than having only a general description, and paper must have a clear application context in which the DFT is applied. This final screening step resulted in 62 primary papers that were further used for full detailed review and analysis. The flow diagram of the whole process is presented in the Figure 2 below.

Figure 2. Study selection process



The number of the papers after each step is presented in Table 1 by the literature databases. From the table you can see that initially the Scopus contributes the largest amount of papers on the topic, followed by ACM. ScienceDirect and IEEE included the least relevant papers. However, considering the primary papers after the last step, where the applications of the DFT in decision support is in focus, ScienceDirect contains highest percentage of relevant papers, followed by IEEE.

DFT Research Focus and Trends

During the paper screening process Step 1 and Step 2, the papers were classified according the application type {*Support*, *Automatic*, *Other*} and research type {*Application*, *Theory*, *Other*}. The class *Other* in both of the types indicates that the paper could not be clearly classified to neither of the other classes in the review process. The classification of the DFT related papers ac-

Table 1. Number of papers by the databases after each step

Step	Scopus	IEEE	ACM	ScienceDirect	Total	Percentage Remaining After the Step
Database Search	477	89	133	91	790	100%
Step 1	308	86	130	90	614	78%
Step 2	251	83	81	84	499	63%
Step 3	22	11	10	19	62	8%
Primary Paper Percentage	5%	12%	8%	21%	8%	

Table 2. Frequency of different types of papers

Classification	Support	Automatic	Others	Total
Application	62	73	0	135
Theory	208	12	134	354
Others	0	10	0	10
Total	270	95	134	499

cording their types allows us to map and analyze the popularity and trends of the DFT research. The focus of the DFT research can be analyzed by cross-tabulating the two paper types to form publication frequency Table 2.

From the Table 2 we can see that the number of published theory papers (354) is about two and a half times the number of application papers (135).

On the other hand, the relationship of the decision support papers frequency to the decision automation papers frequency is at the similar level. Based on that, we can conclude that DFT research has focused, until now, mostly on DFT theory (354) and on decision support (270). The theory of DFT in decision support appears to be clearly the most frequent research topic (208 papers).

Figure 3. Frequency of published DFT papers by application type

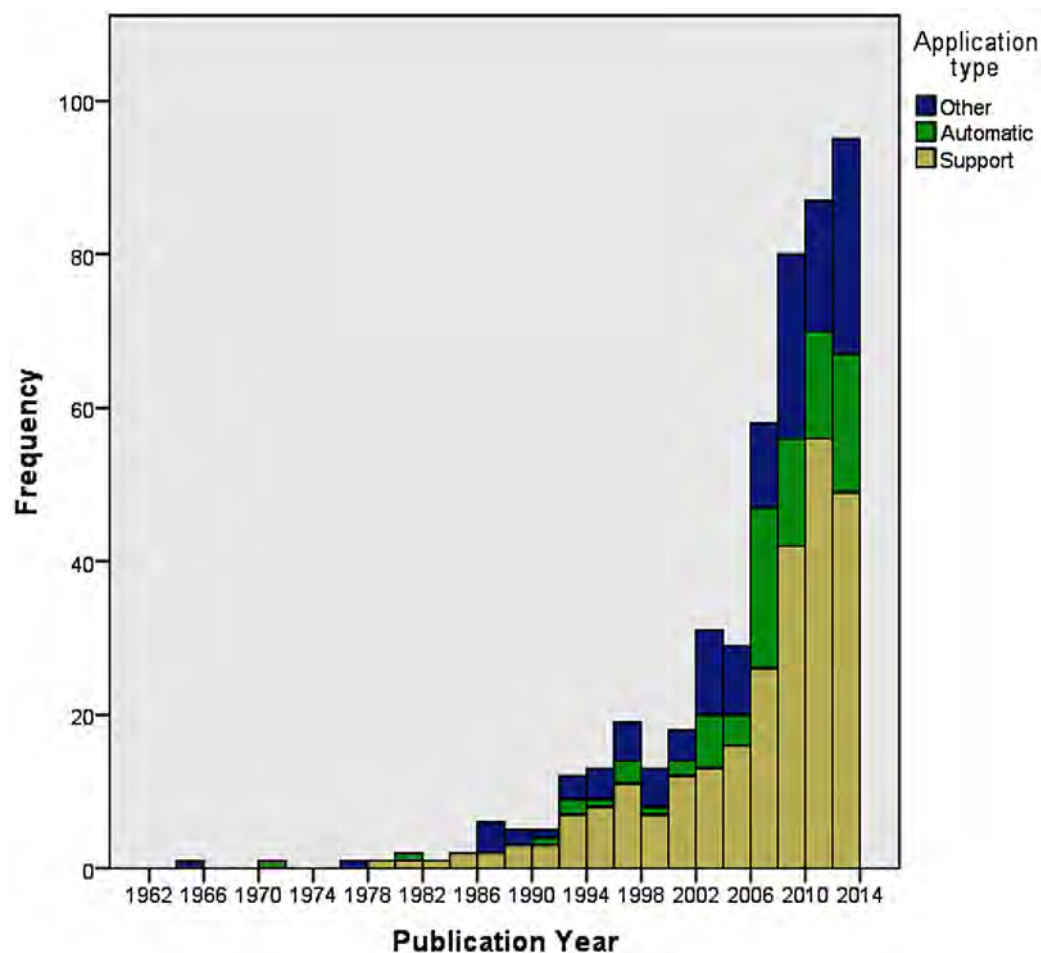
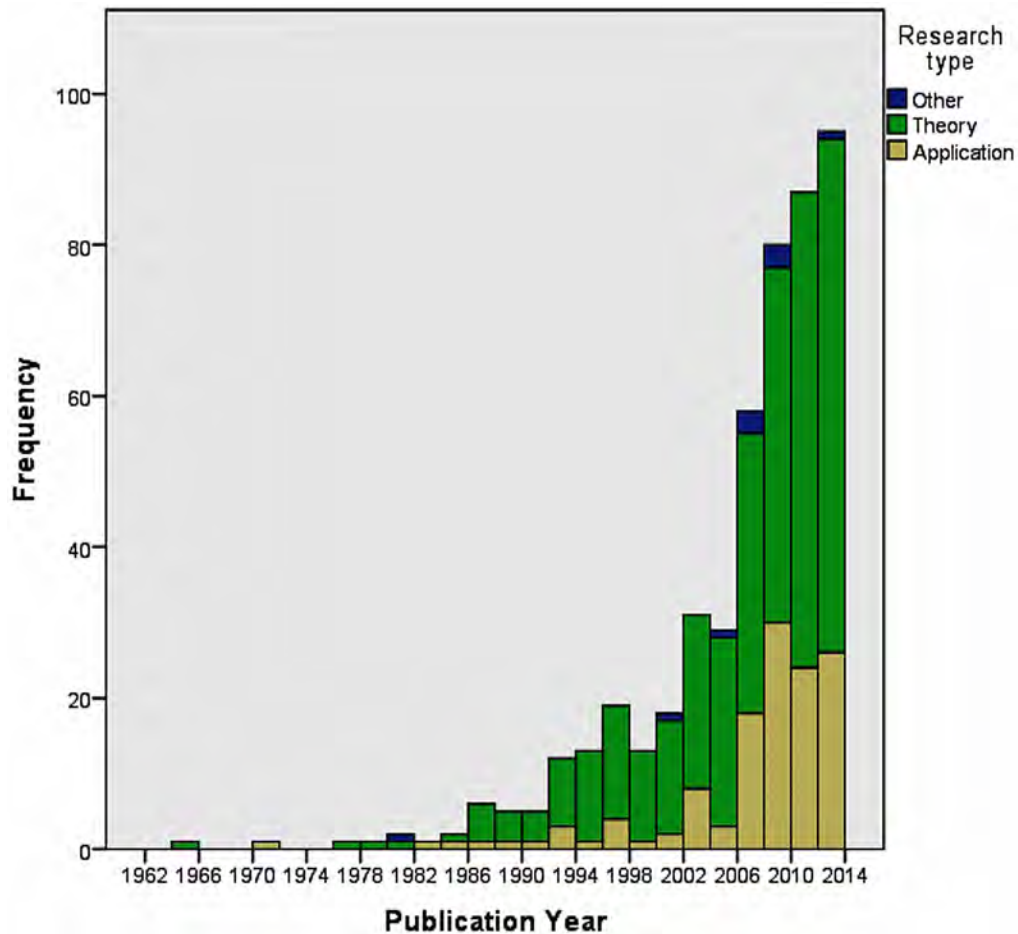


Figure 4. Frequency of published DFT papers by research type



The next interesting question is the trend of research interest in dynamic decision making and DFT during the time. The Figure 3 shows the frequency of the published papers till year 2014 by application type and Figure 4 by research type.

The figure shows clearly how the interest in DFT has been continually increasing during the years. The growth is seems to be almost exponential. Interest in decision support appears to be growing more than in decision automation. This can be caused by the growing interest in decision support systems and decision methods in ICT research during the last years, and also that DFT maybe suits better for supporting human decision making than for decision automation.

The Figure 4 shows that the number of publications on both Application and Theory have been

increasing, relatively approximately at the same phase. This result could imply that while the DFT applications are of growing interest, the theory is still at some level also in developing phase and there are possibilities to improve it.

DFT Application Areas

The analysis of the publication frequencies and trends of research papers clearly showed the increasing interest in developing and applying dynamic decision making and DFT. Growing number of research papers are reporting successful application of DFT to different empirical situations and cases. In order to gain understanding of the extent of these applications, we reviewed and analyzed the 62 decision support application

papers that were identified as the primary studies in our SLR screening Step 3.

The DFT appears to have been widely applied into many areas. For the purpose of the analysis, we categorized the application areas appearing in the primary studies according their context into 8 categories: (1) Energy use management; (2) Decision making behavior and process of Human; (3) Manufacturing and retail industry; (4) Transportation tool allocation or routing; (5) Disaster management; (6) IT technology; (7) Social, Economic and Medical issues; (8) Others. The results of the review and analysis are presented in the Table 3.

Seeing from the context categories of all primary studies, DFT theory is widely used in all sector of society. Broadly speaking, most researchers focus on social, economic and medical industries that account for 10 out of 62 papers. The energy management industry and human decision making behavior are right behind, with 9 relevant papers respectively, followed by manufacturing and retail industry. The difference in numbers of studies across different domains is very small.

The distribution of the application over so wide variety of application areas show that the DFT is not limiting in its application possibilities and that it is gaining more popularity across the different fields of applied sciences.

It was also observed that the primary studies include practical applications of many different DFT types and variations, for example rule-based decision field theory, multi-criteria decision field theory, multi-alternative decision field theory, and time critical decision field theory. The researchers (Roe et al., 2001) rebuild decision field theory as a connectionist network and use it to encompass multi-alternative decision making that is adopted successfully in economic or psychological context (Usher, 2004). The rule-based decision field theory is firstly generated by Johnson et al. (2005) and subsequently apply in routing field to study driver behavior (Gao, 2011) or control driving speed (Zhao, 2011). Lin et al. (2008) propose a dynamic multi-attribute decision making model to deal with the problem of uncertain information and aggregate the multi-period evaluations.

Table 3. Context categories of the DFT decision support application studies

Category	Context	Amount	Percentage
Social, Economic and Medical issues	Trading (2); Stock market; Business risk management; Social security(3); Patient management; Medicine (2)	10	16%
Energy use management	Electricity market (3); Power market; Mineral resource; Water management (2); Petroleum industry; Material recycle	9	15%
Human decision making behavior and process	Market participant behavior (3); Pedestrian behavior; Human behavior (3); Human decision making process(2)	9	15%
Manufacturing and retail industry	Supply chain; Supplier and retailer; Supplier selection (3); Vendor selection; Garment style design; Strategic planning forecast	8	13%
Transportation tool allocation or routing	Vehicle allocation; Port berth allocation; Container dispatching; Routing(4)	7	11%
IT technology	Air combat tactics; Ubiquitous; Game playing; Spam mail mitigation; Future call predication; Multi-sensor detection	6	10%
Disaster management	Disaster management	5	8%
Others		8	13%

FUTURE RESEARCH DIRECTIONS

Our systematic literature review of past DFT related research confirmed that the researchers and practitioners have a growing interest in developing and extending the DFT theory and model further as well as apply it in more widely in new contexts. This reveals the two main research directions, theoretical and practical: extension and validation of the DFT model to be applicable to more complex and realistic decision making situations and extending the practical applications of the DFT to new areas.

In our systematic literature review, we have identified the published papers studying the DFT theory and its applications, and reviewed more deeply the decision support application papers. This approach appeared to be successful for gaining evidence on the trends and current situation of the DFT research. Therefore this type of research would be recommendable to be taken further with more extensive and deeper analysis. For example, the DFT application cases could be analyzed and evaluated for gaining better understanding how well and trustworthily the applicability and efficiency of the DFT has been validated.

CONCLUSION

ICT technology is allowing development of increasingly advanced applications and services for supporting and automating our daily living and related small and big decisions. Therefore human decision making, related theories and models, are gaining more attention. Both researchers and practitioners are interested in finding applicable formal models for modeling, explaining and predicting human decision making that can be used for designing new applications and services. The DFT, as a dynamic decision model, which has more grounding on psychological than economic principles, is able to model natural human decision making processes well in many situations. Its dynamic decision model is in many cases better

than the theories that are relying on static decision models. Our review study revealed that DFT has also practically validated in many reported empirical cases that present wide variety of application contexts. The study also showed that the interest in the DFT has been growing almost exponentially during the last years. The research on both theory and applications are actively going on and expanding. As a rather young theory, DFT has still need for extensions; that was also shown in the trends that our systematic literature review revealed. In its current forms, it is still limited to relatively simple decision making cases that are not covering completely the complexity of real world practical situations and the bounded rationality (cf. Simon, 1955) of human decision making. However, it has a high potential to be further developed and adapted to new kinds of application contexts.

REFERENCES

- Abad, A. G., Jin, J. J., & Son, Y. J. (2014). Estimation of expected human attention weights based on a decision field theory model. *Information Sciences*, 278, 520–534. doi:10.1016/j.ins.2014.03.069
- Bussemeyer, J. R. (1985). Decision making under uncertainty: A comparison of simple scalability, fixed sample, and sequential sampling models. *Journal of Experimental Psychology*, 11, 538–564. PMID:3160815
- Bussemeyer, J. R., & Johnson, J. G. (2008). Micro-process models of decision making. In *Cambridge handbook of computational psychology*, (pp. 302–321). Cambridge University Press. doi:10.1017/CBO9780511816772.014
- Bussemeyer, J. R., & Townsend, J. T. (1993). Decision Field Theory: A Dynamic-Cognitive Approach to Decision Making in an Uncertain Environment. *Psychological Review*, 100(3), 432–459. doi:10.1037/0033-295X.100.3.432 PMID:8356185

- Cheung, R. K., & Muralidharan, B. (1999). Impact of dynamic decision making on hub-and-spoke freight transportation networks. *Annals of Operations Research*, 87, 49–71. doi:10.1023/A:1018909825336
- DeGroot, M. H. (1970). *Optimal statistical decisions*. New York: McGraw-Hill.
- Diederich, A. (1997). Dynamic stochastic models for decision making under time constraints. *Journal of Mathematical Psychology*, 41(3), 260–274. doi:10.1006/jmps.1997.1167 PMID:9325121
- Du, P., & Liu, H. Y. (2010). *Study on air combat tactics decision-making based on bayesian networks*. In *Information Management and Engineering (ICIME), 2010. The 2nd IEEE International Conference on* (pp. 252-256). IEEE.
- Fishburn, P. C. (1988). Expected utility: An anniversary and a new era. *Journal of Risk and Uncertainty*, 1(3), 267–284. doi:10.1007/BF00056138
- Gao, F., & Wang, M. (2011). Modeling E-route Driver Route Choice Behavior under Real-Time Traffic Information. *Journal of Information and Computational Science*, 8(7), 1053–1062.
- Johnson, J. G., & Busemeyer, J. R. (2005). *Rule-based Decision Field Theory: A dynamic computational model of transitions among decision-making strategies*. In *The routines of decision making*. Mahwah, NJ: Erlbaum.
- Kitchenham, B. (2004). Procedures for performing systematic reviews. *Keele University*, 33, 1–26.
- Lin, Y. H., Lee, P. C., & Ting, H. I. (2008). Dynamic multi-attribute decision making model with grey number evaluations. *Expert Systems with Applications*, 35(4), 1638–1644. doi:10.1016/j.eswa.2007.08.064
- Pearson, L. J., Coggan, A., Proctor, W., & Smith, T. F. (2010). A sustainable decision support framework for urban water management. *Water Resources Management*, 24(2), 363–376. doi:10.1007/s11269-009-9450-1
- Roe, M. R., Busemeyer, J. R., & Townsend, J. T. (2001). Multialternative Decision Field Theory: A Dynamic Connectionist Model of Decision Making. *Psychological Review*, 108(2), 370–392. doi:10.1037/0033-295X.108.2.370 PMID:11381834
- Savage, L. J. (1954). *The foundations of statistics*. New York: Wiley.
- Simon, H. A. (1955). A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics*, 69(1), 99–118. doi:10.2307/1884852
- Sucky, E. (2007). A model for dynamic strategic vendor selection. *Computers & Operations Research*, 34(12), 3638–3651. doi:10.1016/j.cor.2006.01.006
- Usher, M., & McClelland, J. L. (2004). Loss aversion and inhibition in dynamical models of multi-alternative choice. *Psychological Review*, 111(3), 757–769. doi:10.1037/0033-295X.111.3.757 PMID:15250782
- Van Horne, J. C., & Parker, G. G. (1967). The random-walk theory: An empirical test. *Financial Analysts Journal*, 23(6), 87–92. doi:10.2469/faj.v23.n6.87
- Vescoukis, V., & Dulamis, N. D. (2011). Disaster management evaluation and recommendation. *Games and Virtual Worlds for Serious Applications (VS-GAMES), 2011 Third International Conference on*, 244-249. doi:10.1109/VS-GAMES.2011.43
- Zhao, G., Wu, C., & Ou, B. (2011). Mathematical modeling of average driver speed control with the integration of queuing network-model human processor and rule-based decision field theory. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 55, No. 1, pp. 856-860). Sage Publications. doi:10.1177/1071181311551178

ADDITIONAL READING

- Badiru, A. B., Simin Pulat, P., & Kang, M. (1993). DDM: Decision support system for hierarchical dynamic decision making. *Decision Support Systems*, 10(1), 1–18. doi:10.1016/0167-9236(93)90002-K
- Campanella, G., & Ribeiro, R. A. (2011). A framework for dynamic multiple-criteria decision making. *Decision Support Systems*, 52(1), 52–60. doi:10.1016/j.dss.2011.05.003
- Cohen, M., Etner, J., & Jeleva, M. (2008). Dynamic decision making when risk perception depends on past experience. *Theory and Decision*, 64(2-3), 173–192. doi:10.1007/s11238-007-9061-3
- Du, P., & Hai-yan, L. (2010). *Study on air combat tactics decision making based on bayesian networks*. The 2nd IEEE International Conference on Information Management and Engineering (ICIME), pp. 252–256. doi:10.1109/ICIME.2010.5478340
- Grimaila, M. R., & Badiru, A. (2013). A hybrid dynamic decision making methodology for defensive information technology contingency measure selection in the presence of cyber threats. *Operations Research*, 13(1), 67–88. doi:10.1007/s12351-010-0102-2
- Khaji, M. R., Shafaei, R., Mohebbi, S., & Aghaie, A. (2009). A system dynamics approach to decision process in supply chain. *IEEE/INFORMS International Conference on Service Operations, Logistics and Informatics (SOLI '09)*, pp. 681–686. doi:10.1109/SOLI.2009.5204019
- Lee, S., Son, Y., & Jin, J. (2008). Decision field theory extensions for behavior modeling in dynamic environment using bayesian belief network. *Information Sciences*, 178(10), 2297–2314. doi:10.1016/j.ins.2008.01.009
- Li, J., & Li, Y. (2012). Cognitive model based fashion style decision making. *Expert Systems with Applications*, 39(5), 4972–4977. doi:10.1016/j.eswa.2011.10.017
- Martinelli, G., Eidsvik, J., & Hauge, R. (2013). Dynamic decision making for graphical models applied to oil exploration. *European Journal of Operational Research*, 230(3), 688–702. doi:10.1016/j.ejor.2013.04.057
- Peng, M., & Zhang, L. M. (2013). Dynamic decision making for dam-break emergency management-part 1: Theoretical framework. *Natural Hazards and Earth System Sciences*, 13(2), 425–437. doi:10.5194/nhess-13-425-2013
- Peng, M., & Zhang, L. M. (2013). Dynamic decision making for dam-break emergency management-part 2: Application to tangjiashan landslide dam failure. *Natural Hazards and Earth System Sciences*, 13(2), 439–454. doi:10.5194/nhess-13-439-2013
- Steyvers, M., Lee, M. D., & Wagenmakers, E. (2009). A bayesian analysis of human decision making on bandit problems. *Journal of Mathematical Psychology*, 53(3), 168–179. doi:10.1016/j.jmp.2008.11.002
- Tsao, Y., Liao, L., & Sun, P. (2013). Dynamic decision making for an inventory system with time-varying demand. *Mathematical and Computer Modelling of Dynamical Systems*, 19(5), 401–416. doi:10.1080/13873954.2013.777747
- Wei-bing, L., & Li, M. (2011). *Dynamic decision making in social security system*. *International Conference on Management Science and Industrial Engineering (MSIE)*, pp. 76–79. doi:10.1109/MSIE.2011.5707509
- Xi, H., Son, Y., & Lee, S. (2010). *An integrated pedestrian behavior model based on extended decision field theory and social force model*. In *Proceedings of the 2010 Winter Simulation Conference*, 824–836. doi:10.1109/WSC.2010.5679108
- Xiang, Y., & Poh, K. (2002). Time-critical dynamic decision modeling in medicine. *Computers in Biology and Medicine*, 32(2), 85–97. doi:10.1016/S0010-4825(01)00036-1 PMID:11879822

Yuan, W., Guan, D., Hung, L. X., Lee, Y., & Lee, S. (2006). *A trust model with dynamic decision making for ubiquitous environments*. *14th IEEE International Conference on Networks (ICON '06)*, Vol 1, pp. 1-6. doi:10.1109/ICON.2006.302565

KEY TERMS AND DEFINITIONS

Decision Field Theory (DFT): A dynamic-cognitive theory and model of human decision making.

Dynamic Decision Making: Decision making under uncertainty in a changing environment where the consequences of the earlier decisions and actions affect to the present decisions.

Effectively Communicating With Group Decision Support Systems Using Information Theory

D**Jamie S. Switzer***Colorado State University, USA***Ralph V. Switzer***Colorado State University, USA*

INTRODUCTION

This entry describes the results from a case study conducted using information theory to examine the effectiveness of communicating using group decision support system (GDSS) technology. At its most basic level, information theory provides the means to measure the efficiency of communication systems. Using information theory as the theoretical foundation, this entry examines how the use of GDSS facilitated computer-mediated communication (CMC) for one particular business with respect to entropy, redundancy, and noise, which are key components in information theory.

BACKGROUND

As originally articulated, information theory is actually a mathematical theory grounded in the field of electrical engineering designed to evaluate the performance of communication systems (Mahowald, Fedorenko, Piantadosi, & Gibson, 2013; Ziemer & Tranter, 2002). Claude Shannon was an electrical engineer and mathematician working at Bell Laboratories in the 1940's. In 1948 he published his landmark paper, "A Mathematical Theory of Communication" (Shannon, 1948). This seminal work was concerned with the transmission and storage of information (Lafrance, 1990), and provided "the analytical

tools to evaluate the amount of information contained in message signals, and to compare the performance of actual systems" (Carne, 1999, p. 172). According to Gallager (2001), Shannon's theory established a conceptual basis for modern digital communications, particularly with regard to data compression, data encryption, and data correction (Gappmair, 1999).

Information theory has been applied to many disparate disciplines, such as computer science, statistics, cybernetics, physiology, psychology, library science, biology, physics, economics, music, and art (Asadi, 2015; Dahling, 1962; De-Fleur & Larsen, 1987; Overstreet, 1984; Pierce, 1980). In the field of communication, information theory has been applied to such areas as speech, linguistics, forensics, broadcasting, journalism, and even animal communication as studied by animal scientists (McCowan, Doyle, & Hanser, 2002; Scott-Phillips, 2015; Stephens, Barrett, & Mahometa, 2013; Watt & Krull, 1974).

There is some disagreement among scholars as to the applicability of information theory to the field of communication, in the non-technical, non-engineering sense (Cherry, 1957; Devlin, 1999; Weaver & Weaver, 1965; Young, 1987). Shannon himself initially cautioned against applying his theory to human communication (Haken & Portugali, 2015, p. 33; Rogers & Valente, 1993). But Shannon, in conjunction with Warren Weaver, did express some "optimism as to its wider ap-

plicability” (Warner, 2001, p. 24). In subsequent years other communication scholars have also explored more general applications of information theory (Pierce, 1980; Rogers & Valente, 1993). Indeed, Shannon’s simple yet elegant model of communication systems forms the basis of many introductory communication courses.

At its most basic level, information theory provides the means to measure the efficiency of communication systems. An information source sends a message (of any type) using a transmitter that carries a signal through a channel. The receiver then receives that signal, and the message arrives at its destination. There are additional factors, however, impacting the effective operation of that process, including entropy, redundancy, and noise.

Entropy, as defined by information theory, is concerned with the many different ways a message can be constructed, depending on the circumstance. An ambiguous or vague message, for example, has a high degree of entropy because the receiver cannot accurately interpret the message that the sender intended, and there is a measure of uncertainty and unpredictability regarding the message (Heath & Bryant, 2000). A message that is clear and concise will reduce the amount of entropy in a given situation.

Redundancy also reduces entropy. The more the same information is conveyed in a particular message, the more certain the receiver is of the interpretation of the message as the sender originally intended. There must be a balance of entropy and redundancy in an effective communication, but that balance is often impacted by the amount of noise present in the channel used to convey the message (Severin & Tankard, 2001).

Noise is simply some sort of interference that occurs in the transmission of the message between sender and receiver (Perry, 2002). Noise can take many forms, including extraneous information that causes a loss of intended information. The noisier the channel, the greater the need for redundancy (Cragan & Shields, 1998; Stephens et al., 2013). Anything that comes along with the signal that was not put there by the information source is noise.

Group Decision Support Systems (GDSS) are computer-based systems “designed to help committees, project teams, and other small groups with activities such as problem identification and analysis, decision making, planning, creativity, conflict management, negotiation, and meeting management” (DeSanctis et al, 2008, p. 552). They structure, store, facilitate, and distribute human communication, and aggregate views of a group of people (Ishizaka & Nemery, 2013).

Research has shown that, in general, GDSS are effective tools for generating ideas and plans (Eden & Ackermann, 2014; Power, 2002) and identifying, collecting, organizing and interpreting the thoughts of the participants (Roszkiewicz, 2007). GDSS facilitates working interactively and dynamically with collective group data. All GDSS software allows for the anonymous generation of ideas, shared and parallel input, and complete and accurate record keeping.

ISSUES AND PROBLEMS

A large technology company contracted with a Carnegie Research I university to facilitate GDSS meetings with its employees, and to conduct a case study to analyze the effectiveness of the use of GDSS for the company based on the results of the GDSS sessions. Case study is not a methodological choice but a choice of what is to be studied (Stake, 2000); here the choice was to study the impact of using the GDSS system with the company’s employees.

A GDSS system can be used virtually or with co-location of the participants. In this instance, the business involved engaged in a series of computer-mediated activities while they were co-located. Chen and Linn (2012) found that group decision-making is more effective when participants are face-to-face rather than dispersed. The GDSS meetings were conducted at a local university in one of its computer labs.

The business’ employees agreed unanimously (n=45) that using GDSS was more effective than

traditional “flip chart” meetings. They were also unanimous in finding the tool easy to use and understand. On a scale of 1-10 (10 being the best), 87% of the meeting participants rated the entire GDSS experience an 8 or higher.

When asked to identify what they particularly liked about collaborating using GDSS, the feedback was consistent among the employees. The anonymity of the input, the ability to generate many ideas in a non-threatening setting, the opportunity to stimulate participation and interaction, the ease of use of the product, and the efficiency of the use of groupware technology over more traditional forms of brainstorming meetings were the most important aspects of the GDSS experience.

Al Shishany and Adams (2013) found that anonymity plays a significant role in fostering discussions. However, Scott, Quinn, Timmerman, and Garrett (1998) found that some members of the groups they studied reported feeling frustrated at not being able to figure out who was saying what. In this particular case study, however, that desire was not expressed. Directly related to the concept of anonymity is the articulation of the computer-mediated collaborative meeting participants’ pleasure at being able to contribute in an environment they felt was non-threatening.

Using GDSS has been proven to increase user participation (Adla, Zarate, & Soubie, 2011). The use of GDSS is intended, among other things, to promote the opportunity to generate a lot of ideas from participants since they can input their thoughts using the computer without interrupting another person (Olson & Olson, 2003).

The use of GDSS in a computer-mediated collaborative meeting environment can greatly increase efficiency (Khan, 2014). As Lococo and Yen state, “Efficient sharing of ideas can be transformed into shared understanding and into shared priorities” (1998, p. 91). Participants in this particular case study concurred that the use of groupware technology was in general more efficient than traditional forms of meetings.

There was very little negative feedback regarding the actual use of GDSS in the collaboration.

The comments involved wanting a clearer understanding at the outset of the goals of the meeting, a feeling that the process was too structured, and a concern that management would not use the information generated in the session to improve the division.

The impressions of the actual meeting room environment had much more of a negative impact on the GDSS participants. On a scale of 1-10 (with 10 being the best), over 25% rated the physical environment of the lab a 6 or below. There was tremendous consistency in the feedback: the glare off the glass over the monitor made it very difficult to see the screen, the workstation ergonomics were poor and the chairs were too far apart, the ambient noise of the computers was very loud, and the room temperature was either a) too hot, or b) too cold.

SOLUTIONS AND RECOMMENDATIONS

GDSS facilitates transactional computer-mediated communication. “Transactional” implies a give-and-take situation; an interpersonal communication relationship in which parties alternate in their roles as sender, receiver, and information processor, thereby exchanging information. “Mediated” signifies that media technologies are involved. In this particular case study, the GDSS technology allowed the participants to input data into the system (sender), read the other meeting participants’ comments (receiver), and then discuss the information and come to consensus (information processor). The process helped the group understand the issues facing them (French, 2007) and produced potential solutions.

Also, it is important to note that this case study occurred in a real-life setting and not in the artificial confines of a laboratory experiment. As Gopal and Prasad state, “...technology cannot be studied outside its social context” (2000, p. 512). The collaborators were not strangers thrown together randomly, but team members who

worked together and knew each other well. The use of zero-history groups in GDSS research has been criticized (Putnam & Stohl, 1996), making the case for “the study of intact natural groups embedded in larger structures” (Scott, 1999, p. 464), as is the case here.

We would argue that the use of GDSS in a computer-mediated collaborative setting reduces the amount of entropy and noise possible while increasing the opportunity for redundancy, thereby eliminating potential barriers to effective communication.

As previously discussed, the concept of entropy has to do with the different ways a message can be constructed. Vague or ambiguous messages have a high degree of entropy. In fact, all computer-mediated communications have the potential for heightened entropy because of the nature of CMC itself. CMC relies heavily on text-based messages, which can increase the amount of ambiguity in a message (Klein, Clark, & Herskovitz, 2003) because of the environment is mediated and not face-to-face (Adla et al., 2011). There are no visual cues, no way “...of allowing the receiver to send the electronic equivalent of...an ‘uh-uh,’ or any other audio-visual reactions which play such a critical role in face-to-face interaction” (Crystal, 2001, p. 30). As Scott states, “the challenge is to understand how the technology (or distribution method) such as a computer can change the meaning of the message” (Scott, 2009, p. 28).

Anonymity, too, can increase entropy. Knowing who sent a particular message can assist the receiver in interpreting that message accurately. For example, what if the message, “we are all going to be fired from our jobs next week,” scrolls up on the screen during a GDSS session. The meaning of the message itself is quite obvious, yet there would be a great deal of uncertainty among the participants. Imagine the thoughts going through everyone’s mind! Knowing who sent that message, however, can help alleviate the uncertainty. If the participants knew the sender was a well-known company joker, the message would be interpreted as just another prank on the

joker’s part. If, however, the participants knew the sender was the president of the company, the message would be interpreted in a completely different way. However, one of the strongest features of GDSS systems is allowing the participants to communicate anonymously to promote dialog (McLeod, 2011).

Using GDSS in a computer-mediated collaborative setting encourages redundancy, which in turn reduces entropy. The software permits everyone to view the comments of all of the participants in real-time. As evidenced by the comments generated in this case study, there was a tremendous amount of redundancy. Perhaps someone’s memory was jogged when they saw a particular comment and was reminded that s/he too wanted to make just such a remark. Or maybe people wanted to make sure they entered their “me too” as well. Regardless of the rationale, GDSS affords the perfect opportunity for a considerable amount of redundancy within the computer-mediated collaborative structure.

As noted earlier, noise is interference in the transmission of a message; anything that comes along with the signal that was not originally put there by the sender. In a traditional (face-to-face) collaborative environment, there are any number of possible distractions that could be construed as noise. People interrupting each other or shouting to be heard over someone else talking would most definitely impact the receipt of a message. Noise – literally.

But the use of a GDSS to collaborate eliminates noise. Everyone is free to input as many comments as they choose without being interrupted or shouted down. The message is clear and concise and received just as the sender intended.

The specific groupware characteristics the employees identified as effective (anonymity, safe setting, brainstorming and idea generating, and efficiency) are precisely why groupware products are used by corporations in meetings and in the decision-making process (Scott, Allen, Rogelberg, & Kello, 2015). In this case study, employees were able to provide management with valuable

input in an effective and efficient manner. This is also consistent with groupware theory. The few negative comments were organizational issues and not GDSS issues.

FUTURE RESEARCH DIRECTIONS

Information theory, as originally articulated by Shannon, is linear and mechanistic (Perry, 2002). A suggestion for future research is to examine the role of GDSS in computer-mediated collaborative settings using more transactional theories and models. Another area for future research is to reexamine information theory itself with respect to more modern computer-mediated collaborative technologies such as GDSS.

CONCLUSION

GDSS can be a very powerful tool for organizations. Used properly, the technology has tremendous potential to generate ideas and assist in decision making to improve the functioning of corporations. While “the same technology will not provide the same results with each group and in each setting” (Boiney, 1998, p. 343), collaborating using GDSS has many benefits, including the reduction of entropy and noise and the increased opportunity for redundancy as predicted by information theory.

Pollock and Kanachowski (1993) define a GDSS as a system where group members use computers interactively to support the group’s decision-making capacity. The use of GDSS has also been shown to produce more creative solutions as well as increasing the commitment of participants to decisions made in the GDSS sessions (Golden & Pineo, 2014). Based on the case study results, the GDSS obviously met this particular company’s needs effectively and efficiently.

Although some research suggests there is inconclusive evidence GDSS actually facilitates computer-mediated collaboration, we believe in

this case the use of the GDSS greatly enhanced the collaborative experience. Based on the results of the case study, the company and its remote division were extremely satisfied with the results. According to Poole and Holmes (1995), the strength of GDSS comes from the technology’s ability to enhance communication and information exchange, complex information processing tasks, and coordination and organization of group activities. These strengths were evident in the results from this case study, and have very practical applications.

The information theory concepts of entropy, redundancy, and noise are highly relevant in examining computer-mediated collaborations using GDSS. The success of a GDSS intervention depends upon several different factors, including the nature of the technology used (Hiltz, Dufner, Fjermestad, Kim, Ocker, Rana, & Turoff, 2012). As the number of communication channels available expands with the diffusion of mobile devices and noise levels in organizational environments increases, using information theory is a highly relevant approach for communication scholars (Stephens et al., 2013) to understand the impact of different technologies on the communication process.

REFERENCES

- Adla, A., Zarate, P., & Soubie, J. (2011). A proposal of toolkit for GDSS facilitators. *Group Decision and Negotiation*, 20(1), 57–77. doi:10.1007/s10726-010-9204-8
- Al Shishany, A., & Adams, J. (2013). Anonymity and group task-conflict in GDSS supported meetings. *American Academic & Scholarly Research Journal*, 5(3).
- Asadi, M. (2015). Information theory research trend: A bibliometric approach. *SLIS Connecting*, 4(1).

- Boiney, L. G. (1998). Reaping the benefits of information technology in organizations: A framework guiding appropriation of group support systems. *The Journal of Applied Behavioral Science*, 34(3), 327–367. doi:10.1177/0021886398343006
- Carne, E. B. (1999). *Telecommunication topics: Applications of functions and probabilities in electronic communications*. Upper Saddle River, NJ: Prentice Hall PTR.
- Chen, J. V., & Linn, K. (2012). User satisfaction with group decision making process and outcome. *Journal of Computer Information Systems*, 52(4), 30–39.
- Cherry, C. (1957). *On human communication: A review, a survey, and a criticism*. Cambridge, MA: MIT Press.
- Cragan, J. F., & Shields, D. C. (1998). *Understanding communication theory: The communicative forces for human action*. Boston: Allyn and Bacon.
- Crystal, D. (2001). *Language and the internet*. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9781139164771
- Dahling, R. L. (1962). Shannon's information theory: The spread of an idea. In W. Schramm (Ed.), *Studies of innovation and of communication to the public*. Stanford, CA: Stanford University, Institute for Communication Research.
- DeFleur, M. L., & Larsen, O. N. (1987). *The flow of information: An experiment in mass communication*. New Brunswick, NJ: Transaction Books.
- Devlin, K. (1999). *Infosense: Turning information into knowledge*. W.H. Freeman and Company.
- Eden, C., & Ackermann, F. (2014). Joined-up policy making: Group decision and negotiation practice. *Group Decision and Negotiation*, 23(6), 1385–1401. doi:10.1007/s10726-013-9375-1
- French, S. (2007). Web-enabled strategic GDSS, e-democracy and Arrows theorem: A Bayesian perspective. *Decision Support Systems*, 43(4), 1476–1484. doi:10.1016/j.dss.2006.06.003
- Gallager, R. G. (2001). Claude E. Shannon: A retrospective on his life, work, and impact. *IEEE Transactions on Information Theory*, 47(7), 2681–2695. doi:10.1109/18.959253
- Gappmair, W. (1999, April). Claude E. Shannon: The 50th anniversary of information theory. *IEEE Communications Magazine*, 37(4), 102–105. doi:10.1109/35.755458
- Golden, K. M., & Pineo, P. P. (2014). Decision-making in organizations: A case study of the use of GDSS in university planning. In E. Nikoi & K. Boateng (Eds.), *Collaborative communication process and decision making in organizations* (pp. 280–302). Hershey, PA: Business Science Reference. doi:10.4018/978-1-4666-4478-6.ch015
- Gopal, A., & Prasad, P. (2000). Understanding GDSS in symbolic context: Shifting the focus from technology to interaction. *Management Information Systems Quarterly*, 24(3), 509–546. doi:10.2307/3250972
- Haken, H., & Portugali, J. (2015). *Information adaptation: The interplay between Shannon information and semantic information in cognition*. Springer International Publishing.
- Heath, R. L., & Bryant, J. (2000). *Human communication theory and research: Concepts, contexts, and challenges*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Hiltz, S. R., Dufner, D., Fjermestad, J., Kim, Y., Ocker, R., Rana, A., & Turoff, M. (2012). Distributed group support systems: Theory development and experimentation. In G. M. Olson, T. W. Malone, & J. B. Smith (Eds.), *Coordination theory and collaboration technology* (pp. 473–506). Psychology Press.
- Ishizaka, A., & Nemery, P. (2013). A multi-criteria group decision framework for partner grouping when sharing facilities. *Group Decision and Negotiation*, 22(4), 773–799. doi:10.1007/s10726-012-9292-8

- Khan, R.A. (2014). GDSS: Tool for effective decision making. *Artificial Intelligent Systems and Machine Learning*, 3(6).
- Klein, E. E., Clark, C. C., & Herskovitz, P. J. (2003). Philosophical dimensions of anonymity in group support systems: Ethical implications of social psychological consequences. *Computers in Human Behavior*, 19(3), 355–382. doi:10.1016/S0747-5632(02)00053-5
- Lafrance, P. (1990). *Fundamental concepts in communication*. Englewood Cliffs, NJ: Prentice Hall.
- Lococo, A., & Yen, D. C. (1998). Groupware: Computer supported collaboration. *Telematics and Informatics*, 15(1-2), 85–101. doi:10.1016/S0736-5853(98)00006-9
- Mahowald, K., Fedorenko, E., Paintadosi, S. T., & Gibson, E. (2013). Info/information theory: Speakers choose shorter words in predictive contexts. *Cognition*, 126(2), 313–318. doi:10.1016/j.cognition.2012.09.010 PMID:23116925
- McCowan, B., Doyle, L. R., & Hanser, S. F. (2002). Using information theory to assess the diversity, complexity, and development of communicative repertoires. *Journal of Comparative Psychology*, 116(2), 166–172. doi:10.1037/0735-7036.116.2.166 PMID:12083612
- McLeod, P. L. (2011). Effects of anonymity and social comparison of rewards on computer-mediated group brainstorming. *Small Group Research*, 42(4), 475–503. doi:10.1177/1046496410397381
- Olson, G. M., & Olson, J. S. (2003). Groupware and computer-supported cooperative work. In J. A. Jacko & A. Sears (Eds.), *The human-computer interaction handbook: Fundamentals, evolving technologies and emerging applications* (pp. 583–595). Mahwah, NJ: Lawrence Erlbaum Associates.
- Overstreet, R. (1984). Information theory and oral interpretation. *The Southern Speech Communication Journal*, 49(Winter), 143–154. doi:10.1080/10417948409372596
- Perry, D. K. (2002). *Theory and research in mass communication: Contexts and consequences*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Pierce, J. R. (1980). *An introduction to information theory: Symbols, signals and noise*. New York: Dover Publications Inc.
- Pollock, C., & Kanachowski, A. (1993). Application of theories of decision making to group decision support systems (GDSS). *International Journal of Human-Computer Interaction*, 5(1), 71–94. doi:10.1080/10447319309526056
- Poole, M. S., & Holmes, M. E. (1995). Decision development in computer-assisted group decision making. *Human Communication Research*, 22(1), 90–127. doi:10.1111/j.1468-2958.1995.tb00363.x
- Power, D. J. (2002). *Decision support systems: Concepts and resources for managers*. Westport, CT: Quorum Books.
- Putnam, L. L., & Stohl, C. (1996). Bona fide groups: An alternative perspective for communication and small group decision making. In R. Y. Hirokawa & M. S. Poole (Eds.), *Communication and group decision making* (pp. 147–178). Thousand Oaks, CA: Sage. doi:10.4135/9781452243764.n6
- Rogers, E. M., & Valente, T. W. (1993). A history of information theory in communication research. In J. R. Schement & B. D. Ruben (Eds.), *Between communication and information* (pp. 35–56). New Brunswick, NJ: Transaction Publishers.
- Roszkiewicz, R. (2007). GDSS: The future of online meetings and true digital collaboration? *Seybold Report: Analyzing Publishing Technologies*, 7(1), 13–17.
- Scott, C., Allen, J. A., Rogelberg, S. G., & Kello, A. (2015). Five theoretical lenses for conceptualizing the role of meetings in organizational life. In J. A. Allen, N. Lehmann-Willenbrock, & S. G. Rogelberg (Eds.), *The Cambridge handbook of meeting science* (pp. 634–662). Cambridge University Press. doi:10.1017/CBO9781107589735.027

- Scott, C. R. (1999). The impact of physical and discursive anonymity on group members multiple identifications during computer-supported decision making. *Western Journal of Communication*, 63(4), 456–487. doi:10.1080/10570319909374654
- Scott, C. R., Quinn, L., Timmerman, C. E., & Garrett, D. M. (1998). Ironic uses of group communication technology: Evidence from meeting transcripts and interviews with group decision support systems users. *Communication Quarterly*, 46(3), 353–374. doi:10.1080/01463379809370107
- Scott, E. (2009). The digital research cycle: How subject matter, textual style, technology and audience combine to shape communication. *Florida Communication Journal*, 2(37), 23–41.
- Scott-Phillips, T. C. (2015). *Why human communication is different, and how language evolved to make it special*. Palgrave Macmillan.
- Severin, W. J., & Tankard, J. W. (2001). *Communication theories: Origins, methods, and uses in the mass media*. Longman.
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(3), 379–423, 623–656. doi:10.1002/j.1538-7305.1948.tb01338.x
- Stake, R. E. (2000). Case studies. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 435–454). Thousand Oaks, CA: Sage Publications.
- Stephens, K. K., Barrett, A. K., & Mahometa, M. J. (2013). Organizational communication in emergencies: Using multiple channels and sources to combat noise and capture attention. *Human Communication Research*, 39(2), 230–251. doi:10.1111/hcre.12002
- Warner, J. (2001). *Information, knowledge, text*. Lanham, MD: The Scarecrow Press, Inc.
- Watt, J. H., & Krull, R. (1974). An information theory measure for television programming. *Communication Research*, 1(1), 44–69. doi:10.1177/009365027400100103
- Weaver, C. H., & Weaver, G. L. (1965). Information theory and the measurement of meaning. *Speech Monographs*, 32(4), 435–447. doi:10.1080/03637756509375467
- Young, P. (1987). *The nature of information*. Praeger.
- Ziemer, R. E., & Tranter, W. H. (2002). *Principles of communications: Systems, modulation, and noise*. John Wiley & Sons, Inc.

ADDITIONAL READING

- Ackerman, M. S. (2000). The intellectual challenge of CSCW: The gap between social requirements and technical feasibility. *Human-Computer Interaction*, 15(2), 179–203. doi:10.1207/S15327051HCI1523_5
- Ackermann, F. (1994). Working together through groupware. In P. Lloyd (Ed.), *Groupware in the 21st century: Computer supported cooperative working toward the millennium* (pp. 162–168). Westport, CT: Praeger.
- Andriessen, J. H. E. (2012). *Working with groupware: Understanding and evaluating collaboration technology*. London: Springer.
- Baltes, B. B., Dickson, M. W., Sherman, M. P., Bauer, C. C., & LaGanke, J. S. (2002). Computer-mediated communication and group decision making: A meta-analysis. *Organizational Behavior and Human Decision Processes*, 87(1), 156–179. doi:10.1006/obhd.2001.2961
- Barkhi, R. (2005). Information exchange and induced cooperation in GDSS. *Communication Research*, 32(5), 646–678. doi:10.1177/0093650205279352

- Barnes, S., & Greller, L. M. (1994). Computer-mediated communication in the organization. *Communication Education*, 43(2), 129–142. doi:10.1080/03634529409378970
- Briggs, R., & Nunamaker, F. (1994). Getting a grip on groupware. In P. Lloyd (Ed.), *Groupware in the 21st century: Computer supported cooperative working toward the millennium* (pp. 61–72). Westport, Connecticut: Praeger.
- Burnett, A. (1994). Computer assisted creativity. In P. Lloyd (Ed.), *Groupware in the 21st century: Computer supported cooperative working toward the millennium* (pp. 197–205). Westport, CT: Praeger.
- Chang, W. L., & Lo, Y. P. (2012). A social network based group decision support system. *International Journal of Mobile Communications*, 10(1), 41–64. doi:10.1504/IJMC.2012.044522
- Contractor, N. S., & Seibold, D. R. (1993). Theoretical frameworks for the study of structuring processes in group decision support systems: Adaptive structuration theory and self-organizing systems theory. *Human Communication Research*, 19(4), 528–563. doi:10.1111/j.1468-2958.1993.tb00312.x
- Cover, T. M., & Thomas, J. A. (2006). *Elements of information theory*. Hoboken, NJ: John Wiley & Sons.
- Creighton, J., & Adams, J. (1998). *CyberMeeting: How to link people and technology in your organization*. New York: American Management Association.
- DeFleur, M. L., & Larsen, O. N. (1987). *The flow of information: An experiment in mass communication*. New Brunswick: Transaction Books.
- Dennis, A. R., Nunamaker, J. F., & Vogel, D. (1990). A comparison of laboratory and field research in the study of electronic meeting systems. *Journal of MIS*, 7, 107–135.
- DeSanctis, G. (2008). The Minnesota GDSS research project: Group support systems, group processes, and outcomes. *Journal of the Association for Information Systems*, 9(10), 551–608.
- DeSanctis, G. L. (1993). Shifting foundations in group support systems research. In L. M. Jessup & J. S. Valacich (Eds.), *Group support systems: New perspectives* (pp. 97–111). New York: Macmillan.
- Forgionne, G. A. (2003). An architecture for the integration of decision making support functionalities. In M. Mora, G. A. Forgionne, & J. N. Gupta (Eds.), *Decision making support systems: Achievements, trends and challenges for the new decade* (pp. 1–19). Hershey, PA: Idea Group Publishing. doi:10.4018/978-1-59140-045-5.ch001
- Hattori, F., Ohguro, T., Yokoo, M., Matsubara, S., & Yoshida, S. (1999). Socialware: Multiagent systems for supporting network communities. *Communications of the ACM*, 3(42), 55–61. doi:10.1145/295685.295701
- Heath, R. L., & Bryant, J. (2000). *Human communication theory and research: Concepts, contexts, and challenges*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Holtham, C. (1994). Groupware: Its past and future. In P. Lloyd (Ed.), *Groupware in the 21st century: Computer supported cooperative working toward the millennium* (pp. 3–14). Westport, CT: Praeger.
- Jackson, M. H., & Poole, M. S. (2003). Idea generation in naturally occurring contexts: Complex appropriation of a simple group procedure. *Human Communication Research*, 29(4), 560–591.
- Jessup, L. M., Connolly, T., & Galegher, J. (1990). The effects of anonymity on GDSS group process with an idea-generating task. *Management Information Systems Quarterly*, 14(3), 313–321. doi:10.2307/248893

- Jessup, L. M., Connolly, T., & Tansik, D. A. (1990). Toward a theory of automated groupwork: The deindividuating effects of anonymity. *Small Group Research*, 21(3), 333–348. doi:10.1177/1046496490213003
- Jessup, L. M., & Tansik, D. A. (1991). Group problem solving in an automated environment: The effects of anonymity and proximity on group process and outcome with a group decision support system. *Decision Sciences*, 22(2), 266–279. doi:10.1111/j.1540-5915.1991.tb00346.x
- Johansen, R., Sibbet, D., Benson, S., Martin, A., Mittman, R., & Saffo, P. (1991). *Leading business teams: How teams can use technology and group process tools to enhance performance*. Reading, MA: Addison Wesley.
- Lam, S. S. K., & Schaubroeck, J. (2000). Improving group decisions by better pooling information: A comparative advantage of group decision support systems. *The Journal of Applied Psychology*, 85(4), 565–573. doi:10.1037/0021-9010.85.4.565 PMID:10948801
- Liff, A. (1998). Fostering online collaboration and community. *Association Management*, 50(9), 33–38, 92.
- Limayem, M., Banerjee, P., & Ma, L. (2006). Impact of GDSS: Opening the black box. *Decision Support Systems*, 42(2), 945–957. doi:10.1016/j.dss.2005.08.004
- Lucas, H. C. (2000). *Information technology for management*. Boston: Irwin McGraw-Hill.
- McLeod, P. L. (1999). A literary examination of electronic meeting system use in everyday organizational life. *The Journal of Applied Behavioral Science*, 35(2), 188–206. doi:10.1177/0021886399352004
- Munkvold, B. E. (2003). *Implementing collaboration technologies in industry: Case examples and lessons learned*. London: Springer. doi:10.1007/978-1-4471-0073-7
- Nikoi, E. (2014). *Collaborative communication processes and decision making in organizations*. Hershey, PA: IGI Global. doi:10.4018/978-1-4666-4478-6
- Orlikowski, W. J., & Hofman, J. D. (2003). An improvisational model for change management: The case of groupware technologies. In T. W. Malone, R. Laubacher, & M. S. Morton (Eds.), *Inventing the organizations of the 21st century* (pp. 265–281). Cambridge, MA: The MIT Press.
- Pierce, J. R. (1980). *An introduction to information theory: Symbols, signals & noise*. NY, NY: Dover Publications, Inc.
- Poole, M. S., Holmes, M., Watson, R., & DeSanctis, G. (1993). Group decision support systems and group communication: A comparison of decision making in computer-supported and nonsupported groups. *Communication Research*, 20(2), 176–213. doi:10.1177/009365093020002002
- Postmes, T., & Lea, M. (2000). Social processes and group decision making: Anonymity in group decision support systems. *Ergonomics*, 43(8), 1252–1274. doi:10.1080/00140130050084978 PMID:10975183
- Recio-Garcia, J. A., Quijano, L., & Diaz-Agudo, B. (2013). Including social factors in an argumentative model for GDSS. *Decision Support Systems*, 56, 48–55. doi:10.1016/j.dss.2013.05.007
- Reinig, B. A., & Shin, B. (2002). The dynamic effects of group support systems on group meetings. *Journal of Management Information Systems*, 19(2), 303–325.
- Rogers, E. M., & Valente, T. W. (1993). A history of information theory in communication research. In J. R. Schement & B. D. Ruben (Eds.), *Between Communication and Information* (pp. 35–56). New Brunswick, NJ: Transaction Publishers.
- Severin, W. J., & Tankard, J. W. (2001). *Communication theories: Origins, methods, and uses in the mass media*. NY, NY: Longman.

Shannon, C. E., & Weaver, W. (1998). *A mathematical theory of communication*. Chicago: University of Illinois Press.

Sosik, J. J., Avolio, B. J., Kahai, S. S., & Jung, D. I. (1998). Computer-supported work group potency and effectiveness: The role of transformational leadership, anonymity, and task interdependence. *Computers in Human Behavior*, 14(3), 491–511. doi:10.1016/S0747-5632(98)00019-3

Turban, R. (1988). *Decision support and expert systems: Managerial perspectives*. NY, NY: Macmillan Publishing Company.

Valacich, J. S., Dennis, A. R., & Nunamaker, J. F. Jr. (1991). Electronic meeting support: The group systems concept. *International Journal of Man-Machine Studies*, 34(2), 261–282. doi:10.1016/0020-7373(91)90044-8

Verdu, S. (1998). Fifty years of Shannon theory. *IEEE Transactions on Information Theory*, 44(6), 2057–2078. doi:10.1109/18.720531

KEY TERMS AND DEFINITIONS



Communication: Human interaction with others to deliver information.

Computer-Mediated Communication: Human interaction using technological devices.

Entropy: The many different ways a message can be constructed depending on the circumstance.

Group Decision Support Systems: Technology that facilitates people working interactively and dynamically with collective group data.

Information Theory: A theory that provides the means to measure the efficiency of communication systems.

Noise: Interference that occurs in the transmission of a message.

Redundancy: Conveying the same information more than once in a message so the interpretation of the message is clear.

Evolutionary Algorithms for Global Decision Tree Induction

Marek Kretowski

Bialystok University of Technology, Poland

Marcin Czajkowski

Bialystok University of Technology, Poland

INTRODUCTION

Data mining (Fayyad, Piatetsky-Shapiro, Smyth & Uthurusamy, 1996) can reveal important and insightful information hidden in data. However, appropriate tools and algorithms are required to effectively identify correlations and patterns within the data. Decision trees (Kotsiantis, 2013) represent one of the main techniques for discriminant analysis prediction in knowledge discovery. The success of tree-based approaches can be explained by their ease of application, fast operation, and effectiveness. Furthermore, the hierarchical tree structure, in which appropriate tests from consecutive nodes are sequentially applied, closely resembles a human way of decision making. All this makes decision trees easy to understand, even for inexperienced analysts.

Despite 50 years of research on decision trees, many problems still remain (Loh, 2014) such as searching only for a locally optimal split in the internal nodes; appropriate pruning criterion, efficient analysis of cost-sensitive data or performing multi-objective optimization. To help resolve some of these problems, evolutionary algorithms (EAs) (Michalewicz, 1996) has been applied to decision tree induction (Barros et al., 2012). The strength of this approach lies in the global search for splits and predictions. It results in higher accuracy and smaller output trees compared to popular greedy decision tree inducers.

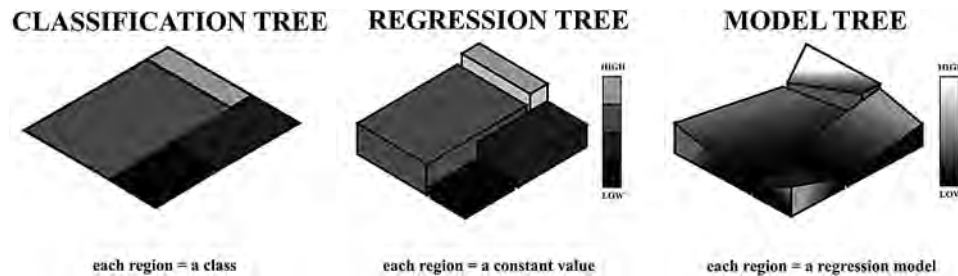
The purpose of this chapter is to illustrate the application of EAs to the problem of deci-

sion tree induction. The objectives are to show that evolutionary optimization, compared to the greedy search, may result in finding globally optimal solutions, whose complexity is significantly smaller and the prediction is highly competitive. We will cover the global induction of classification, regression, and model trees.

BACKGROUND

We may find different variants of decision trees in the literature (Loh, 2014). They can be grouped according to the type of problem they are applied to, the way they are induced, or the type of structure. Tree predictors can be used to classify existing data (classification trees) or to approximate real-valued functions (regression trees) (see Figure 1). In each leaf, classification tree assigns a class label (usually the majority class of all instances that reach that particular leaf), while the regression tree holds a constant value (usually an average value for the target attribute). A model tree can be seen as an extension of the typical regression tree (Quinlan, 1992). The constant value in each leaf of the regression tree is replaced in the model tree by a linear (or nonlinear) regression function. To predict the target value, the new tested instance is followed down the tree from a root node to a leaf using its attribute values to make routing decisions at each internal node. Next, the predicted value for the new instance is evaluated based on a regression model in the leaf.

Figure 1. An illustration of predicted values of the classification, regression, and model trees



Examples of predicted values of classification, regression, and model trees are given in Figure 1.

The gray level color of each region represents a different class label (for a classification tree), and the height corresponds to the value of the prediction function (regression and model trees).

Inducing optimal decision tree is known to be NP-complete (Naumov, 1991). Consequently, practical decision-tree learning algorithms are based on heuristics such as greedy algorithms where locally optimal splits are made in each node. The most popular tree-induction is based on the top-down approach (Rokach & Maimon, 2005). Top-down induction starts from the root node, where locally best split (test) is searched, according to the given optimality measure (e.g. Gini, towing or entropy rule for classification tree and least squared or least absolute deviation error criterion for regression tree). Next, the training instances are redirected to the newly created nodes and this process is repeated for each node until some stopping-rule is satisfied. The recursive partitioning of the dataset may lead to the data over-fitting, therefore, the decision tree pruning (Esposito, Malerba & Semeraro, 1997) is applied to improve the generalization power of the predictor. For an alternative approaches to greedy decision tree induction, like e.g. bottom-up, please refer to framework proposed by Barros et al. (2014).

Most of tree inducing algorithms partition the feature space with axis decision borders (Sheth & Deshpande, 2015). These types of trees are called univariate decision trees. Split at each non-terminal node usually involves single feature. For a continuous-valued feature usually an inequality

test with binary outcomes is applied and for a nominal attribute mutually exclusive groups of attribute values are associated with outcomes. One of the first and most well-known solution that can be applied to classification and regression problem is CART (Breiman, Friedman, Olshen & Stone, 1984) system. Good representatives of univariate inducers are also systems developed by Quinlan: C4.5 (1993) for classification and M5 (1992) for regression.

When more than one feature is taken into account to build a test in non-terminal node, then we deal with multivariate decision trees. The most common form of such a test is an oblique split, which is based on a linear combination of features (hyper-plane). The decision tree which applies only oblique tests is often called oblique or linear, whereas heterogeneous trees with univariate, linear and other multivariate (e.g., instance-based) tests can be called mixed decision trees (Llora & Wilson, 2004). It should be emphasized that computational complexity of the multivariate induction is generally significantly higher than the univariate induction. OC1 (Murthy, Kasif & Salzberg, 1994) is a good examples of multivariate decision tree system.

Inducing the decision tree with greedy strategy usually leads to suboptimal solutions. They search only for locally best splits at each node which does not guarantee the globally best solution. One of the alternatives is the ensemble of trees (Seni & Elder, 2010), which is created by the induction of different trees from the training sample. Ensemble classifiers like Random Forests (Breiman, 2001) induce many decision trees whose predictions

are combined to make the overall prediction for the forest (a collection of trees). Multiple models improve predictive performance, however, single-tree comprehensibility is lost. Different approaches propose look-ahead algorithms like APDT (Shah & Sastry, 1999) for classification and LLRT (Vogel, Asparouhov & Scheffer, 2007) for regression.

Evolutionary computing techniques are proven to be effective at escaping local optima and are able to successfully solve a general class of difficult computational problems. The global induction is mainly represented by systems based on an evolutionary approach (Barros et al., 2012; Barros, Carvalho & Freitas, 2015), however, in the literature there are some attempts of e.g. ant colony optimization algorithms (Boryczka & Kozak, 2010; Otero et al., 2012).

The global approach to decision tree induction was initially investigated in genetic programming (GP) community. The tree-based representation of solutions in a population makes this approach especially well-suited and easy for adaptation to decision tree generation. For the classification trees, the first attempt was made by Koza (1991), where author presented GP-method for evolving LISP S-expressions corresponding to simple decision trees. Next, univariate trees were evolved by Nikolaev and Slavov (1998) and Tanigawa and Zhao (2000), whereas Bot and Langdon (2000) proposed a method for induction of classification trees with limited oblique splits. Among genetic approaches for univariate decision tree induction there is GATree system proposed by Papagelis and Kalles (2001) and GAIT developed by Fu, Golden, Lele, Raghavan and Wasil (2003).

In the literature, there are relatively fewer evolutionary approaches for the regression and model trees than for the classification trees. In TARGET solution proposed by Fan & Gray (2005), authors propose to evolve a CART-like regression tree with simple genetic operators. Later solutions focus on evolving model trees with linear models: E-Motion (Barros, Duncan & Basgalupp, 2011) and non-linear models: GPMCC (Potgieter & Engelbrecht, 2008), in the leaves.

GLOBAL INDUCTION OF DECISION TREES

This section discuss the general concept of a standard evolutionary algorithm and its application for decision tree induction. It is a common algorithmic framework for both genetic algorithms (GAs) and genetic programming (GP) and can be applied to build a decision tree classifier and regression tree predictor.

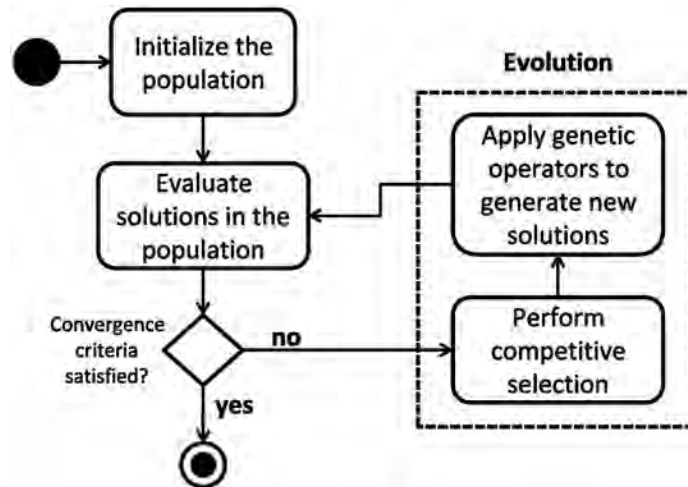
Evolutionary Algorithms

Evolutionary algorithms (Michalewicz, 1996) belong to a family of metaheuristic methods which represent techniques for solving a wide variety of difficult optimization problems. They provide a general framework (see Figure 2) which is inspired by biological mechanisms of evolution. The algorithm operates on individuals which compose a current population. Each individual represents a candidate solution to the target problem. Individuals are assessed using a measure named the fitness function which measures their performance and those with higher fitness have usually higher probability of being selected for reproduction. Genetic operators such as mutation and crossover influence new generations of individuals, producing new offspring. This guided random search (offspring usually inherits some traits from its ancestors) is stopped when some convergence criteria is satisfied. Evolutionary algorithms can also be combined with local search techniques. This synergy, known as Memetic Algorithms (Chen, Ong, Lim & Tan, 2011) allows speeding up and focusing the evolutionary search.

Evolutionary Induction of Decision Tree

Evolutionary algorithms which are designed to solve difficult optimization problems can be successfully applied to the problem of decision tree induction. There are, however; a few factors that need to be consider in order to exploit the full potential of evolutionary induction.

Figure 2. A general framework of evolutionary algorithm



Representation and Initialization of Individuals

The type of EA may be identified by the way the individuals in the populations are represented. A genetic algorithm is typically considered when solutions are encoded in a fixed-length linear string. The tree-encoding schemes usually imply genetic programming (GP), where the solution encodes data and functions (Woodward, 2003), however; the border between different types of EAs is vague and debatable.

When the individual is a univariate tree, each test in a non-terminal node concerns only one attribute (nominal or continuous valued). In a case of a continuous-valued attribute typical inequality tests are applied and only precalculated candidate thresholds (Fayyad & Irani, 1992) are considered as potential splits. A candidate threshold for the given attribute is defined as a midpoint between such a successive pair of examples in the sequence sorted by the increasing value of the attribute, in which the examples are characterized by different classes (or predicted values in case of regression trees). Although this solution works better with classification problem, it still may limit the number of possible splits and focuses the search process for regression trees. For a nominal attribute at least one value is associated with each branch. It means

that an inner disjunction is built into the induction algorithm. Individuals can also be represented by the oblique decision trees with multivariate tests or mixed decision trees (Kretowski & Grześ, 2007).

In general, initial population should be generated randomly to provide enough diversity of individuals. Due to the large search space, seeding the initial population with one of greedy solutions, discussed in the chapter, can enhance the quality of the search and shorten the execution time. To prevent the EA from being trapped in local optima, initial individuals can be induced on a random subsample of the training data and the potential tests in internal nodes can be searched on a random subset of attributes.

Genetic Operators

Genetic operators are the two main forces that form the basis of evolutionary systems and provide a necessary diversity and novelty. Tree-based representation requires developing specialized operators corresponding to the classical mutation and cross-over. Application of the operators can modify the tree structure, tests in internal nodes and models in the leaves (in case of the model trees).

Crossover is inspired by the sexual reproduction of living organisms. The crossover operator attempts to combine elements of two existing in-

dividuals (parents) in order to create a new (novel) solution but with some features of its “parents”. Each crossover starts with selecting the positions (nodes) in two affected individuals. In the most straightforward variant, the subtrees starting in the selected nodes are exchanged, which corresponds to the classical crossover from genetic programming (Koza, 1991). When the non-terminal nodes are chosen and the number of outcomes is equal, tests or branches associated with randomly chosen nodes can also be exchanged.

Mutation operator, also inspired by the biological process of the same, makes random changes in some places of selected individuals. Possibilities of mutation operator depend on the node type (leaf or internal node). For a non-terminal node a few variants exist:

- Existing test can be modified or a new test of the same or different type can be selected,
- Oblique test can be simplified or an axis-parallel test can be transformed into oblique one,
- Node can be transformed (pruned) into a leaf,
- One sub-tree can be replaced by another sub-tree from the same node.

Leaf modification makes sense only if it contains instances from different classes (or different predictive values). The leaf is transformed into an internal node and a new test is randomly chosen. More mutation variants for the leaf can be applied to the model trees. Linear model in the leaf can be extended, simplified or changed by adding, removing or replacing randomly chosen attributes.

Additional local search components can be built into the mutation-like operator. Memetic extension can be used to search for the split in the internal node, for example to find locally optimal test on given attribute or shift the splitting threshold at continuous-valued feature. While performing mutation in the leaves of the model

trees, the multivariate linear regression models can also be locally optimized (Czajkowski & Kretowski, 2012).

Fitness Function

Selection acts as a force which increases the quality of the population. For this reason, mechanism of selection usually requires a quantitative measure to assess individuals in the population. Such an objective function is called a fitness function and in practice it is a very important and sensitive element of the evolutionary system. It drives the evolutionary search process by measuring how good a single individual is in terms of meeting the problem objective.

In context of decision trees a direct minimization of the prediction performance measured on the learning set usually leads to the over-fitting problem and poor performance on unseen, testing observations, because the trees are overgrown. Therefore, we need to balance the reclassification quality (classification tree) or predictive error (regression tree) and the complexity of the tree.

For the classification trees an idea that is similar to the one used in cost complexity pruning in the CART system (Breiman et al., 1984) can be applied. The fitness function that is maximized can have the following form:

$$Fitness(T) = Pred(T) - \alpha * (Comp(T) - 1.0),$$

where $Pred(T)$ is usually equal the reclassification quality (Kretowski & Grześ, 2005) of the tree T and α is the relative importance of the predictor complexity. In the simplest form the tree complexity $Comp(T)$ can be defined as the tree size which is usually considered to be the number of nodes. The penalty associated with the classifier complexity increases proportionally with the tree size and prevents over-fitting. Subtracting the value 1.0 eliminates the penalty when the tree is composed of only one leaf. This general form of

the fitness function can be adjusted to different requirements. The term $Comp(T)$ can be modified to reflect not only the number of nodes but also the complexity of the tests.

In case of regression and model trees, the statistical model-selection criteria that are based on likelihood (e.g. AIC proposed by Akaike (1974) or BIC proposed by Schwarz (1978)) works well as fitness functions (Fan & Gray, 2005). These measures of goodness of fit penalize the model complexity (size of the tree). There are some resemblances in general formula of information criteria with the one from CART system as both construct a weight formula. The $Pred(T)$ is based on the maximum of the likelihood function and the term $Comp(T)$ can also be viewed as a penalty for overparametrization. For both information criteria, the penalty term depends not only on the predictor size but also on the number of observations. In case of multivariate trees, the penalty term may also be adjusted to take into account the number of attributes in the internal nodes and the size of the regression model in the leaves (for the model trees).

In addition, many fields like for example medicine or finance need to take into account, besides prediction error, the costs of features (due to costly measurements) or misprediction-cost (due to costs of wrong decision). The term cost-sensitive is used in the field (Turney, 1995) for the classification problem and the recent survey of cost-sensitive decision tree induction algorithms is available (Lomax & Vadera, 2013). The modification of the fitness function allows selecting features in accordance to their cost and optimizing the misprediction cost instead of the prediction error.

SOLUTIONS AND RECOMMENDATIONS

This section briefly recalls the Global Decision Tree (GDT) system whose general structure fol-

lows a typical EA framework (Michalewicz, 1996) with an unstructured population and a generational selection. Its framework can be used for the evolutionary induction of all types of trees: classification, regression and model (Kretowski & Grześ, 2007, Czajkowski & Kretowski, 2014). Among a family of tree inducers with homogeneous tree representations (univariate, multivariate, regression and model), the GDT system is capable of inducing heterogeneous trees (Czajkowski & Kretowski, in press). Such mixed decision tree can self-adapt its nodes representation to the currently analyzed data during the evolutionary induction.

The GDT solution concept can be applied in many real-life applications, such as finance (Czajkowski & Kretowski, 2015) and medicine (Grześ & Kretowski, 2007). The cost-sensitive approach of GDT is also available as well as various multi-objective optimization strategies that base on a weight formula and lexicographic analysis. Finally, to improve the performance of GDT system a hybrid MPI+OpenMP parallel implementation for classification (Czajkowski, Jurczuk & Kretowski, 2015) and regression/model trees (Czajkowski & Kretowski, 2016) was also proposed.

FUTURE RESEARCH DIRECTIONS

There are a number of promising directions for future research. In particular, there is always a strong motivation for speeding the evolutionary induction of decision trees up. Application of new parallel frameworks like CUDA (Strnad & Nerat, 2016) or Hadoop (White, 2009) is especially important in the context of modern data mining applications, where huge learning sets and the big data need to be analyzed. Another possibility of speeding up and focusing evolutionary search is application of the self-adaptive parameters in EA. Researchers should start to establish the relationship between EAs' parameter settings and dataset features.

Different multi-objective optimization strategies should also be considered like the Pareto-dominance approach. Finally, there are many challenges with application global decision trees for spatio-temporal, streaming datasets or for survivor and risk analysis.

CONCLUSION

In this chapter, the global induction of classification and regression trees is presented. Specialized genetic operators and memetic extensions impact the structure of the predictor, tests in internal nodes and, in case of model trees: regression planes in the leaves. Suitable defined fitness function allows focusing the search process and efficiently generating competitive predictors.

Every approach has drawbacks and limitations, however, in this particular case these drawbacks are well known and easily understood. Firstly, globally induced decision trees are obviously slower than greedy counterparts. Secondly, the one, final solution is not guaranteed as the evolution process will tend to run on indefinitely (if the predefined ending condition is not specified). On the other side, because the evolutionary process is progressive, intermediate answers can be harvested at practically any time. The most important is, however, the fact, that globally induced decision trees are able to avoid local optima, evolved trees are significantly smaller, less complex and usually generate better predictive model, also in terms of predictive accuracy, than the greedy counterparts.

REFERENCES

- Barros, R. C., Basgalupp, M. P., Carvalho, A. C., & Freitas, A. A. (2012). A survey of evolutionary algorithms for decision-tree induction. *IEEE Transactions on Systems, Man and Cybernetics. Part C, Applications and Reviews*, 42(3), 291–312. doi:10.1109/TSMCC.2011.2157494
- Barros, R. C., Carvalho, A. C., & Freitas, A. A. (2015). *Automatic Design of Decision-Tree Induction Algorithms*. Springer International Publishing. doi:10.1007/978-3-319-14231-9
- Barros, R. C., Jaskowiak, P. A., & Carvalho, A. C. (2014). A framework for bottom-up induction of oblique decision trees. *Neurocomputing*, 135, 3–12. doi:10.1016/j.neucom.2013.01.067
- Barros, R. C., Ruiz, D. D., & Basgalupp, M. (2011). Evolutionary model trees for handling continuous classes in machine learning. *Information Sciences*, 181(5), 954–971. doi:10.1016/j.ins.2010.11.010
- Boryczka, A., & Kozak, J. (2010). Ant colony decision trees – a new method for constructing decision trees based on ant colony optimization. *Lecture Notes in Artificial Intelligence*, 6421, 373–382.
- Bot, M., & Langdon, W. (2000). Application of genetic programming to induction of linear classification trees. In R. Poli et al. (Eds.), *Lecture Notes in Computer Science: Vol. 1802. EuroGP 2000* (pp. 247–258). doi:10.1007/978-3-540-46239-2_18
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. doi:10.1023/A:1010933404324
- Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). *Classification and regression trees*. Wadsworth International Group.
- Chen, X. S., Ong, Y. S., Lim, M. H., & Tan, K. C. (2011). A multi-facet survey on memetic computation. *IEEE Transactions on Evolutionary Computation*, 15(5), 591–607. doi:10.1109/TEVC.2011.2132725
- Czajkowski, M., Jurczuk, K., & Kretowski, M. (2015). A Parallel Approach for Evolutionary Induced Decision Trees. MPI+OpenMP Implementation. *Lecture Notes in Artificial Intelligence*, 9119, 340–349.

- Czajkowski, M., Jurczuk, K., & Kretowski, M. (2016). Hybrid Parallelization of Evolutionary Model Tree Induction. *Lecture Notes in Artificial Intelligence*, 9692, 370–379.
- Czajkowski, M., & Kretowski, M. (2012). Does memetic approach improve global induction of regression and model trees? In L. Rutkowski et al. (Eds.), *Lecture Notes in Computer Science: Vol. 7269. EC 2012* (pp. 174–181). doi:10.1007/978-3-642-29353-5_20
- Czajkowski, M., & Kretowski, M. (2014). Evolutionary Induction of Global Model Trees with Specialized Operators and Memetic Extensions. *Information Sciences*, 288, 153–173. doi:10.1016/j.ins.2014.07.051
- Czajkowski, M., & Kretowski, M. (2015). Cost-sensitive global model trees applied in loan charge-off forecasting. *Decision Support Systems*, 74, 57–66. doi:10.1016/j.dss.2015.03.009
- Czajkowski, M., & Kretowski, M. (in press). The Role of Decision Tree Representation in Regression Problems - an Evolutionary Perspective. *Applied Soft Computing*.
- Esposito, F., Malerba, D., Semeraro, G., & Kay, J. (1997). A comparative analysis of methods for pruning decision trees. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(5), 476–491. doi:10.1109/34.589207
- Fan, G., & Gray, J. B. (2005). Regression tree analysis using TARGET. *Journal of Computational and Graphical Statistics*, 14(1), 206–218. doi:10.1198/106186005X37210
- Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., & Uthurusamy, R. (1996). *Advances in knowledge discovery and data mining*. AAAI Press.
- Fayyad, U. M., & Irani, K. B. (1992). On the handling of continuous-valued attributes in decision tree generation. *Machine Learning*, 8(1), 87–102. doi:10.1007/BF00994007
- Fu, Z., Golden, B., Lele, S., Raghavan, S., & Wasil, E. (2003). A genetic algorithm based approach for building accurate decision trees. *INFORMS Journal on Computing*, 15(1), 3–22. doi:10.1287/ijoc.15.1.3.15152
- Grześ, M., & Kretowski, M. (2007). Decision tree approach to microarray data analysis. *Biocybernetics and Biomedical Engineering*, 27(3), 29–42.
- Kotsiantis, S. B. (2013). Decision trees: A recent overview. *Artificial Intelligence Review*, 39(4), 261–283. doi:10.1007/s10462-011-9272-4
- Koza, J. (1991). Concept formation and decision tree induction using genetic programming paradigm. In H. Schwefel et al. (Eds.), *Lecture Notes in Computer Science: Vol. 496. PPSN* (pp. 124–128). doi:10.1007/BFb0029742
- Kretowski, M., & Grześ, M. (2007). Evolutionary induction of mixed decision trees. *International Journal of Data Warehousing and Mining*, 3(4), 68–82. doi:10.4018/jdwm.2007100104
- Llora, X., & Wilson, S. (2004). Mixed decision trees: Minimizing knowledge representation bias in LCS. In K. Deb et al. (Eds.), *Lecture Notes in Computer Science: Vol. 3103. GECCO 2004* (pp. 797–809). doi:10.1007/978-3-540-24855-2_94
- Loh, W. (2014). Fifty years of classification and regression trees. *International Statistical Review*, 83(3), 329–348. doi:10.1111/insr.12016
- Lomax, S., & Vadera, S. (2013). A survey of cost-sensitive decision tree induction algorithms. *ACM Computing Surveys*, 45(2), 16, 1–16.
- Michalewicz, Z. (1996). *Genetic algorithms + data structures = evolution programs*. Springer. doi:10.1007/978-3-662-03315-9
- Murthy, S., Kasif, S., & Salzberg, S. (1994). A system for induction of oblique decision trees. *Journal of Artificial Intelligence Research*, 2, 1–33.

- Naumov, G. E. (1991). NP-completeness of problems of construction of optimal decision trees. *Soviet Physics, Doklady*, 36(4), 270–271.
- Nikolaev, N., & Slavov, V. (1998). Inductive genetic programming with decision trees. *Intelligent Data Analysis*, 2(1-4), 31–44. doi:10.1016/S1088-467X(98)00005-5
- Otero, F. E. B., Freitas, A. A., & Johnson, C. G. (2012). Inducing decision trees with an ant colony optimization algorithm. *Applied Soft Computing*, 12(11), 3615–3626. doi:10.1016/j.asoc.2012.05.028
- Papagelis, A., & Kalles, D. (2001). Breeding decision trees using evolutionary techniques. *Proc. of ML*, 393–400.
- Quinlan, J. (1992). Learning with continuous classes. *Proc. of AI*, 343–348.
- Quinlan, J. (1993). *C4.5: Programs for machine learning*. Morgan Kaufmann.
- Rokach, L., & Maimon, O. (2005). Top-down induction of decision trees classifiers - a survey. *IEEE Transactions on Systems, Man and Cybernetics. Part C, Applications and Reviews*, 35(4), 476–487. doi:10.1109/TSMCC.2004.843247
- Seni, G., & Elder, J. (2010). *Ensemble methods in data mining: Improving accuracy through combining predictions*. San Rafael, CA: Morgan and Claypool.
- Shah, S., & Sastry, P. S. (1999). New algorithms for learning and pruning oblique decision trees. *IEEE Transactions on Systems, Man and Cybernetics. Part C, Applications and Reviews*, 29(4), 494–505. doi:10.1109/5326.798764
- Sheth, N. S., & Deshpande, A. R. (2015). A Review of Splitting Criteria for Decision Tree Induction. *Fuzzy Systems*, 7(1), 1–4.
- Strnad, D., & Nerat, A. (2016). Parallel construction of classification trees on a GPU. *Concurrency and Computation*, 28(5), 1417–1436. doi:10.1002/cpe.3660
- Tanigawa, T., & Zhao, Q. (2000). A study on efficient generation of decision trees using genetic programming. *Proc. of GECCO*, 1047–1052.
- Turney, P. (1995). Cost-sensitive classification: Empirical evaluation of a hybrid genetic decision tree induction algorithm. *Journal of Artificial Intelligence Research*, 2, 369–409.
- Vogel, D., Asparouhov, O., & Scheffer, T. (2007). Scalable look-ahead linear regression trees. *Proc. of ACM SIGKDD*, 757–764.
- White, T. (2009). *Hadoop: The definitive guide*. O'Reilly Media, Inc.
- Woodward, J. R. (2003). GA or GP? That is not the question. *Proc. of IEEE CEC*, 1056–1063. doi:10.1109/CEC.2003.1299785

ADDITIONAL READING

- Bacardit, J., & Llorca, X. (2013). Large-scale data mining using genetics-based machine learning. *WIREs Data Mining Knowledge Discovery*, 3(1), 37–61. doi:10.1002/widm.1078
- Eiben, A. E., & Smit, S. K. (2011). Parameter tuning for configuring and analyzing evolutionary algorithms. *Swarm and Evolutionary Computation*, 1(1), 19–31. doi:10.1016/j.swevo.2011.02.001
- Freitas, A. A. (2002). *Data Mining and Knowledge Discovery with Evolutionary Algorithms*. Springer-Verlag New York. doi:10.1007/978-3-662-04923-5
- Koziol, M., & Wozniak, M. (2009). Multivariate Decision Trees vs. Univariate Ones. *Advances in Intelligent and Soft Computing*, 57, 275–284. doi:10.1007/978-3-540-93905-4_33
- Malerba, D., Esposito, F., Ceci, M., & Appice, A. (2004). Top-down induction of model trees with regression and splitting nodes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(5), 612–625. doi:10.1109/TPAMI.2004.1273937 PMID:15460282

Mukhopadhyay, A., Maulik, U., Bandyopadhyay, S., & Coello, C. A. C. (2014). A Survey of Multiobjective Evolutionary Algorithms for Data Mining: Part I. *IEEE Transactions on Evolutionary Computation*, 18(1), 4–19. doi:10.1109/TEVC.2013.2290086

Nasridonov, A., Lee, Y., & Park, Y. H. (2014). Decision tree construction on GPU: Ubiquitous parallel computing approach. *Computing*, 96(5), 403–413. doi:10.1007/s00607-013-0343-z

Zhao, H., Sinha, A. P., & Bansal, G. (2011). An extended tuning method for cost-sensitive regression and forecasting. *Decision Support Systems*, 51(3), 372–383. doi:10.1016/j.dss.2011.01.003

KEY TERMS AND DEFINITIONS

Classification Tree: A decision tree applied to the classification problem, where each leaf node is labeled with one class label. Classification tree is used to predict membership of cases or instances in the classes of a categorical dependent variable from their measurements on one or more predictor variables.

Decision Tree: A decision tree is a graph that uses a branching method to illustrate every possible outcome of a decision. Each branch of the decision tree represents a possible decision or occurrence. The tree structure shows how one choice leads to the next, and the use of branches indicates that each option is mutually exclusive.

Decision Tree Pruning: An approach that reduces the size of decision tree by removing parts of the tree that provide little power to classify/predict the data. This technique improves the generalization power of the decision tree and reduces the over-fitting.

Global Induction: A method of decision tree generation, where both the tree structure and all tests are searched at the same time; usually based on evolutionary approach in contrast to top-down induction.

Mixed Decision Tree: A decision tree with heterogeneous representation. The tests in internal nodes may be univariate or multivariate whereas the leaves may hold a constant value or a regression model.

Model Tree: An extension of regression trees where the constant value in the terminal node is replaced by the regression plane.

Multivariate Decision Tree: A decision tree with tests involving several attributes. The most common form of such tests is an oblique split, which is based on a linear combination of features (hyper-plane).

Regression Tree: A decision tree applied to the regression problem, where the dependent variable is continuous. Each leaf of the tree usually holds a constant value (usually an average for the target attribute of all instances that reach this particular leaf).

Top-Down Induction: A recursive method of decision tree generation. It starts with the entire input dataset in the root node where a locally optimal test for data splitting is searched and branches corresponding to the test outcomes are created. The test searches and data splitting are repeated in the created nodes unless the stopping condition is met.

Univariate Decision Tree: A decision tree which partitions the feature space with axis-parallel hyper-planes. Each split at non-terminal node involves single feature.

A Family Review of Parameter–Learning Models and Algorithms for Making Actionable Decisions

Chun-Kit Ngan

The Pennsylvania State University, USA

INTRODUCTION

Learning optimal decision parameters over time series (Bisgaard & Kulahci, 2011; Chatfield, 2001) to make actionable recommendations is an important topic which has gained significant interest in the past decade. In financial markets, stock investors analyze economic indicators to predict the occurrence of a bear market bottom or a bull market top to determine when they should buy or sell their stocks in order to gain the maximal profit. In medical diagnosis, physicians measure diabetes patients' blood-glucose-level (BGL) threshold to detect when the patients suffer hypo- and hyperglycemia. In energy industry, microgrid engineers investigate electricity peak demands to decide when the load shedding of electric power within the microgrid should be executed in order to minimize energy costs and maximize customers' savings while preserving the desired quality of service (QoS) in terms of power interruption. The goal of these time-series data analyses is to enable domain experts to understand the problems and lead them into a better decision-making and actionable recommendation.

Consider one particular example above that the stock investors are interested in determining when a bear market bottom will occur so that they can buy the stocks in the lowest price and then sell the stocks in the future to gain the maximal profits. To achieve this purpose, financial specialists suggest that a bear market bottom (Stack, 2009) has occurred if some changes of the financial indexes, such as the S&P 500 falls at least 20% and the Consumer Confidence Index declines at

least 30 points, are detected. Note that these decision parameters, e.g., 20% and 30 points, may be given by the domain experts or learned from the formal mathematical computations. To support such a decision-making event, it is important for domain experts to develop a data analytical methodology that is reliable and useful for their decision-making. The main challenge in such a methodology is how accurately, expressively, and effectively to model, analyze, and compute those time series data such that domain experts can analytically learn the optimal parameters to make a better decision. This is exactly the focus of this chapter.

BACKGROUND

To support the above decision-making events, there are two approaches, i.e., qualitative and quantitative, that data analysts often use to help domain experts learn optimal decision parameters. A qualitative analysis is a domain-knowledge-based approach that requires domain experts to base upon their past experience and observation to analyze scenarios, from which the experts determine the decision parameters. For example, in the stock market, financial specialists have identified a set of stock indicators that can be used to determine the bear market bottom. The indicators include the S&P 500 percentage decline (SPD), Coppock Guide (CG), Consumer Confidence Index drop (CCD), ISM Manufacturing Survey (ISM), and Negative Leadership Composite "Distribution" (NLCD). If these indicators satisfy the

DOI: 10.4018/978-1-5225-2255-3.ch186

pre-defined, parameterized conditions, e.g., $SPD < -20\%$, $CG < 0$, etc., it signals that the best period for the investors to buy the stocks, e.g., the S&P 500 Index Fund, is approaching. Often these parameters may reflect some realities since they are set by the domain experts based on their past experiences, observations, intuition, and domain knowledge. However, the suggested parameters may not always be accurate because those parameters are static, but the scenarios that the experts deal with are always dynamic in nature. Thus data analysts need to develop a class of mathematical models and algorithms that can be used to help the domain experts learn the decision parameters dynamically to fit the need of these scenarios.

A quantitative analysis is a machine-learning-based approach (Bell, 2014) that enables data analysts to apply mathematical models and algorithms to learn parameters on a large amount of datasets for decision making. For instance, the logistic regression models (Bierens, 2008; Cook, et. al., 2000; Dougherty, 2011; Hansen, 2014; Heij, et. al., 2004; Rawlings, et. al., 2012; SL, et. al., 2014) are often used to predict the occurrence of an event (0 or 1) by learning parametric coefficients of the logistic distribution function of the explanatory variables. This is done based upon the historical data by applying nonlinear regression models and the Maximum Likelihood Estimation (MLE) (Myung, 2003). However, using the machine-learning-based approach to learn parameters is computationally complicated and time consuming because each dataset has a large number of attributes (i.e., the curse of dimensionality) (Bellman, 2003 and 2015) that require the data analysts to take a long time to select the significant ones before they can do the parameter learning. Hence domain experts' knowledge is necessary to help the analysts identify and select the significant attributes to avoid the curse of dimensionality before learning the parameters.

Thus this chapter addresses the approach to close the above research gaps. More specifically, the chapter addresses the following two research questions:

1. How to develop the “time-point-based” mathematical models and learning algorithms that combine the strengths of both qualitative and quantitative approaches to learn optimal decision parameters to make actionable recommendations.
2. How to implement the experimental case studies that evaluates the performance of these developed hybrid-based parameter-learning models and algorithms.

HYBRID-BASED PARAMETER-LEARNING MODELS AND ALGORITHMS

To answer the above research questions, the authors have developed a family of parameter-learning models and algorithms, i.e., Expert Query Parametric Estimation (EQPE) models and Checkpoint algorithms (Ngan, et. al., 2010, 2011, 2012 & 2013), which combines the strength of both qualitative and quantitative methodologies to complement each other to learn optimal decision parameters in an efficient manner. This family of models and algorithms relies on domain expertise to select attributes and conditions against the data, from which the family of EQPE models and Checkpoint algorithms can learn decision parameters efficiently. This class of models and algorithms can be divided into two categories: Single-Event (SE) and Multi-Event (ME). The SE-EQPE model, using the Checkpoint algorithm, learns one decision parameter to detect the occurrence of a single event, e.g., the bear market bottom or the bull market top. The ME-EQPE model is the further extension of the SE-EQPE model, using the (Multidimensional) M- or (Relaxed) R-Checkpoint algorithm, to learn multiple decision parameters simultaneously to detect the occurrence of multiple events in sequence, e.g., the bear market bottom and then the bull market top. More specifically, the M-Checkpoint algorithm is a brute-force approach that learns multiple decision parameters optimally at their inter-related time points for multi-events.

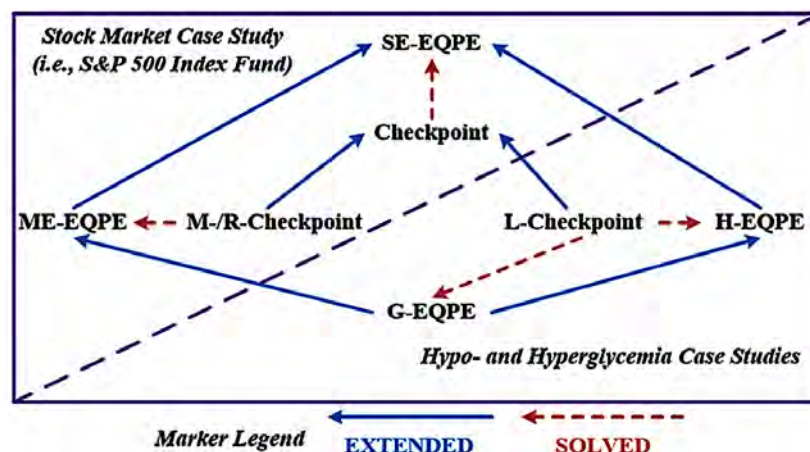
This approach should be used if the data analysts focus on learning the optimal decision parameters and do not need to concern about the high time complexity and computational cost. Oppositely, the R-Checkpoint algorithm is a heuristic methodology that learns multiple decision parameters, which are fairly close to the optimal parameters learned from the M-Checkpoint algorithm, and maintains a satisfactorily low time complexity. This approach should be adopted if the data analysts emphasize on learning the reasonably good parameters in a shorter time.

However, the family of the above EQPE models and Checkpoint algorithms has two important issues. First the models and the algorithms only solve the problems that can solely take the real-value time series, whereas some other specific problems may need the binary time series as the inputs to signal the occurrence of some events. Second the models and the algorithms only consider the conjunctive constraints but cannot solve the specific problems that require the considerations of disjunctive and biconditional constraints.

Consider one example again that the physicians measure diabetes patients' BGL threshold to detect when the patients suffer hypo- and hyperglycemia (American Diabetes Association, 2014 & 2015). First to determine the BGL detection threshold, the binary time series that signals the symptoms of hypo- and hyperglycemia is needed. Second

when the diabetes patients' BGLs are higher or lower than the given threshold, the patients may start suffering at least one or more symptoms of hypo- or hyperglycemia respectively but probably not all. These symptomatic conditions are indeed disjunctive binary constraints, which are used to evaluate whether or not the symptoms of hypo- or hyperglycemia occur in the diabetes patients. Third to make certain the hypo- or hyperglycemic symptoms occurred due to the monitoring constraints satisfied, i.e., the diabetes patients' BGLs lower or higher than the given threshold, and vice versa, biconditional constraints are needed. For instance, if a patient suffers from fatigue, dizziness, confusion, headaches, blurred vision, etc., it implies that the patient's BGL drops below or exceeds above the threshold not because of other illnesses. Similarly, if the patient's BGL becomes lower or higher than the threshold, it implies that he or she suffers from the hypo- or hyperglycemic symptoms. To solve this specific type of problems, the authors have further extended their previous SE model and algorithm to develop the (Hypoglycemic) H-EQPE model and the Linear L-Checkpoint algorithm (Ngan & Li, 2015) to learn one BGL threshold to detect the occurrence of hypoglycemia. Based on the H-EQPE and the ME-EQPE models, the authors further developed the (Glycemic) G-EQPE model to learn the two BGL thresholds to detect the occurrence of both

Figure 1. A family of parameter-learning models and algorithms: EQPEs and checkpoints



hypo- and hyperglycemic events in sequence. Figure 1 demonstrates the relationships among all the developed EQPE-based models and Checkpoint-oriented algorithms.

To show the effectiveness and the efficiency of the developed models and algorithms shown in Figure 1, the authors compare their methods with the domain-knowledge-based and the machine-learning-based approaches. As a proof of concept, the authors have conducted the two experimental case studies, i.e., the stock market and the glycemia, respectively. The authors show that their approaches are more effective and produce the results that are superior to those of the two other approaches mentioned above. More specifically, in their experiments, the authors find that their algorithms outperform the domain experts' recommendations and the logistic regression models.

The rest of the chapter is organized as follows. In the fourth section, the authors describe their SE-EQPE model and the Checkpoint algorithm with the "Stock Market Case Study" results in detail. Using the same case study as an example, the authors illustrate how to formulate the ME-EQPE problem and illustrate the performance of their *M-/R-Checkpoint* algorithms in the fifth section. In the sixth section, the authors demonstrate the formulation of the H-EQPE and G-EQPE models with the "Hypo- and Hyperglycemia Case Studies" results using the *L-Checkpoint* algorithm. In the seventh section, two prospective applications that are solved by the models and algorithms are described. The eighth and ninth section briefly summarizes the future work and concludes the chapter respectively.

SE-EQPE MODEL AND CHECKPOINT ALGORITHM

The goal of a SE-EQPE problem is to find optimal values of decision parameters that maximize an objective function over historical time series. For a SE-EQPE problem being constructed, the authors need to define a model for it. The authors assume

that the time domain $\mathbf{T}$ is represented by a set of natural numbers: $\mathbf{T} = \mathbf{N}$, and that the authors are also given a vector of n real-valued parameter variables $(p_1, p_2, \dots, p_n)$.

Single-Event Expert Query Parametric Estimation (SE-EQPE) Problem: A SE-EQPE problem is a tuple $\langle S_M, P_M, C_M, U \rangle$, where $S_M = \{S_1, S_2, \dots, S_k\}$ is a set of k parametric monitoring time series, $P_M = \{p_1, p_2, \dots, p_n\}$ is a set of n real-value monitoring parametric decision variables, C_M is a parametric monitoring constraint in S_M and P_M , and U is a time utility function.

Single-Event Expert Query Parametric Estimation (SE-EQPE) Solution: A solution to the SE-EQPE problem $\langle S_M, P_M, C_M, U \rangle$ is $\argmax O(p_1, p_2, \dots, p_n)$, i.e., the values of monitoring parameters, $p_1, p_2, \dots, p_n$, that maximize O , where O is the objective function corresponding to U .

To solve the SE-EQPE problem means to find the optimal values of decision parameters $p_1, p_2, \dots, p_n$ to maximize $O(p_1, p_2, \dots, p_n)$. The optimal values of decision parameters $p_1, p_2, \dots, p_n$ can be determined by using the Checkpoint algorithm.

The Checkpoint algorithm uses a KD tree data structure and a searching algorithm (Bentley, 1975 & 1979; Samet, 2006) to find the optimal decision parameters. The details of the algorithm can be found in (Ngan et. al., 2010 & 2011).

Theorem 1: For N parameter vectors in the dataset, the Checkpoint algorithm correctly computes a SE-EQPE solution, i.e., $\argmax O(p_1, p_2, \dots, p_k)$, where O is the objective function of the SE-EQPE problem, with the complexity $O(kN \log N)$.

Stock Market Case Study

The authors assume that the investor buys the S&P 500 index fund at the decision variable time t and sell it at the given t_s , which is the last day of the given training data set. The earning function $SP(t_s)/SP(t) - 1 \in \mathbf{R}$ is the utility, which is maximized by choosing the optimal value $t \in \mathbf{T}$, where $SP(t_s)$ and $SP(t)$ are the sell and buy value

Table 1. Time-series data

Time Series S	Abbreviation
Percentage decline in S&P 500	SPD(t)
Coppock Guide	CG(t)
Points drop in Consumer Confidence	CCD(t)
ISM Manufacturing Survey	ISM(t)
Number of consecutive days in Bear Market “DISTRIBUTIOIN” of Negative Leadership Composite	NLCD(t)

Table 2. SE-EQPE problem and solution formulation for the S&P 500 index fund

Problem: $\langle S_M, P_M, C_M, U \rangle$, where $S_M = \{SPD, CG, CCD, ISM, NLCD\}$ $P_M = \{p_1, p_2, p_3, p_4, p_5\}$ $C_M = SPD(t) < p_1 \wedge CG(t) < p_2 \wedge CCD(t) < p_3 \wedge ISM(t) < p_4 \wedge NLCD(t) > p_5$ $U = SP(t_s)/SP(t) - 1$ Solution: $\operatorname{argmax} O(p_1, p_2, p_3, p_4, p_5) \triangleq U(t)$

of the S&P 500 index fund at the time t_s and t respectively. The SE-EQPE problem and solution for the stock market case study can be constructed by putting the considered time series, parameters, constraints, and functions to the definitions shown in Table 1 and 2.

The values of the optimal decision parameters can be determined by using the Checkpoint algorithm. Using the Checkpoint algorithm, the authors can obtain the optimal decision parameters (i.e., -29.02, -20.01, -26.61, 49, 70) and the maximal earning (i.e., 53.37) from the training dataset for the stock market case study. The time complexity of the MLE for the logistic regression model is $O(k^2N)$, where k is the number of decision parameters, and N is the size of the learning dataset. For the Checkpoint algorithm, the complexity is $O(kM \log N)$. Using the decision parameters from the financial expert (i.e., -20%, 0, -30, 45,

180 days), the logistic regression model, and the Checkpoint algorithm, the “Best Buy” opportunities in stock and their earnings are shown in Table 3. Note that the Checkpoint algorithm considerably outperforms both the financial expert’s criteria and the logistic regression model.

ME-EQPE MODEL AND M-/R-CHECKPOINT ALGORITHM

For solving the formulations which the utility is determined from the multiple time points, a ME-EQPE model is proposed. The model assists users in finding optimal values of decision parameter vectors that maximize an objective function and satisfy their corresponding parametric constraints over historical, multivariate time series. Using the “Stock Market Case Study” as an example,

Table 3. Investors’ earning of the S&P 500 index fund from the test data set¹

Decision Approach	Best Buy	S&P 500 Index	Earning%
Financial Expert’s Criteria	10/09/08	909.92	1.03
Logistic Regression Model	11/26/08	887.68	3.56
Checkpoint Algorithm	03/10/09	719.6	27.8

Table 4. Time-series data

Time Series S	Abbreviation
Percentage decline and increase in S&P 500	SPD(t), SPI(t)
Coppock Guide	CG(t)
Points drop and increase in Consumer Confidence	CCD(t), CCI(t)
ISM Manufacturing Survey	ISM(t)
Number of consecutive days in Bear Market “DISTRIBUTIOIN” of Negative Leadership Composite	NLCD(t)
Negative Leadership Composite “Selling Vacuum”	NLCSV(t)
Time Utility Earning at the time points t_1 and t_2 , i.e., the index fund is bought at t_1 and sold at t_2 of the learning data set	Earning(t_1, t_2)

the authors illustrate how to formulate the ME-EQPE model and illustrate the performance of their M -/ R -Checkpoint algorithms in this section. The detail of the model and both of the algorithms can be found in (Ngan, et. al., 2012 & 2013). The ME-EQPE problem and solution for the stock market case study can be constructed by putting the considered time sequence vectors, parameter vectors, constraints, and functions to the definitions shown in Table 4 and 5.

The values of the optimal decision parameters can be determined by using the learning algorithms, M - and R -Checkpoint algorithm. The time complexity of the MLE for the logistic regression model is $O(mk2^N)$, where m is the number of

events, k is the number of parameteric coefficients, and N is the size of the learning dataset. For the R -Checkpoint algorithm, the complexity is $O(QkN \log N)$, where Q is the total number of m -event, time-point combinations which yields the top- Q time utility U in the learning dataset. Using the logistic regression model and the R -Checkpoint algorithm, the “Best Buy” and “Best Sell” opportunities in this investment and their earnings are shown in Table 6. The results show that the earning obtained from the logistic regression methodology is 0.52% and from the R -Checkpoint is 8.68% that is almost 17 times higher than the earning obtained from the logistic regression approach. On the contrary, even if the

Table 5. ME-EQPE problem and solution formulation for the S&P 500 index fund

Problem: $\langle S_M, P_M, C_M, U \rangle$, where $S_M = \{ \vec{S}_1(t_1), \vec{S}_2(t_2) \}$, for $\vec{S}_1(t_1) = (\text{SPD}(t_1), \text{CG}(t_1), \text{CCD}(t_1), \text{ISM}(t_1), \text{NLCD}(t_1))$ and $\vec{S}_2(t_2) = (\text{SPI}(t_2), \text{CG}(t_2), \text{CCI}(t_2), \text{ISM}(t_2), \text{NLCSV}(t_2))$ $P_M = \{ \vec{P}_1, \vec{P}_2 \}$, for $\vec{P}_1 = (p_1, p_2, p_3, p_4, p_5)$ and $\vec{P}_2 = (q_1, q_2, q_3, q_4, q_5)$ $C_M = \{ C_{M_1}, C_{M_2} \}$, for $C_{M_1} = \text{SPD}(t_1) < p_1 \wedge \text{CG}(t_1) < p_2 \wedge \text{CCD}(t_1) < p_3 \wedge \text{ISM}(t_1) < p_4 \wedge \text{NLCD}(t_1) > p_5$ and $C_{M_2} = \text{SPI}(t_2) \geq q_1 \wedge \text{CG}(t_2) \geq q_2 \wedge \text{CCI}(t_2) \geq q_3 \wedge \text{ISM}(t_2) \geq q_4 \wedge \text{NLCSV}(t_2) > q_5$ $U = f_2(t_2) * f_1(t_1) - 1$, for $f_2(t_2) = \text{SP}(t_2)$ and $f_1(t_1) = \frac{1}{\text{SP}(t_1)}$
Solution: $\text{argmax } O(\vec{P}_1, \vec{P}_2) \triangleq U(t_1, t_2)$

Table 6. Investors' earning of the S&P 500 index fund from the test data set²

Decision Approach	Best Buy	S&P 500 Index	Best Sell	S&P 500 Index	Earning%
Logistic Regression Model	10/14/08	998.01	09/03/09	1003.24	0.52%
M-Checkpoint Algorithm	10/31/08	968.75	09/07/10	1091.84	12.71%
R-Checkpoint Algorithm	10/06/08	1056.89	09/24/10	1148.67	8.68%

* The financial experts do not provide the decision parameters that can be used to determine the "Best Buy" and "Best Sell" opportunity in the sequence of occurrence.

M-Checkpoint algorithm can guarantee the optimality and yield a better earning, i.e., 12.71%, from the testing data set, but the efficiency of the R-Checkpoint algorithm, i.e., $O(QkN \log N)$, is much better than that of the M-Checkpoint algorithm, i.e., $O(N^m)$.

H-/G-EQPE MODEL AND LINEAR L-CHECKPOINT ALGORITHM

In this section, the authors demonstrate the formulation of the H-EQPE and G-EQPE models with the "Hypo- and Hyperglycemia Case Studies" results using the L-Checkpoint algorithm. The L-Checkpoint algorithm uses a two-stage approach: (1) a simple linear search to eliminate all the time points that do not satisfy a set of all binary-value symptoms and (2) a KD tree data structure and a searching algorithm to find the optimal decision parameters.

H-EQPE Model

First the authors formulate the H-EQPE problem in the case study of detecting the symptomatic events of hypoglycemia that are shown in Table 7.

Here is a brief explanation of Table 7. The H-EQPE problem can be formed by: (1) S_M - a set of one parametric monitoring time series (i.e., BGL(t) for blood glucose level); (2) S_B - a set of five binary symptomatic time series (i.e., F(t) for fatigue, D(t) for dizziness, C(t) for confusion, H(t) for headaches, and BV(t) for blurred vision); (3) P_M - a set of one monitoring parametric decision variables (i.e., p_l - a threshold of blood glucose level); (4) P_B - a set of five binary-value symptomatic conditions (i.e., $p_1', p_2', p_3', p_4', p_5'$ with either true or false value for fatigue, dizziness, confusion, headaches, and blurred vision, respectively); (5) C_G - a biconditional global constraint meaning that a patient's blood glucose level at time point t is no higher than the threshold p_l if

Table 7. H-EQPE problem and solution formulation for detecting the symptomatic events of hypoglycemia

Problem: $\langle S_M, S_B, P_M, P_B, C_G, U \rangle$, where $S_M = \{BGL(t)\}$ $S_B = \{F(t), D(t), C(t), H(t), BV(t)\}$ $P_M = \{p_l\}$ $P_B = \{p_1', p_2', p_3', p_4', p_5'\}$ $C_G = BGL(t) \leq p_l \leftrightarrow (F(t) = p_1' \vee D(t) = p_2' \vee C(t) = p_3' \vee H(t) = p_4' \vee BV(t) = p_5')$ $U = \sum_{t'=t_s}^{t_E} U'(t')$ Solution: $\text{argmax } O(p_l) \underline{\text{def}} U(t)$
--

and only if the patient is suffering at least one hypoglycemia symptomatic events (i.e., fatigue, dizziness, confusion, headaches, blurred vision); and (6) U - a time utility function that calculates the sum of the symptomatic events of hypoglycemia occurred in a time period (from t_s - the first time point, to t_e - the last time point when the patient suffers from the hypoglycemic events). To solve the H-EQPE problem means to find the optimal value of decision parameter p_i to maximize $O(p_i)$ - the objective function. The optimal value of decision parameter p_i can be determined by using the learning algorithm, i.e., L -Checkpoint.

G-EQPE Model

To learn both hypo- and hyperglycemic alarm thresholds in sequence, the authors developed the G-EQPE model that is further extended from the H-EQPE and ME-EQPE models.

The authors construct the formulation of the G-EQPE problem and solution, shown in Table 8 and 9, for detecting the symptomatic events of both hypo- and hyperglycemia in sequence.

The values of the optimal detection parameters (i.e., p and q) can be determined by using the L -Checkpoint algorithm. Based on the threshold set by the medical experts, the logistic regression

model, and the L -Checkpoint algorithm, the performance evaluation for detecting the hypo- and hyperglycemic events from the testing dataset is shown in Table 10. Note that the L -Checkpoint algorithm considerably outperforms the medical experts' criteria and the logistic regression model. From the results, the authors find that the accuracy of the L -Checkpoint algorithm is higher than that of the medical experts' criteria and the logistic regression model respectively. The L -Checkpoint algorithm also obtains the sensitivity that is higher than that of the medical experts' criteria and the logistic regression model respectively. The L -Checkpoint algorithm has a satisfactory specificity as well, which is almost the same as that of the other two approaches.

PROSPECTIVE APPLICATIONS

In addition to the above case studies, the models and the algorithms can be applied to other domains. In this section, the authors briefly describe two other problems that the models and algorithms can be used.

Bradycardia and Tachycardia Detection

A normal person's typical heartbeat rate is between 60 and 100 beats a minute. If the heartbeat rate is less than 60 or greater than 100 beats per minute, a person may suffer bradycardia and tachycardia respectively. However, different persons may have different health conditions, so this typical range cannot be applied to every individual. In order to solve this problem, the EQPE-based models and the Checkpoint-oriented algorithms can be used to learn the thresholds for detecting the occurrence of these two events respectively.

Electric Power Load Shedding

Population growth and economic development are among the key factors that lead to a higher total

Table 8. Time series in detecting the symptomatic events of hypo- and hyperglycemia

Time Series	Abbreviation
Blood Glucose Level	BGL(t)
Fatigue	F(t ₁)
Dizziness	D(t ₁)
Confusion	C(t ₁)
Headaches	H(t ₁)
Blurred Vision	BV(t ₁)
Excessive Thirst	ET(t ₂)
Leg Cramps	LC(t ₂)
Rapid Pulse	RP(t ₂)
Frequent Urination	FU(t ₂)
Dry Mouth	DM(t ₂)

Table 9. G-EQPE problem and solution formulation for detecting the symptomatic events of hypo- and hyperglycemia

Problem: $\langle S_M, S_B, P_M, P_B, C_G, U \rangle$, where $S_M = \{ \vec{S}_1(t_1), \vec{S}_2(t_2) \}$, for $\vec{S}_1(t_1) = \text{BGL}(t_1)$ and $\vec{S}_1(t_2) = \text{BGL}(t_2)$ $S_B = \{ \vec{S}_1'(t_1), \vec{S}_2'(t_2) \}$, for $\vec{S}_1'(t_1) = (F(t_1), D(t_1), C(t_1), H(t_1), BV(t_1))$ and $\vec{S}_2'(t_2) = (ET(t_2), LC(t_2), RP(t_2), FU(t_2), DM(t_2))$ $P_M = \{ \vec{P}_1, \vec{P}_2 \}$, for $\vec{P}_1 = (p)$ and $\vec{P}_2 = (q)$ $P_B = \{ \vec{P}_1', \vec{P}_2' \}$, for $\vec{P}_1' = (p_1', p_2', p_3', p_4', p_5')$ and $\vec{P}_2' = (q_1', q_2', q_3', q_4', q_5')$ $C_{G_1} = \{ C_{G_1}, C_{G_2} \}$, for $C_{G_1} = \text{BGL}(t_1) \leq p \leftrightarrow (F(t_1) = p_1' \vee D(t_1) = p_2' \vee C(t_1) = p_3' \vee H(t_1) = p_4' \vee BV(t_1) = p_5')$ and $C_{G_2} = \text{BGL}(t_2) \geq q \leftrightarrow (ET(t_2) = q_1' \vee LC(t_2) = q_2' \vee RP(t_2) = q_3' \vee FU(t_2) = q_4' \vee DM(t_2) = q_5')$ $U = \sum_{t^{(1)}=t_s}^{t_E} U^{(1)}(t^{(1)}) + \sum_{t^{(2)}=t_s}^{t_E} U^{(2)}(t^{(2)})$ Solution: $\text{argmax } O(\vec{P}_1, \vec{P}_2) \underline{\text{def}} U(t_1, t_2)$
--

Table 10. Performance evaluation for the detection of hypo- and hyperglycemic events from the testing dataset

Decision Approach	Accuracy	Sensitivity	Specificity
Medical Expert's Criteria - 70 mg/dL	85.50%	47.90%	100.00%
Logistic Regression Model - (-17.932, -18.773)	93.89%	78.01%	100.00%
L-Checkpoint Algorithm - 99 mg/dL	99.20%	100.00%	98.80%
Decision Approach	Accuracy	Sensitivity	Specificity
Medical Expert's Criteria - 200 mg/dL	88.96%	73.46%	100.00%
Logistic Regression Model - (-23.205, 0.074)	61.92%	52.12%	68.89%
L-Checkpoint Algorithm - 172 mg/dL	99.34%	98.42%	100.00%

electricity consumption and a peak demand usage which result in a rising energy cost to consumers, as it is more expensive to generate electric power for the peak demand. Thus to learn an optimal peak demand bound over historical and projected electric power demands is necessary. As the demand usage exceeds the bound, some electricity loads are shed by shutting down some electric ac-

count units so that the peak demand charge can be controlled. Thus this optimal peak demand bound can be used to monitor the prospective demand usage in any time interval in order to minimize energy costs and maximize customers' savings while preserving the desired quality of service (QoS) in terms of power interruption.

FUTURE RESEARCH DIRECTIONS

One important issue is that the EQPE-based models and the Checkpoint-oriented algorithms are only able to solve the binary decision-making problems, e.g., buy or sell a stock, suffer from hypo- or hyperglycemia, etc. However, there are still many open scenarios, which involve in multiple decision outcomes, for example, how to determine the status of a fetal heart rate, which could be *Normal*, *Suspect*, or *Pathologic*. This is the future work that the authors will explore in their research team.

CONCLUSION

In this chapter, the authors describe and explain a family development of the parameter-learning models and algorithms: EQPEs and Checkpoints. This class of models and algorithms combines the strength of both qualitative and quantitative methodologies to complement each other to learn optimal decision parameters in an efficient manner to make actionable recommendations. To demonstrate the effectiveness and the efficiency of the models and algorithms, the authors have conducted the two experimental case studies, i.e., the stock market and the glycemia, respectively. The authors show that their approaches are more effective and produce the results that are superior to those of the two other approaches mentioned above.

REFERENCES

- American Diabetes Association. (2014). *Hyperglycemia (High Blood Glucose)*. Retrieved March 13, 2016, from <http://www.diabetes.org/living-with-diabetes/treatment-and-care/blood-glucose-control/hyperglycemia.html>
- American Diabetes Association. (2015). *Hypoglycemia (Low Blood Glucose)*. Retrieved March 13, 2016, from <http://www.diabetes.org/living-with-diabetes/treatment-and-care/blood-glucose-control/hypoglycemia-low-blood.html>
- Bell, J. (2014). *Machine Learning: Hands-On for Developers and Technical Professionals*. Indianapolis, IN: John Wiley & Sons, Inc. doi:10.1002/9781119183464
- Bellman, R. (2003). *Dynamic Programming*. Mineola, NY: Dover Publications, Inc.
- Bellman, R. (2015). *Adaptive Control Processes: a Guided Tour*. Princeton, NJ: Princeton University Press.
- Bentley, J. L. (1975). Multidimensional Binary Search Trees Used for Associative Searching. *Communications of the ACM*, 18(09), 509–517. doi:10.1145/361002.361007
- Bentley, J. L. (1979). Multidimensional Binary Search Trees in Database Applications. *IEEE Transactions on Software Engineering*, 5(04), 333–340. doi:10.1109/TSE.1979.234200
- Bierens, J. (2008). *The Logit Model: Estimation, Testing, and Interpretation*. Department of Economics, The Pennsylvania State University. Retrieved from April 03, 2015, http://grizzly.la.psu.edu/~hbierens/ML_LOGIT.PDF
- Bisgaard, S., & Kulahci, M. (2011). *Time Series Analysis and Forecasting by Example*. Hoboken, NJ: John Wiley & Sons, Inc. doi:10.1002/9781118056943
- Chatfield, C. (2001). *Time-Series Forecasting*. Boca Raton, FL: Chapman & Hall/CRC.
- Cichosz, S. L., Frystyk, J., Hejlesen, O. K., Tarnow, L., & Fleischer, J. S. L. (2014). A Novel Algorithm for Prediction and Detection of Hypoglycemia Based on Continuous Glucose Monitoring and Heart Rate Variability in Patients with Type 1 Diabetes. *Journal of Diabetes Science and Technology*, 8(4), 731–737. doi:10.1177/1932296814528838 PMID:24876412

- Cook, D., & Associates. (2000). *Binary Response and Logistic Regression Analysis. NSF/ILI Project: Beyond Traditional Statistical Methods*. Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison. Retrieved May 23, 2013, from <http://www.stat.wisc.edu/~mchung/teaching/MIA/reading/GLM.logistic.Rpackage.pdf>
- Dougherty, C. (2011). *Introduction to Econometrics* (4th ed.). New York, NY: Oxford University Press.
- Hansen, B. E. (2014). *Econometrics. Department of Economics, University of Wisconsin*. Retrieved June 12, 2012, from <http://www.ssc.wisc.edu/~bhansen/econometrics/Econometrics.pdf>
- Heij, D., De Boer, P., Franses, P. H., Kloek, T., & Van Dijk, H. K. (2004). *Econometric Methods with Applications in Business and Economics*. New York, NY: Oxford University Press.
- Myung, J. (2003). Tutorial on Maximum Likelihood Estimation. *Journal of Mathematical Psychology*, 47(01), 90–100. doi:10.1016/S0022-2496(02)00028-7
- Ngan, C. K., Brodsky, A., & Lin, J. (2010). *Decisions on Multivariate Time Series: Combining Domain Knowledge with Utility Maximization*. The 15th IFIP WG8.3 International Conference on Decision Support Systems, Lisbon, Portugal.
- Ngan, C. K., Brodsky, A., & Lin, J. (2011). *A Service Framework for Querying, Monitoring, and Learning Multivariate Time Series*. The 13th International Conference on Enterprise Information Systems, Beijing, China.
- Ngan, C. K., Brodsky, A., & Lin, J. (2012). An Event-Based Service Framework for Querying, Monitoring, and Learning Multivariate Time Series. *Lecture Notes in Business Information Processing*, 102, 208–223.
- Ngan, C. K., Brodsky, A., & Lin, J. (2012). *R-Checkpoint Algorithm for Multi-Event Decision Making over Multivariate Time Series*. In *Fusing Decision Support Systems into the Fabric of the Context* (Vol. 238, pp. 209–220). IOS Press.
- Ngan, C. K., Brodsky, A., & Lin, J. (2013). Multi-Event Decision Making over Multivariate Time Series. Special Issue on Collaborative Decision Support Systems. *International Journal of Information and Decision Sciences*, 5(3), 263–282. doi:10.1504/IJIDS.2013.055925
- Ngan, C. K., & Li, L. (2015). H-EQPE Model and L-Checkpoint Algorithm: A Decision-Guidance Approach for Detecting Hypoglycemia of Diabetes Patients. *International Journal of Decision Support System Technology*, 7(4), 20–35. doi:10.4018/IJDSST.2015100102
- Rawlings, R., Yuan, L., Shi, H., Brehm, W., Pop-Busui, R., & Nelson, P. (2012). Dynamic Stress Factor (DySF): A Significant Predictor of Severe Hypoglycemic Events in Children with Type 1 Diabetes. *Journal of Diabetes & Metabolism*, 3, 177–188. doi:10.4172/2155-6156.1000177 PMID:24349871
- Samet, H. (2006). *Foundations of Multidimensional and Metric Data Structures*. San Francisco, CA: Morgan Kaufmann.
- Stack, J. B. (2009). Technical and Monetary Investment Analysis. *InvesTech Research*, 9(03), 1–8.

ADDITIONAL READING

- Celikyilmaz, A., & Turksen, B. (2010). Decision Making with Imprecise Parameters. *International Journal of Approximate Reasoning*, 51(8), 869–882. doi:10.1016/j.ijar.2010.06.002

Hwang, C. L., & Masud, A. S. M. (1979). *Multiple Objective Decision Making: Methods and Applications. Lecture Notes in Economics and Mathematical Systems* (Vol. 164). Springer-Verlag Berlin Heidelberg. doi:10.1007/978-3-642-45511-7

Hwang, C. L., & Yoon, K. (1981). *Multiple Attribute Decision Making: Methods and Applications. Lecture Notes in Economics and Mathematical Systems* (Vol. 186). Springer-Verlag Berlin Heidelberg. doi:10.1007/978-3-642-48318-9

Jie Hu, J., Peng, Y., & Xiong, G. (2007). *Robust Parameter Decision-Making Based on Multidisciplinary Knowledge Model. Modeling Decisions for Artificial Intelligence* (pp. 431–442). Springer Berlin Heidelberg.

Katahira, K. (2015). The Relation between Reinforcement Learning Parameters and the Influence of Reinforcement History on Choice Behavior. *Journal of Mathematical Psychology*, 66, 59–69. doi:10.1016/j.jmp.2015.03.006

Lence, S. H., & Hayes, D. J. (1994). Parameter-Based Decision Making Under Estimation Risk: An Application to Futures Trading. *The Journal of Finance*, 49(1), 345–357. doi:10.1111/j.1540-6261.1994.tb04434.x

Wang, Z., Yi, D., Duan, X., Yao, J., & Gu, D. (2011). *Measurement Data Modeling and Parameter Estimation (Systems Evaluation, Prediction, and Decision-Making)*. Boca Raton, FL: CRC Press.

Xu, H. (2009). Robust Decision Making and its Applications in Machine Learning. *Doctoral Dissertation*.

KEY TERMS AND DEFINITIONS

Checkpoint Algorithm: An algorithm to learn one decision parameter to detect the occurrence of a single event.

Glycemic Expert Query Parametric Estimation Model: A model to learn two blood-glucose-level (BGL) thresholds to detect the occurrence of both hypo- and hyperglycemia in sequence.

Hypoglycemic Expert Query Parametric Estimation Model: A model to learn one blood-glucose-level (BGL) threshold to detect the occurrence of hypoglycemia.

Linear L-Checkpoint Algorithm: An algorithm to learn one blood-glucose-level (BGL) threshold to detect the occurrence of hypoglycemia.

Multi-Event Expert Query Parametric Estimation Model: A model to learn multiple decision parameters simultaneously to detect the occurrence of multiple events in sequence.

Multidimensional M-Checkpoint Algorithm: A brute-force algorithm to learn multiple decision parameters optimally at their inter-related time points for multi-events.

Relaxed R-Checkpoint Algorithm: A heuristic algorithm to learn multiple decision parameters, which are fairly close to the optimal parameters learned from the M-Checkpoint algorithm, and to maintain a satisfactorily low time complexity.

Single-Event Expert Query Parametric Estimation Model: A model to learn one decision parameter to detect the occurrence of a single event.

ENDNOTES

¹ The test data set is from 07/01/2005 to 06/30/2009 that is the sell date of the fund with the value of 919.32.

² The test data set is from 02/01/2004 to 03/31/2011.

Informed Decision Making With Enterprise Dynamic Systems Control

Sérgio Luís Guerreiro

Instituto Superior Técnico, University of Lisbon, Portugal & INESC-ID, Portugal

INTRODUCTION

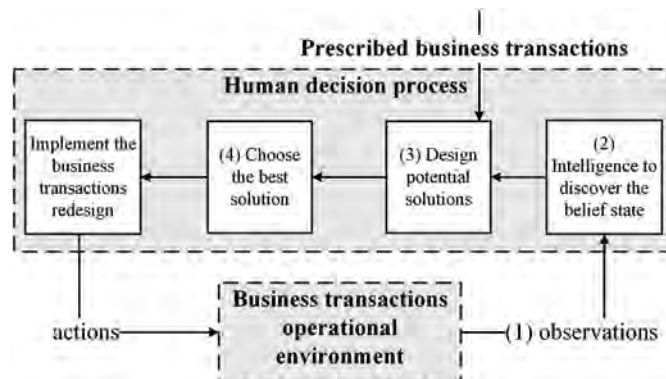
The information systems (IS) are designed, implemented and managed using abstractions layers to cope with the huge organizational complexity that is nowadays posed and also to facilitate the discussion between the different stakeholders of an organization (Laudon & Laudon, 2012) that have diverse perspectives and interpretations of it. Those discussions drive to the classical requirements elicitation stage that aims at identifying the best short-, mid- or long-term models to view, understand and operate the organization and to facilitate the forthcoming IS transformations. In this line, abstraction is a powerful intellectual tool that, in a given instant in time and context, allows to leave some other details for further analysis. In the subsequent instants of time and contexts, the abstraction level decreases and forces the stakeholders to further specify the models. Accordingly, Hoogervorst (2009) explains that business processes models are the result of applying design constraints for a particular organizational reality, which recalls to useful abstractions in order to share a common understanding between the stakeholders that have diverse interpretation of it. In fact, in the IS domain, business processes models (OMG, 2013; Archimate, 2015) are frequently used to describe the way that operations are expected to happen while the actors perform their activities. However, the business transaction models *per se*, are not sufficient and do not guarantee that the business actors perform them accordingly during operation. This unconformance phenomenon (Van der Aalst, 2011) occurs by many and diverse rea-

sons, organizational actors perform workarounds at operation time that could be extremely different from the previous prescribed business transaction models. Operation is the collective activity of all the elements within the organization and in the surrounding environment. It encompasses both the productions performed by the elements within the organization and the interactions with the organizational bounds (Dietz, 2006).

Hence, an actor is autonomous in deciding what to do next, and thus misalignments occur between the business transaction models and actor's operation. Moreover, business actors, individually and/or collectively, operate the organization and also administrate and steer it, by means of observing the state of the world and then acting with purpose to change its state. Moreover, an organizational actor is simultaneous a controller agent and a controlled agent within an enterprise. This reason is why steering the operation of business transactions, by the mean of the correct business rules, is strongly needed nowadays on organizations.

As depicted in Figure 1, organizations require steering for continuous verifying if the desired models are satisfied and then to take purposeful actions to correct them. In line, systems control area identifies the need to construct a classic cycle of observation, decision and control to guarantee that the operation of a system satisfies within the desired conditions (Franklin et al., 2009). Accordingly with these principles, Figure 1 enforces an informed decision-making process, which in practice is a steering cycle with the following counterparts: observation (cf. Figure 1(1)), assessing the environment (cf. Figure 1(2)), designing the

Figure 1. Steering cycle: the business transactions are the organizational objects to be controlled, the observed variables are its states and the control variables are the business transactions redesign initiatives



potential solutions (cf. Figure 1(3)) and choosing the best solution (cf. Figure 1(4)). These counterparts recall to the management competences and are mainly human based. Nevertheless, in this paper, we argue and show how automatic tools deliver support to the managers, aiding at some point in their decision-making tasks.

Moreover, organizational steering is most of the time considered as an independent and isolated organizational add-on component that reacts according with the behavior of the part of the organization that is supposed to control (COBIT, 2007). For instance, the General Systems Theory (Bertalanffy, 1969), the Viable System Model (Beer, 1981) and the recent Enterprise Governance proposals (Hoogervorst, 2009; Hoogervorst & Dietz, 2008).

From the literature, organizational steering is related with the ability to control, within a bounded effort, the operation of the enterprise towards a desired prescription whenever changes or perturbations occur. Steering the organizational operation from *a priori* prescribed models derives from the classical control engineering theories. These approaches are still valid for business information systems domain but require contextual adaptation for dealing with holistic concerns such as models change management

Following the proposals of (Guerreiro et al., 2012; Guerreiro & Tribolet, 2013), in these article the authors state that, due to the organizational

complexity, behavior-based approaches are insufficient because it is impracticable to entire specify the dynamics of the system to be controlled without a constructional perspective of the business transaction models. To produce decisions about which action to enact, the understanding of the essential dynamic of the enterprise is crucial.

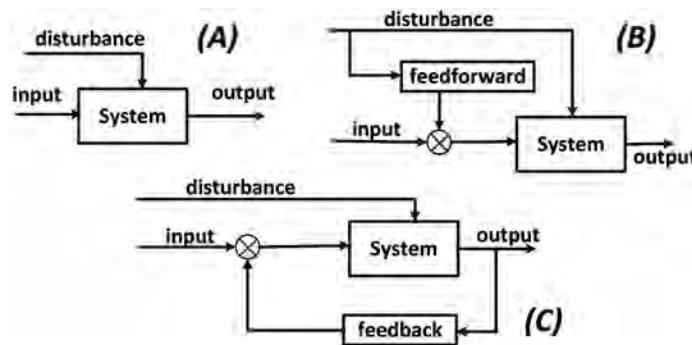
The proposed solution takes advantage of recent advances in the domain of IS ontology, in specific the Enterprise Engineering (Dietz et al., 2013) and the DEMO theory and methodology (Dietz, 2006), to present a steering solution. It triggers two different control actions whenever a misalignment between the business transactions models and operation are identified: (1) a change in the business transaction models to mitigate the misalignment or (2) a change in the business rules because deviation is considered innovative and thus it should be incorporated in the dynamic of the organization.

BACKGROUND

Classical Dynamic Systems Control

From the perspective of classic control concepts (Franklin et al., 2009) the system that we want to control is the execution of the business transactions. The purpose of a control system is to react whenever the disturbance affects the behavior of

Figure 2. Design pattern of control systems. (A) Without control, (B) feed forward control and (C) feedback control



the system or whenever a new input is established. By other words, when the system is not producing the desired output for the imposed input. Control act in the input at the same time as the disturbance is affecting the system.

Figure 2 depicts classical design patterns for a control system. In the top, (A), it shows a system that is not controlled. The disturbance always affects the output delivered by the system. In this pattern, it is not possible to guarantee the behavior of the system output. In the middle, (B), a feed forward pattern that shows that the system input changes accordingly with the actual disturbance. Therefore, the system dynamics it not included in the control actuation. At the bottom of the Figure 2, (C), a feedback control pattern calculates the system input accordingly with the actual misalignment obtained between the output and input. In this pattern, the control actuation calculation takes into consideration the disturbance and the system dynamics. Because the system output depends on the disturbance imposed in the system and on the system dynamics itself.

Organizational Steering

In reality, many classical dynamic control systems exist within an organization and many different scientific perspectives are actually available to the manager, for instance, General Systems Theory (Bertalanffy, 1969), the Viable System Model (Beer, 1981) and the recent Enterprise

Governance proposals (Hoogervorst, 2009; Hoogervorst & Dietz, 2008). Some examples are the access control models (Ferraiolo et al., 2001) that are responsible to grant or revoke the user's access to the different artifacts that exists in an organization. Other example is the business rules that are responsible to maintain the organizational operation within predefined goals (OMG, 2013). Moreover, in a broader scope, the Enterprise Governance that specifies the design restrictions and the subsequent design for the organizational models (Hoogervorst, 2009). Also, in the IT industrial context, the efforts presented by the well spread ITIL (OGC, 2011), which is a set of good practices to be applied on infrastructures, operation and maintenance of IT services, shows a solution that prescribes and steers the operation and a continuous change management processes. In this line, COBIT (COBIT, 2007) prescribes a framework to enforce IT with control mechanisms, using good practices, policies, procedures, practices and organizational structures. COBIT bridges the gap between business risks, control needs and technical aspects. As the main goal, the undesired events are identified and corrected.

Organizational steering is related with the ability to control, within a bounded effort, the operation of the enterprise towards a desired prescription whenever changes or perturbations occur. In line with this concern, Guerreiro et al. (2012) integrates the DSC concepts with the EE concepts to understand, design and implement the

EDSC. More recently, in (Guerreiro & Tribolet, 2013) the EDSC solution details the control for the actor's activity, checking workarounds between the prescribed models and the observations.

Decision-Making

Decision-making is a management competence (Shewhart, 1980) that encompasses: the intelligence to discover the organizational problems, the design of potential solutions, the choosing of the best solution, the implementation of the solution and the verification if the new solution fulfills the desired goals. These stages occur in many levels of organizational management, *e.g.*, project management, operational management, middle management, etc.

Multiple endogenous and exogenous factors promote the need to enforce a continuous decision-making process, for instance, requirements change, legal changes or fraud attempts. In response to these multiple changes, it is necessary to have native decision-making capabilities that continuously find innovative solutions to adapt the organizational operation to be more efficient and effective. In this context, the study of mechanisms to engineer the informed decision-making (Weber, 1987) are key competence for the success of the organization's management. When combining decision-making with business processes, under complex process-based environments, raise the following challenges *(i)* inability to map the current operational observations with the current state where the organization actually is (Montiel & Bickel, 2012), *e.g.*, when actors perform workarounds (Alter, 2014) and override the previous defined prescriptions then the manager need to collect more information to interpret what, in fact, was executed; and *(ii)* incomplete observations (Cassandra, 1998), *e.g.*, because its too expensive to collect information, or, if the business processes are partially performed in paper by humans and partially machine-based. Therefore, in the majority of the situations, the management should support

their decisions in partial information about the surrounding environment (also named as partial observable environments).

D

The EDSC Related Concepts

The concepts of this section derive from the synthesis of the related literature, with a focus to the core EE concepts (Dietz et al., 2013) and DSC (Franklin et al., 2009).

Transition space and state space: An enterprise is a highly complex system of which large parts, like the state of mind of the people in organization, are fundamentally unknowable. In (Bertalanffy, 1969) two distinct spaces are proposed: state space and transition space. The state space is the set of allowable states of a system. The transition space is the set of allowable sequences of transitions of a system. Moreover, to be conformable with EE concepts, we hold the requirement that every state transition is only dependent on the actual state.

Actors: The actors, human or machine, are part of the enterprise and are organised in a network where the individual and collective views of the enterprise coexist (Dietz et al., 2013). Actors have action freedom and act accordingly with their purposes and orchestrations (Winograd, 1986), organized in social systems. Some portions of the enterprise tasks can be automated by software systems while humans perform others. An actor executes different activities over time. For performing an activity, an actor should fulfill tacitly or explicitly, an actor role. An actor is autonomous in deciding what to do next. An user differs from the actor concept because it represents the corresponding access of a given actor role for a given software system. Typically, a relation between user and actor role exists in the organization.

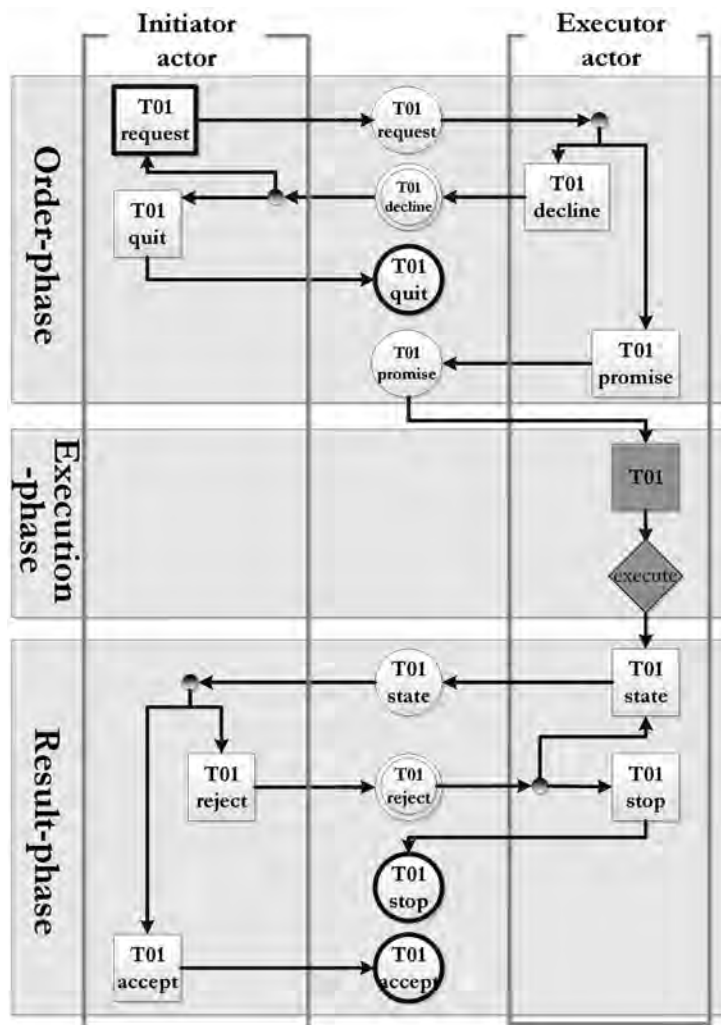
Business Transaction: A business transaction is the result of applying design constraints for a particular organizational reality, valid over a given period of time, and is useful to share a common understanding between the stakeholders that have a diverse interpretation of it. As proposed in EE

(Dietz, 2006), a business transaction involves (1) actor role definitions, in order to specify who is responsible for each part of the transaction, who initiates it and who executes it, (2) a transition space definition, and (3) a state space definition. The state space is the set of allowable states of a system. The transition space is the set of allowable sequences of transitions of a system. State transitions are not dependent on their previous sequence or on the previous states but only on the actual one. When we refer to run-time business transactions, we are referring to the instances of the business transactions model that are execut-

ing at a precise and single instant in time. Many instances of a business transaction model could be executed at the same time in an organization.

Following the Ψ -theory (Dietz, 2006), two distinct actor roles are identified in the standard pattern of a transaction: the Customer and the Producer. The goal of performing such a transaction pattern is to obtain a new fact. The transactional pattern, depicted in Figure 3, is performed by a sequence of coordination and production acts that produces a new service or product, encompassing three distinct phases: (1) order phase with coordination and production acts of request (rq), promise

Figure 3. The DEMO standard pattern of a transaction between two actors with separation between communication and production acts (Adapted from Dietz, 2006)



(pm), decline (dc) and quit (qt), (2) execution phase that includes production act of execution (ex) of the new fact itself and (3) result phase that includes coordination and production acts of state (st), reject (rj), stop (sp) and accept (ac).

Following this definition of business transaction, we introduce the three fundamental dimensions for a business transaction space. The idea is to incorporate the previous concepts into a unique representation, namely:

- **State Space:** Representing the set of allowable states of a system.
- **Transition Space:** Representing the set of allowable sequences of transitions of a system.
- **Actor Role Space:** Representing the set of allowable competences, authorities and delegations of a system.

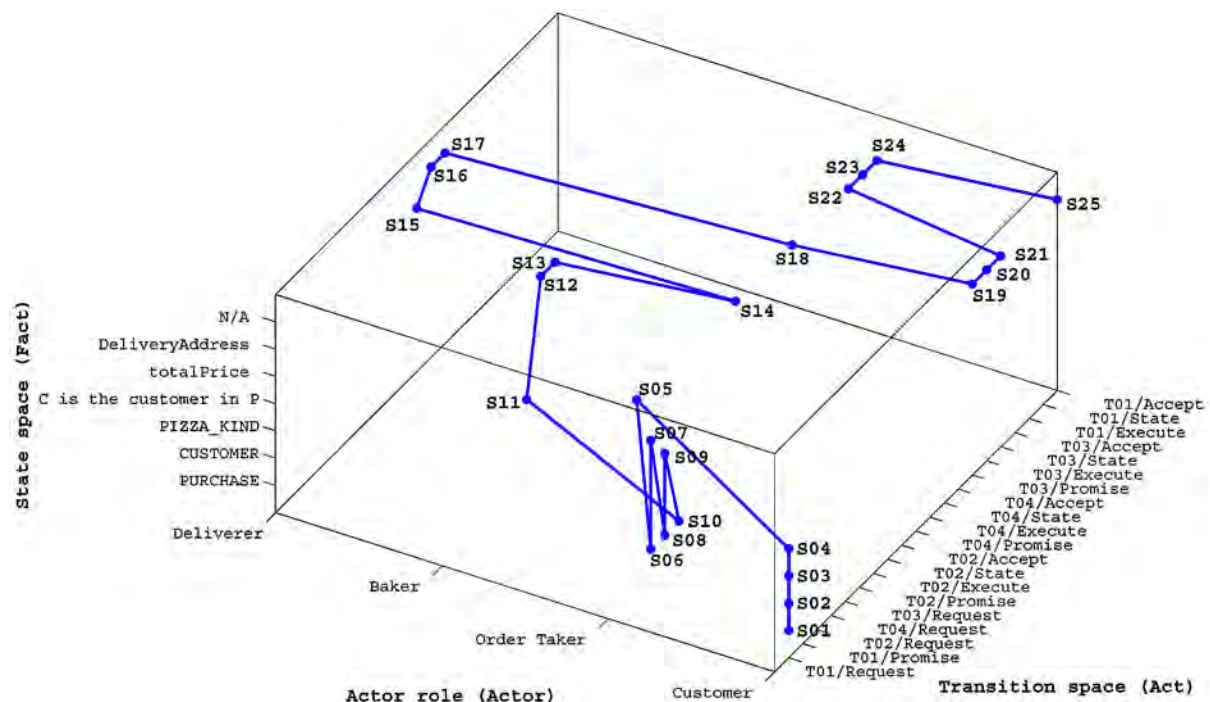
Consequently, each state of the business transaction trajectory is defined by a set of triple in the

form: $S = \{ \langle \text{Transition}_i ; \text{State}_i ; \text{ActorRole}_i \rangle, \dots, \langle \text{Transition}_N ; \text{State}_N ; \text{ActorRole}_N \rangle \}$.

The goal of defining a business transaction space, cf. Figure 4, is to limit the expected execution space where an organization should operate its business transactions instances. If, during operation, a trajectory defined within this space is respected then the organizational purposes are fulfilled otherwise the steering functions should be invoked. Therefore, the business processes change decisions are supported by the alignment (or misalignments) from the expected execution space.

Business transaction model and instance: In the scope of business transactions, a difference exists between the model definition and its instances. A model restricts the design freedom and is relevant to share a common understanding between the different stakeholders of an organization (Op't Land et al., 2009). However, they are not sufficient to be operated by the actors. A model implementation is required, and that implementation is instantiated

Figure 4. 3D graphical representation obtained joining transition space, state space and actor roles space of four business transactions. The line represents the expected execution space.



exactly while the actors perform their activities. Moreover, instances are of cornerstone importance in an organization because its existence is revealed while actors perform actions. An actor role is hard defined in the model, but a user is the one that instantiates the model.

Observable/unobservable operation: Is a core concern for controlling any system. To enable control actions, the actor roles, the state space and the transition space modeled in the business transactions models need to be observed during the business transactions instances execution. There are control variables of an enterprise system that are observable, while others are unobservable. Hence, not all the state space and transition space of the enterprise is observable directly in the operation of an enterprise (Franklin et al., 2009). Observation/non-observation is included in our organizational control solution. This concept implies that an organization does not observe everything about the world. Where, the world is the enterprise and the entire surrounding environment. When a control variable is not observable then a fully informed control decision and actuation is not be performed.

Controllable/uncontrollable operation: Two types of control variables exist in a system, those that are controllable and those that are uncontrollable (Franklin et al., 2009). The control action results in changes in the prescribed business transactions models (negative control) or changes in the business rules (positive control). It is a continuously process running in the enterprise. In the negative control, the prescribed business transactions are changed to avoid the recurrence of unintended operations. In the positive control, innovation is recognized as positive and the deviations from the prescribed transactions models are incorporated in the new prescription. In other words, the observed misalignment is valued as being a more innovative way of operating the organization and thus it is used to define a new organizational prescription, in what constitutes in fact continuous organizational learning

Run-time: Instances exist while the actors perform their activities. When an instance is executing, at a single instant in time, it is run-time. When the models are being designed is called design-time. Moreover, the concept of time is always present in the dynamic of the enterprise. Hence, time should also be considered in the enforcement of a solution for controlling the enterprise. The transition space and state space of an instance evolve via events that are produced at run-time.

EDSC: ONTOLOGICAL BUSINESS TRANSACTIONS STEERING

This section details an ontological proposal to enforce the EDSC approach for business transactions when workarounds occur. Resorting the DEMO theory and methodology, a DEMO Organization Construction Diagram (OCD) is presented using the concepts from the previous section. The concepts competences are mapped into a set of ontological transactions and textual described.

Figure 5 presents two feed forward control diagrams that have the previous presented concepts in mind. The top part designs a non-formal specification. The aim is to share the idea of feedback loops in the scope of the organization operation. Informal blocks and arrows depict the control components and information flow. The solution controls the activity of the actors, checking misalignments between the prescribed models and observations. The observed control variables are used to trigger the EDSC. Introducing the metaphor of the Human Central Nervous Systems (CNS) and Peripheral Nervous System (PNS), PNS grounds on the ability to control using a systemic view of the business transactions operations, checking if complies with the ex-ante business transactions and access control models. The result obtained is one control action: (1) a grant or revoke access to the activities that are currently attempted and/or (2) a change to the prescribed models. CNS grounds in the ability

to control using a systemic view of the historical transactions, checking if complies with the ex-ante business rules. The result is one of the following control actions: (1) a change in the business rules, (2) a change in the business transaction model or (3) a change in the access control model. When needed PNS is able to send an order directly to PNS. For instance, new government laws demanding immediate effect.

This solution integration allows the design of a non-singular solution, because any set of business transactions designed in any business domain could benefit from this solution, and not only a particular subset of business domains.

In the bottom of Figure 5 is detailed the OCD for EDSC proposal, using the concepts from the previous section. The goal of an OCD diagram is to design the ontology of a given solution regarding the actors, the production banks, the coordination banks, the boundary and the overall information flow. The proposed concepts in the previous section are enforced by the competences of a set of ontological transactions. The rationale behind these ontological transactions is:

- The prescribed models regarding business rules, business transaction models and actor's access are supported by T01, T02 and T03,
- Managing of business rules, business transactions models and actor's access are supported by T11, T12 and T13,
- The control of current operation, actor's access and business rules are enforced by T05, T06 and T07,
- And the current operation is observed by T04.

Within the EDSC boundary, the goal is to observe the operation and then to act conformal. In detail, the business rule manager (A01) is responsible to self-initiate the business rule definition transaction (T01) and to initiate the business rule management transaction (T11). Similarly, the model manager (A02) is responsible to self-

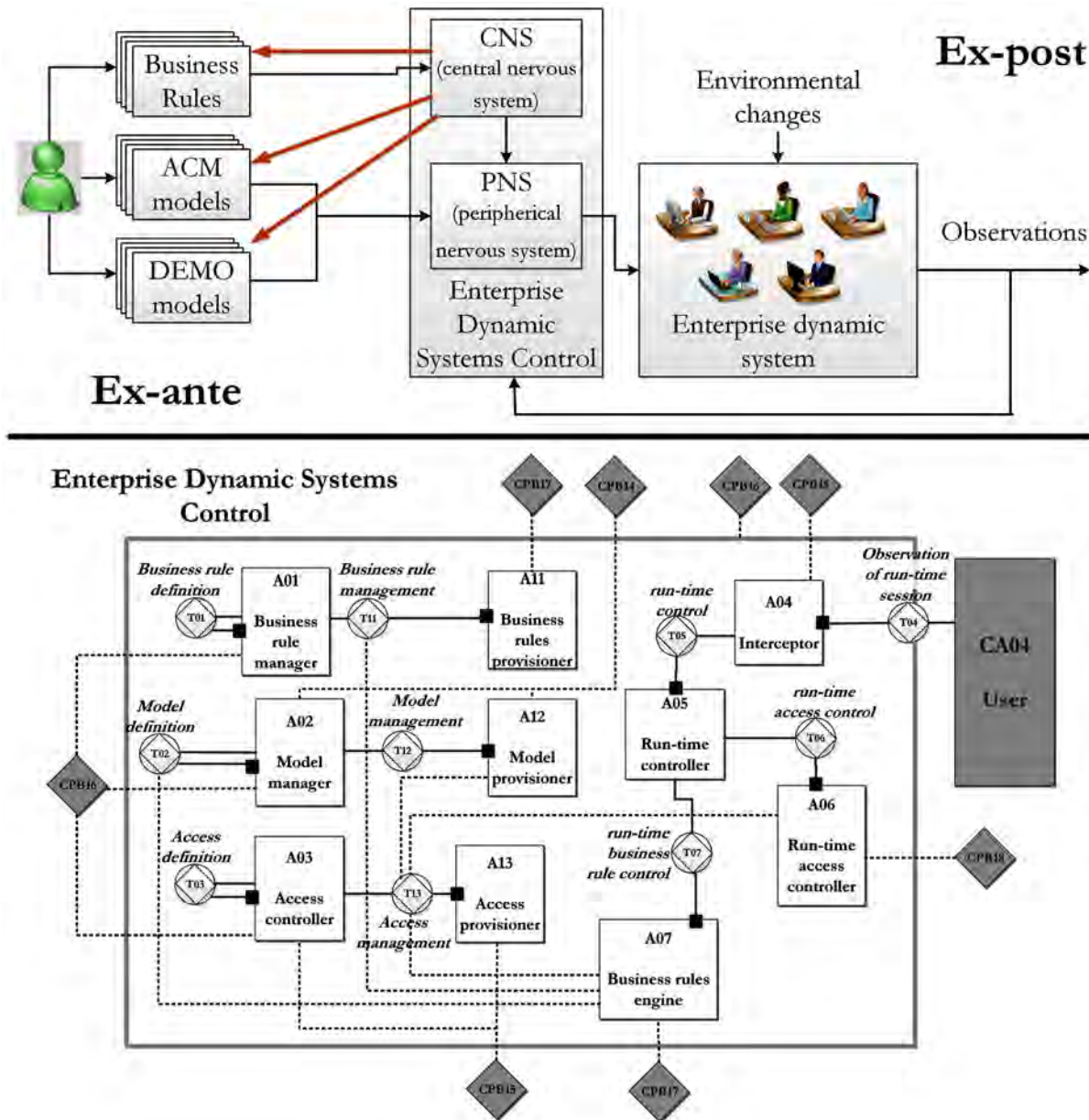
initiate the model definition transaction (T02) and to initiate the model management transaction (T12). Finally, the access controller (A03) is responsible to self-initiate the access definition transaction (T03) and to initiate the access management transaction (T13). The idea of having a self-initiate transaction is to force the definition of the controller prescriptions. If no references are established then no control act can be performed. In other words, references are considered herein as the control bootstrap. The control actuation in the dynamic of the operation occurs by the instantiation of one of the following transactions: business rule management (T11), model management (T12) or access management (T13).

The observation initiates the observation of run-time session transaction (T04). An actor might be a person or a machine. Ontologically, transaction T04 means the connection of the EDSC kernel with the operation of business transaction instances that are compatible with DEMO standard pattern. In practice, it means that observation is a mandatory requirement to enable EDSC.

Once, the observation over the DEMO business transaction space S is established then control is a matter of evaluating the observations and deciding which the correct action to be taken is. That way, every business transaction instance that we want to control must be connected with the boundary of the EDSC. The idea is that S is operated outside EDSC. EDSC is to be integrated with already existing operating systems.

Finally, the run-time controller executes T05 initiating two parallel transactions: run-time access control (T06), control if the access to the sessions should be granted or revoked, and run-time business rule control (T07), control if the prescribed business rules are satisfied. CPB16 refers to the Period that the EG is being considered, by other words, a period where control is valid, hence an information link exists with EDSC boundary.

Figure 5. EDSC. In the top: using free design figures. In the bottom: the OCD DEMO ontological design.



FUTURE RESEARCH DIRECTIONS

Researching organizational control systems is a paramount venture that drives to manifold applications. In fact, diverse impact areas are identified, *e.g.*, Governance Risk and Compliance, business processes run-time compliance verification and access control (Guerreiro et al., 2011), stochastic solutions for decision-making using Markov theo-

ries (Guerreiro, 2013; 2014; 2015), modeling of enterprises considering the risk factors (Gaaloul & Guerreiro, 2015). Moreover, the following research threads are currently open:

1. **Development of Business Transactions Operation Control Using a Cloud Computing Solution:** Develop a software solution, using cloud computing, that offers

the design, operation and control of real business transactions. The idea is to present to the end users a set of business transactions templates, along with a production platform where the business transactions instances operate (Guerreiro et al., 2013).

2. **Risk Dynamics Combined With Business Transactions:** During the business operation many risks may manifest when comparing the prescribed business transactions with its instances. The manifestation of a risk is a non-conformance behavior during the execution of transactions. Some risks are negative and should be avoided (or even revoked), while others are positive and could be incorporated in an evolvable risk pattern repository. However, when managers get aware about a change in the operational conditions, it is often too late to enforce a change in the business transactions prescriptions. Therefore, a new business process compliance solution, able to evolve along with the real-time occurrence of risks, is needed.
3. **Simulation Environments:** Simulators allow the testing of new strategic governance initiatives, in order to understand, in medium and long terms, which are the impacts of each strategy in the organization. Furthermore, a short-term perspective, simulators help the decision making process at low granularity level, and allow the sharing of a common vision about an organizational operation reality among the different stakeholders (Guerreiro, 2013; 2014; 2015).

CONCLUSION

This article conceptualizes and identifies open research challenges in the scope of informed decision-making applied to business processes execution environments. Control is used to cope the workarounds that occur while actors operate. A workaround occurs when an actor decide to adapt, improvise, or perform other change to one

or more aspects of an existing model. In some situations, a workaround could indicate new, and innovative, ways of actors performing their duties. It is not necessarily harmful for the organization.

The designed solution uses DEMO theory and methodology to elicit the set of purposeful requirements for enforcing control in the operation of run-time business transactions. Three feed forward control loops are identified: (1) the business rules compliance, (2) the access control compliance and (3) the predefined business models compliance. The application of these three feed forward control loops is diverse. The abstraction inherited by ontological solutions delivers a powerful intellectual tool that, in a given instant in time and context, allows to leave some details for further analysis. In this context, our design has allowed the discussion between peers and narrowed the design freedom for the solution. Moreover, this design has been triggering the awareness to enforce control mechanisms (using conceptual and stochastic solutions) in the operational environment of an organization between the peers.

ACKNOWLEDGMENT

This work was supported by national funds through Fundação para a Ciência e a Tecnologia (FCT) with reference UID/CEC/50021/2013.

REFERENCES

- Alter, S. (2014). Theory of workarounds. *Communications of the association for information systems. Universal Access in the Information Society*, 34(55), 1041–1066.
- Archimate. (2015). *The Open Group: Archimate version 2.1*. Retrieved November 3, 2015, from <https://www2.opengroup.org/ogsys/catalog/U132>
- Beer, S. (1981). *Brain of the Firm: The Managerial Cybernetics of Organization*. New York: John Wiley & Sons, Inc.

- Bertalanffy, L. (1969). *General Systems Theory*. New York: George Braziller.
- Cassandra, A. R. (1998). A survey of POMDP applications. In *Working Notes of AAAI 1998 Fall Symposium on Planning with Partially Observable Markov Decision Processes* (pp. 17-24). AAAI.
- COBIT. (2007). *Control objectives for information and related technology (cobit). Edition 4.1*. Rolling Meadows, IL: IT Governance Institute. Retrieved November 20, 2013, from <http://www.isaca.org/cobit.htm>
- Dietz, J. (2006). *Enterprise Ontology – Theory and Methodology*. Springer Berlin Heidelberg; doi:10.1007/3-540-33149-2
- Dietz, J., Hoogervorst, J., Albani, A., Aveiro, D., Babkin, E., Barjis, J., & Winter, R. et al. (2013). The discipline of enterprise engineering. *International Journal of Organisational Design and Engineering*, 3(1), 86–114. doi:10.1504/IJODE.2013.053669
- Ferraiolo, D., Sandhu, R., Gavrila, S., Kuhn, D., & Chandramouli, R. (2001). Proposed NIST standard for role-based access control. *ACM Transactions on Information and System Security*, 4(3), 224–274. doi:10.1145/501978.501980
- Franklin, F., Powell, D., & Emami-Naeini, A. (2009). *Feedback control of dynamic systems* (6th ed.). Reading, MA: Addison-Wesley Publishing Company.
- Gaaloul, K., & Guerreiro, S. (2015). A Risk-Based Approach Supporting Enterprise Architecture Evolution. In *The Practice of Enterprise Modeling* (pp. 43–56). Springer International Publishing. doi:10.1007/978-3-319-25897-3_4
- Guerreiro, S. (2013). Business rules elicitation combining Markov decision process with DEMO business transaction space. In *Proceedings of 15th IEEE CBI 2013* (pp. 13-20). doi:10.1109/CBI.2013.11
- Guerreiro, S. (2014). Decision-making in partially observable environments. In *Business Informatics (CBI), 2014 IEEE 16th Conference on* (Vol. 1, pp. 159-166). IEEE. doi:10.1109/CBI.2014.15
- Guerreiro, S. (2015). Engineering the decision-making process using multiple markov theories and DEMO. In *Advances in Enterprise Engineering IX* (pp. 19–33). Springer International Publishing. doi:10.1007/978-3-319-19297-0_2
- Guerreiro, S., & Tribolet, J. (2013). *Conceptualizing Enterprise Dynamic Systems Control For Run-Time Business Transactions*. Paper presented at the 21st ECIS 2013, Utrecht, Netherlands.
- Guerreiro, S., van Kervel, S., & Babkin, E. (2013). Towards devising an architectural framework for Enterprise Operating Systems. In *Proceedings of 8th ICSOFT-PT 2013* (pp. 578-585). doi:10.5220/0004597805780585
- Guerreiro, S., Vasconcelos, A., & Tribolet, J. (2011). Dynamic business transactions control, an ontological example: organizational access control with DEMO. In *Proceedings of the 3rd KEOD 2011, SSEO 2011* (pp. 549-554). doi:10.5220/0003702905490554
- Guerreiro, S., Vasconcelos, A., & Tribolet, J. (2012). *Enterprise dynamic systems control enforcement of run-time business transactions (EEWC 2012)* (Vol. 110). Springer-Verlag. doi:10.1007/978-3-642-29903-2
- Hoogervorst, J. (2009). *Enterprise Governance and Enterprise Engineering*. Springer Berlin. doi:10.1007/978-3-540-92671-9
- Hoogervorst, J., & Dietz, J. (2008, July). Enterprise Architecture in Enterprise Engineering. *Enterprise Modelling and Information Systems Architectures*, 3, 3–11.
- Laudon, K., & Laudon, J. (2012). *Management Information Systems* (12th ed.). Prentice Hall.

Montiel, L., & Bickel, J. (2012). A simulation-based approach to decision making with partial information. *Decision Analysis*, 9(4), 329–347. doi:10.1287/deca.1120.0252

OGC. (2011). *Office for Government Commerce, ITIL v3*. London: Information Technology Infrastructure Library.

OMG. (2013). *Object management group. Semantics of business vocabulary and business rules*. Retrieved November 20, 2013, from <http://www.omg.org/spec/SBVR/1.0/PDF>

Op't Land, M., Proper, E., Waage, M., Cloo, J., & Steghuis, C. (2009). *Enterprise architecture: Creating value by informed governance*. Springer Berlin.

Shewhart, W. (1980). *Economic Control of Quality of Manufactured Product*. American Society for Quality.

Van der Aalst, W. (2011). *Process mining: Discovery, conformance and enhancement of business processes*. Springer Science & Business Media.

Weber, M. (1987). Decision making with incomplete information. *European Journal of Operational Research*, 28(1), 44–57. doi:10.1016/0377-2217(87)90168-8

Winograd, T. (1986). A language/action perspective on the design of cooperative work. In *Proceedings of the ACM conference on Computer-supported cooperative work* (pp. 203–220). New York: ACM doi:10.1145/637069.637096

ADDITIONAL READING

Rozinat, A., & van der Aalst, W. (2008). Conformance checking of processes based on monitoring real behavior. *Information Systems Journal*, 33(1), 64–95. doi:10.1016/j.is.2007.07.001

KEY TERMS AND DEFINITIONS

Actor: The actors, human or machine, are part of the enterprise and are organized in a network where the individual and collective views of the enterprise coexist.

Business Transaction: A model representation of a given organizational reality that is valid within a specific timeframe.

Control: The ability to drive, within a bounded effort, the operation of the enterprise towards a desired prescription whenever changes or perturbations occur.

Decision-Making: A management competence that encompasses: the intelligence to discover the organizational problems, the design of potential solutions, the choosing of the best solution, the implementation of the solution and the verification if the new solution fulfills the desired goals.

Observation: The capability to be aware about the internals of the organization and about its surrounding environment.

Operation: The collective activity of all the elements within the organization and in the surrounding environment.

Workaround: How agents with some degree of behavioral discretion decide whether to follow established practices and what to do when exceptions, anomalies and mishaps occur.

Managerial Tools and Techniques for Decision Making

Davood Askarany

University of Auckland, New Zealand

INTRODUCTION

The main aim of this chapter is to explain what considered as managerial tools and techniques for decision making in the area of management accounting and how they developed and evolved over the time. This chapter is intended to help us document the innovation, evolution and the adoption of a variety of relatively new management accounting techniques and practices in organisations. The chapter also looks at primary tasks/services performed by managerial techniques and tools to capture a wider range of techniques/tools that contribute to the conduction of managerial tasks/services but may not be listed as managerial tools in the literature. In doing so, the article first reviews the introduction of management accounting techniques of the past few decades in the literature and then investigates their implementations in practice. Exploring the levels of organizational satisfactions with adopted management accounting innovations in practice, the article finally discusses the level of associations between the adoption of management accounting innovations and organisational satisfaction

BACKGROUND

There is no universal consensus on what techniques constitute management accounting practice and innovations (Cadez & Guilding, 2008). It is argued that many management accounting techniques are drawn from other disciplines such as engineering and economics (Miller, 1998; Miller, Kurunmäki, & O'Leary, 2008). According to

Miller, et al. (2008), practices such as standard costing, discounted cash flow (DCF), break-even analysis, and much more have been drawn from disciplines other than accounting and then adapted, and constituted as the core of accounting. Contributing to our understanding of what techniques constitute management accounting practices and innovations, this chapter reviews the most prevalent management accounting changes introduced in the literature over the past 70 years. It then examines the adoption and diffusion of latest management accounting innovations in organisations. And finally, the chapter discusses the levels of association between organizational satisfactions and adopted management accounting innovations in practice.

MANAGEMENT ACCOUNTING INNOVATIONS

We may refer to relatively new managerial techniques (introduced over the past three decades) as 'innovation' in this chapter. Rogers (2003) defines innovation as an idea, practice, or object that is perceived as new by an individual or another unit of adoption. Further, he suggests that if the individual has no perceived knowledge about an idea and sees it as new, it is an innovation. Likewise, Damanpour and Gopalakrishnan (1998) define innovation as the adoption of an idea or behaviour new to the organisation. The common criterion in any definition of innovation is newness. According to Rogers (2003), newness in an innovation might be expressed not only in terms of new knowledge, but also in terms of

DOI: 10.4018/978-1-5225-2255-3.ch188

the first persuasion, or decision to adopt. Wolfe (1994) explains the diffusion of innovation as a way the new ideas are accepted (or not) by those to whom they are relevant. Rogers (2003) extends this definition to consider diffusion as a process by which an innovation is communicated through certain channels among the members of a social system. In line with above definitions, we may refer to the process of the evolution and the adoption of relatively newer managerial techniques as 'diffusion' in this chapter.

LITERATURE REVIEW

Johnson and Kaplan (1987, p.12) claim that most of current managerial techniques have been developed during the nineteenth century and the first quarter of the twentieth century. They list the most popular techniques and practices developed by 1925 as follows: cost accounts for labour, material, and overhead; budgets for cash, income, and capital; flexible budgets, sales forecasts, standard costs, variance analysis, transfer prices, and divisional performance measures. According to Chandler (1977), management accounting systems (MAS) first appeared in the United States during the nineteenth century. Johnson & Kaplan (1987) reported that before World War I, the Du Pont Company was using almost all of the management accounting procedures for planning and controlling purposes, known until the 1980s. As with the USA, according to Fleischman & Tyson (2006), the use of cost accounting information was first used in a managerial and purposeful fashion in the New England textile industry in the UK during the 1800s. Similar to other countries such as United States, Canada and Great Britain, trade associations in several industries in France were using uniform costing systems from the end of the nineteenth century to the 1940s (Lemarchand, 2002). Bourguignon, Malleret & Nørreklit (2004) report that 'tableau de bord' as a performance measurement technique that focuses on both financial and non-financial

information was used in France during the first half of the twentieth century.

One of important management accounting innovations that was developed in the first half of the 20th century was called Grenzplankostenrechnung (GPK). GPK is a German management accounting technique, developed in the late 1940s and 1950s, proposed to provide reliable and accurate information on products and services' costs. GPK has also been referred to as Marginal Planned Cost Accounting, Flexible Analytic, Cost Planning and Accounting, and Flexible Margin Costing (Friedl, 2005; Sharman, 2003). Georges Perrin method (GPM) is another old management accounting technique created (in French) by Georges Perrin (1891–1958), who believed GPM could meet the emerging needs for cost calculations for the periods after World War Two (Alcouffe, Berland, & Levant, 2008).

However, the growing scope and speed of technological changes and global competitions in 1980s, has resulted in higher demand for more accurate and advanced managerial techniques to provide organisations with more accurate information for the purpose of planning and control decisions (D Askarany & Yazdifar, 2012; D Askarany, Yazdifar, & Askary, 2010; R. S. Kaplan, 1984). That is why Kaplan (1994) suggests that the 1980s and 1990s have seen a revolution in terms of innovations in management accounting theory and procedures. Echoing this observation, Björnenak & Olson (1999) suggest that during the 1980s and 1990s there has been a rich supply of management accounting innovations in the literature.

According to Askarany (2012) and Hagerty (1997)9), the major developments of managerial techniques and practices since the 1950s can be listed as follows:

- **The 1950s:** Discounted cash flows, Total quality management, Cusum charts and Optimum transfer pricing.
- **The 1960s:** Computer technology, Opportunity cost budgeting, Zero-based budgeting, Decision trees, Critical path

scheduling, and Management by objectives. We can refer to computerized or electronic spreadsheets for business accounting as an important tool developed in the 1960s (Mattessich, 1961). However, electronic spreadsheets made their first appearance for personal computers in 1979 in the form of VisiCalc, as an application to help with accounting tasks (Baker & Sugden, 2003, P.1). Fleischman & Tyson (2006) further consider the 1950s and 1960s as the advent of responsibility accounting with an emphasis on cost control via standard costs.

- **The 1970s:** Information economics and agency theory, Just-in-time scheduling, Strategic business units, Experience curves, portfolio management, Materials resource planning, Diversification, Matrix organisation and Product repositioning. VisiCalc Spreadsheet can be considered as an important tool developed in the 1970s (Baker & Sugden, 2003; Power, 2004).

Fleischman and Tyson (2006) argue that Johnson and Kaplan were badly misguided in claiming that nearly most of current management accounting techniques and practices have been developed before 1925. They emphasize that Johnson and Kaplan failed to appreciate the value and the importance of new developments that had occurred after 1925 and before 1970s. Fleischman & Tyson (2006) highlights innovations introduced prior to the 1970 including the development of a number of new mathematical techniques as follows: linear programming for determining profit maximization or cost minimization in the cases of constrained resources; improved methodologies for the investigation of variances; probability analysis under conditions of uncertainty, direct costing, Programming and Budgeting System (PPBS), Program Evaluation and Review Technique (PERT), Zero-Based Budgeting (ZBB).

- **The 1980s:** Activity-based costing (ABC), activity-based management (ABM), Strategic management accounting (SMA), Activity management (AM), Life cycle costing (LCC), Target costing, Value-added management, Theory of constraints, Vertical integration, Private labels, Benchmarking and Supply Chain management (D Askarany, 2014).

Besides above techniques, Lotus 1-2-3, SuperCalc, PlanPerfect, Quattro Pro, VP-PLANNER, AsEasyAs and Excel Spreadsheets can be considered as some of important tools developed in the 1980s (Baker & Sugden, 2003; Power, 2004). Electronic Data Interchange (EDI) and barcoding are important tools that became popular in the 1980s (Premkumar, Ramamurthy, & Nilakanta, 1994). It should be noted that IBM has made a great contribution to the application and diffusion of most of management accounting tools (including the above's) and other computerized and electronic systems such as Enterprise Resource Planning (ERP), Decision Support System (DSS), Material requirements planning (MRP) and Economic Order Quantity (EOQ).

Adding to the above list, Dugdale & Colwyn (1998) include the application of the theory of constraints (TOC) labelled as 'Throughput accounting' as other advanced management techniques developments of 1980s.

- **The 1990s:** Business process reengineering, Quality functional deployment, Balanced scorecard, Outsourcing, Gain-sharing, Core competencies, Time-based competition and Learning organisation and ERP (D Askarany, 2015).

Expanding the above list, we may also include 'Lean accounting' as another development of 1990s (Kennedy & Widener, 2008). The main purpose of Lean accounting is to reduce steps in transaction processing, eliminate standard costs in

favour of actual costs and discontinue cost allocations (Kennedy & Widener, 2008). In describing lean accounting, Kennedy & Widener (2008) refer to Kaizen costing as continuous improvement and reducing time by eliminating waste and reduction of costs that are the main principles of lean accounting. Another management innovation that was introduced in 1990s to deal with overproduction is called 'Lean Manufacturing' (Holweg, 2007; Yang, Hong, & Modi, 2011). According to Yang et al. (2011), lean manufacturing represents a multifaceted concept that may be grouped together as distinct bundles of organizational practices and could include JIT, total quality management, total preventative maintenance, and human resource management, pull, flow, low setup, controlled processes, productive maintenance and involved employees.

Besides above techniques, following managerial techniques may also be considered as further developments of advanced management innovations in the 1990s: Key performance indicators (KPIs), Resource accounting and budgeting (RAB), Functional analysis, Resource management (Lapsley & Wright, 2004) and Resource Consumption Accounting (RCA). RCA has initially emerged as a management accounting approach around 2000 and is largely based on the German management accounting approach Grenzplankostenrechnung (GPK) which also allows for the use of activity-based drivers (Tse & Gong, 2009; White, 2009; Wikipedia, 2013).

- **The 2000s:** Adding to the above developments, we may also consider Time-Driven Activity-Based Costing (TDABC) as further development of advanced management techniques of 2000s (D. Askarany, 2015; D. Askarany & Smith, 2014; R. Kaplan & Anderson, 2007). We can expand the above list by including the Second Generation Balanced Scorecard, the Third Generation Balanced Scorecard and the Performance Prism as further developments of management accounting innovations in the 2000s

((Lawrie & Cobbold, 2004). There is also some discussion on the fourth and fifth generations of balanced score card in recent years (Christesen, 2008).

The above literature review lists a considerable number of management techniques available for practitioners. However, not all of these techniques have received adequate attention by practitioners or even been considered as pure management accounting techniques by some scholars. Reviewing the development of cost and management accounting innovations of past three decades, Askarany (2012) identifies the most popular management accounting techniques addressed in the literature and introduced to practitioners as follows:

- Activity Based Costing (ABC).
- Activity Based Management (ABM).
- Balanced Scorecard (BSC).
- Benchmarking.
- Value chain analysis.
- Total quality management (TQM).
- Strategic management accounting (SMA).
- Target Costing (TC).
- Risk measurement.
- Re-engineering.
- Outsourcing.
- Life cycle costing (LCC).
- Just-in-time (JIT).

Besides above techniques and tools, there are also some approaches and tasks that received considerable attention in recent years and can be viewed through the lenses of managerial tools and techniques such as 'sustainability', 'integrated reporting', 'data mining', and 'customer relationship management'.

We can also look at primary tasks/services performed by managerial techniques and tools and consider to include any technique/tool (not listed above) that contribute to the conduction of these tasks/services as other managerial tools (Spraaakman, O'Grady, Askarany, & Akroyd, 2015). The followings are some examples of

primary tasks/services performed by managerial tools and techniques:

Costing, budgeting, performance measurement, risk assessment, investment evaluation, customer evaluation, profit analysis, rate and volume analysis, business metrics development, revenue and sale forecasting, price modelling, product profitability, industry or client segment reporting, cost analysis, cost–benefit analysis, break-even point analysis, cost-volume-profit analysis, price analysis, life cycle cost analysis, client profitability analysis, capital budgeting, buy vs. lease analysis, make vs. outsource analysis, strategic planning (e.g. Do-View), strategic management advice, internal financial presentation and communication, financial forecasting, annual budgeting, sensitivity analysis, cost allocation and variance analysis.

Many advocates of relatively new managerial techniques suggest that the adoption of these techniques could contribute to the overall performance of organisations and increase organisational satisfaction (Tatsiopoulos & Panayiotou, 2000; Thyssen, Israelsen, & Jørgensen, 2006; Tornberg, Jämsen, & Paranko, 2002). For example, Maiga & Jacobsred (2008) find that the extent of ABC implementation in organisations is significantly and positively associated with quality, cost, profitability and cycle-time improvements. As another example, Banker & Mashruwala (2007) and Vera-Muñoz et al. (2007) find a significantly positive correlation between employee satisfaction and improved profitability (due to the adoption of managerial technique) in organisations. However, it is not very clear how these techniques have been prevailed in practice. It is also not very clear whether (or not) organisations are satisfied with the performance of their implemented managerial techniques and whether (or not) there is a link between organisational satisfaction and the diffusion of managerial techniques in practice. These questions are addressed in the following section.

DIFFUSION OF MANAGEMENT ACCOUNTING INNOVATIONS IN PRACTICE

Examining the extent of implementation of management accounting innovations in practice, Askarany (2012) finds the adoption rates for some of popular management accounting innovations in Australia as follows: ABC 19%, ABM 11%, BSC 23%, Benchmarking 35%, SMA 13% and TC 15%.

In another study Yazdifar & Askarany (2009) report the adoption rates for the above management accounting innovations in New Zealand and the UK as follows:

- **New Zealand:** ABC 22.5%, ABM 16.9%, BSC 21.4%, Benchmarking 35.2%, SMA 28.2% and TC 18.3%.
- **UK:** ABC 15.2%, ABM 15.2%, BSC 31.8%, Benchmarking 34.8%, SMA 15.2% and TC 16.7%.

Askarany (2012) further explores the level of organizational satisfactions with their implemented management accounting techniques in Australia to see if the adoption of management accounting innovations has contributed to organizational satisfaction. He reports the level of satisfaction of Australian firms with their implemented cost and management accounting systems as follows: Very dissatisfied 7%, Dissatisfied 13%, Moderately satisfied 32% and Very satisfied 9%.

Furthermore, Askarany's (2012) study finds no significance association between organizational satisfaction and the levels of implementation of any of above management accounting innovations in practice. Thus, his findings do not support the notion that the higher the levels of adoption of management accounting innovations, the higher the levels of organizational satisfaction. Though the above findings are not in line with what advocates of management accounting innovations (e.g. Maiga & Jacobs, 2008) believe, the results

are consistent with those findings (e.g. Banker, Bardhan, & Chen, 2008) which suggested that the adoption of management accounting innovation by itself does not improve organizational performance. Given that the majority of these so called management accounting innovations was introduced more than 10-20 years ago, finding no significance association between organizational satisfaction and the levels of implementation of these management accounting innovations could also imply that these techniques have lost their relevance to cope with the new requirements of current organisations.

SOLUTIONS AND RECOMMENDATIONS

These results represent research views situation in Australia, New Zealand and the United Kingdom. However, the experience of other countries, especially of the USA that is a leader in this field and also European developed countries, could vary significantly. For instance, Comparing U.S. and German Cost Accounting Methods, Krumwiede & Suessmair (2007) find a significant difference between U.S. and German organisations (both manufacturing and non-manufacturing organisations) in terms the level of satisfaction with their adopted cost and management accounting systems. Their findings show that German firms have rated their overall satisfactions with their adopted management accounting techniques as 77% (for non-manufacturing firms) to 78% (for manufacturing firms) compared to 23% (for manufacturing firms) to 26% (for non-manufacturing firms) for U.S. firms.

FUTURE RESEARCH DIRECTIONS

Reporting relatively low adoption rates for management accounting innovations as well as low level of organisational satisfaction with the

performance of such innovations suggest further studies into the appropriateness and capability of recent management accounting innovations. In other words, the perceived advantages of management accounting innovations may be inadequate to pursue their adoptions in practice. Or, similar to traditional management techniques, management accounting innovations may have some shortcomings that could prevent them to improve organisational satisfactions.

Despite the introduction and use of majority of management accounting innovations described in this chapter, many organisations worldwide have gone through receivership or collapsed due to either inadequate attention to risk management, performance measurement, overproduction, environment, and stakeholders' needs and preferences or dealing with them in isolation (not altogether and not at the same time). So, it seems to be a need for a dynamic integrated management accounting innovation to simultaneously pay attention to several important factors such as performance management and risk management, overproduction, environment, society and stakeholders' need and preferences.

Inadequate attention to risk management, performance measurement, overproduction, environment, and stakeholders' needs and preferences or dealing with them in isolation are among the problems of the future and could lead to economic crises and obstruction for balanced economic development. So, it is necessary to improve management and management accounting techniques as management accounting should be an amortization means for economic crises. Logistic analysis and overproduction management theory, for instance, could be used in management process for the elimination of potential overproduction. As another example, risk factors, environment and stakeholders' needs and preferences, could be added to balanced scorecard perspectives to simultaneously pay attention to all key performance and key risk indicators in organisations.

CONCLUSION

The purpose of this chapter is to get a clear picture of the scope and extent of recently developed cost and management accounting innovations in the literature as well as their implementation in practice. The chapter identifies the most popular management accounting innovations of past several decades in the literature and reports on the latest adoption rates of most recent management accounting innovations as well as their associations with organisational satisfactions in some selected countries.

The current chapter identifies some popular cost and management accounting changes/innovations of past three decades as follows: performance measurement and BSC techniques, ABC, TDABC, valued added concepts, ABM, TQM, strategic management, risk management, benchmarking, re-engineering, and target costing. However, from the implementation perspective, the chapter suggests that recent management accounting innovations are not widely implemented in practice.

Reviewing the level of association between organizational satisfaction and the adoption of management accounting innovations, the chapter further reports no significant association between organizational satisfaction and the implementation of any of above management accounting innovations. Considering the cost and benefit philosophy, this chapter suggests that perceived advantages of these relatively new management accounting innovations could be inadequate to pursue their adoptions in practice.

This chapter provides a historical review on the evolution and the diffusion of a variety of relatively new management accounting techniques and practices in organisations in general. However, in terms of the level of organisational satisfaction with adopted management accounting innovations, the results are mainly from studies carried out in limited countries: Australia, New Zealand and the United Kingdom and to some extent the USA and Germany. The findings show that the level

of satisfaction with management accounting innovations varies significantly across the countries. Further studies are recommended to explore the experiences of other countries with the adoption of management accounting techniques in practice.

REFERENCES

- Alcouffe, S., Berland, N., & Levant, Y. (2008). Actor-networks and the diffusion of management accounting innovations: A comparative study. *Management Accounting Research*, 19(1), 1–17. doi:10.1016/j.mar.2007.04.001
- Askarany, D. (2012). An Exploratory Study Into the Implementation of Management Accounting Changes in Australia. *Journal of Modern Accounting and Auditing*, 8(5), 628–640.
- Askarany, D. (2014). Characteristics of innovation and the diffusion of benchmarking. *Problems and Perspectives in Management*, 12(4), 563–572.
- Askarany, D. (2015). The Adoption of Balanced Scorecard through the Lenses of Contingency and the Diffusion Theories. *Global Review of Accounting and Finance*, 6(1), 140–150. doi:10.21102/graf.2015.03.61.09
- Askarany, D. (2015). Time-Driven Activity-Based Costing. In *Cost Accounting: A Managerial Emphasis* (15th ed.; p. 182). Hagerstown, MD: Pearson-GEX Publishing Service.
- Askarany, D., & Smith, A. (2014). Cost Benefit Analyses of Organic Waste Composting Systems through the Lens of Time-Driven Activity-Based Costing. *Journal of Applied Management Accounting Research*, 12(2), 59–73.
- Askarany, D., & Yazdifar, H. (2012). An Investigation into the Mixed Reported Adoption Rates for ABC: Evidence from Australia, New Zealand and the UK. *International Journal of Production Economics*, 135(1), 430–439. doi:10.1016/j.ijpe.2011.08.017

- Askarany, D., Yazdifar, H., & Askary, S. (2010). Supply chain management, activity-based costing and organisational factors. *International Journal of Production Economics*, 127(2), 238–248. doi:10.1016/j.ijpe.2009.08.004
- Baker, J., & Sugden, S. J. (2003). Spreadsheets in Education –The First 25 Years. *Spreadsheets in Education*, 1(1), 18–43.
- Banker, R., Bardhan, I. R., & Chen, T.-Y. (2008). The role of manufacturing practices in mediating the impact of activity-based costing on plant performance. *Accounting, Organizations and Society*, 33(1), 1–19. doi:10.1016/j.aos.2006.12.001
- Banker, R. D., & Mashruwala, R. (2007). The Moderating Role of Competition in the Relationship Between Nonfinancial Measures and Future Financial Performance. *Contemporary Accounting Research*, 24(3), 763–793. doi:10.1506/car.24.3.4
- Bjørnønenak, T., & Olson, O. (1999). Unbundling management accounting innovations. *Management Accounting Research*, 10(4), 325–338. doi:10.1006/mare.1999.0110
- Bourguignon, A., Malleret, V., & Nørreklit, H. (2004). The American balanced scorecard versus the French tableau de bord: The ideological dimension. *Management Accounting Research*, 15(2), 107–134. doi:10.1016/j.mar.2003.12.006
- Cadez, S., & Guilding, C. (2008). An exploratory investigation of an integrated contingency model of strategic management accounting. *Accounting, Organizations and Society*, 33(7-8), 836–863. doi:10.1016/j.aos.2008.01.003
- Chandler, D. A. (1977). *Visible Hand: The management revolution in American business*. Cambridge, MA: Harvard University Press.
- Christesen, D. A. (2008). *The Impact of Balanced Scorecard Usage on Organization Performance* (Ph.D.). University of Minnesota. Retrieved from <http://search.proquest.com/docview/304595608>
- Damanpour, F., & Gopalakrishnan, S. (1998). Theories of organizational structure and innovation adoption: The role of environmental change. *Journal of Engineering and Technology Management*, 15(1), 1–24. doi:10.1016/S0923-4748(97)00029-5
- Dugdale, D., & Colwyn, T. J. (1998). Throughput accounting: Transforming practices? *The British Accounting Review*, 30(3), 203–220. doi:10.1006/bare.1997.0062
- Fleischman, R., & Tyson, T. (2006). The History of Management Accounting in the U.S. In *Handbooks of Management Accounting Research* (Vol. 2, pp. 1071-1089). Elsevier.
- Friedl, G. H.-U. B. (2005). Relevance Added: Combining ABC with German Cost Accounting. *Strategic Finance*, 86(12), 56–61.
- Hagerty, M. R. (1997). A Powerful Tool for Diagnosis and Strategy. *Journal of Management Consulting*, 9(4), 16–25.
- Holweg, M. (2007). The genealogy of lean production. *Journal of Operations Management*, 25(2), 420–437. doi:10.1016/j.jom.2006.04.001
- Johnson, & Kaplan, R. S. (1987). *Relevance lost*. Harvard Business School Press.
- Kaplan, R., & Anderson, S. (2007). The innovation of time-driven activity-based costing. *Coastal Management*, 21(2), 5–15.
- Kaplan, R. S. (1984). The Evolution of Management Accounting. *Accounting Review*, 59(3), 390–418.
- Kaplan, R. S. (1994). Management Accounting (1984-1994): Development of new practice and theory. *Management Accounting Research*, 5(3-4), 247–260. doi:10.1006/mare.1994.1015
- Kennedy, F. A., & Widener, S. K. (2008). A control framework: Insights from evidence on lean accounting. *Management Accounting Research*, 19(4), 301–323. doi:10.1016/j.mar.2008.01.001

- Krumwiede, K., & Suessmair, A. (2007). Comparing U.S. and German Cost Accounting Methods. *Management Accounting Quarterly*, 8(3), 1–9.
- Lapsley, I., & Wright, E. (2004). The diffusion of management accounting innovations in the public sector: A research agenda. *Management Accounting Research*, 15(3), 355–374. doi:10.1016/j.mar.2003.12.007
- Lawrie, G., & Cobbold, L. (2004). Third-generation balanced scorecard: Evolution of an effective strategic control tool. *International Journal of Productivity and Performance Management*, 53(7), 611–623. doi:10.1108/17410400410561231
- Lemarchand, Y. (2002). The military origins of the French management accounting model: A return to the mechanisms of accounting change. *Accounting History*, 7(1), 23–57. doi:10.1177/103237320200700103
- Maiga, A. S., & Jacobs, F. A. (2008). Extent of ABC Use and Its Consequence. *Contemporary Accounting Research*, 25(2), 533–566. doi:10.1506/car.25.2.9
- Mattessich, R. (1961). Budgeting Models and System Simulation. *Accounting Review*, 36(3), 384–397. doi:10.2307/242869
- Miller, P. (1998). The margins of accounting. *European Accounting Review*, 7(4), 605–621. doi:10.1080/096381898336213
- Miller, P., Kurunmäkii, L., & OLeary, T. (2008). Accounting, hybrids and the management of risk. *Accounting, Organizations and Society*, 33(7-8), 942–967. doi:10.1016/j.aos.2007.02.005
- Power, D. J. (2004). *A Brief History of Spreadsheets*. Academic Press.
- Premkumar, G., Ramamurthy, K., & Nilakanta, S. (1994). Implementation of Electronic Data Interchange: An Innovation Diffusion Perspective. *Journal of Management Information Systems*, 11(2), 157–186. doi:10.1080/07421222.1994.11518044
- Rogers, E. M. (2003). *Diffusion of innovations*. New York: Free Press.
- Sharman, P. A. (2003). Bring on German Cost Accounting. *Strategic Finance*, 85(6), 30–38.
- Spraakman, G., O’Grady, W., Askarany, D., & Akroyd, C. (2015). Employers’ Perceptions of Information Technology Competency Requirements for Management Accounting Graduates. *Accounting Education: An International Journal*, 24(5), 403–422.
- Tatsiopoulos, I. P., & Panayiotou, N. (2000). The integration of activity based costing and enterprise modeling for reengineering purposes. *International Journal of Production Economics*, 66(1), 33–44. doi:10.1016/S0925-5273(99)00080-8
- Thyssen, J., Israelsen, P., & Jørgensen, B. (2006). Activity-based costing as a method for assessing the economics of modularization A case study and beyond. *International Journal of Production Economics*, 103(1), 252–270. doi:10.1016/j.ijpe.2005.07.004
- Tornberg, K., Jämsen, M., & Paranko, J. (2002). Activity-based costing and process modelling for cost-conscious product design: A case study in a manufacturing company. *International Journal of Production Economics*, 79(1), 75–82. doi:10.1016/S0925-5273(00)00179-1
- Tse, M. S. C., & Gong, M. Z. (2009). Recognition of Idle Resources in Time-Driven Activity-Based Costing and Resource Consumption Accounting Models. *The Journal of Applied Management Accounting Research*, 7(2), 41–54.
- Vera-Muñoz, S. C., Shackell, M. B., & Buehner, M. (2007). Accountants Use of Causal Business Models in the Presence of Benchmark Data: A Note. *Contemporary Accounting Research*, 24(3), 1015–1038. doi:10.1506/car.24.3.12
- White, L. (2009). Resource Consumption Accounting: Manager-Focused Management Accounting. *Journal of Corporate Accounting & Finance*, 20(4), 63–77. doi:10.1002/jcaf.20501

Wikipedia. (2013). *Resource consumption accounting*. Retrieved from http://en.wikipedia.org/wiki/Resource_consumption_accounting

Wolfe, R. A. (1994). Organizational innovation: Review, critique and suggested research directions. *Journal of Management Studies*, 31(3), 405–431. doi:10.1111/j.1467-6486.1994.tb00624.x

Yang, M. G., Hong, P., & Modi, S. B. (2011). Impact of lean manufacturing and environmental management on business performance: An empirical study of manufacturing firms. *International Journal of Production Economics*, 129(2), 251–261. doi:10.1016/j.ijpe.2010.10.017

Yazdifar, H., & Askarany, D. (2009). A comparative investigation into the diffusion of management accounting innovations in the UK, Australia and New Zealand. *Chartered Institute of Management Accountant*, 5(12), 1–11.

ADDITIONAL READING

Al-Omiri, M., & Drury, C. (2007). A survey of factors influencing the choice of product costing systems in UK organizations. *Management Accounting Research*, 18(4), 399–424. doi:10.1016/j.mar.2007.02.002

Anderson, S. W. (2007). *Managing Costs and Cost structure throughout the Value Chain. Handbook of Management Accounting Research*. A. G. H. Christopher S. Chapman, Michael D. Shields (1st ed., Vol. 2, pp. 481–506). UK: Elsevier.

Anderson, S. W., & Young, S. M. (1999). The impact of contextual and process factors on the evaluation of activity-based costing systems. *Accounting, Organizations and Society*, 24(7), 525–559. doi:10.1016/S0361-3682(99)00018-5

Askarany, D. (2003). An Overview of the Diffusion of Advanced Techniques. In B. F. Tan (Ed.), *Advanced Topics in Global Information* (Vol. 2, pp. 225–250). London: IDEA Group Publishing. doi:10.4018/978-1-59140-064-6.ch012

Baird, K. (2007). Adoption of activity management practices in public sector organizations. *Accounting and Finance*, 47(4), 551–569. doi:10.1111/j.1467-629X.2007.00225.x

Berliner, C., & Brimson, J. (1988). *Cost Management for Today's Advanced Manufacturing: The CAM-I Conceptual Design*. Boston: Harvard Business School Press.

Birkinshaw, J., Hamel, G., & Mol, M. (2008). Management innovation. *Academy of Management Review*, 33(4), 825–845. doi:10.5465/AMR.2008.34421969

Boonstra, J. J., & Vink, M. J. (1996). Technological and organizational innovation: A dilemma of fundamental change and participation. *European Journal of Work and Organizational Psychology*, 5(3), 351–375. doi:10.1080/13594329608414865

Bork, H. P., & Morgan, M. J. (1993). Is ABC really a Need not an Option? *Management Accounting Research*, 71(8), 26–27.

Chandler, D. A. (1977). *Visible Hand: The management revolution in American business*. Cambridge: Harvard University Press.

Chapman, C. S., Hopwood, A. G., & Shields, M. D. (2006). *Handbooks of Management Accounting Research* (Vol. 1, pp. 1–561). UK: Elsevier.

Cooper, R. G., & Kleinschmidt, E. J. (1990). New product success factors: A comparison of kills versus successes and failures. *R & D Management*, 20(1), 47–63. doi:10.1111/j.1467-9310.1990.tb00672.x

Currie W.L. (1999), Revisiting management innovation and change programmes: strategic vision or tunnel vision?, Omega Elsevier, Volume 27, Issue 6.

Damanpour, F., & Aravind, D. (2012). Managerial Innovation: Conceptions, Processes, and Antecedents. *Management and Organization Review*, 8(2), 423–454. doi:10.1111/j.1740-8784.2011.00233.x

Damanpour, F., & Wischnevsky, J. D. (2006). Research on innovation in organizations: Distinguishing innovation-generating from innovation-adopting organizations. *Journal of Engineering and Technology Management*, 23(4), 269–291. doi:10.1016/j.jengtecman.2006.08.002

Dikolli, S. S., Kinney, W., & Sedatole, K. (2007). Measuring Customer Relationship Value: The Role of Switching Cost*. *Contemporary Accounting Research*, 24(1), 93–132. doi:10.1506/N658-3452-4X2L-0847

Fagerberg, J. (2005). Innovation: A guide to the literature. In J. Fagerberg, D. C. Mowery & R. R. Nelson (Eds.), *The Oxford Handbook of innovations*: 1–26. Oxford: Oxford University

Gosselin, M. (1997). The Effect of Strategy and Organizational Structure on the adoption and implementation of Activity Based Costing. *Accounting, Organizations and Society*, 22(2), 105–122. doi:10.1016/S0361-3682(96)00031-1

Handbooks of Management Accounting Research Chapman, written by Hopwood, & Shields (2006).

Johnson, H. T., & Kaplan, R. S. (1987). *Relevance Lost—The Rise and Fall of Management Accounting*. Boston, Massachusetts: Harvard Business School Press.

Kelly, K. O. (2007). Feedback and Incentives on Nonfinancial Value Drivers: Effects on Managerial Decision Making. *Contemporary Accounting Research*, 24(2), 523–529. doi:10.1506/Q5X8-5912-W613-0UTL

Langfield-Smith, K. (1997). Management control system and strategy: A critical review. *Accounting, Organizations and Society*, 22(2), 207–232. doi:10.1016/S0361-3682(95)00040-2

Monden Y., Hamada K. (1991), Target Costing and Kaizen Costing in Japanese Automobile Companies, *Journal of Management Accounting Research*, Vol.3, Fall 1991

Schneeweiss, C. (1998). On the applicability of activity based costing as a planning instrument. *International Journal of Production Economics*, 54(3), 277–284. doi:10.1016/S0925-5273(98)00003-6

Schoch, H., Yap, T. H., Hye, L. M., & Beng, A. K. (1994). Activity Based Costing in the Electronics Industry: The Singapore Experience. *Small Business and Entrepreneurship*, 11(2), 28–37. doi:10.1080/08276331.1994.10600454

Yazdifar, H., & Askarany, D. (2012). A comparative study of the adoption and implementation of target costing in the UK, Australia and New Zealand. *International Journal of Production Economics*, 135(1), 382–392. doi:10.1016/j.ijpe.2011.08.012

KEY TERMS AND DEFINITIONS

Activity-Based Costing (ABC): An approach to costing that focuses on activities as the fundamental cost objects. It uses the cost of these activities as the basis for assigning costs to other cost objects such as products, services, or customers.

Activity-Based Management (ABM): Use of ABC concepts to facilitate the identification and reduction of non-value-added activities.

Balanced Scorecard: An integrated strategic performance management framework that helps organisations translate strategic objectives into relevant performance measures, by linking non-financial measures with a financial perspective in four areas of performance concerned with: financials, internal process, customers and innovation & learning.

Benchmarking: The search for industry best practice that will lead to superior performance. It emphasises an outward focus and seeks to improve performance by learning from the experience of effective organisations.

Life Cycle Costing: A costing method that tracks and accumulates the actual costs attribut-

able to each product from its initial research and development to its final customer servicing and support in the marketplace.

Strategic Management Accounting (SMA):

A focus on the analysis of the external environment which mandates corrections and adjustments to the internal control systems structures and decision support systems that are vital for the survival of organisations. SMA has an orientation towards

the organisation's environment such as suppliers, customers, and its competitive position relative to both existing and potential competitors.

Target Costing: A form of costing system in which the manufacturing of a product or the provision of a service is restricted within a pre-determined total cost ceiling so that a competitive price is achieved.

D

A Nature-Inspired Metaheuristic Approach for Generating Alternatives

Julian Scott Yeomans
York University, Canada

INTRODUCTION

Decision-making in the “real world” involves complex problems that tend to be riddled with competing performance objectives and possess requirements which are very difficult to incorporate into any underlying decision support models (Brugnach, Tagg, Keil, De Lange, 2007; Janssen, Krol, Schielen, Hoekstra, 2010; Mowrer, 2000; Walker, Harremoes, Rotmans, Van der Sluis, Van Asselt, Janssen, Krayen von Krauss, 2003). While an optimal solution might provide the theoretically best answer to a mathematical model, in general, it will not be the best solution to the fundamental “real” problem since there are invariably unmodelled objectives and unquantifiable issues not incorporated in the problem formulation (Brugnach *et al.*, 2007; Gunalay, Yeomans, 2012; Gunalay, Yeomans, Huang, 2012; Janssen *et al.*, 2010; Loughlin, Ranjithan, Brill, Baugh, 2001). Consequently, it is preferable to generate a number of different alternatives that provide multiple, disparate perspectives to any particular problem (Imanirad, Yeomans, 2014; Matthies, Giupponi, Ostendorf, 2007; Yeomans, Gunalay, 2011). Preferably these alternatives should all possess good (i.e. near-optimal) objective measures with respect to the modelled objective(s), but be as fundamentally different as possible from each other in terms of the system structures characterized by their decision variables (Yeomans, 2011).

To address this option creation need, several approaches collectively referred to as *modelling-to-generate-alternatives* (MGA) have been developed (Loughlin *et al.*, 2001; Yeomans, Gunalay, 2011; Yeomans, 2012). The principal motivation

for MGA is to create a small set of alternatives that are as maximally different from each other in the decision space as possible, yet are still considered “good” with respect to all of the modelled objective(s) (Yeomans, 2011; Yeomans, 2012). By adopting a maximally different method, the resulting alternative solution set is likely to provide very different perspectives with respect to any unmodelled issues, while simultaneously providing different choices that all perform somewhat similarly with respect to the modelled objectives (Gunalay, Yeomans, 2012; Gunalay *et al.*, 2012; Walker *et al.*, 2003; Yeomans, 2011).

In this chapter, it is shown how a modified version of the metaheuristic Firefly Algorithm (FA) of Yang (2009; 2010) can be used to efficiently generate a set of maximally different solution alternatives. Yang (2010) has demonstrated that, for optimization and calculational purposes, the FA is more computationally efficient than the more commonly-employed enhanced particle swarm, genetic algorithm, and simulated annealing metaheuristic procedures. Thus, this FA-based MGA procedure can be considered very computationally efficient (Imanirad, Yeomans, 2014). This demonstrates the MGA proficiencies of the FA-based approach for constructing multiple, maximally different solution alternatives to the highly non-linear optimization problem of Loughlin *et al.* (2001).

BACKGROUND

While this section provides a brief synopsis of the steps involved in the FA process, more specific details can be found in Yang (2009; 2010). The

Algorithm 1.

```

Objective Function  $F(\mathbf{X})$ ,  $\mathbf{X} = (x_1, x_2, \dots, x_d)$ 
Generate the initial population of  $n$  fireflies,  $\mathbf{X}_i$ ,  $i = 1, 2, \dots, n$ 
Light intensity  $I_i$  at  $\mathbf{X}_i$  is determined by  $F(\mathbf{X}_i)$ 
Define the light absorption coefficient  $\gamma$ 
while ( $t < \text{MaxGeneration}$ )
    for  $i = 1: n$ , all  $n$  fireflies
        for  $j = 1: n$ , all  $n$  fireflies (inner loop)
            if ( $I_i < I_j$ ), Move firefly  $i$  towards  $j$ ; end if
            Vary attractiveness with distance  $r$  via  $e^{-\gamma r^2}$ 
        end for  $j$ 
    end for  $i$ 
    Rank the fireflies and find the current global best solution  $\mathbf{G}^*$ 
end while
Postprocess the results

```

FA is a nature-inspired, population-based meta-heuristic that employs the following three idealized rules: (i) All fireflies within a population are unisex, so that one firefly will be attracted to other fireflies irrespective of their sex; (ii) Attractiveness between fireflies is proportional to their brightness, implying that for any two flashing fireflies, the less bright one will move towards the brighter one. Attractiveness and brightness both decrease as the distance between fireflies increases. If there is no brighter firefly within its visible vicinity, then a particular firefly will move randomly; and (iii) The brightness of a firefly is determined by the landscape of the objective function. Namely, for a maximization problem, the brightness can simply be considered proportional to the value of the objective function. Based upon these three rules, the basic operational steps of the FA are summarized within the pseudo-code of Algorithm 1 Yang (2010).

In the FA, there are two important issues to resolve: the variation of light intensity and the formulation of attractiveness. For simplicity, it can always be assumed that the attractiveness of a firefly is determined by its brightness which in turn is associated with the encoded objective function. In the simplest case, the brightness of a firefly at a particular location $\mathbf{X}$ would be its

calculated objective value $F(\mathbf{X})$. However, the attractiveness, β , between fireflies is relative and will vary with the distance r_{ij} between firefly i and firefly j . In addition, light intensity decreases with the distance from its source, and light is also absorbed in the media, so the attractiveness should be allowed to vary with the degree of absorption. Consequently, the overall attractiveness of a firefly can be defined as

$$\beta = \beta_0 \exp(-\gamma r^2)$$

where β_0 is the attractiveness at distance $r = 0$ and γ is the fixed light absorption coefficient for a specific medium. If the distance r_{ij} between any two fireflies i and j located at $\mathbf{X}_i$ and $\mathbf{X}_j$, respectively, is calculated using the Euclidean norm, then the movement of a firefly i that is attracted to another more attractive (i.e. brighter) firefly j is determined by

$$\mathbf{X}_i = \mathbf{X}_i + \beta_0 \exp(-\gamma(r_{ij})^2)(\mathbf{X}_i - \mathbf{X}_j) + \alpha \mathbf{e}_i.$$

In this expression of movement, the second term is due to the relative attraction and the third term is a randomization component. Yang (2010) indicates that the randomization parameter, α , is normally selected within the range $[0,1]$ and $\mathbf{e}_i$ is

a vector of random numbers drawn from either a Gaussian or uniform (generally $[-0.5, 0.5]$) distribution. It should be noted that this expression is a random walk biased toward brighter fireflies and if $\beta_0 = 0$, it becomes a simple random walk. The parameter γ characterizes the variation of the attractiveness and its value determines the speed of the algorithm's convergence. For most applications, γ is typically set between 0.1 to 10 (Yang, 2010). In any given optimization problem, for a very large number of fireflies $n \gg k$ where k is the number of local optima, the initial locations of the n fireflies should be distributed relatively uniformly throughout the entire search space. As the FA proceeds, the fireflies would converge into all of these local optima (including the global ones). By comparing the best solutions among all these optima, the global optima can easily be determined. Yang (2010) demonstrates that the FA will approach the global optima when $n \rightarrow \infty$ and the number of iterations t , is set so that $t \gg 1$. In reality, the FA has been found to converge extremely quickly.

Two important limiting or asymptotic cases occur when $\gamma \rightarrow 0$ and when $\gamma \rightarrow \infty$. For $\gamma \rightarrow 0$, the attractiveness is constant $\beta = \beta_0$, which is equivalent to having a light intensity that does not decrease. Thus, a firefly would be visible anywhere within the solution domain. Hence, a single (usually global) optima can easily be reached. If the inner loop for j in Figure 1 is removed and X_j is replaced by the current global best G^* , then this implies that the FA becomes a special case of the accelerated particle swarm optimization (PSO) algorithm. Subsequently, the computational efficiency of this special case of the FA is equivalent to that of enhanced PSO. Conversely, when $\gamma \rightarrow \infty$, the attractiveness is essentially zero in the sightline of other fireflies. This is equivalent to the case where the fireflies randomly roam throughout a very thick foggy region. No other fireflies are visible and each firefly roams in a completely random fashion. This case corresponds to a completely random search method.

As the FA operates between these two extremes, it is possible to adjust the parameters α and γ so that the FA can outperform both the random search and the enhanced PSO algorithms. Furthermore, the FA can find both the global optima as well as the local optima concurrently which holds computational and efficiency advantages for MGA purposes (Yeomans, Gunalay, 2011). Another additional advantage of the FA for MGA implementation is that different fireflies essentially work independently of each other and FA are thus better than genetic algorithms and PSO for MGA because the fireflies can aggregate more closely around each local optimum.

MAIN FOCUS OF THE ARTICLE

Most mathematical programming algorithms originating in planning research literature have focused almost entirely on producing optimal solutions for single-objective formulations or, equivalently, generating noninferior solution sets for multi-objective problems (Loughlin *et al.*, 2001; Zechman, Ranjithan, 2004). While such methods may efficiently generate solutions to the derived mathematical models, whether their results actually establish “best” approaches to the underlying “real” problems is debatable.

In most “real world” decision-making, there are numerous system requirements and objectives that are never incorporated into the modeling, or even apparent, when the problems are being initially formulated (Brugnach *et al.*, 2007; Walker *et al.*, 2003). Furthermore, it may never be possible to explicitly convey all of the subjective considerations because there are frequently numerous competing, adversarial, incompatible stakeholder groups. Therefore most subjective aspects remain unquantified and unmodelled in the construction of any corresponding decision models. This is a common occurrence in situations where the final decisions are constructed based not only upon clearly stated and modelled objectives, but also

upon fundamentally subjective, socio-economic and political goals and stakeholder preferences (Yeomans, Gunalay, 2011). Numerous “real world” examples of this type of incongruent modelling duality are described in the literature (Baugh, Caldwell, Brill, 1997; Brill, Chang, Hopkins, 1982; Loughlin *et al.*, 2001; Zechman, Ranjithan, 2004).

Modelling to Generate Maximally Different Alternatives

When unmodelled objectives and unquantified issues exist, different approaches are required in order to not only search the decision space for the noninferior set of solutions, but also to explore the decision space for *inferior* alternative solutions to the modelled problem. In particular, any search for good alternatives to problems known (or suspected) to contain unmodelled objectives must focus not only on the non-inferior solution set, but also necessarily on an exploration of the problem’s inferior region.

To illustrate the implications of an unmodelled objective on a decision search, assume that the optimal solution for a quantified, single-objective, maximization decision problem is X^* with corresponding objective value ZI^* . Now suppose that there exists a second, unmodelled, maximization objective $Z2$ that perhaps subjectively reflects environmental/political acceptability. Let the solution X^a , belonging to the noninferior, 2-objective set, represent a potential best compromise solution if both objectives could somehow have been simultaneously evaluated by the decision-maker. While X^a might be viewed as the best compromise solution to the real problem, it would clearly appear inferior to the solution X^* in the quantified model, since it must be the case that $ZI^a \leq ZI^*$.

This observation implies that when unmodelled objectives are factored into the decision making process, mathematically inferior solutions for modelled problems can potentially be optimal for the real problems. Therefore, when unmodelled objectives and unquantified issues might exist,

different approaches are required in order to not only search the decision space for the noninferior set of solutions, but also to simultaneously explore the decision space for inferior alternative solutions to the modelled problem. Population-based procedures such as the **FA** permit concurrent searches throughout a feasible region and thus prove to be particularly adept methods for searching through a problem’s decision space.

The primary motivation behind MGA is to produce a manageably small set of alternatives that are quantifiably good with respect to known modelled objective(s) yet are **maximally different** from each other within the decision space. In so doing, the resulting solution set is likely to provide truly different alternatives that all perform somewhat similarly with respect to the modelled objective(s) yet very differently with respect to any unmodelled issues. By generating these maximally different solutions, the decision-makers can explore alternatives that may satisfy the unmodelled objectives to varying degrees of stakeholder acceptability. Obviously the solution-setters must then conduct a subsequent comprehensive comparison of the alternatives to determine which options would most closely satisfy their very specific circumstances. Thus, an **MGA** approach should necessarily be considered as one of decision support rather than of explicit solution determination.

In order to properly motivate an **MGA** search procedure, it is necessary to provide a more formal definition of the goals of maximal difference in the MGA process (Loughlin *et al.*, 2001; Yeomans, Gunalay, 2011). Suppose the optimal solution to an original mathematical model is X^* with objective value $Z^* = F(X^*)$. Then the following maximal difference model can be solved to determine an alternative solution that is **maximally different** from X^* :

$$\text{Max } \Delta = \sum_i |X_i - X_i^*|$$

Subject to: $X \in D$

$$|F(X) - Z^*| \leq T$$

where Δ is a difference function (shown as a simple absolute difference in this instance), D is the original mathematical model's feasible domain and T is a target deviation specified in relation to the original optimal function value Z^* . T is a user-supplied value that represents how much of the inferior region is to be explored for alternative solutions.

The FA-based MGA procedure is designed to generate a small number of maximally different alternatives by iteratively adjusting the value of T and using the FA to solve the corresponding, new maximal difference problem instance. Each desired solution alternative undergoes the common search procedure of the FA. The survival of solutions depends upon how different they are from all of the other previously generated alternatives in the decision space and by how well the solutions perform with respect to the modelled objective(s).

Computational Testing of the Firefly Algorithm Used For MGA

As described above, planners generally prefer to be able to select from a set of “near-optimal” alternatives that differ significantly from each

other in terms of the system structures characterized by their decision variables. In order to create this set of alternative planning options, extra target constraints that varied the value of T were placed into the original model formulation in order to force the generation of solutions that were maximally different from the initial optimal solution. The application of this MGA solution creation approach will be demonstrated using a 100-peak multimodal optimization problem taken from Loughlin *et al.* (2001).

The mathematical formulation for the multimodal test problem of Loughlin *et al.* (2001) is:

$$\begin{aligned} \text{Maximize } F(x, y) &= \sin(19\pi x) + \frac{x}{1.7} + \sin(19\pi y) \\ &+ \frac{y}{1.7} + 2 \\ 0.0 &\leq x \leq 1.0 \end{aligned}$$

$$0.0 \leq y \leq 1.0$$

The feasible region corresponding to this problem contains 100 peaks separated by valleys with the increasing amplitudes for both the peaks and valleys as the values of the decision variables

Table 1. Objective values and solutions for the 11 maximally different alternatives

Increment	1% Increment Between Alternatives			2.5% Increment Between Alternatives		
	$F(x,y)$	x	y	$F(x,y)$	x	y
Optimal	5.14	0.97	0.97	5.14	0.97	0.97
Alternative 1	5.10	0.98	0.97	5.01	0.87	0.87
Alternative 2	5.05	0.87	0.98	4.89	0.66	0.87
Alternative 3	5.00	0.76	0.98	4.77	0.87	0.45
Alternative 4	4.99	0.98	0.87	4.65	0.33	0.97
Alternative 5	4.91	0.98	0.76	4.50	0.98	0.02
Alternative 6	4.89	0.55	0.97	4.43	0.02	0.98
Alternative 7	4.89	0.98	0.55	4.29	0.98	0.02
Alternative 8	4.74	0.34	0.98	4.13	0.02	0.99
Alternative 9	4.69	0.98	0.24	4.02	0.99	0.02
Alternative 10	4.64	0.13	0.98	3.87	0.01	0.98

increase from their lower bounds of (0,0) toward their upper limits at (1,1). For the design parameters employed in this specific problem formulation, the mathematically optimal solution of $F(x, y) = 5.146$ occurs at point $(x, y) = (0.974, 0.974)$ (Loughlin *et al.*, 2001). The incremental MGA difference approach outlined in the previous section was used to produce the optimal solution and the 10 maximally different solutions shown in Table 1.

As outlined earlier, many “real world” problem applications tend to be riddled with incongruent performance requirements that are very difficult to quantify. Consequently, it is preferable to create several quantifiably good alternatives that concurrently provide very different perspectives to the potentially unquantifiable and unmodelled performance design issues during the policy formulation stage. The unique performance features captured within these dissimilar alternatives can result in very different system performance with respect to the unmodelled issues, thereby incorporating the unmodelled issues into the actual solution process. These two examples have demonstrated how an MGA modelling perspective can be used to generate multiple, good alternatives via the very computationally efficient FA that satisfy required system performance criteria according to prespecified bounds and yet remain as maximally different from each other as possible in the decision space. In addition to its alternative generating capabilities, the MGA procedure has simultaneously performed exceedingly well with respect to its role in function optimization. It should be explicitly noted that the overall best solution produced by the MGA procedure for the test problem is indistinguishable from the one determined in Loughlin *et al.* (2001).

FUTURE RESEARCH DIRECTIONS

In this chapter, an algorithm was presented that illustrated how the computationally efficient FA could be employed to generate numerous,

maximally different, near-best solutions. In this MGA role, the FA produces multiple alternatives possessing the required problem features, with each generated solution providing a different perspective. Since FA techniques have been applied to a wide spectrum of problem types, the practicality of this MGA approach can clearly be extended into numerous different planning environments. One such area of application would be to problems which contain significant degrees of stochastic uncertainty. Stochastic MGA approaches have only recently been considered within the research literature (Imanirad, Yang, Yeomans, 2016; Yeomans, Imanirad, Yang, 2014). Furthermore, although the FA provides one highly-efficient, computational procedure, there are numerous other nature-inspired metaheuristics that could be similarly employed. The computational efficacy of these extensions will provide the foundation for numerous future research studies.

CONCLUSION

Decision-making in the “real world” can be very complicated and frequently influenced by numerous unquantifiable issues, unmodelled objectives and uncertain factors. These uncertain and competing components oblige the decision-makers to assimilate several incompatible elements into the solution process prior to determining a final decision resolution. With so much uncertainty, it is doubtful that any single solution could ever be constructed that concurrently fulfills all of the incongruent system requirements. Thus, any ancillary modelling that supports the decision process needs to simultaneously consider all of these features while being malleable enough to capture the influences of any inherent planning uncertainty. FA-based MGA supplies an appropriate computational approach to support such “real world” decision-making applications.

REFERENCES

- Baugh, J. W. Jr, Caldwell, S. C., & Brill, E. D. Jr. (1997). A mathematical programming approach for generating alternatives in discrete structural optimization. *Engineering Optimization*, 28(1), 1–31. doi:10.1080/03052159708941125
- Brill, E. D. Jr, Chang, S. Y., & Hopkins, L. D. (1982). Modelling to generate alternatives: The HSJ approach and an illustration using a problem in land use planning. *Management Science*, 28(3), 221–235. doi:10.1287/mnsc.28.3.221
- Brugnach, M., Tagg, A., Keil, F., & De Lange, W. J. (2007). Uncertainty matters: Computer models at the science-policy interface. *Water Resources Management*, 21(7), 1075–1090. doi:10.1007/s11269-006-9099-y
- Gunalay, Y., & Yeomans, J. S. (2012). (in press). Generating alternatives using simulation-optimization combined with niching operators to address unmodelled objectives in a waste management facility expansion planning case. *International Journal of Operations Research and Information Systems*.
- Gunalay, Y., Yeomans, J. S., & Huang, G. H. (2012). Modelling to generate alternative policies in highly uncertain environments: An application to municipal solid waste management planning. *Journal of Environmental Informatics*, 19(2), 58–69.
- Imanirad, R., Yang, X.-S., & Yeomans, J. S. (2016). Stochastic Decision-Making in Waste Management Using a Firefly Algorithm-Driven Simulation-Optimization Approach for Generating Alternative. In S. Koziel, L. Leifsson, & X.-S. Yang (Eds.), *Recent Advances in Simulation-Driven Modeling and Optimization* (pp. 299–323). Heidelberg, Germany: Springer. doi:10.1007/978-3-319-27517-8_12
- Imanirad, R., & Yeomans, J. S. (2014). Biologically-Inspired Optimization Algorithms for Modelling to Generate Alternatives. In J. Wang (Ed.), *Encyclopedia of Business Analytics and Optimization* (pp. 373–383). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-5202-6.ch033
- Janssen, J. A. E. B., Krol, M. S., Schielen, R. M. J., & Hoekstra, A. Y. (2010). The effect of modelling quantified expert knowledge and uncertainty information on model based decision making. *Environmental Science & Policy*, 13(3), 229–238. doi:10.1016/j.envsci.2010.03.003
- Loughlin, D. H., Ranjithan, S. R., Brill, E. D. Jr, & Baugh, J. W. Jr. (2001). Genetic algorithm approaches for addressing unmodeled objectives in optimization problems. *Engineering Optimization*, 33(5), 549–569. doi:10.1080/03052150108940933
- Matthies, M., Giupponi, C., & Ostendorf, B. (2007). Environmental decision support systems: Current issues, methods and tools. *Environmental Modelling & Software*, 22(2), 123–127. doi:10.1016/j.envsoft.2005.09.005
- Mowrer, H. T. (2000). Uncertainty in natural resource decision support systems: Sources, interpretation and importance. *Computers and Electronics in Agriculture*, 27(1-3), 139–154. doi:10.1016/S0168-1699(00)00113-7
- Walker, W. E., Harremoes, P., Rotmans, J., Van der Sluis, J. P., Van Asselt, M. B. A., Janssen, P., & Krayen von Krauss, M. P. (2003). Defining uncertainty – a conceptual basis for uncertainty management in model-based decision support. *Integrated Assessment*, 4(1), 5–17. doi:10.1076/1413-4155(03)00011-1
- Yang, X. S. (2009). Firefly algorithms for multimodal optimization. *Lecture Notes in Computer Science*, 5792, 169–178. doi:10.1007/978-3-642-04944-6_14

Yang, X. S. (2010). *Nature-inspired metaheuristic algorithms* (2nd ed.). Frome, UK: Luniver Press.

Yeomans, J. S. (2011). Efficient generation of alternative perspectives in public environmental policy formulation: Applying co-evolutionary simulation-optimization to municipal solid waste management. *Central European Journal of Operations Research*, 19(4), 391–413. doi:10.1007/s10100-011-0190-y

Yeomans, J. S. (2012). Co-evolutionary simulation-driven optimization for generating alternatives in waste management facility expansion planning. *Journal of Computational Methods in Sciences and Engineering*, 12(1/2), 111–127.

Yeomans, J. S., & Gunalay, Y. (2011). Simulation-optimization techniques for modelling to generate alternatives in waste management planning. *Journal of Applied Operational Research*, 3(1), 23–35.

Yeomans, J. S., Imanirad, R., & Yang, X.-S. (2014). A Stochastic Nature-Inspired Approach for Modelling-to-Generate-Alternatives. *Journal of Applied Operational Research*, 6(1), 30–38.

Yeomans, J. S., & Yang, X.-S. (2014). Municipal Waste Management Optimization Using A Firefly Algorithm-Driven Simulation-Optimization Approach. *International Journal of Process Management and Benchmarking*, 4(4), 363–375. doi:10.1504/IJPMB.2014.065518

Zechman, E. M., & Ranjithan, S. R. (2004). An evolutionary algorithm to generate alternatives (EAGA) for engineering optimization problems. *Engineering Optimization*, 36(5), 539–553. doi:10.1080/03052150410001704863

ADDITIONAL READING

Baetz, B. W., Pas, E. I., & Neebe, A. W. (1990). Generating alternative solutions for dynamic programming-based planning problems. *Socio-Economic Planning Sciences*, 24(1), 27–34. doi:10.1016/0038-0121(90)90025-3

Chang, S. Y., Brill, E. D., & Hopkins, L. D. (1982). Efficient random generation of feasible alternatives: A land use example. *Journal of Regional Science*, 22(3), 303–313. doi:10.1111/j.1467-9787.1982.tb00754.x

Chang, S. Y., Brill, E. D. Jr., & Hopkins, L. D. (1982). Use of mathematical models to generate alternative solutions to water resources planning problems. *Water Resources Research*, 18(1), 58–64. doi:10.1029/WR018i001p00058

De Kok, J. L., & Wind, H. G. (2003). Design and application of decision support systems for integrated water management; Lessons to be learnt. *Physics and Chemistry of the Earth*, 28(14-15), 571–578. doi:10.1016/S1474-7065(03)00103-7

Gidley, J. S., & Bari, M. F. (1986). Modelling to generate alternatives. *ASCE Water Forum*, 86, 1366–1373.

Gunalay, Y., & Yeomans, J. S. (2011). An evolutionary simulation-optimization approach for addressing unmodelled objectives in waste management facility expansion planning. *Lecture Notes in Management Science*, 3, 207–221.

Gunalay, Y., & Yeomans, J. S. (2011). An efficient modelling-to-generate-alternatives approach for addressing unmodelled issues and objectives in public environmental planning. *Asian Journal of Information Technology*, 10(3), 122–128. doi:10.3923/ajit.2011.122.128

Hipel, K., & Ben-Haim, Y. (1999). Decision making in an uncertain world: Information-gap modeling in water resources management. *IEEE Transactions on Systems, Man and Cybernetics. Part C, Applications and Reviews*, 29(4), 506–517. doi:10.1109/5326.798765

Huang, G. H., Baetz, B. W., & Patry, G. G. (1996). A grey hop, skip and jump method for generating decision alternatives: Planning for the expansion/utilization of waste management facilities. *Canadian Journal of Civil Engineering*, 23, 1207–1219. doi:10.1139/196-930

- Imanirad, R., Yang, X.-S., & Yeomans, J. S. (2012). A computationally efficient, biologically-inspired modelling-to-generate-alternatives method. *Journal on Computing*, 2(2), 43–47.
- Imanirad, R., Yang, X.-S., & Yeomans, J. S. (2012). Nature-inspired algorithm for the concurrent generation of alternatives. *Journal on Computing*, 2(3), 101–106.
- Imanirad, R., Yang, X.-S., & Yeomans, J. S. (2013). Modelling-to-Generate-Alternatives Via the Firefly Algorithm. *Journal of Applied Operational Research*, 5(1), 14–21.
- Imanirad, R., Yang, X.-S., & Yeomans, J. S. (2013). A Biologically-Inspired Metaheuristic Procedure for Modelling-to-Generate-Alternatives. *International Journal of Engineering Research and Applications*, 3(2), 1677–1686.
- Imanirad, R., Yang, X.-S., & Yeomans, J. S. (2017). Environmental Decision-Making Under Uncertainty Using a Biologically-Inspired Simulation-Optimization Algorithm for Generating Alternative Perspectives. *International Journal of Business Innovation and Research*, 11(1), 38–59. doi:10.1504/IJBIR.2016.077609
- Liebman, J. C. (1976). Some simple-minded observations on the role of optimization in public systems decision-making. *Interfaces*, 6(4), 102–108. doi:10.1287/inte.6.4.102
- Linton, J. D., Yeomans, J. S., & Yoogalingam, R. (2002). Policy planning using genetic algorithms combined with simulation: The case of municipal solid waste. *Environment and Planning. B, Planning & Design*, 29(5), 757–778. doi:10.1068/b12862
- MacDonald, M. (1996). Bias issues in the utilization of solid waste indicators. *Journal of the American Planning Association*, 62(2), 236–242. doi:10.1080/01944369608975687
- Rubenstein-Montano, B., Anandalingam, G., & Zandi, I. (2000). A genetic algorithm approach to policy design for consequence minimization. *European Journal of Operational Research*, 124(1), 43–54. doi:10.1016/S0377-2217(99)00123-X
- Rubenstein-Montano, B., & Zandi, I. (1999). Application of a genetic algorithm to policy planning: The case of solid waste. *Environment and Planning. B, Planning & Design*, 26(6), 791–907. doi:10.1068/b260893
- Sprouse, W. L., & Mendoza, G. A. (1990). Modeling to generate alternatives: A Shawnee National Forest example. *Computers, Environment and Urban Systems*, 14(3), 203–211. doi:10.1016/0198-9715(90)90009-I
- Xiao, N., Bennett, D. A., & Armstrong, M. P. (2002). Using evolutionary algorithms to generate alternatives for multiobjective site-search problems. *Environment & Planning A*, 34(4), 639–656. doi:10.1068/a34109
- Yeomans, J. S. (2002). Automatic generation of efficient policy alternatives via simulation-optimization. *The Journal of the Operational Research Society*, 53(11), 1256–1267. doi:10.1057/palgrave.jors.2601439
- Yeomans, J. S. (2008). Applications of simulation-optimization methods in environmental policy planning under uncertainty. *Journal of Environmental Informatics*, 12(2), 174–186. doi:10.3808/jei.200800135
- Yeomans, J. S., & Huang, G. H. (2003). An evolutionary grey, hop, skip, and jump approach: Generating alternative policies for the expansion of waste management facilities. *Journal of Environmental Informatics*, 1(1), 37–51. doi:10.3808/jei.200300005

Yeomans, J. S., Huang, G. H., & Yoogalingam, R. (2003). Combining simulation with evolutionary algorithms for optimal planning under uncertainty: An application to municipal solid waste management planning in the regional Municipality of Hamilton-Wentworth. *Journal of Environmental Informatics*, 2(1), 11–30. doi:10.3808/jei.200300014

Yeomans, J. S., & Imanirad, R. (2012). Modelling to generate alternatives using simulation-driven optimization: An application to waste management facility expansion planning. *Applications of Mathematics*, 3(10), 1236–1244.

KEY TERMS AND DEFINITIONS

Firefly Algorithm: The Firefly Algorithm is a computationally efficient, nature-inspired, population-based metaheuristic that derives its solution approach based upon the characteristics of fireflies.

Heuristics: Approximation schemes used in problem-solving to generate good, though not necessarily optimal, solutions to mathematical programming problems.

Maximally Different Solutions: “Good” solution alternatives should possess near-optimal objective measures with respect to all of the known modelled objectives, but be fundamentally different from each other in terms of the system structures characterized by their decision variables. A difference model is employed to generate alternatives that are as far apart in the decision space as possible. The resulting alternative solution set of MGA provides disparate choices that all

perform well with respect to the known modelled objectives, yet very differently with respect to any unknown, unmodelled and/or unquantified issues. Hence, these solutions will provide entirely different perspectives to the original problem.

Meta-Heuristics: High-level, overarching heuristic approaches that have wide-ranging applicability to many different mathematical programming problems.

Modelling to Generate Alternatives (MGA): A modelling approach to systematically provide a set of “good” alternatives with respect to all of the problem’s modelled objectives. The primary motivation for MGA is to produce a manageably small set of alternatives that are good with respect to the known modelled objectives yet as different as possible from each other in the decision space – namely the solution set should provide maximally different alternatives.

Nature-Inspired Meta-Heuristics: Meta-heuristics whose fundamental solution characteristics have been motivated by phenomena occurring in the natural environment.

Stochastic MGA: Modelling to Generate Alternatives in which some or all of the parameters, objectives, constraints and/or other problem characteristics are expressed in some form of uncertainties, probability distributions or some other stochastic representation.

Unmodelled Issues: Aspects of a mathematical problem not captured during the construction and formulation of its corresponding mathematical/computer model.

Unquantified Issues: System objectives and requirements that are neither explicitly apparent nor included in the problem formulation stage—it is impossible to express these aspects quantitatively.

Preferences, Utility, and Stochastic Approximation

Yuri P. Pavlov

Bulgarian Academy of Sciences, Institute of Information and Communication Technologies, Bulgaria

Rumen D. Andreev

Bulgarian Academy of Sciences, Institute of Information and Communication Technologies, Bulgaria

INTRODUCTION

A complex system with human participation like “human-process” is characterized with active assistance of the human in the determination of its objective and description, and in decision-taking during its development. The construction of a mathematically grounded model of such a system is faced with the problem of shortage of mathematical precise information that presents the human activity. Often, in complex processes there is a lack of measurements or even clear scales for the basic heuristic information. On this level of investigations the decisions is close to the art to choose the right decision among great number of circumstances and often without associative examples of similar activity. The basic common source of information here are the expressed human preferences. A solution of this problem is to seek interpretation and expression of different aspects of the complex system through expert analysis and description of the expert’s preferences as an element of the system.

The presentation of human preferences analytically with utility functions is a good possible approach for their mathematical description. It is the first step in realization of a human-centered value driven design process and decision making, whose objective is to avoid the contradictions in human decisions and to permit mathematical calculations. In this approach the human being has the role of decision maker (DM). If the subjective and probability uncertainty of DM preferences is

interpreted as some stochastic noise, stochastic programming can be used for recurrent evaluation of the utility function in the sense of von Neumann with noise (uncertainty) elimination. In fact this is pattern recognition of the positive and negative DM’s answers in regarding to his preferences expressed as comparisons between lotteries in the gambling approach. The utility evaluation is human-computer dialog between a decision maker and computer-based evaluation tool. It concerns mathematically machine learning, since its basis is the axiomatic approach to decision making theory and stochastic approximation. The latter presents unique stochastic recurrent procedures suitable for computer programming.

The American psychologists Griffiths and Tenenbaum (2006) by analyzing intuitive evaluations in the conditions of repetitive life situations have proved the statistical optimality of human assessment. The major idea is that the new data is interpreted in the framework of a probability model built in their consciousness. That means that the Bayesian approach is a natural basis on which human beings formed their decisions, using their previous empirical experience. In such case the *utility theory* and its prescription to make decision based on value (utility) model as mathematical representation of the preferences has another scientific validation of the axiomatic approach in *decision making*. This modeling and its implementation in design is one of the directions of *value driven design*. The latter is a system engineering strategy, which enables multidisciplinary

design optimization by providing designers with an objective function.

People's preferences contain uncertainty of probabilistic nature due to qualitative type both of the empirical expert information and human notions. A possible approach for solution of these problems is *stochastic programming* (Aizerman et al., 1970). The uncertainty of the subjective preferences could be taken as an additive noise that could be eliminated, as is typical in the stochastic approximation procedures and *machine-learning* based on the stochastic programming.

The objective of the article is to present an innovative approach to value driven modeling of management (control) that bases on preference-oriented decision making. It is described a decision technology that realizes measurement (value-based evaluation) of human's objective-oriented preferences as analytic utility function. The latter is used in human-centered modeling of management/control that bases on the value-oriented determination of requirements and preferences of a human being in representation of complex processes. The utility theory and stochastic approximation are possible solution of this problem that results in a value-based approach to modeling of complex systems. It is demonstrated the application of the described value driven approach to management modeling in design of portfolio optimal control.

BACKGROUND

In complex processes and situations, there is a lack of measurements or even clearly identifiable scales for the basic heuristic information. Internal human expectations and heuristic are generally expressed by qualitative preferences. That is why the common sources of information in such a basic level are the human preferences. Probability theory, stochastic programming and expected utility theory address decision making under these conditions (von Neumann & Morgenstern, 1953; Raiffa, 1968; Pfanzagl, 1970; Fishburn, 1970). The mathematical description on such a fundamental

level requires basic mathematical terms like sets, relations and operations over them. In this aspect of mathematical descriptions it is entered the theory of measurements and scaling and utility theory.

From practical point of view the empirical system of human preferences relations is an algebraic system with relations. *System with relations* (SR) is called the set A in conjunction with a set of relations $R_i, i \in I, I = \{1, 2, 3, \dots, n\}$ defined over it and denoted by $(A, (R_i), i \in I)$. In this manner it is introduced a structure in the set A (Pfanzagl, 1970). Relation of *congruency* is called a relation of equivalency ($\approx$) (reflexive, symmetric and transitive) defined over the basic set A , if the property of *substitution* is satisfied, i.e. from the fulfillment of relations $(x_1, x_2, x_3, \dots, x_{hi}) \in A^{hi}$ and $(x_j \approx y_j)$ for every $j=1, 2, 3, 4, \dots, h_i$ it follows that $R_i(x_1, x_2, x_3, \dots, x_{hi}) = R_i(y_1, y_2, y_3, \dots, y_{hi})$ for $\forall i, i \in I$. By definition the relation of equivalency ($\approx_1$) is coarser than the equivalency ($\approx_2$), if the inclusion $(\approx_1) \subseteq (\approx_2)$ is satisfied. It is known that there exists the coarsest relation ($\approx_A$) over SR $(A, (R_i), i \in I)$. This means that if two elements are in congruency relation $(x \approx_A y)$, then they are undistinguishable, with respect to the properties, in the set A , described with the set of relations $((R_i), i \in I)$. If the set A is factorized by the coarsest congruency ($\approx_A$), then in the factor set $A/\approx_A$ the congruency ($\approx_A$) is in fact equality ($=$). A SR $(A, (R_i), i \in I)$, in which the congruency ($\approx_A$) is the coarsest is called *irreducible*. In this case SR $(A/\approx_A, (R_i), i \in I)$ is irreducible.

A homomorphism is an image $f, f: A \rightarrow B$ between two SR $(A, (R_i), i \in I)$ and $(B, (S_i), i \in I)$ from the same type, for which $\forall i, i \in I$ and $(x_1, x_2, x_3, \dots, x_{hi}) \in R_i$ is satisfied $R_i(x_1, x_2, x_3, \dots, x_{hi}) \Leftrightarrow S_i(f(x_1), f(x_2), f(x_3), \dots, f(x_{hi}))$.

DEFINITION: Every homomorphism from irreducible empirical system into the number system SR $(A, (Q_i), i \in I)$ is called k -dimensional scale.

The empirical system of relations SR is an object from the reality with the properties described by the relations in SR, while the numbered system

of relations SR is a mathematical object which reflects the interesting properties in the real object.

In the definition for the scale the correspondence $f_\theta: A \rightarrow R^k$ is not simply defined (R^k is the k -ary Cartesian product of the set of the real numbers). In general sense, there exists entire class of scales converting the irreducible empirical system of relations SR $(A, (R_i), i \in I)$ into the number system SR $(R^k, (S_i), i \in I)$. This class of homomorphism (scales) is denoted by $\mathfrak{N}(A, R^k)$ (every homomorphism is injective in its essence, because the empirical system is irreducible and surjective with regard to $f(A)$). Let A_0 be a subset of A . We denote by $G_A(A_0)$ all injective homomorphism, inclusion (*partial endomorphism*) from SR $(A_0, (R_i), i \in I)$ in SR $(A, (R_i), i \in I)$. If a scale $f_\theta \in \mathfrak{N}(A, R^k)$ is given, then we can characterize the whole class of scales $\mathfrak{N}(A, R^k)$ in the following way: $\mathfrak{N}(A, R^k) = \{\gamma \circ f_\theta / \text{where } \gamma \in G_R^k(f_\theta(A))\}$. Or in other words two scales are equivalent with precision to a partial endomorphism $\gamma \in G_R^k(f_\theta(A))$. The elements of $G_R^k(f_\theta(A))$ are called *admissible manipulations* of the scale f_θ . An example is the measurement of the temperature. The scale $f_\theta(\cdot)$ is the temperature in Celsius. Then every partial endomorphism in this example $\gamma \in G_R^k(f_\theta(A))$ is an affine correspondence of the type $\gamma(x) = ax + b$, $a \in R$, $b \in R$ and $a > 0$, measurements done in Kelvin for example.

From the definition of the measurement and scale it follows that there are infinitely many types of scales. In informal terms measurement is an operation in which a given state of the observed object is mapped to a given denotation. An example of basic scale is the so-called *nominal scale* which is an expression of the equivalence of two phenomena only. Let X is the set of alternatives ($X \subseteq R^m$). Let x and y be two alternatives $((x, y) \in X^2)$. For this weakest scale the following axioms are valid:

1. $((x \approx y \vee \neg x \approx y) \equiv 1) \wedge ((x \approx y \wedge \neg x \approx y) \equiv 0) \wedge x \approx x$;
2. $(x \approx y \Rightarrow y \approx x)$;
3. $((x \approx y \wedge x \approx z) \Rightarrow y \approx z)$.

Here $(\approx)$ denotes equivalence and $\neg(\approx)$ is the opposite (non-equivalence). In this scale only the Kronecker symbol may be used as a measure.

In the case when the observed phenomenon allows not only to distinguish the differences between states but to compare them by preference a stronger scale needs to be used – the *ordering scale*. The preference in the ordering scale is denoted by $(x \succ y)$. In accordance with a tradition, $x \succ y$ is taken to represent “ x is better than y ”. In this scale together with the above three axioms two more are satisfied:

4. $\neg(x \succ x)$ for $\forall x \in X$, $((x \succ y) \Rightarrow \neg(y \succ x))$;
5. $(x \succ y \wedge y \succ z) \Rightarrow x \succ z$.

If incomparable alternatives exist, then the scale is called a scale of *partial ordering*. Under these five axioms an analytical representation by value function $u(\cdot)$ is searched for. A “*value function*” is a function $u(\cdot)$ for which it is fulfilled $((x, y) \in X^2, x \succ y) \Leftrightarrow (u(x) > u(y))$. In this definition, in addition to axioms (4, 5), weak connectedness is also assumed $\neg(x \approx y) \Rightarrow ((y \succ x) \vee (x \succ y))$. The definition could be restricted to the following formula in the case of partial ordering $((x, y) \in X^2, x \succ y) \Rightarrow (u(x) > u(y))$. The assumption of existence of a value function $u(\cdot)$ leads to the *negatively transitive* and *asymmetric* relation $(\succ)$ (*weak order*). Consequently the assumption of existence of a value function $u(\cdot)$ leads to the existence of: asymmetry $((x \succ y) \Rightarrow \neg(x \succ y))$, transitivity $((x \succ y) \wedge (y \succ z) \Rightarrow (x \succ z)$, and transitivity of the “indifference” relation $(\approx)$. It is proven in Fishburn (1970) that the existence of a “*weak order*” $(\succ)$ over X leads to the existence of a “*strong order*” over $X/\approx$ (“*strong order*” is a “*weak order*” for which is fulfilled $(\neg(x \approx y) \Rightarrow ((x \succ y) \vee (x \prec y)))$). Consequently the assumption of existence of a value function $u(\cdot)$ leads to the existence of: asymmetry $((x \succ y) \Rightarrow \neg(x \succ y))$, axiom 4), transitivity $((x \succ y) \wedge (y \succ z) \Rightarrow (x \succ z)$, axiom 5) and transitivity of the “indifference” relation $(\approx)$ (axiom 3).

Depending on the type of the function – continuous, partially continuous or discrete – there are

different types of scales measuring the above relations. A transformation with an arbitrary monotonous function (*admissible manipulations* $\gamma, \gamma \in G_B(f_0(A))$) of the ordinal scale leads to an ordinal scale. When using those scales, apart from comparison by magnitude, we can search the minimum and maximum of the function as feasible mathematical operations. In this manner the reader can move from the language of binary relations and preferences to the language of control criteria as objective value function.

If together with the ordering of the alternatives the distance between them can be evaluated, we can talk about *interval scale*. For these scales the distances between the alternatives have the meaning of real numbers. The mathematical expectation depends on the origin of the scale and thus is unfeasible for these scales. The transition ($\gamma \in G_B(f_0(A))$) from one interval scale to another is achieved with affine transformation $\mathbf{x} = \mathbf{a}\mathbf{y} + \mathbf{b}$, $(\mathbf{x}, \mathbf{y}) \in \mathbf{X}^2$, $\mathbf{a} > 0$, $\mathbf{b} \in \mathbf{R}$. Among these type of scales is also the measurement of the utility function by the “gambling approach” (Raiffa, 1968; Keeney & Raiffa, 1993). It is emphasized that here the calculations are done with numbers related to the distances between the alternatives.

A stronger scale is the *ratio scale*. This is an interval scale with fixed origin $\mathbf{x} = \mathbf{a}\mathbf{y} + \mathbf{b}$, $(\mathbf{x}, \mathbf{y}) \in \mathbf{X}^2$, $\mathbf{a} > 0$. Such is the scale for measuring weight. A utility function could be normed between 0 and 1 with a affine transformation and by this way transformed in *ratio scale*. For these scales the mathematical expectation is sensible evaluation and has physical meaning. For these scales in addition the following axioms are satisfied:

6. $(x = y \wedge z > 0) \Rightarrow ((x + z) > y)$;
7. $x + y = y + x$;
8. $(x = y \wedge z = q) \Rightarrow (x + z = y + q)$;
9. $q + (x + y) = (q + y) + x$.

The *absolute scale* is the most powerful. For it the zero and one are absolute and it is one of a kind and unique scale.

Depending on the way by which the empirical information for the preferences of the DM is derived the functions representing the DM's preferences $(x \succ y)$ as *value or utility* are defined (Keeney & Raiffa, 1993). Let $\mathbf{X}$ is the set of alternatives ($\mathbf{X} \subseteq \mathbf{R}^m$). A “value” function is a function ($u^*: \mathbf{X} \rightarrow \mathbf{R}$) for which it is fulfilled: $((x, y) \in \mathbf{X}^2, x \succ y) \Leftrightarrow (u^*(x) > u^*(y))$. A weaker case is when the following condition is fulfilled instead: $((x, y) \in \mathbf{X}^2, x \succ y) \Rightarrow (u^*(x) > u^*(y))$. For a finite set of alternatives and partial ordering (axioms 4, 5) there always exists such a function with precision up to monotonous transformation. In this manner from the language of binary relations and preferences is moved to the language of control criteria. Let $\mathbf{A}_u^*$ and $\mathbf{B}_u^*$ are the sets:

$$\mathbf{A}_{u^*} = \{(x, y) \in \mathbf{R}^{2m} / (u^*(x)) > u^*(y)\},$$

$$\mathbf{B}_{u^*} = \{(x, y) \in \mathbf{R}^{2m} / (u^*(x)) < u^*(y)\}.$$

If there is a function $\mathbf{F}(x, y)$ of the form $\mathbf{F}(x, y) = \mathbf{f}(x) - \mathbf{f}(y)$, positive over $\mathbf{A}_{u^*}$ and negative over $\mathbf{B}_{u^*}$, then the function $\mathbf{f}(x)$ is a value function equivalent to the empirical value function $u^*(.)$. Such approach permits the use of stochastic “pattern recognition” for solving the problem. In the deterministic case it is true that $\mathbf{A}_{u^*} \cap \mathbf{B}_{u^*} = \emptyset$. In the probabilistic case it is true that $\mathbf{A}_{u^*} \cap \mathbf{B}_{u^*} \neq \emptyset$ (Pavlov & Andreev, 2013).

From practical point of view the empirical system of relations is $\mathbf{SR}(\mathbf{X}, (\approx), (\succ))$, where $(\approx)$ can be considered as the relation “indifferent or equivalent”, and $(\succ)$ is the relation “prefer”. We will look for equivalency of the empirical system with the system of relations ($\mathbf{R}$ -real numbers, $(=)$, $(>)$). The “indifference” relation $(\approx)$ based on $(\succ)$ is defined by $((x \approx y) \Leftrightarrow \neg((x \succ y) \vee (y \succ x)))$.

MAIN FOCUS

Value based decision making bases on preferences representation by utility/value functions. As a direction of value driven design the normative ap-

proach and in particular the utility/value functions permit inclusion of the DM mathematically in the modeling and in the main objective function. It is desirable that the objective function presents all important attributes of a system being designed and outputs a score. At the whole system level, the objective function, which performs this assessment of value, is called a “value model”. In this aspect the Utility theory is the normative approach in decision making in the conditions of risk and uncertainty.

Utility Function and Measurement Scale

There are different systems of mathematical axioms that give sufficient conditions of a utility function existence. The most famous of them is the system of *von Neumann and Morgenstern's axioms* (Fishburn, 1970). According to *Utility theory* let X be the set of alternatives and P is a set of probability distributions over X and $X \subseteq P$. A utility function $u(\cdot)$ is any function, which fulfills the following:

$$(p \succsim q, (p, q) \in P^2) \Leftrightarrow (\int u(\cdot) dp \geq \int u(\cdot) dq).$$

The interpretation is that the integral of the utility function $u(\cdot)$ is a measure concerning the comparison of the probability distributions p and q defined over X . The notation $(\succsim)$ expresses the preferences of DM over P including those over $X (X \subseteq P)$. It is known that the assumption of existence of a value function $u(\cdot)$ leads to the “negatively transitive” and “asymmetric” relation $(\succsim)$ and to transitivity of the relation $(\approx)$. So far we are in the *ordering scale*. The assumption of equivalence with precision up to affine transformation has not been included. In other words there is only a value function. For value, however, the mathematical expectation is unfeasible, but we underline that the mathematical expectation is included in the definition of the utility function. For this reason it is accepted that $(X \subseteq P)$ and that P is a convex set $((q, p) \in P^2 \Rightarrow (\alpha q + (1-\alpha)p) \in P, \text{ for } \forall \alpha \in [0,1])$. Then utility $u(\cdot)$ is determined with the accuracy of an affine transformation (interval scale) (Fishburn, 1970):

$\forall \alpha \in [0,1]$. Then utility $u(\cdot)$ is determined with the accuracy of an affine transformation (interval scale) (Fishburn, 1970):

PROPOSITION 1: If $((x \in X \wedge p(x)=1) \Rightarrow p \in P)$ and $((q, p) \in P^2 \Rightarrow ((\alpha p + (1-\alpha)q) \in P, \alpha \in [0,1]))$ are realized, then the utility function $u(\cdot)$ is defined with precision up to an affine transformation: $(u_1(\cdot) \approx u_2(\cdot)) \Leftrightarrow (u_1(\cdot) = au_2(\cdot) + b, a > 0)$.

Now it is obvious why in practice the gambling approach is used to construct the Neumann utility in *interval scale*. The reason is that to be in the interval scale the set of the alternatives for comparison must be convex. A “lottery” is called every discrete probability distribution over X . It is denoted as $\langle x, y, \alpha \rangle$ the lottery: α is the probability of the appearance of the alternative x and $(1-\alpha)$ - the probability of the alternative y . In practice the utility function is measured by the gambling approach, comparisons between lotteries (Keeney & Raiffa, 1993). Weak points of this approach are violations of the transitivity of the relations and the so called “certainty effect” and “probability distortion” identified by Kahneman and Tversky (Cohen & Jaffray, 1988). The violations of the transitivity of the relation $(\approx)$ also leads to declinations in the utility assessment. All these difficulties explain the DM behavior observed in the Allais Paradox (Allais, 1953).

Stochastic Approximation and Utility Evaluation

It is assumed that $(X \subseteq P), ((q, p) \in P^2 \Rightarrow (\alpha q + (1-\alpha)p) \in P, \text{ for } \forall \alpha \in [0,1])$ and that the utility function $u(\cdot)$ exists. Two sets are used, motivated by Proposition 1:

$$A_{u^*} = \{(\alpha, x, y, z) / (\alpha u^*(x) + (1-\alpha)u^*(y)) > u^*(z)\},$$

$$B_{u^*} = \{(\alpha, x, y, z) / (\alpha u^*(x) + (1-\alpha)u^*(y)) < u^*(z)\}.$$

The notation $u^*(\cdot)$ in the sets A_{u^*} and B_{u^*} is the empirical DM's utility function. The utility

function is evaluated by recognizing the above sets with stochastic recurrent procedures (Pavlov & Andreev, 2013).

The following proposition is in the foundation of the used *stochastic approximation*.

PROPOSITION 2: Denote by A_{u^*} the set $A_u = \{(\alpha, x, y, z) / (\alpha u(x) + (1-\alpha)u(y)) > u(z)\}$. If $A_{u_1} = A_{u_2}$ and $u_1(\cdot)$ and $u_2(\cdot)$ are continuous functions than is true $(u_1(\cdot) = au_2(\cdot) + b, a > 0)$.

The approximation of the utility function is constructed by pattern recognition of the set A_u and is machine learning based on the DM's preferences. The machine learning is a probabilistic pattern recognition, because $(A_{u^*} \cap B_{u^*} \neq \emptyset)$ and the utility evaluation is a stochastic approximation with noise (uncertainty) elimination (Aizerman et al., 1970). The following presents the evaluation procedure: *The DM compares the "lottery" $\langle x, y, \alpha \rangle$ with the simple alternative $z, z \in Z$ ("better- $\setminus, f(x, y, z, \alpha) = 1$ ", "worse- $\setminus, f(x, y, z, \alpha) = -1$ " or "can't answer or equivalent- $\sim, f(x, y, z, \alpha) = 0$ ", $f(\cdot)$ denotes the qualitative DM's answer). This determines a learning point $((x, y, z, \alpha), f(x, y, z, \alpha))$. The following stochastic algorithm constructs the utility polynomial approximation $u(x) = \sum_i c_i \Phi_i(x)$:*

$$c_i^{n+1} = c_i^n + \gamma_n \left[f(t^{n+1}) - \overline{(c^n, \Psi(t^{n+1}))} \right] \Psi(t^{n+1}),$$

$$\sum_n \gamma_n = +\infty, \sum_n \gamma_n^2 < +\infty, \forall n, \gamma_n > 0.$$

The following notations are used: $t = (x, y, z, \alpha)$, $\psi_i(t) = \psi_i(x, y, z, \alpha) = \alpha \Phi_i(x) + (1-\alpha) \Phi_i(y) - \Phi_i(z)$, where $(\Phi_i(x))$ is a family of orthogonal polynomials. The coefficients c_i^n take part in the polynomial presentation $g^n(x) = \sum_{i=1}^n c_i^n \Phi_i(x)$ and

$$\begin{aligned} (c^n, \Psi(t)) &= \alpha g^n(x) + (1-\alpha)g^n(y) - g^n(z) \\ &= G^n(x, y, z, \alpha) \end{aligned}$$

is a scalar product. The line above $\bar{v} = \overline{(c^n, \Psi(t))}$ means: $(\bar{v} = 1)$, if $(v > 1)$, $(\bar{v} = -1)$ if $(v < -1)$ and $(\bar{v} = v)$ if $(-1 < v < 1)$. The following decomposition of the expert function $f(x, y, z, \alpha)$ is true:

$$f(t^{n+1}) = \left[D'(t^{n+1}) + \xi^{n+1} \right].$$

In the formula ξ denotes the uncertainty in the expert's answers. The DM relates intuitively the "learning point" (x, y, z, α) to the set A_u with probability $D_1(x, y, z, \alpha)$ or to the set B_u with probability $D_2(x, y, z, \alpha)$. The probabilities $D_1(x, y, z, \alpha)$ and $D_2(x, y, z, \alpha)$ are the mathematical expectation of $f(x, y, z, \alpha)$ over the set of positive answers- A_u and over the set of negative answers- B_u , respectively: $D_1(x, y, z, \alpha) = M(f/x, y, z, \alpha)$, if $M(f/x, y, z, \alpha) > 0$, $D_2(x, y, z, \alpha) = (-M(f/x, y, z, \alpha))$, if $M(f/x, y, z, \alpha) < 0$. Let $D'(x, y, z, \alpha)$ are the random value: $D'(x, y, z, \alpha) = D_1(x, y, z, \alpha)$, $M(f/x, y, z, \alpha) > 0$; $D'(x, y, z, \alpha) = (-D_2(x, y, z, \alpha))$, $M(f/x, y, z, \alpha) < 0$; $D'(x, y, z, \alpha) = 0$, $M(f/x, y, z, \alpha) = 0$. It is approximated $D'(x, y, z, \alpha)$ by a function of the type: $G(x, y, z, \alpha) = (\alpha g(x) + (1-\alpha)g(y) - g(z))$, $g(x) = \sum_i c_i \Phi_i(x)$ and the function $g(x)$ is an approximation of the utility $u(\cdot)$. The $G^n(x, y, z, \alpha)$ is positive over A_u and negative over B_u depending on the degree of approximation of $D'(x, y, z, \alpha)$. The convergence of the stochastic procedure is described in Pavlov & Andreev (2013). The learning points $((x, y, z, \alpha), f(x, y, z, \alpha))$ could be set with a pseudo random sequence.

SOLUTIONS AND RECOMMENDATIONS

Optimal Portfolio Utility Allocation

The proposed approach is demonstrated in a complex economical problem with human participation. The optimal portfolio allocation problem is discussed repeatedly in the economic literature (Touzi & Tourin, 2012). The construction of the objective utility function remains aside of discussions. In this example is assumed that the outcome set is a two-attribute product set $V \times W$, where V designates the first attribute - the amount π_t , $(X_t \pi_t, \pi_t \in [0, 1])$ invested in a risky process and W designates the quantity of money in BGN's. It is found that preferences on W do not depend on the particular deterministic level at which $v \in V$ is fixed. A convenient implication of this utility independence is that changing v does not affect rank-ordering in W . This determines the form of the two attribute utility $U(v, w) = U(v, w^0) + [U(v, w^*) - U(v, w^0)]U(v^0, w)$ (Keeney & Raiffa, 1993). Each of these single attribute utility functions is evaluated based on consumer's preferences (Figure 1). The ciseaux line is a sample of stochastic optimal control.

Consider a non-risky asset S_0 and a risky one S . The Black-Scholes dynamic model represents a stochastic differential equation determined by:

$$dS_t^0 = S_t^0 r dt \quad \text{and} \quad dS_t = S_t \mu dt + \sigma dW_t.$$

In the formula, μ and σ are constants ($r=0.03$, $\mu=0.05$ and $\sigma=0.3$) and W is one dimensional Brownian motion. By X_t it is denoted the state space vector of the controlled process. The investment policy is defined by a progressively adapted process $\pi = \{\pi_t, t \in [0, T]\}$, where π_t represents the risky amount $(X_t \pi_t, \pi_t \in [0, 1])$ invested at any moment t . The remaining wealth $(X_t - \pi_t X_t)$ is non-risky invested. The time period T is 50 weeks. The self-financing satisfies the stochastic differential equation (Touzi & Tourin, 2012):

$$\begin{aligned} dX_t^\pi &= \pi_t X_t^\pi \frac{dS_t}{S_t} dt + (X_t^\pi - \pi_t X_t^\pi) \frac{dS_t^0}{S_t^0} \\ &= (r X_t^\pi + (\mu - r) \pi_t X_t^\pi) + \sigma \pi_t X_t^\pi dW_t \end{aligned}$$

The amount $X_t^{x, \pi}$ is the solution of the stochastic differential equation with initial wealth (x) at time t . The main objective of the DM is maximization of the expected utility at the final moment T ,

$$V(t, x) := \sup_{\pi \in [0, 1]} E[U(X_T^{t, x, \pi})].$$

Figure 1. Utility objective function and optimal control

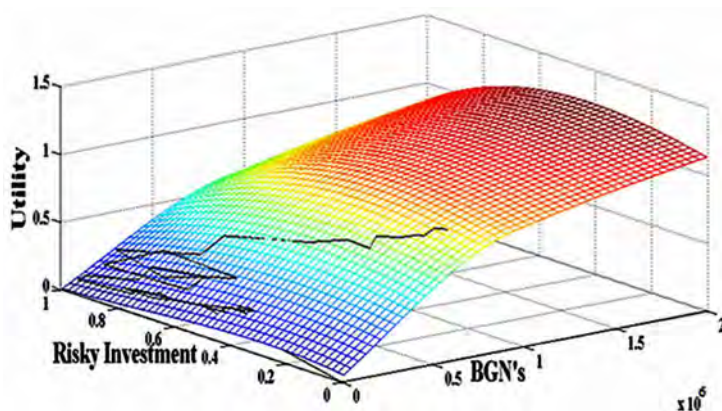


Figure 2. Hamilton-Jacobi-Bellman function

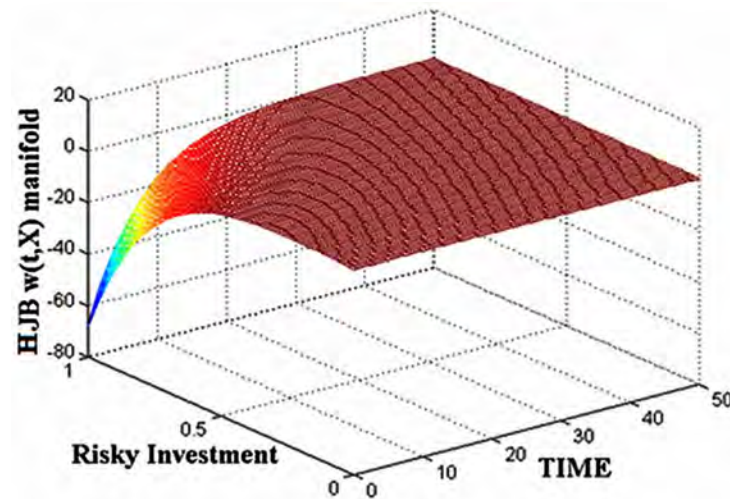
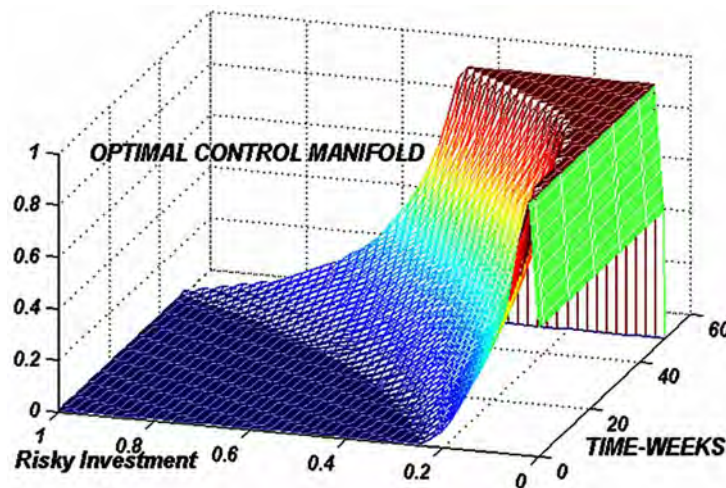


Figure 3. Control MANIFOLD



The control input is the amount π_t invested in the risky process. The optimal control is determined from the Hamilton-Jacobi-Bellman partial differential equation

$$\frac{\partial w}{\partial t}(t, x) + \sup_{\pi \in [0,1]} [(rx + (\mu - r)\pi x) \frac{\partial w}{\partial t}(t, x) + \frac{1}{2} \sigma^2 \pi^2 x^2 \frac{\partial^2 w}{\partial x^2}(t, x)] = 0,$$

following the dynamical programming principle. By passing through generalized solution of the

Black-Scholes stochastic differential equation it is determined a polynomial approximations of the Hamilton-Jacobi-Bellman (HJB) function $w(t, X)$. From HJB surface it is determined the control law manifold $\pi(t, X)$. They are shown in Figures 2 and 3. In Figure 4 it is shown the risk process without control.

The stochastic control is started in 30 different initial points; from 1000 BGN's to 30000 BGN's. In the next Figures 5 and 6, the processes with optimal stochastic control and a sample of 30 optimal control laws are shown.

Figure 4. Process without control

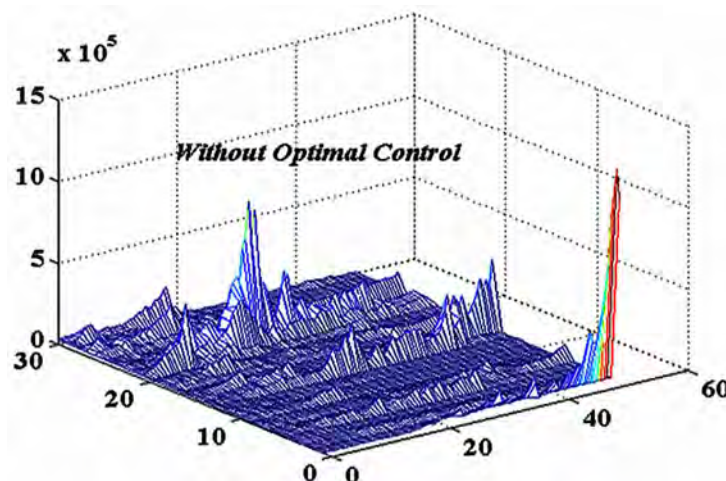
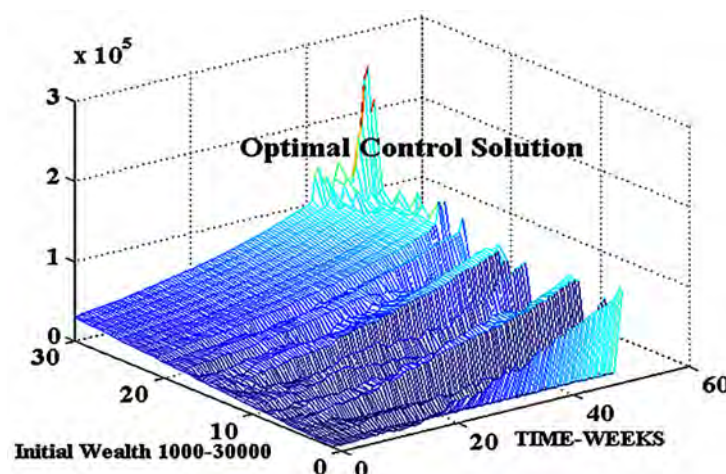


Figure 5. Process with control



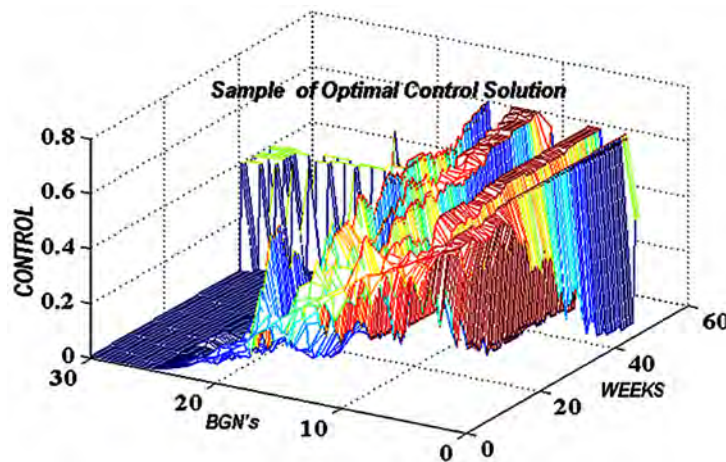
The optimal control is in agreement with the DM's preferences and is significant only for this DM.

FUTURE RESEARCH DIRECTIONS

The preferences representation of the DM is an essential stage in the construction of models used in decision theory and operations research. The publications concerning preferences utility representations and their applications in economic, medicine, ecology, engineering etc. quickly grew

in the last decade (La Red Martinez & Acosta, 2015; Wakker, 2010; Kangas, 2015). One of the main obstacles in this direction is the lack of mathematical algorithms suitable for computer realization. The proposed approach is a possible solution. A serious challenge in the utility theory application is the possibility for decomposition of the multi-attribute utility into single-attribute functions (Keeney & Raiffa, 1993). A perspective idea for a new interpretation of the notion „independence” and possibilities for utility decomposition is proposed by Abbas and Bell (2011).

Figure 6. Samples of optimal control



New achievement in this direction will facilitate considerably the application of the utility theory.

Several non-expected utility theories have been developed in response to the transitivity violations of the preference (Kahneman & Tversky, 1979; Wakker, 2010). Among these theories the rank dependent utility model and cumulative prospect theory are currently the most popular. In the rank dependent utility the decision weight of an outcome is not just probability associated with this outcome. It is a function of both probability and the rank of the alternative. The declination of the probability assessment, the probability distortion has a S-shaped form closed to symmetry. The mentioned peculiarities give a hint for investigations of the process of preferences evaluation as a function of both subjective probability (in the lottery) and rank of the alternative (Machina, 2009). In the classical theory and in the neo-Bayesian approach the utility function is determined in the sense of von Neumann, with precision up to affine transformation. The problem appears when the „state of natures” is connected with human appreciation described with subjective probabilities (Fishburn, 1970; Schmeidler, 1989). This theoretical findings permit deduction that the gambling utility evaluation remain valid, having in mind the certainty effect (Cohen & Jaffray, 1988). The proposed stochastic procedures permit

sufficiently precise evaluation of the utility functions as a function of both probability and rank (Pavlov & Andreev, 2014). The certainty effect is considerably eliminated due to the nature of the stochastic procedures and due to the S shaped form of the probability distortion. The uncertainty of subjective preferences could be regarded as a noise that is eliminated as typical for the stochastic approximation. Thus the polynomial approximation is approximation of the von Neumann utility defined in classical and neo Bayesian theory with precision up to affine transformation.

CONCLUSION

The suggested approach can be regarded as a realization of the prescriptive *decision making*. The utility function is an abstraction presented in the limits of the normative approach, the axiomatic systems of von Neumann. Here is revealed the existence of a mathematical expectation measured in the interval scale on the base of the DM's preferences. This mathematical expectation could be interpreted as an approximation of the Von Neumann's utility function.

This stochastic approach permits inclusion of the DM in the objective function mathematically and in the modeling of complex processes. The

analytical representation of the human preferences and the stochastic machine learning allow practitioners to take advantage of individual application of achievements of decision making theory in various fields of human activities.

REFERENCES

- Aizerman, M. A., Braverman, E., & Rozonoer, L. (1970). Potential Function Method in the Theory of Machine Learning. Moscow, Russia: Nauka. (in Russian)
- Allais, M. (1953). Le Comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'école américaine. *Econometrica*, 21(4), 503–546. doi:10.2307/1907921
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Bertoni, M., Eres, H., & Isaksson, O. (2011). Criteria for assessing the Value of Product Service System Design Alternatives: An Aerospace Investigation. *Functional Thinking for Value Creation*, 141–146.
- Castagne, S., Curran, R., & Collopy, P. (2009). Implementation of value-driven optimization for the design of aircraft fuselage panels. *International Journal of Production Economics*, 117(2), 381–388. doi:10.1016/j.ijpe.2008.12.005
- Cohen, M., & Jaffray, J.-Y. (1988). Certainty Effect Versus Probability Distortion: An Experimental Analysis of Decision Making Under Risk. *Journal of Experimental Psychology. Human Perception and Performance*, 14(4), 554–560. doi:10.1037/0096-1523.14.4.554
- Fishburn, P. (1970). *Utility Theory for Decision-Making*. New York: Wiley.
- Griffiths, T. L., & Tenenbaum, J. B. (2006). Optimal prediction in everyday cognition. *Psychological Science*, 17(9), 767–773. doi:10.1111/j.1467-9280.2006.01780.x PMID:16984293
- Hall, D., & Davis, R. (2007). Engaging multiple perspectives: A value-based decision making model. *Decision Support Systems*, 43(4), 1588–1604. doi:10.1016/j.dss.2006.03.004
- Keeney, R., & Raiffa, H. (1993). *Decision with Multiple Objectives: Preferences and Value Trade-offs*. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9781139174084
- La Red Martinez, L. D., & Acosta, J. C. (2015). Review of Modeling Preferences for Decision Models. *European Scientific Journal*, 11(36), 1857–7881.
- Pavlov, Y., & Andreev, R. (2014). Subjective Probability and Expected Utility: A Stochastic Approximation Evaluation. *Indian Journal of Applied Research*, 4(3), 491–497. doi:10.15373/2249555X/MAR2014/155
- Pavlov, Y. P., & Andreev, R. D. (2013). *Decision Control, Management, and Support in Adaptive and Complex Systems: Quantitative Models*. Hershey, PA: IGI Global. doi:10.4018/978-1-4666-2967-7
- Pfanzagl, J. (1971). *Theory of Measurement*. Wurzburg-Wien, Austria: Physical-Verlag. doi:10.1007/978-3-662-41488-0
- Raiffa, H. (1968). *Decision Analysis*. New York: Addison-Wesley Reading Mass.
- Von Neumann, J., & Morgenstern, O. (1953). *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press.

ADDITIONAL READING

- Abbas, A., & Bell, D. (2011). One-Switch Independence for Multiattribute Utility Functions. *Operations Research*, 59(3), 764–771. doi:10.1287/opre.1100.0911
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2), 263–291. doi:10.2307/1914185

Kangas, A. et al.. (2015). *Decision Support for Forest Management. Managing Forest Ecosystems*, 30. Springer. doi:10.1007/978-3-319-23522-6

Machina, M. (2009). Risk, Ambiguity, and the Rank-Dependence Axioms. *The American Economic Review*, 99(1), 385–392. doi:10.1257/aer.99.1.385

Shmeidler, D. (1989). Subjective probability and expected utility without additivity. *Econometrica*, 57(3), 571–587. doi:10.2307/1911053

Touzi, N., & Tourin, A. (2012). Optimal Stochastic Control, Stochastic Target Problems and Backward SDEs. *Business Economics (Cleveland, Ohio)*, 29.

Wakker, P. (2010). *Prospect Theory For Risk and Ambiguity*. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511779329

KEY TERMS AND DEFINITIONS

Decision Making: A complex and incompletely understood process which addresses the qualitative or intuitive aspects of human behavior and makes a claim for the potential of computer-based applications to a decision maker's qualitative cognitive processes.

Expected Utility Theory: It is the kernel of normative approach for answering of the question how people should rationally make choice under uncertainty.

Measurement Theory: A branch of applied and theoretical mathematics that is useful in measurement and data analysis.

Stochastic Programming: Subfield of mathematical programming that considers optimization in the presence of uncertainty.

Value Based Decision: Enable the assessment of a value for every design option so that options can be rationally compared and a choice taken. At the whole system level, the objective function which performs this assessment of value is called a “value model”.

Value-Driven Modelling: A modelling strategy that bases on an environment which enables optimization by providing designers with an objective function. The objective function inputs all the important attributes of the system being designed, and outputs a score.

Rough-Set-Based Decision Model for Incomplete Information Systems

Safiye Turgay

Sakarya University, Turkey

Orhan Torkul

Sakarya University, Turkey

Tahsin Turgay

Sakarya University, Turkey

INTRODUCTION

The increasing use of computers and information are getting necessity with increased expertise and digital data storage, that becomes compulsory and importance for usage. Rapidly increasing data traffic has revealed a major case for efficient manner with the use of the big data. Developed algorithm and suggested model can help using of the data in database structure and evaluate and process of the data for taking managerial decisions with rough sets approach. Rough sets derived from the fuzzy logic approach to perform the analysis of the incomplete information and uncertain structure. In 1982, this approach proposed by Pawlak, for discover the knowledge from big databases which is used in classification of the attributes recently. Also, Rough set is made as incomplete, inadequate and vague information organizing; it makes it suitable for data analysis (Pawlak, 1994, 1995).

Databases use this data and evaluate managerial decisions in the process of data mining has become imperative that we give the name of the emergence of the field. The rough set is a concept derived from the fuzzy logic approach to carry out the analysis of structures with uncertain data mining techniques. The decision will be developed in conjunction with computerized decision support model, giving more efficient automation systems with algorithms that are targeted. Our suggested decision support system covers the inputs, user

knowledge and expertise, outputs and decision components. In addition, data access, interactive mode, adaptability and flexible mode provides the solve problems and make decision process for certain and uncertain data with suggested rough set based algorithm structure.

This paper presents rough set based decision model, a process in which the suggested algorithm and decision support model and evaluate with the knowledge in databases and the knowledge received externally. The following sections include some necessary literature review and the rough based decision model approach. Then follow the future trends and the conclusion.

BACKGROUND

Rough set theory is an extension of set theory which proposed by Pawlak (1991) for describe and classify the incomplete or insufficient information. Besides it is mathematical tool that overcome the uncertainties and doubts. Also it verifies logic, and allows inconsistent data and no certainty to the discovery of incomplete implications. It is made as incomplete, inadequate and vague information by organizing. Rough set organizes the suitable data for analysis.

In real-world applications may includes the some uncertain and incomplete attributes in the knowledge representation systems in dynamic

situation, for this reason knowledge discovery and processing is very important for decision system. Meanwhile it is supported as a framework for conceptualizing and analyzing certain and uncertain types of data that is a powerful tool for discovering patterns with upper and lower approximations. Some of the studies used the rough set theory with minimum vertex cover problem (Chen et al. 2015; Chen et al. 2016); interval-valued information systems (Leung et al. 2008); intuitionistic fuzzy sets (Zhang et al. 2016; Huang et al. 2016) for knowledge discovery in feature selection (Huang et al. 2016) and rule induction (Lin et al. 2015). Shu & Qian (2015) and Yao & Zhao (2008) used the rough set theory for attribute reduction in pre-processing of the data mining and knowledge discovery. Macia-Perez et al. (2015) proposed the formal expansion of the rough set theory based algorithm for detection of the abnormal behaviour in outlier.

Nowadays, rough set theory has been demonstrated to be useful in many research areas such as knowledge discovery (Polkowski et al., 2010, Zhong, 1998; Sun et al., 2012), machine learning (Ananthanarayana et al., 2003; Pedrycz, 2013), decision analysis and support (Yao, 2010; Yao and Zhao, 2008; Pedrycz, 2014), expert systems (Tsumoto, 1998), and pattern recognition (Swinarski and Skowron, 2003). Rough set theory is also seen in the literature, which is used in conjunction with some heuristic algorithms with optimization applications (Changseok et al. 2010 and Yumin et al. 2010). Liu et al (2011) and (Nauman et al. 2016) used the probabilistic model to evaluate the clusters with rough sets approach. Zhong and Skowron (2001) proposed a rule of database in data mining. Liu et al (2011) and Fang et al. (2016) used the rough probabilistic model to evaluate the cluster approach with rough sets approach. As an example of feature extraction, Salamon and Lopez-Sanchez (2012) proposed the state-based classifier with the rough set theory for inference feature. Derrac et al (2012) used the snapshot feature in the selection process of selecting fuzzy rough set theory and proposed the algorithm with amplifi-

ers. Also rough sets used in wide applications in real life, such as evaluating of the credit ratings in the global banking industry (Chen, 2012) and developed the decision-making aid with hybrid models; tourism demand assessment (Carey et al., 2008;), identifying customer behaviours (James and Hshiung, 2010; Spiric et al, 2014) and medical applications (Tseng(Bill) et al., 2016).

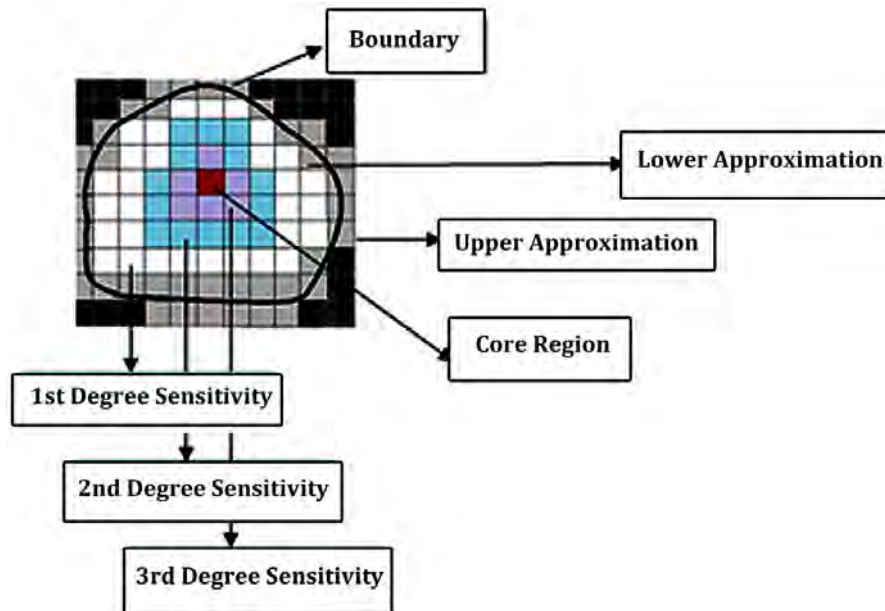
Cuckoo based optimization approach is the new method for the global optimization. It inspired from the Cuckoo bird family life style. It has been proposed by Yang and Deb (2009). Rajabiou (2011) reviewed the cuckoo based optimization algorithm; Ouaraab et al. (2014) developed and the realized the initial test model for the discrete form. This study combined the rough set theory and cuckoo search algorithm for decision making model with usable and correct rules.

ROUGH SET BASED DECISION MODEL

Rough set theory is a mathematical method which used in reasoning and expert systems for information extraction. As in rough set theory is a structure in which to accept certain restrictions.

Rough set theory comes about of the incomplete, inadequate and vague information which is organizing the data for analysis. This study reviews the rough set based containing data sets that both types of missing attribute values in the trivial losses and obtains the rules, recommends issuing a new algorithm. This algorithm can be implemented the dataset and lead to obtaining appropriate rules. Rough sets, data reduction, feature selection, identification of hidden relationships in data and the conversion of the decision rule is concerned with the suggested model. In particular, machine learning, decision analysis used by obtaining information and data from databases, expert systems, decision support systems in which, wide range application areas. In particular, data evaluation and analysis of process uses the incomplete and uncertain data as fuzzy logic not statistical probability

Figure 1. Rough set components



distribution. Rough set theory uses the upper and lower approach (Pawlak et al., 1995; Yao, 2001; Avşar, 2007). Suggested algorithm designed in decision support system which provides support to managers through advanced analysis models in the granting of semi- structured and unstructured qualified decision. Also, it is comprised of three main components which are data, model, and dialogue. Analyzing the data model created and rough set approach used in the decision support system. In particular, data evaluation and analysis process used the statistical probability distribution. Using the upper and lower rough set theory approach preferred the determination of the rules. Fig.1 represents the classical rough set schematic structure. It is shown the boundary; lower and upper approximation; sensitivity degree borders and core region. It can show and help to understand how uncertain or incomplete data evaluate with the rough set approach easily (Pawlak, 1991; Guo and Chankong (2002). Determination and evaluation of the upper and lower bound degrees we can use the cuckoo optimization algorithm for the best solution in a short time.

The Cuckoo based algorithm inspired from the Cuckoo bird family life style. It has been proposed by Yang and Deb in (2009). The algorithm matches the nest and similarly each individual attribute of the pattern corresponds to a Cuckoo-egg. The general system-equation of the Cuckoo algorithm based on the general system-equation of the random-walk algorithms with Levi flight which is a random walk that defined with step-lengths and have a certain probability distribution which steps are being isotropic and random. It is given in Eq. 1;

$$X_{g+1,i} = X_{g,i} + \alpha \otimes \text{levy}(\lambda) \quad (1)$$

Egg laying and breeding represent their main characteristics which has been the basic motivation for development of this new evolutionary optimization algorithm. Mature cuckoos and eggs are the main components of the suggested algorithm. The main aim is survival competition some of the Cuckoos or their eggs, demise. The survived cuckoo societies immigrate to a better environment and start reproducing and laying eggs.

Listing 1. Cuckoo algorithm pseudo code (Yang and Deb, 2009)

```

Begin
Objective function  $f(\mathbf{x})$ ,  $\mathbf{x}=(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)^T$ ;
Initial population of  $n$  host nests  $\mathbf{x}_i$  ( $i=1, 2, \dots, n$ );
    While (t < MaxGeneration) or (stop criterion)
        Get a cuckoo (say i) randomly by Levyflights;
    Evaluate its quality/fitness  $F_i$ ;
        Choose a nest among  $n$  (say j) randomly;
        if ( $F_i > F_j$ )
            Replace j by the new solution;
        End
        A fraction (pa) of worse nests are abandon and
        new once are built
        Keep the best solutions (or nests with quality solutions);
        Rank the solutions and find the current best;
    End while
Post-process results and visualization;
End

```

Cuckoos' survival effort hopefully converges to a state that there is only one cuckoo society, all with the same profit values. During the analysis of the system, following steps are considered.

- Determination of the information system.
- Definition of the distinctive Matrix.
- Discrimination relationship with Cuckoo algorithm.
- Realizing of the reduction.
- Forming of the decision rules.
- Classification of the new features.

Determination of the Information System

Information system is a table in which rows consist of the units (samples) and in column includes the each unit features. S represents the information system and $S=\langle U, A, V \rangle$ consists of the U is a Units set, A symbolizes the conditional attributes of the each unit, V represents the each units attribute value. Any character of the attribute value is

$$a : U \rightarrow V_a, \quad \forall a \in A \quad (2)$$

(Lee ve Vatchsevanos, 2002). V_a is the a 's value set.

Definition of the Distinctive Matrix

In S information system, $M(A)=(m_{ij})_{n \times n}$ represents the distinctive matrix for A attributes. $M(A)$ can shown in

$$M(A) = \left\{ \begin{matrix} \Phi \\ a \in A : a(x_i) \neq a(x_j) \end{matrix} \right\} \quad (3)$$

$M(A)$ has distinctive symmetric matrix characteristics. Each element of the x_i and different x_j values compromise the $M(A)$. Distinctive matrix shows the binary comparison result of the attributes different values.

Applying Cuckoo Algorithm for Discrimination of Relationships

In rough set discrimination relationship represents two approach which are upper and lower approaches. We can able to use the Cuckoo optimization algorithm to determination of the alternative upper and lower solutions for beginning points. Rules induced from the lower approximation concept certainly describe the rules, which are called certain. On the other hand, rules induced from the upper approximation of the concept describe the concept possibly, so these rules are called possible.

Rough set approach is based on two concepts called upper and lower approach defined as follows:

- Elements are strictly belong to related set.
- Elements are likely to belong to related set.

$\underline{X}$ represents the lower approach which includes the certain that the interior value and it combines the primary cluster.

$$\underline{AX} = \{x_i \in U \mid [x_i]_{Ind(A)} \subset X\} \quad (4)$$

$\overline{X}$ upper approach, in the form of demonstrations and the intersection of the empty set's combination of the non-primary cluster with X.

$$\overline{AX} = \{x_i \in U \mid [x_i]_{Ind(A)} \cap X \neq \emptyset\} \quad (5)$$

Realizing of the Reduction

- **Reduced Feature Sets:** It represents the data structure in information system with minimum feature set values.
- **Core Feature Sets:** It occurs from the reduced features set's shared elements.
- **Similarity Matrix:** Reduced feature sets used the similarity matrix with reduced feature sets elements. It is logical function and used the AND and OR functions.

Reduced feature sets and information system can defined lesser than feature. Realizing minimum feature sets represent the rule base.

Forming of the Decision Rules

$D = (U, A \cup \{d\}, V)$ decision system is represented by D's degree of the $r(d)$. d decision features' values equal to the counts of the set elements, which is including the degree of D $r(d)$ are indicated with $r(d)$, the value of d is equal to the number of elements in a set of decision attributes. d decision feature set value is V_d decision feature values set is $\{V_A^1 \cup \dots \cup V_A^{r(d)}\}$. d decision feature, U units set is the

$$CLASS_A(d) = \{X_A^1 \dots X_A^{r(d)}\}$$

and

$$X_A^k = \left\{ x \in U \mid d(x) = V_d^k \right\}, 1 \leq k \leq r(d) \quad (6)$$

X_A^k is the k th decision feature value and decision class. Two unit values has the same decision feature value belong to the same decision class and represents with the U unit set is $\{X_A^1 \cup \dots \cup X_A^{r(d)}\}$.

Validation Test of the Features Dependencies

Dependency of the feature is between coefficient of the A conditional features set and d decision features that are represented with $\gamma(A, d)$.

$$\gamma(A, d) = \frac{|POS_A(d)|}{|U|} \quad (7)$$

$\gamma(A, d)$ have between 0 and 1 values. $\gamma(A, d) = 1$ is the decision feature value which depends on the A conditional feature sets. If $\gamma(A, d) \leq 1$ then d decision feature values depends on the semi A conditional features set. In Listing 2, suggested algorithm of rough set based decision model is introduced.

Listing 2. Algorithm of cuckoo search based rough set decision model

Input: Each m feature valued and C class set member's n objects, U missing quantitative value data sets.

Output: Strict and possible rough rule set

Step 1: Label of class consider the discrete feature sub set and divide each other. Same C_i class's objects represent X_i .

Step 2: $v^{(i)}$ represents the each $Obj^{(i)}$ objects quantitative values for each A_i $i=1$ to n until $j=1 \dots m$. Function values shown as follows:

$$\left(\frac{p_{j_1}^{(i)}}{R_{j_1}} + \frac{p_{j_2}^{(i)}}{R_{j_2}} + \dots + \frac{p_{j_l}^{(i)}}{R_{j_l}} \right) \quad (8)$$

R_{jk} represents the A_j 's k th region strict value and $l (=|A_j|)$ A_j shows the region count. (*) or (?) symbols symbolize the missing values.

Step 3: Missing has unclear value with the primary cluster of qualifications. If, for this the exact possible value '=' equity class into the possible incomplete (c) format ; For trivial value (*) imprecise into each missing class equality (u) format ; Lost value for the (?) It was the object located in the same area as unclear fuzzy missing class into the equation (1) is placed. Possible values is calculated as follows: to be precise =, and 0;

$$P_{Ajk} = \text{Min } p_{jk}^{(i)} \quad (9)$$

The standard Cuckoo Search algorithm includes the three basic rules (Yang and Deb, 2009):

D

- Each cuckoo lays one egg at a time, and dumps it in a randomly chosen nest;
- The best nests with high quality of eggs (solutions) will carry over to the next generations;
- The number of available host nests is fixed, and a host can discover an alien egg with a probability $p_a \in [0, 1]$.

For simplicity, this last assumption can be approximated by a fraction p_a of the n nests being replaced by new nests (with new random solutions at new locations).

When generating new solutions x (t+1) for, say cuckoo i, a Levy flight as a local random walk is performed

$$x_i^{(t+1)} = x_i^{(t)} + \alpha \otimes L(s, \lambda), \quad (10)$$

where $\alpha > 0$ is the step size which should be related to the scales of the problem of interest. The L function as a global random walk is

$$L(S, \lambda) = \frac{\left(\sin \left(\frac{\pi}{2} \right) \right)}{s^{1+\lambda}} \quad (s \gg s_0 > 0) \quad (11)$$

In most cases, we can use $\alpha = O(1)$. The product $\otimes$ means entry-wise multiplications. Levy flights essentially provide a random walk while their random steps are drawn from a Levy distribution for large steps:

$$x_i^{(t+1)} = x_i^{(t)} + \alpha \otimes H(p_a - \epsilon) \otimes (x_j^{(t)} - x_k^{(t)}), \quad (12)$$

which has an infinite variance with an infinite mean. Here the consecutive jumps/steps of a cuckoo essentially form a random walk process which obeys a power-law step length distribution with a heavy tail.

In this case, the host bird can either throw the egg away or abandon the nest so as to build a

completely new nest in a new location. For simplicity, this last assumption can be approximated by a fraction α of the n nests being replaced by new nests (with new random solutions at new locations). For a maximization problem, the quality or fitness of a solution can simply be proportional to the objective function.

Step 4: Start with $q = 1$. Then q determines the count of the number of the current feature with lower approach.

Cuckoo lays one egg at a time, and dumps it in a randomly chosen nest with Levy approach in this step with below loops:

```

Obtain  $C_{1,...,k}$  group solutions in the
solution space
    For  $i=1$  to  $K$ 
        Situation( $C_i$ ) =

$$\frac{\sum_{x \in C_i} f(x)}{\text{solution count}}$$

    End
aim = ArgMin $_{i \in \{1, \dots, K\}}$  average $_{(C_i)}$  [Minimization problem]
h_situation = ArgMin $i \in C_{\text{aim}}$   $f(x)$  [Minimization problem]
    For  $i=1$  to  $N$ 
        For  $k=1$  to  $d$ 
             $x_{i,k} = h\_situation\ k + \lambda * (x_{i,k} - h\_situation\ k)$ 
        End
    End
End
    
```

Step 5: Each X_i class consider the q feature in A subset value with lower approach calculate as follows:

$$A_*(x) = \{ (A_k(Obj^{(i)}), P_{A_k}(Obj^{(i)})) \mid 1 \leq i \leq n, Obj^{(i)} \in X_i, A_k^c(Obj^{(i)}) \subseteq X_i, 1 \leq k \leq |A(Obj^{(i)})| \} \quad (13)$$

A feature sub set includes the $Obj^{(i)}$, which is derived the missing similarity class value that represents $A_k^c(Obj^{(i)})$. It is the k th strict part.

Step 6: Each missing values for $Obj^{(i)}$ object calculated the following steps are followed:

- If $Obj^{(i)}$ object is the A th k combinations, which is applied lower approach to the subset, then obtained the uncertain value as follows:

$$\frac{\sum_{Obj^{(r)} \in A_k^c(Obj^{(r)})} v_j^{(r)} x p_{jk}^{(r)}}{\sum_{Obj^{(r)} \in A_k^c(Obj^{(r)})} p_{jk}^{(r)}} \quad (14)$$

$v_j^{(r)}$ is the quantity value for A_j in $Obj^{(r)}$ and $f_{jk}^{(i)}, R_B^k$ is the membership value for $v_j^{(r)}$. Thus, predicted $Obj^{(i)}$ values changed in rough sets. Rough missing equation class has zero membership values ($Obj^{(i)}, u$) or deleted ($Obj^{(i)}, l$) changed with ($Obj^{(i)}, c$). Rough missing approach evaluates the similarity class values again. In addition $Obj^{(i)}$ function applies to same process with the lower approach.

- $Obj^{(i)}$ object compares the one more than values with lower approach and postpone the prediction of the missing value until determine of the new features.

Step 7: $q = q + 1$. Then if $q > m$ then repeat Steps 5- 7.

Step 8: $Obj^{(i)}$ object has multiple missing value, which considers the maximum numerical expression to estimate the significance of uncertain value object. For estimate and process of the data go to Step 6 (a).

Step 9: Each missing a certain rules derived with the lower approach and next future data determined with the rough set approach.

Step 10: Final rule's effectiveness measures then compares the same exact condition rules and delete the equal and smaller.

Step 11: If $q = 1$ then do. q is used to count the number of missing upper approach.

Cuckoo lays one egg at a time, and dumps it in a randomly chosen nest with Levy approach in this step with below loops:

```

Obtain  $C_{1,...,k}$  group solutions in the
solution space
  For  $i=1$  to  $K$ 
    Situation( $C_i$ ) =
      
$$\frac{\sum_{x \in C_i} f(x)}{\text{solutioncount}}$$

    End
  aim = ArgMax $_{i \in \{1, \dots, K\}}$  average $_{(C_i)}$  [Maximi-
  zation problem]
  h_situation = ArgMax $_i \in C_{\text{aim } f(x)}$  [Maximiza-
  tion problem]
  For  $i=1$  to  $N$ 
    For  $k=1$  to  $d$ 
       $x_{i,k} = h\_situ-$ 
      ation  $k + \lambda * (x_{i,k} - h\_situation\ k)$ 
    End
  End

```

Step 12: Each X_l class's q quality, each subset A missing degree calculate with upper approach as follows:

$$A^*(x) = \{ (A_k(Obj^{(i)}), \mu_{B_k}(Obj^{(i)})) | 1 \leq i \leq n, A_k^c(Obj^{(i)}) \cap X_l \neq \varphi, A_k^c(Obj^{(i)}) \not\subseteq X_l, 1 \leq k \leq |A(Obj^{(i)})| \} \quad (15)$$

$A(Obj^{(i)}), Obj^{(i)}$ a set of equations derived missing subset of the class that contains $A_k^c(Obj^{(i)})$, $A(Obj^{(i)})$ the k -th class defined part of the equation.

Step 13: Rough set approach in each $Obj^{(i)}$ object of the missing upper method applied to imprecise (trivial or loss), the following steps are:

- $Obj^{(i)}$ object with missing upper approach, is assigned the k -th sub set of A feature set R_B^k in the combination of the area, which represents the missing $Obj^{(i)}$ uncertain value as follows:

$$\frac{\sum_{Obj^{(r)} \in A_k^c(Obj^{(r)}) \& Obj^{(r)} \in X_l} v_j^{(r)} x p_{jk}^{(r)}}{\sum_{Obj^{(r)} \in A_k^c(Obj^{(r)}) \& Obj^{(r)} \in X_l} p_{jk}^{(r)}} \quad (16)$$

- $Obj^{(i)}$ function, if one or more missing value is presented than finding of the feature process postponed until the missing information is estimated to have identified them.

Step 14: $q = q + 1$ and if $q > m$ then repeat Steps 11-13 otherwise Stop.

Step 15: Each X_l class missing rough class is used to upper approach for suitable degree as follows:

$$P(A_k^c(Obj^{(i)})) = \frac{\sum_{Obj^{(r)} \in A_k^c(Obj^{(r)}) \& Obj^{(r)} \in X_l} p_{jk}^{(r)}}{\sum_{Obj^{(r)} \in A_k^c(Obj^{(r)})} p_{jk}^{(r)}} \quad (17)$$

Step 16: $Obj^{(i)}$ function represents the missing rough low approach, multiple rough equality missing class introduces, use a maximum suitability measure for estimating the value of certain non-object. Estimates and processing to Step 13 (a) is performed as follows.

Step 17: Possible rules from the rough set approach evaluate subset of missing each Class A and the estimated size of the objects are derived. In addition, the measure of effectiveness for future data determined from the possible values of equality classes.

Step 18: Possible essential part of the rules are the same, which both the effectiveness and appropriateness of the measure value are equal to or less than clear, that specialized and other possible precise and rules determined.

FUTURE TRENDS

The development of decision support one of the key issues in the uncertainties and in-completed situations. The decision strategy developed in this paper provides a way for designing rough set theory are application independent, and improves the performance of decision system with possible proper decision with uncertainty. The experiments have shown that the proposed approach is effective and promising. This study focuses on an algorithm that provides valuable quantitative inferences from incomplete data sets with cuckoo search algorithm based rough set theory for decision rules. Suggested algorithm of rough set based decision model will be used in decision support system for real life applications in various aspects.

For future research, it is noted that cuckoo search algorithm based rough set based decision model involves not just modelling the data but also modelling operations on the data. Also different methodological approach in hybrid structure will be developed such as matrix based; entropy based; multi integer programming approach and feature extraction approach for alternative making decision style.

CONCLUSION

This article discusses a variety of issues in adapting rough set approach in distributed decision system. This study shows how rough set approach can be introduced to different aspects of rule execution from event detection to different situations with sensitivity analysis. Our suggested model includes the determination of information system; defini-

tion of the distinctive matrix; determination of the discrimination relationship; realizing of the reduction; forming of the decision rules; classification of the new features.

Our next work, which considers as a rough set theory with sensitivity analysis in different factors. Our suggested cuckoo search algorithm based rough set model can be applied to distributed decision system. The conditional attribute structure, discernibility matrix is shown in detail. We discussed the distributed decision information system for uncertainty and incomplete data. In addition, this algorithm is applied to data sets containing only the missing attribute value of lost species; it is found that increasing the number of rules obtained in this way and reduces the success rate. Missing attribute the increase in the number formed by the success rate of decline in large-scale and useful information obtained due to the increase in the number of rules.

REFERENCES

- Ananthanarayana, V. S., Narasimha, M. M., & Subramanian, D. K. (2003). Tree structure for efficient data mining using rough sets. *Pattern Recognition Letters*, 24(6), 851–862. doi:10.1016/S0167-8655(02)00197-6
- Avşar, G. (2007). *Extraction of Fuzzy Rules from Incomplete Data with Do not Care and Lost Value by Rough Sets* (Master's Thesis). Firat University, Graduate School of Natural and Applied Sciences, Department of Computer Engineering.
- Carey, G., Law, R., & Mok, H. M. K. (2008). Analyzing and Forecasting Tourism Demand: A Rough Sets Approach, *Journal of Travel Research*, 46(3), 327-338.
- Changseok, B., Wei-Chang, Y., Yuk Ying, C., & Sin-Long, L. (2010). Feature selection with Intelligent Dynamic Swarm and Rough Set. *Expert Systems with Applications*, 37(10), 7026–7032. doi:10.1016/j.eswa.2010.03.016

- Chen, J., Lin, Y., Li, J., Lin, G., Ma, Z., & Tan, A. (2016). A rough set method for the minimum vertex cover problem of graphs. *Applied Soft Computing*, 42(May), 360–367. doi:10.1016/j.asoc.2016.02.003
- Chen, J., Lin, Y., Lin, G., Li, J., & Ma, Z. (2015). The relationship between attribute reducts in rough sets and minimal vertex covers of graphs. *Information Sciences*, 325, 87–97.
- Chen, Y. S. (2012). Classifying credit ratings for Asian banks using integrating features election and the CPDA-based rough sets approach. *Knowledge-Based Systems*, 26, 259–270. doi:10.1016/j.knosys.2011.08.021
- Derrac, J., Cornelis, C., Garcia, S., & Herrera, F. (2012). Enhancing evolutionary instance selection algorithms by means of fuzzy rough set based feature selection'. *Information Sciences*, 186(1), 73–92. doi:10.1016/j.ins.2011.09.027
- Fang, B. W., & Hu, B. Q. (2016). Probabilistic graded rough set and double relative quantitative decision-theoretic rough set. *International Journal of Approximate Reasoning*, 74(July), 1–12. doi:10.1016/j.ijar.2016.03.004
- Guo, J., & Chankong, V. (2002). Rough set-based approach to rule generation and rule induction. *International Journal of General Systems*, 31(6), 601–617. doi:10.1080/0308107021000034353
- Huang, B., Guo, C. X., Li, H. X., Feng, G. F., & Zhou, X. Z. (2016, September). An intuitionistic fuzzy graded covering rough set. *Knowledge-Based Systems*, 107(1), 155–178. doi:10.1016/j.knosys.2016.06.006
- Huang, C. C. Tseng (Bill), T., & Tang, C. Y. (2016). Feature extraction using rough set theory in service sector application from incremental perspective. *Computers & Industrial Engineering*, 91, 30–41.
- James, J. H. L., & Gwo-Hshiung, T. (2010). A Dominance-based Rough Set Approach to customer behavior in the airline market. *Information Sciences*, 180(11), 2230–2238. doi:10.1016/j.ins.2010.01.025
- Lee, S., & Vatchsevanos, G. (2002). An application of rough set theory to defect detection of automotive glass. *Mathematics and Computers in Simulation*, 60(3-5), 225–231. doi:10.1016/S0378-4754(02)00017-4
- Leung, Y., Fischer, M. M., Wu, W. Z., & Mi, J. S. (2008). A rough set approach for the discovery of classification rules in interval-valued information systems. *International Journal of Approximate Reasoning*, 47(2), 233–246.
- Lin, S. H., Huang, C. C., & Che, Z. X. (2015). Rule induction for hierarchical attributes using a rough set for the selection off a green fleet. *Applied Soft Computing*, 37(December), 456–466. doi:10.1016/j.asoc.2015.08.016
- Liu, D., Li, T., & Ruan, D. (2011). Probabilistic model criteria with decision-theoretic rough sets. *Information Sciences*, 181(17), 3709–3722. doi:10.1016/j.ins.2011.04.039
- Macia-Perez, F., Berna-Martinez, J. V., Oliva, A. F., & Ortega, M. A. A. (2015). Algorithm for the detection of outliers based on the theory of rough sets. *Decision Support Systems*, 75(July), 63–75. doi:10.1016/j.dss.2015.05.002
- Nauman, M., Azam, N., & Yao, J. T. (2016). A three-way decision making approach to malware analysis using probabilistic rough sets. *Information Sciences*, 374, 193–209.
- Ouaarab, A., Ahiod, B., & Yang, X.-S. (2014, June). Discrete Cuckoo Search Algorithm for The Travelling Salesman Problem. *Neural Computing & Applications*, 24(7), 1659–1666. doi:10.1007/s00521-013-1402-2

- Ouaarab, A., Ahiod, B., & Yang, X.-S. (2014). Improved and Discrete Cuckoo Search for Solving the Travelling Salesman Problem. In Cuckoo Search and Firefly Algorithm: Theory and Applications. Academic Press.
- Pawlak, J., Grzymala-Busse, Z., Slowinski, R., & Ziarko, W. (1995). Rough sets. *Communications of the ACM*, 38(11), 89–95. doi:10.1145/219717.219791
- Pawlak, Z. (1982). Rough Sets. *International Journal of Computer and Information Sciences*, 11(5), 341–356. doi:10.1007/BF01001956
- Pawlak, Z. (1991). *Rough Sets, Theoretical Aspects of Reasoning about Data*. Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Pawlak, Z., & Skowron, A. (1994). Rough Set Rudiments. *The International Workshop on Rough Sets and Soft Computing*, 72.
- Pedrycz, W. (2013). *Granular Computing: Analysis and Design of Intelligent Systems*. Boca Raton, FL: CRC Press/Francis Taylor. doi:10.1201/b14862
- Pedrycz, W. (2014). Allocation of information granularity in optimization and decision making models: Towards building the foundations of granular computing. *European Journal of Operational Research*, 232(1), 137–145. doi:10.1016/j.ejor.2012.03.038
- Polkowski, L., Tsumoto, S., & Lin, T. Y. (Eds.). (2010). *Rough Set Methods and Applications*. Berlin: Physica-Verlag.
- Rajabioun, R. (2011). Cuckoo Optimization Algorithm. *Applied Soft Computing*, 11(8), 5508–5518. doi:10.1016/j.asoc.2011.05.008
- Salamo, M. & Loipez-Sanchez, M., (2012). Rough set based approaches to feature selection for Case-Based Reasoning classifiers. *Pattern Recognition Letters*, 32(2), 280–292.
- Shu, W., & Qian, W. (2015). An incremental approach to attribute reduction from dynamic incomplete decision systems in rough set theory. *Data & Knowledge Engineering*, 100, 116–132.
- Spiric, J. V., Stankovic, S. S., Docic, M. B., & Popovic, T. D. (2014). Using the rough set theory to detect fraud committed by electricity customers. *International Journal of Electrical Power & Energy Systems*, 62(November), 727–734. doi:10.1016/j.ijepes.2014.05.004
- Sun, L., Xu, J. C., & Tian, Y. (2012). Feature selection using rough entropy-based uncertainty measures in incomplete decision systems. *Knowledge-Based Systems*, 36, 206–216. doi:10.1016/j.knosys.2012.06.010
- Swiniarski, R. W., & Skowron, A. (2003). Rough set methods in feature selection and recognition. *Pattern Recognition Letters*, 24(6), 833–849. doi:10.1016/S0167-8655(02)00196-4
- Tseng, T.L., Huang, C.C., Fraser, K., & Ting, H.W. (2016). Rough set based rule induction in decision making using credible classification and preference from medical application perspective. *Computer Methods and Programs in Biomedicine*, 127, 273–289.
- Tsumoto, S. (1998). Automated extraction of medical expertsystem rules from clinical databases based on rough set theory. *Information Sciences*, 112(1-4), 67–84. doi:10.1016/S0020-0255(98)10021-X
- Yang, X.-S., & Deb, S. (2009). Cuckoo search via Levy flights. In *Proc. of World Congress on Nature & Biologically Inspired Computing* (pp. 210–214). IEEE Publications. doi:10.1109/NABIC.2009.5393690
- Yao, Y. Y. (2001). Information granulation and rough set approximation. *International Journal of Intelligent Systems*, 16(1), 87–104. doi:10.1002/1098-111X(200101)16:1<87::AID-INT7>3.0.CO;2-S

Yao, Y. Y. (2010). Three-way decisions with probabilistic rough sets. *Information Sciences*, 180(3), 341–353. doi:10.1016/j.ins.2009.09.021

Yao, Y. Y., & Zhao, Y. (2008). Attribute reduction in decision-theoretic rough set models. *Information Sciences*, 178(17), 3356–3373. doi:10.1016/j.ins.2008.05.010

Yumin, C., Duoqian, M., & Ruizhi, W. (2010). A rough set approach to feature selection based on ant colony optimization. *Pattern Recognition Letters*, 31(3), 226–233. doi:10.1016/j.patrec.2009.10.013

Zhang, Y., Li, T., Luo, C., Zhang, J., Chen, H. (2016). Incremental updating of rough approximation in interval-valued information systems under attribute generation. *Information Sciences*, 373, 461–475.

Zhong, N., Dong, J. Z., & Ohsuga, S. T. Y. L. (1998). An incremental probabilistic rough set approach to rule discovery. *The IEEE International Conference on Fuzzy Systems*, 2. doi:10.1109/FUZZY.1998.686243

Zhong, N., & Skowron, A. (2001). A rough set-based knowledge discovery process. *International Journal of Applied Mathematics and Computer Science*, 11, 603–619.

ADDITIONAL READING

Hu, B. (2016). Three-way decision spaces based on partially ordered sets and three-way decisions based on hesitant fuzzy sets. *Knowledge-Based Systems*, 91, 16–31. doi:10.1016/j.knsys.2015.09.026

Nina, F. R. C. (2007). “On Applications of Rough Sets Theory to Knowledge Discovery”, Doctor of Philosophy Thesis in Computing and Information Sciences and Engineering, University of Puerto Rico.

Rissino, S. Lambert-Torres (2009). “Rough Set Theory—Fundamental Concepts, Principles, Data Extraction, and Applications”, *Data Mining and Knowledge Discovery in Real Life Applications*, Julio Ponce and Adem Karahoca, eds., Vienna, Austria: I-Tech, 438.

Wei, D., Zhao, Y., & Zhou, X. (2006). “A Rough Set Approach to Incomplete and Fuzzy Decision Information System”, *Proceedings of the 6th World Congress on Intelligent Control and Automation*, June 21 - 23, 2006, Dalian, China.

Yang, X. S., & Deb, S. (2013). Multi objective cuckoo search for design optimization. *Computers & Operations Research*, 40(6), 1616–1624. doi:10.1016/j.cor.2011.09.026

KEY TERMS AND DEFINITIONS

Alternative Set Theory: It means system set theory that related the positive set theory and constructive set theory.

Attribute: Refers with decision table in rough set which is divided into two disjoint groups called condition and decision attributes (action, results, outcome, etc.).

Cuckoo Search Algorithm: A method of global optimization based on the behaviour of cuckoos was proposed by Yang & Deb (2009). The breeding behaviour types are, laid their eggs in the host nests; if not detected and destroyed, the eggs are hatched to chicks by the hosts.

Decision Rules: It determines decision with rules under certain and uncertain conditions.

Decision Support System: It is computer based information system that support decision making activities with inputs, user knowledge and expertise, outputs and decision components.

Incomplete Information: Three types of incomplete data consists of attribute values which are lost values; attribute missing values; irrelevant concept data in attribute.

Indiscernibility Relation: It is a central concept of Rough Set Theory which relates between two or more objects identical relation in subset of the attributes.

Information System: It consists of objects and attributes that shown in table with rows-objects and columns-attributes.

Lower Approximation: It consists of all objects which surely belong to the set.

Rough Relations: Collection of such relations is closed under different binary compositions

such as, algebraic sum, algebraic product etc. for uncertainty and incomplete data.

Rough Set Theory: It is first described by Zdzislaw I. Pawlak in early 1980's. Every object of the universe of discourse some information (data, knowledge) is associated with lower and upper approximation.

Sensitivity Analysis: Analysis of the uncertainty output in system is evaluated by the different certain or uncertain input.

Upper Approximation: It contains all objects which possibly belong to the set.

Using Receiver Operating Characteristic (ROC) Analysis to Evaluate Information –Based Decision–Making

D**Nan Hu***University of Utah, USA*

INTRODUCTION

Business operators and stakeholders often need to make decisions such as choosing between A and B, or between yes and no. These decisions include, but are not limited to, whether to invest in project A versus project B, or whether to continue running a company. These are often made by using a classification tool or a set of decision rules. For example, banks often use credit scoring systems to classify lending companies or individuals into a high or low risk of default, thus helping to decide whether to grant a loan. One important question businesses need to answer is how accurate the information based on these classification tools can help them make a correct decision, or how correctly they can be used to discriminate between two groups of subjects. In this chapter, we address this important issue by presenting accuracy parameters for assessing classification tools such as test modalities, scoring systems, and prediction models. Specifically, we introduce the receiver operating characteristics (ROC) curve as a statistical tool to evaluate these modalities. The ROC curve is widely used in business optimization analysis, health policy making, clinical studies, and health economics (Kampfrath & Levinson, 2013). In the Background section, we give updated examples of using the ROC related methods for assessing decision-makings based on our most current literature review. In the Main Focus section of this chapter, we provide mathematical definitions of the classification accuracy parameters, and describe the procedure to obtain an ROC curve. In addition, we present recent statistical developments

DOI: 10.4018/978-1-5225-2255-3.ch192

in ROC curve methodologies and applications of ROC analysis in a diversity of research areas.

BACKGROUND

Business classification tools include scoring systems, predictive models, and quantitative test modalities. A classification tool is useful in business analytics only if it is shown to distinguish entities with a certain condition from those without that condition. For instance, a credit scoring system is a valuable classification tool for bankers when it can accurately classify between companies with default status (cases) and without default status (controls). A perfect test modality would categorize all default companies as cases and all non-default companies as controls. However, in practice, almost none of the testing modalities can make such a perfect classification. This implies that misclassifications can always exist and the correct classification rate may vary from one test to another. Thus, assessing classification performance among different test modalities is always a necessary step in making important business-related decisions.

MAIN FOCUS

We first define accuracy parameters of binary classification tools, and then extend the evaluation method to test modalities with continuous or discrete ordinal values. By applying accuracy parameters and ROC analysis, business analysts

can easily examine the expected downstream harms and benefits of positive and negative test results based on these test modalities, and directly link the classification accuracy to important decision making (Cornell, Mulrow & Localio, 2008).

Accuracy Parameters for Classification and Decision Making

The accuracy of decision making should be measured by comparing the decision taken by a business to the choice that would be taken in order to maximize its benefit. In this section, we introduce two basic accuracy parameters, sensitivity and specificity, and two misclassification measures, the false positive rate and false negative rate. We define accuracy parameters in the context of classifying the default status of borrowers (companies that apply for a loan). Let S denote the dichotomous true default status such that $S = 0$ represents “no default,” and $S = 1$ indicates “default.” Let Y be the value of a test modality or scoring system. We suppose that Y is also binary such that $Y = 1$ denotes the test positive for default, and $Y = 0$ indicates the test negative. In reality, companies with a positive test result are often refused for a loan. The *sensitivity* of the binary test Y is defined as the probability of test positive among companies with default status ($S = 1$). Mathematically, this probability can be expressed as

$$\text{Sensitivity} = \Pr(Y = 1 \mid S = 1),$$

where the symbol $\mid$ denotes the statistical concept of *conditioning*, the definition of which can be found in introductory statistics books such as Wasserman (2004), Chap. 1. The sensitivity of a test is also known as the *true positive rate (TPR)*. Another important accuracy parameter is the *specificity* of Y , which is defined as the probability of test negative when the default status is absent. This probability is given by

$$\text{Specificity} = \Pr(Y = 0 \mid S = 0).$$

Specificity is often used interchangeably with the *true negative rate (TNR)* in the literature. Both sensitivity and specificity are correct classification rates of a test. Since such a test may also misclassify subjects, error rates are of interest as well.

There are also two types of misclassification rates. The first is the *false positive rate (FPR)*, which is defined as the probability of test positive when the default status is absent. Mathematically,

$$\text{FPR} = \Pr(Y = 1 \mid S = 0).$$

A false positive occurs when a “refusal of loan” decision is made to companies that would never default. By examining the definitions of *FPR* and *specificity*, we note that $\text{FPR} = 1 - \text{specificity}$.

Another misclassification rate is the *false negative rate (FNR)*, which is the probability of test negative when the default status is present. This rate can be expressed by

$$\text{FNR} = \Pr(Y = 0 \mid S = 1).$$

A false negative occurs when a loan is granted to a company that later defaults on the loan. Also, we note that $\text{FNR} = 1 - \text{sensitivity}$.

Table 1 summarizes the aforementioned accuracy and misclassification parameters. The rows of this two-by-two table are split by the true default status ($S = 1$ versus $S = 0$), and columns are classified by test results ($Y = 1$ versus $Y = 0$). In each of the four cells defined by S and Y , the top row displays the frequency of the cell and the bottom row lists the mathematical equation for the accuracy parameter or misclassification rate corresponding to that cell.

Comparing Test Modalities With Binary Values

In the process of decision making, business analysts often have several candidate test modalities with binary values without knowing which modality has the best classification accuracy.

Table 1. Frequencies and mathematical equations of accuracy measures for a binary test (Y)

Test Results	Positive (Y=1)	Negative (Y=0)	Total
True Status			
Positive (S=1)	True positives A Sensitivity = $A/(A+B)$	False positives B False positive rate = $B/(A+B)$	A+B
Negative (S=0)	False negatives C False negative rate = $C/(C+D)$	True negatives D Specificity = $D/(C+D)$	C+D
Total	A+C	B+D	A+B+C+D

For example, banks often want to minimize the chance for granting loans to companies that could default. In another words, banks wish to obtain the test modality that leads to the highest sensitivity. In contrast, when evaluating new projects for investment, business owners prefer to reduce the chance of a poor investment to the greatest extent possible. That is, they need a test modality that produces the greatest specificity.

In order to compare different test modalities, historical data based on a relatively large sample is needed. Unbiased sampling from the study population is also required. To compare two different modalities, two types of study design can be adopted: unpaired design and paired design. Under the unpaired design, the two test modalities are applied to two different groups of subjects or companies. Thus, each subject experiences only one of the two tests. Under the paired study design, every subject is tested by both modalities (Pepe, 2003).

In practice, business analysts may give priority to comparing either sensitivity or specificity between tests. In order to obtain statistical significance of the comparisons, a hypothesis test on two different proportions should be used. Under the unpaired design, test for the proportions of two candidate classification modalities is asymptotically a two-sample normal test, or Z-test. Thus, when the sample size is large, we can always perform a Z-test to compare sensitivities or specificities. However, when the sample size is small, Fisher's exact test should be used. Under the paired design, McNemar's test for paired binary data should be used to compare candidate

modalities. Statistically, this is a Chi-squared test with one degree of freedom. The aforementioned hypothesis tests can be easily performed using commercial software such as SAS (SAS Institute, Inc., Cary, NC, USA) and STATA (Stata Inc., College Station, TX, USA), and the free statistical package R (www.r-project.org).

ROC Curves

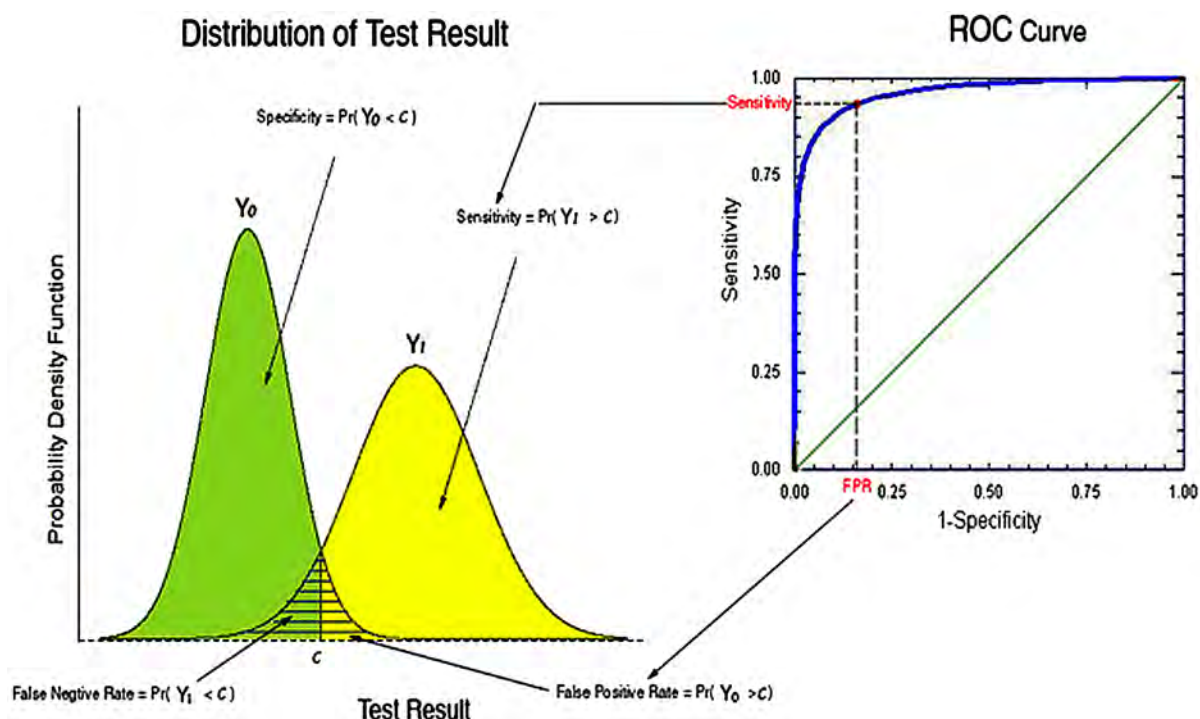
Many test modalities are measured on a continuous or discrete ordinal scale. When such tests are evaluated, we cannot define their sensitivity and specificity without dichotomizing their distributions. A standard and popular tool for evaluating accuracy of these modalities is the receiver operating characteristic (ROC) curve, which has been studied or described in detail in literature such as Swets and Pickett (1982), Swets (1988), Hanley (1989), Begg (1991), Chock, Irwig, Berry & Glasziou (1997), Zhou, Obuchowski & McClish (2002), and Pepe (2003). To dichotomize the continuous distribution, we usually need to specify a *testing threshold* or *cut-point* upon which to base our decision. In this section, we use an example of classifying the default status to illustrate features of the ROC curves. There are two dimensions of an ROC curve. The first is the gold standard (GS) of the test, also known as the reference standard (RS), which is the true status being classified. It is often assumed that the GS for each observation can be confirmed without error. In assessing a credit scoring system, the true default status for each company is the GS, which can be determined based on the bank's guidelines

for default. Although a GS is binary in nature, it can also take on more than two categories. For example, in personal health assessment, the true status could be “excellent,” “very good,” “good,” “fair,” or “poor.” In order to use ROC analysis to assess risk prediction models, investigators often need to dichotomize the categorical variable. Perneger & Courvoisier (2011), in their study of assessing multi-attribute health utility measures, dichotomized the true health status as “excellent” or “very good” versus “good,” “fair,” or “poor.” Classifiers are another aspect of ROC curves. They are the measurements being used for classifications. To classify the default status, the classifier could be a credit scoring system or a risk predictive score estimated from a regression model. When the single threshold value for classifiers needed to make a decision is not known *a priori*, an ROC curve provides a useful description of accuracies for the classifiers. The ROC curve plots a test’s sensitivity versus its FPR (or 1-specificity) at each cut-point c . As the cut-point defining the positivity changes across all

possible values of the test scale, the ROC curve is plotted. In other words, an ROC curve tracks all points of sensitivity and FPR pairs at all possible threshold values. The number of points on the ROC curve is determined using the possible number of available thresholds. For continuous test results with a known distribution, there are an infinite number of cut-points from which we can choose to dichotomize the distributions of cases and controls. In this situation, the resultant ROC curve is smoothed with an infinite number of points.

To mathematically define an ROC curve, let Y be a random variable denoting the value of a test modality with the convention that higher values of Y are correlated with a higher probability of being a case. Let S denote the GS status such that $S = 1$ indicates a case and $S = 0$ indicates a control. Y_D and Y_C represent test results for cases and controls, respectively. When a cut-point value c is used to dichotomize the distribution of Y , the sensitivity and FPR are mathematically given by:

Figure 1. Distribution of test result for cases/controls and the corresponding ROC curve



$Sensitivity = Pr(Y > c \mid S = 1) = S_D(c) = 1 - F_D(c)$, and

$$FPR = Pr(Y > c \mid S = 0) = S_C(c) = 1 - F_C(c),$$

where $F_D(c)$ and $F_C(c)$ represent the distribution functions of Y_D and Y_C , respectively. If we denote the FPR at cut-point c as $FPR(c) = S_C(c) = v$, then the ROC function is defined as the corresponding sensitivity at $FPR = v$. Mathematically, this is a composite function of S_D and S_C in the following form:

$$ROC(v) = S_D(c) = S_D(S_C^{-1}(v)),$$

where $S_C^{-1}(\cdot)$ denotes the inverse function of S_C such that $S_C^{-1}(v) = c$ is equivalent to $S_C(c) = v$.

Figure 1 shows the process of making an ROC curve, and the relationship between the ROC curve and distributions of test modality Y . To recap, Y_D and Y_C are used to denote the test results for cases and controls, respectively. We assume that the test modality here is the predicted risk score from a regression model, cases represent companies that would default, and controls are companies that would not default. In Figure 1, the distribution for controls is located to the left of the distribution for cases because cases usually have higher estimated risk scores. There is an overlapping of the two distributions as represented by the shaded area. At a given threshold c , the sensitivity of the test can be visually represented by the area under the distribution of Y_D from c to the maximal value of Y_D . This sensitivity is shown on the y-axis of the ROC curve for the corresponding risk score (the right panel of Figure 1). In addition, the shaded area on the right side of c represents the probability that Y_C is greater than c , which is the FPR of the risk score and corresponds to the x-axis of the ROC curve. As the cut-points are changed over an infinite number of all possible test values, a continuous ROC curve from point (0, 0) to (1, 1) within the unit square is obtained.

ROC curves are simple and straightforward graphical tools that convey comprehensive infor-

mation about classification accuracies. Because an ROC curve is a plot across all possible decision thresholds, it does not require the specification of threshold values to generate the curve (Zweig and Campbell, 1993). It is independent of the number of cases that occur among the whole population, and it is invariant with respect to monotone transformations that keep the rank of the test score for each individual. ROC curves are monotonic and non-decreasing. An ROC curve that is located closer to the upper and left boundary of the unit square indicates better performance of the test.

Area Under the Curve

The area under the ROC curve (AUC) is an indicator of the overall accuracy of classification tools. The AUC can be interpreted in the following three ways (Zhou, Obuchowski & McClish, 2002): (1) the average sensitivity over all possible values of specificity; (2) the average specificity across all possible values of sensitivity; and (3) the overall concordance of the test values with the gold standard. Specifically, the AUC equals the probability that an underlying test value Y will assign a greater probability of being a case than that of being a control (Greenland, 2008). In this sense, the AUC is equivalent to the concordance C -statistic proposed by Harrell (2001).

Choosing the Optimal Decision Threshold

Test modalities with continuous values, by themselves, do not usually constitute compelling evidence that leads directly to either a positive or a negative decision (Swets, 1992). Rather, a cut-point needs to be specified in order to define a positive or negative decision for or against the condition being classified. In practice, a cut-point can be set anywhere along the range of the test scale such that values above the cut-point uniformly lead to a positive decision, and values below it result in a negative decision. Although an ROC curve is plotted across all possible cut-points of the test,

accuracy parameters (i.e., sensitivity and specificity) cannot be determined unless a cut-point is fully specified. This implies that the resultant test accuracies are linked to the testing threshold used in the study (Zweig & Campbell, 1993).

Choosing a particular decision threshold that is best for a given purpose or for a specific situation is not a trivial task, since many aspects of the classification need to be considered and balanced. In the area of health economics, for example, it has been argued that all related economics techniques and principles should be used to identify a testing threshold (Phelps & Mushlin, 1988; Zweig & Campbell, 1993; Laking, Lord & Fischer, 2006). First, sensitivity and specificity, as performance criteria, are highly correlated (Greenland, 2008; Sanghera, Orlando & Roberts, 2013). Thus, they cannot be considered separately, and the optimal threshold should have an appropriate balance between sensitivity and specificity (Zweig & Campbell, 1993). Consider the breast cancer screening test. While high sensitivity is required in such tests, a low specificity (or high FPR) usually cannot be tolerated, since otherwise a large proportion of healthy women will be referred to invasive confirmative tests such as biopsy. In addition, cut-points should be specified in the most cost-effective way. Many authors proposed to use costs associated with each classification rate as the major criteria. For example, Greenland (2008) pointed out that costs are a crucial element in the optimization process, and proposed the use of a cost-related loss function to define decision rules and specify optimal thresholds. In health economics studies, many authors have suggested that decision criteria should incorporate not only the cost of misclassifications but also information regarding cases and controls. Phelps and Mushlin (1988) proposed a criterion called “expected value of information” (VOI), which is a comprehensive cost-related parameter that combines decision accuracies and epidemiologic information about disease prevalence. The VOI method defines the net benefit of a decision related to a certain disease, which is the difference between the total utility of

a decision and the total cost of the decision. Then, a formula is applied to ensure that the optimal net benefit of a test modality is identified for a given disease prevalence (Altman & Bland, 1994). The results are then used to create a tangent on the ROC curve, and the point where the tangent and the ROC curve meet is the optimal cut-point. This process can be extended to accommodate a range of prevalence values, sensitivities, and specificities.

Extension to Time-Dependent ROC Analysis

When the binary decision space is a function of time, the traditional ROC analysis introduced in previous sections will no longer be adequate. For example, in assessing the performance of credit scoring systems for detecting default risk, a company’s default status can change over time. In order to accommodate this situation and obtain appropriate evaluations, a time-dependent ROC analysis should be used to assess the accuracies of the credit scoring system in a time-varying fashion (e.g., to assess the accuracy for predicting default status three years after the loan is granted). In medical decision making, time-dependent ROC analyses can be applied to evaluate the prognostic accuracy of various biomarkers for the early detection of cancers. If a prognostic tool is clinically accurate enough, doctors could make decisions based on the prognostic modality and give special treatment priority to patients with bad prognostic results. Hu & Zhou (2010) and Hu (2013) summarized the up-to-date statistical methods in time-dependent ROC analysis, and readers who are particularly interested in this issue are referred to their paper for an overview.

Applications

ROC analyses can also be used to evaluate a certain component or parameter in a credit scoring system or a prediction model. Garanin et al (2014) utilized ROC curve analysis for assessing the performance of credit decision making by testing the classifica-

tion accuracy among loan applicants. The authors used the logistic regression model to evaluate the probability of default, and used the ROC curve to test credit scoring and used logistic regression model to test the variables selection in scorecard (Garanin, Lukashevich, & Salkutsan, 2014). Vlasselaer and colleagues used the area under the ROC curve to evaluate the benefit of including certain predictor variables in the Anomaly Prevention using Advanced Transaction Exploration (AP-ATE) model that united both intrinsic features and net-work based features (Vlasselaer et al., 2015). The goal of using the APATE method is to better detect fraudulent credit card transactions by analyzing the customer past spending history and behaviors into useful meaningful features and make comparisons of those feature with a new incoming transaction. From deriving recency, frequency and monetary (RFM) characteristics-customer past incoming transactions and spending history, intrinsic features can be generated. The authors found that, when the model included only 9 network variable, the model reached an AUC of 0.92. However, when the model included all of the social network and RFM variables, the model reached the highest AUC scores of 0.987. They, thus, concluded that RFM variables are good predictors of fraud and there is a strong bind between intrinsic aspects and net-work based feature.

ROC curve analysis is also widely used in examining the performance and utility of a credit scoring system. Byanjankar and colleagues used ROC analysis to evaluate the neural network credit scoring system for detecting the borrowers' credit-worthiness in the peer-to-peer (P2P) lending (Byanjankar, Heikkila, & Mezei, 2015).

In addition, ROC analysis is intensively used in assessing the performance on recently proposed ensemble models for predicting the loan risk. Goyal & Kaur (2016) was seeking for optimum results in credit risk by using support vector machine (SVM), the random forest network, and

Tree model for Genetic Algorithm, and combining these three models (Ensemble model) to analyze the loan risk (Goyal & Kaur, 2016). They cross-validate the results by using 70% of their data as the training set and the rest 30% as the validation set. Accuracy of model was performed based on ROC and the AUC. The authors claimed that the accuracy rate was the highest, in terms of AUC, for the ensemble model, which optimize the result for loan risk prediction. Yao & Lian (2016) also proposed a SVM based Ensemble Model, named SVM-BRS. Their model combined the random subspace, boosting and SVM classifier. The authors also use AUC under the ROC curve as the metric to evaluate the classification accuracy of different models. The authors found that SVM-BRS has a better performance than a single model because SVM-BRS as it had the highest AUC of 77.4%. Figini and colleagues conducted the ROC analysis and use AUC as the metric to compute weights in aggregation probabilities in the weighting ensemble scheme, these authors also used the AUC for evaluating the prediction accuracy. Regis and Artes assessed credit card risks using multi-state Markov models by studying the attributes of different state transition in client-institution relationships over time, and proposed score models for different purposes (Regis & Artes, 2016). The authors compared the results of logistic regression to the results of multi-state Markov models using a Brazilian financial institution database (Regis & Artes, 2016). The category of a client was classified as bad when the credit score is lower or equal to a cut-point c , and as a good when the credit score is higher than c . They used area under the ROC curve to evaluate the quality of model. They authors found that the multi-state Markov models had the better performance than logistic regression models in predicting the risk of default risk, but logistic regression models had better performance when predicting the loan cancellation risk.

Applications of ROC Analysis in Areas Other Than Financial Decision Making

In addition to applications in business analytics and optimization, ROC analysis is widely implemented in medical prognosis, diagnosis (Zhou, Obuchowski & McClish, 2002; Pepe, 2003), and psychometrics (Swets, 1986; Dolle, Schulte-Körne, O'Leary, von Hofacker, Izat & Allgaier, 2012). ROC analyses have recently been extended to several business-related fields such as the analysis of economic evaluations (Laking, Lord & Fischer, 2006; Sutton, Cooper, Goodacre & Stevenson, 2008; Sanghera, Orlando & Roberts, 2013), assessment of financial parameter-derived classifiers for deciding if cloud computing should be used in projects (Kornevs, Minkevica & Holm, 2012), analysis of health administrative data (Nachev, Hill, Barry & Stoyanov, 2010; Benchimol, Manuel, To, Griffiths, Rabeneck & Guttmann, 2011), financial distress analysis (Shams, Sheikhi & Sheikhi, 2011), and examining customer churn predictions (Ballings, Van den Poel & Verhagen, 2012), analyzing credit risk of small and medium enterprises (SMEs) in China (Chen, Wang & Wu, 2010).

Particularly, with a growing focus on comparative effectiveness research and personalized medicine, ROC analysis will play a more important role in health care decision-making. Specific applications of ROC analysis in health care research include predictive model validation, biomarker diagnostics/prognostics, responder analysis in patient-reported outcomes, and comparisons of alternative treatment options (Alemayehu & Zou, 2012).

FUTURE TRENDS

With the rapid increase in the number of new decision making tools and risk predictive models, the assessment of test modalities will become a key element in business analytics, policy making,

and health economics. Benchimol, Manuel, To, Griffiths, Rabeneck & Guttmann (2011) claim that sensitivity, specificity, and the C-statistic (AUC) are among the areas of priority when designing and assessing the quality of validation studies. With the recent development of statistical methodologies in ROC regression (Pepe, 2003), AUC regression (Dodd & Pepe, 2003), and time-dependent ROC methods (Hu & Zhou, 2010), business analysts and policy makers can perform more specific analyses when assessing decision making (e.g., the covariate effect on classification accuracies and time-specific predictions).

CONCLUSION

Classification modalities, including decision making rules, test scoring, and risk predictive models, are quantitative tools that can help businesses make correct decisions. Assessing accuracy is the most important procedure in validating these tools. Sensitivity and specificity are two important aspects of classification accuracy. For a test modality with continuous or discrete ordinal values, an ROC curve plots its sensitivity versus FPR (1-specificity) across all possible decision thresholds, and the area under the ROC curve (AUC) summarizes the overall performance of the modality. To find the optimal threshold value is a challenging task, and often requires the consideration of several aspects of decision making. In presenting this topic to business owners, operating analysts, and public policy makers, we suggest that ROC analysis will become a fundamental tool in making important business and policy decisions.

REFERENCES

Alemayehu, D., & Zou, K. H. (2012). Applications of ROC analysis in medical research: Recent developments and future directions. *Academic Radiology*, 19(12), 1457–1464. doi:10.1016/j.acra.2012.09.006 PMID:23122565

- Altman, D. G., & Bland, J. M. (1994). Statistics notes: Diagnostic tests. 1: Sensitivity and specificity. *BMJ: British Medical Journal*, 308(6943), 1552. doi:10.1136/bmj.308.6943.1552 PMID:8019315
- Ballings, M., Van den Poel, D., & Verhagen, E. (2012). *Improving customer churn prediction by data augmentation using pictorial stimulus-choice data. In Management intelligent systems* (pp. 217–226). Springer.
- Begg, C. B. (1991). Advances in statistical methodology for diagnostic medicine in the 1980's. *Statistics in Medicine*, 10(12), 1887–1895. doi:10.1002/sim.4780101205 PMID:1805316
- Benchimol, E. I., Manuel, D. G., To, T., Griffiths, A. M., Rabeneck, L., & Guttmann, A. (2011). Development and use of reporting guidelines for assessing the quality of validation studies of health administrative data. *Journal of Clinical Epidemiology*, 64(8), 821–829. doi:10.1016/j.jclinepi.2010.10.006 PMID:21194889
- Byanjankar, A., Heikkila, M., & Mezei, J. (2015). Predicting Credit Risk in Peer-to-Peer Lending: A Neural Network Approach. *Computational Intelligence, 2015 IEEE Symposium Series on*, 719 - 725. doi:10.1109/SSCI.2015.109
- Chen, X., Wang, X., & Wu, D. D. (2010). Credit risk measurement and early warning of SMEs: An empirical study of listed SMEs in China. *Decision Support Systems*, 49(3), 301–310. doi:10.1016/j.dss.2010.03.005
- Chock, C., Irwig, L., Berry, G., & Glasziou, P. (1997). Comparing dichotomous screening tests when individuals negative on both tests are not verified. *Journal of Clinical Epidemiology*, 50(11), 1211–1217. doi:10.1016/S0895-4356(97)00122-4 PMID:9393377
- Cornell, J., Mulrow, C. D., & Localio, A. R. (2008). Diagnostic test accuracy and clinical decision making. *Annals of Internal Medicine*, 149(12), 904–906. doi:10.7326/0003-4819-149-12-200812160-00011 PMID:19075211
- Dodd, L. E., & Pepe, M. S. (2003). Semiparametric regression for the area under the receiver operating characteristic curve. *Journal of the American Statistical Association*, 98(462), 409–417. doi:10.1198/0162145030000198
- Dolle, K., Schulte-Körne, G., OLeary, A. M., von Hofacker, N., Izat, Y., & Allgaier, A. K. (2012). The Beck Depression Inventory-II in adolescent mental health patients: Cut-off scores for detecting depression and rating severity. *Psychiatry Research*, 200(2), 843–848. doi:10.1016/j.psychres.2012.05.011 PMID:22657953
- Figini, S., Savona, R., & Vezzoli, M. (2016). Corporate Default Prediction Model Averaging: A Normative Linear Pooling Approach. *Intelligent Systems in Accounting, Finance & Management*, 23(1-2), 6–20. doi:10.1002/isaf.1387
- Garanin, D. A., Lukashevich, N. S., & Salkutsan, S. V. (2014). The Evaluation of Credit Scoring Models Parameters Using Roc Curve Analysis. *World Applied Sciences Journal*, 30(8), 938–942.
- Goyal, A., & Kaur, R. (2016). Loan Prediction Using Ensemble Technique. *International Journal of Advanced Research in Computer Communication Engineering*, 5(3), 523–526.
- Greenland, S. (2008). The need for reorientation toward cost-effective prediction: Comments on 'evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond. *Statistics in Medicine*, 27(2), 199–206. doi:10.1002/sim.2995 PMID:17729377
- Hanley, J. A. (1989). Receiver operating characteristic (ROC) methodology: The state of the art. *Critical Reviews in Diagnostic Imaging*, 29(3), 307–335. PMID:2667567
- Harrell, F. E. (2001). *Regression modeling strategies: With applications to linear models, logistic regression, and survival analysis*. Springer. doi:10.1007/978-1-4757-3462-1

- He, Y. H., Chen, Y. C., Jiang, G. X., Huang, H. E., Li, R., Li, X. Y., & Cheng, Q. et al. (2012). Evaluation of anthropometric indices for metabolic syndrome in Chinese adults aged 40 years and over. *European Journal of Nutrition*, 51(1), 81–87. doi:10.1007/s00394-011-0195-2 PMID:21479941
- Hu, N. (2013). Evaluating the over-time prognostic performance of biomarkers for cancer prognoses using time-dependent receiver operating characteristic (ROC) curve. *Transl Med (Sunnyvale)*, 3(2), 45–46.
- Hu, N., & Zhou, X.-H. (2010). A review of time-dependent ROC curve for evaluating the prognosis capacity of biomarkers and semiparametric regression methods. *Proceeding of Joint Statistical Meeting*, 3336–3347.
- Kampftrath, T., & Levinson, S. (2013). Brief critical review: Statistical assessment of biomarker. *Clinica Chimica Acta*, 419, 102–107. doi:10.1016/j.cca.2013.02.006 PMID:23428592
- Kornevs, M., Minkevica, V., & Holm, M. (2012). Cloud computing evaluation based on financial metrics. *Information Technology and Management Science*, 15(1), 87–92. doi:10.2478/v10313-012-0013-8
- Laking, G., Lord, J., & Fischer, A. (2006). The economics of diagnosis. *Health Economics*, 15(10), 1109–1120. doi:10.1002/hec.1114 PMID:16652389
- Nachev, A., Hill, S., Barry, C., & Stoyanov, B. (2010). Fuzzy, distributed, instance counting, and default artmap neural networks for financial diagnosis. *International Journal of Information Technology & Decision Making*, 9(06), 959–978. doi:10.1142/S0219622010004111
- Pepe, M. S. (2003). *The statistical evaluation of medical tests for classification and prediction*. New York: Oxford University Press.
- Perneger, T. V., & Courvoisier, D. S. (2011). Exploration of health dimensions to be included in multi-attribute health-utility assessment. *International Journal for Quality in Health Care*, 23(1), 52–59. doi:10.1093/intqhc/mzq068 PMID:21084324
- Phelps, C. E., & Mushlin, A. I. (1988). Focusing technology assessment using medical decision theory. *Medical Decision Making*, 8(4), 279–289. doi:10.1177/0272989X8800800409 PMID:3185180
- Regis, D. E., & Artes, R. (2016). Using multi-state Markov models to identify credit card risk. *Production*, 26(20), 330–344. doi:10.1590/0103-6513.160814
- Sanghera, S., Orlando, R., & Roberts, T. (2013). Economic evaluations and diagnostic testing: An illustrative case study approach. *International Journal of Technology Assessment in Health Care*, 29(1), 53–60. doi:10.1017/S0266462312000682 PMID:23298526
- Shams, M. F., Sheikhi, Z., & Sheikhi, M. (2011). Application of var methodology to risk management in the stock market in iran. *International Journal of Business and Management*, 7(1), 70–77.
- Sutton, A. J., Cooper, N. J., Goodacre, S., & Stevenson, M. (2008). Integration of meta-analysis and economic decision modeling for evaluating diagnostic tests. *Medical Decision Making*, 28(5), 650–667. doi:10.1177/0272989X08324036 PMID:18753686
- Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science*, 240(4857), 1285–1293. doi:10.1126/science.3287615 PMID:3287615
- Swets, J. A. (1992). The science of choosing the right decision threshold in high-stakes diagnostics. *The American Psychologist*, 47(4), 522–532. doi:10.1037/0003-066X.47.4.522 PMID:1595983

Swets, J. A., & Pickett, R. M. (1982). *Evaluation of diagnostic systems: Methods from signal detection theory*. New York: Academic Press.

Vlasselaer, V. V., Bravo, C., Caelen, O., Eliassi-Rad, T., Akoglu, L., Snoeck, M., & Baesens, B. (2015). APATE: A novel approach for automated credit card transaction fraud detection using network-based extensions. *Decision Support Systems*, 75, 38–49. doi:10.1016/j.dss.2015.04.013

Wasserman, L. (2004). *All of statistics: A concise course in statistical inference*. Springer. doi:10.1007/978-0-387-21736-9

Yao, J., & Lian, C. (2016). A New Ensemble Model based Support Vector Machine for Credit Assessing. *International Journal of Grid and Distributed Computing*, 9(6), 159–168. doi:10.14257/ijgdc.2016.9.6.16

Zhou, X.-H., Obuchowski, N. A., & McClish, D. K. (2002). *Statistical methods in diagnostic medicine* (Vol. 712). New York: Wiley. doi:10.1002/9780470317082

Zweig, M. H., & Campbell, G. (1993). Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. *Clinical Chemistry*, 39(4), 561–577. PMID:8472349

KEY TERMS AND DEFINITIONS



AUC (C-Statistic): The area under an ROC curve that summarize the overall probability for correct classification.

Diagnostics Test: A quantitative test modality that is used to discriminate cases of interest from non-cases (controls).

False Negative Rate: The probability that a diagnostic test classify incorrectly a case as a control.

False Positive Rate: The probability that a diagnostic test incorrectly classify a control as a case.

Gold Standard: A standard that can specify the true status being evaluated without error (a.k.a., reference standard).

ROC Curve: A curve that plots a diagnostic test's sensitivity versus its false positive rate across all possible threshold values for defining positivity.

Sensitivity: The probability that a diagnostic test can correctly identify a true case.

Specificity: The probability that a diagnostic test can correctly specify a non-case.

Category D

Digital Literacy

Digital Literacy

D

Anirban Ray
UNC Wilmington, USA

INTRODUCTION

Decoding digital literacy is a descriptive act of interpreting, reinterpreting, and understanding the relationship between the terms digital and literacy in the expanding space of information and communication technologies (ICTs). While the idea of literacy reveals a long evolutionary past associated with the term literate, the construct of digital, as we use it today, is shaped by the use of the digits, 0 and 1, in the 1930s and 1940s to represent computer data—a practice that eventually came to be known as digital. With the emergence of the Internet and the Web as the dominant systems of information organization and knowledge creation, the concept of literacy was broadened from its original notion of skills in reading and writing to developing cultural, historical, social, and technical awareness—a shared assumption critical to and closely associated with the understanding of ICTs and their use as well. The shift has influenced the definition of literacy as “primarily a technology of which records are the end products” (Clancy, 1993, p. 20). Although contemporary discourse in digital literacy assumes a much expanded scope of understanding than a product view of technology, the deterministic tendencies are evident in instances in which digital literacy is viewed as a set of benchmark skills. Broadly speaking, digital literacy is couched in both “conceptual” as well as “standardized operational” definitions (Lank-shear & Knobel, 2008, p. 2), the key distinction being the former places digital literacy within the multiplicity of frameworks and models, while the latter measures and observes skills and performances that advance the “standards” of being digitally literate.

BACKGROUND

In 1981 *The Washington Post* first pioneered the concept that demanded “special skills” to use and manage computers (Warschauer 111) and invented the term “computer literacy.” Later, extension of the term “literacy” included “information literacy,” “digital literacy,” and “media literacy” to broaden the idea of skills. Paul Gilster (1997) in his pioneering book, *Digital Literacy*, popularized digital literacy as a shorthand for understanding and using information in multiple formats “from a wide range of sources presented via computers” (p.33). He operationalized and extended the term throughout the book, postulating that “digital literacy is about mastering ideas, not keystrokes” (p.1)—a call to attention between a “special kind of mindset or thinking” and “limited technical skills” (Bawden, 2008, p.19) premised on tasks and performances on the other. According to Gilster, digital literacy is about developing a critical approach toward using digital sources and forming awareness about our “expanded ability” (p.31) to connect with people and information using these sources. Over the years, digital literacy has addressed the split through skill and knowledge perspectives. Evidently, the skill construct affirms the neutrality thesis of technologies in which technologies are understood as means or instruments that need to be learned; conversely, the knowledge model ascertains technologies as more complex systems, not free of social, cultural, and political biases.

Despite these prevalent articulations, the challenges of defining digital literacy stem from a lack of consensus building among stakeholder disciplines, including education, communication studies, English, media studies, library infor-

mation studies and computing. The problem is further compounded by competing interpretive frameworks and theoretical models (Boechler et al., 2014) that stake claims on the scope and application of digital literacy. Considering the value of addressing the diverse views, the scholars have framed a discourse around digital literacy to accommodate dominant perspectives. These perspectives coalesce the domain-specific views into two broad categories that are identified as conservative, sometimes called traditional, and skeptical (Aviram & Eshet-Alkalai, 2006; Boechler et al., 2014). The former is uncritical of existing literature and accepts it in face value, privileging an instrumental view of digital literacy implicated in the notion of acquiring threshold or generic set of technical skills. This perpetuates the standardized paradigm of skill acquisition, a method common in educational institutions that aligns pedagogy through traditional conceptualizations of computer literacy (Ferrari, Punie & Redecker, 2012), information literacy (Mackey and Jacobson, 2011), and network literacy (McClure, 1994). Notwithstanding the widespread adoption of the view in curricular mapping and technology developments, the assumption is challenged as a didactic model that stabilizes teaching and learning as a set of prescriptive and durable practices that have fixed unities of time and place in which the role of technology is regarded as neutral.

The skeptical or functional approach, on the other hand, gained prevalence as a reaction to the conservative approach. The underlying thesis favors contextualization of digital literacy and by extension digital technologies, reframing digital literacy as a plural concept. As an alternative strategy, it recognizes that digital literacy cannot replace traditional learning but can enhance the learning environment. The thinking here is that the functional approach potentially erases the dichotomies between digital and print literacies by emphasizing the hidden aspects of “learning styles, multiple intelligences, personality types,” and capacities (Aviram & Eshet-Alkalai, 2006). The perspective coincides with the idea of meta-

literacy and value adds a plural approach to digital literacy discourse, facilitating a strategic inclusion of multiple critical conditions such as cross-cultural contexts (Thatcher, 2010), privacy and surveillance (Reilly, 2016), and situated learning within the wider conceptual framework.

The skeptical formulation questions the skill paradigm but also situates literacy beyond cognitive processes of reading, writing, and information seeking. In that it underlines the social dimension of literacy, emphasizing critical exchange and application of thoughts and ideas between individuals. Reframing literacy along these lines was done by a group of scholars in the 1980s and 1990s who called it “the New Literacy Studies” (NLS); there are still others who focus “on more recently developed literacy practices which are often (but not always) associated with ‘new technologies’ like computer and the Internet” (Jones & Hafner, 2012, p.13) and define it as “new literacies.” Digital literacy functions as a type of new literacies among several others, like computer literacy, Internet literacy, network literacy or hyper-literacy, and media literacy; other analogs include, Web literacy and game literacy (Buckingham, 2008); library literacy and reproduction literacy (Koltay, 2011); ICT skills, e-Skills, and ICT literacy (Lee, 2014), which all share common conceptual assumptions.

Digital literacy incorporates a strong social component reimagined through concepts like user, access, practice, consumption, interpretation, and production that gain emphasis within the contemporary literacy discourse. Importantly, there are four basic assumptions of new literacies that help to conceptualize digital literacy within a larger framework of literacy: (a) innovations in ICTs require new skills, competencies, awareness, and strategies of use; (b) new literacies develop continually as their defining technologies change (c) literacy components empower individuals as global citizens; (d) new literacies are multi-dimensional and multi-modal and their understanding positively impact our social participation (Leu et al., 2007). These assumptions underscore the critical perspective articulated by Paul Gilster (1997)

and later explained by David Bawden (2008). Similarly, others have emphasized the importance of approaching digital literacy and production from feminist perspectives (see Hawisher and Selfe, 1999; Wajcman, 2004), raising questions like online equity and subject positions, whether technological innovation equals advancement of literacy or how costs of digital technologies hinder or facilitate the learning process across social, cultural, or institutional spaces.

From a critical digital perspective, it is difficult to imagine digital literacy as a stand-alone term; the critical dimension includes the internal “faculties of analysis and judgement applied to the content, usage, and artefact of the technology” while the external critical element refers to the cultural and historical circumstances of the wider field of technology than computers (Hinrichsen and Coombs, 2013). Thus, the idea of literacy in digital literacy continues to be redefined by innovations of digital technologies and signals an opportunity to reconstruct a definition based on contextualized meaning and practices than checklist of skills.

Bawden (2008) offers an updated version of digital literacy by suggesting that “it is not sensible to reduce it to a finite number of linear stages” (p. 28). For him, the six-stage linear model for information literacy formulated by American Library Association in 1989, which had a considerable influence on the scholarship and conceptualization of digital literacy, does not adequately characterize contemporary understanding of digital literacy. Bawden formulates a digital literacy framework in the light of the changing circumstances and awareness for rapid technological innovation. It includes (a) underpinnings as literacy *per se* and computer/ICT literacy, (b) background knowledge as complex of information and nature of information resources, (c) central competencies as basic skills such as reading and understanding digital and non-digital formats, knowledge assembly, evaluation of information, etc., and (d) attitudes and perspectives as underlying concept connecting modern idea of literacy with the values of older

literacy through independent learning and moral/social responsibilities (p.29-30).

Similarly, Yoram Eshet-Alkalai (2004) encapsulates a framework with five types of digital literacy: (a) photo-visual literacy to interact with the visual-graphical interface (b) reproduction literacy to reproduce meanings by repurposing and recombining preexisting information elements; (c) information literacy to assess and evaluate the credibility of online information; (d) branching literacy to navigate the non-linear Web structure; (e) socio-emotional literacy to safeguard personal interests within the expanded scope of collaboration and networking offered by the ICTs. The framework was further updated with the real-time digital skill characterized by the ability of users to manage and respond to real-time, high-speed, and quick-response digital scenarios and genres (Eshet, 2012) such as gaming and simulation.

Both these models expand the scope of digital literacy from practice-orientated to (critical) knowledge-orientated conceptualizations. They index the idea of literacy as a discourse, embedding a complex understanding of the relationships between the tool, the environment, and the actor. The frameworks position digital literacy as a “survival skill” and explore the idea of social membership within the digital space. Most importantly, the shift from a tool-centered toward a critical knowledge orientated framework exemplifies learning as a social phenomenon external to tools or residing outside of tools and machines.

GLOBAL DIGITAL LITERACY PROGRAMMES AND POLICIES

From a policy standpoint, the emphasis placed on digital literacy training is enormous across the global societies. Initiatives by various national and international organizations, governments, and educational institutions support the continuing relevance of digital literacy. The content, strategies, and development of training programs and materials are often structured in conjunction with

the demands and needs of the target populations. For example, a macro perspective on literacy was adopted by UNESCO to guide the digital literacy policy framework that defined literacy as “a continuum of learning in enabling individuals to achieve their goals, to develop their knowledge and potential, and to participate fully in their community and wider society” (UNESCO, 2004). Further, UNESCO’s Information for All Programme³ (IFAP) recognizes digital literacy as “life skill,” outlining objectives squarely related to the more critical idea of digital literacy—“ICT skills, civic skills, learning to learn skills, and participation of adults in lifelong learning” (UNESCO, 2011). The objectives, such as these, are aptly coordinated with the support from national governments and agencies through centralized policies. Evidently, these policies vary across the geographical boundaries both from conceptualization and implementation standpoints.

In the United States, the mainstreaming of digital literacy is guided by several plans of action formulated by governmental and non-governmental agencies. According to the *Fact Sheet: Digital Literacy* report 2011, “28 percent of Americans do not use the Internet at all” (commerce.gov). Toward that end, the transformative task of building discourse communities for participating in information economy is largely facilitated by strategic interventions at K-12 levels and adult community learning centers such as the Literacy Information and Communication Systems (LINCS), a national leadership initiative of the U. S. Department of Education, Office of Career, Technical Adult Education (OCTAE), Division of Adult Education and Literacy (DAEL) (lincs.ed.gov). The compelling idea behind digital literacy education is to ensure digital inclusion by raising the quality of general awareness and critical thinking among learners—a contention echoed by the Association of American Colleges and Universities (AAC&U) in their publication, *Greater Expectations* that envisions college graduates in the role of “integrative thinkers who can see connections in seemingly disparate information

and draw on a wide range of knowledge to make decisions” (aacu.org).

Other western nations also emphasize the importance of digital education as part of their broader educational goals. For instance, the United Kingdom has in place the Digital Inclusion Strategy with the aim to reduce digital exclusion. The main implications of the plan are outlined in terms of skill acquisition, ICT connectivity, and equal accessibility. The government has adopted the Action 15 document with the stated purpose to “Collaborate with partners across public, private and voluntary sectors to help people go online” (gov.uk). On a larger scale the European Union (EU) has undertaken the Digital Literacy 2.0 funded in the “Lifelong Learning Programme” to bring Web2.0 affordances within the ambit of everyday life. The salience of digital literacy is recognized as “one of the key competences to ensure social cohesion, active citizenship and personal fulfillment” (digital-literacy2020.eu). The most comprehensive plan for enhancing citizens’ digital knowledge was envisioned by the European Commission Institute for Prospective Technological Studies. The institute’s goals are premised on the notion that “participation in the digital domain is no longer a question of ‘haves’ or ‘have nots,’ but rather an issue of competence” so much so that digital literacy is accepted as one of the eight key competencies for “Lifelong learning.” Within this context, digital literacy is identified as a “transversal key competence” that enables acquisition of other key competences like language, mathematics, learning to learn, cultural awareness, etc. The European Digital Competence Framework, the primary

instrument of implementation, contains 21 competences structured according to 5 areas: information, communication, content-creation, problem solving, and safety. A lot of emphases are laid for utilizing digital literacy for job searching process and improving employable qualifications (openeducationeuropa.eu).

Digital literacy penetration in South America presents a very different picture altogether. Ac-

cording to a World Economic Forum global technology report, the social and economic impacts resulting from ICTs are still somewhat low. Despite the efforts made by the government agencies to digitally connect its population, problems with infrastructure, political instability, regulatory policies and above all low skill base pose serious challenges (Bilbao-Osorio et al., 2013, p. 15). The ICT Development Index (IDI) 2011 of the major economies shows Argentina ranked at 56, Brazil at 60, and Mexico at 79 respectively (www.itu.int). Argentina in 2006 made a longstanding and comprehensive commitment to introduce digital literacy in educational sector especially at the secondary level. The highlights of the policy include improving computer infrastructure in classroom, training teachers to encourage adoption of digital technologies in pedagogy, developing appropriate educational content to integrate technologies, and implementing One Laptop per Child (OLPC). The Digital Literacy National Campaign, started in 2006, promotes digital literacy in conjunction with Encuentro TV channel that regularly broadcasts quality educational content to schools. Further, the creation of the national portal, www.educ.ar, facilitates teacher training, connects schools to the information superhighway, and develops and delivers digital content for education purposes (Lujambio et al., 2008, p.80-81).

Compared to South American countries, Asian economies perform better when it comes to diffusion of digital literacy. A quick survey of the Association of Southeast Asian Nations (ASEAN) reveals that all eight member-nations have progressed in the direction of digitizing social and cultural life of their citizens (Bilbao-Osorio et al., 2013, p. 15). For instance, The Digital *Saksharta Abhiyan* (DISHA) or National Digital Literacy Mission (NDLM) envisaged by the government of India in 2015 aims to bring at least one member from the household to participate and train in digital literacy. Given the complex diversity and huge population of India, the program enables the participating families to nominate one member to undergo the certificate training. In the Indian

context, digital literacy is defined as “the ability of individuals and communities to understand and use digital technologies for meaningful actions within life situations.” The introduction of the concept proceeds through two staggered levels, (a) appreciation of digital literacy which includes orientating with and operating digital devices like mobile phones and tablets, and (b) basic digital literacy which enables active citizenship through participating in e-governance (ndlm.in). In addition to government initiatives, there are numerous private agencies and organizations involved in the idea of digital literacy for the commons. The “Hole-in-The Wall Education Project (HiWEP) started by a private individual in the late 90s has over 300 learning stations for educating poor and disadvantaged children in India and Africa. Using a principle known as “Minimally Invasive Education,” HiWEP installs networked computers or learning stations in public places like streets, market places, and playgrounds with the purpose to attract and encourage young children to learn ICTs in independent, self-organizing, and sometimes collaborative environments (Mitra & Dangwal, 2010). The project has received a lot of traction among underprivileged communities where technology access is a chronic problem. HiWEP has been praised by various international organizations including UNESCO and continues to remain a crucial instrument for spreading digital literacy.

Global diffusion of digital literacy is based on the idea of strategically exposing the populations to the practical and ever changing dynamics of digital technologies. Regardless of global economic status, the digital literacy policy discourse is dominated by three fundamental optics: reach, penetration, and applicability. Reach entails closing the digital gap, penetration involves targeting to specific needs, and applicability concerns transference of competence for performing activities within social contexts. In the case of the developed nations, the quantum of applicability is very diverse and complex and is thus central to policy implementation. For most developing nations, however, the challenges are capacity

building and task prioritization and hence reach and penetration act as vital indicators. Interestingly, the global diffusion pattern reveals distinct strategies, varying across geographic, cultural, and political scales. For instance, while the EU focuses on employability aspects, the United States is more concerned with early exposures at institutional levels from K-12 upwards. Similarly in the case of South America, the emphasis lies on infrastructure while India, being a collectivist culture, attempts to create a large “family” of digital users by targeting individual households.

ISSUES AND CHALLENGES

Lack of consensus among researchers is a major hindrance to developing a theory of digital literacy that could be translated to educational contexts to serve the current generations of students. Eshet-Alkalai (2004) suggests that “indistinct use of the term causes ambiguity, and leads to misunderstanding, misconceptions, and poor communication.” In the absence of a consolidated definition, educators are faced with the challenges of curricular development and identifying proper assessment techniques consistent with students’ learning objectives; an allied problem is reorienting teachers toward embracing a philosophy of digital literacy that can ultimately address the practical and social needs of students. As the notion of literacy shifts from a text-based syntactic to a graphic and link based semantic conceptualization (Nielsen, 1993), digital literacy must be understood as a moving target because digital technologies evolve rapidly. This has led to question the assumptions whether the so called ‘digital natives’ are truly equipped to understand and use ICTs in their current iterations. The problem is more acute at the K-12 level where in absence of a unified digital literacy curriculum, the institutions are capitulating to what is referred to as the “standards movement” (Trotter, 1997 in Boechler et al., 2014) conceived by the International Society for Technology in Education (ISTE)—an

organization dedicated to “leveraging the use of technology in K-12 education” (Boechler et al., 2014). Thus, as ‘screen becomes the dominant site of texts,’ emphases are laid on measuring concepts of digital literacy through qualitative, quantitative, normative, and formative assessment structures such as, self-reported surveys, Software Recognition Test (SRT), Educational Activities Checklist (EAC), Recreational Experience Scale (RES) (see, Boechler et al., 2014); other formulations include code breaking, text-participating, text-using, and text-analyzing (see, Hinrichsen & Coombs, 2013) and operational digital literacy, usage digital literacy, communication and interaction digital literacy, and creation digital literacy (see Lee, 2014). The strategic integration of these techniques in curricular mapping remain vital as also properly training and upskilling teachers who ultimately impart classroom knowledge.

Another important aspect worth considering is digital literacy is not just an educational construct; it must also be fundamentally approached as a “social fairness issue” (Seale, 2009). The viewpoint generates awareness about digital divide and ethnocentric biases— two critical concepts surrounding the social aspect of digital literacy. In the prevailing circumstances social, political, and commercial activities are increasingly structured on ICTs and therefore for population with low digital literacy competence are at a risk of further marginalization. This divide or gap can potentially disengage individual from active citizenship, creating practical barriers for activities like accessing health and government information, public service information, engaging through social media, learning in mediated environments such as Massive Open Online Courses (MOOCs). In this connection, digital literacy education or DLE has significant implications in addressing the issue of digital divide. The primary goal of DLE is to “support learners’ knowledge and skill construction process through education and practices to enhance their digital literacy” (Lee, 2014). DLE aims to incorporate the social learning paradigm in life situations.

As indicated earlier, each society shapes and constructs definition of digital literacy according to its own social environment. However, many societies consider that other cultures must imitate their patterns of appropriation of digital technologies as “best practices” (Thatcher, 2010, p. 170)—a symptom of ethnocentric oversight. Digital literacy perceptions in global societies are influenced by numerous differentials including infrastructure, regulations, access, community participation, schools versus adult education, and individualism versus collectivism. It would therefore be a gross mistake to assume a universal approach. Consequently, effective measures can be developed to prevent ethnocentric biases in cross-cultural communications. According to Thatcher, one must be sensitized to the need to understand the rhetorical nature of the digital medium itself, to configure the characteristics of the medium to individual purpose, demands, and constraints, to assess the situation in the target culture, and finally to align to communication strategies to the expectations of target culture (p.169). This provides a functional framework that can negotiate the differences across disparate cultural configurations.

FUTURE RESEARCH DIRECTIONS

Current research in digital literacy reveals a solid direction toward developing a working definition of digital literacy and creating assessment tools for educational training. While establishing some type of functional metrics is important, it is equally important to build a corpus of scholarship addressing global discourse communities. Most research currently is confined to micro aspects of digital literacy, focusing mainly on theoretical constructs sometimes discounting the fact that in an information economy the scale of operation is not just regional or national but is global as well. Therefore, given the digital literacy landscape of

shared creativity and involvement in networked activities, it is time that conversations focused on exploring differences and trends between the west and the east, between the developed and the developing nations. Additionally, studies should refine the understanding of digital divide itself since most research on the digital divide tends to make a broad generalization using “multivariate analyses of several individual properties and aggregating them to produce properties of collectives” to support explanation (van Dijk 10). Generally, most approaches lack in conceptual clarity on one hand, and the idea of localization of use on the other. For example, in the west digital divide is defined in terms of use whereas in most other places it is still a matter of access to digital technologies. They also ignore or gloss over cultural factors and focus more on the overall use of technologies and their characteristics. Thus a repurposing of outcomes are warranted to take into consideration the more nuanced elements of digital literacy studies.

CONCLUSION

Digital literacy is multi-dimensional and no single context, culture or society has patent over its definition. Both conservative (skill) and skeptical (knowledge) orientations of digital literacy are still emerging, or at best tentative hypotheses and therefore extensive analyses are warranted before utilizing them as conclusive models. Knowledge paradigm enables individual agency by referencing one’s situation unlike the skill model where the user adopts a top-down task orientated structure. Digital literacy is indeed a crucial “life skill” whose salience cannot be overestimated as society transitions from essentially a linear to a hyper-linear mode of information processing. The digital literacy arc has shifted from its original focus on computers to technologies and to finally the idea of human agency.

REFERENCES

- Aviram, A. & Eshet-Alkai, Y. (2006). Towards a theory of digital literacy: three scenarios for the next steps. *European Journal of Open Learning, Distance and E-Learning*, 1.
- Bawden, D. (2008). Origins and concepts of digital literacy. In *Digital literacies: Concepts, policies and practices* (pp. 17-33). New York: Peter Lang Publishing.
- Bilbao-Osorio, B., Dutta, S., & Lanvin, B. (Eds.). (2013). *The Global Information Technology Report 2013 Growth and Jobs in a Hyperconnected World*. World Economic Forum. Retrieved from http://www3.weforum.org/docs/WEF_GITR_Report_2013.pdf
- Boechler, P., Dragon, K., & Wasniewski, E. (2014). Digital Literacy Concepts and Definitions. *International Journal of Digital Literacy and Digital Competence*, 5(4), 1–18. doi:10.4018/ijdlldc.2014100101
- Clanchy, M. T. (1993). *From memory to written record: England, 1066-1307* (2nd ed.). Cambridge, MA: Blackwell.
- Dijk, J. V. (2005). *The deepening divide: Inequality in the information society*. Thousand Oaks, CA: Sage Publication.
- DLit 2.0. (2015). *Digital Literacy 2020*. Retrieved from <http://www.digital-literacy2020.eu/content/sections/index.cfm>
- Eshet, Y. (2012). Thinking in the Digital Era: A Revised Model for Digital Literacy. *Issues in Information Science and Information Technology*, 9, 267–276.
- Eshet-Alkalai, Y. (2004). Digital literacy: A conceptual framework for survival skills in the digital era. *Journal of Educational Multimedia and Hypermedia*, 13(1), 93–106.
- European Commission. (2014). *A common European Digital Competence Framework for Citizens*. Retrieved from <http://openeducationeuropa.eu/sites/default/files/DIGCOMP%20brochure%202014%20.pdf>
- Ferrari, A., Punie, Y., & Redecker, C. (2012). Understanding Digital Competence in the 21st Century: An Analysis of Current Frameworks. *Lecture Notes in Computer Science 21st Century Learning for 21st Century Skills*, 79-92.
- Gilster, P. (1997). *Digital Literacy*. New York: Wiley.
- Greater Expectations. A New Vision for Learning as a Nation Goes to College. (2002). Association of American Colleges and Universities (AAC&U). Retrieved from <https://www.aacu.org/sites/default/files/files/publications/Greater-Expectations.pdf>
- Hinrichsen, J., & Coombs, A. (2014). The five resources of critical digital literacy: A framework for curriculum integration. *Research in Learning Technology*, 21(0). doi:10.3402/rlt.v21.21334
- International Telecommunication Union. (2012). *Measuring the Information Society*. Retrieved from http://www.itu.int/en/ITU-D/Statistics/Documents/publications/mis2012/MIS2012_without_Annex_4.pdf
- Jones, R. H., & Hafner, C. A. (2012). *Understanding digital literacies: A practical introduction*. London: Routledge.
- Koltay, T. (2011). The media and the literacies: Media literacy, information literacy, digital literacy. *Media Culture & Society*, 33(2), 211–221. doi:10.1177/0163443710393382
- Lankshear, C., & Knobel, M. (Eds.). (2008). *Digital literacies: Concepts, policies and practices*. New York: Peter Lang.

Leu, D. J., & Zawilinski, L. (2007). The new literacies of online reading comprehension. *New England Reading Association Journal*, 43(1). Retrieved from <http://search.proquest.com.liblink.uncw.edu/docview/206029040?accountid=14606>

LINCS. (n.d.). *Digital Literacy Initiatives*. Retrieved from <http://lincs.ed.gov/about-lincs>

Lujambio, D., Roveri, F., Fernández, C., Kozenitzky, I., Escobar, M. V., Fascendini, F.,... Dachesky, M. (2008). Argentina. *Global Information Society Watch*. Retrieved from <https://www.giswatch.org/hi/node/65>

Mackey, T. P., & Jacobson, T. E. (2010). Reframing Information Literacy as a Metaliteracy. *College & Research Libraries*, 72(1), 62–78. doi:10.5860/crl-76r1

McClure, C. R. (1994). Network literacy: A role for libraries? *Information Technology and Libraries*, 13(2), 115–125.

Mitra, S., & Dangwal, R. (2010). Limits to self-organising systems of learning-the Kalikuppam experiment. *British Journal of Educational Technology*, 41(5), 672–688. doi:10.1111/j.1467-8535.2010.01077.x

National Digital Literacy Mission. (2015). Department of Electronics and Information Technology and Ministry of Communications and Information technology, Government of India. Retrieved from <http://ndlm.in/overview-of-ndlm.html>

Nielsen, J. (1993). *Usability engineering*. San Diego, CA: Morgan Kaufman.

Seale, J. (2009, December). Digital Inclusion BETA: A Research Briefing. *The Teaching and Learning Research Programme*. Retrieved from <http://tel.ioe.ac.uk/inclusion/digital-inclusion-research-briefing/>

Thatcher, B. (2010). Understanding Digital Literacy across Cultures. In R. Spilka (Ed.), *Digital Literacy for Technical Communication*. New York: Routledge.

UNESCO. (2011). *Policy Brief: Digital Literacy in Education*. UNESCO Institute for Information Technologies in Education. Retrieved from <http://iite.unesco.org/pics/publications/en/files/3214688.pdf>

Wajcman, J. (2004). *TechnoFeminism*. Cambridge, UK: Polity Press.

Warschauer, M. (2004). *Technology and social Inclusion: Rethinking the digital divide*. Cambridge, MA: MIT Press.

ADDITIONAL READING

Amant, K. S., & Olaniran, B. A. (2011). *Globalization and the digital divide*. Amherst, NY: Cambria Press.

Dijk, J. V., & Deursen, A. V. (2014). *Digital skills: Unlocking the information society*. New York, NY: Palgrave Macmillan. doi:10.1057/9781137437037

Hartley, J. (2011). *The uses of digital literacy*. New Brunswick, NJ: Transaction.

Hawisher, G. E., Selfe, C. L., Guo, Y., & Liu, L. (2006). Globalization and agency: Designing and redesigning the literacies of cyberspace. *College English*, 68(6), 619–636. doi:10.2307/25472179

Lankshear, C., & Knobel, M. (2011). *New literacies: Everyday practices and classroom learning*. Maidenhead: Open University Press.

Strover, S. (2014). The US digital divide: A call for a new philosophy. *Critical Studies in Media Communication*, 31(2), 114–122. doi:10.1080/15295036.2014.922207

Svensson, P., & Goldberg, D. (Eds.). (2015). *Between Humanities and the Digital*. Cambridge, Massachusetts: The MIT Press.

Wiesinger, S., & Beliveau, R. (2016). *Digital Literacy: A Primer on Media, Identity, and Evolution of Technology*. New York, Bern, Berlin: Peter Lang. doi:10.3726/978-1-4331-2821-9

KEY TERMS AND DEFINITIONS

Digital Divide: Disparities among individuals, generations, societies, and cultures resulting from unequal access to and use of computer and Internet technologies and digital infrastructures. Besides suggesting a physical quantity of presence or absence of technologies within a context, the divide implies a difference in quality of use of digital technologies within the same context.

Digital Native: A post-millennial term describing individuals who are part of the digital age from birth.

Hyperliteracy: A systematic process of finding, linking, and retrieving information by developing both critical and functional knowledge of the non-sequential structure of the Web. The primary focus is on understanding the structure of the Web

as providing freedom of organizing information by means of linking of ideas in a nonlinear way.

Information Literacy: A set of competencies associated with identifying the need for information, locating appropriate information, evaluating information, and utilizing information to participate effectively in cultural and social contexts. Recognized as a lifelong process of self-directed learning, information literacy underlies the role of informed citizenship through a proper understanding and use of digital technologies for fulfilling academic, professional, and personal goals.

Reproduction Literacy: Abilities to recreate and repurpose existing digital contents including text, sound, images, graphics, and videos into a new format using digital production capabilities.

Digital Literacy for the 21st Century

D**Hiller A. Spires***North Carolina State University, USA***Casey Medlock Paul***North Carolina State University, USA***Shea N. Kerkhoff***North Carolina State University, USA*

INTRODUCTION

In the past few decades, technology has spanned the globe, connected people in a whole new way. As a result, citizens of all countries have not only had to learn to use new technology, but also learn how to interact with one another. Skills that comprise these abilities have been combined under the term “digital literacy.” The purpose of this chapter is to (a) define digital literacy and its changing nature, (b) discuss implications of digital literacy on contemporary schooling, (c) demonstrate the impact of digital literacy on digital citizenship, and (d) analyze the implications of digital literacy on educational equity.

BACKGROUND

Almost two decades ago, Gilster (1997) defined digital literacy as the “ability to understand and use information in multiple formats from a wide range of sources when it is presented via computers” (p. 1). At this time, the Internet was in its infant stages. More than a decade later with Internet usage in full swing, Fieldhouse and Nicholas (2008) asserted that terms like *literacy* and *fluency* can be used to describe how users find and evaluate information within digital environments. Digital literacy involves any number of digital reading and writing techniques across multiple media forms, including: words, texts, visual displays,

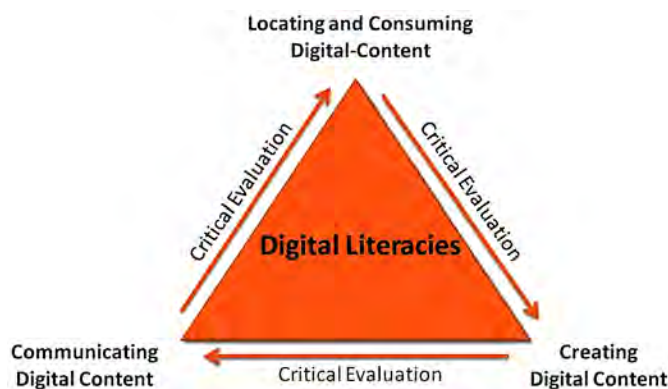
motion graphics, audio, video, and multimodal forms. In the same way that literate individuals can negotiate print text through the processes of reading and writing, literate users of technology are able to consume and produce digital compositions. There are many cognitive processes at work, along a continuum from *consumption* to *production* when a reader is immersed with digital content. The digital context is challenging for all readers due to the fluid nature of the Web and the demand for critical judgments (Spires & Estes, 2002) as the reader makes decisions about how to locate information as well how to discern the reliability and credibility of that same information.

WHAT IS DIGITAL LITERACY?

Spires and Bartlett (2012) have divided the various intellectual processes associated with digital literacy into three categories: (a) locating and consuming digital content, (b) creating digital content, and (c) communicating digital content (see Figure 1). Learners must develop evaluative dispositions as they navigate digital content. A discerning mindset is essential in order to interact with online resources with accuracy. Without critical evaluation, the learner may easily be directed by the technology rather than the learner directing the inquiry.

DOI: 10.4018/978-1-5225-2255-3.ch194

Figure 1. Digital literacy practices involve the ability to locate and consume, create, and communicate digital content, while simultaneously employing a process of critical evaluation
Adapted from Spires & Bartlett (2012)



Locating and Consuming Digital Content

It is essential to develop the skills to locate, comprehend and consume digital content on the Web. Central to being effective with the Web is strategically searching for information and evaluating its accuracy and relevancy (Leu et al., 2008). There is consensus that effective Web search skills must be developed for educational success in a digital society, and instruments such as The Teaching Internet Comprehension to Adolescents (TICA) checklist can ensure that students have the necessary prerequisite Web search skills (Leu et al., 2008). However, more challenging is how to incorporate the effective teaching and development of Web search skills in the classroom (Moraveji et al., 2011). Nevertheless, some important skills are considered necessary for locating and using digital content: domain knowledge, a working knowledge of how to use search engines, basic literacy skills, and a general knowledge of resources available on the Web (Moraveji et al., 2011). In addition to building on the ability to craft productive Web search terms, search lessons should involve direct modeling of the use of search techniques, differentiating between domain names, and querying sites for accuracy and transparency.

Creating Content

Digital content is easily created by teachers and students alike through multiple media and a variety of Web 2.0 tools. The implementation of digital content may be an important and effective method of enhancing teaching and learning (Bakkenes, Vermunt, & Wubbles, 2010), enabling teachers to embrace the 21st century skills that students are expected to master. Digital resources can also free up teachers, allowing them to spend more time facilitating student learning and less time lecturing. Allowing students to create and consume digital content in the classroom may increase engagement while also encouraging the development of skills needed for a technological society. For example, students can create video content with easy-to-use video editors such as *Animoto*, *WeVideo*, and *Powtoon*, just to name a few. Because there is a low bar for technical expertise, students can spend more time on the quality of the content rather than learning the process of a new tool. An added benefit is that the products look polished and professional. Although the creation of digital content is becoming increasingly simple, personalization of learning will require teachers to locate and utilize a variety of digital resources to meet the needs of every learner. Personalization will also

put a heavier emphasis on asking students to show mastery of learning by producing digital content. This generative process requires more time from teachers in terms of designing appropriate rubrics for performance-based learning.

Communicating Digital Content

Digital content must be communicated effectively in order to be a useful educational medium. Using social networking sites like *Facebook*, *Twitter*, and *Instagram* requires users to understand and manipulate information in multiple formats. Web 2.0 tools are social, participatory, collaborative, easy to use, and facilitate the creation of online communities. Being able to communicate digital content using mobile devices such as cellphones and tablets provides convenience and immediacy to the communication process for teachers and students. Additionally, it provides access to an infinite set of people and digital content resources globally to enrich the learning experience. This type of communication affords the possibilities of more customization and personalization for individual learners' interests and needs, which has the potential to increase student engagement in academic learning.

A popular type of digital communication is the act of curating. The capacity to curate at a sophisticated level, both in terms of content and visual appeal, is quickly becoming a necessity for educators who engage in online teaching and learning (Thompson, 2015). The word curate comes from the Latin root *Curare*, or "to cure," and historically has meant "to preserve" (Mihailidis & Cohen, 2013). As students learn to be creators and curators of digital content, there is some evidence that it contributes to their ability to be critical readers of digital texts (O'Byrne, 2012). The word curate derives from the Latin root *Curare*, or 'to cure.' To curate, historically, has meant to take charge of or organize, to pull together, sift through, select for presentation, to heal and *to preserve*. Within digital spaces, organizing and preserving online content is the purview of the individual (Mihailidis

& Cohen, 2013). This online communication trend has created a need to understand how individuals select, sort, synthesize and display content within these spaces.

The Changing Nature of Digital Literacy and Learners

Contemporary education is permeated by the millennial generation, also referred to as Generation Y and the Net Generation. This group is defined as those individuals who were born between the early to mid-1980s and the early to mid-2000s, possessing the following traits: confident, team oriented, conventional, pressured and achieving (Howe & Strauss, 2000). This generation, bigger than previous generations, is entering the workforce and contributing to a shift in our society (Winograd & Hais, 2011). This generation is immersed in a world of multimodality, or how individuals make meaning with different modes, such as print, video, speech, music, or gesture. At the heart of multimodality, is semiotics, which is the study of signs (Kress & Van Leeuwen, 1996). As society has shifted from written to visual texts in contemporary culture, more demand has been placed on teachers to learn how to make instructional changes that take these shifts into account. Leu and his colleagues (Leu et al., 2015) used the term *deictic* to refer to the changing nature of literacy, which is prompted by the constantly changing technologies within our society. By all accounts, these changes will continue to take place since the total number of Internet users is at over 3 billion worldwide and growing.

Digital Citizenship

As technology has spread across the globe, our world has become more connected than ever. This has created a global virtual world that all technology users inhabit, and as a result, technology users have had to learn how to become "digital citizens" (Isman & Canan Gungoren, 2014). Although there are various definitions of this term,

the definitions are similar; they express that first and foremost, a digital citizen must be able to use technology intelligently. Furthermore, one should understand cultural and societal issues as they relate to technology; as a result, digital citizens demonstrate various characteristics. For example, Isman and Canan Gungoren (2014) state:

[They] practice legal and ethical behavior; advocate and practice safe, legal, and responsible use of information and technology; exhibit a positive attitude toward using technology that supports collaboration, learning and productivity; demonstrate personal responsibility for lifelong learning; and exhibit leadership for digital citizenship. (p. 73)

In order to foster the development of these skills, various organizations have begun to develop models and programs designed to assist in educating people on digital citizenship. For example, ISTE published a model listing behaviors associated with digital citizenship (Brichacek, 2014). Such behaviors include “no stealing or damaging others’ digital work, identity or property;” “using digital tools to advance learning and keeping up with changing technologies;” “protecting personal information from forces that might cause harm;” and “equal digital rights and access for all” (Searson, Hancock, Soheil, & Shepherd, 2015, p. 731). Another non-profit organization, *iKeepSafe*, worked with Microsoft and AT&T to develop an online questionnaire that measures digital safety skills and attitudes in six areas, known as the BEaPRO index: balancing digital usage, practicing ethical digital usage, protecting personal information, maintaining healthy and safe relationships, building a positive reputation, and achieving online security (Searson, Hancock, Soheil, & Shepherd, 2015; *iKeepSafe*, 2015).

Still, there is much work to be done in developing global digital citizenship. The findings from *iKeepSafe*’s questionnaire indicated “although many individuals *want* to foster good digital citizenship practices, most have limited knowledge

about how to do so” (Searson, Hancock, Soheil, & Shepherd, 2015, p. 733, emphasis in original). These authors provide a list of recommendations and actions needed to help further global digital citizenship. They suggest that both national and local leadership organizations, such as public policy agencies, law enforcement, and industry leaders, work together in order to tackle the issue. Furthermore, they recommend that educational institutions begin to provide professional development for teachers in order to educate teachers as to how they can teach their students to be digital citizens. They also maintain that stakeholders must be held accountable for privacy and safety of community members, and reported incidents should inform digital citizenship education services and policy development.

Although digital citizenship is a fairly new concept, it is one that is highly important in our globalized, virtual world. It involves not only competent technology use, but also responsible and ethical use of the web. Digital citizenship is largely considered an aspect of digital literacy, and many organizations are working to understand how to include it in digital literacy education.

Digital Literacy and Educational Equity

The digital divide is a gap in access to or usage of ICTs between people, demographic groups, or countries (OECD, 2001). In other words, the global digital divide is one of access to the Internet and also one of users’ competence with ICTs. Access to ICTs continues to be divided within countries as well as among countries and is often associated with socioeconomic status. As of January 2015, only 42% of the world was active Internet users with Canada holding the highest percentage of 93% and India holding the lowest percentage of Internet users at 19% (Kemp, 2015). Access and usage are related in that lack of access leads to less practice digital literacy skills, whereas more access leads to more opportunities to practice.

Problems of access include cost of computers and subscriptions, broadband width of the Internet, and restrictedness of content (Tongia, 2005). Lack of access can be seen at the country-level, such as governments censoring content on the Internet and restricting what sources and what information citizens can obtain. Lack of access can also be seen at the demographic level when certain demographic groups are able to spend more time on the Internet than other groups. In the US, research has shown that students from underprivileged schools spent less time using ICTs even though the amount of computers and broadband width were similar across schools (Leu et al., 2015). One reason for this could be that digital literacy is not tested on government issued assessments tied to funding; therefore, time is spent on what is tested in order to score higher on the assessments and receive needed funding (Leu et al., 2015). This phenomenon has implications for future K-12 assessments.

SOLUTIONS AND RECOMMENDATIONS

Digital Literacy and the Impact on Contemporary Schooling

As technology has become more integral to students' lives, there has been an ever-increasing digital "home-school divide" (Honan, 2006, p. 41); students are using technologies outside of school that are not available in school, while educators struggle to effectively use what technology they have in their classrooms (Henderson, 2011). There is still great debate on exactly how to integrate digital literacy instruction into traditional instruction, and many studies have been and are still being conducted in an attempt to understand how best to bridge the two together (Kervin, Verenikina, Jones, & Beath, 2013; Henderson, 2011; Walsh, 2010; 2008).

Nevertheless, there is little debate on the value of these skills; many countries have begun

to reform their education programs to include better digital education. Some countries even have standards and requirements for students to achieve digital literacy. In 2008, Australia began its Digital Education Revolution in order to equip schools, teachers, and students with the technology necessary to provide a quality digital education. England has Computing Programmes of Study (United Kingdom Dept. of Education, 2013) as part of its National Curriculum, with part of its stated goal that "pupils become digitally literate—able to use, and express themselves and develop their ideas through, information and communication technology—at a level suitable for the future workplace and as active participants in a digital world" (Purpose of Study section, para. 1). The International Society for Technology in Education (ISTE; 2007) has also developed standards for students, teachers, and administrators.

Not only has digital literacy changed educational standards, but it has also changed the content that must be taught in schools. Although today's students' are often considered "digital natives" (Prensky, 2001), they are not necessarily able to use these digital tools in a knowledgeable or critical way (Jones et al., 2010). Students therefore must be taught such skills and how to use technology effectively (Leu et al., 2015), including evaluating and critically analyzing information.

Students must also be taught about cyber safety, "digital footprints," and how to be responsible online (Osborne & Connely, 2015). In fact, many educational programs are now including standards that foster the teaching of digital responsibilities, such as respecting copyright laws, using valid information, and following safe and ethical behaviors when online. (American Association of School Librarians, 2007; ISTE, 2007). Government organizations are also making sure such education is available to students. For example, Qatar's Ministry of Information and Communications Technology, known as ictQATAR (2015), works alongside teachers and parents to teach children Internet responsibility and safety.

Digital literacy has had—and is continuing to have—an impact on contemporary education. Information is readily available to students, and educators are working to teach adolescents how to use this information effectively, ethically, and responsibly. One organization, the Partnership for 21st Century Learning, was developed in order to help foster 21st century learning for students through collaborative partnerships. The 21st Century Learning Framework (Partnership for 21st Century Learning, 2009) has been used in the U.S. as well as other countries to support the inclusion of 21st century skills in education. Although educators are still trying to discover exactly how digital literacy fits into the classroom, it is clear that digital literacy has already greatly altered modern education.

FUTURE RESEARCH DIRECTIONS

Future research should focus on clarifying best practices for teaching students how to navigate digital environments effectively. Specifically, teachers need to know how to help students locate, create and communicate digital content in productive and ethical ways. Additionally, teachers need best practices for how to integrate game-based learning into their classrooms and support students as they navigate virtual spaces related to content learning. One emerging trend is Online Reading Comprehension Assessments (ORCA), in which students' capacity to conduct effective information searches is assessed in a controlled Web environment (Leu et al., 2015). Online and offline reading require different skills, so assessments must be sensitive to the distinctions.

CONCLUSION

In this chapter our aim was to provide a definition of digital literacy and how it is evolving, discuss the implications of digital literacy on contemporary schooling, demonstrate the impact of digital

literacy on digital citizenship, and analyze the implications of digital literacy on educational equity.

REFERENCES

- American Association of School Librarians. (2007). *Standards for the 21st Century Learner*. American Library Association. Retrieved from <http://www.ala.org/aasl/standards>
- Antonio, A., & Tuffley, D. (2014, July 7). Digital literacy in the developing world: A gender gap. *The Conversation*. Retrieved from <http://theconversation.com/digital-literacy-in-the-developing-world-a-gender-gap-28650>
- Bakkenes, I., Vermunt, J. D., & Wubbles, T. (2010). Teacher learning in the context of educational innovation: Learning activities and learning outcomes of experience teachers. *Learning and Instruction*, 20(6), 533–548. doi:10.1016/j.learninstruc.2009.09.001
- Brichacek, A. (2014). *Infographic: citizenship in the digital age*. Retrieved from: <https://www.iste.org/explore/articleDetail?articleid=192>
- Bunker, B. (2010). Retrieved from <http://iitp.nz/files/201001%20Digital%20Literacy%20Research%20Report.pdf>
- Coiro, J., & Dobler, E. (2007). Exploring the online reading comprehension strategies used by sixth-grade skilled readers to search for and locate information on the Internet. *Reading Research Quarterly*, 42(2), 214–257. doi:10.1598/RRQ.42.2.2
- Fieldhouse, M., & Nicholas, N. (2008). Digital literacy as information Savvy: The road to information literacy. In M. Knobel & C. Lankshear (Eds.), *Digital literacies concepts, policies and practices* (pp. 43–72). New York, NY: Peter Lang Publishing.
- Gilster, P. (1997). *Digital literacy*. New York: Wiley and Computer Publishing.

- Hargittai, E., & Walejko, G. (2008). The participation divide: Content creation and sharing in the digital age. *Information Communication and Society*, 11(2), 239–256. doi:10.1080/13691180801946150
- Henderson, R. (2011). Classroom pedagogies, digital literacy and the home-school digital divide. *International Journal of Pedagogies and Learning*, 6(2), 152–161. doi:10.5172/ijpl.2011.152
- Honan, E. (2006). Deficit discourses within the digital divide. *Engineers Australia*, 41(3), 36–43.
- Howe, N., & Strauss, W. (2000). *Millennials rising: The next great generation*. New York, NY: Vintage Books.
- ictQATAR. (2015). *Digital literacy*. Retrieved from <http://www.ictqatar.qa/en/digital-society/digital-literacy>
- iKeepSafe. (2015). *About BEaPRO*. Retrieved from: <http://ikeepsafe.org/be-a-pro/info/>
- International Society of Technology in Education (ISTE). (2007). *ISTE Standards*. Retrieved September 23, 2015, from <http://www.iste.org/standards/iste-standards/>
- Isman, A., & Canan Gungoren, O. (2014). Digital citizenship. *TOJET: The Turkish Online Journal of Educational Technology*, 13(1), 73–77.
- Jones, C., Ramanau, R., Cross, S., & Healing, G. (2010). Net generation or digital natives: Is there a distinct new generation entering university? *Computers & Education*, 54(3), 72–732. doi:10.1016/j.compedu.2009.09.022
- Kemp, S. (2015, January 21). Digital social & mobile worldwide in 2015. *We Are Social*. Retrieved from wearesocial.net/tag/statistics
- Kervin, L., Verenikina, I., Jones, P., & Beath, O. (2013). Investigating synergies between literacy, technology and classroom practice. *Australian Journal of Language and Literacy*, 36(3), 135–147.
- Kress, G., & Van Leeuwen, T. (1996). *Reading images: The grammar of visual design*. London: Routledge.
- Leu, D. J., Coiro, J., Castek, J., Hartman, D., Henry, L. A., & Reinking, D. (2008). Research on instruction and assessment in the new literacies of online reading comprehension. In C. Collins-Block, S. Parris, & P. Afferbach (Eds.), *Comprehension instruction: research based best practices* (pp. 321–346). New York: Guilford Press.
- Leu, D. J., Forzani, E., Rhoads, C., Maykel, C., Kennedy, C., & Timbrell, N. (2015). The new literacies of online research and comprehension: Rethinking the reading achievement gap. *Reading Research Quarterly*, 50(1), 37–59.
- Mihailidis, P., & Cohen, J. N. (2013). Exploring Curation as a core competency in digital and media literacy education. *Journal of Interactive Media in Education*. Retrieved from: <http://www.jime.open.ac.uk/articles/10.5334/2013-02/>
- Moraveji, N., Morris, M. R., Morris, D., Czerwinski, M., & Riche, N. (2011). ClassSearch: Facilitating the development of Web search skills through social learning. In *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems* (pp. 1797–1806). New York: ACM Press. doi:10.1145/1978942.1979203
- O’Byrne, I. (2012). *Facilitating critical evaluation skills through content creation: Empowering adolescents as readers and writers of online information* (unpublished dissertation). University of Connecticut.
- OECD. (2001). *Understanding the digital divide*. Retrieved from <http://www.oecd.org/sti/1888451.pdf>
- OECD. (2011). *PISA 2009 results: Students on the line: Digital technologies and performance*. Paris: OECD.

Osborne, N., & Connelly, L. (2015). Managing Your Digital Footprint: Possible Implications for Teaching and Learning. In *Proceedings of the 2nd European Conference on Social Media 2015: ECSM 2015* (p. 354). Academic Conferences Limited.

Partnership for 21st Century Learning. (2009). *Framework for 21st Century Learning*. Retrieved from: <http://www.p21.org/our-work/p21-framework>

Prensky, M. (2001). Digital natives, digital immigrants. *On the Horizon*, 9(5), 1–6. doi:10.1108/10748120110424816

Reynolds, R., & Chiu, M. M. (2015). Reducing digital divide effects through student engagement in coordinated game design, online resource use, and social computing activities in school. *Journal of the Association for Information Science and Technology*.

Searson, M., Hancock, M., Soheil, N., & Shepherd, G. (2015). Digital citizenship within global contexts. *Education and Information Technologies*, 20(4), 729–741. doi:10.1007/s10639-015-9426-0

Spires, H., & Bartlett, M. (2012). *Digital literacies and learning: Designing a path forward*. Friday Institute White Paper Series. NC State University.

Spires, H., & Estes, T. (2002). Reading in web-based learning environments. In C. Collins Block & M. Pressley (Eds.), *Comprehension instruction: Research-based best practices* (pp. 115–125). New York: Guilford Press.

Thompson, T. L. (2015). Digital doings: Curating work–learning practices and ecologies. *Learning, Media and Technology*, 1–21.

Tongia, R. (2005). Access to ICTs for education. In B. Bracey & T. Culver (Eds.), *Harnessing the potential of ICT for education: A multistakeholder approach* (pp. 143–152). New York: UN ICT Task Force.

United Kingdom Department of Education. (2013). *National Curriculum in England: Computing programmes of study*. Retrieved from <https://www.gov.uk/government/publications/national-curriculum-in-england-computing-programmes-of-study/>

Walsh, M. (2008). Worlds have collided and modes have merged: Classroom evidence of changed literacy practices. *Literacy*, 42(2), 101–108. doi:10.1111/j.1741-4369.2008.00495.x

Walsh, M. (2010). Multimodal literacy: What does it mean for classroom practice? *Australian Journal of Language and Literacy*, 33(3), 211–239.

Winograd, M., & Hasi, M. D. (2011). *Millennial momentum: How a new generation is remaking America*. Piscataway, NJ: Rutgers University Press.

KEY TERMS AND DEFINITIONS

Digital Citizenship: The capacity to conduct oneself in a responsible and ethical manner within public digital environments.

Digital Content: Content that uses information and communication technologies.

Digital Curation: The capacity to select, sort, synthesize and display digital content.

Digital Divide: The gap and access to or usage of ICTs between people, demographic groups or countries.

Digital Footprint: An individual's profile that is depicted to others through the Web.

Digital Literacy: The ability to locate, create, and communicate digital content.

Online Reading Comprehension: The ability to locate reliable sources on the Internet and synthesize for multiple purposes.

Web 2.0 Tools: Technology tools that allow interactivity among users and digital content.

Digital Literacy in Theory and Practice

D**Heidi Julien***State University of New York at Buffalo, USA*

INTRODUCTION

The concept of digital literacy must be understood in the context of “literacies” writ broadly. Contemporary understandings of literacy have expanded the traditional definition that includes reading and writing (possibly also including numeracy and orality), to include interpretive and creative abilities or competencies across a range of texts, in written and other forms. Text, in its contemporary sense, would include the written word, whether rendered on paper or digitally, as well as film and multi-medias. Competencies with texts of any kind are culturally situated, and therefore to be literate is to have the ability to make meaning within particular social conditions (Hoechsmann & Poyntz, 2012). Thus, meaning-making competency for economically privileged youth in a Western urban setting will differ markedly from the meaning-making by adults in a traditional agricultural milieu half-way around the globe with little access to networked communications. In Western industrialized societies, social communication practices via digital means, including interpretation, production, and dissemination, are now commonplace; the degree to which people have the abilities required to participate in these practices can be considered “digital literacy.”

BACKGROUND

Digital literacy, from a pragmatic point of view, is the set of skills, knowledge and attitudes required to access digital information effectively, efficiently, and ethically. It includes knowing how to evaluate digital information, and how to use it in decision-making. This definition is

a useful one, but it is one among many. Jaeger, Bertot, Thompson, Katz, and DeCoster (2012), for example, suggest that “digital literacy encompasses the skills and abilities necessary for access once the technology is available, including a necessary understanding of the language and component hardware and software required to successfully navigate the technology” (p. 3). For Jaeger and colleagues, digital literacy expands notions of the digital divide (a continuing challenge, even in wealthy nations), to add the ability to use technology, in addition to having access to it. They note that “digital literacy” came into its own in the 1990s, and they give credit to Gilster (1997) for moving the concept beyond the lists of information-handling skills articulated by national library associations in various countries, and for emphasizing information understanding and use. For Jaeger et al. (2012), “information literacy” is a subset of digital literacy.

Another perspective is that information literacy is the broader concept, since “information” need not be digital in format. The concept of information literacy has usually emphasized the contextual nature of information seeking, as well as the importance of information quality (Koltay, 2011). For some (e.g., Hobbs, 2010), information creation is an important aspect of digital literacy; that additional aspect relates digital literacy to the term “media literacy” which is also a commonly used term. There is no doubt that conceptual confusion is evident in this area, in which ICT (Information and Communication Technologies) literacy, computer literacy, computational literacy, technological literacy, information literacy, information fluency, digital literacy, transliteracy, and media literacy overlap in their meanings, and are employed differently by different authors and

DOI: 10.4018/978-1-5225-2255-3.ch195

agencies. As noted above, related concepts include literacy (basic reading and writing) and visual literacy, in addition to metaliteracy (a reframing of information literacy that emphasizes participatory online environments (Mackey & Jacobson, 2011)). Bawden (2008) focuses on competencies, suggesting that digital literacy consists of competency in internet searching, hypertext navigation, knowledge assembly, and content evaluation. Koltay (2011) believes that these competencies include notions of critical thinking (a traditional conceptual foundation of information literacy), knowledge assembly (collecting quality information), as well as publishing and communicating information. A broad definition of digital literacy is offered by Martin (2006, p. 19):

Digital Literacy is the awareness, attitude and ability of individuals to appropriately use digital tools and facilities to identify, access, manage, integrate, evaluate, analyse and synthesize digital resources, construct new knowledge, create media expressions, and communicate with others, in the context of specific life situations, in order to enable constructive social action; and to reflect upon this process.

Bawden (2008) notes that,

Digital literacy touches on and includes many things that it does not claim to own. It encompasses the presentation of information, without subsuming creative writing and visualization. It encompasses the evaluation of information, without claiming systematic reviewing and meta-analysis as its own. It includes organization of information but lays no claim to the construction and operation of terminologies, taxonomies and thesauri. (p. 26)

Conceptual confusion is exacerbated since the preferred term of the European Commission (2007) is media literacy, with a particular focus on critical awareness of commercially-produced information (Koltay, 2011, p. 217). A specific

emphasis on discerning the perspectives, intent, and quality of commercial information is not generally the focus of digital literacy discussions in the United States, for example. UNESCO uses the term “media and information literacy,” and it focuses on the need to empower citizens with essential knowledge about the functions of media and information systems in democratic societies. Digital and media literacy is viewed as contributing to sustainable human development, participatory civic societies, sustainable world peace, freedom, democracy, good governance, and fostering of intercultural knowledge and mutual understanding. Such lofty goals place considerable intellectual, political, and practical burdens on a concept such as digital literacy. From the UNESCO perspective, media and information literacy is core to freedom of expression and information, empowering citizens to understand functions of media and other information providers, to enable critical evaluation of content, and to facilitate citizens to make informed decisions as users and producers of information and media content.

Karpati (2011), reflecting a UNESCO perspective, states that digital literacy includes “the use and production of digital media, information processing and retrieval, participation in social networks for creation and sharing of knowledge, and a wide range of professional computing skills” (p. 1), broadening the scope of this concept to include high-level technological competence. UNESCO is particularly focused on the relevance of digital literacy to enhance employability, and lifelong learning, with an obvious goal towards human economic and social development. For Karpati (2011), the most important aspects of digital literacy are “accessing, managing, evaluating, integrating, creating, and communicating information individually or collaboratively in a networked, computer supported, and web-based environment[s] for learning, working, or leisure” (p. 4). Karpati cites the UNESCO’s Annual World Report 2009, Information Society Policies (UNESCO, 2009), which focuses on the relevance of the digital divide, and digital literacy,

in developing nations. In these contexts, digital literacy is considered critical to development of basic literacy and to lifelong learning (Karpati, 2011, p. 6).

Hobbs (2010), writing in the U.S. context, also uses the term “digital and media literacy.” This term is defined very broadly to include the “full range of cognitive, emotional and social competencies that includes the use of texts, tools and technologies; the skills of critical thinking and analysis; the practice of message composition and creativity; the ability to engage in reflection and ethical thinking; as well as active participation through teamwork and collaboration” (p. 17). For Hobbs, what is particularly relevant is the capacity for digital and media literacy to empower people to critically analyze the agendas inherent in information sources, and to advocate for minority or marginalized points of view. She clearly takes a competency view of digital and media literacy, noting that the concept includes the following skills: the ability to access and share information using media and technology, the ability to critically analyze and evaluate information, the ability to create information in sophisticated ways, the ability to reflect on information and communication from an ethical perspective, and the ability to work individually or with others to share information in all contexts (personal, workplace, and at all community levels) (Hobbs, 2010, p. 19). Wohlsen (2014) reiterates the lack of a clear definition for digital literacy, but makes the point that information creation is an important element of the concept.

In the United States, digital literacy tends to be contextualized in terms of the ongoing digital divide, and so it is viewed as important for digital inclusion in communities (Institute of Museum and Library Services, 2012). Digital inclusion assumes the ability to appreciate the benefits of ICTs, and means that citizens are able to use ICTs to access educational, economic, and social opportunities. This concern for expansion of opportunity echoes that expressed by UNESCO. The American Library Association Digital Literacy

Task Force (2013) defines digital literacy as “the ability to use information and communication technologies to find, understand, evaluate, create, and communicate digital information, an ability that requires both cognitive and technical skills” (Digital Literacy Task Force, 2013, p. 2). This report expands the concept to include information stewardship, communication with others, civic participation, and democratic engagement. For the American Library Association, digital literacy is conceived in global terms, and its importance is underlined by tying digital literacy to its potential role in helping the U.S. compete in global economic, educational, and intellectual contexts. The educational link is also made between digital literacy and the Common Core State Standards Initiative for U.S. schools, which actually focuses on “media”. The Digital Literacy Task Force (2013) affirms the critical role of libraries in promoting digital literacy and in supporting programs which develop digital literacy, in partnership with other organizations and institutions.

DIGITAL LITERACY TRAINING

The importance of digital literacy is widely recognized internationally, and top-down efforts in many nations to encourage digital literacy were evident from the 1990s forward. For example, the New Zealand Computer Society has stated that digital literacy is “an essential life skill and right of every... citizen” (Bunker, 2010, p. 5). Further, the Society states that addressing ICT competence in the workforce could increase productivity by \$1.7 billion (in time saved); that ICT competence improves employment opportunities, overcomes isolation, builds confidence and leads to further learning; and, it recommends that national governments should take leadership roles in developing digital literacy among citizens. Digital literacy training initiatives around the world fall on a continuum: basic initiatives, which are the most common, focus on developing basic computing skills applied to everyday simple

tasks. A second level, found less often, focuses on using basic functionality of key applications (word processing, spreadsheets, presentation tools, email, web searching). At the third level, which is rarely found, the focus is on developing confident use of digital tools and facilities to identify, access, manage, integrate, evaluate, analyze, and synthesize digital resources; construct new knowledge; create media expressions; use the net for transactions; and, develop awareness of security issues. Most digital literacy initiatives are derived from centralized policies at national or regional levels, and are strategically linked with government objectives such as developing an information society, and bridging the digital divide. Therefore, centralized initiatives tend to focus on disadvantaged groups such as the elderly, disabled, or unemployed. Other digital literacy initiatives focus on related strategic goals, such as social cohesion, immigrant integration, supporting lifelong learning, and supporting optimal use of online government services. Usability and accessibility, as components of digital literacy, tend not to be emphasized. Often, digital literacy initiatives are partnerships between governments and other institutions. Non-governmental bodies partnering on such initiatives tend to be motivated by a desire to help disadvantaged groups, while private companies may well be motivated by perceived opportunities to grow their market share of products, such as broadband in rural areas, hardware or software purchase by the elderly or disabled, and improving workforce competencies (Shapiro, 2009).

A recent analysis of 470 digital literacy initiatives and survey data in the EU ("Digital Literacy European Commission Working Paper", n.d.) found that digital literacy programs have grown, especially for young people. In addition, the analysis determined that more efforts are required to develop digital literacy skills among older and disadvantaged groups, and that developing trust and confidence in digital transactions remains challenging. This report notes that investment in digital literacy programs have shown positive

outcomes, including expanding internet access and use, encouraging non-users or less-skilled users, and providing training based on user preferences for informal learning approaches and focusing on daily life activities and interests. The report concludes that there remains a significant proportion of the population that does not use the Internet, and that potential exists to focus on differences in quality of use.

In the United States, digital literacy is being supported by a number of organizations and agencies. An online portal, DigitalLiteracy.gov (<http://www.digitalliteracy.gov/>), provides resources to communities and organizations to support local digital literacy initiatives. Individuals can also access training resources on that portal to develop their own digital literacy skills. Digital literacy efforts in schools have been spurred by legislation. In the U.S., the No Child Left Behind Act and the related Enhancing Education Through Technology Act require technological literacy for all children, which enhances digital literacy efforts, especially for children from families without online access (American Library Association, 2012, para. 23). Also in the education context, UNESCO has been active in discussions about technology in education (*cf.* UNESCO's Grünwald Declaration on Media Education, 1982), has worked to develop international media education guidelines, and has developed a media and information literacy curriculum (American Library Association, 2012, para. 20). There is little question that digital literacy is important to education at all levels, and is especially relevant in inquiry-based learning. Many universities worldwide have incorporated digital literacy outcomes into strategic planning efforts, although the degree to which these outcomes are evident in their graduates remains uncertain.

There has been a significant growth in resources to support digital literacy learning, provided by a range of public and private organizations and partnerships. Examples include the Public Library Association's Digital Learn hub (<http://digitallearn.org/>), MediaSmarts in Canada (<http://mediasmarts.ca/>), the LinkAmerica's Founda-

tion Digital Literacy site (<http://www.ictliteracy.info/#>), Microsoft's Digital Literacy site (<https://www.microsoft.com/en-us/digitalliteracy/overview.aspx>) and Google's Digital Literacy and Citizenship Curriculum (<https://www.google.com/goodtoknow/web/curriculum/>). Some sites focus on internet safety for children and youth, such as the UK Safer Internet Centre (<http://www.saferinternet.org.uk/>). U.S. Digital Literacy (n.d.) is an example of a professional development resource designed by teachers for teachers. The site provides definitions and a multiplicity of learning opportunities for teachers who wish to learn about this topic.

CHALLENGES

Digital literacy is recognized as critical to positive health outcomes (of particular importance when so much health information is now obtained online), workforce development, and participative governance (since participatory citizenship is dependent upon relatively sophisticated information finding skills). Digital literacy goes beyond social networking, and increasingly, governments are delivering information and services online, and online only, which requires citizens to be at least minimally digitally literate in order to access that information. In addition, participation in the 'commons' and in 'civil society' depends on citizens' ability to find and evaluate information. Digital literacy is also recognized as an essential competency for job performance, since information gathering, manipulation, and application are key work tasks. Those without good digital literacy skills will be marginalized in private and public life, including employment. A recent report (Head, 2012) suggests that employers in the U.S. are not generally pleased with the digital literacy skill set exhibited by employees newly graduated from university. These young people rely on superficial information searches, lack the skills and perseverance to conduct sophisticated and in-depth information searches, and fail to bring

information from a variety of sources together in useful ways.

Of increasing concern is the growing recognition that digital literacy skills are not developed through experience alone. The key role for formal digital literacy training efforts is twofold: to ameliorate the digital divide, and to emphasize the role of critical analysis in communicative practice (Hoechsmann & Poyntz, 2012, 147). Effective and efficient information-finding skills take time and effort to learn. Information is organized in complex ways, and can be difficult to evaluate. There is growing consensus that, for most people, confidence in information skills exceeds actual skill level. Skills deficits are especially apparent for effective information finding skills and information evaluation skills. Many people do not understand the context of information—how or why it is produced, nor the purposes for which different types of information are made available. Thus, critical evaluation is difficult. Research shows that students entering post-secondary education typically are “surf savvy” but not “search savvy,” and many students, and people in general, do not understand how to evaluate the information they find (Nicholas et al., 2009). Authority is assessed within seconds by dipping and cross-checking across different sites and by relying on favored brands (e.g., Google). The speed of web searching indicates that little time is spent evaluating information for relevance, accuracy or authority. In addition, many people do not understand (or respect) ethical boundaries on using others' ideas and writing, relying on cut and paste techniques to bring disparate information together. It seems clear that there is significant potential for learning institutions at all levels to play important roles to expand the interpretive repertoires of their learners, and to develop a “questioning and reflective approach that recognizes the social and cultural implications of the technologies, institutions, and texts” (Hoechsmann & Poyntz, 2012, 149).

If many people with access to computers and networks remain digitally illiterate, how can this status quo be addressed? Presumably, digital

literacy skills can be learned from teachers in school; however, despite curricular mandates in many jurisdictions, and the presence of certified school librarians in some schools, we know that actual skills continue to be low (Julien & Barker, 2009; Gross & Latham, 2012). Teachers are not necessarily digital literate themselves, so it may be unrealistic to expect them to impart skills they do not yet possess. Teachers' classroom time also is limited, and digital literacy skills tend not to be tested. What is not tested is less likely to be emphasized. When parents are digitally literate, children may learn these skills, but many parents lack this knowledge to pass along. Often, opportunities to develop digital literacy skills exist at post-secondary educational institutions, but instruction is not systematic, may be very limited, and may not be done well (Julien, 2006). The same concerns apply to opportunities at public libraries and other community centers.

Other challenges to developing widespread digital literacy skills include unfounded beliefs about the relative skills and understandings of so-called "digital natives," and assumptions about the value of experience with ICTs without formal learning of skills. Competence with social media or quick Google searches does not necessarily translate into sophisticated information evaluation skills. Another challenge is assumptions about the capacity of libraries to play significant roles in developing citizens' digital literacy skills. Libraries often face severe resource challenges, library administrators may place relatively little emphasis on client training in digital literacy, library customers may lack confidence in the potential for librarians to contribute to digital literacy training, and librarians may be poorly prepared for instructional work (Julien & Hoffman, 2008). In most jurisdictions there is also limited coordination between school teachers and librarians in public libraries and academic libraries, as well as insufficient numbers of teacher-librarians in schools; in many jurisdictions globally, librarian positions in schools have been eliminated entirely. Where potential exists for librarians across contexts to work together to develop community capacity in

digital literacy, actual cooperation or coordination is rare.

FUTURE RESEARCH DIRECTIONS

Research in digital literacy focuses largely on the ways in which it can be further developed among specific populations or in specific contexts, and there remains substantial work to do in this area. For instance, there are nuances related to the learning and expression of digital literacy that will differ between adults and children, and between disparate cultural, educational, and workplace settings. Certainly a focus on outcomes of digital literacy education is also warranted, and should be a significant concern for societies globally. In addition, the digital literacy landscape will evolve with changes in information technologies, and these changes will merit research attention.

CONCLUSION

Despite these challenges, the list of potential benefits arising from individual and community digital literacy is lengthy, and the value of digital literacy is significant. Some perspectives and agendas focus on overtly political outcomes of widespread digital literacy, including enhancing democracy, world peace, and empowering previously marginalized groups politically and socially. Digital literacy certainly has the potential to contribute to far-reaching and important personal and societal consequences. Thus, increasing focus on development of digital literacy, however defined, should be a policy priority for all sectors.

REFERENCES

American Library Association. (2012). *Digital literacy – ALA Digital Literacy Task Force draft report*. Retrieved November 13, 2015 from http://connect.ala.org/files/94226/digilitreport2012_COMMENT%20DRAFT_9%2018%2012.pdf

Bawden, D. (2008). Origins and concepts of digital literacy. In C. Lankshear & M. Knobel (Eds.), *Digital literacies: Concepts, policies and practices* (pp. 17–32). New York: Peter Lang.

Bunker, B. (2010). *A summary of international reports, research and case studies of digital literacy*. New Zealand Computer Society Inc. Retrieved November 13, 2015 from <http://www.iitp.org.nz/files/201001%20Digital%20Literacy%20Research%20Report.pdf>

Digital Literacy European Commission Working Paper and Recommendations from Digital Literacy High-Level Expert Group. (n.d.). Retrieved November 13, 2015 from www.ifap.ru/library/book386.pdf

Digital Literacy Task Force. Office for Information Technology Policy. American Library Association. Conclusions and Recommendations for Digital Literacy Programs and Libraries. (2013). Retrieved November 13, 2015 from <http://www.districtdispatch.org/2013/06/ala-task-force-releases-recommendations-to-advance-digital-literacy/>

European Commission. (2007). *A European approach to media literacy in the digital environment*. Retrieved November 13, 2015 from <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM%3A2007%3A0833%3AFIN%3AEN%3APDF>

Gilster, P. (1997). *Digital literacy*. New York: Wiley.

Gross, M., & Latham, D. (2012). What's skill got to do with it?: Information literacy skills and self-views of ability among first-year college students. *Journal of the American Society for Information Science and Technology*, 63(3), 574–583. doi:10.1002/asi.21681

Head, A. (2012). *Learning curve: How college graduates solve information problems once they join the workplace*. Project Information Literacy Research Report, October 16, 2012. Retrieved November 13, 2015 from http://projectinfoit.org/images/pdfs/pil_fall2012_workplacestudy_fullreport_revised.pdf

Hobbs, R. (2010). *Democracy and media literacy: A plan of action*. Boulder, CO: Aspen Institute.

Hoechsmann, M., & Poyntz, S. R. (2012). *Media literacies: A critical introduction*. Chichester, UK: Blackwell Publishing. doi:10.1002/9781444344158

Institute of Museum and Library Services. University of Washington, International City/County Management Association. (2012 January). *Building digital communities: A framework for action*. Washington, DC: Institute of Museum and Library Services. Retrieved November 13, 2015 from http://www.imls.gov/assets/1/AssetManager/BuildingDigitalCommunities_Framework.pdf

Jaeger, P. T., Bertot, J. C., Thompson, K. M., Katz, S. M., & DeCoster, E. J. (2012). The intersection of public policy and public access: Digital divides, digital literacy, digital inclusion, and public libraries. *Public Library Quarterly*, 31(1), 1–20. doi:10.1080/01616846.2012.654728

Julien, H. (2006). A longitudinal analysis of information literacy instruction in Canadian academic libraries. *Canadian Journal of Information and Library Science*, 29(3), 289–313.

Julien, H., & Barker, S. (2009). How high school students evaluate scientific information: A basis for information literacy skills development. *Library & Information Science Research*, 31(1), 12–17. doi:10.1016/j.lisr.2008.10.008

Julien, H., & Hoffman, C. (2008). Information literacy training in Canada's public libraries. *The Library Quarterly*, 78(1), 19–41. doi:10.1086/523908

Karpati, A. (May 2011). *Policy brief: Digital literacy in education*. UNESCO Institute for Information Technologies in Education. Retrieved November 13, 2015 from <http://iite.unesco.org/pics/publications/en/files/3214688.pdf>

Koltay, T. (2011). The media and the literacies: Media literacy, information literacy, digital literacy. *Media Culture & Society*, 33(2), 211–221. doi:10.1177/0163443710393382

Mackey, T. P., & Jacobson, T. E. (2011). Reframing information literacy as a metaliteracy. *College & Research Libraries*, 72(1), 62–78. doi:10.5860/crl-76r1

Martin, A. (2006). Literacies for the digital age. In A. Martin & D. Madigan (Eds.), *Digital literacies for learning* (pp. 3–25). London: Facet.

Nicholas, D., Huntington, P., Jamali, H. R., Rowlands, I., & Fieldhouse, M. (2009). Student digital information-seeking behaviour in context. *The Journal of Documentation*, 65(1), 106–132. doi:10.1108/00220410910926149

Shapiro, H. (2009). *Supporting digital literacy public policies and stakeholder initiatives. Topic report 4. Conclusions and recommendations based on reviews and findings*. Danish Technological Institute. Retrieved November 13, 2015 from https://joinup.ec.europa.eu/sites/default/files/files_epractice/sites/Topic%20Report%204%20-%20Conclusions%20and%20recommndations%20based%20on%20reviews%20and%20findings.pdf

UNESCO. (1982). *The Grünwald declaration on media education*. Retrieved from http://www.unesco.org/education/pdf/MEDIA_E.PDF

UNESCO. (2009). *Information society policies. Annual world report*. Retrieved from http://portal.unesco.org/ci/en/files/29547/12668551003ifap_world_report_2009.pdf/

U.S. Digital Literacy. (n.d.). Retrieved November 13, 2015 from <http://digitalliteracy.us/>

Wohlsen, M. (2014). Digital literacy is the key to the future, but we still don't know what it means. *Wired*. Retrieved November 13, 2015 from <http://www.wired.com/2014/09/digital-literacy-key-future-still-dont-know-means/>

ADDITIONAL READING

Andreae, J., & Anderson, E. L. (2011). Re-conceptualizing access. *Communications in Information Literacy*, 5(2), 74–81.

Bawden, D. (2001). Information and digital literacies: A review of concepts. *The Journal of Documentation*, 57(2), 218–259. doi:10.1108/EUM0000000007083

Beagle, D. D. (2012). The emergent information commons: Philosophy, models, and 21st century learning paradigms. *Journal of Library Administration*, 52(6-7), 518–537. doi:10.1080/01930826.2012.707951

Bernsmann, S., & Croll, J. (2013). Lowering the threshold to libraries with social media. The approach of Digital Literacy 2.0, a project funded in the EU Lifelong Learning Programme. *Library Review*, 62(1), 53–58. doi:10.1108/00242531311328168

Bradley, C. (2013). Information literacy in the programmatic university accreditation standards of select professions in Canada, the United States, the United Kingdom and Australia. *Journal of Information Literacy*, 7(1), 44–68. doi:10.11645/7.1.1785

Catts, R. (2012). Indicators of adult information literacy. *Journal of Information Literacy*, 6(2), 4–18. doi:10.11645/6.2.1746

Clark, L., & Visser, M. (2011). Digital literacy takes center stage. *Library Technology Reports*, 47, 38.

Feezell, J. T., Kahne, J., & Lee, N.-j. (2012). Digital media literacy education and online civic and political participation. *International journal of communication (Online)*, 1+.

- Ferran-Ferrer, N., Minguillón, J., & Pérez-Montoro, M. (2013). Key factors in the transfer of information-related competencies between academic, workplace, and daily life contexts. *Journal of the American Society for Information Science and Technology*, 64(6), 1112–1121. doi:10.1002/asi.22817
- Ferro, E., Gil-Garcia, J. R., & Helbig, N. C. (2011). The role of IT literacy in defining digital divide policy needs. *Government Information Quarterly*, 28(1), 3–10. doi:10.1016/j.giq.2010.05.007
- Ferro, E., Helbig, N. C., & Gil-Garcia, J. R. (2011). The role of IT literacy in defining digital divide policy needs. *Government Information Quarterly*, 28(1), 3–10. doi:10.1016/j.giq.2010.05.007
- Gee, J. P. (2012). Digital games and libraries. *Knowledge Quest*, 41, 60.
- Gross, M., & Latham, D. (2012). What skill got to do with it?: Information literacy skills and self-views of ability among first-year college students. *Journal of the American Society for Information Science and Technology*, 63(3), 574–583. doi:10.1002/asi.21681
- Gutiérrez-Martín, A. G., & Tyner, K. (2012). Media literacy in multiple contexts. *Communication: Journal for Communication Sciences in Southern Africa*, 19(38), 10–12. doi:10.3916/C38-2012-02-00
- Hamilton, B. J. (2009). Transforming information literacy for nowgen students. *Knowledge Quest*, 37, 48.
- Hicks, A. (2013). Cultural shifts. *Communications in Information Literacy*, 7(1), 50–65.
- Hobbs, R. (2011). *Digital and media literacy: Connecting culture and classroom*. Thousand Oaks: Sage.
- Jaeger, P. T., Bertot, J. C., Thompson, K. M., Katz, S. M., & DeCoster, E. J. (2012). The intersection of public policy and public access: Digital divides, digital literacy, digital inclusion, and public libraries. *Public Library Quarterly*, 31(1), 1–20. doi:10.1080/01616846.2012.654728
- Jayakar, K., & Park, E.-A. (2012). Funding public computing centers: Balancing broadband availability and expected demand. *Government Information Quarterly*, 29(1), 50–59. doi:10.1016/j.giq.2011.02.005
- Jones, R. H., & Hafner, C. A. (2012). *Understanding digital literacies: An introduction*. New York: Routledge.
- Knutsson, O., Blåsjö, M., Hållsten, S., & Karlström, P. (2012). Identifying different registers of digital literacy in virtual learning environments. *The Internet and Higher Education*, 15(4), 237–246. doi:10.1016/j.iheduc.2011.11.002
- Lankshear, C., & Knobel, M. (Eds.). (2008). *Digital literacy: Concepts, policies and practices*. New York: Peter Lang Publishing.
- Lowe, M. (2012). Information literacy 2011: A selection of 2011's literature on IL. *Codex (2150-086X)*, 1(4), 46–74.
- Mills, K. A. (2010). A review of the digital turn in the new literacy studies. *Review of Educational Research*, 80(2), 246–271. doi:10.3102/0034654310364401
- Saunders, L. (2012). Faculty perspectives on information literacy as a student learning outcome. *Journal of Academic Librarianship*, 38(4), 226–236. doi:10.1016/j.acalib.2012.06.001
- Tyner, K. (2009). Media literacy and the tyranny of the narrative. *Afterimage*, 37, 3.
- Weiner, S. (2011). Information literacy and the workforce: A Review. *Education Libraries*, 34(2), 7–14.

KEY TERMS AND DEFINITIONS

Digital Divide: Inequalities between people with access to digital technologies and those without such access. Access may include access to hardware, software, internet connections, and possessing the skill set needed to make use of these technologies.

Digital Literacy: The set of skills, knowledge and attitudes required to access, create, use, and evaluate digital information effectively, efficiently, and ethically.

Digital Native: A person who has interacted with digital technology for most of his or her life.

Encouraging Digital Literacy and ICT Competency in the Information Age

D**Kijpokin Kasemsap***Suan Sunandha Rajabhat University, Thailand*

INTRODUCTION

The growing prominence of the Internet as educational tool requires research regarding learners' digital literacy (Greene, Yu, & Copeland, 2014). Nowadays, students autonomously acquire their digital literacy and are adept at using various ICT tools to enrich their daily leisure lives (Ting, 2015). Digital literacy includes the ability to search for information and to integrate that information while monitoring progress toward achieving educational goals (Bråten, Britt, Strømsø, & Rouet, 2011). Digital natives often engage themselves in the use of ICT tools and in accessing, creating, and sharing both text and videos on the Web 2.0 (Junco, 2012). The ability of digital natives to embrace ICT suggests that they possess a certain level of digital literacy (Ng, 2012).

Competency refers to the ability resulting from individual's knowledge, skills, characteristics, and attitude in executing work to achieve success (Malinina, 2015). ICT plays a critical role in enhancing the quality of education (Vitanova, Atanasova-Pachemska, Iliev, & Pachemska, 2015). Within the context of 21st century skills, the importance of being digitally competent is reflected in the international and national policies for the educational ICT utilization (Kozma, 2008). These policies for educational ICT utilization have introduced ICT competency in the national and school curricula (Aesaert, Vanderlinde, Tondeur, & van Braak, 2013), such as the integration of ICT competences in the educational curricula. ICT competency standards practically define the achievement expectations for students (Thomas & Knezek, 2008).

This article aims to bridge the gap in the literature on the thorough literature consolidation of digital literacy and ICT competency. The extensive literatures of digital literacy and ICT competency provide a contribution to practitioners and researchers in order to maximize the impact of digital literacy and ICT competency in the information age.

BACKGROUND

Technology Acceptance Model (TAM), such as Unified Theory of Acceptance Use of Technology (UTAUT), explains the degree of acceptance of the utilization of information technology (IT) toward adopting the technological infrastructure (Nchunge, Sakwa, & Mwangi, 2013). TAM helps managers and decision makers to evaluate the success of the acceptance of technology to the organization, and motivate users to accept the systems. UTAUT identifies four key factors (i.e., performance expectancy, effort expectancy, social influence, and facilitating conditions) and four moderators (i.e., age, gender, experience, and voluntariness) concerning behavioral intention toward utilizing technology in organizational contexts (Venkatesh, Thong, & Xu, 2016).

Digital literacy refers to the variety of literacies associated with the use of new technologies (Mohammadyari & Singh, 2015). Digital literacy is a fundamental life skill in today's knowledge economy and information society (Bawden, 2001). Digital literacy constitutes new practices rather than new instances of established practices (Simpson & Obdalova, 2014). Proficiency in

digital literacy refers to the ability to read and write using online sources, and includes the ability to select sources relevant to the task, synthesize information into a coherent message, and communicate the message with an audience (Bulger, Mayer, & Metzger, 2014). Appel (2012) defined digital literacy as the ability to find and analyze information by using computers.

Digital literacy is a broad concept encompassing the different aspects, and its development follows a continuum from the acquisition of instrumental skills to that of strategic competence and cognitive skills (Calvani, Fini, Ranieri, & Picci, 2012). Digital literacy is the awareness, attitude and ability of individuals to appropriately utilize the digital tools to identify the digital resources, construct the new knowledge, create the media expressions, and communicate with others (Martin, 2005). Hatlevik and Christophersen (2013) used the term digital competence to describe the acquisition and processing of digital information and the ability to produce the digital information.

Competency is made up of knowledge, skills, and attitude (Malinina, 2015). ICT competency is considered as the educational outcomes (Thomas & Knezek, 2008). ICT competency refers to knowledge, skills, and ability to take advantage of ICT for the purpose of gathering, processing, and presenting the information in support of activities among different groups of people (Albirini, 2006). Traditional methods of teaching ICT are not an effective way for learners to acquire ICT competencies or to gain more positive ICT perceptions (Goktas, Yildirim, & Yildirim, 2008). Instead, learners should interact with new information in ways that enable the active inquiry to promote the useful learning (Daugherty, 2005). To gain ICT competency, learners should be given opportunities to create their own meaning-making processes in order to establish their own knowledge (Goktas & Demirel, 2012).

ADVANCED ISSUES OF DIGITAL LITERACY AND ICT COMPETENCY

This section emphasizes the overview of digital literacy and ICT competency, the encouragement of digital literacy in the information age, and the encouragement of ICT competency in the information age.

Overview of Digital Literacy

Twenty-first century learning skills require the ability to use the Internet technology (Greene et al., 2014). The prominent role the Internet plays in home and classroom lives demands careful attention to its link to student knowledge gains (Greene et al., 2014). The Internet, as a text, consists of multiple print, images, videos, and interactive simulations, all used to communicate with the subsequent effects upon cognition (Collins & Halverson, 2009). While it is important to consider how the Internet utilization affects the students' different cognitive processes (Reinking, 2005), it is important to consider how different cognitive processes influence how students engage with the Internet (Strømsø & Bråten, 2010).

Digital literacy is an important determinant to consider as the number of electronic learning (e-learning) tools has expanded to incorporate the Web 2.0 innovations, such as blogs, podcasts, and wikis (Mohammadyari & Singh, 2015). Digital literacy is a significant determinant of attitudes toward computer-assisted language learning (Oz, Demirezen, & Pourfeiz, 2015). Ullrich et al. (2008) stated that the rapid spread of these tools has meant that individuals often have had to train themselves in how to use these tools. Individuals with a high level of digital literacy have been better able to leverage these new tools to self-manage their training and execute their continuing education activities in an informal setting, toward reducing the interruption to their working lives (Hargittai, 2010).

Besides the technical awareness, digital literacy includes the social and cognitive skills required in the digital environment (Huerta & Sandoval-Almazan, 2007). Eshet-Alkalai (2004) stated that digital literacy includes five skills: photo-visual skills, reproduction skills, branching skills, information skills, and socio-emotional skills. The key motivator behind the growing use of these tools has been the ability to quickly incorporate the material about new developments in a field into the training material, which is important for the fields, which are affected by technology-related issues, such as privacy, security, and standards (Arbaugh & Duray, 2002).

Overview of ICT Competency

ICT competency refers to the learning-process oriented competence used in the complex, authentic, and unpredictable situations, and is underpinned by the technical and application ICT knowledge and skills (Aesaert et al., 2013). Markauskaite (2007) indicated that ICT competency refers to the interactive use of cognitive capabilities and technical capabilities in order to accomplish the cognitive information and ICT-based tasks. Digital information processing and digital communication are recognized as ICT competencies to be measured because these are identified as two themes in the national and international ICT frameworks (Voogt & Roblin, 2012).

Application of ICT initiates the new opportunities in arranging the educational environment (Malinina, 2015). Various types of professional development programs concerning ICT implementation have been organized for in-service teachers toward upgrading ICT competency among teachers and bringing the change to their teaching practices, such as integrating ICT in classroom (Borko, 2004). In order to produce the effective learning through ICT utilization, students should develop the technological competencies, and teachers should develop the teaching, learning, and technological competencies (Pineida, 2011).

Encouragement of Digital Literacy in the Information Age



Digital literacy refers to the skills and knowledge in using computers in a hypermedia environment (Ting, 2015). Digital literacy refers to the cognitive processes that individuals partake in during the utilization of computer-based, multimodal information (Mayer, 2005). Digital literacy, as a complex frame of specific diversified capabilities, represents an actual phenomenon within the social matrix, sourced in the developing potential of the digital technology and the required information literacy (Javorský & Horváth, 2014). Institutions need to place greater value on digital literacy, and better prepare their students and their own organizational processes to thrive in an age of digital knowledge practices (Littlejohn, Beetham, & McGill, 2012).

In terms of the ways in which digital literacy is acquired, adolescents, in particular, engage in a broad range of computerized activities, including doing homework, searching and gathering information on the Internet, using social media platforms to communicate with friends, watching videos on YouTube, and playing educational video games (Appel, 2012). Educational computer games can motivate students to develop the basic competencies and encourage challenging themselves to be better and learn the additional knowledge related to the important tasks (Kasemsap, 2017a). Social media tools are open to anyone, whereas reaching the traditional media often requires a lot of money and a good network of media industry contacts (Kasemsap, 2016a). The use of social media has created the highly effective communication platforms where any user, virtually anywhere in the world, can freely create the content and disseminate this information in real time to a global audience (Kasemsap, 2017b).

For ICT tools, especially for those intended for entertainment purposes, students are often more skilled and adept at using them than their teachers, such as when the ICT tools are used to edit videos and upload them to Facebook or Youtube (Gu, Zhu,

& Guo, 2013). Students autonomously learn the instrumental skills and knowledge of computers and the Internet outside formal education (Eynon & Malmberg, 2011). Because students are raised in such the networked digital environment, their patterns of thinking and communication, notions of learning, needs for control, and even their personal and social values have been shaped by this networked digital environment (Gu et al., 2013). There has been an increasing interest in the ways that young people are using the Internet and other new technologies in their everyday lives and how such use may enhance the informal and formal learning opportunities (Lim, Zhao, Tondeur, Chai, & Tsai, 2013).

The advent of a knowledge-based society, where economic wealth depends on individuals' ability to deal with the abundance of information and to adapt to an ever-changing working environment, makes digital literacy an obvious concept for examining the individual adoption of IT (Mohammadyari & Singh, 2015). Digital literacy empowers individuals to communicate with others, work more effectively, and increase the individual's productivity, particularly with those who have the same skills and proficiency levels (Martin, 2008). Digital literacy reduces the individuals' inclination to unfavorably regard their achievements (Eastin & LaRose, 2000), which should make them more confident about their expected performance.

Self-regulated learning (Winne & Hadwin, 2008) skills, inclusive of making effective plans, and controlling these plans, as well as the strategies used to enact those plans and the learning that results (Azevedo & Jacobson, 2008), are likely to be the critical components of digital literacy. Some researchers have equated digital literacy with search literacy, or searching for information online and information literacy (Hockley, 2012). For example, cognitive overload and disorientation are two primary reasons why students struggle to effectively search the Internet (Gerjets, Scheiter, & Schuh, 2008).

Encouragement of ICT Competency in the Information Age

Regarding ICT competency encouragement, blogs can be utilized as an online learning tool to provide students with access to the significant information, to enlarge the students' understanding of specific issues, and to direct students to investigate the additional materials (Luehmann, 2008). Using blogs in ICT courses can increase the students' perceived technological competencies, while promoting their dynamic engagement (Goktas & Demirel, 2012). Blogs can encourage engagement in the educational activities and involve the interactive communities of learners (Goktas & Demirel, 2012).

Individuals who are regarded as having ICT competency must be able to produce necessary documents, find out solutions to problems, and choose proper ICT tools for problem solving and effective work (Malinina, 2015). Creating and maintaining a weblog using the created blogging software is an easy process, and an instructor can use this learning format to publish the course materials and post the announcements, presentations, timetables, and other information on the Web toward increasing the students' ICT competency (Wassell & Crouch, 2008).

FUTURE RESEARCH DIRECTIONS

The classification of the extensive literature in the domains of digital literacy and ICT competency will provide the potential opportunities for future research. Human capital is the collective skills, knowledge, or other intangible assets of individuals that can be used to create the economic value for the individuals, their employers, or their community (Kasemsap, 2016b). Continuing professional development is the training and education that continues throughout an individual's career in order to improve the skills and knowledge (Kasemsap, 2017c). An examination of linkages

among digital literacy, ICT competency, human capital, and continuing professional development in the workplace would seem to be viable for future research efforts.

CONCLUSION

This article highlighted the overview of digital literacy and ICT competency, the encouragement of digital literacy in the information age, and the encouragement of ICT competency in the information age. Ability to use technology is a critical prerequisite for understanding information communicated through that technology. Digital literacy and ICT competency are the abilities to locate, organize, understand, evaluate, and analyze information using computers and digital technology. Digital literacy and ICT competency have led to the great increases in information that can be conveniently and quickly accessed and effectively facilitate the collaboration and sharing of knowledge. While employability is an obvious driver, developing learners who can learn in the digital society is the significant role for universities and colleges.

Digital literacy and ICT competency look beyond functional IT skills to describe a richer set of digital behaviors, practices, and identities. Digital literacy and ICT competency are a set of academic practices supported by diverse learning methods, such as electronic learning and game-based learning. Encouraging discussion about supporting students with digital skills and the associated impact on the teaching staff roles helps broaden the awareness of both digital literacy and ICT competency across educational institutions. The staff and students involved do not need to be technology experts; effective communication skills, educational flexibility, and an eagerness to learn are crucial toward encouraging both digital literacy and ICT competency.

Benefits of digital literacy and ICT competency include the enhanced capacity to remain abreast of technology developments and effectively

utilize technology to increase productivity and competitiveness in all sectors of the economy and to develop the innovative networks, products, and services for the rapidly growing ICT marketplace. The encouragement of digital literacy and ICT competency is essential for modern organizations that seek to serve suppliers and customers, increase business performance, strengthen competitiveness, and achieve regular prosperity in the information age. Thus, it is required for modern organizations to encourage their digital literacy and ICT competency and develop a strategic plan to regularly investigate their practical improvements toward satisfying customer requirements. Encouraging digital literacy and ICT competency has the potential to enhance organizational performance and achieve strategic goals in the information age.

REFERENCES

- Aesaert, K., Vanderlinde, R., Tondeur, J., & van Braak, J. (2013). The content of educational technology curricula: A cross-curricular state of the art. *Educational Technology Research and Development*, 61(1), 131–151. doi:10.1007/s11423-012-9279-9
- Albirini, A. (2006). Teachers' attitudes toward information and communication technologies: The case of Syrian EFL teachers. *Computers & Education*, 47(4), 373–398. doi:10.1016/j.compedu.2004.10.013
- Appel, M. (2012). Are heavy users of computer games and social media more computer literate? *Computers & Education*, 59(4), 1339–1349. doi:10.1016/j.compedu.2012.06.004
- Arbaugh, J., & Duray, R. (2002). Technological and structural characteristics, student learning and satisfaction with web-based courses an exploratory study of two on-line MBA programs. *Management Learning*, 33(3), 331–347. doi:10.1177/1350507602333003

- Azevedo, R., & Jacobson, M. (2008). Advances in scaffolding learning with hypertext and hypermedia: A summary and critical analysis. *Educational Technology Research and Development*, 56(1), 93–100. doi:10.1007/s11423-007-9064-3
- Bawden, D. (2001). Information and digital literacies: A review of concepts. *The Journal of Documentation*, 57(2), 218–259. doi:10.1108/EUM0000000007083
- Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher*, 33(8), 3–15. doi:10.3102/0013189X033008003
- Bråten, I., Britt, M. A., Strømsø, H. I., & Rouet, J. F. (2011). The role of epistemic beliefs in the comprehension of multiple expository texts: Toward an integrated model. *Educational Psychologist*, 46(1), 48–70. doi:10.1080/00461520.2011.538647
- Bulger, M. E., Mayer, R. E., & Metzger, M. J. (2014). Knowledge and processes that predict proficiency in digital literacy. *Reading and Writing*, 27(9), 1567–1583. doi:10.1007/s11145-014-9507-2
- Calvani, A., Fini, A., Ranieri, M., & Picci, P. (2012). Are young generations in secondary school digitally competent? A study on Italian teenagers. *Computers & Education*, 58(2), 797–807. doi:10.1016/j.compedu.2011.10.004
- Collins, A., & Halverson, R. (2009). *Rethinking education in the age of technology: The digital revolution and schooling in America*. New York, NY: Teachers College Press.
- Daugherty, A. L. (2005). Motivating online information literacy students. *Academic Exchange Quarterly*, 9(4), 286–289.
- Eastin, M. S., & LaRose, R. (2000). Internet self-efficacy and the psychology of the digital divide. *Journal of Computer-Mediated Communication*, 6(1), 25–56.
- Eshet-Alkalai, Y. (2004). Digital literacy: A conceptual framework for survival skills in the digital era. *Journal of Educational Multimedia and Hypermedia*, 13(1), 93–106.
- Eynon, R., & Malmberg, L. E. (2011). A typology of young people's Internet use: Implications for education. *Computers & Education*, 56(3), 585–595. doi:10.1016/j.compedu.2010.09.020
- Gerjets, P., Scheiter, K., & Schuh, J. (2008). Information comparisons in example-based hypermedia environments: Supporting learners with processing prompts and an interactive comparison tool. *Educational Technology Research and Development*, 56(1), 73–92. doi:10.1007/s11423-007-9068-z
- Goktas, Y., & Demirel, T. (2012). Blog-enhanced ICT courses: Examining their effects on prospective teachers' ICT competencies and perceptions. *Computers & Education*, 58(3), 908–917. doi:10.1016/j.compedu.2011.11.004
- Goktas, Y., Yildirim, Z., & Yildirim, S. (2008). A review of ICT related courses in pre-service teacher education programs. *Asia Pacific Education Review*, 9(2), 168–179. doi:10.1007/BF03026497
- Greene, J. A., Yu, S. B., & Copeland, D. Z. (2014). Measuring critical components of digital literacy and their relationships with learning. *Computers & Education*, 76, 55–69. doi:10.1016/j.compedu.2014.03.008
- Gu, X., Zhu, Y., & Guo, X. (2013). Meeting the “digital natives”: Understanding the acceptance of technology in classrooms. *Journal of Educational Technology & Society*, 16(1), 392–402.
- Hargittai, E. (2010). Digital na(t)ives? Variation in Internet skills and uses among members of the “net generation”. *Sociological Inquiry*, 80(1), 92–113. doi:10.1111/j.1475-682X.2009.00317.x

- Hatlevik, O. E., & Christophersen, K. A. (2013). Digital competence at the beginning of upper secondary school: Identifying factors explaining digital inclusion. *Computers & Education*, 63, 240–247. doi:10.1016/j.compedu.2012.11.015
- Hockley, N. (2012). Technology for the language teacher: Digital literacies. *English Language Teaching Journal*, 66(1), 108–112. doi:10.1093/elt/ccr077
- Huerta, E., & Sandoval-Almazan, R. (2007). Digital literacy: Problems faced by telecenter users in Mexico. *Information Technology for Development*, 13(3), 217–232. doi:10.1002/itdj.20071
- Javorský, S., & Horváth, R. (2014). Phenomenon of digital literacy in scope of European cross-curricular comparison. *Procedia: Social and Behavioral Sciences*, 143, 769–777. doi:10.1016/j.sbspro.2014.07.468
- Junco, R. (2012). The relationship between frequency of Facebook use, participation in Facebook activities, and student engagement. *Computers & Education*, 58(1), 162–171. doi:10.1016/j.compedu.2011.08.004
- Kasemsap, K. (2016a). Utilizing social media in modern business. In I. Lee (Ed.), *Encyclopedia of e-commerce development, implementation, and management* (pp. 2171–2182). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-9787-4.ch156
- Kasemsap, K. (2016b). Analyzing the roles of human capital and competency in global business. In S. Sen, A. Bhattacharya, & R. Sen (Eds.), *International perspectives on socio-economic development in the era of globalization* (pp. 1–29). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-9908-3.ch001
- Kasemsap, K. (2017a). Mastering educational computer games, educational video games, and serious games in the digital age. In R. Alexandre Peixoto de Queirós & M. Pinto (Eds.), *Gamification-based e-learning strategies for computer programming education* (pp. 30–52). Hershey, PA: IGI Global. doi:10.4018/978-1-5225-1034-5.ch003
- Kasemsap, K. (2017b). Professional and business applications of social media platforms. In V. Benson, R. Tuninga, & G. Saridakis (Eds.), *Analyzing the strategic role of social networking in firm growth and productivity* (pp. 427–450). Hershey, PA: IGI Global. doi:10.4018/978-1-5225-0559-4.ch021
- Kasemsap, K. (2017c). Encouraging continuing professional development and teacher professional development in global education. In R. Cintron, J. Samuel, & J. Hinson (Eds.), *Accelerated opportunity education models and practices* (pp. 168–202). Hershey, PA: IGI Global. doi:10.4018/978-1-5225-0528-0.ch008
- Kozma, R. (2008). Comparative analysis of policies for ICT in education. In J. Voogt & G. Knezek (Eds.), *International handbook of information technology in primary and secondary education* (pp. 1083–1096). New York, NY: Springer-Verlag. doi:10.1007/978-0-387-73315-9_68
- Lim, C. P., Zhao, Y., Tondeur, J., Chai, C. S., & Tsai, C. C. (2013). Bridging the gap: Technology trends and use of technology in schools. *Journal of Educational Technology & Society*, 16(2), 59–68.
- Littlejohn, A., Beetham, H., & McGill, L. (2012). Learning at the digital frontier: A review of digital literacies in theory and practice. *Journal of Computer Assisted Learning*, 28(6), 547–556. doi:10.1111/j.1365-2729.2011.00474.x
- Luehmann, A. L. (2008). Using blogging in support of teacher professional identity development: A case study. *Journal of the Learning Sciences*, 17(3), 287–337. doi:10.1080/10508400802192706
- Malinina, I. (2015). ICT competencies of foreign languages teachers. *Procedia: Social and Behavioral Sciences*, 182, 75–80. doi:10.1016/j.sbspro.2015.04.740
- Markauskaite, L. (2007). Exploring the structure of trainee teachers' ICT literacy: The main components of, and relationships between, general cognitive and technical capabilities. *Educational Technology Research and Development*, 55(6), 547–572. doi:10.1007/s11423-007-9043-8

- Martin, A. (2005). DigEuLit – a European framework for digital literacy: A progress report. *Journal of eLiteracy*, 2(2), 130–136.
- Martin, A. (2008). Digital literacy and the “digital society”. In C. Lankshear & M. Knobel (Eds.), *Digital literacies: Concepts, policies and practices* (pp. 151–176). New York, NY: Peter Lang.
- Mayer, R. E. (2005). Cognitive theory of multimedia learning. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 31–48). New York, NY: Cambridge University Press. doi:10.1017/CBO9780511816819.004
- Mohammadyari, S., & Singh, H. (2015). Understanding the effect of e-learning on individual performance: The role of digital literacy. *Computers & Education*, 82(1), 11–25. doi:10.1016/j.compedu.2014.10.025
- Nchunge, D. M., Sakwa, M., & Mwangi, W. (2013). Assessment of ICT infrastructure on ICT adoption in educational institutions: A descriptive survey of secondary schools in Kiambu County Kenya. *Journal of Computer Science & Information Technology*, 1(1), 32–45.
- Ng, W. (2012). Can we teach digital natives digital literacy? *Computers & Education*, 59(3), 1065–1078. doi:10.1016/j.compedu.2012.04.016
- Oz, H., Demirezen, M., & Pourfeiz, J. (2015). Digital device ownership, computer literacy, and attitudes toward foreign and computer-assisted language learning. *Procedia: Social and Behavioral Sciences*, 186, 359–366. doi:10.1016/j.sbspro.2015.04.028
- Pineida, F. O. (2011). Competencies for the 21st century: Integrating ICT to life, school and economical development. *Procedia: Social and Behavioral Sciences*, 28, 54–57. doi:10.1016/j.sbspro.2011.11.011
- Reinking, D. (2005). Multimedia learning of reading. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 355–374). New York, NY: Cambridge University Press. doi:10.1017/CBO9780511816819.024
- Simpson, R., & Obdalova, O. A. (2014). New technologies in higher education – ICT skills or digital literacy? *Procedia: Social and Behavioral Sciences*, 154, 104–111. doi:10.1016/j.sbspro.2014.10.120
- Strømsø, H. I., & Bråten, I. (2010). The role of personal epistemology in the self-regulation of Internet-based learning. *Metacognition and Learning*, 5(1), 91–111. doi:10.1007/s11409-009-9043-7
- Thomas, L. G., & Knezek, D. G. (2008). Information, communications, and educational technology standards for students, teachers, and school leaders. In J. Voogt & G. Knezek (Eds.), *International handbook of information technology in primary and secondary education* (pp. 333–348). New York, NY: Springer-Verlag. doi:10.1007/978-0-387-73315-9_20
- Ting, Y. L. (2015). Tapping into students’ digital literacy and designing negotiated learning to promote learner autonomy. *The Internet and Higher Education*, 26, 25–32. doi:10.1016/j.iheduc.2015.04.004
- Ullrich, C., Borau, K., Luo, H., Tan, X., Shen, L., & Shen, R. (2008). *Why Web 2.0 is good for learning and for research: Principles and prototypes*. Paper presented at the 17th International World Wide Web Conference (WWW 2008), Beijing, China. doi:10.1145/1367497.1367593
- Venkatesh, V., Thong, J. Y. L., & Xu, X. (2016). Unified theory of acceptance and use of technology: A synthesis and the road ahead. *Journal of the Association for Information Systems*, 17(5), 328–376.

Vitanova, V., Atanasova-Pachemska, T., Iliev, D., & Pachemska, S. (2015). Factors affecting the development of ICT competencies of teachers in primary schools. *Procedia: Social and Behavioral Sciences*, 191, 1087–1094. doi:10.1016/j.sbspro.2015.04.344

Voogt, J., & Roblin, N. P. (2012). A comparative analysis of international frameworks for 21st century competences: Implications for national curriculum policies. *Journal of Curriculum Studies*, 44(3), 299–321. doi:10.1080/00220272.2012.668938

Wassell, B., & Crouch, C. (2008). Fostering connections between multicultural education and technology: Incorporating weblogs into preservice teacher education. *Journal of Technology and Teacher Education*, 16(2), 211–232.

Winne, P., & Hadwin, A. (2008). The weave of motivation and self-regulated learning. In D. Schunk & B. Zimmerman (Eds.), *Motivation and self-regulated learning: Theory, research, and applications* (pp. 297–314). Mahwah, NJ: Lawrence Erlbaum Associates.

ADDITIONAL READING

Adetimirin, A. E. (2012). ICT literacy among undergraduates in Nigerian universities. *Education and Information Technologies*, 17(4), 381–397. doi:10.1007/s10639-011-9163-y

Apperley, T., & Walsh, C. (2012). What digital games and literacy have in common: A heuristic for understanding pupils' gaming literacy. *Literacy*, 46(3), 115–122. doi:10.1111/j.1741-4369.2012.00668.x

Bierstaker, J., Janvrin, D., & Lowe, D. J. (2014). What factors influence auditors' use of computer-assisted audit techniques? *Advances in Accounting*, 30(1), 67–74. doi:10.1016/j.adiac.2013.12.005

Davies, R. S. (2011). Understanding technology literacy: A framework for evaluating educational technology integration. *TechTrends*, 55(5), 45–52. doi:10.1007/s11528-011-0527-3

Duncan, L. G., McGeown, S. P., Griffiths, Y. M., Stothard, S. E., & Dobai, A. (2016). Adolescent reading skill and engagement with digital and traditional literacies as predictors of reading comprehension. *British Journal of Psychology*, 107(2), 209–238. doi:10.1111/bjop.12134 PMID:26094956

Ellison, T. L. (2015). Artifacts as stories: Understanding families, digital literacies, and storied lives. *Journal of Adolescent & Adult Literacy*, 59(5), 511–513. doi:10.1002/jaal.506

Ferro, E., Helbig, N. C., & Gil-Garcia, J. R. (2011). The role of IT literacy in defining digital divide policy needs. *Government Information Quarterly*, 28(1), 3–10. doi:10.1016/j.giq.2010.05.007

Fisher, D., & Frey, N. (2015). Checking for understanding digitally during content area learning. *The Reading Teacher*, 69(3), 281–286. doi:10.1002/trtr.1407

Gastelú, C. A. T., Kiss, G., & Domínguez, A. L. (2015). Level of ICT competencies at the university. *Procedia: Social and Behavioral Sciences*, 174, 137–142. doi:10.1016/j.sbspro.2015.01.638

Gil-Flores, J., Torres-Gordillo, J. J., & Perera-Rodríguez, V. H. (2012). The role of online reader experience in explaining students' performance in digital reading. *Computers & Education*, 59(2), 653–660. doi:10.1016/j.compedu.2012.03.014

Hatlevik, O. E., Ottestad, G., & Throndsen, I. (2015). Predictors of digital competence in 7th grade: A multilevel analysis. *Journal of Computer Assisted Learning*, 31(3), 220–231. doi:10.1111/jcal.12065

- Hung, Y. H., Chang, R. I., & Lin, C. F. (2015). Survey of software literacy, behavior and personal traits of freshmen accounting majors. *British Journal of Educational Technology*, 46(5), 1064–1069. doi:10.1111/bjet.12333
- Kim, G. M. (2016). Transcultural digital literacies: Cross-border connections and self-representations in an online forum. *Reading Research Quarterly*, 51(2), 199–219. doi:10.1002/rrq.131
- Lin, S., & Overbaugh, R. C. (2013). Autonomy of participation and ICT literacy in a self-directed learning environment (SDLE). *Quality & Quantity*, 47(1), 97–109. doi:10.1007/s11135-011-9505-2
- Lin, S., Shih, T. H., & Lu, R. (2013). ICT proficiency and gender: A validation on training and development. *International Journal of Technology and Design Education*, 23(2), 179–190. doi:10.1007/s10798-011-9173-5
- Lockley, T. (2013). Answers to outstanding questions about Japanese student ICT competencies and a glance into a mobile future. *The Asia-Pacific Education Researcher*, 22(4), 603–617. doi:10.1007/s40299-013-0063-3
- Manderino, M., & Castek, J. (2016). Digital literacies for disciplinary learning: A call to action. *Journal of Adolescent & Adult Literacy*, 60(1), 79–81. doi:10.1002/jaal.565
- Oliver, G., Chawner, B., & Liu, H. P. (2011). Implementing digital archives: Issues of trust. *Archival Science*, 11(3), 311–327. doi:10.1007/s10502-011-9167-9
- Ozsevec, T. (2011). Computer literacy of Turkish preservice teachers in different teacher training programs. *Asia Pacific Education Review*, 12(1), 13–21. doi:10.1007/s12564-010-9123-z
- Park, E. Y., & Nam, S. J. (2014). An analysis of the digital literacy of people with disabilities in Korea: Verification of a moderating effect of gender, education and age. *International Journal of Consumer Studies*, 38(4), 404–411. doi:10.1111/ijcs.12107
- Price-Dennis, D., Holmes, K. A., & Smith, E. (2015). Exploring digital literacy practices in an inclusive classroom. *The Reading Teacher*, 69(2), 195–205. doi:10.1002/trtr.1398
- Rambousek, V., Štípek, J., & Wildová, R. (2015). ICT competencies and their development in primary and lower-secondary schools in the Czech Republic. *Procedia: Social and Behavioral Sciences*, 171, 535–542. doi:10.1016/j.sbspro.2015.01.158
- Scuotto, V., & Morellato, M. (2013). Entrepreneurial knowledge and digital competence: Keys for a success of student entrepreneurship. *Journal of the Knowledge Economy*, 4(3), 293–303. doi:10.1007/s13132-013-0155-6
- Spires, H. A. (2015). Digital game-based learning: What's literacy got to do with it? *Journal of Adolescent & Adult Literacy*, 59(2), 125–130. doi:10.1002/jaal.424
- van Tryon, P. J. S. (2013). The instructional designer's role in forming university-community partnerships in digital literacy. *TechTrends*, 57(1), 52–58. doi:10.1007/s11528-012-0631-z
- Voogt, J., Erstad, O., Dede, C., & Mishra, P. (2013). Challenges to learning and schooling in the digital networked world of the 21st century. *Journal of Computer Assisted Learning*, 29(5), 403–413. doi:10.1111/jcal.12029
- Zimmermann, A., & Ravishankar, M. N. (2014). Knowledge transfer in IT offshoring relationships: The roles of social capital, efficacy and outcome expectations. *Information Systems Journal*, 24(2), 167–202. doi:10.1111/isj.12027

KEY TERMS AND DEFINITIONS

Blog: The website, similar to an online journal, that includes chronological entries made by individuals.

Competency: The important skill that is needed to do a job.

Information Technology: The set of tools, processes, and associated equipment employed to collect, process, and present the information.

Internet: The worldwide computer network that provides information on very many subjects and enables users to exchange messages.

Knowledge: The state of knowing about or being familiar with something.

Learning: The activity of obtaining knowledge.

Literacy: The knowledge of the particular subject, or the particular type of knowledge.

Technology: The use of scientific knowledge to solve practical problems, especially in industry and commerce.

D

Information Needs of Users in the Tech Savvy Environment and the Influencing Factors

Mudasir Khazer Rather
University of Kashmir, India

Shabir Ahmad Ganaie
University of Kashmir, India

INTRODUCTION

Information environment is enough rich, characterized by a growth in information sources as well as providers, a variety of approaches and techniques for accessing information, and a redundancy of content from multiple sources. In this “overloaded” information environment, many information users tend to experience a sense of insufficiency in locating the precise information which leads to anxiety. In this complex information environment, understanding the way individuals choose to satisfy their information needs takes on new urgency. Insight into information seeking can be gained by understanding how users seek information sources and how they locate the desired information to meet their needs (Chandra, Lynn, Lawrence & Lillie, 2007). The concept of information needs was coined by an American information scientist Robert S. Taylor in his article “The Process of Asking Questions” published in *American Documentation* (Now is *Journal of the American Society of Information Science and Technology*). There are many definitions of information need. According to Case (2012) information need is a recognition that your knowledge is inadequate to satisfy a goal that you have. He explains that “having information” is not the same as “being informed.” Information need is one of the cognitive needs of humanity. Information need determines information-seeking behavior and these concepts harmonize one another.

Information need is influenced by a number of factors. It is revealed from the literature that ‘information scattered in too many sources’ and that too in multi-formats is the problem often faced by users. For fulfilling the information needs, users access different sources of information Sources. Scientists, engineers and technologists in general use encyclopedias, handbooks, textbooks, periodicals, abstracts, indexes, standards, patents, etc. for their research and development works. He showed that information needs of scientists, engineers and technologists are equally based on their knowledge about those sources of information and accessibility of these information sources (Gayatri, 2006). Post Graduate students and Research Scholars mostly use journals, library books and textbooks for completing their course work (Fidzani, 1998). The information needs of teachers were found to be mostly related to guidance on administrative procedures, having lesson plans ready, mechanisms for evidencing work, etc (Williams, 2005). Further, information needs of the General people are found to be varied. The areas in which they needed information are diverse. These range from the information needs of the farmers, to that of the petty traders, artisans, blacksmiths, weavers, painters, fishermen, postmasters, labors, adult learners etc (Kadli & Kumbar, 2011). This chapter provides an overview of information needs of users, their types and also the various factors influencing the information needs of users in the digital age.

DOI: 10.4018/978-1-5225-2255-3.ch197

BACKGROUND

Developments in the present tech savvy information world have been witnessed only since when the communication of information has been speedy and rapid. Information is the fundamental unit of communication. Communication is the transmission of information in the form of signs and symbols which gave birth to the concept of data. Data is a codified and communicable symbolic representation of entities, properties and their states. They have content (representation) and form (record) that allow storage, retrieval, transfer, aggregation, and analysis. Data can turn into information if they are put into a context and given meaning. Information is a string of signs and symbols that can be interpreted as a message. It can be transmitted in the form of signals. Information is any sort of event that changes the state of a dynamic system. The meaning of this concept varies in various contexts. Information is closely correlated to notion of data, message, knowledge, wisdom, meaning understanding, perception, communication etc (Silvio, 2006). According to some authors, data are understood to be symbols that have not yet been interpreted, information is data with meaning, and knowledge is what facilitates people to allocate meaning and thereby generate information. Data have generally been taken as simple facts that can be structured to develop information. Information is the English word which is apparently derived from the Latin stem (*informatio*): this noun is derived from the verb “informare” (to inform) in the sense of “to give form to the mind”, “to discipline”, “instruct”, “teach”. *Inform* itself comes from the Latin verb *informare*, which means to give form, or to form an idea of. Furthermore, Latin itself already contained the word *information* meaning concept or idea, but the extent to which this may have influenced the development of the word *information* in English is not clear (Wikipedia, 2013).

A number of authors have given their views and opinions regarding the concept of information. According to Cawkell, (2003), “Information is an assemblage of data in a comprehensible form

capable of communication and use”. While as Martin (1995) believes Information is that which adds to or modifies knowledge structure. Singh (2007) reveals that Information seems to be everywhere. Information is being encoded in the genes, disseminated by media of communication, exchanged in conversation [discussion], contained in all sorts of things, libraries are overflowing with it, institutions are bogged down by it, and people are overloaded with it, still no one seems to know exactly what information is. Adeoti-Adekeye (1997) explains that there are three major fields of information which have traditionally been divided and separated. The first is the literature field of libraries and archives, where information has been put into recorded form. The second is the document field of information centers and record centers, where information has been collected and organized but perhaps not seriously evaluated in the same sense as in the literature field. The third information field is the data field of computers, telecommunications and automated information systems where the information is often numerical or structured.

Data vs. Information

Data refers to raw, unevaluated facts, figures, symbols, objects, events, etc. Data may be a collection of facts lying in storage, like a telephone directory or census records.

Information is data that have been put into a meaningful and useful context and communicated to a recipient who uses it to make decisions. Information involves the communication and reception of intelligence or knowledge. It appraises and notifies surprises and stimulates, reduces uncertainty, reveals additional alternatives or helps eliminate irrelevant or poor ones, and influences individuals and stimulates them to action. An element of data may constitute information in a specific context; for example, when you want to contact your friend, his or her telephone number is a piece of information; otherwise, it is just one element of data in the telephone directory (Babu, Singh & Sachdeva, 2013).

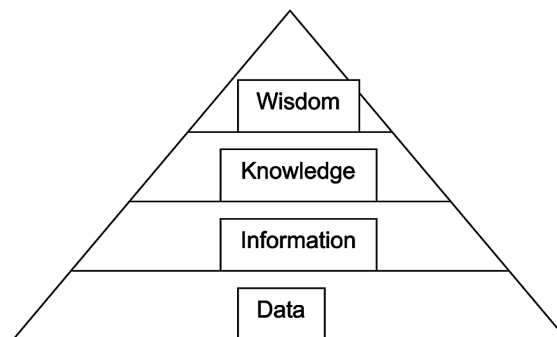
Knowledge is the outcome of information and has been defined in a number of ways. Al-Salti and Hackney, (2011) defines knowledge as “a product of human reflection and experience”. Moreover, it has been considered as a powerful tool to develop better decision making and innovation. Recently, many researchers have stressed that knowledge is an important factor to gain and sustain a competitive advantage. There are two main dimensions of knowledge: explicit and tacit. Explicit knowledge is transmittable knowledge that exists in symbolic or written form and stored in readily accessible media such as manuals, documentations, procedures and program codes. Tacit knowledge, on the other hand, is knowledge that resides in the minds of people but difficult, or even in some cases impossible, to be expressed in verbal, symbolic and written form such as insights, expertise and experience. Such knowledge cannot be transferred through a written document, and yet it is very important in the organization. The above discussion reveals that Knowledge is the outcome of information which itself is considered as the meaningful data.

Data, Information, Knowledge, and Wisdom

Many authors and researchers around the globe have co-related the concept of data, information with knowledge and wisdom. According to Tuomi (2000), when the information is interpreted or put into context, or when meaning is added to it, it is converted into knowledge. There are a number of variations of this widely accepted theme. The general idea is that data is somewhat less than information and that information is less than that of knowledge. Furthermore, it is understood that data is necessary for the creation of information and knowledge emerges only when we have knowledge (Figure 1).

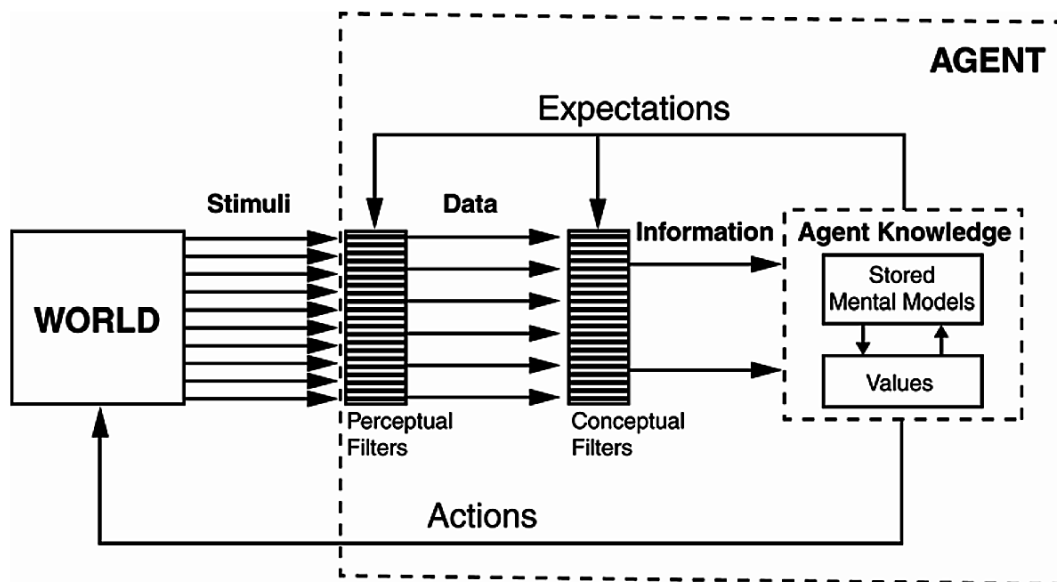
Figure 1 is in accordance with what Singh, (2007) divulges that there is an established hierarchical liaison between data, information and knowledge. Data can be defined easily as “raw”

Figure 1. Relation b/w data, information, knowledge, and wisdom



facts, which can be expressed in terms of numbers, symbols, text, images or voice, etc., representing quantities, actions and objects. But it is complicated to describe knowledge and discriminate it from information. Data becomes “information” when it is placed into some context. Information reduces ambiguity and changes one’s way of thinking. Knowledge has become the most important asset or resource in contrast to information or data as it is not easily identified, understood, classified, organized, shared, measured or to imitate. Pantzar (2000) highlights that Data and information are components of the knowledge transmitted within the information society, in such a manner that data forms the level of the elements. Data is raw, simply exists and has no implication beyond its existence. It can subsist in any form, usable or not. It does not have meaning of itself. The transformation of data into information is thus a process of reception, recognition and conversion. Information and knowledge are probably the very concepts that have been confused most in the information society debate. Knowledge is to be understood as a phenomenon that is larger than information but uses information as its building material. If knowledge is “knowing how” to do something, wisdom is “knowing why, what and how” to do something. Wisdom also extends to the application of knowledge in action. A simplistic representation of the relationship between wisdom and knowledge is captured in the following expression:

Figure 2. Data to knowledge



Wisdom = Knowledge + Ethics + Action.

Helina and Harmaakorpi, (2008) are of the opinion that:

Data is the representation of an object. Information is the aggregation of data into something that has meaning (semantics) through interpretation by human or automated processes and Knowledge is that which is derived and inferred from assimilating information against perceived context, experience or business rules.

The schema depicted in Figure 2 reflects data, information and knowledge as distinct kinds of economic goods, each possessing a specific type of utility. The utility of data resides in the fact that it can carry information about the physical world; that of information, in the fact that it can modify an expectation or a state of knowledge; finally, that of knowledge in the fact that it allows an agent to act in adaptive ways in and upon the physical world. Telephone books are paradigmatically data goods; specialized newsletters, being more selective, exemplify information goods; and brain surgery can be thought of as knowledge good (Boisot & Canals, 2003).

Understanding the concept of data, information and knowledge is becoming increasingly important in relation to the design and development of electronic information. Information is a critical resource in the operation and management of organizations. Timely availability of relevant information is vital for effective performance of managerial functions such as planning, organizing, leading, and control. As and when the citizens of any nation become aware with the concept of information or in other words become information literate, it ensures the paramount research and development of the realm.

Features of Information

The features of good information are relevance, timeliness, accuracy, cost-effectiveness, reliability, usability, exhaustiveness, and aggregation level. Information is relevant if it leads to improved decision making. It might also be relevant if it reaffirms a previous decision. If it does not have anything to do with one's problem, it is irrelevant. Timeliness refers to the currency of the information presented to the users. Currency of data or information is the time gap between the occurrence of an event in the field until its presentation to

the user (decision maker). When this amount of time is very short, we describe the information system as a *real-time* system. Accuracy is measured by comparing the data to *actual* events. The importance of accurate data varies with the type of decisions that need to be made (Babu, Singh, & Sachdeva, 2013).

Value of Information

Information has a great impact on decision making, and hence its *value* is closely tied to the decisions that result from its use. Information does not have an absolute universal value. Its value is related to those who use it, when it is used, and in what situation it is used. In this sense, information is similar to other commodities. Information supports decisions, decisions trigger actions, and actions affect the achievements or performance of the organization (Babu, Singh, & Sachdeva, 2013).

Importance of Information

Good information is essential for effective operation and decision making at all levels in academics as well as business. Every society is surely an information society and every organization is an information organization. Therefore, information is a basic resource like materials, money and personnel. Information can be considered either as an abstract concept (ideas) or as a commodity, usually in the form of letters and reports. Essentially, therefore, information has become a critical resource, just like energy, both of which are vital to the wellbeing of individuals and organizations in the modern world (Adeoti-Adekeye, 1997). Kaye (1995) mentions that most managers would agree that good information is essential to the success of an organization. If an organization is to survive and prosper, it must understand both its own internal workings and the nature of the environment to which it has to adapt and respond. Good information improves decision making, increases efficiency and provides a competitive edge to the organization which knows more than the opposition.

Information Needs and Its Types

Information environment is enough rich, characterized by an increase in information sources as well as information providers, a variety of approaches and methods for accessing information, and a redundancy of content from multiple sources. In this “overloaded” information environment, many information users tend to experience a sense of insufficiency in locating the precise information which leads to anxiety. In this complex information environment, understanding the way individuals choose to satisfy their information needs takes on new urgency. Insight into information seeking can be gained by understanding how users seek information sources and how they locate the desired information to meet their needs (Chandra, Lynn, Lawrence & Lillie, 2007). The concept of information needs was coined by an American information scientist Robert S. Taylor in his article “The Process of Asking Questions” published in American Documentation (Now is Journal of the American Society of Information Science and Technology (Agropedia, 2015). There are many definitions of an information need. According to Krippendorf (1990) information need is a recognition that your knowledge is inadequate to satisfy a goal that you have. He explains that “having information” is not the same as “being informed.” Therefore, the problem is not in obtaining information but, rather in understanding the information that you do have. Each individual’s need is formed by the actual situation and by the way the individual defines that situation (Snunith & Sarah, 2007). Campbell, (2000) defines an information need as “the perception of a lack of information that provokes one to then develop a need for it.” Some information researchers incorporated the study of information need into the larger concept of information behavior which also includes information giving, seeking and use (Pettigrew, Fidel & Bruce, 2001), while others took the information need and believe it to be representative of the whole information-seeking process (Westbrook, 1993). Users’ information needs have indeed changed (and are still chang-

ing) as a result of the emergence and expansion of the electronic form in which information content is being made available for users' access and use (Kebede, 2002).

Numerous studies have been carried out which divulge that information needs change according to the nature of subject field and requirement of users. Some of which are:

Political Respondents need information mostly while preparing for debates, speeches and questions (Alemna & Skouby, 2000). Information needs of people with Medical professionals are seen to be complex as the needs change over time; variations in the desire for information; and there are differences in how information is needed due to the physical and psychological state of different patients (Mark, Janet & Nicole 2003). It is also found from the literature that Students in universities may need information in the following types:

- Educational information need,
- Health information needs,
- Employment information needs,
- Social needs,
- Political information needs (Silvio, 2006)

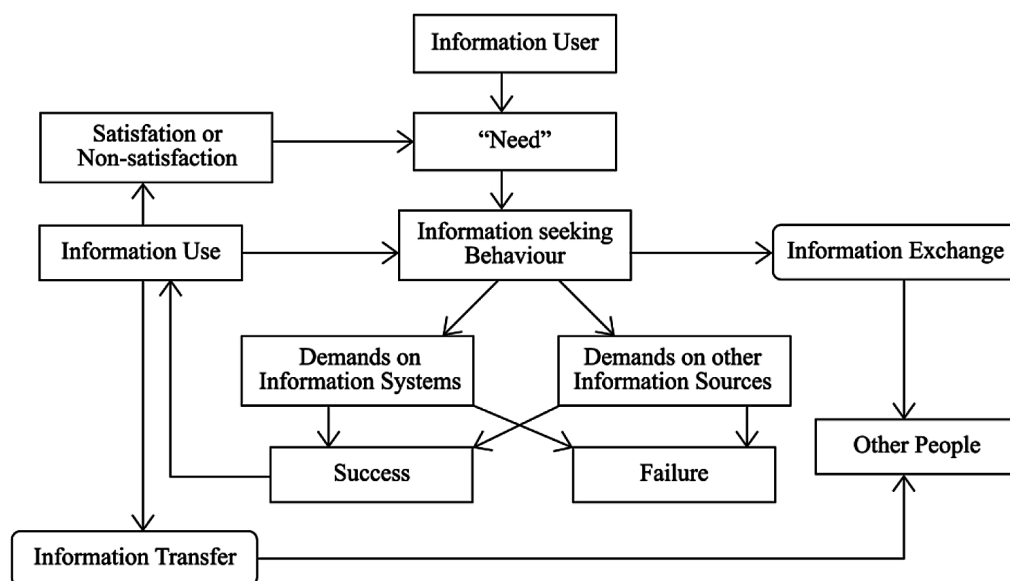
The information needs of teachers were found to be mostly related to guidance on administrative procedures, having lesson plans ready, mechanisms for evidencing work, etc (Williams, 2005). Further, information needs of the General people are found to be varied. The areas in which they needed information are diverse. These range from the information needs of the farmers, to that of the petty traders, artisans, blacksmiths, weavers, painters, fishermen, postmasters, labors, adult learners etc. Their needs, however, crystallized into the following major areas:

- Agriculture information needs.
- Health information needs.
- Political information needs.
- Educational information needs.
- Economic information needs.
- Social information needs (Momodu, 2002).

Use of Information Sources to Fulfill Information Needs

For fulfilling the information needs, users access different sources of information Sources. Scien-

Figure 3. Generation of information needs among users



tists, engineers and technologists in general use encyclopedias, handbooks, textbooks, periodicals, abstracts, indexes, standards, patents, etc. for their research and development works. He showed that information needs of scientists, engineers and technologists are equally based on their knowledge about those sources of information and accessibility of these information sources (Gayatri, 2006). Post Graduate students and Research Scholars mostly use journals, library books and textbooks for completing their course work (Fidzani, 1998). Wilson, (2006) shows through a model (Figure 3), the mechanism that helps to know how the information needs generates among the users of an information system as:

Types of Information Needs

Information age has driven every user towards a way where he needs information in various types and formats. As a result, multiple information sources have been generated to satisfy the user information needs. These needs have been categorized according to the users and their requirements into several types in different fields of knowledge. Moreover, it has been a topic of numerous research papers and information professionals from time to time.

Broadly, there are three types of needs:

- **Expressed or Articulated Need:** The need that is expressed.
- **Unexpressed Needs:** The user is aware of the needs but does not like to express it.
- **Dormant Needs:** The user is unaware about his need. However, Information service provider brings to light their needs.

However, information needs have been categorized into different types by various authors globally. Some of which are as:

Shenton (2007) divulges that information needs (Figure 4) can be divided into five broad categories:

1. Needs that are known to the individual but not to the information professionals.
2. Needs that are known to both parties.
3. Needs that are known to the information professional but not the individual.
4. Needs that are misunderstood by the individual.
5. Needs that are not known to either the individual or the information professional.

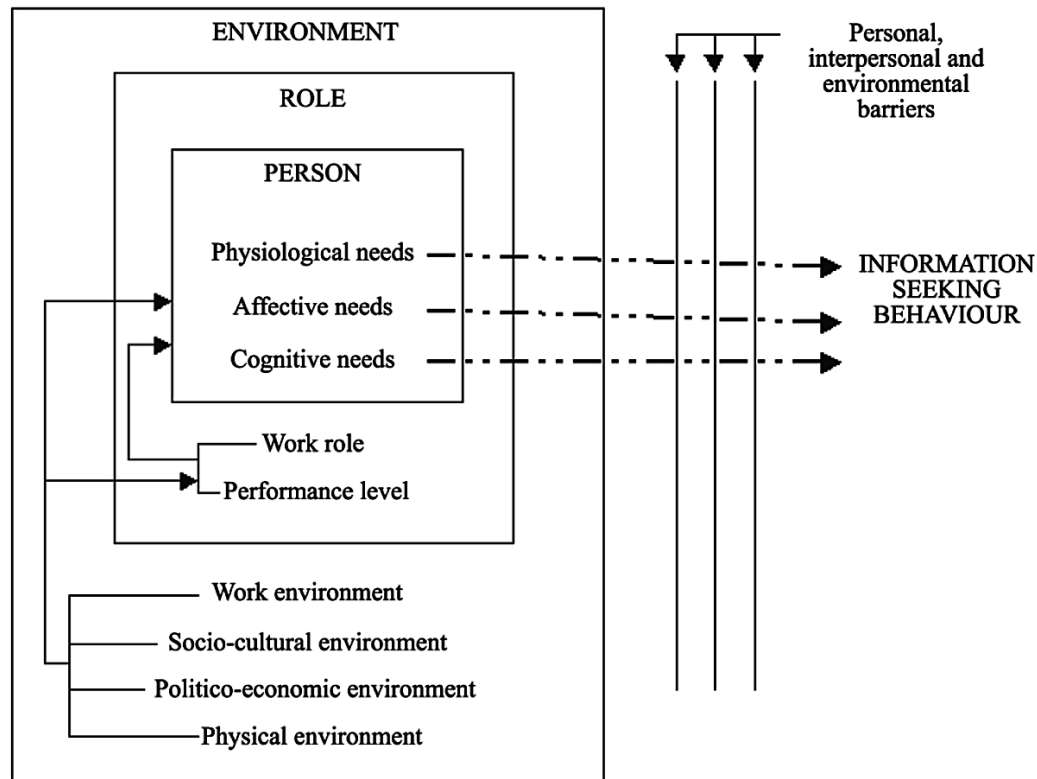
Khoir, Du & Koronios (2014) while discussing about the information needs of general people conclude that their need categories include: housing, schooling, health, banking and finances, driver

Figure 4. Types of information need as represented in a Johari Window

	Misunderstood by the individual	Known to the individual	Not known to the individual
Known to the information professional	Di <i>Misguided needs</i>	B <i>Expressed needs</i>	C <i>Inferred needs</i>
Not known to the information professional	Dii <i>Misguided needs</i>	Ai <i>Unexpressed needs</i> Aii <i>Independently- met needs</i>	E <i>Dormant or delitescent needs</i>

Figure 5. Types of Information needs

Source DOI: 10.1108/00220410610714895



licenses, government-related issues, legal issues and practical information. While as, Alemna and Skouby (2000) reveal with regard to the type of information often sought by the MPs, these were varied and interesting. A broad categorization and ranking of the information types relate to the needs of the society. However, when considered in terms of hierarchy, most of the MPs considered information on rural development, agriculture, and human rights more important than information on foreign affairs and military.

Wilson has divided the information needs of a person according to the circumstance of his related environment (Work environment, Socio-cultural environment, Politico-economic environment) into three main types (Figure 5):

1. Physiological needs.
2. Affective needs.
3. Cognitive needs.

Factors Influencing Information Needs and Seeking Behavior of Users and Their Types

Information need is one of the cognitive needs of humanity. Information need determines information-seeking behavior and these concepts harmonize one another. Information need and information-seeking behavior are affected by many factors. It is observed from the data that 'information scattered in too many sources' and that too in multi-formats is the problem often faced by users. The users sometimes face problems such as needed information is not available in library, incomplete information in sources, do not know how to use online catalogue, do not know how to use electronic resources, lack of information skills to search and internet speed is slow (Kadli & Kumber, 2011). Some early studies suggest that demographic factors such as tenure, experience,

and education affect information seeking behavior (Keller & Holland, 1978; O'Reilly, 1982). Some of the authors are of the opinion factors influencing information seeking behavior of users are of the following types:

1. **External Factors:** These are the factors that include social, organizational, time, navigation, project, demographics and other external factors.
2. **Internal Factors:** These include the factors related to the internal ability and skills of user like the previous knowledge, information literacy, level of experience, self-precision, self- efficiency etc.

Broadly speaking, most of the authors and researchers found that factors influencing the information seeking behavior of users are of two main categories (Figure 6 and Figure 7) viz:

- Micro
- Macro factors

Both the types are discussed as follows.

Micro Factors

The literature related to the factors responsible for influencing information seeking behavior of users reveal that demographics (gender, nationality and age), discipline, level of study, type of enrolment, stage of study, pedagogy etc are the micro factors of information-seeking behavior.

Demographic Characteristics

Some demographic characteristics that directly or indirectly influence the Information seeking behavior of users are as:

- **Gender:** Gender is believed to play an important role in the information seeking behavior of users. Some studies show female have higher social-cultural, and psychologi-

cal adaptation scores than male. However, others reveal that women have lower competency in using computerized library resources and technologies. As most database searching requires pervasive computer use and familiarity, it could affect their ability on Information seeking behavior There-fore, males were more satisfied and confident and had fewer difficulties than females. As revealed by Al-Muomen, Morris and Maynard (2011), gender issues were also considered problematic by some focus group attendees. Many universities in the Arab world, segregate students in the classrooms according to their gender. Male and female students were separated at Kuwait University in 1996. According to the responses of one of the attendees “segregating genders on the campus had led to a big deterioration in the level of education in the country”. Another noted that since genders are now separated in campus and says “sending a mixed group of males and females as a group to work in the library on a research project has obvious consequences”.

- **Nationality:** Studies reveal that level of adjustment of students from different nationalities have a good effect on Information seeking behavior. It could be because of the cultural, language and education system that may be highly influencing the information seeking of users.
- **Age:** Age of the users can affect the information seeking behavior of users. They learn skills and become more educated with age in an academic institution. Postgraduate students are usually older than undergraduate students, they may have advanced level of library skills information literacy and may be more aware with online recourses and search engine which have been used during undergraduate degree. Meanwhile, an old person may lose his remembrance which may prove adverse to his information seeking behavior.

- **Financial Problem/ Income:** If students are wealthier and financially well, they can have better electronic gadgets and get more services than those that are financially weak. Chen (2010) indicated that wealthier immigrants are more adaptive to the host country. It may be because they can visit host country more and had more opportunities to participate to more ceremonies of the host country which makes their social adjustment easier and hence influences their information seeking behavior.

Cultural Aspects

Cultural factors may be viewed as those aspects of culture that members of cultural groups have acquired, intentionally or unintentionally, and carry with them where ever they go. When cultural factors of one group or one individual interface with another culture, it is quite likely that some form of dissimilarity will occur. Such dissension offers the potential for misunderstanding and in the information seeking environment it frequently leads to less-than-successful learning experiences for those who are cultural outsiders and leads to frustration, loss of motivation, and decline in self-esteem and individual value. It is good to say that the role played by cultural factors is highly influencing the information seeking process of students (Chen, 2010).

Information Literacy

A person must be able to recognize when information is needed and have the ability to locate, evaluate, and use effectively the needed information. To produce such skilled learners' schools colleges and universities appreciate and integrate the concept of information literacy into their learning processes and that they play a leadership role in equipping individuals and institutions to take benefit of the opportunities inherent within the

information society. Information literate people are those who have learned how to learn. They know how to learn because they know how knowledge is organized, how to find information and how to use information in such a way that others can learn from them. They are people prepared for lifelong learning, because they can always find the information needed for any task or decision at hand (Susie, 2005). Hence, Information literacy greatly influences the information seeking behavior of users.

IT Skills

The more competent the students attain their IT skills, the more highly they are able to locate suitable resources and to use the online databases. They are also more aware of the e-resources and experience fewer problems while using computers and networks for online searching purposes than students who have less IT skills.

Psychological Factor

Psychology plays a pivotal role in influencing the information seeking behavior of students. Before the relevant information is retrieved the students must overcome possible barriers, which sometimes are psychological. They must experience the situation as gratifying enough and themselves as competent enough to actually take the final decision to seek information. Motivation and interest influence the way information is used and evaluated. More the student is interested in the topic, the more efficiently he seeks information about it. In the use of information systems just technical skills are not enough, positive attitude and self-confidence are needed in order to excel with the systems. Emotional aspects like feelings of dissatisfaction, annoyance, information overload, confrontation to new information and computer loathing may form barriers to the search process (As cited by Heinstrom, 2003).

Academic Influence

Teachers also influence the information seeking behavior of students by offering guidance on how to conduct a literature review, guide with their research process, and help them on how to use information resources. Teachers can also recommend journals and papers by renowned authors and can give projects and assignments that student require in their research. Available literature also reveal that academics do influence students information seeking behavior and play an important role in encouraging students to explore a wide range of information resources. The importance of academic influence on searching behavior has also been noted by other studies as revealed by Barrett (2005) and Tenopir (2003) (As cited by Al-Muomen, Morris & Maynard, 2011).

Macro Factors

The Macro Factors that influence the information needs of users are discussed as under:

Information Resource Design

Information resource design is a term given by Urquhart and Rowley (2007) to cover all the aspects relevant to the devise of training materials to facilitate students locate and use information resources successfully. In focus group discussions with graduate students, Urquhart and Rowley found that much training material provided by the library or staff members often lacked detail. In this study, students attending the focus groups also expressed concern about the lack of, and variability in, information resource training materials. For example, two students from Social Sciences pointed out that some academic staff used the same materials for both Bachelor and Masters' programmes without recognizing the need for it to be pitched at different levels.

Availability and Constraints to Access

Availability of good computer systems and internet access is one of the main influential factors related to information seeking behavior of students. Besides, slow internet access, electricity disruptions and database connection failures discourages the students in seeking precise information. Problems in obtaining usernames and passwords from a particular library to have access to its collection are other issue related to the students information seeking.

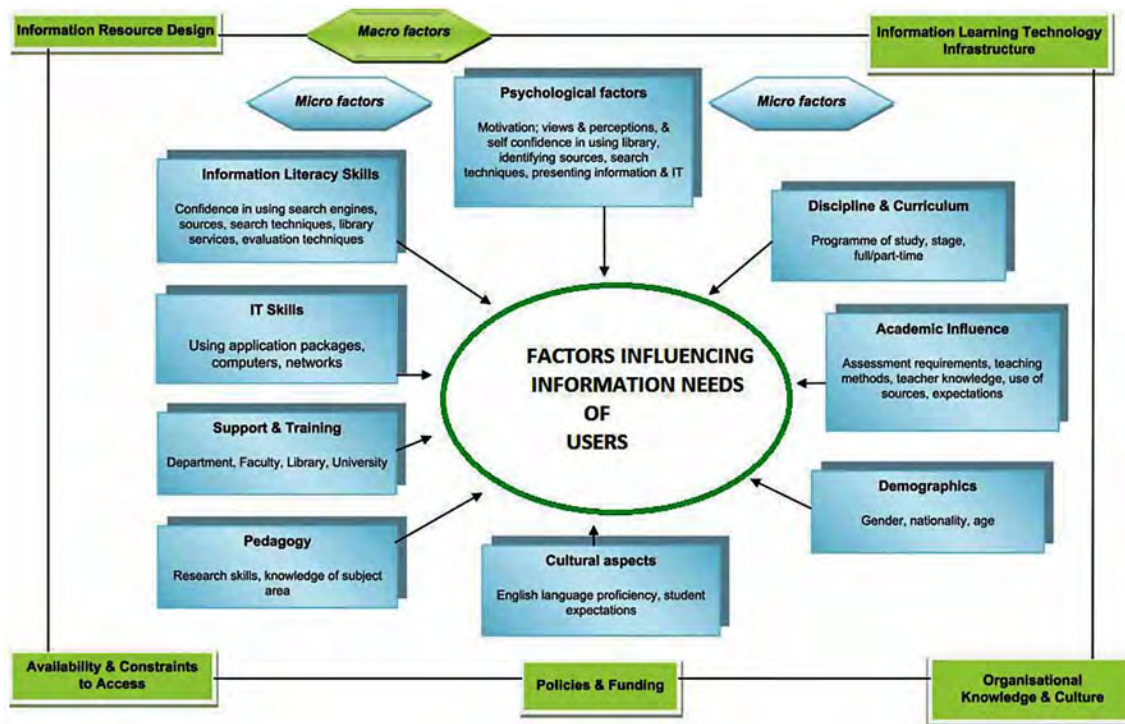
Policies and Funding

We are in an era when people cannot wait and every user needs information instantly. It calls for the Libraries to provide more services and facilities to its users and the libraries should not be just used as a quiet place of study. It is only possible when libraries possess enough financial resources to facilitate better services to its users. Lack of financial resources can become a major problem as this limited the purchase of books and access to databases. One academic solution suggested is that this could be partially overcome by having shared access to materials held in a consortium across universities

Organizational Knowledge and Culture

Organizational culture is another issue that surely will affect the information seeking behavior of students and users of its information sources. The time of working hours and the facilities like Photostat and printing as well as scanning, often needed by a student, can either enhance the skill of information seeking behavior of a student or can prove otherwise. Obviously, barriers and obstacles such as these will affect what, how and when information is sought.

Figure 6. Factors influencing the information needs of users

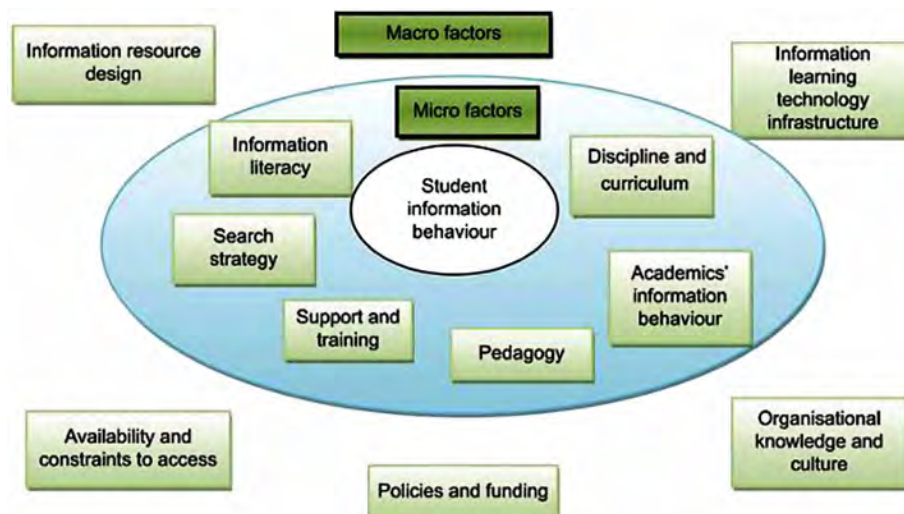


Information Learning Technology Infrastructure

Efficient virtual learning environment and use of webinars also have good influence on the information behavior of users. In their research in Kuwait

University, Urquhart and Rowley (2007) found that students praised the use of virtual learning environments for providing access to materials which enabled them to tailor their learning and to learn at their own speed. In this case, the majority of graduate students expressed concern over

Figure 7. Factors influencing the information needs of users



the lack of sufficient virtual learning materials and reported that the low usage of electronic information resources available through their library portals and virtual systems was, in part, a consequence of this.

FUTURE RESEARCH DIRECTIONS

Information needs of users have indeed changed (and are still changing) as a result of the emergence and extension of the electronic form in which information content is being made available for users' access and use. Information needs of users help in understanding and examining their Information Seeking Behavior and prove useful to select appropriate methodologies and adopt best techniques to carry out research in more systematic manner. These help in better understanding of tasks, activities, available services, infrastructure and problems faced by users while searching their needed information thereby help administrators and higher authorities to take right decisions, acquire relevant information sources, offer best services, implement tech savvy infrastructure and satisfy the information needs of users. This study has a vast scope for future research. It can be extended to study and highlight the information needs of Professors, Lecturers, Teachers, Post Graduate & Under Graduate Students associated with teaching and learning processes in different Colleges and Universities globally. Further, research can be extended towards exploring information retrieval and information seeking models.

CONCLUSION AND SUGGESTION

Information needs are highly related to the fields of study of users and shows a remarkable diversion in their nature. Further, knowing the information needs of various users in diverse fields of study will surely help to understand the information seeking behavior of users. There is great diversity

in the information needs of users depending upon the field of study with which they are associated. In general, these needs are Physiological needs, Affective needs and Cognitive needs. Information need determines the information-seeking behavior of a user and is affected by many factors. There are two main types of factors viz Macro and Micro Factors. Macro Factors include Information resource design, Availability and constraints to access, Policies and funding etc. while as Micro Factors include demographics (gender, nationality and age), discipline, level of study, type of enrolment etc.

REFERENCES

- Adeoti-Adekeye. (1997). The importance of management information systems. *Library Review*, 4(6), 318–327. doi:10.1108/01435129510772283
- Agropedia. (2015). *Information need*. Retrieved from: <http://agropedia.iitk.ac.in/content/information-needs>
- Al-Muomen, N., Morris, A., & Maynard, S. (2012). Modelling information-seeking behavior of graduate students at Kuwait University. *The Journal of Documentation*, 68(4), 430–459. doi:10.1108/00220411211239057
- Al-Salti, Z., & Hackney, R. (2011, September 27). Factors impacting knowledge transfer success in information systems outsourcing. *Journal of Enterprise Information Management*, 24(5), 455–468. doi:10.1108/17410391111166521
- Alemna, A., & Skouby, K. (2000). An investigation into the information needs and information-seeking behavior of members of Ghanas legislature. *Library Management*, 21(5), 235–240. doi:10.1108/01435120010324815
- Babu, R. A., Singh, Y. P., & Sachdeva, R. K. (2013). *Natural Resources Management and Environment Department*. Retrieved from: www.fao.org/nr/index_en.htm

- Boisot, M., & Canals, A. (2004, January 01). Data, information and knowledge: Have we got it right? *Journal of Evolutionary Economics*, 14(1), 43–67. doi:10.1007/s00191-003-0181-9
- Campbell, I. (2000). *The ostensive model of developing information-needs*. Glasgow University. Retrieved from: www.dcs.gla.ac.uk/~iain/papers/phd.pdf
- Case, D. O. (2012). Looking for Information – A Survey of Research on Information Seeking, Needs and Behavior. Bingley, UK: Emerald Group Publishing.
- Cawkell, T. (2003). *Information*. In *International encyclopedia of library and information Science* (2nd ed.). London: Routledge.
- Chandra, P., Lynn, S., Lawrence, O., & Lillie, R. (2007). What is enough? Satisfying information needs. *The Journal of Documentation*, 63(1), 74–89. doi:10.1108/00220410710723894
- Chen, W. (2010). Internet-Usage Patterns of Immigrants in the Process of Intercultural Adaptation. *Cyberpsychology, Behavior, and Social Networking*, 13(4), 387–399. doi:10.1089/cyber.2009.0249 PMID:20712497
- Fidzani, B. (1998). Information needs and information-seeking behavior of graduate students at the University of Botswana. *Library Review*, 47(7), 329–340. doi:10.1108/00242539810233459
- Gayatri, M. (2006). Information-seeking behavior of undergraduate biology students: A comparative analysis of first year and final year students in University College Dublin. *Library Review*, 54(2), 86–99. doi:10.1108/00242539810233459
- Heinstrom, J. (2003). Five personality dimensions and their influence on information behavior. *Information Research*, 9(1). Retrieved from <http://InformationR.net/ir/9-1/paper165.html>
- Helina, M. H., & Harmaakorpi, V. (2008). Data, information and knowledge in regional innovation networks: Quality considerations and brokerage functions. *European Journal of Innovation Management*, 11(1), 103–124. doi:10.1108/14601060810845240
- Kadli, J., & Kumbar, B. D. (2011). Faculty Information-Seeking Behaviour in the Changing ICT Environment: A Study of Commerce Colleges in Mumbai. *Library Philosophy and Practice*. Retrieved from <http://www.webpages.uidaho.edu/~mbolin/kadli-kumbar.htm>
- Kaye, K. (1995). The importance of information. *Library Management*, 16(5), 6–15. doi:10.1108/01435129510772283
- Kebede, G. (2002). The changing information needs of users in electronic information environments. *The Electronic Library*, 20(1), 14–21. doi:10.1108/02640470210418227
- Keller, R., & Holland, W. (1978). Individual characteristics of innovativeness and communication in research and development organizations. *The Journal of Applied Psychology*, 63(6), 672–759. doi:10.1037/0021-9010.63.6.759
- Khoir, S., Du, J. T., & Koronios, A. (2014). Study of Asian Immigrants' Information Behaviour in South Australia: Preliminary Results. In *iConference 2014 Proceedings* (pp. 682–689). doi:10.9776/14316
- Krippendorff, K. (1990). Models and Metaphors of Communication. *Media and Communication, Construction of Realities*. Retrieved from http://repository.upenn.edu/asc_papers/276
- Mark, H., Janet, H., & Nicole, J. (2003). Information needs of people with multiple sclerosis and the implications for information provision based on a national UK survey. *Aslib Proceedings*, 55(5/6), 290–303. doi:10.1108/00012530310498860

- Martin, W. J. (1995). *The global information society*. Aldershot, UK: Aslib Gower.
- Momodu, M. (2002). Information needs and information seeking behavior of rural dwellers in Nigeria: A case study of Ekpoma in Esan West local government area of Edo State, Nigeria. *Library Review*, 51(8), 406–410. doi:10.1108/00242530210443145
- O'Reilly, C. A. (1982). Variations in decision makers use of information sources: The impact of quality and accessibility of information. *Academy of Management Journal*, 25(4), 756–771. doi:10.2307/256097
- Pantzar, E. (2000). Knowledge and wisdom in the information society. *Foresight*, 2(2), 230–236. doi:10.1108/14636680010802573
- Pettigrew, K., Fidel, R., & Bruce, H. (2001). Conceptual frameworks in informational behavior. *Annual Review of Information Science & Technology*, 3(5), 43–78.
- Shenton, A. K. (2007). Viewing information needs through a Johari Window. *RSR. Reference Services Review*, 35(3), 487–496. doi:10.1108/00907320710774337
- Silvio, D. H. (2006). The information needs and information seeking behavior of immigrant southern Sudanese youth in the city of London, Ontario: An exploratory study. *Library Review*, 55(4), 259–266. doi:10.1108/00242530610660807
- Singh, S. P. (2007). What are we managing – knowledge or information? *Vine*, 37(2), 169–179. doi:10.1108/03055720710759946
- Snunith, S., & Sarah, K. (2007). Information needs of North American immigrants to Israel. *Journal of Information Communication and Ethics in Society*, 5(2/3), 185–205. doi:10.1108/14779960710837641
- Susie, A. (2005). From prescribed reading to the excitement or the burden of choice: Information literacy: foundation of e-learning. *Aslib Proceedings*, 57(2), 181–190. doi:10.1108/00012530510589146
- Tuomi, I. (2000). Data is More Than Knowledge: Implications of the Reversed Knowledge Hierarchy for Knowledge Management and Organizational Memory. *Proceedings of the Hawaii International Conference on System Sciences*, 3(2), 45. Retrieved from: <http://www.jstor.org/stable/40398446>
- Urquhart, C., & Rowley, J. (2007). Understanding student information behavior in relation to electronic information services: Lessons from longitudinal monitoring and evaluation, part 2. *Journal of the American Society for Information Science and Technology*, 58(8), 1188–1197. doi:10.1002/asi.20562
- Westbrook L. (1993). User needs: A synthesis and analysis of current theories for the practitioner. *RQ*, 32(4), 541–549.
- Wikipedia. (2013). *Establishing a management information system*. Author.
- Williams, P. (2005). Using information and communication technology with special educational needs students: The views of frontline professionals. *Aslib Proceedings*, 57(6), 539–553. doi:10.1108/00012530510634262
- Wilson, T. (2006). On user studies and information needs. *The Journal of Documentation*, 62(6), 658–670. doi:10.1108/00220410610714912

ADDITIONAL READING

- Al-Ansari, H. (2006). Internet use by the faculty members of Kuwait University. *The Electronic Library*, 24(6), 791–80. doi:10.1108/02640470610714224
- Alazemi, T. R. (2015). *Users' Information Seeking Behaviors, their Interactions and Experience with the Academic Library Web Interface*. University of Salford. Retrieved from: <http://usir.salford.ac.uk/36705/1/Final%20Thesis.pdf>

Caroline, F. T., & Cees, A. W. (2010). Developing scales for information-seeking behavior. *The Journal of Documentation*, 66(1), 46–69. doi:10.1108/00220411011016362

Ganaie, S. A., & Khazer, M. (2015). Diversity of information sources in the digital age an overview. *Journal of Advancements in Library Sciences*, 2(2), 53–61.

Singh, S. P. (2007). What are we managing – knowledge or information? *Vine*, 37(2), 169–179. doi:10.1108/03055720710759946

Snunith, S., & Sarah, K. (2007). Information needs of North American immigrants to Israel. *Journal of Information Communication and Ethics in Society*, 5(2/3), 185–205. doi:10.1108/14779960710837641

Tanackovic, F. S., Horvatic, I. F., & Badurina, B. (2015). European Union information in an acceding country. *Library Hi Tech*, 33(1), 143–158. doi:10.1108/LHT-10-2014-0103

KEY TERMS AND DEFINITIONS



Information: Information is an assemblage of data in a comprehensible form capable of communication and use.

Information Behavior: Information behavior refers to how people approach and handle information. **Information Literacy:** The ability to know when there is a need for information, to be able to identify, locate, evaluate, and effectively use that information for the issue or problem at hand.

Information Need: The perception of a lack of information that provokes one to develop a need for it.

Information Seeking Behavior: A special case of problem solving which, includes recognizing and interpreting the information problem, establishing a plan of search, conducting the search, evaluating the results, and if necessary, iterating through the process again.

Knowledge: When the information is interpreted or put into context, or when meaning is added to it, it is converted into knowledge.

A Maturity Model for Digital Literacies and Sustainable Development

Ravi S. Sharma

Nanyang Technological University, Singapore

Lin G. Malone

Nanyang Technological University, Singapore

Chong Guan

SIM University, Singapore

Ambica Dattakumar

Nanyang Technological University, Singapore

INTRODUCTION

It is a given that the world is now becoming increasingly digitalised. However the speed at which this digitisation has occurred, has led to unequal progression amongst societies. A key aspect of digitisation is the notion of “digital inclusion”; the empowerment of individuals through digital participation. Successful initiatives, supported by digital literacy, have enabled those that are isolated to gain on a social and economic front (Sharma & Mokhtar, 2006). This paper recounts the role of digital literacies in supporting participative, and therefore sustainable, development. Taking a historical development perspective, the paper concludes with a maturity model that links digital policies with socio-economic well-being.

Building on the pioneering work of Gilster (1997), Belshaw (2012) offers a comprehensive definition of modern literacies in digital society:

Literacies involve the mastery of simple cognitive and practical skills. To be ‘literate’ is only meaningful within a social context and involves having access to the cultural, economic and political structures of a society. In addition to providing the means and skills to deal with written texts, literacies bring about a transformation in human

thinking capacities. This intellectual empowerment happens as a result of new cognitive tools (e.g. writing) or technical instruments (e.g. digital technologies). (p.90)

It has been suggested that digital inclusion and participation enables the grassroots to be engaged, bridging some of the prevailing socio-economic disadvantages (SEDs) that exist within societies, as well as across countries (Armenta et al., 2012). This is the fundamental premise of digital literacies – the set of skills and tools that will empower individuals and groups to participate fully in the increasingly digital future and hence bridge the disparities in socio-economic opportunities.

BACKGROUND

The Evolution of Digital Literacy

Lanham (1995) first conceptualised digital literacy as the ability to comprehend information, regardless of the medium. This definition focused on the user’s ability to navigate between the various online and offline mediums. Since this original conceptualisation, the term digital literacy has evolved along with pervasive Information and

Communication Technology (ICT) in society. While Lanham created awareness of the need to comprehend the transformations brought about by the incorporation of ICTs, it was Gilster (1997) who popularised the concept of digital literacy and its emergence as a critical skill. His portrayal of digital literacy as “mastering ideas, not keystrokes” (1997, p.15), positioned the concept to focus more on cognitive ability, as opposed to competencies. This was considered a milestone, as society rapidly digitised and network effects arising from social media led to the development of social capital as a socio-economic advantage.

Building on this, Eshet-Alkalai (2004) presented five survival skills for the digital era: photo-visual literacy, reproduction literacy, information literacy, branching literacy and social-emotional literacy. Of these five digital literacies, four of them are largely based on specific digital skills. As the contrasting element, socio-emotional literacy is of particular interest. The definition of socio-emotionally literate users offered by Eshet-Alkalai (2004) is individuals who are able to work with others, sharing and evaluating information and knowledge, in order to construct new knowledge. This refers to the participation and communication that occurs in the digital world, as well as the opportunities offered via this medium. Where participation leads to collective intelligence, new knowledge may be developed. By situating socio-emotional literacy as a digital literacy skill, the Internet and new media present a new cultural environment, with its own unique values and practices for engagement.

The socio-cultural dimension of digital literacy is further discussed by Bélisle (2006). Although her work focuses on a re-conceptualisation of literacy and not merely digital literacy, Bélisle’s research is important as it explains the changes to society as a result of the digital knowledge revolution. In fact, it could be said that Bélisle (2006) truly grasped the essence of the changes to the concept of literacy within the digital society. Bélisle (2006) examines three dimensions of literacy: Functional, Socio-cultural and Transformational. Functional

literacy refers to the basic skills required to lead a day-to-day life. In the conventional sense, this refers to the skills of reading, writing, speaking and listening. In relation to digital literacies, this includes the ability to perform operational computer skills, such as input, output and searching, but also the ability to understand when and where each skill set is relevant. This dimension of digital literacy could be read in parallel with Lanham’s original concept.

Bélisle’s second dimension of literacies is the socio-cultural. Literacy ultimately serves to address a purpose; it “[gives] access to, and understanding of, the structures of power and authority through mastery of written texts and numbers” (Bélisle, 2006, p.53). Socio-cultural literacy includes knowledge of a society’s values, attitudes, practices and conventions; and an understanding of where each of these apply. This is important in relation to digital literacy, as the digital world provides new channels for participation and communication. Literacy is only meaningful when contextualised to the cultural fabric of society; the socio-cultural dimension of digital literacy enables individuals to immerse themselves within and to participate in social and economic structures of digital society. Hence, when referring specifically to digital literacy, it may be more accurate to consider the socio-cultural dimension of literacy as a “socio-economic” function. This would better capture its *impact on* and *empowerment of* users in online communities.

The final dimension of literacy which Bélisle describes is the transformational dimension of digital literacy. This “brings a profound enrichment and eventually entails a transformation of human thinking capacities” (Bélisle, 2006, p.54). The individuals’ intellectual empowerment through literacy may have the power to transform society, especially where creative cognitive ability leads to the creation of new cognitive tools (Bélisle, 2006). If Bélisle’s transformative digital literacy is viewed alongside Eshet-Alkalai’s socio-emotional literacy, the online world opens up new opportunities for collaboration and creation. This ultimately

brings new knowledge to society, transforming it and those within it.

More recently, Belshaw (2012) has utilised the term digital literacies, rather than digital literacy, both to avoid reducing it to a finite outcome, as well as to address the complexity of the concept. Belshaw (2012) presents eight, non-hierarchic elements of digital literacies: Cultural, Cognitive, Constructive, Communicative, Confident, Creative, Critical and Civic. Within these elements, it can be seen that digital literacies have evolved from a mere set of skills, to encompass cognitive ability, to facilitate cultural engagement, and to enable critical analysis. Continuing the evolutionary path of digital literacies, Belshaw's research (2012) can be seen as particularly prominent, especially where it links the element of "Cultural" to that of "Civic". This emphasizes participation, social justice and civic responsibility; digital literacies are not simply about functional ability but also about the inclusion, participation and empowerment that result from socio-cultural interaction. Belshaw's work draws heavily on Bélisle's; his definition of literacies feeds directly into Bélisle's proposed three

dimensions of digital literacy. Linking Belshaw's work to Bélisle (2006), transformative digital literacy can go beyond the ability to transform an individual through self-enhancement to transforming societies through these "entitlements" gained by individuals.

MAIN FOCUS OF THE ARTICLE

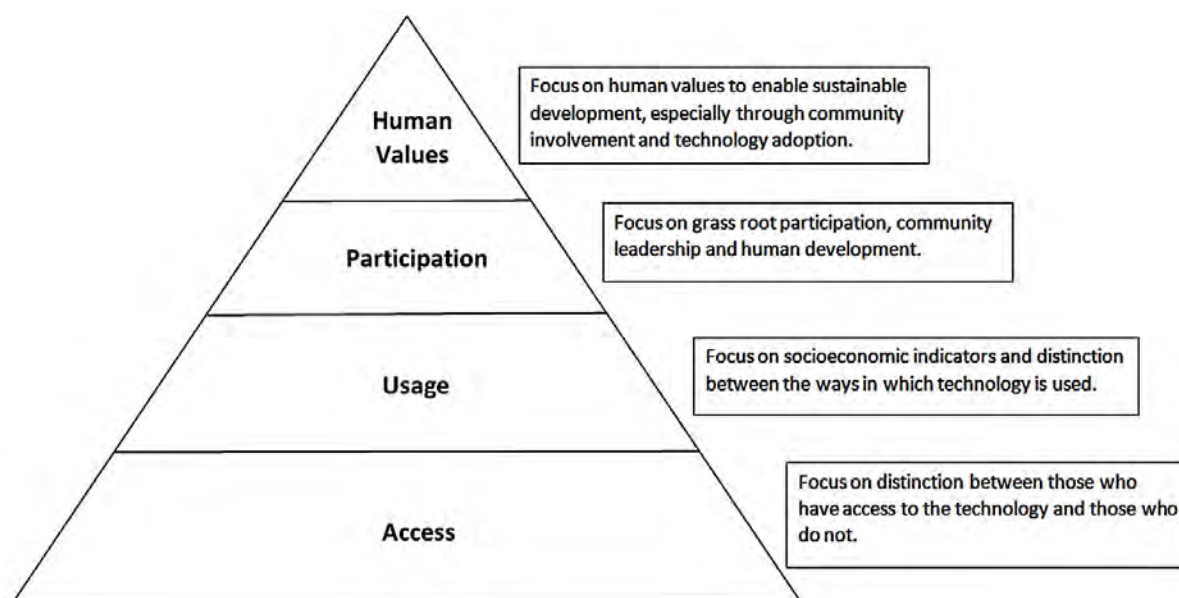
Issues, Controversies, Problems

Levels of Digital Development

The development of a digital society takes place in levels, each of which lays the foundations for the succeeding ones. In the context of the digital divide, Armenta, Serranob, Cabrera and Conte (2012) propose four levels to development; each centred on a key tension point (see Figure 1).

The first level addresses the problem of access, distinguishing between those who have access to the infrastructure and those who do not. The second level focuses on usage, in terms of the social-economic indicators governing usage, while

Figure 1. Levels of digital development



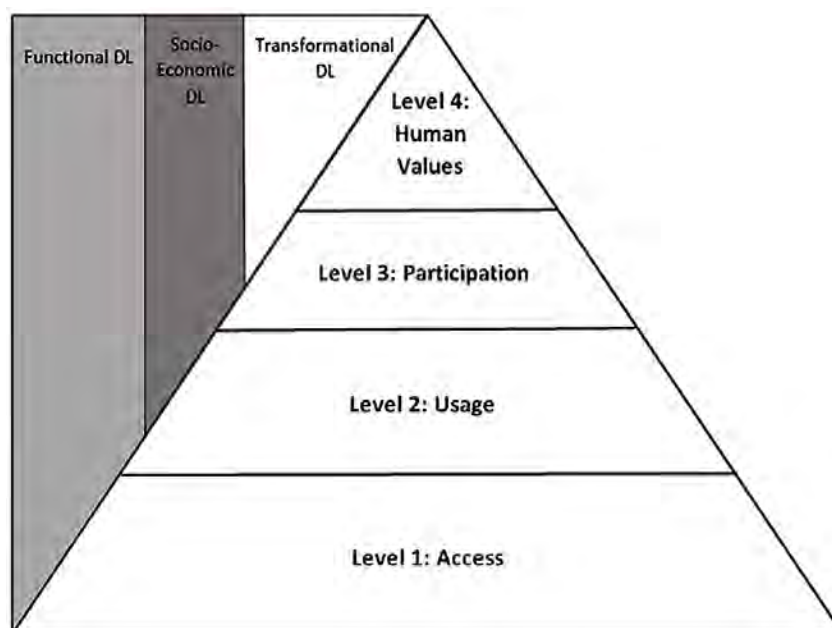
the third level addresses the role of participation within the digital divide and focuses on the need for human development through participation.

Armenta et al. (2012) further suggest that we are currently entering the fourth level of digital development, where digital inclusion must be mediated by human values. At this level, there comes a focus on the need for community involvement and technology adoption. Harding (2016) argues that while the digital world had promised to create a level playing field by creating equal access to information, such information possesses a different cost to different individuals in accordance to their socio-economic status. This highlights the importance of digital inclusion to enable a level playing field. While each level of the developmental model seeks to discourage digital exclusion, it is the fourth level of human values which has the potential to ensure that as society progresses, the socio-economically disadvantaged (SEDs) do not get marginalised and dis-enfranchised. Thus, the links between the civic and cultural elements in Belshaw's model draws parallels to the fourth level of "human values" in work by Armenta et al. (2012) on digital inclusion.

Further analysis of Armenta et al.'s model (2012) has shown each level to reflect different dimensions of digital literacies. This can be seen through adopting a modified lens of Bélisle's (2006) digital literacies and applying it to Armenta et al.'s model (see Figure 2).

A juxtaposition of Bélisle's work (2006) examines three dimensions of functional, socio-cultural and transformational literacies. This provides further clarification of the four developmental levels in Figure 1, by linking digital literacy with digital development. Access (level 1 of development) may be mapped directly to the outcome of functional literacy as without the former, one cannot possibly achieve the latter. Usage and participation (levels 2 and 3) normally lead to socio-cultural (or interchangeably socio-economic) participation. The subtle difference between usage and participation is that while usage is transactional, participation also involves the production of social (emotional) capital. Finally, at the highest level of human values, individuals and society transforms into knowledge societies (again, from the foundations of information societies enabled in levels 1 and 2).

Figure 2. Model of digital development as supposed by digital literacies



As the fundamental argument, bridging digital disparities to access and usage of technologies must be supported by individuals' functional abilities to utilise these tools. Going further into the model, it can be seen that usage still draws upon the socio-economic function of digital literacies. Armenta et al. (2012) discusses how at the second level, applications suited to the needs of the population were incorporated to bridge an existing digital usage divide. There are ample opportunities to develop context relevant applications, opening up new ways of gaining socio-economic status. It may also be said that usage of these applications draw on the socio-economic dimension of digital literacies, as one is able to participate meaningfully within society.

Mansell and Tremblay (2013) have suggested that human development is the process of improving and increasing each individual's choices. The socio-economic dimension of digital literacies can indeed enable this by opening up opportunities for grass-root participation and community leadership through creating meaning around these activities. This increases inclusion in society, and may enable improvement of one's economic status. Morris and Morris (2013), for example, have revealed that Internet access and usage could reduce the socio-economic gap in knowledge and participation. Beyond the socio-economic dimension, level three of participation also supports the transformative capacity of digital literacies. The digital environment holds a myriad of potential for users who are empowered to explore them; digital literacies can open up opportunities for individuals to transform themselves, their abilities and their circumstances.

This final level of human values is demonstrative of transformational digital literacy beyond self-enhancement to enable societal development. As discussed by Belshaw (2012), the digital environment creates new civic responsibilities for users. Relating back to Armenta et al.'s model, the fourth level of human values must comprise an element of civic responsibility to enable society's transformation through community involvement

and technology adoption. By drawing on human values, this level will promote digital inclusion for all.

Digital Maturity and Sustainable Development

"Knowledge societies will not really be worthy of the name unless the greatest possible number of individuals can become knowledge producers rather than mere consumers of already available knowledge." (UNESCO, 2005, p.189). Indeed, UNESCO's further probing (Mansell & Tremblay, 2013) highlights the need for knowledge societies to be based on inclusion to ensure their sustainability. The 2013 report underscores the importance of human values for digital inclusion and participation. The need to respect, welcome and appreciate new ideas is key to building an innovative and sustainable society.

Prior work by Sharma and his associates highlight key characteristics of knowledge societies: 1) they are necessary and sufficient conditions for growth in the knowledge economy, 2) they have high knowledge absorptive capacity and complex chains of creation, production and distribution, and 3) they consist of a sustainable learning community which emphasises innovation (Sharma, Ng, Dharmawirya & Lee, 2008). Four knowledge pillars actively used by the World Bank's Knowledge Assessment Methodology (information infrastructure, economic and institutional governance, education and human capital, and innovation system) have been effective in deriving a set of best practices in developing knowledge policies (cf Sharma et al., 2008; 2009; Chandrasekar & Sharma, 2010). However, these studies did not examine participation gaps arising from the uneven distribution of resources in developing skills and literacies throughout the world, and the lack of transparency in the way digital literacies shape perceptions of the world. At this juncture of the new participatory culture, the key question of "how we can guarantee that the rich opportunities afforded by the expanding

digital landscape are made available to all,” still remains unanswered.

Knowledge societies which are based on an institutional framework that supports collective security system and wider welfare can lead to sustainable development (Spangenberg, 2005a; 2005b). As a knowledge society is based on the need for knowledge distribution, access to information and skills to transform information into knowledge, inclusion within such societies is imperative. This may bridge the existing disparities within society and subsequently, promote sustainable development, where progress benefits all and no one gets left behind.

Synthesising the discussion in the previous section with the gap in prior work identified, the relationship between knowledge societies, digital inclusion and digital literacies may be examined through a proposed *Digital Literacies Maturity*

Model (DLMM) (see Figure 3 below). A maturity model is an assessment tool used to evaluate an entity (e.g., an organization) or a process. The term “maturity” relates to the degree of optimization of practices. Typically, a maturity model formally defines steps and result metrics which are applied to the management of best practices and active optimization of processes (Humphrey, 1988). Our proposed DLMM combines the World Bank’s four knowledge policy pillars with the four levels of digital development and provides a framework on how to promote sustainable development and socio-economic well-being.

Please note that the colouring of the cells in Figure 3 correspond to the different levels of literacy identified in Figure 2 above. Each cell in the matrix has a focus theme that characterises knowledge policy for a given pillar and level of development. For example, at the most basic level of

Figure 3. Digital literacies maturity model

	GOVERNANCE	INFRASTRUCTURE	EDUCATION & HUMAN CAPITAL	INNOVATION
HUMAN VALUES	<i>Sustainable Growth & Development</i>	<i>Community Involvement & Socio-economic Returns</i>	<i>Civic Responsibility</i>	<i>Collective Intelligence</i>
PARTICIPATION	<i>Grassroots Participation</i>	<i>Human Development</i>	<i>Community Participation & Community Leadership</i>	<i>Knowledge Exchange</i>
USAGE	<i>Training</i>	<i>Applications</i>	<i>Skills & Experience</i>	<i>Information Economy</i>
ACCESS	<i>Universal Service Obligation</i>	<i>Devices & Networks</i>	<i>Utility</i>	<i>Efficiency & Effectiveness</i>
DL Maturity Model Key	Functional Digital Literacy	Socio-economic Digital Literacy	Transformational Digital Literacy	

digital development (the access level), governance policies mainly focus on universal service obligations extracted by the regulator from licensed service providers so that no segment of society is excluded from access to basic digital services such as telephony, Internet, cable television and so on. The infrastructure dimension focuses on improving access to devices and networks and the training is geared towards utilitarian functions. Most innovative efforts are engaged in improving access, efficiency and effectiveness. Assuming prevalence of internet access is achieved, the next level focuses on the usage. Government policies promoting skill-based training and applications are essential to the development of an information economy. When both internet usage and access becomes common place, the next question to be asked is how to accelerate digital participation, such as grassroots participation, community leadership and knowledge exchange. Likewise at the highest level of digital development (the human values level), innovation policies must seek to capture the collective intelligence of the masses (sometimes known as the wisdom of the crowds).

However, the field success of the DLMM is dependent on the research methodology chosen. In empirical research conducted on digital literacies (Bonfadelli, 2002; Hargittai & Hinnant, 2008), quantitative data analysis appears to be the preferred research methodology. However, when examining, for example, Hargittai's (2009) web-oriented survey measures to gauge digital literacies, the research may be adequate in capturing functional digital literacy, but its methodology completely overlooks the socio-economic and transformation dimensions of literacies, which is a cause of concern. As ICT becomes increasingly user-centric, functional ability becomes less of an issue as compared to the ability to utilise technology in meaningful ways. In contrast, Canada has provided a vision for 21st century citizens to possess and maintain values that enable them to effectively communicate "to flourish in groups of individuals with multiple-perspectives who care deeply about a topic and are empathetically

responsive to each other's perspectives" (Hoechsmann & DeWaard, 2015, p.19).

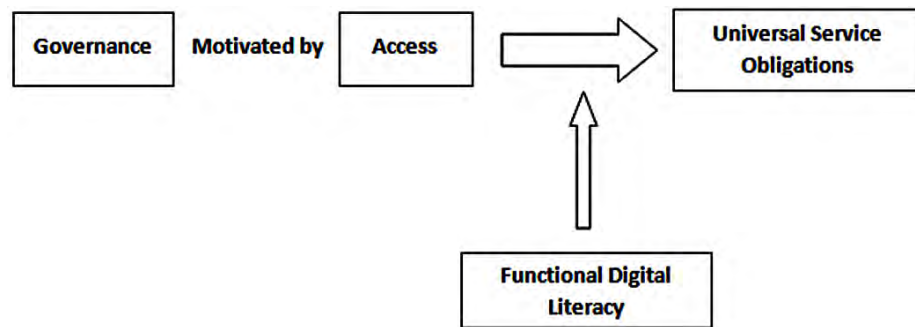
The use of DLMM allows an economy to have its digital literacies and sustainable development processes assessed according to a conceptual framework with a clear set of benchmarks. Maturity is indicated by achieving a particular "Maturity Level", which provides a consistent set of measurements for researchers. For policy makers, the DLMM allows a knowledge society to compare its maturity level with other economies, or other parts of their own economy and derive specific recommendations for improvements. As such, a recommended approach to the field usage of the DLMM would be to create narratives, which allows space for quantitative hypotheses as well as data collection (Sharma et al. 2012). The creation of narratives and the application of the DLMM to these narratives could yield more in-depth and interesting results as compared to other methodologies. The details included through the choice of narrative approach thus become more applicable in relation to the levels of maturity (access, usage, participation, values) than the policies. For knowledge societies, it would be argued with choice narratives that government policies have been situated at the fourth level of human values in the digital development model. Whether this has led to sustainable development can be tested by tracking specific policy indicators of governance, infrastructure, education, and innovation.

SOLUTIONS AND RECOMMENDATIONS

Implications of DLMM for Digital Policies

Policies which encourage the continued evolution of knowledge societies must be constructed on the twin principles of inclusion and be supported by digital literacies to best enable the formation of sustainable knowledge societies. As the flowchart below illustrates (see Figure 4), governance can

Figure 4. Flowchart mapping the relationship between knowledge society pillars, digital development levels, and digital literacies



bridge the access divide through their universal service obligations, ensuring that the Internet is made equally accessible to all citizens as a digital entitlement. Infrastructure, built with the intention of bridging the access divide, can create accessible devices and networks for all. Utility may be enabled where access is supported by education and human capital. Innovation systems which enable access can create an efficient and effective society. However, these access policies must still be supported by functional digital literacy for users. Access is redundant when functional ability is absent.

Socio-economic literacy builds on functional literacy, to create meaning behind activity. On the second level, digital inclusion can be enabled through promoting usage. Government policies can enable usage through training, however these activities only have meaning when viewed through the lens of socio-economic literacy. Infrastructure may create applications tailored to the needs of the population, however it is with socio-economic digital literacy that users are motivated to utilise them. Skills obtained through formal education and life-long learning must be similarly motivated by the socio-economic dimension of literacy which creates meaning around the activity; the functional ability to use software and applications is only meaningful when understanding its socio-economic uses, purposes and relevance to the workforce. Human Capital provides the opportunity to gain experience with using technology. It is

important for individuals to use these experiences appropriately in a given context. For example, an individual can progress from being a passive receiver of technology to an active information consumer with effective searching skills. At the highest level, he or she can apply computational thinking skills and use the information intelligently to create value. This, again, results from digital literacies' socio-economic dimension. Finally, when innovation is supported by usage, it will facilitate the creation of the information economy, as the first step toward knowledge societies.

Through empowering people with socio-economic and transformational literacies, this may open up opportunities for digital citizens to improve their socio-economic status. The third level of digital inclusion focuses on the participation divide. Governments seeking to close the digital divide by bridging the participation gap may implement policies to promote grassroots participation supported by socio-economic digital literacy to create meaningful supporting participation. Infrastructure with the goal of achieving participation can incentivise human development, especially if it is supported by socio-economic literacy. Education and human capital, when directed toward bridging the participation divide, can enable community participation and create community leadership, as supported by socio-economic literacy. Finally, the key to the development of knowledge societies is innovation-supported participation. This can cre-

ate exchange of knowledge, which is catalyzed by transformational literacy.

Transformational literacies are especially key to the fourth level of human values. With transformation literacy as a goal, governance supporting human values can enable sustainable growth and development. Infrastructure with human values can allow community involvement and subsequently, social capital returns. Education and human capital supported by human values will create a sense of civic responsibility toward society; and innovation supported by human value will lead to the expression of collective intelligence.

FUTURE RESEARCH DIRECTIONS

When viewing the digital maturity model alongside Bélisle's conceptualisation of literacies (2006), it is apparent that digital literacies are necessary to support participation and subsequently, inclusion. The need to promote digital inclusion, especially through imparting digital literacies, becomes increasingly important for policy makers

in light of the value placed on knowledge within the modern economy and the drive for countries to construct sustainable growth and development.

Frameworks and measures conceptualised by government and policy makers have been developed with the intention of bridging the digital divide. The European Union and Canada, for instance, have developed conceptual frameworks of digital literacies and measures, with a focus on education and creating the 21st century citizen. When developing frameworks through which to build knowledge societies, it is important that governments and policy makers keep in mind that inclusion must serve as the foundation to such societies, not merely an afterthought. The Digital Literacies Maturity Model presented in this article is intended as a map to the creation of sustainable knowledge societies.

In this regard, the Maturity Model may be re-presented as a table of questions for policy makers. This will enable them to determine whether policies under development place adequate emphasis on digital inclusion (see Figure 5).

Figure 5. DLMM for policy making

	GOVERNANCE	INFRASTRUCTURE	EDUCATION & HUMAN CAPITAL	INNOVATION
HUMAN VALUES	<i>How can governance promote human values?</i>	<i>How can infrastructure promote human values?</i>	<i>How can education and human capital promote human values?</i>	<i>How can innovation promote human values?</i>
PARTICIPATION	<i>How can governance promote participation?</i>	<i>How can infrastructure promote participation?</i>	<i>How can education and human capital promote usage?</i>	<i>How can innovation promote participation?</i>
USAGE	<i>How can governance promote usage?</i>	<i>How can infrastructure promote usage?</i>	<i>How can education and human capital promote usage?</i>	<i>How can innovation promote usage?</i>
ACCESS	<i>How can governance promote access?</i>	<i>How can infrastructure promote access?</i>	<i>How can education and human capital promote access?</i>	<i>How can innovation promote access?</i>

By examining policy-making decisions in this manner, digital development may thus be promoted. In addition, at every level, these decisions must be supported by the knowledge of what kinds of digital literacies are necessary to fulfil the requirements of that level. This will enable the formation of knowledge societies on the basis of inclusion.

CONCLUSION

This article has provided a link between digital development and digital literacies. It highlights that in order to reach the level of human values, there needs to be a shift from functional literacy to transformational literacy. This means that the current and future generations need to be supported with other competencies to enable them to be part of sustainable knowledge societies.

Digital literacy can promote inclusion, and reduce socio-economic disparities by promoting participation at an economic level. In order to enable widespread inclusion, knowledge societies need to develop in line with digital inclusion policies. The maturity model presented in this paper brings together these two concepts, offering policy makers the opportunity to identify key areas of development, and ensuring a digitally enabled future.

REFERENCES

- Armenta, A., Serrano, A., Cabrera, M., & Conte, R. (2012). The new digital divide: The confluence of broadband penetration, sustainable development, technology adoption and community participation. *Information Technology for Development*, 18(4), 345–353. doi:10.1080/02681102.2011.625925
- Bélisle, C. (2006). Literacy and the digital knowledge revolution. In *Digital literacies for learning*. Facet Publishing.
- Belshaw, D. (2012). *What is 'digital literacy'? A Pragmatic investigation* (Doctoral dissertation). Durham University.
- Bonfadelli, H. (2002). The Internet and Knowledge Gaps: A Theoretical and Empirical Investigation Survey. *European Journal of Communication*, 17(1), 65–84. doi:10.1177/0267323102017001607
- Chandrasekar, G., & Sharma, R. S. (2010). Analysing knowledge disparity and value creation: Towards a K-Gini coefficient. *International Journal of Knowledge-Based Development*, 1(3), 242–262. doi:10.1504/IJKBD.2010.035661
- Eshet-Alkalai, Y. (2004). Digital literacy: A conceptual framework for survival skills in the digital era. *Journal of Educational Multimedia and Hypermedia*, 13(1), 93.
- Gilster, P. (1997). *Digital literacy*. New York: Wiley Computer Publications.
- Harding, C. (2016). The New Digital Divide. *Readwrite*. Retrieved from <http://readwrite.com/2016/03/04/new-digital-divide>
- Hargittai, E. (2009). An update on survey measures on web-oriented digital literacy. *Social Science Computer Review*, 27(1), 130–137. doi:10.1177/0894439308318213
- Hargittai, E., & Hinnant, A. (2008). Digital Inequality: Differences in Young Adults Use of the Internet Survey. *Communication Research*, 35(5), 601–621. doi:10.1177/0093650208321782
- Hoechsmann, M., & DeWaard, H. (2015). Mapping digital literacy policy and practice in the Canadian education landscape. *Media Smarts*. Retrieved from <http://mediasmarts.ca/sites/mediasmarts/files/publication-report/full/mapping-digital-literacy.pdf> (Accessed on 16 March 2016)
- Humphrey, W. S. (1988). Characterizing the software process: A maturity framework. *IEEE Software*, 5(2), 73–79. doi:10.1109/52.2014

Lanham, R. A. (1995). Digital literacy. *Scientific American*, 273(3), 198–199.

Mansell, R., & Tremblay, G. (2013). Renewing the Knowledge Societies Vision: Towards Knowledge Societies for Peace and Sustainable Development. Report prepared for the WSIS+10 Review for the Communication and Information Sector, UNESCO and for presentation in the “Knowledge Societies, Stakeholder Accountability for Sustainable Development” Panel at the UNESCO WSIS+10 Conference, Paris, France.

Morris, D. S., & Morris, J. S. (2013). Digital Inequality and Participation in the Political Process: Real or Imagined. *Social Science Computer Review*, 1(5), 589–600. doi:10.1177/0894439313489259

Sharma, R. S., Iqbal, M. I. N. A., & Victoriano, M. M. (2012). On the use of benchmarking and good practices for knowledge management for development. *Knowledge Management Research & Practice*, 11, 1–15.

Sharma, R. S., & Mokhtar, I. A. (2006). Bridging the digital divide in Asia – Challenges and solutions. *International Journal of Technology. Knowledge in Society*, 1(3), 15–30.

Sharma, R. S., Ng, E. W. J., Dharmawirya, M., & Lee, C. K. (2008). Beyond the Digital Divide: A Conceptual Framework for Analyzing Knowledge Societies. *Journal of Knowledge Management*, 12(5), 151–164. doi:10.1108/13673270810903000

Sharma, R. S., Samuael, E. M., & Ng, E. W. J. (2009). Beyond the digital divide: Policy analysis for knowledge societies. *Journal of Knowledge Management*, 13(5), 373–286. doi:10.1108/13673270910988178

Spangenberg, J. H. (2005a). Economic sustainability of the economy: Concepts and indicators. *International Journal of Sustainable Development*, 8(1), 47–64. doi:10.1504/IJSD.2005.007374

Spangenberg, J. H. (2005b). Will the information society be sustainable?: Towards criteria and indicators for a sustainable knowledge society. *International Journal of Innovation and Sustainable Development*, 1(1), 85–102. doi:10.1504/IJSD.2005.008082

UNESCO. (2005). *Toward Knowledge Societies*. UNESCO Publishing. Retrieved from <http://unesdoc.unesco.org/images/0014/001418/141843e.pdf>

ADDITIONAL READING

Coiro, J., Knobel, M., Lankshear, C., & Leu, D. J. (Eds.). (2014). *Handbook of research on new literacies*. Routledge.

European Commission. (2016). The Digital Economy & Society Index (DESI). Retrieved 18th April, 2016, from <https://ec.europa.eu/digital-single-market/en/desi>

Hsieh, J. J. P. A., Rai, A., & Keil, M. (2011). Addressing digital inequality for the socioeconomically disadvantaged through government initiatives: Forms of capital that affect ICT utilization. *Information Systems Research*, 22(2), 233–253. doi:10.1287/isre.1090.0256

Lankshear, C., & Knobel, M. (2008). *Digital literacies: Concepts, policies and practices* (Vol. 30). Peter Lang.

Ramírez, R. (2007). Appreciating the contribution of broadband ICT with rural and remote communities: Stepping stones toward an alternative paradigm. *The Information Society*, 23(2), 85–94. doi:10.1080/01972240701224044

Van Dijk, J. A. (2005). *The deepening divide: Inequality in the information society*. Sage Publications.

KEY TERMS AND DEFINITIONS

Digital Development: The various levels of the digital divide which must be bridged to promote wide spread digital usage within a society.

Digital Literacy Maturity Model: A consistent set of measurements for researchers to study digital literacy in accordance to the level of digital development within knowledge societies.

Digital Literacy: Digital literacy consists of competence in the basic skills to utilize digital technologies, an understanding of how these competences may be utilised to create context to practices and subsequently to participate socially, culturally and economically, and it allows for the intellectual empowerment of individuals to transform society.

Functional Digital Literacy: The basic competences or skills necessary to engage in the digital society.

Knowledge Societies: Societies which possess the necessary and sufficient conditions for growth in the knowledge economy, have high knowledge absorptive capacity and complex chains of creation, production and distribution, and consist of sustainable learning communities which emphasise innovation.

Socio-Economic Digital Literacy: the ability of digital users to engage in the social and economic structures of the digital society.

Sustainable Development: The process of developing while ensuring that the development does not promote existing inequalities, nor does it hinder the resources and abilities for future generations to continue to develop and progress.

Transformational Digital Literacy: The empowerment of digital users, to be transformed intellectually, which ultimately changes society.

D

Teaching Media and Information Literacy in the 21st Century

Sarah Gretter

Michigan State University, USA

Aman Yadav

Michigan State University, USA

INTRODUCTION

Multimedia platforms such as blogs, social networks, forums, and video sharing websites have become a key component of communication in the 21st century. Ranging from flash news, popular press, and activism to trends, scandals, and advertising, these platforms have also become a repository of media and information in today's hyper-connected society. Both individuals and media professionals often create, curate, and circulate content (i.e., user-generated content) in digital media spaces, thereby saturating media spaces with images and information that shape our digital culture (Gleason, 2013). Possessing the competencies to understand how information is conveyed through digital media is therefore an important skill to empower citizens to recognize its functions and effects on human communication. To address these objectives, the United Nations Educational, Scientific, and Cultural Organization (UNESCO) developed a media & information literacy (MIL) framework that encompasses the skills needed by 21st century citizens to critically evaluate information communicated through different media sources (UNESCO, 2013). Drawing from the UNESCO MIL framework, this chapter examines the skills needed by individuals to evaluate information presented through digital mass media, and discusses the role that educators can play in its instruction. The next section provides a brief history of MIL, followed by a look at the specific competencies that compose MIL.

Then, the chapter proceeds to discussing the role of educators in MIL instruction and concludes with the implications of MIL implementation in educational settings.

BACKGROUND

What does it mean to be literate in the 21st century? While literacy has traditionally been contained to reading and writing skills, communication in the 21st century has expanded these customary views of literacy into an ever-evolving concept (Hobbs & Jensen, 2009). In today's world, unfiltered information is available across multiple media platforms, such as television or newspapers, but more particularly on the Internet. Because media and other information providers are instrumental in shaping the perceptions, beliefs, and attitudes of individuals in today's digital age (Guzzetti & Lesley, 2015), being literate in today's society therefore includes being able to read, write, and communicate across a range of platforms, tools, and media. As a result, individuals need to master an array of literacy skills beyond basic reading and writing abilities (Livingstone et al., 2014). Citizens who are not aware of how media and information systems function are more likely to accept media messages as facts, while individuals who possess media and information literacy skills are able to evaluate and draw their own conclusion from the constant flow of mediated information (Potter, 2004).

Media & Information Literacy (MIL) is an umbrella term that bridges information literacy with media literacy. Because the Internet is a digital platform that hosts multitudes of archives of mediatized information, navigating 21st century digital information implies the convergence of different sets of skills to address the challenges of our globalized world. Modern information systems are complex and multifaceted, and require individuals to be informed and engaged citizens in order to make sense of the mediatized information that surrounds them. MIL thus describes the 21st century skills that individuals need to critically evaluate information via various media sources and to become critical consumers of information (UNESCO, 2013). UNESCO (2013) articulated that media and information literate citizens should understand the importance of accessibility to information, know how to evaluate its veracity, and use it in ethical ways. Additionally, they should understand media functions and purposes, and engage with them for self-expression. For instance, individuals should be able to distinguish when media and information are used either for entertainment, decision making, problem solving, learning, or communicating with others. They should also understand how these purposes are related to the roles and the functions that different media play, and that based on these functions, different media adhere to different professional and ethical standards. With this understanding comes the ability to practice one's own digital skills to engage with media and information for personal purposes, such as creating user-generated content, evaluating the credibility of a source, or communicating with others. This conceptual view of MIL is represented in figure 1. below.

The social implication of being media information literate in the 21st century is informed participation in digital communication (Jenkins, 2009; Lee, 2013; UNESCO, 2013). MIL skills allow users to move from being passive consumers of digital information and media to being actively engaged in the information systems that shape their culture (Lankshear & Knobel, 2008).

For example, an informed media and information literate person would recognize and react to media biases when present, would engage in an ethical manner with online social exchanges, or would participate in a digital culture by creating content relevant to that culture. MIL skills not only foster individuals' critical thinking and engagement with contemporary issues, but also allow them to take part in our era's "participatory culture" (Jenkins, 2009). A participatory culture allows users to communicate through the creation of content to actively use media to engage audiences (Jenkins, 2009). Therefore, participants who are equipped with MIL skills can help shape today's digital society and draw their own conclusions from the media and information that structures their culture, instead of simply accepting these media messages as unchangeable facts. Hence, possessing MIL skills can further the gap between those who participate in the culture, and those who do not because they have not acquired the necessary analytical and technical skills to do so (Jenkins, 2009).

UNESCO has focused on issues of media literacy since the 1960's, but acknowledged in the early 2000's that technology was changing the role of media in society and that soon, individuals would need to possess new skills to make sense of new types of communication and ways to access information (Frau-Meigs, 2007). In knowledge societies, information includes and depends on the process of communication, and as a result, media literacy and information literacy are intrinsically connected (Lau, 2013). Livingstone et al. (2008) argued that despite their traditionally divergent disciplinary backgrounds, the object of inquiry in both media literacy and information literacy started to be united to understand "the public's understanding of and effective engagement with media, information and communication technologies of all kinds" (p. 2). Koltay (2011) declared that one of the most salient commonalities between media literacy and information literacy was the analytical and critical thinking skills needed to interact with media and information.

Figure 1. A conceptual model of media and information literacy

Source: Grizzle et al. (2013)



UNESCO (2013) conjoined both media literacy and information literacy using the overlap of critical and analytical thinking practices in a new media & information literacy (MIL) framework. UNESCO recognized that for some, information literacy would be considered the broader area of study with media literacy as part of it, while for others media literacy would be seen as the broader field with information literacy as part of

it. UNESCO's MIL framework, therefore, sought to bridge these two terms together in light of the converging platforms that constitute 21st century communication systems, such as Internet-based media and information platforms, and proposed MIL as the umbrella term encompassing all other literacies, such as visual literacy and digital literacy. MIL is a broad transliteracy concept because it is meant to adapt to the ever-evolving

technologies and communication needs of a specific time or place. It can be applied to traditional media (e.g., books, print newspapers, television) as well as to contemporary media (e.g., Internet, online media), and offers the flexibility to be used in future information systems, regardless of their technology base or scope. Overall, MIL represents a set of analytical competencies that can transfer from one communication system to another and reflect the technological needs of 21st century digital communications. Given the scope of this encyclopedia, this chapter focuses specifically on digital communication and technology in the 21st century while acknowledging that MIL can be applied to and practiced in more traditional communication systems as well.

MEDIA AND INFORMATION LITERACY

A 21st Century Competency

Thanks to web-based technologies and digital media in the 21st century, individuals have more and more opportunities to become creators of knowledge and to take part in societal issues rather than simply being passive consumers of information (Buckingham, 2015; Lankshear & Knobel, 2008). UNESCO's media & information literacy framework provides a set of competencies, or in other words a set of knowledge, skills, and attitudes, that aim to empower citizens to critically evaluate and understand the contents of today's media and information systems (UNESCO, 2013). Because MIL relies on a combination of technical, analytical, and creative skills, it is a literacy that transcends media and type of information. That is, one can apply MIL skills while reading a novel or a newspaper, while watching television shows and advertisements, while surfing the Internet, or while engaging with others on social media. Because of the multi-faceted aspect of online communications in the 21st century and the increasing reliance on the multimodal transmission

of information (e.g., visual, textual, audio, etc.), this chapter emphasizes MIL competencies in the context of digital communication.

It is important to note that in the context of 21st century communication, the term media corresponds to the means by which information is delivered, and can be both professionally produced content—typically labeled as “the media”—or user-generated content. Furthermore, UNESCO (2013) described information literacy as the skills required to seek, evaluate and create information for personal, social, educational, or professional needs (Wilson et al., 2013). Together, being literate in terms of media and information, therefore, involves the ability to recognize the power of media, to evaluate the content of the information they convey, and to produce user-generated content (Moeller et al., 2011; Wilson et al., 2013). UNESCO (2013) summarized these MIL components under three main competencies: access and retrieval, understanding and evaluation, and creation and sharing (UNESCO, 2013). These competencies, along with their sub-skills, are represented in figure 2.

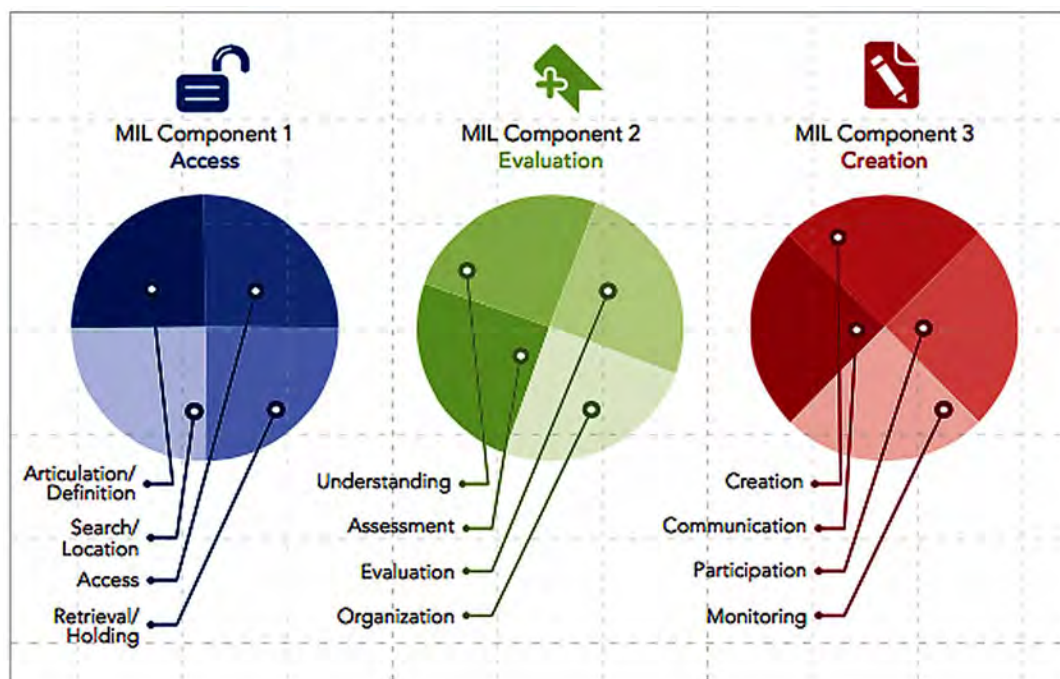
While analytical skills are central to MIL to evaluate media content, developing the retrieval and the creative aspect of MIL skills also relies on a set of technical skills that are intrinsically linked to technology and information systems in the 21st century (Davies, 2011; Jenkins, 2009; Wilson et al., 2013). Understanding how to access and retrieve information, along with knowing how to create and share personal content online depends on the mastery of technological and computing abilities. These skills are reviewed in the following section.

Media and Information Literacy in a Technology-Driven World

In today's digitally connected world, technology is pervasive from the use of social media (e.g., Facebook, Youtube, Snapchat, etc.) to the widespread presence of mobile devices (e.g., smartphones, laptops, tablets, etc.). A recent survey found that 95% of teenagers go online on

Figure 2. MIL components

Source: UNESCO, 2013



a daily basis, 72% of all teens spend time with friends via social media, and 79% of all teens use instant messages with their friends (Lenhart, et. al, 2015). These staggering numbers highlight the need for young individuals to understand issues related to privacy and security on the Internet, as the risks they encounter in relation to anonymity, persuasion, or cyber-bullying grow exponentially (Livingstone & Brake, 2010).

The key to developing individuals' competencies in MIL involves creating awareness of how computerized communication and web-based digital media work. As young citizens become better informed about how information travels over the Internet, they will better understand its implications for privacy and security of information in digital communications. One of the prerequisites for MIL in a digital world is to form a solid foundation of technological skills—often referred to as ICT (Information & Communication Technology) skills (Wilson et al., 2013). For that reason, MIL also has a natural connection to components of computational thinking (CT), which has been sug-

gested to be a 21st century cognitive and creative thinking skill that combines problem solving with principles of computing (Wing, 2006; DeSchryver & Yadav, 2015). For instance, one computational thinking idea that can support the development of MIL involves using “computational tools to analyze and study data and working with large data sets to analyze, visualize, and draw conclusions from trends” (College Board, p. 1). One of the key components of MIL is individuals' ability to access and critically evaluate information, particularly how information can be manipulated to deliver specific messages. As people engage in computational thinking activities to analyze and visualize data, they also develop important MIL practices of how information can be visualized to convey specific messages. Specifically, as individuals create data visualizations or communicate their ideas using a variety of media, they learn to negotiate how to best represent information to their audience, which would be determined by the essence of the message they want to convey for a specific purpose. This process of evaluat-

ing and presenting information to convey the gist of an idea allows individuals to critically assess the authenticity of the information encountered through digital media.

Recently, many nations have come to recognize the need for their youth to acquire technological skills and to prepare them to successfully manage 21st century information. For instance, analyzing scientific claims online, locating relevant research, building arguments from a variety of resources, or producing appropriate media contents are examples of standards now expected from students in the United States (Governors Association Center for Best Practices, 2010; International Society for Technology in Education, 2015; Next Generation Science Standards, 2013). These standards aim to benefit students' development of academic skills as well as personal and social skills in a digital world. A number of national and international educational organizations have also acknowledged the convergence and importance of student competencies in MIL and the need for students for being responsible users of digital media and information (Governors Association Center for Best Practices, 2010; International Society for Technology in Education, 2015; Next Generation Science Standards, 2013; Partnership for 21st century, 2014).

As an example, the national computer science curriculum in the United Kingdom highlighted the imperative need for students to “understand a range of ways to use technology safely, respectfully, responsibly and securely, including protecting their online identity and privacy; recognize inappropriate content, contact and conduct, and know how to report concern” (Microsoft, 2014, p. 42). Similarly, within the United States of America, College Board has launched a new advanced placement course entitled Computer Science Principles (CSP) that focuses on six computational thinking practices and seven big ideas of computer science. The CSP framework is designed to allow students to understand how digital manipulation and storage of personal information has implications for privacy and security of that data (College Board,

2014). Specifically, the CSP framework focuses on advancing students' understanding of how the Internet functions and grasping how computing can impact people and societies. While not labeled as MIL skills, there is considerable overlap between media & information literacy practices outlined in the UNESCO framework and the components in both the UK computing curriculum and the US computer science principles frameworks.

However, while using technology effectively to access and assess information and to effectively use media for communication is now considered an essential 21st century skill for students, teaching these digital skills requires educators to exhibit the same competencies themselves (Wilson et al., 2014). For MIL to become a leading practice in the 21st century and for students to acquire these skills, there is an urgent need to prepare educators in MIL instruction. The following section addresses the qualities that educators should possess to successfully integrate MIL in their teaching.

Educators' Role in Media and Information Literacy Instruction

In today's technology-driven and media-saturated world, educators will play a critical role in addressing the disparities that are emerging between those who can—or cannot—find, analyze, and critically evaluate information and media content to participate in their digital environments (Wilson et al., 2013). UNESCO in particular, envisioned that “initial focus on teachers is a key strategy to achieving a multiplier effect: from information-literate teachers to their students and eventually to society at large” (Wilson et al., 2013, p. 17). Preparing teachers to teach MIL skills is an essential step in order to promote 21st century MIL competencies at a large scale. The UNESCO MIL curriculum framework for teachers describes the competencies that educators should possess and provides examples for educators on how to embed media and information literacy practices in their teaching (Wilson et al., 2013).

According to the UNESCO, teachers should understand the role of media and information in their daily lives; understand media content and its uses; know how to access information effectively and efficiently; critically evaluate information; be familiar with both new and traditional media formats; and be able to situate the sociocultural context of media content (Wilson et al., 2013). UNESCO also identified various dimensions, including policy, curriculum and assessment, pedagogy, and professional training that need to be put into practice for the MIL framework for teachers to be successful (Wilson et al., 2013). For example, educators need to examine educational policies and national standards to better understand what role MIL might play in their own educational context. They also need to learn how to apply MIL to their teaching and understand students' interactions with media and information in their daily lives. In addition, educators must understand how MIL can be integrated in their school curriculum and how they can assess their students' acquisition of MIL skills.

However, promoting MIL skills among students and implementing them in their own teaching can be difficult for teachers to achieve on their own even if they themselves might possess these MIL skills. Progressing from being media and information literate to knowing how to embed MIL in pedagogical practices requires support at various institutional levels, from teacher training for pre-service teachers, to professional development for inservice teachers or educators in non-traditional educational settings. In order to address the growing need for teacher training in media literacy and information literacy (Martens, 2010), a UNESCO international expert group working on the MIL framework suggested that there is an important need to understand teachers' conceptions of media and information literacy (Pérez Tornero, 2008). While a number educational reform initiatives, such as the Next Generation Science Standards in the United States or the computer science curriculum in the United Kingdom, have recently integrated MIL-related concepts in their student

standards, implementing MIL in teacher education and teacher professional development may require time and institutional changes. The following section offers recommendations that educators can flow on their own as a first step towards the goal of promoting MIL to their students.

RECOMMENDATIONS FOR EDUCATORS

Given the current focus on implementing MIL in educational settings, it is critical that teachers understand and adapt to 21st century media and information in order to prepare learners to use technology in their personal, academic, and future professional lives. The following set of recommendations represent active steps that educators can take in order to accomplish that goal.

1. MIL is an attitude. Encourage your students to habitually question the information they encounter in their daily lives as well as the media through which the information is conveyed.
2. MIL is a literacy that transcends types of communication (e.g., books, radio, television, social networks). Practice it with your students both with traditional and new media.
3. MIL is knowledge about how information is transmitted. Teach your students how information is passed on through different media forms.
4. MIL is an investigative activity. Show your students the necessary steps to access and retrieve information based on their needs.
5. MIL is awareness. Help your students develop an awareness of how traditional and digital media providers transform and distort information to convey specific messages.
6. MIL is an evaluative skill. Model to your students how they can assess the veracity of information transmitted through different media.

7. MIL is creative. Encourage and guide students' creation of media to participate in their digital culture of interest.

In addition, organizations such as the International Society for Technology in Education (<http://www.iste.org/standards/ISTE-standards/standards-for-students>), and the Partnership for 21st Century Skills (http://www.p21.org/storage/documents/P21_framework_0515.pdf) further explicate some of the MIL competencies that both students and educators should possess to successfully evaluate, use, and create digital information and media in the 21st century. We recommend readers to look at their standards in order to obtain a complete picture of the current landscape in MIL-related skills.

SUMMARY AND FUTURE DIRECTIONS

This chapter discussed the importance of media & information literacy for digital communication in the 21st century. It provided a brief history of the MIL concept, as well as a discussion of the competencies that compose MIL. We also explored the essential role that teachers play in teaching MIL skills in educational settings and presented recommendations for educators to take the first step in embedding MIL principles in their teaching. Overall, this chapter presented an adaptive set of competencies that will help individuals situate themselves within and contribute to today's digital culture.

Future directions should explore inclusion of MIL within computing concepts, which has been highlighted as an area of importance in the classroom by many scholars (Barr & Stephenson, 2011; Eisenberg et al., 2005; Felini, 2015; Hobbs & Jensen, 2009; Perkovic et al., 2010; Qualls & Sherrell, 2010; Thomas, 2004; Wing, 2006; Yadav et al., 2014). Specifically, future work in this field should examine how exposure to computational thinking constructs also develops students' MIL

skills. As discussed previously, future research should also help to better understand teachers' conceptions of MIL in order to identify potential challenges and barriers to its integration in teacher education and teacher professional development. Furthermore, future research should also look at the implementation and effectiveness of MIL trainings, such as the UNESCO curriculum for teachers; both at the institutional and the individual level. At the institutional level, research is needed to understand the role that the MIL curriculum could play in teacher education programs, either as a domain-general or as a domain-specific set of skills required of future teachers. At the individual level, studies should look at teachers' engagement with MIL on a personal level in order to understand teachers' attitudes towards embedding MIL in their teaching. These future directions will contribute to UNESCO's overall mission to promote MIL on a global level, while supporting learners' acquisition of these vital 21st century skills.

REFERENCES

- Barr, V., & Stephenson, C. (2011). Bringing computational thinking to K-12: What is involved and what is the role of the computer science education community? *ACM Inroads*, 2(1), 48–54. doi:10.1145/1929887.1929905
- Buckingham, D. (2015). Do we really need media education 2.0? Teaching media in the age of participatory culture. In T. Lin, V. Chen, & S. C. Chai (Eds.), *New media and learning in the 21st century: A socio-cultural perspective* (pp. 9–21). Singapore: Springer. doi:10.1007/978-981-287-326-2_2
- College Board. (2014). *Advanced Placement Computer Science Principles: Curriculum framework*. Retrieved from: <http://secure-media.collegeboard.org/digitalServices/pdf/ap/ap-computer-science-principles-curriculum-framework.pdf>

- Davies, R. S. (2011). Understanding technology literacy: A framework for evaluating educational technology integration. *TechTrends*, 55(5), 45–52. doi:10.1007/s11528-011-0527-3
- DeSchryver, M. D., & Yadav, A. (2015). Creative and computational thinking in the context of new literacies: Working with teachers to scaffold complex technology-mediated approaches to teaching and learning. *Journal of Technology and Teacher Education*, 23(3), 411–431.
- Eisenberg, M. B., Lowe, C. A., & Spitzer, K. L. (2004). *Information literacy: Essential skills for the information age*. Westport, CT: Greenwood.
- Felini, D. (2015). Crossing the bridge: Literacy between school education and contemporary cultures. *Research on Teaching Literacy Through the Communicative and Visual Arts*, 2, 19–25.
- Frau-Meigs, D. (2007). *Media Education. A Kit for Teachers, Students, Parents and Professionals*. Retrieved from http://portal.unesco.org/ci/en/ev.php-URL_ID=27056&URL_DO=DO_TOPIC&URL_SECTION=201.html
- Gleason, B. (2013). #Occupy Wall Street: Exploring Informal Learning with Twitter. *The American Behavioral Scientist*, 57(7), 966–982. doi:10.1177/0002764213479372
- Governors Association Center for Best Practices, Council of Chief State School. (2010). *Common Core State Standards for English Language Arts*. Retrieved from: <http://www.corestandards.org/ELA-Literacy/>
- Grizzle, A., Moore, P., Dezuanni, M., Asthana, S., Wilson, C., Banda, F., & Onumah, C. (2013). *Media and information literacy: policy and strategy guidelines*. UNESCO. Retrieved from: <http://unesdoc.unesco.org/images/0022/002256/225606e.pdf>
- Guzzetti, B., & Lesley, M. (2016). *Handbook of Research on the Societal Impact of Digital Media*. Hershey, PA: IGI Global. doi:10.4018/978-1-4666-8310-5
- Hobbs, R., & Jensen, A. (2009). The past, present, and future of media literacy education. *The Journal of Media Literacy Education*, 1(1), 1–11.
- International Society for Technology in Education (ISTE). (2015). *ISTE standards for students*. Retrieved from https://www.iste.org/docs/pdfs/20-14_ISTE_Standards-S_PDF.pdf
- Jenkins, H. (2009). *Confronting the challenges of participatory culture: Media education for the 21st century*. Boston, MA: MIT Press.
- Koltay, T. (2011). The media and the literacies: Media literacy, information literacy, digital literacy. *Media Culture & Society*, 33(2), 211–221. doi:10.1177/0163443710393382
- Lankshear, C., & Knobel, M. (2008). *Digital literacies: Concepts, policies and practices* (Vol. 30). New York, NY: Peter Lang.
- Lee, A. Y. (2013). Literacy and competencies required to participate in knowledge societies. In A. Lee, J. Lau, T. Carbo, & N. Gendina (Eds.), *Conceptual Relationship of Information Literacy and Media Literacy in Knowledge Societies*. Paris: UNESCO. Retrieved from http://www.observatorioabaco.es/biblioteca/docs/371_UNESCO_INFORMATIONLITERACY_013.pdf
- Lenhart, A., Smith, A., Anderson, M., Duggan, M., & Perrin, A. (2015). *Teens, technology, and friendship*. Retrieved from: http://www.pewinternet.org/files/2015/08/Teens-and-Friendships_FINAL2.pdf
- Livingstone, S., & Brake, D. R. (2010). On the rapid rise of social networking sites: New findings and policy implications. *Children & Society*, 24(1), 75–83. doi:10.1111/j.1099-0860.2009.00243.x
- Livingstone, S., van Couvering, E., & Thumim, N. (2014). Converging traditions of research on media and information literacies: Disciplinary and methodological issues. In *Handbook of research on new literacies* (pp. 103–132). New York, NY: Routledge.

Martens, H. (2010). Evaluating media literacy education: Concepts, theories and future directions. *Journal of Media Literacy Education*, 2(1), 1–22.

Microsoft. (2014). *Computer science in the national curriculum*. Retrieved from: http://www.slideshare.net/Microsofteduk/computer-science-in-the-national-curriculum?from_action=save

Moeller, S., Joseph, A., Lau, J., & Carbo, T. (2011). *Towards media and information literacy indicators*. Paris: UNESCO. Retrieved from: [http://www.unesco.org/new/en/communication-and-information/resources/publications-and-communicationmaterials/publications/full_list/towards-information-literacy-indicators/\(3.8.2014\)](http://www.unesco.org/new/en/communication-and-information/resources/publications-and-communicationmaterials/publications/full_list/towards-information-literacy-indicators/(3.8.2014))

National Governors Association Center for Best Practices, Council of Chief State School. (2010). *Common Core State Standards for English Language Arts*. Retrieved from <http://www.corestandards.org/ELA-Literacy/>

Next Generation Science Standards (NGSS). (2013). *The Next Generation Science Standards*. Retrieved from: <http://www.nextgenscience.org/next-generation-science-standards>

Partnership for 21st Century Skills. (2014). *Framework for state action on global education*. Retrieved from http://www.p21.org/storage/documents/Global_Education/P21_State_Framework_on_Global_Education.pdf

Pérez Tornero, J. M., Celot, P., & Varis, T. (2007). *Study on the current trends and approaches to Media Literacy in Europe*. European Commission. Retrieved from: http://ec.europa.eu/culture/library/studies/literacy-trends-report_en.pdf

Perković, L., Settle, A., Hwang, S., & Jones, J. (2010). A framework for computational thinking across the curriculum. In *Proceedings of the fifteenth annual conference on Innovation and technology in computer science education* (pp. 123–127). Ankara, Turkey: Association for Computing Machinery. doi:10.1145/1822090.1822126

Potter, W. J. (2004). *Theory of media literacy: A cognitive approach*. Thousand Oaks, CA: Sage. doi:10.4135/9781483328881

Qualls, J. A., & Sherrell, L. B. (2010). Why computational thinking should be integrated into the curriculum. *Journal of Computing Sciences in Colleges*, 25(5), 66–71.

Thomas, N. P. (2004). *Information literacy and information skills instruction: Applying research to practice in the school library media center*. Westport, CT: Libraries Unltd Incorporated.

United Nations Educational, Scientific, and Cultural Organization (UNESCO). (2013). *Global Media and Information Literacy (MIL) Assessment Framework: Country Readiness and Competencies*. Retrieved from: <http://unesdoc.unesco.org/images/0022/002246/224655e.pdf>

Wilson, C., Grizzle, A., Tuazon, R., Akyempong, K., & Cheung, C. K. (2013). *Media and information literacy curriculum for teachers*. UNESCO. Retrieved from: <http://www.unesco.org/new/en/communication-and-information/resources/publications-and-communication-materials/publications/full-list/media-and-information-literacy-curriculum-for-teachers/>

Wing, J. M. (2006). Computational thinking. *Communications of the ACM*, 49(3), 33–35. doi:10.1145/1118178.1118215

Yadav, A., Mayfield, C., Zhou, N., Hambrusch, S., & Korb, J. T. (2014). Computational thinking in elementary and secondary teacher education. *ACM Transactions on Computing Education*, 14(1), 1–16. doi:10.1145/2576872

ADDITIONAL READING

Hobbs, R., & Jensen, A. (2009). The past, present, and future of media literacy education. *The Journal of Media Literacy Education*, 1(1), 1–11.

Jenkins, H. (2009). *Confronting the challenges of participatory culture: Media education for the 21st century*. Boston, MA: MIT Press.

Koltay, T. (2011). The media and the literacies: Media literacy, information literacy, digital literacy. *Media Culture & Society*, 33(2), 211–221. doi:10.1177/0163443710393382

Lankshear, C., & Knobel, M. (2008). *Digital literacies: Concepts, policies and practices* (Vol. 30). New York, NY: Peter Lang.

Partnership for 21st Century Skills. (2014). Framework for state action on global education. Retrieved from http://www.p21.org/storage/documents/Global_Education/P21_State_Framework_on_Global_Education.pdf

United Nations Educational, Scientific, and Cultural Organization (UNESCO). (2013). Global Media and Information Literacy (MIL) Assessment Framework: Country Readiness and Competencies. Retrieved from: <http://unesdoc.unesco.org/images/0022/002246/224655e.pdf>

Wilson, C., Grizzle, A., Tuazon, R., Akyempong, K., & Cheung, C. K. (2013). Media and information literacy curriculum for teachers. UNESCO. Retrieved from: <http://www.unesco.org/new/en/communication-and-information/resources/publications-and-communication-materials/publications/full-list/media-and-information-literacy-curriculum-for-teachers/>

Wing, J. M. (2006). Computational thinking. *Communications of the ACM*, 49(3), 33–35. doi:10.1145/1118178.1118215

KEY TERMS AND DEFINITIONS

Competencies: A set of knowledge, skills and attitudes.

Computational Thinking: A way of thinking and solving problems based on computer science concepts.

Media: The diverse body of media technologies that reach broad audiences through mass communication.

Media and Information Literacy: The skills required to critically access and assess mediated information while understanding media functions in our daily lives.

Information: The facts or data conveyed or represented via various media.

Participatory Culture: A culture in which individuals are engaged with media instead of simply being passive consumers of information.

Twenty-First Century Skills: A set of skills that individuals need to succeed in the 21st century.

Nigerian Undergraduate Students' Computer Competencies and Use of Information Technology Tools and Resources for Study Skills and Habits' Enhancement

D

Adekunle Olusola Otunla
University of Ibadan, Nigeria

Caleb Okoro Amuda
University of Ibadan, Nigeria

INTRODUCTION

Many educational institutions have adopted policies favoring the implementation of modern technology prompted by Information and Communication Technology (ICT). Modern technologies like the Internet, mobile tele-communication and World Wide Web (www) has become innovative tools that could engage students in continuous learning across homes, classes, campuses, offices, e.t.c. Thus, the role of technology using the Internet, cloud computing and networked computers in teaching and learning is turning teachers from providers of information to become facilitators of learning. As a result of technology-driven learning approaches there is a transition of teaching and learning communities from teacher-focused with traditional approaches to student-centered learning which is largely Technology-driven through the use of tools and resources that surpasses any other previous technology. Consequently, Davis (2010) affirms that technology integration within and outside the classroom is modifying the learning environment from teacher-centred to learner-centred with opportunities for personalized learning experiences. Bodys (2005) also reported that ICT tools and resources allowed students to be more individually active in the learning process and become more independent in making

decisions about how and what they need to learn using electronics learning resources.

Globally, different types of Computer-Based teaching and learning approaches have been developed to achieve the desired learning objectives and outcomes. Examples of computer technology applications, tools and resources for teaching and learning that cuts across all educational levels include; computer-assisted instruction (CAI), computer-assisted learning (CAL), e-learning, interactive video, multi-channel learning, virtual learning, virtual fieldtrips laboratories, virtual libraries, web conferencing, web chatting, digital story-telling, asynchronous online discussion (AOD) flip learning, e-mail communication and other forms of electronics and mobile learning among many other emerging learning technologies. Any of the listed technologies could be combined diversely into 'Learning Management Systems' (LMSs).

BACKGROUND

ICT integration among higher institution students particularly in Nigeria has attracted attentions through various studies. For example, Odusanya and Bamgbala (2002) reported ICT uses among final year students at the University of Lagos-

DOI: 10.4018/978-1-5225-2255-3.ch200

Nigeria; Jagboro (2003) reported on postgraduate students' use of the Internet to search for academic materials at Obafemi Awolowo University, Ile-Ife, Nigeria. Similarly, Ajuwon (2003) reported high rate of regular use of the Internet by medical and nursing students at the University College Hospital, Ibadan, Nigeria and Bello, Arogundade, Sanusi, Ezeoma, Abioye-Kuteyi and Akinsola (2004) also reported that a lower percentage of respondents demonstrated good knowledge of computers and IT at Obafemi Awolowo University Teaching Hospital, Ile-Ife, Nigeria. Ezekoka and Nwosu (2010) also reported that ICT has been found to be very useful in the teaching and learning processes among Nigerian students because of the extensive capacity to store and manipulate information as well as its unmatched ability to serve simultaneously many individual students in different locations as supplementary to classroom instruction. While writing of ICT policy framework, Adomi and Kpangban (2010) observed that the National Policy on Education (Federal Republic of Nigeria, 2004) recognizes the prominent role of ICTs in the modern world and advocated for its integration into the Nigeria educational system especially at the post primary and higher education levels.

Further, Otunla (2013) also stress that educational planners in Nigeria have seen the need for innovative instructional materials and media integration into the school system at all levels of Nigerian education. Adeoye (2013) noted higher institution students can learn through a virtual classroom environment where they can compose e-mails to conduct research via the Internet. The author further observed that the influence of technology on teaching and learning is enormous and could increase to an unimaginable and unpredictable level as new innovations are being introduced to the school system. Otunla and Baiyelo (2013) also observed that computers and Information Communication Technology (ICT) tools and resources such as the Internet, multimedia design, digital motion pictures, digital photography, web broadcasting; and digital print-

ing, teleconferencing, video-conferencing e.t.c. have greatly influenced mode of transaction and the way people relate in the last one and a half decades. Therefore, ICT has permeated every aspect of human existence because it has become an indispensable tool for daily living and especially the educational system. Otunla and Akinyemi (2014) noted that application and integration of mobile technologies in higher education could enhance effective communication between lecturers and students within and outside the four walls of the classroom. More recently, Ezekoka (2015) conducted a survey involving a sample of sixty lecturers in a Nigerian public university on availability of ICT facilities that could promote collaborative learning to increase the level of students' online participation. The author presented some problems encountered by the participants which hindered their optimal performance in online collaborative learning.

At the continental level Uys, Nleya and Molelu (2004) based on an outcome of three case studies of technological adoption within the African continent and abroad suggested that technology needs in Africa could be implemented within a combination of strategically developed framework for technology innovation diffusion. Subsequently, Mtebe, Dachi and Raphael (2011) from their findings on use of technology to improve teaching and learning outcomes in higher education, identified two key issues which are; the need to address problem of university teachers' readiness to accept change in their belief-system about teaching and the use of technology, as well as barriers to access required technologies in terms of hardware and software. Further, Nagunwa, and Lwoga (2012) enumerated some initial challenges faced in their university while they made an attempt to implement computer-based curricula by adopting e-Learning approach and presented the strategic re-establishment of an e-Learning approach. Therefore, it is glaring that higher institutions in Nigeria and African continent at large are taking advantage of the digital media and technologies to support and restructure their mode of teaching

and instructional delivery to prepare students for the world-of-work.

However, computer literacy and skills development are very crucial when it comes to integration of Information Technology (IT) tools, applications and resources in higher education. Without IT competencies and confidence in computer skills students' learning and research opportunities are very limited. Thus, acquiring basic skills in computing and communication media technologies by students is very essential in order for them to function in their academic activities, engage in personal studies and active learning using new technologies. Mcmillan (1996) identified the following characteristics of computer literacy, which connote that the individual; knows how to use: word processing software, an e-mail and a browser for Internet navigation and is capable of; registering or downloading information on at least an auxiliary storage or some other external saving units, recuperate information and print it elsewhere. Stein, Craig and Scollary (1997) further corroborate the need for functional definition which is tailored toward behavioural 'competencies' the authors added a cognitive dimension which defines computer competencies as; the ability to use ICT to identify and search efficiently for specific information in order to build knowledge and develop critical and creative thinking.

Another related technology skill required for knowledge deepening and knowledge creation is 'information literacy' (IL). The College of Dupage Library (CoDL) defines information literacy as the ability to locate, evaluate and use effectively the needed information (CoDL, 2002). Information literacy is also regarded as the set of skills needed to locate, retrieve, analyze and use information. Such that Andretta (2005) made a comparison of standards that was tagged 'Information literacy standards: A practioner's guide' as proposed by various agencies, organizations and professional bodies on skills that people must master before they will be able to perform all the necessary functions to become information literate. The organizations include; American Library Association (ALA),

Association of College and Research Libraries, Society of College, National and University libraries (SCNUL) and Australian and New Zealand Institute for Information Literacy (ANILIL). King (2007) also argues that the attributes of information literacy are multi-dimensional such that it involves employment of both traditional and modern information technology skills to retrieve, manage and present information in an ever widening array of information sources. Therefore, tertiary institution students must become information literate for them to be able to handle acquisition, management and analyzing large quantities of information and data that are widely available from different sources and more importantly from the Internet and the World Wide Web (www). Consequently, as earlier observed students and Internet users generally, need some special computer skills and competencies to be able to handle more effectively the ever-increasing digital information that could be annexed for educational purposes.

STATEMENT OF THE PROBLEM

As a result of the growth in the modern communication technology, and the need for today's higher education learners to become life-long learners through the use of IT tools and resources in preparation for the world-of-work. Moreover, IT is changing the role of tertiary institution students from that of passive-learners to active-learners within the learning environment, so as for them to keep up with current information and trends in their future professional practices. However, computer competencies and skills in the effective use of IT tools, applications and resources becomes a vital ingredient for students' digital mastery, active participation and visibility on the World Wide Web. Further, students are required to be technological-capable in the use and application of IT tools and resources that are abundantly available via the superhighway. Therefore, this chapter, investigate the computer competences and use of information technology

tools and resources in enhancing study skills and habits among undergraduate students in Rivers State, Nigeria.

RESEARCH QUESTIONS

1. What is the level of computer operational skills and competencies of undergraduate students in River State, Nigeria?
2. What is the level of competencies in the use of Information Technology applications and tools among undergraduate students in River State, Nigeria?
3. What is the extent of undergraduate students' use computer tools and software applications as part of their personal learning activities?
4. What is the extent of undergraduate students' use of Information Technology tools and resources as part of their personal learning activities?
5. What is the pattern of undergraduate students' computer competence in Information Technology tools and resources in relation to study skills and habits enhancement?

METHODOLOGY

The study adopted Ex-post facto research design of survey type:

- **Population:** The population of this study was undergraduate students in Rivers State, Nigeria.
- **Sampling Technique and Sample:** The study involved a sample of four hundred and fifty (450) undergraduate students were randomly sampled from three Universities with 150 undergraduate students each from Rivers State University of Science and Technology, Ignatius Ajuru University of Education and University of Port Harcourt all located in Rivers State, Nigeria.

- **Instrumentation:** Three researcher-designed validated instruments were used for data collection. Each of the instruments contains; (A) demographic information section in addition to section (B) that contains items measuring the variables of interest i.e. computer competency, ICT tools and resource use and study habit enhancement. They are:

- **Undergraduate Students' Computer Competency Questionnaire (USCCQ):** It contains a-25 item on computer competency with 3-point Likert scale ratings of 'Highly Competent' (3), 'Slightly Competent' (2) and 'Not Competent' (1).
- **Undergraduate Students' Information Technology tools and resource Use Questionnaire (USITRSQ):** It contains a 25-item on use of ICT tools and resources with 3-point Likert scale ratings of 'Regularly' (3), 'Rarely' (2) and 'Never' (1) respectively and lastly,
- **Undergraduate Students' Information Technology Study Skills and Habit Questionnaire (USITSSHQ):** A 30-item undergraduate students' use of information technology tools and resources to enhance study habit with 4-point Likert scale response options, of Strongly Agree (4); Agree (3); Disagree (2); Strongly Disagree (1).

The three instruments were trial-tested for validation on a group of students who were not part of the sample for the study; using Cronbach Alpha reliability test. The following values were recorded; USCCQ - $r=0.75$, USITRSQ - $r=0.70$ and USITSSHQ - $r=0.62$. The reliability values were considered to be valid and acceptable for the study.

Method of Data Collection and Data Analysis Technique

Primary data was collected directly from the sample covered in the study at the three higher

Table 1. Mean rating on level of computer operational skills and competencies of undergraduate students in Rivers State, Nigeria

Items	Mean	SD	Decision
Starting a computer	2.3384	.78055	*
Launching an application software e.g. Microsoft Word, Excel, etc.	2.6237	.62224	*
Typing document using Microsoft (MS) Word,	2.6667	.67879	*
Slide preparation using MS Power Point	2.4293	.66184	*
Data input on Microsoft Excel	2.3409	.78753	*
Data analysis using Microsoft Excel	1.9798	.84170	*
Charts creation using Microsoft Excel	2.2879	.63482	*
Data input using Microsoft Power Point	2.6288	.56576	*
Power Point Slideshow for group discussion	2.2247	.75473	*
Graphic design application tools using Corel Draw	2.1616	.87678	*
Creation of objects in Corel Draw	2.6263	.66502	*
Computer shut down processes	2.3182	.70784	*
Grand Mean	2.5429	.58736	*

Competency level was set at 1.50 and above, while Incompetence was set at 1.49 and below. Thus, '*' implies competent and '#' implies Incompetent on the decision column.

institutions located in Rivers State, Nigeria. Mean ratings and standard deviation; as well as analysis of variance (ANOVA) and regression analysis were used to answer the research questions.

RESULTS AND DISCUSSIONS

Research Question 1

What is the level of computer operational skills and competencies of undergraduate students in River State, Nigeria?

The grand mean score of 2.54 in Table 1 indicated that undergraduate students covered in this study were competent in using computer technology alongside IT tools, applications and resources in their knowledge formation, generation and acquisition. The finding implies that the undergraduates are active and creative in their quest to learn and acquire knowledge using digitally information systems to generate and gather

relevant information. Finding further implies that the undergraduates in Rivers State, Nigeria are compliant to the digital literacy in line with Mcmillan (1996) and Stein, Craig and Scollary (1997).

Research Question 2

What is the level of competencies in Information Technology applications and tools among undergraduate students in River State, Nigeria?

The grand mean score of 2.64 in Table 2 implies that undergraduate students in Rivers State Nigeria were competent in IT applications and tools alongside accompanying resources with the exception to Log in an e-mail account that recorded the least mean rating ($M=2.22$, $SD=.83$); this may not be unconnected with the fact that many undergraduates hardly operate institutional e-mails or maintain personal active e-mail, but depends on third party such as friends, cyber cafe and business centre operators. However, finding

Table 2. Mean rating on level of competencies in information technology applications and tools

Items	Mean	SD	Decision
Log in an e-mail account	2.2247	.83438	*
Composing e-mail communication	2.5126	.57133	*
Composing e-mail with attachment	2.6439	.57102	*
Sending e-mail with attachment	2.3965	.56648	*
Browsing educational information and resources using Universal Resources locator (URL) e.g. http// www.	2.4470	.62010	*
Browsing Journal articles	2.4798	.61796	*
Downloading e-book on Internet	2.7045	.62121	*
Downloading e-leaning document	2.5758	.71993	*
Downloading pictures on educational concepts	2.7096	.59022	*
Undertaking online application e.g. admission form processing	2.6995	.57200	*
Undertaking online registration e.g. course registration	2.6364	.61136	*
Submission of class assignment via e-mail	2.4419	.60308	*
Submission of class project via e-mail	2.6843	.55926	*
Grand mean	2.6429	.58736	*

Competency level was set at 1.50 and above, while Incompetence was set at 1.49 and below. Thus, '*' implies competent and '#' implies Incompetent on the decision column.

implies that undergraduate students in Rivers State Nigeria satisfied all the competencies highlighted by Mcmillan (1996) as well as Stein, Craig and Scollary (1997) on e-mail communications, Internet browsing and Information Technology.

Research Question 3

What is the extent of undergraduate students' use of computer tools and software applications as part of their personal learning activities?

The grand mean score of 2.32 as indicated in Table 3 implies that undergraduate students in Rivers State, Nigeria regularly engage in the use of computer tools and software applications in the course of their learning activities. Finding, further confirms that undergraduates in Rivers State, Nigeria are not only computer literate in line with Mcmillan (1996) and Stein, Craig and Scollary (1997) but are actively engaging in behavioural indicators that goes with such competencies

Research Question 4

What is the extent of undergraduate students' use of Information Technology tools and resources as part of their personal learning activities?

The grand mean score of 2.32 in Table 4 literarily implies that undergraduate students in Rivers State, Nigeria regularly use IT tools and resources as part of their personal learning activities and thereby technologically-capable. More so, they perceived themselves to possess information literacy skills needed to locate, retrieve, analyze and use information.

Research Question 5

What is the pattern of undergraduate students' computer competence in Information Technology tools and resources in relation to study skills and habits enhancement?

Table 3. Mean rating of extent of undergraduate students' use of computer tools and software applications as part of their personal learning activities

Items	Mean	SD	Decision
Starting a computer	2.1818	.65779	
Launching an application software e.g. Microsoft Word, Excel, etc.	1.9571	.73307	#
Typing document using Microsoft (MS) word,	2.4066	.59437	*
Slide preparation using MS PowerPoint	2.3283	.63110	*
Data input using Microsoft Excel	2.1818	.81894	*
Data analysis using Microsoft Excel	1.9040	.74350	#
Charts creation using Microsoft Excel	2.3258	.65406	*
Data input in power point	2.2121	.81471	*
Power Point slide show for group discussion	2.3696	.78701	*
Graphic design application tools in Corel Draw	2.3813	.68514	*
Creation of objects in Corel Draw	2.4798	.72363	*
Computer shut down processes	2.3384	.78055	*
Grand mean	2.3179	.6991	

High level of use was set at 1.50 and above, while low level of use was set at 1.49 and below. Thus, “*” implies high level of use and “#” implies low level of use on the decision column.

Table 4. Mean rating on extent of undergraduate students' use of IT tools and resources as part of their personal learning activities

Items	Mean	SD	Decision
Log in on e-mail account	2.3081	.60883	*
Composing and sending e-mail communications	2.5480	.67896	*
Composing e-mail with attachment	2.3283	.67378	*
Sending e-mail with attachment	2.4192	.63328	*
Browsing for educational information and resources using Universal Resources Locator (URL)	2.3409	.74115	*
Browsing for Journal articles	2.3561	.62601	*
Downloading e-book on Internet	2.3586	.66576	*
Downloading e-learning document	2.4672	.67613	*
Downloading pictures on educational concepts	2.4268	.69150	*
Undertaking online application e.g. admission	2.4394	.70719	*
Undertaking program registration e.g. course registration	2.3333	.77296	*
Submission of class assignment via e-mail	2.3763	.65784	*
Submission of class project via e-mail	2.1793	.71912	*
Grand mean	2.3179	.6991	

High level of use was set at 1.50 and above, while low level of use was set at 1.49 and below. Thus, “*” implies high level of use and “#” implies low level of use on the decision column.

Table 5. Summary of the regression analysis on the contribution of undergraduate students' computer and it competencies to study skills and habits enhancement

Descriptive Statistics			
Variable	Mean	Std. Deviation	N
Study Habits and Skills	83.7854	9.28872	396
Competencies	82.5859	7.61047	396

Table 6. Model summary and ANOVA

Model		Sum of Squares	Df	Mean Square	F	Sig.
1	Regression	4378.691	1	4378.691	58.084	.000
	Residual	29702.064	394	75.386		
	Total	34080.755	395			

R 0.358

R² 0.128

Adj. R² 0.126

SE 8.68251

Table 7. Relative contribution of predictor on criterion variable

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	47.655	4.761		10.010	.000
	Competence	.437	.057	.358	7.621	.000

Table 5 and Table 6 present the patterns of use of Information Technology tools and resources with the observed variance in students' study skills and habits which accounted for approximately 12.6% of the variance of student study skills and habits ($R^2 = .128$, Adjusted $R^2 = .126$). The result on ANOVA further indicated that pattern of use of IT tools and resources was a significant predictor of students' study skills and habits enhancement ($F_{(1, 394)} = 58.084$, $p = .000$). Table 7 reveals that the Beta values for pattern of use of Information Technology tools and resources ($\beta = .358$; $t = 7.621$, $p = .000$) was significant at .05 alpha level in terms of predicting student study skills and habits.

Findings establish the fact that undergraduate students in Rivers State, Nigeria are perceived to develop skills needed to locate, retrieve, analyze

and use information which partially demonstrates indexes of the College of Dupage Library (CoDL, 2002) definition which stated that information literacy is directly related to the ability to gather and assess relevant information as well as problem solving abilities. The pattern of use of computers and IT tools and resources which accounted for approximately 12.6% of the variance of undergraduate students' study skills and habits was a significant predictor that also agrees with the assertion of King (2007) on multi-dimensional attributes of information literacy, which employ traditional and modern information technology skills to retrieve, manage and present information.

Findings from this study further reveal that undergraduate student' competence and engagement in computer and Internet activities significantly contributes or enhances their

personal and study habits and skills. The finding partly agrees with the submission of Bodys (2005) on the ICT tools and resources allowed students to be more individually active and become more independent in making decisions about their learning.

FUTURE RESEARCH DIRECTIONS

Future research directions could be tailored towards ascertaining various information literacy skills among tertiary institution students in different parts of Nigeria which could be a large scale assessment using case studies alongside in-depth observational techniques.

CONCLUSION

In conclusion this study has revealed that there is no doubt that undergraduate students perceived themselves to possess competencies and skills needed to locate, retrieve, analyze and use information and that that information literacy is directly related to the ability to gather and assess relevant information using information technology tools. The pattern of use of computers and IT tools and resources to retrieve, manage and present information shows that higher institution students in Rivers State, Nigeria are technology-capable which invariably was perceived to enhance their study skills and habits

RECOMMENDATIONS

The study recommends that:

- Undergraduate students should take advantage of the opportunities provided by the Internet computing and Information and Technology for personal study and academic activities in preparation for the real world-of-work.

- University teachers and faculties in higher institutions of learning should integrate into their teachings such activities that compel students to engage in computer and Internet use as part of classroom interactions.
- Undergraduate students should sharpen their skills Internet computing so as to become more familiar with computer platforms, open source application, IT tools and OERs that could assist them their academic life both for their studies and research.

REFERENCES

- Adeoye, B. F. (2013). Recent Developments in Technology Utilization in Teaching and Learning. In U. M. O. Ivowi (Ed.), *Seeking Total Quality Education in Schools* (pp. 273-291). Lagos: CIBN Press Limited.
- Adomi, E. E., & Kpangban, E. (2010). *Application of ICTs in Nigeria Secondary Schools: Library Philosophy and Practice*. Digital commons@ University of Nebraska-London. Retrieved June 15, 2015, from: <http://digitalcommons.unl.edu/libphilprac/345>
- Ajuwon, G. A. (2003). Computer and Internet use by first year clinical and nursing students in a Nigerian teaching Hospital, British Medical Companion. *Medical Information and Decision Making*, 3(1), 10–15. doi:10.1186/1472-6947-3-10
- Andretta, S. (2005). *Information Literacy: A Practitioner's guide*. Oxford, UK: Candos Publishing. doi:10.1533/9781780630755
- Bello, I. S., Arogundade, F. A., Sanusi, A. A., Ezoma, I. T., Abioye-Kuteyi, E. A., & Akinsola, A. (2004). Knowledge and utilization of information technology among healthcare professionals and students in Ile-Ife, Nigeria: A case study of a University teaching hospital. *Journal of Medical Internet Research*, 6(4), e45. doi:10.2196/jmir.6.4.e45 PMID:15631969

- Body, W. (2005). *The Integration of Information and Communication Technology in Scottish School*. An Interim Report by HM Inspectors of Education. Retrieved June 15, 2014 from: http://en.wikibooks.org/wiki/ICT_in_Education/introduction.htm
- College of Dupage Library. (2002). *Student Learning Outcomes for Information Literacy Instruction Programme*. Retrieved June 15, 2015, from: <http://www.cod.edu/library/services/faculty/infolit.learning>
- David, V. (2006). *Concepts of Information Technology*. Cape Town: ICDL Foundation Kenilworth.
- Davis, M. R. (2010). E-Learning seeks a custom fit. *Education Week Digital Directions*, 3(2), 18–19.
- Ezekoka, G. K. (2015). *Maximizing the effects of Collaborative Learning through ICT*. International Educational Technology Conference (IETC). Retrieved online 2/6/2016 at: <http://www.sciencedirect.com/science/article>
- Ezekoka, G. K., & Nwosu, S. N. (2010). Identification of Information and Communication Technology Skills of Teachers in Imo State Secondary Schools. *Journal of Studies in Education*, 1(1), 244–250.
- Federal Republic of Nigeria. (2004). *National Policy on Education Lagos (5th ed.)*. Lagos: NERDC Press.
- Jagboro, K. O. (2003). A study of Internet usage in Nigerian universities: A case study of Obafemi Awolowo University, Ile-Ife, Nigeria. *First Monday: Peer Reviewed Journal on the Internet*. Retrieved online at: <http://www.firstMonday.dk/issues/issues8=2/Jagboro>
- King, L. (2007). *Information Literacy of Incoming Undergraduate Arts Students at the University of Western Cape: Assessment of competencies and proficiencies* (Unpublished Doctoral Thesis). The University of Western Cape.
- Mcmillan, S. (1996). Literacy and computer literacy: Definitions and comparisons. *Computers & Education*, 27(3-4), 161–170. doi:10.1016/S0360-1315(96)00026-7
- Mtebe, J. S., Dachi, H., & Raphael, C. (2011). Integrating ICT into teaching and learning at the University of Dar es Salaam. *Distance Education for Empowerment and Development*, 32(3), 289-294.
- Nagunwa, T. P., & Lwoga, E. T. (2012). Developing an e-Learning strategy to implement medical competency-based curricular: Experience from Muhimbili University of health and Allied Sciences. *International Conference on Education and e-Learning Innovations (ICEELI)*. Retrieved online 2/6/2016 at: http://www.researchgate.net/publication.261524397_development_of_eLearning
- Odusanya, O. O., & Bamgbala, O. A. (2002). Computing and information technology skills of final year medical and dental students at the college of medicine, University of Lagos, Nigeria. *Postgraduate Medical Journal*, 9(4), 189–193. PMID:12690677
- Otunla, A. O. (2013). Operational Processes for Computer-Mediated Learning in Management of Large Surgical Theatres, Laboratories, and Teaching Sessions in Medical and Health Sciences. In U. M. O. Ivowi (Ed.), *Seeking Total Quality Education in Nigeria* (pp. 313-327). Lagos: CIBN Press Limited.
- Otunla, A. O., & Akinyemi, J. O. (2014). Technology Integration in the Classroom: Report of an Asynchronous Online Discussion among a Group of Nigerian Graduate Students. In B. F. Adeoye & L. Tomei (Eds.), *Effects of Information Capitalism and Globalization on Teaching and Learning*. IGI Global. Retrieved online 22/02/2015 at: <http://www.igi-global.com>

Otunla, A. O., & Baiyelo, T. D. (2012). A study on the competency and utilization patterns of ICT Tools Among Medical Undergraduates Students in Three Nigerian Universities. *Journal of Educational Media and Technology*, 17(2), 42–57.

Stein, A., Craig, A., & Scollary, A. (1997). Preparatory IT practices and skills of transition in business students. *Australasian Journal of Educational Technology*, 13(1), 40–53. doi:10.14742/ajet.1918

Uys, P. N., Nleya, P., & Molelu, G. B. (2004). Technological innovations and management strategies for higher education in Africa: Harmonizing reality and idealism. *Educational Media International*, 41(1), 67–80. doi:10.1080/0952398032000105120

ADDITIONAL READING

American Association of School Librarians. (2001) Information Literacy. Position paper Retrieved May 10, 2015, from: <http://www.ala.org/aasl/positions/psinfoit.html>

Andretta, S. (2005). *Information Literacy: A Practitioner's guide*. Oxford: Candos Publishing. doi:10.1533/9781780630755

Californian State University Commission. (2000). Learning resources and Instructional Technology, Work Group on Information Competence. Information Competence in the CSU: A report CSUN, Californian State University. Retrieved July 27, 2014 <http://library.CSUN.edu/susan.curzon/factsheet.html>

Condie, R., & Munro, B. (2007). The impact of ICT in schools: a landscape review. British Educational Communications and Technology Agency (BECTA) Research publications. Retrieved July 20, 2015, from: www.becta.org.uk

KEY TERMS AND DEFINITIONS

Information Literacy: Information Literacy is the ability to locate, evaluate and use effectively the needed information and it is also regarded as the set of skills needed to locate, retrieve, analyze and use information (The College of Dupage Library, 2000).

Information Technology (IT): IT is a broad term which covers all the aspects of the uses of computer technology which includes not only hardware and software, but also communication technology; linking computer systems, software engineering, administration and supporting the infrastructure to manage and deliver information (David, 2006).

The Roles of Digital Literacy in Social Life of Youth

Dragana Martinovic

University of Windsor, Canada

Viktor Freiman

Université de Moncton, Canada

Chrispina S. Lekule

St. Augustine University of Tanzania, Tanzania

Yuqi Yang

University of Windsor, Canada

INTRODUCTION

This chapter contains updated findings related to social aspects of digital activities of youth (Martinovic, Freiman, Lekule, & Yang, 2014). Computers, mobile devices, and the Internet are increasingly used in everyday social practices of youth. Recent statistics reveal that in September 2015, there were 1.01 billion daily active users and on average 894 million mobile daily active users of Facebook (Facebook, 2015). To be successful in school, work, and in socializing, youth need to competently use digital tools and define, access, understand, create, and communicate digital information. Being able to evaluate digital information, develop perceptions of, and respect for, social norms and values for functioning in the digital world, without compromising one's own privacy, safety, or integrity is also important. The competencies and skills that new generations require to be successful in the digital era are largely still not being taught in schools. Results of this chapter will provide the following:

- Address the social prospects of Information and Communication Technology (ICT) use among youth;

- Describe the online behavior of young people through the paradoxical nature of the Internet that provides opportunities for social development but introduces risks;
- Inform educators and youth services about which factors to consider in designing flexible, innovative, and inclusive programs for young people to enable them to successfully function in the era of the Internet, new media, and computer technologies.

BACKGROUND

In the past 15 years or so, ICTs became increasingly accessible in most countries. The ICTs like personal computers, cell phones, and the Internet can be used for both in-school and out-of-school activities, and are particularly suitable for connecting individuals and communities globally (Beetham, McGill, & Littlejohn, 2009). Using these tools appropriately so that one can live, learn, and work in a digital society, is broadly defined as being digitally literate (Beetham, 2010). However, by and large, these competencies and skills are not being taught in schools (Martinovic, Freiman, & Karadag, 2011). For

example, Jenkins (2006) finds that youth are not taught how to participate in social online practices (e.g., in information sharing and collaboration) despite dangers for unskilled users, whereas some authors (Martinovic & Magliaro, 2007; Noveck, 2000) emphasize importance of understanding a paradoxical nature of the Internet, where one can be confronted with limitless information, while obtaining less knowledge; where access is relatively cheap, but the environment is increasingly commercialized; and where communities do form, but atomization prevails.

Livingstone (2008) describes the dichotomy of optimistic and pessimistic opinions coming from academics and media on how ICTs affect young people:

- Optimists emphasize the new opportunities for self-expression, sociability, community engagement, creativity, and new literacies. They envision change in social dynamics, with youth involvement in the co-creation of innovative and counter-consumerist cultures both locally and globally. Public policy makers and educators see opportunities for engagement in collaborative learning and various online government services. Xie (2014), for example found that mobile communication through social networking sites, amplifies social capital of teenagers.
- Pessimists associate the behavior of youth on social networking sites with loss of privacy and lack of shame. They look at social networking as time-wasting and socially isolating activities that may have far-reaching negative effects on the safety of youth. Others fear that youth growing up in the digital age may be barren of understanding emotional nuances and reading social cues, and may lack empathy (Stout, 2010). They note that cyberbullying results in changes in behavior and deep emotional problems among its victims (Mitchell, Ybarra, & Finkelhor, 2007; Ybarra, Diener-West, & Leaf, 2007), and that the effects of Internet

communication both with peers and with strangers on well-being may be particularly adverse for lonely adolescents (Valkenburg & Peter, 2007). Patton et al. (2014) warns against increase of youth violence in on-line space, which includes cyber-bullying, -gangs, -stalking, and -suicide. Moreover, young people recognize that using the Internet for schoolwork may encourage taking shortcuts, cheating, laziness, low school morale, and may hinder development of study skills (Ben-David Kolikant, 2010).

Based on these examples, a digitally literate person should be able to navigate between opportunities and traps created by introducing different ICTs in one's everyday social life.

SOCIAL ONLINE PRACTICES OF YOUTH

Issues, Controversies, Problems

Current international statistics reveal that about 87.9% of North Americans, together with 73.5% Europeans use the Internet (Internet World Stats, 2014). Extreme growth in Internet use in Africa, Middle East, and Latin America in the last 15 years is measured in thousands of percent. Young Canadians seem to follow such common worldwide trends. One Canadian study, Young Canadians in a Wired World (Media Awareness Network, 2001-2012) looked at the online behavior, attitudes, and opinions of more than 5,200 children and youth from grades 4 to 11 in French- and English-language schools in every province and territory. The report revealed that in 2005, young Canadians were almost universally connected to the Internet through the household computer, personal computer, and/or cell phone, and were active users of the ICTs. While the participants in this study described their online experiences as generally positive and socially rewarding, they also reported being exposed to inappropri-

ate content and risky situations such as bullying and sexual harassment which may compromise their integrity and privacy. For this reason the participants highlighted the need to learn how to distinguish credible from non-credible information on the Internet.

In the further text we address the issues of formation of identity, friendship, participatory culture, and political engagement in the context of the ICT use. These areas of social development of young people are described in the literature as dependent on their age.

Identity Formation

Pre-School and Primary School Children

At the very young age, children discover virtual world when playing. Aguiar and Taylor (2015) investigated how preschool children would differentiate the social affordances of a virtual character that simulates social behaviors and those of a stuffed animal they used to play with. It was found that the children relate to the stuffed dog as a friend, while the virtual dog, even when sophisticatedly programmed appear to be entertaining rather than the relationship partner.

Between ages 7-12 which we define as the stage of emerging digital literacy, digital life experiences of children are highly dependent on what ICT are available at home and in school. At the same time, access and use are restricted by parents, educators, laws, etc.

Use of Internet by 39 fourth-graders (ages 9-11) from a small New Hampshire town was studied by Henke and Fontenot (2007). The researchers found that the children were already veteran Internet users (i.e., having at least two years of experience) and that they were able to identify the persuasive intent of commercial Websites, and to distinguish between the informational and entertainment functions of non-profit and government Websites. The children were unwilling to substitute other social activities for Internet use, which was influenced by teachers, parents, and older siblings.

Adolescents

Issues of identity and privacy become even more important in the increasing socialization of youth aged 12-18. Hundley and Shyles (2010) conducted focus-group interviews with 80 West Coast and East Coast U.S. teenagers (ages 12-16) to ascertain their perceptions and awareness of digital technology. The authors learned that the Internet, and particularly social networking sites, played an important social role in the young people's lives: Cell phones helped them to connect to one another; computer games, to have fun; and social networking sites, to stay in touch and connected. Some geographical and cultural patterns emerged: While Facebook was almost unknown in the West (mostly Hispanic) and popular in the East (mostly white), MySpace was the overwhelming favorite among all participants (with West Coast youth more active). The number of online friends varied from 6 to 1000, averaging 85-200. Favorite activities on social networking sites were "talking/meeting with friends," "staying in touch," "updating profile," "checking other profiles," "checking messages," "texting friends," and "multitasking" (e.g., simultaneously using a computer, listening to the music, and watching TV).

For adolescents, electronic media were powerful socializing agents, but they did not completely replace other sources for social interaction. Hundley and Shyles's (2010) participants liked to combine face-to-face and online socializing; their online social groups mirrored their real-life communities. The participants were aware of online risk and saw the danger of cyber-attacks on their privacy, but were ready to accept the risk as they were careful not to disclose any personal information. However, some participants acknowledged they were occasionally dishonest when providing information on Websites (e.g., hiding their age), but did not see a problem with this.

Valkenburg, Schouten, and Peter (2005) surveyed 600 adolescents and found that the Internet plays an important role in their identity exploration and creation. The authors identified different self-

presentational strategies used by these youth. For instance, younger adolescents, girls, and extroverts significantly more often pretended to be older than did the older adolescents, boys, and introverts. Boys and introverts presented themselves more often as a macho person, whereas girls, younger adolescents, and extroverts presented themselves more often as a more beautiful person. Finally, boys presented themselves more often than girls as a real-life acquaintance or as a fantasy character.

Based on the focus group interviews with 127 Italian young Internet users (i.e., middle and high schools students of age 11–20), Bosca et al. (2015) concluded that adolescents do not speak favourably about people who on the Internet post their altered photos or lie about their appearance.

Harman, Hansen, Cochran, and Lindsey (2005) argued that misuse of the Internet can affect various aspects of children's social lives. Harman et al. conducted a study of 187 predominantly European or American sixth-, seventh-, and eighth-grade students, aged 11–16. The children who reported the most deceitful behavior on the Internet (e.g., pretending to be older) had poorer social skills, lower levels of self-esteem, higher levels of social anxiety, and higher levels of aggression (i.e., inappropriate assertiveness), regardless of gender. Frequent use of the Internet, however, did not affect the children's social skills, self-esteem, social anxiety, and aggression.

However, Livingstone (2008) noted a difference between how younger and older teenagers created their online profiles. Younger teens had a playful approach and created profiles that were decorative, stylish, and elaborate. Older teens, on the other hand, created plainer profiles that emphasized their relationships with others (e.g., links to their friends' profiles and group photos). The danger for this latter group lay in their lack of knowledge of when or how to draw a line between what to say in private and in public medium. In addition, Livingstone found that how the teenagers classified their friends (e.g., friends from school, friends from holidays, real friends, old friends, best friends, good friends, 'hi-and-bye' friends)

and the corresponding granulation of closeness with them were not adequately matched by the privacy features of the social networking sites (i.e., 'public' or 'private'). The author suggested that teenagers struggle with online privacy because of this mismatch between their needs and the features of the sites. Moreover, inadequate, confusing, or poorly designed site settings, along with the teens' limited Internet literacy, often left the teens unclear about their control over who could see what about them.

Nonetheless, as shown by both anecdotal and research evidence, youth are very protective of their digital spaces and they make a sharp distinction between situations in which they participate in formal learning and those in which they socialize and learn informally. For example, Marwick, Murgia-Diaz, and Palfrey (2010) argued that young people may not perceive the Internet as a public space. In many ways, "online spaces like MySpace and Facebook are seen as private social spaces where young people can engage in personal talk, gossip, 'hanging out,' flirting, sharing secrets, and all the other social practices that they engage in with their peers offline" (p. 61). From this perspective, youth have redefined the concepts of 'private' and 'public'; they want to keep some digital spaces private from the watching eyes of adults (be they parents, teachers, or marketers) so that they can freely socialize there under their own terms. This stand for also confirmed in Bosca et al.'s (2015) study, although the Italian adolescents also mentioned using Internet to connect to parents. The authors suggested that family relationships may benefit from the Internet communication between parents and children, as long as the parents recognize adolescents' need for autonomous use of online spaces and let them demonstrate their digital skills, which can also improve the status of their children in the family.

Recent study of online behavior of 11–15 year old youth conducted by Afshar et al. (2015) expresses concerns that when shifting to the online world, adolescents can develop social phobia which drives them to having a hidden identity,

while potentially damaging their natural and real identity.

Internet use by adolescents with developmental and social differences was explored by Bannon, McGlynn, McKenzie, and Quayle (2015). The authors conducted focus groups with 36 young people (aged 13–18) who were all identified with Autistic Spectrum Disorder, moderate learning difficulty, and/or social, emotional and behavioural difficulties. Bannon et al. identified aspects of development of identity, social connectedness, and competence (e.g., in specific skills as well as a sense of self-competence) through their participants' use of the Internet. These individuals not only developed in these three areas, but purposefully used the Internet to build social ties to others and avoid sense of anxiety of offline communication. While this segment of adolescent population used the Internet in a similar way and with similar benefits as youth in general, they may need some extra measures to be put to ensure their online safety.

Young Adults

Mahiri (2004) described how youth nowadays use electronically mediated popular culture to produce, consume, and propagate personal/cultural meanings, individual pleasures, and desires that inform, challenge, and often counter well-established societal norms and forms. One example can be seen in Seibel-Trainor (2004), who described a case of a college-educated, single young woman, a member and a regular contributor to The Gossamer Project, who described herself as a “media junkie” (p. 127) who spent 2-5 hours a day reading and writing for The Gossamer Project. This woman, Barbara, followed the events and characters from the series, and as other fans on the Website did, wrote about them extensively. Barbara produced character studies, wrote commentaries on each episode, and predicted what would happen in the next episode. She posted her stories about main characters from the series and received her own fan mail. This was all done under the scrutiny of the television network, which tried to block

the writing of fans like Barbara. However, this attempt of the TV network generated strong opposition from the fans, who organized a maze of connections from their personal Web pages to the Gossamer Project Website to protect Gossamer Project contributors from intellectual property lawsuits. Seibel-Trainor concluded that the Internet removes the distinction between the writer and the reader; writers like Barbara write for their cyber-friends and other fans; they do not accept simply being consumers of media.

West, Lewis, and Currie (2009) interviewed 16 undergraduate students, 21-26 years old, who reported having on average 82 real-life and 200 Facebook friends. All participants had more friends on Facebook than in real life. Facebook friends consisted of real-life friends, cousins, and friends of friends, all of whom were of similar age to the respondents. The students generally did not like the idea of having parents as friends on Facebook, and some were even happy that their parents did not know how to use the Internet, thus allowing them to keep the details of their social life private from their family. The authors suggested that youth who participate in social networking sites have a different notion of what is private and what is public, a distinction that cannot be described in a clear-cut binary way.

In summarizing research on new media spaces, Zemmels (2012) took a stance that nowadays, “identity is largely constructed through media engagement” (p. 17). From that perspective, the construction of identity happens through the active use of media where young people do not only consume, gather, and distribute media content, but add their personal voice in producing media. This creates ‘new social operating system’ where “networked individuals” are constituted within “networked publics,” and where ICT increasingly shape social practices (Zemmels, 2012, p.17).

Redefinition of Friendship

Social networking sites provide increasingly more attractive functions to their users. A well-

developed Facebook or MySpace profile provides a look into the owner's life in a way that the person wants to be perceived by his or her peers (Maranto & Barton, 2010). Users normally search for those who share their labels and/or interests, and request to add them as 'friends.' Through this snowball effect, an individual can quickly become part of a network of many friends.

Intensive engagement of youth in electronic communication, such as texting, instant messaging, and Internet chats, influences the development and expression of their emotions. Some authors (see Stout, 2010), caution that due to the declining of face-to-face communication within the electronic age, the younger generation may lack exposure to real-life experiences that help them develop empathy, understand emotional nuances, and read social cues like facial expressions and body language. As a result, for today's teenagers and preteens, friendship is apparently realized through the exchange of cell phone texting and instant messaging, or on Facebook walls and MySpace bulletins. There are also some noted gender differences in use of the social networking sites—girls use such sites to reinforce pre-existing friendships, whereas boys use them to flirt and make new friends (Lenhart & Madden, 2007); technology is changing the nature of young people's friendships.

However, after Mesch (2009) conducted a survey with around 1,000 adolescents in Israel, he concluded that the youth seem to maintain two separate networks of online and off-line relationships. The type of communication channel adolescents used mostly depended on the strength and origin of social relationship between them. Those who first met face-to-face, who lived in each other's vicinity, and who felt that their relationship is strong and trustworthy, preferred to communicate directly, face-to-face and over the phone.

Participatory Culture

The Internet has shattered the traditional hierarchical boundaries between producers and consumers

of information. Nowadays, millions of youth are both reading and creating material posted on Websites related to topics of their interest. For Jenkins (2006), this type of engagement presents an example of a participatory culture where young people become actively involved in constructing their own virtual spaces from which they can critically examine the real space of adults. Social networks' participants "are also becoming active producers of new media and distributing them in global networked publics" (Zemmel, 2012, p. 17).

In an attempt to find overlaps/connections between online and offline youth cultures, Wilson and Atkinson (2005) investigated how the Internet has been integrated into the everyday subculture of Canadian youth. After analyzing two cases, the 'Rave' and the 'Straightedge', the authors concluded that the Internet is used as a device for bringing people together in real time and as an enabler of subcultural resistance. For example, music has a particular strong bonding role in youth communities; it is useful in providing group identity and in promoting a distinct lifestyle. The availability of various media formats, such as MP3 or MPEG audio files placed on Websites for free download play a great role in this matter. Various social groups, formally labeled by the mainstream culture as misfits, succeed in winning a cultural space, albeit a virtual one. Wilson and Atkinson concluded that youth culture in the age of the Internet could be viewed not only as "more fragmented, diffuse, and neo-tribal than traditional subcultures...but also as more cohesive in the sense that virtual connections can enhance local relationships" (p. 305).

Political Engagement

Through social networking Websites (e.g., Facebook, MySpace), youth are emerging as important factors in political and social change. In their study of 1875 middle-school students from Hong Kong, Seoul, Singapore, Taipei, and Tokyo, Lin, Cheong, Kim, and Jung (2010) investigated youth engagement with social issues. The researchers

found that about half of the teens read news on the Internet, visited charity, environment, or government Websites, voted to reflect their opinions, and/or signed a petition online (one in five youth). While teens preferred to use these virtual spaces to discuss entertainment, sports, or shopping, social networking sites also served as “potential mobilizing forces” (p. 852). For youth who are already interested in politics and are civic-minded, the Internet may function as an amplifier of engagement, albeit in “unconventional ways” (p. 843).

However, Lin et al. (2010) also referred to other researchers (e.g., Livingstone, Bober, & Helsper, 2004) who noted that the U.K. teens they studied had little interest in political engagement on the Internet because they thought it was not very likely that politicians would ever listen to them. Yet Livingstone et al. (2004), too, found that 54% of the teens had visited either charity, government, environment, and human rights sites or Websites for improving conditions at school or work.

Nowadays, youth are involved in international groups where they discuss and organize events related to environmental (e.g., the Earth Hour) and political (e.g., Relay For Life) causes. In other words, youth deploy the technology to spread their political and social views, promote environmental awareness, and raise funds for their favorite causes or charities (Maranto & Barton, 2010). Still, as Wilson and Atkinson (2005) pointed out, violent confrontations between individuals and groups at public spectacles also get arranged in advance through the Internet. Also, students who are influenced by the existence of social peer in-groups in their schools (e.g., geeks, jocks, fashionistas) may take the same exclusionary attitude on the Internet and resist connecting with those who belong to groups other than their own.

SOLUTIONS AND RECOMMENDATIONS

Whatever we say and write about the youth and their habits, connectivity remains the key, which has led some authors to name them the connected

generation (Mason & Rennie, 2006)—that is, the generation that is involved in a wide spectrum of online activities, including buying online. While the positive attributes of connectivity are well recognized, the authors see it also as a trap, especially for youth; instant messaging could be a distraction to learning, extensive use of chats and multiuser discussions could be associated with impaired academic achievement and social isolation, as well as privacy could be lost through providing access to personal data to unintended recipients.

Providing solutions and recommendations is problematic, as digital literacy phenomena are quickly changing, as the ICTs and practices that surround them change. Davies and Cranston (2008) proposed using social networking sites as tools to encourage youth to become more socialized and to learn something informally. They acknowledged, however, that “something” learned may well be undesirable, ranging from relatively benign commercial advertisements to content that does not contribute to healthy social and cognitive development. Thus, young people need supports in developing “appropriate skills and resilience to navigate online social networking risks and opportunities” and developing “shared understandings of safe and reasonable online behaviour patterns” (p. 3). This suggestion extends beyond the simple blocking of social networking sites and instead encourages working with youth as an alternative. In a recent paper, Vodanovich et al. (2015) point at the lack of holistic approach to fully grasp the impact of social networking sites on digital natives’ well-being, in particular their social competence. The authors point at an important role that ubiquitous spaces play in a complex processes of youth becoming independent from their parents, gaining social acceptance from their peers and establishing a coherent identity. Hence, the role of adults is to understand, appreciate and in turn support digital natives in making the all-important transition to adulthood within this new environment.

FUTURE RESEARCH DIRECTIONS

Our work has allowed us to identify issues related to youth's understanding and respecting online social norms and values. But these issues easily become more complex and subtle when we take into consideration meta-social skills (identity formation, friendship, self-esteem, autonomy, etc.). We found that technologies do affect the development of well-rounded citizens of modern society in different ways, and that this development is not easy to capture and assess because no ready-to-use and evidence-based developmental frameworks exist. There is a conceptual and methodological gap between the need for such frameworks and reliable data available to support their development and validation across cultures, age groups, social groups, and individuals.

CONCLUSION

Young people nowadays face challenges that require development of healthy social behavior with regard to online identity formation and protection of privacy. Some positive aspects are seen in the rising of a participatory culture among youth, whereby they use virtual spaces to critically examine the real space of adults. Youth subcultures emerge, and often-marginalized youth find ways to exercise creative expression and shape their own identity. Moreover, the ICTs allow young people to emerge as legitimate social and political actors. Using unconventional ways of engagement, young people surface as political beings and as individuals interested in civic and environmental issues. Some ICT-related addictions (e.g., to computer games, to Facebook, to text messaging) may disappear on their own, while for some problems, interventions may be needed. For example, adolescents may spend less time playing computer games when they start developing an interest in dating and socializing face-to-face. It seems that youth need to be

educated about the importance of maintaining a balance between their real and virtual lives. Organizing outings and clubs where youth play, compete, and socialize may counteract to some extent the allure of the technology that youth are exposed to.

REFERENCES

- Afshar, M., Zeraati, M., & Damaliamiri, M. (2015). Youth and identity crisis in online technology. *European Psychiatry, 30*(1), 1299. doi:10.1016/S0924-9338(15)31017-8
- Aguiar, N. R., & Taylor, M. (2015). Childrens concepts of the social affordances of a virtual dog and a stuffed dog. *Cognitive Development, 34*, 16–27. doi:10.1016/j.cogdev.2014.12.004
- Beetham, H. (2010). *Review and Scoping Study for a cross-JISC Learning and Digital Literacies Programme: Sept 2010*. Bristol: JISC.
- Beetham, H., McGill, L., & Littlejohn, A. (2009). *Thriving in the 21st century: Learning literacies for the digital age (LLiDA project)*. The Caledonian Academy, Glasgow Caledonian University.
- Ben-David Kolikant, Y. (2010). Digital natives, better learners? Students beliefs about how the Internet influenced their ability to learn. *Computers in Human Behavior, 26*(6), 1384–1391. doi:10.1016/j.chb.2010.04.012
- Borca, G., Bina, M., Keller, P. S., Gilbert, L. R., & Begotti, T. (2015). Internet use and developmental tasks: Adolescents point of view. *Computers in Human Behavior, 52*, 49–58. doi:10.1016/j.chb.2015.05.029
- Davies, T., & Cranston, P. (2008). *Youth work and social networking: Final research report*. Retrieved from <http://blog.practicalparticipation.co.uk/wp-content/uploads/2009/08/fullYouth-Work-and-Social-Networking-Final-Report.pdf>
- Facebook. (2015). *Our history*. Retrieved from <http://newsroom.fb.com/company-info/>

- Harman, J. P., Hansen, C. E., Cochran, M. E., & Lindsey, C. R. (2005). Liar, liar: Internet faking but not frequency of use affects social skills, self-esteem, social anxiety, and aggression. *Cyberpsychology & Behavior*, 8(1), 1–6. doi:10.1089/cpb.2005.8.1 PMID:15738687
- Henke, L., & Fontenot, G. (2007). Children and Internet use: Perceptions of advertising, privacy, and functional displacement. *Journal of Business and Economics Research*, 5(11), 59–66.
- Hundley, H. L., & Shyles, L. (2010). US teenagers perceptions and awareness of digital technology: A focus group approach. *New Media & Society*, 12(3), 417–433. doi:10.1177/1461444809342558
- Internet World Stats. (2012). Retrieved from <http://www.internetworldstats.com/stats2.htm>
- Jenkins, H. (2006). *Confronting the challenges of participatory culture: Media education for the 21st century*. The John D. and Catherine T. MacArthur Foundation series on digital media and learning. Retrieved from digitallearning.macfound.org/.../JENKINS_WHITE_PAPER.PDF
- Lenhart, A., & Madden, M. (2007). *Social networking websites and teens*. Pew Research Center. Retrieved from http://www.pewinternet.org/~media/Files/Reports/2007/PIP_SNS_Data_Memo_Jan_2007.pdf.pdf
- Lin, W.-Y., Cheong, P. H., Kim, Y.-C., & Jung, J.-Y. (2010). Becoming citizens: Youths civic uses of new media in five digital cities in East Asia. *Journal of Adolescent Research*, 25(6), 839–857. doi:10.1177/0743558410371125
- Livingstone, S. (2008). Taking risky opportunities in youthful content creation: Teenagers use of social networking sites for intimacy, privacy and self-expression. *New Media & Society*, 10(3), 393–411. doi:10.1177/1461444808089415
- Livingstone, S., Bober, M., & Helsper, E. J. (2004). *Active participation or just more information? Young people's take up of opportunities to act and interact on the Internet*. London: London School of Economics and Political Science.
- Mahiri, J. (2004). New literacies in new century. In J. C. Mahiri (Ed.), *What they don't learn in school: Literacies in the lives of urban youth*. New York, NY: Peter Lang Publishing Inc.
- Maranto, G., & Barton, M. (2010). Paradox and promise: MySpace, Facebook, and the sociopolitics of social networking in the writing classroom. *Computers and Composition*, 27(1), 36–47. doi:10.1016/j.compcom.2009.11.003
- Martinovic, D., Freiman, V., & Karadag, Z. (2011). *Child and Youth Development beyond Age 6—Transitions to Digitally Literate Adults*. Unpublished report, Ministry of Child and Youth Services, Ontario, Canada.
- Martinovic, D., Freiman, V., Lekule, C., & Yang, Y. (2014). Social Aspects of Digital Literacy. In *The Third Edition of Encyclopedia of Information Science and Technology*. IGI Global.
- Martinovic, D., & Magliaro, J. (2007). Computer networks and globalization. *Brock Education Journal*, 16(2), 29–37.
- Marwick, A. E., Murgia-Diaz, D., & Palfrey, J. (2010). *Youth, privacy and reputation: Literature review*. The Berkman Center for Internet & Society Research Publication Series. Retrieved from <http://ssrn.com/abstract=1588163>
- Mason, R., & Rennie, F. (2006). *Elearning: The key concepts. Key Guides*. Oxon, UK: Routledge.
- Media Awareness Network. (2001–2012). *Young Canadians in a Wired World*. Retrieved from <http://mediasmarts.ca/research-policy>
- Mesch, G. S. (2009). Social context and communication channels choice among adolescents. *Computers in Human Behavior*, 25(1), 244–251. doi:10.1016/j.chb.2008.09.007
- Mitchell, K. J., Ybarra, M., & Finkelhor, D. (2007). The relative importance of online victimization in understanding depression, delinquency and substance use. *Child Maltreatment*, 12(4), 314–324. doi:10.1177/1077559507305996 PMID:17954938

- Noveck, B. S. (2000). Paradoxical partners: Electronic communication and electronic democracy. In P. Ferdinand (Ed.), *The Internet Democracy and Democratization* (pp. 18–36). London, UK: Frank Cass. doi:10.1080/13510340008403643
- Patton, D. U., Hong, J. S., Ranney, M., Patel, S., Kelley, C., Eschmann, R., & Washington, T. (2014). Social media as a vector for youth violence: A review of the literature. *Computers in Human Behavior*, 35, 548–553. doi:10.1016/j.chb.2014.02.043
- Seibel-Trainor, J. (2004). Critical cyberliteracy: Reading and writing the x-Files. In J. C. Mahiri (Ed.), *What they don't learn in school: Literacies in the lives of urban youth*. New York, NY: Peter Lang Publishing Inc.
- Stout, H. (2010, Apr. 30). Antisocial networking? *The New York Times, Fashion and Style*. Retrieved from <http://www.nytimes.com/2010/05/02/fashion/02BEST.html>
- Valkenburg, P. M., & Peter, J. (2007). Internet communication and its relation to well-being: Identifying some underlying mechanisms. *Media Psychology*, 9(1), 43–58. doi:10.1080/15213260709336802
- Valkenburg, P. M., Schouten, A. P., & Peter, J. (2005). Adolescents identity experiments on the Internet. *New Media & Society*, 7(3), 383–402. doi:10.1177/1461444805052282
- Vodanovich, S., Shen, K., & Sundaram, D. (2015). Social competence of digital natives: Impact of Social Networking Sites (SNS) use. In *AMCIS 2015 Proceedings* (pp. 1–9). Association for Information Systems.
- West, A., Lewis, J., & Currie, P. (2009). Students Facebook friends: Public and private spheres. *Journal of Youth Studies*, 12(6), 615–627. doi:10.1080/13676260902960752
- Wilson, B., & Atkinson, M. (2005). Rave and Straightedge, the virtual and the real: Exploring online and offline experiences in Canadian youth subcultures. *Youth & Society*, 36(3), 276–311. doi:10.1177/0044118X03260498
- Xie, W. (2014). Social network site use, mobile personal talk and social capital among teenagers. *Computers in Human Behavior*, 41, 228–235. doi:10.1016/j.chb.2014.09.042
- Ybarra, M., Diener-West, M., & Leaf, P. (2007). Examining the overlap in Internet harassment and school bullying: Implications for school intervention. *The Journal of Adolescent Health*, 41(6), 42–50. doi:10.1016/j.jadohealth.2007.09.004 PMID:18047944
- Zemmels, D. R. (2012). Youth and new media: Studying identity and meaning in an evolving media environment. *Communication Research Trends*, 31(4), 4–22. Retrieved from <http://search.proquest.com/docview/1349959575?accountid=14592>

ADDITIONAL READING

- Albirini, A. (2008). The Internet in developing countries: a medium of economic, cultural and political domination. *International Journal of Education and Development using Information and Communication Technology (IJEDICT)*, 4(1), 49–65.
- American Library Association Presidential Committee on Information Literacy. (1989). *Final Report*. Chicago, IL: American Library Association.
- Aufderheide, P. (1992). Media Literacy. A Report of the National Leadership Conference on Media Literacy. Washington, DC: Aspen Institute. Retrieved from http://www.medialit.org/reading_room/article356.html

- Bannon, S., McGlynn, T., McKenzie, K., & Quayle, E. (2015). The positive role of Internet use for young people with additional support needs: Identity and connectedness. *Computers in Human Behavior*, 53, 504–514. doi:10.1016/j.chb.2014.11.099
- Bauer, J., & Kenton, J. (2005). Toward technology integration in the schools: Why it isn't happening. *Journal of Technology and Teacher Education*, 13(4), 519–546.
- Bawden, D. (2008). Origins and concepts of digital literacy. In C. Lankshear, and M. Knobel, M. (Eds.), *Digital Literacies: Concepts, policies and practices* (pp. 17–32). New York, NY: Peter Lang.
- Bunker, B. (2010). *A Summary of international reports, research and case studies of digital literacy: Including implications for New Zealand of adopting a globally-recognised digital literacy standard*. New Zealand: Computer Society Inc.
- Dijst, M. (2004). ICTs and accessibility: An action space perspective on the impact of new information and communication technologies. In M. Beuthe, V. Himanen, A. Reggiani, & L. Zamparini (Eds.), *Transport Developments and Innovations in an Evolving World* (pp. 27–46). Berlin: Springer. doi:10.1007/978-3-540-24827-9_3
- Eugster, P. T., Felber, P. A., Guerraoui, R., & Kermarrec, A.-M. (2003). The many faces of publish/subscribe. *ACM Computing Surveys*, 35(2), 114–131. doi:10.1145/857076.857078
- Huang, C., & Lin, C. (2011). Enhancing classroom interactivity and engagement: CFL Learners perceptions of the application of Web 2.0 technology. *British Journal of Educational Technology*, 42(6), E141–E144. doi:10.1111/j.1467-8535.2011.01219.x
- International Visual Literacy Association. (2009). What is 'Visual Literacy'? Retrieved from <http://www.ivla.org/org/whatvislit.htm>
- Knutsson, O., Blåsjö, M., Hållsten, S., & Karlström, P. (2012). Identifying different registers of digital literacy in virtual learning environments. *The Internet and Higher Education*, 15(4), 237–246. doi:10.1016/j.iheduc.2011.11.002
- Koltay, T. (2011). The media and the literacies: Media literacy, information literacy, digital literacy. *Media Culture & Society*, 33(2), 211–221. doi:10.1177/0163443710393382
- Livingstone, S., & Helsper, E. J. (2007). Gradations in digital inclusion: Children, young people and the digital divide. *New Media & Society*, 9(4), 671–696. doi:10.1177/1461444807080335
- Martin, A. (2005). DigEuLit—a European framework for digital literacy: A progress report. *Journal of eLiteracy*, 2, 130–136.
- Martin, A. (2006). Literacies for the Digital Age. In A. Martin & D. Madigan (Eds.), *Digital literacies for learning* (pp. 3–25). London, UK: Facet.
- Martin, A. (2009). Digital literacy for the Third Age: Sustaining identity in an uncertain world. *eLearning Papers*, 12. Retrieved from www.elearningpapers.eu
- Oliver, M., & Gourlay, L. (2012). Un-defining digital literacies: Students' day-to-day engagements with technologies. EdTech Conference 2012, NUI Maynooth, 31st May 2012. Retrieved from <http://www.slideshare.net/lesleygourlay/oliver-gourlay-keynote-ed-tech-maynooth-1>
- Opgenhaffen, M., & d'Haenens, L. (2012). Heterogeneity within homogeneity: Impact of online skills on the use of online news media and interactive news features. *Communications—the European Journal of Communication Research*, 37(3), 297–316.
- Richtel, M. (2010, November 20). E-Mail Gets an Instant Makeover. *The New York Times*. Retrieved from http://www.nytimes.com/2010/12/21/technology/21email.html?_r=0

Schnellert, G., & Keengwe, J. (2012). Digital technology integration in American public schools. *International Journal of Information and Communication Technology Education*, 8(1), 36–44. doi:10.4018/jicte.2012070105

Simsek, E., & Simsek, A. (2013). New literacies for digital citizenship. *Journal of Educational Technology*, 4(3), 126–137.

Watson, J. A., & Pecchioni, L. L. (2011). Digital natives and digital media in the college classroom: Assignment design and impacts on student learning. *Educational Media International*, 48(4), 307–320. doi:10.1080/09523987.2011.632278

Wilson, E. K., Wright, V. H., Inman, C. T., & Matherson, L. H. (2011). Retooling the social studies classroom for the current generation. *Social Studies*, 102(2), 65–72. doi:10.1080/00377996.2010.484445

Wodzicki, K., Schwammlein, E., & Moskaliuk, J. (2012). Actually, I wanted to learn: Study-related knowledge exchange on social networking sites. *The Internet and Higher Education*, 15(1), 9–14. doi:10.1016/j.iheduc.2011.05.008

KEY TERMS AND DEFINITIONS

Active Users of the ICTs: Individuals, groups and organizations that utilize e-mail, instant messaging, chat rooms, access and download media content, write blogs, and play online games.

Developmental Stages of Digital Literacy in the Social Domain: Stages through which an individual, over time, acquires the digital literacy skills necessary to be considered competent at using digital tools and at defining, accessing, understanding, creating, and communicating digital information.

Digital Literacy (as a Complex Term): One of many constructs that involve the word “literacy”; it includes elements drawn from information literacy, media literacy, and visual literacy.

Digital Literacy (as Ability): Ability to use digital technologies appropriately for learning, working, and functioning in a modern society.

Information and Communication Technologies: ICTs include personal computers, laptops, tablets, PDAs, cell phones, smart phones, computer networks, and the Internet.

Information Literacy: Emphasizes the need for careful retrieval and selection of information available in the workplace, at school, and in all aspects of personal decision-making, especially in the areas of citizenship and health.

Media Literacy: Ability to decode, evaluate, analyze and produce both print and electronic media.

Participatory Culture: A culture in which artistic expression and civic engagement are valued, oriented towards creating and sharing one’s creations.

Social Capital in Online Communication: Social capital developed through online activities and participation in online social networks (e.g., text messaging, chatting, content sharing, organizing group activities, commenting, and tagging).

Social Capital: Quality of group membership. Measured by for example, the frequency of face-to-face interactions with close friends, the number of offline acquaintances, sense of loneliness or lack thereof, community involvement, civic and political participation, and interpersonal trust.

Visual Literacy: Ability to see, discriminate, and interpret the visible natural or artificial objects and symbols in the environment.

Toward a Working Definition of Digital Literacy

Margaret-Mary Sulentic Dowell
Louisiana State University, USA

INTRODUCTION

Digital literacy (Alkali & Amichai-Hamburger, 2004; Bawden, 2008; Buckingham, 2006; Gilster, 1997) is a broad, umbrella term that pertains to the use of literacy skills defined as reading, writing, listening, speaking, composing, communicating, and interacting within digital environments. For example, accessing information and sending information via the internet such as viewing and posting YouTube videos or creating, sending, and receiving e-mails is digital literacy. As well, anime, manga, blogging, fandom blogging, texting, tweeting, designing memes, sharing headcannons, and other forms of creating ideas and communicating perspectives through social media platforms such as Facebook, twitter, Tumblr, and myriad others ways to share thoughts and opinions over the internet or in cyberspace, all qualify as digital literacy (Beach, 2012; Black, 2005; Booth, 2012; Martin & Madigan, 2006; Kist, Tollafield, & Dagistan, 2014; Rodesiler, 2015).

Also referred to as *new literacies* (Coiro, Knobel, Lankshear, & Leu, 2007; 2008; Hagood (2009), Knobel & Lankshear, 2014; Lankshear & Knobel, 2006; Street, 1998), digital literacy implies both the technical ability and emotional skill level needed to generate thought and communicate in multiple formats within digital environments (Elshet-Alkalai, 2004; Landham, 1995). In particular, both the consumption and generation of text and the practices used to create and consume them, formally and informally, both outside and within school, broadly define new literacies. According to Hagood,

New literacies consist of several characteristics: (1) multimodalities, which include linguistic as well as visual, gestural, and auditory texts, (2) situated social practices, which are culturally, linguistically, and textually based, and (3) identities, which connect text users to text uses. (2009, p. 1)

BACKGROUND: A WORKING DEFINITION OF DIGITAL LITERACY

Digital literacy is a complex combination of skill sets defined as the knowledge, technical skills, use, actions, and behaviors that individuals utilize with existing myriad digital and technological devices and resultant forms of communication that have become an integral part of so many people's daily lives. For instance, every day use of a cellphone, smartphone or perhaps a tablet involve digital literacy skill sets. Thus, digital literacy implies the mastery of both the tools and embedded use of technology in personal lives. Digital literacy is also the recognition that digital forms of literacy and the aforementioned skill sets play an essential, critical role in educational and work settings, particularly the reasoned awareness regarding the content that is created and its use in digital literacy formats. Lanham (1995) argues that the notion of being literate has extended from the ability to read and write, speak and listen, to the ability to understand information that is available and accessible in multimodal ways. Being literate in the 21st century requires being skilled at interpreting complex images and discerning the "syntactical subtleties of words" (Lanham, 1995,

p. 161). Whereas historically literacy has included the notion of both composing and comprehending language, defined as the ability to speak, listen, read, and write, literacy has evolved to include a “social practices” approach from scholars within the emerging educational field of New Literacies (Barton, Hamilton, & Ivanic, 2000; Gee, 2001; Street, 1993).

Similar to Lanham, Alkali and Amichai-Hamburger (2004) provide an inclusive, comprehensive definition of digital literacy specifically referring to much more than the ability and skill needed to simply use and navigate a digital device or software, rather, digital literacy requires complex thinking skills as well as critical decision-making ability. The kind of multifaceted thinking required of digital literacy equates to the knowledge beyond how to operate digital devices, access information, or utilize software and refers to the complicated cognitive abilities, requisite motor skills, as well as the sociological and emotional maturity needed to navigate digital environments effectively, usefully, and appropriately. A conceptual model detailed by the authors proposes that digital literacy encompasses five essential digital skills: photo-visual skill – gaining understanding from graphic displays of information; reproduction skill – wherein individuals employ digital reproduction expertise to recreate or craft innovative, significant materials from pre-existing materials; branching skills – defined as building and fashioning knowledge from non-linear, hypertextual navigation); information skills – both assessing and gauging the quality and legitimacy of information), and socio-emotional skills – the mindfulness of the tacit rules that exist and are in place in cyberspace and being able to apply this awareness in online cyberspace communications (Alkali & Amichai-Hamburger, 2004).

Expansion of Digital Literacy

Digital literacy has significant implications for society as a whole and for learning and teaching specifically as new forms of communicating

become the norm (Davies & Merchant, 2009; Gilsten, 1997; Hunter & Caraway, 2014; Pool, 1997). As a practical illustration of what was once a new form of communication becoming the norm, twenty years ago, most households in the United States (US) had landlines. Currently, many US homeowners have discontinued a landline and exclusively use a cell phone. In fact, the various platforms and constructs of social media for the current generation of students equates to most students using at least one form of social media as part of their daily lives (Jenkins, 2006; Lapp, Fisher, Frey, & Gonzalez, 2014; New Media Consortium, 2007). These students will matriculate to a career and the work force, taking with them the expectations of using digital literacies.

As an example of the burgeoning use of digital literacy, Rideout, Foehr and Roberts, assert that, “Eight to eighteen-year-olds spend more time with media than in any other activity besides (maybe) sleeping—an average of more than 7½ hours a day, seven days a week.” (2010, p. 1). Rideout, Foehr and Roberts surveyed a nationally representative sample of 2,002 3rd–12th grade students, ages 8–18 in the US, which contained a subsample of 702 respondents who voluntarily completed week-long (seven-day) media use diaries. Conducted from October 20, 2008, through May 7, 2009, they report that 8 to 18 year olds were spending approximately 6.5 hours with media (cell phones, internet, television, gaming, music) but because these students were adept at multitasking, they were engaged in closer to 8.5 hours of media content daily. According to the authors, based on a survey sample from 2005 to 2010, young people between the ages of 8–18 increased their media consumption by an hour and 17 minutes to 7.5 hours daily, the average work day for most adults. Considering that adults tend to work 40 hour weeks within a five day span, the consumption detailed in Rideout, Foehr & Roberts’ work, represents an incredible amount of time spent with digital literacy. Clearly, the research of Black, (2008), Jenkins (2006), Lapp, Fisher, Frey, and Gonzalez (2014), the New Media

Consortium (2007), Rideout, Foehr & Roberts (2010), and Squire (2008) highlight the pervasiveness of digital literacy practices and depth of participation within a specific segment of US population. Researchers in the United Kingdom (Martin & Madigan, 2006; Street, 1984), Canada (Gilster, 1997), the Netherlands (Kuiper & Volman, 2008), and Australia (Lankshear & Knobel, 2006; 2008; 2012; Ng, 2006; 2010; Thibaut, 2015) also investigate the meaning, use, and prevalence of digital literacy, and postulate that expanding usage will occur and rapid developments will be accessed by younger and younger consumers of digital literacy. The implications for society at large are significant.

DIGITAL LITERACY WITHIN A SOCIO-CULTURAL FRAMEWORK

As an overarching framework, Russian theorist, Vygotsky's (1978) sociocultural theory of learning and the social nature of learning coupled with the mediated aspects of human learning situate the discussion of digital learning. Vygotsky's view of learning is rooted in the social interactions between people and resultant learning that may occur, referred to as the zone of proximal development. French theorist Baudrillard (1990), provides a compatible framework based on his notion of media and culture in the ways that technological progress affects social change. Both Vygotsky and Baudrillard focus on social interactions, providing a link to the powerful ways digital literacy is situated in socio-cultural environs.

Similarly, the work of Beach (2012) and Hagood (2000) informs the perspective that learning is situated, occurs within a socio-cultural context that is meaningful to the learner(s), and not confined to classroom walls and school. Each advocate for literacy that is social, collaborative, and relational, requisites of digital literacy.

Drawing upon the work of Project Tomorrow (2012), Beach addresses the outside versus inside school contexts of leaning, supporting learning

as social activity and occurring in environments that are meaningful and constructed by learners. Beach confronts a "false binary between uses of digital tools for school and non-school social purposes, adolescents are increasingly adopting tools to support their learning." (2012, p. 46). According to Beach, real differences between in and out of school literacy points to students engaging in reading and writing within schools to fulfill graded assignments versus reading and writing outside of school as opportunities for both self-expression and identity construction as well as a social communication activity and seeking specific information that is personally meaningful. The social aspect of students' literacy outside of school is rooted in socially-constructed communication, and essentially addresses primary constructs of writing – audience and purpose – first established by Graves (1983). In essence, Beach confronts the notion of informal learning, learning marked by where it occurs, challenging the less than perception of where learning and literacy occurs and for what purposes.

Hagood (2000) argues that the literacy practices that occur outside of formal schooling are equally important as those that are practiced within formal schooling. Perceiving an evolving literacy landscape that encompasses digital literacy forms and formats, Hagood explicates how literacy is at a crossroad, acknowledging traditional views of literacy that are socio-cultural in nature but maintaining that new literacies are evolving into deep social practices that emphasize social and cultural relationships.

In many educational contexts, social media is a viable means to engage students in their learning creating possibilities to collaborate and offering opportunities to connect with peers; the digital literacies afforded by social media also provides ample opportunity for informal learning (Beach, 2012; 2012; Jenkins, Clinton, Purushotma, Robnson & Weigel, 2007; Junco, Heiberger & Loken, 2011; Kilinc, Evans, & Korkmaz, 2012; Lanksher, 1987; Street, 1984; Thibaut, 2015). In fact, there is a long tradition of literacy practices

wherein communicating effectively with others in meaningful ways is situated outside schools and traditional modes of learning (Heath, 1983).

NEW LITERACIES, NEW ISSUES: THE DIGITAL LITERACY IMPACT

Situated within the field of education, three prominent digital literacy issues have surfaced. Those issues are:

1. The dissonance between digital natives and digital immigrants,
2. Some forms of digital literacy enjoy acceptance and legitimacy, and
3. Access to digital literacy formats.

Digital Native or a Digital Immigrant?

Digital natives or the E-generation (Jones & Flannigan, 2006) and digital immigrants, oftentimes identified as baby-boomers and the generations of adults who grew up in a world of print media (Jones & Flannigan, 2006), are defined broadly as those who have acquired digital literacy skill as children versus those who have acquired digital literacy skill in adulthood (Anderson, 2002; Bennett, Maton & Kervin; 2008; Jones & Flannigan, 2006; Prensky, 2001). This is a generational issue regarding when within a lifetime literacy is acquired and is noteworthy.

Prensky (2001), credited with coining the terms digital native – digital immigrant, posits that education must keep pace with the rapid evolution and expansion of digital literacy in order to adequately prepare children for careers and a workforce that will expect sophisticated forms of digital literacy knowledge and skill. Prensky illustrates the dichotomy of traditional teaching practice with the demands of digital literacy which will continue to evolve and expand, calling for a reassessment of literacy teaching and learning.

Generational shifts in literacy acquisition, increasingly sophisticated information processing, and rapidly evolving forms of digital literacy

uniquely impacts teacher preparation. In essence, 21st century teacher preparation programs and teacher educators face a conundrum; these preparation programs are preparing teacher education candidates for an unknown future. And these future teachers are charged with preparing children for unknown digitally literate futures and career requirements.

What Counts as Literacy?

Certain forms of digital literacy enjoy legitimacy and are both expected and accepted within school environments while other forms are not deemed suitable yet appear more prevalent outside of school (Alverman & Hagood, 2000; Beach, 2012; Black, 2005; Black 2008; Jenkins 1992; Kilinc, Evans & Korkmaz, 2012; Mills & Chandra, 2011, Squire, 2008). In the ultra-traditional model of schooling, typified by nine-month agrarian calendars, separated subjects, and the factory model of attendance wherein the most children appear to have in common and how they are grouped is a birthdate, computer generated assignments, word processed writing, and research via the Internet or web are accepted practice, within limits. Teachers still require children to generate writing on paper, some internet sites are restricted within schools, and certain digital literacy forms of expression are not accepted. Related to the claims of Prensky (2001), the aforementioned researchers recommend that what constitutes literacy and literacy practice broaden to include new and evolving forms of literacy. In essence, traditional literacy practice should expand to include new literacies such as anime blogs, gaming, memes, fan fiction, headcannons, and multimodal forms of expression. Considering how the Internet alone has shifted literacies, emerging literacies must be valued. As literacy shifts at an increasingly rapid rate, these advances must be incorporated into learning environments. Blogs are capable of replacing the traditions of journaling; memes can be accepted ways to respond to events, headcannons can serve as a viable methods of responding to literature, anime and manga can be viewed as appropriate

forms of creative expression, and gaming within classrooms can be suitable practice. As digital literacy evolves, new and yet-to-be created forms and practices of digital literacy will need to be incorporated into schooling, career expectations and the work place.

Digital literacy experts propose that the current generation of teenagers, defined as digital natives of the E-Generation, are more likely to acquire or have acquired the digital competencies needed to effectively traverse multi-dimensional and quickly evolving digital environments. For many adults who grew up in a world of books, magazines, and print literacy, negotiating cyberspace can be uncomfortable. Prensky (2001) claims that individuals who lack digital literacy struggle to live in an unfamiliar world.

Access to Opportunity: The Digital Divide or Digital Literacy as a New Gatekeeper

Access to digital literacy, referred to as the digital divide, has emerged as an issue (DiMaggio & Hargittai; 2001; Flood, Heath & Lapp, 2015; Norris, 2001; Servon, 2008; Warschauer, 2004). In particular, the digital divide threatens to widen existing achievement gaps and further marginalize students who lack access to digital forms of literacy. While access is disparate, especially in urban, rural, and suburban areas, the differences between countries are alarming given rhetoric about global economy and Earth as a global village.

Consider that Coiro, Knobel, Lankshear and Leu report that 99% of US classrooms have access the Internet, the European Union reports 96% Internet access compared to Brazilian schools estimating 26% access and Greece and Poland having 13 to 31% Internet access, while only 5% of Mexican schools were projected to have web access; accessibility rates in sub-Saharan Africa could not be calculated (2008, p. 4). While schools exhibit one view of accessibility, homes and communities can present a differing perspective. Access is tied to economics, whether within or

between geographical areas of a country. Infrastructures are expensive and economics impact access. Despite the promise of global connectivity and immediacy, the issue of access is a gatekeeper to the opportunities afforded by digital literacy.

Coiro, Knobel, Lankshear and Leu recognize the possibilities afforded by digital literacy when they state: "A global network such as the Internet makes it possible to develop and immediately disseminate a new technology of literacy to every person who chooses to access it." (2008, p. 5). However, *choosing* to access and *being able* to access digital literacy is closely tied to economics. In much that same way that older, traditional notions of literacy were tied to having to access books and materials with which to practice, income can impact accessibility to digital literacy opportunities, especially in poor rural and urban schools in the US.

IMPLICATIONS

While digital literacy contains the promise of enhanced learning opportunity, DiMaggio and Hargittai postulate "Internet access is an important resource and inequality in Internet access is a significant public policy issue." (2001, p. 3). As economics emerges as a gatekeeper to digital literacy access, communities and governing structures whether they be schools, school system, families, or countries, will have to grapple with how to provide access to the opportunities digital literacy affords.

Digital literacy has huge implications for teaching, learning, and career opportunities, and these issues correlate to economics. Beach (2012) reiterates that educational practice in the US impacts digital literacy acquisition, calling for a redefined, flexible criteria for assessing student learning that includes myriad forms rather than relying only on criteria created for and linked with traditional print literacies. This implies a teaching force that is current with digital literacies, administration that understands what digital literacy is

and how it functions within school environments, and school systems that invest in digital technologies. Lewis extends Beach's argument, stating that teachers and digital literacy researchers, "need to know what writers of new literacies *do* when they write—what they think about and how they negotiate the demands of new forms and processes of writing." (2007, p. 229).

Economics, access, and education are topics impacted by digital literacy in the US. As traditional literacy, especially print literacy, was once deemed a necessity, Eshet-Alkalai (2004) reminds us that digital literacy skills are becoming basic literacy survival skills for the 21st century. Therefore, educational practices and the use and acceptance of digital literacy skills can also have an impact on work force development.

FUTURE RESEARCH DIRECTIONS

Given that technology and digital literacy will continue to expand, three areas of research should be considered. Research investigating the dissonance between digital natives and digital immigrants, especially in school environments and the workforce should be explored. Research about how children at different ages use digital literacy should be continued. As well, research into why and how some forms of digital literacy enjoy acceptance and legitimacy, particularly in school environs, should be conducted. Finally, access to digital literacy formats, whether it be in schools, in communities, in geographical areas or within the borders of countries should be investigated as well as any restrictions to access that may fall along gender, ethnic/racial, and economic lines.

CONCLUSION

Regarding digital literacy, change is a constant. Shifts in digital literacy within the past 20 years have greatly impacted how people communicate, how teachers teach and children learn in the US and

elsewhere, and the expectations and demands of career and work. Physical space has been removed as a construct and communication and creativity have flourished. Given the rapid evolution of digital literacy, while we do not know how it may evolve, we can know it will evolve, exponentially and in ways not yet imagined.

REFERENCES

- Alkali, Y., & Amichai-Hamburger, Y. (2004). Experiments in digital literacy. *Cyberpsychology & Behavior*, 7(4), 421–429. doi:10.1089/cpb.2004.7.421 PMID:15331029
- Alverman, D., & Hagood, M. (2000). Fandom and critical media literacy. *Journal of Adolescent & Adult Literacy*, 43(5), 436–446.
- Anderson, N. (2002). New Media and new media literacy: The horizon has become the landscape – new media are here. Report produced by Cable in the Classroom.
- Barton, D., Hamilton, M., & Ivanic, R. (2000). *Situated literacies: Reading and writing in context*. New York, NY: Routledge.
- Baudrillard, J. (1990). Mass media culture. In *Revenge of the crystal: Selected writings on the modern object and its destiny, 1968–1983*. Academic Press.
- Bawden, D. (2008). Origins and concepts of digital literacy. In C. Lankshear & M. Knobel (Eds.), *Digital literacies: Concepts, policies and practices*, (pp. 17–32). New York, NY: Peter Lang.
- Beach, R. (2012). Uses of digital tools and literacies in the English language arts classroom. *Research in the Schools*, 19(1), 45–59.
- Bennett, S., Maton, K., & Kervin, L. (2008). The digital natives debate: A critical review of the evidence. *British Journal of Educational Technology*, 39(5), 775–786. doi:10.1111/j.1467-8535.2007.00793.x

- Berners-Lee, T. (1989). *Information Management: A Proposal*. Retrieved from <http://www.w3.org/History/1989/proposal.html>
- Black, R. W. (2005). Access and affiliation: The literacy and composition practices of English-language learners in an online fan fiction community. *Journal of Adult & Adolescent Literacy*, 49(2), 118–128. doi:10.1598/JAAL.49.2.4
- Black, R. W. (2008). Just don't call them cartoons: The new literacy spaces of anime, manga, and fanfiction. In J. Coiro, M. Knobel, C. Lankshear, & D. Leu (Eds.), *Handbook of research on new literacies* (pp. 583–610). New York, NY: Routledge.
- Booth, S. E. (2012). Cultivating knowledge sharing and trust in online communities for educators. *Journal of Educational Computing Research*, 47(1), 1–31. doi:10.2190/EC.47.1.a
- Buckingham, D. (2006). Defining digital literacy. *Digital Kompetanse: Nordic Journal of Digital Literacy*, 1(4), 263–276.
- Coiro, J., Knobel, M., Lankshear, C., & Leu, D. (2008). Central issues in new literacies and new literacies research. In J. Coiro, M. Knobel, C. Lankshear, & D. Leu (Eds.), *Handbook of research on new literacies* (pp. 1–21). New York, NY: Routledge.
- Davies, J., & Merchant, G. (2007). Looking from the inside out: Academic blogging as new literacies. In C. Lankshear & M. Knobel (Eds.), *New literacies sampler* (pp. 167–197). New York, NY: Peter Lang.
- Davies, J., & Merchant, G. (2009). *Web 2.0 for schools: Social participation and learning*. New York: Peter Lang.
- DiMaggio, P., & Hargittai, E. (2001). From the 'digital divide' to 'digital inequality': Studying Internet use as penetration increases. Princeton, NJ: Center for Arts and Cultural Policy Studies, Woodrow Wilson School, Princeton University.
- Eshet-Alkalai, Y. (2004). Digital literacy: A conceptual framework for survival skills in the digital era. *Journal of Educational Multimedia and Hypermedia*, 13(1), 93.
- Flood, J., Heath, S. B., & Lapp, D. (2015). *Handbook of Research on Teaching Literacy through the Communicative and Visual Arts, Volume II: A Project of the International Reading Association*. Routledge.
- Gee, J. (2001). Identity as an analytic lens for research in education. In W. Secanda (Ed.), *Review of research in education* (Vol. 25, pp. 99–125). Washington, DC: American Educational Research Association.
- Gilster, P. (1997). *A primer on digital literacy*. Mississauga, Canada: John Wiley & Sons.
- Graves, D. (1983). *Writing: Teachers and students at work*. Portsmouth, NH: Heinemann.
- Hagood, M. (2000). New times, new millennium, new literacies. *Reading Research and Instruction*, 39(4), 311–328. doi:10.1080/19388070009558328
- Heath, S. B. (1983). *Ways with Words: Language, life, and work in communities and classrooms*. New York: Cambridge University Press.
- Hunter, J., & Caraway, H. (2014). Urban Youth Use Twitter to Transform Learning and Engagement. *English Journal*, 103(4), 76–82.
- Jenkins, H., Clinton, K., Purushotma, R., Robnson, A., & Weigel, M. (2007). *Confronting the challenges of participatory culture: Media education for the 21st century*. Chicago, IL: The McArthur Foundation.
- Jenkins. (1992). *Textual poachers: Television fans and popular culture*. New York, NY: Routledge.
- Jenkins. (2006). *Fans, bloggers, and gamers: Exploring anticipatory culture*. New York, NY: NYU Press.

- Jones, B., & Flannigan, S. L. (2006). Connecting the digital dots: Literacy of the 21st century. *EDUCAUSE Quarterly*, 29(2), 8–10.
- Junco, R., Heiberger, G., & Loken, E. (2011). The effect of Twitter on college student engagement and grades. *Journal of Computer Assisted Learning*, 27(2), 119–132. doi:10.1111/j.1365-2729.2010.00387.x
- Kilinc, E., Evans, R. T., & Korkmaz, U. (2012). Aligning Facebook and Twitter with Social Studies Curriculum. In P. Resta (Ed.), *Proceedings of Society for Information Technology & Teacher Education International Conference 2012* (pp. 517–521). Chesapeake, VA: Association for the Advancement of Computing in Education (AACE).
- Kist, W., Tollafeld, K. A., & Dagistan, M. (2014). Leading ourselves (tweets optional): An analysis of selected Twitter users. *Journal of Adolescent & Adult Literacy*, 58(4), 317–326. doi:10.1002/jaal.362
- Knobel, M., & Lankshear, C. (2014). Studying New Literacies. *Journal of Adolescent & Adult Literacy*, 58(2), 97–101. doi:10.1002/jaal.314
- Kuiper, E., & Volman, M. (2008). The web as a source of information for students in K-12 education. In J. Coiro, M. Knobel, C. Lankshear, & D. Leu (Eds.), *Handbook of research on new literacies* (pp. 241–266). New York, NY: Routledge.
- Lanham, R. (1995). Digital literacy. *Scientific American*, 273(3), 160–161.
- Lankshear, C., & Knobel, M. (2006). *New literacies: Everyday practices and classroom learning* (2nd ed.). New York, NY: Peter Lange.
- Lankshear, C., & Knobel, M. (2008). *Digital literacies: Concepts, policies and practices* (Vol. 30). Peter Lang.
- Lapp, D., Fisher, D., Frey, N., & Gonzalez, A. (2014). Students can purposefully create information, not just consume it. *Journal of Adolescent & Adult Literacy*, 58(3), 182–188. doi:10.1002/jaal.353
- Lewis, C. (2007). New literacies. In C. Lankshear, M. Knobel, C. Bigum, & M. Peters (Eds.), *A new literacies sampler* (pp. 229–237). New York, NY: Peter Lang.
- Lynn, J. (2010). *Internet users to exceed 2 billion this year*. Reuters. Retrieved from <http://www.reuters.com/article/us-telecoms-internet-idUSTRE69I24720101019>
- Marsh, J. (2005). *Popular culture, new media and digital literacy in early childhood*. Psychology Press. doi:10.4324/9780203420324
- Martin, A., & Madigan, D. (2006). *Digital literacies for learning*. London: Facet.
- Mills, K. A., & Chandra, V. (2011). Microblogging as a literacy practice for educational communities. *Journal of Adult and Adolescent Literacy*, 55(1), 35–45.
- New Media Consortium. (2007). *The Horizon Report*. Austin, TX: Author.
- Ng, W. (2006). Web-based technologies, technology literacy and learning. In L. W. H. Tan & R. Subramaniam (Eds.), *Handbook of Research on Literacy in Technology at the K1–2 Level* (pp. 94–117). Hershey, PA: Idea Group. doi:10.4018/978-1-59140-494-1.ch006
- Ng, W. (2010). Empowering students to be scientifically literate through digital literacy. In S. Rodrigues (Ed.), *Multiple Literacy and Science Education: ICTs in Formal and Informal Learning Environments* (pp. 11–31). Hershey, PA: IGI Global Publishing. doi:10.4018/978-1-61520-690-2.ch002

Norris, P. (2001). *Digital divide: Civic engagement, information poverty, and the Internet worldwide*. Cambridge University Press. doi:10.1017/CBO9781139164887

Pool, C. R. (1997). A New Digital Literacy: A Conversation with Paul Gilster. *Educational Leadership*, 55(3), 6–11.

Prensky, M. (2001). Digital natives, digital immigrants part 1. *On the horizon*, 9(5), 1–6. doi:10.1108/10748120110424816

Project Tomorrow. (2012). *Mapping a personalized learning journey: K-12 students and parents connect the dots with digital learning: Speak Up 2011 National Findings*. Retrieved from <http://tinyurl.com/cq7lrvq>

Rideout, V., Foehr, U., & Roberts, D. (2010). *Generation M²: Media in the lives of 8- to 18-year-olds*. Menlo Park, CA: Henry J. Kaiser Family Foundation.

Rodesiler, L. (2015). The Nature of Selected English Teachers Online Participation. *Journal of Adult and Adolescent Literacy*, 59(1), 31–40. doi:10.1002/jaal.427

Servon, L. J. (2008). *Bridging the digital divide: Technology, community and public policy*. John Wiley & Sons.

Squire, K. (2008). Video-game literacy: A literacy of expertise. In J. Coiro, M. Knobel, C. Lankshear, & D. Leu (Eds.), *Handbook of research on new literacies* (pp. 635–670). New York, NY: Routledge.

Street, B. (1984). *Literacy in theory and practice*. London, UK: Cambridge University.

Street, B. (1993). *Cross-cultural approaches to literacy*. New York, NY: Cambridge University Press.

Thibaut, P. (2015). Social network sites with learning purposes: Exploring new spaces for literacy and learning in the primary classroom. *Australian Journal of Language and Literacy*, 38(2), 83–94.

Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Warschauer, M. (2004). *Technology and social inclusion: Rethinking the digital divide*. MIT Press.

ADDITIONAL READING

Alvermann, D., Moon, J., & Hagood, M. (1999). *Popular Culture in the Classroom: Teaching and Researching Critical Media Literacy*. *Literacy Studies Series*. Newark, DE: International Reading Association.

Baskin, C., & Anderson, N. (2008). Learning management systems and virtual learning environments: A higher-education Focus. In J. Coiro, M. Knobel, C. Lankshear, & D. Leu (Eds.), *Handbook of research on new literacies* (pp. 973–998). New York, NY: Routledge.

Coiro, J., Knobel, M., Lankshear, C., & Leu, D. (2008). *Handbook of research on new literacies*. New York, NY: Routledge.

Marshall, J. P. (2008). Gender in online communication. In J. Coiro, M. Knobel, C. Lankshear, & D. Leu (Eds.), *Handbook of research on new literacies* (pp. 491–520). New York, NY: Routledge.

Mortensen, T. (2008). Of a divided mind: Weblog literacy. In J. Coiro, M. Knobel, C. Lankshear, & D. Leu (Eds.), *Handbook of research on new literacies* (pp. 449–466). New York, NY: Routledge.

Quellmalz, E., & Haertel, G. (2008). Assessing new literacies in science and math. In J. Coiro, M. Knobel, C. Lankshear, & D. Leu (Eds.), *Handbook of research on new literacies* (pp. 941–972). New York, NY: Routledge.

Snyder, I., & Bulfin, S. (2008). Using media in the secondary English classroom. In J. Coiro, M. Knobel, C. Lankshear, & D. Leu (Eds.), *Handbook of research on new literacies* (pp. 805–838). New York, NY: Routledge.

KEY TERMS AND DEFINITIONS

Anime: Hand drawn Japanese computer animation categorized by colorful graphics, vibrant characters, and fanciful themes.

Blog: Typically an interactive forum or site for discussion or information-sharing created by an individual or group, centered on differing genres such as politics, music, education, health, travel etc., published on the web where entries, referred to as posts allow visitors and members to leave comments about a topic of interest and create a networking community of users.

Cyberspace: Virtual environment where digital communication over computer networks occurs; typical used to represent the many ways ideas, information, and communication are shared via the Internet and networking sites.

Digital Environments: A virtual or cyber-generated environment accessed or created through the use of one or more digital devices such as a computer, tablet, or a cellular phone.

Digital Literacy: A broad term that refers to the use of literacy skills – reading, writing, listening, to communicate and interact within digital environments and/or using devices and cyberspace to compose and comprehend thoughts.

Fan Fiction: Individuals/fans who write new stories using the characters and setting of published, popular media and share via the web.

Headcanons: From the term “canon,” which refers to events and development that happen in the actual published/aired/official text of a book, television show, movie, etc.; it is something that someone wants to believe is true or enjoys speculating on in a particular fandom based on evidence from the canon or other factors, but is generated and actually exists within their own mind.

Internet: The networked system of private, public, business and governmental mainframe, personal, and wireless computer networks utilizing the Internet protocol suite (TCP/IP) to connect devices globally.

Manga: Japanese graphic novels created in Japan and written in Japanese.

Meme: Typically an imitated image such as a photograph that is used to transmit an idea by adding text via the web.

New Literacies: The technical ability and emotional skill level needed to generate thought and to communicate in multiple formats in digital environments.

World Wide Web: A global publicly available hyperlinked information system, accessed through the Internet, also called simply The Web or WWW credited to Berners-Lee (1989).

Category E

Economics

ICT Investments and Recovery of Troubled Economies

E

Ioannis Papadopoulos

Metropolitan College Thessaloniki, Greece

Apostolos Syropoulos

Greek Molecular Computing Group, Greece

INTRODUCTION

Over the last few years, many countries faced an unprecedented crisis that negatively shaped their economic and social landscape. Here, the term “crisis” means a general turbulence in economy, which is expressed through various negative events in the public and the private Sector of the economy. Common denominator is the radical increase of unemployment that leads to negative social phenomena, poverty and forced immigration. The discussion that follows is based on (Blanchard & Johnson, 2012).

Indicatively, in Public sector we have reduced GDP (Gross Domestic Product - the total amount of values produced in economy in a year) that is usually translated into potential personnel reductions (or in the best case there are no new hirings). Salaries and pensions are suffering cuts, and the same holds for the budget of public investments (new roads, hospitals, schools, etc.). Additionally, the state faces difficulties in supporting the public debt [the accumulated amount of money that the state has borrowed in the past (loans) from international financial markets in order to support its everyday needs that also include paying back other older loans], constant deficits (the public spending is higher than the collected taxes) and incapability of borrowing new loans from international markets. This happens because of the insolvency of the country based on the fact that the annual debt payments due are not supported from the overall operation of the whole economy and from its fu-

ture development prospects. Accordingly, in the private sector, firms face reduced demand for their products, difficulties in borrowing money from banks, increased tax payments and other negative facts that lead them to reduce personnel, reduce their program of investments and hold money because of the uncertainty prevailing in economy.

The opposite of crisis is development. The term “development” means the expansion of an economy (i.e., new jobs, increasing of salaries and pensions, new public and private investments, new prospects and other positive social and economic events).

One of the most important elements of this brief analysis is the term investment. In simply words, an investment is a dedicated amount of money used to create something, with the hope that this will return in the future some profit (typically a multiply of the initial amount) and will support the economy in its expansion. It is the cornerstone of the prosperity of a society and an economy. Investments are separated in public (roads, schools, energy infrastructure, hospitals, etc.) or private ones (new buildings, new businesses, new production lines, collaboration with public sector, etc.). Development of an open economy can only be achieved through investments. In other words, investments are done so to create wealth. Investments in *information and computer technology* (ICT henceforward) are highly important because they facilitate transactions, reduce bureaucracy and accelerate the operational procedures of state and businesses.

The global financial crisis of 2008 triggered a chain of events that revealed many long term structural problems well established in the foundations of the economy of a series of countries such as Greece, Ireland, and Portugal. Such problems were well hidden for many years and because of the global prosperity their importance was underestimated. Because of this, they could not affect the basic economic figures. Such countries had the ability for years to receive loans from international markets (large investment banks such as Goldman Sachs or JP Morgan) and with this method they could proceed in their everyday operational function. But after the crisis of 2008, the matter of insolvency (the degree to which a country is capable of paying back their loans) turned into a highly significant factor. Based on this, the following chain was triggered:

- The global financial crisis (which was primarily initiated from the collapse of the 4th largest investment bank in the USA, Lehmann Brothers) led to solvency crisis.
- The solvency crisis led to a borrowing crisis (global markets were not willing to borrow some countries because of the doubts they had over their ability to pay back these loans).
- Therefore some countries were not able to receive new loans from financial markets.
- This situation led to a debt crisis, meaning that these countries were not able to receive new loans in order to pay back older ones.
- Eventually, the debt crisis led to public expenditures cuts and an increase of taxes which in turn led to an economical and social crisis.

BACKGROUND

The declination of an economy can be triggered from various factors. Officially it is said that an economy declines (i.e., is getting into recession) if the GDP in two consecutive quarters has a

negative growth. The most significant examples of recession periods over the last century was the American great recession of 1929 - 1933 and the events that took place in the Eurozone after the global financial crisis of 2008. Once the economy declines, a series of negative phenomena are taking place at an economical and social level. All such factors have been shortly mentioned in the previous section. A general reference for studying investments is the (Bodie, Kane, & Marcus, 2013).

In general and indicatively we can mention the following factors that lead an economy to declination:

- **Economic Crisis:** A crisis has not the same form across countries and/or chronological periods. A very good example is the Eurozone debt crisis which was initiated after the global financial crisis of 2008. In Ireland, the crisis was initiated from private banks and then was expanded to the public sector (because government decided to assist private banks with billion of euros to avoid bankruptcy which led to increased deficits). In Greece, the crisis was initiated from the state (the Greek government could not borrow money from international markets) and then expanded to private banks. This was happened because these banks had exposure to the Greek state debt since they had lent the state by purchasing state bonds in the past. There are also other cases that underline this ambiguous approach (e.g., Cyprus and Russia are two such cases).
- **Uncertainty Based on Political Factors:** The most recent example is the situation of Brexit. Investors, firms, individuals and other economic entities are looking for a stable political and economical environment. Major decisions such as the referendum for leaving EU, create uncertainty and instability that can push major firms to change location or individuals to reduce their demand for goods and ser-

vices. Therefore this pushes the economy to decline (meaning recession and therefore reduced incomes and increased unemployment).

- **External or Internal Shocks from Demand or Supply Side:** Sometimes specific technical reasons can create economic instability. Various turbulences from demand or supply side can take place because of major changes in economy. For example changes can occur in consumers' preferences, shortage in raw materials, future expectations from consumers and/or firms, restrictions in exports or imports, capital controls in banks and numerous other reasons. All these, can cause instability which can lead to a crisis.
- **Increased Public Deficits:** In some cases governments spend more in public expenditures (hospitals, schools, pensions/salaries, defense budget, public investments, that is, new roads, ports etc.) than they receive from taxes. They do so based on macroeconomical planning in which they consider the long term benefits are fully covering the short term damage because of the multiplying economic and investment effects in economy. In some cases, they do so simply because their policy makers apply low quality planning which comes along with irresponsibility. If this is the case, the national budget may show public deficits which in turn place additional burdens over the historical debt of the country. This in turn may lead to borrowing problems. This last case, leads to an increase of taxes in the private sector with all the negative effects this may cause. The result is a potential decline of the economy which initially was triggered from public deficits. The Greek crisis is a typical example of this sort of crisis.
- **Reduced Competitiveness:** Innovations in industry, new methodologies, new and complex productive procedures create

intense global competitiveness and a demanding environment for businesses and states. New markets and challenges constantly appear while firms and states that cannot be adjusted in time in this totally new environment see a reduction of their market share in the global arena. If this is the case, the exports are going to decrease while imports will increase (X-M). This triggers instability in the Balance of Trade (BOT) that causes budgetary and macro-economic problems. When this happens, the economy declines and triggers all the negative effects in society. To reverse this development, states and firms must rapidly take decisions and proceed in new visionary plans in order to create new, competitive and efficient market procedures and products that will allow them to strongly stand in global markets.

While an economy can be declined due to various and some times to nonexistent technical reasons (i.e uncertainty), recovery is much more difficult to be realized. It is the same just like saying that a building needs years to be constructed but it is just a matter of seconds to be demolished. Once the declination occurred, meaning that an economy gets into recession for many semesters / years, public policy makers and firms' owners must focus to the recovery which must be realized as soon as possible. The more it holds the greater are the negative results to citizens and to society as a whole and the harder the way is to getting out of the tunnel.

Indicatively, the following fundamental methods can be used for getting an economy back to the root of growth:

1. New private investments, domestic and foreign (FDI - Foreign Direct Investments) in critical sectors of the economy that historically appear competitive advantage in the global market. What is the impact of an investment in economy? Consider the

calm surface of a lake. Consider now a big stone being released above the lake. Once the stone touches the water, homocentric, three-dimensional wavy (outbound) formations are being triggered. Now consider not just one but many stones being released over the lake at the same time. The same multiplying effect holds in the real economy. The more investments are being created the larger the multiplying impact stands for the economy. Many new investments create jobs, wealth and supervalues which leverage the economy initially to moving out of the recession (recovery) and subsequently to further development and prosperity. For example in Greece, the country suffered the most since the global financial crisis of 2008, new investments can be realized in sectors with competitive advantage. This is the case for boutique tourism, renewable sources of energy, shipping, mediterranean foods, services (private education, private health, informatics etc.) and other.

2. Intervention in Fiscal Policy from the ministry of finance. This intervention can come through various forms but the most well known are the following ones: Firstly, with new public investments. This is in line with the so called Keynesian policies. Famous 20th century economist John Maynard Keynes, supported the idea that when an economy comes to declination the state must support the economic circuit by increasing the public expenditures focusing in critical functions of economy i.e building new roads, ports, other important infrastructure so as to accelerate the transactions in the economy and as a consequence to move the economy again. This approach assisted US economy to move out of the recession in 1933. The Hoover Dam (amongst other infrastructure investments) was constructed in this framework. Secondly, by reducing the public consumption in carefully selected public activities which appeared as counterproductive activi-

ties with increased bureaucracy and limited efficiency. This must be executed carefully, by bypassing such activities through the re-routing of the withdrawn resources to other efficient public activities that will actively assist the economy to making reboot. Lastly, interventions in public salaries and pensions where necessary.

3. Intervention in Monetary Policy from the central bank (in EU is the ECB). This intervention basically comes in three forms: Firstly, through the basic interest rate. When the basic interest rate, determined from the central bank, is being lowered, this gives the signal to commercial banks to lower their own borrowing interest rates (which are being added on the basic rate - i.e if for example the basic rate = 1% and the required profit from a bank is 3% then the commercial rate from this bank is $1+3=4\%$) and making the money cheaper. In this way, banks increase the number of provided loans to individuals and firms and thus increase the circulation of money in the economy. This increased circulation can be rooted to investments and to the production of new goods and services. Secondly, by increasing the circulating monetary mass. This is the total number of printed notes and money (cash, deposits, reserves etc) circulating across the economy (in banks, individuals, firms, state and other entities). The Central bank by using some techniques (called open market actions) can increase the monetary mass and thus creating artificial demand in economy over the short time. Inflationary phenomena may be appeared but these can be absorbed at a later time over the mid-long term. Such phenomena are barely appearing when the economy suffers crisis. Lastly, through the combination of the two methods which is the most common in the modern world.
4. Cooperation between the state and private sector in healing the primary reasons that led this economy to decline and designing

the future. This can be realized with interventions in various critical sectors of the economy, in institutions and in the public services that cooperate at everyday basis with the private economy. In general, needs a national general visionary plan that will be executed from both state and leading firms so as to get the problematic economy back on track of development.

ICT INVESTMENTS AND ECONOMIC RECOVERY

For many years the ICT sector was, and of course it is expected to remain, one of the largest employers. In the USA there are many ICT companies that have thousands of employers while it is expected to see an increase in job offers. Naturally, similar elements can be said for other countries. Thus the ICT sector can contribute to an economy by directly creating new jobs.

Every now and then, telecommunication companies introduced new forms of home and business Internet connection. Typically, new connections offer higher speeds and better connections while the price of the new technology is always a bit higher when compared to the current technology. But gradually, the price is getting lower, allowing multiplying positive effects in economy. In order to offer a new technology, communication companies need to invest money so to attract more new customers or to convince existing customers to upgrade to new packages. Thus the new technology does not only offer a better internet home or business connection but increases also the GDP, the incomes and the potential prosperity.

Today all modern states offer some sort of online service for citizens to submit their tax declaration thus avoiding queues and making the whole process easier, faster, and more efficient. More generally, many public services have become available online and some of them through mobile phones. Online bank account access is an example of a service that is both online and

through mobile phones. What is more interesting is that new online or mobile phone services and applications have been appeared. All these are crucial elements that apart from the fact that make the life of citizens easier, allow foreign and domestic investments to be accelerated but also attract new ones. In order to ensure the smooth operation of these new services it is necessary to make investments in new IT infrastructure, hire people, etc., thus contributing to the growth of the economy not mentioning the millions of working hours-saved annually by avoiding bureaucratic procedures and queues. Finally, all these systems will need a regular maintenance framework which allows the further creation of new jobs and new incomes as well.

Impact sourcing is a new “industry” whose purpose is to hire people that will perform digital tasks such as transcribing audio files and editing product databases. It is a typical process of business outsourcing which can boost economic development. Typically, these digital tasks are called microworks and employees complete them by using a web-based interface. This way, skillful people from poor countries can find a job and work for a big company without the need to relocate. We have such examples in India or China. This idea has been introduced by Samasource, a San Francisco based company.

Today almost all companies have a web site where their products and their operations are presented. In many cases, the companies have web stores where customers can directly purchase the company’s products. For example, all publishers have web stores while most companies that produce shoes and/or clothing have web stores too. In addition, most companies have profiles in popular social media sites while they run promotional competitions through their web site or their social media profiles. Naturally, companies need either to have a dedicated ICT department or outsource these tasks to ICT companies that specialize in this area. All these create never thought before business opportunities.

SOLUTIONS AND RECOMMENDATIONS

It is next to impossible to give a specific solution that can be applied to every possible troubled economy. Each economy has its peculiarities and thus is difficult to specialize the general solution. Thus instead of giving a general solution, we will discuss how IT investments can help a specific troubled economy to recover. In our case, the trouble economy will be Greece's economy and the specific solution suited for Greece. The country is facing a debt crisis and has no access to markets for more than seven years. We will not discuss what led Greece in this negative situation. Instead it is better to focus to the future and discuss how the country can possibly get out of this crisis. We have already outlined the benefits of investing to ICT and now will explain how such an investment can improve a troubled economy.

First of all, it is rather important to present a few facts about the use of ICT in Greece. In the end of 2014 65% of Greek households had a broadband connection when the EU average was 76%. Also, on the same year only 9% of companies were selling their goods via web shops while the EU average was 15%. In addition, 84% of enterprises are using e-government services while the EU average is 88%. These figures clearly show that there is plenty room for improvement. In addition to these characteristics, there are some serious drawbacks in the current situation. First, it seems that computer literacy is high enough, nevertheless, it is not that complete (e.g., most people know next to nothing about computer hardware and internet connectivity). Also, ICT systems are poorly maintained. Moreover, there is no plan for the use of public information and data. But the most crucial problem is that in every change of ministers and/or government a new political plan is adopted. From a purely technical point of view, there was no central planning on what computer platforms to use so that different ministries use different systems and platforms. Obviously, a good solution to avoid this sort of chaos is to

adopt a specific open source computer platform and switch all existing systems to this platform. Apart from this obvious move, one can propose the following interventions as well:

- The modernization of public administration (e.g., reduction of bureaucracy by adopting a paperless office).
- Reconnection of citizens with the state (e.g., e-unification of public services and systems).
- Adoption of a new digital personal card for all citizens that will provide them full access to all public services.
- Open access to public data.
- Adoption of cloud, etc.

These points could create new jobs, facilitate transactions, save working hours and make the various public facilities more attractive to use by citizens and firms. They will also attract new foreign investments that will boost the economy with all the positive multiplying economic effects that this may cause. In addition, it is mandatory that the government should urge enterprises to enter the e-market and people to prefer e-shopping (a side-effect of it would be a reduction of tax evasion). Also, the government should strive to have skillful citizens, which naturally means that school curricula must be adapted accordingly.

FUTURE RESEARCH DIRECTIONS

We have explained what a troubled economy is and how ICT investments can help such an economy to stand on its own two feet. Since there are no general recipes on what exactly should be done in every possible case, it would quite interesting to try to devise some method (or "algorithm") that could guide someone to choose the best solution for each particular case. Naturally, such a method should take into consideration a number of factors such as the level of training of citizens, the current technological infrastructure and achievements

(i.e., one should propose solutions that make sense), the potential availability of loans that can be rooted to ICT projects, the general legal framework, the level of international collaborations from firms and state, the general investment and tax environment that can attract ICT global firms to move resources in this country etc.

CONCLUSION

Economic recovery is not something that can be achieved overnight. The government of a troubled economy or the administration of an international organization that assists troubled economies to recover (i.e IMF, EU etc), needs to carefully plan the possible recovery of the economy. Clearly, an economy can recover when many medium and large scale investments are being made at a constant rate in sectors with competitive advantage. However, it seems that amongst others, ICT investments are quite viable mainly because they are not so demanding in terms of funding. For example, in order to build a shipyard it is necessary to find a good location, build the appropriate facilities, and install the required equipment. However, in the case of ICT investments, one needs a room, a few computers and some trained personnel. This core infrastructure can massively assist entire organizations and public services accelerating transactions and creating room for multiplying positive economic phenomena. Thus when a nation invests in ICT, actually invests in the future and one can expect only growth of the economic output and prosperity for its citizens.

REFERENCES

Blanchard, O., & Johnson, D. R. (2012). *Macroeconomics* (6th ed.). Harlow, UK: Pearson.

Bodie, Z., Kane, A., & Marcus, A. J. (2013). *Investments* (10th ed.). McGraw-Hill Education.

Kvochko, E. (2013, April 11). *Five ways technology can help the economy*. Retrieved July 12, 2016, from World Economic Forum: <https://www.weforum.org/agenda/2013/04/five-ways-technology-can-help-the-economy/>

Staats, F. G. (2012, December). *The Microwork Solution*. Retrieved July 12, 2016, from Harvard Business Review: <https://hbr.org/2012/12/the-microwork-solution>

Tsakanikas, A., Danchev, S., Giotopoulos, I., Korra, E., & Pavlou, G. (2014, December). *ICT Adoption and Digital Growth in Greece*. Retrieved July 13, 2016, from http://iobe.gr/press_dtl.asp?EID=59

ADDITIONAL READING

Caribbean, E. C. (2010). *ICT for growth and equality: renewing strategies for the information society*. Retrieved September 20, 2016, from Comisión Económica para América Latina y el Caribe: <http://www.cepal.org/en/publications/3008-ict-growth-and-equality-renewing-strategies-information-society-summary>

Maryska, M., Doucek, P., & Kunstova, R. (2012). The Importance of ICT Sector and ICT University Education for the Economic Development. *Procedia: Social and Behavioral Sciences*, 55, 1060–1068. doi:10.1016/j.sbspro.2012.09.598

Parkin, G. (2011). *Digital Marketing: Strategies for Online Success*. London: New Holland Publishers.

Saunders, A., & Cornett, M. (2015). *Financial Markets and Institutions*. New York: McGraw-Hill Education.

Stiglitz, J. E., & Rosengard, J. K. (2015). *Economics of the Public Sector* (4th ed.). New York: W. W. Norton & Company.

KEY TERMS AND DEFINITIONS

Economic Crisis: A situation in which the economy of a country experiences a sudden downturn brought on by a financial crisis.

Economic Recovery: A period of increasing business activity signaling the end of a recession.

Economy: The large set of inter-related production and consumption activities that aid in determining how scarce resources are allocated.

ICT: An umbrella term that includes any communication device or application, computer and network hardware and software, satellite systems and so on.

Investment: An asset that is purchased with the hope that it will generate income or appreciate in the future.

Uberization (or Uberification) of the Economy

E

Nabyla Daidj

Telecom Ecole de Management, France

INTRODUCTION

The management of technological innovation is one of the most demanding challenges today (Dodgson et al., 2008). The external environment characterized by globalization, convergence, competitive/market uncertainty, time-to-market pressure, shortening product lifecycles is also based on knowledge, information, fast-changing technology and an innovative economy. In recent years a series of innovations and trends have changed the way people perceive technology. The global availability of the Internet, along with innovations (products, services and applications) explain certain aspects of the dynamics of the innovation process, the diffusion of technology and the development of various platforms (product and service marketplace, social networking platform, content platform).

The Uber taxi-booking smartphone app, matching those who have cars with people who need rides quickly, was created in 2009 in the US and has been progressively launched all around the world. Since this pioneering app, several other start-ups operating in various activities (transportation, flower and food delivery, events, home services, legal services) have adopted a Uber-like business model (BM) and more and more companies are looking to disrupt regulated industries, such as banking or healthcare.

Uber has been a driving force behind the emergence of a new kind of platform connecting consumers and providers in real time and organizing information without any ownership of the products concerned. Uber, based on trust (MacDonald, 2016), has changed the rules of the

economic game and its success has inspired the term “uberization”.

The objective of this chapter is to gain a more precise understanding of uberization and in particular its theoretical scope. How would we combine uberization with other accepted and widely used concepts in economics and strategy, such as platforms, multi-sided markets, externalities and business ecosystems? Is uberization a relevant concept, an all-purpose word with a multitude of meanings or just a buzz word?

This chapter attempts to answer these questions and is structured as follows. This first section introduces the topic and key concepts and presents a brief overview of the evolution of the Uber phenomenon to a larger trend named uberization or uberification. The second section discusses the conceptual and theoretical framework to analyze the development of uberization. It develops a strategic and economic perspective. In section 3, several recommendations are proposed. Section 4 suggests directions for further research. Section 5 concludes.

BACKGROUND

From Uber to Uberization

Uber (the mobile application) is not uberization (Table 1). In recent years, a number of studies have been performed to gain further insight into the uberization phenomenon, notably empirical publications. The particular attention given to uberization (even in non-specialized press) shows how its importance has grown year by year since

Table 1. From Uber to uberization

Uberization of Work - Employment	“But of all the ways that Uber could change the world, the most far-reaching may be found closest at hand: your office. Uber, and more broadly the app-driven labor market it represents, is at the center of what could be a sea change in work, and in how people think about their jobs”. (Manjoo, 2015) “There has been a lot of debate about how online platforms have changed the nature of work. In some cases, on-demand companies have been harshly criticized for making employer like demands on workers but denying them basic benefits and protections”. (Bernard, 2016)
Uberization of Finance and Banking	“We are on the verge of the Uberization of finance, which will bring multiple new opportunities but also a range of new risks (...). Uber is a high-tech middleman that is making the intermediaries of the past obsolete. The financial world is one of the most mediated industries on the planet, and that is precisely what is about to change. (Karabell, 2015).
Uberization of Business Schools	“What could more surely come next is the appearance of low cost and/or no frills business schools, focusing on the essentials and offering options. That’s partially the idea of online business schools, with some success, but it never really diffused to brick-and-mortar or hybrid models. With the strong discussions on fees in many countries of the world, there is now room for the emergence of a new type of business schools.” (Therin, 2015).

Source: specialised web sites.

the end of the 2000s. The definitions given in the general and specialized press are generally broad or applied to a particular industry or activity.

Uberization should lead to a transformation of the entire economy. Uberization is a phenomenon based on on-demand services (enabled by the Internet and smartphones) and peer-to-peer platforms combining imitable features such as dynamic pricing sometimes called “surge” pricing (Horpedahl, 2015; Smith, 2016), mobile payment, rating systems, algorithmic and data management (Hall et al., 2015).

Towards a New Kind of Capitalism or Its End?

Uberization may turn out to be a concept in the future seen as a new, complex and dynamic form of liberalization (economy) and a more perfect supply and demand equilibrium (market).

Some authors suggest the evolution of capitalism towards ‘postcapitalism’ (Mason, 2015) or ‘platform capitalism’ based on a physical and digital transformation of how goods and services are produced, shared and delivered and how customers (final consumers) engage directly with each other through on-demand platforms (Davis, 2016); others consider uberization “as a form of populism” (Kyrou, 2015, p. 106).

Some experts are even predicting the end of capitalism with the development of sharing, collaborative and participatory practices. However, the rules governing these practices activities (Cusumano, 2015) should be compatible with the market economy and capitalist system. Experts already believe that the potential for growth of the sharing and on-demand economy will be significant in the future. This nascent economy will set the stage for reshaping the economic system, renewal of the market relationships between economic actors, changes in the nature and structure of work (labor considered as a commodity) and work organization and consumer behavior shifts.

The uberization phenomenon calls for a multidimensional framework for understanding its development and impact including the following levels of analysis.

Country Level

The impact of uberization may differ in complexity from country to country depending on the national legislation concerned. The questions raised by uberization have become more pressing as they are closely related to a wide variety of laws and regulations including labor law, antitrust and competition law, and corporate law. The uberization of the economy is widely debated

all around the world and reactions differ from one country to another. In the European Union, UberPop (where the Uber service is directly challenging the taxi market), has given rise to many disputes. Consequently, “a number of regulators in Member States took actions based on existing legal frameworks that resulted in administrative and/or criminal charges against Uber drivers and management (e.g. Netherlands, Portugal, France, Spain and Germany). As a response, Uber submitted complaints to the European Commission” (Azevedo & Maciejewski, 2015, p. 1). In India, the situation is quite different where it was not so much the regulatory and legislative framework that posed problems, rather the Uber mode of payment. Finally, Uber has accepted cash payments for rides. In the United States, Uber claims “that it is preempted from all regulation because it is an information service provider under the Telecommunications Act of 1996.” (Elliott, 2016, p. 743).

Industry and Market Level

There are several factors that determine the market structure of a particular industry: buyers and sellers (number, interactions between them, bargaining power), prices, production and selling processes, and product/service differentiation based on innovation. Uberization affects all sectors of the economy, encourages the emergence of new players, creates new market opportunities and changes the competitive environment and the degree of competition to a certain extent. But as mentioned by Gabel (2016), “entry into the taxi industry involves few risks. Entrants have lower costs than the incumbents, sunk costs are small, and modern technology makes it easy to hail a cab using the Internet. Despite large scale entry and low barriers to entry, monopoly power persists. The persistence of monopoly power illustrates that new technologies may not quickly eviscerate monopoly power.” (p. 527).

Company Level

“Traditional companies now have another set of competitors to worry about – Internet startups in the “sharing economy.” These new companies are actually Web platforms that bring together individuals who have underutilized assets with people who would like to rent those assets short-term” (Cusumano, 2015, p.32). Consequently, uberization may lead most of the incumbent large companies, more often than in the past, to review their future strategy in order to adapt to the arrival of several new players and to achieve a sustainable competitive advantage based mainly on differentiation. There is a risk today that temporary (or transient) advantage is becoming the rule rather than the exception (McGrath, 2013; Carpenter et al., 2014; Daidj & Aras, 2015).

Platform Level

Uber can be considered as an ICT platform and uberization refers then to a platform-based economy reshaping firms’ and sectors’ boundaries. In recent years, the phenomenon of the platform has been observed and the concept of the platform has been thoroughly examined by several scholars (see next section) who mostly focus on specific theoretical aspects of this issue or investigate particular industries.

UBERIZATION: A REVIEW OF LITERATURE

Theoretical studies on uberization are still limited even if scholars have recognized the importance of the phenomenon which could be analyzed through the lens of marketing, strategy, finance, international business, human resources management, information systems and labor law. Uberization is a complex, dynamic and multi-dimensional issue. Probably due to the many disciplines involved in uberization – spanning a diverse field from orga-

nization and management research to computer science and information systems – a unified body of research on uberization and the resulting social, economic and strategic implications remains a matter of debate.

Various points of view of different scholars in two main disciplines are presented below as interaction and interdependence between economics and strategic management have become frequent and multi-faceted. The economics literature has addressed more specifically the challenging platform issue by defining several types of platform (one-sided, two-sided and multiple sided markets) and through the concept of externalities. In the strategy literature, uberization typically revolves around innovation, firms' strategies, business models, business ecosystems and relationships between actors shaping the emergence and the development of different types of cooperation and competition (i.e. coopetitive practices). This section is dedicated to emerging issues with regards uberization of the economy.

Uber: Innovation and Economic Impact on Business Models

Christensen in his book, "The Innovator's Dilemma" (1997) split new technology into two categories: sustaining and disruptive. Sustaining technology relies on incremental improvements to an already established technology. Disruptive technology is a term used for the first time to describe a new technology that unexpectedly overturns the dominant technology in the market sector and that forces changes in industry frontiers, business processes and business models (BMs). Disruption is rarely the result of a single innovation but occurs when two or more technologies converge.

But innovation is not necessarily technological. According to the Oslo manual (OECD, 2005), an innovation is the implementation of a new or significantly improved product (good or service), or process, a new marketing method, or a new organizational method in business practices, workplace organization or external relations. "If

anything, "uberizing" probably means "trapping" a set of innovative procedures – geo-location, online payments and workforce management and distribution – into an "app-accessible service" or a "sweatshop", according to its critics, with lower entry barriers since people monetize resources they already own." (Aloisi, 2015, p. 10).

Regarding BMs, they evolve as the technology matures. "Disrupters often build business models that are very different from those of incumbents" (Christensen et al., 2015, 49). But can Uber be considered as a disruptive innovation and/or a disruptive business model? Nevertheless the following question remains unanswered today. According to Christensen et al., (2015), "Uber is clearly transforming the taxi business in the United States. But is it *disrupting* the taxi business? According to the theory, the answer is no. Uber's financial and strategic achievements do not qualify the company as genuinely disruptive—although the company is almost always described that way" (p. 47).

Uberization, Platforms, and Multi-Sided Markets

Since the beginning of the 2000s, the platform issue has been studied from an economic point of view by several scholars (Caillaud & Jullien, 2001; Armstrong, 2002; Evans, 2003; Ferrando et al, 2003; Rochet & Tirole, 2006; Armstrong, 2006; Cortade, 2006; Evans & Schmalensee, 2010; Weyl, 2010). Platforms are based on technologies that provide support for, and interact with, products and services of other firms. Gawer and Cusumano have highlighted the critical role played by platforms (Gawer & Cusumano, 2002; Gawer & Cusumano, 2008; Baldwin & Clark, 2000).

More recently, Staykova and Damsgaard (2014) considered that one-sided, two-sided and multi-sided platforms should be distinguished clearly (Table 2). "[...] one-sided platforms differ from one-sided markets which function predominantly as resellers (Hagiu and Wright, 2011). One-sided platforms facilitate the communication between

Table 2. From one-sided to multi-sided platforms

One-Sided Platform	Two-Sided Platform	Multi-Sided Platform
• Danske Bank (with Mobilpay) Barclays (Pingit) • Betfair (the world's biggest betting exchange) previous activity before allowing other bookies to offer bets in competition with one another.	• The game console sector (consoles, video games) • Credit cards • Yellow pages	• Exchange platforms (peer-to-peer platforms): Airbnb, Uber • Product marketplace: Amazon, ebay • Payments platforms: Paypal • Social networks platforms: Facebook

Source: Elaborated by the author.

the users of the platform who form one distinctive group of consumers which exhibit same-side network effects and have interchangeable roles. Thus, they differ from the two-sided platforms which link two distinctive groups of users (consumers and merchants) with strong cross-side network externalities” (see below).

Multi-sided platforms allow direct interactions between two or more distinct sides and each side is affiliated with the platform. Affiliation refers to the fact that users on each side have to make platform-specific investments that are necessary in order for them to be able to directly interact with each other. The investment could be “a fixed access fee, expenditure of resources or an opportunity cost” (Hagiu & Wright, 2015, p. 6). According to Hagiu and Wright (2015), Uber can be considered as a multi-sided platform (MSP) enabling independent contractors or professionals to deal directly with clients. MSPs reduce transaction, coordination and search costs between two or more distinct groups of users (Staykova & Damsgaard, 2015).

Staykova and Damsgaard's (2014) approach is also of interest because it allows a better understanding of platform evolution as it introduces changes in the platform and development dynamic within each stage as the competition conditions evolve based on evolutionary economic concepts.. A platform is defined “as one-sided since the new product or service is still developing. As the platform becomes two-sided during the next stage, the level of competition slowly rises, but the preference is given to cooperation. The competition intensifies as the platform matures and becomes multi-sided.” (Staykova & Damsgaard, 2014, p. 12).

Uberization, Network Externalities, and Externalities

The platform industry is generally characterized by the existence of high network externalities. Network analysis is often related to the theory of positive network effects which describes a positive correlation between the number of users of a network for a specific good/service and the utility of this good/service. “There are many products for which the utility that a user derives from consumption of the good increases with the number of other agents consuming the good” (Katz & Shapiro, 1985, p. 424). Network externalities are a well-known phenomenon in the IT sector, though they exist elsewhere (Arthur, 1989; David, 1986; Foray, 1989, 2002).

Two types of network externalities can also be distinguished: direct and indirect externalities. Direct externalities appear when the satisfaction (utility) stemming from the consumption of a product or use of a service directly derives from the number of individuals already consuming the product or using the service. It is thus an intra-group externality. The more the product is used, the more value it acquires in relation to others. Indirect externalities emerge when the demand for a product in a market depends on the offer of another product without which the first cannot function. The appearance of this type of externality is conditioned by the existence of two complementary products forming a ‘system’ product or a platform (Daidj & Isckia, 2009). In this context, two categories of agents are distinguished, and the value of a product for one category is positively

correlated to the number of agents on the other side of the market. This is an inter-group externality. The platform thus internalizes the indirect externalities which can usually be divided into two types: membership (or affiliation) externalities and usage externalities (Rochet & Tirole, 2004). Uber success can be explained partly by its capability to benefit from network externalities, mainly in recruiting drivers. Its development strategy has aimed to reach this critical size both at local and international level.

Uberization is also closely related to scalability (taking advantage of network effects) and to a specific kind of externalities, named reputation externalities. Reputation mechanisms could reduce inefficiencies in markets with asymmetric information and assure quality of service. Even if reputation effects can be extended beyond market platforms (Noskoy & Tadelisz, 2015), reputational externalities represent a key to the success of a peer-to-peer market place based on experiences made by individuals. Reputational risks should not be underestimated; a good reputation is crucial leading to the setting up of reputational feedback mechanisms and ratings.

At a more general level, economists make a distinction between what are called positive and negative externalities. Externalities represent a major cause of market failure. Rahel (2016) has analyzed the economic impact of Uber on the taxi industry and the potential negative externalities (efficiency, quality, and safety) that come with this new technology.

Uberization and Business Ecosystems

The 2000s have been characterized by the emergence of a “new form” of network organization: the business ecosystem (BE). This type of network spans a variety of industries and it is often associated with platforms. Moore (1996) defines the business ecosystem as a business community that brings together various inter-dependent players who belong to different sectors. BE refers to an

undefined, non-geographic network (often based on a platform” or “on platforms”) characterised by cooperative and “open” practices (among suppliers, organizations and customers) in order to co-create value (Daidj, 2013; Daidj, 2015).

Electronic platforms play a strategic role in enhancing value creation within the business ecosystem by sustaining input from various stakeholders. Even if the platform needs a leader, firms are embedded within business ecosystems, the performance of which influences the success and survival of all their member firms. Uber, through the Uber app, has been considered as an independent transportation provider creating a new kind of business ecosystem including, in particular, independent third party transportation partners (driver partners). Smith (2016) has used the expression of “uber-all ecosystem” (p. 389).

More generally, the uberization of the economy has led to the development of intermediaries within each sector and to new partnerships between main players.

SOLUTIONS AND RECOMMENDATIONS

Companies are moving forward with digital transformation at varying paces and experiencing varying levels of success. Since its creation, Uber has radically disrupted an entire industry (transportation) and has continued to transform many industries around the world, removing economic, social and behavioral barriers. All stakeholders are affected by Uber’s development: regulators, incumbents, services providers, employees (for some professions), consumers etc. All of them will have to face several challenges. Challenges are competitive, technological and organizational.

As the previous sections have shown, multi-sided platforms and resulting peer-to-peer platforms, such as Uber, are complex phenomena and corresponding research is very diverse. Accordingly, several theoretical perspectives can be applied to investigate the related phenomena.

FUTURE RESEARCH DIRECTIONS

Uberization is an evolving concept and researchers should be aware that uberization issues are strongly interconnected and that multi-disciplinary approaches will be required to address them in a comprehensive manner. Uberization is at a very early stage and will impose market and technological developments and will require changes to work patterns. Further work is required to better understand the impact of uberization on performances within business ecosystems, including case studies with different types of impact assessments. Further research into the relationships between the main actors will expand the base of evidence to reflect changes that have occurred in recent years in the business landscape in a context of uberization.

CONCLUSION

This chapter has highlighted the main challenges related to uberization. This fast-growing phenomenon is seen today as an important feature of economic transformation. Uberization exposes both traditional groups to direct competition with new entrants (especially peer-to-peer platform operators). The regulation by the public authorities based on the principle of equal conditions of competition for all operators on the market does not want to give new entrants a free rein. Uberization also raises questions about a radical reform of capitalism and of the market economy. Much of the interest in the uberization economy focuses on these online platforms that put forward the virtues of the sharing economy. But facilitating exchanges (of resources and/or competencies) on a platform (such as Uber) does not mean that the human being is put at the center of the economic and social development. Herein lies the heart of the problem.

REFERENCES

- Aloisi, A. (2015). *Commoditized Workers. Labour Issues arising from a Case Study Research on a Set of Online Platforms in the 'On-Demand/Gig Economy'*. Retrieved February 16, 2016 from <http://ssrn.com/abstract=2637485>
- Armstrong, M. (2002). *Competition in two-sided markets*. Mimeo. Oxford, UK: Nuffield College, Oxford University.
- Armstrong, M. (2006). Competition in Two sided Markets. *The Rand Journal of Economics*, 37(3), 668–691. doi:10.1111/j.1756-2171.2006.tb00037.x
- Arthur, W. B. (1989). Competing technologies, increasing returns, and lock-in by historical events. *The Economic Journal*, 99(394), 116–131. doi:10.2307/2234208
- Azevedo, F., & Maciejewski, M. (2015). *Social, Economic and legal consequences for Uber and similar transportation network companies (TNCs)*. Briefing, PE 563.398. Brussels, Strasbourg: European Parliament.
- Baldwin, C. Y., & Clark, K. B. (2000). *Design Rules. The Power of Modularity* (Vol. 1). Cambridge, MA: MIT Press.
- Bernard, T. S. (2016, March 4). Tax Tips for Those Who Make Money in the Gig Economy. *The New York Times*. Retrieved March 12, 2016 from http://www.nytimes.com/2016/03/05/your-money/tax-tips-for-those-who-make-money-in-the-gig-economy.html?_r=0
- Botsman, R., & Rogers, R. (2010). *What's Mine Is Yours: The Rise of Collaborative Consumption*. New York, NY: HarperBusiness.
- Caillaud, B., & Jullien, B. (2001). Competing cybermediaries. *European Economic Review*, 45(4-6), 797–808. doi:10.1016/S0014-2921(01)00123-4

- Carpenter, M., Daidj, N., & Moreno, C. (2014). Game Console Manufacturers: The end of sustainable competitive advantage? *Communications & Stratégies*, 1(94), 39–60.
- Christensen, C. M. (1997). *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail*. Boston, MA: Harvard Business School Press.
- Christensen, C. M., Raynor, M., & McDonald, R. (2015). What is Disruptive Innovation? *Harvard Business Review*, 93(12), 44–53. PMID:17183796
- Cortade, T. (2006). A Strategic Guide on Two-Sided Markets Applied to the ISP Market. *Communications & Stratégies*, 61(1), 17–35.
- Cusumano, M. A. (2015). How Traditional Firms Must Compete in the Sharing Economy. *Communications of the ACM*, 58(1), 32–34. doi:10.1145/2688487
- Daidj, N. (2013). The evolution of new entrants' strategies in the media sector in a context of convergence. The case of Apple, Google and Microsoft. In S. Ben Letaifa, A. Gratacap, & T. Isckia (Eds.), *Understanding business ecosystems, How firms succeed in the new world of convergence* (pp. 165–192). Bruxelles: De Boeck.
- Daidj, N. (2015). *Developing Strategic Business Models and Competitive Advantage in the Digital Sector*. Hershey, PA: IGI Global. doi:10.4018/978-1-4666-6513-2
- Daidj, N., & Aras, G. (2015). Sustainability issue in CSR and strategy: sustainable or temporary competitive advantage in today's dynamic environment. In G. Aras (Ed.), *Sustainable Markets for Sustainable Business. A Global Perspective for Business and Financial Markets* (pp. 163–187). Aldershot, UK: Gower Publishing Limited.
- Daidj, N., & Isckia, T. (2009). Entering the Economic Models of Game Console Manufacturers. *Communications & Stratégies*, 73(1), 23–42.
- David, P. A. (1986). Understanding the Economics of QWERTY: the Necessity of History. In W. P. Parker (Ed.), *Economic History and the Modern Economist*. Oxford, UK: Basil Blackwell Ltd.
- Davis, G. F. (2016). What Might Replace the Modern Corporation? Uberization and the Web Page Enterprise. *Seattle University Law Review*, 39, 501–515.
- Davis, G. F. (2016). *The Vanishing American Corporation: Navigating the Hazards of a New Economy*. Ann Arbor, MI: Berrett-Koehler Publishers.
- Davis, J. (2015). *Drive at Your Own Risk: Uber's Misrepresentations to UberX Drivers About Insurance Coverage Violate California's Unfair Competition Law*. 56 B.C.L. Rev. 1097. Retrieved February 17, 2016 from <http://lawdigitalcommons.bc.edu/bclr/vol56/iss3/7>
- Dodgson, M., Gann, D., & Salter, A. (2008). *The management of technological innovation: Strategy and practice*. Oxford, UK: Oxford University Press.
- Elliott, R. E. (2016). Sharing App or Regulation Hack(ney)?: Defining Uber Technologies, Inc. *The Journal of Corporation Law*, 41(3), 727–753.
- Evans, D. S., & Schmalensee, R. (2010). Failure to Launch: Critical Mass in Platform Businesses. *Review of Network Economics*, 9(4), 1–26. doi:10.2202/1446-9022.1256
- Farrell, M. (June 21, 2014). Uber battle goes to heart of Cambridge. *Boston Globe*. Retrieved January 17, 2016 from: <https://www.bostonglobe.com/business/2014/06/20/uber-battle-goes-heart-cambridge/CpqHJGPP1LogjKrXH31u7O/story.html>
- Ferrando, J., Gabszewicz, J. J., Laussel, D., & Sonnac, N. (2003). *Two-Sided Network Effects and Competition: An Application to Media Industries*. Unpublished paper. Louvain: Université Catholique de Louvain.

- Foray, D. (1989). Les modèles de compétition technologique. Une revue de la littérature. *Technological competition models: A literature review. Revue d'Economie Industrielle*, 48(1), 16–34. doi:10.3406/rei.1989.2244
- Foray, D. (2002). Complex Systems and Collective Adoption: The Role of Networks and Partnerships as an Endogenous Mechanism to Reduce Dynamic Transaction Costs in the Context of Systemic Innovations. In de la Mothe & Link (Eds.), *Networks, Alliances and Partnerships in the Innovation Process* (pp. 191–198). Boston, MA: Kluwer.
- Fréry, F., Lecocq, X., & Warnier, V. (2015). Competing With Ordinary Resources. *Sloan Management Review*, 3(56), 69–77.
- Gabel, D. (2016). Uber and the Persistence of Market Power. *Journal of Economic Issues*, 50(2), 527–534. doi:10.1080/00213624.2016.1179060
- Gawer, A., & Cusumano, M. A. (2002). *Platform Leadership: How Intel, Microsoft, and Cisco Drive Industry Innovation*. Boston: Harvard Business School Press.
- Gawer, A., & Cusumano, M. A. (2008). How Companies Become Platform Leaders. *MIT Sloan Management Review*, 49(2), 28–35.
- Geradin, D. (2015). *Uber and the Rule of Law: Should Spontaneous Liberalization Be Applauded or Criticized?* George Mason Law & Economics Research Paper N°15-53. Retrieved January 30, 2016 from <http://ssrn.com/abstract=2693683>
- Hagiu, A., & Wright, J. (2015). *Multi-Sided Platforms*. Harvard Business School. Working Paper 15-037.
- Hall, J., Kendrick, C., & Nosko, C. (2015). *The Effects of Uber's Surge Pricing: A Case Study*. Working Paper. Retrieved January 17, 2016 from http://faculty.chicagobooth.edu/chris.nosko/research/effects_of_uber's_surge_pricing.pdf
- Horpedahl, J. (2015). Ideology Uber Alles? Economics Bloggers on Uber, Lyft, and Other Transportation Network Companies. *Econ Journal Watch*, 12(3), 360–374.
- Karabell, Z. (2015, November 6). The Uberization of Money. *The Wall Street Journal*. Retrieved January 28, 2016 from <http://www.wsj.com/articles/the-uberization-of-finance-1446835102>
- Katz, M. L., & Shapiro, C. (1985). Network externalities, competition, and compatibility. *The American Economic Review*, 75(3), 424–440.
- MacDonald, C. (2016). Uber Is Built on Trust. *IEEE Technology and Society Magazine*, 35(2), 38–39. doi:10.1109/MTS.2016.2554440
- Manjoo, F. (2015, January 28). *Uber's Business Model Could Change Your Work*. Retrieved February 19, 2016 from http://www.nytimes.com/2015/01/29/technology/personaltech/uber-a-rising-business-model.html?_r=0
- Mason, P. (2015). *Postcapitalism. A Guide to Our Future*. London, UK: Penguin Books.
- McGrath, R. G. (2013). *The End of Competitive Advantage: How to Keep Your Strategy Moving as Fast as Your Business*. Boston: Harvard Business Review Press.
- Noskoy, C., & Tadelisz, S. (2015). *The Limits of Reputation in Platform Markets: An Empirical Analysis and Field Experiment*. NBER Working Paper N°20830. Retrieved September 21, 2016 from: <http://faculty.haas.berkeley.edu/stadelis/EPP.pdf>
- OECD. (2005). *Oslo Manual. Guidelines for collecting and interpreting innovation data*. Paris: OECD and Eurostat.
- Rahel, S. (2016). *Economics of the Taxi Industry: An Uber Shake-up*. Honors Theses AY 15/16. Paper 28. Retrieved September 16, 2016 from http://repository.uwyo.edu/honors_theses_15-16/28

Rochet, J. C., & Tirole, J. (2006). Two-Sided Markets: A Progress Report. *The Rand Journal of Economics*, 37(3), 645–666. doi:10.1111/j.1756-2171.2006.tb00036.x

Smith, W. (2016). The Uber-All Economy of the Future. *The Independent Review*, 20(3), 383–390.

Staykova, K., & Damsgaard, J. (2014). A Model of Digital Payment Infrastructure Formation and Development: The EU Regulator's Perspective. In *Proceedings of the 13th International Conference on Mobile Business, ICMB 2014* (vol. 38). Atlanta, GA: Association for Information Systems. AIS Electronic Library (AISeL).

Staykova, K., & Damsgaard, J. (2015). A Typology of Multi-sided Platforms: The Core and the Periphery. In *European Conference on Information Systems (ECIS) 2015 Proceedings*. Atlanta, GA: Association for Information Systems. AIS Electronic Library (AISeL).

Teece, D. (2010). Business Models, Business Strategy and Innovation. *Long Range Planning*, 43(2/3), 172–194. doi:10.1016/j.lrp.2009.07.003

Therin, F. (2015). *Of Uberization and Business schools (and toilet paper)*. EMLV, Management School Paris-La Défense. Retrieved September 30, 2015 from <http://www.emlv.fr/en/of-uberization-and-business-schools-and-toilet-paper>

Weyl, G. (2010). Price Theory of Multi-sided Platforms. *American Economic Review*, 100(4), 1642–1672.

ADDITIONAL READING

Cook, J. (2015). Uber's internal charts show how its driver-rating system actually works. *Business Insider*. Retrieved September 5, 2016 from: <http://www.businessinsider.com/leaked-charts-showhow-ubers-driver-rating-system-works-2015-2?r=UK&IR=T>

Cramer, K., & Krueger, A. (2016). *Disruptive Changes in the Taxi Business: The Case of Uber*. National Bureau of Economic Research. Retrieved February 19, 2016 from <http://www.nber.org/papers/w22083>

Evans, D. S., & Schmalensee, R. (2007). The Industrial Organization of Markets with Two-Sided Platforms. *Competition Policy International*, 3(1), 151–179.

Goel, S. (2014). *Capitalism Versus the Sharing Economy*. (Unpublished working paper). College Writing Spring 2014. Retrieved March 2, 2016 from <http://escholarship.org/uc/item/8xd4m19w>

Rochet, J. C., & Tirole, J. (2003). Platform Competition in Two-Sided Market. *Journal of the European Economic Association*, 1(4), 990–1029. doi:10.1162/154247603322493212

KEY TERMS AND DEFINITIONS

Business Ecosystem: Interorganizational network (coalition) which brings together actively involved organizations who belong to different sectors, but share the same interests, values and common goals.

Disruptive Technology: Disruptive technology or innovation create major changes in industry frontiers, business processes and business models.

Network Effects: They are related to correlation between the number of users of a network for a specific good/service and the utility of this good/service.

On-Demand Economy: Refers more and more to uberization closely related also to sharing practices spreading across several sectors as banks, insurance, tourism, transport, food delivery, connected health or retail.

Online Platform (Company): Integrated web-based platform company allowing people (consumers, “independent contractors”) who want to share assets online.

Peer-to-Peer Platform: Platform whereby two individuals can communicate, interact (and sell) directly with each other, without intermediation by a third-party.

Sharing Economy: known also as collaborative consumption referring to peer-to-peer-based sharing of access to goods and services.

Uber: The name of a technology company who has launched a smartphone-based app allowing car-sharing service (connecting ride-seekers and UberX drivers). Uber's development has inspired the term 'uberization'.

E

Category E

Educational Technologies

Adaptive Hypermedia in Education

E**Vehbi Turel***The University of Bingol, Turkey*

INTRODUCTION

Since the definitions of the pertinent terms such as multimedia, multiple media, interactive multimedia and hypermedia are given in a previous study (Turel, 2015a, pp. 2495-96), in this article, only the definition of adaptive hypermedia (AH) is focussed on, and the role of adaptive hypermedia (AH) in education is concentrated on. The aim of this article is: (a) to give the definition of AH and state what AH means, and (b) to explore the role of AH in education at this digital age, in which the majority of learners are generally digitally fluent and competitive (Turel 2015b, Gros et al., 2012, pp. 190-210) although some claims otherwise (Bullen, Morgan & Qayyum, 2011, pp. 1-24). Pedagogically and epistemologically, educational institutions (i.e. nursery, primary, secondary and high schools, colleges, vocational schools and colleges, life-long learning centres, adult education centres, and universities) should respond to such learning demands and differences to accommodate the digital-literate, wise and efficient learning style preferences of today's learners by providing AH learning materials for them. More frankly, educational institutions have to use and provide AH learning materials for their learners in order to be competitive in this digital age (Turel, 2014a; Turel, 2013; Duncan-Howell, 2012).

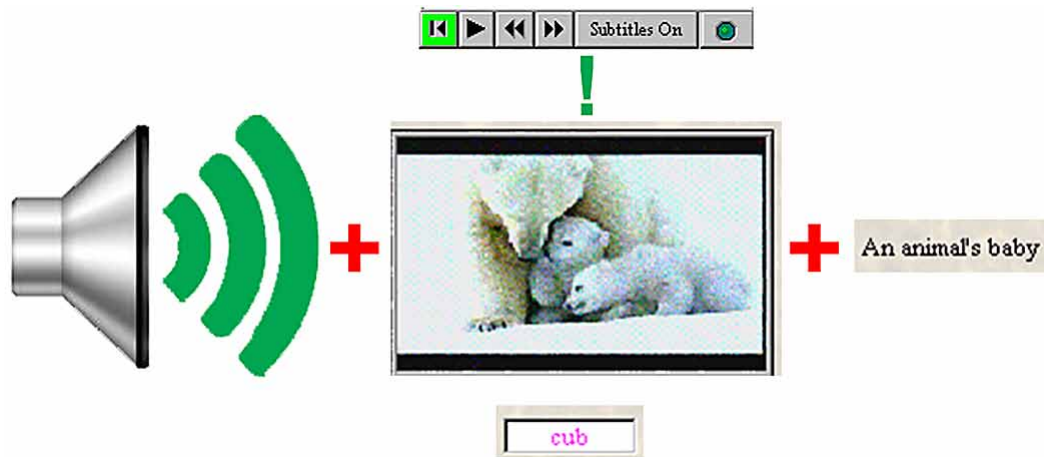
BACKGROUND

When educational (computer) technologists speak of adaptive hypermedia (AH), mostly one thing comes to mind. It is the use, combination and delivery of digital video, audio/sound, text,

visuals (i.e. pictures/images/photographs, graphics, tables, figures), animations, hyperlinks, optimum combinations, instructions etc. on the same digital platform, which are totally computerised and under computer as well as learners' control. This digital platform also enables learners / users to make preferences, record these preferences, their individual needs and learning goals, and then uses them throughout interaction with the learners in order to meet their personal needs so that they can learn better (Turel, 2015a, p. 2497; Turel 2015b). In other words, AH (a) is a digital environment where a wide range of digital elements are combined and delivered on the same environment through hyperlinks (Figures 1 and 2), (b) has a learner model where learners can make preferences, record these preferences, their individual needs and their learning goals, and (c) uses the learner model to adapt the contents of the hypermedia according to the learners' needs (based on the data provided and the preferences made by the learners through the learner model) (Brusilovsky 2012, p. 46; Brusilovsky, 2007, Brusilovsky & Millán, 2007; Brusilovsky, Eklund & Schwarz, 1998). It is because of this 'adaptation feature' that it is now called 'adaptive hypermedia' (AH). AH is relatively a new direction in the field of educational technology (Brusilovsky 2012, p. 46), consists of different models (Kahraman et al. 2013, p. 60) and can be classified according to its application areas such as Educational Adaptive Hypermedia, which is the most popular area for research (Brusilovsky, 1996). To sum up, when the combination and delivery of a wide range of digital elements on the same digital platform offers 'personalised learning', then such a digital platform is called AH.

DOI: 10.4018/978-1-5225-2255-3.ch205

*Figure 1. An example of hypermedia where optimum combinations (i.e. text, video, audio, visuals, control buttons, text boxes etc.) can be provided on the same digital platform
Turel, 2015b, p. XXXIII.*



*Figure 2. An example of hypermedia as well as hyperlinks. The “word in red colour” (i.e. abide by) contains a hyperlink. When it is clicked on, the provided definition is displayed for the learners
Turel, 2015b, p. XXXVIII.*



ADAPTIVE HYPERMEDIA IN EDUCATION

In terms of education, outstanding differences between AH and conventional materials (CMs) exist. These are: (a) the combination and instantaneous delivery of different digital elements on the same digital platform, (b) being navigational and interactive, (c) user control and ease of use

and (d) offering ‘personalised learning’. The objective of educational AH is to design and create pedagogically sound and epistemologically flexible learning environments which (a) not only supports a wide range of learners who are diverse in terms of abilities, disabilities, levels, interests, backgrounds, and other characteristics, but also (b) enables learners to make maximum use of the available personalised interactive and

simultaneous learning material (Shute & Zapata-Rivera, 2012; Brusilovsky, 1996), which is also cost-effective. We can now look at these as well as the other aspects of educational AH.

Adaptive Hypermedia Has More Features

AH has a wide range of combined flexible elements and features. These can make (a) input comprehensible, (b) learning enjoyable (Soboleva & Tronenko 2002, p. 493; Trinder, 2002, p. 75) and (c) adapt the same learning material to the needs of a wide range of diverse learners (Shute & Zapata-Rivera, 2012; Brusilovsky, 1996). All of these can, as a whole, result in better learning (K Govindasamy, 2013; Stepp-Greany 2002, p. 172). On the other hand, CMs (i.e. traditional books, off-air materials, tape cassettes and videotapes) do not have many such elements. For example, traditional books only feature texts, pictures/graphics/figures, indexes and inter-textual citations although 'Choose Your Own Adventures books' series (e.g. *The Cave of Time*, *House of Danger*) are also explicitly designed as 'hyper-textual'. In the same way, off-air radio-programs and tape cassettes feature 'audio-only', while off-air TV programmes and videotapes feature both sound and visuals. Those created for learning/teaching purposes such as tape cassettes and videotapes are sometimes accompanied with learner's books. Similarly, traditional hypermedia (i.e. multimedia, interactive multimedia) has its own limitations. For example, traditional hypermedia provides all learners with the same contents and hyperlinks regardless of their wide range of profiles, needs and individual differences.

Not only is learning a multi-channel phenomenon, but also learning-style preferences are a wide range of - which can be described as learners' natural, habitual and preferred (i.e. biological, developmental and conscious or subconscious) ways of processing, acquiring and representing (new) knowledge and skills -. While CMs provide a limited learning environment, AH provides a

multidimensional, multi-sensory and interactive environment in which not only input can be presented in different ways (Herron et al., 2002, p. 37; Leffa, 1992, p. 66), but also input can be adapted to the learners' individual characteristics and needs. Not only can this meet the needs of learners who are diverse in terms of abilities, disabilities, levels, interests, backgrounds, and other characteristics, but this can also meet the needs of different individual learners who have different learning-style preferences such as auditory, visual, tactile, kinaesthetic, group and individual (Carson & Longhini, 2002, p. 408). For instance, while the provision of audio-only meets the needs of auditory-learners, video and (supplementary) visuals can meet the needs of visual-learners (Turel, 2014b; McLoughlin, 1999, pp. 222, 229). Equally, typing a word or sentence, clicking on a choice, recording their own voice, or dragging and dropping a word, sentence or picture meets the needs of tactile- and kinaesthetic- learners.

AH features the combination of diverse elements which provides interactivity. This facilitates the negotiation of meaning, and it is necessary in learning (Harrington & Levy, 2001, p. 15; Hegelheimer & Chapelle, 2000, p. 42), as learners can access hyperlinks, glossaries, feedback, subtitles, answers, and optimum flexible, customised and conditional combinations (Turel, 2015a; Turel, 2012, pp. 40-41) immediately, and find out what and why they have not understood, and the underlying assumptions.

All of these efficient features of AH facilitate learning; draw learners' attention to their weaknesses and certain aspects of input and can result in depth processing, which are necessary conditions for learning. For example, the combination of different elements can make input comprehensible. This is a requirement of the comprehension input hypothesis and theory, which state that learners acquire only when they understand, and therefore "consider intake synonymous with comprehensible input" (Krashen; 1984; p. 21; Tschirner, 2001, p. 311). It is also a requirement of the social learning theory, which "posits that

people's judgements about their potential ability to perform well or to cope in a situation actually affect their efforts..." (Robinson, 1991, p. 157). The person perception theory, which requires us to avoid focussing learners' attention on their weaknesses so that they do not develop negative judgements about their ability to perform (ibid: 157), and the social-psychological theory and the socio-educational model, which focus on the role of attitudes and motivation in learning (Masgoret & Gardner, 2003, pp. 124, 127, 158-9; Gardner, 1985, p. 158) also require us to provide comprehensible input for learners. Comprehensible input can also be considered as a requirement of the cognitive load theory and working memory, as incomprehensible input can cause risk of overload on limited working memory (Kalyuga, 2000, p. 161; Sweller, 1999).

The combination of optimum different elements on the same digital environment in AH can draw learners' attention to certain salient input. This is a requirement of the noticing hypothesis and the attention theory, which suggest that paying/drawing attention to (specific) forms in the input is necessary for learning (Nicholas, Lightbown & Spada, 2001, p. 721; Williams, 2001, p. 335). This can be implemented more effectively especially in AH environments, as the contents can be adapted to the needs of different individuals who are diverse.

The combination of optimum different elements on the same digital platform in AH is also the requirement of the dual-coding theory and the generative theory of multimedia, as it provides learners with more than one concurrent mode/element which targets to teach one thing. It is learned better than those coded only in one mode because dual-code provides more paths of recall, which can aid to build two types of recall cues in memory (Ginther, 2002, pp. 133-67; Moreno & Mayer, 2002) although this might not always be the case with all combinations (Amrhein et al., 2002, pp. 843-57; Kalyuga, 2000, pp. 162-63), as it also requires considering the requirements of the cognitive load theory and working memory.

All Elements Are on the Same Platform

While working with CMs requires learners to use additional sources (i.e. accompanying auxiliary books, answer keys, dictionaries, grammar books etc.), which might be frustrating, distracting and time-consuming for learners (Leffa 1992, pp. 70-72), in AH, all elements are on the same digital platform and learners find such a combination and delivery motivating (Soboleva & Tronenko, 2002, p. 483; Ayres, 2002, pp. 247-48). As a result, AH provides considerably greater opportunities for ease of use and learner-control than have been possible with CMs. For example, Leffa's (1992, p. 71) study revealed that "the electronic glossary... was not only more efficient but also seemed to demand no previous training..." and "...was significantly superior to the traditional dictionary"; as less time needed for the electronic mode (ibid, p. 72) and it resulted in better comprehension (ibid, p. 70). Likewise, AH is more efficient as well as more cost-effective in comparison to traditional hypermedia, as the former adapts the learning contents to the needs of different learners who are diverse in terms of abilities, disabilities, levels, interests, backgrounds, and other characteristics.

Learners can instantly access different parts of AH (i.e. video, audio, optimum combinations, feedback, glossaries, available sections, syntax, lexis, subtitles etc.), which are combined and delivered on the same digital platform, all of which are good opportunities to exercise control over input (Tschimer, 2001, pp. 312-3) unlike with CMs.

These positive aspects of AH can enable learners (a) to find out the difficulties, the right solutions, what the rules are; (b) to analyse the mistakes that have been made; and (c) to find out why they have made such mistakes by assessing their answers, recording and scoring them, pointing out and explaining mistakes (Mangiafico, 1996, p. 52). This capability provides more real-world learning contexts and more authentic and interactive tasks. Ashward (1996, p. 80) states that "the ability to display such a variety of resources,

to link them together, and to combine all these resources with tutorial programs on the computer provide a highly sophisticated, yet potentially easy-to-use and easy-to-author medium for developing education materials in any subject". Equally, these positive aspects are highlighted by Fox (1995, p. 43), and learners also appreciate them (Brett, 1997, pp. 46-7).

Adaptive Hypermedia Gives Learners Control

AH provides learners with the opportunity to learn any subject at their own pace, without fear of making mistakes in the presence of a teacher or other learners. It is a tension free environment in which learners can use AH environments individually, as no teachers and friends are present. Especially, during self-study, AH is a private and flexible workplace where learners can take risks; work in their own place (Tschirner, 2001, p. 307), in their own time, at the pace they need, and in the way they enjoy; as it gives them the control (Soboleva & Tronenko 2002, p. 493; Trinder, 2002, p. 75) and provides them with different choices, tasks and feedback. Moreover, the contents of AH are adapted to the needs of individual learners. All these positive aspects can make them feel more comfortable and might result in promoting development of self-confidence and provoking working hard. It is due to these reasons that learners do not complain about the fear of making mistakes. Conversely, they express their comfort of working with hypermedia. Ease of use - in relation to pacing and control- (Trinder, 2002, p. 75; Brett, 1997, pp. 46-7) and "user-friendliness" - at an operational level - (Fleta et al., 1999, p. 55) are some hypermedia features found most useful. User-control is a feature that has an impact on learning effectiveness (Peter, 1994, pp. 157-8), which is also considered an essential principle in instructional design (Hoven, 1999, p. 91) and a condition that is necessary in autonomous materials (Watts, 1997, p. 7). It is also a requirement of autonomous learning and development, the main

way of life-long learning, and the requirement of the autonomous learning theory, which "demands that learners take control of their learning" (Voller, 1997, p. 106).

Adaptive Hypermedia Motivates Learners

As AH offers learners many elements on the same digital environment, it is highly motivating. When learners make mistakes, this does not even demotivate them because they have the opportunity to receive instant and meaningful feedback, and practise as many times as they want and need to. Learners find hypermedia motivating and they are engaged, active, attentive and interested in hypermedia when they work with it (Soboleva & Tronenko, 2002, p. 483; Lyall & McNamara, 2000, p. 135). This is an important factor in learning especially during autonomous study because "how to engage the interest of the learner and so sustain his motivation to learn" is considered "a related problem for the writer of self-study materials" (Frankel, 1987, p. 53).

Most importantly, (1) motivation is "directly linked to achievement" according to the socio-educational model' (Masgoret & Gardner, 2003, p. 129), which suggests that integrativeness and attitudes towards the learning situation cause learners' motivation to learn input and motivation is responsible for achievement (ibid: 24; Gardner, 1985, p. 158). (2) Motivation is also common to all models of learning (i.e. the acculturation model, the conscious reinforcement model, the intergroup model, the interactionist model, the LMR-plus Model, the monitor model, the social context model, the social psychological model, the elaboration theory model, ibid: 25-66).

Adaptive Hypermedia Prepares, Directs, and Guides Learners

Since AH (a) enables the combination and delivery of different elements on the same digital platform more effectively (Turel & Kılıc, 2014; Turel 2012,

pp. 35, 45) such as supplying a non-linear editing facility (Tschirner, 2001, p. 307) and (b) also adapts the contents to the needs of individual learners, it enables materials writers to create effective and interactive applications directly to their customers' specifications, their needs, interests and learning styles (Cauldwell, 1996, p. 526). AH can prepare learners more effectively and adequately for input, as it empowers materials writers (1) to provide learners with different elements such as unknown items, technical words and concepts (i.e. terminology), the objective of the topic, syntax, special features of text type/input, short audio messages, short video clips, sample sentences, graphics, animations, visuals, simplified written versions of the text; short information about the topics, the speakers of the existing audio and video clips, their roles, how they interact, the content and cultural differences (Turel, 2014c). (2) Since featuring, delivering and combining different elements on the same digital platform, AH can guide learners more effectively. For example, when learners make a mistake, they receive adapted instant and meaningful feedback. This can enable them to find out why they have made the particular mistakes, how they can overcome such difficulties in future occasions, and improve new strategies. Due to these reasons, AH is considered efficient for self-study use (Soboleva & Tronenko, 2002, p. 483). Additionally, if learners are instructed about which strategies they need to follow in which situations, they can be directed and guided (Debski & Gruba, 1999, pp. 219-20). Similarly, learners' attention can be directed to metacognitive strategies such as planning, directing attention, self-monitoring, self-evaluation and similar strategies which are effective for exploitation of the facility itself. This is important because learners can be taught to "use appropriate comprehension strategies" (Goh, 2000, p. 71) and it is mostly needed during self-study, as learners are by themselves (Debski & Gruba, 1999, pp. 219-20).

AH can qualify learners to overcome potential sources of difficulties in the context of learning

any subject. For example, while learning a foreign language, unfamiliar items, proper names, syntax, fast speech, and unfamiliar accents can be taught more efficiently (Tschirner; 2001; p. 306). For instance, the meanings of unfamiliar items can be provided through hyperlinks or glossaries. These can be explained in through synonyms or antonyms. Likewise, learners' attention can be drawn to cognates, false cognates and polysemous words, which are useful for vocabulary acquisition and helpful for understanding, and improving listening and other skills (Vidal, 2003, p. 80). Additionally, simple sentences and short paragraphs featuring unknown items can be provided. In some cases, they can be explained throughout visuals, audio, video or optimum combinations. Their equivalents in first language can be given if learners are monolingual. Unfamiliar proper names can also be given in advance so that learners will not have difficulty. Such names can be given through pictures/cartoons that have labels and instructions. Similarly, cultural differences, syntax and lexis can be illustrated through simple samples, pictures, audio or video clips (Turel, 2014c, Turel & Kılıc, 2014).

Equally, while studying a different culture, religion, geography, science or any other topics, the pertinent input and relevant features such as cultural difficulties, costumes, traditions, figures, shapes and the assumptions underlying them can be presented and taught more efficiently through providing customised/adapted and combined comprehensible input and interactive tasks with instantaneous feedback.

Adaptive Hypermedia Provides Efficient and Instant Feedback

In CMs, feedback is normally given in learner's books or answer-key books. When learners have difficulty, they can access them. Although this is useful, it is very limited, as (a) it is not instant and (b) it consists of restricted elements such as text, pictures and graphics. In AH, feedback

is (a) immediate, (b) can consist of different elements, which meet both learners' visual and acoustic needs resulting in learning and (c) can be conditional/adapted. Such feedback can help learners to find out what and why they could not understand and overcome the difficulties causing them not to comprehend, which facilitates meaning negotiation (Williams, 2001, p. 337; Smith, 2003, pp. 39-40), draws attention and raises conscious/metacognitive awareness. This can guide and lead learners to develop new and effective strategies, which is one of the targets that material writers want to, and need to, fulfil especially in autonomous materials. This also draws learners' attention to the input, which is necessary for learning (Nicholas, Lightbown & Spada, 2001, p. 721; Williams, 2001, p. 335).

Adaptive Hypermedia Meets the Needs of Different Learners

It is a known fact that all learners do not have the same background and abilities. While some learners have high abilities and know quite a lot about target input/topic, other learners may not. It is also a de facto that not only do AH enable materials writers to make use of different elements, which can make input comprehensible and create gradual and different tasks, but it also enable material writers to adapt the learning contents to the needs of different learners who are diverse in terms of abilities, disabilities, levels, interests, backgrounds, and other characteristics. As a result, different learners can find what is most appropriate for them or a way of working which is most convenient for them. For example, learners with high ability can be provided with higher-level exercises and tasks, while those with low ability can be provided with lower-level exercises and tasks. Similarly, if some learners find audio/video clips too fast, then they can slow down the clips or alternatively they can be provided with the slow versions.

FUTURE RESEARCH DIRECTIONS

E

In spite of the above mentioned advantages of AH, many hypermedia applications on the market are not adequate. They contain certain features that are neither pedagogically nor psychologically sophisticated (Türel 2014; Turel, 2010, p. 399; Trinder, 2002, pp. 69-84). The problems fundamentally stem from hypermedia developers, not the pertinent technology itself (Turel & McKenna, 2012, pp.188-189) although technology has its own limitations, as well.

Therefore, AH applications need to be better developed. The pertinent findings in the field need to and ought to be fully implemented. For example, many AH applications do not feature proper self-assessment tests, although self-assessment tests are an essential element for learning especially during self-study. Similarly, AH applications need to give flexibility to learners. To overcome such shortages, we need to have more comprehensive and flexible AH evaluation criteria so that AH applications can be evaluated based on these criteria although some useful studies have been conducted to this end (Elissavet & Economides 2003; Ferney & Waller 2001). We also need to have enhanced AH development models.

AH should be fully integrated into education alongside CMs. It is striking that hypermedia is still not widely used (Turel, Calik, & Doganer, 2015a, Turel, Calik, & Doganer, 2015b, Turel, 2014a). Priority should be given to young learners, as they have more positive perceptions towards such technology (Türel, 2014). AH developers should further develop fully efficient AH applications in all subjects for all-level learners.

Future investigations need to develop better AH models, find out how to design each element as well as a whole more precisely, and improve efficient, comprehensive and flexible evaluation criteria. These will eventually result as a whole in creating ideal AH environments that facilitate learning.

CONCLUSION

Although AH has a vital role in learning, this is not the only issue that we need to be aware of. To ensure that AH plays its role in learning in true sense, efficient and fully adapted hypermedia environments need to be designed and created. Not only does this require the selection of the right input (i.e. in terms of age, level, contents and acquisition), but it also requires the adapted effective design of every single available element. These can only be achieved in true sense by the active and full participation of the pertinent experts (Turel & McKenna, 2013, pp. 189-190). Only such participation ensures that cost-effective AH can be created. These are very important and all have their role in the effectiveness of AH in education, the final aim of which is to enable learners to learn and acquire the necessary 21st century knowledge, skills and competencies.

REFERENCES

- Amrhein, P. C., McDaniel, M. A., & Waddill, P. (2002). Revisiting the picture-superiority effect in symbolic comparisons: Do pictures provide privileged access? *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 28(5), 843–857. doi:10.1037/0278-7393.28.5.843 PMID:12219794
- Ayres, R. (2002). Learner attitudes towards the use of CALL. *Computer Assisted Language Learning*, 15(3), 241–249. doi:10.1076/call.15.3.241.8189
- Brett, P. (1997). A comparative study of the effects of the use of multimedia on listening comprehension. *System*, 25(1), 39–53. doi:10.1016/S0346-251X(96)00059-0
- Brusilovsky, P. (1996). Methods and techniques of adaptive hypermedia. *User Modeling and User-Adapted Interaction*, 6(2-3), 87–129. doi:10.1007/BF00143964
- Brusilovsky, P. (2007). Adaptive Navigation Support. In P. Brusilovski, A. Kobsa, & W. Nejdl (Eds.), *The Adaptive Web: Methods and Strategies of Web Personalization* (Vol. 4321, pp. 263–290). Berlin: Springer-Verlag. doi:10.1007/978-3-540-72079-9_8
- Brusilovsky, P. (2012). Adaptive Hypermedia for Education and Training. In P. J. Durlach & A. M. Lesgold (Eds.), *Adaptive Technologies for Training and Education* (pp. 46–68). Cambridge, UK: CUP. doi:10.1017/CBO9781139049580.006
- Brusilovsky, P., Eklund, J., & Schwarz, E. (1998). Web-based education for all: A tool for development adaptive courseware. *Computer Networks and ISDN Systems*, 30(1), 291–300. doi:10.1016/S0169-7552(98)00082-8
- Brusilovsky, P., & Millán, E. (2007). User models for adaptive hypermedia and adaptive educational systems. In P. Brusilovski, A. Kobsa, & W. Nejdl (Eds.), *The Adaptive Web: Methods and Strategies of Web Personalization* (Vol. 4321, pp. 3–53). Berlin: Springer-Verlag. doi:10.1007/978-3-540-72079-9_1
- Bullen, M., Morgan, T., & Qayyum, A. (2011). Digital learners in higher education: Generation is not the issue. *Canadian Journal of Learning and Technology*, 37(1), 1–24.
- Carson, J. G., & Longhini, A. (2002). Focussing on learning styles and strategies: A diary study in an immersion setting. *Language Learning*, 52(2), 401–438. doi:10.1111/0023-8333.00188
- Cauldwell, R. T. (1996). Direct encounters with fast speech on CD-AUDIO to teach listening. *System*, 24(4), 521–528. doi:10.1016/S0346-251X(96)00046-2
- Debski, R., & Gruba, P. (1999). A qualitative survey of tertiary instructor attitudes towards project-based CALL. *Computer Assisted Language Learning*, 12(3), 219–239. doi:10.1076/call.12.3.219.5715

- Duncan-Howell, J. (2012). Digital mismatch: Expectations and realities of digital competency amongst pre-service education students. *Australasian Journal of Educational Technology*, 28(5), 827–840. doi:10.14742/ajet.819
- Elissavet, G., & Economides, A. A. (2003). An Evaluation Instrument for Hypermedia Courseware. *Journal of Educational Technology & Society*, 6(2), 31–44.
- Ferney, D., & Waller, S. (2001). Reflections on multimedia design criteria for the international language learning community. *Computer Assisted Language Learning*, 14(2), 145–168. doi:10.1076/call.14.2.145.5781
- Fleta, B. M., Sabater, C. P., Salom, L. G., Guillot, C. P., Monreal, C. S., & Turney, E. (1999). Evaluating multimedia programs for language learning: A case study. *ReCALL*, 11(3), 50–57.
- Fox, J. (1995). Multimedia for language learning: Some Course Design Issues. *Computer Assisted Language Learning*, 8(1), 31–44. doi:10.1080/0958822950080103
- Fox, J., Romano-Hvid, B. C., & Sheffield, J. S. (1992). *New perspectives in Modern Language Learning*. Sheffield, UK: Group UK, Employment Department.
- Frankel, F. (1987). Self-study materials: Involving the learner. In M. Geddes & G. Sturtridge (Eds.), *Individualisation* (pp. 52–60). Hong Kong, China: Modern English Publications Ltd.
- Gardner, R. C. (1985). *Social psychology and second language learning: The role of attitudes and motivation*. Baltimore, MD: Edward Arnold.
- Ginther, A. (2002). Context and Content Visuals and Performance on Listening Comprehension Stimuli. *Language Testing*, 19(2), 133–167. doi:10.1191/0265532202lt225oa
- Goh, C. C. M. (2000). A cognitive perspective on language learners listening comprehension problems. *System*, 28(1), 55–75. doi:10.1016/S0346-251X(99)00060-3
- Govindasamy, K. M. (2013). Embodied agent in tutor role: Effects on field dependent and independent low achievers' retention and perceived science self efficacy beliefs. *Journal of Educational Multimedia and Hypermedia*, 22(3), 273–297.
- Gros, B., Garcia, I., & Escofet, A. (2012). Beyond the Net Generation Debate: A Comparison of Digital Learners in Face-to-Face and Virtual Universities. *International Review of Research in Open and Distance Learning*, 13(4), 190–210. doi:10.19173/irrodl.v13i4.1305
- Harrington, M., & Levy, M. (2001). CALL begins with a C: Interaction in computer-mediated language learning. *System*, 29(1), 15–26. doi:10.1016/S0346-251X(00)00043-9
- Hegelheimer, V., & Chapelle, C. A. (2000). Methodological Issues in Research on Learner-Computer Interactions in CALL. *Language Learning & Technology*, 4(1), 41–59.
- Herron, C., Dubreil, S., Corrie, C., & Cole, S. P. (2002). A classroom Investigation: Can video improve intermediate-level French language students ability to learn about a foreign culture? *Modern Language Journal*, 86(1), 36–53. doi:10.1111/1540-4781.00135
- Hoven, D. (1999). A model for listening and viewing comprehension in multimedia environments. *Language Learning & Technology*, 3(1), 88–103.
- Kahraman, H. T., Sagioglu, S., & Colak, I. (2013). A novel model for web-based adaptive educational hypermedia systems: SAHM (supervised adaptive hypermedia model). *Computer Applications in Engineering Education*, 21(1), 60–74. doi:10.1002/cae.20451
- Kalyuga, S. (2000). When using sound with a text or picture is not beneficial for learning. *Australasian Journal of Educational Technology*, 16(2), 161–172. doi:10.14742/ajet.1829
- Krashen, S. D. (1984). *Principles and practice in second language acquisition*. Oxford, UK: Pergamon.

- Leffa, V. J. (1992). Making foreign language texts comprehensible for beginners: An experiment with an electronic glossary. *System*, 20(1), 63–73. doi:10.1016/0346-251X(92)90008-Q
- Lyall, R., & McNamara, S. (2000). Learning tool or potplant stand? Students opinions of learning from a CAL program in a distance education context. *Australasian Journal of Educational Technology*, 16(2), 126–146. doi:10.14742/ajet.1827
- Mangiafico, L. F. (1996). *The relative effects of classroom demonstration and individual use of interactive multimedia on second listening comprehension* (Unpublished Ph.D. Thesis). Vanderbilt University, Nashville, TN.
- Masgoret, A. M., & Gardner, R. C. (2003). Attitudes, motivation, and second language learning: A meta-analysis of studies conducted by Gardner and associates. *Language Learning*, 53(1), 123–163. doi:10.1111/1467-9922.00212
- McLoughlin, C. (1999). The implications of the research literature on learning styles for the design of instructional material. *Australasian Journal of Educational Technology*, 15(3), 222–241. doi:10.14742/ajet.1859
- Moreno, R., & Mayer, R. E. (2002). Verbal redundancy in multimedia Learning: When reading helps listening. *Journal of Educational Psychology*, 94(1), 156–163. doi:10.1037/0022-0663.94.1.156
- Nicholas, H., Lightbown, P. M., & Spada, N. (2001). Recasts as feedback to language learners. *Language Learning*, 51(4), 719–758. doi:10.1111/0023-8333.00172
- Peter, M. (1994). *Investigating into the design of educational multimedia: Video, interactivity and narrative* (Unpublished Ph.D. Thesis). Open University, London, UK.
- Robinson, G. L. (1991). Effective feedback strategies in CALL: Learning theory and empirical research. In P. Dunkel (Ed.), *Computer-Assisted Language Learning and Testing: Research Issues and Practice* (pp. 155–167). New York, NY: Newbury House.
- Shute, V. J., & Zapata-Rivera, D. (2012). Adaptive educational systems. In J. D. Paula & M. L. Alan (Eds.), *Adaptive technologies for training and education* (pp. 7–27). Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9781139049580.004
- Smith, B. (2003). Computer-mediated negotiated interaction: An expanded model. *Modern Language Journal*, 87(1), 38–57. doi:10.1111/1540-4781.00177
- Soboleva, O., & Tronenko, N. (2002). A Russian multimedia learning package for classroom use and self-study. *Computer Assisted Language Learning*, 15(5), 483–499. doi:10.1076/call.15.5.483.13470
- Stenton, T. (1998). Hypermedia: a new consensus for the 1990's. In K. Cameron (Ed.), *Multimedia CALL: Theory and Practice* (pp. 11–16). Exeter, Elm Bank Publications.
- Stepp-Greany, J. (2002). Student Perceptions on Language Learning in a Technological Environment: Implications for the New Millennium. *Language Learning & Technology*, 6(1), 165–180.
- Sweller, J. (1999). *Instructional design*. Melbourne, Australia: ACER.
- Trinder, R. (2002). Forum: Multimedia in the business English classroom: The learners point of view. *Computer Assisted Language Learning*, 15(1), 69–84. doi:10.1076/call.15.1.69.7291
- Tschirner, E. (2001). Language acquisition in the classroom: The role of digital video. *Computer Assisted Language Learning*, 14(3-4), 305–319. doi:10.1076/call.14.3.305.5796

- Turel, V. (2010). Advanced Turkish. *ReCALL*, 22(3), 396–401. doi:10.1017/S0958344010000224
- Türel, V. (2012). Design of feedback in interactive multimedia language learning environments. *Linguistik Online*, 54.
- Türel, V. (2013). The use of educational technology at tertiary level. *Hacettepe University Journal of Education*, 28(2), 482–496.
- Türel, V. (2014). Learners' perceptions towards interactive multimedia environments. *Hacettepe University Journal of Education*, 29(3), 167–183.
- Turel, V. (2014a). Teachers computer self-efficacy and their use of educational technology. [TOJDE]. *Turkish Online Journal of Distance Education*, 15(4), 130–149. doi:10.17718/tojde.81990
- Turel, V. (2014b). The use and design of supplementary visuals for the enhancement of listening skills in hypermedia. In C. M. Akrivopoulou & N. Garipidis (Eds.), *Human rights and the impact of ICT in the public sphere: Participation, democracy, and political autonomy* (pp. 268–291). Hershey, PA: IGI-Global; doi:10.4018/978-1-4666-6248-3.ch016
- Turel, V. (2014c). Design of cultural differences in hypermedia environments. *International Journal of Human Rights and Constitutional Studies*, 2(2), 150–170. doi:10.1504/IJHRCS.2014.062766
- Turel, V. (2015a). Hypermedia and its' role in learning. In M. Khosrow-Pour (Ed.), *Encyclopedia of information science and technology* (3rd ed.). Hershey, PA: IGI-Global. doi:10.4018/978-1-4666-5888-2.ch243
- Turel, V. (2015b). Listening, multimedia and optimum design. In V. Turel (Ed.), *Intelligent Design of Interactive Multimedia Listening Software* (pp. xx–xxxiiv). Hershey, PA: IGI-Global. doi:10.4018/978-1-4666-8499-7.ch003
- Turel, V., & Kılıc, E. (2014). *The Inclusion and Design of Cultural Differences in Interactive Multimedia Environments*. In *Human Rights and the Impact of ICT in the Public Sphere: Participation, Democracy, and Political Autonomy*. Hershey, PA: IGI-Global.
- Turel, V., & McKenna, P. (2013). Design of Language Learning Software. In B. Zou et al. (Eds.), *Computer-Assisted Foreign Language Teaching and Learning: Technological Advances* (pp. 188–209). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-2821-2.ch011
- Vidal, K. (2003). Academic listening: A source of vocabulary acquisition? *Applied Linguistics*, 24(1), 56–89. doi:10.1093/applin/24.1.56
- Voller, P. (1997). Does the teacher have a role in autonomous language learning? In P. Benson & P. Voller (Eds.), *Autonomy & Independence in Language Learning*. New York: Addison Wesley Longman Ltd.
- Watts, N. (1997). A learner-based design model for interactive multimedia language learning packages. *System*, 25(1), 1–8. doi:10.1016/S0346-251X(96)00056-5
- Williams, J. (2001). The effectiveness of spontaneous attention to form. *System*, 29(3), 325–340. doi:10.1016/S0346-251X(01)00022-7

ADDITIONAL READING

- Bailey, C., Hall, W., Millard, D. E., & Weal, M. J. (2007). Adaptive hypermedia through contextualized open hypermedia structures. *Journal ACM Transactions on Information Systems*, 25(4), 1–36.
- Cristea, A. (2005). Authoring of adaptive hypermedia. *Journal of Educational Technology & Society*, 8(3), 6–8.

Ewais, A., & Troyer, O. D. (2013). Usability evaluation of an adaptive 3D virtual learning environment. *International Journal of Virtual and Personal Learning Environments*, 4(2), 16–31. doi:10.4018/jvple.2013010102

Jones, L. C. (2015). Effects of annotations on inferring meaning within a listening comprehension environment. In V. Turel (Ed.), *Intelligent design of interactive multimedia listening software* (pp. 1–26). Hershey, PA: Information Science Reference; doi:10.4018/978-1-4666-8499-7.ch001

Kang, T. (2015). The effectiveness of multiple media tools in L2 listening: A meta-analysis. In V. Turel (Ed.), *Intelligent Design of Interactive Multimedia Listening Software* (pp. 246–275). Hershey, PA: Information Science Reference; doi:10.4018/978-1-4666-8499-7.ch010

Knutov, E., De Bra, P., & Pechenizkiy, M. (2009). AH 12 years later: A comprehensive survey of adaptive hypermedia methods and techniques. *New Review of Hypermedia and Multimedia*, 15(1), 5–38. doi:10.1080/13614560902801608

Rada, R. (1995). Hypertext, multimedia and hypermedia. *New Review of Hypermedia and Multimedia*, 1(1), 1–21. doi:10.1080/13614569508914658

Retalis, R., & Papasalouros, A. (2005). Designing and generating Educational Adaptive Hypermedia Applications. *Journal of Educational Technology & Society*, 8(3), 26–35.

Shapiro, A. M. (2008). Hypermedia design as learner scaffolding. *Educational Technology Research and Development*, 56(1), 29–44. doi:10.1007/s11423-007-9063-4

Turel, V., & McKenna, P. (2015). Design of multimedia listening software: Instructions, tasks, texts, and self-assessment tests. In V. Turel (Ed.), *Intelligent design of interactive multimedia listening software* (pp. 142–169). Hershey, PA: Information Science Reference; doi:10.4018/978-1-4666-8499-7.ch006

Wang, V. C. X., Russo, M. R., & Dennett, S. (2013). Electronic education and lifelong learning. [IJAVET]. *International Journal of Adult Vocational Education and Technology*, 4(1), 46–60. doi:10.4018/javet.2013010104

KEY TERMS AND DEFINITIONS

Adaptive Hypermedia (AH): The combination and delivery of digital elements (i.e. text, sound/audio, visuals, animation, video, feedback, instructions, self-assessment tests etc.) on the same digital platform that adapts the learning contents to the needs of different learners who are diverse in terms of abilities, disabilities, levels, interests, backgrounds, and other characteristics.

Hyperlinks: A button, image, icon etc. in a digital environment on which learners can click to navigate to another part of the environment.

Hypermedia: IMM is also called hypermedia.

Hypertext: Hypermedia text retrieval system that enables learners to access particular media types in certain locations or files in applications, webpages or other digital environments.

Interactive Multimedia (IMM): The combination and delivery of digital elements on the same digital platform which have links between elements in the form of buttons, hotspots or hyperlinks to create an interactive application in which users can navigate.

Multimedia: The combination and delivery of digital elements on the same digital platform.

Multiple Media: The use of different tools such as television, the tape recorder, video, the overhead projectors, slide projector etc.

Automatic Item Generation

E**Mark Gierl***University of Alberta, Canada***Hollis Lai***University of Alberta, Canada***Xinxin Zhang***University of Alberta, Canada*

INTRODUCTION

As the importance of technology in society continues to increase, countries require skilled workers who can produce new ideas, make new products, and provide new services. The ability to create these ideas, products, and services will be determined by the effectiveness of our educational programs. Education provide students with the knowledge and skills required to think, reason, communicate, and collaborate in a world that is shaped by knowledge services, information, and communication technologies (e.g., Binkley, Erstad, Herman, Raizen, Ripley, Miller-Ricci, & Rumble, 2012; Darling-Hammond, 2014). Educational testing has an important role to play in helping students acquire these foundational skills. Educational tests, once developed almost exclusively to satisfy demands for accountability and outcomes-based summative testing, are now expected to provide teachers and students with timely, detailed, formative feedback to directly support teaching and learning. To meet these teaching and learning directives, formative principles are beginning to guide our educational testing practices. Formative principles can include any assessment-related activities that yield constant and specific feedback to modify teaching and improve learning, including administering tests more frequently (Black & Wiliam, 1998, 2010). But when testing occurs frequently, more test items are required. These additional test items must be

created efficiently and economically while maintaining a high standard of quality. Fortunately, this requirement for frequent and timely educational testing coincides with the dramatic changes occurring in instructional technology. Developers of local, national, and international educational tests are now implementing computerized tests at an extraordinary rate (Beller, 2013). Computerized testing offers many important benefits to support and promote key principles in formative assessment. Computers permit testing on-demand thereby allowing students to take the test at any time during instruction; items on computerized tests are scored immediately thereby providing students with instant feedback; computerized tests permit continuous administration thereby allowing students to have more choices about when they write their exams. In short, computers are helping infuse formative principles into our testing practices to support teaching and learning.

Despite these important benefits, the advent of computerized testing has also raised formidable challenges, particularly in the area of test item development. Educators must have access to large numbers of diverse, high-quality test items to implement computerized testing because items are continuously administered to students. Hundreds of items are needed to develop the test item banks necessary for computerized testing. Unfortunately, test items, as they are currently created, are time consuming and expensive to develop because each individual item is written, initially, by a content

specialist and, then, reviewed, edited, and revised by groups of content specialists (Gierl & Lai, 2016a; Rudner, 2010). Hence, item development has been identified as one of the most important problems that must be solved before we can fully migrate to computerized testing because large numbers of high-quality, content-specific, test items are required (Haladyna & Rodriguez, 2013; Webb, Gibson, & Forkosh-Baruch, 2013).

One promising test item development method that may be used to address this challenge is with *automatic item generation* (AIG) (Gierl & Haladyna, 2013; Irvine & Kyllonen, 2002). AIG is a relatively new but rapidly evolving research area where cognitive and psychometric modeling practices guide the production of tests that include items generated with the aid of computer technology. Research on AIG has adopted different perspectives, including the use of natural language processing and rule-based artificial intelligence (e.g., Gütl, Lankmayr, Weinhofer, & Höfler, 2011; Moser, Gütl, & Lui, 2012), frame-semantic representations (e.g., Cubric & Toasic, 2010; Higgins, Futagi, & Deane, 2005), schema theory (e.g., Singley & Bennett, 2002), and semantic web-rule language (Zoumpatianos, Papasalouros, & Kotis, 2011). The purpose of this chapter is to describe and illustrate the most practical method for generating test items, which is template based. By template-based AIG, we mean methods that draw on *item models* to guide the generative process. Gierl and Lai (2013, 2016a, 2016b) developed a three-step process for template-based AIG. In step 1, content specialists create a cognitive model for AIG. A cognitive model is a representation that highlights the knowledge, skills, and content required to generate new test items. In step 2, an item model is developed to specify where the cognitive model content is placed in each generated item. An item model is a template that highlights the variables in a test item that can be manipulated to produce new items. In step 3, computer algorithms place the cognitive content into the item model. With this process, hundreds of items can be created from a single item model.

The purpose of this chapter is to describe how AIG can be used to generate test items using the selected-response (i.e., multiple-choice) format. To ensure our description is both concrete and practical, we illustrate template-based item generation using an example from the complex problem-solving domain of the medical health sciences. The chapter is concluded by describing two directions for future research.

BACKGROUND

Gierl and Lai (2013, 2016a, 2016b) described a three-step approach for template-based AIG. In step 1, a content specialist creates a cognitive model for AIG. In step 2, an item model is developed to specify where the cognitive model content is placed in each generated item. In step 3, algorithms place the cognitive content into the item model.

Step 1: Identify Content for Item Generation

To begin, the content for item generation is identified by the content specialists. This content is identified using design principles and guidelines that highlight the knowledge, skills, and abilities required to solve problems and perform tasks in a specific domain. A cognitive model for AIG is a representation that organizes the cognitive- and content-specific information into a structured representation of how the content specialist expects that examinees will think about and solve problems. Recently, Gierl and Lai (2016b) proposed the *key features* cognitive model for AIG. With this model, item generation is guided by the relationships among the key features specified in the cognitive model. It is used when the attributes or features of a task are systematically combined to produce meaningful outcomes across the item feature set. The use of constraint programming in step 3 of the AIG process ensures that the relationships among the features yield meaningful items.

The key features cognitive model is most suitable for measuring the examinees' ability to assemble and apply key features within a domain as well as to solve problems using these key features.

Step 2: Create Item Models

In step 2, an item model is developed to specify where the content from the cognitive model must be placed to generate new items. An item model is a template that specifies the features in an item that can be manipulated to generate new items. Item models (LaDuca, Staples, Templeton, & Holzman, 1986) specify which parts of the assessment task can be manipulated for item generation. For a selected-response item type, it includes the stem, the options, and the auxiliary information. The stem is the part of an item which formulates context, content, and/or the question the examinee is required to answer. The options contain the alternative answers with one correct option and one or more incorrect options or distractors. Auxiliary information includes any additional material, in either the stem or option, required to generate an item, including digital media such as text, images, tables, diagrams, sound, and/or video. The stem and options can be divided further into elements. These elements are denoted as non-numeric values (i.e., strings) and numeric values (i.e., integers). By systematically manipulating the elements, many items can be generated from one item model.

Step 3: Generate Test Items Using Computer Technology

In step 3, computer algorithms place the cognitive model content specified in step 1 into the item model developed in step 2. Different types of software have been written to generate test items (e.g., Gütl et al., 2011; Higgins, 2007; Higgins et al., 2005; Singley & Bennett, 2002). Gierl, Zhou, and Alves (2008) introduced and demonstrated the use of the computer program IGOR (which is the acronym for **I**tem **G**enerat**OR**) for template-

based AIG. IGOR is a JAVA-based program designed to assemble test items using the items models from all combinations of elements specified in the cognitive model. Once the content is specified in step 1 and the template is created in step 2, IGOR systematically places the content into the template to produce new items as part of step 3. IGOR has been used to generate items in the content areas of mathematics, science, social studies, architecture, logical reasoning, dentistry, nursing, medical education, business management, and non-verbal reasoning.

The Unique Problem of Generation Distractors for Selected-Response Items

For the selected-response format, items must not only include a stem with a corresponding correct option, but also include a set of incorrect options or distractors. These incorrect options are typically designed from a list of plausible but incorrect alternatives linked to misconceptions or errors in thinking, reasoning, and problem solving. Using a traditional item development approach, a set of distractors are created by a content specialist that is specific to each item. But because AIG produces hundreds of items, an alternative approach is needed to create a correspondingly large number of plausible but erroneous distractors.

Three different methods can be used to generate distractors. The first method for generating distractors is often employed when the elements manipulated as part of the item generation process are first specified. Recall, elements are the variables identified by content specialist and then manipulated using computer algorithms to produce new items. A distractor can be specified as an element that contains content that is related to the correct option but still yields an incorrect response. To identify the content for this element, content specialists identify a list of plausible but incorrect options that are appropriate for all possible items generated with a given item model. Then, distractors are randomly selected from

this pool of plausible but erroneous content and added to each generated item. Hence, this method is called the *distractor pool method with random selection*. It is based on the assumption that a pool of plausible distractors can be created. A sample of these plausible distractors are selected at random to complete the item generation process. The strength of this method is its simplicity. This method can yield large numbers of distractors. The weakness of this method resides with the two strong assumptions required to use this approach. First, it must be assumed that all pooled distractors are equally plausible for all generated items. Equal plausibility is a strong and, in many cases, restrictive assumption. Second, there is little reasoning to guide how distractors are paired with the correct option because pairing is achieved with random assignment. The distractor pool method with random selection is most appropriate for case-specific item generation, where a well-defined set of distractors are known to be plausible across a large number of the generated item.

To improve the plausibility of the distractors, rules and rationales that yield errors or misconceptions can be used to create distractors. Distractor rationales are short description provided by the content specialist to ensure that the reasoning which underlies each option is explicitly stated. When rationales are available, distractors can be systematically generated such that each distractor matches with the rationale. Hence, this approach is called the *systematic generation with rationales method*. It is based on the assumption that algorithms, rules, and procedures can first be articulated by content specialists and then used to create plausible but incorrect alternatives. The strength of this method is that the distractors are much more specific and, hence, plausible and appropriate, especially when compared to the previous method. There is also a precedent for this approach as systematic generation using distractor rationale has been used in quantitative- and rule-based content areas, such as mathematics and science, to produce incorrect options (Haladyna

& Rodriguez, 2013). This method also has some weaknesses. For instance, plausible but incorrect rationales must first be created for each option by the content specialist. This process can be time consuming. Algorithmic or rule-based manipulations must also be permissible for every incorrect option. In some content areas and in some reasoning and problem-solving situations, generic rules or robust algorithms may not be available to guide the development of distractor rationales.

To overcome the weaknesses of the first two methods, a third method for distractor generation is also available. It is called *systematic distractor generation* (Lai, Gierl, DeChamplain, & Boulais, 2015). This method is closely linked to the cognitive modeling approach to AIG. Recall that to generate an item stem and correct option, specific information must be collected in the form of a cognitive model (i.e., step 1 in the three-step AIG process). The systematic distractor generation method also involves specifying information related to errors and misconceptions in the form of a cognitive model in order to generate plausible but incorrect options. The process of modeling incorrect options is more complex compared to modeling the correct option because the item stem is constantly changing in a generative item development system. As a result, the small number of constraints required to ensure that information presented in the stem yields a correct response must be counterbalanced with a much larger number of constraints required to ensure the information presented in the incorrect options is erroneous but plausible. Also, more than one incorrect option is required for a selected-response item. Typically, three, sometimes four, incorrect options must be produced for each generated item. Systematic distractor generation has many strengths. For example, it can be used in diverse testing situations and across many content areas because of its generality. It is also an adaptive method because distractors are selected conditionally using the associate information in the stem and the correct option. Finally, the features that lead to the correct

and incorrect options are explicitly identified. This method also has weaknesses. Systematic distractor generation requires that specific information be collected for each distractor. Because selected-response items contain two or more incorrect options, this specific information is required for each item in order to model the distractors. Hence, it is a time-consuming method for generating the distractors because data must be collected from content specialists, these data must be structured for distractor generation, and then computer code must be written to assemble the content for each option in a selected-response item.

AN EXAMPLE FROM THE MEDICAL HEALTH SCIENCES

To illustrate the three-step approach to AIG, an example from the medical health sciences is presented. This example is drawn from the content area of general surgery where the examinee is required to diagnosis problems that could arise from a serious abdominal injury. This example uses a key features cognitive model with systematic distractor generation.

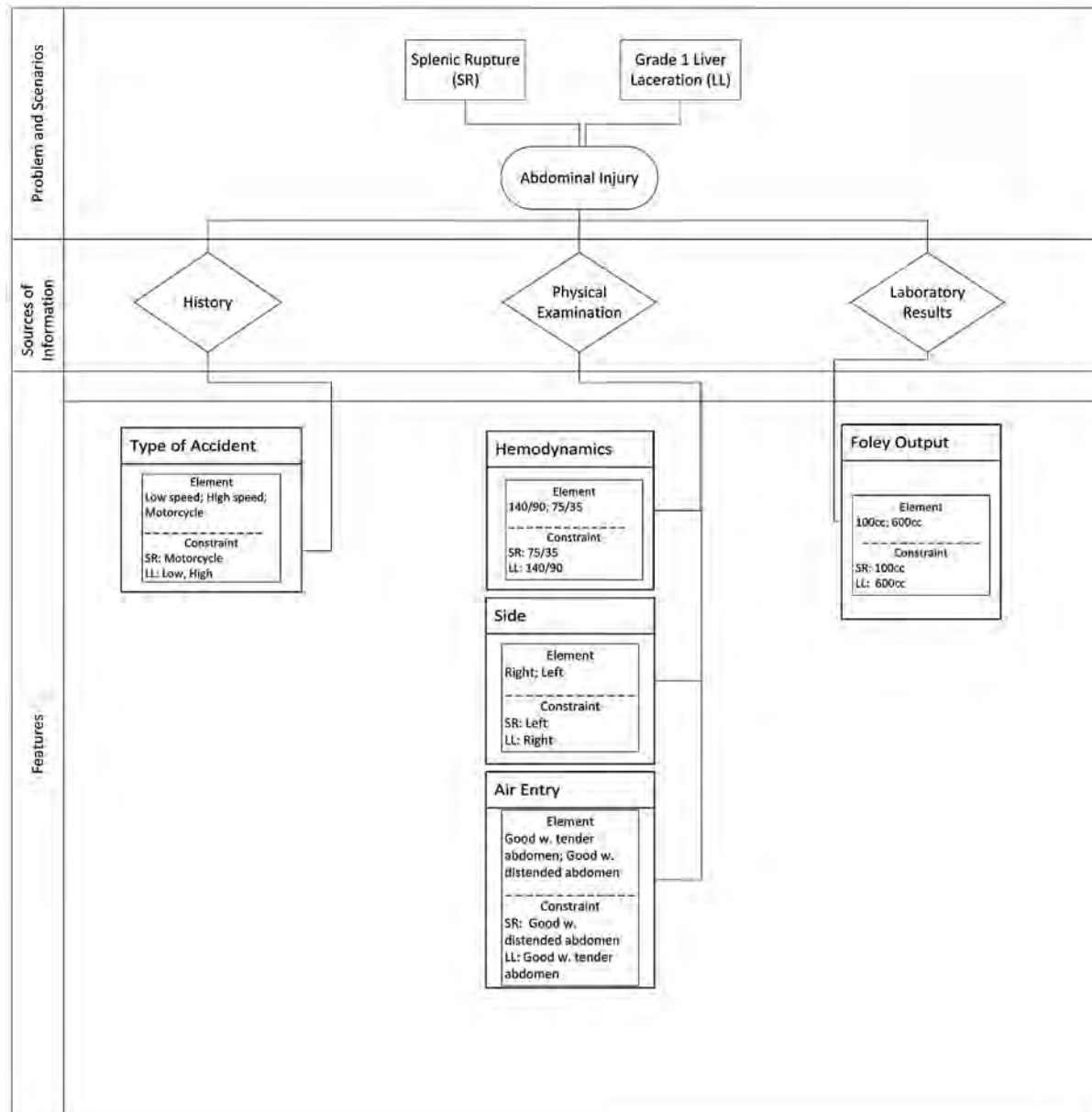
In step 1, a cognitive model for AIG is created. The model in this example was created inductively by asking two 3rd year surgical residents to identify and describe the information that would be used by a surgeon to solve two serious medical problems (i.e., splenic rupture and Grade 1 liver laceration) related to abdominal injury. Three types of information are specified in the cognitive model for AIG. This information is presented as separate panels in Figure 1. The top panel identifies the main concept or problem and its associated scenarios. In this example, the problem is abdominal injury. Two associated scenarios are splenic rupture (SR) and Grade 1 liver laceration (LL). The middle panel specifies the relevant sources of information presented with the problem. In this example, three sources of information were presented: History, physical examination, and laboratory results. The bottom panel highlights

the salient features, which includes the elements and constraints. The first component for a feature is the element. Elements contain content specific to each feature that can be manipulated for item generation. The second component for a feature is the constraint. Constraints serve as restrictions that must be applied during the generation process to ensure that content in the elements are combined in a meaningful way so useful items can be generated. For the Figure 1 example, five features (i.e., type of accident, hemodynamics, side, air entry, and Foley output) were identified across the three sources of information. Each feature specifies the elements and constraints. For instance, the type of accident feature has three elements, low speed, high speed, and motorcycle. These elements contain constraints. For example, splenic rupture is only related to the motorcycle accident (i.e., SR: Motorcycle) whereas the low and high speed accidents are related to the Grade 1 liver laceration (i.e., LL: Low, High).

In step 2, the item model is created (see Table 1). The item model contains the template that will be used to generate test items. It also specifies the elements for the features. Five features are included. The Type of Accident feature has three elements. The remaining four features—Hemodynamics, Side, Air Entry, Foley Output—all contain two elements. Taken together Figure 1 and Table 1 describe the content combinations required to produce the correct option in the cognitive model. Splenic rupture is the correct option, for example, when the elements are: motorcycle accident at highway speeds (Type of Accident), blood pressure is 75/35 and heart rate is 140 (Hemodynamics), left side (Side), good air entry and a large distended abdomen with guarding (Air Entry), and 100cc of bloody urine (Foley Output).

The same logic is also used to produce the incorrect options using the systematic distractor generation method. Distractor 3 is diaphragmatic rupture. This diagnosis shares some, but not all, of the elements with splenic rupture. For instance, a diaphragmatic rupture can be associated with the elements motor vehicle accident at highway speeds

Figure 1. An abdominal injury cognitive model



(Type of Accident), blood pressure is 75/35 and heart rate is 140 (Hemodynamics), and 100cc of bloody urine (Foley Output). But a diaphragmatic rupture is not associated with pain on either the left or the right side (Side) and does not necessarily display physical examination results related to air entry and abdominal pain (Air Entry), as presented in the cognitive or item models. As a

result, diaphragmatic rupture is a plausible but erroneous option because it shares some but not all of the features related to splenic rupture. The same logic used to code the elements in diaphragmatic rupture is also used to code pneumothorax, cardiac tamponade, and aortic rupture in order to produce a cognitive model for the incorrect options. Then, an incorrect option is selected from

Table 1. An item model for generating abdominal injury test items

Stem	A 25-year-old male is involved in a [Type of Accident]. Emergency Medical Services resuscitates him with 2L crystalloid and transports him to your tertiary centre. When he arrives his [Hemodynamics]. He has a Glasgow Coma Scale score of 14. He is complaining of lower-rib pain on his [Side]. On examination, he has [Air Entry]. [Foley Output]. What is the most likely diagnosis?
Elements	Type of Accident: 1: highway speed roll over 2: bicycle accident where he hits a barricade 3: motorcycle accident at highway speeds Hemodynamics: 1: blood pressure is 75/35 and his heart rate is 140 2: blood pressure is 140/90 and his heart rate is 90 Side: 1: left side 2: right side Air Entry: 1: good air entry and a large distended abdomen with guarding 2: good air entry, mildly tender abdomen, and no guarding Foley Output: 1: A foley catheter emits 100cc of bloody urine 2: A foley catheter emits 600cc of urine
Options	1: Splenic rupture 2: Grade 1 liver laceration 3: Diaphragmatic rupture 4: Pneumothorax 5: Cardiac tamponade 6: Aortic rupture 7: Epidural hematoma
Key	

this distractor model for inclusion in the generated item whenever it shares some but not all of the elements with the correct option.

In step 3, IGOR assembles the content specified in an item model, subject to elements and constraints articulated in the cognitive model for both the correct and incorrect options. Iterations are conducted in IGOR to assemble all possible combinations of elements and options, subject to the constraints. In the current example, IGOR generated 62 abdominal injury items. The model used in this example is comparatively small and is provided for illustrative purposes only in our chapter. Item generation in operational testing programs typically include 5-7 different features per model where each feature has 2-10 different elements thereby producing much larger numbers of items.

FUTURE RESEARCH DIRECTIONS

Two areas of future research are recommended. First, methods exist for developing cognitive models for AIG (Gierl & Lai, 2016a). But no methods exist for evaluating the reliability and validity of these cognitive models. While Gierl and Lai (2016b) described two different types of cognitive models, many more cognitive models will likely be identified as researchers and practitioners begin to generate items in different domains and content areas. Hence, future research is required to develop methods for evaluating cognitive models as well as for expanding on the types of cognitive models that can be used for AIG. Second, new methods should be developed that allows feedback to be produced as part of the generative process. That is, the solution and rationale should be produced for

the correct and incorrect options during the three-step item generation process. This outcome would require a significant modification to the current item models and IGOR software. The benefits, however, would be significant. Effective formative assessment permits teachers to identify students' problem-solving strengths and weaknesses so they can adjust their instruction to overcome the weaknesses (Quellmalz & Pellegrino, 2009). But a test must contain a large number of items with carefully selected incorrect options related to a single concept in order to pinpoint different types of weaknesses, problems, and/or misconceptions. Item generation could be conducted so the solutions and rationales for the incorrect options are *designed* from a list of plausible but incorrect alternatives linked to common misconceptions or errors thereby providing teachers with this type of pinpoint precision. Therefore, additional research should be undertaken to refine the item models and the IGOR software so solutions and rationales that provide teachers with specific feedback can be produced as part of the item generation process.

CONCLUSION

As the importance of technology continues to increase, countries require skilled workers who can produce new ideas, products, and services. The ability to create these outcomes will be determined by the effectiveness of our educational programs. Education provides students with the knowledge and skills required to think, reason, communicate, and collaborate in a world that is increasingly shaped by knowledge services, information, and communication technologies. Computerized testing will foster the development of this highly skilled workforce by helping students acquire these essential skills. Computers allow educators to administer tests more frequently so students can receive specific continuous feedback as they develop their skills. But when tests are given more frequently, a constant supply of test items is needed. Automatic item generation is a

rapidly evolving research area where cognitive and psychometric theories are used to constantly produce tests that contain items that are generated with computer technology. The purpose of this chapter was to describe and illustrate a template-based method for generating test items to address the challenging problem of rapidly and economically producing large numbers of high-quality, content-specific, test items required to support computerized testing.

REFERENCES

- Beller, M. (2013). Technologies in large-scale assessments: New directions, challenges, and opportunities. In M. von Davier, E. Gonzalez, I. Kirsch, & K. Yamamoto (Eds.), *The Role of international large-scale assessments: Perspectives from technology, economic and educational research* (pp. 25–45). New York: Springer. doi:10.1007/978-94-007-4629-9_3
- Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., Miller-Ricci, M., & Rumble, M. (2012). Defining twenty-first century skills. In P. Griffin, B. McGaw, & E. Care (Eds.), *Assessment and teaching of 21st century skills* (pp. 17–66). New York: Springer. doi:10.1007/978-94-007-2324-5_2
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74. doi:10.1080/0969595980050102
- Black, P., & Wiliam, D. (2010). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 92(1), 81–90. doi:10.1177/003172171009200119
- Cubric, M., & Tomic, M. (2010). *Towards automatic generation of e-assessment using semantic web technologies*. Paper presented at the International Computer-Assisted Assessment Conference (CAA 2010), Southampton, UK.

- Darling-Hammond, L. (2014). *Next generation assessment: Moving beyond the bubble test to support 21st century learning*. San Francisco, CA: Jossey-Bass.
- Gierl, M. J., & Haladyna, T. (2013). *Automatic item generation: Theory and practice*. New York: Routledge.
- Gierl, M. J., & Lai, H. (2013). Using automated processes to generate test items. *Educational Measurement: Issues and Practice*, 32(3), 36–50. doi:10.1111/emip.12018
- Gierl, M. J., & Lai, H. (2016a). Automatic item generation. In S. Lane, M. Raymond, & T. Haladyna (Eds.), *Handbook of test development* (2nd ed.; pp. 410–429). New York: Routledge.
- Gierl, M. J., & Lai, H. (2016b). The role of cognitive models in automatic item generation. In A. Rupp & J. Leighton (Eds.), *The Wiley handbook of cognition and assessment: Frameworks, methodologies, and applications* (pp. 124–145). New York: Wiley.
- Gierl, M. J., Zhou, J., & Alves, C. (2008). Developing a taxonomy of item model types to promote assessment engineering. *The Journal of Technology, Learning, and Assessment*, 7(2). <http://www.jtla.org> Retrieved January 20, 2011
- Gütl, C., Lankmayr, K., Weinhofer, J., & Höfler, M. (2011). Enhanced Automatic Question Creator – EAQC: Concept, development and evaluation of an automatic test item creation tool to foster modern e-education. *Electronic Journal of e-Learning*, 9, 23–38.
- Haladyna, T. M., & Rodriguez, M. C. (2013). *Developing and validating test items*. New York: Routledge.
- Higgins, D. (2007). *Item Distiller: Text retrieval for computer-assisted test item creation*. Educational Testing Service Research Memorandum (RM-07-05). Princeton, NJ: Educational Testing Service.
- Higgins, D., Futagi, Y., & Deane, P. (2005). *Multilingual generalization of the Model Creator software for math item generation*. Educational Testing Service Research Report (RR-05-02). Princeton, NJ: Educational Testing Service.
- Irvine, S. H., & Kyllonen, P. C. (2002). *Item generation for test development*. Hillsdale, NJ: Erlbaum.
- LaDuca, A., Staples, W. I., Templeton, B., & Holzman, G. B. (1986). Item modeling procedures for constructing content-equivalent multiple-choice questions. *Medical Education*, 20(1), 53–56. doi:10.1111/j.1365-2923.1986.tb01042.x PMID:3951382
- Lai, H., Gierl, M., DeChamplain, A., & Boulais, A. (2015, April). *Systematically generating plausible distractors: Field-testing results and psychometric implications from item generation*. Paper presented at the annual meeting of the Canadian Conference on Medical Education, Vancouver, BC.
- Moser, J. R., Gütl, C., & Liu, W. (2012). *Refined distractor generation with LSA and stylometry for automated multiple choice question generation*. Paper presented at the 25th Australasian Joint Conference-Advances in Artificial Intelligence (AL 2012), Sydney, Australia doi:10.1007/978-3-642-35101-3_9
- Quellmalz, E. S., & Pellegrino, J. W. (2009). Technology and testing. *Science*, 323(5910), 75–79. doi:10.1126/science.1168046 PMID:19119222
- Rudner, L. (2010). Implementing the Graduate Management Admission Test Computerized Adaptive Test. In W. van der Linden & C. Glas (Eds.), *Elements of adaptive testing* (pp. 151–165). New York, NY: Springer.
- Singley, M. K., & Bennett, R. E. (2002). Item generation and beyond: Applications of schema theory to mathematics assessment. In S. H. Irvine & P. C. Kyllonen (Eds.), *Item generation for test development* (pp. 361–384). Mahwah, NJ: Erlbaum.

Webb, M., Gibson, D., & Forkosh-Baruch, A. (2013). Challenges for information technology supporting educational assessment. *Journal of Computer Assisted Learning*, 29(5), 451–462. doi:10.1111/jcal.12033

Zoumpatianos, K., Papasalouros, A., & Kotis, K. (2011). *Automated transformation of SWRL rules into multiple-choice questions*. Paper presented at the FLAIRS Conference 11, Palm Beach, FL.

ADDITIONAL READING

Bormuth, J. (1969). *On a theory of achievement test items*. Chicago: University of Chicago Press.

Drasgow, F., Luecht, R. M., & Bennett, R. (2006). Technology and testing. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 471–516). Washington, DC: American Council on Education.

Embretson, S. E., & Yang, X. (2007). Automatic item generation and cognitive psychology. In C. R. Rao & S. Sinharay (Eds.) *Handbook of statistics: Psychometrics*, Volume 26 (pp. 747–768). North Holland, UK: Elsevier.

Geerlings, H., Glas, C. A. W., & van der Linden, W. J. (2011). Modeling rule-based item generation. *Psychometrika*, 76(2), 337–359. doi:10.1007/s11336-011-9204-x

Gierl, M. J., Lai, H., Fung, K., & Zheng, B. (2016). Using technology-enhanced processes to generate items in multiple languages. In F. Drasgow (Ed.), *Technology and testing: Improving educational and psychological measurement* (pp. 109–127). New York: Routledge.

Lane, S. Raymond, M. & T. Haladyna (2016). *Handbook of test development* (2nd Ed.). New York: Routledge.

Luecht, R. (2013). An introduction to assessment engineering for automatic item generation. In M. J. Gierl & T. Haladyna (Eds.), *Automatic item generation: Theory and practice* (pp. 59–76). New York: Routledge.

Popp, E. C., Tuzinski, K., & Fetzer, M. (2016). Actor or avatar? Considerations in selecting appropriate formats for assessment content. In F. Drasgow (Ed.), *Technology and testing: Improving educational and psychological measurement* (pp. 79–103). New York: Routledge.

Schmeiser, C. B., & Welch, C. J. (2006). Test development. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 307–353). Westport, CT: National Council on Measurement in Education and American Council on Education.

KEY TERMS AND DEFINITIONS

Automatic Item Generation: A process of using item models to generate test items with the aid of computer technology.

Cognitive Model: A representation that highlights the knowledge, skills, problem-solving processes and/or content an examinee requires to answer test items.

Distractor Pool Method with Random Selection: A method for creating distractors when the distractors created from a list and then the list is used to randomly select plausible but erroneous content for each generated item.

Elements: Variables in the item model that can be modified to create new test items.

Item Model: A template that highlights the features in an item that must be manipulated to generate new items.

Key Features Cognitive Model: A model used to guide item generation based on the relationships

among the key features specified in the cognitive model, which include the attributes or features of a task are systematically combined to produce meaningful outcomes across the item feature set.

Systematic Distractor Generation: A method for generating distractors where specific informa-

tion related to errors and misconceptions is used to create plausible but incorrect options.

Systematic Generation with Rationales

Method: A method to systematically create distractors when rationales are used to produce a list of incorrect options.

E

Challenges in Developing Adaptive Educational Hypermedia Systems

Eileen O'Donnell

Trinity College Dublin, Ireland

Liam O'Donnell

Dublin Institute of Technology, Ireland

INTRODUCTION

Traditional educational hypermedia systems afford learners the “one size fits all” approach to learning (Brusilovsky, 2003, 2004; Chatti, Jarke, & Specht, 2010; Hsieh, Lee, & Su, 2013). In the “one size fits all” approach to learning each student in every cohort of students is given access to the same learning objects in the same way as every other student who is studying the same course. The learning objects or learning content stays static regardless of the learning requirements of different students.

The objective of Adaptive Educational Hypermedia Systems (AEHS) is to afford learners the opportunity to engage with learning content which has been specifically designed to meet the learning requirements of each individual learner by adapting the content and the user interface to suit the needs of a specific user. AEHS could be used in the education of learners at all stages of their education from junior school to post graduate level. AEHS could also be used in organisations for continuous professional development or training for compliance purposes, for example, first aid or manual handling. Software engineering for AEHS commences with a thorough study of the requirements of the proposed system. AEHS, as proposed systems, are very complex systems to design as the software engineer has to design a system to enable non-technical and technical educator authors to design adaptive learning courses for use by students. Therefore, the design and development of AEHS are very complicated, time

consuming, and expensive. This article reviews a few of the challenges encountered in the design and development of these complex systems and some of the challenges encountered by educators who propose to use AEHS with their students.

The background section of this article provides the reader with brief definitions and discussions on the concept of AEHS and positions AEHS in the larger research area of E-Learning or Technology Enhanced Learning. The main body of the paper outlines some of the challenges encountered in the development and use of AEHS including: the classification of different categories of learners; the sourcing of suitable educational materials or learning resources; gauging the impact that AEHS have on the learning experience of end users; and student access to open and editable user models/profiles. Followed by sections on the following: solutions and recommendations; future research directions and the conclusion.

BACKGROUND

“Adaptive hypermedia systems build a model of the goals, preferences and knowledge of each individual user, and use this model throughout the interaction with the user, in order to adapt to the needs of that user” (Brusilovsky, 2001, p. 87). AEHS build a model of each individual student, and use the information from this model to determine the adaptive learning experiences to be created for each student.

The following sections provide more information on AEHS and include definitions of the terms Adaptive Education (AE), Adaptive Educational Hypermedia (AEH) and Adaptive Educational Hypermedia Systems (AEHS). E-learning and Technology Enhanced Learning (TEL) pertain to various forms of teaching and learning through the use of technology (O'Donnell & O'Donnell, 2015) and access to the Internet. TEL in the context of this article can be used synonymously with E-Learning. AEHS is a form of E-Learning which goes beyond the “one size fits all” approach to E-Learning by adapting the content to suit the learning requirements of individual learners.

Adaptive Education (AE)

Adaptive Education (AE) can be defined as an educational experience that adapts to suit the learning requirements of each individual learner. The purpose of AE is to provide learners with learning resources which have been specially selected to suit their specific learning needs.

Adaptive Educational Hypermedia (AEH)

AEH is electronic content which can be used in the provision of adaptive education. Software developers and educational providers are continuously

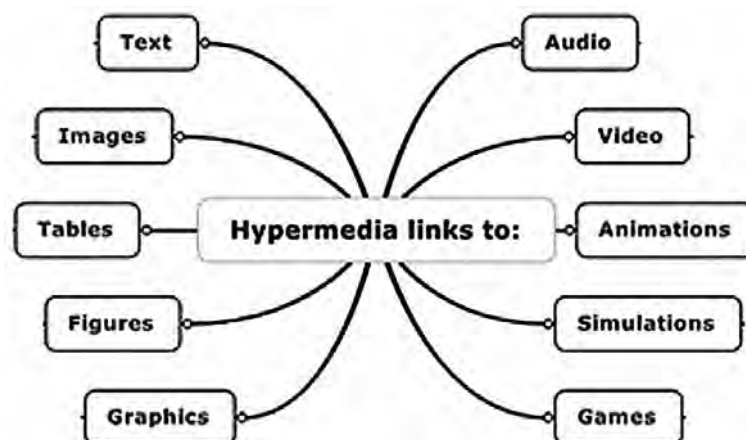
exploring how technology can be used to enhance the learning experience of students.

In a study of AEH authoring tools, Gaffney, Staikopoulos, O'Keeffe, Conlan, and Wade (2014) suggest that “AEH authoring tools have however not been as successfully adopted as was initially expected” (p. 416). Further research is required to explore why AEH authoring tools have not been as successfully adopted as initially expected.

Adaptive Educational Hypermedia Systems (AEHS)

Adaptive Educational Hypermedia Systems (AEHS) are designed and developed to deliver adaptive educational experiences to students. Some examples of AEHS include: GRAPPLE (De Bra et al., 2012; Glahn et al., 2011); AHA! (De Bra, Stash, Smits, Romero, & Ventura, 2007; Knutov, De Bra, & Pechenizkiy, 2009); and ELM-ART (Weber & Brusilovsky, 2001). AEHS use many different mediums of electronic content. Hypertext is a section of online text or an online paragraph of information that has been embedded with links to other content. Hypermedia is electronic content which includes links to many different mediums of content as shown in Figure 1, such as: text; images; tables; figures; graphics; audio; video; animations; simulations; or games.

Figure 1. Hypermedia can be linked to many different mediums of content



The purpose of AEHS is to provide educators with the appropriate toolset to present learners with educational resources that have been specifically selected to suit their individual learning requirements. AEHS or Adaptive Hypermedia Systems for Education (AHSE) have the potential to facilitate personalised technology enhanced learning (Muntean & McManis, 2004). AEHS Authoring Tools (ATs) are created to enable non-technical educator authors to design and create adaptive learning resources.

CHALLENGES ENCOUNTERED IN THE DEVELOPMENT OF AEHS

The development of AEHS is a very difficult process (Baig, 2014). There are many challenges to consider when designing these systems. A few of the challenges are outlined in this paper but others do exist. AEHS are very complex systems because they are designed to be used by many educators and many students who will possess various different knowledge levels of the use of Information and Communications Technology (ICT). There are numerous considerations and challenges involved in designing and developing systems for use by non-technical authors. There are also several challenges involved in developing systems for use by technically competent authors who may require the affordance of complicated functionality. In addition, the designers and developers of AEHS

have to meet the challenges of facilitating the functionality required by educators who employ various different pedagogical teaching methods (O'Donnell, Lawless, Sharp, & Wade, 2015) and learning theories (O'Donnell, Lawless, Sharp, & O'Donnell, 2015).

Adding adaptive functionality to an educational system does not necessarily improve the system, sometimes the adaptive functionality can cause users to lose control of the system (Brusilovsky, 1996). Adding adaptive functionality to an existing Learning Management System (LMS) does not necessarily improve the students' learning experiences. The students may not understand why they are getting access to different learning experiences to those of their peers. And, the educators may lose control over the adaptive functionality resulting in the students not achieving the expected learning outcomes.

Instructional designers and academics require a number of skills to develop positive learning experiences (O'Rourke & Martin, 2011). As shown in Figure 2, some of these required skills may include: pedagogical knowledge (Hirumi, Appelman, Rieber, & Van Eck, 2010; Koh & Chai, 2014); critical and creative thinking (Baum & Newbill, 2010); knowledge of instructional development (Stes, Coertjens, & van Petegem, 2010); awareness of aesthetic principles (Yanchar & Gabbitas, 2011); experience in instructional design (van Rooij, 2010); competence in the use of ICT (O'Rourke & Martin, 2011); among others.

Figure 2. Some of the skills required to develop positive learning experiences

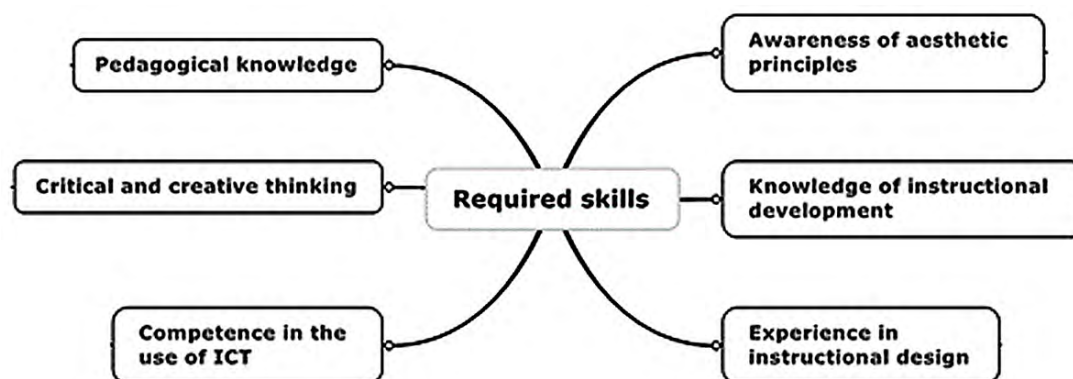
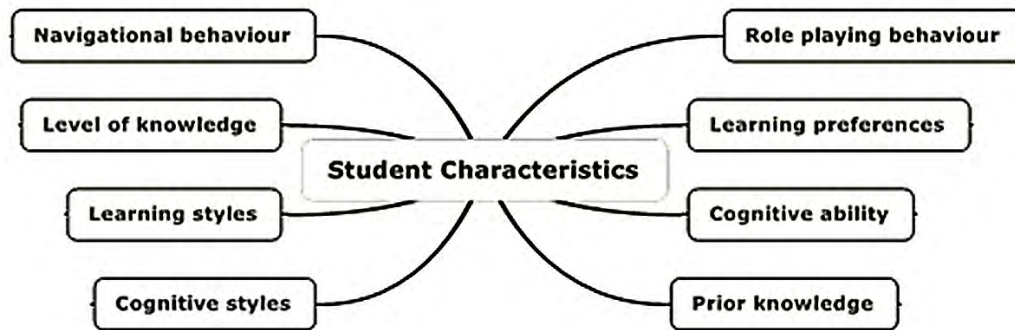


Figure 3. Some of the student characteristics that can be used to classify different types of learners



The skill set required and the challenges encountered will markedly increase as instructional designers attempt to develop positive learning experiences that are also adaptable to each learner's specific learning requirements.

The use of AEHS to support learners can pose challenges to learners with respect to cognitive load and self-regulated learning, Moos (2014) and Kuo and Huang (2009) proposed the use of scaffolding in the development of AEHS to support learners. In the design and development of AEHS both the academic users and the student users (learners) have to be considered and facilitated. This dual requirement increases the complexity in designing and developing AEHS.

The Classification of Different Categories of Learners

Classification of learners is necessary to inform adaptive E-Learning models (Musumba, Oboko, & Nyongesa, 2013). User profiles (Brusilovsky & Millan, 2007) are necessary to store the information collected on each user. Profiling the learners' present knowledge informs academics about the gaps in knowledge and the level at which this information should be focused (Ajmal, Hamidullah, Rahman, & Khan, 2011).

Classification can be based on various different student characteristics as outlined in Figure 3, including the following: level of knowledge (Musumba et al., 2013); learning styles (Akbulut & Cardak, 2012); cognitive styles (Lo, Chan, & Yeh,

2012); prior knowledge (Chen, Chen, & Sun, 2014; O'Donnell, Sharp, Wade, & O'Donnell, 2012); learning preferences (O'Donnell et al., 2012); cognitive ability (O'Donnell et al., 2012); navigational behaviour (Brusilovsky, 2007; O'Donnell et al., 2012); role playing behaviour (Peeters, Bosch, Meyer, & Neerincx, 2014); among others.

The Sourcing of Suitable Educational Materials or Learning Resources

Another challenge encountered in the development of AEHS is the identification of suitable learning resources or objects to be used in the system. Learning objects would also have to be classified to enable retrieval and reuse by AEHS. While educational materials are frequently made available freely online, these resources are available in a number of different formats and may not be easily adopted for use in AEHS. The sourcing of suitable learning resources to facilitate the creation of Adaptive Hypermedia Systems is labour intensive (Levacher, Lawless, & Wade, 2014) and sometimes it is simply not possible to identify appropriate learning resources to use. The creation of learning resources for use in adaptive systems is a time consuming and complex process (Nurjanah & Davis, 2012), labour intensive and expensive. A research study conducted by O'Donnell (2008) was undertaken to establish lecturers' views on the use of ICT in higher education. The study participants were lecturers from various disciplines within the Dublin Institute of Technology. 41 lecturers

participated in this study. One of the findings was that only 15% of the lecturers surveyed agreed that they had sufficient time to create E-Learning resources (O'Donnell, 2008).

Gauging the Impact That AEHS Have on the Learning Experience of End Users

It is very difficult to state the impact that adaptive systems have had on the learning experience of the end user (Mulwa, Lawless, Sharp, Arnedillo-Sanchez, & Wade, 2010; Mulwa, Lawless, Sharp, & Wade, 2012) and “the jury is still out as regards evidence that ICT supports learning” (Livingstone, 2012, p. 19). The findings of Griff and Matter (2013) from the evaluation of LearnSmart (an adaptive online learning system) was that it had no overall effect. Because it is very difficult to gauge the impact that adaptive systems have had on the learning experience of the end users it may be challenging to encourage more educators to engage with the time consuming task of creating courses using AEHS.

There is a need through controlled evaluations of adaptive technologies to gauge the value added to improving student learning (Shute & Zapata-Rivera, 2012). Brusilovsky, Karagiannidis and Sampson (2004) proposed a layered evaluation approach be used when evaluating adaptive learning systems. Until AEHS are used in mainstream education the full impact of adaptive courses on the learning experience of students cannot be evaluated. Further investigations are necessary to gauge the impact that adaptive systems have on the learning experience of end users (learners, students, trainees, employees).

Student Access to Open and Editable User Models/Profiles

Adaptive systems gather some of the data to be stored in the user model by tracking a user's engagement with the adaptive system (Brusilovsky et al., 2008). Some adaptive systems automatically

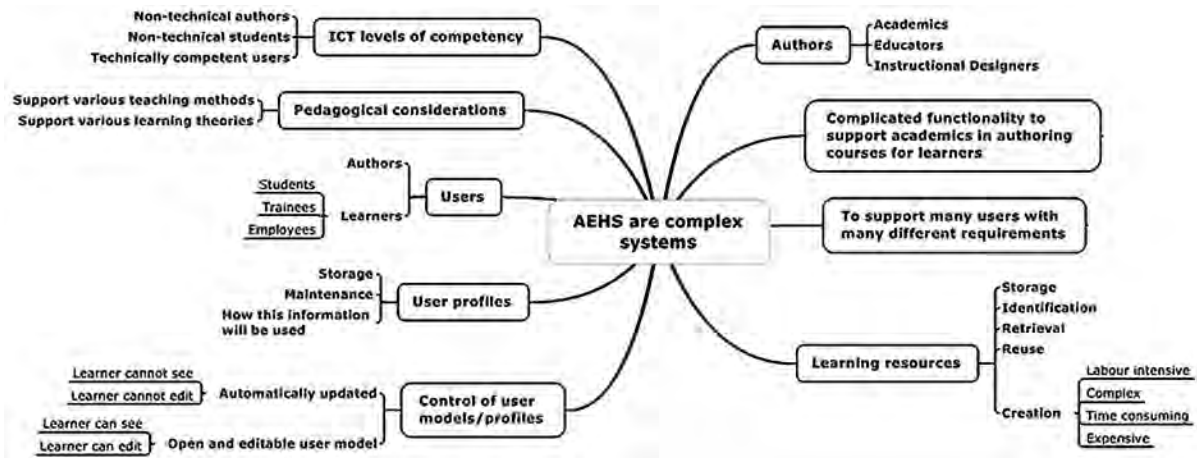
update the user model so that the user never sees the changes made to their personal user model. In addition, in some adaptive systems users are not given the opportunity to edit their own personal user model for fear it will interfere with the personalisation processes of the system. Ahn, Brusilovsky and Han (2015) state that many arguments have been presented “in favour of open and editable user models” (p. 202).

Another challenge to the developers of AEHS is whether or not to provide the functionality to afford educators the opportunity to enable the users to see and edit their own user model. Some educators may be in favour of allowing their students to see and edit their own user model, while others may prefer to keep control of the user models through system generated automatic updates. In a study of self-regulated personalised learning, Steiner, Nussbaumer, and Albert (2009) found that “Up until now, the structures and algorithms underlying personalisation have been completely kept back from the user. The learner is completely unaware of the personalisation process that is taking place behind the scenes...” (p. 650).

In a survey of forty academics conducted by O'Donnell, Sharp, Wade and O'Donnell (2012) in response to the question “Would you trust the decision making algorithms in an authoring tool to determine the most suitable learning activities for each individual student?” (p. 9), only 10% of the academics surveyed responded “yes” to this question. Further along in the same study, O'Donnell et al state “The fact that only 10% of academics surveyed would trust the decision making algorithms is a finding of statistical significance that requires further investigation” (O'Donnell et al., 2012, p. 16). If academics find it hard to trust the decision making algorithms there is a possibility that learners would also like more information about the structures and algorithms used to determine the personalisation process.

Ashman, Brailsford, Cristea, Sheng, Stewart, Toms and Wade (2014) suggest that users should be informed about the personal information that is being collected and how this information is

Figure 4. AEHS are complex systems to design and develop



used in educational and other settings. In a study conducted by Staikopoulos, O’Keeffe, Rafter, Walsh, Yousuf, Conlan and Wade (2014), 30.6% of the students surveyed wanted more control over the content selected by the personalised course.

The amount of control to be allocated to the user of the personalised systems (Parra-Santander & Brusilovsky, 2015) and the right of the user to edit their own user model (Ahn et al., 2015) pose further challenges in the design and development of AEHS. The amount of control allocated to the user and the right of the user to edit their user model require further investigation to establish some guidelines for those challenged with the development of AEHS. AEHS are complex systems to design and develop and present many challenges as illustrated in Figure 4.

SOLUTIONS AND RECOMMENDATIONS

Designers and developers of AEHS require further feedback from the academic community and student body to inform future developments of AEHS authoring tools. Instructional designers and developers of E-Learning courses require additional supports and encouragement to engage with AEHS authoring tools. Further evaluations of existing adaptive courses are required to gauge the

impact that adaptive courses have on the learning experience of students to inform the designers and developers of AEHS and the educators who design and develop adaptive courses.

FUTURE RESEARCH DIRECTIONS

Further research is required to explore why AEH authoring tools have not been as successfully adopted as initially expected. AEHS are not yet sufficiently developed for use by non-technical authors. Further work is required on the development of AEHS. Further investigations are necessary to gauge the impact that adaptive systems have on the learning experience of end users (learners, students, trainees and employees). Further investigations are also required into the amount of control that should be allocated to the end user, and the right of the user to see and edit their user model/profile.

CONCLUSION

AEHS are very complex systems to design and develop. While the concept of AEHS is very promising, the realisation is very challenging, expensive, and time consuming. Some educators feel that they have insufficient time to engage in

the creation of e-learning resources while others feel that they could not trust the decision making algorithms which determine the most suitable learning resources to present to each student. The challenges encountered in the development of AEHS are many and varied as outlined in this paper. Further research is necessary to investigate the true potential of AEHS.

REFERENCES

- Ahn, J., Brusilovsky, P., & Han, S. (2015). *Personalized search: reconsidering the value of open user models*. Paper presented at the 20th International Conference on Intelligent User Interfaces, Atlanta, GA. doi:10.1145/2678025.2701410
- Ajmal, M., Hamidullah, H., Rahman, F., & Khan, B. (2011). Course book evaluation through instructional design parameters in the system of distance education. *International Journal of Academic Research*, 3(1), 638–644.
- Akbulut, Y., & Cardak, C. (2012). Adaptive educational hypermedia accommodating learning styles: A content analysis of publications from 2000 to 2011. *Computers & Education*, 58(2), 835–842. doi:10.1016/j.compedu.2011.10.008
- Ashman, H., Brailsford, T., Cristea, A., Sheng, Q., Stewart, C., Toms, E., & Wade, V. (2014). The ethical and social implications of personalization technologies for e-learning. *Information & Management*, 51(6), 819–832. doi:10.1016/j.im.2014.04.003
- Baig, F. (2014). Quality assessment of COFALE: An authoring tool for Adaptive Educational Hypermedia Systems. *Journal of Quality and Technology Management*, X(11), 95–108.
- Baum, L., & Newbill, P. (2010). A journey through a Jamestown-Era native American village. *TechTrends*, 54(5), 27–37. doi:10.1007/s11528-010-0434-z
- Brusilovsky, P. (1996). Methods and techniques of adaptive hypermedia. *User Modeling and User-Adapted Interaction*, 6(2-3), 87–129. doi:10.1007/BF00143964
- Brusilovsky, P. (2001). Adaptive hypermedia. *User Modeling and User-Adapted Interaction*, 11(1/2), 87–110. doi:10.1023/A:1011143116306
- Brusilovsky, P. (2003). Developing Adaptive Educational Hypermedia Systems: from Design Models to Authoring Tools. In T. Murray, S. Blessing, & S. Ainsworth (Eds.), *Authoring Tools for Advanced Technology Learning Environments* (pp. 377–410). Dordrecht: Kluwer Academic Publishers. doi:10.1007/978-94-017-0819-7_13
- Brusilovsky, P. (2004). *Knowledge Tree: A distributed architecture for adaptive e-learning*. Paper presented at the International World Wide Web Conference - 13th international www conference - Session: Adaptive e-learning systems, New York, NY.
- Brusilovsky, P. (2007). Adaptive navigation support. In P. Brusilovsky, A. Kobsa, & W. Nejdl (Eds.), *The adaptive web* (pp. 263–290). Berlin: Springer-Verlag. doi:10.1007/978-3-540-72079-9_8
- Brusilovsky, P., Karagiannidis, C., & Sampson, D. (2004). Layered evaluation of adaptive learning systems. *International Journal of Continuing Engineering Education and Life-long Learning*, 14(4/5), 402–421. doi:10.1504/IJCEELL.2004.005729
- Brusilovsky, P., & Millan, E. (2007). User models for adaptive hypermedia and adaptive educational systems. In P. Brusilovsky, A. Kobsa, & W. Nejdl (Eds.), *The Adaptive Web*. Berlin: Springer. doi:10.1007/978-3-540-72079-9_1
- Brusilovsky, P., Sosnovsky, S., Lee, D., Yudelso, M., Zadorozhny, V., & Zhou, X. (2008). *An open integrated exploratorium for database courses*. Paper presented at the ITiCSE '08, Madrid Spain. doi:10.1145/1384271.1384280

- Chatti, M. A., Jarke, M., & Specht, M. (2010). The 3P Learning Model. *Journal of Educational Technology & Society*, 13(4), 74–85.
- Chen, J., Chen, M., & Sun, Y. (2014). A tag based learning approach to knowledge acquisition for constructing prior knowledge and enhancing student reading comprehension. *Computers & Education*, 70, 256–268. doi:10.1016/j.compedu.2013.09.002
- De Bra, P., Smits, D., van der Sluijs, K., Cristea, A., Foss, J., Glahn, C., & Steiner, C. (2012). GRAPPLE: Learning Management Systems Meet Adaptive Learning Environments. In A. Peña-Ayala (Ed.), *Intelligent and Adaptive ELS, SIST 17* (pp. 133–160). Berlin: Springer-Verlag.
- De Bra, P., Stash, N., Smits, D., Romero, C., & Ventura, S. (2007). *Authoring and Management Tools for Adaptive Educational Hypermedia Systems: The AHA! Case Study*. In *Studies in Computational Intelligence (SCI)* (Vol. 62, pp. 285–308). Berlin: Springer-Verlag.
- Gaffney, C., Staikopoulos, T., O’Keeffe, I., Conlan, O., & Wade, V. (2014). *A training framework for Adaptive Educational Hypermedia authoring tools*. Paper presented at the Open Learning and Teaching in Educational Communities 9th European Conference on Technology Enhanced Learning, EC-TEL 2014, Graz, Austria. doi:10.1007/978-3-319-11200-8_35
- Glahn, C., Steiner, C., de Bra, P., Docq, F., O’Donnell, E., Verpoorten, D., ... Stash, N. (2011). *GRAPPLE (Generic Responsive Adaptive Personalized Learning Environment): Second empirical evaluation in academic settings*. Retrieved from <http://ebookbrowse.net/d9-5-wp9-finalevaluation-v1-0-pdf-d633058884>
- Griff, E., & Matter, S. (2013). Evaluation of an adaptive online learning system. *British Journal of Educational Technology*, 44(1), 170–176. doi:10.1111/j.1467-8535.2012.01300.x
- Hirumi, A., Appelman, B., Rieber, L., & Van Eck, R. (2010). Preparing instructional designers for game-based learning: Part 1. *TechTrends*, 54(3), 27–37. doi:10.1007/s11528-010-0400-9
- Hsieh, T., Lee, M., & Su, C. (2013). Designing and implementing a personalized remedial learning system for enhancing the programming learning. *Journal of Educational Technology & Society*, 16(4), 32–46.
- Knutov, E., De Bra, P., & Pechenizkiy, M. (2009). AH 12 years later: A comprehensive survey of adaptive hypermedia methods and techniques. *New Review of Hypermedia and Multimedia*, 15(1), 5–38. doi:10.1080/13614560902801608
- Koh, J., & Chai, C. (2014). Teacher clusters and their perceptions of technological pedagogical content knowledge (TPACK) development through ICT lesson design. *Computers & Education*, 70, 222–232. doi:10.1016/j.compedu.2013.08.017
- Kuo, Y. H., & Huang, Y. M. (2009). MEAT: An authoring tool for generating adaptable learning resources. *Journal of Educational Technology & Society*, 12(2), 51–68.
- Levacher, K., Lawless, S., & Wade, V. (2014). Slicepedia: Content-agnostic slicing resource production for Adaptive Hypermedia. *Computer Science and Information Systems*, 11(1), 393–417. doi:10.2298/CSIS121223014L
- Livingstone, S. (2012). Critical reflections on the benefits of ICT in education. *Oxford Review of Education*, 38(1), 9–24. doi:10.1080/03054985.2011.577938
- Lo, J., Chan, Y., & Yeh, S. (2012). Designing an adaptive web-based learning system based on students cognitive styles identified online. *Computers & Education*, 58(1), 209–222. doi:10.1016/j.compedu.2011.08.018
- Moos, D. (2014). Setting the stage for the metacognition during hypermedia learning: What motivation constructs matter? *Computers & Education*, 70, 128–137. doi:10.1016/j.compedu.2013.08.014

- Mulwa, C., Lawless, S., Sharp, M., Arnedillo-Sanchez, I., & Wade, V. (2010). *Adaptive educational hypermedia systems in technology enhanced learning: a literature review*. Paper presented at the ACM. doi:10.1145/1867651.1867672
- Mulwa, C., Lawless, S., Sharp, M., & Wade, V. (2012). *The Evaluation of Adaptive Technology-Enhanced Learning Systems*. Paper presented at the The World Conference on E-Learning in Corporate, Government, Healthcare and Higher Education E-LEARN 2012, Montreal, Quebec, Canada.
- Muntean, C., & McManis, J. (2004). Adaptive e-learning systems: Evaluation issues. *Transactions on Automatic Control and Computer Science*, 49(63), 193–198.
- Musumba, G., Oboko, R., & Nyongesa, H. (2013). Agent-based adaptive e-learning model for any learning management system. *International Journal of Machine Learning and Applications*, 2(1). doi:10.4102/ijmla.v2i1.6
- Nurjanah, D., & Davis, H. (2012). *Improving the workspace awareness of authors in asynchronous collaborative authoring of learning designs*. Paper presented at the EdMedia 2012 - World Conference on Educational Media and Technology, Denver, CO.
- O'Donnell, E. (2008). Can e-learning be used to further improve the learning experience to better prepare students for work in industry (Masters in Information Systems for Managers). Dublin: Dublin City University. Retrieved from <http://arrow.dit.ie/buschmanoth/1>
- O'Donnell, E., Lawless, S., Sharp, M., & O'Donnell, L. (2015). Learning theories: pedagogical strategies for Massive Open Online Courses (MOOCs) in higher education. In E. McKay & J. Lenarcic (Eds.), *Macro-Level Learning Through Massive Open Online Courses (MOOCS)*. Strategies and Predictions for the Future. doi:10.4018/978-1-4666-8324-2.ch006
- O'Donnell, E., & O'Donnell, L. (2015). Technology-Enhanced Learning: Towards providing supports for PhD students and researchers in higher education. In *Curriculum Design and Classroom Management: Concepts, Methodologies, Tools, and Applications* (Vols. 1–3, pp. 242–262). Hershey, PA: Information Science Reference.
- O'Donnell, E., Sharp, M., Wade, V., & O'Donnell, L. (2012). *Academics' views on personalised e-learning in higher education*. Paper presented at the International Conference on Engaging Pedagogy, Institute of Technology, Blanchardstown, Dublin, Ireland. Retrieved from <http://arrow.dit.ie/cgi/viewcontent.cgi?article=1033&context=buschmancon>
- O'Donnell, E., Lawless, S., Sharp, M., & Wade, V. (2015). A review of personalised e-learning: Towards supporting learner diversity. *International Journal of Distance Education Technologies*, 13(1), 22–47. doi:10.4018/ijdet.2015010102
- O'Rourke, S. L., & Martin, M. (2011). Instructional design of a distance education cultural awareness course to enhance currency and authenticity. *British Journal of Educational Technology*, 42(5), 875–882. doi:10.1111/j.1467-8535.2010.01103.x
- Parra-Santander, D., & Brusilovsky, P. (2015). User-controllable personalization: A case study with SetFusion. *International Journal of Human-Computer Studies*, 78, 43–67. doi:10.1016/j.ijhcs.2015.01.007
- Peeters, M., Bosch, K., Meyer, J., & Neerincx, M. (2014). The design and effect of automated directions during scenario-based training. *Computers & Education*, 70, 173–183. doi:10.1016/j.compedu.2013.07.039
- Shute, V., & Zapata-Rivera, D. (2012). Adaptive educational systems. In P. Durlach & A. Lesgold (Eds.), *Adaptive Technologies for Training and Education* (pp. 7–27). Cambridge University Press. doi:10.1017/CBO9781139049580.004

Staikopoulos, A., O'Keeffe, I., Rafter, R., Walsh, E., Yousuf, B., Conlan, O., & Wade, V. (2014). AMASE: A framework for supporting personalised activity-based learning on the web. *Computer Science and Information Systems*, 11(1), 343–367. doi:10.2298/CSIS121227012S

Steiner, C., Nussbaumer, A., & Albert, D. (2009). Supporting self-regulated personalised learning through Competence-Based Knowledge Space Theory. *Policy Futures in Education*, 7(6), 645–661. doi:10.2304/pfie.2009.7.6.645

Stes, A., Coertjens, L., & van Petegem, P. (2010). Instructional development for teachers in higher education: Impact on teaching approach. *Higher Education*, 60(2), 187–204. doi:10.1007/s10734-009-9294-x

van Rooij, S. W. (2010). Project management in instructional design: ADDIE is not enough. *British Journal of Educational Technology*, 41(5), 852–864. doi:10.1111/j.1467-8535.2009.00982.x

Weber, G., & Brusilovsky, P. (2001). ELM_ART: An adaptive versatile system for web-based instruction. *International Journal of Artificial Intelligence in Education*, 12, 351–384.

Yanchar, S., & Gabbitas, B. (2011). Between eclecticism and orthodoxy in instructional design. *Educational Technology Research and Development*, 59(3), 383–398. doi:10.1007/s11423-010-9180-3

ADDITIONAL READING

Acampora, G., Gaeta, M., & Loia, V. (2011). Combining multi-agent paradigm and memetic computing for personalized and adaptive learning experiences. *Computational Intelligence*, 27(2), 141–165. doi:10.1111/j.1467-8640.2010.00367.x

Al-Azawei, A., & Badii, A. (2014). State of the art of learning styles-based adaptive educational hypermedia systems (Ls-Baehss). *International Journal of Computer Science and Information Technology*, 6(3), 1–19. doi:10.5121/ijcsit.2014.6301

Arora, A., Raisinghani, M., Thompson, L., & Leseane, R. (2011). Personality scales and learning styles: Pedagogy for creating an adaptive web-based learning system. *International Journal of Online Pedagogy and Course Design*, 1(1), 29–49. doi:10.4018/ijopcd.2011010103

Bulu, S., & Pedersen, S. (2012). Supporting problem-solving performance in a hypermedia learning environment: The role of students prior knowledge and metacognitive skills. *Computers in Human Behavior*, 28(4), 1162–1169. doi:10.1016/j.chb.2012.01.026

Caravantes, A., & Galán, R. (2011). Generic educational knowledge representation for adaptive and cognitive systems. *Journal of Educational Technology & Society*, 14(3), 252.

Colace, F., Santo, M. D., & Greco, L. (2014). E-Learning and personalized learning path: A proposal based on the adaptive educational hypermedia system. [iJET]. *International Journal of Emerging Technologies in Learning*, 9(2), 9–16. doi:10.3991/ijet.v9i2.3211

Cristea, A., & Ghali, F. (2011). Towards adaptation in e-learning 2.0. *New Review of Hypermedia and Multimedia*, 17(2), 199–238. doi:10.1080/13614568.2010.541289

De Bra, P., Smits, D., van der Sluijs, K., Cristea, A., Foss, J., Glahn, C., & Steiner, C. (2012). GRAPPLE: learning management systems meet adaptive learning environments. In A. Peña-Ayala (Ed.), *Intelligent and Adaptive ELS, SIST 17* (pp. 133–160). Berlin, Heidelberg: Springer-Verlag.

- Glahn, C., Steiner, C., De Bra, P., Docq, F., & O'Donnell, E. (2010). GRAPPLE (Generic Responsive Adaptive Personalized Learning Environment): Second documentation and training for GRAPPLE users. Retrieved from <http://grapple-project.org/public-files/deliverables/D9.4-WP9-SecondTrainingReport-v1.1.pdf>
- Kahraman, H. T., Sagioglu, S., & Colak, İ. (2013). A novel model for web-based adaptive educational hypermedia systems: SAHM (supervised adaptive hypermedia model). *Computer Applications in Engineering Education*, 21(1), 60–74. doi:10.1002/cae.20451
- Mahnane, L., Trigano, P., Tayeb, L. M., & Benmimoun, A. (2011). Designing an adaptive hypermedia system based on the use of psycho pedagogical criteria. [IJCSI]. *International Journal of Computer Science Issues*, 8(2), 669–677.
- Mampadi, F., & Mokotedi, A. P. (2012). Towards effective combination of prior knowledge and cognitive styles in adaptive educational hypermedia systems. [IJET]. *International Journal of Emerging Technologies in Learning*, 7(3), 11–18. doi:10.3991/ijet.v7i3.2079
- Medina-Medina, N., Molina-Ortiz, F., & García-Cabrera, L. (2011). Adaptation and user modeling in hypermedia learning environments using the SEM-HP model and the JSEM-HP tool. *Knowledge and Information Systems*, 29(3), 629–656. doi:10.1007/s10115-010-0357-1
- Mulwa, C., Lawless, S., Ghorab, M. R., O'Donnell, E., Sharp, M., & Wade, V. (2011). A framework for the evaluation of adaptive information retrieval systems through implicit recommendation. In S. Andrews, Polovina, S., Hill, R. and Akhgar, B. (Ed.), *Proceedings of the International Workshop on Task Specific Information Retrieval, TSIR 2011, at the 19th International Conference on Conceptual Structures* (Vol. 6828/2011, pp. 366-374). University of Derby, England: Springer.
- O'Donnell, E., Mulwa, C., Sharp, M., & Wade, V. (2013). Web-mediated education and training environments: a review of personalised interactive e-learning resources. In E. McKay (Ed.), *ePedagogy in Online Learning: New Developments in Web Mediated Human Computer Interaction*. Hershey, New York: IGI Global. doi:10.4018/978-1-4666-3649-1.ch012
- O'Donnell, E., Sharp, M., Wade, V., & O'Donnell, L. (2013). Challenges encountered in creating personalised learning activities to suit students learning preferences. In Y. Kats (Ed.), *Learning Management Systems and Instructional Design: Best practices in online education* (pp. 263–287). Hershey, Pennsylvania, USA: IGI Global. doi:10.4018/978-1-4666-3930-0.ch014
- O'Donnell, E., Sharp, M., Wade, V., & O'Donnell, L. (2014). Personalised e-learning: the assessment of students' prior knowledge in higher education. In V. Wang (Ed.), *Handbook of research on education and technology in a changing society*. Hershey, New York: IGI Global. doi:10.4018/978-1-4666-6046-5.ch055
- Somyürek, S. (2015). The new trends in adaptive educational hypermedia systems. *International Review of Research in Open and Distance Learning*, 16(1). doi:10.19173/irrodl.v16i1.1946
- Souhaib, A., Mohamed, K., Kamal Eddine, E. K., & Ahmed, I. (2010). Adaptive hypermedia systems for e-learning. [IJET]. *International Journal of Emerging Technologies in Learning*, 5(SI3), 47–51. doi:10.3991/ijet.v5s3.1287
- Steiner, C., Hillemann, E., Verpoorten, D., Kleinermann, F., Pekczynski, P., & O'Donnell, E. (2010). GRAPPLE (Generic Responsive Adaptive Personalized Learning Environment): Refinement and improvement of evaluation guidelines. Retrieved from <http://www.grapple-project.org/public-files/deliverables/D8.2b-WP8-Evaluation-Guidelines-v1.0.pdf>

Yang, J., Huang, Z. X., Gao, Y. X., & Liu, H. T. (2014). Dynamic learning style prediction method based on a pattern recognition technique. *IEEE Transactions on Learning Technologies*, 7(2), 165–177. doi:10.1109/TLT.2014.2307858

KEY TERMS AND DEFINITIONS

Adaptive Education (AE): The purpose of Adaptive Education (AE) is to provide learners with learning resources that have been specially selected to suit the specific learning requirements of each individual.

Adaptive Educational Hypermedia (AEH): Electronic content which is used in the provision of adaptive education.

Adaptive Educational Hypermedia Systems (AEHS): Systems that provide educators with

the appropriate toolset to present learners with educational resources that have been specifically selected to suit their individual learning requirements. AEHS are designed and developed to deliver adaptive educational experiences to students.

E-Learning: Facilitating teaching and learning through the use of technology and access to the Internet.

Hypermedia: Electronic content which includes links to many different mediums of content, such as: text; tables; figures; graphics; images; audio; video; animations; simulations; interactive games.

Hypertext: A section of online text or an online paragraph of information that has been embedded with links to other content.

Technology Enhanced Learning (TEL): Learning which is enhanced through the use of technology and the Internet.

Computational Thinking in Innovative Computational Environments and Coding

Alberto Ferrari

University of Parma, Italy

Agostino Poggi

University of Parma, Italy

Michele Tomaiuolo

University of Parma., Italy

INTRODUCTION

The concept of Computational Thinking has been discussed for several decades and in recent years has been brought to the attention of the scientific community by Jeanette Wing. Her article presents Computational Thinking as “a way of solving problems, designing systems, and understanding human behavior that draws on concepts fundamental to computer science.” (Wing, 2006).

CT is a cognitive process involving logical reasoning that embraces the ability to think algorithmically using abstraction, decomposition and generalization.

The importance of CT places it among the basic skills for 21st century, together with reading, writing and calculation, that every person will have to master, so it is important to teach it already in primary school. As the invention of printing facilitated the spread of the three Rs (reading, writing and arithmetic) technology must lead to the spread of CT.

In our high-technological society teaching CT concepts in all level of instruction allows individuals to participate more equitably in society overcoming the differences now present in mastering these skills. Everyone should be able to apply computational strategies in each domain and understand what problems may be treated automatically.

CT is drawing fundamentally on concepts from computer science, it is not programming, but programming, meant as analysis and solution of problems, allows to highlight all CT features. Easy-to-use computational environments foster students in their first approach to coding and their use are becoming more common in school and university.

This chapter will introduce the research on CT and, in particular, the works on innovative computational environments, and will describe the situation of the education to CT in high school and in academic courses.

BACKGROUND

There is no single definition of CT and the following is the work of several authors who point out different aspects. A common feature is that CT represents a set of skills that are part of a cognitive process related to deal and seek solutions to problems. Abstraction, decomposition and generalization are common features of most of these works.

The term “Computational Thinking” was adopted for the first time by Seymour Papert, with reference to the LOGO programming language (Papert, 1996). According to his theory of “constructionism”, programming is a valu-

able educational tool that provides the cognitive artifacts necessary to the human mind to build a representation of the world with which it interacts.

“Think like a computer scientist” means being able to deal with a problem, to design a system and understand human behavior using the fundamental concepts and tools of Computer Science (CS). The power of our “mental” tools is amplified by the power of our “metal” tools (Wing, 2008).

There is still confusion over an acceptable definition for the term, in its essence, CT can be viewed as a re-foundation of “algorithmic thinking”. In the 1950s and 1960s, algorithmic thinking was understood as a mental attitude, enabling the description of a generic problem as a translation of some input data to output data, and the formalization of the required translation as an algorithm. CT is built on the same basis, with the inclusion of the ability to think about a complex problem at different levels of abstractions, to use mathematical methods, and to analyze data and the complexity of solutions, i.e., to study how a certain algorithm scales when the size of the problem grows.

CT can be defined as the ability to solve complex problems by applying the logic of computing paradigm; it is a set of cognitive skills, concepts and techniques of computer science related to problem solving.

In the 2010 workshop on “The Scope and Nature of Computational Thinking” participants developed a vivid discussion on what “CT for everyone” might mean. Wing describes CT as a bridge between science and engineering. Since it deals with thinking processes and abilities that can be applied to different disciplines, it represents a sort of meta-science. Moursund et al. (National Research Council, 2010) suggest a close relation between CT and Procedural Thinking, as developed by Seymour Papert in *Mindstorms* (Papert, 1980).

CT revolution goes much deeper than the use of computer in everyday work; it is changing the way we think. Among the features of the CT we find the ability to formulate a problem and

represent its solution writing an algorithm and compare the solution with others in order to assess its efficiency.

Another important feature is the ability to represent and organize data logically, using abstraction, generalization and modeling concepts, and identifying patterns within these.

Data abstraction, identification of common features and functionality comes into play both when we have to decide what to abstract and when we have to define the level of abstraction. The process of abstraction must then be followed by a process of analysis to verify the correctness of the assumptions and the quality of the result obtained.

Abstraction is probably the most significant process in CT: abstraction is used to define patterns, to generalize, to allow an object to represent many; abstraction allows you to extract common properties, to scale and then to deal with complexity; abstractions are the mental tools of computing.

Decomposition, the ability to decompose a problem into sub-problems of smaller size, is another important aspect. Decomposing a problem usually leads to the recognition of patterns and generalization, and therefore it also leads to the ability to design algorithms. Pattern recognition, the ability to identify similarities or common differences, is the basis of problem solving and algorithm design.

Generalization of pattern and abstraction allows to represent an idea or a process in general terms and then to be able to use this idea to solve other problems of the same nature. Algorithm design is finally the ability to develop a step-by-step solution for a given problem.

Bundy (2007) focuses, among other things, on the ability to process large amounts of data, Big Data, and describe a new model of science that defines e-science.

As we have seen, there are various definitions of CT and each one highlights certain features. A common factor among the various interpretations is the ability to solve problems using a systematic approach: problem solving.

The attitudes to be developed in order to reach the proposed objectives include the confidence to manage complexity and to deal with complex problems, the ability to address open problems and handle situations of ambiguity and the practice of teamwork and the burden of negotiating with colleagues to evaluate different proposals.

In the principal notes of the website “Exploring Computational Thinking” (Google, 2016a) Google asserts that CT is essential to the development of computer applications but at the same time it is an important factor to support problem solving across other scientific disciplines including math, science, and the humanities. Skills of CT allow students to identify a relationship between subjects as well as between school and life outside of the classroom.

CT concepts are present in various aspects that characterize Computing as a science and as a discipline. The concepts, techniques and informatics principles lend themselves as a starting point to educate to CT.

With the development of the CT trend, other views emerged. In particular, Denning (2009) argues that the term is neither unique nor representative of the whole of CS; he underlines the necessity of distinguishing CS from CT; the latter in fact includes only a part of the larger set of skills and knowledge possessed by a computer scientist.

After defining the fundamental principles and core practices of CS Denning ranks CT not among the fundamental principles but in the practices, and he defines it as a lens for looking at the world and interprets it as an algorithmic transformation from input to output information. According to Denning (1989, 2003), CS is a combination of engineering, mathematics, and science. In its essence, it can be described as the science of information processes.

CT can also be used to support problem solving across many disciplines, including math, science, and the humanities and is influencing research in nearly all disciplines, both in the sciences and the humanities (Bundy, 2007)

The term CT has also been used in the past and already in the 1980’s many scientists agreed

as computing was become a third leg of science in addition to theory and experiment so that the computation was essential for the development of the sciences.

Many are the examples of CT application in fields such as biology, chemistry, physics, Perkovic and Settle (2010) present examples of CT in more general terms and in all academic disciplines. Today we can talk about e-science, science that operates in distributed network environments and that uses high-performance computer on big data sets.

CT puts in relation the approach to the problems with the objective of proposing a solution that can be implemented on computer tools. The implementation is therefore closely linked to the use of programming languages. Coding is therefore one of the tools which are best suited to the teaching of the CT concepts.

INNOVATIVE COMPUTATIONAL ENVIRONMENTS

CT is not programming, but programming languages and coding are excellent vehicles for gaining access to basic concepts of CT.

Coding promotes the procedural thought and the process of decomposition, concepts that can be applied also in different contexts. Coding is only one aspect of CT but it is important as a medium to face fundamental concepts as abstraction, decomposition and the algorithmic approach to problem solving (Papert, 1980).

Villani (2015) underlines the value of coding as a basic educational discipline and describes programming languages and mathematics as “languages that help man in his struggle for the understanding of the world”.

Table 1 presents a list of computational environments adopted in project whose purpose is to introduce CT concepts through programming and problem solving.

Already in the 60’s Papert at MIT develops LOGO, a very powerful language that allows

Table 1. Computational environments

Project	Features	Interaction
LOGO	Easy-to-use programming language	Text
Scratch	Design and invent stories. Graphical manipulation of elements	Blocks
Snap!	Recursive blocks	Blocks
Alice	Storytelling in 3D environment	Sprites
App Inventor	Multimedia applications for Android devices	Blocks
Blockly	API. Engine for block application	Blocks

experts to create complex applications, but at the same time has a low threshold of access that allows its use even by primary school children. Mathematics, geometry and logic principles are learned easily through the turtle graphics; turtle, in the first version, was a robot moving on a surface via commands given through a computer. The spread of language and its use in education have a sharp increase in the years '80 thanks to versions of LOGO where the turtle was moving on the screen of a personal computer. Papert (1980), talking about his concept of constructionism, asserts that computer is an important simulation tool and that it is very important in education because it provides the fundamental cognitive artifacts for learning.

According to Papert program languages should be easy to get started ("low floor") and must give opportunities to develop complex projects ("high ceiling").

After a first period of great clamor in introducing coding concepts in education, the interest has gradually waned probably due to environments and languages too difficult to use in this context.

In recent years, there has been a revitalization of interest in introducing programming concepts since the early years of school. Consequently, to facilitate teachers and students, many projects started.

Scratch is developed by Michael Resnick and the Lifelong Kindergarten Group at the MIT Media Lab. Resnick (2009) presents Scratch as a tool for design, create, invent. Digital natives in most cases read but not write technology, so they need

to learn some type of programming to become creators of technology. In Scratch, programming is done by graphical manipulation of elements and not through the classical textual syntax. Users compose scraps of code, script, selecting predefined blocks and connecting them together to form a puzzle that represents the algorithmic solution of the problem. The ease of interaction with the environment makes it also suitable for children, but its use has spread in information technology introductory courses for students in high school and university (Brennan, 2013). Scratch sprites are an example of computer objects with various kind of characteristics, not only graphics ones. Characteristics and behaviors are defined by scripts and in Scratch is also present a simple message exchange mechanism. (Maloney et al., 2010).

Snap! is another programming environment that extends Scratch functionality allowing users to build new blocks (formerly his name was Build Your Own Blocks) and adding new capabilities, such as recursive blocks, which make it an environment aimed at a more advanced target such as high school students or university freshmen.

Snap! is implemented in JavaScript and this is a favorable aspect because it does not require any software installation as it works within any browser (Garcia et al., 2012).

Alice (Alice, 2016) is a 3D programming environment designed to be a student's first exposure to object-oriented programming. It allows students to learn fundamental programming concepts dragging and dropping graphic tiles to create simple

video games (Cooper et al., 2000). Bishop-Clark (2007) studied students using Alice in an introductory computing courses and found that they learn the programming concepts and gain a better appreciation of the complexity of programming compared with ones that start coding using text based languages.

The examples presented, like most projects for the diffusion of CT through coding, are based on graphical programming languages that give users the ability to build programs by composing graphical blocks together, to form a puzzle that represents the algorithmic solutions of problems. This approach, with respect to text-based programming, enables users to avoid issues of syntax and focus on the critical processes of designing, creating, and inventing.

The founding idea of block programming is to provide a graphical interface with blocks of diverse types. The user/programmer can combine the blocks in various ways through drag and drop operations, in such a way to form a puzzle representing the solution to a given problem. The blocks represent the basic elements of the language and they are differentiated by form and color, to easy their identification and usage.

Novice programmers must fight two battles in writing their first programs: the logical battle and the syntax battle. In environments based on block programming is impossible to introduce syntactic errors and this allows users to focus entirely on the logics of assigned problems and their solutions. In fact, the composition of blocks is rigidly constrained by existing slots, which represent the syntactic constraints of the language.

Most of initiatives prior described are directed specially toward primary school and they are based on ludic activities. The MIT App Inventor project (Wolber et al., 2011) is instead directed at a more advanced target. It is a block-based programming tool that enables the creation of simple multimedia application that can be executed on Android devices.

The Google Blockly environment (Google, 2016b) is built along principles similar to those

of the most widespread graphical language like Scratch (Resnick et al., 2009; Maloney et al., 2010), SNAP! (Garcia et al., 2012), MIT App Inventor (Wolber et al., 2011). Blockly is the engine running underneath many of projects that today propose block programming as the simple environment to introduce people of all ages to programming. There are now hundreds of educational projects that use Blockly library, the most important is probably code.org (Code.org, 2016): “Every student in every school should have the opportunity to learn Computer Science” is the opening sentence of this important project that involves tens of millions of young people (Wilson, 2014).

Although most educational projects and experiences proposing a block programming approach are oriented to a very young audience, projects directed toward high school students, like previous presented APP Inventor, are beginning to appear; in these cases, block programming is intended to help students get over the initial learning curve so that they can use more easily text-based programming languages.

Weintrop and Wilensky (2015, 2015b) compares conceptual understanding in block-based and text-based programming and through cognitive interviews and surveys to students found blocks-based programming to be easier because of the shapes and colors of the blocks, the drag-and-drop composition mechanism, and the ease of browsing the blocks library. Some points of strength and weakness are analyzed, as they are perceived by high school students which are allowed to move from a textual programming environment to a block-based one. Students identify the drag and drop composition interaction and the ease of browsing the language as contributing to the perceived ease-of-use of tools based on blocks-based programming. Being less powerful is indicated on the other end as drawback to blocks-based programming compared to the conventional text-based approach.

Price and Barnes (2015) in their study claim that block programming interfaces compared with text based ones can significantly improve novice

performance and understanding of fundamental programming concepts.

Starting from Papert's idea concerning programming languages, the evolution of the block programming can lead to broadening the horizon for applications of this type of language.

Ferrari et al. (2016) present a project that extends this visual paradigm to Object Oriented Programming.

Although, for complex applications, text-based languages will be more comfortable to use than those based on graphical syntax. Writing large programs in a visual programming environments is cumbersome, the same Google Blockly authors argue: "Please do not attempt to maintain the Linux kernel using Blockly".

After overcoming the first coding difficulties and when the programming concepts have been mastered, students should migrate to a conventional text-based language. For this reason, it is important already in the early examples to show the textual code, expressed in a conventional programming language, that correspond to the puzzle of blocks that represent the algorithmic solution of the problem.

Block-based programming environments are proposed as alternative way of teaching coding for novice programmers and are becoming increasingly common in introductory programming courses.

In addition to the discussion on the graphical or textual programming there is an important and unanswered question on which programming language is best for novices and what features make a language more or less accessible to beginners. Stefik and Hanenberg (2014) refuse the possibility of "One Language to Rule Them All" because of the variety of human perception and problem domains and even a perfect language, if it were created, may not be adopted. They present "the programming language wars" as a social ill causing problems in Computer Science discipline especially for new programmers.

COMPUTATIONAL THINKING IN HIGH SCHOOL AND ACADEMIC COURSES

Today, in schools and universities, only few specific courses are related to CT concepts. These concepts are, or should, be treated in "Information Education" courses.

Unfortunately for a long time, especially in high schools, CS has been considered only from the point of view of computer literacy, i.e., the ability to interact with computer systems and use specific programs.

The situation of Informatics education in Europe is not ideal. The title of a recent study, edited by ACM, is quite evocative: "Informatics education: Europe cannot afford to miss the boat" (ACM, 2013). It starts from the consideration that both governments and citizens in EU are conscious of living in a so-called "Information Society". But too often they satisfy themselves with the notion that learning to use digital media (i.e., digital literacy) is enough. Instead, digital literacy has to be considered as a very basic practical skill, and not an adequate intellectual tool for facing the new challenges.

To distinguish the more basic skills from the real science behind IT, the report adopts the term "Informatics", which is already common in continental Europe, but is acquiring the meaning of "Computer Science" also in the Anglo-Saxon world. Informatics is a scientific discipline that was born before actual computing machines, with the studies of Turing and Church in the 1930s. In the following decades, with its development into a full-fledged discipline, Informatics became an interesting mix of mathematical theory and electrical engineering, together with many seminal concepts of its own, including: algorithms; concepts of performance and complexity; data structures; concurrency, parallelism, and distribution; formal languages; abstractions.

The report summarizes its conclusions in few main points, which essentially underline and explain the need of education in both digital literacy and Informatics. Informatics is described as “a major enabler of technology innovation, the principal resource for Europe’s drive to become an information society, and the key to the future of Europe’s economy”. While some good progress is being made in teaching digital literacy, nevertheless more emphasis should be put on ethical issues, for a proper use of information technology resources. Instead, Informatics education, is sorely lacking in most European countries, and the situation is depicted as being worse than in the 70s and 80s. Without rapid changes, Europe will never be a major developer of information technologies, but a mere consumer, thus harming both its economy and the education of its new generation of citizens.

For achieving the desired objectives, the report advances some guidelines for the definition of new curricula. Those should take into account two main principles: “Informatics education must not just dwell on imparting information to students” and “it must draw attention to aspects of informatics that immediately appeal to young students, to encourage interaction, to bring abstract concepts to life through visualization and animation; a typical application of this idea is the careful use of (non-violent) games” (Overmars, 2004; Vassilev & Mutev, 2014).

CS researchers and professionals supported the idea of a more widespread diffusion of Informatics as a scientific discipline, with dedicated hours at least in secondary schools.

According to many reforms, elements of CS should be introduced since the first year in both technical schools and applied sciences secondary schools. In practice, however, the situation is changing at a very slow pace and possibly not improving, even with respect to Informatics.

An important research report (Google & Gallup, 2015) reports that in U.S. many students, parents and also teachers do not properly distinguish between computer science activi-

ties and general computer literacy; courses that school administrators consider to be computer science often lack programming. But another study by Horizon Media states that Americans now view computer science as a basic skill. 86% of interviewed say knowing how to use a computer is “just as important as knowing how to read and write.”

The importance of introduction of computer science in high school is that students who learn computer science in high school are 6 times more likely to major in it and computing occupations are now the first source of new wages in U.S. (Bureau of Labor Statistics, 2016). For this reason, in 2016 the U.S. administration is investing 150 million dollars in tech education projects to launch innovative training and placement models to develop tech talent as a way to keep and create jobs in local economies.

FURTHER RESEARCH DIRECTIONS

To overcome the ambiguity that has often led to regard computer literacy as the main objective of CT, research in this area is directed towards projects that aim at easing the learning of its core skills, while reducing the difficulties that students face in achieving these goals. Emphasis is on concepts and methodological approaches to problem solving, while coding becomes more and more a medium to achieve these objectives, not the ultimate goal. New educational development environments, as well as programming languages proposed, aim to drastically reduce the syntactic difficulty and to facilitate the first approach to programming and design (Kölling, 2008; Resnick et al., 2009).

Specific courses about CT concepts will increasingly be separated by CS courses, CT will become a specific discipline even if linked to more general IT aspects. In fact, ‘Computational thinking is influencing research in nearly all disciplines, both in the sciences and the humanities’ (Bundy, 2007).

CONCLUSION

The main aim of Computational Thinking is not to foster and simplify the development of applications. However, the proposals for new environments, tools and languages, which ease the first approach to programming, can be a starting point after which the other fundamental aspects of Computational Thinking, including abstraction and modeling, can be introduced. Their use will be specifically relevant in educational settings where students face at first the typical syntactic difficulties related with coding, but then they also have to deal with the more advanced challenges of design and testing.

REFERENCES

- ACM. (2013). *Informatics education: Europe cannot afford to miss the boat*. Retrieved July 18, 2016, from <http://europe.acm.org/iereport/ie.html>
- Alice. (2016). *Alice Software Website*. Retrieved July 18, 2016, from <http://www.alice.org>
- Bishop-Clark, C., Courte, J., Evans, D., & Howard, E. (2007). A quantitative and qualitative investigation of using Alice programming to improve confidence, enjoyment, and achievement among non-majors. *Journal of Educational Computing Research*, 37(2), 193–207. doi:10.2190/J8W3-74U6-Q064-12J5
- Brennan, K. (2013). Learning computing through creating and connecting. *Computer*, 46(9), 52–59. doi:10.1109/MC.2013.229
- Bundy, A. (2007). Computational Thinking is pervasive. *Journal of Scientific and Practical Computing*, 1(2), 67–69.
- Bureau of Labor Statistics. (2016). *Employment Projections*. Retrieved July 18, 2016, from <http://www.bls.gov/emp/tables.htm>
- Code.org. (2016). *Code.org Website*. Retrieved July 18, 2016, from <http://www.code.org>
- Cooper, S., Dann, W., & Pausch, R. (2000). Alice: A 3-D tool for introductory programming concepts. *Journal of Computing Sciences in Colleges*, 15, 107–116.
- Denning, P.J. (2003). Great principles of computing. *Communications of the ACM*, 46(11), 15–20. doi:10.1145/948383.948400
- Denning, P.J. (2009). The profession of IT Beyond Computational Thinking. *Communications of the ACM*, 52(6), 28–30. doi:10.1145/1516046.1516054
- Denning, P.J., Comer, D. E., Gries, D., Mulder, M. C., Tucker, A., Turner, A. J., & Young, P. R. (1989). Computing as a discipline. *Communications of the ACM*, 32(1), 9–23. doi:10.1145/63238.63239
- Ferrari, A., Poggi, A., & Tomaiuolo, M. (2016). Object Oriented Puzzle Programming. *Mondo Digitale*, 64. Retrieved July 18, 2016, from http://mondodigitale.aicanet.net/2016-3/Didamatica-Sessioni/Programmazione/paper_39.pdf
- Garcia, D., Segars, L., & Paley, J. (2012). Snap! (build your own blocks): Tutorial presentation. *Journal of Computing Sciences in Colleges*, 27(4), 120–121.
- Google. (2016a). *Exploring Computational Thinking Website*. Retrieved July 18, 2016, from <http://www.google.com/edu/computational-thinking/>
- Google. (2016b). *Blockly Software Website*. Retrieved July 18, 2016, from <https://developers.google.com/blockly/>
- Google & Gallup. (2015). *Images of Computer Science: Perceptions Among Students, Parents and Educators in the U.S.* Retrieved July 18, 2016, from <https://services.google.com/fh/files/misc/images-of-computer-science-report.pdf>
- Kölling, M. (2008). Using BlueJ to introduce programming. In *Reflections on the Teaching of Programming* (pp. 98–115). Berlin, Germany: Springer. doi:10.1007/978-3-540-77934-6_9

- Maloney, J., Resnick, M., Rusk, N., Silverman, B., & Eastmond, E. (2010). The Scratch Programming Language and Environment. *ACM Transactions on Computing Education*, 10(4), 16. doi:10.1145/1868358.1868363
- National Research Council. (2010). *Report of a Workshop on the Scope and Nature of Computational Thinking*. The National Academies Press.
- Overmars, M. (2004). Teaching Computer Science through Game Design. *Computer*, 37(4), 81–83. doi:10.1109/MC.2004.1297314
- Papert, S. (1980). *Mindstorms: Children, Computers, and Powerful Ideas*. New York, NY: Basic Books.
- Papert, S. (1996). An exploration in the space of mathematics educations. *International Journal of Computers for Mathematical Learning*, 1(1), 95–123. doi:10.1007/BF00191473
- Price, T. W., & Barnes, T. (2015). Comparing Textual and Block Interfaces in a Novice Programming Environment. In *Proceedings of the eleventh annual International Conference on International Computing Education Research* (pp. 91–99). ACM. doi:10.1145/2787622.2787712
- Resnick, M., Maloney, J., Monroy-Hernández, A., Rusk, N., Eastmond, E., Brennan, K., & Kafai, Y. et al. (2009). Scratch: Programming for All. *Communications of the ACM*, 52(11), 60–67. doi:10.1145/1592761.1592779
- Settle, A., & Perkovic, L. (2010). *Computational thinking across the curriculum: a conceptual framework*. College of Computing and Digital Media, DePaul University. Retrieved July 18, 2016, from <http://compthink.cs.depaul.edu/FinalFramework.pdf>
- Stefik, A., & Hanenberg, S. (2014). The programming language wars: Questions and responsibilities for the programming language community. In *Proceedings of the 2014 ACM International Symposium on New Ideas, New Paradigms, and Reflections on Programming & Software* (pp. 283–299). ACM. doi:10.1145/2661136.2661156
- Vassilev, T. I., & Mutev, B. I. An Approach to Teaching Introductory Programming Using Games. *Proceedings of International Conference on e-Learning*, 246–253. doi:10.4018/ijhcitp.2015010103
- Villani, C. (2015). *Birth of a Theorem: a mathematical adventure*. Macmillan.
- Weintrop, D., & Wilensky, U. (2015). Using commutative assessments to compare conceptual understanding in blocks-based and text-based programs. In *Proceedings of the eleventh annual International Conference on International Computing Education Research* (pp. 101–110). ACM. doi:10.1145/2787622.2787721
- Weintrop, D., & Wilensky, U. (2015). To block or not to block, that is the question: students' perceptions of blocks-based programming. In *Proceedings of the 14th International Conference on Interaction Design and Children* (pp. 199–208). ACM. doi:10.1145/2771839.2771860
- Wilson, C. (2015). Hour of code---a record year for computer science. *ACM Inroads*, 6(1), 22–22. doi:10.1145/2723168
- Wing, J. M. (2006). Computational Thinking. *Communications of the ACM*, 49(3), 33–35. doi:10.1145/1118178.1118215
- Wing, J. M. (2008). Computational thinking and thinking about computing. *Philosophical transactions of the royal society of London A: mathematical, physical and engineering sciences*, 366(1881), 3717–3725.
- Wolber, D., Abelson, H., Spertus, E., & Looney, L. (2011). *App Inventor*. O'Reilly Media.

ADDITIONAL READING

- Kelleher, C., & Pausch, R. (2005). Lowering the Barriers to Programming: A Taxonomy of Programming Environments and Languages for Novice Programmers. *ACM Computing Surveys*, 37(2), 83–137. doi:10.1145/1089733.1089734

Lye, S. Y., & Koh, J. H. L. (2014). Review on teaching and learning of computational thinking through programming: What is next for K-12? *Computers in Human Behavior*, 41, 51–61. doi:10.1016/j.chb.2014.09.012

National Research Council. (2011). *Report of a Workshop on the Pedagogical Aspects of Computational Thinking*. The National Academies Press.

Scott, J., & Bundy, A. (2015). Creating a new generation of computational thinkers. *Communications of the ACM*, 58(12), 37–40. doi:10.1145/2791290

Shein, E. (2014). Should everybody learn to code? *Communications of the ACM*, 57(2), 16–18. doi:10.1145/2557447

KEY TERMS AND DEFINITIONS

Algorithmic Thinking: A method for solving problems based on the clear definition of the steps needed.

Coding: The process of writing, testing, debugging and maintaining the source code of software applications.

Computational Thinking: A method for solving problems mainly used for the development of computer applications, but that can also be used to support problem solving across all academic disciplines.

Constructivist Theory: An active process in which learners construct new ideas or concepts based upon their current/past knowledge.

E-Science: A computationally intensive science that either is accomplished in massive distributed network environments, or uses large data sets whose processing required distributed computing infrastructures.

Graphical Programming Language: A programming language in which the source code is itself graphical in nature and does not mainly consist of text. It is also known as visual programming language.

Computer Agent Technologies in Collaborative Learning and Assessment

Yigal Rosen

Harvard University, USA

INTRODUCTION

In recognition of the importance of collaborative and problem solving skills, educators are realizing the need for effective and scalable learning and assessment solutions to promote the skillset in educational systems. In the settings of a comprehensive collaborative problem solving assessment, each student should be matched with various types of group members and must apply the skills in varied contexts and tasks. One solution to these assessment demands is to use computer-based (virtual) agents to serve as the collaborators in the interactions with students. The chapter presents the premises and challenges in the use of computer agents in the assessment of collaborative problem solving and describes how human and computer agent collaborative assessments are used in international learning and assessment project Animalia.

BACKGROUND

Collaborative problem solving is recognized as a core competency for college and career readiness. Students emerging from schools into the workforce and public life will be expected to work in teams, cooperate with others, and resolve conflicts in order to solve the kinds of problems required in modern economies. They will further need to be able to use these skills flexibly with various group compositions and environments (Care, & Griffin, 2014; Griffin, Care, & McGaw, 2012; O'Neil, & Chuang, 2008; Rosen, & Rimor, 2012). Educational programs have focused to a greater

extent on the advancement of learning and the assessment of collaborative problem solving as a central construct in theoretical and technological developments in educational research (National Research Council, 2011, 2013; OECD, 2013a). Collaborative skills are included within the major practices in the 2014 U.S. National Assessment of Educational Progress (NAEP) Technology and Engineering Literacy (National Assessment Governing Board, 2013). In this assessment program, students are expected to show their ability in collaborating effectively with computer-based (virtual) peers and experts and to use appropriate information and communication technologies to collaborate with others on the creation and modification of knowledge products. Similarly, the Israeli national program of adopting the educational system to the 21st century illustrates a multi-year program with the goal of leading the implementation of innovative pedagogy and assessment in schools, including collaboration, communication, and problem solving (Israel Ministry of Education, 2011). Collaborative problem solving is one of the areas that the Organisation for Economic Co-operation and Development (OECD) emphasized for major development in the Programme for International Student Assessment (PISA) in addition to scientific, math, and reading literacy for the 2015 assessment. Collaborative problem solving refers to problem solving activities that involve collaboration among a group of individuals (O'Neil, Chuang, & Baker, 2010; Zhang, 1998). In the PISA 2015 Framework (OECD, 2013b), collaborative problem solving competency is defined as "the capacity of an individual to effectively engage in a process whereby two or more

DOI: 10.4018/978-1-5225-2255-3.ch209

agents attempt to solve a problem by sharing the understanding and effort required to come to a solution and pooling their knowledge, skills, and efforts to reach that solution” (p. 6). This definition; treats the competency as conjoint dimension collaboration skills and the skills needed to solve a problem. For the PISA assessment, the focus is on individual capacities within collaborative situations. Thus, the effectiveness of collaborative problem solving depends on the ability of group members to collaborate and to prioritize the success of the group over individual successes. At the same time, this ability is still a trait in each of the individual members of the group. Development of a standardized computer-based assessment of collaborative problem solving skills, specifically for large-scale assessment programs, remains challenging. Unlike some other skills, collaborative problem solving typically requires using complex performance tasks, grounded in varied educational domains, with interaction among students. These factors can affect the level of control that can be applied to ensure accurate assessment of students.

In this chapter, an operational definition of collaborative problem solving refers to “the capacity of an individual to effectively engage in a group process whereby two or more agents attempt to solve a problem by sharing knowledge and understanding, organizing the group work and monitoring the progress, taking actions to solve the problem, and providing constructive feedback to group members.”

COMPUTER AGENT TECHNOLOGIES

Collaboration can take many forms, ranging from two individuals to large teams with predefined roles. For assessment purposes, collaboration can also be performed using simulated agents playing the role of team members, using computer or humans as team members. Thus, a critical distinction is whether all team members are human or some are computer agents. There are advantages and limita-

tions for each method, which are outlined below. The Human-to-Human (H-H) approach provides an authentic human-human interaction that is a highly familiar situation for students. Students may be more engaged and motivated to collaborate with their peers. Additionally, the H-H situation is closer to the collaborative problem solving situations students will encounter in their personal, educational, professional, and civic activities. However, because each human will act independently, the approach can be problematic because of individual differences that can significantly affect the collaborative problem solving process and its outcome. Therefore, the H-H assessment approach of collaborative problem solving may not provide sufficient opportunity to cover variations in group composition, diversity of perspectives, and different team member characteristics in a controlled manner for accurate assessment of the skills on an individual level. Also, computer agent technology can contribute to efficiency in data collection by dramatically decreasing the assessment time with strategic dialogue management and rapid immersion in the collaborative context. Simulated team members using a preprogrammed profile, actions, and communication can potentially provide the coverage of the full range of collaboration skills with sufficient control. In the Human-to-Agent (H-A) approach, collaborative problem solving skills are measured by pairing each individual student with a computer agent or agents that can be programmed to act as team members with varying characteristics relevant to different collaborative problem solving situations. Group processes are often different depending on the task and could even be competitive. Use of computer agents provides a component of non-competitiveness to the collaborative problem solving situation, as it is experienced by a student. Additionally, if the time-on-task is limited, time spent establishing common ground or discussing non-task relevant work may lower group productivity. As a result of these perceived constraints, a student collaborating in H-H mode may limit significantly the extent to which collaborative problem solving dimensions,

E

such as shared understanding, are externalized through communication with the partner. The agents in H-A communication can be developed with a full range of capabilities, such as text-to-speech, facial actions, and optionally rudimentary gestures. In its minimal level, a conventional communication media, such as text via emails, chat, or a graphic organizer with lists of named agents can be used for H-A purposes. However, collaborative problem solving in H-A settings deviates from natural human communication delivery. The dynamics of H-H interaction (timing, conditional branching) cannot be perfectly captured with agents, and agents cannot adjust to idiosyncratic characteristics of humans. For example, human collaborators can propose unusual, exceptional solutions; the characteristic of such a process is that it cannot be included in a system following an algorithm, such as H-A interaction.

Research shows that computer agents have been successfully used for tutoring, collaborative learning, co-construction of knowledge, and collaborative problem solving (Biswas, Jeong, Kinnebrew, Sulcer, & Roscoe, 2010; Graesser et al., 2008; Millis et al., 2011). A computer agent can be capable of generating goals, performing actions, communicating messages, sensing its environment, adapting to changing environments, and learning (Franklin & Graesser, 1996). One of the examples for computer agent use in education is a teachable agent system called Betty's Brain (Biswas, Leelawong, Schwartz, & Vye, 2005; Leelawong, & Biswas, 2008). In this system, students teach a computer agent using a causal map, which is a visual representation of knowledge structured as a set of concepts and their relationships. Using their agent's performance as motivation and a guide, students study the available resources so that they can remediate the agent's knowledge and, in the process, learn the domain material themselves. Operation ARA (Cai et al., 2011; Millis et al., 2011) uses animated pedagogical agents that converse with the student in a game-based environment for helping students learn critical-thinking skills and scientific

reasoning within scientific inquiry. The system dynamically adapts the tutorial conversations to the learner's prior knowledge. These conversations, referred to as "trialogs" are between the human learner and two computer agents (student and teacher). The student learns vicariously by observing the agents, gets tutored by the teacher agent, and teaches the student agent.

A focused study has been conducted to investigate differences in student collaborative problem solving (CPS) performance in H-A and H-H modes (Rosen, 2014; Rosen, & Foltz, 2014; Rosen, & Mosharraf, 2015). Study participants included 179 students; age 14; from the United States, Singapore, and Israel. In all, 136 students participated in the H-A group and 43 participated in the H-H group (43 additional students participated in the H-H setting, acting as 'collaborators' for the major H-H group). Specifically in H-H assessment mode, students were randomly assigned into pairs to work on the CPS task. Because the H-H approach required pairs of students working together in a synchronized manner, the number of pairs was limited. This is due to the characteristics of technology infrastructures in participating schools. The students were informed prior to their participation in the study whether they would collaborate with a computer agent or a classmate. In a case of H-H setting, the students were able to see the true name of their partner. Students were exposed to identical collaborative problem solving assessment tasks and were able to collaborate and communicate by using identical methods and resources. However, while in the H-A mode students collaborated with a simulated computer-driven partner, and in the H-H mode students collaborated with another student to solve a problem.

The findings showed that students assessed in H-A mode outperformed their peers in H-H mode in their collaborative skills. Collaborative problem solving with a computer agent involved significantly higher levels of shared understanding, progress monitoring, and feedback. The results further showed no significant differences in other student performance measures to solve the problem

with a computer agent or a human partner, although on average students in H-A mode applied more attempts to solve the problem, compared to the H-H mode. A process analysis of the chats and actions of the students showed that in H-A group the students encounter significantly more conflict situations (e.g., disagreements) than in the H-H group. However, it was found that students in H-H setting were engaged in significantly more situations in which the partner proposed different solutions for a problem.

PROJECT ANIMALIA

An investigation to look further at the premises of human and computer agent technology in the context of learning and assessing collaborative problem solving skills was undertaken through collaboration between World ORT, the Ministry of Education in Israel, and researchers at Pearson. Animalia is an online mini-course designed to promote students' collaborative problem solving skills in the context of complex ecosystems (Bakken, Bielinski, Johnson, & Rosen, 2015). Animalia is a simulated fictitious village with a serious environmental problem. The fish in a local river are dying and the cause(s) of this problem is unknown. A team of scientists from an international organization (i.e. students) must determine the root cause of the problem and recommend sound, scientifically-based solutions.

The foundation of the problem-solving activity relies on understanding specific science concepts such as pH, oxygen saturation, nitrogen levels, the biological effects of pollutants, ecology, frog life cycles, algae physiology, as well as cause and effects in ecosystems. To create a more realistic scenario, team members will be given multiple and sometimes misleading, irrelevant, or contradictory pieces of information about what was happening in Animalia. It was up to the students to determine the authenticity of the data and to judge the credibility of the source. Each team member have had access to only some information about the overall

situation so that determining the root cause of the problems in Animalia require each team member to share his/her information with the others. Team members were encouraged to distill the information they receive into its essential components and to hypothesize what they think are the causes of the problem(s). Team members were also given feedback on each other's ideas and then form a consensus opinion. Finally, each team summarized findings and recommendations in a written report and an oral presentation. The teams described the cause(s) and provided recommendations to solve the environmental problems in Animalia. Learning was assessed through short quizzes and collaborative tasks. Additionally pre- and post-test of science concepts and collaborative skills were administered at the beginning and at the end of the mini-course. Collaborative problem solving proficiency was assessed at the beginning of Animalia through computer agent task ZooQuest (Rosen, 2014; Rosen, & Mosharraf, 2015).

The development team designed this mini-course to be completed in approximately 14 hours. Even though this is a collaborative task, due to the constraints of physical distance and time zone differences, students typically worked independently and communicate asynchronously (at different times) with their team members in other locations. The only two times students were online at the same time was when the teams discussed townspeople interview videos and gave their final presentations using Google Hangouts. Animalia content was provided in English, and English was spoken between partner schools. This gave students the real-world experience of collaborating with people from different cultures, with different levels of English proficiency, different levels of science knowledge, and across time zones. The content and related activities were provided online via a designated Moodle website. Each team have had its own "course" in the Moodle system which enabled team members to share information with each other, but prevented them from seeing information from other teams. Each session was available to students sequentially

during the project. When they logged in, students saw only those sessions that have been released up to that point. Across sites, teams were asked to stick to a common schedule so that all team members are progressing through the project work at the same pace.

Teachers play a key role in supporting this complex project. Teaching collaborative problem solving skills might be intimidating for teachers who typically deliver content-focused science instruction. But because these skills are fundamental to college and career success, teachers must help students improve upon these skills. Therefore, Animalia includes several professional development sessions to guide teachers on implementation of the course, as well as through the process.

Because students are working in teams from different time zones, teachers need to help coordinate schedules so that students are working on the same content at the same time. It is essential that teachers support and encourage their students as they progress through the sessions. Teacher's actions help students stay engaged and positive. There are times when the teacher will need to help your students troubleshoot any social or technical difficulties that may arise.

Student interaction within each group is necessary for successful completion of the task. Each student takes on a specific scientist role, and each role will gather its own unique information that must be shared with their team members to arrive at solutions or recommendations for the Animalia problem. It is important to emphasize this with students and be sure that each one understands what tools can be used to get connected and stay engaged.

Students must interact on discussion boards throughout the task. It is important that these discussions remain positive and on-task. Teachers monitor these discussions, and tell students so that they know you will be monitoring their progress. Students' awareness of their teacher's actions help prevent the posting of inappropriate or offensive remarks.

The study participants included 220 students, all 14 years old, from Israel, Italy, Bulgaria, Czech Republic, and Spain. The data were collected from February to April 2015. Recruitment of participating schools was achieved through collaboration with World ORT based on the following criteria: (a) population of 14-year-old students proficient in English, and (c) sufficient technology infrastructure (e.g., computers per student, high-speed Internet). In all, 108 students participated in low-CPS condition (i.e., low-CPS proficiency of all team members, as measured by CPS pre-test), and 112 participated in mixed-CPS condition (i.e., mid.-high CPS proficiency of two team members and low CPS proficiency among the other two team members). The students in both conditions were assigned to 4-person teams, two each from two different partner schools. No significant differences were found in science concepts pre-test scores between participants in the two groups. This similarity in student science knowledge allowed comparability of student results in Animalia between the two groups.

Preliminary findings indicated that students participated in a mixed-CPS condition significantly outperformed the students in low-CPS condition in their science concepts post-test, as well in the written report and oral presentation scores.

FUTURE RESEARCH DIRECTIONS

Team composition plays a significant role in collaborative settings (Kreijns, Kirschner, & Jochems, 2003; Nelson, 1999; Rosen, & Rimor, 2009). Collaborative problem solving performance is compromised to the extent that the division of labour is unintelligent, subgoals are not achieved, the group goals are blocked, and there are communication breakdowns. Collaborative problem solving tasks with high interdependency are very sensitive to group composition. One team member who has low competency can dramatically decrease the performance of the entire team and

force other team members to compensate in order to achieve team goals. An overly strong leader can prevent other team members from manifesting their talents. A meaningful collaborative interaction rarely emerges spontaneously, but requires careful structuring of the collaboration to promote constructive interactions. Incorporating collaborative problem solving assessment, as a tool for structuring student assignment to more optimal collaborative learning experiences such as Animalia is one example of potential use of computer agent based assessment. More research is needed to explore further these conditions and to more thoroughly understand their impact on outcomes. In this pilot, new insights were gained about collaborative problem-solving, but replications of this study will shed additional light and help to further inform the work of educators, researchers, and test developers who are motivated to help students discover and refine their own collaborative problem solving skills.

CONCLUSION

Students can demonstrate collaborative problem solving in tasks where assignments are distributed among team members, progress and results are integrated and shared, and products are presented jointly. Task structuring approaches aim to create optimal conditions for collaborative problem solving assessment by designing and scripting the situation before the interaction begins. It may include varying the characteristics of the participants (e.g., the size and composition of the group, or the roles), the availability and characteristics of communication and collaboration tools (e.g., the use of a phrase-chat or graphical tools), and the organization of the task. The collaborative problem solving assessment methods described in this chapter offer one of the few examples today of an approach in assessing collaborative problem solving skills. Collaborative assessments bring new challenges and considerations for the design of effective assessment approaches because they move the field beyond standard item

design tasks. The assessment must incorporate concepts of how humans solve problems in situations where information must be shared and must incorporate considerations of how to control the collaborative environment in ways sufficient for valid measurement of individual and team skills (Rosen, 2014). The quality and practical feasibility of these measures are not yet fully documented. However, these measures can rely on the abilities of technology to engage students in interaction, to simulate others with whom students can interact, to track students' ongoing responses, and to draw inferences from those responses. Group composition is one of the important issues in assessments of collaborative skills (Webb, 1995; Wildman et al., 2012). Overcoming possible bias of differences across groups by using computer agents or other methods becomes even more important within international large-scale assessments where cultural boundaries are crossed. New psychometric methods should be explored for reliable scoring of an individual's contribution to collaborative processes and solutions, such as stochastic and social network analyses, hidden Markov models, and Bayesian knowledge tracing (Soller, & Stevens, 2008; von Davier, & Halpin, 2013). Current research suggests that by using computer agents in a collaborative problem solving task, students are able to show their collaborative skills at least at the level of that of their peers who collaborate with human partners. Although human-to-agent interaction might not be regarded as equal to human-to-human collaboration, the advancing technology of computer agents makes the use of avatars a viable way to simulate collaboration, and this can offer researchers more control than is available with real human collaboration (OECD, 2013b; Rosen, & Wolf, 2014). However, each approach to assessment of collaboration still involves limitations and challenges that must be considered in the design of the assessments. Further research can continue to establish comprehensive validity evidence and generalization of findings both in H-A and H-H collaborative problem solving settings.

REFERENCES

- Bakken, S., Bielinski, Johnson, C., & Rosen, Y. (2015). Animalia: Collaborative science problem solving learning and assessment. In Y. Rosen, S. Ferrara, & M. Mosharraf (Eds.), *Handbook of Research on Technology Tools for Real-Life Skill Development* (pp. 360-384). Hershey, PA: Information Science, IGI Global.
- Biswas, G., Jeong, H., Kinnebrew, J. S., Sulcer, B., & Roscoe, A. R. (2010). Measuring self-regulated learning skills through social interactions in a teachable agent environment. *Research and Practice in Technology-Enhanced Learning*, 5(2), 123–152. doi:10.1142/S1793206810000839
- Biswas, G., Leelawong, K., Schwartz, D., & Vye, N. (2005). Learning by Teaching: A New Agent Paradigm for Educational Software. *Applied Artificial Intelligence*, 19(3-4), 363–392. doi:10.1080/08839510590910200
- Cai, Z., Graesser, A. C., Forsyth, C., Burkett, C., Millis, K., Wallace, P., & Butler, H. et al. (2011). Trialog in ARIES: User input assessment in an intelligent tutoring system. In W. Chen, & S. Li (Eds.), *Proceedings of the 3rd IEEE International Conference on Intelligent Computing and Intelligent Systems* (pp.429-433). Guangzhou: IEEE Press.
- Care, E., & Griffin, P. (2014). An approach to assessment of collaborative problem solving. Special Issue: Assessment in Computer Supported Collaborative Learning. *Research and Practice in Technology Enhanced Learning*, 9(3), 367–388.
- Graesser, A. C., Jeon, M., & Dufty, D. (2008). Agent technologies designed to facilitate interactive knowledge construction. *Discourse Processes*, 45(4), 298–322. doi:10.1080/01638530802145395
- Griffin, P., Care, E., & McGaw, B. (2012). The changing role of education and schools. In P. Griffin, B. McGaw, & E. Care (Eds.), *Assessment and Teaching 21st Century Skills* (pp. 1–15). Heidelberg: Springer. doi:10.1007/978-94-007-2324-5_1
- Israel Ministry of Education. (2011). *Adapting the educational system to the 21st century*. Ministry of Education.
- Kreijns, K., Kirschner, P. A., & Jochems, W. (2003). Identifying the pitfalls for social interaction in Computer-Supported Collaborative Learning environments: A review of the research. *Computers in Human Behavior*, 19(3), 335–353. doi:10.1016/S0747-5632(02)00057-2
- Leelawong, K., & Biswas, G. (2008). Designing Learning by Teaching Systems: The Betty's Brain System. *International Journal of Artificial Intelligence in Education*, 18(3), 181–208.
- Millis, K., Forsyth, C., Butler, H., Wallace, P., Graesser, A. C., & Halpern, D. (2011). Operation ARIES! A serious game for teaching scientific inquiry. In M. Ma, A. Oikonomou, & J. Lakhmi (Eds.), *Serious games and edutainment applications* (pp. 169–195). London: Springer-Verlag. doi:10.1007/978-1-4471-2161-9_10
- National Assessment Governing Board. (2013). *Technology and engineering literacy framework for the 2014 National Assessment of Educational Progress*. Washington, DC: National Assessment Governing Board.
- National Research Council. (2011). *Assessing 21st Century Skills*. Washington, DC: National Academies Press.
- National Research Council. (2013). *New directions in assessing performance of individuals and groups: Workshop summary*. Washington, DC: National Academies Press.

- Nelson, L. (1999). Collaborative problem-solving. In C. M. Reigeluth (Ed.), *Instruction design theories and models* (pp. 241–267). Mahwah, NJ: Lawrence Erlbaum Associates.
- O'Neil, H. F. Jr, & Chuang, S. H. (2008). Measuring collaborative problem solving in low-stakes tests. In E. L. Baker, J. Dickieson, W. Wulfeck, & H. F. O'Neil (Eds.), *Assessment of problem solving using simulations* (pp. 177–199). Mahwah, NJ: Lawrence Erlbaum Associates.
- O'Neil, H. F. Jr, Chuang, S. H., & Baker, E. L. (2010). Computer-based feedback for computer-based collaborative problem solving. In D. Ifenthaler, P. Pirnay-Dummer, & N. M. Seel (Eds.), *Computer-based Diagnostics and Systematic Analysis of Knowledge* (pp. 261–279). New York: Springer-Verlag. doi:10.1007/978-1-4419-5662-0_14
- OECD. (2013a). *OECD skills outlook 2013: First results from the survey of adult skills*. OECD Publishing.
- OECD. (2013b). *PISA 2015 Collaborative Problem Solving framework*. OECD Publishing.
- Rosen, Y. (2014). Comparability of conflict opportunities in human-to-human and human-to-agent online collaborative problem solving. *Technology, Knowledge and Learning*, 19(1-2), 147–174. doi:10.1007/s10758-014-9229-1
- Rosen, Y., & Foltz, P. (2014). Assessing collaborative problem solving through automated technologies. *Research and Practice in Technology Enhanced Learning*, 9(3), 389–410.
- Rosen, Y., & Rimor, R. (2009). Using collaborative database to enhance students' knowledge construction. *Interdisciplinary Journal of E-Learning and Learning Objects*, 5, 187–195.
- Rosen, Y., & Rimor, R. (2012). Teaching and assessing problem solving in online collaborative environment. In R. Hartshorne, T. Heafner, & T. Petty (Eds.), *Teacher education programs and online learning tools: Innovations in teacher preparation* (pp. 82–97). Hershey, PA: Information Science Reference, IGI Global.
- Rosen, Y., & Wolf, I. (2014). *Learning and Assessing Collaborative Problem Solving Skills*. Paper presented at the International Society for Technology in Education (ISTE) conference, Atlanta, GA.
- Soller, A., & Stevens, R. (2008). Applications of stochastic analyses for collaborative learning and cognitive assessment. In G. R. Hancock & K. M. Samuelson (Eds.), *Advances in latent variable mixture models* (pp. 109–111). Charlotte, NC: Information Age Publishing.
- Von Davier, A. A., & Halpin, P. F. (2013, December). *Collaborative problem solving and the assessment of cognitive skills: Psychometric considerations*. Research Report ETS RR-13-41. Educational Testing Service.
- Webb, N. M. (1995). Group collaboration in assessment: Multiple objectives, processes, and outcomes. *Educational Evaluation and Policy Analysis*, 17(2), 239–261. doi:10.3102/01623737017002239
- Wildman, J. L., Shuffler, M. L., Lazzara, E. H., Fiore, S. M., Burke, C. S., Salas, E., & Garven, S. (2012). Trust development in swift starting action teams: A multilevel framework. *Group & Organization Management*, 37(2), 138–170. doi:10.1177/1059601111434202
- Zhang, J. (1998). A distributed representation approach to group problem solving. *Journal of the American Society for Information Science*, 49(9), 801–809. doi:10.1002/(SICI)1097-4571(199807)49:9<801::AID-ASIS>3.0.CO;2-Q

ADDITIONAL READING

Avery Gomez, E., Wu, D., & Passerini, K. (2010). Computer-supported team-based learning: The impact of motivation, enjoyment and team contributions on learning outcomes. *Computers & Education*, 55(1), 378–390. doi:10.1016/j.compedu.2010.02.003

Baylor, A. L., & Kim, Y. (2005). Simulating instructional roles through pedagogical agents. *International Journal of Artificial Intelligence in Education*, 15, 95–115.

Brannick, M. T., & Prince, C. (1997). An overview of team performance measurement. In M. T. Brannick, E. Salas, & C. Prince (Eds.), *Team performance assessment and measurement: Theory methods and applications* (pp. 3–16). Mahwah, NJ: Lawrence Erlbaum Associates.

Graesser, A. C., Lu, S., Jackson, G. T., Mitchell, H., Ventura, M., Olney, A., & Louwerse, M. M. (2004). AutoTutor: A tutor with dialogue in natural language. *Behavior Research Methods, Instruments, & Computers*, 36(2), 180–193. doi:10.3758/BF03195563 PMID:15354683

Griffin, P., & Care, E. (Eds.). Assessment and Teaching of 21st Century Skills: Vol. 2. *Methods and Approach*. Dordrecht: Springer.

Rosen, Y., & Mosharraf, M. (2015). Computer agent technologies in collaborative assessments. In Y. Rosen, S. Ferrara, & M. Mosharraf (Eds.). *Handbook of Research on Technology Tools for Real-Life Skill Development* (pp. 319–343). Hershey, PA: Information Science, IGI Global.

Stahl, G. (2006). *Group cognition: Computer support for building collaborative knowledge*. Cambridge, MA: MIT Press.

KEY TERMS AND DEFINITIONS

Agent: Either a human or a computer-simulated participant in a collaborative problem solving group.

Collaboration: Coordinated, synchronous activity that is the result of a continued attempt to construct and maintain a shared conception of a problem.

Collaborative Problem Solving: The capacity of an individual to effectively engage in a group process whereby two or more agents attempt to solve a problem by sharing knowledge and understanding, organizing the group work and monitoring the progress, taking actions to solve the problem, and providing constructive feedback to group members.

Computer Agent: An avatar with a preprogrammed profile, actions and communication. Computer agents can be capable of generating goals, performing actions, communicating messages, sensing environment, adapting to changing environments, and learning.

Problem Solving: Cognitive processing directed at achieving a goal when no solution method is obvious to the problem solver.

Cost-Effective 3D Stereo Visualization for Creative Learning



R. S. Kamath

Chatrapati Shahu Institute of Business Education and Research, India

R. K. Kamat

Shivaji University, India

INTRODUCTION

Technology-enhanced learning which has become a reality with the pervasive penetration of Information and Communication Technology (ICT) in almost all the walks of higher learning is by now not a new concept, but is still quite new in many educational institutions and settings. The pragmatic view by many researchers, first hand entails, that the chalk-and-talk environment is being both less and less relevant and effective to 21st Century digital age students, and does little or no justice to the learning of academically underprepared students (Dongale, Patil & Kamat, 2015). Especially in the domain of 'Engineering Education' it is clearly evident that Progress in computers and various technologies have changed traditional methods for teaching. Previously dominated by simulation and animation, now the educators are realizing that both the above said tools and techniques alone cannot substantiate true real-sense learning for users. This explores need for more advanced technologies in order to improve learning. In this context, the VR technology has found numerous applications in the field of education. The growth dynamics VR area reveals it's market size to \$407.51 million which will encompass more than 25 million users by 2018 (Marketsandmarkets.com, 2015). The main challenge of VR technology, however is the exorbitant cost due to the inherent sophisticated hardware and software which inhibits its inculcation in the education paradigm. In the backdrop

of above, we present a cost effective 3D stereo visualization system conceived, designed and developed by us for creative learning in the most cost effective manner.

The chapter is structured as follows: We open up with a brief overview of technology inculcation in education, which showcases the gradual progression from simple simulation and animation techniques to more sophisticated ones like VR. We also present the very notion of VR for the benefit of the broad audience of the chapter. The focus then shifts to the VR tool we developed, its system architecture, technical features and cost effectiveness. The manuscript then actually portrays setting the experimental environment for VR based pedagogy and thereby highlights its potential role in presenting the insight in realization of experiential learning in different domains.

BACKGROAUND: PROGRESSION OF EDUCATION TECHONOLOGY FROM SIMULATION, ANIMATION TO VIRTUAL REALITY

With the embryonic digital age, there has been intense discussion all over the globe, particularly in the last decade about the use of technology for personalizing the learning environment. VR is the fascinating area in computer application research (Vafadar, 2013). In recent years, 3D technologies in modeling, printing and stereoscopic have symbolized the true cutting edge in educational

systems (Dalgarno et al., 2010). The use of 3D glasses, stereoscopic 3D content and virtual environments in all curriculum areas to improve 21st century teaching and learning has been the buzz word in all the spheres of academics (Alpaslan & Sawchuk, 2004). The 3D in the Classroom has been clearly the winner over its 2D counterpart in improving the teaching-learning in the classroom. An interesting account of all these developments at the global level has been summarized in the following paragraph.

Pedagogic experts believe that merely theoretical explanation without actual implementation makes learning experience invaluable. Instead of listening only to lectures, if students get real experience in a virtual wrapper can achieve the learning outcomes. In this context VR is real winner and gained immense popularity in the education spheres. It provides a visually appealing technique for presentation of teaching material. It motivates student community by encouraging active participation rather than passivity. For example, a computer-based flight simulator in which pilots can attain flying skills in the absence of a real airplane ought to instill the right kind of skills. Many studies have been conducted on the applications and effectiveness of virtual Reality in education and training. Studies show that a virtual environment can stimulate learning because of it's a tight coupling between illustrative and experiential information (Hamada, 2008). Yahaya incorporated immersive VR technology in creating learning environment (Yahaya, 2004). His investigations indicate that learners gets engaged in real world problems associated with VR environment and it really helps in gaining the subject understanding. Elomar has explained the use of VR technology in learning environment through the 'experiencing of real phenomena'; new educational possibilities by the integration of education with VR (Elomar, 2012).

Dalgarno et al. have explained applications of 3D immersive virtual worlds in education and its implementation in education institutions across New Zealand and Australia (Dalgarno et al., 2010). The authors of the above referred study

have discussed overall research design with results from the Australian/New Zealand perspective. The main finding of this study is the variety of ways of using 3D virtual worlds by academicians.

Yet another paper by Piovesan et al., explained the application of VR in education. This research presents educational software, which permits students to manipulate objects in 3D with a simple interface (Piovesan et al., 2012). The software referred in the research is based on VRML for the design of models and PHP for web publishing. Manseur has discussed the use of VR tools and its applications in science and engineering education (Manseur, 2005). In particular, the VRML for model design is presented and subsequently the development of visualization tools for education has been depicted. Researcher has explained the combination of VRML with other software tools to create interactive VR solutions to support teaching and learning.

Teaching with 3D technologies is attracted by students' community as confirmed from research from case studies around the world. Learning accompanied with 3D video, interactions and simulations in virtual environment is attaining significant increase in performance, retention, abstract concept mastery, and more.

Well known VR tools developed in this context are (engagingeducation.net, 2015):

- **3D Ladibug Document Camera:** A dynamic document camera that can show objects and manipulate them in 3D stereoscopic by using the 3D software and hardware
- **Presente 3D:** An add-on to Microsoft PowerPoint, using which students can attach 3D stereoscopic to their presentations
- **Kid Pix:** An image processing tool that can be used to create pictures and videos in 3D anaglyph
- **Hasbro My3D:** A nifty gadget that facilitates students to get the immersive experiences sensation like walking through the Solar System

- **3D Books:** Facilitating students understanding by both the pictures and the content in 3D

There is increasingly good number of domains apart from education where VR has been finding its applications. Diverse fields including psychology, medicine, neuroscience, and physical and occupational therapy, the ICT MedVR group explores and evaluates areas where VR can add value over traditional assessment and intervention approaches. Areas of specialization are in using VR for mental health therapy, motor and cognitive skills rehabilitation, assessment, and clinical skills training. (Medical Virtual Reality, 2015). VR is beginning to be used extensively in built environment education. 3D virtual environment can provide a rich, interactive and engaging educational context that supports experimental learning. (Nicholos et al., 2012) VR has also its impact in various fields like entertainment, manufacturing industries, military bodies in addition to education. In the field of education 3D graphics plays major role (Kamath & Kamat, 2013). 3D models are very useful to make acquainted students with features of various shapes and objects. Many games have been designed using 3D images that the user needs to intermingle with in order to gain knowledge of a certain lesson (Bowman et al., 2006). Interaction with models increases a user's curiosity and makes learning more excitement. Simulations and visualization of complex data can be extensively done by using VR. The VR Technology has the ability to assist in the teaching process too, enabling students to view and interact with concepts they are working in 3D immersive environment (Steinicke, 2009). The user makes use of VR hardware to move and explore within an onscreen virtual environment as if in reality moving within a place in the actual world. Though VR paves the way for enhancing the visualization ability thereby making its headway in growing number of niche domains in addition to education, still it has to overcome many challenges for its adoption as discussed in the following section.

Impediment in VR Adoption and Way Forward

E

As presented in previous section VR leads to number of benefits, but its adoption is easier said than done. Particularly its espousal in education is challenging due to number of factors such as the customization of the technology per learner's need, lack of apt skill set amongst the faculty who are mainly digital immigrant and catering the needs of the computationally intensive software centric applications with the equally capable hardware platform. However amongst the three main challenges, the cost barrier is the main factor inhibiting the growth of VR in education. We aimed towards the cost reduction without compromising the performance and designed a 3D visualization method, an important facet of VR for active learning (Kamath, et al., 2012). We attempted developing a cost effective technique for achieving VR based active learning environment. With a combination of inexpensive hardware and simple-to-use software, students can enjoy the excitement of eye-popping 3D visualizations by employing our software suite on their fairly low end personal computers. The visualization solution developed is capable to browse the ASCII files containing the details of objects and the corresponding 3D scenes are rendered. 3D stereo visualization is a prime part of our software suite that accounts for virtual reality. This visualization solution offers inexpensive 3D interface to learner. The details of the software suite are presented in the following sections.

STRUCTURAL DESIGN OF 3D VIRTUAL REALITY LEARNING ENVIRONMENT

Traditionally the models of 3D objects are designed in modeling software. Corresponding file holds the geometric information as well as topological information of the object. This results in to very heavy file size. Since the topological details of

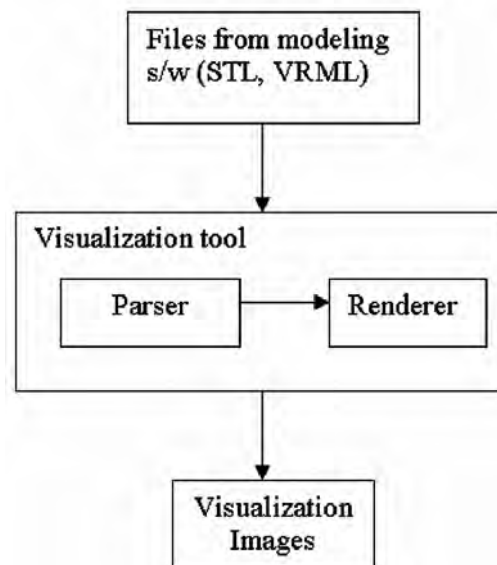
object are not required for visualization, executing such files for stereo visualization mode leads to time-consuming operation on a general-purpose computer. We eliminated the computationally intensive part of the traditional systems by designing a visualization solution which parses the ASCII files generated by any modeling software and retrieves only the data sets required for visualization in order to get desired display. We explain the structural design of the suite in the following section with the help of system architecture, sequence of operations and the method of exploring 3D vision in virtual learning environment (Kamath & Kamat, 2010).

System Architecture

The block diagram of the framework is shown in the in Figure 1. The visualization tool developed displays the model for 3D visualization by executing following modules:

- Computer Aided Design (CAD) systems are the source of design data. The objects are designed in CAD are exported to ASCII format. Our tool uses these ASCII files such as Stereo Lithography (STL) and Virtual Reality Modeling Language (VRML) files as input.
- The inputted STL or VRML file is given for parsing. Parser retrieves the required dataset from the inputted file for the visualization. These dataset stored in the form of scene graph structures. 3D scene is generated by rendering scene graph structures with the help of OpenGL. Rendering, i.e. the process of displaying the objects on the screen is accomplished in this manner.
- The Visual C++ programming language platform is used for the development of the visualization tool. It acts as an interface for STL and VRML models in 3D visualization.

Figure 1. Proposed research architecture



Technical Features of ASCII Files

As mentioned earlier the models designed in CAD system are exported to standard ASCII file formats. Information of the respective model is stored in the form of list of triangles. Coordinates of triangle vertices and other details associated with display are included in these ASCII files. The software suite processes the STL and VRML files for rendering the object. STL file contains facet-based representation of solid objects. The code snippet of ASCII STL file is given in Figure 2. It defines the surface of an object with set of

Figure 2. Syntax of ASCII STL file

```

solid roueascii.stl
  facet normal 1.207 0.927 -0.223
    outer loop
      vertex 3.500 -35.391 4.659
      vertex 4.034 -37.205 0.000
      vertex 3.835 -36.946 0.000
    endloop
  endfacet
  ...
endsolid
  
```

triangles. VRML is a scene description language (Carey et al.). It encloses all the details required for visualization of the model. Figure 3, shows the syntax of VRML file.

Sequence of Operations

A flow chart shown in Figure 4 explains the sequence of tasks implemented during the development of visualization solution.

The sequence of operations is as follows:

1. Initially the objects or models which are designed in CAD are exported into ASCII file formats which are input to this visualization tool. This tool imports the corresponding datasets and processes the same for visualization.
2. Inputted ASCII file is parsed and dataset required for display are retrieved by parser module. It results in to storage of model's details in structures for the next task.
3. Display structures are created during parsing. These structures contain set of triangles definition as well as other details related to the display of CAD model.

4. These scene-graph structures are displayed on the screen. Here OpenGL graphics tool is used for the display of objects.
5. The prime feature of the software suite is to provide a 3D stereovision system that allows user to explore datasets with cost effective way explained next to this.

Cost Effective 3D Visualization

Main feature of the software suite developed is cost effectiveness. The learner can view and interact with 3D visualizations in stereo mode with the help of fairly low cost red-blue glasses on a general-purpose computer. Thus the software suite achieves passive stereo vision (Wormell et al., 2007). This technique allows the perception of depth when viewed through colored red/blue glasses. Here left and right eye images are overlays in different colors and create a stereoscopic picture. The image component sent to brain by each eye with the help of a color filter and the result in three dimensions is interpreted by the brain. This is the less-expensive way of projecting in 3D, using a single digital projector (Zelle & Figura, 2004).

Figure 3. Syntax of ASCII VRML file

```
#VRML V1.0 ascii

Separator {
  Coordinate3 {
    point [
      0.553640 1.107283 0.000000,
      0.536024 1.107283 0.000000,
      0.536024 1.107283 0.196850
      .... - -
      .... - .]
    IndexedFaceSet {
      coordIndex [
        0, 1, 2, -1,
        3, 4, 5, -1,
        ..... - -
        ..... - .]
      }
    }
}
```

Figure 4. System development flowchart

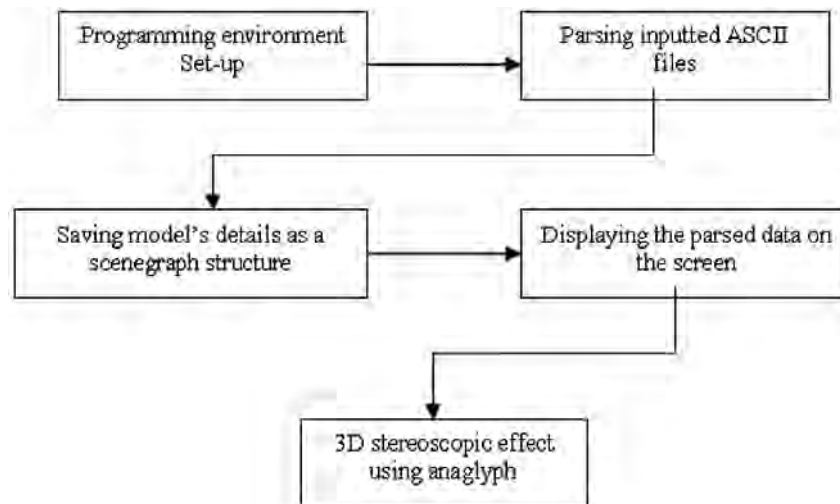
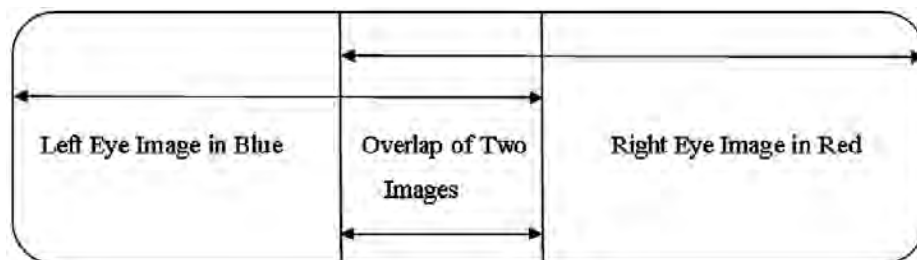


Figure 5. Blended image for two eyes



We follow the time-parallel principle for the stereo vision solution. The principle is based on providing both eye views to the viewer at a time and for directing each view to the appropriate eye it uses optical techniques. This needs the viewer to use red-blue eyewear. Simultaneous presence of both images on a screen, thus a time-parallel method is achieved.

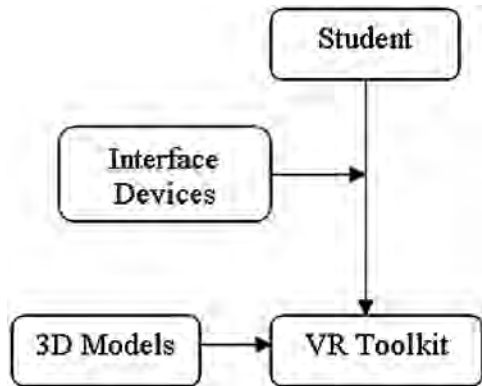
Colored filter over each eye results in one eye image which has been rendered in red and the other in blue. We use monochrome stereo images in the software suite i.e. the superimposition of left and right eye views which are in blue and red respectively. Single image is formed by combining left and right eye images. Figure 5 shows the above referred blending of images. The three-dimensional effect is perceived by viewing this blending through eyeglasses of corresponding colors but in a reversed manner. Thus the rendering has is

easy to achieved using simple image processing techniques and the same is less expensive.

EXPERIMENTAL SETUP FOR VR BASED LEARNING

As seen above thus the boundaries in traditional education can be broken by VR technology to create a world of imagination (Elomar, 2012). This requires time, effort and thoroughly intricate methods to fiddle with the technology for the learning purposes. As evidenced from our development, with the use of simulation and special hardware enables learners to occupy in a virtual environment. For the construction of a virtual reality system the need of special hardware and software includes;

Figure 6. VR based learning environment



1. **Virtual Reality Modeling:** This includes models of real world environments created using AutoCAD, 3Dstudio etc.
2. **Virtual Reality Software:** VR toolkits, software to support ample of applications.
3. **Virtual Environment:** Using display monitor and stereo glasses, user can view and interact with entities in VR environment. Stereo Glasses, allows the user to view true 3-D stereo depth of computer generated images. Gloves, allow the user to interact with virtual environment through finger control.

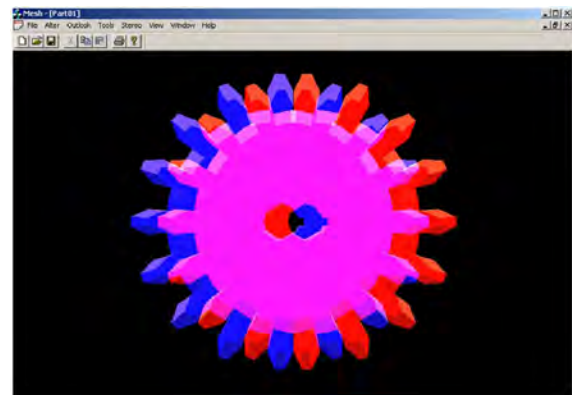
Figure 6 explains the VR based learning environment in education (Kamath & Kamat, 2011). Initially models are designed using CAD software. After the design the details of the model are saved as ASCII format. VR tool reads this file and presents CAD data in virtual space. A learner can visualize the 3D model by using this tool. Since VR is a computer simulation of a natural environment, interaction with a 3D model is more natural than 2D.

Left eye image drawn in blue is superimposed with right eye image which is in red. This superimposition when viewed through red-blue glasses the three-dimensional effect is perceived. Authors recommend the Red-Blue glasses the one shown in Figure 7 to perceive 3D effect. Figure 8 depicts the display of doubly rendered model in red and blue as output of the VR toolkit designed by us.

Figure 7. Red blue glass for 3D stereo display



Figure 8. 3D stereoscopic display of an object



FUTURE RESEARCH DIRECTIONS

The assimilation of VR, simulation and pedagogy provides a natural guide for future research. We are currently working on integration of social networking and visualization suite to imbibe creativity in e-learning scenario. Thus the future research and our further work aim at realization of social classroom in a flipped mode with 24 x 7 availability, which will exploit collaboration, teamwork and creativity amongst the learners and facilitates them to learn on their own, at the pace they prefer to grasp. Coupling of our software suite to the social networking platforms will assist us to research on the learning behavior of the students from different perspectives such as social integration, working in team as well as motivational and inspiration issues towards grasping new concepts.

Following are the set of additional facilities that would increase the supremacy and usability of this software suite;

- Openly support other CAD file formats such as STEP (Standard for the Exchange of Product model data), IGES (Initial Graphics Exchange Specification) etc.
- Implement auto stereoscopic vision that is perceiving 3D effect without the use of any external interface
- Supporting interface for other VR devices such as head mounted display, data gloves, simulators etc
- Picking and selection of desired part of the displayed model
- Changing the display properties for the selected parts of a rendered model

We are also working on the optimizing the software suite for a smart phone environment especially with the Android OS. This will help the learners to use this valuable learning aid on a move.

SIGNIFICANCE AND CONCLUSION

Virtual Reality has a pivotal role in bringing in meaningful teaching-learning ambience in education and training. Research on educational applications of VR, shows that VR technology can significantly improve the effectiveness of teaching by allowing learners to experience theoretical knowledge. In addition, it nurtures imagination, innovation, problem-solving of the learners thereby harnessing their creativity. We presented in this chapter a cost effective 3D stereo visualization solution which has great potential in education for active learning. This tool can be executed on a fairly general purpose low end computing platform. Students have the benefit of learning in such a VR based environment since it renders the display in the most personalized manner and thereby inculcate experiential learning.

Major advantages of using such a virtual reality tool in pedagogy are:

- VR based learning promotes a learner's ability to conceptualize in a three-dimensional space, and provide a sense of achievement whilst doing so.
- It makes the learning environment easy to use, cost effective, and convenient for users.
- The learners can study and test theoretical details in a virtual environment. Relevant information conveyed to students more efficiently than traditional techniques with the help of depth perception.
- VR has the prospective to revolutionize and improve the ways in which students are trained.
- VR grabs and holds the attention of students. It allows extreme close-up examination of an object.
- Engineering students can use less expensive projection or flat screen desktop stereoscopic solutions for 3D visualization and interaction in virtual environment (Sampaio et al., 2010)
- Medical students can learn spatial cues and sense of depth using Z co-ordinate while viewing material (Al-khalifah et al., 2006)
- Stereoscopic 3D Visualization has its impact in almost every area in education (Nicholos et al., 2012)

The development of this visualization suite was mainly an investigation on the technical practicality of a Virtual Reality System. We explored CAD as a medium to support early conceptual design through rapid prototyping of mechanical models. We thus explored Virtual Reality as a potential design prototyping environment in which prototypes of designs can be constructed, communicated and visually evaluated. The designed interface serves the visualization and evaluation of CAD geometry. Through a series of experiments, we thus implemented a method of achieving an

important feature of VR, stereoscopic display, in a cost-effective way for the most aspiring young learners. Interactive immersive less expensive 3D visualization of CAD object is the major outcome of this research which will no doubt yoke the creativity of the students who are now essentially digital natives. We are further working on coupling the software suite to the social networking platforms to experiment on the learning behavior of the students from different perspectives such as social integration, working in team as well as motivational and inspiration towards grasping new concepts.

REFERENCES

- Al-khalifah, A. H., McCrindle, R. J., Sharkey, P. M., & Alexandrov, V. N. (2006). Using virtual reality for medical diagnosis, training and education, *Proc. 6th Intl Conf. Disability, Virtual Reality & Assoc. Tech.* doi:10.1515/IJDHD.2006.5.2.187
- Alpaslan, Z. Y., & Sawchuk, A. A. (2004). Three-dimensional interaction with auto stereoscopic displays. *Proceedings of the Stereoscopic Displays and Virtual Reality Systems XI Symposium.*
- Black, N. (2015). *5 Stereoscopic 3D Resources for Elementary Students*. Retrieved 14 October 2015, from <http://www.engagingeducation.net/wordpress/5-stereoscopic-3d-resources-for-elementary-students/>
- Bowman, D. A., Chen, J., & Chadwick, A. (2006). New directions in 3D user interfaces. *The International Journal of Virtual Reality*, 5(2), 3–14.
- Carey, R., & Bell, G. (n.d.). *The annotated VRML 97 reference manual*. Retrieved from <http://accad.osu.edu/~pgerstma/class/vnv/resources/info/AnnotatedVrmlRef/appd.htm>
- Dalgarno, Carlson, & Tynan. (2010). 3D immersive virtual worlds in higher education: An Australian and New Zealand scoping study. *Proceedings ASCILITE Sydney*.
- Dongale, T., Patil, S., & Kamat, R. (2015). Learning by Simulations. *International Journal of Quality Assurance in Engineering and Technology Education*, 4(2), 13–25. doi:10.4018/IJQAETE.2015040102
- Elomar. (2012). *Virtual Reality Technology as a Didactical and Pedagogical Resource in Distance Education for Professional Training*. INTECH-Open Science.
- Engagingeducation.net. (2015). Retrieved from <http://www.engagingeducation.net/wordpress/5-stereoscopic-3d-resources-for-elementary-students/>
- Hamada, M. (2008). An example of virtual environment and web-based application in learning. *The International Journal of Virtual Reality*, 7(3), 1–8.
- Indeptheducation.com. (2015). Retrieved from <http://www.indeptheducation.com/wp/infographic-3d-enhances-learning/>
- Kamath, R.S., Dongale, T.D., & Kamat, R.K. (2012). Development of Virtual Reality Tool for Creative Learning in Architectural Education. *International Journal of Quality Assurance in Engineering and Technology Education*, 2(4), 16–24.
- Kamath, R. S., & Kamat, R. K. (2010). An Effective Stereo Visualization System Implementation for Virtual Prototyping. *International Journal on Computer Science and Engineering*, 2(7), 2316–2321.
- Kamath, R. S., & Kamat, R. K. (2011). Simple, low cost 3D Stereo Visualization Technique for mechanical engineering learners. *Proceedings of the IETEC'11 Conference*, 50–56.
- Kamath, R. S., & Kamat, R. K. (2013). Development of an Intelligent Virtual Environment for Augmenting Natural Language Processing in Virtual Reality Systems. *International Journal of Emerging Trends & Technology in Computer Science*, 2(3), 198–203.

Manseur, R. (2005). Virtual reality in science and engineering education, *Frontiers in Education, FIE '05. Proceedings 35th Annual Conference*. doi:10.1109/FIE.2005.1612051

Marketsandmarkets.com. (2015). Retrieved from <http://www.marketsandmarkets.com/Market-Reports/augmented-reality-virtual-reality-market-1185.html>

Medvr.edu. (2015). Retrieved from <http://medvr.ict.usc.edu/>

Nicholos, A. N., & Jones, B. L. (2012). Virtual Reality Learning: Effects in College and Training Environment. *66th EDGD Mid year conference proceedings*.

Piovesan, S. D., Passerino, L. M., & Pereira, A. S. (2012). Virtual Reality as a Tool in the Education. *IADIS International Conference on Cognition and Exploratory Learning in Digital Age*.

Sampaio, A. Z., Henriques, P. G., & Martins, O. P. (2010). Virtual reality technology used in civil engineering education. *The Open Virtual Reality Journal*, 2(1), 18–25. doi:10.2174/1875323X01002010018

Steinicke, F., Bruder, G., Hinrichs, K., Jerald, J., Frenzy, H., & Lappey, M. (2009). Real walking through virtual environments by redirection techniques. *Journal of Virtual Reality and Broadcasting*, 6(2).

Vafadar. (2013). Virtual Reality: Opportunities and Challenges. *International Journal of Modern Engineering Research*, 3(2), 139-1145.

Wormell, D., Foxlin, E., & Katzman, P. (2007). Improved 3D interactive devices for passive and active stereo virtual environments. *Proceedings of the 13th. Eurographics Workshop on Virtual Environments*.

Yahaya, R. A. (2006). *Assessing the Effectiveness of Virtual Reality Technology as part of an Authentic Learning Environment*. IEEE Computer Society of India. doi:10.1109/ICALT.2006.1652420

Zelle, Z. M., & Figura, C. (2004). Simple, low cost stereo graphics: VR for everyone. *Proceedings of the SIGCSE'04*. doi:10.1145/971300.971421

ADDITIONAL READING

Kamath, R. S., & Kamat, R. K. (2012). *A Monogram on Design and Implementation of Spatially and Temporally Efficient Visualization Algorithms*. Lap Lambert.

KEY TERMS AND DEFINITIONS

OpenGL: A software interface to graphics hardware for displaying object on the screen.

Parser: A module receives VRML or STL file as input, extracts the required data sets for visualization and creates display structure.

Passive Stereo: An inexpensive technique uses red-blue eyewear to perceive 3D effect.

Prototyping: An early model built to act as a thing to be replicated.

Renderer: A computer program which makes an object appears on the screen.

Scene Graph: An arrangement of logical and special representation of a graphical scene.

Stereo Visualization: A computer generated scene viewed using stereo glass to get three-dimensional effect.

Three-Dimensional: A set of geometric three parameters height, width and depth which describes an image.

Virtual Reality: A computer graphics technology that offers a simulated three-dimensional world.

Could Educational Technology Replace Traditional Schools in the Future?

E**John K. Hope***University of Auckland, New Zealand*

INTRODUCTION

During a recent presentation, Professor Karen Willcox, Professor of Aeronautics and Astronautics and co-Director of the Centre for Computational Engineering at the Massachusetts Institute of Technology, showed a painting of a university classroom in the Middle Ages with a teacher on a podium at the front of a room lecturing to a group of students, one of whom was asleep. Her next picture was of a twenty first century classroom with a teacher at the front lecturing to students, some of whom appeared to be uninterested. Her point was that little has changed in university classrooms for centuries, yet the affordances of e-learning suggest that university education could change more in the next fifty years than any change in the previous five hundred years.

BACKGROUND

The scenario portrayed above is also visible in the compulsory schooling sector where the majority of compulsory education still occurs in classrooms with a teacher, but it could be argued that technology-induced change is already occurring. Information and communication technology has penetrated even the most remote schools in developed countries and is infiltrating schools in some of the least developed countries in the world. Futurists have predicted the demise of many brick and mortar universities resulting from the affordances of technology that allow learning to occur anytime, anywhere and in whatever format the learner desires. What of bricks and mortar

schools? The growth of virtual education is not limited to adult education, schools are already incorporating virtual education into their educational programmes and some parents are opting for home based virtual education in preference to traditional schooling. Marketing of educational technology has attracted some of the biggest names in business and the education market has become an important component of the world economy, so change is inevitable.

The presence of technology in a classroom should not be seen as an indicator of change in the learning process; the early use of computers in schools was merely replacing teacher instruction with computer instruction, termed programmed learning. In both cases the instruction was top down, based on the view that the teacher or computer programmer was the font of knowledge whose task was to deliver information for the student to memorise. This instructional process can be useful for the delivery of some learning, such as presenting factual information, but the exponential growth of knowledge resulting since the computer revolution began, and the ease of access to that knowledge via technology, makes knowledge transmission possible without the need for a teacher in a classroom. It is easier to keep a computer up to date with the latest knowledge than to keep a teacher up to date. So what is driving changes to the form of schooling as it has traditionally been known?

Changes to Schooling

Globalisation has changed the range of knowledge, skills and dispositions required for a satisfying and

productive life. Traditional employment areas such as book-keeping have all but disappeared, new vocations such as mainframe computer engineers were created and then outlived their usefulness, and the world economy has lurched from one crisis to another due to influences beyond the control of any one individual. Soft skills are now needed to compete in a rapidly changing world and education must adapt to meet these changes.

Learners have changed too. The so-called Generation Z students are now in compulsory schooling. There is a debate about the definition of Generation Z. Geck (2006) indicated that the term is used to describe the students who born in or after 1990, Shatto and Erwin (2016) stated that Generation Z follows the Millennials, while Seemiller and Grace (2016) said Generation Z's students were born between 1995 and 2010. According to Seemiller and Grace (2016), Generation Z students are "loyal, thoughtful, compassionate, open-minded, and responsible" (p. 8). Viewed from family perspectives, Seemiller and Grace suggest that the personality characteristics of Generation Z are unique, because Generation Z students are raised by Generation X parents with an emphasis on individual responsibility and independence.

Although the dates that define Generation Z may differ, scholars have reached consensus that the world of Generation Z is shaped by the internet (Bassiouni, & Hackley, 2014). Generation Z are the first generation to spend the whole of their life in a world dominated by social media use. Viewed from social media perspectives, a Pew Research Center (2014) study indicated that millennials are "digital natives—the only generation for which these new technologies are not something they've had to adapt to" (para. 7). They have more reliance on social media such as Facebook than their older generation. An example of the influence of the internet can be found by viewing YouTube clips¹ show Generation Z babies trying to flick pictures in paper magazines to enlarge the picture or turn the page, in marked contrast to earlier generations who knew to turn a physical page.

As with any new technologies, there can be positive and negative effects on people. Social media links make it easier to connect with other people anywhere, anytime, but can be addictive. Social media addiction is a recently defined psychological phenomena spawned by issues with teenagers who cannot be parted from their smart-phone for more than an hour and whose health and school performance can suffer as a result ². Generation Z students are now in schools, bringing with them different learning needs to those of the generations before them. Their social lives are heavily influenced by social media and they cannot see why use of social media should not be normal within their education. Again, education must adapt to the changed needs of this different generation of learners.

As the baby boomer and subsequent generations age, many become financially independent, have increased free time and can turn to education once again to enhance the quality of their lives. Adult learners are either well qualified people who want to study later in life for interest, rather than to qualify for a vocation, or they may be adults who left school before earning the qualifications needed for success in life. In both cases, adult students may look to study at secondary school level, not by physically attending school where they would not fit socially, but studying virtually. As a result, Scandinavian countries have led the world in providing pathways into education for the needs of these adult learners, including use of technology (Biagetti, & Scicchitano, 2013; Boeren, Nicaise, & Baert, 2010; Rubenson, & Elfert, 2015). Adult learners who were schooled decades ago need different approaches to meet their technology education needs (Roche, 2016) as many lack the background in technology of younger generations.

As globalisation has changed the nature of work, technology has influenced the nature of learning and learners themselves have changed, there is a need for schools to change. The affordances of educational technology are likely to drive this change, particularly with new forms of virtual education (Robinson & Aronica, 2015).

The Schooling of the Future

No one person can predict the future, least of all this author, but the remainder of this chapter is devoted to suggestions of what some scholars think the future of compulsory schooling might be, seen through the lens of educationally focused information and communication technology development. Four areas will be examined in the text that follows. The implications of social media for education, blended/virtual learning and educational gaming are learning strategies to be examined in detail, followed by the implications for technological education leadership in schools.

FUTURE RESEARCH DIRECTIONS

Social Media and Education

Social media such as Facebook, WhatsApp, Twitter, LinkedIn and the like are accessed daily by most adults in developed countries. According to Duggan's (2015) survey in the United States, 70% of Facebook users, 59% of Instagram users, 27% of Pinterest users and 22% of LinkedIn users visit these platforms daily. Social media have also gained dedicated followers in schools, firstly in secondary schools and now increasingly in primary schools, despite their use in schools often not being encouraged by school authorities. Facebook states that the age limit for users is thirteen years, yet recent research has found usage at elementary level to be common. For example, McDonald, Laxman and Hope (2015) found that most nine to 12 year old students used Facebook 'illegally' by mis-stating their age or sometimes supported by their parents/caregivers who allow their children to use their Facebook account, despite concerns about frequently reported issues such as cyberbullying, sexual exploitation and social media addiction. By their nature, social media software packages were designed for social use so their take-up by educators has been understandably limited. However some studies

have found that school students often used social media for education related purposes. Roblyer, McDaniel, Webb, Herman and Witty (2010) found that up to 50% of students communicated about schoolwork while online using social media, the implication being that when social media are an integral part of a student's social life, this affordance should be utilised for formal educational purposes (Baran, 2010).

Benefits suggested from social media use in education include encouraging students to engage more freely with their learning and feeling less constrained when talking about it (Tynes, 2007), sharing knowledge with other students (Mazman & Usluel, 2010) and facilitating more introverted students to speak up online where they would be reluctant to do so in a face to face group situation (Larkin, 2016). To mitigate the disadvantages of social media, including the likelihood of social distractions when using social media in class, educational social media sites are available such as Twiducate.com, Classroom 2.0, Ipernity and Ning. Being virtual, these websites work both in school and out of school, blurring both physical and time divisions between the outside world and the classroom. Future use possibilities already being explored include international social network collaborations between students to develop international understanding, acceptance of difference and building international friendships. Social media are often used at all hours of the day and night, obviating the time change difficulties and financial costs that dogged earlier synchronous attempts for students to communicate between differing time zones using telephone and television links.

Blended Learning and Virtual Learning

As noted above, despite predictions of the demise of some physical higher education institutions due to the rapid growth of virtual learning, most school students still learn in a face to face school environment. No futurist found in a recent literature

search predicted the demise of schools in the near future. While as yet unknown new technologies can suddenly arrive and change the world, as did invention of the internet, current trends suggest that the near future of schooling is most likely to be founded on increased use of blended learning, where students can receive complementary face to face to face and virtual instruction (Miron & Gulosino, 2016). One high achieving Pennsylvania school district has utilised blended learning so successfully that it is used as an exemplar in an Entrepreneurship and Technology Innovations class at Harvard Business School, this use of an education example at Harvard Business School being an indication of the close link between the future of schooling and future business. The blended model adopted in Pennsylvania facilitates a move towards the previously idealistic aim of one to one individualised instruction and has seen voluntary enrolment of 300 students in the first year climb to 1100 students in year three. Blended learning does not suit all students. Initial results of the Philadelphia project show that when compared with traditional classes, two thirds of the blended programme students improved more than the traditional peer group, but one third did not improve, or did not do as well (Davis, 2016).

Another indication of the future for virtual learning is the arrival of MOOCs (Massive Open Online Courses) provided tuition free courses in almost any subject imaginable by high ranked universities in many countries. Although not necessarily providing a recognised qualification there has been massive take up of these courses by adults, enrolment numbers being in the millions for MOOCs provided by top universities such as edX MOOCs provided by Harvard and MIT. Although take up is high, completion rates can be as low as 5%, the main problem stated as being student time constraints. Research into this issue has found that adult students are beginning to adapt to MOOC style learning by using the MOOC as an initial source to be supplemented by other resources, utilising only the portion of the MOOC relevant to their needs, and finding that the ability to pick

up small components of the course at different times helped in mitigating the time constraints of modern life (Veletsianos, Reich, & Pasquini, 2016). This research highlights the affordances of virtual education such as anywhere, anytime, affordances that are also applicable to schooling. A study of edX revealed that 39% of edX MOOC participants were teachers, either enrolling for in-service professional development or to help in their pre-service certification endeavours (Ho et al., 2015). AboutEdu Inc., a non-profit organisation listing MOOCs available for the school sector, lists 33 MOOC providing institutions available to elementary and high school students, and/or MOOCs providing professional development opportunities for teachers. AboutEdu Inc. requests for MOOC assistance came from many countries with USA, China and India being the countries most frequently requesting MOOCs, a graphic illustration of the increasing global ubiquity of virtual learning.

A report by the United States National Center for Education Statistics found that virtual and blended schools numbers were growing, despite some negative reports about their impact on student learning. 447 virtual schools were listed in 2013-14, attended by 262,000 students, with another 26,000 students attending 87 blended schools (Miron & Gulosino, 2016). This report noted concern that the graduation rate from virtual schools was 41%, from blended schools 37%, compared with 81% for all schools. Other concerns included high teacher: student ratios (35:1 for virtual schools, 32:1 for blended schools, 16:1 traditional public schools), and virtual school rolls included smaller proportions of minority and low-income students. Three quarters of the virtual schools were private schools, a major difference from the traditional school sector, making comparisons difficult. But this report and another from Stanford University that found cyber-charter schools had an “overwhelmingly negative impact” (Center for Research on Education Outcomes, 2015, p. 32) on student achievement suggests that private sector involvement in the virtual/blended school

development has not always produced positive results. The issue of the increasing trend towards for-profit schooling will be discussed further in a later section of this chapter.

One other component of blended learning is making rapid and seemingly more positive inroads to schooling, particularly in the United States. Open education teaching resources are free to teachers and no longer are limited to teachers sharing their personal planning and ideas on the internet. One example is EngageNY, a collection of materials published free by the state of New York. With 45 million downloads and in excess of thirteen million users, this customizable material is only one example of many that are helping teachers to cope with high workloads and increased accountability. Users of this particular resource are not limited to the United States. Although primarily aligned with United States education standards, their listed users include tens of thousands of teachers from the Philippines, Canada, India, United Arab Emirates and the United Kingdom (New York State Education Department, n.d.). Other open education providers such as UnboundEd (www.unbounded.org) are appearing, increasing the likelihood that open education resources will become even more popular in the future.

Gamification

Almost all school age students in developed and developing countries are aware of digital recreational games, and most play them at some point in their lives. The motivation effect, fun and challenge of recreational digital games such as Minecraft have universal appeal that prompts teachers to wonder if there are classroom applications of gamification that could improve learning. Ever since computers became common in classrooms more than 30 years ago, games have been used in classrooms. Beginning with very simple rote learning, drill and practice applications such as teaching basic mathematical facts with small electronic rewards, newer games are much more sophisticated and are increasing being seen as having potential for

school use. A large scale United States survey of 500,000 students, teachers and parents conducted by a non-profit organisation revealed that in 2015, 48% of teachers surveyed reported using games in the classroom compared with 23% in 2010, and the number of teachers using online videos had increased from 47% in 2010 to 68% in 2015 (Speak Up, 2015). Use does not necessarily equate to achieving improved learning, but the same survey found that 48% of teachers, 50% of high school students, 60% of elementary students and 40% of parents considered educational gaming a component of their ideal future school, suggesting that educational gaming is seen to have value in schooling. Students identified reasons why they saw value in educational online videos including the ability to watch again as needed, easier to understand, connected to the real world, fits learning style, easy to access on mobile devices and more engaging.

Not all results were positive. The Speak Up survey also revealed that teachers were reluctant to use digital learning, and needed professional development to enable them to make maximum use of digital learning, such as when to incorporate them into the learning unit and how to sort out the educationally useful from the plethora of low quality games available on the net. Education administrators were cautious about the use of educational gaming and few had planned appropriate professional development for teachers. Despite this caution, big-money developers are introducing game-based programmes designed to develop 21st century problem solving skills and having the potential use for more authentic assessment purposes, such as measuring skills less readily measured by traditional methods. For example, Microsoft has developed an education version of Minecraft that could be used in this way. At the 2015 American Education Research Association conference, presentations included studies that investigated the merging of physical movement with virtual understanding, understanding students' virtual thinking, and virtual gardening, examples of emerging technologies

that utilise digital learning to supplement more traditional learning methodology, blurring the lines between recreational play and school-based learning (Herold, 2015).

Pokemon Go, an app that is higher tech version of the traditional recreational game, is creating news headlines as it sweeps the world while this chapter is being written. The game requires users to get out into their community and connect the real world with the virtual Pokemon world. Some of the media headlines are negative, noting safety issues for students, the warring component and at higher Pokemon levels, the requirement to pay. Conversely, incredible levels of motivation are generated, students have to get out into their communities and there is a community educational component built into the game. This has also made educators think about the potential of the model for educational applications (Doran, 2016).

Hard research demonstrating the affordances of digital games is sparse, but a meta-analysis of digital game research for K-16 students investigated the cognitive, intrapersonal and interpersonal learning domains (Clark, Tanner-Smith, & Killingsworth, 2014). Overall, analysis of 57 studies showed the digital game studies analysed to be more effective in all three learning domains than the non-game instruction measured. The authors of the research note the difficulty in all such meta-analyses across different studies, the results cannot necessarily be generalised to all other situations, only to the situations measured in the studies considered in the meta-analysis. Nevertheless, this study is one indicator of the potential of digital game based learning, particularly for supporting higher order learning outcomes.

For-Profit Education

For-profit education was mentioned above due to its increasing move into the compulsory schooling sector. When 2015 annual global expenditure on educational technology hardware reaches US\$15 billion (Molnar, 2016), big business takes notice. Similarly, when virtual education creates ac-

cess to the learners of the world at the click of a mouse, big business takes notice again. For-profit education has proliferated far beyond the private international school model and coaching classes of paid education of the past, with millions of school students now paying for part of their education via virtual means. This growth suggests that for-profit education growth is not limited to hardware sales, it also encompasses delivery of education, but not always as effectively as claimed. Henry Levine from Teachers College Columbia sees for-profit education as a whole making more claims about effectiveness and cost efficiency than are justified (Gustke, 2010).

A sample of the issues that have accompanied the growth of for-profit virtual education were documented earlier in this chapter, but there are also some examples where for-profit education has been beneficial. Microsoft's launch of the education version of Minecraft is one example that is currently available free. For profit corporations donate money to facilitate research into the educational use of technology, or provide some of their products without charging. Corporations have the resources to front technological and organizational innovation (Carey, 2010). They may also close educational market gaps left by traditional education groups, and provide a second chance for students ignored by non-profit education (Chung, 2012).

Educational Leadership

Earlier generations of school leaders could focus entirely on curriculum and assessment issues, their only contact with technology being chalk and writing equipment. Today's school leaders still have to focus on curriculum and assessment, but ignore the influence of electronic technology on curriculum and assessment of student learning at their peril. Provision of the most appropriate hardware, selection from the plethora of software packages available and choosing virtual providers are new school leadership tasks. Parental demands for school access to the latest

educational technology, internet security issues and the critical need for educational technology professional development of teachers all concern school leaders of today and are likely to take even more of leaders' time in the future. A 2016 survey by the Education week Research Center found significantly more teachers finding tailored professional development conducted by sharing with other teachers, collaborative planning and having educational technology coaches available in their classrooms meeting their professional development needs better than traditional forms of professional development, such as read-made educational district and for-profit company provided professional development (Flanigan, 2016).

One example that should concern school leaders is a summary of several American surveys by credible organisations, such as the Council of Chief State School officers, which found high stakes test score results varied dependent on the type of technology used to take the tests (Herold, 2016). Overall, there were suggestions that higher scores occurred when desktop computers were used, lower scores when laptops were used and lower again scores when tablets were used. One reason suggested was screen size, but results were not definitive. This example illustrates the difficulties facing school leaders when trying to decide what is best for their students.

However technology affordances also increase the tools available to school leaders that will assist them to deal with these new challenges. One example is the John's Hopkins university partnership with a non-profit organisation called Digital Promise to provide free access to a tool that provides objective evaluation information about the efficacy of technological offerings to schools (Cavanagh, 2015).

CONCLUSION

This chapter began by quoting Professor Wilcox noting that little had changed in schooling over the past 500 years. In response, this chapter provides examples of how the world has changed due to glo-

balisation and e-learning becoming mainstream, the changing needs and aspirations of the current generation of learners and the rapid development of new technologies to assist learning. All of the examples quoted suggest that traditional face-to-face schooling will change.

Hauser and Koutouzios (2009) state "it is arguable that technology has had a greater impact on elementary and secondary school education during the last twenty-five years than virtually any other development in American society" (p. 245). As some schools in the present are already doing, most schools of the future will most likely need to embrace educational use of social media, blended and virtual learning and educational gaming as learning tools and may incorporate for-profit components into their offerings. Educational leadership will change, with much more focus on research justified use of educational technology and continuous professional development of teachers to enable them to cope with rapid technology instigated change. But none of the evidence reviewed here suggests that schools as such will disappear. The economic and social survival of nations depends on the quality of education of its people so some form of compulsory schooling will be needed, be it virtual or not. Perhaps Professor Wilcox's lecture in ten years' time will compare a medieval school with a virtual or partly virtual school, a clearly different learning environment minus a teacher at the front delivering factual knowledge.

REFERENCES

- Baran, B. (2010). *Facebook as a formal instructional environment*. Retrieved from <http://onlinelibrary.wiley.com.ezproxy.auckland.ac.nz/doi/10.1111/j.1467-8535.2010.01115.x/pdf>
- Bassiouni, D. H., & Hackley, C. (2014). Generation Z childrens adaptation to digital consumer culture: A critical literature review. *Journal of Customer Behaviour*, 13(2), 113–133. doi:10.1362/147539214X14024779483591

- Biagetti, M., & Scicchitano, S. (2013). The determinants of lifelong learning incidence across European countries (evidence from EU-SILC dataset). *Acta Oeconomica*, 63(1), 77–97. doi:10.1556/AOecon.63.2013.1.5
- Boeren, E., Nicaise, I., & Baert, H. (2010). Theoretical models of participation in adult education: The need for an integrated model. *International Journal of Lifelong Education*, 29(1), 45–61. doi:10.1080/02601370903471270
- Cabral, J. (2011). Is Generation Y addicted to social media? *The Elon Journal of Undergraduate Research in Communications*, 2(1), 5–14.
- Carey, K. (2010). Why do you think they're called for-profit colleges. *The Chronicle of Higher Education*, 56(41), A88.
- Cavanagh, S. (2015). *New tool helps districts evaluate ed-tech companies' claims of 'Evidence'*. Retrieved from <https://marketbrief.edweek.org/marketplace-k-12/a-tool-for-judging-ed-tech-companies-product-research/>
- Center for Research on Education Outcomes. (2015). *Online charter school study*. Retrieved from <https://credo.stanford.edu/pdfs/OnlineCharterStudyFinal2015.pdf>
- Chung, A. S. (2012). Choice of for-profit college. *Economics of Education Review*, 31(6), 1084–1101. doi:10.1016/j.econedurev.2012.07.004
- Clark, D., Tanner-Smith, E., & Killingsworth, S. (2014). *Digital games, design and learning: A systematic review and meta-analysis (executive summary)*. Menlo Park, CA: SRI International.
- Davis, M. (2016). Harvard Business School examines K-12 blended learning. *Education Week*, 35(35), 27.
- Doran, L. (2016). *As Pokemon Go becomes a sensation, Ed. Experts weigh pros and cons*. Retrieved from http://blogs.edweek.org/edweek/DigitalEducation/2016/07/pokemon_go_evolves_into_overnight.html?cmp=eml-enl-dd-news1
- Duggan, M. (2015). *Mobile messaging and social media 2015*. Retrieved from <http://www.pewinternet.org/2015/08/19/mobile-messaging-and-social-media-2015/>
- Flanigan, R. (2016). Ed-Tech coaches becoming steadier fixtures in classrooms. *Education Week*, 35(35), 31–32.
- Geck, C. (2006). The generation Z connection: Teaching information literacy to the newest net generation. *Teacher Librarian*, 33(3), 19–23.
- Gustke, C. (2010). *E-learning industry on the rise*. Retrieved from <http://www.edweek.org/dd/articles/2010/10/20/01ebiz.h04.html>
- Hauser, G. M., & Koutouzos, D. W. (2009). Technology training and professional development of school leaders in the USA: The critical need for reform. In S. Khine & I. Saleh (Eds.), *Transformative leadership and educational excellence* (pp. 245–266). Rotterdam, The Netherlands: Sense Publishers.
- Herold, B. (2015). *Frontiers of digital learning probed by researchers*. Retrieved from <http://www.edweek.org/ew/articles/2015/05/06/frontiers-of-digital-learning-probed-by-researchers.html>
- Herold, B. (2016). *Digital device choices could impact common-core test results, studies finding*. Retrieved from <http://www.edweek.org/ew/articles/2016/07/20/digital-device-choices-could-impact-common-core-test.html>
- Ho, A., Chuang, I., Reich, J., Coleman, C., Whitehill, J., & Northcutt, C. ... Petersen, R. (2015). *HarvardX and MITx: Two Years of Open Online Courses Fall 2012-Summer 2014*. Available at SSRN: <http://ssrn.com/abstract=2586847> or 10.2139/ssrn.2586847
- Larkin, P. (2016). *Our students can't be "quiet" on social media in school*. Retrieved from http://blogs.edweek.org/edweek/reinventing_k12_learning/2016/06/our_students_cant_be_quiet_on_social_media_in_school.html?cmp=eml-enl-dd-news3

- Mazman, S. G., & Usuel, Y. K. (2010). Modeling educational usage of Facebook. *Computers & Education*, 55(2), 444–453. doi:10.1016/j.compedu.2010.02.008
- McDonald, C., Laxman, K., & Hope, J. (2016). An exploration of the contexts, challenges, and competencies of pre-teenage children on the internet. *International Journal of Technology Enhanced Learning*, 8(1), 1–25. DOI: 10.1504/IJTEL.2016.075949
- Miron, G., & Gulosino, C. (2016). *Virtual schools report 2016: Directory and performance review*. National Education Policy Center, School of Education, University of Colorado Boulder, CO: National Education Policy Center.
- Molnar, M. (2016). *Spending on Ed-Tech hardware hits \$15B worldwide, report finds*. Retrieved from <https://marketbrief.edweek.org/marketplace-k-12/spending-on-education-hardware-up-7-percent-worldwide-report-finds/?cmp=eml-enl=ii-news1>
- Pew Research Center. (2014). *Millennials in adulthood: Detached from institutions, networked with friends*. Retrieved from <http://www.pewsocial-trends.org/2014/03/07/millennials-in-adulthood/>
- Robinson, K., & Aronica, L. (2015). *Creative schools: The grassroots revolution that's transforming education*. New York, NY: Viking.
- Roblyer, M. D., McDaniel, M., Webb, M., Herman, J., & Witty, J. V. (2010). Findings on Facebook in higher education: A comparison of college faculty and student uses and perceptions of social networking sites. *The Internet and Higher Education*, 13(3), 134–140. doi:10.1016/j.iheduc.2010.03.002
- Roche, S. (2016). Building on shifting sands: The challenges of meeting the learning needs of adults in a rapidly changing world. *International Review of Education*, 62(3), 247–251. doi:10.1007/s11159-016-9569-2
- Rubenson, K., & Elfert, M. (2015). Adult education research: Exploring an increasingly fragmented map. *European Journal for Research on the Education and Learning of Adults*, 6(2), 125–138. doi:10.3384/rela.2000-7426.rela9066
- Seemiller, C., & Grace, M. (2016). *Generation Z goes to college*. New York, US: John Wiley & Sons.
- Shatto, B., & Erwin, K. (2016). Moving on from Millennials: Preparing for Generation Z. *Journal of Continuing Education in Nursing*, 47(6), 253–254. doi:10.3928/00220124-20160518-05 PMID:27232222
- Speak Up. (2015). *From print to pixel: The role of videos, games, animations and simulations within K-12 education*. Retrieved from: <http://www.tomorrow.org/speakup/SU15AnnualReport.html>
- Tynes, B. M. (2007). Internet safety gone wild: Sacrificing the educational and psychological benefits of online social environments. *Journal of Adolescent Research*, 22(6), 575–584. doi:10.1177/0743558407303979
- Veletsianos, G., Reich, J., & Pasquini, L. A. (2016). The life between big data log events: Learners strategies to overcome challenges in MOOCs. *AERA Open*, 2(3), 1–10. doi:10.1177/2332858416657002 PMID:26942210

ADDITIONAL READING

- Berney, S., & Betrancourt, M. (2016). Does animation enhance learning: A meta-analysis. *Computers & Education*, 101(October), 150–167. doi:10.1016/j.compedu.2016.06.005
- Hope, J. K. (2016). New learning for new students. In V. C. X. Wang (Ed.), *Handbook of Research on Learning Outcomes and Opportunities in the Digital Age* (pp. 819–837). Hershey, PA: IGI Global; doi:10.4018/978-1-4666-9577-1.ch036

Khenissi, M., Essalmi, F., Jemni, M., Graf, S., & Chen, N. (2016). Relationship between learning styles and genres of games. *Computers & Education*, 101(October), 1–14. doi:10.1016/j.compedu.2016.05.005

Patrick, S., Worthen, M., Frost, D., & Gentz, S. (2016). Promising state policies for personalized learning. Vienna, VA: International Association for K-12 Online Learning (iNACOL).

Schwartz, B., & Caduri, G. (2016). Novelty in the use of social networks by leading teachers in their classes. *Computers & Education*, 102(November), 35–51. doi:10.1016/j.compedu.2016.07.002

KEY TERMS AND DEFINITIONS

Blended Learning: Traditional face to face learning combined with technology mediated instruction.

E-Learning: Any learning conducted via electronic technology (computers, smartphones and the like).

Gamification: Electronic video games designed for teaching and learning rather than purely entertainment.

Virtual Education: Another name for e-learning, but referring more to electronically aided learning conducted without any face-to-face components.

ENDNOTES

¹ <https://www.youtube.com/watch?v=aXV-yaFmQNk>

² See Cabral (2011) for research into social media addiction.

Development of Communication Skills through Auditory Training Software in Special Education

E**Eduardo C. Contreras***Autonomous University of Coahuila, Mexico***Isis I. Contreras***Saltillo Institute of Technology, Mexico*

INTRODUCTION

The primary function of language is communication and this is per excellence, of verbal nature; manifesting itself as the defining feature of men as a rational being capable of communicating through a system of signs (language) used by social communities. People with hearing loss do not have access to natural language, hampering their oral and/or written communication. Some of them have never heard any sounds; in others, the hearing loss appeared later, so they have the need for auditory therapy.

Hearing impaired individuals who have received a cochlear implant, as well as those who need to use an electronic device to improve their hearing, should receive auditory therapy to teach the proper use of the devices. Those who never had access to a proper hearing do not recognize the meaning of the sound, and must learn to associate the sound with its concept. On the other hand, people who had access to hearing have the problem of oral communication, so it is necessary to receive auditory training therapy and must attend therapy in a special education (Merzenich, Pandya, & Tremblay, 2005).

This paper presents a study in Saltillo, Mexico in 2014, which is intended to verify the development of communication for people with hearing loss using Auditory Training Software (ATS), which allows to develop listening skills in a motiva-

tional context and influencing positive behavior in training tasks. The ATS helps to develop the skills of identification of presence or absence of sound, identify and discriminate a syllable from similar pairs considering difficulty levels and distinguish individual words and sentences or phrases.

This ATS has been useful in the community as an important support educational tool for students and teachers, it goes beyond teaching materials because it becomes relevant in the teaching-learning process, appropriation of linguistic symbols, communication (oral and written), personal identity (auditory phonetic), among others.

However, the use of ATS should be customized according to the user profile considering their language, age, education, level of curricular and cultural competition, type and degree of hearing loss, type of disability, among others; the ATS development is complex, it is difficult to find freeware. Currently, there are several applications for Auditory Training (AT) in different languages and even some adapted to the educational needs of other countries, for example, from English to Spanish (Moore & Amitay, 2007).

This paper contributes with a description to develop an educational software created by students and teachers of Computer Science, as well as a teaching tool to assess the user's progress, useful for teachers of Special Education and an appropriate research method to the disciplines mentioned.

BACKGROUND

AT is necessary for the development of communication for people with hearing impairment. ATS has designed to help streamline the complexity of projects and tasks as well as facilitates team collaboration, to develop human skills (the ability to interact and motivate), understand concepts and develop ideas.

Some basic concepts of the subject are mentioned below.

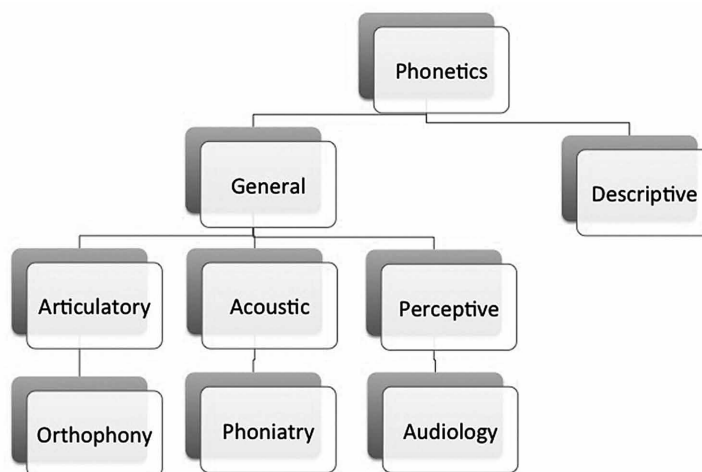
- **Prelocution Deafness:** Hearing impairment is one that occurs before the child has acquired spoken language and for whom the development of speech and hearing may be affected in different ways.
- **Post-Lingual Deafness:** Hearing impairment that occurs after learning spoken language.
- **Phonetics:** The study of the sounds made by the human voice in speech; production by the speaker and reception / perception by the listener, (Llisterri, 1991, p.15).
- AT is the process to teach people to understand the meaning of sounds. During this training, auditory stimulation is provided

to people to learn, to identify, distinguish and conceptualize sounds.

- Auditory-verbal method is a process that uses multisensory rehabilitation methods such as Ling's oral phonological system, Calvert's Multisensory System and the Van Uden reflective Method.

In the AT, the application of the theory of phonetics is very important because it is the basis of our communication system. The AT process uses an experimental method consisting of three stages to achieve the goal of communication. This process is described below (see Figure 1). In the first stage it is necessary to differentiate the General Phonetics (production and perception of sounds) from Descriptive Phonetics (different known languages); it is important General Phonetics because it describes the communication process. The second stage is refers to the study of the sound production by the sender is the Articulatory Phonetic, the Acoustic Phonetics studies transmission and perception of the message. Finally, Perceptive Phonetics, studies hearing and perception of the message by the receiver also used in the synthesis and voice also recognition between human-computer. The application of

Figure 1. Phonetics in the communications process
Source: Prepared, 2016.



the knowledge of phonics to re-educate children or adults with language disorders is known as Orthophony; the medical discipline that is related to voice disorders is the Phoniatriy, lastly, the Audiology is focused on assessing whether there is hearing loss that involves lack of understanding of the message received by the receiver. Every single person, from birth, goes through this process, either naturally or with special training because of a deficiency in the process of communication.

The practice of auditory training in the 21st century is considered in terms of the collaborative, interdisciplinary research of neuroscience, cognitive science, and auditory science (Kricos & McCarthy, 2007). Since the last century there have been published manuals that provide methods and material with basic lessons to train people with hearing impairments, particularly those who use cochlear implants or hearing aids (Rochette, Pescheux, & Bigand, 2008). People who use a hearing device or a cochlear implant need AT (González & Torre, 2006). It is important to understand that even though a child uses an appropriate hearing device, calibrated and in good condition, will not have the same access to auditory information as a child who hears normally does.

The auditory-verbal approach used years ago was proposed based on a group of sequential and structured strategies that rely only on the sense of hearing to provide access to linguistic information. Generally, the educational intervention is performed as described by Pérez (2010), who mentions the necessary resources for training and appropriate intervention for children with various forms of hearing impairment. For this, various methods are used in the lessons proposed for training; generally, they provide material for contrasting sounds, recognizing sounds, sounds of the words, prayers and songs, and rhythmic patterns. Rochette & Bigand (2010) proposed a training program similar to the previous one, which has a variety of stimuli of auditory perception on cognitive operations in children with hearing impairment; it is based on the training of the cognitive operations in the perceptive processes.

The auditory verbal therapy has been currently provided with the support of the computer, as shown by the study conducted in Germany (Strehlow, 2006), which used computer-based methods in order to measure and train the temporary sound processing methods and the stimulation through phonemes in children with dyslexia, to develop skills for spelling and reading. In the study by Russoa, Nicola, Zeckerc, Hayesa & Krausa (2005) they observed that the influence of AT alters the neural structure of speech perception brainstem. They used perceptual AT software in nine patients clinically diagnosed with learning language-based.

In another study by Stacey, Raine, O'Donoghue, Tapper, Twomey & Summerfield (2010) to evaluate the effectiveness of using an AT software in eleven adults with cochlear implants, focused to discriminate vowels and sentences, they found higher significance in discrimination of consonants in speech. Toledo (2005) proposed the use of Speech Analyzer (Summer Institute of Linguistics), software that allows recording and analyzing speech sounds, as a possible tool for studying speech therapy by testing the pronunciation of the consonants in people with speech problems. The test showed that these people could perceive the phonetic accent before applying the rules of orthographic accent. At the University of Konstanz a German prototype to improve the ability to discriminate sounds phonemes and syllables in people with mild cognitive impairment and mild Alzheimer's disease; the effectiveness of this training program to mitigate or even reverse cognitive decline was assessed. Zhou, Sim, Tan, and Wang (2012), mentioned the effectiveness and efficacy of a game for mobile devices (MOGAT) in children with cochlear implant, to develop his musical ability was demonstrated. The discrimination of the tone in a pleasant, intuitive and motivating way was improved.

It is currently offered in the market a variety of educational software for the hearing impaired, as they are: "Computer-Assisted Speech Perception Testing" and "Training at the Sentence Level", or CASPERSent, which is a multimedia program

whose main purpose is the sentence-level speech-perception training and testing for persons with hearing loss. Another programs are similar to the previous “Tracking Computer Assisted Simulation” and “Computer Assisted Speech Training (CATS)”. “Computer-Assisted Speech Training (CAST)”, was originally designed for adults with cochlear implants, but it can be adapted for use with adult hearing aid wearers.

Other commercial programs that can be used for people with hearing loss are: “Seeing and Hearing Speech”, “Sound and Way Beyond”, “Sound Scape” and “Speech Perception Assessment and Training System (SPATS)”.

There are some apps that might help with Auditory Processing Disorder like: “Auditory Reasoning”, “Phonetic Birds”, “Auditory Memory Ride”, “Sound Match”, “Hear Coach”, “Auditory Processing Studio” and “Articulation Games”; these apps can also help with central auditory processing disorder (CAPD) or other related disorders (e.g., receptive language disorder or autism).

The studies previously mentioned have different approaches, some of them use similar methods to the one presented in this chapter, but they do not cover the specific needs of Special Education in México and the above applications are not in the Spanish language. For this reason it became necessary to develop an application that meets the needs of the region of Northeast, México.

MAIN FOCUS OF THE ARTICLE

The approach is to use an appropriate method to provide AT in special education schools for children with hearing loss who have the need to develop communication skills, as well as the use of auditory information to teach them how to use their residual hearing to build relationships in their community.

In the studied problematic, two important elements have been identified: Children with hearing loss and AT educators.

- **Children with Hearing Loss:** Prelingual hearing loss and deafness can occur from birth or childhood; it means that they have never heard a sound in their entire lives. The Greek philosopher Aristotle said “all those who are born deaf are also silent, unable to speak and to rise to abstract and moral ideas”. Consequently, these children have problems of communication and education because this deficiency does not allow the full development of communicative language skills, this due to the lack of association of the signifier and the signified from the linguistic symbol as they are not able to connect sounds with its meaning. This problem also causes deficit in vocabulary, altered articulation of words, delayed morphosyntactic structure, attention deficit, and a poor oral and written communication. These individuals need a specific and planned intervention to help them access education. It is important to use the verbal hearing oral teaching method that allows deaf children (with hearing aids or cochlear implants) to structure their thinking and to acquire an inner language, which is the main base of educational work.
- **AT Educators:** Most educators in public special education centers in México do not use the software as a teaching resource because they do not have specialized software available. The existing software is foreign (in English), and the one provided in Spain contains vocabulary and pronunciation different from the Mexican ones. Software developers in México generally focus on the commercial, because this is not considered profitable. Generally, public institutions of higher education are the ones support the development of educational software for various public schools through projects financed by governmental institutions. Another problem identified in educators is that most of them are unaware

of the proper use of the software and how to evaluate the learner with it. Generally, the educator traditionally continues using the audio-verbal method without the support from ICT's, showing an attitude of technophobia and resistance to change. Added to this, the computer equipment in schools is scarce or obsolete.

Research Method

The research method used in this project is described below; using a conceptual map the system is represented, the study elements are identified in it and with those the problem is defined as: Children with hearing loss need to receive AT according with their level of hearing, reinforced by visual stimulation and interactive tasks that increase their motivation, in order to develop their communication skills.

The elements to define the method of study are identified, which are shown in the following functional relationship:

Communication of a children with hearing loss = f (listening skills, AT software, learning behavior).

These elements are very important in developing the research method because they are variables that contain the hypotheses, the basis for raising objectives and the research questions.

Based on the above a synthetic table is generated (table of methodological consistency), as shown in Table 1.

From this it is checked whether there is correspondence between the variables that are going to related, and between the approach of the objectives and hypotheses, when correct, one proceeds to design the Consistency Matrix (see Table 2) based on the theoretical framework. This matrix will help us develop our instrument of evaluation and measurement of study in a methodical and systematic way (Rivas, 2004). One continues with the collection and statistical processing of data for the results of the study. The research is a study case of the explanatory, qualitative and quantitative type because it uses observation in the population sample.

Sample Description

The study was conducted at the Center for Multiple Attention "Benito Juárez TV", where 11 hearing loss children were selected from among

Table 1. Methodological consistency table, where are shown objectives, hypothesis, research questions and research variables

Particular Objective	Research Questions	Variable	Hypothesis
O1. To evaluate, through software, the development of hearing skills in verbal auditory processing method in the rehabilitation of persons with hearing impairment to enable them to acquire oral communication.	RQ1. How much can the software rehabilitate people with hearing impairment on the development and understanding of sounds, syllables and words, to achieve an acceptable level of communication?	Hearing ability	H1. The multimedia-learning environment improves the development of listening skills in individuals with hearing impairment to develop their communication.
O2. To verify that the use of ear training software is an adequate tool for verbal teaching method in oral communication.	RQ2. Do software elements; multimedia, user interaction, the tasks of the auditory-verbal method relate to communication skills?	Ear training software	H2 There is a positive relationship between software ear training and communication development.
O3. To explore the behavior of natural language learning in the individual with a disability when stimulated in the process of communication.	RQ3. With the stimulus presented in the multimedia educational software, to what extent does it influence attitudes, motivation and attention on learning to speak in the hearing impaired?	Learning Behavior	H3 If people with hearing impairments are stimulated through multimedia software; they present a positive performance in auditory learning to develop their communication skills.

Source: Prepared, 2016.

Table 2. Methodological consistency matrix, where are shown research variables, factors, indicators and dimensions

Variable	Factor	Indicator	Dimension
Auditory skills	Sounds	Presence / absence	9
		Sound identification	8
		Sound discrimination	7
	Syllables	Identifying syllable speech	6
		Discrimination syllable speech	5
		Identifies / discriminates a syllable in cognate pairs	4
		Syllable comprehension	3
	Words	Identifies isolated words	2
		Identifies words in a sentence	1
Learning behavior	Behavioral attitude	Enthusiasm	5 - Always
		Interest	4 - Almost always
		Constancy	3 - Sometimes
	Motivation	Active participation	2 - Rarely
			1 - Never
	Attention	Timing	Minutes
Ear training software	Multimedia elements	Audio	5 – Very good
		Picture	4 - Good
		Video	3 - Regular
		Text	2 - Minimum
	User interface	Usability	1 - None
		Different skill levels	
	Auditory-Verbal Method	Manipulate interface objects	
		Interface consistent	
		Performs the tasks	
		Schedules the tasks	
		Stimulates the audition	
		Evaluates the progress	

Source: Prepared, 2016.

45 students in preschool and elementary school, aged between 4 and 12 years. The rest of them present intellectual disability. The therapy was given one hour a day, twice a week.

The Auditory-Verbal Method

In most special schools in México, the audio-verbal approach for the development of oral communication is used. It aims to ensure that each auditory handicap student acquires oral communication through the stimulation of the auditory pathway, thereby promoting and achieving high levels of intellectual development and social interaction.

The auditory-verbal method suggests an individual education according to the needs of each subject, along with periodic assessments of

progress in each subject, as well as hearing aid evaluation, calibration and proper management for those who make use of it and are subject to the auditory method (Cervera & Ygual, 1998); (Estabrooks, 2006). The use of this method in the classrooms is described below:

Level I: *Sound Level Consciousness.* The first requirement for AT is that the subject perceives the existence of sound, and to be able to differentiate between presence and absence of sound.

Level II: *Sound Alert.* The aim is that the subject is interested in the sound, to pay attention to it and locate it.

Level III: *Thick auditory discrimination.*

The aim is that the subject responds to the language in the undifferentiated period, i.e.; by the emotional tone and not the meaning, and that differentiates through hearing the sounds, noises, voice and language all its acoustic characteristics.

An auditory-verbal approach includes the use of a system of sequential and structured strategies that rely only on the sense of hearing, in order to provide access to linguistic information.

Auditory Training Software Description “AudiTS” for Developing Communication Skills

The instructor tells the beginning user that the training he will receive is gradual and that it consists on four modules. Text instructions are displayed on screen and then an image menu is shown to select the desired module. If the user has received this training before, he will give continuity to it selecting the appropriate module. See Figure 2.

The first three modules are presented in the following environments: animals (pet, farm, wild); transportation (land, air, sea); home; the street; nature.

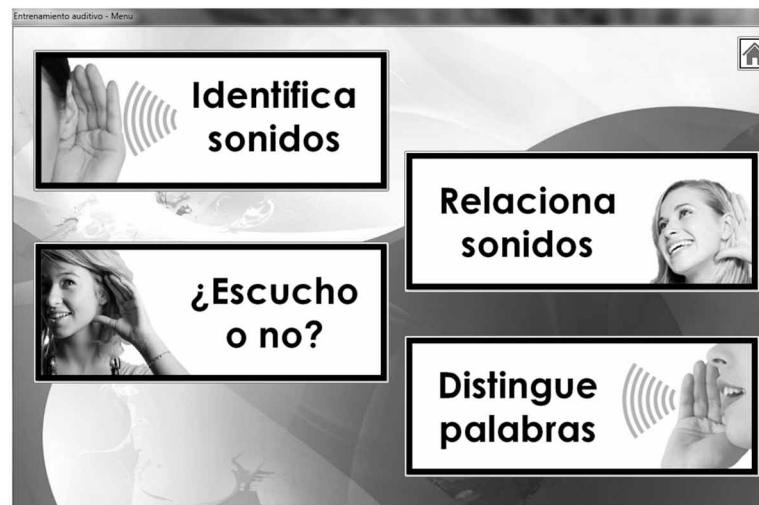
In the first module the user identifies the absence or presence of sound to choose a subject of the aforementioned environments. The first two modules contain an additional menu to select an item in the group. For example, if the user selects “farm animals”, an image is displayed on the screen of the computer and AudiTS can produce a random sound corresponding to the image; the user must answer whether or not there is sound by clicking on one of two buttons (yes or no).

The software validates whether or not there was sound. If the user succeeds, a visual message is displayed congratulating him, and if not, he is prompted to try again with another image; he will press a “Next” button to continue with another object. A digital counter is displayed in the bottom right of the screen that shows the ratio of hits between attempts to meet the user’s progress.

In the second sound identification module, the user selects an item from the proposed environments. An image is displayed and a sound is played. For example, the image of a cat is displayed and a meow is heard, this is with the intention of associating meaning with the signifier. The training consists in knowing the common sounds at home, street and nature, as well as animals and means

Figure 2. The menu of auditory training software (AudiTS)

Source: Prepared, 2016.



of transportation. In this section AudiTS does not validate answers because the child also interacts with the instructor.

In the third module the discriminant is applied in the same environments as in the first module. A topic is selected, four images are displayed and a sound associated with an image is produced; the user must validate the significance of sound with the significant image by clicking on the proper image. The images displayed in groups of four are always random and there is correspondence only with the sound, which the user must learn. The software performs the same tasks as in the first module: validates response, it shows a success or failure message and displays user performance.

The fourth module is divided into two areas: identifying and discriminating one syllable and the ability to identify and discriminate isolated words and contained words in sentences or phrases. To provide training about syllables a video showing an instructor pronouncing various syllables of cognate pairs is displayed, considering levels of difficulty and reinforcing learning with a text showing the syllable at the same time. Isolated words are presented in a video pronounced by an instructor emphasizing their joint and relating its meaning with an image. The words contained in the sentences are also shown in a video, and they are reinforced showing a text on the screen to

highlight the word in the sentence. In this module the instructor validates the child's performance relying on the report template.

Description of the Evaluation Instrument

The template to assess the child's performance is constructed from the matrix of methodological consistency, as shown in Table 3. The instructor records the qualification that the child obtained in each session in the column for each indicator.

Results obtained after about 4 months of effective use of AudiTS during a school period of eight months, excluding vacations and holidays, (see Table 2 for interpreting the results).

In Listening Skill the average obtained in the sound factor is 7; for syllables, 4; for words, 3. Children learned globally 30% of the vocabulary contained in the application (30 of 90 words) obtaining better performance in sound recognition.

In the Learning Behavior mode factor obtained in Behavioral Attitude is 4; motivation, 5; and the attention period was increased from 3 to 15 minutes.

In the AudiTS variable, the average for the multimedia elements factor is 5; user interface is 4 and auditory- verbal method is also 4.

Table 3. Template where the therapist monitors progress of the person receiving the therapy

Indicator	Qualification						
Presence/absence of sound	6	7	5	6	7	5	6
Sound identification	6	5	6	7	6	7	6
Sound discrimination	5	7	6	7	6	5	6
Identifies syllables	3	4	3	2	3	4	3
Discrimination syllables	2	4	3	4	3	3	4
Identifies/ Discriminates a syllable in cognate pairs	3	4	2	3	2	3	4
Syllable comprehension	2	3	4	2	3	2	3
Identifies isolated words	3	2	3	4	3	3	2
Identifies words in a sentence	2	1	2	3	2	1	2
Words comprehension in a sentence	2	1	2	3	2	3	2

Source: Prepared, 2016.

SOLUTIONS AND RECOMMENDATIONS

The solution presented modifies the traditional way of evaluating the child in therapy. The assessment tool is proposed to also consider elements, such as attitude, motivation and attention, as these are very important for a person to accept the therapy.

The template also suggests how to make the performance evaluation of a commercial computer program for AT.

The recommendations of the proposed solution are:

- The use of this software requires the supervision of a qualified person because it is a study that needs to evaluate quantitative and qualitative factors under observation.
- The software allows to evaluate the performance of the hearing ability of the user, but implies the need for the proposal template where the therapist monitors progress of the person receiving the therapy, which can be done manually or using a spreadsheet.
- The recommendation to use the software will reduce the technological gap in which these vulnerable groups currently are, and will be conducive to encourage the authorities and teachers in the use and benefit that it has IT in education and training.
- Another recommendation is to establish a link between the departments of computer science in public schools and special education for the development of educational software according to their specific needs.

FUTURE RESEARCH DIRECTIONS

The future trend in developing countries will not only bring the TIC's to the special education classrooms, but it will also be necessary to use information sciences mainly in the implementation and use of knowledge in educational organizations

due to the complex problems they show, as it is not easy to develop adaptive skills in them.

Currently conducting research projects in information science demand multidisciplinary task as it is in this case; phonetics, special education, computer science and psychology.

To offer computer science teachers and students resources for research and development of software, related to phonetics, speech recognition and aphasia.

The problematic identified should be resolved by the systems theory of management to interpret organizational behavior due to the complexity of the AT. This theory helps to the managers and teachers to understand the concepts: knowledge, adaptation, relationships and environment.

CONCLUSION

The element of therapy which had the greatest influence on children was the positive behavior, quickly identifying sounds and to a lesser extent the recognition of words. This is the area where more work is recommended.

Several special education teachers have accepted the implementation of the proposed program of AT. The viability depends on the approval of education authorities, as it is not only an economic resource issue, but it means also to prepare teachers for the acceptance of change and the use of this technology.

REFERENCES

Cervera, J. F., & Ygual, A. (1998). Entrenamiento de la percepción auditiva en niños con trastornos del lenguaje. *EDETANIA Estudios y propuestas de educación*, 15(1), 1-11. Recovered from <http://empresa.rediris.es/pub/bscw.cgi/d304118/Entrenamiento%20de%20la%20percepci%C3%B3n%20auditiva%20en%20ni%C3%B1os%20con%20trastornos%20del%20lenguaje.pdf>

- Estabrooks, W. (2006). *Auditory-verbal therapy and practice*. Washington, DC: Alexander Graham Bell Association for the Deaf and Hard of Hearing.
- Galloway, J. P. (2009). Contemporary Issues in Teaching and Learning with Technology. In M. Khosrow-Pour (Ed.), *Encyclopedia of Information Science and Technology*. IGI Global. doi:10.4018/978-1-60566-026-4.ch119
- González, I., & Torre, G. (2006). Guía de recursos de la deficiencia auditiva. APADA-Asturias, Oviedo.
- Kricos, P., & McCarthy, P. (2007). From ear to there: A historical perspective on auditory training. *Seminars in Hearing*, 28(2), 89–98. doi:10.1055/s-2007-973435
- Leirer, V. M., Glockner, F., Elbert, T., & Kolassa, I. T. (2009). An auditory computer-based training for mild cognitive impairment and mild Alzheimer's Disease - German prototype of the Brain Fitness Program. *Proceedings of the 2nd International Conference on Pervasive Technologies Related to Assistive Environments*.
- Llisterri, J. B. (1991). *Introducción a la fonética: el método experimental*. Barcelona: Anthropos.
- Mele, C., Pels, J., & Polese, F. (2010). A Brief Review of Systems Theories and Their Managerial Applications. *Service Science*, 2(1-2), 126-135. 10.1287/serv.2.1_2.126
- Merzenich, M., Pandya, P., & Tremblay, K. L. (2005). Roundtable discussion. *Plasticity and auditory training*. *Seminars in Hearing*, 26(3), 144–148. doi:10.1055/s-2005-916377
- Moore, D. R., & Amitay, S. (2007). Auditory Training: Rules and Applications. *Seminars in Hearing*, 28(2), 99–109. doi:10.1055/s-2007-973436
- Pérez, M. (2010). *La deficiencia auditiva*. Revista Digital “Innovación y Experiencias Educativas”. Granada, España.
- Rivas, L. A. (2004). *Cómo hacer una tesis de maestría?* Taller Abierto.
- Rochette, F., & Bigand, E. (2010). Auditory training in deaf children. Rethinking physical and rehabilitation medicine. *Collection de L'Académie Européenne de Médecine de Réadaptation*, 193-201.
- Rochette, F., Pescheux, P., & Bigand, E. (2008). Entrainement auditif et éducation auditive des enfants sourds. *Revue de Neuropsychologie*, 13(2), 1–19.
- Russo, N., Nicola, T., Zecker, S., Hayesa, E., & Krausa, N. (2005). Auditory training improves neural timing in the human brainstem. *Behavioural Brain Research*, 156(1), 95–103. doi:10.1016/j.bbr.2004.05.012 PMID:15474654
- Stacey, P. C., Raine, C. H., ODonoghue, G. M., Tapper, L., Twomey, T., & Summerfield, A. Q. (2010). Effectiveness of computer-based auditory training for adult users of cochlear implants. *International Journal of Audiology*, 49(1), 347–356. doi:10.3109/14992020903397838 PMID:20380610
- Strehlow, U., Haffner, J., Bischof, J., Gratzka, V., Parzer, P., & Resch, F. (2006). Does successful training of temporal processing of sound and phoneme stimuli improve reading and spelling? *Germany. European Child & Adolescent Psychiatry*, 15(1), 21–29. doi:10.1007/s00787-006-0500-4 PMID:16514506
- Sullivan, J. R., & Thibodeau, L. M. (2010). *Computer-based auditory training to improve speech recognition in noise by children with hearing impairment* (Doctoral Dissertation). University of Texas at Dallas, Richardson, TX.
- Toledo, G. A. (2005). Uso del speech analyzer para la enseñanza de la ortofonía, la fonética y la fonología españolas. *Revista de filología*, 293-304.
- Zhou, Y., Sim, K. C., Tan, P., & Wang, Y. (2012). MOGAT: Mobile Games with Auditory Training for Children with Cochlear Implants. *Proceedings of the 20th ACM international conference on Multimedia*, 429-438. doi:10.1145/2393347.2393409

ADDITIONAL READING

Chang, M., Iizuka, H., Naruse, Y., Ando, H., & Maeda, T. (2014). Proceedings of the 5th Augmented Human International. Conference, Article No. 28, ACM New York, NY, USA. doi:10.1145/2582051.2582079

Horvitz, E., Kadie, C., Tim Paek, T., & Hovel, D. (2003). Models of Attention in Computing and Communication: From Principles to Applications. *Communications of the ACM*, 46(3), 52. doi:10.1145/636772.636798

Lopes, M., Mendes, T. C., Batista, A. P., Rademaker, P., Leite de Carvalho, F., & Tabith, A. (2013). Pediatric phoniatry outpatient ward: Clinical and epidemiological characteristics. *Brazilian Journal of Otorhinolaryngology*, 79(2). doi:10.5935/1808-8694.20130029 PMID:23670320

Lozano, A. I., & Romero, E. (2010). Adquisición de las habilidades lingüísticas y cognitivas, relevancia para el aprendizaje del lenguaje escrito. *Umbral Científico*. 16(1) 8-12. Recovered from <http://www.redalyc.org/articulo.oa?id=30418644002>

KEY TERMS AND DEFINITIONS

E

Aphasia: An impairment of language, affecting the production or comprehension of speech and the ability to read or write.

Center for Multiple Attention: Schools for students with disabilities to develop various skills known in México as CAM.

Consistency Matrix: A Chart to verify the logical coherency in a research method and to develop a tool for data collection.

Deaf: The person whose hearing is not functional for ordinary life.

Hypoacoustic: A person whose hearing, though poor, is functional with or without prostheses.

Methodological Consistency: Research method, which consists in relating the basic elements in a table to verify the logical relationship of the study.

Motivational Context: Cognitive affective factor present in every act of learning.

Digital Storytelling in Language Classes

Mehrak Rahimi

Shahid Rajaei Teacher Training University, Iran

INTRODUCTION

Storytelling is an effective and unique way of communication that takes the listeners into an imaginary world, motivates them to think about the happenings and phenomena and sometimes even induces them to do things (Handler-Miller, 2004). Storytelling makes connections between a storyteller and the audience (Smyth, 2005) and thus lets both express and share their emotions and thoughts.

Story is an old-fashioned human experience (McDrury & Alterio, 2003) and “a natural component of society and culture” (Frazel, 2011, p. 6). Stories are made, told, read, and listened to by people with different tastes and views (Jones, 2006) for transferring knowledge and/or sharing experience. The social value of storytelling and writing lies in the fact that stories can link past, present, and future generations (Chung, 2007) of micro and macro cultures and pave the way for the nation to preserve their cultural heritage. It is believed that as the food people consume makes their bodies, the stories they hear construct their minds (Wright, 2008) and form their viewpoints on the world and life.

Recent studies on human’s memory show that the brain saves the events in the form of scenarios (Schank, 1990) or chains of happenings. The human’s brain is story-oriented and experiences are kept like storyboards and thus are not very easily forgotten. This feature of brain and memory has inspired practitioners and educationist to take advantage of storytelling as a technique for training, teaching, and learning (Handler-Miller, 2004). Storytelling can improve teacher quality (McDrury & Alterio, 2003) and is an effective teaching tool that helps students understand intricate and

complex experiences and concepts (Sadik, 2008). Stories prepare learners for communication; make them literate; engage them in an entertaining way (Huffaker, 2004); help them learn language forms; and expand their vocabulary stock (Wang, Li, & Dai, 2008). It is empirically verified that stories can promote students’ problem solving skills (Radbakhsh, Mohammadifar, & Kianersi, 2013); increase their attention and motivation (Mostafazadeh, 2010); lower the physiological and emotional anxiety (Zarei, et al., 2013); and increase their self-esteem (Soltani, Arian, & Angoji, 2013) while making learning more joyful and less frustrating (Rahimi & Soleymani, 2015a).

Stories convey their message when they are told orally, written in words, or are drawn as images. The most primitive way of recording and transferring stories into the future was cave wall drawings. This mechanism of knowledge transferring was altered fundamentally by the invention of writing and then printing press. In the 21st century, technology and its varying forms have had a far reaching influence on the way stories are told, stored, and shared. Traditional forms of storytelling now have evolved into modern ways of storytelling called digital storytelling (Frazel, 2011), where stories are told by the combination of narration, music, images, texts, and movies. In other words, the audiences do not just listen to a story or read it, they have this opportunity to listen, read, watch, and enjoy the combination of different media in the environment of digital devices.

This type of storytelling has attracted the attention of researchers especially pedagogues to scrutinize its impact on learning different subjects. Significant characteristics of digital stories such as flexibility, universality, and interactivity have made them a practical and powerful technologi-

DOI: 10.4018/978-1-5225-2255-3.ch213

cal tool in instruction. In this way, listening to stories can grant the opportunity for students to enhance their engagement in problem solving and deep learning as well as working collaboratively in teams.

Although digital storytelling has been found to be a valuable technique of teaching, language teachers have shown some resistance to use it in their teaching due to certain reasons including

- A lack of confidence in their ability to tell stories or read storybooks aloud.
- A feeling that the language in storybooks was too difficult.
- A feeling that the content of storybooks was sometimes too childish.
- A lack of understanding about the true value of using storybooks.
- A lack of understanding of how to use storybooks and of time to prepare a plan of work (Ellis & Brewster, 2014, p. 6).

One way to overcome these resistances is providing teachers with required resources, giving them technical and pedagogical support to develop appropriate teaching materials, and providing them with other teachers' experiences (Ellis & Brewster, 2014, p. 6). To give more guidance to language teachers, this chapter deals with basic features of digital stories, the educational values of digital stories, and how they can be made and used in teaching. The chapter thus is organized in four sections dealing with theoretical framework of digital stories, the salient features of digital stories, educational values of digital storytelling, and the making of digital stories.

BACKGROUND

Joe Lambert, Bernajean Porter, and Dana Atchley are the pioneers and founders of digital storytelling (Frazel, 2011). The term *digital storytelling* (DST) was coined by Dana Atchley, who adapted and improved her personal stories through the use

of digital media (McLellan, 2006). Later, Lambert was astonished by the way people can easily, fast, and with a little price capture their stories using this technique (Robin, 2008). In 1994, the San Francisco Digital Media Center, which later became the Center for Digital Storytelling (CDS), was founded by Joe Lambert, Dana Atchley, and Nina Mullen. Since then, this center has been responsible for developing, training and helping individuals who are interested in creating digital stories and telling personal stories (Robin, 2008).

It is hard to give an exact definition of digital storytelling because it has been used in many different ways. Digital storytelling employs the art of storytelling and technological tools using graphics, images, music and sound mixed together with the creator's own voice (Porter, 2004). It is the act of merging unmoving images with a narrated soundtrack including both music and voice (Bull & Kajder, 2004). Lambert (2009) defines digital storytelling as sharing one's story through several medium of voice, imagery, music, text, sound, video and animation. Based on a five-part definition, digital stories should (Kittle, 2009):

- Include a compelling narration of a story;
- Provide a meaningful context for understanding the story being told;
- Use images to capture and/or expand upon emotions found in the narrative;
- Employ music and other sound effects to reinforce ideas;
- Invite thoughtful reflection from their audience(s) (p. 169).

Digital stories can be made by both students and teachers by narrating different types of stories from personal experiences to historical events or literary works. The stories can have different topics based on the creators' interests and needs (Gyabak & Godina, 2011). Digital stories can be made very easily with different types of movie maker software programs or online platforms.

The main underlying theoretical basis of digital stories is Cognitive Theory of Multimedia Learn-

ing (CTML) (Mayer, 2005). CTML basically focuses on understanding how learning takes place through cognitive processes such as selecting (transferring some incoming data to working memory for further processing), organizing (classifying images and words into pictorial and verbal models respectively in the working memory), and integrating (connecting the pictorial, verbal, and schemata activated from the long term memory).

CTML works based on the assumption that the working memory's storage capacity is limited and the instructional design should minimize working memory load to utilize the greater capacity of long-term memory. This is possible through eliminating the unrelated information and chunking the related information in the long-term memory schema (Renkl & Atkinson, 2003). Hence, the visual information processing channel might become overloaded when students process on-screen graphics and on-screen text at the same time (Mayer, 2001). Using narration can reduce the cognitive load in the visual channel by processing words in the verbal channel (Figure 1).

Another scientific theory that can be associated with digital storytelling is 'mediation'. Based on sociocultural theory "human forms of mental activity arise in the interactions we enter into with other members of our culture and with the specific experiences we have with the artifacts produced by our ancestors and by out contemporaries"

(Lantolf, 2000, p. 79). The process of mediation in second language learning involves "1. mediation by others in social interaction, 2. mediation by self through private speech, and 3. Mediation by artifacts (for example, tasks and technology) (Ellis, 2008, p. 270).

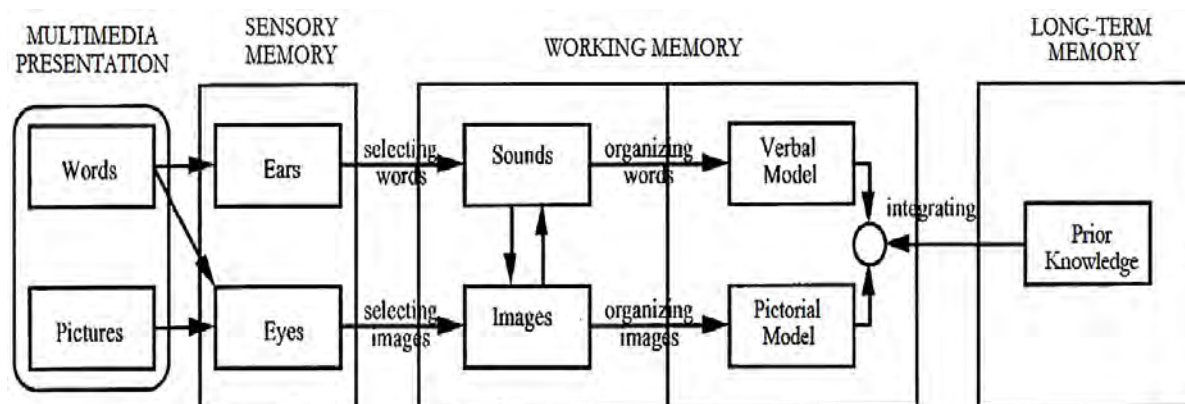
It is evident that computer-based multimedia can help students learn better than traditional classrooms (Bagui, 1998). Based on the theories mentioned, the success of multimedia in language learning can be attributed to three basic factors:

- The dual coding or using more than one code during the learning process (Najjar, 1996); as two media improve learning better than one (Mayer & Anderson, 1991).
- Presenting the information in a non-linear format (in a hypermedia format).
- Interactivity of multimedia in comparison with traditional methods and its flexible learning environment (Najjar, 1996).

Educational Value of Digital Storytelling

Some characteristics of digital storytelling such as flexibility, universality, and interactivity have made it a beneficial tool in learning and teaching (Park & Seo, 2009). Digital stories can engage students in problem solving (Kajder, 2004) and

Figure 1. Cognitive Theory of Multimedia Learning (CTML)
Mayer, 2005, p. 37.



deep learning (Pounsford, 2007). Further, they can make learners technologically competent (Ware, 2006) and increase their collaboration and group working (Di Blas, 2010). Digital stories can enhance learners' creativity (Park & Baek, 2009) and make them interested in learning (Tsou, Wang, & Tzeng, 2006). They utilize both teachers and students' imagination (Robin & Pierson, 2005) and enable teachers to integrate technologies and stories into the teaching program (Porter, 2006) more effectively. Digital stories can be applied in the classroom to motivate students (Coiro, 2003) and notify, teach or communicate with them (Robin, 2008) owing to its discontinuity, nonlinearity, and autonomy (Signes, 2008).

Digital stories provide students with authentic scenarios linked to their personal experiences and thus makes the content easier to be understood (Neo & Neo, 2010) while engaging both teachers and students actively in the class procedure (Dockter, Haug, & Lewis, 2010). Teachers have found multimedia-rich digital stories to be powerful learning tools for instruction across curricular subjects (Robin, 2008). It is believed that digital stories can develop five various skills consisting of (Robin, 2008):

- **Digital Literacy:** The ability to communicate with a community to discuss subjects and catch information.
- **Global Literacy:** The ability to read, respond, and interpret messages from a global view.
- **Technology Literacy:** The ability to apply computers and other technology.
- **Visual Literacy:** The ability to understand through visual images.
- **Information Literacy:** The ability to find and evaluate information.

It is also believed that using digital stories in the classroom has certain benefits such as (Frazel, 2011):

- Creating individualized digital stories as a class assignment.
- Engaging and motivating students to learn the content.
- Addressing the learning needs of K-12 students.
- Providing opportunities to apply emerging technologies as part of the students' learning.
- Supporting team teaching and learning across the curriculum.
- Providing an active instructional format and promoting group activities in the classroom.

Additionally, twelve unique characteristics of digital stories have been suggested by Schafer (2004) to have a vital role in learning and teaching including: interactivity, coherence, concreteness, involvement, cognitive effort, continuity, control, structure, vitality, spatiality, collaboration and immersion.

Digital devices are increasingly being used in education with the progress of technology. Some studies have been done on scrutinizing the value of digital stories in education in the last decade. The positive impact of using digital stories on literacy skills (Kajder, 2004), listening comprehension (Verdugo & Belmonte, 2007), understanding theoretical concepts (Coventry, 2008), learning of school subjects such as English, science, math, and social science (Sadik, 2008), creativity and cultural awareness (Benmayor, 2008), learning achievement, writing skill, verbal skills, critical thinking (Yuskel, Robin, & McNeil, 2011), learning motivation (Yang & Wu, 2012), reading comprehension and vocabulary learning (Chuang & Kue, 2013), literacy skills (Abdollahpour & Asadzadeh Maleki, 2012), listening comprehension (Soureshjani & Etemadim 2012), and listening anxiety (Rahimi & Soleymani, 2015a) is now evident.

How to Make Digital Stories

Making digital stories is not a difficult job, especially by the availability of lots of software programs made for producing movies and animation, both offline and online. Producing digital stories has two distinct parts: one part is related to language and content including the topic, difficulty level, and the length of the story; another part is related to technical and artistic aspects of making the multimedia file. Generally, seven elements have been identified by scholars in producing digital stories including (Lambert, 2010):

1. Point of view (what of the story) which shows the intention of the author of the story or the main point of it.
2. Dramatic question (why of the story) provokes the audience's inquisitiveness, will be answered by the end of the story and keeps the audience members' attention.
3. Emotional connection to the listener which involves the audience in the story or writing of the story.
4. Economy avoids overloading the viewer with excessive use of visuals and/or audio.
5. Pacing provides a rhythm to the story and copes with how slowly or quickly it moves.
6. The gift of voice (narration of the text) helps the audience understand the story through personalization of the narration.
7. The accompanying soundtrack (music) supports the story with appropriate music.

Digital stories can be made by a variety of media editing software including Movie Maker, iMovie, Photo Story, and Corel VideoStudio Pro (Skinner & Hagood, 2008). Digital stories can differ in length; but most stories generally last between two and five minutes. It is suggested that after minute 3, attention to the multimedia and thus comprehension declines (Rahimi & Soleymani, 2015b).

Generally speaking three stages are necessary in making digital stories: Preparation, production, and presentation.

Preparation stage is the stage before creating the story, where the audience and the way of presenting the story are determined. The story map for writing the script should be prepared here as well. In addition, the creator or the storyteller determines the type of final product (video or podcast) and how it will be presented to the audience (online or face to face). In the classroom situation, the teacher with the help of students decides about the theme of the stories and the type of audience. Then, they organize preliminary materials and plan for digital storytelling project/assignment. The teacher will decide at this point whether to have students work as individuals, in small groups, or as a whole class.

As a part of preparation, the teacher may choose to create an introductory digital story about the selected topic and have students brainstorm ways to find meaning, point of views, or emotional connections they had with the narrator. In this part the language teacher pays careful attention to the content of the story by considering the difficulty level of the story and the number of new words and grammatical points it has. One way to take these into account is calculating the readability of the story based on the available indices. The difficulty level of the story should have the features of 'the comprehensible input' and thus be in the range of students' language repertoire, or competence.

In *production stage*, the digital story file is produced by the teacher or students. Once all the resources and the storyboard are ready, students begin production phase. Students should be familiar with the media editing software to create multimedia stories. However, it is possible for the students to create an audio story, or a podcast using audio production software. The teacher acts as a mentor, and aids students with activities such as putting the slides in order or timing the slides. Music and sound effects may be employed. Most

of the required media can be found from different websites and programs. Enthusiastic students create much of their own media.

The prepared digital stories should be reviewed and assessed by other colleagues or peers before the files are going to be shared. Based on the provided feedback the file(s) would be edited. Any problematic content including words, and grammatical structures would be modified at this stage.

Presentation stage is the last but the most critical step among all in which prepared files will be used in the classroom or shared through the Internet (Frazel, 2011). For presentation, the digital story should be saved onto a file sharing website or written onto a CD or DVD.

There are numerous variations of this process and many genres of digital stories and digital story making applications are available (Frazel, 2011).

Digital Storytelling Teachers

In order to make digital stories the following steps should be followed by teachers:

Stage One

1. Selecting a topic for the story based on students' interest and needs.
2. Deciding on the purpose of the story. Is it going to inform, convince, provoke, or improve language skills?
3. Deciding on the point of view of the story.
4. Writing a script for the story.
5. If the story is taken from English literature, the text should undergo linguistic filters to make the text appropriate for the students based on their language proficiency level.

Stage Two

1. Searching for image resources for the story including pictures, drawings, or photographs.
2. Finding music resources and sound effects.
3. Importing images into a media editing program (e.g. Movie Maker).

4. Modifying images and/or image order, if necessary.
5. Mixing the narration and images and creating necessary special effects and graphics using the facilities of the software.
6. Saving the digital story file onto a DVD (MP4) to be watched and listened to by technological devices (such as cell phones, tablets, or laptops).

Stage Three

1. Reviewing of the digital story by colleagues.
2. Using the received feedback to improve and revise the story.
3. Finalizing and preparing the digital story to be performed in the classroom (Frazel, 2011).

Teachers can use different types of questions for developing critical thinking activities to accompany digital storytelling projects. These questions are useful for getting students to think about different parts of the story and improve both their general and language knowledge. Followings are sample questions that can be asked

- What are other possible ways to tell this story?
- What do we expect the audience to tell us at the end of this story?
- What is the role of music in this story?
- What additional information might the audience need to understand the story better?
- What happens if the music and special effects (e.g., graphics) are removed?

Language teachers can use digital stories in teaching four language skills, that is listening, speaking, reading, and writing. The teachers can use this invaluable teaching content to arise students' interest to go through a language task or work on digital stories as independent language tasks. Different types of exercises and activities

can be designed to be done during listening and watching the files or in the phase after it.

Digital Storytelling Students

Morra (2014) has given the detailed description of how teachers and students can make digital stories together in the classroom by specifying an 8-step cycle (Figure 2).

Start with an Idea

Like all stories, digital stories should begin with an idea such as student's questions or topic of a text or lesson. When the basic ideas/themes are discussed and agreed upon, the teacher or students should convert them to concrete materials such as a draft, an outline or a concept map.

Research/Explore/Learn

Students should gather information about the topic they have selected. Different resources are recommended to find more information about the topic; the Internet seems to be the more useful one.

However, atlases, encyclopedias, and periodicals are recommended to be searched as more authentic and trustworthy resources.

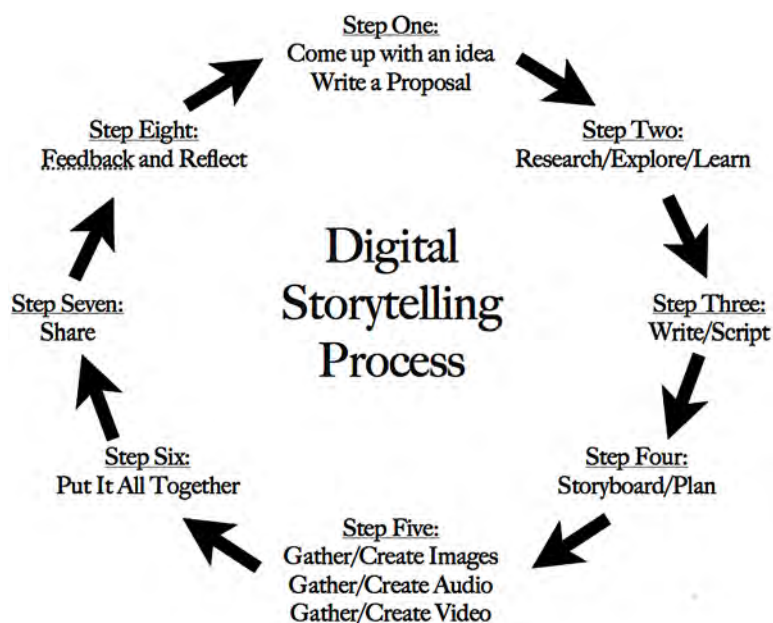
Write/Script

Based on the general theme and the specific information found, the students start writing the script. This is the first draft of the story. The script can be reviewed by peers and the teacher for corrections and improvement. So it is possible that different drafts are written and re-written before the final script is produced. It is noteworthy that script writing of the story should follow the process of writing including drafting, revising, editing, and producing the final product based on the feedback received. As a result digital stories are suitable language tasks for collaborative writing in language classes.

Storyboard/Plan

The writing will be converted into storyboards by the help of which students prepare a *graphic*

*Figure 2. The 8-step cycle of making digital stories
Morra, 2014.*



representation of the script with specified characters and sequences of events.

Storyboards can be made on paper or using media producer software (Figure 3).

As the storyboard is the basis of the story, it is recommended that students spend quality time on this stage to produce storyboards that are understandable and manageable to be made. Otherwise they may face serious problems in producing the multimedia file.

Gather and Create Images, Audio, and Video

To change the storyboard into a real story, images, audios, and videos are required. Students should follow the steps of the storyboard carefully and find the necessary media to make the digital story.

Put It All Together

At this stage students put everything together based on the storyboard in the technology environment. They have to use the affordances of the media producing software to make the digital story as attractive as possible. In other words, the pictures, narration, and the background music will be mixed while special digital effects are added in the environment of the software. This stage needs much of taste and artistry from the side of students.

Share

When the story is made, it should be shared with the audience. This can be done online using the available platforms or offline as digital video files copied on DVDs.

Reflection and Feedback

At the end, the students reflect on their own work and give feedback to the works of others. Online or in-class sessions can be used to help students get feedback and opinion about their made stories (Kajder et al, 2005). In this phase students can do lots of activities including discussions and debates and writing critical paragraphs.

SOLUTIONS AND RECOMMENDATIONS

This chapter aimed at making language teachers familiar with a handy technological artifact that can be easily made both by teachers and students and used in language classes. Digital stories have been reported to be valuable instructional materials in mainstream education, however, their miraculous role in promoting language learning is just being revealed.

Figure 3. A simple storyboard
Bird, 2011



Digital stories can be used in teaching and practicing almost any type of language activity including macro language skills (listening, speaking, reading, and writing) and micro skills (pronunciation, vocabulary, and grammar). As multimedia files, they can improve the capacity of working and short term memory and make things being remembered longer. As collaborative tasks, they let language learners participate in making stories and then acquire new points by cooperation and collaboration. Further, as they are symbolic artifacts, they can enhance the cognitive process and impact the way students think and process information. One possible benefit of human/technology interaction “is *transfer* of the partnered skill to the individual such that a particular mental operation that entails technology is mastered to a greater degree as a result of the partnership” (Lantofl, 2000, p. 93).

It is noteworthy to mention that without necessary skills and knowledge base of making digital stories their effect in language classes may be spoiled, as a part of teacher Technological Pedagogical and Content Knowledge (TPACK) is being able to integrate technology in instruction based on sound pedagogical theories and practices.

FUTURE RESEARCH DIRECTIONS

This study investigated the educational values of digital stories in mainstream education in general and teaching English as a foreign language in particular. It also familiarized language teachers with how to make and use digital stories in language classes.

Language educationists and teachers are urged to do empirical studies on the theoretical guidelines this chapter or similar works offer to verify the value of digital stories in language classes. This may include studies done to compare the effects of digital story in teaching receptive and productive language skills as well as the impact of digital storytelling teachers vs. digital storytelling students.

Comparative quantitative and qualitative studies on the way teachers/students with different personal characteristics make and use digital stories in teaching and learning English are required to be examined. The mediating role of nationality and the context of teaching and learning can also be investigated.

Other types of multimedia contents such as multimedia stories and posters are also required to be introduced to teachers and their effects be examined empirically.

CONCLUSION

This chapter presents some guidelines on what digital stories are and how they can be made to be used in teaching, particularly in language classes. The chapter opens with an introduction on the key role of literature and especially stories in peoples’ life and education. It further goes on with probing into the value of digital storytelling as an innovative technique in pedagogy. The chapter ends with listing educational values of digital stories and how teachers can make and use them in their teaching.

REFERENCES

- Abdollahpour, Z., & Asadzadeh Maleki, N. (2012). The impact of exposure to digital flash stories on Iranian EFL learners’ written reproduction of short stories. *Canadian Journal on Scientific and Industrial Research*, 3(2), 40–53.
- Bagui, S. (1998). Reasons for increased learning using multimedia. *Journal of Educational Multimedia and Hypermedia*, 7(7), 3–18.
- Benmayor, R. (2008). Digital storytelling as signature pedagogy for the new humanities. *Arts and Humanities in Higher Education*, 79(2), 188–204. doi:10.1177/1474022208088648

- Bull, G., & Kajder, S. (2004). Digital storytelling in the language arts classroom. *Learning and Leading with Technology*, 32(4), 46–49.
- Chuang, W., & Kuo, F. (2013). Improving reading comprehension among Taiwanese EFL young EFL young learners using digital stories. *Proceedings of 2013 Asian Literacy Conference*, 134–146.
- Chung, S. K. (2007). Art education technology: Digital storytelling. *Art Education*, 60(2), 17–22.
- Coiro, J. (2003). Reading comprehension on the internet: Expanding our understanding of reading comprehension to encompass new literacies. *The Reading Teacher*, 56(5), 458–464.
- Di Blas, N., Paolini, P., & Sabiescu, A. (2010). Collective digital storytelling at school as a whole-class interaction. *Proceedings of the 9th international Conference on interaction Design and Children*, 11–19. doi:10.1145/1810543.1810546
- Dockter, J., Haug, D., & Lewis, C. (2010). Redefining rigor: Critical engagement, digital media, and the new English/Language arts. *Journal of Adolescent & Adult Literacy*, 53(February), 418–420. doi:10.1598/JAAL.53.5.7
- Ellis, G., & Brewster, J. (2014). *Tell it again: The storytelling handbook for primary English language teachers*. British Council.
- Ellis, R. (2008). *The study of second language acquisition* (2nd ed.). Oxford, UK: Oxford University Press.
- Frazel, M. (2011). *Introduction to Digital Storytelling*. ISTE.
- Gyabak, K., & Godina, H. (2011). Digital storytelling in Bhutan: A qualitative examination of new media tools used to bridge the digital divide in a rural community school. *Computers & Education*, 57(4), 2236–2243. doi:10.1016/j.compedu.2011.06.009
- Handler-Miller, C. (2004). *Digital Storytelling: A creator's guide to interactive entertainment*. Oxford, UK: Elsevier.
- Huffaker, D. (2004). Spinning yarns around the digital fire: Storytelling and dialogue among youth on the internet. *Information Technology in Childhood Education Annual*, 1, 63–75.
- Kittle, P. (2009). Student engagement and multimodality: Collaboration, schema, identity. In A. Herrington, K. Hodgson, & C. Moran (Eds.), *Teaching the new writing: Technology, change, and assessment in the 21st-Century classroom* (p. 169). New York: Teachers College.
- Lambert, J. (2009). *Digital storytelling: Capturing lives, creating community*. Berkeley, CA: Digital Diner Press.
- Lambert, J. (2010). *Digital Storytelling cookbook*. Available online at: <http://www.storycenter.org/storage/publications/cookbook.pdf>
- Lantolf, J. P. (2000). Second language learning as a mediated process. *Language Teaching*, 33(02), 79–96. doi:10.1017/S0261444800015329
- Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions. *Educational Psychologist*, 32(1), 1–19. doi:10.1207/s15326985ep3201_1
- Mayer, R. E. (2001). *Multimedia learning*. New York, NY: Cambridge University Press. doi:10.1017/CBO9781139164603
- Mayer, R. E. (2005). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (pp. 31–48). New York: Cambridge University Press. doi:10.1017/CBO9780511816819.004
- Mayer, R. E., & Anderson, R. B. (1991). Animations need narrations: An experimental test of dual-coding hypothesis. *Journal of Educational Psychology*, 83(4), 484–490. doi:10.1037/0022-0663.83.4.484
- McDrury, J., & Alterio, M. (2003). *Learning through storytelling in higher education: Using reflection and experience to improve learning*. London: Kogan Page.

- McLellan, H. (2006). Digital storytelling in higher education. *Journal of Computing in Higher Education*, 19(1), 65–79. doi:10.1007/BF03033420
- Meadows, D. (2003). Digital storytelling: Research-based practice in new media. *Visual Communication*, 2(2), 189–193. doi:10.1177/1470357203002002004
- Morra, S. (2014). *8 steps to great digital storytelling*. Available online at: <http://edtechteacher.org/8-steps-to-great-digital-storytelling-from-samantha-on-edudemic/>
- Mostafazadeh, F. (2010). Storytelling: A new clinical education method. [in Persian]. *Research in Medical Teaching*, 2(2), 53–58.
- Najjar, L. J. (1996). Multimedia information and learning. *Journal of Educational Multimedia and Hypermedia*, 5(2), 129–150.
- Neo, M., & Neo, T. K. (2010). Students' perceptions in developing a multimedia project within a constructivist learning environment: A Malaysian experience. *The Turkish Online Journal of Educational Technology*, 9(1), 176–184.
- Park, E. J., & Seo, J. H. (2009). Applying digital storytelling technique to website navigation for improving emotional user experience. *Proceeding of the International Association of Societies of Design Research*, 4125–4128.
- Porter, B. (2006). Beyond words: The craftsmanship of digital products. *Learning and Leading with Technology*, 28–31.
- Pounsford, M. (2007). Using storytelling, conversation and coaching to engage: How to initiate meaningful conversations inside your organization. *Strategic Communication Management*, 11(3), 32–35.
- Radbakhsh, N., Mohammadifar, M., & Kiani-Ersi, F. (2013). The impact of games and stories on children's creativity. [in Persian]. *Innovation and Creativity in Humanities*, 2(4), 177–195.
- Rahimi, M., & Soleymani, E. (2015a). The impact of digital storytelling on language learners' listening anxiety. *Quarterly of Educational Psychology*, 6(1), 87–111.
- Rahimi, M., & Soleymani, E. (2015b). *Digital Storytelling in Language Classes Discovering the Power of Corel® VideoStudio® Pro X7*. Tehran: Rash Publications.
- Renkl, A., & Atkinson, R. (2003). Structuring the transition from example study to problem solving in cognitive skill acquisition: A cognitive load perspective. *Educational Psychologist*, 38(1), 15–22. doi:10.1207/S15326985EP3801_3
- Robin, B., & Pierson, M. (2005). *A multilevel approach to using digital storytelling in the classroom*. Digital Storytelling Workshop, SITE 2005, University of Houston. Available online at: <http://faculty.coe.uh.edu/brobin/homepage/SITE2005-article.htm>
- Robin, B. R. (2008). Digital storytelling: A powerful technology tool for the 21st century classroom. *Theory into Practice*, 47(3), 220–228. doi:10.1080/00405840802153916
- Sadik, A. (2008). Digital storytelling: A meaningful technology-integrated approach for engaged student learning. *Educational Technology Research and Development*, 56(4), 487–506. doi:10.1007/s11423-008-9091-8
- Schafer, L. (2004). Models for Digital Storytelling and interactive narratives. *The Proceedings of the 4th International Conference on Computational Semiotics for Games and New Media*, 148–155.
- Schank, R. (1990). *Tell me a story: Narrative and intelligence*. Evanston, IL: Northwestern University Press.
- Signes, C. G. (2008). *Practical Uses of Digital Storytelling*. Digital Storytelling/Relato Digital. Available online at: http://www.uv.es/gregoric/DIGITALSTORYTELLING/DS_files/DST_15_ene_08_final.pdf

Smyth, J. (2005). *Storytelling with young children*. Watson, ACT: Early Childhood.

Soltani, M., Arian, K., & Angaji, L. (2013). The impact of group storytelling on female second grade primary-school students' self-esteem. [in Persian]. *Quarterly of Educational Psychology*, 29(9), 95–108.

Soureshjani, K., & Etemadi, N. (2012). Listening comprehension success among EFL preschool children using internet-based materials. *Journal of Social Sciences and Humanities*, 7(1), 243–251. Retrieved from <http://journalarticle.ukm.my/5025/>

Tsou, W., Wang, W., & Tzeng, Y. (2006). Applying a multimedia storytelling website in Foreign Language Learning. *Computers & Education*, 47(1), 17–28. doi:10.1016/j.compedu.2004.08.013

Verdugo, D. R., & Belmonte, I. A. (2007). Using digital stories to improve listening comprehension with Spanish young learners of English. *Language Learning & Technology*, 11, 87–101.

Wang, D., Li, J., & Dai, G. (2008). A pen and speech-based storytelling system for Chinese children. *Computers in Human Behavior*, 24(6), 2507–2519. doi:10.1016/j.chb.2008.03.014

Wright, A. (2008). *Storytelling with children* (2nd ed.). Oxford University Press.

Yuksel, P., Robin, B., & McNeil, S. (2011). Educational uses of digital storytelling all around the world. In M. Koehler & P. Mishra (Eds.), *Proceedings of Society for Information Technology & Teacher Education International Conference* (pp. 1264–1271). Chesapeake, VA: AACE.

Zarei, K., Parand, Z., Seyedfatemi, N., Khoshbakht, F., Haghani, H., & Zarei, M. (2013). The impact of storytelling on physiologic anxiety, anxiety and social anxiety of school-age hospitalized patients. [In Persian]. *Medical Surgical Nursing Journal*, 2(3-4), 115–121.

ADDITIONAL READING

Alameen, G. (2011). Learner digital stories in a Web 2.0 age. *TESOL Journal*, 2(3), 355–369. doi:10.5054/tj.2011.259954

Alexander, B. (2011). *The new digital storytelling: Creating narratives with new media*. Santa Barbara: Praeger.

Alexander, B., & Levine, A. (2008). Web 2.0 storytelling: Emergence of a new genre. *EDUCAUSE Review*, 43(6), 40–56.

Azmi, A. L., & Nuraihan, M. D. (2013). Pre-service ESL teachers' perceptions of parody integration in digital stories. *World Applied Sciences Journal*, 21, 28–35.

Bird, R. (2011). 3D character animation. Retrieved from <https://software.intel.com/en-us/articles/3d-character-animation-part-1>

Farmer, L. (2004). Using technology for storytelling: Tools for children. *New Review of Childrens Literature and Librarianship*, 10(2), 155–168. doi:10.1080/1361454042000312275

Li, L. (2007). Digital Storytelling: Bridging traditional and digital literacies. In T. Bastiaens & S. Carliner (Eds.), *Proceedings of World Conference on ELearning in Corporate, Government, Healthcare, and Higher Education 2007* (pp. 6201–6206). Chesapeake, VA: AACE; Available online at <http://www.editlib.org/p/26774>

Pounsford, M. (2007). Using storytelling, conversation and coaching to engage: How to initiate meaningful conversations inside your organization. *Strategic Communication Management*, 11(3), 32–35.

Tucker, G. (2006). First person singular: The power of digital storytelling. *Screen Education*, 42, 54–58.

KEY TERMS AND DEFINITIONS

Collaborative Storytelling: A kind of digital storytelling by which groups of students work on certain themes to produce and share their stories with others by collaborative technologies such as online story making platforms.

Digital Storytelling: Sharing one's story through several medium of voice, imagery, music, text, sound, video and animation using technological tools.

Movie Makers: Software programs that let the users combine pictures, sound, music, and movie to make multimedia content.

Multimedia: The combination of several media (sound, image, graphics, movies, music) to present content.

Multimedia Cognitive Theory of Learning (MCTL): The underlying theoretical basis of digital storytelling that deals with how learning takes place through cognitive processes such as selecting, organizing, and integrating.

Multimedia Learning: The kind of learning that focuses on how people learn more deeply from words and graphics than from words alone.

Storytelling: Narrating sequences of events in an artistic manner for a group of audience orally or in written words.

Teacher-Made Digital Stories: Digital stories that are made by teachers based on the objectives of the lesson to teach a certain point/objective.

Distance Teaching and Learning Platforms

E

Linda D. Grooms

Regent University, USA

INTRODUCTION

The knowledge explosion, the increased complexity of human life, and the ubiquitous, 24/7 nature of technology coupled with the globalization of the marketplace herald the need to embrace the most effective methods and formats of teaching and learning. Currently providing powerful educational opportunities, the science and technology of distance learning continues to multiply at unprecedented rates. Where historically traveling from village to village verbally disseminating knowledge was the only process of training those at a distance, today's learners eagerly embrace the rapidly expanding web-based delivery systems of the 21st century, which offer a plethora of educational alternatives. So with this rapidly changing distance educational landscape, one must question, what exactly is distance teaching and learning, how has it evolved, and what is its future?

BACKGROUND

In very simplistic terms, distance learning is just that--learning that occurs at a distance (Rumble & Keegan, 1982; Shale, 1990; Shale & Garrison, 1990) or that which is characterized by a separation in geographical proximity and/or time (Holmberg, 1974, 1977, 1981; Kaye, 1981, 1982, 1988; Keegan, 1980; McIsaac & Gunawardena, 1996; Moore, 1973, 1980, 1983, 1989a, 1989b, 1990; Ohler, 1991; Sewart, 1981; Wedemeyer, 1971). In his 1986 theory of transactional distance, Moore (Moore & Kearsley, 1996) defined distance not only in terms of place and time but also in terms of structure and dialogue between the learner and the instructor. In this theory, distance

becomes more pedagogical than geographical. As structure increases, so does distance. As dialogue increases, distance declines thus establishing the foundational role interaction plays in the distance learning environment. Saba (1998) furthered this concept concluding,

... the dynamic and systemic study of distance education has made 'distance' irrelevant, and has made mediated communication and construction of knowledge the relevant issue So the proper question is not whether distance education is comparable to a hypothetical 'traditional,' or face-to-face instruction, but if there is enough interaction between the learner and the instructor for the learner to find meaning and develop new knowledge. (p. 5)

To facilitate greater interaction in the geographically and/or organizationally dispersed distance environment, today the convergence or fusion of technologies enable individuals to overcome the barrier of separation, affording institutional and learner opportunity to transcend intra- and inter-organizational boundaries, time, and even culture. By definition, the paradigm of distance, online, or e-learning revolutionizes the traditional environment; however, even with this change, learning, which involves some manner of interaction with content, instructor, and/or peers, remains at the core of the educational process.

Although imperative in both environments, research shows these three types of interaction to be the hub of the ongoing traditional versus distance argument. Traditionalists often fear that with anything other than face-to-face instruction, interaction somehow will decrease thus making learning less effective, when in reality, numerous

DOI: 10.4018/978-1-5225-2255-3.ch214

studies have revealed no significant difference in the learning outcomes between traditional and distance courses (Russell, 1999). In fact, distance courses have been found to “match conventional on-campus, face-to-face courses in both rigor and quality of outcomes” (Pittman, 1997, p. 42). Despite these findings, critics still abound.

Two distinguishing characteristics of the non-traditional environment--individualized learning and flexibility--often arouse suspicion and caution among traditionalists (Grooms, 2000). Many are convinced that with any form of study outside the confines of the typical brick and mortar, “every vestige of intellectual rigor [will] disappear into oblivion.... [These skeptics interpret] individualized learning as individualized isolation, especially from faculty, and they look on flexibility as no more than a synonym for escape from regulation and responsibility” (Gould, 1972, p. 9).

In contrast, with their introduction of Equivalency Theory, Simonson, Schlosser, and Hanson (1999) accentuated the concept of equivalency as “central to the widespread acceptance of distance education” (p. 72) thus supporting Keegan’s (1989) call for parity in quality, quantity, and status. Further, recognizing the need to bring integrity and prestige to the field, Shale and Garrison (1990) suggested building a framework based not on isolation but upon interdependence, which would imply that distance learning would merely become an alternative method for delivering traditional content with the context dictating the type of interaction required. So how did we get to where we are now?

MAIN FOCUS OF THE ARTICLE

Distance Learning Evolution

As previously mentioned, distance learning has been with us in one form or another virtually since the creation of time. For years, itinerant teachers traveled from village to village verbally disseminating information to those hungry for knowledge;

however, the invention of Guttenberg’s printing press in 1440, made possible serious distribution of learning to larger numbers of people.

Capitalizing on this broader use of print media, correspondence study became a popular form of distance education, the first record of which was in 1728 when Caleb Philipps’ advertised the introduction of shorthand (Battenberg as cited in Baath, 1980 & Holmberg, 1986). Often conjuring thoughts of isolation and autonomy, this record of instruction mirrored those images. In fact, in this account there was no mention of interaction of any type other than what was inherent with the content.

Over a hundred years later in his 1833 Swedish advertisement, although not directly stated, Meuller’s offer to study composition seems to be the first to imply some form of exchange between the student and teacher. More definitively, in 1840, the most acknowledged root of distance learning explicitly employing learner-instructor interaction began in the United Kingdom. Using passages from the Bible, Isaac Pitman taught shorthand (Baath, 1980; Holmberg, 1974; Kaye, 1988; Rumble, 1986), yet this time, once learners transcribed these passages, they were returned for correspondence with the teacher via the penny post, thus some called it postal teaching (Dewal, 1988).

As evidenced in these early days of pure correspondence education, any offered guidance transpired through some form of dispatched communication such as the mail (Wedemeyer, 1971) and student contact, even with the instructor, was not necessarily encouraged. This is clearly seen in Keegan’s (1980) classic article, *On Defining Distance Education*, where he documented that in its strictest sense, pure correspondence study advocates specified that “students enrol [sic] with them because they ‘want to be left alone’” (p. 31). Directly challenging this belief, Holmberg (1982) later advocated that “any post-graduate distance study must have a truly communicative character if more is meant than merely providing reading lists and odd comments on students’ work” (p. 259).

While print remained the primary mode of distance learning until the 1920s and 30s, the introduction of radio broadcasts soon followed with television and satellite delivery systems initiating the labor pains for the birth of the current online technological revolution. Prior to the advent of the World Wide Web (WWW) in the early 1990s, interaction continued to transpire primarily between the learner and content with occasional interaction between the learner and the instructor through such means as telephone and videoconferencing.

As the distance learning landscape continued to evolve, learner-instructor interaction became increasingly important, thus catapulting the first of two significant paradigm shifts. While some (Daniel & Marquis, 1979) were accentuating the importance of getting the right independence-interaction mixture, others heralded the positive impact (e.g. Cookson, 1989; Robinson, 1981) and, even more so, the critical nature of the tension between the role of technology and type of educational process with its accompanying interaction in the distance learning environment (e.g. Bocarnea, Grooms, & Reid-Martinez, 2006; Grooms, 2000, 2003; Grooms & Reid-Martinez, 2011, 2012, 2013).

The Current State of Affairs

To be embraced, any new mode or method of education such as distance learning with its multiple technological changes must do more than merely emulate the status quo of traditional education. The virtual environment of the 21st century claims to do just that. While offering flexibility from traditional geographical proximity and time constraints (Barnes & Greller, 1994; Harasim, 1990; Hiltz & Johnson, 1990; Kaye, 1989; Moore, 1983), computer-mediated communication (CMC) (Harasim, 1993; Kaye, 1989; McIsaac & Gunawardena, 1996) and other information and communication technology (ICT) serve as excellent participation equalizers or what

Szecszy (2011) refers to as “leveling devices.” These unprecedented technological advances increasingly blur the line between traditional face-to-face learning and that which occurs at a distance.

While multiple studies have indicated there is no significant difference between distance and traditional learning effectiveness, the geographical dispersion of people, shifting market conditions, and rapid technological changes continue to compel transformation in the way we do business both in the marketplace and in the halls of academe. Promising to deliver increased access, quality, and efficiency of learning in an ever-growing competitive market (Benoit, Benoit, Milyo, & Hansen, 2006), as it moved online, the technology of higher education altered both teaching and learning (Kapitzke, 2000) and instructor and student roles (Stadtlander, 1998). With this in mind and regardless if its now considered online, e- or m-, learning is now mediated through synchronous (interactive/real time) or asynchronous (delayed interaction) means and the most critical consideration continues to remain in aligning the task, the delivery method, and the delivery format.

Distance Learning Delivery Methods

Almost 150 years following the advent of postal teaching and the first record of any form of learner-instructor interaction, Linda Harasim (1989), a pioneer in the online classroom, clearly differentiated three delivery methods that she believed distinguished traditional, distance, and online education: *one-to-many*, as in the traditional lecture method when one instructor addresses many students; *one-to-one*, as in the tutorial method; and *many-to-many*, a collaborative process with students learning from each other, with or without an instructor. In the first method, learners are mere passive recipients of knowledge and information, whereas in the latter two, they are actively involved in the learning process. A clear shift in the role of the instructor transpired: from information dispenser to one who facilitates an environment

where knowledge “emerges from active dialogue among those who seek to understand and apply concepts and techniques” (Hiltz, 1990, p. 135).

Although the traditional face-to-face environment is time-place dependent, it allows for the implementation of all three delivery methods contingent upon the instructional task. In contrast, the distance and online environments are time-place independent and mediated, facilitating flexibility and reflective response; however, in like manner they also align method with context. When appropriate, distance learning uses the *one-to-one* (e.g., print media, programmed instruction, adaptive educational hypermedia, or even 3D and immersive virtual worlds) or *one-to-many* (e.g., audio teaching, streaming video, podcasting, or

even videologging such as YouTube) methods, while online classes additionally employ the *many-to-many* concept, which forces the second paradigm shift: the need for peer interaction which often uses such platforms as chat rooms, instant messaging, or one of the many web conferencing options.

Distance Learning Delivery Formats

Distance learning formats typically corral into four arenas: print, audio, video, or digital (see Table 1) with all serving as viable options. Apart from media or technology differences, communication within these formats can either be one-way with the learner taking a passive role or two-way with

Table 1. Distance learning delivery formats

Media	Passive One-Way	Interactive Two-Way	
		Asynchronous Delayed	Synchronous Real Time
Print	• Syllabi • Texts • Instructor notes • Study guides • Workbooks • Fax		
Audio	• Radio • Audiotape • CD Rom • Voicemail		• Telephone • Audio Conferencing
Video	• Videotapes • DVD • Film • Cable, Broadcast, and Digital Television (one-way audio and video)		Satellite Videoconferencing (two-way audio and one-way video)
Digital	• CAI (computer assisted instruction) – using the computer as a self-contained teaching machine • Webcasts • Streaming Video • Podcasting & its varied forms (e.g. Vodcasting & Phoningcasting) • Some forms of blogging such as moblogging, linklogging, photologging, edublogging, and videologging such as YouTube	• CMI (computer-mediated instruction) • E-mail • Texting • Bulletin Boards (Threaded Discussions; Newsgroups) • Listservs • Wikis • Web-based instruction (e.g., Blackboard) • Cloud Technology (e.g., Google Apps) • Open Source Learning Platforms (e.g., Moodle) • Social Networking (e.g., Facebook, LinkedIn, and Google+) • Other forms of blogging and many of its variations including microblogging (e.g., Twitter, Tumblr, and SnapChat)	• Web Conferencing (e.g., Webinars, Google Hangouts, Blackboard Collaborate, and GotoMeeting) • Chat Rooms • Instant Messaging • Shared Whiteboards • Interactive Optical Sensory Technology • 3D and Immersive Virtual Worlds • Adaptive educational hypermedia

the learner interacting with either the instructor or peers.

Whether using syllabi, texts, instructor notes, or other forms of print, today this medium still remains a significant component in distance learning. In addition, where telephones, audio conferencing, and video technology augmented print in the latter part of the 20th century, today the exponential proliferation of the Internet has provided a plethora of independent and collaborative learning opportunities, both synchronous and asynchronous. Regardless of the format, researchers (Beetham & Sharpe, 2007; Rourke & Coleman, 2011) caution that rather than technology driving pedagogy, pedagogy should drive technology with the ultimate goal always of learning.

Most recently with the plethora of multiple media at hand, many educational institutions have turned to massive open online courses (MOOCs), while others have employed a blended or hybrid learning approach combining some form of the traditional with the digital. Osguthorpe and Graham (2003) summarized the latter approach very well.

Those who use blended approaches base their pedagogy on the assumption that there are inherent benefits in face-to-face interaction (both among learners and between learner and instructor) as well as understanding that there are some inherent advantages to using online methods in their teaching. Thus, the aim of those using blended learning approaches is to find a harmonious balance between online access to knowledge and face-to-face interaction. The balance ... will vary for every course [based upon the] instructional goals, student characteristics, instructor background, and online resources ... No two courses will be exactly the same [The goal] is to ensure that the blend involves the strengths of each type of learning environment and none of the weaknesses. (p. 228)

Citing three typical reasons for using a blended approach--more effective pedagogy, increased convenience and access, and increased cost ef-

fectiveness (Graham, Allen, & Ure, 2005)--it is easy to see why progressively more institutions are implementing such a format.

Further blurring the line between traditional and distance learning, the blending and convergence of technologies continue to increase the fluidity and ubiquity of education. Today with the tap or swipe of a finger, learners have virtually instant access to a plethora of information via the Internet, which Klinger and Coffman (2011) claim provides a transformational highway for distance learning. Given the mobility of netbooks, smartphones, and tablets, geographical proximity becomes a non-issue as learners are no longer bound by context and can learn anytime and anywhere with social media producing an ecosystem ripe for interaction. Through time, the mobility of distance learning has remained constant; however, the delivery format has opened virtual horizons far beyond traveling from village to village. Foreshadowing this technological explosion, futurists McCain and Jukes (2001) projected that "at a certain point, the boundaries between reality and virtual reality will collapse because of the increased sophistication and transparency of these powerful, fused technologies" (p. 60). With the introduction of such things as cloud computing, adaptive educational hypermedia, interactive optical sensory technology, and three-dimensional virtual and immersive learning environments, that day has already arrived. So, where do we go from here?

FUTURE RESEARCH DIRECTIONS

The future of distance learning will continue to reflect the exponentially growing technological advances. In an April 1965 *Electronics* magazine article, Gordon Moore, cofounder and Chairman Emeritus of Intel, the world's largest silicon chip manufacturing company, predicted that the number of transistors on an integrated circuit would double every 12-18 months. While in 1975, he amended this timeframe to roughly every 24 months, it has been accurate in excess of one half century. At

present, what has become known as Moore's Law shows no signs of slowing down; although, in a 2000 interview with *Time* magazine, Moore did acknowledge that at some point in the next two or three generations, this growth would perhaps slow down to doubling every five years.

Exploring other technological alternatives, since the late 1950s some "wondered whether a better approach to miniaturization might be to 'grow' single molecules that functioned as electronic circuits or components. [They believed] that such molecules might be faster and smaller [than the traditional silicon chip] ... and they might also be easier to make" (Kelly & Mody, 2015, 8). While this movement has waxed and waned over the years, in 2010, researchers from Japan and Michigan Technological University claimed they had succeeded in building a molecular computer that could "replicate the inner mechanisms of the human brain, repairing itself and mimicking the massive parallelism that allows our brains to process information like no silicon-based computer can" (Borghino, 2010, 1). Four years later, Sebastian (2014) unveiled that IBM scientists had created "what some have claimed is the most advanced neuromorphic (brain-like) computer chip to date" (1).

Even with such remarkable advancements, we are quite a distance from reaching the apex of technological discovery, which history reveals directly affects distance teaching and learning. In reflection, these new discoveries "may very well prove the long-term solution to validate Moore's law well into the next century" (Borghino, 2010, p.2).

CONCLUSION

Whether one views distance as geographical or pedagogical, as the above suggests, the technological explosion of the 21st century provides unprecedented opportunity to render distance virtually irrelevant. If Moore's law holds fast and technology continues to double every 18-

24 months, the exponential growth of distance learning will continue to catapult educators into uncharted territory. The question remains, will there still be a place for the staunch traditionalist or will education continue to be so radically transformed that traditional education is hardly recognizable. Regardless of the medium (print, audio, video, or digital), method (one-to-many, one-to-one, or many-to-many), format (passive or interactive), or even device (PC, smartphone, tablet, or watch), interactive learning continues to remain at the core of the distance educational process.

Negroponte (1995) poignantly prophesied, "Distance means less and less in the digital world. In fact, an Internet user is utterly oblivious to it. On the Internet, distance often seems to function in reverse" (p. 178). Supporting this prediction, science and technology have triggered physical distance to become truly irrelevant. So, what does the future hold? Where *will* distance learning be in 5, 10, or even 20 years from now? Some would postulate that we are limited only by our imaginations.

REFERENCES

- Baath, J. A. (1980). *Postal two-way communication in correspondence education: An empirical investigation*. Malmö, Sweden: LiberHermods.
- Barnes, S., & Greller, L. M. (1994). Computer-mediated communication in the organization. *Communication Education*, 43(2), 129–142. doi:10.1080/03634529409378970
- Beetham, H., & Sharpe, R. (Eds.). (2007). *Rethinking pedagogy for a digital age: Designing and delivering e-learning*. London: Routledge.
- Benoit, P. J., Benoit, W. L., Milyo, J., & Hansen, G. J. (2006). *The effects of traditional versus web-assisted instruction on learning and satisfaction*. Columbia, MO: University of Missouri Graduate School.

- Bocarnea, M. C., Grooms, L. D., & Reid-Martinez, K. (2006). Technological and pedagogical considerations in online learning. In A. Schorr & S. Seltsmann (Eds.), *Changing Media Markets in Europe and Abroad: New Ways of Handling Information and Entertainment Content* (pp. 379-392). Lengerich: Pabst Science Publishers.
- Borghino, D. (2010, May 10). Scientists create organic 'molecular computer'. *Gizmag* [New and Emerging Technology News]. Retrieved from <http://www.gizmag.com/organic-molecular-computer/15041/>
- Cookson, P. S. (1989). Research on learners and learning in distance education: A review. *American Journal of Distance Education*, 3(2), 22-34. doi:10.1080/08923648909526661
- Daniel, J. S., & Marquis, C. (1979). Interaction and independence: Getting the mixture right. *Teaching at a Distance*, 14, 29-44.
- Dewal, O. S. (1988). Pedagogical issues - distance education. *Prospects*, 18(1), 63-73. doi:10.1007/BF02192959
- Gould, S. B. (1972). Prologue: Prospects for non-traditional study. In S. B. Gould & K. P. Cross (Eds.), *Explorations in non-traditional study* (pp. 1-12). San Francisco: Jossey-Bass.
- Graham, C. R., Allen, S., & Ure, D. (2005). Benefits and challenges of blended learning environments. In M. Khosrow (Ed.), *Encyclopedia of information science and technology* (pp. 253-259). Hershey, PA: Information Science Reference. doi:10.4018/978-1-59140-553-5.ch047
- Grooms, L. D. (2000). Interaction in the computer-mediated adult distance learning environment: Leadership development through online education. *Dissertation Abstracts International*, 61(12), 4692A.
- Grooms, L. D. (2003). Computer-mediated communication: A vehicle for learning. *International Review of Research in Open and Distance Learning*, 4(2). doi:10.19173/irrodl.v4i2.148
- Grooms, L. D., & Reid-Martinez, K. (2011). Sustainable leadership development: A conceptual model of a cross-cultural blending learning program. *International Journal of Leadership Studies*, 6(3), 412-429.
- Grooms, L. D., & Reid-Martinez, K. (2012, November). *Building a Community of Leaders through Intrapersonal, Interpersonal, and Intercultural Communication*. Presented at the annual convention of the National Communication Association, Orlando, FL.
- Grooms, L. D., & Reid-Martinez, K. (2013, November). *Cross-Cultural Learning for Leadership Resilience and Sustainability*. Presented at the 15th annual International Leadership Association Conference, Montreal, Canada.
- Harasim, L. (1989). On-line education: A new domain. In R. Mason & A. Kaye (Eds.), *Mindweave: Communication, computers and distance education* (pp. 50-62). New York: Pergamon Press.
- Harasim, L. M. (1990). Online education: An environment for collaboration and intellectual amplification. In L. M. Harasim (Ed.), *Online education: Perspectives on a new environment* (pp. 39-64). New York: Praeger.
- Harasim, L. M. (1993). Networked: Networks as social space. In L. M. Harasim (Ed.), *Global networks: Computers and international communication* (pp. 15-34). Cambridge, MA: The MIT Press.
- Hiltz, S. R. (1990). Evaluating the virtual classroom. In L. M. Harasim (Ed.), *Online education: Perspectives on a new environment* (pp. 133-169). New York: Praeger.
- Hiltz, S. R., & Johnson, K. (1990). User satisfaction with computer-mediated systems. *Management Science*, 36(6), 739-764. doi:10.1287/mnsc.36.6.739
- Holmberg, B. (1974). *Distance education: A short handbook*. Malmö, Sweden: Hermods.

- Holmberg, B. (1977). *Distance education: A survey and bibliography*. London: Kogan Page.
- Holmberg, B. (1981). *Status and trends of distance education*. London: Kogan Page.
- Holmberg, B. (1982). Distance study at the post-graduate level: Graduate study at a distance requires greater attention to communication with the student. In J. S. Daniel, M. A. Stroud, & J. R. Thompson (Eds.), *Learning at a distance: A world perspective* (pp. 258–260). Edmonton, AB, Canada: Athabasca University/International Council for Correspondence Education.
- Holmberg, B. (1986). *Growth and structure of distance education*. Wolfeboro, NH: Croom Helm.
- Kapitzke, C. (2000). The sociability and spatiality of online pedagogy and collaborative learning in an educational media and technology course. *Journal of Educational Technology & Society*, 3, 344–441.
- Kaye, A. (1981). Media, materials and learning methods. In A. Kaye & G. Rumble (Eds.), *Distance teaching for higher and adult education* (pp. 48–69). London: Croom Helm.
- Kaye, A. (1982). Using the media for adult basic education. In A. Kaye & K. Harry (Eds.), *Using the media for adult basic education* (pp. 9–29). London: Croom Helm.
- Kaye, A. (1988). Distance education: The state of the art. *Prospects*, 18(1), 43–54. doi:10.1007/BF02192957
- Kaye, A. (1989). Computer-mediated communication and distance education. In R. Mason & A. Kaye (Eds.), *Mindweave: Communication, computers and distance education* (pp. 3–21). New York: Pergamon Press.
- Keegan, D. (1989). Problems in defining the field of distance education. In M. G. Moore & G. C. Clark (Eds.), *Readings in principles of distance education* (pp. 8–15). University Park, PA: The Pennsylvania State University.
- Keegan, D. J. (1980). On defining distance education. *Distance Education*, 1(1), 13–36. doi:10.1080/0158791800010102
- Kelly, K. F., & Mody, C. C. M. (2015, September 25). Whatever happened to the molecular computer. *IEEE Spectrum*. Retrieved from <http://spectrum.ieee.org/biomedical/devices/whatever-happened-to-the-molecular-computer>
- Klinger, M. B., & Coffman, T. L. (2011). Emphasizing diversity through 3D multi-user virtual worlds. In G. Kurubacak & T. V. Yuzer (Eds.), *Handbook of research on transformative online education and liberation: Models for social equality* (pp. 86–106). Hershey, PA: Information Science Reference. doi:10.4018/978-1-60960-046-4.ch005
- McCain, T., & Jukes, I. (2001). *Windows in the future: Education in the age of technology*. Thousand Oaks, CA: Corwin Press.
- McIsaac, M. S., & Gunawardena, C. N. (1996). Distance education. In D. H. Jonassen (Ed.), *Handbook of research for educational communications and technology* (pp. 403–437). New York: Simon & Schuster Macmillan.
- Moore, G. (2000, June 19). Our Technology: Gordon Moore Q & A. *Time*, 155(25), 99. Retrieved from <http://www.time.com/time/magazine/article/0,9171,997241,00.html> PMID:11765518
- Moore, G. E. (1965, April 19). Cramming more components onto integrated circuits. *Electronics*, 38(8), 114–117.
- Moore, M. (1983). The individual adult learner. In M. Tight (Ed.), *Adult learning and education* (pp. 153–168). London: Croom Helm.
- Moore, M. G. (1973). Toward a theory of independent learning and teaching. *The Journal of Higher Education*, 44(9), 661–679. doi:10.2307/1980599
- Moore, M. G. (1980). Independent study. In R. D. Boyd & J. W. Apps et al. (Eds.), *Redefining the discipline of adult education* (pp. 16–31). San Francisco: Jossey-Bass.

- Moore, M. G. (1989a). Editorial: Three types of interaction. *American Journal of Distance Education*, 3(2), 1–7. doi:10.1080/08923648909526659
- Moore, M. G. (1989b, May). *Effects of distance learning: A summary of the literature*. Paper presented for Congress of the United States Office of Technology Assessment, Washington, DC.
- Moore, M. G. (1990). Correspondence study. In M. W. Galbraith (Ed.), *Adult learning methods: A guide for effective instruction* (pp. 345–365). Malabar, FL: Robert E. Krieger Publishing.
- Moore, M. G., & Kearsley, G. (1996). *Distance education: A systems view*. Boston: Wadsworth Publishing.
- Negroponte, N. (1995). *Being digital*. New York: Alfred A. Knopf.
- Ohler, J. (1991). Why distance education? *The Annals of the American Academy of Political and Social Science*, 514(1), 22–34. doi:10.1177/0002716291514001003
- Osguthorpe, R. T., & Graham, C. R. (2003). Blended learning environments: Definitions and directions. *The Quarterly Review of Distance Education*, 4(3), 227–233.
- Pittman, V. (1997). Distance education exchange. *The Journal of Continuing Higher Education*, 45(2), 42–43. doi:10.1080/07377366.1997.10400321
- Robinson, B. (1981). Support for student learning. In A. Kaye & G. Rumble (Eds.), *Distance teaching for higher and adult education* (pp. 141–161). London: Croom Helm.
- Rourke, A., & Coleman, K. (Eds.). (2011). *Pedagogy leads technology: Online learning and teaching in higher education: New technologies, new pedagogies*. Champaign, IL: Common Ground Publishing.
- Rumble, G. (1986). *The planning and management of distance education*. New York: St. Martin's Press.
- Rumble, G., & Keegan, D. (1982). Introduction. In G. Rumble & K. Harry (Eds.), *The distance teaching universities* (pp. 9–14). London: Croom Helm.
- Russell, T. L. (1999). *The no significant difference phenomenon as reported in 355 research reports, summaries and papers: A comparative research annotated bibliography on technology for distance education*. Raleigh, NC: North Carolina State University.
- Saba, F. (1998). Is distance education comparable to “traditional education”? *Distance Education Report*, 2(5), 5.
- Sebastian, A. (2014, August 7). IBM cracks open a new era of computing with brain-like chip: 4096 cores, 1 million neurons, 5.4 billion transistors. *Extreme Tech: Latest Technology News*. Retrieved from <http://www.extremetech.com/extreme/187612-ibm-cracks-open-a-new-era-of-computing-with-brain-like-chip-4096-cores-1-million-neurons-5-4-billion-transistors>
- Sewart, D. (1981). Distance teaching: A contradiction in terms? *Teaching at a Distance*, 19, 6–18.
- Shale, D. (1990). Toward a reconceptualization of distance education. In M. G. Moore, P. Cookson, J. Donaldson, & B. A. Quigley (Eds.), *Contemporary issues in American distance education* (pp. 333–343). New York: Pergamon Press.
- Shale, D., & Garrison, D. R. (1990). Education and communication. In D. R. Garrison & D. Shale (Eds.), *Education at a distance: From issues to practice* (pp. 23–29). Malabar, FL: Robert E. Krieger Publishing Company.
- Simonson, M., Schlosser, C., & Hanson, D. (1999). Theory and distance education: A new discussion. *American Journal of Distance Education*, 13(1), 60–75. doi:10.1080/08923649909527014
- Stadtlander, L. M. (1998). Virtual instruction: Teaching an online graduate seminar. *Teaching of Psychology*, 25(2), 146–148. doi:10.1207/s15328023top2502_20

Szecszy, E. M. (2011). Building knowledge without borders: Using ICT to develop a binational education research community. In G. Kurubacak & T. V. Yuzer (Eds.), *Handbook of research on transformative online education and liberation: Models for social equality* (pp. 67–85). Hershey, PA: Information Science Reference. doi:10.4018/978-1-60960-046-4.ch004

Wedemeyer, C. A. (1971). Independent study. In L. C. Deighton (Ed.), *The encyclopedia of education* (pp. 548–557). New York: Macmillan Company & The Free Press.

ADDITIONAL READING

Bandyopadhyay, A., Pati, R., Sahu, S., Peper, F., & Fujita, D. (2010). Massively parallel computing on an organic molecular layer. *Nature Physics*, 6(5), 369–375. doi:10.1038/nphys1636

Bhagat, K. K., Wu, L. Y., & Chang, C.-Y. (2016). Development and validation of the perception of students towards online learning (POSTOL). *Journal of Educational Technology & Society*, 19(1), 350–359.

Bozkurt, A., Akgun-Ozbek, E., Yilmazel, S., Erdogdu, E., Ucar, H., Guler, E., & Aydin, C. H. et al. (2015). Trends in distance education research: A content analysis of journals 2009–2013. *International Review of Research in Open and Distributed Learning*, 16(1), 330–363. doi:10.19173/irrodl.v16i1.1953

Burge, E., Gibson, C. C., & Gibson, T. (Eds.). (2011). *Flexible pedagogy, flexible practice: Notes from the trenches of distance education*. Edmonton, Alberta: Athabasca University.

Ceruzzi, P. E. (2005, July). Moores law and technological determinism: Reflections on the history of technology. *Technology and Culture*, 46(3), 584–593. Retrieved from <http://www.jstor.org/stable/40060905> doi:10.1353/tech.2005.0116

Cheney, A., & Sanders, R. L. (Eds.). (2011). *Teaching and learning in 3D immersive worlds: Pedagogical models and constructivist approaches*. Hershey, PA: Information Science Reference. doi:10.4018/978-1-60960-517-9

Gedik, N., Kiraz, E., & Ozden, M. Y. (2013). Design of a blended learning environment: Considerations and implementation issues. *Australasian Journal of Educational Technology*, 29(1), 1–19. doi:10.14742/ajet.6

Kurubacak, G., & Yuzer, T. V. (Eds.). (2011). *Handbook of research on transformative online education and liberation: Models for social equality*. Hershey, PA: Information Science Reference. doi:10.4018/978-1-60960-046-4

Lê, T., & Lê, Q. (Eds.). (2012). *Technologies for enhancing pedagogy, engagement and empowerment in education: Creating learning-friendly environments*. Hershey, PA: Information Science Reference. doi:10.4018/978-1-61350-074-3

Lister, M. (2014). Trends in the design of e-learning and online learning. *Journal of Online Learning and Teaching*, 10(4), 671–680.

Liu, H.-C., & Yen, J.-R. (2014). Effects of distance learning on learning effectiveness. *Eurasia Journal of Mathematics, Science, & Technology Education*, 10(6), 575–580. doi:10.12973/Eurasia.2014.1218a

Logofătu, B., & Vișan, A. (2015). New trends in the educational area: Case study regarding the usability of google apps tools within the department for distance learning. *Elearning & Software for Education*, 2, 526–531. doi: 026X1517110.12753/2066

Somyürek, S. (2015). The new trends on adaptive educational hypermedia systems. *International Review of Research in Open and Distributed Learning*, 16(1), 221–241. doi:10.19173/irrodl.v16i1.1946

KEY TERMS AND DEFINITIONS

Cloud Computing: Technologies that facilitate the sharing of digital files over the Internet.

Correspondence Learning: A form of distance learning using dispatched or one-way communication.

E-Learning: Learning that occurs electronically.

Equivalency: Distance learning that possesses equality with learning experienced in the face-to-face venue.

Face-to-Face Learning: Learning that is time-place dependent.

mLearning: Learning that occurs through the use of mobile technology.

Optical Sensory Technology: Technology that provides the ability to track user input within an augmented or virtual reality.

Three-Dimensional Virtual Learning: A computer-based simulated environment.

Time-Place Dependent: Education that transpires in the same location at the same time.

Time-Place Independent: Learning that does not rely on geographical proximity or time.

Traditional Study: Face-to-face learning.

E

Do Usability Design Features of a Mobile Game Influence Learning?

Rex Perez Bringula

University of the East, Philippines

Edison Cabrera

University of the East, Philippines

Princess B. Calmerin

University of the East, Philippines

Eduardo A. Lao

University of the East, Philippines

Christian Gerard Sembrano

University of the East, Philippines

Angelita D. Guia

University of the East, Philippines

Joan P. Lazaro

University of the East, Philippines

Alexis John M. Rubio

University of the East, Philippines

Annaliza E. Catacutan

National University, Philippines

Marilou N. Jamis

National University, Philippines

Lalaine P. Abad

Department of Education, Philippines

INTRODUCTION

Mobile learning allows learners to construct their own learning experience (Bandalaria, 2007). This form of learning empowers students to develop their own skills and knowledge (Sharples, Taylor & Vavoula, 2007). Students, through the use of mobile learning, may create opportunities to learn anytime and anywhere (Martin & Ertzberger,

2013; Sanchez, Mendoza, & Salinas, 2009) and can connect that learning experience to real life situations. Recently, games have been integrated in mobile platforms. Educational technology developers combined the entertaining components of games with educational contents in order to develop games for pedagogical purposes.

However, the existing threads of discussion on serious game usability do not provide evidence as

to which of the usability features of a mobile game can influence student learning. This chapter aims to address this issue. A serious game in disaster response was developed and utilized by first-year students. The goal of this paper is to find answers to the following questions:

1. What is the knowledge of the disaster response participants before and after playing the game?
2. What is the learning gain of the students after playing the game?
3. What are the perceptions of the participants in the usability of the game in terms of content, ease of use, usefulness, and aesthetics?
4. Is there a significant difference between the knowledge of students on disaster response before and after playing the game?
5. Do the usability factors, singly or in combination, influence the knowledge of the participants in natural disaster response?

BACKGROUND

Mobile games can be utilized to channel information and learning. Games with educational content may engage people to play and learn at the same time (Muratet, Torguet, Jessel, & Viallet, 2009). Studies have documented the positive results of mobile games as educational tools. Batson & Feinberg (2006) showed that students who used educational games had a positive learning experience. Mobile educational games also improve the motivation of students to learn and to have positive attitudes towards learning (Tuzun, Yilmaz-Soylu, Karakus, Inal, & Kizilkaya, 2009; Salter, Pittaway, Swabey, Capstick, & Douglas, 2012). Chow, Woodford & Maes (2011) also revealed that the use of mobile educational games improved students' retention, critical thinking skills, and understanding of the content of an introductory statistics course (Chow et al., 2011). Moreover, it was proven that these educational materials can improve learners' problem solving

skills and can promote collaboration among them (Sanchez & Olivares 2011). Recently, Connolly, Boyle, MacArthur, Hainey, & Boyle (2012) summarized the results of 129 studies that investigated the impacts and outcomes of serious games on learning and students' classroom engagement. The researchers consistently found that playing serious games were associated to knowledge acquisition, content understanding, and affective and motivational outcomes.

Learning of students could also be measured in terms of learning gain. Learning gain is the measurement of performance in a test as indicated by the percentage points a student can gain from the first/previous test to the second/recent test (Colt, Davoudi, Murgu, & Rohani, 2011; Steif & Dollár, 2009). Rodrigo et al. (2013) utilized this measure to show that learning gains could be influenced by the ability of the students to solve problems. Bringula, Alvarez, Evangelista, & So (in press) used this measure to determine the impact of a mobile learning software on the mathematics performance of students. They found out that after using the mobile learning software, students increased their mathematics performance by 41%.

Various design models and guidelines were proposed in an attempt to guide educational game developers to develop serious games (e.g., Billi et al., 2010; Bringula, Alcid, Bandril, De Guzman, & Lopez, 2014). The goal of these models and guidelines is to balance the entertaining and educational components of the mobile game (Kreutzer, Marks, & Bowers, 2015). Usability of mobile games was explored in an attempt to warrant that the end product will be functional and playable (Olsen, Procci, & Bowers, 2011; Warren, Jons, & Lin, 2011). In an educational point of view, the purpose of the exploration of serious game usability is to ensure knowledge transfer (Kreutzer, Marks, & Bowers, 2015).

There is no universal definition of usability (Sindhuja & Dastidar, 2009). This is because usability is dependent on the content and nature of the systems being investigated. In mobile games

designed for educational purposes (that is, serious games), the usual metrics of usability—efficiency, effectiveness, and user satisfaction—may not be the appropriate criteria for a usable mobile game (Thomas, Schott, & Kambouri, 2003). Researchers suggested that educational game developers must balance learning component of games while maintaining the fun in the games (Thomas et al., 2003; Ibrahim, Vela, Rodriguez, Sanchez, & Zea, 2012). The content of the game can focus on the knowledge or skills that are intended to be inculcated or transferred (Thomas et al., 2003). Derakhshan (2009) showed that content was an important feature of a mobile learning environment. Kutluk and Gülmez (2014) disclosed that good organization and easy to navigate content were desirable characteristics of mobile learning environment. Shchiglik, Barnes, Scornavacca (2016) showed that the game content quality should be accurate, relevant, and easy to understand. Assuring the quality of game content can attract and motivate students to learn (Alqahtani & Mohammad, 2015).

The game must educate the users while they have fun playing the game. The game must be simple and enjoyable so that it captivates the imagination of the player (Ibrahim et al., 2012). Davis (1989) defined ease of use as a system quality factor which is related to user-friendliness of application design. In other words, ease of use is equated to the effort exerted in using a technology (Bringula, 2016). Tan, Ooi, Sim, & Phusavat (2012) showed that ease of use can affect intention to use mobile learning. Hence, ease of use influences the attitudes and behaviour of people towards the technology (Abu-al-Aish & Love, 2013; Alqahtani & Mohammad, 2015). In the context of mobile games, developers must ensure that the independent components of a game are easy to use (Olsen et al., 2011). In turn, learners may continue to use the game (Alqahtani & Mohammad, 2015). In a recent study, Shchiglik et al. (2016) showed that ease of use of games can be measured in terms of the following — easy to use, easy to learn and

operate, users know what to do, easy to navigate, easy to find things in the game.

The usefulness of the game is anchored to its game objective. The clarity of game objective is achieved when a player can explain the game (Bringula et al., 2014; Ibrahim et al., 2012). If learners can state the purpose of the game, they have an idea of its usefulness. The learners' perception of the usefulness of the game can influence their intention to use the game (Tan et al., 2012). Therefore, perceived usefulness is the expectation of users towards a new technology as well as the view that it would bring satisfaction and enhanced performance. Chittaro & Sioni (2015) provided evidence that a game in disaster response must contain useful recommendations in terms of safety, evacuation, and injury reduction.

Lastly, aesthetics is the most discussed factor of usability. In game design, aesthetics is the overall quality of visual design (Bringula et al., 2014). Game objects, text layout, buttons, graphics, and colors constitute the visual design of the game. Gamers are more particular in this aspect than the rest of other game design considerations. This is because the environment in mobile devices is different from the personal computers. In another study, Alqahtani & Mohammad (2015) included the adaptability of the application to the screen size of mobile devices, suitability of the design in landscape and portrait modes, and contentment with the design of the application as items of aesthetics. The findings of Shchiglik et al. (2016) are consistent to that of Alqahtani and Mohammad (2015).

However, it is still unclear which usability design factors may influence student learning. Current studies focus on the usability of serious games. Absent in the literature is the evidence as to whether these game usability features could influence user learning. The main focus of this article is to determine if game usability design features could influence learning. It is hypothesized that usability factors, singly or in combination, do not influence the knowledge of the participants in natural disaster response

THE EXPERIMENT AND MEASUREMENT OF LEARNING

A one-group pretest-posttest experimental design (Figure 1) was conducted by the researchers to address the main focus of the chapter. The study was conducted at the National University, Manila. Ninety-eight participants (34 female, 64 male, average age is 17.7 years old) have utilized a mobile android-based game named H.E.A.T. on disaster response (Figure 2). The participants used the game during their laboratory sessions for three consecutive days. The laboratory session lasted for one hour. This is the intervention period (denoted by *X*). A day before the intervention period, a pretest (denoted by *O*) was conducted to determine the prior knowledge of the participants on how to respond during a disaster. A day after the intervention period, a posttest (denoted by *O*) was administered. The experiment ran for five consecutive days. Each test contained 20 items.

THE USABILITY GAME DESIGN CRITERIA

The usability game design criteria were pretested to ensure its reliability and validity. Students from the University of the East assessed the software. It contained four dimensions which consisted of content, ease of use, usefulness, and aesthetics. Content was composed of three items. The first item refers to the completeness of the game. It measured the scope of the game in terms of different forms of natural calamities. The reliability of the game content is the third item of this dimension.

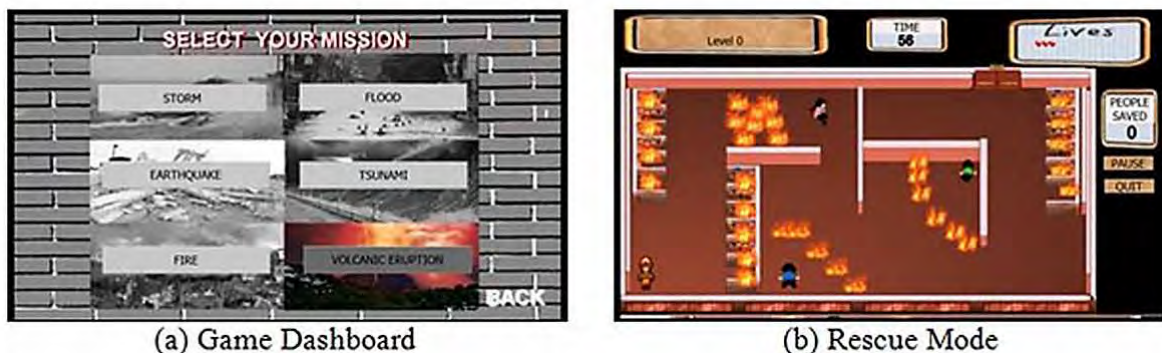
Ease of use was composed of six items. It evaluated if it may be easy to use the game controls and determined if all controls are functional. Learners must also feel that the game is easy to play. They must not be confused or lost when playing the game. They can also evaluate the game in terms of how easy it is to move from one game screen to another. The last item of this criterion asked the learners to rate the overall ease of using the game.

Usefulness, which is composed of four items, intended to measure the perceived educational value of the game in terms of disaster response. It

Figure 1. One-group pretest-posttest design

<i>O</i>	<i>X</i>	<i>O</i>
Knowledge on disaster response before using the game (pretest)	Use of the mobile game on disaster response (three days)	Knowledge on disaster response after using the game (posttest)

Figure 2. The game utilized in the study



determined the perceptions of the learners in terms of how much they learned from the game. Moreover, it contains an item that examined whether the game gave useful information and tips. The last criterion was aesthetics. It consisted of seven items relating to attractiveness of screen layout, comfort when playing games, appropriateness of colors, legibility of font sizes and font styles, and overall attractiveness of the game.

RESULTS

It was found that the pretest score on storm had the lowest percentage at 50% while pretest score in earthquake had the highest percentage at 77%. After the intervention period, all test scores were higher. Posttest scores in earthquake, flood, and tsunami are all 81%. The overall mean also went up from 59% to 74%. Students had the most learning gain (l.g.) in terms of knowledge in flood (l.g. = 60%). Overall, the learning gain is 35%.

The same sample *t*-test revealed that there is a significant difference on the posttest and pretest scores in fire ($d = 0.66$, $t(97) = 4.55$),

flood ($d = 0.83$, $t(97) = 8.40$), tsunami ($d = 0.60$, $t(97) = 5.20$), and volcanic eruption ($d = 0.56$, $t(97) = 4.80$). All results are significant at 0.05 level of confidence. Hence, the results are unlikely to have arisen from sampling error. On the other hand, the differences on the scores are not significant in storm ($d = 0.06$, $t(97) = 0.46$, $p = 0.644$) and earthquake ($d = 0.16$, $t(97) = 1.50$, $p = 0.138$).

Usefulness had the highest mean rating of 4.18 (s.d. = 0.75), while aesthetics had the lowest mean rating of 3.05 (s.d. = 0.86). Respondents agreed that the content of the game (mean = 3.92, s.d. = 0.69) is acceptable and useful (mean = 4.18, s.d. = 0.75). The standard deviations of all criteria showed that the responses are not dispersed.

The responses in the open-ended questions were analyzed by taking keywords. The keywords were then grouped to determine the themes. The results are shown in Table 1. It was disclosed that the five themes emerged. Three out of the five themes are covered in the current study. Respondents commented that the game is easy to use and play. Also, the controls of the game need to be improved. In term of aesthetics, 27 respondents

Table 1. Results of open-ended questions

Ease of Use • easy to use and play (2 times mentioned) • controls need to be improved (4 times mentioned)
Usefulness • respondents learned a lot of lessons about disasters (6 times mentioned) • helped provide knowledge on what to do during calamities (8 times mentioned) • gave tips on what to do during calamity • very informative (6 times mentioned) • good game for students (2 times mentioned) • gave awareness about disasters • useful (2 times mentioned)
Aesthetics • graphics needs to be improved (27 times mentioned)
Fun • very enjoyable to play (2 times mentioned) • fun (3 times mentioned) • happy • boring
Effectiveness • tasks proved a bit time-limited • walking of character is slow (14 times mentioned) • game occupied a large memory (3 times mentioned)

advised that the graphics of the game need to be upgraded. Twenty-six respondents agreed that the game is informative and useful. The fun component of the game is included as six participants expressed enjoyment with the game while one participant expressed otherwise. Contrary to the study of Thomas et al. (2003), effectiveness is found to be a component of serious game usability. This factor includes the time to complete the game and the memory space used in the game.

The result of multiple regression of learning gain on game usability criteria showed that the *p*-values of Content ($b = 0.01, p = 0.93$), Ease of Use ($\beta = -0.02, p = 0.87$), and Usefulness ($\beta = 0.08, p = 0.48$) are all greater than 0.05 level of confidence. Thus, aesthetics ($\beta = 0.227$), with an associated *p*-value of 0.025, was the only predictor of learning gain. Aesthetics accounts for 4% (Adj. R^2) in the variability of learning gain. The result is unlikely to have arisen from sampling error ($F(1,96) = 5.20, p < 0.05$).

DISCUSSION

The result of the experiment shows that the students have a considerable amount of knowledge on how to respond to disasters. It is interesting to note that 77% of the questions relating to earthquake response were answered correctly. This is explained by the fact that earthquake drills were conducted at school, local, and national levels at the time the study was conducted. This is in relation to the earthquake preparedness program of the government. This explains why the students have the highest percentage on earthquake items.

Meanwhile, students had the lowest percentage score on typhoon preparedness. On this item, students only had a passing mark on the pretest. This finding is unexpected considering that the Philippines is frequently devastated by typhoons. According to the Philippine Statistics Authority (2014), an average of nine typhoons visited the Philippines from 1993 to 2013. This explains the reason behind this unexpected result. Filipinos

are too attuned to typhoons to the extent that they do not mind preparing for them. In fact, it was reported that the strength of Typhoon Haiyan was underestimated by the locals, resulting to poor planning (Jibiki, Kure, Kuri, & Ono, 2016).

The result signals a need to conduct a study on the attitudes of Filipinos toward typhoons. This is not addressed in the current study. A game may be developed to assess the attitudes of Filipinos toward typhoons. The game may then respond depending on the assessed attitudes.

It is interesting to note that all test items increased following the intervention period. The test items for storm remained passing and surfaced as having the lowest score. Students had the highest learning on flood preparation. This is a good indication since flooding can be a consequence of other natural disasters, including typhoons. The overall learning gain of 35% means that students acquired 35 percentage points out of 100 percentage points that they could have gained from the pretest to posttest. This means that students acquired more than one-third of the possible points. Moreover, a learning gain of 30% is the defining minimum value at which the educational intervention could be regarded as effective (Hake, 1998; Prather, Rudolph, & Brissenden, 2009). In general, the game made it possible for students to know more about disaster preparedness.

After using the game, the posttest scores proved higher than the pretest scores. These indicate that students had positive learning gains. Further statistical tests confirmed that the differences in test items in fire, flood, tsunami, and volcanic eruption are significant. This shows that students learned to be more prepared in times of fire, flood, tsunami, and volcanic eruption. Furthermore, it also implies that the software can be utilized as an instructional material for disaster preparedness and risk reduction.

The evaluation of the usability design features of the game showed that students agreed that the information provided by the game is reliable. The students likewise favorably rated the software in terms of ease of use. According to the learners, it

is easy to learn how to play the game because all controls are functional. Students also perceived that they learned a lot from the game. The game helped them to understand natural disasters and how to respond appropriately. The result of open-ended questions showed that 26 respondents commented that the game is informative and helpful. This confirms the numerical rating on usefulness.

Meanwhile, they agreed to a lesser extent that the game is appealing. The responses of the 27 respondents on the open-ended question were in agreement that there is a need to improve the game's graphics. The nature of the participants may explain the findings. It is worth noting that the respondents of this study are information technology students who are knowledgeable in computer games. They may compare the game to other commercially-available games with better graphics and designs. Nonetheless, the evaluation of the respondents can serve as a basis to improve it in terms of screen layout, colors, and font sizes.

It was also revealed that six students enjoyed playing the game. They viewed the game as informative and helpful, as well as fun and enjoyable. The "fun" component of the game is not included in the evaluation. Future studies may include this factor in evaluating a game on disaster response. It is worth noting that students are also concerned with slow movement of the game character. The students recommended that the time to finish a task is not realistic. They were also concerned with the memory space it used when installed in a mobile device.

The result of the regression analysis revealed that it is justifiable to focus on the aesthetic components of the game. It was disclosed that aesthetics is the only predictor that could influence student learning in disaster response. Bringula & De Leon (2014) said that aesthetics is a "human" side design consideration. It is an "experiential" factor where "users have to use the software in order to judge the quality of the software" (Bringula & De Leon, 2014, p. 198). In short, this is the factor that users directly experience throughout the game.

Based on this finding, it is important that serious games developers consider the visual

content of the game in terms of its game objects, text layout, buttons, graphics, and colors. It is recommended that game objects be made clear. The text styles and sizes should be legible. Players must instinctively know the purpose of the buttons. The screen layout must fit the screen size of mobile devices. It is essential that colors blend the overall environment of the game. These aesthetic considerations can ensure the comfort of playing a serious game.

FUTURE RESEARCH DIRECTIONS

Aesthetics is found to be a significant predictor of learning in a mobile game. However, it accounts for only 4% of learning disaster response. In light of this finding, there are other usability design factors not considered in the study that might be investigated. The results of the experiment disclosed that fun and effectiveness were the two themes emerged from the open-ended questions. Future research may include these factors in the research instrument and quantitatively analyze their contributions in explaining the usability of the game. Informal interviews with students showed that the game can include more levels with different difficulties and surprises to capture the interest of the students. These design considerations will keep the students playing and will entice them to play the game again.

A study on different samples is also recommended. Children, who are among the most vulnerable sector of society, can be included as participants in future studies. The game can be utilized to teach them how to respond during natural disasters.

CONCLUSION

On the basis of these findings, the null hypothesis stating that the usability factors do not influence the knowledge of the participants in natural disaster response is partially rejected. It is also concluded

that the game was helpful in increasing the knowledge of the students to respond during fire, flood, tsunami, and volcanic eruption. It concluded that the game was not helpful during storm and earthquake situations. The game did not significantly contribute to student knowledge on responding to storms since they are already attuned to storms. They did not achieve a significant learning gain in earthquake test items because earthquake drills were conducted in local, institutional, and national levels prior to this study.

REFERENCES

- Abu-al-Aish, A., & Love, S. (2013). Factors influencing students acceptance of m-learning: An investigation in higher education. *The International Review of Research in Open and Distributed Learning*, 14(5), 82–107. doi:10.19173/irrodl.v14i5.1631
- Alqahtani, M., & Mohammad, H. (2015). Mobile applications' impact on student performance and satisfaction. *The Turkish Online Journal of Educational Technology*, 14(4), 102–112.
- Bandalaria, M. (2007). Impact of ICTs on open and distance learning in a developing country setting: The Philippine experience. *International Review of Research in Open and Distance Learning*, 8(1), 1–13. Retrieved from <http://www.irrodl.org/index.php/irrodl/article/view/334/793> doi:10.19173/irrodl.v8i1.334
- Batson, L., & Feinberg, S. (2006). Game designs that enhance motivation and learning for teenagers. *Electronic Journal for the Integration of Technology in Education*, 5, 34–43.
- Billi, M., Burzagli, L., Catarci, T., Santucci, G., Bertini, E., Gabbanini, F., & Palchetti, E. (2010). A unified methodology for the evaluation of accessibility and usability of mobile applications. *Universal Access in the Information Society*, 9(4), 337–356. doi:10.1007/s10209-009-0180-1
- Bringula, R. P. (2016). Factors affecting web portal information services usability: A canonical correlation analysis. *International Journal of Human-Computer Interaction*, 1–13. doi:10.1080/10447318.2016.1199180
- Bringula, R. P., Alcid, A. S., Bandril, L. B. P., De Guzman, A. E., & Lopez, L. J. C. (2014). Development of game design guidelines. In A. Turnip (Ed.), *2014 2nd International Conference on Technology, Informatics, Management, Engineering & Environment* (pp. 234–239). Danvers, MA: IEEE. doi:10.1109/TIME-E.2014.7011624
- Bringula, R. P., Alvarez, J. N., Evangelista, M., & So, R. (in press). Learner-interface interactions with a mobile-assisted learning in mathematics: Effects on and relationship between mathematics performance. *International Journal of Mobile and Blended Learning*.
- Bringula, R. P., & De Leon, A. N. (2014). The role of trust in web-based election system usability. In M. Indrawan-Santiago, M. Steinbauer, H.-Q. Nguyen, A. M. Tjoa, I. Khalil, & G. Anderst-Kotsis (Eds.), *iiWAS 2014: The 16th International Conference on Information Integration and Web-based Applications & Services* (pp. 196–199). Danvers, MA: ACM. doi:10.1145/2684200.2684317
- Chittaro, L., & Sioni, R. (2015). Serious games for Serious games for emergency preparedness: Evaluation of an interactive vs. a non-interactive simulation of a terror attack. *Computers in Human Behavior*, 50, 508–519. doi:10.1016/j.chb.2015.03.074
- Chow, A. F., Woodford, K. C., & Maes, J. (2011). Deal or no deal: Using games to improve student learning, retention and decision-making. *International Journal of Mathematical Education in Science and Technology*, 42(2), 259–264. doi:10.1080/0020739X.2010.519796

- Colt, H. G., Davoudi, M., Murgu, S., & Rohani, N. Z. (2011). Measuring learning gain during a one-day introductory bronchoscopy course. *Surgical Endoscopy*, 25(1), 207–216. doi:10.1007/s00464-010-1161-4 PMID:20585964
- Connolly, T. M., Boyle, E. A., MacArthur, E., Hainey, T., & Boyle, J. M. (2012). A systematic literature review of empirical evidence on computer games and serious games. *Computers & Education*, 59(2), 661–686. doi:10.1016/j.compedu.2012.03.004
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *Management Information Systems Quarterly*, 13(3), 319–340. doi:10.2307/249008
- Derakhshan, N. (2009). *Student and faculty perceptions of the features of mobile learning management systems in the context of higher education* (Unpublished doctoral dissertation). Oklahoma State University.
- Hake, R. (1998). Interactive engagement versus traditional methods: A six-thousand student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, 66(1), 64–74. doi:10.1119/1.18809
- Ibrahim, A., Vela, F. L. G., Rodriguez, P. P., Sanchez, J. L. G., & Zea, N. P. (2012). Playability guidelines for educational video games: A comprehensive and integrated literature review. *International Journal of Game-Based Learning*, 2(4), 18–40. doi:10.4018/ijgbl.2012100102
- Jibiki, Y., Kure, S., Kuri, M., & Ono, Y. (2016). Analysis of early warning systems: The case of super-typhoon Haiyan. *International Journal of Disaster Risk Reduction*, 15, 24–28. doi:10.1016/j.ijdrr.2015.12.002
- Kreutzer, C., Marks, M., & Bowers, C. (2015). A pedagogical approach to usability in serious games. In C. Stephanidis (Ed.), *HCI International 2015* (pp. 43–48). Springer International Publishing; doi:10.1007/978-3-319-21380-4_8
- Kutluk, F. A., & Gülmez, M. (2014). Research about mobile learning perspectives of university students who have accounting lessons. *Procedia: Social and Behavioral Sciences*, 116, 291–297. doi:10.1016/j.sbspro.2014.01.210
- Martin, F., & Ertzberger, J. (2013). Here and now mobile learning: An experimental study on the use of mobile technology. *Computers & Education*, 68, 76–85. doi:10.1016/j.compedu.2013.04.021
- Muratet, M., Torguet, P., Jessel, J.-P., & Viallet, F. (2009). Towards a serious game to help students learn computer programming. *International Journal of Computer Games Technology*. doi:10.1155/2009/470590
- Olsen, T., Procci, K., & Bowers, C. (2011). *Serious games usability testing: How to ensure proper usability, playability, and effectiveness*. Paper presented at First International Conference: DUXU 2011, Orlando, FL. doi:10.1007/978-3-642-21708-1_70
- Philippine Statistics Authority. (2014). *Tropical cyclone frequency*. Retrieved from <http://philfsis.psa.gov.ph/index.php/id/15/matrix/J30FSTII>
- Prather, E. E., Rudolph, A. L., & Brissenden, G. (2009). Teaching and learning astronomy in the 21st century. *Physics Today*, 62(10), 41–47. doi:10.1063/1.3248478
- Rodrigo, M. M. T., Ong, A., Bringula, R. P., Basa, R. S., Dela Cruz, C., & Matsuda, N. (2013). Impact of prior knowledge and teaching strategies on learning by teaching. *AIED Workshop on Simulated Learners: AIED 2013 Workshop Proceedings*, 4, 71–84. Retrieved from <http://ceur-ws.org/Vol-1009/0408.pdf>
- Salter, S., Pittaway, J., Swabey, K., Capstick, M., & Douglas, T. (2012). Using an online interactive game to enhance the learning outcomes for first year tertiary students. *Creative Education*, 3(06), 1–8. doi:10.4236/ce.2012.326114

Sanchez, J., Mendoza, C., & Salinas, A. (2009). Mobile serious games for collaborative problem solving. In B. K. Wiederhold & G. Riva (Eds.), *The Annual Review of Cybertherapy and Cybermedicine 2009* (vol. 144, pp. 193-197). Amsterdam: Studies in Health Technology and Informatics (SHTI), IOS Press.

Sanchez, J., & Olivares, R. (2011). Problem solving and collaboration using mobile serious games. *Computers & Education*, 57(3), 1943-1952. doi:10.1016/j.compedu.2011.04.012

Sharples, M., Taylor, J., & Vavoula, G. (2007). A theory of learning for the mobile age. In R. Andrews & C. Haythornthwaite (Eds.), *The Sage handbook of e-learning research* (pp. 221-247). London: Sage Publications. doi:10.4135/9781848607859.n10

Shchiglik, C., Barnes, S. J., & Scornavacca, E. (2016). The development of an instrument to measure mobile game quality. *Journal of Computer Information Systems*, 56(2), 97-105. doi:10.1080/08874417.2016.1117368

Sindhuja, P. N., & Dastidar, S. G. (2009). Impact of the factors influencing website usability on user satisfaction. *The IUP Journal of Management Research*, 8(12), 54-66.

Steif, P. S., & Dollár, A. (2009). Study of usage patterns and learning gains in a web-based interactive static course. *The Journal of Engineering Education*, 98(4), 321-333. doi:10.1002/j.2168-9830.2009.tb01030.x

Tan, W.-H. G., Ooi, K.-B., Sim, J.-J., & Phusavat, K. (2012). Determinants of mobile learning adoption: An empirical analysis. *Journal of Computer Information Systems*, 52(3), 82-91.

Thomas, S., Schott, G., & Kambouri, M. (2003). Designing for learning or designing for fun? Setting usability guidelines for mobile educational games. In J. Attewell & C. Savill-Smith (Eds.), *Learning with mobile devices: Research and development* (pp. 173-181). London: Learning and Skills Development Agency.

Tuzun, H., Yilmaz-Soylu, M., Karakus, T., Inal, Y., & Kizilkaya, G. (2009). The effects of computer games on primary school students achievement and motivation in geography learning. *Computers & Education*, 52(1), 68-77. doi:10.1016/j.compedu.2008.06.008

Warren, S., Jones, G., & Lin, L. (2011). Usability and play testing: The often missed assessment. In L. Annetta & S. C. Bronack (Eds.), *Serious Educational Game Assessment: Practical Methods and Models for Educational Games, Simulations and Virtual Worlds* (pp. 131-146). Rotterdam, The Netherlands: Sense Publishers. doi:10.1007/978-94-6091-329-7_8

ADDITIONAL READING

Attewell, J., & Savill-Smith, C. (2004). *Learning with mobile devices: Research and development*. London: Learning and Skills Development Agency.

Grimus, M., & Ebner, M. (2015). Learning and teaching with mobile devices: An approach in higher secondary education in Ghana. *International Journal of Mobile and Blended Learning*, 7(2), 17-32. doi:10.4018/ijmbl.2015040102

Ha, K.-M. (2016). Disasters can happen to anybody: The case of Korea. *Environmental Impact Assessment Review*, 57, 1-9. doi:10.1016/j.eiar.2015.11.002

Hersh, M., & Leporini, B. (2013). An overview of accessibility and usability of educational games. In C. Gonzalez (Ed.), *Student usability in educational software and games: Improving experiences* (pp. 1-40). Hershey, PA: IGI Global; doi:10.4018/978-1-4666-1987-6.ch001

Hoehle, H., Aljafari, R., & Venkatesh, V. (2016). Leveraging Microsofts mobile usability guidelines: Conceptualizing and developing scales for mobile application usability. *International Journal of Human-Computer Studies*, 89, 35-53. doi:10.1016/j.ijhcs.2016.02.001

Schmitz, B., Klemke, R., Walhout, J., & Specht, M. (2015). Attuning a mobile simulation game for school children using a design-based research approach. *Computers & Education*, 81, 35–48. doi:10.1016/j.compedu.2014.09.001

KEY TERMS AND DEFINITIONS

Aesthetics: A game attribute that relates to the overall visual design of a serious game.

Content: A game design feature that involves completeness and reliability of information provided by the game.

Disaster: An event that is caused by accidents or natural calamities that disrupts order and may result in damage of property or loss of life.

Ease of Use: A game design feature that measures the effort of the user in playing a game.

Mobile Game: A game that runs on any handheld portable device.

Serious Game: A type of game that is both entertaining and educational.

Usefulness: A game feature with a goal is to transfer knowledge or skills to its users.

Educational Technology and Intellectual Property

E**Lesley S. J. Farmer***California State University – Long Beach, USA*

INTRODUCTION

In today's digital world, leaders and other educators can manipulate a wide variety of information for authentic projects. In the process, everyone needs to acknowledge the idea creators and their intellectual property. As technologies have expanded, and their production has become more sophisticated, the legal regulations surrounding their use have become more complex. With the advent of interactive social media and increased resource sharing, as well as growth in distance learning opportunities, complying with the legal use of information technology can be daunting. In any case, leaders and other educators should be aware of the more important aspects of technology-related copyright laws and regulations. This chapter provides an overview of copyright law and fair use for educational research purposes.

LEGAL BACKGROUND

A central aspect of education is intellectual pursuit and the recognition of great minds. Yet teachers bemoan the rise in cheating, which technology facilitates. On their part, students have a more lax attitude about intellectual property. Particularly with media and crowdsourcing, which foster collaborate knowledge generation, identifying the originator of an idea can be difficult to ascertain. Furthermore, the informational content itself may be dynamic so that the authorship and copyright may change over time.

Although intellectual property is sometimes used interchangeably with copyright, the former is

a broader concept. Copyright protects creative and original ideas that are recorded in tangible form. Other U. S. intellectual property laws deal with trademarks, patents, trade secrets, and licenses.

Copyright laws seek solutions to give authors fair compensation for sharing their work. Begun as a way to give scientists and inventors lead time to prevent others from using their work without permission, copyright laws in the United States have become more far-ranging, both in terms of expanded formats as well as issues of authorship and access. According to current law, the copyright owner has the exclusive right, and can authorize others, to reproduce, distribute, display, publicly perform, and make derivative works based on the original work. The duration of the copyright term has lengthened over the years, starting from a length of 28 years (as established in 1790) to 70 years after the death of the author, according to the 1998 Act.

Publishing has also impacted copyright over time. Reporters increasingly get personal credit and remuneration for their contributions, which historically were considered solely work for hire. Publishers create copyright agreements to cover authorship rights based on format. Multimedia copyright laws can be very specific: restricting resizing or other image manipulation, stipulating the length of music or video that can be copied legitimately. Fortunately, education falls under the umbrella of Fair Use, so restrictions are loosened up a bit in order to support personal research.

The chief statute driving copyright law is the Copyright Act of 1976, which became effective in 1978. Several factors were included for the first time in this piece of legislation: a codification of

fair use, the right for an author to receive copyright for an unpublished work, and the divisibility of authors' rights. The Act includes definitions, delineates what is copyrightable, and describes copyright rights and limitations.

The Digital Millennium Copyright Act (DMCA) was added to the Act in 1998, largely to conform to international treaties (note that no international copyright law exists) that dealt with technological issues, particularly online material. DMCA limits database company liability, and addresses digital preservation.

Educators and leaders also need to know about the 2002 Technology, Education, and Copyright Harmonization (TEACH) Act, which impacts copyright usage in distance education or in cases where digital information is transmitted as a supplement face-to-face instruction. Displays and performances can be disseminated only for the period of the course and only to those students who are enrolled in the course. Likewise, if teachers copy an article for a face-to-face class, then they can link to the same article online, depending on the magazine database license agreement. A better solution is for the teacher to provide the citation, and ask the students to access the article themselves from the library's database collection. However, the teacher should *not* download the whole magazine issue just because it is technically possible; that action probably does not comply with copyright law.

Other laws exist to support copyright law such as anti-piracy. For instance, the 1997 No Electronic Theft Act expands criminal prosecution of copyright infringements to individuals who do not benefit commercially. The Family and Entertainment Copyright Act of 2005, updated and absorbed the 1997 Act about theft, and also addressed artists' rights.

Particularly with the advent of social media in which students might produce and disseminate information publicly, copyright applications are stricter. A few examples of technology-relevant copyright practices follow.

- Images should not be resized, cropped, or changed in context without explicit permission.
- Photos of recognizable people require written permission if they are to be broadcast.
- Scanning or digitizing work is considered reproducing it, and requires prior permission if shared publicly. Other rights may be covered by patent law, such as copying code.
- Music and video downloads can be very problematic. It is wisest to use pre-approved sites such as iTunes.
- Computer program appearance, graphics and sound may be protected from copying, depending on the company's licensing agreements.
- Slander and libel can occur on social networking sites such as MySpace and Facebook; students might not realize that they can be held legally responsible for their comments, and even arrested and prosecuted.
- Information about health and legal issues should include a disclaimer so the author is not held legally responsible in case the reader uses that information and experiences negative results.
- Each database aggregator and disseminator (such as video streaming) has a unique licensing agreement that covers copyright issues. A good rule of thumb is to apply the most conservative guidelines in order to avoid case-by-case decisions.

Further complicating matters, different countries have different copyright guidelines, so information accessed from around the world may be subject to conflicting laws. International aspects also impact copyright in several ways. At present, few copyright laws exist. How, then, are royalties handled when publishers are international or offer translations of works? They are done by international treaties or agreements, usually on a

case-by-case basis. Right now, Disney licenses some merchandise abroad, which products are not allowed in the U.S. because of copyright restrictions; foreign entities can sell them to U.S. buyers, who cannot buy the items stateside. Such inequities need to be addressed.

This situation is exacerbated in light of transnational social networking sites such as Facebook and Google, who are reluctant to comply with the most stringent regulations when other countries may have laxer laws. The International Copyright Act of 1891 was the first step in extending copyright to works of foreign nationals – of selected countries. The World Intellectual Property Organization (WIPO), established in 1967 under the auspices of UNESCO, serves as a global forum to address issues of intellectual property services, policy, information and cooperation. Currently, they are discussing the need for libraries to be able to provide all of their users with access to information, particularly to facilitate countries' economic development (IFLA, 2014).

Legal cases test the boundaries of existing copyright law. The U.S. Copyright Office's Fair Use Index (<http://www.copyright.gov/fair-use/>) provides searchable summaries of major fair use decisions. For instance, *Authors Guild, Inc. v. Google Inc.*, No. 13-4829-cv (2d Cir. Oct. 16, 2015) contested Google's to digitally duplicate books from library collections. The court defended Google's practice as covered under Fair Use because the process was transformative: the search function added value, and the materials were for libraries. Nor did the digitized clips jeopardize market value. Similarly, in *Authors Guild, Inc. vs. HathiTrust*, 755 F. 3d 87, 90-91 (2d Cir. 2014), the court upheld the practice of creating read-aloud versions of resources so they would be accessible for individuals with vision impairments; the decision was based on the Americans with Disabilities Act. On the other hand, when publishing houses sued Georgia State University for posting unlicensed works for students to access, even after using a fair use checklist;

the university was found at fault for not complying with licensing agreements as they impacted copyright (*Cambridge University Press v. Mark P. Becker* No. 1:08-cv-01425-ODE (N.D. Ga. March 21, 2016)).

As seen above, copyright law changes over time as it tries to respond to changes in technology and societal behaviors. Every three years, the Library of Congress makes decisions about copyright waivers under DMCA; for instance, DVD snippets can be "ripped" for educational purposes, and electronic literary works may be "cracked" to enable read-aloud functionality for individuals with visual impairments (Hobbs, 2016). Other copyright issues, such as the appropriate copyright handling of orphaned works (where the author cannot be identified or contacted), remain. Crowdsourced materials also constitute a copyright headache, and one can only imagine how future formats and content interaction will impact and maybe redefine copyright.

Derivative works gains another dimension in the copyright world as individuals transform existing recorded (thus copyrighted) materials (Ahmeti, 2015). The copyright holder may create derivative works such as adaptations, revisions, translations, extensive excerpts, and different formats of the same work; if others want to make a derivative work, they must ask for permission from the holder. However, if a person *transforms* the work, then copyright permission is not needed; transformation requires adding something new for a further or different purpose rather than a substitution of the original work; example include parodies, mash-ups, remixing. While sampling may be transformative, it should be noted that in the music world, digital sampling is very restrictive.

When in doubt, users should act conservatively. On the other hand, these copyright issues can overwhelm educators, who then may be reluctant to incorporate digital sources, in particular, for fear of the "copyright police." As a result, learners may have little access to valuable open education materials.

FAIR USE

Fortunately, educational institutions enjoy more freedom in the use of copyrighted materials because of the intent of that use: teaching and personal research without market compensation. These fair use exemptions are listed in Section 107 of 17 United States Code 106:

1. The purpose and character of the use, including whether such use is of commercial;
2. Nature or is for nonprofit educational purposes;
3. The nature of the copyrighted work;
4. Amount and substantiality of the portion used in relation to the copyrighted work as a whole; and
5. The effect of the use upon the potential market for or value of the copyrighted work.

The expansion of digital resource has further complicated copyright laws and Fair Use. The ease and speed of digital duplication can tempt the most honest user. Especially when software “suites” facilitate repurposing of information (such as turning a PowerPoint into an outline word processed document), it can be difficult to explain how to use digital materials within copyright limits. How does format define information? How does multimedia incorporation change the nature and intellectual property of each information element? These issues can lead to valuable learning moments during which students can understand the nuanced nature of digital information. The National Council of Teachers of English has developed a code of best practices in fair use for media literacy education, which can serve as a useful set of guidelines (<http://www.ncte.org/positions/statements/fairusemedialiteracy>).

Certainly, students and teachers alike need to cite sources accurate to credit the original creators, but it is no substitute for asking permission. Basically, asking permission requires finding the author of the work, identifying the specific information to be used (full citation, including

which section of the document), the purpose of the use, and the extent of use (e.g., how, where and frequency). Copyright clearinghouses (e.g., <http://www.copyright.com>) can also help users obtain permission.

Fortunately, user-friendly guidelines about fair use are available. The American Library Association for Information Technology Policy has developed copyright educational tools that focus on copyright and fair use: <http://www.ala.org/advocacy/pp/pub/copyright>. The Copyright Advisory Office of Columbia University Libraries’ Information Services has developed a useful fair use checklist that can help the education community respect intellectual property while researching and generating ideas: <https://copyright.columbia.edu/basics/fair-use/fair-use-checklist.html>.

PROJECT CHECKLIST

As multimedia and digital resources are incorporated into presentations, particularly as these products are broadcast online, the school community needs to follow more stringent copyright guidelines. The Washington State Library (2008) has developed a checklist of factors to consider when thinking about using and publishing information. Their first point is probably the most salient: ownership of physical and digital objects doesn’t guarantee use rights.

- Who owns the material? Have you made a good faith effort to find the owner? Who controls access to the physical items? Who will control access to the digital items?
- Are there written agreements as to construction, ownership, and access to the digital collections?
- Determine copyright and permissions status; are the materials owned outright? Can the rights to publish the materials digitally be obtained?
- What percentage of the materials will need research or requests for permissions? What

will be the added cost of devoting staff time to obtaining permissions?

- Is the material in the public domain or covered by copyright that your organization has legally obtained?
- Remember, any original student work cannot be digitized or published without a signed permission form from the student or their guardian if they are a minor.
- Address copyright, permission to copy on your own digital site. Post a statement to guide use of materials on your digital Web site.
- Provide contact information for users of your Web site.

THE CREATIVE COMMONS

A growing interest in the creative “information commons” reflects a philosophy that “information yearns to be free” and that accessible information leads to expanded discourse, knowledge and progress. Public domain documents have exemplified this philosophy for over a century, and government publications carry on this tradition in the spirit of civic engagement.

A welcome alternative is the Creative Commons (<http://www.creativecommons.org>), which enables people to upload and share data with the understanding that any use requires cited acknowledgment, and any changes to the data also need to be uploaded and cited within the Creative Commons. This proactive strategy recognizes the benefits of collaborative knowledge building. In 2005 the International Federation of Library Associations and Institutions (IFLA) called upon the World Summit on the Information Society Summit to promote a global Information Commons where all people could access and disseminate information without restrictions. Two main approaches to the creative commons have emerged: 1) contributing and modifying original information, and 2) facilitating the *sharing* of information.

The idea of a creative commons can be expanded across systems, disciplines, even national boundaries. Creative Commons provides an extensive directory of federated repositories (http://wiki.creativecommons.org/Content_Directories). One of the advantages of federated collaborations is that each site typically controls copyright and licensing issues. Nevertheless, these consortia initiatives require careful, thorough planning, not only in terms of technical requirements but also in terms of fiscal and governance issues.

A recent development related to the Creative Commons is the Open Textbook Initiative. In an effort to make textbooks more affordable, the Higher Educational Opportunity Act of 2008 supported the development of open textbooks. Textbook publishers, such as Flat World Knowledge, offer free online access to textbooks under a Creative Commons license, and low-cost fees for printing copies. In addition, these textbooks are often repurposed so that instructors can essentially create customized texts for their courses. Faculty are more apt to access such materials through repositories than publishers, such as Cool4Ed, Curriki, and OpenLearn. While open textbooks certainly attracts students, some instructors are concerned that materials have not been adequately reviewed so might not be high quality. On their part, some authors are not so keen on this practice because they get little or no fiscal remuneration for the materials they create. Instead, they must be satisfied with possible increase in visibility and reputation, and derive satisfaction from contributing to the field in a timely manner.

Similarly, “open source” is a growing approach to software development, which impacts intellectual property rights, mainly in the areas of patents and licensing. The term emerged with Netscape’s code, which was openly released in 2006 for others to use freely. At this point, the largest repository of open source code, SourceForge, has over than three million registered users. Open source code has been the basis for many social networking applications such as wikis and blogs.

TECHNOLOGY-BASED INFORMATION PACKAGES

Individuals and institutions routinely “package” information together as a way to synthesize the best resources for their clientele. For instance, course “readers” have been a mainstay for years. Copyright permission has always been a part of that production process. Several general copyright issues loom when dealing with technology aspects of packaged information. As libraries and other educational entities jump on the curation band wagon, especially in the digital realm, they need to be aware of copyright issues.

- **Content:** Does the entity have permission to copy, download, digitize, modify, excerpt, or package information? What is the copyright for each document? What limitations exist on such practices: e.g., length of time to keep the copy, number of copies (sometimes within a specified timeframe), extent of copying (such as percentage of the entire work), access rights, purpose limitations (such as referral vs. personal research vs. resale), format.
- **Format:** Packaging information, such as software installation instructions, for internal use, is a much different issue than scanning published cartoons for a public website. Even if the entity gets permission to copy an image, that permission might not extend to resizing or cropping the image. Copyright owners might not permit content to be reformatted or repurposed with reason, since changing context can result in different interpretations of the information. Likewise, pinning up candid student pictures in the faculty mail room requires a different level of permission than broadcasting those same pictures on the Internet, particularly if any under-age students are shown. Indeed, because of stalking and other criminal behaviors, few institutions show captioned pictures of their staff, particularly in online environments.

- **Liability:** Entities need to read licensing agreements carefully to make sure that their packaging efforts comply with the legal language. Users need equitable access, proper authorization, and confidentiality; moreover, no profits can be incurred (Farb & Riggio, 2004). Once entities start to extract and synthesize information, even with permission, they stand to risk being sued. For instance, health information might be followed by some user with unpleasant results. Especially in the areas of law and health, entities need to make sure that they post a disclaimer that they are not legal or health professionals (unless so licensed). Even the software used to package information is likely to entail legal right for its use, particularly if the product is web-based and disseminated externally. The North Carolina State Library Web Portal Collection Development Policy addresses a number of these issues (<http://www.digitalnc.org/about/policies/digitization-guidelines/>).

In addition, specific types of information packages have unique copyright implications, as detailed here.

Courseware

Most campuses now use learning management systems or courseware as a means to access course materials or facilitate communication. Faculty may spend hundreds of hours developing and implementing their course materials: creating their own content, selecting and linking other existing resources, sequencing content, facilitating student interaction, and providing feedback. Who owns the content?

Generally, their affiliated institution owns the course and its content as part of the faculty contract; it is considered work for hire. However, each institution or educational system determines intellectual property and accompanying copyright agreements. Can the faculty make money inde-

pendently from that course material? Usually only if they transform the existing work. Adding an Amazon link to the instructor's campus website so students can buy an instructor's publications can be asking for trouble. Can another faculty member copy the course entirely, and teach with it? They need permission from the copyright holder, which is probably the campus administration; faculty should ask about the appropriate protocol ahead of time – and administration should inform all faculty about course intellectual property rights from the start. The American Association of University Professors Copyright, Distance Ed & Intellectual Property does a good job of explaining implications of technology for teaching and learning, and provides sources to help faculty protect their intellectual property (<https://www.aaup.org/issues/copyright-distance-ed-intellectual-property>).

What about copyright and fair use of MOOCs (Massive Open Online Courseware)? Institutions anticipate unlimited participation and open access; however, they are increasingly likely to include carefully worded licensing agreements even though they provide free access to students. In other words, one may look but not copy; fair use might not apply to non-MOOC educators. To be safe, instructors should simply link to the original material or online course rather than download or copy it (Hui-Wen & Chao-Chen, 2012).

Repositories

Increasingly, institutions and organizations are developing digital repositories, either storing data or storing the documents themselves. In both cases, these services manage and disseminate digital resources. The scope of repositories can range from a single program to international consortia. The value of repositories lies in the quality of use of their content, so identifying desired kinds of documents and collecting high-quality materials are key functions. Of special interest now are repositories of data sets; U.S. federal agencies now require a data management plan as part of their grant proposals, so most academic libraries

either establish an in-house repository or join a data management consortium. Major technology management issues of a repository follow:

- Maintaining the data without damage or alteration: storage and security requirements.
- Providing “physical” access to the data, including extracting of it from the archive: authentication, verification, identification, metadata harvesting software.
- Ensuring that the user can understand and interpret the data: display/rendering software and preservation planning.
- Ensuring long-term stability: technology planning and data maintenance (Wheatley, 2004).

In structuring repositories, entities need to consider three layers: storage, database, and application, each of which may have copyright implications. The Academic and Research Libraries of the American Library Association developed an institutional repository SPEC toolkit to help librarians with policies and procedures (<http://www.arl.org/storage/documents/publications/repository-services-report-jan09.pdf>).

One promising practice in digitizing materials is systemwide technology initiatives, usually in the form of federated repositories of archival materials. DSpace (<http://www.DSpace.org>) is an Open Archive Initiative that provides guidelines for digitizing, cataloging, storing, and disseminating unique digital sources. This open source solution emphasizes the need for technical expertise and planning. Nevertheless, the *content within* the repository still needs copyright oversight. Duranceau and Kriegsman provide guidance on open access policies for institutional repositories (http://www.ala.org/alcts/sites/ala.org.alcts/files/content/resources/papers/ir_ch05_.pdf).

Knowledge management (KM) consists of managing the knowledge of an organization, usually collecting, storing, organizing, and disseminating explicit information such as documents with the intent that others will use that

information: an internal repository. Particularly in today's society where in-house knowledge may be proprietary and employees may leave after a short time, making tacit knowledge explicit through the use of a knowledge management system offers an effective way to coalesce expertise. Because they are internal, such KM repositories normally do not require copyright permission, although institutions should set up policies and procedures that delineate levels of access and authorization protocols. As some of the documents may be permitted to be access externally, dealing with copyright permissions from the start is a good practice.

E-Publishing

For decades, in order to inform their constituents, institutions and organizations have created documents for their own members and for the public. These might be as simple as a guide to the library or as sophisticated as a peer-reviewed journal. Be it an original document or a compilation, copyright law must be observed.

For example, Stanford University Library's HighWire Press publishes almost a thousand e-journals. The press prides itself in its efforts to incorporate multimedia, hyperlinks, interactivity, and powerful search engines. Although they publish some free online articles, the press hosts digital content on behalf of many publishers, each of which has specific terms of use. HighWire's notice states: "Unless explicitly stated otherwise, content on HighWire's publishing platform, including content accessible without charge, cannot be copied, re-purposed, displayed on other websites, reprinted, redistributed, entered into a database, modified, used to create derivative works or otherwise re-used without the specific permission of its publisher" (<http://highwire.stanford.edu/about/terms-of-use.dtl>).

In a few cases, one of the impetuses is financial gain, but for the majority of e-publishers the goal is dissemination of information on topics of interest to their clientele. To this end, open access journals have become quite attractive: they are

relatively inexpensive to create (except for labor costs of editing and layout) and disseminate, can include digital features not available in print format such as sound, and may be more timely to publish than print versions. The Scholarly Publishing and Academic Resources Coalition (SPARC) (<http://www.arl.org/sparc>), founded in 1998, exemplifies a collaborative library approach to disseminating research. Internationally, the Budapest Open Access Initiative (2002) asserts that

Removing access barriers to this literature will accelerate research, enrich education, share the learning of the rich with the poor and the poor with the rich, make this literature as useful as it can be, and lay the foundation for uniting humanity in a common intellectual conversation and quest for knowledge.

On the other hand, copyright issues connected with maintaining a permanent collection of the e-publications may be overlooked. Furthermore, if indexing services do not receive a copy of the journal, then access to the information will become even more limited. The Association of College and Research Libraries provides a scholarly communication toolkit, which specifically supports a more open system of digital scholarship (http://acrl.ala.org/scholcomm/?page_id=42). Charles W. Bailey, Jr. maintains the most extensive overview about digital scholarship publications (<http://www.digital-scholarship.org/about/overview.htm>).

Linking Sources

The "off-the-street" online user would like a single-stop searching tool that would link all relevant material: primary sources, secondary sources, print, web-based, audiovisual. Increasingly, libraries are using "discovery" tools that approximate this process; a first search typically results in a list of articles and a list of books and media, with a number of additional criteria (e.g., format, date, language, full-text, etc.) to filter results. The new cataloging standards are built on

the Functional Requirements for Bibliographic Records (FRBR) conceptual model, which facilitates linking individual data elements of a source with other sources (Howarth, 2012). The Open Linked Data Project further extends this idea through the semantic web. While an exhaustive one-stop tool is probably not feasible in the near future (despite Google's intention), smaller-scale reference linking has become an attractive way to add value to information, and facilitate information services. There are several ways to link sources:

- Between citations (e.g., databases)
- Within sources (e.g., hyperlinks)
- Between sources and applications (e.g., course management systems)
- Between sources and services (e.g., e-reserves)

Usually, metadata provides the basis for these actions. Digital Object Identifiers (DOI) provide an international standards-based “system for persistent and actionable identification and interoperable exchange of managed information on digital networks” (<http://www.doi.org>).

Links may be categorized as either static (created as a permanent link) or dynamic (created in response to user action). Normally, information packagers pursue dynamic solutions where linking can occur without full control of the resources. This approach is attractive in theory, but may threaten vendors who are less comfortable about open access. Furthermore, as entities create documents that include links, be it at the citation or source level, they need to consider how the source information is captured and authenticated, processing between links, and hosting services. All entities and related protocols need to be interoperable as well as legally compliant (Van de Sompel and Hochstenbach, 1999). Certainly, links within the same website are permissible, but sometimes links to a different website may require permission from the owner of that site; the law is not clear about that issue.

SOCIAL MEDIA

Social media signals a new level of complexity in copyright: the interactive Internet. It fosters collaborative work, which results in collective intelligence. Determining the author and ways to recompense the use of social media works presents a real challenge legally. Furthermore, these documents are often dynamic in nature so that the content may change moment to moment. What, indeed, is the basis of copyright? At best, it would need to be based on a time-stamped version that acknowledges both the time of the creation as well as the time that someone accessed it to use it.

Some people assume that all document on the Internet are free, which assumption is clearly wrong, as discussed above. In some cases, the document creators function on a creative commons agreement, or publish a statement of use. Wikipedia serves as a good example of a community-built wiki that addresses copyright issues. Wikipedia is a good example. It has what it calls a “copyleft” agreement, which states: “Wikipedia content can be copied, modified, and redistributed so long as the new version grants the same freedoms to others and acknowledges the authors of the Wikipedia article used” (<http://en.wikipedia.org/wiki/Wikipedia:Copyrights>). Most of Wikipedia's articles are co-licensed under the Creative Commons Attribution-Sharealike 3.0 Unported License and the GNU Free Documentation License. Wikipedia states that each image notes its legal usage basis.

Blogs are usually the work of a person or group, and are in a recorded form, so they follow normal copyright laws by default. That argument would then also apply to comments that followers make. Of course, the blogger can provide further guidance, be it a statement about being in public domain or a explicit clause that bars anyone from reproducing any of the material in any format.

Podcasts and videocasts consist of compressed audio and video files that are then broadcast by the creator or an aggregator such as iTunes. Again, these are recorded documents so are assumed to

fall within copyright laws. ‘Casts can be easily downloaded and shared, so it is wise to contact the originator or from a reputable Internet site. Difficulties tend to arise when someone uses copyrighted sources (particularly commercial music and videos) to create a mash-up or other kind of derivative work. It can be as simple as videotaping a toddler dancing to a Bruce Springsteen song, and then uploading it to YouTube. Unless the videotaper got written permission from the publisher (or who else owned the copyright) to use that song, the broadcasting that that video constitutes illegal usage if for no other reason than that the song writer and producer should be paid royalties every time the video with the music is played. If the videotaper wants to avoid legal action, he or she should either get permission or not upload or broadcast the video; instead, that video can be recorded for home use, and friends can come over to see the bouncing baby bopper. The Podcasting legal guide (http://wiki.creativecommons.org/Podcasting_Legal_Guide) by Vogel, Marlick and the Berman Center provides current in-depth information on legal uses of these media.

Social bookmarking sites, such as Delicious and LibraryThing, provide a means to store and describe (“tag”) URLs for later retrieval and use. Similar to a bibliography, a list of lists falls under copyright law. Therefore, the creator of each list has the right to determine who can use it and the conditions of its use. To some degree, that can be controlled through the settings that the originator chooses. Usually, each overarching social bookmarking site has a policy about the use of the site as a whole and the use of individual bookmark collections. Increasingly, social media websites such as StumbleUpon integrate social bookmarking, web searching, and blogging.

RSS means “really simple syndication,” and an RSS feed is a means to gather information from other websites to a central location. Every single website that offers an RSS service has its own copyright policy about legal use. Of course, if the originating news is in the public domain and is copied from an open source or creative commons/

copyleft site, then that information can be used freely. One can even broadcast that information via his or her own RSS feed; however, it is usually safer to keep one’s personal RSS feeder private and avoid copyright problems altogether. In terms of copyright law, “really simple” is not really simple.

File sharing is a central feature of social networking, as evidenced by the high use of Flickr and YouTube in particular. Without explicit permission, only the owner of the files can reproduce with work, create derivative works, distribute copies of the work to the public, display or perform the work publicly (including by digital transmission). Therefore, such files cannot be integrated into a school DVD for sale or placed on an institutional website or played at a public meeting. In short, one should read each file-sharing site’s copyright guidelines before downloading documents (Talab & Butler, 2007). Fortunately, some social networking sites include as a searching filter Creative Commons/public domain so that the user can appropriate the file in good conscience.

Intellectual property issues become even stickier when one joins social network sites (SNS) (Wauters, Lievens & Valcke, 2014). Often one has to register to gain access to the network’s files, which translates into providing personal information as the cost for such membership. These websites usually have terms of use contracts, but users tend not to assess the implications and consequences of sharing their own information, such as selling personal information to SNS advertisers. Users should be aware of these practices, and, at the least, check for privacy setting options. At present, SNSs offer little protection to users, and government regulators do not stringently monitor or enforce intellectual property rights. *Caveat emptor*.

Even more fundamentally, crowdsourcing challenges the current notions of intellectual property and copyright (Almeti, 2015). If an editor solicits content, monitors and reviews each input, explains the terms of contributions clearly from the start, and produces a “fixed” work, then copyright is usually not a problem. When the work is created

on a SNS, with each person able to add and edit each other, with no end date and no recorded agreement, then copyright is likely to default to the site's creator and administrator, but the situation can be messy, especially if some kind of commercial or economic benefit is incurred. A Creative Commons license stated at the start of the enterprise is probably the best solution. Technology that can support micro-royalty payments is improving, but the logistics are still daunting. As mentioned before, clear and legally "tight" contracts stipulated from the beginning, and signed before anyone contributes, remain the best solution for the time being.

ADMINISTRATIVE ISSUES

Because everyone has to comply with copyright law, it behooves everyone to be knowledgeable about it, but it can be difficult to keep current. In the final analysis, administrators are responsible for such compliance. Understanding copyright, particularly as it applies to technology, enables administrators to address the risks of technology as they optimize its benefits. The Consortium for Educational Technology for University Systems (<http://www.cetus.org/intellectual-property/>) suggests ways that administrators can oversee intellectual property.

1. Foster intellectual creation, its dissemination, and its use.
2. Develop, implement and enforce technology policies and procedures that optimally support the mutual rights of the academic community.
3. Facilitate the legal and ethical access, storage, and attribution of intellectual works by the entire academic community in coordination with information providers.
4. Review and revise intellectual property policies and procedures in light of technology and cultural changes that impact the academic community.

Digital Rights Management

E

As digital resources grow in number and percentage in institutions, the function of digital asset management has grown in complexity. How do institutions keep track of all their electronic resources as well as make sure that they are effectively stored and retrieved? Especially in today's networked world, digital asset management has to address interactivity options and social networking features, such as push technology, RRS feeds, comment/messaging options, incorporation of faculty repositories, enterprise mash-ups, and user-customizable folksonomy "shells." While some of these features are low-cost plug-ins, their management and incorporation into the library's digital collection system can involve sophisticated technical support, which is usually *not* inexpensive. Furthermore, such customizations need to be well documented and maintained.

Within that scope, the field of digital rights management (DRM) has become increasingly important. Broadly speaking, DRM deals with the exact rights that each digital content has: who holds those rights and under which circumstances, who has authority to access that content and how that can be insured as well as preventing non-authorized people access (Calhoun, 2005). Complying with intellectual property regulations is very difficult; so digital rights management technologies are being employed to control content use. While automated systems conveniently take care of authentication issues and facilitate fair royalties compensation, they can also jeopardize privacy rights and leak into discriminatory profiling practices. E-resources sometimes are not device-neutral, so rights are sometimes given for just one operating system, which belies the concept of intellectual property. Installation of DRM-protected content can be burdensome, and may malfunction, preventing authorized access. At the other end of the process, when the access key to DRM-protected content is lost or the device becomes obsolete, the content itself becomes unreadable, even if it has been paid for

legally. DRM can also pose problems for users with special needs. Furthermore, these DRM digital tools are not standardized. In short, DRM technology may be unavoidable when purchasing some digital resources, but needs to be carefully used (Houghton-Jan, 2007).

Even without considering DRM technology, digital rights management can be complicated. Libraries are usually the body that deals with commercial digital resources, and increasingly deal with institutional digital documents as part of knowledge management initiatives. Within that scope, subscription databases (including e-book aggregators) constitute the major copyright effort because each vendor has a unique set of license agreements. Farb and Riggio (2004) list basic elements of most contracts: scope, completeness of content, duration, warranties, indemnification, access, confidentiality, sharing, archiving, disability compliance, and usage statistics. Increasingly, libraries are considering “leasing with an option to buy” licenses as a way to insure access through backfile ownership; licensing a database with no copyright to the content beyond the date of the license is not attractive.

Training

Faculty need to become aware of intellectual property issues, and their impact on education. However, with awareness sometimes comes fear and over-caution about integrating technology into their instruction. The wiser choice is to empower faculty to safely and successfully navigate copyright law waters. Trainers can share examples of copyright compliance, and have faculty discuss copyright case studies and scenarios such as downloading streaming video or examining licensing fees (Russell, 2010). Other helps include permission forms and checklists of good practices, such as: citing sources thoroughly, linking to sources rather than posting the full text, asking permission to copy sources, checking database terms of use, and using Creative Commons and public domain sources (Disclafani & Hall, 2012; Shin, 2015).

To help the educational community remain in compliance, librarians sometimes provide training about copyright because of their role as information managers. Institutions should also maintain a reviewed bibliography of copyright resources that can be consulted easily. Some good websites that explain copyright follow.

- **U.S. Copyright Office (<http://www.copyright.gov>):** Provides copyright forms and regulations, and a searchable database of copyright registrations.
- **American Library Association Copyright (<http://www.ala.org/advocacy/copyright>):** Good information and links about legislation, intellectual property, and international copyright activities.
- **Columbia University Libraries Information Services Copyright Advisory Office (<http://copyright.columbia.edu>):** Well-respected site targeted to librarians and educators.
- **Consortium for Educational Technology for University Systems (<http://www.cetus.org/>):** University coalition of resources on intellectual property
- **University of Texas (<http://www.utsystem.edu/offices/general-counsel/intellectual-property>):** Many links on copyright issues; they have a good crash course tutorial at <http://www.lib.utsystem.edu/copyright/>.
- **Stanford University Libraries Copyright & Fair Use (<http://fairuse.stanford.edu>):** Focuses on fair use, the public domain, and the permissions process.
- **Copyright for Educators (<http://www.koce.org/classroom/copyright.htm>):** Public television site designed to help educators learn about fair use.
- **Intellectual Property Legal Center (<http://www.cetus.org/fairindex.html>):** Provides information on several aspects of intellectual property, with a separate section on ebooks.

- **Cybercrime (<http://www.cybercrime.gov>):** Computer crime and intellectual property section of the Criminal Division of the U. S. Department of Justice.
- **Software & Information Industry Association (<http://www.siiia.net/Divisions/IP-Protection-Services>):** Their Intellectual Property Division campaign tries to balance enforcement with education.

FUTURE TRENDS

Copyright law tends to react to new technologies and practices, so it is unlikely that it will pro-actively change. Nevertheless, technology will continue to change and expand, so the law will need to respond to, and accommodate, such changes. In addition, technology has also impacted the interchange of information, from the local to the international level, which will affect copyright law directly and indirectly.

Both technology as a whole and the phenomenon of resource sharing is already impacting the publishing model as it develops new models of production and diffusion. Not only do commercial publishers need to address digital formats and ereaders with uneven standards and proprietary operating systems, but they have to deal with unique collations of content, non-commercial competitors, all of which impact the various copyright options. Particularly as product buyers and licensees share these documents, such as libraries and organizations, determining what is a just royalty can be problematic. Micro-payments for access and downloading or printing are being incorporated, but the models for fair rights still need to be hammered out.

Interactive technology corners the market in terms of complex copyright issues. As noted

above, determining the recordable version of a document when multiple users can edit it almost simultaneously can be a copyright nightmare. As mentioned above, documents and repositories are increasingly the products of collective intelligence. In the face of such dynamics, it seems almost laughable to consider that current copyright law allows for proprietary rights for 70 years after the author's death in some cases.

The core concept of a work also needs further scrutiny as works are manifested in so many different formats and appearances. Right now, an e-book might be licensed to be readable on a single device; is that fair if the content itself exists separately from that device? How different does the document have to be in order to be considered a separate work? As the content, container, and metadata elements are considered separately, will copyright elements follow? Especially as web 3.0 emphasizes the relations among documents, those relationship themselves might fall under copyright just as bibliographies and links can.

As this practice expands in global environments, copyright law may need to become more regulated on the international scale so that authors can be fairly compensated. Theoretically, virtual worlds are associated with the country in which the software is based, but such boundaries may also come into question as documents are created in those virtual realities by people in other countries.

This issue of global access also brings up the issue of universal access – for individuals with special needs as well as those with language and linguistic differences. For example, in 2013 the WIPO Marrakesh Treaty required countries to provide technological protection measures to control access to, and use of, digital copyrighted works by people who are print disabled. To what extent is a simpler version or a translation of a document considered a derivative work, especially if it is the result of an automated application? To what extent is automatic captioning covered by copyright? To what extent can fair use be applied?

Particularly as digital programs automate such variations, copyright law will need to be revisited once again.

CONCLUSION

How does the spirit of the U.S. Constitution “to promote the progress of science and useful arts, by securing for limited times to authors and inventors the exclusive right to their respective writings and discoveries (Article 1, Section 1, Clause 8) manifest itself in the 21st century? The countries forefathers could not have imagined the Internet and the semantic web.

Nevertheless, educators and other researchers have the strength of fair use behind them so they can promote an effective learning environment, be it face-to-face or virtual. In addition, collaborative practices such as the Creative Commons promote intellectual sharing in order to advance their fields.

Technology has significantly expanded the intellectual arena in terms of production and sharing of information. In response, copyright law has tried to codify intellectual property rights to balance the rights of the creator and appropriate access to information. Such laws are often format-specific, reflecting both the nature of the medium as well as the perspectives of their producers; however, the content and format are splitting apart so that each digital element may be considered separately. Dissemination models have also changed drastically, so how the information buyer uses that information constitutes another can of copyright worms. The international dimension that technology facilitates adds another layer of copyright issues. Furthermore the dynamic nature of socially constructed information insures that copyright law will always lag behind technology and social behavior. Underlying principles must be the mainstay since practices will continue to vary according to changing contexts.

At the least, leaders and other educators need to keep abreast of copyright issues, and try to comply with them within the context of their use.

REFERENCES

- Ahmeti, E. (2015). Creation of copyright derivative works through digital services. *Proceedings of the Multidisciplinary Academic Conference*, 1-7.
- Budapest Open Access Initiative. (2002). Retrieved October 30, 2010, from <http://www.soros.org/openaccess/read.shtml>
- Calhoun, T. (2005, March). DRM: The challenge of the decade. *Campus Technology*, 18-20.
- Disclafani, C. B., & Hall, R. (2012). Stop saying no: Start empowering copyright role models. *Journal of Library & Information Services in Distance Learning*, 6(3-4), 251–264. doi:10.1080/1533290X.2012.705151
- Farb, S., & Riggio, A. (2004). Medium or message? A new look at standards, structures, and schemata for managing electronic resources. *Library Hi Tech*, 22(2), 144–152. doi:10.1108/07378830410524576
- Hobbs, R. (2016). Lessons in copyright activism: K-12 education and the DMCA 1201 exemption rulemaking process. *International Journal of Information and Communication Technology Education*, 12(1), 50–63. doi:10.4018/IJICTE.2016010105
- Houghton-Jan, S. (2007). Imagine no restrictions. *School Library Journal*, (June): 53–54.
- Howarth, L. (2012). FRBR and linked data: Connecting FRBR and linked data. *Cataloging & Classification Quarterly*, 50(5-7), 763–776. doi: 10.1080/01639374.2012.680835

Hui-Wen, H., & Chao-Chen, C. (2011). Research on copyright issues of opencourseware materials. *Journal of Educational Media & Library Sciences*, 48(3), 401–405.

International Federation of Library Associations and Institutions. (2014). *Lyon Declaration*. The Hague, Netherlands: International Federation of Library Associations and Institutions.

Russell, C. (2010). The best of copyright and VideoLib. *Library Trends*, 58(3), 349–357. doi:10.1353/lib.0.0095

Shin, S. (2015). Teaching critical, ethical and safe use of ICT in pre-service teacher education. *Language Learning & Technology*, 19(1), 181–197.

Talab, R., & Butler, R. (2007). Shared electronic spaces in the classroom. *TechTrends*, 51(1), 12–14. doi:10.1007/s11528-007-0004-1

Van de Sompel, H., & Hochstenbach, P. (1999). Reference linking in a hybrid library environment. *D-Lib Magazine*, 5(4). doi:10.1045/april99-van_de_sompel-pt1

Wauters, E., Lievens, E., & Valcke, P. (2014). Towards a better protection of social media users: A legal perspective on the terms of use of social networking sites. *International Journal of Law & Information Technology*, 22(3), 254–294. doi:10.1093/ijlit/eau002

Wheatley, P. (2004). *Institutional repositories within the context of digital preservation*. York, UK: Digital Preservation Coalition.

KEY TERMS AND DEFINITIONS

Copyright: Laws that regular the use of the work of a creator.

Crowdsourcing: Practice of enlisting the help of many people in order to generate content, particularly in online environments.

Creative Commons: A nonprofit organization that seeks to make creative work available for others to build on while respecting intellectual property rights.

Digital Rights Management: Access control technologies used to protect the copyright of electronic media.

Fair Use: The legal conditions under which one can use copyrighted material without obtaining permission and paying royalties.

Intellectual Property: Any intangible asset that consists of human knowledge and ideas; the ownership of ideas.

License: Permission granted by authority to exercise certain rights and privileges that would otherwise constitute an illegal act.

Open Access: Free availability on the public Internet.

Open Source: Coding that is freely available to the public, usually applied to software development.

Social Media: Interactive web; enables people to collaborate and share online.

Employing Educational Robotics for the Development of Problem-Based Learning Skills

Nikleia Eteokleous

Frederick University Cyprus, Cyprus

INTRODUCTION

The technological improvements within the robotics field and its expansion to various fields such as medicine, industry and education, calls for robotics integration within the educational practice as learning tools. Robotics in the classroom has taken a global momentum especially because of its positive contributions in the teaching of science, technology, engineering and mathematics (STEM) (Benitti, 2012). Additionally, research has shown that robotics integration in education promotes the development of various non-cognitive skills, however extremely important life skills. For example, reasoning, problem solving, tinkering, sequencing, computational thinking, decision making, scientific investigation, collaboration, knowledge construction, critical thinking, creativity, communication (Bers, Ponte, Juelich, Viera & Schenker, 2002; Benitti, 2012; Chambers & Carbonaro, 2003; Eteokleous, 2016; Miglino, Lund, & Cardaci, 1999; Resnick, Berg, & Eisenberg, 2000; Williams, Ma, & Prejean, 2010).

Educational systems are responsible in preparing students (future citizens) for this ever-changing Hi-Tech, globalized, interconnected world. Numerous 21st century skills are reported in the literature as important to be developed by future citizens as the means to address the needs and demands of the society. The 21st century skills have been outlined and described by various researchers and reports (e.g. Ananiadou & Claro, 2009; Bybee & Fuchs, 2006; Griffin & Care, 2005; Mojika, 2010; Rotherham & Willingham, 2010; Trilling & Fadel, 2009), and can be summarized

DOI: 10.4018/978-1-5225-2255-3.ch217

as follows: communication, collaboration, critical thinking, problem solving, knowledge construction, creativity – innovation, self-directed learning, global citizenship and digital literacy. The changes in the global competition and collaboration, the focus on service economy, as well as the information growth, constitute the development of the 21st century skills extremely important. Given the aforementioned, the workforce needs have changed, the job tasks and type of work are changing and consequently the required skills are changing.

Problem solving and digital literacy is one of them and robotics and programming are becoming important elements within the educational settings. The students need to be provided with the opportunities to experience tinkering, fabrication, design and create technological artifact & interactive objects, construct their own meaningful projects, experience the scientific method of inquiry (Bers, 2008a; Bers, 2008b; Bers, Matas & Libman, 2013; Bernstein, Mutch-Jones, Cassidy, Hamner, & Cross, 2016; Eteokleous, 2016). Consequently, educators need to design the appropriate learning environments where students have the opportunity to develop the aforementioned skills.

Main Aim

Robotics activities are related to addressing a problem, and usually problems in authentic, real situations. The students are given a driving question and are requested to solve a “problem”. Having noticed this connection in relation to the pressing need to develop 21st century skills, the

current study evaluates the integration of robotics as an educational tool within the teaching and learning process where the problem based learning (PBL) method and the interdisciplinary approach are intertwined. Specifically, robots are used as cognitive-learning tools in order to apply the problem based learning method in early elementary grades (2nd and 3rd graders) in curricular-integrated activities (interdisciplinarity). More importantly, the study aims to examine whether the integration of robotics as cognitive-learning tools influence the development of the following PBL skills: creativity - innovation, critical thinking, and collaboration.

BACKGROUND

Educational Robotics

The idea of robotics integration in education has been around for more than 20 years (Miglino, Lund, & Cardaci, 1999; Papert, 1980). However, the great revolution in the field of educational robotics has been achieved throughout the last decade, where robotics escaped the laboratory and made efforts to connect to education (Chambers, & Carbonaro, 2003). The robotics materials (building blocks/bricks, sensors and motors) are perceived as toys by the children and research revealed that regardless of age, educational background and interests, students consider working with robots to be “fun” and “interesting”. (Chambers & Carbonaro, 2003; Williams and Prejean, 2010). Numerous research studies suggest that robotics integration for educational purposes is an effective teaching method; arguing that if robotics activities are appropriately designed and implemented have great potential to significantly improve and enhance the teaching and learning process (Benitti, 2012; Bauerle, & Gallagher, 2003; Bers et al., 2002; Eteokleous, Demetriou, & Stylianou, 2013; Papert, 1993).

Research has shown that robotics integration in education promotes the development of student higher-order thinking skills such as application,

synthesis, evaluation, problem solving, decision making, and scientific investigation (Bers et al. 2002; Chambers & Carbonaro, 2003; Resnick, Mojica, 2010; Berg, & Eisenberg, 2000). In order to achieve the above, robotics need to be integrated as tools and not as subject matters in the educational practice. When robotics is integrated as a subject matter, as an autonomous entity, and not within a well-designed lesson plan, there is limited educational potential and value. On the other hand, robotics integration as a learning tool, in selected teaching cases exploits its full potential; therefore it upgrades and enhances the teaching and learning process and promotes school transformation (Eteokleous, et al., 2013). The intention of this approach is not to learn how to use the robotics package, and its programming software, but to use it as a tool within a specific educational context to achieve learning objectives. In other words, robotics is employed as a tool to teach and deliver concepts within various subject matters such as Mathematics (Whitehead, 2010), Engineering (Craig, 2014), Science (Vollstedt, 2005), Physics, and even in non-technology related fields such as Biology, Psychology (Bers, Ponte, Juelich, Viera & Schenker, 2002; Eguchi, 2007, Eteokleous, et al., 2013; Craig, 2014). *Robotics integration* in the teaching and learning practice is defined as the use of robotics by students as a tool that enhances their learning experience and supports the achievement of specific learning goals (Ward, et al., 2012; Eteokleous, et al., 2013).

Problem-Based Learning (PBL) and Robotics

Problem-based learning is an instructional method characterized by the use of “authentic” problem sets, as contexts for students to develop critical thinking and problem solving skills, and acquire the necessary course concepts. Along the same lines, problem based learning is defined as “a systematic teaching method that engages students in learning knowledge and skills through an extended inquiry process structured around complex, au-

thentic questions and carefully designed products and tasks” (Markham, 2003, p. 4). Additionally, PBL is defined as the approach that challenges students to learn through engagement in a real problem or situation (Domin, 1999; Duch, 1995; Grant, 2009). Students are presented with real-world, multidisciplinary problems that demand critical thinking, engagement, and collaboration. Given the aforementioned, it is a challenging and demanding process, requesting students to use various types of thinking in order to realize, capture and solve the problem given.

Usually robotics activities are related to addressing a problem, and sometimes problems in authentic, real situations. The students are given a driving question and are requested to program the robots in order to perform a number of activities, aiming to solve a “problem” and/or achieve a goal (Gibbon, 2007; Varnado, 2005). There are also studies showing improvements in problem solving skills as well as self-regulation, recognition, orientation skills (Baker, Nugent, Grandgenett, & Adamchuk, 2012; Karim, Lemaignan, S., Mondada 2015). Having noticed this connection, the current study aims to bring together the PBL approach and robotics integration in the educational settings, in order to examine the role of robotics in promoting the development of specific problem-based skills (creativity-innovation, critical thinking and collaboration) in early elementary grades.

The learning environment where robotics are employed as educational tools includes exercises and activities where the problem-solving process is integrated. Therefore, the students are actively involved in designing, building and organizing the entire process in order to reach to a solution (to solve the problem). Additionally, in order to solve the problem, the students need to construct the appropriate robotic model as well as program it. Construction and programming provide instant feedback to students, where the students can take corrective actions through experimentation and experience the process of testing and re-testing (Alimisis, 2009; Arlegui, Menegatti, Moro, Pina, 2008).

The Educational Robotics Package

There is a great variety of educational robotics packages (BeeBots, BlueBots, Thymio, Cubelets, Roamer, LegoWeDo, Engino, Lego Mindstorms). For the purposes of this research the BeeBots robotics package was employed. The reasons for choosing the BeeBots were due to their characteristics and ease of use given the fact that the students did not have any other experience with robotics. The BeeBot is a colorful, easy-to-operate, and friendly little robot that invites students to experience sequencing, estimation, problem-solving, and many more. It can be programmed to move forward and backward for 15 centimetres, and to turn 90° right and left. The robot can store up to 43 steps. It has also the Pause button, which is considered as a step. It can be programmed in order to perform various educational exercises using specific floor mats (i.e. alphabet mat, geometry shapes mat, treasure island mat). The floor mats are necessary since the BeeBots are programmed to perform various activities using floor mats of different concepts.

RESEARCH METHODOLOGY

A descriptive case study approach was employed collecting both quantitative and qualitative data (Cohen, Manion, & Morrison, 2008; Yin, 2003). Classroom interventions were designed and implemented for two months in two classes: 2nd and 3rd grade. The population of the study was 43 primary education students: 21 2nd graders and 22 3rd graders. Two elementary school teachers and two teacher assistants closely collaborated with the Robotics Academy in designing and delivering the lessons. Specifically, The Robotics Academy at Frederick University (Department of Education) (<http://akrob.frederick.ac.cy>, <https://www.facebook.com/AkadimiaRompotikis>), is a research and educational unit that aims to promote and conduct research in the area of robotics education. The Robotics Academy had the overall responsibility

Table 1. The floor mats developed by students

Grade	Subject Matters	Floor Mats
2 nd	Mathematics	Solid shapes
	Language and Linguistics	Our school map
	Arts	<i>Developed the floor mats</i>
3 rd	Language and Linguistics	The Alphabet
		Word Search Puzzle
		Professions of the Past
	Arts	<i>Developed the floor mats</i>

to professionally train and educationally and scientifically support the teachers and the assistants, and provide the educational robotics packages. They attended one-week intensive professional development training on educational robotics. The teachers were given the flexibility to choose the disciplines (topics) that they thought it would better fit to robotics.

The classroom intervention duration was 5 weeks, involved various disciplines (interdisciplinarity) and was divided in three phases. During the 1st phase the students were given an interactive presentation regarding robotics, its usefulness and value in our daily life. At the 2nd phase, during the Arts Course, the students collaborated to design and develop the floor mats to be used as educational material within other disciplines. Specifically, the 2nd graders developed two floor mats: The Solid Shapes and The School to be used in Mathematics and Health Education respectively. The 3rd graders developed four different floor mats: The Word Search Puzzle, Professions of the Future, The Similes and The Alphabet to be used in the Language and Linguistic Course and Geography (See Table 1). Finally, during the 3rd phase the teachers delivered the lessons designed where robotics employed as cognitive learning tools through various exercises.

Questionnaires and focus groups with the students, as well as classroom observations conducted by the researcher's team were the main methods of data collection. Three different instruments were used, where each instrument measures different

problem based learning skill: 1) innovation - creativity, 2) collaboration, and 3) critical thinking. The instruments were taken from the Buck Institute of Education (www.bie.org) and were adjusted for the purposes of this study. They were translated in Greek and pilot tested. Specifically, two teachers and five students participated in the pilot study. The authors took into consideration the teachers' and students' comments and accordingly adjusted the instruments.

The instruments were used in order to develop the three data collection methods: questionnaires, observations and focus groups. Pre- and post-questionnaires were given to students. The questionnaires were different for each grade due to the age of the students (as suggested by the Buck Institute of Education). For the 2nd graders a simple form of questionnaire was given, where a combination of smiley / sad faces and phrases (as responses) were used. However, for the 3rd graders a different form of questionnaire was used; they were required to address various statements using a 5 likert-type scale. (See see Table 2).

The observation templates and the focus groups protocols were developed based on the same parameters as the questionnaires in order to record and evaluate the development of the PBL skills as well as to overall examine the teaching intervention. Observations were conducted throughout all lessons delivered. Additionally, four focus groups were conducted (two focus groups for each grade) by the completion of the classroom interventions. A total of 15 students participated

Table 2. The PBL skills questionnaires

Grade	Questionnaires	Parameters	Questions	Question Type
2 nd	Critical Thinking		6	3 point - Likert scale
	Collaboration		5	
	Innovation – Creativity		6	
3 rd	Critical Thinking (4)	Analyzing driving question and begin inquiry	2	5 point - Likert scale
		Gather and evaluate information	2	
		Use evidence and criteria	3	
		Justify choices	2	
	Collaboration (6)	Takes responsibility	4	
		Helps the team	4	
		Respects others	2	
		Makes and follows arrangements	4	
		Organizes work	2	
		Works as a whole team		
	Innovation – Creativity (4)	Define the creative challenge	1	
		Identify resources of information	1	
		Generate and select ideas	4	
		Present work to others	1	

at the focus groups: seven 2nd grade students and eight 3rd grade students. The students were selected on a voluntarily basis. Through the focus groups the researcher's team was also aiming to capture students' reactions, experiences and opinions regarding the teaching intervention.

The quantitative data collected from the questionnaires were used to perform descriptive (frequencies, percentages, cross tabulations) and inferential statistics (t-test, ANOVA). The SPSS package (version 19) was used. The qualitative data collected through observations and focus groups were analyzed by using the continuous comparison of data approach (Maykut & Morehouse, 1994).

MAIN FOCUS OF THE PAPER

Through the activities designed, the students were involved in the extremely interesting process of programming and controlling the robots. The activities performed with the robots besides being enjoyable and creative; they promoted learning

by playing, and specifically the development of some of PBL skills. Specifically, the analysis of the pre- and post-questionnaires revealed the development of the creativity-innovation and the critical thinking skills. However, the collaboration skills were not developed as much as the aforementioned skills. This is due to minor conflicts observed within the teams while trying to design and develop the floor mats as well as to program the BeeBots. Students were so excited in using the robots that the majority of them were very anxious in holding, touching and programming them. In some cases it was impossible for students to hold turns in using the BeeBots. Finally, no statistical significant differences were revealed in regards to age and gender.

The observations and the focus groups complemented and further explained the results of the questionnaires in regards to the development of PBL skills. Additionally, several issues were revealed in regards to teachers' and students' role, difficulties encountered, student attitudes and opinions.

The observations assisted also in gaining and evaluating information on issues that cannot easily and/or directly addressed through the questionnaires. The observations showed that during the lessons where the robots used; more time was needed than the allocated time for each teaching period (40 minutes). The aforementioned was observed in almost all courses (i.e. Maths, Health Education, Language and Linguistics) besides the Arts. Also, observations revealed that although students followed the directions given by the teachers and addressed the exercises and problems given, they wanted to be granted time to “play” with the BeeBot. One of them mentioned “...we want to develop the different paths for the BeeBot...”. The robots intrigued their imagination and creativity in developing by themselves the routes for the BeeBot to follow. They also reported that they wanted to develop activities for other students to solve (program the robots).

The role of teachers was extremely important in helping the students, guiding, monitoring and facilitating the teaching and learning process. Additionally, the presence of a teacher assistant deemed necessary in order to provide any assistance needed to the teacher, since in some cases the students were anxious and excited, given the hands-on character of the activities and the student-centered learning environment developed.

Finally, the focus groups helped the researcher looked for themes that helped her in gaining better understanding and more information on students’ experiences and opinions. The majority of the students reported that the educational robotics experience was extremely enjoyable and very interesting. Overall, the students examine various concepts (interdisciplinary approach) by programming the BeeBot, having the chance to develop various knowledge and skills through the process of programming. The students experience the multifaceted process of problem solving and decision-making, as well as cultivate collaborative and exploration skills.

FUTURE RESEARCH DIRECTIONS

E

The results of the study provide the foundation to further investigate robotics integration in the teaching and learning process. Some more studies were designed and implemented in relation to the aforementioned end, where different educational robotics packages (Lego WeDo, Lego EV3, mBots) are used for the development of problem-solving based skills. Additionally, BeeBots and Lego WeDo are being employed within a newly designed innovative curriculum developed by Frederick University Robotics Academy (<https://www.facebook.com/AkadimiaRompotikis/>) in order to examine the development of computational thinking, self-regulation and spatial skills in elementary schools students. Last but not least, the impact of robotics in developing self-confidence and self-fulfillment to special needs students is planned to be examined the upcoming semester.

CONCLUSION

Overall, the study reveals the great potential of integrating robotics as a cognitive-learning tool across disciplines in order to achieve problem-based learning skills. It examines the integration of robotics within a formal in-classroom setting in contrast to other studies that examine in-formal and non-formal educational settings. Additionally, it focuses on early elementary grades, whereas other studies focus on 5th and 6th graders and even more on secondary education. The results of the study aim to promote research in the field of educational robotics in order to further examining and defining the appropriate learning pedagogies and teaching approaches to be employed when robotics is integrated as cognitive-learning tool. It also highlights the need for teachers to be professionally trained, supported and guided. Finally, the study suggests the value of adopting appropriate robotics packages, learning pedagogies, and teaching

approaches when robotics and programming are integrated within classroom activities. The focus should be to engage teachers in using robotics and programming in their classrooms across disciplines.

REFERENCES

- Alimisis, D. (2009). Teacher Education on Robotics-Enhanced Constructivist Pedagogical Methods. School of Pedagogical and Technological Education.
- Ananiadou, K., & Claro, M. (2009). *21st Century Skills and Competences for New Millennium Learners in OECD Countries*. OECD Education Working Papers, No. 41. OECD Publishing. DOI:10.1787/218525261154
- Arlegui, J., Menegatti, E., Moro, M., & Pina, A. (2008). Robotics, Computer Science curricula and Interdisciplinary activities. *Workshop Proceedings of SIMPAR. Intl. Conf. on Simulation, Modeling and Programming for Autonomous Robots*, 10-21.
- Barker, S. B., Nugent, G., Grandgenett, N., & Adamchuk, I. V. (2012). *Robots in K-12 Education: A New Technology for Learning*. Information Science Reference. IGI Global. doi:10.4018/978-1-4666-0182-6
- Bauerle, A., & Gallagher, M. (2003). Toying With Technology: Bridging the Gap Between Education and Engineering. In C. Crawford et al. (Eds.), *Proceedings of Society for Information Technology & Teacher Education International Conference 2003* (pp. 3538-3541). Chesapeake, VA: AACE.
- Becker, H. J. (1993). Teaching with and about computers in secondary schools. *Association for Computing Machinery*, 36(5), 69-55. doi:10.1145/155049.155066
- Becker, H. J., & Ravitz, J. L. (2001, March). *Computer Use by Teachers: Are Cuban's Predictions Correct?* Paper presented at the 2001 Annual Meeting of the American Educational Research Association, Seattle, WA.
- Benitti, F. (2012). Exploring the educational potential of robotics in schools: A systematic review. *Computers & Education*, 58(3), 978-988. doi:10.1016/j.compedu.2011.10.006
- Bernstein, D., Mutch-Jones, K., Cassidy, M., Hamner, E., & Cross, J. (2016). *Robots and Romeo and Juliet: Studying Teacher Integration of Robotics into Middle School Curricula*. Paper presented at the 2016 Annual Meeting of the American Educational Research Association, Washington, DC.
- Bers, M. U., Ponte, I., Juelich, C., Viera, A., & Schenker, J. (2002). Teachers as Designers: Integrating Robotics in Early Childhood Education. *Information Technology in Childhood Education Annual*, (1): 123-145.
- Bielaczyc, K., & Collins, A. (1999). Learning communities in classrooms: A reconceptualization of educational practice. In C. Reigeluth (Ed.), *Instructional-design theories and models: A new paradigm of instructional theory* (pp. 269-292). Mahwah, NJ: Erlbaum.
- Bybee, R. W., & Fuchs, B. (2006). Preparing the 21st century workforce: A new reform in science and technology education. *Journal of Research in Science Teaching*, 43(4), 349-352. doi:10.1002/tea.20147
- Chambers, J. M., & Carbonaro, M. (2003). Designing, Developing, and Implementing a Course on LEGO Robotics for Technology Teacher Education. *Journal of Technology and Teacher Education*, 11(2), 209-241.

- Cognition and Technology Group at Vanderbilt. (2003). Connecting learning theory and instructional practices: Leveraging some powerful affordances of technology. In H. O'Neil & P. Perez (Eds.), *Technology application in Education: A learning view* (pp. 173–209). Mahwah, NJ: Erlbaum.
- Cohen, L., Manion, L., & Morrison, K. (2008). *Research methods in education* (5th ed.). London: Routledge.
- Craig, D. C. (2014). Robotics Programs: Inspiring Young Women in STEM. In J. Koch, B. Polnick, & B. Irby (Eds.), *Girls and Women in STEM* (pp. 153–174). Charlotte, NC: Information Age Publishing, Inc.
- Cuban, L., & Pea, R. (1998, February). *The Pros and Cons of Technology in the Classroom*. Paper presented at the Funder's Learning Community Meeting, Palo Alto, CA.
- Domin, D. (1999). A review of laboratory instruction styles. *Journal of Chemical Education*, 76(4), 543–547. doi:10.1021/ed076p543
- Duch, B. J. (1995). What is problem-based learning? *About Teaching: A newsletter of the Center for Teaching Effectiveness*, 47. Retrieved October 7 2013, from: <http://www.udel.edu/pbl/cte/jan95-what.html>
- Earle, R. S. (2002, January-February). The Integration of Instructional Technology: Promises and Challenges. *ET Magazine*, 42(1), 5–13. Available from: <http://BooksToRead.com/etp>
- Eguchi, A. (2007). Educational Robotics for Elementary School Classroom. In *Proceedings of Society for Information Technology & Teacher Education International Conference 2007* (pp. 2542–2549). Chesapeake, VA: AACE.
- Eteokleous, N. (2016). *Developing 21st Century Skills Through an Educational Robotics Curriculum*. Paper presented at the 2016 Annual Meeting of the American Educational Research Association, Washington, DC.
- Eteokleous, N., Demetriou, A. Y., & Stylianou, A. (2013). The Pedagogical Framework for Integrating Robotics as an Interdisciplinary Learning – Cognitive Tool. In J. Roselli & E. Gulick (Eds.), *Information and Communications Technology: New Research* (pp. 141–158). Nova Science Publishers, Inc.
- Gibbon, L. W. (2007). *Effects of Lego Mindstorms on convergent and divergent problem-solving and spatial abilities in fifth and sixth grades students* (Doctoral dissertation). Retrieved from: ProQuest Dissertation and Theses database.
- Grant, M. (2009, April). Understanding projects in project-based learning: A student's perspective. Paper presented at Annual Meeting of the American Educational Research Association, San Diego, CA.
- Griffin, P., & Care, E. (Eds.). (2015). *Assessment and Teaching of 21st Century Skills, Methods and Approach*. Dordrecht: Springer; doi:10.1007/978-94-017-9395-7
- Haugland, W. S. (2000). Early Childhood classrooms in the 21st century: Using Computers Maximize Learning. *Young Children*, 12–18.
- Jonassen, D. H. (1999a). Designing Constructivist Learning Environments. In C. Reigeluth (Ed.), *Instructional Design Theories and Models: A new paradigm of Instructional Theory* (pp. 215–239). Mahwah, NJ: Erlbaum.
- Jonassen, D. H. (1999b). *Computer as Mindtools in Schools: Engaging Critical Thinking* (2nd ed.). Columbus, OH: Prentice Hall.
- Karim, M., Lemaignan, S., & Mondada, F. (2015). A review: Can robots reshape K-12 STEM education? *Proceedings of the 2015 IEEE - International Workshop on Advanced Robotics and its Social impacts*. doi:10.1109/ARSO.2015.7428217
- Larmer, J., & Mergendoller, R. J. (2010). 7 Essentials for Project-Based Learning. *Educational Leadership*, 68(1).

- Markham, T. (2003). *Project-based learning handbook (2nd ed.)*. Novato, CA: Buck Institute for Education.
- Maykut, P., & Morehouse, R. (1994). *Beginning Qualitative Research, A Philosophic and Practical Guide*. London: The Falmer Press.
- Miglino, O., Lund, H. H., & Cardaci, M. (1999). Robotics as an Educational Tool. *Journal of Interactive Learning Research*, 10(1), 25–47.
- Mojica, K. D. (2010). *Ordered effects of technology education units on higher-order critical thinking skills of middle school students* (Doctoral dissertation). Retrieved from: ProQuest Dissertation and Theses database.
- Papert, S. (1980). *Mindstorms: Children, Computers, and Powerful Ideas*. New York, NY: Basic Books.
- Papert, S. (1993). *Mindstorms: children, computers, and powerful ideas* (2nd ed.). New York, NY: Basic Books.
- Resnick, M., Berg, R., & Eisenberg, M. (2000). Beyond black boxes: Bringing transparency and aesthetics back to scientific investigation. *Journal of the Learning Sciences*, 9(1), 7–30. doi:10.1207/s15327809jls0901_3
- Rotherham, J. A., & Willingham, D. T. (2010). “21st-Century” Skills: Not New, but a Worthy Challenge. *American Educator*, 34(1), 17–20.
- Salomon, G., Perkins, D. N., & Globerson, T. (1991). Partners in Cognition: Extending Human Intelligence with Intelligence Technologies. *Educational Researcher*, 20(3), 2–9. doi:10.3102/0013189X020003002
- Varnado, T. E. (2005). *The effects of a technological problem-solving activity on FIRST(TM) LEGO(TM) League participants’ problem solving style and performance* (Doctoral dissertation). Retrieved from: ProQuest Dissertation and Theses database.
- Vollstedt, A-M. (2005). *Using robotics to increase student knowledge and interest in science, technology, engineering and math* (Master’s thesis). Retrieved from: ProQuest Dissertation and Theses database.
- Ward, R., Miller, J. L., Sienkiewicz, R., & Antonucci, P. (2012). ITEAMS: Increasing the self-identification for girls and underserved youth in pursuing STEM careers. *Journal of Systemics, Cybernetics & Informatics*, 10(1), 95–99.
- Whitehead, S. H. (2010). *Relationship of robotic implementation on changes in middle school students’ beliefs and interest toward science, technology, engineering and mathematics* (Doctoral Dissertation). Retrieved from: ProQuest Dissertation and Theses database.
- Williams, D., Ma, Y., & Prejean, L. (2010). A Preliminary Study Exploring the Use of Fictional Narrative in Robotics Activities. *Journal of Computers in Mathematics and Science Teaching*, 29(1), 51–71.
- Yin, R. (2003). *Case study research: Design and Methods*. Thousand Oaks, CA: Sage Publications.

ADDITIONAL READING

- Alimisis, D., Arlegui, J., Fava, N., Frangou, S., Ionita, S., Menegatti, E., & Pina, A. et al. (2010). Introducing robotics to teachers and schools: experiences from the TERECoP project. In J. Clayson & I. Kalas (Eds.), *Proceedings for Constructionism 2010* (pp. 1–13). Paris: American University of Paris. Barker.
- Arlegui, J., Menegatti, E., Moro, M., & Pina, A. (2008). Robotics, Computer Science curricula and Interdisciplinary activities, In Proceedings of the TERECoP Workshop “Teaching with robotics, Conference SIMPAR 2008”, Venice.

Barker, B. S., & Ansorge, J. (2007). Robotics as means to increase achievement scores in an informal learning environment. *Journal of Research on Technology in Education*, 39(3), 229–243. doi:10.1080/15391523.2007.10782481

Benitti, F. B. V. (2012). Exploring the educational potential of robotics in schools: A systematic review. *Computers & Education*, 58(3), 978–988. doi:10.1016/j.compedu.2011.10.006

Bers, M. (2008a). *Blocks to Robots: Learning with Technology in the early childhood classroom*. New York: Teachers College Press.

Bers, M. (2008b). Engineers and Storytellers: Using robotics manipulatives to develop technological fluency in early-childhood. In O. Saracho & B. Spodek (Eds.), *Contemporary perspectives on science and technology in early childhood education* (pp. 105–125). Charlotte, NC: Information Age.

Bers, M., Matas, J., & Libman, N. (2013). Linvot ULehibanot, To Build and to Be Built: Making Robots in Kindergarten to Explore Jewish Identity. *Diaspora, Indigenous, and Minority Education: Studies of Migration, Integration, Equity, and Cultural Survival*, 7(2), 164–179. doi:10.1080/15595692.2013.787062

Bishop, A.P., Bertram, B.C., & Lunsford, K.J. & al. (2004). Supporting Community Inquiry with Digital Resources. *Journal of Digital Information*, 5(3).

Blikstein, P. (2013). Digital fabrication and 'making' in education: The democratization of invention. In J. WalterHerrmann & C. Bóching (eds.), *FabLabs: Of Machines, Makers and Inventors* (pp. 1–21). Bielefeld: Transcript Publishers.

Detsikas, N., & Alimisis, D. (2011). Status and Trends in Educational Robotics Worldwide with Special Consideration of Educational Experiences from Greek Schools. Proceedings of the International Conference on Informatics in Schools: Situation, Evolution and Perspectives (ISSEP, Comenius University, Bratislava, Slovakia.

diSessa, A. A. (2000). *Changing minds: computers, learning, and literacy*. Cambridge, MA: MIT Press.

Druin, A., & Hendler, J. (2000). *Robots for kids: exploring new technologies for learning experiences*. San Francisco: Morgan Kaufman/Academic Press. Lawhead.

Moro, M., & Alimisis, D. (2009). From the Logo Turtle to the Tiny Robot Turtle: practical and pedagogical issues. Proceedings of the 5th Pan-Hellenic Conference "ICT in Education, Syros (Greece).

KEY TERMS AND DEFINITIONS

BeeBots: A colorful, easy-to-operate, and friendly little robot that can be used by students from the age of four to the age of ten, depending on the level of programming difficulty of each exercise.

Educational Robotics: The use of robotics in the teaching and learning process (in the educational practice) as a subject matter and/or as a cognitive learning tool.

Floor Mat: A paper or foam based mat (mainly located on the floor or high-table) which is separated in blocks of 15X15 centimeters. Each floor mat represents a specific theme/ concept and each block depicts different sub-theme/ sub-concept related to the overall theme/ concept. Each floor mat may have from 9 to 24 blocks (or even more!) Floor mats can be developed for various subject matters.

Interdisciplinarity: It is the employment of various disciplines. The purpose is to relate/ combine two or more subject matters (or academic disciplines) into one exercise/ activity in achieving specific educational objectives.

Problem Based Learning (PBL): The students are involved in a learning process where they needs to solve a specific problem using various knowledge and skills. This approach can be used within various subject matters even.

Robotics Packages: In the education market there are various robotics packages that allow users/students to construct and program robots. These packages mainly contain bricks, wires, sensors and a visual programming software.

Twenty (21st) Century Skills: The skills that considered to be important and necessary for

student to develop in survive and succeed in the 21st century. Those are: problem solving, critical thinking, collaboration, communication, global citizenship, digital literacy, knowledge construction, creativity, innovation, self-directed learning.

From Digital Exclusion to Digital Inclusion for Adult Online Learners

E

Virginia E. Garland

University of New Hampshire, USA

INTRODUCTION

In this second decade of the new millennium, it is becoming clear that nations are pressuring their citizenry to become more technologically skilled in order for them to compete globally in the quest for political, economic, military and scientific success. Educational leaders are also aware of the need to provide more “tech savvy” students who can be successful academically, socially, and professionally in the international arena (Garland & Tadeja, 2013). Adult learners are taking a greater role in that process of becoming better academic and digital citizens. For instance, “Goal 2020” of the United States Department of Education is a higher education reform policy initiative focused on the United States leading “the world in the proportion of college graduates” with “at least 8 million additional adults [who] will need to return to college and earn associate and bachelor’s degrees by the year 2020” (Gast, 2013, p.17). But the goal of digital literacy for all the world’s people is hampered by many factors.

There are wide disparities in Information Science and Communication Technologies (ICT) skills between digitally excluded and digitally included online learners both nationally and internationally. There is an enormous and even widening gap between Internet users in developed versus developing nations (Pick & Azari, 2008). There are also digital exclusion issues in the United States and in other economically advanced countries. Secondary level students in the United States are increasingly taking online courses; but those public school students who live in low socio-economic areas, who speak a language other

than English, who are considered minorities, and who are disabled are more likely to lack academic technology skills (Garland, 2015).

In addition, a significant number of adult learners across the globe are a new group of “digitally excluded” students. Distance education courses are increasingly being taken for academic, professional, and personal reasons by students who are over twenty-five, yet many are struggling to compete with their “digital native” peers. According to one study, “Expanding adult digital literacy is essential for confronting vulnerable adults’ issues of exclusion and marginalization that are increasingly being amplified by the digital mediation of modern social life.” (Jacobs et al, 2014, p 626) Adult learners also need full inclusion in the technological demands of online education.

This chapter focuses on current trends in digital exclusion of adult learners and provides some solution strategies for ICT directors, higher education administrators, online instructors, and the older students they serve.

BACKGROUND

The most recently available statistical data indicate that online degree programs have a significant impact on adult learners. Over one fourth of all higher education students are taking online courses. According to the National Center for Educational Statistics (NCES) report, “Distance Education,” 25.8% of all undergraduate and graduate students were “enrolled exclusively” (12.5%) or “enrolled in some” (13.3%) distance education courses in 2012 (NCES, Distance Education,

para 3). NCES indicated in its “Back to school statistics” report that “In 2013 there were about 12.2 million college students under age 25 and 8.2 million students 25 years old and older” (NCES, Back to school statistics, para 16). By the year 2024, these enrollments are projected to almost double, with over 23 million “total fall enrollment in degree-granting postsecondary institutions,” the majority of which, over 13 million, are expected to be women (NCES, Digest of Education Statistics, table 303.40). In addition, the number of adult students taking non-degree online programs is also increasing. McCallum (2012) states that there are over 90 million students over 25 years old who are taking post-secondary studies in the United States alone. To put a human face on these statistics, consider the case of an online adult learner named Rosa.

In the fall of 2015, the author was introduced to Rosa (identifying information removed) in Texas at a national leadership conference, which focused on the accomplishments of undergraduate and graduate students. Rosa, whose primary language is Spanish, is the first in her family to go to college. Despite her academic success in being awarded a partial scholarship for her undergraduate studies, Rosa confided in me that she was struggling with financial and time management issues in trying to complete the online courses of her program. Rosa is a recently divorced, single mother of two young children. She spends most of her time working and taking care of her family. Rosa and her children have had to move back in with her parents. She does not have the resources to afford a smart phone and the latest computer upgrades. Although she is working part-time and saving money to make the tuition payments of her current degree program, Rosa feels that she might have to discontinue her undergraduate studies. It is likely that this academically talented adult student will drop out of college.

Digital exclusion of adult learners is associated with technology, socio-economic and gender factors. Sadly, Rosa fits the profile of many digitally excluded online adult learners. She is a mature

student, over the age of twenty-five, who does not have the requisite technology skills and tools to be able to easily access the Internet and to effectively use the information and communication digital tools prevalent in online courses. Rosa is also a non-traditional student, a single mother who was drawn to the flexibility in her schedule promised by the lure of distance education programs.

Why do adult students like Rosa want to take online courses? They find that online courses are appealing because of their “anytime, anywhere” access. Britt states “students desire more control over their education along with a flexible schedule in the learning process which is a distinct advantage of online education” (2015, p. 400). But some mature students are “digitally excluded,” unfamiliar with and perhaps unable to afford the newer technologies that are used as instructional tools in online courses (Garland, 2009).

Non-traditional students need not only technology training but also access to current digital tools in order to be successful online learners and more effective professionals. According to Gast (2013), the reasons for adult learners’ desire to take online programs include “the need for updated skills to compete in a knowledge based economy, a change in demographics due to immigration and higher retirement ages, technological advances bringing the classroom to the student, and a globalization of the higher education system. Most recently, online and for-profit institutions have been a primary beneficiary of this growing enrollment trend among adult learners” (p. 18). Indeed, online degree programs have almost doubled in a recent ten year period, “62.4% of colleges offered online degree programs at the end of 2012 which is up significantly from 32.5% in 2002” (Britt, 2015, p. 399).

Although the number of online degree programs at more traditional colleges in the United States is increasing, some previously programs in distance education universities are in decline. Two distance education giants, the University of Phoenix and Corinthian Colleges, have sharply decreasing enrollments and increasing public scru-

tiny. At the University of Phoenix, the largest for-profit university in the United States, “enrollment numbers have been cut in half from five years ago, down from 460,000 to 213,000...[and] Corinthian Colleges, another major for-profit system, closed its doors amid legal action and accusations of financial manipulation towards students” (Jackson, 2015, para 5 & 6). One reason for enrollment declines at the University of Phoenix is competition from other distance education programs, but another reason is the attrition of adult learners in online courses.

SOCIO-ECONOMIC, GENDER, AND TECHNOLOGICAL ISSUES

Despite the exponential growth of both distance education courses and the number of adult learners who are taking them, the dropout rate of students over twenty-five years of age is high (Park and Choi, 2009). Key factors that inhibit the academic success many adult learners in online courses are socio-economic, gender and technology based. These issues are helpful in understanding why certain groups of adults are especially underserved in their need for digital inclusion. They include the poor, mostly women, who have minimal literacy skills and speak English as a second language. The Literacy, Language, and Technology Research Group (LLTR) from Portland State University recently studied (Jacobs et al., 2014) the need for “adult digital literacy learning” in a technologically advanced “multiliterate world.” These researchers found that the “digital newcomer” is “economically vulnerable, under-served, and high-need adult population (i.e., low-income, low-literate, elderly, and English Speakers of Other Languages) who are new to, or have limited experience with, technology” (p. 625).

Constraints on the Adult Learner

If working adult learners, many of whom are single mothers like Rosa, had more time to take

affordable online courses, they would be less likely to drop out. Based at one of the world’s largest online universities, University of Phoenix, faculty members Donnelly and Kovacich (2014), studied adult student attrition in community college based online courses in the Boston, Massachusetts area. The adult students of their study have the same characteristics of those researched by the LLTR group in Portland, Oregon. Donnelly and Kovacich found that poor, working women with children are likely to fail or drop out, stating that “Over 80% of community college students work as well as attend school, are raising children, and care for relatives... attrition rates for online courses at community colleges are 15% to 50% higher than the same courses provided in face-to-face environments” (p. 35). In a study of adult learners in Hellenic Open University, Greece’s “only university offering exclusively distance education courses,” Angelaki and Mavroidas (2013) found that women were more likely to experience “negative emotions... mainly associated to the unknown for students [sic] methodology of distance learning, the requirements of the course, the familial, professional and social responsibilities of the students as well as to the difficulties stemming from using technological tools” (p. 88).

Online adult learners who determine that their online courses are a “waste of time and money” are more likely to drop out. Gast (2013) contends that fiscal and time management factors are key in understanding why adult learners fail to complete online courses. The latter researcher studied economically disadvantaged, mature students, and found that “Time and finances are cited as the most common barriers faced by adult students.” (p. 18). Even if adult students can afford their online programs, they might not have the time needed to complete the required assignments. Romero and Barbera (2011) found that there are time constraints for the older learner who has more work and family responsibilities than his or her younger peer. Similarly, Herbold (2012) states that “time” is valued more by the “non-traditional, older students...With restrictions on

their time due to life's other responsibilities, adult learners are better able to successfully complete a course of study if the conditions for learning are optimized" (pp. 117-118). Solutions for delivering effective online instructional strategies that make the best use of the adult learner's limited time and financial resources are provided in the "Solutions and Recommendations" section of this chapter.

Technology Skills and the Adult Learner

With access to effective, wireless technology tools that deliver a curriculum that is relevant, the adult student has a greater chance of success in online courses (Garland 2015). In her study of 69 graduate students in two online education courses, Herbold (2012) also found that "Students could access the course from any location they had Internet access, which could be anywhere with a portable smart device. The place could also be a fixed location such as a desk computer or laptop at home, work, or public places with Wi-Fi like the library, coffee shop, hotel or truck stop. These mature students can access the classroom from almost anywhere their busy and involved lives take them" (p.122).

However, even if adult online learners have Internet access tools, such as the "portable smart device" described by Herbold (2012), they might still be inexperienced in "digital literacy" skills. Many people born before the "iGeneration" have smart phones, but they are more likely to use them for communication and social media purposes than as tools in the academic assignments of online courses (Garland, 2009). According to Jacobs et al (2014), "engagement with Web 2.0 involves a new ethos or way of understanding one's relationship to knowledge construction...[The] dichotomy of digital natives and digital immigrants has been found to be overly simplistic and in need of more nuanced understandings of what it means to be able to use the computer and Internet" (p. 625). Thus, a discussion of the term "digital inclusion" must include not only access to the Internet, but also the technology skills needed for the adult learner to be fully engaged in the online course.

SOLUTIONS AND RECOMMENDATIONS

These are the major intervention strategies needed to ensure success for the adult online learner:

- Technology training and support for adult online learners in the use of the digital tools used in distance education programs
- Organizational and time management skill training for adult students to become more self-directed learners
- Curriculum design and teaching methods training for online instructors and academic coordinators
- Increased financial assistance opportunities for those most in need, the at risk, non-traditional student
- Expanded and affordable childcare programs, especially for working mothers who are also online learners

In addition to the policy initiatives at the international and national levels to make online courses more affordable and accessible to adult learners, there should be local efforts to engage the mature student. Institutions of higher education must develop and implement policies that better ensure the success of adult learners, whose enrollment projections, particularly those for women, are greater than those of their younger counterparts.

Attrition is decreased if colleges and universities support the acquisition of academic and technology skills that enable the mature student to be successful. The online course curriculum should also be meaningful to the adult learner. According to Park and Choi (2009), adult learners were less likely to drop out of online courses if they had "organizational support" and that they viewed the course to be relevant to their personal and professional lives. ICT directors and university administrators should be aware that "Adult students have many commitments such as family and work in addition to school and must create the time and energy to manage them all. Schools need to understand how outside factors can hinder

the adult learning process and take steps to fully understand the plight of adult learners” (Donnelly and Kovacich, 2014, p. 40). Adult student advising programs and financial assistance opportunities would benefit the mature learner. But the adult learner must also be engaged in self-directed learning.

Stine (2010) found these three factors as key to the success of adult learners in distance education: high levels of student skills in technology, academics, and individual learning. One model program for preparing adult learners for online courses is the Learner Web project of Portland State University for digitally excluded, “high need” adults, “designed to give learners a self-directed learning experience...the shifts learners undergo as they gain experience and confidence with digital tools can help educators develop more robust systems for supporting vulnerable populations” (Jacobs, et al, 2014). Even non-vulnerable adult learners need to be self-directing in their online coursework. According to Jacobs’ (2012) study of online graduate students, “to address the adult learning characteristics of being autonomous, self-directing and self-responsible, students were given the latitude to select the activities they preferred and that would best meet their individual needs” (p. 122). Thus, self-directed learning is an approach that is valuable in motivating the older online learner.

The adult student who takes more control of his or her own online learning would also benefit from effective time management skills. In their study of community college students who dropped online courses, Donnelly and Kovacich (2014) found that those who had higher attrition rates also had little time to focus on studying because of other tasks, such as work and family obligations. In addition to suggesting that some online adult learners study more effectively in the morning, the University of Phoenix faculty members, Donnelly and Kovacich (2014), along with Romero and Barbera (2011), recommend that more interactive course design features would also enhance student engagement in synchronous online learning activities. Adult

students who were interviewed by Donnelly and Kovacich (2014) in the community college online courses study stated that they did not drop out of online courses if they were effective in organizing their course schedules and assignment deadlines.

Mature students need other motivational tools, such as having achievable goals and knowing that the instructor is meeting their needs for collaboration and evaluative feedback. Adult students are more goals oriented than their younger peers. Ekmekci (2013) found that adult students will find online courses “a waste of time” unless they are taught by knowledgeable instructors with clearly stated goals that are relevant to students’ professional lives. The goals of online courses, such as the completion of multimedia group projects, are more easily attainable with formative assessment strategies.

According to Leong (2011), other instructional strategies to enhance the success of online adult learners include the instructors’ use of both collaborative activities to improve “social presence” and immediate feedback on assignments to increase “cognitive absorption” of the content being taught. Ke (2010) believes that there are three types of “presence” needed in the online course: cognitive, social and instructional. Online dialogues and prompt, effective feedback by the online instructor are viewed by Ke as ways to increase these three types of “presence.” The need for a “social presence” between instructor and student, and often between student and student, cannot be underestimated. In their study of distance education programs for adult learners in Greece, Angelaki and Mavroidis explain that “The expanded use of new tools related to Information and Communication Technologies (ICT) facilitates communication, cooperation and dialogue among participants in distance learning courses” (2013, p. 78). In addition to the value of social presence is the importance of cognitive presence. Leong’s and Ke’s findings are similar to that of Tyler-Smith (2006), who argues that “cognitive overload” is the most significant reason that the adult “eLearners” he studied in New Zealand

dropped out within the first weeks of the online courses they were taking.

The adult learner also needs to be able to afford the tuition of distance education degree programs. Despite enormous national investment in higher education, adult learners in the United States are those with the largest student debt. According to The Hechinger Report (2015), “The reason that the ‘average’ student’s debt is so high – almost \$23,000 – is because the figure includes the loans of graduate students, who are permitted to borrow unlimited amounts from the federal government up to the cost of attendance. Sixty-five percent of 2012 graduates who borrowed \$50,000 or more were graduate students” (para 5). In addition, Gast (2013) found that student loan debt is a significant inhibitor to adult student enrollment.

The single mother who is at risk of dropping out needs not only financial aid but also improved social and family support. In a study for the Institute for Women’s Policy Research, Gault, Reichlin, and Roman (2014) recommend that federal and state student financial aid programs apportion more assistance to low-income adult learners. These researchers also advocate “more inclusive campus policies,” to meet the needs of single mothers, such as increased child care subsidies and support as well as “single-parent housing.”

FUTURE RESEARCH DIRECTIONS

The attrition of adult learners in higher education online courses is an emerging field of research. Longitudinal studies of distance education programs and the mature students they serve are needed. Political science, adult motivational and women’s studies theorists can inform ICT and educational practitioners on the reasons for the alarmingly high dropout rates for adult students in distance education programs in community colleges, which have received an infusion of federal assistance in the past few years.

In the United States, policy makers and politicians are making a wide variety of recommenda-

tions to alleviate the alarming increase of student debt. One policy highly criticized is the 2010 Health Care and Education Reconciliation Act (a section of which is the Patient Protection and Affordable Care Act or “Obama Care”). Although this legislation invested two billion dollars in community colleges and “career training, student attrition and student debt have reached all time highs. Political scientists and educators need to further study the projected effectiveness of financial assistance or student debt relief programs.

Adult motivation theories also shed further light on the ways to ensure the success of non-traditional students in online courses. In a study of 203 university students, Sogunro (2015) found eight factors that motivate adult learners in higher education: effective instructional methods, solid curriculum, course relevance, interactive classroom, formative and prompt assessment, student self-directedness, positive learning environment, and meaningful academic advising. All of these factors can apply to the online course; especially those focused on curriculum design, interactive technology tools and feedback methods. And, adult learners particularly need a curriculum that is relevant and advising that is supportive. More research should be conducted on the application of adult motivational factors to the design and implementation of online courses in order to reduce the high rate of adult learner attrition.

Another future research direction is that of studying factors which affect the academic success rate of women, particularly single mothers, in online courses. The Institute for Women’s Policy Research has completed valuable research on the topic of college affordability for low-income adult learners. The work of this policy group should be supported and continued.

CONCLUSION

In the political arena, debate continues on the affordability of higher education in the United States. For students in many other economically

advanced countries, such those discussed in this chapter, New Zealand and Greece, national legislation results in the reduction of much of the cost of public higher education. But in the United States, policies on the national and institutional levels are especially needed to support the efforts of economically disadvantaged adult students who are at greatest risk of not completing their educational programs. Rising levels of adult student debt for online courses relate to the increasing attrition of mature learners in distance education programs in the United States. Although there has been recent litigation in California on the mismanagement of online student tuition and recruitment practices (Jackson, 2015), much more needs to be done to ensure that institutions of higher education which offer distance education programs are successful in meeting the academic and financial needs of all students.

Colleges and universities which offer online courses should act more honestly and effectively at not only fair practices in student recruitment and in financial assistance, but also in focusing those policies on accommodating the needs of adult students. Prospective students who are over twenty-five, many of whom are working single mothers, desire to further their education through online courses because they can conveniently access them anytime, anywhere (Garland, 2015). But, these underserved adults are among the most likely to drop out of online courses (Donnelly and Kovacich, 2014). According to Gast (2013), institutions should be marketing their distance education courses with adults in mind, supporting non-traditional students with specialized services, preparing mature students for post-degree or post-certificate professional work, and making their tuition more affordable. Higher education leaders should be more proactive in developing tutoring programs, offering financial and childcare services, and providing advisors who are sensitive to the needs of prospective or enrolled adult students.

Online instructors themselves can do much to design and implement online courses that are

meaningful to the adult learner. The collaborative efforts of both online teachers and their motivated, technology trained students are needed for adult learners like Rosa to transition from “digital exclusion” to “digital inclusion.”

REFERENCES

- Angelaki, C., & Mavroidis, I. (2013). Communication and social presence: The impact on adult learners' emotions in distance learning. *European Journal of Open Distance and E-Learning*, 16(1), 78–93.
- Britt, M. (2015). How to better engage online students with online strategies. *College Student Journal*, 49(3), 399–404.
- Donnelly, W., & Kovacich, J. (2014). A phenomenological investigation of the problem of adult student attrition in community college online courses. *Exchange*, 3(1), 34–43.
- Ekmekci, O. (2013). Being there: Establishing instructor presence in an online learning environment. *Higher Education Studies*, 3(1), 29–38. doi:10.5539/hes.v3n1p29
- Garland, V. E. (2009). Wireless technologies and multimedia literacies for K-12 education. In L. Hin & H. Subramaniam (Eds.), *Handbook of Research on New Media Literacy at the K-12 Level*. IGI Global.
- Garland, V. E. (2015). The digitally excluded learner: Strategies for success. In M. Khosrow-Pour (Ed.), *Encyclopedia of Information Science and Technology* (3rd ed.; pp. 2400–2408). Hershey, PA: Information Science Reference. doi:10.4018/978-1-4666-5888-2.ch233
- Garland, V. E., & Tadeja, C. (2013). *Educational leadership and technology: Preparing administrators for a digital age*. New York: Taylor and Francis.

- Gast, A. (2013). Current trends in adult degree programs: How public universities respond to the needs of adult learners. *New Directions for Adult and Continuing Education*, 2013(140), 17–25. doi:10.1002/ace.20070
- Gault, B., Reichlin, L., & Roman, S. (2014). *College affordability for low-income adults: Improving returns on investment for families and society. Report#C412*. Washington, DC: Women's Policy Research.
- Herbold, K. (2012). Giving student choice in online learning environments: Addressing Adult Learner Needs. *The International Journal of Technology Knowledge in Society*, 7(5), 117–125.
- Jackson, A. (2015). It looks like this for-profit college could be in financial trouble. *Business Insider*. Retrieved November 11, 2015 from: <http://www.businessinsider.com/university-of-phoenix-enrollment-numbers-are-down-2015-3>
- Jacobs, G. E., Castek, J., Pizzolato, A., Reder, S., & Pendell, K. (2014). Production and consumption: A closer look at adult digital literacy acquisition. *Journal of Adolescent & Adult Literacy*, 57(8), 624–627. doi:10.1002/jaal.293
- Ke, F. (2010). Examining online teaching, cognitive, and social presence for adult students. *Computers & Education*, 55(2), 808–820. doi:10.1016/j.compedu.2010.03.013
- Leong, P. (2011). Role of social presence and cognitive absorption in online learning environments. *Distance Education*, 32(1), 5–8. doi:10.1080/01587919.2011.565495
- McCallum, C. (2012). *The perception of adult learners concerning their satisfaction of their educational experiences in a midwestern community college* (Unpublished doctoral dissertation). Western Michigan University. Retrieved 1-3-16 from ProQuest Dissertations & Theses database.
- National Center for Educational Statistics. (n.d.a). *Back to school statistics*. Retrieved from <http://nces.ed.gov/fastfacts/display.asp?id=372>
- National Center for Educational Statistics. (n.d.b). *Digest of educational statistics*. Retrieved from https://nces.ed.gov/programs/digest/d14/tables/dt14_303.40.asp
- National Center for Educational Statistics. (n.d.c). *Distance learning*. Retrieved from <https://nces.ed.gov/fastfacts/display.asp?id=80>
- Park, J., & Choi, H. (2009). Factors influencing adult learners' decision to drop out or persist in online learning. *Journal of Educational Technology & Society*, 12(4), 207–217.
- Pick, J. B., & Azari, R. (2008). Global digital divide: Influence of socioeconomic, governmental, and accessibility factors on information technology. *Information Technology for Development*, 14(2), 91–115. doi:10.1002/itdj.20095
- Romero, M., & Barbera, E. (2011). Quality of learners time and learning performance beyond quantitative time-on-task. *International Review of Research in Open and Distance Learning*, 12(5), 125–137. doi:10.19173/irrodl.v12i5.999
- Sogunro, O. A. (2015). Motivating factors for adult learners in higher education. *International Journal of Higher Education*, 4(1), 22–37.
- Stine, L. (2010). Basically unheard: Developmental writers and the conversation on online learning. *Teaching English in the Two-Year College*, 38(2), 132–148.
- The Hechinger Report. (2015). *Heaviest Debt Burdens Fall On 3 Types of Students*. Retrieved from: <http://www.usnews.com/news/articles/2015/06/08/heaviest-college-debt-burdens-fall-on-3-types-of-students>

Tyler-Smith, K. (2006). Early attrition among first time eLearner: A review of factors that contribute to drop-out, withdrawal and non-completion rates of adult learners undertaking eLearning programmes. *MERLOT Journal of Online Learning and Teaching*, 2(2), 73–85.

ADDITIONAL READING

Calvin, J., & Winfrey-Freeburgh. (2010). Exploring adult learners' perceptions of technology competence and retention in web-based courses. *The Quarterly Review of Distance education*, 11(2).

Chang S., Shieh R., Liu E. & Yu, P. (2012). Factors influencing women's attitudes towards computers in a computer literacy training program. *Turkish Online journal of Educational Technology* 11(4), 177-187.

Cox, M. J., Niederhauser, D. S., Castillo, N., McDougall, A. B., Sakamoto, T., & Roesvik, S. (2013). Researching IT in education. *Journal of Computer Assisted Learning*, 29(5), 474–486. doi:10.1111/jcal.12035

Epstein, D., Nisbet, E., & Gillespie, T. (2011). Whos responsible for the digital divide? Public perceptions and policy implications. *The Information Society*, 27(2), 92–104. doi:10.1080/01972243.2011.548695

Freeman, B. (2012). Using digital technologies to redress inequities for English language learners in the English speaking mathematics classroom. *Computers & Education*, 59(1), 50–62. doi:10.1016/j.compedu.2011.11.003

Garland, V. E. (2010). Emerging technology trends and ethical practices for the school principal. *Journal of Educational Technology Systems*, 38(1), 39–50. doi:10.2190/ET.38.1.e

Lee, Y., Choi, J., & Kim, T. (2012). Discriminating factors between completers of and dropouts from online learning courses. *British Journal of Educational Technology*, 44(2), 328–337. doi:10.1111/j.1467-8535.2012.01306.x

Mok, K. H., & Leung, D. (2012). Digitalisation, educational and social development in Greater China. *Globalisation, Societies and Education*, 10(3), 271–294. doi:10.1080/14767724.2012.710118

Nartgun, S. (2011). Relationships between open education students' economic profiles and their use of the Internet in education. *Turkish Online Journal of Distance Education*, 12(4), 179–200.

Poellhuber, B., Anderson, T., & Roy, N. (2011). Distance Students readiness for social media and collaboration. *International Review of Research in Open and Distance Learning*, 12(6), 102–125. doi:10.19173/irrodl.v12i6.1018

Rovai, A. P. (2003). In search of higher persistence rates in distance education online programs. *The Internet and Higher Education*, 6(1), 1–16. doi:10.1016/S1096-7516(02)00158-6

Yoo, S. J., Huang, W. D., & Wenhao, D. (2013). Engaging online adult learners in higher education: Motivational factors impacted by gender, age, and prior experiences. *Journal of Continuing Higher Education*, 61(3), 151–164. doi:10.1080/07377363.2013.836823

From Digital Natives to Student Experiences With Technology

Sue Bennett

University of Wollongong, Australia

Linda Corrin

University of Melbourne, Australia

INTRODUCTION

The term ‘digital native’ was popularized by Prensky (2001) as a means to distinguish young people who were highly technologically literate and engaged. A ‘digital native’ can be defined as an individual who has grown up immersed in digital technology and is technologically adept and interested. The digital native is described in direct contrast to the ‘digital immigrant’, who having been exposed to digital technology later in life is fearful of it, mistrustful and lacks the skills to use technology adeptly. According to Prensky’s (2001) vision, all young people who have grown up since the widespread advent of the personal computer can be considered digital natives, and, by elimination, all older people are digital immigrants.

It is argued that the existence of the digital native makes dramatic educational reforms necessary because traditional education systems do not, and can not, cater for the needs and interests of young people. As a result, outdated schools and universities and outmoded teaching simply alienate students from learning, leaving them disengaged and disenchanted by education’s alleged failure to adapt to the new digital world. By implication, education must be transformed by technology, coupled with new pedagogies. Although this argument is a familiar one to those acquainted with the broader educational technology literature, the digital native hypothesis provides a new basis

for claims for revolutionary educational change through technology integration.

Recent research has revealed that the term is misapplied when used to generalized about an entire generation, and instead indicates that only a small sub-set of the population fits this characterization. This research shows significant diversity in the technology skills, knowledge and interest of young people, and suggests that there are important ‘digital divides’ which are ignored by the digital native concept.

This chapter charts the development of the digital native idea and the debate that has surrounded it. It provides an account of the research and conceptual work it has stimulated, and suggests future directions research may take in the coming decades.

BACKGROUND

The idea of the digital native appears to have first emerged in an essay entitled *Declaration of the Independence of Cyberspace* by Barlow (1995) in which he admonished parents with the charge: “You are terrified of your own children, since they are natives in a world where you will always be immigrants” (p.12). Papert (1996), in *The Connected Family*, similarly evokes a rift between parents and children, and teachers and students, portraying older generations as being both afraid of computers and technically incompetent. Clearly,

the idea of a digital generation gap was gaining currency at this time.

Regardless of its exact provenance, it has been Prensky who popularized the term 'digital native' in his widely cited 2001 article, *Digital Natives, Digital Immigrants*. Around the same time, Tapscott (1998) had put forward the similar notion of 'the Net Generation', while social commentators coined the term 'Millennials' as a generational label (Howe & Strauss, 2000). Since then a proliferation of less widely used epithets has appeared, all attempting to capture the essence of the same phenomenon (e.g., Generation C, Google Generation, Nintendo Generation, etc.).

In short, the idea of the digital native captured the imaginations of teachers, parents, journalists, commentators and academics. Closer examination of Prensky's arguments, particularly in his influential 2001 paper, reveals little in way of evidence to substantiate his claims, however. He relies on anecdotes, conjecture and speculation. Nonetheless his ideas have often been uncritically repeated and cited as if fact. Similar arguments purportedly based on evidence provide few details of the data collection methods and analysis processes, thwarting critical scrutiny of these studies (e.g., Tapscott, 1998; Palfrey & Gasser, 2008). This presents a significant challenge in assessing the quality of this research.

It was a few years after Prensky's 2001 paper before researchers began to seriously address his claims, apparently galvanized by dissatisfaction with his arguments. Since that time a significant body of international research has largely debunked the idea of a uniformly technically savvy generation. Instead it suggests that the label 'digital native' likely only applies to a small minority of the population. Of much greater interest is the wide diversity of technology use uncovered by this research. These differences are often thought of as 'digital divides' because they highlight significant gaps between the ways individuals and/or communities engage with technology. These gaps present an ongoing challenge to those con-

cerned with equity and justice in education, and in society more broadly.

More recently there have been attempts to redefine and rehabilitate the term 'digital native'. In fact, this emerged in Dede's (2005) argument that aptitude with technology is not necessarily related to age but to other personal characteristics. In recent years Prensky (2009) has also seemed to resile from his earlier sharp distinctions, praising rather than criticizing the role of the teacher. Nevertheless the original divisive idea remains potent.

In the next section we turn to examine some of the research evidence that has emerged in response to the idea of the digital native.

RESEARCHING 'DIGITAL NATIVES'

Researching Technology Use

In the mid 2000s researchers began to investigate some of Prensky's key claims about digital natives. The initial area of focus was on determining whether, in fact, digital technologies were as extensively used within younger generations of the population as was supposed by the digital native thesis (e.g. Kennedy, Krause, Judd, Churchward & Gray, 2006; Kvavik, Caruso & Morgan, 2004; Oliver & Goerke, 2007). These studies set about to establish the extent of access to and ownership of a wide range of technologies, and to discover the extent to which they were used for particular activities. In short, researchers wanted to know who was using what technology, how often and for what purposes. Similar research had already been conducted, for example through studies of children's use of technology in and out of school (e.g., Downes, 2002; Kent & Facer, 2004; Kerauwalla & Crook, 2002), but these studies were not specifically driven by the digital native concept. Related work was also being conducted in disciplines outside of education, such as youth studies, cultural studies and media studies, but again these did not relate to the digital native idea (e.g.,

Livingstone & Helsper, 2007; Selwyn, 2003). These studies do, however, suggest that there was a broader appeal to research along these lines.

Early ‘digital natives’ studies tended to use survey methods to collect data from large populations, often of higher education students. In this exploratory work researchers attempted to gain a broad perspective by collecting data from participants who are relatively easy to access with a focus on phenomena relatively easy to measure through self-report (e.g., Kennedy et al, 2006; Kvavik, Caruso & Morgan, 2004). While questions about access to technologies and frequency of use are common features of these studies, many have gone further to gauge skills, interests and preferences, have included multiple age ranges rather than only younger people, and in some cases incorporated qualitative methods to complement quantitative data. One of the most notable surveys has been the ECAR series in the United States, which has run since 2004 with consistently large sample sizes of college students (see Dahlstrom, Brooks, Grajek & Reeves, 2015 for the latest report). Similar studies from around the world have contributed to a developing understanding of technology use, particularly among young people (e.g., Jones, Ramanaua, Cross & Healing, 2010; Kennedy et al., 2009; Oliver & Goerke, 2007; Corrin, Lockyer & Bennett, 2010; Margaryan, Littlejohn & Vojt, 2011; Thompson, 2013; Teo, 2015; Akçayır, Dündar & Akçayır, 2016; Henderson, Finger & Selwyn, 2016; Šorgo, Bartol, Dolničar & Boh Podgornik, 2016).

In sum, the main findings of these studies have been as follows:

1. There is near universal adoption of some technologies (e.g., mobile phones).
2. Some technologies have not been widely adopted, for example, RSS feeds and some forms of social media. The reasons for this are not clear, however. Perhaps, some technologies are too specialized, overly technical, or judged to be less useful.
3. There are indicators of some differences due to age, gender, socio-economic background, year level of study, and discipline of study (at university or college), although findings are not consistent across all studies or all technologies.
4. The studies trace how some technologies are abandoned, for example, because they are superseded in favor of alternatives (e.g., the demise of MySpace and the rise of Facebook, and the shift from dial-up to broadband Internet access).
5. Skills, knowledge and interests are highly varied when comparing individuals. Findings suggest that individuals adapt their technology use to suit their needs and interests and the contexts they engage in.
6. Younger people often have lower skill and knowledge levels than what might be expected based on the digital native hypothesis.

A common conclusion from these studies is that, while there appear to be some age-related factors, diversity is often higher within age groups than between them. It is also important to note that while large-scale survey studies can indicate patterns, the measures used are relatively crude and their accuracy is limited by participants’ abilities to recall and estimate their usage. There is a need for qualitative studies that are capable of exploring technology use in greater depth and with sensitivity to individuals’ contexts. There are also, to date, few studies from developing countries and of less affluent communities, making the global situation difficult to discern.

In short, the research conducted thus far suggests that only a small minority of the population can be considered ‘digital natives’, even disregarding age as a factor to include technologically adept older people. People adopt technologies for a wide range of reasons and have diverse patterns and habits, and the skills they develop are often narrow and highly contextualized (i.e., fit for a particular purpose). As a result, it would be wrong

to generalize about a section of a population on the basis of how they use technology, and in particular on the basis of presumed exposure to technology.

Implications for Education

Prensky (2001) posed the problem for contemporary education as follows: “Our students have changed radically. Today’s students are no longer the people our educational system was designed to teach” (p. 1). This pronouncement was based on the assumption that all young people were digital natives being held back by an outdated education system. If, however, not all young people are digital natives, only some, and there is significant diversity within the population with regard to technological prowess, then the problem for education is somewhat different. The challenge of how education can cater appropriately for learners remains, but it is made more complicated by the fact that learners comprise a diverse rather than homogenous group. A further challenge for public education is that if some students are disadvantaged by virtue of their socio-economic situations, then how can an inclusive education system address that disadvantage?

Concerns about a digital divide between the ‘haves’ and ‘have nots’ first emerged in relation to differences in access to technology (Warschauer, 2004). As technology became cheaper and easier for ordinary people to obtain, the focus shifted to differences in the skills and knowledge people have to make effective use of technology (Selwyn, 2004; Warschauer, 2004). And as ideas about what it means to be digitally literate have changed, this has seen a move away from a focus on developing people’s technical skills to a focus on developing their capacities to use technology responsibly, creatively and innovatively. This poses questions for education about how students can be equipped with these more sophisticated skills and understandings.

The infusion of digital technologies into everyday life has also raised questions about the relation between technology in education and out,

particularly amongst those who speculate about how the high levels of motivation exhibited by young people while gaming or socializing online might be employed in learning (Prensky, 2001; Tapscott, 1999). This reflects a wider conversation about how Web 2.0 technologies might be integrated into education, and warnings that their application might not be straightforward because of fundamental differences between informal learning and formal educational contexts (Dohn, 2009).

These discussions indicate that while the original digital native hypothesis is not a sound basis for recommending or planning educational change, differences in the ways technologies are used and their increasing prevalence in society continue to raise important questions for education. These are questions that need to be informed not only by empirical evidence gleaned from further research studies, but also by theories that help us to explain the phenomena and thereby better understand it.

Theoretical Perspectives

Just as the original proposal of the digital native lacked empirical foundations, theoretical underpinnings were also absent. However, as the research agenda has developed, casting doubt on the general nature of the claims and in doing so revealing people’s diverse engagements with technology, researchers have begun to conceptualize both the nature of the debate itself and to propose theoretical constructs that might help to explain the phenomena and frame future investigations.

The debate itself has been described as an academic form of a ‘moral panic’, a concept widely used in the social sciences (Bennett, Maton & Kervin, 2008). A moral panic, as described by Cohen (1972), occurs when a particular group is seen as a threat to societal norms. Importantly, the concern inspired exceeds the supporting evidence. Thus, the lack of evidence base and the extreme language used in arguments for the existence and importance of digital natives is consistent with a

moral panic. This characterization is useful because it helps to explain how the idea gained such prominence on the basis of flimsy evidence. It also explains how the form for the debate stymied genuine academic discussion until the emergence of empirical research.

More recently, researchers have proposed that this empirical evidence provide the stimulus for developing more sophisticated ways of thinking about and researching people's technology use (e.g., Bennett & Maton, 2010). Drawing on a range of sociological theories, these authors argue that concepts related to social networks (Castells, 2001; Wellman, 2002), social practices (Bourdieu, 1990) and the nature of knowledge and education (Bernstein, 1999) are critical to advancing understanding in this area.

SOLUTIONS AND RECOMMENDATIONS

Although research on young people's use of technology has consistently found diversity in areas such as access, use and ability with technology, there are several ways educators can address these differences within educational environments. In acknowledging that diversity exists, educators and institutions should be encouraged to investigate the profile of students in their own context. Such understandings can inform effective design of technology-based activities and how best to articulate these to students. This includes making clear the purpose and value of using particular technologies in the context of learning activities (So, Choi, Lim & Xiong, 2012). At an institutional level, caution should be exercised when setting strategy and targets around levels of use of technology to avoid the use of technology for technology's sake.

While the digital native concept emphasizes the high levels of engagement with technology in young people's everyday lives, research has revealed much lower levels of adaption of these technologies to support learning. The technology

choices of young people to support their studies tend to focus on pragmatic functions like speed, convenience, and efficiency (Henderson, Selwyn, Finger & Aston, 2015; Thompson, 2013). Questions have also been raised about whether young people have the requisite pedagogical knowledge to inform innovative technology choices to support their learning (Margaryan et al., 2011). Consequently, assumptions cannot be made that young people will have the knowledge and skills to be able to adapt and use a wide range of technology in an educational context. As a result, educators have an important role to play in influencing and stimulating young people's engagement with technology (Henderson, Finger & Selwyn, 2016). In order to do this, educators need support to develop their own digital literacy (Ng, 2012).

Greater attention regarding enhancing digital literacy for young people in the context of education is required. Despite the increasing trend in young people's exposure to and ownership of technology, ownership of technological devices alone does not directly impact levels of digital literacy (Šorgo, Bartol, Dolničar & Boh Podgornik, 2016). Recommendations from recent research consistently call for increased support for digital literacy in learning environments broadly, but also specifically within the context of particular learning activities (Buzzard et al., 2011; Toliver, 2011; Thompson, 2013). The form, timing, and extent of this support needs to be customized to the specific context, but also be flexible enough to cater for the diverse experiences and skills of the audience.

FUTURE RESEARCH DIRECTIONS

Future research into people's technology uses and choices will continue to monitor new developments, sparked by emerging technologies and changing patterns of adoption and use. In the short term, one focus will be on the impact of Web 2.0 technologies and their proposed capacity for democratizing participation in technology-based

activities. More generally, the trend towards greater online connectivity through new services and devices will continue, and so pose further questions for researchers about digital divides and digital inclusion across societies. The role technology plays in supporting learning that happens in formal and informal educational contexts is another area of research that will provide insights that can inform design of learning activities and technology support.

Future research will also require a commitment to developing more sophisticated understandings of technology use and choice. As noted above, in-depth qualitative research will be needed to provide insights into the diversity uncovered by recent surveys. Longitudinal studies that observe students' technology choices over time are also needed to understand the factors that influence changes in interaction and transitions between technologies (Corrin, Bennett & Lockyer, 2013; Henderson, Finger & Selwyn, 2016). Findings from this work will enable the field to transcend simplistic labels and thereby truly account for the rich array of activities and practices with technology. These are developments that can underpin discussions about what role technology can and should play in education such that the best learning outcomes can be achieved for all students.

CONCLUSION

To conclude, although misguided in its attempt to characterize a whole generation of young people, the idea of the digital native has been helpful in drawing educators' and researchers' attention to the under-researched area of young people's technological experiences and preferences. It has stimulated a very productive and promising avenue for educational technology research that has the potential to lead to better informed decision-making about technology and to improved teaching and learning. The challenge for educators moving forward is to develop ways to

make effective use of technology in teaching and learning while acknowledging and addressing issues around students' diversity in technology use and ability. This requires sophisticated thinking and careful consideration (Bennett & Maton, 2011) of the role of technology in teaching and learning that moves beyond a reliance on simplistic, generational labelling.

REFERENCES

- Akçayır, M., Dündar, H., & Akçayır, G. (2016). What makes you a digital native? Is it enough to be born after 1980? *Computers in Human Behavior*, 60, 435–440. doi:10.1016/j.chb.2016.02.089
- Barlow, J. (1996). *Declaration of the Independence of Cyberspace*. Retrieved July 30, 2011 from <http://homes.eff.org/~barlow/Declaration-Final.html>
- Bennett, S., & Maton, K. (2010). Beyond the digital natives debate: Towards a more nuanced understanding of students technology experiences. *Journal of Computer Assisted Learning*, 26(5), 321–331. doi:10.1111/j.1365-2729.2010.00360.x
- Bennett, S., & Maton, K. (2011). Intellectual field or faith-based religion: Moving on from the idea of 'digital natives. In M. Thomas (Ed.), *Deconstructing digital natives: young people, technology and the new literacies* (pp. 169–185). New York: Routledge.
- Bennett, S., Maton, K., & Kervin, L. (2008). The digital natives debate: A critical review of the evidence. *British Journal of Educational Technology*, 39(5), 775–786. doi:10.1111/j.1467-8535.2007.00793.x
- Bernstein, B. (1999). Vertical and horizontal discourse: An essay. *British Journal of Sociology of Education*, 20(2), 157–173. doi:10.1080/01425699995380
- Bourdieu, P. (1990). *The Logic of Practice*. Cambridge, UK: Polity.

- Buzzard, C., Crittenden, V. L., Crittenden, W. F., & McCarty, P. (2011). The Use of Digital Technologies in the Classroom: A Teaching and Learning Perspective. *Journal of Marketing Education*, 33(2), 131–139. doi:10.1177/0273475311410845
- Castells, M. (2001). *The Internet Galaxy*. New York: Oxford University Press. doi:10.1007/978-3-322-89613-1
- Cohen, S. (1972). *Folk Devils and Moral Panics*. London: MacGibbon & Kee.
- Corrin, L., Bennett, S., & Lockyer, L. (2013). Digital Natives: Exploring the diversity of young people's experience with technology. In R. Huang, Kinshuk, & J. M. Spector (Eds.), *Reshaping Learning - The Frontiers of Learning Technologies in Global Context* (pp. 113-138). New York: Springer-Verlag.
- Corrin, L., Lockyer, L., & Bennett, S. (2010). Technological diversity: An investigation of students technology use in everyday life and academic study. *Learning, Media and Technology*, 35(4), 387–401. doi:10.1080/17439884.2010.531024
- Dahlstrom, E., Brooks, D. C., Grajek, S., & Reeves, J. (2015). *ECAR Study of Students and Information Technology*. Louisville, CO: ECAR.
- Dede, C. (2005). Planning for neomillennial learning styles: Implications for investments in faculty and technology. In D. Oblinger & J. Oblinger (Eds.), *Educating the Net Generation* (pp. 15.1-15.22). Boulder, CO: EDUCAUSE. Retrieved March 31, 2006, from <http://www.educause.edu/educatingthenetgen>
- Dohn, N. (2009). Web 2.0: Inherent tensions and evident challenges for education. *Computer-Supported Collaborative Learning*, 4(3), 343–363. doi:10.1007/s11412-009-9066-8
- Downes, T. (2002). Blending play, practice and performance: Children's use of computer at home. *Journal of Educational Enquiry*, 3(2), 21–34.
- Henderson, M., Finger, G., & Selwyn, N. (2016). What's used and what's useful? Exploring digital technology use(s) among taught postgraduate students. *Active Learning in Higher Education*, 1–13.
- Henderson, M., Selwyn, N., Finger, G., & Aston, R. (2015). Students everyday engagement with digital technology in university: Exploring patterns of use and usefulness. *Journal of Higher Education Policy and Management*, 37(3), 308–319. doi:10.1080/1360080X.2015.1034424
- Howe, N., & Strauss, W. (2000). *Millennials Rising: The Next Great Generation*. New York: Vintage.
- Jones, C., Ramanaua, R., Cross, S., & Healing, G. (2010). Net generation or Digital Natives: Is there a distinct new generation entering university? *Computers & Education*, 54(3), 722–732. doi:10.1016/j.compedu.2009.09.022
- Kennedy, G., Dalgarno, B., Bennett, S., Gray, K., Waycott, J., Judd, T.,... Chang, R. (2009) *Educating the Net Generation - A Handbook of Findings for Practice and Policy*. Australian Learning and Teaching Council. Retrieved October 19, 2009, from http://www.altc.edu.au/system/files/resources/CG6-25_Melbourne_Kennedy_Handbook_July09.pdf
- Kennedy, G., Krause, K., Judd, T., Churchward, A., & Gray, K. (2006). *First Year Students' Experiences with Technology: Are They Really Digital Natives?* Melbourne, Australia: University of Melbourne. Retrieved April 10, 2007, from http://www.bmu.unimelb.edu.au/research/munatives/natives_report2006.rtf
- Kent, N., & Facer, K. (2004). Different worlds? A comparison of young peoples home and school ICT use. *Journal of Computer Assisted Learning*, 20(6), 440–455. doi:10.1111/j.1365-2729.2004.00102.x
- Kerawalla, L., & Crook, C. (2002). Childrens computer use at home and at school: Context and continuity. *British Educational Research Journal*, 28(6), 751–771. doi:10.1080/0141192022000019044

- Kvavik, R. B., Caruso, J. B., & Morgan, G. (2004). *ECAR Study Of Students And Information Technology 2004: Convenience, Connection, and Control*. Boulder, CO: EDUCAUSE Center for Applied Research.
- Livingstone, S., & Helsper, E. (2007). Gradations in digital inclusion: Children, young people, and the digital divide. *New Media & Society*, 9(4), 671–696. doi:10.1177/1461444807080335
- Margaryan, A., Littlejohn, A., & Vojt, G. (2011). Are digital natives a myth or reality? University students use of digital technologies. *Computers & Education*, 56(2), 429–440. doi:10.1016/j.compedu.2010.09.004
- Ng, W. (2012). Can we teach digital natives digital literacy? *Computers & Education*, 59(3), 1065–1078. doi:10.1016/j.compedu.2012.04.016
- Oliver, B., & Goerke, V. (2007). Australian undergraduates use and ownership of emerging technologies: Implications and opportunities for creating engaging learning experiences for the Net Generation. *Australasian Journal of Educational Technology*, 23(2), 171–186. doi:10.14742/ajet.1263
- Palfrey, J., & Gasser, U. (2008). *Born digital: Understanding the first generation of digital natives*. New York: Basoc Books.
- Papert, S. (1996). *The Connected Family: Bridging the Digital Generation Gap*. Atlanta, GA: Longstreet Press.
- Prenksy, M. (2001). Digital Natives, Digital Immigrants. *On the Horizon*, 9(5), 1–6. doi:10.1108/10748120110424816
- Prensky, M. (2009). H. sapiens digital: From digital immigrants and digital natives to digital wisdom. *Innovate*, 5(3). Retrieved February 4, 2009, from <http://innovateonline.info/index.php?view=article&id=705>
- Selwyn, N. (2003). ‘Doing IT for the kids’: Re-examining children, computers and the ‘information society’. *Media Culture & Society*, 25, 351–378.
- Selwyn, N. (2004). Reconsidering political and popular understandings of the digital divide. *New Media & Society*, 6(3), 341–362. doi:10.1177/1461444804042519
- So, H., Choi, H., Lim, W. Y., & Xiong, Y. (2012). Little experience with ICT: Are they really the Net Generation student-teachers? *Computers & Education*, 59(4), 1234–1245. doi:10.1016/j.compedu.2012.05.008
- Šorgo, A., Bartol, T., Dolničar, D., & Boh Podgornik, B. (2016). Attributes of digital natives as predictors of information literacy in higher education. *British Journal of Educational Technology*.
- Tapscott, D. (1998). *Growing up Digital: The Rise of the Net Generation*. New York: McGraw-Hill.
- Tapscott, D. (1999). Educating the Net Generation. *Educational Leadership*, 56(5), 6–11.
- Teo, T. (2015). Do digital natives differ by computer self-efficacy and experience? An empirical study. *Interactive Learning Environments*, 1–15.
- Thompson, P. (2013). The digital natives as learners: Technology use patterns and approaches to learning. *Computers & Education*, 65, 12–33. doi:10.1016/j.compedu.2012.12.022
- Toliver, F. (2011). My students will Facebook me but won’t keep up with my online course: The challenges of online instruction. *American Communication Journal*, 13(1), 59–81.
- Warschauer, M. (2004). *Technology and Social Inclusion: Rethinking the Digital Divide*. Cambridge, MA: MIT Press.
- Wellman, B. (2002). Little boxes, glocalization, and networked individualism? In M. Tanabe, P. van den Besselaar, & T. Ishida (Eds.), *Digital Cities II: Computational and Sociological Approaches* (pp. 10–25). Berlin: Springer. doi:10.1007/3-540-45636-8_2

ADDITIONAL READING

Buckingham, D. (2007). *Beyond Technology: Children's Learning in the Age of Digital Media*. Cambridge: Polity Press.

Centre for Education Research and Innovation. (2010). Are the new millennium learners making the grade: Technology use and educational performance in PISA. France: Organisation for Economic Co-operation and Development (OECD).

Cuban, L. (2001). *Oversold and underused: Computers in the classroom*. Cambridge, MA: Harvard University Press.

Kennedy, G., Dalgarno, B., Bennett, S., Gray, K., Judd, T., Waycott, J.,... Krause, K. (2009). *Educating the Net Generation: Implications for Learning and Teaching in Australian Universities*. Australia: The University of Melbourne, Charles Sturt University, Griffith University, The University of Sydney, University of Wollongong. Published under Creative Commons Attribution-Noncommercial-ShareAlike 2.5. Available online: http://www.altc.edu.au/system/files/resources/CG6-25_Melbourne_Kennedy_Final%20Report_July09_v2.pdf

Selwyn, N. (2010). *Schools and schooling in the digital age*. London: Routledge.

Thomas, M. (2011). *Deconstructing digital natives*. New York: Routledge.

KEY TERMS AND DEFINITIONS

Digital Divide: Digital divides are gaps between individuals or groups due to differences in their access to digital technologies. Access

refers to more than physical access, including also the ability to use technologies effectively. Divisions may occur due to factors such as age, gender, race/ethnicity, socio-economic status and/or geographic location.

Digital Generation Gap: The digital generation gap refers to the proposed gap between children and adults (especially parents and teachers) due to young people's natural ability to adapt to new technologies more successfully than older generations.

Digital Immigrant: A digital immigrant is a person born before the widespread adoption of computers and has had to adopt digital technology later in life. Digital immigrants are considered to be less technically able than digital natives and it is argued that they can never develop the same level of technology skills and knowledge as digital natives.

Digital Inclusion: Digital inclusion refers to mindsets, strategies and initiatives that seek to ensure that all people in society have equitable access to technology regardless of their personal circumstances. It is underpinned by the belief that access to technology and the ability to use it effectively are important to citizenship and social cohesion.

Digital Native: In its original sense, a digital native is a person who has grown up after the widespread introduction of the personal computer and therefore been immersed in digital technology. It is claimed that by virtue of this exposure digital natives think, behave and learn differently to older generations. More recently the term has been redefined by some to refer to a person of any age who is highly adept with technology.

ICT Eases Inclusion in Education

E**Dražena Gašpar***University of Mostar, Bosnia and Herzegovina*

INTRODUCTION

Access to education was defined as a fundamental human right in the framework of the Declaration on the Human Rights from 1948. However, the World Report on Disability written in 2011 by the World Health Organization and the World Bank estimates that there are between 93 and 150 million school-aged children with disabilities around the world (UNESCO, 2014). Unfortunately, the fact is that many of these children are completely excluded from educational opportunities, even primary education. Also, because of their learning difficulties, great numbers of children do not have equal access to education.

The inclusion of children with special needs in regular classes creates an entire range of challenges within a given school system and requires the application of new methods and forms of work that are appropriate to each child.

Information and Communication Technologies (ICT) should be increasingly involved in the educational system to improve the quality of teaching, and to provide new experiences during the teaching and learning processes. In this paper term ICT covers all the technologies used to communicate, create, store and manage information. With use of ICT it becomes possible to meet the specific needs of different groups of students, including students with special needs. For students with special needs, ICT can offer numerous ways to remove obstacles as they try to participate in the teaching and learning of the curriculum.

This chapter presents brief analyses of different supportive technologies, such as hardware and software solutions, Web 2.0 technologies, virtual learning environments (VLEs), virtual

worlds, and other similar technologies. The aim of this chapter is to show the potential of ICT in education, especially when facilitating student inclusion. This chapter will also stress some open issues, including limitations in interactions, communication, and learning. ICT can provide new opportunities in inclusive education, and despite all of its potential limitations, ICT should be considered as a key tool to promote equity in educational opportunities.

BACKGROUND

In this paper, the definition of inclusive education, as set out by UNESCO, is adopted: *Inclusive education is a process of strengthening the capacity of the education system to reach out to all learners... As an overall principle, it should guide all education policies and practices, starting from the fact that education is a basic human right and the foundation for a more just and equal society* (UNESCO, 2014, p.11). Coupled with the process of inclusion, the term “special educational needs” (SEN) is often used across Europe. This term is frequently adopted to specify *learners who encounter barriers to learning, either temporarily or in the long term* (EADSNE, 2013, p.6). This definition stresses that the term SEN covers not just learners with disabilities, but all learners who, for various reasons, do not make expected progress for their age. According to this concept, learners with SEN comprise a wider group of students than those with disabilities alone and there are some estimates that these individuals account for around 20% of the school-age population (EADSNE, 2013, p.6).

DOI: 10.4018/978-1-5225-2255-3.ch220

ICTs that are used to support children, young people, and adults with disabilities, are commonly referred to as assistive technologies (AT), although there is no one single internationally accepted definition for this term. The British Assistive Technology Association (BATA), a social enterprise that focuses on AT for inclusion in education, defines AT as *any item, equipment, hardware, software, product or service which maintains, increases or improves the functional capabilities of individuals of any age, especially those with disabilities, and enables them more easily to communicate, learn, enjoy and live better, more independent lives* (BATA, 2015, para.2). In this chapter, the BATA definition of AT is used.

Research into assistive learning technologies has grown significantly in the last decade. Today, research into specific topics can be found in many books and journals worldwide and is the focus of many specialist publications from a variety of disciplines. As such, including all research related to ICT – or, more precisely, to AT – in this short review is in the advance lost battle. Therefore, this review should be understood simply as the author’s own choice of sources related to research on ICT in inclusive education, mostly from last five years.

Recent researches have shown that ICT, in its various forms, is decreasing the gap in education and enabling the inclusion of students with special educational needs in classrooms with their classmates; furthermore, AT help students largely reach their educational goals (Brodin, 2010). Several studies point to the idea that ICT could help SEN students, particularly students with reading and/or writing disabilities, through word processors, word prediction programs, spell and grammar checks, voice recognition, text-to-speech (TTS) programs, planning and organizing tools, etc. (Anderson et al., 2009; Maor et al., 2011; Peterson-Karlan, 2011).

Also, a number of studies have shown that all teachers should be familiar with the use of ICT for SEN students because students who benefit from ICT can be found in every classroom (Anderson et al., 2009; Starcic, 2010). According to the parents in Brodin’s (2010) study, ICT was

used only to a limited extent, and they complained about the teacher’s lack of knowledge about ICT use and tools.

In the last few years, the term “accessible technology” has been adopted to define technologies designed to allow learners to use mainstream technology without any disadvantages, as opposed to AT, which specifically support SEN learners (Barres et al., 2013; McKnight & Davies, 2012). Of course, between those two technologies, considerable overlap exists. Namely, ICT developers apply inclusive design approaches, so their technology can be used instead of AT; using *methods designed to make technologies or information accessible, therefore may be counted as assistive technology approaches, as the approach is assistive, even if the technology is not necessarily so* (McKnight & Davies, 2012, p.13).

ICT SUPPORT FOR INCLUSIVE EDUCATION

The implementation of inclusive education within the regular school system involves a number of activities that take place throughout the entire school practice of all of its participants. The inclusion of children with special needs in regular classes requires the application of new methods and forms of work appropriate to each child (Gašpar & Vetma, 2014). This process is, by itself, very complex and its results are usually not instantly visible. ICT support could ease that process for all stakeholders and make the results more visible. When inclusive education is in question, all aspects of the use of ICT become important. ICT could provide SEN students with the following key inclusive benefits (Winter & O’Raw, 2010, p.87):

- Better control over their own learning experience.
- Students can participate and contribute more fully in classroom activities, and they can also complete assignments independently.

- Students can interact, to a greater extent, with their typical peers, ultimately improving their social skills and enhancing their acceptance.

ICT Classifications

There are different approaches and classifications about what inclusive ICTs in education comprise. The reason for this lies in the fact that ICTs/ATs are very diverse, numerous, and are constantly undergoing change. In one of its documents, UNESCO recognized that inclusive ICTs for education include (UNESCO, 2014, p.11):

- **Mainstream Technologies:** Commercial products available in the market and dedicated to all individuals (computers, Web browsers, word processors, whiteboards, mobile phones and etc.).
- **Assistive Technologies:** Specific technologies that enable SEN learners to access and use mainstream technologies (medical aids, learning aids such as screen readers, alternative keyboards, augmentative and alternative communication devices, and etc.).
- Compatibility between AT products and mainstream technologies.
- Accessible media and formats, such as mainstream publication formats (MS Word, PowerPoint, and structured and tagged PDF files) or HTML5 (Hypertext Markup Language), videos with captioning, DAISY (Digital Accessible Information System) books, EPUB, etc.
- Accessible digital learning content and instructional delivery systems, such as those found in online learning environments, in the classroom, or in learners' management systems.

However, there are other approaches, like the ones used at LEARN (Leading English Education and Resource Network), a non-profit organiza-

tion that primarily serves the public and private Anglophone and Aboriginal Youth and Adult education sectors of Québec, Canada (LEARN, 2013). LEARN organized learning aid in two main groups: computer and peripherals and learning and communication aid. A different approach is used by Futurelab at NFER – the National Foundation for Educational Research in the United Kingdom (FUTURELAB, 2009). Futurelab organized AT in six categories: mobile technologies, audiovisual tools, online communities, podcasts, blogs, wikis, games and learning platforms.

The focus of this chapter is primarily on specific the ICT used for inclusion education (AT), rather than on the general accessibility of mainstream ICT, especially when hardware and software programs are in question. The currently developed AT classification follows the main chapter's idea and comprises three main types of AT:

- AT based on hardware solutions,
- AT based on software solutions, and
- AT based on Web technologies.

Hardware AT solutions are comprised of devices that ease one's access to a computer (PC or laptop), such as substitutes to the standard keyboard (ergonomic keyboard, keyboard with large keys, pen-touch keyboard, adjustable keyboard, touchless keyboard, left-handed/one-handed keyboard, Braille keyboard, voice-activated keyboard, illuminated keypad, etc.), a substitute to a standard screen (cursor/pointer tracking, screen reader – Braille, and speech synthesizer, etc.), or a substitute to a standard mouse (trackball, switches, mouth-activated pointing device, foot-activated pointing device, head- or eye-activated pointing device, voice-activated control device, touch screen, mouse emulator, etc.). Also, hardware AT solutions include other devices (tablet, mobile phone, digital camera, scanners, text [Optical Character Recognition, or OCR] scanner, barcode scanner with synthesized voice, devices with symbol-based communication, and TTS output, etc.) that could be used as input devices or for data capture and

communication support to an adapted learning environment. For example, OCR is a method of converting text from paper format to an electronic version, usually by using a scanner. Accordingly, books, printed worksheets, even photographs with graphics and text can be converted to digital form and read aloud using TTS.

Software AT solutions are comprised of software-enabling oral/written communication (word processors, voice recognition, TTS, word prediction, etc.), remote communication software (e-mail software with icons), software support for specific subjects (spelling/grammar check, dictionary, visual dictionary, digital calculator, digital microscope, digital encyclopedia, notation software for music, sequencer music software, synthesizer, paint or image editing software), and software support for planning and organizing (graphics organizers, electronic agenda, digital portfolio, etc). For instance, some of the built-in features in standard word processing software can support students who have difficulty with written language and processing. For example, spell check helps students with dysgraphia and other learning disabilities. Also, the autocorrect feature can be enabled or disabled depending on students' strengths and needs. Grammar check helps students identify awkward grammatical constructions like passive sentences. TTS add-ins support auditory proofing before students submit their work, while numerous free TTS add-ins for Microsoft Word are also available (Hobgood & Ormsby, 2014).

Web technologies that could be used in inclusive education are comprised of social networks (Facebook, Twitter, LinkedIn, etc.), Website creation tools (blogs, wikis, etc.), collaborative tools (virtual learning environments), three-dimensional (3D) tools (virtual worlds, virtual touch, etc.), audio tools (audio creating, audio sharing, etc.), video tools (video creating, video sharing, video streaming, etc.), text-based tools (note-taking and document creating software, discussion forums, etc.), and image-based tools (image sharing, online whiteboards, drawing tools,

etc.). Web-based systems provide opportunities for young people to access multimedia-rich resources, which can support SEN learners, and also provide a richer learning experience for all. For instance, for students who do well with written products, online text platforms like blogs and wikis can increase motivation by offering the promise of an attractive product with a "real" audience. Some blogging sites offer teachers the ability to create a classroom blog linked to individual student blogs (Hobgood & Ormsby, 2014).

Virtual Learning Environment

A Virtual Learning Environment (VLE), also known as Course Management Systems (CMS) or Learning Management Systems (LMS), is a system through which learning materials are delivered to students via the Web. These systems include assessment, student tracking, collaboration, and communication tools (OXFORD ORC, 2015). These environments also offer students access wherever they are, 24 hours a day, 7 days a week, 365 days a year. Because of that, educational institutions can offer lessons not only to full-time students, but also to workers studying part time.

There are different types of VLE and they fit into any one of the following three categories (OXFORD ORC, 2015, para.2):

- Off-the-shelf environments, such as Blackboard or WebCT.
- Open source environments (often free to use and adapt, but where support is charged for), such as Moodle. Figure 1 is an example of a Moodle screen.
- Bespoke (developed by institutions for their own individual needs).

Virtual World

A virtual world is a computer-based, online community environment that is designed and shared by individuals so that they can interact in a custom-built, simulated world. Today, users interact in

Figure 1. Example of a Moodle screen (MOODLE, 2015)

Source: Moodle, 2015.



the virtual world using graphical models called avatars. Avatars are controlled using input devices (keyboard, mouse, and other specially designed simulation gadgets). All virtual worlds possess the qualities of persistence and interactivity. This enables the users to explore the inherent benefits of socialization and allows them to study human nature and users' abilities (TECHOPEDIA, 2015).

The educational activities that can be carried out in a virtual world are quite diverse (Dalgarno & Lee, 2012, p.241):

- Exploration (simulation of a library, school, or church).
- Practicing with theoretical concepts through interaction (with other avatars or with objects in simulated locations).
- Carrying out activities that are difficult to perform in the real world, either for their difficult implementation or dangerousness (e.g., simulation of a medical intervention, handling chemicals).
- "Role playing" activities, assigning different roles to each student.
- Playing serious games, in which the student learns in a playful way.

Some of the popular virtual worlds are: Second Life, OpenSim, Unity, and Active Worlds.

One example of the use of a virtual world in education is the "Accessibility in Virtual Worlds" project, which is aimed at blind students, whose positions are indicated by sounds, thus enabling navigation and interaction with peers, both blind and sighted (Sheehy, 2010). Another example that uses virtual worlds for inclusion is Brigadoon. It is an island in Second Life that serves as a therapeutic place for people with Asperger syndrome (autism); it facilitates these students' social interactions by using avatars in a controlled environment (Biever, 2007). Figure 2 is an example of a Second Life screen.

SOLUTIONS AND RECOMMENDATIONS

The process of implementing ICT in inclusive education is very complex, resource consuming (specifically of people, time, and material resources), and it does not always feature clearly visible and measurable results. The latest OECD PISA (Program for International Student Assessment)

Figure 2. Example of a Second Life screen
(Second Life, 2015).
Source: Second Life, 2015.



report issued on September 15, 2015 says that even countries which have invested heavily in information and communication technologies (ICT) for education have seen no noticeable improvement in their performances in PISA results for reading, mathematics or science (OECD, 2015a, para.2). Conversely, although only 42% of 15-year-old students in Korea and 38% in Shanghai–China stated that they used computers at schools, students in both Korea and Shanghai–China were among the top performers in digital reading and computer-based mathematics tests in the OECD PISA in 2012. On the other hand, in countries where students demonstrated more Internet use at school for schoolwork, students’ performance in reading generally declined between 2000 and 2012 (OECD, 2015b, p.15). The results of the OECD PISA report show that ICTs have not yet been widely adopted in formal education. *But, where they are used in the classroom, their impact on student performance is mixed, at best. In fact, PISA results show no appreciable improvements in student achievement in reading, mathematics or science in the countries that had invested heavily in ICT for education* (OECD, 2015b, p.15).

One of the possible interpretations of these findings could be that educators need additional time and effort to learn how to use ICT in the classroom

and, at the same time, they need to stay focused on student learning. Because of that, the PISA report results should not be interpreted in such a way that investment in and implementation of ICTs in education should be stopped. On the contrary, these results point to the need for critical thinking about how to implement and integrate ICTs in education, particularly in inclusive education.

Although in most countries, gaps in computer access between advantaged and disadvantaged students has lessened between 2009 and 2012, the results from the PISA report show that *once the so-called “first digital divide” (access to computers) is bridged, the remaining difference between socio-economic groups, in the ability to use ICT tools for learning is largely explained by the difference observed in more traditional academic abilities* (OECD, 2015b, p.16). The conclusion is that in order to create equal opportunities in the digital world, it is better to ensure that all learners achieve a basic level of proficiency in reading and mathematics than to get expanded access to high technology (OECD, 2015b, p.16).

It is obvious that relationships between learners, ICT, and the learning process are very complex and loose, so any real contributions that the ICT could make to teaching and learning have yet to be fully researched.

There are a lot of open issues related to the process of ICT implementation in inclusive education, such as:

- How to ensure full engagement, dedication, and participation of all relevant stakeholders (governments, teachers, [SEN] learners, parents...).
- How to ensure full cross-sectorial and intergovernmental agency cooperation and coordination.
- How to avoid harmful aspects of Internet use (information overload, plagiarism, fraud, violations of privacy, online bullying, etc.).
- Development, implementation, and monitoring of realistic strategies related to ICTs in inclusive education.
- Overestimation of digital skills among learners, teachers, and parents.
- Hidden resistance of teachers, learners, and parents (everyone supports innovation, but not in his/her school and not for his/her children).
- The benefits of “good practice” cases are mostly insufficient to mobilize support (the costs for “bad practice” are concentrated).
- The relatively low quality of educational software and courseware.
- The loose standards in developing educational software and courseware.

Some of the recommendations for coping with the listed issues are as follows:

- Previous experiences with the implementation of ICTs in inclusive education, whether good or bad, should be something to learn from.
- The government should have to ensure the framework for successful implementation of ICT in inclusive education. That framework includes appropriate policy documents (strategies, action plans, etc.) developed in coordination with all rel-

evant stakeholders (governments, ministries, agencies, teachers, parents, etc.). Also, government responsibility includes developing adequate ICT infrastructure (especially communication infrastructure, such as providing broadband access to the Internet), additional support to educate teachers and parents about ICT in inclusive education, and enabling access to information about available ICT tools and standards (Websites).

- The foundation skills necessary for the digital age should be taught, and not always with the use of ICT.
- Teachers have to become not just leaders in the implementation of ICTs in inclusive education, but also equal partners in their design.
- Teachers are crucial in ensuring the successful exploitation of ICT in inclusive education, but they require education and training to equip them to succeed.
- Teachers should combine their efforts and experience to improve /develop teaching and learning methodologies featuring fully integrated ICT.
- ICT should support the teaching and learning process. The focus of teachers should be on learners, teaching, and learning methods.

ICT can be a valuable tool for SEN learners who are more vulnerable to the digital divide and to exclusion from educational opportunities. If ICT is implemented and integrated into teaching and learning processes in an appropriate way, ICT can improve students' quality of life by increasing their participation and reducing social exclusion.

FUTURE RESEARCH DIRECTIONS

Further research related to ICT in inclusive education should be focused on the following directions:

- Forming interdisciplinary teams composed of teachers, learners, parents, ICT professionals, pedagogues, psychologists, etc., in designing and developing inclusive ICT tools.
- Forming interdisciplinary teams composed of teachers, learners, parents, ICT professionals, pedagogues, psychologists, etc. in developing new and improving existing teaching and learning methods.
- Forming interdisciplinary teams composed of teachers, learners, parents, ICT professionals, pedagogues, psychologists, etc. in developing and implementing standards for inclusive ICT.
- Research the use of tablet and mobile phones in inclusive education.
- Research new methods of interaction with ICT, which are achieved by body gestures, eye movement, and via the brain and nervous system.

CONCLUSION

This chapter briefly described the types of technologies used, and the ways in which they can be used, in inclusive education. Although all aspects of the use of ICT are important, many researchers have observed the importance of how ICT is applied – meaning how it is integrated with teaching methodologies. All past experiences of ICT implementation in inclusive education have demonstrated that providing the technology itself is not enough. Teachers are essential in that process, but they require additional education and training, as well as more information on how to choose appropriate ICT tools. Namely, the use of ICT in inclusion is a lifelong requirement that extends beyond the classroom. Some learners require that information is provided in a form that they can easily understand (in an easy-to-read or symbol format), or on an accessible Website that can be read aloud using screen-reading technology. Other

students might need a device that speaks, and thus supports communication and social interaction. By allowing users to learn at their own pace, ICT can also encourage less able students while enhancing their self-confidence.

New ICT devices, like tablets and smart phones, as well as new methods of interacting with technology via body gestures, touch, eye movement, and even directly via the brain and nervous system, are providing new opportunities, as well as potential barriers, for interaction, communication, and learning, thus opening the door to new research opportunities. However, the general conclusion is that ICT should be considered as a key tool for promoting equity in educational opportunities, and that access to appropriate ICTs should be considered as an essential human right.

REFERENCES

- Anderson, C. L., Anderson, K. M., & Cherup, S. (2009). Investment vs. return: Outcomes of special education technology research in literacy for students with mild disabilities. *Contemporary Issues in Technology & Teacher Education*, 9(3), 337–355.
- Barres, G. D., Carrión, C. Z., & Delgado, R. (2013). *Technologies for Inclusive Education: Beyond Traditional Integration Approaches*. Hershey, PA: IGI Global. doi:10.4018/978-1-4666-2530-3
- BATA - British Assistive Technology Association. (2015). Retrieved August, 14, 2015 from <http://www.bataonline.org/further-assistive-technology-definition>
- Biever, C. (2007). Web removes social barriers for those with autism. *New Scientist*, 26–27.
- Brodin, J. (2010). Can ICT give children with disabilities equal opportunities in school? *Improving Schools*, 13(1), 99–112. doi:10.1177/1365480209353483

Dalgarno, B., & Lee, M. J. W. (2012). Exploring the relationship between afforded learning tasks and learning benefits in 3D virtual learning environments. *Proceedings of the 29th ASCILITE Conference*, 236–245.

EADSNE – European Agency for Development in Special Needs Education. (2013). *Information and communication technology for inclusion*. Odense, Denmark: European Agency for Development in Special Needs Education.

FUTURELAB. (2009). *Using digital technologies to promote inclusive practices in education - a Futurelab handbook*. Retrieved April, 22, 2015 from http://archive.futurelab.org.uk/resources/documents/handbooks/digital_inclusion3.pdf

Gašpar, D., & Vetma, S. (2014). ICT in promoting inclusive practice in education. In *Proceedings of International Conference on e-Education - ICeE* (pp. 55-62). Mostar, Bosnia and Herzegovina: Faculty of Information Technologies, University “Džemal Bijedić”.

Hobgood, B., Ed, D., & Ormsby, L. (2014). *Inclusion in the 21st-century classroom: Differentiating with technology*. Retrieved July, 12, 2015 from <http://www.learnnc.org/lp/editions/every-learner/6776>

LEARN - Leading English Education and Resource Network. (2013). *Types of Assistive Technology*. Retrieved July, 14, 2015 from http://www.learnquebec.ca/en/content/pedagogy/insight/intech/assistive_technology/AssisTechTypes.htm

Maor, D. C. J., & Drewry, R. (2011). The effectiveness of assistive technologies for children with special needs: A review of research-based studies. *European Journal of Special Needs Education*, 26(3), 283–298. doi:10.1080/08856257.2011.593821

McKnight, L., & Davies, C. (2012). *Current Perspectives on Assistive Learning Technologies – 2012 review of research and challenges within the field*. Oxford, UK: The Kellogg College Centre for Research into Assistive Learning Technologies.

MOODLE. (2015). Retrieved September, 9, 2015 from <http://school.demo.moodle.net/mod/forum/view.php?id=158>

OECD. (2015a). *New approach needed to deliver on technology's potential in schools*. Retrieved September, 22, 2015 from <http://www.oecd.org/education/new-approach-needed-to-deliver-on-technologys-potential-in-schools.htm>

OECD. (2015b). *Students, Computers and Learning: Making the Connection - Executive Summary*. Retrieved September, 22, 2015 from http://www.keepeek.com/Digital-Asset-Management/oecd/education/students-computers-and-learning/executive-summary_9789264239555-2-en#page1

OXFORD ORC. (2015). *Learn about Virtual Learning Environment/Course Management System content*. Retrieved August, 30, 2015 from <http://global.oup.com/uk/orc/learnvle/>

Sheehy, K. (2010). Virtual worlds, inclusive education The TEALEAF framework. In K. Sheehy, R. Ferguson, & G. Clough (Eds.), *Controversial Issues in Virtual Education: Perspectives on Virtual Worlds* (pp. 1–15). New York, NY: Nova Science Publishers.

Staricic, A.I. (2010). Educational Technology for the Inclusive Classroom. *Turkish Online Journal of Educational Technology*, 9(3), 26-37.

TECHOPEDIA. (2015). *Definition - What does Virtual World mean?* Retrieved May, 10, 2015 from <https://www.techopedia.com/definition/25604/virtual-world>

UNESCO. (2014). *Model Policy for Inclusive ICTs in Education for Persons with Disabilities*. Retrieved, June, 11, 2015 from <http://www.unesco.org/new/en/communication-and-information/resources/publications-and-communication-materials/publications/full-list/model-policy-for-inclusive-icts-in-education-for-persons-with-disabilities/>

Winter, E., & O’Raw, P. (2010). *Literature Review of the Principles and Practices relating to Inclusive Education for Children with Special Educational Needs*. National Council for Special Education.

ADDITIONAL READING

Beacham, N., & McIntosh, K. (2014). Student teachers attitudes and beliefs towards using ICT within inclusive education and practice. *Journal of Research in Special Educational Needs*, 14(3), 180–191. doi:10.1111/1471-3802.12000

Benigno, V., Bocconi, S., & Ott, M. (2007). Inclusive education: helping teachers to choose ICT resources and to use them effectively. *eLearning Papers*, 6, 1-13.

Blamire, R. (2009). ICT Impact Data at Primary School Level: the STEPS approach. In Scheuermann, F., & Pedro, F. (eds.) *Assessing the effects of ICT in education*, 199-211. European Union/OECD. France

Bocconi, S., Kampylis, P., & Punie, Y. (2013). Framing ICT-enabled Innovation for Learning: The case of one-to-one learning initiatives. *Europe European Journal of Education*, 48(1), 113–130. doi:10.1111/ejed.12021

Chantry, J., & Dunford, C. (2010). How do computer assistive technologies enhance participation in childhood occupations for children with multiple and complex disabilities? A review of the current literature. *British Journal of Occupational Therapy*, 73(8), 351–365. doi:10.4276/030802210X12813483277107

Corn, J., Tagsold, J. T., & Argueta, R. (2012). Students with special needs and 1:1 computing: a teachers perspective. *Journal of Research in Special Educational Needs*, 1(2), 217–223. doi:10.1111/j.1471-3802.2012.01251.x

Edyburn, D. L. (2013). Critical issues in advancing the special education technology evidence base. *Exceptional Children*, 80(1), 7–24.

Edyburn, D. L. (Ed.). (2015). *Efficacy of Assistive Technology Interventions*. Bingley, PA: Emerald Group Publishing Limited. doi:10.1108/S2056-769320151

European Agency for Development in Special Needs Education. (2013). *European and International Policy for Supporting ICT for Inclusion*. Odense, Denmark: European Agency for Development in Special Needs Education European Commission (2013). *Survey of Schools: ICT in Education – Benchmarking Access, Use and Attitudes to Technology in Europe’s Schools*. EC Brussels

Foley, A., & Ferri, B. A. (2012). Technology for people, not disabilities: Ensuring access and inclusion. *Journal of Research in Special Educational Needs*, 12(4), 192–200. doi:10.1111/j.1471-3802.2011.01230.x

Friesen, E. L., Walker, L., Layton, N., Astbrink, G., Summers, M., & De Jonge, D. (2015). Informing the Australian government on AT policies: ARATAs experiences. *Disability and Rehabilitation. Assistive Technology*, 10(3), 236–239. doi:10.3109/17483107.2014.913711 PMID:24796214

Graham, R., & Richardson, W. (2012). Leveling the playing field: Assistive technology, special education and a Canadian perspective. *American International Journal of Contemporary Research*, 2(1), 6–15.

Klotz, M. B., & Koch, C. (2015). Assistive Technology Advances in Education. *National Association of School Psychologists. Communique*, 43(8), 1–8.

McKeown, S., & McGlachon, A. (2015). *Brilliant Ideas for Using ICT in the Inclusive Classroom* (2nd ed.). New York, PA: Routledge.

Morgan, R. K., & Olivares, K. T. (Eds.). (2012). *Quick Hits for Teaching with Technology: Successful Strategies by Award-Winning Teachers*. Bloomington, PA: Indiana University Press.

Murchland, S., & Parkyn, H. (2010). Using assistive technology for schoolwork: The experience of children with physical disabilities. *Disability and Rehabilitation. Assistive Technology*, 5(6), 438–447. doi:10.3109/17483107.2010.481773 PMID:20446900

Politis, Y., Deveril, D., Baldiris Navarro, S., Avila, C., de Lera, E., Monjo, T., & Goodman, L. (2014). *Introducing the Inclusive Learning Handbook: an OER for teachers and policy makers*. Valencia, PA: IATED Academy Publishing.

Ravonne, A., Green, R. A., & Blair, V. (2011). *Keep It Simple: A Guide to Assistive Technologies*. Santa Barbara, PA: Libraries Unlimited.

Reid, G., Strnadová, I., & Cumming, T. (2013). Expanding horizons for students with dyslexia in the 21 st century: Universal design and mobile technology. *Journal of Research in Special Educational Needs*, 13(3), 175–181. doi:10.1111/1471-3802.12013

Sampath, H., Sivaswamy, J., & Indurkha, B. (2010). Assistive systems for children with dyslexia and autism. *ACM Sigaccess Accessibility and Computing. SIGACCESS Newsletter*, 96, 32–36.

Sharma, F. R., & Wasson, S. G. (2012). Speech recognition and synthesis tool: Assistive technology for physically disabled person. *International Journal of Computer Science and Telecommunications*, 3(4), 86–91.

Söderström, S. (2013). Digital Differentiation in Young Peoples Internet Use—Eliminating or Reproducing Disability Stereotypes. *Future Internet*, 5(2), 190–204. doi:10.3390/fi5020190

Steel, E. J., & de Witte, L. P. (2011). Advances in European Assistive technology service delivery and recommendations for further improvement. *Technology and Disability*, 28(3), 131–138.

Wastiau, P., Blamire, R., Kearney, C., Quittre, V., Van de Gaer, E., & Monseur, C. (2013). The Use of ICT in Education: A survey of schools in Europe. *European Journal of Education*, 48(1), 11–27. doi:10.1111/ejed.12020

KEY TERMS AND DEFINITIONS

Assistive Technologies (AT): Information Communication Technologies that are used to support people with special needs.

Inclusive Education: Education tailored to the needs of each learner, regardless of her/his physical, intellectual, social or other conditions.

Information Communication Technologies (ICT): Term ICT covers all the technologies used to communicate, create, store and manage information, including computers, the Internet, broadcasting technologies and telephony.

Students With Special Needs (SEN): Learners with difficulties in learning, temporarily or permanently, that cause that they do not make expected progress in education for their age.

Virtual Learning Environments (VLEs): The specific form of educational system that includes tools for communication, collaboration, student tracking, assessment and so on, all based on Web platform.

Virtual World: Computer based simulation of real world where users can create a personal avatar which explore the virtual world, participate in its activities and communicate with other avatars.

Web Technologies: All technologies based on any effective computer network (Local Area Network, Wide Area Network) that make it possible for different computers to communicate and share resources.

The Infusion of Technology Within the Classroom Facilitates Students' Autonomy in Their Learning

Fariel Mohan

University of Trinidad and Tobago, Trinidad and Tobago

Garry Soomarah

University of Trinidad and Tobago, Trinidad and Tobago

INTRODUCTION

The use of technology in education is not something new or innovative but rather an asset in assisting our students with their cognitive skills. Many people have argued that the benefit of technology infused into education is non-existent or even limited to have any significant impact to truly advocate its use in the mainstream curriculum. To have an impact, some (Niess, 2005; Shulman, 1986) suggest that in order for technology to become an integral component or tool for learning, science and mathematics pre-service teachers must develop an overarching conception of their subject matter with respect to technology and

what it means to teach with technology. Others (Hooper, 1991; Rieber & Welliver, 1989) state the full potential of any educational technology can only be realized when educators progress through five steps or phases: Familiarization, Utilization, Integration, Reorientation, and Evolution, otherwise, the technology will likely be misused or discarded. It is further stated that the traditional perspective of educational technology focuses on either the technology itself or a teacher's instruction which is only the first three phases as shown in Figure 1.

The motivation behind this experiment was to investigate if technology can be used to assist students to focus on constructing his/her own

Figure 1. A model of adoption of both "idea" and "product" technologies in education

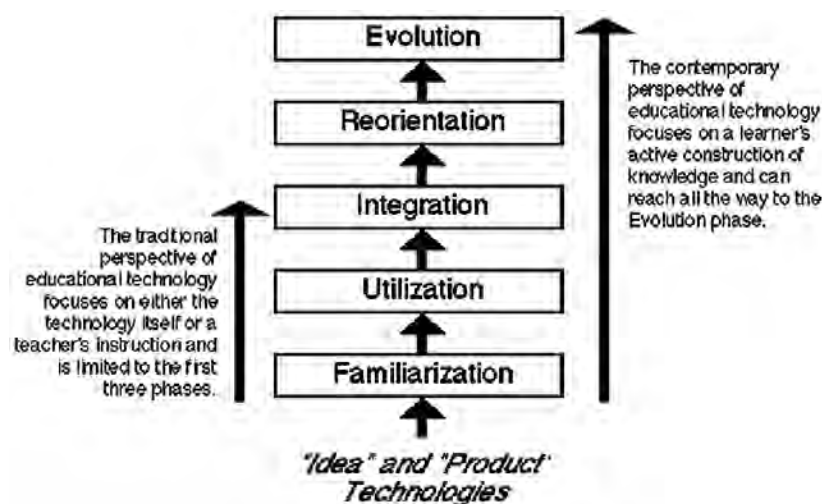


Figure 2. Summary of CSEC maths pass rate (2009 – 2011)

CXC CSEC Pass Rates:			
Mathematics examination			
Average score for June examination			
Year	Total % students passing Paper 1	Total % students passing Paper 2	Total % students achieving a passing grade overall
2011	62%	17.5%	36%
2010	57%	24%	41%
2009	62%	26%	45%

knowledge without developing an overarching conception of their subject matter. The success of this experiment was determined by an improvement in the pass rate in teaching. Research was done in order to determine which teaching subject will be the subject of this experiment. Researchers have noted that many students entering university often fail 1st year mathematics which may be due to a poor mathematics background (Eng, Li Li, & Julaihi, 2009; Rylands, & Coady, 2009; Whannell, & Allen, 2012). In the Caribbean, there is also a general weakness in students' mathematics background as stated by a former deputy principal of the University of the West Indies, Mona campus (Green-Evans, 2005). Students who complete high school write an official examination from Caribbean Secondary Education Certificate (CSEC) in order to get accepted into a university. Figure 2 shows a table summarising the results of all the Caribbean high school students who wrote CSEC mathematics for 3 years. This performance re-enforce the fact that students have a weakness in mathematics, especially in paper 2, the written examination. Paper 1 is multiple choice examination (Caribexams, 2004).

Based on these findings, the subject chosen for the experiment was 1st year Mathematics at the University of Trinidad & Tobago (UTT). Technology was used to build a virtual classroom (VC) to

assist students with their cognitive skills ensuring that the subject matter in teaching mathematics did not require an overarching conception. Instead the VC and physical classroom was part of a blended approach. The VC provided an environment for students to focus on expressing their existing knowledge of mathematics, reflecting on this knowledge and thinking in order to provide prompt feedback. At the end of the experiment, students provided suggestions to enhance the VC which were considered and other VCs were designed.

BACKGROUND

The relatively recent introduction of new technology into mainstream schooling was widely expected to penetrate and transform teaching and learning across the curriculum. As noted by (Kinach, 2002), teacher educators must challenge their pre-service teachers' habitual ways of thinking about subject matter and subject matter teaching. This is especially important in the teaching of mathematics since for many years there has been increasing concerns about students failing achievement in mathematics and their negative attitudes towards mathematics, despite its importance in the 'world today' (Gresham, 2007). Some educators believe that students develop attitudes and

emotional reactions towards mathematics from as early as 9 years old (McLeod, 1992), which are seldom ambivalent; rather, they are either positive or negative, with negative attitudes persisting well into adulthood (Brady & Bowd, 2005). Research into the factors that impact upon students' success in, and attitudes towards mathematics point to mathematics anxiety as one of them (Shores & Shannon, 2007).

The aim of this project was to investigate whether infusing technology into education can advocate use in the mainstream due to an increase in pass rate. If there was an increase of the pass rate of 1st year mathematics, it suggests this approach had a positive impact in teaching and further work is required. The VC presented teaching matter through reflective questions emphasizing on background knowledge which focused on 9 years old students, e.g. fractions. The VC was designed so that it was free, accessible 24/7, fosters collaboration, anonymous and most importantly to ensure that students feel comfortable thus encouraging students to share their understanding and thoughts on a question. The students were given an incentive of 10% of final mark awarded based on the number of comments on the VC. For example, more than 150 comments was awarded 10%. Even though the students were extrinsically motivated into participating in the beginning for the marks, their attitude towards the VC changed in which they sought greater autonomy in the learning process. This drive towards autonomy was made possible because of the many benefits derived from using the VC such as the students took control of his/her learning which was demonstrated by student's frequent comments.

This VC provided an environment where students were able to share their thoughts without being ridiculed as well as learn at his/her own pace. Thus motivating students to build on whatever existing knowledge they had which according to Maslow (Huit, 2007) is achieved by satisfying their needs according to his hierarchal model. Based on his theory it is perceived that once ones

basic needs are met they are able to move on to the other level of development. Maslow's motivation theory suggest that an individual can only move up the pyramid based on the needs of the previous level being met or satisfied. Due to their needs being met, students' level of competency proportionally increased, therefore their self-efficacy in the given study increased. A contributing factor to this was anonymity which allowed everybody using the VC to share their thoughts. This allowed students to be bold and take risks in asking questions which otherwise may not have been asked thus the VC acted as a form of scaffolding.

Scaffolding acts as a stepping stone in which students build on previous knowledge before moving on to the next level. This is achieved when students are comfortable in sharing their existing level knowledge and building on it from their peers or instructor, who is an expert. According to Vygotsky (1978) scaffolding can act as a form of social and instructional support needed by students who are learning new material. After students have achieved understanding of the specific concept the scaffold is taken away so that they can move on to the next level. Vygotsky (1978) also spoke of the Zone of Proximal Development (ZPD) which he defined as the distance between the actual developmental level as determined by independent problem-solving and the level of potential development as determined through problem-solving under adult guidance, or in collaboration with more capable peers. The VC results suggest that based on the components of Vygotsky's (ZPD), the instructor's guided problem solving questions motivated the students to collaborate resulting in just over 2000 entries in one of the experiment. Hence they were able to move from Bloom's (1956) cognitive taxonomy of knowledge to that of analyzing, evaluating and creating which are higher order cognitive skills in the taxonomy. The improved understanding by the students will result in more students passing thereby increasing the pass rate.

DESIGN OF EXPERIMENTS

The experiment was designed to help students discover his/her current knowledge and then construct new knowledge by building on his/her existing knowledge. A VC was built as a support to the traditional classroom teaching using the blended approach. The sample was 1st year UTT students and all 68 students used the VC. The objective was to improve the students' understanding thereby improving the pass rate for UTT 1st year Mathematics for Technicians. The pass rate was 25 – 40%.

All students had to use the VC and some students wanted a choice. The 2nd VC was from 100 students who chose either VC or answer the questions on paper. The objective was to determine if the improvement to the pass rate for Mathematics for Technicians Level 1 was due to technology or not.

Students approached the instructor to use technology to teach without a physical classroom. This was the 3rd VC with 28 repeating students who were scheduled for classes 5:00 pm – 9:00 pm (difficulty in getting transport after 9:00 pm). The objective was to improve the pass rate for Mathematics for Technicians Level 1 regardless of the constraint of no traditional teaching.

Some students reoriented the VC and a new VC was evolved through them. The 4th VC was totally the students' idea and design. The objective was to allow students and instructor to guide the VC by the postings. Also social and entertaining postings such as videos and pictures were allowed. This VC had 100 students.

1ST VIRTUAL CLASSROOM

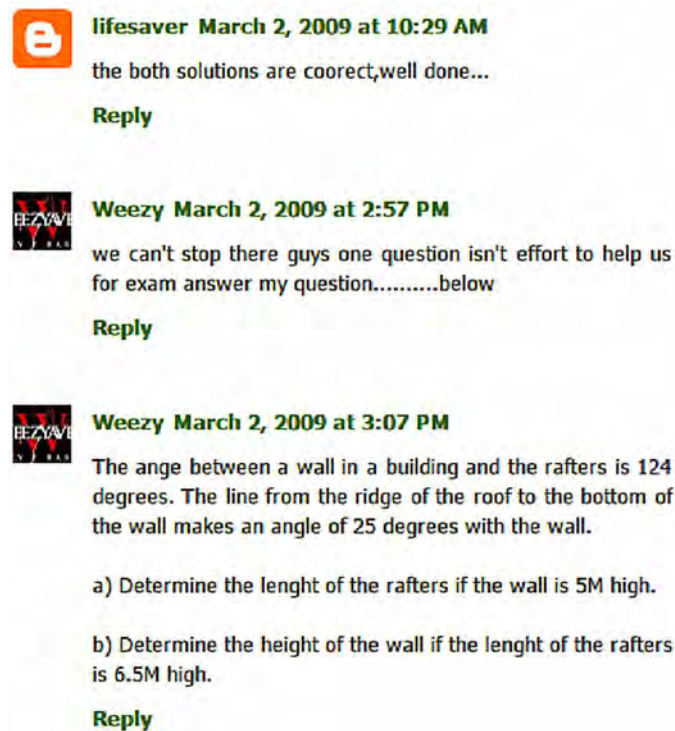
The 1st VC was designed to incorporate technology with learning mathematics by providing a classroom outside the physical classroom which was not affected in any way. The VC, a blog from blogger, was designed using Web 2.0 technology. All 68 students had to be registered anonymously

to this one blog. The anonymity promotes honest opinions as it combats fear of judgement and embarrassment of correction, thus these barriers which exist in a physical classroom no longer were limitations. A student was allowed to register as different users to create suspense and interest. The 68 students were all the 1st year Mechanical, Chemical & Petroleum engineering students. Each discipline was taught separately but by the same instructor and had a common final examination. As an incentive to encourage the students to make comments and post questions on the VC, 10% (of final mark) was awarded based on his/her VC usage. For example, more than 150 postings was awarded the full 10%.

The VC was guided by the instructor who posted questions hoping to engage the students in reflecting and making his/her comments. The comment displays the student current knowledge and can be an eye opener into background knowledge. Writing a comment will also have an additional benefit in providing an opportunity for a student to express his/her knowledge through writing thus enhancing writing skills. Some students understanding of the content lead them to provide feedback for other students' comments thus allowing that student to take the role of the instructor. The opportunity for a student to be the instructor motivates that student to continue using the VC. This is based on the work of Steinberg and Maslow (1943) who state the importance of motivation playing an integral part in educational development of student autonomy in their learning. An incorrect comment provided an opportunity for another student to analyse and rebuild that incorrect knowledge. This dialog thus breaks down the barriers between students and instructor thereby shifting the role of instructor. An example from 2 VC users, Lifesaver and Weezy, is shown in Figure 3.

Problem solving which is inherent in the study of mathematics has been demonstrated to strongly influence students' attitudes, both positive and negative (Debellis & Goldin, 2006; Hannula, 2002). The VC strived to portray a positive outlook

Figure 3. Dialog between students



in an attempt to connect to the individual level each student was at. Students were encouraged in the physical classroom to use the VC by assuring them prompt feedback. The instructor was registered as a very unintelligent student asking a lot of questions. This provided an avenue for students to think and answer these questions. Figure 4, shows the number of comments that a post on a given day received (Mohan, 2008). At the end of the semester, the students said the marks for VC usage motivated them to use the VC. The number of students who passed the final examination increased from 40% to 67%.

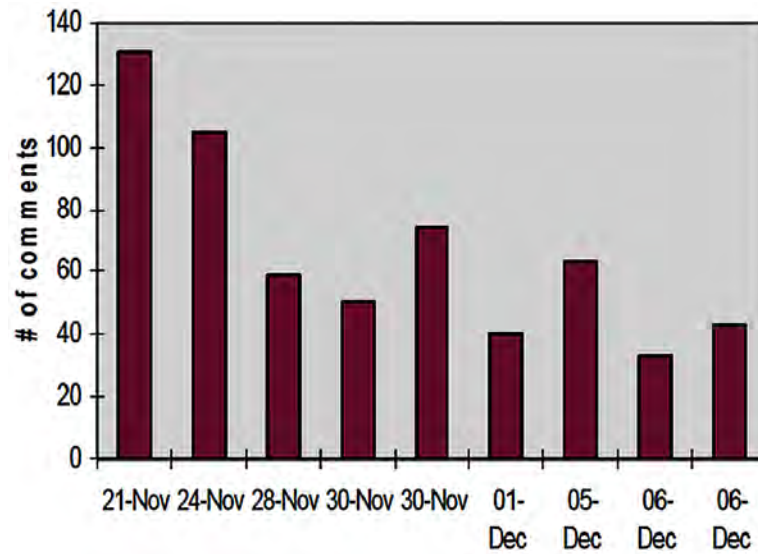
2ND VIRTUAL CLASSROOM

The 2nd VC provided the 100 students with a choice, to choose to use the VC or not to use the VC. The students who choose not to use the VC, they were given all the instructor's questions posted and had to write their answers and submit in order to

be evaluated for the 10%. All the students were Mechanical, Chemical or Petroleum engineering students and was taught by the same instructor at the same time in the auditorium. An examination was given before using the VC to get an idea of mathematical background which was used for the initial questions posted. The VC option was chosen by 80 students while 20 students chose no VC. The idea of giving choice was to let the student take ownership of his/her own learning. This experiment aimed at investigating whether the needs of students with respect to technology as a scaffolding tool, can aid students' performance. The VC users benefited by collaboration and learning at own pace. A student can choose to read the questions and comments, correct an incorrect comment, answer a posted question or ask a question. The anonymous student was motivated to learn at his/her pace and constructed new knowledge without the judgment of peers.

This trust in ownership and autonomy in learning is highly advocated by constructivist theorist

Figure 4. Graph showing usage



Jean Piaget “who articulated mechanisms by which knowledge is internalized by learners. He suggested that through processes of accommodation and assimilation, individuals construct new knowledge from their experiences.” According to Piaget, assimilation is when individuals incorporate new experiences into an already existing framework without changing that framework which can be faulty if their understanding is not clear. Whereas, accommodation “is the process of reframing one’s mental representation of the external world to fit new experiences. Accommodation can be understood as the mechanism by which failure leads to learning: when we act on the expectation that the world operates in one way and it violates our expectations, we often fail, but by accommodating this new experience and reframing our model of the way the world works, we learn from the experience of failure, or others’ failure.” This was further cemented by students being able to submit questions in their VC which promoted creative, critical and problem solving skills in the cognitive domain. While in the affective domain they developed social competencies like respect, appreciation and caring about the views of their classmates. This, in today’s society is an important

ingredient in companies with respect to teamwork and being a team player.

The VC served as a flexible, accessible and convenient community providing the ingredients necessary for collaboration. Figure 5 shows an example of a dialog from 2 VC users, Sparkle and marz. Sparkle commented and marz commented on Sparkle’s comment. This demonstrated marz reflected on sparkle’s comment and new knowledge evolved hence marz commented on sparkle’s comment (Mohan, 2011).

“The learning community was like an extended family and the friends that I made here became the most important reason for me to come to class and to continue with my college education”

Figure 5. Dialog among students

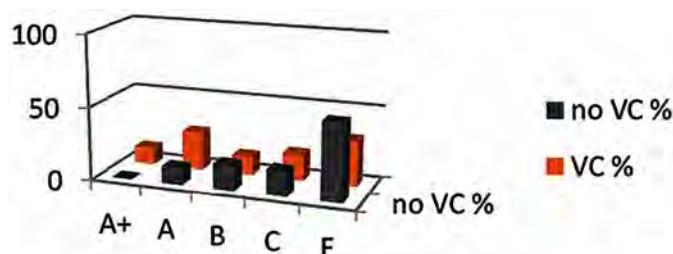
sparkle said...

$5\cos x + 4x^5 + 8$
integrated form is $5 \sin x + \frac{4}{6} x^6 + 8x + c$

marz said...

SPARKLE CAN I ASK WHY U INTEGRATED
THE 8 isn't that a constant
THIS IS HOW I WUD HAVE INTEGRATED IT.
add 1 to the power
 $4x^6 + 5 \sin x + c$
because c is the constant

Figure 6. Graph comparing performance virtual classroom vs. no virtual classroom



(Hesse and Mason, 2003). Learning communities create learning environments where students are not expected to be passive listeners, taking notes and memorizing facts, but instead are expected to work together, reading, writing, talking, and relating their learning to their daily lives.

Figure 6 shows the comparison of the % of each grade. It can be seen that 0% of the students got A+ in the no-VC approach while 12% of the students got A+. The % of failure must be also noted. The results suggests the VC does cause an increase in performance and also a deeper understanding displayed by three times the A's were achieved with a lot of A+s. The results obtained from the students who chose no VC had a pass rate increase from 40% to 48% while the students who chose VC had a pass rate increase from 40% to 69%.

3RD VIRTUAL CLASSROOM

The 3rd VC focused on removing the traditional classroom for teaching. This approach was based on the university of Wisconsin-Milwaukee (Garnham & Kaleta, 2002) suggestion that students learn more in blended courses than they do in comparable traditional class sections. Blended course instruction offers more choices for content delivery and can be more effective than courses that are either fully online or fully classroom-based (Singh, 2003). A group of 28 repeaters (students who failed the course) could not attend the scheduled 5:00 pm – 9:00 pm repeaters class due to transportation problem. A physical classroom was available once a week for a lunchtime

period only and this was used for either official assessments or tutorials. The instructor used the knowledge gained by reading comments from the previous VC to address foundational problems with these students. Once these foundational problems were cleared, focus was more on the identifying the learning style of the student.

Figure 7 shows the result of the repeating students. The normal pass rate at UTT for 1st year Mathematics was 40% and the result reflects an increase in pass rate to 82%.

4TH VIRTUAL CLASSROOM

The 4th VC demonstrated students evolving when they decided they wanted to build the VC. They selected a social media, Socialgo, to build the VC. The students also included posting of videos and pictures in the VC. The students asked to also act as an instructor in posting reflecting and thinking questions. The identities of these students remained anonymous. Figure 8 shows the positive

Figure 7. Graph showing performance using blended approach

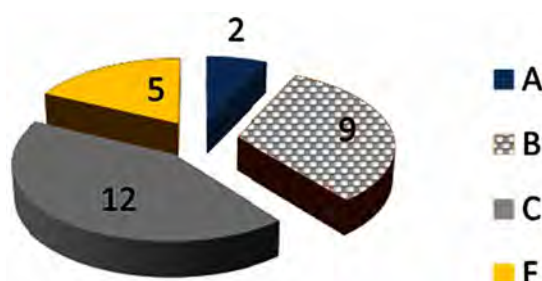
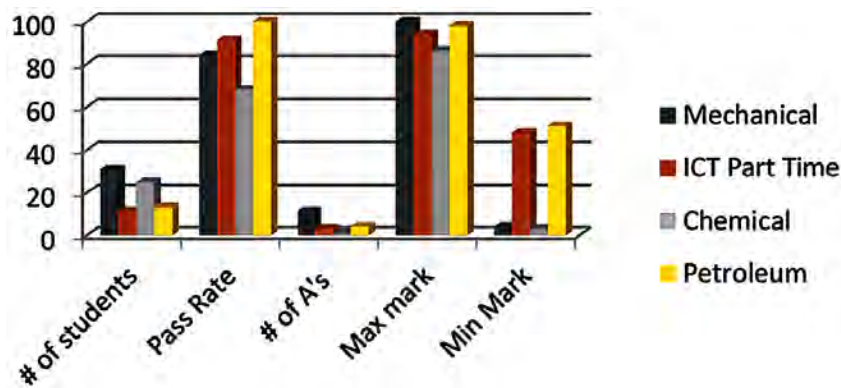


Figure 8. Graph showing performance using social virtual classroom



results in terms of improved performance. The average pass rate of the 4 disciplines showed an increase in pass rate from 40% to 62% with 2 individual disciplines above 80% pass rate.

Table 1 (Mohan, 2010) show a summary of the results from the 4th VC. Even though the non-academic parts, videos and pictures comments were both more than 100 while the academic part, blogs comments was 87, the blog postings have an average of more than 31 comments to each posting. This demonstrates that the postings stimulate thoughts from other students. The results suggest that the non-academic postings stimulated interest which lapped over to the academic posting. The students' academic postings were more than 25% of the postings.

Mohan (2013) demonstrates that when a student has ownership of his/her learning and is motivated, he/she can be very creative. For example Tom registered himself as Dr. MathsBsc-MscPhd, who depicted a brilliant student but Tom

also registered himself as Big Bird, who was an extremely weak student. In a similar manner, the instructor was registered as the instructor, using the instructor's name, and was also registered as a weak student, Black Eye-Peas. Several guidelines were followed in the VC. These guidelines and the motivation for each one are given in Table 2 (Mohan, 2013).

Another important aspect of the VC was that many topics were available at the same time. A student can move at his/her individual pace of learning without interfering with another. This was especially important since time flexibility in learning a topic was essential. In Figure 9, the graph shows a summary of a portion of the 2020 comments, grouped by the time posted, timestamp (Mohan, 2011). A VC must be time convenient as shown in Figure 8.

Figure 9. Graph showing usage at certain times

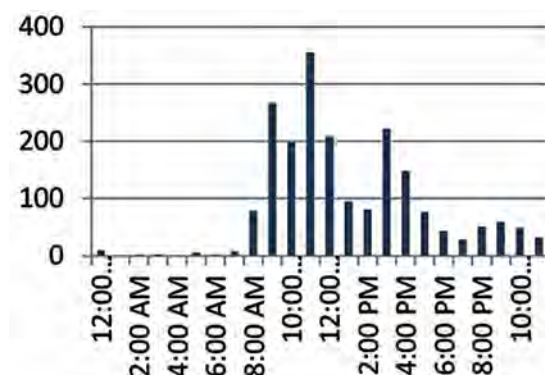


Table 1. Comments generated for the different type of items

	Created	Comments Generated
Forum	17	47
Pictures	106	12
Blogs	87	2069
Videos	105	231

Table 2. Guidelines for the social virtual classroom

Guideline	Rationale
Allow students to register as anonymous users	This guideline which ensures that all registered users are anonymous is intended for students who may have a social, cultural, or academic difference which may contribute to low esteem of themselves. It also caters for students who have a barrier due to peer pressure or other circumstance. This was used to establish collaboration.
Post stimulating fundamental questions	In order to help students extract their knowledge and realise what they know and hence what they do not know, the questions posted by the instructor were not about solving questions but rather stimulating, thinking and opinionated questions. This approach was also intended to encourage collaboration among students.
Establish that every comment is valuable	A student's comment which is participation was encouraged. The accuracy of that comment was not considered. This approach was used in an attempt to relate to all students' different backgrounds.
Motivate rebuilding of comments by collaboration	This guideline emphasised collaboration since strength can be obtained by variety. When a student freely shares his/her knowledge and other students contribute by making comments to that knowledge, the student can reflect on those comments which may then give him/her the opportunity to re-create their knowledge.
Provide quick feedback	The guideline of quick feedback was aimed at encouraging students that other students are important and need to be heard so that they could expect a quick response. In most cases, a comment would be answered within 2 days by either a student or an instructor.
Ensure 24/7 and absolutely free	An important consideration in this experiment was to ensure the approach was convenient and affordable. This approach must be adaptable to each student's constraints in terms of time, place and cost. An internet-based application has the features of 24/7, anytime and anywhere. <i>SocialGo</i> is a free social networking application.
Attract student interest by posting a video on the social community	A student can feel a sense of satisfaction if they are the first to post a video for other students to see. This guideline is intended to help those students who are weaker academically so that they can still participate in building the social community. The feeling of belonging and creating may have positive influences on a student.
Engage students through interesting non-work postings	Since these digital native students are so visual, relating to a video or social aspect is always appealing. If the student can be engaged with the videos then the student may have a spill over and may also engage with the subject matter.
Encourage students to provide feedback to one another	The students were provided the opportunity to comment on one another's comments thereby enabling the typical teacher-student role to be shifted since now the student can take up the role as the teacher. This approach is two-fold. The first is the virtual classroom requires a lot of time from the instructor, if only for the instructor to correct or make suggestions to comments. The confidence of a student being able to teach or identify weaknesses in the knowledge of others may boost a student's self esteem. The second is the students getting feedback.
Enforce that comments can be questions, answers, suggestions or illustrations	The guideline assumes that a student's knowledge is varied hence collaboration by answering or illustrating in the manner that is unique to that student can be beneficial. Any type of comment, questions, answers, suggestions or illustrations were all considered equally valuable. The idea is that the students can build on each other's knowledge
Reward students based on the quantity of comments, not on the quality	The guideline of rewarding students (by awarding marks based on usage) is based on rewards which are given when students are playing computer games. The reward is really to motivate the student to keep trying and the more the student tries, the more points (marks) will be obtained.

This VC demonstrated that the student can posted thinking questions that led into discussions at interesting times such as 2:00 am. The student feedback was prompt and many users were correcting or improving comments. At the end of the experiment, the students were asked to fill out a survey. Each question had four responses strongly agree, agree, disagree and strongly disagree. The

survey was divided into two categories using the VC and the community. Some of the questions from each of these two categories are shown in the Table 3 below. The names of the students were not collected in the survey.

From Table 3 (Mohan, 2013), it can be seen that 80% of the students strongly agreed that being anonymous was important to them, suggesting

Table 3. Student feedback

Using the Social Virtual Classroom	Strongly Agree	Agree	Disagree	Strongly Disagree
Were you encouraged to think outside the classroom by using the social virtual classroom?	40	38	22	0
Did reading other users' comments help you in your understanding?	59	30	11	0
Was being anonymous important to you?	80	17	3	0
Did the social virtual classroom help you to build more confidence in your work?	45	23	32	0
Would you prefer more courses to use this approach in teaching?	69	31	0	0
The Inclusion of the Social Community				
Did you believe that the social aspects of the social virtual classroom were helpful?	65	17	18	0
Was looking at other users' videos interesting to you?	44	30	21	5

that the anonymous feature was a very important feature. 59% of the students stated that reading other users' comments helped in understanding the subject matter. Another interesting result was that 69% of the students preferred more courses to use the VC approach in teaching. On the negative side, 32% of the students did not agree that the VC helped build more confidence and 22% of the students said that they were not encouraged to think outside the classroom by using the VC.

FUTURE RESEARCH DIRECTIONS

These experiments reflected that VCs led to an increase in the pass rate of 1st year Mathematics at the University of Trinidad & Tobago. Four VC were used. The 1st VC, all students had to use and the pass rate was 67%. The 2nd VC, students were given choice and the pass rate was 69%. Future research direction is to investigate whether there is any relationship between the frequency of the VC usage and the grade obtained. Another future research would be to investigate whether anonymity is a key factor in the VC. The VC should be instructor independent hence future research should confirm this by repeating the experiment but change the traditional classroom instructor.

The 2nd VC offered choice of with or without technology and the pass rate was increased by 29% as compared to 8% respectively. A VC with technology was more than three times successful as without technology. That result is quite significant therefore further research must be to repeat this experiment to confirm technology does indeed have a positive impact in teaching.

The 3rd VC was for repeating students with the constraint of no traditional classroom. Future research directions is to confirm the result is accurate by repeating a number of times. This can lead to an interesting option for teaching repeating students.

The 4th VC incorporated non-academic postings eg videos and pictures. Future research direction can be to investigate if culture and social postings can contribute in further enhancing the pass rate.

CONCLUSION

The experiments described in this chapter suggest that VCs do cause an increase in the pass rate of mathematics. The VCs was designed to stimulate students to think by posting stimulated opinionated questions. In the 2nd VC that the students who

chose no VC were given the same stimulated opinionated questions on paper which resulted in an increase from 40% to 48%. The 1st and 2nd VCs got a greater increase of 67% and 69% respectively suggesting that technology assisted students with their cognitive skills. This was possible with an environment fostering students' collaboration may motivate students to analyse and evaluate their existing knowledge. Thus enabling each student to build on their individual knowledge demonstrating self-centred learning. The success of the experiment was based on improved pass rate which was accomplished by the 4 VCs but the most impressive VC was the one 3rd VC which increased to 82%. This was the smallest group with 28 students and this was also repeating students. Another observation from the experiments was the students indicated that they were more inclined to use the VC since they were anonymous to other students. The prompt feedback encouraged the students to use the VC more frequently thus allowing them to evaluate and analyse their knowledge meta-cognitively.

REFERENCES

- Bloom, B. S. (1956). Taxonomy of educational objectives. In *Handbook I: The cognitive domain*. New York: David McKay.
- Brady, P., & Bowd, A. (2005). Mathematics anxiety, prior experience and confidence to teach mathematics among pre-service education students. *Teachers and Teaching: Theory into Practice*, 11(1), 37–46.
- Caribexams. (2004). *Caribbean Examination Council (CXC) mathematics pass rates*. Retrieved from http://www.caribexams.org/m_pass_rates
- Debellis, V., & Goldin, G. (2006). Affect and meta-affect in mathematical problem solving: A representational perspective. *Educational Studies in Mathematics*, 63(2), 131–147. doi:10.1007/s10649-006-9026-4
- Eng, T. H., Li Li, V., & Julaihi, N. H. B. (2009). *A Case Study of 'High-Failure Rate' Mathematics Courses and its' Contributing Factors on UiTM Sarawak Diploma Students*. Conference on Scientific & Social Research 2009, Paper 1.
- Garnham, C., & Kaleta, R. (2002). Introduction to Hybrid Courses. *Teaching with Technology Today*, 8(6). Retrieved September 22, 2009 from <http://www.uwsa.edu/ttt/articles/garnham.htm>
- Green-Evans, V. (2005). *Maths failures continue at UWI, review planned*. Retrieved Wednesday, November 25, 2015, Observer staff reporter, http://www.jamaicaobserver.com/news/74399_Maths-failures-continue-at-UWI--review-planned
- Gresham, G. (2007). A study of mathematics anxiety in pre-service teachers. *Early Childhood Education Journal*, 32(2), 181–188. doi:10.1007/s10643-007-0174-7
- Hannula, M. (2002). Attitude towards mathematics: Emotions, expectations and values. *Educational Studies in Mathematics*, 49(1), 25–46. doi:10.1023/A:1016048823497
- Hesse, M., & Mason, M. (2005). The case for Learning Communities. *Community College Journal*, 76(1).
- Hooper, S. (1992). Cooperative Learning and Computer-Based Instruction. *Educational Technology Research and Development*, 40(3), 21–38. doi:10.1007/BF02296840
- Huitt, W. (2007). *Maslow's hierarchy of needs*. Educational Psychology Interactive. Valdosta, GA: Valdosta State University. Retrieved from, <http://www.edpsycinteractive.org/topics/regsys/maslow.html>
- Kinach, B. (2002). Understanding and learning-to-explain by representing mathematics. *Journal of Mathematics Teacher Education*, 5(2), 153–186. doi:10.1023/A:1015822104536
- Maslow, A. H. (1943). A Theory of Human Motivation. *Psychological Review*, 50(4), 370–396. doi:10.1037/h0054346

McLeod, D. B. (1992). Research on affect in mathematics education: a reconceptualization. In D. A. Grouws (Ed.), *Handbook of Research on Mathematics Learning and Teaching* (pp. 575–596). New York: Macmillan.

Mohan, F. (2008). yourMathsCorner: A blog-based approach to learning prerequisite mathematical knowledge at the tertiary level. *International Journal of Mathematics and Computers in Simulation*, 2, 95–101.

Mohan, F. (2010). *Using Social Networking Software to Increase Students' Participation in a Virtual Classroom*. In J. Herrington & C. Montgomerie (Eds.), *Proceedings of EdMedia: World Conference on Educational Media and Technology 2010*. Association for the Advancement of Computing in Education (AACE).

Mohan, F. (2011). Building a cultural community classroom to connect instructors with students. *ICALT 2011, IEEE International Conference on Advanced Learning Technologies*, 147 - 149. doi:10.1109/ICALT.2011.49

Mohan, F. (2013). *Using a Social Learning Community to Actively Engage Students' Participation in a Virtual Classroom*. In *Cases on E-Learning Management: Development and Implementation* (pp. 50–70). Academic Press.

Niess, M. L. (2005). Preparing teachers to teach science and mathematics with technology: Developing a technology pedagogical content knowledge. *Teaching and Teacher Education*, 21(5), 509–523. doi:10.1016/j.tate.2005.03.006

Rieber, L., & Welliver, P. (1989). Infusing Educational Technology into Mainstream Educational Computing. *International Journal of Instructional Media*, 16(1), 21–32.

Rylands, L., & Coady, C. (2009). Performance of students with weak mathematics in first-year Mathematics and Science. *International Journal of Mathematical Education in Science and Technology*, 40(6), 741–753. doi:10.1080/00207390902914130

Shores, M. L., & Shannon, D. M. (2007). The Effects of Self-Regulation, Motivation, Anxiety, and Attributions on Mathematics Achievement for Fifth and Sixth Grade Students. *School Science and Mathematics*, 107(6), 225–236. doi:10.1111/j.1949-8594.2007.tb18284.x

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14. doi:10.3102/0013189X015002004

Sibawu, S. (2013). The zone of proximal development in the learning of Mathematics. *South African Journal of Education*, 22(2), 2013.

Singh, H. (2003). Building effective blended learning programs. *Educational Technology*, 43(6), 51–54.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Whannell, R., & Allen, B. (2012). First year Mathematics at a regional university: Does it cater to student diversity? *The International Journal of the First Year in Higher Education*, 3(2), 45–48. doi:10.5204/intjfyhe.v3i2.125

ADDITIONAL READING

De Wever, B., Van Keer, H., Schellens, T., & Valcke, M. (2007). Applying multilevel modelling to content analysis data: Methodological issues in the study of role assignment in asynchronous discussion groups. *Learning and Instruction*, 17(4), 436–447. doi:10.1016/j.learninstruc.2007.04.001

Díaz, A. L., & Entonado, B. F. (2009). Are the Functions of Teachers in e-Learning and Face-to-Face Learning Environments Really Different? *Journal of Educational Technology & Society*, 12(4), 331–343.

Driscoll, A., Jicha, K., & Hunt, A., Tichavsky, & L. Thompson, G. (2012). Can Online Courses Deliver In-class Results? A Comparison of Student Performance and Satisfaction in an Online versus a Face-to-face Introductory Sociology Course. *American Sociological Association*, 40(4), 312–331.

Dziuban, C. D., Hartman, J., Juge, F., Moskal, P. D., & Sorg, S. (2006). Blended learning enters the mainstream. In C. J. Bonk & C. Graham (Eds.), *The Handbook of Blended Learning: Global Perspectives, Local Designs* (pp. 195–209). San Francisco, CA: Pfeiffer Publications.

Rabab'h, B. S., & Veloo, A. (2015). Prediction of Mathematics Learning Strategies on Mathematics Achievement among 8th Grade Students in Jordan. *Asian Social Science*, 11(2), 276–283.

Rovai, A. A. (2002). A preliminary look at the structural differences of higher education classroom communities in traditional and ALN courses. *Journal of Asynchronous Learning Networks*, 6(1), 41–56.

Simmons, G. R. (2014). Business Statistics: A comparison of Student Performance in three Learning Modes. *Journal of Education for Business*, 89(4), 186–195. doi:10.1080/08832323.2013.836470

Yaratan, H., & Kasapoglu, L. (2012). Eight grade students attitude, anxiety, and achievement pertaining to mathematics lessons. *Procedia: Social and Behavioral Sciences*, 46, 162–171. doi:10.1016/j.sbspro.2012.05.087

KEY TERMS AND DEFINITIONS

Anonymous: Anonymous meant identity was unknown thus a student was free to share his/her understanding. It also provided a mystery within the VC. Many students registered as different user in order to get more usage. Thus developing an opportunity for a student to be creative.

Autonomy: The student is given many opportunity to make decisions thus taking control of his/her own learning. The student decides if to share his/her understanding or if to reflect and then comment on another student's comment or if to disguise as two different users.

Collaboration: Collaboration in the VC is building team spirit. It is an opportunity for students to come together and share their understandings and help build one another so that everyone has the same understanding which will lead them to pass mathematics.

Virtual Classroom Usage: Encouraging students to share his/her understanding so if that understanding was incorrect, it can be reconstructed and become correct. The usage is how many comments were made so the more comments the more a reward.

Virtual Classroom: The VC was an online environment where students came together for learning and/or teaching. This environment facilitated individual personal needs like convenience in time and location, motivating students to share understanding by making comments and engaging students in dialog and feedback and most important providing an opportunity for students to fix his/her background knowledge.

Integrated Paper-Based and Digital Learning Material for Smart Learners

E

Sabrina Leone

Università Politecnica delle Marche, Italy

SMART LEARNING, SMART LEARNERS, AND INCLUSIVE LEARNING

Smart learners are lifelong learners (Leone, 2010) whose potential is unleashed by the seamless use of smart technologies (i.e., smartphones, tablets, tablet PCs, sensor network nodes, contact-less smart cards, RFID and QR codes) to access huge amounts of open resources and connections, anywhere anytime. Personal, and personalised, smart technologies increase a learner's independence in a novel way, and makes the context for engaging in study more tailored and potentially self-directed (Middleton, 2015).

Smart learning encompasses any teaching and learning approach that flawlessly accommodates technology and enhances practice through social uses of new inclusive spaces. Fruitful interactions are the core of a smart learning cycle and are supported by a smart learning environment (Liu, Huang & Chang, 2015). Indeed, the current digital learning environment is gradually evolving into the smart learning environment (Li, Chang, Kravcik, Popescu, Huang, Kinshuk & Chen, 2015), that is a user-friendly space that facilitates easily accessible, appealing and effective learning. A learning environment may be considered smart when it includes adaptive technologies or innovative features and capabilities that improve understanding and performance. Specifically, features of smartness are

1. Conversational support for learners, teachers and designers,

2. Dynamic updating of student profiles, resources and databases, and
3. Automatic [re-]configuration of interfaces to adjust to different learners and learning situations (Spector, 2014).

The concept of smart learning includes that of ubiquitous learning (uLearning). The only difference stands in the higher power of next generation (smart) technologies that are creating disruptive learning landscapes and learners' profiles.

Today's students expect always-on, available-anywhere information and personalised, multichannel learning. The term "classroom" is becoming more figurative than literal (IBM, 2015). In a smart learning environment the physical and virtual dimensions merge (Liu, Huang & Chang, 2015), and learning is inclusive.

Inclusive education is an essential component of lifelong learning; it is concerned with an individual's effective participation in society and with the achievement of his/her full potential. The affordances of new educational technologies can enable the development of uLearning environments and of multimodal learning contents that foster inclusion, personalisation and interaction, provided that a learner-centred and technology-enhanced approach is adopted.

Internationally, inclusive education is increasingly understood more broadly as a change, in a holistic approach, that supports and welcomes diversity (in race, economic status, social class, ethnicity, language, religion, gender, sexual orientation and ability) amongst all learners (UNESCO, 2009).

Since learning takes place in many contexts, formal, non-formal and informal, inclusive and quality education become synonyms and are vital for the development of more inclusive societies.

Specifically, quality learning is characterised by two important components: the learner's cognitive development, and the promotion of values and attitudes of active citizenship and/or of creative and emotional development.

An inclusive curriculum is based on the four pillars of education for the 21st century – learning to know, to do, to be and to live together (Delors et al., 1996). Promoting inclusion means stimulating discussion, encouraging positive attitudes and improving educational and social frameworks. This involves changes in content, approaches, structures and strategies in order to provide all learners with flexible and personalised learning to meet individual needs, abilities and learning styles.

uLearning, supported by the growing diffusion of wireless smart technologies and institutional policies, is becoming more and more a modality of flexible and participatory learning to be adopted in and out of the classroom exploiting smartphones, tablets, tablet PCs, sensor network nodes, contactless smart cards, RFID (El-Bishouty, Ogata, & Yano, 2007) and QR codes.

Thanks to this technological growth, a personal learning environment could be embedded in everyday life (Ogata & Yano, 2004) and become a Computer Supported Ubiquitous Learning (CSUL) environment, characterised by permanency, accessibility, immediacy, interactivity, situatedness and adaptability (Curtis, Luchini, Bobrowsky, Quintana, & Soloway, 2002; Leone & Leo, 2011a). Learning theories for CSUL are authentic learning (Brown, Collins & Duguid, 1989), situated learning (Lave, & Wenger, 1991) and learning by doing (Schank, 1995).

It is widely acknowledged that information and communication technologies (ICT) enrich the learning experience (UNESCO, 2012). Anyhow, the focus has to be placed on learning, rather than

on technology in itself. In a technology-enhanced learning approach, the advantages arising from the integration of ICT in the learning curriculum have to be assessed within the learning experience, the usefulness of learning and its enhancements (Leone, 2008; Leone & Leo, 2011a).

Pedagogical and psychological researchers have debated for decades on a common understanding of “effective learning”. According to recent literature (Bulu & Yildirim, 2008; Calvani, 2006; Ellis, 1999; Liu, Huang & Chang, 2015; Wasson, 2007), social interaction among learners is a major element of the learning process, indeed, it can decisively impact on learning outcomes (Agostinho, Lefoe, & Hedberg, 1997).

Cooperation is an essential factor in the construction of an “effective learning” environment since it engages students in knowledge construction through interaction and negotiation with their peers. Cooperation enables learners to discuss, argue, agree and reflect on ideas, principles and knowledge. In the design of a suitable –situated, real – learning environment prior attention has to be paid to knowledge construction and effective learning, that is to learning relevant for learners (Johnson & Johnson, 1994).

Smart learning emphasises technology-embedded learning to enhance cooperation and interaction between learners (Bae, Shin, Kim & Choi, 2015).

This work aims to illustrate the *QRcode* format, a framework that supports uLearning by the integration of paper-based and digital learning material through Quick Response (QR) code. The format was devised within the research project *Learning4All* (2009-2012) and was validated by several learning experiences of English as a foreign language (EFL) for different clusters (Leone, 2014). Subsequently, the format was selected as element of techno-pedagogical innovation in the *Eureka* project (2012-2014), a network of 11 schools in Apulia, Italy, for the enhancement of curriculum continuity from middle into high school.

INTEGRATED PAPER-BASED AND DIGITAL LEARNING MATERIAL: LITERATURE REVIEW

Paper and traditional books have been used as basic tools in developing knowledge-intensive tasks and learning (Chao & Chen, 2009). However, a paper textbook can be combined with ubiquitous technologies in a whole to deepen reading comprehension and to enrich it with audio, video and grammar, vocabulary and cultural/technical/professional in-depth contents.

Paper-based learning material has been successfully enhanced by multimedia contents in experiences on annotation conducted through digital pen (Chao & Chen, 2009; Lai, W.-C., Chao & Chen, 2007). More recently, practitioners and researchers have shown growing interest for the potential of QR code as ubiquitous learners' tool, and several learning experiences have been conducted in different contexts: outdoor education (Lai, Chang, Wen-Shiane, Fan & Wu, 2013; Law & So, 2010) and outdoor students' assessment specifically (Conejo, Perez de la Cruz, Barros, Galvez & Garcia-Viñas, 2013); integrated paper-based and digital learning materials to enhance listening comprehension in foreign language learning (Law & So, 2010), to enhance reading comprehension (Chen, Chia-En Teng, & Lee, 2010;), to provide support in maths homework (McCabe & Tedesco, 2012), to convey directions to English language learners and recordings for students who have difficulty in reading (Shumack, Lewis, Simmons & Carpenter, 2013), and to encourage students' interaction during face-to-face lectures (Law, 2013). Most common uses include access to web sites with course information and study materials (Bobeva & Hopkins, 2012). Nevertheless, little literature (Leone & Leo, 2011a; Leone & Leo, 2011b; Leone, 2012; Leone, 2014) is available about the principles of instructed learning in the use of paper-based learning material integrated with digital material through QR code within a definite learning format or model.

Thanks to its two dimensions (vertical and horizontal), the QR code can store greater amounts of information and services (i.e., website addresses, text, contact details) (Ramsden, 2009; Savarani & Clayton, 2009) that learners can readily, anywhere and anytime, decode by a mobile device with an embedded camera and code reading software installed. Besides fostering flexibility of provision, the integration of QR codes with paper-based learning material also offers personalisation of learning because different learning styles and approaches to the use of ICT for learning can be accommodated.

THE QR CODE FORMAT

QRcode is a technology-enhanced learning format that was devised within the research project *Learning4All* (2009-2012). The project investigated on how an aware adoption of ICT can contribute to improve the quality of teaching, in particular for students with special needs (UNESCO, 2009). *Learning4All* included seven research units: Politecnico di Milano - coordinator -, IMATI CNR di Genova and the Università di Bari, di Bologna, di Perugia, del Salento and Politecnica delle Marche (UNIVPM).

The *QRcode* format consists in the integration of paper-based and digital learning material through QR code, aimed at facilitating personalised and flexible (anytime, anywhere) learning. Multimedia contents are coded by the teacher in QR code and are then easily accessible by a decoding software which is supplied with many mobile devices or is downloadable free (Leone & Leo, 2011b). The *QRcode* format is potentially adaptable to all disciplines and was implemented in several learning experiences of EFL for different clusters, with good results: in February-June 2010 in three different scenarios (adult lifelong learners, in-service teachers and secondary school students) (Leone & Leo, 2011a), in February-March 2011 with 4 classes of a secondary school and in March-

April 2012 with 2 classes of a secondary school for surveyors (in an interdisciplinary module EFL – Building Construction) (Leone, 2014). Further, as element of techno-pedagogical innovation in the *Eureka* project (2012-2014), the format was implemented in the following sessions: in October-November 2012, initial training of 80 in-service teachers from the many partner institutes; in April-May 2013 and in January-February 2014, 2 interdisciplinary paths with 8 3rd-year classes of a middle school together with 8 1st-year classes of high school.

Needs, Objectives, and Competences

General learning objectives of the *QRcode* format are: learning to learn, promoting the deepening of contents in different contexts and by different means, favour interdisciplinarity of knowledge domains, favour continuity between in-class and out-of-class activity (integration between formal and informal learning). Specific learning objectives are to: facilitate the acquisition of basic, communicative and multimedia, and digital competences; favour a better general/specific comprehension of one or more of the topics developed; support motivation, participation and interest for a discipline; promote cooperative and collaborative learning. The *QRcode* format can develop basic and higher cognitive and social competences (knowing and understanding the topic treated, communicating adequately in relation to the context, retrieve information from various sources and re-elaborate it in a personal way, meet a new culture in a view of cultural pluralisme), crosscurricular (autonomous study and in-depth analysis, problem solving, relational skills), of citizenship and digital (understanding the affordances of ICT for knowledge sharing and collaborative construction) skills.

Learning Strategies

The *QRcode* format can be developed through exercises, brainstorming (encoding in QR code

of forums and wikis), collaborative and cooperative learning, learning by doing, problem solving, webquest, game based learning, simulations in immersive virtual environments. The format can be an effective inclusive learning tool since it facilitates personalised and flexible learning if a learner-centred approach is adopted, an approach attentive to each learner's diverse needs (UNESCO, 2009). A metacognitive work that allows to enhance autonomy and responsibility could be a first step towards inclusion.

Organisation

The physical places to implement the *QRcode* format are the classroom with a wireless connection or a hot spot, where learners operate, individually or in a group, on paper-based and digital learning material, by the various mobile devices available, and any other place by mobile devices with an Internet connection. The learning environment is, therefore, both physical (classroom and class) and virtual (web-based, with interactive activities, and/or 3D) according to what the teacher adopts. The length of the implementation can vary in relation to the number and the extension of the uLearning modules of the designed curricular path. Necessary human resources are the teacher, that defines strategies, objectives, contents and activities, and the ICT technician, that installs/checks the QR code decoder software in the mobile devices in use, checks the Internet connection and deals with troubleshooting.

Evaluation and Assessment

The *QRcode* format is monitored at the beginning and at the end of its implementation, in order to evaluate its effectiveness and rating from the participants' point of view (evaluation), and in terms of achievement of learning objectives (assessment). For the evaluation, the data are collected (1) by entry and exit anonymous surveys, to record students' digital skills, motivation and expectations on the format (entry survey), and students' feedback (exit survey); (2) by interviews, to record

teachers' motivations for the choice of the format, expectations and context of implementation (entry interviews) and teachers' achieved results and impressions (exit interviews). Assessment is carried out by close, semi-structured and/or open tests, and it consists in an entry test (prior knowledge and competences), formative tests (at the end of each learning unit) and a summative test (at the end of the learning path).

Recommendations for the Optimal Implementation of the Format

The *QRcode* format is a very flexible tool. Nevertheless, thorough organisation and management of hardware and software are success key factors of the experience. Moreover, relevant elements are: a learner-centred and technology-enhanced approach, drawn on constructivism (Jonassen & Land, 2000); an adequate familiarisation of the students with learning tools and environments; check of the accessibility of the necessary devices; modulation of space/time of class and autonomous work; collaboration with other teacher to create an interdisciplinary path; engagement of the educational institution as a whole as stakeholder of techno-pedagogical innovation.

SWOT Analysis

Strengths of the *QRcode* format are the provision of personalised and flexible learning. Weaknesses are initial technological difficulties that can cause demotivation and disengagement. Opportunities are many: opening to a wide range of learning paper-based and digital materials; use of open software and learning materials; possible effective integration of the format in the curriculum; possible spur for interdisciplinary activities. Vice-versa, possible threats are a teacher-centred approach, that risks to void the flexibility of the format, and the limits of the participants' mobile devices (speed of the processors, small displays, Internet connection costs), even though the technological standard of next generation (smart) devices is high.

CHARACTERISTICS AND OUTCOMES OF THE RESEARCH EXPERIENCES WITH THE FORMAT



All the experiences of implementation of the *QRcode* format have been conducted in Italy for:

1. Nine different clusters of learners of EFL:
 - a. In February-June 2010 in a refresher course for secondary school teachers (23 participants, 10 weekly 3-hour lectures), a language certification course for Italian secondary school students (upper classes, 16 participants, 17 3-hour lectures twice a week) and a course for Italian adult beginners (15 participants, 20 weekly 3-hour lectures). Participants were 54 in all. The author was teacher-facilitator and tutor of the research experience at the same time;
 - b. In February-March 2011 (4 weeks, 9 hours in all for each class) with 4 classes (second- and forth-year students) of a secondary school, 85 learners in all. The author was tutor of the research experience, in support of the two curricular teachers;
 - c. In March-April 2012, on the basis of the preceding research experiences, an implementation enriched with cooperative work, on an interdisciplinary module EFL-Building Construction (4 weeks, 3 hours a week for each class) with 2 upper classes of a secondary school for surveyors, 45 learners in all. The author was the coordinator of the research experience and English teacher, at the same time;
2. Three clusters of learners in interdisciplinary paths within the *Eureka* project (2012-2014) (the author was the coordinator of the techno-pedagogical innovation of the project, one of the teachers' trainers and tutor in the last two experiences):

- a. In October-November 2012 in the initial training of 80 in-service teachers from the many partner institutes (2 in-presence 4-hour workshops and the ongoing distributed community of practice *Comunità di Pratica Progetto Eureka*, in the eLearning space www.elearnigplace.it/corsi);
- b. In April-May 2013 (4 weeks, 9 hours a week for each class) an interdisciplinary path on Italian, music, history, geography, French and maths with 4 3rd-year classes of a middle school together with 4 1st-year classes of high school, 193 learners in all;
- c. In January-February 2014 (4 weeks, 9 hours a week for each class) an interdisciplinary path on citizenship education, geography, Italian, the arts and music with 4 3rd-year classes of a middle school together with 4 1st-year classes of high school, 187 learners in all.

In all the implementations the participants used their own mobile devices; in addition laptops were available at schools. All the courses were developed in uLearning on curricular in-presence and distance learning modules. Personalisation was realised by a great variety of graded learning materials (on the basis of the entry assessment

test and evaluation survey, and proposed in hard copy and pdf file, both integrated with in-depth digital contents in QR code), web-based interactive activities and self-assessment.

In the interdisciplinary implementation EFL-Building Construction of March-April 2012 (Leone, 2012) the teaching team defined strategies, objectives, contents and activities, and enriched the experience with cooperative learning, either in presence and in eLearning (*PlaLE – Personal language Learning Environment*, in <http://www.elearningplace.it/corsi/course/category.php?id=14>), with which learners had already familiarised in EFL since the beginning of the school year.

Specifically, objectives of the implementation of the *QRcode* format in all cooperative and interdisciplinary implementations were metacognition, the synergic enhancement of disciplinary, technical-professional (when applicable) and language (when applicable) skills, the acknowledgement of the importance of cooperative learning in the participants' personal growth, the acquisition of new learning tools for cooperative work, the activation of students' participation in group work, the enhancement of peer-scaffolding.

Contents included domain knowledge, micro-language (if applicable), communicative functions (for foreign languages), inherent grammar and vocabulary, and cultural in-depth topics.

Figure 1. Integrated activities in “Talking about the past”

Week 1/class 2: Wednesday 9th (5B) and Thursday 10th (5A) February
Expressing a sequence of events - past tenses revision (I)

Activity 1 - An apartment in London.
 Read the following text and turn it into the past.



This apartment is in the heart of Chelsea, one of the most fashionable areas of London. It has two beautifully furnished bedrooms and a fully fitted kitchen. It is not far from Buckingham Palace, the Royal Albert Hall, Knightsbridge, Sloan Square and the Natural History and Victoria and Albert Museums. The nearest underground stations are South Kensington and Sloan Square. The minimum stay is seven nights. There is a direct dial telephone and satellite tv. A maid service is guaranteed five days a week. You can also have a wireless broadband Internet connection. There is also off-street parking for residents.

Activity 2
 unit 9 - 10: ex. 3 - the simple past
 of regular verbs + ex. 5 - the
 simple past of irregular verbs



Figure 2. Content of Activity 2 (ex. 3), in QR code in Figure 1

Module B
Units 9-10

GRAMMAR
3. The Simple Past tense of regular verbs
Write down the sentences using the prompts given. Remember to use the Simple Past. Then click on OK.

1. Greetings and reply
2. The past tense of 'to be'
3. The Simple Past tense of regular verbs
4. Questions with 'How'
5. The Simple Past tense of irregular verbs
6. Similarity
7. Don't speak too fast!
8. Let's talk about holidays
9. My first novel

Practice **Test**

Module B > Units 9-10 > 3. The Simple Past tense of regular verbs

Margaret Layton, Marina Spiazzi, Marina Tavella - Slides ZANICHELLI

You / to order / a pizza for dinner / last night?	
What subject / Mark / to study / yesterday afternoon?	
We / to wait / for them in front of the cinema / last night.	
Miranda / to watch / a beautiful film on TV / yesterday.	
David / to live / in New York / last year?	
You / not to lock / the door / this morning.	
Bob / not to invite / Polly / to his birthday party.	
Liz / to look / very tired / last Sunday.	

Glossary | Grammar | Print | Exit

Figures 1 and 2 show an extract of the activities proposed in EFL for the communicative function “Talking about the past” and for those codified in *QR code*, respectively.

In all the 9 experiences in EFL, learners carried out an extensive range of tasks in relation to the four language skills (see details in Leone & Leo, 2011a; Leone, 2012; Leone, 2014).

As a whole, results of all the experiences emphasise that learners’ motivation and active participation grew as their awareness, self-confidence and familiarisation with the learning environment improved. In the exit interviews the teachers, too, expressed full satisfaction with the students’ participation and learning results (even learners who were usually demotivated progressed in terms of cognitive, digital and crosscurricular skills), and with their personal and professional empowerment.

The main difficulties were technical-organisational ones and in a few occasions due to the teacher’s mismatching approach; in particular:

- Teenagers showed to be false digital natives; as a result, even though an extensive phase of familiarisation preceded the start of the experience, during the first and second week 40% and 10% of the time available respectively was forcedly dedicated troubleshooting on the mobile devices;
- Because of a recurrent insufficient Internet connection in some of the classrooms used, it was necessary to move the learners in alternative locations;
- In some sessions a lack of the already configured devices occurred (participants simply forgot to bring them);
- In some sessions learners were rather passive and disoriented because of discontinuity in the learning path (decontextualised activities) and insufficient teacher’s facilitation (guide, scaffolding and feedback during the tasks);
- Insufficient engagement and interest for the experience within the teachers’ institutes.

The results obtained from the implementation of the *QRcode* format in the different scenarios show an important positive impact on different aspects: the uLearners' disciplinary skills, management of ICT tools, satisfaction levels with the flexibility and personalisation of learning, as well as on the contents proposed and cross-curriculum objectives such as developing autonomy, building learning confidence, empowerment, a positive attitude and motivation towards a new way of learning, results which come to support the effectiveness of the model implemented (Leone & Leo, 2011a; Leone, 2012).

FUTURE RESEARCH DIRECTIONS

In the next future, the *QRcode* format could be extensively used in different grades of instruction and in teachers' continuous professional training. In addition, a pilot implementation of the format with augmented reality is starting. In order to fully exploit the potential of the format, scaffolding might be provided as an added service to tackle large margins of improvement that have emerged in the participants' digital skills, in their approach to teaching and learning, in their approach to the adoption of ICT in the curriculum as a chance of holistic change and growth towards smart inclusive education.

CONCLUSION

Personal, and personalised, smart technologies increase lifelong learners independence in a novel way, makes the context for engaging in study disruptive, more tailored and potentially self-directed, and allow for access to huge amounts of open resources and connections, anywhere anytime.

The affordances of smart technologies can enable the development of uLearning environ-

ments and of multimodal learning contents that foster inclusion, personalisation and interaction, provided that a learner-centred and technology-enhanced approach is adopted.

The *QRcode* format has been devised with the aim of providing personalisation and empowerment to ubiquitous lifelong learners facilitating the acquisition of some of the skills necessary for the 21st century.

The format is very flexible for both the teacher-facilitator and the learners, and can be enriched with interdisciplinary and cooperative learning experiences.

The outcomes of all the research experiences carried out to date highlight an important positive impact on the participants' disciplinary competences, acquisition of skills in the use of new digital tools, satisfaction for flexibility and personalisation of learning and for the contents and the crosscurricular competences, and validate the potential of the implemented framework. However, success keys of the experience are: (1) punctual organisation and management of the necessary hardware and software; (2) a suitable students' familiarisation with the learning environment (technology, tools and learning approach); (3) a learner-centred approach; (4) a technology-enhanced, rather than technology-driven, approach.

REFERENCES

- Agostinho, S., Lefoe, G., & Hedbeg, J. (1997). Online Collaboration and Problem Solving for Effective Learning: A Case Study of a Post Graduate University Course. *Proceedings of AUSWEB97 The Third Australian World Wide Web Conference*. Lismore: Southern Cross University.
- Bae, Y., Shin, S., Kim, M., & Choi, I. (2015). A Study on Instructional Design of Global Connected Learning Based on Smart Education. *Advanced Science and Technology Letters*, 103, 261-265.

- Bobeve, M., & Hopkins, D. (2012). Embedding QR Codes in Education Contexts. Proceedings of Innovation in challenging times: The ABS Learning & Teaching Conference. Manchester Metropolitan University Business School.
- Brown, J. S., Collins, A., & Duguid, P. (1989). *Situated Cognition and the Culture of Learning. In Educational Researcher* (pp. 32–42). Academic Press.
- Bulu, S. T., & Yildirim, Z. (2008). Communication Behaviors and Trust in Collaborative Online Teams. *Journal of Educational Technology & Society*, 11(1), 132–147.
- Calvani, A. (2006). *Reti, comunità e conoscenza. Costruire e gestire dinamiche collaborative*. Trento: Erikson.
- Chao, P.-Y., & Chen, G.-D. (2009). Augmenting paper-based learning with mobile phones. *Interacting with Computers*, 21(3), 173–185. doi:10.1016/j.intcom.2009.01.001
- Chen, N. S., Chia-En Teng, D., & Lee, C. H. (2010). Augmenting Paper-based Reading Activities with Mobile Technology to Enhance Reading Comprehension. *Proceedings of WMUTE2010*. doi:10.1109/WMUTE.2010.39
- Conejo, R., Perez de la Cruz, J. L., Barros, B., Galvez, J., & Garcia-Viñas, J. I. (2013). Context-aware Assessment Using QR-codes. *Proceedings of Research in Engineering Education Symposium 2013*.
- Curtis, M., Luchini, K., Bobrowsky, W., Quintana, C., & Soloway, E. (2002). Handheld Use in K-12: A Descriptive Account. *Proceeding of the International Workshop on Wireless and Mobile Technologies in Education*. Los Alamitos, CA: IEEE Computer Society. doi:10.1109/WMTE.2002.1039217
- Delors, J. (1996). *Learning: the Treasure Within. Report to UNESCO of the International Commission on Education for the Twenty-first Century*. Paris: UNESCO.
- El-Bishouty, M. M., Ogata, H., & Yano, Y. (2007). PERKAM: Personalized Knowledge Awareness Map for Computer Supported Ubiquitous Learning. *Journal of Educational Technology & Society*, 10(3), 122–134.
- Ellis, R. (1999). *Learning a second language through interaction*. Amsterdam: John Benjamins Publishing Company. doi:10.1075/sibil.17
- IBM. (2015). *Education for a smarter planet*. Retrieved December 5, 2015 from http://www.ibm.com/smarterplanet/us/en/education_technology/ideas/
- Johnson, D. W., & Johnson, R. (1994). *Learning Together and Alone: Cooperative, Competitive and Individualistic Learning*. Englewood Cliffs, NJ: Prentice-Hall.
- Jonassen, D. H., & Land, S. M. (2000). *Theoretical Foundations of Learning Environment*. Lawrence Erlbaum Associates.
- Lai, H.-C., Chang, C.-Y., Wen-Shiane, L., Fan, Y.-L., & Wu, Y.-T. (2013). The implementation of mobile learning in outdoor education: Application of QR codes. *British Journal of Educational Technology*, 44(2), E57–E62. doi:10.1111/j.1467-8535.2012.01343.x
- Lai, W.-C., Chao, P.-Y., & Chen, G.-D. (2007). The Interactive Multimedia Textbook: Using A Digital Pen to Support Learning for Computer Programming. *Proceedings of ICALT, Seventh IEEE International Conference on Advanced Learning Technologies (ICALT 2007)*, 742–746. doi:10.1109/ICALT.2007.245
- Lave, J., & Wenger, E. (1991). *Situated Learning. Legitimate Peripheral Participation*. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511815355
- Law, C., & So, S. (2010). QR codes in education. *Journal of Educational Technology Development and Exchange*, 3(1), 85–100.

- Law, R. (2013). Interactive lectures using quick response codes. *Proceedings of the 18th ACM conference on Innovation and technology in computer science education (ITiCSE '13)*. ACM.
- Leone, S. (2008). The use of new technologies in advanced Italian classes. I. Olney, G. Lefoe, J. Mantei, & J. Herrington (Eds.), *Proceedings of the Second Emerging Technologies Conference 2008* (pp. 120-129). Wollongong: University of Wollongong.
- Leone, S. (2010). F2F learning vs eLearning: the lifelong learner's point of view. *Proceedings of INTED 2010*.
- Leone, S. (2012). Il format *QRcode* in un'esperienza di apprendimento cooperativo e interdisciplinare. *Proceedings di Didamatica 2012*.
- Leone, S. (2014). The *QRcode* format: a tool for inclusive, personalised and interdisciplinary learning experiences. In M. Khosrow-Pour (Ed.), *Encyclopaedia of Information Science and Technology* (3rd ed.). Hershey, PA: IGI Global.
- Leone, S., & Leo, T. (2011a). The synergy of paper-based and digital material for ubiquitous foreign language learners. *Knowledge Management & E-Learning: An International Journal*, 3(2).
- Leone, S., & Leo, T. (2011b). Il format *QRcode* in Learning for All. *Proceedings di Didamatica 2011*.
- Li, Y., Chang, M., Kravcik, M., Popescu, E., Huang, R., Kinshuk, & Chen, N.-S. (Eds.). (2015). *State-of-the-Art and Future Directions of Smart Learning*. Singapore: Springer.
- Liu, X., Huang, R., & Chang, T.-W. (2015). Design of theoretical model for smart learning. In *State-of-the-Art and Future Directions of Smart Learning*. Singapore: Springer.
- McCabe, M., & Tedesco, S. (2012). Using QR Codes and Mobile Devices to Foster an Inclusive Learning Environment for Mathematics Education. *International Journal of Technology and Inclusive Education*, 1(1), 37–43. doi:10.20533/ijtie.2047.0533.2012.0006
- Middleton, A. (2015). Thinking about smart learning. In *Smart learning: Teaching and learning with smartphones and tablets in post compulsory education*. Sheffield, UK: Media-Enhanced Learning Special Interest Group and Sheffield Hallam University.
- Ogata, H., & Yano, Y. (2004). *Context-Aware Support for Computer-Supported Ubiquitous Learning*. Paper presented at the 2nd IEEE International Workshop on Wireless and Mobile Technologies in Education, JhongLi, Taiwan.
- Ramsden, A. (2009). *The potential of QR Codes in Education*. *JISC Emerge Users & Innovation Programme Meeting*. Retrieved August 12, 2009, from <http://www.slideshare.net/andyramsden/potential-qr-codes-education-emerge-2009-presentation>
- Saravani, S.A., & Clayton, J.F. (2009). A conceptual model for the educational deployment of QR codes. *Same places, different spaces. Proceedings Ascilite Auckland 2009*.
- Schank, C. (1995). *What We Learn When We Learn by Doing*. Technical Report No. 60. Northwestern University, Institute for Learning Sciences.
- Shumack, K., Lewis, T., Simmons, K., & Carpenter, L. (2013). QR Codes in K12 and Teacher Education. In R. McBride & M. Searson (Eds.), *Proceedings of Society for Information Technology & Teacher Education International Conference 2013* (pp. 3789-3790). Chesapeake, VA: AACE.

Spector, J. M. (2014). Conceptualizing the emerging field of smart learning environments. *Smart Learning Environments*, 1(1), 1–10. doi:10.1186/s40561-014-0002-7

UNESCO. (2009). *Policy Guidelines on Inclusion in Education*. Paris: UNESCO.

UNESCO. (2012). *Education for All Global Monitoring Report 2012*. Paris: UNESCO.

Warschauer, M. (Ed.). (1996). *Telecollaboration in foreign language learning*. Honolulu, HI: University of Hawai'i Second Language Teaching and Curriculum Center.

Wasson, B. (2007). Design and Use of Collaborative Network Learning Scenarios: The DoCTA Experience. *Journal of Educational Technology & Society*, 10(4), 3–16.

ADDITIONAL READING

Barr, R., & Tagg, J. (1995). From Teaching to Learning: A New Paradigm for Undergraduate Education. *Change Magazine*, 2(12), 8–12. Retrieved January 11, 2008, from www.cic.uiuc.edu/resources/deo/paradigm.html

Bereiter, C. (2002). *Education and Mind in the Knowledge Age*. Mahwah, NJ: Lawrence Erlbaum Associates.

Bull, G. & Anstey, M. (2010). *Evolving pedagogies: Reading and writing in a multimodal world*. Carlton South: Education Services Australia.

Chapelle, C. A. (2001). *Computer Applications in Second Language Acquisition. Foundations for teaching, testing and research*. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9781139524681

European Commission. (2000). *E-learning – Designing Tomorrow's Education*. Brussels: European Commission.

European Commission. (2001). *Making a European area of lifelong learning a reality*. Brussels: European Commission.

Felder, R. M., & Silverman, L. K. (1988). Learning and Teaching Styles in Engineering Education. *English Education*, 78(7), 674–681.

Gimeno, J. (Ed.). (2010). *Saberes e incertidumbres sobre el curriculum*. Madrid: Morata.

Keegan, D. (2004). *The incorporation of mobile learning into mainstream education and training*. Paper presented at The 18th Asian Association of Open Universities Annual Conference.

Koehler, M. J., & Mishra, P. (2009). What is technological pedagogical content knowledge? *Contemporary Issues in Technology & Teacher Education*, 9(1), 60–70.

Kukulska Hulme, A., Sharples, M., Milrad, M., Arnedillo-Sánchez, I., & Vavoula, G. (2009). Innovation in Mobile Learning: A European Perspective. *International Journal of Mobile and Blended Learning*, 1(1), 13–35. doi:10.4018/jmbl.2009010102

Latour, B. (2005). *Reassembling the social: an introduction to Actor-network theory*. New York: Oxford University Press.

Leone, S., & Guazzaroni, G. (2010). Pedagogical sustainability of interoperable formal and informal learning environments. In F. Lazarinis, S. Green, & E. Pearson (Eds.), *Developing and utilizing e-learning applications*. IGI-Publisher. doi:10.4018/978-1-61692-791-2.ch001

Leone, S., Leo, T., & Chen, N.-S. (2010). An integrated model of synchronous cyber assessment and blended learning environment for foreign language learners. *Proceedings of ICALT 2010*, Sousse (Tunisia), 5–7 July 2010. doi:10.1109/ICALT.2010.38

Levy, M. (1997). *Computer-Assisted Language Learning: Context and Conceptualization*. Oxford: Clarendon Paperbacks.

Middleton, A. (Ed.). (2015). *Smart learning: Teaching and learning with smartphones and tablets in post compulsory education*. Sheffield, UK: Media-Enhanced Learning Special Interest Group and Sheffield Hallam University.

Molyneux, P., & Godinho, S. (2012). This is my thing!: Middle years students engagement and learning using digital resources. *Australasian Journal of Educational Technology*, 28(8), 1466–1486. doi:10.14742/ajet.782

Nalder, J. (2008). The dawn of uLearning. Master thesis. Retrieved March 10, 2010 from <http://www.scribd.com/doc/12398804/The-dawn-of-uLearning-Jonathan-Nalder-Masters-thesis>

Nix, J. (2005). The development of mobile learning for smartphones. *Proceedings of IADIS International Conference Applied Computing*, Algarve, Portugal, February 22-25, 2005.

Nunan, D. (1991). *Language teaching methodology*. UK: Prentice Hall.

Ramírez, E., Clemente, M., Cañedo, I., & Martín, J. (2012). Incorporating Internet resources into classroom practice: Secondary school teacher action plans. *Australasian Journal of Educational Technology*, 28(8), 1433–1450. doi:10.14742/ajet.780

Riischoff, B. & Ritter, M. (2001). Technology-enhanced Language Learning: Construction of Knowledge and Template-based Learning in the Foreign Language Classroom. *Computer Assisted Language Learning* (14/3-4), 219-32.

Trentin, G. (2005). From “formal” to “informal” e-Learning through knowledge management and sharing. *Journal of e-Learning and Knowledge Society*, 1(2), 209-217.

Wexler, S., Brown, J., Metcalf, M., Rogers, D., & Wagner, E. (2008). *360° Report: Mobile learning*. Santa Rosa, USA: eLearning Guild.

KEY TERMS AND DEFINITIONS

CSUL (Computer Supported Ubiquitous Learning) Environment: Everyday ubiquitous learning informed by the theories of authentic learning (Brown, Collins & Duguid, 1989), situated learning (Lave, & Wenger, 1991) and learning by doing (Schank, 1995).

Formal Learning: Hierarchically structured, chronologically graded educational system running from primary through to tertiary institutions.

Inclusive Learning: Learning that allows to meet not only special needs, but also diverse needs (e.g., different learning styles).

Informal Learning: Unstructured learning that allows persons to acquire attitudes, values, skills and knowledge from daily experience, within the individual's environment (i.e., family, friends, peer groups, etc.).

Lifelong Learning: A holistic vision of learning in different contexts (formal, non-formal and informal) and throughout life, based on the evolution of provider-driven education toward personalised learning and aiming at improving knowledge, skills and competencies within a personal, civic, social and/or employment-related outlook.

QR (Quick Response) Code: A bidimensional code (it displays information in both vertical and horizontal directions) that can hold larger amounts and different kinds of contents (e.g., website addresses, texts, numerical information, contact details) than a normal bar code (monodimensional). The information stored in a QR code can be readily decoded and accessed by a mobile device with an embedded camera and free code reading software installed.

QRcode Format: A technology-enhanced learning format to be implemented in a learner-centred learning environment to offer inclusive, personalised (different learning styles and goals) and flexible (anytime, anywhere) learning by the integration of paper-based and digital learning materials through QR code.

Ubiquitous Learning: Wireless learning supported by a large number of cooperative small nodes with computing and/or communication capabilities (e.g., handheld devices, sensor network

nodes, contact-less smart cards, RFID and QR codes) and characterised by high mobility and embeddedness.

E

Leveraging Technology–Enhanced Teaching and Learning for Future IS Security Professionals

Ciara Heavin

University College Cork, Ireland

Karen Neville

University College Cork, Ireland

Sheila O’Riordan

University College Cork, Ireland

INTRODUCTION

The use of social media technologies to connect with peers/colleagues is prevalent amongst students and practitioners alike. These technologies are being used to share ideas, content, resources, and experiences for both social and professional purposes. However, modern learning environments do not always implement the latest technologies and are therefore failing to support the needs and career expectations of Generation 2020. Thus, technology enhanced learning is proving invaluable in creating interactive collaborative learning environments that can address the needs of future graduates. The social business gaming platform considered in this chapter leverages the social networking concept in an academic environment. This study was undertaken in order to develop Information Systems (IS) security skillsets through the creation and facilitation of social business gaming. The online business game required students to apply what they have learned to problem situations to further develop their understanding of IS security (ISS) topics. The problems posed required learners to prove their understanding of the material being taught in the traditional lecture, and then apply what they had learned in an online environment, allowing students to both collaborate and compete against their peers in a

series of challenges. The game was utilised as a part of the continual assessment process to evaluate group interaction, role-playing, competition and learning in an ISS assignment and facilitate the students to measure their own performances of understanding. Thus, the game was not just an assessment mechanism for grades, but also a learning tool. This chapter focuses on a group of final year undergraduate students completing Bachelor of Science in IS and outlines the online ISS environment used in the study.

BACKGROUND

Organisations actively use simulated environments to both test (e.g. psychometric) and train (e.g. virtual trading of stocks and case study analysis) employees. Medical and scientific educators actively promote the learning of these disciplines through simulation and modeling tools (Quellmalz & Pellegrino, 2009) but to date social gaming has not been widely applied as a learning aid for business and IS (security) graduates. This chapter endeavours to leverage social media technology to enhance and support the learning and assessment mechanisms utilised in an undergraduate final year ISS module with the objective of providing students with a practical proactive knowledge of

DOI: 10.4018/978-1-5225-2255-3.ch223

the implementation and management of ISS in business, an increasingly important and understudied topic (White et al., 2013). The chapter is structured as follows; the subsequent section considers the area of learning, focusing on the weaknesses associated with traditional learning and highlighting how learning tools may overcome many of these. Following this, the nature of ISS education is presented and the workplace of the future is considered with particular emphasis placed on the need for business graduates with skills in social media technology. The research approach is then outlined. The case is presented and discussed and finally attention is attributed to the conclusions of the study.

Traditional Approaches in Teaching and Learning

Traditional learning approaches dominate third level education, however, more recently these practices are complemented by alternative approaches to teaching and assessment. This includes the use of Web 2.0 technologies (i.e. podcasts, social network sites, media sharing platforms, etc.) as a means of active learning, to further support and engage the learner (Cao et al., 2013). Traditional learning, also known as the teacher-centered paradigm, is regarded as a learning environment that encourages passive learning (Barr & Tagg, 1995), does not develop problem-solving skills, and ignores the individual needs of the students (Hannum & Briggs, 1982). It could be argued that advances in technology, such as multimedia and virtual simulations, have left the traditional classroom trailing behind, with learners expecting more and more. Social media provides a solution to these problems by incorporating the collaborative attributes associated with Web 2.0 technologies (Schneckenberg, 2009). Instructors are redesigning learning and assessment mechanisms by leveraging the dynamic interactive capabilities that social media can provide, which ultimately helps improve the essential skills students require to become business and ISS professionals. The

widely accepted criticism of the teacher centred model is that the ‘what’ rather than the ‘how’ of the instruction is delivered (Goodlad, 1984). It is argued that problem-solving and other intellectual skills are difficult to incorporate into the traditional environment due to the very nature of the educational system. Factors such as space, the grouping of students according to grades, and the duration and size of classes all hinder the desired environment. Technology enhanced learning is not the ‘silver bullet’ solution to the problems encountered in the education system, but it can provide a necessary balance to some of the limitations experienced with the traditional approach.

Technology Enhanced Teaching and Learning

Technology enhanced learning refers to the enhancement of learning using information and communication technology (ICT), with the added benefit of helping students to develop new skills with digital tools along the way (Klemke & Specht, 2013). Technology, in this instance, plays a significant role in making learning more effective, efficient, or enjoyable (Goodyear & Retalis, 2010). In the literature it is often termed *e-learning* and it can support the educator and learner in a number of ways. For example, differing learning styles can be catered for so that educators can reach more students in a variety of ways, this subsequently enhances the number of students able to learn the course material (Sulcic & Lesjak, 2001). It is imperative when an organization or university decides to implement an e-learning initiative that they develop an effective solution that recognizes the need for good learning practices, by incorporating good design and development guidelines (Sulcic & Lesjak, 2001) – such as the learning dimensions advocated by Reeves and Reeves (1997) for interactive learning and collaboration. Active learning approaches, such as case-based learning and problem-solving, have long been advocated as ways of fostering deeper learning (Healy & Neville, 2009; Boyce

et al. 2001). Similarly, for many years, organisations have been using problem-solving scenarios such as business simulations to both test and train employees. Simulations enhance the learner's logical reasoning, numeric abilities, and spatial thinking through real problem-solving scenarios. Realising the potential of such methods, however, requires active engagement from educators and learners alike (Healy & McCutcheon, 2008). For many educators, the lack of appropriate materials, learning management, assessment techniques, and guidance are often perceived as barriers to student or employee engagement. With the 'right' underlying pedagogical approach social media technology provides educators with the technical platform to overcome these well-cited issues. Quellmalz et al. (2013, p. 1111) suggest that "engaging students in interactive assessments may provide a better estimate of their more complex inquiry practices than active or static formats" – providing third levels educators with an impetus to deliver a more complete learning experience.

Social Media Enabled Learning

Extant learning theories support the view that student learning is enhanced through opportunities to work collaboratively (Prince, 2004), and virtual learning environments actively support this form of learner collaboration (Peat, 2000). Web 2.0 has revolutionized traditional media content and the way people communicate and interact. Consequently, social media technologies have the potential to support and enhance teaching and learning in higher education (Cao & Hong, 2011; Hajli et al., 2013). In fact, social media gives learners a chance to manipulate their learning environment and to participate actively in the learning process (Hrastinski, 2009). It is through these collaborative technologies that students and knowledge workers will gain enhanced insight in the knowledge at their disposal. These tools will also enable information workers to locate and connect people with certain expertise across organizations, bringing people, systems, and data

into alignment faster to respond to challenges and take advantage of competitive opportunities. In an educational context, social media, when properly facilitated and framed can help expand the potential for learning over time by creating more student connections (Chen & Bryer, 2012). Valjtaga and Fielder (2009, p58) widely support the use of social media technology as a means of skilling students in preparation for the 'real world', "we need to construct opportunities for participants in higher educational settings to practice the advancement of self-directing intentional learning projects." However, research is needed in the area of education for the use of social media applications as online learning environments, and the learning affordances they may offer (Selwyn, 2007). Currently this area is lacking rigorous and carefully conducted research (Cao & Hong, 2011). One area of the IS undergraduate curriculum that attracts considerable attention is IS security education, as it has several practical 'real world' implications in the form of accessibility of sensitive data (Sauls & Gudigantala, 2013). Extant research highlights the importance of ISS education in terms of developing an entry level understanding of security from an IT infrastructure perspective as well as the complementary analytical skills and problem solving skills (Sauls & Gudigantala, 2013; White et al. 2013). The next section considers ISS as a topic which benefits from active learning approaches through the use of e-learning environments.

IS Security (ISS) in Undergraduate Education

In just a few decades, the use of IS has formalised information management and streamlined the administration of organisations (Galliers & Newell, 2001; Dhillon, 2006). However, in the realms of IS security, one of the fundamental problems for an organisation is to choose the right kind of environment to function in. While many organisations have engaged in identifying security issues and as a result developed appropriate ISS policies, there is a clear mismatch between policies and

what is done in practice. Therefore theories-in-use have degrees of effectiveness which are learned (Mattia & Dhillon, 2003). Espoused theory and theory-in-use are a part of the double-loop learning concept which creates a mindset that consciously seeks out security problems in order to resolve them. This results in changing the underlying governing variables, policies, and assumptions of either the individual practitioner, function or the organisation. To help address this, the literature has identified the importance of producing qualified and competent graduates to manage the security challenges facing organisations today (Sauls & Gudigantala, 2013). Hence, it is vital that IS/IT security and disaster recovery are treated as core components in business IS curriculms (Patten & Harris, 2013; White et al., 2013; Kim et.al, 2006). Considering the complexity of the subject area, it is evident that teaching the know-how and know-what of an ISS course to IS undergraduates requires a hands-on approach to adequately deal with some of the concepts and underlying principles (Ilvonen, 2013; Sauls & Gudigantala, 2013). Thus, it is necessary to go beyond the traditional learning environment where the 'knowing what' or declarative knowledge base is core to the learning experience (Quellmalz et al., 2013). This includes providing real-world experiences, using real industry situations, scenarios, and exercises, by training students from both a proactive and reactive viewpoint (Woodward et al., 2013). Exploring the 'know-how', ensures that students acquire a more complete and practical experience, better preparing them as prospective managers of tomorrow (Ilvonen, 2013).

IS Graduates of the Future

The workplace is continuously evolving and adapting to support organizational and technological change (Lee & Brand, 2005). Companies and universities will experience daunting challenges as they compete for and produce the next generation of IS talent. Generation 2020s or Generation Zs are characterized by the pervasiveness and avail-

ability of technology in every aspect of their lives (Geck, 2007). Instead of educating this generation about the Internet, the focus will be providing them with the skills to leverage the value of technology (Geck, 2007). The new workplace environment expects its knowledge workers to be proactive and show initiative, collaborate smoothly with others, take responsibility for their own professional development, and be committed to high quality performance standards. Graduates will have to use IT hardware and software intensively in a richly enabling business process environment to fulfil their roles. Indeed in 'Talent 2020' strategy, Deloitte (2012) refer to the development of their 'onboarding programs', which involve the provision of technology and resources to new employees as a means of enabling their efficient and effective engagement with the workplace. Notably, college students practice poor security behaviors and fail to properly use computer security tools (Mensch & Wilkie, 2011), thus emphasising the need for IS students to gain a critical understanding of the global issues of information security and assurance from their college education (White et al., 2013). With this in mind, 'knowing how' through active learning has become a formative part of the IS curriculum as the nature of technologies and work practices have changed. In addition, by utilizing technology in teaching and learning, it encourages student engagement and better prepares students for the work environments already embracing such technologies (Doyle et al., 2015). Subsequently, we endeavour to investigate how potential IS graduates leverage social media technology in a manner that enables them to appreciate, first-hand, the 'real world' implications of ISS. Essentially moving from the teacher-centred paradigm to the learner-centred paradigm where assessment is devised based on the student's learning needs and their motivations (Ilvonen, 2013).

Research Approach

This research study outlines the adoption of a blended approach that encompasses both the

traditional approach and the use of social media tools as part of an online game facilitated by IS Security teachers within a university setting. The case selected for this research study was an undergraduate ISS module within University College Cork (UCC), Ireland. This module was selected because of the progressive approach of the associated facilitators to develop, customize, and blend traditional learning approaches and e-learning technologies. The researchers examined the development and implementation of an online game designed to allow students to leverage their classroom acquired know-what in the area of IS security in a simulated 'real world' environment, namely, social media technology Facebook. Table 1 provides a description of the game.

The objective to the ISS Game was to establish a fictitious security company with a social media presence. Data was collected through the student groups, each group submitted a report as the continuous assessment deliverable as part of the overall module. These group reports detailed their communications (e.g. group meetings and communications), and provided an account of their

fictitious company assets, these comprised group generated content including photos, recipes and limericks, and the security attacks made by the company on other groups in the class. This data was analyzed based on the group's ability to leverage the material they learned in class and apply it in the online game scenario, the project outcomes are outlined in Table 2. The performance of each of the groups was reflected in their project grades. Significantly, the feedback on the ISS game from the class was hugely positive, particularly in terms of making, what can be perceived as a 'boring' or far removed topic to undergraduate students, more interesting, engaging, and important. The authors investigated the degree to which the benefits of the game met the learning needs of the students. The analysis also expanded the on-going design of the game to provide an innovative approach to learner support that is more akin to the true essence of social learning. The next sections provide a description of the ISS module and the social game. This is followed by the results of the student's participation and an analysis of the results.

Table 1. Social ISS game description

Student groups are required to form a fictitious security consultancy through, for example, Facebook. The consultancy will be composed of the 6 group members and provide a landing page only viewable to the public and used to document the everyday operations of the consultancy. A hidden section will be used to collect, store and share security resources, tools and information about 5 corporate security breaches. Access to this section of the 'company' will be submitted to the Coordinators.

Students will therefore complete the following, illustrate solutions when appropriate:

1. Create a Fictional consultancy and provide public information about the company and the different (max 6) employees.
2. Select 5 security breaches and critically evaluate / discuss the breaches in terms of:
 - a. ISS Controls before and after the breach
 - b. Business impacts
3. Use Facebook and or Twitter to document the work undertaken to investigate the breaches selected.
4. Compare and contrast the selected 5 or more breaches and the techniques used by the ISS groups to protect corporate assets.
5. Store an asset electronically in the secured section of the Facebook page, on 6 USB keys and printed copies of the page. Create and document the ISS strategy used to protect your corporate asset.
6. A copy of the assets as well as the company details, logins / access to the corporate network and twitter accounts are to be submitted ...
7. Select one or more other student groups to determine if their ISS controls can be bypassed. Document the strategy used by your group to test another groups controls. The strategy used should be documented using a company / group twitter account.

Note: Extra marks will be awarded to the group/company which can obtain a copy of an asset protected by another group. A maximum of 2 assets will be rewarded.

SOCIAL GAMING ENVIRONMENT

This game was created to facilitate and support understanding and learning of the links between ISS and its business applications, essentially to stimulate the 'right' type of learning (Carless, 2007). The creation of fictitious companies and corporate espionage components presented students with the opportunity to play the role of ISS practitioners protecting and simultaneously targeting corporate boundaries in an attempt to acquire and protect assets. At the start of the teaching period 2012/2013, the class (72 students) was divided into teams of 6 members. Each group was expected to meet at least once a week. Project Teams were asked to record minutes of all meetings and day-to-day operations through Twitter and these were kept for review by the coordinators throughout the year (protected Tweets with group and coordinator access). Facebook and Twitter were used to co-ordinate the work effort; thus planning of ISS tasks and workloads well in advance formed an essential element of the overall assignment.

Objective and Outcome of the ISS Game

The game was structured as part of the lecture series (24*2 hours) to gradually build knowledge of the ISS subject domains while simultaneously simulating 'real world' situations when the groups are asked to deliver a series of requirements to determine their level of understanding of the topics (Table 2) discussed in class. Thus following Wilson's (2012) recommendation that if instructors choose to use an active learning environment there needs to be controls in place to guarantee that students are somewhat knowledgeable and prepared before their arrival in the classroom or in this case social media gaming environment. At the end of each assessment submission the goals for the next submission and lecture were set, based on the level of knowledge and understanding demonstrated by students up to that point. Groups submitted and

presented their deliverables at agreed deadlines. This enabled a post-mortem evaluation approach (Kasi et al., 2008). Groups were also required to select 5 security breaches and critically evaluate/discuss the breaches in terms of the controls used before and after the breach. The business impacts were also investigated. These evaluations were then presented and discussed in class. In some instances 3 or 4 groups investigated the same case. However each had their own view regarding how the company reacted to and learned from a reported (published) incident. Case analysis and discussion was the traditional and preferred form of an in-class exercise. This component was enhanced through the use of social media to store documents (reports, articles, videos and slides) and Twitter searches using # to find the discussions which occurred in real-time. Table 2 outlines how these topics were taught, applied, and assessed through the ISS business game.

Student groups submitted their secret/assets at the beginning of the game (Term 1). However while the same technologies, discussed in class and reviewed by the groups, empower ISS practitioners they also empower hackers and hacking organisations to subjugate different types of information systems. This threat became part of the game as students adopted the role of hackers and targeted the secret/asset of another groups to gain extra marks. Thus supporting extant research which suggests that teaching students to defend against IS attacks also develops their attacking skills (Ilvonen, 2013).

Table 3 outlines the results of the hacking component of the game. The groups primarily tried to use man in the middle and password bypass attacks. The most successful attack was conducted by group 1. The group emailed the class using a fake email address from one of the coordinators: Coordinator1@gmail.ucc.ie requesting that students email their secrets before 5pm on the day of the agreed 'secret' submission. 26 students emailed their group's secrets to the fake account. This earned group 1 two extra bonus marks due to the limit of 2 and the fact that by the time the

Table 2. Topics taught and game outcomes

Generative Topics: Analysing ISS requirements; developing ISS strategies; creating ISS plans; ISS controls, computer forensics, compliance, SecSDLC (systems development life-cycle); secure Development, Designing Audit reporting & project management			
Through-lines	Performances	Assessment	Outcomes
How do you address ISS business problems?	Students will understand how to identify assets & allocate the right controls	Related exercises of case organizations	Analysis of Real cases & solutions used to protect assets
How do you identify the ISS requirements of a business?	Students will see the connection between ISS controls & business value	Provide the requirements of a business case	Business impacts of security breaches will be measured.
How should you address analysis requirements?	Students will understand how to analyse problems	Illustrate a logical view of a solution	ISS strategies are selected, discussed & evaluated
How can you use what you have learned to build a solution?	Students will understand the building blocks of an ISS solution	Provide step-by-step guide incorporating a Security life cycle	ISS plan developed
How do you use solve a security breach?	Students will address business & ISS threats, controls & disaster recovery	Provide a fake solution to a security breach based on a case description	Computer forensics of a breach presented
How do you use build an ISS report & apply lessons learned?	Students will understand the skills necessary to convey technical ISS issues to mixed audiences	Feedback on presenting lessons learned from case assessments	Audit review conducted of a selected case
Understanding			
Goals	How to:		
Students will understand the process of 'Forming fictitious ISS companies'.	Submit Company through Facebook , roles, Twitter diary & asset	Emailed to Lecturer early in Term 1	Details of Company, Structure & secret submitted
Students will understand how to apply a divide & conquer approach to an ISS problem	Develop a strategy to protect the group asset. Target other groups to obtain their asset	Presentation through preliminary report & reported through protected tweets.	Full ISS solution submitted for review & feedback
Students will understand how to develop ISS strategies to protect corporate boundaries	Present a walkthrough of the key elements of their strategies: failures & successes	Presentation of solution, with detailed feedback provided	ISS solution presented & reviewed
Students will learn to compete against their peers through ISS bypass attempts.	Students will understand the complete process of protecting & targeting an organization	Construction & presentation of the final versions of solutions.	ISS solution tested internally & externally (penetration tests).

project concluded (despite being consistently targeted by other 11 groups) did not have their corporate secret stolen. This attack was also used to illustrate well published cases of man-in-the-middle attacks. That is despite the fact that this topic was covered during class (2 weeks before

the attack occurred) to the amusement of the majority of the students, due to the simplicity of the attack, 26 students fell for it. This reinforced the importance of controls such as SETA (security, education, training and awareness). As outlined in Table 3 groups were successful in acquiring

Table 3. Attack attempts (Step 7 of Table 1)

Grade	Group	Attack Type	Attempt	Success	Failure	Hacked	Secret Lost
85%	1	Man in Middle	4	26/72	3	4 Org Secrets	None Stolen
72%	2	Password bypass	6	1/72	5	1 Org Secret	None
62%	3	Man in Middle	4	0/72	4	0 Org Secrets	None
68%	4	USB Acquired	2	1/72	1	1 Org Secret	Secret Stolen
75%	5	All Attempted	8	0/72	8	0 Org Secrets	None
60%	6	Password bypass	1	1/72	1	1 Org Secret	Secret Stolen
59%	7	Password Guess	1	1/72	0	1 Org Secret	None
64%	8	Login left open	1	1/72	0	1 Org Secret	Secret Stolen
70%	9	Man in Middle	12	0/72	12	0 Org Secret	Secret Stolen
63%	10	SETA Failure	4	1/72	3	1 Org Secret	Secret Stolen
54%	11	None attempted	0	0/72	0	0 Org Secret	Secret Stolen
50%	12	Password Bypass	27	0/72	27	0 Org Secret	Secret Stolen

another group's asset as well as unsuccessful in protecting their own corporate assets.

In a repeated exercise, conducted with 57 students in the 2015/16 term, Table 4 outlines the number of attack attempts made by individual students and the number of assets the students were able to steal overall, based on 44 survey responses. On average, the students attempted 6 attacks and were able to steal at least one asset, though this varied across the groups.

Figure 1 displays a breakdown of the companies that were attacked during the 2015/16 term exercise. It is evident that some companies were targeted more than others during the exercise.

Of the students surveyed, only some of them showed an awareness of the loss of their asset (see Figure 2). To illustrate this, taking the 44 student responses in Table 4, it was claimed that 61 assets were stolen overall, however, of the 43 student responses about the awareness of this loss, 72% of the students surveyed claimed no loss of an asset and 28% recognised that they had actually lost an asset.

Surveying the students enabled greater insight into the value of the exercise for learning. When asked whether their own attack attempts improved the protection of the group/company's asset (Figure 3), 88% of students (out of 43 responses) replied in the affirmative. Student's had learned through attack attempts how to better improve their own security. For example, one student commented

Table 4. Attack attempts and stolen assets (2015)

	Average No.	Total No.	Responses
Attack attempts	6	258	44
Stolen assets	1	61	44

"We made sure we logged off our computers in the labs at all times when left unattended after getting access to other rival company computer accounts that way". Likewise, different strategies used by students were noted "As we composed a phishing email, it made us more aware that other groups would be doing the same so we were extra careful when receiving emails, correctly identifying them as false".

Discussion

Problem-solving skills require the use of a number of different learning strategies and types of knowledge. In this case, the students leveraged the concepts and strategies considered in the lecture in order to create a fictional company, roles (hacker and security officer), and to apply technical (bypass Facebook controls) and no-technical hacking (social engineering). Students established their secure company and attacked another company/student group using social media. Students were exposed to a number of security threats, enabling them to protect their assets while assuming the

Figure 1. Survey of 44 students (2015): question “what companies did you attack?”

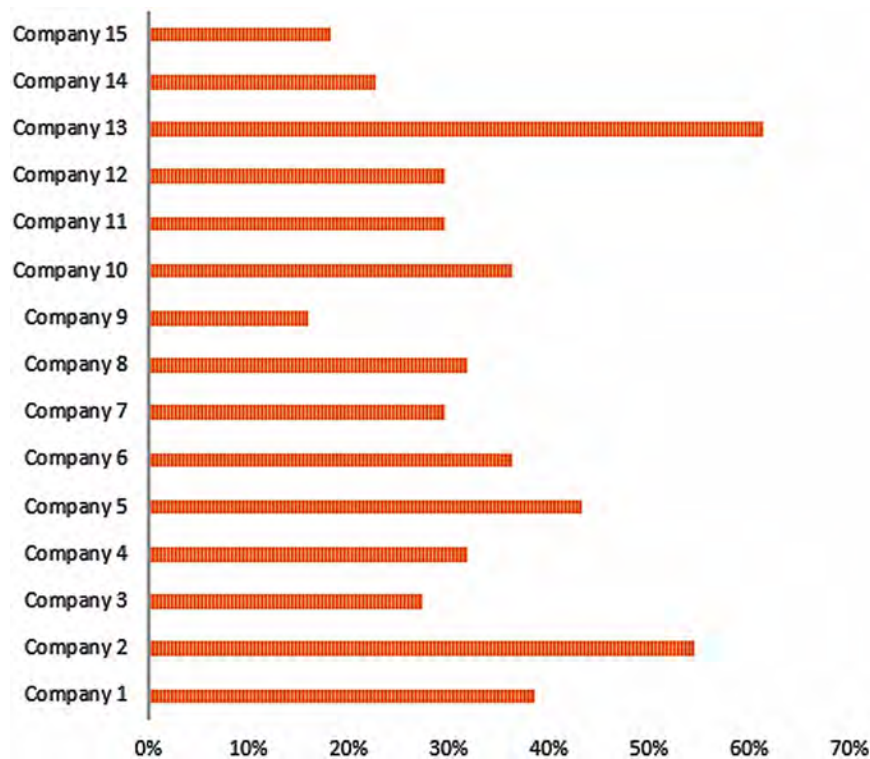
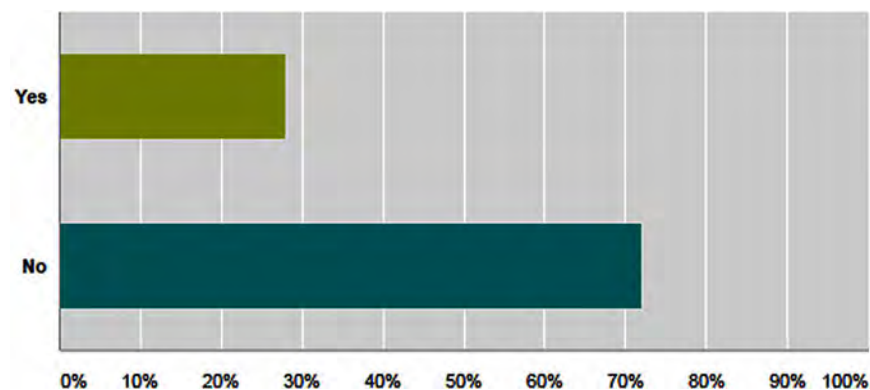


Figure 2. Survey of 43 students (2015): question “was your asset stolen?”

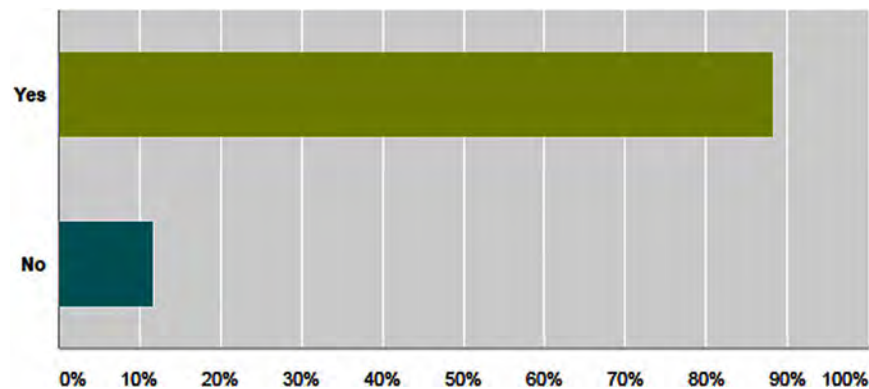


role of hacker. Students became very competitive as they endeavored to steal marks as well as company assets.

The learner’s own experience, internal mental models, and other ‘cognitive structures’ are necessary to ‘construct’ their own knowledge when faced with new information or different situations. This game reinforced the need to recognise that the technical side of ISS is a part of, but not the

answer to, the different ISS challenges. Knowledge and expertise of the technologies are deemed both necessary and valuable in terms of alleviating ISS risks. However in order to achieve this, ISS students must be familiar with critical business processes as well as ISS business impacts. Technological changes, in both secure hardware and software, are as constant as the increase in the number of threats to corporate ISS. Forgetting the most basic

Figure 3. Survey of 43 students (2015): question “did your attacking attempts improve the protection of your group/company asset?”



types of attacks and the potential for employee mistakes are common issues for organisations in general. These errors were experienced by the student groups (Table 3). The mistakes made, as well as adopting the role of hacker, reinforced the material taught in class.

FUTURE RESEARCH DIRECTIONS

In their study Tay and Allen (2011, p. 153) purport that

Staff (educators) saw both the necessity of including greater use of social media in teaching and, at the same time, believed that neither social media technologies themselves, nor the informal and personal cultures of use that students had developed, would necessarily mean that this innovation would – without close attention to pedagogic design – reliably improve students’ outcomes.

While the use of social media technology in this study proved to positively enhance the student learning experience, it is imperative that academia continues to engage with social media technologies in order to further understand and leverage their capabilities, as well as ensuring that this convergence of the traditional and new provides students with a fulfilling learning experience ultimately preparing them for the challenges of

the workplace. Notably this was the first time the game was utilised as part of continuous assessment for an ISS module. The social gaming approach will be used in the future however the lecturers acknowledge the importance of leveraging student grades as part of the student group evaluation process in order move beyond the reported ‘novel’ and positive perceptions associated with the use of social media as part of this class.

Cao et al. (2013) purport that social media is being widely used as a mechanism for learning and assessment in third level institutions. In this study, an assumption was made that students were familiar with social media technology from a social perspective, subsequently they were required to leverage this platform as a means of meaningfully engaging with ISS issues. White et al. (2012) persuade that ISS awareness and education is core to IS education and they contend that this should be supported by education in preventative education and incidence response. The social game presented in this study could be further developed to provide students with the ‘know-what’ and ‘know-how’ to devise preventative security and response strategies.

Like Cao and Hong’s (2011) study of social media utilization in teaching, this research focused on the delivery and assessment of one ISS undergraduate module in one university. In order to further understand the impact of utilizing social media technology as a means of an active

learning approach, a longitudinal study over a number of semesters would be necessary to fully investigate the strengths and weaknesses of this learning strategy (Mikropoulos and Natsis, 2011). Thus, this exercise was repeated with students during the 2015/16 term and preliminary findings were presented along with the findings from 2013/14. This time, students were surveyed after the exercise to better understand the impact of the game on their learning. Overall, students were positive about the learning impacts of the game and were able to identify multiple strategies for attacking and protecting company assets. It would also be beneficial to apply a similar social media gamed-based assessment in another undergraduate programme as a means of exploring the strength of this approach in another setting. It is only then that we can really begin to understand the true impact of the use of social media as part of the active learning approach.

CONCLUSION

It is important that educators begin to place social media technology at the core of IS third level learning and assessment approaches going forward. It is through leveraging these technologies as part of the active learning approach that will better prepare our graduates for the workplace. Undoubtedly, universities and companies face daunting challenges as they compete for the best talent in 2020. They will need to attract new talent, train, retain, and create an engaged workforce. Workers will mainly be knowledge workers, utilizing the Internet, computer technology, communications technology, and knowledge processing platforms as an integral part of their work. Working in physical office spaces with colleagues will diminish, giving way to virtualized teams in distant geographical locations. Third level educators and employers will need to work together to ensure that graduate employees have access to leading edge connectivity tools and the appropriate skills to replace the lack of face-to-face communication.

Organizational leaders are keenly aware that the workplace is changing and are already recruiting a new breed of employee. They are adapting their workplace policies and strategies to appeal to all generations. Therefore IS educators must adapt to the changing needs of industry and students in developing competitive skill-sets through the design of a curriculum that excites students, trains them [IS students] with hands-on exposure, and provides them with the necessary skills to achieve success in the IT industry (Sauls & Gudigantala, 2013). As “current university graduates will become tomorrow’s protectors of data and systems” (White et al., 2013, p. 14).

REFERENCES

- Boyce, G., Williams, S., Kelly, A., & Yee, H. (2001). Fostering Deep and Elaborative Learning and Generic (Soft) Skill Development: The Strategic Use of Case Studies in Accounting Education. *Accounting Education*, 1(10), 37–60. doi:10.1080/09639280121889
- Cao, Y., Ajjan, H., & Hong, P. (2013). Using social media applications for educational outcomes in college teaching: A structural equation analysis. *British Journal of Educational Technology*, 44(4), 581–593. doi:10.1111/bjet.12066
- Cao, Y., & Hong, P. (2011). Antecedents and consequences of social media utilization in college teaching: A proposed model with mixed-methods investigation. *Horizon*, 19(4), 297–306. doi:10.1108/10748121111179420
- Carless, D. (2007). Learning-oriented assessment: Conceptual bases and practical implications. *Innovations in Education and Teaching International*, 44(1), 57–66. doi:10.1080/14703290601081332
- Deloitte. (2012). *Talent 2020: Surveying the Talent Paradox from the Employer Perspective*. Retrieved on January 31, 2014, from http://www.deloitte.com/assets/Dcom-Azerbaijan/Site%20SMF/EN/Events/US_Talent2020_September2012_09142012.pdf

- Doyle, C., Sammon, D., & Neville, K. (2015). Building an evaluation framework for social media-enabled collaborative learning environments (SMECLEs). *Journal of Decision Systems*, 24(3), 298–317. doi:10.1080/12460125.2015.1054694
- Geck, C. (2007). The generation Z connection: Teaching information literacy to the newest net generation. *Toward a 21st-Century School Library Media Program*, 235.
- Goodlad, J. (1984). *A Place called school*. New York: McGraw-Hill Book Co.
- Goodyear, P., & Retalis, S. (2010). *Technology-enhanced learning*. Sense Publishers.
- Hajli, M., Bugshan, H., Lin, X., & Featherman, M. (2013). From e-learning to social learning—a health care study. *European Journal of Training and Development*, 37(9), 851–863. doi:10.1108/EJTD-10-2012-0062
- Hannum, W., & Briggs, L. (1982). How does instructional system design differ from traditional instruction? *Educational Technology*, 22(1), 9–14.
- Healy, M., & McCutcheon, M. (2008). Engagement with Active Learning: Reflections on the Experiences of Irish Accounting Students. *Irish Accounting Review*, 1(15), 1–49.
- Healy, M., & Neville, K. (2009). A Teaching Case: Towards Bridging Disciplinary Divides in IT Education. *Proceedings of the 17th European Conference on Information Systems (ECIS 2009)*.
- Iivonen, I. (2013). Information security assessment of SMEs as coursework – learning information security management by doing. *Journal of Information Systems Education*, 24(1).
- Kim, Y., Hsu, J., & Stern, M. (2006). An update on the IS/IT skills gap. *Journal of Information Systems Education*, 17(4), 395.
- Klemke, R., & Specht, M. (2013). Technology Enhanced Learning. *Proceedings of the Fourth International Conference on e-Learning (eLearning-2013)*.
- Lee, S. Y., & Brand, J. L. (2005). Effects of control over office workspace on perceptions of the work environment and work outcomes. *Journal of Environmental Psychology*, 25(3), 323–333. doi:10.1016/j.jenvp.2005.08.001
- Mattia, A., & Dhillon, G. (2003). *Applying Double Learning to Interpret Implications for Information Systems Security Design*. IEEE Systems, Man and Cybernetics Conference, Washington, DC. doi:10.1109/ICSMC.2003.1244262
- Mikropoulos, T. A., & Natsis, A. (2011). Educational virtual environments: A ten-year review of empirical research (1999–2009). *Computers & Education*, 56(3), 769–780. doi:10.1016/j.compedu.2010.10.020
- Patten, K. P., & Harris, M. A. (2013). The Need to Address Mobile Device Security in the Higher Education IT Curriculum. *Journal of Information Systems Education*, 24(1), 41.
- Peat, M. (2000). Towards first year biology online: A virtual learning environment. *Journal of Educational Technology & Society*, 3(3), 203–207.
- Prince, M. (2004). Does active learning work? A review of the research. *The Journal of Engineering Education*, 93(3), 223–231. doi:10.1002/j.2168-9830.2004.tb00809.x
- Quellmalz, E. S., Davenport, J. L., Timms, M. J., DeBoer, G. E., Jordan, K. A., Huang, C. W., & Buckley, B. C. (2013). *Next-generation environments for assessing and promoting complex science learning*. Academic Press.
- Quellmalz, E. S., & Pellegrino, J. W. (2009). Technology and testing. *Science*, 323(5910), 75–79. PubMed doi:10.1126/science.1168046
- Reeves, T. C., & Reeves, P. M. (1997). Effective Dimensions of Interactive Learning on the World Wide Web. In B. H. Khan (Ed.), *Web-Based Instruction* (pp. 59–66). Englewood Cliffs, NJ: Educational Technology Publications, Inc.

Sauls, J., & Gudigantala, N. (2013). Preparing Information Systems (IS) Graduates to Meet the Challenges of Global IT Security: Some Suggestions. *Journal of Information Systems Education*, 24(1).

Schneckenberg, D. (2009). Web 2.0 and the empowerment of the knowledge worker. *Journal of Knowledge Management*, 6(13), 509–520. doi:10.1108/13673270910997150

Selwyn, N. (2007). *Web 2.0 applications as alternative environments for informal learning - a critical review*. Retrieved on May 16, 2012 from, <http://www.oecd.org/dataoecd/31/37/39459090.pdf>

Sulcic, V., & Lesjak, D. (2001). DE in Slovenia: Where are we? *Proceedings of the 9th European Conference on Information Systems (ECIS 2001): "Global Co-operation in the New Millennium"*, 1087-1097.

Tay, E., & Allen, M. (2011). Designing social media into university learning: Technology of collaboration or collaboration for technology? *Educational Media International*, 48(3), 151–163. doi:10.1080/09523987.2011.607319

Valjataga, T., & Fielder, S. (2009). Supporting students to self-direct intentional learning projects with social media. *Education et Sociétés*, 12(3), 58–69.

White, G. L., Hewitt, B., & Kruck, S. E. (2013). Incorporating Global Information Security and Assurance in IS Education. *Journal of Information Systems Education*, 24(1).

Wilson, L. V. (2012). An Assessment Of Learning Outcomes Based On A Comparison Of Active Learning And Traditional Lecture Pedagogical Styles In A Legal Environment Classroom. *Southern Journal Of Business And Ethics*, 4(101), 110.

Woodward, B., Imboden, T., & Martin, N. L. (2013). An Undergraduate Information Security Program: More than a Curriculum. *Journal of Information Systems Education*, 24(1), 63.

ADDITIONAL READING

Moran, M., Seaman, J., & Tinti-Kane, H. (2011). *Teaching, Learning, and Sharing: How Today's Higher Education Faculty Use Social Media*. Babson Survey Research Group.

KEY TERMS AND DEFINITIONS

Information System Security (ISS): The study of defending information, hardware, software resources against unwarranted attack.

Knowledge Workers: A person with a specialized skillset e.g. software engineer, architect, financial analyst.

Online Learning Environment: Online (web-based) channels utilised to support student learning and assessment.

Social Business Gaming: The use of online social gaming as a means of measuring student performance as part of their on-going learning process.

Social Media Technology: Technologies such as Facebook, LinkedIn and Twitter which facilitate user engagement and user-generated content.

Student Assessment and Learning: Knowing what content and skills students have mastered.

Workspace Technology: The platforms and supporting technologies utilised by Knowledge Workers in their day-to-day work routine.

Liberating Educational Technology Through the Socratic Method

E**Frank G. Giuseffi***Lindenwood University, USA*

INTRODUCTION

Digital literacy and technology are instruments for human communication and behavior (Lemke, 2010). The skills and attributes the human person needs for responsible citizenship, and work-performance, is being re-defined by, what Dede (2010) called “information and communication technologies.” In the educational world, November (2012) claimed that a revolution is happening where teachers are harnessing the uses of technology in their courses. Moreover, the influence technology is having on society is fundamentally changing the nature and functions of schools (Lever-Duffy, McDonald, & Mizell, 2010). While one should celebrate the positive results of the role technology has given students in their educational endeavors, our celebration must be tempered with caution. Recent research suggest that suggests that technology, in the form of laptops, has not raised student achievement in any significant way (Goodwin 2011; Hu 2007). If this is the reality that confronts us, we are then pressed to respond. The question for educators is what kind of response? What are the answers to this problem that seems to be growing? This chapter, per the author, suggests that a return to something used in antiquity may be the answer. This chapter explores the use of the Socratic method as a teaching technique that can give direction to the lack of pedagogical vision in the great One to One debate currently confronting schools.

BACKGROUND

In 2009, Arne Duncan, U.S. Secretary of Education exhorted public schools nationwide to implement technology in public schools (Lemke, 2010). He indicated at a national consortium that “good teachers can utilize new technology to accelerate learning and provide extended learning opportunities for students” (Lemke, 2010, p. 245). As one example of this desire to increase educational technology, schools began to invest millions of dollars in One to One laptop programs (Goodwin, 2011).

However, even before this speech by Duncan, issues were raised concerning One to One laptop programs. Hu (2007) indicated that school districts in New York and elsewhere were seeing One to One laptop programs as major obstacles to student learning. As early as 2007, the United States Department of Education found that there was “no difference in academic achievement between students who used educational software programs for math and reading and those who did not” (Hu, 2007). Studies in Texas and Michigan showed mixed results in student achievement when it came to the effectiveness of laptop programs (Goodwin, 2011).

One cannot also discount the influence teaching has on successful laptop programs (Stansbury, 2010). Studies published in the *Journal of Technology, Learning and Assessment* at Boston College’s Lynch School of Education indicated that “the most important factor of all is the teaching practices of instructors – suggesting school

laptop programs are only as effective as the teachers who apply them” (Stansbury, 2010). This is further confirmed from results of a recent study of 997 schools in the United States indicating that one of the factors that added to successful laptop programs was teacher training (Goodwin, 2011).

Hence, educators must seize upon the notion that before laptops are given to students, a commitment to teacher training is needed (November, 2010). Norris and Soloway (2010) echoed this sentiment; they wrote: “To make the computer an essential tool in the classroom, and thus to realize the potential value added from technology, we need to redefine the curriculum in terms of what gets taught, and we need to redefine how it gets taught” (p. 1). Indeed, Pearlman (2010) made the bold claim that simply putting computers in the hands of students is not a solution, but actually “reinforces the old teacher-directed whole group instruction” (p. 127). The common experiences of schools that have embraced laptop programs has been to add on the technology to the same lesson assignments, instead of changing the nature of the lesson assignments. Students have been given the technology, but the lessons have not changed, resulting in the laptop becoming high-priced notebooks (November, 2010).

THE SOCRATIC METHOD

Socrates continually asked insightful questions that reflected the reality that learning came from within (Cookson, 2009). This can be an important teaching technique. As Paul and Elder (2007) noted,

Teachers, students, or indeed anyone interested in probing thinking at a deep level can and should construct Socratic questions and engage in Socratic dialogue. The purpose of using Socratic questioning in teaching may be to probe student thinking; to determine the extent of their knowledge on a given topic, issue, or subject; to model Socratic questioning for them; or to help them analyze a concept or line of reasoning (Pg. 36).

To fully articulate this argument, we first turn to a working definition of the Socratic method. While various definitions exist, the chapter, per the author, puts forth the following,

In the Socratic method, the classroom experience is a shared dialogue between the teacher and students in which both are responsible for pushing the dialogue forward through questioning. The “teacher” or leader of the dialogue asks probing questions in an effort to expose the values and beliefs which frame and support the thoughts and statements of the participants in the inquiry. The students ask questions as well, both of the teacher and each other (Reich, 2003) (P. 1).

The intention is to advance the discussion through dialogue and to uncover, dissect and critically examine accepted positions. This is fundamentally done through thought-provoking and specific questioning. This is clearly the way of Socrates (Cookson, 2009). He [Socrates] would engage his students through questioning and examination of beliefs (Gose 2009; Morrell 2004). This kind of dialectical practice that Socrates espoused leads one to the “good life” (Yengin and Karahoca, 2012). According to Cookson (2009) ultimately, Socrates

Believed learning came from within and that the best and most lasting way to bring latent knowledge to awareness was through the process of continual questioning and unconventional inquiry. For Socrates, answers were always steps on the way to deeper questions (P. 1).

Socratic questions are broken down into basically three kinds: Spontaneous, Exploratory and Focused (Paul & Elder, 2007). The spontaneous question is motivated from a genuine curiosity on the teacher’s part. They are unplanned and arise from conversations that take various paths. Examples could be asking for evidence, asking for others opinions on a given topic, or examples based on a point made during the discussion.

According to Paul and Elder (2007), this kind of question fosters a “Socratic Spirit.” One asks exploratory questions to discover what students are thinking on a particular subject or question. These questions help to broach topics or to get a sense of students’ attitudes or dispositions on a topic. The best and most effective exploratory questions, however, are carefully constructed before the dialogue begins. As Paul and Elder (2007) wrote,

However, for the greatest success some preplanning or pre-thinking is helpful. For example, one could construct a list of possible questions to ask at some point in the discussion. Another preparation technique is to predict students’ likeliest responses and frame some follow-up questions.

Lastly, focused questioning is often done within the context of a larger curriculum, needing to clarify or examine specific issues. The facilitator and students look at a specific topic, subject or idea. The dialogue is more structured and directed (Paul & Elder, 2007).

Let us now proceed to look at the aforementioned types of questions asked based on a Socratic discussion of Aristotle’s *Nicomachean Ethics* in my high school History of Ideas class in the Spring of 2016. We first see spontaneous questioning occur in the following. In Book II, Chapter six, Aristotle wrote,

But not every action nor every passion admits of a mean; for some have names that already imply badness, e.g. spite, shamelessness, envy and in the case of actions adultery, theft, murder; for all of these and suchlike things imply by their names that they are themselves bad, and not the excesses or deficiencies of them (Aristotle, trans. 1941, p. 959).

Teacher: What do we make of this word “shamelessness?” Is shame still a common feeling in today’s society?

Student 1: *The feeling of shame comes from societal pressure. This society prizes individualism. Individuals should not necessarily feel shame if it comes from society’s standards.*

Student 2: *But society’s standards are important.*

Student 1: *Not, however, if it takes away from you as a person.*

Student 2: *You can become a better person from society and having feelings of shame.*

I was initially looking at Aristotle’s concept of the “mean,” however, my interest in “shamelessness” led him or her to ask a spontaneous question on that subject. This then led to further comments on not just shame but its relationship to society and also the importance of the individual.

Staying on the philosophy of Aristotle, the following was an example of exploratory questioning. From Aristotle’s Book II, Chapter 6 of the *Nicomachean Ethics*,

But the intermediate relatively to us is not to be taken so; if ten pounds are too much for the person who is to take it, or too little – too little for Milo, too much for the beginner in athletic exercises. The same is true of running and wrestling. Thus a master of any art avoids excess and defect, but seeks the intermediate and chooses this – the intermediate not in the object but relatively to us (Aristotle, trans. 1941, p. 958).

Teacher: *How does one live morally by finding the intermediate state?*

Student 1: *They don’t go overboard or lend themselves to extremism on a moral issue.*

Student 2: *It depends on the situation you are confronted with. Maybe some things call for extreme measures.*

Teacher: *In what way?*

Student 2: *Well, sometimes governments becomes so tyrannical, revolution is called for. I would say that’s a kind of extremism.*

Teacher: *In a situation like that what would be the intermediate state? In other words, what would citizens who thought the government*

was tyrannical first seek to achieve if they were looking for Aristotle's conception of the intermediate?

Student 1: *Possibly trying to find a compromise with the tyrannical government? I know that sounds a little strange, but maybe those who were feeling abused could work with the government to resolve some of the injustices. Wouldn't that be looking for an intermediate state.*

I explored the issue of "the intermediate" by using moral behavior as an example. Then the discussion evolved into the possible justification, at times, of choosing the extreme.

Lastly, we looked at focused questioning. Here was a discussion based on Aristotle's *Nicomachean Ethics*, Book II, Chapter 7. He wrote,

With regard to feelings of fear and confidence courage is the mean; of the people who exceed, he who exceeds in fearlessness has no name (many of the states have no name), while the man who exceeds in confidence is rash, and he who exceeds in fear and falls short in confidence is a coward (Aristotle, trans. 1941, p. 960).

Teacher: *I am intrigued by the distinctions between courage and rashness. Would coming to the aid of a friend who is being bullied by someone much stronger and larger than you be a courageous or rash act?*

Student 1: *It depends on what you did. I would not use force on someone who is stronger and bigger than me.*

Student 2: *But in those situations one does not worry about oneself.*

Student 3: *Yeah but I am not going to risk my physical well-being over an issue that might be minor. I mean what really constitutes bullying anyway?*

Teacher: *So would it be a rash act to interject in a situation that might cause you physical harm and not be very helpful to the situation?*

Student 2: *I agree that rash acts often counter logic or sense. I would not excuse the scenario you gave. I think coming to the aid of someone even though you may be lacking in some quality is a courageous act.*

I directed the discussion on the differences between courage and rashness, citing an example in order to further illuminate in the students the nature of courage.

ONLINE SOCRATIC DIALOGUE

Studies suggest increased learning in online educational environments that use the Socratic method of teaching (Tucker and Neely, 2010). Moreover, the use of the Socratic method in these same environments bolsters the facilitation of discussion and increased critical thinking skills (Jumper, 2016).

Let us proceed, then, to describe how this can be done in a High School American History class that incorporates an asynchronous, One to One lap top program. In an online learning scenario with each student assigned a lap top, the instructor assigns *The Gettysburg Address*, one of the most well-known presidential speeches in American History; it was delivered by President Abraham Lincoln on November 19, 1863, several months after the three-day American Civil War battle in Gettysburg, Pennsylvania. The speech would be read by attaching a credible, respected History link to the school's Learning Management System (LMS). Prior to the assignment the teacher would have:

- Pretested the students to ensure that previous material had been mastered
- Articulated the learning outcome in specific detail
- Ensured that technologies were available to all students

- Set up content modules for students to use as resources of information (Duffy, McDonald & Mizell, 2010).

Hrastinski (2008) argued that there are three necessary forms of “communications” in asynchronous education the instructor should be aware of: 1) content-related; 2) planning of tasks, and 3) social support. In content-related communication, students (e-learners) must be able to read the assigned content, ask questions, and share their thoughts. In planning of tasks, there are assigned responsibilities, collaboration and conflict resolution. Lastly, in social support, there must be teamwork, positive feedback, and technical support.

In the following examples we will clearly see content-related communication. That is not to say, however, that the Socratic practitioner should not pay close attention to planning of tasks and social support (Gose, 2008). Also, whether in the traditional style of Socratic dialogue or its application in an online program, support for the students is essential. Both instructor and student must be conscious of group cohesion, respect for each other, and emotional support. This will be discussed further in the chapter.

The description of the speech is then followed by three sets of Socratic questions mentioned earlier - spontaneous, exploratory and focused. The exploratory and focused questions will first be presented. These questions can be asked by the instructor for students to ponder and reflect on prior to the dialogue. In fact, having some answers to these questions will help the students develop better spontaneous questions during the dialogue. The following elucidates, through examples, the three types of Socratic questions:

Exploratory

1. What is equality?
2. What does it mean to be “conceived in liberty?”

3. What is the relationship between liberty and equality?

E

Focused

1. What is the “great task” Lincoln is asking the American people to continue?
2. How would you describe Lincoln’s “new birth of freedom?”
3. Do you think this speech does the men who died on that battlefield honor? Why or why not?

Once the students have reflected on these questions, the teacher assigns a time and day for discussion; this is where the spontaneous questioning can take place. The teacher can first ask students about their own definitions of equality and liberty, inviting them to offer specific examples. The teacher can also investigate students’ answers on the connection between liberty and equality – challenging the notion that they are harmonious. This is where the “Socratic Spirit” is developed and a meaningful dialogue occurs. It should be noted that this educational experience would be asynchronous, where teacher and student are not communicating at the same time (Duffy, McDonald & Mizell, 2010). While one could argue that this presents a problem to Socratic dialogue, there are actually positive aspects to this as will be discussed later.

Let us see what that spontaneous Socratic conversation might look like while using laptops via an online discussion board; this platform is where teachers can post information and communicate with students (Ward, 2012). In an asynchronous Socratic dialogue, it is imperative that the teacher set clear procedures and rules. In developing these rules, there will be questions that would need to be answered. For instance, what is the time frame for responding to someone’s question? Also, it will be important to know how frequently each student is expected to post on the online discussion board (Ward, 2012).

The discussion begins with the teacher (facilitator) identified as TCH. Although not indicated on this page, the comments from each individual would include date and time of question or comment.

TCH: *Can somebody offer a definition of “equality?”*

ST#1: *In my mind, equality is the idea that all people are respected before the law.*

ST#2: *I agree, but also that all citizens of a country are protected before the law.*

TCH: *So are both of you arguing that governments or lawmakers grant equality to people?*

ST#1: *Yes, I would. At least a certain kind of equality is granted by governments.*

TCH: *Let us look at the Gettysburg Address again. Lincoln says “dedicated to the proposition that all men are created equal.” Is this implying that the equality that the founders conceived was something they granted or was found in nature?*

ST#2: *I think one can interpret from the document that they are found in nature.*

TCH: *Whether found in nature or granted by governments, the Founders certainly believed these concepts were important to the American Founding. Is there a relationship between equality and liberty?*

ST#3: *Yes. I was trying to come up with a definition of equality that includes liberty. For instance, a regime can promote the idea that all people are free to earn a living, worship in their own way and reside where they want to; this is a kind of equality, isn't it?*

Hence, we see the progression of the lesson take this form:

Figure 1.



With this example, we see a Socratic dialogue about equality via the use of a laptop program. The nature of the Socratic dialogue stays the same as the classical one described earlier. However, unlike the classical conception of the Socratic method, the Socratic method discussion via the Forum is, as we have described earlier, asynchronous - students and instructor can take some time reflecting on a question or comment before they participate. During this “reflective time” students could have re-read the Gettysburg Address, done further research on Abraham Lincoln, or developed better questions for the dialogue. While students are doing this, instructors must be aware of the second form of communication described by Hrastinski (2008), social support. For example, during the reflective time, the teacher could offer positive feedback to the questions and comments that were already asked, present further scholarly analyses by Lincoln scholars, or describe further information on asking insightful questions.

However, it must be noted, that the teacher must be careful not to offer too much information, since he or she runs the risk of disrupting the original topic in the dialogue. Second, students who are not as confident in their speaking ability or who are shy can benefit from using the forum, via their laptops, to voice their opinions. Lastly, students who are apprehensive with their answers can take the necessary time to review and edit what they wish to articulate (Duffy, McDonald, & Mizell, 2010).

In essence, we find four stages that must occur in a One to One Laptop program that uses the Socratic method of teaching and learning. First, there must be a preliminary stage where the lesson is assigned, pre-assessing and frontloading information has occurred, learning objective has been presented and laptops are available to all students. Second, there must be a clear indication of all three modes of communication – content-related, planning of tasks and social support. Third, Socratic questions should be delineated into three question subgroups argued by Paul & Elder (2007) – spontaneous, exploratory and fo-

Figure 2.



cused. Lastly, during the reflective time, teachers must be attentive to giving social support to their students; this will help with content-mastery and achievement.

FUTURE RESEARCH DIRECTIONS

The global possibilities to this educational experience are innumerable. Improvements in communication technologies caused a “death of distance” where the actual physical space between peoples from other countries is made obsolete (Zhao, 2009). Students from all over the United States and the world can participate in a Socratic discussion on the Gettysburg Address, and any other topic in American History. Therefore, it is necessary to continue research on the impact the Socratic method of teaching and learning has on One to One programs. This can be a school-wide initiative or a small sample using action research methods. Questions to consider are:

- Does the use of the Socratic method in One to One laptop programs increase student test scores?

- Does the use of the Socratic method in One to One laptop programs improve teacher pedagogy?
- Does the use of the Socratic Method in One to One laptop programs enhance Self-Directed Learning?

Asking these and other questions will continue the important conversation about teaching, technology and learning in the 21st century.

CONCLUSION

As modern educators, we tend to quickly embrace those things we think will improve student learning. Our faith and confidence in advances in educational technologies is no different. Educators and other professionals quickly invest in the latest technologies to hopefully improve learning outcomes and teacher performance. However, the pre-thinking, and pre-planning that comes with the purchase of these technologies is often set aside or not fully addressed. We should accept that technology like medicine is constantly changing and improving. However, that acceptance is guided by a respect and recognition of those ideas

and educational practices that have informed humankind for centuries. One of those practices has been the Socratic method. We are correct to look forward to the future of technology and the many great things it can offer the world, but we would also do well to capture and hold on to those ideas and practices that have proven beneficial to the intellectual development of humankind.

REFERENCES

- Aristotle,. (1941). *Basic works of Aristotle* (W. D. Ross, Trans.). New York, NY: Random House.
- Cookson, P. W. (2009). *What would Socrates Say?* Retrieved from <http://www.ascd.org/publications/educational-leadership/sept09/vol67/num01/What-Would-Socrates-Say%C2%A2.aspx>
- Dede, C. (2010). Comparing frameworks for 21st century skills. In J. Bellanca & R. Brandt (Eds.), *21st century skills: Rethinking how students learn* (pp. 243-272). Bloomington, IN: Solution Tree Press.
- Goodwin, B. (2011, February). Research says one to one laptop programs are no silver bullet. *Educational Leadership*, 68(5). Retrieved from http://www.ascd.org/publications/educational_leadership/feb11/vol68/num05/One-to-One_Lap-top_Programs_Are_No_Silver_Bullet.aspx
- Gose, M. (2009). When Socratic dialogue is flagging: Questions and strategies for engaging students. *College Teaching*, 57(1), 45–49. doi:10.3200/CTCH.57.1.45-50
- Hrastinski, S. (2008, November). *Asynchronous and synchronous e-learning*. Retrieved from <http://er.educause.edu/articles/2008/11/asynchronous-and-synchronous-elearning>
- Hu, W. (2007, May 4). Seeing no progress, some schools drop laptops. *The New York Times*. Retrieved from http://www.nytimes.com/2007/05/04/education/04laptop.html?_r=0
- Ilker, Y., & Karahoca, A. (2012). “What is Socratic method?” *The analysis of Socratic method through “self determination theory” and “unified learning model”*. Paper presented at the 2nd World Conference on Innovation and Computer Science, Singapore.
- Jumper, S. (2016). Retrieved from http://www.academia.edu/4068653/USING_SOCRATIC_TECHNIQUES_TO_INCREASE_LEARNING_OUTCOMES_IN_HIGHER_EDUCATION
- Lemke, C. (2010). Innovation through technology. In J. Bellanca & R. Brandt (Eds.), *21st century skills: Rethinking how students learn* (pp. 243-272). Bloomington, IN: Solution Tree Press.
- Lever-Duffy, J., McDonald, J. B., & Mizell, A. P. (2005). *Teaching and learning with technology*. Boston, MA: Pearson Education, Inc.
- Morris, C., & Soloway, E. (2010, May). One to one computing has failed our expectations. *District Administration*. Retrieved from <http://www.districtadministration.com/article/one-one-computing-has-failed-our-expectations>
- November, A. (2010). Technology rich, information poor. In J. Bellanca & R. Brandt (Eds.), *21st century skills: Rethinking how students learn* (pp. 275-283). Bloomington, IN: Solution Tree Press.
- November, A. (2012). *Who owns the learning*. Bloomington, IN: Solution Tree Press.
- Paul, R., & Elder, L. (2007). Critical thinking: The art of Socratic questioning. *Journal of Developmental Education*, 31(1), 32–37.
- Pearlman, B. (2010). Designing new learning environments. In J. Bellanca & R. Brandt (Eds.), *21st century skills: rethinking how students learn* (pp. 117-147). Bloomington, IN: Solution Tree Press.
- Speaking of Teaching. (2003, Fall). *Stanford University Newsletter on Teaching*. Retrieved from http://web.stanford.edu/dept/CTL/Newsletter/socratic_method.pdf

Stansbury, M. (2010). One to one computing programs only as effective as their teachers. *eSchool News*. Retrieved from <http://www.eschoolnews.com/2010/02/16/11-programs-only-as-good-as-their-teachers/>

Tucker, J. P., & Neely, P. W. (2010, June). Using Web Conferencing and the Socratic Method to Facilitate Distance Learning. *International Journal of Instructional Technology and Distance Learning*, 7(6), 15–22. Retrieved from http://www.itdl.org/Journal/Jun_10/article02.htm

Ward, D. (2012, December). *Socratic method and online teaching*. Retrieved from <https://www.google.com/webhp?sourceid=chrome-instant&ion=1&espv=2&ie=UTF-8#q=socratic+method+and+online+teaching>

Zhao, Y. (2009). *Catching up or leading the way: American education in the age of globalization*. Alexandria, VA: ASCD.

ADDITIONAL READING

Bassett, P. F., & Thorn, C. (Eds.). (2004). *Looking ahead: independent school issues & answers*. Gilsum, NH: Avocus Publishing.

Chafee, J. (2012). *Thinking critically*. Boston, MA: Wadsworth Cengage Learning.

Copeland, M. (2005). *Socratic circles: Fostering critical and creative thinking in middle and high school*. Portland, ME: Stenhouse Publishers.

Isenberg, S. (2007). *Applying andragogical principles to internet learning*. Youngstown, NY: Cambria Press.

KEY TERMS AND DEFINITIONS

Alan November: Current writer and theorist on the role technology has on teaching and learning.

Asynchronous: Online education where students and teacher work toward a shared learning objective by communicating with each other at different times.

Critical Thinking Skills: The capacity to understand and analyze ideas through a systematic form of thinking that applies logic, dialectic, self-reflection, and disciplined questioning.

Digital Literacy: The interpretive understanding of technological social networks that create communication and the uses and applications of social media.

Discussion Board: A social network forum that allows for communication and shared inquiry between individuals.

Educational Technology: A description of any learning program and lesson assignment that incorporates technology in its learning outcome and teaching methodology.

Learning Management System: Software programs that capture and documents information for educational purposes.

Socratic Method: A method of teaching that applies a dialectical exchange between persons who wish to arrive at a shared understanding concerning a particular question, thought or idea.

Online Academia

Magdalena Bielenia-Grajewska
University of Gdansk, Poland

INTRODUCTION

The aim of this contribution is to investigate the key notions characterizing the academia of the twenty-first century, especially the sphere of online learning. To present the mentioned relations and complexities within academic learning itself, the *5S Model of Online Academia* has been created and discussed by this author. The proposed model focuses the discussion about online academia on such elements as subject, situation, spirit, stakeholder, senses and subject, and their subcategories, to elaborate on the multilevel nature of universities. This framework provides an insight into the current complexities of online academia and offers an insight into its future perspectives.

BACKGROUND

There are some important factors of modern academia that shape its current state. First, academia is characteristic of its dual nature, it *has been an institutionalized space struggling to secure time for thought, consideration and the slower, time-consuming and lengthy scholarly and scientific conduct deliberately detached from the faster pace of capitalist production, media, politics and their ideological apparatuses; at the same time, it has been a symbol of and an instrument of modern progress, where individual academics and scientists have formed disciplinary associations and alliances, and advocated (to various degrees, and in diverse incarnations), socio-political, economic, scientific and cultural change* (Vostal, 2016: 7). Modern academia can also be discussed by looking at it as a complex adaptive system. It

can be characterized, among others, by nonlinear behavior, visible in disproportionate responses. Other important features of academia are independence, intelligence, learning and self-organization. Moreover, universities are the places of faculty disagreements and different points of control (Rouse, 2016). Academia can also be studied through the prism of its key determinants. Examining the growing role of technology in the life of universities and high schools, non-living entities such as computer systems, hardware, software and mobile technologies determine the way teaching and research are conducted. These notions are studied in the work by Davey and Tatnall (2012) who discuss the notion of technological adoption, focusing on school management software. In the past, traditional dissemination channels were used to gather and share knowledge. Nowadays there are *networked scholars* who use participatory technologies and online social networks in their research (Veletsianos, 2016).

ONLINE ACADEMIA- KEY ELEMENTS AND CHARACTERISTICS

Modern education differs from what could be observed some years ago; teaching and learning of the twenty-first century involve not only various learners in terms of their age, gender and background, but also diversified methods of encoding and decoding knowledge. As has already been mentioned, modern academia relies on technology in all spheres of activity. Murphy, Kalbaska, Horton-Tognazzini and Cantoni (2015) discuss in their contribution four categories of online learning: resources, tutorials, courses and

MOOCs. Apart from different materials and tools used in online learning, one of the key notions is connected with *online ludic corporate identity*, represented in the growing popularity of ludism among the broadly understood stakeholders. It is visible in, among others, fun and relaxation that accompany shopping or telling jokes and funny stories by workers (Bielenia-Grajewska, 2016). Analysing the sphere of teaching and learning, *ludic academia*, responding to the needs of stakeholders, is connected with using, for instruction, the tools associated in the past exclusively with leisure activities. One example is the application of online games in teaching and learning. Since games are not associated with spending free time as it was in the past, the appearance of serious games has changed instruction in the twenty-first century. The usage of serious games can be examined by using the 5P Model of Studying Serious Games (Bielenia-Grajewska, 2016). The first element is called *Problem-solving* and is connected with information processing and decision making. In the case of educational games, it can enhance the capacity of students to analyse and make decisions in virtual reality. The second notion is *Personae*; and is connected with elements reflecting organizational identity. In the case of educational games used for academia, verbal and non-verbal representations create the identity of a given university and show its mission and vision. The third element- *People*- stresses the lack of fixed hierarchy in games and interactions among people representing different levels on the real organizational ladder. Thus, serious games may be the place where students and professors interact in a more informal way. The next notion is called *Proficiency*; serious games are often offered in both synchronous and asynchronous ways, in a relaxed atmosphere and, consequently, their efficiency is higher than standard learning tools. The last element is *Persuasion*. Since the purpose of serious games is to advertise products or services or to evoke certain behaviours, they can be used

effectively in academia to attract new students or make current learners more interested in the instruction. For example, with the appearance of MOOCs and other educational offers available online, modern education is not restricted in terms of geography, prior levels of knowledge or types of accessible educational tools as it was before. Consequently, learners, teachers and instruction environments of the twenty-first century fail easy categorization. As has been mentioned earlier in this contribution, e-universities have become more and more popular nowadays. The reasons for this are mainly connected with technological development, including the growing popularity of the Internet, and the consequent needs and expectations of stakeholders towards educational services. Thus, nowadays students have the opportunity to have their courses run on educational platforms (e.g. Moodle), take part in massive online open courses (MOOCs) and participate in webinariums or chats with professors. The access to these resources is open or restricted only to the students of a given university or the participants of a given course. Although open online courses have many advantages, offering education to all interested in learning, Schulze, Leigh, Sparks & Spinello (2017) draw attention to the fact that MOOCs are characterized by low completion rates, taking into account the high number of registered users. Apart from MOOCs, gamification is used in online academia. It is defined as using the techniques and thinking characteristic of games for making people behave in a given way, promoting some learning styles and solving problems. In the sphere of education, such elements as points, virtual awards or leaderboards are used in courses to stimulate the interest of students in the learning content (Bielenia-Grajewska, 2015b). It should be mentioned, however, that apart from concentrating on education directed at many participants, online learning also focuses on individuals at the same time, offering online tutoring and mentoring (Berg, 2010).

5S Model of Online Academia

To study the complexities of offering tuition on the web, the 5S Model of Online Academia has been created and discussed by this author. The author has presented the key elements of modern academia (present in both its online and offline versions), studying their role in education by paying special attention to e-learning.

SUBJECT

The first element of the studied model- Subject- includes what constitutes the topic of learning. Subject is connected with students' interests, needs and expectations towards the curriculum, with such frequent sub-notions as *salaries*, solutions, stepping stones, skills and schemes. Starting with salaries, selecting subjects is often connected with discrepancies in remuneration levels and the potential increase in salaries together with one's higher skills or experience. Salary may also be the determinant of choosing a course for vocational training, with not the best but the cheapest option

being selected by participants. *Solutions*, on the other hand, encompass the ways problems and challenges can be dealt with. They stimulate a better understanding of the topic. In the case of online learning, solutions are often discussed in forums or chat rooms. *Stepping stones* are connected with stages of accomplishing educational goals. In the case of online courses, they often mark the level of knowledge acquired in a given stage and in a given period of time, or one necessary to embark on a new learning stage, in a way forcing students to learn systematically. *Skills* also constitute an important aspect of learning. In the era of online schooling, people can gain new knowledge also on the net, without leaving their houses or offices. The restrictions related to geographical barriers do not apply to the same extent as in the past, with the material ones being relevant also only to some extent since many online courses are offered free of charge. *Schemes* are related to consecutiveness in education. The necessity to repeat knowledge is solved in e-learning. Online courses are effective for novel and complicated materials that require repetition since students can listen to the tape or watch the video as many times as necessary.

Figure 1. 5S model of online academia (created by this author)



SITUATION

Situation is broadly understood as the learning environment. From the micro sphere, it can be understood as a course or class, whereas the more macro approach focuses on learning policies and pan-national regulations. It envelops source, saturation point (of information), season, scene and 3 *Situational S-components*: Storm, Silence and Struggle. The first sub-element is *Source*, being connected with the origin of a given situation. It encompasses the reasons why something happened. In the learning context, it can be connected with the rationale why someone wanted to start a new course or the whole curriculum. The next one is *saturation point*, being linked with the amount of information that has so far been available for future students; the knowledge they possess determines their interest in engaging in the offered type of schooling. In addition, an offer that in a way exceeds the mentioned saturation point and does not evoke any interest among the target audience is substituted with the more novel one. Saturation point may also be treated as the number of professionals possessing this educational profile and already present on the job market. *Season* is associated with determinants related to time of day, days, months and other time-related factors connected with how different notions determine organizational performance. For example, daily students usually study from Monday to Friday, with most of their courses planned for mornings and afternoons, whereas extramural students usually frequent courses during weekends or evenings. In comparison with standard forms of teaching and learning, online courses offer the possibility to study when one can and wants, taking into account time differences and one's professional duties or household chores. Even if a webinarium is planned, students who cannot attend it synchronously can later watch the recorded video. *Style* reflects the way one deals with knowledge, with the way instruction is coded and decoded. In the case of education, teacher's style may determine the way a given matter is understood. When the manner

of instruction is not favored by students, even a simple matter may be difficult to follow. It holds especially true in the case of online learning when it is difficult to observe if students are interested and pursue their studies without the control of an instructor. *Simulation* is connected with causing certain reactions, usually among learners. When a response is caused, then more complicated forms of interaction can be made. Traditional and online courses should stimulate participants to search for new sources of knowledge, to find answers for their queries and to pose new questions. *Scene* is connected with the geographical location of learning, determining when interaction takes place. In the case of online instructions, scene is not linked to one geographical location. An online course has an Internet address in order to be found in the web jungle. Learning may also be discussed through the prism of 3 *Situational S-components*, such as *Storm*, *Silence* and *Struggle*, mainly mirroring the type of situations in learning, with silence used to denote unproblematic conditions and standard ways of interactions. Storm, on the other hand, depicts crisis or changes in organizations offering tuition or among students or teachers, with struggle used to describe teachers' or students' engagement in protecting their workplace and place of study.

E

SENSES

Senses encompass such notions as Speech, Scent, Sound and Sign. *Speech*, being the verbal representation of communication, can be sub-categorized into standard communication, slang and symbolic language. Standard communication includes accepted forms of interaction, being used in everyday performance of organizations. It mainly encompasses literal forms, making it possible to express one's point of view effectively. In addition to standard forms of communication used in teaching, students and teachers often rely on symbolic language, allowing for a more subjective type of expression. The mentioned figurative interaction encompasses metaphors, metonymies,

and allegories. In addition to literal and non-literal dialogic representation, students and teachers rely on slang or professional dialects. Since the verbal sphere of education constitutes the most often used channel of communication, attention is paid to the way words determine modern instruction. This thought can also be used to discuss the introduction of new terms in any discipline, showing, for example, how, among different linguistic tools, metaphors play important functions in managing knowledge. Relying on well-known domains, they are effective in coding and decoding knowledge. Analyzing specialized communication, professionals relying on symbolic language communicate quickly and effectively. Since metaphors offer efficient communication, dense in terms of signs and words used, it turns out to be effective in online communication that is connected with being economical in using linguistic representation. Metaphors are also important at the organizational level, showing the profile of company activities and organizational culture (Bielenia-Grajewska, 2015a). The performance of refreshing the mind is connected with several ways the application of metaphors in modern teaching can be understood. One method is to look at symbolic language from the learners' perspective. Metaphors make complicated and novel concepts easier to understand since they relate to well-known concepts and phenomena. In addition, metaphors make the teacher's profession interesting; teachers searching for new metaphors in their instruction make their own job adventurous. The next sense- *Scent*- is associated with the olfactory dimension of communication and learning, represented in sensory marketing, with things like special scents characteristic of educational settings. This dimension may also be a part of a given lesson, such as one devoted to presenting some dishes representative of a given culture. In the case of online learning, scent has to be substituted with other forms of expression, often relying more on words and sounds to evoke similar feelings. *Sounds* represent the auditory sphere of educational interactions. In the case of online learning, the speed of online connection

determines the quality of voice. *Signs*, although they can have different dimensions, are often used in the pictorial sphere of representation.

SPIRIT

Spirit is connected with learning atmosphere and teaching rules. The first approach to learning spirit and its role for online instruction can be studied through such notions as Standards, Sanctions, Space and System. *Standards* are mainly understood as the set of rules and expectations that both instructors and learners have to meet in order to fulfill their roles. At the micro level, it is connected with respecting regulations connected with frequenting online courses, such as punctuality in the case of assignments, observing net etiquette on forums, etc. *Sanctions*, understood as an element of social control, are discussed through the prism of violating classroom codes of behavior and the consequences related to these acts of negligence. In the case of academia, they may result in suspension or relegation. *Space* is connected with the arrangement of learning place. In the case of academia, traditional universities are often supplemented with virtual classrooms. *System* can be researched through the notion of learning culture. Examining academia, it may be university culture that is recognized by both learners and instructors. There are different types of learning culture that do not foster knowledge exchange. One of them is *sweep learning culture*, observable in educational organizations that do not face problems and do not solve them, but rather only try to survive till the semester ends. This sort of atmosphere facilitates dissatisfaction of students and learners, being represented especially by a high rate of dropouts. Another type of learning spirit is *surface learning culture* that is superficial and does not rely on deep values. In this case, only basic materials are offered, with the lack of student immersion into the broader context of disciplines. The next type is *stream learning culture*, following the most popular model

of teaching and running organizations, built on widely accepted rules and values, and likely to be met in many learning entities. In the latter one, there is limited place for own ideas and creativity.

STAKEHOLDERS

Stakeholders constitute the last dimension that can be divided into inner and outer subtypes. As far as subcategorization is concerned, the inner dimension encompasses *Staff* (teachers, school owners) and *Students*, whereas the outer one can be classified as *Shoppers* (encompassing the ones searching for educational offers before making decisions); both groups may be viewed through the perspective of status. *Status* may reflect such notions as one's position within a studied hierarchy. In the case of staff, it concerns one's place on an educational ladder, whereas shoppers are often labelled by taking into account their social or professional position and potential interest in the offered educational products or services.

FUTURE RESEARCH DIRECTIONS

Manning, Wong, and Tatnall (2010) discuss that nowadays many universities combine traditional learning with e-learning platforms, using both face-to-face interactions and distance instructions. It can be expected that in the future even more courses will be offered online, taking into account the expectations and needs of students who prefer online learning or their geographical location does not allow them to study in a traditional way. Moreover, Sancassani, Casiraghi, Corti and Trentinaglia (2014) stress the role of MOCCs in improving the quality of traditional learning as well as its cost. Taking the last money-related factor into account, it can be predicted that more educational institutions will be interested in offering instruction that is characterized by economic efficiency for both educators and students.

CONCLUSION

The growing popularity of technology and diversified demands and expectations of academia stakeholders have led to crucial changes in the way universities perform nowadays. As has been discussed in this contribution, online academia is connected with offering its tuition and running its communication on the Internet. The proposed model offers a complex insight of how online academia can be investigated.

REFERENCES

- Berg, G. A. (2010). *Cases on Online Tutoring, Mentoring, and Educational Services: Practices and Applications: Practices and Applications*. Hershey, PA: IGI. doi:10.4018/978-1-60566-876-5
- Bielenia-Grajewska, M. (2015a). The role of figurative language in knowledge management. In M.Khosrow-Pour (Ed.), *Encyclopedia of Encyclopedia of Information Science and Technology* (pp. 4728-4736). Hershey, PA: IGI. doi:10.4018/978-1-4666-5888-2.ch464
- Bielenia-Grajewska, M. (2015b). Jak "rozegrać" zajęcia, czyli o roli gamifikacji w dydaktyce akademickiej. In D. Becker-Pestka & E. Kowalik (Eds.), *Wyzwania współczesnej pedagogiki* (pp. 245–256). Gdańsk: Wyższa Szkoła Bankowa.
- Bielenia-Grajewska, M. (2016). Serious games and their application in creating corporate identity. In R. Nakatsu, P. Ciancarini, & M. Rauterberg (Eds.), *Handbook of Digital Games and Entertainment Technologies* (pp. 1–18). Singapore: Springer. doi:10.1007/978-981-4560-52-8_53-1
- Davey, B., & Tatnall, A. (2012). Using ANT to guide technological adoption: The case of School Management Software. *International Journal of Actor-Network Theory and Technological Innovation*, 4(4), 47–61. doi:10.4018/jantti.2012100103

- Manning, K., Wong, L., & Tatnall, A. (2010). Aspects of e-Learning in a University. *International Journal of Actor-Network Theory and Technological Innovation*, 2(4), 43–52. doi:10.4018/jantti.2010100105
- Murphy, J., Kalbaska, N., Horton-Tognazzini, L., & Cantoni, L. (2015). Online Learning and MOOCs: A Framework Proposal. In I. Tussyadiah & A. Inversini (Eds.), *Information and Communication Technologies in Tourism 2015. Proceedings of the International* (pp. 847–858). Cham: Springer. doi:10.1007/978-3-319-14343-9_61
- Rouse, W. B. (2016). *Universities as Complex Enterprises: How Academia Works, Why It Works These Ways, and Where the University Enterprise Is Headed*. Hoboken, NJ: Wiley. doi:10.1002/9781119245872
- Schulze, A. S., Leigh, D., Sparks, P., & Spinello, E. (2017). Massive Open Online Courses and Completion Rates: Are Self-Directed Adult Learners the Most Successful at MOOCs? In F. Topor (Ed.), *Handbook of Research on Individualism and Identity in the Globalized Digital Age* (pp. 24–49). Hershey, PA: IGI Global; doi:10.4018/978-1-5225-0522-8.ch002
- Veletsianos, G. (2016). *Social Media in Academia: Networked Scholars*. New York: Routledge.
- Vostal, F. (2016). *Accelerating Academia: The Changing Structure of Academic Time*. New York: Palgrave Macmillan. doi:10.1057/9781137473608
- Bielenia-Grajewska, M. (2012). Linguistic aspects of informal learning in corporate online social networks. In V. P. Dennen & J. B. Myers (Eds.), *Virtual Professional Development and Informal Learning via Social Networks* (pp. 93–112). Hershey, PA: IGI Publishing. doi:10.4018/978-1-4666-1815-2.ch006
- Bielenia-Grajewska, M. (2014). Corporate Online Social Networks and Company Identity. In K. Michael & R. Abbas (Eds.), *Privacy and Security Issues in Social Network. Encyclopedia of Social Network Analysis and Mining* (pp. 270–275). New York: Springer. doi:10.1007/978-1-4614-6170-8_337
- Kyei-Blankson, L., Blankson, J., Ntuli, E., & Agyeman, C. (2016). *Handbook of Research on Strategic Management of Interaction, Presence, and Participation in Online Courses*. Hershey, PA: IGI Publishing. doi:10.4018/978-1-4666-9582-5
- Niess, M. L., & Gillow Wiles, H. (2015). *Handbook of Research on Teacher Education in the Digital Age*. Hershey, PA: IGI Publishing. doi:10.4018/978-1-4666-8403-4
- Raisinghani, M. S. (2013). *Curriculum, Learning, and Teaching Advancements in Online Education*. Hershey, PA: IGI Publishing. doi:10.4018/978-1-4666-2949-3
- Sancassani, S., Casiraghi, D., Corti, P. & Trentinaglia, N. (2014). MOOC, OER e l’approccio “flipped classroom”: due case study di transizione in ambito scolastico e aziendale, *Form@re, Open Journal per la formazione in rete*, 1/14, 49–59.
- Tatnall, A., & Davey, B. (2003). ICT and Training: A Proposal for an Ecological Model of Innovation. *Journal of Educational Technology & Society*, 6(1), 14–17.

ADDITIONAL READING

Betts, K., Kramer, R., & Gaines, L. L. (2013). *Online Faculty and Adjuncts: Strategies for Meeting Current and Future Demands of Online Education Through Online Human Touch Training and Support*. Hershey, PA: IGI Publishing. doi:10.4018/978-1-4666-2949-3.ch007

Unnithan, C., & Tatnall, A. (2014). Actor-Network Theory (ANT) Based Visualisation of Socio-Technical Facets of RFID Technology Translation: An Australian Hospital Scenario. [IJANTTI]. *International Journal of Actor-Network Theory and Technological Innovation*, 6(1), 31–53. doi:10.4018/ijantti.2014010103

KEY TERMS AND DEFINITIONS

E

Academia: The academic world constituting of teachers, students and other stakeholders participating in teaching and learning.

E-Learning: Learning by using tools and materials available online.

Online Academia: Teaching and learning conducted by the stakeholders of the academic world on the Internet.

Online Learning Propelled by Constructivism

Kathaleen Reid-Martinez
Oral Roberts University, USA

Linda D. Grooms
Regent University, USA

INTRODUCTION

Augmenting communication in and among those in the academic, business, and military communities, the exponential advancement of science and technology has availed vast amounts of information to virtually millions of people around the globe. In conjunction with this knowledge explosion has been a growing concern for the democratization of the learning process, with constructivism driving much of the educational agenda, most particularly in online distance education. This article examines the resurgence of the constructivist approach to teaching and learning, its convergence with rapidly changing technological advances, and its relationship to future trends in online pedagogy.

BACKGROUND

While the constructivist method has been highly emphasized in the recent literature for online distance education (Brown, L. 2014; Bryant & Bates, 2015; Holzweiss, Joyner, Fuller, Henderson, & Young, 2014; Lê & Lê, 2012; “Learning Theories”, 2014; Mbatl & Minnaar, 2015; Symeonides & Childs, 2015; Thorne, 2013), it is not a new approach to learning. Presenting an early example, Socrates facilitated discourse with students asking directed questions to assist them in realizing the weaknesses in their logic and critical thinking. This enabled learners to share in the responsibility of their learning through active participation while

negotiating meaning in the creation of shared understanding. In contrast, medieval professors in later Western culture most often served as primary repositories of information along with the scrolls and velum texts found in the limited number of physical libraries available to educators. With the lecture serving as the quickest and easiest way to disseminate information to both small and large groups of individuals, it was both an efficient and effective delivery method in the shaping and forming of student knowledge, quickly becoming the standard for traditional education.

MAIN FOCUS OF THE ARTICLE

Resurgence of Constructivism

While the lecture method was the norm of information delivery for centuries in Western culture, the knowledge explosion arising from the latter part of the 20th century demanded more active learner participation. In light of this constant and rapid flux of information and knowledge, students became life-long learners compelled to use metacognitive skills to constantly evaluate and assimilate new material into their respective disciplines. As this implies, knowledge was no longer viewed as a fixed object; rather, learners constructed it as they experienced and co-created an understanding of various phenomena by collaborating and working with peers and professors as well as with the information. Now, rather than strictly acquiring

DOI: 10.4018/978-1-5225-2255-3.ch226

ing information, Duffy and Cunningham (1996) explicated that “learning is an active process of constructing ... knowledge and ... instruction is a process of supporting that construction” (p. 171).

Based on the work of Kidd (1973), Long (1983), Moore (1989), and Palmer (1993), Grooms’ (2000) *Learner Interaction Model* (see Figure 1) illustrates that in the constructivist culture, the learner perpetually interacts with these three components of learning--content, facilitator or professor, and peers--each mutually and non-discriminately influencing the other.

Critical in this process is recognizing the shifting role of the professor who becomes the *guide on the side* or content facilitator and is no longer the proverbial *sage on the stage* or content provider. The student’s role also has changed from being a passive receiver of information to an active participant in the knowledge-making process (Weller, 1988), aligning with Bandura’s (1977, 1994) concept of the autonomous learner, an important dimension of the constructivist model. Table 1, based upon an earlier model from Reid-Martinez, Grooms, and Bocarnea (2009) and Reid-Martinez and Grooms (2015), delineates these two approaches to learning.

Of special interest in the above listing is the role of community. The constructivist approach

recognizes that students do not learn strictly within the limited confines of a local educational institution, but rather within the broader international and global context of their personal lives extended through social media and multiple technologies. Consequently, the boundaries between the educational institution and the larger community become blurred creating its own unique set of opportunities and challenges.

As people work collaboratively in the learning activities and new technologies, they bring multiple worldviews and experiences to each situation often creating a plethora of perspectives. During this collaborative learning process, they must negotiate and generate meaning and solutions to problems through shared understanding. Thus, education moves from a single, solitary pursuit of knowledge to a collaborative learning community that shapes and informs responses to the environment. As noted by Fuller and Söderlund (2002), this challenges the common metaphor of the university as a self-contained village.

Rapidly Changing Distance Learning Technologies

Over the years, educators have experimented with and successfully employed multiple media for

Figure 1.
2000 Grooms, L. D.

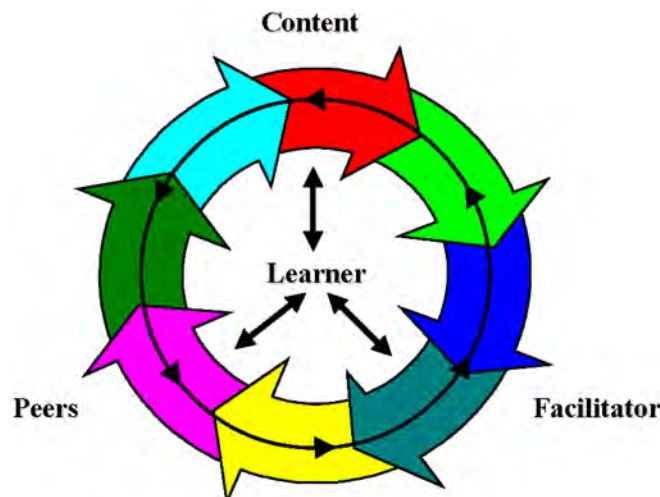


Table 1.

Approaches to Learning		
	Traditional	Constructivist
Professor	Sage on the Stage	Guide on the Side
	Content Provider	Content Facilitator
Learner	Passive Recipient	Active Participant
Knowledge	Fixed Object	Fluid
Organization of Learning	Ordered & Structured	Open & Often Chaotic
Communication	Uni-directional	Multi-directional
Primary Resource	Text & Professor	Multiple Sources
Method	Lecture	Active Process
Media	Print	Blended
Format	Structured & Individualized	Adaptive & Collaborative
Activities	Goal-Oriented	Problem-Centered
Focus of Learning	Knowledge & Understanding	Application, Analysis, Synthesis, & Evaluation
Assessment	Recall	Alternative Assessment
Community	Local Educational Institution	Integrated with Life in Global Contexts

distance learning, and today, as much as in the past, they continue to stress that pedagogy must drive technology (Rourke & Coleman, 2011). As early as the 18th century, print material was used and even today still serves an important role in distance education even as it gives way to more reliance on technology and web-based resources for collaborative development of knowledge that incorporates the diversity of learners and their contexts. After the 1930s, other media became significant with audio--including radio and audiotapes--and video--including film, public broadcasting, and cable--dominating much of the 20th century.

Much of this education was one-way based on a mass communication or one-to-many educational model. Basically, it was a rigid structure with information flowing in one direction, from the powerful and knowledgeable instructor reaching to the individual or even the larger group of students. It included elements of limited feedback through the use of such things as the penny post in the 19th century and the addition of the telephone and fax in the 20th century. Limited opportunities for face-to-face interaction were also incorporated

with some programs. Thus much of distance learning during these times remained mainly non-interactive.

By the 1990s, the advent of the Internet presented new opportunities in distance education. The result was the evolution of a new type of collaborative learning, in which the potential for interaction between the professor and the learner increased exponentially with wide-area networks accommodating synchronous and asynchronous communication. While exploring computer-mediated activities of the online learning environment, Santoro (1996) highlighted three broad categories: (a) computer-assisted instruction, which allows the computer to serve as “teacher” by structuring information delivered to the human user; (b) computer-based conferencing, which includes e-mail, interactive messaging, and group conference support systems; and (c) informatics, which refers to online public access libraries and interactive remote databases. This proliferation of the Internet unlocked the door for educational institutions to reach beyond their four walls making services accessible to students around the world through online activities.

Although the communication technologies of the 21st century--print, audio, video, digital, and the Internet--cover a broad spectrum of distance education mediums, this exponential growth in science and technology has catapulted the Internet into rapidly becoming the preferred delivery platform. Since 1995, researchers such as Cotton (1995) and others have been tracking this information along with scholars such as Bocarnea, Grooms, and Reid-Martinez (2006), Grooms and Reid-Martinez (2013), and Reid-Martinez and Grooms (2012, 2015). They continue to explore not only the trends in distance education but also the understanding of and the issues involved in aligning the environment with student needs. Typical factors include (a) the characteristics of the discipline, (b) the degree of interactivity sought in the distance learning process, (c) learner characteristics, (d) instructor traits, (e) the expansiveness of the distance education initiative, (f) the desired level of accessibility and flexibility related the delivery capacity of learning platforms and smart mobile devices as well as other methods of dissemination, and lastly (g) the availability of technical support.

In addition to the global reach of the Internet, the lines *among* communication technologies have swiftly blurred. Today in the convergence of technologies, computers, telephones, and cameras are no longer distinct entities, but can be found bundled into one small handheld gadget through the fusion of technology (McCain & Jukes, 2001). These smaller fused devices create more mobility and simultaneously provide mixed realities through virtual immersive environments embedded within traditional spaces. Continuing advances, such as that found in interactive optical sensory technology, feed this growing world that fuses the virtual with the physical (Rolf, 2012). Through this fusion, communicating with students and colleagues has become more instantaneous, integrated and complex. While vastly expanding the means of interaction and feedback, it demands greater capacity and understanding of the multiple communication modalities.

Connected to these new technologies is the capacity to enhance adaptive individualized learning. As noted by Allen and Seaman (2013), adaptive learning helps overcome the barriers to online learning, a common concern of many academic leaders. As we see in the research of such scholars as Yang, Gamble, Hung, and Lin (2014), critical thinking can be enhanced through adaptive learning in the online environment. These new adaptive learning technologies accelerate and enhance learners' problem solving ability. The fusion of technological capacity for adaptive learning with collaborative technology platforms results in individuals operating at a more advanced level and collectively harnessing greater learning and problem solving abilities from all participants.

With such rapid technological advances, today's educators are dropped into what Jacque Ellul (1964) described as the intersection of tension between humanity and technology. This struggle with the latent and manifest, intended and unintended consequences of technology exists as students and professors wrestle with cloud computing, three-dimensional immersive learning environments, and other rapidly expanding web opportunities.

Such technology facilitates greater flexibility and customization in the learning process. Lead Learning Designer at IBM, Don Morrison (2004) demonstrated how the learning process can be established within parameters and policies that most appropriately align with the primary strengths and weaknesses of each medium. He noted that among others, cost, time constraints, delivery speed, and infrastructure help determine appropriate application. Morrison's work also pointed to ways in which educational models can be designed to marry traditional and online means of moving from the simple to the more complex methods of learning.

As the above suggests, these new electronic forms of communication have forced a paradigm shift in education. This move is most avidly seen in distance learning, where even the terminology has shifted from distance education to words such

as online or e-learning. Such terms more clearly indicate the way in which learners can use multiple media to easily collaborate through a continuous integration of knowledge and social capital.

FUTURE RESEARCH DIRECTIONS

As previously discussed, the rapid growth of technology continues to herald unprecedented opportunity for distance learning, and when wed to a constructivist approach, it presents opportunities for online pedagogy that can transcend traditional modes of education. From this marriage emerges three primary factors that define the new online pedagogy: (a) *community* development--the ability to build networks and communities that cross time and geographic boundaries, (b) *structure*--the technological ability to manage vast amounts of information, and (c) *collaborative* opportunities--for shared knowledge and wisdom building in response to the complexities of a global society (Reid-Martinez, 2006).

Community

As Bocarnea et al. (2006) note, today's technologies launch a new paradigm of online learning and pedagogy, which has the potential to be communal in nature. Primarily, these technologies allow for interaction between students and professors, students and peers, and the broader community in unprecedented ways. For example, students today have greater instructor and peer access through social media and e-learning platforms. Indeed the study of mobile technology for learning in environments of high action and great distance as found in the work of Black and Hawkes (2011) points to the now even more ubiquitous capacity of mobile learning (m-learning). In turn, this poses the question of expectation--whether conscious or not--regarding ubiquity of instructor presence and community development. It also enhances the ubiquity of the student in the learning process. Indeed, learning is no longer "just in time," but

with adaptive learning and other capacities that allow for learning that is "just for you" and "just with you" through wearable smart devices and other advancing technologies. This contemporary and developing technology now allows learning to be fluid with the learner in a simultaneous and continuous nature (Reid-Martinez, 2015).

These new technological advances raise the question of boundaries in the learning process. While advancing technologies provide unprecedented opportunities for networking and building strong virtual learning communities, they do transcend geographical boundaries and normal hours of operation as well as far beyond the duration of students' formal education. As early as 2002, Young highlighted the differences between the boundaries embedded in his traditional face-to-face class and that which he encountered online. This suggests that guidelines following best practices to manage the continuous nature of virtual learning experiences are essential to prevent online instructors and students from feeling overwhelmed by the 24-7, ubiquitous opportunity for interaction.

In addition, this communal nature of the virtual learning environment provides opportunities for students to bring their local community contexts into the learning experience in direct ways as well as immediately allowing them to apply what they have learned through their study. For example, students in leadership programs can be employed full-time in leadership positions and take their learning experiences directly into their work environment through well-designed course assignments. The professor is no longer someone whom the student must wait to see in class later in the week, but rather is readily available in e-learning and m-learning platforms to serve as consultant and mentor as the student applies the principles studied that week. The professor has become the *guide on the side*. This triangulation of student-professor-content points to the need for well-designed learning experiences developed from a constructivist perspective to meet the challenges and needs of today's students. Indeed, unit-

ing the new technology with this approach meets the needs of contemporary students working in rapidly changing and highly demanding global environments.

Additionally, rapid advancement of technology creates a moving target challenge for course developers who often find themselves reacting to the technological advances rather than proactively establishing the technology's relationship to the learning process. While scholars such as Schweizer (1999) noted unsound pedagogy and inadequate design in early online courses, others such as Wang and Newlin (2002) reinforced the importance of incorporating the opposite as critical predictors of successful student performance. Just a short time later, others such as Beetham and Sharpe (2007) and Rourke and Coleman (2011) continued to reinforce the critical role of pedagogy in using technology. Such approaches assure that good pedagogy is the driver of learning, not new technologies. This helps resolve Bocarnea et al.'s (2006) observation and concern that theory typically "follows technology in desperate attempts to describe the impact of an already existing and rapidly fading ... technological reality" (p. 385).

This posits that staying focused on strategy and content design remains the dominant challenge. Online pedagogy, the science *of* and *about* online education, provides perspective to assist in focusing and maintaining the balance necessary for creating excellent online learning experiences.

Structure

Heralded just over 20 years ago by Negroponte (1995), the information age is collapsing on itself as the amount of online information is becoming unbearable. After the scramble to have everything digitized, the primary challenge today is how to create meaningful knowledge from such massive amounts of data. Quality of knowledge, in contrast to quantity, drives the heart of this concern. In light of this overload, *structure* is essential to online knowledge development.

Related to structure, is the development of open-source initiatives. While most often understood as the software that is open for use and modification by the public, the phrase has become a recognized attribute ascribed to multiple endeavors, such as knowledge-building. The open-source nature of online initiatives pushes a new model for managing learning and knowledge-building through the communal process. It allows diverse individuals from various locations to combine information from multiple sources into distributed knowledge networks. Through this open-source structure, participants interact to share experiences and knowledge, thereby expanding their awareness of new concepts and differing approaches to problem-solving as they modify the information in the open-source environment and re-distribute it back to fellow participants (Bocarnea et al., 2006).

As this suggests, through interaction, participants build complex webs of knowledge in the open-source cyberspace. The technology provides the structure to create and maintain webs of knowledge, and it also grants ease of access globally to those interested in that knowledge. In the process, knowledge is given away to others who in turn begin to use it in multiple ways while beginning the next evolution of knowledge development as they add to and transform the knowledge base they accessed through the open-source structure. With this transformation through this structural capacity is the transference of power and control that becomes less centrist and more distributed globally.

Collaborative Knowledge-Building

As noted above, interaction is the key for the development of open-source knowledge-building. While scholars such as Cederbom and Paulsen (2001) posited learning as a behavioral change, others hold that it is simply when learners meet needs and establish goals for attaining knowledge (Ponton & Carr, 2000). Referring to this process as an implied contract, Keirns (1999), along with the above scholars, suggested that if online learn-

ing is used, structure is critical to allow students to advance in their knowledge.

Because “engaging the learner in reflective and collaborative thought processes ... results in the most effective learning, whether the setting is a traditional classroom or an online environment” (Cox & Cox, 2008), the design of the online course becomes the principal structure to assure learners’ goals are achieved. In that course design, one way of attaining the collaborative thought processes is to incorporate structured interaction. Not only does this interaction provide for collaborative thoughts and knowledge development, it provides multiple opportunities for faculty to prompt students’ critical thinking. In keeping with this, researchers such as Pelech and Pieper (2010) went so far as to clearly lay out, among other approaches, the roles of visual literacy, bridging questioning, and kinesthetic activities in applying Bloom’s taxonomy for constructivist learning.

While these dimensions of critical thinking undergird contemporary collaborative capacity in online learning, other psychological dimensions are becoming more prevalent in understanding how critical thinking and collaboration can increase. For example, Rolf (2012) surfaced the complexities of virtual collaboration in his study of contemporary technologies’ role in creating mixed realities. Another example is Mabrito’s (2011) study of vicarious interaction--that is, observing not participating in peer as well as faculty-peer online interaction--which can generate more idea awareness in a constructivist learning environment and has potential impact on how collaboration occurs.

A common way to collaborate in the online environment is through interaction. Blair (2002) suggested that stronger relationships are forged through increased interaction frequency. Increased interaction relates to higher learner commitment due to the socialization the learner experiences as a participant in the knowledge-building process. Thus, learner perception of interaction plays an important role in student achievement, satisfaction, and quality of learning.

Again, in the collaborative nature of the constructivist online culture, interaction perpetually occurs between learners and content, learners and instructors, learners and peers, and learners and external experts with each type of interaction reinforcing and fostering collaborative knowledge-building. Teaching disappears and “communication of information rules, where information is available to all and in abundance” (Brown, T. H., 2015, p, 228). With this in mind, online course design must include best structures to capitalize on this collaborative interaction.

CONCLUSION

As the above suggests, the advent of online learning education has not just provided opportunity to disseminate information in a new medium, it has radically adjusted the distance learning paradigm in terms of distribution methods, community building, knowledge development, and learning. The use of 21st century technology is rapidly closing the gap of the communication immediacy essential in developing communities of practice for knowledge-building. With their open-source networks, these new technologies encourage and actively support constructivist pedagogy in the distance education paradigm. Most of all, distance education through its constructivist pedagogy and contemporary technologies has the technical capacity to fulfill its greatest potential, which is to reach every learner who desires to participate in the knowledge-building process. The result can be a democratization of education not previously seen, allowing for shifts in power and control throughout societies.

REFERENCES

- Allen, I. E., & Seaman, J. (2013). *Grade Change: Tracking Online Education in the United States*. Babson Survey Research Group and Quahog Research Group. Retrieved from <http://www.onlinelearningsurvey.com/reports/gradechange.pdf>
- Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ: Prentice Hall.
- Bandura, A. (1994). Self-efficacy. In V. S. Ramachaudran (Ed.), *Encyclopedia of human behavior* (Vol. 4, pp. 77–81). New York: Academic Press.
- Beetham, H., & Sharpe, R. (Eds.). (2007). *Rethinking pedagogy for a digital age: Designing and delivering e-learning*. London: Routledge.
- Black, J., & Hawkes, L. W. (2011). Making the case for mobile collaborative learning for reading comprehension on handheld computers. In A. Rourke & K. Coleman (Eds.), *Pedagogy leads technology: Online learning and teaching in higher education: New technologies, new pedagogies* (pp. 143–158). Champaign, IL: Common Ground Publishing.
- Blair, J. (2002). The virtual teaching life. *Education Week*, 21, 31–35.
- Bocarnea, M. C., Grooms, L. D., & Reid-Martinez, K. (2006). Technological and pedagogical considerations in online learning. In A. Schorr & S. Seltsmann (Eds.), *Changing media markets in Europe and abroad: New ways of handling information and entertainment content* (pp. 379–392). Pabst Science Publishers.
- Brown, L. (2014). Constructivist learning environments and defining the online learning community. *Journal of Science Education and Technology*, 9(4), 1–6.
- Brown, T. H. (2015). A reflection on Barber, Donnelly, and Rizvi (2013): “An avalanche is coming: Higher education and the revolution ahead”. *International Review of Research in Open and Distributed Learning*, 16(4), 227–234. doi:10.19173/irrodl.v16i4.1952
- Bryant, J., & Bates, A. J. (2015). Creating a constructivist online instructional environment. *Tech Trends. Linking Research & Practice to Improve Learning*, 59(2), 17–22. doi:10.1007/s11528-015-0834-1
- Cederblom, J., & Paulsen, D. W. (2001). *Critical reasoning: Understanding and criticizing arguments and theories* (5th ed.). Belmont, CA: Wadsworth.
- Cotton, C. (1995). Time-and-place independent learning: The higher education market for distance learning emerges. *Syllabus*, 8(5), 37–39.
- Cox, B., & Cox, B. (2008). Developing interpersonal and group dynamics through asynchronous threaded discussions: The use of discussion board in collaborative learning. *Education*, 128(4), 553–565.
- Duffy, T. M., & Cunningham, D. J. (1996). Constructivism: Implications for the design and delivery of instruction. In D. H. Jonassen (Ed.), *Handbook of research for educational communications and technology* (pp. 170–198). New York: Simon Schuster Macmillan.
- Ellul, J. (1964). *The technological society* (J. Wilkinson, Trans.). New York: Vintage Books.
- Fuller, T., & Söderlund, S. (2002). Academic practices of virtual learning by interaction. *Futures*, 34(8), 745–760. doi:10.1016/S0016-3287(02)00018-6
- Grooms, L. D. (2000). Interaction in the computer-mediated adult distance learning environment: Leadership development through online education. *Dissertation Abstracts International*, 61(12), 4692A.
- Grooms, L. D., & Reid-Martinez, K. (2013, November). *Cross-Cultural Learning for Leadership Resilience and Sustainability*. Presented at the 15th annual International Leadership Association Conference, Montreal, Canada.

- Holzweiss, P. C., Joyner, S. A., Fuller, M. B., Henderson, S., & Young, R. (2014). Online graduate students perceptions of best learning experiences. *Distance Education*, 35(3), 311–323. doi:10.1080/001587919.2015.955262
- Keirns, J. (1999). *Designs for self-instruction: Principles, processes, and issues in developing self-directed learning*. Needham Heights, MA: Allyn & Bacon.
- Kidd, J. R. (1973). *How adults learn*. Chicago: Follett Publishing.
- Lê, Q., & Lê, M. (2012). New challenges in web-based education. In T. Lê & Q. Lê (Eds.), *Technologies for enhancing pedagogy, engagement and empowerment in education: Creating learning-friendly environments* (pp. 58–65). Hershey, PA: IGI Global. doi:10.4018/978-1-61350-074-3.ch005
- Long, H. B. (1983). *Adult learning: Research and practice*. New York: Cambridge.
- Mabrito, M. (2011). Lurking and learning: A study of vicarious interaction in the online classroom. In A. Rourke & K. Coleman (Eds.), *Pedagogy leads technology: Online learning and teaching in higher education: New technologies, new pedagogies* (pp. 105–116). Champaign, IL: Common Ground Publishing.
- Mbati, L., & Minnaar, A. (2015). Guidelines towards the facilitation of interactive online learning programmes in higher education. *International Review of Research in Open and Distributed Learning*, 16(2), 272–287. doi:10.19173/irrodl.v16i2.2019
- McCain, T., & Jukes, I. (2001). *Windows in the future: Education in the age of technology*. Thousand Oaks, CA: Corwin Press.
- Moore, M. G. (1989). Editorial: Three types of interaction. *American Journal of Distance Education*, 3(2), 1–7. doi:10.1080/08923648909526659
- Morrison, D. (2004). *What do instructional designers design?* Retrieved from <http://www.morrisonco.com>
- Negroponte, N. (1995). *Being digital*. New York: Alfred A. Knopf.
- Palmer, P. J. (1993). *To know as we are known: Education as a spiritual journey*. San Francisco: Harper Collins.
- Pelech, J., & Pieper, G. (2010). *The comprehensive handbook of constructivist teaching: From theory to practice*. Charlotte, NC: Information Age.
- Ponton, M. K., & Carr, P. B. (2000). Understanding and promoting autonomy in self-directed learning. *Current Research in Social Psychology*, 5(19). Retrieved from <http://www.uiowa.edu/~grpproc/crisp/crisp.5.19.htm>
- Reid-Martinez, K. (2006). *What's that in your hand?* Presented at the 2006 General Assembly, International Conference of Educators, Indianapolis, IN.
- Reid-Martinez, K., & Grooms, L. D. (2012, November). *Building a community of leaders through communication using a constructivist approach*. Presented at the annual convention of the National Communication Association, Orlando, FL.
- Reid-Martinez, K., & Grooms, L. D. (2015). Constructivism as the driver of 21st century online distance education. In M. Khosrow-Pour (Ed.), *Encyclopedia of Information Science and Technology* (3rd ed.; pp. 2229–2238). Hershey, PA: IGI Global; doi:10.4018/978-1-4666-5888-2.ch216
- Reid-Martinez, K., Grooms, L. D., & Bocarnea, M. C. (2009). Constructivism in online distance education. In *Encyclopedia of Information Science and Technology* (2nd ed.; Vol. 2, pp. 701–707). Hershey, PA: IGI Global. doi:10.4018/978-1-60566-026-4.ch114

Rolf, D. (2012). Mixed realities: Human interaction technologies. In T. Lê & Q. Lê (Eds.), *Technologies for enhancing pedagogy, engagement and empowerment in education: Creating learning-friendly environments* (pp. 209–216). Hershey, PA: IGI Global. doi:10.4018/978-1-61350-074-3.ch018

Rourke, A., & Coleman, K. (Eds.). (2011). *Pedagogy leads technology: Online learning and teaching in higher education: New technologies, new pedagogies*. Champaign, IL: Common Ground Publishing.

Santoro, G. M. (1996). What is computer-mediated communication? In Z. L. Berge & M. P. Collins (Eds.), *Computer mediated communication and the on-line classroom* (Vol. 1, pp. 11–27). Cresskill, NY: Hampton Press.

Schweizer, H. (1999). *Designing and teaching an on-line course: Spinning your web classroom*. Boston: Ally & Bacon.

Symeonides, R., & Childs, C. (2015). The personal experience of online learning: An interpretive phenomenological analysis. *Computers in Human Behavior*, 51, 539–545. doi:10.1016/j.chb.2015.05.015

Theories, L., & Engagement, S. (2014). Article. *ASHE Higher Education Report*, 40(6), 15–36. doi:10.1002/aehe.20018

Thorne, J. A. (2013). Biblical online education: Contributions from constructivism. *Christian Education Journal*, 10(1), 99–109.

Wang, A., & Newlin, M. (2002). Predictors of performance in the virtual classroom. *T. H. E. Journal Online*, 29(10), 21–22, 26–28.

Weller, H. G. (1988). Interactivity in microcomputer-based instruction: Its essential components and how it can be enhanced. *Educational Technology*, 28(2), 23–27.

Yang, Y.-T. C., Gamble, J. H., Hung, Y.-W., & Lin, T.-Y. (2014). An online adaptive learning environment for critical-thinking-infused English literacy instruction. *British Journal of Educational Technology*, 45(4), 723–747. doi:10.1111/bjet.12080

Young, J. R. (2002). The 24-hour professor: On-line teaching redefines faculty members' schedules, duties, and relationships with students. *The Chronicle of Higher Education*, 48(38), A31–A33.

ADDITIONAL READING

Burge, E., Gibson, C. C., & Gibson, T. (Eds.). (2011). *Flexible pedagogy, flexible practice: Notes from the trenches of distance education*. Edmonton, Alberta: Athabasca University.

Denton, D. W. (2012, July/August). Enhancing instruction through constructivism, cooperative learning, and cloud computing. *TechTrends*, 56(4), 34–41. doi:10.1007/s11528-012-0585-1

Fosnot, C. (1996). *Constructivism: Theory, perspectives, and practice*. New York, NY: Teachers College Press.

Gazi, Z., & Aksai, F. (2011). *Handbook of online pedagogy*. Saarbrücken, Germany: LAP LAMBERT Academic Publishing.

Gülseçen, S. (2012). Digital learning environments and student centered curriculum in a university context. In T. Lê & Q. Lê (Eds.), *Technologies for enhancing pedagogy, engagement and empowerment in education: Creating learning-friendly environments* (pp. 209–216). Hershey, PA: IGI Global. doi:10.4018/978-1-61350-074-3.ch009

Hai-Jew, S. (2012). *Constructing self-discovery learning spaces online: Scaffolding and decision making technologies*. Hershey, PA: Information Science Reference. doi:10.4018/978-1-61350-320-1

McHaney, R. W. (2012). Knowledge spaced for online discovery learning. In S. Hai-Jew (Ed.), *Constructing self-discovery learning spaces online: Scaffolding and decision making technologies* (pp. 72–94). Hershey, PA: IGI Global. doi:10.4018/978-1-61350-320-1.ch005

Shi, Y., Fan, S., & Yue, Y. (2012). Empowering students in computer-supported education. In T. Lê & Q. Lê (Eds.), *Technologies for enhancing pedagogy, engagement and empowerment in education: Creating learning-friendly environments* (pp. 198–207). Hershey, PA: IGI Global. doi:10.4018/978-1-61350-074-3.ch017

Trifonas, P. P. (2012). *Learning the virtual life: Public pedagogy in a digital world*. New York: Routledge.

KEY TERMS AND DEFINITIONS

Autonomous Learner: An individual who takes responsibility for his or her learning.

Collaborative Learning: The process in which individuals negotiate and generate meaning and solutions to problems through shared understanding.

Computer-Assisted Instruction: The computer serves as the “teacher” by structuring information delivered to the human user.

Computer-Based Conferencing: E-mail, interactive messaging, and group conference support systems.

Constructivism: An approach in which students share responsibility for their learning while negotiating meaning through active participation in the co-creation of shared understanding within the learning context.

Distributed Knowledge: Information dispersed throughout a community of practice and not held by any one individual.

Informatics: Online public access libraries and interactive remote databases.

Interaction: mutual communicative exchange between individuals.

Science Animation and Students' Attitudes

E**Sivasankar Arumugam***Sri Venkateswara College of Education, India***Nancy Nirmala***Christ the King Matric Higher Secondary School, India*

INTRODUCTION

Teaching is a kind of social engineering that does not deal with lifeless machines and hard wares; it cultivates tender minds into brave hearts that in turn with brimming confidence is going to build the society. In this endeavour every student is important, every detail present in each topic is essential and every batch of students is precious. The digital transformations that are taking place in educational arena had opened newer avenues for the teachers, learners, administrators and researchers in the form of animation. Animation in its own virtue along with enthusiasm of digital native learners had grown leaps and bounds. The need for animation in third world countries seems to be pinning as teacher- pupil ratio is alarmingly high when comparing with many of the Western counterparts and majority of the education machinery is examination ridden. When the teacher has to run behind the content and ensure zero failure it becomes imperative to teach science with animation. And the metamorphosis that animation could take as per the projection of experts is very promising. This study makes an experimental approach with science animation in secondary school classes and its impact through achievement and attitude.

Meaning of Animation

Etymologically animation has got Latin origin animatio from animare which means the condition of being alive or giving life. Rapid display of images, pictures or frames is called as animation.

The technique of capturing successive frames of pictures or positions of toys or models that create an illusion of movement while the movie is shown as a sequence gives life to animation. In other words a collection of static images joined together and shown consecutively so that they appear to move is called as animation.

Evolution of Animation

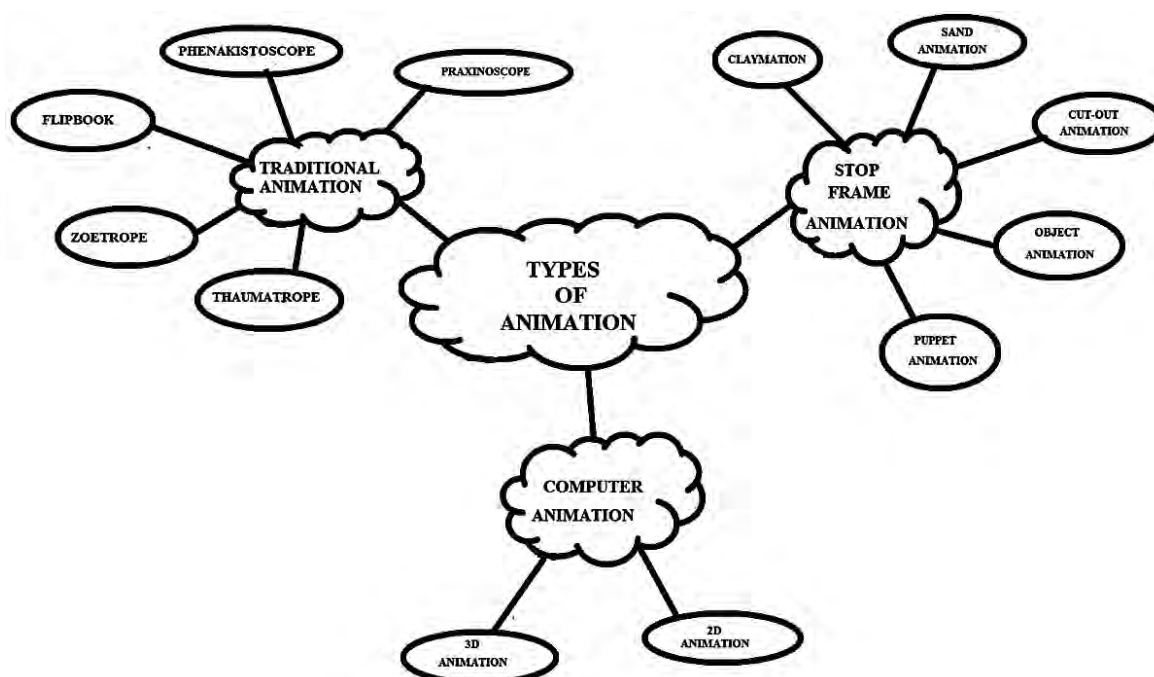
Animation's origin can be tracked right from 1824 with the wide usage of thaumatrope that was largely given credit to John Aryton of Paris. Since then the field of animation had experienced various meaningful adaptations and thereby can be classified into three wide verticals namely (i) Traditional animation, (ii) Stop frame animation and (iii) computer animation (2D and 3D). Traditional animation includes Thaumatrope, Phenakistoscope, Zoetrope, Flip book, and Praxinoscope.

Thaumatrope

In early 19th century, Thaumatrope was very popular. Etymologically Thaumatrope has got Greek origin Thagma meaning marvel and tropos meaning turning. Two pieces of string with two different pictures on each side, like as a tiger and a cage, is attached to a small disk. Twirling of strings quickly between the fingers seems to combine into a single image. This is due to persistence of vision, the physiological phenomenon that ensures imagery rests in the eyes and brain for a small fraction of a second even after the vision is blocked or the object is removed.

DOI: 10.4018/978-1-5225-2255-3.ch227

Figure 1.



Phenakistoscope

Phenakistoscope was the earliest animating device invented simultaneously by Joseph Plateau of Belgium and Simon Von Stampfer of Austria in the year 1831. In Greek Phenakezein means to deceive or to cheat. It composed of a disk with series of images pasted or drawn on radii of the disk at different distance from the centre. The Phenakistoscope would be placed in front of a mirror and rotated. While the phenakistoscope rotates, a viewer can look through the slits at the reflection of the drawings that are visible like a flash when a slot passes by the viewer's eye and ensures the illusion of animation.

Zootrope

William George Horner in 1834 suggested the concept of Zootrope. Etymologically Zootrope has got Greek origin Zeo meaning life and tropos meaning turning. Zootrope literally means "wheel of life". The principle is same that of phenakistoscope. It is a cylindrical rotating instrument with

many frames of images printed on a paper strip kept over the inner circumference. The observer watches through vertical slits around the sides and view the moving images on the opposite side as the cylinder rotates. As it rotates the frames between the viewing slits moves in the opposite direction of the picture on the other side and serves as a micro shutter. It does not require the use of a mirror to view the illusion, and as it has cylindrical shape it can be viewed by many at a time.

Flip Book

John Barnes Linnet, introduced the first flip book in 1868 and named it as Kineograph (moving picture). It was a book which had a series of images, when flipped shows that the drawings are moving. It is the simplest way of making animation without a camera.

Praxinoscope

In Greek praxein means action and scope means watcher. Praxinoscope had got two variants namely

non-projective and projective. Non-projective praxinoscope is the successor to zoetrope invented in 1877 by a French science teacher, Charles-Emile Reynaud. The narrow slit in the zoetrope was replaced by an inner radius of mirrors, kept inside so that the reflections of the images seemed to be more or less stationary in position as the wheel is rotated. Non-projected praxinoscope gives images that are less distorted and brighter than the zoetrope gives. In 1889 Reynaud introduced an improvised version which could project pictures on a screen from a longer roll of images. This projected praxinoscope was used for the benefit of larger audience.

Stop Frame Animation

During late nineteenth century stop frame animation had got its lime light. Stop Frame Animation is a film making technique to make a physically manipulated object appears to move on its own, this is by using individual camera shots of a different movement. Putting these shots together results in an animation impact. There are many types of stop frame animation namely clay animation, sand animation, cut-out animation, object animation, puppet animation, graphic animation and so on. Even though it is a time consuming process it leaves an impact on learning in a positive way and enhance look and feel.

Clay Animation

Clay animation is one of the many forms of stop frame animation invented in the year 1897 in which each of animated piece is made of a malleable substance, usually plasticine a type of clay. In claymation (short for clay animation) each object is made of clay or pliable material usually surrounded with skeleton wire called an armature. An armature can be built to support the clay figure using wire, tin foil, Styrofoam balls and Popsicle sticks. The series of capture with malleable figures ensures the impact of animation.

Sand Animation

Sand animation, otherwise called as sand art, in which an artist forms a series of images with sand. At first an artist applies sand to a surface like glass and then forms images by drawing lines and figures in that sand with his hands. In this sand animation overhead projector or light box is used as an aid by the performer. Then the animator moves the sand around on a back lighted or front lighted piece of glass to create each frame and on sequential grab the effect of animation is bestowed over sand images.

Cut-Out Animation

Cut-out animation is one of the stop frame animation techniques using flat characters, props and background cut from materials such as paper, card, stiff fabric or even photographs. The characters are fragmented into smaller pieces and captured the pieces by giving small movements manually, at each step, to create the illusion of movement. The cut-outs may be of paper drawings, photographs or any 3D objects. The cut-out animation process is a tedious one as it needs manual intervention that is the animator has to adjust the pieces for every frame.

Object Animation

The animation created with the use of any non-drawn objects like toys, blocks, dolls, etc. that are not fully malleable, such as clay or wax and not formed to look like a living or non-living character.

Puppet Animation

The puppets or toys which are flexible to move their parts, so that they can be repositioned between frames to create the illusion of motion are used in this variant. When the frames are played in rapid sequence the puppets are getting life. From the ancient times puppets are used for story

telling in ceremonies and later it came to the form of animation.

Computer Animation

The animator uses software to draw models and objects. The animator can animate irrespective of physical laws like gravity, mass and force. Early digital computer animation may be the brain child of Bell Telephone laboratories during 1960's. Lawrence Livermore of National laboratory had also indulged in developing early digital computer animation. There are two types of computer animation 2D animation and 3D animation. During 1976 the concept of 3D animation caught up and development of Computer Generated Imagery (CGI) technologies had given life to 3D games and game engines.

2D Animation

The 2D animation images are created or edited using 2D bitmap graphics or 2D vector graphics. Analogue computer animation, power point animation and flash animation are some of the applications of 2D animation.

3D Animation

In 1972, Frederic Parker created the first 3D human face model. In 1973, Edwin Catmull one of the founders of Pixar Animation studio created digitized hands. 3D animation is digitally modelled and manipulated by an animator. 3D animation is carried out through step by step process.

As in any scientific venture 3D animation also requires lots of planning. A creative artist with unquenchable imagination prepares the digital space. The 3D animation software packages like Maya and 3Ds Max exploit the use of node system. Node system is interesting, flexible and robust at the same time it is challenging when it comes to execute command. The software keeps in mind every command that is given in a linear fashion and calculates how things would look like with

the first node and keeps in track and moves in to subsequent frames. First the digital artist creates image planes that will describe the exact objects in different axes and integrates to create the three dimensional model that is made up of wire frame structure. The Euclidean geometry model depicts X axis as horizontal, Y as vertical and Z as depth, that involves lot of mathematical equation and algorithms. In this stage in Maya the artist can view the wire frame structure by pressing 4 or he can get the other rendering options like smooth shaded, smooth shaded with hardware texturing and smooth shaded with lights.

After finalising the structural part of the model the surface is textured as per the need of the animation theme. The appropriate light arrangement is provided on the objects and rendering is the part that ensures improvisation every time.

Once the visuals are finalised the story board developed by the director guides the animation engineer to what extent the objects stay and move in the screen through timing schedule. The animator provides the commands to the objects to make required moments followed by rendering and editing. At last the voice over takes place with music or description addition and the 3D animation is finalised through analysis by the stake holders. Thus making 3D animation is the cyclic process that offers improvisation always.

BACKGROUND

Teaching of Science Through Animation

It is needless to underline the importance of Science in the digital era. The accumulation of new scientific inventions are taking place in one hand, the dynamic principles of curriculum always push forward the developments at the earliest to the curriculum on the other hand. And it is evident that many school teachers teach in schools what they studied in their college that stands testimony to the rapid developments that are taking place in the

ambit of science teaching. It is the paramount task before the science teacher to explain the increasingly abstract topics in a jammy academic year full of hectic activities. This situation warrants the resetting of the compass of work ethics for science teachers. In this pressure cooker situation the search for better instructional media that will ensure greater individualisation and better illustration becomes inevitable. The present generation kids are glued to television animation and keep spending more time in animated games even. It is not pathological but has become the symptom of normalcy in the society where both the parents are employed and children left with only virtual world. Many experimental studies in this area give some meaningful findings for the usage of animation to teach science. Some topics like skeletal structure, illustration of crystal lattices, molecular labelling of chemical kinetics, internal structure and function of many internal organs like eyes, kidneys, heart and so on can be better explained with animation and aid the learners to take it to their long term memory and better achievement.

Psycho-Pedagogy Behind Animation

The effective use of technology for optimum realisation of educational objectives always seems to be the task on the table for majority of educational researchers. The advances in cognitive psychology provide convincing theoretical backdrop to use multimedia and web based technologies to improve learner participation, enhance learning outcomes and to extend active retention with better memory traces. As per Mayer (1991) through multimedia learning learners indulge in three important cognitive activities like selecting, organising and integrating. By these activities the learners build connections between verbal based models and visual based models with enriched prompts and cues.

It is the assumption that animation provides separate systems for processing pictorial and verbal materials. In this dual channel assumption each channel has got limited amount of material that

can be processed at a time. Meaningful learning is ensured only when these mental connections between pictorial and verbal representations are complete. The alternate school of thought which highlights the cognitive overload is diminishing day by day. As per these theorists the learners intended cognitive processing exceeds the available cognitive capacity of the learners. This argument is losing its ground as more and more scientifically designed and pedagogically superior educational animations becoming runaway success.

Julie Bauer (2002) is of opinion that graphics have been used from the past to portray things that are inherently spatio-visual. Concepts that are metaphorically spatio-visual can be better explained with moving graphics than static graphics when it conforms to congruence principle of learning. For real time reorientation of time and space animation is the tool that stands taller in spatio-visual ambience. Multiple representation principle, contiguity principle, split attention principle, individual differences principle and coherence principle if applied judiciously in educational animation tastier will be the fruits of education.

Need and Significance of the Study

The classroom trend changes from manual to technological and therefore the conventional way of teaching is not enough for the children. Today the children need active participation in the classroom. They need their teachers to be a motivator, active participant etc., since the availability of many software make the teachers feel free and easy to design their own style. Even though it is the additional workload for the teachers the learning outcome or evaluation gives some satisfactory improvement (Sajid Musa 2015).

The main intension of animation class is not only in creating curiosity but also helps to sustain their interest and attain curricular objectives. The concepts which they learnt through animation withstand the retroactive inhibition of memory which makes their studies a productive one for their future aspects.

The animation class not only makes the children to learn, but it makes them to understand. It creates some positive attitude among the learners to pursue education with ease (Falode 2016). This study is the need of the hour to assess whether animation has got positive educational outcomes for teaching and learning communities.

The effectiveness of an educational animation largely depends on the soundness of its psychological constructs. Animation in education motivates the learner in an interesting way in which traditional method of "Chalk & Talk" is avoided and also minimises actual classroom demonstration & lecture. While using educational animations not only creativity & novelty is increased among learners monotony is also reduced. Animation creates liveliness in the classroom situation and prolongs retention, it keeps the learners attentive and thereby learners can easily understand the concepts of the subject easily when compared to traditional method of learning. Many of the complex ideas become easier while using animation and also matches the cognitive demands of a learning task. (Tversky et al. 2002)

Animation has the potential to attract the concentration of learners in their topics through visualization of content. It's an enjoyable and memorable one where animation leaves a lasting impression and form a stronger relationship, making them more likely to respond. Watching a video is easier than reading a lengthy document or book; it not only saves the time but also makes sure your message get across. Present both auditory and visual stimulus to the students, allowing easier cognitive tagging and encoding of information besides triggering the imagination of learners.

Limitations

The educational effectiveness of any visuals largely depends on how widely it is used by the instructor rather than how well learners learn when animation is used. So it lies with the instructor to maximize the utility of animated media to supplement the face to face instruction. Many a

times out of vested interest or technical inertia the instructor in developing countries are miles away to make up their mind to go for animation.

Mostly real life learning is lost in animation. Students are fostered with virtual ecology. The individual differences prevailing among the learners may not be effectively addressed with animations that are targeting generalised perspectives.

Review of Related Literature

Review of related literature provides innumerable studies in this area, for brevity a few studies are highlighted hereunder followed by Meta synthesis.

Angelos Konstantinidis and others (2013) introduced four Web 2.0 tools; Blogger, Strip Generator, Go! Animate, and Google Forms, that are free and easy to use, in an effort to motivate teachers with low technological skills in integrating them into their instruction practices. In their descriptive method they found that the aforementioned tools comprised the curriculum in a blended-learning professional development course for in-service teachers that attracted many favourable comments from the participants.

Lirong Xiao (2013) made a survey through inquiry forms about animation content in education and found that in animation field, although some software companies have developed their individual production toolboxes or platforms for animation content in education, there is lack of relevant research from the perspective of animation techniques.

Sederick C. Rice (2013) used interactive animations to enhance teaching, learning, and retention of respiration pathway concepts in face-to-face and online high school, undergraduate, and continuing education learning environments. The study found that the content used for the research supports the development of more inquiry-based classroom and distance-learning environments that can be facilitated by teachers/instructors, it also improves retention of important respiration subject content and problem-based learning skills for students.

Gokhan Aksoy (2012) analysed the impacts of animation technique on academic performance of students in the “Human and Environment” unit present in the Science and Technology course of the seventh grade in primary education with 58 students. The study revealed that animation technique is very effective when comparing with traditional teaching methods in terms of improving students’ achievement.

Waqur Un Nisha Faizi and others (2012) identified the main causes of increasing attention towards animation course in different institutions of Karachi through survey method. The researchers found that majority of the respondents had a view that animation course ready to meet the need and requirement of modern times. Some respondents opined that students of animation class lead a happy and prosperous life at the end of the course and animation course grooms the ability of decision making among students.

Animesh K. Mohapatra (2013) probed the extent of computer animations in teaching membrane transport to pre-service teacher trainees’ and the extent of understanding of concepts and functions in membrane transport. Many a times educators find it difficult to teach various membrane transport processes, as many of the resources are two dimensional whereas membrane transport is four dimensional in nature. Much research finding support that viewing the processes in three dimensions helps better learning, and animations are perhaps effective visualization tools for learners and help to retain for long-term. It was concluded that animations with movements can provide learners with explicit dynamic information that is either implicit or unavailable in static graphics. Therefore, it was recommended to use computer animation to transform students to active mode.

As it had been observed that this particular field of study has got immense potential and the available literature also has got spaces to fill with this study on “Science Animation and Students Attitude – An Analytical study” has been taken up.

METHODOLOGY

Assumption of the Study

Gaining of knowledge through learning is a day-to-day process which we experienced from our childhood. All the people don’t have an opportunity to engage in learning through schools or colleges throughout their life. Apart from the educational institution students learn a lot from their environment which are unorganized and perhaps unauthentic. The systematic method of constructing scientific and verified knowledge is received only from the schools, so one should be aware of what they learnt and it should be useful to their life. The purpose of education is to make a learner to be a learner for their life time that is they should not be filled with bookish content alone and they must know how to convert the content to their practical utility.

The engaged time of the institution must be better transformed into academic learning time. The relatively small portion of time is only going to create a large impact in the cognitive structure of an individual. The tools those are useful to make better memory traces with the learner can be complete only with animation.

Thus the researcher by this assumption made the following research questions.

1. Do animated lessons trigger the learners’ interest?
2. Does animation find useful place in curriculum?
3. Do animated lessons increase the span of attention?
4. Does animation make learning a joyous one?
5. Do animated lessons ensure better understanding of the content?
6. Can students achieve better with animated lessons?

Objectives of the Study

The following are the objectives

1. To measure the span of attention of the students studying through animated lessons.
2. To find out the curiosity in the learners towards animation lessons.
3. To analyse the difference in achievement between conventional class and class with animated lessons.
4. To evaluate the level of understanding of students attending the class with animated lessons.
5. To evaluate the relationship between achievement and attitude of students attending the class with animated lessons.
6. To find out the relationship between the attitude and its sub variables towards animation.
7. To evaluate the effectiveness of the class with animated lessons in science subject.

Hypotheses of the Study

The following hypotheses were formulated based on the objectives of the study.

1. There is no significant difference between control and experimental group in their pre-test achievement.
2. There is no significant difference between control and experimental group in their post-test achievement.
3. There is no significant correlation between experimental group post test achievement and students attitude towards animation.
4. There is no significant correlation between experimental group post test achievement and the sub variables of attitudes towards animation like narration, visual transformation, interest & attention, easy and rapid understanding and Participatory evaluation.
5. There is no significant correlation between experimental group students' attitude and the sub variables of attitude towards animation

like narration, visual transformation, interest & attention, easy and rapid understanding and Participatory evaluation.

6. There is no significant cross correlation between sub variables of attitude towards animation like narration, visual transformation, interest & attention, easy and rapid understanding and Participatory evaluation.

Research Design

To study the effectiveness of teaching through animation, the pre-test, treatment, post-test equivalent group experimental design was adopted in the study. A sample of 43 students for each control and experimental group were selected by purposive random sampling based on their academic term achievement. The equivalence of mean achievement was taken into consideration to finalize the sample of both groups. The following tools were used in the study.

Tool 1: Achievement test in science constructed & validated by the investigators.

Tool 2: Scale of attitude towards animation lessons constructed & validated by the investigators.

Research Procedure

The population may be school students studying science as one among their subjects, the representative sample was selected from a private school in Kumbakonam of Thanjavur District for both control & experimental groups. The students studying in eighth standard (280) were screened through their achievement in term test. Among the 280 students the investigator adopted purposive sampling technique to select the students who fall in the average category (50% - 60%) that accounted 86 students based on their term test marks. 86 students were randomized and split into 43 each to form the control and experimental groups for the study.

The pre test was administered to both the experimental and control groups and data were

collected to assess the achievement of the students. The topics were selected from science textbook and were taught to the control group with conventional method. For the experimental group the same topics were taught with animated lessons for 30 days. The animated lessons are mostly 2D and 3D animations those are freely available to the users. Both control and experimental group were instructed by the same investigator. For the execution of animated science lesson the investigators had selected animated videos from selected websites that provide free feed. After the treatment period the post test was administered for the sample and data were collected for achievement test. Scale of attitude towards animation lessons was administered on experimental group after post test to assess their attitude about the treatment.

DESCRIPTION OF RESEARCH TOOLS

Construction of Achievement Test

The achievement test was constructed by the investigators from the topics which they planned to teach the students using animation. Investigators reviewed the Tamilnadu State Board Textbook for standard eight and finalized the topic pertaining to second term science and settled with two topics namely skeletal system and atomic structure. The syllabi and their finalization were carried out with the help of research supervisor, four experienced teacher educator and Principal of the school, Module 1 consisting animated video of human body and its movements, joints and types of joints, skeleton and movement of animals. Like wise module 2 consisting law of conservation of mass, law of definite proportion, Dalton's atomic theory, discovery of electron, properties of cathode rays, discovery of protons and Thomson's atomic model.

Achievement Test

Standardization

The test items constructed were subjected to jury's opinion consisting of research supervisor, principal of the school and four experienced teacher educators. To establish the reliability of the test, the investigator adopted split-half method. The reliability of the whole test is found to be 0.86 significant at 0.01 levels. Hence it is concluded that the test is highly reliable. To establish the validity of the test the investigator attempted to find out the correlation co-efficient between the achievement scores in the test and scores they got in their term test by product moment correlation co-efficient method by taking 25 percentage of the total sample of 86 students. The correlation co-efficient is found to be 0.831 significant at 0.01 level. Hence it is concluded that the test has high validity. The content validity of the test was ensured by the panel that acts as jury. The answer books of the students for achievement in science were scored by giving one mark each for a right response of the objective type question, thus a range of 0-30 marks can be secured by the students.

Scale of Attitude Towards Animation Lessons

Attitude may be considered as the state of mind or feeling about anything or anybody. It is essential to estimate the attitude of the learners towards animation lessons and what really matters for them to have animation as the preferred mode of educational delivery. So the investigators would like to know the attitude of the students present in the experimental group towards animation lessons.

The investigators after having meticulous analysis and critical review with various stakeholders had extracted five important dimensions that compose the attitude of the students towards animation lessons as following

1. Narration
2. Visual transformation
3. Interest and attention
4. Easy and rapid understanding
5. Participatory evaluation

As much as 78 questions were developed in these dimensions. These questions were subjected to item analysis and expert regulations. The final version rested with 30 questions spread across all dimensions; the significance of each dimension is given below.

Narration

Narration is the use of a selected methodology or process of giving a written or spoken commentary to unfold a story to an audience. Narration embraces a set of approaches through which the author of the story presents it. Narrative mode includes Narrative point of view: The perspective through which a story is communicated and Narrative voice: The format through which a story is communicated.

In animated lessons the narration or the script board spells success to a greater extent. By narration the investigators mean the way of unfolding the scientific concepts by employing the technological pedagogical principles of learning and also child friendly elements like animated characters that may explain the lessons as video jockey with some sense of humour.

Visual Transformation

The visual elements that are rich in details and better in dimension may offer a meaningful understanding of the point of discussion. By visual transformation the investigators mean the change of frames that may offer more concrete idea about abstract concepts by means of taking the learners to the quint essential part of the topic.

Interest and Attention

Interest may be defined as the feeling one may have to know or to learn more about somebody or something. It may also be defined as an activity or a subject that one may enjoy and spend the free time in doing or studying it.

Attention is the act of carefully listening or thinking about something. By interest and attention the researchers mean a connection of the learner with animated lessons which affects his attitude towards it especially while he receives immense benefit and urge to attend the classes.

Easy and Rapid Understanding

The knowledge that someone has about a particular situation or subject may be termed as understanding. To gain that knowledge in a given topic if the learner is going to take less cognitive effort that ensures easy understanding. If the time that is going to be consumed is less by the learner to understand that may be considered as rapid understanding.

Through this dimension the researchers would like to seek whether animation lessons ensure easy and rapid understanding from the learner's perspectives.

Participatory Evaluation

Participatory evaluation is the partnership approach to evaluation in which the stakeholders actively engage in developing the evaluation and all phases of its implementation. The stakeholders are instructors, learners, decision makers and any other beneficiaries. The process includes identifying relevant questions, planning the evaluation design, selecting appropriate measures for collecting and analysing data (Ann zukoski and Mia Luluquisen).

Table 1.

Group	Sample	Mean	Std. Deviation	t	P value
Post-test (control group)	43	9.419	3.905	21.445	0.000
Post-test (experimental group)	43	25.326	3.037		

In this dimension the researcher would like to estimate the readiness of the learners to voluntarily ensure the progress by answering formative questions available in the particular topic with innate enthusiasm and self regulating their zeal to answer.

Statistical Techniques

Mean, Standard deviation, 't' test, Correlation, Regression, Effect size, Cohen's d and Gain score analysis are the statistical techniques that were employed for data analysis.

ANALYSIS AND INTERPRETATION OF DATA

Testing the Significance of Difference Between the Mean Scores of Pre-Test

The table value of 't' for the degree of freedom with 0.01 level of significance is 2.64, whereas

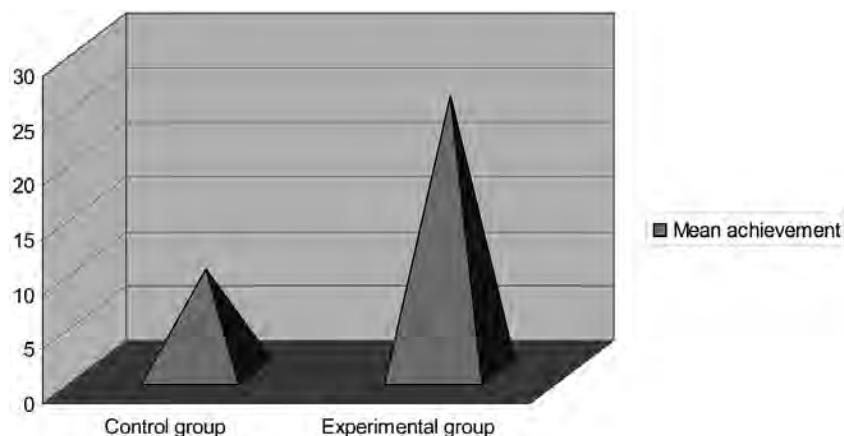
the calculated 't' value is 1.699. This makes it obligatory to accept the null hypothesis and concluded that there is no significant difference between the experimental group and control group students in their pre test achievement in science.

Testing the Significance of Difference Between the Mean Scores of Post-Test

The significance of difference between the mean of control group and experimental group of eighth standard students in their post-test achievement in science is shown in Table 1.

The calculated value of 't' 21.445 is significant at 0.01 level. This makes it mandatory to reject the null hypothesis and concluded that there is significant difference between the experimental group and control group students in their achievement in science in which the experimental group has got higher scores.

Figure 2.



Effect Size Analysis

Effect size analysis was carried out using pooled standard deviation to find out the relative effectiveness of the animated science lessons over the conventional method. The instructional objectives were same for the pre-test and post-test. The effect size for post test achievement is 0.915 that is large effect. Likewise the value of Cohen's d for the comparison of control and experimental group post test achievement in science is 4.543. These findings again prove the fact that students studied through animated lessons have fared better and the treatment is very effective over conventional method for the chosen topics. The average attention span observed by the investigator for experimental group was thirty five out of forty five minutes where as for the control group it was only twenty one minutes.

Gain Score Analysis

The gain score analysis of control and experimental post test achievement in Science revealed that the experimental group has got mean achievement of 25.33 in contrast to 9.42 of control group that is 169% is the gain score of the experimental group

post-test achievement. This substantiates that the treatment with animated lessons seems to be very effective over conventional method of teaching.

Correlation Analysis

There is significant positive correlation between experimental group post test achievement and students attitude towards animation with 0.01 level of significance. This gives way to reject the null hypothesis and concluded that achievement in Science and attitude towards animation lessons with experimental group have meaningful correlation that may be attributed to animation lessons.

Testing the correlation between experimental group post test achievement and the sub variables of attitudes towards animation like narration, visual transformation, interest & attention, easy and rapid understanding and Participatory evaluation

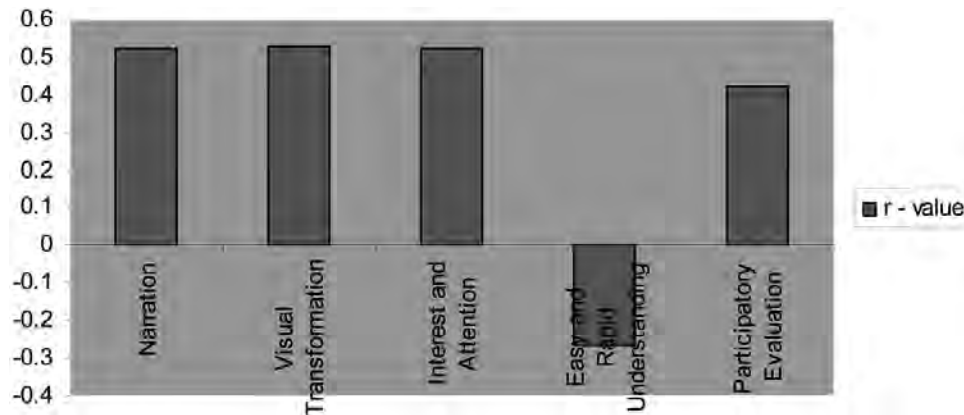
The correlation between experimental group post test achievement and the sub variables of attitudes towards animation like narration, visual transformation, interest & attention, easy and rapid understanding and Participatory evaluation is shown in Table 2.

Table 2.

Variables	N	r
Experimental group post-test vs. Narration	43	0.522**
Experimental group post-test vs. Visual Transformation	43	0.530**
Experimental group post-test vs. Interest and Attention	43	0.522**
Experimental group post-test vs. Easy and Rapid understanding	43	-0.268**
Experimental group post-test vs. Participatory Evaluation	43	0.425**

** Significant at 0.01 level

Figure 3.



There is significant positive correlation between experimental group post test achievement and the sub variables of attitudes towards animation like narration, visual transformation, interest & attention and Participatory evaluation except easy and rapid understanding and makes it mandatory to reject the null hypothesis except for the sub variable Easy and Rapid Understanding.

Testing the correlation between experimental group students' attitude and the sub variables of attitude towards animation like narration, visual transformation, interest & attention, easy and rapid understanding and Participatory evaluation

There is significant positive correlation between experimental group students' attitude and the sub variables of attitude towards animation like narration, visual transformation, interest & attention, easy and rapid understanding and Participatory evaluation except Easy and Rapid Understanding and thereby it is obligatory to reject the null hypothesis.

Testing the cross correlation between sub variables of attitude towards animation like narration, visual transformation, interest & attention, easy and rapid understanding and Participatory evaluation

The cross correlation between sub variables of attitude towards animation like narration, visual transformation, interest & attention, easy and rapid understanding and Participatory evaluation is shown in the Table 4.

There is significant positive cross correlation between sub variables of attitude towards animation like narration, visual transformation, interest & attention, easy and rapid understanding and Participatory Evaluation. This gives way to reject the above said null hypothesis for all the sub variables except Easy and Rapid Understanding.

Regression Analysis

The scores of students' attitude towards animation and its sub variables with respect to achievement

Table 3.

	Narration	Visual Transformation	Interest & Attention	Easy & Rapid Understanding	Participatory Evaluation
Attitude	0.733**	0.949**	0.922**	-0.764**	0.725**

Table 4.

Sub variables	Narration	Visual Transformation	Interest & Attention	Easy & Rapid Understanding	Participation Evaluation
Narration	1	0.849**	0.897**	-0.615**	0.425**
Visual Transformation	0.849**	1	0.988**	-0.856**	0.573**
Interest & Attention	0.897**	0.988**	1	-0.838**	0.530**
Easy & Rapid Understanding	-0.615**	-0.856**	-0.838**	1	-0.519**
Participation Evaluation	0.425**	0.573**	0.573**	-0.519**	1

in science are subjected to regression analysis that yielded the following interpretation.

Predictors: (Constant)- Post Test Achievement of Experimental Group in Science

The method used in this analysis is enter method. The dependent variable is achievement in science and the variables entered are attitude towards animation and its sub variables. The adjusted R square value is 0.364. This indicates that the independent variable in this model account for 36.4% variance in dependent variable that is achievement in science. Both the constant and few of the sub variables has got significance level of 0.01 – 0.05. The β coefficient is significant at 0.05 level. This indicates that the achievement in science, attitude towards animation and its sub variables are positively related except the sub variable easy and rapid understanding.

Findings

1. It is found that there is significant difference between the experimental group and control group students in their achievement in science in which the experimental group that

learnt with animated lessons has got higher scores.

2. It is found that there is significant positive correlation between experimental group post test achievement and student's attitude towards animated lessons which proves that instruction with animated lessons seems instrumental in building a positive attitude towards animation lessons among the majority of the students.
3. It is found that there is significant positive correlation between experimental group post test achievement and the sub variables of attitudes towards animation like narration, visual transformation, interest & attention and Participatory evaluation except easy and rapid understanding. It gives an inference that people who have positive attitude towards narration, visual transformation, interest & attention and Participatory evaluation that may be created out of animated classes had scored better in their achievement test in science.
4. It is found that there is significant positive correlation between experimental group students' attitude and the sub variables of attitude towards animation like narration, visual transformation, interest & atten-

tion, easy and rapid understanding and Participatory evaluation except Easy and Rapid Understanding that gives an insight that most of the sub variables has got right orientation with attitude towards animation and perhaps Easy and Rapid understanding had not been preferred to the extent it had been preferred that the top scorers whereas it seems the poor scorers has preferred this dimension as essential.

5. It is found that there is significant positive cross correlation between sub variables of attitude towards animation like narration, visual transformation, interest & attention and Participatory evaluation except easy and rapid understanding.

Educational Implications of the Study

This study reveals that science instruction with animated lessons is very effective in improving the achievement of students in school science. The overall outcome of the study highlights the effectiveness of teaching with animated lessons with large effect size and gain score in achievement of Science. The attitude of the students towards animation classes is very positive. This study implies that teaching with animated lessons will definitely enhance the outcome of teaching learning process in the following ways,

1. Animated lessons make the learners to be attentive and focused.
2. Animated media offers in-depth details to make the students thoroughly understand the abstract elements present in a particular topic.
3. Visual transformation and Narrative paradigms that are employed in animated lessons seems to be psycho pedagogical in nature to sustain the interest of the learners and to complete the syllabus in time.
4. The effort required by teacher and learner alike can be minimised with animated classes.

5. As animated lessons are self explanatory and informational through multiple media, the impact on memory traces would be long lasting.

FURTHER RESEARCH DIRECTIONS

This study entitled “Science Animation and Students Attitude-An Analytical Study” is an investigation at standard Eighth level. It is suggested that further studies may be conducted in the following areas,

1. It is suggested that teaching through animation is evaluated in other school subjects like Mathematics, Social studies, Language I and II etc.
2. It is suggested that the same study is carried out in all standards of different boards.
3. It is suggested that the same study is carried out in higher education and professional education courses.
4. It is suggested that the influence of other variables than those studied now can also be investigated.
5. It is suggested that efforts should be taken to develop animated lesson for more topics and evaluate their effectiveness.

CONCLUSION

Through animated lessons science can be taught effectively. More and more new information with the support of modern electronic media can be passed to the learners easily. In the information era ocean of information cannot replace the teacher and his personal influence in the mind of the learner. In this method the teacher is having face to face interaction and gives guidance about the veracity of the content to ensure the right learning to take place through self motivated learners. Though the groups in the study are homogenous, the differences in mean score and standard deviation will stand testimony to the fact that, teaching

science through animated lessons is superior to conventional method of teaching. There seems to be no doubt that auto instructional strategies are supplementary as well as main process of learning. Teaching is replaced by learning, ultimately leading to self learning thus teaching science subject through animation fill the gaps of the conventional teaching learning process.

REFERENCES

- Aksoy, G. (2012). The Effects of Animation Technique on the Seventh Grade Science and Technology Course. *Scientific Research.*, 3(3), 304–308.
- Aksoy, G. (2013). Effect of Computer Animation Technique on Students Comprehension. *Mevlana International Journal of Education.*, 3(1), 40–46. doi:10.13054/mije.13.02.3.1
- Ali, A. Z. M., & Madar, A. R. (2010). Effects of Segmentation of Instructional Animation in Facilitating Learning. *Journal of Technological Education and Training.*, 2(2), 15–29.
- Animesh, K. M. (2013). Fostering Pre service Teacher Trainees' Understanding of Membrane Transport with Interactive Computer Animation. *Scientific Research*, 4(10), 640–645.
- Everything About Animation. (n.d.). *Stop motion*. Retrieved from: <http://animationmovies.blogspot.in/p/stop-motion.html>
- Faizi, W. U. N., Shakil, A. F., & Kanwal, S. (2012). The Causes of Increasing Attention towards Animation Course in Different Institutes of Karachi, Pakistan. *International Journal of Academic Research in Business and Social Sciences.*, 2(7), 444–457.
- Falode, O. C. et al.. (2016). Effectiveness of Computer Animation Instructional Package on Academic Achievement of Senior Secondary School Agricultural Science Students in Animal Physiology in Minna, Nigeria. *Bulgarian Journal of Science and Education Policy.*, 10(1), 5–8.
- Gaur, A. S., & Gaur, S. S. (2009). *Statistical Method for Practice and Research*. New Delhi: SAGE Publications India Pvt. Ltd.
- Huang, H.-W., & McConnell, R. (2009). Students Experiences of Technology-Enhanced learning in Two Traditional Teacher Preparation Classrooms. *Journal of Online Learning and Teaching*, 5(3), 522.
- Konstantinidis, A. (2013). Web 2.0 Tools for Supporting Teaching. Turkish online. *Journal of Distance Education*, 14(4), 287–295.
- Mayer, R. E., & Anderson, R. B. (1991). Animations Need Narrations: An Experimental Test of Dual-coding Hypothesis. *Journal of Educational Psychology*, 83(4), 484–490.
- Mohanasundaram, K., & Sivasankar, A. (2010). Blended Learning: A New Horizon. *University News*, 48(3), 1–4.
- Mohanasundaram, K., & Sivasankar, A. (2010). Effective Blends and Economical Blogs: Successful Change Agents of Higher Education to Meet New Challenges. *University News*, 48(28), 86–90.
- Reporter, B. S. (2008). Animation industry to grow at 27% CAGR: NASSCOM. *Business Standard*. Retrieved from: http://business-standard.com/article/technology/animation-industry-to-grow-at-27-cagr-nasscom-108110600063_1.html
- Rice, S.C. (2013). Using Interactive Animation to Enhance Teaching Learning and Retention of Respiration Pathway Concepts in Face to Face and Online High School, Undergraduate and Continuing Education Learning Environments. *Journal of Microbiology & Biology Education*, 14(1), 113 – 115.
- Swarmy, A. M. A. (2010). Internet Awareness and Competence Among High School Students and Teachers. *Edutracks*, 9(7).
- Tversky, B., Morrison, J. B., & Betrancourt, M. (2002). Animations: Can it facilitate? *International Journal of Human-Computer Studies*, 57(4), 247–262. doi:10.1006/ijhc.2002.1017

Utts, J. (2003). A Study Comparing Traditional and Hybrid Internet-Based Instruction in Introductory Statistics Classes. *Journal of Statistics Education*, 11(3).

Wikipedia. (n.d.). *History of animation*. Retrieved from: http://en.wikipedia.org/wiki/History_of_animation

Wordpress. (n.d.). Retrieved from; http://dmizerny.files.wordpress.com/2012/09/img_0982.jpg

Xiao, L. (2013). Animation Trends in Education. *International Journal of Information and Education Technology*, 3(3), 286 – 289.

Zukoski, A., & Luluquisen, M. (2002). *Community Based Public Health Policy and Practice*. Academic Press.

ADDITIONAL READING

Ainsworth, S. (2008). How do Animations Influence Learning? In D. H. Robinson & G. Schraw (Eds.), *Current Perspectives on cognition, Learning, and Instruction: Recent Innovations in Educational Technology that Facilitate Student Learning* (pp. 37–67). Charlotte, NC: Information Age Publishing.

KEY TERMS AND DEFINITIONS

E

Animation: Latin word ‘animatio’ that means the condition of being alive or giving life.

Attitude: The state of mind about anything or anybody.

Computer Animation: The images or objects that are created using computer software subjected to animation.

Effectiveness of Animation: Remarkable achievement of learners by the impact of animation.

History of Animation: The growth and development took place in the field of animation.

Psycho-Pedagogy: Educating the learners through the principles of psychology.

Science Animation: Animation clippings that are used to explain science concepts.

Visual Transformation: Change of visual frames in any video.

Three Cases of Unconventional Educational Uses of Digital Storytelling

Emmanuel Fokides

University of the Aegean, Greece

INTRODUCTION

The narration is the art of using words and actions for the representation of the elements of a story in such a way that the listener's imagination is stimulated (Genette, 1998). More simply put, it is the art of telling a story to an audience, in order to convey important messages. Due to the technological developments, storytelling has become digital; the oral or written story is enhanced by using multimedia and hypermedia elements (Lathem, 2005). Narrations, either digital or conventional, are useful educational tools. Since narrations cause the keen interest of students, this helps them to easily consolidate and assimilate information (Coventry, 2008). They increase the oral and written skills, strengthen critical thinking and the ability to analyze and synthesize information (Ohler, 2006). When students create their own digital stories (individually or as a group), they learn to conduct research on a topic, to ask questions, to organize ideas, to express their views and to make meaningful narratives (Robin, 2006).

There is an extensive literature regarding the educational benefits when using digital storytelling (e.g., Coventry, 2008; Ohler, 2006; Robin, 2008). Disproportionally fewer studies have been conducted examining the potential of this tool in areas where the settings are not strictly instructional or the main objective is not some form of knowledge acquisition. The present study is an attempt to fill that gap, by embracing the standpoint that digital storytelling is a good method for documenting personal experiences, that it can be a form of narrative therapy and that it can help students to discover parts of their

personality (Sawyer & Willis, 2011). Three case studies are presented where digital storytelling was used in a non-mainstream, unconventional way. In all, knowledge acquisition was irrelevant or an insignificant factor; instead, the emphasis was on broader issues that students, as well as teachers, face at school.

BACKGROUND

Focusing on problems that students and teachers face at school, which are not directly related to knowledge acquisition, but affect how the school functions and/or the emotional well-being of students, three areas were of interest: the poor school integration of immigrant students, young students' adjustment to school, and bullying.

In Greece, 10.35% of the total students' population are immigrants (Hellenic Statistical Authority, 2011). Insufficient knowledge of the Greek language and, consequently, low performance in language lessons is a major problem (Retali, 2013). There is also a more important difficulty; that of poor school and/or social integration. Schools could play an important role, but the Greek educational system is not capable of assimilating immigrant students well (Skourtou, Vratsalis, & Govaris, 2004). Therefore, there is a need to help them overcome their adaptation problems.

Coming to primary school for the very first time marks the beginning of a transitional period to children's lives. Rules and routines are different from those they were accustomed in the kindergarten and their status and identity might be affected (Fabian, 2007). Problems may arise

that have short and long-term educational and/or psychological implications (Dockett & Perry, 2009). Behavioral problems are also common (Brooker, 2008). Consequently, finding ways that allow a smooth and quick transition to the school's environment are quite important.

Bullying is a phenomenon that is becoming more and more frequent in Greece's schools. It greatly affects students' psycho-emotional development, their school performance (Manesis & Lambropoulou, 2014) and may lead to serious psychological trauma and dysfunctional social behavior (Galanaki & Vogiatzoglou, 2015). In hypothetical situations children easily express their intentions to help the victim or to report the incident (Rigby & Johnson, 2006), but in reality, only a small percentage actually acts (Salmivalli, Lagerspetz, Bjorkqvist, Osterman, & Kaukiainen, 1996). Studies, in Greece, examining interventions where students actively participated and were not just passive receivers of information are sparse (e.g., Kyriazis & Zacharias, 2015). Thus, there is a need to inform them in a more comprehensive way.

The coming sections present how these issues were dealt using digital storytelling.

MAIN FOCUS OF THE ARTICLE

Case 1: School Integration of an Immigrant Student

In order to help a sixth-grade female immigrant student, having significant adaptation problems, a short project was planned and carried out, at a primary school in Rhodes, Greece, from late October 2014, till mid-March 2015. Twenty students (including the subject) were involved. The main idea of the project was to ask her to develop and present to her classmates an autobiographical digital story, illustrating her thoughts and feelings from her transition from one country to another.

She was selected as the study's subject because: (a) she recently came to Greece from the Domini-

can Republic. There were no other immigrants from that country; therefore, no one could help her and her family during their transitional period, (b) she did not socialize with the other students, whose attitudes toward her were "indifferent", (c) she should be attending high-school, but because of her poor school performance she had to repeat the primary school's last grade, and (d) her adaptation problems had worsened because of (c). From the above, it can be argued that she reflected characteristics and problems arising from the fact that she was an immigrant, therefore she constituted a critical case (Flyvbjerg, 2006). Data were gathered from multiple sources; interviews, direct observations, drawings, and the subject's digital story.

The interviews with the subject's teacher were about her difficulties in social interaction with her classmates. The interviews with the students focused on how they view and interact with her. The interviews with the subject focused on the difficulties she was facing and the level of social interaction with her classmates. The in-classroom observations were focused on her behavior and the attitudes of the other students toward her.

The development of the story lasted for a month, with a total of nine one-to-two hour sessions. It consisted of three parts: (a) "Before leaving" (seven scenes), where her thoughts, feelings, and conversations with relatives and friends were depicted, (b) "The journey" (two scenes) where her first impressions of her new home were illustrated, and (c) "In Rhodes" (two scenes) where her situation at school was portrayed. Her favorite song when she was living in her homeland accompanied the first part, while in the other two parts she used her favorite Greek song. Even though all dialogues and thoughts were written in Greek, they were "spoken" in Spanish, using her own voice. It has to be noted that during the development of the subject's story, the researcher did not intervene in any way. This was done because guidance regarding the structure or the content of the digital story, might have resulted in the alteration of the results.

Her detachment and loneliness were strongly portrayed at the third part of the story (e.g., all students are playing, but she sits alone and no one is talking to her), in contrast to the first, which is full of dialogues with relatives and friends. This contrast, reveals the extent of the lack of interaction and communication with others and her strong feelings of isolation and solitude. Also, the story was -in some sense- bilingual. Texts were in Greek; the narration was in Spanish. This fact, together with the Spanish song in the story's first part, had an impact on students. Photos, language, and music, combined together, helped them in having a better understanding of the subject's culture, as they stated in their interviews.

In their initial interviews, the majority of students (16) expressed the view that they have no dissimilarities with foreigners, with the exception of language. However, their focus was on the language and communication problems and not on the immigrants' feelings. After the presentation of the story, students stated that they gained a better understanding of the immigrants' problems (17 cases). There was also a shift in their focus; from communication problems to the immigrants' feelings of loneliness and isolation.

From the analysis of the initial subject's interview, the strong attachment to her country became clear. Loneliness, unhappiness, anxiety and fear about whether she will be accepted, were her strongest feelings. Significant changes were noted after the intervention; she felt more accepted and, finally, she had started socializing with other students and she was quite happy about it.

The researcher's observations and the teacher's interview also confirmed the positive change in the subject's behavior and socialization. A positive change to the other students' behavior was also observed, the most important being that students acted. The communication barrier was lifted by both sides. Not only the subject was more open in joining groups of students, but also, they were more open in asking her to join them, in various activities.

Case 2: School Adjustment

In order to facilitate students understanding of the functions and rules of school and to change the attitudes and behaviors of those having adjustment problems, a short project was developed, involving the use of a technique to foster positive behaviors, combined with digital storytelling. A total of a hundred and five first-grade students participated, from five primary schools in Athens, Greece. The project lasted from mid-September (schools in Greece start in mid-September), until late October 2015.

Behavioral modeling and, in specific, mimicking was utilized (Akers, 1977). In mimicking, one observes a model that expresses the desired behavior and, subsequently, he/she adopts it (Rogers, 2003). If ready-made stories were used, it would be like lecturing students. Instead, students were asked to develop their own. Since they were not yet able to read and write (at least efficiently), they acted as the "brains" and teachers acted as their "hands". The digital stories' developing software was projected using the classes' video projectors, students saw what available choices they had, collectively determined what to do, and "commanded their teachers to execute their will. In reality, teachers indirectly guided students to certain key points, by constantly asking questions about the conditions and behavioral problems that prevail in their classrooms. The outline of the stories was "A day at school" and the idea was to develop two-part digital stories. On the first, "wrong" students' behavior and dysfunctional classes were depicted. On the second, all problems were resolved and the "ideal" students' behavior was portrayed. This stage of the project lasted for two weeks (seven two-hour sessions).

Multiple sources of evidence were to be used as suggested by Yin (1994) and Paton (1990); pre-, during, and post-stage observations together with pre- and post-stage interviews with teachers and students were used for data collection purposes. This is a form of triangulation, which allows the

verification of interviews while interviews allow the researcher to explore the internal aspects of the underlying behavior (Patton, 1990). The focus of the observations was incidents of poor school adjustment and behavioral problems. The teachers and the students presenting significant adjustment problems were asked about the observed episodes, in order to understand and clarify their intentions and/or interpretations of the events.

During the observations prior to the development of the stories, it was noted that 14 students (9 boys and 5 girls) had considerable adjustment and behavioral problems and became the study's focus students. Each repeatedly exhibited the following categories of problems: (a) lack of self-restraint/discipline, (b) lack of interest or denial of participation in the lessons or in the school activities, and (c) denial to follow rules. Also, 24 students presented some of the above problems, but these were sporadic. The rest of them did not present any problems, or they were negligible. A noteworthy finding of the focus students interviews was that, though they could understand that their behavior was wrong, they could not make the connection between wrong behavior and its consequences, except for the ones that were related to them.

Students found the notion that teachers were their "hands" and that they were the "brains" very interesting and fun, and they were constantly asking to add more scenes to the stories. In the first part of the stories they effortlessly illustrated the characters' wrong behavior at school. What was not expected, but actually happened, was students to easily portray the ideal conditions. All the basic functions and rules that govern school seemed to be understood and the same applied to what was considered appropriate behavior.

During post-stages' observations, a sharp decrease in all problems was noted, but they were not totally eliminated. The results were especially interesting in the focus group, in which the majority of the focus students (12 out of 14) exhibited only minor behavioral issues. Also, teachers and students alike, quite often, referred to the digital

stories during in-classroom conversations and/or arguments. During post-stage interviews, the matter of focus students not being able to make the connection between wrong behavior and its consequences to others was reinstated, to evaluate if any changes had occurred. Indeed, 8 students gave answers that clearly indicated that they understood that there were broader implications.

Case 3: Dealing with Bullying

Since incidents of bullying are becoming more frequent in Greece's schools, an intervention was planned and carried out in the fourth grade of a primary school in Rhodes, Greece, from late October 2015, till late November 2015 (ten two-hour sessions). Twenty-four students were involved. The goal was to inform students about bullying. The main idea was students to work in groups of four, develop bullying stories, present and discuss them to the classroom, and collectively develop one final story. The researcher offered no help or guidance and did not intervene in the process, with one exception. In the final story, he suggested three alternative endings; the victim remained silent and continued to be tormented by the bully, the victim reacted, but failed because the bully was overwhelmingly powerful, and the victim became a bully himself, harassing younger students. Following the presentation of the final story, students were asked to write a short essay, presenting their thoughts and feelings for these endings and the way they would have reacted. Research data were obtained by analyzing the digital stories and the short essays using the iterative coding process (Creswell, 2002) to identify the categories, themes, and patterns that emerged from the data. All stories were viewed once, by two individuals, to identify the main ideas. Then, they were re-viewed in more detail and the ideas were labeled with codes. This process was repeated two more times to reduce overlap and redundancy of the codes until a small set of sub-themes were identified.

Physical together with verbal bullying were the main themes in all scenarios. All stories although simplistic, accurately illustrated what bullying is: the repeatedness of the incidents, the use of violence and bad language, the abuser's overwhelming power, the victim's reluctance to report the events. Emotions, like fear, depression, loneliness, and embarrassment, were also accurately presented. Bullies in all cases were boys. In all but two cases, the victim was a boy. Bystanders were included, but in most cases, there were no dialogues or thoughts accompanying these characters, so it is impossible to know which type of bystanders were portrayed. On a story, an observer helped the victim to beat the bully. The rest of the stories ended with the victim talking to an adult (two to a teacher, two to the victim's mother, one to the headmaster).

The three alternative endings of the final story were, in essence, wrong. Students' essays were evaluated on the basis of the reasoning for rejecting them and the reasoning for selecting their own "right" course of action. The negative emotions one has when being bullied, were enough for rejecting the first two of the alternative endings. The reason for rejecting the third ending was almost totally in line with the phrase: "We do not treat others in the same way they (wrongly) treat us". There were six cases (two boys and four girls) in which students stated that they would talk to an adult only if the bullying situation becomes intolerable, while in the rest, students stated that they would react immediately. The negative emotions of being bullied were once again the reason for seeking help if they were the victims of bullying. Finally, in all cases, talking to a parent, a teacher, the school's headmaster, to a friend or a combination of the above, were considered the "right" reaction if students were bullied or witnesses to bullying.

SOLUTIONS AND RECOMMENDATIONS

Setting aside the satisfactory results that all had, the cases presented in the previous sections share important common features. In all, digital storytelling was used in ways that deviate from its mainstream uses, that is in a strictly instructional setting, for simply telling a story, or for acquiring literacy skills. Taking advantage of storytelling's compatibility with young students' mentality (Robin, 2008), they explored alternative uses that all of them tried to resolve issues that students, as well as teachers, often face at school. Indeed, bullying, immigrants' integration problems and problems during the first days at school, are not rare and dealing with them can be quite difficult.

The rationale behind the three cases was based on two concepts. The first was that students' active participation is fundamental. Although a researcher, and/or a teacher was present, their role was limited; offering technical assistance when and if needed, not intervening at all, or covertly guiding. Social constructivism provided the theoretical foundation. By adopting the Vygotskian perspective, teachers were the ones who -indirectly- guided and supported students (Niesel & Griebel, 2007). By adopting the Piagetian perspective, students collaborated, negotiated, and came to a common consensus on how to develop their stories (Smith, 2012). By avoiding lecturing and direct manipulation, students' stories were not the result of someone imposing his/her thoughts and views on them. Instead, students were free to:

- Visualize their thoughts when developing the stories (Regan, 2008).
- Use, as raw material, pieces of previous knowledge and experiences they might have had.

- Construct their own understanding regarding important issues and socially negotiate them (Ertmer & Newby 2013).
- Embed their assumptions in real-world situations and determine by themselves the “right” course of actions.
- Become communicants of someone else’s thoughts, feelings, and problems (by viewing the digital stories) (Cane, 2010).

As a result, they formed a more comprehensive understanding regarding the issues that were discussed in the digital stories (Lenette, Cox, & Brough, 2013). Rules, instructions, good practices, behavioral patterns, and someone else’s emotions, require being deeply understood, before being accepted, applied, or become one’s own viewpoint.

The study’s second notion was that digital storytelling can be a form of a narrative therapy by helping students to discover parts of their personality (Sawyer & Willis, 2011). The power of stories as therapeutic means has been recognized by psychotherapy decades ago (White and Epston, 1990) The above was considered and utilized in the case of the immigrant student. Reflection on her problems, during the development of her story, and externalizing them during the presentation of the story, held the key to overcoming her problems (Rosenthal, 2003). Reflection and discussion of an issue were also key elements in the other two cases. Literature suggests that, when it comes to digital narrations, stories have to be told directly and unfiltered in order to act as a narrative therapy but also to start a dialogue among students (Harvey & Robinson, 2012). In turn, the discussion of a narrative promotes mutual understanding between students (Caine, 2010). The digital stories offered the basis upon which the process of documenting their experiences, reflecting upon them, and discussing them, was build and facilitated.

One has to be reminded that the researchers (or the teachers) avoided intervening, for purely research reasons; for not altering the results or for letting the students work by themselves. Contrary to that, in real-life situations, the teachers would

not be restricted. Taking advantage of the close affiliation between a teacher and his students, which is fundamental (Hamre & Pianta, 2006), results of such interventions are expected to be even better. That is because students will feel more comfortable in expressing themselves and the teachers, by knowing the background of each student, will be able to guide them more efficiently.

Considering all the above, it can be concluded that digital storytelling offered a simple, yet effective, solution to issues that are otherwise difficult to deal with. This solution can be summarized in just three steps: (a) ask students to develop digital stories on an important issue, (b) let them free to reflect upon that issue while developing their stories, and (c) discuss with them the issue while the stories are presented.

FUTURE RESEARCH DIRECTIONS

The main limitation in all the cases was the small sample size. Also, they were all conducted in Greece. Therefore, their results cannot be easily generalized. Further studies are needed with larger sample sizes and from different educational systems, in order to identify differences or similarities to the findings of the present cases and to obtain more reliable results. In addition, since the duration in all cases was short, longer-term projects can be tested, examining and comparing their results to short-term projects.

It would also be interesting to conduct research using conventional, instead of digital storytelling and compare the results. By doing so, it would be possible to determine if the results can be attributed to the medium used and/or to the methods.

Finally, one has to keep in mind that bullying troubles students of all ages, immigrant students face problems regardless of the level of education they study, and the transition from one level of education to another is always a sensitive period to a young student’s life. Therefore, virtually students of all ages can become target-groups of studies similar to the ones described in the previous sections.

CONCLUSION

In all the cases that were presented, the problems that were studied were dealt efficiently. Furthermore, all projects were short in duration and can be easily applied, without altering the school's timetable. Though it is certain that long term interventions yield good results, time is a crucial factor. Short term innovative interventions are needed because results can be produced right away and problems can be dealt on the spot. In all cases, whatever results were attained, were achieved fast, probably easing the way to follow up, longer term interventions. Also, no specialized equipment was needed and software similar to the one that was used is freely available.

In addition, the simplicity of the cases' design allows similar interventions to be easily applied to kindergarten, as well as to older students. Thus, teachers, as well as policy makers, can consider using their findings when designing similar or more well-organized, long-term interventions.

In conclusion, although the small sample size in all cases constitutes a considerable limitation, nonetheless, results point toward one direction; digital storytelling is a flexible and powerful instrument, an all-in-one tool, that can be used in many and diverse situations, educational or non-educational.

REFERENCES

- Akers, R. L. (1977). *Deviant Behavior: A Social Learning Approach* (2nd ed.). Belmont: Wadsworth.
- Brooker, L. (2008). *Supporting Transitions in the Early Years*. Maidenhead, UK: Open University Press.
- Caine, V. (2010). Visualizing community: Understanding narrative inquiry as action research. *Educational Action Research*, 18(4), 481–496. doi:10.1080/09650792.2010.524820
- Coventry, M. (2008). Cross-currents of pedagogy and technology: A forum on digital storytelling and cultural critique: Introduction. *Arts and Humanities in Higher Education*, 7(2), 165–170. doi:10.1177/1474022208088646
- Creswell, J. W. (2002). *Educational research: Planning, conducting, and evaluating quantitative*. Upper Saddle River, NJ: Prentice Hall.
- Dockett, S., & Perry, B. (2009). Readiness for school: A relational construct. *Australian Journal of Early Childhood*, 34(1), 20–25.
- Ertmer, P. A., & Newby, T. J. (2013). Behaviorism, cognitivism, constructivism: Comparing critical features from an instructional design perspective. *Performance Improvement Quarterly*, 26(2), 43–71. doi:10.1002/piq.21143
- Fabian, H. (2007). Informing transitions. In A. Dunlop & H. Fabian (Eds.), *Informing Transitions in the Early Years: Research, Policy and Practice* (pp. 3–20). Maidenhead, UK: Open University Press.
- Flyvbjerg, B. (2006). Five misunderstandings about case-study research. *Qualitative Inquiry*, 12(2), 219–245. doi:10.1177/1077800405284363
- Galanaki, E., & Vogiatzoglou, P. (2015). Εκφοβισμός/θυματοποίηση στο σχολικό πλαίσιο και μοναξιά [Bullying/victimization in the school context and loneliness]. *Παιδαγωγική*, 44.
- Genette, G. (1998). *Die Erzählung* (Vol. 2). Wilhelm Fink Verlag-München.
- Hagan, W. J. (2007). Case studies from the new entrant classroom: Children's developing repertoires of participation. *New Zealand Research in Early Childhood Education Journal*, 10, 95–104.
- Hamre, B. K., & Pianta, R. C. (2006). Student-teacher relationships. In G. G. Bear & K. Minke (Eds.), *Children's Needs III: Development, Prevention, and Intervention* (pp. 59–71). Washington, DC: National Association of School Psychologists.

- Harvey, H. B., & Robinson, K. (2012). Intercultural storytelling performance in Morocco and the United States. *Storytelling, Self, Society*, 8(3), 180–193.
- Hellenic Statistical Authority. (2011). *Greek Census 2011*. Retrieved from <http://www.statistics.gr/home>
- Kyriazis, N., & Zacharias, P. (2015). Development of an educational digital game for learning social skills and anti-bullying behaviors in schools. *Proceedings of the 4th Panhellenic Conference on ICT in Education*.
- Lathem, S. A. (2005). Learning communities and digital storytelling: New media for ancient tradition. *Proceedings of Society for Information Technology & Teacher Education International Conference*, 2286–2291.
- Lenette, C., Cox, L., & Brough, M. (2013). Digital storytelling as a social work tool: Learning from ethnographic research with women from refugee backgrounds. *British Journal of Social Work*, 1–18.
- Manesis, N., & Lambropoulou, A. (2014). Σχολικός εκφοβισμός: Ενέργειες εκπαιδευτικών για την πρόληψη και την αντιμετώπισή του [School bullying: Training actions for preventing and dealing with it]. *Παιδαγωγική Θεωρία & Πράξη*, 7, 83–89.
- Niesel, R., & Griebel, W. (2007). Enhancing the competence of transition systems through co-construction. In A. Dunlop & H. Fabian (Eds.), *Informing Transitions in the Early Years: Research, Policy and Practice* (pp. 21–32). Maidenhead, UK: Open University Press.
- Ohler, J. (2006). The world of digital storytelling. *Educational Leadership*, 63, 44–47.
- Patton, M. Q. (1990). *Qualitative evaluation and research methods* (2nd ed.). Newbury Park, CA: Sage.
- Phinney, J. S. (1996). Understanding ethnic diversity. *The American Behavioral Scientist*, 40(2), 143–152. doi:10.1177/0002764296040002005
- Regan, B. (2008). Why we need to teach 21st century skills and how to do it. *Multimedia & Internet @ Schools*, 15(4), 10–13.
- Retali, K. (2013). Performance of pupils with immigrant background in Greece: Findings and challenges. In P. Angelidis & H. Hadjisotiriou (Eds.), *Intercultural Dialogue in Education: Theoretical Approaches, Political Origin and Pedagogical Practices*. Athens: Diadrasi.
- Rigby, K., & Johnson, B. (2006). Expressed readiness of Australian schoolchildren to act as bystanders in support of children who are being bullied. *Educational Psychology*, 26(3), 425–440. doi:10.1080/01443410500342047
- Robin, B. (2006). The educational uses of digital storytelling. *Proceedings of Society for Information Technology & Teacher Education International Conference*, 1, 709–716.
- Robin, B. (2008). *The effective uses of digital storytelling as a teaching and learning tool*. In *Handbook of Research on Teaching Literacy Through the Communicative and Visual Arts*. New York: Lawrence Erlbaum Associates.
- Rogers, B. (2003). *Behaviour Recovery: Practical Programs for Challenging Behaviour and Children with Emotional Behaviour Disorders in Mainstream Schools*. Australian Council for Educational Research.
- Rosenthal, G. (2003). The healing effects of storytelling: On the conditions of curative storytelling in the context of research and counseling. *Qualitative Inquiry*, 9(6), 915–933. doi:10.1177/1077800403254888

Salmivalli, C., Lagerspetz, K., Bjorkqvist, K., Osterman, K., & Kaukiainen, A. (1996). Bullying as a group process: Participant roles and their relations to social status within the group. *Aggressive Behavior*, 22(1), 1–15. doi:10.1002/(SICI)1098-2337(1996)22:1<1::AID-AB1>3.0.CO;2-T

Sawyer, C. B., & Willis, J. M. (2011). Introducing digital storytelling to influence the behavior of children and adolescents. *Journal of Creativity in Mental Health*, 6(4), 274–283. doi:10.1080/15401383.2011.630308

Skourtou, E., Vratisalis, K., & Govaris, C. (2004). *Μετανάστευση στην Ελλάδα και Εκπαίδευση: Αποτίμηση της Υπάρχουσας Κατάστασης. Προκλήσεις και Προοπτικές Β* [Immigration in Greece and Education: Assessment of the Existing Situation. Challenges and Prospects for Improvement]. Athens: IMEPO.

Smith, P. J. (2012). Children as teachers: Creating opportunities for children to share expertise with their peers. *New Zealand Journal in Early Childhood Education*, 15, 84–101.

White, M., & Epston, D. (1990). *Narrative Means to Therapeutic Ends*. New York: W. W. Norton & Company.

Yin, R. K. (1994). *Case study research: Design and method* (2nd ed.). Thousand Oaks, CA: Sage publications.

ADDITIONAL READING

Benick, G. (2012). Digital storytelling and diasporic identities in higher education. *Collected Essays on Learning and Teaching*, 5, 147–152.

Benmayor, R. (2008). Digital storytelling as a signature pedagogy for the new humanities. *Arts and Humanities in Higher Education*, 7(2), 188–204. doi:10.1177/1474022208088648

Emert, T. (2014). “Hear a story, tell a Story, teach a story”: Digital narratives and refugee middle schoolers. *Voices from the Middle*, 21(4), 33.

Piaget, J. (1997). *The Moral Judgement of the Child*. Simon and Schuster.

Rogoff, B. (2003). *The Cultural Nature of Human Development*. Oxford: Oxford University Press.

Rogoff, B., Paradise, R., Arauz, M. B., Chavez, M. C., & Angelillo, C. (2003). Firsthand learning through intent participation. *Annual Review of Psychology*, 54(1), 175–203. doi:10.1146/annurev.psych.54.101601.145118 PMID:12499516

Trilling, B., & Fadel, C. (2009). *21st century skills: Learning for life in our times*. John Wiley & Sons.

Vygotsky, L. (1978). Interaction between learning and development. *Readings on the Development of Children*, 23(3), 34–41.

KEY TERMS AND DEFINITIONS

Bullying: The repeated use of violence (verbal or physical) and/or threats, for abusing, intimidating, or dominating others.

Constructivism: A learning theory which argues that humans generate knowledge and meaning from their experiences. Although not a specific pedagogy, is the underlying theme of many education reform movements.

Digital Storytelling: A digital form of a story that combines a conventional story (oral or written) and multimedia and/or hypermedia elements.

Mimicking: The observation of a model that expresses the desired behavior and, subsequently, adopting it.

Narrative Therapy: A form of psychotherapy in which an individual, together with the therapist, co-authors a narrative about himself/herself. Through this process, the values, skills and knowledge one has are identified, so as to effectively confront whatever problems he/she faces.

School Adjustment: The process of adapting to the role of being a student and to various aspects of the school environment.

Social Integration: The process in which all members of the society are engaged in a dialogue to achieve and maintain peaceful social relations. It does not imply or suggest forced assimilation.

E

3D Printing Applications in STEM Education

Norman Gwangwava

Botswana International University of Science and Technology, Botswana

Catherine Hlahla

National University of Science and Technology, Zimbabwe

INTRODUCTION

The chapter focuses on the design of educational toys for early school aged children, based on their anthropometric measurements. It also covers case study applications of 3D printing in engineering design undergraduate studies. Research on existing educational toys and different child development stages was carried out. Concepts were generated from the collected data and the best concepts selected through ranking methods. Dimensioning of the selected concepts was based on the collected anthropometry data. STL files were used to manufacture the chosen concepts by means of 3D printing.

Children are active learners who use the physical environment in a direct, hands-on manner to develop different skills. Toy experts believe that educational toys play a large role in the development of children. They stimulate play, language and reading skills and help children achieve milestones in both gross and fine motor skills. The implementation of ergonomics and the consideration of children's anthropometry dimensions in the design of toys play an important role in ensuring safety and injury risk reduction of children during play.

The National University of Science and Technology (NUST) is exploring 3D printing technology in the lecture room for its BEng program; ready-to-use 3D printed gardening implements, Mass-Customization of Office Mini-Storage Products from 3D Printing and other research

projects. 3D printing enables students in science, technology, engineering and mathematics (STEM) to visualize concepts.

BACKGROUND

Many of the children toys in the global market are imported from other countries, specifically China. The designers of these toys aim at achieving as good anthropometric match for as many potential customers in their country as possible. Thus the toys are custom designed to suit body dimensions of the children in that particular country yet the same toys are being exported and used by children across the world. Accidents and musculoskeletal health problems may occur due to incorrect product dimensions and sizes that do not meet the children's dimensional requirements.

Anthropometric data for children reflect general health status, dietary adequacy and growth and development over time (McDoweliet et al, 2008). Although several researchers have studied the anthropometry of children, they have most related their studies to nutritional, health and growth aspects (Khor et al, 2009). There are a few studies on the importance of child anthropometry in the design of various child products, specifically toys. Anthropometric measurements are necessary to form the data base which is required for the proper sizing of educational toys. Although the idea of considering child anthropometry in the design for child products is not new, the scarcity of avail-

able sources on anthropometric data among early school aged children calls for more anthropometric research so as to customize the children toys.

With respect to higher education needs, STEM education is more demanding in terms of well equipped laboratories, prototyping needs, as well as building experimental conditions that match the practical world. The pass rate at NUST has been low due to inadequate facilities for STEM enrolled students. 3D printing helps students to bridge the gap between the practical STEM world and the lecture room. Ease of recycling prototypes is also very important to keep the STEM training costs low. 3D printing, coupled with modular design concepts, was proved to be the best choice in cost effectiveness.

LITERATURE REVIEW

This section covers recent, historical and empirical reviews laying the foundation for the present study. Information that is relevant for the anthropometric research for the design of educational toys is presented. The section gives an insight into anthropometry, educational toys and 3D Printing which are the main subjects used to meet the objectives of the chapter.

Anthropometry

Anthropometry is the science that measures the range of body sizes in a population (Pheasant & Haslegrave, 2005). In product design, anthropometrics is the use of body measurements to determine the optimum size for products for comfortable and efficient use. Designers integrate the use of anthropometric data in their design process to optimize the usability and functioning of a product while improving comfort and safety. Advances in 3D imaging technologies have facilitated the collection of these measurements and shapes among the elderly or children (Goto, et al 2015). There are two primary types of anthropometric measurements; structural (static) and functional

(dynamic) measurement. Structural measurements are taken while the body is in a static position. These include skeletal dimensions (joint to joint measurement) and soft tissue measures in contour dimensions.

Dynamic measurements are taken while the body is engaged in some kind of activity like driving a car or reaching for objects. Engineering anthropometry is concerned with the application of both types of data to the design of the products people use.

Child Development Stages

Child development is the change or growth that occurs in children and the gaining of skills in all aspects of the child's life (Frost et al, 2008). It is often divided into three main areas; physical, cognitive, and social-emotional development.

Physical development fall into two main categories:

1. **Gross-Motor Development:** Involves improvement of skills using the large muscles in the legs and arms, such activities include running and bike riding.
2. **Fine-Motor Development:** The coordination of small muscles, in movements, usually involving the synchronization of hands and fingers with the eyes. Cutting, grasping, molding and writing are some of the activities that require fine-motor development.

Cognitive development is about how children learn, think and develop ideas. This is one of the areas of development that is strongly influenced by the experiences a child has. For example learning the names of animals is only possible if a child has been told them.

Social-emotional development is a child's ability to understand the feelings of others, control their own feelings and behaviours and get along with peers (Hoffman, 2013). Social and emotional development involves the acquisition of a set of skills. These include the ability to:

1. Identify and understand their own feelings
2. Accurately read and comprehend emotional states in others
3. Regulate their own behaviour
4. Develop empathy for others
5. Establish and sustain relationships

Educational Toys

The early childhood years are generally regarded as the foundation upon which the rest of an individual's life is built (Rutter, 2002). Educational toys are essential for the proper early development of growing children as they facilitate the mental and physical growth of children. Whenever a child plays with toys, they are practicing life skills. Educational toys enhance intellectual, social, emotional, and/or physical development.

Toys can be divided into several groups, depending on the part of the child they help to develop. Kudrowitz (2010) presented a universal classification system to communicate and ideate new toy concepts to toy designers and students. The system consists of two graphical tools that help designers to classify and manipulate toy product concepts.

Good educational construction toys promote convergent thinking while simultaneously setting the stage for creative, divergent play for children. While they follow the instructions to build the model, children gain skill with the construction pieces. Completing the instructions (convergent thinking) gives them a sense of achievement and reinforces confidence.

Interlocking manipulative toys like puzzles challenge the child to improve hand-eye coordination, patience, and an understanding of spatial relationships. Modern day children are growing up in the greatest era of technological advancements, hence there is need for them to stay in touch with technology. An example of these toys is a remote controlled toy car. Children can be inspired by educational toys which touch on everything from basic STEM to the real physical world. Table 1 outlines the educational toys for the different child development stages.

Table 1. Child development educational toys

Child Development	Type of Educational Toy
Physical or Muscle Development	• Magnetic blocks • Writing tools and scissors • Puzzles • Playground equipment
Intellectual Development	• Card games • Sorting games • Listening games • Books
Creative Development	• Clay • Crayons • Paints • Paper
Sensory Development	• Sand and water toys • Musical instruments • Bubbles
Make-Believe and Social Development	• Dolls • Puppets • Cars and trucks

3D Printing

3D printing is a manufacturing method based on advanced technology that builds up parts, additively, in layers (Ventola, 2014). The 3D printing process starts with a 3D digital model, created using 3D software such as 3D CAD. Table 2 shows different types of 3D printing technologies.

Current and Future Applications of 3D Printing

3D Printing technology is used in the children toy industry (pre-school and primary education), secondary and higher education.

Children Toy Industry

The toy industry is among the biggest beneficiaries of the technology since toys tend to be small and made out of plastic, making it easy to 3D print them. Children can customize their own unique toys by deciding on the features of the toys before they are 3D printed for them. Besides promoting creativity, 3D printing also speeds up and simplifies the making of toys, which in turn lowers the cost.

Table 2. The seven 3D printing/ AM process categories (ASTM, 2014)

Process Type	Brief Description	Related Technologies	Materials
Powder Bed Fusion	Thermal energy selectively fuses regions of a powder bed	Electron beam melting (EBM), selective laser heat sintering (SLS), selective heat sintering (SHS), and direct metal laser sintering (DMLS)	Metals, polymers
Directed Energy Deposition	Focused thermal energy is used to fuse materials by melting as the material is being deposited	Laser metal deposition (LMD)	Metals
Material Extrusion	Material is selectively dispensed through a nozzle or orifice	Fused deposition modeling (FDM)	Polymers
Vat Photopolymerization	Liquid photopolymer in a vat is selectively cured by light-activated polymerization	Stereolithography (SLA), digital light processing (DLP)	Photopolymers
Binder Jetting	A liquid bonding agent is selectively deposited to join powder materials	Powder bed and inkjet head (PBIH), plaster-based 3D printing (PP)	Polymers, Foundry Sand, Metals
Material Jetting	Droplets of build material are selectively deposited	Multi-jet modeling (MJM)	Polymers, Waxes
Sheet Lamination	Sheets of material are bonded to form an object	Laminated object manufacturing (LOM), ultrasonic consolidation (UC)	Paper, Metals

Education Sector Applications

The use of 3D printing in education was investigated by Slavkovsky (2012). Since physical models are important for hands-on active learning, 3D printing technology in education has been used since a while and considered for sustainable development, for secondary education in STEM projects and elementary mathematics education (Pearce et al, 2010; Berry et al, 2010; Lipson, 2007; Lacey, 2010). The best way for students to learn and retain information taught is by applying it to real-life situations. 3D Printing technology is being used in the classroom to offer a highly engaging, hands-on way to teach STEM subjects as well as to expose students to technologies of the future from an early age.

Knill and Slavkovsky (2013) illustrate how 3D printing can help to visualize concepts and mathematical proofs. The new 3D printing technologies make the realization of mathematical models more accessible than ever. Visualization has always been an important ingredient for communicating mathematics (Pedersen, 2005).

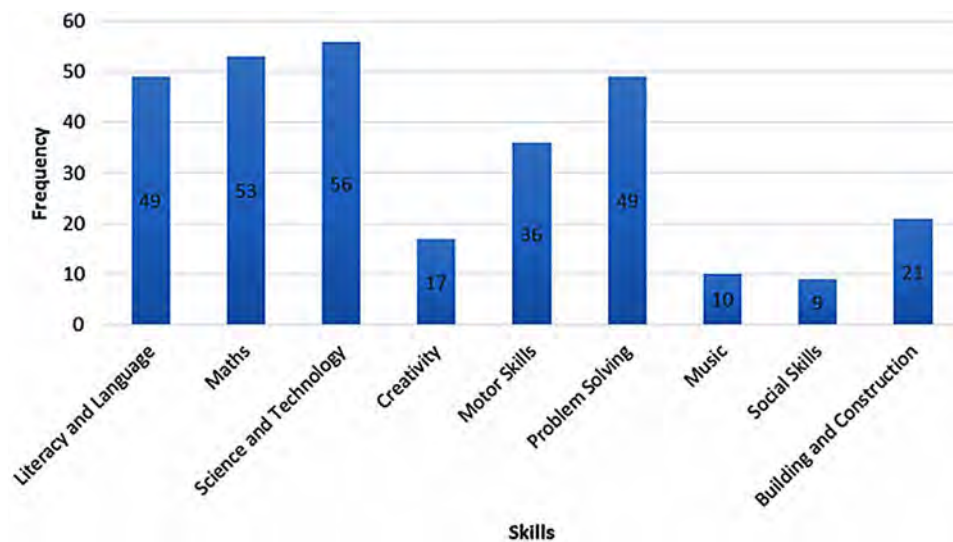
Many higher education institutions have re-worked all their lab activities so that they are based on a 3D printing platform. FDM 3D printers are used for most lab workbenches and are customized by installing automation hardware with display interfaces and input/output electronics. These upgrades give students valuable experience using real-world automation equipment.

Schelly et al (2015) investigated the potential of open-source (OS) technologies in an educational setting, given the combination of economic constraints affecting most educational environments. The OS aspect proves to be particularly relevant in the educational setting, as it decreases cost of access and encourages active participation, innovation and improvement through real experience with design and fabrication.

SOLUTIONS AND RECOMMENDATIONS

The section outlines the application of anthropometrics for early school aged children in the design of educational toys. It also covers case

Figure 1. Graph showing the frequency of each skill



study applications of 3D printing in engineering undergraduate studies. Research on existing educational toys and different child development stages was carried out. Concepts were generated from the collected data and the best concepts selected through ranking methods.

Pre- and Primary School Toy Concepts

Relevant data for the design of educational toys was gathered through the use of questionnaires, consultation and ergonomic study. The sample population used included individuals who usually spend reasonable time with children, especially mothers who have information concerning the toys that children play with. The following are some of the questions that were presented in the research instruments and their responses.

1. What are the key skills/educational values that children should acquire from playing with toys?

The majority of the respondents selected literacy and language, maths, science and technology, problem solving and motor skills as one of their choice. Figure 1 summarizes the responses.

Toy property rating for a number of toy properties that are relevant in the design of a toy were outlined in the questionnaire and the respondents had to rate the relevance of each property. Figure 2 summarizes the responses. Some of the responses augment the “Let’s Play!” projects at the University of Buffalo which formulated a number of universal design guidelines for toys (Hinske et al, 2008; Hengeveld et al, 2007).

1. Figure 3 shows a pie chart deduced from the toy attribute survey data. This helps in comparing the value of importance and prioritizing the technical attributes.
2. Based on the various research methods outlined previously, Table 3 summaries the educational requirements of early school aged children and examples of ideal toys to meet those educational requirements.
3. An anthropometric research was carried out in order to come up with children body dimensions that are essential in the design of educational toys.

Table 4 summarizes the mean and standard deviation for the various anthropometry dimensions in a typical source population composed of early school aged children between the age of 6 and 7

Figure 2. Graphical representation of the responses

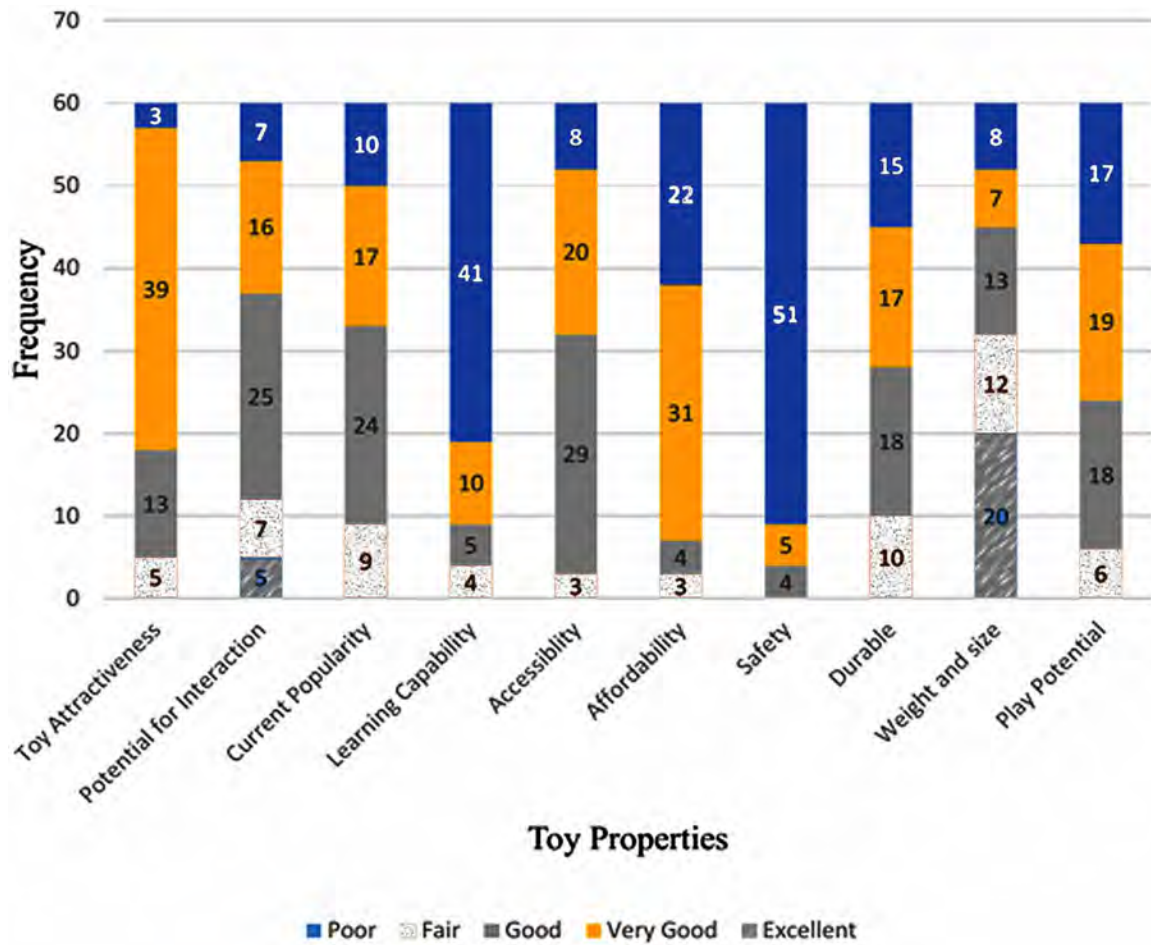


Figure 3. Pie Chart for toy attributes

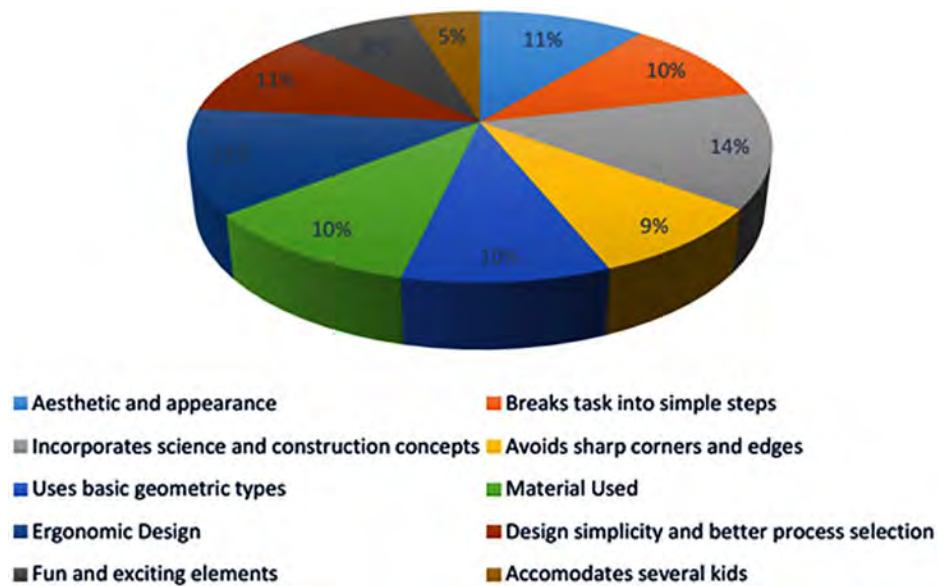


Table 3. Educational toy requirements

Research Methods	Educational Values/Requirements	Educational Toy Types
Questionnaire	- Scientific Concepts - Maths and Problem solving - Literacy and language	- Electronic Toys - Puzzles - Computer Toys - Dolls - Drawing Sets
Consultation	- Read and write - Informal/ Formal Communication - Mathematical calculation	- Building Toys - Mechanical Toys - Make-believe Toys - Art and craft Toys - Word Construction Toys - Phonic Toys
School Syllabus/ Primary school Curriculum	- Spoken English/ Communication/ Identifying - Addition and subtraction of number up to 20 - Division and multiplication - Counting - Cognitive and physical development - Tell clock times - Art and craft/Drawing	- Clock toys - Reading Cards - Abacus - Art Material - Story Books
Literature Review	- Cognitive - Social - Physical development	- Construction Toys - Puzzles - Musical Instruments - Pattern Making Toys - Transportation Toys - Jumping Ropes

Table 4. Anthropometry dimensions

years. Children in this age group will be attending primary school, either in grade 1 or grade 2.

Developed Toy Concepts

Toy concepts for three educational values were proposed. The three educational values deduced from Table 3 are:

- Scientific Concepts
- Mathematics and Problem Solving
- Literacy and Language Skills.

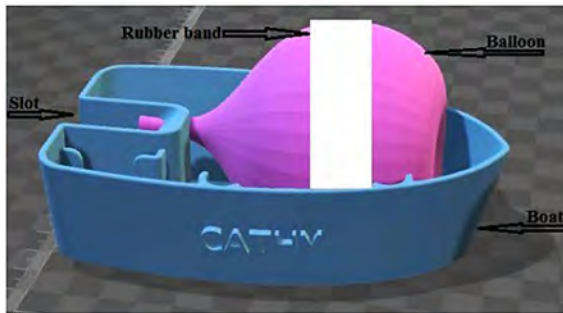
The dimensions of the final toy models for 3D printing were based on the anthropometric data presented in Table 4.

Concept for Scientific Toy

The Balloon Boat in Figure 4 was designed to move on water by the deflation of a balloon. The child blows up the balloon, closes the balloon opening by fingers then places the boat on top of water,

Anthropometry Dimension	Mean, μ (cm)	Standard Deviation, SD
Standing Height	109,6	2,5
Sitting Height	78,2	1,8
Standing Shoulder Height	81,6	1,3
Sitting Shoulder Height	55,2	1,5
Standing Eye Height	98,9	2,2
Sitting Eye Height	72,1	1,3
Sitting Elbow Height	40,5	1,6
Chest Breadth	19,3	0,8
Shoulder Breadth	29,8	1,0
Elbow-Hand Length	31,3	1,5
Shoulder-Elbow Length	24,7	1,9
Palm Length	13,4	0,3
Palm Width	6,2	0,1
Thumb Width	1,5	0,1
Thumb Thickness	0,9	0,1
Minimum Holding Distance Between Thumb and Index finger	2,1	0,2
Holding Distance Between Thumb and Middle Finger	5,4	0,3
Forward Reach	47,3	1,3

Figure 4. Scientific toy



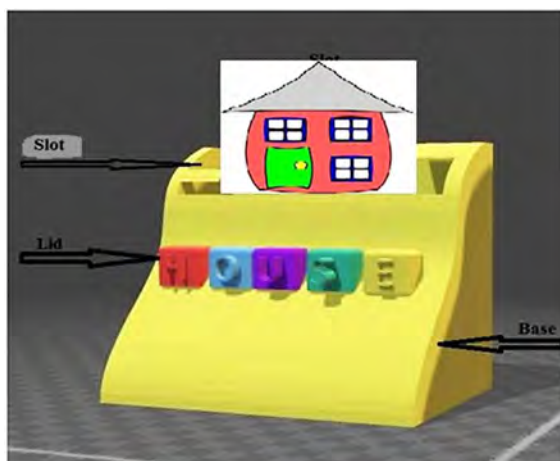
remove their finger and observe what happens. A rubber band is added on top to secure the balloon.

- **Advantages:**
 - Incorporates fun scientific concept
 - Offer opportunity to learn something new
 - Active participation of child during use
- **Disadvantages:**
 - Use of very light material for the boat
 - Require a wide container of water
 - Frequent balloon replacement

Concept for Literacy and Language Toy

Reading cards and construction toys are used to acquire reading and spelling skills that will

Figure 5. Literacy and language toy



obviously lead to better communication skills. The word construction tool and reading card are combined to come up with toy shown in Figure 5.

E

- **Advantages:**
 - Word construction skills
 - Spelling skills and reading skills
 - Overall improvement of language skills
- **Disadvantages:**
 - Does incorporate writing skills

Mathematics and Problem Solving Concept

The concept shown in Figure 6 combines the counting frame and the mathematics calculator to come up with the mathematics and problem solving toy.

Engineering Design Course Instruction

3D printing can serve as the best platform to deliver lectures and typical design assignments to undergraduate students in engineering. Two

Figure 6. Mathematics and problem solving toy

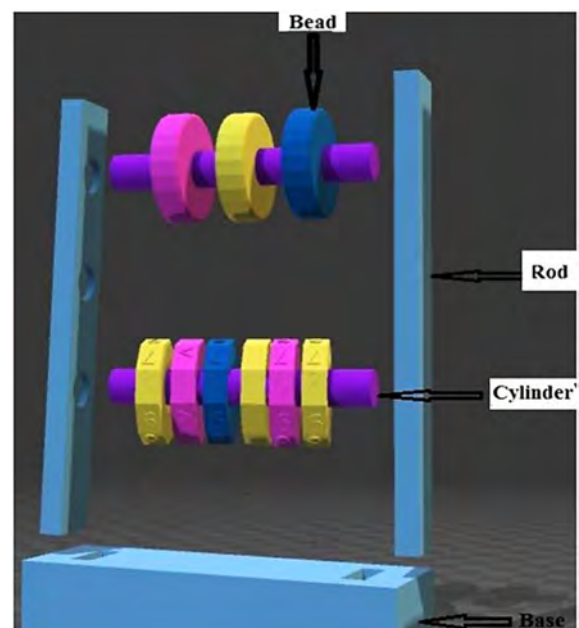
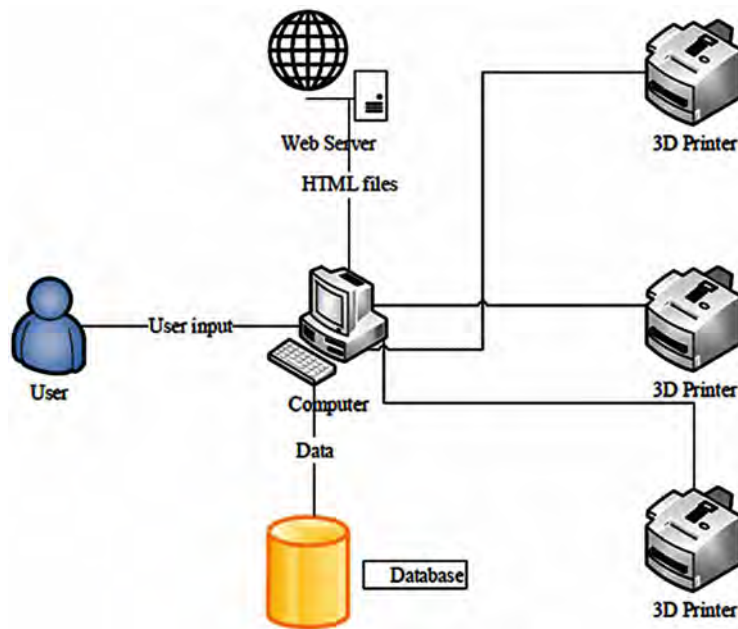


Figure 7. System architecture



projects are illustrated: online 3D printing mass customization platform and ready-to-use 3d printed gardening implements.

3D Printing Mass-Customization Design Project

3D printing offers opportunities for mass customization. An online system which can be used in customizing 3D printed products is under development at NUST. The system automatically generates the product as specified by the customer. Students and course instructors use the system to promote innovation and design thinking.

System actors for the Mass Customization include the administrator, and users. The administrator administers the system, adds, modifies and deletes product design templates. Users can modify and customise product templates. The system architecture and home page are shown in Figure 7 and 8 respectively. A similar application was conducted at Griffith University (Loy, 2014). The research considered the potential of 3D printing as an eLearning tool for design education and the role of eMaking in bringing together the virtual and the physical in the design studio.

Table 5 shows the product template, a cubic penholder that the customer uses to personalise the diameter and colour of the pen. It can be used for crayons, pencils and markers.

3D Printed Gardening Implements Design Project

Research on possible gardening implements design was carried out as part of an undergraduate engineering design course. Having explored a variety of possible designs, the best design of the gardening implements was chosen and optimized for 3D printing. The printed models can be further subjected to the testing phase to see if they meet the design requirements and to determine their efficiency. After satisfactory performance, the digital models of the designed gardening implements were uploaded on a website. The website consists of ready to print 3D gardening implements and farmers from all over the world can access them hence eliminating the need for farmers to be experts in CAD software for designing their own implements.

Three leading factors to consider during the design of 3D printed gardening implements were



Figure 8. The home page

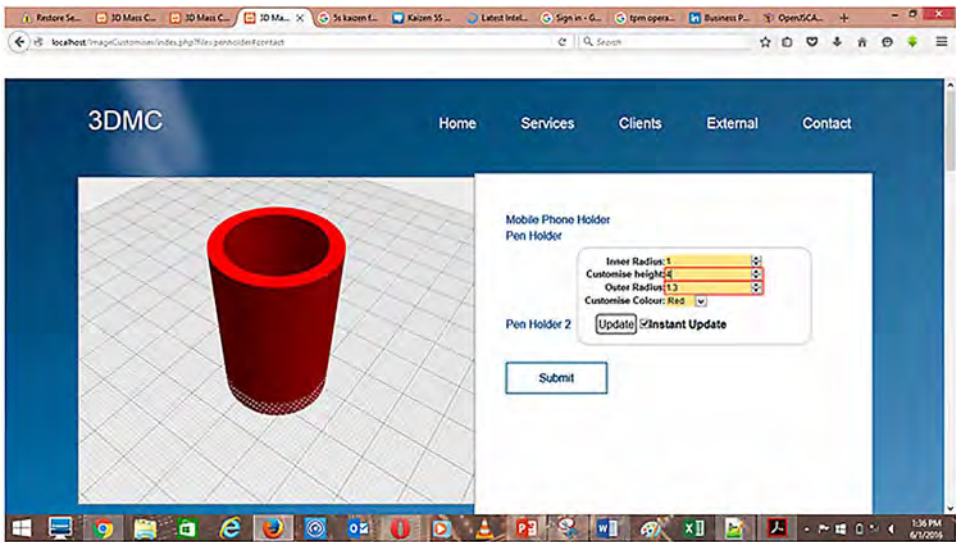


Table 5. Output system for penholder template

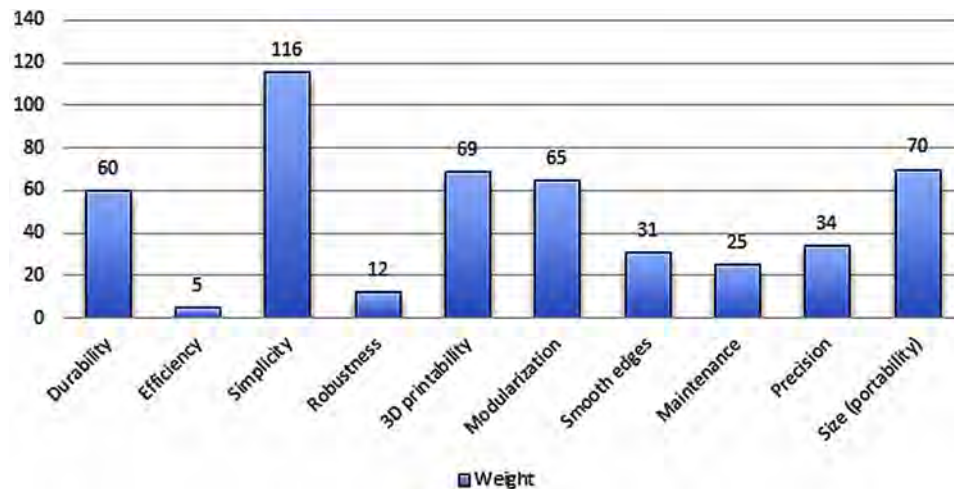
Parameter Input	System Output
Customise Colour: Blue Pen Radius 0.5 Update ☐ Instant Update Submit	
Customise Colour: Yellow Pen Radius 1 Update ☒ Instant Update Submit	

found to be simplicity, portability and durability. These weights are shown in Figure 9.

One chosen concept based on design attributes is illustrated in Figure 10. The design has a spinner for the sprinkler to be mounted on top of the sprinkler and will be rotating under the water

pressure distributing water evenly throughout a 360° angle. Due to the pressure of the water there is need for a mechanism to firmly hold the two together while leaving room for free rotation of the spinner.

Figure 9. Design attributes for gardening implements



Website for Gardening Implements Ready for 3D Printing

A website which anyone could access from anywhere to upload and/or download STL CAD files ready for printing was designed as part of the engineering design course. The domain name for the website is <http://anenyasha3dprints.com/>. The website is shown in Figure 11.

FUTURE RESEARCH DIRECTIONS

3D printing has three major weaknesses which are time consuming, mechanically weak printed models, and limited material selection.

Future research should concentrate on making the process faster as well as multi- material and colour printing. This would reduce the inventory

of 3D printers in laboratories whilst enabling students to build as many models within short time frames. Product specific research will boost the database of design for manufacture and assembly (DFMA) rules for 3D printing as well as designing ready for use 3D printed products.

The wide use of smart phones and data connectivity has caused a lot of boom in the design of mobile applications (Apps) for software products that traditionally demanded more resources from desktop computers. More mobile Apps can be designed in future that focus on 3D printing. There are quite a number of softwares that can be used for non-designers to customise their products. They are both online and offline. Programmers can make use of the free libraries offered by OpenJscad, OpenCascade, PythonOCC and many others to design products that are customisable.

Figure 10. Sprinkler design for 3D printing

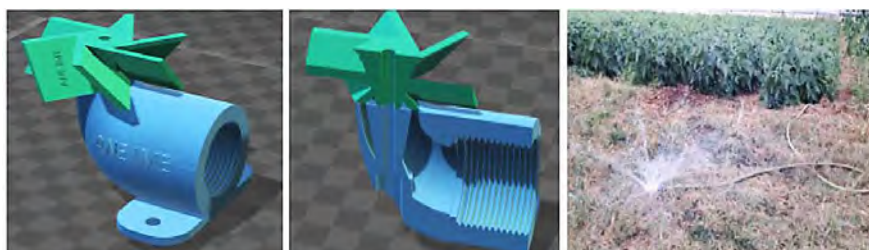
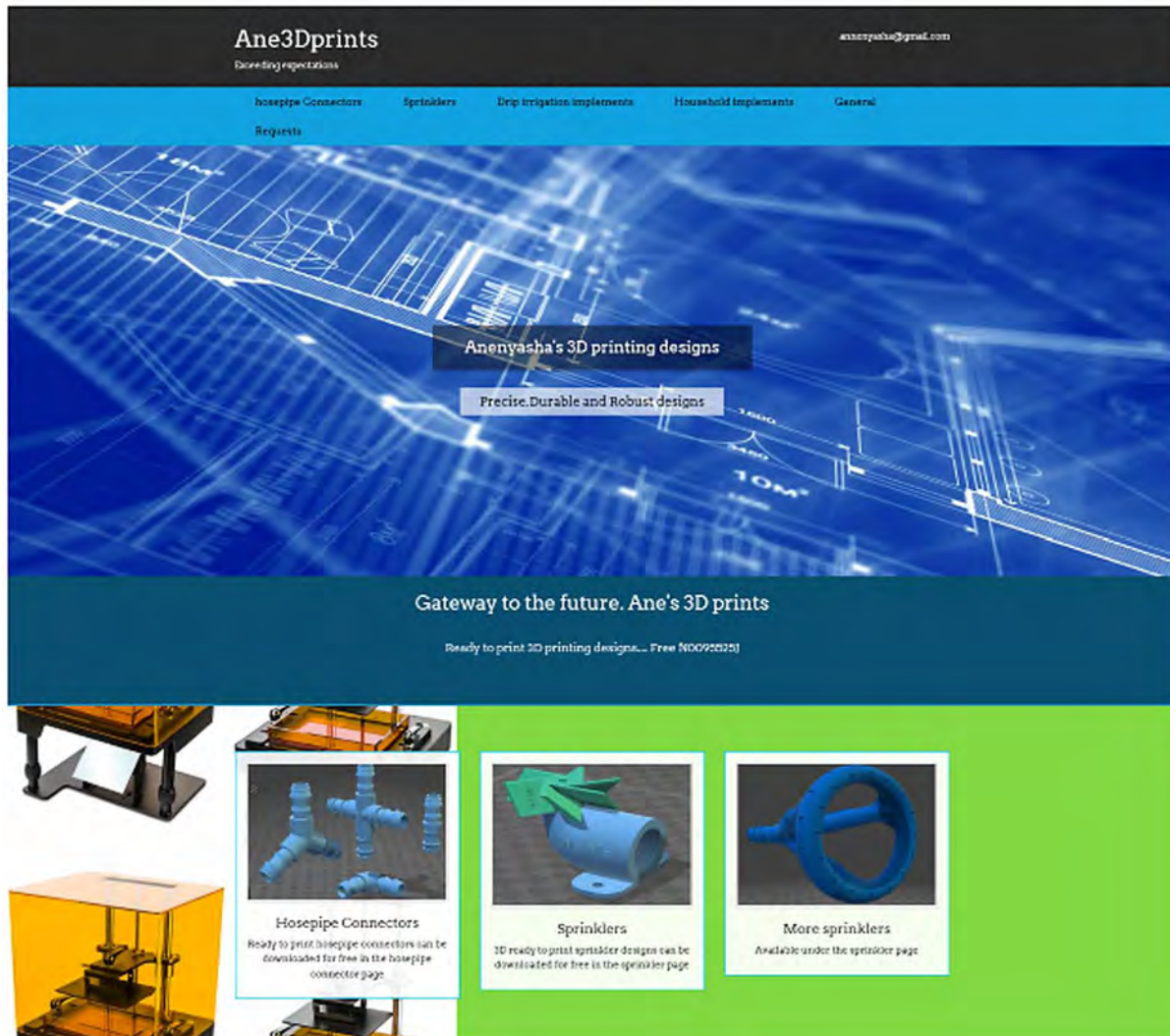


Figure 11. Website for gardening implements ready for 3D printing



Nowadays, e-Learning systems are on the increase and more focus is being directed towards mobile platforms. Since 3D printing is highly promising in the education sector, future research should focus on integrating 3D printing Apps to e-Learning systems. A number of CAD vendors are now offering cloud based solutions to their industrial and educational software solutions. Major examples are AutoDesk and Solid Works. This helps in democratising design and CAD modelling since the approach eliminates the need for high speedy computers. Therefore, picking up lessons from the trend to design 3D printing Apps

that can design digital models as well as actually command 3D printers from the students' remote comfort will be a perfect integration to the boom of e-Learning systems.

There is also a limited range of educational toys custom designed for children with disabilities. These children play with toys that do not suit their conditions in term of educational value, anthropometric dimensions, social and mental development. Therefore the design and customisation of educational toys for these children is a very important consideration for further work.

CONCLUSION

The chapter outlined anthropometric research for the design of educational toys which can be 3D printed for ready use. It has also shown how the technology can be used to cost-effectively allow institutions to offer their students state-of-the-art laboratory platforms, which match industry practice. 3D printing is proving to be the best platform for in-class instruction as well as for carrying out practical world design assignments. This was illustrated through two design projects carried out by engineering students in a design course; 3D printing mass-customisation for mini-office products and design of 3D printed ready to use gardening implements. Future research prospects have been highlighted, which include integration of 3D printing into the emerging mobile e-Learning platforms, as well as taking advantage of the cloud platform to offer Apps that allow digital object modeling for 3D printing.

REFERENCES

- Berry, R. Q., Bull, G., Browning, C., Thomas, D. D., Starkweather, K., & Aylor, J. H. (2010). Preliminary considerations regarding use of digital fabrication to incorporate engineering design principles in elementary mathematics education. *Contemporary Issues in Technology & Teacher Education*, 10(2), 167–172.
- Canessa, E., & Zennaro, M. (Eds.). (2013). Low-Cost 3D Printing for science, education and Sustainable Development. International Centre for Theoretical Physics.
- Kudrowitz, B. M., & Wallace, D. R. (2010). *The play pyramid: a play classification and ideation tool for toy design*. Academic Press.
- Frost, J. L., Wortham, S. C., & Reifel, S. (2008). *Play and child development* (3rd ed.). Upper Saddle River, NJ: Pearson.
- Goto, L., Lee, W., Song, Y., Molenbroek, J., & Goossens, R. (2015). Analysis of a 3D anthropometric data set of children for design applications. *Proceedings 19th Triennial Congress of the International Ergonomics Association*, 9-14. Retrieved July 5, 2016, from <http://www.iea.cc/congress/2015/1418.pdf>
- Hengeveld, B., Voort, R., Van, B., Hummels, C., & De Moor, J. (2007). Designing for diversity: Developing complex adaptive tangible products. *Proceedings of the 1st International Conference on Tangible and Embedded Interaction (TEI)*, 155-158. Retrieved November 9, 2016, from https://www.researchgate.net/publication/221308444_Designing_for_diversity_developing_complex_adaptive_tangible_products
- Hinske, S., Langheinrich, M., & Lampe, M. (2008). *Towards guidelines for designing augmented toy environments*. Retrieved November 9, 2016, from www.vs.inf.ethz.ch/publ/papers/hinske2008dis.pdf
- Khor, G. L., Noor Safiza, M. N., Jamalludin, A. B., Jamaiah, H., Geeta, A., Kee, C. C., & Ahmad, F. Y. et al. (2009). Nutritional status of children below five years in Malaysia: Anthropometric analyses from the Third National Health and Morbidity Survey III (NHMS, 2006). *Malaysia Journal of Nutrition*, 15(2), 121–136. PMID:22691811
- Knill, O., & Slavkovsky, E. (2013). *Illustrating mathematics using 3D printers*. Retrieved, July 20, 2016, from <https://arxiv.org/abs/1306.5599>
- Lacey, G. (2010). 3D printing brings designs to life. *TechDirections*, 70(2), 17–19. Retrieved from connection.ebscohost.com/c/articles/53992338/3d-printing-brings-designs-life
- Lipson, H. (2007). *Printable 3D models for customized hands-on education*. Paper presented at Mass Customization and Personalization (MCPC), Cambridge, MA.

Loy, J. (2014). eLearning and eMaking: 3D Printing blurring the digital and the physical. *Education Sciences*, 4(1), 108-121. Retrieved November 8, 2016, from www.mdpi.com/2227-7102/4/1/108

McDowell, M. A., Ogden, C. L., & Flegal, K. M. (2008). Anthropometric reference Data for Children and Adults: United States, 2003–2006. *National Health Statistics Reports*, (10): 22. PMID:25585443

Pearce, J. M., Blair, C. M., Kaciak, K. J., Andrews, R., Nosrat, A., & Zelenika-Zovko, I. (2010). 3D printing of open source appropriate technologies for self-directed sustainable development. *Journal of Sustainable Development*, 3(4), 17–28. doi:10.5539/jsd.v3n4p17

Pedersen, S. A., Mancosu, P., & Jorgensen, K. F. (2005). Visualization, Explanation and Reasoning styles in Mathematics. Springer.

Pheasant, S., & Haslegrave, C. M. (2005). Bodyspace: Anthropometry, ergonomics and the design of work. CRC Press.

Rutter, M. (2002). The interplay of nature, nurture and developmental influences: The challenge ahead for mental health. *Archives of General Psychiatry*, 59(11), 996–1000. doi:10.1001/archpsyc.59.11.996 PMID:12418932

Schelly, C., Anzalone, G., Wijnen, B., & Pearce, J. M. (2015). Open-Source 3-D Printing Technologies for Education: Bringing Additive Manufacturing to the Classroom. *Journal of Visual Languages and Computing*, 28, 226–237. doi:10.1016/j.jvlc.2015.01.004

Slavkovsky, E. (2012). *Feasability study for teaching geometry and other topics using three-dimensional printers* (Master's Thesis). Harvard University.

Ventola, C. L. (2014). Medical applications for 3D Printing: Current and projected uses. *Pharmacy and Therapeutics*. 39(10), 704-711. Retrieved July 25, 2016, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC189697/>

ADDITIONAL READING

Berman, B. (2012). 3-D printing: The new industrial revolution. *Business Horizons*, 55(2), 155–162. doi:10.1016/j.bushor.2011.11.003

Bhusnure, O. G., Gholve, V. S., Sugave, B. K., Dongre, R. C., Gore, S. A., & Giram, P. S. (2016). 3D printing and pharmaceutical manufacturing: Opportunities and challenges. *International Journal of Bioassays*, 5(1), 4723–4738. doi:10.21746/ijbio.2016.01.006

Burriss, K., & Burriss, L. (2011). Outdoor play and learning: Policy and practice. *International Journal of Education Policy & Leadership*, 6(8), 1–12.

Cheng, M., & Johnson, J. (2010). Research on childrens play: Analysis of developmental and early education journals from 2005 to 2007. *Early Childhood Education Journal*, 37(4), 249–259. doi:10.1007/s10643-009-0347-7

Frei, P., Su, V., Mikhak, B., & Ishii, H. (2000). *curlybot*: Designing a new class of computational toys. *Proceedings of the Conference on Human Factors in Computing Systems*, New York, 1-6, April 2000. Retrieved November 8, 2016, from <http://www.cs.uml.edu/~fredm/courses/91.548-spr03/papers/>

Gattouffie, K., & Reisman, S. (2007). Mass customization research: Trends, directions, diffusion intensity, taxonomic and framework. *International Journal of Flexible Manufacturing System*, pp. 637-666.

Ibrahim, A. M. S., Jose, R. R., Rabie, A. N., Gerstle, T. L., Lee, B. T., & Lin, S. J. (2015). Three dimensional printing in developing countries. *Plastic and reconstructive surgery global open* 3 (7): e443. Retrieved July 21, 2016, from <http://nrs.harvard.edu/urn-3:HUL.InstRepos:21462570>

Kelly, B. (2012). The State of STEM and jobs, US News and World Report. Retrieved July 7, 2016, from. National Math and Science Initiative. Increasing the Achievement and Presence of Under-Represented Minorities in STEM Fields. Retrieved July 19, 2016, from,

Priovolou, M. (2012). Let's go to the Movies! Learning math through creativity and role playing, Proceedings of the European Conference on Games Based Learning; 2012, 378-383.

Schubert, C., Van, L. M. C., & Donoso, L. A. (2014). Innovations in 3D printing: A 3D overview from optics to organs. The British Journal of Ophthalmology, 98(2), 159–161. Retrieved July 25 2016 PubMed doi:10.1136/bjophthalmol-2013-304446

Soni, R. 2006. Ergonomics and Its Relationship with Anthropometry. Retrieved July 17, 2016, from <http://soni2006.hubpages.com/hub/ergonomics-and-anthropometry>

KEY TERMS AND DEFINITIONS

3D Printing: A radically different manufacturing method based on advanced technology that builds up parts, additively, in layers at the

sub millimeter scale. Also known as Additive Manufacturing.

Anthropometry: The science that measures the range of body sizes in a population.

CAD: A software for 2D and 3D digital modelling of objects.

Educational Toy: A toy that helps a child learn something good, something that will help in the future.

e-Learning: Learning conducted via electronic media, typically on the Internet.

Ergonomic Design: A safe and comfortable design or product achieved through the use of anthropometric data to optimize the usability and functionality.

Mass Customization (MC): A business strategy that aims to provide customers with individualized products at near mass production efficiency.

STEM: A curriculum based on the idea of educating students in four specific disciplines — science, technology, engineering and mathematics — in an interdisciplinary and applied approach.

STL (STereoLithography): A standard file type (triangulated representation of a 3D CAD model) used by most additive manufacturing systems.

Tools, Pedagogical Models, and Best Practices for Digital Storytelling



Jari Multisilta

Tampere University of Technology, Finland

Hannele Niemi

University of Helsinki, Finland

INTRODUCTION

Sharing photos and short videos with others has become increasingly popular among youth. Using social media services, users share events and moments from their daily lives. Östman (2015) defined this phenomenon as *life-publishing*. Examples of life-publishing include the growing use of Snapchat and Periscope social media services among youth. According to Piwek and Joinson (2016), Snapchat users mainly share “selfies,” and they mostly use the service at home. Although sharing videos is a common activity among youth, schools are not using digital videos for learning. There is a need to study the pedagogical models that could be used in designing classroom activities involving the use of digital videos.

In this chapter, digital video storytelling refers to learning activities that involve the creation and use of digital video. According to Ladeira, Marsden, and Green (2011, p. 431), “digital storytelling typically seeks to preserve and disseminate real-life, non-fiction stories.” In a learning context, digital storytelling involves the creation and distribution of content that is used in the learning process as well as the interaction between the users of the content. Digital storytelling that includes user-generated content has been used in preserving personal experiences (Ladeira, Marsden & Green, 2011), mobile collaborative live video production, such as in an event in which a Video Jockey (VJ) mixes the video feed using

the audience (Engström, Esbjörnsson & Juhlin, 2008), and in collaborative learning (Niemi, Harju, Vivitsou, Viitanen, Multisilta & Kuokkanen, 2014; Niemi & Multisilta, 2015; Tuomi & Multisilta, 2010; Wolf & Rummmler, 2011). In this chapter, digital video storytelling will be discussed in the context of learning.

Digital video storytelling can be seen as an approach to learning twenty-first century skills. Taking advantage of the creative potential of modern communication technologies, students can work together, explore their ideas, and become creators, producers, and active learning participants. Twenty-first century skills have become a key topic on the agendas of education systems worldwide. Educators are required to seek new forms of teaching and learning for the future. The challenge is determining how to motivate students to learn and become engaged in learning. Digital video storytelling can assist in motivating students by bringing technologies they use in their free time into the school environment.

In this chapter, pedagogical models, examples, best practices, and outcomes that illustrate how students become engaged and motivated when using digital storytelling in knowledge creation in cross-cultural settings will be presented. The results are based on the empirical data and findings from several international pilots.

A review of existing tools and practices for digital video storytelling will be presented. The results show that students can become highly

engaged in learning through digital storytelling. The research data indicate that engagement in digital video storytelling is a combination of a joy of learning (fun) and a commitment to hard work.

BACKGROUND

The use of videos on the Internet has been expanding rapidly in the last few years. Although the most popular web video content is related to music videos and entertainment, web videos can have several educational uses. Khan Academy (www.khanacademy.com) is an example of a web video service that has a large collection of educational videos. According to Talbert (2012, para. 7), “Khan Academy is a collection of video lectures that give demonstrations of mechanical processes.” Considerable debate has taken place regarding the pedagogical model used at Khan Academy (Prensky, 2011; Talbert, 2012; Thompson, 2011). The main criticism is that Khan Academy is not supporting a constructivist learning model in which learners actively create knowledge using activities that support knowledge construction.

The creation of video stories by the learners themselves is considered a more effective way of using video in learning. According to Correia et al. (2005, p. 1), “the ability to have constant access through mobile devices allows a new way of doing cinematographic narratives that can enhance the participants’ experience in a significant way.”

Video stories can be interactive (Ladeira, Marsden and Green, 2011) or generated in real time with scripting (Vaucelle & Davenport, 2004). Storytelling platforms can also support automatic story creation (Multisilta & Mäenpää, 2008; Zsombori, et al., 2011). Multisilta et al. created a mobile social media service for community created videos (Multisilta & Mäenpää, 2008). MoViE, the *Mobile Video Experience Platform*, is a research platform for studying how people create video stories and how they share and learn with mobile social media. MoViE has been used both in primary and secondary schools (Tuomi &

Multisilta, 2010) as well as in higher education (Kiili, Multisilta, Suominen & Ketamo, 2010).

Digital stories have also been used to preserve, reflect on, and share the life experiences of people who do not have access to personal computers and who are living in rural areas in South Africa (Bidwell, Reitmaier, Marsden & Hansen, 2011). Digital story creation may also be able to provide a means of exploring self-identity through the sharing and group construction of digital video stories (Vaucelle & Davenport, 2004).

LEARNING USING DIGITAL VIDEO STORYTELLING

Pedagogical Models

In this subchapter, two pedagogical models for using digital videos and storytelling will be presented. The aim of the pedagogical models is to provide teachers with a sound basis for applying digital video storytelling tools in their own classrooms. The global sharing pedagogy and video inquiry learning models are both based on theoretical work that has been evaluated in empirical research projects.

As a pedagogical method, digital storytelling builds on learner-centered approaches that can improve students’ learning in several ways (Kearney, 2009; Yang, 2012). According to Niemi et al. (2014), learning with digital storytelling is seen as a socially and culturally related process that takes place in the interaction between a learner and material tools, psychological tools, or other human beings (Vygotsky, 1978). In this sense, it builds on the constructivist learning model. Learners play a central role in exploring and building knowledge by using tools available in the digital learning environment. Students also interact with psychological tools when using language, brainstorming, or creating stories. Learning with others can take place when creating video stories with peers and watching stories that other students have made. When planning and making

digital stories collaboratively, students can become aware of their own knowledge and experiences and reflect on and share these experiences with others. Watching other students' stories can also create new perspectives on topics and promote the understanding of a certain phenomenon (Niemi et al., 2014).

Global sharing pedagogy supports Vygotsky's ideas of learning (Vygotsky, 1978). According to Vygotsky, learning happens in social activity and higher mental functions are mediated by tools and signs. In addition, learning happens in the zone of proximal development, that is the difference between what a learner can do without help and what he or she can do with help. The help can be provided by the teacher or by the peer learners. In addition, Vygotsky pointed out that there is a close connection between learning, thinking, and the language. The language can be seen as a tool for both social activity and thinking.

In global sharing pedagogy, learning is a mediated activity with tools, signs, and social interaction. According to Niemi et al. (2014), there are four mediators of learning in the global sharing pedagogy: 1) *learner-driven knowledge and skills creation*, 2) *collaboration*, 3) *networking*, and 4) *digital media competencies and literacies*. The mediators contribute to the learners' engagement in the learning activity. Niemi et al. (2014) illustrated that engagement in the digital video storytelling activity could be divided into two components: joy of learning and hard work. With the use of digital tools, learning can happen both in social context and through individual learning.

Inquiry learning is defined as "an approach to learning that involves a process of exploring the natural or material world and that leads to asking questions, making discoveries, and rigorously testing those discoveries in the search for new understanding" (de Jong & van Joolingen, 2008, p. 458). In an earlier study, the researchers found that students find it difficult to argue and reflect on what they see in the video clips and video stories they and their peers have produced; however, argumentation, reflection, and commenting are important 21st century skills (Niemi et al., 2014).

In video inquiry learning, the learners and teachers capture mobile video recordings of events and phenomena that prompt questions from students and that serve as a basis for inquiries and collaborative learning in the Science, Technology, Engineering, and Mathematics (STEM) disciplines. Video inquiry learning is closely related to higher order thinking skills. According to Lewis and Smith (1993), higher order thinking consists of problem solving, critical thinking, creative thinking, and decision making. They explained that "higher order thinking occurs when a person takes new information and information stored in memory and interrelates and/or rearranges and extends this information to achieve a purpose or find possible answers in perplexing situations" (Lewis & Smith, 1993, p. 136). The definition clearly describes the process of inquiry video learning in which learners create video clips or video stories and argue, reflect, and comment on the content of the videos they and their peers have recorded and shared through social media.

Tools

There are different types of learning tools that can be used in digital video storytelling: 1) Learning Management Systems (LMS), 2) stand-alone video editors, and 3) video sharing sites. Several learning environments and LMS include an option to upload video files and share files with other users; however, such systems are not designed for digital video storytelling because they do not have editing or remixing features and their collaboration features are limited. Instead, their purpose is to manage class activities, deliver content, provide a platform for submitting exercises, and assess student activities. The stand-alone video editor software do not support sharing and collaboration. They provide a wide range of effects and tools for adding music, titles, and subtitles to video.

In this subchapter, five video sharing sites are compared based on a set of requirements that are typical for classroom use. The video sharing sites were selected based on the discussions with teachers. YouTube (www.youtube.com), Vimeo (www.vimeo.com),

vimeo.com), and WeVideo (www.wevideo.com) are popular video sharing services; TeacherTube (www.teachertube.com) is one of the oldest video sharing sites for teachers and MoViE (cicero-movie.edu.helsinki.fi) is one of the more recent video sharing sites targeted for learning.

The following features were reviewed:

- Access (how to sign on to the service).
- Content moderation.
- Editing capabilities.
- Possibility to do annotations to the content.
- Possibility to use analytical data from the user activities.
- Privacy settings.
- Existence of content grouping and channels.
- Limitation of what kind of content that can be uploaded to the site.
- Pricing model.
- Pedagogical support.

It is important to have an authentication system, but there are different regulations in different countries, such as how old children should be to have an Internet identity. For classroom use, it is important that the teacher is able to moderate the content to prevent the use of inappropriate content. Digital video storytelling often requires the remixing and editing of videos. In many cases, the learners create their stories by recording several short video clips that they edit and remix later. This can be done using a stand-alone video editor or video editing features provided by the video sharing service. Annotations allow for adding small comments over the video. These comments could be used as questions, or they could include important information related to the learning content. For example, a teacher could add an annotation to the video to remind students to pay attention to a specific event in the video. Statistical data provide information to the users regarding the use of the content and could potentially be very useful for teachers, but, so far, the video sharing sites do not have effective learning analytics tools.

Privacy, content, and cost are also important features for video sharing sites that are used in schools. All the reviewed systems have privacy settings for individual videos and groups. It is important to be able to create working areas for a class or a group of students, so that they have their own space where they can collaborate and share the videos. When schools and teachers are considering which video sharing site to use, it may be important to investigate what type of content is used in the service. Although a closed working area would be created on the site for students, the existence of entertainment content in the same system decreases the credibility of the system. Many systems are free or have a free trial option; however, the free access may include.

Most video sharing sites allow registered users to upload videos to the service and to set privacy limitations for the videos they have uploaded. Moreover, most video sharing sites are not designed for learning. Many video sharing sites allow users to watch videos, but uploading your own content requires registration either with the service or using a Facebook or Google identity. For schools, this may be a problem because in many countries 13 years is the minimum age that is required to sign up for Facebook or Google (Table 1).

YouTube, WeVideo, and MoViE have a video editor that can be used to do simple trims (YouTube), remixing (MoViE), or full video edits with sound and effects. YouTube and MoViE have a useful feature for annotating the videos. By using annotations, students can add subtitles that appear and disappear at certain times over the video. By using annotations, students could ask questions and make comments on the video. Some video sharing sites provide analytics tools for the owner of the video clip. Typically, the owner can see how many times the video has been watched and, in some cases, from where the watchers are geographically. Tools that could provide learning analytics to teachers would be very useful, but the systems do not have this feature yet. Moderation is also an important characteristic for teachers,

Table 1. Access and moderation of the content

	Access	Moderation
YouTube	Google UID	No
Vimeo	Vimeo UID, Facebook UID	Channels can have moderators.
WeVideo	Facebook UID	No
MoViE	MoViE UID	Yes
TeacherTube	TeacherTube UID	Content moderated by the service staff; content flags.

since it helps in preventing inappropriate content from being shared on the site. For TeacherTube, the service providers moderate the content. In MoViE and Vimeo, the teacher or channel owner can moderate the content. TeacherTube also has a red flag system in which users can flag inappropriate content for review by the service providers (Table 2).

All the reviewed video sharing sites have a method to create a specific working area, which is called a channel, album, project, group, or collection, depending on the system. The basic idea of the working area is to separate access for a certain group of people, who can share and access materials within. Only MoViE and TeacherTube have pedagogical concepts included in the service. TeacherTube includes features from LMS (such as a class management), but it is effective in utilizing the creation and delivery of video content (Table 3).

Best Practices

A digital video story can be created on almost any topic using several different methods. Table 4

shows examples of the video storytelling projects that the researchers have been implementing. In total, over 2000 video clips were created during the interventions. Some of the topics were performed similarly to a drama, that was scripted with roles and actors selected and camera angles planned well ahead. Other topics mimicked documentaries and included interviews of the subjects. In some classes, it was not possible for the students to appear in the video clips. In these cases, students created animations using self-drawn paper dolls, or they created and presented a series of slides with narration. As a result, in all of the videos the students' voice was strong. The teachers and the researchers helped the students complete the manuscript (Penttilä, Kallunki, Niemi & Multisilta, 2016). In addition, the students needed help in selecting the shooting locations so that the sounds from the surrounding environment did not significantly affect the final outcome. Many videos were shaky because they were recorded using a mobile device that was not attached to a tripod. The students showed creativity, such as in adding a background soundtrack from a music player to the video clips while recording.

Table 2. Edit, annotations, analytics, and privacy

	Edit/Remix	Annotations	Analytics	Privacy
YouTube	YouTube editor	Yes	Google analytics	Videos and channels can be private.
Vimeo	No	No	Simple statistics	Videos can be private.
WeVideo	Editor	No	No	Yes
MoViE	Limited editing	Yes	Google analytics available for research purposes.	Yes
TeacherTube	No	No	No	Videos can be private.

Table 3. Grouping, content, price, and pedagogical concepts

	Groups	Content	Price	Pedagogical Concepts
YouTube	Channels	Everything/not school content only.	Free	No
Vimeo	Channels, albums	Everything/not school content only; channel owner can set the content for all or mature audiences.	Basic is free.	No
WeVideo	Projects	Everything/not school content only.	Basic is free.	No
MoViE	Groups	School-related content.	Free	GSP, VIL, questionnaires.
TeacheTube	Groups, channels, collections.	School-related content.	Free, advertisements	Lessons, questionnaires, classroom management, badges.

Table 4. Examples of digital video stories

Topic	Exercise	Example
Local history	Create a video story of a historical event in your own school, neighborhood, or city.	• A drama of Greek mythology. • An interview of a local Nobel prize winner. • A paper doll animation of the immigration of Europeans to the USA in 1820-1930
Influencing opinion	Create a video story of an issue or cause for which you would like to raise awareness.	• A drama of homelessness. • The use of animals in testing cosmetics.
Chemistry lab	Document the chemical experiment you are doing.	Close-up video clips of a chemical process (for example, how to create a superball using polymerization).
Physics	Create video clips of a physical phenomena.	Video clips of circular motion (the clips can be recorded outside the classroom).
Mathematics	Do your own Khan Academy video.	Create a video of a solution of a mathematical exercise using a pen and paper.
Biology	Study the circulation of water.	A narrated PowerPoint presentation with video clips recorded at the local lake or river.

SOLUTIONS AND RECOMMENDATIONS

Niemi and Multisilta (2015) studied the methods of engagement in collaborative video work for learning. Based on these authors' findings, students can become highly engaged in a learning activity through digital storytelling. In addition, engagement can be defined by two components: an emotional aspect (joy of learning or fun) and a commitment to hard work. Using digital video storytelling, students learn several 21st century skills when producing learner-driven content. In addition, working in groups is important for student motivation and enthusiasm. The digital environment was shown to be the best predictor of

both engagement and learning outcomes (Niemi & Multisilta, 2015).

The findings provide evidence that students enjoy creating digital videos. They work actively, seeking new knowledge and constructing their videos using different information sources. A general observation is that they also self-evaluate their work. It was observed that commenting on others' work was difficult for students. Giving and receiving feedback is not self-evident. Although students find it interesting to watch others' videos, making active comments is still fairly uncommon.

For the Video Inquiry Project (VIP), design-based research and development was conducted to establish a broadly scalable approach for K-12 learners and teachers to capture mobile video

recordings of events and phenomena that prompt questions that can serve as a basis for inquiries and collaborative learning in the STEM disciplines (Multisilta et al., 2014). During the VIP project, it was found that digital media, including video, can provide an important nexus attracting joint attention of peer learners or learners-and-a-teacher, catalyzing learning conversations in-the-moment. Video can also play an important bridging function, connecting and spawning learning events across settings (including school and home) and generating discussions about math, science, and engineering. These conversations provide opportunities to ask questions, express confusion, share perspectives, and provide explanations. Learner-generated videos also provide a sense of agency and ownership that is engaging.

There are several kinds of tools that can be used for digital video storytelling. Some tools can be used only for editing the video content while others have capabilities to share the video to others and collaborate with other learners. Most of the tools are available on the web, but have not been designed to be used for learning. In addition, there has been a wide discussion on if schools should block the access to YouTube and similar sites (Storm, 2012) because of the availability of distracting content. It should be on the teacher's responsibility and right to select a pedagogical model and tools and apply them in the classroom in a meaningful way. So far, global sharing pedagogy and video inquiry learning models have been illustrated to provide engagement for STEM learning.

FUTURE RESEARCH DIRECTIONS

There is a growing need to seek new methods to understand the learning process in digital learning environments. Digital learning environments can be designed so that they collect various data from users while they use the environment and interact with other learners; however, these data are often either not available or not very useful for teach-

ers. Learning-related user data could be utilized so that they provide support in understanding the learning process in a much deeper way. In order to do so, learning data should be collected and analyzed in relation to the content. The data should be visualized so that teachers could immediately understand their meaning in relation to the class and the lesson. New methods and tools for collecting, analyzing, and visualizing learning analytics data in inquiry-video-based learning are needed.

General analytics tools, such as Google Analytics, can be used to evaluate the use of a digital video storytelling service, but it only provides largely superficial information regarding the learning process. Generic data collected by the existing analytics systems include web navigation data, such as page hits, the number of visitors to a page, and the time spent on the page. By utilizing data from the content and from the learning activities, teachers could understand the learning process in a more meaningful way.

CONCLUSION

By using digital video storytelling, students can engage in a variety of learning activities that could also be done outside the classroom, at home, at a sports field, or at a museum. Students already use videos often in their free time, but schools are not utilizing the expertise students have in publishing their lives in a digital space. By utilizing digital video storytelling as a tool in a classroom, students learn twenty-first century skills, such as creativity, problem solving, and collaboration.

Students need teachers' guidance in creating the manuscript and planning the video shooting locations. The stories can be dramas, plays, animations, narrated presentations, documentaries, and interviews. In the stories, the student's own voice is strong.

In the chapter, five video sharing sites were compared based on a set of requirements that are typical for classroom use. All the tools were available on the Internet. None of the tools reviewed

will support all classroom requirements. In the end, it is up to the teacher to decide the best tool for the particular case in the classroom.

REFERENCES

- Bidwell, N. J., Reitmaier, T., Marsden, G., & Hansen, S. (2010). Designing with mobile digital storytelling in rural Africa. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1593–1602). New York, NY: ACM. doi:10.1145/1753326.1753564
- Correia, N., Alves, L., Correia, H., Romero, L., Morgado, C., Soares, L.,... Jorge, J. A. (2005, May). InStory: a system for mobile information access, storytelling and gaming activities in physical spaces. In *Proceedings of the 2005 ACM SIGCHI International Conference on Advances in computer entertainment technology* (pp. 102–109). New York, NY: ACM. doi:10.1145/1178477.1178491
- de Jong, T., & van Joolingen, W. R. (2008). Model-facilitated learning. In J. M. Spector, M. D. Merrill, J. Van Merriënboer, & M. P. Driscoll (Eds.), *Handbook of Research on Educational Communications and Technology* (pp. 457–468). New York, NY: Lawrence Erlbaum Associates.
- Engström, A., Esbjörnsson, M., & Juhlin, O. (2008, September). Mobile collaborative live video mixing. In *Proceedings of the 10th international conference on Human computer interaction with mobile devices and services* (pp. 157–166). New York, NY: ACM.
- Kearney, M. (2011). A learning design for student-generated digital storytelling. *Learning, Media and Technology*, 36(2), 169–188. doi:10.1080/17439884.2011.553623
- Kiili, K., Multisilta, J., Suominen, M., & Ketamo, H. (2010). Learning experiences of mobile social media. *Int. J. Mobile Learning and Organisation*, 4(4), 346–359. doi:10.1504/IJML.2010.037533
- Ladeira, I., Marsden, G., & Green, L. (2011). Designing interactive storytelling: a virtual environment for personal experience narratives. In *Human-Computer Interaction—INTERACT 2011* (pp. 430–437). Berlin: Springer. doi:10.1007/978-3-642-23771-3_32
- Lewis, A., & Smith, D. (1993). Defining higher order thinking. *Theory into Practice*, 32(3), 131–137. doi:10.1080/00405849309543588
- Multisilta, J., Kallunki, V., Ojalainen, J., Penttilä, J., Liu, A., Eduard, K., & Pea, R. (2014). *Engagement in Inquiry-Based Learning with Mobile Devices*. Paper presented at the Nordic Research Symposium on Science Education NFSUN 2014, Helsinki, Finland.
- Multisilta, J., & Mäenpää, M. (2008). Mobile video stories. In *Proceedings of the 3rd International Conference on Digital interactive Media in Entertainment and Arts, DIMEA '08* (pp. 401–406). New York, NY: ACM. Retrieved March, 16, 2016, from <http://doi.acm.org/10.1145/1413634.1413705>
- Niemi, H., Harju, V., Vivitsou, M., Viitanen, K., Multisilta, J., & Kuokkanen, A. (2014). Digital Storytelling for 21st-Century Skills in Virtual Learning Environments. *Creative Education*, 5(09), 657–671. doi:10.4236/ce.2014.59078
- Niemi, H., & Multisilta, J. (2014). Global is becoming everywhere: global sharing pedagogy. In H. Niemi, J. Multisilta, L. Lipponen, & M. Vivitsou (Eds.), *Finnish Innovations and Technologies in Schools: Towards New Ecosystems of Learning* (pp. 35–48). Rotterdam: Sense Publishers. doi:10.1007/978-94-6209-749-0_3
- Niemi, H., & Multisilta, J. (2015). Digital story telling promoting 21st century skills and students' engagement. *Technology, Pedagogy and Education*. doi:10.1080/1475939X.2015.1074610
- Östman, S. (2015). “Millaisen päivityksen tästä sais?” Elämänjulkaisijuuden kulttuurin omaksuminen. Jyväskylä: Nykyluktuurin tutkimuskeskus. (In Finnish)

Penttilä, J. S. M., Kallunki, V. A. J., Niemi, H. M. & Multisilta, J. A. (2016) A Structured Inquiry into a Digital Story: Primary School Students Report the Making of a Superball. *International journal of mobile and blended learning (IJMBL)*, 8(3), pp. 19-34

Piwek, L., & Joinson, A. (2016). What do they snapchat about? Patterns of use in time-limited instant messaging service. *Computers in Human Behavior*, 54, 358–367. doi:10.1016/j.chb.2015.08.026

Prensky, M. (2011, Aug-Sep). Khan Academy. *Educational Technology*. Retrieved March, 16, 2016, from http://www.marcprensky.com/writing/Prensky-Khan_Academy-EdTech-Jul-Aug2011.pdf

Storm, S. (2012). YouTube Subtracts Racy and Raucous to Add a Teaching Tool. *The New York Times*. Retrieved March, 16, 2016, from http://www.nytimes.com/2012/03/10/education/youtube-finds-a-way-off-schools-banned-list.html?_r=0

Talbert, R. (2012). The trouble with Khan Academy. *The Chronicle*. Retrieved March, 16, 2016, from <http://chronicle.com/blognetwork/castingoutnines/2012/07/03/the-trouble-with-khan-academy/>

Thompson, C. (2011, July). How Khan Academy is changing the rules of education. *Wired Digital*. Retrieved March, 16, 2016, from http://www.wired.com/2011/07/ff_khan/

Tuomi, P., & Multisilta, J. (2010). MoViE: Experiences and attitudes—Learning with a mobile social video application. *Digital Culture & Education*, 2(2), 165–189.

Vaucelle, C., & Davenport, G. (2004). A system to compose movies for cross-cultural storytelling: textable movie. In *Technologies for Interactive Digital Storytelling and Entertainment* (pp. 126–131). Berlin: Springer. doi:10.1007/978-3-540-27797-2_17

Vygotsky, L. (1978). *Mind in society* (M. Cole & V. John-Steiner, Eds.). Cambridge, MA: The MIT Press.

Wolf, K. D., & Rummler, K. (2011). Mobile learning with videos in online communities: the example of draufhaber.tv. *MedienPädagogik*, 19, 1-13. Retrieved March, 3, 2016, from <http://www.medienpaed.com/19/#wolf1105>

Yang, Y. F. D. (2012). Multimodal composing in digital storytelling. *Computers and Composition*, 29(3), 221–238. doi:10.1016/j.comp-com.2012.07.001

Zsombori, V., Frantzis, M., Guimaraes, R. L., Ursu, M. F., Cesar, P., Kegel, I., & Bulterman, D. C. et al. (2011, June). Automatic generation of video narratives from shared UGC. In *Proceedings of the 22nd ACM conference on Hypertext and hypermedia* (pp. 325-334). New York, NY: ACM. doi:10.1145/1995966.1996009

ADDITIONAL READING

Banchi, H., & Bell, R. (2008). The many levels of inquiry. *Science and Children*, 46(2), 26–29.

Frazel, M. (2010). *Digital storytelling guide for educators*. Eugene, Oregon: International Society for Technology in Education.

Hakkarainen, P. (2011). Promoting meaningful learning through video production-supported PBL. *Interdisciplinary Journal of Problem-Based Learning*, 5(1), 34–53. doi:10.7771/1541-5015.1217

Harju, V., Viitanen, K., & Vivitsou, M. (2014). Digital storytelling in Finnish schools. In H. Niemi, J. Multisilta, L. Lipponen, & M. Vivitsou (Eds.), *Finnish innovations and technologies in schools: Towards new ecosystems of learning* (pp. 49–56). Rotterdam: Sense Publishers.

Kearney, M. (2011). A learning design for student-generated digital storytelling. *Learning, Media and Technology*, 36(2), 169–188. doi:10.1080/17439884.2011.553623

Lambert, J. (2013). *Digital storytelling: Capturing lives, creating community*. New York, NY: Routledge.

McGee, P. (2015). *The instructional value of digital storytelling: Higher education, professional, and adult learning settings*. New York, NY: Routledge.

Ohler, J. B. (2013). *Digital storytelling in the classroom. New media pathways to literacy, learning and creativity*. Thousand Oaks, CA: Corwin. doi:10.4135/9781452277479

KEY TERMS AND DEFINITIONS

Constructivist Learning Model: Learners actively create knowledge using activities that support knowledge construction.

Digital Video Storytelling: Learning activities that involve the creation and the use of digital video.

Global Sharing Pedagogy: Learning is seen as a mediated activity with tools, signs, and social interaction. Global sharing pedagogy has four mediators of learning: 1) learner-driven knowledge and skills creation, 2) collaboration, 3) networking, and 4) digital media competencies and literacies. The mediators contribute to the learners' engagement in a learning activity.

Inquiry Learning: Learning activities where learners search and construct new knowledge by exploring the world or the phenomena, by asking questions, making hypothesis, and testing the hypothesis.

Life Publishing: The use of social media services in which the users are sharing events and moments from their daily lives.

Twenty-First Century Skills: A set of knowledge and skills that are or will be important to succeed in the future world. Examples of twenty-first century skills include creativity, communication, collaboration, digital literacy, and problem-solving skills.

Video Inquiry Learning: Learning is based on the investigation of questions, scenarios, or problems using videos that students create and share using digital video storytelling tools. The videos prompt questions and serve as a basis for inquiries and collaborative learning.

The Use of Podcasting/ Vodcasting in Education

E**Athanasios T. Stavrianos***2nd Technical Vocational School of Xanthi, Greece***Apostolos Syropoulos***Greek Molecular Computing Group, Greece*

INTRODUCTION

The terms podcasting and vroadcasting refer to automatically downloadable audio and video files. Typically a podcast is an MP3 file whereas a vodcast can be any popular compressed video file. In 2004 the term podcasting was first mentioned in an article in the newspaper *The Guardian* (Hammersley, 2004). The term podcasting derives from iPod, the device that was first used to download and play podcasts. The inventors of the technology are Dave Winer and Adam Curry (Brown & Green, 2007). At that time, Winer was a software developer and an RSS evangelist while Curry was an MTV vj. Rich Site Summary, or just RSS, is a format for delivering regularly changing web content. Initially, podcasting was used for personal entertainment or information but soon it became clear that it could be used in education. Since its introduction the technology became very popular and this can be seen by the number of downloads. In April of 2006 ten million podcasts were downloaded while in November of the same year 17 million podcasts were downloaded.

The terms podcasting and vroadcasting (the vod part comes from *Video On Demand*) refer to a process. In particular, when there is an event, one has to capture the song, the interview, etc. The result can be either an audio file or a video file. Today's video capturing devices can produce high definition video that is stored using a compressed codec so there is no need to re-encode the video, something that was quite common in the early

days. Then, one had to post this audio or video file to a web site or a blog and using an RSS (Rich Site Summary) envelope (RSS is a format for delivering regularly changing web content). Obviously, the author had to inform people who might be interested in the new content. This is done automatically if people subscribe to an RSS feed. Nowadays one can use her smartphone to read the RSS feed and then to download the new content.

Today one can use a smartphone to record a video file or an audio file. Also, one can use a digital camera to record a video file. When one prepares a podcast, then it is recommended to use the "Audacity" open source software for audio editing. On the other hand, the open source project "pitivi" can be used to edit video files.

BACKGROUND

Podcasting was primarily used in tertiary education to make lectures available to students in order to clarify difficult parts and emphasize important ones. Later on, podcasts were replaced by vodcasts. Currently, there are four kinds of vodcasts that are used in education: lecture-based, enhanced, supplementary, and worked examples (Kay, 2012). A lecture-based or "substitutional" vodcast is a recording of an entire lecture. Thus students can experience what happened in the lecture hall without actually being physically present. An enhanced vodcast is video footage of a slideshow (e.g., Powerpoint or Beamer presentations) that is

presented with an audio explanation. Supplementary vodcasts are designed to augment the teaching and learning of some courses and may include administrative support, real-world demonstrations, summaries of lectures or textbook chapters, or additional material designed to broaden or deepen student understanding.

There are also other ways to classify vodcasts. For example, depending on whether a vodcast is offered in segments or not, one can talk about segmented vodcasts or non-segmented vodcasts, respectively. In addition, the pedagogical strategy can be used to categorize vodcasts. In particular, there are three different teaching approaches. The first is called receptive viewing and includes vodcasts to be viewed by students in a passive manner (i.e., like watching a movie). There are problem-solving vodcasts that are designed to explain and help students in learning how to solve problems and exercises related to their courses. Naturally, such vodcasts are useful for people who study science, mathematics, or engineering. A third category includes vodcasts that are created by students for students.

Although vodcasts seem to be quite popular today, there are a few questions related to their use in education. The first question is whether students are ready for this technology and the second question is whether this technology is actually useful. The first question has been tackled by (Walls, Kucsera, Walker, Acee, McVaugh, & Robinson, 2010) among others. First we need to note that today most if not all students own laptop or desktop computers and smartphones, which can be used to listen to music, to watch videos, take pictures, shoot videos, and so on. Thus many of the devices of the past (e.g., iPods, mp3 players, pocket digital cameras, and so on) have been replaced by smart-phones. This simply means that all students have the potential to create, download, and watch vodcasts. It has been suggested that vodcasting can improve student learning outcomes. This suggestion is largely based on Mayer's cognitive theory of multimedia learning (Mayer, 2001). According to this theory

“an individual's information processing system includes separate cognitive channels to process visual/pictorial and auditory/verbal stimuli; in this respect, learning is obtained by integrating information between such channels” (Walls, Kucsera, Walker, Acee, McVaugh, & Robinson, 2010). In different words, there is a limit to what a learner can achieve. Thus when a learner is presented with a large amount of pictures, images, talks, sounds, etc., hence she might fail to comprehend most of this information. Not so surprisingly, vodcasts can solve this problem! How? Each student will have the opportunity to watch vodcasts in the comfort of their room or in any other place that may suit a student. More importantly, students can control the speed and pace by which they will watch specific content.

Of course vodcasting is not a panacea and so it has certain limitations. For example, when students are already using a variety of resources in their studies (e.g., text-books, lectures notes, etc.), then adding one more kind or resource may create some sort of cognitive overload. Although students own smart-phones and other related technological “gadgets”, still they use them for entertainment and not for their studies. Therefore, it is not obvious that students would consider using their smartphones in their studies, particularly in institutions that have not used vodcasts in their educational resources. It has also been argued that the use of vodcasting may provide a justification and excuse for students to skip classes. However, some studies have revealed that students do not consider a vodcast as a substitute for attending a lecture. After all, one cannot ask a vodcast questions!

According to findings provided by (Walls, Kucsera, Walker, Acee, McVaugh, & Robinson, 2010) students do not find vodcasts particularly useful in their studies. They think that supplementary vodcasts contribute something in their learning. Also, students utilize vodcasts in rather different ways and in rather different circumstances. According to this study, they might utilize vodcasts during trips, while eating or exercising, and while study-

ing. Based on these findings one could conclude that students are not particularly ready. However, students become ready when they realize that a particular technology can be really beneficial in their studies.

A study that is documented by (Fernandez, Simo, & Sallan, 2009) examined the usefulness of vodcasting (the authors use the term video podcasting which has been substituted by vodcasting). Based on a number of principles of good practices aiming to improve students' learning process in undergraduate studies, the authors examined if these are achieved with the use of vodcasting. Briefly, these principles are:

1. Active learning is more effective than passing learning;
2. Learning demands students to be focused and aware of the importance of what they study;
3. Students learn more effectively when they have reasonable and positive goals that coincide with the goals of the instructor;
4. New knowledge must be connected to previous knowledge;
5. Information must be organized in personally meaningful ways;
6. Students need feedback to check their learning progress;
7. Unlearning what has been learned is more difficult than learning new things;
8. The way students are assessed affects the way they study;
9. Mastering a skill or a body of knowledge takes a lot of time and energy;
10. Learning to apply known ideas to new contexts is quite difficult;
11. High expectations may lead to great achievements;
12. Teachers have to balance levels of intellectual challenge and instructional support;
13. Motivation to learn is never steady; and
14. Teachers and students must always interact.

Students that participated in this study hoped that vodcasts would render attending classes useless. However, this was not the case as vodcasts cannot replace a textbook, which is something these students learned soon after the experiment started. Of course, all students found out that vodcasts can be used in any place and at any time, something that is particularly convenient. Most students used vodcasts in order to get prepared for new lectures. This way they knew the difficult parts of the lecture so they could prepare questions in order to clarify things. In general, this experiment showed that vodcasts are particularly useful.

So far we discussed the use of vodcasts in tertiary education. However, their usefulness in secondary education has also been studied (Coutinho & Rocha, 2010). In this study the authors present the "Geomcasting" project whose goal was the creation of vodcasts and *screencasts* (i.e., video made of screenshots that is accompanied by narration) by students so to help other students to pass the descriptive geometry class. The students involved created content for different parts of the class and the resulting content was uploaded to a blog site for future use. The authors of the study noticed that students were quite happy with their participation in the project and also those that used the content in their own studies found it particularly useful.

THE EZCAST CASE

So far we have given a general presentation of the technology and its use in education. In passing we mentioned how one can create content and make it available but we did not say anything about modern platforms that really facilitate the creation and sharing processes of vodcasts. In October of 2015, the first named author attended a presentation that took place in an auditorium of the Université libre de Bruxelles (ULB).¹ The presentation was about innovative educational technologies that the

teaching staff of the university developed and is actually using. In particular, the attendees were introduced to the *EZcast* (pronounced *Easy Cast*) web application that allows lecturers or teachers to share media contents with an audience. In addition, it was demonstrated how university students are using and interacting with the platform. Also, they had the chance to watch sample videos and learn about the hardware infrastructure used to record, manage, and upload multimedia content. Naturally, the presentation did not give the full details of every aspect of the project, nevertheless, it was enough for most participants to start studying all relevant details of the platform.

The Development Process

EZcast (EZcast wiki, 2014) along with its front-office component EZplayer (Roland, 2013), was officially launched by the Université libre de Bruxelles at the beginning of the academic year 2013-2014 and was made available to all students. The project evolves rapidly and this means that new releases will be available soon.

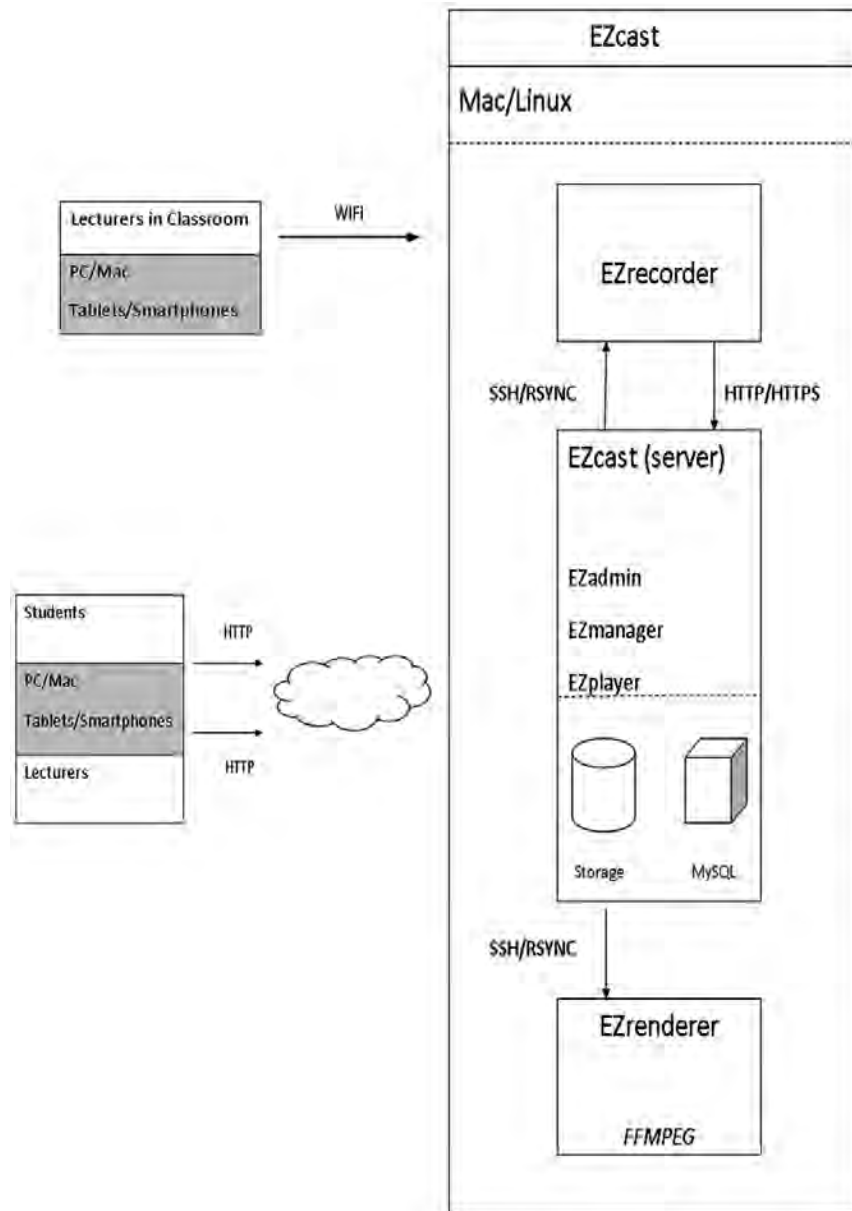
The implementation principles of the EZcast platform are a mixture of a design-based research approach (Collective, 2003) and a systemic user-centric approach (Roland, 2013). Design-based research aims at improving the educational practices using methods such as iterative cycles of analysis, design, enactment and redesign, constant collaboration between researchers and practitioners (Wang & Hannafin, 2005) leading to the production of a device (in this case the EZcast platform) and at the same time the introduction of new principles and theories. Based on this theoretical framework, ULB conducted a survey for three years collecting quantitative (5399 students answered questionnaires) and qualitative data (52 students were interviewed). The survey measured the way students use a vodcasting tool in an academic course and investigated the tool's degree of acceptance. The survey showed that vodcasting was not just a supporting tool for taking notes but also a medium of study giving options

to interact with the delivered knowledge (Roland, 2013). These results led to the formation of a team of people consisting of teachers, educational designers, developers, researchers, and teachers that acted as the α -testers of the first version of EZplayer that was developed in early 2013. 500 students acted as β -testers and their feedback was used to produce the second bug-free, *responsive-design*, and more interactive version of EZplayer.

A Deeper Look Into EZcast

EZcast is an open source web platform that allows teachers and lecturers to record lectures. Then these recordings can be made available to students via a web media player. Students have access to high quality educational multimedia content consisting not only of recorded audio and video but also by audio-video-synced slide presentations. On the other hand, EZcast allows educational material producers to easily create, edit and distribute it to targeted audience through their web browser. Each user of the platform is assigned a role. The two main roles supported by the platform (i.e., students and instructors) have access to EZplayer, an “*enriched web audiovisual player*” (Roland, 2013). This player has two main features: a set of audio visual handling tools that facilitate the easy transition from an AV stream to an “audio-slideshow” stream and the ability to accelerate or to decelerate a stream, sharing a link in a specific time, etc. In addition, students can add temporal bookmarks to specific frames of a video, thus they are able to tag them with a title, a description, and add keyword fields so to annotate or summarize the lecture that is recorded. Teachers have also the ability to add their own tags but students can modify these teacher-tags. In this way, every student can create lists of “official” and personal tags, which they can be used when performing searches so to retrieve educational material on demand. Moreover, since February 2015, an innovative *discussion function* has been added to the EZplayer. Users can now pinpoint specific framesets of the video thus creating a discussion

Figure 1. A general view of the EZcast Infrastructure
(Jansens & Wijns, 2015)



thread on which everyone can contribute with a comment and then vote for the best replies.

Although it is not our intention to fully describe all the features of each component of the main platform, still it is beneficial for the reader to have a general idea of the platform's infrastructure. Complete documentation, user tutorials, and installation guides concerning EZcast can be found at <http://ezcast.ulb.ac.be/>

Infrastructure

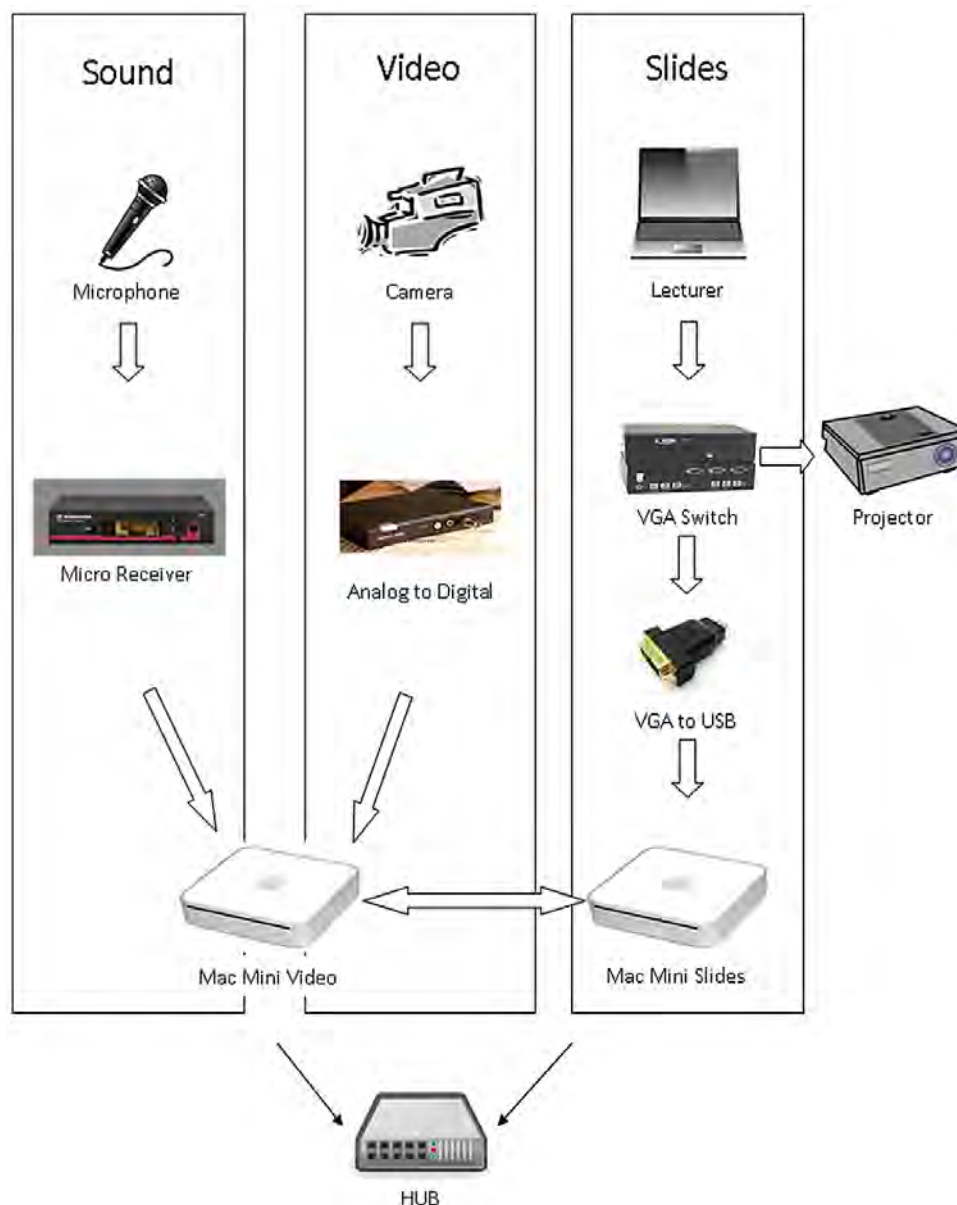
EZcast is a logical application framework composed of five distinct entities: EZrecorder, EZadmin, EZmanager, EZrenderer, and EZplayer (EZcast wiki, 2014). Different users can have access to four web interfaces: EZrecorder, EZadmin, EZmanager, EZplayer. As seen in Figure 1, the five entities interact with each other using

various protocols and links, but main user roles (i.e., instructors and students) mostly interact with EZplayer (i.e., watching, sharing and commenting audiovisual material), and EZrecorder (instructors recording and uploading their videos).

The applications EZadmin, EZmanager, and EZplayer actively communicate with each other so they must be installed on the same server, which

is supposed to run some flavor of UNIX or Linux. EZrecorder allows a user to record videos and publish them to a specific audience by directly communicating with EZmanager. The EZcast framework is freely available as an open source project under LGPL v.3 from 2 git repositories (EZcast² and EZrecorder³).

*Figure 2. Auditorium Hardware Infrastructure
(Jansens & Wijns, 2015)*



An Overview of the Five Entities

EZrecorder allows users to monitor their recordings automatically and autonomously. After a *secure authentication*, lecturers can choose the *album* they want to record, the title of the course, and available streaming formats: Audio-Video, audio-slideshow, or audio-video slideshow. The interface provides full recording management options by giving access to camera controls and predefined scenes. At the end of recording, the sequence can be published directly on the Internet or placed in a private album.

Media files sent to EZrecorder typically have been created in existing auditoriums of the institution equipped with the required hardware infrastructure. As seen in Figure 2, sound captured by a microphone is sent to an amplifier and then to a computer in order to synchronize it with the video. A permanently installed camera sends the captured video to a device that converts it from an analog to a digital format before sending it to the same computer (Mac mini Video). A VGA switch is used so to split the produced slideshows from a lecturer's computer. The output signal is transmitted simultaneously to the auditorium projector and a VGA-to-USB converter before it is fed to another computer (Mac mini Slide).

EZmanager Provides the ability to manage recorded albums, recordings from EzRecorder and submit videos from a user's computer. EzManager offers four modes of content publishing: EZplayer, RSS feeds, direct download, or embed code.

EZplayer This is an online audiovisual player allowing users (teachers and students) to interact directly with educational content. The available options are shown in Figure 2. In particular, transition through various types of streaming, video handlers, visual bookmarks, discussion threads at specific moments of the video, advanced search.

EZrenderer Processes the recordings and video submissions. Users can choose a jingle and titling to be incrustated before the video, and decide on the output aspect ratio and video quality.

EZadmin This is the web interface for users and course administration. For instance, it allows administrators to create new users and courses, add new renderers, enable or disable recorders in classrooms.

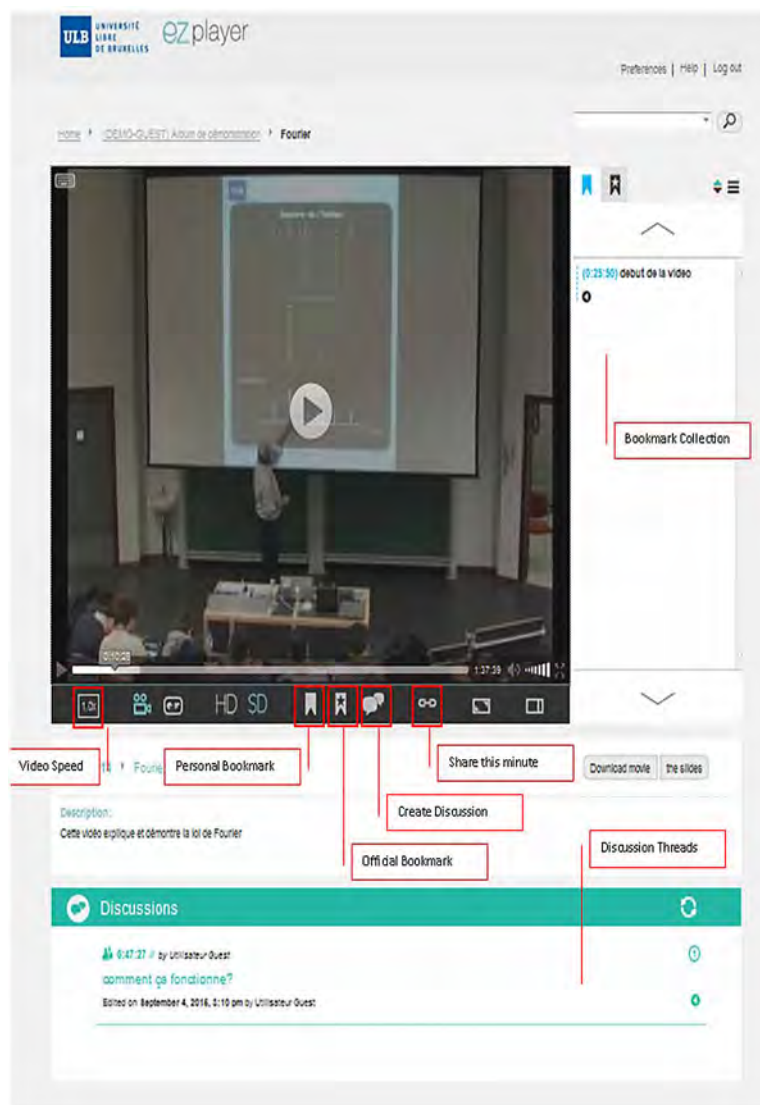
SOLUTIONS AND RECOMMENDATIONS

EZcast is a standalone application that must be installed on a server managed by the institution that will eventually run it. Also, it should only be available to targeted audience that belongs to the university in the broad sense of the word. On the other hand, there are ready-to-go educational web tools available to anyone, offering similar player features but with less recording and administrative options. Some of them are as follows:

- TEACHEM (<http://www.teachem.com/>)
- TEACHERS TUBE (<http://www.teacher-tube.com>)
- KHANACADEMY (<https://www.khanacademy.org/>)
- TED (<http://www.ted.com/>), etc.

The tools mentioned above belong to a wide class of tools that are particularly useful for distributing educational content. In addition, most of them offer class management options, student evaluation methods (quizzes, tests, and so on), communication with parents, etc. Even though they are not categorized as vodcasting applications, multimedia content (video, images, slideshows) can be embed or uploaded from educators and learners to the community. In most cases, users passively watch videos through a player with some basic commenting options available. Content-user interactivity features such as commenting on specific moments of the video, quiz questions on the video, and bookmarking framesets for future use, are currently being introduced to some of these applications (Teachem, Khnacademy).

*Figure 3. Basic features of EZplayer
(Jansens & Wijns, 2015)*



FUTURE RESEARCH DIRECTIONS

We have presented the idea of vodcasting, its general use in education, and a particular tool that can be used to create and distribute vodcasts. The next big thing in video is truly interactive video. Currently, some blu-ray movies provide some kind of interactivity. Thus, generating interactive educational videos may be the big thing in educational technology. Naturally, this would demand the creation of video-capturing, editing,

and uploading tools. One more related idea, would be to consider 3D videos, especially videos that describe lab activities.

CONCLUSION

We have presented scientific evidence to support that vodcasting can be beneficial to learners. In particular, they supplement their teaching material nicely. Students become easily quite familiar

with vodcasting tools and seem to give a vote of confidence on their ease of use in addition to their usefulness in their studies. By introducing to vodcasts certain features that are very common to Web 2.0 tools (e.g., some form of interactivity, the ability to comment certain framesets, bookmarking abilities, alteration between vodcasts and podcasts) has turned them into a valuable learning tool. So vodcasting has become the new interactive, note-taking, knowledge acquisition, on demand tool.

EZcast is a new and very promising platform that allows users to create, edit, and upload vodcasts. This tool and other similar tools should be used daily in the educational process.

REFERENCES

- Brown, A., & Green, T. (2007). Video Podcasting in Perspective: The History, Technology, Aesthetics, and Instructional Uses of a New Medium. *Journal of Educational Technology Systems* September, 36(1), 3-17.
- Collective, T. D.-B. (2003). Design-Based Research: An Emerging Paradigm for Educational Inquiry. *Educational Researcher*, 32(1), 5-8. doi:10.3102/0013189X032001005
- Coutinho, C. P., & Rocha, A. M. (2010). Screenshot and Vodcast: An Experience in Secondary Education. In D. Gibson & B. Dodge (Eds.), *SITE 2010-Society for Information Technology & Teacher Education International Conference* (pp. 1043-1050). Academic Press.
- EZcast wiki. (2014). Retrieved July 12, 2016, from <https://github.com/ulbpodcast/ezcast/wiki>
- Fernandez, V., Simo, P., & Sallan, J. M. (2009). Podcasting: A new technological tool to facilitate good practice in higher education. *Computers & Education*, 53(2), 385-392. doi:10.1016/j.compedu.2009.02.014
- Hammersley, B. (2004, February 12). *Audible revolution*. Retrieved July 15, 2016, from The Guardian: <https://www.theguardian.com/media/2004/feb/12/broadcasting.digitalmedia>
- Jansens, M., & Wijns, A. (2015). *EZcast infrastructures*. Retrieved July 12, 2016, from http://ezcast.ulb.ac.be/files/EZcast_Infrastructure_doc.pdf
- Kay, R. H. (2012). Exploring the use of video podcasts in education: A comprehensive review of the literature. *Computers in Human Behavior*, 28(3), 820-831. doi:10.1016/j.chb.2012.01.011
- Lonn, S., & Teasley, S. D. (2009). Podcasting in higher education: What are the implications for teaching and learning? *The Internet and Higher Education*, 12(2), 88-92. doi:10.1016/j.iheduc.2009.06.002
- Mayer, R. E. (2001). *Multimedia Learning*. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9781139164603
- Roland, N. (2013). *La recherche scientifique au service de l'expérience utilisateur: retour sur le processus de conception d'EZplayer*. Retrieved from <http://uptyv.univ-poitiers.fr/program/campus-europeen-d-ete-2013etnbux-designetnb-l-experience-utilisateur-au-service-des-apprentissagesetnb/video/3882/la-recherche-scientifique-au-service-de-l-experience-utilisateur-retour-sur-le-processus-de-conce>
- Walls, S. M., Kucsera, J. V., Walker, J. D., Acee, T. W., McVaugh, N. K., & Robinson, D. H. (2010). Podcasting in education: Are students as ready and eager as we think they are? *Computers & Education*, 54(2), 371-378. doi:10.1016/j.compedu.2009.08.018
- Wang, F., & Hannafin, M. J. (2005). Design-based research and technology-enhanced learning environments. *Educational Technology Research and Development*, 53(4), 5-23. doi:10.1007/BF02504682

ADDITIONAL READING

Braun, L. W. (2007). *Listen Up! Podcasting for Schools and Libraries*. Medford, NJ: Information Today.

Hubackova, S., & Golkova, D. (2014). Podcasting in Foreign Language Teaching. *Procedia: Social and Behavioral Sciences*, 143, 143–146. doi:10.1016/j.sbspro.2014.07.376

Morris, T., Tomasi, C., & Terra, E. (2008). *Podcasting For Dummies* (2nd ed.). Hoboken, NJ: Wiley Publishing.

Ng, W. (2015). *New Digital Technology in Education: Conceptualizing Professional Learning*. Cham, Switzerland: Springer International Publishing. doi:10.1007/978-3-319-05822-1

Zheng, L. (2016). Reflection on Design-Based Research: Challenges and Future Direction. In Y. Li, M. Chang, M. Kravcik, E. Popescu, & R. Huang (Eds.), *Kinshuk, et al., State-of-the-Art and Future Directions of Smart Learning* (pp. 293–296). Singapore: Springer. doi:10.1007/978-981-287-868-7_35

KEY TERMS AND DEFINITIONS

Educational Technology: Computer software and hardware that is used for educational purposes.

Embed Code: A block of HTML code that is embed in a target page pointing back to the source page and creating relevant content.

LGPL: The *GNU Lesser General Public License* and details how open source software can be freely copied, distributed and modified.

Podcast: An audio file that is posted to some site and which is automatically downloaded by subscribers of the site.

Responsive Design: A web design approach aimed at allowing web content to be viewed properly on all devices, using CSS3 and HTML.

Screencast: A kind of vodcast that consists of screenshots and narrated text.

Secure Authentication: The process of identifying an individual, usually based on a username and password.

Video File Format: A type of file format for storing digital video data on a computer system.

Vodcast: A video file that is posted to some site and which is automatically downloaded by subscribers of the site.

ENDNOTES

¹ The visit in the university was part of a short term exchange of pupils in Belgium. This was part of an EU funded Erasmus+ Project between 7 secondary schools from Austria, Belgium, Greece, Portugal, Romania, Slovenia, and Turkey.

² <https://github.com/ulbpodcast/ezcast>

³ <https://github.com/ulbpodcast/ezrecorder>

The Vital Importance of Faculty Presence in an Online Learning Environment



Ni Chang

Indiana University – South Bend, USA

INTRODUCTION

One of the instructional methods, which is clamored by students and which could arguably provide high quality educational opportunities, is faculty presence, as it makes possible the interaction between instructor and students and between students and students in a virtual learning environment (O'Reilly, 2009). Online instructors and academic administrators in higher education cannot simply hold an assumption that quality online courses or student learning could largely depend upon good Internet connectivity, high quality equipment, solid content knowledge of instructors (Welch & Napoleon, 2015), and beautifully designed online courses. The presence of an online instructor cannot ever be neglected or marginalized in online students' learning success. Therefore, highly significant is to address roles that instructors play in an online learning environment in order to underscore the crucial importance of faculty presence in the success of student learning.

BACKGROUND

The Paradigm Shift

An ever increasing number of colleges and universities are transferring courses from face-to-face (F2F) classroom meetings to online learning environments, as students seek out different sources for their educational experience (Welch & Napoleon, 2015). More than 6.7 million students

reportedly took at least one online course during the fall semester of 2011, 570,000 more students enrolled themselves in distance education than those in the previous year (Welch & Napoleon, 2015). Recently, more than 60% of administrators primarily in charge of academics at more than 2,800 colleges and universities in the United States made clear that shifting courses from F2F meetings to online was critical to their long-term strategies (Allen & Seaman, 2013).

With the ever expanding online education and given that online instruction differs distinctively from the traditional F2F instruction (Roman, Kelsey, & Lin, 2010), roles that an online instructor play in a virtual learning environment deserve a great deal of attention, as they underline the necessity of teacher presence in an online learning environment.

Hernández et al.'s (2010) study focused on the roles an instructor played in both e- and traditional learning environments. The researchers performed a comparative analysis of students' perceptions with 33 participants involved in a F2F traditional teaching while 23 students engaged in an online environment. Both of the groups taught by the same instructor. Hernández et al found students' perceptions varied regarding the roles that the instructor played in the F2F and online contexts. Generally, F2F group valued the instructor's role in the learning process more highly than the online group. The findings suggest online instructors ought to make additional efforts to better student learning.

reported that students of distance education classes performed poorly and some even were not able to complete online courses. Furthermore,

there seems to have higher dropout rates within online courses than F2F settings, which might be due to a lack of support from instructor and peers and which might be due to students feeling emotionally isolated (Artino & Jones, 2012; Dabbagh & Kitsantas, 2012). The reported findings offer a strong indication that online instructors need to meet the needs of students (Orso & Doolittle, 2012; Welch, Napoleon, Hill, & Rommell, 2014) by playing a variety of roles in online learning classrooms. Hence, in the paradigm shift, a crucial need is to understand roles that online instructors play in a virtual learning environment (Dennen, Darabi, & Smith, 2007) in order to highlight the paramount importance of faculty presence in an online learning environment.

Faculty Presence in Online Learning Environment

Faculty presence originates from. These researchers termed teacher presence as “teaching presence” and explained, “The concept of teaching presence is constitutively defined as having three categories – design and organization, facilitating discourse, and direct instruction” (p. 1). This explanation makes it clear that the presence of an online instructor embraces more than just answering emails and making announcements. With the presence of an online instructor, students would not feel like that they are situated in a “ghost town” (Online Learning Insight, 2012).

According to (O'Reilly (2009)), there are five interaction points, through which online instructors could connect or interact with students. These five interaction points are announcements, email, discussion forums, feedback summaries, and chat sessions. Roles online instructors play through these avenues could also reflect the online instructors' professional teaching dispositions, comprised of instructors' beliefs, values, and attitudes (Welch & Napoleon, 2015).

noted, “[D]ispositions have also been discussed as affective qualities including empathy,” which means that online instructors should make an effort

to know students and to motivate them to learn. Motivation is closely related to students' affect for learning, comprised of students' attitudes, beliefs, and values toward learning (McCroskey, Richmond, & McCroskey, 2006). Students' affective learning is inseparable from teacher presence governed by instructors' beliefs, values, and attitudes.

Students' Affective Learning

Instructors' beliefs, values, and attitudes also affect their verbal and non-verbal immediacy. Verbal immediacy is primarily concerned with ways instructors talk and facilitate student learning in the traditional classroom (Chang, 2011a), whereas nonverbal immediacy involves behaviors that are only observable to receivers or communicators, such as smile, “eye contact, body position, physical proximity, body movement” (Richmond, Gorham, & McCroskey, 1987 in Velez & Cano, 2008, p. 77). Instructor immediacy behaviors communicate positive relational affect (Velez & Cano, 2008): when there exist instructor immediacy behaviors, students feel close to their instructor (Christophel, 1990) and feel motivated to learn (Christophel, 1990; Velez & Cano, 2008). Students have a propensity to take satisfaction responding to questions and actively conceptualize and internalize knowledge (Krathwohl, Bloom, & Masia, 1964).

However, when instructor verbal and non-verbal immediacy are only shown through video lectures, motivation may not be as powerful as when there is an interaction between instructor and student. In a text-based teaching and learning environment, such as discussion asynchronous forums, non-verbal immediacy is non-existent, so is verbal. Then, if instructors are “hiding” from students, a sense of insecurity and overwhelming feeling arises, which could be detrimental to otherwise enthusiastic learning desires.

It is apparent that in an online learning environment, an instructor should make an effort to increase students' affect for learning (McCroskey et al., 2006). noted that learning not only is emo-

tionally oriented, but also cultivated by interacting with other people, in particular, with a course instructor. The presence of an instructor could make happen interactive communication between an instructor and students (Anderson et al., 2001). Therefore, addressing roles that instructors play is extremely necessary, as it could highlight the vital importance of teacher presence in an online learning environment.

MAIN FOCUS: ONLINE INSTRUCTORS' ROLES

Roles instructors play as facilitators for social instruction are critical in creating positive online learning environments and in promoting students' academic engagement (Cho & Cho, 2014) and are vitally important to students' learning satisfaction (Fedynich, Bradley, & Bradley, 2015). In the following text, the roles of an e-instructor are characterized horizontally by two categories: Pedagogical Efficacy, which chiefly focuses on

the promotion of students' cognition (8 roles in total) and Affective Promotion, which largely focuses on the promotion of students' affective learning (19 roles in total). In the category of Pedagogical Efficacy, there are two subtitles, namely, Knowledge Building (5 roles in total) and Instructional Preparation (2 roles in total). In the category of Affective Promotion, there are three subtitles, namely, Purposeful Commitment (9 roles in total), Purposeful Organization (4 roles in total) and Meaningful Management (6 roles in total).

These roles are also set apart vertically across the two categories of Pedagogical Efficacy and Affective Promotion by three distinct stages, namely, Course Development (7 roles in total), Course Delivery (18 roles in total), and Course Completion (2 roles in total) (see Table 1) address these roles one by one, the author will present them in the form of stages in order of Course Development, Course Delivery, and Course Completion. When each of the online instructor's roles is addressed in the following text, one of the subtitles along with the name of that particular role is placed

Table 1. Responsibilities and accountabilities of an E-instructor

Stage	Pedagogical Efficacy		Affective Promotion		
	Knowledge building	Instructional Preparation	Purposeful Commitment	Purposeful Organization	Meaningful Management
Course Development	<i>Inquiry</i>	<i>Decision Making</i>		<i>Course Conversion</i>	
	<i>Practice</i>	<i>Instructional Planning</i>		<i>Constructing</i>	
				<i>Redesigning and Revising</i>	
Course Delivery	<i>Learning Lecturing</i>		<i>Ongoing Assessment Ownership</i>	<i>Receiving</i>	<i>Assisting</i>
	<i>Individualized Instruction</i>		<i>Encouraging</i>		<i>Open Mindedness</i>
			<i>Monitoring Soliciting</i>		<i>Flexibility</i>
			<i>Reminding</i>		<i>Announcement</i>
			<i>Modifying</i>		<i>Acknowledging</i>
			<i>Clarifying</i>		<i>Patience</i>
Course Completion		<i>Informed Decision</i>	<i>Reflective Practice</i>		

in parentheses for easy reading. For example, (Knowledge Building/*Inquiry*) is referred to the role of inquiry an online instructor plays, which appears under Knowledge Building in the category of Pedagogical Efficacy.

COURSE DEVELOPMENT

With respect to teaching presence, according to Anderson et al. (2001), this stage cannot be omitted. During the Course Development stage, in the category of Pedagogical Efficacy, the instructor assumes four roles, ranging from those of gaining technological skills to those of getting the course ready for teaching. That is, the instructor is responsible for acquiring necessary and useful technological skills (Seaton & Schwier, 2014) (Knowledge Building /*Inquiry*) and familiarizes the learned skills through practice (Knowledge Building/*Practice*). During this time, the instructor also needs to engage in research to decide the content of a course plan (Instructional Preparation/*Decision Making*). According to, the role an online instructor plays is to select and filter information for student learning. Carefully determining what information needs to be offered to online students is supported by, who found from interviewing eight e-instructors that the information covered in virtual learning environments was not equivalent to that in F2F settings. One of the interviewees noted, “I went from about 13 individual classes or modules to about six modules” (Wilson et al., 2003). Followed by the decision making, the instructor needs to lay out a course plan appropriate for the students’ learning needs (Chang, 2009b, 2014; Yang & Cornelius, 2005) (Instructional Preparation/*Instructional Planning*). Online instructors need to facilitate well-considered discussion, which ought to be meaningful to students and which can maintain students’ desire for an in-depth study of concepts (Yang & Cornelius, 2005).

During this course preparation stage, students’ affective learning cannot be neglected. An e-instructor plays three roles in this process (Affective Promotion). The online instructor needs to keep in mind that designing an online course is by no means “curriculum conversion” (Roman et al., 2010), because this practice is insufficient to guide students in their acquisition of knowledge in a self-controlled manner (Chang, 2009c) (Purposeful Organization/*Course Conversion*). Learning through a virtual classroom seems intimidating to some and confusing to others. To minimize the degree of apprehension and anxiety, useful are an analysis of the learning environment and offering details of requirements and expectations (Chang, 2011a, 2011b). Such practices are intended to avoid the phenomenon that “an instructional design project may produce a theoretically sound but practically unable product” (Tessmer, 1990, p. 56) (Purposeful Organization/*Constructing*). The unable product could be viewed as a fabulously built online course, but in the end the course might make students feel isolated, frustrated, and intimidated due to a lack of timely and useful assistance (Chang, 2009b, 2014).

O’Reilly (2009) reported that teacher presence reflected through interactive communication between an instructor and students could lead to remove some feelings of disconnect by online students. While instructors are present and interactive, students tend to be encouraged to learn. When an online instructor is present, monitoring students’ learning, autonomy-supportive teaching takes place, which could, in turn, promote students’ motivation and inspire independent and critical thinking (Reeve, 2009). With teacher presence, if a course is repeated in a following semester, an instructor is able to easily engage in modifying *redesigning and revising* the course (Purposeful Organization) (Chang, 2009b, 2014) in hopes to further positive affect student learning. According to, nearly all cognitive oriented teaching and learning is involved with an affective component.

COURSE DELIVERY

It is through the stage of Course Delivery that interactions between an instructor and students vastly transpire. Anderson et al. (2001) found this stage could not be omitted. In other words, this stage should see much interaction between an instructor and students rather than the amount of interaction is minimized, as a great deal of apps and software, such as online video lectures, podcast, etc., is utilized to dominate course instruction.

Specifically, in this stage, the category of Pedagogical Efficacy, an instructor undertakes four roles and works as an academic guide to render needed and useful help to students (Chang, 2011a, 2011b). Along with interacting with students, the instructor needs to play a role of a learner, who continuously and consistently acquires knowledge relating to content areas and to technological skills (Knowledge Building/*Learning*). At the very beginning of this stage, *lecturing* (Knowledge Building/*Lecturing*) students to prepare them for online learning is of essence (Chang, 2009b, 2014; Yang & Cornelius, 2005). During the lecture, the instructor needs to be cautious not to drive students away by making a poor presentation, as some students may feel unsettled by the idea of independent learning (O'Reilly, 2009; Online Learning Insight, 2012). Therefore, needed is a clear elaboration of how to engage in independent study, how to manage one's time, and how to self-regulate one's learning, and how to navigate through an online course as the introduction of a course at the beginning of course delivery.

In the course delivery stage, the online instructor cannot "rest" and let the online course run by itself (Chang, 2011a). Nor can the instructor only conceptualize that his or her responsibility is to respond emails about logistics and/or students' questions. This stage should see teaching presence and social presence, which are expected to bring about cognitive presence (Anderson et al., 2001) and affective learning (Chang, 2011a, 2011b). Communication between an instructor and students plays an important role in shaping students'

views and their approach to learning (Armstrong, 2010). *Individualized instruction* (Knowledge Building) transpires as each communication is tailored to specific needs (Deutsch, 2013). Such practice could work like a fertile ground for individualized instruction, where an instructor could provide students with targeted informational and explanatory feedback (Chang, 2011a, 2011b, 2014; Rhode, 2008), encouraging them to think deeper. This process can provoke students' thinking and provide students with opportunities to advance their learning, by revising and relearning concepts (Chang, 2014; Nicol & Macfarlane-Dick, 2006). The researchers furthered that feedback that was informational and explanatory was low-stakes assessment, helping students see what deficiencies existed in relation to criteria. If receiving useful and meaningful feedback regularly, students can become less uncertain and anxious about their learning than without (O'Reilly, 2009).

The provision of feedback on individual assignments can, in turn, simultaneously provide useful lens for an instructor to know how to further improve the course design and teaching strategies (Chang, 2011a). This kind of interaction with students is *ongoing assessment* (Purposeful Commitment under the category of Affective Promotion), which is essential to high quality teaching and learning. Ongoing assessment allows for providing scaffolding in order to motivate students to construct knowledge. Pivotal is the instructor's appropriate guidance and support, which helps develop students' skills of self-responsibility (Yang & Cornelius, 2005), self-motivation, and a sense of autonomy (Reeve, 2009). These components are all crucial to students' affective learning (Chang, 2011a).

Students' affect for learning during the course delivery stage apparently plays an essential role in the success of their learning. It is thus not surprised to see that an online instructor plays 15 roles in this aspect, which are three times more than those in the Course Development. To cultivate students' affective learning, an e-instructor should be purposefully committed to employing various

strategies to encourage learners to be owners of their own learning (Reeve, 2009) (Purposeful Commitment/*Ownership* and Purposeful Commitment/*Encouraging*). Thus, students may want to actively participate in online discussions and activities (Roberson, 2013). *Monitoring* (Purposeful Commitment) students' learning and then appropriately *assisting* (Purposeful Commitment) them in their learning would help them ease the transition and boost their self-confidence in online learning (Chang et al., 2012). Furthermore, to gradually help students transit from the familiar to the unfamiliar learning environment, the instructor needs to frequently send out email, *reminding* (Purposeful Commitment) students of matters requiring their attention (Chang, 2009b, 2014). In the reciprocal interaction with students, it is fundamental for the instructor to understand that one student's question may be representative of others' and that emerging problems in the process of teaching and learning may become a potent opportunity for the instructor to reexamine the course design and instructional strategies. These can be the basis for the instructor to clarify (Purposeful Commitment/*Clarifying*) topics under discussion and to modify course content as well as ongoing approaches to teaching (Chang, 2009b, 2014) (Purposeful Commitment/*Modifying*).

To further set up an emotionally supportive learning environment, assessments of the effect of course design and delivery are continuous throughout each semester. The instructor purposefully organizes a course by evaluating it based on unexpected email messages and communications with students in various forms such as online discussions, dialogues, and chats. For example, in *receiving* email messages sent by students (Purposeful Organization), an instructor needs to read them carefully, as he or she may be able to learn which assignments are most or least helpful to students. In communicating with students, an instructor could also take advantage of every available opportunity to not only provoke students' thinking, but also solicit students' reactions toward online course delivery or arrangement

(Purposeful Organization/*Soliciting*). Information gained through such a process is conducive to the improvement of instruction.

In interacting with students, it is beneficial for an online instructor to let students know that assistance is readily available to them via email or phone calls (Meaningful Management/*Assisting*). In a geographically isolated learning setting, such an act of an instructor would make e-students feel comfortable, as it demonstrates that the instructor sincerely cares about their online learning (Chang, 2009b, 2014). To support and sustain students' affective learning, the instructor must also maintain an open mind in order to learn about and understand students' needs and learning levels (Meaningful Management/*Open Mindedness*).

During the Course Delivery stage, many unexpected events may occur. It requires the instructor to remain flexible and creative (Meaningful Management/*Flexible and Creative*) to appropriately deal with emergent matters and to promote student learning. In addition, the instructor also needs to utilize emails or Announcement in a course management system (CMS) to inform e-students of class-related information and business that is unrelated to course content (Meaningful Management/*Announcement*).

Giving appropriate and relevant praise could positively encourage and further motivate students to learn (Rhode, 2008). Therefore, the instructor needs to acknowledge the efforts made by students, as the acknowledgement can encourage their continuous endeavors and promote their intrinsic motivation (Meaningful Management/*acknowledging*) (Chang, 2009c). *Patience* plays an important role in online teaching as well (Meaningful Management). Sometimes, an explanation of an assignment needs to be repeated. Sometimes, one concept may require several elaborations. If students fail to understand points embedded in readings, the instructor should ask them to re-read required class materials and/or offer guidance, if needed (Chang, 2009b, 2014).

COURSE COMPLETION

During this stage, an instructor assumes two roles. One is concerned with decision making, which is in the category of Pedagogical Efficacy while the other is self-reflection, which is included in the category of Affective Promotion. Although Anderson et al. (2001) did not include this stage, I believe they are not only essential, but also crucial in the success of online instruction.

The first role is that an online instructor needs to discern and analyze the data collected through the entire semester so that he or she can make decisions about next semester's teaching and learning (*Instructional Preparation/Informed Decision*). The collected data may include students' email, students' feedback about a course, students' survey results, and informal conversations between an instructor and students. These resources may enable the instructor to seek answers to questions of course improvement. Some of these questions may be: What has been achieved in the past semester? What has been viewed as a failure or failures? What lessons should be drawn to improve future online instruction? The instructor answers these questions through *reflective practice* (Purposeful Commitment) to showcase his or her purposeful examination of performance in order to know how to improve instruction.

FUTURE RESEARCH DIRECTIONS

Technology has developed rapidly, so is distance education or e-learning (Margalina, De-Pablos-Heredero, & Botella, 2014). However, alarming is a higher drop-out rate of online learners than that of traditional learners (Rauscher & Cronje, 2005). In order to increase retention rates and to prevent from an increase in online learners' drop-out rate, noted technology should not be the target of blame for poor students' learning. How a course is arranged and designed and how frequently communication takes place between an instructor and students matter more than what technology can

do. Clark (1994) made an analogy: "... media are mere vehicles that deliver instruction, but do not influence learner achievement any more than the truck that delivers our groceries causes changes in our nutrition" (in Rauscher & Cronje, 2005, p. 107). In bringing about students' learning in good quality, online instructors should pay more attention to *how and why* technology is used than to what technology needs to be included in teaching and learning. That is, online instructors should be fully aware of how "to be there" for students while technology delivers a course, as what instructors do or do not do could indeed have an effect on the success or failure of student learning.

One aspect that determines students' learning success or failure is students' motivation to learn. Yet, in a virtual learning environment, the level of students' motivation to learn ought be higher than that in a F2F setting (. Unfortunately, it does not seem the case. ascribed the high dropout rates of e-learners to a lack of motivation. As such, technology alone cannot possibly turn around the situation. It definitely require the presence of an online instructor in a virtual learning environment to employ various strategies to increase students' motivation to learn, which poses pedagogical challenges to online instructors in transitioning from F2F to online settings (Park, Johnson, Vath, Kubiskey, & Fishman, 2013). It is always a challenge for instructors to know how to interact with online learners to build positive relationships with them (Roman et al., 2010). Therefore, it is necessary for colleges and universities to methodically devise a comprehensive plan that can aptly offer quality training and support (Deutsch, 2013; Park et al., 2013) for would-be-e-instructors and online instructors to enhance quality learning (Chang, 2011a; Roman et al., 2010). Good quality training enables online instructors not only to understand online instructional tools, but also develop knowledge as to what they could do to be present and how to attend to students' affective learning. Such training should be a professional development and a prerequisite for success in online teaching (Margalina et al., 2014; Roman et al., 2010). In this

way, faculty can develop knowledge and skills to seriously and carefully attend to students' affective learning (Chang, 2011a). Teaching and learning are not simply about delivering information and/or completing assigned tasks, but they are concerned about students' state of emotion in learning, which leads to their motivation to learn.

CONCLUSION

Online learning isolates learners and the instructor, leaving the community members feeling lonely and, sometimes, frustrated, which could vastly impede teaching and learning. To reduce the level of discomfort, the instructor must be "visible" (Chang, 2009b, 2014) by assuming various roles. Therefore, not only could an instructor promote students' cognition, but also attends to their affective learning. This article addresses the vital importance of faculty present in a virtual learning environment by focusing on various roles that an online instructor should play in an entire process of teaching and learning.

There are eight roles total in the category of Pedagogical Efficacy, which are intended to promote student learning. These roles range from assisting students in gaining technological skills to teaching them academic content. This category also demonstrates that the instructor is responsible for acquiring technological skills and becomes familiar with newly learned skills through practice.

There are 19 roles in the category of Affective Promotion—twice as many roles as in Pedagogical Efficacy. Affective learning has been recognized as more potent and influential than academic efforts in student learning (Rauscher & Cronje, 2005), as students' affective learning is just as critical a dimension of learning as cognition. The roles that an online instructor assumes in this category include interactive communication between an instructor and students, encouragement, acknowledgement, and other logistics. Attending to students' affective learning via every possible

avenue is likely to motivate students to learn and support their cognition, as students' emotions do affect their quality learning (Rauscher & Cronje, 2005); Instructors' scaffolding in a virtual learning environment is also essential to students' success of learning. In short, all these roles require an instructor to be present and visible in an online learning environment, as technology alone cannot promote students affect for learning.

REFERENCES

- Allen, I. E., & Seaman. (2013). *Changing course: Ten years of tracking online education in the United States*. Babson Survey Research Group.
- Anderson, T., Rourke, L., Garrison, D. R., & Archer, W. (2001). Assessing Teaching presence in a Computer Conference Environment. *Journal of Asynchronous Learning Networks*, 5(2), 1–17.
- Armstrong, D. (2010). *A qualitative study of undergraduate students' approaches, perceptions, and use of online tools*. Academic Press.
- Artino, A. R., & Jones, K. D. (2012). Exploring the complex relations between achievement emotions and self-regulated learning behaviors in online learning. *Internet and Higher Education*, 15, 170-175.
- Chang, N. (2009b). *Facilitating roles an instructor undertakes in a virtual learning environment* (2nd ed.). Hershey, PA: Information Science Reference.
- Chang, N. (2009c). Can students improve their learning with the use of an instructor's extensive feedback assessment process? *International Journal of Instructional Technology and Distance Learning*, 6(5), 32.
- Chang, N. (2011a). *Formative assessment and feedback with teacher immediacy behaviors in an e-text-based context*. Hershey, PA: IGI Global. doi:10.4018/978-1-61520-983-5.ch015

- Chang, N. (2011b). Pre-service teachers' views: How did e-feedback through assessment facilitate their learning? *Journal of Scholarship of Teaching and Learning*, 11(2), 16–33.
- Chang, N. (2014). *Roles of online instructors apt for students' cognitive and affective learning* (3rd ed.). Hershey, PA: Information Science Reference.
- Cho, M. H., & Cho, Y. (2014). Instructor scaffolding for interaction and students' academic engagement in online learning: Mediating role of perceived online class goal structures. *Internet and Higher Education*, 21, 25–30.
- Christophel, D. M. (1990). The relationships among teacher immediacy behavior, and student motivation, and learning. *Communication Education*, 39(4), 321–340. doi:10.1080/03634529009378813
- Dabbagh, N., & Kitsantas, A. (2012). Personal learning environments, social media, and self-regulated learning: A natural formula for connecting formal and informal learning. *Internet and Higher Education*, 15, 3–8.
- Dennen, V. P., Darabi, A., & Smith, L. J. (2007). Instructor-learner interaction in online courses: The relative perceived importance of particular instructor actions on performance and satisfaction. *Distance Education*, 28(1), 65–79. doi:10.1080/01587910701305319
- Deutsch, N. M. (2013). *Skills teaching online vs. face-to-face classes*. Retrieved from <http://blog.wiziq.com/online-teaching-skills>
- Fedynich, L., Bradley, K. S., & Bradley, J. (2015). Graduate students' perceptions of online learning. *Research in Higher Education Journal*, 27, 1–13.
- Hernández, A. B., Gorjup, M. T., & Cascón, R. (2010). The role of the instructor in business games: A comparison of face-to-face and online instruction. *International Journal of Training and Development*, 14(3), 169–179. doi:10.1111/j.1468-2419.2010.00350.x
- Krathwohl, D. R., Bloom, B. S., & Masia, B. B. (1964). *Taxonomy of educational objectives: Handbook II: Affective domain*. New York: David McKay Co.
- Margalina, V. M., De-Pablos-Heredero, C., & Botella, J. L. M. (2014). Achieving job satisfaction for instructors in e-learning: The relational coordination role. *International Journal of Human Capital and Information Technology Professionals*, 6(4), 64–79. doi:10.4018/IJHCITP.2015100104
- McCroskey, J. C., Richmond, V. P., & McCroskey, L. L. (2006). *The role of communication in instruction: The first three decades*. Mahwah, NJ: Lawrence Erlbaum.
- Morrison, D. (2012). *Strategies for online instructors: Understanding the needs of the online learner*. Retrieved from <http://onlinelearninginsights.wordpress.com/2012/08/20/strategies-for-online-instructors-understanding-the-needs-of-the-online-learner>
- Nicol, D., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199–218. doi:10.1080/03075070600572090
- O'Reilly, K. (2009). Faculty presence promotes quality of education in the online asynchronous classroom. *Contemporary Issues In Education Research*, 2(3), 53–58. doi:10.19030/cier.v2i3.1087
- Orso, D., & Doolittle, J. (2012). Instructor characteristics that affect online student success. *Faculty Focus*. Retrieved from <http://www.facultyfocus.com/articles/online-education/instructorcharacteristics-that-affect-online-student-success/>
- Park, G., Johnson, H., Vath, R., Kubiskey, B., & Fishman, B. (2013). Examining the roles of the facilitator in online and face-to-face professional development contexts. *Journal of Technology and Teacher Education*, 2(2), 225–245.

- Rauscher, W. J., & Cronje, J. C. (2005). Online with Krathwohl: Affective aspects of learning in an online environment. *SAJHE*, 19(3), 104–118.
- Reeve, J. (2009). Why teachers adopt a controlling motivating style toward students and how they can become more autonomy supportive. *Educational Psychologist*, 44(3), 159–175. doi:10.1080/00461520903028990
- Rhode, J. (2008). *Roles and responsibilities of the online instructor*. Retrieved from <http://www.slideshare.net/jrhode/roles-and-responsibilities-of-the-online-instructor>
- Rippe, C. (2009). Using rubrics to improve teaching, learning, and retention in distance education. *Faculty Focus*. Retrieved from <http://facultyfocus.com/wp-content/uploads/images/facultydev-in-distanced.pdf>
- Roberson, K. (2013). Teaching techniques that establish relevance, promote autonomy. *Faculty Focus*. Retrieved from <http://www.facultyfocus.com/articles/effective-teaching-strategies/motivating-students-with-teaching-techniques-that-establish-relevance-promote-autonomy/>
- Roman, T., Kelsey, K., & Lin, H. (2010). Enhancing online education through instructor skill development in higher education. *Online Journal of Distance Learning Administration*, 8(4).
- Seaton, J. X., & Schwier, R. (2014). An exploratory case study of online instructors: Factors associated with instructor engagement. *International Journal of E-Learning & Distance Education*, 29(1), 1–16.
- Tessmer, M. (1990). Environment analysis: A neglected stage of instructional design. *Educational Technology Research and Development*, 38(1), 55–64. doi:10.1007/BF02298248
- Velez, J. J., & Cano, J. (2008). The relationship between teacher immediacy and student motivation. *Journal of Agricultural Education*, 49(3), 76–86. doi:10.5032/jae.2008.03076
- Welch, A., & Napoleon, L. (2015). Professional teaching dispositions of online instructors: Why they matter. *Procedia: Social and Behavioral Sciences*, 171, 584–589. doi:10.1016/j.sbspro.2015.01.164
- Welch, A., Napoleon, L., Hill, B., & Rommell, E. (2014). Virtual teaching dispositions scale© (vtlds): A multidimensional instrument to assess teaching dispositions in virtual classrooms. *Journal of Online Learning and Teaching*, 10(3).
- Wilson, D., Varnhagen, S., Krupa, E., Kasprzak, S., Hunting, V., & Taylor, A. (2003). Instructors' adaptation to online graduate education in health promotion: A qualitative study. *Journal of Distance Education*, 18(2), 1–15.
- Yang, Y., & Cornelius, L. F. (2005). Preparing instructors for quality online instruction. *Online Journal of Distance Learning Administration*, 8(3).

ADDITIONAL READING

- Alvarez, I., Espasa, A., & Guasch, T. (2012). The value of feedback in improving collaborative writing in an online learning environment. *Studies in Higher Education*, 37(4), 387–400. doi:10.1080/03075079.2010.510182
- Henrie, C. R., Halverson, L. R., & Graham, C. R. (2015). Measuring student engagement in technology-mediated learning: A review. *Computers & Education*, 90, 36–53. doi:10.1016/j.compedu.2015.09.005
- Hung, W. C., & Jeng, I. (2013). Factors influencing future educational technologists intentions to participate in online teaching. *British Journal of Educational Technology*, 44(2), 255–272. doi:10.1111/j.1467-8535.2012.01294.x

Krause, U., & Stark, R. (2010). Reflection in example- and problem-based learning: Effects of reflection prompts, feedback and cooperative learning. *Evaluation and Research in Education*, 23(4), 255–272. doi:10.1080/09500790.2010.519024

KEY TERMS AND DEFINITIONS

Affective Promotion: Encompasses endeavors and strategies made by an e-instructor in fostering students' emotional involvement in e-learning and in setting up an emotionally supportive learning environment to facilitate student learning.

Instructional Preparation: Is related to avenues in which an e-instructor is engaged to make decisions based on information at hand as well as collected through previous experiences of working with students in order to help plan instruction suited to learners' needs.

Knowledge Building: Refers to an e-instructor, who keeps professionally up-to-date through self-development and learning alongside students and who attains technological knowledge and skills by attending relevant workshops and the frequent interaction with a computer.

Meaningful Management: Refers to an e-instructor who manages a course in ways that may help ease students' unnecessary frustration resulting from their being situated in a novel learning environment. This type of course management

aims to promote students' affective learning in the virtual classroom.

Pedagogical Efficacy: Refers to the growth and development of both faculty and students concerning academics-oriented knowledge and skills ranging from content-specific areas to technological skills through efforts exerted by an e-instructor.

Purposeful Commitment: Refers to an e-instructor who is committed to helping students become owners of their own learning by the instructor becoming visible through various means in the shared virtual classroom in order to support learning.

Purposeful Organization: Refers to an e-instructor who is committed to helping students become owners of their own learning, achieved when the instructor becomes visible through various means in the virtual classroom.

Reflective Practice: Refers to an e-instructor's consistent behaviors in assessing the course by an ongoing, even daily, basis as well as at the end of a semester in order to motivate learners to succeed in learning.

Teacher Presence: Refers to an instructor, who learns new technology, acts as an instructional developer, designs and organizes an online course prior to its commencement, who is there for students, scaffolds and facilitates learning, and promotes students' affect for learning during course delivery, and who evaluates and reflects on his or her performance to modify it for improvement.

Category E

Electrical Engineering

Mechanisms of Electrical Conductivity in Carbon Nanotubes and Graphene

E**Rafael Vargas-Bernal***Instituto Tecnológico Superior de Irapuato, Mexico*

INTRODUCTION

In the search for alternative materials to semiconductor materials used commonly in electronics such as silicon, germanium, gallium arsenide, gallium phosphide, etc., researchers around the world have been developing carbon-based materials with ideal electrical properties to operate with high efficiency in nanoelectronics. Carbon nanotubes (CNTs) and graphene represent two technological options for these innovative materials, which can be used either individually, or in composite or hybrid materials as electrical filler. They offer electrical properties such as high electrical conductivity and high dielectric permittivity, which can be tuned by synthesis, doping, functionalization, etc. These qualities can be exploited in applications such as interconnects, electronic devices such as field-effect transistors, batteries, fuel cells, supercapacitors (Yusoff, 2015), electrodes for touch screens (Zheng, 2015), flexible transparent memory circuits, materials for electrostatic discharge (ESD) and electromagnetic interference (EMI) shielding (Vargas-Bernal, 2015c), etc.

This chapter will review the most important electrical transport mechanisms associated with the electrical conductivity of carbon nanotubes and graphene, since these can be used in individual way or within composite or hybrid materials, with the aim of discovering the origin of their extraordinary electrical properties than have been used, are being used, and will be used in diverse technological applications. The effect of a set of technical variables related with electrical behavior of carbon nanotubes and graphene, and associated with the electrical conductivity such as band gap,

intrinsic mobility, percolation threshold, electrical conductivity, and dielectric permittivity, are also discussed.

BACKGROUND

Electrical conduction can be defined as the movement of electrical carriers through a transmission medium. A transmission medium is a material substance that transmits or guides through of itself electromagnetic waves. This movement of carriers generates an electrical current in response to an electrical field. Moreover, in each type of material, different mechanisms of electrical conduction are presented. For example, electrons are electrical carriers in metals, and the Ohm's law is the mathematical relationship used to determine the mathematical expression between the electrical current (I) and the applied potential difference (V) between a pair of ends of the material (Bird, 2014):

$$I = \frac{V}{R} = VG, \quad (1)$$

where R and G are electrical resistance and electrical conductance, respectively. Thus, one or more electrons from each atom can move freely within the metal, since they are loosely bound to the atom in the higher level of the valence band. These electrons are incorporated to the conduction band as electrical carriers due to the potential difference, and therefore, an electrical current is generated. An electrical current is a flow of electrical charge carried out regularly by moving electrons through a medium.

DOI: 10.4018/978-1-5225-2255-3.ch233

Electrical conductivity (σ) also called specific conductance can be defined as the ability of a material for conducting an electrical current. In three-dimensional conductor materials, the electrical conductance can be mathematically expressed as:

$$G = \frac{A}{\rho L} = \frac{Wt}{\rho L} = \frac{Wt\sigma}{L}, \quad (2)$$

where A is the cross-sectional area, L is the length, W is the width, t is the thickness, and, ρ and σ are electrical resistivity and electrical conductivity of the material, respectively. Two different types of electrical conductivities can be found in materials: surface conductivity and bulk conductivity. Surface conductivity or sheet conductance quantifies the electrical conductance of thin films with uniform thickness nominally. This represents the rate between the electrical conductivity of the material, and the thickness of the thin film. Therefore, it is mathematically expressed as:

$$G_s = \frac{t}{\rho} = \sigma t, \quad (3)$$

whose units are square per Ohm or Siemen square or denoted by sq/Ω or $\square/\Omega$ or $\text{S}\cdot\text{sq}$ or $\text{S}\cdot\square$, which is dimensionally equal to an Siemen. Bulk conductance, specific electrical conductance, or volume conductivity (σ) is expressed in units of Siemens per meter (S/m).

Materials can be electrically classified in accordance with their conductivities as conductive, static conductive, or static dissipative (Grady, 2011). The surface conductivity regimes for each are approximately greater than $10^{-4} \text{ sq}/\Omega$, 10^{-6} – $10^{-4} \text{ sq}/\Omega$, and 10^{-12} – $10^{-6} \text{ sq}/\Omega$, respectively. Moreover, the corresponding volume conductivity regimes are > 0.1 , 10^{-3} – 10^{-1} , and 10^{-9} – 10^{-3} S/m . Two applications can be identified in accordance with the value of conductivity: electromagnetic interference (EMI) shielding and electrostatic dissipation (ESD). EMI shielding uses materials with

high conductivity, while ESD requires materials with low conductivity.

In materials such as insulators and semiconductors, there is an energy range called forbidden band or band gap, where electron energy states cannot exist between the top of the valence band and the bottom of the conducting band.

MAIN FOCUS OF THE ARTICLE

Carbon nanotubes and graphene has a structure of conjugated system, where a system of connected p -orbitals with delocalized electrons in atoms, presents alternating single and multiple bonds, which in general may lower the overall energy of the system and increase stability. Two p -orbitals form a π -bond. A π -bond is a covalent chemical bond, where two lobes of one atomic orbital are overlapped to other two lobes of the other atomic orbital involved. Therefore, the π -electrons do not belong to a single bond or atom, but rather to a group of atoms (Jug, 2001). The conjugation can be viewed as the overlap of one p -orbital with another across a sigma bond. A simple model of the energy levels can be considered as a quantum mechanical problem of a one-dimensional particle, representing the movement of a π -electron along a long conjugated chain of carbon atoms as found in carbon nanotubes and/or graphene. In this model, the lowest possible absorption energy corresponds to the energy difference between the highest occupied molecular orbital (HOMO), and the lowest unoccupied molecular orbital (LUMO). Almost all electronic transitions in conjugated π -systems are carried from a bonding molecular orbital (MO) to an antibonding MO (π to π^*), but electrons from non-bonding Lone pairs (pair of valence electrons that are not shared with another atom) can also be promoted to a π -system MO (n to π^*) in charge-transfer complexes. A HOMO to LUMO transition is carried out by an electron if it is allowed by the selection rules for electromagnetic transitions. Thus, the electrical conductivity is guaranteed at using carbon nanotubes and/or

graphene in electronic devices. Carbon nanotubes can achieve electrical conductivities between 100 and 200×10^3 S/m, while graphene can achieve an electrical conductivity between 1,738 S/m and 100×10^6 S/m.

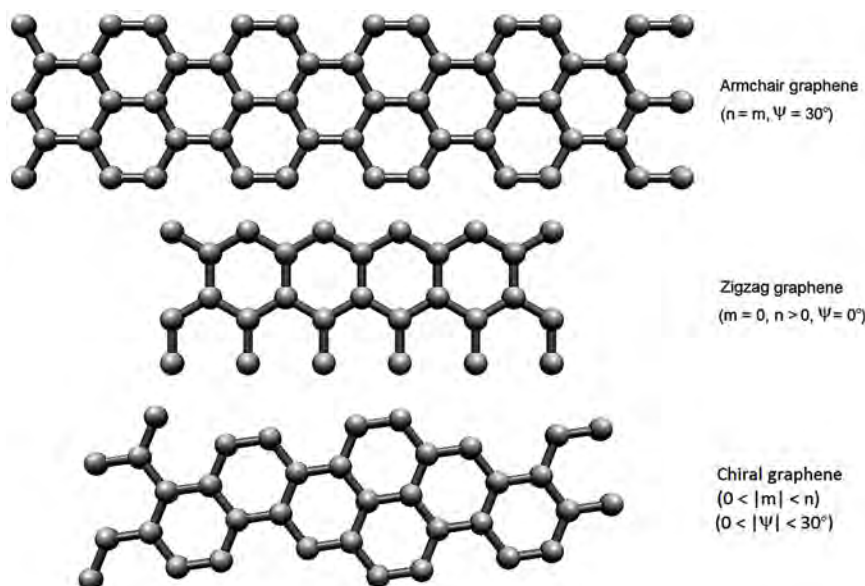
The hexagonal lattice, found in the carbon nanomaterials, possesses the longest mean free path of any known material, in the order of microns. Therefore, a ballistic transport is presented in these materials, since the distance that an electron can travel freely without bumping into anything, or having its path disrupted by scattering is large, which reduces their electrical resistance even at room temperature.

Graphene is an allotrope of carbon, formed by a simple sheet of graphite, which is one-atom thick; with carbon atoms arranged in a regular hexagonal pattern (see Figure 1). It has an extremely low weight, since, a graphene sheet with an area of one square meter; this material only weights 0.77 grams. As a semi-metallic material, this material is a semiconductor, and also it has high electron mobility at room temperature. Graphene has no bandgap, since its conduction and valence bands fulfill the Dirac points. It is possible to induce a

small bandgap in graphene by doping. A material without band gap can convert all wavelengths of light to electrons, with energy levels that even did not found in semiconductor materials, and thus, graphene is a wonderful candidate for use in photovoltaic cells. Graphene has high energy density and/or highest current density (a million times that of copper), that favors the storage capacity and rate of charging/discharging, which make it suitable for batteries and electrical supercapacitors. It has the highest intrinsic mobility (100 times more than silicon) between $15,000 \text{ cm}^2 \cdot \text{V}^{-1} \cdot \text{s}^{-1}$ and $200,000 \text{ cm}^2 \cdot \text{V}^{-1} \cdot \text{s}^{-1}$ nearly unaffected by the temperature (10K and 100K), a fact that hints at defect scattering, being the dominant scattering mechanism (Tanaka, 2014). Graphene has a unique combination of properties that is ideal for next-generation electronics, including mechanical flexibility, high electrical conductivity, and chemical stability.

The graphene presents a very high electrical conductivity, thanks to its zero-overlap semi-metallic behavior, with holes and electrons as charge carriers. The electrons in carbon atoms are distributed in the following way: 2 in the inner shell and 4 in the outer shell. Three outer electrons

Figure 1. Different types of graphene sheets in accordance with their chiral indices (n,m) and chiral angle (ψ)



are connected to other 3 carbon atoms, and the fourth electron is left free to be exploited in the electrical conductivity. These free electrons are called pi (π) electrons and they are located in both sides of the graphene sheet. Moreover, these electrons overlap, and improve the carbon-to-carbon bonds in graphene, leading to electronic properties established by the bonding and anti-bonding, of the orbitals associated with them.

Three different types of sheets of graphene can be identified: armchair, zigzag and chiral, in accordance with the chiral indices, as is illustrated in Figure 1 (Vargas-Bernal, 2015b). Zigzag graphene nanoribbons have a metallic behavior, while armchair graphene nanoribbons can have behavior either metallic or semiconducting. This last depends of the number (N) of atoms across the width where k is the number of nanoribbons presented by the material, and therefore, when $N=3k-1$ a metallic behavior is obtained, if $N=3k$ a semiconducting behavior is presented, and finally, if $N=3k+1$, a moderate semiconducting behavior is achieved (see Figure 2).

Another important parameter which defines the electrical properties of graphene is its bandgap. Graphene has no bandgap, since its conduction and

valence bands fulfill the Dirac points. The Dirac points are six locations in momentum space, on the edge of the Brillouin zone, divided into two non-equivalent sets of three points. The two sets are labeled as K and K' , as is illustrated in Figure 3. Electrons through the graphene propagate according to the tight-binding model (electronic band structure), where they are energetically dispersed as a traveling wave within the medium. The mathematical relationship for determining 2D energy dispersion of the electrons in π bands of graphene is given as (Das, 2015):

$$\varepsilon_{2D} = \pm \gamma_0 \sqrt{1 + 4 \cos \left(\frac{\sqrt{3} k_x a}{2} \right) \cos \left(\frac{k_y a}{2} \right) + 4 \cos^2 \left(\frac{k_y a}{2} \right)} \quad (4)$$

where $\gamma_0 = 3.033$ eV is the nearest-neighbor overlap energy. The ε - k (energy versus wave vector) relationship of the electrical carriers is linear for lowest energies near the six corners of the 2D hexagonal Brillouin zone, which leads to zero effective mass for electrons and holes, as shown in Figure 3 (Shuai, 2012).

Figure 2. Electrical behavior of the graphene related with the number of atoms across the width of the nanoribbon of graphene and the wave vector k

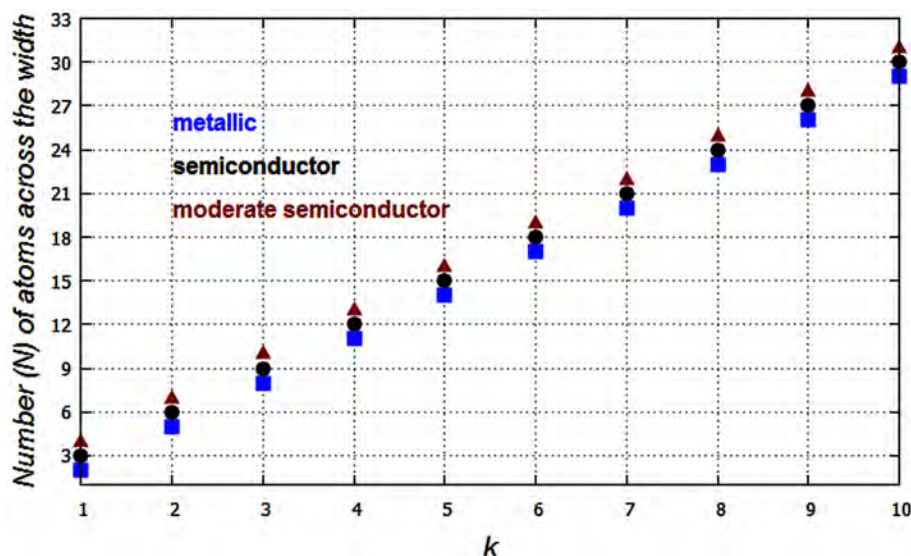
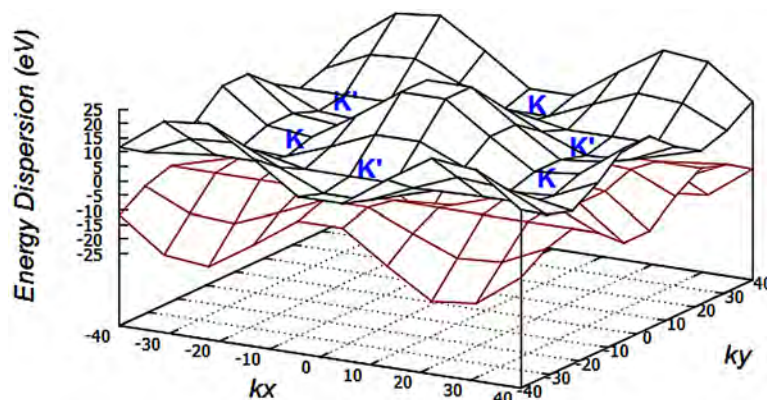


Figure 3. Band structure of graphene (ϵ - k relationship of graphene nanoribbons)

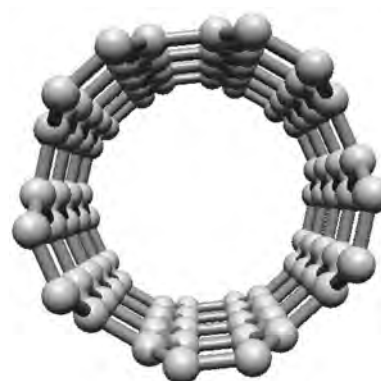
A carbon nanotube is an allotrope of carbon that can be conceptualized as a layer of graphite with one-atom-thick called graphene, which is wrapped as a cylinder. One, two or more layers of graphene can be concentrically wrapped giving place to a single-wall nanotubes (SWNTs), double-wall nanotubes (DWNTs) or multi-wall nanotubes (MWNTs). The wrapping of a graphene sheet can be identified by means of the indices (n, m) , where n and m denote integer numbers representing the number of unit vectors along directions in the honeycomb crystal lattice. When $m = 0$, the carbon nanotubes are called zigzag nanotubes, if $n = m$ they are called armchair nanotubes, and in otherwise, they are called chiral nanotubes, as depicted in Figure 4.

The diameter of an ideal carbon nanotube is determined from its (n, m) chiral indices as:

$$d = \frac{\sqrt{3}a_{CC}}{\pi} \sqrt{n^2 + nm + m^2} \quad (5)$$

where $a_{CC} = 0.142$ nm is the carbon-carbon bond length, as is illustrated in Figure 5. The electrical properties of the carbon nanotubes are directly proportional to the value of the diameter, that is, great values of diameter imply high conductivity and vice versa. The chiral angle is defined as:

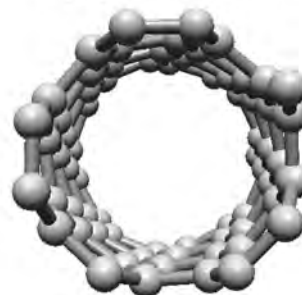
Figure 4. Different types of carbon nanotubes according with chiral indices



Armchair Carbon Nanotube

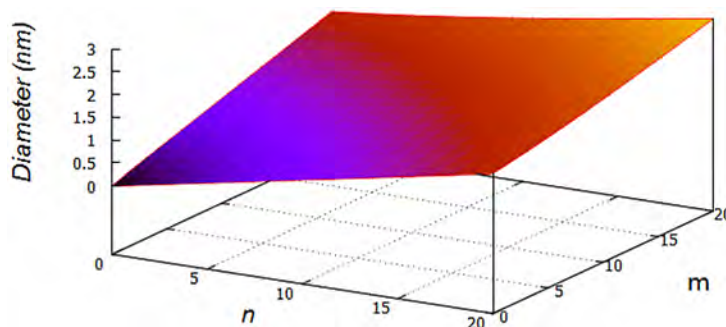


Zigzag Carbon Nanotube



Chiral Carbon Nanotube

Figure 5. Estimation of the diameter of carbon nanotubes according to the chiral indices



$$\psi = \cos^{-1} \left(\frac{2n + m}{2\sqrt{n^2 + nm + m^2}} \right) \quad (6)$$

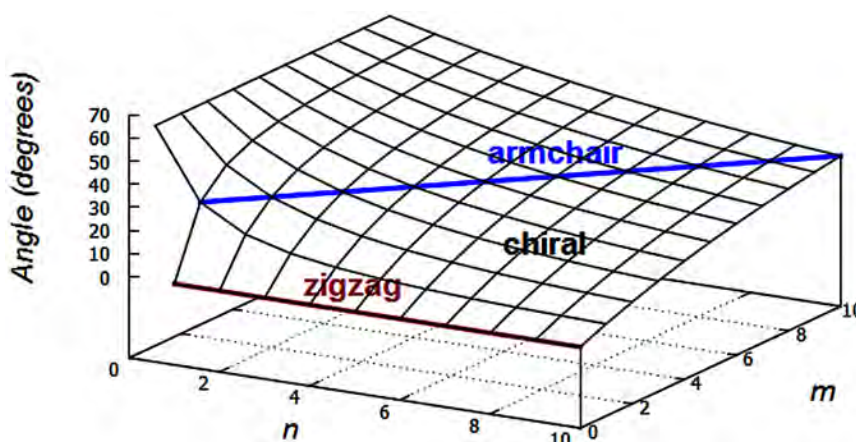
It allows distinguishing three classes of carbon nanotubes with different electrical properties: armchair ($n=m$, $\psi=30^\circ$) with a metallic behavior, zigzag ($m=0$, $n>0$, $\psi=0^\circ$) with a semiconducting behavior, and chiral ($0<|m|<n$, $0<|\psi|<30^\circ$) with a moderate semiconducting behavior, as shown in Figure 6.

In accordance with Kataura's graph, the band gap energy in a carbon nanotube is inversely proportional to the diameter of the nanotube. In a SWNT, their band gap goes from 0 to 2 eV and their electrical conductivity presents metallic or semiconducting behavior. For a given (n, m) nanotube, when $n = m$, the nanotube has a me-

tallic behavior; if $n-m$ is a multiple of 3, then the nanotube is semiconducting, with a small bandgap, and otherwise the nanotube presents a moderate semiconducting behavior with band gap (Vargas-Bernal, 2012). The chirality has strong influence on electrical properties in nanotubes with small diameter. Theoretically, carbon nanotubes can carry an electrical current density of the order of $4 \times 10^{13} \text{ A/m}^2$, whose value is more than 1,000 times greater than that of metals, such as copper, which reduce the effect of electromigration.

In carbon nanotubes, electrons travel only along the tube's axis, and therefore, these materials are referred to as one-dimensional conductors. At room temperature, the dominant scattering mechanism of transport of electrons implies the emission of optical phonons and it is called ballistic transport. When in a carbon nanotube and graphene nanorib-

Figure 6. Electrical classification of the carbon nanotubes according to the chiral angle and indices



bon, the electrons in transit are not scattered with too many phonons and the devices have about 100 nm long, then the transport of electrons depends on the nanoribbon edge structure and the electron energy. The maximum electrical conductance of a single-walled carbon nanotube is estimated as two times (sheet with two sides) the value than that of a single ballistic quantum channel, that is, $2G_0$ where $G_0 = 2e^2/h$ or $7.748091734625 \times 10^{-5} \text{ S}$, where e is the electrical charge of the electron and h is the Planck's constant. At high temperatures, a different mechanism of transport of electrons is presented in carbon nanotubes called hopping. Hopping is a thermally activated process in which an electron moves from one site to another. The electron loses all information about its phase in the process, and it does not exist, coherence among the amplitudes for finding an electron at different lattice sites.

Electrical transport properties of one-dimensional or quasi-1D nanomaterials such as quantum wires, nanofibers, and carbon nanotubes are been subjected of intense scrutiny with the aim of introducing these materials into electronic applications. Semiconducting carbon nanotubes are used in nano-electronic devices, while metallic carbon nanotubes are applied in nano-interconnects (Das, 2015). The electrical transport in semiconducting carbon nanotubes can be ballistic (tunneling) (see

Figure 7), or diffusive (scattering) (see Figure 8). In carbon nanomaterials, tunneling can be defined as a quantum mechanical phenomenon, where an electron tunnels through or cross a potential energy barrier (space without matter), which it classically could not be surmount. Mechanism of diffusion based on scattering appears when defects, impurities or lattice vibrations are found in the carbon nanomaterials. Diffusion is the movement of electrons from a region of high concentration to a region of low concentration, where variables such as pressure or temperature are involved.

During ballistic transport, carbon nanomaterials do not dissipate heat (Frank, 1998). Electrical conduction in carbon nanomaterials is quantized due to their one-dimensionality, and the number of allowed electronic states is limited. In other words, charge carriers are transmitted through discrete conduction channels without experiencing scattering due to impurities, local defects or lattice vibrations (Vargas-Bernal, 2015a). In addition, the electrons do not find electrical resistance, and no energy dissipation occurs in the conduction channel. For armchair carbon nanotubes, there are two sub-bands that cross the Fermi level, and to semiconducting carbon nanotubes, these sub-bands do not cross the Fermi level. Thus, there are two conducting channels and each band accommodates two electrons of opposite spin (White,

Figure 7. Electrical ballistic transport in carbon nanomaterials



Figure 8. Electrical diffusive transport in carbon nanomaterials



1998). Conduction of electrons in armchair carbon nanotubes experience an effective disorder that is averaged over the tube's circumference, which increases the electron's mean free paths with nanotube's diameter, and therefore, the ballistic transport in the carbon nanotube is guaranteed. The electrical conductivity of individual carbon nanotubes under ballistic conduction achieves the value of 100×10^6 S/m for single-walled and 3.33×10^6 S/m for multi-walled carbon nanotubes, respectively (Li, 2007).

Functionalization of carbon nanomaterials is necessary to interface it with other moieties in order to expand the scope of its electrical/electronic applications (Sreeprasad, 2013). Unfortunately, this also modulates its electrical properties, and therefore, a wide knowledge related with the effect of carrier scattering, carrier concentration, charge polarity, quantum-capacitance enhanced doping, energy levels, transport mechanisms, and orbital hybridization of energy-bands, must be developed. In (Sreeprasad, 2013), the authors have included factors such as covalent bonds, adsorption, π - π bonds, and lattice incorporation to model its influence on electrical characteristics. They discovered that three different mechanisms governing the electrical properties can be identified: (1) conversion of carbon's hybridized state, (2) dipole interactions enhanced via quantum capacitance, and (3) orbital hybridization with an interfacing molecule.

The use of dopants in carbon nanomaterials such as metals increases the electrical conductivity and its electrical current capacity. Ag nanowires have been used to generate hybrid fibers with graphene oxide (Xu, 2013) for applications in high-performance materials, chemical supercapacitors, ultralight aerogels, sensors and catalysts.

Nanocomposites for electrical applications and based on carbon nanotubes or graphene and polymers, must achieve a certain percolation threshold or a minimum volume fraction, to form a cluster of carbon nanotubes or graphene sheets connected, which will operate as electrical pathways, between a pair of electrodes, to carry

out the electrical transport of electrical carriers (Hu, 2014). In this case, the electrical conductivity implies a tunneling of electrons (non-ohmic behavior), one by one, through the polymer found between two neighboring carbon nanotubes or two neighboring graphene nanoribbons.

Carbon nanomaterials can be used as electrical fillers in composite materials or hybrid materials for electronic applications. These materials must be dispersed in polymeric matrix to avoid the formation of agglomerates of fillers. The electrical transport mechanism in these materials is due to tunneling between electrical fillers, since polymers are regularly electrical insulators (Wen, 2012). When aligned carbon nanomaterials are used in a composite material based on polymer (Gong, 2015), anisotropic electrical conductivity in the longitudinal and transverse directions is obtained due to: 1) nanomaterials are distortable, and the Van der Waals forces are presented, and 2) there is structural distortion of the fillers. In addition, aligned graphene in composite materials based on polymers can provide high dielectric constants of over 14000 with 3 wt% of graphene at 1 KHz thanks to the charge accumulation favored by the aligning (Yousefi, 2014).

Surfactants can be used to exfoliate and disperse carbon nanomaterials to be used in composites based in polymers (Mittal, 2015). These additives allow reduce the percolation threshold (minimum percentage to form an interconnected network between electrical fillers), and therefore, a high electrical conductivity is easily achieved, either using high boiling point solvents or low boiling point solvents.

SOLUTIONS AND RECOMMENDATIONS

Different mechanisms are contributing to the electrical conductivity; a complete analysis must be realized to determine the percentage of its influence on electrical properties of the carbon nanomaterials. In real applications, ballistic and diffusive

transports are completely influenced by different perturbations from fabrication and operation of the carbon nanomaterials and their composites or hybrid materials. Models of conductivity electrical must be developed to understand the types of transport presented in aligned carbon nanotubes (Yousefi, 2014; Gong, 2015), segregated structures (Du, 2011), compacted materials composed of multiple carbon nanomaterials (Marinho, 2012), and mixes of graphene and carbon nanotubes in composite and hybrid materials.

Until now, in the case of composite materials, diverse strategies have been proposed to reduce the percentage of the CNTs and/or graphene (electrical fillers) to be incorporated in composites based on polymers or metals with the aim of obtaining extraordinary electrical properties (Min, 2010, Jomaa, 2015). The most critical factor implied in the reduction of the filler is the reduction of the percolation threshold or the minimum value of carbon nanotubes and/or graphene used to achieve the electrical conductivity in composite or hybrid materials. Among these techniques that reduce the quantity of filler used in a composite or a hybrid material, can be mentioned: surfactant treating, acid treating, and chemical functionalization. The surfactants lower the surface tension of carbon nanomaterials, which preventing the formation of aggregates of fillers, and thus, an increasing of electrical conductivity is obtained. The cleaning of carbon nanomaterials by means of acids increases electrical conductivity of them, decreasing the concentration of impure materials found on surface of these materials. However, an excessive time in acid can damage the structure of the fillers. A better dispersion can be achieved with functionalization since it changes the hybridization of the carbon nanomaterials, and thus, electrical properties are also changed.

FUTURE RESEARCH DIRECTIONS

This chapter has presented an overview of the most significant advances in the research on electrical transport mechanism that can be exploited

in carbon nanomaterials such as graphene and carbon nanotubes. More research must be realized to discover the effect of the ambient on electrical conductivity of both carbon nanomaterials as was developed in (Roch, 2015). A lot of work must be realized to model electrical transport in carbon nanomaterials that have been functionalized to be used in electronic applications.

Carbon nanomaterials will continue offering a highly disruptive technology to be exploited in the interconnection of electronic devices, and the fabrication of electronic devices based either in individual elements or as composite and hybrid materials. These materials will be used in the next generation of electrical wires, with the aim of replacing traditional metal ones based on copper or aluminum (Lekawa-Raus, 2014).

CONCLUSION

In this article, different mechanisms involved in the electrical conductivity of carbon nanotubes and graphene, either working individually or within a composite or hybrid materials, have been discussed. In the case of composite or hybrid materials can be distinguished two mechanisms: the formation of a conducting network interconnecting, carbon nanotubes and/or graphene sheets, and the hopping of the electrons within carbon nanomaterials. Carbon nanomaterials can present ballistic transport (perfect material) or diffusive transport (when defects, impurities and lattice vibrations are found). New mechanisms are presented when carbon nanomaterials are functionalized: hybridized states of the carbon atoms, dipole interactions based on quantum capacitance, and orbital hybridization in the interface.

ACKNOWLEDGMENT

This work is part of a project that is being financed by CONACYT-Mexico (project 152524) and supported by PRODEP and ITESI (project 284.15-PD).

REFERENCES

- Bird, J. (2014). *Electrical Circuit: Theory and Technology* (5th ed.). New York: Routledge.
- Das, D., & Rahaman, H. (2015). *Carbon Nanotube and Graphene Nanoribbon Interconnects*. Boca Raton, FL: CRC Press.
- Du, J., Zhao, L., Zeng, Y., Zhang, L., Li, F., Liu, P., & Liu, C. (2011). Comparison of Electrical Properties between Multi-Walled Carbon Nanotube and Graphene Nanosheet/High Density Polyethylene Composites with a Segregated Network Structure. *Carbon*, 49(4), 1094–1110. doi:10.1016/j.carbon.2010.11.013
- Frank, S., Poncharal, P., Wang, Z. L., & de Heer, W. A. (1998). Carbon Nanotube Quantum Resistors. *Science*, 280(5370), 1744–1746. doi:10.1126/science.280.5370.1744 PMID:9624050
- Gong, S., Zhu, Z. H., & Meguid, S. A. (2015). Anisotropic Electrical Conductivity of Polymer Composites with Aligned Carbon Nanotubes. *Polymer*, 56, 498–506. doi:10.1016/j.polymer.2014.11.038
- Grady, B. P. (2011). *Carbon Nanotube-Polymer Composites: Manufacture, Properties, and Applications*. Hoboken, NJ: John Wiley & Sons. doi:10.1002/9781118084380
- Hu, K., Kulkarni, D. D., Choi, I., & Tsukruk, V. V. (2014). Graphene-Polymer Nanocomposites for Structural and Functional Applications. *Progress in Polymer Science*, 39(11), 1934–1972. doi:10.1016/j.progpolymsci.2014.03.001
- Jomaa, M. H., Masenelli-Varlot, K., Seveyrat, L., Lebrun, L., Dib Jawhar Dib, M. C., Beyou, E., & Cavaillé, J. Y. (2015). Investigation of Elastic, Electrical and Electromechanical Properties of Polyurethane/Grafted Carbon Nanotubes Nanocomposites. *Composites Science and Technology*, 121, 1–8. doi:10.1016/j.compscitech.2015.10.019
- Jug, K., Hilbert, P. C., & Shail, S. (2001). σ - π Energy Separation in Modern Electronic Theory for Ground States of Conjugated Systems. *Chemical Reviews*, 101(5), 1477–1500. doi:10.1021/cr990328e PMID:11710230
- Lekawa-Raus, A., Patmore, J., Kurzepa, L., Bulmer, J., & Koziol, K. (2014). Electrical Properties of Carbon Nanotube based Fibers and Their Use in Electrical Wiring. *Advanced Functional Materials*, 24(24), 3661–3682. doi:10.1002/adfm.201303716
- Li, Q. W., Li, Y., Zhang, X. F., Chikkannanavar, S. B., Zhao, Y. H., Dangelewicz, A. M., & Zhu, Y. T. et al. (2007). Structure-Dependent Electrical Properties of Carbon Nanotube Fibers. *Advanced Materials*, 19(20), 3358–3363. doi:10.1002/adma.200602966
- Marinho, B., Ghislandi, M., Tkalya, E., Koning, C. E., & de With, G. (2012). Electrical Conductivity of Compacts of Graphene, Multi-Wall Carbon Nanotubes, Carbon Black, and Graphite Powder. *Powder Technology*, 221, 351–358. doi:10.1016/j.powtec.2012.01.024
- Min, C., Shen, X., Shi, Z., Chen, L., & Xu, Z. (2010). The Electrical Properties and Conducting Mechanisms of Carbon Nanotube/Polymer Nanocomposites: A Review. *Polymer-Plastics Technology and Engineering*, 49(12), 1172–1181. doi:10.1080/03602559.2010.496405
- Mittal, G., Dhand, V., Rhee, K. Y., Park, S.-J., & Lee, W. R. (2015). A Review on Carbon Nanotubes and Graphene as Fillers in Reinforced Polymer Nanocomposites. *Journal of Industrial and Engineering Chemistry*, 21, 11–25. doi:10.1016/j.jiec.2014.03.022
- Roch, A., Greifzu, M., Talens, E. R., Stepien, L., Roch, T., Hege, J., & Leson, A. et al. (2015). Ambient Effects on the Electrical Conductivity of Carbon Nanotubes. *Carbon*, 95, 347–353. doi:10.1016/j.carbon.2015.08.045

Shuai, Z., Wang, L., & Song, C. (2012). Theory of Charge Transport in Carbon Electronic Materials. Springer. doi:10.1007/978-3-642-25076-7

Sreeprasad, T. S., & Berry, V. (2013). How Do the Electrical Properties of Graphene change with Its Functionalization? *Small*, 9(3), 341–350. doi:10.1002/sml.201202196 PMID:23169614

Tanaka, K., & Iijima, S. (Eds.). (2014). *Carbon Nanotubes and Graphene* (2nd ed.). Kidlington: Oxford, Elsevier.

Vargas-Bernal, R. (2015a). Performance Analysis of Interconnects based on Carbon Nanotubes for AMS/RF IC Design. In M. Fakhfakh, E. Tlelo-Cuautle, & M. H. Fino (Eds.), *Performance Optimization Techniques in Analog, Mixed-Signal, and Radio-Frequency Circuit Design* (pp. 336–363). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-6627-6.ch014

Vargas-Bernal, R. (2015b). Advances in Computational Modeling of Electronic Devices based on Graphene. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 5(1), 109–116. doi:10.1109/JETCAS.2015.2398219

Vargas-Bernal, R. (2015c). Performance Analysis Electromagnetic Interference Shielding based on Carbon Nanomaterials used in AMS/RF IC Design. In M. Fakhfakh, E. Tlelo-Cuautle, & M. H. Fino (Eds.), *Performance Optimization Techniques in Analog, Mixed-Signal, and Radio-Frequency Circuit Design* (pp. 268–294). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-6627-6.ch011

Vargas-Bernal, R., & Herrera-Pérez, G. (2012). Carbon Nanotube- and Graphene based Devices, Circuits and Sensors for VLSI Design. In VLSI Design, (pp. 41-66). Rijeka, Croatia: Intech.

Wen, M., Sun, X., Su, L., Shen, J., Li, J., & Guo, S. (2012). The Electrical Conductivity of Carbon Nanotube/Carbon Black/Polypropylene Composites prepared through Multistage Stretching Extrusion. *Polymer*, 53(7), 1602–1610. doi:10.1016/j.polymer.2012.02.003

White, C. T., & Todorov, T. N. (1998). Carbon Nanotubes as Long Ballistic Conductors. *Nature*, 393(6682), 240–242. doi:10.1038/30420

Xu, Z., Liu, Z., Sun, H., & Gao, C. (2013). Highly Electrically Conductive Ag-doped Graphene Fibers as Stretchable Conductors. *Advanced Materials*, 25(23), 3249–3253. doi:10.1002/adma.201300774 PMID:23640822

Yousefi, N., Sun, X., Lin, X., Shen, X., Jia, J., Zhang, B., & Kim, J. K. et al. (2014). Highly Aligned Graphene/Polymer Nanocomposites with Excellent Dielectric Properties for High-Performance Electromagnetic Interference Shielding. *Advanced Materials*, 26(31), 5480–5487. doi:10.1002/adma.201305293 PMID:24715671

Yusoff, A. R. M. (Ed.). (2015). *Graphene-based Energy Devices*. Wiley-VCH.

Zheng, Q., & Kim, J.-K. (2015). *Graphene for Transparent Conductors: Synthesis, Properties and Applications*. New York, NY: Springer. doi:10.1007/978-1-4939-2769-2

ADDITIONAL READING

Mahmoodi, M., Arjmand, M., Sundararaj, U., & Park, S. (2012). The Electrical Conductivity and Electromagnetic Interference Shielding of Injection Molded Multi-Walled Carbon Nanotube/Polystyrene Composites. *Carbon*, 50(4), 1455–1464. doi:10.1016/j.carbon.2011.11.004

Matsumoto, K. (Ed.). (2015). *Frontiers of Graphene and Carbon Nanotubes: Devices and Applications*. Tokyo, Kanto: Springer. doi:10.1007/978-4-431-55372-4

Sadasivuni, K. K., Ponnamm, D., Kim, J., & Thomas, S. (Eds.). (2015). *Graphene-based Polymer Nanocomposites in Electronics*. Heidelberg, Baden-Württemberg, Springer. doi:10.1007/978-3-319-13875-6

Shafraniuk, S. (2015). *Graphene: Fundamentals, Devices, and Applications*. Boca Raton, Florida: Pan Stanford Publishing. doi:10.1201/b18258

Tiwari, A., & Syväjärvi, M. (Eds.). (2015). *Graphene Materials: Fundamentals and Emerging Applications*. Hoboken, New Jersey: John Wiley & Sons. doi:10.1002/9781119131816

KEY TERMS AND DEFINITIONS

Ballistic Conduction: or **Ballistic Transport:** The transport of electrons in a medium having negligible electrical resistivity or the highest electrical conductivity.

Band Gap: An energy difference (in electron volts) between the top of the valence band and the bottom of the conduction band in insulators and semiconductors.

Carbon Nanomaterials: Nanostructures of carbon such as fullerenes, carbon nanotubes, nanofibers and graphene with unique physico-chemical properties with multiple technological applications.

Carbon Nanotubes: Allotropes of carbon with a cylindrical nanostructure of length-to-diameter of up to 132,000,000:1, which have unusual properties and valuable for nanotechnology, electronics, optics and other fields of materials science and technology.

Diffusive Transport: The movement of electrons from a region of high concentration to a region of low concentration where variables such as pressure or temperature are involved.

Electrical Conductivity: The physical property that quantifies how strongly a given material opposes the flow of electrical current. A low resistivity indicates a material that readily allows the movement of electrical charge.

Electronic Device: Device that accomplishes its purpose controlling the flow of electrons applied to digital electronics, analog electronics, microelectronics, optoelectronics, or integrated circuits.

Graphene: A two-dimensional, crystalline allotrope of carbon whose atoms are densely packed in a regular sp^2 -bonded atomic-scale chicken wire (hexagonal) pattern composed by a one-atom thick layer of graphite.

Interconnect: A path of material that connects two elements or components in an integrated circuit, through which electrical current is transported.

Quantum Tunneling or Tunneling: A quantum mechanical phenomenon, where an electron tunnels through a band gap, which it classically could not surmount.

Category E

Electronic Business

E-Business and Big Data Strategy in Franchising

Ye-Sho Chen

Louisiana State University, USA

INTRODUCTION

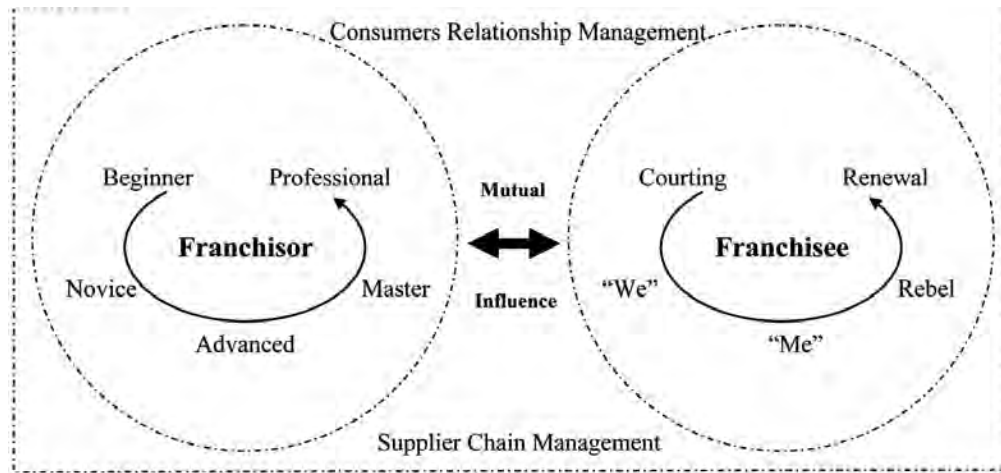
Franchising as a global growth strategy is gaining its popularity (Justis and Judd, 2002; Thomas and Seid, 2000; Chen and Justis, 2006). For example, according to the statistics of the China Chainstore & Franchise Association, China has over 4,500 franchises and chain store companies creating more than 5 million jobs nationwide and the country's top 100 franchises generated total sales of \$66 billion with the total number of stores across these 100 franchises at 124,086 (U.S. Commercial Service, 2016). The popularity of franchising continues to increase, as we witness an emergence of a new e-business model, Netchising, which is the combination power of the *Internet* with big data (O'Donnell, 2014; Saunders, 2015; Franchise Update, 2016) for global demand-and-supply processes and the international *franchising* arrangement for local responsiveness (Chen, Justis, and Yang, 2004; Chen, Chen, and Wu, 2006). In his best seller, *Business @ the Speed of Thought*, Bill Gates (1999) wrote: "Information Technology and business are becoming inextricably interwoven. I don't think anybody can talk meaningfully about one without talking about the other." (p. 6) Gates' point is quite true when one talks about e-business and big data strategy in franchising. Thus, to see how e-business and big data can be "meaningfully" used in franchising, one needs to know how franchising really works.

BACKGROUND: BUILDING THE FRANCHISOR/FRANCHISEE RELATIONSHIP

Franchising is "a business opportunity by which the owner (producer or distributor) of a service or a trademarked product grants exclusive rights to an individual for the local distribution and/or sale of the service or product, and in return receives a payment or royalty and conformance to quality standards. The individual or business granting the business rights is called the *franchisor*, and the individual or business granted the right to operate in accordance with the chosen method to produce or sell the product or service is called the *franchisee*." (Justis & Judd, 2002, pp. 1-3) Developing a good relationship between the franchisor and the franchisee is the key for a successful franchise (Justis & Judd, 2002). Figure 1 describes how to build a good franchisor/franchisee relationship.

The franchisor needs to learn continuously for the growth of the franchise. The learning process is developed through five stages (Justis and Judd, 2002): (1) Beginner – learning how to do it; (2) Novice – practicing doing it; (3) Advanced – doing it; (4) Master – teaching others to do it; and (5) Professional – becoming the best that you can be. Once reaching the Advanced stage, most preceding struggles have been overcome. However, further challenges will arise as the franchise continues growing. This is especially true once the system reaches the "Professional" stage, where various

Figure 1.



unpredicted and intricate problems could arise. Bud Hadfield (1995), the founder of Kwik Kopy franchise and the International Center of Entrepreneurial Development, aptly stated: “The more the company grows, the more it will be tested.” (p. 156). To capture the learning process, a counter-clockwise round arrow surrounding the franchisor is used to depict the increasing intensity of learning as the franchisor continues to grow.

The franchisee also goes through five stages of franchisee life cycle (Schreuder, Krige, and Parker, 2000): (1) Courting: both the franchisee and the franchisor are eager with the relationship; (2) “We”: the relationship starts to deteriorate, but the franchisee still values the relationship; (3) “Me”: the franchisee starts to question the franchisor that the success so far is purely of his/her own work; (4) Rebel: the franchisee starts to challenge the franchisor; and (5) Renewal: the franchisee realizes the “win-win” solution is to continue working with the franchisor to grow the system. Similar to the franchisor, a counter-clockwise round arrow surrounding the franchisee is used in Figure 1 to depict the increasing intensity as the franchisee continues growing.

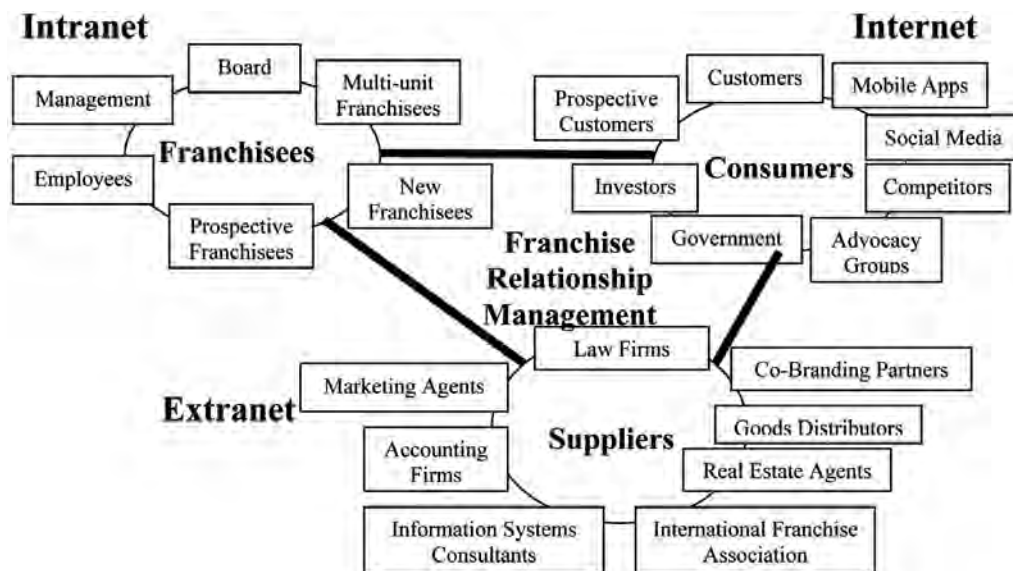
As the franchisee progresses through the life cycle, the good relationship gradually develops an influencing process (Justis & Vincent, 2001), depicted in Figure 1 with a bi-directional arrow.

By going through the processes of learning and influencing, supported by big data analytics such as consumer connection (FB, 2014) & boosting engagement (TFO, 2014), picking store locations (Thau, 2014), and tracking franchisee happiness (TFO, 2014), both the franchisor and the franchisee gain the progressive working knowledge of relationship management with the consumers and suppliers. The franchisor, the franchisee, the consumers, and the suppliers in Figure 1 are surrounded with dashed lines, indicating that there is no limit to the learning process.

E-BUSINESS STRATEGY IN FRANCHISING

With the advancement of Internet technology, franchise companies are adapting e-business strategies for perfecting the franchisor/franchisee relationship to grow their franchises globally. Figure 2 is a visual depiction of deploying e-business strategy in franchising. This community of franchise companies, consumers, and suppliers can be virtually connected for relationship management as follows: (1) collaboration with consumers through Internet, enabling the franchisor and the franchisees to build up relationships with customers, prospective customers, investors, competitors, social media,

Figure 2. E-Business Strategy in Franchise Relationship Management



mobile apps, advocacy groups, and government; (2) intra-enterprise collaboration through Intranet, enabling the franchisor to build up relationships with the board of directors, multi-unit franchisees, new franchisees, prospective franchisees, franchisor management and employees; and (3) collaboration with suppliers through Extranet, enabling the franchisor and the franchisees to build up relationships with members and affiliates of international franchise association, law firms, co-branding partners, goods distributors, real estate agents, information systems consultants, accounting firms, and marketing agents.

Harnessing the E-Business Strategy Around the Customer Service Life Cycle

Table 1 shows a customer-service-life-cycle (CSLC) (Ives and Mason, 1990) e-business strategy in franchising (Chen, Chong, and Justis, 2002) for relationship management depicted in Figure 2. Here we define the franchisee as the customer of the franchisor and the franchisee's customer as the customer's customer of the franchisor. The stages of CSLC are based on two well-known

franchising books by Justis and Judd (2002) and Thomas and Seid (2000).

There are four major components in the e-business strategy: (1) Benchmarking the Requirements and Acquisition stages. The CSLC model shown in Table 1 is a comprehensive guide for a franchise to develop its web site, especially at the stages of Requirements and Acquisition. The model may be used to compare a franchise's e-business strategy with its competitors. As the industry progresses, best practices based on the CSLC model will evolve and become a standard for benchmarking and websites enhancements. (2) Helping the franchisees serve their customers in the Ownership stage without the Internet encroachment. There is a rich collection of studies in e-business in franchising (Chen, Chong, and Justis, 2002) showing how the Internet can help the franchisees serve their customers in the Ownership stage, including "Marketing & Promoting the Franchise Products/Services" and "Managing the Franchise System". (3) Cultivating the Ownership and Renewal/Retirement stages with effective knowledge management. As mentioned earlier, the greatest challenge in the Ownership stage is to build up the relationship between the franchisor and franchisee. To cultivate the Owner-

Table 1. The Customer-Service-Life-Cycle (CSLC) E-business strategy in franchising

CSLC	Sub-stages	Internet Strategy	Intranet Strategy	Extranet Strategy
Requirements	Understanding How Franchising Works	Using the web site as the friendly customer relationship management tool to address customer concerns at various stages (Chen, Chong, and Justis, 2002), e.g., providing useful information on financing and showing how the franchise system may help finance the franchise investment Benchmarking and enhancing the web site continuously (Chen, Chong, and Justis, 2002; Chen, Justis, and Chong, 2008), e.g., identifying frequently the best practices of web deign in the industry and improving the web site accordingly	Transforming the organizational structure and corporate culture to fit the e-business operation pushed by the Intranet systems (Zeng, Chen, and Huang, 2008), e.g., designing an environment for more team work opportunities and establishing e-learning environment for the employees	Partnering with suppliers to enhance the various stages of CSLC continuously (Chen, Justis, and Wu, 2006), e.g., a franchise system may need to partner with banks to deliver good services at the stage of “Financing the Franchise Business”. Aligning the Internet and Intranet Strategy with reputable Application Service Providers (ASP) having focused businesses reengineering around the stages of CSLC (Chen, Ford, Justis, and Chong, 2001). For example, Statability.com is a “visionary Web-based Reporting” portal for the hospitality industry. It has the focused business reengineering around the stage of “Managing the Franchise System”. Its focused service is being respected by franchise companies in the hospitality industry, as is evidenced from the ever-increasing list of its client base, including Hilton and Marriott.
	Investigating Franchise Opportunities			
	Obtaining Franchisee Prospectus			
	Making the Choice			
Acquisition	Preparing Business Plan			
	Financing the Franchise Business			
	Signing the Contract			
Ownership	Marketing & Promoting the Franchise Products or Services		Helping the franchisees make sales and serve their customers, e.g., via mobile apps and social media, with proper policies dealing with the Internet encroachment issues	
	Managing the Franchise System		Cultivating the franchisor/ franchisee relationship with effective knowledge management tools (Chen, Chong, Justis, 2000a,b; Chen, Hammerstein, and Justis, 2002; Chen, Seidman, and Justis, 2005), e.g., basic communications support, distance learning, and centralized franchise applications such as employee recruitment and online ordering	
	Building the Relationship between the Franchisor and the Franchisee			
Renewal or Retirement	Becoming a Professional Multi-unit Franchisee or Retiring from the Franchise System			

ship stage so that "Professional" franchisee can advance to the Renewal stage instead of retiring, Chen, Chong, and Justis (2000) suggest building an Intranet-based Franchising Knowledge Repository. The Repository provides a framework based on which a franchise system may transform into a learning organization. (4) Partnering with the

"disruptive technology" providers to enhance the CSLC stages. Innovative entrepreneurs will reengineer their franchise businesses around the CSLC model shown in Table 1. Their ability to track, analyze, and leverage the buying behaviors of their customers in the CSLC sub-stages is their real competitive advantage. For example, Statability.

com is a “visionary Web-based Reporting” portal for the hospitality industry. In terms of the CSLC model, Statability.com is a focused business which reengineers around the sub-stages of Ownership. Its “disruptive technology” of reporting is adopted by many franchises in the hospitality industry. As discussed earlier, partnering with those “disruptive technology” providers will make the franchise system more competitive.

Aligning the CSLC-Based E-Business Strategy with Application Service Providers

Although Internet technology can help deploy the franchise’s e-business strategy, the immediate question is: at what cost? Because of the need for e-business processes to monitor the linkage of internal information technologies with external processing and services, the e-business investment could be very expensive and complicated. Many franchise companies, especially small ones, find it financially difficult to invest in the e-business technologies; however, a new type of service in e-business called Application Service Providers (ASP) promises to make e-business more economical and affordable to the franchise systems. The concept of subscribing information technologies through ASPs has special appeal in the franchising industry because an ASP can duplicate success for other similar franchises quickly and inexpensively (Chen, Ford, Justis, and Chong, 2001). When aligning the CSLC-based e-business strategy with ASPs, a franchise company should focus on (Chen, Ford, Justis, and Chong, 2001): (1) Develop an overall vision of the *applications*, including software and hardware, needed for the company. (2) Determine what applications and the specific *services*, e.g., to be available 24 hours a day and 7 days a week with 99.999% of reliability, you want an ASP to host, which have to be clearly defined in the Service Level Agreement. (3) Evaluate ASP *providers*, i.e., vendors who provide the applications services, using flexibility and trust relationship as the two primary factors.

An Attention-Based Framework for the Franchisee Training

The third industrial revolution, combining information technology with globalization, produces an environment where everyone is facing the problem of information overload. Simon (1971) spoke for us all when he said that ‘a wealth of information creates a poverty of attention.’ (p.41) Getting the franchisee’s attention on training in an information rich world is a major challenge. Ocasio (1997) proposed an attention-based theory of the firm, which allows the firm to shield off irrelevant information and gain access to information relevant to what the firm focuses on. According to Ocasio (1997), attention is defined to “encompass the noticing, encoding, interpreting, and focusing of time and effort by organizational decision-makers on both (a) issues: the available repertoire of categories for making sense of the environment: problems, opportunities, and threats; and (b) answers: the available repertoire of action alternatives: proposals, routines, projects, programs, and procedures.” (p.188) Ocasio (1997) further classifies attention into three principles: (1) focus of attention, what decision makers do primarily depends on the selective issues and answers they focus attention on; (2) situated attention, what decision makers focus on and do depends primarily on the particular contextual environment they are located in; and (3) structural distribution of attention, how decision makers attend to the particular contextual environment they are in depends on how the firm’s attention structure (including rules, resources, and relationships) channels and distributes various issues, answers, and decision makers into specific communications and procedures.

In the context of franchising, what do focus of attention, situated attention, and attention structures look like? How does a franchise design an attention-based training program for the franchisees? We propose an attention-based framework in Table 2 for the franchisee training. Such a framework has two dimensions. The first dimension is the franchisee life cycle, consisting of Beginner

Table 2. An Attention-based framework for the franchisee training

Situating Attention: Relationship Touchpoints		Attention Structures: Demand & Supply Value Networks				
		Customers	Franchisee Outlet	Franchisor Headquarters	Suppliers & Partners	Franchise Community
Franchisee Life Cycle	Beginner in the Courting Phase: Beginner Guide	Focus of Attention: Learning how to become a franchisee ■ Understanding how franchising works ■ Investigating franchise opportunities ■ Obtaining franchisee prospectus ■ Making the choice ■ Preparing business plan ■ Financing the franchised business ■ Signing the contract				
	Novice in the “We”-Phase: Practicing	Focus of Attention: Practicing how to do activities such as: ■ How to get training and services from the headquarters ■ How to find a good site ■ How to find suppliers ■ How to work with the franchisor ■ How to work with fellow franchisees				
	Advanced in the “Me”-Phase: Doing	Focus of Attention: Doing activities such as: ■ How to acquire and keep customers ■ How to hire, train, and fire employees ■ How to manage inventory ■ How to manage the back office operations				
	Master in the Rebel Phase: Teaching Others	Focus of Attention: Teaching others how to do activities such as: ■ How to teach others ■ How to work as team ■ How to do the bulleted processes above for Beginner, Novice, and Advanced franchisees				
	Professional in the Renewal Stage: Creative Learning and Innovation	Focus of Attention: Becoming the best he/she can be by: ■ Learning to creatively improve activities such as: ■ How to cut the cost of the operations ■ How to increase the profit of the operations ■ How to acquire other franchises and brands ■ Looking for opportunities for innovation such as: ■ Are there any new growth opportunities we can create based on our intangible assets of demand & supply value chains? How do we avoid the loss of this new venture? ■ Are there any partnership opportunities with our customers and suppliers so that their customers could become ours and vice versa. ■ What are the major concerns in the communities and how can we help to deal with them and build a good media relationship also?				

in the Courting Phase, Novice in the “We”-Phase, Advanced in the “Me”-Phase, Master in the Rebel Phase (since the rebel ones tend to be those who know the system well and are capable of influencing others to follow them), and Professional in the Renewal Phase. It is vital for relationship building to understand which stage the franchisee is situated and allocate appropriate resources at different touch-points to help them perform their focuses of attention. The second dimension is the demand-and-supply value networks (Chen, Justis, and Wu, 2006), the attention structures of the franchise, consisting of customers, franchisee

outlet, franchisor headquarters, suppliers and partners, and franchise community. The main body of the framework is the focus of attention of the franchise of different levels.

BIG DATA STRATEGY IN FRANCHISING

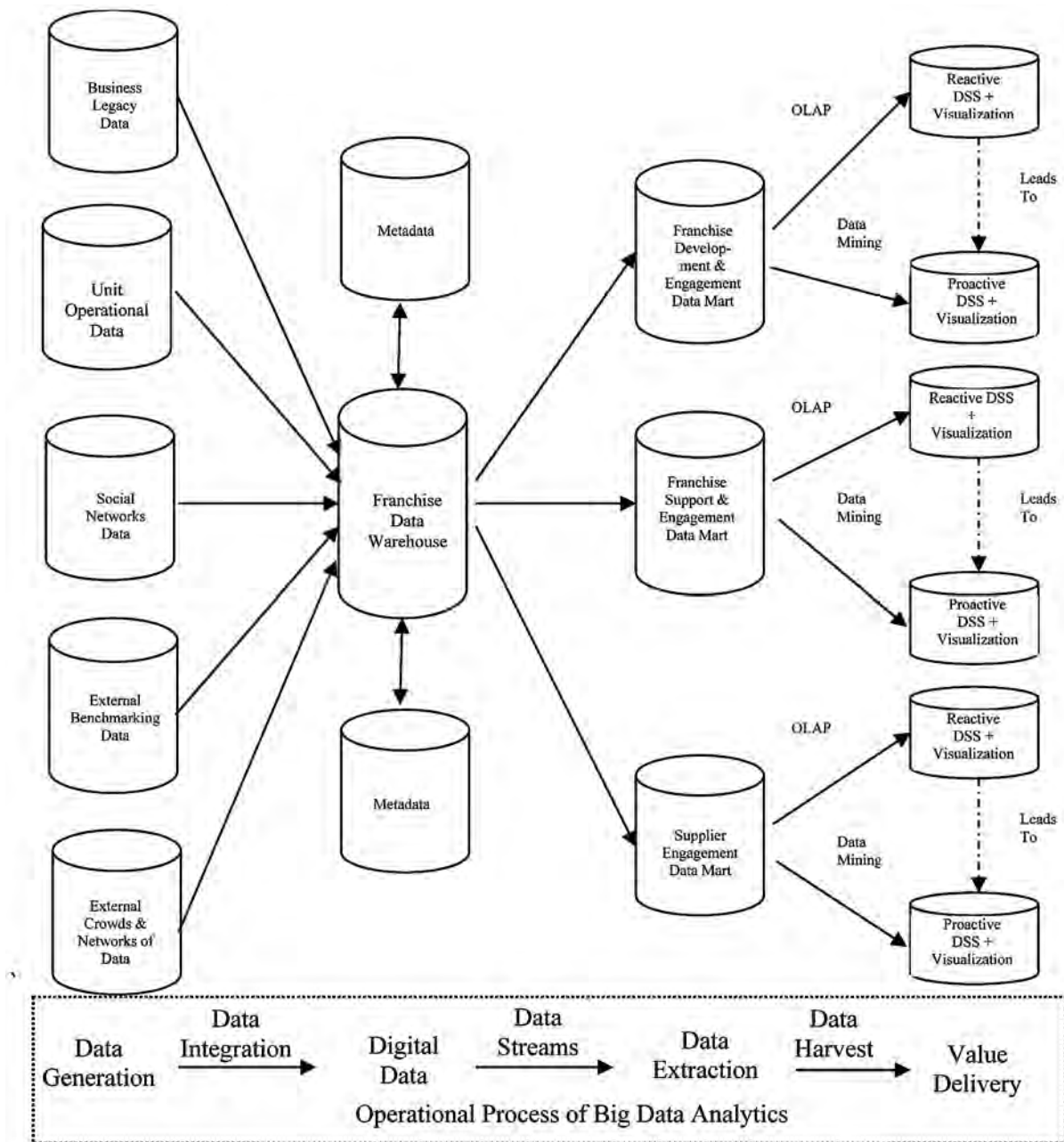
Leveraging big data for business growth is getting popular in franchising (O'Donnell, 2014; Saunders, 2015; Franchise Update, 2016). Here are some illustrative examples. Franks (2014)

studied Frozen Yogurts' big data practices and suggested empowering users such consumers and franchisees with direct access to the data and let them mix-and-match data and tools to create customized, and creative, analytic sandboxes. van Rijmenam (2015) studied how McDonald's leveraging big data, including using different metrics in their drive-thru system to optimize the customer experience; using predictive analytics to make effective changes in business and people practices; and making organizational changes in their menus, training programs and supply chain. Webster & Hume (2016) offered a framework of touchpoints to maximize the consumer engagements through social media. Kacker & Perrigot (2016) provided some insights on franchise use of the professional social media network LinkedIn.

As shown in Figure 2, Tables 1 and 2, there are many touchpoints of big data within the franchise business community where the franchisor and the franchisee can influence each other. An architecture of big data strategy in franchising is shown in Figure 3. The architecture consists of seven steps of operational process of big data analytics:

1. **Data Generation:** Data of different formats (e.g., numeric, text, audio, and video) may be generated from internal and external sources (Piccoli & Pigni, 2013). Internal data sources include unit operational data, generated during the daily activities at the franchisor headquarters; business legacy data, embedded in the activities that have been working well or gradually adapted since the franchise system came into existence; and social networks data, captured during important franchise meetings such as franchisee training and supplier annual conferences. External data sources include external benchmarking data, used to compare the relationship management activities in the franchise community; and external crowds and networks data, captured from social media such as LinkedIn and used for recruiting prospective franchisees.
2. **Data Integration:** Data of different sources are reconciled and integrated with respect to well-defined formats for easy retrieval. This is a very time consuming process that involves the activities such as recovery, cleansing, extracting, filtering, conditioning, scrubbing, and loading.
3. **Digital Data:** The formatted data is digitalized and stored in the franchise data warehouse with layers of metadata for efficient and effective data administration.
4. **Data Streams:** Digital data is segmented (Pigni, Piccoli & Watson, 2016) according to why it is segmented; where the segment is located; who is the identifier of the segment; what causes the segment to be created; how the event behind the segment is initiated and completed; and when the segment will be activated for value creation.
5. **Data Extraction:** Several data marts are created with data extracted (Piccoli & Pigni, 2015) from the data warehouse based on various franchisee/customer-centered segmentations. Data transformation is needed here that involves the activities such as exploration, replication, propagation, summary, aggregate, and metadata.
6. **Data Harvest:** Data analysis is needed here that involves the use of effective tools such as online analytical processing (OLAP), data mining (Chen, Zhang, & Justis, 2005; Chen, Justis, & Chong, 2008), Adobe Analytics and Predictive Analytics (Aman, 2016).
7. **Value Delivery:** Data harvest producing various relationship performance indicators with strong data visualization capabilities for decision making via decision support systems (DSS).

Figure 3.



FUTURE RESEARCH DIRECTIONS: CYBER SECURITY AND SUSTAINABILITY

There are two future trends necessary for franchises to pay attention to: cyber security and sustainability. Cyber security is the protection of

franchise information systems from damage or theft. It is necessary for the franchisor to incorporate cyber security management in implementing the e-business and big data strategy, e.g., setting up privacy protection policies (Webstersey & Melodia, 2013) and data security risk management (Adler, Demicco & Neiditz, 2015). In responding

to the growing demand for sustainability research and practices, PwC and APICS Foundation (PwC, 2014) surveyed 500 supply chain executives and developed a collaborative framework of sustainable supply chains to help companies prioritize their sustainability strategies. The framework consists of an end-to-end plan identifying environmental, social, economic, and ethical priorities for each of the following five processes: design and plan, source, make, deliver, and use and return. Such a holistic framework, when appropriately implemented, can help develop the franchise socially responsible brands which are important for millennials (Krishnan, et al, 2014).

CONCLUSION

Franchising has been popular as a growth strategy for businesses; it is even more so in today's global and e-commerce world (Chen, Chen, and Wu, 2005) with big data. The essence of franchising lies in managing the good relationship between the franchisor and the franchisee. In this paper we showed e-business and big data strategy plays an important role in growing and nurturing such a good relationship. Specifically, we discussed: (1) managing the franchisor/franchisee relationship through the CSLC approach, where organizational learning is believed to be the key to building the good relationship; (2) harnessing the e-business strategy around the CSLC approach, where four major components are discussed: benchmarking the requirements and acquisition stages, helping the franchisees serve their customers in the ownership stage and avoiding Internet encroachment, cultivating the ownership and renewal/retirement stages with effective knowledge management, and partnering with the "disruptive technology" providers to enhance the CSLC stages continuously; (3) aligning the CSLC-based e-business strategy with application service providers, where trust relationship is the major issue; and (4) leveraging big data to grow the franchise. Two future trends, cyber security and sustainability, were also discussed.

REFERENCES

- Adler, J., Demicco, M., & Neiditz, J. (2015). Critical Privacy and Data Security Risk Management Issues for the Franchisor. *Franchise Law Journal*, 35(1), 79.
- Aman, F. (2016). Naked Marketing: The Big Data Marketing Technology Stack. *Informatica Blog*. Retrieved from blogs.informatica.com/2016/02/12/naked-marketing-the-big-data-marketing-technology-stack/#fbid=ZIFRuYJY5pc
- Chen, Y., Chen, G., & Wu, S. (2006). A Simonian Approach to E-business Research: A Study in Netchising. *Advanced Topics in E-Business Research: E-Business Innovation and Process Management*, 1.
- Chen, Y., Chong, P., & Justis, R. T. (2000a). Information Technology Solutions to Increase Franchise Efficiency and Productivity. *Proceedings of the 2000 Franchise China Conference and Exhibition*.
- Chen, Y., Chong, P., & Justis, R. T. (2000b). Franchising Knowledge Repository: A Structure for learning Organizations. *Proceedings of the 14th Annual International Society of Franchising Conference*.
- Chen, Y., Chong, P., & Justis, R. T. (2002). E-Business Strategy in Franchising: A Customer-Service-Life-Cycle Approach. *Proceedings of the 16th Annual International Society of Franchising Conference*.
- Chen, Y., Ford, C., Justis, R. T., & Chong, P. (2001). Application Service Providers (ASP) in Franchising: Opportunities and Issues. *Proceedings of the 15th Annual International Society of Franchising Conference*.
- Chen, Y., Hammerstein, S., & Justis, R. T. (2002). Knowledge, learning, and capabilities in franchise organizations. *Proceedings of the 3rd European Conference on Organizational Knowledge, Learning, and Capabilities*.

- Chen, Y., & Justis, R. T. (2006). *Chinese Franchise Brands Going Globalization*. 1st Annual China Franchise International Summit, International Franchise Academy, Zhuhai, China.
- Chen, Y., Justis, R. T., & Chong, P. P. (2008). Data mining in franchise organizations. In J. Wang (Ed.), *Data Warehousing and Mining: Concepts, Methodologies, Tools, and Applications* (pp. 2722–2733). doi:10.4018/978-1-59904-951-9.ch169
- Chen, Y., Justis, R., & Wu, S. (2006). Value networks in franchise organizations: A study in the senior care industry. *Proceedings of the 20th Annual International Society of Franchising Conference*.
- Chen, Y., Justis, R. T., & Yang, H. L. (2004). Global E-Business, International Franchising, and Theory of Netchising: A Research Alliance of East and West. *Proceedings of the 18th Annual International Society of Franchising Conference*.
- Chen, Y., Seidman, W., & Justis, R. T. (2005). Strategy and docility in franchise organizations. *Proceedings of: 19th Annual International Society of Franchising Conference*.
- FB. (2014). The big data trends you need to know about. *Franchise Business*. Retrieved from www.franchisebusiness.com.au/news/the-big-data-trends-you-need-to-know-about
- Franks, B. (2014). What Frozen Yogurts Can Teach Us About Big Data Analytics. *Forbes*, 19(November). Retrieved from www.forbes.com/sites/teradata/2014/11/19/what-frozen-yogurts-can-teach-us-about-big-data-analytics-2
- Gates, B. (1999). *Business @ the Speed of Thought: Succeeding in the Digital Economy*. Warner Books.
- Hadfield, B. (1995). *Wealth Within Reach*. Cypress Publishing.
- Ives, B., & Mason, R. O. (1990). Can Information Technology Revitalize Your Customer Service? *The Academy of Management Executive*, 4(4), 52–69. doi:10.5465/AME.1990.4277208
- Justis, R. T., & Judd, R. J. (2002). *Franchising*. DAME Publishing.
- Justis, R. T., & Vincent, W. S. (2001). *Achieving Wealth Through Franchising*. Adams Media Corporation.
- Kacker, M., & Perrigot, R. (2016). Retailer use of a professional social media network: Insights from franchising. *Journal of Retailing and Consumer Services*, 30, 222–233. doi:10.1016/j.jretconser.2016.01.020
- Krishnan, V., Sullivan, U. Y., Aurand, T. W., & Gordon, G. L. (2014). Socially Responsible Brands: An Assessment Among Millennials. *Journal of Contemporary Business Issues*, 56.
- Lindsey, M. K., & Melodia, M. S. (2013). *Data Protection and Privacy in Franchising: Who Is Responsible?* Academic Press.
- Ocasio, W. (1997). Towards an attention-based view of the firm. *Strategic Management Journal*, 18, 187.
- Piccoli, G., & Pigni, F. (2013). Harvesting external data: The potential of digital data streams. *MIS Quarterly Executive*, 12(1), 143–154.
- Piccoli, G., & Pigni, F. (2015). *Selecting Digital Data Stream Winners*. SIM Advanced Practices Council.
- Pigni, F., Piccoli, G., & Watson, R. (2016). Digital Data Streams: Creating value from the real-time flow of big data. *California Management Review*, 58(3), 5–25. doi:10.1525/cmr.2016.58.3.5
- PwC. (2014). *Sustainable Supply Chains: Making Value the Priority*. Retrieved from www.pwc.com/us/en/operations-management/publications/sustainable-supply-chain.jhtml

Schreuder, A. N., Krige, L., & Parker, E. (2000). The Franchisee Lifecycle Concept – A New Paradigm in Managing the Franchisee-Franchisor Relationship. *Proceedings of the 14th annual International Society of Franchising Conference*.

Simon, H. A. (1971). Designing organizations for an information rich world. In *Computers, Communications, and the Public Interest*. Baltimore, MD: The Johns Hopkins Press.

Thau, B. (2014). How Big Data Helps Chains Like Starbucks Pick Store Locations -- An (Unsung) Key To Retail Success. *Forbes*, 24(April). Retrieved from www.forbes.com/sites/barbarathau/2014/04/24/how-big-data-helps-retailers-like-starbucks-pick-store-locations-an-unsung-key-to-retail-success/#50dae77e21c3

Thomas, D., & Seid, M. (2000). *Franchising for Dummies*. IDG Books.

TFO. (2014). Big Data: Transforming the Franchise Industry. *Top Franchise Opportunities*. Retrieved from <http://topfranchiseopportunities.com/blog/big-data-transforming-the-franchise-industry>

U.S. Commercial Service. (2016). *2016 Top Markets Report Franchising Country Case Study: China*. Retrieved from http://trade.gov/topmarkets/pdf/Franchising_China.pdf

van Rijmenam, M. (2015). From Big Data to Big Mac; how McDonalds leverages Big Data. *Dataflog*. Retrieved from <https://dataflog.com/read/from-big-data-to-big-mac-how-mcdonalds-leverages-b/403#!prettyPhoto>

Webster, G. B., & Hume, M. (2016). Analysing the Role of Social Media in Dialogue Marketing and Management as a Contemporary Franchising Local Area Marketing Technique. *Competitive Social Media Marketing Strategies*, 20.

Zeng, Q., Chen, W., & Huang, L. (2008). E-Business Transformation: An Analysis Framework Based on Critical Organizational Dimensions. *Journal of Tsinghua Science and Technology*, 13(3), 408–413. doi:10.1016/S1007-0214(08)70065-8

ADDITIONAL READING

FB. (2014). The big data trends you need to know about. *Franchise Business*, December 9 (www.franchisebusiness.com.au/news/the-big-data-trends-you-need-to-know-about)

Piccoli, G., & Pigni, F. (2013). Harvesting external data: The potential of digital data streams. *MIS Quarterly Executive*, 12(1), 143–154.

TFO. (2014). Big Data: Transforming the Franchise Industry. *Top Franchise Opportunities*, March 10 (<http://topfranchiseopportunities.com/blog/big-data-transforming-the-franchise-industry>)

KEY TERMS AND DEFINITIONS

Customer Service Life Cycle: Serving customers based on a process of four stages: Requirements, Acquisition, Ownership, and Retirement. Many companies are using the approach to harness the Internet to serve the customers.

Franchisee Life Cycle: The stages a franchisee goes through in the franchise system: Courting, “We”, “Me”, Rebel, Renewal.

Franchisee: The individual or business who receives the business rights and pay the royalties for using the rights.

Franchising: A business opportunity based on granting the business rights and collecting royalties in return.

Franchisor/Franchisee Learning Process: The stages of learning, including Beginner, Novice, Advanced, Master, and Professional.

Franchisor/Franchisee Relationship Management: The vital factor for the success of a franchise, including: Knowledge, Attitude, Motivation, Individual Behavior, and Group Behavior.

Franchisor: The individual or business who grants the business rights.

Facilitating Interaction Between Virtual Agents Through Negotiation Over Ontological Representation

Fiona McNeill

Heriot-Watt University, UK

INTRODUCTION

It is universally acknowledged that the problem of integration of information across large communities is a difficult and pressing one, particularly when these communities are disparate, widespread and not under centralised control, such as in the Semantic Web (Berners-Lee et al, 2001). The simplest solution to this problem is the enforcement of a single ontology: a single view of the world. However, if the agents interacting are from different organisations or fields, attempting to use a single ontology is usually neither practical nor desirable. Users need to develop a representation that is best suited to their own problems and they need to maintain and update that representation locally. Even if all users do subscribe to a single ontology, integration problems still exist, as changes and updates are made and users tune their ontologies to fit their own needs.

The problem of ontology matching has been widely studied and powerful solutions are available (see (Shvaiko & Euzenat, 2013; Euzenat & Shvaiko, 2013) for a comprehensive survey. However, the ontologies considered are almost always taxonomies, and the problem of ontology matching is concerned with relating a single term in one ontology to one or more terms in another ontology: for example, a term *car* in one ontology may relate to a term *automobile* or a term *carriage* in another ontology. Much less considered is the problem of relating compound terms such as first-order terms or database entries with multiple fields: for example, a term

car(make,model) in one ontology relating to a term *automobile(model,year,brand)* in another. In such situations we still have the problem of relating the single terms contained within these compound terms – e.g., this matching depends on knowing that *car* may be related to *automobile* and that *make* may be related to *brand*. But we must also consider the overall relation of the compound terms, which requires not only semantic but also structural matching.

Another drawback of traditional ontology matching in an online environment is that it tends to assume full knowledge of all relevant ontologies and is generally performed off-line, prior to interaction. These are the assumptions made by the main evaluation processes for Ontology Matching, such as the Ontology Alignment Evaluation Initiative (OAEI¹). But in large, fast-moving agent communities, or situations where some information may be confidential, we cannot assume that we can have full knowledge of any agent or service we may interact with, nor is it possible to perform the matching off-line if we may not know prior to interaction which agents will need to interact.

In this paper, we introduce our theory of on-the-fly, structured matching and briefly describe the ORS system, which we have developed to implement this theory. Our central hypothesis is that representation – as well as vocabulary and beliefs – must be treated as a fluent and that automated, dynamic matching techniques that can map between structured terms are necessary for full integration of disparate ontologies (Bundy et al, 2006).

BACKGROUND

In a system such as the Semantic Web, where there is no centralised control, we cannot have a complete global overview of the agents and data in the system. Agents may join and leave the system freely and they will all have their own ontologies and data that may be large and complex and may be confidential. We cannot hope for a complete description of the relations between every agent in the system. Our approach is therefore not to consider how such a system can be controlled but how an individual agent can successfully make its way in such a system, interacting with the agents that it needs to interact with, even if these agents are not using the same ontological terms or representations, and even if it is not known in advance of the interaction which agents these will be.

Although many existing ontologies are simple taxonomies, and matching these ontologies is a crucial task, we believe that this kind of matching cannot be sufficient. Agents that are capable of interacting in complex and unpredictable environments need to be able to plan, and planning agents need far richer descriptions of the world: not only taxonomies of classes but also relations and functions between these classes, and planning rules describing how to influence the world. Uniformity of these relations, functions and rules can no more be assumed than uniformity of terms within taxonomies, and therefore matching between these structured objects is just as crucial as the more frequently addressed problem of matching between simple terms within a taxonomy.

Matching large ontologies at run-time, particular ones that contain structured terms, is generally not feasible, but we make this problem tractable by only fixing mismatches when this is demonstrated to be necessary. Since interactions may be frequent and fleeting, there may not be much value in matching the full ontologies, since the interaction that is desired may only require a very small part of the ontologies, and even if inconsistencies exist between the ontologies, these may not lead to communication breakdown between agents during

a particular interaction. Our approach is therefore to diagnose mismatches and refine the ontologies accordingly only when these mismatches directly cause communication breakdown.

To this end, we developed ORS (the Ontology Repair System)². This is a tool that an individual agent (which we name PA – the planning agent) can make use of as an aid when communication breaks down. ORS tracks the course of the communication between PA and any agents it may be interacting with (we name these SA – service-providing agents). If communication proceeds successfully then ORS does not need to be utilised. However, if communication breaks down, ORS begins the diagnostic process, analysing the communication so far and prompting PA to ask further questions in order to pin down a specific mismatch between the ontologies of PA and SA, which is then corrected.

The benefit of ORS is therefore that it allows an agent to interact successfully with other agents, even when their ontologies are mismatched in important ways, and even when this mismatch is between complex, structured ontological objects as well as when this mismatch is between simple terms. It works on the fly and fully automatically even when interactions are unpredictable and unforeseen.

ONTOLOGICAL MISMATCH

Planning agents require ontologies that contain three different kinds of objects, which entail three different kinds of mismatches:

1. Purely semantic mismatch, where the mismatch is between words or phrases – for example, *car* is matched to *auto*: this is the problem that is covered by conventional ontology matching.
2. Structural mismatch, where the mismatch is between structured terms (such as relations and functions) – for example, *car(Make,Model)* is matched to *car(Make)*,

or $car(Make, Model, Year)$ is matched to $car(Make, Model, Date)$.

3. **MISMATCHES** of planning rules, where one agent has a different idea of the conditions and effects of performing an action to another – for example,

```
BuyTicketAction: wants_to_
travel (Me, Destination) → has_
ticket (Me, Destination) matched
to BuyTicketAction: wants_
to_travel (Me, Destination)
and has_money (Me) → has_
ticket (Me, Destination) 3.
```

Of these, the first point is only considered incidentally, due to the large body of work that already addresses this issue; our emphasis is strongly on the second and third points.

Structured Matching

The problem of structural mismatch within ontological mismatch is crucial to successful interaction of agents or services. Not only are the utterances of agents usually structured, but service invocations are also necessarily structured, and their automatic interaction requires structured matching of just the type our work addresses; therefore, semantic matching alone cannot be sufficient.

In our work, we consider quantifier-free, first-order terms; that is, predicates with some number of arguments ≥ 0 . Most common service invocations, such as those expressed in BPEL⁴, as well as most types of database entries, can be expressed in such a way. Our techniques are therefore very widely applicable.

The space of possible mismatches between one first-order term and another, more general, first-order term is described by the theory of abstraction (Giunchiglia and Walsh, 1992). They describe four kinds of mismatch:

- **Propositional Abstraction:** A term is matched to one with fewer arguments – for example, $car(Make, Model)$ maps to $car(Make)$.
- **Predicate Abstraction:** A term is matched to one with a more general predicate – for example, $car(Make)$ maps to $vehicle(Make)$.
- **Domain Abstraction:** A term is matched to one with a more general type of argument – for example, $car(Make, Second-hand-dealer)$ maps to $car(Make, Dealer)$.
- **Precondition Abstraction:** A term is removed from a rule – for example, $has(money, Me) \rightarrow owns(car(Make), Me)$ maps to $has(money, Me) \ \& \ has(id, Me) \rightarrow owns(car(Make), Me)$.

By inverting these relationships, we obtain four dual refinement operators. These abstraction and refinement operations are sufficient for describing most ways in which quantifier-free, first-order terms can be related. Non-identical terms must be either synonymous, more general, less general or unrelated. The case where they are synonymous is dealt with by semantic matching. There are a limited number of attributes of first-order terms: they have predicate names, numbers of arguments, orders of arguments and types of arguments. Variations in these attributes gives the above classification of how first-order terms may be mismatched.

WORKED EXAMPLE

Consider the interaction between a planning agent (PA) which wishes to buy a ticket, and an agent acting as the front-end to a ticket-selling service (the service providing agent – SA). PA contacts SA with the following message:

PA: buy(pa, london, edinburgh)⁵

indicating that PA wishes to buy a ticket between London and Edinburgh. It is likely that SA will have some conditions on selling tickets, for example that the buyer has money, and must verify that these conditions hold. SA will therefore respond with a question:

SA: *money(pa,Amount)?*

This indicates that SA wishes PA to find a suitable instantiation of the variable *Amount* such that the relation *money(pa,Amount)* becomes true. But imagine that PA has money represented as a predicate *money(Agent,Amount,Currency)*, and can instantiate this to *money(pa,100,dollars)*. Perhaps this agent is used to operating in many different currency zones, whereas SA only operates in one currency zone, so their designers chose different representations. This question from the SA is considered a *surprising question* because it does not directly match any of the preconditions in the PA's rule.

PA must therefore respond negatively to SA's enquiry:

PA: *no*

SA: *fail: buy(pa,london,edinburgh)*

However, PA is able to analyse the problem and alter its ontology appropriately. In this instance, it is obvious that the problem is that *money/2* is mismatched to *money/3*. PA can determine that the two arguments of SA's *money/2* match the first two arguments of its *money/3*, and hence the third argument – *Currency* – is redundant. It therefore alters its definition of *money* from *money(Agent,Amount,Currency)* to *money(Agent,Amount)*, and alters all instantiations of this predicate accordingly. This means that it is removing information (that the currency is dollars) that may be important in subsequent interactions. This could be resolved by retaining a copy of the old *money/3*, whilst using the updated *money/2* in the current interaction. Additionally, it is making the assumption that the *Amount* referred to by SA

is in dollars. If the assumption in SA's design that money is in dollars is implicit, it cannot be accessed automatically. Confirmation of this would therefore involve user intervention.

The interaction can then be resumed and this time PA will be able to respond appropriately.

(Note that several important steps, such as discussion of price, have been omitted from this simplified example).

THE ONTOLOGY REPAIR SYSTEM

In this section, we describe the role of ORS through giving a step-by-step description of how it is used by an agent (PA) within a multi-agent, semantic-web-like environment.

1. PA has a goal, or is provided by the user with a goal, and uses its ontology to form a plan to achieve this goal. The steps of this plan will normally involve interaction with other agents. For example, if PA has to organise a journey, it will need to interact with (among others) a ticket-selling agent. Each of PA's planning rules will describe the circumstances under which it believes these other agents (SAs) will perform their roles (e.g., a ticket-selling agent will sell a ticket if it is provided with an appropriate destination and sufficient money). If there are no mismatches, the PA's rule will be compatible with the SA's rule which the SA is actually using to perform the task.
2. In order to achieve the goal, PA will execute each step by communicating with the appropriate SA.
3. PA will request the service it requires from the agent, and then wait for a response. This could be an indication that the service has been performed, but it will usually be further questioning from the SA, which is trying to establish whether or not the preconditions are satisfied, to confirm that it can perform the service.

4. If this process results in the service being successfully performed, PA proceeds to the next step of the plan and the functionality of ORS is not required at this step. However, if failure occurs, PA invokes ORS to attempt to track down the cause of this failure.
5. ORS will analyse all the questioning that has occurred so far, giving particular attention to questions the SA asked that were not expected. Expected questions would tie in with the preconditions of PA's rule for the relevant service; any questions deviating from these, either substantially or through a structural difference, would be a *surprising question (SQ)*. An example from the previous section is the question *money(pa,Amount)?* being asked when the question *money(pa,Amount,dollars)?* was expected; alternatively, a completely unexpected question might be asked.
6. ORS's diagnostic algorithms (briefly described below) use this information, prompting the PA to ask for further information if necessary, to diagnose the problem

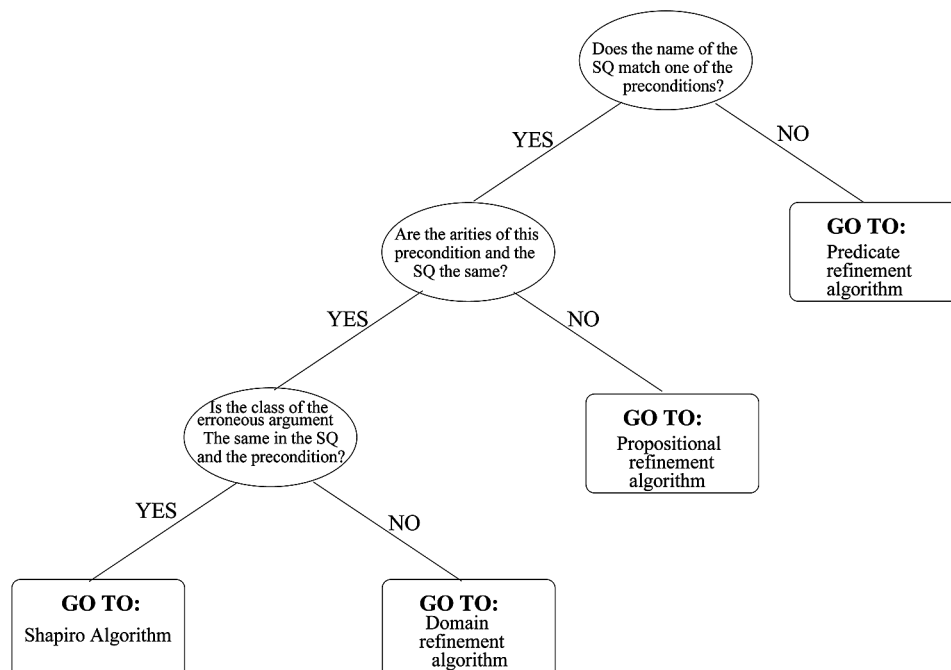
and to repair PA's ontology accordingly. Occasionally, the fault can be tracked down to a particular ontological object (in the worst case, the whole rule), but an appropriate fix cannot be found (we know there is a mismatch but do not have enough information to determine exactly what it is). This ontological object is then marked as unusable.

7. PA uses its updated ontology to replan from its current state (so earlier successful plan steps will not be repeated) and the process begins again. This time, it is hoped that the problematic interaction will be more successful, though it may, of course, fail again due to other mismatches, which in their turn must be identified and diagnosed.

FAULT REPAIR AND DIAGNOSIS

The diagnostic process works through a series of algorithms. The flow chart in Figure 1 describes one of these algorithms; this particular flow chart is invoked when the diagnostic process has

Figure 1. Flow chart illustrating part of the diagnostic algorithm



already determined that a precondition of a rule is at fault. This algorithm determines exactly how it is at fault – is it incorrectly named? is one of the arguments wrong? is the number of arguments wrong? When the diagnostic process has narrowed down what kind of mismatch it is, it next needs to determine the details of the mismatch. Predicate, propositional and domain refinement mentioned in Figure 1 all refer to types of mismatches outlined above; the Shapiro algorithm⁶ is used when there is no structural mismatch but an incorrectly instantiated fact. It determines how this incorrect fact came to be believed. Further details of these processes and the other diagnostic algorithms can be found in (McNeill and Bundy, 2007).

NEGOTIATING REPAIRS

The basic version of ORS makes the assumption that only PA makes repairs: if mismatches were found, it was SA's version which was preferred. In practice, however, this may not always be the best option. It may be that making the proposed change would be very costly to PA – for example, it may require changing an ontological object that is frequently used, and changing it may complicate many other interactions, or the proposed change may involve removing detail that is of high value to PA. PA making the diagnosed repair is not the only way to make the interaction possible: SA making the inverse of the repair would also create an understanding between them that would allow the action to be performed. If, in these situations, it was the case that making the inverse change was not costly to SA then the greatest good would be achieved by SA performing the change rather than PA. We refer to this as *maximising social welfare*.

ORS is a system designed to be used by a particular agent and we do not make any assumptions about the agents with which it interacts, so we cannot assume that they also have access to ORS. If they don't then they cannot implement repairs themselves and negotiation is impossible: PA must either perform the repair itself or decline

the interaction. However, in cases where SA does have access to ORS, negotiation is possible.

In order to reflect the idea that there is a greater cost associated with changing some parts of the ontology than others, we have introduced the notion of *protection*. An ontological object that is marked as protected will invoke a higher cost if it needs to be altered than one not. There are two levels of protection - high and low. Protection may be applied to an entire ontological object (e.g., a predicate definition or a rule) or to a part of it (e.g., a particular argument of a predicate or a particular precondition in a rule).

With this extended version of ORS, if PA diagnoses a repair that would lead to the alteration of any protected object, it attempts to negotiate with SA to see if SA would implement the inverse repair instead. Because PA is already providing the mismatch information, SA does not need have access to any diagnostic ability, merely the ability to repair its ontology appropriately. In an open environment, SAs will make such decisions on whatever basis they chose. For the purposes of our implementation, we designed our own SAs which would be prepared to perform the mismatches if they had no relevant protection, and would refuse if they had relevant protection at any level. If SA refuses to perform the inverse repair, PA will make the repair itself if it has only low protection; if it has high protection, it will mark SA as inappropriate and try to re-plan.

We have developed a negotiation protocol through which PA and SA can discuss possible allocations of repairs. The negotiation protocol specifies the set of permissible offers and possible agreements. In addition, it provides the agents with the essential vocabulary and the rules for a successful negotiation. The negotiation outcome - which agent will perform the repair - is always the result of mutual agreement. The strategy that each agent follows is based on a utility-based model: each agent forms its utility for the negotiated repair, and proceeds to negotiate according to this utility. In this way the agents express their preferences. We have also introduced the

notion of an *expert agent*, where if PA and SA are not sure which of their versions is preferable, they can get the advice of an impartial agent. An expert agent can be any agent (other than PA or SA), nominated by either agent, and will be relevant to the particular repair. For example, if the concern is about a predicate representing *money*, a *banking agent* may be consulted.

The lifecycle of negotiation is as follows.

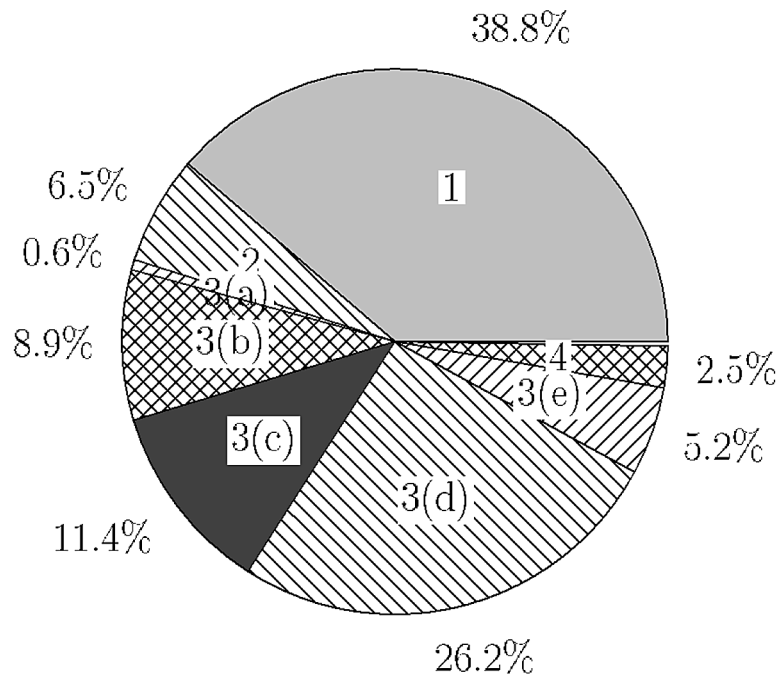
- PA calculates a cost for the repair in the range [0 2]. A cost of zero indicates that no protection is violated, a cost of 1 indicates that low protection is violated and a cost of 2 indicates that high protection is violated.
- PA then calculates the *utility* of repair, which is the cost of the action not being performed less the cost of performing it (currently this is set at 2 for all actions).
- If the cost of repair is higher than zero, PA initiates the negotiation process, sending a suggested repair to SA.
- Once SA receives this, it calculates its own cost and utility scores for implementing this repair. In our current implementation, SAs calculate this score in the same manner as PA; in an open environment, they can choose their own method of calculating it, providing the utility score is in [0 2]. If its utility score is zero, it has no incentive to perform the repair and will send a rejection notice to PA. If its utility score is greater than zero, it will enter into negotiation with PA, sending back a *proposal* for interaction, which contains this utility score.
- PA then calculates the *social welfare* score for either agent performing the repair, where the social welfare of an agreement is a measure of the utility this agreement has for the system in total.
- If the social welfare is maximised by PA performing the repair, it will do so. If the social welfare is maximised by SA performing the repair, PA will request SA to do so. If the social welfare is equal either way then it

is not clear which of the agents should perform the repair (SA can agree or decline). In this situation, the agents may choose to consult an expert agent, which will indicate a preference. It is acceptable for either agent to make the repair. However, since the benefits of the different representations is not clear, we introduce the notion of the *expert agent* to assist in such situations. Both agents have the choice as to whether to accept this recommendation; neither is obliged to recognise the expert agent as an appropriate authority. If both agents do agree to take the expert agent's opinion into account, the utility of the agent with the preferred version is increased by 1, and the negotiation continues on that basis.

In order to make our approach flexible, and applicable to agents in an open environment that have not been designed with this specific interaction in mind, we created an ontology which defines the negotiation protocol, so that this is available to all agents without it being hard-coded into the agent.

1. ORS could refine the mismatch
2. ORS could not refine the mismatch but straightforward improvements to ORS would allow it to solve it.
3. ORS could not refine the mismatch because:
 - a. ORS did not have sufficient functionality;
 - b. This mismatch is outside the current scope of ORS;
 - c. This mismatch is irrelevant to an automated system – this is usually a change to commenting or formatting;
 - d. This mismatch could not occur in the representation ORS currently deals with;
 - e. This mismatch could not be highlighted in a planning context.
4. The information we had about the mismatch was insufficient to determine which category it would fall into.

Figure 2. Piechart illustrating the percentage of mismatches that fall into the following categories⁷



EVALUATION

The standard approach to evaluation of ontology matching, outlined in the OAEI⁸ is to input two full ontologies and evaluate how many of the mismatches between them are correctly or incorrectly diagnosed, or are missed. Such evaluation makes no sense for ORS, since ORS is designed not to discover and patch all mismatches but only those that are impeding interaction. A better metric for ORS, therefore, is to evaluate how frequently it can facilitate interaction that would have failed without access to its functionality.

FUTURE RESEARCH DIRECTIONS

We examined mismatches between several large ontologies and analysed which of these mismatches were described by mismatches ORS could diagnose and refine, and which were not. The pie chart below illustrates the results. Although this highlights that there is much ORS cannot currently do, we believe these results to

be encouraging. This is the first step towards a new approach to ontology mismatch and ORS is a prototype system, so the fact that it can already tackle 38.8% of mismatches bodes well.

Evaluation of the negotiation extension demonstrated that the agents always acted so as to maximise our definition of maximal social welfare. In addition, we used computational tree logic to prove that: deadlock is never reached; when a call for a proposal is received, the receiving agent can always refuse this call or make a proposal; when a proposal is received, the receiving agent can always either refuse it or accept it.

Full details of our evaluation can be found on the ORS website⁹.

CONCLUSION

In this article, we have introduced ORS, a new approach to ontology mismatch which aims to resolve miscommunication between agents, where this occurs due to ontology mismatch. ORS works by diagnosing mismatches and repairing

ontologies during runtime, but only where this is demonstrated to be necessary. ORS is designed to be a tool that an agent interacting in an uncertain world can rely on to assist it when communication breaks down due to misunderstandings.

We have described the kinds of mismatches ORS can diagnose and refine and have briefly outlined promising evaluation results. We have many ideas for increasing the functionality of ORS which should greatly improve these evaluation results, including broadening the scope of the kinds of ontologies ORS can deal with and building in improved semantic matching techniques by incorporating existing semantic matches. Further details of all aspects of ORS, together with full information about our plans for future work, can be found in (McNeill and Bundy, 2007).

REFERENCES

- Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. *Scientific American*, 284(5), 34–43. doi:10.1038/scientificamerican0501-34 PMID:11396337
- Bundy, A., McNeill, F., & Walton, C. (2006). On repairing reasoning reversals via representational refinements. In *Proceedings of FLAIRS 2006 (The Florida AI Research Society Conference)*. AAAI Press.
- Euzenat, J., & Shvaiko, P. (2013). *Ontology matching* (2nd ed.). Springer. doi:10.1007/978-3-642-38721-0
- Giunchiglia, F., & Walsh, T. (1992). A theory of abstraction. *Artificial Intelligence*, 57(2-3), 323–389. doi:10.1016/0004-3702(92)90021-O
- McNeill, F., & Bundy, A. (2007). Dynamic, automatic, first-order ontology repair by diagnosis of failed plan execution. *International Journal on Semantic Web and Information Systems*, 3, 1–35.
- Shvaiko, P., & Euzenat, J. (2013). Ontology matching: State of the art and future challenges. *IEEE Transactions on Knowledge and Data Engineering*, 25(1), 158–176. doi:10.1109/TKDE.2011.253

KEY TERMS AND DEFINITIONS

Ontology: Formal representation of domain knowledge, containing a class hierarchy and possibly relations and functions between these classes. ORS: Ontology repair system which equips a planning agent to proceed successfully with problematic interaction by identifying and repairing any ontological problems which may have led to the difficulties.

Planning Agent (PA): An agent which responds to goals through forming plans to achieve them and then enacting these plans through interaction within a domain.

Semantic Mismatch: Mismatch between two ontological terms where the structure is the same but the meanings of the words within the terms is not.

Service Providing Agent (SA): An agent which is able to provide a service to a planning agent, providing the correct conditions are met.

Structural Mismatch: Mismatch between two ontological terms where the meanings of the words within the terms is the same but their structure is not (two ontological objects may be mismatched both semantically and structurally).

Structured Ontological Term: An ontological term, such as a relation or function, the meaning of which is determined not only by the meanings of the words in the term but also by their structure.

Surprising Questions: When two agents are communicating with a particular goal in mind, a question that is posed by one agent which is not thought to be pertinent to the situation by the other agent is considered by that agent to be a surprising question.

ENDNOTES

- ¹ <http://oaei.ontologymatching.org/>
- ² <http://dream.inf.ed.ac.uk/projects/dor/>
- ³ Note that the $\rightarrow$ in these rules implies the performance of an action: the predicates on the left-hand side must be true before the action is performed; the predicates on the right hand side are made true after the action is performed.
- ⁴ Web Services Business Process Execution Language Version 2.0, <http://docs.oasis-open.org/wsbpel/2.0/wsbpel-v2.0.pdf>
- ⁵ Note that we use a lower case initial letter to indicate a constant and an upper case initial letter to indicate a variable. In theory, the name of the variable can be anything, but we use the convention of using the type name

so that the type information is immediately apparent. Thus *money(pa,Amount)* indicates that the predicate *money* relates a specific agent named *pa* has an unknown value which must be of type *Amount*.

- ⁶ This is so named as it is loosely inspired by Shapiro's procedure for debugging programs by tracing back to find the original source of the problem.

- ⁷ The differences between these categories and the process by which we determine how to categorise mismatches are not obvious at first glance. Full details of how this categorisation is done can be found on the ORS website (above).

- ⁸ <http://oaei.ontologymatching.org/>
- ⁹ <http://dream.inf.ed.ac.uk/projects/dor/>

On-Line Credit and Debit Card Processing and Fraud Prevention for E-Business

E

James Williams

University of Pittsburgh, USA

INTRODUCTION

Credit and debit cards are Americans' top choice for online shopping, with 48 percent in 2014 preferring to use credit cards, 30 percent using debit cards and 12 percent using PayPal. (Total System Services, Inc., 2014). With their Visa and MasterCard logos, debit cards look like credit cards, but they do not draw money from the same source as credit cards. Debit cards, sometimes called a check or bank card, draw funds from the user's bank checking account, not a line of credit although prepaid debit cards do not access the user's bank account. Many debit cards are actually dual debit/credit cards. Users can use these dual debit cards as one or the other. If it is used as a debit card, the user must enter the personal identification number (PIN) to authorize the transaction (Velocity Payment services, 2016). The use of EMV (Europay, MasterCard, and Visa) chips on cards has pushed more cyber-criminals to attacking on-line (Card Not Present) for fraudulent purposes.

The 10 largest U.S. merchant acquirers ranked by general purpose transactions tied to PIN- and signature-based debit cards and credit cards accounted for 64.61 billion transactions in 2013 (Federal Trade Commission, 2012). Accepting credit and debit cards is essential for any e-commerce Web site. Processing credit and debit cards over the Internet is one of the fastest growing segments of business transactions. This type of transaction or "card-not-present" transaction requires a special type of merchant account. Also, if someone steals a debit card number, an entire bank account is vulnerable. Although prepaid debit cards are also available (Wikipedia, 2016). Debit cards are

basically processed in the same manner as credit cards (Velocity Payment services, 2016). In the early days of credit and debit card usage, to accept such cards, a merchant needed a merchant account through a bank. But today there are a number of services, generally referred to as credit/debit card processors or merchant account services, which will permit a merchant to accept credit and debit card payments online without their own merchant account. There are actually three different methods for processing credit and debit card payments using a merchant account service. These are:

1. **Real-Time Processing:** Real-time processing allows e-commerce merchants to link their e-commerce shopping cart with a gateway merchant service which will automatically process card payments.
2. **Virtual Terminal (Online Interface):** An e-commerce merchant can also process card transactions, manually, 24 hours a day by logging in online and submitting a secure form through a merchant account interface. A merchant can use this to process card payments while taking the customer's information over the phone if the merchant is able to access the Internet at high speed while talking to the customer.
3. **Automated Recurring Billing (ARB):** Some e-commerce merchant services need to charge customers on a monthly or account threshold basis. Some merchant account services allow the merchant to set the time interval or account threshold level and some services allow a merchant to upload multiple subscriptions using a batch file like Microsoft Excel.

DOI: 10.4018/978-1-5225-2255-3.ch236

PayPal is generally accepted as the most widely used online merchant account service with more than 150 million users across the world. VeriSign operates a competing service called Payflow that is typically used by merchants with a high volume of transactions each month. Although the number of merchant account service providers continues to increase, some of the more popular one are listed below (Smith, 2016; Williams and Premchaiswadi, 2009):

- Flagship Merchant Services
- Gomerchant Merchant Accounts
- Merchant Accounts Express
- Cayan
- Electronic Transfer Inc.
- E-Commerce Exchange
- NorthAmerican Bancard
- Charge.com
- TSYS
- Free AuthNet
- Merchant Credit Card
- Payment Depot
- Helcim
- Transfirst
- Dharma

Companies that sell merchandise and services over the Internet are referred to as e-tailers or e-commerce merchants. These credit and debit card processing services make it easy for e-tailers to start accepting cards for purchases of their products and services.

BACKGROUND

Who Are the Participants in On-Line Credit/Debit Card Purchases?

Consumers and Merchants

The consumer is an individual or organization that has the intent of making a purchase. They have money or credit and they desire goods and

services. The merchant is the one with the goods and services and is looking to sell them to consumers. Consumers are motivated to select a particular merchant by things such as price, service, selection or preference. But the merchant's primary motivation is to make money by selling the goods or services for more money than they paid for them. This money between what they bought it for and what they sold it for is called their margin. There are several different methods to exchange money for products and services such as bartering, cash, checks, debit cards, installment payments or credit cards. When credit and debit cards are used, the consumer and the merchant both have banks that they are working with to process the credit/debit card payment transactions.

Issuing Bank

Consumers typically get their credit/debit cards from a bank or credit union, called the "issuing bank." Sometimes an issuing bank is simply called an "issuer." An issuing bank may not only be associated with major credit card brands such as American Express, MasterCard and Visa, but also with credit cards called "private label credit cards." These are the ones that department stores or shops offer, such as Sears and Target credit cards. Issuing banks are lending institutions that support these credit cards by granting and managing extended credit or maintaining bank accounts for debit cards. The purpose of the issuing bank is to grant a credit or debit card directly to a consumer. The issuer typically has a consumer fill out an application, check the applicant's credit and banking history and maintain their account. The issuing bank is the one that decides what a consumer's credit limit is, based on credit history and current debt load. There are thousands of issuing banks in the United States. In Canada and the United Kingdom as well as most other countries in the world there are far fewer banks. Issuing banks make money on the interest the consumer pays on outstanding balances from previous purchases, and they get a portion of every purchase

a consumer makes with the credit card from a merchant. Debit networks charge an interchange fee that consists of a percentage and a flat transaction fee. Debit networks such as NYCE, STAR and ACCEL maintain computer networks to route transaction information among businesses and banks. A debit network charges an interchange fee and a switch fee to route transaction information over its network. The debit network interchange is very similar in structure and complexity to the interchange fees charged by Visa and MasterCard (Cardfellow, 2016).

Acquiring Bank

The acquiring bank represents the e-commerce merchant. The acquiring bank processes all of the merchant's credit/debit card payments with the associations (American Express, MasterCard, Visa, etc.), and provide the merchant with reconciliation data and other financial tools. The acquiring bank also makes money on every transaction a merchant processes. There are many acquiring banks in the United States and abroad, and merchants are free to move from one acquirer to another. Merchants typically select their acquiring bank based on the amount of money, called basis points, they charge per transaction.

Payment Processors and Gateway Services

Theoretically, e-commerce merchants can connect directly to their acquiring bank, but there are a number of reasons why they may not want, or be able, to do so. There are technical and business requirements for conducting the payment process for credit/debit cards and most merchants don't want to deal with these requirements. As an alternative, they use a third party to process credit/debit card payments for them and their acquiring bank. These third parties are called credit/debit card payment processors and gateway services. Credit/debit card payment processors offer the physical infrastructure for the merchant to com-

municate with the acquiring banks and the credit card associations. They connect all the credit/debit card payment participants together. This permits even very small banks to offer merchant services that they could not provide otherwise. Some credit/debit card payment processors also provide acquiring bank services directly to the merchant.

Gateway services provide merchants physical infrastructure as well. They generally offer technology and integration services among all the participants. The gateway service providers charge the merchant a transaction fee or basis points for their services. These fees are in addition to the credit/debit card payment processor fees the merchant is already paying. If a merchant decides to use a gateway service provider they have to set up accounts with an acquirer. The acquirer can be an acquiring bank or a credit/debit card payment processor that offers acquiring bank services.

Credit Card Associations

The credit card associations such as Visa, MasterCard International, American Express, Discover, etc. are responsible for establishing the procedures and policies for how credit card transactions, services and disputes are handled. They are bound by national banking laws and provide the money that covers some of the fraud that occurs within their membership. Each of the credit card associations operate somewhat differently and even within the same association they may operate differently in different parts of the world.

For example, Visa has regions that operate somewhat autonomously. There is Visa U.S.A., Visa Europe, Visa Asia, etc. Each of these regions has slightly different rules, tools and services. Visa does not actually issue credit cards to consumers; they use issuing banks to issue credit cards that are branded as "Visa." MasterCard International is somewhat different from Visa in that there is one association for the entire world with all regions using the same basic structure, policies, and management procedures. MasterCard International also uses issuing banks to issue credit cards

to consumers that are branded as “MasterCard.” American Express differs by acting as the issuer for all American Express branded credit cards. American Express is one global organization with regional coverage. American Express also differs from Visa and MasterCard in permitting merchants to set up direct connections for performing the acquiring functions.

How the On-line Credit/Debit Card Payment Process Works

When a merchant makes a sale over the internet; the card number, the amount of the sale, and the merchant identification (ID) are transmitted from the merchant’s establishment or the internet Web site over the credit/debit card processor’s computer network. The credit/debit card processor can either be a bank or a merchant account service company called a credit/debit card processor that does nothing but provide credit/debit card processing services as discussed above.

From the credit/debit card processor’s network the transaction is transmitted to the credit/debit card processing company’s computer network. If the customer is using MasterCard, for example, the transaction will go to MasterCard’s computer network. Then, the electronic transaction is sent to the bank that issued the credit/debit card to the customer. The bank’s computer system checks the account and verifies that the customer has adequate funds to cover the purchase. The bank’s computer system then sends the merchant an authorization over these same networks. Although the sale is complete, the transaction is not complete since no actual money has been exchanged.

At the end of the business day the merchant account service (credit/debit card processor) sends that day’s charges to the credit/debit card network, e.g. MasterCard, for processing. The transactions are transmitted via the merchant’s credit/debit card processor service to the credit/debit card network. Individual transactions are then extracted and sent back to the individual cardholders’ banks. The issuing banks then debit the cardholders’ accounts

and make appropriate payments to the merchant’s credit/debit card processor through the Federal Reserve Bank’s Automated Clearing House.

The credit/debit card processor then credits the merchant’s bank account for the transaction amount, minus its fees for the transaction. Those fees are also used for paying transaction fees to the issuing bank and the credit/debit card processing network. Despite the use of computers, it can take two business days before the merchant’s account is credited (Murdock, 2006).

Opening a Merchant Account

In order to accept credit/debit cards, a merchant can open a merchant account with a bank. However, many banks have gotten out of the credit card processing business, and those that remain are often reluctant to service small businesses, particularly ones with limited operating histories. Many small businesses must therefore go through a specialized credit/debit card processor or an independent sales organization, commonly referred to as an “ISO.” Whether a merchant uses a bank, ISO or a credit/debit card processor, they need a merchant account before they can accept credit/debit card payments.

An ISO or an Independent Sales Organization is an entity that acts more or less as a middle man, helping formulate a Bank or Bank/Credit or Debit Card Processor alliance. Within such an arrangement, an ISO has an agreement to sell the services of the Bank or Bank/ Credit Card Processor alliance, and is allowed to mark up the Fees and sign up merchants.

- **Tier I - ISO:** Also known as a Super ISO, Wholesale ISO, Full Liability ISO, and a Full Service ISO, a Tier I - ISO always does their own underwriting and risk-assessment and assumes full chargeback liability for their merchants and provide full technical support.
- **Tier II - ISO:** These are shared liability ISOs. Usually, they do not do their own

underwriting, or require underwriting approval from the ISO or credit card processor with which they are contracted. They provide technical support capabilities and they also have the support from the ISO or credit/debit card processor with which they are contracted.

- **Tier III - ISO:** These are usually comprised of a few salespeople with no technical support to provide to their merchants, Tier III - ISOs also do take any responsibility for any chargeback risk.

Typical Information Required for a Merchant Account

Getting the required information together before applying for a merchant account can save time during the application process. Although different merchant account providers have different requirements typically what follows are required in order to obtain a merchant account:

1. Business checking account
2. A copy of a voided check
3. A copy of the company's Articles of incorporation, business license or reseller license. The purpose of this is to prove the applicant is a legitimate business.
4. Pictures of business office and location
5. Have a web site and URL
6. Photocopy of the merchant's return policy
7. Provide business references
8. Photocopy of recent tax returns
9. Site inspection (pictures of your inventory).
10. A photocopy of the applicant's drivers license (Secondary verification of ID)

Technology Requirements for Processing Credit/Debit Cards on Web Sites

The following are considered to be the technology requirements and best practices for e-commerce Web Sites that accept credit/debit card payments (Federal Trade Commission. (2016).

Create a Secure Payment Web Site

This is needed to protect credit/debit card data and other sensitive information from hackers during the credit/debit card transaction process. Identity theft and credit/debit card fraud are occurring more frequently on the Internet, and merchants must ensure that their customers are protected from internet criminals. Merchants can help secure their site by having a secure socket layer certificate, or SSL. SSL encrypts information being entered on the merchant site as it is sent across the Internet.

Utilize a Compatible Shopping Cart Application

This is required to make sure the merchant's shopping cart application can communicate with the merchant's credit/debit card payment-processing gateway. There are several different types of credit/debit card payment gateways, and each has a set of standards that must be followed. Many of the free shopping cart application software packages do not support all of the available credit/debit card payment gateways. A merchant should check with their merchant account provider or their shopping cart documentation to make sure that all the components will work together. Shopping cart applications fall into two basic categories namely: Local shopping carts that merchants can install on their own Web servers, and third-party shopping carts that run on a provider's web site. If a merchant decides to install his own shopping cart software, he will have a variety of software packages from which to choose. Three popular ones are Miva Merchant, BigCommerce, and Agoracart.

If a merchant would rather not install shopping cart software on his own web site, there are a number of third-party options available. When a merchant utilizes third-party shopping cart software, the merchant must place a link on his web site to the third party's web site where the application exists. This link takes customers to the merchant's offsite shopping cart software.

Provide E-mail Message Encryption

If a merchant plans on accepting orders and sending or receiving credit/debit card information via email, the merchant will need to encrypt the information that is transmitted. PGP, which stands for “Pretty Good Privacy,” is the most common form of email encryption. PGP encrypts an email when it is sent and decrypts the email when it has reached the intended recipient. If e-commerce merchants plan to use PGP, they will also need to make sure that their email clients support it.

Utilize a Firewall

If a merchant stores customer credit/debit card numbers or other personal information on his server, it is necessary to have a site-wide firewall to protect this information.

Use Anti-Virus Software and Update It Often

This will prevent most of the hacker’s attempts to invade the merchant’s Web site and steal personal information such as credit/debit card numbers.

Use Anti-Virus Software

This is another important way to protect the merchant’s network and computer systems from outside intruders.

Regularly Download and Install Security Updates:

ware performance and security can be optimized by installing all service and security updates.

After merchants have implemented these basic technology requirements, they are ready to offer their customers an easy way to purchase their merchandise or services. Merchants can also give their customers comfort in knowing that they are providing a safe and secure method for making credit/debit card payments and providing other personal information.

How Credit/Debit Card Payments are Accepted and Processed Online

If most of a merchant’s business is conducted on the Internet, Real-Time processing is the appropriate solution. When a customer who is using a merchant’s Web site is finished shopping and is ready to pay, typically the customer simply clicks on a “Check Out” button which is a link to a secure page where customers type in their credit/debit card information. After a few seconds, a message will then appear showing whether the credit/debit card has been accepted or declined. Real-Time credit/debit card processor or merchant account service providers will have an online database containing all of the credit/debit card transactions for a merchant which makes month-end accounting and balancing simple. Real-Time processing is the best solution for those who plan on having a high volume of daily transactions. Real-Time processing helps to automate the payment acceptance process, unlike in retail establishments where entering credit card information must be done manually. Most Real-Time solutions are coupled with a “Virtual Terminal” that allows a merchant to process Mail Order/Telephone Order (MOTO) orders manually via a web browser from any location that has access to the Internet.

The process a credit/debit card transaction goes through is fairly complicated; however it generally only takes a few seconds. The steps below illustrate how credit/debit card transactions are typically processed using a Real-Time credit card processing service (Google, 2013):

1. Using the merchant’s shopping cart Web interface, customers select “check out” with the items they placed into their shopping cart or selected from an order form on a merchant’s Website.
2. Customer then selects “credit card” or “debit card” as their method of payment.
3. The customer’s Web browser then connects to the Merchant’s website host’s secure server, and brings up the secure payment form.

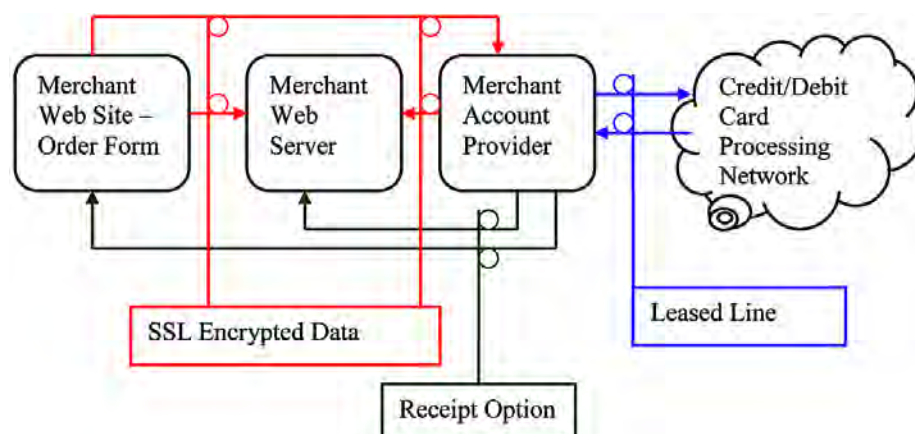
4. Customer enter their credit/debit card information on the secure payment form, and authorize the transaction by clicking a “Complete Order” or “Continue” type of button.
5. The credit/debit card transaction information is transmitted to the Website host’s secure server using SSL encryption.
6. The merchant’s secure server connects to the merchant’s processing bank either via a secure payment gateway (a third party which provides the connection to the processing bank), or directly (some credit/debit card processors have their own proprietary secure payment gateway and therefore do not require a third party to provide this service).
7. The credit/debit card processor service sends the transaction to the credit card/debit card processor network, such as Visa, MasterCard, or Flagship Merchant Services and the validity of the card and availability of funds is confirmed.
8. If the credit/debit card transaction is approved, an authorization code is returned to the credit card processor service or to the Secure Payment Gateway from the credit/debit card association network or debit card issuing bank.
9. The authorization is encrypted by the Payment Gateway or credit/debit card processor and transmitted in encrypted form to the secure Web server of the merchant, which permits fulfillment of the order.
10. The merchant’s secure Web server then sends the customer’s Web browser a confirmation receipt.
11. The amount due for the credit/debit card transaction is moved from the card holder’s bank to the merchant’s credit/debit card processing bank. The merchant’s credit/debit card processing bank transfers the money to the merchant’s local bank within 2 to 3 business days.

Figure 1, below, illustrates the technological components of a typical credit card processing system

Minimizing Internet Credit Card Fraud

Although there are no certified global figures on losses from credit/debit card fraud a Nilson Report in July 2015 showed credit card and debit card fraud resulted in losses amounting to \$16.31 billion during 2014. Card issuers and merchants incurred 62% and 38% of those losses, respectively, with the following transactional breakdown: Card issuer losses occur mainly at the point of sale from counterfeit cards while merchant losses occur

Figure 1. Typical credit card processing system



mainly on card-not-present (CNP) transactions on the Web, at a call center or through mail order (The Nilson Report, 2015).

For US buyers, credit card fraud does not pose a significant problem, as their loss is limited to \$50 but debit card fraud can cost a user and financial institutions much more. But for merchants that is not the situation. When a consumer indicates an instance of fraud, the disputed amount is removed from the merchant's account and credited back to the customer. This "chargeback" typically comes with a standard fee of \$15 per instance. At highest risk are downloadable software and entertainment, and high ticket items such as airline tickets, computers, and diamonds.

Payment gateways in the US have developed sophisticated fraud checking, but it has not halted credit/debit card fraud. To protect themselves, merchants can capture IP addresses of purchasers (must inform them that they are doing so), carefully examine purchases made from free e-mail addresses, those with different shipping and billing addresses, bounced e-mail order confirmations, no-existent telephone numbers, and large middle-of-the-night transactions. Merchants must also be cautious about shipping to Eastern European countries with a history of fraudulent transactions and telephone the buyer before shipping high ticket items.

Sophisticated Security Required to Prevent Credit/Debit Card Fraud

Online merchants have been forced to develop sophisticated security protections that far exceed the normal security approval process by the credit/debit card companies (NCR, 2015). Currently, credit card companies only verify whether a credit card number is correct and then match the number against the customer's billing address but the criminals can make sure the number is correct and that the addresses match. Cybercrime in all forms shows no signs of decreasing in the near future. For example, MSNBC reported that Visa quietly informed select merchants that 485,000

credit card numbers were stolen from a major e-tailer in January 1999 and in 2008 Bank of America notified thousand of card holders that their MasterCard information had been compromised. E-tailers (Web Merchants) find themselves in a difficult position regarding credit card fraud (Federal Bureau of Investigation, 2014).

Credit/Debit Card Fraud Solution Approaches

While there does not appear to be any simple solutions, experts believe that potential cyber-criminals will soon begin to reconsider committing credit/debit card fraud. This type of criminal activity has, for a long time, been considered too small to bother with, but using credit/debit cards fraudulently is quickly becoming "identity theft" -- which was recently defined as a serious federal felony. Cyber-criminals do leave digital fingerprints and can get caught. There are a number of approaches used by criminals to commit credit/debit card fraud and there are a number of procedures implemented to deter their attempts at credit/debit card fraud. These are discussed below.

Security Codes

An important Internet security feature that now appears on the back of most Visa/MasterCard and Discover cards, and on the front of American Express cards is a security code. This code is generally a three or four-digit number which provides a cryptographic check of the information embossed on the card. The security code helps validate that the customer placing an online order actually has the credit card in his/her possession, and that the credit/debit card account is legitimate. The security code is only printed on the card and it is not contained in the magnetic stripe information nor does it appear on sales receipts or billing statements. The goal is to make certain that the customer must have the card in his/her possession in order to use this code. Since Card Security Codes are not scanned into standard credit card

readers, in theory, these numbers are only visible to the customer. Visa, MasterCard, Discover and American Express now require Internet commerce sites to obtain the security code for all cards that have a code printed on them. In order for a card transaction to be accepted and processed, this code is required as part of the transaction data.

Unfortunately, the cyber-thieves are also using advanced techniques to ascertain critical information about stolen card numbers. They have developed software that can determine which bank issued a card, harvest the three-digit card verification number and determine the available credit-card limit as well as get personal information such as address and telephone number about the owner.

Credit Card “Skimming”

Criminal gangs recruit individuals who work within restaurants, hotels and retail outlets. The recruits are given battery powered electronic devices known as “skimmers” that read and capture all of the credit or debit card’s details in the few seconds that it takes to swipe the card through the card reader machine. When customers pay their bill, their card is first swiped through the legitimate credit card machine, but then it is also swiped through the “skimmer” reader. The recruits then pass the “skimmer” machines onto counterfeiters, who pay the recruits for their part in the crime. Once the “skimmer” machines have been given to the counterfeiters, they download the information onto a computer and produce a fake clone of the card. The “cloned” card is used to illegally purchase products and services. Skimming is costing card users worldwide millions of dollars in phony charges, as stolen clones are sold and used in the United States and elsewhere around the world. The introduction of EMV chips on credit cards has made the skimming process nearly impossible.

Single Use Credit Card Numbers

Some credit card companies have a new security and privacy offering which utilizes the concept of disposable card numbers. With this system, customers can get unique card numbers linked to their card account each time they make a purchase online. This allows the customer to avoid transmitting their “real” card numbers. The single-use numbers don’t work for recurring charges but they also don’t work for cyber-thieves who try to make multiple purchases

Smart Card Technology for On-Line Purchasing

Newer “smart cards” are embedded with a computer chip containing a digital certificate. A digital certificate consists of basic information about the cardholder’s digital identity. It contains elementary personal information such as the cardholder’s name, e-mail address and digital signature. The digital signature is nothing more than a series of numbers called a public key which forms the basis of encryption algorithms. Unlike a written signature, a digital signature has two purposes. It authenticates who the cardholder is legally and it also allows the cardholder’s messages to be encrypted.

Because the smart card chips are programmable, “smart chip technology” is flexible, and designed for multiple applications. These cards are inserted into a, typically free, smart card reader plugged into the user’s or merchant’s computer. The card, together with a PIN number, allows consumers to buy on the Internet using their digital certificate. The card allows access to an online wallet, which contains information such as shipping and ordering information. This information is automatically transmitted to the merchant’s online order forms.

Address Verification System (AVS)

E-commerce merchants can utilize an Address Verification System (AVS) for consumers from the United States. An AVS takes the consumer's ZIP code and the numbers in the street address, and compares them with the numbers in the card billing address. If they agree, the transaction is authorized; if they do not agree, the transaction is flagged as suspicious or in some cases not allowed, depending upon the merchant's preference. Using AVS lowers the merchant's discount rate, and can protect against stolen cards where the thief has the card number, but not a correct address.

Telephone Number Authentication

A telephone number authentication service can provide a decrease in the number of fraudulent transactions that pass through an on-line e-commerce web site. Most cyber-criminals are not willing to provide their real telephone number to complete a transaction and many, if asked for a telephone number will simply exit the transaction. There are a number of services that will provide real time telephone number authentication. These services can determine whether a telephone number is real, no longer in service, stolen or a legitimate working number at the address given by the user.

Telephone Call Verification

Telephone Call Verification works by automatically calling an online end-user's telephone number at the same time the end-user is making a transaction on a website. The user while on the website answers the phone and is provided a one-time personal identification number (PIN) presented via the web interface; an otherwise anonymous online end-user will be able to confirm that the person who received the phone call and the person who is interacting on the website are the same person.

Customer Transaction and IP History Databases Checks

Another approach for detecting online fraud is to compare a transaction with previous transactions made for a given card number and make sure it fits the pattern of use. There are companies that provide real time checks of cards with databases of millions and, in some cases billions, of records to detect anomalies (Conlin, 2007). This type of service will score a card transaction based on all the intelligence it has gathered both about the transaction and former purchases. In addition, online fraud detection solutions based on a combination of IP reputation analysis and a mutual collaboration network has proved successful. IP reputation uses geo-location and proxy detection by providing relevant information about the IP's historic behavior, both legitimate and suspicious (Wikipedia, 2016).

Intelligent Credit/Debit Card Fraud Detection

An intelligent card detection system monitors card transactions as they occur by gathering data from the current and previous transactions and uses this data to compute a transaction score for the current transaction. The algorithms used to compute such a score are called classifiers. Typically, high scores for transactions are more likely to be fraudulent than low scores thus transaction scores are compared to a threshold and the score is classified as normal or fraudulent. Card fraud detection is a complicated problem involving many input variables such as time, transaction amount, merchant, merchant category code, country, etc. acquired from multiple transactions in a sequence. A classifier computes a fraud score based on a number of these variables.

Two basic approaches have been used in developing classifiers, namely, neural networks and Bayesian decision methods. A neural network is a nonlinear function which takes multiple input

variables and computes a score from them. A neural network consists of a series of interconnected neurons similar to the structure of the brain. The interconnections have weights assigned to them and the input neurons (input nodes) to a network represent each continuous variable and every value that a categorical variable can take. The Bayesian approach to card fraud detection is based on probability theory. Research has identified a number of characteristics derived from card transactions which tend to be predictive of fraud. A Bayesian approach computes the probability distributions for each of these card transaction characteristics using a process called evidence integration to compute a fraud probability from the individual characteristic probabilities (Velocity Payment services, 2016).

Regulatory and Legislative Issues

Management of information risk is now tied to regulatory mandates. Since 1999, laws enacted at the federal and state levels have forced companies to be extremely careful in protecting the confidentiality and reliability of medical, financial and other sensitive information stored on their computer systems. Failure to comply with these mandates can lead to civil and criminal penalties, lawsuits and related litigation costs and, of course, damage to reputations. Although the earlier laws focused on financial and healthcare companies, two of the most recent laws, namely, the 2002 Sarbanes-Oxley Act and the California Data Protection Law (SB 1386), broadened the scope of companies that are required to comply. The Gramm-Leach-Bliley Act (GLB), also known as the Gramm-Leach-Bliley Financial Services Modernization Act, is an Act of the United States Congress that stipulates that every financial institution must protect the security and confidentiality of its customers' personal information. The Federal Trade Commission in conjunction with several other federal and state agencies along with the Federal Bureau of Investigation (FBI) is the

federal agency responsible for enforcement of these laws and mandates.

Federal Trade Commission (FTC)

The FTC deals with issues that are related to the economic life of every American citizen and business. It is the only federal agency with both consumer protection and competition jurisdiction across all sectors of the economy including e-commerce. The FTC is charged with law enforcement and protecting consumers' as well as business' interests by sharing its expertise with federal and state legislatures and U.S. and international government agencies; developing policy and research tools through hearings, workshops, and conferences; and creating practical educational programs for consumers and businesses in a global marketplace with constantly changing technologies (Federal Trade Commission. (2016).

Managing Information Risk

Managing information risk must be integrated with a merchant's overall risk management strategy. The technology infrastructure; including servers, network monitors, and firewalls, needs to be assessed and managed in terms of its relation to people, operations, supply chains and other business drivers. Some of the steps involved with information technology (IT) risk management include paying attention to human factors, putting proper security policies in place, identifying critical assets and fostering better communication and an enterprise-wide perspective among IT managers and risk managers. Bringing together IT, risk management, internal audit, legal and human resources to address information management risk issues produce a consensus to the identification of threats, the areas of operation (ranked in order of most critical and sensitive) that could be affected by a threat, potential financial or reputational loss, and the most cost-effective way to reduce the risk (National Institute of Standards and Technology, 2007).

An Information Risk Assessment

A risk assessment should be performed by any merchant accepting credit or debit card payments and this assessment should examine the following risk factors:

- **System Characteristics:** Assess and identify the resources and information that constitute the systems used for financial purposes and identify the business systems jointly with management personnel, IT personnel and users.
- **Threat Identification:** Conduct interviews and utilize work-group sessions with key management team members, technology administrators and system users to uncover potential threat agents that may impact the confidentiality, integrity and availability of information stored in databases and files.
- **Vulnerability Identification:** Conduct a technical assessment to detect vulnerabilities and to assess how effective the controls are for preventing unauthorized access,
- **Control Analysis:** Assess countermeasures regarding items such as firewalls, encryption, web server access policies, password policies, backup and recovery procedures, change-management procedures, currency of software, hardware maintenance and the physical environment.
- **Insurance Gap Analysis:** Assess current insurance policies in terms of coverage for financial loss arising out of unauthorized access or use of confidential information, damage to third-party software or data as well as damage to the business network or databases and files. The risk assessment can not only help identify the critical areas of risk to be addressed, but can also be used to recommend best practices to remedy the risk.

Credit/Debit Card Fraud Preventive Steps for Online Business Owners

When a merchant physically accepts a credit/debit card, and the charge is authorized, and the merchant has conformed to card regulation, the merchant will get paid, even if a stolen card is used. But, the liability for fraud shifts from the card issuer to the merchant for 'Card Not Present' sales (Internet sales, mail order, and telephone/fax order). After a card processor or registration service approves a card transaction, the merchant should perform additional checks, as fraudulent orders are sometimes approved.

The following methods and techniques (NCR, 2015); (Federal Trade Commission, 2012); (Jepson, (2016) can be utilized to protect an e-commerce merchant against credit/debit card fraud. Typically, a combination of methods is the best approach.

Follow the procedures recommended by your credit/debit card processor and the related card organizations:

- **Authorization Approval:** Make sure you get it from the issuing bank.
- **Address Verification System (AVS):** AVS is only available for the and partially available in four European countries to verify the address matches.
- **Card Verification Methods (CVM):** Security Codes: VISA = CVV2, MasterCard = CVC2, and American Express = CID use a security code of 3 or 4 extra digits, a pin for debit cards.
- **Payer Authentication Programs:** Authentication programs (Verified by Visa and MasterCard's SecureCode) use personal passwords to ensure the identity of the online card user.
- **Real-Time Authorization:** Credit/debit card information is sent to the processor for immediate approval.

- **Bin Check:** The first 6 digits of the credit card are called the Bank Identification Number (BIN).
- **Calling The Card-Issuing Bank:** Call the card-issuing bank, to verify the charge.
- **Different Bill And Ship To Addresses:** Use Google to search for the numeric street address, street name, and zip code.
- **Negative Historical File:** Keep a database of prior fraud attempts, problem customers, charge back records, and customers receiving refunds.
- **Shared Negative Historical File:** Combined negative historical databases by several merchants.
- **Positive Database File:** This file contains a list of good customers.
- **Credit Service Database:** A credit database service, such as Equifax (www.equifax.com), Experian (www.experian.com), and Trans Union (www.tuc.com) are most useful for high-dollar value items.
- **Customizable Merchant Rules:** The merchant sets up rules to stop or flag specific orders for review.
- **Fraud Scoring Systems:** Weights, points or probabilities are assigned for different elements of a transaction (IP Address, free-email account, time of day, AVS results, amount of sale, type of products ordered, shipment method, different shipping/billing addresses, certain zip codes, etc) to generate a fraud score to indicate the likelihood of fraud.
- **Pattern Detection:** Check if multiple orders are placed shipping to the same address, but different credit/debit cards were used. Check orders for an unusually high quantity of a single item.
- **Alternate Thank You Page:** If an order is being shipped to a non-English speaking country, display an alternate thank you page. Require the customer to fax either a photo of the credit card or a Xerox of his/her credit card bill.
- **Custom Built Software:** Some merchants have branded their software, displaying the customer's name in the software.
- **Preventative Data Checking Measures:** Check the data fields to determine if the buyer is a real person. Check if the ZIP Code the customer listed really exists. Check if the customer's e-mail address is formatted properly. Check for incomplete names or an address like 123 Main Street.
- **Free Email Accounts:** There is a much higher incidence of fraud from free email services. Many fraudsters use free email addresses to remain anonymous.
- **Reverse IP Address Checks:** Check if the IP address matches the email address and physical billing address of the customer. The IP address identifies the location of the server where the order was placed. Numerical IP addresses can be checked through programs such as WsPing32.
- **Anonymous and Open Proxy IP Addresses:** IP addresses can also be forged thereby hiding the forged IP address's true location of the fraudster. Organized credit/debit card fraud rings often use anonymous proxies.
- **Checking Telephone Numbers:** The web sites at <http://www.freeality.com/finde.htm>, <http://www.theultimates.com/>, <http://www.anywho.com>, <http://nt.jcsm.com/ziproundacx.asp>, and <http://nt.jcsm.com/ziproundacx.asp> provide tools to match the telephone area code to a postal zip code, reverse telephone directories, search for email addresses, maps, directions, etc. The merchant can call directory assistance to determine if the number on the order phone matches their number.
- **Fax Orders:** When a credit/debit card order is received by fax, require the customer to also fax copies of both sides of the credit/debit card and a copy of their state-issued ID, or driver's license.

- **International Orders:** Some countries have very bad reputations for fraud. Your bank or credit card processor can provide a list of high-risk countries. High risk countries include developing nations like Indonesia, Malaysia, Benin, Nigeria, Pakistan, Israel, Egypt, and Eastern European countries. Placing an international phone call to the issuing bank may make sense for large orders and/or ask the customer to contact the merchant by phone or email for shipping costs. A fraudster may consider this too much contact, and decide to go elsewhere.
- **Calling the Customer:** Calling customers is not only an excellent way to detect fraud, but it can also be a valuable part of your customer service.
- **Web Site Information:** Make sure the order form includes fields to enter the CVV2 verification code imprinted on the card, the name of the card-issuing bank, and the bank's toll-free telephone number, the customer's telephone number and email address.
- **Processing Orders:** Do not ship any order until the charge can be verified by additional checks.
- **Use Temporary Activation Codes:** If the merchant wants to process orders immediately, issue thirty-day temporary validation keys for downloaded software.
- **Anti-Fraud Groups:** Become educated by attending seminars offered by credit/debit card companies and card processors. Organizations such as www.antifraud.com and www.wiscomcomputing.com offer help. These groups also offer tips, databases of stolen credit/debit cards, and web look up tools.
- **File a Complaint With the FTC and the FBI:** If you detect fraud or have been a victim of fraud, file a complaint with the FTC at <https://www.ftccomplaintassistant.gov/> and the FBI's Internet Crime Complaint

Center or IC3, a partnership of the FBI and the National White Collar Crime Center at <http://www.fbi.gov/majcases/fraud/internetschemes.htm>.

FUTURE RESEARCH DIRECTIONS

Machine learning technologies such as artificial intelligence are currently being investigated that can detect and prevent fraud in a variety of financial transactions, including credit cards, debit cards, ATMs, and e-payments. The machine learning system can be built into a client's infrastructure and a machine learning model can be created using existing data and retrained as conditions change, forming an integrated system that allows a client to combat fraud before it happens. By identifying legal transactions that have a high probability of being followed by a fraudulent transaction, an institution can take pro-active measures. While machine learning and stream computing technologies currently can't predict the future, as they continue to be developed they will allow financial institutions to make smarter decisions and work to prevent fraud before it happens.

CONCLUSION

Based on past performance and predictions for the future, it seems safe to say that purchasing goods and services over the internet will continue to increase. This is because it is more efficient for the merchants and they can reach a much larger audience than using the face-to-face, in-store methods of the past. But like most uses of technology, there are individuals who find ways to use the technology for criminal purposes. This has been the case when utilizing credit or debit cards for purchasing goods and services over the internet. Thus, a sort of battleground has evolved between the e-commerce merchants along with their customers and the cyber-criminals. As new technological security methods are implemented by merchants to protect themselves and their cus-

tomers, the cyber-criminals attempt to find ways through or around these technological barriers. If past events are any indication of the future, this battle is not over and merchants and their customers must continue to find secure methods to combat the criminals attempting to fraudulently steal financial and other personal information for their own financial gain.

REFERENCES

- Best 10 Merchant Services. (2009). *Best 10 Merchant Services*. Retrieved May 8, 2016 from <http://www.best10merchantservices.com/>
- Cardfellow. (2016). *Debit Card Transaction Fees: Signature Vs PIN*. Retrieved May 9, 2016 from <https://www.cardfellow.com/debit-card-transaction-fees/#DebitNetworkFees>
- Conlin, J. (2007, May 11). *Credit card fraud keeps growing on the Net*. Retrieved May 9, 2016 from Web Site: <http://www.nytimes.com/2007/05/11/your-money/11iht-credit.1.5664687.html>
- Constantin, L. (2015, April 14). *Web App Attacks, PoS Intrusions and Cyber-espionage Top Causes of Data Breaches*. Retrieved May 10, 2016 from <http://www.computerworld.com/article/2909519/web-app-attacks-pos-intrusions-and-cyberespionage-top-causes-of-data-breaches.html>
- Cooper, A. (2016). *Is it safe to shop online with a debit card?* Retrieved on May 3, 2016 from <http://money.howstuffworks.com/personal-finance/online-banking/safe-to-shop-online-with-debit-card.htm>
- Faughnan, J. G. (2007, November 23). *International Net-Based Credit Card/Check Card Fraud with Small Charges*. Retrieved May 10, 2016 from <http://www.faughnan.com/ccfraud.html#Introduction>
- Federal Bureau of Investigation. (2014). *Internet Crime Complaint Center (IC3)*. Retrieved May 9, 2016 from <http://www.ic3.gov/default.aspx>
- Federal Trade Commission. (2012). *Protecting Against Credit Card Fraud*. Retrieved May 9, 2016 from <https://www.consumer.ftc.gov/articles/0216-protecting-against-credit-card-fraud>
- Federal Trade Commission. (2016). *About the Federal Trade Commission*. Retrieved May 12, 2016 from <http://www.ftc.gov/ftc/about.shtm>
- Google. (2013, September 10). *Computerized Method and System for Providing a Continuing Education Credit Card*. Retrieved May 8, 2016 from the Web Site: <http://www.google.com/patents/US8533108>
- Jepson, T. (2016). *Merchant Credit Card Fraud: 31 Ways to Minimize Credit Card Fraud*. Retrieved May 9, 2016 from <http://www.wiscocomputing.com/articles/ccfraud.htm>
- Murdock, K. (2006). *Credit Card Processing: How It All Works*. Retrieved May 12, 2016 from <http://www.practicalecommerce.com/articles/168-Credit-Card-Processing-How-It-All-Works>
- National Institute of Standards and Technology. (2007, July 3). *Guide for Conducting Risk Assessments*. Retrieved May 12, 2016 from <http://csrc.nist.gov/publications/nistpubs/800-30/sp800-30.pdf>
- NCR. (2015). *Card fraud detection - Comparison of Detection Technologies*. Retrieved May 9, 2016 from <http://www.ncr.com/financial-services/enterprise-fraud-prevention>
- Smith, R. (2016). *How Credit Card Processing Works*. Retrieved May 10, 2016 from Web Site: <http://www.smithfam.com/news/nov99d.html>
- TechAdvisory.org. (2014, June 13). *10 Security Best Practices*. Retrieved May 10, 2016 from <http://www.techadvisory.org/2014/06/10-security-best-practices/>
- The Nilson Report. (2016). *Charts and Graphs*. Retrieved May 7, 2016 from https://www.nilsonreport.com/publication_chart_and_graphs_archive.php?1=1&year=2013

TopTenReviews. (2016). *Credit Card Processing Services Review*. Retrieved May 10, 2016 from <http://credit-card-processing-review.toptenreviews.com/>

Total System Services, Inc. (2014). *Consumer Payments Study*. Retrieved May 4, 2016 from http://tsys.com/Assets/TSYS/downloads/rs_2014-consumer-payments-study.pdf

Velocity Payment services. (2016). *How Debit Card Processing Works*. Retrieved on May 5, 2016 from Web Site <http://paymentservices.velocitysc.com/learnabout/howdebitcardprocessingworks>

Wikipedia. (2016). *Debit Card*. Retrieved from https://en.wikipedia.org/wiki/Debit_card

Williams, J., & Premchaiswadi, W. (2009). *On-Line Credit Card Payment Processing and Fraud Prevention for e-Business*. In *Encyclopedia of E-Business Development and Management in the Global Economy*. IGI Global.

ADDITIONAL READING

Guides, F. (2009). *Credit Card Skimming*. Retrieved from <http://www.fraudguides.com/business-credit-card-skimming.asp>

Internet Crime Complaint Center (IC3). (2009). Retrieved January 21, 2009 from <http://www.ic3.gov/default.aspx>

Montague, D. A. (2004). *Fraud Prevention Techniques for Credit Card Fraud*. Victoria, Canada: Trafford Publishing.

Stoneburner, G., Goguen, A., & Feringa, A. (2009). *Risk Management Guide for Information Technology Systems Recommendations of the National Institute of Standards and Technology*. Retrieved January 22, 2009, from <http://csrc.nist.gov/publications/nistpubs/800-30/sp800-30.pdf>

KEY TERMS AND DEFINITIONS

Acquiring Bank: The bank that represents the e-commerce merchant and processes all of the merchant's credit card payments with the credit card associations.

Credit Card Processor: A third party utilized to process credit card payments for merchants and their acquiring bank.

Credit Card: A card issued by banks, savings and loans, retail stores, and other businesses that can be used to borrow money or buy products and services on credit.

Cyber-Criminal: An individual who commits a crime using a computer and the internet to steal a person's identity such as credit card information.

Debit Card: A bank issued card used for cash transactions, but is not a credit card. In a debit card transaction, the amount of a purchase is withdrawn from the available balance in the cardholder's bank account. If the available funds are insufficient, the transaction is not completed. It is also called asset card in the US.

E-Commerce: The buying and selling of goods and services on the Internet.

Fraud: An act of deception for the purpose of unlawful financial gain using stolen credit card information.

Issuing Bank: The bank that issues consumers their credit cards.

Merchant Account: A legally binding contract wherein an acquiring bank extends a line of credit to a merchant who desires to accept payment using credit cards.

Service Gateway: This is another name for a credit card processor.

Skimming: This is a type of fraud wherein the numbers on a credit card are recorded and transferred to a duplicate card.

SSL: SSL is an abbreviation for Secure Sockets Layer, a protocol developed for transmitting documents over the Internet using a cryptographic system that uses two keys to encrypt data; namely a public key known to everyone and a private or secret key known only to the recipient of the document.

Category E

Electronic Commerce

Adoption and Use of Mobile Money Services in Nigeria

Olayinka David-West

Pan-Atlantic University, Nigeria

Immanuel Ovemeso Umukoro

Pan-Atlantic University, Nigeria

Omotayo Muritala

Pan-Atlantic University, Nigeria

INTRODUCTION

The Bill and Melinda Gates Foundation [BMGF] (2013) reports that more than two billion people globally are outside the formal financial sector and are either financially excluded or underserved. Increasing access to quality and affordable financial services accelerates the well-being of households, communities, and economies especially those in the developing world. One promising way to deliver financial services to the poor both profitably and at scale is through digital payment platforms delivered through mobile telephony. With many challenges and difficulties, including infrastructure, policy and regulatory constraints, associated with access to financial services through the bricks and mortar model, driving financial inclusion using mobile money has been considered a veritable approach, especially for rural dwellers.

The Central Bank of Nigeria licensed twenty-one mobile money operators (see Table 1) to provide mobile money services to millions of Nigerians as a means of bridging the gap between the financially served, the underserved and un-served. However, unlike some Kenya where 70% of the population are active mobile money subscribers with just six licensed mobile money operators, Nigeria, with more MMOs has less

than a million (about 0.8%) adult as active mobile money subscribers (EFInA, 2014) indicating low adoption and acceptance of mobile money, most prominent in rural areas where bank penetration is low (Kama & Adigun, 2013, p.33; KPMG, 2015, p.13).

MNOs and financial institutions remain key drivers of mobile services although this is dependent on the licensing and regulatory framework in place. To determine which of these categories is first mover in terms of deployment is largely the dictate of regulatory authorities (Lal & Sachdev, 2015). In Nigeria, MNOs provide voice, data, auxiliary services including mobile payments, in partnership with banks and other third parties. MNOs have the infrastructure and capacity to operate at scale, yet, great challenges to adoption and use of services continue to threaten operations and growth of the sector.

This chapter explored causes of mobile money adoption in Nigeria alongside recommendations to policy makers, mobile money operators, agents and institutions promoting financial inclusion. The state of Nigeria's mobile money services is established through a quantitative analytical approach using data from EFInA's Access to Finance and InterMedia's Financial Inclusion Insight nationwide surveys.

Table 1. Registered Mobile Money Operators in Nigeria

Mobile Money Service	Company	Website	Status	Model
ReadyCash	Parkway Projects	http://www.readycash.com.ng/	Active	Non Bank-led
UMo	UBA AfriPAY	www.umo.net	Inactive	Bank-led
GTMobileMoney	GTBank	http://www.gtbank.com/mobile-money	Active	Bank-led
FirstMonie	FirstBank	http://www.firstmonie.com	Active	Bank-led
PocketMoni	eTranzact	http://www.pocketmoni.com	Active	Non Bank-led
Paga	PagaTech	http://www.pagatech.com	Active	Non Bank-led
Fortis Mobile Money	Fortis MFB	http://www.fortismobilemoney.com	Active	Non Bank-led
*909# Mobile Money	Stanbic IBTC	https://web.909wallet.com/	Active	Bank-led
Monitise	Monitise MM Nigeria Ltd	http://monitiseimmigeria.com.ng	Inactive	Non Bank-led
FetsWallet	FETSWALLET	https://www.fetswallet.com	Active	Non Bank-led
Ecobank Mobile Money	Ecobank	http://www.ecobank.com/mobilemoney.aspx	Active	Bank-led
QikQik	Eartholeum	http://www.eartholeum.com/about-qikqik-1.html	Active	Non Bank-led
Teasy Mobile	Teasy Mobile	http://teasymobile.com	Active	Non Bank-led
Mimo	Mkudi Limited	https://www.mimo.com.ng/about	Active	Non Bank-led
EazyMoney	Zenith Bank	http://www.eazymoney.com.ng	Active	Bank-led
PIDO	PayCom	http://www.paycom-ng.com/	Active	Non Bank-led
Vcash	Virtual Terminal Network	https://www.virtualterminalnetwork.com/	Active	Non Bank-led
Cellulant Wallet	Cellulant Nigeria Ltd.	https://www.cellulantwallet.com	Active	Non Bank-led
Zinternet Mobile Money	Zinternet Nigeria Limited		Inactive	Non Bank-led
Fidelity Mobile	Fidelity Bank	https://www.fidelitybank.ng/mobile-money/	Active	Bank-led
Zoto Mobile Money	Hedonmark Management Services Ltd	www.zoto.com.ng	Active	Non Bank-led

Source: Central Bank of Nigeria

BACKGROUND

Financial Inclusion and Exclusion

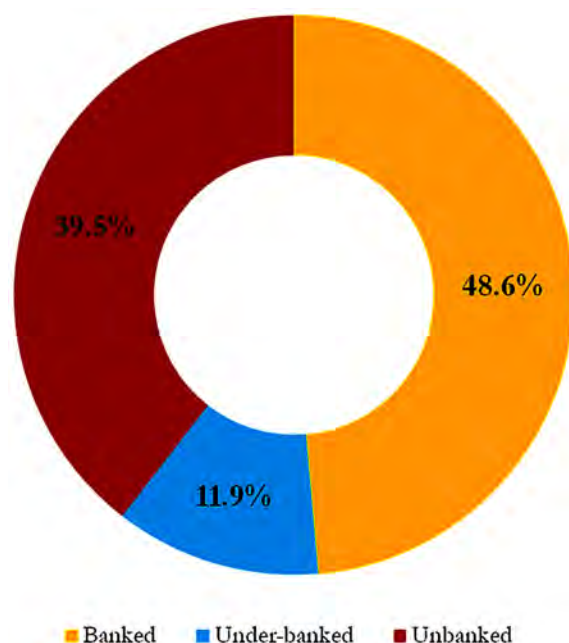
Financial inclusion and exclusion have received global attention because of their impacts on development. A common thread to several definitions of financial inclusion (Devlin, 2005; Sarma, 2008; Sarma & Pais, 2011; Demirguc-Kunt & Klapper, 2012; World Bank, 2014; Park & Mercado, 2015; Demirguc-Kunt, Klapper, Singer, & Van Oudheusden, 2015) is its link to ownership of an account with a formal financial institution and the conduct of financial transactions using same account within the last ninety days (Allen et al, 2013; InterMedia, 2015, p.2). Sarma (2008) defines financial inclusion as a process that ensures the ease of access, availability and use of the formal financial system for all members of an economy. Although financial exclusion may be voluntary or involuntary (Park & Mercado, 2015), formal financial services should be readily available and accessible to all. Consequently, Kasproicz and Rhyne (2013) define financial inclusion as a state in which everyone who can use financial services (including disabled, poor, and rural populations), has access to quality financial services at afford-

able prices, in a convenient manner, and with dignity for the clients.

The Global Findex (2014) posited that an inclusive financial system is critical to reducing poverty and achieving inclusive economic growth as getting more people to participate in the financial system opens up potential for business expansion, investment in children's education, ability to absolve financial shocks, combat financial risks associated with individual and business transactions and empowering women economically to have control over their finances. Thus, financial inclusion is critical to inclusive growth as access to finance can enable economic agents to make longer-term consumption and investment decisions, participate in productive activities, and cope with unexpected short-term shocks (Park & Mercado, 2015). Sarma and Pais (2011) proposed an index of financial inclusion that aggregates access, availability and use as measures of a country's financial inclusion standing. Absence of these three key factors is what leads to financial exclusion (Peruta, 2015).

According to EFINA (2014) about 40% of the adult population (18+ years) are excluded (unbanked); about 12% are informally included (under-banked) – using informal financial in-

Figure 1. Financial access status of adult Nigerians



stitutions such as co-operatives, thrifts, money guards, etc.

Intermedia (2015, p.3) reports that two in five Nigerian adults have account with a formal financial institution, accounting for 37% financial inclusion. Although the deep in financial inclusion recorded in 2015 might be explained by the deteriorating security situation due to persistent insurgency (InterMedia, 2015, p.3) and worsening economic situations, Nigeria still faces a daunting task of meeting the 80% financial inclusion target by 2020.

Why Mobile Money?

Peruta (2015) defines mobile money as a financial service that offers electronic accounts (e-wallets) in which customers can deposit, withdraw cash and manage electronic money. Micheni, Lule and Muketha (2013) see mobile money as a technology that uses wireless network infrastructure to facilitate the exchange of money value between various actors involved in a transaction. In contrast to the broad term – digital financial services

(DFS) which refers to financial services provided through an electronic platform such as mobile phones, electronic cards, the internet, etc. (InterMedia, 2015, p.2), mobile money services are built and delivered through mobile phones and mobile telephone networks. Mobile phones are rapidly transforming the way people use and access financial services and while more than a billion people around the world do not have a basic bank account, they do own a cell phone which provides a dramatic opportunity to expand financial access to poor people and achieve financial inclusion (CGAP, 2009a, p.15; 2009b, p.4).

Considered as the most affordable and widely spread technology device around the world, use of mobile phones for mobile money service has challenged traditional methods by creating new, convenient and accessible financial systems (David-West, 2016, p.256) especially in economies where financial regulations are high and do not provide enough room for the poor who may not readily have the requirements to get into the formal financial service stream.

Leveraging on mobile phone penetration opens financial services to millions of people without access to traditional banking services around the world (The Global Findex, 2014; Martinez, Hidalgo & Tuesta, 2013). Mobile money services can be accessed with a non-feature (basic) mobile telephone while leveraging on a robust mobile telephone network spread connecting both rural and urban areas. Zimmerman, Pearse and Bau (2016) opine that mobile money services are a solution to financial insecurity as digitization of cash transfers eliminates the risk of transporting large amounts of banknotes and coins. Furthermore, extra expenses, cash leakage from corruption, barriers of affordability, accessibility (Mbiti & Weil, 2011; Donovan, 2012; and Peruta, 2015, p.5), cost and strict banking requirements such as means of identification are eliminated.

Mobile money is not without opportunities to financial institutions and its operators. CGAP (2009a, p.15) reports that mobile phone banking and other technologies are already dramatically

reducing the costs that banks and microfinance institutions (MFIs) incur in providing financial services to low-income populations. With much lower costs and more convenient services, formal financial services become more attractive to poor people than the informal which are costlier and less secured.

Adoption and Acceptance of Mobile Money

Mobile money services have been in the Nigerian financial market for over 5 years. The InterMedia (2016a, p.1) cross-country comparison report shows that as at 2016, Nigeria has 13 active MMOs, Bangladesh with 2, Indonesia and Pakistan with 8 each, Uganda with 7, while Kenya and Tanzania have 6 mobile money providers each. Only India with 14 has more mobile money providers than Nigeria. However, the adoption of mobile money in Nigeria seems to be the lowest when compared to other countries with fewer numbers of providers despite the 85% mobile telephony penetration (InterMedia, 2016a, p.1).

EFInA (2014) reported that 98.8% of Nigeria's adult population (18+ years) have never used mobile money services while 0.4% have discontinued use of mobile money services. Although about 0.8% are current users as at the time of

the study, only 0.5% representing less than half a million adults Nigerians are registered mobile money users in a country of over 95 million adults.

As at 2015, only about 12% of adult Nigerians were aware of mobile money providers while 13% were aware of mobile money as a concept (InterMedia, 2015a, p.3). Mobile money adoption in Nigeria still remains low and even if the 12% awareness was to be converted to use, only an insignificant 0.3% growth will be recorded (InterMedia, 2015a, p.13). A cursory look at the adoption of mobile money in Nigeria reveals that mobile money is still encased within the banking population. In a nationwide survey conducted by EFInA (2014), it was found that across all financial access strands, about 0.8% users of mobile money (registered and unregistered) are from the banked population.

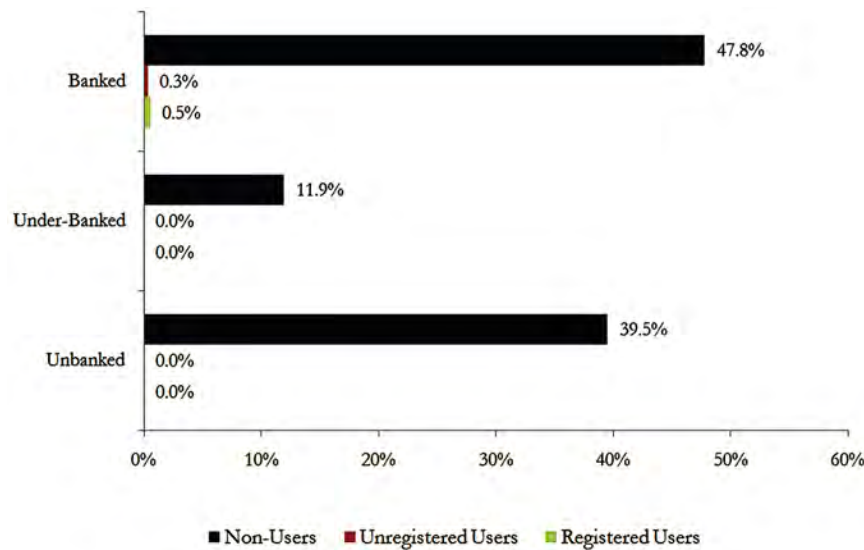
In 2015, the apex bank estimated the annual mobile money transactions at \$25m, representing 0.005% of total gross domestic product in the period and could be considered low when compared with Uganda where in 2013, mobile money transaction was around \$7.2 billion; \$1.9b in Tanzania and \$7.38b in Kenya, (Ondiege, 2015; Commonwealth, 2014).

Although 36% of Nigerian adults are digitally included, with their registered accounts offering some form of digital access (InterMedia, 2016b,

Figure 2. Mobile money adoption across the adult population



Figure 3. Mobile money use per financial access strand



**This figures are based on EFINA 2014 National Financial Inclusion Study covering 22,044 respondents*

p.19) to financial services, mobile money adoption remains very low. EFINA (2014) reported mobile money access level of 0.8% while Inter-Media (2016b, p.39) reports that in 2015, access to mobile money was 0.9% of which 0.6% are registered users and 0.5% as active users. While the number of mobile money deployments has experienced explosive growth, the number of active mobile money users has not grown on the same trajectory (Davidson & McCarty, 2015, p.2; David-West, 2016, p.257).

Mobile Money Inhibitors

While Nigeria's mobile money market remains saturated with operators in light of other evolving markets and only second to India, low level of adoption and acceptance poses threat to investors and potential users who may be discouraged from adoption and use given the absence of network effects. Understanding causes of low adoption and acceptance is thus instrumental in devising ways of harnessing the potentials of mobile money, not just in Nigeria but also in similar markets facing low adoption and acceptance. Peruta (2015, p.6)

opines that understanding the adoption and assimilation of an innovation such as mobile money is best understudied using models of innovation diffusion, technology acceptance and theory of planned behaviour as they provide theoretical foundation for factors that either promote or inhibit the acceptance and use of a technology.

From existing models, the following factors have been identified as possible inhibitors to mobile money adoption and use:

- **Complexity of Technology:** Technology sophistication has been identified as a hindrance to the adoption and use of a technology/innovation, mobile money inclusive. Models of information systems use such as Davis' (1989) technology acceptance model, Roger's (1995) diffusion of innovation, DeLone and McLean's (2003) IS success model and others confirm the significant relationship between ease of use of a technology and its associated use. Thus, the more complex a technology is perceived to be, the less likely it will be adopted and used. One of the problems associated with non-

adoption/use of mobile money in Nigeria is that many do not understand nor know how the technology works (InterMedia, 2015a, p.17). Davidson and McCarty (2015, p.3) reiterated that most customers do not understand the mechanics of performing transactions and are apprehensive to try something as novel as mobile money. Breaking this barrier requires technology providers designing simple interfaces, simplifying the processes required to initiate and complete a transaction using mobile money services. Increasing complexity of a technology impedes its adoption and use especially when the target users are unskilled and uneducated.

- **Security/Trust:** Shi (2011, p.36) opines that use of mobile financial services is hugely hinged on trust, confidentiality, data integrity, authentication and non-repudiation from the providers. Consequently, MMOs' inability to guarantee these measures translates into poor adoption and use. Trust, integrity and security are integral measures that must be provided if consumers are expected to adopt and use mobile money. When customers do not trust the operator's brand or network, they become hesitant to conduct transactions on such platforms (Davidson & McCarty, 2015, p.2). It should be noted that trust and security are quickly ascertained by intending adopters as they often inquire from peers, friends and family members. This is well documented as social influence under the unified theory of acceptance and use of technology by Venkatesh, *et al* (2003).
- **Education/Financial Literacy:** Davidson and McCarty (2015, p.2) submit that often times customers are aware of mobile money service, but do not understand how it could be beneficial to them. Peruta (2015, p.22) reports that education and financial literacy remain key elements in mobile money adoption as a financial inclusion

instrument. When customers understand mobile money's functionalities and usefulness, they learn how to transact. This typically requires a process of education carried out by an agents, friends or family members (Davidson & McCarty, 2015, p.3). Increasing financial literacy and mobile money adoption can be triggered using mobile agents as instruments of sensitization by playing the dual role of agents and financial educators in their catchment areas (Davidson & McCarty, 2015, p.20).

- **Awareness:** First, when people are unaware of the availability of both financial and mobile money services, adoption and use are usually near zero. Davidson and McCarty (2015, p.3) posit that it is not enough that consumers know the name of the mobile money service or even what mobile money but awareness campaigns must build understanding to help users see how it is both relevant and beneficial to them which is fundamental to their behavioural change towards facilitating adoption and use. Lack of product and service awareness reflects the low level of adoption of mobile money service in Nigeria. From the 2013 survey by NOI Polls, 59% (about 6 out of 10) of Nigerians were completely unaware of the mobile money services while the Ericsson ConsumerLab (2016) reports 42% in the level of mobile money services awareness.

The 2014 financial inclusion survey on Nigeria conducted by EFINA highlights some of the identified inhibitors reported across literature as accounting for the low adoption of mobile money in Nigeria. Figure 4 shows some self-reported inhibitors to mobile money use among those who had previously adopted the innovation but discontinued its use.

While it is true that the technology offers effective, efficient and secured ways of transacting, studies (Micheni *et al*, 2013; Davidson & McCarty, 2015; Peruta, 2015) have not shown that

the technology is without setbacks. One major user self-reported setback to its sustained use is lack of knowledge/understanding of the functionalities and efficient use. About 51% of the early adopters indicated complexity as one of the reasons for discontinuity. Although the technology may not be as sophisticated as it is perceived by early adopters, the level of user education and previous experience with similar technological innovation becomes a factor when it comes to adoption and continuous intention to use (Peruta, 2015; Davidson & McCarty, 2015, p.3).

Other inhibitors to use of mobile money include non-ownership of mobile phones, unavailability of mobile money agents, unstable network, proximity to banking facility, lack of trust, preference for cash and others.

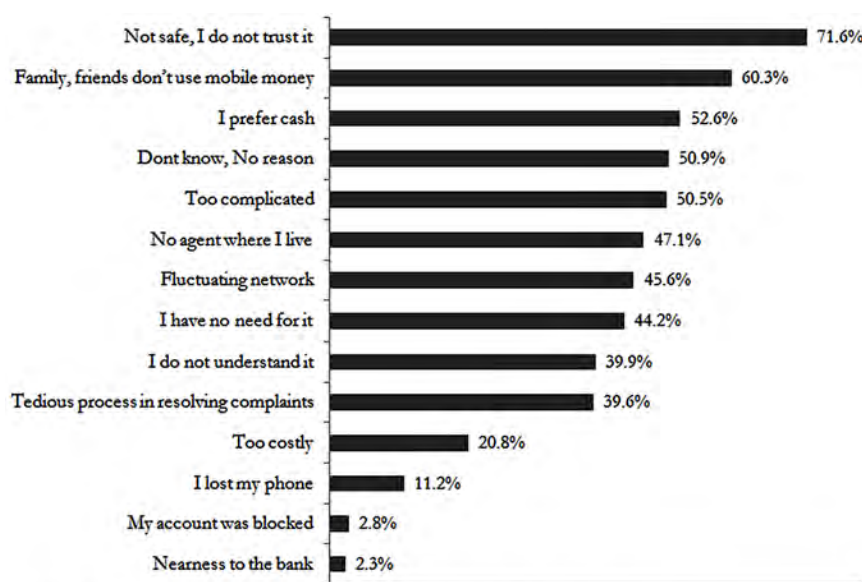
Figure 5 reveals some factors responsible for non-adoption of mobile money services among adult Nigerians who have never used mobile money services.

EFInA (2014) survey reported 87.3% of adult Nigerians as those unaware of mobile money services. Among these, 59.1% lived in rural ar-

eas where majority of the unbanked low-income earners are concentrated.

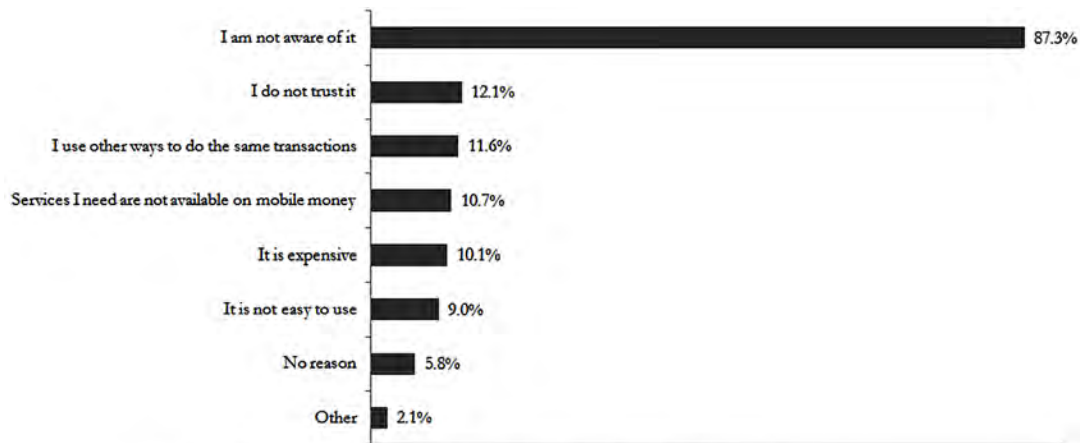
About 12.1% of non-adoption of mobile money is accounted for by lack of trust in the innovation which is ranked highest among factors that accounted for discontinued use of mobile money. EFInA (2015) *Mobile Money Customer Insights* identified fraud, network and transaction failure, limited number of agents and inadequate infrastructure as challenges experienced by active users of mobile money. Trust is a necessary precondition for trying mobile money and must be high, since for most users, their first interaction with a mobile money service will be to hand over cash to MMOs or their agent representatives (Davidson & McCarty, 2015, p.4). Intending adopters of a technological innovation often inquire from peers, friends and family members who have adopted or had precious/related experience with such technology and once reports relating to trust and security are not satisfactory, intention to use and subsequent adoption are negatively influenced. Davidson and McCarty (2015, p.2) further buttress that as customers lose trust in an operator's brand,

Figure 4. Causes of mobile money use discontinuance



**This figures are based on EFInA 2014 National Financial Inclusion Study covering 22,044 respondents*

Figure 5. Inhibitors of mobile money adoption



**This figures are based on EFINA 2014 National Financial Inclusion Study covering 22,044 respondents*

they are hesitant to conduct financial services on the platform.

Another inhibitor reported by non-adopters is the use of services which do not meet their financial services needs. David-West (2016, p.259) reports that although mobile money schemes offer services such as person-to-person remittances, mobile airtime purchases, utility and retail payments, savings etc., the duplication of services across providers is indicative of either low product innovation capabilities or unmet customer value propositions. Services offered on mobile money platforms must be able to provide customers with more benefits than other alternatives provided through the traditional brick and mortar platform or other digital financial services such as internet banking, ATMs and POS services otherwise, attracting customers remains difficult. Identifying the right customer segment and tailoring financial products that come with better comparative advantage when compared to existing products available to the target market segment is a step in the right direction. Such comparative advantage must be characterised by ubiquitous access, queue avoidance, and alternative payment methods (Shi, 2011, p.56).

Other factors affecting mobile money adoption include unavailability of preferred services

and products, and transaction cost. Although McKay and Pickens (2010) posit that the cost of using mobile money services is about 19% less than charges by banks, putting in perspective that majority (88%) of Nigerians live below \$2.50/day poverty line (Intermedia, 2015a) implies that intending adopters will be put away by the extra service charge.

Drivers of Mobile Money

In driving mobile money adoption, mass sensitization on the availability and customer-tailored services are pre-requisite. Davidson and McCarty (2015, p.3) argue that it is insufficient for consumers to know what mobile money is or the name of various service providers. Awareness campaign must champion the understanding of mobile money in order to influence positive behavioural change towards adoption and use. Costs associated with mobile money use must be less than cost charged by banks. Peruta (2015, p.6) further reports that the growing network of agents and mobile telephone density is an effective bridge to the access gap created by lack of widespread bank branches. GSMA (2012) reports that across 28 countries surveyed, there were more mobile money agents than bank branches. This further highlights the importance

of mobile money to closing financial exclusion gaps. Other drivers of adoption include timely access to financial assets, purchase possibilities irrespective of time and location, and as alternative to cash payments (Shi, 2011, p.34). These are subsumed and elaborated below:

- **Transaction Cost:** Micheni, *et al.* (2013) posit that transaction cost negatively influences adoption of mobile money and if cost is affordable and acceptable, there would be ease of adoption and use. However, cost is relative to the target market segment under consideration. When mobile money is deployed as a tool for financial inclusion, the target market segment is predominantly the rural poor who are unemployed, low education, financially illiterate and unwilling to incur cost on the simplest form of transaction – if possible, will appreciate such innovation to come free of charge.
- **Facilitating Conditions:** These refer to the organisational and technical resources and capabilities deployed to support users' exploitation of a system. Beyond providing infrastructural and technical support for customers to optimize the use of a system/technology, facilitating conditions also include support such as incentives, training/sensitizations, and customer support offered by technology and service providers to facilitate the adoption and use of a product or service – in this case, *mobile money*.
- **Comparative Advantage:** Many customers are not technologically confident, with low level of financial literacy (Micheni, *et al.*, 2013), unbanked with irregular or low income and often lack access to formal financial channels who in their vulnerability would readily trust mobile money providers to manage their money. Consequently, they want to stay close to their monies – which can be best provided for them in the form of facilitating conditions that offer them easy and fast access to their money.
- **Compatibility:** Shi (2011, p.35) reports that the compatibility of mobile payments with consumers' purchase transactions, habits, and preferences correspondingly influence further adoption and diffusion of mobile money. For mobile money systems to be adopted and used, consumers must be able to integrate such payment systems into their daily life which is an important aspect of compatibility.
- **Perceived Usefulness:** Across information systems research, the usefulness of a technology in meeting user needs is considered as a pre-condition for adoption, diffusion and use. Researchers in financial inclusion studies consider this as value proposition that the provider must clearly define in order to convince users of the need to adopt and use an innovation. David-West (2016, p.262) emphasised that the statement of access to the equivalency with cash in the mobile money space summarises customer value proposition. When value proposition is not clearly defined, perceived usefulness may be elusive, and adoption hindered.
- **Accessibility:** With liberalisation of the telecommunications sector, mobile telephony adoption has overcome socio-economic barriers to exceed banking penetration in under a decade (David-West, 2016, p.257). The delivery of financial services through this platform provides wider reach to millions of customers with access to mobile phones. Ubiquity of mobile money services is among several benefits that consumers expect to harness which further accelerates adoption and diffusion. This is likened to comparative advantage of Roger's (1995) diffusion of innovation theory. Thus, consumers must be able to see the relative advantage of mobile money over traditional and/or other form of banking services for adoption and use the technology.

Supporting Evidence From Available Data

From the EFINA (2014) data, evidence of perceived drivers of mobile money adoption are illustrated in Figure 6.

Users self reported benefits of mobile money services provide insights into its comparative advantage in deepening financial inclusion as compared to the brick and mortar banking. Fast and ease of transacting, social influence and safety and being alternative saving platform as shown in Figure 6 are reasons users remained committed to using mobile money. Although social influence

accounts for about 30%, the profile of the financially excluded - rural, low income, female, low education, unemployed and youthfulness (ages 18–35) combined to make the use of mobile money a viable tool for both domestic and international remittances amongst others especially in places where access and availability of traditional banking infrastructure is lacking.

About 38.7% of unregistered mobile money users would likely consider adoption once they have understanding of how it works, assured of safe transaction, proximity of mobile money agents, and mobile money attaining critical mass.

Figure 6. Reasons for use mobile money services

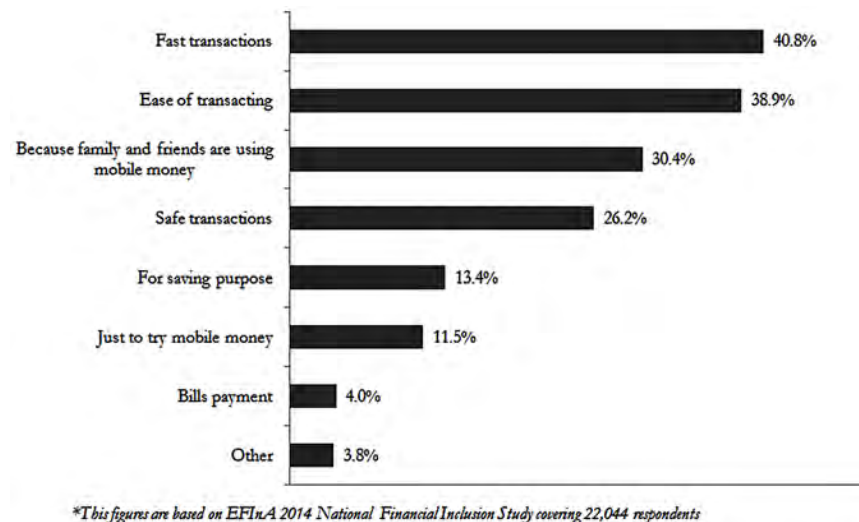


Figure 7. Mobile money adoption intention

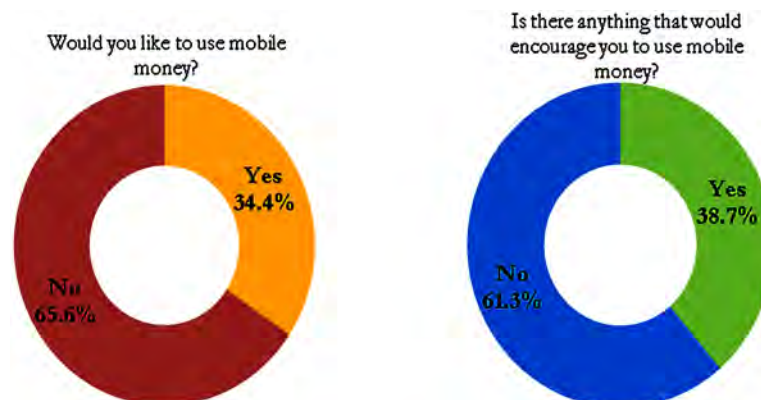
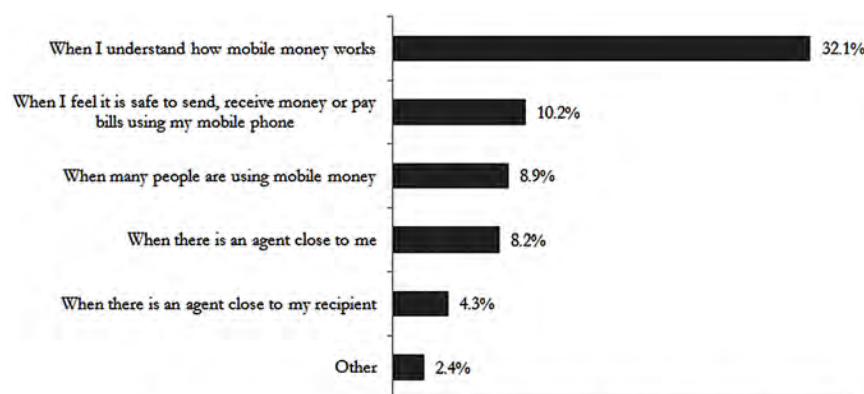


Figure 8. Facilitators to mobile money adoption



*This figures are based on EFINA 2014 National Financial Inclusion Study covering 22,044 respondents

Mobile Money Adoption: Far-Striking the Acceptance Process

Mobile money benefits transcend enhancing financial inclusion to include building cashless and cash-light economies, curbing corruption associated with financial transactions and offers diverse perspectives to monitoring cash flow among individuals and across destinations. Ease and convenience of transacting are added advantages that come with mobile money adoption and use. This argument is premised on the behavioural decision theory (Micheni *et al*, 2013, p.10) in which cost and benefits are significant to perceived usefulness and ease of use. Thus, if the cost of mobile money is affordable and there are several benefits relative to other financial service channels, adoption and use will be high.

Building customers' trust is key to achieving critical mass. Davidson and McCarty (2015, p.4) posit that trust is a necessary precondition for trying mobile money (both as a technology and as a service), which must be high, since for most users, their first interaction with a mobile money service will be to hand over cash. Known mobile money brand, extensive above-the-line advertising, trustworthy agents, and positive word of mouth all build trust. However, the most effective way to gain a customer's trust is to ensure that customers' experience with the service is good,

most importantly at the early stages of adoption; else, customers are unlikely to become regular users. One way of winning customers' trust is by offering credible services which according to Micheni, *et al* (2013, p.11) entails protection of information or systems from unsanctioned intrusions or outflows that affect mobile money adoption (Wang *et al.*, 2003).

The licensing of MNOs as mobile money providers can increase adoption and use among millions of customers registered on mobile telecommunications network (InterMedia, 2016a, p.1). Unfortunately, Nigeria's regulatory regime explicitly excludes MNOs as main providers of mobile money services. Efforts to close this gap and retain control of financial transactions under the regulatory scope of the Central Bank of Nigeria has resulted in the introduction of agency banking (David-West, 2016, p.260). As Nigeria's financial regulation advocates for mobile money as a tool for diving financial inclusion, more favourable regulations that offer win-win business models are needed.

FUTURE RESEARCH DIRECTIONS

The study explored factors inhibiting mobile money adoption in Nigeria, with the means and approaches by which the existing financial exclu-

sion gap among the financially served, underserved and un-served, could be bridged. However, while there are several factors identified as inhibitors to mobile money adoption in Nigeria, the study did not provide the impact level of each factor, which should feed regulators, policy makers and industry players in determining how best to deploy their limited resources among several factor requirements, and obtain mobile money adoption at scale. In light of this, future studies should deploy a predictive model to determine the relative importance of mobile money inhibitors.

CONCLUSION

In closing financial exclusion gaps, the adoption and use of mobile money presents the potentials for bridging access barriers, make financial services available, affordable and ultimately facilitate use of mobile money. David-West (2016, p.271) stressed that without the successful complementary initiatives and active stakeholders participation, mobile money cannot be perceived as the magic bullet of financial inclusion. As MMOs strive to improve service quality, access and adoption, government's involvement must be pronounced in driving of financial literacy, agency banking, microfinance banking, rural enterprise development as well as mobile money awareness. The MMOs must on their part strive to improve on service reliability, quality, security and partner with the mobile network operators whose data and voice services have become a daily dosage supporting the lives of millions of Nigerians.

ACKNOWLEDGMENT

This work was supported by the Bill & Melinda Gates Foundation [OPP1133684].

REFERENCES

- Allen, F., Demircuc-Kunt, A., Klapper, L., & Martinez, P. S. (2013). *Foundations for financial inclusion*. World Bank Policy. Research Working Paper.
- BMGF. (2013). *Fighting poverty, profitably: Transforming the economics of payments to build sustainable, inclusive financial systems*. Retrieved July 14, 2016 from <https://docs.gatesfoundation.org/Documents/Fighting%20Poverty%20Profitably%20Full%20Report.pdf>
- CGAP. (2009a). *Advancing Financial Access for the World's Poor. Annual report 2009*. Washington, DC: The World Bank Group.
- CGAP (2009b). *Financial Access 2009: Measuring access to financial services around the world*. Washington, DC: The World Bank Group.
- Commonwealth. (2014). The mobile money revolution in East Africa. *Commonwealth Broadband Forum*. Retrieved July 16, 2016 from http://www.cto.int/media/events/pst-ev/2014/Broadband%20Forum/Presentation%20PDF/Godfrey%20W%20Kyama_Central%20Bank%20of%20Kenya%20School%20of%20Monetary%20Studies.pdf
- David-West, O. (2016). The path to digital financial inclusion in Nigeria: Experiences of Firstmonie. *Journal of Payment Strategy and Systems*, 9(4), 256–273.
- Davidson, N., & McCarty, M. Y. (2015). *Driving Customer Usage of Mobile Money for the Unbanked*. Retrieved July 14, 2016 from <http://www.gsma.com/mobilefordevelopment/wp-content/uploads/2012/03/drivingcustomerusagefinal-lowres.pdf>
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *Management Information Systems Quarterly*, 13(3), 319–340. doi:10.2307/249008

- DeLone, W. H., & McLean, E. R. (2003). The DeLone and Mclean model of information systems success: A ten year update. *Journal of Management Information Systems*, 19(4), 31–47.
- Demirgüç-Kunt, A., & Klapper, L. (2012). *Measuring Financial Inclusion*. The Global Findex Database. World Bank Policy Research Working Paper 6025.
- Demirguc-Kunt, A., Klapper, L., Singer, D., & Van Oudheusden, P. (2015). *The Global Findex Database 2014: Measuring Financial Inclusion around the World*. Policy Research Working Paper 7255. World Bank.
- Devlin, F. (2005). A detailed study of financial exclusion in the UK. *Journal of Consumer Policy*, 28(1), 75–108. doi:10.1007/s10603-004-7313-y
- Donovan, K. (2012). *Mobile money for financial inclusion* (Conference Edition). Information and Communications for Development. doi:10.1596/9780821389911_ch04
- Dzokoto, V. (2013). *Making sense of mobile money in urban Ghana: Personal, business, social and financial inclusion prospects*. Research report. IMTFI. Retrieved July 17, 2016 from http://www.imtfti.uci.edu/files/docs/2014/dzkoto_final_report_feb2014.pdf
- EFinA. (2014). *Key Findings: The Financial Excluded Population in Nigeria*. EFinA: Did You Know Series, Series Three. Retrieved May 14, 2016 from <http://www.efina.org.ng/assets/ResearchDocuments/A2F-2014-Docs/Updated/EFinA-Access-to-Financial-Services-in-Nigeria-2014-Survey-Key-FindingswebsiteFINAL.pdf>
- EFinA. (2015). *Mobile Money Customer Insights: key findings from focus group discussions with users and non-users of mobile money*. Retrieved July 17, 2016 from <http://efina.org.ng/assets/ResearchDocuments/OtherResearch/Mobile-Money-Customer-Insights-Key-Findings-from-EFinA-FGDs.pdf>
- Ericsson ConsumerLab. (2016). *Financial services for everyone: Bridging the gap between the banked and the unbanked in Sub-Saharan Africa*. Retrieved July 19, 2016 from <http://www.ericsson.com/m-commerce/insights/consumerlab#subsaharan>
- GSMA. (2012). *Mobile money for the unbanked*. Retrieved July 14, 2016 from http://www.gsma.com/mobilefordevelopment/wp-content/uploads/2012/10/2012_MMU_Annual-Report.pdf
- InterMedia. (2015). *Nigeria – Quicksights Report: FII Tracker Survey*. Author.
- InterMedia. (2016a). *Cross-country comparison: Financial services in developing markets*. Author.
- InterMedia. (2016b). *Nigeria – Wave 3 Report FII Tracker Survey*. Author.
- Kama, U., & Adigun, M. (2013). *Financial inclusion in Nigeria: Issues and challenges*. Occasional Paper No. 45. Retrieved May 3, 2016 from <http://www.cenbank.org/out/2014/rsd/occasional%20paper%20no.%2045%20issues%20and%20challenges.pdf>
- Kasprowicz, P. & Rhyne, E. (2013). *Looking through the demographic window: Implications for financial inclusion*. Centre for Financial Inclusion, Publication 18.
- KPMG. (2015). *Sector report: Banking in Sub-Saharan Africa*. Retrieved July 19, 2016 from <https://www.kpmg.com/Africa/en/IssuesAndInsights/Articles-Publications/General-Industries-Publications/Documents/Banking%20in%20Africa%202015.pdf>
- Lal, R., & Sachdev, I. (2015). *Mobile money services – Design and development for financial inclusion*. Harvard Business School, Working Paper 15-083
- Martínez, C. H., Hidalgo, X. P., & Tuesta, D. (2013). *Demand factors that influence financial inclusion in Mexico: Analysis of the barriers based on the ENIF survey*. BBVA Research, 13/37 Working Paper.

- Mbiti, I., & Weil, D. (2011). *Mobile banking: The impact of M-Pesa in Kenya*. NBER Working Paper 17129.
- Mbiti, M. (2011). *Mobile banking: The impact of M-Pesa in Kenya*. NBER Working Paper 17129.
- McKay, C., & Pickens, M. (2010). *Branchless banking 2010: Who's served? At what price? What's next?* Washington, DC: The World Bank Group.
- Micheni, E. M., Lule, I., & Muketha, G. M. (2013). Transaction cost and facilitating conditions as indicators of the adoption of mobile money services in Kenya. *International Journal of Advanced Trends in Computer Science and Engineering*, 2(5), 9–15.
- NOI Polls. (2013). *Adoption of Mobile Money Services Slow Due to Low Public Awareness*. Retrieved July 17, 2016 from <http://www.noi-polls.com/root/index.php?pid=249&ptid=1&parentid=13>
- Park, C., & Mercado, R. V. (2015). *Financial inclusion, poverty, and income inequality in developing Asia*. ADB Economics Working Paper Series No. 426. Retrieved July 14, 2016 from <http://www.adb.org/sites/default/files/publication/153143/ewp-426.pdf>
- Peruta, M. D. (2015). *Mobile money adoption and financial inclusion objectives: A microeconomic approach through a cluster analysis*. GREDEG Working paper, No.2015-49. Retrieved May 3, 2016 from <http://www.gredeg.cnrs.fr/working-papers/GREDEG-WP-2015-49.pdf>
- Rogers, E. (1995). *Diffusion of innovation*. New York: Free Press.
- Sarma, M. (2008). *Index of financial inclusion*. Indian Council for Research on International Economic Relations, Working Paper No 215. Retrieved June 22, 2016 from http://icrier.org/pdf/Working_Paper_215.pdf
- Sarma, M., & Pais, J. (2011). Financial inclusion and development. *Journal of International Development*, 23(5), 613–628. doi:10.1002/jid.1698
- Shi, X. (2011). *Exploring factors that hinder the adoption of mobile services in China: A qualitative user analysis with special focus on mobile financial services* (Master's thesis). School of Economics, Aalto University. Retrieved June 29, 2016 from https://aaltodoc.aalto.fi/bitstream/handle/123456789/782/hse_ethesis_12650.pdf?sequence=1&isAllowed=y
- The Global Findex. (2014). Expanding financial inclusion for Women and the poor. Infographic: Global Findex 2014. *Gender and Income*. Retrieved June 1, 2016 from <http://www.worldbank.org/en/programs/globalfindex/infographics/infographic-global-findex-2014-gender-income>
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Towards a unified view. *Management Information Systems Quarterly*, 27(3), 425–478. doi:10.2307/3250981
- Wang, Y. S., Wang, Y. M., Lin, H. H., & Tang, T. I. (2003). Determinants of user acceptance of internet banking: An empirical study. *International Journal of Service Industry Management*, 14(5), 501–519. doi:10.1108/09564230310500192
- Zimmerman, J., Pearse, J., & Bau, S. (2016). 5 ways an emergency mobile wallet can lead to financial inclusion. *CGAP: Advancing financial inclusion to improve the lives of the poor*. Retrieved July 10, 2016 from <http://www.cgap.org/blog/5-ways-emergency-mobile-wallet-can-lead-financial-inclusion>

ADDITIONAL READING

David-West, O., Umukoro, I. O., & Muritala, O. (2016). Impact of consumers' community and individual dynamics on financial inclusion in Nigeria. *AIB-SSA 2016 Conference Proceedings*. Available at <http://www.aibssa.org/wp-content/uploads/2016/09/AIB-SSA-2016-Conference-Proceedings.pdf>

Etim, A. S. (2014). Mobile banking and mobile money adoption for financial inclusion. *Research in Business and Economics Journal*, 9, 1–13.

Krishna Kishore, S.V., & Sequeira, A.H. (2016). An empirical investigation on mobile banking service adoption in rural Karnataka. *Sage Open*, 1-26. DOI: 10.1177/2158244016633731

McKay, C. (2011). Ghana: Aiming for interoperability in branchless banking. *Journal of Economic Literature*, 35, 34–44.

Ondiege, P. (2015). Regulatory impact on mobile money and financial inclusion in African countries - Kenya, Nigeria, Tanzania and Uganda. *Centre for Global Development*. Retrieved July 17, 2016 from <http://cgdev.org.488elwb02.blackmesh.com/sites/default/files/Background-Paper-Financial-Inclusion-Ondiege.pdf>

Villasenor, J. D., West, D. M., & Lewis, R. J. (2016). Advancing Equitable Financial Ecosystems. *The 2016 Brookings Financial and Digital Inclusion Project Report*. Available at https://www.brookings.edu/wp-content/uploads/2016/08/fdip_20160816_project_report.pdf

KEY TERMS AND DEFINITIONS

Financial Inclusion: This is a state in which all bankable members of an economy have ease of access to and use available formal financial services at an affordable cost.

Financially Excluded: These are unbanked members of an economy without access to any form of financial services.

Financially Served: This refers to banked individuals with access to a bank account either with a deposit money bank (DMB) or microfinance bank (MFB) or who make use of services provided by other financial institutions such as insurance companies or pension schemes.

Financially Under-Served: These are members of an economy who are under-banked with access to informal financial services only.

Mobile Money: The use of a mobile phone to conduct financial services using electronic accounts (e-wallet).

Mobile Money Adoption: This refers to a state in which mobile money service is accepted as either an alternative or supplement to existing financial service products.

Mobile Money Operator (MMO): This is a licensed mobile money service provider that develops and deploys financial services through mobile phones and mobile telephone networks.

Mobile Money Services: These include the range of financial services offered by MMOs, such as cash-in, cash-out, utility/bill payments, remittances (domestic and international), airtime top-up, savings, credits, etc.

E-Commerce Models, Players, and Its Future

E**Liguo Yu***Indiana University – South Bend, USA*

INTRODUCTION

Wikipedia defines e-commerce (electronic commerce) as the trading or facilitation of trading of products or services using the Internet. E-commerce is becoming a major business model in developed countries. For example, in US, the e-commerce sales as ratio of total retail sales reached 7.5% in December 2015; in China, the e-commerce sales as ratio of total retail sales reached 9.7% in July 2015; and it is estimated that globally, the e-commerce sales will reach \$ 1.5 trillion in 2018. Furthermore, e-commerce is growing and will continue to grow, especially in developing countries.

E-commerce plays an important role in our modern economy (Goldfarb, Greenstein, & Tucker, 2015), and its effect goes beyond economy and business. E-commerce is becoming an integral part of our society and our life. It is therefore necessary to clearly understand e-commerce operations and impacts. In this chapter, we examine various e-commerce models, major players, and the environmental ecosystem factors in order to understand how e-commerce has shaped our new way of conducting business and how it is going to influence our future. The remaining of this chapter is organized as follows. Section 2 reviews e-commerce history. Section 3 introduces various e-commerce models. Section 4 describes the major e-commerce players in the global business. Section 5 describes emerging models and emerging market. Section 6 describes the future trend of e-commerce. Conclusions are in Section 7.

BACKGROUND

Self-sufficiency was considered a primitive lifestyle in early human society. Once communications and interactions became popular in human society, trading gradually replaced self-sufficient economy (Watson, 2006). The advantages of commerce/trading over self-sufficient economy include the increase of productivity of human being and the increase of diversity of consuming products/services. Commerce (trading), as a basic activity of human society, can be traced back to prehistoric time and its evolution can be categorized into the following stages.

- **Barter System:** Barter system is the direct exchange of products or services with other products or services without using a medium. This kind of trading model can be traced back to 150 thousand years ago (Watson, 2006).
- **Currency System:** In currency system, standardized exchange medium, money, was introduced. With the support of currency, trading became more convenient. The development of original currency, metal coins, is believed to have originated about 500 BC.
- **Banknote:** The first development of a local banknote (paper currency) began about 700 AD in China. With its many advantages over metal coins, paper currency or paper money was gradually adopted nationally in China around 1100 AD.

- **E-Currency:** Electronic money was introduced within the modern banking system. Examples of e-currency include bank deposits, fund transfer, and electronic payment. Using e-currency, money balance and transaction history are stored in a computer system and users who have the access key could manage their money. The access key could be a bank card or an online password. The first bank card was introduced in 1946 by John Biggins, a banker in New York, and the first online banking system was introduced in 1983 by the Bank of Scotland.
- **E-Commerce:** The first e-commerce model was proposed and built in 1970s, which was long before the wide use of the Internet. It was created in student laboratories of Stanford University and Massachusetts Institute of Technology (Markoff, 2005). In the following half century, various e-commerce models are proposed, tested, and implemented (Tkacz & Kapczynski, 2009; Palmer, 1988; Kelly, 2005).

E-commerce, like any other internet-based applications, evolves with the revolution of information technology (Rayport & Jaworski, 2002). On an e-commerce platform, not only payments can be transacted electronically, placing orders, processing orders, and tracking orders all can be executed electronically. After about 150 thousand years of evolution, human beings are entering a completely different business era, which is characterized by e-commerce.

E-COMMERCE MODELS

There are many ways a business activity could be conducted online. The most commonly seen e-commerce models include B2B, B2C, and C2C, which are responsible for majority of the online business transactions. These common models together with some other models are described below.

B2B Model

In B2B (Business-to-Business) model, one business purchases products from another business and then sells the products to the final customer (Haig, 2001; King, 2015). In this model, business operations, such as product ordering, payment transactions, and order tracking, could all be accomplished online. This B2B model significantly reduces the business cost compared with the traditional business purchasing process. Table 1 lists some of the most visited B2B websites globally¹. We should not be surprised to see that most commonly visited B2B websites are in China, because China is currently the world's largest manufacturer. The retail business worldwide is importing from China. Therefore, B2B e-commerce could facilitate international trading.

B2C Model

In B2C (Business-to-Consumer) model, a business sells its products directly to the final customer (Elliott, 2002). In this model, consumers visit a B2C website and purchase products. Table 2

Table 1. Top B2B websites measured with global traffic rank

Company	Website	Global Traffic Rank	Languages	Home Country
Alibaba	alibaba.com	81	Multiple languages	China
IndiaMart	indiamart.com	1,145	English	India
Made-in-China.com	made-in-china.com	1,170	Multiple languages	China
DHgate.com	dhgate.com	1,186	Multiple languages	China

Table 2. Top B2C websites measured with global traffic rank

Company	Website	Global Traffic Rank	Languages	Home Country
Amazon	amazon.com	6	Multiple languages	USA
Alibaba	tmall.com	37	Chinese	China
JD.com	jd.com	88	Chinese	China
Walmart	Walmart.com	110	English	USA

lists some of the most visited B2C websites globally². Although Alibaba's Tmall is considered a B2C model, its operation is quite different from the others. For Amazon, JD.com, and Walmart, they are the sellers, which means they purchase products from manufacturers, store products in their own warehouses, and when orders are placed by consumers, they need to ship the products. For Alibaba's Tmall, it is just a B2C platform. Manufacturers and retailers could open a store in Tmall and Alibaba is responsible for payment transactions and the store owner is responsible for shipment. Therefore, Alibaba does not directly handle products, including purchasing, warehousing, and shipping.

C2C Model

In C2C (Consumer-to-Consumer) model, a consumer buys a product directly from another consumer (Turban, King, & Wang, 2003). Table 3 lists some of the most visited C2C websites globally. Most C2C websites, such as eBay and Taobao of Alibaba, offer services to connect consumers (sellers and buyers) of a product. In contrast, Craigslist operates on a different business model, through which it offers classified advertising services to get consumers connected.

Other Models

In addition to the three aforementioned e-commerce models, there are other Internet-based business models. Some people consider them as e-commerce and others do not. These models include C2B (Consumer-to-Business), B2G (Business-to-Government), G2B (Government-to-Business), and G2C (Government-to-Citizen). For example, in C2B model, consumers could offer their products/services online and companies can buy and pay for the products/services. Examples of C2B websites include fotolia.com, GoogleAdsense, and Amazon Affiliate Program.

Vertical e-commerce is a business model that the sellers are specialists of a kind of products and the e-commerce website is targeted to a specific group of consumers. For example, beauty.com is a cosmetics vertical e-commerce website; drugstore.com is a medicine vertical e-commerce website; zappos.com is a shoe and clothing vertical e-commerce website; and zulily.com is a vertical e-commerce website featuring clothing, toys, and home products. Although usually vertical e-commerce business is much smaller than the horizontal e-commerce business, such as Amazon, Alibaba, and eBay, which sell products of various categories, there are many vertical e-commerce websites and they play important roles in the retail industry.

Table 3. Top C2C websites measured with global traffic rank

Company	Website	Global Traffic rank	Languages	Home Country
Alibaba	taobao.com	12	Chinese	China
eBay	ebay.com	23	Multiple languages	USA
Craigslist	craigslist.org	71	Multiple languages	USA
Etsy	etsy.com	130	English	USA

4. THE BIG E-COMMERCE PLAYERS

E-commerce, like any other internet-based business, can grow fast and big without the locality limitations. Currently, the two biggest e-commerce companies worldwide are Amazon and Alibaba. Their business are described below.

Amazon

In 1995, Jeff Bezos launched Amazon.com as an online bookstore (Stone, 2013). After 20 years growth, it now becomes the most valued e-commerce company in the world. Although Amazon is mainly operated in the United States, its business has been expanded all around the world, even in the most competitive marketplace like China.

Amazon typically follows a B2C business model. It purchases products from the manufacturers, store products in its own warehouse,

and deliver the products when orders are placed by consumers. The product categories listed in Amazon include books, audio/video/games, electronics, clothing, toys, cosmetics, sports, appliance, furniture, office supplies, auto care, and more. Amazon is well known for its fast delivery and high rate of customer satisfaction.

Globally, Amazon is actively growing its business. Currently, products can be shipped to more than 75 countries that participate in the Amazon Global program. Amazon also produces consumer electronics, such as Kindle e-book reader, Fire tablet, and Fire phone. Moreover, Amazon's web service is one of the most important cloud computing providers in the world. Table 4 shows the business and finance growth of Amazon in the past 18 years.

By the end of 2015, Amazon is ranked No. 9 of the public corporations worldwide based on market capitalization. The most important fact

Table 4. Amazon's growth (data source: various websites on the Internet)

Year	Number of Employees (Thousand)	Annual Active Users (Million)	Annual Revenue (\$ Billion)	Market Capitalization (\$ Billion)
1997	0.26	1.5	0.148	2.36
1998	1.2	6.2	0.610	28.11
1999	3	14	1.640	31.04
2000	4.5	20	2.762	8.32
2001	6	25	3.122	6.82
2002	7	30	3.933	10.51
2003	8	40	5.264	24.23
2004	9	45	6.921	20.78
2005	12	50	8.490	21.55
2006	13.9	66	10.710	18.11
2007	15.8	76	14.835	37.36
2008	20	88	19.166	28.28
2009	25	105	24.509	60.30
2010	35	130	34.204	81.57
2011	55	164	48.077	93.49
2012	80	200	61.093	127.66
2013	115	237	74.452	172.47
2014	150	270	88.988	170.47
2015	220	304	107.01	282.25

about Amazon is that its business is still expanding at about 19% year to year rate of revenue growth.

Alibaba

Alibaba is an international internet giant. Fairly to say, Alibaba is not only an e-commerce company. It provides many different services to e-commerce business and consumers. Therefore, Alibaba is an e-commerce environment, in which different kinds of business operations could be executed.

Alibaba group was founded in 1999 by Jack Ma (Liu & Avery, 2009). It provides various e-commerce solutions, such platforms for B2B, B2C, and C2C. It also provides payment transaction service through its own payment system, Alipay. In addition, Alibaba contracts with many package delivery companies to provide logistics services. Big data technology is used in these services to improve the business efficiency of its partners.

That is not all. Alibaba has many other business, including travel arrangement, advertisement, news and media, music, sport, health, entertainment, financial, banking, and insurance. Alibaba also has its software products and cloud computing services. Moreover, Alibaba is actively investing in foreign countries and foreign companies, including both public and private companies. Examples are Zulily, Groupon, Snapchat, Jet.com, and Magic Leap of USA, and Paytm, Snapdeal of India.

To understand the volume of business transactions conducted on Alibaba's platforms, let's look at the sales of Singles' Day (11.11) shopping festival created by Alibaba. In November 11, 2015, a total of \$14.3 billion sales were made at Alibaba's platforms, which is more than the combinations of Black Friday sales and Cyber Monday sales in USA. Moreover, Alibaba has a 45% year to year rate of revenue growth, which is faster than its American counterpart Amazon.

As stated by the founder and executive chairman of Alibaba, Jack Ma, Alibaba is building an internet-based business ecosystem that is going to connect the world together, from cities to cities and from villages to villages. Jack ma calls his plan

EWTO (Electronic World Trading Organization). To help the readers better understand Alibaba's internet-based business ecosystem, Table 5 compares Alibaba's business with some corresponding US companies. It is worth to note that Alibaba is expanding its business both domestically and internationally. Currently Alibaba has connected buyers and sellers of more than 240 countries. Globalization is at the core of Alibaba's strategic plan over the next decade.

Other Players

Besides Amazon and Alibaba, other major e-commerce companies in the world include JD.com of China and eBay of USA. JD.com is similar to Amazon and operates with a B2C business model, and it is currently the major competitor of Alibaba in China. In contrast, eBay is an America-based C2C e-commerce website. eBay has successfully developed its business around the world, except China, where it lost its battle against Alibaba's Taobao (Ou & Davison, 2009), at least for now. Table 6 summarizes the latest financial data of Amazon, Alibaba, JD.com, and eBay. The data is retrieved in February 2016. These four e-commerce companies are responsible for about 46% of the global e-commerce GMV (Gross Merchandise Volume)³.

It is worth making the following remarks about the data in Table 6. First, Alibaba and eBay are e-commerce platforms and accordingly their revenues do not include the values of products traded on their platforms. On the contrary, the revenue of Amazon and JD.com include all the product values traded on their websites. Second, it is expected that Alibaba and JD.com could have higher growth rates in the next few years and that is why their market capitalizations is relatively higher compared to their net incomes.

Table 5. Alibaba's internet-based ecosystem (data source: various websites on the Internet)

Business Name	Service Description	Operation Method	Corresponding USA Company/Service
Alibaba.com	B2B (international)	Platform	NA
1688.com	B2B (domestic)	Platform	NA
TMall	B2C (domestic, international import)	Platform	Amazon
Aliexpress	B2C (international export)	Platform	Amazon
Taobao	C2C (domestic)	Platform	eBay
Alipay	Online payment	Directly owned	PayPal
Cainiao	Logistics	Collaborated	UPS
Alitrip	Travel arrangement	Directly owned	Priceline
Juhuasuan	Group and discount shopping	Directly owned	Groupon
AutoNavi	Map	Directly owned	Google map
Youku Tudou	Video sharing and broadcasting	Directly owned	YouTube
SCMP	Newspaper	Directly owned	Washington Post
Alimama	Online advertisement	Directly owned	Google
YunOS	Mobile operating system	Directly owned	Android
UCWeb	Mobile web browser	Directly owned	Safari
Sogou	Mobile search engine	Directly owned	Google
Aliyun	Cloud computing	Directly owned	Amazon Web Service
Baozun	E-commerce service	Partially owned	Vteams
Sina	Online news services	Partially owned	The Huffington Post
Weibo	Micro blogging	Partially owned	Twitter
Momo	Social networking	Partially owned	Facebook
Kuaidi	Online taxi dispatch	Partially owned	Uber
Meizu	Smart phone manufacture	Partially owned	Apple

Table 6. Comparisons of the financial results of the big e-commerce companies (data source: various websites on the Internet)

Business Name	Number of Employee	Monthly Active User	Annual GMV	Annual Revenue	Annual Net Income	Market Capitalization
Amazon	230,800	244 M	\$225.6 B	\$107 B	\$596 M	\$ 261 B
Alibaba	34,985	405 M	\$376 B	\$11.7 B	\$5.38 B	\$168 B
JD.com	94,615	131 M	\$71.4 B	\$28 B	-\$131 M	\$36 B
eBay	11,600	162 M	\$82 B	\$8.59 B	\$2.23 B	\$28.5 B

EMERGING MODELS AND EMERGING MARKET

Although the world e-commerce business is largely shared by Amazon, Alibaba, and several other big players, no one business and no one business

model can dominate the market forever. Some new e-commerce models have surfaced recently. The e-commerce future is unknown, especially in the emerging market.

One new e-commerce model is represented by Jet.com. It is an e-commerce website launched

in July 2015. The unique feature of Jet.com is its smart shopping cart, which can help the customer to build a shopping list with more potential savings. According to its business strategy, Jet.com makes no profit through selling products. Instead, the income is generated through member fees. Collecting membership fee is not rare in e-commerce business (Monteiro, Swatman, & Tavares, 2003). Other e-commerce companies, like Vipshop, Sephora, also have paid membership programs.

O2O (Online-to-Offline) is considered another important emerging e-commerce model (Roberts & Zahay, 2012). This model combines online shopping and offline (physical store) shopping together and provides unique shopping experiences. For example, a product could be ordered online and picked up from or delivered by a local physical store; or a service, such as movie ticket, club membership, could be ordered online and consumed offline locally. Therefore, O2O model is also called local e-commerce. It can be integrated into local business and accordingly it is well supported by the local government. O2O is becoming a more and more popular life style in China. Examples of O2O services in China include Meituan.com, 58.com. Internet giants Alibaba and Baidu also have their O2O business, which are Koubei and Nuomi, respectively. Relatively speaking, O2O business is still under development in USA. Groupon can be classified into this field.

Another important trend in e-commerce industry in the past years is that more and more customers like to shop with mobile devices. For example, in the quarter ended December 2015, Alibaba's mobile penetration rate is about 67% and this number was only 19.7% two years ago. Mobile devices, such as smart phones and tablets, provide convenient ways and pleasant experiences for users to shop online. With devices connected online all the time, the customers would like to shop more frequently. Therefore, e-commerce companies are shifting their core business from desktop computers to mobile devices.

Although e-commerce is a common business model in developed countries, especially in USA and China, it is still under development in many

other countries. There are great business potentials around the world. Using India as an example, it is estimated that its e-commerce could reach \$100 billion by 2019. Currently, Amazon, Alibaba, and India's domestic e-commerce companies, such as Snapdeal, Flipkart are fighting hard to grow their market share there.

FUTURE RESEARCH DIRECTIONS

The key factors to e-commerce, just like to any other business, are user addition and user retention. That is why JD.com is collaborating with Tencent to get more users from Wechat, a communication platform and Alibaba is building its ecosystem to maintain a sustainable customer base. Friendly user interface, secure and convenient payment method, fast and quality delivery are other important factors to improve user retention rate. For new generations of consumers, shopping is no longer the single purpose of shopping online; shopping experience could be more important than the product. Like the traditional shopping mall, entertainment, social networking, education, and more, could all be built into e-commerce. An online expectant mother club incorporating all these features to allow pregnant women to socialize, obtain parenting knowledge and experience, and most likely, share baby products shopping information, is a perfect example of social e-commerce. Another examples of social e-commerce is the online stores built on top of social networking platforms, such as Facebook and WeChat. In this model, products/service are sold to friends or followers. This is a kind of peer to peer business model and some business cost, such as advertisement, could be significantly reduced. Most importantly, with the growing usage of mobile devices that connect users all the time, social e-commerce could become more and more popular.

E-commerce companies and customers are important players in an e-commerce environment. However, they are not the only players. No business can be operated without the influence of other

factors, such as government policies, international laws, cultures, religions, and geopolitics. The roles played by different parties, such as governments and other organizations, in an e-commerce environment could be complex and their influence may shape the future business rules and models of e-commerce. A detailed study in this area is surely needed for the future research. For example, Amazon originally only collected sales tax from a few states; after a long battle between Amazon and the state governments, it now collects sales tax from 28 states. Another example is Alibaba. Currently, the stores at Alibaba's Tmall and Taobao platforms are exempted from the 4% business tax that is unavoidable for offline physical stores. No one is expected this policy will remain forever. Clearly, any future changes of this tax rule may have effect on Alibaba's business. In addition, globalization is the theme of the 21st century, to expand business in developing countries, especially those unrest regions with cross cultural and religious conflicts, is a challenging task for Amazon, Alibaba, and any ambitious e-commerce companies.

E-commerce is growing and is growing at an unprecedented speed. Its impact on our life is extraordinary. Commerce started in the prehistoric time and the business model has not changed much for thousands of years. At one time, all stores were small with merchandise sold to local consumers. But, industrial revolution finally led to the changes of our commerce model. In the 20th century, chain stores, department stores, and especially supermarkets, gradually replaced small family stores, and our lifestyle was changed. Now, less than 100 years later, information technology revolution is going to change again our commerce model, which in turn will affect our life.

7. CONCLUSION

In this chapter, we discussed e-commerce models, its big players, and its future trend. E-commerce is becoming an important integral part of our

economy and our life. E-commerce originated and developed in the era of internet and will continue to develop with the evolution of information technology. It is hard to predict the future. However, one thing we are sure is that our lifestyle will always change, including the way of doing business.

E-commerce, like any other information technology supported business activities, will play an important role in shaping our society and life. We are living in a time full of possibilities. The potential of e-commerce is yet to be realized.

REFERENCES

- Elliott, S. (2002). *Electronic Commerce: B2C strategies and models*. John Wiley & Sons.
- Goldfarb, A., Greenstein, S. M., & Tucker, C. E. (2015). *Economic Analysis of the Digital Economy*. University of Chicago Press. doi:10.7208/chicago/9780226206981.001.0001
- Haig, M. (2001). *The B2B E-Commerce Handbook: How to Transform Your Business-to-Business Global Marketing Strategy*. Kogan Page Publishers.
- Kelly, K. (2005). We are the Web. *Wired Magazine*, 13(8), 113–123.
- King, K. A. (2015). *The Complete Guide to B2B Marketing: New Tactics, Tools, and Techniques to Compete in the Digital Economy*. Pearson Education.
- Liu, S., & Avery, M. (2009). *Alibaba: The Inside Story behind Jack Ma and the Creation of the World's Biggest Online Marketplace*. Collins Business.
- Markoff, J. (2005). *What the Dormouse Said: How the Sixties Counterculture Shaped the Personal Computer Industry*. Penguin.
- Monteiro, J., Swatman, P. M., & Tavares, L. V. (2003). *Towards the Knowledge Society: Ecommerce, Ebusiness and Egovernment*. Springer. doi:10.1007/978-0-387-35617-4

Ou, C. X., & Davison, R. M. (2009). Why eBay lost to TaoBao in China: The Glocal Advantage. *Communications of the ACM*, 52(1), 145–148. doi:10.1145/1435417.1435450

Palmer, C. (1988). Using IT for competitive advantage at Thomson Holidays. *Long Range Planning*, 21(6), 26–29. doi:10.1016/0024-6301(88)90155-0

Rayport, J. F., & Jaworski, B. J. (2002). *Introduction to E-Commerce*. McGraw-Hill/Irwin.

Roberts, M., & Zahay, D. (2012). *Internet Marketing: Integrating Online and Offline Strategies*. Cengage Learning.

Stone, B. (2013). *The Everything Store: Jeff Bezos and the Age of Amazon*. Random House.

Tkacz, E., & Kapczynski, A. (2009). *Internet-Technical Development and Applications* (Vol. 64). Springer Science & Business Media. doi:10.1007/978-3-642-05019-0

Turban, E., King, D., & Wang, J. (2003). *Introduction to E-Commerce*. Upper Saddle River, NJ: Prentice Hall.

Watson, P. (2006). *Ideas: a History of Thought and Invention, from Fire to Freud*. Harper Perennial.

ADDITIONAL READING

Brandt, R. L. (2011). *One Click: Jeff Bezos and the Rise of Amazon.com*. Penguin.

Hashemi, S., & Hashemi, B. (2004). *Anyone Can Do It: Building Coffee Republic from Our Kitchen Table-57 Real-Life Laws on Entrepreneurship*. John Wiley & Sons.

Hunt, B. (2011). *Convert: Designing Web Sites to Increase Traffic and Conversion*. John Wiley & Sons. doi:10.1002/9781118255827

King, D., & Lee, J. (2000). *Electronic Commerce: A Managerial Perspective* (Vol. 1). Upper Saddle River, NJ: Prentice Hall.

Nahai, N. (2014). *Website Branding for Small Businesses: Secret Strategies for Building a Brand, Selling Products Online, and Creating a Lasting Community*. Skyhorse Publishing, Inc.

Plant, R. T. (2000). *Ecommerce: Formulation of Strategy*. Prentice Hall PTR.

Reynolds, J., & Mofazali, R. (2000). *The Complete E-Commerce Book: Design, Build, and Maintain a Successful Web-Based Business*. Cmp Books.

Rognerud, J. (2008). *Ultimate Guide to Search Engine Optimization: Drive Traffic, Boost Conversion Rates and Make Lots of Money*. Jon Rognerud SEO.

Sung, T. K. (2006). E-commerce critical success factors: East vs. West. *Technological Forecasting and Social Change*, 73(9), 1161–1177. doi:10.1016/j.techfore.2004.09.002

KEY TERMS AND DEFINITIONS

B2B (Business-to-Business): An e-commerce activity involves two business.

B2C (Business-to-Consumer): An e-commerce activity in which a product is sold by a business and to a consumer.

C2C (Consumer-to-Consumer): An e-commerce activity in which a product is sold by one consumer to another consumer.

E-Commerce: The business activity of trading products or services online.

GMV (Gross Merchandise Volume): The total sales of the values of products by a retailer in a specific period of time.

Horizontal E-Commerce: A website that sells various categories of products.

Internet-Based Business Ecosystem: A business environment built on top of the Internet to facilitate trading between sellers and buyers and provide services for various business activities.

Local E-Commerce: An O2O model that is integrated with local business, where products/services are ordered online and consumed locally.

Market Capitalization: The total market value of a company's outstanding shares.

MAU (Monthly Active User): The average number of users that conduct business activities on an e-commerce platform in one month period.

O2O (Online-to-Offline): An e-commerce activity that combines online shopping with offline (physical store) shopping.

Social E-Commerce: An e-commerce built on top of social networking platforms, where products/services are sold to friends or followers in the social network.

Vertical E-Commerce: A website that sells a specific category of products.

ENDNOTES

¹ <http://www.b2bsiteranking.com/>

² <http://www.alexa.com/>

³ <https://www.internetretailer.com/2015/07/29/global-e-commerce-set-grow-25-2015>

Electronic Payment Frameworks

E

Antonio Ruiz-Martínez
University of Murcia, Spain

Oussama Tounekti
University of Murcia, Spain

Antonio F. Skarmeta
University of Murcia, Spain

INTRODUCTION

The use of electronic payment (e-payment) systems to make payments for the access to contents and services is more and more usual. Currently, for making e-payments we can use systems such as Paypal, Bitcoin, Ripple, Apple Pay, EMV. Indeed, the use of cryptocurrencies as Bitcoin has supposed a revolution in the e-payments arena (Hileman, 2014; Vigna & Casey, 2015).

This diversity of payment options with the support of different business models (pay-per-use, flat rate, the incorporation of discounts, tickets or loyalty information, etc) arises the need of systems that support different issues regarding the purchase such as the discovery of the payment information associated to the product/service with payment options supported, the negotiation of the payment option to make the payment, and the reception of a receipt of the transaction. To cope with these issues, the definition of e-payment frameworks has been proposed. These frameworks aim to facilitate, along the purchase process, the exchange of payment information and the use of different payment instruments in an easy way at the same time it guarantees interoperability, trust and security (Jaffe & Boyera, 2015; Ruiz-Martínez et al., 2012).

This article presents what an e-payment framework is, its main features and benefits, what the main e-payment frameworks proposed so far are and, the current initiatives that are being devel-

oped. For each framework developed so far, the article presents its key features and differences with previous works. The presentation of the different frameworks proposed so far will allow the reader to understand the evolution of these frameworks and how different features have been incorporated along the time. Once previous works have been presented, the article introduces the most recent work in this field, that is, the work that is being developed by W3C with its Web Payment Activity Initiative. Then, we present future research directions by indicating the main challenges to be addressed. Thus, the reader will have a broad vision of e-payment frameworks.

BACKGROUND

This section provides some information on the different phases that comprises a payment process. The knowledge of this process is fundamental to understand the different elements that are defined in an e-payment framework.

Payment Process

The process of making purchase on the Internet is divided into the following stages (Maes et al., 1999): need identification, product brokering, merchant brokering, negotiation, purchase and delivery, and product service and evaluation. Thus, once the user realizes that he/she has a need, it

gathers information regarding the product and the merchant, then, client can negotiate the conditions with the vendor. Once the conditions have been agreed, the purchase and delivery of the product is made. Finally, several processes related to user satisfaction could be performed.

Recently, the W3C's Web payments activity initiative (W3C, 2014) considered that the simplest purchase comprises four steps: publishing an asset, publishing a license, publishing an offer, and performing the purchase.

In both cases, we can point out that, in order to make a purchase, we have to manage two kinds of information: information describing the product/service and its merchant, and information regarding the payment process: payment systems available to perform the payment, receipts, etc.

ELECTRONIC PAYMENT FRAMEWORKS

This section aims to provide the definition of what an e- payment framework is, what its main features are and its benefits.

Definition

In the literature there are different definitions of what a e-payment framework is. Next, the most relevant are presented.

Fischer et al. (2002) consider that an electronic payment framework aims to integrate payment instruments into e-commerce applications and should, at least, address the specificity of the payment instrument and the underlying protocol.

In SEMPER Framework for Secure Electronic Commerce (Schunter, 1999; Lacoste et al., 2000), it is considered that, from a technical point of view, a payment framework describes a generic model for performing e-commerce transactions between two parties through a set of services (supporting, transfer, exchange, commerce services) enable the integration of any payment protocol and product in such as way that applications were not restricted

to the use of specific protocols or proprietary solutions or technology.

Ruiz-Martínez (2015) mentions that the purpose of a web payment framework is to facilitate the purchase process and the exchange of payment information, supporting that different payment instruments can be used and guaranteeing interoperability, security, and trust.

In this paper a payment framework is defined as a set of extensible elements and/or services, generally, organized within a set of layers and that is designed to facilitate the use of different payment systems in a unified manner during a payment process. Thus, for the developers of payment-based applications, the incorporation of new payment systems is easier, transparent and can enhance user's trust (Gefen et al., 2003; Wareham et al., 2005).

Main Features

The different features that an e-payment framework should offer are:

- *The support of different payment instruments and their underlying protocols.* It should support different kinds of payment models such as credit or debit, using different payment instrument such as credit cards, e-checks or digital cash, and different payment protocols (Bitcoin, EMV, etc). This support should be made through a generic Application Programming Interface (API).
- *The support of different kinds of devices for making the payment and storing payment information.* The payment could be made through desktop computers, smartphones, ebooks readers or automobiles. It should also support the use of different devices that could increase security, such as secure elements or secure storages.
- *The specification and exchange of the payment information associated to the purchase.* Payment information (prices, pay-

ment systems supported, etc) should be described in some format that allows its automatic processing and exchange.

- *The negotiation of the payment systems, prices and conditions of the purchase.* Customer and merchant might support different payment systems and be interested in different conditions that should be negotiated at the beginning of the payment process.
- *Digital receipts and loyalty information.* During a purchase process it is also important that vendor provides a receipt of the transaction. Additionally, some loyalty information could be sent. This information could also be used later to obtain better conditions in the next purchases such as better prices, additional services, etc.
- *Extensibility.* The framework should support the inclusion of new payment systems in a seamless, transparent and uniform way.
- *Security.* It should provide secure mechanisms to guarantee the exchange of payment information and identify the different participants in the system. It is specially relevant to guarantee non-repudiation and the secure identification of the parties.
- *Privacy.* Purchase sensitive information should be protected.

The design of an e-payment framework should be made taking the following principles into account (Sporny, 2014): broad individual empowerment, maximize competition, respectful of the law, broadly implementable and data portability.

Benefits

The main benefit of introducing a payment framework is that the development of applications and solutions based on payment will reduce its development costs and increase confidence in its security (Lacoste, 2000) (Ruiz-Martínez et al., 2012) (Ruiz-Martínez, 2015). This kind of framework can increase interoperability between the

different stakeholders and the different payment methods (Web Payments Interest Group Charter, 2016), which will suppose a broader use of payments, more competition in the payment area and will prevent that a payment vendor establishes a monopoly (Web Payments Interest Group Charter, 2016). Furthermore, it can also enhance user experience and trust (Hope-Bailie et al., 2015; Gefen et al., 2003) and facilitate that customer and merchant can use their preferred payment instrument and schemes (Ruiz-Martínez, 2015; Hope-Bailie et al., 2015).

PAYMENT FRAMEWORKS PROPOSED

This section presents the main frameworks that have been proposed so far. We present them in a chronological way to understand how they have evolved and the different features that they have been incorporated over time.

The main frameworks proposed so far are: Joint Electronic Payment Initiative (JEPI), Secure Electronic Marketplace for Europe (SEMPER), W3C Common Markup for micropayment per-fee-links, Internet Open Trading Protocol (IOTP), Meng and Zhang's proposal and Ruiz-Martínez et al.'s proposal (PayFrameworks) (Ruiz-Martínez et al., 2012). Next, we describe briefly each framework.

Joint Electronic Payment Initiative (JEPI)

JEPI was a joint proposal between W3C and CommerceNet consortium. Its aim was to automate the selection of the payment protocol most suitable for client and merchant for any given transaction. For this purpose, JEPI defined two protocols as extensions of HTTP (see Figure 1): Protocol Extension Protocol (PEP) and Universal Payment Preamble (UPP).

PEP is a generic negotiation protocol that allows that a user agent and a Web server can know the extension modules they support each

other and negotiate the different parameters that will be used in these modules. In JEPI, PEP is the cornerstone for UPP.

UPP (UPP,1997) is the core of JEPI and consists of a specific extension module for negotiating payment information used in a transaction: payment instrument, brands, and payment protocol. Each payment system is considered a PEP extension module and it is identified by means an URL. Furthermore, JEPI also defines a mechanism for initiating the payment system chosen.

This framework did not address different issues such as security, the negotiation payment systems (it was oriented to the negotiation of payment instruments), different payment models and operation modes and, finally, the definition of payment information for being embedded in the Web or APIs for wallets.

Secure Electronic Marketplace for Europe (SEMPER)

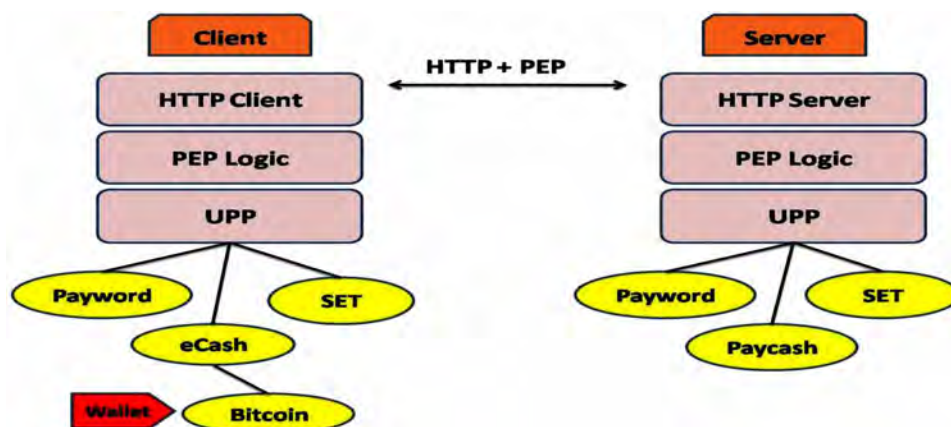
SEMPER was a project partially funded by the European Union and it was formed by a consortium around 20 companies, financial institutions, European universities and technological/research institutes. Its purpose was to work in different issues related to e-commerce in insecure networks such as legal, technical and social issues. They aimed to define an open architecture that was

independent of the network architecture, operating system and software used. This architecture supports its evolution through the inclusion of new components by means of plug-ins. Its architecture is divided in four layers: business, exchange, transfer and supporting services.

Within its objectives, one of the most important ones was the design and development of a generic payment service (see Figure 2) that provides a framework that allows business applications the support of payment in a transparent way without being to know the different particularities of the different payment methods. The main component of this service is a hierarchy of APIs for the transfer of value. This hierarchical API reflects the separation according different payment models. Thus, this hierarchy consists of a common API for all payment methods and specific extensions for each model. The different models in which this API is divided are digital money and account-based systems (see Figure 2).

In addition to these interfaces, this generic service supports other features such as mechanisms for selecting the payment instrument for the transaction, information services, control and management (that makes possible to support the development of business applications that use the generic service), tools and the needed framework for incorporating any payment system in the

Figure 1. JEPI framework



generic payment service. Each specific payment system implements some of these APIs.

The main strength of this proposal is the definition of its wallet from the generalization of the different kind of payment systems. This framework is more oriented to the development of desktop applications than Web applications. Then, it does not define how payment information can be incorporated to the Web and it does either define how to encapsulate payment messages and how the wallet exchange them with the server wallet.

W3C Common Markup for Micropayment Per-Fee-Links

The W3C Micropayment Working Group proposed the Common Markup for micropayment per-fee-links (CMMPFL). The goal of this specification was to define an extensible mechanism that allowed embedding, in a Web page, all the needed information to initialize a micropayment system (amounts and currencies, payment systems, etc) when a payment is required to access a Web content. Thus, it aimed at different micropayment wallets could co-exist in an interoperable way.

This specification, to the best of our knowledge, is the first specification in introducing the concept of payment link or per-fee-link. The goal of these links is to make easy the purchase process of an

electronic product (Web page, image, etc) as much as possible: as simple as clicking on a Web link.

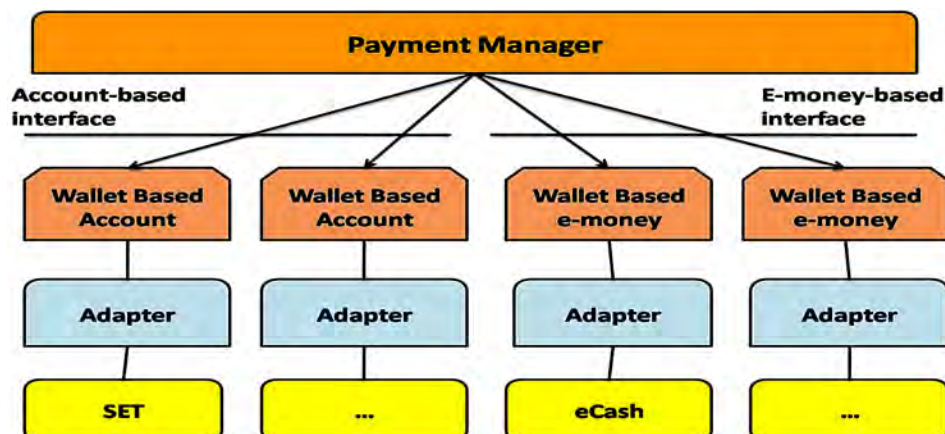
Thus, this specification was centered on how merchant sent, in the HTML web page, payment information to the client through the Web browser. This information could be included in the Web page by means of Java Applets or ActiveX components or by means of RDF. For the management of this information, it was proposed the use of a module called per fee link handler, which invokes to the different wallets that implement the specific payment protocols. However, this project did not define a Web API for invoking the wallet to perform payment process.

Internet Open Trading Protocol (IOTP)

IOTP was submitted to the Internet Engineering Task Force (IETF) (Burdett, 2000) with the objective of becoming it in a standard and interoperable protocol to make business on the Internet. Namely, its main goal was to provide a payment system independent protocol to make B2C transactions and that encapsulates any payment protocol.

The main IOTP feature to point out is that the payment process was designed so that it was independent of the payment system and defined how the messages of the e-payment system are encapsulated in IOTP. Associated to the protocol,

Figure 2. SEMPER payment framework



an API that considers different types of transaction (purchase, refund, value exchange, authentication, withdrawal, deposit, inquiry and ping) was defined.

In the IOTP v1.0 some issues to be improved were detected and were specified as a set of requirements for the version 2.0 of the protocol. The issues found were: the offer request is not specified in the protocol, it does not support the negotiation of prices, it does not use the XML Digital Signature (XMLDsig) scheme for the support of e-signature in IOTP. However, IOTP 2.0 version was not developed.

PayFrameworks

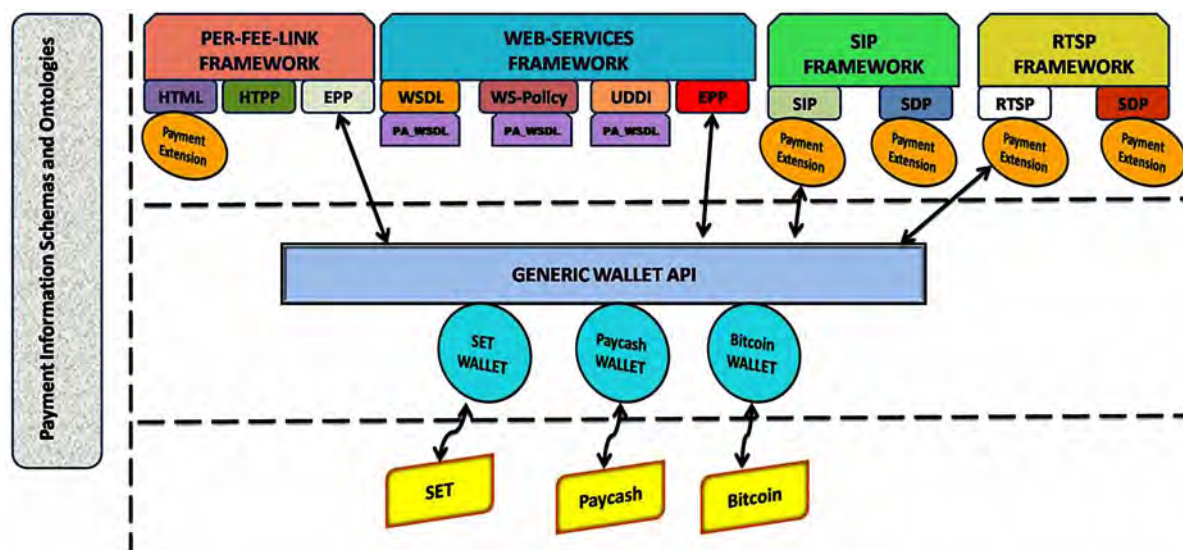
Payment frameworks for the purchase of electronic products and services (PayFrameworks) (Ruiz-Martínez et al., 2012) defines a new comprehensive framework (see Figure 3) that aims support the purchase of the main kind of electronic products and services (e-products hereafter). Namely, it considers Web products, Web services, Session Initiation Protocol (SIP) sessions and streaming sessions based on the use of the Real Time Streaming Protocol (RTSP).

For the purchase of e-products accessed by means of HTTP, the per-fee-link framework and the web services framework have been defined (see Figure 3). Both use the Extended Payment Protocol (EPP) (Ruiz et al., 2009), an oriented-session payment protocol, for the exchange of payment information and payment protocol messages. Furthermore, it defines extensions to HTML, WSDL, WS-Policy and UDDI to include payment information.

For SIP and RTSP a set of extensions to these protocols has been proposed (Ruiz-Martínez et al., 2008; Ruiz-Martínez et al., 2016). These extensions are used to incorporate payment information to these protocols.

Independently of the kind of protocol used, all extensions made to specify payment information are based on the use of the same ontology and the same set of XML schemas (Payment Information Schemas and Ontologies layer in Figure 3). Furthermore, to facilitate that the implementation of an e-payment system can be used in the different frameworks, all the frameworks use a generic API that have been defined to support any e-payment system (see in Figure 3 how the different protocols connect with the Generic Wallet API layer).

Figure 3. PayFrameworks



DISCUSSION

After analyzing the main frameworks proposed, we can find the evolution they have experimented. The first framework (JEPI) was only centered on exchanging payment information through HTTP. Then, other elements have been considered such as embedding payment information in the Web (CMMPFL or PayFrameworks) by means of its description in some language that allows its automatic processing (CMMPFL, IOTP, PayFrameworks), even some semantic information has been defined to be able to perform reasoning (PayFrameworks). In these frameworks is fundamental also the definition of generic APIs that allow the support of different payment systems, protocols and instruments (SEMPER, IOTP, PayFrameworks).

However, these frameworks have not been standardized and are not being currently used. The main reasons we can found is that, when these frameworks were defined, the use of e-payment systems on the Internet was not extended, some technologies were not mature enough or its use is not extended (semantics, security) and that in the definition of these frameworks all stakeholders did not participate, which make difficult that these frameworks were broadly extended (Jaffe & Boyera, 2015; Ruiz-Martínez, 2015).

CURRENT E-PAYMENT FRAMEWORK INITIATIVES

Nowadays, the use of e-payments is quite extended and mobile payments are growing and are becoming an important mean to make payments on the Internet. Indeed, Garnet predicts by 2017, that in the US, 50% of US Digital Commerce Revenue will be driven by the customer's engagement behaviour (Garnet, Inc, 2015). However, in e-payments there are several challenges that have arisen such as an important shopping cart abandonment, high level fraud in credit card payments, and an important number of payment solutions available but that

cannot be offered on all merchants' sites or on all devices (W3C, 2014). In order to cope with these problems and make easy payments on the Web, the W3C launches the Web payments Activity. In the next section we present the work W3C Web Payment Activity is making through Web Payments Interest Group, Web Payments Working Group and Web Payments Community Group.

W3C's Web Payment Activity

The W3C has launched the Web Payments Activity with the aim of making payments easy and more secure on the Web, improve interoperability and foster competition and innovation in web payments. Namely, the goals this activity pursues are (W3C, 2015):

1. Support the choice of user's preferred payment solution independently of the device used.
2. Make easy merchants can support a number of payment solutions in a transparent way.
3. Easier application monetization for developers.
4. Less entry barriers for new payment solution providers.
5. Decreasing online fraud.
6. Supporting in an effective way micropayments and paywalls.

This activity is being developed by the Web Payments Interest Group, Web Payments Working Group and Web Payments Community Group. It started at the end of 2014 and it is expected to be finished by the end of 2017. Its roadmap can be found in (Sporny et al., 2015).

Conceptually, the first version of the payment framework proposed in this specification contains the layers that are shown in Figure 4: transport layer, payment instruments negotiation layer and payment instrument layers. There are also other supporting modules/layers for the layers we have just mentioned such as Data/User privacy, protocols and flows information, and data format and API.

The goal of the *transport layer* is to ensure convergence among payment solutions covering the possibility of making transactions online through HTTP or offline through protocols based on proximity as NFC, BTLE, etc.

The *payment instruments negotiation layer* will support that client and merchant can agree the payment instruments to be used.

The *payment instrument layer* is responsible for working with a specific payment instrument layer. They have considered different kind of payment instruments: token-based payments, push-based payments, credit card-based payments, cryptocurrencies and coupons/loyalty. This layer should also support new kinds of payment instruments that could appear in the future. This layer, apart from managing payment instruments, needs to manage other issues related to the payment process such as the management of credential, the authentication between parties, the access to secure hardware where different information could store and manage in a more secure way (information of the payment instrument, credentials or information to perform authentication such as keys and certificates), and the generation of receipts associated to the transaction.

Additionally, to these layers, the frameworks defined two layers that are used as building blocks: *data/user privacy module* that is in charge of managing the information exchanged between the parties in order to preserve user privacy and,

the *protocols & flows, information, data format & API* module that is responsible for managing the protocols and the flow of information, the information that needs to be exchange between the parties and the API for its use.

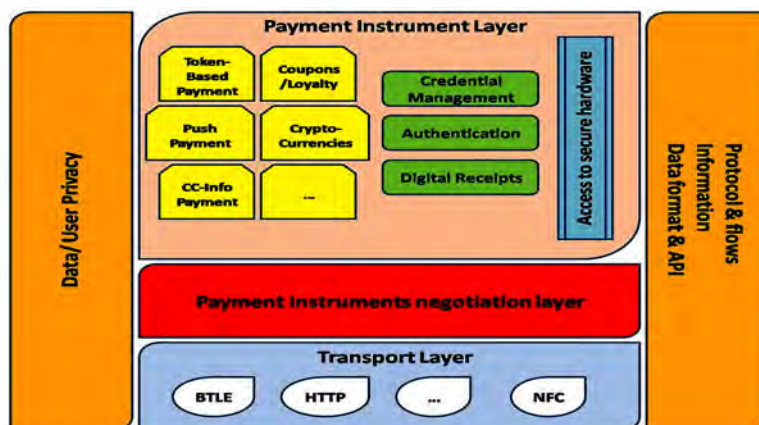
In these layers we can point out the presence of wireless transport protocols (for mobile payment systems), the use of secure hardware to improve payment security and the management and preservation of user personal information.

This initiative considers all the elements introduced in previous works regarding payment frameworks. Furthermore, some technologies and its use such as security and semantic information technologies are more mature. Therefore, the authors consider that if the work made in this initiative is supported by the different stakeholders in a few year, this framework could be standardized and the use and interoperability of (mobile) e-payment systems could be a success.

FUTURE RESEARCH DIRECTIONS

Many of the future research directions coincide with W3C's work and objectives since it aims to address the different issues that a comprehensive Web payment framework should satisfy and in its definition main stakeholders are involved. Next, we present the different future research directions.

Figure 4. W3C draft payment framework



First, a Payment API that supports different kind of payment methods and that payments can be made using different types of devices should be defined. Thus, it is necessary to research to provide an API that can be used both desktop computers and mobile devices in a similar way through different websites, which is fundamental for reducing card abandonment. The case of payment in mobile devices is specially challenging due to the limitations of these devices compared to a desktop computer. It is also important that the specification is agreed by main stakeholders so that, finally, is developed and used in web browsers.

Second, related to the previous issue, it is required to investigate which is the best approach (plug-in, extensions, independent application, etc) to incorporate payment software that implements the mentioned API into web browsers to support different payment protocols and instruments. This process should be made in a simple way and that allows the user establish payment preferences in an easy way. Therefore, it is also necessary to investigate in the different issues regarding usability and trust in the use of this software.

As third future research direction we should consider the incorporation of secure hardware devices to increase payment security. This issue is especially important in mobile devices since it requires the definition of a layer/middleware that is independent of the operating system.

Four, we should investigate in how in the definition of new vocabularies W3C is defining it is possible reuse existing terms and relationships that are already defined in some quite extended vocabularies such as GoodRelations (Ashraf et al., 2011). In the event these the terms cannot be reused and combined into a single vocabulary, we should investigate in how to make a mapping between terms of the different vocabularies.

As final future research direction we have identified to guarantee security and privacy in the different exchanges. Thus, we should investigate how the framework, on the one hand, is able to protect that exchanged information cannot be modified or observed, and on the other hand,

is able to provide enough evidences to perform correctly non-repudiation processes. During transactions is also fundamental the definition of mechanisms so that the different parties can be identified and that we can establish some level of trust that can be used to decide whether we make a payment with an entity that we have not made purchases previously.

CONCLUSION

Currently, there are different (mobile) e-payment systems to perform a purchase on the Internet. In this situation where multiple options are available, e-payment frameworks arise in order to facilitate the discovery of payment information, payment systems and the negotiation and selection of the payment system to perform a purchase.

This paper has characterized payment frameworks and introduced the main electronic payment frameworks that have been defined so far as well as the most recent activity conducted by the W3C through its Web payment initiative. This analysis shows that electronic payment frameworks have gained in complexity and now more elements are needed in order to facilitate the discovery of payment information and perform the payment process. However, technology now is more mature than before and all the stakeholders are interested in the support of e-payments. Therefore, now we are in the conditions so that the development of the payment framework is a success. This will require overcoming some challenges regarding the definition of generic APIs, the registration and use of wallets, and interoperability, security, trust and privacy issues.

REFERENCES

Ashraf, J., Cyganiak, R., O'Riain, S., & Hadzic, M. (2011). *Open ebusiness ontology usage: investigating community implementation of goodrelations*. Academic Press.

- Burdett, D. (2000). *Internet Open Trading Protocol - IOTP Version 1.0*. IETF. doi:10.17487/rfc2801
- Chung, E., & Dardailler, D. (1997). *White Paper: Joint Electronic Payment Initiative* (W3C note). W3C.
- Davidson, K., & Kawatsura, Y. (2000). *Digital Signatures for the v1.0 Internet Open Trading Protocol (IOTP)*. IETF.
- Dulai, T., Jaskó, S., & Tarnay, K. (2013). IOTP and Payment Protocols. In *Research and Development in E-Business through Service-Oriented Solutions* (pp. 20–45). IGI Global. doi:10.4018/978-1-4666-4181-5.ch002
- Eastlake, D., 3rd & Smith, C. (2000). *Internet Open Trading Protocol (IOTP) HTTP Supplement*. IETF.
- Gartner, Inc. (2015). Gartner Says By 2017, U.S. Customers' Mobile Engagement Behavior Will Drive Mobile Commerce Revenue to 50 percent of U.S. *Digital Commerce Revenue*.
- Gefen, D., Karahanna, E., & Straub, D. W. (2003). Inexperience and experience with online stores: The importance of TAM and trust. *Engineering Management. IEEE Transactions on*, 50(3), 307–321.
- Hileman, G. (2014). From Bitcoin to the Brixton Pound: History and Prospects for Alternative Currencies (Poster Abstract). In R. Böhme, M. Brenner, T. Moore, & M. Smith (Eds.), *Financial Cryptography and Data Security* (pp. 163–165). Springer Berlin Heidelberg. doi:10.1007/978-3-662-44774-1_13
- Hope-Bailie, A., Jacobs, I., Sporny, M., & Adler, P. (2015). *A Vision for a Web Payments Architecture* (W3C Editor's Draft). W3C. Retrieved from <https://www.w3.org/Payments/IG/Vision>
- Jaffe, J., & Boyera, S. (2015, January). Now is the time for web payment standards. *The Banker*, 118–119.
- Kawatsura, Y., Hiroya, M., & Beykirch, H. (2004). *Payment Application Programmers Interface (API) for v1.0 Internet Open Trading Protocol (IOTP)*. IETF.
- Lacoste, G., Pfitzmann, B., Steiner, M., & Waidner, M. (2000). *SEMPER - Secure Electronic Marketplace for Europe* (Vol. 1854). Berlin: Springer Berlin Heidelberg. doi:10.1007/b75215
- Lacoste, G., Pfitzmann, B., Steiner, M., & Waidner, M. (2000). The Payment Framework. In *SEMPER - Secure Electronic Marketplace for Europe* (pp. 185–211). Springer Berlin Heidelberg.
- Maes, P., Guttman, R. H., & Moukas, A. G. (1999). Agents that buy and sell. *Communications of the ACM*, 42(3), 81–91, ff. doi:10.1145/295685.295716
- Meng, B., & Zhang, H. (2005). An Electronic Commerce System Prototype and Its Implementations. In *Proceedings of the The Fifth International Conference on Computer and Information Technology* (pp. 966–970). IEEE Computer Society. doi:10.1109/CIT.2005.71
- Ruiz-Martínez, A. (2015). Towards a web payment framework: State-of-the-art and challenges. *Electronic Commerce Research and Applications*, 14(5), 345–350. doi:10.1016/j.elerap.2015.08.003
- Ruiz-Martínez, A., Canovas, O., & Gomez-Skarmeta, A. F. (2009). Design and implementation of a generic per-fee-link framework. *Internet Research*, 19(3), 293–312. doi:10.1108/10662240910965360
- Ruiz-Martínez, A., Cánovas Reverte, Ó., & Gómez-Skarmeta, A. F. (2012). Payment frameworks for the purchase of electronic products and services. *Computer Standards & Interfaces*, 34(1), 80–92. doi:10.1016/j.csi.2011.05.007

Ruiz-Martínez, A., Sánchez-Laguna, J. A., & Gómez-Skarmeta, A. F. (2008). Towards a Generic Payment Framework for Internet Media-on-Demand Services Based on RTSP. In *22nd International Conference on Advanced Information Networking and Applications, 2008. AINA 2008* (pp. 826–833). doi:10.1109/AINA.2008.54

Ruiz-Martínez, A., Sánchez-Laguna, J. A., & Skarmeta, A. F. (2016). Extending SIP to support payments in a generic way. *Computer Standards & Interfaces*, 46, 23–36. doi:10.1016/j.csi.2016.01.002

Schiedung, J. (2004). *Analysing and modelling of IOTP business transactions using coloured Petri Nets and BPEL*. Academic Press.

Schunter, M., Waidner, M., & Whinnett, D. (1999). The SEMPER Framework for Secure Electronic Commerce. In D. M. Nüttgens & P. D. D. h c A.-W. Scheer (Eds.), *Electronic Business Engineering* (pp. 169–185). Physica-Verlag HD. doi:10.1007/978-3-642-58663-7_11

Sporny, M. (2014). *Web Payments Design Principles* (W3C Community Group Draft Report). W3C. Retrieved from <https://web-payments.org/specs/source/design-principles/>

Sporny, M., Jacobs, I., Hope-Bailie, A., & Adler, P. (2015). *Web Payments Roadmap 1.0* (W3C Editor's Draft). W3C. Retrieved from <https://www.w3.org/Payments/IG/Roadmap/>

Vigna, P., & Casey, M. J. (2015). *The Age of Cryptocurrency: How Bitcoin and Digital Money Are Challenging the Global Economic Order*. New York: St. Martin's Press.

W3C. (2014). *Web Payments Activity*. Retrieved March 21, 2016, from <https://www.w3.org/Payments/>

Wareham, J., Zheng, J. G., & Straub, D. (2005, January 25). Critical themes in electronic commerce research: A meta-analysis. *Journal of Information Technology*, 20(1), 1–19. doi:10.1057/palgrave.jit.2000034

Web Payments Interest Group Charter. (2016). Retrieved March 22, 2016, from https://www.w3.org/2014/04/payments/webpayments_charter.html

ADDITIONAL READING

Ateniese, G., Faonio, A., Magri, B., & de Medeiros, B. (2014). Certified Bitcoins. In I. Boureanu, P. Owesarski, & S. Vaudenay (Eds.), *Applied Cryptography and Network Security* (pp. 80–96). Springer International Publishing.

Huang, Z., & Benyoucef, M. (2013). From e-commerce to social commerce: A close look at design features. *Electronic Commerce Research and Applications*, 12(4), 246–259. doi:10.1016/j.elerap.2012.12.003

Turban, E., King, D., Lee, J. K., Liang, T.-P., & Turban, D. C. (2015). *Electronic Commerce*. Cham: Springer International Publishing. doi:10.1007/978-3-319-10091-3

Turban, E., King, D., Lee, J. K., Liang, T.-P., & Turban, D. C. (2015). Electronic Commerce Payment Systems. In *Electronic Commerce* (pp. 521–559). Springer International Publishing. doi:10.1007/978-3-319-10091-3_11

Vandezande, N. (2013). *Mobile Wallets and Virtual Alternative Currencies Under the EU Legal Framework on Electronic Payments (SSRN Scholarly Paper No. ID 2325410)*. Rochester, NY: Social Science Research Network.

Vigna, P., & Casey, M. J. (2015). *The Age of Cryptocurrency: How Bitcoin and Digital Money Are Challenging the Global Economic Order*. New York: St. Martin's Press.

KEY TERMS AND DEFINITIONS

Electronic Payment Framework: A set of extensible elements and/or services, generally,

organized within a set of layers and that is designed to facilitate the use of different payment systems in a unified manner during a payment process.

Electronic Payment Protocol: A protocol to make electronic payments.

Electronic Payment System: A system that establishes how to pay for goods or services electronically during the execution of a financial transaction.

Micropayment: A payment where the amount to be paid is very small.

Paywall: A system that only gives access to web contents to those users that have a paid subscription.

User Payment Agent: It is being used as a synonym of wallet.

Wallet: A software or service that supports one or more payment systems.

Factors Determining E-Shopping Compliance by Nigerians

E**Adamkolo Mohammed Mohammed Ibrahim***University of Maiduguri, Nigeria***Md. Salleh Hj. Hassan***Universiti Putra Malaysia, Malaysia***Sarina Yusuf***Universiti Putra Malaysia, Malaysia*

INTRODUCTION

Online shopping (e-shopping) is increasingly gaining acceptance in many societies despite the myriad of circumstances and factors that influence its acceptance (Venkatesh, Thong & Xu, 2012). These factors often border around consumers' cognitive perceptions toward service quality and technology adoption expectancies (Chen & Macredie, 2010). The literature on e-commerce in Nigeria focuses on the advantages and constraints of online shopping, with little attention to the factors that influence e-shopping compliance. This chapter believes that a clear understanding of the effects of the factors would provide e-tailers with the opportunity to formulate tailored strategies to making e-shopping accepted by consumers.

The primary concept of this study is adoption of e-shopping. E-shopping refers to the act of purchasing products or services over the Internet (Business Directory, 2016; Lims & Dubinsky, 2004). Liao and Cheung (2000) described e-shopping as "virtual shopping over the Internet" (p. 3). According to Alam and Noor (2009), the adoption of technology, innovation or information and communication technology (ICT) is the means to enable businesses to compete on a global scale, with improved efficiency, and closer customer and supplier relationships. Therefore, acceptance of e-shopping by Nigerians is regarded as a crucial

condition for consumers and e-tailers to overcome challenges associated with conventional shopping (Tarute & Gatautis, 2014).

This study was performed based on the extended version of the unified theory of acceptance and use of technology (UTAUT2) model perspective. A new conceptual model was designed, incorporating constructs from the service quality (SERVQUAL) model (see Parasuraman, Zeithaml & Berry, 2002). AlAwadhi & Morris, (2008) and Venkatesh, Morris & Ackerman (2000) support such modifications. The urge to understand the factors affecting online shopping from both cognitive (UTAUT2) and SERVQUAL perspectives prompted this hybridization. This chapter was structured into literature review, methodology, results and discussion and conclusion sections.

BACKGROUND

Online Shopping in Nigeria

Online shopping is gradually becoming trendy, especially among the elites, middle-income earners, students and technocrats in Nigeria (Aminu, 2013). About a decade earlier, however, online shopping was much less heard about (Chang & Samuel, 2004) much less of adoption. Most e-tailers in the country offer a wide range of as-

sorted products and services online. However, most Nigerian consumers are slow at complying (accepting) online shopping (FOTN, 2015).

The literature suggests that the low level of online shopping acceptance among Nigerians might have links with cognitive, service quality and other social and economic factors (Amina, 2013; FOTN, 2015). These factors are believed to affect the consumers' behavior to purchase products online (Chukwu & Uzoma, 2014). Some of popular e-tailers include Konga, Jumia, Glamor and Manna Stores (Aminu, 2013).

LITERATURE REVIEW

The Concept of Consumer Acceptance of Technology

This study adopted the extended version of the unified theory of acceptance and use of technology (UTAUT) model developed by Venkatesh, Morris, Davis & Davis (2003). Venkatesh, Thong & Xu (2012), formulated the extended UTAUT (i.e., UTAUT2) model. The model focuses on consumer adoption of technology. It was developed based on a comprehensive synthesis of past information and communication technology (ICT) acceptance research. The UTAUT model has four key variables namely, performance expectancy (PE), effort expectancy (EE), social influence SI, and facilitating conditions (FC). Those constructs are theorised to influence behavioral intention to use ICT or ICT use. Those four constructs are retained in the UTAUT2 model's consumer technology acceptance and use context.

The literature defines performance expectancy as the degree to which using a particular technology provides benefits to consumers in performing certain activities (online). Effort expectancy is defined as the degree of ease linked with consumers' use of technology. Social influence is defined as the degree to which consumers perceive that important others or people that matter to them (e.g., family, friends and colleagues) believe that they

should use a particular technology. The literature defines facilitating conditions as the consumers' perceptions of the resources and support available to guarantee smooth use of a technology (AlAwadhi & Morris, 2008).

Performance expectancy, effort expectancy and social influence are theorised to influence use behavior. In the original UTAUT2 model, individual difference variables namely, age, gender and experience are theorised to moderate various relationships. However, the moderation apparatus has been jettisoned in this chapter because this study is preliminary. Venkatesh, et al. (2012) incorporated three additional constructs, namely hedonic motivation (HM), price value (PV) and habit (HT).

Prior studies have discovered that many consumers use a technology chiefly for the fun and joy they derive from it (Brown & Venkatesh, 2005; van der Heijden, 2004). Hence, the researchers incorporated hedonic motivation construct into the model. Hedonic motivation has been found to influence ICT use in online purchase context (Chan, Gong, Xu & Thong, 2008). The researchers define hedonic motivation "as the fun or pleasure derived from using a technology" (p. 161). Brown and Venkatesh (2005) and Thong, et al (2006) explained that construct plays an important role in determining technology acceptance and use.

Previous studies have established that consumers usually bear the monetary cost of online transaction (Thong, et al., 2006) rather than employees. Chan, et al. (2008) suggests that the cost and pricing structure influence consumers' technology use. For instance, there is evidence that the popularity of instant messaging (IM) services of WhatsApp social media application in Nigeria is due to the low (virtually free) cost relative to other types of mobile Internet applications such as short messaging services (SMS) (FOTN, 215). In UTAUT2, price value is conceptualized together with the quality of products or services to determine the perceived value of products or services (Zeithaml, 1988; Venkatesh, et al., 2012). Cited in Venkatesh, et al. (2012), Dodds, et al.

(1999) defines price value as “consumers’ cognitive trade-off between the perceived benefits of the applications and the monetary cost for using them” (161). Venkatesh, et al. (2012) posits that the price value of using a technology is positive when the benefits are perceived to be greater than the monetary cost.

Habit has been introduced into the model because prior research found that some consumers tend to use a technology with automaticity (e.g., Kim & Malhotra, 2005; Venkatesh, et al., 2012). Habit is defined as the extent to which consumers tend to use a technology to perform a behavior (shopping) automatically because of learning (prior experience) (Limayem, et al., 2007). Kim and Malhotra (2005) and Limayem, et al. (2007) found that habit influences technology use.

Findings from a cross-cultural study of ICT adoption performed by Oshlyansky, Cairns and Thimbleby (2007) reveal that performance expectancy, effort expectancy and social influence predict use intention. Šumak, Polančič and Heričko (2010) found social influence has a significant impact on students’ behavioral intention to use moodle; and, behavioral intention is a powerful predictor of the use of ICT (Baptista & Oliveira, 2015). In a research conducted by Cheng, Liu, Song and Qian (2008), the validity of the UTAUT model was investigated using 313 intended users of Internet banking in China. The result identified performance expectancy and social influence as strong predictors of behavioral intention (Taiwo & Downe, 2013).

The researcher performed some critical adjustments as is normally accepted in literature (e.g., AlAwadhi & Morris 2008) in the model by incorporating three constructs, namely reliability (RL), empathy (EM) and responsiveness (RS) from the service quality (SERVQUAL) model (see Parasuraman, et al., 2002). This chapter deems those adjustments as necessary because the literature suggests that service quality factors significantly determine online shopping acceptance (e.g., Venkatesh, et al., 2000).

The Concept of Service Quality

As a marketing strategy, service quality (SERVQUAL) is a concept that has aroused considerable interest in the research literature because of the difficulties in defining and measuring it (Parasuraman, et al., 2002). Service quality has been defined severally. A commonly used definition is the one that defines service quality as the extent to which a service meets customers’ needs or expectations (Lewis & Mitchell, 1990). Service quality is also the difference between customer expectations of service and perceived service (van Iwaarden, van der Wiele, Ball & Millen, 2003). The literature underscores that if expectations are greater than performance, then perceived quality is less than satisfactory; hence, customer dissatisfaction occurs (Parasuraman, et al., 2002). Measuring service quality allows for comparison between pre- and post-service changes, for the location of quality-related problems and for the establishment of clear standards for service delivery (Lewis & Mitchell, 1990).

The SERVQUAL instrument has been adjudged the predominant method used to measure consumers’ perceptions of service quality. The modified version of the model, which was adopted by this study, has five generic dimensions or factors: tangibles, reliability, responsiveness, assurance and empathy (van Iwaarden, et al., 2003). The reviewed model has five key constructs (van Iwaarden, et al., 2003; Parasuraman, et al., 2002) as follows:

1. **Tangibles:** Physical facilities, equipment and appearance of personnel;
2. **Reliability:** Ability to perform the promised service dependably and accurately;
3. **Responsiveness:** Willingness to help customers and provide prompt service;
4. **Assurance:** Knowledge and courtesy of employees and their ability to inspire trust and confidence, competence, courtesy, credibility and security; and

5. **Empathy:** Caring, individualised attention that the organization provides to its customers, as well as access, communication, and understanding the customer.

Hypotheses of This Study

- H₁:** There was significant relationship between performance expectancy and e-shopping acceptance
- H₂:** There was significant relationship between effort expectancy and e-shopping acceptance
- H₃:** There was significant relationship between social influence and e-shopping acceptance
- H₄:** There was significant relationship between facilitating conditions and e-shopping acceptance
- H₅:** There was significant relationship between hedonic motivation and e-shopping acceptance
- H₆:** There was significant relationship between price value and e-shopping acceptance
- H₇:** There was significant relationship between habit and e-shopping acceptance
- H₈:** There was significant relationship between reliability expectancy and e-shopping acceptance
- H₉:** There was significant relationship between empathy and e-shopping acceptance
- H₁₀:** There was significant relationship between responsiveness expectancy and e-shopping acceptance

METHODOLOGY

The research instrument consisted of a 40-item questionnaire that was pre-tested and administered face-to-face to a sample of 380 students selected from a population of 2,750 undergraduate students on campus at Yobe State University Damaturu. Since e-shopping is a novel online activity in the society, it was presumed that not all students might have experienced shopping online. Therefore, purposive sampling method was adopted. Moreover, only students that had been confirmed to have had experienced online shopping at least once were selected for participation in the study.

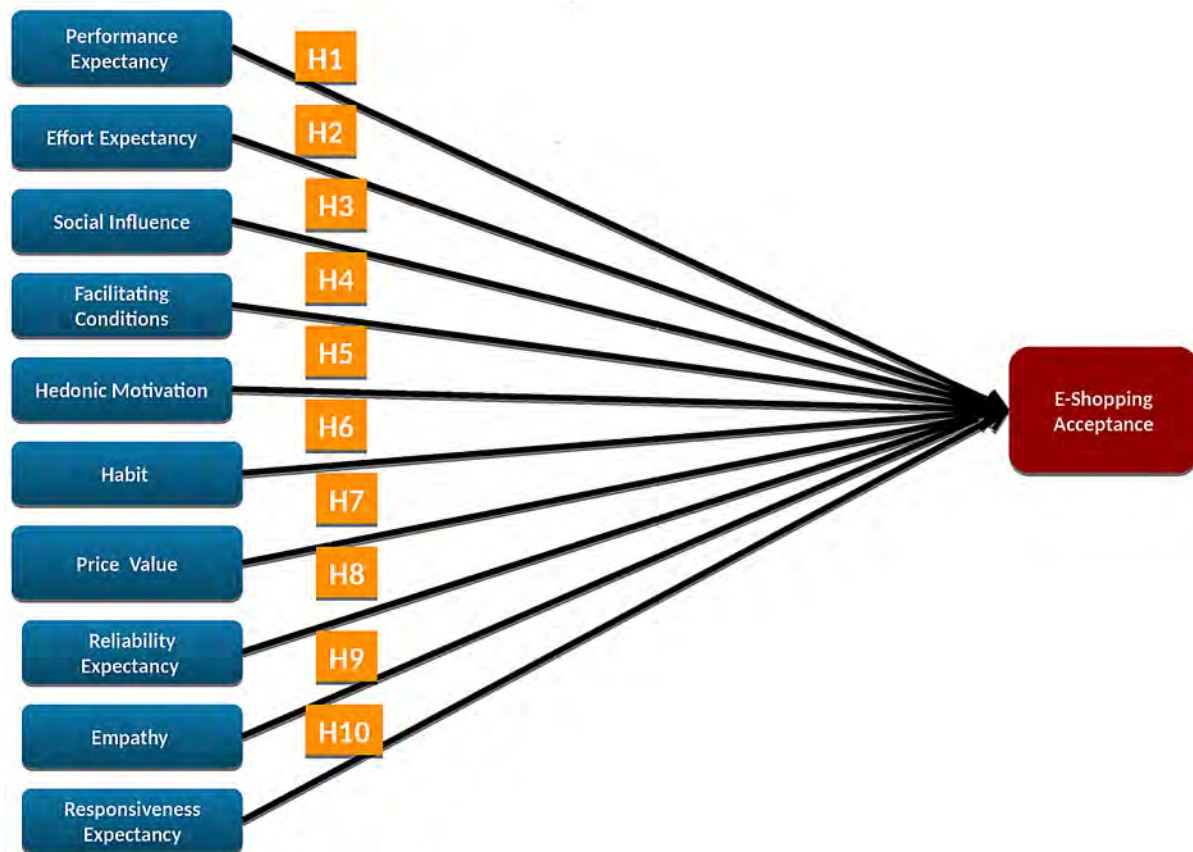
Data collection lasted for two weeks, and with 337 questionnaires retrieved, an 88.7% return rate was achieved. However, data of only 300 valid cases were analysed since 37 cases were eliminated during data sorting and cleansing due to incomplete responses and response inconsistency. The descriptive analysis was performed using the statistical package for social sciences (SPSS) software version 22.0 and five-point Likert Scale was used to gauge respondents' perceptions. Moreover, a committee of expert scholars validated the questionnaire. The degree of reliability of the items in the scale ranged from $\alpha = .755$ to $\alpha = .876$.

Furthermore, this study was performed based on the UTAUT2 model perspective. The model was modified by incorporating three additional predictors, namely reliability expectancy, empathy and responsiveness expectancy from the SERVQUAL model (see Figure 1).

Table 1. Performance expectancy as factor affecting acceptance of e-shopping

Measurement Scale			
S/No	Item	Mean	SD
1	Doing e-shopping increases my chances of achieving things that are important to me in the online purchasing process.	2.53	.836
2	I can save time in the online purchasing process when I do e-shopping.	2.43	1.04
3	I find doing e-shopping very useful in the online purchasing process.	2.32	.858
Overall Mean		2.43	.911

Figure 1. The conceptual model of this study, illustrating the 10 independent variables and the dependent variable
(adopted with adjustments from Venkatesh, et al. (2012))



RESULTS

Performance Expectancy

The moderate mean value ($M = 2.53$, $SD = .836$) of the first item in the scale indicates many of the respondents believed that conducting e-shopping would increase their chances of achieving that were important to them in the online purchasing process. Other items were less important (see Table 1).

Effort Expectancy

The item that gauged the respondents' belief that learning how to do online shopping was easy for them scored the highest mean value in the scale, even though the intrinsic value of the mean score

was moderate ($M = 2.55$, $SD = 1.06$). Other items were moderately important (see Table 2).

Social Influence

The finding indicates the level of the respondents' perception that people who were important to them could affect their acceptance of online shopping is low. Other items were less important (see Table 3).

Facilitating Conditions

The respondents' perception that access to technical help that could guarantee smooth performance of online shopping affected their acceptance of e-shopping is moderate. All other items were moderately important (see Table 4).

Table 2. Effort expectancy as factor affecting acceptance of e-shopping

Measurement Scale			
S/No	Item	Mean	SD
1	Learning how to do e-shopping is easy for me.	2.55	1.06
2	It is easy for me to become skilful at doing e-shopping.	2.51	.975
3	I find doing e-shopping easy to use.	2.50	.982
	<i>Overall Mean</i>	<i>2.52</i>	<i>1.01</i>

Table 3. Social influence as factor affecting acceptance of e-shopping

Measurement Scale			
S/No	Item	Mean	SD
1	People whose opinions I value prefer that I do e-shopping.	1.95	.831
2	People who influence my behavior think that I should do e-shopping.	1.91	.793
3	People who are important to me think that I should do e-shopping.	1.83	.733
	<i>Overall Mean</i>	<i>1.90</i>	<i>.786</i>

Table 4. Facilitating conditions as factors affecting acceptance of e-shopping

Measurement Scale			
S/No	Item	Mean	SD
1	I have the knowledge necessary to do e-shopping.	2.77	1.12
2	I feel comfortable doing e-shopping.	2.72	1.05
3	I have the resources necessary to do e-shopping.	2.24	.836
	<i>Overall Mean</i>	<i>2.58</i>	<i>1.00</i>

Hedonic Motivation

The respondents' perception whether they believed that doing e-shopping was enjoyable was high, even though the actual value of the mean score was moderate ($M = 2.60$, $SD = 1.12$). Concisely, many of the respondents believed that they felt entertained when they do e-shopping (see Table 5).

Price Value

Many of the respondents liked searching for cheap products online. This item scored the highest mean value in the scale. However, the actual mean value of the item is moderate ($M = 2.62$, $SD = 1.06$). In short, not so many of the respondents believed that

the price value of online products of the e-shops was appealing (see Table 6).

Habit

The finding shows most of the respondents did not believe that doing e-shopping had become their habit as indicated by the very low or just low mean values (see Table 7).

Reliability Expectancy

Many of the respondents believed that purchasing products online was reliable as indicated by the moderate mean values. However, not so many of the respondents believed that online stores keep to time ($M = 2.41$, $SD = .733$) (see Table 8).

Table 5. Hedonic motivation as factor affecting acceptance of e-shopping

Measurement Scale			
S/No	Item	Mean	SD
1	Doing e-shopping is enjoyable.	2.60	1.12
2	Doing e-shopping is very entertaining.	2.56	1.11
3	Doing e-shopping is fun.	2.36	1.00
Overall Mean		2.51	1.08

Table 6. Price value as factor affecting acceptance of e-shopping

Measurement Scale			
S/No	Item	Mean	SD
1	I like to search for cheap product deals online.	2.62	.106
2	I can save money by examining the prices of products for e-shopping online.	2.08	.912
3	Online shopping offers better value for my money.	2.07	.799
Overall Mean		2.26	.601

Table 7. Habit as factor affecting acceptance of e-shopping

Measurement Scale			
S/No	Item	Mean	SD
1	Doing e-shopping has become a habit for me.	1.90	.787
2	I am addicted to doing e-shopping.	1.58	.496
3	I must do e-shopping.	1.48	.501
Overall Mean		1.65	.595

Table 8. Reliability expectancy as factor affecting acceptance of e-shopping

Measurement Scale			
S/No	Item	Mean	SD
1	You get what you ordered from online stores.	2.90	1.07
2	Online stores are dependable because the price shown on the website is the actual amount billed.	2.69	1.11
3	When online stores promise to do something by a certain time, they do so.	2.41	.733
Overall Mean		2.67	.971

Empathy

Many of the respondents regarded the service quality of the e-tailers as empathetic as indicated by the moderate the mean values. However, not so

many of the respondents believed that the service quality of the e-tailers was any empathetic ($M = 2.37$, $SD = .995$) (see Table 9).

Table 9. Empathy as factor affecting acceptance of e-shopping

Measurement Scale			
S/No	Item	Mean	SD
1	Employees of online stores are polite to customers.	2.94	1.01
2	Employees of online stores provide customers individual attention.	2.72	.971
3	When customers have problems, employees of online stores are sympathetic and reassuring to customers.	2.37	.995
Overall Mean		2.68	.992

Responsiveness Expectancy

Many of the respondents believed that the online stores were responsive to consumers' needs. This is indicated by the moderate mean values. However, not many of the respondents believed that the online stores were any responsive to customers' needs ($M = 2.16$, $SD = .842$) (see Table 10).

DISCUSSION

Not quite many of the respondents ($M = 2.43$, $SD = .911$) believed that doing shopping online would yield remarkable outcomes (performance expectancy) to them (refer to Table 1). However, quite a number of them ($M = 2.52$, $SD = 1.01$) believed that doing shopping online would be effort-free, and easy (refer to Table 2). Similarly, not quite many ($M = 1.90$, $SD = .786$) of the respondents believed that important others would influence their e-shopping compliance (refer to Table 3). This suggests that social relationship is not important in influencing their e-shopping compliance.

Contrary to the result of social influence construct, quite many ($M = 2.58$, $SD = 1.00$) of the respondents believed that access to technical help would influence their e-shopping compliance (refer to Table 4). This suggests that consumers prefer to shop online when they are sure of the availability of someone that would help them in case they were stuck while shopping online. Similarly, quite many ($M = 2.51$, $SD = 1.08$) of the respondents believed that they enjoyed shopping online (refer to Table 5). This suggests that consumers are interested to shop online.

Contrary to the result of hedonic motivation construct, not quite many ($M = 2.26$, $SD = .601$) of the respondents believed that the price value of products promoted online did not influence their e-shopping compliance (refer to Table 6). This suggests that consumers are not impressed by the prices of products tagged on the websites of e-stores. Similarly, not many of the respondents believed that they did e-shopping because it had become their habit (refer to Table 7). This suggests that consumers prefer to do e-shopping not necessarily because they are addicted to it.

Table 10. Responsiveness expectancy as factor affecting acceptance of e-shopping

Measurement Scale			
S/No	Item	Mean	SD
1	Employees of online stores understand customers' needs.	2.94	1.01
2	Employees of online stores respond to customer requests promptly.	2.72	.971
3	Employees of online stores have customers' best interests at heart.	2.16	.842
Overall Mean		2.61	.941

Quite many ($M = 2.67$, $SD = .971$) of the respondents believed that the services of the online stores were reliable (refer to Table 8). This suggests that consumers believe that the service delivery quality of the online stores is reliable and dependable. Similarly, quite many ($M = 2.68$, $SD = .992$) of the respondents believed that the online stores were empathetic to customers' needs (refer to Table 9). This indicates that customers believe that online stores treat customers' needs with utmost care.

Similarly, quite many ($M = 2.61$, $SD = .941$) of the respondents believed that the online stores were responsive to customers' needs by keeping their words and being prompt (refer to Table 10). This suggests that customers believe that online keep promise. However, not so many ($M = 2.28$, $SD = .825$) of the respondents believed that doing e-shopping was any favorable to them. This suggests that not so many people do online shopping. This may be associated with the relatively low level of broadband penetration in Nigeria, relatively low level of access to online applications and particularly, the novelty of online shopping in the country (Emmanuel, 2015; FOTN, 2015).

FUTURE RESEARCH DIRECTIONS

This chapter recommends that future research should focus on investigating e-shopping compliance by consumer group categories, rather than general consumers. For example, there is need to understand the e-shopping readiness (acceptance) by public (civil) servants, employees of private (corporate) organizations, urban mothers, politicians, academicians, etc. Conducting exploratory studies to understand the online shopping characteristics of multiple groups of Nigerian consumers could be of immense benefits to both e-tailers and researchers as follows, given that the online shopping market is still new in the country (see Emmanuel, 2015). E-tailers would be able to determine the specific areas to channel their business resources mostly and understand their

target audience (consumers) better. Researchers would be able to identify salient issues, extract cogent variables, adopt appropriate and relevant research methodologies and theoretical models and then design cutting-edge conceptual frameworks that could yield better results. In addition, this chapter recommends that future research should adopt this model to conduct a study in different contexts, e.g., in the United Kingdom (UK), Ghana, Malaysia, the United States (US), etc using a qualitative perspective.

It may be necessary to extend the critical investigation into the exploitation of social networking sites by online retail stores to enhance cost-effective access online retail services. This chapter argues that social media platforms provide one of the most cost-effective promotion services to the online stores. In addition, publicizing and promoting products and services via social media ensures virtually the widest market coverage in the country (see FOTN, 2015). Finally, it may be necessary to direct future research attention on the correlation between consumer demographic characteristics and service quality.

CONCLUSION

Most of the respondents perceived e-shopping positively even though their acceptance of online shopping was rather moderate. However, their level of favorable perception of e-shopping could increase if, *ceteris paribus*, factors such as access and affordability of Internet services, technology use skills, prior e-shopping experience, trustworthiness and reliability of online stores were addressed.

Furthermore, the findings indicate that three out of the 10 factors investigated, namely Empathy ($M = 2.68$, $SD = .992$), Reliability Expectancy ($M = 2.67$, $SD = .971$) and Responsiveness Expectancy ($M = 2.61$, $SD = .941$) have been discovered to influence online shopping acceptance critically. Whereas, the influence of Facilitating Conditions ($M = 2.58$, $SD = 1.00$), Effort Expectancy ($M =$

2.52, SD = 1.01), Hedonic Motivation (M = 2.51, SD = 1.08) and Performance Expectancy (M = 2.43, SD = .911) was weak. The influence of Price Value (M = 2.26, SD = .601), Social Influence (M = 1.90, SD = .786) and Habit (M = 1.65, SD = .595) on e-shopping compliance was unimportant. These findings suggest that online stores' service quality (SQ) dimensions are the most important factors that affect online consumers' acceptance of e-shopping in Nigeria. Therefore, this chapter concludes consumers' perceptions of e-tailers' service quality influence e-shopping compliance more than cognitive perceptions do.

REFERENCES

- Alam, S. S., & Noor, M. K. M. (2009). ICT adoption in small and medium enterprises: An empirical evidence of service sectors in Malaysia. *International Journal of Business and Management*, 4(2), 112–125. doi:10.5539/ijbm.v4n2p112
- AlAwadhi, A., & Morris, A. (2008). The use of the UTAUT model in the adoption of e-government services in Kuwait. *Proceedings of the 41st Annual Hawaii International Conference on System Sciences*, (pp. 219-229). IEEE. doi:10.1109/HICSS.2008.452
- Aminu, S. A. (2013). Challenges militating against adoption of online shopping in retail industry in Nigeria. *Journal of Marketing Management*, 1(1), 23–33.
- Baptista, G., & Oliveira, T. (2015). Understanding mobile banking: The unified theory of acceptance and use of technology combined with cultural moderators. *Computers in Human Behavior*, 50, 418–430. doi:10.1016/j.chb.2015.04.024
- Brown, S. A., & Venkatesh, V. (2005). Model of adoption of technology in the household: A baseline model test and extension incorporating household life cycle. *Management Information Systems Quarterly*, 29(4), 399–426.
- Chan, K. Y., Gong, M., Xu, Y., & Thong, J. Y. L. (2008). Examining user acceptance of SMS: An empirical study in China and Hong Kong. *Proceedings of 12th Pacific Asia Conference on Information System*.
- Chang, J., & Samuel, N. (2004). Internet shopper demographics and buying behavior in Australia. *Journal of American Academy of Business*, 5(1/2), 171–176.
- Cheng, D., Liu, G., Song, Y.-F., & Qian, C. (2008). Adoption of internet banking: An integrated model. *Proceeds of the 4th International Conference on Wireless Communications, Networking and Mobile Computing*.
- Chukwu, B. I. & Uzoma, I. C. (2014). Impact of social media networks on consumer patronage in Nigeria: A study of Jumia and Konga Nigeria Limited. *European Journal of Business and Management*, 6(30), 63-70.
- Emmanuel, O. (2015). Facebook's daily users more in Nigeria than South Africa, Kenya, others. *Premium Times*. Available at <http://www.premiumtimesng.com/business/189835-facebook-daily-users-more-in-nigeria-than-south-africa-kenya-others.html>
- FOTN. (2015). *Freedom on the Net Report 2015: Nigeria*. Available at <https://freedomhouse.org/report/freedom-net/2015/nigeria>
- Heijden, H. (2004). User acceptance of hedonic information systems. *Management Information Systems Quarterly*, 28(4), 695–704.
- Kim, S. S., & Malhotra, N. K. (2005). A longitudinal model of continued IS use: An integrative view of four mechanisms underlying post-adoption phenomena. *Management Science*, 51(5), 741–755. doi:10.1287/mnsc.1040.0326
- Lewis, B. R., & Mitchell, V. W. (1990). Defining and measuring the quality of customer service. *Marketing Intelligence & Planning*, 8(6), 11–17. doi:10.1108/EUM0000000001086

- Liao, Z., & Cheung, M. T. (2000). Internet-based e-shopping and consumer attitudes: An empirical study. *Information & Management*, 38(5), 299–306. doi:10.1016/S0378-7206(00)00072-0
- Limayem, M., Hirt, S. G., & Cheung, C. M. K. (2007). How habit limits the predictive power of intentions: The case of IS continuance. *Management Information Systems Quarterly*, 31(4), 705–737.
- Lims, H., & Dubinsky, A. J. (2004). Consumers perceptions of e-shopping characteristics: An expectancy-value approach. *Journal of Services Marketing*, 18(7), 500–513. doi:10.1108/08876040410561839
- McCarty, J. A., & Shrum, L. J. (1993). The role of personal values and demographics in predicting television viewing behavior: Implications for theory and application. *Journal of Advertising*, 22(4), 77–101. doi:10.1080/00913367.1993.10673420
- Oshlyansky, L., Cairns, P., & Thimbleby, H. (2007). Validating the Unified Theory of Acceptance and Use of Technology (UTAUT) tool cross-culturally. *Proceedings of the 21st BCS HCI Group Conference*.
- Parasuraman, A., Zeithaml, V. & Berry, L. (2002). SERVQUAL: a multiple-item scale for measuring consumer perceptions of service quality. *Retailing: Critical Concepts*, 64(1), 140.
- Sumak, B., Polančič, G., & Heričko, M. (2010). An empirical study of virtual learning environment adoption using UTAUT. *Mobile Hybrid and On-Line Learning, 2010, ELM 10. Proceedings of 2010 Second International Conference on* (pp. 17-22). IEEE.
- Taiwo, A. D., & Downe, A. G. (2013). The theory of user acceptance and use of technology (UTAUT): A meta-analytic review of empirical findings. *Journal of Theoretical and Applied Information Technology*, 49(1), 1817–3195.
- Tarute, A. & Gatautis, R. (2013). ICT impact on SMEs performance. *Procedia - Social and Behavioral Sciences*, 110(2014), 1218 – 1225.
- Thong, J. Y. L., Hong, S. J., & Tam, K. Y. (2006). The effects of post-adoption beliefs on the expectation-confirmation model for information technology continuance. *International Journal of Human-Computer Studies*, 64(9), 799–810. doi:10.1016/j.ijhcs.2006.05.001
- van der Business Directory. (2016). *Marketing strategy*. Available at <http://www.businessdirectory.com/definition/marketing-strategy.html>
- Van Iwaarden, J., van der Wiele, T., Ball, L., & Millen, R. (2003). Applying SERVQUAL to web sites: An exploratory study. *International Journal of Quality & Reliability Management*, 20(8), 919–935. doi:10.1108/02656710310493634
- Venkatesh, V., Morris, M. G., & Ackerman, P. L. (2000). A longitudinal field investigation of gender difference in individual technology adoption decision-making processes. *Organizational Behavior and Human Decision Processes*, 83(1), 36–60. doi:10.1006/obhd.2000.2896 PMID:10973782
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *Management Information Systems Quarterly*, 27(3), 425–478.
- Venkatesh, V., Thong, J. Y. L., & Xu, X. (2012). Consumer acceptance and use of information technology: Extending the Unified Theory of Acceptance and Use of Technology¹. *Management Information Systems Quarterly*, 36(1), 157–178.
- Zhou, Z., Jin, X., & Fang, F. (2014). Moderating role of gender in the relationships between perceived benefits and satisfaction in social virtual world continuance. *Decision Support Systems*, 65, 69–79. doi:10.1016/j.dss.2014.05.004

ADDITIONAL READING

Ayo, C. K., Adewoye, J. O., & Oni, A. A. (2011). Business-to-consumer e-commerce in Nigeria: Prospects and challenges. *African Journal of Business Management*, 5(13), 5109–5117.

Chiejina, C., & Olamide, S. E. (2014). Investigating the significance of the ‘pay on delivery’ option in the emerging prosperity of the Nigerian e-commerce sector. *Journal of Marketing and Management*, 5(1), 120–135.

Falode, B. O., Amubode, A. A., Adegunwa, M. O., & Ogunduyile, S. R. (2016). Online and offline shopping motivation of apparel consumers in Ibadan metropolis, Nigeria. *International Journal of Marketing Studies*, 8(1), 150–160. doi:10.5539/ijms.v8n1p150

Salimon, M. G., Yusoff, R. Z., & Mokhtar, S. S. M. (2016). The influence of e-satisfaction, e-trust and hedonic motivation on the adoption of e-banking and its determinants in Nigeria: A pilot study. *Mediterranean Journal of Social Sciences*, 7(1), 54–63.

KEY TERMS AND DEFINITIONS

Acceptance: This term simply refers to an individual’s resolution to adopt a technology.

B2B: This acronym means business-to-business and it refers to online commercial activity directly between two or more business organizations (as opposed to between a firm and direct consumers).

B2C: This acronym means business-to-consumers and it refers to online commodity retail directly between a firm and consumers (as opposed to B2B).

Compliance: This term is a synonym of acceptance.

Consumers: This term refers to individuals that purchase products online for direct use (consumption).

E-Shopping: This term means electronic shopping and it refers to purchasing products and services online.

E-Tailers: This term means electronic retailers. It refers to business organizations that engage in retailing products online.

Online: This term can be defined from two major perspectives namely adjective and adverb. (1) Adjective: It refers to Internet-based applications such as website, web portal, e-data base, social networking sites and even cloud technology (such as Dropbox, Google Drive, etc.). (2) Adverb: It means performing a given task via the Internet or Internet-related platforms.

Enterprise Interoperability

E**Ejub Kajan***State University of Novi Pazar, Serbia*

INTRODUCTION

EC (Electronic Commerce) represents one of the major driving forces to build an electronic society. In the past enterprises faced with interoperability problems, first in the era of closed systems, later on in the infant era of open systems and ACME (A Company that Make Everything)-like vendor's solutions (Kajan, 2014). Since then, many interoperable frameworks have been suggested, developed and implemented; most of them are based on Web services and Semantic Web technologies. With the advent of globalization many organizations look for new partners to reach common goals to improve, for example, production rate, increase market share, refine supply chain, etc. Such new networked organizations are known as Virtual Enterprises (VEs). In a VE it is unlikely that any single partner will decide on the infrastructure, applications, and/or processes to be used. Instead, knowledge sharing around common goals and retaining the autonomy of each partner is crucial (Kajan, et. al., 2016).

In recent days enterprises are facing with a new challenge: everything is connected or it is going to be connected in the near future. In that movement social networking, networked things, smart objects, social communities of Web services, etc., are taking place as stated by Tan et al. "Currently, most social networks connect people or groups who expose similar interests or features. In the near future, we expect that such networks will connect other entities, such as software components, Web-based services, data resources, and workflows" (Tan et al., 2013).

As a consequence the new engineering discipline driven by the torrents of data available today

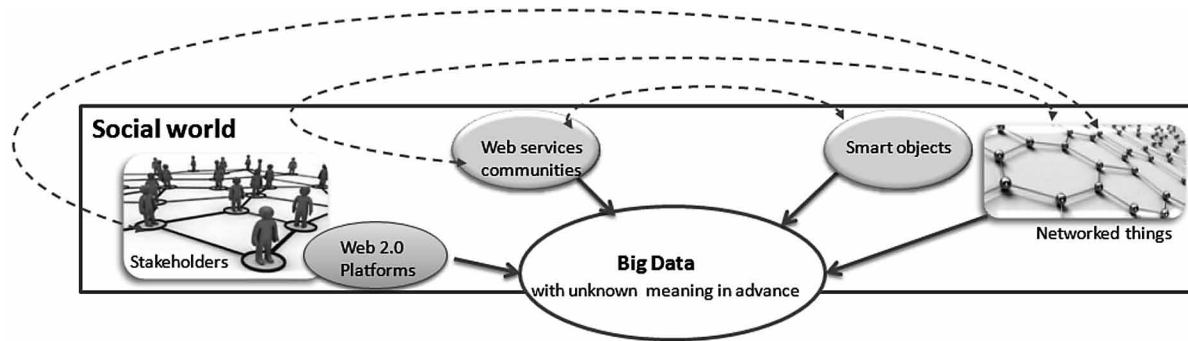
has born: the data science. Data is collected about anything, at any time and any place (van der Aalst, 2014). In such circumstances there is a lot of new heterogeneity issues between business processes, supported applications, data sources, events, and associated data on one side and different hardware, operating systems, database systems, network infrastructure, etc. on the other side, that make huge difficulties and barriers in achieving the full potential of EC. This article gives an overview of main challenges, obstacles, approaches and recent research efforts, and forecasts in order to overcome recent interoperability problems.

THE ISSUES AND CHALLENGES OF ENTERPRISE INTEROPERABILITY

The Web is the backbone of a new social era -a more open, global, ubiquitous, and pervasive platform. People, software, and things are part of a new era in which almost "everything" will be socially connected as shown in Figure 1. We refer to it as the *Social World*. The complexity of social world calls for mutual understanding of all entities involved. Besides many initiatives, ideas, and particular platform-to-platform solutions, connecting all by mutual understanding and self-learning capabilities is yet a big challenge (Dorloff & Kajan, 2012).

Interoperability is the ability of two or more systems or components to exchange information and use the information that has been exchanged in a useful way.¹ In the context of social enterprises (i.e., those that explore data from social world), interoperability is the ability of interactions (exchange of information and services) between

Figure 1. The social world

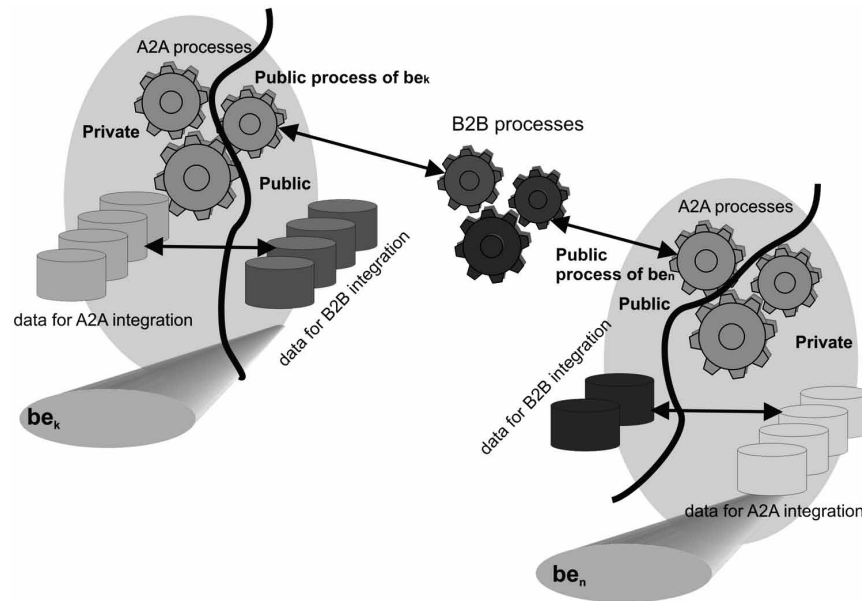


enterprise systems (Verginadis, 2011), but also to explore and exchange information of mutual interest gathered from social world. It has triggered the need of capturing the intrinsic characteristics of the business world (enterprises, their applications and stakeholders that may run various business processes) on the one side, and social and ubiquitous worlds on the other side, for bridging the gap between them. These goals have emphasized by (Romero and Vernadat, 2016) as “Enterprise Interoperability and Networking services will need to become a ‘commodity’ in order to support the building of a hyper-connected world and the seizing of its opportunities for industry (e.g. Industry 4.0)...”.

Most organizations are now modeled according to the principles of Service Oriented Architecture (SOA) for the sake of improving efficiency, agility, and response to changing market needs. SOA supports the integration of several enterprises into an entity usually known as a VE by exposing these organizations’ capabilities as services (Huhns & Singh, 2005). A VE possesses the following characteristics (Narendra, et al., 2013): (i) it is formed for a specific service-oriented process execution (for short- or long-term), and may dissolve once that execution is done; (ii) it is dependent on the nature of the interactions among the participating organizations; (iii) it is typically formed via a joint alignment of strategies among the participating organizations; and (iv) since the participating organizations are autonomous, conflicts would definitely arise.

The main goal of the VE is to provide business entities (BEs) with the ability to establish business collaborations in a way that their public business processes can interact with each other to exchange data. In a business scenario, BEs are usually loosely coupled; that means business processes require ad hoc integration from time to time. Such on-demand integration may experience many conflicts. Problems arise due to the heterogeneity between business processes and data involved, which are both different by nature on the one side and on the other side between underlying IT technologies, which are different by default. In EC, we re-characterize interoperability into A2A and B2B. They have much in common, but they are also different. Business processes inside an enterprise have their private and public parts, as shown in Figure 2. The private part is visible only inside a business entity interacting with other internal business processes whilst the public part acts inside the business entity but also takes place in B2B processes interacting with the public parts of business processes that belong to the other BE. A2A is an important mechanism for BEs in order to achieve business goals, but also serves as a flywheel of B2B efforts of that BE. As much as a BE reaches full A2A, i.e., it becomes a Zero Latency Enterprise (ZLE), i.e., its chances to have successful B2B relationships with other BEs increases (Kajan, 2010). Both A2A and B2B interoperability are victims of this heterogeneity. From that point of view, the main difference between them is the ability to control

Figure 2. The private and public parts of A2A and B2B



heterogeneity, which is much easier in a single business. In a B2B environment, where a large number of potential participants may exist, the autonomy of every BE how to run its own business is circumscribed. Thus B2B experiences face many more difficulties in interoperability than A2A.

Two Web technologies recently knocked on the doors of BEs, exemplified by two paradigms, Web 2.0, as a common name for Web of people using Internet and Web for people-to-people collaboration; and Internet of Things (IoT) that should, by nature, allow machine-to-machine communication, but also envisioning values to EC as well (Qin, et al., 2014). A visionary paper (Maamar, 2003) briefly set the target. Following that vision we set up the research roadmap for Enterprise 2.0 (Faci, et al., 2014), and made some experiments in the areas of business process design and management in social enterprises (Maamar, et al., 2013; Kajan, et al., 2014). These efforts among others gave an excellent opportunity to explore business process modeling, management, design, implementation and evaluation of VEs, and new possibilities to solve conflict resolution in business processes, as well (Maamar, et al., 2013). Ubiquitous Web gave many new opportunities, but also raised new

challenges to reach enterprise interoperability in A2A, B2B, and VE environments (Romero & Vernadat, 2016).

OBSTACLES IN ENTERPRISISE INTEROPERABILITY- PAST, PRESENT, AND FUTURE

Perhaps most readers of this article have forgotten or have never faced with the difficulties to reach any interoperable transaction in the era of closed systems. The era of open systems, first envisioned by OSI and definitively reached by Internet and Web, offers a lot of possibilities to reach enterprise interoperability on decent level. The most recognized and investigated obstacles for enterprise interoperability are semantics conflicts that may grow any time by any way, especially having in mind business relations on the fly (Kajan & Stoimenov, 2005). Conflicts may be of different types and at all levels of business collaboration. They may be caused either by data or business processes themselves, or business rules or business protocols, very often by all together.

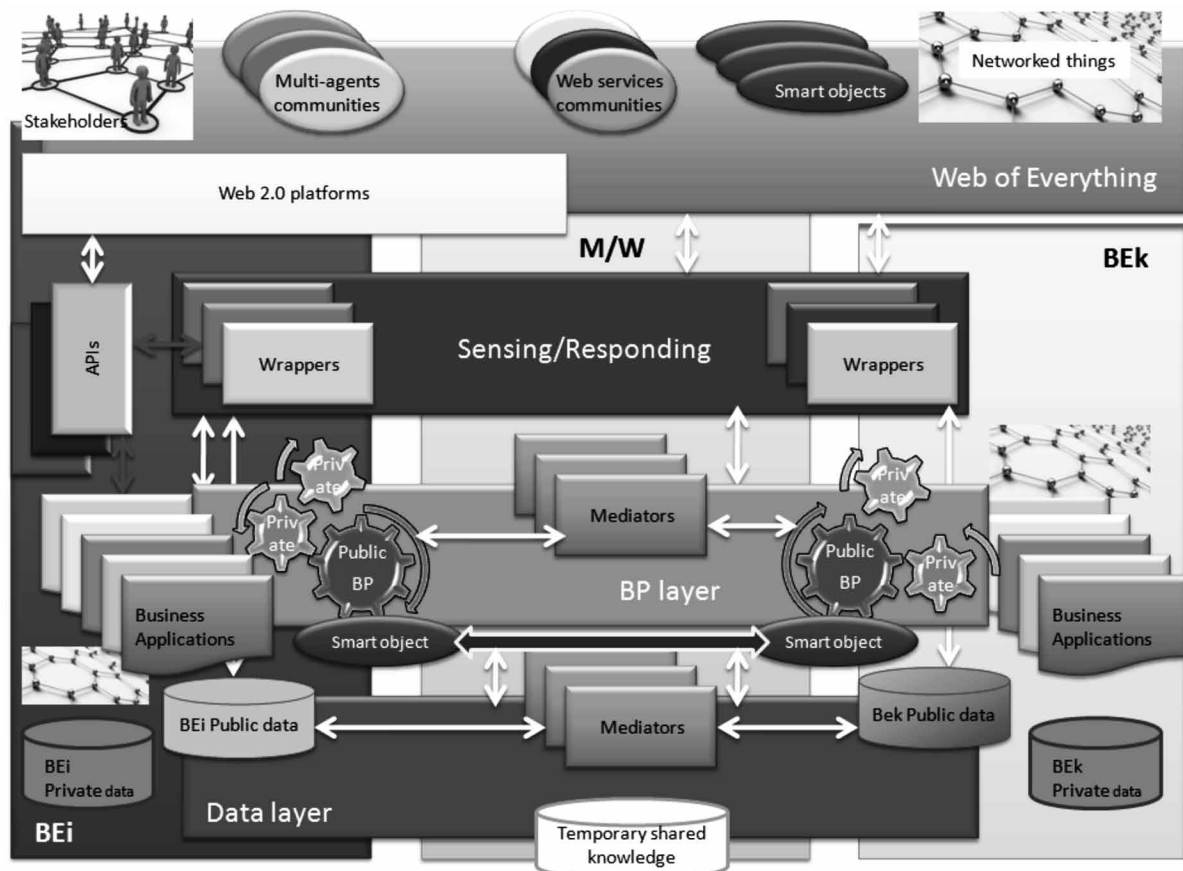
Conflicts between data may appear either in data itself and/or in data schemas. At the data level, conflicts are caused by different representations and interpretations of same data (e.g., expressions, units, and precision) (Kim & Seo, 1991). At the schema level, conflicts are characterized by differences in logical structures and/or inconsistencies of metadata in the same application domain. Examples include names, entity identifiers, schema isomorphism, and other schema discrepancies (Sheth & Kashyap, 1992). Business data also suffers from semantic conflicts. Nevertheless, there are also specific conflicts that may occur between business data caused by different business documents, products and services catalogues and classification schemas as well. There are several reasons for this including, but not limited to, different document names, different

document structures, document elements and their representation, etc. These differences render BEs unable to establish collaboration between their business processes.

Semantic conflicts between business data are very complicated and differ from data integration in distributed heterogeneous databases. Research on business data (Ng, et al., 2000) shows that a typical product schema tree has two basic characteristics: shallowness (two or three levels) and bushiness (too many attributes, at the second level of a tree, most of them N:1 to the root of tree). There is also typically very limited knowledge about local schemas, a huge number of them, and fast evolution of these schemas.

At the business process level non-standardized business terminology, lack of common acceptable and understandable processes (grammar) and lack

Figure 3. Enterprise interoperability in the presence of ubiquitous Web



of common dialog rules (protocols). The problem arises from incompatibility of B2B protocols, i.e. from the heterogeneity of atomic transactions that appear in public part of a business process (Kajan, 2010). In a VE conflicts may be caused by different business rules that participating BEs may express depending of collaboration patterns among them (Kajan, et al., 2016).

By this end we have address the well known interoperability obstacles. When we come into the era of ubiquitous Web, where people, machines, software entities, communities of Web services, agents and multi-agents systems, smart objects, etc., offer opportunities to play important role in enterprise interoperability, a variety of new challenges have raised that should be solved to achieve full, conflict-free interoperability. But, new obstacles for enterprise interoperability have also risen. These include, but not limited to: dynamic nature of social and ubiquitous world, Big Data challenges, stakeholder's autonomy, privacy, sentiments-driven data, etc. A part of such heterogeneous environment has shown in Figure 3. The crossroads between M/W (middleware) and sensing/responding layer becomes one of the most research challenges for bringing enterprise interoperability into full potential.

BACKGROUND ON ENTERPRISE INTEROPERABILITY EFFORTS

The first serious requirement for data integration appeared in big enterprises whose systems consisted of many LANs mutually connected by either private or public infrastructures. In the 1980s, this requirement was satisfied by federated databases (Sheth & Larson, 1990). The problem with that approach was cost; it uses 1:1 relationship between every pair of databases involved in an integration process. Data Warehouses (DWs) offered a cheaper solution (Chaudhury & Dayal, 1997). Problem with DWs is their potential size because they are accumulating all the data during their use. Next step was represented by Data

Marts, (DM) introducing an interface data layer between query technologies and DWs. All of these three database technologies may be implemented in A2A, whilst for B2B they have no particular importance.

The big step forward for data integration in distributed environments has been made by eXtensible Markup Language (XML). The power of XML is based on its extensibility that allows users to create own domain-specific tags using Document Type Definition (DTD) or XML schema languages. XML has not been developed to define data semantics, thus its power is limited to common data syntax.

Another axis for data integration is semantic Web technologies (Berners-Lee, et al., 2001). The pool of human knowledge, as the Web was described by its inventor, has been transformed in recent years into a digital jungle with billions of documents in it. The idea behind the semantic Web is machine-readable Web supported with knowledge-oriented technologies such as Resource Description Framework (RDF) and Ontology Web Language (OWL). Although RDF is intended to give meaning to data by metadata and classes, respectively, it does not cover complete semantic requirements for interoperability. Here is the place where ontology deployment may play an important role.

In data integration, it is very common to use a combination of mediators and wrappers (Wiederhold, 1992), a three-layer architecture consisting of clients, middle tier and heterogeneous databases on the back layer. This integration architecture has been an active research area for almost twenty years. Examples include, but are not limited to, relational database integration (Papakonstantinou, et al., 1995; Yerneni, et al., 1999), B2B integration (Kajan & Stoimenov, 2005), integration of Deep Web data (Zhang, et al., 2012), etc.

The requirements for interoperability between processes also caused the development of middleware technologies. The first of them is Remote Procedure Call (RPC), the technology that allows clients to invoke remote procedure the same way as

local calls do (Birell & Nelson, 1984). The second generation of middleware technologies is based on distributed objects, the typical representatives of which are Common Object Request Broker Architecture (CORBA), Distributed Component Object Model (DCOM), and Enterprise Java Beans (EJB); the critical assessment on these technologies may be found in (Medjahed, et al., 2003).

Significant improvements in process interoperability have been made by Web Services, which has been considered as a key technology for the development of dynamic EC (Chen, et al., 2003). Web service environments are highly dynamic and distributed by nature, thus they required some requirements that should be met in order to achieve success of their composition. These requirements are identified at definition, selection and execution phases of Web services (Sheng et al., 2014). Although a lot of research on this promising technology has been conducted over the decade, the recent overview of these efforts (Sheng et al., 2014) shows that there is no maturity reached yet and it still contains some open issues that should be taken into account. These issues are related to composition that should be dependable, acceptable, autonomous and pervasive, but also to Restful services and data mashup capability, and security support as well.

Business process interoperability is one of the hot research topics for years. It includes analysis of potential technologies (Kajan, 2004; Medjahed, et al., 2003), Semantically Enabled Service-oriented Architectures (SESA) and modeling of choreography components for business process harmonization according to their structure and timing (Cardoso, 2007, Nacer & Aissani, 2014). Despite different approaches, all given research examples assume some kind of ontology deployment including ontology-based representation of product and services a given, for example in (Hepp, 2006; Hepp, 2008). There are several research prototypes of ontology-based middleware. Examples include, but not limited to, Toronto Virtual Enterprise (TOVE) (Fox & Gruninger, 1998), (Bremen University Semantic

Translator for Enhanced Retrieval (BUSTER) (Visser, 2004), Conflict Resolution Environment for Autonomous Mediation (CREAM) (Park & Ram, 2004), etc. Recent efforts that deploy ontologies into VE interoperability arena are motivated by variety of reasons. For example, Enterprise Competency Organization Schema (ECOS) aiming to standardize BE's profiles by shared ontology that may be used in pre-creation phase of VE in order to make the best choice in partner selection (Khilwaki, 2011). Wajid et al. (2013) work on a system to interplay between automatic recognition of business documents and human intervention by means of assisting mapping tool. From the viewpoint of semantic interoperability across VE, (Kajan, et al., 2016) proposes the use of goal based models for VEs in order to define up front the common goals of the VE and avoid ambiguities during enactment.

FUTURE RESEARCH DIRECTIONS

Towards Dynamic Knowledge-Based Interoperability

When we back to the fact that everything is going to be connected, if not yet, we are facing with, more or less, soft or hard, boundaries that exist among a variety of technology-driven networks, e.g. human-, service-, semantic-networks, networks of things and intelligent networks, as well. By this, the virtual world will be extended by a number of more complex physical objects acting in special environments. Collaborations in future real-time collaborations will take place between human beings, computers and facilities cooperating in different domains and environments. This enhances the still high heterogeneity of digital life and brings up new interoperability challenges. – drastically more objects to be identified and characterized, seamless connections between the Internet and radio networks, more complex feedback controlled systems, the need for more powerful and frameworks, concepts, models and monitoring systems.

Because of this heterogeneity, the restricted ability to control such complex interactions may become critical and new risks may arise. A big challenge will be to ensure visibility, common understanding and interoperability on all levels (Dorloff & Kajan, 2012). In consequence additional and even new standards and regulations on all levels of the Interoperability framework will be needed.

Although the research community continues to look into the Web from a technical and computational perspective, a *social perspective* has lately emerged as a direct result of the tremendous advances of Web 2.0 technologies. Web 2.0 technologies (e.g., REST and JSON) and Web 2.0 applications (e.g., LinkedIn, Tweeter, and Facebook) help enterprises tap into the social media so they can reach out to more customers, profile customers, gauge customers' satisfaction levels, and so on. Lee emphasized that the power of Web 2.0 does not rely on its software tools and techniques, but on data that Web 2.0 may bring within (Lee, 2011). He pointed out the six new business models behind Web 2.0 technologies, in which use of these technologies are primary tools to make revenue and survive on the market. In (Bonchi et al., 2011) special emphasis has given to data mining from social networks in order to improve business through Social CRM, recommender systems, collaboration with customers, production forecasting, expert finding, churn (loss of customers) analysis, etc. The Enterprise 2.0 (aka social enterprise) is the one that connects successfully business and social worlds in such way that accommodates the concerns of both the business community and the social community. Such connection requires a meet-in-the-middle environment capable to handle interactions between these two worlds (Buregio, et al., 2015). By now we will use a Social Machine-based approach for such environment (Buregio, et al., 2013). We believe that Social Machines actually represent a promising paradigm to deal with the complexity of this new emerging Web around us, and a practical way to explain each and every entity connected to it. We surely can use the concepts and instruments of

the existing Web 2.0 but we need more and higher sophisticated frameworks, methods knowledge and experience. A conceptual view to an extension of social machine with semantics capabilities using knowledge artifacts is given in (Buregio, et al., 2016). In that concept, each SSM can be viewed as a "knowledge-as-a-service" provider, once it offers different APIs that allow delivering insights into business and social artifacts it acts over. Such environment helps not only to bridge the gap between the business and social worlds but also to optimize the sharing and accessing of knowledge across the entire organization and stakeholders it interacts with. Bringing semantic to the social Web is a step towards intelligent networks that will allow intelligence-to-intelligence communications.

In the center of this concept are intelligent software agents, resident or mobile, armed with various AI skills and capable to act as independent self-proactive software artifacts. Since Kasbah, the first e-commerce agent marketplace (Chavez & Maes, 1996), up to day, research on agents, multiagent systems and agent communication languages is taking much interest in the research community, good overview of which may be found in (Wooldridge, 2009). Such agents are capable to build not only intelligent networks, but also to make a "self-active glue" that will be able to dynamically connect necessary pieces coming from other networks.

CONCLUSION

This article shows several lessons that we learned, which may be summarized as follows. Many efforts have been devoted to overcoming interoperability problems of any kind, but they have not reached their promise yet. Instead of unlimited interoperability inside and between *BEs*, the result was too many mutually incompatible "standards" that formed isolated areas of e-business seeking for new investments over and over again.

Recent research efforts are looking very promising. They are concentrating on two big goals: to provide middleware technologies and frameworks which are capable to act as self-acting dynamic artifacts at the time of invocation and to provide necessary semantic understanding between peers of any kind. Despite of well-defined goals and recent results, the whole thing is still under development. At least but not last, many business processes yet waiting to be socially modeled and managed. Is there a fiction or a forecast?

We believe that the whole idea based on knowledge-enabled self-acting artifacts is not yet another hip and that it will lead the Web and EC to their full potential. Let us finish this article with few visions given in the past. “Now we can imagine the world of people with active machines forming part of the infrastructure. We have only to express a request for bids, or make a bid, and machines will turn a small profit matching the two” wrote Tim Berners-Lee in the foreword of the book about Semantic Web (a.k.a. Web 3.0) (Fensel et al., 2003). For 25th anniversary of the Web San Murugesan cited a computer scientist Alan Kay: “The best way to predict a future is to invent it” emphasizing technological achievements on the Web and encouraging further application development towards more ubiquitous and smarter than current (Murugesan, 2014). Enterprise applications armed with full interoperability features will be there, for sure.

REFERENCES

- Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. *Scientific American*, 284(5), 34–43. doi:10.1038/scientificamerican0501-34 PMID:11396337
- Birrell, A. D., & Nelson, B. J. (1984). Implementing Remote Procedure Calls. *ACM Transactions on Computer Systems*, 2(1), 39–59. doi:10.1145/2080.357392
- Bonchi, F., Castillo, C., Gionis, A., & Jaimes, A. (2011). Social Analysis and Mining for Business Applications. *ACM Transactions on Intelligent Systems and Technology*, 2(3), 22. doi:10.1145/1961189.1961194
- Buregio, V., Faci, N., Kajan, E., Maamar, Z., & Sellami, M. (2016). Bringing Semantics to the Social Web. *Proceedings of 4th International CPMMI Conference*.
- Burégio, V., Maamar, Z., & Meira, S. (2015). An Architecture and Guiding Framework for the Social Enterprise. *IEEE Internet Computing*, 19(1), 64–68. doi:10.1109/MIC.2014.85
- Buregio, V., Meira, S., & Rosa, N. (2013). Social machines: a unified paradigm to describe social web-oriented systems. *WWW'13*, 885–890.
- Cardoso, J. (2007). *Semantic Web Services: Theory, Tools and Applications*. Hershey, PA: IGI Global. doi:10.4018/978-1-59904-045-5
- Chaudhuri, S., & Dayal, U. (1997). An overview of data warehousing and OLAP technology. *SIGMOD Record*, 26(1), 65–74. doi:10.1145/248603.248616
- Chavez, A., & Maes, P. (1996). Kasbah: An agent marketplace for buying and selling goods. In *Proceedings of the International Conference on the Practical Application of Intelligent Agents and Multi-Agent Technology*, 75–79.
- Chen, M., Chen, A., & Shao, B. (2003). The Implications and Impacts of Web Services to Electronic Commerce Research and Practices. *Journal of Electronic Commerce Research*, 4(4), 128–139.
- Dorloff, F.-D., & Kajan, E. (2012). Balancing of Heterogeneity and Interoperability in E-Business Networks: The Role of Standards and Protocols. *International Journal of E-Business Research*, 8(4), 15–33. doi:10.4018/jebr.2012100102

- Faci, N., Maamar, Z., Kajan, E., & Benslimane, D. (2014). Research Roadmap for the Enterprise 2.0 – Issues & Solutions. *Scientific Publications of the State University of Novi Pazar Series A: Applied Mathematics. Informatics and Mechanics*, 6(2), 81–89.
- Fensel, D., Hendler, J., Lieberman, H., & Wahlster, W. (Eds.). (2003). *Spinning the Semantic Web*. MIT Press.
- Fox, M. S., & Gruninger, M. (1998). Enterprise Modeling. *AI Magazine*, 109–121.
- Hepp, M. (2006). Products and Services Ontologies: A Methodology for Deriving OWL Ontologies from Industrial Categorization Standards. *International Journal on Semantic Web and Information Systems*, 2(1), 72–99. doi:10.4018/jswis.2006010103
- Hepp, M. (2008). GoodRelations: An ontology for describing products and services offers on the Web. *Lecture Notes in Computer Science*, 5268, 332–347. doi:10.1007/978-3-540-87696-0_29
- Huhns, M. N., & Singh, M. P. (2005). Service-oriented computing: Key concepts and principles. *IEEE Internet Computing*, 9(1), 75–81. doi:10.1109/MIC.2005.21
- Kajan, E. (2004). The Maturity of Open Systems for B2B. *ACM SIGecom Exchanges*, 5(2), 34–44. doi:10.1145/1120687.1120692
- Kajan, E. (2010). Interoperability Issues of Business Processes: Key Issues and Technological Drivers. In I. Lee (Ed.), *Encyclopedia of E-Business Development and Management in the Global Economy* (pp. 908–917). Hershey, PA: IGI Global. doi:10.4018/978-1-61520-611-7.ch091
- Kajan, E., Faci, N., Maamar, Z., Loo, A., Pljakovic, A., Sheng, Q. Z., & Dustdar, S. (2014). The Network-based Business Process. *IEEE Internet Computing*, 18(2), 63–69. doi:10.1109/MIC.2014.35
- Kajan, E., Narendra, N. C., & Maamar, Z. (2016). Towards Conflict-free Virtual Enterprises. In I. Lee (Ed.), *Encyclopedia of E-Commerce Development, Implementation and Management* (pp. 1116–1129). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-9787-4.ch079
- Kajan, E., & Stoimenov, L. (2005). Toward an Ontology-Driven Architectural Framework for B2B. *Communications of the ACM*, 48(12), 60–66. doi:10.1145/1101779.1101810
- Khilwani, N., Harding, J. A., & Tiwari, M. K. (2011). Enterprise competence organization schema: Publishing the published competences. *Journal of Engineering Manufacture*, 225(6), 921–942. doi:10.1177/09544054JEM2097
- Kim, W., & Seo, J. (1991). Classifying Schematic and Data Heterogeneity in Multidatabase Systems. *IEEE Computer*, 24(12), 12–18. doi:10.1109/2.116884
- Lee, I. (2011). Overview of Emerging Web 2.0-Based Business Models and Web 2.0 Applications in Business: An Ecological Perspective. *International Journal of E-Business Research*, 7(4), 1–16. doi:10.4018/jebr.2011100101
- Maamar, Z. (2003). Commerce, e-Commerce, and m-commerce, what comes next? *Communications of the ACM*, 46(12), 251–257. doi:10.1145/953460.953508
- Maamar, Z., Faci, N., Mostéfaoui, S. K., & Kajan, E. (2013). *Network-based Conflict Resolution in Business Processes* (pp. 132–137). Coventry, UK: IEEE ICEBE.
- Medjahed, B., Benetallah, B., Bouguettaya, A., Ngu, A. H., & Elmargamid, A. K. (2003). Business-to-business interactions: Issues and enabling technologies. *The VLDB Journal*, 12(1), 59–85. doi:10.1007/s00778-003-0087-z
- Murugesan, S. (2014). The Web @ 25 and Beyond. *Computing Now*, 7(7).

- Nacer, H., & Aissani, D. (2014). Semantic web services: Standards, applications, challenges and solutions. *Journal of Network and Computer Applications*, 44, 124–151. doi:10.1016/j.jnca.2014.04.015
- Narendra, N. C., Le, L.-S., Ghose, A., & Sivakumar, G. (2013). Towards an Architectural Framework for Service-oriented Enterprises. *Lecture Notes in Computer Science*, 7759, 215–227. doi:10.1007/978-3-642-37804-1_22
- Ng, W. K., Yan, G., & Lim, E.-P. (2000). Heterogeneous Product Description in Electronic Commerce. *ACM SIGecom Exchanges*, 1(1), 7–13. doi:10.1145/844302.844305
- Papakonstantinou, Y., Gupta, A., Garcia-Molina, H., & Ullman, J. (1995). A query translation scheme for rapid implementation of wrappers. *DOOD*, 95, 161–186.
- Park, J., & Ram, S. (2004). Information systems interoperability: What lies beneath? *ACM Transactions on Information Systems*, 22(4), 595–632. doi:10.1145/1028099.1028103
- Qin, Y., Sheng, Q. Z., Falkner, N. J. G., Dustdar, S., Wang, H., & Vasilakos A. V. (2014). *When Things Matter: A Data-Centric View of the Internet of Things*. CoRR, abs/1407.2704
- Romero, D., & Vernadat, F. (2016). Enterprise Information Systems State of the Art: Past, Present and Future Trends. *Computers in Industry*, 79, 3–13. doi:10.1016/j.compind.2016.03.001
- Sheng, Q. Z., Qiao, X., Vaslikos, A. V., Scabo, C., Bourke, S., & Xu, X. (2014). Web services composition: A decades overview. *Information Sciences*, 280, 218–238. doi:10.1016/j.ins.2014.04.054
- Sheth, A. P., & Kashyap, V. (1992). *So far (schematically), yet so near (semantically)*. IFIP WG2.6 Database Semantics Conference on Interoperable Database Systems, Lorne, Australia.
- Sheth, A. P., & Larson, J. A. (1990). Federated database systems for managing distributed, heterogeneous and autonomous databases. *ACM Computing Surveys*, 22(3), 183–236. doi:10.1145/96602.96604
- Tan, W., Blake, M. B., Saleh, I., & Dustdar, S. (2013). Social-Network-Sourced Big Data Analytics. *IEEE Internet Computing*, 17(5), 62–69. doi:10.1109/MIC.2013.100
- Van der Aalst, W. M. P. (2014). In K. Mertins, F. Benabeu, R. Poler, & J.-P. Bourriers (Eds.), *Data Scientist: The Engineer of the Future* (pp. 13–26). Enterprise Interoperability, VI.
- Verginadis, Y., Apostolou, D., Papageorgiou, N., & Mentzas, G. (2011). OCEAN: An ontology for supporting interoperability service utility in virtual organizations. *Int. J. Networking and Virtual Organizations*, 9(2), 184–209. doi:10.1504/IJNVO.2011.042418
- Visser, U. (2004). Genaral Approach of Buster. *Lecture Notes in Computer Science*, 3159, 37–51. doi:10.1007/978-3-540-28636-3_3
- Wajid, U., Namoun, A., Marin, C., & Mehandijev, N. (2013). Designing and evaluating system of recognition to support interoperability among collaborative enterprises. *Computers in Industry*, 64(5), 598–608. doi:10.1016/j.compind.2013.03.003
- Wiederhold, G. (1992). Mediators in the architecture of future information systems. *IEEE Computer*, 25(3), 38–49. doi:10.1109/2.121508
- Wooldridge, M. (2009). *An introduction to multi-agent systems* (2nd ed.). New York, NY: John Wiley & Sons.
- Yerneni, R., Garcia-Molina, H., & Ullman, J. (1999). Computing capabilities of mediators. ACM SIGMOD99, Philadelphia, PA. doi:10.1145/304182.304221

Zhang, H., Lin, L., & Qiang, B. (2012). A Semantic Wrappers of Deep Web Data Source Based on RDF View. *Journal of Computer Information Systems*, 8(9), 3527–3539.

KEY TERMS AND DEFINITIONS

A2A (Application-to-Application): Process of data and application integration inside an enterprise in order to reach ZLE. Also known as EAI (Enterprise Application Integration) and EII (Enterprise Information Integration).

B2B (Business-to-Business): Type of EC where participating entities are enterprises at both sides. Nowadays, it may be thought as a third wave of EC, where *BEs* should be able to establish and handle their business relationship dynamically and seamlessly on demand.

Business Entity (BE): An enterprise, a part of an enterprise and/or its application that participate in A2A or B2B.

EC (Electronic Commerce): An emerging concept that describes the process of buying, selling, or exchanging data, services and products over the Internet.

Framework: A common template that has well defined functionality in order to solve a problem, e.g. interoperability, and precisely defined inputs and outputs intended for its communication with the external world.

Interoperability: An attribute given to systems, applications and data that assigns their ability to communicate with another systems, applications and data in a manner that they may exchange and mutually use that data.

Middleware: A dynamic self-organized layer, usually acting on the top of the transport layer, that provides uniformity between the lower layers of the framework (hardware, operating systems, etc.), which are different by default, and hides serious, natural discrepancies that exist among applications.

Virtual Enterprise (VE): A temporary alliance of BEs that come together to jointly fulfill common business goals.

ZLE (Zero Latency Enterprise): A *BE* that has been reached the full A2A integration so that there is no data latency between its own business processes.

ENDNOTE

- ¹ IEEE Standard Computer Dictionary: A Compilation of IEEE Standard Computer Glossaries. New York, NY. (1990)

Has Bitcoin Achieved the Characteristics of Money?

Donovan Peter Chan Wai Loon
University of Malaya, Malaysia

Sameer Kumar
University of Malaya, Malaysia

INTRODUCTION

In late 2013 and early 2014, the term bitcoin had made headlines in the global financial news. Designed and developed by Satoshi Nakamoto¹ (Nakamoto, 2008), this “virtual money” was propelled by cyber lobbyist and at the end of 2013 the exchange rate per bitcoin had multiplied more than five times over a space of several weeks. Supporters claim that the fundamental objective of bitcoin is to act as a substitute to the current payment systems and to allow dealings internationally, across currency denominations, without intervention of dominant bodies or central banks, and without the abuse of traditional financial institutions such as banks. According to followers of electronic currencies, government administrations regularly enforce unwanted controls and constraints and central banks may expedite oversupply of money leading to increased inflation. In addition, a large number of supporters also complain on the overpriced fees charged among other claimed exploitations done by banks. The dramatic rise in bitcoin’s value in November 2013 had triggered exceptional interest to it. The trading price of a bitcoin was only USD0.05 when it begun in 2010 and had reached and all time high exceeding USD1, 200.00 (Wood, 2013)

As the main role of Bitcoin’s creation is to function as an alternate form of currency that individuals can use to conduct transactions with each another without the interest of governing bodies such as banks, the main question that arises is

“how well can bitcoin serve the role of currency?” Money is commonly accepted to be a mechanism that functions as a medium of exchange, a unit of account, and a store of value (Halaburda & Gans, 2014). For the purpose of this paper we will be looking at how bitcoin has achieved these three functions.

BACKGROUND

Essentially, bitcoin is a peer-to-peer network version of e-cash that facilitates transactions between parties minus the proof requirement of an appointed third party; i.e. banks or financial institutions. This unit of network is denoted as bitcoin, which is considered by most to be the greatest digital currency to date (Brito & Castillo, 2013; Lo & Wang, 2014).

The advantages that bitcoin (Rogojanu & Badea, 2014) seems to offer are: 1) it is believed to be a completely decentralized method not linked with any predominant bodies, central banks, or recognized payment systems, which are ruled by banks and therefore it is thought to be less susceptible to exploitation or fraud; 2) it has pseudonymous characteristics; 3) it forces no direct charges on transactions and shows the possibility for transaction fees in general to be lesser. A person can use a bitcoin, by sending his or her account number which is the “public key” and a password which is the “private key” for authentication on the public transaction record,

known as the “block chain” (Reid & Harrigan, 2013) Certain individuals or “miners” will then use their computers processing power to authenticate if the transaction is real by solving a rigorous computational task also known as “finding the hash of a nonce” (Lo & Wang, 2014). As a payment for validating the transaction, the first person or party to provide the answer to the task will be compensated with a certain number of bitcoin, accumulating to the available stock of bitcoin and consequently generate money creation.

There is a process in the system, which routinely regulates the computational difficulty of validating transactions to guarantee that each transaction is normally verified within 10 minutes (Bamert, Decker, Elsen, Wattenhofer, & Welten, 2013). However, it is unavoidable that there are some uncertainties in the actual extent of time taken to verify each transaction. The same process also dictates the eventual quantity of bitcoin, which is estimated to be 21 million units by 2140.

BITCOIN AS MONEY

Medium of Exchange

While bitcoin has no inherent value, its price eventually depends on its practicality as a form of money in the consumer market. Indication of bitcoin’s mark in daily business is mostly hearsay, comprising of articles about individuals simply living by using bitcoin or estimations of big statistics of industries that are eager to agree to bitcoin. Majority of the high-ranking traders receiving bitcoins are led by companies in the software and hardware industry concentrated on marketing products dedicated to bitcoin applications and by exchange markets offering facilities to bitcoin investors.

Accurate understanding into the implementation of bitcoin can be acquired from data of the universal record of bitcoin transactions (Vander-vort, 2014). Data from numerous websites show that bitcoin daily transactions count has reached

capacities of tens of thousands but it is widely believed that most of these transactions comprise of activities between speculators, and only a few are actually used for trading of goods and services.

In March 2014, a co-founder of Coinbase Fred Ersham, the prominent digital wallet service, anticipated that 80% of activity on his site was associated with speculation. He also adds that 24,000 vendors are registered with Coinbase (Yermack, 2013). In a global market of billions of consumers, who make numerous commercial transactions everyday, bitcoin seems to have an extremely insignificant market presence.

Another barrier for bitcoin in developing as a general used means of exchange rises from the scarcity of obtaining new bitcoins. A user typically obtains bitcoins from vendors or virtual exchanges (Chowdhury & Mendelson, 2013). The user then needs to find a safe place to store the bitcoins such as digital wallets. Purchases of bitcoins cannot be usually made by credit card or via Paypal, however the user needs to connect his bank account or do a bank transfer to the exchange. The present bitcoin exchanges come with several issues such as low liquidity and to a certain extent execution risks. Another way is for users to mine bitcoin; but this is highly unlikely due to the huge capital investments needed to purchase supercomputers for the mining activity.

Lastly, bitcoin does not allow for users to receive goods and services from a vendor in advance without making prior payment for it. In most current online markets, users are constantly able to make purchases for items and services without money in hand. This is done through credit card facilities backed by the vendor or a third-party credit card company. Till date, there are no bitcoin related credit cards and the idea of such a facility seems to be unknown.

Unit of Account

Bitcoin is faced with a number of challenges for it to become a suitable unit of account. The first reason is due to its great instability. Bitcoin’s value

related to other currencies deviates significantly on daily basis. Merchants that deal with bitcoins often have to readjust prices, a habit that would be expensive to their businesses and complicating to the customer. In an economy that uses bitcoin as the main form of currency such a problem would not happen, but such environment is not available in today's society.

Another associated obstacle occurs due to the difference in "market values" that a bitcoin trades for at any certain period of time. This discrepancy of market prices, range from high to low estimations, causing a bitcoin's price to alter over several transactions within the same day. Many merchants are faced with the problem of changing market values of bitcoin and rely heavily on price aggregations, which does not take into account the merchants and customers actual cost of purchasing or selling a bitcoin at the current time. These circumstances are unimaginable in a present well-developed market as it clearly goes against the law of having a single price.

Possibly the greatest problem for bitcoin to become a generally used unit of account is its comparatively high price per unit. Bitcoin supporters more often than not, underestimate and overlook this important detail. At the present time of writing this paper, the average value of a bitcoin is USD247.88. This forces merchants to price goods with four or more decimals using the bitcoin (BTC) prices. Although supporters argue that the computations are simple and direct, it is difficult to adopt as no other currency in the world so far quotes prices for goods in this manner. General accounting software presently at use also adjusts its calculation to two decimal points for prices of goods. Advocates tend to disregard the failure of bitcoin to create prices that are familiar to customers. Sources are inclined to agree that the current computational data sets allow for bitcoins to be divided and have up to eight decimals places offering the prospect for a big enough number of units.

Store of Value

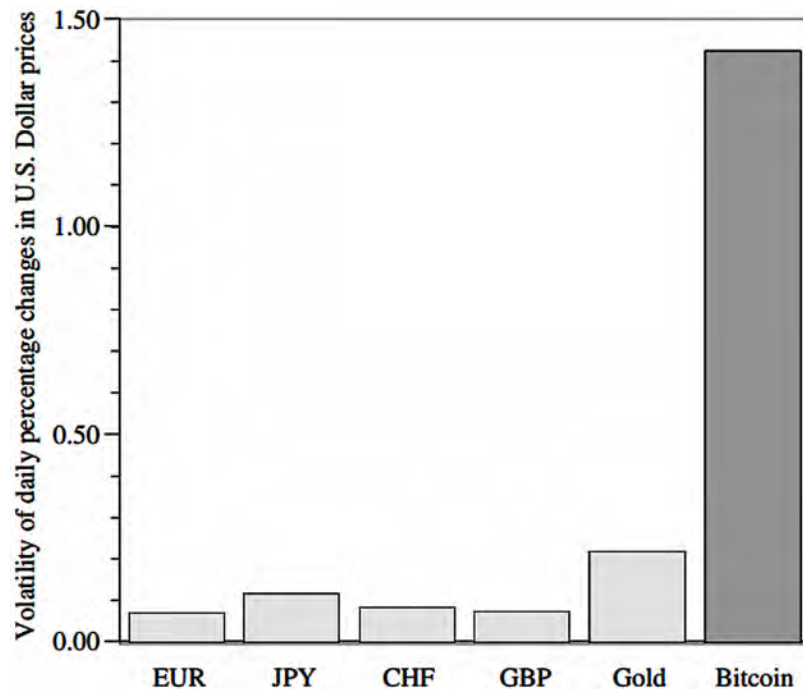
When money acts as a store of value, a person acquires the money and may choose to exchange it for goods or services at a later time to his convenience. When he does this, the person believes to get the same economic worth of what the money was when he had received it.

In the past, considering money as a store of value basically meant safeguarding it against theft. This was done by storing it in hidden places or by depositing it into a bank. Such approaches cannot be used to secure bitcoins as it does not have any physical form. Bitcoins must be kept inside computer software called "digital wallets", which safety has become a main concern of the bitcoin industry. Several digital wallet companies have sought the assistance of third-party insurers to provide some basic deposit insurance. This however does not exclude the customer from assuming the financial risk (Moore & Christin, 2013) should there be a breach in the security of both the digital wallet and insurance company.

If a customer discovers an effective means to keep and protect his bitcoins, he then has to deal with additional issues of handling the threat of bitcoin's high volatility. Figure 1 below depicts the volatility of the bitcoin-dollar exchange rate, using daily data gathered in 2013. For evaluation reasons the graph also depicts other major currencies and their volatilities of the exchange rates and the London Price of gold (Yermack, 2013). The major currencies used were Euro, Yen, British Pound, and Swiss Franc.

Bitcoin recorded the highest exchange rate volatility in 2013 with 142%. Exchange rate volatilities of major currencies were within the range of 7% to 12%. Gold, which is another credible source as a store of value had a volatility of 22% in 2013. In his analysis Yermack (2013) explains that majority of vastly traded shares have volatilities around 20% to 30%, and even high-risk shares seldom demonstrate a volatility as high as 100%.

Figure 1. Bitcoin volatility compared to major currencies and gold
(Yermack, 2013)



Based on Figure 1, it is possible to determine that keeping bitcoins even for a brief a period is dangerous, due to its inconsistency in acting as a store of value.

FUTURE RESEARCH DIRECTIONS

Most major currencies in global economy today are either supported by legal bodies or governments. Therefore, the question is “Who would have confidence in Bitcoin?” as it is not supported by either of the two entities mentioned above. However, people appear to put their trust in bitcoin, as the continuous buying of bitcoins is proportionate to keep it nearly at par with the US dollar. In this section, we will assess if this confidence is correct.

Confidence in Bitcoin

Most people tend to look at bitcoin as more of an investment. Similar to most other investments,

bitcoin is also vulnerable to unreasonable price bubbles and also ungrounded or grounded loss of confidence, which would cause the demand to fall and create a surplus in supply. Confidence in bitcoin may drop due to unanticipated fluctuations in the inflation imposed by software or hardware developers, a government crackdown, the creation of superior competing alternative currencies, or a deflationary spiral (Grinberg, 2011). Confidence may also drop due to technical issues; anonymity of the bitcoin system is exposed, theft of money, or if new transactions are denied by hackers or governments. Certain people put their faith in bitcoin as they trust that bitcoin is not governed by any central organization with the authority to expand the money supply above the inflation rate pre-programmed automatically into the software. Developers may however be able to do so, and even if it is done for a good purpose and with the consent of majority of bitcoin users, it could cause lost of confidence by many. In another aspect, the inflation rate could be increased tremendously that bitcoin suffers

from hyperinflation causing a steep drop of its value. A better rival currency (Barber, Boyen, Shi, & Uzun, 2012) could create a loss of confidence to bitcoin and reduce its value or cause the entire collapse. While bitcoin may be tough to shut down because of its distributed nature, government restrictions may yet be disastrous to the confidence level in bitcoin. Most users do not want to be associated let alone own a currency that is against the law. Hence, the future research must more look into ways and means as to how users and common people could have confidence in this new currency and treat it in at least similar to the traditional currencies.

Deflationary Spiral

Bitcoin may experience deflation that causes certain groups or even businesses to abandon bitcoin. This could possibly lead to a total meltdown of the system or mere depreciation in the value of bitcoin. As the higher limits of bitcoins is set at 21 million, bitcoins may become more treasured as time goes by as the supply of government supported currencies continue to grow. This statement is backed by opinions that people would source for commodities that are scarce and difficult to attain. However, when the prices of bitcoin take a downturn, miners of bitcoins may react by decreasing production leading to reduced demands and ultimately lower the value of it. The final result is underemployment and loss of capital. Consequently, bitcoin related industries might opt to end it and further limit bitcoin's reach. Future research could also look into ways that this e-currency remains stable. As mentioned in the section earlier, one of the biggest issues of bitcoin is its potential

Potential Technology Failures

Technological letdowns could also result in loss of confidence by users and discourage them from performing bitcoin transactions. Listed below are a few examples:

All bitcoin transactions are public. These transactions however are anonymous since there is nothing linking individuals or companies to the accounts in the transactions (Miers, Garman, Green, & Rubin, 2013). Occasionally, people do put their details online, which can be traced back to their virtual identities. There are possible techniques employing arithmetic calculations that can undo the anonymity of the system. Under unforeseen conditions and rapid exposure of this would be most detrimental to bitcoin's value. Just like physical money, bitcoin can be stolen or lost. Storing bitcoins in a personal computer is as risky as having large amounts of cash in a physical wallet. Users are generally advised to make backups and protect bitcoins through trusted online services such as digital wallet companies. Hackers employ an array of methods to conduct theft that target both individuals and large companies alike. Viruses or Trojans will conceal themselves in an individual's computer and steal the bitcoin data without the person even realizing before it's too late. Large-scale attacks can cause a confidence crisis amongst users of bitcoin. Even though bitcoin runs on a distributed network and usually has no specific point to breakdown, it is still prone to attacks by denial of service. Parties in the bitcoin mining system that have sufficient capabilities in computational power can efficiently stop any transaction from occurring. Gaining the required tools to conduct is fairly easy but expensive. This may be done by groups that stand to gain from the disruption of the bitcoin system such as hackers who want to hold bitcoin companies at ransom, governments who want stop bitcoin and even people who have other related interests. The future research has to look into ways and means to identify other potential failures and strengthen security.

CONCLUSION

The concept and innovation of bitcoin is ingenious. However for bitcoin to become more than just a

novelty and establish itself as a widely accepted currency, bitcoin must first have a more stable value. Bitcoin must also overcome the problems of its unconventional pricing mechanism, shortage of vendors who accept it, and the difficult way of obtaining bitcoins. Users of bitcoin will also need to have adequate computer knowledge when dealing with this currency thus creating an added barrier to a wider implementation of bitcoin.

It is uncertain to say what the future holds for bitcoin and its network. Nevertheless, a growing interest among society and scholars on the subject of virtual payment technologies and digital money's incredible potential on e-commerce warrants further research and development of understanding of this area.

REFERENCES

- Bamert, T., Decker, C., Elsen, L., Wattenhofer, R., & Welten, S. (2013). *Have a snack, pay with bitcoins*. Paper presented at the Peer-to-Peer Computing (P2P), 2013 IEEE Thirteenth International Conference on. doi:10.1109/P2P.2013.6688717
- Barber, S., Boyen, X., Shi, E., & Uzun, E. (2012). Bitter to better—how to make bitcoin a better currency. *Financial Cryptography and Data Security*, 399-414.
- Brito, J., & Castillo, A. (2013). *Bitcoin: A primer for policymakers*. Academic Press.
- Chowdhury, A., & Mendelson, B. K. (2013). *Virtual Currency and the Financial System: The Case of Bitcoin*. Working Papers and Research (2013-09).
- Grinberg, R. (2011). Bitcoin: An innovative alternative digital currency. *Hastings Science & Technology Law Journal*, 4, 160.
- Halaburda, H., & Gans, J. (2014). Some Economics of Private Digital Currency. In *Economic Analysis of the Digital Economy*. University of Chicago Press.
- Lo, S., & Wang, J. C. (2014). *Bitcoin as money?* Academic Press.
- Miers, I., Garman, C., Green, M., & Rubin, A. D. (2013). *ZeroCoin: Anonymous distributed e-cash from bitcoin*. Paper presented at the Security and Privacy (SP), 2013 IEEE Symposium on.
- Moore, T., & Christin, N. (2013). Beware the middleman: Empirical analysis of Bitcoin-exchange risk. *Financial Cryptography and Data Security*, 25-33.
- Nakamoto, S. (2008). *Bitcoin: A peer-to-peer electronic cash system*. Academic Press.
- Reid, F., & Harrigan, M. (2013). *An analysis of anonymity in the bitcoin system*. Springer. doi:10.1007/978-1-4614-4139-7_10
- Rogojanu, A., & Badea, L. (2014). The issue of competing currencies. Case study—Bitcoin. *Theoretical and Applied Economics*, 21(1), 103–114.
- Vandervort, D. (2014). Challenges and Opportunities Associated with a Bitcoin-Based Transaction Rating System. *Financial Cryptography and Data Security*, 33-42.
- Vigna, P. (2016). Craig Wright Says He Will Prove He Is Bitcoin Creator 'Satoshi Nakamoto'. *Wall Street Journal*. Retrieved from <http://www.wsj.com/articles/craig-wright-says-he-will-prove-he-is-bitcoin-creator-satoshi-nakamoto-1462309228>
- Wood, R. W. (2013). Bitcoin's \$13.50 To \$1,200 Eleven Month Climb--Now Taxes. *Forbes*.
- Yermack, D. (2013). *Is Bitcoin a real currency? An economic appraisal*. Academic Press.

ADDITIONAL READING

- Androulaki, E., Karame, G. O., Roeschlin, M., Scherer, T., & Capkun, S. (2013). *Evaluating user privacy in bitcoin* *Financial Cryptography and Data Security* (pp. 34–51). Springer. doi:10.1007/978-3-642-39884-1_4

Babaioff, M., Dobzinski, S., Oren, S., & Zohar, A. (2012). *On bitcoin and red balloons*. Paper presented at the Proceedings of the 13th ACM conference on electronic commerce.

Bohannon, J. (2016). The Bitcoin busts. *Science*, 351(6278), 1144–1146. doi:10.1126/science.351.6278.1144 PMID:26965609

Eriksson, B. (2016). A study of Bitcoin as a currency for email-based micro-transactions.

Grinberg, R. (2012). Bitcoin: An innovative alternative digital currency. *Hastings Sci. & Tech. LJ*, 4, 159.

Lischke, M., & Fabian, B. (2016). Analyzing the Bitcoin Network: The First Four Years. *Future Internet*, 8(1), 7. doi:10.3390/fi8010007

Marshall, R. (2016). Bitcoin: Where Two Worlds Collide. *Bond Law Review*, 27(1), 5.

Miers, I., Garman, C., Green, M., & Rubin, A. D. (2013). *Zero coin: Anonymous distributed e-cash from bitcoin*. Paper presented at the Security and Privacy (SP), 2013 IEEE Symposium on. doi:10.1109/SP.2013.34

Wood, R. W. (2013). Bitcoin's \$13.50 To \$1,200 Eleven Month Climb--Now Taxes. *Forbes*.

KEY TERMS AND DEFINITIONS

Bitcoin: A global electronic currency without an intermediary (decentralized).

Block Chain: A public record of all bitcoin transactions mined since the beginning of bitcoin cryptocurrency.

Block: A new group of transactions. Blocks connect the transaction together.

Cold Storage: The safe keeping of private keys disconnected from the Internet.

Cryptocurrency: A non-printed currency produced by solving mathematical problems.

DoubleSpend: Sending bitcoins to two different parties at the same time. Bitcoin mechanism makes it almost impossible to double spend it.

Miner: A computer or group of computers that do bitcoin transactions (adding new transactions or verifying blocks created by other miners. Miners are rewarded with transactions fees.

Mining: Payment processing work for recording payments in the public ledger.

Private Key: Key (cryptographic system) known only to the user.

ENDNOTE

- ¹ The real name of the original developer is often disputed (Vigna, 2016)

The Importance of Electronic Commerce in Modern Business

E**Kijpokin Kasemsap***Suan Sunandha Rajabhat University, Thailand*

INTRODUCTION

Electronic commerce (e-commerce) is the important Internet application that enhances the way of commercial transactions (Chang & Hu, 2012). E-commerce provides businesses to sell their products and services across the globe (Yapar, Bayrakdar, & Yapar, 2015). Given the tremendous success of the Internet and e-commerce in developed countries, emerging economies are quickly embracing information technology (IT) (Agarwal & Wu, 2015). The high efficiency, low trade costs, and simplified transaction process in e-commerce have changed manufacturers' operations, business, management models, and trade models (Lu & Liu, 2015). Multinational corporations often push their branches and suppliers to adopt e-commerce technology to connect with their global production network (Rahayu & Day, 2015).

E-commerce is beneficial to both business and consumer wise as payment and documentations can be completed with greater efficiency (Kasemsap, 2016a). The applications and opportunities for e-commerce are increasing (Huang, Lee, & Lee, 2009). Web technology and fast Internet access allow people to become more familiar with e-commerce (Jung, Cui, Eom, & Yeom, 2014). Information and communication technology (ICT) has fundamentally reshaped the way businesses and consumers can globally communicate, interact, and transact, thus generating the rapid growth of e-commerce (Pires, Stanton, & Salavrakos, 2010). E-commerce has made a significant impact on the traditionally vertical and horizontal relations of the supply chain, as well as the influenced social structure in modern

business (Chen, Yu, & Hsieh, 2013). Both virtual businesses and e-commerce operations enable organizations to have the fluid structures, modify the changing conditions, lower the expenses, and attain the new consumers (Gokmen, 2012).

This article aims to bridge the gap in the literature on the thorough literature consolidation of e-commerce. The extensive literatures of e-commerce provide a contribution to practitioners and researchers in order to maximize the business impact of e-commerce in modern business.

BACKGROUND

E-commerce has become an important factor in determining the future survival of organizations (Holsapple & Singh, 2000). E-commerce has substantial potential to foster the growth of small and medium-sized enterprises (SMEs) in developed and developing countries (Kurnia, Choudrie, Mahbubur, & Alzougool, 2015). Much of the existing literature on the SMEs' adoption of e-commerce technologies seeks to explain the e-commerce adoption behavior from the perspective of technology adoption, using theoretical frameworks, such as the Theory of Reasoned Action (Ajzen & Fishbein, 1975), the Theory of Planned Behavior (Ajzen, 1991), the Technology Acceptance Model (Davis, 1989), and the Diffusion of Innovation (DOI) (Rogers, 1995). Gibbs and Kraemer (2004) explained that the scope of e-commerce adoption across ten different countries is influenced by national factors such as government policy and e-commerce legislation.

DOI: 10.4018/978-1-5225-2255-3.ch243

The emergence of e-commerce presents the new challenges in understanding the consumers' switching intentions in the context of e-commerce in general and online auctions in particular (Lin, Cheng, Wang, & Chang, 2012). E-commerce consists of the purchasing and selling of products or services through electronic systems, such as computer networks and the Internet (Kasemsap, 2016b). E-commerce presents consumers with a convenient way of shopping outside of their local jurisdiction (Alm & Melnik, 2012). Online shopping seems to primarily reflect a desire for time efficiency on the part of the shopper (Anesbury, Nenycz-Thiel, Dawes, & Kennedy, 2016). Experience with online shopping directly increases the consumer's intention to use the Internet to buy products (Soopramanien, 2011).

E-commerce is defined as a reduction in the consumers' search costs (Goldmanis, Hortaçsu, Syverson, & Emre, 2010). E-commerce offers the possibility of accessing markets in a relatively low-risk, low-cost manner (Cloete & Doens, 2008). Several types of e-commerce including business-to-business (B2B), business-to-customer (B2C), and customer-to-customer (C2C) are common on the Internet (Shojaiemehr & Rafsanjani, 2013). E-commerce technologies, particularly those supporting B2B, exchange, include the Internet, e-mail, Electronic Data Interchange (EDI), Electronic Funds Transfer (EFT), and barcodes (Gunasekaran, Marri, McGaughey, & Nebhwani, 2002). Bharadwaj and Soni (2007) indicated that five parts of the e-commerce infrastructure include the flow of information, organizational image, reaction to customer needs, increased sales, and access to new markets.

The rapid development of IT enables an increasing number of consumers to search and book products online first and then to consume them in brick-and-mortar stores (Xiao & Dong, 2015). Existing studies have focused considerable attention on the factors regarding e-commerce adoption (Ghobakhloo, Arias-Aranda, & Benitez-Amado, 2011). The findings are classified into two categories (i.e., the strategic value of certain

IT to top managers and the factors that influence the adoption of IT) (Grandon & Pearson, 2004), emphasizing the importance of the perceived benefits, strategic commitment to e-commerce, external pressures, organizational readiness, information system characteristics, and e-commerce capabilities, among others (Hong & Zhu, 2006).

Several scholars have adopted the process perspective to explore the adoption of e-commerce (Lin, Huang, & Tseng, 2007). Such studies have adopted two alternative perspectives: a factor-oriented perspective (Kirwan, 2009) and a stage-oriented perspective (Rao, Metts, & Monge, 2003). Factor-oriented studies segment the adoption process into several stages, such as the initiation, decision and implementation stages, and investigate the facilitators and inhibitors in each stage (Cui & Pan, 2015). Stage-oriented studies focus on the adoption stages (e.g., presence, transaction portal integration, and enterprise combination) and investigate the features of each stage, such as communication channels and e-commerce functions (Cui & Pan, 2015).

PERSPECTIVES ON ELECTRONIC COMMERCE IN MODERN BUSINESS

This section emphasizes the overview of e-commerce and the importance of e-commerce in modern business.

Overview of Electronic Commerce

E-commerce has made considerable challenges into both large organizations and SMEs (Solaymani, Sohaili, & Yazdinejad, 2012). E-commerce is essential for enterprises because it improves their understanding of customer needs and of the products and services available in the global marketplace (Zhu, 2004). By conducting e-commerce, manufacturers can eliminate the distributors and wholesalers; optimize the internal processes; reduce the production period, inventory, and circulation (Chun, Rhee, Park, & Kim, 2011);

and obtain the higher margins (Laudon & Traver, 2009). Although some manufacturers, such as Dell, General Electric, and Estee Lauder (Yan & Pei, 2010), have achieved tangible improvements in their operational efficiency and revenue generation by adopting e-commerce, some firms have been less successful. Researchers and practitioners have struggled to determine the optimal conditions for the adoption of e-commerce by manufacturers (Oh, Cruickshank, & Anderson, 2009).

Privacy preservation relies on the good faith of individuals and organizations that gather data on the customers in an open electronic environment, such as the Internet (Aimeur, Brassard, Fernandez, & Onana, 2008). For e-commerce to reach its full potential, consumers must have the effective protection when shopping online, and their primary rights must be adequately protected (Dahiyat, 2011). One of the foundations of e-commerce is trust (van Baal, 2015). The importance of trust in building and maintaining the consumer relationships in the online environment is widely accepted in the information system literature (Cheung & Lee, 2006). Trust plays an important role in the formation of dependent relationships represented by online transactions (Li, Pieńkowski, van Moorsel, & Smith, 2012).

Trust has been recognized as a main factor in promoting e-commerce (Son, Narasimhan, & Riggins, 2005). The increasing popularity of open standards in e-commerce demands more research on generalized trust (Qu, Pinsonneault, Tomiuk, Wang, & Liu, 2015). Zaheer and Venkatraman (1994) investigated the electronic linkages in the insurance industry and found that the insurance agencies' trust in the carrier practically increases their business toward the carrier over electronic channels. Hart and Saunders (1998) stated that suppliers' trust in customers increases the suppliers' use of EDI for diverse functions (e.g., inventory advice and order status report). Klein et al. (2007) indicated that mutual trust between trading partners increases IT customization for supply chain coordination and improves the strategic information sharing about capacity planning, production schedule, and marketing strategies.

In e-commerce environments, the reputation of sellers is a crucial issue to the potential buyers in making decisions (Zhang, Wang, & Zhang, 2014). Because of the importance of reputation management, a variety of reputation management models that can be utilized in many e-commerce markets have been developed during the past several years (Zhou, Dresner, & Windle, 2008). For many e-commerce models, reputation management has become one of the critical factors that restrict their development (Rice, 2012). In the electronic business (e-business) market, merchants' reputation can help consumers reduce information asymmetry and increase their acceptance of e-commerce (Ruohomaa & Kutvonen, 2005).

Importance of Electronic Commerce in Modern Business

The increasing popularity of e-commerce is apparent (Spremic & Hlupic, 2007). Firms should employ e-commerce to compete in the global competition of the 21st century (Wu & Fang, 2011). The Internet and related technologies have vastly expanded the variety of products that can be profitably promoted and sold by online retailers (Hinz & Eckert, 2010). To make business organizations grow in the rapidly changing environment, managers must understand the knowledge of e-commerce and keep paying attention to e-commerce development (Shiau & Dwivedi, 2013). Business organizations need their information system and knowledge management to offer the e-commerce solutions (Sharma, Gupta, & Wickramasinghe, 2006).

Internet has a tremendous impact on how companies do business (Frank & Lange, 2007). The use of the Internet as the additional sales channel offers the traditional retailers opportunities to reach the expanded markets while improving the efficiency of their operations (Xia & Zhang, 2010). Firm ability in conducting transaction and collaboration with business partners over electronic channels (e.g., B2B e-commerce) can have the strategic impact on firm performance (Wong, Lai, & Cheng, 2011). B2C e-commerce fundamentally alters the

structure of those transaction costs relevant to the IT structure and the brand architecture in global marketing (Treiblmaier & Strebing, 2008).

The adoption of e-commerce technologies enables SMEs to access larger markets without expanding their physical presence (Quaddus & Hofmeyer, 2007). E-commerce technologies have a strong appeal to SME retailers as e-commerce can make geographic locations and time irrelevant (Premkumar & Roberts, 1999). Due to the impacts of e-commerce technologies on SMEs, e-commerce technologies have become the significant entities of attention within business and entrepreneurship literature (Peltier, Zhao, & Schibrowsky, 2012).

FUTURE RESEARCH DIRECTIONS

The classification of the extensive literature in the domains of e-commerce will provide the potential opportunities for future research. Mobile commerce is the business that is conducted on the Internet through the use of mobile phones or other handheld electronic devices (Kasemsap, 2016c). With the rapid growth and innovations witnessed in the mobile industry, the communication field has greatly transformed. Business communication has greatly eased with the advent of mobile communication. Business communication enables colleagues or business partners to discuss business without necessarily holding meetings. The Internet-enabled phones give access to social media platforms, such as Facebook and Twitter, from anywhere across the globe.

The use of social media improves effective communication, increases thought leadership, enhances sales, and provides better customer service (Kasemsap, 2016d). Professional and business applications of social media platforms can enhance business performance toward reaching strategic goals in the digital age (Kasemsap,

2017a). Web mining is the application of data mining techniques to discover the interesting patterns from web data in order to better serve the needs of web-based multifaceted applications (Kasemsap, 2017b). An examination of linkages among e-commerce, mobile commerce, mobile communication, business communication, social media, and web mining in modern business would seem to be viable for future research efforts.

CONCLUSION

This article highlighted the overview of e-commerce and the importance of e-commerce in modern business. E-commerce consists of the purchasing and selling of products or services through electronic systems, such as computer networks and the Internet. In this modern world of technology, e-commerce becomes a very significant option for many businesses as there are lots of companies that are interested in developing their online stores. E-commerce becomes one of the preferred ways of online shopping as they enjoy their online because of its easiness and convenience.

Because of its convenience, consumers can save their lots of time as well as money by searching their products easily and making purchasing online. The performance of e-commerce is critical for modern organizations that seek to serve suppliers and customers, increase business performance, strengthen competitiveness, and gain sustainable competitive advantage in modern business. Thus, it is urgent for modern organizations to promote their e-commerce and develop a strategic plan to regularly investigate their practical improvements toward satisfying customer requirements. Applying e-commerce has the potential to enhance organizational performance and gain sustainable competitive advantage in modern business.

REFERENCES

- Agarwal, J., & Wu, T. (2015). Factors influencing growth potential of e-commerce in emerging economies: An institution-based N-OLI framework and research propositions. *Thunderbird International Business Review*, 57(3), 197–215. doi:10.1002/tie.21694
- Aimeur, E., Brassard, G., Fernandez, J. M., & Onana, F. S. M. (2008). ALAMBIC: A privacy-preserving recommender system for electronic commerce. *International Journal of Information Security*, 7(5), 307–334. doi:10.1007/s10207-007-0049-3
- Ajzen, I. (1991). The theory of planned behaviour. *Organizational Behavior and Human Decision Processes*, 50(2), 179–211. doi:10.1016/0749-5978(91)90020-T
- Ajzen, I., & Fishbein, M. (1975). *Understanding attitudes and predicting social behavior*. Englewood Cliffs, NJ: Prentice Hall.
- Alm, J., & Melnik, M. I. (2012). Cross-border shopping and state use tax liabilities: Evidence from eBay transactions. *Public Budgeting & Finance*, 32(1), 5–35. doi:10.1111/j.1540-5850.2011.00999.x
- Anesbury, Z., Nenycz-Thiel, M., Dawes, J., & Kennedy, R. (2016). How do shoppers behave online? An observational study of online grocery shopping. *Journal of Consumer Behaviour*, 15(3), 261–270. doi:10.1002/cb.1566
- Bharadwaj, P. N., & Soni, R. G. (2007). E-commerce usage and perception of e-commerce issues among small firms: Results and implications from an empirical study. *Journal of Small Business Management*, 45(4), 501–521. doi:10.1111/j.1540-627X.2007.00225.x
- Chang, Y., & Hu, H. (2012). Security analysis of an electronic commerce protocol using casper/FDR2. *Wuhan University Journal of Natural Sciences*, 17(6), 499–503. doi:10.1007/s11859-012-0877-5
- Chen, L. C., Yu, T. J., & Hsieh, C. J. (2013). KeyGraph-based chance discovery for exploring the development of e-commerce topics. *Scientometrics*, 95(1), 257–275. doi:10.1007/s11192-012-0826-2
- Cheung, C. M. K., & Lee, M. K. O. (2006). Understanding consumer trust in Internet shopping: A multidisciplinary approach. *Journal of the American Society for Information Science and Technology*, 57(4), 479–492. doi:10.1002/asi.20312
- Chun, S. H., Rhee, B. D., Park, S. Y., & Kim, J. C. (2011). Emerging dual channel system and manufacturer's direct retail channel strategy. *International Review of Economics & Finance*, 20(4), 812–825. doi:10.1016/j.iref.2011.02.006
- Cloete, E., & Doens, M. (2008). B2B e-marketplace adoption in South African agriculture. *Information Technology for Development*, 14(3), 184–196. doi:10.1002/itdj.20105
- Cui, M., & Pan, S. L. (2015). Developing focal capabilities for e-commerce adoption: A resource orchestration perspective. *Information & Management*, 52(2), 200–209. doi:10.1016/j.im.2014.08.006
- Dahiyat, E. A. R. (2011). Consumer protection in electronic commerce: Some remarks on the Jordanian electronic transactions law. *Journal of Consumer Policy*, 34(4), 423–436. doi:10.1007/s10603-011-9170-9
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *Management Information Systems Quarterly*, 13(3), 319–340. doi:10.2307/249008
- Frank, U., & Lange, C. (2007). E-MEMO: A method to support the development of customized electronic commerce systems. *Information Systems and e-Business Management*, 5(2), 93–116.

- Ghobakhloo, M., Arias-Aranda, D., & Benitez-Amado, J. (2011). Adoption of e-commerce applications in SMEs. *Industrial Management & Data Systems*, 111(8), 1238–1269. doi:10.1108/02635571111170785
- Gibbs, J. L., & Kraemer, K. L. (2004). A cross-country investigation of the determinants of scope of e-commerce use: An institutional approach. *Electronic Markets*, 14(2), 124–137. doi:10.1080/10196780410001675077
- Gokmen, A. (2012). Virtual business operations, e-commerce & its significance and the case of Turkey: Current situation and its potential. *Electronic Commerce Research*, 12(1), 31–51. doi:10.1007/s10660-011-9084-2
- Goldmanis, M., Hortaçsu, A., Syverson, C., & Emre, Ö. (2010). E-commerce and the market structure of retail industries. *The Economic Journal*, 120(545), 651–682. doi:10.1111/j.1468-0297.2009.02310.x
- Grandon, E. E., & Pearson, J. M. (2004). Electronic commerce adoption: An empirical study of small and medium US business. *Information & Management*, 42(1), 197–216. doi:10.1016/j.im.2003.12.010
- Gunasekaran, A., Marri, H. B., McGaughey, R. E., & Nebhwani, M. D. (2002). E-commerce and its impact on operations management. *International Journal of Production Economics*, 75(1), 185–197. doi:10.1016/S0925-5273(01)00191-8
- Hart, P., & Saunders, C. (1998). Emerging electronic partnerships: Antecedents and dimensions of EDI use from the supplier's perspective. *Journal of Management Information Systems*, 14(4), 87–111. doi:10.1080/07421222.1998.11518187
- Hinz, O., & Eckert, J. (2010). The impact of search and recommendation systems on sales in electronic commerce. *Business & Information Systems Engineering*, 2(2), 67–77. doi:10.1007/s12599-010-0092-x
- Holsapple, C. W., & Singh, M. (2000). Toward a unified view of electronic commerce, electronic business, and collaborative commerce: A knowledge management approach. *Knowledge and Process Management*, 7(3), 151–164. doi:10.1002/1099-1441(200007/09)7:3<151::AID-KPM83>3.0.CO;2-U
- Hong, W. Y., & Zhu, K. (2006). Migrating to Internet-based e-commerce: Factors affecting e-commerce adoption and migration at the firm level. *Information & Management*, 43(2), 204–221. doi:10.1016/j.im.2005.06.003
- Huang, T. C., Lee, T. J., & Lee, K. H. (2009). Innovative e-commerce model for food tourism products. *International Journal of Tourism Research*, 11(6), 595–600. doi:10.1002/jtr.731
- Jung, I. Y., Cui, W., Eom, H., & Yeom, H. Y. (2014). Distributed electronic commerce cluster for small enterprise. *Cluster Computing*, 17(3), 681–689. doi:10.1007/s10586-013-0334-z
- Kasemsap, K. (2016a). Advocating electronic business and electronic commerce in the global marketplace. In S. Dixit & A. Sinha (Eds.), *E-retailing challenges and opportunities in the global marketplace* (pp. 1–24). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-9921-2.ch001
- Kasemsap, K. (2016b). Implementing electronic commerce in global marketing. In I. Lee (Ed.), *Encyclopedia of e-commerce development, implementation, and management* (pp. 591–602). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-9787-4.ch043
- Kasemsap, K. (2016c). Investigating the roles of mobile commerce and mobile payment in global business. In S. Madan & J. Arora (Eds.), *Securing transactions and payment systems for m-commerce* (pp. 1–23). Hershey, PA: IGI Global. doi:10.4018/978-1-5225-0236-4.ch001

- Kasemsap, K. (2016d). Utilizing social media in modern business. In I. Lee (Ed.), *Encyclopedia of e-commerce development, implementation, and management* (pp. 2171–2182). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-9787-4.ch156
- Kasemsap, K. (2017a). Professional and business applications of social media platforms. In V. Benson, R. Tuninga, & G. Saridakis (Eds.), *Analyzing the strategic role of social networking in firm growth and productivity* (pp. 427–450). Hershey, PA: IGI Global. doi:10.4018/978-1-5225-0559-4.ch021
- Kasemsap, K. (2017b). Mastering web mining and information retrieval in the digital age. In A. Kumar (Ed.), *Web usage mining techniques and applications across industries* (pp. 1–28). Hershey, PA: IGI Global. doi:10.4018/978-1-5225-0613-3.ch001
- Kirwan, O. (2009). An action research case study of the facilitators and inhibitors of e-commerce adoption. *Journal of International Business Research*, 2(2), 48–56.
- Klein, R., Rai, A., & Straub, D. W. (2007). Competitive and cooperative positioning in supply chain logistics relationships. *Decision Sciences*, 38(4), 611–646. doi:10.1111/j.1540-5915.2007.00172.x
- Kurnia, S., Choudrie, J., Mahbubur, R. M., & Alzougool, B. (2015). E-commerce technology adoption: A Malaysian grocery SME retail sector study. *Journal of Business Research*, 68(9), 1906–1918. doi:10.1016/j.jbusres.2014.12.010
- Laudon, K. C., & Traver, C. G. (2009). *E-commerce: Business, technology, society*. Upper Saddle River, NJ: Prentice Hall.
- Li, F., Pieńkowski, D., van Moorsel, A., & Smith, C. (2012). A holistic framework for trust in online transactions. *International Journal of Management Reviews*, 14(1), 85–103. doi:10.1111/j.1468-2370.2011.00311.x
- Lin, C., Huang, Y. A., & Tseng, S. W. (2007). A study of planning and implementation stages in electronic commerce adoption and evaluation: The case of Australian SMEs. *Contemporary Management Research*, 3(1), 83–100. doi:10.7903/cmr.44
- Lin, T. C., Cheng, H. K., Wang, F. S., & Chang, K. J. (2012). A study of online auction sellers' intention to switch platform: The case of Yahoo!Kimo versus Ruten_eBay. *Decision Sciences*, 43(2), 241–272. doi:10.1111/j.1540-5915.2011.00347.x
- Lu, Q., & Liu, N. (2015). Effects of e-commerce channel entry in a two-echelon supply chain: A comparative analysis of single- and dual-channel distribution systems. *International Journal of Production Economics*, 165, 100–111. doi:10.1016/j.ijpe.2015.03.001
- Oh, K. Y., Cruickshank, D., & Anderson, A. R. (2009). The adoption of e-trade innovations by Korean small and medium sized firms. *Technovation*, 29(2), 110–121. doi:10.1016/j.technovation.2008.08.001
- Peltier, J. W., Zhao, Y., & Schibrowsky, J. A. (2012). Technology adoption by small businesses: An exploratory study of the interrelationships of owner and environmental factors. *International Small Business Journal*, 30(4), 406–431. doi:10.1177/0266242610365512
- Pires, G. D., Stanton, J., & Salavrakos, I. D. (2010). The interaction of foreign direct investment with electronic commerce in less developed countries. *Forum for Social Economics*, 39(2), 127–143. doi:10.1007/s12143-008-9021-x
- Premkumar, G., & Roberts, M. (1999). Adoption of new information technologies in rural small businesses. *Omega: The International Journal of Management Science*, 27(4), 467–484. doi:10.1016/S0305-0483(98)00071-1
- Qu, W. G., Pinsonneault, A., Tomiuk, D., Wang, S., & Liu, Y. (2015). The impacts of social trust on open and closed B2B e-commerce: A Europe-based study. *Information & Management*, 52(2), 151–159. doi:10.1016/j.im.2014.07.002

- Quaddus, M., & Hofmeyer, G. (2007). An investigation into the factors influencing the adoption of B2B trading exchanges in small business. *European Journal of Information Systems*, 16(3), 202–215. doi:10.1057/palgrave.ejis.3000671
- Rahayu, R., & Day, J. (2015). Determinant factors of e-commerce adoption by SMEs in developing country: Evidence from Indonesia. *Procedia: Social and Behavioral Sciences*, 195, 142–150. doi:10.1016/j.sbspro.2015.06.423
- Rao, S. S., Metts, G., & Monge, C. A. M. (2003). Electronic commerce development in small and medium sized enterprises: A stage model and its implications. *Business Process Management Journal*, 9(1), 11–32. doi:10.1108/14637150310461378
- Rice, S. C. (2012). Reputation and uncertainty in online markets: An experimental study. *Information Systems Research*, 23(2), 436–452. doi:10.1287/isre.1110.0362
- Rogers, E. M. (1995). *Diffusion of innovation*. New York, NY: The Free Press.
- Ruohomaa, S., & Kutvonen, L. (2005). Trust management survey. *Trust Management*, 3477, 77–92. doi:10.1007/11429760_6
- Sharma, S. K., Gupta, J. N. D., & Wickramasinghe, N. (2006). A framework for designing the enterprise-wide e-commerce portal for evolving organizations. *Electronic Commerce Research*, 6(2), 141–154. doi:10.1007/s10660-006-1180-3
- Shiau, W. L., & Dwivedi, Y. K. (2013). Citation and co-citation analysis to identify core and emerging knowledge in electronic commerce research. *Scientometrics*, 94(3), 1317–1337. doi:10.1007/s11192-012-0807-5
- Shojaiemehr, B., & Rafsanjani, M. K. (2013). A multi-agent based model for collective purchasing in electronic commerce. *International Journal of Machine Learning and Cybernetics*, 4(4), 409–417. doi:10.1007/s13042-012-0098-4
- Solaymani, S., Sohaili, K., & Yazdinejad, E. A. (2012). Adoption and use of e-commerce in SMEs. *Electronic Commerce Research*, 12(3), 249–263. doi:10.1007/s10660-012-9096-6
- Son, J. Y., Narasimhan, S., & Riggins, F. J. (2005). Effects of relational factors and channel climate on EDI usage in the customer–supplier relationship. *Journal of Management Information Systems*, 22(1), 321–353.
- Soopramanien, D. (2011). Conflicting attitudes and scepticism towards online shopping: The role of experience. *International Journal of Consumer Studies*, 35(3), 338–347. doi:10.1111/j.1470-6431.2010.00945.x
- Spremic, M., & Hlupic, V. (2007). Development of e-commerce in Croatia: A survey. *Information Technology for Development*, 13(4), 391–409. doi:10.1002/itdj.20068
- Treiblmaier, H., & Strebingner, A. (2008). The effect of e-commerce on the integration of IT structure and brand architecture. *Information Systems Journal*, 18(5), 479–498. doi:10.1111/j.1365-2575.2007.00288.x
- van Baal, S. (2015). Not all seals are equal: An experimental investigation of the effect of third-party seals on purchase probability in electronic commerce. *Electronic Commerce Research*, 15(2), 177–206. doi:10.1007/s10660-015-9180-9
- Wong, C. W. Y., Lai, K. H., & Cheng, T. C. E. (2011). Value of information integration to supply chain management: Roles of internal and external contingencies. *Journal of Management Information Systems*, 28(3), 161–199. doi:10.2753/MIS0742-1222280305
- Wu, C. H., & Fang, W. C. (2011). Combining the fuzzy analytic hierarchy process and the fuzzy Delphi method for developing critical competences of electronic commerce professional managers. *Quality & Quantity*, 45(4), 751–768. doi:10.1007/s11135-010-9425-6

Xia, Y., & Zhang, G. P. (2010). The impact of the online channel on retailers' performances: An empirical evaluation. *Decision Sciences*, 41(3), 517–546. doi:10.1111/j.1540-5915.2010.00279.x

Xiao, S., & Dong, M. (2015). Hidden semi-Markov model-based reputation management system for online to offline (O2O) e-commerce markets. *Decision Support Systems*, 77, 87–99. doi:10.1016/j.dss.2015.05.013

Yan, R. L., & Pei, Z. (2010). Information asymmetry, pricing strategy and firm's performance in the retailer-multi-channel manufacturer supply chain. *Journal of Business Research*, 64(4), 377–384. doi:10.1016/j.jbusres.2010.11.006

Yapar, B. K., Bayrakdar, S., & Yapar, M. (2015). The role of taxation problems on the development of e-commerce. *Procedia: Social and Behavioral Sciences*, 195, 642–648. doi:10.1016/j.sbspro.2015.06.145

Zaheer, A., & Venkatraman, N. (1994). Determinants of electronic integration in the insurance industry: An empirical test. *Management Science*, 40(5), 549–566. doi:10.1287/mnsc.40.5.549

Zhang, H., Wang, Y., & Zhang, X. (2014). The approaches to contextual transaction trust computation in e-commerce environments. *Security and Communication Networks*, 7(9), 1331–1351. doi:10.1002/sec.839

Zhou, M., Dresner, M., & Windle, R. J. (2008). Online reputation systems: Design and strategic practices. *Decision Support Systems*, 44(4), 785–797. doi:10.1016/j.dss.2007.10.001

Zhu, K. (2004). The complementarity of information technology infrastructure and ecommerce capability: A resource-based assessment of their business value. *Journal of Management Information Systems*, 21(1), 167–202.

ADDITIONAL READING

Agudo-Peregrina, A. F., Pascual-Miguel, F. J., & Chaparro-Peláez, J. (2014). It's never the same: The role of homogeneity in online services. *Service Business*, 8(3), 453–464. doi:10.1007/s11628-014-0249-7

Banerjee, P. K., & Ma, L. C. (2012). Routinisation of B2B e-commerce by small firms: A process perspective. *Information Systems Frontiers*, 14(5), 1033–1046. doi:10.1007/s10796-011-9329-6

Bhatnagar, A., Nikolaeva, R., & Ghose, S. (2016). Online market entry: The motivations for imitation across retailer types. *Managerial and Decision Economics*, 37(3), 151–166. doi:10.1002/mde.2708

Capece, G., Calabrese, A., Di Pillo, F., Costa, R., & Crisciotti, V. (2013). The impact of national culture on e-commerce acceptance: The Italian case. *Knowledge and Process Management*, 20(2), 102–112. doi:10.1002/kpm.1413

Chen, M. Y., & Teng, C. I. (2013). A comprehensive model of the effects of online store image on purchase intention in an e-commerce environment. *Electronic Commerce Research*, 13(1), 1–23. doi:10.1007/s10660-013-9104-5

Dakhli, S., & Strauss, R. P. (2013). Should sales taxes be imposed on e-commerce? *Journal of Public Economic Theory*, 15(6), 803–828.

Fauska, P., Kryvinska, N., & Strauss, C. (2014). Agile management of complex goods & services bundles for B2B e-commerce by global narrow-specialized companies. *Global Journal of Flexible Systems Management*, 15(1), 5–23. doi:10.1007/s40171-013-0054-5

Georgiadis, C. K., Stiakakis, E., & Ravindran, A. R. (2013). Editorial for the special issue: Digital economy and e-commerce technology. *Operations Research*, 13(1), 1–4. doi:10.1007/s12351-011-0122-6

- Guo, F., Cao, Y., Ding, Y., Liu, W., & Zhang, X. (2015). A multimodal measurement method of users' emotional experiences shopping online. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 25(5), 585–598. doi:10.1002/hfm.20577
- Hamad, H., Elbeltagi, I., Jones, P., & El-Gohary, H. (2015). Antecedents of B2B e-commerce adoption and its effect on competitive advantage in manufacturing SMEs. *Strategic Change*, 24(5), 405–428. doi:10.1002/jsc.2019
- Humair, F., Humair, L., Kuhn, F., & Kueffer, C. (2015). E-commerce trade in invasive plants. *Conservation Biology*, 29(6), 1658–1665. doi:10.1111/cobi.12579 PMID:26249172
- Otim, S., & Grover, V. (2012). Resolving uncertainty and creating value from the exercise of e-commerce investment options. *Information Systems Journal*, 22(4), 261–287. doi:10.1111/j.1365-2575.2011.00385.x
- Papastavrou, S., Chrysanthi, P. K., & Samaras, G. (2014). Performance vs. freshness in web database applications. *World Wide Web (Bussum)*, 17(5), 969–995. doi:10.1007/s11280-013-0262-0
- Peng, L., & Lai, L. (2014). A service innovation evaluation framework for tourism e-commerce in China based on BP neural network. *Electronic Markets*, 24(1), 37–46. doi:10.1007/s12525-013-0148-0
- Piva, F., & Dahab, R. (2013). E-commerce of digital items and the problem of item validation: Introducing the concept of reversible degradation. *Applicable Algebra in Engineering Communication and Computing*, 24(3/4), 277–308. doi:10.1007/s00200-013-0197-9
- Propheter, G. (2012). Political determinants of state e-commerce sales tax policy. *Politics & Policy*, 40(4), 657–679. doi:10.1111/j.1747-1346.2012.00369.x
- Rezaei, R., Chiew, T. K., & Lee, S. P. (2014). A review on e-business interoperability frameworks. *Journal of Systems and Software*, 93, 199–216. doi:10.1016/j.jss.2014.02.004
- Rosaci, D., & Sarné, G. M. L. (2012). A multi-agent recommender system for supporting device adaptivity in e-commerce. *Journal of Intelligent Information Systems*, 38(2), 393–418. doi:10.1007/s10844-011-0160-9
- Shojaiemehr, B., & Rafsanjani, M. K. (2014). A fuzzy system approach to multilateral automated negotiation in B2C e-commerce. *Neural Computing & Applications*, 25(2), 367–377. doi:10.1007/s00521-013-1491-y
- Sigman, B. P., & Boston, B. J. (2013). Digital discernment: An e-commerce web site evaluation tool. *Decision Sciences Journal of Innovative Education*, 11(1), 29–46. doi:10.1111/j.1540-4609.2012.00367.x
- Sila, I. (2013). Factors affecting the adoption of B2B e-commerce technologies. *Electronic Commerce Research*, 13(2), 199–236. doi:10.1007/s10660-013-9110-7
- Tian, Z., Zhang, Z., & Guan, X. (2013). A new evolution model for B2C e-commerce market. *Information Technology & Management*, 14(3), 205–215. doi:10.1007/s10799-013-0159-8
- Valvi, A. C., & Fragkos, K. C. (2012). Critical review of the e-loyalty literature: A purchase-centred framework. *Electronic Commerce Research*, 12(3), 331–378. doi:10.1007/s10660-012-9097-5
- Walter, F. E., Battiston, S., Yildirim, M., & Schweitzer, F. (2012). Moving recommender systems from on-line commerce to retail stores. *Information Systems and e-Business Management*, 10(3), 367–393.
- Wang, J., & Ngoasong, M. Z. (2012). The internationalization process of Chinese SMEs: Does globalizing wholesale markets play a role? *Strategic Change*, 21(3/4), 143–157. doi:10.1002/jsc.1900

Yassine, A., Shirehjini, A. A. N., Shirmohammadi, S., & Tran, T. T. (2012). Knowledge-empowered agent information system for privacy payoff in eCommerce. *Knowledge and Information Systems*, 32(2), 445–473. doi:10.1007/s10115-011-0415-3

Zhang, H., Lu, Y., Shi, X., Tang, Z., & Zhao, Z. (2012). Mood and social presence on consumer purchase behaviour in C2C e-commerce in Chinese culture. *Electronic Markets*, 22(3), 143–154. doi:10.1007/s12525-012-0097-z

KEY TERMS AND DEFINITIONS

Business: The activity of buying and selling commodities, products, or services.

Competitive Advantage: A superiority gained by an organization when it can provide the same value as its competitors but at a lower price, or can

charge the higher prices by providing the greater value through differentiation.

Electronic Commerce: The activities that relate to the buying and selling of products and services over the Internet.

Information Technology: The development, installation, and implementation of computer systems and applications.

Internet: The worldwide computer network that provides information on many subjects and enables users to exchange the messages.

Marketing: The strategic functions involve in identifying and appealing to the particular groups of consumers, often including activities, such as advertising, branding, pricing, and sales.

Technology: The branch of knowledge that deals with the creation and use of technical methods and their interrelation with life, society, and the environment.

Improving Competencies for the Courier Service Industry in Malaysia

Hoo Yee Hui

Universiti Tunku Abdul Rahman, Malaysia

Yudi Fernando

Universiti Malaysia Pahang, Malaysia

INTRODUCTION

Begun around the year 2000, globalization is being powered by individuals and smaller organizations. Emergence of e-commerce enables the individuals and smaller organizations to do business in a more effectiveness and efficiency way (Bowersox et al., 2013). They increasingly rely on courier services that can complement and contribute to their logistics operation particularly in Asia market due to its high population.

It is crucial for a business to reassess their competencies to operate in a competitive global working environment which is highly volatile and rapidly change. Subsequently, the supply chain is facing a paradigm change in the e-business environment. A company supply chain in an e-business environment can be very complicated typically in Asia Market due to Asia Market is made up of many countries which is vary in culture, religion, political system, legal systems, language and stage of economic development. The use of internet and advancement in technologies enable e-commerce to quickly emerge in courier service industry, linking enterprises, associations, governments and individuals together. The usage of internet has changed the industry strategy in conducting business.

The aim of this chapter is, therefore, to explore the literature on improving competencies for the courier service industry in Malaysia. The competencies of courier services as the ability of third party (3PL) logistics operator to smooth

the supply chain in e-commerce. The focus of the study is not how courier services competencies is generated within an e-commerce environment in Malaysia, but rather on how such competencies manifest itself, and able to ultimately influence the competencies for the Courier Service Industry in Malaysia. Additionally, this study is intended to help shape the future of courier services as a professional entity in the business environment.

BACKGROUND

Globalization affects the way of doing business today. Many industries experienced deregulation, deconsolidation, re-engineering, downsizing, and organization restructuring to meet the need for efficiency in a fast-paced operation with unpredictable supply pattern (Ling, 2014; Sheffi, 2012; Grant et al., 2006).

E-commerce has opened up an utterly new market for players in logistics. In the year 2012, total population in Malaysia has reached 28.7 million and the internet users are 17.7million accounted 65.8% of the population. Among the users of internet in Malaysia, there are 30% of the people who are e-shoppers that engage in purchasing goods and services from various entities on the web (Malaysian Communications and Multimedia Commission, 2012).

Besides that, business of courier services has been stimulated by the development of e-commerce in terms of the growth in parcels volume and in

home delivery. There is a tremendous increase in sales orders delivered by e-commerce companies that accomplish via courier companies, which has resulted in the rapid growth in this niche segment as well as the increase in the struggle for a greater market share in this sector (Hoo et al. 2014).

Sheffi (2012) defined logistics clusters as one of the strategies to support existing local industries, these clusters include third party logistics services providers (3PLs), transportation carriers, warehousing companies, freight forwarders, retailers, distributors, manufacturer; and the operations of companies for whom logistics is a large part of their cost. Courier service is operationally distinguished from the ordinary mail services. It plays a role as third party logistics services providers (3PLs) in Malaysia logistics clusters.

The logistics clusters have been recognized as one of the sectors attain global competitiveness in the Ninth Malaysia Plan and Third Industrial Plan (IMP3) (Khalid, 2009). Unfortunately, the Chairman of Asia Logistics Council, Sri Abdul Rahman Mamat declared that logistics sector is the weakest amongst the four pillars of international trade namely finance, commerce, insurance and logistics in Malaysia (Bernama, 2013).

PROBLEM STATEMENT

The courier services industry plays a key role in e-commerce activities by meeting the delivery fulfillment in the right time, to the right person, at the right place. In a study conducted by Malaysian Communication and Multimedia Commission (MCMC), result shown that the customer satisfaction index towards courier service in 4th quarter of year 2010 is 3.74 from a scale of 1 (total dissatisfaction) to 5 (total satisfaction). According to the Industry Performance report for the year 2012 provided by Malaysian Communication and Multimedia Commission, the quality goal achieved is only 69% which is lower than the expected quality goal, more than 80%.

It is a challenge for courier service providers when facing online order soars, fulfilling customers' order become more complex. Because of the recent increase and focus on e-commerce, without some movement toward the establishment of professional principles for practitioners in this field, there will continue to be a lack of awareness and acknowledgement about the specialized and unique skills and knowledge inherent in the practice of courier service industry.

RESEARCH QUESTIONS

This study seeks to uncover the professional skills and competencies for individuals who work in courier services industry. The main research question is: What are the competencies of courier services in Malaysia? The following two questions will be examined: (1) What are the required professional skills for courier services practitioners? (2) What are the most critical competencies factors in courier services?

OBJECTIVE OF THE STUDY

The learning purposes were listed as below:

1. To identify the competencies that marks a professional in courier services industry.
2. To investigate the critical competitive factors for the courier services industry.

LITERATURE REVIEW

Logistics secured a prominent role as it provides the backbone to facilitate the international trade. With the increase of the global competition and rapid progress of the IT technology, the emergence of e-commerce has spurred the demand for courier service in order to facilitate the delivery of merchandise purchased online.

Courier service is operationally distinguished from the ordinary mail services by some features including speed, security, tracking, signature, specialization and individualization of express services, and swift delivery times, which are optional for daily mail services (Hoo et al., 2014). Besides that, courier service plays an important role in the success of e-tailing business. It offered one stop solution to e-retailer by providing services from warehouse management to goods delivery.

In the consumers' perspective, they look for courier service providers that are able to offer same day delivery. For example, Amazon.com order fulfillment goal is to ship orders on the same day they are placed by the customers (Grant et al. 2006). However, in the e-tailing market, customer prefers free shipping while the retailers tend not to choose faster shipping option because it is expensive for them. E-tailing is directly derived from the word "retailing". Chandra & Sunitha (2012) defined e-tailing as retailing conducted online, over the internet. A survey shown that package shipped per order in e-tailing market increased from 2.3 per order in year 2014 to 2.6 per order in year 2015. (Stevens, L., 2016). This phenomena result a longer shipping time due to longer transit time for courier service provider to reach its receiver.

According to Malaysian Communications and Multimedia Commission (MCMC), there are total 112 licensed courier companies in Malaysia in year 2010. The number of active courier companies only accounted 35 and 28 of them are members of AMEC. AMEC members account for about 85% of the entire value of the Malaysian express market. The number of branches in courier services industry in Malaysia increased from 240 branches in the year 2010 to 266 branches in the year 2013 (Postal and Courier services, 2013) indicated that the demand on this service had increase significantly.

Unlike other transportation service provider in Malaysia's Logistics Cluster which is under contract to local authorities, Land Public Transport Commission (SPAD) and Road Transport Department (JPJ), the courier service industry in

Malaysia is governed by the Malaysian Communications and Multimedia Commission (MCMC) under Law of Malaysia Act 741. Under the Proposed Industry Development Plan for 2015 under MCMC, research on e-commerce impact on Postal and Courier Industry served as one of the events to be launched recently (AMEC, 2015).

Under the Law of Malaysia Act 741, Postal Service Act 2012 is used to regulate the freight rate, license provision, as well as customer protection. Although the rates are being regulated, the act does not standardize the services offered by the courier services providers in Malaysia. This had made courier service industry a dynamic business entities in Malaysia logistics cluster. Wide range of logistics services can be offered by courier services including but not limited to mailing room services, break bulk, customized logistics solutions, warehousing, and shipping services through its fourth party logistics service provider.

In another hand, the launching of National Broadband Initiatives (NBI) in Malaysia encourages its citizen to use e-commerce services for their daily economies activities. Many e-commerce service providers such as *Groupon.com*, *Ensogo.com*, *11 street.com*, *Lazada.com* in Malaysia had started to react on this phenomena and offered logistics related job to ensure the smooth operation of its e-commerce activities (Jobstreet, 2015). Also, the job responsibility of logistics manager in *11 street.com* is to manage the logistics and deliver in Market place model "dropship" (Jobstreet, 2015).

E-commerce is a key driver in the growth of courier service as more and more people buying and selling through the internet (Alminnourliza and Osman, 2012). The development of e-commerce had break the sales limited in time and space, and plays an important role in the construction of information system, logistics, capital flow and information flow (Chen, Q. & Zhang, N., 2014).

Over the past five years (2010-2014), Malaysia's e-commerce market size has seen 31% increase in compound annual growth rate (CAGR) and emerging as the third country that recorded

the highest percentage of Internet users (67%) after Singapore and Brunei in Southeast Asia. The emerging of e-commerce causing the growth in logistic industry, especially the courier segment has seen exponential growth over the past one year. For instance, courier service contributed 60% of POS Malaysia's total earnings in year 2015, as compared to 41% in year 2014 (Retail Asia, 2016).

Viswanadham, Kumar and Gaonkar (2002, p. 1503) defined three modes of e-commerce order fulfillment which are (1) dedicated in-house fulfillment, (2) outsourcing the fulfillment to a third party which offered one stop solution to e-commerce companies as well as maintains the inventory in the warehouse, and (3) supplier drop shipping the items to the customer whereby the dropshipper orders goods from its own supplier and manages its own warehouse and transportation system. Unlike the case in outsourcing, the goods in the warehouse are under the ownership of the dropshipper.

Development of e-commerce causes modifications to logistics chain management strategies and methods of flow control which cover both the process of handling the online and offline sales channel. Initially a niche service, courier services became an accessible service to wider categories of clients coupled with the increase infrastructure of developing countries and rural areas (Georgescu, 2012).

Logistics infrastructure is required to sustain a country's competitive advantage in long term. Serbat & Harun (2011) analyzed that the logistics competencies are used to obtain the competitive advantage by countries that are integrated to global economy. They found that logistics infrastructure is the most effective ways in discriminating between high and low competitive countries which include all the modes of transportation and Information and Communication Technologies (ICT).

Employee skill, experience and knowledge are the intangible assets in the organization. To increase the capability of employees and ensure quality service provide to customer in today's competitive market, job performance management

became an important task for organization. One of the tools to evaluate job performance is using job appraisal. Beatson et al. (2008, p. 170) describe job performance as the following:

When staff enjoys their jobs they are more likely to value the firm's customers and want to ensure that these customers are happy with the service they receive. The way that these staff can do this is by delivering good service and fulfilling customers' needs.

Hence, job performance presented in delivering good service and fulfilling customer needs.

In the study conducted by Gupta and Sharma (2009) on the factors influence job satisfaction level in public sector organizations, various factors influence job satisfaction had been identified and classified into two categories which are environmental factors and physical factors. The following factors are found in environmental factors:

1. Leadership and planning in the organization,
2. Individual participation,
3. Recognition and rewards,
4. Team work and cooperation,
5. Training program.

Result shown that training program was the most important factor as it enabled employees to keep themselves well-equipped with organization change.

Researchers had identified three competitive factors for supply chain strategy which are time, cost and quality under competitive profile (Harrison et al., 2014). Time refers to the order lead time or response time. Cost refers to the total cost of supply and underpins the price changed. Quality refers to the delivery reliability.

Shang and Chin (2012) discovered the importance of maritime service attributes such as transit time, on-time pick up and delivery, reliability of delivery are categorized under reliability of shippers' requirement. Not only that, Jaiswal (2008) believe that reliability is the most vital aspect of

service quality attributes especially for freight forwarders involving in the service sector. Hence, reliability is considered to be a crucial factor for the courier service competencies improving.

Ling (2014) explained further on firm's competitive priorities is determined by its four long-term structural decisions: facility, capacity, technology, and vertical integration, as well as by its four infrastructural decisions: workforce, quality, production planning and control, and organization.

Information technology is one of the important factors for courier service to engage with customer especially dealing with e-commerce transaction. Information technology was defined as the new technology to improve the efficiency of logistics operation (Huang, et al., 2013). Investment and adoption of information technology has become a crucial factor to supply chain efficiency and effectiveness in globalization era.

However, new information and technologies must be adopted by companies in order to integrate the supply chain process. These technologies are including information system such as warehouse management system (WMS), integrated business management systems and transportation management systems (TMS). Never the less, the communication technologies such as radio identification tags (RFID) and global positioning system (GPS) also helps the courier service provider to shorter the transit time and track consignment effectively (Roy, j., 2010). A study conducted by Mohamed (2008) found that using latest technologies to automated tracking of consignments as well as delivery status remain challenges.

Integration of e-commerce and courier service is a key online business strategy which enables the e-retailers to deliver convenient and seamless online shopping experience that customers expect and deserve. Courier services providers plays an important role in the supply chain of e-commerce as they are the last "node" in the e-commerce supply chain. The focus on competencies in courier services industry will spurred innovation, attracting more talent and more money that contributed to the growth of the cluster.

METHODOLOGY

The study applies Delphi methodology to identify the competencies in courier services industry. Delphi research methodology was selected for this study because of its ability to gain consensus about the required professional skills and knowledge of courier service practitioners.

The Delphi method is defined as a method for structuring a group communication process in allowing a group of individuals to deal with a complex problem (Cole, 2008, Linstone & Turoff, 1975). This method was used to identify and assess the required professional competencies for successful performance of courier services professionals.

FUTURE RESEARCH DIRECTIONS

Although there are some increases in the research of the subject of this study within Malaysia context, however, the research in improving competencies for the courier service industry in Malaysia have not reached an optimum level as compared to the Asian counterparts. This study contributes to e-business' and courier services' owners by unveiling a better understandings in Malaysia context toward the competencies that would give the e-commerce and courier service industry a huge boost.

Throughout the literature of this study, it is proposed that courier service providers should pay more attention to those key factors possess a positive correlation that not only enhance the growth of e-commerce industry but also boost the development of courier service sector. Subsequently, success in e-commerce and courier service can be promisingly achieved since the business environment is changing constantly and customer demand is evolving.

CONCLUSION

The evolution of e-commerce will fascinate people to shop online and acquire their desirable goods from other parts of the worlds. This phenomenon will arouse the demand for courier service. Handling of courier service from the delivery process of merchandise purchased online to the arrival of goods at the door steps of online shoppers needed to be expedited.

REFERENCES

- Abdulla, S. A. A. (2009). *Strategies for Competitive Advantage and Supply Chain Management: Synergy Opportunities* (Doctoral dissertation). Texas A&M University.
- Association of Malaysian Express Carriers. (2015). *Industry News Digest, Vol. 67*. Retrieved November 19, 2015, from <http://www.amec.org.my>
- Beatson, A., Lings, I., & Gudergan, S. P. (2008). Service staff attitudes, organisational practices and performance drivers. *Journal of Management & Organization, 14*(02), 168–179. doi:10.1017/S1833367200003370
- Bowersox, D. J., Closs, D. J., & Cooper, M. B. (2002). *Supply chain logistics management* (Vol. 2). New York, NY: McGraw-Hill.
- Burmaoglu, S., & Sesen, H. (2011). Analyzing the Dependency Between National Logistics Performance and Competitiveness: Which Logistics Competence is Core for National Strategy? *Journal of Competitiveness, 3*(4).
- Chandra, P. S., & Sunitha, G. (2012). E-tailing-the mantra of modern retailer's success. *Researchers World, 3*(2), 42.
- Chen, Q., & Zhang, N. (2015). Does E-Commerce Provide a Sustained Competitive Advantage? An Investigation of Survival and Sustainability in Growth-Oriented Enterprises. *Sustainability, 7*(2), 1411–1428. doi:10.3390/su7021411
- Chung, S., & Shan, L. (2012). Customer relationship management and firm performance: An empirical study of freight forwarder services. *Journal of Marine Science and Technology, 20*(1), 64–72.
- Cole, M. (2008). *A critical assessment of professional skills and knowledge in supplier diversity: A Delphi study* (Doctoral dissertation). Antioch University.
- Georgescu, S. D. (2013). The influence of it and e-commerce on the evolution on courier services sector. *Annals of the University of Craiova, 1*.
- Gupta, M., & Sharma, P. (2009). Job satisfaction level among employees: a case study of Jammu region, J&K. *IUP Journal of Management Research, 8*(5), 17.
- Harrison, A., & Van Hoek, R. I. (2008). *Logistics management and strategy: competing through the supply chain*. Pearson Education.
- Hoo, Y. H., Andreen, L. A. Q., Ong, C. P., & Jasmine, S. (2014). Emergence of E-commerce: Consequences to the Courier Service Industry. *Proceedings of the 3rd International Conference on Business and Communication*.
- Huang, S. T., Bulut, E., Duru, O., & Yoshida, S. (2013, August). Service Quality Assessment Based on Customer Satisfaction in International Freight Forwarding Industry: An Empirical Study in East Asia. In *International Forum on Shipping, Ports and Airports (IFSPA) 2013: Trade, Supply Chain Activities and Transport: Contemporary Logistics and Maritime Issues*.
- Jaiswal, A. K. (2008). Customer satisfaction and service quality measurement in Indian call centres. *Managing Service Quality: An International Journal, 18*(4), 405–416. doi:10.1108/09604520810885635
- Jobstreet. (2015). Retrieved September 25, 2015, from <http://www.jobstreet.com.my>

Lee, H. L., & Lee, C. Y. (Eds.). (2007). *Building supply chain excellence in emerging economies* (Vol. 98). Springer Science & Business Media. doi:10.1007/978-0-387-38429-0_16

Li, L. (2014). *Managing Supply Chain and Logistics: Competitive Strategy for a Sustainable Future*. World Scientific Books.

Malaysian Communications and Multimedia Commission. (2010). *Postal Courier Services: Selected Facts and Figures 2H, 2010*. Retrieved November 5, 2015, from <http://www.skmm.gov.my>

Malaysian Communications and Multimedia Commission. (2012). *Hand Phone Users Survey 2012*. Malaysian Communications and Multimedia Commission. Retrieved November 5, 2015, from <http://www.skmm.gov.my>

Muhammad, N. M. N., Isa, F. M., & Kifli, B. C. (2009). Positioning Malaysia as halal-hub: Integration role of supply chain strategy and halal assurance system. *Asian Social Science*, 5(7), 44.

Retail Asia. (2016). *Malaysia is Poised for E-commerce Growth through Better Mobility*. Retrieved March 5, 2016, from <http://www.retailnews.asia>

Roa, S. (2009). *Managing Supply Chain Risks through Outsourcing: Three Studies* (Doctoral dissertation). University of Kentucky.

Roy, J. (2011). Logistics and the Competitiveness of Canadian Supply Chains. *Global Value Chains: Impacts and Implications, Trade Policy Research, Minister of Public Works and Government Services Canada. Section, 4*, 313–333.

Sheffi, Y. (2012). *Logistics clusters: delivering value and driving growth*. MIT Press.

Stevens, L. (2016). *Holiday Online Orders Taking Longer, Study Says*. Retrieved March 2, 2016, from <http://www.wsj.com>

Viswanadham, N., Kumar, V. V., & Gaonkar, R. (2002). Optimal Retail Order Fulfilment Strategies. *Proceedings of the 4th Asian Control Conference*.

ADDITIONAL READING

Crew, M. A., & Kleindorfer, P. R. (Eds.). (2012). *Current directions in postal reform* (Vol. 35). Springer Science & Business Media.

Ghiani, G., Manni, E., & Quaranta, A. (2010). Shift scheduling problem in same-day courier industry. *Transportation Science*, 44(1), 116–124. doi:10.1287/trsc.1090.0293

Ho, J. S. Y., Teik, D. O. L., Tiffany, F., Kok, L. F., & Teh, T. Y. (2012, August). Logistic service quality among courier services in Malaysia. In *International Conference on Economics, Business Innovation* (Vol. 38, pp. 113–117).

Kassim, N., & Asiah Abdullah, N. (2010). The effect of perceived service quality dimensions on customer satisfaction, trust, and loyalty in e-commerce settings: A cross cultural analysis. *Asia Pacific Journal of Marketing and Logistics*, 22(3), 351–371. doi:10.1108/13555851011062269

Malik, S. U. (2012). Customer satisfaction, perceived service quality and mediating role of perceived value. *International Journal of Marketing Studies*, 4(1). doi:10.5539/ijms.v4n1p68

Mota Pedrosa, A. (2012). Customer integration during innovation development: An exploratory study in the logistics service industry. *Creativity and Innovation Management*, 21(3), 263–276. doi:10.1111/j.1467-8691.2012.00648.x

Singh, L. P., Singh, S., & Bhardwaj, A. (2011). Role of logistics and transportation in green supply chain management: An exploratory study of courier service industry in India. *International Journal of Advanced Engineering Technology*, 2(1), 260–269.

Wisniewski, S. C. (1997). *Multinational enterprises in the courier service industry: Aspects of employment and working conditions in selected enterprises* (No. 324860). International Labour Organization.

KEY TERMS AND DEFINITIONS

Courier Service: A third party logistics services provider whereby its primary services are to receive shipments and to deliver shipments in a short transit timeframe.

Delphi Method: One of the research method usually applied by researchers on a complex problem with limited research respondents.

Dropshipper: A sender who orders goods from its own supplier and owned a fully control in its warehouse and transportation system.

E-Shoppers: Consumers who obtain goods and services from various entities without geographical limitation through internet platform.

E-Tailing: Retail business which is conducted online.

Logistics Clusters: A logistics strategic to improve and promote the existing local industries.

Logistics Competencies: The ability of a logistics service provider to do a job in an effectiveness and efficiency way.

E

New Advances in E-Commerce

Khaled Ahmed Nagaty

The British University in Egypt, Egypt

INTRODUCTION

Advances in computer architecture and communication technology have led to introduction of global, dynamic and intelligent e-business models and new tools for doing e-commerce. These models use artificial intelligence techniques and web2 technologies to perform online transactions between businesses and customers. As the number of clients over the Internet grows exponentially enhanced intelligent e-business models are needed to efficiently respond to clients with an optimum selection for their requests. Agent based e-business models use intelligent agents, mobile agents, evolutionary agents, rational agents, data-mining agents to develop agent-based e-commerce models. Recommender systems provide intelligent assistance to online customers is CBR-based or collaborative filtering based e-business models. In B2C models there are a large numbers of buyers and sellers where buyers specify their preferences and sellers specify their products and selling prices. The major role of AI techniques in B2C e-commerce is to come up with the best match between a buyer and seller. Examples of e-commerce systems in B2C using AI techniques are product selection and recommendation, negotiation and auction systems. In B2B the major role of AI techniques is supply chain management to help clients taking the right decisions. Modern tools used by intelligent E-commerce models include: algorithmic trading, decentralized autonomous corporations (DAC), crypto-currency systems, smart contracts, Internet of Things, cloud computing and big data analytics.

BACKGROUND

Intelligent Techniques

Intelligent techniques in E-commerce include:

Intelligent Agents

They are personalized software that function independently and use machine-learning algorithms to analyze data, adapt to their environment in order to achieve the goals on behalf of users. Main agents can use sub-agents to work under their supervision to achieve smaller tasks to help main agents achieve their pre-written goals. These agents can surf the Internet gathering information, searching e-market places looking for products and services satisfying users' preferences. With the help of intelligent agents customers can enter e-negotiations and take part in e-auctions. According to (Nwana, 1996), the intelligent agents important features include:

1. **Autonomy:** Agents achieve their goals independently and act proactively without human intervention.
2. **Cooperation:** Agents communicate with each other and negotiate on certain issues.
3. **Learning:** Agents are able to learn, react or interact with their environment and with other agents.

Intelligent agents allow sellers to track demand and market shares, perform knowledge mining, negotiate and learn from collaboration with buyer agents. Seller agents can gain money by selling

services, they may create children that may either survive or die depending on whether they can get enough business. Broker agents, provider agents and merchant agents sell products and services required by sellers while comparison shopping agents are required for buyers to comparison shop (Sinmao, 1999). These types of agents get the required information from retailers' web sites using semantic web.

Types of intelligent agents include:

1. **Rational Agents:** They are autonomous computer programs that are capable of a goal directed behavior (Russell et al, 2003). The main goal of rational agents is to autonomously select the optimal outcome from possible preferences in a given situation. To achieve this goal the rational agents collect information about their environment which may be obtained from past experiences, they take actions considering obligations and duties then they evaluate the outcome to decide whether the final decision is the optimum one or not. If the final result is not optimum the rational agents adjust their behavior to improve results in the future.
2. **Evolutionary Agents:** They are agents that gain new knowledge from self-reasoning and through knowledge exchange with other agents. Evolutionary agents share information and knowledge with each other within a multi-agent environment regardless of their structures or objectives. The main task for each evolutionary agent is to fulfill one or more of its own goals. Agents are selected for growth from the current population to go into the mating pool. This mating pool is the basis for the next population. The growth mechanism is based on learning mechanism, fuzzy reasoning and ontological reasoning. The most fitting agent is selected based on a self-organizing map.
3. **Data Mining Agents:** They are agents designed to find specific data types and identifying patterns among this data. They generate alerts to organizations if any altera-

tion in the trend paradigm is discovered so effective strategies can be implemented to gain from this alteration or reduce its damage. Data mining agents perform one or more specific tasks, for example they can be programmed to monitor stock prices for a specific range of companies and alerting their users if significant aberrations from historical trends are discovered.

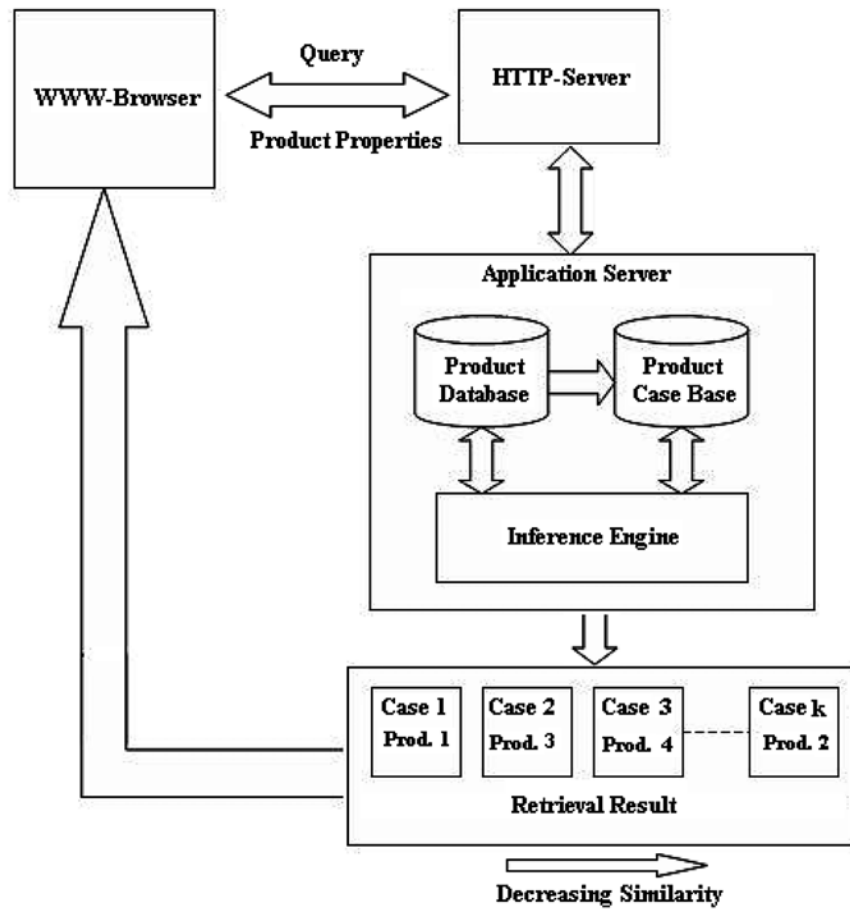
CBR Technique (Prasad, 1995)

Content based reasoning is a problem solving technique that retrieves solutions for previous similar cases from the case-base and adapt them to solve new problems. Each case has two parts: specification part and solution part. The specifications part describes the symptoms of the problem while the solution part describes the causes of this fault. CBR plays an important role in e-commerce applications for product recommendations by using context-sensitive information related to products to help users take the right decision quickly (Kumar et al., 2005). Content-based techniques make use of textual annotations describing products obtained from e-commerce websites (Abbattista, et al., 2002).

CBR works as follows:

1. Take customer preferences. Customers preferences may vague or rate their preferences ranging from 'must have' to 'not important'.
2. Retrieve products with similar offers from case base and submit them to the customer.
3. Customer may accept a similar offer or quit the process if he is no longer interested.
4. Customer may modify his preferences and go to step (ii). Similarity or disparity measures between two products are to identify similar products. Weighted Euclidean distance measure is used to retrieve the most similar product (Wettschereck and Aha, 1995). As an example a CBR-based recommender system in the sports domain is presented in (Prasad and Clementi 2002). Figure 1 shows a CBR e-commerce system.

Figure 1. A CBR e-commerce system
Prasad, 1995.

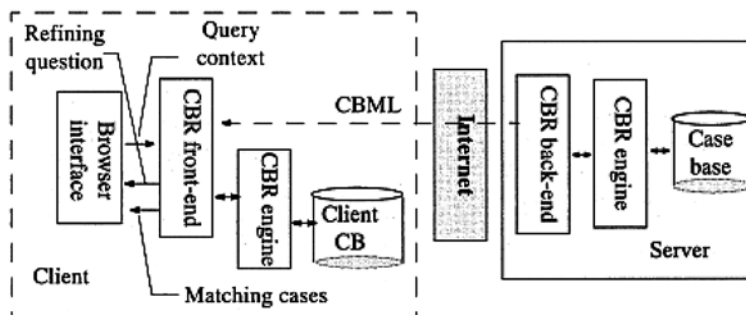


Recently CBR has a significant contribution in developing web-based intelligent systems. Figure 2 is the general architecture of Web-based CBR systems.

ACF Technique (Hayes et al, 2001)

An automated collaborative filtering that is similar to case base reasoning (CBR). While the CBR

Figure 2. General architecture of web-based CBR system
Zhaohao Sun, 2004.



produces a solution the ACF produces recommended components for the target user profile. It is a method to make automatic predictions (filtering) about a user's preference by collecting information about many users' preferences (collaboration). In ACF based recommender web sites, each ACF customer profile contains records about the historical customer's consumptions of items with either explicit or implicit ratings. An ACF case is an incomplete one – it is essentially one row in the utility matrix which is usually quite sparsely populated. ACF main goal is case completion which is an incremental enhancement of the user profile based on user's feedback. When a new user access the E-commerce website, the system uses this information to match the user's preferences with the database of preferences, in order to retrieve a user profile with a preference class closely matches the new user's preferences and extract completion information for the new case profile which is then submitted to the new user. Products that satisfy new user's preferences are then recommended to the new user (Sarwar et al., 2002). ACF has an indefinite iterative recommendation cycle where items recommended for inclusion in a user's profile are determined by the feedback input from this user. The recommendation cycle includes three phases: the *retrieve* phase where similar profiles are retrieved, the *reuse* phase and the *revise* phase.

GBR Approach (Burke 2000)

A goal-based reasoning approach used to search for products similar to those already known to the user. There are always some standard goals for each product domain, where a similarity metric between two products is measured regarding these goals. Different users might have different rankings of goals or different goals altogether. For each goal, a similarity metric between pairs of products defines how closely those two products with respect to that goal. Interaction between goals should be considered and combination of metrics is used to achieve ranking of products.

Supply Chain Management (SCM) is a key element for successful B2B e-commerce. With the exponential growth of Internet users e-commerce enterprises redesign their business models as they are moving to e-commerce environment (Fenstermacher and Zeng, 1999). Most of AI-based SCM problem solvers are agent based where each agent is responsible for one or more activities of the SCM (Prasad, 2003).

Modern Tools

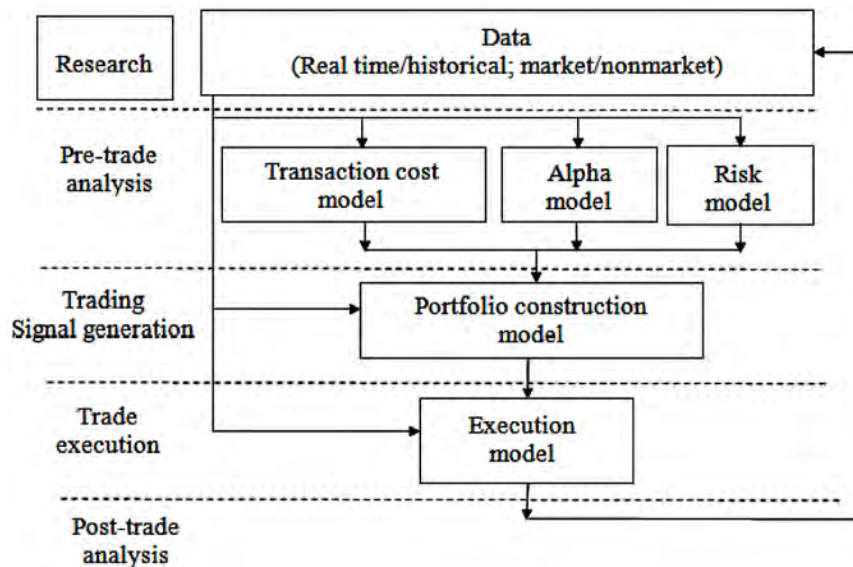
Intelligent E-commerce models based on the above intelligent techniques are adapting to the emerged modern tools namely smart phones and cloud computing. However, tools such as big data analytics and social media are used to provide E-commerce intelligent models with insights required to improve the performance of these models. In this section, the most prominent modern tools and their influence on intelligent E-commerce models are discussed.

Algorithmic Trading

Also known as algo-trading, black box trading or automated trading where fast computers and complex mathematical models are used to automate one or more of the trading process stages. They benefit from statistical patterns within or across financial markets to capture odd prices, detect and exploit rival strategies to optimally execute a defined set of instructions for placing trades to generate profits either by costs saving, client commissions or proprietary trading at a very high speed and frequency that are beyond the capabilities of human minds (Giuseppe Nuti et al., 2011). Figure 3 shows the principal components of an algo-trading system and the order at which they process.

The algorithmic trading system is composed of the pre-trade analysis component, trading signal generation and trade execution components. The pre-trade analysis encompasses any system that uses financial data and involves intelligent

Figure 3. Algorithmic trading components
As adapted from Giuseppe Nuti et al., 2011.



techniques to scan news or tweets on Twitter to forecast future price movement or volatility. The pre-trade analysis includes three mathematical models which are:

1. The alpha model which predicts the financial instruments future behavior to trade.
2. The risk model which evaluates risk levels associated with the financial instruments.
3. The transaction cost model which calculates the potential costs associated with trading the financial instruments.

The trading signal generation creates signals to traders by applying technical analysis to price charts. Trading signal generation consists of the portfolio construction model which decides what portfolio of financial instruments should be owned going forward and in what quantities. The input to this model is the results obtained from the alpha model, risk model and transaction cost model. The trade execution model executes the trades and place orders in one or more exchanges, making several decisions with constraints on trading costs and trading duration.

Mobile Internet Market

(Andreas Kaplan, 2012) Defines mobile marketing as “any marketing activity conducted through a ubiquitous network to which consumers are constantly connected using a personal mobile device”. Mobile marketing is a cheaper, very effective and fast growing method to promote products and services to consumers (Ververidis & Polyzos, 2002). Based on the location of a mobile user which can be obtained using radio-frequency identification (RFID) tags, Bluetooth or GPS information advertisements can be sent to the mobile in real time based on his/her proximity from a store (Li & Du, 2012). In other words, companies with a strong mobile marketing presence and have an optimized website on the Internet will attract more visitors and more likely these visitors will be potential buyers of services or products.

Decentralized Autonomous Corporations

DAO organizations are decentralized organizations that can make decisions without people engagement by using smart contracts. Intelligent

techniques use these rules and protocols to help reduce transaction costs associated with actual legal contracts while providing security using cryptographic authentication to facilitate negotiation and enforces contracts performance.

Cryptocurrency

Cryptocoins are they only means to deal with DAO corporations. No central authority or banks can control cryptocurrencies systems. Bitcoins network is considered the first decentralized digital cryptocurrency system. It is a peer-to-peer (P2P) digital money system where the unit of currency is bitcoin. Electronic payments are implemented by generating transactions which transfer bitcoins directly between participants. Each participant transfers the bitcoin to the next participant by digitally signing a hash of the previous transaction and the public key of the next participant and adding these to the end of the coin. A payee can verify the signatures to verify the chain of ownership. A Proof-of-Work (PoW) system prevents people from double spending the bitcoins and verifies transactions without a central organization. Digital signatures corresponding to sending addresses are used to authenticate transactions, identify users' identities and control creation of new bitcoins. All bitcoin transactions are added to a unique public ledger called block chain. A block chain provides participants with a distributed trustworthy storage and contains every transaction done on the bitcoins P2P network and shared by all participants of this network. Each participant can own a copy of this block chain thus allowing him to monitor any other participant which prevents attacker from falsifying transactions or account balances (Satoshi Nakamoto,2008).

Smart Contracts

Small programs that replace legalese with computer programs that contain pre-written rules and logic in the form of "IF... THEN" statements and behave exactly as expected. Smart contracts

contain digital signatures and time stamps that prevent any potential contention between parties. When two parties digitally signed and deployed a smart contract on the network no unilateral party can cancel this transaction.

Internet of Things (IoT)

This technology forms a worldwide network for e-commerce where all digital devices, objects and human beings are interconnected to share information which reflects objects manufacturers, transportation and consumption and other details of people's life (Li H,2013). IoT opens a new exciting era in the areas of marketing and advertising. It provides companies and organizations with a powerful tool for marketing their products where marketers and advertisers have the opportunity to analyze customers' behaviors and respond to the needs of their audience with appropriate messages at the appropriate time on the appropriate devices. Figure 4 shows the architecture of an IoT e-business model which applies the concept of decentralized autonomous organizations (DAO). It is composed of four modules namely: entities and commodities which are considered the infrastructure of the IoT e-business model and the basic operating mode and transactions mode (Yu Zhang, 2016).

1. **Entities:** IoT integrates an extremely large number of heterogeneous entities such as DAOs, hardware entities and software entities that intended to improve efficiency and bring interconnectivity to all aspects of modern economy. Hardware entities include sensing devices or actuating devices. Software entities include applications that use sensed data, software agents that are responsible for entity's conceptualization and functionality or services. Software entities can search or purchase IoT commodities according to certain conditions.
2. **Commodities:** They are smart properties that can be traded or loaned via block chain using smart contracts. Smart properties in-

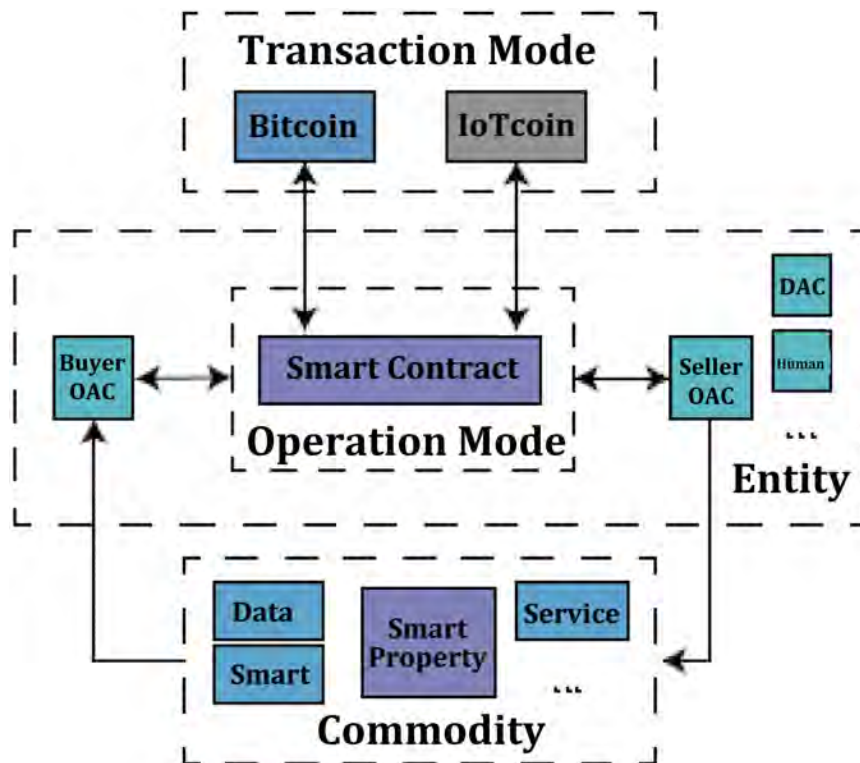
clude physical properties such as cars and houses and non-physical properties such as loans. Their ownership is controlled via block chain and smart contracts.

3. **Transaction Mode:** Bitcoin is adopted as the unit currency and IoTcoin as commodity exchange certificates and compatible with bitcoin. With IoT coin and smart contracts users can exchange the paid data or smart property or make deals with other users on the P2P without the involvement of a third party.
4. **Operation Mode:** The design and signing of the smart contract is accomplished in this stage. When a smart contract is executed Bitcoins and IoT commodities transfers are performed according to the signed smart contract (Yu Zhang, 2016).

Social Media

Facebook and Twitter are popular social media having great impact on the growth of E-commerce market which encourages strategic developers to consider social media in their strategic plans. Key information such as age, geographic location, language, popular phrases, keywords, behaviors, opinions and interests of users can be obtained from social media. Engagement with these users and delivering a needed support to them can influence their perception on a particular product. Based on a study involved 6000 social media users conducted in April 2013, 40% of the users bought their commodities after sharing them on social media websites, 71% are ready to buy products based on the feedback coming from social media while 74% depend on social media to make their

Figure 4. The architecture of IoT e-business
As adapted from Yu Zhang, 2016.



decisions. It was also found that Facebook influences 30.8% of social media users purchasing habits, while LinkedIn and YouTube influence 27% of social media site users respectively (Seave, 2013). These statistical results have encouraged E-commerce enterprises to rely on social media to gain competitive advantage and business values, influence customers perception, increase customer loyalty, satisfaction and retention, most importantly create brand awareness and build reputation then finally improve sales and revenues (Hea et al., 2013) (Melville et al., 2009).

Big Data Analytics

With the ever growing business data, E-commerce enterprises are now considering their data as big assets. Big data has the three characteristics which are volume, velocity and variety (3V's). Volume refers to the size of data in Terabytes, Exabytes or Petabytes while velocity refers to the speed with which data is captured and finally variety refers to the heterogeneous nature of data i.e. structured and unstructured datasets (Uyoyo Zino Edosio, 2014). Big data analytics hold many promises for providing business insights, analyzing consumer behavior as they can have deep insight into customers' needs, predict customers' behaviors, and for adapting of advertisements to fit consumers' different tastes. It can provide a suitable metric to measure the effectiveness in meeting customers' needs. Social media data is considered big data as it holds the 3Vs characteristics of big data. Text mining and semantic analysis are the two common techniques to analyze social media data (Tan et al., 2013). Predictive analysis is another big data tool that is used to identify patterns and learn from historical data to allow E-commerce vendors to predict customers' behavior and build an effective future marketing campaign (Shmueli & Koppius, 2011). Predictive analysis is also used by recommender systems to predict the preference or the rating of a customer towards a specific product while price management strategies use predictive analysis to set future prices.

Cloud Computing

Cloud computing has a significant effect on the performance of intelligent E-commerce models as it provides scalable and flexible services to the E-commerce enterprises on demand. It enables them to perform their computation task on the public cloud using pay-as-you-go style rather than purchasing hardware and software items. Pay-as-you-go pricing scheme is very flexible as it helps enterprises to pay for their consumed resources only, eliminate the costs for hardware and software maintenance, costs of upgrading and the wages for IT personnel. All these help E-commerce enterprises to decrease the cost of system building and maintenance and to maximize the utilization of their resources to get the maximum return on their investment (Danping Wang, 2013). E-commerce enterprises hosting their applications and data on cloud data centers enjoy both high availability i.e. 24x7 and reliability. In addition, cloud computing providers guarantee the scalability of applications running on their servers thus giving the clients the opportunity to scale up or scale down their computing resources if this is needed. Scalability supports e-commerce enterprises to continue offering their services to their clients even if there is sudden growing of demand. Another important advantage provided by cloud computing is storage replication where data is stored in multiple geographic locations to protect it from natural disasters. The huge storage capabilities of the cloud data centers are used to store the explosive growing customers' data which will become a growing problem if these enterprises are using local servers.

INTELLIGENT AND MODERN MODELS

Issues, Controversies, Problems

One of the major problems on the Internet is lack of advising where no human advisors are available to

advice customers online to identify what products better fit their preferences. Some researchers argue that using AI in e-commerce would result in a market being overtaken by machines. Also, there is a belief that decisions taken by human intelligence is much better than decisions taken by artificial intelligence techniques. However, some researches argue that artificial intelligence could be helpful to guide people in taking their right decisions but not replacing the human intelligence. As modern life becomes more complex people have no time to perform business in traditional ways which requires more flexible tools to facilitate people's life in doing e-business. Researchers argue that although modern tools have significantly flourished e-business but issues of adaptation, privacy and security are of main concern.

SOLUTIONS AND RECOMMENDATIONS

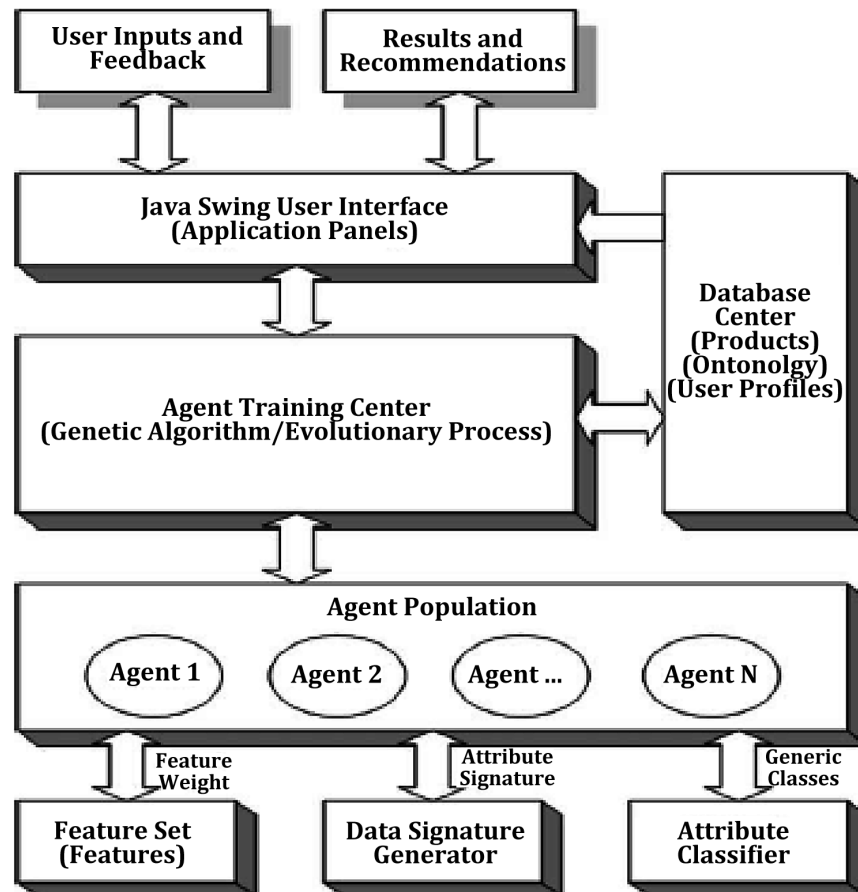
The major advantage of using artificial intelligent techniques in e-commerce business models is advising the customers on the items they want to buy through the Internet. Advising on the internet is important because there are no real persons to advice customers on the internet. This advice helps online customers to browse a large number of products (Prasad, 2003).

1. **Personalized Brokering Agents:** A personalized product-brokering agent is one of the B2C intelligent applications that depend on evolutionary agents to build an efficient agent-based ontology through collaboration and exchange of information with other product-brokering agents. Evolutionary agents in personalized product-brokering application use genetic algorithms and evaluation function to track user's preferences for a specific product. A user profile is required for the product-brokering agent to work efficiently. At the beginning a new product-brokering agent does not have enough information on

the products to recommend any of them to the user. Some keywords about products' features are required for product-brokering agents to start searching for products. To collect this initial information the product-brokering agent relies on other sources such as web crawlers called InfoSpider in (G. Pant, 2002). As an example, in the Handy product-brokering intelligent application an initial population of brokering-agents are created and evolved using genetic algorithms (Steven Guan, 2002). Each agent has a unique name to differentiate it from other agents. Newly created brokering agents acquire their prior knowledge about products using the Internet. Figure 5 shows an evolutionary intelligent agents system.

2. **E-Auctions:** Excellent e-business tool to help B2C, B2B and C2C manage complex and challenge negotiation in real time. E-auctions take place in e-markets where sellers present their goods on the auction website and bidders submit their bids for the product within a certain period. E-auctions allow sellers from any region in the world to offer their bids to the buyer. E-auctions significantly reduce the time required for sending material, product specifications from the buyers, negotiating with the sellers and finally receive quotations. Two main types of e-auctions are available: forward e-auctions and reverse e-auctions. In forward e-auctions, several buyers offer their bids to a seller see Figure 6. The goal of forward e-auction is to push the price upwards to get the highest offered price. In reverse e-auctions, a single buyer let potential sellers to know about his/her intention to buy items they sell through sending emails or making posts on social media see Figure 7. The goal of reverse e-auction is to push the price downwards to get the lowest price to pay.
3. **Recommender Systems:** An e-business tool to predict user responses to items.

Figure 5. Evolutionary intelligent agents system
As adapted from Sheng-Wei Guan et al, 2005.



Recommender systems are either based on the CBR or ACF technology. Items recommended to users by CBR recommender systems are based on the properties of these items where similarity of items is determined by measuring the similarity in their properties. Items recommended to the users by ACF recommender systems are those items preferred by similar users. The two main entities in recommender systems are:

- Users,
- Items.

The relationship between users and entities is a preference relationship where users have preferences for certain items. By using the users' feedback on products the recommender system

Figure 6. Forward auction

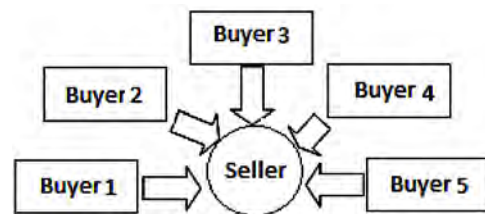
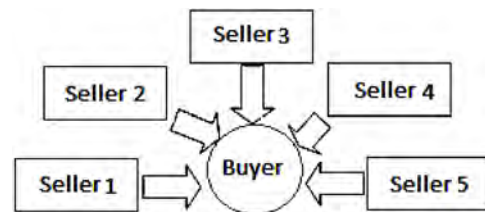


Figure 7. Reverse auction



can adapt automatically and its performance is enhanced comprehensively (Weihong and Yi, 2006). A utility sparse matrix represents the degree of preference of a certain user and to a specific item. Degrees of preferences come from an ordered set of integers for example integers from 1 to 5. Figure 8 shows $M \times N$ utility matrix where M the number of users, N the number of items and d_{ij} is the degree by which user i prefers item j .

Since the utility matrix is sparse, meaning that most of its entries are unknown, the recommender system is required to predict the values for these entries.

4. **E-Procurement:** An e-business tool that facilitates the stages of the procurement process: searching, sourcing, and negotiation with suppliers, ordering, receipt, and post-purchase review (Croom & Brandon-Jones, 2004). E-procurement allows organizations to cut their costs, control budgets, adds transparency to their spending, auditing and approvals. There are three types of e-procurement systems: buyer e-Procurement systems, seller E-procurement systems and online intermediaries (Koorn, Smith and Mueller, 2001).
5. **E-Payment:** Facilitates transfer of money between suppliers and vendors through the Internet. Figure 9 shows a secured e-payment system where data is transmitted on a secured communication channels. Types of e-payments include:
 - a. **Cards:** These include credit cards, debit cards and prepaid cards.

- b. **Online Payments:** Where a customer pays for a product or a service through the Internet.
 - c. **Mobile Payments:** Payments can be made by either sending SMS messages, transmit a PIN number or use WAP to make online payments.
 - d. **Automatic Teller Machines (ATM):** These are fixed stations equipped with communication channels, keyboards, screen monitors to allow clients to make transactions or access information.
 - e. **P2P Payments:** Which allow a person to pay another person using account or prepaid card. A good example for this mechanism is PayPal that was recently purchased by Ebay.
6. **Amazon EC2 (Elastic Compute Cloud):** A Web-based service with practically unlimited number of virtual machines. EC2 allows business subscribers to execute their applications in the Amazon.com computing environment. New York Times and Eli Lilly & Co. have benefitted from using EC2. The former transformed its TIF data to PDF data with a cost of 3000 USD while the second decreased the cost of data analysis to 89 USD which usually spent one million USD before (Danping Wang, 2013). Figure 10 shows the cloud computing E-commerce industry chain. With this structure the E-commerce enterprise will be directly served by the cloud service provider rather than the IT firms group as in traditional E-commerce industry chain.

Figure 8. The utility matrix M users with N items

	Item 1	Item 2	Item 3	Item 4		Item N
User 1	d_{11}	d_{12}	d_{13}	d_{14}		d_{1N}
User 2	d_{21}	d_{22}	d_{23}	d_{24}		d_{2N}
User 3	d_{31}	d_{32}	d_{33}	d_{34}		d_{3N}
User 4	d_{41}	d_{42}	d_{43}	d_{44}		d_{4N}
.....						
User M	d_{M1}	d_{M2}	d_{M3}	d_{M4}		d_{MN}

Figure 9. An e-payment system
As adapted from www.epayments.com.

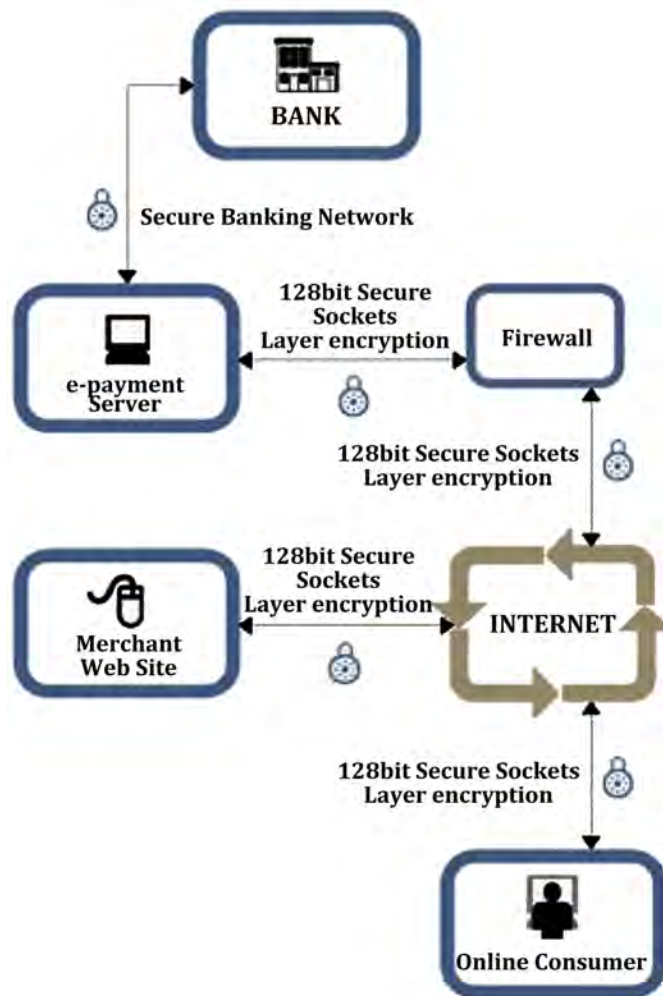
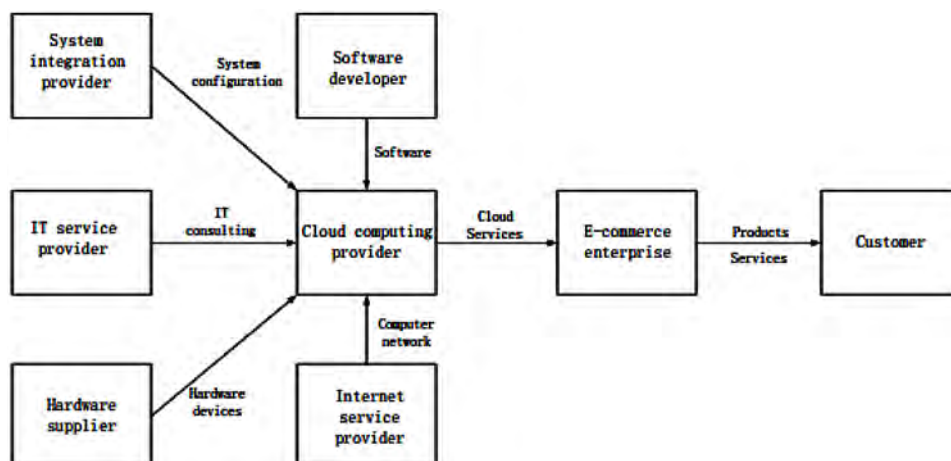


Figure 10. E-commerce industry chain based on cloud computing
As adapted from Danping Wang, 2013.



FUTURE RESEARCH DIRECTIONS

Internet of Things (IoT), cloud computing, big data analytics, social media and mobile apps are the tools going to shape e-commerce in the future. IoT e-business models and their applications in e-business is a promising area of research. New mobile search apps are developing very rapidly and consequently mobile search on the Internet is expanding very fast but the real challenge is how to adapt e-business strategies and models to the mobile environment. Converting unstructured data captured from social media such as tweets and videos to make meaningful insights that benefit E-commerce enterprises is a big challenge. Cloud computing will be the future platform for all real time applications and thus adapting e-business models to the cloud computing environment will be very promising for e-commerce enterprises seeking high availability, reliability and scalability for their online services. Users' privacy protection especially in the public cloud is still an important area of research. Protection of personal information and online transactions of customers will help E-commerce to grow and flourish.

CONCLUSION

In this chapter, the most common intelligent techniques applied in e-commerce models such as intelligent agents, CBR techniques and collaborative filtering technique have been discussed. Their implementation in the most common E-commerce models namely e-brokers, e-auctions, e-procurements, e-payments and recommender systems have been presented. With the advances in communications technology, smart phones technology and distributed computing algorithms new modern tools have been introduced to flourish E-commerce models. Modern tools which include algorithmic trading algorithms, social media, mobile marketing, big data analytics and cloud computing have been presented. Their influences on e-commerce models have also been discussed.

REFERENCES

- Abbattista, F., Degemmis, M., Fanizzi, N., Licchelli, O., Lops, P., Semeraro, G., & Zambetta, F. (2002). Learning User Profiles for Content-Based Filtering in e-Commerce. Atti del Workshop AI*IA su Apprendimento Automatico: Metodi e Applicazioni.
- Burke, R. (2000). Knowledge-Based Recommender Systems. Encyclopedia of Library and Information Science.
- Croom, S., & Brandon-Jones, A. (2004). E-Procurement: Key issues in e-Procurement implementation and operation in the public sector. 13th International Purchasing & Supply Education & Research Association (IPSERA) Conference, Catania, Italy.
- Fenstermacher, K. D., & Zeng, D. (1999). Know Your Supply Chain. *Proceedings of the AAAI-00 Workshop on Knowledge-Based Electronic Markets*.
- Guan, S., Seng, C., & Ngoo, F. Z. (2002). Handy broker: An intelligent product-brokering agent for m-commerce applications with user preference tracking. *Electronic Commerce Research and Applications*, 1(3-4), 314–330. doi:10.1016/S1567-4223(02)00023-6
- Guan, S.-U., Chan, T. K., & Zhu, F. (2005). Evolutionary intelligent agents for e-commerce: Generic preference detection with feature analysis. *Journal of Electronic Commerce Research and Applications*, 4(4), 377–394. doi:10.1016/j.elerap.2005.07.002
- Hayes, C., Cunningham, P., & Smyth, B. (2001). A Case-based Reasoning View of Automated Collaborative. *Proceedings of 4th International Conference on Case-Based Reasoning, ICCBR2001*. doi:10.1007/3-540-44593-5_17

- Hea, W., Zhab, S., & Li, L. (2013). Social media competitive analysis and text mining: A case study in the pizza industry. *International Journal of Information Management*, 33(1), 464–472. doi:10.1016/j.ijinfomgt.2013.01.001
- Kaplan. (2012) If you love something, let it go mobile: Mobile marketing and mobile social media 4x4 Found. *Business Horizons*, 55(2), 129–139.
- Koorn, R., Smith, D., & Mueller, C. (2001). *e-Procurement and Online Marketplaces*. Amsterdam, The Netherlands: Compact. Retrieved on Jul 7 17, 2016 from <https://www.compact.nl/pdf/C-2001-1-Koorn.pdf> 1.
- Kumar, P., Gopalan, S., & Sridhar, V. (2005). Context enabled Multi-CBR based Recommendation Engine for E-commerce. *Proceedings of the 2005 IEEE International Conference on e-Business Engineering (ICEBE'05)*. doi:10.1109/ICEBE.2005.42
- Li, H., Tian, Y., & Liu, Y. (2013). UAI-IOT Framework: A Method Of Uniform Interfaces to Acquire Information from Heterogeneous Enterprise Information Systems. *IEEE International Conference on and IEEE Cyber, Physical and Social Computing*, 724–730.
- Li, K., & Du, T. C. (2012). Building a targeted mobile advertising system for location-based services. *Elsevier Science Direct*, 54(1), 1–8.
- Melville, P., Sindhvani, V., & Lawrence, R. (2009). Social media analytics: Channeling the power of the blogosphere for marketing insight. *Proc. of the WIN*, 1(1), 1–5.
- Nakamoto, S. (2008). *Bitcoin: A Peer-to-Peer Electronic Cash System*. Retrieved June 30, 2016, from <http://www.bitcoin.org/bitcoin.pdf>
- Nuti, G., Mirghaemi, M., Treleaven, P., & Yingseer, C. (2011, November). Algorithmic Trading. *Innovative Technology for Computer Professionals, IEEE*, 44(11), 61–69.
- Nwana, H. (1996). Software agents: An overview. *The Knowledge Engineering Review*, 11(03), 205–244. doi:10.1017/S026988890000789X
- Pant, G., & Menczer, F. (2002). MySpider: Evolve your own intelligent web crawlers. *Autonomous Agents and Multi-Agent Systems*, 5(2), 221–229. doi:10.1023/A:1014853428272
- Prasad, B. (1995). Planning With Case-Based Structures. In *Proceedings of the American Association for Artificial Intelligence (AAAI) Fall Symposium on Adaptation of Knowledge For Reuse*. Cambridge, MA: MIT.
- Prasad, B. (2003). Intelligent Techniques for E-commerce. *Journal of Electronic Commerce Research*, 4(2).
- Prasad, B., & Clementi, J. P. (2002). A Knowledge-Based E-Commerce System for Product Recommendation in Sports Domain. *Proceedings of the International Conference on Information and Knowledge Engineering*.
- Sarwar, B. M., Karypis, G., Konstan, J., & Riedl, J. (2002). Recommender systems for large-scale ecommerce: Scalable neighborhood formation using clustering. *Proceedings of the fifth international conference on computer and information technology*, 1–6.
- Sarwar, B. M., Karypis, G., Konstan, J. A., & Riedl, J. (2000). Analysis of recommendation algorithms for e-commerce. *ACM Conference on Electronic Commerce*, 158–167.
- Seave, A. (2013). *How Social Media Drives Your Customers' Purchasing Decisions*. Retrieved July 10, 2016, from <http://www.forbes.com/sites/avaseave/2013/07/22/how-social-media-moves-consumers-from-sharingto-purchase/>
- Shmueli, G., & Koppius, O. R. (2011). Predictive analytics in information systems research. *Management Information Systems Quarterly*, 35(3), 553–572.

Sinmao, M. V. (1999). *Intelligent Agents & E-Commerce*. Retrieved on January 27, 2009, from <http://www.cis.udel.edu/~wchen/Sinmaom.htm>

Sun, Z., & Finnie, G. (2004). A Unified Logical Model for CBR-Based E-Commerce Systems. *International Journal of Intelligent Systems*, 20(1), 29–46. doi:10.1002/int.20052

Tan, W., Blake, B. & Saleh, I. (2013). Social Network-Sourced Big Data Analytics. *IEEE, I(2)*, 62-69.

Uyoyo Zino Edosio. (2014), Big Data Analytics and its Application in E-Commerce. In *Proceedings E-Commerce Technologies*. University of Bradford.

Ververidis, C., & Polyzos, G. (2002). Mobile marketing using a location based service. In *Proceedings of the First International Conference on Mobile Business*. Prentice-Hall.

Wan, D. (2013). Influences of Cloud Computing on E-Commerce Businesses and Industry. *Journal of Software Engineering and Applications*, 6(06), 313–318. doi:10.4236/jsea.2013.66039

Weihong, H., & Yi, C. (2006). An E-commerce recommender system based on content-based filtering. *Wuhan University Journal of Natural Sciences*, 11(5).

Wettschereck, D., & Aha, D. W. (1995). Weighting Features. In *Proceedings of the First International Conference on Case-Based Reasoning*. New York: Springer.

Zhang, Y., & Wen, J. (2016). The IoT electric business model: Using blockchain technology for the internet of things. Science+ Business Media Springer.

KEY TERMS AND DEFINITIONS

Big Data: A type of structured and unstructured data that is characterized by its big volume usually measured in Terabytes, Exabytes or Petabytes and requires specific hardware and software tools for processing.

Cloud Computing: A real-time distributed computing environment where data and applications are stored on globally distributed data centers which guarantee flexibility, availability, scalability, high storage capabilities and data replication.

E-Commerce: Doing traditional business over the Internet.

Intelligent Techniques: Artificial intelligence techniques that allow computers to perform complex tasks just like humans can do such as: understanding words meaning, decisions making and controlling complex objects.

Internet of Things: The ever growing network of heterogeneous objects such as digital devices, software objects and human beings interconnecting to interact and exchange information.

Mobile Marketing: Broadly engaging with customers through advertisements, messages and apps on mobile devices.

Social Media: Websites and online applications that allow large group of people to communicate, socially interact, exchange information and share opinions.

Online Mediation in E-Commerce Matters



Ángela Coello Pulido
University of Vigo, Spain

INTRODUCTION

Conflict is not an exceptional fact in human being's life but an ordinary phenomenon in relationships between parties who have different points of view and different positions. Society cannot be based in a permanent conflict situation. Therefore, legal systems have standardized some mechanisms which have been designed to solve different sort of disputes between or among individuals or organizations or between or among the firsts and seconds. Some of these systems are hetero-compositives (trial, arbitration) and some of them are auto-compositives (mediation, conciliation, negotiation, etc.).

Mediation is one of the auto-compositives mechanisms created to resolve differences between disputants and it is included in the so-called Alternative Dispute Resolution (ADR), that is to say, it allows an out-of-court solution. This system, which is getting more and more significant, consist of a way in which the parties negotiate by themselves to get an agreement with the help of a third, the mediator, who is impartial and neutral. In mediation table the parties defend their own positions and the mediator tries to emerge the true interests of the parties as only with the base of the interests, the amicable settlement will be possible and effective. As Fisher, Ury and Patton (2012) have pointed in their *Getting to yes. Negotiating an agreement without giving in*, "for a wise solution reconcile interests, not positions" (p. 42).

An important issue should be taken into account: in this way of problem solving, one of the negotiators could be more influential or more powerful than the others. Therefore, the mediator should assume the role of a professional who is

called to balance the leverage between or among negotiators. This is one of the biggest advantages of mediation with regard to traditional negotiation.

Alternative Dispute Resolution (ADR), in general, and mediation, in particular, have been gradually evolving through years to be adapted to the new times. Thus, they became into Online Dispute Resolution (ODR) and online mediation. Online mediation is one of the Online Dispute Resolution (ODR), that is, one of the Alternative Dispute Resolution (ADR) which uses Information and Communications Technology (ICT) as a key element in resolving process. The emergence of ADR as legal resources to problem solving and their effectiveness, motivate to use them when circumstances allow it. Mediation, which is having more and more acceptance, can be carried out by using electronic instruments without losing its essence.

One of the fields where online mediation could display all its charm is e-commerce. E-commerce has some virtues that transform it in a very useful way for market. It uses the Internet to do business. World Wide Web and mobile apps became the protagonist of this new reality. Consequently, consumers and users more easily access the market while companies improve their business.

E-commerce has increased its importance in the last years because of its usefulness in business world. Information and Communications Technology (ICT) is an unquestionable reality and, because of that, online relationships should be developed in an agile, certain and peaceful way. Consequently, obviously it should be considered that if e-commerce allows the development of online relationships, conflicts arising from these relationships also must be resolved online. The

reason is that it would not be logical to approve the e-commerce relational virtues and break them from the moment the conflict arise.

In short, it is clear that online mediation has some important advantages (lower-cost, faster, etc.). However, on the other hand, despite the virtues of online mediation, it could also reveal some disadvantages (difficulties in the use of ICT by some citizen, Internet connection problems, etc.). Therefore, these risks should be taken into account.

The aim of this chapter is to describe why online mediation is an appropriate instrument for conflict solving in e-commerce field and what are some of the advantages and disadvantages of this ODR not only for citizens but also for organizations.

BACKGROUND

The early moments of e-commerce could be found in the late years of 20th century but it was in 21th century when this kind of trade became into a revolution thanks to the improvement of technology and telecommunications. ICT and our social system are closely connected, which makes it impossible to understand one without the other (Garriga Domínguez, 2012, 67).

However, cyberspace is not always a peaceful space. Where information and communication operate in a high, fast and agile level and where contracts can be concluded at the click of a mouse, frequently appear the misunderstandings, the interest clashes and, definitely, the differences between and among the e-commerce users. Misunderstandings, delays, mistakes, late payment and non-payment are some of the conflicts that can arise in e-commerce transactions.

Conflicts are often solved in a judicial process. However, when it is possible, an amicable settlement could be the best option in order to get an effective solution. Especially in the complex area of ICT. Thus, online mediation was born as a system that is adapted to the new times.

In such a context, it could be considered that face-to-face mediation could be more appropriate due to possible communication difficulties and possible problems in information exchanges in online option. Traditional mediation is very virtuous because of the personal meetings. This personal interaction allows the mediator to guide conversations and to build an environment for good communication trying to moderate discussions and helping the parties to approximate their interests. However, the task could be more difficult in online mediation because the absence of personal meeting that could work as a barrier for a good communication. On the contrary, some have thought that online mediation has useful virtues. For instance, it has been argued that online mediation has more similarities than differences with respect to face-to-face mediation (Alzate Sáez de Heredia, 2008, p. 8). It could be very useful when the parties are separated geographically or, even, when communication between them is especially difficult. Furthermore, it has been said that as well as cyberspace provides an environment to observe new sources and patterns of conflict, technology provides the opportunity to consider new approaches to conflict resolution (Katsh, 2001, p. 812).

In opinion of Cortés (2011), “there are some misconceptions about ODR, such as the idea that ODR is only valid for small claims, or that ODR exclusively relies on automated technology, or that ODR can only deal with online disputes” (p. 2).

In any case, the reality is that online mediation is working properly. To that effect, it could be developed by different instruments such as e-mail, video conference or even a virtual platform that the mediator uses to that task; definitely, resources for reproducing sound, images and texts. In fact, written exchanges are not strange in online mediation. If the mediator decides to use a special platform, this instrument should be used not only for video conference but also for upload and download data and documents. Besides, it should have a schedule for mediation program-

ming, a notifications system, instant messaging, text message and a documentation management system. All of this in a confidential environment.

ONLINE MEDIATION IN E-COMMERCE. ITS USEFULNESS

E-commerce provides a number of features that reveal its success and effectiveness. Some of them are its ubiquity, its speed or its ease. Firstly, the ubiquity of e-commerce refers the opportunity of puts somebody in touch with another person who is at great distance from them and at any time. Instead of a physical marketplace, the result is a marketplace, that is to say, a marketplace which is extended beyond traditional boundaries and removed from a temporal and geographic location (Laudon & Traver, 2015, p. 52). Secondly, the speed is a consequence of ubiquity because if e-commerce operates at any place and at any time, many obstacles can be overcome and it can obtain the speed as a result. Lastly, the ease of e-commerce is evidenced by the fact that at the click of a mouse, the transaction can be formalized.

Therefore, e-commerce operates from a global perspective. Thus, it is reasonable to believe that conflicts which emerge in this scenario should be solved by an instrument that has the same useful features. It would not be logical to consider such a useful marketplace if it did not exist an available mechanism with similar advantages for problem solving in this context.

Online mediation allows maintaining important advantages of the e-commerce. Its link with the ICT makes it a perfect candidate for that task. Online mediation shares the ubiquity, speed and ease. Online mediation can work at any place if connection is possible. Besides it is speedy because it avoids the obstacle of geographic distance between the parties and mediator and easy for use to the extent that people is able to use electronic mechanisms correctly.

There is one important idea that should be taken into account in any study of the ODR which is the consideration of technology as a Fourth Party.

This idea was first proposed by Kasth & Rifkin in their *Online Dispute Resolution: Resolving Conflicts in Cyberspace* (Kath & Rifkin, 2001).

Relating to mediation, on the one hand, traditional mediation has, at least, the following three participants: the mediator and two parties in conflict. On the other hand, online mediation has, at least, the following four: the mediator, the two parties in conflict and the ICT which include additional advantages to this auto-compositive ADR.

However, something should be considered: despite the unquestionable advantages of online mediation, it can also have some disadvantages because of its newness. In other words, due to ICT are relatively new technologies, online mediation has a shorter history than traditional mediation. The lack of a high experience brings to light certain difficulties that should be noticed and that can be overcome at the same time that technological progress.

Both advantages and disadvantages will be analyzed in the following.

Advantages of Online Mediation in E-Commerce

Understanding of advantages of online mediation requires the difference between two types: the advantages which come from the use of certain technology and those that result from the inexistence of a face-to-face meeting as in traditional mediation.

Firstly, resulting from the intervention of ICT, the online mediation shares some characteristics with the e-commerce. Therefore, it is very useful in this field.

Before anything, online mediation is useful because of its ubiquity. The ubiquity means the possibility of maintain that contact between parties in conflict without connecting simultaneously but at different times. On the one hand, the global reach allows people located in different geographic points to solve their problems, emerged during e-commerce transactions, avoiding additional costs as trips for instance. On the other hand, the asynchronicity manifestations are found in the

possibility to establish direct contacts at different times depending on the Internet connection of the parties and this has a special importance relating to the time zone if the parties are at a great distance. For instance, this feature of online mediation is useful if the conflict emerges between a consumer and a trader who is established in a different State with a different time zone.

When mediation is carried out through the use of video conference, communication is established at the same time. This would be an appropriate option if the parties are in the same time zone and have the same time availability. However, it could be developed by other ways like e-mail or chat. In the second case, parties can connect depending on their availability.

Moreover, given that online mediation allows the parties to solve their differences whenever and wherever they are, the process is speedy because of it is able to make the most of time, in other words, it is able to optimizing time.

In addition, mediation practiced online tries to be easy by using ICT. These new technologies can facilitate operations. Simple things like documents sharing or record the minutes. But also more complicated matters like permitting the ubiquity that has been mentioned above.

Secondly, beyond the technical issues, online mediation is very useful, for example, in order to avoid the difficulties of a face-to-face meeting. Sometimes people are not able to face a problem solving process in front of their adversary, but they get good results in front of a computer screen.

Disadvantages of Online Mediation in E-Commerce

Just like the advantages, understanding of disadvantages of online mediation requires the difference between two types: the disadvantages which come from the use and access to certain technology and those that result from the inexistence of a face-to-face meeting as in traditional mediation.

The new Information and Communications Technologies are becoming more and more im-

portant and useful. However, some doubts (or disadvantages) exist on these technologies, which are reflected in the online mediation.

One of these disadvantages, which is very important, is the doubt in relation to the secure information exchange. The mediation process needs an information exchange and this requirement could result in the parties' mistrust because of the possibility of digital recording of these details and their retention.

In these circumstances, this issue can result in two problems. On the one hand, the concern of the parties to disclose relevant information. Each of these participants has some data and could think that an online process could result in a loss of control. Consequently, online mediation entails the risk of concealing information that could lead to an effective agreement.

On the other hand, appears the right to privacy. Computers can store a significant amount of information. Automatic processing of personal data has led to the birth of a new human right: the right to informational self-determination which is included in the Third Generation of human rights. Information is power in our society and this power is decisive when computing transforms partial and disperse information into mass and organized information; therefore legal framework of computing has priority interest (Pérez Luño, 2006, p. 32).

Difficulties linked with information are serious because information plays a significant part in any negotiation, including mediation. Information managed at the preparatory stage and its use during the rest of the process could be one of the essential elements of negotiation.

Another disadvantage that could be arise from online mediation with respect to traditional mediation is a possible poor understanding of the correct use of ICT or, even, an almost ignorance (not a complete lack of knowledge about the existence of these technologies because e-commerce requires a minimum). Perhaps, one or more of the parties do not have enough skills or expertise in the use of ICT. Maybe they are not used to

using these technologies. And not just this, but also it is possible a situation of lack of access to these technologies whenever the user wants. Information and Communications Technologies are relatively young. Therefore, not everyone has knowledge about how to use it correctly as well as not everyone accesses it at any time. It is necessary a technological equality (Vilalta Nicuesa, 2013, p. 292).

The lack of skills or expertise or the access difficulties could result in a imbalance between the parties. Maybe some of them are not communicating properly and this makes difficult their participation in mediation or more difficult than the other parties' participation.

There is an important difficulty in online mediation which is the necessity of global standards, that is to say, common rules to every problem solving process. It must be kept in mind that the ubiquity of this method allows the start of both national and international mediation processes. Thus, cross-border processes require the determination of applicable law. Therefore, it could be useful the implementation of common standards for every problems solving process in this field.

Finally, besides, exist other problems in this field that are more general: computer virus, software errors, computer malfunctions, etc.

Secondly, should also refer to disadvantages of online mediation which come from the inexistence of a face-to-face meeting as usually happen in traditional mediation. For instance, this method result impersonal and it is specially the personal meeting one of the greatest virtues of mediation.

SOLUTIONS AND RECOMMENDATIONS

In view of the foregoing, it should be noted that both advantages and disadvantages of online mediation have to do with the access and use of Information and Communications Technology, not with mediation features themselves.

Therefore, on the basis that online mediation has certain advantages as an auto-compositive system created for problem solving, just in the same way of traditional mediation, the problems or disadvantages should be manage properly. Thus, parties should take care of their reserved information, privacy, etc., becoming aware of the real process they are facing.

ICT are very useful in mediation, especially in a field like e-commerce. However, it is recommended to analyze each particular situation at the very beginning of mediation in order to make the decision of going through an online way or a traditional way. Besides, if the circumstances demand it, it is advisable that the parties put an end to online mediation to begin a traditional one.

In addition, it might be considered that the ICT future development could correct some of the disadvantages of online mediation. For instance, solving difficulties as those related to the right to privacy. But others will stay because they are a consubstantial part of the mediation. For example, the difficulties in information exchange.

FUTURE RESEARCH DIRECTIONS

Access to information, global communication, speed transactions, etc. These are some of the things that technology allows to consumers and users, service providers, traders and general public. As ICT, e-commerce is becoming more and more usual, successful and, ultimately, useful. ODR operate in this field with the same benefits.

Consequently, to the extent that ICT develops, also ODR are gaining popularity. According to Rule (2015), "technology is changing not only the way we communicate; it is altering the way we disagree and the way we resolve our disputes" (p. 5).

As a result of the foregoing, it seems that future research will continue in two ways: the technological study and the legal-social study. Besides, the legal-social study will depend on the technological study because of the need for the

regulation of the relationships in the new context of cyberspace to respond to changes as they arise.

Therefore, new communication channels can be implemented in order to improve ORD, in general, and online mediation, in particular, at the same time as a solution for an appropriate regulation is sought. In other words, and referring to e-commerce, this trade can be improved by technological advances and, in the same direction, online mediation will be adapted to the e-commerce users' needs.

But another line of research should be taken into account which is the evolution from ADR to ODR given that the unquestionable ICT emerge and development. ADR have proved their success and effectiveness. Now is time for ODR.

CONCLUSION

E-commerce is here and it has come to stay. This type of trade entail a new space for relationships development in ICT area; therefore it is necessary an effective mechanism for problem solving in this field. This system should be adapted to these technologies; as a consequence, online mediation is an appropriate method in order to face this task.

Online mediation has certain advantages with respect to traditional mediation, that is to say, mediation based on face-to-face meetings. However it has also a number of disadvantages or difficulties. Some of these advantages and disadvantages have emerged as a result of the use of ICT and others as a consequence of avoiding in-person processes. Nevertheless, the most important fact about the link between e-commerce and online mediation is the similar characteristics that they have and that transform online mediation in an appropriate mechanism for problem solving in the sphere of this trade: ubiquity, speed, ease.

In conclusion, if e-commerce has a variety of features that make it so useful and effective, the mechanism for problem solving in this field should have similar advantages. Online mediation has them. Such a useful mechanism as online mediation is part of contemporary advance. Digital Age is here.

REFERENCES

- Alzate Sáez de Heredia, R. (2008). Mediación en línea. *Revista de Mediación*, 1, 6–15.
- Cortés, P. (2011). *Online dispute resolution for consumers in the European Union*. Oxon, UK: Routledge.
- Fisher, R., Ury, W., & Patton, B. (2012). *Getting to yes. Negotiating an agreement without giving in*. London: Random House Business Books.
- Garriga Domínguez, A. (2012). El impacto de las TIC en los derechos humanos. In G. Domínguez (Ed.), *Fundamentos éticos y jurídicos de las TIC*. Navarra, Spain: Thomson Reuters. Aranzadi.
- Katsh, E. (2001). Online Dispute Resolution: Some lessons from the e-commerce revolution. *Northern Kentucky Law Review*, 28(4), 810–821.
- Katsh, E., & Rifkin, J. (2001). *Online Dispute resolution: Resolving Conflicts in Cyberspace*. San Francisco, CA: Jossey-Bass Publishers.
- Laudon, K. C., & Traver, C. G. (2015). *E-commerce 2015. Business. Technology. Society*. Harlow, UK: Pearson.
- Pérez Luño, A. (2006). *La tercera generación de derechos humanos*. Navarra, Spain: Thomson. Aranzadi.
- Rule, C. (2015). Technology and the future of dispute resolution. *Dispute Resolution Magazine*, 4-7.
- Vilalta Nicuesa, A. E. (2013). *Mediación y arbitraje electrónicos*. Navarra, Spain: Thomson Reuters. Aranzadi.

ADDITIONAL READING

- Abney, S. (2011). *Civil Collaborative Law. The road less travelled*. Bloomington, IN: Trafford Publishing.

- Alexander, N. (2009). *International and Comparative Mediation. Legal Perspectives*. Alphen aan den Rijn, The Netherlands: Wolters Kluwer. Law & Business.
- Barona, S., & Esplugues, C. (2014). ADR Mechanisms and their incorporation into global justice in the twenty-first century: some concepts and trends. In Esplugues and Barona (eds.), *Global perspectives on ADR* (pp. 1-52). Cambridge-Antwerp-Portland: Intersentia.
- Bartness, B. A. (2015). Mediation strategies. Unique to commercial transactions (not really). *ACResolution*, 8-9.
- Berguer, K. P. (2009). *Private dispute Resolution in International Business. Negotiation, Mediation, Arbitration.: Vol. II. Handbook*. Alphen aan den Rijn, The Netherlands: Wolters Kluwer. Law & Business.
- Cortés, P. (2010). Un modelo para la acreditación de los sistemas ODR en la Unión Europea. *IDP. Revista de Internet. Derecho y Política*, 10, 4-14.
- Deutsch, M., Coleman, P. T., & Marcus, E. C. (Eds.). (2006). *The handbook of conflict resolution*. San Francisco, CA: John Wiley & Sons.
- Ebner, N. (2008). Trust building in e-negotiation. In L. Brennan & V. E. Johnson (Eds.), *Computer-mediated relationships and trust: managerial and organizational effects*. IGI Global. doi:10.4018/978-1-59904-495-8.ch010
- Esplugues, C. (Ed.). (2014). *Civil and comercial mediation in Europe. Cross-border mediation*. Cambridge, Antwerp, Portland: Intersentia.
- Esplugues, C., Iglesias, J. L., & Palao, G. (Eds.). (2013). *Civil and commercial mediation in Europe. National Mediation. Rules and Procedures*. Cambridge, Antwerp, Portland: Intersentia.
- Esplugues Mota, C., & Barona Vilar, S. (2014). *Global perspectives on ADR*. Cambridge, Antwerp, Portland: Intersentia.
- Fisher, R., & Shapiro, D. (2007). *Building agreement. Using emotions as you negotiate*. London, United Kingdom: Random House Business Books.
- Frey, M. A. (2003). *Alternative methods of dispute resolution*. Clifton Park, NY: Thomson Delmar Learning.
- Kennedy, G. (1998). *Kennedy on negotiation*. Aldershot, United Kingdom: Gower.
- Kennedy, G. (2009). *Essential negotiation. An A-Z Guide*. New York, NY: Bloomberg Press.
- Lax, D. A., & Sebenius, J. K. (1986). *The manager as negotiator*. New York, NY: The Free Press.
- Poblet, M. (Ed.). (2011). *Mobile technologies for conflict management. Online dispute resolution, governance, participation*. New York, NY: Springer. doi:10.1007/978-94-007-1384-0
- Rainey, D. (2014). Third-Party Ethics in the Age of Fourth Party. *International Journal of Online Dispute Resolution*, 1, 37-56.
- Raymond, A. H. (2014). Yeah, but did you see the gorilla? Creating and protecting an informed consumer in cross-border online dispute resolution. *Harvard Negotiation Law Review*, 19(129), 129-171.
- Riskin, L. L. (2003). Decisionmaking in mediation: the new old grid and the new new grid system. *Notre Dame Law Review*, 79, 1, 53 pp.
- Rule, C. (2003). *Online dispute resolution for business. For e-commerce, B2B, consumer, employment, insurance, and other commercial conflicts*. San Francisco, CA: John Wiley & Sons.
- Rule, C. (2008). Making peace on eBay: resolving disputes in the world's largest marketplace. *ACResolution*, 8-11.
- Rule, C. (2015). Dispute resolution and the evolution of customer support. *ACResolution*, 6-7.
- Ury, W. (1993). *Getting past no. Negotiating in difficult situations*. New York, NY: Bantam Books.

Ury, W. (2000). *The third side. Why we fight and how we can stop*. New York, NY: Penguin.

Ury, W. (2007). *The power of a positive no. Save the deal. Save the relationship and still say no*. New York, NY: Bantam Books.

Wahad, M. S. A., Katsh, E., & Rainey, D. (Eds.). (2012). *Online dispute resolution: theory and practice. A treatise on technology and dispute resolution*. The Hague, Netherlands: Eleven international publishing.

Ware, S. J. (2007). *Principles of Alternative Dispute Resolution*. St. Paul, MN: Thomson West.

KEY TERMS AND DEFINITIONS

Alternative Dispute Resolution (ADR): Systems created to get an out-of-court solution to differences between parties who have different points of view and different positions.

Amicable Settlement: A deal which comes from a consensual agreement of the parties in a negotiation process.

Auto-Composition: A way for problem solving which is based in an arrangement emerged from the free will of the opponents.

E-Commerce: The use of the Internet for doing business between and among individuals or organizations or between or among the firsts and seconds.

Information and Communications Technology (ICT): Technologies which have been created to manage and transmit information.

Online Dispute Resolution (ODR): Systems created to get an out-of-court solution to differences between parties who have different points of view and different positions by using Information and Communications Technology (ICT) as a key element in resolving process.

Online Mediation: An Online Dispute Resolution which consists of a way in which the parties negotiate by themselves to get an agreement with the help of a third, the mediator, who is impartial and neutral.

Reputational Mechanism in Consumer-to-Consumer Online Commerce

E**Mikhail I. Melnik***Kennesaw State University, USA*

INTRODUCTION

The emergence of the Internet is, in many ways, an historically similar event to the industrial revolution; it has fundamentally changed the cost of transmitting, presenting, and processing information. However, the new means of transmitting and presenting information do not insure its quality. Early in the Internet development, asymmetry of information presented a significant challenge to the adaptation of the Internet in consumer-to-consumer commerce. However, with the help of simple, intuitive, and transparent instruments such as customer review or rating mechanisms, consumer-to-consumer online commerce has been able to overcome this potentially serious obstacle and flourish, changing the way consumers shop.

BACKGROUND

Asymmetry of Information

Asymmetry of information occurs when the parties involved in a transaction have different access to the relevant for the transaction information. The issue of asymmetry of information has been studied by economists for many decades. In 2001, George Akerlof shared the Nobel Prize in economics with Michael Spence and Joseph Stiglitz for their monumental work in the area of asymmetry of information in economics. His groundbreaking work formulated the problem on the example of used car sales, Akerlof (1970). When selling a used car, the seller knows far more about the car than a perspective buyer, putting the buyer into

a position of disadvantage. The asymmetrically distributed information not only puts the party with less information at a disadvantage but also creates a disincentive for that party to participate in the transaction. This latter characteristic can threaten the growth and development of an impacted market.

In online consumer-to-consumer commerce, the issue is magnified by the fact that there are two channels for the emergence of asymmetry of information. First, in most cases, the perspective purchasers may not have any direct way of examining the goods they intend to acquire online. Second, historically, buyers have been required to submit payments prior to the seller delivering the item. This dynamic implies that the buyer has to rely on the seller for the accuracy of the item's description and compliance with the terms of transaction.

The success of online consumer-to-consumer marketplaces, including one of the largest such platforms – eBay, serves as an excellent and innovative example of addressing the problem of asymmetry of information. Typically, a consumer-to-consumer website merely serves as the platform for its users to sell/purchase various goods and services. This limited role places the burden of compliance with the terms of transaction on the individual participants and therefore presents the possibility of asymmetry of information.

eBay Rating Mechanism

eBay is perhaps the largest and most recognized consumer-to-consumer website today. eBay is set up as a platform for third party users to sell

DOI: 10.4018/978-1-5225-2255-3.ch247

their goods. eBay corporation merely serves as the marketplace and brokers the trades, but the actual compliance with the terms of transaction is left up to the individual users. By this setup, eBay exposes its users to potentially serious asymmetry of information issues, including the accuracy of item description and compliance with the terms of transaction, such as shipping the product, selecting the correct shipping option, shipping it in a timely fashion, etc. The problem is further highlighted by the customary requirement that the buyer submits the payment prior to the seller shipping the item. Despite the obvious nature of the problem, eBay has become a flourishing consumer-to-consumer platform with the quarterly gross merchandise trade approaching nearly 22 billion dollars in the fourth quarter of 2015 (eBay, 2016).

eBay was able to overcome the asymmetry of information problem by creating a simple and intuitive user rating mechanism. In February of 1996, just six months after the launch of the website, eBay introduced its rating system, back then, referred to as feedback (eBay, 1996). After completing a transaction on eBay, each party can rate the performance of the other along three choices: positive, neutral, and negative. eBay converts these individual responses into a continuous variable, the eBay rating, which is simply the difference between the number of positive and negative responses. This rating is then displayed next to the user's id. In the case of sellers, it is easily visible on every listing they place on eBay.

The eBay rating of a user only captures responses left by those users with whom the user had transactional experience on eBay. This reduces the ability of manipulating someone's rating. Furthermore, although eBay users can rate each other after every transaction, the rating only measures responses left by unique users. This implies that if two users participate in multiple transactions between themselves and place multiple positive responses about each other, these multiple responses will count as one, because they are left by the same user. This characteristic of the eBay rating mechanism further enhances the ability to resist manipulation.

Over time, the rating mechanism underwent some changes. Today, each seller's auction not only displays the overall value of the seller's rating, but it also reports the percentage of positive responses. In addition, buyers have the ability to rate sellers on various individual aspects of transaction experience. These may include the accuracy of item description, the quality of communications, expedience of shipping, and shipping and handling charges.

The simplicity of the mechanism not only makes the process of evaluating someone's performance on eBay easy, but it also makes the interpretation of the rating mechanism rather straightforward. This is particularly important for sellers. For buyers, the rating is arguably less important (Melnik and Alm, 2002). The seller's overall rating conveys the information about the net number of praises left by other unique users who had transaction experience with this seller. Thus, the rating can serve as a signal of reputation, and thereby help address the problem of asymmetry of information and promote trust (Resnick and Zeckhauser, 2002).

Other online consumer-to-consumer platforms have developed their own, similar to eBay rating mechanisms. One example is Yahoo Auctions. During 1998 – 2007, Yahoo competed with eBay in the auction space (in the U.S. market) and employed a rating mechanisms that was similarly constructed (Melnik, 2011).

SOLUTIONS AND RECOMMENDATIONS

Empirical Evidence

A rising price (English design) auction outcome can be analyzed as a two-step process. First, an auction may or may not result in a sale. Second, if an auction results in a sale, then its price is determined by the willingness to pay of the bidder with the second highest willingness to pay (assuming there are more than one bid). This

two stage nature allows for the empirical analysis of the effect of the seller's rating to progress along several venues: estimating the effect on the probability of sale, willingness to pay (including censored observations, i.e. observations with no bids), and willingness to pay for sold items only (including a more restricted approach where only observations with multiple bids are present).

Seller's Rating and Buyer's Willingness to Pay

One popular empirical approach has been to estimate the willingness to pay. Since willingness to pay is censored for those listings where no bids were received, i.e. the listings resulting in no sale, a censored regression technique, such as the Tobit maximum likelihood estimation method, became a popular approach. In an eBay auction, the seller has the choice of the starting or opening price. Since opening prices can vary from listing to listing, the proposed variation of the Tobit estimation technique by Amemiya (1984) has become frequently used in studies using eBay data. Amemiya (1984) modified the Tobit model to allow for variability in the censoring point across observations. The use of this variation of the Tobit model allows for grouping of all observations (censored, i.e. with no bids, and uncensored, i.e. with bids) into one dataset. For empirical illustrations of this approach see Dewally and Ederington (2006), Melnik and Alm (2002, 2005), and Reiley et al. (2007).

These studies generally find that the seller's reputation has an impact on the buyer's willingness to pay in online auctions. Dewally and Ederington (2006) use data on collectible comic books and show that a higher reputation increases willingness to pay. Furthermore, they demonstrate that the effect is larger for those items that are not professionally graded (i.e. items with third-party certified product characteristics). Melnik and Alm (2002) focus on a homogeneous product, a 5 dollar U.S. gold coin which derives virtually all of its value from its gold content. They demonstrate that the seller's overall rating has a positive effect on

willingness to pay and their negative reputation has a significant penalty (Melnik & Alm, 2002). Reiley et al. (2007) find that the overall reputation has no statistically significant effect, while the negative reputation significantly reduces the buyer's willingness to pay.

Probability of Sale

Another approach has been to investigate the probability of sale. Here, a simple probit or logit model is employed to investigate how the various characteristics of an auction, including the seller's rating, impact the probability of sale.

Using a logit model, Eaton (2002) demonstrated that negative feedback reduced the probability of sale. The results also showed that the seller's positive rating had no significant impact on the probability of sale (Eaton, 2002). Employing a probit model, Jin and Kato (2008) showed that the seller's positive reputation (feedback) increased the probability of sale while the negative reputation (feedback) decreased it.

Jin and Kato (2008) used collectible sports trading cards in their study and the results demonstrated that a certification by a third party reduced the effect of the seller's negative reputation. Their findings demonstrated that the presence of certification by a credible third party reduces the uncertainty about the item. This reduces the signaling property of the seller's rating, thereby reducing its importance in the buyer's decision.

Sold Items Only

A number of empirical studies focused on sold items only. This approach enables the researcher to directly estimate the willingness to pay of the second highest bidder using the OLS technique. However, the downside of this approach is that it ignores all observations with no bids (or in the case of the more restricted studies, one bid listings as well). It is important to note that in an English auction, the winning bid represents the willingness to pay of the bidder with the second

highest valuation. This implies that listings with a single bid become censored as the second highest valuation is unobserved and lies below the observed bid. As a result, some of the studies focused on successfully completed listings only restricted their data to observations with two or more bids. For example, Houser and Wooders (2006) restricted their estimation to observations with at least two bids and showed that the seller's positive rating increased the winning bid while the negative rating decreased it. Another illustration of an earlier work using sold only auctions was a study by McDonald and Slawson Jr. (2002). This study also found that a higher seller's rating resulted in a higher closing price and an increased number of bids.

Homogeneous vs. Heterogeneous

As previously noted, asymmetry of information presents two issues for perspective buyers in online consumer-to-consumer commerce: uncertainty about the accuracy of the product description and uncertainty about the seller's compliance with the terms of transaction.

One of the main challenges faced by anyone who wishes to conduct an empirical investigation using eBay generated data is the choice of the product for data collection. Some products might exhibit significant variability in item-specific characteristics, presenting challenges for empirical investigations. For example, in the market for used textbooks, some books might be more worn out than others. If one does not control for the quality of each individual textbook, one omits an important regressor.

This induced many researches to focus on items with controlled item-specific characteristics. Such controls can come in a variety of ways. For some products, the mere presence of a photo of the actual item may be sufficient. Although, a photo does convey a lot of information to the perspective bidders, such information is not easily quantifiable into a continuous numerical variable. As a consequence, a frequently favored approach

by empirical researchers was to focus on a product that had easily measurable item-specific characteristics, had a third party evaluation (certification), or was simply one that exhibited no heterogeneity.

Houser and Wooders (2006) selected Pentium processor chips in their study. Since Pentium processors are unlikely to exhibit significant item-specific variability (for each particular processor model), this practically ruled out heterogeneity. Many investigators decided to focus on items with easily measurable quality and third party certifications, e.g. collectible coins. Examples of such studies include Lucking-Reiley et al. (2007), Bajari and Hortacsu (2003), and Melnik and Alm (2002, 2005).

Another approach to address the issue of item-specific heterogeneity is to conduct a controlled experiment. This approach allows not only to control for the item-specific characteristics, but also for the behavior of the seller. One such study was done by Resnick et al. (2006) in which the same established eBay seller sold identical items using two different eBay accounts: their own established account and a new account with a zero rating (established for the purpose of the experiment). The results demonstrated that the difference in willingness to pay was about 8.1% of the selling price in favor of the account with the higher rating (Resnick et al., 2006). This study suggested a rather large penalty being placed on a new seller and highlighted the importance of the eBay rating for sellers. As discussed later, such penalties can play an important role because they increase the costs sellers face when switching to a competing platform, thereby potentially reducing the competitive pressure on the platform itself.

However, item-specific heterogeneity is a double-edged sword as it also presents a unique opportunity to differentiate between the various roles the reputational mechanism plays. In the case of a product with unobserved item-specific characteristics, the seller's reputation signals two important pieces of information: the reliability of the seller in complying with the transaction's terms, and the reliability of the seller in accurately

describing the item. In the case of a homogeneous good, the signaling role of the mechanism is reduced to the first of these properties only.

Melnik and Alm (2005) attempted to analyze these different signaling properties of the seller's reputation on eBay by contrasting uncertified, high quality (AU—almost uncirculated) collectible U.S. Morgan silver dollar coins with the same category coins that were certified by an independent coin grading agency. Their results indicated that the overall seller's rating played an important role in the case of the non-certified coins. However, in the case of the certified coins, the overall rating was insignificant while the negative rating had a statistically significant negative impact on the willingness to pay. It is important to note that the average prices differed sizably between these two groups of coins, with the certified coins being significantly more expensive on average (\$327 versus \$58). The results suggested that the negative rating conveyed the signal of the reliability of the seller when it came to complying with the terms of the transaction. This is because in the case of a more expensive item, the consequences to the buyer of the seller's failure to comply with the terms of the transaction were more severe.

eBay Seller's Rating as a Barrier to Entry

As the preceding discussion illustrates, the eBay rating mechanism has played an important role as a signal of information about the quality and reliability of the seller in an online auction. The rating mechanism enabled eBay to overcome the asymmetry of information problem in online commerce prior to the development of various forms of buyer protection. In this sense, the rating has played an important historical role in the growth and popularity of the eBay website (Brown and Morgan, 2006). However, the implications of the rating mechanism did not stop there.

The eBay rating mechanism has also created a sizable barrier to entry for competing to eBay consumer-to-consumer online platforms. Resnick

et al. (2006) demonstrated that the difference between an established seller and a newcomer can be around 8.1% of the sales price. Melnik and Alm (2002) show a much smaller but still sizable impact of the seller's reputation on the realized auction price. They find the difference in the willingness to pay to be about 4.8% between an established seller and a new one. These empirical findings indicate that an established eBay seller would potentially incur a sizable price penalty if they were to shift from eBay to a competing online auction platform and start there anew. Indeed, this characteristic of eBay rating was tested in the competition with Yahoo Auctions in the U.S. market.

FUTURE RESEARCH DIRECTIONS

Further empirical work is needed to draw more definitive conclusions on the exact effects of seller's reputation in consumer-to-consumer online commerce. This includes understanding the various signaling properties of online reputational mechanisms. As can be concluded from the preceding discussion, reputational mechanisms are also changing over time, presenting a further basis for empirical and theoretical research.

Emergence of third party payment mechanisms also alters the asymmetry of information in consumer-to-consumer online commerce. Credit cards, online payment mechanisms such as PayPal, all reduce some aspects of asymmetry of information, thereby altering the signaling properties of any reputational measures.

CONCLUSION

Empirical evidence of the impact reputational mechanisms have in online consumer-to-consumer commerce is less than fully conclusive. Although most studies have demonstrated that various aspects of the eBay seller rating mechanism have been shown to have statistically significant impact on the realized price and auction outcome, there

are other studies that show the impact to be statistically insignificant. However, several conclusions may be drawn from the empirical literature.

One such conclusion is that negative comments about a seller tend to matter more to the perspective buyers than positive comments. This is seen in both, the magnitude and statistical significance of the coefficients in most studies where the individual aspects of the seller's rating are used (Melnik and Alm, 2005). Another common observation is that the relationship between the seller's rating and the buyer's willingness to pay is non-linear and exhibits diminishing marginal returns. The first few positive comments may significantly matter to the potential buyers, but the marginal effect diminishes with rating. It can be argued that at some level, additional positive comments become insignificant. Finally, if the seller's rating indeed impacts the buyer's willingness to pay, then this rating may serve as a barrier to entry not only for other sellers, but for competing consumer-to-consumer platforms (see Melnik and Alm, 2005; Melnik, 2011).

REFERENCES

- Akerlof, G. (1970). The market for lemons: Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, 84(3), 488–500. doi:10.2307/1879431
- Alm, J., & Melnik, M. (2010). Do eBay sellers comply with state sales taxes? *National Tax Journal*, 63(2), 215–236. doi:10.17310/ntj.2010.2.02
- Amemiya, T. (1984). Tobit models: A survey. *Journal of Econometrics*, 24(1-2), 3–16. doi:10.1016/0304-4076(84)90074-5
- Bajari, P., & Hortacsu, A. (2003). The winners curse, reserve prices, and endogenous entry: Empirical insights from eBay auctions. *The Rand Journal of Economics*, 34(2), 329–355. doi:10.2307/1593721
- Bajari, P., & Hortacsu, A. (2004). Economic insights from internet auctions. *Journal of Economic Literature*, 42(2), 457–486. doi:10.1257/0022051041409075
- Bolton, G., Katok, E., & Ockenfels, A. (2004). How Effective Are Electronic Reputation Mechanisms? An Experimental Investigation. *Management Science*, 50(11), 1587–1602. doi:10.1287/mnsc.1030.0199
- Brown, J., & Morgan, J. (2006). Reputation in online markets: The market for trust. *California Management Review*, 49(1), 61–81. doi:10.2307/41166371
- Bruce, N., Haruvy, E., & Rao, R. (2004). Seller rating, price and default in online auctions. *Journal of Interactive Marketing*, 18(4), 37–50. doi:10.1002/dir.20021
- Cabral, L., & Hortacsu, A. (2010). The dynamics of seller reputation: Evidence from eBay. *The Journal of Industrial Economics*, 58(1), 54–78. doi:10.1111/j.1467-6451.2010.00405.x
- Caillaud, B., & Mezetti, C. (2004). Equilibrium reserve prices in sequential ascending auctions. *Journal of Economic Theory*, 117(1), 78–95. doi:10.1016/j.jet.2003.11.002
- Dellarocas, C. (2003). The digitalization of word-of-mouth: Promise and challenges of online reputation measures. *Management Science*, 49, 1407–1424. doi:10.1287/mnsc.49.10.1407.17308
- Dewally, M., & Ederington, L. (2006). Reputation, certification, warranties, and information asymmetries: Lessons from the online comic book market. *The Journal of Business*, 79(2), 693–730. doi:10.1086/499169
- Dewan, S., & Hsu, V. (2004). Adverse selection in electronic markets: Evidence from online stamp auctions. *The Journal of Industrial Economics*, 52(4), 497–516. doi:10.1111/j.0022-1821.2004.00237.x

- Dini, F., & Spagnolo, G. (2009). Buying reputation on eBay: Do recent changes help? *International Journal of Electronic Business*, 7(6), 581–598. doi:10.1504/IJEB.2009.029048
- Eaton, D. H. (2002). *Valuing information: Evidence from guitar auctions on eBay*. Working Paper. Murray, KY: Murray State University.
- eBay. (1996). *Feedback Forum: Founder's Letter*. Retrieved March 15, 2016 from: <http://pages.ebay.com/services/forum/feedback-foundersnote.html>
- eBay. (2016). *eBay Inc. Reports Fourth Quarter and Full Year 2015 Results*. Retrieved February 19, 2016 from: <https://investors.ebayinc.com/releasedetail.cfm?ReleaseID=952024>
- eBay Inc. (2002). *eBay Announces Fourth Quarter and Year End 2001 Financial Results*. Retrieved January 23, 2014, from http://investor.ebayinc.com/financial_releases.cfm
- eBay Inc. (2014). *eBay Reports Fourth Quarter and Full Year 2013 Results*. Retrieved January 23, 2014, from http://investor.ebayinc.com/financial_releases.cfm
- Houser, D., & Wooders. (2006). Reputation in auctions: Theory, and Evidence from eBay. *Journal of Economics & Management Strategy*, 15(2), 353–369.
- Jin, G. Z., & Kato, A. (2008). Price, quality, and reputation: Evidence from an online field experiment. *The Rand Journal of Economics*, 37(4), 983–1005.
- Katkar, R., & Reiley, D. (2006). Public versus secret reserve prices in eBay auctions: Results from a pokémon field experiment. *Advances in Economic Analysis & Policy*, 6(2), 1–23.
- Klemperer, P. D. (1999). Auction theory: A guide to the literature. *Journal of Economic Surveys*, 13, 227–286.
- Laffont, J., & Maskin, E. (1980). Optimal reservation price. *Economics Letters*, 6, 309–313.
- Levin, D., & Smith, J. (1996). Optimal reservation prices in auctions. *The Economic Journal*, 106, 1271–1283.
- Livingston, J. A. (2005). How valuable is a good reputation? A sample selection model of internet auctions. *The Review of Economics and Statistics*, 87(3), 453–465.
- McDonald, C. G., & Slawson, V. C. Jr. (2002). Reputation in an internet auction market. *Economic Inquiry*, 40(4), 533–650.
- Melnik, M. I., Xu, & Alm. (2009). The choice of opening prices on eBay. *The Manchester School Journal*, 77(4), 411–429.
- Melnik, M. I. (2011). Listing fees and their role on eBay: Evidence from a natural experiment. *Journal of Internet Commerce*, 10(4), 1–21.
- Melnik, M. I. (2015). The Rise of eBay: Confronting the Challenges of Asymmetry of Information and Competition. In T. Tsiakis (Ed.), *Trends and Innovations in Marketing Information Systems* (pp. 293–307). Hershey, PA: IGI Global.
- Melnik, M. I., & Alm, J. (2002). Does a seller's reputation matter? Evidence from eBay auctions. *The Journal of Industrial Economics*, 50, 337–349.
- Melnik, M. I., & Alm, J. (2005). Seller reputation, information signals, and prices for heterogeneous coins on eBay. *Southern Economic Journal*, 72(2), 305–328.
- Melnik, M. I., & Richardson, P. (2011). How seller and product characteristics impact promotion on eBay. *Journal of Promotion Management*, 17(4), 426–437.
- Myerson, R. (1981). Optimal auction design. *Mathematics of Operations Research*, 6, 58–73.
- Pavlou, P., & Dimoka, A. (2006). The nature and role of feedback text comments in online marketplaces: Implications for trust building, price premiums, and seller differentiation. *Information Systems Research*, 17(4), 392–414.

Reiley, D. H. (2000). Auctions on the internet: What's being auctioned and how? *The Journal of Industrial Economics*, 48, 227–252.

Reiley, D. H. (2006). Field experiments on the effects of reserve prices in auctions: More Magic on the Internet. *The Rand Journal of Economics*, 37(1), 195–211.

Reiley, D. H., Bryan, D., Prasad, N., & Reeves, D. (2007). Pennies from eBay: The determinants of price in online auctions. *The Journal of Industrial Economics*, 55(2), 223–233.

Resnick, P., & Dellarocas, C. (2003). *Online Reputation Mechanisms: A Roadmap for Future Research*. Technical report, First Interdisciplinary Symposium on Online Reputation Mechanisms. Retrieved January 23, 2014, from: <http://ccs.mit.edu/dell/symposium.html#publications>

Resnick, P., Zeckhauser, R., Swanson, J., & Lockwood, K. (2006). The value of reputation on eBay: A controlled experiment. *Experimental Economics*, 9(2), 79–101.

Riley, J., & Samuelson, W. (1981). Optimal auctions. *The American Economic Review*, 71, 381–392.

Standifird, S. (2001). Reputation and e-commerce: eBay auctions and the asymmetrical impact of positive and negative ratings. *Journal of Management*, 27(3), 279–295.

U.S. Department of the Census. (2013). *The 2011 E-commerce multi-sector E-Stats report*. Retrieved January 23, 2014, from: <http://www.census.gov/econ/estats/>

Wolford, J. (2013). eBay Seller Sues Buyer Over Negative Feedback. *WebProNews*. Retrieved January 23, 2014, from: <http://www.webpronews.com/ebay-seller-sues-buyer-over-negative-feedback-2013-04>

Zeithammer, R. (2007). Optimal Selling in Dynamic Auctions: Adaptation versus Commitment. *Marketing Science*, 26(6), 859–867.

KEY TERMS AND DEFINITIONS

Asymmetry of Information: Different parties have different access to the relevant information.

English Auction: A rising price auction.

Feedback: Comments posted online.

Online Reputational Mechanism: A structured presentation of feedback about a particular user of a website.

Opening Price: The starting price in an auction.

Reserve Price: The lowest acceptable bid to the seller, also the lowest binding bid.

Willingness to Pay: The maximum value the buyer is willing to pay for a given unit of a product.

Retail Prices and E-Commerce

E**Jihui Chen***Illinois State University, USA*

INTRODUCTION

In the pre-Internet era, consumers relied on media such as Sunday newspapers and flyers for product and price information. Such search process is time-consuming and unlikely to be exhaustive. The existence of incomplete information leads to price dispersion in the marketplace (Stigler, 1961). Recent advances in information technology have dramatically changed the manner in which consumers and businesses gather and transmit information.

Online shoppers enjoy enhanced search capability through effective tools. Search sites, such as Google and Yahoo!, become indispensable for comparison shopping. A recent comScore Media Matrix monthly qSearch™ analysis reports a total of 16.8 billion unique desktop search queries submitted in February 2016.¹ With the rising popularity of mobile devices, from smartphones to tablets, one would only expect a greater utilization of search engines.² Along with ever-increasing competition among online retailers, we would expect prices to converge in the new economy. However, an extensive literature on Internet prices has documented persistent price dispersion in various online markets. In this chapter, I review existing studies on the topic and discuss future research directions in light of recent developments with e-commerce.

BACKGROUND

The Internet provides an ideal setting for empirical studies with abundance of data.³ In this section, I review research on retailer prices on the Internet. In general, we may sort e-retailers into two categories:

web-based e-retailers (Dotcoms), such as eBay and Amazon, who exclusively conduct their business on the Internet and have no physical presence,⁴ and multi-channel retailers (MCR), such as Best Buy's online branch, which is an extension of the brick-and-mortar establishment.

Early studies usually compare online and offline prices on books, CDs, and DVDs sold in the U.S. (Bailey, 1998; Brynjolfsson & Smith, 2000) as these commodities were among the first available on the Internet. Gradually, the empirical literature expands to include a wide variety of products such as airfares (Clemons et al., 2002; Chen, 2006; Chellappa et al., 2011), automobiles (Zettelmeyer et al., 2006), pharmaceuticals (Stylianou et al., 2005), service supply products (Ghose & Yao, 2011), consumer electronics (Baye et al., 2004a, 2004b; Xing et al., 2004), groceries (Gan et al., 2007), and hotel rooms (Delos Santos et al., 2011), to name a few.

Recent studies also include international data. For example, Liu and Tang (2005) study the Chinese book market, Li et al. (2009) study the Australian DVD market, Mizuno et al. (2010) electronics in Japan, Koppius et al. (2004) Dutch flower auctions; Englmaier and Schmoller (2011) football game auctions at HATTRICK, U.K., Zhong and Ong (2011) prepaid phone cards at Taobao.com, China, and Richards et al. (2016) online groceries in U.K., to name a few.

The empirical literature often compares prices and price dispersion between online and brick-and-mortar sellers, and between Dotcoms and MCRs. While most have found lower online than offline prices and lower Dotcoms than MCR prices, there seems no general consensus regarding the level of dispersion across different distribution channels or types of e-retailers. However, it is clear that persistent price dispersion remains on the Internet.

DOI: 10.4018/978-1-5225-2255-3.ch248

Existing studies have developed several measures of price dispersion:⁵ In a given product market,

- **Price Range:** The difference between the highest and the lowest price.
- **Percent Price Range:** The ratio of price range to the lowest price.
- **Coefficient of Variation:** The ratio of the standard deviation to the average price.
- **Standard Deviation:** The fraction of average unit price: defined as the ratio of standard deviation to the mean unit price that is averaged across all products in a given market (Chiou and Pate, 2010, Table 4, p.302).
- **Gini Coefficient:**

$$Gini = 1 + \frac{1}{N} - \frac{2}{\lambda N^2} \sum_{i=1}^N (N+1-i) p_i$$

where p_i is the price of observation i , with $i=1,2,\dots, N$, λ is the mean price. (Gaggero and Piga, 2009).

- **Price Gap:** The price difference between the two lowest-priced firms (Baye et al., 2004a).
- **Atkinson Index:** Defined as

$$= \begin{cases} 1 - \frac{1}{\lambda} \left(\frac{1}{N} \sum_{i=1}^N p_i^{1-\phi} \right)^{\frac{1}{1-\phi}} & \text{for } \forall \phi > 0 \text{ and } \phi \neq 1 \\ 1 - \frac{1}{\lambda} \left(\prod_{i=1}^N p_i \right)^{\frac{1}{N}} & \text{for } \phi = 1 \end{cases}$$

where p_i is the price of observation i , with $i=1,2,\dots, N$; λ is the mean price; and ϕ is the choice parameter (Gaggero and Piga, 2009, p.8, footnote 14).

- **Entropy Index:** Defined as

$$= \begin{cases} \frac{1}{N} \left(\sum_{i=1}^N \left[\left(\frac{p_i}{\lambda} \right)^\phi - 1 \right] \right) & \text{for } \forall \phi > 0 \text{ and } \phi \neq 1 \\ \frac{1}{N} \sum_{i=1}^N \left(\frac{p_i}{\lambda} \right)^\phi \ln \frac{p_i}{\lambda} & \text{for } \phi = 1 \text{ or } \phi = 0 \end{cases}$$

where p_i is the price of observation i , with $i=1,2,\dots, N$; λ is the mean price; and ϕ is the choice parameter (Gaggero and Piga, 2009, p.8, footnote 14).

Explaining E-Retail Prices

The literature has documented a number of explanations for pricing issues observed in various markets, which I summarize below:

- **Branding/Reputation:** Because consumers have to submit the payment before receiving an order, trust plays a crucial role in online shopping. Naturally, risk-averse online shoppers prefer more reputable stores (Smith and Brynjolfsson, 2001). Thus, consumer awareness and sensitivity to branding allow some sellers to charge premia (Baylis and Perloff, 2002; Dinlersoz and Li, 2006), resulting in price dispersion (Chen and Hitt, 2002).⁶
- **Channel Substitution:** As sellers expand to the online channel, various coordination issues arise. While there is some evidence of cannibalization of the Internet channel in the publishing industry (Gentzkow, 2007), the actual outcome varies across product markets, and probably over time as well. For instance, Goolsbee (2001) finds evidence of channel conflict when consumers buy computers between channels; Pozzi (2013) estimates a 13% rise in overall revenue resulting from the establishment of the Internet channel by a supermarket chain, with little impact on traditional sales; Overby and Forman (2015) show that the

diffusion of online sales reduce geographic price dispersion in the wholesale used car market.

- **Competition:** The level of competition may determine sellers' pricing strategies. With only a handful of competitors, high prices can be sustained through tacit collusion. As more sellers join the competition, it would drive down the price as well as price dispersion. Evidence from various online markets suggests that competition leads to lower prices (Baye et al., 2004b; Haynes and Thompson, 2008) and lower price dispersion (Nelson et al., 2007; Gaggero & Piga, 2011). Using ticket-level data, Chakrabartya and Kutlub (2014) find an S-shaped relationship between price dispersion and market concentration (measured by inter-firm, inter-flight, and frequency competitions, respectively).
- **Consumer Heterogeneity:** The existence of consumer heterogeneity allows sellers to engage in price discrimination (Baye and Morgan, 2001; Matsumoto and Spence, 2016). For example, sellers may offer promotions (e.g., e-coupons and rebates) to attract price-sensitive consumers, while charging the full price to loyal customers (Grewal et al., 2011). In recent airline studies, Piga and Bachis (2011) identify different currencies as an effective tool to discriminate travelers residing in different regions, and Kutlu (2015) argues that limited consumer memory may also lead to price dispersion.
- **Dynamic Pricing:** Market efficiency, reduced menu costs, and effective monitoring capability allow e-retails to adjust prices fairly easily according to changing market conditions (Kannan and Kopalle, 2001; Jayaraman and Baker, 2003). Escobari (2012) and Alderighi, Nicolini, and Piga (2015) both study the dynamic pricing of seat inventories, using the U.S. and European data, respectively.
- **Oligopoly Strategies:** Profit-maximizing sellers engage in mixed pricing strategies to avoid fierce competition in the homogeneous product market (Baye and Morgan, 2001; Yan, 2008). This randomization strategy effectively prevents both buyers and competing sellers from figuring out where the lowest price is. Periodical sales are good examples. Sometimes, e-retails engage in product differentiation (Clemons et al., 2002; Ba et al., 2012).
- **Market Efficiency:** Sengupta and Wiggins (2014) study transaction data to compare airfares purchased online and offline and claim that more efficient shopping contributes to the observed lower online fares and thus greater online dispersion. In another airline study, Dana and Orlov (2014) find online purchase reduces market frictions, which in turn improves airlines' operation efficiency, defined by load factors.
- **Market Maturity:** Online sellers' pricing strategies may evolve over time. To capture the effect of evolution on firms' strategies, studies usually collect data at different points in time to examine the effect of market evolution (Pan et al., 2003). Chen's (2006) findings contrast those of Clemons et al. (2002), suggesting price convergence in the online airfare market. Taking an innovative approach, Bock et al. (2007) use data from U.S. and China to proxy different stages of e-commerce maturity. They conclude that more matured online markets are associated with lower prices and lower price dispersion.
- **Price Partitioning:** E-commerce encourages innovative pricing strategies, such as flexible price structure. One popular practice is to separate surcharges, such as shipping and handling fees, from the base price (Xia and Monroe, 2004; Gumus et al., 2013). Such a strategy allows e-retailers to further differentiate from each other. Even if they charge the same price for an

item, e-retailers may still soften competition by varying surcharges. As a result, the final price can be quite different depending on where a buyer shops. For instance, “free-shipping” with minimum purchase is an effective way to increase demand (Brynjolfsson and Smith, 2000; Dinlersoz and Li, 2006).

- **Quality:** Some studies relate price dispersion to product quality. Wolff (2015) examines the online diamond market and finds a positive relationship between dispersion and quality of these gemstones. In an airline panel study, Ater and Orlov (2015) show that flight quality (measured by both flight times and on-time performance) deteriorate as more travelers obtain Internet access.
- **Seller Heterogeneity:** Price dispersion may also be due to differences in seller types. For example, consumers may prefer MCRs to Dotcoms in terms of post-sale services (e.g., returns and exchanges). Carlton and Chevalier (2001) argue that internalization of free-riding problems leads to higher online prices at MCRs. The geographic location of e-retailers also matters in setting prices. In a cross-country study, Baye et al. (2006) find that the average price for consumer electronics sold in the Euro zone rises by 3% after the introduction of Euro in 2000. On price stickiness, Lunnemann and Wintr (2011) argue that price changes occur less often in the U.S. than the selected EU countries.
- **Taxation:** E-retailers are required by law to collect sales tax in the states where they have a physical presence (Anderson et al., 2010). Depending on the shipping address, online shoppers may pay different final prices pertaining to sales tax for otherwise an identical item from the same e-retailer. Recent studies show evidence of cross-border purchases driven by lower sales tax rates available (Ballard

and Lee, 2007; Ellison and Ellison, 2009; Hortaçsu, Martínez-Jerez, and Douglas, 2009; Goolsbee, Lovenheim, and Slemrod, 2010). Using a 4% sales tax cut in New York City as a natural experiment, Hu and Tang (2014) find Internet and catalog sales decline by 15%, although the effects vary across different types of consumers, products, and channels.

FUTURE RESEARCH DIRECTIONS

As the literature moves towards maturity, researchers began to explore new areas relating to e-commerce. Here I discuss several potential directions for future research.

As much has changed during the last decade, the literature on online price and price dispersion is in need for new evidence. It would be interesting to re-investigate the topic using most recent data, especially in the international markets. For example, Navarro et al. (2015) study online price and price dispersion of airfares on the Madrid–New York route from a new perspective, i.e., economic crisis. Furthermore, recent research has turned to transaction data (Sengupta and Wiggins, 2014; Zhao et al., 2015) and structural estimation (Baye et al., 2011; Pozzi, 2012), and still have ample room for further developments in both U.S. and overseas markets.

Understanding consumer behavior is another potential topic of interest. Future research may study consumer pre-purchase search behavior using browsing information, not self-reported surveys (Johnson et al., 2004), and then test whether the result of limited consumer search still prevails today. Understanding the competitive implications of the interactions between search and targeting online by incorporating fine aspects of consumer search (Chen and Sudhir, 2004) should also be useful. In addition, surcharges are non-refundable for rebates or price-matching requests, and, depending on the amount, such surcharges may affect consumer’s

decision regarding refund request, which in turn influences post-purchase behavior (Xia and Monroe, 2004).

Recently, the Marketplace Fairness Act, or Internet taxation, provides another potential avenue for future research. From early days of e-commerce, sales tax rates have been positively affected online purchase, with a declining effect over time (Alm & Melnik, 2005; Anderson et al., 2010; Einav, Knoepfle, Levin & Sundaresan, 2014). However, due to data limitation, existing studies focus on the impact of Internet taxation on the propensity of online purchase, future work may investigate the impact on the magnitude of online spending (Alm & Melnik, 2005).

Finally, the border between online and offline channels becomes increasingly blurry, giving the rise of hybrid operations in retailing. For example, Amazon recently launched its first brick-and-mortar bookstore in Seattle, WA, in late 2015, with more physical locations in the plan. Incidentally, China's Alibaba.com established its first physical store in Tianjin, China, in early 2016.⁷ Such a new trend in e-commerce is to provide consumers an integrated shopping environment, through which the industry has offered an avenue for further research as well.

CONCLUSION

This article presents a useful reference on recent research on Internet price dispersion, and aims to better prepare various interest groups for future challenges and opportunities in the new economy.

REFERENCES

- Alderighi, M., Nicolini, M., & Piga, C. A. (2015). Combined effects of load factors and booking time on fares: Insights from the yield management of a low-cost airline. *The Review of Economics and Statistics*, 97(4), 900–915. doi:10.1162/REST_a_00451
- Alm, J., & Melnik, M. I. (2005). Sales taxes and the decision to purchase online. *Public Finance Review*, 33(2), 184–212. doi:10.1177/1091142104267929
- Anderson, E. T., Fong, N. M., Simester, D. I., & Tucker, C. E. (2010). How sales taxes affect customer and firm behavior: The role of search on the Internet. *JMR, Journal of Marketing Research*, 47(2), 229–239. doi:10.1509/jmkr.47.2.229
- Ater, I., & Orlov, E. (2015). The effect of the Internet on performance and quality: Evidence from the airline industry. *The Review of Economics and Statistics*, 91(1), 180–194. doi:10.1162/REST_a_00442
- Ba, S., Stallaert, J., & Zhang, J. (2012). Online price dispersion: A game theoretic perspective and empirical evidence. *Information Systems Research*, 23(2), 575–592. doi:10.1287/isre.1110.0353
- Bailey, J. P. (1998). *Electronic commerce: Prices and consumer issues for three products: books, compact discs, and software*. Organization for Economic Co-Operation and Development, OCDE/GD(98)4.
- Ballard, C. L., & Lee, J. (2007). Internet purchases, cross-border shopping, and sales taxes. *National Tax Journal*, 60(4), 711–725. doi:10.17310/ntj.2007.4.02
- Baye, M. R., Gao, X., & Morgan, J. (2011). On the optimality of clickthrough fees in online markets. *The Economic Journal*, 121(November), 340–367. doi:10.1111/j.1468-0297.2011.02468.x
- Baye, M. R., Gatti, R., Kattuman, P., & Morgan, J. (2006). Did the Euro foster online price competition? Evidence from an international price comparison site. *Economic Inquiry*, 44(2), 265–279. doi:10.1093/ei/cbj015
- Baye, M. R., & Morgan, J. (2001). Information gatekeepers on the Internet and the competitiveness of homogeneous product markets. *The American Economic Review*, 91(3), 454–474. doi:10.1257/aer.91.3.454

- Baye, M. R., Morgan, J., & Scholten, P. (2004a). Price dispersion in the small and in the large: Evidence from an Internet price comparison site. *The Journal of Industrial Economics*, 52(4), 463–496. doi:10.1111/j.0022-1821.2004.00236.x
- Baye, M. R., Morgan, J., & Scholten, P. (2004b). Persistent price dispersion in online markets. In D. Jansen (Ed.), *The New Economy*. Chicago: University of Chicago Press.
- Baylis, K., & Perloff, J. M. (2002). Price dispersion on the Internet: Good firms and bad firms. *Review of Industrial Organization*, 21(3), 305–324. doi:10.1023/A:1020415304514
- Bock, G., Lee, S., & Li, H. Y. (2007). Price comparison and price dispersion: Products and retailers at different Internet maturity stages. *International Journal of Electronic Commerce*, 11(4), 101–124. doi:10.2753/JEC1086-4415110404
- Brynjolfsson, E., & Smith, M. D. (2000). Frictionless commerce? A comparison of Internet and conventional retailers. *Management Science*, 46(4), 563–585. doi:10.1287/mnsc.46.4.563.12061
- Carlton, D. W., & Chevalier, J. A. (2001). Free riding and sales strategies for the Internet. *The Journal of Industrial Economics*, 49(4), 441–461. doi:10.1111/1467-6451.00157
- Chakrabartya, D., & Kutlub, L. (2014). Competition and price dispersion in the airline markets. *Applied Economics*, 46(28), 3421–3436. doi:10.1080/00036846.2014.931919
- Chellappa, R. K., Sin, R. G., & Siddarth, S. (2011). Price-formats as a source of price dispersion: A study of online and offline prices in the domestic US airline market. *Information Systems Research*, 22(1), 83–98. doi:10.1287/isre.1090.0264
- Chen, J. (2006). Differences in average prices on the Internet: Evidence from the online market for air travel. *Economic Inquiry*, 44(4), 656–670. doi:10.1093/ei/cbj040
- Chen, P., & Hitt, L. M. (2002). Measuring switching costs and the determinants of customer retention in Internet-enabled businesses: A study of the online brokerage industry. *Information Systems Research*, 13(3), 255–274. doi:10.1287/isre.13.3.255.78
- Chen, Y., & Sudhir, K. (2004). When shopbots meet emails: Implications for price competition on the Internet. *Quantitative Marketing and Economics*, 2(3), 233–255. doi:10.1023/B:QMEC.0000037078.02232.76
- Chiou, L., & Pate, J. (2010). Internet auctions and frictionless commerce: Evidence from the retail gift card market. *Review of Industrial Organization*, 36(May), 295–304. doi:10.1007/s11151-010-9248-6
- Clemons, E. K., Hann, I., & Hitt, L. M. (2002). Price dispersion and differentiation in online travel: An empirical investigation. *Management Science*, 48(4), 534–549. doi:10.1287/mnsc.48.4.534
- Dana, J. D. Jr, & Orlov, E. (2014). Internet penetration and capacity utilization in the US airline industry. *American Economic Journal. Microeconomics*, 6(4), 106–137. doi:10.1257/mic.6.4.106
- De los Santos, B., Kosova, R., & Koulayev, S. (2011). *Estimating price discrimination in the online distribution channel*. Working Paper.
- Dinlersoz, E. M., & Li, H. (2006). The shipping strategies of internet retailers: Evidence from internet book retailing. *Quantitative Marketing and Economics*, 4(4), 407–438. doi:10.1007/s11129-006-9010-4
- Einav, L., Knoepfle, D., Levin, J., & Sundaresan, N. (2014). Sales taxes and Internet commerce. *The American Economic Review*, 104(1), 1–26. doi:10.1257/aer.104.1.1
- Ellison, G., & Ellison, S. F. (2009). Tax sensitivity and home state preferences in Internet purchasing. *American Economic Journal. Economic Policy*, 1(2), 53–71. doi:10.1257/pol.1.2.53

- Englmaier, F., & Schmoller, A. (2011). *Determinants and effects of reserve prices in Hatrick auctions*. Working Paper.
- Escobari, D. (2012). Dynamic pricing, advance sales, and aggregate demand learning in airlines. *The Journal of Industrial Economics*, 60(4), 697–724. doi:10.1111/joie.12004
- Gaggero, A. A., & Piga, C. A. G. (2009). *Airline market power and intertemporal price dispersion*. Working paper.
- Gaggero, A. A., & Piga, C. A. G. (2011). Airline market power and intertemporal price dispersion. *The Journal of Industrial Economics*, 59(4), 552–577. doi:10.1111/j.1467-6451.2011.00467.x
- Gan, L., He, S., Juang, T., & Tan, J. (2007). A comparative analysis of online grocery pricing in Singapore. *Electronic Commerce Research and Applications*, 6(4), 474–483. doi:10.1016/j.elerap.2007.02.006
- Gentzkow, M. (2007). Valuing new goods in a model with complementarity: Online newspapers. *The American Economic Review*, 97(3), 713–744. doi:10.1257/aer.97.3.713
- Ghose, A., & Yao, Y. (2011). Using transaction prices to re-examine price dispersion in electronic markets. *Information Systems Research*, 22(2), 269–288. doi:10.1287/isre.1090.0252
- Grewal, D., Ailawadi, K. L., Gauri, D., Hall, K., Kopalle, P., & Robertson, J. R. (2011). Innovations in retail pricing and promotions. *Journal of Retailing*, 87, S43–S52. doi:10.1016/j.jretai.2011.04.008
- Goolsbee, A. (2001). Competition in the computer industry: Online versus retail. *The Journal of Industrial Economics*, 49(4), 487–499. doi:10.1111/1467-6451.00159
- Goolsbee, A., Lovenheim, M. F., & Slemrod, J. (2010). Playing with fire: Cigarettes, taxes, and competition from the Internet. *American Economic Journal. Economic Policy*, 2(1), 131–154. doi:10.1257/pol.2.1.131
- Gümüş, M., Li, S., Oh, W., & Ray, S. (2013). Shipping fees or shipping free? A tale of two price partitioning strategies in online retailing. *Production and Operations Management*, 22(4), 758–776. doi:10.1111/j.1937-5956.2012.01391.x
- Haynes, M., & Thompson, S. (2008). Price, price dispersion and number of sellers at a low entry cost shopbot. *International Journal of Industrial Organization*, 26(2), 459–472. doi:10.1016/j.ijindorg.2007.02.003
- Hortacsu, A., Martinez-Jerez, F. A., & Douglas, J. (2009). The eography of trade in online transactions: Evidence from eBay and MercadoLibre. *American Economic Journal. Microeconomics*, 1(1), 53–74. doi:10.1257/mic.1.1.53
- Hu, Y. J., & Tang, Z. (2014). The impact of sales tax on internet and catalog sales: Evidence from a natural experiment. *International Journal of Industrial Organization*, (January), 32, 84–90.
- Jayaraman, V., & Baker, T. (2003). The Internet as an enabler for dynamic pricing of goods. *IEEE Transactions on Engineering Management*, 50(4), 470–477. doi:10.1109/TEM.2003.820134
- Johnson, E. J., Moe, W. W., Fader, P. S., Bellman, S., & Lohse, G. L. (2004). On the depth and dynamics of online search behavior. *Management Science*, 50(3), 299–308. doi:10.1287/mnsc.1040.0194
- Kannan, P. K., & Kopalle, P. K. (2001). Dynamic pricing on the Internet: Importance and implications for consumer behavior. *International Journal of Electronic Commerce*, 5(3), 63–83.
- Koppius, O. R., van Heck, E., & Wolters, M. J. J. (2004). The importance of product representation online: Empirical results and implications for electronic markets. *Decision Support Systems*, 38(2), 161–169. doi:10.1016/S0167-9236(03)00097-6
- Kutlu, L. (2015). Limited memory consumers and price dispersion. *Review of Industrial Organization*, 46(4), 349–357.

- Li, H., Tang, F., Huang, L., & Song, F. (2009). A longitudinal study on Australian online DVD pricing. *Journal of Product and Brand Management*, 18(1), 60–67. doi:10.1108/10610420910933371
- Liu, Y., & Tang, F. (2005). An empirical analysis on pricing patterns in Chinas online book market. *Journal of International Consumer Marketing*, 18(1/2), 117–136. doi:10.1300/J046v18n01_06
- Lünnemann, P., & Wintr, L. (2011). Price stickiness in the US and Europe revisited: Evidence from Internet prices. *Oxford Bulletin of Economics and Statistics*, 73(5), 593–621. doi:10.1111/j.1468-0084.2011.00657.x
- Matsumoto, B., & Spence, F. (2016). Price beliefs and experience: Do consumers' beliefs converge to empirical distributions with repeated purchases? *Journal of Economic Behavior and Organization*, 126, 243–254.
- Mizuno, T., Nirei, M., & Watanabe, T. (2010). Closely competing firms and price adjustment: Some findings from an online marketplace. *The Scandinavian Journal of Economics*, 112(4), 673–696. doi:10.1111/j.1467-9442.2010.01626.x
- Navarro, J. A., Martínez, M. A., & Trinqucoste, J. (2015). The effect of the economic crisis on the behaviour of airline ticket prices. A case-study analysis of the New York–Madrid route. *Journal of Air Transport Management*, (August): 47, 48–53.
- Nelson, R. A., Cohen, R., & Rasmussen, F. R. (2007). An analysis of pricing strategy and price dispersion on the Internet. *Eastern Economic Journal*, 33(1), 95–110. doi:10.1057/eej.2007.6
- Overby, E., & Forman, C. (2015). The effect of electronic commerce on geographic purchasing patterns and price dispersion. *Management Science*, 61(2), 431–453. doi:10.1287/mnsc.2014.1977
- Pan, X., Shankar, V., & Ratchford, B. T. (2003). The evolution of price dispersion in Internet retail markets. *Advances in Applied Microeconomics: Organizing the New Industrial Economy*, 12, 85–105. doi:10.2139/ssrn.437940
- Piga, C. A. G., & Bachis, E. (2011). Low-cost airlines and online price dispersion. *International Journal of Industrial Organization*, 26(6), 655–667.
- Pozzi, A. (2012). Shopping cost and brand exploration in online grocery. *American Economic Journal. Microeconomics*, 4(3), 96–120. doi:10.1257/mic.4.3.96
- Pozzi, A. (2013). The effect of Internet distribution on brick-and-mortar sales. *The Rand Journal of Economics*, 44(3), 569–583. doi:10.1111/1756-2171.12031
- Richards, T. J., Hamilton, S. F., & Allender, W. (2016). Search and price dispersion in online grocery markets. *International Journal of Industrial Organization*, 47(July), 255–281. doi:10.1016/j.ijindorg.2016.05.004
- Sengupta, A., & Wiggins, S. N. (2014). Airline pricing, price dispersion, and ticket characteristics on and off the Internet. *American Economic Journal. Economic Policy*, 6(1), 272–307. doi:10.1257/pol.6.1.272
- Smith, M. D., & Brynjolfsson, E. (2001). Consumer decision-making at an Internet shopbot: Brand still matters. *The Journal of Industrial Economics*, 49(4), 541–558. doi:10.1111/1467-6451.00162
- Stigler, G. J. (1961). The economics of information. *Journal of Political Economy*, 69(3).
- Stylianou, A., Kumar, R., & Robbins, S. (2005). Pricing on the Internet and in conventional retail channels: A study of over-the-counter pharmaceutical products. *International Journal of Electronic Commerce*, 10(1), 135–148.

Wolff, F. C. (2015). Does price dispersion increase with quality? Evidence from the online diamond market. *Applied Economics*, 47(55), 5996–6009. doi:10.1080/00036846.2015.1061645

Xia, L., & Monroe, K. (2004). Price partitioning on the Internet. *Journal of Interactive Marketing*, 18(4), 63–73. doi:10.1002/dir.20017

Xing, X. L., Tang, F. F., & Yang, Z. L. (2004). Pricing dynamics in the online consumer electronics market. *Journal of Product and Brand Management*, 13(6), 429–441. doi:10.1108/10610420410560316

Yan, R. (2008). Pricing strategy for companies with mixed online and traditional retailing distribution markets. *Journal of Product and Brand Management*, 17(1), 48–56. doi:10.1108/10610420810856512

Zettelmeyer, F., Morton, S., & Silva-Risso, H. (2006). How the Internet lowers prices: Evidence from matched survey and auto transaction data. *JMR, Journal of Marketing Research*, 43(2), 168–181. doi:10.1509/jmkr.43.2.168

Zhao, K., Zhao, X., & Deng, J. (2015). Online price dispersion revisited: How do transaction prices differ from listing prices? *Journal of Management Information Systems*, 32(1), 261–290. doi:10.1080/07421222.2015.1029397

Zhong, Z., & Ong, D. (2011). *Lemony prices: An online field experiment on price dispersion*. Working Paper.

ADDITIONAL READING

Baye, M. R., & Morgan, J. (2013). Pricing on the Internet. In S. N. Durlauf & L. E. Blume (Eds.), *The New Palgrave Dictionary of Economics*. Online Edition. Palgrave Macmillan. doi:10.1057/978-1-349-95121-5_2856-1

Goldfarb, A., Greenstein, S., & Tucker, C. (2015). *University of Chicago Press, Chicago*.

Lieber, E., & Syverson, C. (2012). Online versus offline competition. In M. Peitz & J. Waldfogel (Eds.), *Oxford Handbook of the Digital Economy*. Oxford University Press. doi:10.1093/oxford-hb/9780195397840.013.0008

KEY TERMS AND DEFINITIONS

Integrated Shopping: E-retailers establish physical stores as a way to enhance shopping experience of their customers.

Load Factor: Or passenger load factor, is often used by airlines to measure capacity utilization of their fleets.

Multichannel Retailers: Retailers conduct businesses through multiple distribution channels.

Price Discrimination: Firms charge different prices to different consumers for an identical product.

Price Dispersion: In a homogeneous product market, prices charged by different sellers are different.

Price Partitioning: Retailers divide the total product price into different components – a base price and surcharges including shipping and handling, taxes, and other fees.

Shopbot: A useful online search tool helps shoppers collect product information.

The Marketplace Fairness Act: A pending proposed legislation requires e-retailers to collect sales tax regardless of physical location.

ENDNOTES

¹ Source: comScore Release February 2016 U.S. Desktop Search Engine Rankings. Retrieved May 30, 2016, from <https://www.comscore.com/Insights/Rankings/comScore-Releases-February-2016-US-Desktop-Search-Engine-Rankings>.

² Source: Computer and Internet Use in the United States: 2013, American Community Survey Reports. Retrieved February 27,

2015, from <http://www.census.gov/content/dam/Census/library/publications/2014/acs/acs-28.pdf>.

³ In this literature, studies often use electronic agents to automate data collection (Clay et al., 2001; Baye et al., 2004).

⁴ Supporting services for these e-retailers including call centers and warehouses are not considered as part of the core business.

⁵ Baye and his colleagues define a set of Internet indices, including price gap and

value of information on their website (www.nash-equilibrium.com).

⁶ An extensive literature on online auctions studies the value of seller reputation (Bajari and Hortacsu, 2004; Resnick, Zeckhauser, Swanson, and Lockwood, 2006).

⁷ Source: “Why Would Amazon Open Physical Stores?” by Trefis Team, Forbes, February 11, 2016. Retrieved May 30, 2016, from <http://www.forbes.com/sites/greatspeculations/2016/02/11/why-would-amazon-open-physical-stores/#542a6a203a1a>

Social Commerce Using Social Network and E-Commerce

E**Roberto Marmo***University of Pavia, Italy*

INTRODUCTION

Social network is an online media that supports social interaction in real life. E-commerce regards buying and selling of products or services conducted over internet and other computer networks. While traditional E-commerce supports the transactional and informational aspects of online shopping, Social Commerce fulfils the social aspects of shopping. Combining the advantages offered by social network and e-commerce, it is possible to create a web business platform to generate and increase revenue by turning web visitors into customers. Social Commerce involves social media and user contributions to assist in the online buying and selling of products and services, to design and deliver better customer experiences with the help of technology. This chapter describes some technologies that can help define Social Commerce, it also discusses background, knowledge, challenges and critical factors necessary for successful business. Future trend are described.

BACKGROUND

Social Network is a social structure composed of individuals, organizations, company etc. which are connected by relationships and interactions. Web-based social networks are online communities that allow users to publish resources (personal data, photo, video, blog) and to establish relationships of a different type about business, entertainment, gaming, dating, etc. Usage and diffusion of social networking platform has been

increasing, with hundreds web platforms in the world collecting the information of more than one billion registered users. Therefore, today social networks are used intensely to communicate, share information, make decisions, and do business in many ways. Well-known social networking websites are: Facebook as generic social network, LinkedIn and Viadeo as business social network, Flickr about photo sharing and Google+ as solution from Google search engine. Thus, online social network is a relevant part of human life (Fu, Chen, Liu, & Wang, 2007; Goth, 2008) and it is truly the reflection of today's society.

Consumers generally want to ask their friends or experts about a shopping decision or want to share their new purchase with friends; therefore, conversations regarding shopping happen all the time. Consumers generally are communicating with each other via social network. For selling message and open up new sales opportunities, it is necessary to resound with customers and to take advantage of their social network channels by interacting with them, in order to send business messages to customers and their friends.

Electronic commerce, commonly known as e-commerce or eCommerce, is a type of industry where the buying and selling of products or services is conducted over electronic systems such as internet and other computer networks. E-commerce is not only limited to online sales, but also covers: real-time management of product availability, online payment, delivery tracking and after-sales service. E-commerce sites are online stores which have at least the following elements at the front-office (customer) level:

DOI: 10.4018/978-1-5225-2255-3.ch249

- Electronic catalog listing all products for sale, price and availability;
- Search engine to easily locate a product using search criteria;
- Virtual cart, to trace the purchases of the client along the shopping path;
- Secure online payment via secure transaction;
- Order tracking system.

A back-office system allows the online dealer to organize electronic catalog, modify prices, add or remove product, handle client orders.

SOCIAL COMMERCE AND SOCIAL SHOPPING

Social Commerce is a subset of e-commerce that involves using social media, to connect, listen, understand, and engage people to improve the user shopping experience and to assist in the online buying and selling of products and services. In other words, these applications merge online shopping and social networking (Tedeschi 2006). The social commerce website are sites designed specifically to support social interactions while online consumers shop.

This is the latest solution to a chronic problem for online retailers and shoppers: many shoppers aren't sure what to buy, but they know they won't find it on the sites of mainstream retailers (Tedeschi, 2006).

Social Shopping is another frequent synonyms, but Social Commerce regards online sellers combining social network and e-commerce, social shopping regards collaborative activities by online shoppers. Social Shopping regards how customers use a social commerce website to shop, so it plays to a natural extension of our everyday behaviors, where we go shopping with friends, seek advice from friends and then decide based, at least in part, their feedback. In other words, the distinction

between Social Shopping and Social Commerce is that while social shopping connects customers, Social Commerce connects sellers and customers (Andrew, 2010).

The online shopper can be examined as a prospective user of an emerging social commerce platform. It is the concept of word of mouth from social media marketing (Singh, 2012), aligned to e-commerce.

The term Social Commerce was introduced by Yahoo! in November 2005 (available on line at www.ysearchblog.com/2005/11/14/social-commerce-via-the-shoposphere-pick-lists/) to describe a set of online collaborative shopping tools such as shared pick lists, user ratings and other user-generated content-sharing of online product information and advice. Important dates about history of Social Commerce are visualized at website http://socialtimes.com/social-commerce-infographic-2_b84120.

Social Commerce gives retailers opportunity to leverage benefits of interacting with consumers, discussing directly goods and their features, spreading information about retailer. Within this framework, customers have access to social knowledge and experiences to support users in better understanding their online purchase purposes, and making informed and accurate purchase decisions (Dennison et al. 2009).

A systematic analysis to 64 papers is discussed on (Salvatori, 2014), to summarize the existing evidence concerning the social commerce and outline some open challenges.

Social Principles

Social psychology explains that there are rules that people follow to navigate the decisions that need to be made. Six principles of influence that human beings use in decision-making can be identified (Cialdini, 2006), so it is possible to identify six social principles that can drive sales to Social Commerce success:

1. **Scarcity:** Less is more is perceived as more valuable, so offering limited-time discounts, items with limited availability, limited access and exclusivity;
2. **Affinity:** Shoppers are likely to buy items that friends or VIP figures have recommended;
3. **Reciprocity:** Shoppers are more likely to buy items that they have sampled;
4. **Popularity:** Shoppers follow the crowd, so they are likely to buy an item with reviews or higher ratings;
5. **Authority:** Shoppers are likely to purchase an item recommended by a leader or an expert;
6. **Consistency:** When shoppers aren't sure about a shopping decision, they do what they have been doing based on their buying history.

Tools about social activities can allow sharing the user shopping experience:

- Ratings;
- Recommendations;
- Forum;
- Group buy;
- Purchase sharing.

Online shopping is a voluntary and hedonic activity, and users participate because they are intrinsically motivated. The experience often offers entertainment and fun, which users have been found to appreciate (Mathwick, 2002).

Main Steps

There are main steps in successful Social Commerce strategy:

1. Searching information about product through tags, bookmarks, keywords, ranking created from other people on social network, blog, other websites;
2. Purchasing from sellers who are linking the social presence with e-commerce site, the

online transaction and payment systems are based on e-commerce platform;

3. Sharing information that comments or recommendation the shopping experience with other people, adding social features such as social plugin by Facebook platform.

In this way, it is possible to apply word-of-mouth to e-commerce, by sharing purchasing decisions before, during and after buying. Solutions exist for brands large and small, making their products more available and more convenient for customers.

Customer's roles vary across platforms and can range from only customer, customer generating content (e.g., product reviews and recommendations, consumer-generated media) to being sellers and curators of online stores.

A practical road map for monetizing Social Commerce investment is required. A strong and measurable strategy, a sort of social action business plan, can be obtained in this way:

1. Understand social networks, social business and Social Commerce;
2. Define current status quo, desired goals, timeline;
3. Set up e-commerce web site reached by links from social networking platform, or set up social network pages with e-commerce module;
4. Define your customers, exploring demographics of your target market and learn where and how customers are communicating with each about buying decisions;
5. Build customer base through social media marketing and appropriate use of most popular social networking platform that appeal to your target audience to full advantage;
6. Define promotional schemes so that users recommend other;
7. Define which social platform works for which objectives, define creative ways to share buying content;

8. Choose between popular social networking platforms or specific platform for social commerce;
9. Optimize social network pages to attract clicks and customers;
10. Monitor online reputation, how seller is viewed by customers, addressing contents which are damaging to business;
11. Measure results closely as any other marketing activity;
12. Improve customer service;
13. Learn from the mistakes of others.

Regarding (3), here are two technical strategies: social media on e-commerce platforms, e-commerce on social media platforms. First case regards helping people connect where they buy. It is necessary sharing purchase information across the open web and social network by placing in e-commerce products pages social plugin provided by Facebook: share, like and recommendation buttons. Also, integration with chat is desirable, so friends can shop together and buyers can talk with vendor. Second case regards the availability of e-commerce platforms which makes it possible to buy without leaving the social network they are using to talk to their friends, such as e-commerce tab on Facebook Page.

Regarding (8), on the internet there are online services created specifically to make Social Commerce, traditional social networking platforms have been created for different purposes and were later adapted for Social Commerce.

Regarding (9), some suggestions about most popular social networking platform are explained in following sections.

Regarding (11), as a positive result of investment, a seller wants to increase traffic, conversion rate (percentage of website visitors who buy something on the site) and average order value. Return on Investment (ROI) measures the effect or action of social media on sales, that is, as formula: gain from investment/ cost of investment. A high ROI means the investment gains compare favorably to investment cost. It is hard to evaluate and measure

the benefits brought by social commerce especially from financial perspective (Jussila et al, 2011; Turban et al, 2010). As example, a Facebook Page of company has successfully attracted 1,000 visitors to its e-commerce website after implementing social commerce in its business, but this performance indicator can only demonstrate that 1,000 people know information about the company and its product, there is no guarantee that all or part of them will turn into buyers.

The number of likes on Facebook Page and number of post shared between friends are not particularly meaningful to calculate ROI. In order to track results, it is necessary to define and measure key performance indicators as follows: analytics from popular social networking platform such as Facebook Insights which provides developers and Page owners with metrics about their content, how many users recommend other users, traffic coming from recommendations, sales coming from recommendations.

By understanding and analyzing trends about usage and demographics as well as consumption and creation of content, sellers can be better equipped to improve business and create better experiences on Facebook.

Regarding (12), in order to create a business connection which leads to sell, each seller must listen and quickly respond, handling service issues instantly using social networks. In this way, customer service initiatives can be named as Social CRM (Customer Relationship Management), the next frontier for organizations that want to optimize the power of social interactions to get closer to customers.

Below some technologies that can help business based on social commerce are introduced.

Facebook Page

Facebook is the most popular online social networking service. F-commerce is derived from e-commerce, suggests the use of Facebook as a platform for facilitating and executing sales transactions, either on Facebook itself or exter-

nally using Facebook Open Graph. Online floral retailer 1-800-Flowers has the first online store on Facebook at 11.50 a.m. July 8, 2009.

Facebook provides the Page to build a closer relationship between sellers and customers, as detailed on www.facebook.com/about/pages, it also allows to create global pages with regional featured products. The online store created on Facebook Page lets customers shop and complete purchases without leaving Facebook. An example regards Blomming, a personal Social Commerce made by humans, easily discover and shop unique items in an always updated marketplace, in www.facebook.com/blomming using tab Demo Shop available under the cover photo.

Some suggestions about best business practices using Facebook Page:

- Publish limited-time offers via wall and tabs;
- Highlight a product that appeals to a large number of fans;
- Use photo tab to showcase product images and a gallery of related products;
- Use wall and status updates to post product information, pin a new post to the top of page each week so people notice what's important, attach an image of the product that appears in fans' wall;
- Include a link in every post, so interested users can directly go to the point of purchase;
- Give users a voice to comment on the items they like best, encourages conversation around product on the wall;
- Create tabs to offer an incentive suitable for fan base growth, such as to provide exclusive offers to its fans who clicked on the button Like;
- Allow users to state preferences on products by creating polls in which users can vote for their favourites items;
- Provide users the power to express themselves creates a feeling of value, such as allowing download and publications of photo.

As described on a study about Thai consumers' intention to buy toward Facebook commerce is described on (Suraworachet, 2012), belief in people who like a Facebook Page, belief in people who like a photo of an item on e-commerce catalog and belief in friends who like a Facebook Page have significant impact on consumers' attitude.

Furthermore, attitude and perceived ease of use also have a significant effect on consumers' purchasing intention on e-commerce inside Facebook.

Pinterest

Pinterest is a pinboard-style photo-sharing website that allows users to create and manage theme-based image collections. Pinterest has found an optimal balance between inspirational browsing and shopping, making an effective distribution channel. It is possible to create branded pages for brands, stores and boutiques. If users are already discovering products through Pinterest but going off to merchant sites to transact, Pinterest should own that transaction and offer a consistent user experience. Pinterest offers a separate shopping channel (shop.pinterest.com), useful to focus on commerce.

Twitter

Twitter is an online social networking and micro-blogging service that enables users to send and read tweets, which are text messages limited to 140 characters. Twitter feed can be used to promote an individual product or to drive traffic to e-commerce website. It is useful to provide offers within time-sensitive campaigns. Optimizing times when the greatest number of followers see a tweet is critical to garnering the largest number of impressions. 140 character limit for each tweet requires to make offer and product descriptions short and direct. Checking out popular niche hashtags it is useful to find top influencer (a person who has upwards of a few thousand followers) related to keywords that describe product for sale. The influencer can tweet to their followers about e-commerce web

pages, this approach can drive a lot of people to check out what they're recommending and even make a purchase.

YouTube

YouTube is a video-sharing website. It is best served for creating awareness and interest about a campaign and initiatives. Some suggestions to create a video that encourages purchases: video can describes product manufacturing, video can suggests how to use an object, it is possible to create a branded channel, video can addresses a consumer pain point including music or comedy and paying attention to the open and close. It is important to place links within a video to suggest the path to point of purchase. By embedding videos outside of YouTube on the web it is possible to create a viral brand experience reaching many customers.

ISSUES AND PROBLEMS

The main doubts that must be discredited are that Social Commerce users are not in a mood for shopping when they socialize. Issues like privacy and fair use of data pose a formidable challenge, so the question regard whether Social Commerce is disruptive or intrusive. Social Commerce requires a wide variety of information to target and promote products, it requires analysis of customer preferences and tastes, it is not always possible to obtain similar information due to privacy setting and lack of data. Credit card safety and prevention of online fraud are also relevant problems, such as traditional e-commerce. Social networking platforms are not feeling with information and privacy security, so it is necessary a special effort to create a store where the user can buy with confidence and without getting in trouble.

A model that captures the essential features of the Social Commerce business is described in (Kim, 2013) to see whether Social Commerce is sustainable, and discuss conditions for stable evolution of the industry. The model is investigate

about the possibility that the tragedy of commons occurs in the industry due to an excessive competition in market share.

Predictable products like books, CDs, and electronics can be bought easily. But fashion lagged behind. Also in case of recommendations and suggestions from friends, consumers want a tactile, in-person experience when it comes to garments.

User graphic interface play a relevant rule, adding social tools requires re-design the web page and pay attention to create visual form having more buttons, icons, link ecc. So, customer can lose attention, feel confused and he rapidly moves to another website.

The price fairness perception had a significant influence on the repurchase intention. The illusion of control did not play a significant role in the point that the Social Commerce buying experience affected the price fairness perception. As a result of analysing the influences (Kim, 2013) on the price fairness perception by classifying the types of price fluctuations, there was a significance influence only in case of unfavourable type to consumers.

SOLUTIONS AND RECOMMENDATIONS

Sellers are directly connected by online social networks with customers, so today sellers are perceived as individuals instead of firms. Social Commerce cannot be arranged using only internet technology, because it concerns people influencing people in case of shopping online. Therefore, it is necessary to acquire skills on science of human relationships and social psychology that deals with how people think about, influence and relate to one another (Goleman, 2007).

Trust is particularly important in Social Commerce transactions, due to the difficulty of taking legal action in cases of dispute or fraud and the potential for conflicts caused by differences in culture and business environments. Trust is usually about trusting of recommendations from

friends and from other consumers. To deal with these issues, sellers need to establish high levels of trust with current and potential customers (Turban, 2012).

Therefore, sellers must be conversational, involving nurturing relationships, building long-standing trust with customers. In this way, it is possible to gain more prospects, convert them into buyers, and foster them as brand advocates for your business.

Today, mobile device is playing a much greater role in the average shopping experience (Hayden, 2014). To capitalize on immense m-commerce (mobile e-commerce) opportunity, sellers must develop an application (app) or specific version of website for mobile device, possibly including mobile payment solutions.

Each seller must help customers to promote his brand as they recommend products and services to their friends on preferred social networks, absolutely necessary to give to customers opportunity to share information about product with their social media contacts.

A social network between sellers in a large online Social Commerce marketplace can create a virtual shopping mall, because each seller creates his or her own shop, and network ties between sellers are directed hyperlinks between their shops. Allowing sellers to connect generates considerable economic value, the network's value lies primarily in making shops more accessible to customers browsing the marketplace, and the sellers who benefit the most from the network are not necessarily those who are central to the network but rather those whose accessibility is most enhanced by the network (Andrew, 2010).

In order to mitigate the risk of disrupting and irritating users who do not want to participate in social networks, a separate shopping channel is desirable to be adopted by this kind of user.

Shopping recommendation to customers can be based on style and preferences, these recommendations are computer generated, some are handpicked by friends, etc.

The design of a Social Commerce website should implement effective local social network features to connect with and engage its users, mediate their identities, and empower them in this rising participatory culture (Sun, 2006).

A model and a set of principles for Social Commerce design are introduced and design features are summarized as: usability, information quality, website quality, service quality and playfulness. The proposed and justified Social Commerce design model are based on: Individual, then Conversation, then Community, then Commerce. A specific table lists the design features as they apply or not to two target Social Commerce websites: Amazon and Starbucks Facebook.

As suggested in (Huang, 2013) management needs to identify its existing e-commerce and social networking applications and capabilities. They must then decide on the way to develop their Social Commerce strategy: to add social features to their e-commerce platform, or to add commerce features to their social network platform.

FUTURE RESEARCH DIRECTIONS

As Social Commerce scales, the biggest challenge will be surfacing signal buried in noise, so it is necessary new trend to create relevant content, newsfeed about shopping, specific guidelines for user interface design.

Technologies such as augmented reality will be combined with Social Commerce, allowing shoppers to visualize fashion items on themselves and solicit feedback from friends.

Future and emerging trends regards research model for investigating the impact of online communication of consumers, which produce social word of mouth on trust building mechanisms, increasing the level of trust on new products (Hajli, 2013).

One of the main reasons for participating in Social Commerce is to share and exchange information about shopping with other users.

Challenging issue regards framework to protect private information on social network (Carminati et al, 2009) and developing social customer relationship software taking care of privacy.

Social networking giants will develop competitive technological features to provide e-commerce in their platforms, in order to obtain a percentage of the price paid when a user of the social network buys something.

Technologies such QR codes and geo-tagging will bring social commerce to street commerce using mobile device.

Recommendation aims to recommend products from e-commerce websites to users at social networking sites in coldstart situations, a problem which has rarely been explored before. A major challenge is how to leverage knowledge extracted from social networking sites for cross-site cold-start product recommendation.

CONCLUSION

Today's consumer relies on input from a trusted community when making buying decisions, generally are communicating with each other via social network. Social Commerce is a business phenomenon that needs to be better understood, because it power commerce and drive sales combining social networking and e-commerce online platforms. This approach involves a lot of problems, so determining relevant factors and effective strategies to utilize social network for all kind of sellers is no easy.

Social Commerce, if approached properly, can be an extremely effective tool when it comes to the success of business, because it helps to boost e-commerce website traffic, reach out to a wider audience in shorter time and lower cost, increase brand awareness and it can increase product sales.

Due to wide variety of methods and strategies proposed which can be adopted, companies should look out for challenges and plan well when adopting social commerce into their existing business.

With new and upcoming technologies, the biggest change for Social Commerce marketers will be a shift in focus from branding to lead generation. they also ease the sales process by providing ways to effectively interact with customers.

REFERENCES

- Andrew, T. S., & Olivier, T. (2010). Deriving value from Social Commerce networks. *JMR, Journal of Marketing Research*, 47(2), 215–228. doi:10.1509/jmkr.47.2.215
- Carminati, B., Ferrari, E., & Perego, A. (2009). Security and Privacy in Social Networks. *Encyclopedia of Information Science and Technology*, 7, 3369-3376.
- Cialdini, R. (2006). *Influence: the psychology of persuasion*. New York, NY: HarperBusiness.
- Dennison, G., Bourdage-Braun, S., & Chetuparambil, M. (2009). Social Commerce defined. White paper #23747. Research Triangle Park, NC: IBM Corporation. Retrieved from <http://www-01.ibm.com/software/genservers/commerce/socialcommerce>
- Fu, F., Chen, X., Liu, L., & Wang, L. (2007). Social dilemmas in an online social network: The structure and evolution of cooperation. *Physics Letters. [Part A]*, 371(1-2), 58–64. doi:10.1016/j.physleta.2007.05.116
- Goleman, D. (n.d.). *Social Intelligence: The new science of human relationships*. New York: Bantam.
- Goth, G. (2008). Are social networking growing up? *IEEE Distributed Systems Online*, 9(2), 3. doi:10.1109/MDSO.2008.4
- Hajli, M., Khani, F., & Hajli, M. (2013). Establishing trust in Social Commerce through social word of mouth. *Proceedings of 7th International Conference on ECDC e-Commerce in Developing Countries*, 1-22. doi:10.1109/ECDC.2013.6556738

Hayden, T. (2014). *The mobile commerce revolution: business success in a wireless world*. Que Publishing.

Huang, Z., & Benyoucef, M. (2013). From e-commerce to Social Commerce: A close look at design features. *Electronic Commerce Research and Applications, Special Issue on Social Commerce Part 2*, 12(4), 246–259. doi:10.1016/j.elerap.2012.12.003

Jussila, J., Karkkainen, H., & Lein, M. (2011). Benefits of Social Media in business-to-business customer interface in innovation. *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*, 167–174. doi:10.1145/2181037.2181065

Kim, D. (2013). Under what conditions will Social Commerce business models survive? *Electronic Commerce Research and Applications, Special Issue on Social Commerce Part 2*, 12(4), 69–77. doi:10.1016/j.elerap.2012.12.002

Kim, Y., Lim, K., & Jung, N. (2013). A Social Commerce customers' price fairness perception affects their repurchase intention. *Proceedings of 15th International Conference on ICACT Advanced Communication Technology*, 1042–1048.

Mathwick, C. (2002). Understanding the online consumer: A typology of online relational norms and behavior. *Journal of Interactive Marketing*, 16(1), 40–55. doi:10.1002/dir.10003

Salvatori, L., & Marcantoni, F. (2015). Social Commerce: A Literature Review. *Proceedings of the 2015 Science and Information Conference*, 257–262. doi:10.1109/SAI.2015.7237152

Shen, J. (2012). Social Comparison, Social Presence, and Enjoyment in the Acceptance of Social Shopping Websites. *Journal of Electronic Commerce Research*, 13(3), 198–212.

Singh, S. (2012). *Social media marketing for dummies*. Hoboken, NJ: Wiley.

Sun, H. (2011). Designing for Social Commerce experience as cultural consumption. *Proceedings of the 4th international conference IDGD on Internationalization, Design and Global Development*, 402–406. doi:10.1007/978-3-642-21660-2_45

Suraworachet, W. (2012). The Study on the Effect of Facebook's Social Network Features toward Intention to Buy on F-commerce in Thailand. *Proceeding of 12th International Symposium on Applications and the Internet (SAINT)*, 245–250. doi:10.1109/SAINT.2012.46

Tedeschi, B. (2006). Like Shopping? Social Networking? Try Social Shopping. *New York Times*. Retrieved September 13, 2013 <http://www.nytimes.com/2006/09/11/technology/11ecom.html>

Turban, E., Bolloju, N., & Liang, T. (2010). Social commerce: an e-commerce perspective. *Proceedings of the 12th International Conference on Electronic Commerce: Roadmap for the Future of Electronic Business*, 33–42.

Turban, E., Strauss, J., & Lai, L. (2012). *Social Commerce*. New York, NY: Pearson Education.

ADDITIONAL READING

Baghdadi, Y. (2013). From e-Commerce to Social Commerce: A Framework to Guide Enabling Cloud Computing. *Journal of Theoretical and Applied Electronic Commerce Research*, 8(3), 12–38. doi:10.4067/S0718-18762013000300003

Chingning, W., & Zhang, P. (2012). The evolution of Social Commerce: An examination from the people, business, technology, and information perspective. *Communications of the AIS*, 31, 105–127.

Collier, M. (2012). *Social media commerce for dummies*. Hoboken, N.J.: Wiley.

Kang, Y. R. K., & Park, C. P. C. (2009). Acceptance Factors of Social Shopping. In *Proceedings of the 11th International Conference on Advanced Communication Technology*, 2155-2159.

Lawson, J., & Schepp, D. (2014). *Kick ass Social Commerce for e-preneurs: it's not about likes it's about sales*. BenBella Books. Marsden, P. *Social commerce: monetizing social media*. Retrieved April 2014 on http://digitalintelligencetoday.com/documents/Syzygy_2010.pdf

Marsden, P., & Chaney, P. (2012). *The Social Commerce handbook: 20 secrets for turning social media into social sales*. New York: McGraw-Hill.

Singh, S. (2012). *Social media marketing for dummies*. Hoboken, N.J.: Wiley.

Turban, E., Strauss, J., & Lai, L. (2012). *Social Commerce*. New York, NY: Pearson Education.

Web site retrieved May, 2014: www.socialmediatoday.com, www.thenextweb.com, www.businessinsider.com, www.bazaarvoice.com/research-and-insight/social-commerce-statistics, <http://blogs.gartner.com>

Xiaoling Sun, X., Zhao, & Y., Zhu, Q. (2012). Understanding the niche strategies adopted by social commerce websites. In *Proceedings of the 2012 International Conference on Information Management, Innovation Management and Industrial Engineering*, 418-421. doi:10.1109/ICIIM.2012.6339690

KEY TERMS AND DEFINITIONS

E-Commerce: Buying and selling of products or services conducted over internet and other computer networks.

Influencer: People who potentially have the power to convince others to stay or switch to other brand.

Mobile Social Commerce: Integration of social media with e-commerce on mobile platforms.

Social Commerce: An emerging approach of e-commerce that involves using social network to supports social interaction, to assist in the online buying, to facilitate discussion between seller and between buyer and seller or buyer and his friends.

Social CRM: The integration of social media into customer relationship management as tool focused on optimizing customer lifetime value.

Social Media Marketing: The process of gaining website traffic or attention through social media sites. Social media marketing programs usually center on efforts to create content that attracts attention and encourages readers to share it with their social networks.

Social Network Analysis: Mathematical technique developed to understand structure and behaviour between members of social system, to map relationships between individuals in social network.

Social Networking: Grouping of individuals into specific groups or communities.

Social Shopping: A natural extension of our everyday behaviors, where people go shopping with friends.

Trust: Psychological status of depending on another person or organization to achieve a planned goal.

Word-of-Mouth: Passing of information from person to person by oral or written communication.

An Update on Bitcoin as a Digital Currency



Cecilia G. Manrique

University of Wisconsin – La Crosse, USA

Gabriel G. Manrique

Winona State University, USA

INTRODUCTION

“The Evolution of Virtual Currencies: Analyzing the Case of Bitcoin,” a manuscript co-authored by Cecilia G. Manrique with Gabriel G. Manrique was recently published as a chapter in the book *Information and Communication Technologies in Public Administration*. It was written at the height of the recent bitcoin controversy when its value (in relation to the US dollar) underwent wild swings. It has been more than a year since the research for that chapter was conducted, and much has transpired since then affecting bitcoin’s acceptability as a virtual currency. It is the purpose of this paper to present an update on the status of bitcoin as a currency and to determine its stability and ability to become a real one.

This chapter will consist of four parts: a background based on a previous work, a review of what has been written about the topic to date, an update on the current status of bitcoin, and a future trends outlook on the acceptability of bitcoin as a currency. It is the purpose of this entry to shed further light on what has transpired regarding bitcoin since its very volatile and controversial period in 2013-2014 and to extrapolate on its possible future directions. Much of the controversy surrounding bitcoin at that time stemmed from several factors: its level of acceptability as a medium of exchange which is a crucial element for any aspiring currency, real or virtual; its various uses some of which had been said to facilitate criminal enterprises such as Silk Road; and its lack of stability raising questions about its suitability as a store of value. In its early stages,

bitcoin was applauded as a new and innovative currency because it could serve as a medium of exchange and a store of value with no bank and government intervention. By 2013 however, its reputation and seeming potential were ravaged by scandal, theft and turbulent prices. Has anything changed since then to restore the luster of bitcoin?

BACKGROUND

This encyclopedia entry comes on the heels of a recently published chapter entitled “The Evolution of Virtual Currencies: Analyzing the Case of Bitcoin” published in the book *Information and Communication Technologies in Public Administration*. The research and writing for it took place at the height of the bitcoin controversy. At that time in 2013-2014 the value of bitcoin skyrocketed and waned at such a volatile rate that it made for good fodder to many newspaper and magazine articles. In the month leading to bitcoin reaching its highest value

vis-à-vis the US dollar of \$1147 on December 4, 2013, bitcoin tripled in value only to decline to \$830 a month later, a decline in value of more than a quarter. By early April 2014 bitcoin had lost another 47%. As of this writing, bitcoin’s value seems to have settled in the \$250 range (BPI, 2015).

The Manrique and Manrique chapter took a look at the evolution of virtual currencies and focused attention upon bitcoin. It studied the case of bitcoin in terms of its potential as a legal tender taking into consideration the various char-

DOI: 10.4018/978-1-5225-2255-3.ch250

acteristics that make for what would constitute legal tender in society. The positive and negative impacts of transacting in bitcoin were laid out. A description of the international exchanges that bitcoin engendered was also provided. But more importantly the chapter explored the direction that regulation might take. At that time regulatory agencies and groups were just at the early stages of “thinking about” what needs to be done and were not ready to recommend, let alone enact, rules on virtual currency including the regulation of bitcoin. Thus the chapter concluded with some of the areas of research that would provide fertile fields for further research (Manrique & Manrique, 2015).

Literature Review

As a result of the expansion of, and accompanying controversies associated with virtual currencies in general and bitcoin in particular, many books about virtual currencies have been written and reviewed. A discussion of a few of those books and articles will take place in the next few paragraphs.

Bitcoin as a payment network or digital currency was said to have been developed by a person (or group) that goes by the name of Satoshi Nakamoto. A 2008 work attributed to him laid the groundwork for the first specification and proof of what the bitcoin concept was all about. (Nakamoto, 2008) If one were interested in the technology behind bitcoin one can consult the work by Menezes, et.al, which is a handbook on cryptography describing the process of turning ordinary information into hidden language that needs some code to decode it and make it readable again. Although it does not apply the theory to bitcoin since it was written before the conceptualization of bitcoin itself the reader will gain an understanding of the technology that goes into the way transactions are undertaken using virtual currency (Menezes, 1996)

It is estimated that in 2014 alone about 200 pieces had been written about cryptocurrency and that dozens more will have appeared in 2015.

Many of them are geared towards the basics of understanding the bitcoin phenomenon because they are “befuddled” (Barski, 2014) while others are written for the beginner who would like to dabble in bitcoin transacting. (Franco, 2014) Several column pieces have been published by the Wall Street Journal and the Economist over the past year that give one a good impression of the direction the discussion has been going. And a television special on bitcoin was hosted by Morgan Spurlock as recently as early 2015. The Spurlock television special takes a very elementary look at bitcoin and provides background but not much more to the serious researcher. It serves the purpose of exposing the general public to the existence of virtual currencies, their uses and the possibility of adoption in the foreseeable future. Thus it serves an educational purpose by bringing a new and innovative topic to the attention of the ordinary citizen. (Spurlock, 2015)

Two of the books that have been written about bitcoin are entitled *Digital Gold* written by Nathaniel Popper and *The Age of Cryptocurrency: How Bitcoin and Digital Money are Challenging the Global Economic Order* by Paul Vigna and Michael Casey, two writers whose articles in the Wall Street Journal served as sources for much of the research that went into the original Manrique and Manrique chapter.

David Kushner’s review of *Digital Gold* in the Wall Street Journal takes a look at the nefarious characters that Nathaniel Popper writes about in his “inside story” about the misfits and millionaires who have tried repeatedly and extensively to re-invent money with their support of virtual currencies but more specifically of bitcoin. Kushner wrote that the cast of characters portrayed by Popper may have a tendency to turn off readers to the possibilities of virtual currencies because the portrayals tend to focus on the machinations of the bitcoin underworld. Prominent among the characters accounted for in the book is Ross Ulbricht who was responsible for building Silk Road. Silk Road is considered by many as the largest black market site on the web where transactions

on drugs and drug-trafficking, money laundering, and other criminal enterprises were widely being undertaken. The criminal network activity in Silk Road used two main elements: Tor which promotes “anonymity online” via its free software and open network that prevents the tracking of browsing and website visits; and bitcoin which is the digital equivalent of unmarked banknotes (Bitcoin Buccaneers, 2015). Eventually, the criminal activity that took place in Silk Road was investigated and prosecuted by federal authorities, regardless of where the criminal activity took place. A federal jury convicted Ulbricht in February 2015 and in May sentenced him to life in prison.

Aside from Ulbricht, Kushner focuses on Popper’s choice of the Winklevoss brothers, Cameron and Tyler, as major players in the unfolding cryptocurrency drama. The Winklevoss brothers were involved in a lengthy legal dispute with Mark Zuckerberg over ownership of the idea for Facebook. They have since invested a good portion of their share of the Facebook settlement in bitcoin giving them claim to approximately 1% of all bitcoins in the market. For Kushner the rest of the cast of characters in the bitcoin saga based on Popper’s book remains basically the same and does not bode well for bitcoin’s hoped for role as the global change agent that it was advertised to be. What the developments and the players in the financial markets show is that the evolving “digital cash system is beginning to resemble the old one” (Kushner, 2015).

In *The Age of Cryptocurrency*, Vigna and Casey present a more optimistic view of the cryptocurrency’s role in the financial system. In fact they seem to indicate that beyond the virtual currency itself, venture capitalists supporting bitcoin are interested in how the blockchain technology behind bitcoin can be used to track various forms of financial transactions and asset transfers cheaply and securely. It has great potential to impact the financial world. Blockchain can be likened to a permanent accounting ledger that keeps track of all transactions associated with bitcoin. The potential for applying blockchain technology to different

kinds of transactions serves as a “foundation for a growing number of startups” (Vigna and Casey, 2015). While some may see this development as undermining existing financial institutions, Vigna and Casey see positive developments in the concurrent evolution of bitcoin and other cryptocurrency technology alongside existing financial institutions and businesses. This may portend a revolution or evolution in the internet of things and the internet of values.

Other works such as that of Miers, et. al. looks at alternatives to bitcoin as a result of the security problems encountered with its use. Zerocoin, according to them, serves as such an alternative, and maybe even a better one, because of the anonymity and privacy it provides. (Miers, 2013)

UPDATE

More than a year has passed since the authors researched developments in bitcoin. Since then, while much has transpired, it appears that issues surrounding those which seemed controversial at the time, have since calmed down. One of those major concerns, the volatility of bitcoin’s value, has been reduced by the relatively smaller range within which bitcoin traded during the first 6 months of 2015 compared to its trading range in the first six months of 2014 - 75% v. 167% (BPI, 2015). This section of the chapter provides an update on the status of bitcoin as a currency and its suitability and ability to become a legitimate currency. It focuses on the accompanying blockchain technology which is deemed to be the future in value-exchange protocol just as TCP/IP was for communications protocol.

Bitcoin possesses useful qualities that one looks for in a currency: it is hard to earn; it is limited in supply; and it is easy to verify. Likewise its stability and reliability are very important. It therefore caused tremendous anxiety in the market when Bitstamp (a bitcoin exchange) reported that 19,000 bitcoin units had vanished from the exchange due to a hacking attack. This led to a temporary halt

in trading and immediately brought into question the stability and reliability of the virtual currency (The Magic of Mining, 2015).

However, those optimistic about bitcoins are betting that the technology behind it called blockchain, which makes it difficult to hack, is where the future lies. To hack blockchain would require owning 51% of bitcoin's computing capacity which would be prohibitive in terms of hardware, software and electricity used. One estimate of these costs is at least \$425M (Magic of Mining, p. 58). Furthermore the energy used to support the tremendous computing power bitcoin operations entail is being generated in Inner Mongolia where energy costs are low due to poor compliance with environmental regulations. This relatively inexpensive source of energy for computing power is another advantage bitcoin has over would-be hackers. However, one must recognize that there is always the possibility that a government recognizing its monopoly power can takeover in Inner Mongolia and potentially make things difficult and expensive for bitcoin users everywhere.

Despite its problems bitcoin cryptocurrency technology is still the fastest growing sector in the Information Technology (IT) industry. Now that bitcoin's breakneck speed of growth and development may have ended and seems to have leveled, there appears to be a consensus forming among its proponents, and a resulting code of conduct among developers, that will tend to block radical changes to the current structure of bitcoin technology. It is inevitable that there will be break away developments in bitcoin. (Athey, 2016) Those who attempt such would most likely be connected to bitcoin via bitcoin blockchains – in essence forming sidechains allowing for continued experimentation and expansion from the original structure. Likewise there are those who are working towards private, permissioned blockchains veering away from the original public, permissionless ones developed for bitcoin. Adding that layer of privacy can make more adapters of the technology feel better about its security. (Bheemaiah, 2016)

Thus there seems to be two trends developing: one is the growing acceptability of bitcoin as a currency beyond just the early enthusiasts; and two is the growing usefulness of blockchain technology outside of its use in cryptocurrency to other areas not envisioned before. These are the topics of the next few paragraphs.

Bitcoin as Currency

As a currency one can see the growing acceptance of bitcoin in the market place. It is estimated that more than 100,000 institutions now accept bitcoins as a means of undertaking transactions. Businesses accepting bitcoin include Time and “Microsoft in select marketplaces, in the Windows Store, or in stores that house Xbox Games, Xbox Music or Xbox Video” (Hernandez, 2014). Dell has announced that Bitcoin would be accepted as a payment method for dell.com technology purchases. And UK merchants accepting bitcoin can be found in <http://bitcoinacceptedhere.org.uk/> while products and places where bitcoins are accepted can be found at <https://www.cryptocoinsnews.com/tag/bitcoin-accepted-here/>.

Amagi Metals has been accepting bitcoin since 2012, and cryptocurrencies comprise roughly 30-40% of its business, which is all done online, so it is already further along towards being a complete bitcoin business than most other retailers. And it is in 2017 that the owner of the business indicates that it will turn away business that is transacted in dollars and those not transacted in bitcoin.

St. Petersburg, a city on Florida's west coast, is slowly becoming a bitcoin hotbed, with its growing affection for the cryptocurrency. The St. Petersburg Museum of History has started to accept payment in bitcoin, the first in the state to accept the cryptocurrency. Going hand-in-hand with that is another prominent bitcoin-related event in the area: the 2014 Bitcoin Bowl – formerly known as the Beef O' Brady Bowl which was meant to boost bitcoin's acceptance in the city (Vigna & Casey 2014a).

Other uses of bitcoin have been in the realm of charitable giving. In 2013, immediately after Typhoon Yolanda struck the Philippines, the charitable organization Save the Children began accepting donations via BitPay, a company that processes the donations. Bitpay converts the bitcoin gifts into dollars thus avoiding the uncertainty over price volatility. The use of Bitpay is also advantageous for both donors and recipients especially for those accepting small gifts of a few dollars or less because there are no transaction fees incurred. This is unlike donating using credit cards where transaction fees may be charged. The American Red Cross and Green Peace are also engaged in the growing acceptance of the cryptocurrency. Its usefulness in the non-profit sector is tremendous because transactions can carry lower costs than credit or debit based payments. It thus attracts a whole new pool of givers (Silverman, 2015).

The development of San Francisco based 21 Inc. which has accumulated the largest investment by a startup in the digital currency sector has bolstered confidence in the future of bitcoin. It is named after the 21 million limit of the bitcoin algorithm that determines how much bitcoin is released into the system. Its goal is to come up with software and hardware products designed to drive the mainstream adoption of bitcoin (Casey, 2015a). Widely-recognized and respected names such as Lawrence H. Summers, former Treasury Secretary, joined 21 Inc.'s advisory board. Cisco Systems invested an undisclosed amount in the company for a line of chips for smartphones that will be able to "mine" the digital currency (Heavy Hitter, 2015). Goldman Sachs, JP Morgan Chase and Co., Yuan Capital, Qualcomm Inc., eBay Inc., Dropbox Inc., Expedia Inc., and Zynga Inc. are just some of the companies that are honing in on the new technology that can come out of the heavy investment in cryptocurrency technology.

Blockchain Technology

Followers of developments in this area are especially excited about the potential of blockchain

technology. It can eventually be used to support important new uses in transactional activities such as lawyer-free smart contracts to tamper-proof online voting systems (Casey 2015b, C2). This can lead to the further development of the "Internet of things" and the "internet of value," terms which have been used to describe the "myriad smart, internet-connected apps communicating with servers, networks and each other in order to optimize their operation, maintenance and energy usage without direct human intervention" (Casey, 2015a).

It will bring bitcoin into the realm of a mass-marketed phenomenon. In the financial transactions sector, private companies typically handle the transfer and sales of shares with information systems that use spreadsheets maintained by lawyers who verify transactions by hand. If such were replaced by the process based on bitcoin's blockchain technology it would mean real time settlement instead of waiting a few days for transactions to be verified and cleared. This can reduce the risk of counterparty failure and free up billions of dollars that are sidelined during the waiting period. With bitcoin technology getting the Nasdaq nod by the appointment of a "block chain technology evangelist," Fredrik Voss, Nasdaq takes the lead in efforts to increase the use of such technology in the NY Stock Exchange (Hope and Casey, 2015a). It is possible that other exchanges will follow suit.

Blockchain technology and the increasing acceptance of Bitcoin will likely transform how people buy and sell goods and transfer money across the globe. However, bitcoin's limited use beyond the 100,000 products and places that accept them still needs to be addressed. And it is thought that much of the reluctance in its adoption is because of security concerns, criminal usage, and price volatility issues. It is difficult to trade in bitcoin if there are not enough buyers and sellers at exchanges. It will require more interest from institutional investors including hedge funds and proprietary traders. Liquidity is also widely scattered across exchanges in HongKong, Europe and

the United States. It is said that for firms to get engaged what will be required are more sophisticated infrastructure which 21 Inc. investment hopes to address (above) and regulatory clarity and certainty which has been paved by New York state (Hope and Casey, 2015b).

Bitcoin got another boost in July of 2014 when the New York State Department of Financial Services released a proposal to regulate digital currencies in the state. It got greater strength when in June of 2015 the regulation was finally approved. It is the first of its kind in terms of the regulation of virtual currencies for those operating in the state of New York. Pertinent aspects of the regulation affect those that exchange virtual currencies for dollars. They have to apply for a license; have to engage in reporting; and comply with rules regarding customer/consumer protection, anti-money laundering, capital adequacy, change of ownership and cyber security (Casey, 2015a).

Such regulations have both supporters and protesters. Supporters applaud these regulations as an opportunity to bring credibility, legitimacy and confidence to a six year technology marred by concerns about security and criminality. Protesters decry the regulations as the state's clumsy attempts that will cause emerging technologies in this area to be "doomed even before they get out of the cradle" (Casey, 2015a).

However in the natural evolution of things, the bitcoin phenomenon will lead to all kinds of linked cryptocurrencies and blockchain uses as stores of value, means of exchange, mechanisms for transferring assets and verifying transactions – all functions of a legitimate and accepted currency which can still be found in bitcoin. But more than the currency aspect, bitcoin has unleashed a wave of financial innovations, considered a new branch of the IT industry which is not yet fully understood in terms of the other rising uses of blockchain technology: transacting in other assets such as stocks and bonds and/or notarized e-messaging. Eventually, it may be that bitcoin's value is not so much as a currency but in the block chain technology it has spawned.

It is therefore likely that bitcoin acceptability as a currency will continue to grow but at the same time newer, more innovative uses for the blockchain technology that goes along with it will be discovered and developed thus taking advantage of this growing field.

FUTURE RESEARCH DIRECTIONS

It is very interesting to monitor the evolution of a concept and a reality in a cryptocurrency like bitcoin. It is certainly something that warrants continued monitoring since it seems to be gaining acceptance now. It may not completely replace the dollar but its technological value especially in business and finance in the age of mobile technology will certainly keep the attention of those in the forefront of developments in the technological field. And the authors will continue to monitor bitcoin in its evolution.

But more importantly the movement in the field seems to be geared towards the promise that blockchain technology brings to the economic and financial realm. When layers of redundancy can be removed it augurs well for the speed of transactions in a fast-paced world. It is a development that needs to be monitored as well.

CONCLUSION

This update certainly enhances our knowledge and understanding of where bitcoin and blockchain technology are at this point in their evolution in the technology field.

Bitcoin no longer hogs the limelight today. Gone are the days when scandal and controversy seemed to plague it. The volatility has seemed to have settled and the focus of attention has been on the stability of this cryptocurrency as a medium of exchange and store of value little touched by the ravages of the market. Its underlying block chain technology is moving towards further uses that can help it find a niche in the world of technol-

ogy innovation. From a management perspective in whatever field - accounting, finance, law - the direction towards which things are moving not only in the bitcoin realm but in the blockchain field spells efficiency with the removal of redundant layers and the verification of transactions being undertaken in a split second. Such speed and efficiency are after all what the world of technology brings to people's daily lives. Such is the future promise of this type of technological development.

REFERENCES

- Athey, S. (2016). What Is The Current State Of Bitcoin And Blockchain Technology? *Forbes*. Retrieved from <http://www.forbes.com/sites/quora/2016/01/28/what-is-the-current-state-of-bitcoin-and-blockchain-technology/#44945b301b1c>
- Barski, C., & Wilmer, C. (2014). *Bitcoin for the Befuddled*. No Starch Press. Retrieved from <https://www.nostarch.com/bitcoin>
- Bheemaiah, K. (n.d.). *Blockchain 2.0: The Renaissance of Money*. Retrieved March 28, 2016 from <http://www.wired.com/insights/2015/01/block-chain-2-0/>
- Buccaneers, B. (2015, January 17). *The Economist*, p. 33.
- Bitcoin Price Index (BPI). (n.d.). Retrieved June 27, 2015, from <http://www.coindesk.com/price/>
- Casey, M. J. (2015a, June 4). New York Gets Bitcoin License Rules. *The Wall Street Journal*, p. C3.
- Casey, M. J. (2015b, March 11). Noted Names Put Cash in Bitcoin Startup 21 Inc. *The Wall Street Journal*, pp. C1-C2.
- Franco, P. (2014). *Understanding Bitcoin: Cryptography, Engineering, and Economics*. Wiley. doi:10.1002/9781119019138
- Heavy Hitter Joins 21 Inc. (2015, May 19). *The Wall Street Journal*, p. C3.
- Hernandez, P. (2014). *Microsoft: Bitcoin Accepted Here*. Retrieved June 24, 2015 from <http://www.eweek.com/cloud/microsoft-bitcoin-accepted-here.html>
- Hope, B., & Casey, M. J. (2015a, May 11). Bitcoin Technology Gets Nasdaq Test? *The Wall Street Journal*, p. C2.
- Hope, B., & Casey, M. J. (2015b, April 6). Large Trading Firms See Opportunity in Bitcoin. *The Wall Street Journal*, pp. C1, C6.
- Kushner, D. (2015, June 15). Cash for the New Century. *The Wall Street Journal*, p. A15.
- Nakamoto, S. (2008). *Bitcoin: A peer-to-peer electronic cash system*. Retrieved March 27, 2016 from <http://www.cryptovest.co.uk/resources/Bitcoin%20paper%20Original.pdf>
- The Magic of Mining. (2015, January 20). *The Economist*, pp. 58, 60.
- Manrique, C. G., & Manrique, G. G. (2015). The Evolution of Virtual Currencies: Analyzing the Case of Bitcoin. In C. G. Reddick & L. Anthopoulos (Eds.), *Information and Communication Technologies in Public Administration* (pp. 213-234). Boca Raton, FL: CRC Press, Taylor & Francis Group. doi:10.1201/b18321-13
- Menezes, A. J., Van Oorschot, P. C., & Vanstone, S. A. (1996). *Handbook of applied cryptography*. CRC Press. doi:10.1201/9781439821916
- Miers, I., Garman, C., Green, M., & Rubin, A. D. (2013, May). Zerocoin: Anonymous distributed e-cash from bitcoin. In *Security and Privacy (SP), 2013 IEEE Symposium on* (pp. 397-411). IEEE. Retrieved March 27, 2016 from <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6547123>
- Popper, N. (2015). *Digital Gold: Bitcoin and the Inside story of the Misfits and Millionaires Trying to Reinvent Money*. New York: Harper.
- Silverman, R. E. (2015, April 13). Brother, Can You Spare a Bitcoin? *The Wall Street Journal*, p. R9.

Spurlock, M. (2015, February 19). *Morgan Makes “Cents” Out of Bitcoin*. CNN Special.

Vigna, P., & Casey, M. (2015). *The Age of Cryptocurrency: How Bitcoin and Digital Money are Challenging the Global Economic Order*. New York: St. Martin's Press.

Vigna, P., & Casey, M. J. (2014). BitBeat: Bitcoin Accepted Here; Dollars, Not So Much. *WSJ online*. Retrieved June 24, 2015, from: <http://blogs.wsj.com/moneybeat/2014/08/22/bitbeat-bitcoin-accepted-here-dollars-not-so-much/>

ADDITIONAL READING

Bitten. (2014). *The Economist*. p. 62. Bollen, Rhys. The Legal Status of Online Currencies: Are Bitcoins the Future? Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2285247

Bray, D. A., & Konsynski, B. R. (2006, November 1) “Virtual Worlds, Virtual Economies, Virtual Institutions” Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=962501

Casey, M. J. (2013a, January 17). Coin of Kings: Bitcoin Scores. *The Wall Street Journal*. p. C1. CIO Report. *CIO Explainer: What is Blockchain?* Retrieved March 28, 2016 from <http://blogs.wsj.com/cio/2016/02/02/cio-explainer-what-is-blockchain/>

Fowler, G. A. (2014, February 19). Want to Buy Everyday Stuff with Bitcoin: A User's Guide. *The Wall Street Journal*. pp. D1, D3.

Galston, E. (2014, February 28). The Bitcoin Paradox that Undid Mt. Gox, *The Wall Street Journal*. p. A13.

Queenan, J. (2014, January 5). I'm Up on Bitcoins, Even if They're Down. *The Wall Street Journal*. p. C11.

Wallace, Benjamin. (2011, November 23). The Rise and Fall of Bitcoin. *Wired Magazine*.

KEY TERMS AND DEFINITIONS

21 Inc.: At present, the largest start-up company invested primarily in bitcoin technology.

Bitpay: A means of payment via bitcoin, used by non-profit organizations.

Blockchain: The technology that allows for permanent tracking of all bitcoin transactions.

Cryptocurrency: Currency that is created through digital processes and algorithms.

Internet of Things: The myriad of sensors, digital equipment, and networks that connect equipment, machinery, tools, etc. together for the ease of human transaction and interaction.

Ross Ulbricht: Businessman convicted for conducting criminal activities on Silk Road using bitcoin.

Silk Road: Website that was infamous for anonymous criminal activity using bitcoin.

Use of Bitcoin for Internet Trade



Sadia Khalil

NUST School of Electrical Engineering and Computer Science, Pakistan

Rahat Masood

NUST School of Electrical Engineering and Computer Science, Pakistan

Muhammad Awaiz Shibli

VisionIT, USA

INTRODUCTION

The advent of the digital currency systems has revolutionized the concept of money transfer by allowing the internet based creation, storage and transference of money. In the past few years, the digital currency systems have emerged as an efficient means of money transfer. They have received worldwide adoption by providing a medium of exchange based on mathematical operations and by taking the currencies out of the control and manipulation of the governments. In addition to being used in the e-commerce and commercial sectors, the digital currencies have also attracted a large population of the earth which cannot get access to the formal banking systems. The crypto-currencies, being one of their types, involve different cryptographic functions for their creation and transference, in a trusted and secure environment. The use of crypto-currencies has progressed from a virtual concept to reality by the evolution of Bitcoin. The success of this concept has led to the creation of many other crypto-currencies which include Litecoin, PeerCoin, Namecoin, Quarkcoin, Primecoin and Zetacoin (Stevenson, 2013). Bitcoin, along with the other crypto-currency systems, is very popular in the business world and the global economy, due to its decentralized peer-to-peer architecture. In comparison with the other payment platforms, which maintain a private communication network for sending and receiving money, Bitcoin uses the internet as its medium of transference.

By Looking critically into the Bitcoin protocol, we can find some weaknesses that can be violated by the attackers for malicious purposes. In the past few years, a lot of vulnerabilities have been exploited causing the users to lose their bitcoins (L., n.d.), (Blasco, 2013), (arXiv, 2014). Matthew Wilson et al. (Yelowitz, 2014) analyze the characteristics of the Bitcoin users based on the Google search data and found that illegal activities and programming enthusiast are related to Bitcoin search but no correlation was found with political and investment motives. As of March 2014, bitcoins of worth 502,081,166.11\$ have been stolen (L., n.d.). Based on the empirical analysis of Bitcoin exchange risks, it is found that the failure rate of bitcoin exchanges is 40% (Christin, 2013). Mt. Gox that was considered to be the largest Bitcoin exchange, got bankrupt in February 2014, allegedly due to theft, resulting in the loss of 850,000 bitcoins, out of which 20,000 were later recovered (https://en.bitcoin.it/wiki/Mt._Gox, n.d.).

BACKGROUND

The Bitcoin¹ protocol was first introduced in 2009 by a pseudonymous developer Satoshi Nakamoto (Nakamoto, 2008). Since then, it has been widely adopted as a payment procedure for many e-commerce businesses as well as regular stores. This crypto-currency along with the others,

DOI: 10.4018/978-1-5225-2255-3.ch251

is considered to be a convenient way of achieving the open source peer-to-peer money. It operates in the cyberspace and requires Bitcoin wallets for storage purposes as well as for the generation of Bitcoin addresses. At the time of this writing, the Bitcoin market capitalization is \$5.8 billion (Crypto-Currency Market Capitalizations, n.d.). Keeping in view its frequent usage, Bitcoin ATMs have been deployed in various parts of the world to facilitate its users (Bitcoin ATM News, n.d.). In comparison with the Visa transactions, where the transaction speed is 2000 tps (transactions per second) and PayPal which has 115tps transaction speed, the Bitcoin network is restricted to 7 tps (Scalability, n.d.). In spite of these statistics, the advantages of Bitcoin transaction over other transaction mechanisms like PayPal, Western Union and M-Paisa etc. cannot be neglected. It gives the users the advantage of carrying out instant, anonymous and irrevocable transactions with very low transaction fees. The original Bitcoin paper (Nakamoto, 2008) presents a brief overview of the architecture and the protocol but a lot of details are missing in it. With the passage of time, a number of suitable changes and ideas have been suggested through the Bitcoin Improvement Proposals (BIPs) which are incorporated after being approved by the Bitcoin community.

In the last few years, researchers all over the world are working on Bitcoin security and there is still a need for a comprehensive assessment of attacks that are targeting the Bitcoin transactions. In this chapter, we investigate the Bitcoin protocol in detail. We have analyzed the Bitcoin architecture and its major components. We then review the Bitcoin protocol considering a use case scenario to demonstrate how a Bitcoin transaction takes place. The vulnerabilities and attacks section heuristically show how attacks like double spending, selfish mining, compromising anonymity and malware attacks can be carried out in the currently deployed versions of Bitcoin protocol. A comparison between different crypto-currencies with respect to their features and possible attacks is also presented.

BITCOIN ARCHITECTURE

In comparison with the traditional currencies that depend on a trust based model for their creation, circulation and transference, Bitcoin relies on a Proof-of-work (discussed in detail later in this chapter) based peer-to-peer model.

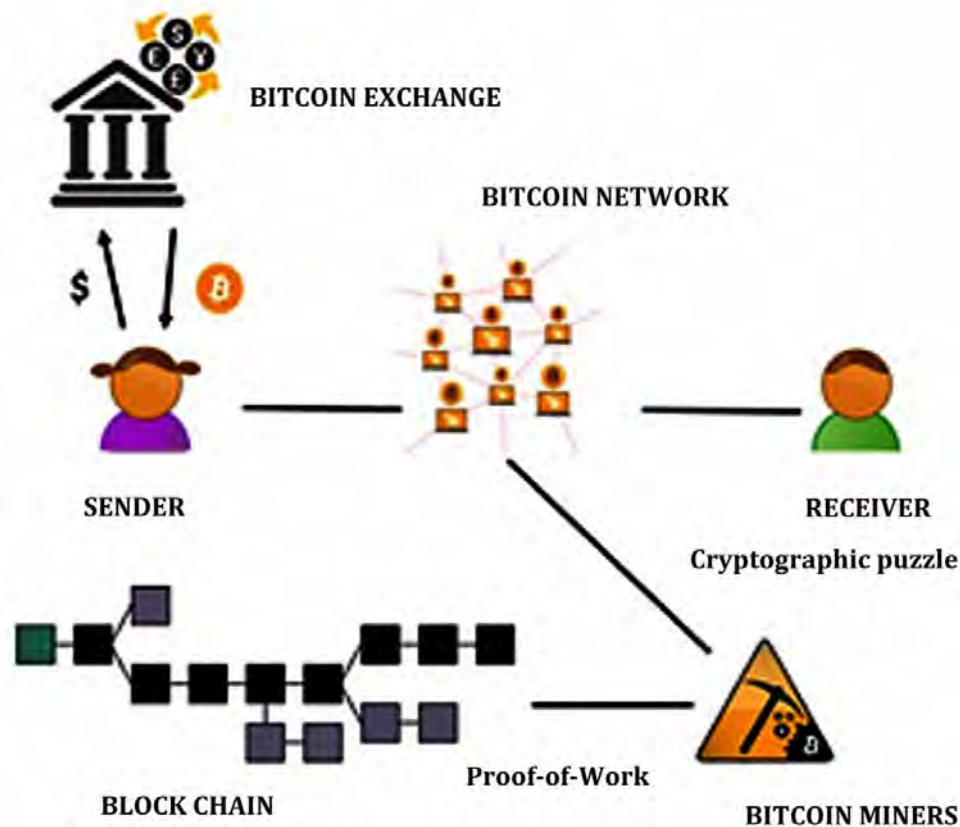
Being a crypto-currency, the Bitcoin protocol makes use of the hash functions and public key cryptography for the generation and transmission of bitcoins. A single bitcoin can be regarded as a series of digital signatures. For sending bitcoins to another entity, the sender digitally signs a hash of the involved previous transactions out of which the bitcoins are sent to the receiver as proof of possession of those bitcoins. The receiver can easily verify the series of digital signatures. Figure 1 shows the major components that constitute the Bitcoin Architecture.

The Bitcoin exchanges are not an inherent part of the Bitcoin protocol, however, they are one of the means for obtaining the bitcoins. The Bitcoin users can trade the bitcoins in exchange of the traditional or digital currency. Another way is to get bitcoins personally by asking the possessor of the bitcoins to transfer them to the buyer's Bitcoin wallet.

The Bitcoin users require Bitcoin wallets to carry out the transactions. Just like real life wallets, which are used to keep cash, Bitcoin wallets are also responsible of keeping record of the bitcoins sent and received by the owner. Bitcoin wallets generate a public/private key pair for carrying out a transaction.

The Bitcoin miners are the entities in the Bitcoin network that possess computational resources to compete in the mining process. Mining is the mechanism through which the transactions are recorded in the block chain. Mining involves a complex cryptographic mechanism known as Proof-of-Work (PoW). The motivation for the miners to perform mining is the Bitcoin reward that is granted to each miner. This reward is the source of creation of new bitcoins in the system. Initially, its value was 50 BTCs and is set in a way

Figure 1. Bitcoin architecture



that it is halved after every 4 years (Controlled Supply, n.d.).

Individual mining is tedious and the miner has to be efficient enough to be the first one to solve the PoW puzzle successfully. If an entity is not interested in mining individually, it can join a mining pool. Joining a mining pool increases the probability of a miner in solving the PoW as the miners are assigned smaller cryptographic puzzles to solve and are able to combine their computational power with that of others in the mining pool.

In addition to the reward, the miners also receive a very small amount of bitcoins as the transaction fees. If a transaction is processed in a way that it draws bitcoins from many addresses, then greater transaction fees is associated with it due to the large transaction size. When a miner successfully generates a block, the information

related to all of the transactions is incorporated in it and that miner can obtain the transaction fees.

The Block chain serves as a global ledger for keeping a record of all the confirmed transactions that have taken place since the first ever Bitcoin transaction.

BITCOIN PROTOCOL

The Bitcoin wallets generate a public/private key pair for carrying out a transaction. The hash of the public key is referred to as Bitcoin address which can be used to send or receive transactions. The Private Key is used to digitally sign the transaction in order to add sender's identity. All the transactions are signed and verified using the elliptic curve digital signature algorithm ECDSA. In such signature algorithms, public

key is derived by the multiplication of the base point of the elliptic curve with the private key. The elliptic curve used in Bitcoin is secp256k1 which is recommended in SEC 2 (Standards for efficient cryptography) (Research, 2000). Besides having the knowledge of the public key and the base point, it is not possible for anyone to derive the private key. This factor adds to the security of ECDSA. The ECDSA signatures also require a random parameter for each separate transaction made with the same private key.

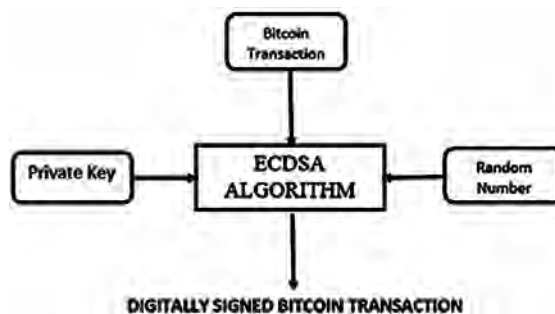
Figure 2 explains how the Bitcoin transactions are digitally signed. A Bitcoin transaction comprises of two parts: input and output. Each part embeds certain scripts which control the future spending of the bitcoins that are being sent. The input of a transaction connects to a previous output. The Input consists of hashes of the previous transactions that client had with other users in the network. Amount of the current transaction is assigned out of these transactions. The purpose of taking hashes is to reference the output of the transactions from which this transaction is funded. It also comprises of a script called *scriptSig* which refers to the number of arguments that are expected by a script. The output consists of the amount to be sent to the receiver signed by the receiver's public key and the change to be sent back to the sender. A short script called *scriptPubKey* is also a part of the output which specifies the conditions under which the output could be redeemed. The script contains the hash of the ECDSA public key and a signature validation routine. For a valid transaction, the output script should evaluate to be true given the *scriptSig* provided in the input.

Each transaction is identified by a unique transaction ID which is the SHA-256 hash of the entire transaction.

Proof of Work

Bitcoin transactions are verified through the Proof-of-work which is used as a countermeasure for double spending attacks (discussed in the next section). The receiver of the transaction

Figure 2. Digital signatures in Bitcoin transactions



requires confirmation in the form of a proof from the majority of the nodes, that the amount is not double spent. This proof is provided using the PoW system which is based on the idea of Adam Back's Hashcash (Back, 2002). Hashcash ensures that the requester has spent considerable amount of CPU resources before receiving the services.

Bitcoin mining involves finding the PoW, in which a nonce is calculated by the miners using a random number generator (RNG) algorithm ran at a colossal rate. The nonce is calculated corresponding to an automatically generated challenge/difficulty. The block including the nonce when hashed under SHA 256, generate a string whose initial bits are consecutive zeroes (as specified in the challenge). The hash of the block (including the nonce) should satisfy the challenge/difficulty. The computational resources required for doing the PoW are directly proportional to the number of required zeroes. The first miner whose RNG spits out the correct nonce is the miner who gets the reward for solving the PoW. Any node of the Bitcoin network can become a miner if it possesses reasonable computing resources.

Once the PoW is solved, the block is broadcast to the network. The block will be accepted by the nodes only if the transactions are valid and are not double spent. Broadcasting the block also informs the other Bitcoin miners to quit mining this block and start working on another one. The verified block is then appended to the block chain. Each block is a reference to the previous one and results in formation of a continuous chain which serves as public ledger.

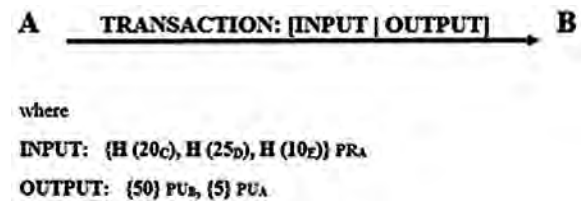
The average time for mining is about 10 minutes. Therefore, the memory pool of each node will contain all the unlogged transactions that have taken during the time period of these 10 minutes. For every 2016 blocks that have been mined, the network checks the performance of the miners. If it is greater than normal, the difficulty level of the PoW is increased. The difficulty in solving this cryptographic puzzle provides security as it ensures that no malicious entity can flood the block chain with conflicting blocks.

Suppose a merchant running an online business accepts bitcoins payments. In order to receive the payments, Bitcoin address of the merchant should be known to all of the customers. Practically, it is quite inefficient to have a separate transaction for each bitcoin transfer, considering the small size of such transactions. In this regard, Bitcoin protocol allows splitting and combining of transactions. Generally, there can be one input if previous transaction is large enough for the amount to be sent. If the previous transactions are small and amount to be sent is large, then there will be multiple inputs combining previous smaller transactions. Figure 3 explains a scenario of a Bitcoin transaction in which an entity **A** wants to send 50 bitcoins to entity **B**. **A** has received 20 BTCs from **C**, 25 BTCs from **D** and 10 BTCs from **E**. As discussed earlier, input part of the transaction contains hashes of the previous transactions which is digitally signed with private key of **A** ($P R_A$). Output contains the amount which is to be sent to **B** and the change amount that is to be sent back to **A**. In order to ensure that the desired amount is only received by **B**, the amount is encrypted using public key of **B** ($P U_B$). The change amount is encrypted using public key of **A** ($P U_A$).

For the validation of a transaction following steps take place:

1. Each node gets the transaction that was broadcast by the sender.
2. Each node checks the formatting of the transaction and confirms if the transaction is already present in the block chain. The

Figure 3. An entity **A** wants to send 50 BTCs to another entity **B**. Previously, **A** has received 20 BTCs from **C**, 25 BTCs from **D** and 10 BTCs from **E**.



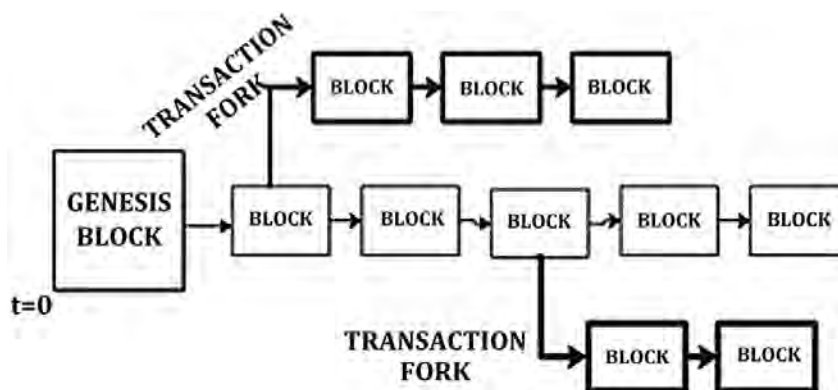
unlogged transactions are temporarily stored in memory pools (memory buffers) of each node.

3. The miners compete in the formation of a block by participating in the mining process. PoW puzzle is calculated which involves a distributed time stamping mechanism to prevent double spending. This work requires a lot of computational power and hence miners are rewarded with a small percentage of transaction fee. If the nonce that solves the PoW is successfully found, then all the transactions in the memory pool are transferred to the block, which is then broadcast. The purpose of the broadcast is to tell other nodes to start working on a new block.
4. The nodes validate the block by confirming the PoW and it is added to the block chain. The miners work on the construction of new blocks.

Conflict Resolution and Consensus

If two miners succeed in finding the PoW simultaneously, then after broadcasting, some of the nodes will receive one of the blocks whereas others will receive the other one. This results in the formation of a fork as shown in fig 4. The nodes with the first block will work on it and keep a copy of the second one. Similarly, the nodes that receive the second block first will mine further blocks on this block and keep a copy of the first block. The conflict will be solved after a new PoW is found and one branch becomes longer than the other one. In such case, the blocks in the shorter branch will

Figure 4. Forks in block chain



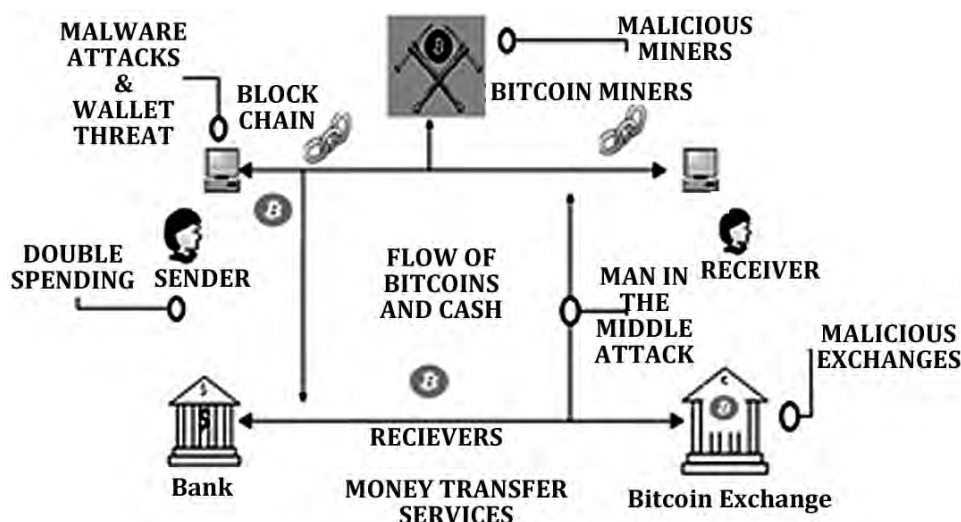
be discarded and the miners will shift to working on the longer one. The discarded blocks are called orphan blocks.

VULNERABILITIES AND POSSIBLE ATTACKS

Despite being an efficient and convenient way of electronic funds transfer, the Bitcoin transactions are becoming a target of many known and zero day attacks. No holistic solution has been proposed up till now to cater these attacks. This leads to the need of certain modifications in the protocol by

the consent of the Bitcoin community. Figure 5 illustrates the vulnerabilities particular to the main entities in the Bitcoin architecture. The flow starts from a sender who has installed a Bitcoin wallet on his system in order to send some bitcoins to the receiver. The sender's bitcoin wallet is prone to malware attacks and theft which can result in the loss of private keys generated by the wallet. The sender can trick the receiver by double spending the amount of bitcoins he is going to send to the receiver. Hence, the sender can deprive the receiver from the sent bitcoins and can reuse those bitcoins. Malicious mining is a major threat in Bitcoin transactions. If malicious miners have a

Figure 5. Vulnerabilities in Bitcoin architecture



large hash rate, then they can easily add conflicting transactions in block chain and can steal bitcoins. Bitcoin transactions which were initially claimed to be completely anonymous, can reveal the identity of the communicating ends. In order to buy bitcoins from exchanges, bitcoin users have to provide some personal information. This personal information can be revealed through man in the middle attack and malicious exchanges. Figure 5 shows some of the attacks that are possible in Bitcoin transactions.

Double Spending Attacks

While a transaction is taking place, there is a possibility that the sender is sending the same bitcoins to two receivers at a time. This leads to a problem known as double spending. Double spending attacks occur when a user assigns same amount of bitcoins (BTCs) to more than one user at a time, allowing users to spend the same amount of bitcoins twice. One of the possibilities of double spending is that the attacker tricks the vendor by sending him a transaction which actually cannot be redeemed. In order to avoid detection, attacker has to replace the transaction as well as the block in the block chain which contains that transaction. This requires a lot of computational power due to the re-computation of that block as well as all subsequent blocks in the block chain.

The Bitcoin system uses hash based PoW schemes to protect against the double spending attacks, but the attackers have found ways to bypass such schemes.

The chances of a double spending attack being successful depend on the attacker's hash rate (<https://en.bitcoin.it/wiki/Double-spending>, n.d.). If an attacker can control more than half of the network's hash rate, then this attack is called 51% attack. In this attack, the attacker makes a transaction to the vendor and meanwhile keeps on mining conflicting block in which a double spending transaction is included. In the presence of some conflicting transactions, the miners have to vote for the valid block by appending further

blocks to the block the miners think is valid. The longest block chain is accepted whereas the other one is discarded. The attacker can get the false blocks verified by the network, by generating blocks in order to keep the chain longer than that of the legitimate peers (Drainville, 2012). After receiving the BTCs, the vendor waits for some confirmations and then provides the service. Once the service is provided, the attacker releases the fork by overtaking the number of blocks on the legitimate chain. Figure 4 highlights two conflicting blocks resulting in the formation of a transaction fork.

Selfish Mining

Mostly, the Bitcoin miners work in the form of groups called mining pools and share the revenue according to their computational power. This honest mining process is the backbone of Bitcoin protocol. A lot of research has been done to analyze the behavior of bitcoin miners. According to the researchers, Ittay Eyal et al. (Siler, 2013), Bitcoin mining is no longer incentive compatible. A large number of selfish miners can join a mining pool and get their mining reward larger than their fair share.

In this attack, if the selfish miners find a block 'X', they do not reveal it to the network, keeping it secret. All the other miners keep on finding that block, as without discovering it, they cannot move to the next block 'Y'. In the meantime, the selfish miners start working on next blocks ('Y', 'Z' etc.). When the honest miners find block 'X' and distribute it to the network, the selfish miners immediately release all the work they have done so far, which might be several blocks (i.e. 'X->Y->Z...'). As the revenue goes to the miner of longer chain, selfish miners win over the honest miners and get the incentive (Arvind Narayan, 2013).

The revenue that the selfish miners get increases super-linearly with the size of the group. So, the colluding miners usually try to add more miners into their group to increase their cumulative computational power. This snowball scenario

helps them to avoid the situation in which the honest miners find the block before the selfish miners would find the second block, due to enhanced computational resources.

Anonymity

Initially, when the Bitcoin protocol was designed, it was claimed that the protocol provides complete anonymity of both the sender and the receiver and they cannot be traced back. But recently, the researchers have found ways to identify the sender and the receiver by using some extra computational power. The Bitcoin community itself states that the current implementation is not very anonymous (Anonymity, n.d.). The Bitcoin protocol is now considered to be pseudo-anonymous.

Each Bitcoin transaction has a list of inputs and outputs. The input contains previous transactions so that the miners would verify that the bitcoins are not already spent (to avoid double spending). The transactions usually have two outputs; one output belongs to the receiver while the other output contains address of sender, to return the extra bitcoins. Using the reference to previous transactions from the input, transaction graphs can be built. These graphs can help in tracking the sender (Moser, 2013).

People usually have to provide their personal information such as, copy of National ID card etc., to bitcoin exchanges, in order to buy bitcoins. Also, the transactions are stored publically in the block chain, so anyone can analyze these transactions to find the links between the previous owners of bitcoins.

Malware Attacks

The cyber-criminals use malicious software to attack systems that have installed the Bitcoin wallets. In most cases, the purpose is to steal critical information. They try to expand the attack by targeting more and more machines, creating a botnet.

The Bitcoin wallets are the digital place to store the private keys used to access bitcoins. If the keys are compromised, all of the bitcoins are lost. To attack the bitcoin wallets, the attacker

writes some malicious code and spreads it to the botnet. The malware steals data from the wallet.dat file which contains sensitive information related to the keys and user preferences. Therefore, it is a common security practice to encrypt this file. However, it has been identified that some of the recent malware detected have key logging capability which lets them steal the password to decrypt the required file. Khelios and IRC are some known malwares that do this job for the attackers (Blasco, 2013). MtGox, allegedly got bankrupt due to this attack. It lost 850,000 bitcoins, out of which, 750,000 bitcoins belonged to customers (arXiv, 2014).

Apart from attacking the Bitcoin wallets, some malware are designed to attack the computational resources of different machines. The incentive driven mining process motivates the miners to increase their computational power in order to get more and more reward. The attackers have now found a way to win the competition. They write and install Bitcoin daemons on victims' machine and connect it to the mining pool. The attacker then uses victims' computational power to mine bitcoins. As, the computational power is enhanced, the chances of solving the PoW increase, resulting in increase in miners' incentive. Such malware is spread through fake emails, Skype, and phishing websites. Some of the known malware are Zeus, Dorkbot and Ufasoft (Blasco, 2013).

COMPARISON WITH OTHER CRYPTO-CURRENCIES

Bitcoin became the pioneer crypto-currency by introducing the concept of decentralized peer-to-peer cryptography based digital currency. The idea of Bitcoin has paved the way for the creation of many other crypto-currencies which have been collectively termed as the 'altcoins'. The altcoins are considered to be modified and improvised versions of Bitcoin but still lag behind in terms of acceptance, market capitalization and liquidity. Table 1 highlights distinguishing features of Bitcoin along with top three altcoins with respect to market capitalization.

Table 1. Table showing salient features of Bitcoin and other crypto-currencies

Crypto-Currency	Features				Advantages	Disadvantages
	Hashing Algorithm	Mining Process	Current Mining Reward	Maximum Supply		
Bitcoin (https://en.bitcoin.it/wiki/Main_Page , n.d.)	SHA 256	Proof-of-work	25 coins (halves after every 210,000 coins)	21million	• Greatest market capitalization among crypto-currencies. • Widely accepted by businesses.	• Requires extensive computational resources (ASICs) for mining (in comparison with Litecoin, CPU mining is not possible in Bitcoin) • Prone to various attacks. (51% attack, selfish mining, compromising anonymity).
Litecoin (Litecoin-Open Source P2P Digital Currency, n.d.)	SCRYPT*	Proof-of-work	50 coins (halves after every 840,000 coins)	84million	• Greatest market capitalization among altcoins. • Faster transaction confirmation. • Higher transaction volume in the blockchain. • Market entry costs are very low which allow anyone with a computer and internet to mine litecoins.	• Lesser market acceptance. • Exposed to similar attacks as of bitcoin (51% attack, selfish mining, and compromising anonymity).
Peercoin (http://peercoin.net)	SHA 256	Proof-of-work + Proof-of-Stake **	25 coins (halves after 16 times increase in the network) but as the reward decreases, reward will be proportional to the miner's stake in the currency.	No limit (designed to reach an annual 1% inflation rate)	• No limit on maximum supply. • Lesser chances of 51% attack/monopoly or due to proof-of-stake system (as new coins are generated based on the holding of the individual.) • Energy efficiency (lesser power consumption required for proof-of-stake as compared to proof-of-work that involves resource intensive hashing functions).	• Lesser market acceptance. • Rich get richer. • People get rewarded for hoarding peercoins.
Namecoin (http://www.econinfosec.org)	SHA 256	Proof-of-work	25 coins (halves after every 210,000 coins)	21million	• Same implementation as Bitcoin. • Can be used for money transfer as well as for storing information (DNS or identification/authorization) in the blockchain. • Provides decentralized DNS to prevent internet censorship.	• Lesser market acceptance. • Users have to pay for network fees along with transaction fees. • Prone to attacks like 51% attack and compromising anonymity.

* SCRYPT is slightly simpler algorithm as compared to SHA 256 which is less susceptible to ASICs (Application Specific Integration Circuits) designed for Bitcoin mining. (<http://citeseerx.ist.psu.edu>). It is designed to make mining accessible to everyone without the requirement of computational resources. (Script.CC, n.d.)

** A hybrid of proof-of-work and proof-of-stake is used by Peerecoin in the mining process. According to the design, proof-of-work is used only in the initial generation of coins, but in the long term proof-of-stake would be used. Proof-of-stake reduces the chances of 51% attack by allowing coin generation based on miner's holding in the network. (<http://peercoin.net>)

Based on the features of each of the crypto-currencies, table 2 illustrates a list of crypto-currencies along with the possibilities of different attacks associated with them. The tick mark ✓ indicates the existence of a particular attack, whereas cross mark × indicates its absence. As discussed in section IV, Bitcoin is prone to attacks like double spending, selfish mining, compromising anonymity and malware attacks etc. Altcoins are designed either to provide additional functionalities on the top of the Bitcoin protocol or to fix the issues found in the protocol. Litecoin transactions have faster block generation rate and provide greater resilience to double spending attacks. As a security measure against double spending attacks, it is a standard practice for a merchant to wait for a sufficient number of blocks/confirmations to be added to its transaction's block. Litecoins transactions can have more confirmations during the same time that is used with Bitcoin, therefore the probability of double spending attacks is lesser. Transactions in Litecoin, Peercoin and Namecoin are all pseudo-anonymous like Bitcoin. Malware attacks can be observed in all crypto-currencies and can be avoided by adopting efficient end user security practices. Peercoin prevents 51% attacks through proof-of-stake scheme which allows miners to generate coins based on their share in the network. A successful double spending attack would require an attacker to possess a very large number of coins, which is practically infeasible. Peercoin prevents selfish mining by providing a fair distribution of coins scheme through proof-of-stake.

FUTURE DIRECTIONS

Digital currencies like Bitcoin are of growing interest to economists, cyber security experts, speculators and media. Never in history have people experienced a decentralized money transfer system that offered anonymous and unregulated transactions to be carried out with very less transaction fees. World's unbanked population finds no choice other than Bitcoin for sending money to other people in the world. On a larger scale, the worldwide national adoption of decentralized and virtual currencies will result in economic neutrality and political transparency.

It cannot be said with surety that Bitcoin or any other crypto-currency will be able to replace the real currency but there is a great desire for the digital currencies. Human beings are considered to be the weakest link in the security chain, therefore there is a dire need of a decentralized and trustless system as it is better to trust on mathematical operations than humans. A large number of credible researchers and entrepreneurs are investing their time and money in this domain. The advantages that digital currency systems provide over physical monetary systems cannot be overlooked. However, certain economic and security challenges act as a hurdle in world-wide adoption of virtual currencies. In order to ensure successful and world-wide adoption in future, security weaknesses in the protocol need to be dealt with. Moreover, price fluctuations are causing inconvenience for businesses accepting bitcoins as they have to adjust their prices accordingly.

Table 2. Crypto-currencies and their possible attacks

Crypto-Currency	Attacks			
	Double Spending	Selfish Mining	Compromising Anonymity	Malware Attacks
Bitcoin	✓	✓	✓	✓
Litecoin	×	✓	✓	✓
Peercoin	×	×	✓	✓
Namecoin	✓	✓	✓	✓

In short, Bitcoin is evolving with progress in economic and technological conditions and its future is quite vague. If the Bitcoin community puts its efforts in resolving the inherent security issues in its protocol, then Bitcoin surely has the capability of being a strong competitor of real currency.

CONCLUSION

In this chapter, we have studied the Bitcoin protocol from a critical perspective and have highlighted various vulnerabilities that need to be addressed. We have presented a holistic survey of attacks that can take place during the Bitcoin transactions. We also discuss a comparative analysis of Bitcoin and other crypto-currencies by highlighting their specific features, advantages and disadvantages and possible attacks. Based on our analysis, it is suggested that changes need to be made to the Bitcoin protocol by the consensus of the open source community in order to overcome the weaknesses.

REFERENCES

- Ahamad, Nair, & Varghese. (2013). *A Survey on Crypto Currencies*. Retrieved November 11, 2015, from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.428.7952&rep=rep1&type=pdf>
- Anonymity. (n.d.). Retrieved October 4, 2015, from <https://en.bitcoin.it/wiki/Anonymity>
- Back, A. (2002). *Hash Cash-A Denial of Service Countermeasure*. Retrieved June 24, 2015, from <http://www.hashcash.org/papers/hashcash.pdf>
- Bitcoin ATM News. (n.d.) Retrieved October 4, 2015, from <http://www.coindesk.com/technology/bitcoin-atm/>
- Certicom Research. (2000). *SEC2: Recommended Elliptic Curve Domain Parameters*. Retrieved January 28, 2016, from <http://www.secg.org/SEC2-Ver-1.0.pdf>
- Christin, T. M. (2013). *Beware the middleman: Empirical analysis of bitcoin-exchange risk*. In *Financial Cryptography and Data Security* (pp. 25–33). Springer.
- Controlled Supply. (n.d.). Retrieved 4 June 2015, from https://en.bitcoin.it/wiki/Controlled_supply
- Crypto-Currencies Market Capitalizations. (n.d.) Retrieved January 28, 2016, from <http://coinmarketcap.com/>
- Drainville, D. (2012). *An analysis of the bitcoin electronic cash system*. Waterloo, Canada: University of Waterloo.
- Emerging Technology from the arXiv. (2014). *The Troubling Holes in MtGox's Account of How It Lost \$600 Million in Bitcoins*. Retrieved October 4, 2015, from <http://www.technologyreview.com/view/526161/the-troublingholes-in-mtgoxs-account-of-how-it-lost-600-million-in-bitcoins/>
- Eyal, I., & Sirer, E. G. (2014). Majority is not enough: Bitcoin mining is vulnerable. In *Financial Cryptography and Data Security* (pp. 436–454). Springer Berlin Heidelberg. doi:10.1007/978-3-662-45472-5_28
- Jamie Blasco. (2013) *How cybercriminals are exploiting Bitcoin and other virtual currencies*. Retrieved October 4, 2015, from <http://www.alienvault.com/open-threatexchange/blog/how-cybercriminals-are-exploiting-bitcoin-and-other-virtual-currencies>
- Kalodner, H., Carlsten, M., Ellenbogen, P., Bonneau, J., & Narayanan, A. (n.d.). *An empirical study of Namecoin and lessons for decentralized namespace design*. Retrieved November 11, 2015, from http://www.econinfosec.org/archive/weis2015/papers/WEIS_2015_kalodner.pdf
- King, S., & Nadal, S. (2012). *PPCoin: Peer-to-Peer Crypto-Currency with Proof-of-Stake*. Retrieved November 11, 2015, from <http://peercoin.net: http://peercoin.net/assets/paper/peercoin-paper.pdf>

Litecoin. (n.d.). Retrieved November 11, 2015, from <https://litecoin.org/>

Love, D. (n.d.). *\$500 Million Worth of Bitcoin Has Been Stolen Since 2010*. Retrieved October 4, 2015, from <http://www.businessinsider.com/how-many-bitcoins-have-been-stolen-2014-3>

Moser, M. (2013). *Anonymity of bitcoin transactions*. Paper presented at Münster Bitcoin conference, Münster, Germany.

Mt. Gox. (n.d.). Retrieved 4 June, 2015, from [https://en.bitcoin.it/wiki/Mt. Gox](https://en.bitcoin.it/wiki/Mt._Gox)

Nakamoto, S. (2008). *Bitcoin: A peer-to-peer electronic cash system*. Retrieved from <http://www.cryptovest.co.uk/resources/Bitcoin%20paper%20Original.pdf>

Narayan, A. (2013). *Why the Cornell Paper on Bitcoin Mining is Important*. Retrieved June 10, 2015, from <https://freedom-to-tinker.com/blog/randomwalker/whythe-cornell-paper-on-bitcoin-mining-is-important/>

Scalability. (n.d.). Retrieved October 4, 2015, from <https://en.bitcoin.it/wiki/Scalability>

Script.CC. (n.d.). Retrieved November 11, 2015 from <https://script.cc/>

Stevenson, J. (2013). *Bitcoins, litecoins, what coins? A global phenomenon*. Google Commerce Ltd.

Yelowitz, A., & Wilson, M. (2015). Characteristics of Bitcoin users: An analysis of Google search data. *Applied Economics Letters*, 22(13), 1030–1036. doi:10.1080/13504851.2014.995359

ADDITIONAL READING

Barber, S., Boyen, X., Shi, E., & Uzun, E. (2012). Bitter to better—how to make bitcoin a better currency. In *Financial cryptography and data security* (pp. 399–414). Springer Berlin Heidelberg. doi:10.1007/978-3-642-32946-3_29

Moore, T. (2013). The promise and perils of digital currencies. Androulaki, E., Karame, G. O., Roeschlin, M., Scherer, T., & Capkun, S. (2013). Evaluating user privacy in bitcoin. In *Financial Cryptography and Data Security* (pp. 34–51). Springer Berlin Heidelberg.

Poelstra, A. (2014). *A Treatise on Altcoins*.

KEY TERMS AND DEFINITIONS

Altcoins: All the crypto-currencies other than the Bitcoin.

Crypto-Currencies: The type of digital currencies that use cryptography for their creation, security and transference.

Digital Currencies: The currencies that use internet for their transference.

Digital Signature: A mathematical scheme for ensuring the authenticity of a message. The sender encrypts the message using its private key in order to create a digital signature. By doing this, the sender adds his identity to the message since the private key is only owned by that particular sender.

Proof-of-Work: A defense mechanism against Denial of Service attacks in which the service requester has to perform some mathematical task before accessing the services.

Public Key Cryptography: An encryption technique in which each entity possesses a key pair (public/private key). In comparison with symmetric encryption in which single key is shared among the sender and the receiver, the public key cryptography involves two keys without the need of sharing them.

ENDNOTE

- ¹ The term bitcoin will be used in two contexts throughout the paper. Bitcoin (with 'B' in capitals) refers to the protocol suggested by Satoshi Nakamoto whereas bitcoin refers to the currency or the amount (often abbreviated as BTCs).

Category E

Electronic Services

Determining Impact of Demographics on Perceived Service Quality in Online Retail

Prateek Kalia

I. K. Gujral Punjab Technical University, India

Penny Law

Regenesys Business School, South Africa

Richa Arora

Regenesys Institute of Management, India

INTRODUCTION

Throughout the World Internet users, buyers and businesses are growing at an exuberant speed. Prevalence of computers and internet as one of the most influential technologies and its integration with business has made sale of goods and services through websites a profitable and low cost affair (Doostar, Akbari, & Abbasi, 2013). Businesses have realized that they can use internet as a powerful tool to increase overall service offerings (Griffith & Palmer, 1999). On the other hand, customers can avail benefit like convenience, availability of wide variety of product/service, competitive prices, extensive information, comparing alternatives etc. However, most of the retailers are selling similar products and gaining competitive advantages solely based on a cost leadership strategy is difficult (Jun, Yang, & Kim, 2004; Shankar, Smith, & Rangaswamy, 2003). In this scenario, researchers have pointed out that superior service quality can be critical for Internet retailers' long-term success (Fassnacht & Koese, 2006; Zeithaml, Parasuraman, & Malhotra, 2002). However, perception of the service quality can significantly differ between different customers, leading to difference in their satisfaction and future behavior (Sánchez-Pérez, Sánchez-Fernández, Marín-Carrillo, & Gázquez-Abad, 2007). Many

researchers have highlighted how demographic factors can influence customers' preference of online store visit (Phang, Kankanhalli, Ramakrishnan, & Raman, 2010), information search behavior (Kalia, Singh, & Kaur, 2016), consumer's online buying behavior (Li, Kuo, & Russell, 1999), differentiation of web-shoppers and non-shoppers (Karayanni, 2003) and evaluation of the e-service quality (Barrera, García, & Moreno, 2014). Ganesan-Lim, Russell-Bennett, & Dagger (2008) also mentioned in their literature review that individual consumers perceive service differently therefore quality perceptions may vary from one segment of the population to another. Acknowledging the fact that demographic information is essential for segmentation and targeting (McCarty & Shrum, 1993) or relevant in formulation of marketing or product strategy by internet retailers (Chang & Samuel, 2006), this study tries to understand whether significant difference in perceived service quality (PSQ) exist within demographic characteristics of online shoppers, such as education, age, gender, monthly income, occupation and marital status. This article is organized as follows: a literature review relevant to service quality and demographic effects on service quality perceptions is done to develop hypotheses. Then methodology and results are discussed. At the end conclusion and managerial implications are drawn.

BACKGROUND

Online Service Quality

To measure customer perceptions of service quality in service and retailing organizations Parasuraman, Zeithaml, & Berry (1988) developed a 22-item survey research instrument called SERVQUAL. Later, through focus group research with online shoppers, Zeithaml, Parasuraman, & Malhotra (2000) developed a framework for consumer evaluation of electronic service quality, known as e-SERVQUAL. They defined e-service quality (e-SQ) as, “the extent to which a website facilitates efficient and effective shopping, purchasing, and delivery” (Zeithaml et al., 2000). Their framework considered 11 dimensions of e-SQ i.e. access, ease of navigation, efficiency, flexibility, reliability, personalization, security/privacy, responsiveness, assurance/trust, site aesthetics, and price knowledge. On the basis of SERVQUAL number of scales for measuring online service quality were developed in subsequent years; for example, WebQual 1.0 (S. Barnes & Vidgen, 2000), PIRQUAL (J. Francis & White, 2002), WebQual 4.0 (Barnes & Vidgen, 2003), E-S-QUAL and e-RecS-Qual (Parasuraman, Zeithaml, & Malhotra, 2005), E-A-S-QUAL (M. Kim, Kim, & Lennon, 2006), eTransQual (Bauer, Falk, & Hammerschmidt, 2006) and eSELFQUAL (Ding, Hu, & Sheng, 2011). There was origin of some independent scales like SITEQUAL (Yoo & Donthu, 2001), WebQual (Loiacono, Watson, & Goodhue, 2002), IRSQ (Janda, Trocchia, & Gwinner, 2002), .comQ (Wolfinbarger & Gilly, 2002) and eTailQ (Wolfinbarger & Gilly, 2003). Number of researchers have highlighted why consumer perception of online service quality is important (Cai & Jun, 2003; Cheng, Wang, Lin, Chen, & Huang, 2008; J. E. Francis & White, 2002; Gounaris, Dimitriadis, & Stathakopoulos, 2005; Janda et al., 2002; Jun et al., 2004; Lee & Lin, 2005; Yang & Jun, 2002; Yoo & Donthu, 2001) and how service quality can significantly affect attributes like, loyalty (Dai, Haried, & Salam, 2011; Ding et al., 2011; J. Kim, Jin, & Swin-

ney, 2009; Srinivasan, Anderson, & Ponnnavolu, 2002; Swaid & Wigand, 2009; Wolfinbarger & Gilly, 2003; Zehir, Sehitoglu, Narcikara, & Zehir, 2014), satisfaction (Bauer et al., 2006; Cho & Park, 2001; Ding et al., 2011; J. Kim et al., 2009; S. Kim & Stoel, 2004; Szymanski & Hise, 2000; Wolfinbarger & Gilly, 2003; Yang, Peterson, & Cai, 2003), customer retention (Wolfinbarger & Gilly, 2003), perceived value (Bauer et al., 2006; Zehir et al., 2014), attitude toward the website (Wolfinbarger & Gilly, 2003), behavioral intentions (Collier & Bienstock, 2006), and service enjoyment and commitment (Dai et al., 2011).

Demographics and Service Quality

Different motivations drive shoppers to react differentially to diverse marketing messages (Moe, 2003). For instance, shoppers may search for product information actively or casually browse a webstore. This difference can be ascribed with customer's demographics and these variables offer valuable insights into ‘who consumers are’ and ‘what they need’ (Kalia, 2016a; Phang et al., 2010). For segmentation and targeting, researchers consider demographic information fundamentally necessary (McCarty & Shrum, 1993) and suggested to understand the affect of demographics like age, income and gender with respect to customer perceptions of quality (Lim et al., 2008). This section reviews prior research on the impact of demographics on service quality perceptions. As compared to previous studies, where two, three or maximum four demographic variables are undertaken; this study has comprehensively considered six demographic variables i.e. education, age, income, occupation, marital status and gender.

Education

Researchers observed that education level influence the evaluation of service quality (Min & Khoon, 2013). Vrechopoulos et al. (2001) discovered that Internet shoppers are mostly University graduates and postgraduates, simi-

larly Kalia (2016) reviewed that online shoppers are well educated and less resistant to change, more open-minded, venturesome, cosmopolitan in outlook, socially mobile, self-confident and mature. Barrera et al. (2014) highlighted that people without a university degree score perceived quality service higher than people with a university degree. Kumbhar (2011) argued that on the basis of education level perceived service quality, perceived value from e-banking service and overall satisfaction in e-banking vary among customers. Collectively, the preceding discussion leads to the following hypothesis.

H1: *Perceived service quality in online retail differ according to education.*

Age

Consumer age affects service quality perceptions (Lim et al., 2008). Reason to shop online vary in Internet shoppers according to their age, for example middle-aged customers (24-44) primarily shop online for 'convenience' than 'price' and 'product selection' (Chang & Samuel, 2006). On the other hand, older shoppers are more likely to prefer search/deliberation over hedonic browsing than younger shoppers (Phang et al., 2010). Pretorius and Smit (2010) observed difference in overall website satisfaction across the age groups and concluded that younger individuals, on average, are more satisfied with websites. Similarly, Barrera et al. (2014) found that people under 24 years perceive a greater service quality than those over 24 years. Few other studies also confirmed that internet shoppers are young and fall within the age bracket of 21-30 years (Kalia, 2016a) or 25-44 years (Vrechopoulos et al., 2001). Khare (2011) observed that Indian customers' perceptions towards the service quality of multinational banks differ across age categories, Kumbhar (2011) noticed that (Khare, 2011) perceived service quality, perceived value from e-banking service and overall satisfaction in e-banking differ with age group of the customer and Butler et

al. (1996) observed that age adequately predict one component of perceived hospital quality. In sum, the above discussion leads to the following hypothesis.

H2: *Perceived service quality in online retail differ according to age.*

Income

Internet shoppers who are motivated to purchase online vary in income. Middle-incomes (Aus\$40,000 to \$69,999) tend to be primarily influenced to purchase online for the reason of 'convenience' relative to 'price' and 'product selection' (Chang & Samuel, 2006). Kumbhar (2011) observed difference in perceived service quality, perceived value from e-banking service and overall satisfaction in e-banking due to varying income level of the customers. At large, online shoppers have higher or above average household/disposable income incomes (Kalia, 2016a). Vrechopoulos et al. (2001) also found that internet shoppers have average monthly household income of 300000-2000000 GRD. Above discussion lead to following hypothesis.

H3: *Perceived service quality in online retail differ according to income*

Occupation

Past studies have affirmed that perceived service quality, perceived value from e-banking service and overall satisfaction in e-banking differ by profession of the customers (Kumbhar, 2011). Vrechopoulos et al. (2001) discovered that Internet shoppers are largely scientists, private employees or freelancers. Kalia (2016) also reviewed that majority of online shoppers are wealthy or dual-career families with small children. Following hypothesis is proposed in the light of above discussion.

H4: *Perceived service quality in online retail differ according to occupation*

Marital Status

Marital status has been found to moderate the relationship between usefulness, enjoyment, external characteristics and reliability (Doostar et al., 2013) and Internet shoppers are mostly single (Vrechopoulos et al., 2001). Bhatnagar, Misra, & Rao (2000) concluded that marital status had no significant effect, except hardware category where marital status did have a significant effect. Majority of studies supported the fact that:

H5: *Perceived service quality in online retail differ according to marital status.*

Gender

Researchers have found that gender is influential in the evaluation of service quality (Min & Khoon, 2013) and Internet shoppers who are motivated to purchase online vary in gender or gender adequately predict components of perceived quality (Butler et al., 1996). Certain studies observed that women have a higher valuation of the service quality of Web sites than men (Barrera et al., 2014) and primarily, female are influenced to purchase online for the reason of 'convenience' relative to 'price' and 'product selection' (Chang & Samuel, 2006). Females also express greater overall website satisfaction (Pretorius & Smit, 2010). There have been studies which confirmed that gender moderates the relationship between usefulness, enjoyment, external characteristics and reliability (Doostar et al., 2013). Certain studies found that Internet buyers tend to be male than females, because women perceive a higher level of risk and engage in high exploratory behavior while buying online than men (Kalia, 2016a; Vrechopoulos et al., 2001). Khare (2011) found that Indian customers' perceptions towards the service quality of multinational banks differ between the two genders. On the basis of above literature review following hypothesis is proposed.

H6: *Perceived service quality in online retail differ according to gender.*

E

METHODOLOGY

Under methodology, research goal, sample and data collection are talked about. Subsequently, results of Kruskal-Wallis (H Test) to check significant difference in perceived service quality in online retail with respect to different demographic factors have been discussed. At the end results of post hoc test have been shared.

RESEARCH GOAL

There have been previous studies which have partially discussed demographics and service quality but there is shortage of studies which comprehensively and directly discuss the impact of demographics on perceived service quality in online retail. Further, this study is conducted in India, which is one of the most promising e-commerce market in Asia-Pacific region (eMarketer, 2015). India's online retail industry is estimated to grow to healthy 50-55 percent CAGR of Rs 504 billion by 2015-16; this indicates that Indian e-retail market has huge potential for future growth. However, due to sudden boom, competition in e-commerce market in India is getting intense (Kalia, 2016c); so much that e-retailers in India are adapting their business model to sustain and win (Kalia, 2015). Under such conditions, this research tries to determining the impact of demographics (education, age, income, occupation, marital status and gender) on perceived service quality in online retail.

SAMPLE AND DATA COLLECTION

An online survey was developed to test the hypothesis. Mail containing link to questionnaire was sent to respondents in three capital cities of

Northern India i.e. Chandigarh, Delhi/National Capital Region and Jaipur. Selection of these cities was based on eBay.in Census 2012 which identified these cities among top ten e-commerce hubs (Ebay.in, 2012). Hence, it was appropriate to collect data from respondents from these cities, who have made atleast one online purchase in past six months from popular e-retailers of India and gone through real experience of an online purchase. Snowball sampling (Malhotra, 2007) was deployed i.e. an initial group of respondents were selected randomly and subsequent respondents were selected based on the referrals. Considering 300 cases as good sample size (Comrey & Lee, 1992; Tabacnik & Fidell, 1996; Vanvoorhis & Morgan, 2007), total 308 responses have been used for data analysis. E-retailers in this research denotes business-to-consumer e-commerce companies selling products to consumers in categories like clothing & accessories, books & magazines, mobile phones & accessories, auto accessories & parts, memory cards, pen drives & HDD, watches, laptops & computer peripherals, shoes & other footwear, movies & music (CD/DVD) and home appliances. Scale developed by Jun et al. (2004), consisting twenty-one items to measure perceived online retail service quality has been used. Demographic profile of the respondents has been depicted in Table 1.

SCALE RELIABILITY AND CONSTRUCT VALIDITY

Cronbach's Alpha value of 0.928 confirmed the reliability of 21 item scale. KMO score of 0.926 (Kaiser & Rice, 1974) and Bartlett's Test score under 0.001 (significant) indicated that data is suitable for factor analysis. Items were checked for their communality scores and all the items scored greater than permissible limit i.e. <.30. Four factors accounted for 59.2 percent of the total variance. To examine the dimensions principal component factor analysis with varimax rotation was used. For purification of items many

researchers used 0.5 cut-off for loading scores and deleted items with primary factor loadings less than 0.5 (Cai & Jun, 2003; J. E. Francis & White, 2002; Jun et al., 2004; Long & McMellon, 2004). On similar lines primary factor loadings less than 0.5 were filtered out and total nineteen items were retained out of initial twenty-one items. Since all items have a factor loading greater than 0.5, therefore its confirmed that constructs are "unidimensional and factor distinct, and the items used to operationalize the constructs are all loaded onto a single factor" (Chong, 2013). Four factors were derived i.e. ease of use/attentiveness, access, reliable/prompt response and credibility/security. Cronbach's alpha reliability analysis results for these four factors extracted were 0.80,

Table 1. Demographic findings (n = 308)

Demographic Variables		n	%
City	Delhi/NCR	108	35.1
	Jaipur	93	30.2
	Chandigarh/Tricity	107	34.7
Education	Under Graduate	52	16.9
	Graduate	82	26.6
	Postgraduate	109	35.4
	Professional	65	21.1
Age	18-24 Years	137	44.5
	25-31 Years	127	41.2
	32-38 Years	32	10.4
	39 Years & Above	12	3.9
Gender	Male	168	54.5
	Female	140	45.5
Monthly Income	Less than 50,000	66	21.4
	50,001-80,000	72	23.4
	80,001-1,10,000	34	11.0
	1,10,001- 1,40,000	28	9.1
	More than 1,40,000	108	35.1
Profession	Business	28	9.1
	Service	173	56.2
	Student	107	34.7
Marital Status	Married	72	23.4
	Unmarried	236	76.6

0.82, 0.81 and 0.75 respectively. Cronbach α coefficient scores mentioned above indicated good reliability (Table 2).

ANALYSIS AND RESULTS

Kolmogorov-Smirnov one sample test revealed that the data needs application of non-parametric tests. Therefore, data has been analyzed by deploying Mann-Whitney U Test and Kruskal-Wallis H test (Table 3 and Table 4).

It was observed that no significant difference exists in perceived service quality in online retail with respect to education. Therefore, hypotheses H1 is rejected. The result is coherent with findings of Ilias et al. (2009) they found no difference in students' satisfaction towards service quality determinants and overall service quality on the basis of semester of studies and Phang et al. (2010), they reported that no significant difference exist between the shoppers adopting a search/deliberation strategy and those adopting a hedonic browsing strategy within different education groups.

No significant difference was found in perceived service quality in online retail with respect to age in this study. Therefore, hypotheses H2 is rejected. This finding is consistent with the results of Doostar et al. (2013), they discovered that age don't moderates the relationship between usefulness, enjoyment, external characteristics and reliability and Min and Khoon (2013), they discovered that age factor does not make any significant difference in the key elements of the service quality evaluation.

No significant difference was observed in perceived service quality in online retail with respect to income in this study. Therefore, hypotheses H3 is rejected. This result is coherent with findings of Ganesan-Lim et al., (2008) i.e. they noticed no differences in service quality perceptions on the basis of income. Phang et al. (2010) also concluded that no significant difference exist between the shoppers adopting a search/deliberation strategy and those adopting a hedonic browsing strategy within different income groups.

However, Kruskal-Wallis (H Test) values indicated that significant difference exist within occupational categories with respect to perceived service quality dimensions of "access" ($H(2) = 7.066, p = 0.029$) and "credibility/security" ($H(2) = 7.247, p = 0.027$). Hence, H4 has been accepted.

No significant difference was observed in perceived service quality in online retail with respect to marital status in this study. Therefore, hypotheses H5 is rejected. This result is similar to the conclusions of Kalia (2016) who profiled online shoppers and suggested that marital status has no significant effect.

It was observed that no significant difference exists in perceived service quality in online retail with respect to gender. Therefore, hypotheses H6 is rejected. There have been studies which support above finding, for example Ganesan-Lim et al. (2008) found no differences in service quality perceptions on the basis of gender, Ilias et al. (2009) concluded that there is no difference in students' satisfaction towards service quality determinants and overall service quality on the basis of gender, Kumbhar, (2011) noticed that there is no difference in perception of service quality, perceived value and overall satisfaction in e-banking by gender and Phang et al. (2010) brought out that there is no significant difference between males and females in their preference of search/ deliberation or hedonic browsing

POST HOC

On the basis of significant result obtained within different occupational categories, a post hoc test was applied. Kruskal-Wallis (H Test) values indicated that significant difference exist between businessmen and servicemen with respect to perceived service quality dimensions of "access" ($H(1) = 7.262, p = 0.007$) and "credibility/security" ($H(1) = 7.288, p = 0.007$). Also significant difference was found between businessmen and students with respect to perceived service quality dimension of "credibility/security" ($H(1) = 5.231, p = 0.022$). However, no significant difference

Table 2. Factor analysis results

Dimensions	Cronbach Alpha	% of Variance	Items	Components			
				1	2	3	4
Reliability	.816	5.588	The quantity and quality of the product/service I received was exactly the same as that I ordered.	.077	.013	.732	.173
			The product/service I ordered was delivered to me within the time promised by the online retailer.	.267	.125	.723	.183
			The billing process was accurately handled and its records were kept accurately.	.266	.200	.731	.044
			The online retailer responded to my inquiry promptly.	.376	.273	.518	.222
			When the online retailer promised to e-mail or call me by a certain time, it did so.	.175	.419	.538	.246
Access	.828	7.334	The Web site showed its street and e-mail addresses, and telephone and fax numbers.	.229	.582	.339	.105
			If I want to, I could easily contact a customer service representative over the telephone.	.428	.592	.278	.154
			The online retailer offered multiple ordering options such as telephone or mail options.	.276	.720	.178	.035
			For more information, I could turn to the online retailer's chat rooms, bulletin boards, or others.	.104	.781	.049	.345
			The Web site had a message area for customer questions and comments.	.257	.572	.033	.466
Ease of Use/ Attentiveness	.805	41.490	The Web site address was easy to remember.	.580	.175	.378	.095
			The organization and structure of online catalogs were logical and easy to follow.	.722	.209	.214	.158
			All the terms and conditions (e.g. payment, warranty, return policies) were easy to read/understand.	.679	.186	.198	.199
			The contents in the Web site were easy to understand.	.808	.227	.168	.117
			I received a personal "thank you" note via e-mail or other media after I placed an order.	.506	.225	.152	.305
Credibility/ Security	.753	4.839	I felt secure in providing personal information for online purchase.	.417	.182	.205	.567
			I felt the risk associated with online transactions was low.	.269	.009	.183	.678
			The Web site showed how long the online retailer has been in this online business.	.036	.344	.068	.749
			I received special rewards and discounts as promised from doing business with the online retailer.	.099	.158	.278	.660

between servicemen and students with respect to perceived service quality dimensions of "access" and "credibility/security" was found (Table 5).

FUTURE RESEARCH DIRECTIONS

Present study tried to decipher the differences in perceived service quality (PSQ) within demo-

graphic characteristics of online shoppers in India and found that significant difference in perceived service quality (PSQ) exist within occupational categories. Finding "what constitutes an appealing e-service quality vis-à-vis occupation can be interesting research. Another, important and recent phenomenon to explore can be how "Webographics" (Grossnickle & Raskin, 2001; Kalia, 2016b)

Table 3. Demographics and perceived service quality in online retail

Perceived Service Quality dimensions/ Demographics	Reliability					Access					EOU/Attentiveness					Credibility/Security				
	N	Mean Rank	χ^2	df	p-Value	N	Mean Rank	χ^2	df	p-Value	N	Mean Rank	χ^2	df	p-Value	N	Mean Rank	χ^2	df	p-Value
Education	52	168.63	2.124	3	.547	52	168.38	2.335	3	.506	52	165.76	1.393	3	.707	52	165.83	3.323	3	.344
	82	146.04				82	158.65				82	153.61				82	150.04			
	109	155.42				109	149.58				109	154.60				109	160.93			
	65	152.32				65	146.42				65	146.45				65	140.29			
Age	137	151.47	1.588	3	.662	137	156.72	.188	3	.980	137	151.91	1.321	3	.724	137	152.36	6.095	3	.107
	127	157.65				127	153.43				127	156.12				127	162.11			
	32	164.03				32	150.73				32	166.42				32	154.97			
	12	130.33				12	150.46				12	135.13				12	97.08			
Monthly family income	66	135.14	6.250	4	.181	66	161.17	.856	4	.931	66	152.53	3.173	4	.529	66	148.84	4.583	4	.333
	72	158.85				72	149.58				72	156.03				72	151.93			
	34	174.43				34	156.51				34	170.62				34	171.91			
	28	172.38				28	160.34				28	170.14				28	179.71			
	108	152.52				108	151.55				108	145.55				108	147.65			
Occupation	28	164.82	.504	2	.777	28	192.82	7.066	2	.029*	28	182.63	3.179	2	.204	28	196.71	7.247	2	.027*
	173	154.68				173	145.92				173	152.68				173	148.36			
	107	151.50				107	158.34				107	150.09				107	153.37			
Total	308																			

*Significant at 0.05

Table 4. Demographics and perceived service quality in online retail

Perceived Service Quality Dimensions/Demographics		Gender		Marital Status	
		Male	Female	Married	Unmarried
Reliability	N	168	140	72	236
	Mean Rank	157.71	150.65	163.22	151.84
	Sum of Ranks	26495.50	21090.50	11751.50	35834.50
	Mann-Whitney U	11220.500		7868.500	
	Wilcoxon W	21090.500		35834.500	
	Z	-.698		-.954	
	Asymp. Sig. (2-tailed)	.485		.340	
Access	N	168	140	72	236
	Mean Rank	155.15	153.73	147.35	156.68
	Sum of Ranks	26064.50	21521.50	10609.50	36976.50
	Mann-Whitney U	11651.500		7981.500	
	Wilcoxon W	21521.500		10609.500	
	Z	-.140		-.782	
	Asymp. Sig. (2-tailed)	.888		.434	
EOU/ Attentiveness	N	168	140	72	236
	Mean Rank	156.34	152.29	149.27	156.10
	Sum of Ranks	26265.50	21320.50	10747.50	36838.50
	Mann-Whitney U	11450.500		8119.500	
	Wilcoxon W	21320.500		10747.500	
	Z	-.401		-.574	
	Asymp. Sig. (2-tailed)	.688		.566	
Credibility/ Security	N	168	140	72	236
	Mean Rank	157.53	150.86	155.13	154.31
	Sum of Ranks	26465.50	21120.50	11169.00	36417.00
	Mann-Whitney U	11250.500		8451.000	
	Wilcoxon W	21120.500		36417.000	
	Z	-.660		-.069	
	Asymp. Sig. (2-tailed)	.509		.945	

*Significant at 0.05

can effect service quality or how demographics and webographics interact and affect e-service quality.

CONCLUSION AND IMPLICATIONS

This research establishes in general that no significant difference in perceived service quality (PSQ) exist within demographic characteristics

of online shoppers. However, one variable within which significant difference in perceived service quality (PSQ) exist is occupation. This study has various implications.

This is one of the first known study to probe the differences in perceived service quality (PSQ) within demographic characteristics of online shoppers in India. E-commerce growth in Asia-Pacific region will be driven by China, India and

Table 5. Post hoc: occupation and perceived service quality in online retail

		Access					Credibility/Security				
		N	Mean Rank	χ^2	df	p-Value	N	Mean Rank	χ^2	df	p-Value
1	Business	28	128.29	7.262	1	0.007*	28	128.29	7.288	1	0.007*
	Service	173	96.58				173	96.58			
2	Business	28	79.04	2.840	1	.092	28	82.93	5.231	1	0.022*
	Student	107	65.11				107	64.09			
3	Service	173	136.34	1.207	1	.272	173	138.78	.208	1	.649
	Student	107	147.22				107	143.28			

*Significant at 0.05

Indonesia with the latter two markets clocking growth at 129.5% and 65.6%, respectively, in 2015 (eMarketer, 2015; Kalia et al., 2016). This signals that India is one of the most promising e-commerce market. Despite this fact there is dearth of research which examine the profiles of Asian e-shoppers (Kau, Tang, & Ghose, 2003).

Therefore, a study of Indian online shoppers' behaviour will help the e-businesses to contrive appropriate marketing strategies. Significant difference within occupational categories with respect to perceived service quality dimensions of "access" and "credibility/security" signals that one size can't fit all. Therefore, while proposing scales to measure perceived service quality researchers have to be more careful. Further, results of post hoc test indicated significant difference between businessmen and servicemen with respect to perceived service quality dimensions of "access" and "credibility/security" and difference between businessmen and students with respect to perceived service quality dimension of "credibility/security". This situation brings an opportunity for an online retailer to create a simplified website interface which appeals to all occupational categories. Secondly, web retailers can utilize the demographic information filled by the user during first time registration on the website to optimize the website according to visitors "preferences and abilities". Third, web retailers can create two or more websites and deploy cookies or other identifiers to direct user to the appropriate website.

On the basis of the fact that occupation is having an effect on service quality availed by online consumers, there is a possibility of developing different product and service offers that suits or surpass the preferences identified to be above average within an occupation (Hasslinger, Hodzic, & Opazo, 2007).

REFERENCES

- Barnes, S., & Vidgen, R. (2000). WebQual : An Exploration of Web-site Quality. *Proceedings of the Eighth European Conference on Information Systems*.
- Barnes, S.J., & Vidgen, R. (2003). Measuring Web site quality improvements: A case study of the forum on strategic management knowledge exchange. *Industrial Management & Data Systems*, 103(5), 297–309. doi:10.1108/02635570310477352
- Barrera, R. B., García, A. N., & Moreno, M. R. (2014). Evaluation of the e-service quality in service encounters with incidents: Differences according to the socio-demographic profile of the online consumer. *Revista Europea de Dirección Y Economía de La Empresa*, 23(4), 184–193. <http://doi.org/10.1016/j.redes.2014.09.004>
- Bauer, H. H., Falk, T., & Hammerschmidt, M. (2006). eTransQual: A transaction process-based approach for capturing service quality in online shopping. *Journal of Business Research*, 59(7), 866–875. doi:10.1016/j.jbusres.2006.01.021

- Bhatnagar, A., Misra, S., & Rao, H. R. (2000). On risk, convenience, and Internet shopping behavior. *Communications of the ACM*, 43(11), 98–105. doi:10.1145/353360.353371
- Butler, D., Oswald, S. L., & Turner, D. E. (1996). The effects of demographics on determinants of perceived health-care service quality. The case of users and observers. *Journal of Management in Medicine*, 10(5), 8–20. doi:10.1108/02689239610146508 PMID:10166035
- Cai, S., & Jun, M. (2003). Internet users perceptions of online service quality: A comparison of online buyers and information searchers. *Managing Service Quality*, 13(6), 504–519. doi:10.1108/09604520310506568
- Chang, J., & Samuel, N. (2006). Why Purchase Online? An Empirical Study of Australian Internet Shoppers. *Studies in Business and Economics*, 12(1), 69–79. Retrieved from <http://qspace.qu.edu.qa/handle/10576/6861>
- Cheng, J. M.-S., Wang, E. S. T., Lin, J. Y. C., Chen, L. S. L., & Huang, W. H. (2008). Do extrinsic cues affect purchase risk at international e-tailers: The mediating effect of perceived e-tailer service quality. *Journal of Retailing and Consumer Services*, 15(5), 420–428. doi:10.1016/j.jretconser.2007.11.001
- Cho, N., & Park, S. (2001). Development of electronic commerce user-consumer satisfaction index (ECUSI) for Internet shopping. *Industrial Management & Data Systems*, 101(8), 400–405. doi:10.1108/EUM00000000006170
- Chong, A. Y.-L. (2013). Mobile commerce usage activities: The roles of demographic and motivation variables. *Technological Forecasting and Social Change*, 80(7), 1350–1359. doi:10.1016/j.techfore.2012.12.011
- Collier, J. E., & Bienstock, C. C. (2006). Measuring Service Quality in E-Retailing. *Journal of Service Research*, 8(3), 260–275. doi:10.1177/1094670505278867
- Comrey, A. L., & Lee, H. B. (1992). *A First Course in Factor Analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Dai, H., Haried, P., & Salam, A. F. (2011). Antecedents of online service quality, commitment and loyalty. *Journal of Computer Information Systems*, 52(2), 1–11. Retrieved from <http://www.iacis.org/jcis/articles/JCIS52-2-1.pdf>
- Ding, D. X., Hu, P. J.-H., & Sheng, O. R. L. (2011). e-SELFQUAL: A scale for measuring online self-service quality. *Journal of Business Research*, 64(5), 508–515. doi:10.1016/j.jbusres.2010.04.007
- Doostar, M., Akbari, M., & Abbasi, R. (2013). Impact of Demographic Characteristics on Relationship between Customers ' Perceived Service Quality and Websites ' Services in Electronic Markets. *International Research Journal of Applied and Basic Sciences*, 5(5), 530–537.
- Ebay.in. (2012). *ebay.in Census 2012 - India under a lens*. Retrieved March 9, 2013, from <http://pics.ebaystatic.com/aw/pics/in/aboutebay/ebay-census-guide-2012.pdf>
- eMarketer. (2015). *Asia Pacific Is Home to Majority of World Retail Ecommerce Market*. Retrieved March 26, 2016, from <http://www.emarketer.com/Article/Asia-Pacific-Home-Majority-of-World-Retail-Ecommerce-Market/1013352>
- Fassnacht, M., & Koesel, I. (2006). Quality of Electronic Services: Conceptualizing and Testing a Hierarchical Model. *Journal of Service Research*, 9(1), 19–37. doi:10.1177/1094670506289531
- Francis, J., & White, L. (2002). PIRQUAL: A scale for measuring customer expectations and perceptions of quality in internet retailing. In *AMA Winter Educators' Conference: Marketing Theory and Applications* (pp. 263–270).
- Francis, J. E., & White, L. (2002). Exploratory and Confirmatory Factor Analysis of the Perceived Internet Retailing Quality Model. In *ANZMAC 2002 Conference Proceedings* (pp. 1791–1796).

- Gounaris, S., Dimitriadis, S., & Stathakopoulos, V. (2005). Antecedents of Perceived Quality in the Context of Internet Retail Stores. *Journal of Marketing Management*, 21(7–8), 669–700. doi:10.1362/026725705774538390
- Griffith, D. A., & Palmer, J. W. (1999). Leveraging the web for corporate success. *Business Horizons*, 42(1), 3–10. doi:10.1016/S0007-6813(99)80043-3
- Grossnickle, J., & Raskin, O. (2001). *Handbook of Online Marketing Research*. New York: McGraw-Hill.
- Hasslinger, A., Hodzic, S., & Opazo, C. (2007). *Consumer Behaviour in Online Shopping*. Retrieved from <http://hkr.diva-portal.org/smash/get/diva2:231179/FULLTEXT01.pdf>
- Ilias, A. I., Hasan, H. F. A., & Rahman, R. A. (2009). Student Satisfaction and Service Quality: Any Differences in Demographic Factors? *International Business Research*, 1(4), 131–143. doi:10.5539/ibr.v1n4p131
- Janda, S., Trocchia, P. J., & Gwinner, K. P. (2002). Consumer perceptions of Internet retail service quality. *International Journal of Service Industry Management*, 13(5), 412–431. doi:10.1108/09564230210447913
- Jun, M., Yang, Z., & Kim, D. (2004). Customers perceptions of online retailing service quality and their satisfaction. *International Journal of Quality & Reliability Management*, 21(8), 817–840. doi:10.1108/02656710410551728
- Kaiser, H. F., & Rice, J. (1974). Little Jiffy, Mark IV. *Educational and Psychological Measurement*, 34(1), 111–117. doi:10.1177/001316447403400115
- Kalia, P. (2015). Tope-Retailers of India: Business model and components. *International Journal of Electronic Marketing and Retailing*, 6(4), 277–298. doi:10.1504/IJEMR.2015.073450
- Kalia, P. (2016a). Demographic Profile of Online Shoppers: An Overview. *Indian Journal of Economics and Development*, 12(1a), 37–42. doi:10.5958/2322-0430.2016.00032.9
- Kalia, P. (2016b). Determining effect of webographics on customer's purchase frequency in e-retail. *Journal of Internet Banking and Commerce*, 21(2), 1–24. Retrieved from <http://www.icommercecentral.com/open-access/determining-effect-of-webographics-on-customers-purchase-frequency-in-eretail.php?aid=78448>
- Kalia, P. (2016c). Tsunami E-Commerce in India: The Third Wave. *The Global Analyst*, 5(7), 47–49.
- Kalia, P., Kaur, N., & Singh, T. (2016). E-Commerce in India: Evolution and Revolution of Online Retail. In S. Dixit & A. K. Sinha (Eds.), *E-Retailing Challenges and Opportunities in the Global Marketplace* (pp. 99–120). Hershey, PA: IGI Global; doi:10.4018/978-1-4666-9921-2.ch005
- Kalia, P., Singh, T., & Kaur, N. (2016). An Empirical Study of Online Shoppers' Search Behaviour with Respect to Sources of Information in Northern India. *Productivity : A Quarterly Journal of the National Productivity Council*, 56(4), 353–361.
- Karayanni, A. D. (2003). Web-shoppers and non-shoppers: Compatibility, relative advantage and demographics. *European Business Review*, 15(3), 141–152. doi:10.1108/09555340310474640
- Kau, A. K., Tang, Y. E., & Ghose, S. (2003). Typology of online shoppers. *Journal of Consumer Marketing*, 20(2), 139–156. doi:10.1108/07363760310464604
- Khare, A. (2011). Customers perception and attitude towards service quality in multinational banks in India. *International Journal of Services and Operations Management*, 10(2), 199. doi:10.1504/IJSOM.2011.042517

- Kim, J., Jin, B., & Swinney, J. L. (2009). The role of e-tail quality, e-satisfaction and e-trust in online loyalty development process. *Journal of Retailing and Consumer Services*, 16(4), 239–247. doi:10.1016/j.jretconser.2008.11.019
- Kim, M., Kim, J.-H., & Lennon, S. J. (2006). Online service attributes available on apparel retail web sites: An E-S-QUAL approach. *Managing Service Quality*, 16(1), 51–77. doi:10.1108/09604520610639964
- Kim, S., & Stoel, L. (2004). Apparel retailers: Website quality dimensions and satisfaction. *Journal of Retailing and Consumer Services*, 11(2), 109–117. doi:10.1016/S0969-6989(03)00010-9
- Kumbhar, V.M. (2011). Customers' Demographic Profile and Satisfaction in E-Banking Services : A Study of Indian Banks. *International Journal for Business, Strategy & Management*, 1(1), 1–9.
- Lee, G.-G., & Lin, H.-F. (2005). Customer perceptions of e-service quality in online shopping. *International Journal of Retail & Distribution Management*, 33(2), 161–176. doi:10.1108/09590550510581485
- Li, H., Kuo, C., & Russell, M. G. (1999). *The Impact of Perceived Channel Utilities, Shopping Orientations, and Demographics on the Consumer's Online Buying Behavior*. Retrieved January 26, 2015, from <http://doi.wiley.com/10.1111/j.1083-6101.1999.tb00336.x>
- Lim, C. G., Russell-Bennett, R., & Dagger, T. (2008). The Impact of Service Contact Type and Demographic Characteristics on Service Quality Perceptions. *Journal of Services Marketing*, 22(7), 550–561. doi:10.1108/08876040810909677
- Loiacono, E. T., Watson, R. T., & Goodhue, D. L. (2002). *Webqual : A Measure of Website Quality* (pp. 432–438). American Marketing Association.
- Long, M., & McMellon, C. (2004). Exploring the determinants of retail service quality on the Internet. *Journal of Services Marketing*, 18(1), 78–90. doi:10.1108/08876040410520726
- Malhotra, N. K. (2007). *Marketing Research: An Applied Orientation (5th ed.)*. Delhi: Prentice-hall Of India Pvt Ltd. doi:10.1108/S1548-6435(2007)3
- McCarty, J. A., & Shrum, L. J. (1993). The Role of Personal Values and Demographics in Predicting Television Viewing Behavior : Implications for Theory and Application. *Journal of Advertising*, 22(4), 77–101. doi:10.1080/00913367.1993.10673420
- Min, S., & Khoon, C. C. (2013). Demographic Factors in the Evaluation of Service Quality in Higher Education: International Students' Perspective. *International Review of Management and Business Research*, 2(4), 994–1010. doi:10.5539/ijms.v6n1p90
- Moe, W. W. (2003). Buying, Searching, or Browsing: Differentiating Between Online Shoppers Using In-Store Navigational Clickstream. *Journal of Consumer Psychology*, 13(1–2), 29–39. doi:10.1207/S15327663JCP13-1&2_03
- Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1988). SERVQUAL : A Multiple-Item Scale for Measuring Consumer Perceptions of Service Quality. *Journal of Retailing*, 64(1), 12–40.
- Parasuraman, A., Zeithaml, V. A., & Malhotra, A. (2005). E-S-QUAL: A Multiple-Item Scale for Assessing Electronic Service Quality. *Journal of Service Research*, 7(10), 1–21. doi:10.1177/1094670504271156
- Phang, C. W., Kankanhalli, A., Ramakrishnan, K., & Raman, K. S. (2010). Customers preference of online store visit strategies: An investigation of demographic variables. *European Journal of Information Systems*, 19(3), 344–358. doi:10.1057/ejis.2010.32
- Pretorius, H., & Smit, E. (2010). Internet apprehension and website perceptions : demographic and webographic effects. In SAMRA 2010 Conference. Johannesburg, South Africa: South African Marketing Research Association. Retrieved from <http://www.samra.co.za/wp-content/uploads/2015/06/Pretorius-H.-Smit-E.-Internet-apprehension-and-website-perceptions.pdf>

- Sánchez-Pérez, M., Sánchez-Fernández, R., Marín-Carrillo, G. M., & Gázquez-Abad, J. C. (2007). Service Quality in Public Services as a Segmentation Variable. *Service Industries Journal*, 27(4), 355–369. doi:10.1080/02642060701346771
- Shankar, V., Smith, A. K., & Rangaswamy, A. (2003). Customer satisfaction and loyalty in online and offline environments. *International Journal of Research in Marketing*, 20(2), 153–175. doi:10.1016/S0167-8116(03)00016-8
- Srinivasan, S. S., Anderson, R., & Ponnnavolu, K. (2002). Customer loyalty in e-commerce: An exploration of its antecedents and consequences. *Journal of Retailing*, 78(1), 41–50. doi:10.1016/S0022-4359(01)00065-3
- Swaid, S. I., & Wigand, R. T. (2009). Measuring the Quality Of E-Service: Scale Development and Initial Validation. *Journal of Electronic Commerce Research*, 10(1), 13–28.
- Szymanski, D. M., & Hise, R. T. (2000). E-satisfaction: An initial examination. *Journal of Retailing*, 76(3), 309–322. doi:10.1016/S0022-4359(00)00035-X
- Tabacnik, & Fidell, L. S. (1996). *Using Multivariate Statistics* (3rd ed.). New York: HarperCollins.
- Vanvoorhis, C. R. W., & Morgan, B. L. (2007). Understanding Power and Rules of Thumb for Determining Sample Sizes. *Tutorials in Quantitative Methods for Psychology*, 3(2), 43–50. doi:10.20982/tqmp.03.2.p043
- Vrechopoulos, A. P., Siomkos, G. J., & Doukidis, G. I. (2001). Internet shopping adoption by Greek consumers. *European Journal of Innovation Management*, 4(3), 142–153. doi:10.1108/14601060110399306
- Wolfenbarger, M., & Gilly, M. C. (2002). comQ: Dimensionalizing, Measuring and Predicting Quality of the E-tail Experience (No. 02–100). Cambridge, MA: Academic Press.
- Wolfenbarger, M., & Gilly, M. C. (2003). eTailQ: Dimensionalizing, measuring and predictingetail quality. *Journal of Retailing*, 79(3), 183–198. doi:10.1016/S0022-4359(03)00034-4
- Yang, Z., & Jun, M. (2002). Consumer Perception of E-Service Quality: From Internet Purchaser and Non-purchaser Perspectives. *The Journal of Business Strategy*, 19(1), 19–41. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&profile=ehost&scope=site&authtype=crawler&jrnl=08872058&AN=36541870&h=wtTKHP5iLBc7uq52RBO1HoYAuIxW0EUjC-QSOeLhHISmo1CRnNz5bFJgnMIRD/TUQ1bOQAJKvONc55oDVXb9qQ==&crl=c>
- Yang, Z., Peterson, R. T., & Cai, S. (2003). Services quality dimensions of Internet retailing: An exploratory analysis. *Journal of Services Marketing*, 17(7), 685–700. doi:10.1108/08876040310501241
- Yoo, B., & Donthu, N. (2001). Developing a Scale to Measure the Perceived Quality of An Internet Shopping Site (SITEQUAL). *Quarterly Journal of Electronic Commerce*, 2(1), 31–47.
- Zehir, C., Sehitoglu, Y., Narcikara, E., & Zehir, S. (2014). E-S-Quality, Perceived Value and Loyalty Intentions Relationships in Internet Retailers. In *Procedia - Social and Behavioral Sciences* (Vol. 150, pp. 1071–1079). Elsevier B.V. <http://doi.org/doi:10.1016/j.sbspro.2014.09.120>
- Zeithaml, V. A., Parasuraman, A., & Malhotra, A. (2000). *A Conceptual Framework for Understanding e-Service Quality: Implications for Future Research and Managerial Practice* (Vol. 115). Marketing Science Institute.
- Zeithaml, V. A., Parasuraman, A., & Malhotra, A. (2002). Service Quality Delivery through Web Sites: A Critical Review of Extant Knowledge. *Journal of the Academy of Marketing Science*, 30(4), 362–375. doi:10.1177/009207002236911

ADDITIONAL READING

Dixit, S., & Sinha, A. K. (2016). *E-Retailing Challenges and Opportunities in the Global Marketplace*. Hershey, PA: IGI Global; doi:10.4018/978-1-4666-9921-2

Khosrow-Pour, M. (2006). *Encyclopedia of E-Commerce, E-Government, and Mobile Commerce* (Vols. 1–2). Hershey, PA: IGI Global; doi:10.4018/978-1-59140-799-7

Khosrow-Pour, M. (2010). *E-Commerce Trends for Organizational Advancement: New Applications and Methods*. Hershey, PA: IGI Global; doi:10.4018/978-1-60566-964-9

KEY TERMS AND DEFINITIONS

Access: Ease of contact and information gathering from the online retailer (Jun et al., 2004).

Attentiveness: The extent to which online retailers provide personalized services to their customers (Jun et al., 2004).

Credibility: Trustworthiness and believability of the online retailers (Jun et al., 2004).

Digital Buyers: Consumers who directly buy goods or services from a seller over the Internet using a web browser.

Ease of Use: Easy navigation of online systems, well-organized and well-structured online catalogs, concise contents, and easy-to-understand terms and conditions in e-retailers Web sites (Jun et al., 2004).

E-Retailer: Virtual merchant that sells goods and services to the public via online transactions.

E-Service Quality: The extent to which a website facilitates efficient and effective shopping, purchasing, and delivery (Valarie A Zeithaml et al., 2000).

Reliable/Prompt Response (2004): The ability to perform the promised service accurately dependably, promptly, and timely. Jun et al.

Security: Concerned with online transaction safety and customer privacy (Jun et al., 2004).

The Impact of Carbon Nanotubes and Graphene on Electronics Industry

E**Rafael Vargas-Bernal***Instituto Tecnológico Superior de Irapuato, Mexico***Gabriel Herrera-Pérez***Instituto Tecnológico Superior de Irapuato, Mexico***Margarita Tecpoyotl-Torres***Universidad Autónoma del Estado de Morelos, Mexico*

INTRODUCTION

A great technological opportunity has been introduced with the discovery of carbon nanomaterials where carbon nanotubes and graphene are their main representatives, which are completely involved in the science, technology and applications on new electronic devices, as an interesting alternative to those devices based on conventional semiconductor materials. They have attracted significantly attention of the scientific and industrial communities thanks to their unique electrical, optical, thermal, mechanical, and chemical properties. Nevertheless, carbon nanotubes and graphene due to their insignificant size, these require be homogeneously embedded into dielectric and light-weight, matrices, which are regularly of the polymeric type giving place to polymer-matrix composite materials (Das, 2013). Several models have been developed to predict the behavior of electrical conductivity of these composite materials based on carbon nanomaterials (Vargas-Bernal, 2013). Various chemical strategies for achieving synthesis of these composites have been proposed around the world, but these are not studied in this paper. The future trend in electronics industry consists in using individual carbon nanotubes and graphene sheets or set of them to design devices, circuits or systems. A lot of electronic devices have been fabricated based

on carbon nanotubes and graphene: field-effect transistors, diodes, analog and digital circuits, sensors (biosensors, gas sensors, etc.), solar cells, batteries, supercapacitors, flexible displays, etc. In this paper, the impact of the carbon nanotubes and graphene in electronic devices (discrete devices and/or integrated circuits), optoelectronic devices, photovoltaic devices, energy storage devices, and sensors, is briefly reviewed. This paper has been divided as follows: basic concepts about carbon nanotubes and graphene are described in section entitled Background. Next, different applications in the electronic industry are discussed in section entitled Applications of the Carbon Nanotube and Graphene. A comparison between the performance of the carbon nanotubes and graphene is presented in the next section. Future research directions are described in an additional section. Finally, conclusions about this study are given in the end section.

BACKGROUND

Carbon nanomaterials possess unique properties that can be exploited electrical, thermal, chemical and mechanically to provide applications in areas such as composite materials, energy storage and conversion, sensors, drug delivery, field emission devices, and nanoscale electronic components.

Three different morphologies between carbon nanomaterials can be distinguished: carbon nanotubes, fullerenes and graphene. Carbon nanotubes and graphene have been used more extensively in the electronic industry. Next, a brief description of their properties is made with the purpose of knowing the advantages of these materials for electronic applications. The basic concepts of each type of material have been separated for a better description.

Carbon Nanotubes

A carbon nanotube can be defined as a set of cylinder-shaped graphite sheets. Carbon nanotubes (CNTs) can be categorized by the number of graphite layers in their structure: Single-Wall Nanotubes (SWNTs) containing a single layer (see Figure 1), Double-Wall Nanotubes (DWNTs) with two layers, and Multi-Wall Nanotubes (MWNTs) that contain more of two layers (see Figure 2). Two main physical properties are associated directly with electronic applications: high thermal and electrical conductivity (Vargas-Bernal, 2012). Carbon nanotubes and their compounds exhibit extraordinary electrical properties useful for electronic organic materials, and can be used for a wide range of new and existing applications such as conductive plastics, flat-panel displays either as flexible displays or touch screens, micro-and nano-electronics (transistors), radar-absorbing coating, ultra-capacitors, solar cells, batteries with improved lifetime, hydrogen storage cells, conductors, smart textiles, electrochemical sensors, biosensors, and gas sensors. A more detailed study of their properties is found in Vargas-Bernal, 2012.

Graphene

Graphene is a two-dimensional allotrope of carbon with hexagonal lattice in atomic-scale, as shown in Figure 3 (Vargas-Bernal, 2015c). This material is a thin layer of pure carbon where each atom has four bonds, one σ -bond with each of its three neighbors and one π -bond that is oriented out the plane. Just

as the carbon nanotube, graphene has many extraordinary properties such as 100 times stronger than steel by weight, conducts heat and electricity with great efficiency and is almost transparent. From the electronic point of view, it presents the effect of a bipolar transistor, ballistic transport of electrical charge, and large quantum oscillations. A lot of electronic devices have been developed

Figure 1. Structure of the single-wall carbon nanotube (SWNT)

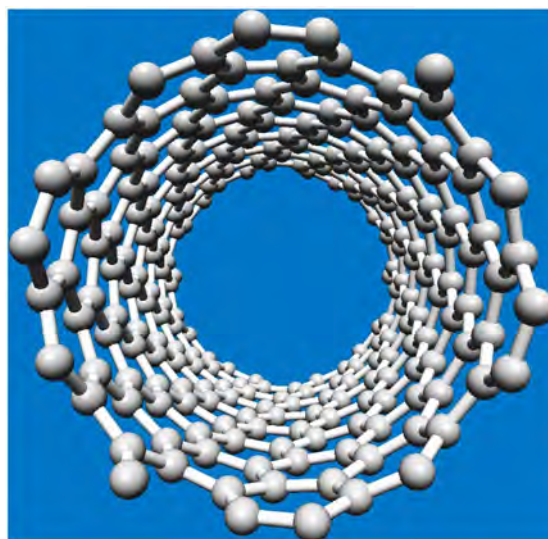


Figure 2. Structure of the multi-wall carbon nanotube (MWNT)

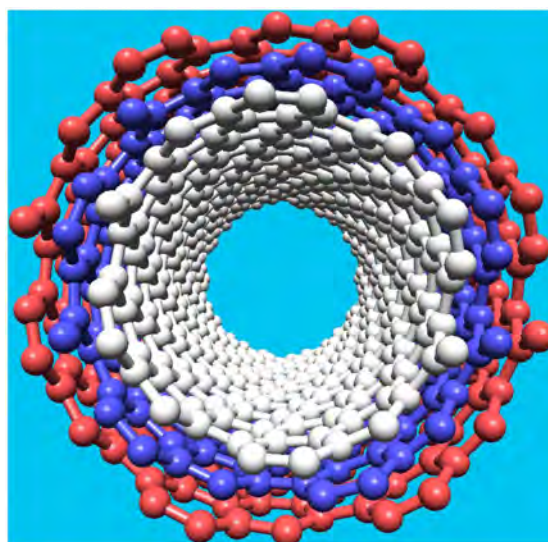
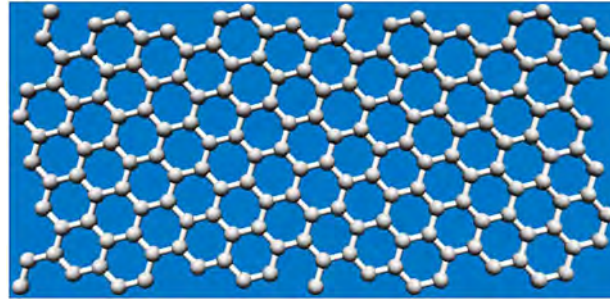


Figure 3. Structure of the graphene



using graphene in areas such as integrated circuits, sensors, optoelectronics, photovoltaics, and energy storage. A more detailed study of their properties is found in Vargas-Bernal, 2015c.

APPLICATIONS OF CARBON NANOTUBES AND GRAPHENE

Material scientists and engineers have been developing new architectures and devices based on carbon nanomaterials. A brief description of the multiple applications that carbon nanotubes and graphene have provided in the electronics industry is presented below. The applications of each material have been separated for a better description.

Carbon Nanotubes

Carbon nanotubes compounds are used in electronic packaging (IC trays and wafer carriers), transport trays for hard disk drives (HDD) and components, and ICs test and burn sockets, with the aim of providing protection from electrostatic discharges (ESD), homogeneous electrical conductivities on the surface, overheating, dimensional stability, recyclability, wafer hazing, ionic outgassing, static decay time, and high cleanliness requirements. Carbon nanotubes and graphene embedded in polymeric matrices can increase the efficiency of the electromagnetic interference (EMI) shielding to circuits operating to very high frequencies (Vargas-Bernal, 2015a). Moreover, carbon nanotubes represent a better alternative

to provide efficient and high-speed, electrical interconnects among devices within integrated circuit (Vargas-Bernal, 2015b), since they can offer reduction to the effects presented by copper, such as electromigration, large values of parasitic elements, large delays, and high thermal dissipation. Interconnect faults can be presented during device operation of integrated circuits originated by open-circuits and short-circuits. The possibility of self-repairing in integrated circuits is being studied through the use of carbon nanotubes dispersed in an insulating fluid to repair open-circuit interconnect faults by applying electric field across the gap between a pair of materials that have been disconnected (Sambandan, 2012).

Regularly, CNTs have been used as channels in (FET) field-effect transistors or CNTFETs either *p*-type or *n*-type (Vargas-Bernal, 2012). When CNTs are used to fabricate integrated circuits (ICs), there exist problems to obtain similar properties due to that metallic and semiconducting CNTs and both types of nanotubes can be found mixed during their synthesis process. A modularized method for fabricating ICs by using individual CNTs, has been proposed and experimentally demonstrated (Pei, 2014). Logic functions including AND, NOT, OR, NAND, NOR, and XOR have been designed based on carbon nanotubes. Digital circuits based on CNTFETs have the potential to improve the energy-delay product with respect to the conventional circuits based on silicon. The first computer based on CNTFETs was verified by running an operating system based on multitasks and was reviewed using 20 different instructions

of a commercial microprocessor with an instruction set of the type Interlocked Pipeline Stages (MIPS) (Shulaker, 2013).

The exhaustive use of portable and flexible systems has led to the development of power sources, such as primary batteries, rechargeable batteries (notebook computers and mobile phones), and supercapacitors (De Volder, 2013). These devices must be easily implemented in products such as smart clothings, flexible displays, transdermal delivery patches, etc. Carbon nanotubes embedded in a composite are considered very attractive materials for these applications due to their high electrical conductivity, large surface-to-volume ratio, as well as, by their thermal and mechanical properties (Wang, Z. 2014). Carbon nanotubes act as electrical fillers in thin films to generate lightweight current collectors and electrodes used in flexible alkaline batteries such as those based on lithium-ions (Li, S. 2011). Composites based on carbon nanotubes and vanadium oxide, can be used in binder-free flexible electrodes for supercapacitors, which offer large thickness, controlled porosity, high electrical conductivity and cyclic stability (Boukhalfa, 2012). In this application, carbon nanotubes reduce the electrical capacitance and allow the development of electrodes with high cycling rates for high-power applications, where charge-discharge processes of voltage are required in very short times. Three different groups of energy storage device are exploiting carbon nanotubes in their design: (1) industrial and stationary applications, (2) automotive and transportation applications, and (3) portable and personal electronics applications (Li, X. 2013).

Properties of carbon nanotubes such as large surface-to-volume ratio, high electrical conductivity and good catalytic activity are being applied in the development of counter electrodes in solar cells. By example, vertically aligned carbon nanotubes (VACNTs) possess high degree of order, controllability, and easy manipulation to be considered as strategic elements in energy conversion (Li, S. 2011) such as organic solar cells. Well-aligned CNTs improve the electrical properties

of the solar cells increasing the ion transport and reducing the charge transfer resistance.

CNTs have high mobility, high on/off ratio, excellent mechanical properties, and transparency. Therefore, they can be considered as a serious candidate to be used as base material to replace indium tin oxide (ITO) in electrodes of flexible displays (Yu, 2011). Devices for imaging based on visible light and X-ray can make use of organic photodetectors using thin-film transistors based on carbon nanotubes (Takahashi, 2013). These devices take advantage of the high mobility of carbon nanotube transistors for operating to low voltages, which enormously reduces the power consumption.

By example, potentiometric biosensors based on single-walled carbon nanotubes can behave as transducers ion-to-electron and aptamers can be used as recognition elements (Zelada-Guillén, 2012). This type of biosensors can be prepared either covalently or non-covalently, with the possibility of obtaining simplicity, stability, short analysis time and better detection limit or high sensitivity, respectively. The contribution of the carbon nanotubes is associated with the possibility of realizing detection in real-time and label-free, as well as high selectivity, high selectivity, which are difficult of achieving with other materials (Justino, 2013).

Graphene

Three main properties can be exploited in graphene for electronic applications: ultra-high mobility, low electric field, and high saturation velocity (Wang, Z.X. 2012). Thus, graphene is considered a promising material to design field-effect transistors (GFETs) used in high-frequency analog applications such as phase detectors, frequency doublers, and radio-frequency mixers (Lin, 2011). These transistors can achieve cutoff frequency of ~ 500 GHz due to the high mobility of carriers exploited in the graphene (Vargas-Bernal, 2015c). Even oscillator circuits based on devices operating to THz are being designed (Díaz, 2014), where

graphene is the electrical-channel to interconnect the terminals of drain and source of transistors. Frequency multipliers based on field-effect transistors (GFETs) were introduced in 2009, and these have been used as frequency-doubling devices (Ramón, 2012) operating in a 3-GHz bandwidth with low capacitance and high mobility for applications in security, short-range communication, and molecular spectroscopy.

Materials such as graphene present problems to be used as switch in logic circuits, since it has not cutoff current due to its ambipolar curve characteristic (Vargas-Bernal, 2015c), that is, a symmetric relation between I_{DS} and V_{GS} (drain-to-source current against gate-to-source voltage) is presented by their devices. Moreover, graphene is less prone than copper, to effects such as electromigration, low thermal conductivity, low mechanical strength, and high corrosion, which favor graphene to be considered as a better material for interconnects in integrated circuits (Prasad, 2014).

Graphene is being used to develop transparent electrodes to substitute those electrodes based on indium tin oxide (ITO) by taking advantage of ultrathin films with electrical and transparent behavior to the electrical current applied (Yu, 2011). Carbon nanotubes can be used in electron emission screens transverse based on graphene (Lei, 2013), with the aim of minimizing electrostatic shielding induced by the bulk substrate.

High sensitivity to the electrical current (semiconductor with bandgap zero) is exploited in biosensors based on graphene FETs, thanks to the absence of cutoff current and the characteristic of a high mobility of the electrical carriers during the phase of sensing (Pumera, 2011). Graphene due to their electro-activity and transparency is mainly used for enzymatic biosensing, DNA sensing, and immunosensing. Biosensors based on graphene FETs present low working potentials, high sensitivities, low detection limits, and long-term stabilities (Kuila, 2011).

Graphene has a relatively good, frequency response, and its light weight suggests their use in audio speakers and microphones as diaphragms

subjected to pressure waves in air by mechanically vibrating (Zhou, 2013). Highly sensitive Hall effect sensors can be developed thanks to the high electron mobility in the graphene, to be used in DC current transformers for special applications (Petruk, 2014) or scanning Hall probe microscopy (Sonusen, 2014).

Nowadays, a lot of interest is being put in the development of solar cells with low-cost production, simple fabrication and high energy conversion efficiency. A solar cell requires a counter electrode material with low sheet resistance, high reduction catalytic activity, good chemical stability, and low production costs. Graphene represents an alternative material to fabricate counter electrodes for solar cells with the possibility of matching properties such as electrical conductivity and the reduced catalytic activity found in platinum. Nitrogen-doped carbon nanomaterials can act as metal-free electrodes with electrocatalytic activity, long-term operation stability, and tolerance to effects such as crossover/poisoning, which can be exploited in the fabrication of solar cells (Xue, 2012). The performance of counter electrode is directly related with short-circuit current, open-circuit voltage, fill factor, and power conversion efficiency of a solar cell.

Graphene electrodes can be used to replace conventional electrodes based on indium tin oxide (ITO) for flexible displays (Yu, 2011). The use of graphene is due to their high transmittance, high optical transparency, high electrical conductivity and extraordinary mechanical properties (Jo, 2012). These electrodes can be used in transistors, light-emitting diodes, solar cells and flexible devices.

The use of graphene in photovoltaic devices, fuel cells, batteries, supercapacitors, and devices for hydrogen generation offer an interesting alternative to fulfill the increasing global energy demand (Bonaccorso, 2015). The advantage of using 2D materials is completely related with the ease of creating and designing layered artificial structures with multiple electrical properties. In particular, electrodes for supercapacitors and

lithium-ion batteries, gas storage tanks, as well as electrocatalysts and catalyst supports for fuel cells, are being developed (Candelaria, 2012). Different types of supercapacitors based on 2D structures have been proposed: electrochemical double-layer capacitors, pseudo-capacitors, and asymmetric supercapacitors (Huang, 2012). Moreover, graphene-based supercapacitors can be designed using five different types of nanoarchitectures (Salunkeh, 2014): (1) graphene doped with heteroatoms, (2) activated graphene, (3) composites based on metal oxide and graphene, (4) composites graphene-polymer, and (5) asymmetric supercapacitors based on graphene. Graphene is attractive as electrode material for energy storage thanks to their high specific surface area, interconnected porous structure, pore size similar to the electrolyte ions, good wettability toward the electrolyte, and high electrical conductivity.

COMPARISON BETWEEN CARBON NANOTUBES AND GRAPHENE

A great debate about which of the two allotropes have greater significance in the electronics industry has arisen among researchers who work either with developments based on carbon nanotubes or those working with graphene. CNTs as graphene are often regarded rivals in electronic applications, however, they are completely related, since carbon nanotubes are the zipped version of the graphene and their properties are compatible. Carbon nanotubes can be used as a reinforcing rebar for graphene during its manufacturing process. Moreover, the carbon nanotubes can be heated to produce graphene.

A performance comparison between FET devices based on carbon nanotubes and those based on graphene has shown that devices based on graphene have higher mobility and higher current density than those based on carbon nanotubes (Biswas, 2011). In addition, devices based on carbon nanotubes present a high on/off ratio with respect to those based on graphene. Metal electrodes are necessary to connect devices based on

CNTs, while that these are not required by devices based on graphene. Moreover, CNTs have lower transmittance, lower electrical conductivity and higher sheet resistance than graphene. However, both materials are highly flexible. Graphene and carbon nanotubes must be fully processed in a standard, reliable and repeatable manner to be used in consumer electronics on a large scale. In addition, they must demonstrate high yield, high speeds, and low power consumption to be led to the commercial applications.

The main obstacle to obtain practical applications with carbon nanotubes has been hampered by the intrinsic difficulty to connecting them within an electronic device. In the case of graphene, it has zero bandgap, which complicates their switching properties and therefore, its use in digital circuits.

FUTURE RESEARCH DIRECTIONS

Carbon nanotubes and graphene are being incorporated into existing and future electronic devices, and as a result they are being considered as strategic materials to substitute to the silicon in the next-generation of electronic products. Moreover, carbon nanomaterials have not one single area of application, and therefore, the researchers have found difficulty in knowing the limits that these materials have in the electronics industry. Thus, a lot of novel and interesting applications will be designed in the next twenty years, which will make use of carbon nanomaterials. During this time is hoped that money and expertise will be invested around of the world to develop and to research carbon nanomaterials applied in electronics industry.

Novel manufacturing processes must be developed to scale up them and lowering production costs to achieve the maximum performance of the graphene with respect to other materials. New methods to reduce power dissipation and increase the stability in devices directed to flexible electronics must be developed to lead to the carbon nanotubes into large-scale production. A significant decreasing in off-current magnitude, degree of

hysteresis, variation in threshold voltage and bias stress degradation presented by transistors based on carbon nanotubes must be achieved. For both materials, new strategies to develop composite or hybrid materials are necessary with the aim of guaranteeing the efficient use of carbon nanotubes and graphene by exploiting all their physical properties in different electronic applications.

CONCLUSION

This article provides a small review of the multiple applications of the carbon nanotubes and graphene, both in present as well as a projection of the future applications directed to the electronic sector. The next generation of electronic devices requires using materials with very high performance to fulfill the extraordinary requirements imposed by the novel applications. The graphene and carbon nanotube market directed to the electronics industry will rise by a CAGR of 60%. Thus, it is expected that an exponential boom related with research and technological development about graphene and carbon nanotubes could be prolonged until the year 2020. Physical properties such as thermo-electric conduction, high surface area to volume ratio, and mechanical strength have inspired a huge quantity of applications in electronic industry. Until now, a lot of applications have been designed and some of them are found in the industry. The success of the carbon nanomaterials depends of the ordering these materials at hierarchical scales, starting from large-scale dispersions and films and it finishes into ordered macrostructures and nanoscale devices in the near future.

ACKNOWLEDGMENT

The first author want thank the economical support to the research project financed by CONACYT-Mexico (152524) and by the Tecnológico Nacional de México under project 1671, both projects are associated with composites based on polymers and carbon nanotubes.

REFERENCES

- Biswas, C., & Lee, Y. H. (2011). Graphene versus Carbon Nanotubes in Electronic Devices. *Advanced Functional Materials*, 21(20), 3806–3826. doi:10.1002/adfm.201101241
- Bonaccorso, F., Colombo, L., Yu, G., Stöller, M., Tozzini, V., Ferrari, A. C., & Pellegrini, V. et al. (2015). Graphene, Related Two-Dimensional Crystals, and Hybrid Systems for Energy Conversion and Storage. *Science*, 347(6217), 1246501. doi:10.1126/science.1246501 PMID:25554791
- Boukhalfa, S., Evanoff, K., & Yushi, G. (2012). Atomic Layer Deposition of Vanadium Oxide on Carbon Nanotubes for High-Power Supercapacitor Electrodes. *Energy & Environmental Science*, 5(5), 6872–6879. doi:10.1039/c2ee21110f
- Candelaria, S.L., Shao, Y., Zhou, W., Li, X., Xia, J., Zhang, J.-G., Wang, Y., Liu, J., Li, J. & Cao, G. (2012). Nanostructured Carbon for Energy Storage and Conversion. *Nano Energy*, 1(2), 195–220. doi:10.1016/j.nanoen.2011.11.006
- Das, T. K., & Prusty, S. (2013). Graphene-based Polymer Composites and Their Applications. *Polymer-Plastics Technology and Engineering*, 52(4), 319–331. doi:10.1080/03602559.2012.751410
- De Volder, M. F. L., Tawfick, S. H., Baughman, R. H., & Hart, A. J. (2013). Carbon Nanotubes: Present and Future Commercial Applications. *Science*, 339(6119), 535–539. doi:10.1126/science.1222453 PMID:23372006
- Díaz, E., Miralles, K., Dominguez-Adame, F., & Gaul, C. (2014). Spin-dependent Terahertz Oscillator based on Hybrid Graphene Superlattices. *Applied Physics Letters*, 105(10), 103109. doi:10.1063/1.4895673
- Huang, Y., Liang, J., & Chen, Y. (2012). An Overview of the Applications of Graphene-based Materials in Supercapacitors. *Small*, 8(12), 1805–1834. doi:10.1002/smll.201102635 PMID:22514114

- Jo, G., Choe, M., Lee, S., Park, W., Kahng, Y. H., & Lee, T. (2012). The Application of Graphene as Electrodes in Electrical and Optical Devices. *Nanotechnology*, 23(11), 112001. doi:10.1088/0957-4484/23/11/112001 PMID:22370228
- Justino, C. I. L., Rocha-Santos, T. A. P., Duarte, A. C., & Rocha-Santos, T. A. P. (2013). Advances in Point-of-Care Technologies with Biosensors based on Carbon Nanotubes. *Trends in Analytical Chemistry*, 45, 24–36. doi:10.1016/j.trac.2012.12.012
- Kuila, T., Bose, S., Khanra, P., Mishra, A. K., Kim, N. H., & Lee, J. H. (2011). Recent Advances in Graphene-based Biosensors. *Biosensors & Bioelectronics*, 26(12), 4637–4648. doi:10.1016/j.bios.2011.05.039 PMID:21683572
- Lei, W., Li, C., Cole, M. T., Qu, K., Ding, S., Zhang, Y., & Milne, W. I. et al. (2013). A Graphene-based Large Area Surface-Conduction Electron Emission Display. *Carbon*, 56, 255–263. doi:10.1016/j.carbon.2013.01.004
- Li, S., Luo, Y., Lv, W., Yu, W., Wu, S., Hou, P., & Cheng, H.-M. et al. (2011). Vertically Aligned Carbon Nanotubes grown on Graphene Paper as Electrodes in Lithium-Ion Batteries and Dye-Sensitized Solar Cells. *Advanced Energy Materials*, 1(4), 486–490. doi:10.1002/aenm.201100001
- Li, X. & Wei, B. (2013). Supercapacitors based on Nanostructured Carbon. *Nano Energy*, 2(2), 159–173. doi: 10.1016/j.nanoen.2012.09.008
- Lin, Y.-M., Valdes-Garcia, A., Han, S.-J., Farmer, D. B., Meric, I., Sun, Y., & Jenkins, K. A. et al. (2011). Wafer-Scale Graphene Integrated Circuit. *Science*, 332(6035), 1294–1297. doi:10.1126/science.1204428 PMID:21659599
- Pei, T., Zhang, P., Zhang, Z., Qiu, C., Liang, S., Yang, Y., & Peng, L.-M. et al. (2014). Modularized Construction of General Integrated Circuits on Individual Carbon Nanotubes. *Nano Letters*, 14(6), 3102–3109. doi:10.1021/nl5001604 PMID:24796796
- Petruck, O., Szewczyk, R., Ciuk, T., Strupiński, W., Salach, J., Nowicki, M., & Trzcinka, K. et al. (2014). Sensitivity and Offset Voltage Testing in the Hall-Effect Sensors made of Graphene. In R. Szewczyk, C. Zieliński, & M. Kaliczyńska (Eds.), *Recent Advances in Automation, Robotics and Measuring Techniques* (pp. 631–640). Springer. doi:10.1007/978-3-319-05353-0_60
- Prasad, N., Kumari, A., Bhatnagar, P. K., Mathur, P. C., & Bhatia, C. S. (2014). Current Induced Annealing and Electrical Characterization of Single Layer Graphene grown by Chemical Vapor Deposition for Future Interconnects in VLSI Circuits. *Applied Physics Letters*, 105(11), 113513. doi:10.1063/1.4895802
- Pumera, M. (2011). Graphene in Biosensing. *Materials Today*, 14(7-8), 308–315. doi:10.1016/S1369-7021(11)70160-2
- Ramón, M. E., Parrish, K. N., Chowdhury, S. F., Magnuson, C. W., Movva, H. C. P., Ruoff, R. S., & Akinwande, D. et al. (2012). Three-Gigahertz Graphene Frequency Doubler on Quartz Operating beyond the Transit Frequency. *IEEE Transactions on Nanotechnology*, 11(5), 877–883. doi:10.1109/TNANO.2012.2203826
- Salunkhe, R. R., Lee, Y.-H., Chang, K.-H., Li, J.-M., Simon, P., Tang, J., & Yamauchi, Y. et al. (2014). Nanoarchitected Graphene-based Supercapacitors for Next-Generation Energy-Storage Applications. *Chemistry (Weinheim an der Bergstrasse, Germany)*, 20(43), 13838–13852. doi:10.1002/chem.201403649 PMID:25251360
- Sambandan, S. (2012). Self Repair in Circuits – Automating Open Fault Repair in Integrated Circuits using Field-Induced Aggregation of Carbon Nanotubes. *IEEE Transactions on Electron Devices*, 59(6), 1773–1779. doi:10.1109/TED.2012.2191557
- Shulaker, M. M., Hills, G., Patil, N., Wei, H., Chen, H.-Y., Wong, H.-S. P., & Mitra, S. (2013). Carbon Nanotube Computer. *Nature*, 501(7468), 526–530. doi:10.1038/nature12502 PMID:24067711

- Sonusen, S., Karci, O., Dede, M., Aksoy, S., & Oral, A. (2014). Single Layer Graphene Hall Sensors for Scanning Hall Probe Microscopy (SHPM) in 3300 K Temperature Range. *Applied Surface Science*, 308, 414–418. doi:10.1016/j.apsusc.2014.04.191
- Takahashi, T., Yu, Z., Chen, K., Kiriya, D., Wang, C., Takei, K., & Javey, A. et al. (2013). Carbon Nanotube Active-Matrix Backplanes for Mechanically Flexible Visible Light and X-Ray Imagers. *Nano Letters*, 13(11), 5425–5430. doi:10.1021/nl403001r PMID:24143966
- Vargas-Bernal, R. (2015a). Performance Analysis of Electromagnetic Interference Shielding based on Carbon Nanomaterials used in AMS/RF IC Design. In M. Fakhfakh, E. Tlelo-Cuautle, & M. H. Fino (Eds.), *Performance Optimization Techniques in Analog, Mixed-Signal, and Radio-Frequency Circuit Design* (pp. 268–294). Hersey, PA: IGI Global; doi:10.4018/978-1-4666-6627-6.ch011
- Vargas-Bernal, R. (2015b). Performance Analysis of Interconnects based on Carbon Nanotubes for AMS/RF IC Design. In M. Fakhfakh, E. Tlelo-Cuautle, & M. H. Fino (Eds.), *Performance Optimization Techniques in Analog, Mixed-Signal, and Radio-Frequency Circuit Design* (pp. 336–363). Hersey, PA: IGI Global; doi:10.4018/978-1-4666-6627-6.ch014
- Vargas-Bernal, R. (2015c). Advances in Computational Modeling of Electronic Devices based on Graphene. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 5(1), 109–116. doi:10.1109/JETCAS.2015.2398219
- Vargas-Bernal, R., & Herrera-Pérez, G. (2012). Carbon Nanotube- and Graphene based Devices, Circuits and Sensors for VLSI Design. In E. Tlelo-Cuautle & S. Tan (Eds.), *VLSI Design* (pp. 41–66). Rijeka, Croatia: Intech. doi:10.5772/38743
- Vargas-Bernal, R., Herrera-Pérez, G., Calixto-Olade, M. E., & Tecpoyotl-Torres, M. (2013). Analysis of DC Electrical Conductivity Models of Carbon Nanotube-Polymer Composites with Potential Application to Nanometric Electronic Devices. *Journal of Electrical and Computer Engineering*, 179538. doi:10.1155/2013/179538
- Wang, Z., Wu, Z., Bramnik, N., & Mitra, S. (2014). Fabrication of High-Performance Flexible Alkaline Batteries by Implementing Multiwalled Carbon Nanotubes and Copolymer Separator. *Advanced Materials*, 26(6), 970–976. doi:10.1002/adma.201304020 PMID:24510667
- Wang, Z. X., Zhang, Z. Y., & Peng, L. M. (2012). Graphene-based Ambipolar Electronics for Radio Frequency Applications. *Chinese Science Bulletin*, 57(23), 2956–2970. doi:10.1007/s11434-012-5143-x
- Xue, Y., Liu, J., Chen, H., Wang, R., Li, D., Qu, L., & Dai, L. (2012). Nitrogen-doped Graphene Foams as Metal-Free Counter Electrodes in High-Performance Dye-Sensitized Solar Cells. *Angewandte Chemie International Edition*, 51(48), 12124–12127. doi:10.1002/anie.201207277 PMID:23124846
- Yu, W. J., Lee, S. Y., Chae, S. H., Perello, D., Han, G. H., Yun, M., & Lee, Y. H. (2011). Small Hysteresis Nanocarbon-based Integrated Circuits on Flexible and Transparent Plastic Substrate. *Nano Letters*, 11(3), 1344–1350. doi:10.1021/nl104488z PMID:21322606
- Zelada-Guillén, G. A., Sebastián-Avila, J. L., Blondeau, P., Riu, J., & Rius, F. X. (2012). Label-Free Detection of *Staphylococcus aureus* in Skin using Real-Time Potentiometric Biosensors based on Carbon Nanotubes and Aptamers. *Biosensors & Bioelectronics*, 31(1), 226–232. doi:10.1016/j.bios.2011.10.021 PMID:22154169
- Zhou, Q., & Zettl, A. (2013). Electrostatic Graphene Loudspeaker. *Applied Physics Letters*, 102(22), 223109. doi:10.1063/1.4806974

ADDITIONAL READING

- Ding, L., Zhang, Z., Pei, T., Liang, S., Wang, S., Zhou, W., & Peng, L.-M. et al. (2012). Carbon Nanotube Field-Effect Transistors for Use as Pass Transistors in Integrated Logic Gates and Full Subtractor Circuits. *ACS Nano*, 6(5), 4013–4019. doi:10.1021/nn300320j PMID:22482426
- Franklin, A. D., Luisier, M., Han, S.-J., Tulevski, G., Breslin, C. M., Gignac, L., & Haensch, W. et al. (2012). Sub-10 nm Carbon Nanotube Transistor. *Nano Letters*, 12(2), 758–762. doi:10.1021/nl203701g PMID:22260387
- Iwan, A., & Chuchmała, A. (2012). Perspectives of Applied Graphene: Polymer Solar Cells. *Progress in Polymer Science*, 37(12), 1805–1828. doi:10.1016/j.progpolymsci.2012.08.001
- Kim, B., Chung, H., & Kim, W. (2012). High-Performance Supercapacitors based on Vertically Aligned Carbon Nanotubes and Non-aqueous Electrolytes. *Nanotechnology*, 23(15), 155401. doi:10.1088/0957-4484/23/15/155401 PMID:22437007
- Liu, F., Lee, C. W., & Im, J. S. (2013). Graphene-based Carbon Materials for Electrochemical Energy Storage. *Journal of Nanomaterials*, 642915. doi:10.1155/2013/642915
- Liu, J., Xue, Y., Zhang, M., & Dai, L. (2012). Graphene-based Materials for Energy Applications. *MRS Bulletin*, 37(12), 1265–1272. doi:10.1557/mrs.2012.179
- Park, H., Afzali, A., Han, S.-J., Tulevski, G. S., Franklin, A. D., Tersoff, J., & Haensch, W. et al. (2012). High-Density Integration of Carbon Nanotubes via Chemical Self-Assembly. *Nature Nanotechnology*, 7(12), 787–791. doi:10.1038/nnano.2012.189 PMID:23103933
- Wang, H., Nezich, D., Kong, J., & Palacios, T. (2009). Graphene Frequency Multipliers. *IEEE Electron Device Letters*, 30(5), 547–549. doi:10.1109/LED.2009.2016443
- Zhu, X., Zhu, Y., Murali, S., Stoller, M. D., & Ruoff, R. S. (2011). Nanostructured Reduced Graphene Oxide/Fe₂O₃ Composite as a High-Performance Anode Material for Lithium Ion Batteries. *ACS Nano*, 5(4), 3333–3338. doi:10.1021/nn200493r PMID:21443243
- Zhu, Z., Garcia-Gancedo, L., Flewitt, A. J., Xie, H., Moussy, F., & Milne, W. I. (2012). A Critical Review of Glucose Biosensors based on Carbon Nanomaterials: Carbon Nanotubes and Graphene. *Sensors (Basel, Switzerland)*, 12(5), 5996–6022. doi:10.3390/s120505996 PMID:22778628
- Zurutuza, A., & Marinelli, C. (2014). Challenges and Opportunities in Graphene Commercialization. *Nature Nanotechnology*, 9(10), 730–734. doi:10.1038/nnano.2014.225 PMID:25286257

KEY TERMS AND DEFINITIONS

Carbon Nanomaterials: Nanostructures of carbon such as fullerenes, carbon nanotubes, nanofibers and graphene with unique physico-chemical properties with multiple technological applications.

Carbon Nanotubes: Allotropes of carbon with a cylindrical nanostructure of length-to-diameter of up to 132,000,000:1, which have unusual properties and valuable for nanotechnology, electronics, optics and other fields of materials science and technology.

Composites: Materials made from two or more constituents materials with significantly different physical and/or chemical properties, that when are combined, produce a material with different characteristics from the individual components.

Current Density: The electric current per unit area of cross section, and as vectorial quantity it has magnitude, direction, and sense.

Electrical Conductivity: The physical property that quantifies how strongly a given material opposes the flow of electrical current. A low resistivity indicates a material that readily allows the movement of electrical charge.

Electromagnetic Shielding: The action of reducing and/or isolating the electromagnetic field in a space by blocking the electromagnetic waves with barriers made of conductive and/or magnetic materials.

Graphene: A two-dimensional, crystalline allotrope of carbon whose atoms are densely packed in a regular sp^2 -bonded atomic-scale chicken wire (hexagonal) pattern composed by a one-atom thick layer of graphite.

IC Design: A subfield of electrical engineering, encompassing the particular logic and circuit

design techniques required to design integrated circuits, or ICs. ICs consist of miniaturized electronic devices built into an electrical network on a monolithic semiconductor substrate by photolithography.

Interconnect: A path of material that connects two elements or components in an IC, through which electrical current is transported.

Microelectronics: The study and manufacture (or microfabrication) of very small electronic designs and components normally in micrometer-scale or smaller by means of semiconductors.

E

Integrating Content Authentication Support in Media Services

Anastasia N. Katsaounidou

Aristotle University of Thessaloniki, Greece

Charalampos A. Dimoulas

Aristotle University of Thessaloniki, Greece

INTRODUCTION

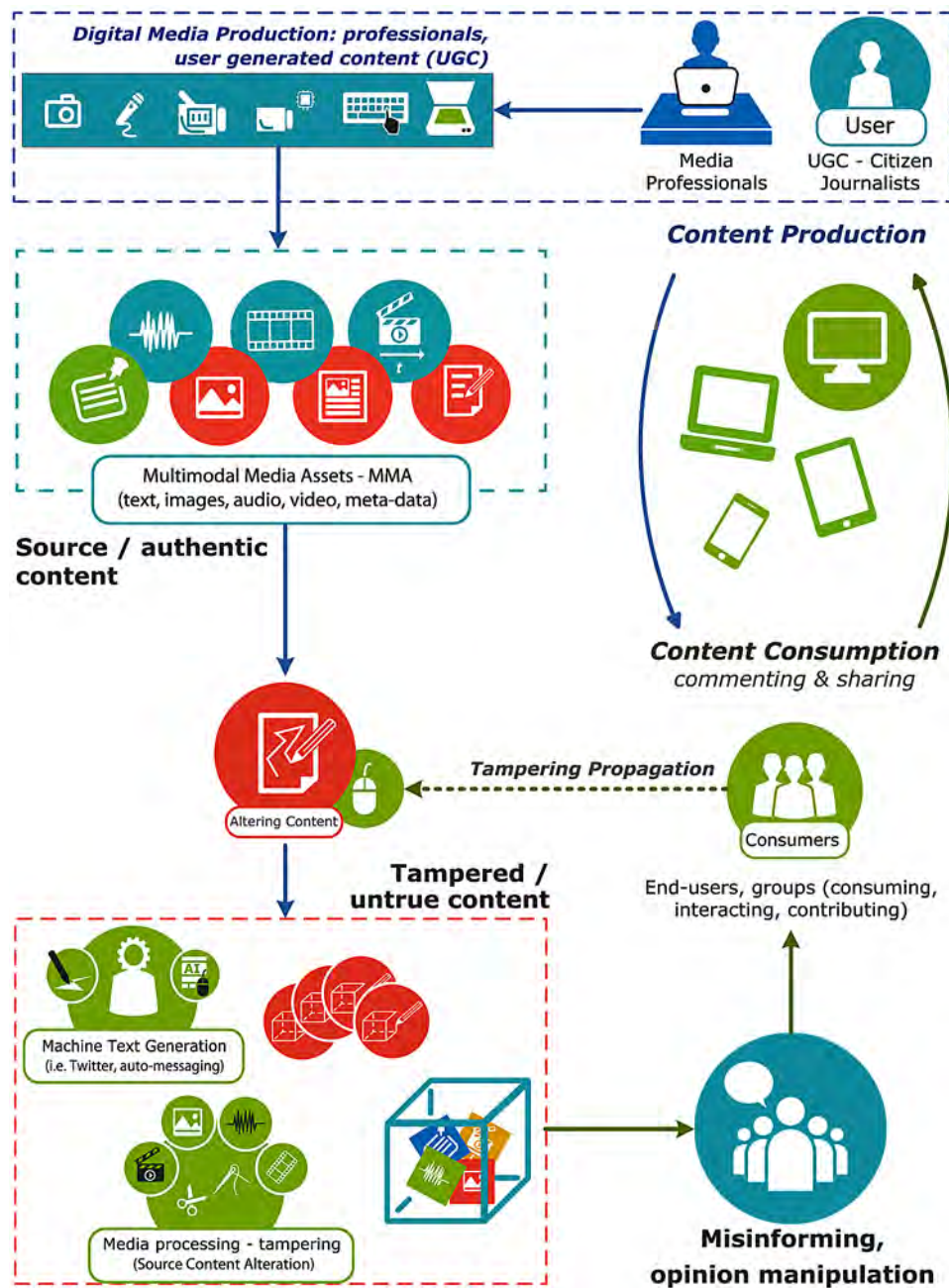
The tremendous evolution of Information and Communication Technologies (ICT) and the low cost of the digital media devices have fueled the widespread expansion of the so-called User Generated Content (UGC). Social networking has become a popular way for users to meet and interact online through text and audiovisual content (photos, sounds, videos, etc.) that is produced and distributed in real time. Traditionally, users share news and multimodal content through social media, while simultaneously discover information on the Web. Among others, trust is considered to be one of the crucial factors of information capturing and dissemination. In the road from Web 2.0 to Web 3.0 and beyond, the quality and credibility of the recorded, shared and broadcasted content is controversial (Ljung & Wahlforss, 2008; Matsiola, Dimoulas, Kalliris & Veglis, 2015). Many (easy to use) multimedia capturing and processing tools (desktop applications and online /cloud services) are currently available and can be exploited literately at any time and place through mobile devices. This “processing at the fingertip” vision familiarizes average users with multimodal media production, processing and management tasks (Dimoulas, Veglis & Kalliris, 2014, 2015).

The domination of digital content over traditional analog media (i.e. films and tapes) has given rise to a number of new information security challenges. Digital content can be processed,

intentionally altered or falsified and redistributed relatively easy. This has important consequences for governmental, commercial, social and professional media organizations that rely on digital information (Stamm, Wu & Liu, 2013). Hence, mass communication and journalistic processes can be associated to unwanted content tampering, construction of fake evidences, sharing and propagation of untrue stories. In particular, the universality of “digital news reporting” has turned the evaluation of shared media into a field of prime importance, focusing on the automatic detection of manipulated and misused Web content (Mendoza, Poblete & Castillo, 2010). Its aim is to lay the basis for a future generation of tools that could assist media professionals in the process of verification. According to Figure 1, where the problem definition of the discussed topic is presented, media alteration involves all content types (text, images, audio, video, etc.) that are encountered in today’s Multimodal Media Assets (Dimoulas et al., 2014; Katsaounidou, 2016).

Content alteration can be conducted by anyone involved in the media production processes (media professionals, UGC–citizen journalist, etc.), as the bluish arrows in Figure 1 indicate. Once information falsification occurs without being noticed by the users, uncontrolled propagation of untrue stories may appear as a side effect of the contemporary need for timely and immediate informing. Hence, tampered information can be massively shared /propagated by end-users/con-

Figure 1. Problem definition: (multi)-media tampering as part of digital informing /infotainment



sumers (greenish-dotted arrows in Figure 1). The outmost target of the current chapter is to describe a collaborating model for overall supporting content authentication through dedicated computerized environments. In this context, users' and journalists' training (and their valuable feedback) holds a key role towards the integration and unification

of applicable media veracity services (and their associated learning resources). Thus, algorithms, methodologies and related ground-truth data-sets would be continuously updated, progressed and adapted to the specific needs of the encountered application scenarios.

BACKGROUND

Multimodal items can be evaluated and verified with a variety of methods. In principle, different authentication strategies are associated to various content entities, based on their distinct communication and operation attributes. Context-aware is the simplest authenticity analysis approach that can be applied in most media types, purposing to detect and pull abnormal information out of the content by exploiting human observation and cognition abilities (Silverman, 2013; Katsaounidou, 2016). Missing or unexpected shadows /reflections are common examples in the case of tampered images (Johnson & Farid, 2007; Farid, 2009; Krawetz, 2007). Syntax errors and/or absence of meaning indicate potential tampering in text messaging (Zubiaga & Ji, 2014). Abruptly interrupted background sound, unexpected intonation, and pitch changes form “novelty detection” indicators that reveal possible audio forgery regions (Dimoulas & Symeonidis, 2015; Gupta, Cho & Kuo, 2012; Hua, Zhang, Goh & Thing, 2016; Kotsakis, Mislou, Kalliris & Matsiola, 2015). Various combinations of the aforementioned audiovisual parameters are used in video /multimedia forgery detection (Angelov, Sadeghi, Tehran & Ramezani, 2011; Ho & Li, 2015; Kompatsiaris, Merialdo & Lian, 2012; Thakur, 2014). In this context, both machine-assisted and human computing methods exist (Figure 2), where the former utilize Natural Language Processing and Machine Learning algorithms, while the latter rely on crowdsourcing and experts’ suggestions (Silverman, 2013). Furthermore, semi-automated methods are used to detect cases in which the deception was committed on the semantic level, such as reusing older images (and other media) in different, irrelevant contexts. There are also approaches that do not exclusively rely on contextual or high-level information (i.e. format- and production-related metadata), but also examine structural features and content-based low level descriptors (i.e. frequency of letters/words in text; visual parameters like color, structure, shape; video motion activity metrics; audio

volume, spectral features, etc.) (Farid, 2009; Ho & Li, 2015; Katsaounidou, 2016; Thakur, 2014; Wang, 2009).

In practice, most of the aforementioned approaches require users’ active participation for drawing conclusions, while others incorporate learning-based decision systems (although they are not such common in real-world use scenarios). Nevertheless, enormous research progress has been conducted regarding the implementation of fully-/semi-automated content authentication algorithms. The purpose of the so-called forensic techniques is to trace a multimedia file processing history, to detect potential manipulation and generally to identify all possible forgeries. These methods can be grouped into five basic categories: a) device fingerprints-based that search for inconsistencies in the associated capturing patterns; b) compression- and coding-specific fingerprints that examine the presence of multiple compression levels; c) content statistics-based (i.e. image /noise histograms) that detect outliers indicating abnormal behavior; d) manipulation-specific fingerprints that investigate specific processing trails; and e) techniques that search for physical inconsistencies in the multimedia content (i.e. unexpected geometric artifacts).

As presented in Figure 1, tampering can be applied to any content entity that can be part of today’s multimedia information exchange. Following the above main directions, various methods and algorithms have been implemented for the different media types, both at research and application level. For instance, Twitter has become the quintessential platform to follow and be aware of breaking news and ongoing events. The fast pace of the tweeting streams favors the appearance of fake reports that often challenge the identification of accurate information to get rid of hoaxes. This endangers the trustability of tweets, as the veracity of some pieces of information can seem questionable on many occasions (Zubiaga & Ji, 2014). As Twitter grows more popular, new automations on ranking the credibility of the users and their posts are needed. Textual features, like the tweet-

length (number of containing characters/words), the frequency of question/exclamation marks or the number of the included uppercase characters, form typical evaluation parameters (Boididou, Papadopoulos, Kompatsiaris, Schifferes & Newman, 2014). Sentiment analysis (by counting the number of words with predefined emotional strength) could help in checking inconsistencies between the discussed topic and the composed messages, whereas user-related features can be also engaged (number of friends /followers, account verification status, etc.). Another approach is to classify posts in different topics, which has proven to increase authenticity recognition accuracy through statistical analysis. Thus, tweets' variance and the overall impact of noisy data are eliminated (Jin, Cao, Zhang & Zhang, 2015).

With millions of users capturing, recreating and sharing multimedia events, experiences and places around the world through social media, images and videos have flooded the Web, proving their power to influence society (Papadopoulos, Cesar, Shamma & Kelliher, 2015). Contemporary software has made the manipulation of photographs easier to carry out and harder to uncover than ever before, but technology also enables new methods of detecting doctored images. Aiming at restoring the wanted trust on digital images, Digital Image Forensics (DIF) has emerged as a sub-field of Digital Image Processing, taking advantage of the tradition and the pioneering research of the associated scientific societies. DIF has actually introduced and established the aforementioned categories of content-based media authentication (Farid, 2009): a) camera-based techniques that exploit artifacts, introduced by the camera lens, sensor, or on-chip post-processing (Johnson and Farid, 2007; Ho & Li, 2015); b) format-based techniques that leverage the statistical correlations, introduced by a specific lossy compression scheme (Farid, 2008; Luo, Qu, Huang, and Qiu, 2007; Ye, Sun & Chang, 2007); c) pixel-based techniques that search for related statistical anomalies at the pixel level (Mahdian & Saic, 2008; Zampoglou et al., 2015); d) model-based techniques that search

for specific tampering artifacts, detecting anomalies in the three-dimensional interaction between physical objects, light, and the camera (Johnson & Farid, 2007) e) geometric techniques that make measurements of objects in the world and their positions relative to the camera (Farid, 2009).

Nowadays, there is a rapidly growing number of audio and video files available on the Web, either on related social media (i.e. YouTube, Vimeo, SoundCloud, etc.) or through Web-sites of broadcasters, news and media organizations (Dimoulas et al., 2014, 2015; Dimoulas & Symeonidis, 2015; Ioannidou, Apostolidis, Collyda & Mezaris, 2015). While research on audio forgery detection has also gained significant attention, studies in this field are rather limited (compared to those in image and video). The aforementioned categories also apply in this case, with feature-based audio fingerprinting to be the most common single-side authentication approach; prior information can also be exploited (in the form of invisible metadata) in double-sided watermarking techniques (Dimoulas & Symeonidis, 2015; Gupta et al., 2012; Ho & Li, 2015; Hua et al. 2016). Likewise, video forgery detection can engage both active (double-sided) and passive (single-sided) approaches (Sowmya & Chennamma, 2015). Watermarking and signature-related techniques fall into the former category, while context- and content-based evaluation is engaged in the latter. Video analysis is usually conducted at pixel, block, frame and scene level, searching for all spatial, temporal and spatiotemporal alterations. Spatial (intra-frame) tampering is similar to image-forgery, where foreground/background object attack might occur (removal, addition, modification, etc.); it is usually detected through image/frame statistics (spatial features). Motion-pictures tinkering includes frames (and objects) drop/deletion, copy, insertion, swapping and shuffling, where motion-related temporal and spatiotemporal features can be also exploited (Sowmya & Chennamma, 2015; Ho & Li, 2015; Thakur, 2014; Richao, Gaobo & Ningbo, 2014; Wang, Li & Ma, 2014; Wang, 2009). Video authentication is more difficult than image and audio

counterparts, due to the higher (spatiotemporal) dimension of the video-sequences that is associated to increased data-rates and computational load demands. For similar reasons, video tampering is also more difficult and less effective compared to image forgery (it takes much more processing/time or else alterations could be more easily detected).

MEDIA AUTHENTICATION DEMANDS IN REAL-WORLD PRACTICAL SCENARIOS

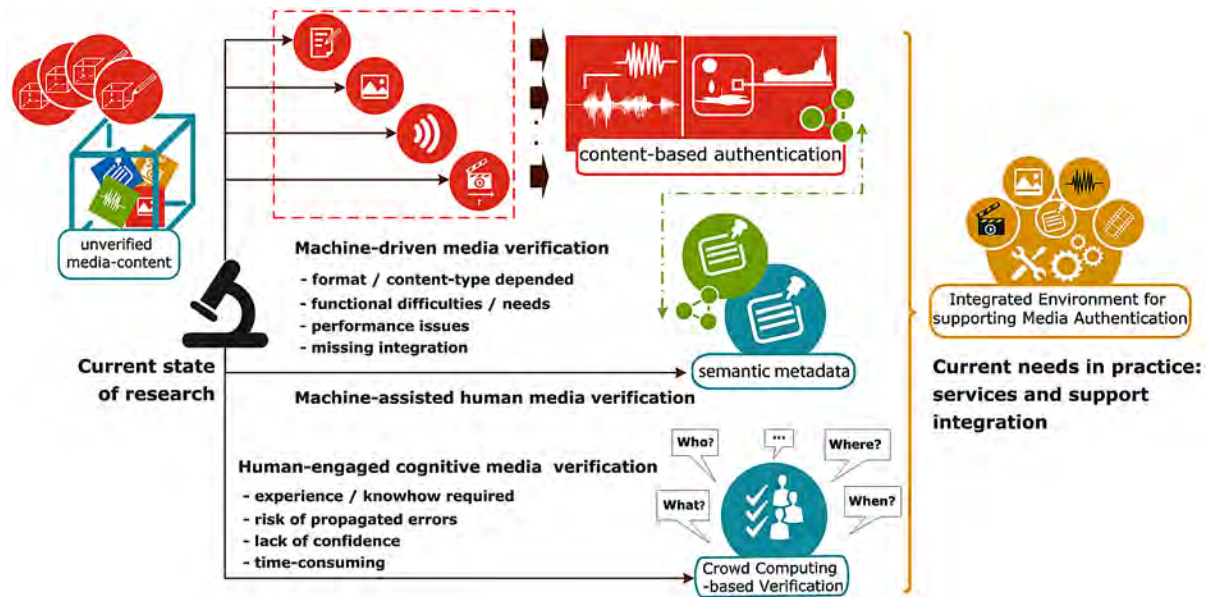
Media Veracity Methods and Tools: Issues, Controversies, Problems

Despite the outstanding progress that has been conducted in research terms, content forgery tools that can be useful to media users and professionals are quite limited. Automated authentication algorithms can detect a huge amount of content when it comes to queries on breaking news events, but arguably only human-beings can sift through (in real time) and make sense of content in an efficient way, (Silverman, 2013). Adopting best practices suggested by the Verification handbook, various services such as the Crowd Crafting¹ and Verily² platforms can be exploited in the direction of context-aware crowd computing approach. Additional, tools have been implemented to find contact details and profiles of users who are active on social media, such as WebMii³, which can be used to identify fake profiles. Findexif.com⁴ and Jeffrey's Exif Viewer⁵ can be utilized in device-fingerprint forgery detection. Both reveal photos' Interchange Information (Exchangeable Image File, known as EXIF) that most JPEG-storing digital cameras use for annotating common capturing parameters (shutter speed, exposure, F number, etc.). Specialized DIF tools are also available through online/free platforms dealing with most image forgery problems/techniques (e.g. Forensically⁶, created by Jonas Wagner; Multimedia Forensics Tool⁷, implemented in the framework of the REVEAL⁸ project). Far less

activity is observed regarding audio and video veracity services. Characteristic examples are the Forensic Audio Analysis System⁹ with the EdiTracker¹⁰ plugin (in the former case) and the ongoing project InVID¹¹ (in the latter). InVID authenticates the reliability and accuracy of news-worthy videos through shot/scene segmentation and concept detection strategies. Hence, despite the availability of the aforementioned applications/services, there is lack of an integrated system (both in term of methods and content modalities) for supporting media authentication from all of its aspects in real world scenarios (as indicated in the right part of Figure 2).

Consequently, media professionals and citizen journalists still rely mostly on traditional methods while verifying news reporting information (and its dissemination). The arising problem derives from the complexity of the algorithms and their outputs that require technical know-how and "digital literacy", in order to be comprehensible and usable by the average user. Indeed, recent research on image authenticity evaluation confirmed the expected difficulties and weaknesses of those services, while it also revealed that most of participants (including active UGC/Web-2.0 users and media professionals) were not even aware of the existence of such tools (Katsaounidou, 2016). It was also proven that authentication recognition accuracy increases when involved users are properly instructed and truly exploit related algorithms (proving their usefulness), especially after being familiarized and comprehend the provided indications. Nevertheless, while most of the forgeries are generally detected by at least one method, the achieved authentication performance remains at low level when algorithms are used separately, increasing also the appearance of false positive results (cases where genuine images are considered to be fake /tampered). Hence, considering the harmful consequences of falsified news and media content propagation, there is a profound need for such advantageous techniques to be combined into an integrated framework.

Figure 2. Available media forgery methods /modalities and the need for integration



Solutions and Recommendations: Integrating Media Authentication Support

The preceding analysis advocates that state-of-research content authentication algorithms have to be deployed in real-world practical scenarios, accommodating various users' and media application needs. All media experts, professional journalists and ordinary users should be trained in the "fresh technologies" shaping this new field. Indeed, the implementation of such a learning framework should put great emphasis on receiving the wanted feedback, with ultimate goal the collaborative development, upgrading, integration and maintenance of unified content verification services through user-friendly (online) environments. For instance, unification, in terms of formalized visualization and representation of the extracted results, is really necessary for the average user, who has difficulties in comprehending the scientific and operational characteristics of the algorithms, therefore in interpreting the provided indications. In this way, more standardized corrective remarks can be obtained through users' interaction as useful feedback towards algorithmic and interfacing

improvements. Furthermore, integration is also desirable for the various content types of today's multimodal posts and shares, where multiple tools need to be applied for the different modalities in a collaborative fashion. Also, it would be really useful to be able to track the accuracy of all the available methods/modalities with respect to the various thematic topics and content categories. In this context, more accurate decision making would be possible by selecting the method expecting to have optimum performance in each specific case, or by properly combining/weighting multiple algorithms via modular-ensemble systems.

It has been observed that manipulated content overwhelms Internet and often reappears on the agenda at different time-instances, drawing users' attention and having impact on public opinion. For this reason, the construction of an online repository would be quite useful in the direction of documenting known forgery cases, helping users to be aware of the already detected falsified information. These data-sets could be also exploited in best-practice adoption and training-scenarios, aiming at conveying the current know-how possessed by experts to regular users. As presented in Figure 3, average users, media professionals and forgery

Figure 3. The proposed integration for supporting media authentication



specialists are being part of the proposed Media Authentication Network, purposing to acquire, exchange and disseminate knowledge on content authenticity and media veracity strategies. The recommended model implicates continuous analysis, classification and documentation of the recorded samples, fed by the associated submissions, queries and comments on both authentic and manipulated content, deriving from all the involved users' feedback. Additional classification metadata can be inserted in terms of subject/topic, creators' and media organizations identity, location (geo-tag) and time information, re-share's methods, the associated falsifying occurrence ratings and others. Hence, a gradually constructed ground-truth would be available for future reference and use, where, besides context-aware authentication, algorithmic and hybrid approaches could be also supported through training and feedback processes.

FUTURE RESEARCH DIRECTIONS

On the road to Web 3.0 and beyond, the ultimate goal is to be able to deploy semantic processes for recognizing and semantically conceptualizing digital content at various levels. Nowadays many

related services are continuously launched and progressed, which apparently would have a positive impact on the efficiency of future media veracity tools. For instance, this can be served through content recognition and concept understanding, followed by the selection of category-optimal forgery detection algorithms, as previously explained. Also, context inconsistencies between the extracted semantics and the associated results of the various authentication methods and algorithms (or between the content artifacts of the various modalities) are expected to be very helpful. For instance, language and emotion classification outcomes can be evaluated in term of consistency with the surrounding environment and context of the events (Kotsakis et al., 2014, 2015). Similarly, content recognition can be exploited for matching and aligning different streams representing the same multimedia event, favoring semantic metadata propagation of matched media, from one to another (Dimoulas & Symeonidis, 2015). Hence, media veracity metadata can be initially extracted from documents that can be more easily and accurately analyzed in terms of authenticity evaluation and then propagated to all the linked data. Progress on related fields like human biometric analysis and contemporary time-, location- and context-aware

ubiquitous technologies could further expedite authentication support, through the extraction and enhancement of user- and situation-related metadata (i.e. who, where, what, when).

From a different perspective, upcoming semantic recognition and conceptualization services are mostly based on the use of machine learning algorithms, heavily relying on the availability of appropriate ground-truth data-sets. Hence, dedicated labeling could be extracted by the proposed model /repository that could serve in assisting the implementation /training of general-purpose semantic services. Considering that massive content is associated to social media posts and shares, it is obvious that enormous amounts of annotations can be also obtained in various levels, through direct and indirect tasks (i.e. registering users' selections on specific browsing /interacting scenarios) (Dimoulas et al., 2015). Hence, relatively easily obtained ground-truth could be further exploited in other digital content forgery applications (i.e. law and justice affairs, sports, politics, sciences, show-biz, history and culture, health and medicine, and others). On the other hand, as technology progresses, content tampering becomes easier and more effective, while anti-forensic techniques also evolve, aiming at preventing or hardening proper forensic investigation. Similarly, the upcoming trend of Robot Journalism (i.e. gathering and publishing news/data without human intervention) can be turned into a subject of information tampering (both unintentionally and deliberately), making even more crucial the need for proper users' informing/training. Thus, a continuous effort is needed in order to be able to keep up with the pace of the new developments, therefore to efficiently deploy media authentication, which can be also considered as a proof of concept for the proposed model.

CONCLUSION

The present chapter examines a variety of content authentication methods that are intended for

helping average users and media professionals in securing themselves by fake information and preventing the spread of misinformation. Undoubtedly, there has been an exponential progress in content veracity strategies, by implementing dedicated fully-/semi-automated authenticity evaluation techniques. However, this know-how has been mainly developed at research level, without providing related straightforward tools to the users involved in media production and consumption services, in various real-world scenarios. Hence, there is an urgent need for further supporting media authentication by exploiting all the available methods in properly integrated computerized environments. The implementation of a Media Authentication Network is proposed, where on-demand training (and feedback) on the new technologies is considered of major importance, enabling users to collaborate with media and forgery experts towards adoption, refinement and widespread dissemination of best practices. These synergies would also propel better comprehension of the involved tools and algorithms, in order to be fully exploited in practice, gaining valuable feedback for further improvements. Thus, a continuously updated online repository, containing documented examples, learning resources and media veracity computing services, could be adaptively accommodated in order to better support the specific needs of the different users, in various application scenarios.

REFERENCES

Angelov, P., Sadeghi-Tehran, P., & Ramezani, R. (2011). An approach to automatic real-time novelty detection, object identification, and tracking in video streams based on recursive density estimation and evolving Takagi–Sugeno fuzzy systems. *International Journal of Intelligent Systems*, 26(3), 189–205. doi:10.1002/int.20462

- Boididou, C., Andreadou, K., Papadopoulos, S., Dang-Nguyen, D. T., Boato, G., Riegler, M., & Kompatsiaris, Y. (2015, September). Verifying multimedia use at mediaeval 2015. *Proceedings of the MediaEval 2015 Multimedia Benchmark Workshop*.
- Boididou, C., Papadopoulos, S., Kompatsiaris, Y., Schifferes, S., & Newman, N. (2014, April). Challenges of computational verification in social multimedia. In *Proceedings of the companion publication of the 23rd international conference on World wide web companion* (pp. 743-748). International World Wide Web Conferences Steering Committee. doi:10.1145/2567948.2579323
- Dimoulas, C., Veglis, A., & Kalliris, G. (2014). Application of Mobile Cloud-Based Technologies in News Reporting: Current Trends and Future Perspectives. In J. Rodrigues, K. Lin, & J. Lloret (Eds.), *Mobile Networks and Cloud Computing Convergence for Progressive Services and Applications* (pp. 320-343). IGI. doi:10.4018/978-1-4666-4781-7.ch017
- Dimoulas, C., Veglis, A., & Kalliris, G. (2015). Audiovisual Hypermedia in the Semantic Web. In M. Khosrow-Pour (Ed.), *Encyclopedia of Information Science and Technology* (3rd ed.; pp. 7594-7604). IGI. doi:10.4018/978-1-4666-5888-2.ch748
- Dimoulas, C. A., & Symeonidis, A. L. (2015). Syncing shared multimedia through audiovisual bimodal segmentation. *MultiMedia, IEEE*, 22(3), 26-42. doi:10.1109/MMUL.2015.33
- Farid, H. (2009). Image forgery detection. *IEEE Signal Processing Magazine*, 26(2), 16-25. doi:10.1109/MSP.2008.931079
- Gupta, S., Cho, S., & Kuo, C. C. J. (2012). Current developments and future trends in audio authentication. *MultiMedia, IEEE*, 19(1), 50-59. doi:10.1109/MMUL.2011.74
- Ho, A. T., & Li, S. (Eds.). (2015). *Handbook of Digital Forensics of Multimedia Data and Devices*. John Wiley & Sons. doi:10.1002/9781118705773
- Hua, G., Zhang, Y., Goh, J., & Thing, V. (2016). Audio Authentication by Exploring the Absolute-Error-Map of ENF Signals. *IEEE Transactions on Information Forensics and Security*, 11(5), 1003-1016. doi:10.1109/TIFS.2016.2516824
- Ioannidou, A., Apostolidis, E., Collyda, C., & Mezaris, V. (2015). A web-based tool for fast instance-level labeling of videos and the creation of spatiotemporal media fragments. *Multimedia Tools and Applications*, 1-40.
- Jin, Z., Cao, J., Zhang, Y., & Zhang, Y. (2015). *MCG-ICT at MediaEval 2015: Verifying Multimedia Use with a Two-Level Classification Model*. Academic Press.
- Johnson, M. K., & Farid, H. (2007, January). Exposing digital forgeries through specular highlights on the eye. In *Information Hiding* (pp. 311-325). Springer Berlin Heidelberg. doi:10.1007/978-3-540-77370-2_21
- Katsaounidou, A. (2016). *Content authenticity issues: detection (and validation) techniques of untruthful news stories from humans and machine*. (Unpublished Master Thesis). Post-Graduate Program of the School of Journalism and Mass Communications, Aristotle university of Thessaloniki.
- Kompatsiaris, Y., Merialdo, B., & Lian, S. (Eds.). (2012). *TV Content Analysis: Techniques and Applications*. CRC Press.
- Kotsakis, R., Dimoulas, C., Kalliris, G., & Veglis, A. (2014). Emotional Prediction and Content Profile Estimation in Evaluating Audiovisual Mediated Communication. *International Journal of Monitoring and Surveillance Technologies Research*, 2(4), 62-80. doi:10.4018/IJMSTR.2014100104
- Kotsakis, R., Mislow, A., Kalliris, G., & Matsiola, M. (2015). Feature-Based Language Discrimination in Radio Productions via Artificial Neural Training. In *Proceedings of the 10th Audio Mostly*. ACM. doi:10.1145/2814895.2814928

- Krawetz, N. (2007). A pictures worth digital image analysis and forensics. *Black Hat Briefings*, 1-31.
- Ljung, A., & Wahlforss, E. (2008). *People, Profiles and Trust: On Interpersonal Trust in Web-Mediated Social Spaces*. Lulu.com.
- Luo, W., Qu, Z., Huang, J., & Qiu, G. (2007, April). A novel method for detecting cropped and recompressed image block. In *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on* (Vol. 2, pp. II-217). IEEE. doi:10.1109/ICASSP.2007.366211
- Mahdian, B., & Saic, S. (2008). Blind authentication using periodic properties of interpolation. *IEEE Transactions on Information Forensics and Security*, 3(3), 529–538. doi:10.1109/TIFS.2004.924603
- Matsiola, M., Dimoulas, C., Kalliris, G., & Veglis, A. (2015). Augmenting user interaction experience through embedded multimodal media agents in social networks. In *Social Media and the Transformation of Interaction in Society*. IGI Global. doi:10.4018/978-1-4666-8556-7.ch010
- Mendoza, M., Poblete, B., & Castillo, C. (2010, July). Twitter Under Crisis: Can we trust what we RT? In *Proceedings of the first workshop on social media analytics* (pp. 71-79). ACM. doi:10.1145/1964858.1964869
- Richao, C., Gaobo, Y., & Ningbo, Z. (2014). Detection of object-based manipulation by the statistical features of object contour. *Forensic Science International*, 236, 164–169. doi:10.1016/j.forsciint.2013.12.022 PMID:24529789
- Sidiropoulos, P., Mezaris, V., Kompatsiaris, I., Meinedo, H., Bugalho, M., & Trancoso, I. (2011). Temporal video segmentation to scenes using high-level audiovisual features. *IEEE Transactions on Circuits and Systems for Video Technology*, 21(8), 1163–1177. doi:10.1109/TCSVT.2011.2138830
- Silverman, C. (Ed.). (2013). *Verification handbook*. European Journalism Centre.
- Sowmya, K. N., & Chennamma, H. R. (2015). A Survey On Video Forgery Detection. *International Journal of Computer Engineering and Applications*, 9(2), 17–27.
- Stamm, M. C., Wu, M., & Liu, K. R. (2013). Information forensics: An overview of the first decade. *Access, IEEE*, 1, 167–200. doi:10.1109/ACCESS.2013.2260814
- Thakur, M. K. (2014). *Tampered videos: detection and quality assessment* (Unpublished Doctoral dissertation). Jaypee Institute of Information Technology, Department of Computer Science Engineering & Information Technology. Retrieved from <http://shodhganga.inflibnet.ac.in/handle/10603/44603>
- Wang, Q., Li, Z., Zhang, Z., & Ma, Q. (2014). Video Inter-frame Forgery Identification Based on Optical Flow Consistency. *Sensors & Transducers*, 166(3), 229.
- Wang, W. (2009). *Digital video forensics* (Ph.D. dissertation). Department of Computer Science, Dartmouth College, Hanover, NH.
- Zampoglou, M., Papadopoulos, S., & Kompatsiaris, Y. (2015, June). Detecting image splicing in the wild (WEB). In *Multimedia & Expo Workshops (ICMEW), 2015 IEEE International Conference on* (pp. 1-6). IEEE.
- Zubiaga, A., & Ji, H. (2014). Tweet, but verify: Epistemic study of information verification on Twitter. *Social Network Analysis and Mining*, 4(1), 1–12. doi:10.1007/s13278-014-0163-y

ADDITIONAL READING

- Benevenuto, F., Magno, G., Rodrigues, T., & Almeida, V. (2010, July). Detecting spammers on twitter. In *Collaboration, electronic messaging, anti-abuse and spam conference* (Vol. 6, p. 12). CEAS.

- Cheong, F., & Cheong, C. (2011). Social Media Data Mining: A Social Network Analysis of Tweets During The 2010-2011 Australian Floods. *PACIS*, 11, 46–46.
- Cozzolino, D., Poggi, G., & Verdoliva, L. (2015). Efficient Dense-Field Copy–Move Forgery Detection. *Information Forensics and Security. IEEE Transactions on*, 10(11), 2284–2297.
- Farid, H. (2009). Photo fakery and forensics. *Advances in Computers*, 77, 1–55. doi:10.1016/S0065-2458(09)01201-7
- Golbeck, J. (2009). Trust and nuanced profile similarity in online social networks. [TWEB]. *ACM Transactions on the Web*, 3(4), 12. doi:10.1145/1594173.1594174
- Hermida, A., & Thurman, N. (2008). A clash of cultures: The integration of user-generated content within professional journalistic frameworks at British newspaper websites. *Journalism practice*, 2(3), 343–356.
- Johnson, M. K., & Farid, H. (2005, August). Exposing digital forgeries by detecting inconsistencies in lighting. In *Proceedings of the 7th workshop on Multimedia and security* (pp. 1–10). ACM. doi:10.1145/1073170.1073171
- Kobayashi, M., Okabe, T., & Sato, Y. (2010). Detecting forgery from static-scene video based on inconsistency in noise level functions. *IEEE Transactions on Information Forensics and Security*, 5(4), 883–892. doi:10.1109/TIFS.2010.2074194
- Palfy, J., Darjaa, S., & Pospichal, J. (2013, December). Dysfluent speech detection by image forensics techniques. In *Automatic Speech Recognition and Understanding (ASRU), 2013 IEEE Workshop on* (pp. 96–101). IEEE. doi:10.1109/ASRU.2013.6707712
- Popescu, A. C., & Farid, H. (2005, January). Statistical tools for digital forensics. In *Information Hiding* (pp. 128–147). Springer Berlin Heidelberg.
- Redi, J. A., Taktak, W., & Dugelay, J. L. (2011). Digital image forensics: A booklet for beginners. *Multimedia Tools and Applications*, 51(1), 133–162. doi:10.1007/s11042-010-0620-1
- Ren, Y., Fan, M., Ye, D., Yang, J., & Wang, L. (2015). Detection of double MP3 compression Based on Difference of Calibration Histogram. *Multimedia Tools and Applications*, 1–16.
- Seo, E., Mohapatra, P., & Abdelzaher, T. (2012, May). Identifying rumors and their sources in social networks. In *SPIE Defense, Security, and Sensing* (pp. 83891I–83891I). International Society for Optics and Photonics. doi:10.1117/12.919823
- Siapera, E., & Veglis, A. (Eds.). (2012). *The handbook of global online journalism*. John Wiley & Sons. doi:10.1002/9781118313978
- Tzelepis, C., Galanopoulos, D., Mezaris, V., & Patras, I. (2015). (in press). Learning to detect video events from zero or very few video examples. *Image and Vision Computing*. doi:10.1016/j.imavis.2015.09.005
- Zubiaga, A., Ji, H., & Knight, K. (2013) Curating and contextualizing twitter stories to assist with social newsgathering. In: *Proceedings of IUI 2013*, pp 213–224 doi:10.1145/2449396.2449424

KEY TERMS AND DEFINITIONS

Anti-Forensics: Anti-forensic techniques are content actions aiming at preventing or hardening proper forensic investigation.

Biometrics: Biometric systems measure and analyze various persons’ psycho-physiological parameters (fingerprint, face-expressions, stress-factors, etc.) aiming at evaluating their behavioral status in various cognitive task and/or surveillance scenarios.

Computer Forensics Science: A research field focusing on content examination with the aim of identifying, preserving and recovering the truth of digital media.

Digital Forensics: A branch of forensic science encompassing the recovery and investigation of material found in digital devices, often in relation to computer crime.

Digital Image Forensics (DIF): This field emerged as a sub-field of Digital Image Processing (DIP), aiming at providing tools for images tampering investigation.

Machine Learning (ML): Scientific discipline that investigates algorithms and methods aiming at giving machines the ability to learn from experience (without being explicitly programmed), in order to respond autonomously on specific tasks and automate various data-handling processes.

Multimodal Media Assets (MMA): The kind of multimedia content commonly encountered in today's posts and shares, incorporating mixed time-based and page-based media (text, images, audio, video, etc.) along with their metadata.

Natural Language Processing (NLP): scientific discipline utilizing Machine Learning and Computational Linguistics, aiming at giving

computers/machines the ability to perceptually interact through human (natural) languages.

User Generated Content (UGC): Any form of content created and shared by users of an online system or service.

E

ENDNOTES

- ¹ <http://crowdcrafting.org/>
- ² <https://veri.ly/>
- ³ <http://webmii.com/>
- ⁴ <http://www.findexif.com/>
- ⁵ <http://regex.info/exif.cgi>
- ⁶ <https://29a.ch/photo-forensics/>
- ⁷ <http://reveal-mklab.iti.gr/reveal/verify.html>
- ⁸ <http://revealproject.eu/>
- ⁹ <http://www.forensicav.ro/software.htm>
- ¹⁰ http://speechpro-usa.com/product/forensic_analysis/editracker
- ¹¹ <http://www.invid-project.eu/>

IT Service Management Architectures

Torben Tambo

Aarhus University, Denmark

Jacob Filtenborg

Aarhus University, Denmark

INTRODUCTION

In a world of ever increasing importance of digital business models with strong elements of outsourcing, heterogeneous IT landscapes, and cloud services, IT service management (ITSM) must be founded on strong approaches of governance (Pröhl & Zarnekow, 2015; Iden & Eikebrokk, 2013; Ali et al., 2013;). Digital business models mean that companies potentially both can act as providers and users of IT services (Richardson & Mahfouz, 2014), and companies change from being suppliers or distributors of traditional physical product to provide or rely on a mesh of digital services. ITSM architectures cover the system architecture, information architecture, and enterprise architecture enabling IT services and IT services deliverables (Braun & Winter, 2007; Eder & Nag, 2001). Important for these architectures are the alignment with fundamental operational agreements, mainly Service Level Agreements (SLA) and operational frameworks such as Information Technology Information Library (ITIL) (Holland, 2015), COBIT (Saran, 2015), Microsoft MOF, HP ITSM and IBM ITPM (Iden & Eidebrokk, 2013). ITSM architectures most generally reflect the services stipulated by ITIL (Gama et al., 2013) although an extension of ITIL is seen with the introduction of the Service Integration and Management (SIAM) framework (Armes et al., 2015).

As many ITSM architectures are relatively well-defined on single management dimensions of ITIL, there is a general lack of architectural under-

standing across the different service management dimensions. Complex services crossing several ITIL management layers and other added services suffer from lack of a single ITSM architecture with this resulting in customized and hard-coded solutions for monitoring, reporting and enabling of transparency. Data collection for e.g. incident management might remain loosely connected to management systems for continuity, capacity or security management. Configuration Management DataBases (CMDB) and Service Catalogues are widely expected to connect services and management systems, but are often incomplete, inconsistent, under-managed, over-managed, overaged or lagging behind with precision.

The claim of this article is that the modern digital business must connect all its silos of service management systems in well planned architectural systems potentially also connected to the general enterprise architecture of the company.

This article is based on a case study in a major IT services provider, where it was realized that service management had grown apart from the enterprise architecture in terms of highly customized solutions and services not using sufficient scale of magnitude and uniformity of processes, services and reporting. A services architecture was made from available but disjoint sources in order to create transparency and efficiency in the services production. The services architecture is important to realize digital business models whether being producer of IT services as core business or being a digitally enabled business.

BACKGROUND

Most companies and organizations are increasingly dependent on information technology (IT) and IT services. As technology is becoming more distant and commoditized in the form of cloud services and fragmented between highly specialized vendors, management of services is getting comparatively more critical (Amanatullah et al., 2013). IT Service Management (ITSM) is thus the adaptation of the right services to the right business.

Services Fundamentals

The fundamental concept of creation of business value of IT is normally expressed as services (Lahtela et al., 2014). Services are building blocks of service systems. Services are inclined to interact with work as work systems (Alter, 2014a; Alter, 2014b) with a need to recognize ‘service’ as a highly multi-faceted term with numerous contextually-linked meanings. Moreover, services are defined by constituting elements of technology, practices, and skills and operable elements of initiation, execution and termination. Frameworks such as ITIL (Holland, 2015; Randone, 2012; Kaschanki & Toland, 2006) define services from levels of criticality and impact for business, especially drawing up the concept of Service Level Agreement (SLA) that over time has transcended from IT into business in general.

Quality of service (QoS), and henceforth fulfillment of business expectations, is fundamentally defined as the ability to meet SLA KPI’s (Dubois, 2014). QoS is thus not a subjective judgement but an objective, agreed, defined fact normally also being measurable. This leaves the subjective issue open. Popularly, there can be disputes on services as they are subjectively perceived “bad”, but do meet SLA. Garschhammer et al. (2001) define QoS as independent of providers and clients, but at the same level as service access, service functionality and service management. QoS consists the several elements of service expectations stipulated in

the SLA, e.g. availability, scalability – capacity management, ‘detect and correct’, and the two ‘front ends’ of ITIL: Incident management and problem management (Soomro & Hesson, 2012; Franke et al., 2014).

Service design is the activity of composing the right set of services from available or new services based on business customs, technologies, people, skills and context. Garschhammer et al. (2001) discuss service design based on telecom services and describe the service design process as based on a service life-cycle analysis, mapping interactions, describing functional attributes, defining roles, object and relations, and ending with a service model.

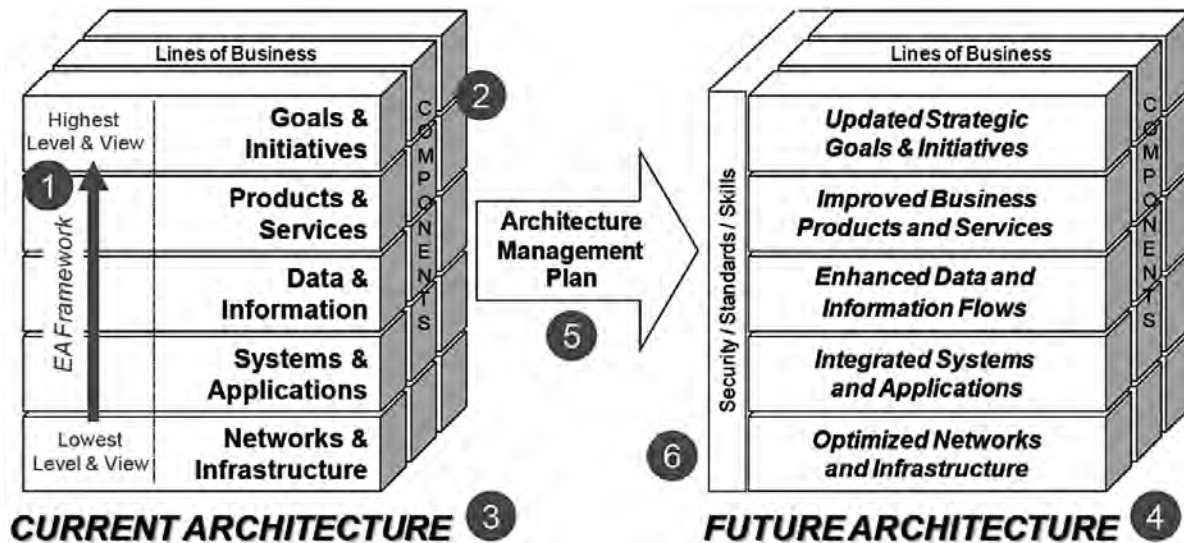
In designing services, appropriate design for the optimal solution of a given requirement is normally the point of departure. It is however much more critical to ensure reusability of services through harmonization, standardization, adaptation, scalability and security. Wullenweber et al. (2008) discuss standardization of services in relation to potential outsourcing thereby underlining the importance of services being movable across organizational barriers.

Infrastructure and Enterprise Architecture Fundamentals

Infrastructures have traditionally seen as the supporting structures for IT services consisting of technology and assets (Ali et al., 2013; Närman et al., 2014). Infrastructures are in such a perspective company internal, also considering Alter’s Work System Model (Alter, 2014a). Increasingly benefits have been realized from outsourcing of parts and pieces of infrastructure. Especially application management, server-services, security services, and more labor intensive services of incident management and call centers.

Cloud architectures and service virtualization mean that services are provided by a shared infrastructure. Technical as well as human efforts in services are thus to be considered as relative mobile which in turn can be discussed as fragmentation

Figure 1. The EA3 cube of enterprise architecture methodology
Bernard, 2012.



of services, distance between the place of service delivery and service consumption.

Enterprise architecture is focused on documenting and managing change in the relationship between business strategy, organization and technology (Zachman, 1996; Bernard, 2012). Enterprise architecture connects technology from its lowest level and upwards towards products and services as described in Figure 1.

TOGAF is an industry leading framework for prescribing change processes in structured and documented phases (Vicente et al., 2013a) using a cyclical approach to ensure continuous update and improvement of the enterprise architecture documentation. Archimate is a modelling language for describing processes and artefacts associated with enterprise architecture modelling and management processes (Band et al., 2015; Vicente et al., 2013b).

Services, Architectures, and State of the Art

The literature and cases on applying enterprise architecture on IT service provider organizations is limited. Braun & Winter (2007) discuss enterprise architecture and services integration but with the main aim of extending enterprise architecture with

service of service-oriented architecture (SOA), i.e. this is more a technical design than a strategic supplier perspective.

Conceptual integration of enterprise architecture and ITIL is suggested by Gama et al. (2014) to ensure simplification of IT governance processes and avoid risks of ambiguity in providing services (Jäntti & Hotti, 2015).

ITIL has largely been regarded as state of the art for production of IT services. Especially in its full and idealized implementation, ITIL provide a heavily detailed and extensive framework on the operational foundations and best practices within IT services. It is worth noticing that ITIL is a proposition for IT services and normally needs local adaption to the service context needed. In that sense ITIL address the key operational requirements, but it also address transitions with due reflections on both infrastructure and service transition risk management, transparency, mitigation of problems, and stabilization.

MAIN FOCUS

According to the enterprise architecture model above, the perspective of this article is that IT beneficially can be regarded as a line-of-business

by its own in any aggregate of enterprise and thereby IT logically must have an enterprise architecture of its own. Key elements of this is the commercial, organizational, HR and technological transformation from the fragment/value-provider perspective of IT service providers (internal or external) and to reviewing own capabilities and competencies with the methodologies normally reserved for analysis and solutions to clients.

Issues, Controversies, Problems

Of main caveats for IT transformation are appropriate self-perceptions of precision and specificity of its services and added value to own business models as well as clients' business models. This is complicated by IT landscapes as generally being predominantly heterogeneous (Heininger et al., 2013) with long development time and long phase out time of given services and technologies. A closer relation between service catalogues and enterprise architecture is worthwhile investigating.

Combinations of requirements from fragmented clients, in case of services providers, or extent and differentiation of value chains, in case of internal providers, create complexity given variations in infrastructure, applications and financial restrictions. Aggregate services are thus more complex – and the solution is probably not to divest service fragments to yet new providers.

Seen from provider perspective, services clients are likely to be less and less loyal, aim for perceived commoditized services, and accept multi-vendor service landscapes with having full insight into consequences for complexity and necessary capabilities for governance. Here, ITIL will come to an end in management of complexity, and SIAM-like approaches must be considered.

So, the aim is to investigate the formation of architectures positively supporting ITSM in multi-client, multi-vendor landscapes. Below a case study is presented using a transformational process of a service provider seeking to manage complexity and transformation.

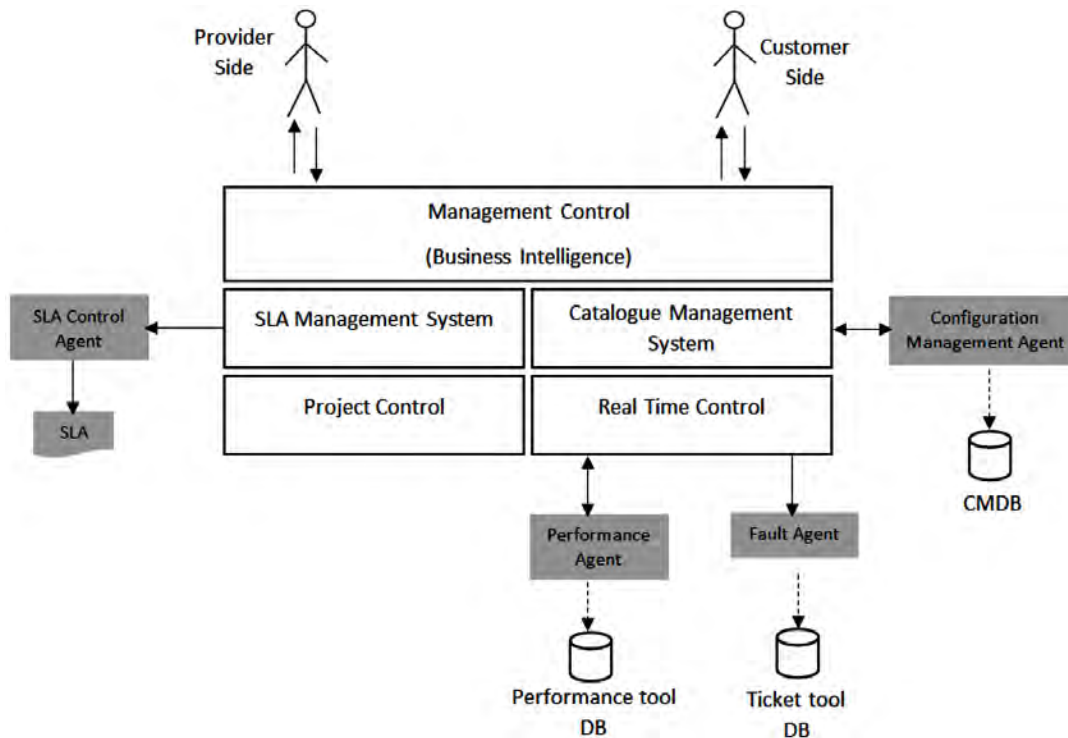
Case Study

A larger IT services provider, anonymized as ABIT, had been transitioning from a few clients to engage with several larger clients across Europe. ABIT had a generally good reputation in the market mainly due to highly customized service contracts, adaptive approach to customers momentary need, and even more customized report services. Services included most of the ITIL palette of services and associated systems operations and management activities. ABIT could also deliver software development services, but this is more a secondary activity. The highly customized service contracts had impacts in the organization in form of:

1. Complex reporting of SLAs to clients that was customized and labor intensive – more than 3500 KPIs were identified.
2. There was not a direct translation from ITSM, systems management software and infrastructural monitoring to a reporting system.
3. Manually processed SLA reports were delivered to clients retrospectively with no online reporting to clients.
4. Reporting data were ambiguous and frequently disputed both internally in ABIT and between clients and ABIT.

To overcome the issues of IT services monitoring and reporting, ABIT started to review the options for design of an enterprise architecture for its IT services. Thus ABIT's approach had typically been to review its services as an external extension of the clients' enterprise architecture, e.g. applications and services, information and data. Figure 2 is describing the architectural draft with the rationale that Service Catalogue-based SLAs drive business processes either in terms of ongoing and agreed services, but also transformational projects and projects related to problem management.

Figure 2. Relation between ITSM architectures of services and change



The realized issue was to review itself as a producing company that could benefit from an enterprise architecture approach. According to the principles of Ross et al. (2006) it was decided as critical to obtain a unified enterprise architectural strategy for the mainline of internal operations although with option to adapt to a diversified enterprise architecture in various critical client driven areas.

As the immediate objective of the company was to develop a unified enterprise architecture for reporting on services, it was necessary to define a relevant starting point. As an IT services provider, ABIT sells server, storage and network capacity along with management and support services for standard and bespoke software systems. This defines key artifacts of the enterprise architecture. The services are operated via a number of management systems. Operations management systems (ITOM) monitor infrastructure. ITSM systems register mainly incidents and problem management. The individual customer service

contracts, SLA KPI's, and the available service catalog form a level of business objectives that are to be executed through ITOM and ITSM systems. Portfolio management, long-term planning and transitional projects are managed in non-standardized systems using mostly ad-hoc project models, although transitional initiatives draw on inspiration from both ITIL and TOGAF.

Results

The case led to a number of significant achievements for ABIT. The unified enterprise architecture not only created a more efficient service delivery organization, it also created more precise and well-defined services of the IT Service Catalogue. It was realized that the diversified approach to customers was actually not what the customers preferred; the customers preferred clear links between expected services, the providers ability to produce these services, and the report provided on the services. More specifically, the hours spend

on reporting to each customer was drastically reduced – in a broad scale from 100 to 10 hours per customer either per month or per quarter. This left more resources to qualified services of getting better insight in the customers' expectations and needs. Identified new needs or identified obsolete services could far better be presented to customers combining the IT Service Catalogue and the IT Service Management Enterprise Architecture.

SOLUTIONS AND RECOMMENDATIONS

The tradition among IT service providers to consider themselves as a production apparatus with a distinct enterprise architecture is relatively new and remain yet to permeate industry. It is a question of a redefined mindset: To perceive IT as a business with a business model of its own rather than see IT as a fragment of other business models. Typically, IT has expressed itself in terms of solutions for improvement of other business models, but in this context, IT can improve IT.

From the case study, a generalized approach can be derived. The approach can be described as a transition from the fragmented and client-only-centric to the IT-intrinsic perception of the IT service provider as a fully workable and responsible provider with its own enterprise architecture.

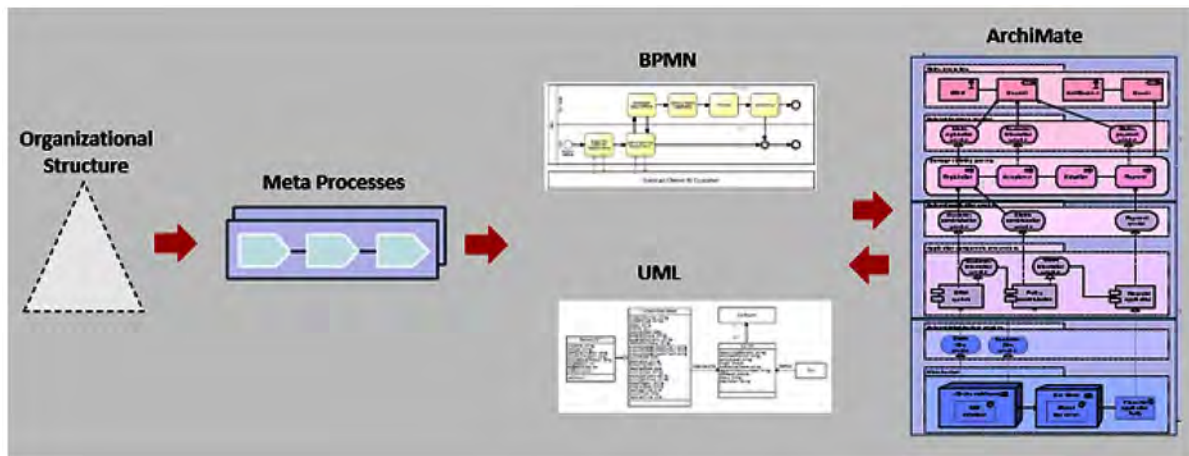
The approach consisted of a set of metaprocesses, illustrated in figure 3. In clear terms these processes entailed the following activities:

1. Define the organizational structure of the client organization, thereafter define the providing organization of ABIT. The overlap between these formed the focal organization of the intended enterprise architecture, with the sum of the providing organization as the target organization for the to-be enterprise architecture.
2. Metaprocesses were defined in terms of
 - a. Service catalogue development and service catalogue services measurability (SLA).
 - b. CMDB artefact (CI – configuration items) development and relations between CI and the service catalogue.
 - c. The organizational responsibility, capability and alignment on governance of metaprocesses.
 - d. The customer service design according to requested services, service catalogue and ABIT's CMDB infrastructure (unified).
 - e. The customer service SLA (diversified).
3. Found services in processes described using BPMN notation.
4. Convert artefacts to a UML representation as an enterprise architecture repository.
5. Combine metaprocesses, the BPMN, and the UML in Archimate to represent the final enterprise architecture.
6. From known SLA KPI's, a relevant reporting and analytics framework can be created combining distributed service repositories.

Table 1.

Strategy	To provide leading and cost-efficient IT services to high-end customers
Business	Appropriately combine skills, processes, technologies, standards and customer collaboration
Information	Any SLA related to ITSM, ITOM or change processes must be represented in the information repository of enterprise. Any process and technology should be reviewed as information provider to the overall management processes
Systems	There should be a direct connection between service catalogues and systems although systems need also
Networks	Largely describing connectivity between place-of-use and place-of-make, thereby supporting cloud, distributed and local service architecture

Figure 3. Relationship between organization, metaprocesses and enterprise architecture formation



The approach is indicated in Figure 3.

Using Bernard's (2012) EA3 model this would be expressed as shown in Table 1.

Importantly is to express the designed enterprise architecture as both documentation and planning. In terms of IT services organizations, it is critical to extend this with reasonable considerations of the interactions, relationships and agreements with the client organizations mainly expressed in terms of SLA and service catalogues.

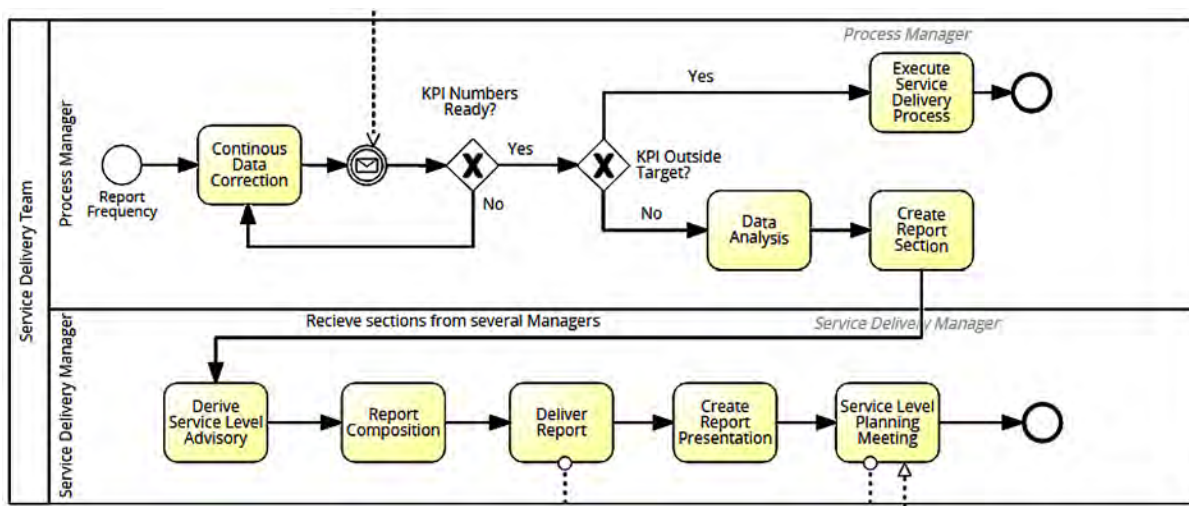
Furthermore, validation of the stated models is by itself fundamental. Below is the model developed from the case study above. The model is

narrowly focused on iteratively ensuring a precise relationship between the organizational entity of the delivered services, the services as defined by the service catalogue, the SLA, the KPI, and reactive processes initiating service execution, data collection, and data reporting.

An excerpt of the developed processes supporting the ITSM enterprise architecture is illustrated in Figure 4.

Figure 4 is highlighting that services are managed as a continuous flow of data underpinning the KPIs of the service catalogue. A number of organizational roles within ABIT were clarified as

Figure 4. Service delivery reporting and adaptation



a part of the process. Especially ‘process managers’ have technical responsibilities, whereas ‘service delivery managers’ have client responsibilities. SLA KPIs are in such a way constructed to continuously confirm the validity of the designed enterprise architecture, or give reason to modify or update the enterprise architecture.

The architecture of ITSM should follow the general architecture of the desired business objectives. In reshaping ITSM through architectural thinking, it is highly a matter of having IT “taking its own medicine”, and reflect specifically on:

- Accept dual EA purpose: One of the service provider, one of the impact of a service to the clients business.
- Exploit IT-takeaways to IT business, e.g. “big data analytics on ITSM and ITOM”, “BPM on IT processes”, “CRM analytics on service quality”.
- Aim for unification rather than diversification.

At the longer term, ABIT intends to adapt to the SIAM (Armes et al., 2015) framework. Here ABIT can get a clearer picture of service constructs, definition of service delivery complexity, and requirements for collaborative service delivery environments. SIAM can overcome the lack of understanding of complexity in ITIL – although ITIL’s definitions and obligations of roles remain precious. SIAM can probably not ensure the architectural uniformity at the inner lines of ABIT. Therefore it is still valuable to consider enterprise architectures for ITSM providers.

FUTURE RESEARCH DIRECTIONS

The case study above is more focused on ITIL than SIAM (Armes et al., 2015). Future research can benefit by including aspects of SIAM in to the architectural design, i.e.

- Multi-vendor services,

- Supplier management including supplier life-cycle management,
- End-to-end service management,
- Advanced portfolios and service catalogs,
- Integrated toolsets,
- Providers insight in clients operations.

A key issue in the presented study is that the focal/client companies are likely not to relate themselves deeply into the ITSM reporting but rather to the providers’ ability to “foresee the unexpected”, e.g. shifts in load, new security threats, patching conflicts. As the outlined architectural guidelines are mostly focused on quantitative, “as-is” reporting, new approaches would be needed to support advanced prediction analytics, early warnings, and automated remediation. Such systems and solutions do exist (SAP, 2016; Oracle, 2016), but are largely product-centric, disconnected from the overall ITSM landscape, and not taking complex, distributed SIAM multi-vendor service environments into account.

CONCLUSION

This article has described the development of a series of approaches in moving from fragmented and ad hoc based IT service management to a unified approach of enterprise architecture for IT service management. The architecture included an information architecture made in UML. A business process architecture made in BPMN and a modelling approach in Archimate. The resulting IT Service Management architecture can serve as a template for the creation of overarching service architectures for governance of operational and monitoring technologies linking technology and services.

This article demonstrates the value of considering IT services as a full and independent business of its own with an appropriate enterprise architecture. This opposes the view traditionally offered by IT service vendors as having managing fragments of clients’ business models without full

inclusion in the clients enterprise architecture, and likewise without specifically to utilize enterprise architectural views in the providers own line of business. The Ross et al. (2006) operating model of Unification is making strongest sense related to standardized business processes of ITIL, ISO2000 and common Operations Frameworks like MOF

A positive learning is that the suggested approach is as applicable to in-house IT operations as it is to IT service providers. In-house IT as well as service providers have to deal with fragmentation of vendors and sprawling service catalogues, whereas the overall service requester must be supported with appropriate means and tools for knowing, what is being paid for and what to expect. The recently introduced IT4IT framework (Josey, 2015) furthermore supports the findings of the necessity of designing IT for IT.

REFERENCES

- Ali, S. M., Soomro, T. R., & Brohi, M. N. (2013). Mapping information technology infrastructure library with other information technology standards and best practices. *Journal of Computer Science*, 9(9), 1190–1196. doi:10.3844/jcssp.2013.1190.1196
- Alter, S. (2014a). *Disentangling service: Using a work system perspective to reconcile different but overlapping portrayals of service and service systems*. University of San Francisco Working Paper.
- Alter, S. (2014b). *A unified operational view of service, service systems, and service science*. INFORMS Annual Meeting, San Francisco.
- Amanatullah, Y., Lim, C. U., Ipung, H. P., & Juliandri, A. (2013). Toward cloud computing reference architecture: Cloud service management perspective. In *ICT for Smart Society (ICISS), 2013 International Conference on* (pp. 1-4). IEEE.
- Armes, D., Engelhart, N., McKenzie, P., & Wiggers, P. (2015). *SIAM: Principles and Practices for Service Integration and Management*. Van Haren.
- Band, I., Engelsman, W., Feltus, B. C., Paredes, S. G., & Diligens, D. (2015). *Modeling Enterprise Risk Management and Security with the ArchiMate®*. Academic Press.
- Bernard, S. (2012). *An Introduction to Enterprise Architecture* (3rd ed.). Bloomington, IN: Authorshouse.
- Braun, C., & Winter, R. (2007). Integration of IT service management into enterprise architecture. In *Proceedings of the 2007 ACM symposium on Applied computing* (pp. 1215-1219). ACM. doi:10.1145/1244002.1244267
- Dubois, E. (2014, June). Information Systems for the Governance of Compliant Service Systems. In *International Conference on Advanced Information Systems Engineering* (pp. 1-11). Springer International Publishing. doi:10.1007/978-3-319-07881-6_1
- Eder, M., & Nag, S. (2001). *Service Management Architectures Issues and Review*. RFC Editor. doi:10.17487/rfc3052
- Gama, N., Sousa, P., & da Silva, M. M. (2013). Integrating enterprise architecture and IT service management. In *Building Sustainable Information Systems* (pp. 153-165). Springer US. doi:10.1007/978-1-4614-7540-8_12
- Garschhammer, M., Hauck, R., Hegering, H. G., Kempter, B., Radisic, I., Rolle, H.,... Nerb, M. (2001). Towards generic service management concepts a service model based approach. In *Integrated Network Management Proceedings, 2001 IEEE/IFIP International Symposium on* (pp. 719-732). IEEE. doi:10.1109/INM.2001.918076
- Heininger, R., Böhm, M., & Krcmar, H. (2013). Managing Heterogeneity in IT Service Management: Towards a Research Agenda. In *SIG Service Science (SIG SVC) pre-ICIS Workshop*.
- Holland, K. (2015). *An introduction to Service Integration and Management and ITIL*. Axelos.

Iden, J., & Eikebrokk, T. R. (2013). Implementing IT Service Management: A systematic literature review. *International Journal of Information Management*, 33(3), 512–523. doi:10.1016/j.ijinfomgt.2013.01.004

Jäntti, M., & Hotti, V. (2015). Defining the relationships between IT service management and IT service governance. *Information Technology Management*, 1–10.

Josey, A. (2015). *The Open Group IT4IT™ Reference Architecture, Version 2.0*. Berkshire, UK: The Open Group.

Lahtela, A., Hotti, V., & Salomaa, H. (2014). Service Support in IT Governance, IT Management and Enterprise Architecture Context. In *The Fourth International Conference on Digital Information Processing and Communications (ICDIPC2014)* (pp. 166–172). The Society of Digital Information and Wireless Communication.

Närman, P., Buschle, M., & Ekstedt, M. (2014). An enterprise architecture framework for multi-attribute information systems analysis. *Software & Systems Modeling*, 13(3), 1085–1116. doi:10.1007/s10270-012-0288-2

Oracle. (2016). *Managing Automated Database Maintenance Tasks*. Retrieved from https://docs.oracle.com/cd/E11882_01/server.112/e25494/tasks.htm#ADMIN0235

Pröhl, T., & Zarnekow, R. (2015). Towards real-time IT service management systems: In-situ analysis of events and incidents using SAPHANA. *Technische Berichte des Hasso-Plattner-Instituts für Softwaresystemtechnik an der Universität Potsdam*, 171.

Randone, C. (2012). Enterprise Architecture, IT Service Management, and Service-Oriented Architecture: Relationships, Approaches, and Operative Guidelines. Part 1. *Journal of Enterprise Architecture*, 8(2), 45–55.

Richardson, B. F., & Mahfouz, A. Y. (2014). Strategic Integration of IT and Business Service Management. *Communications of the IIMA*, 9(2), 7.

SAP. (2016). *SAP Standard for System Monitoring and Alerting*. SAP Solution Manager 7.1. Retrieved from <https://support.sap.com/dam/library/SAP%20Support%20Portal/support-programs-services/methodologies/support-standards/e2e-standard-for-system-monitoring-and-alerting.pdf>

Saran, C. (2015). *The pros and cons of running your IT department as a business*. Computer-weekly.com

Vicente, M., Gama, N., & da Silva, M. M. (2013a). Using archiMate and TOGAF to understand the enterprise architecture and ITIL relationship. In *Advanced Information Systems Engineering Workshops* (pp. 134–145). Springer Berlin Heidelberg. doi:10.1007/978-3-642-38490-5_11

Vicente, M., Gama, N., & da Silva, M. M. (2013b). Using ArchiMate to Represent ITIL Metamodel. In *Business Informatics (CBI), 2013 IEEE 15th Conference on* (pp. 270–275). IEEE. doi:10.1109/CBI.2013.45

Winter, R., & Fischer, R. (2007). Essential Layers, Artifacts, and Dependencies. *Journal of Enterprise Architecture*, 1–12.

Wullenweber, K., Beimborn, D. & Weitzel, T., 2008. The impact of process standardization on business process outsourcing. *Information Systems Frontiers*, 10(2), 211–224.

Zachman, J. (1996). *Concepts for the framework for enterprise architecture*. Los Angeles, CA: Zachman International.

ADDITIONAL READING

Betz, C. T. (2011). *Architecture and patterns for IT service management, resource planning, and governance: Making shoes for the cobbler's children*. Elsevier.

Dibbern, L. (2016) IT Service Management using ITIL® and UML, 2nd edition: A guide to IT service management and ITIL® V3, 2011 Edition. CreateSpace Independent Publishing Platform; 2 edition.

Moeller, R. R. (2013). *Executive's guide to IT governance: improving systems processes with service management, COBIT, and ITIL*. John Wiley & Sons. doi:10.1002/9781118540176

Simon, D., & Schmidt, C. (Eds.). (2015). *Business architecture management: Architecting the business for consistency and alignment*. Springer. doi:10.1007/978-3-319-14571-6

Vicente, M., Gama, N., & da Silva, M. M. (2013). Using ArchiMate to Represent ITIL Metamodel. In *Business Informatics (CBI), 2013 IEEE 15th Conference on* (pp. 270-275). IEEE. doi:10.1109/CBI.2013.45

KEY TERMS AND DEFINITIONS

Architecture: In Information Systems research, a fundamental or aggregated artefact of technological, organizational and processual character.

CMDB: Configuration Management Database, the key repository in ITIL for items to be serviced. The individual items are known as CI, Configuration Items. The CMDB can be supported by specialized or general software. Some items can be retrieved automatically via automated networked IT services information collection, e.g. using the SNMP protocol.

Enterprise Architecture: The business activity of ensuring appropriate relationships between business strategy, and supporting technologies. Enterprise Architecture is mostly connected to

information technology, but is not limited to this. Any relationship between business and technology can be encompassed in enterprise architecture. The activities are generally two-fold: To document the existing portfolio of technology, and to plan technological transitions.

IT Service Catalogue: The service catalogue is a “handbook” describing, what can be delivered of services from any, in-house or external, IT service provider. The service catalogue aims at adjusting expectations between requesters and providers and contributes to control cost of services. Furthermore, providers are “protected” from supplying services, although desired by the client, where no services exist e.g. caused by competencies, technology, logistics.

IT Service Management: ITSM is the discipline of organizing and managing IT services to meet expectations of clients and ensure budgetary objectives. ITSM entails long term planning to safeguard relevant technology and human resources. Furthermore, ITSM is aimed at adapting the IT Service Catalogue to fit the demand from clients and the given financial constraints.

ITIL: Information Technology Information Library. A management framework for IT services and infrastructure management. Originally developed within the UK government, but as of today, a de facto standard globally for IT services provision and governance. ITIL is mostly adapted to local requirements and is rarely found fully and precisely implemented. ITIL is reflected in the ISO/IEC 20000 standard.

SIAM: Service Integration and Management – a governance framework for complex, multi-vendor IT services environments. SIAM is designed to overcome issues with ITIL on responsibilities, contract management, cross-organizational collaboration and heterogeneous environments.

Category E

Enterprise Resource Planning

Business Intelligence Impacts on Design of Enterprise Systems

Saeed Rouhani

University of Tehran, Iran

Dusanka Milorad Lecic

Levi9 Global Sourcing Balkan, Serbia

INTRODUCTION

In this information age, the approach to decision support as an individual system, such as decision-support systems (DSS), has been replaced by a new viewpoint of intelligent software and systems (Mehdi Ghazanfari, Rouhani, Jafari, & Taghavifard, 2009). Based on this new approach, enterprise systems are designed to have Business Intelligence (BI) as an umbrella concept which covers various enabler tools and capabilities in the form of non-functional requirements (Rouhani & Zare Ravasan, 2015b). Most organizations still experience a lack of business intelligence (BI) in their decision-making processes when implementing enterprise systems. The current state-of-the-art in decision support focuses on the intelligence requirements of enterprise systems as important quality aspects, along with other functional and non-functional needs. But the literature lacks studies on the evaluation of these intelligence requirements (Rouhani & Zare Ravasan, 2015a).

In this book chapter, business intelligence and enterprise systems literature are reviewed. Also based on the latest researches, the position of BI on these systems is discussed. In the following, through the study of BI capabilities and proposing them as non-functional, the BI Impacts on the design of enterprise systems and software would be described and the direction for future research and insights for information systems development would be prescribed.

BACKGROUND

In this chapter, it will be given literature review about business intelligence and enterprise systems. The lack of a large number of researchers will limit this work on newer research period.

Business Intelligence

The term Business Intelligence was introduced to describe a set of concepts and methods to improve business decision-making by using fact-based, computerized decision support systems. This term is introduced by Howard Dresner of the Gartner Group (Nylund, 1999). The first definition introduced BI as a management philosophy and tool that help organizations to manage and refine business information to make effective decisions (Gbosbal & Kim, 1986).

Business intelligence presents the process through which organizations take advantage of virtual and digital technology to collect, manage and analyze structural or non-structural data. In other words, the technology and commercial processing procedures in the decision-making are supported through the extraction, integration and analysis of data (Berson & Smith, 1997).

The purpose of business intelligence is to help control the resources and the information flows of the business which exists in and around the organization. BI makes a large contribution to the required intelligence and knowledge of the

organizations' management by identifying and processing data in order to explain their hidden meanings (Azoff & Charlesworth, 2004).

As a total definition, Lönnqvist & Pirttimäki (2006) stated that "Business intelligence as a term can be used when referring to the following concepts":

1. Related information and knowledge of the organization, which describe the business environment, the organization itself, the conditions of market, customers and competitors and economic issues;
2. A system and a systematic process by which organizations obtain, analyze and distribute the information for making decisions about business operations.

Eckerson realized that business intelligence must be in able to provide the production reporting tools, end-user query and reporting tools, on-line analytical processing (OLAP), dashboard/screen tools, data mining tools and planning and modeling tools (Eckerson, 2010).

Bose (2009) believed that the role of BI is preparing the right information to the right people at the right time to improve decision making, hence improve managerial proceeding and enterprise performance. Generally, the main focus in managerial approach is on the process of gathering data from internal and external sources and of analyzing them to produce relevant information to the decision making process (M. Ghazanfari, Jafari, & Rouhani, 2011; Petrini & Pozzebon, 2009). In technological point of view, BI has been proposed as an instrument of analysis, providing automated decision making about business conditions, sales, customer demand, product preference, and so on. It uses a different type of analysis and mathematical, statistical and artificial intelligence, as well as data mining and OLAP to take a decision in more suitable format (M. Ghazanfari, et al., 2011; Petrini & Pozzebon, 2009). Ghazanfari et al. (2011) and Rouhani et al. (2012) claimed that there is

one more approach in BI definition called system enabler in which the main focus is on value added features on supporting information.

Enterprise Systems

Enterprise systems (ES) which also called Enterprise Information Systems, have emerged as a promising tool used for integrating and extending business processes across the boundaries of business functions at both intra-organizational and inter-organizational levels in the past decade. ES provides an IT platform that enables industrial organizations to integrate and coordinate their business processes; it is considered a revolutionary advance in the continuous evolution of computer applications in business and industry (Xu, 2011). Enterprise systems as a commercial software packages, promise the integration of all the information flowing through a company, from customer information, financial and accounting information to human resource information and so on.

In the 1990s, enterprise systems have been mainly used for managing the physical assets of an enterprise but today it is known that the knowledge is a compilation of an enterprise's invisible assets (Xu, 2011). Increasing requirements for extended enterprises have also stimulated the integration of the knowledge management function into enterprise system for knowledge asset management (Xu, 2011). The concept of integration enterprise system and knowledge management becomes a strategic initiative for providing competitive advantages to enterprises. Entire Resource Planning is considered a significant step in the evolution of ES.

Types of enterprise systems include Enterprise Resource Planning (ERP), Supply Chain Management (SCM), Supplier Relationship Management (SRM) and Product Lifecycle Management (PLM). The enterprise resource planning system integrates software applications in the same way as the business processes (such as purchasing, finance, human resources and inventory manage-

ment) are integrated. Within an ERP system, the software modules may communicate and share data. Each module has few applications that perform the required functions to execute particular end-to-end business processes. ERP applications support various operational and administrative tasks. These applications may also support a number of different industries, including oil and gas, banking and so on (Rouhani & Zare Ravasan, 2015a).

BUSINESS INTELLIGENCE OF ENTERPRISE SYSTEMS

Nowadays, the individual system approach applied to decision support systems (DSS) has been substituted by a new environmental approach. In the past, decision support systems were independent systems in an organization. They had a frail relationship with other systems. However, now enterprise systems are the foundation of an organization. And practitioners design and implement business intelligence as umbrella concept creates a decision-support environment for management in enterprise systems (Alter, 2004).

The increasing trend to use intelligent tools in business systems has increased the need for BI evaluation of enterprise systems. There are some delimited efforts to evaluate BI, but always they consider BI as tools or separated systems to enterprise systems.

In the year of 2008, Elbashir et al. suggested measuring the effects of business intelligent systems on the business process and provided some effective methods for the measurement (Elbashir, Collier, & Davern, 2008). One year later Lin et al. have developed a performance assessment model for business intelligent systems using analytic network process (ANP) (Lin, Tsai, Shiang, Kuo, & Tsai, 2009). In that way, business intelligence viewed as an independent system. Ghazanfari et al. (2011) made a model for the business intelligence assessment of Enterprise Systems. Their model can be applied to assess and rank enterprise

software based on their business intelligence capabilities. Further, Rouhani et al. (2012) proposed an evaluation model of business intelligence for enterprise systems using fuzzy TOPSIS. They utilized fuzzy TOPSIS approach to rank ESs, based their BI potentials. In the same year Popovic et al. (2012) proposed the model based on relationships between maturity, information quality, analytical decision-making culture and the use of information for decision-making as significant elements of the success of BI systems. In addition, Duan, and Da Xu (2012) pointed out the challenges and opportunities to smoothly connect industrial informatics to enterprise systems.

In 2013, Işık et al. have suggested a partial least squares model in BIS success domain (Işık, Jones, & Sidorova, 2013). This model emphasizes that decision environment has influence on the relationship between BI success and BI capabilities.

An intelligent approach to the evaluation and selection of enterprise systems includes four aspects relevant to the evaluation and selection of the most suitable enterprise systems. The first aspect of the approach is a set of evaluations that are executed in order to select supporting BI software with a high intelligence level, on the basis of the criteria of BI competency and capability. The second dimension includes analysis of the functional and non-functional suitability of the options on the basis of the requirements of the organization. The third aspect of the proposed approach is the maximization of the integration and relationships of the software and systems in the enterprise. The fourth aspect of the proposed approach is a quantitative evaluation of the implementation factors of cost and time (Rouhani & Zare Ravasan, 2015a).

BI AS REQUIREMENTS

The literature that considers the evaluation of the functional and non-functional suitability of the alternative enterprise software and systems using various criteria lacks. All previous researches con-

centrated on requirement engineering and software quality. They don't cover all aspects of enterprise systems evaluation, selection and implementation. Another observation based on the review of the literature (Jadhav & Sonar, 2009) is that although the functional criteria for software selection are altered for different software packages, other criteria related to the quality, cost and benefits, vendor, hardware and software requirements, the opinions of different stakeholders about the software package, and the output characteristics of the software package are universal and can be used for evaluation of any software package.

The functional requirements are the functions of the enterprise systems that are expected by the stakeholders and end-users. Wide-ranging industry research, in-depth interviews with users and analysis of current versions of software systems are methods for exploring the functional requirements.

The non-functional requirements are features of the system that are not covered by its functional description but are related to the capability and resiliency of the system. Researchers and practitioners have developed categories for the non-functional requirements, from different viewpoints.

Jadhav and Sonar (2011) classified non-functional requirements based on ISO/IEC9126 which has been withdrawn by the ISO/IEC 25010 in 2012 on the quality, technical, vendor, output and opinion categories.

BI capabilities are not related to specific enterprise system module or function and are related to every system with different mission. Consequently BI criteria can be discussed as non functional requirements of enterprise systems.

The intelligence evaluation criteria of enterprise systems and software are the competencies and capabilities of enterprise systems in the field of decision support and BI. There is no comprehensive list of intelligence criteria. Therefore, the authors reviewed various studies about decision support and BI and used their recent research (Rouhani & Zare Ravasan, 2015b). As a contribution,

basic BI capabilities that can be included in the list of non-functional requirements of enterprise systems and should be considered in design stages are defined below:

1. **Groupware:** Groupware is a shared tool for disseminating and sharing data, information and knowledge which facilitates collaborative communication and group decision-making. Groupware is also provides required infrastructure for the team and group working such as video conferences and documentation tools in a team working environment. Groupware has been regarded as one of the required factors in achieving BI competency in working systems (Marinoni, Higgins, Hajkowicz, & Collins, 2009; Reich & Kapeliuk, 2005).
2. **Optimization Technique:** It refers to ESs capability in supporting complex arithmetic analysis either using regular techniques such as Simplex and goal programming or meta-heuristic methods and algorithms such as Artificial Neural Network (ANNs), Genetic Algorithm (GA), Ant Colony (AC), and so on (Azadivar, Truong, & Jiao, 2009; Nie, Zhang, Liu, Zheng, & Shi, 2009).
3. **Learning Technique:** It refers to ESs learning capability in making decisions on the basis of prior decisions and the capability to learn from the historical data through discovering pattern and rules in decision making process (Power & Sharda, 2007; Ranjan, 2008).
4. **Import Data:** Data integration has been considered as one of the most important infrastructural requirements in the context of decision-making capabilities of ESs which refers to the ESs capability to extract, and load required data to its database and convert it to an understandable format (Alter, 2004; Quinn, 2009).
5. **Export Data:** It refers to ESs capability in exporting data and reports to other information systems, software packages, and other

- facilities such as Personal Digital Assistants (PDAs), mobile cell phones, and so on (Shi et al., 2007).
6. **Simulation Models:** In order to cut the costs and risks of doing real tests in operational environments, organizations need facilities that enhance a simulation of the reality and analyze the potential impacts of the events and relevant risks. This capability can help decision-makers either in a decision-making process and outcomes (Quinn, 2009; Zhan, Loh, & Liu, 2009).
 7. **Visual Graphs:** It refers to ESs capability in preparing user friendly and graphical reports and even video or 3D graphics to users (Azadivar, et al., 2009; Li, Shue, & Lee, 2008).
 8. **Summarization:** It refers to ESs capability in a summarization of the information while listing the main points in a brief and also in a comprehensive manner. This capability is of more importance in reporting features of ESs (Hemsley-Brown, 2005; Power, 2008).
 9. **Backward and Forward Reasoning:** Organizations' decision makers need ESs that justify the rational and reason of the decisions proposed by the system. Such a facility can help organizations in building trust atmosphere with regard to the results suggested by the system (Evers, 2008; Gottschalk, 2006).
 10. **Knowledge Reasoning:** Knowledge is the result of high-level ESs, formed by logical rules and support inferences in decision making. The capability of inference by machine is one aspect of BI competencies of ESs. Providing reasons based on machine inference (expert systems) in organizational decision making has been considered as an important and novel characteristic in BI by scholars and practitioners (du Plessis & du Toit, 2006; Evers, 2008).
 11. **Alarms and Warnings:** It refers to ESs capability in providing alarms and warnings in pre-defined thresholds which are substantially common in large integrated ESs such as ERPs. This capability can help decision makers proactively respond to risky situations (Power, 2008; Ross, Dena, & Mahfouf, 2009; Xiaoshuan, Zetian, Wengui, Dong, & Jian, 2009).
 12. **Dashboard/Recommender:** Every organization, regarding its unique goals and business requirements, needs some sorts of Key Performance Indicators (KPIs) differ from others. This criterion indicates ES capability in providing effective and tailor-made dashboards for new cases in organizations' different hierarchical levels (Bose, 2009; Hedgebeth, 2007).
 13. **Combination of Experiments:** Tacit and explicit knowledge of human resource should be used to verify the information of ESs. The capability of acquisition and combination of managers' and employees' experiments is an important characteristic for decision support. Historically, combination of experiments in knowledge management processes is classified as a significant requirement of decision making in organizations (Ross, et al., 2009).
 14. **Environment and Situation Awareness:** It refers to ESs' capability in extracting environmental information such as technology trends, changes in rules and regularity, rivals, suppliers and customers related indicators and also situation-specific information such as the time, place, person, challenges, possibilities, and so on, to provide more accurate results (du Plessis & du Toit, 2006; Raggad, 1997).
 15. **OLAP:** Online Analytical Processing tools enable users to interactively analyze multi-dimensional data from multiple perspectives which has been regarded as one of the most important capabilities of BI systems (Berzal, Cubero, & Jiménez, 2009; Rivest et al., 2005; Shi, et al., 2007).
 16. **Data Mining Techniques:** Data mining, a brand new and interdisciplinary field of computer science is the process of discovering

new patterns from large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics and database systems which is considered as one of the most important capabilities of BI in ESs (Berzal, et al., 2009; Shi, et al., 2007).

17. **Web and Email Channels:** World Wide Web (www) and the Internet have transformed the way people communicate and disseminate information. This media along with the emerging web 2.0 and semantic web has been considered as one of the main types of media for publishing organizational reports on the web and should be taken into account in BI capabilities of ESs. This criterion also encompasses the capability of automatically sending required information and reports to the pre-defined email list (Anderson, Jolly, & Fairhurst, 2007; Wen, Chen, & Pao, 2008).
18. **Mobile Channel:** Today, regarding the penetration of mobile devices in people everyday life, ESs should be empowered by mobile channel access to system functionalities and support managers by reports on their phones and handsets (Cheng, Lu, & Sheu, 2009; Power, 2008; Wen, et al., 2008).
19. **Intelligent and Multi-Agents:** Intelligent agent is an artificial agent operating in a software environment for doing pre-defined tasks which could be regarded as another BI capability of ESs. Also, some systems composed of multiple interacting intelligent agents known as a multi-agent system for doing complicated tasks (Gao & Xu, 2009).
20. **MCDM Tools:** Whether in our daily lives or in professional settings, there are typically multiple conflicting criteria that need to be assessed in making decisions. So, ESs should be enabled with Multi-Criteria Decision-making (MCDM) tools to manipulate this situation (İç & Yurdakul, 2009; Marinoni, et al., 2009).

As it is obvious above criteria or requirements are not related to specific enterprise system module

or function and are applicable in every system with different business scope and they have affect on quality of decision support services. Therefore they would be classified as non functional requirements of enterprise systems. But meeting all these non functional requirements are not mandatory and they can be meted in different level of BI maturity and development.

SOLUTIONS AND RECOMMENDATIONS

A series of methodologies and processes can be used to design and develop an enterprise system. Traditional developers have used an engineering approach such as the system development life cycle (SDLC), which is a systematic procedure of developing an information system through stages that occur in sequence. And recently many developers use Object-Oriented Analysis and Design (OOAD) for enterprise system engineering (Rumbaugh, 2003).

But in both mentioned approaches of system design, requirement engineering is critical process. This process focuses on the functional requirements and in next step selects the non functional level to deliver that functions. For as much as, the key to successful enterprise systems is an exact design, BI should be considered in selection of those non functional especially.

Based on the mentioned BI capabilities as requirements, new disciplines in the stage of requirements specification would be exist. And system designers should think more about business intelligence platform and tools and also should elaborate BI capabilities as non-functionalities of their desirable system. Requirement engineering gets new aspects and can be tailored and improved by the decision support vision in enterprise system. As current research contribution, the BI-Based design area of enterprise systems proposed in below:

- **“What Happened” Analysis:** Visual graphs, Summarization, Import data from

other systems, Export reports to other systems, OLAP, Web channel, Mobile channel, E-mail channel, Alarms and warnings.

- **“Why Happened” Analysis:** Backward and forward reasoning, Knowledge reasoning, Environmental awareness, Simulation models, Dashboard.
- **“What Next” Analysis:** Group decision-making (Groupware), Optimization technique, Learning technique, Combination of experiments, Intelligent agent, MCDM tools, Recommender.

FUTURE RESEARCH DIRECTIONS

Considering the current review and recommendation, some respectful and applicable suggestion can be presented for implication as follows: first, in every enterprise system development methodology a major step can be added for BI requirement analysis and design. Second, determination of the most needed BI non-functional requirement to reach targeted decision support benefits in enterprise systems has been clear, therefore the organizations have a roadmap to implement the right BI not in the form of independent tools but in the form of embedding requirements in enterprise systems. Third, in acquiring and selection of enterprise systems the proposed criteria would be useful as evaluation criteria.

Based on mentioned findings, it is possible to make some insightful recommendations for future research. First, it is possible to establish the requirement engineering best criteria to do list as a standard. Second, presenting a customized information system development methodology based on BI requirements. Finally, it is recommended to consider and test the effectiveness model of those BI capabilities based on ES user satisfaction data.

CONCLUSION

Business intelligence includes information and knowledge of the organization, and also system and systematic process of decision support in business context. Nowadays, BI capabilities have been transformed to the new criteria for evaluation and accusation of enterprise systems and software. In this chapter, business intelligence and enterprise system concepts were reviewed and the BI position in these system was discussed. The recent efforts show this position as innovative requirements of enterprise systems. Therefore based on this point of view, novel and modern non-functional requirements (BI capabilities) should be embedded in enterprise systems and software. For this reason, the design stage of traditional system development life cycle has the responsibility about decision support besides the functional support and should consider and elaborate the proposed BI requirements. As current research contribution, the BI-Based design area of enterprise systems should think about: “What happened analysis”, “Why happened analysis” and “What next analysis” non-functional requirements.

REFERENCES

- Alter, S. (2004). A work system view of DSS in its fourth decade. *Decision Support Systems*, 38(3), 319–327. doi:10.1016/j.dss.2003.04.001
- Anderson, J. L., Jolly, L. D., & Fairhurst, A. E. (2007). Customer relationship management in retailing: A content analysis of retail trade journals. *Journal of Retailing and Consumer Services*, 14(6), 394–399. doi:10.1016/j.jretconser.2007.02.009
- Azadivar, F., Truong, T., & Jiao, Y. (2009). A decision support system for fisheries management using operations research and systems science approach. *Expert Systems with Applications*, 36(2), 2971–2978. doi:10.1016/j.eswa.2008.01.080

- Azoff, M., & Charlesworth, I. (2004). *The new business intelligence*. A European Perspective, Butler Group, White Paper.
- Berson, A., & Smith, S. J. (1997). *Data warehousing, data mining, and OLAP*. McGraw-Hill, Inc.
- Berzal, F., Cubero, J. C., & Jiménez, A. (2009). The design and use of the TMiner component-based data mining framework. *Expert Systems with Applications*, 36(4), 7882–7887. doi:10.1016/j.eswa.2008.11.033
- Bose, R. (2009). Advanced analytics: Opportunities and challenges. *Industrial Management & Data Systems*, 109(2), 155–172. doi:10.1108/02635570910930073
- Cheng, H., Lu, Y. C., & Sheu, C. (2009). An ontology-based business intelligence application in a financial knowledge management system. *Expert Systems with Applications*, 36(2), 3614–3622. doi:10.1016/j.eswa.2008.02.047
- du Plessis, T., & du Toit, A. (2006). Knowledge management and legal practice. *International Journal of Information Management*, 26(5), 360–371. doi:10.1016/j.ijinfomgt.2006.06.003
- Duan, L., & Da Xu, L. (2012). Business intelligence for enterprise systems: A survey. *Industrial Informatics. IEEE Transactions on*, 8(3), 679–687.
- Eckerson, W. W. (2010). *Performance dashboards: measuring, monitoring, and managing your business*. John Wiley & Sons.
- Elbashir, M. Z., Collier, P. A., & Davern, M. J. (2008). Measuring the effects of business intelligence systems: The relationship between business process and organizational performance. *International Journal of Accounting Information Systems*, 9(3), 135–153. doi:10.1016/j.accinf.2008.03.001
- Evers, M. (2008). An analysis of the requirements for DSS on integrated river basin management. *Management of Environmental Quality: An International Journal*, 19(1), 37–53. doi:10.1108/14777830810840354
- Gao, S., & Xu, D. (2009). Conceptual modeling and development of an intelligent agent-assisted decision support system for anti-money laundering. *Expert Systems with Applications*, 36(2), 1493–1504. doi:10.1016/j.eswa.2007.11.059
- Gbosbal, S., & Kim, S. K. (1986). Building effective intelligence systems for competitive advantage. *Sloan Management Review (1986-1998)*, 28(1), 49.
- Ghazanfari, M., Jafari, M., & Rouhani, S. (2011). A tool to evaluate the business intelligence of enterprise systems. *Scientia Iranica*, 18(6), 1579–1590. doi:10.1016/j.scient.2011.11.011
- Ghazanfari, M., Rouhani, S., Jafari, M., & Taghavi-fard, M. (2009). ERP Requirements for supporting management decisions & Business Intelligence. *The ICFAI University Journal of Information Technology*, 3.
- Gottschalk, P. (2006). Expert systems at stage IV of the knowledge management technology stage model: The case of police investigations. *Expert Systems with Applications*, 31(3), 617–628. doi:10.1016/j.eswa.2005.09.063
- Hedgebeth, D. (2007). Data-driven decision making for the enterprise: An overview of business intelligence applications. *Vine*, 37(4), 414–420. doi:10.1108/03055720710838498
- Hemsley-Brown, J. (2005). Using research to support management decision making within the field of education. *Management Decision*, 43(5), 691–705. doi:10.1108/00251740510597716
- İç, Y. T., & Yurdakul, M. (2009). Development of a decision support system for machining center selection. *Expert Systems with Applications*, 36(2), 3505–3513. doi:10.1016/j.eswa.2008.02.022
- Işık, Ö., Jones, M. C., & Sidorova, A. (2013). Business intelligence success: The roles of BI capabilities and decision environments. *Information & Management*, 50(1), 13–23. doi:10.1016/j.im.2012.12.001

- Jadhav, A. S., & Sonar, R. M. (2009). Evaluating and selecting software packages: A review. *Information and Software Technology*, 51(3), 555–563. doi:10.1016/j.infsof.2008.09.003
- Jadhav, A. S., & Sonar, R. M. (2011). Framework for evaluation and selection of the software packages: A hybrid knowledge based system approach. *Journal of Systems and Software*, 84(8), 1394–1407. doi:10.1016/j.jss.2011.03.034
- Li, S. T., Shue, L. Y., & Lee, S. F. (2008). Business intelligence approach to supporting strategy-making of ISP service management. *Expert Systems with Applications*, 35(3), 739–754. doi:10.1016/j.eswa.2007.07.049
- Lin, Y.-H., Tsai, K.-M., Shiang, W.-J., Kuo, T.-C., & Tsai, C.-H. (2009). Research on using ANP to establish a performance assessment model for business intelligence systems. *Expert Systems with Applications*, 36(2), 4135–4146. doi:10.1016/j.eswa.2008.03.004
- Lönnqvist, A., & Pirttimäki, V. (2006). The measurement of business intelligence. *Information Systems Management*, 23(1), 32–40. doi:10.1201/1078.10580530/45769.23.1.20061201/91770.4
- Marinoni, O., Higgins, A., Hajkowicz, S., & Collins, K. (2009). The Multiple Criteria Analysis Tool (MCAT): A new software tool to support environmental investment decision making. *Environmental Modelling & Software*, 24(2), 153–164. doi:10.1016/j.envsoft.2008.06.015
- Nie, G., Zhang, L., Liu, Y., Zheng, X., & Shi, Y. (2009). Decision analysis of data mining project based on Bayesian risk. *Expert Systems with Applications*, 36(3), 4589–4594. doi:10.1016/j.eswa.2008.05.014
- Nylund, A. (1999). Tracing the BI family tree. *Knowledge Management*, 60.
- Petrini, M., & Pozzebon, M. (2009). Managing sustainability with the support of business intelligence: Integrating socio-environmental indicators and organisational context. *The Journal of Strategic Information Systems*, 18(4), 178–191. doi:10.1016/j.jsis.2009.06.001
- Popovič, A., Hackney, R., Coelho, P. S., & Jaklič, J. (2012). Towards business intelligence systems success: Effects of maturity and culture on analytical decision making. *Decision Support Systems*.
- Power, D. J. (2008). Understanding data-driven decision support systems. *Information Systems Management*, 25(2), 149–154. doi:10.1080/10580530801941124
- Power, D. J., & Sharda, R. (2007). Model-driven decision support systems: Concepts and research directions. *Decision Support Systems*, 43(3), 1044–1061. doi:10.1016/j.dss.2005.05.030
- Quinn, N. W. T. (2009). Environmental decision support system development for seasonal wetland salt management in a river basin subjected to water quality regulation. *Agricultural Water Management*, 96(2), 247–254.
- Raggad, B. G. (1997). Decision support system: Use IT or skip IT. *Industrial Management & Data Systems*, 97(2), 43–50. doi:10.1108/02635579710161331
- Ranjan, J. (2008). Business justification with business intelligence. *Vine*, 38(4), 461–475. doi:10.1108/03055720810917714
- Reich, Y., & Kapeliuk, A. (2005). A framework for organizing the space of decision problems with application to solving subjective, context-dependent problems. *Decision Support Systems*, 41(1), 1–19. doi:10.1016/j.dss.2004.05.001
- Rivest, S., Bédard, Y., Proulx, M. J., Nadeau, M., Hubert, F., & Pastor, J. (2005). SOLAP technology: Merging business intelligence with geospatial technology for interactive spatio-temporal exploration and analysis of data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 60(1), 17–33. doi:10.1016/j.isprsjprs.2005.10.002

Ross, J. J., Dena, M. A., & Mahfouf, M. (2009). A hybrid hierarchical decision support system for cardiac surgical intensive care patients. Part II. Clinical implementation and evaluation. *Artificial Intelligence in Medicine*, 45(1), 53–62. doi:10.1016/j.artmed.2008.11.010 PMID:19112011

Rouhani, S., Asgari, S., & Mirhosseini, S. V. (2012). Review study: Business intelligence concepts and approaches. *American Journal of Scientific Research*, 50, 62–75.

Rouhani, S., Ghazanfari, M., & Jafari, M. (2012). Evaluation model of business intelligence for enterprise systems using fuzzy TOPSIS. *Expert Systems with Applications*, 39(3), 3764–3771. doi:10.1016/j.eswa.2011.09.074

Rouhani, S., & Zare Ravasan, A. (2015a). (in press). Multi-objective model for intelligence evaluation and selection of enterprise systems. *Int. J. Business Information Systems. Article*.

Rouhani, S., & Zare Ravasan, A. (2015b). A practical framework for assessing business intelligence competencies of enterprise systems using fuzzy ANP approach. *Int. J. Applied Decision Sciences*, 8(1), 52–82. doi:10.1504/IJADS.2015.066559

Rumbaugh, J. (2003). *Object-oriented analysis and design*. OOAD.

Shi, Z., Huang, Y., He, Q., Xu, L., Liu, S., Qin, L., & Zhao, L. et al. (2007). MSMinera developing platform for OLAP. *Decision Support Systems*, 42(4), 2016–2028. doi:10.1016/j.dss.2004.11.006

Wen, W., Chen, Y., & Pao, H. (2008). A mobile knowledge management decision support system for automatically conducting an electronic business. *Knowledge-Based Systems*, 21(7), 540–550. doi:10.1016/j.knosys.2008.03.029

Xiaoshuan, Z., Zetian, F., Wengui, C., Dong, T., & Jian, Z. (2009). Applying evolutionary prototyping model in developing FIDSS: An intelligent decision support system for fish disease/health management. *Expert Systems with Applications*, 36(2), 3901–3913. doi:10.1016/j.eswa.2008.02.049

Xu, L. D. (2011). Enterprise systems: State-of-the-art and future trends. *Industrial Informatics. IEEE Transactions on*, 7(4), 630–640.

Zhan, J., Loh, H. T., & Liu, Y. (2009). Gather customer concerns from online product reviews-A text summarization approach. *Expert Systems with Applications*, 36(2), 2107–2115. doi:10.1016/j.eswa.2007.12.039

ADDITIONAL READING

Ghazanfari, M., Jafari, M., & Rouhani, S. (2011). An Assessment model for Intelligence Competencies of Accounting Information Systems. *International Journal of Scientific & Engineering Research*, 2(4), 1–6.

Loshin, D. (2012). *Business intelligence: the savvy manager's guide*. Newnes Publishing.

KEY TERMS AND DEFINITIONS

Business Intelligence: Business intelligence, or BI, is a general term that refers to a range of capability and tools used to analyze an organization's raw data and transform them to actionable knowledge by data mining, online analytical processing, querying, reporting and so on.

Enterprise System: Enterprise system (ES) is large-scale application software package that support business processes, resource management, information flows, reporting, and data analytics in complex organizations and in scale of enterprise.

Functional Requirement: Functional requirement that define specific behavior or function of system which come from business needs. The plan for implementing functional requirements is detailed in the system design.

Nonfunctional Requirements: A non-functional requirement is a system requirement that specify criteria that can be used to judge the operation of a system, rather than specific behaviors. These requirements determine the quality of the system or software. The plan for implementing non-functional requirements is detailed in the system architecture.

Requirement: A requirement is a singular documented physical and functional need that a particular design, product or process must be able to perform. It is a statement that identifies a necessary attribute, capability, characteristic, or quality of a system for it to have value and utility to a customer, organization, internal user, or other stakeholder.

System Design: Systems design is major phase of information system development and is the process of defining the architecture, components, modules, interfaces, and data for a system to satisfy specified requirements.

Deployment of Enterprise Architecture From the Activity Theory Perspective



Tiko Iyamu

Cape Peninsula University of Technology, South Africa

Irja Naambo Shaanika

Namibia University of Science and Technology – Windhoek, Namibia

INTRODUCTION

Enterprise architecture (EA) is defined as an integrated and holistic vision of a system's fundamental organisation, embodied in its elements (people, processes, applications, and so on), their relationships to each other and to the environment, and the principles guiding its design and evolution (Janssen & Kuk, 2006). According to Kaisler, Armour and Valivullah (2005), EA identifies the main components of the organisation, its information systems, the ways in which these components work together in order to achieve defined business objectives and the way in which the information systems support the business processes of the organisation. Kamogawa and Okada (2009) asserts that the compelling need of EA is to enable strategic business goals and organisations to derive strategic outcomes from EA in terms of operational excellence.

Organisations do admit and acknowledge the importance of service delivery to their clients and citizens respectively. However, many clients and citizens continue to be dissatisfied with the type of services that they get. In attempt to get solution, some countries opt for transformation of their governments' activities into e-governments (Janssen and Kuk 2006; Marawar, Kale and Araspure 2010; Mohamed et al, 2012). According to Lee et al, (2013) the conception of Government-wide EA is the result of e-government considerations. Notably, EA is promoted as a key tool in the

transformation and modernisation of country governments (Madsen & Heje, 2009).

The rationale for the deployment (development and implementation) of the EA may vary from one organisation to another. However, the underlying aim is to provide a better structure in order to effectively manage IT-related projects and development activities across an organisation (Janssen and Kuk, 2006). Mohamed et al. (2012) posit that reasons for EA adoption include reducing the cost of IT and business operations by identifying duplications and opportunities for reuse and enabling interoperability and providing technical and managerial standards for agencies.

In some countries, the Government-wide EA's aim is for each ministry's investments in IT to be aligned with government-wide policy goals (Lee et al., 2013). The author argues that, government Ministries experience challenges in the planning, development and implementation of their information systems and supporting technologies. Government-wide enterprise architecture is seen as the strategy to eliminate inconsistencies and duplication of efforts in information systems across government Ministries. It enables improved citizens and business relationships, where by citizens can interact with Ministries as integrated businesses (Janssen and Cresswell 2005). The objectives of this study was to presents the influencing factors in the deployment of EA in organisations. This will assist organisations including government to assess and examine how

EA could be deployed for efficiency and effectiveness in addressing the challenges of uniformity integration, of processes, systems and technologies within their organisations.

The remainder of this paper is structured into four sections. In the first section, activity theory is discussed. The second section presents analysis of EA deployment from the perspective of activity theory. The third section presents a model and discussion on the factors which influences the deployment of EA in an organisation. A conclusion is finally drawn in the last section.

BACKGROUND

Activity Theory as a Lens

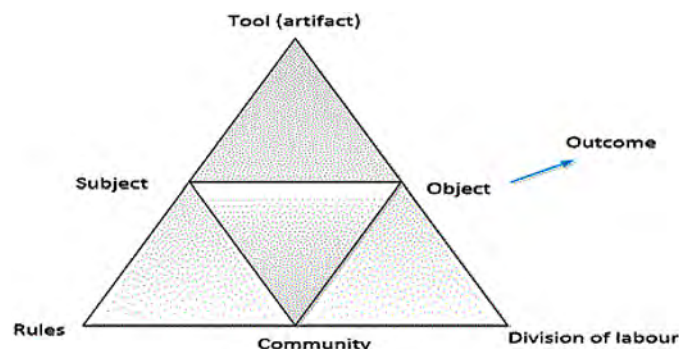
This study uses Activity Theory (AT) to analyse the factors that influence the deployment of EA. The AT is a socio- technical theory that is concerned with the development of social activities. As shown in Figure 1 below, AT consists of six main components. According to Golsorkki, Rouleau, Seidl and Vaara (2010), the theory “conceptualises the on-going construction of activity as a product of activity systems comprising the subject; the community within which subject interacts with other subjects; the tools that mediate between subjects, community and objective” (p. 127). In activity theory, subject is referred to as any living being with needs. As noted by Kaptelinin and

Nardi (2006) in activity theory not every actor is a subject. According to Kaptelinin and Nardi (2006) in AT a subject have needs that can be met only by being and acting in the world. Therefore non-living actors such as cars, computers cannot be regarded as subjects. Actors use tools to pursue their goals, and perform their activities. Uden (2007) stated that tools can be physical entities, such as a hammer or psychological such as language, culture or ways of thinking.

AT is used in information systems and technology studies primarily because the development, implementation, and management of systems and technologies are regarded as social technical activities. During the development and implementation of information systems and technologies in organisations, many actors are involved. Their involvement is governed by the organisation’s rules, cultures, values, and norms. The governance is mainly to ensure and enforce commitment, focus, and order; critical aspects that are needed in order to achieve organisational goals and objectives. However, rules can come in varying degrees, and can constrain or liberate the activity of a system (Yamgata-Lynch, 2010).

Also, information systems and technologies are developed and used according to the actors’ knowledge and specialisations. AT defines specialisations as a form of division of labour. Uden (2007) stated that the divisions of labour are concerned with how tasks are divided between community members and also the division of

*Figure 1. Activity theory
Engestrom et al., 1999.*



power and status. Due to the powers that actors possess, they can enable or constrain systems and technologies' activities. AT provided a theoretical lens which was used to analyse the activities and processes of the actors in the deployment of EA.

Analysis of EA Deployments

The development and implementation of information systems and information technologies is a challenging process. In organisations, process involves IT specialists, stakeholders and business staff with diverse and often conflicting goals. The EA provides approaches and strategies to deal with such challenges. Through its various domains which include business, information, application and technical architecture, EA provides a holistic view of the Enterprise's operations, processes and activities. In general terms, EA specifies how information systems and technologies are related to the overall business processes and organisational activities. The deployment of EA enables an organisation to support, manage and govern their information systems and their supporting technologies to achieving and improving on effectiveness and efficiency. This manifests to minimising risks and inconsistencies through principles, standards and governance of technologies. Also, this addresses the challenges caused by constant changes as a result of changing environmental trends and business processes, which inevitably leads to obsolescence in IT. EA enforces change from the perspectives of the current to the future, at different levels and across its architectural domains. This change also helps to identify redundancies and address across the IS/IT infrastructures, for organisation's needs, goals and objectives.

The EA has been deployed in many different environments and in various ways in recent years. Yet, it continues to strive in the challenges it was purposefully deployed to addressing. From the perspective of AT, this section presents analysis of the underlining guide and scope within which the EA is deployed in organisations. This includes understanding the Enterprise, Organisational

Units, the effect of Change in context, the necessity of Information Systems and Technologies, and the recipients of services:

The Enterprise

The enterprise represents the bigger organisation with all the connections to the external environment which includes business partners, suppliers, debtors, stakeholders and customers. Within context, the enterprise mandates and delegates the power to manage the selection, development, implementation, support and management of processes, information systems and technologies across its business units. Its responsibility is to ensure that the required IS/IT complies with the defined ICT policies and standards, and to advice on best practices and strategies in the deployment, support and management of information systems and technologies.

As IT needs arise within individual units, the enterprise communicates them to the respective authorities within the environment. However, not every employee is satisfied with the communication channels that are used to communicate the IT needs of the organisation. Some employees sometimes do feel that management discussed matters without involving them from the very beginning. As an implication, some IS/IT artefacts that are deployed are not being supported or reluctantly by some employees in the organisation. This also results resistant towards the organisation's projects.

Within the enterprise, the selection, development and implementation of technologies involved actors from different backgrounds. Depending on the actors' roles and responsibilities, some used their political power to influence and manipulate decisions. As a result, some decisions were not made for the best interest of the whole organisation but rather to serve personal interests.

Organisational Units

The Organisational Units are defined accordance to the businesses and activities of the organisation. The most common among organisations are

Information Technology (IT) and Human Resource (HR), mainly because of their essentiality. In an attempt for an organisation to achieve its goals and objectives, the various Units adopts IT artefacts to support and enable their activities and processes. However, some of Units are often challenged with their ICT's operations in an organisation. Some challenges include the development and implementation of the supporting technologies, compatibility and integrations issues. Such challenges prohibit the Units from attaining the potential of their investment on IT.

The IT Unit takes cognisance of the operations, processes and activities of Units in providing them with support and enablement. To do this, the focal actor, which often the IT Chief Information Officer or Director must have a good understand of the Units, and have relevant skills-set under his/her auspices. In some instances, the employees specialised in more than one area. As a result, some employees belong to more than one network within the organisational Units. Their participation in the organisation's projects is therefore influenced by their involvement in other networks.

In some organisations, many actors are not aware of the underlining roles and responsibilities of the IT Unit. Some employees from the business Units indicate that IT is supposed to support their activities, irrespective of what it is including telephone and other technological artefacts. This is due to lack of communication between the units (business and IT). Also, each unit sometimes feel that they are more superior to the others. As a result there exist competitions between the units instead of working together for the interest of the organisation. EA must is expected to bridge the gap between units, using communication as one of its tools.

Information Systems and Technologies

The Units and the Enterprise at large make use of IT artefacts for various reasons, such as to collaborate and share information from person-to-person, unit-to-unit and organisation-to-organisation. Such collaborations ensure that the organisation

have real -time information that are required by employees and clients to make informed decisions.

The selection, development and implementation of EA require knowledge and skills to succeed. As such, the employees interpret and understand the technologies based on their knowledge and skills. EA skills in the many organisations are acquired through formal education or through on-the-job training. However, knowledge and expertise is a challenge, but remain an essential aspect of the deployment of EA. In an attempt to resolve this challenge, some organisations do outsource the activities of EA to other organisations. Interestingly, the challenge persists because not every activities of the EA can be outsourced due to its strategic nature. Irrespective, insiders are required in the implementation and support of those activities that are outsourced.

Also the deployment of IS/IT depends on both technical and non-technical factors for efficiency and effectiveness. However, in some units of many organisations, the technical factors are often more glorified than the non-technical factors. Thus, non-technical factors, such as organisational culture and structure need to be represented and understood in the deployment of EA in organisations. This includes the creation and communication of employees' roles and responsibilities across the structure. Depending on the organisational culture, the activities and processes are legitimised, and therefore influence the deployment of EA.

The selection, development and implementation activities and processes are guided and informed by the business rules, policies and standards. The use of standards and policies must be well understood across the IT and business units. Standards and policies enable and can also constrain uniformity and order necessary factors and activities, in achieving organisational goal and objectives.

Change

Organisational processes and activities are constantly changing. The change comes from different sources, and is caused by the internal and external

factors, such as globalisation, new markets and international laws. These changes impact the way society lives and conducts businesses. The impact can be both negative and positive, depending on how the change is handled. The organisation's survival and competitive advantage also depends on how efficient and effective it manages and responds to change, using the EA.

Information systems and technologies effectively respond to its purpose according to their life span. For example, hardware such as servers, printers and computers function effectively for a period of three to four years. Due to the fast changing technologies, organisations often find it hard to keep up with the latest technologies. As a result, most organisations are operating with obsolete technologies. The application of obsolete technologies is often a challenge and as such, the technologies are not compatible and *integratable* with others.

As activities and processes within society change, they have an impact on IS/T. As such, IS/T need to be re-engineered and re-designed to support and enable new processes and advancements in the society. Change in an environment is communicated to respective authorities. The communication is sometimes carried based on understanding and interest of the actor. This makes actors critically important in the deployment of EA.

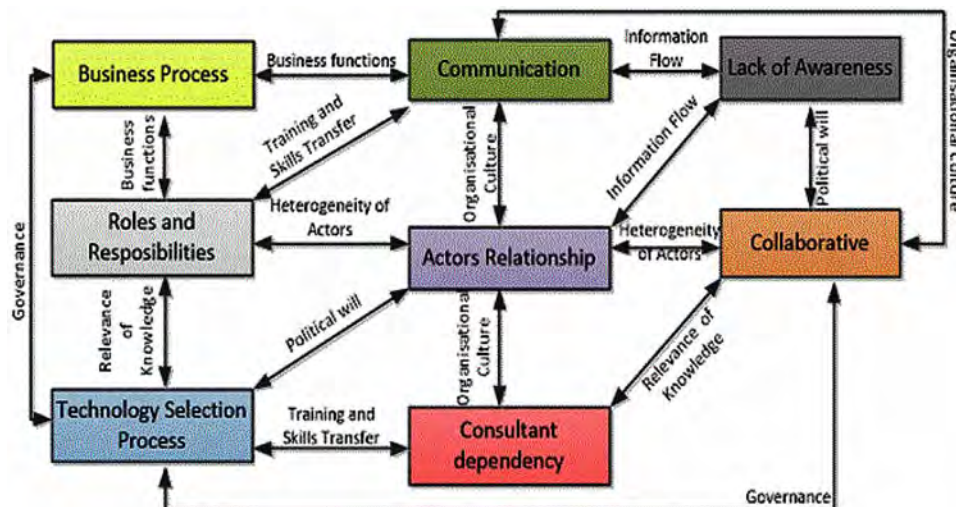
MODEL FOR INFLUENCING FACTORS

The above analysis was carried out, using Activity Theory. Based on the analysis, some factors were found to be of significant influence in the development, implementation and management of EA within context and relevance of organisations' environments. The factors include business process, communication, lack of awareness, roles and responsibilities, actors' relationships, collaborative technology selection processes and consultant dependency, as depicted in the Model (Figure 2). The Model illustrates the relationship, interconnectivity and dependence of the factors on each other. The discussion on the factors should be read with the figure to get a better understanding on how the factors influence the deployment of EA in an organisation.

Business Process

EA is process oriented, and as such, it depends on business and IT processes in its deployment. Business processes define how the activities are executed and by who. They are used as guidelines by employees. Processes are of significance to the organisations' effective operations. How processes are communicated and interpreted by employees,

Figure 2. Factors influencing EA deployment



does impacts the deployment of EA, which manifest in IT processes. IT processes include selection, development and implementation of technologies in organisations.

Processes are defined based on the various business functions, which defines the individual business's daily operations. The operations are the activities that uniquely differentiate the business units within an organisation. Due to business functions' interdependency, processes are interconnected across the organisations. Interdependency calls for an understanding between the various business functions and processes. Not considering and a lack of understanding of business functions interdependence during the selection, development and implementation of information systems and technologies leads to process misalignments.

Communication

Communication plays a vital role in the deployment of EA, which has impact on the selection, development and implementation of information systems and technologies in an organisation. How business process, needs and technology needs are communicated to the employees impacts their involvement in the organisational activities. Communication is done according to the policies and standards defined. This ensures a governed interaction among the employees. The communication process is influenced by the organisational structure, which influence how EA is deployed in the organisation. Employees at the operational level communicate on a daily basis as they interact with each other when carrying out their tasks.

Communication can promote a positive environment where there is sharing of information and knowledge. This leads to employees' full participation in organisational activities and as a result, new and fresh ideas are brought on the table. On the other hand, a lack of communication among the employees can hinder organisations' deployment of the EA. When the organisation's processes and activities are not well communicated, employees

are reluctant to participate in the deployment of EA. Reluctance leads to conflicts and demoralisation among the employees, in the process.

Lack of Awareness

To achieve the organisation's objectives and goals through the deployment of the EA, business processes, principles and standards are essentially important, and must be communicated to the employees. However, the communication process is not always constructive and done as expected. As such, it results to lack of awareness among employees, in some cases. Due to the lack of awareness, activities are not always carried out as per defined standards, in the deployment of the EA. Awareness is impacted by information flow in the organisation.

Information flow is done from top-down, from management to operational employees as well as from bottom-up, from the operational employees to the management. Effective and efficient information flow enables rapid business decision making and information sharing. The design, development and implementation of information systems aims to establish the value and importance of using information effectively across the various units of an organisation, as well as the need to attain collaboration with customers.

Roles and Responsibilities

Employees participate in the deployment of the EA according to their roles and responsibilities. Roles and responsibilities are defined to ensure that employees carry out the tasks allocated to them. The roles and responsibilities are discussed between the employees and the management. The negotiations between the employees and management is to ensure that both parties are aware and understand what is expected of them and to resolve roles and responsibilities conflicts, in specific task, such as deployment of the EA.

Roles and responsibilities are assigned to employees in accordance to relevance factors, such as skills and knowledge. However, some employees are knowledgeable in more than one business function. As a result, there is heterogeneity of actors within organisations. Due to the heterogeneity of actors, an actor can partake in various projects relating to the development, implementation, maintenance and usage of systems and technologies. The involvement of the actors within a project leads to the creation of actors' network. In actor network theory, a network is an interconnection of both technical and non-technical actors. Actors can belong to more than one kind of network. The actors' performance in a network impacts their involvement in other networks. As such, due to the commitments, they might not fully participate in other networks.

Actor Relationship

Technical and non-technical actors depend on each other through their interaction, in the deployment of EA in their organisations. This could hold to the symmetry argument by actor network theory, which says that human and non-human actors are equal, and should be treated as such. Even though ANT has been criticised for such philosophical assumptions, the argument seems to reflect from the deployment of EA, in that humans depends on technologies and processes for EA to be deployed. Technical actors are used as tools to enable and support employees when carrying out their activities. This is to promote efficiency and effectiveness within organisations. On the other hand, the technical actors were selected, developed and maintained by the non-technical actors through the means of other technical actors. Actors' relationship is of importance for the organisations' operations. Without their interactions, it would be impossible to achieve the organisational goals and objectives.

Collaborative

Actors collaborate with each other on the basis of their relationship. The collaboration that happens between individuals or groups impacts the deployment of EA. Some of the collaborations are carried out through the means of information technology, such as email, which are subjectively read and understood. However, the use of such technology was viable and feasible from two different perspectives; to eradicate duplication and enact synchronisation, including support and the enabling of real-time data across business units.

Collaboration is crucial as business units are dependent on each other when carrying out their activities. Through collaborations, employees share ideas and resolve disagreements. This is seen as an effective way to support employees within the organisations.

Technology Selection Process

Technologies are selected according to relevance to business unit or the organisation. Also, the focal actor can decide on technology selection based on personal interest. Such process has impact on how EA is deployed in an organisation. The selection process involves actors from business and IT settings, their knowledge about business functions and information systems and technologies. This promotes understanding and collaborations between business and IT units within the organisations. The lack of cooperation between the two teams leads to conflict within the various projects. Rivalry on many issues during the selection, development and implementation process results into misalignments among business and IT processes.

The selection process needs to be controlled and managed accordingly. This is not often the case, which makes a stronger case for EA and governance. Governance helps to maintain the scope and boundaries within which employees

carry out their activities. Through governance, the principles, policies and standards are developed. Principles are guiding statements that guide employees' actions on a daily basis. Employees' interactions are done according to the defined organisational principles. Good principles lead to good professional work ethics and a collaborative working environment.

The purpose of standards is to ensure uniformity and consistency across the organisation. Process uniformity and consistency leads to ease of interaction among the employees as they are able to relate to one another about the processes in place. This ensures that the developed and implemented information systems and technologies are fit for supporting business processes. In addition, technology selection process is influenced by politics. Politics is a manifestation of actors' relationships. Because of their roles and responsibilities, employees use their power to influence the selection process. In most cases it is carried to satisfy individuals or groups' personal interests rather than the organisation's business interest.

Consultancy Dependency

Technologies selected are either developed in-house or are outsourced. The rationale for outsourcing is due to the lack of technical know-how within the organisation. Outsourcing enables organisations to acquire technological artefacts from other organisations that have the necessary skills and competencies. This allows the employees to focus and concentrate on other projects they are capable of managing.

Also, outsourcing entails the process of skills transfer by the consultants to the organisations' employees. In an attempt to acquire such skills, during the development and implementation process, employees are often involved and trained on how to use and maintain the technologies. Due to the specialised nature of information systems and technology, in most cases skills transfer does not always take place as expected. This can be due to the systems complexities, and as such there is

limited time for the employees to learn, adopt and acquire the necessary skills about the technologies.

Also the consultants feel that they are the custodians of the application hence they should be the ones responsible for applications and systems maintenance. This has a serious influence on the collaboration among the organisational employees and the consultants. As such, some actors do not fully collaborate and this negatively impacts skills and knowledge transfers within the organisation. As a result, the organisations constantly depend on the consultants for support as they are more knowledgeable about the implemented technologies.

FUTURE RESEARCH DIRECTIONS

Organisations strive to obtain value from their supporting technologies; however the management of these various technologies can be challenging. These have negative impacts on service delivery. As IT becomes part of every organisation there is a need for governance and management approaches of information systems artefacts. In doing so, the influencing factors become critically important. Deployment of strategic approaches such as EA enables organisations and governments administrations to derive values and return on investments from their technologies. Thus, EA support and enable end-to-end of processes and activities, towards actualisation of objectives. This is attainable through its various domains. EA is not a once off project as business as well as technology requirements are constantly changing. Its success depends on the organisation support and understanding of various factors.

Understanding the impacting factors is also a challenge on its own. There is need for deeper interpretation of how factors impacts each other and technologies. Thus, the activity theory was employed in the analysis, purposely to enhance our understanding of factors, which could influence the deployment of EA in an organisation.

CONCLUSION

This chapter contributes to the ineffective implementation and management of information systems, as well as lack of guidelines for defining hard and soft boundaries, the interfaces and the alignments of all components of the enterprise. As a result, organisations fail to plan and anticipate the business processes and technical innovations and this has caused IT to become a costly investment for the government. In face of these challenges, this chapter employs the lens of AT to understand and examine the activities which impacts the deployment of EA in organisations.

REFERENCES

- Engeström, Y., Miettinen, R., & Punamäki, R. L. (Eds.). (1999). *Perspectives on Activity Theory*. Cambridge University Press. doi:10.1017/CBO9780511812774
- Golsorkki, D., Rouleau, L., Seidl, D., & Vaara, E. (2010). *Strategy as practice*. New York: Cambridge University Press.
- Janssen, M., & Cresswell, A. (2005) Enterprise architecture integration in e-government. *Proceedings of the 38th Hawaii International Conference on System Sciences*, (pp. 1-10). IEEE.
- Janssen, M., & Kuk, G. (2006) A Complex Adaptive System Perspective of Enterprise Architecture in Electronic Government. *Proceedings of the 39th Hawaii International Conference on System Sciences* (pp. 1-10). IEEE. doi:10.1109/HICSS.2006.6
- Kaisler, S., Armour, F., & Valivullah, M. (2005) Enterprise Architecting: Critical Problems. *Proceedings of the 38th Hawaii International Conference on System sciences* (pp. 224b-224b). IEEE. doi:10.1109/HICSS.2005.241
- Kamogawa, T., & Okada, H. (2009) Enterprise Architecture Create Business Value. *Proceedings of the Ninth Annual International Symposium on Applications and the Internet* (pp. 205-208). IEEE. doi:10.1109/SAINT.2009.46
- Kaptelinin, V., & Nardi, B. (2006). *Acting with technology: Activity theory and interaction design*. Massachusetts Institute of Technology.
- Lee, Y. J., Kwon, Y. I., Shin, S., & Kim, E. J. (2013) Advancing government-wide enterprise architecture: A meta-model approach. *Proceedings of the 15th International Conference Advanced Communication Technology (ICACT)*, (pp. 886-892). IEEE.
- Madsen, K., & Heje, J. (2009) Enterprise architecture in government: Fad or future? *Proceedings of the 42nd Hawaii International Conference on System Sciences* (pp. 1-10). IEEE.
- Marawar, T., Kale, S., & Araspure, K. (2010) E-governance. *Proceedings of the International Conference on Data Storage and Data Engineering* (pp. 183-187). IEEE.
- Mohamed, M., Galal-Edeen, G., Hassan, H., & Hasanien, E. (2012) An evaluation of enterprise architecture frameworks for E-government. *Proceedings of the 2012 International Conference on Computer Engineering & Systems* (pp. 255-260). IEEE. doi:10.1109/ICCES.2012.6408524
- Uden, L. (2007). Activity theory for designing mobile learning. *International Journal Mobile Learning and Organisation*, 1(1), 81–102. doi:10.1504/IJMLO.2007.011190
- Yamgata-Lynch, L. C. (2010). *Activity Systems Analysis Methods: Undersatanding Complex Learning Environments*. London: Springer Science. doi:10.1007/978-1-4419-6321-5

ADDITIONAL READING

Anderson, L., & Stillman, J. (2013). Making Learning the Object: Using Cultural Historical Activity Theory to Analyze and Organize Student Teaching in Urban High-Needs Schools. *Teachers College Record*, 115(3), n3.

Arbab, F., de Boer, F., Bonsangue, M., Lankhorst, M., Proper, E., & van der Torre, L. (2015). Integrating Architectural Models-Symbolic, Semantic and Subjective Models in Enterprise Architecture. *Enterprise Modelling and Information Systems Architectures*, 2(1), 40–57.

Cameron, B. H., & McMillan, E. (2013). Analyzing the current trends in enterprise architecture frameworks. *Journal of Enterprise Architecture*, 9(1), 60–71.

Chen, R., Sharman, R., Rao, H. R., & Upadhyaya, S. J. (2013). Data Model Development for Fire Related Extreme Events: An Activity Theory Approach. *Management Information Systems Quarterly*, 37(1), 125–147.

Niemi, E., & Pekkola, S. (2013, January). Enterprise architecture quality attributes: A case study. In *System Sciences (HICSS), 2013 46th Hawaii International Conference on* (pp. 3878-3887). IEEE. doi:10.1109/HICSS.2013.201

Roth, S., Hauder, M., Farwick, M., Breu, R., & Matthes, F. (2013). Enterprise Architecture Documentation: Current Practices and Future Directions. In *Wirtschaftsinformatik* (p. 58).

Shaanika, I., & Iyamu, T. (2015). Deployment of Enterprise Architecture in the Namibian Government: The Use of Activity Theory to Examine the Influencing Factors. *The Electronic Journal of Information Systems in Developing Countries*, 71.

Simon, D., Fischbach, K., & Schoder, D. (2013). An exploration of enterprise architecture research. *Communications of the Association for Information Systems*, 32(1), 1–72.

Yates, A. (2013). *Compulsive exercise and the eating disorders: Toward an integrated theory of activity*. Routledge.

KEY TERMS AND DEFINITIONS

Activity Theory: A socio-technical theory concerned with human activities development and their consciousness.

Deployment: Development and implementation of Information Systems and Technologies IS/IT.

Enterprise: Represents the bigger organisation with all the connections to the external environment which includes business partners, suppliers, debtors, stakeholders and customers.

Enterprise Architecture: Systematic approach for governance of organisational artefacts.

Ministry: A legal entity operating fully under government control and is not profit driven. Used as a case in this study.

ERP Systems Benefit Realization and the Role of ERP-Enabled Application Integration

E

Joseph K. Nwankpa
Miami University, USA

INTRODUCTION

One of the key challenges facing businesses after ERP systems implementation is discerning how to realize pre-implementation benefits. ERP systems are complex software solutions that integrate information and business processes within and across functional areas of business (Davenport, 2000). These systems represent a major departure from the legacy systems and functional information systems that were widespread in the past. Some organizations have successfully implemented and benefited from their ERP system deployments (Nwankpa & Roumani, 2014) and have indeed achieved operational efficiencies and other far-reaching positive changes (Jones et al., 2008), while other organizations are left to grapple with ways to translate pre-implementation expectations into actual post-implementation benefits (Gattiker & Goodhue, 2005). Despite a large body of ERP research literature from a number of different perspectives most published research continues to struggle to adequately explain these mixed results in post-implementation outcomes and benefits (Markus et al., 2000; Nwankpa et al., 2013). As organizations continue to invest in ERP systems, the overarching question for management becomes how they can optimally realize the potential benefits from their ERP system.

ERP systems as platform technologies provide not only a common business process within the organization but also create an integrated platform that permits the adoption and integration of third party non-ERP applications (Liu et al., 2013;

Nwankpa et al., 2013). Thus, organizations with ERP systems can leverage on this information superiority and integrate additional non-ERP applications such as e-commerce applications, customer relationship management (CRM) systems and supply chain planning systems. This level of integration enabled by an ERP system can extend functionalities such as connecting a website to an ERP system as well as advancing information visibility across an organization's value chain (Nwankpa et al., 2013). For instance, integrating CRM and ERP applications can improve operational efficiencies by enabling value-chain processes to adjust promptly to each other (Liu et al., 2013). Firms with CRM applications and ERP systems are able to leverage the CRM applications' ability to extract customer information from multiple customer touch points as well as ERP systems ability to configure product offerings, scheduling, order fulfilment and interdepartmental information exchange (Liu et al., 2013). Given the critical role ERP-enabled third party application integration, studies on ERP benefits that combine ERP-enabled application integration are needed to develop a better grounded theoretical understanding and devise more effective ERP benefit realization practices.

To this end, this study sets out to examine a central research question that has not been adequately investigated in the ERP benefit literature: (i) Is there a positive implication of ERP-enabled application integration on overall ERP benefit? And if yes, what are the antecedents of ERP-enabled application integration? This study at-

tempts to answer this question by conceptually and empirically testing an integrated research model that combines the framework of ERP-enabled application integration and overall ERP benefits with survey data collected from employees in a wide range of United States firms.

The rest of the article is arranged as follows. The next section reviews the extant research on existing ERP system factors, ERP-enabled integration and ERP benefits, and based on this review the study will develop research hypotheses and propose a theoretical model. Next, a description of the research design and data collection, as well as data analyses using structural equation modeling is presented. Next, a discussion and conclusions will be presented.

BACKGROUND

ERP System Benefit

Earlier studies on ERP benefits had mixed results. Poston and Grabski (1986) used archival data to examine a group of firms before and after ERP implementation and found no post-implementation general financial improvement. However, a subsequent study found that the financial performance of non-ERP adopters actually decreased over time when compared to those of ERP adopters that remain unchanged (Hunton et al., 2003). Mabert et al. (2000) examined ERP system benefits in US manufacturing firms and found that although ERP use was pervasive in the US manufacturing sector, the system did not lead to significant operating expense reduction. On the contrary, Shang and Seddon (2002) demonstrated that the implementation of ERP systems resulted in significant operational benefits. They developed five dimensions of ERP benefits namely, operational, managerial, strategic, IT infrastructure and organizational and concluded that ERP benefit realization was in fact a continuous process with specific benefits realized at different rates in different core processes and in different organizations. O'Leary (2004) identified

tangible and intangible ERP benefits and argued that benefits vary across industries, especially intangible benefits. Gattiker and Goodhue (2005) examined ERP benefits by applying a two-phased model that delineated intermediate ERP benefits and overall ERP benefits. They operationalized and measured intermediate benefits as "task efficiency" and "coordination improvement". These benefits, they argue, will occur at the functional and operational levels within the firm but will contribute to the firm's overall ERP benefits (Gattiker & Goodhue, 2005) and concluded that realizing the intermediate benefits is critical to attaining the overall ERP benefits. Similarly, Chou and Chang (2008) decomposed these intermediate benefits in a bid to gain more insight and understanding and found that customization and organizational mechanisms are key drivers of intermediate ERP benefit while reaffirming the role of intermediate benefits as predictors of overall ERP benefit.

ERP-Enabled Application Integration

ERP-enabled application integration refers to the extent of real-time communication between an ERP system and another non-ERP application. Such integration enables an organization to increase the visibility of an ERP system by leveraging the functionalities of third party applications with the cross-functional information exchange and integration platform of an existing ERP system (Nwankpa et al., 2013). Application integration enables organizations to remain agile and responsive to their companies needs and keep pace with technological leaps and innovations. Through application integration, firms can extend and enhance functionalities, exchange information with a firm's existing ERP system and better increase visibility throughout the entire value chain (Rai et al., 2006). The extant literature has rendered some support for ERP-enabled application integration. For instance, firms adopting both an ERP system and a CRM application achieved significant improvement compared to firms adopting only an ERP system (Aral et al., 2006).

RESEARCH MODEL AND HYPOTHESES DEVELOPMENT

Building on the background literature discussed above, the following section presents the research model underlying our study in Figure 1. The reasoning for the key links in the model and the specific hypotheses are discussed below.

Task Efficiency

The degree to which task efficiency is achieved with an existing ERP system remains an important indicator of how well the system aligns with organizational goals and objectives. Thus, if management believes deploying an ERP system has led to significant gains in organizational task efficiency, they may seek to extend such benefits by integrating additional ERP-enabled applications. For instance, Liu et al. (2013) argued that real-time information sharing enabled by ERP systems led organizations to the integration of CRM applications. Thus, organizations are able to better leverage customers' information gathered by the CRM with the existing in-house information enabled by the existing ERP system. Similarly, integrating ERP systems with supply chain applications can improve operational competence by advancing information availability across an organization's value chain (Nwankpa et al., 2013). In sum, task efficiency is expected to influence overall ERP benefit and should be a strong predictor of ERP-enabled application integration as firms will seek to extend the functionalities of the ERP systems if they view the system as leading to improvement in organizational task and responsibilities. Thus, this study proposes the following hypotheses:

- H1:** For an ERP implementing firm, task efficiency will positively affect overall ERP benefit.
- H2:** For an ERP implementing firm, task efficiency will positively affect ERP-enabled application integration.

Current ERP Performance

Prior research on ERP system performance has revealed important findings for ERP implementing firms. Wei (2008) argued that to evaluate the performance of an ERP system, organizations need to first identify key performance indicators and expectations. Current ERP performance has also been linked with application integration and extension of ERP functionalities. Willis and Willis-Brown (2002) noted that the underlying ERP system needs to be stable and operating at an acceptable level of performance for third party application integration to take place. If the core ERP system is underperforming or unstable, firms may be hesitant to integrate and extend applications and functionalities as such applications rely extensively on the ERP system platform. Thus, stabilizing the core ERP system, enables organizations to successfully leverage the common database structure of the system. Such firms can extend the reach of their ERP with add-ons such as SCM, data warehouse applications and web-enabled applications (Nwankpa et al., 2013). Therefore, this study contends that current ERP performance will impact ERP-enabled application integration as organizations will view the existing ERP performance as an indicator of how well other ERP-enabled applications align with the ERP. Thus, the study proposes the following hypothesis:

- H3:** For an ERP implementing firm, current ERP performance will positively affect ERP-enabled application integration.

Extent of ERP Implementation

ERP systems have modular and divisible characteristics that create a variety of options to consider during ERP implementation. The extent of ERP implementation has come under scrutiny as researchers attempt to explain variance in ERP benefit realization (Karimi et al., 2007; Nwankpa et al., 2013). For instance, the examination of

the impact of ERP investment on market value of firms found that firms implementing ERP systems with greater functional, organizational and geographic scope had positive and higher shareholders returns while firms implementing ERP projects with lesser scope generated negative returns (Ranganathan et al., 2006). Similarly, the extent of ERP implementation has been found to influence business process outcomes. Karimi et al. (2007) noted that firms will require a certain level of depth in the initial ERP implementation to create the platform necessary to facilitate higher business process outcomes while prior research found that the extent of ERP implementation created the suitable platform for organizational ease of use of subsequent technologies (Nwankpa et al., 2013). Thus, the extent of ERP implementation this study argues will underpin how add-ons and ERP-enabled applications can be supported and integrated. This leads to the following hypothesis:

H4: For a firm that has implemented an ERP system, extent of ERP implementation will positively affect ERP-enabled application integration.

Coordination Improvement

A number of researchers have identified coordination improvement as an intermediate benefit of an ERP system (Gattiker & Goodhue, 2005; Chou & Chang, 2008). Chou and Chang (2008) found that coordination improvement was a product of organizational machinery and how well the ERP is allied with the organization. Extending the ERP system beyond the core applications can improve coordination and optimize processes for business units that deal directly with customers. For instance, integrating the ERP core with a CRM can facilitate coordination at the customer level and at the firm level as each application automatically notifies the other of customer related business processes such as customer requests, order fulfilments production scheduling and accounts and payments (Liu et al., 2013). In summary,

coordination improvement can be a predicator of ERP-enabled application integration and an indicator of how well an organization will achieve overall ERP benefit. This leads to the following hypotheses:

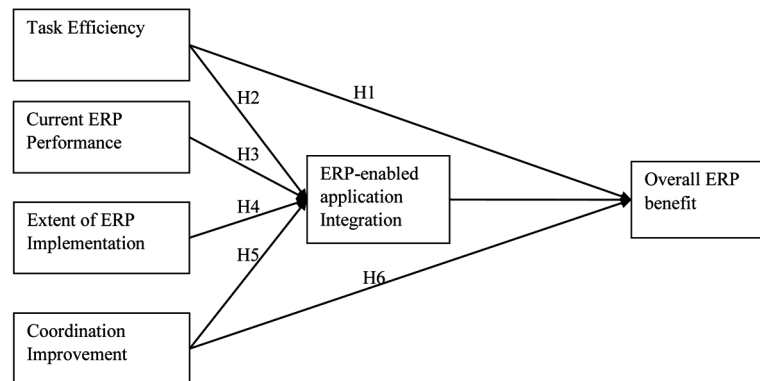
H5: For an ERP implementing firm, coordination improvement will positively affect ERP-enabled application integration.

H6: For an ERP implementing firm, coordination improvement will positively affect overall ERP benefit.

Impact of ERP-Enabled Application Integration on Overall ERP Benefit

Current literature suggests that platform technologies such as an ERP system enables the adoption and integration of subsequent applications and add-ons (Karimi et al., 2007; Nwankpa et al., 2013). These additional applications allow firms to extend functionalities and remain agile, strategic and responsive to changing technological opportunities. Willis and Willis-Brown (2002) argued that for organizations to realized full ERP benefits, they need to view the initial ERP implementation as a foundation laying exercise that creates the opportunity for organization to integration emerging technologies. Such a view was reinforced by Bread and Sumner (2004) who noted that the integration of CRM and e-commerce capabilities are valuable to firms due to the enabling base technology and backbone support provided by the ERP systems. Firms who do not extend the functionalities of their ERP systems or integrate their ERP systems with additional applications may not lay claims to the same overall ERP benefits (Nwankpa et al., 2013; Liu et al., 2013). These system add-ons provide very specific functionalities capable of creating a wide range of different capabilities that firms can attach to their ERP to meet and respond to emerging business challenges (Goodhue et al., 2009). Thus, this paper argues that in a post-implementation phase, ERP-enabled application integration will enable firms to achieve higher overall ERP ben-

Figure 1. Research model



efits by providing add-on solutions and specific non-critical functionalities yet necessary to meet emerging organizational needs. This leads to the following hypothesis:

H7: For an ERP implementing firm, ERP-enabled application integration will positively affect overall ERP benefit.

RESEARCH METHODS

Sample

The firm is the unit of analysis in this study as such the subjects are IT decision makers within the organization such as, the chief information officer (CIO), chief technology officer (CTO) or vice president of IT operations. These executives are typically tasked with the decisions making processes of their organizations technology needs. Dun and Bradstreet's Million Dollar database – a directory of executives was used to identify a random sample of IT executives. This database provides contact information of executives in various positions in firms in the United States and was used to identify a random sample of information systems executives holding IT positions in companies the United States. The survey instrument was developed and adapted after a thorough literature review and interviews with some IT managers. Questionnaires were mailed to

1250 corporations in the United States that were selected randomly from Dun and Bradstreet's Million Dollar database. A total of 287 responses were returned in the first mailing. Phone calls were made to about 550 randomly selected non-respondents and where possible voice mail messages were left explaining the nature of the research and requesting the contacts participation. An additional 58 usable surveys were received from a second mail providing a total of 345 responses. Out of 345 responses 317 were usable, resulting in an actual response rate of 25.3%. Table 1 provides sample characteristics.

Whenever possible, this study used previously validated measures and adapted them in the context of ERP system benefit and ERP-enabled application integration. In the absence of existing scales, theoretically grounded new scales were developed. The items for measuring the intermediate ERP benefits namely task efficiency and coordination improvements were adapted from Chou and Chang (2008) to fit the context of ERP-enabled application integration. Similarly, the items for measuring current system performance were adapted from Nwankpa et al. (2013) while the items measuring extent of ERP implementation were adapted from Karimi et al. (2007). Conversely, the three items measuring overall ERP benefit were adapted from Chou and Chang (2008). In the absence of an existing scale measuring ERP-enabled application integration, these measurement items were developed based on prior literature on ERP-

Table 1. Sample characteristics

Classification	Number of Respondents	(%) Respondents
Firm's Size (Number of Employees)		
less than 1000	107	33%
1001 - 5000	116	37%
Above 5000	94	30%
Industry		
Construction	45	14%
Utilities	39	12%
Manufacturing	73	23%
Service	64	20%
Telecommunication	23	7%
Wholesale and Retail	59	19%
Other	14	4%
Time Elapsed Since ERP Implementation		
1 - 3 years	75	23%
3 - 6 years	81	26%
6 - 10 years	85	27%
More than 10 years	76	24%
Job Title of Respondents		
CIO	179	57%
CTO	69	22%
Vice president of IT Operations	24	7%
IT Director	45	14%
Average number of years in position	5.1 years	

enabled adoption (Nwankpa et al., 2013; Liu et al., 2013). These items were measured with multiple indicators coded on a seven-point Likert scale (See Appendix). In addition to the focal constructs, the study measured demographic characteristics of each respondent and the time elapsed since the ERP implementation.

Construct validity can be assessed using convergent validity and discriminant validity.

Convergent validity was tested using two criteria (Fornell & Larcker, 1981). First, all indicator loadings should be significant and exceed 0.7 and second, the average variance extracted (AVE) by each construct should exceed the variance due

to the measurement error for that construct. As shown in Table 2, all of the items exhibit a loading higher than 0.7 on their respective construct, and as shown in Table 3 were greater than 0.70, thus satisfying both conditions for convergent validity. The variance inflation factor (VIF) examined the existence of excessive multicollinearity among latent constructs. All VIF measures were below the 5.0 recommended minimum (Kline 1998) and the highest VIF value was 2.76. Discriminant validity was assessed using three tests. First, an examination of cross factor loadings (Table 2) indicates good discriminant validity, because the loading of each item on its assigned construct is greater than its loadings on all other constructs (Chin, 1998). Second, the correlations among the constructs are below the 0.85 threshold (Kline,

Table 2. Loading and cross-loadings

	TE	CEP	EEI	CI	EAI	OEB
TE1	0.923	0.344	0.344	0.325	0.329	0.451
TE2	0.902	0.322	0.329	0.375	0.336	0.455
TE3	0.911	0.371	0.361	0.364	0.367	0.423
CEP1	0.233	0.901	0.287	0.465	0.313	0.297
CEP2	0.217	0.918	0.289	0.433	0.317	0.385
CEP3	0.246	0.916	0.256	0.454	0.319	0.306
EEI1	0.344	0.434	0.925	0.398	0.433	0.374
EEI2	0.327	0.398	0.936	0.432	0.437	0.379
EEI3	0.389	0.362	0.945	0.412	0.372	0.367
CI1	0.415	0.412	0.217	0.942	0.377	0.211
CI2	0.396	0.431	0.239	0.915	0.321	0.234
CI3	0.357	0.412	0.243	0.909	0.233	0.228
EAI1	0.402	0.451	0.432	0.337	0.951	0.407
EAI2	0.412	0.475	0.398	0.376	0.943	0.414
EAI3	0.397	0.481	0.404	0.388	0.939	0.376
EAI4	0.441	0.426	0.432	0.391	0.947	0.332
OEB1	0.343	0.412	0.327	0.456	0.356	0.939
OEB2	0.386	0.471	0.323	0.417	0.371	0.931
OEB3	0.457	0.391	0.378	0.423	0.377	0.918

TE: Task Efficiency, CEP: Current ERP Performance, EEI: Extent of ERP Implementation

CI: Coordination Improvement, EAI: ERP-enabled Application Integration, OEB: Overall ERP Benefit.

Table 3. Construct correlations, consistency and reliability

	AVE	CR	α	TE	CEP	EEI	CI	EAI	OEB
TE	0.92	0.97	0.94	0.96					
CEP	0.91	0.98	0.96	0.38	0.95				
EEI	0.89	0.95	0.92	0.41	0.49	0.94			
CI	0.92	0.98	0.94	0.39	0.42	0.41	0.96		
EAI	0.89	0.95	0.93	0.43	0.39	0.43	0.44	0.94	
OEB	0.91	0.98	0.93	0.35	0.47	0.42	0.39	0.39	0.95

1998), suggesting discriminant validity. Third, the square root of the AVE from a construct is greater than the correlations among the construct and all other constructs in the model (Table 3) (Fornell & Larcker, 1981).

DATA ANALYSIS AND RESULT

Data analysis and empirical validation of our hypotheses were done with partial least square (PLS) analysis. PLS was selected because it enabled specification and testing of path models with latent constructs and because it places minimal restriction on the measurement scales, sample size and residual distribution (Chin & Newsted, 1999). Further, PLS allows latent constructs to be modeled as formative indicators as was the case with our data. SmartPLS 2.0 software was used for the analysis. SmartPLS 2.0 performs bootstrapping analysis to assess the statistical significance of

the loading and path coefficients (Ringle et al., 2005). Consistent with prior research using PLS models, this study analyzed the research model in two stages (Hulland, 1999; Chin, 1998). The first stage involved the assessment of the reliability and validity of the measurement model and the second stage involved the assessment of the structural model (Hulland, 1999).

Structural Model

In PLS analysis, examining the structural paths and the R-square scores of the endogenous variables assesses the explanatory power of the structural model. Figure 2 shows the results of the structural path analysis. Overall, all of the 7 paths are significant, with a p-value of less than

0.05. Hypothesis 1 received significant support. Results show that for an ERP implementing firm, task efficiency will positively affect overall ERP benefit ($\beta = 0.29$, $p < 0.05$). As indicated in

Figure 2. Research model with result

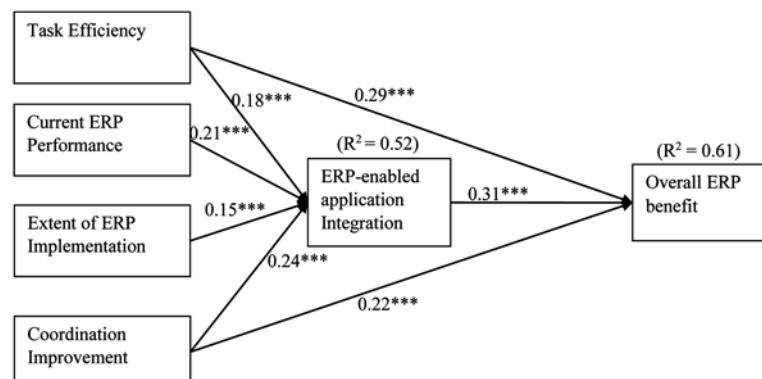


Figure 2, the influence of task efficiency, current ERP Performance, extent of ERP Implementation and coordination improvement on ERP-enabled application integration accounted for 52 percent variance of ERP-enabled application integration. Thus, providing support for Hypothesis 2 ($\beta = 0.18$, $p < 0.01$), Hypothesis 3 ($\beta = 0.21$, $p < 0.01$), Hypothesis 4 ($\beta = 0.15$, $p < 0.01$) and Hypothesis 5 ($\beta = 0.24$, $p < 0.01$). Similarly, Hypothesis 6 results indicate that coordination improvement will positively influence overall ERP benefit ($\beta = 0.22$, $p < 0.01$) while Hypothesis 7 received significant support with result showing that ERP-enabled application integration positively influenced overall ERP benefit ($\beta = 0.31$, $p < 0.01$). The model explained 61% of the variance in overall ERP benefit.

FUTURE RESEARCH DIRECTIONS

Given that respondents came from a single country, ERP benefits, national and cultural particularities need to be analyzed in future research. One suggestion, will be introducing organizational culture as a moderator in the influence of ERP benefits. Second, this study adopts a cross-sectional view in measuring the constructs. Thus, such a design may not adequately capture the interactions among the constructs and cannot establish causality. Future research might find it useful to measure these variables from multiple points in time. A more rigorous test of our model will involve a longitudinal study. Such a study may enrich the findings of our results as well as establish the causality of argument. Nonetheless, the theory in this study suggest that the relationships tested in the research model are causal in nature.

CONCLUSION

As firms continue to invest in ERP systems to integrate information and streamline business process across functional areas, it has become increasingly critical to understand how expected ERP

benefits can be attained. Drawing upon contingent resource-based theory framework and grounding our research in practitioner interviews, this study develops a more comprehensive model, which examines both antecedents and consequences of ERP-enabled application integration and extends the nomological net of constructs that examine overall ERP benefits. The results from the empirical evidence show that task efficiency, current ERP performance, extent of ERP implementation and coordination improvement positively influence ERP-enabled application integration. The study also finds that task efficiency, coordination improvement and ERP-enabled application integration are key determinants of overall ERP benefits. These results contribute to a stream of research that examines the overall ERP benefit at the post-implementation stage.

REFERENCES

- Aral, S., Brynjolfsson, E., & Wu, D. J. (2006). Which came first, it or productivity? The virtuous cycle of investment and use in enterprise systems. *The 27th International Conference on Information Systems (ICIS)*, 1819-1839.
- Beard, J. M., & Sumner, M. (2004). Seeking strategic advantage in the post-net era: Viewing ERP systems from the resource-based perspective. *The Journal of Strategic Information Systems*, 13(2), 129-150. doi:10.1016/j.jsis.2004.02.003
- Chin, W., & Newsted, P. (1999). Structural Equation Modeling Analysis with Small Samples Using Partial Least Squares. In *Statistical Strategies for Small Sample Research*. Sage Publications.
- Chin, W. W. (1998). Issues and Opinion on Structural Equation Modeling. *Management Information Systems Quarterly*, 22(1), vii-xvi.
- Chou, S.-W., & Chang, Y.-C. (2008). The implementation factors that influence the ERP (enterprise resource planning) benefits. *Decision Support Systems*, 46(1), 149-157. doi:10.1016/j.dss.2008.06.003

- Davenport, T. H. (2000). *Mission Critical: Realizing the Promise of Enterprise Systems*. Boston, MA: Harvard Business School Press.
- Fichman, R. G. (2004). Going Beyond the Dominant Paradigm for Information Technology Innovation Research: Emerging Concepts and Methods. *Journal of the Association for Information Systems*, 8(5), 314–355.
- Fornell, C., & Larcker, D. F. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *JMR, Journal of Marketing Research*, 18(1), 39–50. doi:10.2307/3151312
- Gattiker, T. F., & Goodhue, D. L. (2005). What happens after ERP implementation: Understanding the impact of inter-dependence and differentiation on plant-level outcomes. *Management Information Systems Quarterly*, 29(3), 559–585.
- Goodhue, D. L., Chen, D. Q., Boudreau, M. C., Davis, A., & Cochran, J. D. (2009). Addressing Business Agility Challenges with Enterprise Systems. *MIS Quarterly Executive*, 8(2), 73–87.
- Hulland, J. (1999). Use of partial least squares (PLS) in strategic management research: A review of four recent studies. *Strategic Management Journal*, 20(2), 195–204. doi:10.1002/(SICI)1097-0266(199902)20:2<195::AID-SMJ13>3.0.CO;2-7
- Hunton, J. E., Lippincott, B., & Reck, J. L. (2003). Enterprise resource planning (ERP) systems: Comparing firm performance of adopters and non-adopters. *International Journal of Accounting Information Systems*, 4(3), 165–184. doi:10.1016/S1467-0895(03)00008-3
- Jones, M. C., Zmud, R. W., & Thomas, D. C. (2008). ERP in Practice: A Snapshot of Post-Installation Perception and Behaviors. *Communications of the Association for Information Systems*, 1(2), 437–462.
- Karimi, J., Somers, T. M., & Bhattacharjee, A. (2007). The Impact of ERP Implementation on Business Process Outcomes: A Factor-Based Study. *Journal of Management Information Systems*, 24(1), 101–134. doi:10.2753/MIS0742-1222240103
- Kline, P. (1998). *Principles and practice of structural equation modelling*. New York: The Guildford Press.
- Liu, A. Z., Liu, H., & Xu, S. X. (2013). How do competitive environment moderate CRM value? *Decision Support Systems*, 56, 462–473. doi:10.1016/j.dss.2012.11.003
- Markus, M. L., Axline, S., Petrie, D., & Tanis, C. (2000). Learning from Adopters Experience with ERP – Successes and Problems. *Journal of Information Technology*, 15, 245–265. doi:10.1080/02683960010008944
- Nwankpa, J. K., & Roumani, Y. (2014). Understanding the link between organizational learning capability and ERP system usage: An empirical examination. *Computers in Human Behavior*, 33, 224–234. doi:10.1016/j.chb.2014.01.030
- Nwankpa, J. K., Roumani, Y., Brandyberry, A., Guffrida, A., Hu, M., & Shanker, M. (2013). Understanding the Link between Initial ERP System and ERP-Enabled Adoption. *Information Resources Management Journal*, 4(26), 18–39. doi:10.4018/irmj.2013100102
- OLeary, D. E. (2004). Enterprise resource planning systems: An empirical analysis of benefits. *Journal of Emerging Technologies in Accounting*, 1(1), 63–72. doi:10.2308/jeta.2004.1.1.63
- Rai, A., Patnayakuni, R., & Seth, N. (2006). Firm Performance impacts of digitally enabled supply chain integration capabilities. *Management Information Systems Quarterly*, 30(2), 225–246.

Ranganathan, C., & Brown, C. V. (2006). ERP investments and the market value of firms: Toward an understanding of influential ERP project variables. *Information Systems Research*, 17(2), 145–161. doi:10.1287/isre.1060.0084

Ringle, C. M., Wende, S., & Will, A. (2005). SmartPLS 2.0.M3 (beta). Hamburg, Germany: University of Hamburg. Available at www.smartpls.de

Shang, S., & Seddon, P. B. (2002). Assessing and managing the benefits of enterprise systems: The business managers perspective. *Information Systems Journal*, 12(4), 271–299. doi:10.1046/j.1365-2575.2002.00132.x

Wei, C. C. (2008). Evaluating the performance of an ERP system based on the knowledge of ERP implementation objectives. *International Journal of Advanced Manufacturing Technology*, 39(1-2), 168–181. doi:10.1007/s00170-007-1189-3

Willis, T. H., & Willis-Brown, A. H. (2002). Extending the value of ERP. *Industrial Management & Data Systems*, 102(1), 35–38. doi:10.1108/02635570210414640

KEY TERMS AND DEFINITIONS

Coordination Improvement: The degree of real-time communication and information flow among cross-functional units within an organization.

Current ERP Performance: The degree to which the existing ERP system is able to meet the requirements of the current business functions and activities as well as anticipated future requirements.

ERP Benefit: The degree to which an ERP system improves organizational performance and objectives.

ERP Systems: Information system packages that integrate information and business processes within and across functional areas of business.

ERP-Enable Application Integration: The extent of real-time communication between an ERP system and another non-ERP application.

Extent of ERP Implementation: The degree and scale of initial ERP implementation and deployment.

Task Efficiency: The ability to achieve organizational task with minimal resources.

APPENDIX: MEASUREMENT ITEMS

Task Efficiency

1. Due to the ERP implementation, employees such as buyers, planners, and production supervisors need less time to do their job (Based on Chou & Chang, 2008).
2. ERP saves time in jobs like production, material planning and production management (Based on Chou & Chang, 2008).
3. ERP helps employees like buyers, planners, and production supervisors to be more productive (Based on Chou & Chang, 2008).

Current ERP Performance

At the time of integrating an ERP-enabled application:

1. Our ERP system performance was at a level that was considered successful (Based on Nwankpa et al., 2013).
2. Our ERP system was performing to our satisfaction (Based on Nwankpa et al., 2013).
3. Our ERP system was a success in terms of performance (Based on Kull & Narasimhan, 2010; Nwankpa et al., 2013).

Extent of ERP Implementation

1. Functional scope of implementation of your selected ERP System: Accounting/Finance, Manufacturing, Planning/Scheduling, Human Resources, Sales/Distribution, Logistics/Inventory Control, Other (please specify) (Based on Karimi et al., 2007).
2. Scope of implementation of your selected ERP: Department/Division, Multiple departments/divisions, Entire company, Multiple companies, Other (please specify) (Based on Karimi et al., 2007).
3. Geographical extent of implementation: Single site, Multiple sites, National, Global (Based on Karimi et al., 2007).

Coordination Improvement

1. ERP helps to adjust to changing conditions among different units of the firm (Based on Chou & Chang, 2008).
2. ERP has improved the coordination among different units of the firm (Based on Chou & Chang, 2008).
3. ERP facilitates the integration of important information among different units of the firm (Based on Chou & Chang, 2008).
4. ERP helps to synchronize among different units of the firm (Based on Chou & Chang, 2008).

ERP-Enabled Application Integration

At the time of integrating an ERP-enabled application:

1. Our ERP system enabled our organization to share data, communicate and implement other functional specific applications (Developed).
2. Our ERP system lead to the integration and exchange of data with other applications with specific functionalities (Developed).
3. Our ERP system enabled our firm to implement applications that go beyond the boundaries of the ERP system (Developed).
4. Our ERP system created the platform needed to integrate additional applications (Developed).

Overall ERP Benefit

1. In terms of ERP's business impacts on the organization, the ERP system has been a success (Based on Chou & Chang, 2008).
2. ERP has seriously improved the organization's overall business performance (Based on Chou & Chang, 2008).
3. ERP has had a significant positive effect on this organization (Based on Chou & Chang, 2008).

From On–Premise ERP to Cloud ERP

E**Karim Mezghani***Al Imam Mohammad Ibn Saud Islamic University (IMSIU), Saudi Arabia & University of Sfax, Tunisia*

INTRODUCTION

With the development of cloud computing, a cloud-based ERP begins to emerge as an alternative to the on-Premise solution. According to Grabski, Leech and Schmidt (2011), “Cloud computing has the potential to radically change the ERP environment. The data and the application are no longer housed on-Premise; rather, a vendor provides access to the application which can be customized to meet the user’s needs and the vendor also hosts the data securely somewhere on the Internet...Many research questions surround this evolutionary approach to ERP systems”. Arnesen (2013) adds that ERP vendors “are in the process of developing hosted or cloud solutions as the market moves to a cloud environment”. Thus, according to Mezghani (2014), “cloud ERP seems to become a real substitute to on-Premise ERP and firms would be likely “pushed” to switch toward the cloud solution”.

One of the concerns linked to ERP implementation is alignment. Indeed, ERP systems affect almost all business processes and even some strategic choices of a firm. That is why many researches focus on alignment concerns when studying ERP implementation (Hong and Kim, 2002; Yaseen, 2009; Mezghani and Mezghani, 2014). Nevertheless, as reported above, firms are more and more interested in switching toward cloud ERP (Mezghani, 2014). By the way, “their business strategy, IT strategy, business processes and information technologies shall be re-aligned” (Li, Wang, Wu, Li and Wang, 2011). Indeed, in a cloud computing environment, misalignments can lead to decreased operating efficiency and losses

for organizations (Géczy, Izumi and Hasida, 2012). However, “no cloud computing service vendor can satisfy the complete functional information system requirements of an enterprise” (Li et al., 2011). These authors add that “sometimes, enterprises have to simultaneously use software services distributed in different clouds in conjunction with their intra-IS. These bring great challenges for business–IT alignment of an enterprise in the cloud computing environment”. Also, with cloud ERP, data and applications are provider-hosted. So, the way ERP is managed may be redefined and, then, achieving alignment might be more challenging.

Thus, this chapter aims to expose the major challenges and issues linked to ERP systems alignment. Some tendencies and best practices are also proposed for firms to overcome the alignment challenges, mainly in a cloud computing environment.

BACKGROUND: ON-PREMISE ERP AND ALIGNMENT CONCERNS

Importance of Alignment When Implementing On-Premise ERP

Since the development of the Strategic Alignment model (SAM) by Henderson and Venkatraman (1993), many definitions were proposed for IT-business alignment. Based on previous studies, mainly those of Reich and Benbasat (2000), Mezghani and Mezghani (2014) proposed a definition for the ERP strategic alignment. Thus, these authors consider that such alignment exists when:

- “The ERP implementation integrates business strategy (in terms of strategic choices).
- The business strategy considers the ERP characteristics (mainly the benefits and limitations).

This alignment is defined as a pattern of co-variation due to the specificities of ERP systems that are standard and adaptable at the same time”.

Keeping ERP implementation aligned with business is an important matter because:

- **Getting ERP Benefits Depends on Alignment:** Many firms do not gain ERP benefits because they do not consider such systems when making strategic choices (Yaseen, 2009). Indeed, since ERP systems integrate all business processes, they can be considered as useful tools to access and analyze several data. Also, as ERP systems are developed based on *best practices*, firms can gain such practices if they align their business processes with the ERP ones.
- **ERP Modules Are Standard:** As mentioned below, adapting the business processes according to the ERP ones can be beneficial to firms. However, such alignment can be risky since it can affect key processes or generate resistance to change (Davenport, 1998). So, firms can choose to adapt ERP modules rather than business processes. Although useful, this alternative is also risky because it can lead to technical bugs. In all cases, project team should manage both alternatives carefully to keep alignment with low risk.
- **ERP Implementation Is a Strategic Investment:** ERP project is a costly project, time consumer and affects all business processes, so top management should consider ERP implementation as a strategic investment and priority (Grant, 2003). Besides, Lee and Myers (2004) affirm that the strategic context may change during an ERP project. Then, it is crucial to follow up continuously

the alignment state and make the required adjustments unless ERP project would face delays (Lee & Myers, 2004).

Although the importance of alignment in ERP implementation, achieving such alignment is a true challenge. That is why, several studies tried to provide ideas and practices that can be useful to realize this alignment.

Social Dimension vs. Technical Dimension for ERP Alignment

When talking about successful ERP implementation, it is necessary to focus on factors that contribute to alignment (Kidd, 2011). To achieve such alignment, the project team should focus on both ERP and business processes adaptation (Hong & Kim, 2002; Mezghani & Mezghani, 2007). However, both alternatives are risky.

When choosing a technical perspective (ERP adaptation), it is important to note that, although ERP systems are adaptable, making a deep customization can cause technical bugs and affect the integrated *best practices* (Mezghani & Mezghani, 2007). In fact, a technology-oriented project can be a key reason of failure in ERP implementation (Kidd, 2011). Indeed, ERP systems cover all business processes and are developed based on *best practices* of business management, so they should be managed as organizational projects that involve all people over the firms (Mezghani & Mezghani, 2007; Kasemsap, 2015). Then, many authors suggest focusing on business processes adaptation. Indeed, implementing an ERP can be integrated as part of a BPR (business process re-engineering) project that aims to improve business processes and align with ERP ones. Nevertheless, this alternative is very risky as it can affect the key processes and then some competitive advantages (Davenport, 1998). Employees may also resist to such adaptations mainly when organizational change is not well managed (Kasemsap, 2015)

Regarding benefits and risks of both alternatives, it is argued that alignment should be achieved

by well-defined technical and organizational adaptations with a particular attention to social concerns (Mezghani & Mezghani, 2014).

Indeed, according to the alignment duality approach developed by Reich and Benbasat (1996, 2000), many technical and intellectual tools could be used when making efforts to reach alignment. But such alignment cannot be concretized in absence of the social dimension that characterizes the interactions between IT specialists and business managers. These interactions involve communication, training, top management support, users' involvement and other social practices which let managers understand the business and technical requirements of ERP projects so adaptations could be deployed carefully (Mezghani & Mezghani, 2014). Such practices refer to management of changes (technological and organizational ones) in order to facilitate implementation and get expected benefits of ERP.

The importance of the social dimension can also be revealed when talking about skills re-

quired to ERP implementation and alignment. Indeed, in addition to IT skills, it is argued that interpersonal skills are important antecedents of alignment as they favor interactions between IT and business managers (Mezghani and Mezghani, 2014). In sum, one can understand the effects of the social dimension on ERP alignment through its components as shown in Table 1.

So, it appears that social dimension should be considered in addition to the technical one in order to achieve the required alignment when reducing the risks linked to system and processes adaptations. However, when deploying an ERP on a cloud, the required adaptations may change and then the way to consider alignment may be viewed differently.

Table 1. Social dimension and ERP alignment

Social Dimension Components	Effects on ERP Alignment	Authors
Top management support	As ERP projects involve many technical and organizational changes, the team project need to be continually supported by providing required financial and human resources to manage such changes (reduce change resistance, deploy costly customizations,...). Top managers who support ERP projects are more likely to consider such systems when making strategic choices.	(Yaseen, 2009; Silvius, De Haes and Van Grembergen, 2010; Kufandirimbwa, Hapanyengwi and Kabanda, 2012; Mezghani and Mezghani, 2014; Shatat, 2015)
Communication	Refers to information sharing among managers. It is considered as a key factor in ERP projects. Communication can promote mutual understanding between IT specialists and business ones. This helps to well adapt both ERP systems and business processes so better alignment is achieved.	(Reich and Benbasat, 2000; Kidd, 2011; Mezghani and Mezghani, 2014; Kasemsap, 2015)
Users' involvement	Involving potential users in ERP implementation helps to better identify their needs for closer alignment. Also, such involvement stimulate ERP acceptance so change resistance is reduced.	(Mezghani and Mezghani, 2007; Kidd, 2011; Kasemsap, 2015)
Training	When firms invest in training to implement an ERP system, this involves they are giving a strategic dimension to the ERP project. Users' training helps also to stimulate ERP acceptance et to prepare them for the "new" business processes (in case of processes redesign).	(Mezghani and Mezghani, 2007; Kidd, 2011; Kasemsap, 2015)
Interpersonal skills	Since ERP projects require change management practices that involve many social aspects (communication, support, involvement,...), business and IT managers need to develop interpersonal skills to well achieve such practices and to favor mutual understanding.	(Mohamed and McLaren, 2009; Mezghani and Mezghani, 2014)

ERP ALIGNMENT IN A CLOUD COMPUTING ENVIRONMENT

Cloud ERP: Definition and Overview

Cloud ERP involves using ERP on a cloud so ERP modules are got as “software as a service” (Saas). Considered as a real substitute to on-Premise ERP, “Cloud-based ERP systems are projected to cannibalize approximately 18% of the revenues of the traditional ERP market by 2016” (Das and Dayal, 2016). According to these authors, the total cost of a cloud ERP ownership can be 30%–60% less than that of an on-premise one over a 10-year period. A recent study conducted by Panorama Consulting (2016) indicates that 56% “of the respondents deploying cloud technology reported implementation cost savings of at least 20-percent”.

In fact, cloud ERP systems propose a set of benefits that make it interesting to switch toward them. Nevertheless, as cloud-based solutions, they are associated with a wide range of risks that may make managers hesitant to adopt them. From previous studies comparing cloud ERP to On-Premises ERP, Mezghani (2014) presented the main benefits and risks linked to cloud ERP implementation (Table 2).

According to Mezghani (2016), it is necessary to achieve cloud ERP alignment in order to get linked benefits. Indeed, in a cloud computing environment, misalignments can lead to decreased

operating efficiency and losses for organizations (Géczy et al., 2012). Moreover, since cloud ERP deployment involves a third party (the provider), achieving alignment becomes a real challenge.

Cloud ERP Deployment and the Alignment Challenge

As a new topic, cloud ERP alignment did not yet involve considerable studies. However, the analysis of recent studies on cloud ERP deployment reveals that such alignment is a real concern that needs to be considered and studied deeply.

When examining the determinants of cloud ERP selection and adoption, Das and Dayal (2016) developed a research model to identify the main determinants. One of the prominent conclusions made by these authors after doing empirical investigations is that “the client should examine the organizational fit of the cloud ERP in terms of technical compatibility and strategic value. If there is a mismatch, the organization will not benefit from the investment in the long run”. The authors admit that alignment is crucial for the success of cloud ERP deployment. Moreover, Das and Dayal (2016) think that firms should maintain their in-house systems if they deal with sensitive business processes or are unable to perform customizations.

In a recent exploratory study aiming at examining the alignment challenge within a firm involved in a perspective of switching toward cloud ERP,

Table 2. Benefits and risks of cloud ERP

Cloud ERP Benefits	Cloud ERP Risks
• Cloud ERP Is Cost Saver: Moving to cloud ERP reduces costs linked to licensing, maintenance and upgrading (Makkar and Meenakshi, 2012). • Cloud ERP Is Time Saver: Less time is spent in monitoring IT infrastructure. Besides, tests and trainings are conducted from the first steps of the project which can help to save time (Elragal and Elkommos, 2012). • High Scalability: According to Arnesen (2013), “you can add or reduce users as your needs change, which works especially well for seasonal businesses or companies on a high-growth path”. • Accessibility: “As a cloud based solution, cloud ERP modules can be accessed anytime and anywhere through Internet” (Mezghani, 2014).	• Security: Regarding cloud-based solutions, many managers perceive that the confidentiality and security of business data are not guaranteed (Benlian and Hess, 2011). • Internet Connectivity: “If you lose connectivity because of a natural disaster or cyber-attack or if the vendor goes out of business, you may have no access to the system or data” (Arnesen, 2013). • Dependency: “Even if choosing On-Premises ERP, firms risk to become dependent on vendors since such systems need regular maintenances and updates. This dependency seems to be greater with cloud based ERP” (Mezghani, 2014). Indeed, according to Arnesen (2013), cloud ERP vendors retain the data and firms have to pay vendors on time or they may lose data access.

Mezghani (2016) revealed a set of concerns linked to alignment.

First, the author noted that cloud ERP adoption involves a “double challenge” since adopters should care about the cloud aspect and the ERP aspect. Indeed, it is largely admitted that ERP implementation is a difficult matter. Thus, when associated to the cloud computing, such implementation becomes more challenging. By the way, cloud ERP implementation should be planned strategically in a business-IT alignment perspective (Mezghani, 2016). Moreover, “cloud computing as a new technology is involved in IT strategy. When planning IT implementation, the managers of an enterprise take cloud service as another resource apart from intra-IS. Intuitively, it enlarges the technological scope and extends the scope of IT governance. With high efficiency and flexibility, it also enhances the system competencies” (Li et al., 2011). These authors add that, from a SAM perspective, this change would have impacts on business strategy as well as business processes. According to Géczy et al. (2012), “it is important to align the organization’s functional and operating model with the cloud-based model of utilization of information technology resources and services. Misalignments lead to decreased operating efficiency and losses for organizations”.

Second, Mezghani (2016) revealed that cloud ERP adoption is also linked to technical concerns as integration with current applications and customizations. Such concerns involve alignment efforts in terms of technical fit. Indeed, “without possibilities of customization, users and organizations may encounter substantial re-training and adjustment costs” (Géczy et al., 2012). However, as customers neither own the infrastructure nor run the applications, cloud ERP may fail to give customer greater sense of ownership to make the required changes (Salleh, Teoh and Chan, 2012). Besides, cloud service provider has its general operational processes that could be adopted in most cases and provided to as many subscribers as possible but would not fit all individual requirements (Li et al., 2011). Arnesen (2013) adds that

“though configuration of cloud ERP is available to all customers, major customizations usually aren’t allowed so the vendor can maintain the upgrade path”. According to Peng (2015), “an integrated solution from one single ERP vendor may often not satisfy all business needs of companies... different cloud vendors may use very different technologies and platforms to deploy and host their ERP packages. From a technical perspective, cloud vendors may not be able or willing to customize their ERP system to allow seamless integration between the package and other cloud or on-premise applications”. Thus, it is difficult to customize a cloud ERP mainly when its components are purchased from various providers. Moreover, it is difficult to change cloud vendors even in the case of dissatisfaction (Dutta, Peng and Choudhary, 2013). Then, a weak level of alignment would favor the dependency concern considered as a crucial issue. In fact, a survey research conducted by Dutta et al. (2013) among highly experienced IT professionals involved in developing and implementing cloud based solutions showed that most of the “top ten risks” are linked to the dependency concern (difficulty to change cloud provider, privacy of data, unauthorized access to enterprise data,...). This implies that misalignment in cloud ERP deployment may not only create customization concerns but also generate dependency-linked problems. The importance of such issue is noted by Das and Dayal (2016) who stated that clients should assess the technical compatibility of cloud ERP before adoption unless they would fail in their investment.

Finally, as an enterprise system, cloud ERP requires changes within business processes. This means that firms would perform alignment efforts according to an external unit (the provider) (Mezghani, 2016). This would increase the cloud ERP risks mainly those related to privacy and provider dependency. Moreover, Das and Dayal (2016) recommend maintaining an on-Premise ERP if sensitive business processes are to be involved in a cloud ERP adoption. Also, processes adaptations may require users training to the

new practices. Dutta et al. (2013) note here that inadequate users' training on cloud services and usage could lead to resistance to technical and organizational changes.

All such ideas fit with those advanced by Li et al. (2011) who affirmed that cloud-based solutions alignment is "more complex than that of past cases". Such complexity becomes greater when cloud is not only linked to some functional software but also associated to a wide system as ERP.

SOCIAL AND TECHNICAL DIMENSIONS FOR CLOUD ERP ALIGNMENT: FUTURE RESEARCH DIRECTIONS

Regarding the alignment challenge linked to cloud ERP adoption, researchers and practitioners need to develop tools and frameworks focusing on meeting such challenge. A dual perspective based on social and technical dimensions can be proposed as an alternative for further studies.

On the one hand, integration with the existing IT architecture and customization at several levels are crucial to meet the changing demands of users and the network availability (Géczy et al., 2012). Li et al. (2011) propose a structured approach that involves mapping the processes in order to identify what they call "collaboration points" to facilitate integration of business processes into the cloud.

On the other hand, even if these adaptations may be rapidly run when adopting a cloud-based ERP, users' acceptance of such changes is not guaranteed (Salleh et al., 2012). Also, firms, especially small ones, would not have enough bargaining power to let the cloud providers customize the processes to meet one's specific requirements (Li et al., 2011). By the way, the social dimension appears as a facilitator factor as for on-Premises ERP projects mainly when cloud ERP deployment, compared to in-house one, involves a third party. In fact, according to Peng (2015), potential failure of cloud ERP should be attributed to human problems.

The possible effects of the social dimension are summarized in Table 3.

Table 3. Social dimension and cloud ERP alignment

Social Dimension Components	Effects on Cloud ERP Alignment	Authors
Top management support	"From a social perspective, prioritization of cloud ERP adoption means that such adoption is considered at the strategic level (strategic fit). In fact, given the security concerns linked to cloud computing and the possibility to integrate modules from multiple providers, cloud ERP adoption needs to be managed strategically. From an operational level, prioritization means that top management allocates the required resources for customization and organizational change".	Mezghani (2016)
Communication	Since cloud ERP adoption involves a third party (the provider), communication should be crucial to favor mutual understanding between managers and the provider and identify the alignment possibilities. Decisions about cloud adoption are complex due to the multiple challenges so regular communication between IT specialists and business managers is needed.	Katsanos (2014), Mezghani (2016)
Users' involvement	In cloud ERP projects, users' involvement is considered among the first steps that help to better identify requirements and reduce resistance to technological and organizational changes.	Makkar and Meenakshi (2012), Mezghani (2016)
Training	"The client should invest in training the employees to familiarize them with the cloud ERP application, business process modifications and change management resulting from the implementation efficiently. If managers ignore training aspects of cloud ERP, cloud ERP implementation can be inefficient"	Das and Dayal (2016)
Interpersonal skills	When to adopt cloud ERP, managers would focus more on business issues of cloud-based solutions rather than technical ones since technical adaptations would be made by the providers. The managers would need to communicate and negotiate frequently with these providers. Thus, they would need a set of interpersonal skills to be better communicators and more convincing.	Mezghani (2016)

Based on a case study conducted within a firm intending to switch toward cloud ERP, Mezghani (2016) identified many components of the social dimension in the interviewee's discourse. Terms as "discussions", "meetings", "training" and "support" are largely pronounced and linked to "organizational needs", "compatibility" and "customization". This shows that, empirically, the social dimension would play a crucial role in the alignment efforts.

Further research focusing on the social dimension components would be useful to develop more ideas about such components and propose other determinants of cloud ERP alignment. In fact, "studies on cloud-based ERP are very limited" (Peng, 2015). The existing ones focus mainly on describing benefits and risks without reflection to alignment concerns although such concerns were presented as key factors for on-Premises solutions success.

CONCLUSION

Alignment concerns have been largely considered as critical issues when talking about IT implementation. In a cloud environment, such concerns should be more problematic since cloud computing adoption involves external actors. Moreover, Parappallil, Zarvic and Thomas, (2012) state that there is an on-going trend toward studies focusing on alignment concerns linked to cloud computing due to "the enormous increase in IT outsourcing and investments in Cloud Computing services".

In this chapter, we tried to present the main alignment concerns linked first to on-Premise ERP and then cloud ERP. It was shown that, in a cloud environment, such concerns are more challenging since a third party (providers) is closely involved in cloud ERP deployment. Indeed, performing alignment efforts in this case may be hampered by the providers themselves or cause alignment to external models.

To overcome such challenges, we proposed to emphasize on a dual perspective that combines

technical efforts to social ones. Further studies may focus on the components of the social dimension as determinants of cloud ERP alignment. On the one hand, the social dimension is largely considered as enabler of on-Premise ERP alignment. On the other hand, recent studies employed such dimension as determinant of cloud ERP adoption.

Since more and more firms are pushed toward cloud ERP, conducting empirical and explanatory studies focusing on alignment concerns would be possible. Then, research frameworks could be developed toward helping practitioners overcoming such concerns.

REFERENCES

- Arnesen, S. (2013, February). Is a Cloud ERP Solution Right for You? *Strategic Finance*, 45-50.
- Benlian, A., & Hess, T. (2011). Opportunities and risks of software-as-a-service: Findings from a survey of IT executives. *Decision Support Systems*, 52(1), 232-246. doi:10.1016/j.dss.2011.07.007
- Das, S., & Dayal, M. (2016). Exploring determinants of cloud-based enterprise resource planning (ERP) selection and adoption: A qualitative study in the Indian education sector. *Journal of Information Technology Case and Application Research*, 18(1), 11-36. doi:10.1080/15228053.2016.1160733
- Davenport, T. H. (1998). Putting the enterprise into the enterprise system. *Harvard Business Review*, 76(4), 121-131. PMID:10181586
- Dutta, A., Peng, G. C., & Choudhary, A. (2013). Risks in enterprise cloud computing: The perspective of IT experts. *Journal of Computer Information Systems*, 53(4), 39-48. doi:10.1080/08874417.2013.11645649
- Elragal, A., & Elkommos, M. (2012). In-House versus In-Cloud ERP Systems: A Comparative Study. *Journal of Enterprise Resource Planning Studies*, 2012, 1-13.

- Géczy, P., Izumi, N., & Hasida, K. (2012). Cloudsourcing: Managing cloud adoption. *Global Journal of Business Research*, 6(2), 57–70.
- Grabski, S. V., Leech, S. A., & Schmidt, P. J. (2011). A review of ERP research: A future agenda for accounting information systems. *Journal of Information Systems*, 25(1), 37–78. doi:10.2308/jis.2011.25.1.37
- Grant, G. G. (2003). Strategic alignment and enterprise systems implementation: The case of Met-alco. *Journal of Information Technology*, 18(3), 159–175. doi:10.1080/0268396032000122132
- Henderson, J. C., & Venkatraman, N. (1993). Strategic alignment: Leveraging information technology for transforming organizations. *IBM Systems Journal*, 32(1), 4–16. doi:10.1147/sj.382.0472
- Hong, K. K., & Kim, Y. G. (2002). The critical success factors for ERP implementation: An organizational fit perspective. *Information & Management*, 40(1), 25–40. doi:10.1016/S0378-7206(01)00134-3
- Kasemsap, K. (2015). Implementing Enterprise Resource Planning. In M. Khosrow-Pour (Ed.), *Encyclopedia of Information Science and Technology* (3rd ed.; pp. 798–807). Hershey, PA: Information Science Reference. doi:10.4018/978-1-4666-5888-2.ch076
- Katsanos, K. (2014). *Strategic Alignment Facilitates Better Cloud Computing Decisions*. Retrieved August 30, 2015, from <http://pivotpoint.io/en-us/article/strategic-alignment-facilitates-better-c#.VgEgtZfdW3U>
- Kidd, P. T. (2011). Enterprise Information Systems: Aligning and Integrating Strategy, Technology, Organization and People. In *Enterprise Information Systems: Concepts, Methodologies, Tools and Applications* (pp. 314–328). Hershey, PA: Business Science Reference.
- Kufandirimbwa, O., Hapanyengwi, G., & Kaban-da, G. (2012). State of ICT-Business Alignment: A Case of Zimbabwe. *International Journal of IT/ Business Alignment and Governance*, 3(2), 1–20. doi:10.4018/jitbag.2012070101
- Lee, J. C., & Myers, M. D. (2004). Dominant actors, political agendas, and strategic shifts over time: A critical ethnography of an enterprise systems implementation. *The Journal of Strategic Information Systems*, 13(4), 355–374. doi:10.1016/j.jsis.2004.11.005
- Li, Q., Wang, C., Wu, J., Li, J., & Wang, Z. Y. (2011). Towards the business–information technology alignment in cloud computing environment: An approach based on collaboration points and agents. *International Journal of Computer Integrated Manufacturing*, 24(11), 1038–1057. doi:10.1080/0951192X.2011.592994
- Makkar, G. D., & Meenakshi, B. (2012). EAAS – ERP as a service. *Journal of Information and Operations Management*, 3(1), 141–145.
- Mezghani, K. (2014). Switching toward Cloud ERP: A research model to explain intentions. *International Journal of Enterprise Information Systems*, 10(3), 48–64. doi:10.4018/ijeis.2014070104
- Mezghani, K. (2016). Exploring the strategic alignment concern when switching toward cloud ERP. *Proceedings of the 27th conference of International Business Information Management Association*.
- Mezghani, K., & Mezghani, L. (2007). Alignement des processus ERP et des processus organisationnels: Le cas du processus commercial d’une entreprise. *eJournal of Digital Enterprise*, (19).
- Mezghani, K., & Mezghani, L. (2014). Effects of Business Managers Skills on Enterprise Resources Planning Strategic Alignment. *American Journal of Business and Management*, 3(1), 15–20. doi:10.11634/216796061403518

Mohamed, S., & McLaren, T. S. (2009). Probing the gaps between ERP education and ERP implementation success factors. *AIS Transactions on Enterprise Systems*, 1(1), 8–14.

Panorama Consulting. (2016). *2016 Report on ERP Systems and Enterprise Software*. Retrieved August 30, 2016, from <http://panorama-consulting.com/resource-center/2016-erp-report/>

Parappallil, J. J., Zarvic, N., & Thomas, O. (2012). A Context and Content Reflection on Business-IT Alignment Research. *International Journal of IT/Business Alignment and Governance*, 3(2), 21–37. doi:10.4018/jitbag.2012070102

Peng, G. C. (2015). Is It Now the Right Time to Migrate Enterprise Systems to the Cloud? In M. Khosrow-Pour (Ed.), *Encyclopedia of Information Science and Technology* (3rd ed.; pp. 1070–1077). Hershey, PA: Information Science Reference. doi:10.4018/978-1-4666-5888-2.ch102

Reich, B. H., & Benbasat, I. (1996). Measuring the linkage between business and information technology objectives. *Management Information Systems Quarterly*, 20(1), 55–81. doi:10.2307/249542

Reich, B. H., & Benbasat, I. (2000). Factors that influence the social dimension of alignment between business and information technology objectives. *Management Information Systems Quarterly*, 24(1), 81–113. doi:10.2307/3250980

Salleh, S. M., Teoh, S. Y., & Chan, C. (2012). Cloud Enterprise Systems: A Review Of Literature And Its Adoption. *Proceedings of the Pacific Asia Conference on Information Systems*, 76.

Shatat, A. S. (2015). Critical Success Factors in Enterprise Resource Planning (ERP) System Implementation: An Exploratory Study in Oman. *Electronic Journal of Information Systems Evaluation*, 18(1), 36–45.

Silvius, A. G., De Haes, S., & Van Grembergen, W. (2010). Explorative Study on the Influence of National Cultures on Business/IT Alignment Maturity. *International Journal of IT/Business Alignment and Governance*, 1(2), 26–45.

Yaseen, S. G. (2009). Critical factors affecting enterprise resource planning implementation: An explanatory case study. *International Journal of Computer Science and Network Security*, 9(4), 359–363.

ADDITIONAL READING

Chebrolu, S. B. (2011). Assessing the relationships among cloud adoption, strategic alignment and IT effectiveness. *Journal of Information Technology Management*, 22(2), 13–29.

Gangwar, H., Date, H., & Ramaswamy, R. (2015). Understanding determinants of cloud computing adoption using an integrated TAM-TOE model. *Journal of Enterprise Information Management*, 28(1), 107–130. doi:10.1108/JEIM-08-2013-0065

Kearns, G. S., & Sabherwal, R. (2007). Strategic Alignment Between Business and Information Technology: A Knowledge-Based View of Behaviors, Outcome, and Consequences. *Journal of Management Information Systems*, 23(3), 129–162. doi:10.2753/MIS0742-1222230306

Lequeux, J. L. (2009). De l'ERP à l'EERM en mode Cloud. *IT-EXPERT*, (78), 6–12.

Mezghani, K., Ayadi, F., & Aloulou, W. (2014). Effects of Business Managers Skills: A Study of ERP Strategic Alignment in some Arab SME's. In K. Todorov & D. Smallbone (Eds.), *Handbook of Research on Strategic Management in Small and Medium Enterprises* (pp. 421–439). Hershey, PA: Business Science Reference. doi:10.4018/978-1-4666-5962-9.ch020

Park, S. C., & Ryoo, S. Y. (2013). An empirical investigation of end-users switching toward cloud computing: A two factor theory perspective. *Computers in Human Behavior*, 29(1), 160–170. doi:10.1016/j.chb.2012.07.032

KEY TERMS AND DEFINITIONS

Cloud ERP: Putting ERP modules on a cloud so ERP software is offered as a service. The cloud ERP is managed by the provider.

Enterprise Resources Planning (ERP): A software package that is composed of standard modules connected directly to a single database and could cover all business processes of a firm.

ERP Alignment: The coherence between the ERP requirements and the firm's ones.

Interpersonal Skills: The ability of a manager to work with a team, to understand, guide and motivate others.

On-Premise ERP: An ERP system installed locally in the firm's servers and computers. Such system is generally managed by the IT staff.

Social Dimension: Related to the interactions between the business staff and the IT specialists (IT staff and providers).

Technical Dimension: Related to all kinds of customizations and adaptations needed to operate with the ERP system.

Socio–Technical Change Perspective for ERP Implementation in Large Scale Organizations

E**Jessy Nair***PES University, India***D. Bhanusree Reddy***VIT University, India***Anand A. Samuel***VIT University, India*

INTRODUCTION

Successful implementation of an Information System (IS) is vital for sustaining and enhancing the competitive position of an organization (Gunasekaran, 2005; Jing & Qiu, 2007). Management thinking and practice reflected the trend of implementing IS by relating Information Technology (IT) based development to the strategic needs of the business to develop a competitive advantage (Galliers & Baets, 1998; Al-Mudimigh, Zairi & Al-Mashari, 2001). As computing concepts and technology advanced with time, the scope of the business systems widened to become Enterprise Resource Planning (ERP). ERP enables today's fiercely competitive business environment that requires deeper interaction between customers and organization to coordinate entire value-chain of the organization. Hence disjointed functional departments are integrated as process-oriented cross functional departments, generically known as value chain system. From technological perspective, ERP evolved from legacy systems to more flexible tiered client-server architecture and a software product that represents the final stage of an evolution towards integration, originating from IT supported manufacturing (Klaus, Rosemann & Gable, 2000). ERP facilitates in achieving greater benefits from the databases and ensures that the

system environment is built following an open system approach (Al-Mashari, 2003) enabling companies to standardize business processes with ERP system and more easily endorse best practices. By creating more efficient processes, companies can concentrate their efforts on serving their customers (Gunasekaran, 2005; Laframboise, & Reyes, 2005), maximizing profits (Laframboise, & Reyes, 2005) and reducing costs (Gunasekaran, 2005) for the organization. As ERP continues to evolve into a real-time planning tool, it will play a more strategic role in helping companies achieve their business objectives. ERP has grown from coordination of manufacturing processes to integration of enterprise-wide back-end processes and in the internet era it evolves to become the database backbone for an organization's web based front-end technology to stay connected to customers.

Organizations are yet to realize efficiencies and cost-savings as originally planned. Most organizations that are implementing ERP do not find success and struggle with ERP implementations. Meta Group reports 70 per cent failure rate of ERP implementation projects. However, adopters have not been intimidated by risks of implementation which is indicative of the boom in ERP market fueled by globalization, midmarket growth and other factors. The ERP software market reached

\$47.7 billion by 2011, a compound annual growth rate of almost 11% according to AMR Research Inc. The technical system (Lee, 2000; Lee, 1999; Bostrom & Heinen, 1977) of the organization comprising of ERP and its required infrastructure and business processes acts as an enabler in achieving the organization's strategic goals with successful implementation of ERP. ERP implementation is not only a technical system imperative but has to synergize with the social system of the organization. Lee (2004) states that as social systems seek information from technical systems, the technical systems too pose its own organization requirements on the social system. From this standpoint, ERP implementation research can be classified into two major groups where ERP deployment corresponds to a technical system and ERP organizational intervention corresponds to a social system. Though ERP implementation is highly researched, a framework illustrating all its dimensions to enable the organizational decision makers to configure the most suitable combination of variables for a research theme is lacking. The objective of this chapter is to develop a General Morphological Analysis (GMA) framework to identify and illustrate research dimensions for ERP implementation research. GMA framework is a highly structured and illustrative method to identify all possible combinations of dimensions and variables for creating models of systems and processes, which are usually non-quantifiable. Extensive literature review is carried out to categorize the dimensions and its variables of ERP implementation using GMA framework. The framework serves as a map to choose research themes on ERP implementation and a suitable configuration is formulated from the map with socio technical change as a research theme.

The following section describes ERP as the backbone of an organization's technology infrastructure by detailing the ERP evolution. In this section ERP evolution is reviewed comprehensively by drawing out the significance of an unified enterprise essential for the present competitive business environment. The definitions of ERP are

then collated through literature review to detail the evolution of ERP. The section describes the means by which ERP concept has undergone a change from technical to socio technical perspective. The next section describes a morphological field that maps out various research themes for ERP implementation. Then a particular configuration of ERP dimensions is chosen as an example of how the framework can be used by decision makers to identify themes around ERP organizational intervention. Finally, we discuss implications for future research and draw conclusions.

BACKGROUND: UNDERSTANDING THE NEED FOR A UNIFIED ENTERPRISE VIEW THROUGH ERP

Earliest legacy systems in organizations were developed to manage the transactions on a routine basis and came to be known as Transaction Processing Systems (TPS) specifically used in organization's tactical operations. The period of 1960s saw organizations centralizing their computing systems and used software packages like inventory control (IC) to automate their inventory control systems (Rashid, Hossain & Patrick, 2001). The manufacturing area focused on traditional process of management of large stocks of inventory, hence IS were specifically developed to cater to this demand. Material Requirements Planning (MRP) an earliest form of computerized IS business application is a production planning and control system for managing inventory in organizations. During 1970, MRP packages were extended with further applications in order to offer complete support for the entire production planning and control cycle (Klaus, Rosemann & Gable, 2000). The augmented version of MRP resulted in Manufacturing Requirement Planning (MRP) II which integrated traditional core functions of an organization (Yusuf & Little, 1998).

With the evolution of technology, applications were added to suit the needs of organizations which are moving away from functional

view to a process view (Al-Mashari, 2003). In the process view, organizations reorganize their structures, standardize the processes and facilitate seamless flow of information to the stakeholders by integrating the information and processes of distinct subsets of the organization (Gattiker & Goodhue, 2000). ERP emerged towards end of 1980's as a derivative of MRP and is a technology system that enables standardizing an organization's processes by integration (Shehab, Sharp, Supramaniam & Spedding, 2004). ERP includes functionalities such as human resources planning, decision support applications, regulatory control, quality, elements of supply chain management and maintenance support and the like. ERP advocates believe that it can combine business process optimization in an organization and IT solutions into one integrated solution (Chung & Snyder, 1999) and support the backend functions of the organization.

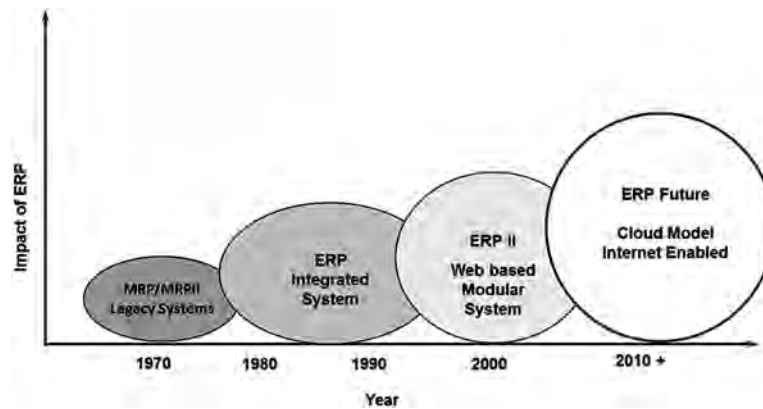
Capital intensive industries were first movers in using core modules of ERP software. With globalization, the service sectors too have realized its significance as organizations are increasingly finding it vital to keep in touch with stakeholders like suppliers, customers and partners and hence have started adding extensions to ERP in the domain of Supply Chain Management (SCM) and Customer Relationship Management (CRM). Hence ERP vendors are more focused in providing ERP systems by way of analytical applications for example SCM, CRM, Sales Force Automation (SFA), Manufacturing Execution System, Inventory Management System and so forth. This view is based on the idea that an ERP system is not simply a tool that provides a single output, but rather an infrastructure that supports capabilities of all other information tools and processes utilized by a firm (Bendoly & Schoenherr, 2005). This encompasses what is often referred to as the entire value chain of the enterprise, from prospect and customer management through order fulfilment and delivery (Shehab, Sharp, Supramaniam, & Spedding, 2004). This era of ERP systems came to be known as ERP II or the extended ERP, a

concept originally conceived by Gartner Group in 2000 and defines it as a transformation of ERP into next-generation enterprise systems. ERP II enables organization to meet customer demands through online electronic business (e-business). According to Chaffey (2007), though ERP systems were initially developed for managing the integration of processes and communication at operational level, with sophisticated solutions provided by vendors, tactical and strategic organizational levels are being integrated. Hence, ERP has become the backbone of organizations operations by connecting to the front-end operations through e-business and allowing organizations to achieve competitive advantage.

ERP systems of early era that were based on mainframe systems model moved to client/server solution model (Rao, 2000) and scalable relational database management systems (RDBMS). This has contributed to the ease of deployment of ERP systems (Gupta, 2000). New era ERP systems architecture is based on open source technology, incorporating internet in their applications for organizations to stay connected with their stakeholders since collaborating with partners is the key to success of the company. The concept of utilizing a web browser or an internet medium to access an ERP system is known as ERP on clouds. Cloud technology allows organizations to use various models to virtualize or store software, networks and communication systems and infrastructure with a vendor also known as a service provider. Though the idea seems to be favorable for business at large, the concept of implementation of ERP on cloud is still in its infancy. The evolution of ERP is depicted in Figure 1.

Once an organization (large or small and medium business) selects ERP as a technology infrastructure to be implemented, the deployment method is planned by the organization. Parr and Shanks' (2000) comprehensive literature has detailed ERP implementation approaches as comprehensive, middle road and vanilla. Koch (2007) categorizes these approaches as big bang, franchising and slam dunk respectively and defines

*Figure 1. Evolution of ERP systems
Adapted from (Nair, Reddy & Samuel, 2011).*



these concepts as follows: Big Bang method of installation is chosen by organizations when legacy systems are swapped with ERP system installed for the entire organization. This method of organizing projects requires tremendous resources from the organization in terms of investment, people and time. Adopting new standard processes and adaptation to the new software system by the entire organization has never been a success story for a majority of the organizations. In the Franchising strategy, ERP systems are independent and only common processes are linked across the organization. The merit of this category of ERP deployment is that core businesses of the organization are not affected during the course of the project since pilot installation happens at units where such a project would not disrupt the routine processes. Finally, the slam dunk method is where a few key processes are bought under the preview of the ERP system. This strategy of ERP project is adopted by small organizations which prefer to see their systems up and running and so avoid the radical reengineering of processes in the organization.

The outcome of organizational change due to ERP system intervention results in tangible and intangible benefits to the organizations. The tangible benefits from ERP implementation have a bottom-line impact on profitability by way of reduction in inventory, administrative costs, operational costs, cycle-time reduction and the like.

Intangible benefits gained through ERP system such as user satisfaction, customer satisfaction, efficient communication of information and increased efficiency in users are rather difficult to quantify. Therefore to take an extensive advantage of the tangible and intangible benefits, the technological dimension and social dimension from an organization context have to be synergetic. Table 1 consolidates benefits of ERP from the literature on the basis of dimensions such as strategic, operational, tangible, and intangible (Gattiker & Goodhue, 2000; Shang & Seddon, 2000; Murphy & Simon, 2002; Holsapple & Sena, 2003).

ERP DEFINITIONS: TECHNICAL TO SOCIOTECHNICAL CHANGE PERSPECTIVE IN LARGE SCALE ORGANIZATIONS

ERP is an application software system that unifies a large scale organization through its technology imperative by integrating island like functional departments to a cross-functional process oriented enterprise organized itself by the concept of value-chain systems. Therefore information which is a critical resource and indispensable to the organization for decision making can be leveraged by organization stakeholders seamlessly across geographical boundaries. This way ERP

Table 1. Benefits of ERP implementation

	Tangible / Quantitative	Intangible / Qualitative
Operational	○ Reduction in inventory ○ Administrative costs ○ Operational costs ○ Cycle-time reduction ○ Productivity Improvement ○ Quality improvement	○ User satisfaction ○ Efficient communication of information and coordination ○ Increased efficiency in users
Strategic	○ Process standardization ○ Process integration ○ Product Innovation ○ Cost Leadership ○ Customer Supplier Linkages	○ Customer satisfaction ○ Improved system reliability ○ Higher data quality ○ Greater agility in implementing new businesses ○ Decision making ○ Learning curve ○ Competitive Advantage

enables the organization to better deploy and manage resources and increase overall effectiveness of the organization. Davenport (1998) refers to ERP as the “most important IT development in the 1990s” accepted by the business world. Its main value is providing the opportunity to integrate the entire organization and share information through a common database repository. Table 2 describes the transformation of ERP as a concept from technology imperative to a change agent causing intervention on the organization consequently resulting in social intervention. This change from a technology imperative comprising of technology and business processes to social imperative to include dimensions such as organization and users is illustrated in Table 2. Though ERP implementation is highly researched, a framework identifying all possible combinations of dimensions and variables for creating models of ERP system implementation for evolving a research theme is lacking. This is notwithstanding Orlikowski & Baroudi (1991) state that IS research to a large extent “is concerned with the ongoing relations among information technology, individuals, and organizations (p.6)”. Nevertheless, Yusuf, Gunasekaran & Abthorpe (2004) emphasize this statement in ERP context and point that its implementation must be managed as a program

of wide-ranging organisational change initiatives rather than as a software installation effort. The following section describes a morphological field that maps out various research themes for ERP implementation.

GENERAL MORPHOLOGICAL ANALYSIS OF ERP IMPLEMENTATION IN LARGE SCALE ORGANIZATION

ERP implementation is not just an installation of software but a project where the organization has to consider various issues that are organizational and stakeholder oriented. Shehab, Sharp, Supramaniam & Spedding (2004) emphasize that organizations should adopt a strategic approach by reengineering their business processes to adapt to world class built-in ERP software functionalities developed by vendors. This will enable in bringing about “deep changes” in *de-facto* methods of doing work in the organization resulting in shake up the roles and responsibilities of important people. ERP implementation seeks to integrate different components of an organization- functional units, business processes, people (Ranganathan & Brown, 2006; Gunasekaran, 2005), technology (Ranganathan

Table 2. Evolution of ERP definition

ERP Definitions	Source
Enterprise Systems are commercial software packages that promise the seamless integration of all the information flowing through a company- financial and accounting information, human resource information, supply chain information, customer information.	(Davenport,1998)
An Enterprise Resource Planning system is a packaged business software system that allows a company to: ➤ Automate and integrate the majority of its business processes ➤ Share common data and practices across the entire enterprise ➤ Produce and access information in a real-time environment	(Deloitte Consulting, 1999)
ERP systems are about people, not processes or technology. The organization goes through major strategic transformation during this period. The management of this change for ERP implementation requires careful planning.	(Bingi, Sharma & Godla, 1999)
ERP is a phenomenon leading to a complete renewal of enterprise IT infrastructure resulting in significant consequences in economic, technical and social aspects.	(Markus & Tanis, 2000)
ERP systems are the software tools used to manage enterprise data and provide information to those who need it, when they need it.	(Gunasekaran, 2005)
ERP is more of an organizational and social aspect than technical.	(Hanafizadeh & Ravasan, 2011)
ERP imposes its own logic, best practices and structure on a company's strategy, culture, organization of departments and structure and hence pushes for and demands a huge amount of change management.	(Poti, Bhattacharyya, & Kamalanabhan, 2011)
ERP systems are defined from the perspective of stakeholder groups by the perceptions they hold in the organisation's management of the change process which accompanies the implementation.	(Williams, Williams & Morgan, 2013)

& Brown, 2006; Gunasekaran, 2005), organizational issues (Gunasekaran, 2005) and partners. The components behave as a unified whole and achieve superior firm performance (Ranganathan & Brown, 2006; Gunasekaran, 2005). ERP implementation is a significant challenge due to integration of a variety of legacy systems to ERP deployments. To facilitate the users in adopting the system efficiently and effectively due to changed and standardized processes results in a socio technical change management. Nevertheless, Keen (1981) states that many IS implementation have been successful from a technical perspective but organizational failures. To identify the themes for ERP research a General Morphological Analysis (GMA) is used. GMA is a generalized framework that can accommodate and facilitate the study of various aspects of ERP implementation in a multi-dimensional space.

GMA is a method for structuring and investigating total set of relationships contained in multidimensional, usually non-quantifiable, complex problems (Ritchey, 2006). Pioneered by Fritz Zwicky at the California Institute of Technology during the 1930s and 1940s, it relies on a constructed parameter space, linked by way of logical relationships, rather than on causal relationships and a hierarchal structure (as cited in Ritchey, 2006). The dimensions identified are placed in columns and variables of each dimension are clustered under the associated dimension in columns. This visual relationship is called as a morphological field (Ritchey, 2006) and a combination of variables from each of the dimension is called a configuration or a morphotype. Three broad steps of developing a morphological framework are: *messes*- complex and unstructured issue in hand to be studied, *problems*-here the is-

sue has a defined form therefore its dimensions and its corresponding variables are identified and represented in columns and the last level is called a *puzzle*- a well-defined and well-structured problem with a specific solution that somebody can work out (as cited in Ritchey, 2002). There have been attempted studies in areas of modelling complex socio-technical systems (Ritchey, 2002), accountability in life cycle of products and packaging introduced in the market (Ritchey, 2006), virtual organization contexts (Shekar & Ganesh, 2007) and knowledge transfer research (Kumar & Ganesh, 2009).

ERP implementation calls for socio technical change in the organization. A morphological framework can be a guidepost to identify all possible dimensions and its associated variables to formulate a *morphotype* or *puzzle* to carry out further research to reduce the risk in ERP implementation. An extensive literature survey to understand different dimensions of ERP implementation using a combination of keywords such as ERP, enterprise systems, ERP implementation, ERP project, ERP implementation framework and models was carried out. The review results are classified based on two central themes or *messes* of ERP implementation research; ERP deployment and ERP intervention as demonstrated in

Table 3 and Table 4 respectively. A representative list of variables drawn from literature are listed in columns and clustered under their respective dimensions (see Table 3 and 4). ERP deployment as detailed in the earlier section of the chapter calls for organization to plan for ERP implementation based on the size of the organization, type of module, method of deployment and implementation strategy. GMA for ERP deployment will be of scope to practitioners to develop themes for future research for effective deployment of ERP. However ERP intervention analysis (see Table 4) is a highly researched area by academics and benefits the top-management in anticipating the risks associated to ERP implementation. The dimensions of ERP intervention can be categorized as stages of ERP software implementation, theoretical perspectives, ERP frameworks and contexts.

Socio Technical Change: A Morphotype of ERP Intervention

The morphology based framework shown in Table 4 provide not only a holistic map of ERP intervention in an organization but also can be used a tool to understand and analyze different aspects of ERP interventions. This is perhaps the first attempt to use morphology based approach

Table 3. General morphological analysis map of ERP deployment

ERP Deployment			
Industry	Modules	Method of Deployment	Deployment Strategies
Large scale enterprises	Human Resources	Big Bang	On - Premise ERP Installation
	Customer Relationship Management		
	Supply Chain Management	Phased (by function or location)	Hybrid Model
	Procurement		
	Project Management		
Medium and Small scale enterprises	Financials	Modular	Cloud Model
	Sales Management		
	Sales Order Processing		
	Product Life Cycle Management	Hybrid	On-Demand/Software as a Service(SaaS)
	Business Intelligence		

Table 4. General morphological analysis map of ERP intervention

ERP Intervention			
Stages	Perspectives	Frameworks	Contexts
Chartering Phase, Project Phase, Shakedown Phase, Onward & Upward Phase	Process	De-Lone & McLean Model	Benefits & Risks
	Variance		Organizational Change
	Business-IT Alignment		
Initiation, Adoption/Adaption, Acceptance, Routinization, Infusion	Actor-Network	Balanced Scorecard	Organizational Environment
	Technology Acceptance		Stakeholders-Internal & External
Project Preparation, Business Blueprint, Realization, Final Preparation, Go Live and Support	User Acceptance	Leavitt's Diamond Model	Stakeholders-Internal & External
	Structuration		
	Socio-Technical System		

to formulate different research themes on ERP intervention in an organization. The morphological map of ERP intervention given in Table 4 has four dimensions: perspective, stages, frameworks and contexts having 3, 8, 3 and 4 variables respectively. By interconnecting different instances across the four dimensions of the framework one can come up with large number of possible notions to ERP interventions. From Table 5 it is observed that we can have $3 \times 8 \times 3 \times 4 = 288$ morphotypes as themes for research for ERP intervention. ERP implementation is a collaborative process and intervention created by the technology triggers requirements of change in the social elements of the organization (Nair, Reddy & Samuel, 2011). Hence ERP intervention cannot be perceived in isolation as a technology change but as a socio technical change in an organization.

Socio technical change is a foundational area to assess impact of ERP intervention in an organization. Social factors are characteristics of the social context of the organization (Kumar & Ganesh, 2009), while technical refers to hardware and software requirements (Whitworth, 2009) corresponding to ERP implementation. The morphological map of ERP intervention is analyzed and synergistic subthemes are highlighted in Table 5. Leavitt's (1965) diamond model is a suitable framework to analyze ERP intervention on internal organizational environment dimensions: organiza-

tion stakeholders, internal organization, processes and technology (Bostrom & Heinen, 1977; Lee, 2000; Hanafizadeh & Ravasan, 2011). The four stages of Markus & Tanis (2000) called as the emergent process model for ERP implementation consists of four stages (chartering phase, project phase, shakedown phase and onward and upward phase). This software life cycle model posits the necessary actions to be deliberated by decision makers in each phase of the ERP project where actions in each phase are mentioned *a priori* (Nair, Reddy, & Samuel, in press). The socio technical change morphotype formulated in this article having subthemes of ERP life cycle, socio technical perspective, Leavitt's diamond model and organizational change can form as a basis of future research in this area.

FUTURE RESEARCH DIRECTIONS

The scope of dimensions that can be explored by both the academicians and practitioners through morphological framework is extensive. Further refinement of the morphological framework can be taken up by using rigor in gathering literature to include top quality practitioner articles and more databases for search purposes to identify more dimensions for study. Since this framework using morphology is probably the first in area of

Table 5. A morphotype of socio technical change in the morphological field of ERP intervention

ERP Intervention			
Stages	Perspectives	Frameworks	Contexts
Chartering Phase, Project Phase, Shakedown Phase, Onward & Upward Phase	Process	De-Lone & McLean Model	Benefits & Risks
	Variance		
Initiation, Adoption/Adaption, Acceptance, Routinization, Infusion	IT Diffusion	Balanced Scorecard	Organizational Change
	Business-IT Alignment		Organizational Environment
	Actor-Network		
	Technology Acceptance		
Project Preparation, Business Blueprint, Realization, Final Preparation, Go Live and Support	Structuration	Leavitt's Diamond Model	Stakeholders -Internal & External
	Socio Technical System		

ERP implementation research; refinement of this ERP implementation morphology can be a scope of further studies. New avenues of research can be explored through dimensions such as deployment models like Software as a Service (SaaS) in cloud concept since ERP software trends are orienting towards this area and issues arising in ERP implementation in small and medium size industries.

CONCLUSION

ERP is the backbone of technology infrastructure of an organization and has evolved from a technical system to a socio technical system. ERP implementation as a technological intervention ensues organizational change. There are several studies associated to performance analysis of ERP implementation and are extensively researched. The objective of this article is to use General Morphological Analysis (GMA) framework to develop research themes to analyze the performance of ERP implementation. Hence a novel morphological analysis framework is developed in this article that can be used to formulate research concepts in ERP implementation based on perspectives,

dimensions and corresponding variables. The article identifies two morphological subfields: ERP deployment and ERP interventions for technical and social perspectives of ERP implementation respectively. To a researcher or a practitioner, the framework serves as a map to choose research themes from the morphological fields of ERP deployment and ERP interventions. Application of the morphological field is demonstrated by identifying a suitable morphotype from the map with socio technical change as the theme.

REFERENCES

- Al-Mashari, M. (2003). Enterprise resource planning (ERP) systems: A research agenda. *Industrial Management & Data Systems*, 103(1), 22–27.
- Al-Mudimigh, A., Zairi, M., & Al-Mashari, M. (2001). ERP software implementation: An integrative framework. *European Journal of Information Systems*, 10(4), 216–226. doi:10.1057/palgrave.ejis.3000406
- AMR Research Inc. (n.d.). *AMR Market Sizing Report: 2006 – 2011*. Retrieved December 11, 2013, from <http://www.timereaction.com/papers/amrerp.pdf>

- Bendoly, E., & Scoenherrer, T. (2005). ERP System and Implementation – Process Benefits Implications for B2B e- Procurement. *International Journal of Operations & Production Management*, 25(4), 304–319. doi:10.1108/01443570510585516
- Bingi, P., Sharma, M. K., & Godla, J. K. (1999). Critical Issues Affecting an ERP Implementation. *Information Systems Management*, 16(5), 7–15. doi:10.1201/1078/43197.16.3.19990601/31310.2
- Bostrom, R. P., & Heinen, J. S. (1977). MIS Problems and Failures: A Socio- Technical Perspective Part I: The Causes. *Management Information Systems Quarterly*, 1(3), 17–32. doi:10.2307/248710
- Chaffey, D. (Ed.). (2007). Managing E-Business Infrastructure. E-Business and E-Commere Management – Strategy, Implementation and Practice. Pearson.
- Chung, S., & Snyder, C. (2000). ERP adoption: A technological evolution approach. *International Journal of Agile Management Systems*, 2(1), 24–32. doi:10.1108/14654650010312570
- Davenport, T. (1998). Putting Enterprise into the Enterprise system. *Harvard Business Review*, 76(4), 1–11. PMID:10181586
- Deloitte Consulting. (1999). *ERP's second wave: Maximizing the value of ERP-enabled processes*. Author.
- Galliers, R. D., & Baets, W. R. (1998). *Information technology and organizational transformation: innovation for the 21st century organization*. John Wiley & Sons, Inc.
- Gattiker, T. F., & Goodhue, D. L. (2000, January). Understanding the Plant Level Costs and Benefits of ERP: Will the Ugly Duckling Always Turn Into a Swan? In R. Sprague Jr. (Ed.), *Proceedings of the 33rd Hawaii International Conference on System Sciences* (pp. 1-10). Los Alamitos, CA: IEEE Computer Society Press. doi:10.1109/HICSS.2000.926907
- Gunasekaran, A. (2005). Enterprise Information Systems & Organizational Competitiveness. Editorial Preface, Inaugural Issue. *International Journal of Enterprise Information Systems*, 1(1), i–vi.
- Gupta, A. (2000). Enterprise resource planning: The emerging organizational value systems. *Industrial Management & Data Systems*, 100(3), 114–118. doi:10.1108/02635570010286131
- Hanafizadeh, P., & Ravasan, A. Z. (2011). A McKinsey 7S model-based framework for ERP readiness assessment. *International Journal of Enterprise Information Systems*, 7(4), 23–63. doi:10.4018/jeis.2011100103
- Holsapple, C. W., & Sena, M. P. (2005). ERP plans and decision-support benefits. *Decision Support Systems*, 38(4), 575–590. doi:10.1016/j.dss.2003.07.001
- Jing, R., & Qiu, X. (2007, June). A study on critical success factors in ERP systems implementation. In Service Systems and Service Management (Ed.), *International Conference on Service Systems and Service Management* (pp. 1-6). IEEE. doi:10.1109/ICSSSM.2007.4280192
- Keen, P. G. (1981). Information systems and organizational change. *Communications of the ACM*, 24(1), 24–33. doi:10.1145/358527.358543
- Klaus, H., Rosemann, M., & Gable, G. (2000). What is ERP? *Information Systems Frontiers*, 2(2), 141–162. doi:10.1023/A:1026543906354
- Koch, C. (2007). *The ABCs of ERP*. Retrieved October 1, 2012, from <http://www.cio.com/research/erp/edit/erpbasics.html?action=print>
- Kumar, J. A., & Ganesh, L. S. (2009). Research on knowledge transfer in organizations: A morphology. *Journal of Knowledge Management*, 13(4), 161–174. doi:10.1108/13673270910971905
- Laframboise, K., & Reyes, F. (2005). Gaining competitive advantage from integrating enterprise resource planning and total quality management. *Journal of Supply Chain Management*, 41(3), 49–64. doi:10.1111/j.1055-6001.2005.04103005.x

- Leavitt, H. J. (1965). Applied Organizational Change in Industry: Structural, Technological and Humanistic Approaches. In J. G. March (Ed.), *Handbook of Organizations* (pp. 1144–1170). Chicago: Rand McNally.
- Lee, A. S. (1999). Inaugural editor's comments. *Management Information Systems Quarterly*, 23(1), v–xi.
- Lee, A. S. (2000). Researchable Directions for ERP and Other New Information Technologies. Editors Comments. *Management Information Systems Quarterly*, 24(1), iii–viii.
- Lee, A. S. (2004). Thinking about Social Theory and Philosophy for Information Systems. In *Social Theory and Philosophy for Information Systems* (pp. 1–26). Chichester, UK: John Wiley & Sons.
- Markus, M. L., & Tanis, C. (2000). The Enterprise System Experience - from adoption to success. In *Framing the Domains of IT Management: Projecting the Future Through the Past* (pp. 173–207). Pinnaflex Educational Resources Inc.
- Murphy, K. E., & Simon, S. J. (2002). Intangible benefits valuation in ERP projects. *Information Systems Journal*, 12(4), 301–320. doi:10.1046/j.1365-2575.2002.00131.x
- Nair, J., Reddy, D. B. S., & Samuel, A. A. (2011). A Framework for ERP System Implementation: Socio-Technical Change Management Model Perspective. *International Journal of Business Derivatives*, 1(1), 185–207. Retrieved from <http://ssrn.com/abstract=1924671>
- Nair, J., Reddy, D. B. S., & Samuel, A. A. (2015). Conceptualizing Dimensions of Enterprise Resource Planning Systems Success: A Socio-Technical Perspective. *International Journal of Enterprise Information Systems*.
- Orlikowski, W. J., & Baroudi, J. J. (1991). Studying Information Technology in Organizations: Research Approaches and Assumptions. *Information Systems Research*, 2(1), 1–28. doi:10.1287/isre.2.1.1
- Parr, A. N., & Shanks, G. (2000). A model of ERP project implementation. *Journal of Information Technology*, 15(4), 289–304. doi:10.1080/02683960010009051
- Poti, S., Bhattacharyya, S., & Kamalanabhan, T. J. (2011). Social and Cultural Challenges in ERP Implementation: A Comparative Study Across Countries and Cultures. *International Journal of Information Systems and Social Change*, 2(4), 44–67. doi:10.4018/jissc.2011100104
- Ranganathan, C., & Brown, C. V. (2006). ERP Investments and the Market Value of Firms. *Information Systems Research*, 17(2), 145–161. doi:10.1287/isre.1060.0084
- Rao, S. S. (2000). Enterprise resource planning: Business needs and technologies. *Industrial Management & Data Systems*, 100(2), 81–88. doi:10.1108/02635570010286078
- Rashid, M., Hossain, L., & Patrick, J. (2001). The Evolution of ERP Systems: A Historical Perspective. In L. Hossain, J. Patrick, & M. Rashid (Eds.), *Enterprise Resource Planning: Global Opportunities & Challenges* (pp. 1–16). Idea Group Publishing.
- Ritchey, T. (2002). *Modelling Complex Socio-Technical Systems using Morphological Analysis*. Adapted from an address to the Swedish Parliamentary IT Commission, Stockholm, December 2002. Retrieved from www.swemorph.com/pdf/it-webart.pdf
- Ritchey, T. (2006). Problem structuring using computer-aided morphological analysis. *The Journal of the Operational Research Society*, 57(7), 792–801. doi:10.1057/palgrave.jors.2602177

- Shang, S., & Seddon, P. B. (2000, August). A comprehensive framework for classifying the benefits of ERP systems. *Americas Conference on Information Systems*, 1005-1014.
- Shehab, E. M., Sharp, M. W., Supramaniam, L., & Spedding, T. A. (2004). Enterprise resource planning: An integrative review. *Business Process Management Journal*, 10(4), 359-386. doi:10.1108/14637150410548056
- Shekhar, S., & Ganesh, L. S. (2007). A morphological framework for virtual organizations. *IIMB Management Review*, 19(4), 355-364.
- Whitworth, B. (2009). *The Social Requirements of Technical Systems*. In *Handbook of Research on Socio-Technical Design and Social Networking Systems* (pp. 1-22). IGI Global. doi:10.4018/978-1-60566-264-0
- Williams, J., Williams, M. D., & Morgan, A. (2013). A teleological process theory for managing ERP implementations. *Journal of Enterprise Information Management*, 26(3), 235-249. doi:10.1108/17410391311325216
- Yusuf, Y., & Little, D. (1998). An empirical investigation of enterprise-wide integration of MRP II. *International Journal of Operations & Production Management*, 18(1), 66-86. doi:10.1108/01443579810192916
- Kwon, T. H., & Zmud, R. W. (1987). Unifying the fragmented models of information systems implementation. In J. R. Boland & R. Hirshheim (Eds.), *Critical issues in information systems research* (pp. 227-251). New York: John Wiley.
- Lyytinen, K. (1987). Different Perspectives on Information Systems: Problems and Solutions. *ACM Computing Surveys*, 19(1), 5-46. doi:10.1145/28865.28867
- Orlikowski, W. J., & Baroudi, J. J. (1991). Studying Information Technology in Organizations: Research Approaches and Assumptions. *Information Systems Research*, 2(1), 1-28. doi:10.1287/isre.2.1.1
- Robey, D., Ross, J. W., & Boudreau, M. C. (2002). Learning to implement enterprise systems: An exploratory study of the dialectics of change. *Journal of Management Information Systems*, 9(1), 17-46.
- Tavana, M. (2011). *Managing Adaptability*. Intervention, and People in Enterprise Information Systems; doi:10.4018/978-1-60960-529-2

ADDITIONAL READING

- Davenport, T. (1998). Putting Enterprise into the Enterprise system. *Harvard Business Review*, 76(4), 1-11. PMID:10181586
- Davenport, T. H. (2000). *Mission Critical: Realizing the Promise of Enterprise Systems*. Boston, MA: HBR Press.
- Hammer, M., & Stanton, S. (1999). How Process Enterprises Really Work. *Harvard Business Review*, 108-118. PMID:10662000

KEY TERMS AND DEFINITIONS

Enterprise Resource Planning (ERP)

System: ERP Systems are *defacto* standardized application system software that centralizes database of the organization and brings about organizational change during its implementation to achieve organization goals.

Morphological Framework: An illustrative mapping of dimensions of a problem area and clustering the sub-dimensions against the dimensions to configure the problem area for research.

Socio Technical Change Perspective: The change brought about through intervention of technology on the symbiosis of social (organization stakeholders, organization structure, policies, culture, values, norms, beliefs) and technical (business processes and technological infrastructure) dimensions of an organization.

Techno Organizational Change: The change brought about during the implementation of new technology in the organization's structure, policies, culture, values, norms and beliefs.

Value-Chain: Leveraging technology to seamlessly integrate primary functional departments of the organization to process-oriented cross-functional departments to add value by way of service to end customer.

E

Category E

Entrepreneurship

Current Scenario of Youth Entrepreneurship in India

E**Neeta Baporikar***Namibia University of Science and Technology, Namibia & University of Pune, India*

INTRODUCTION

Many young people cannot find employment. This has become particularly acute since the education explosion in early 2000's and onset of the financial crisis in 2008. These outcomes are both inefficient and inequitable. Evidence shows that the unemployed are unhappier, more likely to experience a range of health issues, and face difficulties in integrating back into the labour market place (Bell & Blanchflower, 2009). For young people, the effects of unemployment may be particularly scarring. Evidence suggests that a spell of youth unemployment increases the likelihood of poorer wages and unemployment in later life (Blanchflower & Oswald, 1998). Such outcomes also have pronounced social costs. It represents a loss of potential output and leads to increased costs to the taxpayer apart from causing social disruptions. One potential way of integrating young people into the global economy is to increase youth entrepreneurship. World Bank estimates that by 2015, there will be about 3 billion youths less than 25 years old and a big portion of that will be in India. Hence, promoting youth entrepreneurship will not only help in reducing unemployment but more importantly show young people that they have alternatives to create their own destiny? Moreover, youth entrepreneurship has become a topic of interest for research scholars and also a subject of major concern for the Government. While youth entrepreneurship is an under-explored field, the main factor for its growing attention is the increased number of unemployed young people. Furthermore entrepreneurship is seen as

a channel for the talents of many highly educated young people to explore their potential and cash their business acumen. The core of this chapter is to provide a contemporary perspective on youth entrepreneurship.

BACKGROUND

In spite of the increasing recognition of entrepreneurship as a source of job creation, regional development, and economic dynamism in a rapidly globalizing world, there has been no systematic attempt to look at it from a youth angle. According to Holland's (1997) theory, people are attracted to work environments that conform to their personality orientation. Chell (2008) suggests that personality traits of entrepreneurs may be important for entrepreneurship. Shepherd et al., (2009) posit that the personality traits of the entrepreneur may explain entrepreneurial failure. According to Splaver (1977) it is important for you to have a good understanding of yourself and your personality, if you are to make intelligent career plans. (Ciavarella et al., 2004; Zhao et al., 2010) points out that personality trait have a direct effect on entrepreneurial performance measures. Personality traits of entrepreneurs may have a different effect on firm performance in case the firms are innovative (Zhao et al., 2010). Penrose (1959) emphasized that carrying out similar tasks has important implications for cognitive processes of human beings. Simon's (1947) gave the idea of bounded rationality which refers to human limitation to process information. Under the as-

sumption of bounded rationality, therefore, past experience influences the processing of incoming information. Witt (2000) explained that events in the environment are only perceived and interpreted along some specific associative lines. In this sense, a cognitive frame is a schematic representation of an individual's perception of the environment built through prior learning and adaptation. Gardner (1983) argues that intelligence refers to both the personal decisions and potentials of individuals. This potential comes out or develops according to cultural environment, values and opportunities. Baum et al. (2001) reported a positive relationship between practical intelligence and entrepreneurial processes and entrepreneurial characteristics, which proposes that practical intelligence is one of the strengths underlying high performance enterprises.

Although youth may not seem like a likely group for entrepreneurship because of their inexperience and lack of finances, they may not have an interest in entrepreneurship. Even if they have, intention is one thing, action is another. The proportion of youth involved in self-employment varies across countries, which may indicate variations in barriers and opportunities and labour market conditions. Societal outlook towards youth entrepreneurship is also a deciding factor. In some parts of India, where the youth comes from business families background then the motivation levels as well as support is high as compared to youth from other background. While young and adult entrepreneurs tend to have similar motivations for entrepreneurial activities, they often have different approaches and run different types of businesses. As observed, young entrepreneurs generally operate smaller businesses than adults. Youth-operated businesses are also more likely to be concentrated in certain industries with low barriers to entry and low capital requirements such as construction, information and communication and other services firms and tend to focus on narrow product lines. Accordingly, many young entrepreneurs focus exclusively on local markets because of familiarity and because they lack the

knowledge about opportunities in other markets and how to take advantage of them (Chigunta, 2002). However, at the same time, they are likely to be more open than adult entrepreneurs to international activity. Through grounded theory, review of policy documents and secondary data analysis the aim of this chapter is to provide a contemporary perspective on youth entrepreneurship with focus on India. Moreover, capacity of the ecosystem to promote innovation derives from its diversity of talents, trust across social barriers, motivations that rise above short term rationality, and social norms that promote rapid, "promiscuous" collaboration and experimentation among individuals (Hwang & Horowitz 2012). Indian entrepreneurship development shows that this distinct social behavior is highly present in the Indian ecosystem. (Baporikar, 2013)

Understanding Youth Entrepreneurship

Before looking at the potential benefits of promoting youth entrepreneurship, it is important to have an understanding of what 'youth entrepreneurship' is. 'Youth' is defined by the United Nations as those between 15-24 years of age. 'Entrepreneurship' is a generic term that subsumes many issues. It has, therefore, been defined in very many ways. Rabboir (1995) – quoted in Schnurr and Newing (1997) - lists twenty definitions of 'entrepreneurship' from various authorities on the subject. He concludes that efforts to reach a consensus on the subject have not been successful and various analysts are changing their definitions as work, study and experience in the field evolve. Given the elusive definition of entrepreneurship, it is increasingly recognised that what is of great consequence is not what 'entrepreneurship' is or who 'entrepreneurs' are, but rather what they do or the ways in which different types of people, at different stages in the lives of their enterprises will respond to assistance of various types (Schnurr & Newing, 1997; Harper, 1996). Drawing upon the above definition of entrepreneurship, and for

the purpose of hereto, 'youth entrepreneurship' is defined as the "practical application of enterprising qualities, such as initiative, innovation, creativity, and risk-taking into the work environment (either in self-employment or employment in small start-up firms), using the appropriate skills necessary for success in that environment and culture" (Schnurr & Newing, 1997). This definition assumes the following: young individuals developing and making full use of their own abilities, alone or in groups; young people defining their own problems, identifying solutions and finding resources to realize their vision; and young people realizing their own potential and vision, growing in confidence and taking active roles in their own communities.

Youth Entrepreneurship and Self-Employment

It is now widely accepted that there are many good reasons to promote entrepreneurship among young people. While caution should be exercised so that entrepreneurship is not seen as a 'mass' or wide-ranging solution which can cure all society's social ills, as many experts such as Curtain (2000) warn, it has a number of potential benefits. An obvious, and perhaps significant one, is that it creates employment for the young person who owns the business. This is especially the case in an economy subject to rationalization, change and restructuring. Many experts believe that this could bring back the alienated and marginalized youth into the economic mainstream (Curtain, 2000; White & Kenyon, 2000). There may also be a direct effect on employment if new young entrepreneurs hire fellow youths from the 'dole' queues (Curtain, 2000). In this way, entrepreneurship could help address some of the socio-psychological problems and delinquency that arise from joblessness. Youth-run enterprises (YREs) also provide valuable goods and services to society, especially the local community (OECD, 2001). Stone, et al. (n.d) in a survey of YREs in Minnesota, USA, observed that they build houses, publish books,

run restaurants, staff child care centres, provide business services, and offer others services and goods. This results in the revitalization of the local community.

It has also been observed that new small firms tend to raise the degree of competition in the product market, thereby bringing gains to consumers (Curtain, 2000). In addition, the enterprises may create linkages between youth entrepreneurs and other economic actors, such as through subcontracting, franchising, and so on (White and Kenyon, 2000). Youth entrepreneurship also promotes innovation and resilience as it encourages young people to find new solutions, ideas and ways of doing things through experience-based learning (OECD, 2001; White and Kenyon, 2000). In certain circumstances, young entrepreneurs may be particularly responsive to new economic opportunities and trends. This is especially important given the on-going globalization process. It is increasingly accepted that youth entrepreneurs can present alternatives to the organization of work, the transfer of technology, and a new perspective to the market (White & Kenyon, 2000).

In a broader sense, 'entrepreneurship', when treated as 'enterprise', helps young women and men develop new skills and experiences that can be applied to many other challenges in life. In the Oxford Thesaurus, 'enterprise' is defined as "resourcefulness, initiative, drive, imagination, enthusiasm, zest, dash, ambition, energy, energy, vitality, boldness, daring, audacity, courage, get up and go..." It is worth noting that these attributes are generally associated with youth. In that sense, youth have "the qualities of resourcefulness, initiative, drive, imagination, enthusiasm, zest, dash, ambition, energy, boldness, audacity, courage..." (Schnurr & Newing, 1997:2)

Bennell (2000) in this regard argues that the challenge for governments, NGOs and international bodies seeking to improve youth livelihoods is to "tap into the dynamism of young people and build on their strong spirit of risk-taking". Moreover, the success of the 'new economy'-however defined-is dependent on the promotion

of a culture of entrepreneurship. It has been observed that youth have the capacity to understand it and be its pioneers. This is reflected in high youth participation in internet business start-ups (OECD, 2001; Curtain, 2000). Given this situation, the promotion of youth enterprise in general and youth entrepreneurship in particular is vital. The importance of this promotion should also be seen in the context of improving social attitudes towards entrepreneurship. Collectively, these influences are referred to as an enterprise culture. White and Kenyon (2000) define an 'enterprise culture' as a "set of attitudes, values and beliefs operating within a particular community or environment that lead to both 'enterprising' behaviour and aspiration towards self-employment" (p.18).

Blockades to Youth Entrepreneurship

A lot of constraints and barriers to youth entrepreneurship have been identified like lack of capital, poor infrastructure, strict and cumbersome Government regulations, social attitudes, lack of skills, inadequate entrepreneurship education, lack of work experience, under capitalization, lack of networks, market barriers, lack of guidance and awareness etc. These constraint or barriers prevents some from turning ideas into projects. Apart from these the other barriers could be the culture, milieu, upbringing, gender bias, risk aversion and the individual persona.

Steps to Overcome Barriers

Entrepreneurship can play a role in supporting employment creation and reducing the pressures on labour market is now a well-established fact. However, there are no quick fixes to ensure that all youth are integrated into the labour market. While some countries do a better job than others, all countries face pressures and challenges in helping their youth. Yet, in Indian scenario there is a greater need to: Conduct competitions and youth fairs/melas, Develop entrepreneurship

skills, Develop entrepreneurship infrastructure, Equipping with life skills, Foster entrepreneurial mindset, Institutional Development, Investment Services, Knowledge Management, Provide information, advice, coaching and mentoring and Provide financial support.

SOLUTIONS AND RECOMMENDATIONS

There is need to stimulate policy debate on the factors that encourage youth entrepreneurship. The obstacles that stand in the and the policy measures that can support youth entrepreneurship needs to be revisited to suit the changing times and that too keep in in mind the contextual and geographical country based requirements. In general the critical issues that need be addressed immediately include: access to micro-credit for young, as a major constraint to the growth of youth enterprises is lack of access to finance. Further, lack of sufficient collateral, experience and biases further disadvantage young people. It is also important to note that many micro-credit schemes, especially youth credit schemes, have failed in many countries. The overall message from the failure of these schemes suggests that success or failure in terms of financial viability and servicing the poor, in this case young people largely depends on the design of the programme (Curtain, 2000).

To overcome the lacuna regarding micro financing the need is to treat youth as 'clients' and not mere programme 'beneficiaries', shift the focus from 'product-centered' to 'customer-based' programmes, adopt innovative approaches for security, collateral, business plan evaluation, etc., and strengthen financial systems management is required.

The second is the establishment of training and business development services through integrated approach by skills training, business counseling, and mentor support, access to working space, business expansion support, and creating support

networks. The training should also be extended to service providers in order to improve their professional and technical competence, especially in the areas of programme conception, design, implementation and evaluation. Appropriate and suitable training models need to be adopted.

Moreover, in those countries currently lacking explicit policies on youth, there is urgent need to design compressive national youth policies that indicate the directions a country intends giving to the development of its young people. Significantly, such policies should be properly integrated with key macro and sectoral policies in order to avoid treating youth livelihoods and entrepreneurship as an isolated activity. This has to be done in conjunction with intermediary and support institutions as it has radical implications. Depending on context, there is need to clearly redefining the role(s) of: Government, NGOs, Private Sector, Donor Agencies and Local Community

FUTURE RESEARCH DIRECTIONS

For effective advancement of youth entrepreneurship there are many critical components. Particularly in LDCs, there is also urgent need for a prioritization of scarce resources. While the general view is that poor countries lack sufficient resources to invest in the promotion of youth enterprise development, a review of the literature suggests that even that which is available is not properly used due to lack of priorities and strong accountability mechanisms. There is also very little research on the qualities and particular needs of youth who want to become self-employed in both developed and developing countries. Especially in developing countries, there is currently a dearth of empirical data on the informal sector, and on the participation of youth in the sector and the extent to which the existing policy and institutional framework impacts on youth entrepreneurship. So there is urgency to undertake evaluative studies on the youth enterprise development process, Tracer studies of entrepreneur training programmes,

institutional development needs, diagnosis of a variety of institutions in the public and private sectors, need to be embarked on. Research to identifying opportunities for youth in value chains and high value markets, is another interesting area for further research.

CONCLUSION

Indian is poised for growth, but its success depends on its ability to develop the skills and talents of its greatest asset: young people. Every year, millions of youth become potential entrepreneurs, yet the field—like any other field—is challenged to get people whose skills match its needs. Youth entrepreneurship is also not a panacea for solving the youth unemployment problem but it does have a role to play in facilitating both - employment and entrepreneurship development, hence it can be a part of the response. While context clearly matters, to maximize effectiveness and efficiency, policy should target resources on young people with the best chance of success, provide sufficient support to allow them to start businesses outside the low entry barrier but high competition sectors, and provide integrated packages of complementary support rather than one-shot instruments.

Increasingly, educators and scholarly work finds that for young people to succeed in today's rapidly changing and globalized world, they need an educational foundation that includes more than just academic and technical skills (Baporikar, 2014). Youth critically need life and employability skills as a way to prepare them to be motivated and confident decision-makers who can overcome adversity and realize their potential. Various stakeholders across cross-section indicate that the following core life skills are essential for all entrepreneurs especially youth entrepreneurs to possess, regardless of their enterprise size, location and responsibility level: self-confidence, ability to manage emotions, personal responsibility, respecting self and others, cooperation and teamwork, communication and interpersonal skills,

creative thinking, critical thinking and problem solving, decision-making, conflict management. These can be termed as critical success factors for youth entrepreneurship (Baporikar, 2016, 2017). If policies and programmes for youth entrepreneurship can bring scope and scale to bear on the youth employment challenge, it can fulfill its own needs for an economic development, equity while empowering young people around the world to acquire the means and achieve the stability they need to enter adulthood with renewed hope for a bright future.

REFERENCES

- Baporikar, N. (2013). Entrepreneurship in a Modern Networked Indian Economy. *International Journal of Asian Business and Information Management*, 4(4), 48–66. doi:10.4018/ijabim.2013100104
- Baporikar, N. (2014). *Entrepreneurial Education (Process Of Creating Entrepreneurs)*, Himalaya Publishing House. Mumbai, India: ISBN No.
- Baporikar, N. (2016). Youth Entrepreneurship in India: Outlook and Obstacle. In U. Aung & P. Ordoñez de Pablos (Eds.), *Managerial Strategies and Practice in the Asian Business Sector* (pp. 308–332). Hershey, PA: Business Science Reference; doi:10.4018/978-1-4666-9758-4.ch016
- Baporikar, N. (2017). Critical Review of Entrepreneurship in Oman. In P. Zgheib (Ed.), *Entrepreneurship and Business Innovation in the Middle East* (pp. 147–174). Hershey, PA: IGI Global; doi:10.4018/978-1-5225-2066-5.ch008
- Baum, R. J., Locke, E. A., & Smith, K. G. (2001). A multidimensional model of venture growth. *Academy of Management Journal*, 44(2), 292–303. doi:10.2307/3069456
- Bell, D., & Blanchflower, D. (2009). *What Should Be Done about Rising Unemployment in the UK?* IZA, Discussion Paper No. 4040.
- Bennell, P. (2000). *Improving Youth Livelihoods in SSA: A review of policies and programs with particular emphasis on the link between sexual behavior and economic well-being*. IDRC.
- Blanchflower, D., & Oswald, A. J. (1998). *Entrepreneurship and the Youth Labour Market Problem: A Report for the OECD*. Paris: OECD.
- Chell, E. (2008). *The Entrepreneurial Personality*. Oxford, UK: The Psychology Press. doi:10.4324/9780203938638
- Chigunta, F. (2002). *Youth Entrepreneurship: Meeting the Key Policy Challenges*. Oxford, UK: Oxford University.
- Ciavarella, M. A., Buchholtz, A. K., Riordan, C. M., Gatewood, R. D., & Stokes, G. S. (2004). The Big Five and venture survival: Is there a linkage? *Journal of Business Venturing*, 19(4), 465–483. doi:10.1016/j.jbusvent.2003.03.001
- Curtain, R. (2000). *Towards a Youth Employment Strategy*. United Nations on Youth Employment.
- Gardner, H. (1983). *Frames of mind: the theory of multiple intelligences*. Basic Books.
- Harper, C. (1996). From the Editor. *Small Enterprise Development Journal*.
- Holland, J. L. (1997). *Making Vocational Choices*. Odessa, FL: Psychological Assessment Resources, Inc.
- Hwang, V. W., & Horowitz, G. (2012). *The Rainforest: The Secret to Building the Next Silicon Valley*. Create Space Independent Publishing Platform.
- OECD. (2001). *Putting the Young in Business: Policy Challenges for Youth Entrepreneurship*. OECD.
- Penrose, E. T. (1959). *The Theory of the Growth of the Firm*. New York: John Wiley.
- Schnurr, J., & Newing. (1997). *A Conceptual and Analytical Framework for Youth Enterprise and Livelihood Skills Development: Defining an IDRC Niche*. IDRC.

Shepherd, D. A., Wiklund, J., & Haynie, J. M. (2009). Moving forward: Balancing the financial and emotional costs of business failure. *Journal of Business Venturing*, 24(2), 134–148. doi:10.1016/j.jbusvent.2007.10.002

Simon, H. A. (1947). *Administrative Behavior: A Study of Decision-making Processes in Administrative Organizations*. Chicago, IL: Macmillan.

Splaver, S. (1977). *Your personality and your career*. New York, NY: Julian.

Stone, J.R. (n.d.). *Youth Run Enterprises: Successes, Barriers and Policy Implications*. Academic Press.

White & Kenyon. (2000). *Enterprise-Based Youth Employment Policies*. Geneva: Strategies and Programmes. Report to ILO.

Witt, U. (2000). Changing cognitive frames - changing organizational forms: An entrepreneurial theory of organizational development. *Industrial and Corporate Change*, 9(4), 733–755. doi:10.1093/icc/9.4.733

Zhao, H., Seibert, S. E., & Lumpkin, G. T. (2010). The relationship of personality to entrepreneurial intentions and performance: A meta-analytic review. *Journal of Management*, 36(2), 381–404. doi:10.1177/0149206309335187

ADDITIONAL READING

Baporikar, N. (2011). *Entrepreneurship Development and Project Management*. Mumbai, India: Himalaya Publishing House.

Baporikar, N. (2013). *Entrepreneurship and Small Industry (Theory and Practice)* (p. 7). Mumbai, India: Himalaya Publishing House.

Baporikar, N. (2014). *Entrepreneurial Education (Process Of Creating Entrepreneurs)*. Mumbai, India: Himalaya Publishing House.

Baporikar, N. (2014). Strategic Management Overview and SME in Globalized World. In K. Todorov & D. Smallbone (Eds.), *Handbook of Research on Strategic Management in Small and Medium Enterprises* (pp. 22–39). Hershey, PA: Business Science Reference. doi:10.4018/978-1-4666-5962-9.ch002

Baporikar, N. (2014). Youth Entrepreneurship in Indian Scenario. *International Journal of Asian Business and Information Management*, 5(2), 74–84. doi:10.4018/ijabim.2014040106

Baporikar, N. (2015). Entrepreneurship in Sultanate of Oman: A Case Approach. *International Journal of Asian Business and Information Management*, 6(4), 14–26. doi:10.4018/IJABIM.2015100102

Baporikar, N. (2015). Role of Entrepreneurship in Networked Indian Economy. In P. Ordóñez de Pablos (Ed.), *Technological Solutions for Sustainable Business Practice in Asia* (pp. 36–59). Hershey, PA: Business Science Reference. doi:10.4018/978-1-4666-8462-1.ch003

Baporikar, N. (2016). *Handbook of Research on Entrepreneurship in the Contemporary Knowledge-Based Global Economy* (pp. 1–608). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-8798-1

Baumol, W. J., Litan, R. E., & Schramm, C. J. (2007). *Good Capitalism, Bad Capitalism, and the Economics of Growth and Prosperity*. New Haven: Yale University Press.

Bhide, A. (2000). *The origin and evolution of new businesses*. New York: Oxford University Press.

Bruyat, C. A., & Julien, P. A. (2001). Defining the field of research in entrepreneurship. *Journal of Business Venturing*, 16(2), 165–180. doi:10.1016/S0883-9026(99)00043-9

Chen, Y. F., & Lai, M. C. (2010). Factors influencing the entrepreneurial attitude of Taiwanese tertiary-level business students. *Journal of Social Behavior and Personality*, 38(1), 1–12. doi:10.2224/sbp.2010.38.1.1

Henderson, R., & Robertson, M. (2000). Who Wants to be An Entrepreneur? Young Adult Attitudes to Entrepreneurship as a Career. *Career Development International*, 5(6), 279–287.

Lichtenstein, G. A., & Lyons, T. S. (2010). *Investing in Entrepreneurs: An Approach for Strengthening your Regional and Community Economy*. Santa Barbara, CA: Praeger/ABC-CLIO.

Lüthje, C., & Franke, N. (2003). The Making of an Entrepreneur: Testing A Model of Entrepreneurial Intent among Engineering Students at MIT. *R & D Management*, 33(2), 135–147. doi:10.1111/1467-9310.00288

Smallbone, D., & Welter, F. (2001). The distinctiveness of entrepreneurship in transition economies. *Journal of Small Business Economics*, 16(4), 249–262. doi:10.1023/A:1011159216578

Wang, C., & Wong, P. (2004). Entrepreneurial interest of university students in Singapore. *Technovation*, 24(2), 161–172. doi:10.1016/S0166-4972(02)00016-0

Yonca, G., & Atsan, N. (2006). Entrepreneurial characteristics amongst university students: Some insights for entrepreneurship education and training in Turkey. *Education + Training*, 48(1), 25–38.

Zimmerer, T. W., & Scarborough, N. M. (2008). *Essentials of entrepreneurship and small business management: the foundations of entrepreneurship*. USA: Pearson education.

KEY TERMS AND DEFINITIONS

Critical Success Factors (CSFs): CSF is greatest possible important circumstance, fact, or influence that contributes for the correct or

desired result of an attempt and it is one of the parts that make up a whole or the one that actively contributes to the production of a result or success or failure of something that is aimed to be achieved or attained.

Government: The organization, machinery, or agency through which a political unit exercises authority and performs functions and which is usually classified according to the distribution of power within it. It is a political system by which a body of people is administered and regulated.

Information Technology: The study or use of systems (especially computers and telecommunications) for storing, retrieving, and sending information.

Initiative: Readiness to engage in daring or difficult activity, the condition of being initiated into some experience or sphere of activity: knowledgeableness.

Knowledge: The fact or condition of knowing something with familiarity gained through experience or association, acquaintance with or understanding of a science, art, or technique, the range of one's information or understanding, the circumstance or condition of apprehending truth or fact through reasoning or the fact or condition of having information or of being learned.

Management: The act or activity or process of looking after and making decisions about something.

Models: An exact representation of something in greatly reduced size for better understanding and adoption if found suitable.

Opportunity: A favorable combination of circumstances, time, and place.

Policies: The science and art of employing, a careful plan or method, the art of devising or employing plans or stratagems toward a goal, an adaptation or complex of adaptations (as of behaviour, metabolism, or structure) that serves or appears to serve an important function in achieving evolutionary success.

Process: A natural phenomenon marked by gradual changes that lead toward a particular result, a natural progressively continuing operation or

development marked by a series of gradual changes that succeed one another in a relatively fixed way and lead toward a particular result or end.

Risk: To take a chance on a possible course of action to achieve a desired result.

Stakeholder: A person with an interest or concern in something, especially a business. Stakeholder is a member of a type of organization or system in which as a member or participant seen as having an interest in its success.

E

Entrepreneurship

Mehmet Eymen Eryilmaz
Uludağ University, Turkey

INTRODUCTION

From the past to today, it has been discussed by scholars of various study of fields (for example in strategic management (Barney, 2002; Booth, 1998; Eryilmaz, 2016), in organization theory (Davis & Marquis, 2005) and in business history (Kurt, 2016)) that whether the field is transformed into an academic discipline or not. In a similar vein, some early (e.g. Vesper, 1988) and recent (e.g. George & Wadhwani, 2006; Urban, 2010) studies in the field asserted that entrepreneurship gained status of an academic discipline.¹ During this study, historical background of entrepreneurship discipline will be examined. Then, some discussions and empirical studies on antecedents and consequences of entrepreneurship will be shared with readers. The study will continue with a section that focuses on the link between entrepreneurship and information and communication technologies. Then, the study will give some information on recent developments and possible future trends in the field. Finally, the study will be ended with a conclusion part.

BACKGROUND

A Brief History of Entrepreneurship and Some Definitions

According to some studies (e.g. Stevenson & Jarillo, 1990), the word of entrepreneurship was invented by an Irish-French economists, *Richard Cantillon* (1680-1734). Cantillon derived the concept of “entrepreneur” from French “entreprendre” which may be translated into English as “to undertake” (Matlay, 2005). Besides, Cantillon who

was named by Jevon as “The Cradle of Political Economy” (Hayek, 2005) stated that entrepreneurial activity includes buying from a certain price and the risk that is relating to selling from an uncertain price. In addition, the French economist *Jean Baptiste Say* (1767-1832) extended definition of Cantillon by adding the statement of “bringing factors of production together” (Stevenson & Jarillo, 1990). Another important contributor to the field, English political economist and philosopher, *John Stuart Mill* (1806-1873) asserted that main element that differentiates an entrepreneur from a manager is bearing of risk (Carland *et al.*, 1984). As consistent with this stream, Gartner (1989: 62) conceptualizes entrepreneurship as “the creation of new organizations”. In addition, *Francis Amasa Walker* (1840-1897), an American economist and educator, was another contributor to the field of entrepreneurship. According to him, an entrepreneur was a person who is born with above average talent with respect to organization and coordination of factors of production. According to him, investor who supplies needed funds and receives an interest from uses of them, and entrepreneur who obtains profit from his/her managerial capabilities, are different concepts (Balachandran & Sakthivelan, 2013).

An Austrian-born American economist, *Joseph Alois Schumpeter* (1883-1950), also brought a breath of fresh air into the field. Schumpeter conceptualizes entrepreneurs and entrepreneurship as change agents in an economy (Jones & Wadhwani, 2006) and as a situation respectively that economy totally improves. According to him, there is innovation in the heart of concept of entrepreneurship. In addition, Schumpeter defines enterprise as making new combinations. As a natural consequence of this definition, entrepreneur is

DOI: 10.4018/978-1-5225-2255-3.ch262

the person who creates these new combinations. At this point, he seemed to feel a need to explain the concept of “new combination”. In term of his idea, there can be various new combinations such as 1) putting a new good or service on market, 2) using a new method of production, 3) entering into a new market that doesn’t have information about the good, 4) finding a new source of input (e.g. raw material/half manufactured goods) and finally, 5) changing structure of market by creating or breaking a monopoly in market (or industrial reorganization). For example, in the previous year, a company announced that it will launch to produce electricity from bamboo in Japan (Milliyet, 2015). This development may be accepted as an example of entrepreneurship in a Schumpeterian manner. Schumpeterian stance differentiates business man and entrepreneur from each other (Carland *et al.*, 1984; Stevenson & Jarillo, 1990).

In 1985, an American entrepreneur, *Gifford Pinchot III* (1942 -) coined the concept of “intrapreneurship”. “Intrapreneurship can be defined as the development, within a large corporation, of internal markets and relatively small autonomous or semiautonomous business units that produce products, services, or technologies by employing the firm’s resources in a unique way” (Dollinger, 2008: 384 cited from Hisrich *et al.*, 1985).

As it can be seen above, there are two main-streams in conceptualizations of entrepreneur and entrepreneurship. The first stream strongly links entrepreneurship with behavior of risk taking. In addition, the second stream that Schumpeter takes the lead principally underlines innovative behavior of entrepreneur. There are also some eclectic approaches in the literature that endeavor to combine these two approaches as well. For example, Johnson (2001: 137) defines entrepreneur as “an individual who takes agency and initiative; who assumes responsibility and ownership for making things happen; is both open to and able to create novelty; who manages the risks attached to the process; and who has the persistence to see things through to some identified end-point, even when faced with obstacles and difficulties”. In

a similar manner, for Shane and Venkataraman (2000), entrepreneurship may associate with both new and extant organizations. According to the author of the study, since behaviors of risk taking and innovation are associated with each other and complimentary, a combination of these two streams to conceptualize entrepreneurship seems to be more appropriate. Almost every innovation trial bears a risk. For example, pen phone of Siemens seems to attract less attention than expected.

Antecedents and Consequences of Entrepreneurial Behavior

It seems that there are many triggers of entrepreneurial intent and causes of entrepreneurial success and failure at macro and micro levels. Although it is sometimes criticized by some scholars; some macro level factors such as culture, religion, trust, financial and educational systems, political and legal institutions may have some impacts on entrepreneurship processes (De Clercq *et al.*, 2013; Gohman, 2012; Jones & Wadhwani, 2006; Valdez & Richardson, 2013). For example, De Clercq *et al.* (2013) hypothesized that the relationship between two variables such as individual resources (e.g. human, social and financial capital) and new business activity is moderated by formal institutions (educational and financial systems) and informal institutions (culture and trust). Their findings showed that the hypotheses are partially supported. In a similar vein, according to Hefner, success in business of overseas Chinese may be explained by traditionally strong ties among members of Chinese families and necessity of being successful as a minority group. In a similar manner, Walker showed us in her study that how slavery and institutionalized racism in US before civil war limited entrepreneurial opportunities of African Americans. The same study also indicated that African Americans performed some entrepreneurial activities to delegitimize these institutions (Jones & Wadhwani, 2006 cited from Hefner, 1998 & Walker, 1986). In a similar vein, Turkey is 51st and 56th in “2015-2016 The Global Competitive-

ness Index” and innovation and sophistication factors subbranch of it respectively (Sala-i Martin *et al.*, 2015). According to the author, one of the most important factors for this relative failure in innovation and entrepreneurship is some problems in higher education system.

On the other hand, another stream that mainly takes support from the field of psychology investigates the relationship between micro factors such as personality traits², gender etc. and entrepreneurial intentions and success/failure (Stevenson & Jarillo, 1990). For example, Rauch and Frese (2007) found a positive and significant relationship between entrepreneurial behavior and some personality traits such as generalized self-efficacy, innovativeness, need for achievement, need for autonomy, proactive personality and stress tolerance. In a similar vein, Engelen *et al.* (2015) investigated the relationship between overconfidence of chief executive officer (CEO) and orientation of entrepreneurship and they found a positive relationship that is moderated by market dynamism. Some studies criticize the approach of personality traits to entrepreneurship. According to Gartner (1989), there aren't significant personality differences between entrepreneurs and non-entrepreneurs. Therefore, the correct question isn't “who is an entrepreneur?” to reveal entrepreneurs. Instead, the behavioral stance to entrepreneurship recommends that the question of “what does an entrepreneur do?” should be asked. On the other hand, a meta-analytic research on links between personality traits and entrepreneurial behaviors showed that personality of entrepreneurs is important. Some previous studies couldn't find a direct relationship between personality traits and entrepreneurial intent, behavior or success since some variables mediate to this relationship (Rauch & Frese, 2006). Finally, family environment and child rearing style of a parent (even toys that are given to a child) may affect future success of a child in entrepreneurship as well. For example, in many countries, majority of parents often have an interfering style in their relationships with their children. This style may affect decision making

skills and future entrepreneurial capacity of children in a negative direction.

One group of factor without another one seems to be inadequate. For example, if only macro factors were enough to understand entrepreneurship, how would we explain success differences in processes of entrepreneurship of people who live in same conditions?. However, entrepreneurial behavior in some countries that is supported by some macro level factors is stronger. Therefore, a combined set of factors (macro and micro levels) may be the best alternative to examine and explain entrepreneurial intents and behaviors (Rasmussen & Sorheim, 2006).

Again, entrepreneurial behaviors may have consequences at macro and micro levels. For example, Schumpeter claimed that entrepreneurial activity is the essence of economical development of a country (Carland *et al.*, 1984; Shane & Venkataraman, 2000). In a similar vein, some scholars discuss or empirically support the impact of entrepreneurial activities on competitive advantage of a country (Matlay, 2005). In addition, it may be expected that entrepreneurs may trigger productivity at national level by augmenting competition (van Stel *et al.*, 2005). Finally, various studies assume and empirically support that activities of entrepreneurship have some links with regional and local economic development as well (e.g. Baptista *et al.*, 2005; Malecki, 1993; Rasmussen & Sorheim, 2006). At a micro level, one of the possible consequences of a successful entrepreneurial activity is personal accumulation of wealth.³ Indeed, wealth accumulation is used in academic research as an indicator of entrepreneurial success (Jones & Wadhvani, 2006). Naturally, entrepreneurial behaviors may generate good or bad personal reputation for their owners.

Information and Communication Technologies and Entrepreneurship

“Information and Communication Technologies (ICT)” can be defined as ‘electronic means of capturing, processing, storing, and communicating

information” (Heeks, 1999: 3). ICTs may be input, output or a part of entrepreneurship processes. As an input, ICT may be used to increase entrepreneurial competencies. For example, Sinkovics *et al.* (2004) suggest that candidates of entrepreneurship need three different competencies such as intra-personal, inter-personal and organizational competencies to be successful. Some ICT tools such as mobile phones, internet discussion groups, video conferencing and computer simulations can enable competencies to be gained by students. According to Sinkovics *et al.* (2004), uses of ICT tools in international entrepreneurship education provide some benefits to students such as being familiar to values of counterparts in other countries and being familiar to new ICT tools. In addition, ICTs can create new business areas for entrepreneurs. For example, one of the newest concepts associated with ICT is “nomophobia”. According to a definition, “nomophobia is the modern fear of being unable to communicate through a mobile phone (MP) or the internet” (Yıldırım, 2014: 8 cited from King *et al.*, 2014: 28). Some health centers were founded in United States to treat addiction of people to mobile phones (Karahasan, 2012).

In addition, ICTs may be a part of entrepreneurial process (or production technology). Extant entrepreneurs can access data and information by ICT that they need to increase their performance. For example, a foreign entrepreneur who plans to enter into Turkish market can collect data and information about Turkish market and consumer behaviors from web. In parallel with this example, after a literature review, Ongori and Migiro (2010) assert that uses of ICT ease to get into international markets of small and medium sized organizations. ICTs are used by not only business organizations but also social organizations. For example, many aiding organizations in Africa obtain benefits from ICTs (Molony, 2007). Therefore, ICTs may be beneficial tools for social entrepreneurship as well.

Finally, ICT may be output of an entrepreneurial process as well. Some organizations may choose to produce either tangible (components,

computers and networks) or intangible (software, web pages) ICTs as an output (Heeks, 1999). WhatsApp that were bought by Facebook in 2014 is a good example of ICT as an output of entrepreneurial processes.

FUTURE RESEARCH DIRECTIONS

Continuous technological developments in internet economy cause emergence of new concepts in the field of entrepreneurship such as “e-entrepreneurship” and “netpreneurship”. Today, everything slips back to virtual environment. Wars, loves and so many things are experienced by people in virtual environments now. Therefore, emergence of concepts such as netpreneurship in the field of entrepreneurship is natural. Netpreneurs, they are sometimes called as “Ontrepreneurs”, are entrepreneurs who have ability to run a business largely on internet (Balachandran & Sakthivelan, 2013). At this point, it can be investigated that whether candidates of netpreneurship need different psychological traits and educational backgrounds from candidates of traditional entrepreneurship or not. In addition, generation of people seems to be another important factor that may determine netpreneurship tendency of people. Therefore, it can be researched that whether there are significant differences between members of generation X and Y in terms of tendency of being netpreneur.

In addition to the relationships of entrepreneurship and ICT, the field of entrepreneurship contains some other points of development as well. When the “Management and Organization Studies (MOS)” literature is examined, it can be observed easily that although a limited sensitivity on context has emerged recently (for example in subfield of organizational behavior (Johns, 2006; Morgeson *et al.*, 2010)), the literature is largely context free. A similar evaluation is valid for the literature of strategic entrepreneurship as well. Strategic entrepreneurship is a popular subfield that combines competitive advantage seeking behavior of strategic management and opportunity

seeking behavior of entrepreneurship to reach desirable financial performance (Ireland *et al.*, 2003). On the other hand, scholars of the field seem to overwhelmingly focus on activities of strategic entrepreneurship in developed countries. However, the findings of some previous studies showed us that behaviors of strategic entrepreneurship in emerging economies have unique characteristics (Bruton *et al.*, 2013). Therefore, an increase in number of strategic entrepreneurship studies that are conducted in emerging countries⁴ will be beneficial to understand this context and its effects on strategic entrepreneurship better.

Another recent trend in the field of entrepreneurship seems to make a strong emphasis on emotions (e.g. Cardon *et al.*, 2012; Podoynitsyna *et al.*, 2012; Wolfe & Shepherd, 2015). The studies of entrepreneurship that focus on cognitive side of decision making processes largely has overlooked impact of affect until recently. However, some scholars recently assert that entrepreneurship isn't only a cognitive act (Klein *et al.*, 2013) and affects and cognition often shape processes of decision making jointly.⁵ In parallel with this idea, Hayton *et al.* (2012) proposes that affects foster working memory that stores information that is used as an infrastructure during entrepreneurial decisions. Therefore, studies that discuss and investigate the roles of positive and negative affective situations on various steps of entrepreneurial process will be interesting. In addition, face to face relationships of entrepreneurs often ease to reflect emotions. At this point, consequences of virtual work environment on moods of entrepreneurs may be investigated.

In social sciences, it is often mentioned that a linguistic turn (Alvesson & Kärreman, 2000) is experienced. With this turn, language seems to regain its importance that it lost in periods of Enlightenment and Romanticism (Bonet & Saquet, 2010). According to the view, language isn't solely a basic tool to exchange ideas. It is also an instrument to build the reality. One of the sub branches of this view is studies that focus on rhetoric. Rhetoric is "the ability of 'seeing' the available means of persuasion" in terms of Aristotle (1991:

13). Entrepreneurs need to persuade themselves first and then, some other different actors such as families, inner circles and investors (such as angel investors and venture capitals). Even though various studies (e.g. Brooks *et al.*, 2014; Chen *et al.*, 2009) have recently emerged in the literature that endeavor to link entrepreneurship with rhetoric, number of these studies is still so limited. In addition to these, another research question may be in this point that how and when will ICTs affect rhetorical capacities of entrepreneurs?.

Further, neurosciences are just beginning to create impact on the social sciences (Martin de Holan, 2014). In addition, there is a strong interest (both sides from critical and proponent)⁶ towards neurosciences in MOS field. Many studies (e.g. Lindebaum, 2016; McDonald & Tang, 2014) that have been published in prestigious MOS journals seem to support this claim to a large extent. Some of these studies take a stand on proponents' side; others express their doubts about neuroscience based MOS studies. For example, Lindebaum and Zundel (2013) profess that some MOS researchers who try to link MOS and neurosciences often adopt a very reductionist stance and endeavor to explain some very complex phenomena (such as leadership, management development) with quite simple mechanisms. In a similar vein, entrepreneurship researchers appear to begin fostering sympathy to the mind and its workings. However, entrepreneurship researchers have largely focused on attributes and behavior of entrepreneurs until very recently instead of what and how entrepreneurs think and which factors direct entrepreneurs to think in that ways (Martin de Holan, 2014). Therefore, neuroscientific studies on entrepreneurship that will take some limitations of using views of neurosciences in social sciences into account will be interesting.

Finally, MOS has focused on couple who are working together in a common workplace. Recently, some part of the literature of entrepreneurship has begun to show interest to copreneurs. According to Greenhaus and Callanan (2006), copreneurs are couples who focus on the same entrepreneurial activities. At this point, some

studies that explore antecedents and (positive and negative) consequences of copreneurship may be attractive.

CONCLUSION

In this study, the relationships between two popular concepts, “entrepreneurship” and “information and communication technologies”, were discussed. According to the study, ICTs can be input, output or as a part of entrepreneurial processes. On the other hand, when ICTs are used in a negative way, they can be harmful for entrepreneurial success as well. Therefore, appropriate uses of ICTs must be guaranteed.

REFERENCES

- Alvesson, M., & Kärreman, D. (2000). Taking the linguistic turn in organizational research: Challenges, responses, consequences. *The Journal of Applied Behavioral Science*, 36(2), 136–158. doi:10.1177/0021886300362002
- Aristotle,. (1991). *Aristotle on Rhetoric: A Theory of Civic Discourse* (G. A. Kennedy, Trans.). New York: Oxford University Press.
- Balachandran, V., & Sakthivelan, M. S. (2013). Impact of information technology on entrepreneurship (e-entrepreneurship). *Journal of Business Management & Social Sciences Research*, 2(2), 51–56.
- Baptista, R., Escaria, V., & Madruga, P. (2005). *Entrepreneurship, regional development and job creation: the case of Portugal*. Max Planck Institute Discussion Papers on Entrepreneurship, Growth and Public Policy. Retrieved 28 March, 2016 from <https://papers.econ.mpg.de/egp/discussionpapers/2005-06.pdf>
- Barney, J. (2002). Strategic management: From informed conversation to academic discipline. *The Academy of Management Executive*, 16(2), 53–57. doi:10.5465/AME.2002.7173521
- Bonet, E., & Sauquet, A. (2010). Rhetoric in management and in management research. *Journal of Organizational Change Management*, 23(2), 120–133. doi:10.1108/09534811011031283
- Booth, C. (1998). The problems and possibilities of reflexivity in strategy. *Electronic Journal of Radical Organisation Theory*, 4(1). Retrieved January 7, 2009 from http://merlin.mgmt.waikato.ac.nz/ejrot/Vol4_1/booth.pdf
- Brooks, A. W., Huang, L., Kearney, S. W., & Murray, F. E. (2014). Investors prefer entrepreneurial ventures pitch by attractive men. *Proceedings of the National Academy of Sciences of the United States of America*, 111(12), 1–5. doi:10.1073/pnas.1321202111 PMID:24616491
- Bruton, G. D., Filatotchev, I., Si, S., & Wright, M. (2013). Entrepreneurship and strategy in emerging economies. *Strategic Entrepreneurship Journal*, 7(3), 169–180. doi:10.1002/sej.1159
- Cardon, M.S., Foo, M., Shepherd, D. & Wiklund, J. (2012, January). Exploring the hearth: entrepreneurial emotion is a hot topic. *Entrepreneurship Theory and Practice*, 1-10.
- Carland, J. W., Hoy, F., Boulton, W. R., & Carland, J. A. C. (1984). Differentiating entrepreneurs from small business owners: A conceptualization. *Academy of Management Review*, 9(2), 354–359.
- Chen, X., Yao, X., & Kotha, S. (2009). Entrepreneur passion and preparedness in business plan presentations: A persuasion analysis of venture capitalists funding decisions. *Academy of Management Journal*, 52(1), 199–214. doi:10.5465/AMJ.2009.36462018

- Davis, G. F., & Marquis, C. (2005). Prospects for organization theory in the early twenty-first century: Institutional fields and mechanisms. *Organization Science*, 16(4), 332–343. doi:10.1287/orsc.1050.0137
- De Clercq, D., Lim, D.S.K. & Oh, C.H. (2013, March). Individual-level resources and new business activity: the contingent role of institutional context. *Entrepreneurship Theory and Practice*, 303-330.
- Dollinger, M.J. (2008). *Entrepreneurship: Strategies and Resources* (4th ed.). Marsh Publications.
- Engelen, A., Neumann, C. & Schwens, C. (2015, September). Of course I can: the effect of ceo overconfidence on entrepreneurially oriented firms. *Entrepreneurship Theory and Practice*, 1137-1160.
- Eryılmaz, M. E. (2016). Bir araştırmacının naçizane gözlemleri: İlgili dergilerin editör ve hakemleri stratejik yönetim alanı yazarlarından ne beklerler? [Humble reflections of a researcher: what do editors and reviewers of relevant journals expect from authors of strategic management field?]. *Turkish Journal of Management*, 1(1), 9–22.
- Gartner, W.B. (1989, Summer). “Who is an entrepreneur?” is the wrong question. *Entrepreneurship Theory and Practice*, 47-67.
- Gohman, S.F. (2012, March). Institutions, latent entrepreneurship, and self-employment: an international comparison. *Entrepreneurship Theory and Practice*, 295-321.
- Green, S. E. (2004). A rhetorical theory of diffusion. *Academy of Management Review*, 29(4), 653–669.
- Greenhaus, J.H., & Callanan, G.A. (2006). *Encyclopedia of Career Development*. Thousand Oaks, CA: Sage Publications. doi:10.4135/9781412952675
- Hayek, F. A. (2005). *The Trend of Economic Thinking: The Essays of Political Economists and Economic History*. Routledge.
- Hayton, J.C. & Cholakova, M. (2012, January). The role of affect in the creation and intentional pursuit of entrepreneurial ideas. *Entrepreneurship Theory and Practice*, 41-67.
- Healey, M. P., & Hodgkinson, G. P. (2014). Rethinking the philosophical and theoretical foundations of organizational neuroscience: A critical realist alternative. *Human Relations*, 67(7), 765–792. doi:10.1177/0018726714530014
- Heeks, R. (1999). *Information and communication technologies, poverty and development*. Institute for Development Policy and Management. Working Paper Series, Paper No. 5. Retrieved 11 June, 2016 from, http://unpan1.un.org/intradoc/groups/public/documents/NISPAcee/UNPA_N015539.pdf
- Ireland, R. D., Hitt, M. A., & Sirmon, D. G. (2003). A model of strategic entrepreneurship: The construct and its dimensions. *Journal of Management*, 29(6), 963–989. doi:10.1016/S0149-2063(03)00086-2
- Johns, G. (2006). The essential impact of context on organizational behavior. *Academy of Management Review*, 31(2), 386–408. doi:10.5465/AMR.2006.2020867
- Johnson, D. (2001). What is innovation and entrepreneurship? Lessons for larger organisations. *Industrial and Commercial Training*, 33(4), 135–140. doi:10.1108/00197850110395245
- Jones, G., & Wadhwani, R. D. (2006). *Entrepreneurship and business history: renewing the research agenda*. Harvard Business School Working Papers, 07-007. Retrieved 26 March, 2016 from <http://www.hbs.edu/faculty/Publication%20Files/07-007.pdf>
- Karahasan, F. (2012). Siz de nomofobi kurbanı olabilirsiniz. *Milliyet*. Retrieved 10 July, 2015 from <http://www.milliyet.com.tr/siz-de-nomofobi-kurbani-olabilirsiniz/ekonomi/ekonomiyazardetay/13.12.2012/1641067/default.htm>

- Kerr, W. R., & Nanda, R. (2011). Financing constraints and entrepreneurship. In D. B. Audretsch, O. Falck, S. Heblich, & A. Lederer (Eds.), *Handbook of Research on Innovation and Entrepreneurship* (pp. 88–103). Cheltenham, UK: Edward Elgar. doi:10.4337/9781849807760.00015
- Klein, P. G., Mahoney, J. T., McGahan, A. M., & Pitelis, C. N. (2013). Capabilities and strategic entrepreneurship in public organizations. *Strategic Entrepreneurship Journal*, 7(1), 70–91. doi:10.1002/sej.1147
- Klotz, A.C. & Neubaum, D.O. (2016, January). Research on the dark side of personality traits in entrepreneurship: observations from an organizational behavior perspective. *Entrepreneurship Theory and Practice*, 7-17.
- Kurt, M. (2016). Bir disiplin olarak işletmecilik tarihi: Doğuşu, gelişimi, Türkiye’de potansiyeli ve zorlukları üzerine [Business history as a discipline: on its origins, progress, prospects and challenges in Turkey]. *Turkish Journal of Management*, 1(1), 41–57.
- Lindebaum, D., & Zundel, D. (2013). Not quite a revolution: Scrutinizing organizational neuroscience in leadership studies. *Human Relations*, 66(6), 855–877. doi:10.1177/0018726713482151
- Lindebaum, K. (2016). Critical essay: Building new management theories on sound data? the case of neuroscience. *Human Relations*, 69(3), 537–550. doi:10.1177/0018726715599831 PMID:27041766
- Malecki, E. J. (1993). Entrepreneurship in regional and local development. *International Regional Science Review*, 16(1/2), 119–153. doi:10.1177/016001769401600107
- Martin de Holan, P. (2014). Its all in your head: Why we need neuroentrepreneurship. *Journal of Management Inquiry*, 23(1), 93–97. doi:10.1177/1056492613485913
- Matlay, H. (2005). Researching entrepreneurship and education: Part 1: what is entrepreneurship and does it matter?. *Education + Training*, 47(8/9), 665–677.
- McDonald, P., & Tang, Y. (2014). Neuroscientific insights into management development: Theoretical propositions and practical implications. *Group & Organization Management*, 39(5), 475–503. doi:10.1177/1059601114550712
- Miller, D. (2015, January). A downside to the entrepreneurial personality?. *Entrepreneurship Theory and Practice*, 1-8.
- Milliyet. (2015). *Japonlar Bambu Ağacından Elektrik Üretecek*. Retrieved 5 August, 2015 from <http://www.milliyet.com.tr /japonlar-bambu-agacindan-elektrik/ekonomi/detay/2092132/default.htm>
- Molony, T. (2007). I dont trust the phone; it always lies: Trust and information and communication technologies in Tanzanian micro-and small enterprises. *Informational Technologies and International Development*, 3(4), 67–83. doi:10.1162/itid.2007.3.4.67
- Morgeson, F. P., Dierdorff, E. C., & Hmurovic, J. L. (2010). Work design *in situ*: Understanding the role of occupational and organizational context. *Journal of Organizational Behavior*, 31(2-3), 351–360. doi:10.1002/job.642
- Ongori, H., & Migiro, S. O. (2010). Information and communication technologies adoption in SMEs: Literature review. *Journal of Chinese Entrepreneurship*, 2(1), 93–104. doi:10.1108/17561391011019041
- Podoyntsina, K., Van Der Bij, H. & Song, M. (2012, January). The role of mixed emotions in the risk perception of novice and serial entrepreneurs. *Entrepreneurship Theory and Practice*, 115-140.
- Rasmussen, E. A., & Sorheim, R. (2006). Action-based entrepreneurship education. *Technovation*, 26(2), 185–194. doi:10.1016/j.technovation.2005.06.012

- Rauch, A., & Frese, M. (2006). Meta-analysis as a tool for developing entrepreneurship research and theory. *Entrepreneurship: Frameworks and Empirical Investigations from Forthcoming Leaders of European Research. Advances in Entrepreneurship, Firm Emergence and Growth*, 9, 29–51. doi:10.1016/S1074-7540(06)09003-9
- Rauch, A., & Frese, M. (2007). Lets put the person back into entrepreneurship research: A meta-analysis on the relationship between business owners personality traits, business creation, and success. *European Journal of Work and Organizational Psychology*, 16(4), 353–385. doi:10.1080/13594320701595438
- Sala-i-Martin, X., Crotti, R., Battista, A., Hannon, M. D., Galvan, C., Geiger, T., & Marti, G. (2015). Reaching Beyond the New Normal: 3 Findings from the Global Competitiveness Index 2015–2016. In *The Global Competitiveness Report 2015-2016*. Retrieved 11 June, 2016 from http://www3.weforum.org/docs/gcr/2015-2016/Global_Competitiveness_Report_2015-2016.pdf
- Shane, S., & Venkataraman, S. (2000). The promise of entrepreneurship as a field of research. *Academy of Management Review*, 25(1), 217–226.
- Sinkovics, R. R., Bell, J., & Deans, K. R. (2004). Using information communication technology to develop international entrepreneurship competencies. *Journal of International Entrepreneurship*, 2(1/2), 125–137. doi:10.1023/B:JIEN.0000026909.49686.a5
- Stevenson, H. H., & Jarillo, J. C. (1990). A paradigm of entrepreneurship: Entrepreneurial management. *Strategic Management Journal*, 11, 17–27.
- Urban, B. (2010). Entrepreneurship as a discipline and field of study. In B. Urban (Ed.), *Frontiers in Entrepreneurship* (pp. 33–62). Springer. doi:10.1007/978-3-642-04502-8_2
- Valdez, M.E. & Richardson, J. (2013, September). Institutional determinants of macro-level entrepreneurship. *Entrepreneurship Theory and Practice*, 1149-1175.
- van Stel, A., Carree, M., & Thurik, R. (2005). *The effect of entrepreneurial activity on national economic growth*. Max Planck Institute Discussion Papers on Entrepreneurship, Growth and Public Policy. Retrieved 28 March, 2016 from <http://www.econstor.eu/bitstream/10419/19995/1/2005-04.pdf>
- Vesper, K. H. (1988). Entrepreneurial academics-how can we tell when the field is getting somewhere? *Journal of Business Venturing*, 3(1), 1–10. doi:10.1016/0883-9026(88)90026-2
- Wassener, B. (2010). New term for emerging economies is suggested. *New York Times*. Retrieved 27 March, 2016 from http://www.nytimes.com/2010/11/16/business/global/16eagles.html?_r=0
- Wolfe, M.T. & Shepherd, D.A. (2015, July). What do you have to say about that? Performance events and narratives' positive and negative emotional content. *Entrepreneurship Theory and Practice*, 895-925.
- Yıldırım, C. (2014). *Exploring the Dimensions of Nomophobia: Developing and Validating a Questionnaire Using Mixed Methods Research* (Master's Thesis). Iowa State University. Retrieved 11 June, 2016 from <http://lib.dr.iastate.edu/cgi/viewcontent.cgi?article=5012&context=etd>

ADDITIONAL READING

- Alvarez, S. A., & Barney, J. B. (2007). Discovery and creation: Alternative theories of entrepreneurial action. *Strategic Entrepreneurship Journal*, 1(1/2), 11–26. doi:10.1002/sej.4

Amit, R., Glosten, L., & Muller, E. (1993). Challenges to theory development in entrepreneurship research. *Journal of Management Studies*, 30(5), 815–833. doi:10.1111/j.1467-6486.1993.tb00327.x

Bailetti, T. (2012). Technology entrepreneurship: overview, definition, and distinctive aspects. *Technology Innovation Management Review*, February, 5-12.

Bird, B. (1988). Implementing entrepreneurial ideas: The case for intention. *Academy of Management Review*, 13(3), 442–453.

Crant, J. M. (1996). The proactive personality scale as a predictor of entrepreneurial intentions. *Journal of Small Business Management*, 34(3), 42.

Guth, W. D., & Ginsberg, A. (1990). Introduction to corporate entrepreneurship. *Strategic Management Journal*, 11, 5–15.

Heeks, R. (1998). *Management Information and Management Information Systems*. London: School of Oriental and African Studies, University of London.

Kollman, T., Kuckertz, A., & Stöckman, C. (2010). *E-Entrepreneurship and ICT Ventures: Strategy, Organization and Technology*. Hershey, New York: IGI Global. doi:10.4018/978-1-61520-597-4

Kuriyan, R. (2008). *Outsourcing Development: The State, Entrepreneurship, and Information Technologies in India*. California. Berkeley: University of California.

Mumba, D. (2014). Empowering Women Entrepreneurs in Africa: Investigating Information Access and Use of Information and Communication Technologies by Women-Owned Enterprises in Zambia. Canada, Ontario: The University of Western Ontario.

Naude, W. (2013). Entrepreneurship and economic development: theory, evidence and policy. *IZA Discussion Paper*, No. 7507. Retrieved 11 July, 2016 from <http://ftp.iza.org/dp7507.pdf>

KEY TERMS AND DEFINITIONS

Copreneurs: Couples who focus on the same entrepreneurial activities (Greenhaus & Callanan, 2006).

Entrepreneurship (In a Classical Manner): Associated with bringing factors of production together and taking the risk that is stemmed from this activity (Stevenson & Jarillo, 1990).

Information and Communication Technologies (ICT): “Electronic means of capturing, processing, storing, and communicating information” (Heeks, 1999: 3).

Netpreneurs: People who have an ability to operate a business on internet to a great extent (Balachandran & Sakthivelan, 2013).

Nomophobia: A fear of being unable to communicate through a mobile phone or the internet” (Yıldırım, 2014: 8 cited from King *et al.*, 2014: 28).

Ontpreneurs: Entrepreneurs who are working online. This concept is often accepted as synonym with netpreneur by many scholars in the field of entrepreneurship.

Schumpeterian Entrepreneurship: Making new combinations.

ENDNOTES

¹ At this point, there are also some studies (e.g. Rauch & Frese, 2006) in the literature that have some doubts about whether the field of entrepreneurship is a mature discipline or not.

² This stream seems to attribute a positive meaning to personality to a large extent. However, there are some studies (e.g. Klotz & Neubaum, 2016) that attract our attentions to the darker sides of personality (such as

Machiavellianism, psychopathy, narcissism etc.) and their relationship with entrepreneurial consequences. In addition, some scholars (e.g. Miller, 2015) discuss that some personality traits (such as being energetic and high self confidence) that seems to be positive for entrepreneurship at the first point may transform into negative ones (such as aggressiveness and narcissism) later.

³ In effect, there also seems to be a reversed relationship between these variables as well. For instance, Kerr and Nanda (2011) discuss that a desire for personal wealth may trigger the choice of being an entrepreneur.

⁴ According to the Spanish Bank BBVA, emerging countries isn't a monolithic group.

BBVA tends to divide emerging countries into subgroups such as "EAGLEs" and "NEST" (Wassener, 2010). Therefore, research in emerging countries may be conducted in terms of these subgroups.

⁵ In addition, entrepreneurs often experience a mixed of emotions during their decision making processes instead of experiencing only one emotion (Podoyntsyna *et al.*, 2012).

⁶ In effect, there is one another position in MOS field about neurosciences. Critical realist position (e.g. Healey & Hodgkinson, 2014) seems to adopt an intermediate stance in this heated discussion.

Entrepreneurship as the Vantage Point

E

Madhu Kishore Raghunath Raghunath Kamakula
GVP College of Engineering (Autonomous), India

Chandra Sekhar Patro
GVP College of Engineering (Autonomous), India

INTRODUCTION

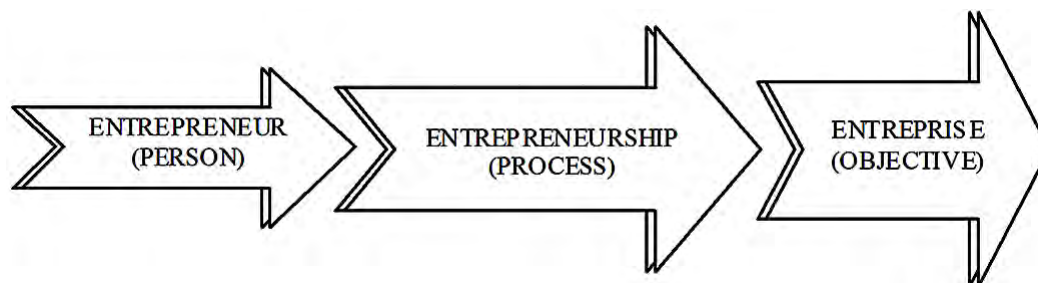
Entrepreneurship is the dynamic process of creating incremental wealth. This wealth is created by individuals who assume the major risks in terms of equity, time, and/or career commitment of providing value for some product or service. The product or service itself may or may not be new or exceptional but value must somehow be infused by an entrepreneur (Shapero, 1975). Entrepreneurship is an elusive concept and has been defined differently by different authors. Some of them view it as 'risk-bearing'; others call it as 'innovations', and yet others consider it as 'thrill seeking'. Today, an entrepreneur is an innovator or developer who recognizes and seizes opportunities; converts those opportunities into workable/marketable ideas; adds value through time, effort, money, or skills; assumes the risks of the competitive marketplace to implement these ideas; and realizes the rewards from these efforts.

An entrepreneur is the aggressive catalyst for change in the world of business. He or she is an independent thinker who dares to be different in a

background of common events. The characteristics are; personal initiative, the ability to consolidate resources, management skills, a desire for autonomy, and risk taking. Other characteristics include aggressiveness, competitiveness, goal-oriented behaviour, confidence, opportunistic behaviour, intuitiveness, reality-based actions, the ability to learn from mistakes, and the ability to employ human relations skills. Entrepreneurship is an essential element for economic progress as it manifests technology in different ways: a) by identifying, assessing and exploiting business opportunities; b) by creating new firms and/or renewing existing ones by making them more dynamic; and c) by driving the economy forward - through innovation, competence, job creation - and by generally improving the wellbeing of society.

The accessibility of Information technology is a necessity for any entrepreneurial activity. Today technology has its impact on individuals and communities as entrepreneurship does, and hence IT cannot be neglected. In the past two decades, advanced industrial countries with new technology have become driving force of entrepreneur-

Figure 1. Concept of entrepreneurship



DOI: 10.4018/978-1-5225-2255-3.ch263

ship creating economic development (Tavakoli, 2013). Given that entrepreneurship and impact of technology on it have evolved into a distinctive force, which is in line with changes in business optimization.

BACKGROUND

The entrepreneurial revolution has taken hold across the globe and has undeniably impacted the world of business. The younger generation of the 21st century is becoming the most entrepreneurial generation since the Industrial Revolution. There are various personal characteristics, cognitions, and social conditions that affect an individual's choice to pursue entrepreneurial activities. Greenberger & Sexton (1988) discussed five reasons why an entrepreneur wants to start a business. First, they would like to take a chance in the market. Second, they believe their managerial skills are more efficient than those of other people. Third, they believe that their specialty could be developed into an enterprise. Fourth, they have already developed a product or service and they believe that can find a niche in the market. Fifth, this is the only way to start an enterprise and their other options are limited. Ghosh & Kwan (1996) found that there were several reasons that these entrepreneurs started enterprises namely, the individual wants personal growth; they like the challenge; there is more freedom in being a business owner; chance to use one's own knowledge and experience; do not like working for others; they are affected by family or friends; and family tradition. Carter, Gartner, Shaver & Gatewood (2003) explored and compared the reasons that nascent entrepreneurs and non-entrepreneurs to start enterprises and make career choices. The results of indicated self-realization, financial success, roles, innovation, recognition and independence are the main reasons to start enterprises. Wang & Wong (2004) found gender, family experience with business and educational level are significant factors in entrepreneurial interests. Although the concept of

entrepreneurial competencies is used widely by government agencies and others in their drive for economic development and business success, the core concept of entrepreneurial competencies, its measurement and its relationship to entrepreneurial performance and business success is in need of further rigorous research and development in practice (Mitchelmore & Rowley, 2010).

The entrepreneurship approaches are important aspect or components of entrepreneurship development as well as entrepreneurial environment which should be considered, they contribute significantly to the entrepreneurial development success and as well as that of the success of the entrepreneurs. These approaches once again include; ethnicity, location, women, religious, socio-cultural, and youths (Lucky, 2011). Therefore, if these entrepreneurial approaches are being given urgent consideration, they are capable of changing the entrepreneurial scenario of any country and thereby serving as a driver and motivating force in achieving entrepreneurial success. Naudé *et al* (2009) empirically identified the determinants of start-up rates across different sub-national regions and in particular to investigate the role of access to finance on a regional (sub-national) level on start-ups. They find that the most important determinants of start-up rates are profit rates, educational levels, agglomeration as measured by the economic size of a district, and access to formal bank finance. It is also identified that the existence of 'congesting' factors such as increased competition, tougher barriers to entry, monopolistic behavior, and a greater difficulty to be innovative and novel are also influencing. During the last two decades, the development of new technologies and emergence of new business models has enabled the shift from large corporations to small and new ventures (Acs & Audretsch, 1988). Most entrepreneurs simply got tired of working for others, had a great idea they wanted to commercialize, or woke up one day with an urgent desire to build wealth before they retired. So they took the big leap." The country's economic policy environment must be favourable for organizations

to achieve efficiencies in today's global market (Santhi, 2011). It should enable the entrepreneurs to provide a magical touch to an organization, whether in public or private or joint sector, in achieving speed, flexibility, innovativeness, and a strong sense of self-determination. They bring a new vision to the forefront of economic growth of a country.

CHALLENGES FACED BY AN ENTREPRENEUR

In this modern era nothing can be earned easily and the same goes with success as an entrepreneur. Entrepreneurship life is not as easy as it seems the way one sees them in magazines showing them enjoying all facets of life happily. Success waits for those who are ready to face the hurdles and overcome them someday as life goes on (Pride, 2014). The following will make one understand the challenges that an entrepreneur always has to be prepared to face and overcome (Morris, 2014).

1. **Selection the Right Business:** Most of the businesses fail for the reason that they are not real businesses. To do great business (product or service) one should have a great idea. If the business finds a right customer, but the idea chosen to reproduce is illogical everything falls apart. So, every entrepreneur's first challenge lies picking up the right start to go ahead and survive in the market.
2. **Financing the Business:** Financing has always been a key area to be considered for an entrepreneur starts an enterprise. It is the lifeblood of every business organisation, so an entrepreneur should be smart enough to gather funds from different sources, in prior and every entrepreneur should have proper view of where his investments will flow from, which will not hamper the continuity of business.
3. **Finding the Customers:** Finding the right customer is the one most decisive challenge that crops up during the establishment of an enterprise. Enterprises have to be on their toes, as customer is considered the god, and without customers a business is a dead rubber which hurts an enterprise, and right product and service is the solution for finding the right customer. Once an enterprise has the right product or service, customers will grow automatically.
4. **When to Quit a Day Job:** When one plans to start a business, the ultimate question to be answered is to know the right time to quit job. An entrepreneur starts a business just because he wants to, one need to wait for the right time and then take a step ahead. The right time generally is when an individual have a full proof business opportunity or have enough money in the bank that can work on business.
5. **Dealing With Stress for Irregular Incomes:** There is no other way to face off; it's extremely stressful when certainty of business payment is not regular. Living life like that can be extremely stressful, so an entrepreneur need to have the mental fortitude to power on, keep working, and never give up.
6. **Dealing With Competition:** The biggest challenge that comes before every entrepreneur is to deal with stiff competition at every step. Entrepreneurs will find businesses with similar product and trying to keep up with the challenge from the competitor become need of the hour to reap benefits and the best way to flush through this challenge is by maintaining a standard and by grading products which would benefit customers better than competitor's.
7. **Hiring Employees:** Though an enterprise owns the finest technology, encompass the best layout and of course invest a lot of capital, nothing works out unless the business have required manpower to operate them

or to define minutely the force behind the organisational success.

8. **Cash and Time Management:** Management of cash flow is essential for the survival of a business enterprise. At times, many entrepreneurs struggle to pay the creditors, employees, loans, etc. In the same manner, time management might also be the biggest problem faced by the entrepreneurs. If the entrepreneur manages both time and cash simultaneously, the targets could be accomplished.
9. **Marketing Strategy:** Strategy to go with is a question which arises in different phases of a business and this is the next big hurdle which has to be planned in advance as the right strategy sets you up for success. Speaking of a perfect marketing strategy, it can be devised by having a clear awareness about the stage of a product or service in its lifecycle.
10. **Business Expansion:** When an entrepreneur starts an enterprise and is running successfully accomplishing the objectives, after a while a question squeezes him i.e. what further can be achieved and of course why not as a dream never stops at one stage. Entrepreneurs can solve this problem by creating new processes that focus on task delegation. Many entrepreneurs do this at one point or the other that is to delegate the task to subordinates when it needs to be.

MEASURES OF ENTREPRENEURIAL ORIENTATION

A vast majority of scholars agree with the view that the degree of Corporate Entrepreneurship can be measured by three dimensions: innovativeness, proactiveness and risk-taking, as mentioned by (Covin & Slevin, 1991; Miller & Friesen, 1983). However some authors, such as Lumpkin and Dess (1996) argue that five dimensions, should be used to measure entrepreneurship; namely autonomy,

competitive aggressiveness, proactiveness, innovativeness and risk-taking. Kreiser, Marino, and Weaver (2002) suggested including growth orientation as the fifth, independent measurement of entrepreneurial management. Based on the proposition of different authors, the descriptions of each of these dimensions are discussed in detail.

Autonomy

Autonomy refers to the independent action of an individual or a team in bringing forth an idea or a vision. In general, it means the ability and will to pursue opportunities, even though factors such as resource availability, actions by competitive rivals, or internal organizational considerations may change the course of the initiative, but not sufficient to extinguish it (Lumpkin & Dess, 1996). Modern firms are increasingly encouraging entrepreneurship at all levels of the organization. To foster entrepreneurial attitudes and behaviour, managers must give significant discretion to employees. Nyström (1979) described it is principally a decentralized, curious and open-minded organization culture that enables firms to meet the challenge of discovering and forming new possibilities and application areas. Corporations do not carry out their innovation activities in isolation of their research labs, but building and tightening the co-operation with their consumers or even competitors have become ever important (Christensen & Raynor, 2003).

Innovativeness

Based on Schumpeter's concept of entrepreneurship, innovativeness refers to the creation of new products, services, processes, technologies and business models. Economically, innovation is the combination of resources in a new and original way. Entrepreneurially, it is the discovery of a new and better way of doing things. Kreiser, Marino, & Weaver (2002) expanded the definition that by regarding innovativeness as the capability, capacity and willingness of an enterprise to

support creativity and experimentation to solve recurring customer problems. Innovation is not simply about generating creative ideas, but also involves the commercialization, implementation and the modification of existing products, services and new ways to meet market demand via new resource combinations. Accordingly, innovation lays at the heart of the entrepreneurial process and is a means of opportunity exploitation. Innovation is not a characteristic of the individual entrepreneurs, but of their actions (Gartner, 1988).

Proactiveness

Proactiveness reflects an action-orientation with a forward-looking perspective reflected in actions taken in anticipation of future demand (Covin & Slevin, 1989; Lumpkin & Dess, 2001). Kreiser, Marino, & Weaver (2002) define proactiveness as the aggressive execution and follow-up actions to drive an enterprise toward the achievement of its objectives by whatever reasonable means required. Proactive firms constantly seek new opportunities by anticipating future demand and developing products and services in regards of unmet customer needs. They tend to be industry leaders in regards of developing new products, procedures, or technologies (Lumpkin and Dess, 1996). Consequently, they are also likely to be initiators in the creation or discovery of new attributes that lead to an increase in value creation (Foss et al, 2007). As such, proactiveness has certain underlying attributes like the anticipation and quick reaction to opportunities; the attitude to being a pioneer or fast follower; and the high regard for employee initiatives (Stevenson & Jarillo, 1990).

Risk-Management

Before elaborating risk-management, the term propensity to take risk needs to be defined. Risk-taking refers to the willingness to commit significant resources to opportunities that involve a reasonable chance of costly failure. Brockhaus

(1980) has found that some entrepreneurs may be cautious and risk averse under some circumstances and risk-taking in others. While risk bearing is an important element of entrepreneurial behaviour, entrepreneurial managers found to be „carefully brave” that is they tend to take risk grudgingly and only after they have made valiant attempts to spread their risks on capital sources and resource providers (Stevenson, 2006). Although all venturing attempts face uncertainty and the possibility of painful mistakes such problems take a more acute form for entrepreneurial managers, vis-à-vis small business founders (Aldrich & Martinez, 2001). Hence, the measurement of the extent to which individuals differ in their willingness to take risk is fraught with difficulty, especially when it is based on subjective evaluation.

Growth Orientation

A considerable body of literature has demonstrated that growth orientation in itself represents an entrepreneurial characteristic (Cooper, Woo, & Dunkelberg, 1989). Vesper (1980) pointed out that many business owners never intend their business to grow over what they consider to be a controllable size. Hence, it is necessary to go beyond the notion of corporate life cycles and stages to conceive of an entrepreneurial firm (Carland, Hoy, & Carland, 1984). Glueck (1980) distinguished between entrepreneurial ventures and what he termed family businesses by focusing on the needs and preferences opposed to those of the business. The critical factor in distinguish entrepreneurial managers from non-entrepreneurial ones, and in particular small business owners, is the presence of a sound and articulated growth objective.

IMPACT OF ENTREPRENEURSHIP IN ECONOMIC DEVELOPMENT

Economic development essentially means a process of upward change whereby the real per capita income of a country increases for a long

period of time. The crucial role played by the entrepreneurs in the western countries has made the people of underdeveloped countries conscious of the significance of entrepreneurship in economic development. India after independence had a challenge of recovering back from having a sluggish economic development. There were many platforms to develop the economy of which few were like direct participation of the government, Regulation of private sector economic, increased exports business and lastly the Act of Entrepreneurship, such as Start-Ups, Joint Ventures.

Entrepreneurship is the act of thinking creatively and en-cashing their opportunity being selfish personally and indirectly helping in Economic development thus possessing more impetus, and the focus has been on entrepreneurial activity in which the individual's involvement has a direct ownership interest in the business. Countries where more entrepreneurship is motivated by an economic opportunity recognized than by necessity have higher levels of income (Lal & Clement, 2005). We are finding "How entrepreneurship is good for economic development?" for this we need to find meaning of entrepreneurship and self employment. Entrepreneurship may not lead to economic development as there is no mechanism to link the activity to development. We have learned from the studies and trend that self-employment declines as economies become more developed. So we can conclude that when economies remove people from self-employment, there can be an increase in development. The important role that an entrepreneurship plays in the economic development of an economy can be put in a more systematic manner as follows:

1. Promotes capital formation by mobilizing the idle saving of the public.
2. Reduces the concentration of economic power.
3. Stimulates the equitable redistribution of wealth.
4. It also induces backward and forward linkages which stimulated the process of economic development in the country.
5. It promotes country's export trade.

INFORMATION TECHNOLOGY IN ENTREPRENEURSHIP

The advent of information technology has changed all the aspects of the world in the third millennium to a great extent and the entrepreneurship in particular. IT contributes to the entrepreneurship in different ways. It is now a widely held view that the world economy has entered a much more complex phase where individual national economies have become inextricably linked. In this new world economy, resources and markets have ceased to have the indelible national identity of the past. With the changes in businesses today, information technology (IT) plays a vital role in business survival.

Deans and Kane (1992) found that information technology (IT) plays an indispensable role in making a company successful under uncertain and turbulent economic conditions. Pearce et al. (2007) introduces information technology in line with social entrepreneurship as leverage for the sustainable development. Entrepreneurial individuals combine many personality traits - innovativeness, risk taking, pro-activeness in the sense of doing what is necessary to realize their ideas combined with shouldering responsibility for success or failure (Covin and Slavin, 1989). Entrepreneurship is defined as the process of creating value by combining a unique mix of the aforementioned concepts in order to take advantage of an opportunity (Morris and Sexton, 1996).

IT TOOLS IN ENTREPRENEURSHIP

In order to strengthen the society of learning and knowledge, Despite a certain setback, with the application of the IT strategy, will support the development strategy of all entities and by

learning from the world's most prominent business practices, it is possible to overcome the development gap and define new development potentials of in the area of entrepreneurship. As specified by (Duncombe, 1999) the tools of IT for Entrepreneurship development are:

1. **E-Mail:** E-mail enables fast and relatively cheap communication internally, between networked computers, and externally, within the locality, regionally and world-wide. An e-mail message, and any accompanying computer files, can be sent via the phone line within a few minutes to any global destination.
2. **Internet and World-Wide Web:** Connection to the internet enables communication with a global network of computers, and access to enormous quantities of information, providing for multi-media content based on text, pictures, sound, graphics and moving images.
3. **Local Networking:** Enterprises gets benefited from creating internal networks by connecting computers together by means of cabling. This enables them to communicate with each other, to share data files, printers, faxes, etc. Computers can be linked within a single location known as a Local Area Network (LAN) or between multiple locations, known as a Wide Area Network (WAN). Internal networks can also be connected to external networks such as the World Wide Web (Internet).
4. **Cloud Computing Tools:** cloud computing tools like Cloud ERP, Marketing automation, inbound marketing, lead tracking, sale cycle management have strengthened the enterprises to go and conduct their business in much larger scale.
5. **Mobile Communications:** Mobile communications based on new digital technology, allows the entrepreneurs to answer queries from the customers immediately, and to reach the staff working away from the of-

fice, no matter in which location they are. In developed countries mobile phones are now part of the business fabric and have been extremely rapid.

6. **Electronic Commerce:** E-commerce will have a significant effect on the Entrepreneurship. E-commerce has already been in use, through the telephones and credit cards, electronic payment and money transfer systems and smart cards. However, it is through the medium of the internet and online services that rapid expansion is likely to take replace the traditional way of doing things.

The Latest emerging tools of Information Technologies for Entrepreneurship that will be making headlines in shaping and development of Entrepreneurship ventures are:

1. **Social Technologies:** Enterprises will also be benefited by improved social networking sites like Facebook, Twitter, etc where information always flows and will be garnered by the one in need of the information.
2. **Cognitive Networks, Big Data:** Communication systems handle volumes of data generated by embedded devices, mobile users, enterprises, contextual information, network protocols, location information and such.
3. **Green Communications:** Green Communications created the Participatory Internet to answer connectivity challenges of network densification, mobility and sustainability.
4. **Smarter Smartphones, Connected Sensors:** The indisputable rock-start of devices is the smartphone, and its future can't be brighter, these ubiquitous communication devices continue evolving, and so are prices which, driven by cost and performance improvements in digital technologies, are falling rapidly.

5. **Network Neutrality, Internet Governance:** The Internet has been operating since its inception under “open” principles, i.e. an open standards-based network that treats all traffic in roughly the same way, i.e. no connection blocking, bandwidth transparency, universal connectivity, and best effort service.
6. **Molecular Communications:** Molecular communication is an emerging paradigm where bio-nanomachines communicate to perform coordinated actions. Molecular communications utilize biological molecules both as carriers and as information.

FUTURE RESEARCH DIRECTIONS

People live in a day and age in which empowerment and individuality are embraced and this approach is gaining momentum. The future without a doubt belongs to entrepreneurs and this is something to be really excited about. The future is headed in a direction we have never seen before and changes are vastly innovative and fast approaching. Consumers are definitely privy to this fact and more and more they are taking their business to a thriving entrepreneur over a corporate entity. If the basic idea of being own boss isn't appealing enough all of the other perks that come along with building a brand that belongs all to his own blood, sweat, tears should seriously make him reconsider his career path. A few of the momentous shift happening in the way of entrepreneurship are namely; Digital wanderers, Personal brands, Bonding together, Heart-Centric businesses, and No more cost controls. This study articulates the potential importance of entrepreneurship and to put forth the various dimensions it paves. But the research on how to establish control and appeal is limited in this study. For further studies, it is important to explore how the entrepreneurs can establish the formal appeal and control system, and keep transparency in establishing the businesses.

CONCLUSION

Entrepreneurship is the lifeblood of any economy. Indian entrepreneurs are more about overcoming barriers, obstacles, inspiring and surmount in their relevant fields. Entrepreneurship is one of the important segments of economic growth and development. Innovation is a key factor that an entrepreneur brings in an overall change through which it is possible for the maximum social good. As it is said that, 'Failures are the stepping stones for success.' An entrepreneur life is not a bed of roses, as they face many obstacles in the way of entrepreneurial achievement.

Information technology has created dramatic developments in various spheres of entrepreneurship and is regarded as most important tool of modern entrepreneurship. Entrepreneurship is requisite of technological development is infrastructure of entrepreneurship. Furthermore, entrepreneurship can tell the condition of a nation, it acts as a central cohesive source of support and stability, not only to the society but also to the whole economy given the impetus of technological tools and their application. Last but not least one can say that, where technological entrepreneurship exists, dwells development, similarly progress and prosperity.

REFERENCES

- Acs, Z. J., & Audretsch, D. B. (1988). Innovation in large and small firms: An empirical analysis. *The American Economic Review*, 78(4), 678–690.
- Aldrich, H. E., & Martinez, M. A. (2001). Many are called, but few are chosen: An Evolutionary Perspective for the Study of Entrepreneurship. *Entrepreneurship Theory and Practice*, 25(2), 41–56.
- Brockhaus, R. H. (1980). Risk taking propensity of entrepreneurs. *Academy of Management Journal*, 23(3), 509–520. doi:10.2307/255515

- Carland, J. W., Hoy, F., & Carland, J. A. C. (1984). Differentiation entrepreneurs from small business owners: A conceptualization. *Academy of Management Review*, 9(2), 345–359.
- Carter, N. M., Gartner, W. B., Shaver, K. G., & Gatewood, E. J. (2003). The career reasons of Nascent entrepreneurs. *Journal of Business Venturing*, 18(1), 13–39. doi:10.1016/S0883-9026(02)00078-2
- Christensen, C. M., & Raynor, M. E. (2003). *The Innovator's Solution*. Boston, MA: Harvard Business School Press.
- Cooper, A. C., Woo, C. Y., & Dunkelberg, W. C. (1989). Entrepreneurship and initial size of firms. *Journal of Business Venturing*, 4(5), 317–332. doi:10.1016/0883-9026(89)90004-9
- Covin, J. G., & Slevin, D. P. (1991). A conceptual model of entrepreneurship as firm behavior. *Entrepreneurship Theory and Practice*, 16(1), 7–25.
- Deans, P. C., & Kane, M. J. (1992). *Information Systems and Technology*. Boston, MA: PWS-Kent Publishing.
- Duncombe, R. (1999). *The Role of Information and Communication Technology for Small, Medium and Micro Enterprise Development in Botswana. Interim results of a research project sponsored by the UK Department for International development (DFID)*. Institute for Development Policy and Management, University of Manchester.
- Foss, K., Foss, N. J., & Klein, P. G. (2007). Original and derived judgment: An entrepreneurial theory of economic organization. *Organization Studies*, 28(12), 1893–1912. doi:10.1177/0170840606076179
- Gartner, W. B. (1988). Who is an entrepreneur? Is the wrong question. *American Journal of Small Business*, 12(4), 11–32.
- Ghosh, B.C., & Kwan, W., (1996). *An analysis of key success factors of SMEs: A comparative study of Shingapore/Malaysia and Australia/New Zealand*. Unpublished.
- Glueck, W. F. (1980). *Business policy and strategic management*. New York: McGraw-Hill.
- Greenberg, D. B., & Sexton, D. L. (1988). An Interactive model of new venture initiation. *Journal of Small Business Management*, 26(3), 1–7.
- Kreiser, P. M., Marino, L. D., & Weaver, K. M. (2002). Assessing the Psychometric Properties of the Entrepreneurial Orientation Scale: A Multi-Country Analysis. *Entrepreneurship Theory and Practice*, 26(4), 17–94.
- Lal, A. K., & Clement, R. W. (2005). Economic Development in India: The Role of Individual Enterprise and Entrepreneurial Spirit. *Asia-Pacific Development Journal*, 12(2), 81–99.
- Lucky, E. O. (2011). Nurturing Entrepreneurship Development in the 21st century. *International Journal of Humanities and Social Science*, 1(9), 219–227.
- Lumpkin, G. T., & Dess, G. G. (1996). Clarifying entrepreneurial orientation construct and linking it to performance. *Academy of Management Review*, 21(1), 135–172.
- Miller, D., & Friesen, P. H. (1983). Strategy making and environment: The third link. *Strategic Management Journal*, 4(3), 221–235. doi:10.1002/smj.4250040304
- Mitchelmore, S., & Rowley, J. (2010). Entrepreneurial competencies: A literature review and development agenda. *International Journal of Entrepreneurial Behavior & Research*, 16(2), 92–111. doi:10.1108/13552551011026995
- Morris, B. (2014). *Challenges Faced By Entrepreneurs Today*. Retrieved from <https://www.deluxe.com/blog/top-10-challenges-faced-entrepreneurs-today-solved>
- Morris, M. H., & Sexton, D. (1996). The Concept of Entrepreneurial Intensity: Implications for Company Performance. *Journal of Business Research*, 36(1), 5–13. doi:10.1016/0148-2963(95)00158-1

Pearce, J. M., Grafman, L., Colledge, T., & Legg, R. (2008). Leveraging information technology, social entrepreneurship, and global collaboration for just sustainable development. In *Proceedings of Annual Conference of National Collegiate Inventors and Innovators Alliance* (p. 201). National Collegiate Inventors & Innovators Alliance.

Pride, V. (2014). *Need for Corporate Entrepreneurship*. Retrieved from <https://boldanddetermined.com/top-10-challenges-youll-face-new-entrepreneur>

Santhi, N., & Kumar, S. R. (2011). Entrepreneurship Challenges and Opportunities in India. *Bonfring International Journal of Industrial Engineering and Management Science*, 1(1), 14–16. doi:10.9756/BIJEMS.1004

Shapero, A. (1975). The displaced, uncomfortable entrepreneur. *Psychology Today*, 9(6), 83–88.

Stevenson, H. H. (2006). A Perspective on Entrepreneurship. Harvard Business School, 1–13.

Stevenson, H. H., & Gumpert, D. E. (1985). The Heart of Entrepreneurship. *Harvard Business Review*, 85–94.

Stevenson, H. H., & Jarillo, J. C. (1990). A paradigm of entrepreneurship: Entrepreneurial management. *Strategic Management Journal*, 11, 17–27.

Tavakoli, A. (2013). Impact of Information Technology on Entrepreneurship Development. *Advances in Environmental Biology*, 7(8), 1421–1426.

Vesper, K. H. (1980). *New venture strategies*. Englewood Cliffs, NJ: Prentice Hall.

Wang, C. K., & Wong, P. K. (2004). Entrepreneurial interest of university students in Singapore. *Technovation*, 24(2), 163–172. doi:10.1016/S0166-4972(02)00016-0

ADDITIONAL READING

Acs, Z.J. (2006). How is Entrepreneurship Good for Economic Growth? *Innovations, MIT Press Journal*. 97-107. Doi:10.1162/itgg.2006.1.1.97

Audretsch, D. B., & Keilbach, M. (2004). Entrepreneurship Capital and Economic Performance. *Regional Studies*, 38(8), 949–959. doi:10.1080/0034340042000280956

Cohen, W. A. (2006). *The Entrepreneur & Small Business Problem Solver* (3rd ed.). Hoboken, NJ: John Wiley & Sons.

Powell, B. (2008). *Making Poor Nations Rich. Entrepreneurship and the Process of Economic Development*. Stanford: Stanford University Press.

Rodrik, D. (2007). *One Economics Many Recipes. Globalization, Institutions, and Economic Growth*. Princeton: Princeton University Press.

Smallbone, D., & Welter, F. (2006). Conceptualising Entrepreneurship in a Transition Context. *International Journal of Entrepreneurship and Small Business*, 3(2), 190–206. doi:10.1504/IJESB.2006.008928

VanStel, A., Carree, M., & Thurik, R. (2005). The Effect of Entrepreneurial Activity on National Economic Growth. *Small Business Economics*, 24(3), 311–321. doi:10.1007/s11187-005-1996-6

KEY TERMS AND DEFINITIONS

Business: A body involved in buying and selling of goods & services to the end-user.

Cash Flow: The difference between the company's cash receipts and its cash payments in a given period. It refers to the amount of money actually available to make purchases and pay current bills and obligations.

Economic Development: Economic Development is an act of creating wealth and wellbeing among society.

Entrepreneur: An Entrepreneur is the impetus which drives the nation to an elevated stage where one has the ability to seize every opportunity put forth to realise dreams.

Entrepreneurship: Entrepreneurship is the process of launching an enterprise in order to generate income as well as enrich economic development.

Information Technology: A creation which makes things easy cum advanced.

Innovation: Innovation is process of creating something new and creative which is a value addition.

E

Entrepreneurship Concept, Theories, and New Approaches

José Manuel Saiz-Alvarez

Tecnologico de Monterrey, Mexico

Martín García-Vaquero

Nebrija University, Spain

INTRODUCTION

In today's world, have a good number of active entrepreneurs is key to determining good corporate health of a country. Entrepreneurs discover and exploit new business opportunities guided by intuition (Saiz-Alvarez, Coduras, & Cuervo-Arango, 2013) and risk control, while motivating social change. Complementary to entrepreneurs, capitalists only invest their money searching for share profitability, and managers, as decision-makers, foster mainly first-order competitive advantages in the company (R&D and innovation) by enhancing competition (Cuervo, Ribeiro, & Roig, 2007). When competition is strong, firms are achieving continuous improvements over time, which benefits their stakeholders.

Contrary to Van Praag and Van Ophem (1995), who affirm entrepreneurship was first studied in Cantillon's *Essai sur la nature du commerce en général*, published in 1755, the concept of entrepreneur (in the sense of merchant) was first studied in Spain in the second book of Tomás de Mercado's *Tratos y contratos de mercaderes y tratantes discidos y determinados* [Deals and Contracts Applied to Certain Merchants and Traders], published in 1569, and reedited with a few tweaks in 1571, as *Suma de tratos y contratos* [Treaty on Deals and Contracts]. Regarding entrepreneurship, Tomás de Mercado approves new business and exchange trade, only if they are consistent with the moral rules based on not taking advantage of others by using abusive prices and

interest rates, or by conducting unfair agreements, both socially and economically.

This connection between moral acts and entrepreneurship continues, with exceptions, until today. In fact, entrepreneurs must be involved in the study of sources for discovering, and achieving new business opportunities, in order to exploit them (Shane and Venkataraman, 2000). As a result, and given the impact of entrepreneurship in a country, «contemporary economic theory recognizes entrepreneurship as an independent factor of production on a more-or-less equal footing with land, labor, and capital» (Hébert and Link, 1989, p. 40). The objective of this work is to define entrepreneurship, and the classical and new theories of entrepreneurship, both from economic and psychological perspectives.

BACKGROUND

ENTREPRENEURSHIP: CONCEPT

Entrepreneurship can be analyzed following a multiple perspective, generating a multidimensional concept (Bula, 2012). While the study of entrepreneurship began for Economics and Managerial Sciences in the second half of the XVI century, the first psychological studies about this issue were published in the 1960s with the seminal works of McClelland (1965), Rotter (1966), and Atkinson (1966). These authors focused their efforts on trying to explain how individual and social motivation is one of the most important psychological factors

to explain entrepreneurship, and also discover that imagination, power distance and willingness for taking risks are key factors for success. Moreover, when the possibility of achievement gets higher, entrepreneurial propensity rate increases (Kalkan & Kaygusuz, 2012).

In order to entrepreneurs to have success, they must identify business opportunities (Stevenson & Jarillo, 1990; Barringer & Ireland, 2006; Timmons, 1999; Mariotti & Glackin, 2010), be able of choosing and managing entrepreneurial careers (Haynie & Shepherd, 2011), and be capable of acting entrepreneurially (McMullen & Shepherd, 2006; Shepherd & Patzelt, 2011) by being adapted to business circumstances given their capacity of resilience to failure.

The mentality of entrepreneurs differs, as they are driven by *entrepreneurial alertness* (Kirzner, 1979) defined as a distinctive set of perceptual and cognitive processing skills directed to opportunity recognition processes. As a result, and given this entrepreneurial alertness, only the most risk-lovers tend to be successful while managing their businesses, as «the entrepreneur always searches for change, responds to it, and exploits it as an opportunity» (Drucker, 1985, p. 25).

Entrepreneurship is the act of innovation involving endowing existing resources with new wealth-producing capacity (Drucker, 1985), as the nature of the decision making context with entrepreneurs' decisions (Alvarez & Barney, 2005). As a result, «the entrepreneur is someone who is specialized in taking responsibility for and making judgmental decisions that affect the location, form, and the use of goods, resources or institutions» (Hébert & Link, 1989, p. 213).

Entrepreneurs are the business core of a company, especially in newly-born firms. Given the above definitions, entrepreneurship can be defined as the type of business strategy focused on the creation of jobs, social wealth, and profit by optimizing the use of productive and commercial resources.

Moreover, Ahmad and Seymour (2008) affirm that entrepreneurship is the phenomenon

associated with entrepreneurial activity, that is, «the enterprising human action in pursuit of the generation of value, through the creation or expansion of economic activity, by identifying and exploiting new products, processes or markets» (p. 9). In this sense, Lumpkin and Dess (1996) affirm that the essential act of entrepreneurship by launching new ventures, either by start-up firms or through internal corporate venturing, and is accomplished by entering into new or established markets with new or existing goods and services. Entrepreneurial activities theoretically guided by some theories of entrepreneurship that will be analyzed in the next section.

THEORIES OF ENTREPRENEURSHIP

The Neoclassical School

The combination of psychological and managerial characteristics in entrepreneurs is fundamental to understand their behavior. In this sense, Hébert and Link (1989), cited by Audretsch (2003), have identified three distinct intellectual traditions in the development of the entrepreneurship contemporary literature: [1] the German Tradition, based on von Thünen (2015[1875]) and Schumpeter (1997[1911]), that has the greatest impact on the contemporary entrepreneurship literature; [2] the Chicago Tradition, rooted on Knight (1921) and Schultz (1980), and [3] the Austrian Tradition, based on von Mises (1949), Kirzner (1979), and Shackle (1968). We can add an emerging intellectual traditions formed by Latin American and Spanish researchers, given the socioeconomic peculiarities of these countries that have been barely studied so far in the economic literature on this issue.

In his classic treatise, *Theorie der wirtschaftlichen Entwicklung. Eine Untersuchung über Unternehmengewinn, Kapital, Kredit, Zins und den Konjunkturzyklus* [Theory of Economic Development. A Study of Entrepreneurial Profit, Capital, Credit, Interest Rate, and the Business

Cycle], Schumpeter (1997[1911]) proposes a theory of creative destruction, where new firms equipped with the entrepreneurial spirit displace less innovative incumbents, leading to economic growth, as a «function of entrepreneurs is to reform or revolutionize the pattern of production by exploiting an invention, or more generally, an untried technological possibility for producing a new commodity or producing an old one in a new way» (Schumpeter, 1977[1942], p. 13). As a result, Schumpeter (1934) defines entrepreneurs by focusing on adopting innovation and new business strategies that can take several forms: [1] the introduction of new methods of production to increase productivity, efficiency, and EBITDA (Earnings before Interests, Taxes, Depreciation, and Amortization); [2] the introduction of new goods or quality thereof; [3] the opening of new markets; [4] the conquest of new sources of supply of new materials or semi-elaborated goods; and [5] the entry to new market niches or sectors. In short, entrepreneurs identify opportunities, assemble required resources, implement a practical action plan, and harvest the reward in a timely, flexible way (Sahlman & Stevenson, 1991).

Regarding the Chicago and Austrian Traditions, entrepreneurs attempt to predict and act upon change within dynamic and uncertain markets (Knight, 1921). According to this School of Thought, free market conditions determine competition on global markets, so public intervention is negative, as it detracts economic resources to private firms. As a result, entrepreneurs are immune from control of rational bureaucratic knowledge (Weber, 1947), so they can be defined as the professionals who discover new business opportunities, after having taken into account commercial and financial risks, and the subsequent creation of new economic activity, often through the creation of new business organizations (Reynolds, 2005).

For this School, entrepreneurs must compete in a globalized business world, by thinking globally and acting locally (“glocalization”). Therefore, entrepreneurs must keep labor costs in line with

productivity, and must achieve a steady increase in production in order to satisfy existing market niches, and enter into new markets. If that steady growth is not achieved, there is a risk of business stagnation, which is detrimental to the company’s stakeholders.

Entrepreneurs, as coordinators and arbitrageurs, as well as the main judgment-based agent of production in the economy (Hébert and Link, 1988, cited by Bula, 2012), are always speculators, as they deal with the uncertain conditions of the future, so it is vital for them to have success in the correctness of their anticipation of uncertain events (Von Mises, 1949).

In a globalized socioeconomic world, entrepreneurs create social and economic wealth by bringing together unique packages of resources to exploit marketplace opportunities (Ireland, Hitt, & Sirmon, 2003) guided by opportunity recognition (Gaglio & Katz, 2001; Shanke & Venkataraman, 2000; Kirzner, 1979), and by creating a previously-unperceived opportunity for profit (Von Mises, 1949). In short, and as Grebel, Pyka, and Hanusch (2001) gather from Schumpeter and the Neo-Schumpeterian School, «no entrepreneur—no innovation — no dynamics — no evolution» (p. 7).

The New Keynesian School

Entrepreneurship is marginal in the new Keynesian school, as labor markets only deal with the development of efficiency wage models. Despite this issue, the implementation of public policies to support private entrepreneurship generates crowding-in effects, although public spending is increased, so new entrepreneurs are created.

By combining macroeconomic government spending data with individual-level entrepreneurship data, Islam (2015) finds a positive relationship between increasing the share of social and public goods at the cost of private subsidies and entrepreneurship, while confirming a negative relationship between total government consumption and entrepreneurial activity.

In our globalized world, informal ventures are increasingly important, and are influenced by economic and political factors affecting entrepreneurship. In this sense, Autio and Fu (2015) study the influence of economic and political institutions on the prevalence rate of formal and informal entrepreneurship across 18 countries in the Asia-Pacific region during the period 2001-2010, and find that the quality of institutions influences on both formal and informal entrepreneurship. Higher quality of economic and political institutions doubles the rates of formal entrepreneurship, and halves the rates of informal entrepreneurship. Moreover, both economic and political institutions have a complementary effect on driving entry into formal entrepreneurship, whereas only direct effects are observed for informal entry.

Entrepreneurs for New Keynesian economies are benefited from public expense and investment. In effect, government's Keynesian functions stimulate demand for goods and services through purchases and income redistribution; while government's Schumpeterian economic policy lowers the cost of innovation through regulatory requirements, R&D subsidies and innovation, and the acquisitions of expensive and new technology. These two economic-based Schools of thought are complemented with psychological-based Schools, as entrepreneurship is formed by a mixture of both economic and psychological thought.

Psychological-Based Schools

Regarding entrepreneurship, five Schools of thought can be distinguished, as follows: [1] *Functionalism*, where theorists try to explain thoughts, the learning process, instincts, and decisions simply describing what entrepreneurs do, without asking how they do it.

In this sense, is of interest to link resilience and functionalism, as rule-governed behavior turns people insensitive to contingencies, so a higher capacity of resilience is achieved; [2] *Behaviorism*, rooted on Functionalism, is focused on both

human and animal behaviors, so it analyzes the reasons why entrepreneurs act; [3] *Gestalt*, where human mind interprets data through several norms, laws, and organizational principles, so partial information can be completed. This completeness is necessary to make decisions with imperfect or asymmetric information, as is normal in the daily operations done by companies; [4] *Humanistic Psychology*, based on human motivation and on satisfaction of needs, has a direct influence on entrepreneurs, as they want to be happy by achieving corporate goals, and [5] *Cognitivism*, as well as Behaviorism, studies the mind through scientific experimentation, but differs from Behaviorism by accepting that psychologists can study and understand the inner workings of mind and mental processes, so psychoanalysis is rejected given its subjectivity.

These psychological theories applied to entrepreneurship complement economic and managerial theories on entrepreneurship, in order to have an overview into the entrepreneurial behavior. Entrepreneurs with better growth prospects have a high need for power and independence (Duchéneau and Orhan, 1998), as the need for achievement and competition is positively correlated with the result of the company. Moreover, entrepreneurs have a greater need for achievement (Box, White, and Barr, 1993).

The influence of psychological aspects on entrepreneurs is also seen in the modern theories of entrepreneurship that will be seen in the next paragraph.

NEW APPROACHES OF ENTREPRENEURSHIP

Jack-of-All-Trades Theory

Theoretically developed by Lazear (2005), and mathematically by Kaiser (2012), entrepreneurs' main task is to organize professional teams ("generalists"/"specialists") by combining intel-

Table 1. Psychological aspects of entrepreneurs

	Risk-Lover	High Necessity for Achievement	Innovation	Desire for Independence	High Necessity for Competence
Anna, Chandler, Jansen, & Mero (2000)		X	X	X	
Box, White, & Barr (1993)		X			X
Collins, Moore, & Unwala (1988)		X		X	
Crainich, Eeckhoudt, & Trannoy (2013)	X				
Davidsson (1988)					X
De Pablo, & Bueno (2004)	X	X	X		X
Duchéneau, & Orhan (1998)				X	X
Dubini (1988)	X	X	X		
Douglas, & Shepherd (1997)	X				
Feesen, & Dugan (1989)				X	X
Genesca, & Venecia (1984)				X	
Jenssen, & Kolvered (1992)		X	X	X	
Jindapon (2013)	X				
Johnson (1990)		X			
Koh (1996)		X		X	
McClelland (1965)		X			
Rusque (2002)		X		X	
Sheinberg, & MacMillan (1988)				X	
Sijbom, Janssen, & Van Yperen (2015)		X			
Smith, Bracker, & Miner (1987)	X		X		
Woo, Cooper, & Dunkelberg(1988)		X		X	

Source: Elaborated from Barba-Sánchez, Jiménez-Zarco, and Martínez-Ruiz (2007), and Marulanda, Correa, and Mejía (2009).

lectual capital, as the sum of human capital, relational capital, and structural capital (Saiz-Alvarez, 2012), physical capital and ideas to innovate, and to create product and services to actual and prospective clients. According to this entrepreneurial theory, only generalists (“jacks-of-all trades”) become entrepreneurs, as they assemble production factors, and have knowledge, skills and working abilities in a large number of areas, especially when entrepreneurs are individuals. As a result, and according to Kaiser (2012): [1] individuals with more balanced skill sets are more likely to become entrepreneurs; [2] the supply of generalists entrepreneurs is smaller for production processes, as they require a large number of independent skills; and [3] generalists entrepreneurs have, on

average, a more balanced human capital strategy than specialists.

O-Ring Theory

Seminally developed by Kremer (1993), and theoretically by Fabel (2004), the O-Ring Theory deals with abilities, and product, task, and service quality in firms, as different abilities and skills are supplied by different individuals with different capacities. Team work is essential in this theory, as if a group member fails to fulfill the task (“specialists”), the entire project can also fail. The more able the team, the more per capita capital is used, so it is preferably to limit the size of the team, as more per capita capital must be used.

Resources— and Capabilities— Based Theory

This theory can be seen as an “inside-out” process of strategy formulation, as the entrepreneur looks at what resources the firm possesses to assess their potential for value generation defined in a business strategy to capture the maximum value in a sustainable way. These financial, physical, human, technological, reputational, and organizational *resources* are inputs to be taken into account for the production process, while *capabilities*, defined as the capacity for a team of resources to perform some task or activity, constitute the main source of the firm’s competitive advantage.

One of the keys to have success when applying this theory into practice is to efficiently apply *routines*, defined as those regular and predictable patterns of activity which are made up of a sequence of coordinated and repeated (*learning-by-doing*) actions done by individuals. As a result, econo-

mies of experience are created, and sustainable and actions guided by appropriability.

Achieving sustainability is fundamental for firm’s reputation and good economic results in the long term. The four main elements for the sustainability derived from resources are: [1] *durability*, defined as the rate at which resources depreciate or become obsolete; [2] *replicability*, when acquisition of resources is done inside the organization. The more complex the organizational routines are, the more difficult it will be to replicate them (Grant, 1991); [3] *transparency*, as imperfect information creates inefficiency-based problems; and [4] *transferability*, when rivals can easily acquire resources and capabilities, so leading positions can be in danger.

Successful firms must hire the best human resources, as well as to renew the level of capability per unit of activity and to constantly invest, if they want to be leaders in their sectors. Complex organizational capabilities pose a huge barrier

Figure 1. Branches of the capability lifecycle

Source: Elaborated from Helfat and Peteraf (2003).

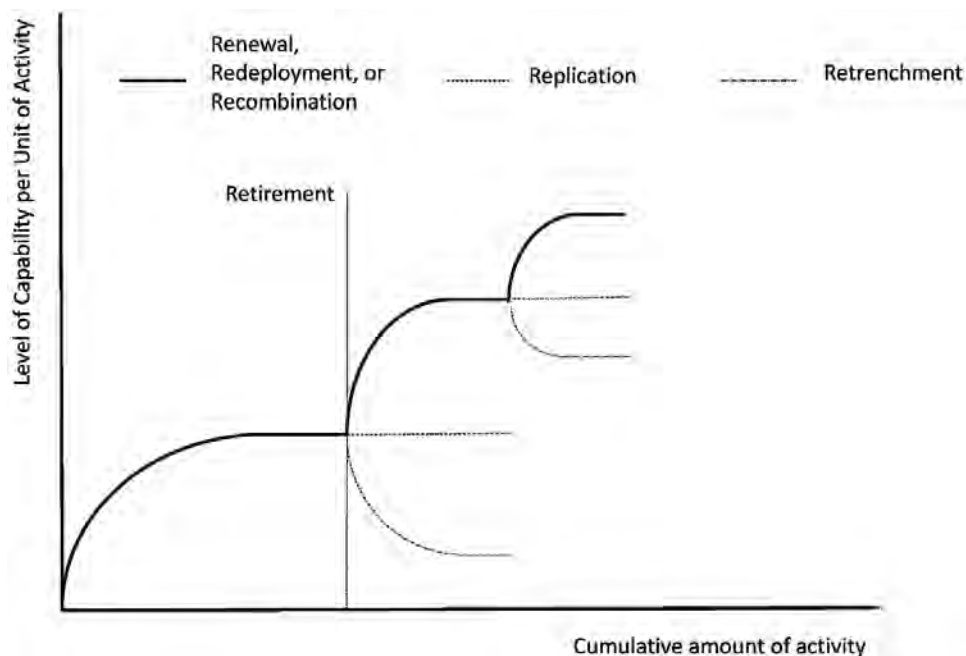
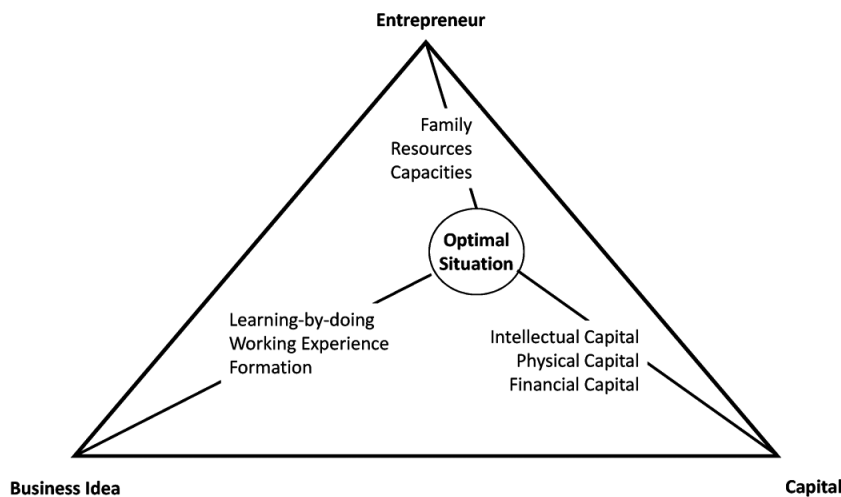


Figure 2. Optimal situation and business success

Source: Own elaboration.



for other firms to enter the market, and especially when economies of experience are well developed.

Theory of the Optimal Triangle

Companies having a high level of business growth and motivation, influence positively on the formation of new business ideas. It is therefore essential that emotional characteristics close to the ability to work and enthusiasm must be developed to take forward the company. Without enthusiasm, mainly in start-ups, it is impossible for firms to survive in the global market.

Business success is positively related to the level of training, participation in previous projects, business experience, and the number of founders. These factors, however, can change by other endogenous variables, such as debt, product prices, stakeholders' attitudes and investments, purchases, rotation sales, and wages, and exogenous factors, such as energy prices, external shocks, and country risk, among others.

We propose to summarize these variables in three key terms: entrepreneur, business ideas, and capital, as shown in Figure 2. Successful entrepreneurs are at the base of the triangle, because they support entrepreneurs, as they are fuelled by intellectual (as the sum of human, relational and

structural capital), physical, and financial capital. The combination of entrepreneurial resources and capacities, working experience by learning-by-doing, formation, and capital determine the optimal situation for the company. As seen in Figure 2, this point is between the vertices "entrepreneur" and "capital", as they are fundamental to have business success.

CONCLUSION

Based on Information and Communication Technologies (ICTs), entrepreneurship can be classified as: [1] *Intraentrepreneurship* or *Corporate venturing* (Sharma and Chrisman, 1999), defined by the type of entrepreneurship focused on fast trucking product development to take advantage of new business opportunities or to assess feasibility of new processes or designs; [2] *Traditional entrepreneurship*, that according to Brunet and Alarcón (2004), can be classified as: [a] *Novice entrepreneurs*, when they manage their first firm; [b] *Habitual entrepreneurs*, that are classified as *serial entrepreneurs*, when they sell their original ventures to buy and develop new ones, and [c] *Portfolio entrepreneurs*, when they do not sell their original ventures and buy others, so a cluster

is created; [3] *Solidarity Entrepreneurship* (Saiz-Alvarez, 2015), where firms are guided by human values based on solidarity and not for profit, and [4] *Social Entrepreneurship*, mainly formed by cooperatives, whose aim is to maximize the impact of the company on the population searching for social wealth. In this sense, and after having analyzed data from 106,484 individuals in 26 nations, Stephan, Uhlaner, and Stride (2015) confirm joint effects on social entrepreneurship of formal regulatory (government activism), informal cognitive (post-materialist cultural values), and informal normative institutions, through socially supportive cultural norms, or weak-tie social capital.

FUTURE RESEARCH DIRECTIONS

- Women Entrepreneurship,
- Social and Solidarity Entrepreneurship (SSE),
- ICTs, Business Incubators, and Business Accelerators,
- Entrepreneurship in Education,
- Entrepreneurship as a Tool for Sustainable Economic Growth.

REFERENCES

- Ahmad, N., & Seymour, R. G. (2008). *Defining Entrepreneurial Activity: Definitions Supporting Frameworks for Data Collection*. OECD Statistics Working Papers, 1. OECD Publishing.
- Alvarez, S. A. (2005). Two Theories of Entrepreneurship. Alternative Assumption and the Study of Entrepreneurial Action. *Foundations and Trends in Entrepreneurship*, 1(3), 105–148. doi:10.1561/03000000003
- Alvarez, S. A., & Barney, J. B. (2005). How Do Entrepreneurs Organize Under Conditions of Uncertainty? *Journal of Management*, 31(5), 776–793. doi:10.1177/0149206305279486
- Anna, A. L., Chandler, G. N., Jansen, E., & Mero, N. P. (2000). Women business owners in traditional and non-traditional industries. *Journal of Business Venturing*, 15(3), 279–303. doi:10.1016/S0883-9026(98)00012-3
- Atkinson, J. W. (1966). Motivational Determinants of Risk-taking Behavior. In J. W. Atkinson & N. T. Feather (Eds.), *A Theory of Achievement Motivation* (pp. 11–30). New York: John Wiley & Sons.
- Audretsch, D. B. (2003). *Entrepreneurship. A Survey of the Literature*. *Enterprise Papers*, 14. Brussels, Belgium: European Commission.
- Autio, E., & Fu, K. (2015). Economic and Political Institutions and Entry into Formal and Informal Entrepreneurship. *Asia Pacific Journal of Management*, 32(1), 67–94. doi:10.1007/s10490-014-9381-0
- Barba-Sánchez, V., Jiménez-Zarco, A. I., & Martínez-Ruiz, M. P. (2007, January). Effects of entrepreneurial motivation and background on new venture's growth. *International Congress on Marketing Trends*. Paris: Ecole Supérieure de Commerce de Paris.
- Box, T. M., White, M. A., & Barr, S. H. (1993). A contingency model of new manufacturing firm performance. *Entrepreneurship. Theory into Practice*, 18(2), 31–45.
- Brunet, I., & Alarcón, A. (2004). Teorías sobre la figura del emprendedor. *Papers*, 73, 81–103.
- Bula, H. O. (2012). Evolution and Theories of Entrepreneurship: A Critical Review on the Kenyan Perspective. *International Journal of Business and Commerce*, 1(11), 81–96.
- Collins, O. F., Moore, D. G., & Unwala, D. B. (1964). *The organization makers: a behavioral study of independent entrepreneurs*. New York: Meredith.

- Crainich, D., Eeckhoudt, L., & Trannoy, A. (2013). Even (Mixed) Risk Lovers are Prudent. *The American Economic Review*, 103(4), 1529–1535. doi:10.1257/aer.103.4.1529
- Cuervo, A., Ribeiro, D., & Roig, S. (Eds.). (2007). *Entrepreneurship: Concepts, Theory, and Perspective*. Springer.
- Davidsson, P. (1988). *Type of man and type of company revisited: a confirmatory cluster analysis approach*. In *Frontier of Entrepreneurship Research* (pp. 88–105). Wellesley, MA: Babson College.
- De Pablo, I., & Bueno, Y. (2004). Rasgos personales que caracterizan a un emprendedor de éxito. *Iniciativa Emprendedora y Empresa Familiar*, 44, 113–125.
- Douglas, E. J., & Shepherd, D. (1997). *Why entrepreneurs create businesses: a utility maximizing response*. In *Frontiers of Entrepreneurship Research* (pp. 185–186). Wellesley, MA: Babson College.
- Drucker, P. F. (1985). *Innovation and Entrepreneurship: Practice and Principles*. New York: Harper Business.
- Dubini, P. (1988). *Motivational and environmental influences on business start-ups: some hints for public policies*. In *Frontiers of Entrepreneurship Research* (pp. 31–45). Wellesley, MA: Babson College.
- Duchéneau, B., & Orhan, M. (2000). *Les femmes entrepreneurs en France: Percée de femmes dans un monde construit au masculin*. Paris, France: Seli Arslan.
- Fabel, O. (2004). Spinoffs of Entrepreneurial Firms: An O-Ring Approach. *Journal of Institutional and Theoretical Economics*, 160(3), 416–438. doi:10.1628/0932456041960614
- Feeser, H. R., & Dugan, K. W. (1989). *Entrepreneurial motivation: a comparison of high and low growth high tech founders*. In *Frontiers of Entrepreneurship Research* (pp. 13–27). Wellesley, MA: Babson College.
- Gaglio, C., & Katz, J. (2001). The Psychological Basis of Opportunity Identification: Entrepreneurial Alertness. *Small Business Economics*, 16(2), 95–111. doi:10.1023/A:1011132102464
- Gaglio, C. M. (1997). Opportunity recognition: review, critique and suggested research directions. In J. Katz (Ed.), *Advances in Entrepreneurship, Firm Emergence and Growth* (pp. 139–202). Greenwich, CT: JAI Press.
- Genesca, E., & Veciana, J. M. (1984). Actitudes hacia la creación de empresas. *Información Comercial Española*, 611, 147–155.
- Haynie, J. M., & Shepherd, D. (2011). Towards a Theory of Discontinuous Career Transition: Investigating Career Transitions Necessitated by Traumatic Life Events. *The Journal of Applied Psychology*, 96(3), 501–524. doi:10.1037/a0021450 PMID:21090884
- Hébert, R. F., & Link, A. N. (1989). In Search of the Meaning of Entrepreneurship. *Small Business Economics*, 1, 39–49. doi:10.1007/BF00389915
- Helfat, C. E., & Peteraf, M. A. (2003). The Dynamic Resource-based View: Capability Life-cycles. *Strategic Management Journal*, 24(10), 997–1010. doi:10.1002/smj.332
- Ireland, R. D., Hitt, M. A., & Sirmon, D. G. (2003). A Model of Strategic Entrepreneurship: The Construct and its Dimensions. *Journal of Management*, 29(6), 963–989. doi:10.1016/S0149-2063(03)00086-2
- Islam, A. (2015). Entrepreneurship and the Allocation of Government Spending Under Imperfect Markets. *World Development*, 70, 108–121. doi:10.1016/j.worlddev.2015.01.002

- Jenssen, S., & Kolvereid, L. (1992). The entrepreneurs' reasons leading to start-up as determinants of survival and failure among Norwegian new ventures. In S. Birley (Ed.), *International perspectives on entrepreneurship research* (pp. 120–133). Amsterdam, Netherlands: Elsevier Science Publishers.
- Jindapon, P. (2013). Do Risk Lovers Invest in Self-protection? *Economics Letters*, 121(2), 290–293. doi:10.1016/j.econlet.2013.08.031
- Johnson, B. R. (1990). Toward a multidimensional model of entrepreneurship: The case of achievement motivation and the entrepreneur, *Entrepreneurship. Theory into Practice*, 14(3), 39–48.
- Kaiser, U. (2012). A Primer in Entrepreneurship. Department of Business Administration, University of Zurich, Switzerland. (unpublished)
- Kalkan, M., & Kaygusuz, C. (2012). The Psychology of Entrepreneurship. In *Entrepreneurship. Made and Educated*. InTech Europe. Retrieved April 30, 2015 from <http://www.intechopen.com/books/entrepreneurship-born-made-and-educated/the-psychology-of-entrepreneurship>
- Kirzner, I. M. (1979). *Perception, Opportunity, and Profit: Studies in the Theory of Entrepreneurship*. Chicago: The University of Chicago Press.
- Koh, H. C. (1996). Testing hypotheses of entrepreneurial characteristics. *Journal of Managerial Psychology*, 11(3), 12–25. doi:10.1108/02683949610113566
- Kremer, M. (1993). The O-Ring Theory of Economic Development. *The Quarterly Journal of Economics*, 108(3), 551–575. doi:10.2307/2118400
- Lazear, E. P. (2004). Balanced Skills and Entrepreneurship. *The American Economic Review*, 94(2), 208–211. doi:10.1257/0002828041301425
- Lazear, E. P. (2005). Entrepreneurship. *Journal of Labor Economics*, 23(4), 649–680. doi:10.1086/491605
- Lumpkin, G. T., & Dess, G. G. (1996). Clarifying the Entrepreneurial Orientation Construct and Linking It to Performance. *Academy of Management Review*, 21(1), 135–172.
- Mariotti, S., & Glackin, C. (2010). *Entrepreneurship*. Upper Saddle River, NJ: Prentice Hall.
- Marulanda, J. A., Correa, G., & Mejía, L. F. (2009). Emprendimiento: Visiones desde las teorías del comportamiento humano. *Revista EAN*, 66, 153–168.
- McClelland, D. (1965). Need of Achievement and Entrepreneurship: A Longitudinal Study. *Journal of Personality and Social Psychology*, 1(4), 389–392. doi:10.1037/h0021956 PMID:14328753
- McMullen, J. S., & Shepherd, D. A. (2006). Entrepreneurial Action and the Role of Uncertainty in the Theory of the Entrepreneur. *Academy of Management Review*, 31(1), 132–152. doi:10.5465/AMR.2006.19379628
- Reynolds, P. D. (2005). Understanding business creation: Serendipity and scope in two decades of business creation studies. *Small Business Economics*, 24(4), 359–364. doi:10.1007/s11187-005-0692-x
- Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs*, 80(1), 1–28. doi:10.1037/h0092976 PMID:5340840
- Rusque, A. M. (2002, November). De la capacidad emprendedora basada en rasgos individuales a la capacidad emprendedora como responsabilidad social. In *The Entrepreneurial Venture*. Boston: McGraw Hill.
- Saiz-Alvarez, J. M. (2012). Ética, capital intelectual y gestión del conocimiento. *Revista EAN*, 73, 44–59.

- Saiz-Alvarez, J. M. (2015). Solidarity Entrepreneurship in the European Knowledge Economies: Keys for a New Paradigm to Be Applied into Developing Nations. In *Entrepreneurship in the Contemporary Knowledge-based Global Economy*. Hershey, PA: Information Science Reference (IGI Global).
- Saiz-Alvarez, J. M., Coduras, A., & Cuervo-Arango, C. (2013). Entrepreneurial Strategy, Innovation and Cognitive Capabilities. What Role for Intuitive SMEs? *Journal of Small Business Strategy*, 21(2), 25–35.
- Scheinberg, S., & McMillan, I. C. (1988). *An all country study of motivations to start a business*. In *Frontiers of Entrepreneurship Research* (pp. 669–687). Wellesley, MA: Babson College.
- Schultz, T. W. (1980). Investment in Entrepreneurial Ability. *The Scandinavian Journal of Economics*, 82(4), 437–448. doi:10.2307/3439676
- Schumpeter, J. A. (1934). *The Theory of Economic Development: An Inquiry into Profits, Capital, Credit, Interest, and the Business Cycle*. Cambridge, MA: Harvard University Press.
- Schumpeter, J. A. (1977). *Capitalism, Socialism, and Democracy*. Harper Collins Publishers. (Original work published 1942)
- Schumpeter, J. A. (1997). *Theorie der wirtschaftlichen Entwicklung: Eine Untersuchung über Unternehmergewinn, Kapital, Kredit, Zins und den Konjunkturzyklus*. Berlin, Germany: Duncker Humblot. (Original work published 1911) doi:10.3790/978-3-428-07725-0
- Shackle, G. L. S. (1968). *Expectations, Investment, and Income*. Oxford Clarendon Press.
- Shane, S., & Venkataraman, S. (2000). The Promise of Entrepreneurship as a Field of Research. *Academy of Management Review*, 25(1), 217–226.
- Sharma, P., & Chrisman, J. J. (1999). Toward a Reconciliation of the Definitional Issues in the Field of Corporate Entrepreneurship, *Entrepreneurship. Theory into Practice*, 23, 11–27.
- Shepherd, D. A., & Patzelt, H. (2011). The New Field of Sustainable Entrepreneurship: Studying Entrepreneurial Action Linking “What Is To Be Sustained” with “What Is To Be Developed?”, *Entrepreneurship Theory into Practice*, 35(1), 137–163.
- Sijbom, R. B. L., Janssen, O., & Van Yperen, N. W. (2015). Leaders Receptivity to Subordinates Creative Input: The Role of Achievement Goals and Composition of Creative Input. *European Journal of Work and Organizational Psychology*, 24(3), 462–478. doi:10.1080/1359432X.2014.964215
- Smith, N. R., Bracker, J. S., & Miner, J. B. (1987). *Correlates of firm and entrepreneurs success in technologically innovative companies*. In *Frontiers of Entrepreneurship Research* (pp. 337–356). Wellesley, MA: Babson College.
- Stephan, U., Uhlaner, L. M., & Stride, C. (2015). Institutions and Social Entrepreneurship: The Role of Institutional Voids, Institutional Support, and Institutional Configurations. *Journal of International Business Studies*, 46(3), 308–331. doi:10.1057/jibs.2014.38
- Stevenson, H. H., & Jarillo, J. C. (1990). A Paradigm of Entrepreneurship: Entrepreneurial Management. *Strategic Management Journal*, 11, 17–27.
- Timmons, J. A. (1999). *New Venture Creation. Entrepreneurship for the 21st Century*. New York: Irwin/McGraw-Hill.
- Tomás de Mercado. (1975). *Suma de tratos y contratos*. Madrid, Spain: Editora Nacional. (Original work published 1571)
- Tomás de Mercado. (2011). *Tratos y contratos de mercaderes y tratantes discidos y determinados*. Valladolid, Spain: Maxtor. (Original work published 1569)
- Van Praag, C. M., & Van Ophem, H. (1995). Determinants of Willingness and Opportunity to Start as an Entrepreneur. *Kyklos*, 48(4), 513–540. doi:10.1111/j.1467-6435.1995.tb01282.x

Von Mises, L. (1949). *Human Action*. New Haven, CT: Yale University Press.

Von Thünen, J. H. (2015). *Der isolierte Staat in Beziehung auf Landwirtschaft und Nationalökonomie* [The Isolated State in Relation to Agriculture and the Domestic Economy]. Berlin, Germany: Wiegandt. (Original work published 1875)

Weber, M. (1947). *The Theory of Social and Economic Organization*. New York: Oxford University Press.

Westhead, P., & Wright, M. (1998). Novice, Portfolio, and Serial Founders: Are They Different? *Journal of Business Venturing*, 13(3), 173–204. doi:10.1016/S0883-9026(97)90002-1

Woo, C. Y., Cooper, A. C., & Dunkelberg, W. C. (1988). *Entrepreneurial typologies: definitions and implications*. In *Frontiers of Entrepreneurship Research* (pp. 165–176). Wellesley, MA: Babson College.

KEY TERMS AND DEFINITIONS

E

Capability: Capacity for a team of resources to perform some task or activity, what constitutes the main source of the firm's competitive advantage.

Entrepreneur: Individual who starts a new business or adopt a new strategy in an existing organization.

Entrepreneurship: Business strategy focused on the creation of jobs, social wealth, and profit by optimizing the use of productive and commercial resources.

Glocalization: Term created by the combination of «global» (think global) and «local» (act local).

Intellectual Capital: Combination of human, structural and relational capital.

Resilience: An individual resistance to failure, usually in business.

Resource: Inputs to be taken into account for the production and commercial processes.

Open Data and High-Tech Startups Towards Nascent Entrepreneurship Strategies

Fotis Kitsios

University of Macedonia, Greece

Maria Kamariotou

University of Macedonia, Greece

INTRODUCTION

Open data can be defined as freely, accessible, online data, which are available and can be reused. They are provided under open access allowance so that the data can be reused without restrictions (Jetzek et al., 2014). They can be used for the development of applications which improve citizens' life. A way to boost the development of innovative applications is by hosting hackathons, workshops and conferences (Zuiderwijk et al., 2015).

A definition for a hackathon determines hackathon as an event where people come together in order to engage in creating and launching a new or finished application, which will solve citizens' problems (Rosell et al., 2014). Hackathons are designed to support the use of open data which will benefit both government and citizens (Sieber & Johnson, 2015). Governments desire to scrimmage both citizens and developers to expand an application using open data which are developed, promoted and distributed through hackathons. According to researchers, there are many motivations to persuade developers to participate in a hackathon (Komssi et al., 2015). By hosting these competitions, governments wish to inform about the concernment and the use of open data while they are also supporting the expansion of new applications. These hackathons are frequently consolidated with prize money for participants (Johnson & Robinson, 2014).

Open data are free and accessible, so is beneficial for researchers, government and society. The benefits of open data could be categorized into political and social. Providing the sharing of scientific data is not only a technological issue. It also includes organizational models and research practices, as well as it contains institutional, legal and economic factors (Sa & Grieco, 2016). Several scientific journals support the reveal of experimental data thus data can be reused, reproduced and confirmed (Hossain et al., 2016). Some examples of these benefits are the increased transparency, the accountability, the participation and self-empowerment of citizens to economic growth and the stimulation of competitiveness and innovation (Viscusi et al., 2014). There are two significant benefits concerning the release of open data. The first regards the increased participation of citizens in government, the transparency and melioration of decision-making and the understanding of smarter government tactics through data. The second benefit is the release of data by startups and new businesses in order to expand innovative applications (Conradie & Choenni, 2014).

Open data contains an ecosystem, which displays the relationship among government and innovators and citizens, which is furthermore involved within the larger environment, such as the economy, legal system, and policy expertise (Zuiderwijk & Janssen, 2014).

Previous researchers only describe the activities which are necessary to host the contest (Grabowski et al., 2015, Lee et al., 2015, Rosell et al., 2014). There is limited previous research not only on what motivates the developers to participate in open data competitions, but also on the benefits and challenges which are caused from the use of open data. Furthermore, researches focus on factors that affect nascent entrepreneurs' decision to create a startup but researchers in the field of open data and hackathons relative researches are limited.

The objective of this chapter is to present a framework in order to examine the impact of motivations, benefits and barriers of the use of open data in the participation in hackathons and to develop a startup based on their applications.

The structure of this chapter is as following: After the introductory section, a theoretical background is provided about open data and hackathons. Results of previous surveys support this background. The chapter ends up with conclusions, suggestions for future research and limitations.

BACKGROUND

Open Data

According to Worthy, (2014), open data are defined as the reuse of information and the publishing of government data in a reclaimable form. Data which have been published in an open data format would be available to the public to be reused without limitations (Attard et al., 2015). So developers in hackathons will have the opportunity to use them in order to develop their applications for citizens.

Open data contains an ecosystem, which displays the relationship among government, innovators and citizens. Also, it is involved within the larger environment, such as the economy, legal system, and policy expertise. The open government ecosystem contains relationships between government and innovators from technology sectors, private industry, and academic institutions. Their aim is to transform the form of new data

standards, to investigate new designs of information systems and new technology platforms that permit the expansion of information or technology resources for the future.

The major actors, who contribute in open data are data providers, open data legislators, open data facilitators and many different kinds of open data users, such as citizens, researchers, journalists, developers, entrepreneurs and academics. These actors have various interests and these interests may vary (Viscusi et al., 2014).

The first actor named "Data Providers", consists of the organizations that provide free data. They are commonly public administrations that have many data but they do not have the abilities or resources to use them in the form of data or expand services with them. These organizations provide data to improve the national economy, allowing businesses and citizens to take advantage of them. The second actor is "Service Providers". They offer services related to data and can improve the revenue by the treatment services. The third actor is "Application Developers". They cooperate with partners developing applications related to open data or use open data as completed with their own data in their applications. They develop applications and service using data dispensing by the previous actors. The next actor is "Application Users". They consume data. This process can be done with the development of applications and services. A user can be a customer, a citizen or a business. A customer is defined as a user who has acquired a commercial application from an application store. A citizen can be defined as a user who uses the application personally as a citizen. A business can be defined as a user who uses the applications in his own business. The final actor is "Infrastructure and Tool Providers". They provide the necessary tools supporting the expansion of applications based on open data. These tools are tools providers, cloud service providers, market place providers (Immonen et al., 2014).

The purpose of a good government is to interact with users or citizens, business, and civil sector organizations. Other relationships are referred among public managers and citizens, civil society

organizations and businesses that allow government to detect what types of government data and information is needed and what data or government information services estimate as transparency.

There are two significant benefits concerning the release of open data. The first regards the increased participation of citizens in government, the transparency and melioration of decision-making and the understanding of smarter government tactics through data. The second benefit is the release of data by startups and new businesses in order to expand innovative applications (Conradie & Choenni, 2014). An essential way for actors to be informed about the use of open data is by hosting hackathons (Zuiderwijk et al., 2015).

The benefits of open data could be categorized to political and social. Some examples of these benefits are the increased transparency, accountability, participation and self-empowerment of citizens to economic growth and stimulation of competitiveness and innovation, to name a few (Viscusi et al., 2014). Another benefit is the accessibility of information to wider approach and the allowance to induce innovative usage and application (Roy, 2014). Additionally, Sieber & Johnson, (2015) categorize benefits as political and social, economic, operational and technical, which increase transparency, accountability and civil participation. Open data supports accountability, enhance trust and boost citizen satisfaction. Citizens can generate better ideas, think about alternatives and make better decisions.

As far as businesses, open data can increase several benefits for them. These benefits contain economic growth, innovation, empowerment, participation and new services. Despite the opportunities which are derived for businesses using open data, many businesses still do not benefit from this use. They deal with difficulties about the accessibility of data or the limited knowledge and resources to make use of the data. The use of open data involves stakeholders, staff, software providers and government actors (Zuiderwijk et al., 2015).

On the contrary, researchers determine barriers and challenges that disturb the use of open data. The barriers are categorized into ten categories. These categories are availability and access, findability, usability, understandability, quality, linking and combining data, comparability and compatibility, metadata, interaction with the data provider and opening and uploading (Zuiderwijk & Janssen, 2014). The barriers which are included in these categories are legislation, information quality, and technology, use and participation (Janssen et al., 2012). Other barriers are about technical issues, privacy or law (Conradie & Choenni, 2014). Challenges and problems could come up to the capability of users and application developers to identify the advantages related to open data. These challenges and problems are related to unfamiliar business idea in order to use data to develop new business, lack of technical readiness to use data sources due to complex data format or interfaces, weakness to generate suitable data sources for application goals, insufficient access of regional data sources for developing applications for local services, unclear licensing of open data, legal issues related to data that are not officially open by decision of owner, technical restrictions mentioned to data publishing platforms, warranty of quality and reliability of data, lack of local data and usefulness of open data (Jaakkola et al., 2014).

There is a need for education, experience from users, citizens and businesses and support from government to use open data (Pope & Greene, 2003). Statistical techniques are used for the collection, analysis and visualization of data. Knowledge about these techniques is limited. So, training is necessary to expand the capabilities of using these techniques (Janssen et al., 2012).

There is a number of dimensions which point out the significance of open data. These dimensions are related to legislative, political, social, economical, institutional, operational and technical aspects. A legislative dimension concentrates on freedom of information acts, open data policies, open government instructions, notes and state-

ments. A political dimension brings the importance of both political evolutions and political differentiation among countries. A social dimension examines the importance of cultural differentiation between countries and discrepancies in agendas related to the social advantages of opening data, such as transparency and accountability into focus. An economical dimension focuses on financial benefits which can be appeared with open data. An institutional dimension emphasizes the ways in which organizations allow and push the publication and adoption of open data. An operational dimension highlights the use of open data and the demand for being able to use open data. A technical dimension emphasizes the importance of open data technologies, open data platforms and open data substructures (Zuiderwijk & Janssen, 2014).

Hackathons

Governments aim to scrimmage citizens and developers who expand an application using open data to participate in contests and to explicate it through contests. By hosting these competitions, governments wish to inform citizens about the concernment and the use of open data while they are also supporting the expansion of new applications. Despite the rate of their popularity, contests and hackathons are a new phenomenon and researchers have not focused on their interest yet. More attention should be paid to these hackathons. Furthermore, researches which highlight the benefits of the contest or hackathon, for both sponsoring governments and hackathon participants are limited. These hackathons are frequently consolidated with prize money for participants and they typically inform, encourage and convince public of the use of open data and their value. Hackathons represent a specific challenge or topic (such as tourism), on which the sponsoring government intends to contribute to the creation of an application for citizens which will meet the market needs (Johnson & Robinson, 2014). Therefore, the most significant challenge of hosting a contest or a hackathon is to convince

developers to generate innovative ideas, which can be transformed to applications for the citizens. Hackathons include activities such as idea generation, software design, implementation and testing, service operation and monitoring (Juell-Skielse et al., 2014). Open data contests or hackathons have the same purpose, but in the first category of competitions developers create applications based on open data. Thus in this view the fact that it is the responsibility of governments to supply with free and easily open data is derived.

Participants can access data through references to data as provided in journal articles, newsletters or other publications, government or institutional announcements, data archives/repositories/systems, web search engines, directories or catalogs, social media, blogs as well as with direct requests to data providers (Schmidt et al., 2016). Especially, examples of open research data which are distributed as government statistics and industry records are GenBank, Protein Data Bank, Sloan Digital Sky Survey, World Data Centers, Global Biodiversity Information Facility, NASA Distributed, Active Archive Centers, International Virtual Observatory Alliance, Digital Earth, PubMedCentral and arXiv (Pasquetto et al., 2015).

Examples of motivations regard the participation in the contest are training, the fairness of the judgment system, the collaboration and new knowledge (Rosell et al., 2014). Fun and enjoyment, intellectual challenge, status and reputation, user need, professional and personal identity, autonomy, learning and skills development, money, extrinsic reciprocity, signaling and career concerns motivate developers to participate in the innovation competitions (Juell-Skielse et al., 2014). Motivations such as fun, learning and collaboration convince developers to generate ideas that include high market and technical uncertainties (Komssi et al., 2015).

Previous researches on what motivates the developers to participate in open data competitions, as well as on the benefits and challenges which are caused by the use of open data are limited. Furthermore, researches focus on factors that

affect nascent entrepreneurs' decision to create a startup but researchers in the field of open data and hackathons relative researches are limited.

Table 1 summarizes the motivations to take part in the competition, the sources for search open data, as well as the benefits and the barriers of the use of open data.

Examples of created services which are based on open data aim to improve the economy, transparency, society, education, citizen life, navigation and mobility, tourism, environment, security and religion (Foulonneau et al., 2014).

The findings of these surveys show that companies use open data for visualization in order to support management and decision making. Statistics and identification of trends are useful because managers can make more effectively decisions. The use of open data can support the prediction of future trends, too. Also, open data can be shared among businesses and other ac-

tors to co-create in service development process. The sharing of open data is an innovative and attractive way to improve services, generating new ideas and develop new applications and services. Furthermore, businesses can use open data to analyze customers' behaviour, preferences and future demands. Researchers have focused on studying how businesses can communicate with customers using Information Technology (Kitsios & Kamariotou, 2016). When businesses provide data to developers, they can implement applications which are based on customers' preferences. Thus, businesses have to communicate with customers and they have to think about how customers can provide these data. It is important for businesses to define why data should be opened both in their policies and their business models. Managers should support a culture for innovation and sharing of open data to gain the benefits of their use (Immonen et al., 2014).

Table 1. Description of the variables

Variables	Questions	References
Motivations to take part in the competition	Fun and Enjoyment	(Juell-Skielse et al., 2014)
	Intellectual Challenge	(Juell-Skielse et al., 2014)
	Status and Reputation	(Juell-Skielse et al., 2014)
	User Need	(Juell-Skielse et al., 2014)
	Professional and Personal Identity	(Juell-Skielse et al., 2014)
	Autonomy	(Juell-Skielse et al., 2014)
	Learning and Skills Development	(Juell-Skielse et al., 2014)
	Money	(Juell-Skielse et al., 2014)
	Reciprocity	(Juell-Skielse et al., 2014)
	Signaling and Career Concerns	(Juell-Skielse et al., 2014)
Sources for search open data	References to data as provided in journal articles	(Schmidt et al., 2016)
	Newsletters or other publications	(Schmidt et al., 2016)
	Government or institutional announcements	(Schmidt et al., 2016)
	Searching in specific data archives/repositories/systems	(Schmidt et al., 2016)
	Web search engines	(Schmidt et al., 2016)
	Directories or catalogs	(Schmidt et al., 2016)
	Social media	(Schmidt et al., 2016)
	Blogs	(Schmidt et al., 2016)
	Direct requests to data providers	(Schmidt et al., 2016)

continued on following page

Table 1. Continued

Variables	Questions	References
Benefits of the use of open data	More transparency	(Janssen et al., 2012)
	Democratic accountability	(Janssen et al., 2012)
	Creation of trust in government	(Janssen et al., 2012)
	Improvement of public policies	(Janssen et al., 2012)
	Fair comparison with other research results	(Janssen et al., 2012)
	Easier access to data and discovery of data	(Janssen et al., 2012)
	Creation of new data based on combining data	(Janssen et al., 2012)
	Sustainability of data (no data loss)	(Janssen et al., 2012)
	The ability to reuse data/not having to collect the same data again and counteracting unnecessary duplication and associated costs (also by other public institutions)	(Janssen et al., 2012)
	Availability of information for investors and companies	(Janssen et al., 2012)
	Optimization of administrative processes	(Janssen et al., 2012)
	New governmental services for citizens	(Janssen et al., 2012)
	New (innovative) social services	(Janssen et al., 2012)
	Development of new products and services	(Janssen et al., 2012)
	Improvement of citizen services	(Janssen et al., 2012)
	Contribution toward the improvement of processes, products, and/or services	(Janssen et al., 2012)
	Economic growth and stimulation of competitiveness	(Janssen et al., 2012)
	Stimulation of innovation	(Janssen et al., 2012)
Barriers of the use of open data	Emphasis of barriers and neglect of opportunities	(Janssen et al., 2012)
	Lack of ability to discover the appropriate data	(Janssen et al., 2012)
	Difficulty in searching and browsing due to no index or other means to ensure easy search for finding the right data	(Janssen et al., 2012)
	Lack of knowledge to make use of or to make sense of data	(Janssen et al., 2012)
	Lack of information	(Janssen et al., 2012)
	No information about the quality of the open data	(Janssen et al., 2012)
	Incomplete information, only part of the total picture shown or only a certain range	(Janssen et al., 2012)
	No license for using data	(Janssen et al., 2012)
	Data available in various forms	(Janssen et al., 2012)
	Data formats and datasets are too complex to handle and use easily	(Janssen et al., 2012)
	Lack of the necessary capability to use the information	(Janssen et al., 2012)
	No process for dealing with user input	(Janssen et al., 2012)
	Having to pay a fee for the data	(Janssen et al., 2012)
	Registration required before being able to download the data	(Janssen et al., 2012)
	Legacy systems that complicate the publicizing of data	(Janssen et al., 2012)
	No uniform policy for publicizing data	(Janssen et al., 2012)
	No resources to publicize data (especially small agencies)	(Janssen et al., 2012)

The results of a survey in 21 teams participated in the final of Travelhack 2013 show that participants were citizens who were interested in developing services, project teams representing businesses and startups which aim to generate businesses from applications. The motivational factors with the highest value were self explanatory, named; Fun and Enjoyment, Intrinsic Intellectual Challenge and Intrinsic Status and Reputation. Also, the majority of the teams planned to finalize the service. An important finding is that money was a weak motivational factor (Juell-Skielse et al., 2014).

Schmidt et al., (2016) provide the results of 1253 participants to the Belmont Forum's Open Data Survey. The results of this survey show that open data are important for research (82%) as well as supporting applications (51%). The majority of participants expressed that the quality of data, the low cost, the ability of reuse and the easy of accessibility were the main benefits using open data. On the other hand, the most significant barriers were the legal constraints, the loss of recognition and the desire to publish results before realizing data. Furthermore, participants discover data in journal articles, web search engines and specific data repositories.

Rosell et al., (2014) present the results of an internal hackathon at AT&T. They observed that the participants felt as entrepreneurs, learning new technologies and having the freedom to develop new applications. Organizers of this hackathon followed a cyclical process which indicates that hackathon gives the opportunity to developers to generate new ideas and expand them into platforms. This expansion is supported by the learning of new technologies which can be useful for the production of new services.

These findings conclude that current studies have started to explore issues related with the use of open data by businesses, governments and developers participating in innovation contests but there are still several unanswered issues.

FUTURE RESEARCH DIRECTIONS

Hackathons support only the first phase of the development of applications. Thus, the understanding of collaborative development of digital services and the support of new mechanisms for increasing the support of the later phases of digital service development are proposed to be further expanded. Moreover, further research should be implemented in providing guidelines to organizers of innovation contests to develop them better to meet participants' needs of specific open data market.

Several technologies, such as cloud computing, semantic web technologies and interoperable applications can be used to support the development of digital services based on open data. However, issues which concern about the owners of data, the payment policies and the security of private data are still unanswered. A type of public catalogue of available open data is necessary to provide details about the most demanded open data.

CONCLUSION

Open data provides new business opportunities, new types of business models and an ecosystem where actors use open data to develop new digital services. Although open data is a new trend, the interest of businesses is high and they should be followed to be competitive, businesses are not yet motivated to provide widely their own data.

In this chapter, many motivations and opportunities were identified for businesses and developers using open data. Also, obstacles limiting the usage of open data were determined. The participation of businesses and developers in hackathons is important because they will understand the significance of open data, the benefits which they will gain and they can co-operate to develop new digital services. The main stakeholders, involved in open data processes are data providers, open

data legislators, open data facilitators and many different kinds of open data users, such as citizens, researchers, journalists, developers, entrepreneurs and academics. Governments ask to engage citizens and developers who build an application using open data promoted and distributed through contests. By organizing these contests, governments hope to inform about the importance and the use of open data while they also encourage the development of new applications and startups. A lot of survey is still necessary to be done to bring under control the barriers and to examine how hackathons can be used to support the co-operation among developers, businesses and governments.

REFERENCES

- Attard, J., Orlandi, F., Scerri, S., & Auer, S. (2015). A systematic review of open government data initiatives. *Government Information Quarterly*, 32(4), 399–418. doi:10.1016/j.giq.2015.07.006
- Conradie, P., & Choenni, S. (2014). On the barriers for local government releasing open data. *Government Information Quarterly*, 31, S10–S17. doi:10.1016/j.giq.2014.01.003
- Foulonneau, M., Martin, S., & Turki, S. (2014). How open data are turned into services? *International Conference on Exploring services science*, 31–39. doi:10.1007/978-3-319-04810-9_3
- Grabowski, S., Grzenda, M., & Legierski, J. (2015). The Adoption of Open Data and Open API Telecommunication Functions by Software Developers. *International Conference on Business Information Systems*, 337–347. doi:10.1007/978-3-319-19027-3_27
- Hossain, M. A., Dwivedi, Y. K., & Rana, N. P. (2016). State-of-the-art in open data research: Insights from existing literature and a research agenda. *Journal of Organizational Computing and Electronic Commerce*, 26(1-2), 14–40. doi:10.1080/10919392.2015.1124007
- Immonen, A., Palviainen, M., & Ovaska, E. (2014). Towards open data based business: Survey on usage of open data in digital services. *International Journal of Research in Business and Technology*, 4(1), 286–295. doi:10.17722/ijrbt.v4i1.197
- Jaakkola, H., Mäkinen, T., & Eteläaho, A. (2014). Open data: opportunities and challenges. *Proceedings of the 15th International Conference on Computer Systems and Technologies*, 25–39.
- Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, adoption barriers and myths of open data and open government. *Information Systems Management*, 29(4), 258–268. doi:10.1080/10580530.2012.716740
- Jetzek, T., Avital, M., & Bjorn-Andersen, N. (2014). Data-driven innovation through open government data. *Journal of Theoretical and Applied Electronic Commerce Research*, 9(2), 100–120. doi:10.4067/S0718-18762014000200008
- Johnson, P., & Robinson, P. (2014). Civic Hackathons: Innovation, Procurement, or Civic Engagement? *Review of Policy Research*, 31(4), 349–357. doi:10.1111/ropr.12074
- Juell-Skielse, G., Hjalmarsson, A., Johannesson, P., & Rudmark, D. (2014). Is the Public Motivated to Engage in Open Data Innovation? *International Conference on Electronic Government, Springer Berlin Heidelberg*, 277–288. doi:10.1007/978-3-662-44426-9_23
- Kitsios, F., & Kamariotou, M. (2016). Critical success factors in service innovation strategies: An annotated bibliography on NSD. *Proceedings of British Academy of Management (BAM), Conference 2016*.
- Komssi, M., Pichlis, D., Raatikainen, M., Kindstrom, K., & Jarvinen, J. (2015). What are hackathons for? *Software, IEEE*, 32(5), 60–67. doi:10.1109/MS.2014.78

- Lee, M., Almirall, E., & Wareham, J. (2015). Open data and civic apps: First-generation failures, second-generation improvements. *Communications of the ACM*, 59(1), 82–89. doi:10.1145/2756542
- Pasquetto, I. V., Sands, A. E., & Borgman, C. L. (2015). Exploring openness in data and science: What is open, to whom, when, and why? *Proceedings of the Association for Information Science and Technology*, 52(1), 1–4. doi:10.1002/pra2.2015.1450520100141
- Pope, J. A., & Greene, W. E. (2003). Developing a Model of Entrepreneurship Style. *Journal of Business and Entrepreneurship*, 15(1), 64–74.
- Rosell, B., Kumar, S., & Shepherd, J. (2014). Unleashing innovation through internal hackathons. *Innovations in Technology Conference, IEEE*, 1–8.
- Roy, J. (2014). Open Data and Open Governance in Canada: A Critical Examination of New Opportunities and Old Tensions. *Future Internet*, 6(3), 414–432. doi:10.3390/fi6030414
- Sá, C., & Grieco, J. (2016). Open Data for Science, Policy, and the Public Good. *Review of Policy Research*, 33(5), 526–543. doi:10.1111/ropr.12188
- Schmidt, B., Gemeinholzer, B., & Treloar, A. (2016). Open Data in Global Environmental Research: The Belmont Forums Open Data Survey. *PLoS ONE*, 11(1), 1–29. doi:10.1371/journal.pone.0146695 PMID:26771577
- Sieber, R. E., & Johnson, P. A. (2015). Civic open data at a crossroads: Dominant models and current challenges. *Government Information Quarterly*, 32(3), 308–315. doi:10.1016/j.giq.2015.05.003
- Viscusi, G., Castelli, M., & Batini, C. (2014). Assessing social value in open data initiatives: A framework. *Future Internet*, 6(3), 498–517. doi:10.3390/fi6030498
- Webster, J., & Watson, R. T. (2014). Analyzing the past to prepare for the future: Writing a literature review. *Management Information Systems Quarterly*, 26, 13–23.
- Worthy, B. (2015). The impact of open data in the UK: Complex, unpredictable, and political. *Public Administration*, 93(3), 788–805. doi:10.1111/padm.12166
- Zuiderwijk, A., & Janssen, M. (2014). The negative effects of open government data-investigating the dark side of open data. *Proceedings of the 15th Annual International Conference on Digital Government Research*, 147–152. doi:10.1145/2612733.2612761
- Zuiderwijk, A., Janssen, M., & Dwivedi, Y. K. (2015). Acceptance and use predictors of open data technologies: Drawing upon the unified theory of acceptance and use of technology. *Government Information Quarterly*, 32(4), 429–440. doi:10.1016/j.giq.2015.09.005
- Zuiderwijk, A., Janssen, M., Poulis, K., & van de Kaa, G. (2015). Open data for competitive advantage: insights from open data use by companies. *Proceedings of the 16th Annual International Conference on Digital Government Research*, 79–88. doi:10.1145/2757401.2757411

ADDITIONAL READING

- Ayele, W., Juell-Skielse, G., & Johannesson, P. (2015). Evaluating Open Data Innovation: A Measurement Model for Digital Innovation Contests. *Proceedings of Pacific Asia Conference on Information Systems*, (pp. 1–16).
- Cena, G., Scanzio, S., & Valenzano, A. (2015). Reliable comparison of clock discipline algorithms for time synchronization protocols. In *IEEE 20th Conference on Emerging Technologies & Factory Automation (ETFA)*, (pp. 1–9). Luxembourg, Luxembourg. doi:10.1109/ETFA.2015.7301461
- Hellberg, A. S., & Hedström, K. (2015). The story of the sixth myth of open data and open government. *Transforming Government, People, Process and Policy*, 9(1), 35–51.

Hielkema, H., & Hongisto, P. (2013). Developing the Helsinki smart city: The role of competitions for open data applications. *Journal of the Knowledge Economy*, 4(2), 190–204. doi:10.1007/s13132-012-0087-6

Juell-Skielse, G., Hjalmarsson, A., Juell-Skielse, E., Johannesson, P., & Rudmark, D. (2014). Con-tests as innovation intermediaries in open data markets. *Information Polity*, 19(3,4), 247-262.

Kassen, M. (2013). A promising phenomenon of open data: A case study of the Chicago open data project. *Government Information Quarterly*, 30(4), 508–513. doi:10.1016/j.giq.2013.05.012

Meijer, R., Conradie, P., & Choenni, S. (2014). Reconciling contradictions of open data regarding transparency, privacy, security and trust. *Journal of Theoretical and Applied Electronic Commerce Research*, 9(3), 32–44. doi:10.4067/S0718-18762014000300004

Scanzio, S., Cumani, S., Gemello, R., Mana, F., & Laface, P. (2010). Parallel implementation of Artificial Neural Network training for speech recognition. *Pattern Recognition Letters*, 31(11), 1302–1309. doi:10.1016/j.patrec.2010.02.003

Zuiderwijk, A., Helbig, N., Gil-García, J. R., & Janssen, M. (2014). Special Issue on Innovation through Open Data: Guest Editors Introduction. *Journal of Theoretical and Applied Electronic Commerce Research*, 9(2), i–xiii. doi:10.4067/S0718-18762014000300001

KEY TERMS AND DEFINITIONS

Actors of Open Data: Data providers, open data legislators, open data facilitators and many different kinds of open data users, such as citizens, researchers, journalists, developers, entrepreneurs and academics.

Barriers of Open Data: The barriers were categorized into ten categories. These categories are availability and access, findability, usability, understandability, quality, linking and combining data, comparability and compatibility, metadata, interaction with the data provider and opening and uploading.

Benefits of Open Data: Increased transparency, accountability, participation and self-empowerment of citizens to economic growth and stimulation of competitiveness and innovation.

Digital Innovation Contest: An event in which third party developers participate to develop and implement the most satisfying service prototype, for a particular aim, based on open data.

Hackathons: Hackathons are events where people come together in order to engage in creating and launching a new or finished application, which will solve citizens' problems.

Open Data Ecosystem: It displays the relationship among government, innovators and citizens, which is involved within the larger environment, such as the economy, legal system, and policy expertise.

Open Data: Open data can be defined as freely, accessible, online data, which are available and can be reused and they are provided under open access allowance so that the data can be reused without restrictions.

Category E

**Environmental Science and
Agriculture**

Carbon Capture From Natural Gas via Polymeric Membranes

E

Nayef Mohamed Ghasem
UAE University, UAE

Nihmiya Abdul Rahim
UAE University, UAE

Mohamed Al-Marzouqi
UAE University, UAE

INTRODUCTION

With the rapid development of technology, there is an increasing demand for fuels. Natural gas is an environmentally friendly, renewable, and clean energy source. It is also the third largest proportion in energy structure throughout the world after coal and oil. The composition of the raw natural gas extracted from producing wells depends on the type, depth, and location of the underground deposit and the geology of the area. Natural gas consists primarily of methane as the prevailing element but it also contains considerable amounts of light and heavier hydrocarbons as well as contaminating compounds of CO_2 , N_2 , Hg, He, H_2S and etc. The presence of acid gases such as CO_2 and H_2S can cause corrosion of pipeline and equipment and they present a major safety risk. Also they reduce the energy content of the gas and affect the selling price of the natural gas. Further in Liquefied Natural Gas (LNG) processing plant, while cooling the natural gas to a very low temperature, the CO_2 can be frozen and block pipeline systems and cause transportation drawback. Consequently, natural gas produced at the wellhead must be processed, i.e., cleaned, before it can be safely delivered to the high-pressure, long-distance pipelines that transport the product to the consuming public.

The traditional method for CO_2 separation is amine scrubbing. Although high product yields

and purities can be obtained, the disadvantage of this method is its high energy consumption, especially during stripper, in combination with high liquid losses due to evaporation of the solvent in the stripper (Naim et al., 2012). In addition, as liquid and gas streams cannot be controlled independently the occurrence of flooding, foaming, channeling and entrainment of the absorption liquid also limits the process. Membrane technology is a promising method to replace the conventional absorption technology. It has a high energy efficiency, is easy to scale-up because of its modular design and it has a high area-to-volume ratio. A limitation can be found in the permeability-selectivity tradeoff relation. Gas-liquid membrane contactor (GLMC) combines the advantages of membrane technology with those of absorption liquid (Ze et al., 2014). In a GLMC the microporous membrane acts as a fixed interface between the feed gas and the absorption liquid without dispersing one phase into another and this decoupling of the gas and liquid phase prevents any momentum transfer occurring across the phase boundary. As a consequence, the operation problems and constraints take place in conventional absorption technology can be resolved. Further the employment of microporous membrane elucidates the permeability-selectivity tradeoff relation drawback challenged in membrane technology. The performance of GLMC as CO_2 absorber and stripper depend upon several

DOI: 10.4018/978-1-5225-2255-3.ch266

factors such as type of membrane, type of absorption liquid, module configurations and process parameters. Understanding the optimistic attributes of these factors on CO_2 separation performance of GLMC is vital important to develop the GLMC that gives the outstanding CO_2 absorption/stripping performance. The focus of this work is to illustrate the potential for the energy efficient and effective separation of CO_2/CH_4 gas mixture via lean solvent and regenerating of the rich solvent through absorption/stripping mechanism taking place in a hollow fiber GLMC process.

BACKGROUND

GLMCs have attracted great interest over the past decade as CO_2 absorber and stripper. In the absorber, CO_2 diffuses from the feed gas through the porous membrane and is then absorbed in the flowing liquid. Then this CO_2 rich liquid circulated from the absorber to the stripper membrane contactor module in which stripped CO_2 will be carried by sweep gas (Figure 1).

In the GLMC, gas and liquid flow on the different side of the microporous membrane and membrane acts only as a barrier between two phases without dispersing one phase to another. In general, when hydrophobic microporous membranes are used in membrane contactors, the gas-liquid interface is immobilized at the opening of the pores of microporous membrane by careful control of the pressure difference between the two phases. For applications in gas-liquid absorption/desorption, the driving force is based on the concentration gradient. The gas molecules to be separated diffuse from the concentrated phase to the gas liquid interface via the membranes pores then contacts the diluted phase on the other side. For instance, in the case of CO_2/CH_4 separation, as shown in Figure 2, CO_2 molecules diffuses from the feed gas side through the membrane and is then absorbed in the selective absorption liquid.

In GLMC the mass transfer process consists three steps in series: the transfer from one phase to the membrane surface, transfer within membrane pores and transfer from other phase interface to the bulk. Figure 3 shows the concentration

Figure 1. Membrane gas absorption/stripper process

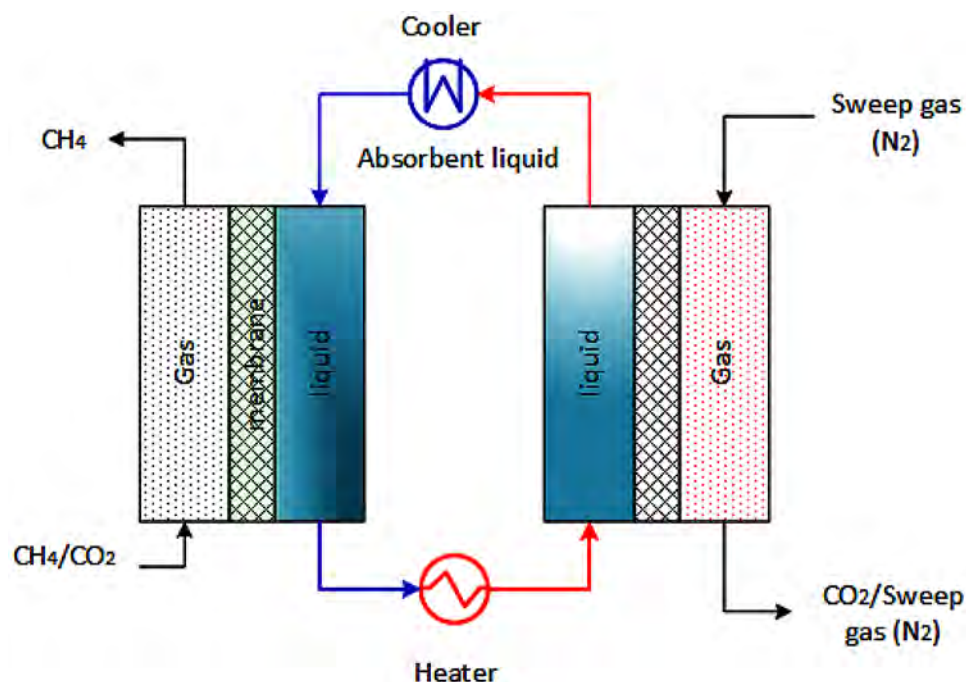
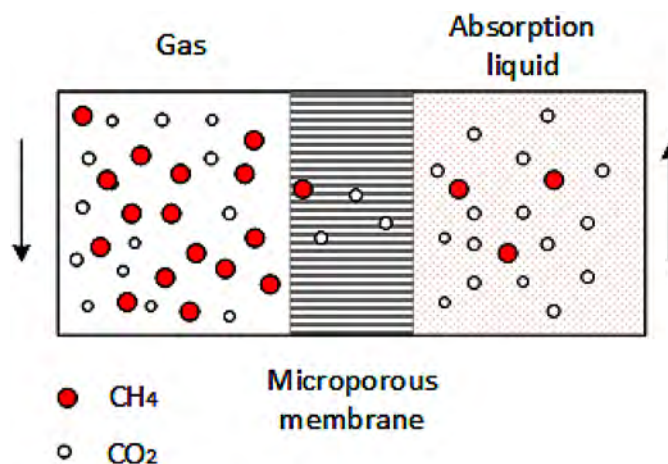


Figure 2. CO₂ molecule transport through microporous membrane

profiles that are formed when species “i” is transferred from the gas phase to the liquid and if hydrophobic membranes are used. In order to describe this mass transfer process between two phases through porous membrane a resistance in series model based on film theory can be used. The resistances to the mass transfer encountered in both cases are those offered by the boundary layers and the membrane and can be drawn, as in Figure 4, by considering an electrical analogy.

As shown in the Figure 4, three resistances exist in the resistance in series model:

- Gas film resistance.
- Liquid film resistance.
- Membrane resistance.

A general expression used to calculate the flux of the species through the membrane is:

$$J = K_{ov}(C_1 - C_2) \quad [1]$$

with

$$\frac{1}{K_{ov}} = \frac{1}{k_G} \frac{d_o}{d_i} + \frac{1}{k_m} + \frac{1}{mEk_L} \quad [2]$$

hence

$$K_{ov} = f(k_G, k_m, k_L, m, E) \quad [3]$$

where J is flux. C_1 and C_2 are concentration of CO₂ in the two phases. K_{ov} is overall mass transfer coefficient. k_G , k_m and k_L are individual mass transfer coefficient of gas phase, membrane and liquid phase respectively. M is physical solubility and E is enhancement factor if any chemical reaction present.

Hence overall mass transfer coefficient in GLMC, be determined by three individual mass transfer coefficients, solubility and enhancement factor which in turn be influenced by so many factors. Firstly, membrane mass transfer coefficient influenced by so many membrane inherent properties such as membrane pore size, porosity and tortuosity which on the other hand controlled by adopted fabrication method. Further membrane mass transfer coefficient also subject to whether the membranes pores are gas filled (ideal non wetted mode) or liquid filled (non-preferred wetted mode). Next, the gas and liquid mass transfer coefficients as well depend in GLMC module configurations and adopted process parameters during separation process such as gas and liquid flow rates, temperature etc. Solubility and enhancement factor depends on the type of solvent used. So detail understanding of these properties is vital important for complete understanding of GLMC. Qi and Cussler were the first to propose the idea of CO₂ absorption by sodium hydroxide in a HFMC. Afterwards GLMC as CO₂ absorber

Figure 3. Concentration profiles formed when species “i” is transferred from one phase to another phase through hydrophobic membrane

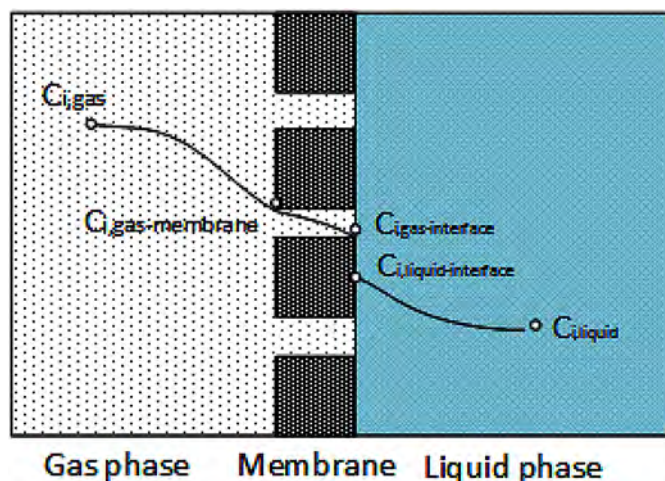
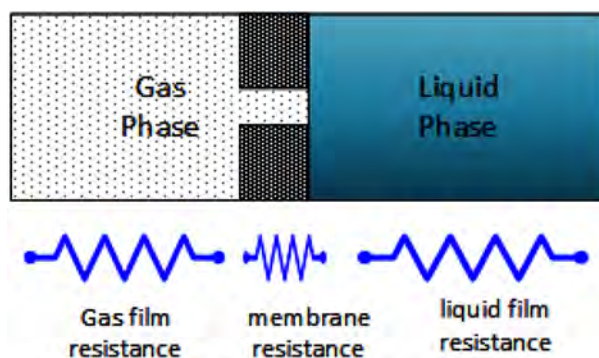


Figure 4. Resistance in series model



has been comprehensively studied by several researchers and exciting experimental and theoretical results have been reported (Ghasem et al., 2011, 2012a, 2012b, 2013, 2014). Currently, efficient and effective CO_2 stripping from liquid absorbents by using membrane contactor has become the target of many researchers in the field and few experimental results have been reported [Rahim et al., 2014, 2015].

MAIN FOCUS OF THE ARTICLE

The schematic diagram in Figure 5 shows the experimental setup used to study the potential of

custom made hollow fiber GLMC for separation of CO_2/CH_4 gas mixture via lean solvent and regeneration of the rich solvent. Homemade and characterized 28% PVDF hollow fiber membranes (Ghasem et al., 2012, 2013) were used to prepare the GLMC modules. Shell and tube type modules were constructed using Perspex tube as shell (transparent) and hollow fiber membranes as tubes. The details of the membrane contactor modules are given in Table 1. Firstly the CO_2 absorption (by using fresh absorption liquids) and stripping (by using manually CO_2 loaded absorption liquids) performance of GLMC was tested separately. Then the absorption experiments were conducted by recycling the absorption liquid several times with

and without the stripping between each recycle to investigate the potential of using GLMC for continuous mode operation.

To study the absorption performance 10/90 vol% CO₂/CH₄ gas mixture was applied to the shell side of the membrane contactor module at different gas inlet flow rate (10 to 100 ml/min) by using mass flow controllers (Alicat Scientific). Different 0.5 M aqueous solutions of primary (MEA), secondary (DEA), hindered (AMP) amines and amine salts (PG) were used as absorption liquid. The liquid feed stream supplied to lumen side of the module at variable flow rate of 10-40 ml/min by using peristaltic pump (Masterflex® L/S®). Also the liquid feed stream was heated for different temperatures; range from room temperature 20 °C to 60 °C by using feedback control heater (WiseStir®). Exit gas concentration was measured using gas chromatography (Shimadzu, Japan). Also the long term absorption performance of absorption liquids was examined by running absorption experiments for 200 minutes. Moreover for PG the ratio between potassium hydroxide and glycine were altered in order to study the effect of solution pH on absorption/stripping performance. Further the recyclability of absorption liquids were studied by recycling the absorption liquid from absorber module with and without stripping between each cycle.

To study the stripping performance the liquid feed stream either coming from the absorber module or manual CO₂ preloaded aqueous solutions (different initial CO₂ loading) were pumped to the stripping module. Pure nitrogen, as a sweep gas,

was made to flow through the module shell side at different gas inlet flow rate (100 to 600 ml/min) by using mass flow controllers (Alicat Scientific). The liquid feed stream supplied to lumen side of the module at variable flow rate of 10-50 ml/min by using peristaltic pump (Masterflex® L/S®). Also the liquid feed stream was heated for different temperatures, range from room temperature 25 °C to 90 °C by using feedback control heater (WiseStir®). A counter-current flow mode was applied for the gas and liquid phases. Variable pressure difference range from 0.01 to 0.05 MPa was applied between the liquid stream and gas stream by using control valve in order to avoid the formation of bubbles on the liquid side and to study the pressure effect. The system was in operation for 30 min to achieve a steady state condition before taking samples. The liquid phase CO₂ concentration at the inlet and outlet of the stripper module was measured by using double chemical titration method (described in the succeeding section), to determine stripping flux and efficiency. Also the sweep gas exit concentration was measured by using gas chromatography (Shimadzu, Japan) in order to make sure the titration results. The experimental CO₂ absorption/stripping efficiency was calculated as:

$$\eta(\%) = \left(\frac{v_i C_i - v_o C_o}{v_i C_i} \right) \times 100 \quad (4)$$

Where C_i and C_o (mol/m³) are the feed gas CO₂ concentration for absorption and liquid phase CO₂ concentration for stripping at the inlet and outlet of the membrane module respectively. v_i and v_o (m³/min) are the feed gas flow rate for absorption and liquid flow rate for stripping at the inlet and outlet of the membrane module respectively

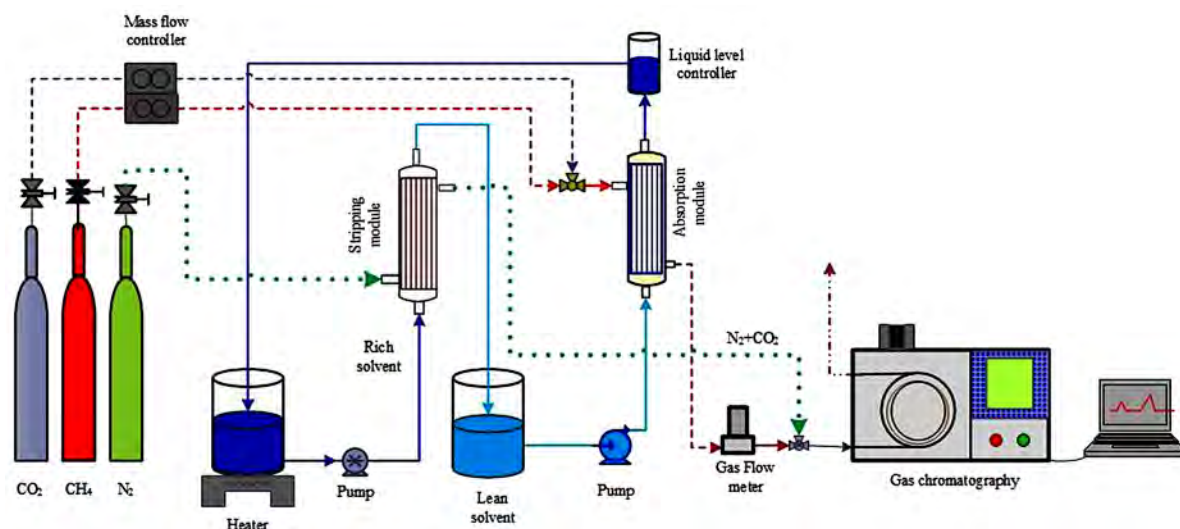
The CO₂ absorption/stripping flux of the module can be calculated as:

$$J_{co_2} = \frac{(v_i C_i - v_o C_o)}{A_i} \quad (5)$$

where J_{co_2} (mol/m² s) is the CO₂ flux and A_i (m²) is inner surface of the hollow fiber membranes.

Table 1. Specifics of the gas liquid membrane contactor module

Fiber	28% PVDF
Module length (mm)	260
Module outer dia. (mm)	15
Module inner dia. (mm)	11
Fiber outer dia. (mm)	1.5
Fiber inner dia. (mm)	0.42
No of fibers	10

Figure 5. CO₂ absorption/stripping experiment setup

Measurement of CO₂ Loading in Liquid Sample

CO₂ concentration in the aqueous solution was determined by double chemical titration method (Seo et al., 1996). An excess amount of 1.0 M NaOH solution was added to the known amount of CO₂ rich liquid sample, converting dissolved CO₂ into the nonvolatile ionic species. Then an excess amount of 1.0 M BaCl₂ solution was added to the solution. The solution then stirred and heated for 3 hour at temperature of 70°C. Absorbed CO₂ precipitates as BaCO₃. The excess NaOH was titrated with 1.0 M HCl solution using phenolphthalein as the indicator. After this titration, using Methyl Orange as the indicator, HCl was added to measure the amount of BaCO₃. The volume of HCl added to neutralize the basic species in the solution was used to calculate the CO₂ loading.

The CO₂ loading was calculated according to the following equation:

$$\alpha = \frac{v_{HCl}}{2 \times v_{sample}} \quad (6)$$

where α = CO₂ loading in mol of CO₂/volume of solution, v_{HCl} = volume of HCl request to neutral-

ize the BaCO₃ in ml, v_{sample} = volume of sample taken for analysis in ml.

SOLUTIONS AND RECOMMENDATIONS

The potential of custom made hollow fiber GLMC for separation of CO₂/CH₄ gas mixture via lean solvent and regeneration of the rich solvent was tested by using different aqueous solutions of primary (MEA), secondary (DEA), hindered (AMP) amines and amine salts (PG) as absorption liquids. The experimental results proved GLMC is promising energy efficient and effective alternative for conventional CO₂ absorption/stripping technology and the type of absorption liquid and operating parameters plays a vital role in eventual absorption/stripping performance given away by GLMC.

CO₂ Absorption Performance of GLMC

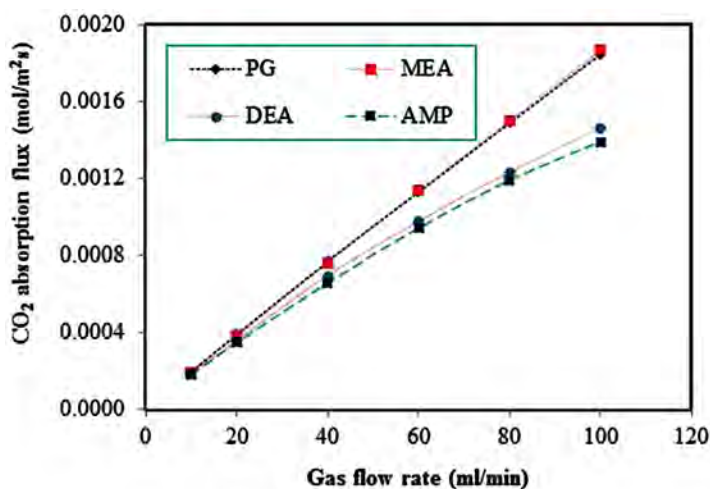
The CO₂ absorption experiments in GLMC revealed the absorption performance order of studied absorbent liquids is PG>MEA> AMP>DEA> regardless of operating conditions adopted (Figure 6). The aqueous amine-based solutions are usu-

ally adopted CO_2 absorption liquids and MEA is the most commonly used amongst them. The structures of alkanolamines include primary, secondary, ternary amines containing at least one OH and amine group such as MEA, DEA and MDEA. The reactivity of amines to CO_2 follows the order primary, secondary and ternary amines. So as obtained MEA has greater absorption performance than the DEA. In addition to these amines, the steric hindrance amines such as AMP also proposed by several researchers. This is because that the steric character reduces the stability of the formed carbamate and easy regenerate the solution. The absorption performance of AMP is better than the DEA as shown in figures. Amino acid salts (AAS) aqueous solutions attract great research interest in recent years as CO_2 absorbent liquid in GLMC because of its prominent characters such as their better affinities towards CO_2 than alkanolamines and their high surface tension because of its ionic nature. Hence mostly used PG absorption performance was compared with other absorbent liquids and its absorption performance was greater than the MEA aqueous solutions. As well even after 8 hours long run the PG did not wet the membrane. Whereas the amine solutions cause membrane wetting averagely after five six

hours of continuous running, this is due to its low surface tension.

Moreover from the absorption experiments it was observed the gas flow rate have a significant effect on CO_2 absorption performance in GLMC. Although as shown in Figure 6 the CO_2 absorption flux increased with gas flow rate the CO_2 removal efficiency decreased with increased gas flow rate. For instance for PG the removal efficiency reduced from 99% to 94% with the increased of gas flow rate from 10 ml/min to 100 ml/min. This can be attributed to decrease in contact time and increase in driving force for mass transfer. Increasing the gas flow rate decreases the residence time of the gas phase in the membrane contactor and hence contact time of the gas phase with liquid. On the other hand the increase in the gas velocity results in the reduction of the boundary layer and the improvement of the total mass-transfer rate. Further it was observed for any absorption liquid the CO_2 absorption performance in GLMC could be enhanced by high liquid flow rates. This enhancement in absorption performance was significant for the absorption liquids which have poor absorption performance than the absorption liquids which have good absorption performance. For PG the removal efficiency increased only by

Figure 6. CO_2 absorption performance of various absorption liquids in GLMC (0.5M aqueous solutions at room temperature flow in lumen side at 10 ml/min; gas mixture 9% CO_2 /91% CH_4 flows in shell side)



3.3%, whereas for DEA the removal efficiency increased by 17% with increase of liquid flow rate from 10 ml/min to 50 ml/min. This is attributed to higher liquid velocity leads to a lower CO_2 concentration in the liquid phase, which in turn results in a higher CO_2 concentration gradient between gas and liquid phase phases. Also increasing the liquid velocity turns the fluid flow form laminar to turbulent this reduces the liquid phase boundary layer.

The effect of temperature on CO_2 absorption performance in GLMC is attributed to collective effect of solubility (physical absorption), chemical reaction (chemical absorption), diffusion and evaporation of absorbent. Further increase in temperature would decrease the viscosity of the solution which is favorable character of GLMC absorbent liquid. It is well known the favored chemical reaction rate and diffusion rate with temperature would enhance the absorption performance. While decrease in CO_2 solubility and an increase in evaporation of absorbent (wetting) with temperature would reduce the absorption performance. Experimentally it was observed MEA and shows no any significant effect of temperature on absorption performance contrary to the results obtained with AMP and DEA. This may be due to for MEA and PG the favored chemical reaction and diffusion with temperature were compensated by the reduced solubility. Also these absorbent have good reactivity even at room temperature (around 95% removal), so the enrichment in absorption performance with temperature was not highlighted. Whereas for AMP and DEA the favored effect of temperature on chemical reaction and diffusion were higher than the reduced effect temperature on CO_2 solubility and so the enrichment in absorption performance with temperature was significant. However all solvents shows some flattening off at elevated temperatures. Because the wetting caused by elevated temperature cannot be neglected in long term operation. Further the thermal degradation of membrane material also should be considered. So, solvent temperature is

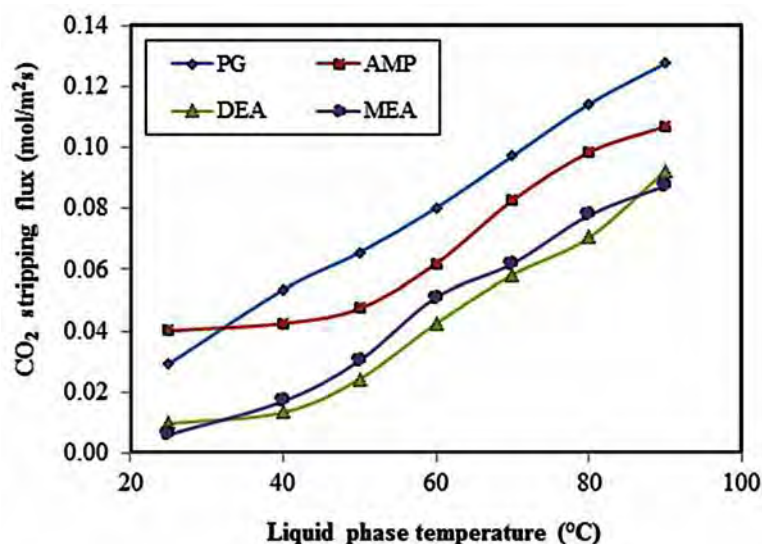
a crucial factor to be controlled very carefully for long term operation performance.

CO_2 Stripping Performance of GLMC

CO_2 stripping performance of PVDF hollow fiber GLMC was compared between primary (MEA), secondary (DEA), hindered (AMP) amines and amine salts (PG). The liquid solutions were manually preloaded with CO_2 for saturation level in order to obtain higher CO_2 flux for comparing purposes. In terms of regeneration AMP and PG perform well as absorbent liquid comparing to MEA and DEA. The results in Figure 7 show the effect of different aqueous solution on stripping flux as function of temperature. The regeneration energy of the absorbents was directly related with the heat of reaction and showed higher values when the binding force between amines and CO_2 molecules was larger. Therefore, absorbents that formed carbamate required larger quantities of heat, and the absorption heats of the MEA and DEA were higher than those of AMP and PG. So AMP and PG shows better stripping performance compare to MEA and DEA. Regardless of type of solvent the CO_2 stripping flux and efficiency increases with temperature rapidly. Because as the reaction rate increases with temperature, the formation of carbamate, which is the final product of CO_2 reaction, become unstable and the energy consumed in regeneration become smaller.

The studies on sweep gas flow rates exposed the gas flow rate have no any significant effect on stripping flux and efficiency regardless of type of solvent. However, when operating at high rich solution temperature, the low sweep gas flow rate, allow the vapor molecules to easily enter through the pores and wet the membrane. This reduces the effective long term operation of the membrane contactor module. So, moderate sweep gas flow rate gives better performance comparing to low sweep gas flow rate. Liquid flow rate shows some notable criteria on CO_2 stripping flux and efficiency. At low temperatures the increase in

Figure 7. Effect of liquid phase temperature on CO_2 stripping flux for four different solvents (saturation initial CO_2 loading, $v_l = 50 \text{ ml/min}$, $v_g = 600 \text{ ml/min}$ and $P_l - P_g = 0.5 \times 10^5 \text{ Pa}$)



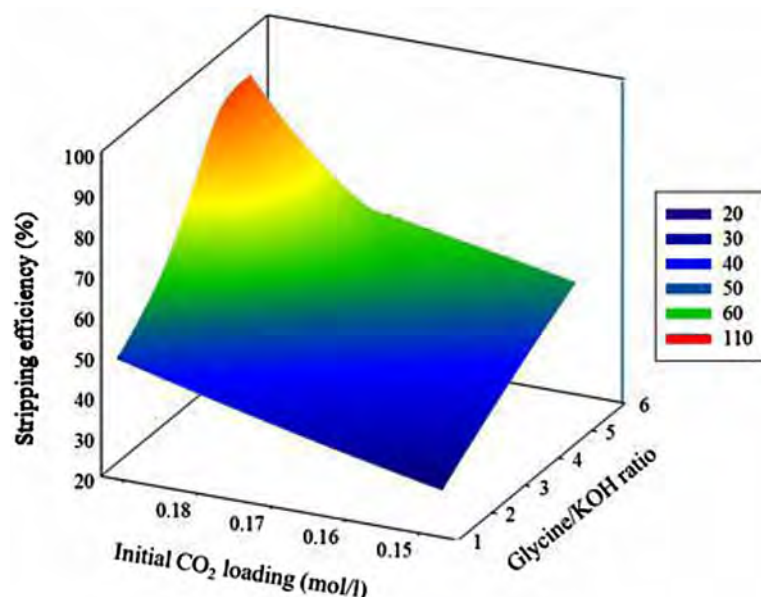
liquid flow rate reduces the stripping efficiency. In contrast at high temperature the increase in liquid flow rate increases the stripping efficiency. Two phenomenon govern the stripping efficiency. The residence time (contact time between gas and liquid phase) and liquid phase boundary layer thickness. At low temperature contact time overtaken by the boundary layer thickness. Lower liquid velocity tend to higher residence time, which lets dissolved CO_2 to shift to gas-liquid interface and results in increasing the driving force of mass transfer. In contrast at high temperature due to low solubility more CO_2 end to be released. So rather than residence time boundary layer thickness is important. Because the liquid flow rate reduces the boundary layer thickness and increase the mass transfer coefficient, at high temperature high flow rates gives the high removal efficiency.

The liquid pressure also has great influence in the stripping performance. The considerable increase in CO_2 flux with liquid pressure can be attributed to the increase in driving force for desorption as a result of an increase in CO_2 concentration. Many researchers proved the CO_2 absorption flux increase drastically with gas phase pressure. So vice versa by increasing the liquid side

pressure the CO_2 desorption flux can be increased. But, even though, hydrophobic PVDF hollow fiber membrane can resist wetting, applying higher pressure in the liquid side can gradually cause wetting. It is well known that partial wetting of membrane can increase mass transfer resistance significantly. So the liquid side pressure should be maintained below the break through pressure.

The effect of absorption liquid pH on CO_2 absorption/stripping performance in gas liquid membrane contactor module was investigated by using asymmetric solutions of PG (i.e. solutions containing different molar amounts of amino acid (glycine) and base (potassium hydroxide)). At lower pH the CO_2 reaction equilibrium shifts towards the release of CO_2 enhancing the partial pressure of CO_2 . So higher molar ratios of amino acid shows better stripping and reduced absorption performance (Figure 8). Subsequently in continuous operation of absorption and followed up by stripping, the stripping performance was controlled by reduced absorption performance because of resultant low initial CO_2 concentrations. Hence for the net effect the molar ratio has to be optimized to achieve both good absorption and subsequent stripping performance.

Figure 8. Effect of different asymmetric solutions on overall efficiency



FUTURE RESEARCH DIRECTIONS

GLMC due to attested several advantages has become the promising alternative for conventional CO₂ absorption/stripping process. Although the GLMC for CO₂ separation has been extensively studied there is still a long way ahead before this technique to completely replace the existing CO₂ separation technology. The main hitch is long term stability of GLMC absorption process due to membrane wetting. The key is employing surface modified super hydrophobic membranes and high surface tension absorption liquids in GLMC absorption/stripping processes. Currently, various techniques such as surface grafting, bore filling grafting, coating/interfacial polymerization and in-situ polymerization are being investigated to improve the surface hydrophobicity. Some works are going on using ionic liquid membranes in GLMC applications. Further as a consequences of seek out for new absorption liquids that not only have high surface tension but also can be regenerable in efficient way, AAS and ionic liquids have attracted great research interest. Most importantly AASs have favorable biodegradation properties,

which make the disposal of these solvents easier and with lower environmental impacts. In spite of all these work still the membrane wetting at elevated temperatures is arguing. Moreover, despite the fact that solvent regeneration is responsible for the major cost component in gas separation processes due to energy consumption, the studies on CO₂ stripping using GLMC have started recently and there are only few reports documented in the open literature. Hence future research directions should focus on super hydrophobic membranes, absorption liquids that can be regenerated in an energy efficient way and eventually suitable membrane-absorbent combination, therefore, GLMC will thrive as a perfectly energy efficient and effective CO₂ absorption/stripping technology, which can replace the current CO₂ separation technology completely.

CONCLUSION

The experimental results proved GLMC is promising energy efficient and effective alternative for conventional CO₂ absorption/stripping technology

and the type of absorption liquid and operating parameters plays a vital role in eventual absorption/stripping performance given away by GLMC. Studied AAS (PG) shows both good absorption and stripping performance. The gas flow rate has a significant effect on CO₂ absorption performance where as it has no effect in stripping performance. Further the CO₂ absorption performance in GLMC could be enhanced by high liquid flow rates. This enhancement was significant for the absorption liquids which have poor absorption performance than the absorption liquids which have good absorption performance. In contrast, the stripping performance enhancement with liquid flow rate depends on liquid temperature. Because the gas-liquid contact time was a key factor to enhance the stripping flux at low temperature while liquid phase boundary layer thickness and associated mass transfer resistance is important at elevated temperature. So by controlling the liquid phase velocity and the length of module at low temperature better stripping performance can be achieved. The effect of liquid temperature on absorption performance in GLMC is not straightforward. Since the liquid temperature cooperatively influence the several factors which determines the absorption performance positively and negatively, it should be handled in care depending on the system used. However by increasing the rich solution temperature the stripping performance can be improved preferably. To improve the stripping performance enhanced CO₂ desorption (based on pH-shift) was studied. In this study the pH values were changed by using asymmetric solutions of PG. Lower pH shifts the reaction equilibrium towards the release of CO₂ and shows better stripping and reduced absorption performance. Thus for the net effect the molar ratio has to be optimized to achieve both good absorption and subsequent stripping performance.

ACKNOWLEDGMENT

The authors would like to acknowledge the financial support provided by the research and graduate study of the United Arab Emirates University (UAEU) and National Research Fund (NRF). Grant number 31N168-UPAR (9) 2013.

REFERENCES

- Ghasem, N., Al-Marzouqi, M. H., & Rahim, N. A. (2012a). Effect of polymer extrusion temperature on poly (vinylidene fluoride) hollow fiber membranes: Properties and performance used as gas-liquid membrane contactor for CO₂ absorption. *Separation and Purification Technology*, 99, 91–103. doi:10.1016/j.seppur.2012.07.021
- Ghasem, N., Al-Marzouqi, M. H., & Rahim, N. A. (2014). Absorption of CO₂ Form Natural Gas via Gas-liquid PVDF Hollow Fiber Membrane Contactor and Potassium Glycinate as Solvent. *Jurnal Teknologi*, 69(9), 121–126. doi:10.11113/jt.v69.3409
- Ghasem, N. M., Al-Marzouqi, M., & Zhu, L. P. (2012b). Preparation and properties of polyether-sulfone hollow fiber membranes with o-xylene as a additive used in membrane contactors for CO₂ absorption. *Separation and Purification Technology*, 92, 1–10. doi:10.1016/j.seppur.2012.03.005
- Ghasem, N. M., Al-Marzouqi, M. H., & Duaidar, A. (2011). Effect of quenching temperature on the performance of poly (vinylidene fluoride) microporous hollow fiber membranes fabricated via thermally induced phase separation technique on the removal of CO₂ from CO₂ gas mixture. *International Journal of Greenhouse Gas Control*, 5(6), 1550–1558. doi:10.1016/j.ijggc.2011.08.012

Ghasem, N. M., Al-Marzouqi, M. H., & Rahim, N. A. (2013). Modeling of CO₂ absorption in a membrane contactor considering solvent evaporation. *Separation and Purification Technology*, 110, 1–10. doi:10.1016/j.seppur.2013.03.008

Naim, R., & Ismail, A. F. (2013). Effect of polymer concentration on the structure and performance of PEI hollow fiber membrane contactor for CO₂ stripping. *Journal of Hazardous Materials*, 250, 354–361. doi:10.1016/j.jhazmat.2013.01.083 PMID:23474409

Rahim, N. A., Ghasem, N., & Al-Marzouqi, M. H. (2014). Stripping of CO₂ from different aqueous solvents using PVDF hollow fiber membrane contacting process. *Journal of Natural Gas Science and Engineering*, 21, 886–893. doi:10.1016/j.jngse.2014.10.016

Rahim, N. A., Ghasem, N., & Al-Marzouqi, M. H. (2015). Absorption of CO₂ from natural gas using different amino acid salt solutions and regeneration using hollow fiber membrane contactors. *Journal of Natural Gas Science and Engineering*, 26, 108–117. doi:10.1016/j.jngse.2015.06.010

Seo, D.-J., & Hong, W.-H. (1996). Solubilities of carbon dioxide in aqueous mixtures of diethanolamine and 2-amino-2-methyl-1-propanol. *Journal of Chemical & Engineering Data*, 41(2), 258–260. doi:10.1021/jc950197o

Ze, Z., & Sx, J. (2014). Hollow fiber membrane contactor absorption of CO₂ from the flue gas: Review and perspective. *Global NEST Journal*, 16, 355–374.

ADDITIONAL READING

Drioli, E., Criscuoli, A., & Curcio, E. (2011). *Membrane Contactors: Fundamentals, Applications and Potentialities: Fundamentals, Applications and Potentialities* (Vol. 11). Elsevier.

Ghasem, N. M., & Al-Marzouqi, M. (2014). Modeling and Experimental Study of Gas-Liquid Membrane Contactor. *Encyclopedia of Information Science and Technology*, Third Edition (10 Volumes) July, 2014. Chapter 537, pages 5442–5453. Doi:10.4018/978-1-4666-5888-2

Li, J.-L., & Chen, B.-H. (2005). Review of CO₂ absorption using chemical solvents in hollow fiber membrane contactors. *Separation and Purification Technology*, 41(2), 109–122. doi:10.1016/j.seppur.2004.09.008

Mansourizadeh, A., & Ismail, A. F. (2009). Hollow fiber gas–liquid membrane contactors for acid gas capture: A review. *Journal of Hazardous Materials*, 171(1-3), 38–53. doi:10.1016/j.jhazmat.2009.06.026 PMID:19616376

Mansourizadeh, A., & Ismail, A. F. (2011). CO₂ stripping from water through porous PVDF hollow fiber membrane contactor. *Desalination*, 273(2-3), 386–390. doi:10.1016/j.desal.2011.01.055

Mulder, M. (1997). *Basic Principles of Membrane Technology*. Kluwer Academic Publishers.

Naim, R., Ismail, A. F., Cheer, N. B., & Abdullah, M. S. (2014). Polyvinylidene fluoride and polyetherimide hollow fiber membranes for CO₂ stripping in membrane contactor. *Chemical Engineering Research & Design*, 92(7), 1391–1398. doi:10.1016/j.cherd.2013.12.001

Rahbari-Sisakht, M., Ismail, A. F., Rana, D., & Matsuura, T. (2013). Carbon dioxide stripping from diethanolamine solution through porous surface modified PVDF hollow fiber membrane contactor. *Journal of Membrane Science*, 427, 270–275. doi:10.1016/j.memsci.2012.09.060

Simioni, M., Kentish, S. E., & Stevens, G. W. (2011). Membrane stripping: Desorption of carbon dioxide from alkali solvents. *Journal of Membrane Science*, 378(1-2), 18–27. doi:10.1016/j.memsci.2010.12.046

Yan, S., He, Q., Zhao, S., Zhai, H., Cao, M., & Ai, P. (2015). CO₂ removal from biogas by using green amino acid salts: Performance evaluation. *Fuel Processing Technology*, 129, 203–212. doi:10.1016/j.fuproc.2014.09.019

KEY TERMS AND DEFINITIONS

AASS: Amino acid salt solution. Chemical solvent derived by mixing amino acid and alkaline hydroxides.

CO₂ Absorption: Operation used in removing acid gas from other gases using liquid solvent.

CO₂ Stripping: Operation used in removing absorbed acid gas from absorbent liquid.

GLMC: Gas Liquid membrane contactor. A device that have bundle of fibers enclosed on a shell in which gas and liquid flow without dispersing with each other.

Hydrophobic: The not-affinity to water.

Natural Gas: Gas consists with more than seventy percent methane.

PVDF: Polyvinylidene fluoride, polymer used in fabrication.

Wetting: Membrane pores filled with liquid instead of gas.

E

Enhancing the Resiliency of Smart Grid Monitoring and Control

Wenbing Zhao

Cleveland State University, USA

INTRODUCTION

Smart grid is one of the hottest research areas in recent years. The development of smart grid is partially driven by the fact that the traditional data communication infrastructure for electric power grid can no longer meet the needs of new developments (Wang, Xu, & Khanna, 2011):

- The recent deregulation would allow many independent parties to enter the utility industry by offering alternative channels for electric power generation, distribution, and trade. This inevitably demands timely, reliable and secure information exchanges among these parties (Bose, 2005).
- The current data communication infrastructure lacks the support for large-scale real-time coordination among different electric power grid health monitoring and control systems, which could have prevented the 2003 massive blackout incident in North America (Birman et al., 2005).
- The use of modern computer networking technology could also revolutionize the everyday electric power grid operations, as shown by the huge benefits of substation automation and the use of Phasor Measurement Units (PMUs) for electric power grid health monitoring (Melliopoulos, 2007).

However, the openness and the ease of information sharing and cooperation brought by smart grid also increased the likelihood of cyber attacks on the electric power grid, as demonstrated

recently by an experiment conducted by the US Department of Energy's Idaho Lab (CNN, 2007). To address such vulnerability, intrusion detection and intrusion tolerance techniques must be used to enhance the current and future data communication infrastructure for the electric power grid. Byzantine fault tolerance is a fundamental technique to achieve the objective (Castro & Liskov, 2002; Zhao, 2014a).

In this chapter, we focus our discussions on the security and reliability of smart grid health monitoring and control. We elaborate in detail the need for Byzantine fault tolerance and the challenges of applying Byzantine fault tolerance into this problem domain. In particular, we investigate experimentally the feasibility of using such sophisticated technology to meet potentially very stringent real-time requirement for the health monitoring and control of smart grid, while ensuring high degree of reliability and security of the system.

BACKGROUND

A Byzantine faulty process may behave arbitrarily. In particular, it may disseminate conflicting information to different components of a system, which constitutes a serious threat to the integrity of a system (Lamport, Shostak, & Pease, 1982). Because a Byzantine faulty process may also choose not to send a message, or refuse to respond to requests, it can exhibit crash fault behavior as well. Consider the scenario that multiple PMUs periodically report their measurement results to a controller for electric power grid health monitor-

ing. When it detects an abnormality, the controller may wish to issue specific control instructions to the actuating devices, such as Intelligent Electronic Devices (IEDs) (Hossenlopp, 2007) located at the same substation as those PMUs to alleviate the problem. Due to the critical role played by the controller, it must be replicated to ensure high availability. Otherwise, the controller would become a single-point of failure. The main components and their interactions are illustrated in Figure 1.

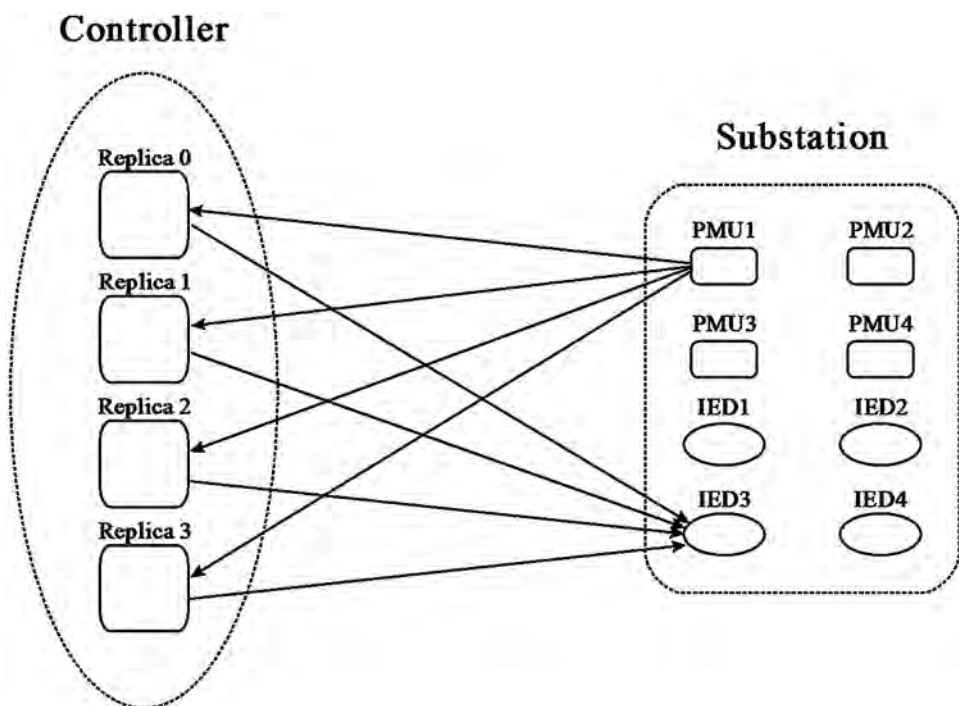
However, the controller replicas, the PMUs, and the IEDs, might be compromised under cyber attacks. Consider the following two scenarios:

- A Byzantine faulty PMU could potentially send inconsistent data to different controller replicas. Without proper coordination among the controller replicas, the state of the replicas might diverge in the former case, which would lead to inconsistent decisions among the replicas.

- A compromised controller replica could send conflicting commands to different IEDs. Without a sound mechanism at each IED, a malicious command might be executed in the latter case, which could lead to the destruction of a generator or a transmission line, as reported by CNN (2007).

Byzantine fault tolerance (BFT) refers to the capability of a system to tolerate Byzantine faults (Lamport, Shostak, & Pease, 1982). If BFT is used, the cyber attacks illustrated above could be defeated provided that the number of compromised controller replicas, f , is below a threshold, and the number of non-faulty PMUs and IEDs are sufficient for the normal operation of the substation. For the client-server system shown in Figure 1, BFT can be achieved by using $3f + 1$ replicas to tolerate up to f faulty replicas and by ensuring all non-faulty replicas to execute the same set of requests in the same order. The latter means that the server replicas must reach an agreement on the set of requests and their relative ordering

Figure 1. The interaction of substation devices (PUMs and IEDs) and the controller replicas



despite the presence of Byzantine faulty replicas and clients. Such an agreement is often referred to as a Byzantine agreement (Lamport, Shostak, & Pease, 1982). The Byzantine agreement among the replicas ensures that a faulty client (i.e., a PMU) cannot cause the divergence of the state of non-faulty controller replicas. Furthermore, a Byzantine agreement must be reached among all non-faulty replicas on each command issued by the controller for reasons to be explained in the next section. Before an IED can accept the command, it must wait until it has collected at least $f + 1$ identical command from different replicas.

We should note that Byzantine fault tolerance has been a hot research area in many other areas, such as Web services (Merideth et al., 2005; Zhao, 2009) and data storage systems (Rhea et al., 2003). Even though the works are in a different context, many insights are useful for BFT controls in electric power grid applications. In particular, the mechanisms designed to cope with the interaction of a replicated object and the un-replicated external entities reported in (Merideth et al., 2005) have been partially incorporated in this work.

BYZANTINE FAULT TOLERANT MONITORING AND CONTROL MECHANISMS

In this work, we choose to use PBFT, a well-known Byzantine agreement algorithm developed by Castro and Liskov (2002). The PBFT algorithm is designed to support client-server applications running in an asynchronous distributed environment with the Byzantine fault model. The implementation of the algorithm contains two parts. At the client side, a lightweight library is in charge of sending the client's request to the primary replica, retransmitting the request to all server replicas on the expiration of a retransmission timer (to cope with the primary faults and network faults), and collecting and voting on the corresponding replies. The main PBFT algorithm is executed at the server side by a set of $3f + 1$ replicas to toler-

ate up to f faulty replicas. One of the replicas is designated as the primary while the remaining are backups. Furthermore, all messages are protected by a digital signature, or an authenticator (Castro & Liskov, 2002) so that a faulty replica or client cannot tamper with the messages and cannot impersonate as another non-faulty replica or client.

The normal operation of the (server-side) PBFT algorithm involves three phases: pre-prepare, prepare, and commit. In the beginning of the pre-prepare phase, the primary multicasts a pre-prepare message containing the client's request, the current view number and a sequence number assigned to the request to all backups. A backup verifies the request message and the ordering information. If the backup accepts the message, it multicasts to all other replicas a prepare message containing the ordering information and the digest of the request being ordered. This starts the prepare phase. A replica waits until it has collected $2f$ prepare messages from different replicas (including the message it has sent if it is a backup) that match the pre-prepare message before it multicasts a commit message to other replicas, which starts the commit phase. The commit phase ends when a replica has received $2f$ matching commit messages from other replicas. At this point, the request message has been totally ordered and it is ready to be delivered to the server application if all previous requests have already been delivered. If the primary or the client is faulty, the Byzantine agreement on the ordering of a request might not be reached, in which case, a new view is initiated, triggered by a timeout on the current message being ordered. A different primary is designated in a round-robin fashion for each new view installed.

For electric power grid health monitoring and control, however, the above BFT algorithm cannot be used directly, because normally the controller replicas collect input from the PMUs and the control commands are issued to IEDs. Furthermore, the updates from PMUs are one-way messages in that the PMUs normally do not wait for an explicit response from the controller. On the other hand, IEDs are acting as the server role

when it receives the control commands from the controller replicas. Table 1 provides a summary of the actions taken by the controller replicas in response to receiving a report from a PMU and to sending of a command to an IED.

On collecting PMU data, the controller replicas engage in a Byzantine agreement for each input message as usual, as shown in Figure 2. However, the message delivery procedure must be modified. When a replica reaches a Byzantine agreement on the message, and it has delivered all previously ordered messages, it invokes the callback function provided by the controller to deliver this message. On the return of the up-call, no message is sent back to the PMU.

Upon issuing a control command, a controller replica does not directly send the command to the target IED. Instead, a round of Byzantine agreement on the command message is conducted, as shown in Figure 3. The procedure is very similar to that of PMU input message ordering, except that the pre-prepare message is triggered by the issuing of a control command rather than the receiving of a client's request, and the command is sent to the target IED when the Byzantine agreement is reached, instead of delivering a request. As mentioned in the previous Section, the target IED

must not accept a control command immediately because the command might have been sent by a faulty controller replica. By waiting until it has received $f + 1$ identical command from different controller replicas, it can guarantee that at least one of them is sent by a non-faulty replica, because at most f replicas can be faulty according to our assumption.

If the replicas operate completely deterministically and in lock-step, the round of Byzantine agreement for the commands to the IEDs is not necessary. However, it is virtually impossible to guarantee lock-stepped execution of the replicas across a network. If the control command contains information such as the time to execute the command, the commands issued by different replicas would contain different timestamps, which would make it impossible for the IEDs to authenticate and compare the commands for acceptance. Therefore, in general, it is necessary for the replicas to reach an agreement on the command to be issued to the IEDs. Here, we assume that a backup replica is able to verify if the command proposed by the primary is valid. A backup would initiate a view change if it deems that the command from the primary is invalid. If a backup replica cannot verify the command issued by the primary,

Figure 2. Normal operation of the BFT algorithm for PMU report handling

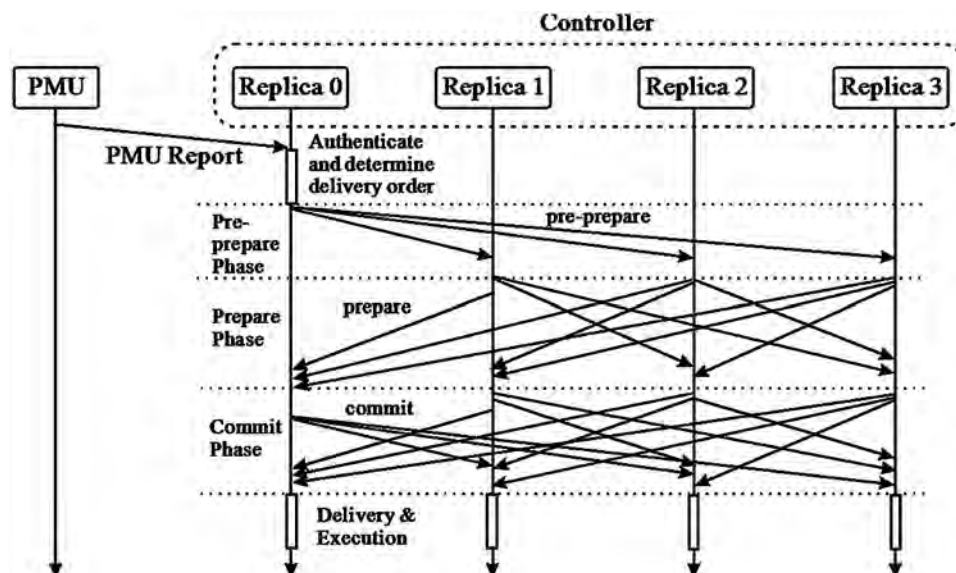


Figure 3. Byzantine agreement (normal operation) on the control command at the controller replicas and the acceptance of the command at an IED

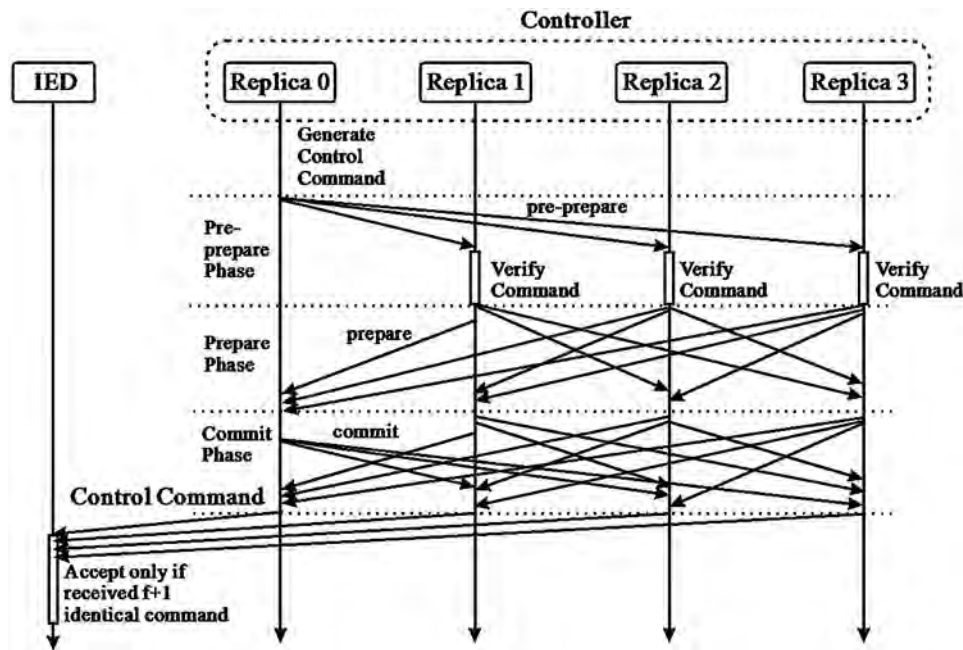


Table 1. A summary of the actions taken by the controller replicas in response to receiving a report from a PMU and to sending of a command to an IED

Event	Actions Taken
On receiving a report from PMU	Byzantine agreement on the report No message is sent back to PMU
On sending of a command to IED	Byzantine agreement on the command prior to sending IED must collect $f+1$ consistent commands from different replicas

more sophisticated mechanisms must be used, as reported in (Zhang et al., 2011).

EXPERIMENTAL ASSESSMENT

The implementation of our Byzantine fault tolerance framework is based on the PBFT library developed by Castro and Liskov (2002). We incorporated the changes necessary for electric power grid monitoring and control as mentioned in the previous Section.

The test-bed consists of 12 PCs in a local-area network (LAN) connected by a 100Mbps Ethernet. Four of the PCs in the LAN are equipped

with Pentium IV 2.8GHz processors and 256MB memory, and the remaining PCs in the LAN each has a single Pentium III 1GHz processor. All PCs on the LAN run the Red Hat 8.0 Linux. The remote PC has one Pentium IV 3.2GHz processor running CentOS 4.5 Linux.

The main objective of the performance evaluation is to assess whether or not the Byzantine fault tolerance mechanisms are efficient enough to meet the real-time communication requirement for power grid health monitoring and control. Consequently, we characterize the response time and jitter of the Byzantine fault tolerant system.

The test application simulates the electric power grid health monitoring and control scenario

as shown in Figure 1. The controller is replicated in the 4 Pentium IV PCs (one replica per PC) and the PMUs and IEDs are run on the remaining 8 Pentium III PCs (a pair of PMU and IED on each PC). During the experiments, up to 8 concurrent PMU-IED pairs are used.

A PMU (as the client) periodically reports its measured data to the replicated controller according to the IEEE 1344 standard. Upon each PMU message received, the controller replicas generate a command and send it to the corresponding IED (collocated on the same node as the reporting PMU). Note that this is done purely for the purpose of performance characterization and might not match the practical usage scenarios. The payload of each PMU report is 14 bytes long. The payload of each control command is set to 128 bytes long.

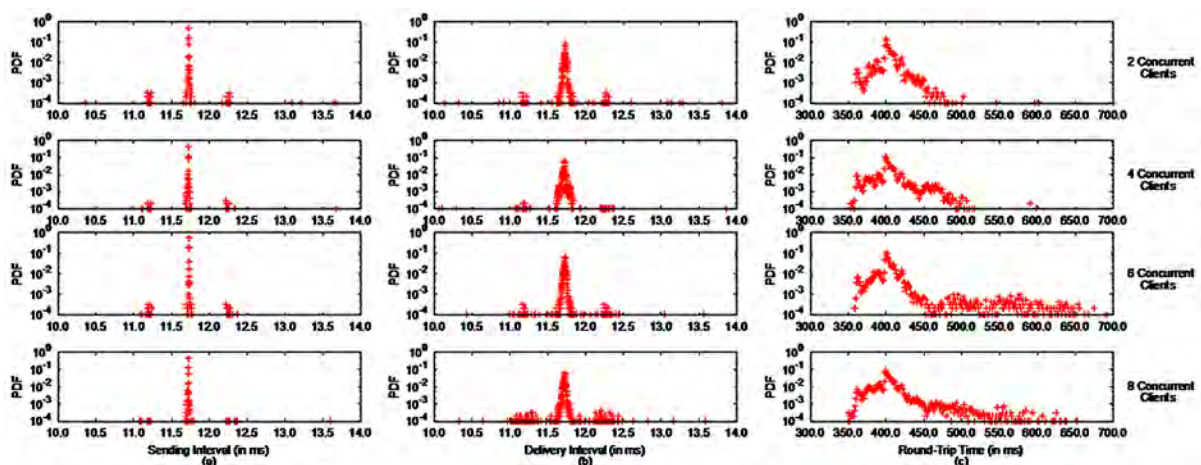
Furthermore, the PCs in our test-bed are not equipped with high-resolution GPS devices, preventing us from directly measuring one-way latencies for PMU reports and control command notifications. Instead, we measure the round-trip time from the sending of a PMU report to the receiving of a command in response to the report at a collocated IED.

To gain insight on the jitters of networking and Byzantine agreement processing delays, we

measure the intervals between two consecutive sending of PMU reports at each PMU and the intervals of consecutive deliveries of the PMU reports at each controller replica, and compare the probability density functions (PDFs) of the sending intervals and the delivery intervals. The PDFs provide a much more detailed and accurate picture on the predictability of the arrival rate of the PMU reports at the controller replicas than using the mean values and standard deviations. For similar reasons, the PDF is used to capture and present the round-trip times. In each run, 10,000 samples are taken.

Figure 4 shows the experimental results under the normal operation condition. The number of concurrent PMU-IED pairs varies from 1 to 8. The PDFs for the sending intervals measured at the PMUs are shown in Figure 4(a). The interval between two consecutive sending is controlled by the `nanosleep()` API provided by Linux. Even though the target interval is 10 milliseconds, the actual intervals vary slightly (with peak value of about 11.6 milliseconds). If there is no jitter in networking and Byzantine agreement processing, the PDFs of the delivery intervals measured at the controller replicas should be identical to those of the sending intervals. The PDFs of the

Figure 4. The measured PDFs with various number of PMU-IED pairs. (a) The report-sending interval at the PMU. (b) The delivery interval at the controller replicas. (c) The round-trip time from sending a PMU report and the receiving of a control command



delivery intervals shown in Figure 4(b) indicate that there is noticeable jitter. However, the jitter is small enough to sustain a 60Hz PMU sampling rate for all scenarios tested (up to 8 concurrent PMU-IED pairs), which is often regarded as the most demanding SCADA requirement (Johnston, 2005). Furthermore, Figure 4(c) shows that the round-trip time is in the sub- millisecond range, again for all scenarios measured, which is more than adequate to ensure urgent sensing data delivered and control command acted upon.

When the primary controller replica is faulty, it may take a significant amount of time (e.g., 2 seconds) for a view change to complete. During this period of time, the controller is basically out-of-service. To address this issue, the controller should periodically send contingency control commands to the IEDs. If an IED does not receive a control command in time, it should resort to the contingency command. We emphasize that this situation, even though not desirable, is far better than the IED executing a command sent by a malicious controller, which can lead to the destruction of critical components of the power grid.

FUTURE RESEARCH DIRECTIONS

There is a large body of research work on how to restructure the current data communication infrastructure for electric power grid, such as (Zhang, Wang, & Xiang, 2015; Birman et al., 2005; Bose, 2005; Hossenlopp, 2007; Melliopoulos, 2007). The SCADA security issues have also attracted worldwide attention (<http://sandia.gov/scada/>). However, the work that targets both the security and reliability aspects of the infrastructure is rarely seen. Our work appears to be the first to assess if it is possible to apply the Byzantine fault tolerance technology for electric power grid health monitoring and control. We plan to carry out more in-depth investigations. In particular, the performance of the BFT technology in the wide-area network environment should be carefully evaluated. Furthermore, we also plan to explore

context-ware and adaptive fault tolerance (Buys et al., 2011) to improve the performance and robustness of our framework. Another research direction is to improve the performance of the proposed framework by adapting our replication protocol specific for SCADA interactions. We call this line of research application-aware Byzantine fault tolerance (Zhao, 2014a; Zhao, 2014b).

The importance of stable sampling rate for networked sensing and control is addressed in (Liberatore, 2006). In (Liberatore, 2006), Liberatore proposed a playback-based method to increase the predictability of the sampling rate. This method seems to be a good candidate to be integrated with the proposed framework to address the potential jittering problem in the wide-area networks.

CONCLUSION

In this chapter, we presented the justification and a feasibility study of applying the Byzantine fault tolerance technology to electric power grid health monitoring. We proposed and implemented the BFT mechanisms needed to handle the PMU data reporting and control commands issuing to the IEDs. We carried out an empirical study to assess the feasibility of using the BFT technology for reliable and secure electric power grid health monitoring and control. We show that under the LAN environment, the overhead and jitter introduced by the BFT mechanisms are negligible, and consequently, Byzantine fault tolerance could readily be used to improve the security and reliability of electric power grid monitoring and control while meeting the stringent real-time communication requirement for SCADA operations.

While the brief out-of-service time (typically in 1-2 seconds) during a view change can be a concern for electric power grid health monitoring and control, additional mechanism, such as the playback scheme proposed in (Liberatore, 2006) could be used to alleviate the problem. In any case, the BFT sensing and control ensures that a PMU report from a compromised PMU

cannot cause the state divergence of the correct controller replica, and a control command from a compromised control replica is never accepted by a correct actuating device such as an IED.

REFERENCES

- Birman, K. P., Chen, J., Hopkinson, E. M., Thomas, R. J., Thorp, J. S., Renesse, R. V., & Vogels, W. (2005). Overcoming communications challenges in software for monitoring and controlling power systems. *Proceedings of the IEEE*, 93(5), 1028–1041. doi:10.1109/JPROC.2005.846339
- Bose, A. (2005). Improved communication/computation infrastructure for better monitoring and control. In *Proceedings of the IEEE Power Engineering Society General Meeting* (pp. 2715–2716). doi:10.1109/PES.2005.1489755
- Buys, J., De Florio, V., & Blondia, C. (2011). Towards context-aware adaptive fault tolerance in SOA applications. In *Proceedings of the 5th ACM international conference on Distributed event-based system*, (pp. 63–74). doi:10.1145/2002259.2002271
- Castro, M., & Liskov, B. (2002). Practical Byzantine fault tolerance and proactive recovery. *ACM Transactions on Computer Systems*, 20(4), 398–461. doi:10.1145/571637.571640
- CNN. (2007). *Staged cyber attack reveals vulnerability in power grid*. Retrieved from <http://www.cnn.com/2007/US/09/26/power.at.risk/index.html>
- Hossenlopp, L. (2007). Engineering perspectives on iec 61850. *IEEE Power and Energy Magazine*, 5(3), 45–50. doi:10.1109/MPAE.2007.365819
- Johnston, R. (2005). *Obtaining high performance phasor measurements in a geographically distributed status dissemination network* (Master's thesis). Washington State University.
- Lamport, L., Shostak, R., & Pease, M. (1982). The Byzantine generals problem. *ACM Transactions on Programming Languages and Systems*, 4(3), 382–401. doi:10.1145/357172.357176
- Liberatore, V. (2006). Integrated playback, sensing and network control. *Proceedings of the 25th IEEE International Conference on Computer Communications*, 1–12.
- Melliopoulos, A. P. S. (2007). Substation automation: Are we there yet? *IEEE Power and Energy Magazine*, 5(3), 28–30. doi:10.1109/MPAE.2007.365815
- Merideth, M., Iyengar, A., Mikalsen, T., Tai, S., Rouvellou, I., & Narasimhan, P. (2005). Thema: Byzantine-fault-tolerant middleware for web services applications. *Proceedings of the IEEE Symposium on Reliable Distributed Systems*, 131–142. doi:10.1109/RELDIS.2005.28
- Rhea, S., Eaton, P., Geels, D., Weatherspoon, H., Zhao, B., & Kubiawicz, J. (2003). Pond: the oceanstore prototype. *Proceedings of the 2nd USENIX Conference on File and Storage Technologies* (Vol. 3, pp. 1–14).
- Wang, W., Xu, Y., & Khanna, M. (2011). A survey on the communication architectures in smart grid. *Computer Networks*, 55(15), 3604–3629. doi:10.1016/j.comnet.2011.07.010
- Zhang, H., Zhao, W., Moser, L. E., & Melliar-Smith, P. M. (2011). Design and implementation of a Byzantine fault tolerance framework for non-deterministic applications. *IET Software*, 5(3), 342–356. doi:10.1049/iet-sen.2010.0013
- Zhang, Y., Wang, L., & Xiang, Y. (2015). Power System Reliability Analysis With Intrusion Tolerance in SCADA Systems. *IEEE Transactions on Smart Grid*.
- Zhao, W. (2009). Design and implementation of a Byzantine fault tolerance framework for Web services. *Journal of Systems and Software*, 82(6), 1004–1015. doi:10.1016/j.jss.2008.12.037

ADDITIONAL READING

- Aiyer, A. S., Alvisi, L., Clement, A., Dahlin, M., Martin, J. P., & Porth, C. (2005). BAR fault tolerance for cooperative services. *Operating Systems Review*, 39(5), 45–58. doi:10.1145/1095809.1095816
- Amir, Y., Coan, B., Kirsch, J., & Lane, J. (2008, June). Byzantine replication under attack. In Dependable Systems and Networks With FTCS and DCC, 2008. DSN 2008. IEEE International Conference on (pp. 197-206). IEEE. doi:10.1109/DSN.2008.4630088
- Amir, Y., Danilov, C., Dolev, D., Kirsch, J., Lane, J., Nita-Rotaru, C., & Zage, D. et al. (2008). Steward: Scaling Byzantine Fault-Tolerant Replication to Wide Area Networks. *IEEE Transactions on Dependable and Secure Computing*, 80–93.
- Bessani, A. N., Alchieri, E. P., Correia, M., & Fraga, J. S. (2008, April). DepSpace: A Byzantine fault-tolerant coordination service. *Operating Systems Review*, 42(4), 163–176. doi:10.1145/1357010.1352610
- Casey, P. R., Jaber, N., & Tepe, K. E. (2011, October). Design and implementation of a cross-platform sensor network for smart grid transmission line monitoring. In *Proceedings of the IEEE International Conference on Smart Grid Communications* (pp. 285-290). IEEE. doi:10.1109/SmartGridComm.2011.6102334
- Castro, M., & Liskov, B. (2000, October). Proactive recovery in a Byzantine-fault-tolerant system. In *Proceedings of the 4th conference on Symposium on Operating System Design & Implementation-Volume 4* (pp. 19-19). USENIX Association.
- Castro, M., & Liskov, B. (2001, July). Byzantine fault tolerance can be fast. In *Dependable Systems and Networks*, 2001. DSN 2001. International Conference on (pp. 513-518). IEEE. doi:10.1109/DSN.2001.941437
- Chai, H., Melliar-Smith, P. M., Moser, L. E., Zhang, H., & Zhao, W. (2013). Toward Trustworthy Coordination of Web Services Business Activities. *IEEE Transactions on Services Computing*, 6(2), 276–288. doi:10.1109/TSC.2011.57
- Chun, B. G., Maniatis, P., & Shenker, S. (2008, June). Diverse Replication for Single-Machine Byzantine-Fault Tolerance. In *USENIX Annual Technical Conference* (pp. 287-292).
- Clement, A., Wong, E. L., Alvisi, L., Dahlin, M., & Marchetti, M. (2009, April). Vol. 9, pp. 153–168). Making Byzantine Fault Tolerant Systems Tolerate Byzantine Faults. In NSDI.
- Cowling, J., Myers, D., Liskov, B., Rodrigues, R., & Shriram, L. (2006, November). HQ replication: A hybrid quorum protocol for Byzantine fault tolerance. In *Proceedings of the 7th symposium on Operating systems design and implementation* (pp. 177-190). USENIX Association.
- Distler, T., & Kapitza, R. (2011, April). Increasing performance in byzantine fault-tolerant systems with on-demand replica consistency. In *Proceedings of the sixth conference on Computer systems* (pp. 91-106). ACM. doi:10.1145/1966445.1966455
- Driscoll, K., Hall, B., Sivencrona, H., & Zumsteg, P. (2003). Byzantine fault tolerance, from theory to reality. In *Computer Safety, Reliability, and Security* (pp. 235-248). Springer Berlin Heidelberg. doi:10.1007/978-3-540-39878-3_19
- Garcia, R., Rodrigues, R., & Preguiça, N. (2011, April). Efficient middleware for byzantine fault tolerant database replication. In *Proceedings of the sixth conference on Computer systems* (pp. 107-122). ACM. doi:10.1145/1966445.1966456
- Hendricks, J., Ganger, G. R., & Reiter, M. K. (2007). Low-overhead byzantine fault-tolerant storage. *Operating Systems Review*, 41(6), 73–86. doi:10.1145/1323293.1294269

Kapitza, R., Behl, J., Cachin, C., Distler, T., Kuhnle, S., Mohammadi, S. V., & Stengel, K. et al. (2012, April). CheapBFT: resource-efficient byzantine fault tolerance. In *Proceedings of the 7th ACM European conference on Computer Systems* (pp. 295-308). ACM.

Kotla, R., & Dahlin, M. (2004, June). High throughput Byzantine fault tolerance. In *Proceedings of the International Conference on Dependable Systems and Networks* (pp. 575-584). IEEE.

Veronese, G. S., Correia, M., Bessani, A. N., & Lung, L. C. (2009, September). Spin One's Wheels? Byzantine Fault Tolerance with a Spinning Primary. In *Proceedings of the 28th IEEE International Symposium on Reliable Distributed Systems* (pp. 135-144). IEEE. doi:10.1109/SRDS.2009.36

Yin, J., Martin, J. P., Venkataramani, A., Alvisi, L., & Dahlin, M. (2002, June). Byzantine fault-tolerant confidentiality. In *Proceedings of the International Workshop on Future Directions in Distributed Computing* (pp. 12-15).

Zhang, H., Chai, H., Zhao, W., Melliar-Smith, P. M., & Moser, L. E. (2012). Trustworthy coordination of Web services atomic transactions. *IEEE Transactions on Parallel and Distributed Systems*, 23(8), 1551-1565. doi:10.1109/TPDS.2011.292

Zhang, Y., Zheng, Z., & Lyu, M. R. (2011, July). BFTCloud: A byzantine fault tolerance framework for voluntary-resource cloud computing. In *Proceedings of the IEEE International Conference on Cloud Computing*, (pp. 444-451). IEEE. doi:10.1109/CLOUD.2011.16

Zhao, W. (2009). Design and implementation of a Byzantine fault tolerance framework for Web services. *Journal of Systems and Software*, 82(6), 1004-1015. doi:10.1016/j.jss.2008.12.037

Zhao, W. (2014a). *Building dependable distributed systems*. John Wiley & Sons. doi:10.1002/9781118912744

Zhao, W. (2014b). Application-Aware Byzantine Fault Tolerance. In *Proceedings of the 12th IEEE International Conference on Dependable, Autonomous and Secure Computing* (pp. 45-50). IEEE. doi:10.1109/DASC.2014.17

Zhao, W. (2015). Optimistic Byzantine fault tolerance. *International Journal of Parallel, Emergent and Distributed Systems*, 1-14 (preprint).

Zhao, W., & Zhang, H. (2009). Proactive service migration for long-running Byzantine fault-tolerant systems. *IET software*, 3(2), 154-164.

KEY TERMS AND DEFINITIONS

Byzantine Fault Tolerance: It refers to the capability of a system to tolerate Byzantine faults.

Intelligent Electronic Device (IED): It is an actuating device that is capable of receiving commands from a controller. Example IEDs include protective relaying devices, and voltage regulators.

Jitter: It refers to the deviation from the periodicity of a sequence of events or signals.

Normal Operation: It refers to the operation of an algorithm during a period that either there is no fault, or the faults do not disrupt its operation. For example, when a backup replica crashes, the PBFT algorithm would still operate as normal.

Phasor Measurement Unit (PMU): It is a device that measures the electrical waves in an electric power grid. The measurements must be synchronized with a global clock, such as a GPS.

Sampling Rate: It is defined as the number of samples taken per unit of time.

SCADA: It is short for Supervisory Control and Data Acquisition. It is a type of industrial control system that monitor and control industrial processes that exist physically.

View Change: It refers to the configuration change of the group of replicas that engage in Byzantine fault tolerance. When the primary is suspected of being faulty, a new view is initiated so that a different replica is elected as the primary for the new view.

E-Waste, Chemical Toxicity, and Legislation in India

Prashant Mehta

National Law University Jodhpur, India

INTRODUCTION

The industrial revolution was a period of dynamic change and dramatic innovation in the history of human society (Ayers, 1999). Across the world, societies are constantly reinventing to manage revolutionary changes that have radically transformed the lifestyle of people. Some of these changes are subtle and barely noticeable, while other changes are blatant and abrupt, like advances in Information and Communication Technology (ICT) and widespread use of Electrical and Electronics Equipment (EEE), which has made human civilization to grow in a more efficient manner.

Following economic liberalization in 1991, the Indian ICT industry has been one of the major drivers of economic progress both in terms of volume and applications. It has assumed the role of providing a forceful leverage to the socio, economic, and technological growth of a developing society (Joseph, 2007). However consumption and production processes of these complex electronic devices are unsustainable, pose a serious challenge to environment and human health, making e-waste one of the largest growing waste streams (Lundgren, 2012a). With waste market getting increasingly global, such waste is illegally exported to crude e-waste recycling hotspots in Asian countries, such as China, India, and Pakistan, and in some African countries, like Ghana and Nigeria (Castillo, 2011). Such illegal trade in e-waste is camouflaged and conducted under the pretext of obtaining 'reusable' equipment or 'donations' from developed nations.

E-waste comprises of ICT and EEE products that are not fit for their originally intended use. It includes computers, its accessories (monitors,

printers, keyboards, and central processing units), typewriters, mobile phones and chargers, remotes, compact discs, headphones, batteries, LCD/Plasma TVs, air conditioners, refrigerators, and other household appliances (Lalchandani, 2010). The increasing 'market penetration' in the developing countries, 'replacement market' in the developed countries (Borthakur & Sinha, 2013), coupled with rapid developments, innovation, miniaturization, and replacement resulted into higher rate of obsolescence of electronics products. It is estimated that in 2014 world-wide 41.8 million metric tonnes (Mt) e-waste was generated and most of it was not collected and treated in environmentally sound manner (Baldé, Wang, Kuehr, & Huisman, 2015). Most of this either end up with municipal waste in landfills or unauthorized recycling yard (Greenpeace Press Report, 2008).

As noted by UNEP in 2005, "Every year, 20 to 50 million tonnes of e-waste is generated world-wide, which could bring serious risks to human health and the environment" (Schwarzer, Giuliani, Kluser, & Peduzzi, 2005). Even though there is no clear data on the quantity of e-waste generated and disposed of each year in India, it is estimated 70 percent of e-waste handled in India is imported. It also estimates that between the years of 2007-2020, domestic television e-waste will double, computer e-waste is expected to increase five-fold, while cell phones will increase eighteen times (Disabled World, 2015). Thus knowledge society of 21st century is creating its own toxic footprint which is most debated issue amongst the environmentalists, environment regulators, world-wide environment forums, governmental, and non-governmental agencies, and policy makers.

DOI: 10.4018/978-1-5225-2255-3.ch268

BACKGROUND OF STUDY

Solid waste management, which is already a mammoth task in India, is becoming more complicated by the invasion of e-waste, which has complex characteristics as it differs chemically and physically from urban or industrial waste. Each wave of technology creates a set of waste previously unknown by humans (Sikdar & Vaniya, 2014) making e-waste management a big issue in both developed and developing countries.

The current practices of e-waste management in developing countries suffer from a many drawbacks such as informal recycling, inadequate legislation, low public awareness of the hazardous nature of e-waste, use of obsolete methods, and inadequate emphasis on the employee's protection (Cobbing, 2008), is jeopardizing people's health and environment (Smith, Sonnenfeld, & Naguib Pellow, 2006a).

Having reviewed literature from various other studies conducted in India and abroad, and understanding the magnitude of this problem, it is time for India to critically review its management of e-waste, to work towards a strategy to create the necessary infrastructure, and mechanisms to support sustainable and environmentally friendly e-waste management besides sensitizing consumers, waste recyclers, and future decision makers on issues like e-waste characteristics, its trans-boundary movement recycling technology, social, and environmental considerations, and toxic effect on health.

DEFINITION OF E-WASTE

Even though there is no standard definition for e-waste, some of the reported definitions of e-waste in literature are mentioned below:

According to the Basel Convention, "Wastes are substances or objects, which are disposed of or are intended to be disposed of, or are required to be disposed of by the provisions of national laws" (Text of Basel Convention, 2014).

According to Basel Action Network (BAN), "E-waste includes a wide and developing range of electronic appliances ranging from large household appliances, such as refrigerators, air-conditioners, cell phones, stereo systems, and consumable electronic items to computers discarded by their users" (Puckett, Byster, Westervelt, Gutierrez, Davis, Hussain, & Dutta, 2002), (Gaidajis, Angelakoglou, & Aktsoglou, 2010).

As per European Directive 2002/96/EC, "Waste electrical and electronic equipment (WEEE), including all components, subassemblies, and consumables, which are part of the product at the time of discarding" (European Parliament, 2003, January 27), (Jain, 2008), (European Parliament, 2012, July 4).

As per European Directive 75/442/EEC, Article I(a), "Any substance or object which the holder discards or is required to discard in compliance with the national legislative provisions". Further it includes all components, subassemblies, and consumables which are part of the product at the time of discarding (Borthakur & Singh, 2012).

According to Organisation for Economic Co-operation and Development (OECD), "Any household appliance consuming electricity and reaching its life cycle end", also referred as composite waste (OECD, 2007).

These differences in definitions, of what constitutes e-waste, have the potential to create disparities in both the quantification of e-waste generation and the identification of e-waste flows across nations. The lack of a precise definition of e-waste is one of the key issues that need to be addressed on an international level (Lundgren, 2012b).

COMPONENTS OF E-WASTE

E-waste is classified as hazardous waste (Tsydenova & Bengtsson, 2011a), and it imposes many challenges on the recycling industry (Smith, Sonnenfeld, & Naguib Pellow, 2006b). Modern electronics can contain up to 60 different elements;

many are valuable, some are hazardous (Third World Network, 1991) and some are both. Several rare elements are also used (Frazzoli, Orisakwe, Dragone, & Mantovani, 2010). The types and amounts of metals used in electronics products vary with evolution of technology. The most complex mix of hazardous substances is usually present in the printed wiring boards (PWBs) that contain valuable metals like copper, silver, gold, palladium, and platinum, brominated flame retardants used in connectors, cathode ray tubes and LCD contain heavy metals like lead and barium, switches and flat screens contain mercury, older capacitors and transformers contain poly chlorinated biphenyl's (PCB's), poly vinyl chloride (PVC) coated copper cables and casing, plastics from computer hardware that release highly toxic dioxins and furans (Sum, 1991). The fraction including iron, copper, aluminum, gold, and other metals in e-waste is over 60 percent, plastics account for about 30 percent, and the hazardous pollutants comprise only about 2.70 percent of waste (Widmer, Oswald-Krapf, Sinha-Khetriwal, Schnellmann, & Bo'ni, 2005) besides alloys that mostly decreases metal's recyclability.

TOXIC ELEMENTS IN E-WASTE

E-waste contains thousands of components made of deadly chemicals, heavy metals, flame retardants, and potentially hazardous substances whose main routes of human exposure are through inhalation, dust ingestion, dermal exposure, and oral intake. Metal toxicity causes breathing difficulties, respiratory irritation, coughing, choking, pneumonitis, tremors, neuropsychiatric problems, convulsions, coma and even death (Yu, Welford, & Hills, 2006 a). Some toxic chemicals found in e-waste are analyzed below.

1. Beryllium (Be) is used as copper-beryllium alloys in computer motherboards, relays, and connectors (Taylor, Ding, Ehler, Foreman, Kaszuba, & Sauer, 2003). Beryllium refining

produces fumes, dust, and oxides which are both acutely and chronically toxic to humans. If inhaled in large concentration, it causes acute lung disease, breathing discomfort, coughing, rapid heartbeat, and death in extreme cases. Its compounds are carcinogenic in nature (IARC, 1993b) and studies have shown that people can still develop beryllium disease (beryllicosis) after many years of last exposure.

2. Cadmium (Cd) is a toxic heavy metal found naturally in very low concentration (Salomons & forstner, 1984). It is used in switches, rechargeable (Ni-Cd) batteries, stabilizers, resistors, and corrosion-resistant alloys. It is released into the environment as powdered dust during crushing and milling of plastics, CRTs, and circuit boards. It is potentially a long-term cumulative poison associated with deficits in cognition, learning, behavior, and neuromotor skills in children, causes severe pain in the joints and spine (Itai-itai disease), affects kidneys and softens bones (osteomalacia and osteoporosis) in humans. There is evidence of the role of cadmium and beryllium in carcinogenicity (Strickland, & Kensler, 1995), (Pruss-Ustun & Corvalan, 2006).
3. Hexavalent Chromium (Cr-VI) is used to protect metal housings in a computer from corrosion. It is very reactive and soluble in water, making it more mobile in environment (Mukherjee, 1998). Its corrosive nature cause skin allergies (dermatitis), damage DNA, liver, kidneys, pulmonary congestion, edema, bronchial maladies including asthmatic bronchitis, and lung cancer (IARC, 1990a).
4. Lead (Pb) as lead oxide comes from breaking of CRT, lead powder is released while removing solder from microchips, and lead fumes comes from high temperature smelting processes exposing the workers (Schutz, Olsson, Jensen, Gerhardsson, Borjesson, Mattsson, & Skerfving, 2005). In unlined

landfills, lead would dissolve in leachate or mix with ground water leading to contamination. It is neurotoxin that exerts toxic effects on the central nervous system (organic affective syndrome), peripheral nervous systems (motor neuropathy), the hemopoietic system (anemia), the genitourinary system (capable of causing damage to all parts of nephron), and the male and female reproductive systems (Harrington, Aw, & Baker, 2003). It affects mental development in children with impaired cognitive function, behavioral disturbances, attention deficits, hyperactivity, conduct problems, and lower IQ.

5. Mercury (Hg) is used as lighting device which illuminates most flat panel monitors. Workers can inhale of mercury vapour or dust which released while breaking and burning of circuit boards and switches. It affects kidneys, immune system, damage to the genitourinary system (tubular dysfunction), central and peripheral nervous systems, reduced fertility, and impairs fetus growth. When inorganic mercury spreads out in the water, it is transformed into toxic methylated mercury by microbial activity, which bio-accumulates, biomagnifies in living organisms, and concentrates through the food chain, particularly by fish (Hu, & Speizer, 2001), (WHO, 1989).
6. Brominated Flame Retardants (BFRs) like Polybrominated biphenyl (PBB), Polybrominated diphenyl ether (PBDE), and Tetrabromobisphenol-A (TBBPA) are chemically persistent organic pollutants (POP) along with toxic antimony trioxide which is used as flame retardants in electronic devices. They release carcinogenic brominated dioxins and furans as gases during fire. PBDE used in transformers and capacitors is bioaccumulative, impair brain function, and can cause liver and malfunctioning of endocrine system (thyroid damage). TBBPA used in printed circuit boards contains bro-

mine that can leach into landfills. Dust on computer cabinets contains BFRs.

7. Polyvinyl Chloride (PVC) is found in circuit boards, cabinets, and insulation on cables. It is hazardous because contains upto 56 percent chlorine which are precursors to polychlorinated di-benzo-p-dioxins and di-benzo-furans (classified as POP under Stockholm Convention) during incineration along with large quantities of hydrogen chloride gas, which when inhaled may leads to cancer, respiratory problems, affect reproductive, and immune system.
8. Polycyclic Aromatic Hydrocarbons (PAH) is generated from e-waste recycling activities and have potential impacts on soil, vegetation, and human health include breathing difficulties, respiratory irritation, coughing, choking, pneumonitis, tremors, neuropsychiatric problems, convulsions, coma and even death (Yu, Welford, & Hills, 2006 b). Epidemiological studies in the past on occupational exposure to PAH, provides sufficient evidence of the role of PAH in the induction of skin and lung cancers (Stewart & Kliehues, 2003).
9. Cobalt (Co) is extensively used in integrated circuits, semi-conductors, magnetic recording media, thin metallic films, and rechargeable batteries. It is mainly absorbed from the pulmonary and the gastrointestinal tracts and cause allergic dermatitis, rhinitis, vomiting, thyroid damage, and impaired vision. Cobalt dust may cause an asthma-like disease with symptoms ranging from cough, shortness of breath, and dyspnea to decreased pulmonary function, nodular fibrosis, permanent disability, and death.

In addition to the hazardous materials e-waste also contain a large number of valuable precious metals like Gold (Au), Silver (Ag), Platinum (Pt), and Palladium (Pd) in concentrations 40 to 50 times richer than there naturally occurring

deposits. These precious metals are extracted using hydrometallurgical processes, where valuable metals first leached into acid or alkali solutions and then they are concentrated by using various methods like precipitation, cementation, solvent extraction etc.

E- WASTE DISPOSAL AND RECYCLING PRACTICES IN INDIA

E-wastes disposal is a big problem faced by many countries including India. It is estimated that, by 2020, India could see nearly 500 percent rise in the number of old computers being dumped (Schluepa, et.al., 2009). E-waste disposal process work in two ways:

1. By removing the hazardous items;
2. By separating recyclable materials.

The three main groups of substances that may be released during recycling and material recovery, and may pose significant human and environmental health risks are:

1. Original constituents of EEE like lead and mercury;
2. Added substance during recovery processes like cyanide; and
3. Hazardous byproducts formed by incineration of e-waste like PAH.

In India, e-waste collection, transportation, segregation, dismantling, recycling, and disposal is done by unorganized small enterprise that is difficult to regulate. They employ untrained labours (more so children and women) who work in poorly-ventilated or enclosed areas without appropriate equipment and technical expertise. The process includes manual disassembly, melting, acid extraction of metals from complex mixtures, and extruding plastics.

Further incineration of printed circuit boards for desoldering and removal of chips exposes

workers to fumes of metals, particularly those in solder (often lead and tin), and other hazardous substances that can be potentially released (Tsydenova, & Bengtsson, 2011b) which pollutes the surrounding air. Inhalation and dust ingestion impose a range of potential occupational hazards including silicosis (Lepawsky, & McNabb, 2010). It exposes them to dangerous slow-poisoning chemicals on a regular basis impacting their health. Computer wastes that are land-filled produces contaminated leachate which eventually pollutes the groundwater, whereas acids and sludge obtained from melting computer chips, if disposed on the ground causes acidification of soil and irreversible damage to environment.

LEGAL FRAMEWORK IN INDIA

The environmentally sound management of used EEE imports, recycling, and their disposal has become the subject of serious discussion, debate, and significant challenge among the government, organizations, environmental groups, and component manufacturers in India. The department related Parliamentary Standing Committee on Science & Technology, under Ministry of Environment & Forests (MoEF) which is also responsible for environmental legislation, in its 192nd report on the 'Functioning of the Central Pollution Control Board (CPCB), which plays important role in drafting guidelines, has concluded that e-waste is going to be a big problem in the future due to modern life style, increase in the living standards of people, and augmentation of economic growth.

The rules and regulations for waste control in India are primarily listed under the aegis of Environmental Protection Act 1986. Despite a wide range of environmental legislation in India there are no specific laws or guidelines for electronic or computer waste (Devi, Shobha, & Kamble 2004). Electronic waste is included under List-A and List-B of Schedule-3 of the Hazardous Wastes (Management & Handling Rules), 1989, as amended in 2000 and 2003. In 2007, separate guidelines

on e-waste management were implemented, but they were voluntary and had limited impact. These guidelines include details such as e-waste composition; recycle, re-use, and recovery potential of items of economic value, identification of possible hazardous contents in e-waste, treatment, disposal options, and the environmentally sound e-waste treatment technologies (Rajya Sabha Unstarred Question No. 1887, 2009). However these rules primarily dealt with industrial waste and lack elements to deal with complex nature of e-waste.

Subsequently in 2008, these rules were amended to include toxic content and made registration mandatory for recyclers. The provision of environmental protection is delegated among various states in India. Following Supreme Court directions (Writ Petition (Civil) No. 657, 1995), the states have notified a set of hazardous waste laws and built a number of hazardous waste disposal facilities in the last ten years. However, the Controller and Auditor General (CAG) report found that over 75 percent of state bodies were not implementing these laws (Writ Petition (Civil) No. 10, 1995), giving rise to sloppy enforcement of e-waste related legislation.

India is also a signatory to the Basal Convention (under UNEP) on the control of Transboundary Movement of hazardous wastes and their disposal but officially opposes enforcement of BAN Amendment (Basel Action Network, 2011). The regulations banning the importation of hazardous waste for disposal are weak and an imported consignment of electronic scrap still comes into the country, as they are not properly classified as plastic scrap or mixed waste. In 2014, India generated 17 lakh tonnes of e-waste increasing growing at rate of 4-5 percent per year. With such exponential growth, the Indian government finally woke up and responded by framing the E-waste (Management and Handling Rules), 2016 with the aim to make re-cycling of e-wastes environmentally friendly. To begin with, the rules put India along with a select club of nations like the United States and many European nations to have legislation to regulate and manage e-waste. These rules recognized the producer's liability

for recycling and reducing e-waste in the country. It will apply to every producer, consumer, and bulk consumer involved in manufacture, sale, purchase, and processing of electronic equipment or components. It also brought disposal of CFL lights under its preview.

While the rules seem impressive on paper, environmentalist argues that there is total oversight of the ground situation. These rules ignore the unorganized sectors where 90 percent of the e-waste is generated. Also there is lack of a safe e-waste recycling infrastructure in the organized sector where only a fraction of the e-waste (estimated 10 percent) finds its way to recyclers due to absence of an efficient take back scheme for consumers. The ministry is giving the producers of EEE a breathing period of one year to set up their collection centers (Kumar & Shah, 2014) and develop technical guidelines for the environmentally sound management of e-wastes. Thus computer, mobile handset, and consumer goods manufacturers, will be required to come up with e-waste collection centers or introduce 'take back' systems. Therefore over reliance on the capacities of the informal sector poses severe risks to the environment and human health.

STRATEGIC INTERVENTION

A smart e-waste management system have to access current e-waste situation, recognize that e-wastes is complex mixture of hazardous substances, reduce the generation of e-waste through smart manufacturing and maintenance, reuse till functioning of EEE, and finally recycle those components that cannot be repaired or refurbished. Recycling and reuse of materials are the next level of potential options to reduce e-waste (Ramachandra & Saira, 2004). Based on current situational analysis in India, following strategic intervention is proposed:

1. Extended Producers Responsibility be introduced that involves collection and disposal of e-waste in environmentally sound manner

- or else stringent financial penalties must be imposed by state governments.
2. Update the legal and institutional framework for e-waste management including effective enforcement of laws, regulations, and standards. Introduce strict liability clause in proposed rules for effectiveness.
 3. Defining responsibilities of prime stakeholders at the level of government, supply chain, consumers, and develop a comprehensive policy that address all issues ranging from production, trade to final disposal, including technology transfers for the recycling of electronic waste.
 4. Raise public awareness, advocate for e-waste management across all stakeholders through public-private-partnerships linkages.
 5. Introduce a concept of 'e-waste exchange' as an independent market instrument offering services for sale and purchase of e-waste.
 6. Tighten import norms and custom procedure at port of entry. Maintain statistical records of imported of EEE goods for further analysis including their final disposal.
 7. Create a facilitative environment for investment in e-waste handling and disposal infrastructure by creating modern e-waste recycling facility and provide tax incentives to make it more effective.
 8. Introduce advance recycling fees. Also setup and operationalise an e-waste fund to benefit those working within this industry.

FUTURE RESEARCH DIRECTIONS

Developing national approach to handle e-waste, strengthening regulatory environment, designing new methods to increase waste collection, integrated modelling concepts to build waste recycling capacities, and building awareness among people can be carried out on larger scale across India. Also study of impact on toxic heavy metals individually as well as their cumulative effect on ecosystem needs to be further explored. Further quantities of waste material that moves between countries,

waste flows within a country and between countries and hazardous substance emissions associated with manual recycling process, social, and its health impact on children and women can also be further investigated. A comparative study of various EEE product categories can made.

CONCLUSION

From the discussion above, it is aptly clear that India faces an enormous task of handling and disposing growing piles of e-waste and its impact on human health (more so of women and children) and natural environment has increased manifold. Therefore policy level interventions should include strong e-waste regulation, tight control on import and export of e-wastes, and facilitation in development of recycling infrastructure. Lack of strict enforcement of legislation is also worrisome situation.

It requires building of public awareness, establishment of institutional infrastructures (including e-waste collection, transportation, treatment, storage, recovery, and disposal) at national and regional levels. Furthermore product end-of-life management should be made a priority during design of new electronic products using green materials, innovation in product technology, life cycle analysis, public outreach, and so on. Beyond conservation of raw materials and energy, there are additional environmental benefits of recycling, such as reduced land disturbance, water use, air emissions, and waste generation which in turn improves efficiency and environment. In conclusion it is time for us to look deep and ahead today as tomorrow it will be too late to act.

REFERENCES

Ayers, S. (1999). *The cultural impact of computer technology*. Yale-New Haven Teachers Institute. Retrieved February 29, 2016 from <http://www.yale.edu/ynhti/curriculum/units/1999/7/99.07.07.x.html>

- Baldé, C. P., Wang, F., Kuehr, R., & Huisman, J. (2015). The global e-waste monitor - 2014. United Nations University, IAS-SCYCLE.
- Basel Action Network (BAN). (2011). *Toxic trade news: "Cochin Port a safe conduit for imported e-waste". "Most aspects of e-waste not regulated in U.S., Va."; "Research identifies U.S electronic wastes as likely source of toxic jewellery imports from China"; "178 countries agree to allow the ban on exports of toxic wastes to developing countries to become law"*. Retrieved March 1, 2016 from <http://www.ban.org/>
- Borthakur, A., & Singh, P. (2012). Electronic waste in India: Problem and Policies. *International Journal of Environment Science*, 3, 354–362.
- Borthakur, A., & Sinha, K. (2013). Generation of electronic waste in India: Current scenario, dilemmas and stakeholders. *African Journal of Environmental Science and Technology*, 7(9), 899–910. doi:10.5897/AJEST2013.1505
- Castillo, M. (2011, January 14). *Electronic Waste: Where does it go and what happens to it?* Retrieved March 1, 2016 from <http://techland.time.com/2011/01/14/electronic-waste-where-does-it-go-and-what-happens-to-it/>
- Cobbing, M. (2008). *Toxic Tech: Not in our backyard. Uncovering the hidden flows of e-waste*. Report from Greenpeace International, Amsterdam. Retrieved March 1, 2016 from <http://www.greenpeace.org/raw/content/belgium/fr/press/reports/toxic-tech.pdf>
- Devi, B. S., Shobha, S. V., & Kamble, R. K. (2004). E-Waste: The Hidden harm of Technological Revolution. *Journal IAEM*, 31, 196–205.
- Disabled World. (2015). *Electronic waste: Medical & health issues*. Retrieved March 3, 2016 from <http://www.disabled-world.com/health/ewaste.php>
- Disposal of e-waste. (2009). Rajya Sabha Unstarred Question No. 1887.
- European Parliament. (2003, January 27). Directive 2002/96/EC of the European Parliament and of the Council on waste electrical and electronic equipment (WEEE). *Official Journal of the European Union*, L, 37, 24–38.
- European Parliament. (2012, July 4). Directive 2012/19/EU of the European Parliament and of the Council on waste electrical and electronic equipment (WEEE). *Official Journal of the European Union*, L, 197, 38–71.
- Frazzoli, C., Orisakwe, O. E., Dragone, R., & Mantovani, A. (2010). Diagnostic health risk assessment of electronic waste on the general population in developing countries scenarios. *Environmental Impact Assessment Review*, 30(6), 388–399. doi:10.1016/j.eiar.2009.12.004
- Gaidajis, G., Angelakoglou, K., & Aktsoğlu, D. (2010). E-waste Environmental Problem and Current Management. *Journal of Engineering Science and Technology, Review*, 3(1), 193–199.
- Greenpeace. (2008). *Take Back Blues, An assessment of e-waste takes back in India*. Retrieved February 29, 2016 from www.greenpeace.org/india/press/reports/recycling-of-electronic-waste
- Guidance Manual for the Implementation of the Organisation for Economic Co-operation and Development (OECD). (2007). *Recommendation C(2004)100 on Environmentally Sound Management (ESM) of Waste, Section 5.2, 12*. Retrieved March 2, 2016 from <http://www.oecd.org/env/waste/39559085.pdf>
- Harrington, J. M., Aw, T. C., & Baker, E. L. (2003). Occupational and environmental health and safety. In A. W. David, M. C. Timothy, D. F. John, & J. B. Edward (Eds.), *Oxford Textbook of Medicine* (4th ed.; Vol. 1, pp. 956–960). New York: Oxford University Press.
- Hu, H., & Speizer, F. E. (2001). Specific environmental and occupational hazards. In Harrison's Principles of Internal Medicine (15th ed.; vol. 2, pp. 2591–2592). McGraw-Hill Inc.

- International Agency for Research on Cancer (IARC). (1990). Chromium. IACR Monograph, 49, 677.
- International Agency for Research on Cancer (IARC). (1993). Beryllium and Beryllium Compounds. International Agency for Research on Cancer (IARC) Monograph, 58, 444.
- Jain, A. (Ed.). (2008). *E-waste: Implications, regulations, and management in India and current global best practices*. New Delhi: Teri Press.
- Joseph, K. (2007). Electronic waste management in India - Issues and Strategies. *Proceedings of the Eleventh International Waste Management and Landfill Symposium*.
- Kumar, R., & Shah, J. D. (2014). Current Status of Recycling Waste Printed Circuit Boards in India. *Journal of Environmental Protection*, 9-16.
- Lalchandani, N. (2010, April 24). E-scare. *The Times of India*.
- Lepawsky, J., & McNabb, C. (2010). Mapping international flows of electronic waste. *Canadian Geographer*, 54(2), 177–195. doi:10.1111/j.1541-0064.2009.00279.x
- Lundgren, K. (2012a). The Global Impact of e-waste: Addressing the Challenge. International Labour Office, Programme on Safety and Health at Work and the Environment (Safe Work), Sectoral Activities Department (Sector). Geneva: ILO.
- Lundgren, K. (2012b). The Global Impact of e-waste: Addressing the Challenge. International Labour Office, Programme on Safety and Health at Work and the Environment (Safe Work), and Sectoral Activities Department (SECTOR). Geneva: ILO, ISBN 978-92-2-126897-0 (print)
- Mukherjee, A. B. (1998). Chromium in the environment of Finland. *The Science of the Total Environment*, 217(1-2), 9–19. doi:10.1016/S0048-9697(98)00163-6 PMID:9695169
- Pruss-Ustun, A., & Corvalan, C. (2006). *Preventing disease through healthy environments: Towards an estimate of environmental burden of disease*. WHO Publication.
- Puckett, J., Byster, L., Westervelt, S., Gutierrez, R., Davis, S., Hussain, A., & Dutta, M. (2002). *Exporting Harm - The High-Tech Trashing of Asia: The Basel Action Network (BAN) Silicon Valley Toxics Coalition*. Seattle, WA: SVTC.
- Ramachandra, T. V., & Saira, V. K. (2004, March). Environmentally sound options for waste management, *Envis Journal of Human Settlements*. Retrieved March 2, 2016 from <http://www.ces.iisc.ernet.in/energy/paper/ewaste/ewaste.html>
- Salomons, W., & Frostner, U. (1984). Metals in hydrocycle. Springer-Verlag. doi:10.1007/978-3-642-69325-0
- Schluepa, M., Hageluekenb, C., Kuehrc, R., Magalinic, F., Maurerc, C., Meskersb, C.,... Wang, F. (2009). *Recycling: From e-waste to resources, Sustainable Innovation and Technology Transfer Industrial Sector Studies (Nairobi and Bonn, UNEP and STeP)*. Retrieved March 2, 2016 from http://www.unep.org/pdf/Recycling_From_e-waste_to_resources.pdf
- Schutz, A., Olsson, M., Jensen, A., Gerhardsson, L., Borjesson, J., Mattsson, S., & Skerfving, S. (2005). Lead in finger bone, whole blood, plasma and urine in lead-smelter workers: Extended exposure range. *International Archives of Occupational and Environmental Health*, 78(1), 35–43. doi:10.1007/s00420-004-0559-5 PMID:15750821
- Schwarzer, S., Giuliani, A., De Bono, G., Kluser, S., & Peduzzi, P. (2005). “E-waste: the hidden side of IT equipment’s manufacturing and use”. *Early Warnings on Emerging Environmental Threats, No. 5, United Nations Environment Programme (UNEP)*. GRID Europe.

- Sikdar, M., & Vaniya, S. (2014). The New Millennium and Emerging Concerns. *International Journal of Scientific and Research Publications*, 4(2).
- Smith, T., Sonnenfeld, D. A., & Naguib Pellow, D. (Eds.). (2006). Challenging the chip: Labor rights and environmental justice in the global electronics industry. Philadelphia, PA: Temple University Press.
- Stewart, B. W., & Kliehues, P. (2003). *World Cancer Report*. Lyon: IARC Press.
- Strickland, P. T., & Kensler, T. W. (1995). Chemical and physical agents in our environment. In *Clinical Oncology*. Churchill Livingstone Inc.
- Sum, E. Y. L. (1991). The recovery of metals from electronic scrap. *Journal of Metallurgy*, 43, 53–61.
- Taylor, T. P., Ding, M., Ehler, D. D., Foreman, T. M., Kaszuba, J. P., & Sauer, N. N. (2003). Beryllium in the environment: A Review. *Journal of Environmental Science and Health. Part A, Toxic/Hazardous Substances & Environmental Engineering*, 38(2), 439–469. doi:10.1081/ESE-120016906 PMID:12638707
- Text of the Basel Convention. (2014). *The Control of Transboundary Movements of Hazardous Wastes and Their Disposal*. United Nations Environment Programme (UNEP). Retrieved March 1, 2016 from <http://www.basel.int/portals/4/basel%20convention/docs/text/baselconvention-text-e.pdf>
- Third World Network. (1991). Toxic Terror: Dumping of Hazardous Wastes in the Third World, Penang, Malaysia. Author.
- Tsydenova, O., & Bengtsson, M. (2011a). Chemical hazards associated with treatment of waste electrical and electronic equipment. *Waste Management (New York, N.Y.)*, 31(1), 45–58. doi:10.1016/j.wasman.2010.08.014 PMID:20869229
- Tsydenova, O., & Bengtsson, M. (2011b). Chemical hazards associated with treatment of waste electrical and electronic equipment. *Waste Management (New York, N.Y.)*, 31(1), 45–58. doi:10.1016/j.wasman.2010.08.014 PMID:20869229
- Widmer, R., Oswald-Krapf, H., Sinha-Khetriwal, D., Schnellmann, M., & Bo'ni, H. (2005). Global perspectives on e-waste. *Environmental Impact Assessment Review*, 25(5), 436–458. doi:10.1016/j.eiar.2005.04.001
- World Health Organization (WHO). (1989). Mercury, Environmental Health Criteria, 86. WHO.
- Yu, J., Welford, R., & Hills, P. (2006 a). Industry responses to EU WEEE and ROHS Directives: Perspectives from China. *Corporate Social Responsibility and Environmental Management*, 13(5), 286–299. doi:10.1002/csr.131
- Yu, J., Welford, R., & Hills, P. (2006 b). Industry responses to EU WEEE and ROHS Directives: Perspectives from China. *Corporate Social Responsibility and Environmental Management*, 13(5), 286–299. doi:10.1002/csr.131

ADDITIONAL READING

- Brigden, K., Labunska, I., Santillo, D., & Allsopp, M. (2005). *Recycling of electronic wastes in China & India: Workplace & environmental contamination*. Greenpeace International.
- Bruke, M. (2007). The Gadget Scrap, Chemistry World, 44–48. Retrieved April 1, 2016 from www.chemistryworld.org
- Fornalczyk, A., Willner, J., Francuz, K., & Cebulski, J. (2013). E-waste as a source of valuable metals. *Archives of Materials Science and Engineering*, 63(2), 87–92.

Gupta, V., & Kumar, A. (2014). E-waste status and management in India. *Journal of Information Engineering and Applications*, 4(9), 41–48.

Khetriwal, D. S., Luepschen, C., & Kuehr, R. (Eds.). (2013). *Solving the e-waste problem: An interdisciplinary compilation of international e-waste research*. United Nations University.

Krishna, R., & Saha, S. (2015). Study paper on e-waste management. Retrieved March 30, 2016 from http://tec.gov.in/pdf/Studypaper/e%20waste%20management_11.08.pdf

Namias, J. (2013). The future of electronic waste recycling in the United States: Obstacles and domestic solutions. Retrieved March 31, 2016 from http://www.seas.columbia.edu/earth/wtert/sofos/Namias_Thesis_07-08-13.pdf

KEY TERMS AND DEFINITIONS

Dioxins and Furan: These are persistent environmental pollutants (POPs), formed as unintentional by-product during e-waste incineration to recover valuable metals.

Electronic Waste (E-Waste): It refers to EEE waste, including all components, subassemblies, and consumables which are part of the product destined for reuse, resale, salvage, recycling or disposal at the time of discarding.

Environmental Hazard: A substance, state or event which has the potential to threaten the surrounding natural environment that adversely affects people's health (Pollution and Natural disasters).

Environmental Sustainability: It is defined as could be defined as a condition of balance, resilience, and interconnectedness that allows human society to satisfy its needs while neither exceeding the capacity of its supporting ecosystems to continue to regenerate the services necessary to meet those needs nor by our actions diminishing biological diversity.

Metal Poisoning: Toxic metals in certain form and dose sometimes imitate the action of an essential element in the body; interfere with the metabolic processes that cause illness (metal poisoning).

Occupational Hazard: It is risk accepted as a consequence of a particular occupation and they encompass chemical, biological, psychosocial, and physical hazards.

Recycling: It is the process of converting waste materials into reusable objects; dismantling, separating fractions, and recovering material in order to reduce the consumption of fresh raw materials, energy usage, air pollution (from incineration) and water pollution (from land filling) from e-waste after the lifespan of the equipment.

Green IT and the Struggle for a Widespread Adoption

E**Edward T. Chen***University of Massachusetts – Lowell, USA*

INTRODUCTION

Since the inception of computers, both for business and personal purposes, there have been multiple environmental issues that resulted from this technology. The complex electronics require significant electricity to operate them, large amounts of energy to keep them cool for prolonged usage, and various chemicals and resources to construct them. Notably, within the last decade, there has been a movement building for the ecologically responsible construction, use and disposal of computer systems and their components, including monitors, batteries and printer cartridges. This initiative is commonly known as Green Information Technology (IT), or Green IT.

As both consumable and enterprise level computing products grows, a need for sustainability arises. A balance between the energy consumption and the provided services is required to ensure the environment can survive the influx of billions and billions of devices. Concepts like the Internet of Things, Big Data, smart devices and phones, and complex business analytics for corporations all drive the need for more connected devices. These devices consume more electricity than ever before and data runs the planet (Murugesan & Gangadharan, 2012; Subburaj, Kulkarni, & Jia, 2014).

The Green IT (green information technology) is the practice of environmentally sustainable computing (McLaughlin, 2013). The lack of regulations, standardizations, and standard operating procedures has left this notion out of the mainstream and under the radar of many organizations' information technology (IT) implementations. Several ideas at different levels have

been proposed over the years. Its current adoption rate is not enough for sustainability. G-Readiness framework combines properties, processes, and components that are well defined and measurable to ensure success in the greening of IT (Molla, Cooper, Corbitt, Deng, Peszynski, Pittayachawan, & Teoh, 2008). Large technology companies have designed, patented, and implemented as a way to offer a differentiated service and a competitive advantage through green IT. Some of their innovations have the potential to be replicated for further successes (Murugesan & Gangadharan, 2012).

BACKGROUND

Though there is not a general consensus on the exact definition of Green IT (also referred to as green computing, green information and communication technologies (ICT), or ICT sustainability), the most commonly accepted definition was coined by San Murugesan, an outspoken university professor, in his 2008 article entitled "Harnessing Green IT: Principles and Practices". Murugesan defined green IT as "the study and practice of designing, manufacturing, using and disposing of computers, servers, and associated subsystems... efficiently and effectively with minimal or no impact on the environment" (Murugesan, 2008). Multiple efforts can be made, both from individual home users as well as those of entire businesses, to reduce the negative impact on the environment from the technology they are using.

The hardware, software, and components that make up technology are always changing and evolving. Some components like computer

processors, are gaining the ability to process information faster while the integrated circuits are getting smaller. Other devices gain new features with each new release and make the older model seem outdated or no longer usable. This perception is particularly accurate with personal technology such as laptops, phones, and tablets. Unused excess hardware accumulates in staggering quantities. In the corporate world, technology is advancing faster than the needs of many businesses. Data centers are filled with high-powered servers and storage devices, which run 24/7/365 in a production environment. Attractive and enticing price points combined with clever marketing presentations convince companies that the deployment of these systems is necessary to solve their IT and IS (information systems) problems (Nguyen, Cheriet, Lemay, Reijs, Mackarel, & Pastrama, 2012).

According to Gartner Research, there are 2 billion computers in use today. They predict that the number of devices and things, items such as thermostats, refrigerators, cars, and other non-traditional computing hardware and sensors, on the Internet could surpass 40 billion by the year 2020 (Akhgar, Pattinson, & Dastbaz, 2015). Greenpeace estimates that if the Internet were a country, it would fall between Japan and Russia, or 5th place, in overall electricity consumption in the world (Cook & Pomerantz, 2015). 50% of the world's population owns a cellular telephone. This number is only going to go up as emerging countries begin to rely on the same technology as First World countries. Tablets are expected to outpace computers in sales and use before the end of this year (Akhgar, Pattinson, & Dastbaz, 2015). The amount of technology in use and the amount of technology that has been cast aside present two challenges for the concept of green IT: reducing energy consumptions of current hardware and finding ways to safely recycle previous hardware that is no longer in use. Stated in a different way, it is solving the two problems of how to reduce CO₂ emissions and how to lower e-waste (Ahmad & Ranka, 2016; Elliot, 2007).

A study was conducted in 2009 to investigate why the lack of growth with implementing and supporting green IT initiatives and standards. It surveyed Chief Information Officers (CIOs) and other IS professionals to find the "barriers" that keep green IT from being implemented. The results of the survey show no surprises, citing a lack of business leadership; the unknown costs versus cost savings for green IT solutions, and the absence of value by turning to green alternatives (Dedrick, 2010). Also uncovered through the same survey was the importance of government incentives or regulations mandating converting to green IT. Without formal direction or instructions to do so, companies are not eager to start the perceived arduous process to switch.

INFORMATION TECHNOLOGY'S GREEN PROBLEM

Energy consumption is a major aspect of IT, with the methods to produce electricity still largely powered by the depletion of fossil fuels such as coal and oil. It is estimated that the carbon dioxide (CO₂) produced by a single desktop computer over its lifetime is 1,096 kilograms (Thomson & van Belle, 2015). This consumption and pollution is amplified by the increasing occurrence of non-efficient software and coding, requiring computers to take longer periods to process finite tasks. A 2009 disputed study found that the average search using the popular search engine Google produced approximately 7 grams of CO₂ and required roughly half of the amount of energy needed to boil a kettle of water (Swaine, 2009; Warman, 2009). A portion of the total electricity consumed by IT is to power the computers and their components, while an even larger portion (30%) is used to cool the computers and their related hardware, particularly in the data centers associated with these information systems (IS) that house the computers and their related components, such as servers and power supplies (Murugesan, 2008; Nguyen, 2012).

The majority of businesses routinely require that their data centers remain cooled at all times at temperatures less than 70 degrees Fahrenheit (21 degrees Celsius) and one study found that only 7% of data centers in the world run at or above a temperature of 75 degrees Fahrenheit (23.8 degrees Celsius) (Mitchell, 2013). The majority of energy spent to cool these facilities utilizes air cooling as opposed to other more efficient methods of cooling and require continuous operation to maintain the set temperature, regardless of whether the IT is even in use. It is estimated that many processors sit idle between 85 to 95% of the day, requiring nearly the same amount of electricity as when active (Shah, 2012). Energy consumption utilized by IT is continually increasing as IT becomes more prevalent in society, with electricity usage by data centers in the United States rising 74% between 2000 and 2010 (van Bussel, Smitter, & Vandepas, 2015).

Computer production requires extensive resources and chemicals, with their associated production facilities consuming large amounts of energy for daily operations as well as climate control. Materials such as lead, mercury, cadmium and hexavalent chromium, combined with large amounts of water, aluminum and plastic, are required to make these machines and their intricate components. With today's pressing need for IT, this technology is more necessary than ever before to ensure each company and household remains capable of performing all manners of tasks and communications both within their community and throughout the world. Despite these high production levels of state-of-the-art equipment, many networks built to last 33 years are seeing themselves disposed of after only 3 to 5 years of use, mainly due to a knee-jerk reaction by their owners to consistently replace and update their 'obsolete' systems for the supposedly required next generations of technology (Ahmad & Ranka, 2016; Ogden, 2013).

These discarded systems are then largely disposed of inappropriately due to minimal disposal regulations and recycling options for IT, resulting

in approximately 80% of the systems being dumped in nearby landfills or exported to developing countries such as China and Pakistan where the disposal regulations are less stringent. In 2008, it was estimated that two-thirds of the estimated 870 million personal computers that will be made within the next five years will end up discarded in landfills (Murugesan, 2008). In 2006, the global production of IT-related waste was estimated at 20 to 50 million tons per year (van Bussel et al., 2015). These inefficient practices regarding IT are contributing towards global warming, the expedited depletion of natural resources, and increasing the waste in landfills, leading to the pollution of both the land as well as water sources in the vicinity.

SOLUTIONS AND RECOMMENDATIONS

Despite this dire situation, many options exist to implement green IT and make computing less of a hazard on the environment. The first practice the majority of companies and households can perform is power management, ranging from simply turning off their computers when not in use to providing a power management features to the IT to automatically reduce its electrical load following a period of inactivity. This feature, known as sleep mode, can reduce costs by 60 to 70% and now comes standard on most new computer systems after lengthy pressure was exerted on the manufacturing industry to standardize power-saving features in their products (Murugesan, 2008). Another prevalent shift in the IT industry is the development of ecological hardware that requires much lower levels of electricity to operate as well as attempts to increase the lifespan of the equipment. An example of this innovative new technology is the Atom series of energy-efficient processors designed by Intel. They are ultra-thin and lightweight and their use in IT reduces the space and energy required on the IS. In addition, the fit-PC by Compulab is a series of fanless per-

sonal computers that are smaller and require less power than traditional IT, often able to function for long periods of time on batteries and requiring lower-power processors such as the Intel Atom.

Monitors that display the IT images have improved significantly in technology and availability, greatly reducing energy costs. Light-emitting diode (LED) displays are now very comparable in cost and beneficial due to their use of significantly less electricity than liquid crystal display (LCD) monitors as well as their predecessor, the cathode ray tube (CRT) display. In 2014, the German ion research company GSI Helmholtz had created the world's most energy-efficient supercomputer, the L-CSC, in Frankfurt as evidenced by the 'Green 500' report released in November of 2014. Despite being incredibly energy-efficient, the computer was also rated the fourth fastest computer in Germany upon its release (Phys.org, 2014), demonstrating that green computers can be just as powerful and effective as standard models. In addition, efficient coding of software and other applications can significantly reduce the time required for computing processes by increasing the software processing speed on each computer, reducing the electricity consumption as well as energy required to cool the IS. The increasing development and availability of green IT hardware and software are helping provide efficient alternatives to traditional IT equipment used throughout the past 30 years.

There is a wide variety of cooling system methods that can be used to maintain a requisite temperature in the spaces that house computer systems. In addition to traditional air conditioning, alternative methods such as air-side economizers for facilities and liquid cooling systems for PC's are showing themselves to be viable options for keeping equipment cool and, in many cases, requiring a fraction of the cost of air conditioning while being several hundred times more efficient. Despite these benefits, it is understood that a massive overhaul of a company's cooling systems, as well as the required investment that would accompany it, are not simple fixes that can be implemented

overnight. Despite these hurdles, the primary solution to high cooling costs that is immediate and is being recommended by leading green companies (such as FedEx, Raytheon and Northrop Grumman) as well as the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) is to simply raise the temperature in data centers. Rather than require a temperature of no warmer than 70 degrees Fahrenheit (F), multiple experts state that all companies should be safely able to raise the temperature of its data centers to 75 degrees F and possibly even as high as 80.6 degrees F to maintain the climate required to protect the IS while simultaneously reducing the energy demands of the facility cooling systems (Mitchell, 2013).

Finally, the most effective solution to reducing energy, hardware, and software costs for all entities utilizing IT seems to be virtualizing equipment. Virtualization is the practice of consolidating servers, desktops, and any other types of equipment to fewer pieces of equipment that can still handle the load requirements. Server virtualization is common in which many virtual servers are hosted on a smaller number of more powerful servers. Instead of having 20 servers throughout the country, entities can invest in three larger server facilities that can still provide full network coverage throughout the company while significantly reducing facility operating, cooling and maintenance costs (Pandi & Somasundaram, 2016). Northrop Grumman is an example of this type of virtualization, eliminating 4,000 physical servers and combining 19 data centers and 81 smaller server rooms into only three facilities (Mitchell, 2013).

Desktop virtualization works in many ways by the same principles. Thin clients are stateless and fan-less computer desktop terminals with no hard drives. They simply link in to the data center to access all desktop capabilities while only drawing one-fifth of the electricity of a traditional desktop computer (Murugesan, 2008; Nguyen, 2012). Even virtual chillers are becoming a frequently-utilized concept, providing climate controls to multiple

facilities within the area at a fraction of the cost each facility would require to cool through individual systems. These concentrated efforts may take considerable planning and expense to arrange in the majority of circumstances, but the initiatives pay off for years in the future with exponentially lower energy and facilities costs to the companies (Pandi & Somasundaram, 2016).

LIMITATIONS

Despite the obvious benefits and wide array of alternatives for green IT that are available to companies now more than ever before, the main deterrent is that there are very minimal laws of compliance and regulation regarding green IT. Instead of strict legislation requiring specific changes to be implemented, the majority of initiatives and alternative options available within the United States and beyond are solely advisory. The United States Environmental Protection Agency (EPA) founded the Energy Star program in 1992, giving it a significant upgrade in 2006 to include computers and other IT, but it is still mainly an optional program and not a requirement for all appliances and equipment to have the green star logo depicting environment-conscious operation. As of 2008, only 26 of the 50 United States had established a statewide recycling program for the proper disposal of older computers. The Green Electronics Council created the Electronic Product Environmental Assessment Tool (EPEAT) as a method to assist prospective IT buyers through the ranking of available for purchase equipment by more than 50 criteria topics, assigning scores based on how green the technology is (Ahmad & Ranka, 2016).

In 2007, President George W. Bush signed Executive Order number 13423, which required all federal agencies to utilize EPEAT when purchasing new computer systems and requiring all vendors contracted by the Federal government to utilize EPEAT in their system purchases. President Obama recently modified this requirement in March of 2015 with Executive Order 13693, stating

that all government procurements of electronics should “meet or exceed specifications, standards, or labels recommended by (the) EPA” (Moodie, 2015). This legislation is expected to reduce the prominence of the EPEAT ratings system by instead requiring companies wishing to do business with the federal government to follow one of the many recommended ratings systems, rather than specifically the most well-known, EPEAT, but at least the directive to procure green equipment remains a priority.

The Restriction of Hazardous Substances (ROHS) in Electrical and Electronic Equipment Directive makes significant restrictions on the European Union market for equipment containing specific amounts of certain hazardous substances (Murugesan, 2008). Since August of 2006, Greenpeace International has produced their quarterly Guide to Greener Electronics, ranking several companies on their overall use (or lack thereof) of green IT. Then in 2007, several manufacturers of computer equipment (including Dell, HP, Microsoft, and Intel) formed the Green Grid, a consortium dedicated to making data centers and information systems more ecologically efficient.

Four general drivers for a company to choose to utilize green IT are economic, regulatory, market opportunity, and finally influence from social, cultural and political pressures (Thomson & van Belle, 2015). Despite the wide variety of laws and regulations aimed to transform companies into becoming more environmentally-conscious, the majority remain overwhelmingly optional for companies to follow. Thus, the primary motivator for companies to go green is economic, strongly appealing primarily to their own self-interests (Bohas & Poussing, 2016). Implementing green methods reduces energy costs. Green practices can extend the life of the IT and decreases the costs of disposal of obsolete equipment. In addition, following green practices will set a positive example for other companies to follow and can greatly increase the public image and reputation of the firm, known as “Green corporate image” (Bohas & Poussing, 2016).

However, many businesses look at green IT through a purely economic mindset. Despite the benefits to the business and the environment, fiscal incentives or government subsidies will have to be involved to aid the businesses in the costly transition towards greener IT (Mitchell, 2013). Through the use of several of the advisory organizations listed above, there is a growing voice for green practices with increasing pressures being placed on both the companies as well as the governments of the world to enact stricter regulations to enable change. There is also a small but growing market opportunity for companies that specialize in green IT consulting to help companies shift to cost effective and ecological green practices. This will likely help increase the prevalence of green computing in the years to come.

Another major benefit of green IT is that for the majority of the process, managerial practices should not have to change significantly in the business settings. The process of shifting systems to greener alternatives will require close coordination and decision-making between the management of each business and the IS personnel involved. Following these shifts, the systems will largely function in a way that does not change the day-to-day aspects of each normal business process. The employees will likely just see a greater emphasis on turning off their computers at the end of the day and other similar green initiatives that are both understandable and significant practices that businesses can implement to help their environment.

Other than the higher costs that are still regularly associated with newer technology, it does not appear to be any subsequent technological problems or complications that will arise as a result of green IT. Alternatively, green methods are designed and implemented specifically to reduce the existing energy problems that have been prevalent for decades. They are consistently being seen as reliable solutions to the disadvantages of outdated technology.

FUTURE RESEARCH DIRECTIONS

The Internet of Things (IoT) promises a lot of innovation through sensors operating and interoperating together. It has already solved problems that previously having no operational solutions. IoT brings with it immense computing capabilities and endless possibilities that combine science fiction with an improved quality of life. The need for a seeming endless amount of storage to keep the big data collected from over 40 billion IoT devices. This alone facilitates the immediate need for green IT. However, this technology and this level of capability can also drive green IT initiatives.

As Big Data becomes a term that continues to grow in usage, the amount of data is also growing. The more data that needs to be stored, the more natural resources it will take to power the servers, keep them cool, and house them in large facilities. One innovative team decided to see if solving green IT can be accomplished by reducing the abundant amount of data that is currently burning up more energy than necessary. While this study still needs further development, it helps to further the cause here of helping to define how to make green IT more accessible to the general IT public. It categorizes green IT into six core components. The first component is longevity of the hardware. How well was the product produced? What is the expected life span? The second component is software optimization, allowing for the hardware to consume less resource to compile and run the software. Third one is power management in finding the right formula and standard to shut off computer resources when not in use. The fourth component is the recycling of materials and not just recycling computers but finding alternative uses for the hardware. Fifth component on the list is telecommuting. An indirect benefit from fewer cars on the road will reduce carbon dioxide emissions. And finally, the sixth component is energy-efficient computing or low-power IT (van Bussel, Smitter, & van de Pas, 2015).

Current and previous research has been focused mainly on trying to better understand how green IT can be codified for implementation. Longitudinal studies have looked at ways that Green IT could be standardized, means to quantify form over function, and even look at the individuals who will need to support it and their impressions from a psychological approach. If green IT is to surpass its current adoption rate, stronger research needs to be conducted by pairing technology with green philosophies. For example, implementing power-saving functionality at the firmware level of a network switch can save energy for the entire network. Sensors abound from the Internet of Things revolution should make collected data meaningful for experiments and publish new findings and standards easier to reach target audiences (Kaushik & Vidyarthi, 2016).

CONCLUSION

Through a series of missteps or misinterpretations, green IT is still struggling to find its place within the technology field. Studies have been conducted; outcomes have been published; and successes have been documented; but something remains missing in this vital area that blends innovations with ecological awareness and respect. The G-readiness framework is probably the most concrete way for a common business to assess the potential of green IT for them based on the factors most important to keeping the business operational. And while the framework might seem daunting at first, there is a structure to it that forces a common sense and logical approach to reduce energy and look at IT in a smarter way.

Companies have adopted the virtualization of servers. Virtualization is not considered a win for green IT, nor did it help to mainstream the concept. Instead, IT professionals everywhere have no idea that when they learn how to run multiple services virtually, they reduce their carbon footprints. Those same IT professionals will be needed to collaborate with their business

counterparts to ensure they can provide innovative solutions to problems as they arise. This will be the cornerstone to the implementation of green IT within a business.

The green IT community is still small and many of the same people have been working together on the different studies and experiments trying to garner a stronger and larger following. Social media should be exploited to ensure the next generation of IT professionals and operators know how to think green right from the beginning. It's the Millennials that will be faced with the green IT challenge, a challenge to sustain a healthy planet through the Internet of Things and the 50 billion devices and zettabytes worth of data that comes with it. Green IT must become the new normal before that time. It is no longer an option, and it is no longer something for someone else to worry about.

Currently, there are not enough green IT regulations implemented through the governments of the world, including that of the United States. They would require at least a small percentage of its implementation by the business sector to kick start the unanimous shift towards fully green computing. Further, until commercial entities are required to comply with green IT practices, the personal-use market will also not attain compliance and yet another decade could likely pass resulting in very minimal improvements. Until the requirements are enacted, the majority of businesses will simply not switch over. Before they are able to devote the requisite time and energy to make this essential shift, they want to see that green IT practices can benefit them significantly.

The concept of Green IT is steadily increasing in prominence and the benefits such as reduced energy costs are making green IT practices very attractive to all businesses. They desire a sense of longevity and a cost-effective IS infrastructure. The omnipresence and convenience of IT in all venues of life factored in with the environmental benefits from computing. The reduced greenhouse gas emissions evidenced from telecommuting, online education enterprises and video telecon-

ferencing, will keep the IT industry strong and a permanent fixture in today's society. Green IT will eventually become the standard to follow, but the implementation will be much slower if major elements do not become mandatory through the use of effective legislation and increasing numbers of entities willing to set the correct industry examples.

REFERENCES

- Ahmad, I., & Ranka, S. (2016). *Handbook of energy-aware and green computing*. CRC Press.
- Akhgar, B., Pattinson, C., & Dastbaz, M. (2015). *Green Information Technology: A Sustainable Approach*. Waltham, MA: Elsevier, Inc.
- Bohas, A., & Poussing, N. (2016). An empirical exploration of the role of strategic and responsive corporate social responsibility in the adoption of different green IT strategies. *Journal of Cleaner Production*, 122, 240–251. doi:10.1016/j.jclepro.2016.02.029
- Cook, G., & Pomerantz, D. (2015). *Clicking Clean: A Guide to Building the Green Internet*. Washington, DC: Greenpeace, Inc.
- Dedrick, J. (2010). Green IS: Concepts and issues for information systems research. *Communications of the Association for Information Systems*, 27(11), 173–183.
- Elliot, S. (2007). Environmentally sustainable ICT: a critical topic for IS research? *Pacific Asia Conference on Information Systems*, 114, 100–112.
- Kaushik, A., & Vidyarthi, D. P. (2016). A green energy model for resource allocation in computational grid using dynamic threshold and GA. *Sustainable Computing: Informatics and Systems*, 9, 42–56.
- McLaughlin, E. (2013). *Definition of green IT (green information technology)*. Retrieved March 30, 2016 from <http://searchcio.techtarget.com/definition/green-IT-green-information-technology>
- Mitchell, R. (2013). IT gets its green on. *Computerworld*, 26–29.
- Molla, A. (2009). Organizational motivations for Green IT: exploring Green IT matrix and motivation models. *Pacific Asia Conference on Information Systems*.
- Molla, A., Cooper, V., Corbitt, B., Deng, H., Peszynski, K., Pittayachawan, S., & Teoh, S. Y. (2008). E-readiness to G-readiness: developing a green information technology readiness framework. *Australasian Conference on Information Systems*, 35, 669–678.
- Moodie, A. (2015). *Is the Obama Administration lowering the bar on green electronics?* Retrieved March 20, 2016 from <http://www.theguardian.com/sustainable-business/2015/apr/23/obama-repeat-epa-technology-green-electronics-council>
- Murugesan, S. (2008). Harnessing Green IT: Principles and practices. *IT Professional*, 10(1), 24–33. doi:10.1109/MITP.2008.10
- Murugesan, S., & Gangadharan, G. (2012). *Harnessing green IT: principles and practices*. Chichester, UK: John Wiley and Sons. doi:10.1002/9781118305393
- Nguyen, K. K., Cheriet, M., Lemay, M., Reijts, V., Mackarel, A., & Pastrama, A. (2012). Environmental-aware virtual data center network. *Computer Networks*, 56(10), 2538–2550. doi:10.1016/j.comnet.2012.03.008
- Ogden, J. (2013). Blog: Green IT – a key to long-term savings. *CIO Asia*. Retrieved June 20, 2016 from <http://www.cio-asia.com/tech/data-center/blog-green-it--a-key-for-long-term-savings/?page=1>
- Pandi, K. M., & Somasundaram, K. (2016). Energy efficient in virtual infrastructure and green cloud computing: A review. *Indian Journal of Science & Technology*, 9(11), 1–8. doi:10.17485/ijst/2016/v9i11/89399

Phys.org. (2014). *German Supercomputer is a world champion in saving energy*. Retrieved March 10, 2016 from <http://phys.org/news/2014-11-german-supercomputer-world-champion-energy.html>

Shah, R. (2012). Eco-friendly IT: Greener approach to IT. *Journal of Management and Business Research*, 2(2), 4–25.

Subburaj, S., Kulkarni, S., & Jia, L. (2014). Green IT: Sustainability by aligning business requirements with IT resource utilization. *International Journal of Communication Networks and Distributed Systems*, 12(1), 30–46. doi:10.1504/IJCND.2014.057986

Swaine, J. (2009). *Two Google Searches 'Produce Same CO₂ as Boiling a Kettle*. Retrieved March 30, 2016 from <http://www.telegraph.co.uk/technology/google/4217055/Two-Google-searches-produce-same-CO2-as-boiling-a-kettle.html>

Thomson, S., & van Belle, J. (2015). Antecedents of Green IT adoption in South African higher education institutions. *The Electronic Journal Information Systems Evaluation*, 18(2), 172–186.

Van Bussel, G., Smitter, N., & Vandepas, J. (2015). Digital archiving, Green IT and environment. Deleting data to manage critical effects of the data deluge. *The Electronic Journal Information Systems Evaluation*, 18(2), 187–198.

Warman, M. (2009). *Green I.T.: How many Google searches does it take to boil a kettle?* Retrieved March 30, 2016 from <http://www.telegraph.co.uk/technology/google/4241791/Green-I.T.-how-many-Google-searches-does-it-take-to-boil-a-kettle.html>

Bohas, A., & Poussing, N. (2016). An Empirical Exploration of the Role of Strategic and Responsive Corporate Social Responsibility in the Adoption of Different Green IT Strategies. *Journal of Cleaner Production*, 122, 240–251. doi:10.1016/j.jclepro.2016.02.029

Kaushik, A., & Vidyarthi, D. P. (2016). A green energy model for resource allocation in computational grid using dynamic threshold and GA. *Sustainable Computing: Informatics and Systems*, 9, 42–56.

Pandi, K. M., & Somasundaram, K. (2016). Energy efficient in virtual infrastructure and green cloud computing: A review. *Indian Journal of Science & Technology*, 9(11), 1–8. doi:10.17485/ijst/2016/v9i11/89399

KEY TERMS AND DEFINITIONS

Carbon Footprint: The total amount of greenhouse gas emissions usually expressed in tons of carbon dioxide (CO₂).

Data Center: A physical or virtual infrastructure used to house a large group of networked computer servers for the remote storage, processing, or distribution of large amounts of data.

Energy Consumption: Amount of energy consumed in a process or system, or by an individual, organization or country.

e-Waste: Any old, end-of-life or discarded electronic or electrical devices or their components.

Green IT: Green information technology is the practice of environmentally sustainable and responsible use of computers and related resources.

ADDITIONAL READING

Ahmad, I., & Ranka, S. (2016). *Handbook of energy-aware and green computing*. Chapman & Hall/CRC Computer and Information Science Series. CRC Press.

Identification of Green Procurement Drivers and Their Interrelationship Using Fuzzy TISM and MICMAC Analysis

Surajit Bag

Tega Industries South Africa Pty Ltd., South Africa

INTRODUCTION

Green procurement is sourcing products and services that cause minimal adverse environmental impacts. It incorporates human health, social and environmental concerns into the search for high quality products and services at competitive prices. Green procurement is generally considered a mammoth task by the procurement managers. Recently focus has been given by regulatory bodies to apply pressure on firms for implementing green programs. Environment protection bodies are regularly organizing seminars and conferences to educate and train managers in such greening initiatives. In some countries the government has developed green specifications for items and mandatory part of tender requirement for public procurement. However, the green procurement programs are still under nascent phase in most developing countries.

The present research is motivated based on the study of Azevedo et al., (2011) where they have pointed potential future research area in exploring the enablers and barriers influencing companies in taking green procurement decisions. Secondly, Appolloni et al., (2014) conducted a review on green procurement considering the time frame between 1996 and 2013 but does not highlight the inter-relationships between the drivers of green procurement practices and they have also kept it under one of future research directions. They have also mentioned the need for strong qualitative and quantitative research to support the progress of green procurement.

The objective of the current study is to identify the leading drivers that influence green procurement programs and determine the interactions among the identified drivers. This chapter is structured into four additional sections. The next section presents the background of the study which helps to identify the green procurement drivers. The third section introduces Fuzzy TISM. Finally, conclusions, limitations and directions of future research are presented.

BACKGROUND

In this section an attempt has been taken to briefly explain the key drivers of green procurement.

Government Policy and Regulations

Governments are among the largest consumers in an economy. The public sector on average spends 45%-65% of their budgets on procurement. Given this substantial purchasing power, governments have enormous leverage to stimulate and drive markets for sustainable production and consumption when they make a determined effort to purchase 'green' products and services. Adopting such an approach is a smart form of procuring goods and service – it not only improves the efficiency of public procurement but also uses the public market power to bring about significant environmental and socioeconomic benefits. Supply chain management operates within a regulatory framework set by National Government and

extended by provinces and local governments to specific policies, legislation and regulations. In South Africa for instance important legislation influencing this function includes the Public Finance Management Act (1999), Preferential Procurement Policy Framework Act (2000), Preferential Procurement Framework Regulations (2001) and National Treasury Regulations (2005). The Municipal Finance Management Act (MFMA) of 2003 governs the financial and supply chain management functions of Local Government. In developing green procurement policies, local government would need to ensure that these policies: are aligned with their existing Supply chain management regulatory frameworks; avoid a clash between the Preferential Procurement regulations and environmental principles or criteria in the policy; incorporate green procurement in all dimensions of the supply chain management cycle; and institutionalize green procurement within the existing structures set out by the regulatory framework. Government policy and regulations positively influences green procurement (Min & Galle 1997; Diabat & Govindan 2011; Hassini et al., 2012; Bag., 2014; Appolloni et al., 2014)

Total Quality Environmental Management

Firms with successful TQEM programs will have more formal mechanisms for interacting with suppliers. Business units with successful TQEM programs exhibit a greater degree of competitive focus and strategic sourcing. In every step of the manufacturing process there will be quality check to avoid rejections and wastage and this will assist in saving natural resources. TQEM positively influences green procurement (Khidir et al., 2010; Diabat & Govindan 2011; Ageron et al., 2012; Dubey et al., 2013; Bag., 2014; Bag & Anand., 2014; Dubey et al., 2014).

Management Support

Management support is important in success of any projects and specially for green procurement

programs where strategic decisions are mainly involved. Management support have positively influenced green procurement (Min & Galle 1997; Zhu et al., 2008; Arslan 2010; Bag., 2014).

Management Review

Management review periodically is necessary to check the progress of green procurement programs and see that timelines are met. The review will capture the bottlenecks, critical paths and develop strategies to find out ways to complete the activities at economical cost within the timeline. As per expert opinion management review positively influence green procurement.

Continuous Education of Employees

Organizations practicing green procurement must have the transformation and diversity manager to carry out the necessary trainings of employees. Training will assist employees in gaining knowledge and deeper understanding of green procurement and its importance in supply chain management. Moreover, the training budget must be utilized carefully in proper training and must be aligned with the company mission and vision so that organization ultimately benefits in the long run. As per expert opinion continuous education of employees positively influences green procurement.

Cross Functional Team Building

In a manufacturing firm there are people from planning, procurement, production, quality assurance, logistics and other functions. Since green procurement involves close coordination with all related supply chain functions therefore it is imperative that organization form a green procurement committee comprising people from all functions, i.e. a cross-functional team to drive the green procurement project. This committee will be responsible for generating weekly progress reports, communicating to internal and external environment, maintain records of consumption

of resources, reduction in energy usage and associated cost savings. As per expert opinion cross functional team building positively influences green procurement.

Organization Culture

Empowering junior managers and creating innovation culture among employees will facilitate green procurement practices. Ultimately the organisational culture, structure and process must support the green procurement practices. As per expert opinion organizational culture positively influences green procurement.

Green Process and Technology

Introducing 'green' tendering criteria can influence the marketplace and result in new entrants in the field of environmental technologies and products - potentially resulting in increased competition and reduced prices. Firms must remove the obsolete machinery and convert to green process and technology (renewable source of energy, reducing utility costs, environmental innovation and source reduction) to drive the green procurement programs. Green process and technology positively influences green procurement (Green et al., 2000; Hassini et al., 2012; Bag., 2014).

Information System Infrastructure

IT Eco-Efficiency and IT Eco-Innovation is important dimension of information system infrastructure required to drive organization wide green programs. Additionally, SAP, ERP, RFID, and data analytics are important tools to derive the MIS reports for assessing the progress of green procurement projects. IT infrastructure positively influence green procurement (Green et al., 1998; Hervani et al., 2005; Bag 2014).

Green Design

Driving the green procurement project starts in the design stage to ensure that products comply

with restrictions on specified chemical substances in parts and materials, while complying with obligations for labeling, information provision and energy-saving standards for finished products. The green design should comply with environmental norms and regulations such as WEEE, RoHS, REACH, EU Directive on Packaging and Packaging Waste. Green Design positively influence green procurement (Mavi et al., 2013).

Re-Use, Re-Engineering, and Recycling of Products and Materials

The 3 Rs' forms the basis of any green procurement programs and enhance cost savings and savings of natural resources. 3Rs' positively influence green procurement (Mavi et al., 2013).

Supplier Flexibility

It is critical for suppliers to adopt flexibility and supply alternate material as per revised bill of material in green procurement projects. The supply risk must be minimized by supplying the alternate material in time at economical costs. Supplier flexibility positively influence green procurement (Mavi et al., 2013; Bag, 2016b)

Suppliers' Capability to Innovate

It is imperative that supplier has the capability of innovation by demonstrating knowledge and willingness by coming with new eco-products at cheaper costs. It positively influences green procurement (Min & Galle., 2001; Chiou et al., 2011; Ageron et al., 2012).

Supply Risk Management

Green procurement adopt some of the best practices which automatically minimizes the supply risks associated with traditional supply chain management. It positively influences green procurement (Ageron et al., 2012; Bag 2014; Bag & Anand., 2014; Dubey et al., 2014).

Trust Building in Suppliers

Trust is a soft factor associated with green procurement. It positively influences green procurement (Ageron et al., 2012; Ji et al., 2014).

Low Supplier Lead Time

Green procurement prerequisites best operational practices such as vendor managed inventory, just in time approach and lean manufacturing to optimize costs. Automatically the supplier lead time becomes low in such cases and suppliers are able to deliver material as per delivery schedule. This involves proper strategy building and careful monitoring. Therefore, supplier lead time influences green procurement (Ageron et al., 2012; Bag, 2014; Bag & Anand., 2014; Dubey et al., 2014).

Customer Satisfaction

Customer satisfaction is the outcome of good green procurement practices and measured as a success parameter in any green programs. It positively influences green procurement (Diabat & Govindan, 2011; Ageron et al., 2012; Bag., 2014; Bag & Anand., 2014; Dubey et al., 2014).

Annual Savings From Green Procurement Practices

Organizations derive both tangible and intangible benefits from green procurement practices. The direct savings from green procurement practices are significant and highly motivates the procurement managers. As per expert opinion annual savings from green purchasing positively influences green procurement.

Annual Saving of Natural Resources

The annual savings of natural resources motivates managers in driving green procurement projects. The key areas where reduction can be achieved are: consumption of coal used to fire boiler, diesel oil in forklifts, steam for running hydraulic

systems and compressor, energy for utility and paper for printing will lead to significant savings of natural resources. As per expert opinion savings of natural resources positively influences green procurement.

Procurement Excellence

Adopting best world class sustainability practices such as green procurement leads to procurement excellence. Therefore, procurement excellence is the outcome and considered as a variable for measuring success of green procurement (Bag, 2016a).

DATA ANALYSIS

The current study intends to develop green procurement theory based on fuzzy total interpretive structural modeling (Fuzzy TISM) approach. Fuzzy TISM is an advanced method and designed in a manner to capture both the statements of respondents as well as logic and interpretation. TISM has greater explanatory power than other established inductive approaches.

Fuzzy Total Interpretive Structural Modeling (FUZZY TISM)

The steps of Fuzzy TISM have been followed as per Khatwani et al., (2015). The data analysis is segregated into following sub-sections:

Start the Decision Making Process

In the current study, the responses of five experts from manufacturing sector have been gathered. The input has been utilized to refine the drivers and further used in modeling.

Selection of Criteria and Sub Criteria of Green Procurement

This is the second step in Fuzzy TISM modeling approach. For the purpose of this study, twenty-

five drivers were identified from literature which influences green procurement practices and finally refined through experts' opinion. The final twenty sub criteria are presented in the previous section.

In the next step the process of gathering responses for fuzzy TISM has been explained in details.

Gathering Responses and Calculation of Aggregated SSIM

For the purpose of study, the responses of five procurement managers from manufacturing sector has been collected for evaluation of interrelationship among selected drivers. These experts are well experienced with the green procurement practices. A joint meeting was organized with these five experts and the research problem and methodology was presented and explained to them in 30 minutes' duration. Further they were requested to fill up the VAXO matrix. The entire process of filling up the VAXO matrix separately by these five experts took almost one hour. From the submitted five separate VAXO matrices, finally the aggregated SSIM matrix was calculated by applying mode. The aggregated SSIM matrix is shown in Table 1 (Appendix).

Calculation of Fuzzy Reachability Matrix

The fuzzy reachability matrix is generated from aggregated fuzzy SSIM matrix which was presented in the above stage. The fuzzy reachability matrix is shown in Table 2 (Appendix).

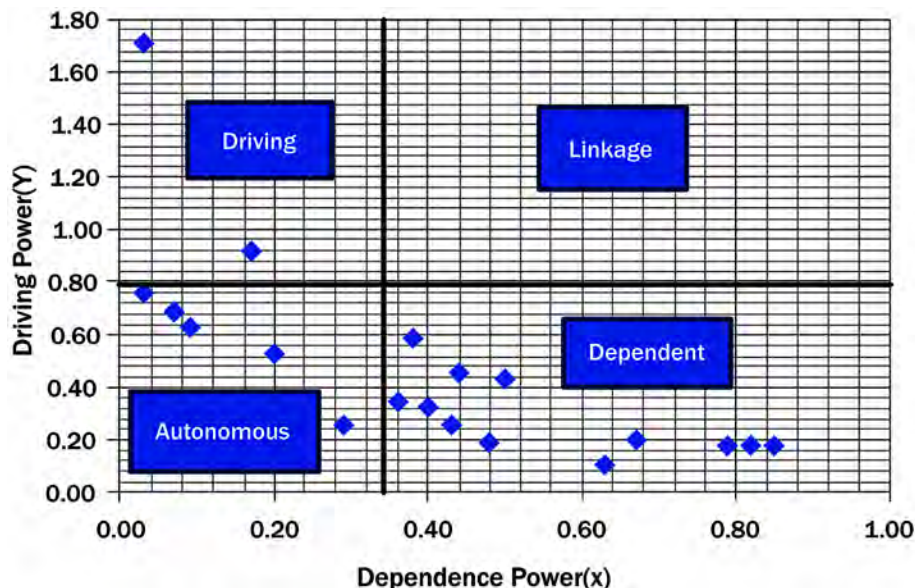
Calculation of Final Fuzzy Reachability Matrix

The final fuzzy reachability matrix is derived from step above and presented in Table 3 (Appendix). From the final fuzzy reachability matrix, the dependence power (X) and driving power (Y) is calculated based on summing of columns and rows and calculating the crisp value. The crisp value of each element is used as an input to develop fuzzy MICMAC analysis.

Driving Power and Dependence Power Matrix (MICMAC) Based on Fuzzy Reachability Matrix From Table 3

See Figure 1.

Figure 1. MICMAC analysis based on crisp values



Discussions Based on MICMAC Analysis Based on Fuzzy Reachability Matrix

Cluster 1 - Autonomous Variables: These variables have a weak drive power and weak dependence power. In this cluster we have six variables. Total Quality Environmental management (E2), Management Support (E3), Management Review (E4), Cross functional team building (E6), Organization Culture (E7) and Information System Infrastructure (E9).

Cluster 2 - Dependence Variables: These variables have a weak drive power but strong dependence power. In this cluster we have twelve variables. Green process and Technology (E8), Green Design (E10), Re-use, Re-engineering and Recycling of products and materials (E11), Supplier Flexibility (E12), Suppliers' capability to Innovate (E13), Supply risk management (E14), Trust building in suppliers (E15), Low supplier lead time (E16), Customer Satisfaction (E17), Annual Savings from green procurement practices (E18), Annual Saving of natural resources (E19) and Procurement Excellence (E20).

Cluster 3 - Linkage Variables: These variables have a strong drive power as well as strong dependence power. Linkage variables are very sensitive and unstable. Any action on these variables will trigger an effect on other variables and also a feedback on themselves. In this cluster we have no variables.

Cluster 4 - Driving Variables: These variables have a strong drive power but weak dependence power. In this cluster we have two variables. Government policy and Regulations (E1) and Continuous education of employees (E5).

Defuzzified Reachability Matrix

The Defuzzified reachability matrix with fuzzy linguistic terms VH, H, as 1 and rest as 0 is shown in Table 4 (Appendix).

Defuzzified Reachability Matrix and Transitivity Check

The paired comparisons are translated in the form of reachability Matrix. The matrix has been also checked for transitivity rule. A final reachability matrix, post transitivity check is presented in Table 5 (Appendix).

Driving Power and Dependence Power Matrix (MICMAC) Based on Defuzzified Reachability Matrix from Table 5

See Figure 2.

Discussion on MICMAC Analysis Based on Defuzzified Reachability Matrix

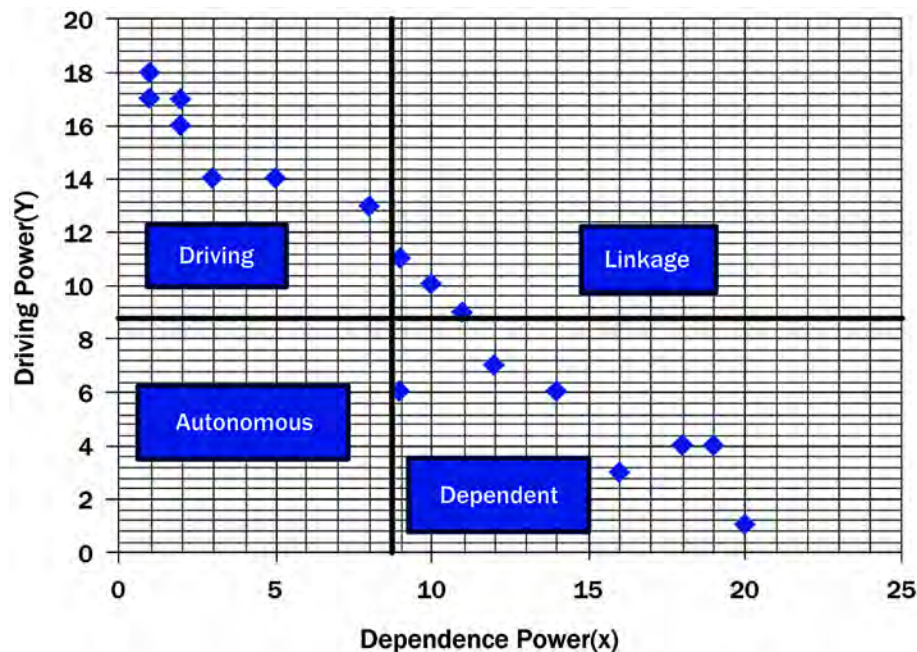
Cluster 1 - Autonomous Variables: These variables have a weak drive power and weak dependence power. In this cluster we have no variables.

Cluster 2 - Dependent Variables: These variables have a weak drive power but strong dependence power. In this cluster we have nine variables. Information System Infrastructure (E9), Suppliers' capability to Innovate (E13), Supply risk management (E14), Trust building in suppliers (E15), Low supplier lead time (E16), Customer Satisfaction (E17), Annual Savings from green procurement practices (E18), Annual Saving of natural resources (E19), Procurement Excellence (E20).

Cluster 3 - Linkage Variables: These variables have a strong drive power as well as strong dependence power. Linkage variables are very sensitive and unstable. Any action on these variables will trigger an effect on other variables and also a feedback on themselves. In this cluster we have three variables. Green Design (E10), Re-use, Re-engineering and Recycling of products and materials (E11) and Supplier Flexibility (E12).

Cluster 4 - Driving Variables: These variables have a strong drive power but weak de-

Figure 2. MICMAC analysis based on Table 6



pendence power. In this cluster we have eight variables. Government policy and Regulations (E1), Total Quality Environmental management (E2), Management Support (E3), Management Review (E4), Continuous education of employees (E5), Cross functional team building (E6), Organization Culture (E7), and Green process and Technology (E8).

Level Partition on Reachability Matrix

The final reachability matrix obtained in Table 5 is now partitioned into different levels. After the first iteration, the driver classified to level 1 is discarded and the partitioning procedure is repeated on the remaining drivers to determine the level 2. These iterations are continued until the level of each driver has been determined and presented in Table 6 (Appendix).

TISM Diagram

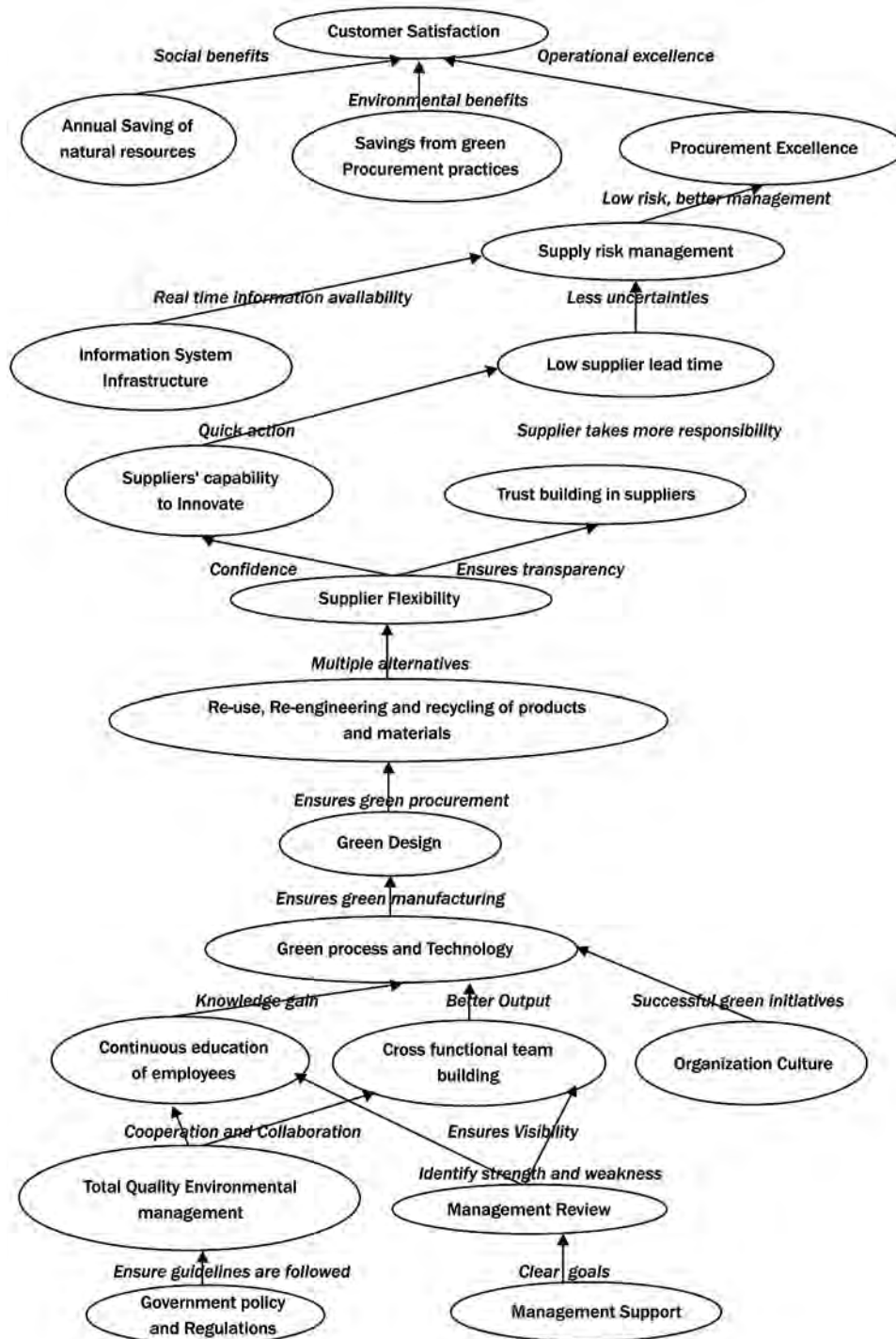
The connective and interpretive information contained in the interpretive direct interaction

matrix and diagraph is used to derive the TISM. The nodes in the diagraph are replaced by the interpretation of elements placed in boxes. The interpretation of the cells of interpretive direct interaction matrix is depicted by the side of the respective links in the structural model. The final TISM model after removing the transitive links is presented in Figure 3.

CONCLUSION

In the current study author proposes Fuzzy TISM method for identifying the interrelationships among elements influencing green procurement. Due to incorporation of fuzziness in TISM the decision makers have the flexibility in assigning the level of influence of pair wise elements. Apart from wider flexibility Fuzzy TISM also enhance the quality of decision making. The findings from the TISM model show that Customer Satisfaction is in level 1, Annual Savings from green procurement practices, Annual Saving of natural resources and Procurement Excellence are in level 2, Supply risk management is in level 3, Information System Infrastructure and Low supplier lead time is

Figure 3. TISM diagraph



in level 4, Suppliers' capability to Innovate and Trust building in suppliers is in level 5, Supplier Flexibility is in level 6, Re-use, Re-engineering and Recycling of products and materials is in level 7, Green Design is in level 8, Green process and Technology is in level 9, Continuous education of employees, Cross functional team building and Organization Culture is in level 10, Total Quality Environmental management and Management Review is in level 11 and Government policy and Regulations and Management Support are in the bottom level.

The driving factors which emerged from Fuzzy TISM and MICMAC analyses are Government policy and Regulations (E1), Total Quality Environmental management (E2), Management Support (E3), Management Review (E4), Continuous education of employees (E5), Cross functional team building (E6), Organization Culture (E7), and Green process and Technology (E8). These elements are the key players which must be considered while planning green procurement programs.

LIMITATIONS AND DIRECTIONS OF FUTURE RESEARCH

The present research has certain limitations. The model is developed purely based on interview with five procurement management experts from the manufacturing sector. In future studies the author proposes to statistically validate the model using big sample size and further extend theory of green procurement.

REFERENCES

Ageron, B., Gunasekaran, A., & Spalanzani, A. (2012). Sustainable supply management: An empirical study. *International Journal of Production Economics*, 140(1), 168–182. doi:10.1016/j.ijpe.2011.04.007

Appolloni, A., Sun, H., Jia, F., & Li, X. (2014). Green Procurement in the private sector: A state of the art review between 1996 and 2013. *Journal of Cleaner Production*, 85, 122–133. doi:10.1016/j.jclepro.2014.08.106

Arslan, T., Yilmaz, V., & Aksoy, H. K. (2010). Structural Equation Model for Environmentally Conscious Purchasing Behavior. *International Journal of Environmental of Research*, 6(1), 323–334.

Azevedo, S. G., Carvalho, H., & Machado, V. C. (2011). The influence of green practices on supply chain performance: A case study approach. *Transportation Research Part E, Logistics and Transportation Review*, 47(6), 850–871. doi:10.1016/j.tre.2011.05.017

Bag, S. (2014). Developing a GSCM model for the Indian rubber goods manufacturing sector (Unpublished Doctoral Dissertation). University of Petroleum & Energy Studies. Available at <http://ils.ddn.upes.ac.in:8001/cgi-bin/koha/opac-authoritiesdetail.pl?authid=13645>

Bag, S. (2016). Green strategy, supplier relationship building and supply chain performance: Total interpretive structural modelling approach. *International Journal of Procurement Management*, 9(4), 398–426. doi:10.1504/IJPM.2016.077702

Bag, S. (2016b). Flexible procurement systems is key to supply chain sustainability. *Journal of Transport and Supply Chain Management*, 10(1), 1–9. doi:10.4102/jtscm.v10i1.213

Bag, S., & Anand, N. (2014). Modeling Green Supply Chain Management framework using ISM and MICMAC analysis. *African Journal of Business Management*, 8(22), 1053.

Chiou, T. Y., Chan, H. K., Lettice, F., & Chung, S. H. (2011). The influence of greening the suppliers and green innovation on environmental performance and competitive advantage in Taiwan. *Transportation Research Part E, Logistics and Transportation Review*, 47(6), 822–836. doi:10.1016/j.tre.2011.05.016

- Diabat, A., & Govindan, K. (2011). An analysis of the drivers affecting the implementation of green supply chain management. *Resources, Conservation and Recycling*, 55(6), 659–667. doi:10.1016/j.resconrec.2010.12.002
- Dubey, R., & Ali, S. (2014). Identification of flexible manufacturing system dimensions and their interrelationship using total interpretive structural modelling and fuzzy MICMAC analysis. *Global Journal of Flexible Systems Management*, 15(2), 131–143. doi:10.1007/s40171-014-0058-9
- Dubey, R., Bag, S., Ali, S., & Venkatesh, V. (2013). Green purchasing is key to superior performance: An empirical study. *International Journal of Procurement Management*, 6(2), 187–210. doi:10.1504/IJPM.2013.052469
- Green, K., Morton, B., & New, S. (1998). Green purchasing and supply policies: Do they improve companies environmental performance? *Supply Chain Management: An International Journal*, 3(2), 89–95. doi:10.1108/13598549810215405
- Green, K., Morton, B., & New, S. (2000). Greening organizations purchasing, consumption, and innovation. *Organization & Environment*, 13(2), 206–225. doi:10.1177/1086026600132003
- Hassini, E., Surti, C., & Searcy, C. (2012). A literature review and a case study of sustainable supply chains with a focus on metrics. *International Journal of Production Economics*, 140(1), 69–82. doi:10.1016/j.ijpe.2012.01.042
- Hervani, A. A., Helms, M. M., & Sarkis, J. (2005). Performance measurement for green supply chain management. *Benchmarking, International Journal (Toronto, Ont.)*, 12(4), 330–353.
- Ji, P., Ma, X., & Li, G. (2014). Developing green purchasing relationships for the manufacturing industry: An evolutionary game theory perspective. *International Journal of Production Economics*.
- Khatwani, G., Singh, S. P., Trivedi, A., & Chauhan, A. (2015). Fuzzy-TISM: A fuzzy extension of TISM for group decision making. *Global Journal of Flexible Systems Management*, 16(1), 97–112. doi:10.1007/s40171-014-0087-4
- Khidir ElTayeb, T., Zailani, S., & Jayaraman, K. (2010). The examination on the drivers for green purchasing adoption among EMS 14001 certified companies in Malaysia. *Journal of Manufacturing Technology Management*, 21(2), 206–225. doi:10.1108/17410381011014378
- Mavi, R. K., Kazemi, S., Najafabadi, A. F., & Mousaabad, H. B. (2013). Identification and assessment of logistical factors to evaluate a green supplier using the fuzzy logic DEMATEL method. *Polish Journal of Environmental Studies*, 22(2), 445–455.
- Min, H., & Galle, W. P. (1997). Green purchasing strategies: Trends and implications. *International Journal of Purchasing and Materials Management*, 33(2), 10–17. doi:10.1111/j.1745-493X.1997.tb00026.x
- Min, H., & Galle, W. P. (2001). Green purchasing practices of US firms. *International Journal of Operations & Production Management*, 21(9), 1222–1238. doi:10.1108/EUM0000000005923
- Zhu, Q., Sarkis, J., Cordeiro, J. J., & Lai, K. H. (2008). Firm-level correlates of emergent green supply chain management practices in the Chinese context. *Omega*, 36(4), 577–591. doi:10.1016/j.omega.2006.11.009

ADDITIONAL READING

- Berry, C. (2011). *The sustainable procurement guide: procuring sustainably using BS 8903*. BSi.
- Theron, C., & Dowden, M. (2014). *Strategic Sustainable Procurement: Law and Best Practice for the Public and Private Sectors*. Do Sustainability.

KEY TERMS AND DEFINITIONS

Green Procurement: Green Procurement is procuring of products and services that cause minimal adverse environmental impacts. It incorporates human health, social and environmental concerns into the search for high quality products and services at competitive prices.

Total Interpretive Structural Modeling (TISM): Fuzzy TISM is an extension of Interpre-

tive structural modeling (ISM) technique. ISM methodology, transforms unclear, poorly articulated models of systems into clear, well-defined models. ISM uses experts to judge the variables, and the relations among the variables are interpreted. ISM depends on the experts' knowledge and familiarity with the firm, its operations, and its industry. ISM generates deep knowledge of the subject and is greatly helpful for practitioners.

APPENDIX



Table 1. Aggregated SSIM matrix

	E 20	E 19	E 18	E 17	E 16	E 15	E 14	E 13	E 12	E 11	E 10	E 9	E 8	E 7	E 6	E 5	E 4	E 3	E 2	E 1
E 1	O(No)	V(VH)	V(VH)	O(No)	O(No)	O(No)	O(No)	O(No)	O(No)	V(VH)	V(VH)	V(L)	V(VH)	O(No)	O(No)	O(No)	O(No)	O(No)	V(VH)	
E 2	V(VH)	V(VH)	V(VH)	V(VH)	V(L)	V(H)	V(VH)	V(H)	V(H)	V(VH)	V(VH)	V(H)	V(VH)	O(No)	V(VH)	O(No)	O(No)	O(No)		
E 3	V(VH)	V(VH)	V(VH)	V(H)	V(L)	V(H)	V(H)	V(H)	V(H)	V(VH)	V(VH)	V(VH)	V(VH)	V(H)	V(VH)	V(VH)	V(H)			
E 4	V(VH)	V(VH)	V(VH)	V(H)	V(H)	V(L)	V(H)	V(L)	V(L)	V(H)	V(H)	V(L)	V(H)	V(L)	V(H)	V(H)				
E 5	V(VH)	V(H)	V(H)	V(L)	O(No)	O(No)	O(No)	O(No)	O(No)	V(H)	V(H)	O(No)	V(H)	O(No)	V(L)					
E 6	V(VH)	V(VH)	V(VH)	V(L)	V(L)	V(L)	V(H)	V(L)	V(H)	V(VH)	V(VH)	V(L)	V(VH)	O(No)						
E 7	V(VH)	V(VH)	V(VH)	V(VH)	V(L)	V(VH)	V(VH)	V(VH)	V(VH)	V(VH)	V(VH)	V(VH)	V(VH)							
E 8	V(VH)	V(VH)	V(VH)	V(VH)	V(VH)	V(VH)	V(VH)	V(H)	V(H)	V(VH)	V(VH)	V(H)								
E 9	V(H)	V(L)	V(L)	V(L)	V(L)	O(No)	V(H)	O(No)	V(H)	V(L)	V(L)									
E 10	V(H)	V(VH)	V(VH)	V(VH)	V(H)	V(H)	V(H)	V(H)	V(H)	V(VH)										
E 11	V(VH)	V(VH)	V(VH)	V(VH)	V(H)	V(H)	V(H)	V(H)	V(H)											
E 12	V(H)	V(H)	V(L)	O(No)	V(VH)	V(H)	V(VH)	V(VH)												
E 13	V(VH)	V(VH)	V(VH)	V(H)	V(H)	V(L)	V(VH)													
E 14	V(VH)	O(No)	O(No)	V(H)	A(VH)															
E 15	V(VH)	V(H)	V(H)	O(No)	V(VH)															
E 16	V(H)	V(L)	V(L)	V(L)																
E 17	A(VH)	A(H)	A(H)																	
E 18	X(VH)																			
E 19	X(VH)																			
E 20																				

Source: Author own compilation

Identification of Green Procurement Drivers and Their Interrelationship

Table 2. Fuzzy reachability matrix

	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12	E13	E14	E15	E16	E17	E18	E19	E20
E 1	1	VH	No	No	No	No	No	VH	L	VH	VH	No	No	No	No	No	No	VH	VH	No
E 2	No	1	No	No	H	VH	No	VH	H	VH	VH	H	H	VH	H	L	VH	VH	VH	VH
E 3	No	No	1	H	VH	VH	H	VH	VH	VH	VH	H	H	H	H	L	H	VH	VH	VH
E 4	No	No	No	1	H	H	L	H	L	H	H	L	L	H	L	H	H	VH	VH	VH
E 5	No	No	No	No	1	L	No	H	No	H	H	No	No	No	No	No	L	H	H	VH
E 6	No	No	No	No	No	1	No	VH	L	VH	VH	H	L	H	L	L	L	VH	VH	VH
E 7	No	No	No	No	No	No	1	VH	VH	VH	VH	VH	VH	VH	VH	L	VH	VH	VH	VH
E 8	No	No	No	No	No	No	No	1	H	VH	VH	H	H	VH	VH	VH	VH	VH	VH	VH
E 9	No	No	No	No	No	No	No	No	1	L	L	No	No	H	No	L	L	L	L	H
E 10	No	No	No	No	No	No	No	No	No	1	VH	H	H	H	H	H	VH	VH	VH	H
E 11	No	No	No	No	No	No	No	No	No	No	1	H	H	H	H	H	VH	VH	VH	VH
E 12	No	No	No	No	No	No	No	No	No	No	No	1	VH	VH	H	VH	No	L	H	H
E 13	No	No	No	No	No	No	No	No	No	No	No	No	1	VH	L	H	H	VH	VH	VH
E 14	No	No	No	No	No	No	No	No	No	No	No	No	No	1	No	No	H	No	No	VH
E 15	No	No	No	No	No	No	No	No	No	No	No	No	No	VH	1	VH	No	H	H	VH
E 16	No	No	No	No	No	No	No	No	No	No	No	No	No	VH	No	1	L	L	L	H
E 17	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	1	No	No	No
E 18	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	H	1	VH	VH
E 19	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	H	VH	1	VH
E 20	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	VH	VH	VH	1

Source: Author own compilation



Table 3a. Final fuzzy reachability matrix

	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11
E 1	(1,1,1)	(0.75,1,0,1,0)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0.75,1,0,1,0)	(0.25,0.5,0.75)	(0.75,1,0,1,0)	(0.75,1,0,1,0)
E 2	(0,0,0.25)	(1,1,1)	(0,0,0.25)	(0,0,0.25)	(0.5,0.75,1,0)	(0.75,1,0,1,0)	(0,0,0.25)	(0.75,1,0,1,0)	(0.5,0.75,1,0)	(0.75,1,0,1,0)	(0.75,1,0,1,0)
E 3	(0,0,0.25)	(0,0,0.25)	(1,1,1)	(0.5,0.75,1,0)	(0.75,1,0,1,0)	(0.75,1,0,1,0)	(0.5,0.75,1,0)	(0.75,1,0,1,0)	(0.75,1,0,1,0)	(0.75,1,0,1,0)	(0.75,1,0,1,0)
E 4	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(1,1,1)	(0.5,0.75,1,0)	(0.5,0.75,1,0)	(0.25,0.5,0.75)	(0.5,0.75,1,0)	(0.25,0.5,0.75)	(0.5,0.75,1,0)	(0.5,0.75,1,0)
E 5	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(1,1,1)	(0.25,0.5,0.75)	(0,0,0.25)	(0.5,0.75,1,0)	(0,0,0.25)	(0.5,0.75,1,0)	(0.5,0.75,1,0)
E 6	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(1,1,1)	(0,0,0.25)	(0.75,1,0,1,0)	(0.25,0.5,0.75)	(0.75,1,0,1,0)	(0.75,1,0,1,0)
E 7	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(1,1,1)	(0.75,1,0,1,0)	(0.75,1,0,1,0)	(0.75,1,0,1,0)	(0.75,1,0,1,0)
E 8	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(1,1,1)	(0.5,0.75,1,0)	(0.75,1,0,1,0)	(0.75,1,0,1,0)
E 9	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(1,1,1)	(0.25,0.5,0.75)	(0.25,0.5,0.75)
E 10	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(1,1,1)	(0.75,1,0,1,0)
E 11	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(1,1,1)
E 12	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)
E 13	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)
E 14	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)
E 15	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)
E 16	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)
E 17	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)
E 18	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)
E 19	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)
E 20	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)
Dependence Power	(1,1,5,75)	(1,75,2,6,5)	(1,1,5,75)	(1,5,1,75,6,50)	(2,75,3,50,8)	(3,25,4,25,8,5)	(1,75,2,25,7)	(5,75,7,5,11)	(4,25,6,10,25)	(6,75,9,12,25)	(7,5,10,13)
Crisp Value	0.03	0.07	0.03	0.07	0.17	0.20	0.09	0.38	0.29	0.44	0.50

Source: Author own compilation

Table 3b. Final fuzzy reachability matrix

	E12	E13	E14	E15	E16	E17	E18	E19	E20	Driving Power	Crisp Value
E 1	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0.75,1,0,1,0)	(0.75,1,0,1,0)	(0,0,0.25)	(5.75,7.50,10.75)	1.71
E 2	(0.5,0.75,1,0)	(0.5,0.75,1,0)	(0.75,1,0,1,0)	(0.5,0.75,1,0)	(0.25,0.5,0.75)	(0.75,1,0,1,0)	(0.75,1,0,1,0)	(0.75,1,0,1,0)	(0.75,1,0,1,0)	(10.50,14.25,16.75)	0.70
E 3	(0.5,0.75,1,0)	(0.5,0.75,1,0)	(0.5,0.75,1,0)	(0.5,0.75,1,0)	(0.25,0.5,0.75)	(0.5,0.75,1,0)	(0.75,1,0,1,0)	(0.75,1,0,1,0)	(0.75,1,0,1,0)	(11.5,15.75,19.25)	0.76
E 4	(0.25,0.5,0.75)	(0.25,0.5,0.75)	(0.5,0.75,1,0)	(0.25,0.5,0.75)	(0.5,0.75,1,0)	(0.5,0.75,1,0)	(0.75,1,0,1,0)	(0.75,1,0,1,0)	(0.75,1,0,1,0)	(8.50,12.50,16.50)	0.69
E 5	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0.25,0.5,0.75)	(0.5,0.75,1,0)	(0.5,0.75,1,0)	(0.75,1,0,1,0)	(4.75,6.75,11.25)	0.92
E 6	(0.5,0.75,1,0)	(0.25,0.5,0.75)	(0.5,0.75,1,0)	(0.25,0.5,0.75)	(0.25,0.5,0.75)	(0.25,0.5,0.75)	(0.75,1,0,1,0)	(0.75,1,0,1,0)	(0.75,1,0,1,0)	(7.75,11,14.25)	0.53
E 7	(0.75,1,0,1,0)	(0.75,1,0,1,0)	(0.75,1,0,1,0)	(0.75,1,0,1,0)	(0.25,0.5,0.75)	(0.75,1,0,1,0)	(0.75,1,0,1,0)	(0.75,1,0,1,0)	(0.75,1,0,1,0)	(10.25,13.5,15.25)	0.63
E 8	(0.5,0.75,1,0)	(0.5,0.75,1,0)	(0.75,1,0,1,0)	(0.75,1,0,1,0)	(0.75,1,0,1,0)	(0.75,1,0,1,0)	(0.75,1,0,1,0)	(0.75,1,0,1,0)	(0.75,1,0,1,0)	(9.25,12.25,14.75)	0.59
E 9	(0,0,0.25)	(0,0,0.25)	(0.5,0.75,1,0)	(0,0,0.25)	(0.25,0.5,0.75)	(0.25,0.5,0.75)	(0.25,0.5,0.75)	(0.25,0.5,0.75)	(0.5,0.75,1,0)	(3.50,5.50,9.50)	0.26
E 10	(0.5,0.75,1,0)	(0.5,0.75,1,0)	(0.5,0.75,1,0)	(0.5,0.75,1,0)	(0.5,0.75,1,0)	(0.75,1,0,1,0)	(0.75,1,0,1,0)	(0.75,1,0,1,0)	(0.5,0.75,1,0)	(7.9,50,13.25)	0.46
E 11	(0.5,0.75,1,0)	(0.5,0.75,1,0)	(0.5,0.75,1,0)	(0.5,0.75,1,0)	(0.5,0.75,1,0)	(0.75,1,0,1,0)	(0.75,1,0,1,0)	(0.75,1,0,1,0)	(0.75,1,0,1,0)	(6.50,8.75,12.50)	0.43
E 12	(1,1,1)	(0.75,1,0,1,0)	(0.75,1,0,1,0)	(0.5,0.75,1,0)	(0.75,1,0,1,0)	(0,0,0.25)	(0.25,0.5,0.75)	(0.5,0.75,1,0)	(0.5,0.75,1,0)	(5.6,75,12.50)	0.35
E 13	(0,0,0.25)	(1,1,1)	(0.75,1,0,1,0)	(0.25,0.5,0.75)	(0.5,0.75,1,0)	(0.5,0.75,1,0)	(0.75,1,0,1,0)	(0.75,1,0,1,0)	(0.75,1,0,1,0)	(5.25,7,10.75)	0.33
E 14	(0,0,0.25)	(0,0,0.25)	(1,1,1)	(0,0,0.25)	(0,0,0.25)	(0.5,0.75,1,0)	(0,0,0.25)	(0,0,0.25)	(0.75,1,0,1,0)	(2.25,2.75,7.25)	0.11
E 15	(0,0,0.25)	(0,0,0.25)	(0.75,1,0,1,0)	(1,1,1)	(0.75,1,0,1,0)	(0,0,0.25)	(0.5,0.75,1,0)	(0.5,0.75,1,0)	(0.75,1,0,1,0)	(4.25,5.5,9.5)	0.26
E 16	(0,0,0.25)	(0,0,0.25)	(0.75,1,0,1,0)	(0,0,0.25)	(1,1,1)	(0.25,0.5,0.75)	(0.25,0.5,0.75)	(0.25,0.5,0.75)	(0.5,0.75,1,0)	(3.4,25,8.75)	0.19
E 17	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(1,1,1)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(1,1,5.75)	0.20
E 18	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0.5,0.75,1,0)	(1,1,1)	(0.75,1,0,1,0)	(0.75,1,0,1,0)	(3.3,75,8)	0.18
E 19	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0.5,0.75,1,0)	(0.75,1,0,1,0)	(1,1,1)	(0.75,1,0,1,0)	(3.3,75,8)	0.18
E 20	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0,0,0.25)	(0.75,1,0,1,0)	(0.75,1,0,1,0)	(0.75,1,0,1,0)	(1,1,1)	(3.25,4,8)	0.18
Dependence Power	(5,7,11.50)	(5,5,7.75,12)	(9,25,12.5,15.50)	(6,25,8.25,12.50)	(6,5,9.5,13.50)	(9,5,13.5,16.75)	(11,75,16,17.75)	(12,16,25,18)	(12,75,17,18.5)		
Crisp Value	0.36	0.40	0.63	0.43	0.48	0.67	0.79	0.82	0.85		

Table 4. Defuzzified reachability matrix

	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12	E13	E14	E15	E16	E17	E18	E19	E20
E 1	1	1	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	1	1	0
E 2	0	1	0	0	1	1	0	1	1	1	1	1	1	1	1	0	1	1	1	1
E 3	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1
E 4	0	0	0	1	1	1	0	1	0	1	1	0	0	1	0	1	1	1	1	1
E 5	0	0	0	0	1	0	0	1	0	1	1	0	0	0	0	0	0	1	1	1
E 6	0	0	0	0	0	1	0	1	0	1	1	1	0	1	0	0	0	1	1	1
E 7	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1
E 8	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1
E 9	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	1
E 10	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1
E 11	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1
E 12	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	1	1
E 13	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1	1	1	1
E 14	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	1
E 15	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	1	1	1
E 16	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	0	1
E 17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
E 18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1
E 19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1
E 20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1

Source: Author own compilation

Table 5. Defuzzified reachability matrix and transitivity check

	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12	E13	E14	E15	E16	E17	E18	E19	E20
E1	1	1	0	0	1*	1*	0	1	1*	1	1	1*	1*	1*	1*	1*	1*	1	1	1*
E2	0	1	0	0	1	1	1*	1	1	1	1	1	1	1	1	1*	1	1	1	1
E3	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1*	1	1	1	1
E4	0	0	0	1	1	1	0	1	1*	1	1	1*	1*	1	1*	1	1	1	1	1
E5	0	0	0	0	1	0	0	1	1*	1	1	1*	1*	1*	1*	1*	1*	1	1	1
E6	0	0	0	0	0	1	0	1	1*	1	1	1	1*	1	1*	1*	1*	1	1	1
E7	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1*	1	1	1	1
E8	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1
E9	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	1*	1*	1*	1
E10	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1
E11	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1
E12	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1*	1*	1	1
E13	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1	1	1	1
E14	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	1
E15	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1*	1	1	1
E16	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	1*	1*	1
E17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
E18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1
E19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1
E20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1

Source: Author own compilation

Table 6. Level partitioning

Elements	Reachability Set	Antecedent Set	Intersection Set	Level
E1	1,2,5,6,8,9,10,11,12,13,14,15,16,17,18,19,20	1	1	XII
E2	2,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20	1,2	2	XI
E3	3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20	3	3	XII
E4	4,5,6,8,9,10,11,12,13,14,15,16,17,18,19,20	3,4	4	XI
E5	5,8,9,10,11,12,13,14,15,16,17,18,19,20	1,2,3,4,5	5	X
E6	6,8,9,10,11,12,13,14,15,16,17,18,19,20	1,2,3,4,6	6	X
E7	7,8,9,10,11,12,13,14,15,16,17,18,19,20	2,3,7	7	X
E8	8,9,10,11,12,13,14,15,16,17,18,19,20	1,2,3,4,5,6,7,8	8	IX
E9	9,14,17,18,19,20	1,2,3,4,5,6,7,8,9	9	IV
E10	10,11,12,13,14,15,16,17,18,19,20	1,2,3,4,5,6,7,8,10	10	VIII
E11	11,12,13,14,15,16,17,18,19,20	1,2,3,4,5,6,7,8,10,11	11	VII
E12	12,13,14,15,16,17,18,19,20	1,2,3,4,5,6,7,8,10,11,12	12	VI
E13	13,14,16,17,18,19,20	1,2,3,4,5,6,7,8,10,11,12,13	13	V
E14	14,17,20	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16	14	III
E15	14,15,16,17,18,19,20	1,2,3,4,5,6,7,8,10,11,12,15	15	V
E16	14,16,17,18,19,20	1,2,3,4,5,6,7,8,10,11,12,13,15,16	16	IV
E17	17	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20	17	I
E18	17,18,19,20	1,2,3,4,5,6,7,8,9,10,11,12,13,15,16,18,19,20	18,19,20	II
E19	17,18,19,20	1,2,3,4,5,6,7,8,9,10,11,12,13,15,16,18,19,20	18,19,20	II
E20	17,18,19,20	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,18,19,20	18,19,20	II

Source: Author own compilation

Load Flow Analysis in Smart Grids



Osman Hasan

National University of Sciences and Technology, Pakistan

Awais Mahmood

National University of Sciences and Technology, Pakistan

Syed Rafay Hasan

Tennessee Technological University, USA

INTRODUCTION

With 19320 TW-hr/yr consumption of electrical energy in the entire world nowadays, the traditional unidirectional power transmission grids are struggling to survive as the number of fluctuations, blackouts and outages is tremendously growing since the last decade (Gao et al., 2012). More reliable and safe distribution networks have become a dire requirement due to the safety and financial-critical nature of electricity these days. For example, a blackout per minute across Silicon Valley costs 75 million and 1 million dollars for Sun Microsystems alone. There are numerous environmental concerns with the present-age power generation methods as well since these methods are largely dependent on fossil fuels, which result in global warming and carbon-dioxide emissions. For example, the United States power system alone is responsible for 40 percent of carbon emission nationwide (Hledik, 2009). Thus renewable energy resources, like solar and wind based solutions, are extensively being advocated throughout the world but the traditional grid does not facilitate their integration in the national grids. Moreover, the traditional power grids are not very efficient in terms of distribution loss management as well. For example, about 17 percent of electrical energy generated in the year 2011 by Pakistan was wasted in distribution systems. Similarly, the problem of electricity theft is also a growing concern in traditional grids.

Smart grids can overcome the above mentioned shortcomings by providing an alternative electric power transmission framework that comprises of Intelligence based Electronic Devices (IED) (Momoh, 2012) for detecting and correcting faults, and advanced metering infrastructure (AMI), to facilitate the integration of multiple renewable energy sources. Some of the distinguishing characteristics of smart grids compared to traditional power grids include:

- **Safety and Reliability:** Smart grids can predict unforeseen situations and autonomously react accordingly to prevent them (e.g., isolating the faulty component of the grid from the entire system (Farhangi, 2010)) and hence improve the safety and reliability (Moslehi and Kumar, 2010) of power distribution and save millions of dollars.
- **Cost-Effectiveness:** Smart grids provide real-time tariff information to the consumers so that they can manage their loads to save energy and costs (Li et al., 2010).
- **Efficiency:** Smart grids allow optional usage of the assets to maximize the efficiency of the grid and thus can have a major performance impact. For example, according to the US Department of Energy (DOE), just a 5% increase in grid efficiency can have the same impact as if fuel and greenhouse gas emissions are eliminated from 53 million cars.

DOI: 10.4018/978-1-5225-2255-3.ch271

- **Security:** Smart grids allow more secure electrical networks, by using tools like smart meters, and thus electricity theft can be minimized (Khurana et al., 2010, Metke and Ekl, 2010).
- **Environmental Friendliness:** Smart grid allows the integration of environmental friendly generation methods and is inline with the recent advancements in renewable energy research (RER) (Ipakchi and Albuyeh, 2009).

Based on above-mentioned capabilities, the National Academy of Engineering listed “electrification as made possible by the grid” as the most significant engineering achievement of the 20th Century.

Due to the inherent randomness of smart grids, including variable loads, peak consumption times and renewable energy sources with generation capacity depending on varying weather conditions, there is a lot of interest in rigorously analyzing the voltages and load profiles for resilient and effective power delivery to the users. Besides providing means for effectively managing the energy distribution, these profiles can be used by the consumers to change their loads by a smart device from anywhere as per their requirements. Load flow analysis (Van Benthem and Doets, 2001) fulfills the above-mentioned requirements and allows us to find the magnitude and phase angle of the voltage and the real and reactive power flowing in each bus of the smart grid and the optimal parameters for various components, like inductors, conductors, transformers, and shunt capacitors. It also provides statistics about the behavior of the system during on-peak and off-peak loads in order to identify and plan the contingencies. Moreover, load flow studies help us in conducting short-circuit fault analysis and in finding the stability and the steady-state operating state of an electric power system by calculating the voltage drop on each feeder, the power flow in all branches and feeder circuits, X/R ratio in line impedances and the voltage at each bus. Finally, load flow studies can determine if system volt-

ages remain within the given specifications and if any of the expensive equipment of the grid is overloaded. The results of load flow analysis are used to make key decisions and ensure a safe and reliable power distribution.

There are various uncertain and random elements associated with the load consumption in smart grids. For example, the usage of consumer appliances depends on weather conditions and the time of the day. The distributed generation and usage of storage cells also plays a key role in varying the electrical demand. Some of the key factors that influence the loads in smart grids and must be taken into consideration for load flow analysis of smart grids include weather conditions, time-of-day, arbitrary disturbances, electricity prices, demand response, storage cells and electric vehicles.

BACKGROUND

There are various uncertain and random elements associated with the load consumption in smart grids. For example, the usage of consumer appliances depends on weather conditions and the time of the day. The distributed generation and usage of storage cells also plays a key role in varying the electrical demand. The smart grid components may fail randomly and either self-repair or need manual repair to restore their operation. Some components may also have back-up protection. Similarly, the influence of electricity prices on the energy demand cannot be neglected as higher prices usually result in the reduction of energy consumption. Moreover, in smart grids, the consumers are more cautious about costs since they can get the real-time tariffs using smart meters. Time-of-Use (TOU) pricing scheme, which offers low off peak rates, encourages consumers to shift their loads to off peak hours. Moreover, electric vehicles (EVs) also greatly influence load profiles since their charging consumes a significant amount of energy and thus is recommended to be done in the off peak times.

The load flow analysis mainly involves studying the behavior of node voltages and the power entering and leaving the nodes in a smart grid while considering the above-mentioned elements of randomness and uncertainty. In this section, we describe some of the key techniques that have been used for load flow analysis.

Numerical Methods

Numerical methods are one of most widely used analysis methods for load flow. Primarily, these methods are used to solve the following nodal equation:

$$I = V \cdot Y_{bus} \quad (1)$$

where I is the N vector of current in each bus, V is the N vector bus voltages and Y_{bus} is the admittance bus matrix. Iterative numerical methods can solve the Equation (1), however, they generate approximate results mainly because the precision of the results is directly proportional to the number of iterations. Some of the frequently used numerical methods for load flow analysis are described in detail below:

Gauss-Seidel Method

The Gauss-Seidel method (Chapra Steven and Canale Raymond, 2008) is a widely used iterative approach for solving linear equations based on an initial guess. In load flow analysis, we develop an admittance matrix using the Kirchhoffs current law and then an iterative method is used to solve the scheduled reactive and real power at each bus, respectively. The final node voltage is determined in an iterative manner by using intermediate values of node voltages obtained after each iteration. This process converges linearly if the initial estimation is close to the unknown value. A variant of modified Gauss-Seidel method has been used for power flow calculations in (Teng, 2002). Gauss-Seidel method shows slower rates of convergence but its main strength is that it does not need to solve a

complex matrix system. Gauss-Seidel method is not suitable for radial distribution systems where there are branch connections between a large set of surrounding buses (Momoh, 2012).

Newton-Raphson Method

The Newton-Raphson (NR) method is another iterative method to solve non-linear load flow equations. The sensitivity matrix is determined from inverting the Jacobian matrix, which consists of the injected power equations. The NR method has been used to solve three phase power flow equations in (Le Nguyen, 1997). This method is very useful for large systems. But it is computationally inefficient because it does not take advantage of the radial distribution systems (Momoh, 2012). Moreover, the method also fails if the Jacobian matrix is singular. Finally, in the case of a low X/R ratio value, the NR method becomes ill conditioned (Momoh, 2012). On the other hand, it offers a fast convergence rate.

Fast Decouple Method

The Fast Decouple method is one of the most effective techniques used in power system analysis and design. Just like the Newton-Raphson method, it also utilizes the Jacobian matrix. On contrary, in this technique small angle approximation is used to calculate the relatively smaller elements of the Jacobian matrix. However, this method shows poor convergence for low values of the X/R ratio (Iwamoto and Tamura, 1981). This method has been successfully used for three phase power flow studies in (Zimmerman and Chiang, 1995).

Besides using the above mentioned mainstream numerical methods for load flow analysis, the other significant contributions in this direction include using the interval arithmetic (Wang and Alvarado, 1992), the holomorphic embedded load flow method (Trias, 2012) and an optimal multiplier method for ill conditioned systems (Iwamoto and Tamura, 1981). Other significant methods include Zero-mismatch method, Ward and Hale method,

Glimm and Stagg method and Secondary adjustment method for load flow analysis and more details about them are discussed in (Stott, 1974).

Simulation Methods

The main idea behind simulation is to construct a computer-based model of the given system and then analyze the desired load flow properties by observing the behavior of the model under different test cases. A number of dedicated simulations packages for analyzing power distribution systems are available and have been used to analyze the following key aspects (Glover et al., 2011):

- Arc flash hazard and fault analysis,
- Circuit breaker duty,
- Demand management,
- Distribution reliability evaluation,
- Power factor correction,
- Power loss computations,
- Components sizing,
- Voltages/VAR optimization,
- Power quality and reliability,
- Harmonic analysis, and
- Fault detection.

Simulation is a very user-friendly analysis approach since the analysis requires test pattern generation only. On the down side, simulation cannot guarantee absolute correctness of analysis because there is always a possibility that a corner case is missed in test patterns used for analysis. Some of the common load flow software are explained below:

Electrical Transient Analyzer Program

The Electrical Transient Analyzer Program (ETAP) simulator, developed by Operation Technology Inc., is one of the most widely used load flow analysis software. It allows the user to utilize the built-in templates to quickly construct a model of the entire electrical network. The ETAP simulator can be used

to analyze integrated power systems (Khan et al., 2009) and has an ability to track up to 10 million load items. In (Sedighizadeh and Rezazadeh, 2008), a genetic algorithm is used for optimal allocation of distributed generation for improved voltage profile and the correctness of results is evaluated by ETAP. The power flow analysis package in ETAP provides both Newton-Raphson and Accelerated Gauss-Seidel method to solve power flow equations. ETAP offers two methods to calculate the X/R ratio. The first method finds the equivalent resistance and reactance of the entire system, to get a single value X/R ratio for a given location. In the second method, individual branch current contributions (each with a separate X/R ratio) are combined into a single X/R ratio. ETAP has also been used to analyze different distribution system models to minimize their power losses (Ramesh et al., 2009). The developers of ETAP claim that ETAP can also serve as a complete smart grid analysis tool.

GridLAB-D

GridLAB-D is a recently developed open-source power system modeling and simulation tool by the Pacific Northwest National Laboratory (PNNL) of the US DOE. It offers distributed energy resource modeling, integration of transmission and distribution systems, SCADA and metering models. An interesting feature is its external links to MySQL, MATLAB, MS-EXCEL and MS-ACCESS (Schneider et al., 2009). This tool offers timing models ranging from a few seconds to decades. GridLAB-D gives a simulation environment that can be incorporated with a variety of data management and analysis tools. It divides the power flow problem in two parts, i.e., 1) transmission and 2) distribution. It uses Gauss-Seidel iterative method to solve power flows at the transmission side while Forward and Backward Sweep (FBS) method to solve power flow problems at the distribution side (Schneider et al., 2009). Instead of the FBS method, the newer versions of GridLAB-D utilize the Gauss Seidel Three Phase Current Injection

method due to the inability of FBS to handle networked distributions (Schneider et al., 2009).

Use of Matlab for Load Flow Simulations

Although MATLAB, which is a high level language for doing intense numerical computations and programming, is not developed particularly for load flow analysis but it can be used for that purpose. Given the general-purpose nature of MATLAB, it provides a very flexible environment for load flow analysis. For example, the NR method is programmed to solve a 5-bus system on MATLAB (Mallick and Hota, 2015). The Power System Analysis Toolbox (PSAT), which is an open source MATLAB toolbox, has been used for power flow calculations (Milano, 2005).

Power World Simulator

The Power World simulator is a commercial grade power system analysis and simulation package developed by the Power World Corporation. It is designed to simulate high voltage power system operations on a time frame ranging from several minutes to several days. The software can solve up to 100,000 buses with high efficiency. It extensively uses state-of-the-art graphics for better and easy understanding. The simulator is mainly based on the Newton-Raphson method for iteratively solving non-linear equations for power flows. Load flow analysis by fast decouple method is also an option available in this simulator. It uses three nested loops to solve power flows.

Simulation and testing are the state-of-the-art analysis techniques; however, as we have seen that they also use the numerical methods for their computations and thus cannot guarantee accuracy. Moreover, the main idea behind simulation and testing methods is to approximate a solution to a query by observing a subset of each probable run. Hence, it is possible that a system bug may not be detected during the simulation-based analysis. Moreover, system models, used in simulation

cannot capture the true random behaviors, such as frequent changes in renewable energy generation, variations in network configurations and the peak loads, which are very frequently encountered in smart grids. In most cases, simulation based methods rely on pseudo random number generation methods for modeling these elements of randomness and the reliability of the analysis is dependent on the quality of these random number generators.

Computational Intelligence

Computational Intelligence (CI) methods provide a very efficient alternative for verifying and analyzing complex systems that exhibit random behavior. Thus, the grid can be controlled more reliably and more rapidly than humans by multi-variable nonlinear optimal controller based on CI, without requiring a mathematical model of the grid.

CI techniques primarily consist of Artificial Neural Networks (ANNs) (Haykin, 1994), evolutionary computation (Back et al., 1997), non-linear programming (Sasson, 1969) and fuzzy logic systems (Takagi and Sugeno, 1985). ANN based load forecaster is one of the most successful applications of CI for predicting load flows in an uncertain environment. Evolutionary computational techniques tend to solve combinatorial optimization techniques by learning and adapting to new situations and are primarily based on Genetic Algorithm (GA), particle swarm optimization (PSO) and ant colony optimization methods (Vlachogiannis et al., 2005). GAs have been used to analyze the Reactive power (VAR) with real-time operation that contains randomness and uncertainty (Bakirtzis et al., 2002). PSO (del Valle et al., 2008) uses simple mechanism that mimics social behavior of bird flocking and fish schooling to guide the particles search for globally optimal solution. In (Miranda and Saraiva, 1991), fuzzy modeling is used for optimal load flow. Fuzzy logic is also used for developing a unified power flow controller for damping the power system oscillations (Eldamaty et al., 2005). Linear, nonlinear, dynamic and integer program-

ming methods have also been used for load flow analysis (Momoh, 2012). The above-mentioned CI based techniques have also been used to analyze various power systems (Saxena et al., 2010) including maintenance scheduling, long-term system expansion and planning and load forecasting. A very comprehensive overview about CI techniques and the advantages and disadvantages of Genetic Algorithm (GA), Simulated Annealing (SA), Artificial Neural Networks (ANNs), Expert Systems (ES) and few other techniques for analyzing power distribution systems is presented in (Saxena et al., 2010).

CI techniques of linear programming, non-linear programming, quadratic programming and Newton based techniques (Sun et al., 1984) have also been successfully used to solve various Optimal Power Flow (OPF) problems (Habibollahzadeh et al., 1989, Aoki et al., 1987). These techniques have certain drawbacks as mentioned in (Abido, 2002), like, non-linear programming has convergence problems and it is complex, quadratic programming techniques have problems with cost approximation. Newton based techniques may have convergence failure due to inappropriate initial conditions, hence, they are sensitive to initial conditions. In (Abido, 2002), a particle swarm optimization technique is proposed to do OPF.

Probabilistic Load Flow

Given the large number of uncertainties involved in the load flow analysis of smart grids, probabilistic analysis methods have also been used in this domain. Probabilistic Load Flow (PLF) is primarily done by using Monte Carlo (MC) simulations and the convolution method through Fast Fourier Transforms (FFT). These MC simulations are very time consuming since a large number of samples are usually used. An extended form of PLF that supports non-linear load flow equations is presented in (Allan et al., 1976) based on numerical methods. Another technique, known as Stochastic Load Flow (SLF) (Vorsic et al., 1991), is based on the assumption that the states of the system

and power flow outputs are normally distributed. SLF deals with short time uncertainties and is only effective for analyzing system operation and thus the reliability of this method has been questioned (Anders, 1989). PLF has been used in (Allan et al., 1974) for analyzing power flows where the nodal loads and generation are defined as random variables and power flow is computed as a probability density function.

Multi-linear simulation algorithm has demonstrated better results than the MC simulation based methods (Da Silva and Arienti, 1990). Similarly, the method of combined cumulants and Gram-Charlier expansion for probabilistic load flow computation (Zhang and Lee, 2004) has reported a significantly improved performance. With this method, the probability density function of the transmission line flows is obtained. It determines the effect of prolonged uncertainty of transmission network. It also provides a new way of calculating probability density function which requires reduced storage, is applicable to larger systems and is faster than Monte Carlo simulations. It is useful in system expansion planning. Moreover, it also ensures better approximation of the cumulative distribution function curve. This method was practically demonstrated on a WSCC (Western Systems Coordinating Council) test system, which consists of 179 buses and 263 lines. The probability of any overloaded line can be easily computed through this method. The probabilistic load flow analysis using an algorithm based on the point estimate method is given in (Su, 2005). It is based on an assumption that the uncertainties of line parameters and injections in the bus can be measured. The method allows any deterministic load-flow program to be used. In order to calculate the statistical moments of load flow solution distribution for a system having m uncertain parameters, $2m$ calculations of load flow are used and the value of the solution is weighted at $2m$ locations. These moments are then used in probability density fittings. This method has been tested on several IEEE test systems and it is verified that performance of point estimate method

is better than Monte Carlo simulation and it also requires less number of computations. A detailed comparison between the cumulants method and the point estimate method is presented in (Li and Zhang, 2009) with respect to the load flow analysis.

Analyzing random branch outages in load flow is another area where probabilistic load flow is considerably used since conventional load flow methods cannot cater such problems (Hu and Wang, 2006). PLF allows us to solve discrete distribution part of each state and output variable. The effect of branch outage is greater on system state than uncertainties caused by nodal power injections. For simplifying the convolution of random variables, moments and cumulants are used. Branch outages are simulated by injecting the virtual power to the related nodes. The resulting distribution is found by convolving continuous and discrete distributions. The Dynamic Stochastic Optimal Power Flow (DSOPF) (Momoh, 2009), which is based on Adaptive Dynamic Programming (ADP) technique, is another stochastic method for PLF. ADP tends to cope with the complex power system problems, which can be predicted under uncertainty conditions, and it is useful where there is not enough prior knowledge. These tools usually ensure robustness, scalability, stochasticity, predictivity, adaptability and acquisition of instantaneous data.

FUTURE RESEARCH DIRECTIONS

All the above-mentioned techniques are found to be quite scalable and user-friendly but cannot guarantee the accuracy of the analysis results. The main reasons behind the inaccuracies in the result include the usage of computer-arithmetic based models, which contain round-off errors, and the sampling based nature of the analysis, i.e., the models are analyzed for a subset of all possible scenarios due to limited computational resources. Given the safety-critical nature of smart-grids, the accuracy of load flow analysis results is the most desirable feature, since an undetected fault

in the smart grid system can have major impact. For example, the analysis inaccuracy limitations have been reported as the main causes behind the 2003 Northeast blackout in the United States and Canada (Poulsen, 2004a),(Poulsen, 2004b) which approximately affected 55 Million people. As a complementary approach, formal methods (Abrial, 2009) which are computer based mathematical analysis tools, can be used to overcome the inaccuracy limitations in the domain of load flow analysis. However, to the best of our knowledge, no prior work regarding the formal load flow analysis exists so far. In order to fill this gap, we recommend to use probabilistic model checking (J. Rutten and Parker, 2004), which is a widely used formal method for analyzing Markovian models, to ensure accurate results of load flow analysis of smart grids.

CONCLUSION

Load flow analysis plays a vital role in safe and effective working of the smart grid system and a number of analysis methods have been used in this domain. This paper presents a brief overview about smart grids and the main factors affecting the loads in smart grids. This information can be utilized to understand the random and uncertain components in smart grids and thus model them appropriately in their load flow analysis, which can be conducted using Numerical methods, Simulation methods, Computational intelligence and probabilistic load flow methods. Each of these analysis methods have their own advantages and disadvantages and they have been highlighted in this chapter.

REFERENCES

Abido, M. (2002). Optimal power flow using particle swarm optimization. *Electrical Power & Energy Systems*, 24(7), 563–571. doi:10.1016/S0142-0615(01)00067-9

- Abrial, J. (2009). Faultless systems: Yes we can! *Computer*, 42(9), 30–36. doi:10.1109/MC.2009.283
- Allan, R., Borkowska, B., & Grigg, C. (1974). Probabilistic analysis of power flows. Electrical Engineers. *Proceedings of the Institution of*, 121(12), 1551–1556.
- Allan, R., Grigg, C., & Al-Shakarchi, M. (1976). Numerical techniques in probabilistic load flow problems. *International Journal for Numerical Methods in Engineering*, 10(4), 853–860. doi:10.1002/nme.1620100412
- Anders, G. J. (1989). *Probability concepts in electric power systems*. Academic Press.
- Aoki, K., Nishikori, A., & Yokoyama, R. (1987). Constrained load flow using recursive quadratic programming. Power Systems. *IEEE Transactions on*, 2(1), 8–16. doi:10.1109/T-ED.1987.22880
- Back, T., Hammel, U., & Schwefel, H.-P. (1997). Evolutionary computation: Comments on the history and current state. Evolutionary computation. *IEEE Transactions on*, 1(1), 3–17.
- Bakirtzis, A. G., Biskas, P. N., Zoumas, C. E., & Petridis, V. (2002). Optimal power flow by enhanced genetic algorithm. Power Systems. *IEEE Transactions on*, 17(2), 229–236.
- Benysek, G., Kazmierkowski, M., Popczyk, J., & Strzelecki, R. (2011). Power electronic systems as a crucial part of smart grid infrastructure-a survey. *Technical Sciences*, 59(4), 455–473.
- Chapra Steven, C., & Canale Raymond, P. (2008). *Numerical methods for engineers*. McGraw-Hill Pub.
- Da Silva, A. L., & Arienti, V. (1990). Probabilistic load flow by a multilinear simulation algorithm. Generation. *Transmission and Distribution*, 137(4), 276–282. doi:10.1049/ip-c.1990.0037
- del Valle, Y., Venayagamoorthy, G. K., Mohagheghi, S., Hernandez, J.-C., & Harley, R. G. (2008). Particle swarm optimization: Basic concepts, variants and applications in power systems. *Evolutionary Computation*, 12(2), 171–195. doi:10.1109/TEVC.2007.896686
- Eldamaty, A., Faried, S. O., & Aboreshaid, S. (2005). Damping power system oscillations using a fuzzy logic based unified power flow controller. In Electrical and Computer Engineering, (pp. 1950–1953). doi:10.1109/CCECE.2005.1557364
- Farhangi, H. (2010). The path of the smart grid. *Power and Energy Magazine, IEEE*, 8(1), 18–28. doi:10.1109/MPE.2009.934876
- Gao, J., Xiao, Y., Liu, J., Liang, W., & Chen, C. (2012). A survey of communication/networking in smart grids. *Future Generation Computer Systems*, 28(2), 391–404. doi:10.1016/j.future.2011.04.014
- Glover, J. D., Sarma, M. S., & Overbye, T. J. (2011). *Power system analysis and design*. CengageBrain.com.
- Habibollahzadeh, H., Luo, G.-X., & Semlyen, A. (1989). Hydrothermal optimal power flow based on a combined linear and nonlinear programming methodology. *Power Systems*, 4(2), 530–537. doi:10.1109/59.193826
- Haykin, S. (1994). *Neural networks: a comprehensive foundation*. Prentice Hall.
- Hledik, R. (2009). How green is the smart grid? *The Electricity Journal*, 22(3), 29–41. doi:10.1016/j.tej.2009.03.001
- Hu, Z., & Wang, X. (2006). A probabilistic load flow method considering branch outages. *Power Systems*, 21(2), 507–514. doi:10.1109/TPWRS.2006.873118
- Ipakchi, A., & Albuyeh, F. (2009). Grid of the future. *Power and Energy Magazine, IEEE*, 7(2), 52–62. doi:10.1109/MPE.2008.931384

- Iwamoto, S., & Tamura, Y. (1981). A load flow calculation method for ill-conditioned power systems. *Power Apparatus and Systems*, (4), 1736–1743.
- Khan, R. A. J., Junaid, M., & Asgher, M. M. (2009). Analyses and monitoring of 132 kv grid using etap software. In *Electrical and Electronics Engineering* (p. I–113). IEEE.
- Khurana, H., Hadley, M., Lu, N., & Frincke, D. A. (2010). Smart-grid security issues. *Security & Privacy, IEEE*, 8(1), 81–85. doi:10.1109/MSP.2010.49
- Le Nguyen, H. (1997). Newton-raphson method in complex form power system load flow analysis. *Power Systems*, 12(3), 1355–1359. doi:10.1109/59.630481
- Li, F., Qiao, W., Sun, H., Wan, H., Wang, J., Xia, Y., & Zhang, P. et al. (2010). Smart transmission grid: Vision and framework. *Smart Grid*, 1(2), 168–177. doi:10.1109/TSG.2010.2053726
- Li, G., & Zhang, X.-P. (2009). Comparison between two probabilistic load flow methods for reliability assessment. In *Power & Energy Society General Meeting*, (pp. 1–7). IEEE. doi:10.1109/PES.2009.5275534
- Mallick, A. K., & Hota, B. S. (2015). *Load flow study in power system*. Retrieved from [http://ethesis.nitrkl.ac.in/2464/1/load flow in power system.pdf](http://ethesis.nitrkl.ac.in/2464/1/load%20flow%20in%20power%20system.pdf)
- Metke, A. R., & Ekl, R. L. (2010). Security technology for smart grid networks. *Smart Grid*, 1(1), 99–107. doi:10.1109/TSG.2010.2046347
- Milano, F. (2005). An open source power system analysis toolbox. *Power Systems*, 20(3), 1199–1206. doi:10.1109/TPWRS.2005.851911
- Miranda, V., & Saraiva, J. (1991). Fuzzy modeling of power system optimal load flow. In *Power Industry Computer Application* (pp. 386–392). IEEE; doi:10.1109/PICA.1991.160606
- Momoh, J. (2012). *Smart grid: fundamentals of design and analysis* (Vol. 63). Wiley; doi:10.1002/9781118156117
- Momoh, J. A. (2009). Smart grid design for efficient and flexible power networks operation and control. In *Power Systems Conference and Exposition* (pp. 1–8). IEEE; doi:10.1109/PSCE.2009.4840074
- Moslehi, K., & Kumar, R. (2010). A reliability perspective of the smart grid. *Smart Grid*, 1(1), 57–64. doi:10.1109/TSG.2010.2046346
- Poulsen, K. (2004a). *Software bug contributed to blackout*. Retrieved from <http://www.securityfocus.com/news/8016>
- Poulsen, K. (2004b). *Tracking the blackout bug*. Retrieved from <http://www.securityfocus.com/news/8412>
- Ramesh, L., Chowdhury, S., Chowdhury, S., Natarajan, A., & Gaunt, C. (2009). Minimization of power loss in distribution networks by different techniques. *Energy and Power Engineering*, 2(1).
- Rutten, J., Kwiatkowska, M. G. N., & Parker, D. (2004). *Mathematical techniques for analyzing concurrent and probabilistic systems*. American Mathematical Society.
- Sasson, A. M. (1969). Nonlinear programming solutions for load-flow, minimum-loss, and economic dispatching problems. *Power Apparatus and Systems*, (4), 399–409.
- Saxena, D., Singh, S., & Verma, K. (2010). Application of computational intelligence in emerging power systems. *Engineering. Science and Technology*, 2(3), 1–7.
- Schneider, K. P., Chassin, D., Chen, Y., & Fuller, J. C. (2009). Distribution power flow for smart grid technologies. In *Power Systems Conference and Exposition, 2009. PSCE'09. IEEE/PES*, (pp. 1–7). IEEE. doi:10.1109/PSCE.2009.4840078

- Sedighizadeh, M., & Rezazadeh, A. (2008). Using genetic algorithm for distributed generation allocation to reduce losses and improve voltage profile. World Academy of Science. *Engineering and Technology*, 37, 251–256.
- Stott, B. (1974). Review of load-flow calculation methods. *Proceedings of the IEEE*, 62(7), 916–929. doi:10.1109/PROC.1974.9544
- Su, C.-L. (2005). Probabilistic load-flow computation using point estimate method. *Power Systems*, 20(4), 1843–1851. doi:10.1109/TPWRS.2005.857921
- Sun, D. I., Ashley, B., Brewer, B., Hughes, A., & Tinney, W. F. (1984). Optimal power flow by newton approach. *Power Apparatus and Systems*, (10), 2864–2880.
- Takagi, T., & Sugeno, M. (1985). Fuzzy identification of systems and its applications to modeling and control. *Systems, Man and Cybernetics*, (1), 116–132.
- Teng, J.-H. (2002). A modified Gauss–Seidel algorithm of three-phase power flow analysis in distribution networks. *Electrical Power & Energy Systems*, 24(2), 97–102.
- Trias, A. (2012). The holomorphic embedding load flow method. In *Power and Energy Society General Meeting*, (pp. 1–8). IEEE.
- Van Benthem, J., & Doets, K. (2001). Higher-order-logic. In *Handbook of Philosophical Logic* (pp. 189–243). Springer.
- Vlachogiannis, J. G., Hatziaargyriou, N. D., & Lee, K. Y. (2005). Ant colony system-based algorithm for constrained load flow problem. *Power Systems*, 20(3), 1241–1249. doi:10.1109/TPWRS.2005.851969
- Vorsic, J., Muzek, V., & Skerbinek, G. (1991). Stochastic load flow analysis. In *Electrotechnical Conference*, (pp. 1445–1448). IEEE.
- Wang, Z., & Alvarado, F. L. (1992). Interval arithmetic in power flow analysis. *Power Systems*, 7(3), 1341–1349. doi:10.1109/59.207353
- Zhang, P., & Lee, S. T. (2004). Probabilistic load flow computation using the method of combined cumulants and gramcharlier expansion. *Power Systems*, 19(1), 676–682. doi:10.1109/TPWRS.2003.818743
- Zimmerman, R. D., & Chiang, H.-D. (1995). Fast decoupled power flow for unbalanced radial distribution systems. *Power Systems*, 10(4), 2045–2052. doi:10.1109/59.476074

ADDITIONAL READING

- Endrenyi, J., Anders, G., & Leite da Silva, A. (1998). Probabilistic evaluation of the effect of maintenance on reliability. an application. *Power Systems*, 13(2), 576–583. doi:10.1109/59.667385
- Fan, J., & Borlase, S. (2009). The evolution of distribution. *Power and Energy Magazine, IEEE*, 7(2), 63–68. doi:10.1109/MPE.2008.931392
- Miceli, R. (2013). Energy management and smart grids. *Energies*, 6(4), 2262–2290. doi:10.3390/en6042262
- Schneider, J., Gaul, A. J., Neumann, C., Hogräfer, J., Wellßow, W., Schwan, M., & Schnettler, A. (2006). Asset management techniques. *Electrical Power & Energy Systems*, 28(9), 643–654. doi:10.1016/j.ijepes.2006.03.007

KEY TERMS AND DEFINITIONS

Asset Management Systems: Asset management of smart grids is a core requirement due to the huge investments involved. For example, according to the US Department Of Energy (DOE), around 1.5 trillion dollars have been invested in the US electricity infrastructure so far. Asset management applies to both tangible and intangible assets.

Distribution Management System: The Distribution Management System (DMS) may be regarded as the control center of the smart grid. The DMS mainly uses the fault location, Geographic Information Systems (GIS) and Outage Management System (OMS) to improve the reliability of the smart grid by reducing outages and sustaining the frequency and voltage levels. The most important role of DMS is to check the faults and isolate the faulty part out of the system. The “intelligent nodes” of the DMS can communicate with one-another periodically and if a fault occurs then they work together to reconfigure the system.

Energy Management System: The Energy Management System (EMS) is used for monitoring and controlling the performance of the generation and transmission system. It allows getting real-time updates from power plants about their conditions and generation parameters. The monitor and control functions are implemented through Supervisory Control and Data Acquisition (SCADA).

Renewable Energy Integration: Integrating various renewable energy sources is the most desirable feature of smart grids. However, this component faces various challenges. Advanced energy storage at the transmission, distribution, and residential levels, Static VAR compensators and synchro-phasors within the transmission grid, dynamic pricing demand response, micro grids, virtual power plants, and smart wind and solar technologies are some of the tools for Renewable Energy Integration (REI).

E

Methodology of Climate Change Impact Assessment on Forests

Mostafa Jafari

Regional Institute of Forest and Rangelands (RIFR), Iran

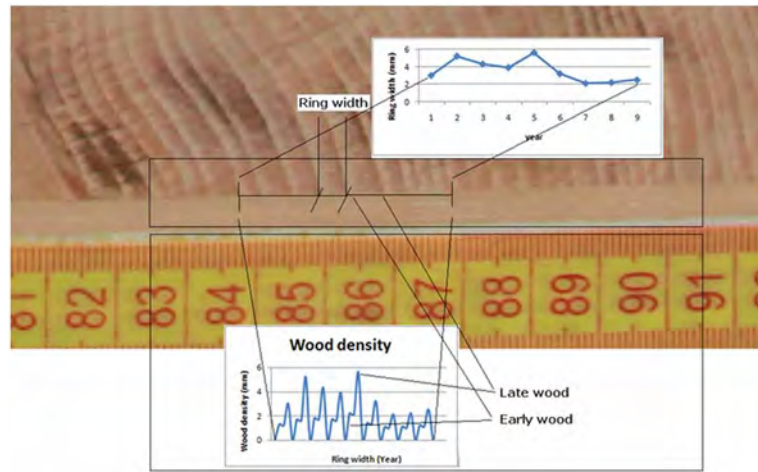
INTRODUCTION

Climate change is one of the main challenging issues in various countries (Jafari, 2013b) in current century. Climate change and climate variability and Global Warming and its' effects on natural resources, plants, animal and in general on human life are among subjects that received attention of scientists and politicians in recent years. Climate change challenges need to be considered in various dimensions (Jafari, 2013c). To both understand the present climate and to predict future climate change, it is necessary to have both theory and empirical observation. Any study of climate change involves the construction (or reconstruction) of time series of climate data. How these climate data vary across time provides a measure (either quantitative or qualitative) of climate change. Types of climate data include temperature, precipitation (rainfall), wind, humidity, evapotranspiration, pressure and solar irradiance (aric, 2008). Climate change assessments and evaluation should be done by using recorded observation data as well as prepared and provided proxy data (Jafari, 2010). Plant ecophysiological study has very important role to recognize climate changes (Jafari, 2007). Trees and also woods can be used as archive of past events. Climate change will strongly affect water resources, plant communities and wildlife in the arid and semi-arid regions (FAO, 2009). Water, environment humidity and temperature are main factors of plant growth. Majority of plant and forest ecosystems on the earth are formed under these two main factors. Whatever amount of humidity and required water are available and also favorable temperature for plant growth cause

plant community reach higher plants and trees and forest ecosystems would develop. In fact plants are important climate indicators. Trees are not an exception. Plants, especially, trees are sensitive to their environmental changes, and tree-ring width is one of the reliable proxies of ambient environmental conditions. Climate and environmental changes affect natural ecosystems as well as planted forests (Kiaee and Jafari, 2014). Investigation of quantity and quality of these growths could help to consider past climatic conditions. Measuring and recording tree rings' widths and its' densities of early woods and late woods can provide valuable data resources to produce time series and consider its correlation with climate factors in the same time periods (Figure 1).

Seasonal changes in temperate climatic region effect on tree rings widths periodically. In spring and summer time plants grow better than unpleasant seasons like fall and winter. The outermost layer of a tree is composed of bark. Bark itself is composed of two tissues: an innermost layer of live phloem, and an outer layer of periderm (the bark 'proper'), which has an outermost layer of waterproofing cork (phellum) which protects the wood to some degree from insects, etc (Figure 2). The cork has its own cambium (phellogen) between the phloem and cork layer. Only the outermost layer of a tree is alive (essentially only the phellogen, phloem, cambium, and maturing xylem of the current year's growth). Consequently, the majority of the trunk does not require gaseous exchange. The bark is punctuated by lenticels, a sort of giant stoma, which allows the thin outermost living layers of the trunk to 'breathe' (Anonymous, 2008a),

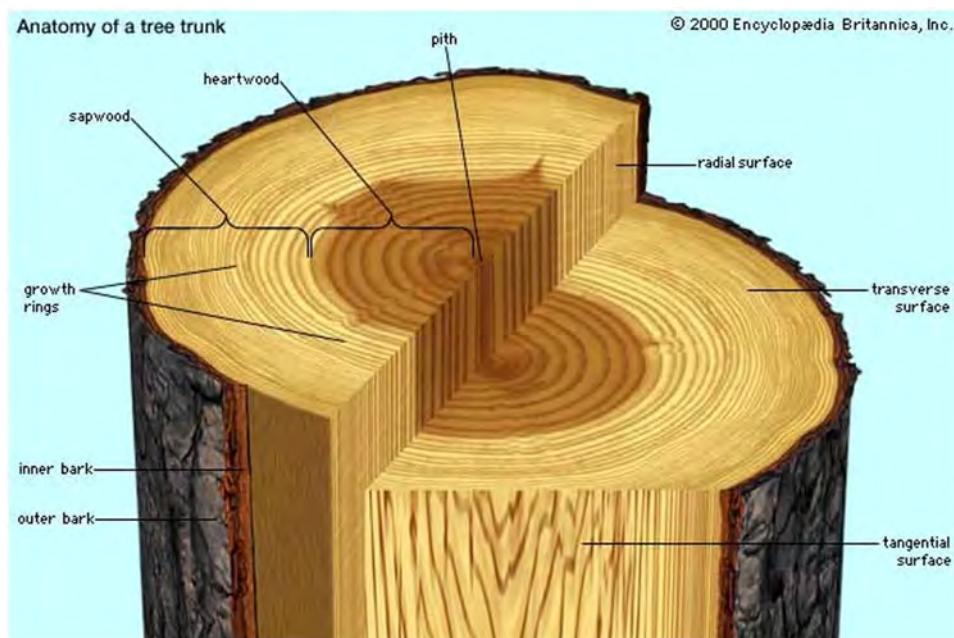
Figure 1. Tree ring width and densities, *Fagus orientalis* (beech tree), Mazandaran province mid-elevation forest (MA II F3)
(Author, 2010)



Growths of the vascular cambium tissue produce wood as secondary xylem production. Sapwood is xylem that conveys water and dissolved minerals from the roots to the rest of the tree. The darker heartwood is older xylem that has been infiltrated by gums and resins and has lost

its ability to conduct water. Each growth layer is distinguished by early wood (springwood), composed of large thin-walled cells produced during the spring when water is usually abundant, and the denser latewood (summerwood), and composed of small cells with thick walls. Growth rings vary in

Figure 2. Anatomy of a tree trunk
(Encyclopedia Britannica, 2000)



width as a result of differing climatic conditions; in temperate climates, a ring is equivalent to one year's growth. Certain conducting cells form rays that carry water and dissolved substances radially across the xylem. Bark comprises the tissues outside the vascular cambium, including secondary phloem (which transports food made in the leaves to the rest of the tree), cork-producing cells (cork cambium), and cork cells. The outer bark, composed of dead tissue, protects the inner region from injury, disease, and desiccation (Encyclopedia Britannica, 2006). A big trunk of a harvested tree can be use as an archive of data and may provide its life long time series (Figure 3)(Jafari, 2010).

Main objectives of dendrochronology are: a) Put the present in perspective of the past, b) Better understand current environmental processes and conditions, and c) Improve understanding of possible environmental issues of the future. To meet these objectives, the exact year of formation of each growth ring must be known: a) Merely counting rings doesn't ensure accurate dating, and b) Crossdating, also known as pattern matching, ensures accurate dating (Sheppard, 2013).

BACKGROUND

Dendrochronology is an accepted and reliable method in considering climate change impact on forest ecosystem through study tree ring widths (Jafari, 2015). The cambium of the trees growing in temperate zones becomes dormant in the falls and reactivates each spring. This leads to annual rings and the vessels produced in the spring are often larger than in the fall; the large vessels allow for rapid sap movement in the spring, whereas the narrow vessels minimize the risk of cavitations under dry conditions in late summer. This leads to a ring-porous pattern in the wood as opposed to the diffuse porous pattern where vessels are more even in size. As trees age the vessels in the center of the stem become air-filled and cease to carry water; they still function for support and storage of waste products, some of which are colored; this is the heartwood in contrast to the sapwood which carries water and is confined to the outer few annual rings (The Ohio State University, 2013).

Recording temperature and using thermometers have only been widely used since around 1850. Thus, the instrumental record for earlier

*Figure 3. Hyrcanian forest research site, an old trunk of Fagus orientalis L. (Asalem, Guilan province forest)
(Jafari, 2010)*



times is quite poor and full of gaps. Essentially nothing is available in the way of quantitative measurements of weather conditions for the time before 1800 A.D. To reconstruct climate change, therefore, we need to use indirect indicators. One source of information is historical records: logs, dairies, lists on when the wine harvest began, reports on when the ice first broke up in a northern river, or when the cherry trees first blossomed. In some cases, such reports go back hundreds of years, although rarely in unbroken sequence. Logs and dairies are treasured finds; they do not exist for most regions of the planet (Anonymous, 2008b).

Climate factors data can be measured by direct observation in different meteorological station (like climatology and synoptic stations) or can be recorded by different instrument in different locations and with different time intervals. This information is more or less confidential for judgments on the past events, and good tools for the future projections.

Applications of Dendrochronology

As definition point of view, the word dendrochronology is composed of: dendro (using trees, or more specifically the growth rings of trees), chrono (time, or more specifically events in past time), and logy (the study of). Applications in dendrochronology include: ecology (insect outbreaks, forest stand structure, past fires) (Jafari, 2012a),

climatology (past droughts or cold periods), geology (past earthquakes, volcanic eruptions), and anthropology (past construction, habitation, and abandonment of societies) (Sheppard, 2013). Also some new terms provided (Jafari, 2013a) for related applications such as, dendro-productivity (Jafari & Khoranke, 2013), dendro-genetic (Jafari *et al.*, 2012b), dendro-medical (Jafari, 2014a).

In region where the seasons provide clear seasonal climate difference, trees develop annual rings of different properties depending on weather, rain, temperature, soil pH, plant nutrition, CO₂ concentration, etc. in different years. These variations are used in dendroclimatology to infer past climate variations. Annual rings width of old trees wood sample is valuable data source as a live archive document for past climate changes, if year of growth and cross dating be well recognized. In case of sample from standing live trees, growth year is identifiable. But in case of wood samples, which could be found in archeological sites, recognizing growth year by producing skeleton graph and cross dating is necessary. Wood samples could show the years of past events like fire, drought or flood in growing site of sample tree. Fire-scarred ponderosa pine (*Pinus ponderosa*) from Ashenfelter Basin, Laramie Peak, Wyoming, (Figure 4). Low to moderate intensity fires that burned through a forest may injure or scar surviving trees, leaving a clear record of their passage. (Swetnam and Baisan, 2002)

Figure 4. Fires may injure or scar surviving trees (Swetnam and Baisan, 2002)



Proxy Data and Climate Change

Climatologists who study past—or paleo—climates (Paleo-climatologists) use the term “proxy” to describe a way that climate change is recorded in nature, within geological materials such as ocean or lake sediments, tree-rings, coral growth-bands, ice-cores, and cave deposits.

For a proxy to be useful, it must first be established that the proxy (i.e. tree-ring width, stable isotope composition of ice, sediment composition) is in fact sensitive to changes in temperature (or some other environmental parameter). This phase of research is known as calibration of the proxy. Perhaps the most frequently used temperature proxy is the relative abundance of microfossils in sediments. That microfossils bear witness to temperature was recognized early in the history of oceanography.

Measuring and recording of tree-ring width is another reliable source of proxy of ambient environmental conditions. When a tree grows at high elevation, near the tree limit, its growth is limited by temperature, and the thickness of its growth rings contains clues about whether the growing season was warm or cold. An equation can then be written relating the changes in ring width to temperature change. Similarly, if the growth is limited by water (say, in a warm semi-arid setting) the ring width can be used to calculate changes in rainfall. Climate proxies have been utilized to provide a semi-quantitative record of average temperature in the Northern Hemisphere back to 1000 A.D (Anonymous, 2008b).

To provide paleo proxy data paleo-climatologists gather proxy data from natural recorders of climate variability such as tree rings, ice cores, fossil pollen, ocean sediments, corals and historical data. By analyzing records taken from these and other proxy sources, scientists can extend our understanding of climate far beyond the 100+ year instrumental record.

Principle sources of the major types of proxy climatic data for palaeoclimatic reconstructions can be categorized as following groups (Jafari,

2010): Glaciological (Ice Cores), Oxygen isotopes, Physical properties, Trace element & micro-particle concentrations

- **Geological:** A. Sediments, 1. Marine (ocean sediment cores), i) Organic sediments (plank-tonic & benthic fossils), Oxygen isotopes, Faunal & floral abundances, Morphological variations, ii) Inorganic sediments: Mineralogical composition & surface texture, Distribution of terrigenous material (provided by river erosion), Ice-rafted debris.
- **Geochemistry:** 2. Terrestrial, Periglacial features, Glacial deposits & erosion features, Glacio-eustatic features (shorelines and sea level changes), Aeolian deposits (sand dunes), Lacustrine deposits/varves (related to the lakes), B. Sedimentary Rocks, Facies analysis, Fossil/microfossil analysis, Mineral analysis Isotope geochemistry.
- **Biological:** Tree rings (width, density, isotope analysis), Pollen (species, abundances), Insects.
- **Historical:** Meteorological records, Parameteorological records (environmental indicators), Phenological records (biological indicators).

Proxy material differs according to its: a) its spatial coverage; b) the period to which it pertains; and c) its ability to resolve events accurately in time (Bradley, 1985). Some proxy records, for example ocean floor sediments, reveal information about long periods of climatic change and evolution, with a low-frequency resolution. Others, such as tree rings are useful only during the last 10,000 years at most, but offer a high frequency (annual) resolution. The choice of proxy record (as with the choice of instrumental record) very much depends on what physical mechanism is under review. As noted, climate responds to different forcing mechanisms over different time scales, and proxy materials will contain necessary cli-

matic information on these to a greater or lesser extent, depending on the three factors mentioned (aric, 2008).

Natural Archives

Growth conditions can be recorded in tree rings. A wide ring could be define as plenty of warm days and sufficient water, a narrow ring means nasty conditions, either a short growing seasons because summer was late in coming (up on the mountain), or a severe water shortage (in the foothills, in areas where water is limiting). The mixture of conditions recorded (time of snow-melt, intensity of winter rain, temperature in June, etc.) depends on what a given tree cares about in terms of growth. Hence, a tree is a “reporter,” and the same is true for all other organisms recording climate change. What a scientist can extract from tree rings depends on how many properties of a ring can be measured (width, density of early wood, density and width of wood grown late in the season), how clever the statistical methods are, and how well the items of interest (say, spring temperature or annual rainfall) are correlated with the properties measured. For instance, special measurements can be made on the isotope chemistry of the wood. This kind of information can yield insights on the composition of the rainfall (from oxygen isotopes) and on the rate of photosynthesis (from carbon isotopes) (Anonymous, 2008b).

While tree growth is influenced by climatic conditions, patterns in tree-ring widths, density, and isotopic composition reflect variations in climate. In temperate regions where there is a distinct growing season, trees generally produce one ring a year, and thus record the climatic conditions of each year. Trees can grow to be hundreds to thousands of years old and can contain annually-resolved records of climate for centuries to millennia.

Tree Rings Measurement Instruments

E

Outcome from dendrochronological research studies played an important role in the early days of radiocarbon dating. Tree rings provided truly known-age material needed to check the accuracy of the carbon 14 dating method. During the late 1950s, several scientists (notably the Dutchman Hessel de Vries) were able to confirm the discrepancy between radiocarbon ages and calendar ages through results gathered from carbon dating rings of trees. The tree rings were dated through dendrochronology.

Even now, tree rings are still used to calibrate radiocarbon determinations. Libraries of tree rings of different calendar ages are now available to provide records extending back over the last 11,000 years. The trees often used as references are the bristlecone pine (*Pinus aristata*) found in the USA and waterlogged Oak (*Quercus* sp.) in Ireland and Germany. Radiocarbon dating laboratories have been known to use data from other species of trees (BETA, 2013).

Borer core samples submitted to the laboratory are registered, followed by preparation of optimal surfaces for analysis across several growth radii of the tree. Subsequently, tree-ring series are measured manually (Figure 5) and registered and or using specially designed measuring devices (Lintab and Aniol) connected to computers (Figure 6), screens and printers. The soft wares used for data storage, cross-correlation and statistical analyses are CATRAS and ITRDBLIB. Samples subject to wood anatomical determination are analyzed with light microscopy and compared against the laboratory’s extensive reference collection of European woody plants (Hammarlund, 2013). Major Equipment in the Biogeography/ Dendrochronology Laboratory currently houses one Velmex Measuring Machine connected to a

Stereozoom Microscope on a boom stand and a microcomputer analysis system. The lab equipment also includes: - incremento borers, - stereozoom microscopes, - belt sanders (4X24"), - desktop and laptop computers, - GPS units, Stihl Chain saw Stihl 046 with a 24" bar, Hand saws, - Hood, - cruiser packs, - Soil Sampling probes, - Munsell Charts, - Measuring tapes, - map tubes, - Paper straws, Poplar core mounts, skeleton plot paper (Indiana State University, 2013).

Discs or Borer Core Samples

The stem cross section is the best way to have a good surface on which to observe tree ring series. These discs can be sometimes obtain in co-operation with foresters when timbering is programmed. In most of the cases, cross dating and then measurements of ring-width as well as densitometry analysis are performed with small cores 4mm in diameter extracted from the tree by an increment borer. In order to avoid dissymmetry in radial tree-growth measurement 2 or 3

cores are extracted on each tree. Consequently, on each sampling site 20 to 45 cores are collected and brought back to the laboratory. Precise cares have to be taken in coring, particularly when densitometry analysis will be performed. The most important is the position of the borer on the trunk. In order to obtain lately an observation surface the most perfectly perpendicular to the long axis of tracheids and fibers, the borer has to be also positioned perpendicularly to trunk axis (TGTC, 2008).

In case which disc of harvested trunk is not available (Figure 7) or it is not possible to take disc, core sampling is an alternative. There are different types of drill instrument for this purpose (Figure 8). Increment borers are instruments that take a small cylindrical core from trees and allow determination if radial growth or age of the tree (Figure 8c). The borer consists of three parts (Figure 8): the case or handle, the borer bit, and the extractor. Increment borers come in various sizes, from 4 inches (101.6 mm) to 30 inches (762 mm) length or more. The ones which are in usual

*Figure 5. Measuring disc sample tree ring widths in the Golestan research centre lab. using manual measurements
(Author, 2010)*



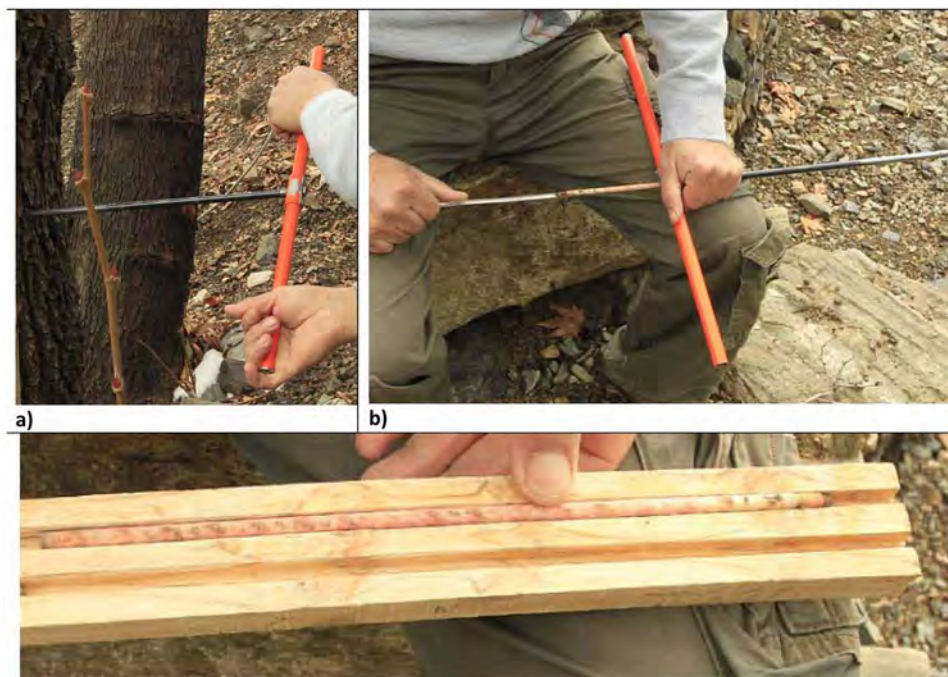
Figure 6. Measuring core sample tree ring widths in the dendrochronology lab. using computer equipped machine
(Author, 2012)



Figure 7. Big trunk disc sample taken during harvesting process
(Author 2010)



Figure 8. Wood sample by increment borer: a) Using borer in the trunk, b) Extracting core sample from borer, c) sample in wooden holder
(Eng Khoshnevis, photos by the author, 2012)



use are in the 8-18 inch range (457.2 -203.2 mm). Smaller borers are used for small trees or where only recent growth is needed from larger trees (Figure 8a, b, and c, Figure 9).

In the field sampling experiment, when the cores remove from the borer, for the safety, it is needed to lay them into an increment core holder that has been pre-glued. Tightly wind the glued cores with cotton string (Figure 10) so as to apply pressure during the drying process (McCarthy, 2008).

Immediate observation of rings on cores extracted from the borer is rarely possible, and moreover rings width measurement quite impossible. Good observation and measurement need a perfect transverse section. After a correct reorientation of the core as the piece of wood was in the trunk, such a section is obtained either by refreshing with a razor blade or polishing the surface selected in order to obtain a plane surface allowing to access to the cell structure (TGTC, 2008).

Measured Data Analysis

A great number of laboratories in different part of the world have been established to study on wood and climate changes. Palaeo-climate changes are on the target of the most of these institutes. The International Tree-Ring Data Bank is maintained by the NOAA Paleoclimatology Program and World Data Center for Paleoclimatology. The Data Bank includes raw ring width or wood density measurements, and site chronologies (growth indices for a site). Tree-ring measurement series from other parameters are welcome as well. Reconstructed climate parameters, including North American Drought, are also available for some areas. Over 2000 sites on six continents are included (WDC, 2008). The objective of the measuring ring widths would be to develop tree-ring records of climate over the past several centuries, to understand inter-annual to century scale variability in climate. This study will improve the capability of understanding environmental variability and key features of the

Figure 9. Borer samples in wooden holder
(Author, 2014)

E



Figure 10. Borer sample and sample holder in the field
(Author, 2012)



regional environment, e.g. persistence of drought, reliability of stream flow (Hughes & Touchan, 1997). Analyzing data by using software by accuracy of 0.01 mm (Robinson & Evans 1980) or more accurate up to 0.001 mm.

Statistical Analysis of the Climate Factors

Tree-rings can provide continuous yearly paleoclimatic records for regions or periods of time with no instrumental climate data. However, different species respond to different climate parameters with, for example, some sensitive to moisture and others to temperature. For example four common species which grow in Northern Ireland and their suitability for climate reconstruction are beech, oak, ash and Scots pine. Beech and ash are the most sensitive to climate, with tree-ring widths more strongly influenced by precipitation and soil moisture in early summer than by temperature or sunshine. Oak is also sensitive to summer rainfall, whereas Scots pine is sensitive to maximum temperature and the soil temperature. The moisture-related parameters, rainfall and the Palmer Drought Severity Index (PDSI), and to a lesser extent, maximum and mean temperatures, can be reconstructed. Reconstructions of climate parameters with tree-rings as proxies may be relatively stable for some seasons such as May–July. The combinations of species are more successful in reconstructing climate than single species (García-Suárez *et al.*, 2009). The development of dendrochronological time series in order to analyze climate-growth relationships usually involves first a rigorous selection of trees and then the computation of the mean tree-growth measurement series. A change in the perspective, passing from an analysis of climate-growth relationships that typically focuses on the mean response of a species to investigating the whole range of individual responses among sample trees (Carrer, 2011).

Crossdating

Primary phase of crossdating work, under a good dissecting microscope, begins by counting backwards from the first known year behind the bark. Using a fine mechanical pencil, place a single dot on each decadal ring (e.g., 2010, 2000, 1990, etc.), place two dots on each 50-year ring (2010, 1960, 1910, etc.), and three dots on each century ring (2010, 1910, 1810, etc.). At this stage, these marks are just temporary year assignments. The actual years will be confirmed after skeleton plotting (McCarthy, 2008).

Using a mm graph paper is first step to draft skeleton plots. The decades are labeled on the x-axis and a vertical line is drawn on a y-scale composed of 10 units. Any ring that is smaller than its neighbor rings (± 3 on either side) gets a line drawn on the paper. If the ring is very small, the line may be 10 units. If the ring is half as small as its predecessor you might code it as a 5, etc. (rings that are coded less than a 5 are rarely useful in crossdating). This is a bit counterintuitive because the longer the line, the smaller the ring. According to the provided skeleton graphs, cross-dating among different samples comparing with control sample would be possible. It is also possible to recognize different years of various samples from different sources for cross-dating.

A more precise method of dating volcanic deposits of recent age is to identify anomalous growth patterns among the annual rings of trees growing at the time the deposits were emplaced. Trees that were injured but not killed by tephra or lahars may show a sequence of narrow rings beginning at the time of impact. “Cross-dating,” the matching of ring-width variation patterns in one tree with corresponding ring patterns in another, should be used to ensure that dating errors are not introduced by missing rings.

Missing rings can often be identified by drawing narrow rings in cross-dating progress. Control tree shows 23 annual rings between the 1472 and

1495 narrow rings (Figure 11), while the tree in the We (Mount St. Helens) tephra-fall zone shows only 20 rings. The series of missing and narrow rings starting with the 1482 ring were caused by tree injury during fallout of layer we. Because of possible missing rings, dates of past volcanic events cannot be determined unequivocally by counting back to a series of narrow rings (Brantley *et al.*, 1986).

Since, the same set of environmental factors influence tree growth throughout a region, the patterns of ring characteristics, such as ring widths, are often common from tree to tree. These patterns can be matched between trees in a process called crossdating (Figure 12), which is used to assign exact calendar year dates to each individual ring (The University of Arizona, 2013). The chronology provides two main types of information:

1. The chronology can be used as a tool for dating events that caused tree death or a

marked change in the appearance of a ring or set of rings. The death date can be used to date the tree cutting involved in the construction of wooden dwellings. Scars can record the timing of events such as fire, flood, avalanche, or other geomorphologic events, while sequences of suppressed or larger rings record events such as insect infestation, effects of pollution, or changes in forest dynamics.

2. The chronology is an average of coherent variations in growth from a number of trees. It enhances the common pattern of variation or “signal” -usually related to climate- while the non-common variance or “noise” is dampened. Chronologies from trees that are sensitive to climate can be used to reconstruct past variations in seasonal temperature, precipitation, drought, stream flow, and other climate-related variables (The University of Arizona, 2013).

Figure 11. Tree-ring cross-dating
(Brantley *et al.*, 1986)

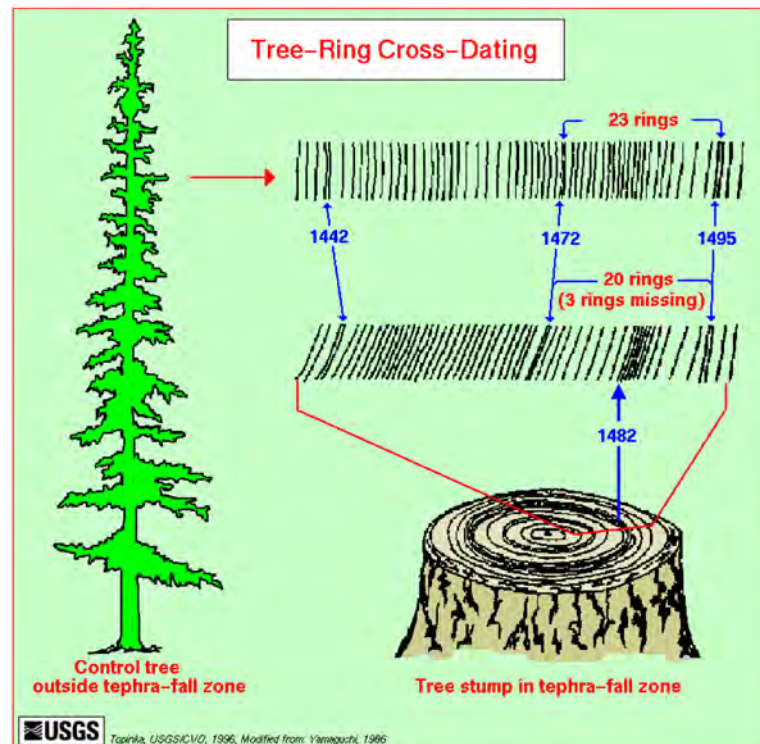
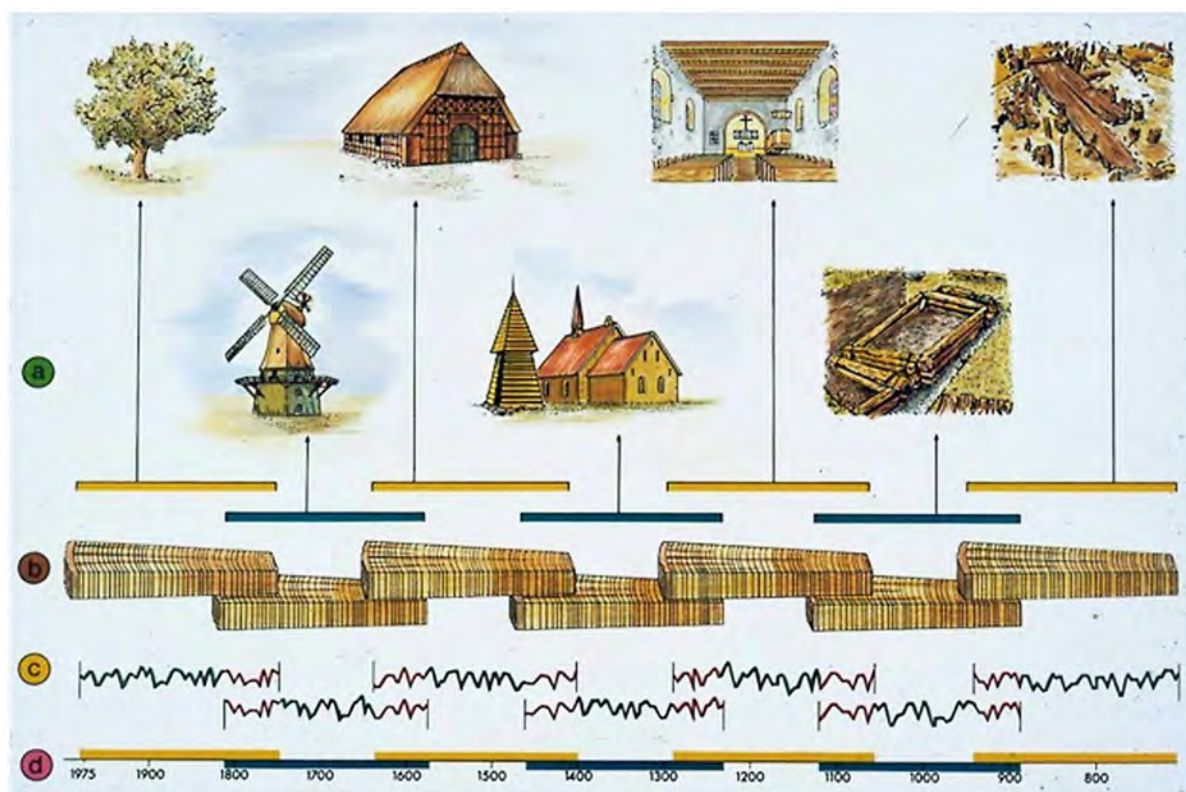


Figure 12. Cross-dating according to the annual ring growth of different samples from live standing trees to old or new wood samples, Illustration showing an example of time series crossdating
(Adapted from Dendrologisches Labor Hamburg, 2015, <http://www.scinexx.de/dossier-detail-186-7.html>)



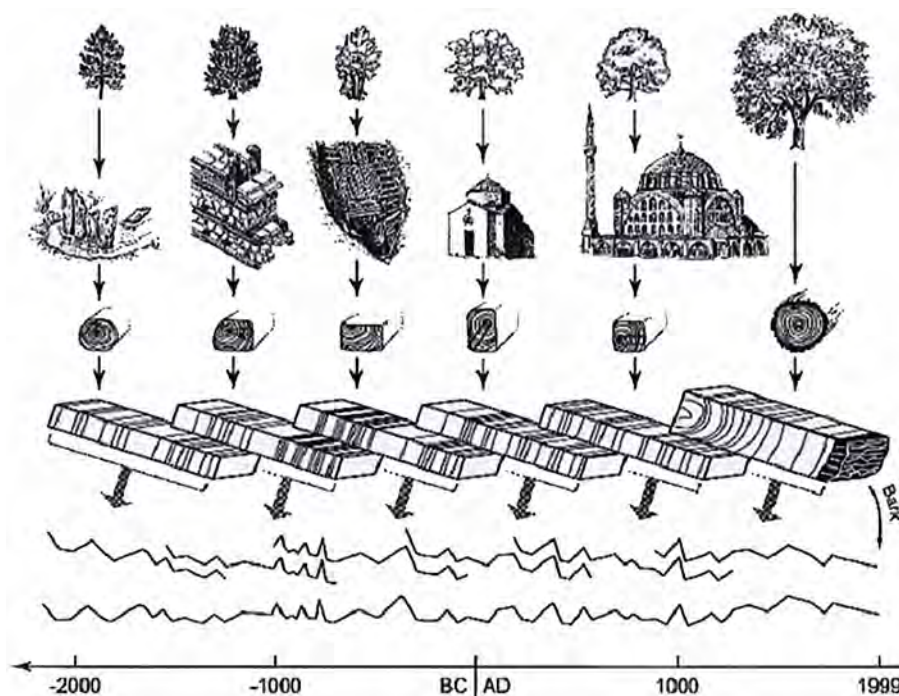
The techniques of dendrochronological study were used to date a spruce coffin board from Pukatawagan Bay, Manitoba received from Manitoba Historical Resources. The sample contained 74 annual rings, although the outermost 16 rings were rotten and not measured. The ring width series from the coffin board was matched against records from living spruce growing near South Indian Lake. Crossdating shows that the coffin board was cut in 1878. The board also contains a ring containing poorly developed latewood (a 'light' ring) in 1817 that is found commonly in spruce records across northern Manitoba (Figure 12)(Scott & Nielsen, 2002).

Using different wood materials of different periods, by cross-dating of the samples comparing with control one, it would be possible to record

and estimate changes from present time back to the ancient area (Figure 13).

Dr. Andrew E. Douglass, an astronomer, developed dendrochronology about 1913. Douglass used a bridging method to create his chronology. First he studied recently cut trees whose dates he knew. This initial step was critical because by knowing the cut date, Douglass knew when each tree added its last growth ring. This, in turn, let him determine the year each tree started growing. The calculation was straightforward: count the dark rings inward and subtract that number from the year the tree was cut. As Douglass matched and recorded ring patterns from trees of different ages, he confirmed that their patterns overlapped during the years the trees simultaneously lived (The University of North Carolina, 2013).

Figure 13. The bridging method, cross-dating of annual rings from present live tree to the past wood samples, for establishing a tree-ring sequence (Kuniholm, 2001)



FUTURE RESEARCH DIRECTIONS

Enhance and improvement of measuring methods on climate change issues in different sector is a crucial and important needs. Dendrochronology study method has several different applications. It is crucial to used dendrochronological method in medical field. Some question need to be considered and if possible answered. How medical needs could link with dendrochronological experiences? What kind of element may be detected on tree rings? How people could benefit of the results? What are the best analysis methods? By using this method we will be able to extend our work in medicine science areas and speed up medical approach with lower cost and economical saving (Jafari, 2014a).

CONCLUSION

The AFOLU sector (AR5, IPCC, WGIII, Chapter 11) is responsible for just under a quarter (~10–12 GtCO₂eq/yr) of anthropogenic GHG emissions mainly from deforestation and agricultural emissions from livestock, soil and nutrient management (robust evidence; high agreement) (Smith *et. al.*, 2014).

Climate change impact will cause changes in biomass production in natural ecosystems. It is a need to consider the vulnerability of Net Primary production (NPP) in forest ecosystem (Jafari, 2014b).

Climate change issue is an important subject in current century. In all possible ways we need to cope with this phenomenon to enhance our understanding knowledge. Dendrochronology as an able and certified study method could be implying in a wide range of applications.

REFERENCES

- Anonymous. (2008a). *Plant Growth, in Science*. Retrieved from http://www.steve.gb.com/images/science/wood_layers.jpg
- Anonymous. (2008b). *Climate Change, Past and Future, Calspace Courses*. Retrieved from <http://earthguide.ucsd.edu/virtualmuseum/climatechange1/cc1syllabus.shtml>
- aric. (2008). *Global Climate Change Student Guide*. Author.
- BETA. (2013). *Beta Analytic, Radiocarbon Dating*. Retrieved from <http://www.radiocarbon.com/tree-ring-calibration.htm>
- Bradley, R. S. (1985). *Quaternary Palaeoclimatology: Methods of Palaeoclimatic Reconstruction*. London: Unwin Hyman.
- Brantley, S., Yamaguchi, D., Cameron, K., & Pringle, P. (1986). Tree-Ring Dating of Volcanic Deposits. *Earthquakes and Volcanoes*, 18(5), 184–194.
- Carrer, M. (2011). Individualistic and Time-Varying Tree-Ring Growth to Climate Sensitivity. *PLoS ONE*, 6(7), e22813. doi:10.1371/journal.pone.0022813 PMID:21829523
- FAO. (2009). *Guidelines for good forestry and range practices in arid and semi-arid zones of the Near East*. Cairo: FAO Regional Office for the Near East.
- García-Suárez, A.M., Butler, C.J., & Baillie, M.G.L. (2009). Climate signal in tree-ring chronologies in a temperate climate: A multi-species approach. *Dendrochronologia*, 27, 183–198. doi:10.1016/j.dendro.2009.05.003
- Hammarlund, D. (2013). *Equipment and sample handling*. Retrieved from <http://www.geol.lu.se/dendro/en/equipment.htm>
- Hughes, M., & Touchan, R. (1997). *Preparation for Collaborative Work on Past Climate in Jordan*. Science Storm, A world of wonder. Retrieved from <http://www.sciencestorm.com/>
- Indiana State University. (2013). *Dendrochronology Lab, Indiana State University*. Retrieved from http://dendrolab.indstate.edu/equip_fac.htm
- Jafari, M. (2007). *Review on needfulness for plant ecophysiological study and investigation on climate change's effects on forest, rangeland and desert ecosystems*. Presented at Workshop, Climate Change in South-Eastern European Countries: Causes, Impacts, Solutions, Orange-rie, Burggarten, Graz, Austria. Retrieved from http://www.intecol.net/pages/002_personal.php?mode=list&table=blog&tb_kind=Research&id=mjafari&vvt=aa&bidx=27531
- Jafari, M. (2010). Climate Change Impacts on Iranian Ecosystems with review on Climate Change Study Methods (Pub. No. 421). RIFR.
- Jafari, M. (2012a). A new approach to dendro-ecological studies: Climate change impact on forest' wood production in Astara (Gilan). *Iranian Journal of Wood and Paper Science Research*, 27(4), 690–706.
- Jafari, M. (2013a). *New method on dendro-chronology study in forest ecosystems; two new approaches: Dendro-productivity and Dendro-genetic*. Paper ID: 1076-ADA2013. Presented in The 3rd International Conference of Asian Dendrochronology Association (ADA2013), Climate Change and tree Rings, Tehran, Iran. Retrieved from <http://ada2013.ut.ac.ir/paper?manu=7148>
- Jafari, M. (2013b). Integration of forestry and rangeland management institutions in the Islamic Republic of Iran. In *Challenges in adopting an integrated approach to managing forest and rangelands in the Near East Region, Part II, Country studies* (pp. 37–67). Cairo: FAO.
- Jafari, M. (2013c, July-December). Challenges in Climate Change and Environmental Crisis: Impacts of Aviation Industry on Human. *Urban and Natural Environments in International Journal of Space Technology Management and Innovation*, 3(2), 24–46.

- Jafari, M. (2014a). *Dendro-Medical as New Approach on Dendrochronological Study method*. Presented in the 9th International Conference on Dendrochronology, as Poster No. PO086, Ref. 1590, ThemeZ01, Melbourne, Australia.
- Jafari, M. (2014b). Change and vulnerability of Net Primary Production (NPP) in Iranian forest, rangeland and desert ecosystems impacted by climate change. *Iranian Journal of Range and Desert Research*, 21(1), 139-153.
- Jafari, M. (2015). Dendrochronology and Climate Change. In *Encyclopedia of Information Science and Technology* (3rd ed., pp. 2917–2930). IGI Global.
- Jafari, M. & Khorankeh S. (2013). Impact of climate and environmental changes on forest ecosystem's productivity (case study: Galugah). *Iranian Journal of Forest and Poplar Research*, 21(1), 166-183.
- Jafari, M., Maghsudloo, M. K., Khanjani Shiraz, B., Karimidoost, A., & Hemmati, A. (2012b). New approach in dendro-genetic method for identification of beech tree variability. *Iranian Journal of Rangelands and Forests Plant Breeding and Genetic Research*, 20(2), 284–294.
- Kiaee, M., & Jafari, M. (2014). Investigation and consideration of forest tree reaction to climate and environmental changes (Case study: Lavizan forest park). *Journal of Plant Research*, 27(1), 130–141.
- Kuniholm, P. I. (2001). *Dendrochronology and Other Applications of Tree-ring Studies in Archaeology*. In *The Handbook of Archaeological Sciences*. London: John Wiley & Sons, Ltd.
- McCarthy, B. C. (2008). *Introduction to Dendrochronology, Lab procedures*. Retrieved from <http://www.plantbio.ohiou.edu/dendro/>
- RMTRR. (2008). *Rocky Mountain Tree-Ring Research (RMTRR)*. Retrieved from <http://www.rmtrr.org/gallery.html>
- Robinson, W. J., & Evans, R. (1980). A microcomputer-based tree-ring measuring system. *Tree-Ring Bulletin*, 40, 59–64.
- Schweingruber, F. H. (1988). *Tree Rings: Basics and Applications of Dendrochronology*. Dordrecht: Reidel; doi:10.1007/978-94-009-1273-1
- Scott, St. G., & Nielsen, E. (2002). Dendrochronological analysis of a coffin board recovered from Pukatawagan, Bay, Manitoba. Terrain Sciences Division, Geological Survey of Canada and Manitoba Geological Survey, Manitoba Industry, Trade and Mines.
- Sheppard, P. R. (2013). *Crossdating Tree Rings*. Laboratory of Tree-Ring Research, The University of Arizona. Retrieved from <http://www.ltrr.arizona.edu/skeletonplot/applications.htm>
- Smith, P., Bustamante, M., Ahammad, H., Clark, H., Dong, H., Elsiddig, E. A., & Tubiello, F. et al. (2014). Agriculture, Forestry and Other Land Use (AFOLU). In *Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, UK: Cambridge University Press.
- Swetnam Thomas, W., & Christopher, H. B. (2002). Fire Regimes in Sierran Mixed-Conifer Forests. In *Sierra Nature Notes*, (vol. 2). Laboratory of Tree-Ring Research, The University of Arizona.
- TGTC. (2008). *Tree growth and the tree-site complex*. TGTC.
- The Ohio State University. (2013). *Woody plants*. Department of Horticulture & Crop Science, The Ohio State University. Retrieved from <http://www.hcs.ohio-state.edu/hcs300/anat2.htm>, <http://hcs.osu.edu/>
- The University of Arizona. (2013). *Paleo Slide Set: Tree Rings: Ancient Chronicles of Environmental Change*. Laboratory of Tree-Ring Research, The University of Arizona.

The University of North Carolina. (2013). *Tree-Ring Dating, Lesson 2.4*. Research Laboratories of Archeology. The University of North Carolina at Chapel Hill. Retrieved from <http://rla.unc.edu/lessons/Lesson/L204/L204.htm>

WDC. (2008). *The World Data Center for Paleoclimatology Mirror Site (WDC)*. Retrieved from <http://wdc.obs-mip.fr/treering.html>

ADDITIONAL READING

Barnola, J. M., Raynaud, D., Korotkevich, Y. S., & Lorius, C. (1987). Vostok ice core provides 160,000 year record of atmospheric CO₂. *Nature*, 329(6138), 408–414. doi:10.1038/329408a0

Bergthorsson, P. (1969). An estimate of drift ice and temperature in Iceland in 1000 years. *Jökull*, 19, 94–101.

Briffa, K. R., & Schweingruber, F. H. (1992). Recent dendroclimatic evidence of northern and central European summer temperatures. In: *Climate Since A.D.1500*. Routledge, London, pp. 366–392. COHMAP Members, 1988. Climatic change of the last 18,000 years: Observations and model simulations. *Science*, 241, 1043–1052.

Davis, M. B. (1963). On the theory of pollen analysis. *American Journal of Science*, 261(10), 899–912. doi:10.2475/ajs.261.10.897

Emmanuel, W. R., Shugart, H. H., & Stevenson, M. P. (1985). Climate change and the broad-scale distribution of terrestrial ecosystem complexes. *Climatic Change*, 7(1), 29–43. doi:10.1007/BF00139439

Fritts, H. C. (1971). Dendroclimatology and dendroecology. *Quaternary Research*, 1(4), 419–449. doi:10.1016/0033-5894(71)90057-3

Fritts, H. C. (1976). *Tree rings and climate* (pp. 80–96). London: Academic Press.

Schweingruber, F. H., Fritts, H. C., Bräker, O. U., Drew, L. G., & Schär, E. (1978). The X-ray technique as applied to dendroclimatology. *Tree Ring Bull.*, 38, 61–91.

KEY TERMS AND DEFINITIONS

AFOLU: Agriculture, Forestry and Other Land Use (AFOLU), main players on emission reduction and mitigation aspect of climate change, chapter 11, WGIII, IPCC AR5 2014.

Climate Change: Global Warming, climate change and climate variability are a definition of deviation of climatic factors from its normal trends mainly impacted by human activities. Global Warming and its' effects on natural resources, plants, animal and in general on human life are among subjects that received attention of scientists and politicians in recent years.

Dendrochronology: Dendrochronology was developed about 1913, and is a (climate change) method to study tree ring widths in terms of time. The word dendrochronology is composed of: dendro (using trees, or more specifically the growth rings of trees), chrono (time, or more specifically events in past time), and logy (the study of). Dendrochronology as an able and certified study method could be implying in a wide range of applications.

Proxy Data: Climate change assessments and evaluation should be done by using recorded observation data as well as prepared and provided proxy data. Paleoclimatologists (climatologists who study past – or paleo – climates) use the term “proxy” to describe a way that climate change is recorded in nature, within geological materials such as ocean or lake sediments, tree-rings, coral growth-bands, ice-cores, and cave deposits.

Tree Rings: Trees growing in temperate climatic region are under seasonal changes. In spring and summer time plants grow better than unpleasant seasons like fall and winter. Each growth layer is distinguished by early wood (springwood), composed of large thin-walled cells produced during the spring when water is usually abundant, and the denser latewood (summerwood), and composed of small cells with thick walls.

Model for Assessment of Environmental Responsibility in Health Care Organizations

E

María Carmen Carnero

University of Castilla-La Mancha, Spain & University of Lisbon, Portugal

INTRODUCTION

Sustainability is considered a paradigm for businesses in the 21st Century (Garcia et al., 2016). Despite this, the existing tools for helping to introduce strategies and manage activities to promote sustainable business are few (Garcia et al., 2016). These deficiencies become more important in Health Care Organizations owing to its particular conditions of resource consumption and waste production. Health Care Organizations are the only type of company which can generate all the classes of waste, from waste without risk to waste that is potentially infectious, carcinogenic, mutagenic, teratogenic or radioactive. The risk to people and to the environment from this waste is much greater if it is not correctly segregated. It is also vital to carry out action to reduce the consumption of limited natural resources such as water and energy, while increasing the protection and conservation of the environment, including reducing the emission of pollutant gases, protecting biodiversity or considering the role of suppliers in action to prevent or reduce waste.

It is, therefore, essential to have objective tools to assist in monitoring environmental sustainability in this type of organization, taking into account a number of factors. That is, by assessing how improvement actions, within a process of continuous improvement, are contributing to improvements in sustainability. However, it is clear that there is little linkage between sustainability reporting and management control systems (Cintra & Carter, 2012).

Nonetheless, despite its importance, the literature on the development of systems for environmental assessment in Health Care Organizations is very limited.

This Chapter therefore sets out a multicriteria assessment system constructed by extension to a fuzzy environment of the Technique for Order Preference by Similarity to Ideal Situation (TOPSIS), to assess the environmental responsibility of a Health Care Organization. This model allows joint evaluation of a significant number of decision criteria, which include any event that may cause adverse effects on water, ground, seas and rivers, wild species or their habitats; it also considers the existence of possible measures to be carried out in Health Care Organizations to minimize the probability of an event, or to eliminate all risk. However, it should be noted that this model is not intended to perform an environmental audit in the field of health care, as it would need to include economic, technical, legal and other criteria, or a system of environmental impact that would require the assessment of a variety of risks and consequences. The aim is to provide a hospital with a model which is easy to apply, with criteria specific to health care, and which allows its responsibility with regard to the environment to be monitored over time. Following the methodology laid down in Carnero (2015), criteria were used that were assessed depending on the number of admissions or annual services provided, making it possible to compare results over time for a single organization, or between organizations. The model has been used in a Public Hospital.

DOI: 10.4018/978-1-5225-2255-3.ch273

BACKGROUND

The literature includes a large number of contributions on environmental questions (Aragones-Beltran et al., 2009; Higgs et al., 2008; Hsu & Hu, 2008; Kang et al., 2007; Lamelas et al., 2008; Liang et al. 2006; Madu, Kuei & Madu, 2002; Pilavachi, Chatzipanagi, & Spyropoulou, 2009; Tzeng & Lin, 2005; Tseng, Lin, & Chiu, 2009; Van Calker et al., 2006). However, these are invariably related to manufacturing, transport or energy companies. In the case of service companies and, in particular, in Health Care Organizations, the contributions are practically non-existent (Carnero, 2015).

Health Care Organizations, places dedicated to the improvement and development of preventive measures in health care, with respect to their users, those who live in the area and their workers, should be involved in minimizing their own environmental impact, as there is a strong correlation between the two (Comunidad de Madrid, 2005). In order to improve environmental sustainability of a Health Care Organization, however, it is vital to monitor sustainability over time for decision making and management of activities that constitute an organization's system processes (Salvado et al., 2015).

A system of environmental assessment should combine many factors, which may be technical, social, political, economic and environmental, which often conflict with one another (Lahdelma et al., 2000); it may also be necessary to include a number of individuals or decision groups, with different perspectives or responsibilities within the Health Care Organization; as well as the need to incorporate a great deal of information, quantitative but in many cases qualitative, relating to uncertainties, scenarios, goals, etc. (Munda, 2005). These characteristics make the use of Multi-Criteria Decision Analysis (MCDA) methods highly suitable for supporting decisions about sustainability (Santoyo-Castelazo & Azapagic, 2014). The fact that the model produced, based on mathematical techniques, is objective, also helps to guarantee public acceptance of the solution or

result obtained (Huang, Keisler, & Linkov, 2011). MCDA methods then, although not suited to all environmental problems, are very convenient in environmental impact assessment, as they give information in a structured fashion, which can be easily interpreted by the decision makers (Neste & Karjalainen, 2013). All this has led to an increase in the literature applying MCDA in the environmental field over the last two decades (Carnero, 2014).

Because of the pressure brought to bear on companies by different stakeholders and by society to address ecological and social sustainability (Garcia et al., 2016) a variety of research has been carried out applying MCDA methods in this area. Gumus (2009) presents a methodology for selection of the most suitable hazardous waste transportation firms using a fuzzy Analytical Hierarchy Process (AHP). A similar technique is used by Heo et al. (2010) to get the weightings of the criteria to establish ex-ante and ex-post stages of renewable energy dissemination programmes in Korea. Reza et al. (2011) combine morphological analysis and AHP to choose new sustainable products from the earliest stages of conception. The study of Chan et al. (2012) is along similar lines, but using fuzzy AHP. Boran et al. (2012) use intuitionistic fuzzy TOPSIS to assess renewable energy technologies for electricity generation, such as photovoltaic, hydro, wind, and geothermal energy in Turkey. Wang et al. (2012) produce a model for selecting of green initiatives in the fashion industry. Vinodh et al. (2014) describe an assessment model to determine the best method for recycling plastics. Pourebrahim et al. (2014) made a selection of criteria and alternatives for conservation development in a coastal zone. Galvez et al. (2015) propose a model combining Mixed Integer Linear Programming optimization and AHP to assess possible scenarios for the implementation of an anaerobic co-digestion logistics network used to create sustainable energy production processes from biogas. Al Garni et al. (2016) use AHP to assess renewable power generation sources including solar photovoltaic,

concentrated solar power, wind energy, biomass, and geothermal, finding that in the case of Saudi Arabia the photovoltaic, followed by concentrated solar power are the first-placed technologies.

The literature which uses MCDA methods for the assessment of environmental sustainability and environmental impact includes the following contributions. Hermann et al. (2007) describe the tool COMPLIMENT, which allows the overall environmental impact of an organization to be found; they combine life cycle assessment, AHP and Environmental Performance Indicators (EPIs). Kaya and Kahraman (2011) use fuzzy AHP to build an environmental impact assessment methodology for urban industrial planning. Viaggi et al. (2011) develop and apply a multicriteria methodology to estimate the environmental effectiveness of European Union agri-environment schemes in Ireland and Emilia-Romagna (Italy). Larimian et al. (2013) use fuzzy AHP to assess environmental sustainability from the point of view of security in different areas in a region of Tehran. Egilmez et al. (2015) assess environmental sustainability in 27 Canadian and US cities using fuzzy multicriteria decision-making models. The research of Zhang et al. (2016) is also related to City sustainability evaluation via MCDA methods, in this case, of 13 cities in China. Khalili and Duecker (2013) describe a methodology for designing a sustainable environmental management system built using ELECTRE. This system serves as a back-up to monitor the efforts made by companies in the area of sustainability, for example through product design, operational development or the modelling of the supply chain. Salvado et al. (2015) use AHP to calculate a sustainability index which allows companies and their supply chains to get information about their own level of economic, social and environmental sustainability. A further review of the literature on environmental questions analysed via multicriteria techniques can be seen in Huang et al. (2011) and Neste and Karjalainen (2013).

However, in the field of environmental assessment of Health Care Organizations, the only contribution is Carnero (2014) which describes

a model using a fuzzy AHP together with utility functions. This model is applied to a recently opened public hospital, giving a utility of 0.764 out of 1, and showing how this type of model can be very positive in the process for certification to standard ISO 14001. Carnero (2015) shows another, more advanced, model also using fuzzy AHP, but assessing new criteria, and most are assessed based on number of care services provided annually, which allows results to be compared over time for one Health Care Organization, or between Health Care Organizations.

MODEL FOR ASSESSMENT OF ENVIRONMENTAL RESPONSIBILITY

The development of new information technologies, together with the development and application of new concepts in environmental sustainability, require a constant updating of the assessment models in this area.

The choice of criteria and subcriteria takes account of Rodríguez et al. (2005), García et al. 2010, Mata et al. (2011), Tejedor (2012), Bon-García (2012), Yanguas (2012), Galdakao-Usansolo Hospital (2012), Carnero (2014) and Carnero (2015). The criteria and subcriteria used in the multicriteria model are:

- Annual water consumption (C1).
- Annual energy consumption. Two subcriteria are considered:
 - Annual consumption (MW/h) by the Hospital of electricity, refrigerating energy, thermal energy and natural gas (C2).
 - Consumption of renewable energies (C3).
- Environmental accidents and incidents (C4). Accidents have potentially serious environmental implications, such as uncontrolled spillages into the water supply, fire, x-ray emissions, spillage of dangerous substances on the ground, leaks or

spillages of natural gas or diesel, spillage of acetylene or refrigerating gas, leaks of ethylene oxide, etc. Incidents are matters that give rise to internal non-compliance with standard ISO 14001 and require an analysis of the cause and corrective action. Although they are not as serious as environmental accidents, they should be analysed and their causes eliminated as quickly as possible. For example, unsuitable storage of waste or chemicals, complaints from patients or neighbours about noise, etc.

- Biodiversity (C5). An assessment is made of the capacity of the organization to adapt to the rural and forest environment, to take care of endangered species, and also of action taken to continuously improve the environmental impact of the organization.
- Activities to promote and spread the environmental message (C6). This includes annual planning of congresses, celebration, exhibitions, promotion of activities related to environmental prevention and protection, spreading of awareness of environmental aims of the Health Care Organization, etc.
- Training and cooperation on environmental matters (C7). The organization is assessed with respect to training programmes among care and non-care staff in the Health Care Organization, and planning of groups to analyse problems and improvements. This also considers the existence of surveys and systems for gathering complaints and suggestions, and for dealing with them efficiently.
- Noise inside and outside. This criterion is made up of the subcriteria:
 - Noise inside the Hospital (C8). This should not exceed 30dBA, and optimum noise is between 15 and 25 dBA.
 - Noise near the Hospital (C9). This should not exceed 55dBA, and optimum noise is between 35 and 45dBA.

- Waste production. Assess annual waste production. It is divided into the subcriteria:
 - Group I waste (C10). This is general waste, with no risk, such as edible oils, plastics, paper and cardboard, clothes, glass, etc.
 - Group II waste (C11). Sanitary waste that may be treated as urban waste.
 - Group III waste (C12). Dangerous waste products, including industrial oil, batteries, non-halogenated solvents, chemical waste, radiology liquid, out-of-date or retired medicines, anatomical remains with formaldehyde, cytostatic waste, etc.

Other waste, in other groups, including radioactive waste, is subject to special legislation which must be complied with, and so is not included in the assessment system.

- Green purchasing (C13). Inclusion in conditions for purchasing products and contracting services, of guarantees of respectful treatment of the environment. Assessment of suppliers with regard to certification with standard ISO 14001. Minimizing consumption of paper, cardboard and plastic in transactions and packages, etc.

All the criteria and/or subcriteria are assessed with respect to the number of annual admissions or services provided by the Health Care Organization. This means that the results can be compared over time for a single organization, and comparisons can also be made between Health Care Organizations (Carnero, 2015).

Although different multicriteria techniques can be used to build the model, in this case, applying fuzzy TOPSIS allows simultaneous valuation of a significant number of criteria.

A fuzzy MCDA will be used, as this allows the uncertainty, ambiguous situations or vagueness of the judgements of the decision makers to be

included (SeongKon et al., 2011). Furthermore, decision makers usually feel more confident in giving interval judgements rather than fixed value judgments (Isaai et al., 2011).

A triangular fuzzy number $\tilde{a} = (l, m, u)$ is defined by the membership function $\mu_{\tilde{a}}(x)$ which satisfies the conditions of normality and convexity. l , m and u are real numbers which satisfy $l \leq m \leq u$. The membership function is defined in Equation (1) (Chang, 1996).

$$\mu_{\tilde{a}}(x) = \begin{cases} \frac{x-l}{m-l} & x \in [l, m] \\ \frac{u-x}{u-m} & x \in [m, u] \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

Kaufmann and Gupta (1988) give the main algebraic operations of triangular fuzzy numbers $\tilde{A} = (l, m, u)$ and $\tilde{B} = (l, m, u)$:

$$(l_1, m_1, u_1) \oplus (l_2, m_2, u_2) = (l_1 + l_2, m_1 + m_2, u_1 + u_2) \quad (2)$$

$$(l_1, m_1, u_1) \ominus (l_2, m_2, u_2) = (l_1 - l_2, m_1 - m_2, u_1 - u_2) \quad (3)$$

$$(l_1, m_1, u_1) \otimes (l_2, m_2, u_2) \approx (l_1 l_2, m_1 m_2, u_1 u_2) \quad (4)$$

$$(l_1, m_1, u_1) \oslash (l_2, m_2, u_2) \approx (l_1 / l_2, m_1 / m_2, u_1 / u_2) \quad (5)$$

$$(l_1, m_1, u_1)^{-1} \approx (1 / u_1, 1 / m_1, 1 / l_1) \quad \text{for } l, m, u > 0 \quad (6)$$

$$\lambda \otimes (l_1, m_1, u_1) \approx (\lambda l_1, \lambda m_1, \lambda u_1), \quad \lambda > 0, \lambda \in \mathbb{R}^+ \quad (7)$$

Fuzzy TOPSIS (FTOPSIS) was proposed by Chen (2000). In a decision problem with criteria $(C_1, C_2, \dots, C_n)$ and alternatives $(A_1, A_2, \dots, A_m)$, the best alternative in FTOPSIS is such that should

have the shortest distance to a fuzzy positive ideal solution (FPIS) and the farthest distance from a fuzzy negative ideal solution (FNIS). The FPIS is calculated using the best performance values for each criterion and the FNIS looks at the worst performance values.

In FTOPSIS the decision makers use linguistic variables to obtain the weightings of the criteria and the ratings of the alternatives. If there is a decision group made up of K individuals, the fuzzy weight and rating of the k th decision maker with respect to the i th alternative in the j th criterion are respectively:

$$\tilde{w}_j^k = (w_{j1}^k, w_{j2}^k, w_{j3}^k) \quad (8)$$

$$\tilde{x}_{ij}^k = (a_{ij}^k, b_{ij}^k, c_{ij}^k) \quad (9)$$

where $i = 1, 2, \dots, m$ and $j = 1, 2, \dots, n$.

The aggregate fuzzy weights $\tilde{w}_{ij}$ of each criterion given by K decision makers are calculated using Equation (10).

$$\tilde{w}_j = \frac{1}{K} \otimes (\tilde{w}_j^1 \oplus \tilde{w}_j^2 \oplus \dots \oplus \tilde{w}_j^K) \quad (10)$$

While to calculate the aggregate ratings of the alternatives Equation (11) is used (Ouma, Opudo, & Nyambenya, 2015).

$$\tilde{x}_{ij} = \frac{1}{K} \otimes (\tilde{x}_{ij}^1 \oplus \tilde{x}_{ij}^2 \oplus \dots \oplus \tilde{x}_{ij}^K) \quad (11)$$

A fuzzy multicriteria group decision-making problem which can be expressed in matrix format is shown in Equation (12) (Chen, 2000).

$$D = \begin{bmatrix} \tilde{x}_{11} & \tilde{x}_{12} & \dots & \tilde{x}_{1n} \\ \tilde{x}_{21} & \tilde{x}_{22} & \dots & \tilde{x}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{x}_{m1} & \tilde{x}_{m2} & \dots & \tilde{x}_{mn} \end{bmatrix}$$

$$\tilde{W} = (\tilde{w}_1, \tilde{w}_2, \dots, \tilde{w}_n) \quad (12)$$

where $\tilde{w}_j$ and $\tilde{x}_{ij}$ are linguistic variables which can be described by triangular fuzzy numbers.

The weightings of the criteria can be calculated by assigning directly the linguistic variables shown in Table 1. The ratings of the alternatives are found using the linguistic variables of Table 2 (Chen, 2000).

The linear scale transformation is used to transform the various criteria scales into a comparable scale. And thus we obtain the normalized fuzzy decision matrix $\tilde{R}$ (Rodrigues, Osiro, & Ribeiro, 2014).

$$\tilde{R} = [\tilde{r}_{ij}]_{m \times n} \quad i = 1, 2, \dots, m; j = 1, 2, \dots, n. \quad (13)$$

where

$$\tilde{r}_{ij} = \left(\frac{l_{ij}}{u_j^+}, \frac{m_{ij}}{u_j^+}, \frac{u_{ij}}{u_j^+} \right) \text{ and } u_j^+ = \max_i u_{ij} \text{ in the case of benefit criteria type}$$

$$\tilde{r}_{ij} = \left(\frac{l_j^-}{u_{ij}}, \frac{l_j^-}{m_{ij}}, \frac{l_j^-}{l_{ij}} \right) \text{ and } l_j^- = \max_i l_{ij} \text{ in the case of cost criteria type}$$

Next, the weighted normalized decision matrix, $\tilde{V}$ is calculated, by multiplying the weightings of the criteria $\tilde{w}_j$, by the elements $\tilde{r}_{ij}$ of the normalized fuzzy decision matrix.

Table 1. Linguistic variables for the weights

Linguistic Variables for the Weights	Fuzzy Number
Very Low (VL)	(0, 0, 0.1)
Low (L)	(0, 0.1, 0.3)
Medium Low (ML)	(0.1, 0.3, 0.5)
Medium (M)	(0.3, 0.5, 0.7)
Medium High (MH)	(0.5, 0.7, 0.9)
High (H)	(0.7, 0.9, 1.0)
Very High (VH)	(0.9, 1.0, 1.0)

(Chen, 2000)

Table 2. Linguistic variables for the ratings

Linguistic Variables for the Ratings	Fuzzy Number
Very Poor (VP)	(0,0,1)
Poor (P)	(0,1,3)
Medium Poor (MP)	(1,3,5)
Fair (F)	(3,5,7)
Medium Good (MG)	(5,7,9)
Good (G)	(7,9,10)
Very Good (VG)	(9,10,10)

(Chen, 2000)

$$\tilde{V} = [\tilde{v}_{ij}]_{m \times n} \quad \text{donde } \tilde{v}_{ij} = \tilde{x}_{ij} \otimes \tilde{w}_j \quad (14)$$

The distances d_i^+ and d_i^- of each weighted alternative from the FPIS and FNIS are calculated using Equations (15) and (16).

$$d_i^+ = \sum_{j=1}^n d_{\nu}(\tilde{v}_{ij}, \nu_{ij}^+) \quad (15)$$

$$d_i^- = \sum_{j=1}^n d_{\nu}(\tilde{v}_{ij}, \nu_{ij}^-) \quad (16)$$

where $d_{\nu}(\tilde{a}, \tilde{b})$ is the distance measured between the fuzzy numbers $\tilde{a}$ and $\tilde{b}$. This distance is calculated from Equation (17) (Rodrigues, Osiro, & Ribeiro, 2014).

$$d_{\nu}(\tilde{a}, \tilde{b}) = \sqrt{\frac{1}{3}[(l_a - l_b)^2 + (m_a - m_b)^2 + (u_a - u_b)^2]} \quad (17)$$

Finally, the closeness coefficient, CC_i , is calculated for each alternative i using Equation (18). This parameter allows the degree of fuzzy satisfaction to be evaluated for each Health Care Organization.

$$CC_i = \frac{d_i^-}{d_i^- + d_i^+} \quad (18)$$

SOLUTIONS AND RECOMMENDATIONS

The model was applied to a Health Care Organization whose mission is working to improve the health of the people it serves, with quality, safety and sustainability. It has 30 medical specialties, 386 beds and 1,744 staff, and provides medical cover to an area with 300,000 inhabitants.

An expert in environmental matters was used to obtain a pairwise comparison matrix for the decision criteria. The decision maker was asked to evaluate the importance of the criteria or sub-criteria that could be assessed by applying the fuzzy scale set out in Table 1. The resulting weightings are shown in Table 3. Next, the procedure is applied to evaluate each alternative (different years to be assessed in the Health Care Organization) via the linguistic variables from Table 2. This gives the fuzzy weighted normalized decision matrix shown in Table 4. The distances d_i^+ and d_i^- of each weighted alternative from the FPIS and FNIS are calculated using Equations (15) and (16), giving the results shown in Table 5.

The closeness coefficient of each alternative is set out in Table 6. From CC the ranking of the three alternatives is Hospital year 3, year 2 and year 1. The best result is obtained by the hospital in the third year assessed. This shows the process of continuous improvement undertaken by the Health Care Organization and allows a global follow-up to be carried out annually.

Table 3. Fuzzy weights

Criteria/Subcriteria	Weightings
C1	(0.900, 1.000, 1.000)
C2	(0.900, 1.000, 1.000)
C3	(0.300, 0.500, 0.700)
C4	(0.900, 1.000, 1.000)
C5	(0.700, 0.900, 1.000)
C6	(0.100, 0.300, 0.500)
C7	(0.300, 0.500, 0.700)
C8	(0.300, 0.500, 0.700)
C9	(0.100, 0.300, 0.500)
C10	(0.100, 0.300, 0.500)
C11	(0.300, 0.500, 0.700)
C12	(0.900, 1.000, 1.000)
C13	(0.300, 0.500, 0.700)

(Created by the author)

FUTURE RESEARCH DIRECTIONS

AHP and fuzzy AHP are the most widely-used multicriteria techniques in the literature in general, and specifically in relation to the environment. However, this Chapter uses the fuzzy TOPSIS technique, due to the ease with which it can assess a large number of criteria and alternatives, but it would be appropriate to validate the environmental assessment models using other techniques successfully applied in a significant number of real cases, such as the Measuring Attractiveness by a Categorical Based Evaluation Technique (MACBETH).

Table 4. The fuzzy weighted normalized decision matrix

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13
Hospital year 1	(0.630, 0.900, 1.000)	(0.630, 0.900, 1.000)	(0.210, 0.450, 0.700)	(0.300, 0.556, 0.778)	(0.350, 0.630, 0.900)	(0.050, 0.210, 0.450)	(0.090, 0.250, 0.490)	(0.129, 0.357, 0.700)	(0.043, 0.214, 0.500)	(0.090, 0.300, 0.500)	(0.270, 0.500, 0.700)	(0.630, 0.900, 1.000)	(0.150, 0.350, 0.630)
Hospital year 2	(0.810, 1.000, 1.000)	(0.810, 1.000, 1.000)	(0.270, 0.500, 0.700)	(0.300, 0.556, 0.778)	(0.490, 0.810, 1.000)	(0.070, 0.270, 0.500)	(0.270, 0.500, 0.700)	(0.129, 0.357, 0.700)	(0.043, 0.214, 0.500)	(0.090, 0.300, 0.500)	(0.270, 0.500, 0.700)	(0.810, 1.000, 1.000)	(0.210, 0.450, 0.700)
Hospital year 3	(0.810, 1.000, 1.000)	(0.810, 1.000, 1.000)	(0.270, 0.500, 0.700)	(0.500, 0.778, 1.000)	(0.350, 0.630, 0.900)	(0.090, 0.300, 0.500)	(0.270, 0.500, 0.700)	(0.129, 0.357, 0.700)	(0.043, 0.214, 0.500)	(0.090, 0.300, 0.500)	(0.270, 0.500, 0.700)	(0.810, 1.000, 1.000)	(0.270, 0.500, 0.700)

(Created by the author)

Table 5. The distance measurement

Alternatives	d_i^+	d_i^-
Hospital year 1	7.035	6.984
Hospital year 2	6.223	7.700
Hospital year 3	6.104	7.813

(Created by the author)

Due to the constant evolution of information communication systems and the development of new technologies, with very dramatic effects on improvements in energy efficiency and consumption of natural resources, this model should be periodically reviewed to include new criteria or to up-date their definitions.

CONCLUSION

Health Care Organizations have an essential responsibility to the environment, being significant consumers of natural resources, as well as producers of large quantities of waste, some of which can cause serious risk to people and the environment unless they are properly handled.

The monitoring of a number of environmental matters is key to assessing continuous improvement in the actions undertaken. However, this monitoring requires objective tools which consider a series of criteria adapted to each organization.

This Chapter therefore presents a multicriteria model for assessing environmental responsibility in Health Care Organizations. The intention is that the model, although based on mathematical tools, should be easy to apply, and should take into account the uncertainties and ambiguities which characterize the real-life decision process. The fuzzy TOPSIS technique was thus used, as it allows a large number of alternatives to be assessed in a simple manner, which means the results obtained can be compared over time.

Table 6. Closeness coefficient of each year assessed

Alternatives	CC
Hospital year 1	0.498
Hospital year 2	0.553
Hospital year 3	0.561

(Created by the author)

To test the utility of the model, it was applied in a public Health Care Organization over three consecutive years, showing how the improvement actions undertaken increase the overall utility of the result.

ACKNOWLEDGMENT

This research was supported by the Junta de Comunidades de Castilla-La Mancha and the European Regional Development Fund under Grant number PPII-2014-013-P.

REFERENCES

- Al Garni, H., Kassem, A., Awasthi, A., Komljenovic, D., & Al-Haddad, K. (2016). A multicriteria decision making approach for evaluating renewable power generation sources in Saudi Arabia. *Sustainable Energy Technologies and Assessments*, 16, 137–150. doi:10.1016/j.seta.2016.05.006
- Aragonés-Beltran, P., Mendoza-Roca, J. A., Bes-Pia, A., García-Melon, M., & Parra-Ruiz, E. (2009). Application of multicriteria decision analysis to jar-test results for chemicals selection in the physical–chemical treatment of textile wastewater. *Journal of Hazardous Materials*, 164(1), 288–295. doi:10.1016/j.jhazmat.2008.08.046 PMID:18829168

- Bon-García, J. V. (2012). *Declaración ambiental año 2011*. Consorci Hospital General Universitari Valencia.
- Boran, F. E., Boran, K., & Menlik, T. (2012). The evaluation of renewable energy technologies for electricity generation in Turkey using intuitionistic fuzzy TOPSIS. *Energy Sources Part B*, 7(1), 81–90. doi:10.1080/15567240903047483
- Carnero, M. C. (2014). *Model for Sustainability in Health Care Organizations*. In J. Wang (Ed.), *Encyclopedia of Business Analytics and Optimization* (Vol. 3, pp. 437–456). Hershey, PA: IGI Global.
- Carnero, M. C. (2015). Assessment of Environmental Sustainability in Health Care Organizations. *Sustainability*, 7(7), 8270–8291. doi:10.3390/su7078270
- Chan, H. K., Wang, X., & Raffoni, A. (2012). An integrated approach for green design: Life-cycle, fuzzy AHP and environmental management accounting. *Procedia: Social and Behavioral Sciences*, 57, 236–242.
- Chang, D. Y. (1996). Applications of the extent analysis method on fuzzy AHP. *European Journal of Operational Research*, 95(3), 649–655. doi:10.1016/0377-2217(95)00300-2
- Chen, C. T. (2000). Extension of the TOPSIS for group decision-making under fuzzy environment. *Fuzzy Sets and Systems*, 114(1), 1–9. doi:10.1016/S0165-0114(97)00377-1
- Cintra, Y. C., & Carter, D. (2012). Internalising sustainability: Reflections on management control in Brazil. *International Journal of Strategic Management*, 12(2), 108–125.
- Comunidad de Madrid. (2005). *Guía para la implantación de sistemas de gestión medioambiental en centros hospitalarios*. Madrid: Dirección General de Promoción y Disciplina Ambiental.
- Egilmez, G., Gumus, S., & Kucukvar, M. (2015). Environmental sustainability benchmarking of the U.S. and Canada metropolises: An expert judgment-based multi-criteria decision making approach. *Cities (London, England)*, 42, 31–41. doi:10.1016/j.cities.2014.08.006
- Galdakao-Usansolo Hospital. (2012). *Declaración Ambiental. Ejercicio 2012*. Available online: <http://www.hospitalgaldakao-usansolo.com>
- Galvez, D., Rakotondranaivo, A., Morel, L., Camargo, M., & Fick, M. (2015). Reverse logistics network design for a biogas plant: An approach based on MILP optimization and Analytical Hierarchical Process (AHP). *Journal of Manufacturing Systems*, 37(3), 616–623. doi:10.1016/j.jmsy.2014.12.005
- Garcia, S., Cintra, Y., Torres, R. C. S. R., & Lima, F. G. (2016). Corporate sustainability management: A proposed multi-criteria model to support balanced decision-making. *Journal of Cleaner Production*, 136, 181–196. doi:10.1016/j.jclepro.2016.01.110
- Gumus, A. T. (2009). Evaluation of hazardous waste transportation firms by using a two step fuzzy AHP and TOPSIS methodology. *Expert Systems with Applications*, 36(2), 4067–4074.
- Heo, E., Kim, J., & Boo, K. J. (2010). Analysis of the assessment factors for renewable energy dissemination program evaluation using fuzzy AHP. *Renewable & Sustainable Energy Reviews*, 14(8), 2214–2220. doi:10.1016/j.rser.2010.01.020
- Hermann, B. G., Kroeze, C., & Jawjit, W. (2007). Assessing environmental performance by combining life cycle assessment, multi-criteria analysis and environmental performance indicators. *Journal of Cleaner Production*, 15(18), 1787–1796. doi:10.1016/j.jclepro.2006.04.004

- Higgs, G., Berry, R., Kidner, D., & Langford, M. (2008). Using IT approaches to promote public participation in renewable energy planning: Prospects and challenges. *Land Use Policy*, 25(4), 596–607. doi:10.1016/j.landusepol.2007.12.001
- Hsu, C. W., & Hu, A. H. (2008). Green supply chain management in the electronic industry. *International Journal of Environmental Science and Technology*, 5(2), 205–216. doi:10.1007/BF03326014
- Huang, I. B., Keisler, J., & Linkov, I. (2011). Multi-criteria decision analysis in environmental sciences: Ten years of applications and trends. *The Science of the Total Environment*, 409(19), 3578–3594. doi:10.1016/j.scitotenv.2011.06.022 PMID:21764422
- Isaai, M. T., Kanani, A., Tootoonchi, M., & Afzali, H. R. (2011). Intelligent timetable evaluation using fuzzy AHP. *Expert Systems with Applications*, 38(4), 3718–3723. doi:10.1016/j.eswa.2010.09.030
- Kang, H. Y., Ogunseitan, O., Shapiro, A. A., & Schoenung, J. M. (2007). A comparative hierarchical decision framework on toxics use reduction effectiveness for electronic and electrical industries. *Environmental Science & Technology*, 41(2), 373–379. doi:10.1021/es060739u PMID:17310694
- Kaufmann, A., & Gupta, M. M. (1988). *Fuzzy Mathematical Models in Engineering and Management Science*. Amsterdam: North Holland.
- Kaya, T., & Kahraman, C. (2011). An integrated fuzzy AHP–ELECTRE methodology for environmental impact assessment. *Expert Systems with Applications*, 38(7), 8553–8562. doi:10.1016/j.eswa.2011.01.057
- Khalili, N. R., & Duecker, S. (2013). Application of multi-criteria decision analysis in design of sustainable environmental management system framework. *Journal of Cleaner Production*, 47, 188–198. doi:10.1016/j.jclepro.2012.10.044
- Lahdelma, R., Salminen, P., & Hokkanen, J. (2002). Locating a waste treatment facility by using stochastic multicriteria acceptability analysis with ordinal criteria. *European Journal of Operational Research*, 142(2), 345–356. doi:10.1016/S0377-2217(01)00303-4
- Lamelas, M. T., Marinoni, O., Hoppe, A., & de la Riva, J. (2008). Suitability analysis for sand and gravel extraction site location in the context of a sustainable development in the surroundings of Zaragoza (Spain). *Environmental Geology*, 55(8), 1673–1686. doi:10.1007/s00254-007-1116-9
- Larimian, T., Zarabadi, Z. S. S., & Sadeghi, A. (2013). Developing a fuzzy AHP model to evaluate environmental sustainability from the perspective of Secured by Design scheme-A case study. *Sustainable Cities and Society*, 7, 25–36. doi:10.1016/j.scs.2012.10.001
- Lee, S. K., Mogi, G., Lee, S. K., & Kim, J. W. (2011). Prioritizing the weights of hydrogen energy technologies in the sector of the hydrogen economy by using a fuzzy AHP approach. *International Journal of Hydrogen Energy*, 36(2), 1897–1902. doi:10.1016/j.ijhydene.2010.01.035
- Liang, Z. H., Yang, K., Sun, Y. W., Yuan, J. H., Zhang, H. W., & Zhang, Z. Z. (2006). Decision support for choice optimal power generation projects: Fuzzy comprehensive evaluation model based on the electricity market. *Energy Policy*, 34(17), 3359–3364. doi:10.1016/j.enpol.2005.06.021
- Madu, C. N., Kuei, C., & Madu, I. E. (2002). A hierarchic metric approach for integration of green issues in manufacturing: A paper recycling application. *Journal of Environmental Management*, 64(3), 261–272. doi:10.1006/jema.2001.0498 PMID:12040959
- Mata, F., García, J. A., Sierra, S., Giordani, V., Gutiérrez, F. J., Márquez, M., & Cerrato, R. J. et al. (2011). *EMAS 2009. Declaración ambiental EP Hospital Costa del Sol*. Consejería de Salud de la Junta de Andalucía.

- Munda, G. (2005). Measuring sustainability: A multi-criterion framework. *Environment, Development and Sustainability*, 2(1), 117–134. doi:10.1007/s10668-003-4713-0
- Neste, J., & Karjalainen T. P. (2013). *A literature review – The use of multi-criteria decision analysis in Environmental Impact Assessment, Annex 7.2.5. Inception report IMPERIA EU LIFE11 ENV/FI/905 Action: B1B Report on the use of MCDA in EIA and SEA, 25.4.2013*
- Ouma, Y. O., Opudo, J., & Nyambenya, S. (2015). Comparison of Fuzzy AHP and Fuzzy TOPSIS for Road Pavement Maintenance Prioritization: Methodological Exposition and Case Study. *Advances in Civil Engineering*, Article ID 140189..10.1155/2015/140189
- Pilavachi, P. A., Chatzipanagi, A. I., & Spyropoulou, A. I. (2009). Evaluation of hydrogen production methods using the analytic hierarchy process. *International Journal of Hydrogen Energy*, 34(13), 5294–5303. doi:10.1016/j.ijhydene.2009.04.026
- Pourebahim, S., Hadipour, M., Mokhtar, M. B., & Taghavi, S. (2014). Application of VIKOR and fuzzy AHP for conservation priority assessment in coastal areas: Case of Khuzestan district, Iran. *Ocean and Coastal Management*, 98, 20–26. doi:10.1016/j.ocecoaman.2014.05.009
- Reza, B., Sadiq, R., & Hewage, K. (2011). Sustainability assessment of flooring systems in the city of Tehran: An AHP-based life cycle analysis. *Construction & Building Materials*, 25(4), 2053–2066. doi:10.1016/j.conbuildmat.2010.11.041
- Rodrigues, F., Osiro, L., & Ribeiro, L. C. (2014). A comparison between Fuzzy AHP and Fuzzy TOPSIS methods to supplier selection. *Applied Soft Computing*, 21, 194–209. doi:10.1016/j.asoc.2014.03.014
- Rodríguez, E., Carmona, M., & Soto, J. (2005). *Memoria sostenibilidad del hospital clínico San Carlos*. Madrid, Spain: Hospital Clínico San Carlos.
- Salvado, M. F., Azevedo, S. G., Matias, J. C. O., & Ferreira, L. M. (2015). Proposal of a Sustainability Index for the Automotive Industry. *Sustainability*, 7(2), 2113–2144. doi:10.3390/su7022113
- Santoyo-Castelazo, E., & Azapagic, A. (2014). Sustainability assessment of energy systems: Integrating environmental, economic and social aspects. *Journal of Cleaner Production*, 80, 1–20. doi:10.1016/j.jclepro.2014.05.061
- Tejedor, J. (n.d.). *Declaración Ambiental Enero de 2011-Diciembre de 2011, hm Hospitales*. Available online: <http://www.hmhospitales.com/grupohm/Paginas/HMHospitales.aspx>
- Tseng, M. L., Lin, Y. H., & Chiu, A. S. F. (2009). Fuzzy AHP-based study of cleaner production implementation in Taiwan PWB manufacturer. *Journal of Cleaner Production*, 17(14), 1249–1256. doi:10.1016/j.jclepro.2009.03.022
- Tzeng, G. H., Lin, C. W., & Opricovic, S. (2005). Multi-criteria analysis of alternative-fuel buses for public transportation. *Energy Policy*, 33(11), 1373–1383. doi:10.1016/j.enpol.2003.12.014
- Van Calker, K. J., Berentsen, P. B. M., Romero, C., Giesen, G. W. J., & Huirne, R. B. M. (2006). Development and application of a multi-attribute sustainability function for Dutch dairy farming systems. *Ecological Economics*, 57(4), 640–658. doi:10.1016/j.ecolecon.2005.05.016
- Viaggi, D., Finn, J. A., Kurz, I., & Bartolini, F. (2011). Multicriteria analysis for environmental assessment of agri-environment schemes: How to use partial information from Mid-Term Evaluations? *Agricultural Economics Research*, 12(1), 5–21.
- Vinodh, S., Prasanna, M., & Prakash, N. H. (2014). Integrated fuzzy AHP–TOPSIS for selecting the best plastic recycling method: A case study. *Applied Mathematical Modelling*, 38(19–20), 4662–4672. doi:10.1016/j.apm.2014.03.007

Wang, X., Chan, H. K., Yee, R. W. Y., & Diaz-Rainey, I. (2012). A two-stage fuzzy AHP model for risk assessment of implementing green initiatives in the fashion supply chain. *International Journal of Production Economics*, 135(2), 595–606. doi:10.1016/j.ijpe.2011.03.021

Yanguas, G. (2012). *Perfil ambiental de España. Informe basado en indicadores*. Ministerio de Agricultura, Alimentación y Medio Ambiental Español. Retrieved may 6, 2013 from <http://www.magrama.gob.es/es/>

Zhang, L., Xu, Y., Yeh, C. H., Liu, Y., & Zhou, D. (2016). City sustainability evaluation using multi-criteria decision making with objective weights of interdependent criteria. *Journal of Cleaner Production*, 131(10), 491–499. doi:10.1016/j.jclepro.2016.04.153

ADDITIONAL READING

Griffiths, J. (2006). Environmental sustainability in the national health service in England. *Public Health*, 120(7), 609–612. doi:10.1016/j.puhe.2006.04.005 PMID:16750230

Herva, M., & Roca, E. (2013). Review of combined approaches and multi-criteria analysis for corporate environmental evaluation. *Journal of Cleaner Production*, 39, 355–371. doi:10.1016/j.jclepro.2012.07.058

Janssen, R. (2001). On the use of multi-criteria analysis in environmental impact assessment in The Netherlands. *Journal of multi-criteria decision analysis - Optimization. Learning and Decision Support*, 10(2), 101–109.

Joshi, H., & Latif, M. A. (2004). Environmental impact assessment of water resource projects with matrix and multi-criteria decision-making methods: A case study. *Journal of Environmental Systems*, 3(3), 207–231. doi:10.2190/E7BK-WAHF-QHF5-V6P6

Mendoza, G. A., & Prabhu, R. (2003). Qualitative multi-criteria approaches to assessing indicators of sustainable forest resource management. *Forest Ecology and Management*, 172(1-3), 329–343. doi:10.1016/S0378-1127(02)00044-0

Munda, G., Nijkamp, P., & Rietveld, P. (1994). Qualitative multicriteria evaluation for environmental management. *Ecological Economics*, 10(2), 97–112. doi:10.1016/0921-8009(94)90002-7

Schaltegger, S., Lüdeke-Freund, F., & Hansen, E. (2012). Business Cases for Sustainability. The Role of Business Model Innovation for Corporate Sustainability. *International Journal of Innovation and Sustainable Development*, 6(2), 95–119. doi:10.1504/IJISD.2012.046944

Uysal, F., & Tosun, Ö. (2012). Fuzzy TOPSIS-based computerized maintenance management system selection. *Journal of Manufacturing Technology Management*, 23(2), 212–228. doi:10.1108/17410381211202205

World Commission on Environment and Development. (1987). *Our common future*. Oxford: Oxford University Press.

KEY TERMS AND DEFINITIONS

Multi-Criteria Decision Analysis: It is a part of operations research that use multiple criteria in decision-making processes providing acceptable compromise solutions when criteria are in conflict. There is a relevant quantity of tools belonging to this category, for example Analytic Hierarchy Process (AHP), Analytic Network Process (ANP), Measuring Attractiveness by a Categorical Based Evaluation Technique (MACBETH), ELimination Et Choix Traduisant la REalité (ELECTRE), Multi-Attribute Utility Theory (MAUT), The Preference Ranking Organization METHod for

Enrichment of Evaluations (PROMETHEE), etc. The MCDA techniques allow to construct objective models to improve understanding of underlying decision processes in the systemic processes.

TOPSIS: The Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) is a multicriteria decision-making approach created by Hwang and Yoon in 1981. It is a compensatory aggregation method based on the concept that the best alternative should have the shortest geometric distance to a positive ideal solution (PIS) and the geometric farthest distance from a negative ideal solution (NIS).

E

Potential Benefits and Current Limits in the Development of Demand Response

Clementina Bruno

University of Piemonte Orientale, Italy

INTRODUCTION

Once upon a time, for many families, electricity was a somehow magic and mysterious stuff allowing houses lighting and appliances operation, whose secrets began just behind the switch or the socket. Other, more informed, users knew that it came from generation plants and “travelled” along a grid towards houses or firms. Nowadays, the role of end users has changed a lot. They have a broader knowledge of the electric system, and a certain awareness of being part of it, in some cases not simply as consumption units. Distributed generation (such as residential photovoltaic production) and demand response mechanisms have transformed (residential, industrial or commercial) users in an active part of the electricity supply chain, so that they are often defined as “prosumer” (Crispim et al., 2014).

In particular, Demand Response (DR) is attracting increasing attention from regulators, policy makers and system operators due to its large potential in supporting and, in some cases, substituting generation in providing flexibility to the system. This corresponds, on the academic side, to an exponential growth of scientific production, with focus on the technical or on the socio-economic features of the issue. This chapter will provide a review of some recent contributions on this topic. Far from being exhaustive of the extremely wide related literature, the aim of this chapter is to provide a general presentation of the issue, briefly discussing the main benefits related to DR, as well as the most relevant regulatory, technological and

socio-economic challenges that can slow down or hinder its development. Therefore, this work will provide an analysis of the impact and issues related to DR from a socio-economic perspective. Moreover, it will also briefly consider the role of technology (especially information and communication technology) in supporting DR implementation and, more in general, the evolution towards “smart” systems. The rest of the chapter is organized as follows. The next section defines DR and illustrates the most relevant benefits of its development. Subsequently, challenges to DR development are discussed and some recommendations are provided. Future research directions and conclusions close the work.

BACKGROUND

The literature provides a wide set of definitions of DR. Quite common across these definitions is the focus on end-users and on the modification in their electricity utilization patterns (see, for instance, the list provided in Eid et al., 2016). For example, in the FERC (Federal Energy Regulatory Commission) website¹ DR is defined as

Changes in electric usage by demand-side resources from their normal consumption patterns in response to changes in the price of electricity over time, or to incentive payments designed to induce lower electricity use at times of high wholesale market prices or when system reliability is jeopardized.

There are several typologies of DR mechanisms, which can be classified following different criteria (Vardakas et al., 2015). Here we propose the most common classification.

- Time-based retail rates (Cappers et al., 2012), also called rate-based or price-based programs (Siano, 2014), or implicit DR (SEDC - Smart Energy Demand Coalition, 2015), provide incentive to end-users to modify their consumption as response to price variations. Price fluctuations are designed to reflect the dynamics of the wholesale market price or the grid tariff, and ultimately, of the cost of the electric service. Prices can be predetermined but different for given time periods or move dynamically depending on the system and market contingencies.
- Incentive-based retail programs (Cappers et al., 2012), also defined as event-based programs (Siano, 2014), reliability-based (Shen et al., 2014) or explicit DR schemes (SEDC, 2015) reward consumers through a payment or a bill credit for a reduction in their consumption. Such mechanisms are activated by the entity managing DR services (users can contract directly with the utility or with an aggregator) in response to particular events affecting the electric system, e.g. network congestion².

Examples of price-based DR programs are:

- Time of Use tariffs, where prices are different but fixed for given time periods (e.g. times of the day or days of the week).
- Critical Peak Pricing, that applies particularly high prices for a limited period (few hours) in response to critical technical or economic/market events.
- Critical Peak Rebate, where consumers are recognized a bill rebate if they reduce their consumption below a pre-specified baseline during critical hours.

- Real Time Pricing, where prices vary dynamically (e.g. every hour) following the wholesale market price and/or the actual generation costs.

Common forms of incentive-based mechanisms, instead, are:

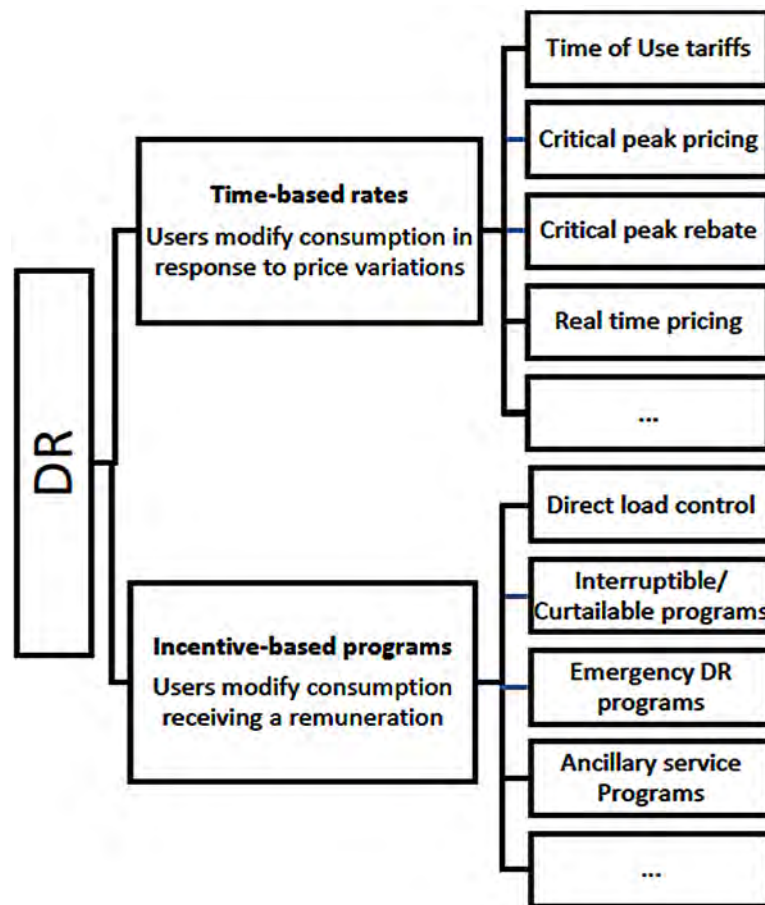
- Direct Load Control, where the utility has the opportunity to manage directly some consumer's equipment (e.g. air conditioning or heating).
- Interruptible/Curtailable programs, where (usually large) users accept that (a part of) their load can be disconnected, in some cases even without notification.
- Emergency DR programs, which provide end-users a compensation to reduce their loads when the system reliability is endangered.
- Ancillary Service Programs, where consumers provide "reserves" by committing to reduce their load in case of necessity.

For further examples or deeper descriptions of DR programs, see, among the others, Cappers et al. (2012), Darby and McKenna (2012), Shen et al. (2014), Siano (2014), Hu et al. (2015), Vardakas et al. (2015). Figure 1 reports some examples of DR programs.

While some mechanisms are well suited even in "traditional" electric systems (e.g. Interruptible/Curtailable programs), other ones present important technological requirements, and can develop their full potential when implemented in smart grid contexts (e.g. Real Time Pricing).

Following Siano (2014; p.462), a smart grid (SG) is "an electric grid able to deliver electricity in a controlled, smart way from points of generation to consumers that are considered as an integral part of the SG since they can modify their purchasing patterns and behavior according to the received information, incentives and disincentives". This definition highlights some relevant peculiarities of SGs with respect to tra-

Figure 1. Examples of DR programs



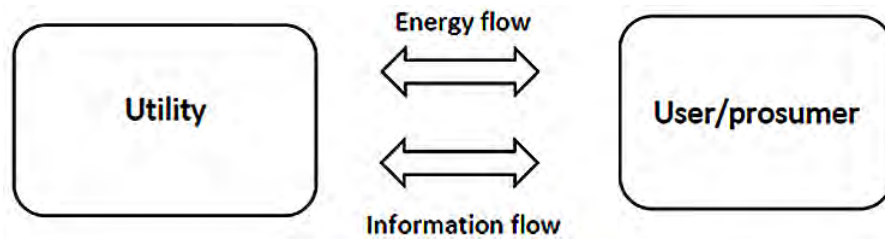
ditional grid structures. First, energy is delivered from generation “points”, rather than generation “plants” (thus the definition encompasses also distributed generation). Second, consumers are active subjects, because they provide services to the electric system, by modifying their behaviour. Third, such modifications are driven by (system or economic) information that consumers receive; however, we must notice that also the information flow to the utility is crucial, since it allows the application of the correct incentive or price schemes. This information exchange relies on the implementation of advanced technologies such as those embedded in smart meters, and is not possible with traditional equipment. Therefore, in SGs, both information and energy can flow from the utility to the users and vice versa, as represented in Figure 2. Finally, the delivery of power is “controlled”

and “smart”, suggesting continuous adjustments of demand and supply to ensure high efficiency in the whole service provision.

DR (especially incentive-based programs) has led to the emergence of a new subject in the electricity value chain, i.e. the aggregator. The aggregator is a market intermediary that “aggregates” the DR capacity from consumers that do not have sufficient dimension, information, knowledge, technology, reliability or availability to participate directly in electricity markets. Usually, the aggregator is an electricity supplier or an independent intermediary (SEDC, 2015; Eid et al., 2015).

The potential positive effects of DR on the electric system and, more generally, on users and on the community as a whole are broadly recognised at the political level. For instance, in the European

Figure 2. Energy and information flows in smart grids



context, the Electricity Directive (2009) and the Energy Efficiency Directive (2012) recognise the opportunities related to DR development and encourage Member States to foster this form of flexibility and to remove potential barriers (e.g. regulatory barriers) to the effective participation of users in the electricity market.

The advantages deriving from the implementation of DR programs are numerous.

Firstly, DR provides system flexibility that can be effectively employed to contrast the intermittency and limited predictability of some Renewable Energy Sources (RES), such as wind and solar. This is especially the case of Real Time Pricing, within the price-based mechanisms, and of several incentive-based programs, such as Direct Load Control (Cappers et al., 2012). Secondly, DR improves the general system reliability, e.g. by reducing demand in case of outages, thus supporting the system recovery, as highlighted by Siano (2014)³. Thirdly, DR programs help flattening consumers' demand, for instance by leading them to reduce consumption or to shift it to off-peak periods. In the short term, this reduces the global generation costs, by limiting the use of expensive peak generation technologies. Moreover, in the long term, this reduces or defers new investments in generation or grid capacity. These cost savings can turn into benefits for consumers. In fact, users participating to DR programs may enjoy lower bills or other compensations for their flexibility. Lower tariffs induced by lower wholesale price, however, will benefit all users, even non-participant ones. Fourthly, non-activation of peak units reduces the related emissions. Fifthly, the "distributed" nature of DR resources is also able to reduce the

network losses associated to transmission and distribution. Finally, a more elastic demand limits the possibility of generators to exert market power. (Batlle and Rodilla, 2009; Darby and McKenna, 2012; Siano, 2014; Shen et al., 2014; Gils, 2014; Hu et al., 2015; Eid et al., 2016).

Several scientific works support the desirability of DR programs, from different perspectives. For instance, Feuerriegel and Neumann (2016), relying on a mathematical model and German data, find that, when DR is employed for load shifting, it can generate savings in the order of 2.83%. Brouwer et al. (2016) employ a simulation tool to evaluate the economic impact of RES penetration for the year 2050. The results show that system costs would increase with RES penetration, due to investments and RES intermittency, but this effect could be counteracted by some options, including DR development. Dupont et al. (2014) run a simulation based on Belgian data and find that DR reduces system costs and emissions while improving reliability. Schleich et al. (2013) analyse econometrically the results of a trial in Austria and show that receiving feedback on consumption can induce households to reduce electricity use by 4.5% on average. Conversely, Torriti (2012) finds that Time-of-Use tariffs in Northern Italy generate lower bills, but higher average consumption. DR effectiveness for load shifting is supported, but limited to morning peaks. Finally, Bradley et al. (2013) rely on an extensive review of the literature and find support to the economic sustainability of DR in UK markets.

Notwithstanding the broad evidence in favour of DR, its development is just in a starting phase and proceeds slowly, with limited consumers'

involvement (Hu et al., 2015). This likely depends on the complexity of adopting and managing DR programs, related to a set of existing challenges hindering DR development. In the following section, some relevant challenges will be discussed.

REGULATORY, TECHNOLOGICAL, AND SOCIO-ECONOMIC CHALLENGES TO DR DEVELOPMENT

Regulatory Challenges: DR Potential and Current Regulation in Europe

In order to understand the potential limits to DR development, we provide an example referred to the European case. SEDC (2015) reports an evaluation of the regulatory framework supporting explicit DR in 2015⁴ for 16 European Countries. Regulation is evaluated with respect to four main “areas”: “consumer access and aggregation”, “program requirements”, “measurement and verification”, “finance and penalties”. Each country receives a score for each area and an overall score resulting from their sum. Higher values correspond to environments more favourable to DR development. Results show that some states have reached an advanced level in the regulation aimed at promoting DR and consumers empowerment, whereas other ones still present relevant inconsistencies between this target and their actual regulatory framework. These differences across countries are highlighted by the diversified grades reported in the last column of table 1, which range from 16 for France and Switzerland (a star indicates existing standardized arrangements involving aggregators) to 3 or 2 for Italy and Spain, respectively.

It is interesting to compare this evaluation with the results reported in Gils (2014). The author provides an assessment of the theoretical DR potential in Europe and some North-African Countries (“theoretical” indicates that the potential includes all equipment available for DR purposes, independently from possible constraints

due to technological, economic, legal issues, etc.). Results are detailed by consumer type and by country, and demand flexibility potential is evaluated either in terms of load reduction (due to load shedding or shifting by delaying consumption) and load increase (related to load shifting by anticipating consumption). Focussing only on the 16 Countries analysed in SEDC (2015), we provide, in the second and third columns of table 1, rankings constructed on the basis of the potential load increase or reduction (originally expressed in MW by Gils, 2014). We can see that these two rankings are quite similar, and their comparison with the last column allows some reasoning. France and UK, that present very high DR potentials and favourable regulatory frameworks, seem to be isolated cases. Several other countries with noticeable potential (Germany, Italy, Spain) present very low SEDC grades. Conversely, countries like Switzerland or Ireland, whose regulatory frameworks are supportive to DR, can exploit only limited potential. Keeping in mind the limitations related to the fact that SEDC (2015) focuses only on incentive-based DR, nevertheless we can conclude that there is an inconsistency between the regulatory evolution and the amount of DR resources that could be exploited. In fact, several states are neglecting the opportunity offered by DR, or, at least, are accumulating an important delay in fostering its development. This happens especially for countries where DR benefits could be large.

In SEDC (2015), the main regulatory limits in promoting DR seem to be related to the area “Consumers participation and aggregation”, for several reasons. For instance, in a number of states, DR resources continue to be excluded from some markets, and generation is still the favourite source of flexibility. In some cases, markets are open to DR, but not to aggregated load, thus limiting the participation of medium or small users. Moreover, the possibility of end-users of freely choosing their DR services provider is often limited.

Other limitations appear also with respect to other “areas”. For example, excessively high

Table 1. DR potential and actual regulation in Europe (our elaboration from SEDC, 2015, and Gils, 2014)

Country	Ranking Based on Potential Load Reduction (Gils, 2014)	Ranking Based on Potential Load Increase (Gils, 2014)	SEDC (2015) Grade
Austria	13	12	10
Belgium	10	11	12
Denmark	14	14	8
Finland	9	10	12
France	2	1	*16
Germany	1	2	6
Ireland	15	15	12
Italy	3	4	3
Netherlands	8	9	10
Norway	11	8	10
Poland	6	6	4
Slovenia	16	16	6
Spain	4	5	2
Sweden	7	7	10
Switzerland	12	13	*16
UK	5	3	12

capacity thresholds to access the market, or other unnecessarily binding requirements (probably a result of older market rules designed for large generators only) in terms of DR events duration or frequency can limit consumers' participation. The fairness and the transparency of the methodologies for measuring the actual consumption reduction constitute another issue, together with the procedures to determine the standard consumption "baseline" (i.e. the starting point for evaluating reductions). Moreover, in some cases, participation to DR programs is subject to the provision of relevant bank guarantees, which can constitute a further barrier.

Some of these points are highlighted also in ENTSO-E (European Network of Transmission System Operators for Electricity, 2015), that also underlines, in a discussion including price- and incentive-based programs, the "lack of effective-

ness of price signals" (p. 5), as well as the limited price variability in some retail markets, especially for medium-small users. This limits both the ability of prices to reflect actual energy costs and the related consumers' responsiveness. Moreover, to ensure small consumers' participation to DR programs, it is necessary that "price signals remains understandable and manageable" (p. 6).

Technological Challenges: Technology Requirements for Effectively Exploiting DR Potential

In order to be effective in providing DR resources, consumers need to interact actively with suppliers or DR services providers. Moreover, they need to be able to react promptly to price or event-driven signals requiring DR activation. This is often not feasible through manual intervention, especially when signals are frequent or provided with very short notice (Cappers et al., 2012).

To deal with this issue, an extended deployment of smart technologies appear necessary, as well as adequate communication and standardization. Below, these concepts are briefly introduced (see Siano, 2014; Shen et al., 2014; ENTSO-E, 2015; Hu et al., 2015; Vardakas et al., 2015).

Among smart technologies, a key role is played by Advance Metering Infrastructures (AMI), systems able to measure electricity usage, to save and analyse the related data, to receive information from devices and to exchange information with utilities. Important components of AMI are smart meters, electronic devices able to measure users' consumption at fixed time intervals (e.g. one hour) and communicate data to suppliers.

In addition, automation technology allows consumers to program "smart" appliances (such as specifically designed washing machines or refrigerators) or control devices (e.g. "smart thermostats") to automatically respond to event or price signals. For instance, "smart thermostats" can adjust the temperature of rooms in response to electricity price variations. Home Area Networks (HAN) can be employed for connecting

all these home devices among each other and with the utilities.

Finally, it is relevant to underline the key role of communication. For an effective use of the communication infrastructure, integration among networks should be promoted, and this would require the adoption of open communication standards. Concerning communication systems (Siano, 2014), they can be one-way (information and DR signals flow from the utility to the end-users), or more expensive but more effective two-way systems, that, additionally, allow utilities to receive feedback in relation to consumers response (e.g. smart meters). They can be wireless (e.g. relying on cellular networks) or wired. The latter can rely, for instance, on power line communications (PLC, either broadband or narrow band), suitable at local level and not requiring new infrastructure building, or on optic fiber communications, for longer distances. The key role of communication highlights the importance of security (Vardakas et al., 2015), either in terms of protection of consumers' sensitive data and privacy, or in terms of preservation of system and market information from external inference or attacks that can create damages or inconvenience.

Economic and Social Challenges

From the purely economic perspective, two major issues, somehow interrelated, concern the provision of sufficient economic incentives to participate to DR programs (ENTSO-E, 2015) and the distribution of investment costs related to enabling technologies. With respect to the latter point, Eid et al. (2016) underlines the problem of correctly splitting DR cost and benefits along the whole electricity supply chain (e.g. among network operators, retailer and consumers), in such a way to create positive business cases for all the involved parties. Failing in providing proper mechanisms to reach this goal generates the risk that none of the parties would chose to make the "first move" (p. 22). The authors report an example related to investments in smart meters. Moreover,

they state that, as the (environmental) benefits of such programs would affect the whole society, this should be the case also for costs. Also Shen et al. (2014), with a similar line of reasoning, suggest that incentives to DR participation and investments could effectively receive public support, in consideration of their positive impact on the society. Hu et al. (2015) states that utilities do not have sufficient incentives to support DR in absence of subsidies. ENTSO-E (2015), however, stresses that general subsidization of DR programs must be minimal and avoid market distortions, since DR "must achieve its full economic potential in fair competition with other sources" (p.8).

Turning, now, to the former point, as highlighted by Cappers et al. (2012), "the opportunities created by the DR service providers must generate sufficient value to customers or else they will eschew these offerings" (p.426). In this perspective, with specific regards to residential users, Darby and McKenna (2012) underlines the key role of rates. Cappers et al. (2012) stress also that a sufficient remuneration is necessary for DR service providers as well, which will be in charge of marketing and managing customers' involvement in DR programs. These points highlight the crucial role played by the design of DR mechanisms: advanced optimization methods provide relevant support in this sense. Vardakas et al. (2015) report a broad survey and discussion of optimization models for DR, highlighting the plurality of available options. For instance, the objective function(s) of the optimization problem may target the minimization of electricity costs or of total consumption, or the maximization of social welfare, or combine two of these objectives. Moreover, based on the form of the objective functions or of the constraints, the optimization problem can be linear or non-linear, while the variable characteristics can lead to integer or mixed-integer optimization problems; each type of problem requires an appropriate technique to be solved. Different sources of complexity can arise in dealing with such models, e.g. problems with no feasible solutions or excessive computational

times. Additionally, game theory provides valuable theoretical support to model the interactions among electric system actors (users or utilities). For a deep discussion and examples of optimization algorithms in the context of DR and SGs, see Vardakas et al. (2015) and the references therein.

Moreover, there are some social and cognitive issues potentially affecting users' willingness to be involved in DR, and this holds especially for residential users. In fact, apart from the gratification that can derive from the idea of having contributed to improve the system reliability (Siano, 2014), users could be affected by some kinds of inconvenience. First, they will need to change some consumption habits: while some activities can be more easily shifted to a different moment in time, other activities, such as cooking (and eating) are less likely to be suitable for variations (Darby and McKenna, 2012). In addition, some consumers could have a limited willingness to accept high levels of automation and external control of their activities by a "Big Brother" (Cappers et al., 2012, p.425). Darby and McKenna (2012) suggest that home automation could effectively involve heating and cooling, as well as electric vehicles, while consumption related to other activities can be managed manually. Moreover, the same contribution highlights that consumers used to flat rates are not likely to switch easily to variable tariffs, and that a gradual passage to Time of Use pricing and subsequently to the more effective, but more complex, Real Time Pricing could represent a feasible path.

Many of the mentioned authors, however, agree on the key role of consumers' education and information in order to overcome most of the social issues described above, as long as the DR services providers will be able to achieve a high level of trust from customers and to provide an adequate protection of their rights and privacy.

SOLUTION AND RECOMMENDATIONS

Throughout this chapter, we have drawn a picture of DR as a promising resource developing in the electric system that, however, still presents many unknown features, which make difficult evaluating its potential, thus slowing down or limiting its development.

As in any other case where something "new" is developing on a broad (in this case global) scale, also with respect to DR it is important that all the involved actors (Governments, regulators, utilities, users, etc.) can make their choices on an informed basis. Therefore, as highlighted in Hu et al. (2015), it is crucial that further studies are promoted and developed.

In my opinion, such studies could focus on several fields and adopt different methodologies (examples can range from game-theoretical studies aimed at identifying the best market design to field trials evaluating the practical consequences of specific projects). A multi-disciplinary approach involving different entities, professional figures and expertise (e.g. regulators and utilities, practitioners and academics) seems the most promising strategy, given the multiplicity of perspectives that need to be considered to effectively deal with such a broad issue.

FUTURE RESEARCH DIRECTIONS

Future research focused on DR can take manifold directions. With reference to the economic and social sciences fields, some (not exhaustive) examples could be the following.

- Direct evaluations of consumers' preferences and willingness to accept contracts

including different kinds of DR programs could be performed by adopting surveys approaches relying on contingent valuation or conjoint analysis. These approaches can be implemented describing hypothetical scenarios to generic consumers, or directed to subjects involved in actual trials or pilot studies. These methods can be designed also to infer the correct level of monetary incentive to be provided.

- DR affects system operations. Efficiency analysis techniques (parametric and non-parametric) can be effectively employed to evaluate the impact of DR on the performance of firms operating in different phases of the supply chain (generation or network management).
- A more general approach based on cost-benefit analysis, adopting a macro-economic perspective, appears crucial for evaluating specific programs before their implementation. Such an approach should be able to incorporate also non-monetary elements, such as environmental benefits related to reduced emissions.

CONCLUSION

Interest in DR as a source of flexibility is increasing from several direction (policy makers, regulators, utilities, academics and users). The potential benefits related to DR development are numerous and extremely interesting. Nevertheless, the implementation of such programs is still in a starting phase and is advancing in a relatively slow way. This is due to the complexity of the management of this new resource. This chapter has analysed the main regulatory, technological and socio-economic challenges to DR development. As a conclusion, the level of knowledge of the issue needs to be expanded in order to support informed decision-making, and multi-

disciplinary approaches can constitute, as this article has discussed, an effective strategy given the multiplicity of perspectives to be considered in dealing with DR issues.

ACKNOWLEDGMENT

This research is supported by the Associazione per lo sviluppo dell'Università nel territorio novarese.

REFERENCES

- Battle, C., & Rodilla, P. (2009). Electricity demand response tools: Current status and outstanding issues. *European Review of Energy Markets*, 3, 1–27.
- Bradley, P., Leach, M., & Torriti, J. (2013). A review of the costs and benefits of demand response for electricity in UK. *Energy Policy*, 52, 312–327. doi:10.1016/j.enpol.2012.09.039
- Brouwer, A. S., Van den Broek, M., Zappa, W., Turkenburg, W. C., & Faaij, A. (2016). Least-cost options for integrating intermittent renewables in low-carbon power systems. *Applied Energy*, 161, 48–74. doi:10.1016/j.apenergy.2015.09.090
- Cappers, P., Mills, A., Goldman, C., Wiser, R., & Eto, J. H. (2012). An assessment of the role mass market demand response could play in contributing to the management of variable generation integration issues. *Energy Policy*, 48, 420–429. doi:10.1016/j.enpol.2012.05.040
- Crispim, J., Braz, J., Castro, R., & Esteves, J. (2014). Smart Grids in the EU with smart regulation: Experiences from the UK, Italy and Portugal. *Utilities Policy*, 31, 85–93. doi:10.1016/j.jup.2014.09.006
- Darby, S. J., & McKenna, E. (2012). Social implication of residential demand response in cool temperate climates. *Energy Policy*, 49, 759–769. doi:10.1016/j.enpol.2012.07.026

Directive 2009/72/Ec Of The European Parliament And Of The Council of 13 July 2009 concerning common rules for the internal market in electricity. (n.d.). Retrieved July 20th, 2016 at: <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32009L0072&from=EN>

Directive 2012/27/Eu Of The European Parliament And Of The Council Of 25 October 2012 On Energy Efficiency. (n.d.). Retrieved July 20th, 2016 At: <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32012L0027&from=EN>

Dupont, B., Dietrich, K., De Jonghe, C., Ramos, A., & Belmans, R. (2014). Impact of residential demand response on power system operation: A Belgian case study. *Applied Energy*, 122, 1–10. doi:10.1016/j.apenergy.2014.02.022

Eid, C., Codani, P., Chen, Y., Perez, Y., & Hakvoort, R. (2015). *Aggregation of demand side flexibility in a smart grid: A review of European market design*. EEM15 – 12th International Conference on the European Energy Market, Lisboa,, Portugal.. doi:10.1109/EEM.2015.7216712

Eid, C., Koliou, E., Valles, M., Reneses, J., & Hakvoort, R. (2016). Time-based pricing and electricity demand response: Existing barriers and next steps. *Utilities Policy*, 40, 15–25. doi:10.1016/j.jup.2016.04.001

European Network of Transmission System Operators for Electricity – ENTSO-E. (2015). *Market design for demand side response*. Policy paper, November 2015. Retrieved September 12th, 2016 at: https://www.entsoe.eu/Documents/Publications/Position%20papers%20and%20reports/entsoe_pp_dsr_web.pdf

Federal Energy Regulatory Commission (FERC). Retrieved from <http://www.ferc.gov/industries/electric/indus-act/demand-response/dem-res-adv-metering.asp>

Feuerriegel, S., & Neumann, D. (2016). Integration scenarios of Demand Response into electricity markets: Load shifting, financial savings and policy implications. *Energy Policy*, 96, 231–240. doi:10.1016/j.enpol.2016.05.050

Gils, H. C. (2014). Assessment of the theoretical demand response potential in Europe. *Energy*, 67, 1–18. doi:10.1016/j.energy.2014.02.019

Hu, Z., Kim, J., Wang, J., & Byrne, J. (2015). Review of dynamic pricing programs in the U.S. and Europe: Status quo and policy recommendations. *Renewable & Sustainable Energy Reviews*, 42, 743–751. doi:10.1016/j.rser.2014.10.078

Schleich, J., Klobasa, M., Gözl, S., & Brunner, M. (2013). Effects of feedback on residential electricity demand – Findings from a field trial in Austria. *Energy Policy*, 61, 1097–1106. doi:10.1016/j.enpol.2013.05.012

Shen, B., Ghatikar, G., Lei, Z., Li, J., Wikler, G., & Martin, P. (2014). The role of regulatory reforms, market changes, and technology development to make demand response a viable resource in meeting energy challenges. *Applied Energy*, 130, 814–823. doi:10.1016/j.apenergy.2013.12.069

Siano, P. (2014). Demand response and smart grids – A survey. *Renewable & Sustainable Energy Reviews*, 30, 461–478. doi:10.1016/j.rser.2013.10.022

Smart Energy Demand Coalition. (2015). *Mapping Demand Response in Europe Today 2015*. Retrieved July 20th, 2016 at: <http://www.smart-energydemand.eu/wp-content/uploads/2015/09/Mapping-Demand-Response-in-Europe-To-day-2015.pdf>

Torriti, J. (2012). Price-based demand side management: Assessing the impacts of time-of-use tariffs on residential electricity demand and peak shifting in Northern Italy. *Energy*, 44(1), 576–583. doi:10.1016/j.energy.2012.05.043

Vardakas, J. S., Zorba, N., & Verikoukis, C. V. (2015). A survey on demand response programs in smart grids: Pricing methods and optimization algorithms. *IEEE Communications Surveys and Tutorials*, 17(1), 152–178. doi:10.1109/COMST.2014.2341586

ADDITIONAL READING

Aalami, H. A., Moghaddam, M. P., & Yousefi, G. R. (2010). Demand response modelling considering interruptible/curtailable loads and capacity market programs. *Applied Energy*, 87(1), 243–250. doi:10.1016/j.apenergy.2009.05.041

Agrell, P. J., Bogetoft, P., & Mikkers, M. (2013). Smart-grid investments, regulation and organization. *Energy Policy*, 52, 656–666. doi:10.1016/j.enpol.2012.10.026

Albadi, M. H., & El-Saadany, E. F. (2008). A summary of demand response in electricity markets. *Electric Power Systems Research*, 78(11), 1898–1996. doi:10.1016/j.epsr.2008.04.002

Brouwer, A. S., Van den Broek, M., Seebregts, A., & Faaij, A. (2015). Operational flexibility and economics of power plants in future low-carbon power systems. *Applied Energy*, 156, 107–128. doi:10.1016/j.apenergy.2015.06.065

Earle, R., Kahn, E. P., & Macan, E. (2009). Measuring the capacity impacts of demand response. *The Electricity Journal*, 22(6), 47–58. doi:10.1016/j.tej.2009.05.014

Feuerriegel, S., & Neumann, D. (2014). Measuring the financial impact of demand response for electricity retailers. *Energy Policy*, 65, 359–368. doi:10.1016/j.enpol.2013.10.012

Katz, J. (2014). Linking meters and markets: Roles and incentives to support a flexible demand side. *Utilities Policy*, 31, 74–84. doi:10.1016/j.jup.2014.08.003

Kondziella, H., & Bruckner, T. (2016). Flexibility requirements of renewable energy based electricity systems – a review of research results and methodologies. *Renewable & Sustainable Energy Reviews*, 53, 10–22. doi:10.1016/j.rser.2015.07.199

Lang, C., & Okwelum, E. (2015). The mitigating effect of strategic behavior on the net benefits of a direct load control program. *Energy Economics*, 49, 141–148. doi:10.1016/j.eneco.2015.01.025

Rochlin, C. (2009). The alchemy of demand response: Turning demand into supply. *The Electricity Journal*, 22(9), 10–25. doi:10.1016/j.tej.2009.09.004

Rodrigues, R., & Linares, P. (2014). Electricity load level detail in computational general equilibrium – Part I – Data and calibration. *Energy Economics*, 46, 258–266. doi:10.1016/j.eneco.2014.09.016

Rodrigues, R., & Linares, P. (2015). Electricity load level detail in computational general equilibrium – Part II – welfare impacts of a demand response program. *Energy Economics*, 47, 52–67. doi:10.1016/j.eneco.2014.10.015

Ruester, S., Schwenen, S., Batlle, C., & Pérez-Arriaga, I. (2014). From distribution networks to smart distribution systems: Rethinking the regulation of European electricity DSOs. *Utilities Policy*, 31, 229–237. doi:10.1016/j.jup.2014.03.007

Wang, Q., Zhang, C., Ding, Y., Xydis, G., Wang, J., & Østergaard, J. (2015). Review of real-time electricity markets for integrating Distributed Energy Resources and Demand Response. *Applied Energy*, 138, 695–706. doi:10.1016/j.apenergy.2014.10.048

KEY TERMS AND DEFINITIONS

Demand Response (DR): Source of flexibility for the electricity system relying on modifications in users' consumption.

Enabling Technologies: With respect to DR, this term indicates those technologies (e.g. equipment, appliances) that allow an effective implementation of DR programs. Examples could be smart meters, allowing communication between users and utilities, or home appliances that automatically react to price or non-price signals by modifying electricity consumption.

Incentive-Based DR: DR mechanisms that rely on consumption modifications induced by non-price signals, for which the involved users receive compensations.

Off-Peak: Points of lower demand (with respect to the peak-load) in a certain time period.

Peak-Load: Within the electric system, it is the point of highest demand in a certain time period. We can have daily peaks, seasonal peaks, etc.

Peak-Shifting: Refers to the possibility of moving demand from peak to off-peak times. This is one of the desirable effects of DR.

Price-Based DR: DR mechanisms that rely on consumption modifications induced by price variations.

ENDNOTES

- ¹ <http://www.ferc.gov/industries/electric/indus-act/demand-response/dem-res-adv-metering.asp>
- ² Siano (2014) identifies as a separated category demand reduction bids, where participant send their bids (offer a demand reductions and request a price) to the utility or the aggregator.
- ³ The authors refers to DR benefits in a SG context, but most of the proposed lines of reasoning are likely to apply in a more general framework.
- ⁴ As mentioned in the report, explicit DR has limited requirements in terms of public investment in technology, therefore regulation plays the more critical role in making DR programs available and suitable for end-users.

Waste Gas End-of-Pipe Treatment Techniques in Italian IPPC Chemical Plants

Gaetano Battistella

ISPRA, Italy

Giuseppe Di Marco

ISPRA, Italy

Carlo Carlucci

ISPRA, Italy

Raffaella Manuzzi

ISPRA, Italy

Federica Bonaiuti

ISPRA, Italy

Celine Ndong

ISPRA, Italy

INTRODUCTION

Due to more stringent emission regulations, very efficient new advanced emission control technologies are required adopting National IPPC (Integrated Pollution Prevention and Control) Permits (below AIA) based on Best Available Technologies (below BAT) Conclusions.

Some of these techniques are operating inside Chemical Plants and Refineries based in Italy, such as Oxidation, Adsorption and Absorption devices. Other techniques (i.e. the ones that are new advanced technologies still in research or in demonstration state), are not subject of this Paper, based on describing running situation inside operating IPPC Chemical Plants and Refineries licensed in Italy at National Level.

This paper includes, but are not limited to, the results of a screening of Italian Chemical IPPC Industries and Refineries up to day, trying

to highlight operating conditions and possible already existing improvements for removal of:

- VOC and other cancer causing and toxic substances;
- Dust, Mercury and heavy metals;
- NO_x and Nitrogen compounds;
- SO_x and Sulphur compounds;
- Chlorides and Fluoride compounds.

The abatement techniques analyzed in this work operate mainly on VOC content reduction, through the use of Oxidizing devices or on inorganic compounds abatement (in addition to VOC), through the use of Absorption or Adsorption devices.

Superior Environmental Protection and Research Institute (below ISPRA) experience, mainly developed as Technical Support to Italian Minister of Environment, Land and Sea (below IMELS),

has allowed to analyze different operative conditions, related with abatement techniques and their application in IPPC permitted plants.

Many pollutants emitted from IPPC plants (according to Environmental Permits limit values) have been identified and charted a profile of possible application for abatement techniques in these plants in their different IPPC categories.

The results of this analysis allow to suggest a possible reconsideration and, also, new assessment for some end-of-pipe devices, in order to find other better defined operational contexts, different from actually Italian provisions and, also, an evaluation of current operational performances of the devices, in order to improve their environmental conditions, consistently with BAT application.

BACKGROUND

In Italy, IPPC Permit is an authorization released for environmental protection purposes, in order to prevent and control pollution ‘at the source’ by means of an integrated authorization, allowing operation of IPPC industrial activities with specified production’s characteristics and dimensions, at both national and regional levels (Battistella, 2013).

The list of the categories of these specific industrial activities is regulated by the Italian Legislative Decree n. 59/2005 and s.m.i. (Italian Legislative Decree n. 152/2006 and s.m.i.) that adopts and endorses the Directive n. 96/61/EC and s.m.i. (Directive 2008/1/EC and s.m.i.) concerning integrated pollution prevention and control (actually recast in the Directive 2010/75/EU).

IPPC permits – by law definition - plan and perform an integrated prevention and control set in the exact point of pollution (‘a la source’), e.g. pollutants are identified, declared, controlled, detected and monitored in the admission/emission points of the IPPC industrial activities, as well as during all activities of industrial plants’ operation (Battistella & Di Marco, 2013a; Battistella & Di Marco, 2013b).

This means authorization of plants’ operation controlling natural resources’ usage, as well as emissions and discharges in the environment inside predefined limit values with prescriptions, adoption of predefined monitoring framework, as self-controls on selected parameters, frequencies and methodologies, with a periodic reporting and planned inspections based on the effective environmental risk (Battistella & Di Marco, 2013a; Battistella & Di Marco, 2013b).

In Italy, AIAs are released by the Competent Authority, as

- By IMELS for national strategic plants;
- By other Authority designed by Region or autonomous Province for others.

In order to accomplish IPPC permits operative performances in terms of assigned limit values, among other provisions, also waste gas end-of-pipe treatment devices are adopted - as well as installed and operated - in order to abate or at least decrease pollutants’ contents (even often dangerous substances and compounds) in *waste-gases* before their emission into open air.

Adopted techniques must be considered equal to, or committed to become under IPPC permits period of duration (more or less 10 years), as Best Available Techniques and their operating performances are described in details into *Reference Documents on Best Available Techniques (BRefs¹)*.

BAT CONCLUSIONS APPLICATION IN ITALIAN IPPC PERMITTED PLANTS

The Italian National Environmental Regulatory Framework for IPPC Installations

As regards to Italian Regulations, Attachment X to Part II of Legislative Decree n. 152/06 ^[9] defines the list of main pollutants to be monitored,

if pertinent, with reference to air emissions for IPPC Plants under operational licensee.

Attachments to Part V of Legislative Decree n. 152/06 list minimum emissions values (relevance thresholds) and maximum emissions values (limit values) and related prescriptions (continuous monitoring, process parameters to monitor, measurements' norms, continuous monitoring control procedures, conformity criteria for measurements, etc.) for pollutant substances emitted in atmosphere - as conveyed emissions - at the stacks and specific indications for some substances (as SO_x, NO_x, CO, Dusts, VOC, etc.) and for some types of plants.

It is relevant to highlight that Legislative Decree n. 152/06 *does not establish specific limit values for emissions of pollutants from end of pipe treatment devices*, as well as *does not even describe specific abatement techniques*.

This is why applied emission limit values are strictly connected with specific pollutants and plant units from where they are generated, as mentioned in Suppl. Annex I at V Part of Italian Legislative Decree n.152/06.

Table 1 shows main pollutants related with operative experience in Italian IPPC plants.

In III Part of mentioned Annex I, emission levels for some specific type of production such

Table 1. Main pollutants related with operative experience in Italian IPPC plants

Pollutant	Class Identity (Annex I, V Part of Italian Legislative Decree n.152/06)	Relevance Threshold (g/h)	Limit Value (mg/Nm³)
Point 1.1. Cancer-causing, reproduction toxic or mutagen compounds (Table A1)			
Benzene	class III	25	5
PHA (summation)	class I	0,5	0,1
Organic chlorated compounds	class III	25	5
Point 1.2. Compounds that can be especially toxic and be accumulated (Table A2)			
Dioxins and Furans	class I	0,02	0,01
PCB	class II	0,5	0,5
Point 2. Inorganic compounds that mainly come up in dust form (Table B)			
Hg	class I	1	0,2
Point 3. Inorganic compounds that mainly come up in vapour or gas form (Table C)			
SO _x	class V	5000	500
NO _x	class V	5000	500
Chlorine	class II	50	5
Chlorine inorganic compounds (as HCl)	class III	300	30
Fluorine and its compound (as HF)	class II	50	5
H ₂ S	class II	50	5
NH ₃	class IV	2.000	250
Point 4. Organic Compounds that come up in dust, vapour or gas form (Table D)			
Organic compounds (included Chlor Aromatics, Halogenic hydrocarbons and Nitro aromatics compounds)	It depends on the specific compound		
5. Total dust			
Total Dust	-	> 5000	50
		>1000 <5000	150

as Non ferrous metal, Sulphuric acid, Chlorine, Fertilizers, polymers are described, while in IV Part emission levels for Refineries are described.

For VOC in Annex III at V Part of Italian Legislative Decree n.152/06, emission levels for particular type of compound are described as shown in the following Table 2.

Screening of IPPC Permitted Chemical Plants and Refineries in Italy

In Italy, at December 2014, are operating 175 activities so called “strategic” (IPPC permits released at national level) and between these there are: 16 Crude Oil Refineries and 46 large Chemical Plants and about 5.560 Plants have received IPPC Permits, while almost 200 are concluding their procedures (no one with critical conditions

to be resolved for EU obligations) in terms of first release (Superior Environmental Protection and Research Institute [ISPRA], 2015).

Figure 1 shows the IPPC Permits situation at the beginning of year 2013 (ISPRA, 2015).

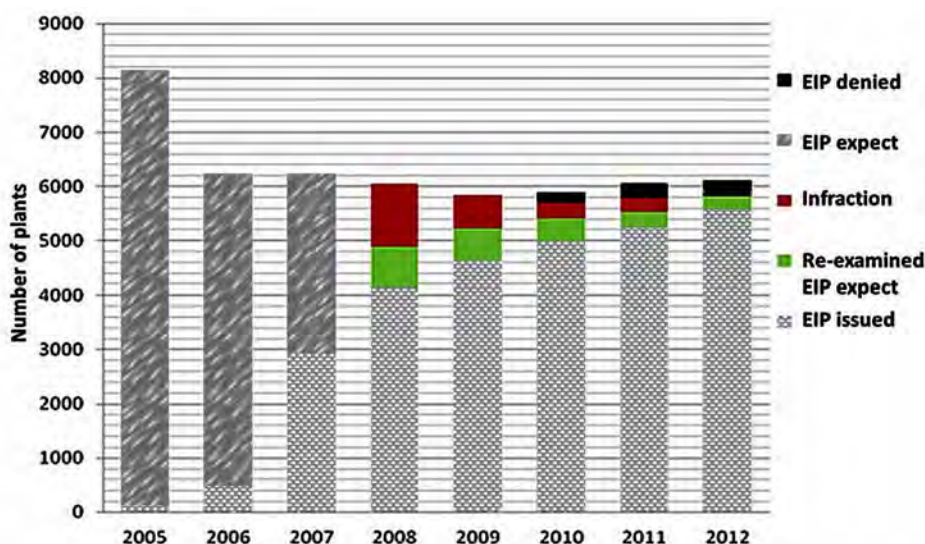
In the Italian procedure to release IPPC permits, ISPRA is on charge of performing the requested comparisons of environmental parameters with the applicable ‘Best Available Techniques’ - contained inside available published BRef documents – in order to evaluate and to propose to the ‘Technical Investigating Group’ appropriate limit values, prescriptions and recommendations to be adopted in the ‘Technical Advice’, to be approved (Battistella, 2013; Battistella & Di Marco, 2013a; Battistella & Di Marco, 2013b).

In this way, also a virtual process is started, in order to compare industrial activities and plants’ performances in term of ‘resource efficiency’,

Table 2. Emission levels for VOC

Pollutant	Identity	Relevance Threshold (g/h)	Limit Value (mg/Nm ³)
VOC	Cancer-causing, reproduction toxic or mutagen compounds	10	2
	Alogenated	100	20

Figure 1. IPPC Permits situation at the beginning of year 2013



trying to optimize the analyzed processes and to reduce their emissions limits.

The complete IPPC Permits release procedure is shown in the following Figure 2 (Battistella, 2013).

In Italy, since its start in 2009, this procedure allowed the results shown in the following figures for abatement of main environmental pollutants (air emissions) and at December 2014, every IPPC Plant have been licensed with IPPC Permits, in terms of first release, concluding authorization procedures (ISPRA, 2015).

1. *For 16 Italian Refineries*, the result in pollution abatement for macro pollutants, such as SO_x, NO_x, CO and PTS, is showed in Figure 3.
2. *For 46 Italian Chemical Plants*, the result in pollution abatement for macro pollutants, such as SO_x, NO_x, CO and PTS, is showed in Figure 4.

BAT Conclusions Application in IPPC Permitted Plants

In Italy, recent regulatory innovations introduced by Legislative Decree n. 46/2014—adopting in Italy

the European Commission ‘Industrial Emissions Directive’ - led to the imposition, during operation of IPPC installations as specified in AIA Permits, of further specific indications/prescriptions as defined in new BAT for each IPPC sector, with revised emission limit values for pollutants released into the environment, as reported in “BAT Conclusions” of new reviewed BRefs (Battistella, Di Marco, Bonaiuti, Carlucci, 2014).

This new approach applies to all IPPC installations subject to AIA Permits (first release) at national level, for which new legislation requires a review and an updating of existing authorizations, through new procedural “review/renewal” mechanisms, now closely linked to the adoption of new technologies as identified and published into new Brefs documents as ‘BAT Conclusions’ (Battistella et al., 2014).

In numerous ‘vertical’ BRefs, or rather specific Reference Documents for each type of production plants, end-of-pipe gas treatments are considered as a BAT for VOC and inorganic compounds abatement.

Almost all of these documents send back to ‘horizontal’ *Reference Document on Best Available Techniques in Common Waste Water and*

Figure 2. IPPC Permits release procedure

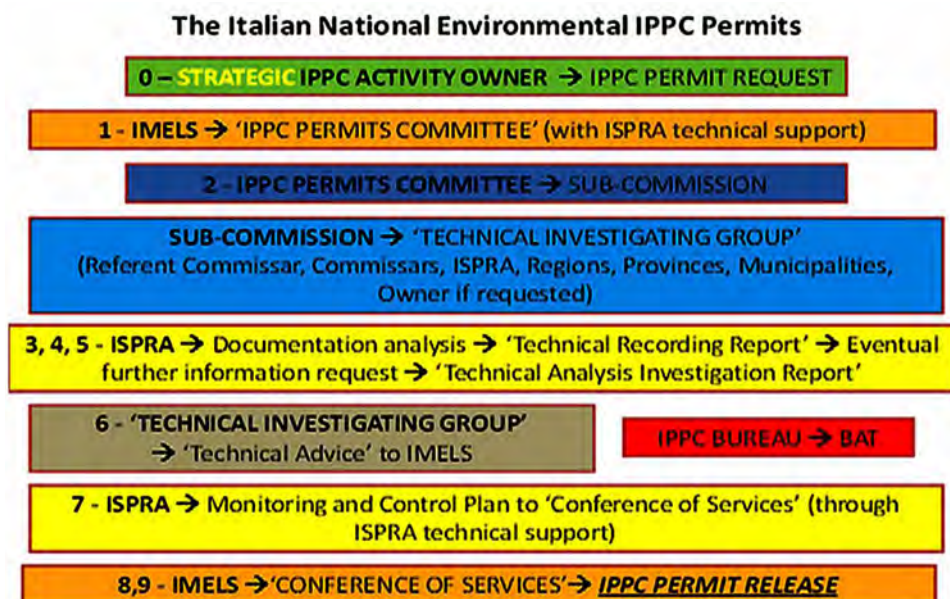


Figure 3. Pollution abatement for macro pollutants in Italian refineries

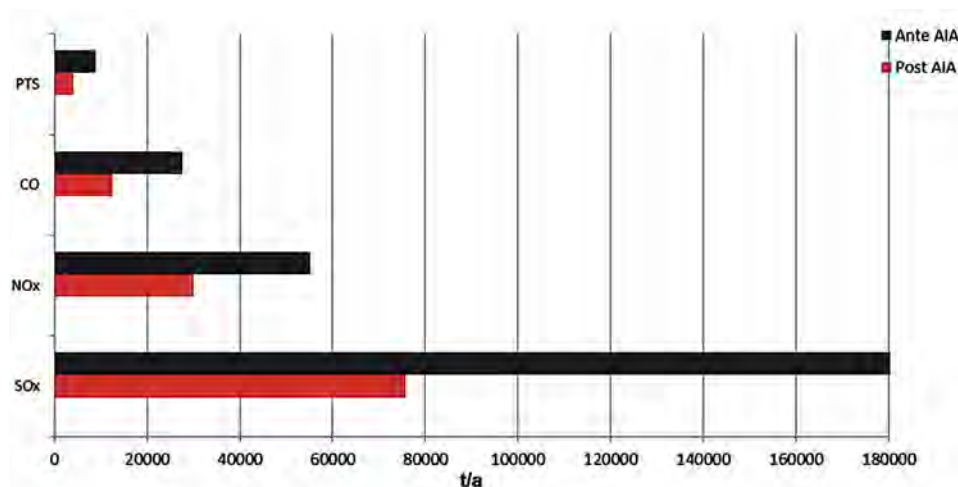
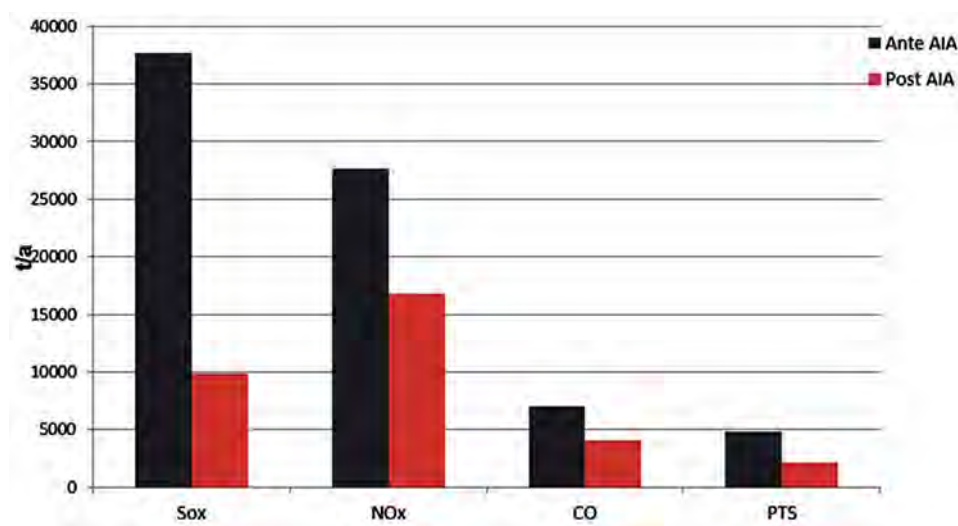


Figure 4. Pollution abatement for macro pollutants in Italian chemical plants



Waste Gas Treatment / Management Systems in the Chemical Sector - February 2003, for achievable emission limits and operative conditions of end of pipe devices.

Design standards are strictly related to properties of gas flow that define operative conditions, achievable performance and cross-media effects.

In Table 3 are described operative conditions and performance and emission levels achievable by implementation of BAT.

The following figures show the state-of-the-art of BAT application for waste-gas cleaning in

Italian AIA IPPC plants (Chemical plants and Refineries)³.

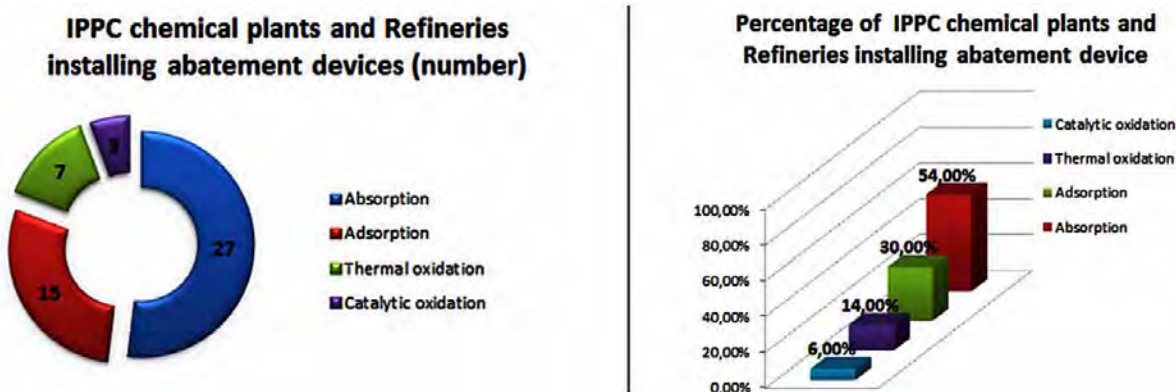
Going towards a complete BAT application for waste-gas cleaning operations, Italian AIA IPPC chemical plants and Refineries actually apply mainly Absorption and Adsorption techniques (respectively 54% and 30% of total AIA IPPC Chemical plants and Refineries adopting BAT⁴ as shown in Figure 5).

Absorption devices are mostly installed in IPPC Chemical Plants due to the properties of emitted pollutants (mainly inorganic ones), whereas Ad-

Table 3. Operative conditions and performance and emission levels achievable by implementation of BAT

Description ²	Thermal oxidation	Catalytic oxidation	Wet scrubbing (Absorption)	Adsorption
	Oxidation with oxygen (air) by heating a gas above its auto-ignition point	Oxidation with oxygen (air) using catalyst to lower auto-ignition point	Mass transfer from gaseous phase into liquid phase	Mass transfer from gaseous phase to solid surface
Application	Emissions from all VOC sources most suitable for autothermal VOC concentrations and final treatment for hazardous substances	Same application as thermal incineration, contaminants restricted to non-poisonous ones. It can be with or without heat recovery	Control of VOC, inorganic compounds and also of dust, depending on variant and scrubbing liquid (water, acidic and alkaline solution). By desorption it's also possible recovery of substances.	Removal of VOC, odorous substances, dioxins, mercury After regeneration it's also possible recovery of substances.
Cross-media effects	CO and NO _x in flue gas Cl- and S-content require flue gas treatment Dioxins can represent a problem if not operated at optimized conditions	Very low NO _x content in flue gas (about 15 mg/Nm ³) Cl- and S- content require flue gas treatment Dioxins are usually not a problem	Waste water to be treated Energy consumption and emissions during regeneration	Waste water generated during regeneration Disposal of adsorbent
Achievable performance [% pollutant removal]	VOC 95-99	VOC 90-99 CO >98 Odour 80-95	VOC 50-99 Inorganic compounds 90-99 SO ₂ 80-99	VOC 80-95 Odour 80-95 H ₂ S 80-95
Achievable emission levels [mg/Nm ³]	TOC 1-4	-	HF <1 HCl <10 (<50 with water) SO ₂ < 40	Hg < 0,05 Dioxins <0,1 ng/ Nm ³ TEQ

Figure 5. IPPC Chemical plants and Refineries installing abatement devices



sorption technique is mostly adopted in Refineries because of the mainly presence of organic emitted pollutants (see Figure 6).

Following graphics (see Figures 7, 8, 9, 10) represent operating data concerning waste gas end-of-pipe treatment techniques as run in licensed Italian Chemical plants and Refineries⁵ in order to represent operative conditions and emissions

scenarios related with application of BAT in Italian IPPC Plants.

The upper value for each board of graphics represent the *maximum IPPC Permits limit values* for each IPPC Category of activities.

The graphics well represents that VOC (as organic volatile compounds or equivalent carbon) are the widespread pollutants and the choice of

Figure 6. Percentage of Italian IPPC plants using 'sorbents technologies'

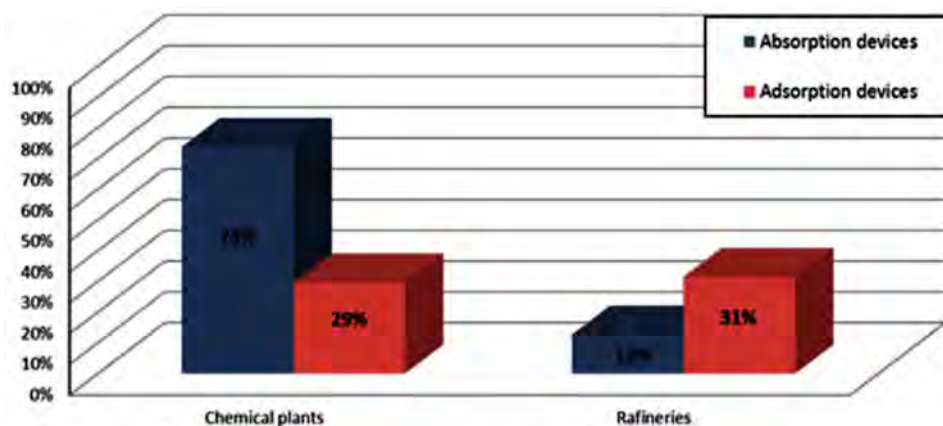


Figure 7. Operating emission data from thermal oxidizers in IPPC permitted plants

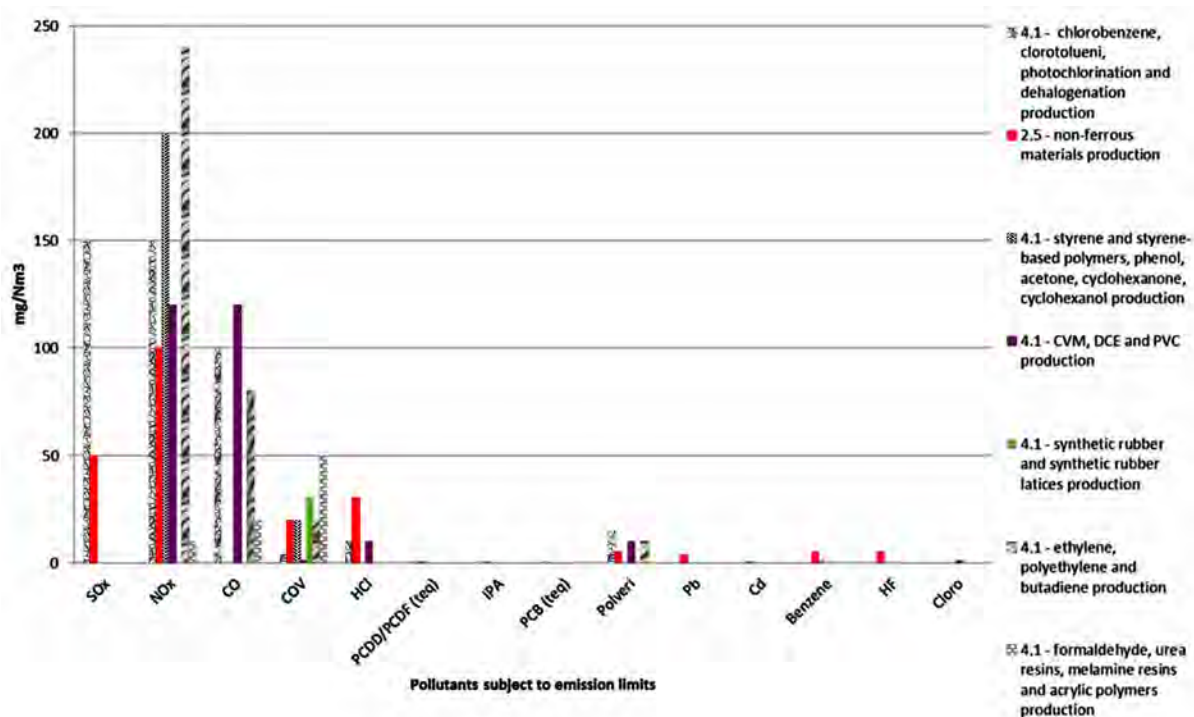


Figure 8. Operating emission data from catalytic oxidizers in IPPC permitted plants

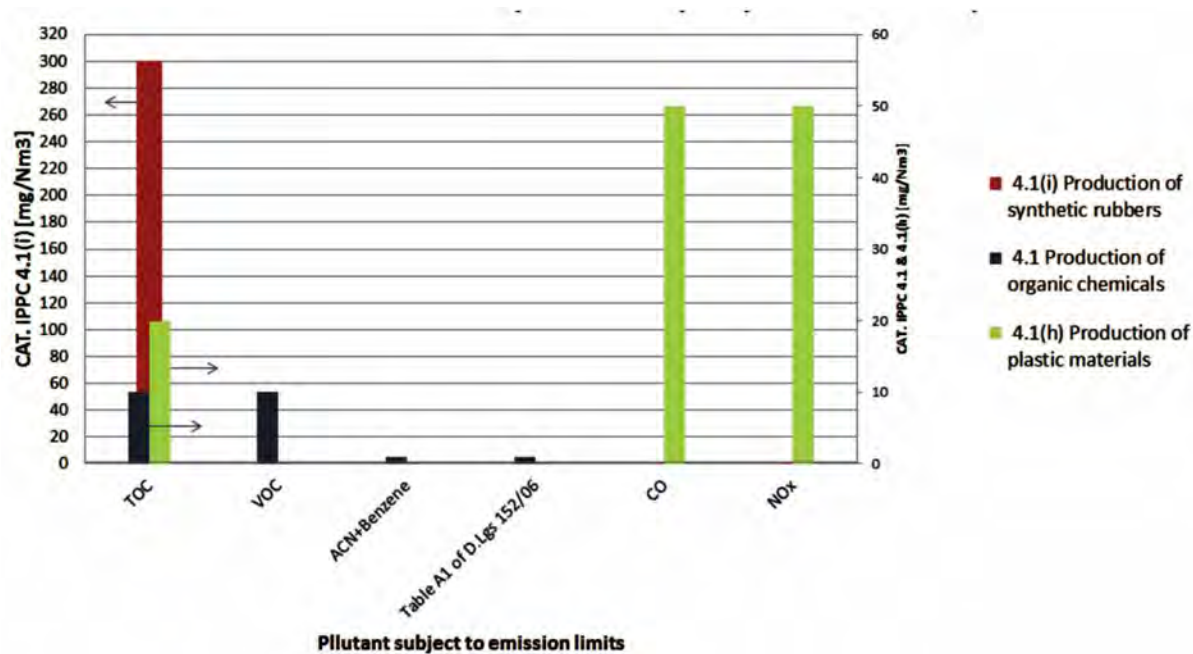


Figure 9. Operating emission data from wet scrubbing devices in IPPC permitted plants

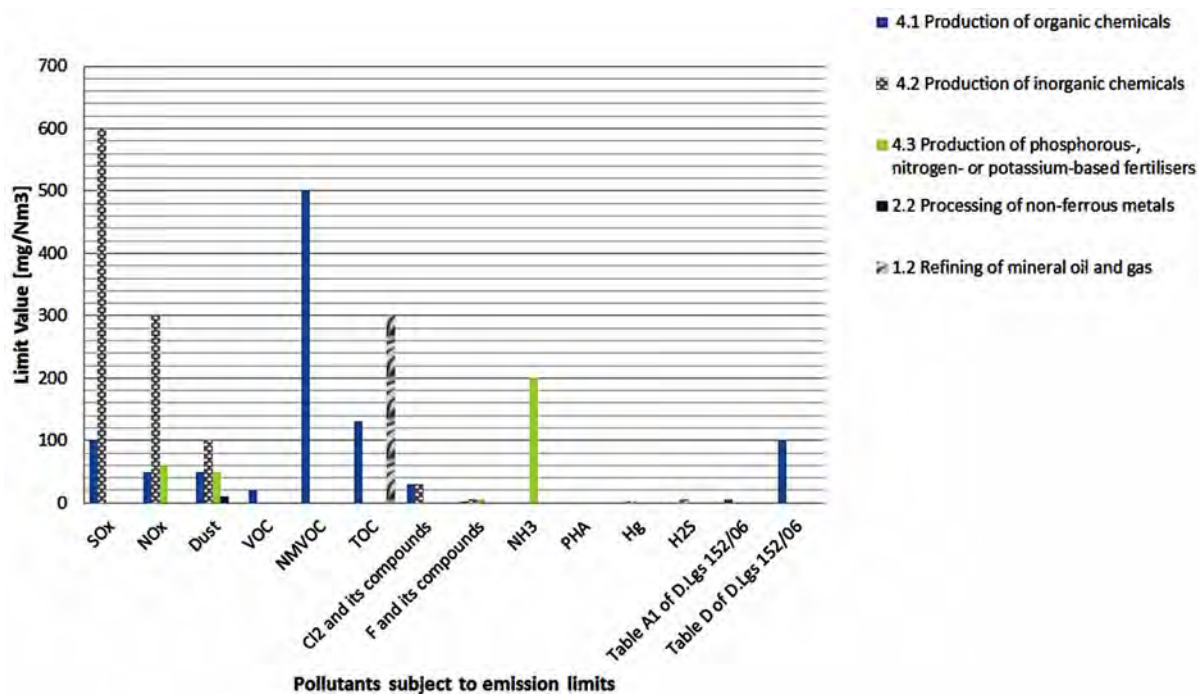
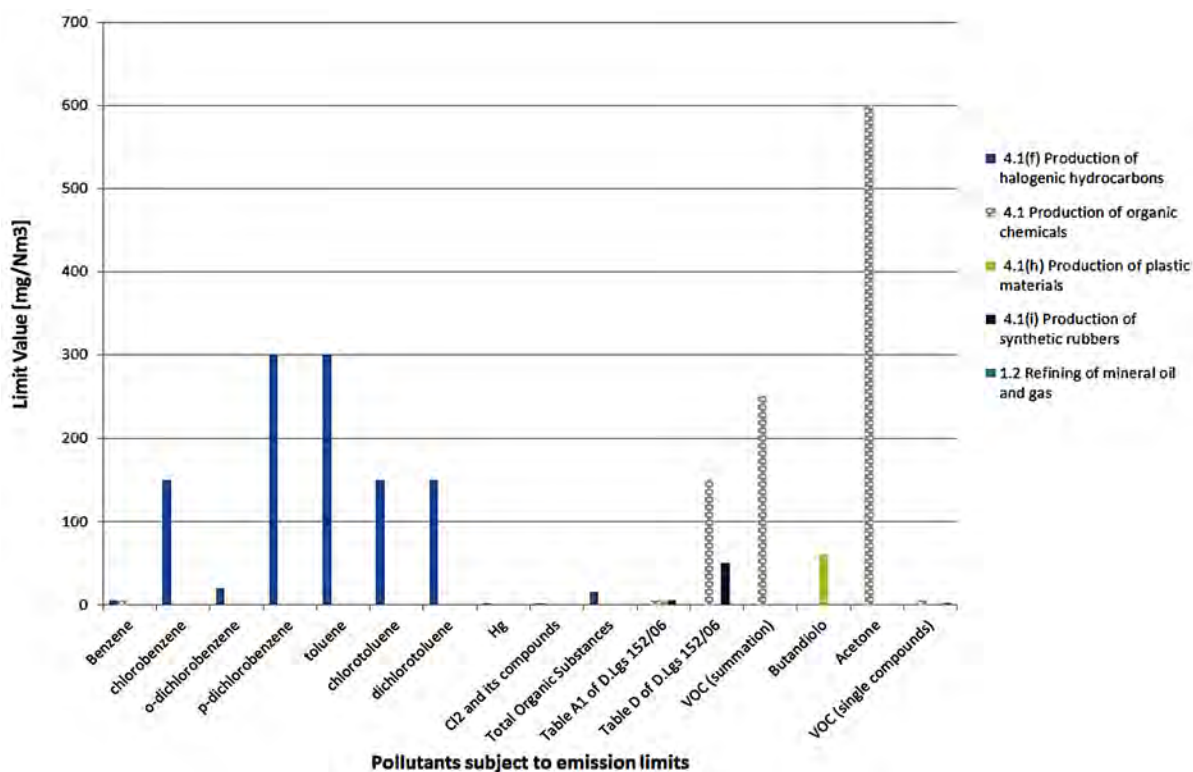


Figure 10. Operating emission data from adsorption devices in IPPC permitted plants



treatment device is strictly connected with the amount of pollutants fed in it.

In the following tables (see Tables 4, 5, 6, 7) are described operating data in 'first' AIA permits partitioned for each IPPC activity: in some cases AIA permit emission limit values are described as a range, in order to represent different limit values within these ranges fixed for different plants.

FUTURE RESEARCH DIRECTIONS

Recent regulatory innovations introduced by Italian Legislative Decree n. 46/2014 ^[9] led to the imposition, during operation of IPPC installations as specified in AIA Permits, of further specific indications as defined in new BAT for each IPPC sector, with revised emission limit values for pollutant released into the environment, as reported in "BAT Conclusions" of new reviewed BRefs.

This new approach applies to all IPPC installations subject to AIA Permits (first released) at national level, for which the new legislation requires an updating of existing authorizations, through new procedural "review/renewal" mechanisms, now closely linked to the identification of new technologies and the publication of BAT Conclusions.

In such a way, a second phase of IPPC permits is starting, after the expiring first phase, taking into account 'BAT Conclusions' issued in sectorial BRefs updating and upgraded and also other environmental aspects, as for example renewed attention towards territory inside which IPPC installations stands, based on a new evaluation of site conditions in terms of production/release of relevant hazardous substances, possibility of soil and groundwater contamination, soil vulnerabilities and all available operational data coming from annual reporting.

Table 4. AIA permits emission limits range for thermal oxidizers in IPPC permitted plants

THERMAL OXIDATION OPERATING DATA IN AIA PERMITS		
IPPC ACTIVITIES	Pollutants	AIA Permits Emission Limits Range (mg/Nm ³)
4.1 Production of organic chemicals + 4.2 Production of inorganic chemicals	SOx	150
	NOx	150
	CO	100
	VOC	4
	HCl	10
	PCDD/PCDF (as TEQ)	0,1
	PHA	0,1
	PCB (as TEQ)	0,1
	Dust	15
2.5.a Production of non-ferrous crude metals from ore, concentrates or secondary raw materials by metallurgical, chemical or electrolytic processes + 4.2 Production of inorganic chemicals	SOx	50
	NOx	100
	TOC	20
	Dust	5
	Pb	3,5
	Cd	0,2
	Benzene	5
	PCDD/PCDF	0,0005
	PHA	0,1
	HF	5
	HCl	30
4.1 Production of organic chemicals	NOx	10 - 240
	VOC	1-30
	CO	20 - 120
	Dust	10
	TOC	50
	Benzene	1
	PCDD/PCDF (come teq)	0,1 ng/Nm ³
	Chlorine	1
	CVM + DCE	0,2 ppm v/v
	HCl	10

For IPPC installations permitted at national level, these new regulations are going to finalize already existing IPPC permits by means of a ‘review and renew’ mechanism, giving also new opportunities for research, and focused mainly on these new aspects:

1. More tight respect – by now mandatory – towards the application of new technologies, to reach new and more stringent limit values, as well as synthetically defined inside ‘BAT Conclusions’ inside new revised BRefs;
2. Greater attention – to be verified from time to time – towards territorial environmental

Table 5. AIA permits emission limits range for catalytic oxidizers in IPPC permitted plants

CATALYTIC OXIDATION OPERATING DATA IN AIA PERMITS		
IPPC ACTIVITIES	Pollutants	AIA Permits Emission Limits Range (mg/Nm ³)
4.1 Production of organic chemicals	TOC	10
	VOC	10
	ACN+Benzene	1
	Table A1 of D.Lgs 152/06	1
4.1(i) Production of synthetic rubbers	TOC	300
4.1(h) Production of plastic materials	TOC	20
	CO	50
	NOx	50

Table 6. AIA permits emission limits range for wet scrubbing devices in IPPC permitted plants

ABSORPTION (WET SCRUBBING) OPERATING DATA IN AIA PERMITS		
IPPC ACTIVITIES	Pollutants	AIA Permits Emission Limits Range (mg/Nm ³)
4.1 Production of organic chemicals	SOx	100
	NOx	50
	Dust	5-50
	VOC	20
	NMVOC	500
	TOC	10-130
	Cl ₂ and its compounds	8-30
	F and its compounds	1
	Table A1 of D.Lgs 152/06	5
	Table D of D.Lgs 152/06	5-100
4.2 Production of inorganic chemicals	SOx	75-600
	NOx	100-300
	Dust	5-100
	Cl ₂ and its compounds	1-30
	F and its compounds	5
	Hg	0,05
	H ₂ S	5
4.3 Production of phosphorous-, nitrogen- or potassium-based fertilisers	NOx	50-60
	Dust	20-50
	F and its compounds	6
	NH ₃	10-200
2.2 Processing of non-ferrous metals	Dust	10
1.2 Refining of mineral oil and gas	TOC	300
	H ₂ S	'Bubble' application
	SOx	'Bubble' application

Table 7. AIA permits emission limits range for adsorption devices in IPPC permitted plants

ADSORPTION OPERATING DATA IN AIA PERMITS		
IPPC ACTIVITIES	Pollutants	AIA Permits Emission Limits Range (mg/Nm ³)
4.1(f) Production of halogenic hydrocarbons	Benzene	5
	chlorobenzene	150
	o-dichlorobenzene	20
	p-dichlorobenzene	150-300
	toluene	300
	chlorotoluene	150
	dichlorotoluene	150
	Hg	0,05
	Cl ₂ and its compounds	1
	Total Organic Substances	15
4.1 Production of organic chemicals	Benzene	5
	Table A1 of D.Lgs 152/06	1-5
	Table D of D.Lgs 152/06	150
	VOC (summation)	250
	Acetone	600
	VOC (single compounds)	5
4.1(h) Production of plastic materials	Table A1 of D.Lgs 152/06	4
	Butandiolo	60
4.1(i) Production of synthetic rubbers	Table A1 of D.Lgs 152/06	5
	Table D of D.Lgs 152/06	50
1.2 Refining of mineral oil and gas	VOC	2-20

aspects, up to now considered only marginally, as managed only inside the respect of the specific applicable standard;

3. A precise reconsideration of the implementation of adequate control and monitoring systems inside IPPC operating installations, finalized also to validate planned controls.

On account of the advantages reachable with a large use of these abatement devices, it's desirable a faster and faster development in application and correct use to achieve BAT performance and emission levels as soon as possible in a more wide part of IPPC plants.

Considering Italian Regulations, it can be suitable the definition of a specific contest for

these kind of devices due to their advantages and peculiarity.

CONCLUSION

The release of IPPC operational permits has given place to hundreds of different scenarios of best available techniques (BAT) implementation in these industrial sites, where environmental pollution is going now under a planned control in order:

1. To enforce BAT application and implementation inside new and already existing IPPC Plants;

2. To strengthen IPPC Permits via harmonization of integrated management of air, water and soil pollutants emissions an emissions to avoid different approaches towards pollution impacts management could allow pollution transfers from one environmental matrix to another.
3. To perform a properly planned and adequate monitoring and data reporting activities and, if needed, also periodic inspections based on the effective environmental risk found.

BAT application, in waste gas cleaning operations, is normally adopted in Italian IPPC chemical plants and Refineries, nevertheless mainly *Absorption systems and Adsorption systems are widely used, as well as respectively 54% and 30% of total Italian IPPC Chemical plants and Refineries* are installing these types of devices:

1. *Absorption devices* are mostly installed in Italian IPPC Chemical Plants (74% absorption vs 29% adsorption);
2. *Adsorption techniques* are mostly adopted in Refineries (31% adsorption vs 13% absorption) due to peculiarity of pollutants emitted (mainly inorganic ones in Chemical plants and organic ones in Refineries).

Considering all installed devices:

1. Absorption devices (such as wet scrubbers) are most diffused with an incident percent-

age of 55,9% relative to total number of devices installed and they are mainly used for abatement of inorganic substances (see Figure 11).

2. *Adsorption and Absorption devices* are still mostly diffused for abatement of VOC, respectively 43.28% and 31.34% of total devices analyzed in this work (see Figure 12).

For these devices operative management is relatively simple and cross media effects are at minimum compared with Thermal and Catalytic Oxidation devices (Battistella et al., 2014).

A correct use and management of waste gas cleaning devices can bring:

- An abatement of organic substances (cancer causing ones too), generally coming from vents, storage tanks and handling of liquid substances containing VOC, with a high efficiency of VOC removal and recovery in case of adsorption devices;
- A strong reduction of inorganic pollutants in the emissions from production processes with simple maintenance and minimum cross media effects (in case of Adsorption and Absorption);
- The recovery of combustion heat and generation of steam (in case of Thermal Oxidation);

Figure 11. Percentage of installed 'end of pipes abatement devices'

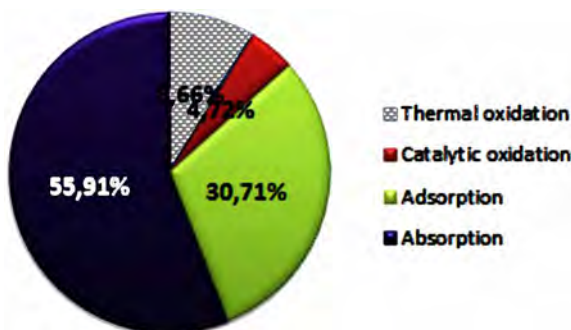
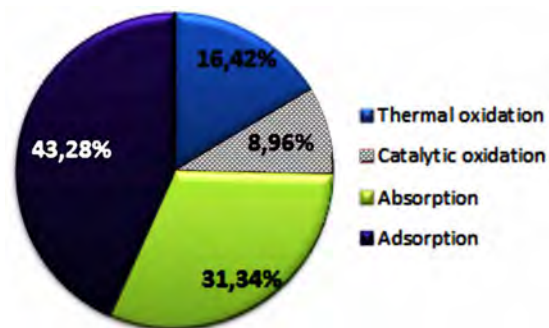


Figure 12. Percentage of installed 'end of pipes devices' for abatement of VOC



- A high thermal efficiency using recuperative and regenerative oxidation (in case of Thermal and Catalytic Oxidation).

REFERENCES

Battistella, G. (2013, October). *L'innovazione tecnologica per le attività di licensing, monitoraggio e controllo degli insediamenti IPPC soggetti ad AIA nazionale*. Paper presented at Veneto Region Educational Course 'Emissione negli impianti di processo', Mestre, Italy.

Battistella, G., & Di Marco G. (2013). *Air emissions measurements in environmental monitoring of IPPC plants*. IGI Global.

Battistella, G., & Di Marco, G. (2013, June). *Some Open Issues in Environmental Monitoring of IPPC Plants Air Emissions Measurements*. Paper presented at 4th Imeko TC19 Symposium on Environmental Instrumentation and Measurements Protecting Environment, Climate Changes and Pollution Control, Lecce, Italy.

Battistella, G., Di Marco, G., Bonaiuti, F., & Carlucci, C. (2014, March). *La Ossidazione Termica come migliore tecnica disponibile per gli Off-Gas*. Paper presented at the AIDII 20th Conference of Industrial Hygiene 'Le giornate di Corvara', Corvara, Italy.

KEY TERMS AND DEFINITIONS

Absorption: Mass transfer between a soluble gas and a solvent – often water – in contact with each other.

Adsorption: Heterogeneous reaction in which gas molecules are retained on a solid surface (adsorbent) that prefers specific compounds to others and thus removes them from effluent streams.

AIA: (Italian IPPC Permit): Operating permit released for environmental protection purposes, in order to prevent and control pollution.

BAT: Best Available Technologies.

BREF: Reference Document on Best Available Technologies.

Catalytic Oxidation: Oxidation process of combustible gases and odorants in a waste gas stream, by heating a mixture of contaminants with air or oxygen above its auto-ignition point. The gas, after passing through the flame area, passes through a catalyst bed. The catalyst has the effect of increasing the oxidation reaction rate, enabling conversion at lower reaction temperatures than in thermal oxidation units.

IMELS: Italian Minister of Environment, Land and Sea.

IPPC: Integrated Pollution Prevent and Control.

ISPRa: Superior Environmental Protection and Research Institute.

PTS: Total Suspended Particulate, the atmospheric particulate matter.

Thermal Oxidation: Oxidation process of combustible gases and odorants in a waste gas stream, by heating a mixture of contaminants with air or oxygen above its auto-ignition point in a furnace and maintaining it at high temperature for sufficient time to complete combustion to carbon dioxide and water.

TOC: Total Organic Carbon, the amount of carbon found in an organic compound.

VOC: Volatile Organic Compound, organic chemicals that have a high vapor pressure at ordinary room temperature. Their high vapor pressure results from a low boiling point, which causes large numbers of molecules to evaporate from the liquid form.

ENDNOTES

¹ Available in the web portal at the URL <http://eippcb.jrc.ec.europa.eu/reference>.

² For further technical descriptions see Section 3 of Reference Document on Best Available Techniques in Common Waste Water and Waste Gas Treatment/Management Systems in the Chemical Sector - February 2003.

³ It were analyzed only data concerning ‘first’ AIA Permits. This scenario could change because of AIA Permits provisions or modifications.

⁴ Some AIA IPPC plants adopt more than one different technique considered as BAT.

⁵ It were analyzed only data concerning first AIA Permits. The scenario could be changed because of AIA Permits provisions or modifications.

E

Category E

Ethics and Social Responsibility

The Foundation of (Business) Ethics' Evolution

E

Ben Tran

Alliant International University, USA

INTRODUCTION

Business education and higher education, according to Brink and Smith (2012), is general face criticism on several fronts and are subject to increasing scrutiny. Pringle and Michel (2007, p. 202) advised that “state legislators, parents, taxpayers, and donors want universities to justify their investments by providing evidence of student learning”. This justification seems warranted given Arum and Roksa’s (2011) compelling evidence demonstrating that undergraduate students are learning little, partly, because of the lack of rigor at institutions of higher education. In addition, possessing an MBA degree and the mastery of MBA subject matter are uncorrelated with career success (Pfeffer & Fong, 2002). Business schools are at a crossroads and it is time to seriously rethink or redesign business education (Datar, Garvin, & Cullen, 2010). The *Wall Street Journal* reported that corporate recruiters are questioning the value of the undergraduate business degree and “they’re looking for candidates with a broader academic background” (Korn, 2012). The purpose of this chapter is to cover the foundation of (business) ethics and the meaning of business ethics. In so doing, this chapter will cover the topic of accreditation in higher educational institutions, and the teaching of business ethics courses in higher educational institutions.

FOUNDATION OF BUSINESS EDUCATION AND ACCREDITATION IN HIGHER EDUCATION INSTITUTIONS

In light of criticisms regarding business education in higher education institutions, it would be prudent for business schools to assure their stakeholders of quality and accountability. Accreditation is one method of holding a program or institution accountable and demonstrating that the program/institution meets at least a minimum quality threshold. The Council for Higher Education Accreditation (CHEA) defines accreditation as “a process of external quality reviewer created and used by higher education to scrutinize colleges, universities and programs for quality assurance and quality improvement” (Eaton, 2011, p. 1). Accreditation serves several roles, two of which include “assuring quality and “engendering private sector confidence” (Eaton, 2011, p. 2-3). CHEA indicates that “accreditation in the United States is about quality assurance and quality improvement. It is a process to scrutinize higher education institutions and programs” (Eaton, 2011, p. 11).

The goal of CHEA is to assure “that accrediting organizations contribute to maintaining and improving academic quality” (Eaton, 2011, p. 9). CHEA’s role is to review and scrutinize the quality and effectiveness of accreditors and *recognize*

them. CHEA does not accredit institutions or programs, rather, CHEA accredits that accreditors. CHEA recognizes sixty institutional and programmatic accrediting organizations, including three [levels (gold, silver, and bronze)] that accredit business programs: the Association to Advance Collegiate Schools of Business (AACSB) International, the Accreditation Council for Business Schools and Programs (ACBSP), and the International Assembly for Collegiate Business Education (IACBE). Hence, a higher educational institution receiving regional accreditation does not necessarily translate to the same higher educational institution receiving one of the three levels of accreditation for its business program.

Currently, there are six regional accrediting agencies for educational institutions in the United States:

1. Middle States Association of Colleges and Schools [(Middle States Association, MSACS, or MSA) which covers educational institutions in New York, New Jersey, Pennsylvania, Delaware, Maryland, the District of Columbia, Puerto Rico, and the US Virgin Islands, as well as schools for American children in Europe, North Africa, and the Middle East),
2. New England Association of Schools and Colleges, Inc. [(NEASC) which covers educational institutions in the six New England states (Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, and Vermont)],
3. Higher Learning Commission (formerly part of the North Central Association of Colleges and Schools and covers educational institutions in Arkansas, Arizona, Colorado, Iowa, Illinois, Indiana, Kansas, Michigan, Minnesota, Missouri, North Dakota, Nebraska, New Mexico, Ohio, Oklahoma, South Dakota, Wisconsin, West Virginia, and Wyoming),
4. Northwest Accreditation Commission (NWAC), formerly known as the Northwest

Association of Accredited Schools, is for primary and secondary schools and Northwest Commission on Colleges and Universities (NWCCU) for postsecondary institutions in Alaska, Idaho, Montana, Nevada, Oregon, Utah, and Washington,

5. Southern Association of Colleges and Schools [(SACS) which covers educational institutions in Alabama, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, Texas and Virginia], and
6. Western Association of Schools and Colleges [(WASC) which covers educational institutions in California, Hawaii, Guam, American Samoa, Micronesia, Palau, and Northern Marianas Islands, as well as schools for American children in Asia].

Attaining and maintaining accreditation may help a business program distinguish itself based on quality. However, accreditation requires a substantial financial investment. Roberts, Johnson, and Groesbeck (2004, p. 112) indicated that “the annual incremental cost increase for even a small school...and can easily exceed \$500,000”.

Standards of Business Ethics

Among the three levels (bronze, silver, and gold) that accredit business programs (AACSB, ACBSP, and IACBE), only two levels are international accreditation bodies for higher education business schools, the AACSB and the ACBSP (Franks & Spalding, 2013). Both AACSB and ACBSP require business schools to incorporate ethics into their curricula, but between these two levels of accreditation, the standards put forth by the AACSB are more detailed than those put forth by the ACBSP. Under the AACSB approach, ethics education is required as part of the general knowledge and skills portion of the standards for undergraduates, and in the management-specific portion of the standards for undergraduate and master's students. Ethics education under this

regime involves four subject areas: (1) responsibility of business and society, (2) ethical leadership, (3) ethical corporate governance, and (4) ethical decision-making ([Association to Advance Collegiate Schools of Business (AACSB, 2004). Under the ACBSP approach, ethical issues are expected to be addressed within several contexts. For example, business ethics is considered to be part of a member school's education in regard to the impact of business on society. In addition, the standards require that ethical awareness and global awareness be included in the student skills that are assessed by member schools. Business ethics is also included as part of the common professional component of a typical business school curriculum [Association of Collegiate Business Schools and Program (ACBSP, 2011)].

Despite these accreditation requirements, the content of business ethics is not specifically defined or circumscribed by either of the major business school accreditation bodies. In their glossary section, for example, the ACBSP provides a definition of ethical behavior that refers to how organization insures that all its decision, actions and stakeholder interactions confirmed to the organization's moral and professional principals. Even though the ACBSP standards acknowledge that an organization's "moral principles" define right and wrong, the standards do not require instruction in moral theory *per se*, and the standards do not offer any guidance as to how those moral and professional principles are or ought to be derived [International, the Accreditation Council for Business Schools and Program (ACBSP, 2011, p. 62)]. Similarly, the AACSB standards suggest that students should be provided with the opportunity to learn and practice "multiple models for ethical decision making" as an important step in supporting their ability to make good personal choices and business decision in the future [Association to Advance Collegiate Schools of Business (AACSB, 2004, p. 18)]. As an assurance of learning guideline, the AACSB offers as possible learning goal the ability of students to "identify an ethical dilemma in a scenario case and apply an

ethics model of framework to propose and defend a resolution" (AACSB, 2007, p. 7).

The Accreditation of Business Education

All three business program accreditors, according to Brink and Smith (2012), are international in scope and, not surprisingly, the age of the accreditor is related to the number of programs it accredits. The AACSB is the oldest of the business school accreditors, founded in 1916 and is also the largest of the accreditors, the ACBSP was founded in 1988, and the IACBE is the newest of the accreditors, founded in 1997. The accreditation guidelines provided by all three accreditors are mission-based. The AACSB changed to mission-linked standards in 1991, whereas the ACBSP and IACBE were mission-based since inception. Many scholars have suggested that the emergence of the ACBSP as a competitor may have been partially responsible for AACSB's change to mission-linked standards (Lowrie & Willmott, 2009; McKenna, Cotton, & Van Auken, 1997; Ramey, 1993). The accreditation guidelines proffered by the three accreditors are similar with respect to the overall content and underlying principles. Juan and Ofori-Dankwa (2006) described several similarities in the mission focus of the three accreditors. Furthermore, Ramey (1993) suggested that ACBSP is a clone of AACSB and Lowrie and Willmott (2009, p. 412) stated that the mission linked approach of AACSB is "emulated by its national competitors, the ACBSP...and IACBE".

First, the accreditation guidelines of the ACBSP and IACBE, in particular, are very similar to one another, perhaps because both accrediting bodies were founded by the same person: Dr. John L. Green, Jr. (IACBE, 2011). Though the contents of the accreditors' guidelines are similar, the rigidity and rigor of the guidelines vary. Second, accreditors differ in the rigor of the faculty qualification requirements. The ACBSP's definition of *academically qualified* (AQ) faculty is similar to the IACBE's definition of *doctorally qualified*

(DQ) faculty, and they both have similar definitions of *professionally qualified* (PQ) faculty. However, the IACBE definitions could be considered less rigorous because they are *suggested principles* rather than *required standards*. The ACBSP and IACBE allow for a greater variety of pathways to becoming A/DQ and PQ compared to the AACSB. Therefore, the AACSB has the most rigorous, and the IACBE has the least rigorous definitions of A/DQ and PQ (Brink & Smith, 2012).

Third, the required percentages of qualified faculty vary. The AACSB, according to Brink and Smith (2012), requires at least 90% of the faculty to be AQ or PQ and 50% to be AQ. The ACBSP criteria vary by undergraduate and graduate programs. The ACBSP requires that at least 80% of undergraduate and 90% of graduate credit hours be taught by AQ or PQ faculty. Therefore, the ACBSP's requires that percentages for overall AQ and PQ faculty is less rigorous than the AACSB's. The ACBSP requires that at least 40% of undergraduate and 70% of graduate credit hours be taught by AQ faculty, and 100% of doctorate credit hours be taught by AQ faculty. The ACBSP faculty qualification standards regarding the percentage of AQ faculty required at the graduate level are more rigorous than the AACSB's standards. Overall, the AACSB has the most rigorous requirements regarding the necessary percentages of qualified faculty, and the IACBE has the least rigorous requirement, offering the principle that there should be at least one full-time DQ faculty for each program. Fourth, accreditors differ with respect to the rigor of research or scholarly activity requirements. The AACSB expects a significant portion of intellectual contributions to be peer-reviewed scholarship, whereas the ACBSP and IACBE are more open to other forms of scholarly activity. In fact, the ACBSP explicitly acknowledges that the AACSB is the premier accreditor for research-oriented programs (Brink & Smith, 2012). In addition, Roller, Andrews, and Bovee (2003) surveyed faculty from AACSB-, ACBSP-, and IACBE-accredited business programs and found that all three groups of faculty rated AACSB

the highest with respect to promoting excellence in research.

Meanings of Business Ethics

Business ethics is a form of applied ethics (Velentzas & Broni, 2010) that examines ethical principles and moral or ethical problems that arise in a business environment (Solomon, 1991). It applies to all aspects of business conduct (Baumhart, 1968; Ferrell, Fraedrich, 1977; Singer, 1991) and is relevant to the conduct of individuals and business organizations as a whole (Bernard, 1972; Donaldson, 1982, p. 36). Applied ethics is a field of ethics that deals with ethical questions in many fields such as technical, legal, business and medical ethics (Preston, 1997, p. 6-11). Business ethics consists of a set of moral principles and values (Jones, Parker, & Bos, 1005, p. 17) that govern the behavior of the organization with respect to what is right and what is wrong (Badiou, 2001; Seglin, 2003). It spells out the basic philosophy and priorities of an organization in concrete terms (French, 1979, 1995). It also contains the prohibitory actions at the workplace (Collier, Esteban, 1007, p. 19; Duska, 1999). It provides a framework on which the organization could be legally governed. With time, certain moral philosophies have helped in the evolution of four basic concepts of ethics. They are deontology, relativism, egoism, and utilitarianism (Velentzas & Broni, 2010). Business ethics is the behavior that a business adheres to in its daily dealings with the world (Borgerson & Schroeder, 2008).

The ethics of a particular business can be diverse (Solomon, 1983). They apply not only to how the business interacts with the world at large, but also to their one-on-one dealings with a single customer (Solomon, 1991). As such, if a company does not adhere to business ethics and breaks the laws, they usually end up being fined (Drucker, 1981). Many companies have broken anti-trust, ethical and environmental laws and received fines worth millions (Velasquez, 1983). The problem is that the amount of money these companies

are making outweighs the fines applied (Green, 1991). The profits blind the companies to their lack of business ethics, and the money sign wins (De George, 1999). Hence, in the increasingly conscience-focused marketplaces, the demand for more ethical business processes and actions, known as ethicism, is increasing (Donaldson, 1982, p. 32). Simultaneously, pressures is applied on industry to improve business ethics through new public initiatives and laws. Thus, business ethics can be both a normative and a descriptive discipline (Abrams, 1954). As a corporate practice and a career specialization, the field is primarily normative. In academic descriptive approaches are also taken. The range and quantity of business ethical issues reflects the degree to which business is perceived to be at odds with non-economic social values. Historically, interest in business ethics accelerated dramatically during the 1980s and 1990s, both within major corporations and within academia (Cory, 1005, p. 11).

The Historical Philosophy of Ethics

As a discrete, self-conscious academic discipline, business ethics is roughly four decades old (Velentzas & Broni, 2010). Richard DeGoerge (2005), on the other hand, dates academic business ethics to the 1970s, identifying Baumhart as a forerunner to a self-conscious academic business ethics. Raymond Baumhart's (1961, 1963, 1968) groundbreaking studies in the 1960s are generally understood to be early contributions to business ethics. Furthermore, prominent contemporary business ethicist Norman Bowie (1999, 2005) dates the field's first academic conference to 1974. Although academic instruction explicitly devoted to the relationship between ethics and commerce can be found in U.S. business schools as early as the first three decades of the 20th century, particularly in Catholic colleges and universities, creation of academic positions dedicated explicitly to business ethics in U.S. business schools tracks closely waves of corporate scandal from the 1980s to the present. Academic business ethicists address

questions that range across the functional areas of business, giving rise to various recognized specialties in business ethics. But despite the wide range of questions pursued, the bulk of the academic literature and discussion is focused more closely on the large corporation whose ownership shares are traded on public exchanges. In this broad sense, ethics in business is simply the application of everyday moral or ethical norms to business (Bennett, 2003; Boylan, 1995), and a philosophical paradigm shares a similar long tradition (Capaldi, 2006; Velentzas & Broni, 2010, pp. 45-56).

FUTURE RESEARCH DIRECTIONS

There is no doubt that business ethics is an important part of the higher education landscape, however, business ethics is a relatively contemporary subject. It was not officially recognized as an applied science until after an academic conference held at the University of Kansas in 1974. The subsequent development of the *Journal of Business Ethics and Business Ethics Quarterly* resulted (DeGeorge, 2005). "Recent years have seen exponential growth in the use of terms such as corporate social responsibility, sustainability, going green, sustainable development, social entrepreneurship, corporate ethics, the triple bottom line, as well as a variety of other terms that imply that businesses have obligations beyond the maximization of profit (AACSB International, N. D.). According to the Aspen Institute, an international non-profit organization dedicated to fostering values-based leadership, the percentage of MBA programs that require at least one ethics, sustainability, business and society or corporate social responsibility course has increased from 34% in 2001 to 54% in 2005 (Aspen Institute, 2005). Ethical missteps such as the savings and loan crises in the late 1980s and the accounting scandals marking the entrance to the 21st century have buttressed the need for instruction in business ethics both outside of the workplace and in higher education. Multiple authors, according to Hasnas

(2013), have cited instances in which courses in business ethics at university or business schools do not have profound transformative effects on student's attitudes about ethics (Wynd & Mager, 1989) or on perceptions of ethical behavior (Davis & Welton, 1991), or do these classes affect their future professional lives (Pamental, 1989). Kraft and Singhapakdi (1991) found that business ethics taught at the undergraduate or even at the graduate level were of little significance in an individuals' overall ethics training.

CONCLUSION

According to Duska (2013), Hasna (2013), and Koehn (2005), teaching ethics is always a challenge, because most students spend most of their time collecting facts and learning how the world works without truly understanding it. Through most of their education, students attain academic success by demonstrating the ability to understand or discover factual information and reproduce it on demand—to read, to remember, and to repeat. But ethics has no determinate body of facts to absorb and retain. Its focus is not on discovery but on evaluation. The skills it requires are the analytical ones of perceiving the relationships among assertions, recognizing inconsistencies, tracing implications, and judging the cogency of arguments—skills that a significant number of students have never developed. If this situation makes teaching ethics a challenge, then, teaching it in a business school is doubly so. Not only business students, but most business school faculty are completely unfamiliar with the techniques of ethical analysis (Duska, 2013; Hasna, 2013; Koehn, 2005). They are trained in quantificational methods and are often gifted empirical problem-solvers. They are skilled at determining the most effective means of attaining specified goals. But the conceptual tools needed to determine which goals are proper to pursue are entirely new to them. The lack of familiarity with the nature of ethical enquiry that is typical of business school students and faculty presents those charged with

teaching ethics in business schools with four major challenges (Hasnas, 2013): 1) the challenge of definition, 2) the challenge of abstraction, 3) the challenge of cultural relativism, and 4) the challenge of integration.

REFERENCES

AACSB International. (n.d.). *Ethics/sustainability research center*. Retrieved on October 15, 2015, available at <http://www.aacsb.edu/resources/ethics-sustainability/about.asp>

Abrams, F. (1954). Management responsibilities in a complex world. In T. H. Carroll (Ed.), *Business education for competence and responsibility*. Chapel Hill, NC: University of North Carolina Press.

Arum, E., & Roksa, J. (2011). *Academically adrift: Limited learning on college campuses*. Chicago, IL: The University of Chicago Press.

Aspen Institute. (2005). *Beyond the grey pin-stripes*. Retrieved on October 15, 2015, available at <http://www.beyondgreypinstripes.org>

Association of Collegiate Business Schools and Program. (2011). *ACBSP standards and criteria for demonstrating excellence in baccalaureate/graduate degree schools and programs*. Retrieved on October 14, 2015, available at <http://www.acbsp.org/download.php?sid=29>

Association to Advance Collegiate Schools of Business. (2004). *Ethics education in business schools: Report of the ethics education task force to AACSB international's board of directors*. Retrieved on October 14, 2015, available at <http://www.aacsb.edu/publications/researchreports/archives/ethics-education.pdf>

Association to Advance Collegiate Schools of Business. (2007). *AACSB assurance of learning standards: An interpretation*. Retrieved on October 14, 2015, available at http://www.aacsb.edu/publications/whitepapers/AACSB_Assurance_of_Learning.pdf

- Badiou, A. (2001). *Ethics: An essay on the understanding of evil*. London: Verso.
- Baumhart, R. (1961). How ethical are businessmen? *Harvard Business Review*, 39(4), 6–9.
- Baumhart, R. (1963). *Exploratory study of businessmen's views on ethics and business* (Doctoral dissertation). Harvard Business School, Harvard University, Boston, MA.
- Bennett, C. (2003). *Ethics in business*. Mason, OH: South-Western.
- Bernard, W. (1972). *Morality*. Harper & Row.
- Borgerson, J. L., & Schroeder, J. E. (2008). Building an ethics of visual representation: Contesting epistemic closure in marketing communication. In M. P. Morland & P. H. Werhane (Eds.), *Cutting edge issues in business ethics* (pp. 87–108). Boston: Springer. doi:10.1007/978-1-4020-8401-0_8
- Bowie, N. E. (1999). *Business ethics: A Kantian perspective*. Oxford, UK: Blackwell Publishers.
- Bowie, N. E., & Werhane, P. H. (2005). *Management ethics*. Malden, MA: Blackwell Publishers.
- Boylan, M. (1995). *Ethical issues in business*. Wadsworth Publishing.
- Brink, K. E., & Smith, C. A. (2012). A comparison of AACSB, ACBSP, and IACBE accredited U.S. business programs: An institutional resource perspective. *Business Education & Accreditation*, 4(2), 1–16.
- Capaldi, N. (2006). What philosophy can and cannot contribute to business ethics. *Journal of Private Enterprise*, 22(2), 68–86.
- Collier, J., & Esteban, R. (2007). Corporate social responsibility and employee commitment. *Business Ethics (Oxford, England)*, 16(1), 19–33. doi:10.1111/j.1467-8608.2006.00466.x
- Datar, S. M., Garvin, D. A., & Cullen, P. G. (2010). *Rethinking the MBA: Business education at a crossroads*. Boston, MA: Harvard Business Press.
- Davis, J., & Welton, R. (1991). Professional ethics: Business students perceptions. *Journal of Business Ethics*, 10(6), 451–463. doi:10.1007/BF00382829
- DeGeorge, R. T. (1999). *Business ethics* (5th ed.). Prentice Hall.
- DeGeorge, R. T. (2005). *History of business ethics*. Paper presented at Global Business Ethics Conference: The Accountable Corporation, Santa Clara, CA.
- Donaldson, T. (1982). *Corporations and morality*. Englewood Cliffs, NJ: Prentice Hall.
- Drucker, P. (1981). What is business ethics? *The Public Interest*, (63), 18–36.
- Duska, R. F. (1999). Employee rights. In R. E. Frederic (Ed.), *A companion to business ethics* (pp. 257–268). Oxford, UK: Blackwell. doi:10.1002/9780470998397.ch21
- Duska, R. F. (2013). What's the purpose of ethics education? *Expositions*, 7(1), 42–52.
- Eaton, J. S. (2011). *An overview of U.S. accreditation*. Retrieved on October 14, 2015, available at <http://www.chea.org/pdf/Overview%20of%20US%20Accreditation%2003.2011.pdf>
- Ferrell, O., & Fraedrich, J. (1997). *Business ethics: Ethical decision making and cases*. Houghton Mifflin.
- Franks, R. A., & Spalding, A. D. (2013). Business ethics as an accreditation requirement: A knowledge mapping approach. *Business Education & Accreditation*, 5(1), 17–30.
- French, P. A. (1979). The corporation as a moral person. *American Philosophical Quarterly*, 16(3), 207–215.
- French, P. A. (1995). *Corporate ethics*. Fort Worth, TX: Harcourt Brace.
- Green, R. (1991). When is “everyone’s doing it” a moral justification? *Business Ethics Quarterly*, 1(1), 75–93.

- Hasnas, J. (2013). Teaching business ethics: The principles approach. *Journal of Business Ethics Education*, 10, 275–304. doi:10.5840/jbee20131014
- International Assembly for Collegiate Business Education (IACBE). (2011). *Dr. John L. Green, Jr. named president emeritus of the IACBE*. Retrieved on October 15, 2015, available at <http://www.IACBE.org/news-releases-president-emeritus.asp>
- Jones, C., Parker, M., & Bos, R. (2005). *For business ethics: A critical text*. London: Routledge. doi:10.4324/9780203458457
- Julian, S. D., & Ofori-Dankwa, J. C. (2006). IS accreditation good for the strategic decision making of traditional business schools? *Academy of Management Learning & Education*, 5(2), 225–233. doi:10.5465/AMLE.2006.21253788
- Koehn, D. (2005). Transforming our students: Teaching business ethics post-Enron. *Business Ethics Quarterly*, 15(1), 137–151. doi:10.5840/beq20051517
- Korn, M. (2012, April 5). Wealth or waste? *Rethinking the value of a business major*. *The Wall Street Journal*, p. B1.
- Kraft, K., & Singhapakdi, A. (1991). The role of ethics and social responsibility in achieving organizational effectiveness: Students versus managers. *Journal of Business Ethics*, 10(9), 679–686. doi:10.1007/BF00705874
- Lowrie, A., & Willmott, M. P. (2009). Accreditation sickness in the consumptions of business education: The vacuum in AACSB standard setting. *Management Learning*, 40(4), 411–420. doi:10.1177/1350507609335851
- McKenna, J. F., Cotton, C. C., & Van Auken, S. (1997). The new AACSB accreditation standards: A prospect of tiering? *Journal of Organizational Change Management*, 10(6), 411–502. doi:10.1108/09534819710190166
- Pamentel, G. (1989). The course in business ethics: Can it work? *Journal of Business Ethics*, 8(7), 547–551. doi:10.1007/BF00382930
- Pfeffer, J., & Fong, C. T. (2002). The end of business schools? Less success than meets the eye. *Academy of Management Learning & Education*, 1(1), 78–95. doi:10.5465/AMLE.2002.7373679
- Preston, D. (1997). Can business ethics really exist? *Computers & Society*, (March), 6–11.
- Pringle, C., & Michel, M. (2007). Assessment practices in AACSB—accredited business schools. *Journal of Education for Business*, 82(4), 202–211. doi:10.3200/JOEB.82.4.202-211
- Ramey, G. W. (1993). Guest editorial—AACSB: Weathering the storm. *Journal of Organizational Change Management*, 6, 8–13.
- Roller, R. H., Andrews, B. K., & Bovee, S. L. (2003). Specialized accreditation of business schools: A comparison of alternative costs, benefits, and motivations. *Journal of Education for Business*, 78(4), 197–204. doi:10.1080/08832320309598601
- Seglin, J. L. (2003). *The right thing: Conscience, profit and personal responsibility in today's business*. Rollinsford, NH: Cambridge University Press.
- Vandivier, K. (1983). Case study—The aircraft brake scandal. In T. Donation & P. Werhane (Eds.), *Ethical issues in business: A philosophical approach* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.
- Velentzas, J., & Broni, G. (2010). Ethical dimensions in the conduct of business: Business ethics, corporate social responsibility and the law. The ethics in business as a sense of business ethics. *International Conference On Applied Economics 2010*, Athens, Greece.
- Wynd, W., & Mager, J. (1989). The business and society course: Does it change student attitudes? *Journal of Business Ethics*, 8(6), 487–491. doi:10.1007/BF00381815

ADDITIONAL READING

- Anderson, D. (1999). Business ethics and the pragmatic attitude. In R. Frederick (Ed.), *A companion to business ethics* (pp. 56–64). Oxford: Blackwell. doi:10.1002/9780470998397.ch5
- Boatright, J. R. (1999). Does business ethics rest on a mistake? *Business Ethics Quarterly*, 9(4), 583–591. doi:10.2307/3857936
- Buchholz, R., & Rosenthal, S. (2001). A philosophical framework for case studies. *Journal of Business Ethics*, 29(1/2), 25–31. doi:10.1023/A:1006438824678
- Cragg, W. (1997). Teaching business ethics: The role of ethics in business and in business education. *Journal of Business Ethics*, 16(3), 231–245. doi:10.1023/A:1017974908203
- Donaldson, T. (2012). The epistemic fault line in corporate governance. *Academy of Management Review*, 37(2), 256–271. doi:10.5465/amr.2010.0407
- Falkenberg, L., & Woiceshyn, J. (2008). Enhancing business ethics: Using cases to teach moral reasoning. *Journal of Business Ethics*, 79(3), 213–217. doi:10.1007/s10551-007-9381-9
- Felton, E., & Sims, R. (2005). Teaching business ethics: Targeted outputs. *Journal of Business Ethics*, 60(4), 377–391. doi:10.1007/s10551-004-8206-3
- Greenberg, J. (1990). Organizational justice: Yesterday, today, and tomorrow. *Journal of Management*, 16(2), 399–432. doi:10.1177/014920639001600208
- Greenberg, J., & Bies, R. (1992). Establishing the role of empirical studies of organizational justice in philosophical inquiries into business ethics. *Journal of Business Ethics*, 11(5-6), 433–444. doi:10.1007/BF00870555
- Headland, T. (1990). Introduction: A dialogue between Kenneth Pike and Marvin Harris. In T. Headland, K. Pike, & M. Harris (Eds.), *Emics and etics: The insider/outsider debate* (pp. 13–27). Newbury Park: Sage.
- Izzo, G., Langford, B., & Vitell, S. (2006). Investigating the efficacy of interactive ethics education: A difference in pedagogical emphasis. *Journal of Marketing Theory and Practice*, 14(3), 239–248. doi:10.2753/MTP1069-6679140305
- James, H. Jr, & Cohen, J. (2004). Does ethics training neutralize the incentives of the Prisoners Dilemma? Evidence from a classroom experiment. *Journal of Business Ethics*, 50(1), 53–61. doi:10.1023/B:BUSI.0000020869.42655.f4
- Norman, W. (2011). Business ethics as self-regulation: Why principles that ground regulations should be used to ground beyond-compliance norms as well. *Journal of Business Ethics*, 102(1), 43–57. doi:10.1007/s10551-011-1193-2
- Quinn, D., & Jones, T. (1995). An agent morality view of business policy. *Academy of Management Review*, 20(1), 22–42.
- Quinton, A. (1964). Contemporary British philosophy. In D. O'Connor (Ed.), *A critical history of western philosophy* (pp. 530–556). New York: The Free Press. doi:10.1007/978-1-349-08051-9_29
- Roca, E. (2007). Introducing practical wisdom in business schools. *Journal of Business Ethics*, 82(3), 607–620. doi:10.1007/s10551-007-9580-4
- Sen, A. (1993). Does business ethics make economic sense? *Business Ethics Quarterly*, 3(1), 45–54. doi:10.2307/3857381
- Sims, R. (2004). Business ethics teaching: Using conversational learning to build an effective classroom learning environment. *Journal of Business Ethics*, 49(2), 201–211. doi:10.1023/B:BUSI.0000015782.99051.d4

Trevino, L., & Victor, B. (1992). Peer reporting of unethical behavior: A social context perspective. *Academy of Management Journal*, 35(1), 38–64. doi:10.2307/256472

Velasquez, M. (2012). *Business ethics: Concepts and cases* (7th ed.). Boston: Pearson.

Victor, B., & Stephens, C. U. (1994). Business ethics: A synthesis of normative philosophy and empirical social science. *Business Ethics Quarterly*, 4(2), 145–155. doi:10.2307/3857486

Webley, S., & Werner, A. (2008). Corporate codes of ethics: Necessary but not sufficient. *Business Ethics (Oxford, England)*, 17(4), 405–415. doi:10.1111/j.1467-8608.2008.00543.x

Zych, J. (1999). Integrating ethical issues with managerial decision-making in the classroom: Product support program decisions. *Journal of Business Ethics*, 18(3), 255–266. doi:10.1023/A:1005871423821

KEY TERMS AND DEFINITIONS

AACSB: The Association to Advance Collegiate Schools of Business.

Academic Business Ethicists: Address questions that range across the functional areas of business, giving rise to various recognized specialties in business ethics.

ACBSP: Association of Collegiate Business Schools and Program.

Accreditation: A process of external quality reviewer created and used by higher education to scrutinize colleges, universities and programs for quality assurance and quality improvement.

Business Ethics: A form of applied ethics that examines ethical principles and moral or ethical problems that arise in a business environment.

CHEA: Council for Higher Education Accreditation.

IACBE: International Assembly for Collegiate Business Education.

Higher Learning Commission: Formerly part of the North Central Association of Colleges and Schools and covers educational institutions in Arkansas, Arizona, Colorado, Iowa, Illinois, Indiana, Kansas, Michigan, Minnesota, Missouri, North Dakota, Nebraska, New Mexico, Ohio, Oklahoma, South Dakota, Wisconsin, West Virginia, and Wyoming.

Middle States Association of Colleges and Schools: Commonly known as Middle States Association, MSACS, or MSA, which covers educational institutions in New York, New Jersey, Pennsylvania, Delaware, Maryland, the District of Columbia, Puerto Rico, and the US Virgin Islands, as well as schools for American children in Europe, North Africa, and the Middle East.

New England Association of Schools and Colleges, Inc.: Also known by the abbreviation NEASC, covers educational institutions in the six New England states (Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, and Vermont).

Northwest Accreditation Commission (NWAC): Formerly known as the Northwest Association of Accredited Schools, is for primary and secondary schools and Northwest Commission on Colleges and Universities (NWCCU) for postsecondary institutions in Alaska, Idaho, Montana, Nevada, Oregon, Utah, and Washington.

Southern Association of Colleges and Schools[(SACS): Covers educational institutions in Alabama, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, Texas and Virginia.

WASC: Western Association of Schools and Colleges (WASC) covers educational institutions in California, Hawaii, Guam, American Samoa, Micronesia, Palau, and Northern Marianas Islands, as well as schools for American children in Asia.

Integrating Sustainability and CSR in the Value Chain of the Information Technology Sector

E

Patricia Martínez García de Leaniz
University of Cantabria, Spain

María Elena García Ruiz
University of Cantabria, Spain

INTRODUCTION

Actually, all industrial sectors are aware of the importance of sustainability in all its business actions by articulating corporate social responsibility (CSR) strategies. The CSR concept refers to the set of commitments that a company acquires to manage its economic, social and environmental impacts and to support the objective of maximizing profits while generating benefits for society, particularly for its stakeholders. Due to their negative impacts in the natural environment in the information technology (IT) sector the interest in CSR and sustainability is even greater (Carter & Rogers, 2008; Dao et al., 2011).

Many international initiatives show the growing importance of socially responsible and sustainable aspects in the IT sector. For instance, the International Telecommunications Union (UIT) is the United Nations specialized agency for ITs. One of the focal points of the UIT is IT, sustainability and climate change. The Global e-Sustainability initiative (GeSI) was born in order to promote sustainable development in the IT sector. This initiative promotes those technologies that allow sustainable development and it is dedicated exclusively to IT sustainability through innovation.

In today's rapidly changing market, IT companies are seeking greater performance, better features and more flexibility at the lowest price (Hervani et al., 2005). Despite the economic recession the IT sector is seeing exceptional

growth. According to CompTIA, the IT sector Trade Association, this sector generated 3.7 \$ trillion income in 2014 with growth rates of 5% for 2015. In this context, demand going forward will be determined by evolving views on perceived value and consumer tastes. In this regard, customers are demanding more out of IT firms that simply a quality product at a low price. The goal for IT companies is to compete on the market through offering distinguished sustainable products (McAfee & Brynjolfsson, 2008). IT companies should be able to create and develop new and innovative products while operating in a sustainable and responsible way. By implementing socially responsible initiatives, IT businesses can build sustainability into the operations and management. This study is aimed at exploring the importance of sustainability and corporate social responsibility in the IT sector.

BACKGROUND

Corporate Social Responsibility

An increasingly important aspect of corporate management in recent decades has been the incorporation of CSR, a construct that emphasizes the obligation of companies to integrate social and environmental parameters into their modus operandi and their long-term strategies. Although some publications present CSR as a new construct,

DOI: 10.4018/978-1-5225-2255-3.ch277

the idea that companies must manage their social and environmental obligations has its roots in a much older debate (Dejean and Gond, 2004). Extant literature in economic and organizational theory has reviewed the role of business in society, and all of them agree that the purpose of business should be broadened beyond only economic benefits (Friedman, 1970) and that a social dimension should be incorporated into corporate performance.

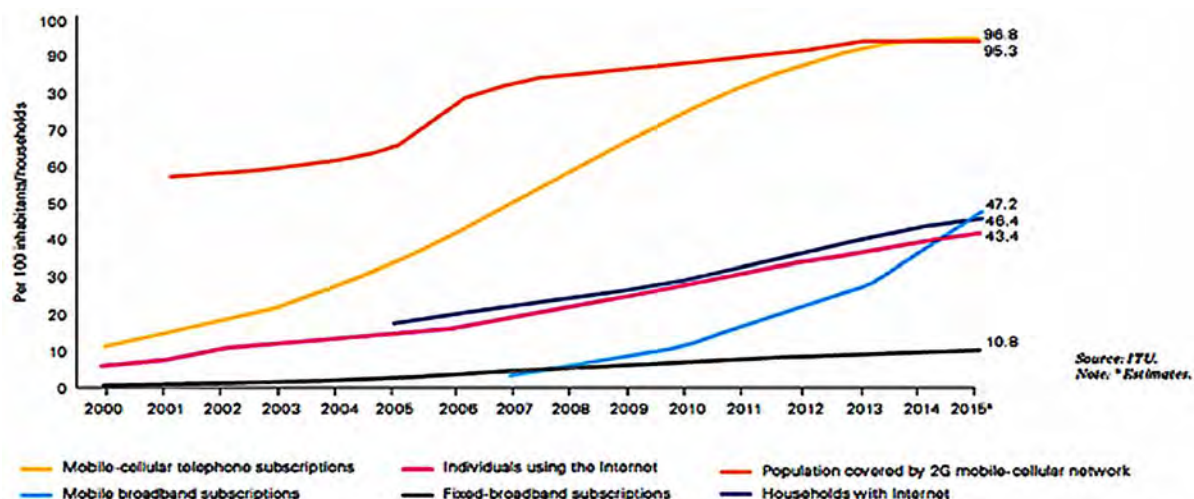
The CSR concept refers to the set of commitments that a company acquires to manage its economic, social and environmental impacts and to support the objective of maximizing profits while generating benefits for society, particularly for its stakeholders. However, one question may arise: What kind of actions can be considered as socially responsible? Depending on the business activity, the sector or country in which the company operates, the applicable law can change and, therefore, also the actions that may be considered as socially responsible. However, all CSR actions have as a starting point to meet the expectations of stakeholders minimizing the risks associated with their activity (other than financial) and “positivizing” the impacts of their activity in the economic, social and environmental sphere.

IT Sector

The IT sector has transformed the way people live, work and learn. From mobile phones and micro-computer chips to the Internet, IT has consistently delivered innovative products and services that are now an integral part of everyday life. Technological progress, infrastructure deployment, and falling prices have brought unexpected growth in IT access and connectivity to billions of people around the world. IT has systematically increased productivity and supported economic growth across both developed and developing countries (McAfee & Brynjolfsson, 2008). Globally 3.200 million people are using the Internet of which 2.000 million are from developing countries. By the end of 2015, there are expected more than 7.000 million mobile cellular subscriptions worldwide (less than 1.000 million in 2000). As shown in Figure 1, mobile broadband subscriptions is the most dynamic market segment with 12 times higher penetration than in 2007 (penetration rate 97%) in comparison with fixed-broadband (11% penetration expected in 2015) (ICT Facts & Figures, 2015).

Figure 1. Growth trend in the IT industry

Source: ICT Facts & Figures (2015)



IT Sector's Value Chain

Before delving into the key issues for a sustainable and responsible management in the IT sector we have to explore the agents involved in the value chain of this industry. As Porter explains (1985) a value chain is a set of activities that a firm operating in a specific industry performs in order to deliver a valuable product or service to the market. The IT value chain involves several agents and activities, from the creation and edition of the content to the service provision and final delivery to users. According to the European Commission (1997) the key players in the value chain of the IT sector are: Manufacturers and integrators of equipment and devices, content and applications developers, network operators, distributors and users. Figure 2 graphically shows the generic value chain that is established for the provision of IT services to end users.

- **Manufacturers and Integrators of Equipment and Devices:** These agents are designers and producers of IT hardware. It is possible to distinguish between manufacturers of equipment (modems, servers, routers, gateways and network interfaces) and handset manufacturers (terminals for transmitting and/or receiving information).
- **Content and Applications Developers:** These actors create the content and adapt it to compatible formats for applications according to user terminal devices. They also develop intermediary applications between the device and network hardware.

- **Network Operators:** They provide, operate and maintain the infrastructure that allows the circulation and transmission of information and content from the sender to the receiver in an effective and efficient manner.
- **Service Providers:** Originally, the provision of IT services was closely linked to the existence of a particular infrastructure or network. It is no longer the case and users can enjoy a wide variety of access services accessing their content provider regardless of their network operator.
- **Distributors:** They are responsible for the distribution and sale of terminal devices to end users. Often, service providers and device manufacturers themselves act as distributors of these devices, so in these cases it is not easy to identify these agents.
- **Users:** They are the final consumers of information services accessing these services through appropriate devices.

CSR ASPECTS IN THE IT VALUE CHAIN

Porter and Kramer (2006) suggest that the value chain of firms can be used to show how CSR can impact companies' strategy. It is essential to explore the structure of the IT value chain and its composition to make an analysis of the impacts and responsibilities of those agents involved. Table 1 summarizes some of the key CSR aspects to consider for each actor in the IT value chain for

Figure 2. Major players in the IT value chain

Source: Lemonche and de la Cuesta (2010)

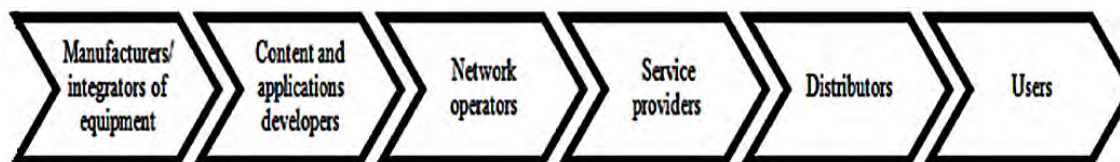


Table 1. CSR aspects for each agent in the IT value chain

Critical Aspects According to CSR Criteria				
Agent	Service/Product	Economy	Society	Environment
Manufacturers and integrators of equipment and devices	Equipment and devices	Quality and reliability of equipment.	Life cycle of products. Delocalized production in developing countries.	Delocalized production in developing countries. Communication in emergencies.
Content and applications developers	Software products	Copyright vs free download. Proprietary software vs. open software.	-	User safety. Child protection. Accessibility.
Network operators	Connectivity services	Network reliability. Communications quality.	Logistics and removal of network equipment. Management equipment suppliers.	Scope of service in excluded areas (universal access).
Service providers	Access and value added services	Customer care.	-	Digital divide. IT literacy. Accessibility.
Distributors	Equipment and devices	Quality vs price.	Distribution logistics. Withdrawal of obsolete equipment.	User safety. Distribution in depressed areas.

Source: Lemonche and de la Cuesta (2010)

each CSR are described in the triple bottom line: economic, environmental and social.

Economic Aspects

Beyond maximizing business benefit associated with each actor in the IT value chain, it is necessary to take into account the specific aspects of their activity, their role in the sector and the needs and expectations of their stakeholders. Thus, a manufacturer or distributor of equipment and devices should pay attention to the quality and reliability of its products and the right quality/price ratio. If, on the contrary, it is a network operator the challenge will be to ensure its customers a high degree of reliability network (understood as the degree of service continuity or absence of cuts, signal degradation or incidents perceived by users). This can be expressed as “quality communication” with regard to both the end user and any other agent whose business depends on the network. In this sense, the greater the impact of a possible lack of communication in the business, the

higher must be the degree of network reliability, for which different levels of service quality and price are designed.

The software and content providers face the debate between the protection of intellectual property rights and the open provision of software and content via Internet to facilitate universal access to information. All actors in the IT value chain must responsibly manage their own value chain and provide good after-sales service.

Environmental Aspects

It is known that all business activities have environmental impacts and so it is in the IT sector (Vachon & Klassen, 2007). However, environmental management in IT firms is especially important for those actors who work directly with physical products (manufacturers, operators and distributors) compared to those other agents whose “products” are basically information and knowledge (content developers and service providers). The IT sector is based mainly on the production

and use of electronic equipment, from the largest infrastructure (telephone exchanges, network data servers or satellite communication stations) to the smallest mobile phone.

Therefore, the responsible and sustainable management of the life cycle of these equipments is key to the proper environmental management in the IT industry (Jørgensen, T.H., 2008). The R+D+I of IT firms seek to develop equipment and devices that achieve maximum energy efficiency with minimum consumption of materials and optimal recovery cycle. In this sense, the environmental responsibility of manufacturers or producers is high and so it is recognized in several legislative initiatives. Environmental responsibility of network operators and distributors will be to maintain the ideal life cycle of these products, selecting less polluting equipment and more responsible suppliers, optimizing energy consumption and contributing to successful recovery of products at the end of their useful life.

The delocalized production in developing countries (mainly Asian) and international supply chains of IT equipments must always maintain high standards of environmental management beyond borders and local laws (Zhu et al., 2005).

Social Aspects

Social issues of the IT value chain are perhaps where it is possible to find the greatest number of synergies among the different actors and at the same time where each of them can provide a more differentiated value. There are indisputable common goals, such as those aimed at defending human rights in the global supply chain, improving the accessibility to information services and combating the digital divide, among others. In addition, public commitments such as declarations of principles or codes of conduct must be accompanied by control and verifications to ensure a real responsible management of these aspects.

The responsibility of IT companies is evident not only in the adequate supply of equipment and provision of IT services to those who can use them,

but also keeping the focus on those individuals and communities that do not have access to ITs and therefore see their quality of life very limited. Each IT agent can provide here their own solutions for the provision of equipments and communication systems in emergency solutions, the extension of communication networks in depressed areas or the research on accessibility solutions.

Finally, we cannot forget that while ITs can greatly benefit individuals they also have some risks related to misuse or abuse that are important to consider. Beyond the necessary security guarantees for the health and safety of users, it is essential for IT actors to manage with prudence the effects of information and communication services in especially sensitive groups such as children. IT actors directly involved in the provision of these services must exercise their responsibility, always keeping in mind their business strategies to their potential users and providing the best controls against abuse, supporting parents and educators in their educational role and always protecting those most in need.

CSR TOOLS FOR IT BUSINESSES

In this section we will deepen into the responsibility of IT firms and the available mechanisms for them to exercise this responsibility. It is necessary to bear in mind that the IT sector differs from other industries in that, in addition to the direct impacts (economic, social and environmental) this sector may affect the impact of other sectors. To explore these issues, we will base our arguments on the impacts classification provided by Lemonche and de la Cuesta (2010). According to these authors, the IT industry has three pathways to impact on economy, society and environment:

- *Direct Impacts* that an IT company has in the development of its activity, including the social and economic impacts on its stakeholders and throughout its value chain.

- *Enablers Impacts*, since the application of technology allows companies in other sectors improve their social and environmental impacts.
- *Systemic Impacts* on global social issues such as inclusion, inequality or access to information.

Thus, we conclude that IT companies have two ways to approach their CSR strategy:

- Practicing their CSR as a company measuring the impacts of their activity.
- Promoting a social and responsible use of their products. That is, promoting the role of the IT sector itself as an engine for social change.

This is the big difference between the IT industry and other business sectors: the possibility of having two different ways to face the challenge of CSR and sustainability. The first course of action is based on using the common goals of companies to improve their responsible management. The second one consists on promoting the responsibility of the IT sector, developing technological improvements in other social areas such as education or health. Below we present some of the CSR approaches considering social, environmental and economic impacts of ITs.

We delve into the systemic impacts of ITs through their effects on other social areas. As a transversal element of all these synergies are green ITs (Sabbaghi and Vaidyanathan, 2012). Among the green IT applications which are more attractive to promote systemic changes we emphasize the dematerialization, virtualization and cloud computing models. These are applications that provide direct and enabling effects but at the same time present a significant potential for changing habits of energy consumption, which forms the basis for the aforementioned systemic effects. Some of the points where IT companies can positively influence to a greater extent are:

- **Health:** The World Health Organization (WHO) recognizes the potential of ITs to achieve more effective health services and improve access to health care. In this regard, the WHO emphasizes the crucial role that technology can provide to the management and access to health services in the provision of health care, health education and communication.
- **Digital Inclusion:** The Digital Agenda for some countries (2013) has established a plan for digital inclusion and employability as a result of contributions of multiple stakeholders (public and private) coordinated with the common goal of improving the quality of life of citizenship and enhancing the competitiveness and positioning of small and medium size companies through the use of ITs.
- **Education:** The emerging “information society” with the globalization context in which we live today exert a great influence, clearly and directly, in the world of education and learning. Within this social scenario, ITs play a key role for the following reasons. First, educational institutions have been subjected to a process of updating in light of new social expectations. Subsequently, they have had to implement new methodological and pedagogical techniques in accordance to the integrated use of new technologies. Secondly, ITs are tools that enable new mechanisms for knowledge transfer.
- **Transparency and Good Governance:** Transparency and participatory democracy are constituent factors of good governance whether in the public, private or institutional setting. The term “good governance” involves new forms of participation of civil society. One of these new ways of civic participation is access to information and thereby proclaiming the right to transparency. ITs are vital to enable all these emerging principles and rights.

- **Sustainability:** According to the “2020 SMART” report (2008), the IT sector is vital in the fight against the climate change, which could allow the reduction of CO2 emissions in 7.8 GT by 2020. This reduction would come from the incorporation of technological improvements in those sectors with the greatest environmental impact. The IT industry has, for instance, the ability to reduce the need to transport people and goods and to optimize energy consumption in lighting systems or factories or buildings.
- **Development Cooperation:** Within the United Nations “Millennium Declaration” is included the specific goal to develop a global partnership (in cooperation with the private sector) to ensure the benefits of new technologies, especially information and communication technologies. This demonstrates the commitment to boost new methodologies as a means to improve the quality of the initiatives in development cooperation.

CSR Management in Delocalized Production

All companies make purchases and not all of them have the same risks. We must remember two of the premises of CSR: that is voluntary and goes beyond legislation. To choose which types of initiatives could be implemented companies must first analyze the following points: (1) The ability to influence the behavior of suppliers and (2) the risks associate with the countries providing the purchases (in terms of environmental protection, corruption or human rights for instance. To minimize the risks associated with purchasing materials outside our borders, we will highlight simple actions that do not require great investment.

A useful starting point could be the elaboration of codes of conduct, codes of purchased or supplier codes which specify the conditions that our suppliers have to meet. As an example IT firms can

ask suppliers if they meet the recommendations of the ILO on child or forced labor. Similarly, firms can request that purchased materials meet some specific certification that verifies the behaviors of suppliers. Firms can also require certifications in some CSR domain such as employees’ working conditions or environmental performance, among others. There is also the possibility of adopting international standards as enforceable and exclusive criteria that suppliers should fulfill, certifiable standards in environmental legislation or socially responsible certified management systems. Some of the most widespread certifiable CSR standards are:

- **SGE21: 2008:** It is the first CSR management system that allows audit processes and achieves a certification in ethical and socially responsible management. This certificate is based on consolidated models such as those related to quality or environmental aspects, but enriching them through a multi-stakeholder vision.
- **Social Accountability SA8000:** Social Accountability International (SAI) is an international non-profit (multi-stakeholder) organization dedicated to developing and implementing socially responsible standards to improve working conditions. It is recognized worldwide and is one of the most prestigious certifications in labor rights.

Internal Management of CSR

Internal CSR affects the management of people within the organization (Dao et al., 2011). It is something that goes beyond human resources management, since the impact of the activity on the lives of employees is part of the social dimension of CSR. IT firms can improve the performance of the company on internal CSR by implementing plans for equal opportunities for male and female workers, designing plans to integrate people at risk of exclusion within the company or promoting

corporate volunteering among employees (giving business hours to collaborate in a local project for instance) (Wirtenberg et al., 2007). In this sense, Colbert and Kurucz (2007) argue that by linking human resources principles to companies' CSR objectives, firms can create lasting industry advantage and build capabilities for positive change.

Environmental Management

Some tools to measure environmental impact are the following international standards that may be interesting for IT firms:

- **ISO/TS 14067: 2013:** Greenhouse gases and carbon footprint of products: It establishes principles, requirements and guidelines for the quantification and communication of carbon footprints on the basis of the emission and removal of greenhouse gases during the life cycle a product.
- **ISO 14046:** Environmental management – Water footprint: This standard establishes the principles, requirements and guidelines for a correct evaluation of the water footprint of products, processes and organizations, from the analysis of their life cycle.

In addition, the IT sector has specific tools to manage their environmental impacts. We highlight the “Toolkit on environmental sustainability in the ICT sector”¹¹. This is an initiative of the ITU that provides comprehensive and detailed information for IT companies on how to achieve sustainability in their operations and management of the organization, through the practical application of international standards and guidelines.

CSR and Reporting

CSR and sustainability reports have been strengthening as a communication tool in the IT sector as evidence of the commitment of this sector to

sustainable development. We detail below the main characteristics that a CSR report must present:

- **Transparency:** It should include both positive and negative information and honesty about the compliance of objectives.
- **Materiality:** The information contained in the report should cover those aspects and indicators that reflect social, environmental and economic significant impacts of the organization.
- **Customization:** The CSR strategy of each IT company should be “unique” as its sustainability report.
- **Performance Indicators:** IT companies have to choose the most appropriate indicators to reflect what they want to communicate. Performance indicators give firms a baseline from which they can make a comparison between planned targets and performance achieved.

FUTURE RESEARCH DIRECTIONS

The authors of this chapter hope that this research will prove to both scholars and practitioners the value of CSR initiatives for the IT industry. Future studies could explore the implementation of these initiatives in each of the key agents forming the value chain of the IT sector (from manufactures and integrators of equipment and devices to network operators and distributors) in order to identify best practices and develop effective strategies of benchmarking. Therefore, enablers and obstacles could be identified in order to put into practice CSR activities.

CONCLUSION

Value chain is increasingly recognized as a key component of corporate responsibility. Managing

the social, environmental and economic impacts of supply chains makes good business sense as well as being the right thing to do. However, the IT value chain consists of continuously evolving markets and relationships. In this chapter, the authors offer practical steps that IT companies can take as the basis to introduce sustainability and CSR aspects into their value chain.

There are numerous reasons why IT companies are willing to introduce sustainability in their value chains. Primary among them it is to ensure compliance with laws and regulations and to adhere to and support international principles for sustainable business conduct. In addition, IT companies are increasingly taking actions that result in better social, economic and environmental impacts because society expects this and because there are business benefits to doing so. By managing and seeking to improve environmental, social and economic performance and good governance throughout value chains, IT companies act in their own interest the interests of stakeholders and the interest of society at large.

In this chapter the authors have explored CSR tools available to the IT sector, taking into account that there are three ways to influence the economy, society and environment distinguishing between direct, enables and systemic impacts. Through these actions and considering these impacts, IT companies can have positive influences on health, digital inclusion, education and transparency, among others. Therefore, the authors highly recommend that IT firms should pay special attention to various fundamental aspects such as their delocalization policies, internal CSR management, environmental management and communication, with the ultimate goal of capturing all their CSR strategy in the sustainability report that ought to contain references, among others, to transparency, customization and performance indicators.

REFERENCES

- Carter, C. R., & Rogers, D. S. (2008). A framework of sustainable supply chain management: Moving toward new theory. *International Journal of Physical Distribution & Logistics Management*, 38(5), 360–387. doi:10.1108/09600030810882816
- Colbert, B. A., & Kurucz, E. C. (2007). Three conceptions of triple bottom line business sustainability and the role for HRM. *Human Resource Planning*, 30(1), 21–29.
- Dao, V., Langella, I., & Carbo, J. (2011). From green to sustainability. Information technology and an integrated sustainability framework. *The Journal of Strategic Information Systems*, 20(1), 63–79. doi:10.1016/j.jsis.2011.01.002
- Dejean, F., & Gond, J. P. (2004). The corporate social responsibility: Strategic issues and research methodologies. *Finance. Contrôle et Stratégie*, 7(1), 5–31.
- Desruelle, P., & Stančík, J. (2014). Characterizing and comparing the evolution of the major global economies in information and communication technologies. *Telecommunications Policy*, 38(8/9), 812–826. doi:10.1016/j.telpol.2014.04.010
- European Commission. (1997). *Green paper on the convergence of the telecommunications, media and information technology sectors and the implications for regulation*. Retrieved December 15, 2015, from: http://www.ab.gov.tr/files/ardb/evt/1_avrupa_birligi/1_6_raporlar/1_2_green_papers/com1997_green_paper_on_convergence_of_telecommunications_and_media_sectors.pdf
- Friedman, M. (1970, September 13). The social responsibility of business is to increase its profits. *The New York Times Magazine*.
- Hervani, A. A., Helms, M. M., & Sarkis, J. (2005). Performance measurement for green supply chain management. *Benchmarking: An International Journal*, 12(4), 330–353. doi:10.1108/14635770510609015

- ICT Facts & Figures. (2015). *The world in 2015*. Telecommunication Development Bureau. International Telecommunication Union. Retrieved September 3, 2015, from: <http://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2015.pdf>
- Jørgensen, T. H. (2008). Towards more sustainable management systems: Through life cycle management and integration. *Journal of Cleaner Production*, 16(10), 1071–1080. doi:10.1016/j.jclepro.2007.06.006
- Lemonche, P., & de la Cuesta, M. (2010). *Impactos de las TIC desde una perspectiva de RSC: Aspectos clave para la gestión responsable de las TIC. Responsabilidad Corporativa y Sostenibilidad*. Cuaderno Red de Cátedras Telefónica.
- McAfee, A., & Brynjolfsson, E. (2008). Investing in the IT that makes a competitive difference. *Harvard Business Review*, 86(7), 98–107. PMID:18271321
- Miller, B., & Atkinson, R. D. (2014). *Raising European Productivity Growth through ICT*. The Information Technology and Innovation Foundation. Retrieved October 6, 2015, from: <http://www2.itif.org/2014-raising-eu-productivity-growth-ict.pdf>
- Porter, M. E. (1985). *Competitive advantage: Creating and sustaining superior performance*. New York: Simon and Schuster.
- Porter, M. E., & Kramer, M. R. (2006). Strategy and society: The link between competitive advantage and corporate social responsibility. *Harvard Business Review*, 84, 12. PMID:17183795
- Sabbaghi, A., & Vaidyanathan, G. (2012). Green information technology and sustainability: A conceptual taxonomy. *Issues in Information Systems*, 13(2), 26–32.
- Stonebraker, P. W., Goldhar, J., & Nassos, G. (2009). Weak links in the supply chain: Measuring fragility and sustainability. *Journal of Manufacturing Technology Management*, 20(2), 161–177. doi:10.1108/17410380910929600
- The Climate Group. (2008). *SMART 2020: Enabling the low carbon economy in the information age*. Retrieved October 12, 2015, from: http://www.smart2020.org/_assets/files/02_smart-2020Report.pdf
- United Nations Organization. (2000). *Millennium Declaration*. Retrieved December 20, 2015, from: <http://www.un.org/en/development/devagenda/millennium.shtml>
- Vachon, S., & Klassen, R. D. (2007). Supply chain management and environmental technologies: The role of integration. *International Journal of Production Research*, 45(2), 401–423. doi:10.1080/00207540600597781
- Wirtenberg, J., Harman, J., Russell, W., & Fairfield, K. (2007). HR's role in building a sustainable enterprise: Insights from some of the world's best companies. *Human Resource Planning*, 30(1), 10–20.
- Zhu, Q., Sarkis, J., & Geng, Y. (2005). Green supply chain management in China: Pressures, practices and performance. *International Journal of Operations & Production Management*, 25(5), 449–468. doi:10.1108/01443570510593148
- ## ADDITIONAL READING
- Martinuzzi, A., Kudlak, R., Faber, C., & Wiman, A. (2011). CSR activities and impacts of the ICT sector. RIMAS Working Papers (5). Vienna University of Economics and Business.
- Schipper, I., & de Haan, E. (2005). *CSR issues in the ICT hardware manufacturing sector. SOMO ICTSectorReportAmsterdam*. Centre for Research on Multinational Corporations.

KEY TERMS AND DEFINITIONS

Corporate Social Responsibility: It is about how companies manage the business processes to produce an overall positive impact on society. This concept goes beyond compliance and engages in actions that appear to further some social good, beyond the interests of the firm and that which is required by law (e.g., waste and pollution reduction processes, educational and social programs...).

Information Technology: It is the application of telecommunications equipment to store, retrieve, transmit and manipulate data in the context of a business.

Sustainability: It is a concept that describes a human development aims at satisfying present needs without compromising the capacity of future generations to satisfy their own needs.

Value Chain: It is a set of activities that a firm operating in a specific industry or sector performs in order to deliver a valuable product or service for the market. The IT value chain involves several agents and activities, from the creation and edition of the content to the service provision and final delivery to users.

E

ENDNOTES

¹ <http://gesi.org>

² <http://www.itu.int/en/about/Pages/default.aspx>

The Morality of Reporting Safety Concerns in Aviation

Kawtar Tani

UCOL, New Zealand

INTRODUCTION

Previous investigation into the morality of actions suggested that individuals' level of cognitive moral development, strongly influences their decisions regarding what is right or wrong, and focuses upon the rights, duties and obligations involved in a particular ethical situation. Using the cognitive moral development framework, this research sought to explore the moral reasoning behind aviation employees' intentions to report wrongdoing in the aviation context. Specifically, looking at whether an association between participants' intentions to intervene in a wrongdoing situation, and their level of moral reasoning exists. This research covers one of the important issues that affect safety management in organisations, and provides opportunities on how it can be improved.

BACKGROUND

The role of employees in the prevention and detection of wrongdoing within organisations has been widely acknowledged (Camerer, 2001; Sawyer, 2005; KMPG, 2008). Up to 80% of incidents are believed to be unreported within organisations (NZALPA, 2005). It can be argued that by reporting wrongdoing, employees could provide information that can be used proactively by organisations to improve safety practice at the workplace. However, viewpoints about reporting vary considerably. Individual values, corporate values, cultural environment and personal benefits are but some of the few elements that decide whether the action of reporting is an important source of

information (Kohn & Kohn, 1988), or whether it is a disloyal act deserving punishment (Lewis, 2001). Whatever la raison d'être of reporting may be, it can be argued that reporters of wrongdoing are highly devoted to their cause, determined to defeat silence and act in a way that they perceive right. Such individuals find themselves balancing conflicting loyalties, obligations and values none of which are right or wrong (Dehn, 1999). There is a conflict between the right to privacy and the right to know; employer and colleagues' loyalty vis à vis the good citizen's duty to uphold the law; and the conflict between individualism and being a team player (Camerer, 2001). Set against these imperatives is the fear of being disloyal and losing the trust of the employer and colleagues, the fear of being wrong, and the fear of accepting responsibility for one's actions and their resulting consequences (Dehn, 1999).

MAIN FOCUS OF THE STUDY

There are two frameworks within which under-reporting of aviation wrongdoing could be explored: a moral framework, and an evolutionary framework. The moral framework assumes that the person witnessing wrongdoing does what they believe to be the 'right thing', within the limits of their understanding of right and wrong. In other words, participants' responses could be related to their level of moral development. Kohlberg's (1976) and Gilligan's (1982) cognitive moral development theories are two prominent theories which propose that individuals' progress in stages of moral reasoning and tend to operate at

a particular stage at any point in time. Kohlberg's cognitive stages of moral development propose that an individual's level of cognitive moral development strongly influences their decision regarding what is right or wrong and focuses upon the rights, duties and obligations involved in a particular ethical situation. Contrarily, Gilligan defines ethical issues mainly in terms of helping others and minimising harm and argues that moral behaviour results from meeting one's obligations and responsibilities to others.

As the level of people's moral judgment can be raised by education (Kohlberg, 1984; Oderman, 2002; Peters, 2015), the first aim of this study was to investigate aviation employees' level of moral reasoning when confronted with wrongdoing situations. If a relationship between reporting intentions and moral development is suggested, then ethics education may be one means by which under-reporting in aviation can be reduced. The following hypothesis was therefore proposed:

Hypothesis 1: Aviation employees indicating they would intervene upon becoming aware of aviation wrongdoing operate at a higher level of moral development than aviation employees who would not intervene in a wrongdoing situation.

The evolutionary framework in which to further explore under-reporting of aviation wrongdoing is that of inclusive fitness. Although people are believed to operate at a particular level of moral reasoning (Flanagan, 1984; Krebs, Vermeulen & Carpendale, 2014), it seems reasonable that their behaviour may be affected by the extent to which the person committing the act of wrongdoing has a close relationship with the person witnessing wrongdoing. Broadly in line with the arguments of Hamilton's (1964) biological concept of inclusive fitness, individuals help those to whom they are related in preference to those with whom they share no genetic ties. In the context of the study reported here, it might be expected to find that

when witnessing wrongdoing, the witness might be less likely to report a 'friend' or a 'family member', than a person with whom they share no such relationship. The second hypothesis was therefore:

Hypothesis 2: Whether participants felt they had to take action in the situation presented to them would differ according to the level to which they are related to the person responsible for wrongdoing.

The primary aims of the current study were therefore: i) to investigate the level of moral reasoning of aviation employees, when faced with evidence of wrongdoing, in relation to their intentions to intervene in wrongdoing situations, and ii) to investigate the effect of the relationship between the witness and the act of wrongdoing in aviation.

METHOD

Participants

One hundred and seventy-five (175) participants were recruited from 142 aviation organisations within New Zealand: flying clubs (91), flight training schools (15), aviation associations (5), air services (16), and air charters (15). Of these participants, 163 held a pilot's license at the time of participation.

Materials and Design

Overview

A scenario based study was designed to investigate participants' moral reasoning and the effect of the relationship between observer and protagonist when evaluating wrongdoing situations in the aviation industry. Using nine separate scenarios, participants were asked what they would do upon

becoming aware of wrongdoing and then to explain their decisions in the context of established theories of moral reasoning.

Investigation of the effect of the relationship between the observer and the protagonist when evaluating wrongdoing situations in the aviation industry was therefore a within-subjects experimental design. Investigation of the relationship between participants' level of moral reasoning and whether they chose to act upon hearing of wrongdoing was correlational design.

Nine scenarios were developed from accident and incident reports published on the Civil Aviation Authority of New Zealand (CAA NZ) website (www.caa.govt.nz). Real-life examples of accidents and incidents were chosen as the basis for developing the scenarios for the current study as they represented the kind of situations that aviation employees are known to have encountered in real life. The accident and incident reports were selected so that the scenarios used in the study could be short, easy to understand, cover different examples of wrongdoing, and be manipulated as required for the purpose of the current study. Any identifiable features in the original reports were omitted from the scenarios (e.g., names, places and dates).

To explore whether aviation employees' actions and reasoning behind these actions were affected by the level to which they personally were related to the protagonist, the scenarios were further divided into three groups. Each group contained three scenarios worded so that the relationship between the participant and the protagonist would be: i) participant not related to the protagonist (e.g., the wrongdoing behaviour was committed by someone that the participant does not personally know); ii) participant related to the protagonist (e.g., the wrongdoing behaviour was committed by a colleague); and iii) participant closely related to the protagonist (e.g., the wrongdoing behaviour was committed by a friend or a family member).

In an attempt to reduce the possibility that the level of relatedness in each scenario would interact with the perceived seriousness of the wrongdoing

described in the scenario, the wrongdoing behaviour in each of the nine scenarios was assigned a seriousness rating of 1 to 10 (where 1 = not very serious, and 10 = very serious) by the researcher, an academic with an interest in the study, and an independent third party. By summing the three seriousness ratings for each scenario, an overall seriousness rating was then calculated for each scenario (possible scores range from 3 to 30, with higher scores indicating higher perceived seriousness). The nine scenarios were arranged in descending order of seriousness and then manipulated so that the total seriousness rating for each level of relatedness (not related = scenarios 1, 2 & 7, related = scenarios 3, 5 & 6, closely related = scenarios 4, 8 & 9) were as similar as possible. The seriousness scores for the scenarios were score 21 for scenarios 1 & 8, score 20 for scenarios 5, 6, 7 & 9, score 14 for scenario 4, and score 13 for scenarios 2 and 3). The order in which scenarios were presented to participants in the actual study was randomly determined.

Development of Response Options for Scenarios

For each of the nine scenarios, participants were requested to answer two questions: question 1 to determine whether participants believed that they would take action in the wrongdoing situation, and question 2 to provide an indication of the participants' moral reasoning associated with their answer to question 1. Therefore participants were first asked (question 1) if they would take action in each of the situations (Yes/No), then select from a series of six response options (question 2) that best explain their answers to question 1.

A series of response options were developed for question 2 to attempt to measure whether a participant's level of moral reasoning was Pre-conventional, Conventional, or Post-conventional. The current study used both Kohlberg's and Gilligan's stage theory of moral reasoning as it has been argued that people may have distinctly different orientations towards moral reasoning,

based upon a 'justice' perspective or a 'care' perspective. As both theories essentially relate to three levels of moral development, a participant's reasoning would be considered at a higher level if, for example, they selected Gilligan's Conventional level rather than Kohlberg's Pre-conventional level of reasoning. Response options for each scenario were labelled K1, K2, K3, G1, G2, and G3, relating to the explanations corresponding to levels 1, 2, and 3, of Kohlberg's and Gilligan's theories of moral reasoning, respectively.

Finally, to enable between-group comparisons, participants were asked to report whether they held a pilot's license, and if so, how long it had been held.

Procedure

A total of 142 businesses in New Zealand were selected using the search engines Google and Yahoo, from which a total of 310 email addresses consisting of 102 personal email addresses (personal email addresses of staff), and 208 general email addresses (email addresses beginning with 'info', 'enquiries', organisation's name, etc.) were extracted. Each personal email address was sent a personalised email invitation to anonymously participate on the online survey. For the general email addresses, the email invitation asked the receiver to anonymously participate and also distribute the email invitation among their colleagues. No questions through which a participant could be identified were included in the interview, and no personal or identifiable information was recorded on the data sheet.

The survey was administered electronically via a website developed by the researcher to which participants were directed to complete the survey.

RESULTS

The level of statistical significance, alpha, was set at $p = .05$ for all statistical tests, and all tests were conducted as two-tailed.

Overall, there were 175 completed surveys. However, as it is not known how many participants were approached (as email respondents were asked to distribute the original invitation to others aviation employees within or outside their organisations), it is not possible to calculate the exact response rate. That said, if only original invitations were considered, this would equate to a response rate of 56%. Although this may seem high, similarly high responses rates have been found in New Zealand aviation research (e.g. 60% in Gilbey, Fifield, & Rogers, 2006).

For each scenario, Chi-Square test for goodness of fit (Elliott & Woodward, 2016) was used to ascertain if there was a statistically predominant response (Yes/No) to question 1. In scenarios 1, 5, 6, and 8, there was evidence that participants were statistically more likely to report that they would take action, but there was no statistically predominant response for scenarios 2, 3, 4, 7, and 9 (i.e., in the latter, participants were equally likely to report that they would take action or not take action).

For each scenario, the relationship between the seriousness ratings and the percentage of participants who stated that they would take action was tested using Spearman's rank order correlation. There was evidence that the relationship between the assigned seriousness ratings and participants' intention to take action was significant, $r_s = .694$, $n = 9$, $p = .04$; that is participants were more likely to take action in the scenarios that were judged to be more serious.

To investigate the relationship between intention to take action and flight experience amongst participants who were pilots, a median split was calculated for the time a pilot license was held (value 4 years). A Chi-Square test using flying experience (low = less than 4 years vs. high = equal to or greater than 4 years) and reporting intention (*Yes/No*) were conducted for each scenario¹. Results of the Chi-Square test of independence indicated that no comparisons achieved statistical significance, which suggested that a pilot's flying

experience did not appear to be related to their decision to take action for any scenario.

To test whether participants were more likely to take action was associated with how close they were to the hypothetical protagonist in each scenario, the number of each participant's *Yes* response was calculated for the 3 scenarios of each level of relatedness (thus, the range of scores for each level of relatedness for each participant was 0-3). Application of Friedman's test (Elliott & Woodward, 2016) suggested evidence of statistically significant changes in the distribution of responses associated with the level of relatedness to the event, $\chi^2 = 65.55, df = 2, p < .001$. Follow-up pair-wise applications of the Wilcoxon Test show that participants were more likely to intervene if they were *not related* ($M = 2.27$) than closely related ($M = 1.71$) to the protagonist, ($Z = 6.46, n = 159, p < .001$), and more likely to take action if they were *not related* ($M = 2.27$) than *related* ($M = 1.75$) to the protagonist ($Z = 6.67, n = 161, p < .001$). There was no statistically significant difference in the likelihood of taking action between *related* ($M = 1.75$) and *closely related* (1.71) ($Z = .68, n = 157, p = .496$).

Due to low expected cell counts for some combinations of responses to question 2, it was not possible to statistically explore the relation-

ship between all possible answers in relation to responses to question 1. Therefore, to test whether participants who answered *Yes* to question 1 for each scenario tended to operate at a higher level of moral reasoning than participants who answered *No*, the level of moral reasoning (i.e., Pre-conventional, Conventional, Post-conventional) was used for question 2, rather than the more detailed levels of either Kohlberg's or Gilligan's moral development theories (thus making the analysis 2×3 , rather than 2×6). Chi-Square tests of independence were then calculated for each scenario using participants' responses to question 1 (*Yes/No*) and their corresponding level of moral reasoning to question 2, Pre-conventional (PreC), Conventional (Conv), Post-conventional (PostC) as data. The results of the Chi-Square tests of independence are shown in Table 1.

With the exception of scenario 6, there was evidence of a significant relationship between participants' decision to become involved and their reason for their decision; that is, participants who reported that they would take action seemed to operate at a higher level of moral reasoning than participants who reported that they would not take action. Table 1 shows that when participants responded that they would take action (*Yes*), the modal explanation for why they did so, in all

Table 1. Relationship between response to question 1 and level of moral reasoning indicated in question 2

Scenario	Chi-Square Test of Independence				Modal Response	
	<i>n</i>	<i>df</i>	X^2	<i>sig</i>	<i>Yes (n)</i>	<i>No (n)</i>
1 nc	166	2	49.81	<.001	Conv (22)	PreC (21)
2 c	145	2	69.95	<.001	Conv (44)	PreC (54)
3 c	149	2	57.80	<.001	Conv (37)	PreC (49)
4 cc	153	2	70.86	<.001	PostC (40)	PreC (38)
5 nc*	71	2	12.99	<.001	Conv (66)	PreC (9)
6 c*	75	2	2.98	.231	Conv (68)	PreC/Conv/postC (6)**
7 nc	156	2	72.84	<.001	Conv (70)	PreC (52)
8 cc	166	2	70.92	<.001	Conv (56)	PreC (32)
9 cc	165	2	53.95	<.001	Conv (55)	PreC (43)

*Fisher's exact test was used, due to one or more expected cell counts being less than 5

** There was no single modal response to question 2 for participants who answered No to question 1; that is, each level was chosen the same amount of times.

except scenario 4, was consistent with conventional reasoning. Contrarily, when participants responded that they would not take action, the modal level of moral reasoning was most likely to be Pre-conventional.

To explore whether participants' level of moral reasoning was related to their degree of relatedness to the protagonist, participants' answers to question 2, indicating their level of moral reasoning for each scenario, were assigned a numerical value (i.e., Pre-conventional = 1, Conventional = 2, Post-conventional = 3). The numerical values were then summed for the three scenarios of each level of relatedness (the range of possible scores for each participant's answer to question 2 would therefore be 3–9 for each level of relatedness, with lower values suggesting lower levels of moral reasoning and vice versa). Application of Friedman's test suggested that there were no statistically significant changes in the distribution of levels of moral reasoning associated with the level of relatedness to the event, $\chi^2 = 3.35$, $df = 2$, $p = .188$. The mean scores were: *closely related* = 6.15 ($SD = 1.45$); *related* = 5.87 ($SD = 1.20$); and *not related* = 5.99 ($SD = 1.47$).

DISCUSSION

The first aim of this study was to explore whether differences in the level of moral reasoning were associated with differences in intentions to intervene in wrongdoing situations as one means of understanding under-reporting in aviation. There was evidence that, for eight of the nine scenarios tested, there was a significant association between whether participants believed that they would intervene upon becoming aware of wrongdoing and their level of moral reasoning. Specifically, with the exception of scenario 6, the modal level of moral reasoning was higher for participants who stated that they would intervene than participants who stated they would not intervene. As such, the relationship between participants' level of moral reasoning and their intentions to

intervene in wrongdoing provides evidence to support Hypothesis 1.

The findings of the current study indicate that when people become aware of wrongdoing in aviation, there is a great deal of uncertainty in their evaluation as to whether they believe that they would or would not become involved (i.e., Yes/No answers). For example, in five of the nine scenarios there was no statistically predominant response for what they would do on becoming aware of wrongdoing – that is, in five of the nine scenarios participants were equally likely to not intervene than to intervene. However, in four of the scenarios there was a statistically predominant response (take action), but nevertheless, the percentage of participants who would not intervene still ranged between 8% and 35%. In principle, this finding confirms that despite reporting of safety concerns being a legal requirement, under-reporting is still one of the key problems facing the aviation industry (NZALPA, 2005). Communicating safety concerns through the appropriate channels, as seen in the case of the Air Adventures accident, could have helped avoid the accident that ultimately occurred (CAA NZ, 2006). Thus, this study again shows that under-reporting is likely to occur and on occasions, it could lead to accidents such as that of Air Adventures.

Although there was a low rate of intentions to intervene in wrongdoing situations, which was arguably consistent with accidents such as the Air Adventures accident, there was however, a strong relationship between perceived seriousness of the wrongdoing and likelihood of a participant acting upon such wrongdoing. Although a causal relationship would need to be determined, this relationship does suggest that emphasising the importance and potential negative implications of any wrongdoing could be a means of raising the level of reporting.

This study also manipulated the level of relatedness between the person assessing the situation (i.e., the participant) and the person(s) involved in wrongdoing, to investigate whether this affected participants' reporting intentions. Hypothesis 2,

stating that participants would differ about whether they felt they had to intervene or not depending on the level to which they were related to the actor in the wrongdoing scenario was supported. The results suggested that overall, participants believed that they were more likely to intervene in wrongdoing situations if they were not related to the person(s) involved in such wrongdoing, than if they were related or closely related to them. The findings showed no difference in the likelihood of intervening if the participant was either related or closely related to the person(s) involved in wrongdoing. As such, this finding supports the evolutionary perspective argued by Hamilton (1964), and demonstrates that 'relatedness' is a factor likely to affect aviation employees propensity to report wrongdoing in aviation.

Arguably, the nature of aviation is that it is indeed a tight knit community where many people who work together may socialise together. For example, "The aviation industry is a unique business made up of a tight-knit community with a special bond in common" (Dawson Aircraft, 2010). This means that in some instances where people become aware of wrongdoing, it is likely that they have a relationship to the protagonist. If this relationship is closely related, or even just related, then the current findings indicate that this could result in higher levels of under-reporting. Indeed, there are reasons to suspect that the lack of communication of concerns prior to the Air Adventures accident was an example of this possibility. The influence of close relationships may prove difficult to overcome, as the aphorism 'not report on people' may be even more likely to occur if there is a close relationship between the wrongdoer and the observer of wrongdoing. Nevertheless, it is possible that ethical education or clearer reporting guidelines could allow the effect of close relationships to be set aside when considering the potential implications of not reporting wrongdoing.

An exploratory test of the relationship between the length of time a pilot had held a pilot's license and their stated intention to intervene failed to

find any evidence of a relationship. However, it is possible that this lack of evidence may have been confounded by other factors; for example, pilots who had held their license longer may in principle be less likely to intervene and also have built more close friendships in their industry. Moreover, the use of median split as a basis for comparison between groups may have failed to reveal that a difference could exist if a different value had been used.

Overall, there was evidence that when people become aware of wrongdoing that involves a person to whom they are closely related, they will be less likely to intervene. In line with the arguments of inclusive fitness concept (Hamilton, 1964), it could therefore be assumed that the more related the individual to the person involved in wrongdoing, the less likely that the individual will take corrective action in the wrongdoing situation.

LIMITATIONS

A number of potential limitations are noted in the current study. First, as participants were a self-selecting sample, this may limit the generalisability of the results to all aviation employees. Second, it should be acknowledged that there may have been differences in interpretation of wrongdoing scenarios presented to participants. Although the scenarios were based on real-life wrongdoing situations and have been reported on the regulatory authority's incident and accident database, there is a possibility that given people's differences in background, education and experiences, not all participants would have viewed the presented scenarios as wrongdoing situations, or agreed with the level of seriousness assigned to each of these scenarios. Third, it has long been recognised that people do not always behave in a way that is consistent with the beliefs that they express (LaPiere, 1934). As such, it is unclear to what extent these hypothetical findings would hold true in real life situations, although later studies have shown that

attitudes are a good predictor of behaviour (e.g., Sheppard, Hartwick, & Warshaw, 1988).

FUTURE RESEARCH DIRECTIONS

Finding evidence that there is a relationship between people's level of moral reasoning and whether or not they would intervene (e.g., report their concerns about something they perceive to be wrongdoing) suggested at least two potential directions for future research. One approach would be to carry out an intervention study using participants that stated they would not take action in wrongdoing situations and provide half of them with, for example, a short course in business and workplace ethics, whilst the other half would serve as a control condition. If, subsequently, there was found to be a causal relationship between level of moral reasoning and intention to intervene, then this would be good evidence that ethics education would be a potentially efficacious means to increase safety in aviation. A second and perhaps more pragmatic approach would be to investigate ways in which information about what should be reported could be made available to aviation employees. Nevertheless, there is some evidence to support the former approach as research has suggested that education may indeed help to improve people's levels of moral reasoning (Rest, 1986; Self, Baldwin, & Wolinsky, 1992; Holm, Nielsen, Norup, Vegner, Guldmann, & Andreasen, 1995), whilst it would be difficult to ensure that a list of what should be reported would be exhaustive.

CONCLUSION

In conclusion, the findings of the current study suggest that there is an association between participants' level of moral reasoning and their likelihood to intervene upon observing wrongdoing in an aviation context. There was clear evidence that whether or not wrongdoing is communicated depends very much upon who it is that observes

it and also upon their relationship to the person committing the wrongdoing. This study therefore makes a significant contribution in uncovering some of the issues that lead to underreporting of wrongdoing. Indeed, findings of this study can almost certainly be used to contribute and improve knowledge in areas other than aviation.

REFERENCES

- Camerer, L. (2001, October). *Summary of panel discussion on whistleblowing: A practical tool in combating corruption*. Paper presented at the 10th International Anti-Corruption Conference, Prague, Czech Republic.
- Civil Aviation Authority of New Zealand. (2006). *Director's media conference statement*. Retrieved from http://www.caa.govt.nz/publicinfo/med_rel_Coroners_Report.htm
- Dawson Aircraft. (2010). Retrieved from: <http://www.aircraftpartsandsalvage.com/index.php>
- Dehn, G. (1999, December). *Whistleblowing to combat corruption*. Paper presented at the Organisation for Economic Co-operation and Development OECD Labour/Management Programme, Paris, France.
- Elliott, A., & Woodward, W. (2016). *IBM SPSS by Example: A Practical Guide to Statistical Data Analysis*. London: Sage Publications.
- Flanagan, O. (1984). *The Science of the mind* (2nd ed.). Cambridge, MA: The MIT Press.
- Gilbey, A., Fifield, S., & Rogers, S. (2006). Comparative optimism: Are air traffic controllers immune? *Human Factors and Aerospace Safety*, 6(2), 89–102.
- Gilligan, C. (1982). *In a different voice: psychological theory and women's development*. Cambridge, MA: Harvard University Press.

- Hamilton, W. D. (1964). The genetical evolution of social behavior. *Journal of Theoretical Biology*, 7(1), 1–52. doi:10.1016/0022-5193(64)90038-4 PMID:5875341
- Holm, S., Nielsen, G. H., Norup, M., Vegner, A., Guldman, F., & Andreasen, P. H. (1995). Changes in moral reasoning and the teaching of medical ethics. *Medical Education*, 29(6), 420–423. doi:10.1111/j.1365-2923.1995.tb02865.x PMID:8594405
- Kohlberg, L. (1976). Moral stages and moralization: Cognitive developmental approach. In T. Lickona (Ed.), *Moral development and behavior: Theory, research, and social issues*. New York: Holt, Rinehart & Winston.
- Kohlberg, L. (1984). *The psychology of moral development*. New York: Harper and Row.
- Kohn, S. M., & Kohn, M. D. (1988). *The labor lawyer's guide to the rights and responsibilities of employee whistleblowers*. Westport, CT: Greenwood Publishing Group.
- KPMG. (2008). *KPMG Forensic's 2008 survey of fraud in Australia and New Zealand*. Retrieved from http://www.kpmg.ch/docs/KPMG_Forensic_Fraud_Survey_2008_Australia_New_Zealand.pdf
- Krebs, D. L., Vermeulen, S., & Carpendale, J. (2014). Structural and situational influences on moral judgment: The interaction between stage and dilemma. In *Handbook of Moral Behavior and Development: Research* (vol. 2). New York: Psychology Press.
- LaPiere, R. (1934). Attitudes versus actions. *Social Forces*, 13(2), 230–237. doi:10.2307/2570339
- Lewis, D. (2001). Whistleblowing at work: On what principles should legislation be based? *Industrial Law Journal*, 30(2), 169–193. doi:10.1093/ilj/30.2.169
- New Zealand Air Line Pilots' Association. (2005, December). 'N'FORMATION Line up, 5.
- Oderman, D. B. (2002). Ethics education in university aviation management programs in the US: Part one—the need. aviation institute. *Journal of Air Transportation*, 8(1), 15–36.
- Peters, R. S. (2015). *Moral Development and Moral Education*. Oxon, UK: Routledge.
- Rest, J. R. (1986). *Moral development: Advances in research and theory*. New York, NY: Praeger Publishers.
- Sawyer, K. R. (2005, September). *The test called whistleblowing*. Paper presented at the National Conference of Whistleblowers Australia 'Whistleblowing: Making it work', Adelaide, Australia.
- Self, D. J., Baldwin, D. C. Jr, & Wolinsky, F. D. (1992). Evaluation of teaching medical ethics by an assessment of moral reasoning. *Medical Education*, 26(3), 178–184. doi:10.1111/j.1365-2923.1992.tb00151.x PMID:1614342
- Sheppard, B. H., Hartwick, J., & Warshaw, P. R. (1988). The theory of reasoned action: A meta-analysis of past research with recommendations for modifications and future research. *The Journal of Consumer Research*, 15(3), 325–343. doi:10.1086/209170
- Tani, K., & Gilbey, A. (2012). Aviation employees' intentions to report safety concerns. *Aeronautica*, 2(2), 8–17.

ADDITIONAL READING

- Blasi, A. (1983). Bridging moral cognition and action: A theoretical view. *Developmental Review*, 3(2), 178–210. doi:10.1016/0273-2297(83)90029-1
- Bommer, M., Gratto, C., Gravander, J., & Tuttle, M. (1987). A behavioral model of ethical and unethical decision making. *Journal of Business Ethics*, 6(4), 265–280. doi:10.1007/BF00382936

Bouville, M. (2008). Whistle-blowing and morality. *Journal of Business Ethics*, 81(3), 579–585. doi:10.1007/s10551-007-9529-7

Derry, R. (1989). An empirical study of moral reasoning among managers. *Journal of Business Ethics*, 8(11), 855–862. doi:10.1007/BF00384528

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research*. Reading, MA: Addison-Wesley.

Fraedrich, J., Thorne, D. M., & Ferrell, O. C. (1994). Assessing the application of cognitive Fraedrich moral development theory to business ethics. *Journal of Business Ethics*, 13(10), 829–838. doi:10.1007/BF00876263

Harris, J. R. (1990). Ethical values of individuals at different levels in the organizational hierarchy of a single firm. *Journal of Business Ethics*, 9(9), 741–750. doi:10.1007/BF00386357

Hegel, G. W. F. (2001). *Philosophy of Right* (S. W. Dyde, Trans.). Canada: Batoche Books Limited. (Original work published 1821)

Kidwell, J. M., Stevens, R. E., & Bethke, A. L. (1987). Differences in ethical perceptions between male and female managers: Myth or reality? *Journal of Business Ethics*, 6(6), 489–493. doi:10.1007/BF00383291

Nisan, M. (1984). Content and structure in moral judgment: An integrative view. In W. M. Kurtines & J. L. Gewirtz (Eds.), *Morality, moral behavior, and moral development*. New York: Wiley.

KEY TERMS AND DEFINITIONS

Conventional Level: For Kohlberg, this level is about acceptance of the rules and standards of one's group. For Gilligan, this level is characterised by responsibility: 'Good' is equated with caring for others.

Ethics: A set of standards of right and wrong, good and bad.

Gilligan's Moral Development Theory: An alternative theory to Kohlberg's theory, proposed by Gilligan (1982) that argues that ethical decisions based on justice and fairness were incomplete. Gilligan called her theory "ethics of caring", it is a stage theory of moral development for women grouped into three major levels: pre-conventional, conventional, and post conventional. For Gilligan, the transitions between the stages are dependent on changes of the self and the environment in which the self lives, rather than the Kohlbergian changes in cognitive capability.

Kohlberg's Moral Development Theory: A theory proposed by Kohlberg (1976) in which he described six stages of moral reasoning grouped into three major levels: pre-conventional, conventional, and post-conventional. Kohlberg believed that individuals could only progress through these stages one stage at a time; one cannot get to a higher stage without passing through the stage immediately preceding it. Individuals could only come to comprehending a moral rational one stage above their own.

Morality: The application of ethical standards.

Post-Conventional Level: For Kohlberg, this level is where ethical principles are considered. For Gilligan, this level is characterised by the focus on eliminating tensions between self and others, and accepting the principle of care as a universal ethical principle which condemns exploitation and hurt in our life and others.

Pre-Conventional Level: For Kohlberg, this level is where behaviour is motivated by anticipation of pleasure or pain. For Gilligan, this level is characterised by caring for the self in order to ensure survival. In this level, women begin to see connections between themselves and others.

Wrongdoing: In the context of this study, wrongdoing is used to imply instances of behaving or acting in a way that is inconsistent with safety practices, not limited to practices that are against the law.

ENDNOTE

- ¹ Although anecdotal evidence suggests that 1000 hours (personal communication) is a useful value for determining experience

(high/low), in the current study, the term low and high are used simply to denote to which half of the medium split a participant was assigned.

Science, Ethics, and Weapons Research

E

John Forge

Independent Researcher, Australia

INTRODUCTION

If it were not for *weapons research*, there would be no predator drones or smart bombs or improvised explosive devices or assault rifles. The insurgents in the Middle East and elsewhere would have no means to fight, and there would be no wars, large or small. Even more importantly, there would be no vast arsenals of thermonuclear weapons capable of ending much of the sentient life on the planet. The world would then most certainly be a safer place. But weapons research is not something new: the gunpowder weaponry of the early modern period was the product of research, as were the torsion catapults in Greece at the time of Philip and Alexander of Macedon. Whatever else is true about weapons research, it is clear that it introduces new (or improved) means of killing and destruction, and this is sufficient to define the activity.¹ This would appear to be a very weighty matter, something that one might imagine philosophers, and others who think about such things, would have had a lot to say; surprisingly, not much at all has been written on the subject, though some explanation of this neglect will be given in this chapter.

The main issue for *ethics* and weapons research centres on the ethical or moral evaluation of the activity: Is it *ever* morally justified to design the means to kill, harm and destroy, and if so, under precisely what circumstances? Turning to science and its relation to weapons research, the question here is the *role* that science plays in weapons research. Perhaps weapons research is a wholly (applied) scientific endeavour or perhaps science is a part of weapons research? Bringing ethics back in, if weapons research is deemed morally

wrong, then is it the case that *whatever* role science plays is also wrong? To answer these questions, three examples will be given which will help to clarify the roles that science can play in weapons research. If weapons research itself is understood as applied science, as it is by Arrigo for instance (Arrigo 2000: 303), then one might expect this to entail the application of theory to the design for new weapons, for true or radical innovation. But there are other ways in which science can inform weapons research, as will be seen presently. Before moving on to these examples, it is worth making some general, and very brief, comments about ethics and the way it can apply to an intellectual activity such as science. This is worth doing because it cannot be assumed that the audience for the present topic is familiar with philosophy or ethics, but it is necessary to have a framework.²

BACKGROUND

A straightforward way to describe ethics is to say that it is a study which deals with what persons ought and ought not to do. It is thus to do with the choices, actions and behaviour of mature competent people. Some of the things that people do do not affect others, other humans, other sentient beings, in any significant way and hence these do not come under the purview of ethics. Those actions that do affect others are, however, open to moral or ethical evaluation: are they right or are they wrong? To resolve that question, one needs to appeal to a moral system. All such systems forbid certain actions, namely those that inflict unjustified harm on others. This is surely intuitive and obvious: *no one* wants to be harmed. It is almost

by definition that no sentient being wants to feel pain - assuming that the pain does not indicate that some medical treatment is working or some such – and to be in pain is one form of being harmed.

Some moral systems require people not only to refrain from harming others but also to provide some positive benefit for them. Jeremy Bentham and John Stuart Mill, the nineteenth century English philosophers, famously believed that one ought to strive to increase the amount of happiness in the world. However, morality is supposed to be *impartial* in the sense that it forbids discrimination in regard to moral action. Prohibitions on harming do not end with family or friends or community or country: *nobody* should be harmed, no one at all. Some critics of the style of morality advocated by Bentham and Mill have pointed out that it is impossible to increase the amount of happiness in the world impartially: no one can make everyone happy! Just how serious this objection is is a matter of ongoing debate. But it is only necessary here to note that this kind of moral system shares the prohibition on harming with the former kind: for the topic at hand, it is clear that the moral evaluation of weapons research, whatever else it might involve, will not be such as to see it as an activity which aims to increase the amount of happiness in the world.

Most philosophers do not believe that the dictates of morality are absolute and cannot be broken in any circumstances. For example, most accept that a moral rule such as “Do not cause pain” has justified exceptions. Clearly, a dentist who inflicts pain on her patient to save his teeth has not done something morally wrong – provided that the patient understands and assents to the treatment. Also, it is generally agreed that it is permissible to cause pain in self-defence, if that is the only way to defend oneself. This leads to the view that *justifiable* exceptions to the overall moral prohibition against harming will be such as to show that the harm inflicted will prevent other harms. Just how this is worked out will vary from case to case, and it is here that much of the hard work in ethical reflection and evaluation takes place. One might think as a basic principle that

the harms prevented should be at least as much or many or as great as the harms caused if there is to be justification, and that therefore it must be necessary to be able to make some informed assessment of what these might be. This brings the discussion back to weapons research, an activity that aims to provide the *means* to harm. It is now necessary to look at some examples.

1. It is well-known that both kinds of nuclear weapon, fission and fusion weapons, were the *direct results* of the application of scientific theory to design. Without advances in nuclear physics in the 1930s and 1940s, the very idea of a nuclear weapon would not have been dreamed up. Rhodes gives an excellent account of the genesis of the idea of a fission weapon, from the speculations of Szilard from 1933 to his work with Fermi in the US, and the ‘memorandum’ written by the émigré scientists Frisch and Peierls in England which made predictions about crucial nuclear parameters. The designs of the two kinds of fission bombs made during the Second World War were determined by the properties of the fissile materials used, matters that were uncovered by painstaking research.³ The idea for a thermonuclear weapon emerged from the atomic bomb project, and was also driven by science and by scientists. Indeed, the fusion reactions that power these weapons must be initiated by ‘fission triggers’, so not only did the research into fission or atomic bombs lead to thermonuclear weapons, the products of the former were integral elements of the latter.⁴

The research leading to nuclear weapons was exceptional in that scientists, including all of those mentioned above, took the initiative in agitating for programmes to investigate the possibility of using the recently discovered fissile materials for a weapon of terrible destruction, to the subsequent regret of some of them. The military is now well-aware of the power of science and has set up agencies to oversee the application of scientific

theory to the design of new weapons systems. The best-known is the US's Defense Advanced Research Projects Agency, or DARPA; notice that, in common with armed forces the world over, the word "defense" appears in its name. There are equivalents in Russia, China, the UK, France, Israel and other countries.

2. In the nineteenth century the weapons that caused the majority of fatalities and injuries in war, artillery and the infantryman's longarm, were transformed by the scientific analysis of existing armaments. Smoothbore cannon and muskets were transformed into steel, breech-loading guns whose recoil was damped and rifles firing clips of cartridges using smokeless powder.⁵ These are instances of science *improving* existing designs. One example will be given as an illustration. All smoothbore weapons which fire balls, musket and cannon balls, are inherently inaccurate because of uncontrollable forces which act on balls as they spin in the air after leaving a gun barrel. This is called the Magnus Effect or Magnus Force, after the German physicist Heinrich Magnus who described it in 1852. It was also investigated by the English physicist Benjamin Robbins, in his book *The New Principles of Gunnery*, published a century before Magnus' work (Steele 1994: 359). The effect is this: unless a ball only spins about the axis in the direction it is travelling, what is called the roll axis in aeronautics, it will experience unequal pressure on either side in the direction of its motion, and hence be deflected in the direction of the axis of spin. In other words, it will not travel straight, and if it actually hits (anywhere near) its target, then this is a matter of luck.

This analysis showed that if only the bullet could be made to spin about the roll axis, accuracy would be greatly increased. The first attempts to do so involved 'rifling' musket barrels, fitting them

with raised 'lands' spiralling down the inside of the barrel which caused the ball to spin about the required axis. There were two problems: the rifled musket was very slow to load because the ball had to be hammered down the barrel, and the barrel became fouled after only a few shots because the powder residue collected around the lands. These problems were solved by three innovations. There were, in the first place, advances in metal working that allowed relatively efficient breech-loading rifles – leakage of gas from around the breech was eventually dealt with by gunsmiths such as Paul Mauser. Then there was progressive improvement in propellant until eventually smoke-less powder became available. Finally, bullets were designed that would not stick in the barrel on the lands. The idea was that the bullet would have two parts; a base that was made out of a relatively soft material that would deform in the lands and impart a twist to a hard core made narrower than the base so that it would not stick in the barrel. After two attempts in England in the 1820s, eventually a French innovation by Captain Claude-Étienne Minié of the French Chasseurs was successful in 1847.⁶ This led to many orders of magnitudes increase in lethality: in the American Civil War over half a million casualties, including two hundred thousand killed, were caused by 'Minié rifles'.

3. The application of scientific theory, either to produce quite new weapon designs or to improve existing ones, is a relatively new phenomena, beginning with the study of ballistics in the seventeenth century (see Hall 2009). However, *systematic* research into weapons design took place at least as far back as the fourth century BCE. Remarkable work was done in that century in relation to the design of the torsion catapult, which marked the beginning of effective siege artillery. A torsion catapult works because of the restoring force exerted by 'springy' material, like rope, which has been stretched. Wrapping rope around a stout piece of wood fixed to a frame, embedding a handle in the

mass of rope and then pulling the handle back will create a torque: the handle will fly back to its resting position when released. It can be therefore seen how, in principle, such a device can be used to cast stones when placed in a basket attached to the handle.

It is not known how the Greek engineers discovered the torsion principle and applied it to designing catapults, but this must have been done *empirically* as the theory was not formulated until much later. However, the design of torsion catapults was codified in treatises which gave general instructions how to build them. These instructions were general because formulae, known as the hole formulae, gave the key parameter for casting stones of a given weight or arrows of a given length. The Greek engineers had discovered formulae that related the optimal size of the hole in the catapult frame in which the wood about which the rope or springs was to be wound to the weight of the stone or length of the arrow. The former is more complex and involves a cube root, which gives an equation that Greek mathematics at the time could not solve exactly. Perhaps still more remarkably, the treatises gave approximation methods for solving the equations in question.⁷

FUTURE DIRECTIONS

If weapons research is morally wrong, then it would appear that the actions of any scientist who contributes to weapons research are also morally wrong. But the matter is not quite so straightforward. A scientist who knowingly, willingly and intentionally engages in weapons research is as open to moral censure as any other weapons researcher, assuming that weapons research is a proscribed activity. It is not so clear, however, that a scientist who does not intend that her work be used to further the ends of weapons research, but sees that it may well in fact do so is also blameworthy. It is less clear still that a scientist who does not foresee this outcome, assuming it comes

about, is to blame. It would not be necessary to address any of these issues if weapons research was not morally wrong, so this matter must now be discussed.

It was mentioned above that the topic has received much less attention than it deserves.⁸ Forge, who has addressed the topic at some length and argued that weapons research is morally wrong under all circumstances, believes that the question of the morality of weapons research has been assimilated to questions about the justness of war, about which a great deal has been written of late, and he believes this is a mistake (Forge 2013: ix). The consensus about war is that war is just or morally permissible – no blame attaches – if it resists aggression. This is the ‘modern view’ which sees just war as more than self-defence, other defence being allowed if the other is the victim of aggression. The implicit corollary is that measures such as weapons research are justified if they aid in the prosecution of a just war or in the pursuit of self-defence or deterrence.⁹ This is the standard rationale or *standard justification* of weapons research. Forge denies that this justification is convincing. However, Kemp, one of the few who have discussed the ethics of weapons research, claims that it is not only permissible to undertake weapons research for defence, and maintains that it is a *civic duty* imposed on scientists to do so (Kemp 1994). Forge and Kemp therefore adopt diametrically opposed viewpoints, and the topic can be explored by comparing and assessing their respective positions.

Kemp claims that since the state has a duty to protect its citizens, and since Hobbes many political philosophers believe that this is the *raison d’être* of the state, then those who are able to assist in this vital endeavour by designing the means for defence have a duty to do so. Note that a civic duty in this sense is not a *moral* duty, as it only applies to a part of the moral community, namely one’s fellow citizens, and hence lacks the (completely) impartial character of morality. Expressed in these terms, Forge’s view is that it is moral duty *not* to undertake weapons research.

Forge does not deny that states have a right to defend themselves, or defend others from aggression, nor does he deny that weapons research can help to achieve these ends. What he does deny is that this is *all* that weapons research does, namely provide the means to resist aggression, whatever the intent behind it. Note also that Kemp's position is compatible with weapons research being permissible while Forge's is not. Which is correct? Kemp's view seems to be convincing, but when weapons research is analysed in more depth, this initial reaction is seen to be wrong.

Hacking maintains that what weapons research produces in the first instance is *knowledge*, instructions for making new or improved weapons, not hardware (Hacking 1986).¹⁰ Weapons research is not the same thing as weapons manufacture. A scientist who successfully discharges her civic duty does not therefore provide a *unique* array of defensive weapons for her country, but a *design*, and any institution with the requisite skills and resources could produce identical weapons if it possesses the design. For instance, the Kalashnikov, the AK-47, the most widely made weapon of all time, has been manufactured from 1947 by the Soviet Union, Russia, Poland, Hungary, East Germany, Bulgaria, Romania, China, North Korea, Yugoslavia, Iraq, Egypt and India, and variants by Finland, Israel and South Africa.¹¹ Mikhail Kalashnikov invented a highly effective assault rifle, but should he be praised for discharging his civic duty or condemned for introducing into the world a terrible new means to kill? Clearly, his innovation spread beyond his own country. Successful weapons tend to become ubiquitous. Both sides in the American Civil War used 'Minié rifles', as did both sides in the Franco-Prussian War of 1871. By the First World War, all protagonists had machine guns, quick-firing artillery as well as reliable rifles. There was less uniformity in the armaments of the Second World War, though all sides had tanks, bombers, fighters, field artillery and so forth, and there is still less uniformity today. This trend is due to the vastly increased pace of weapons research and it is to be expected

that states will strive to copy successful weapons innovations as much as they can. This may not be easy and it may not always be clear how it can be done, but the very fact that many different countries come up with weapons that are remarkable similar shows that it happens. Weapons innovation cannot be contained in time or space.

If successful weapons innovation does not lead to a unique array of weapons, perhaps this does not matter if leads to weapons that can *only* be used to defend against aggression – perhaps Kemp's position can be resurrected along the lines that designing defensive weapons is a civic duty? The role of a defensive weapon, or weapon that is postured defensively, is to actively protect an asset, but this does not necessarily amount to defending against aggression. The work of Luttwak on the levels of strategy shows the relationship between defence and its strict 'opposite' offence is complex and multi-faceted.¹² One implication of Luttwak's analysis is that a state that would embark on a war of aggression needs a range of defensive options. It needs to be able to defend itself, its 'homeland', from retaliation, it needs to defend captured territory and its forces in enemy territory from counter-attack, etc. Thus, if the overall theatre or grand strategy is one of aggression, with the long-term goal of capturing territory or otherwise exerting hegemony, at 'lower' levels defensive tactics and even defensive operations may be necessary at certain times and places, and conversely. Fighting on the Eastern Front in the Second World War, after Hitler invaded the Soviet Union, exemplified this dynamic. At the level of grand strategy, Hitler was waging an aggressive war of conquest, while Stalin was defending. But even as early as the third month of the war, the Germans were mostly on the defensive, with their infantry holding positions while their tanks, the main instrument of the blitzkrieg operations, were being repaired and refitted.¹³ Thereafter, the German offensive re-commenced, stalling finally in November 1941 in front of Moscow, and the next two years the protagonists defended and attacked one another in turn until the Soviet

Union began to prevail. All the weapons both sides had at their disposal were used in all forms of fighting, including tanks that were deployed in mobile defence: there were no weapons *only* used for defence. Kemp's position therefore cannot be salvaged.

Forge maintains that since weapons are, by their very nature, the means to harm, then research which aims to introduce new or improved weapons into the world requires justification, and that the only possible justification is that the weapon in question will prevent more harm than would be the case were it not available.¹⁴ This kind of justification lies behind all references to defence in relation to weapons research, military spending and so forth. But this demand cannot be satisfied: weapons research produces knowledge in the form of designs which can be implemented in the future in circumstances that are unknowable – as the examples discussed above have shown – and since there is no such thing as a weapon that cannot be used to aid aggression, there can be no assurance that an episode of weapons research will produce a weapon that prevents more harm than it causes. It is precisely this assurance that is needed if weapons research is to be justified, hence Forge argues that weapons research is always unjustified. The historical record shows that wars have become more frequent and more deadly since scientific advances began to be applied systematically to weapons research, and so the facts seem to support Forge's position. But these important issues surely need more discussion, not just from philosophers and others who reflect on weapons research, but from the scientists and other experts who actually take part.

CONCLUSION

Returning to focus again explicitly to the role of scientists, if weapons research is morally wrong and without justification, then scientists, just like everyone else, should not take part. A scientist who knowingly and deliberately takes part in an

activity that is morally wrong is as blameworthy as any other participant. However, there are different ways in which scientists can be said to participate. For instance, a scientist can receive a grant from the military even though she does not intend to undertake weapons research and does not see how her work might provide a basis for weapons research. But, the military only sponsors research it thinks might be useful for its own ends, including weapons research, and this is a good reason why scientists should not accept such funding. It is also possible that a scientist knows she is conducting research that could have military application without being sponsored by the military. This happened in the late 1930s when some members at least among the scientific community believed that current advances in nuclear physics could contribute to a weapon of mass destruction. The majority of scientists working at the time, including Hahn and Strassman, the discoverers of nuclear fission, saw themselves as 'pure' research scientists. Szilard sought a moratorium on publishing papers on nuclear physics in order to try to prevent such nuclear weapons research, with limited success (Rose 1998; 92). This raised complex questions about the responsibility of scientists for applications of their work that they do not intend, or even foresee, but these are questions beyond the scope of this chapter.

REFERENCES

- Acheson, D. (1969). *Present at the Creation*. New York: Norton.
- Arrigo, J. (2000). The Ethics of Weapons Research: A Framework for Discourse between Insiders and Outsiders. *Journal of Power and Ethics*, 1.
- Bridger, S. (2015). *Scientists at War: The Ethics of Cold War Weapons Research*. Cambridge, MA: Harvard University Press. doi:10.4159/9780674426085

- Forge, J. (2007). No Consolation for Kalashnikov. *Philosophy Now*, 59.
- Forge, J. (2008). *The Responsible Scientist*. Pittsburgh, PA: Pittsburgh University Press.
- Forge, J. (2013). *Designed to Kill: The Case Against Weapons Research*. Dordrecht: Springer. doi:10.1007/978-94-007-5736-3
- Fotion, N. (1990). *Military ethics*. Stanford, CA: Hoover Institution Press.
- Grandy, D. (1996). *Leo Szilard*. Lanham, MD: University Press of America.
- Greener, W. (1910). *The Gun and Its Development*. London: Cassell.
- Grice, M. (2009). *On Gunnery*. Charleston, SC: Booksurge Publishing.
- Hacking, I. (1986). Weapons Research and the Form of Scientific Knowledge. *Canadian Journal Of Philosophy, Supplementary*, 1, 2.
- Hall, A. (2009). *Ballistics in the Seventeenth Century*. Cambridge, UK: Cambridge University Press.
- Hall, B. (1997). *Weapons and Warfare in Renaissance Europe*. Baltimore, MD: Johns Hopkins University Press.
- Hoddeson, L., Henriksen, P., Meade, R., & Westfall, C. (1993). *Critical Assembly*. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511665400
- Kalashnikov, M. (2006). *The Gun that Changed the World*. Cambridge, UK: Polity Press.
- Kanaher, L. (2007). *AK-47*. Wiley. Hoboken, N.J.
- Kemp, K. (1994). Conducting Scientific Research for the Military as a Civic Duty. In *Ethical Issues in Scientific Research*. New York: Garland.
- Liddell Hart, B. (1973). *History of the Second World War*. London: Pan.
- Luttwak, E. (1987). *Strategy: The Logic of War and Peace*. Cambridge, MA: Harvard University Press.
- Manchester, W. (2003). *The Arms of Krupp*. Boston: Back Bay Books.
- Marsden, E. (1969). *Greek and Roman Artillery: Historical Development*. Oxford, UK: Oxford University Press.
- Marsden, E. (1971). *Greek and Roman Artillery: Technical Treatises*. Oxford, UK: Oxford University Press.
- Rhodes, R. (1986). *The Making of the Atomic Bomb*. Penguin.
- Rhodes, R. (1995). *Dark Sun: The Making of the Hydrogen Bomb*. New York: Touchstone.
- Rihill, T. (2007). *The Catapult*. Yardley, PA: Westholme.
- Rose, P. (1998). *Heisenberg*. Berkeley, CA: California Univ, Press.
- Stahel, D. (2009). *Operation Barbarossa and Germany's Defeat in the East*. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511732379
- Stankiewicz, R. (2000). The Concept of Design Space. In J. Ziman (Ed.), *Technological Innovation as an Evolutionary Process*. Cambridge, UK: Cambridge Univ. Press.
- Steele, B. (1994). Muskets and Pendulums: Benjamin Robins, Leonhard Euler and the Ballistics Revolution. *Technology and Culture*, 35.
- Walter, J. (1979). *The German Rifle*. London: Arms and Armour Press.
- Warren, K. (1989). *Armstrongs of Elswick*. London: MacMillan. doi:10.1007/978-1-349-10994-4
- Weart, S., & Szilard, G. (1978). *Leo Szilard: His Version of the Facts* (Vol. 2). Cambridge, MA: MIT Press.

Zabicki, D. (2006). The French 75. In S. Tucker & P. Roberts (Eds.), *World War I*. Santa Barbara, CA: ABC-CLIO.

ADDITIONAL READING

Forge, J. (2004). The Morality of Weapons Research. *Science and Engineering Ethics*, 10. PMID:15362709

Forge, J. (2007) "What are the Moral Limits of Weapons Research?", *Philosophy in the Contemporary World*, 14.

Forge, J. (2009). Proportionality, Just War Theory and Weapons Innovation. *Science and Engineering Ethics*, 15. PMID:18802788

Forge, J. (2014). The Case Against Weapons Research. *International Journal of Technoethics*, 5.

Gert, B. (2004). *Gert Common Morality*. Oxford: Oxford University Press. doi:10.1093/0195173716.001.0001

Gert, B. (2005). *Morality: Its Nature and Justification* (Revised Edition). Oxford: Oxford University Press. doi:10.1093/0195176898.001.0001

McNeill, W. (1982). *Pursuit of Power*. Chicago: Chicago University Press.

Orend, B. (2005). *The Morality of War*. Peterborough, Ontario: Broadview.

Walzer, M. (1977). *Just and Unjust Wars*. London: Allen Lane.

ENDNOTES

¹ Forge has given the following definition: Weapons research is research carried out with the intention of designing new weapons, or improving the design of existing weapons, or designing or improving the means for

carrying out activities associated with the use of weapons, Forge, 2013: 14.

² The writer's own webpage, www.moralitymatters.net, contains some introductory material, together with more on the topic of this chapter, and also more references.

³ See especially Rhodes 1986; for his discussion of Szilard, see 28-30, and for the Frisch-Peirels Memorandum see 321-325. Weart edited Szilard's papers, which were published in 1978, see Weart and Szilard 1978. Grandy has a more objective account of Szilard's contribution. Forge 2008 Chapter 8 is a short accessible overview of the Manhattan Project and the events leading up to it. For a more technical discussion, see Hodderon *et al.* 1993.

⁴ Rhodes' second book, Rhodes 1995, takes up where his first book ends, at the conclusion of the Second World War, and begins with a discussion of the Soviet atomic bomb project and moves on to the decision by the US to then proceed with thermonuclear weapons. Both books are important because they give background to the decisions to actually go ahead with the respective nuclear weapons research programmes, surely among the most momentous decisions of all time. For more on the decision to proceed with thermonuclear weapons research by one who was actually involved, see Acheson 1969, especially 344-349.

⁵ It was necessary to develop special steels for guns that used the new high explosives, both of which were due to advances in chemistry in the nineteenth century, see Manchester 2003. Manchester's book gives an excellent account of Krupps, the famous, or notorious, weapons manufacturer and innovator. Warren 1989 tells the story of George Armstrong, the English equivalent of Alfred Krupp. The problem of recoil, which had plagued gunners from the very beginnings of artillery, was solved at the end of the century by the French, see Zabecki 2004.

- ⁶ See Greener 1910: 629-632 for the two British attempts. An excellent account of Minié is in Hall 1997: 135-147. Walter 1979 is a comprehensive account of the nineteenth century German rifle innovation, especially in regard to the Mauser.
- ⁷ Marsden wrote two important books on this topic, which have become classics. One gives the historical development of Greek and Roman artillery, while the second reproduces four technical treatises, from the fourth century BCE to the first century CE, which give instructions for building catapults. Ri-hill is a more up to date, and perhaps more readable, general text on the topic, which does not always agree with Marsden.
- ⁸ Sarah Bridger's recently published *Scientists at War: The Ethics of Cold War Weapons Research* has, in spite of the title, almost nothing in it about ethics. It is however an excellent account of the attitudes of many of the leading US weapons researchers between the launching of Sputnik and Star Wars, with special reference to weapons research and the Vietnam War.
- ⁹ Some Just War Theorists, and some others, have argued that certain classes of weapons, especially weapons of mass destruction, are immoral but have not (even) gone on to state the corollary to that proposition that the weapons research that aims to produce weapons of mass destruction is immoral. Fotion, Fotion 1990, gives one of the best discussion of the issue from this restricted viewpoint.
- ¹⁰ This knowledge, in the form of designs, can be more or less well articulated, from the most basic and sketchy idea – for instance, using a chain reaction in a uranium assembly – to an engineering specification that could, in principle, be implemented entirely by machine. For more on this issue, see Stankiewicz 2000.
- ¹¹ For Kalashnikov's story in his own words, which must be taken with a grain of salt, see Kalashnikov 2006. For more details, and more objectivity, see Kanaher 2007.
- ¹² Luttwak 1987 is the main source. But see also Liddell Hart 1973 for various ideas about what could count as a defensive weapon.
- ¹³ See Stahel 200, especially Chapter 7 and the introduction.
- ¹⁴ Forge's main work is Forge 2013, which contains many further references.

Category F

Fuzzy Logic and Soft Computing

Application of Fuzzy Numbers to Assessment Processes



Michael Voskoglou

Graduate Technological Educational Institute (T.E.I.), Greece

INTRODUCTION

In the present work a combination of the COG defuzzification technique and of the TFNs (or TpFNs) is used as an assessment tool. Examples of assessing student problem-solving abilities and basket-ball player skills are also presented illustrating in practice the results obtained. This new fuzzy assessment method is validated by comparing its outcomes in the above examples with the corresponding outcomes of two commonly used assessment methods of the traditional logic, the calculation of the mean values and of the Grade Point Average (GPA) index. Finally, the perspectives of future research on the subject are discussed.

BACKGROUND

The *Fuzzy Numbers* (FNs) play an important role in fuzzy mathematics analogous to the role played by the ordinary numbers in crisp mathematics. This section contains the background from the theory of FNs, which is necessary for the purposes of the present work.

Fuzzy Sets and Logic

The fuzzy sets theory was created in response of expressing mathematically real world situations in which definitions have not clear boundaries. For example, “the high mountains of a country”, “the young people of a city”, “the good players of a team”, etc. The notion of a *fuzzy set* (FS) was introduced by Zadeh (1965) as follows:

1. **Definition:** A FS on the universal set U of the discourse (or a fuzzy subset of U) is a set of ordered pairs of the form $A = \{(x, m_A(x)): x \in U\}$, defined in terms of a *membership function* $m_A: U \rightarrow [0,1]$ that assigns to each element of U a real value from the interval $[0,1]$.

The value $m_A(x)$ is called the *membership degree* of x in A . The greater is $m_A(x)$, the better x satisfies the characteristic property of A . The definition of the membership function is not unique depending on the user’s subjective data, which is usually based on statistical or empirical observations. However, a necessary condition for a FS to give a reliable description of the corresponding real situation is that its membership function’s definition satisfies the common sense. Note that, for reasons of simplicity, many authors identify a FS with its membership function.

A crisp subset A of U can be considered as a FS in U with $m_A(x) = 1$, if $x \in A$ and $m_A(x) = 0$, if $x \notin A$. In this way most properties and operations of crisp sets can be extended to corresponding properties and operations of Fs. For general facts on FS we refer to the book of Klir & Folger (1988).

Fuzzy Logic (FL), based on FS theory, constitutes a generalization and complement of the classical bi-valued logic that finds nowadays many applications to almost all sectors of human activities (e.g. Chapter 6 of Klir & Folger, 1988, Voskoglou, 2011, 2012, 2015 etc.). Due to its nature of characterizing the ambiguous real life situations with multiple values, FL offers, among others, rich resources for assessment purposes,

which are more realistic than those of the classical logic (Voskoglou, 2011, 2012, 2015, etc).

The FL approach for a problem's solution involves the following steps:

- *Fuzzification* of the problem's data by representing them with properly defined FSs.
- *Evaluation of the fuzzy data* by applying principles and methods of FL in order to express the problem's solution in the form of a unique FS.
- Defuzzification of the problem's solution in order to "translate" it in our natural language for use with the original real-life problem.

One of the most popular defuzzification methods is the *Center of Gravity (COG) technique*. When using it, the fuzzy outcomes of the problem's solution are represented by the coordinates of the COG of the membership's function graph of the FS involved in the solution (van Broekhoven & De Baets, 2006).

Fuzzy Numbers

2. **Definition:** A FN is a fuzzy set A on the set R of real numbers with membership function

$m_A: R \rightarrow [0, 1]$, such that:

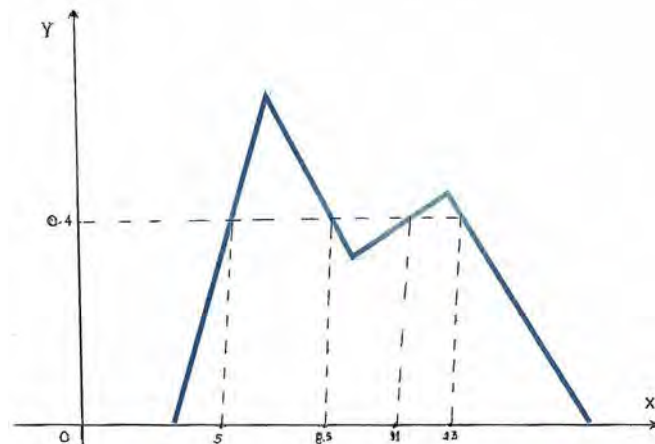
- A is *normal*, i.e. there exists x in R such that $m_A(x) = 1$,
 - A is *convex*, i.e. all its α -cuts $A^\alpha = \{x \in U: m_A(x) \geq \alpha\}$, α in $[0, 1]$, are closed real intervals, and
 - Its membership function $y = m_A(x)$ is a piecewise continuous function.
3. **Counter-Example:** Figure 1 represents the graph of a fuzzy set on R which is not convex and therefore it is not a FN. For example, it can be observed that $A^{0.4} = [5, 8.5] \cup [11, 13]$, therefore $A^{0.4}$ is not a closed interval.

Since the x -cuts A^x of a FN, say A , are closed real intervals, we can write $A^x = [A_l^x, A_r^x]$ for each x in $[0, 1]$, where A_l^x, A_r^x are real numbers depending on x .

The following statement defines a *partial order* on the set of all FNs:

4. **Definition:** Given the FNs A and B we write $A \leq B$ (or $\geq$) if, and only if, $A_l^x \leq B_l^x$ and $A_r^x \leq B_r^x$ (or $\geq$) for all x in $[0, 1]$. Two such FNs are called *comparable*, otherwise they are called *non comparable*.

Figure 1. Example of a non convex fuzzy set on R



5. **Remark:** One can define the four basic arithmetic operations on FNS in the following two, equivalent to each other, ways (Kaufmann & Gupta, 1991):

With the help of their α -cuts and the Representation-Decomposition Theorem of Ralescou - Negoita (Sakawa, 1993, Theorem 2.1, p.16) for FS. In this way the fuzzy arithmetic is turned to the well known arithmetic of the closed real intervals.

By applying the Zadeh's extension principle (Klir & Folger, 1988, Section 1.4, p.20), which provides the means for any function f mapping a crisp set X to a crisp set Y to be generalized so that to map fuzzy subsets of X to fuzzy subsets of Y .

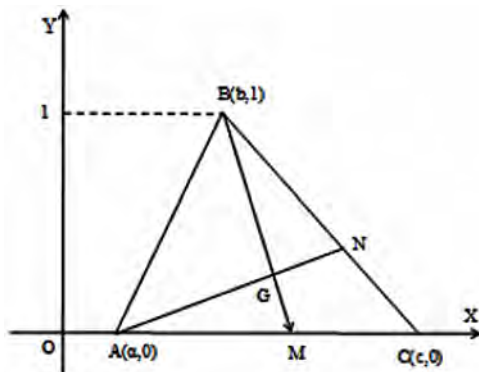
In practice the above two general methods of the fuzzy arithmetic, requiring laborious calculations, are rarely used in applications, where the utilization of simpler forms of FNs is preferred.

For general facts on FNs we refer to the classical on the subject book of Kaufmann and Gupta (1991).

Triangular Fuzzy Numbers (TFNs)

A TFN (a, b, c) , with a, b, c in $\mathbf{R}$ represents the fuzzy statement "the value of b lies in the interval $[a, c]$ ". The membership function of (a, b, c) is zero outside the interval $[a, c]$, while its graph in $[a, c]$ consists of two straight line segments forming a triangle with the OX axis (Figure 2).

Figure 2. Graph and COG of the TFN (a, b, c)



Therefore the analytic definition of a TFN is given as follows:

6. **Definition:** Let a, b and c be real numbers with $a < b < c$. Then the TFN (a, b, c) is a FN with membership function:

$$y = m(x) = \begin{cases} \frac{x-a}{b-a}, & x \in [a, b] \\ \frac{c-x}{c-b}, & x \in [b, c] \\ 0, & x < a \text{ or } x > c \end{cases}$$

The following two Propositions refer to basic properties of TFNs that we are going to use later in this work:

7. **Proposition:** The x -cuts A^x of the TFN $A = (a, b, c)$, $x \in [0, 1]$, are calculated by the formula $A^x = [a + x(b - a), c - x(c - b)]$.

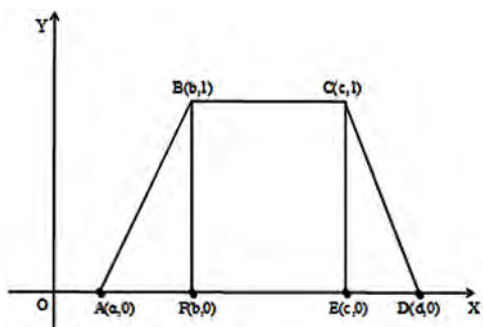
Proof: The x -cut $A^x = \{y \in \mathbf{R}: m(y) \geq x\}$ of the TFN A is a real interval of the form $A^x = [A_l^x, A_r^x]$, with A_l^x in $[a, b]$ and A_r^x in $[b, c]$ such that $m(A_l^x) = m(A_r^x) = x$. Therefore, by Definition 6 we

$$\text{have that } \frac{A_l^x - a}{b - a} = x \Leftrightarrow A_l^x = a + x(b - a). \text{ Similarly, by Definition 6 again, } \frac{c - A_r^x}{c - b} = x \Leftrightarrow A_r^x = c - x(c - b).$$

8. **Proposition:** (Defuzzification of a TFN) The coordinates (X, Y) of the COG of the graph of the TFN (a, b, c) are calculated by the formulas $X = \frac{a + b + c}{3}$, $Y = \frac{1}{3}$.

Proof: The graph of the TFN (a, b, c) is the triangle ABC of Figure 2, with $A(a, 0)$, $B(b, 1)$ and $C(c, 0)$. Then, the COG, say G , of ABC is the intersection point of its medians AN and BM . The proof of the Proposition is easily obtained by calculating the equations of AN and BM and by solving the linear system of these two equations.

Figure 3. Graph of the TpFN (a, b, c, d)



9. **Remark:** (*Arithmetic Operations on TFNs*) It can be shown (Kaufmann and Gupta, 1991) that the two general methods of defining arithmetic operations on FNs mentioned in Remark 5 lead to the following simple rules for the *addition* and *subtraction* of TFNs:

Let $A = (a, b, c)$ and $B = (a_p, b_p, c_p)$ be two TFNs. Then

- The sum $A + B = (a+a_p, b+b_p, c+c_p)$.
- The difference $A - B = A + (-B) = (a-c_p, b-b_p, c-a_p)$, where $-B = (-c_p, -b_p, -a_p)$ is defined to be the *opposite* of B .

In other words, the opposite of a TFN, as well as the sum and the difference of two TFNs are always TFNs. On the contrary, the *product* and the *quotient* of two TFNs, although they are FNs, they are not always TFNs, unless if a, b, c, a_p, b_p, c_p are in $\mathbf{R}^+$ (Kaufmann and Gupta, 1991).

The following two *scalar operations* can be also defined:

- $k + A = (k+a, k+b, k+c)$, $k \in \mathbf{R}$
- $kA = (ka, kb, kc)$, if $k > 0$ and $kA = (kc, kb, ka)$, if $k < 0$.

Trapezoidal Fuzzy Numbers (TpFNs)

A TpFN (a, b, c, d) with a, b, c, d in $\mathbf{R}$ represents the fuzzy statement *approximately in the interval* $[b, c]$. Its membership function $y=m(x)$ is zero

outside the interval $[a, d]$, while its graph in this interval $[a, d]$ is the union of three straight line segments forming a trapezoid with the X-axis (see Figure 3),

Therefore, the analytic definition of a TpFN is given as follows:

10. **Definition:** Let $a < b < c < d$ be given real numbers. Then the TpFN (a, b, c, d) is the FN with membership function:

$$y = m(x) = \begin{cases} \frac{x-a}{b-a} & , \quad x \in [a, b] \\ x = 1 & , \quad x \in [b, c] \\ \frac{d-x}{d-c} & , \quad x \in [c, d] \\ 0 & , \quad x < a \text{ and } x > d \end{cases}$$

It is easy to observe that the TpFNs are generalizations of TFNs. In fact, the TFN (a, b, d) can be considered as a special case of the TpFN (a, b, c, d) with $b=c$.

11. **Remark:** (*Arithmetic Operations on TpFNs*) It can be shown (Kaufmann and Gupta, 1991) that the addition and subtraction of two TpFNs are performed in the same way that it was mentioned in Remark 9 for TFNs. Also, the two scalar operations defined in Remark 9 for TFNs hold also for TpFNs.

12. **Proposition:** (*Defuzzification of a TpFN*) The coordinates (X, Y) of the COG of the graph of the TpFN (a, b, c, d) are calculated by the formulas $X = \frac{c^2 + d^2 - a^2 - b^2 + dc - ba}{3(c + d - a - b)}$, $Y = \frac{2c + d - a - 2b}{3(c + d - a - b)}$.

Proof: We divide the trapezoid forming the graph of the TpFN (a, b, c, d) in three parts, two triangles and one rectangle (Figure 3). The coordinates of the three vertices of the triangle ABE are $(a, 0)$, $(b, 1)$ and $(b, 0)$ respectively, therefore

by Proposition 8 the COG of this triangle is the point $C_1 (\frac{a+2b}{3}, \frac{1}{3})$. Similarly one finds that the COG of the triangle FCD is the point $C_2 (\frac{d+2c}{3}, \frac{1}{3})$. Also, it is easy to check that the COG of the rectangle BCFE, being the intersection point of its diagonals, is the point $C_3 (\frac{b+c}{2}, \frac{1}{2})$. Further, the areas of the two triangles are equal to $S_1 = \frac{b-a}{2}$ and $S_2 = \frac{d-c}{2}$ respectively, while the area of the rectangle is equal to $S_3 = c-b$.

Further, the coordinates of the COG of the trapezoid, being the resultant of the COGs $C_i (x_i, y_i)$, for $i=1, 2, 3$, are calculated by the formulas

$$X = \frac{1}{S} \sum_{i=1}^3 S_i x_i, Y = \frac{1}{S} \sum_{i=1}^3 S_i y_i \quad (1)$$

where $S = S_1 + S_2 + S_3 = \frac{c+d-b-a}{2}$ is the area of the trapezoid (Wikipedia, 2014).

The proof of the Proposition is completed by replacing the above found values of S, S_i, x_i and $y_i, i = 1, 2, 3$, in formulas (1) and by performing the corresponding operations.

MAIN FOCUS OF THE ARTICLE

A method is developed in this Section of using the TFNs and TpFNs as assessment tools, which is validated with the parallel use of two other traditional assessment methods, the calculation of the mean values and of the GPA index.. The examples presented concern the assessment of student and basket – ball player performance. The following definition is needed for this purpose:

13. **Definition:** Let $A_i = (a_{i1}, a_{i2}, a_{i3}, a_{i4}), i = 1, 2, \dots, n$ be TFNs (TpFNs), where n is a non negative integer, $n \geq 2$. Then we define the *mean value* of the A_i 's to be the TFN (TpFN)

$$A = \frac{1}{n} (A_1 + A_2 + \dots + A_n).$$

14. **Example:** The students of two different Departments of the School of Management and Economics of the Graduate Technological Educational Institute (T. E. I.) of Western Greece achieved the following scores (in a climax from 0 to 100) at their common progress exam in the course "Mathematics for Economists I":

First Department (D_1): 100(2 times), 99(3), 98(5), 95(8), 94(7), 93(1), 92(6), 90(5), 89(3), 88(7), 85(13), 82(6), 80(14), 79(8), 78(6), 76(3), 75(3), 74(3), 73(1), 72(5), 70(4), 68(2), 63(2), 60(3), 59(5), 58(1), 57(2), 56(3), 55(4), 54(2), 53(1), 52(2), 51(2), 50(8), 48(7), 45(8), 42(1), 40(3), 35(1).

Second Department (D_2): 100(1), 99(2), 98(3), 97(4), 95(9), 92(4), 91(2), 90(3), 88(6), 85(26), 82(18), 80(29), 78(11), 75(32), 70(17), 64(12), 60(16), 58(19), 56(3), 55(6), 50(17), 45(9), 40(6).

The student performance was characterized by the fuzzy linguistic labels (grades) A, B, C, D and F corresponding to the above scores as follows: A (85-100) = excellent, B (84-75) = very good, C (74-60) = good, D (59-50) = fair and F (<50) = unsatisfactory.

The two Departments performance will be assessed by using the two above mentioned traditional methods first and TFNs next.

- a. **Traditional Methods:** It is straightforward to calculate the mean values of the student scores of the two Departments, which are equal to 72.44 and 72.04 for D_1 and D_2 respectively. Therefore, the two Departments

demonstrated a good (C) *mean performance*, with the performance of D_1 being slightly better.

In contrast to the mean values of student scores, the *Grade Point Average* (GPA) index is a weighted average that measures the two Departments' *quality performance*, by assigning greater coefficients (weights) to the higher scores. Namely, the GPA index is calculated by the formula: GPA

$$= \frac{0n_F + 1n_D + 2n_C + 3n_B + 4n_A}{n} \quad (2) \text{ where } n$$

is the total number of students and n_A, n_B, n_C, n_D and n_F the numbers of students that demonstrated excellent (A), very good (B), good (C), fair (D) and unsatisfactory (F) performance respectively (Swinburne.edu.au, 2014).

In case of the ideal performance ($n_A = n$) formula (2) gives that GPA = 4, while in case of the worst performance ($n_F = n$) it gives that GPA = 0. Therefore, $0 \leq \text{GPA} \leq 4$, which means that values of GPA greater than 2 could be considered as indicating a more than satisfactory performance.

In our case, and in order to calculate the values of the GPA index for the two Departments, Table 1 is formed by depicting the student performance in terms of the grades A, B, C, D and F defined above.

Replacing the data of Table 1 to formula (2) and making the corresponding calculations one finds the same value $\text{GPA} = \frac{43}{17} \approx 2.53$ for the

Table 1. Student performance in terms of the linguistic labels

Grade	D_1	D_2
A	60	60
B	40	90
C	20	45
D	30	45
F	20	15
Total	170	255

two Departments that indicates a more than satisfactory quality performance.

- b. **Using TFNs:** A TFN (denoted by the same letter) is assigned to each linguistic label (grade) as follows: A = (85, 92.5, 100), B = (75, 79.5, 84), C = (60, 67, 74), D = (50, 54.5, 59) and F = (0, 24.5, 49). The middle entry of each of the above TFNs is equal to the mean value of the student scores assigned to the corresponding grade. In this way a TFN corresponds to each student assessing his (her) individual performance. The representation of the linguistic labels A, B, C, D and F by TFNs has the advantage of determining numerically the scores corresponding to each label. In fact, the scores assigned to the above labels in the statement of this Example are not standard, since they may differ from case to case. In a more rigorous assessment, for instance, one could take A(90-100), B (80-89), C(70-79), D (60-69), F(<60), etc.

It becomes now clear that Table 1 gives rise to 170 TFNs representing the progress of the students of D_1 and 255 TFNs representing the progress of the students of D_2 . Consequently, it is logical to accept that the overall performance of each Department is given by the corresponding mean value of the above TFNs (Definition 13). For simplifying our notation, let us denote the above mean values by the letter of the corresponding Department. Then, making straightforward calculations, one finds that

$$D_1 = \frac{1}{170} \cdot (60A + 40B + 20C + 30D + 20F) \approx (63.53, 71.74, 83.47), \text{ and}$$

$$D_2 = \frac{1}{255} \cdot (60A + 90B + 45C + 45D + 15F) \approx (65.88, 72.63, 79.53).$$

The left entries (63.53 and 65.88 respectively) and the right entries (83.47 and 79.53 respectively)

of the TFNs D_1 and D_2 determine approximately only the overall performance of the two Departments to be from good (C) to very good (B). Concerning the middle entries of D_1 and D_2 (71.74 and 72.63 respectively), since the middle entries of the TFNs A, B, C, D and F were chosen to be equal to the means of the scores assigned to the corresponding linguistic grades, they are simply equal to the mean values of these means and therefore they do not give a representative measure of the two Departments performance.

Applying Proposition 7 one finds that the x-cuts of the above two TFNs are $D_1^x = [63.53 + 8.21x, 83.47 - 11.73x]$ and $D_2^x = [65.88 + 6.75x, 79.53 - 6.9x]$ respectively. But $63.53 + 8.21x \leq 65.88 + 6.75x \Leftrightarrow 1.46x \leq 2.35 \Leftrightarrow x \leq 1.61$, which is true, since x is in $[0, 1]$. On the contrary, $83.47 - 11.73x \leq 79.53 - 6.9x \Leftrightarrow 3.94 \leq 4.83x \Leftrightarrow 0.82 \leq x$, which is not true for all the values of x . Therefore, according to Definition 4, the TFNs D_1 and D_2 are not comparable, which means that at this stage one can not decide which of the two Departments demonstrates the better performance.

A good way to overcome this difficulty is to defuzzify the TFNs D_1 and D_2 . By Proposition 8, the COGs of the triangles forming the graphs of the TFNs D_1 and D_2 have x-coordinates equal to $X = \frac{63.53 + 71.74 + 83.47}{3} \approx 72.91$ and $X' = \frac{65.88 + 72.63 + 79.53}{3} \approx 72.68$ respectively.

Observe now that the GOGs of the graphs of D_1 and D_2 lie in a rectangle with sides of length 100 units on the X-axis (student scores from 0 to 100) and one unit on the Y-axis (normal fuzzy sets). Therefore, *the nearer the x-coordinate of the COG to 100, the better the corresponding Department's performance*. Thus, since $X > X'$, D_1 demonstrates a (slightly) better overall performance than D_2 .

15. **Example:** The individual performance of the five players of a basket-ball team who started a game was assessed by six different athletic journalists using a scale from 0 to 100 as follows: P_1 (player 1): 43, 48, 49,

49, 50, 52, P_2 : 81, 83, 85, 88, 91, 95, P_3 : 76, 82, 89, 95, 98, P_4 : 86, 86, 87, 87, 87, 88 and P_5 : 35, 40, 44, 52, 59, 62. The players' performance is characterized by the fuzzy linguistic labels A, B, C, D and F defined in Example 14.

Here the five players' overall performance will be assessed by using the above mentioned two traditional methods first and also by using TFNs and TpFNs next.

a. **Traditional Methods:** Adding the $5 * 6 = 30$ in total scores assigned to the players by the journalists and dividing the corresponding sum by 30 one finds that the *mean performance* of the five players is approximately equal to 60.05, i.e. it can be marginally characterized as good (C).

Further, it can be observed that 14 of the above, 30 in total, scores correspond to the label A, four to B, one to C, four to D and seven to F. Replacing these values to formula (2) one finds that the GPA index is approximately equal to 2.47. Therefore, the five players' overall quality performance is characterized as more than satisfactory.

b. **Using TFNS:** Working as in Example 14, one forms the TFNs A, B, C, D and F. Further, observing the $5 * 6 = 30$ in total player scores it becomes clear that 14 of them correspond to the TFN A, four to B, one to C, four to D and seven to F.

The mean value of the above TFNs (Definition 13) is equal to $M = \frac{1}{30} (14A + 4B + C + 4D + 7F) \approx (58.33, 68.98, 79.63)$. Therefore, the players' overall performance lies in the interval $[58.33, 79.63]$, i.e. it could be characterized from fair (D) to very good (B).

Further, defuzzifying the TFN M (Proposition 8) one finds the value 68.98 for the x-coordinate of the COG of its graph, which finally shows that

the five players demonstrated a good (C) overall performance.

- c. **Using TpFNs:** A TpFN (denoted, for simplicity, by the same letter) is assigned to each basket-ball player P_i as follows: $P_1 = (0, 43, 52, 59)$, $P_2 = (75, 81, 95, 100)$, $P_3 = (75, 76, 98, 100)$, $P_4 = (85, 86, 88, 100)$ and $P_5 = (0, 35, 62, 74)$. Each of the above TpFNs describes numerically the individual performance of the corresponding player in the form (a_1, a_2, a_3, a_4) , where a_1 and a_4 are the lower and upper bounds respectively of his performance with respect to the linguistic labels A, B, C, D and F, while a_2 and a_3 are the lower and higher scores respectively assigned to the corresponding player by the athletic journalists. The mean value of the TpFNs P_i , $i = 1, 2, 3, 4, 5$ (Definition 13), is equal to

$P = \frac{1}{5} \sum_{i=1}^5 P_i = (47, 64.2, 79, 86.6)$, which gives the following information:

- The players' performance, is fluctuated from unsatisfactory ($a_1=47$) to excellent ($a_4=86.6$).
- The overall players' performance is lying in the interval $[a_2, a_3] = [64.2, 79]$, i.e. it can be characterized from good (C) to very good (B).

Further, defuzzifying the TpFN P (Proposition 12) one finds the value

$$\frac{79^2 + (86.6)^2 - (64.2)^2 - 47^2 + 79 * 86.6 - 47 * (64.2)}{3(79 + 86.6 - 47 - 64.2)} \approx 68.84$$

for the x-coordinate of the COG of its graph, which finally shows that the five players demonstrated a good (C) overall performance.

SOLUTIONS AND RECOMMENDATIONS

The outcomes obtained from the application of the four in total assessment methods used in our Examples are depicted in Tables 2 and 3. The data of these Tables show that the outcomes of the two new fuzzy assessment methods (TFNs/TpFNs) are compatible to the corresponding outcomes of the two traditional assessment methods (mean values / GPA index). This is a strong indication that the two fuzzy methods “behave” well (validation of the fuzzy methods), which was also crossed with different problems in earlier author's works (Voskoglou, 2015, etc.). The small numerical differences appeared are due to the different “philosophy” of the four methods (mean and quality performance, bi-valued and fuzzy logic).

The approximation of the overall player performance obtained in the first step of case iii of Example 15 (good to very good) is better than that obtained in the first step of case ii (fair to very good), because the TpFNs, due to the way of their construction, measure more accurately than the TFNs the player performance. However, it is not always easy in practice to use TpFNs instead of TFNs. In Example 14, for instance, there are

Table 2. Outcomes of experiment 14

Method	D_1	D_2	Performance
Mean values	72.44	72.04	Good (C)
GPA index	2.53	2.53	More than satisfactory
TFNs	72.91	72.68	Good (C)

Table 3. Outcomes of experiment 15

Method		Performance
Mean values	60.05	Good (C)
GPA index	2.47	More than satisfactory
TFNs	68.98	Good (C)
TpFNs	68.84	Good (C)

$170+255=425$ students under assessment, which means that one has to form a great number of TpFNs for evaluating their performance, resulting to laborious calculations.

Another advantage of using TpFNs as assessment tools is that they enable, in contrast to TFNs, the comparison of the individual performance of any two persons. In case iii of Example 15, for instance, it is enough to defuzzify the TpFNs P_i , $i=1, 2, 3, 4, 5$ corresponding to the players' individual performance, as we did for P and to apply the same comparison criterion.

FUTURE RESEARCH DIRECTIONS

The new method of using TFNs/TpFNs for student and basket-ball player assessment developed here has the potential of a general assessment method. This means that it could be utilized for assessing a great variety of other human or machine (e.g. Case – Based Reasoning or Decision – Making systems) activities. This is the main target of our future research on the subject.

CONCLUSION

In the present paper a combination of TFNs / TpFNs and of the COG defuzzification technique was used for assessing human skills. Examples for student and basket-ball player assessment were presented illustrating the results obtained and the new fuzzy assessment method was validated by comparing its outcomes in these examples with the corresponding outcomes of two traditional assessment methods, the calculation of the mean values and of the GPA index. The main advantage of the new fuzzy method with respect to the traditional ones is that, due to the FL's property of characterizing the ambiguous assessment situations with multiple values, it fits better to such kind of situations.

REFERENCES

- Kaufmann, A., & Gupta, M. (1991). *Introduction to Fuzzy Arithmetic*. New York: Van Nostrand Reinhold Company.
- Klir, G. J., & Folger, T. A. (1988). *Fuzzy Sets, Uncertainty and Information*. London: Prentice-Hall.
- Sakawa, M. (1993). *Fuzzy Sets and Interactive Multiobjective Optimization*. London: Plenum press. doi:10.1007/978-1-4899-1633-4
- Swinburne.edu.au. (2014). *Grade Point Average Assessment*. Retrieved on October 15, 2014 from: <http://www.swinburne.edu.au/studentadministration/assessment/gpa.html>
- Van Broekhoven, E., & De Baets, B. (2006). Fast and accurate centre of gravity defuzzification of fuzzy systems outputs defined on trapezoidal fuzzy partitions. *Fuzzy Sets and Systems*, 157(7), 904–918. doi:10.1016/j.fss.2005.11.005
- Voskoglou, M. G. (2011). Stochastic and fuzzy models in Mathematics Education, Artificial Intelligence and Management. Lambert Academic Publishing.
- Voskoglou, M. Gr.(2015) Fuzzy Assessment Processes. In M. Roushdy, & T. Nazmy (Eds.), *Proceedings of the 7th IEEE International Conference on Intelligent Computing and Information Systems (ICICIS 15)*. Ain Shams University.
- Voskoglou, M. G. (2012). A Study on Fuzzy Systems. *American Journal of Computational and Applied Mathematics*, 2(5), 232–240. doi:10.5923/j.ajcam.20120205.06
- Wikipedia.org (2014). *Center of mass: A system of particles*. Retrieved on October 10, 2014 from: http://en.wikipedia.org/wiki/Center_of_mass#A_system_of_particles
- Zadeh, L. A. (1965). Fuzzy Sets. *Information and Control*, 8(3), 338–353. doi:10.1016/S0019-9958(65)90241-X

ADDITIONAL READING

Hanss, M. (2005). *Applied Fuzzy Arithmetic, An Introduction with Engineering Applications*. Berlin, Heidelberg: Springer-Verlag.

Jones, A., Kaufmman, A., & Zimmerman, H. J. (1986). *Fuzzy Sets: Theory and Applications, NATO ASI Series, Series C: Mathematical and Physical Sciences* (Vol. 177). Dordrecht, Holland: Reidel Publishing Company.

Subbotin, I. Ya. (2014), Trapezoidal Fuzzy Logic Model for Learning Assessment, *arXiv*, 1407.0823 [math.GM].

Subbotin, I. Ya., Badkoobei, H., & Bilotskii, N. N. (2004). Application of fuzzy logic to learning assessment. *Didactics of Mathematics: Problems and Investigations*, 22, 38–41.

Subbotin, I. Ya., & Bilotskii, N. N. (2014). Triangular fuzzy logic model for learning assessment. *Didactics of Mathematics: Problems and Investigations*, 41, 84–88.

Subbotin, I. Ya, & Voskoglou, M. Gr. (2014), A Triangular Fuzzy Model for Assessing Critical Thinking Skills, *International Journal of Applications of Fuzzy Sets and Artificial intelligence*, 4, 173 -186.

Subbotin, I. Y., & Voskoglou, M. G. (2016). An Application of the Generalized Rectangular Fuzzy Model to Critical Thinking Assessment. *American Journal of Educational Research*, 4(5), 397–403.

Theodorou, J. (2010), *Introduction to Fuzzy Logic*, Tzolas Publications, Thessaloniki, Greece (in Greek language).

Voskoglou, M. Gr. & Subbotin, I. Ya. (2015). Fuzzy Methods for Student Assessment. *International Journal of Education and Information Technology*, 1(1), 20–28.

Voskoglou, M. G. (2013). Assessing Students Individual Problem Solving Skills. *International Journal of Applications of Fuzzy Sets and Artificial Intelligence*, 3, 39–49.

KEY TERMS AND DEFINITIONS

Centre of Gravity (COG) Defuzzification Technique: A defuzzification technique in which the corresponding system's fuzzy outputs are represented by the coordinates of the COG of the level's section contained between the graph of the membership function involved and the OX axis. The COG technique can be used, under a proper manipulation, as an assessment method of the fuzzy system's quality performance.

Defuzzification: The process of representing a system's fuzzy outputs by a crisp number.

Grade Point Average (GPA): A weighted average assigning greater coefficients (weights) to the higher scores and therefore focusing on a system's quality performance.

Fuzzy Arithmetic: The extension of the four classical arithmetic operations of crisp numbers to fuzzy numbers.

Fuzzy Assessment Methods: Assessment methods based on the principles of fuzzy logic.

Fuzzy Logic (FL): A logic based on the concept of fuzzy set that, in contrast to the classical bi-valued logic (yes – no), characterizes a case with multiple values.

Fuzzy Number (FN): A fuzzy set A on the set R of real numbers which is *normal* (i.e. there exists x in R such that $m_A(x) = 1$) and *convex* (i.e. all its α -cuts $A^\alpha = \{x \in U: m_A(x) \geq \alpha\}$, α in $[0, 1]$, are closed real intervals) and whose membership function $y = m_A(x)$ is a piecewise continuous function.

Fuzzy Set (FS): A generalization of the concept of crisp set that gives a mathematical formulation to real situations in which certain definitions have not clear boundaries (e.g. the high mountains of a country). A FS, say A , is characterized by a membership function $y = m_A(x)$ defined on the universal set of the discourse U and taking values in the interval $[0, 1]$, thus assigning to each element x of U a membership degree $m_A(x)$ with respect to A . The closer is $m_A(x)$ to 1, the better x satisfies the characteristic property of A .

Trapezoidal Fuzzy Number (TpFN): A fuzzy number whose membership function's graph forms a trapezoid with the OX-axis. The TpFNs are generalizations of the triangular fuzzy numbers.

Triangular Fuzzy Number (TFN): The simplest form of a fuzzy number, whose membership function's graph forms a triangle with the OX-axis.

F

Application of Soft Set in Game Theory

B. K. Tripathy

VIT University, India

Sooraj T. R.

VIT University, India

Radhakrishna N. Mohanty

VIT University, India

INTRODUCTION

Soft set theory is a new mathematical approach to handle uncertainty based problems. It was introduced in (Molodtsov, 1999). Later on, fuzzy soft sets is defined in (Maji et al. 2001) and provided an application of soft set theory in decision-making problems (Maji et al., 2002). There are several theories which deal with vagueness and ambiguity. Some of these are: probability theory, interval mathematics and fuzzy set theory. All these models have their own drawbacks; such as fuzzy sets (Zadeh, 1965) are completely dependent on membership functions. There is no unique formula to define membership functions. Molodtsov observed that the main reason of these drawbacks is perhaps due to the inadequacy of parameterization tools. That led him to introduce soft set theory, which is a parameterized collection of subsets.

Game theory is defined as the mathematical model of interaction between rational, intelligent decision makers. In this chapter, we mention how game theory and soft sets are related. When we handle real life situations, we can observe that most of the information available is ambiguous or uncertain. Now a days, in most of the games we need to create human models or human behaviour. There exist a lot of approaches to describe human

behavior in game theory. Some of them are pay function, choice functions etc. Molodtsov has introduced s-function (soft function) which keeps all good sides of choice function and eliminates the drawbacks of pay function and choice functions (Molodtsov, 1999). Deli and Cagman has studied further to associate soft set theory in game theory (Deli et al., 2013).

This chapter is further organized into ten sections. The next section contains definitions and notions of soft set theory. In the section three we have given some descriptions about the game theory. In the section four we have provided classification of games. Section five illustrates the representation of games in some subsections subsequently. In section six the definitions and notions related to both game theory and soft set theory are discussed. Section seven provides the definitions and notions about soft sets in decision making. In the section eight we provided the soft sets in decision making based on game theory which further describes about two-person soft game and n-person soft game under the subsections subsequently. Section nine provides future scope of this research work. Finally provide conclusions drawn from our research work. The chapter ends with a bibliography of sources referred for the compilation of our work.

BACKGROUND

Soft set is a parameterized family of subsets defined over a universe associated with a set of parameters.

The definition of soft set is given below.

Definition 1 (Molodtsov, 1999): A pair (F, E) is called a soft set over U iff F is a mapping of E into the set of all subsets of the universal set U ; i.e.

$$F: E \rightarrow P(U) \quad (1)$$

where U is the universal set, E is the parameter set and $P(U)$ is the power set U .

In other words, a soft set over U is a parameterised family of subsets of the universe U . For $e \in E$, $F(e)$ can be called as the set of e -approximate elements of the soft set (F, E) . So, a soft set can be represented as a collection of approximations. The parameter part of the approximation is called as predicate and for each parameter in E and the set containing all the elements of $F(e)$ is called the value set of e in (F, E) .

The pair (U, E) is often regarded as a soft universe. A parameter can be anything adverbial for the elements, such as a number, word, phrase or a sentence which can describe the value set more appropriately.

The way of describing an object in soft set theory differs principally from the way an object is described in classical mathematics. Normally, in classical mathematics a mathematical model of an object is constructed to define the notion of the exact solution of that model. Sometimes the mathematical model becomes too complicated to find the exact solution. So, we need the notion of approximate solution to get rid of these types of problems. However, in soft set theory, we are getting the solution to the problems by the opposite approach. The description of the object will have an approximate nature initially and so that we do not need to use the notion of exact solution.

There are no restrictions on the approximate description of the objects in soft set theory, which makes this model very convenient and can be easily applied in real life problems. We can use any type of parameterisation depending upon the preferences with the help of words and sentences, mappings, functions, real numbers, and so on.

Definition 2 (Maji et al., 2001): The pair (F, E) can be called as a fuzzy soft set over U , where F is a mapping given by

$$F: A \rightarrow I^U \quad (2)$$

where U is the initial universal set, E is the set of parameters, I^U is the power set of all fuzzy subsets of U and $A \subseteq E$.

For any element of E , the possible number of degrees of belongingness are real numbers lying in the interval of $[0, 1]$ and hence infinite.

GAME THEORY

Game theory was introduced in (Von Neumann, 1944). Neumann is also known as the father of game theory. Game theory is a type of decision theory in which one's choice of action is determined after taking into account all possible alternatives available to an opponent playing the same game, rather than just by the possibilities of several outcomes.

In this section we discuss about strategic form games, the representation of games and the classification of games. Strategic form games are also called normal game forms.

Definition 3 (Strategic Form Games): A strategic form game ' Γ ' is a 3-tuple $\langle N, S_i, u_i \rangle$ where $i \in N$, N is the finite set of players, i.e., $N = \{1, 2, \dots, n\}$. S_i 's are the strategy sets of players and

$$u_i : S_1 \times S_2 \times \dots \times S_{n-1} \times S_n \rightarrow \mathbb{R}$$

for $i=1, 2, \dots, n$ are mappings called the utility functions or payoff functions.

The strategies are also called actions or pure strategies. Denoting the Cartesian product $S_1 \times S_2 \times \dots \times S_n$ by S , we have S is the collection of all strategic profiles of the players. The utility of an agent depends not only on its own strategy but also strategy of the rest of the players.

Notions Used in Game Theory

In this section we provide some of the concepts and notations to be used in this chapter.

Utilities

Utility theory is the science of assigning numbers to outcomes in a way that captures the preferences of the players. In any game, utility represents the motivation of players. A utility function for a given player assigns a number for every possible outcome of the game with the property that a higher number implies that the outcome is more preferred. Von Neumann and Morgenstern stated and proved expected *utility maximization theorem* in (Neumann et al., 1944), which establishes that for any rational decision maker that there must exist a way of assigning utility numbers to different outcomes in a way that the decision maker would always choose the option that maximizes his expected utility.

Intelligence

Intelligence is another notion that we use in game theory. This notion connotes that each player in the game knows everything about the game that a game theorist knows, and the player is competent enough to make any inferences about the game that a game theorist can make.

Common Knowledge

The definition of common knowledge as follows: A fact is the common knowledge among the players if every player knows it, every player knows that every player knows it, and so on. That is, every statement of the form “every player knows that every player knows that ... every player knows it” is true and infinitum. (Aumann, 1976)

Bounded Reality

In (Osborne et al., 1994), the concept of bounded rationality is given as “Game theory, in its most common form, assumes that all the players are symmetric; that is, they have identical capabilities of perception and computation. It does not model asymmetries in abilities or perceptions of situations.”

CLASSIFICATION OF GAMES

In this section we provide the broad categories into which Games can be classified. These are,

1. **Non-Cooperative Games and Cooperative Games:** Non cooperative games are those in which the actions of individual players are the primitives. In cooperative games, joint actions of group of players are the primitives.
2. **Games With Perfect Information and Games with Imperfect Information:** If the player is having full information about the past history of the game, the game is said to be of perfect information. Otherwise game is said to be of imperfect information.
3. **Games With Complete Information and Incomplete Information:** A game with complete information is one in which, at the first point in time when the players can begin to plan their moves, some players have private information about the game that other players don't know.

REPRESENTATION OF GAMES

The primary categorization of games is that of cooperative and non cooperative games. Most of the cooperative games are represented by using characteristic function form, while the extensive and the normal forms are used to define non cooperative games. These forms are explained in the following sections.

Extensive Form

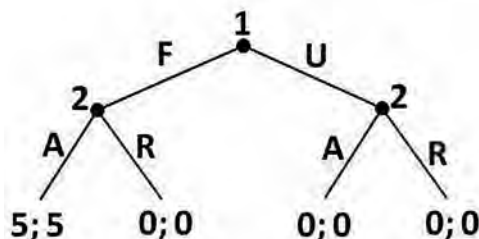
The extensive form can be used to formalize games with a time sequencing of moves. Games here are played on trees. An example is shown below.

In the game tree representation, there are three types of nodes.

1. Root node (initial decision node)
2. Decision node (which are internal nodes)
3. Terminal nodes (Leaf nodes).

When the game actually takes place, the path that represents the sequence of events is called the path of play. Each decision node is labeled with the player who takes a decision at that node. The links that are outgoing at the decision node are labeled with the actions the player may select at that node. Here each vertex or node represents a point of a choice for a player. A player is specified by a number listed by the vertex. The lines out of the vertex represent a possible action for that player. The payoffs are specified at the bottom of the tree. The extensive form can be viewed as a multi-player generalization of a decision tree.

Figure 1. An extensive form game



The extensive form can also capture simultaneous-move games and games with imperfect information. To represent it, either a dotted line connects different vertices to represent them as being part of the same information set (i.e. the players do not know at which point they are), or a closed line is drawn around them.

Normal Form

The normal (Strategic form) game is usually represented by a matrix which shows the players, strategies and payoffs. It can be represented by any function that associates a payoff for each player with every possible combination of actions. Suppose there are two players, one chooses the row and the other chooses the column. Each player has two strategies which are specified by the number of columns and number of rows. The first number is the payoff received by the row player (player 1), second number is the payoff received for the column player (player 2). Suppose that Player 1 plays Up and that Player 2 plays Left. Then Player 1 gets a payoff of 4, and Player 2 gets 3. When a game is presented in normal form, it is presumed that each player acts simultaneously or, at least, without knowing the actions of the other. If players have some information about the choices of other players, the game is usually presented in extensive form.

Characteristic Function Form

In games that possess removable utility separate rewards are not given; rather, the characteristic function decides the payoff of each unity.

Table 1. Normal form game

	Player 2 Chooses Left	Player 2 Chooses Right
Player 1 Chooses Up	4, 3	-1, -1
Player 1 Chooses Down	0, 0	3, 4

The origin of this form is to be found in John von Neumann and Oskar Morgenstern's book; when looking at these instances, they guessed that when a union C appears, it works against the fraction (N/C) as if two individuals were playing a normal game. The balanced payoff of C is a basic function. Although there are different examples that help to determine coalitional amounts from normal games, not all appear that in their function form can be derived from such. Formally, a characteristic function is seen as: (N, v) , where N represents the group of people and $v: 2N \rightarrow R$ is a normal utility. Such characteristic functions have expanded to describe games where there are no removable utilities.

In (Molodtsov, 1991) the concept of optimality principles is defined as a person's mathematical model instead of payoff function. Optimality principle is stated as "An optimal policy has the property that whatever the initial state and initial decision are the remaining decision must constitutes an optimal policy with regard to the state resulting from the first decision". Optimality principle is considered as the generalization of choice function. Choice function is a mapping which associates a set of strategies with a given situation.

GAME THEORY AND SOFT SETS

In this section, we discuss how soft sets are related with game theory. Molodtsov discussed some

applications of soft sets (Molodtsov, 1999). Among these applications, he mentioned how soft sets are related with game theory and operations research. There are a lot of ways to describe human behaviour in game theory. Normally we use payoff function for modeling human behaviour. Payoff function for a player is a mapping from the cross-product of players strategy spaces to the players set of payoffs. i.e., payoff function of a player takes as its input as a strategy profile and yields a representation of payoffs as its output. But the construction of the pay function is a difficult problem. As we have mentioned earlier, there are some strategies which match according to the situations. From these available strategies, we have to choose the best possible solution for each situation. A mapping which associates a set of strategies with a given situation is called choice function. But it has some constraints. The choice function cannot give appropriate description of various compromises and concessions which are typical for such problems. Soft set theory gives an opportunity to construct a new mathematical tool which keeps all good sides of choice function and eliminates its drawbacks. Here, soft set means to point out different strategies of each and every player. Molodtsov has suggested the concept of a s-function. For any set of strategies there exists a set ϵ of optimal choices and the user has to select suitable s-functions for describing human's behaviour under uncertainty. But the question is: which s-function has to be chosen for describing human behaviour?

Table 2. Symbols used

Symbols	Meaning
n	The number of players
S_i	The set of strategies of player 'i'
E	The set of parameters of player 'i'
$M(S)$	The set of all sub sets of the set S
$(F_i, E_i) : M(S) \rightarrow S, (F_i, E_i)$	soft choice function for player 'i'

Molodtsov introduced some approaches to this problem for different types of uncertainties. He has given a formal definition of the game with s -function for modeling a person's behaviour. The important symbols used are shown in Table 2.

If $P \subset S$, that is, P is a subset of admissible strategies and ε is a parameter, $\varepsilon \in E_i$, then $F_i(P, \varepsilon)$ is a set of ε -optimal situations for player 'i'. This type of games are said to be in a normal form with the following notation.

$$\langle (F_i, E_i), S_i, i = 1, \dots, n \rangle.$$

In game theory, Nash equilibrium is a concept to find out the solution of a non cooperative game involving two or more players. Here, each player is assumed to know the equilibrium strategies of the other players and no other player has anything to gain by changing their own strategy. If a player has chosen a strategy and no player can benefit by changing strategies while the other players keep their plans unchanged, then the current set of strategy choices and their corresponding payoffs constitutes a Nash equilibrium.

For a soft game to be in normal form, the analogue of the Nash equilibrium is the following construction.

Definition 4: A Situation $s \in S$ is called a situation of soft-equilibrium $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)$, $\varepsilon_i \in E_i$, if and only if

$$s \in F_i(s_1 \times \dots \times s_{i-1} \times S_i \times s_{i+1} \times \dots \times s_n, \varepsilon_i),$$

for every $i=1,2,\dots,n$. The set of all situations of soft ε -equilibrium is denoted by $N(\varepsilon)$. Then the soft set $(N, E_1 \times \dots \times E_n)$ is called a soft equilibrium. The soft guarantee concept is discussed below.

Suppose if there is only one player and

X is the set of strategies of the player,

E is the set of parameters of the player,

S is the set of situations.

$(F, E): M(S) \rightarrow S$ is a soft choice function of the player.

Point set mapping is a mapping that associates a number of different number of elements of the second set with the same element of the first. Let π be a point set mapping

$$\pi : X \rightarrow M(S) \quad (3)$$

The player knows that if he chooses the strategy $x \in X$, then he will get one of the situations, $s \in \pi(x)$. Consider the soft set $(\text{Gar}F, E)$, where

$$\begin{aligned} \text{Gar}F(X, \varepsilon) &= \{x \in X \mid \pi(x) \subset F(\pi(x), \varepsilon)\}, \\ \pi(x) &= \bigcup_{x \in X} \pi(x) \end{aligned} \quad (4)$$

is a soft guarantee concept. Here, the notion of Stackelberg solution is applied to the soft sets. Suppose there are two players in normal form

$$\langle (F_i, E_i), S_i, i = 1, 2 \rangle.$$

Stackelberg leadership model is a strategic game in economics in which the leading firm moves first and then the follower firms move sequentially.

Player 1 makes the first move. He chooses one strategy $s_1 \in S_1$ and inform player 2 about his choice. Since player 2 knows the strategy s_1 , he chooses his strategy $s_2 \in S_2$ so that

$$(s_1, s_2) \in F_2(s_1 \times s_2, \varepsilon_2).$$

It is assumed that player 1 knows the value of the parameter $\varepsilon_2 \in E_2$. Player 1 considers uncertain the possible choices of player 2. Applying guarantee concept, we will get soft Stackelberg set $(St_1, E_1 \times E_2)$ for player 1, where

$$St_1(S_1 \times S_2, \varepsilon_1, \varepsilon_2) = \{s_1 \in S_1 \mid F_2(s_1 \times S_2, \varepsilon_2) \subset F_1(S_1 \times S_2, \varepsilon_1)\} \quad (5)$$

SOFT SETS IN DECISION MAKING

An application of soft sets in decision making theorem is given in (Maji et al, 2002). In game theory, players have to take decisions according to the situations they face. So, we can use the decision making concept in game theory to take a particular strategy which suits to the situation.

Definition 5: Comparison table of a fuzzy soft set (F, P) is a square table in which number of rows and number of columns is equal. Rows and columns both are labeled by the object names $s_1, s_2, \dots, s_n$ of the universe, and the entries are c_{ij} where c_{ij} represents the number of parameters for which the membership value of h_i exceeds or equal to the membership value of h_j . Clearly, $0 \leq c_{ij} \leq k$ and $c_{ii} = k, \forall i, j$ where k is the number of parameters in P.

Definition 6: Row sum of an object 'h' is denoted by ' r_i ' and is calculated by using the formula

$$r_i = \sum_{j=1}^n c_{ij} \quad (6)$$

Here r_i , indicates the total number of parameters in which h_i dominates all the members of U.

Column sum of an object h_j is denoted by t_j and is calculated by using the formula

$$t_j = \sum_{i=1}^n c_{ij} \quad (7)$$

Definition 7: Score of an object h_j is denoted by

$$t_j \text{ and is calculated by using the formula } S_i = r_i - t_i.$$

The following algorithm is used to select a suitable strategy Mr. X would like to play.

Algorithm 1:

Step 1: Input the fuzzy soft set (F, E).

Step 2: Input the set P of choice parameters of Mr. X which is a subset of E.

Step 3: Consider the fuzzy soft set (F,P) and we write in tabular form.

Step 4: Compute the comparison table of fuzzy soft set (F,P).

Step 5: Compute r_i and t_i for $h_i, \forall i$.

Step 6: Compute the score of $h_i \forall i$.

Step 7: Find $S_i = \max S_i$. Then Mr. X can choose the strategy ' s_i '.

Consider a fuzzy soft set (F, E). Suppose that $P = \{e_1, e_2, e_3, e_4\}$ is the set of choice parameters for Mr. X on the basis of which he wants to choose a strategy from the availability U. Let $s_1, s_2, \dots$ denote the strategies. Tabular representation of the fuzzy soft set (F, P) is shown in Table 3.

The comparison-table of the fuzzy soft set shown in Table 3 is computed and shown in Table 4.

Next, we need to compute row-sum, column sum, score for each ' s_i ' as shown in the Table 5.

In Table 5, the maximum score is 13, scored by the strategy ' s_3 '. From this, we can decide that player X can choose s_3 . In case, he doesn't want to choose ' s_3 ' due to certain reasons, his second choice will be ' s_1 '.

Table 3. Representation of Fuzzy Soft Set

U	e_1	e_2	e_3	e_4
s_1	1.0	0.2	1.0	1.0
s_2	0.4	0.3	0.0	0.1
s_3	1.0	1.0	1.0	0.5
s_4	0.4	1.0	0.2	0.3
s_5	0.6	1.0	1.0	0.2
s_6	0.8	0.0	0.2	0.3

Table 4. Comparison table for the Fuzzy Soft Set (F, P)

U	s ₁	s ₂	s ₃	s ₄	s ₅	s ₆
s ₁	4	3	3	3	3	4
s ₂	1	4	0	1	0	1
s ₃	3	4	4	4	4	4
s ₄	1	4	1	4	2	3
s ₅	2	4	2	3	4	2
s ₆	0	3	0	3	2	4

Table 5. Computation of Row-Sum, Column-Sum and Score for each 's_i'

U	Row-Sum(r _i)	Column-Sum(t _j)	Score(S _i)
s ₁	20	11	9
s ₂	7	22	-15
s ₃	23	10	13
s ₄	15	18	-3
s ₅	17	15	2
s ₆	12	18	-6

SOFT SETS IN DECISION MAKING BASED ON GAME THEORY

Soft games can be applied to problems contain vagueness and uncertainties. The following sections discuss two-person soft games and n-person soft games.

Two Person Soft Games

This was introduced by Deli and Cagman (2013). They created two person soft games with the help of soft payoff functions.

Definition 8: Let U be a set of attributes, P(U) be the power set of U, X, Y are a set of strategies. Then, a set valued function

$$f_s : X \times Y \rightarrow P(U) \quad (8)$$

is called a soft payoff function. For each $(x, y) \in X \times Y$, the value $f_s(x, y)$ is called a soft payoff.

Definition 9: Let X and Y be a set of strategies of player 1 and player 2, U be a set of alter-

natives and $f_{s_k} : X \times Y \rightarrow P(U)$ be a soft payoff function for player k, where $k=1, 2$. Then for each player k, a two person soft game (tps-game) is defined by a soft set over U as

$$S_k = \{((x, y), f_{s_k}(x, y)) : (x, y) \in X \times Y\} \quad (9)$$

The tps-game is played as follows: at a certain time Player 1 chooses a strategy $x_i \in X$, simultaneously Player 2 chooses a strategy $y_j \in Y$ and once this is done each player k receives the soft payoff $f_{s_k}(x_i, y_j)$.

If $X=\{x_1, x_2, \dots, x_m\}$ and $Y=\{y_1, y_2, \dots, y_n\}$, then the soft payoffs of S_k can be arranged in the form of the $m \times n$ matrix shown in Table 6.

Let's consider an example. Let $U=\{u_1, u_2, \dots, u_{10}\}$ be a set of alternatives, P(U) be the power set of U, $X=\{x_1, x_3, x_5\}$ and $Y=\{x_1, x_2, x_4\}$ be the set of strategies of Player 1 and 2 respectively. If player 1 constructs a tps-game as shown in Box 1.

Then the soft payoffs of the game can be arranged as in Table 7.

Box 1.

$$s_1 = \left\{ \begin{aligned} &((x_1, x_1), \{u_1, u_2, u_5, u_8\}), (x_1, x_2), \{u_1, u_2, u_3, u_4, u_4, u_8\}), (x_1, x_4), \{u_3, u_8\}), \\ &(x_3, x_1), \{u_1, u_3, u_7\}), (x_3, x_2), \{u_1, u_2, u_3, u_5, u_6, u_7\}), (x_3, x_4), \{u_1, u_2, u_3\}), \\ &(x_5, x_1), \{u_3, u_4, u_5, u_8\}), (x_5, x_2), \{u_1, u_2, u_3, u_4, u_5, u_6, u_7, u_8\}), (x_5, x_4), \{u_1, u_2, u_3, u_8\}) \end{aligned} \right\}$$

Table 6. The two person soft game

S_k	y_1	y_2	...	y_n
x_1	$f_{s_k}(x_1, y_1)$	$f_{s_k}(x_1, y_2)$	...	$f_{s_k}(x_1, y_n)$
x_2	$f_{s_k}(x_2, y_1)$	$f_{s_k}(x_2, y_2)$	...	$f_{s_k}(x_2, y_n)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_n	$f_{s_k}(x_m, y_1)$	$f_{s_k}(x_m, y_2)$	...	$f_{s_k}(x_m, y_n)$

If Player 1 selects x_3 and player 2 select x_2 , then the value of the game will be a set $\{u_1, u_2, u_3, u_5, u_6, u_7\}$, that is,

$$fs_1(x_3, x_2) = \{u_1, u_2, u_3, u_5, u_6, u_7\}$$

In this case, player 1 wins the set of alternatives $\{u_1, u_2, u_3, u_5, u_6, u_7\}$ and player 2 lost the same set of alternatives. Suppose, if the player 2 constructs a tps-game as shown in Box 2.

Then the soft payoffs of the game can be arranged as in Table 8.

If Player 1 select x_3 and player 2 select x_2 , then the value of game will be a set $\{u_4, u_8\}$, That is

$$fs_2(x_3, x_2) = \{u_4, u_8\}$$

In this case, Player 1 wins the set of alternatives $\{u_4, u_8\}$ and player2 lost $\{u_4, u_8\}$.

N-Person Soft Games

In two person games, only two are participating. But in some games, more than two people are participating. This has led to the concept of n-person games. So, two person games can be extended to n person soft games.

Definition 10: Let U be a set of alternatives, $P(U)$ be the power set of U and X_k is the set of strategies of Player k , ($k=1,2,\dots,n$). Then, for each player, an n-person soft game (nps-game) is defined by a soft set over U as shown in Box 3, where fs_k^n is a soft payoff function of Player k .

Box 2.

$$s_1 = \left\{ ((x_1, x_1), \{u_3, u_4, u_6, u_7\}), (x_1, x_2), \{u_6, u_7\}), (x_1, x_4), \{u_1, u_2, u_4, u_5, u_6, u_7\}), \right. \\ \left. (x_3, x_1), \{u_2, u_4, u_5, u_6, u_8\}), (x_3, x_2), \{u_4, u_8\}), (x_3, x_4), \{u_4, u_5, u_6, u_7, u_8\}), \right. \\ \left. (x_5, x_1), \{u_1, u_2, u_6, u_7\}), (x_5, x_2), \{u_7\}), (x_5, x_4), \{u_4, u_5, u_6, u_7\}) \right\}$$

Box 3.

$$S_k^n = \{((x_1, x_1, \dots, x_n), fs_k^n(x_1, x_2, \dots, x_n)) : (x_1, x_2, \dots, x_n) \in X_1 \times X_2 \times \dots \times X_n\}$$

Table 7. Soft Payoff of the Game for strategy one

S_1	x_1	x_2	x_4
x_1	$\{u_1, u_2, u_3, u_8\}$	$\{u_1, u_2, u_3, u_4, u_5, u_8\}$	$\{u_3, u_8\}$
x_3	$\{u_1, u_3, u_7\}$	$\{u_1, u_2, u_3, u_5, u_6, u_7\}$	$\{u_1, u_2, u_3\}$
x_5	$\{u_3, u_4, u_5, u_8\}$	$\{u_1, u_2, u_3, u_4, u_5, u_6, u_8\}$	$\{u_1, u_2, u_3, u_8\}$

The nps-game is played as follows: at a certain time Player 1 chooses a strategy $x_1 \in X_1$ and simultaneously each Player k ($k=2, \dots, n$) chooses a strategy $x \in X_k$ and once this is done each player k receives the soft payoff $fs_k^n(x_1, x_2, \dots, x_n)$.

Definition 11: Let

$$S_k^n = \left\{ ((x_1, x_1, \dots, x_n), fs_k^n(x_1, x_2, \dots, x_n)) : (x_1, x_2, \dots, x_n) \in X_1 \times X_2 \times \dots \times X_n \right\}$$

be an nps-game. Then, a strategy $x \in X_k$ is called a soft dominated to another strategy $x \in X_k$, if

$$\begin{aligned} fs_k^n(x_1, \dots, x_{k-1}, x_k, x_{k+1}, \dots, x_n) &\supseteq \\ fs_k^n(x_1, \dots, x_{k-1}, x_k, x_{k+1}, \dots, x_n) &\end{aligned} \quad (10)$$

For each strategy $x_i \in X_i$ of player i ($i=1, 2, \dots, k-1, k+1, \dots, n$), respectively.

FUTURE RESEARCH DIRECTIONS

There may be some situations where we need to create objects having same attributes. This can be achieved by incorporating multiset theory with fuzzy soft set and game theory. So, a model can be developed with the capabilities of game theory, fuzzy set, soft set, multiset theory. In fact the research in the direction of application of soft sets in game theory has not been explored much. But as explained in this chapter it has great potential.

Table 8. Soft Payoff of the Game for strategy two

S_2	x_1	x_2	x_4
x_1	$\{u_3, u_4, u_6, u_7\}$	$\{u_6, u_7\}$	$\{u_1, u_2, u_4, u_5, u_6, u_7\}$
x_3	$\{u_2, u_4, u_5, u_6, u_8\}$	$\{u_4, u_8\}$	$\{u_4, u_5, u_6, u_7, u_8\}$
x_5	$\{u_1, u_2, u_6, u_7\}$	$\{u_7\}$	$\{u_4, u_5, u_6, u_7\}$

So, one can try to develop the concepts already in existence or introduce new directions.

CONCLUSION

Game theory deals with any competitive situation and the models are mostly of uncertain nature. Even under this, lack of sufficient number of parameters is a big hurdle which is efficiently handled by soft set theory. In this chapter our focus was to introduce both the topics in brief and mainly deal with applications of soft set theory in achieving better solutions in handling games. Soft set theory gives an opportunity to construct a new mathematical tool which keeps all good sides of choice function and eliminates its drawbacks. We proposed some directions for further research also.

REFERENCES

- Aumann, R. J. (1976). Agreeing to disagree. *Annals of Statistics*, 4(6), 1236–1239. doi:10.1214/aos/1176343654
- Bector, C. R., Chandra, S., & Vijay, V. (2004). Duality in linear programming with fuzzy parameters and matrix games with fuzzy pay-offs. *Fuzzy Sets and Systems*, 146(2), 253–269. doi:10.1016/S0165-0114(03)00260-4
- Deli, I., & Cagman, N. (2013). *Application of soft sets in decision making based on game theory*. arXiv: 1302.4568v4
- Maji, P. K., Biswas, R., & Roy, A. R. (2001). Fuzzy Soft Sets. *Journal of Fuzzy Mathematics*, 9(3), 589–602.

Maji, P. K., Roy, A. R., & Biswas, R. (2002). An application of soft sets in a decision making problem. *Computers & Mathematics with Applications (Oxford, England)*, 44(8-9), 1077–1083. doi:10.1016/S0898-1221(02)00216-X

Molodtsov, D. A. (1991). Optimality principles as a model of player behavior in game theory. Computer Center of Russian Academy of Sciences.

Molodtsov, D. A. (1999). Soft Set Theory First Results. *Computers & Mathematics with Applications (Oxford, England)*, 37(4-5), 19–31. doi:10.1016/S0898-1221(99)00056-5

Myerson, R. B. (1997). *Game Theory: Analysis of Conflict*. Cambridge, MA: Harvard University Press.

Neumann, J. V., & Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton University Press.

Osborne, M. J., & Rubinstein, A. (1994). *A Course in Game Theory*. Oxford University Press.

Roy, R., & Maji, P. K. (2007). A fuzzy soft set theoretic approach to decision making problem. *Journal of Computational and Applied Mathematics*, 203(2), 412–418. doi:10.1016/j.cam.2006.04.008

Zadeh, L. A. (1965). Fuzzy Sets. *Information and Control*, 8(3), 338–353. doi:10.1016/S0019-9958(65)90241-X

KEY TERMS AND DEFINITIONS

Choice Function: Choice function is a mapping which associates a set of strategies with a given situation.

Fuzzy Set: Fuzzy sets are an extension of the classical notion of sets whose elements have degrees of membership.

Game Theory: The study of mathematical models of conflict and cooperation between intelligent rational decision-makers.

Nash Equilibrium: It is a concept to find out the solution of a non cooperative game involving two or more players in game theory.

Payoff Function: Payoff function is used for modeling human behavior. Payoff function for a player is a mapping from the cross-product of players strategy spaces to the players set of payoffs.

S-Function (Soft Function): It is a payoff function defined using notion of soft set.

Soft Game: A game using soft payoff functions.

Soft Set: Soft set is a parameterized family of subsets defined over a universe associated with a set of parameters.

Category G

Gaming

Application of Gamification to Blended Learning in Higher Education

Kamini Jaipal-Jamani

Brock University, Canada

Candace Figg

Brock University, Canada

INTRODUCTION

Digital games have been identified as a teaching strategy that goes beyond conventional notions of pedagogy to create learning environments that feel “new, meaningful, and 21st century” (Johnson, Adams, Becker, Estrada, & Freeman, 2015, p. 34). The motivational aspects of games make them highly applicable in higher education to foster student engagement in learning (Connolly, E.A. Boyle, MacArthur, Hainey, & J.M. Boyle, 2012). However, a recent review shows that digital games are not being utilised to their potential in educational settings (Boyle et al., 2016). Research shows that the inclusion of digital game-based learning (DGBL) is often challenging for educators to implement in formal education learning environments (Faria & Wellington, 2004; Ritzhaupt, Gunter, & Jones, 2010; Van Eck, 2015). This chapter proposes the use of gamification in a blended learning environment as a low-risk strategy to promote more widespread use of DGBL in higher education courses. While gamification is an emerging strategy with many definitions associated with it, there is consensus that it involves “the integration of game elements, mechanics, and frameworks into non-game situations and scenarios” (Johnson, Adams, Becker, Estrada, & Freeman, 2013, p. 20). In this chapter, the relevant literature on blended learning, gaming, and gamification are reviewed, and an empirical case is presented, to illustrate the application of gamification to blended learning in higher educa-

tion. The design and implementation of a gamified blended course are discussed and future research directions are proposed.

BACKGROUND

Blended Learning

A current trend in higher education is the implementation of blended learning (Halverson, Graham, Spring, Drysdale, & Henrie, 2014; Johnson et al., 2015). According to Graham (2013, p. 334), blended learning is commonly used to “denote a combination of traditional face-to-face and online instruction.” Blended learning evolved from the research into effective teaching practices in higher education which indicated that teacher-directed, lecture-style instruction prominent in higher education learning environments often resulted in “surface-learning” of content (Marton & Saljo, 1976). Deep and meaningful learning promotes metacognitive and higher order thinking skills which are supported by learning strategies such as collaborative tasks, engagement in discourse, reflection, and self-regulated learning (Garrison & Vaughan, 2008). Nonetheless, it has been found that some learning strategies tend to be more effective in one learning environment compared to the other. For example, collaborative tasks in face-to-face learning environments provide emotional support through opportunities “to create a sense of community and connectedness more quickly,”

which is lacking in the online environment (Garrison & Vaughan, 2008, p. 28). On the other hand, reflection and discourse through self-regulation of learning is not as strongly supported by face-to-face learning interactions. It is the asynchronous nature of online learning activities that encourages collaborative discourse, reflection, and self-regulation of learning outside of class (Vaughan, Cleveland-Innis, & Garrison, 2013). Blended learning environments are therefore beneficial for mitigating emotional distance (Vaughan et al., 2013) and developing metacognition and higher order learning skills (e.g., S. K. Taradi, M. Taradi, Radic, & Pokrajac, 2005; Tsai, 2014).

Various frameworks have been proposed for designing blended learning. For example, Wu, Tennyson, and Hsia (2010) found that a combination of the following factors: cognitive (e.g., computer self-efficacy), technological (e.g., flexible access to online activities) and social interactions, contributed to learner satisfaction. Another framework that has been most prevalent for informing blended learning design is Garrison and Vaughan's (2008) community of inquiry (Halverson et al., 2014) where social and cognitive presence is addressed in the face-to-face and online environment. According to this framework, social presence is best promoted through learning activities (e.g., wikis, blogs) that engage students in small group discussions, collaborative research, and cooperative group work (Vaughan, Cleveland-Innis, & Garrison, 2013). Cognitive presence in the online environment is developed through inquiry with the community of learners, using ongoing reflection and critical discourse. Many online learning activities, such as WebQuests, videos, simulations, and games provide learners with opportunities to gain knowledge and participate in a community of online inquiry (Figg & Jaipal-Jamani, 2015).

Digital Games

Generally, digital games are identified by a set of common characteristics that include explicit rule systems and players, visual and auditory com-

ponents, action and conflict, players competing against each other, or the game system, and the attainment of an outcome or a clear and quantifiable goal (Gee, 2013; Salen & Zimmerman, 2004). Additionally, it is the interactions among the technology, player thinking, and social interaction that create the gaming experience for players (Shaffer, Squire, Halverson, & Gee, 2005). Three corresponding constructs, rules, play, and culture, as described by Salen and Zimmerman (2004), provide educators with a simple framework to understand how games are designed to promote the gaming experience. Rules refer to the features used to organise the game such as the structure, elements, and mechanics. These design features determine player input and output, what choices players make, and what the outcomes or rewards of the game are. Play describes players' immersion experiences with the game and includes motivation to play and the tactics and strategies used while playing. Culture encompasses the larger context of the game which could be fictional or nonfictional and includes values, narrative, and community building. It is the use of a combination of these game features (rules, play, and culture) that capture the player's attention and enable the player to experience emotional well-being (e.g., happiness, pleasure) and satisfaction of psychological needs such as autonomy, competence, and relatedness (Ryan, Rigby, & Przybylski, 2006; Weber & Shaw, 2009).

Digital Games in Educational Settings

The use of digital games in educational settings has gained momentum because of the many benefits for the workplace and learning attributed to playing games. For example, many of the workplace skills employees are expected to have are similar to the skills that lead to success in gaming, such as strategic thinking, problem solving, and adapting to change (Federation of American Scientists, 2006). Many characteristics of the gaming environment are also purported to support development

of 21st century learning skills or competencies (Granic, Lobel, & Engels, 2014), such as critical thinking and problem solving, collaboration and teamwork, and computer literacy (Trilling & Fadel, 2009). The motivational benefits of digital games, through entertaining and capturing the attention of the player for long periods of time, are well documented (Deterding, Dixon, Khaled, & Nacke, 2011; Gee, 2003). Additionally, digital games have contributed to cognitive, skill, and attitude development (e.g., Egenfeldt-Nielsen, 2007; Vogel et al., 2006). Research also points to the applicability of games in educational settings because the game experience is consistent with experiential views of how students learn (Gee, 2013). As Gee (2013, p. 18) succinctly points out, “games are just well designed experiences in problem solving” where factual knowledge is used as a tool to solve a problem. When digital games are used for educational purposes, characterised by some type of learning that occurs by interaction of the player with the mechanics of the game, this learning experience is called digital game-based learning, or DGBL (Van Eck, 2015). The most prevalent forms of DGBL in education are the use of commercial off-the-shelf games (COTS) which are easily accessible or serious games, ones designed to train or teach specific skills and content. The application of DGBL in higher education settings is not widespread (Boyle et al., 2016) and this situation may be attributed to pedagogical and implementation issues and challenges experienced by instructors.

Issues and Challenges Using DGBL

A pedagogical issue that crops up when using COTS and serious games in educational settings is that commercial games may not meet a wide range of course learning outcomes. For example, the scope of the content covered by a video game may not adequately reflect the content as described in the course curriculum or provide sufficient

scaffolding for learning complex content (Faria & Wellington, 2004; Lean, Moizer, Towler, & Abbey, 2006). This limitation has led to instructors making teaching modifications to teaching strategies and assessment tools. Therefore, serious video games or COTS are often not used as primary instructional activities and are perceived as “most useful as a supplemental resource or a consolidation tool for student learning of [the content]” (Jaipal & Figg, 2009, p. 5). Another pedagogical issue is the type of learning promoted by DGBLs. Instructors wanting to promote reflection and higher order thinking processes in their subject areas are challenged to find suitable games promoting such learning goals. Many serious games primarily support the learning and reinforcement of factual information or acquisition of knowledge (Kiili, 2005) and professional training (e.g., medicine and business; Boyle, et al., 2016) through drill and practice, quizzes and tests, and training simulations.

Factors that contribute to implementation challenges have also been identified in the literature. Technical competence, especially lack of expertise or prior experiences playing games, influences instructor confidence in their ability to incorporate digital games in instruction (Faria & Wellington, 2004, Ritzhaupt et al., 2010; Van Eck, 2015). Instructor perceptions of games as an effective learning strategy, also affect their adoption of games. Research shows that perceptions of games are “based upon [instructor] professional judgement of benefit and risk, rather than on the resources available” (Lean et al., 2006, p. 239). In light of these barriers to faculty adoption of DGBLs, gamification offers instructors a way to take advantage of the motivational and cognitive affordances of games, in spite of modest technical competence levels. Gamification also enables the tailoring of the learning activity to match specific student needs and learning goals with fewer changes to existing instructional practices (Figg & Jaipal-Jamani, 2015).

GAMIFICATION IN HIGHER EDUCATIONAL SETTINGS

Gamification

Gamification is a recent trend in many public sectors, including education, because it incorporates aspects of games that can motivate, increase participation and engagement, and has the potential to improve learning (Dicheva, Dichev, Agre, & Angelova, 2015; Seaborn & Fels, 2015). Admittedly, there is controversy around the use of the term gamification. Critics argue that gamification as described in the literature, with its emphasis on the use of points, badges and leaderboards for task completion and achievement, is really pointsification (Robertson, 2010, para 11). While gamification does include the strategy of pointsification, there is consensus in the education literature that gamification involves the application of a range of game design elements in non-game contexts (Deterding et al., 2011). Seaborn and Fels (2015) further characterise gamification as: 1) being designed for non-entertainment purposes and 2) using elements of games, without creating a “fully-fledged game. In this way, gamified systems are game-like, but not a game” (p. 27). Studies on gamification in educational contexts do incorporate game elements beyond the points system (Dicheva et al., 2015). Besides the motivational game elements identified such as choice and rewards, Kapp (2012) and Stott and Neustaedter (2013) identify other dynamic gaming elements such as “providing learners with permission to fail, encouragement of out-of-box thinking, and fostering a sense of control” (Kapp, 2012, p. xxii) that contribute to rich learning experiences. It should be noted, however, that simplifying the complexity of the game application to a few game mechanics could result in less engagement of students and lead to an alienation of interest (Lee & Hammer, 2011; Stott & Neustaedter, 2013). As well, the use of game elements “is not practical for all learners, all content, all the time—any more than are lectures or textbooks” (Van Eck,

2015, p. 24) and will not meet all learning goals. Therefore, as Lee and Hammer (2011) point out, “we must carefully design gamification projects that address real challenges” (p.4).

Design Framework for Gamification

Kapp (2012) suggests that gamified learning environments are effective when they include these three components: 1) game-based mechanics, 2) aesthetics, and 3) game thinking.

- Game-Based Mechanics:** The game-based mechanics most used in educational settings are the use of badges, point systems, leaderboards, goals, and time limits (Dicheva et al., 2011; Kapp, 2012; Deterding et al., 2011). Applying these game mechanics to non-gaming situations provide the engagement and motivational aspects found in games such as freedom to fail, rapid feedback, progression, and story-telling (Stott & Neustaedter, 2013). For example, incorporating tasks, quests, or challenges that can be repeated with no penalty for failure supports freedom to fail and mastery of content. Experience points provide immediate feedback to players and rewards with badges and levels provide recognition of achievements and a record of progress. Competition through leaderboards acknowledges and recognizes achievements while providing a comparison of performance with others.
- Game Aesthetics:** Game aesthetics refers “to the way the game mechanics and dynamics interact with the game designer’s artistry, to produce cultural and emotional outcomes” (Dicheva et al., 2011, p. 3). It includes the visual and aural presentation of the gamified activity such as the presentation of tasks through a cohesive interface theme, the design of an attractive layout, the use of appropriate graphics, and ease of navigation through the material. These

aesthetic features contribute to feelings of satisfaction and how willingly a player accepts the gamification (Kapp, 2012).

- **Game Thinking:** Game thinking is described as, “thinking about an everyday experience and converting it into an activity that has elements of competition, cooperation, exploration, and storytelling” (Kapp, 2012, p.11). It is a mind-set where the element of fun or challenge is used deliberately to create an experience that motivates learners to engage in desired behaviors or learning goals (Werbach & Hunter, 2012) which are clearly communicated Figg & Jaipal-Jamani, 2015; Robson, Plangger, Kietzmann, McCarthy, & Pitt, 2015). Tasks should provide valuable information or skills, tap into personal interests and passions, provide an element of free exploration and play, result in mastery of skills, and vary in complexity and depth (Deterding, 2011; Van Eck, 2015).

There is consensus in the field that gamification holds potential to motivate and engage learners (Domínguez et al., 2013; Van Eck, 2015; Robson et al., 2015). Adding gamification to blended higher education courses through scaffolded, online learning activities is, however, a new practice which may be complex, and require some effort to make it motivating for students (de-Marcos, Domínguez, Saenz-de-Navarrete, & Pagés, 2014). In the next section, we describe an example of a successful gamified, blended learning environment in higher education.

Blended Gamification in Action: The *TPACK Quest*

The strategy of gamification was used to redesign an instructional technology methods course in a Faculty of Education. The purpose for redesigning the course was to address the problem of lack of motivation among preservice teachers (student teachers) to engage with background information (e.g., textbook readings) assigned to enhance in-

class learning experiences and to deepen their knowledge about teaching with technology, or technological pedagogical and content knowledge (TPACK; Mishra & Koehler, 2006). Drawing on aspects of the design frameworks of Wu et al. (2010) and Vaughan et al. (2013) for elements of blended learning, and Kapp (2012) for gamification components, a gamified, online learning environment, called the *TPACK Quest*, was created using Wordpress, a free blogging tool. To ensure a quality blended learning environment where online learning activities enhanced in-class activities, gamified, online activities addressed cognitive, technological, and social interactions respectively (Wu et al., 2010; Vaughan et al., 2013). For example, cognitive skills were developed by having preservice teachers learn about the characteristics and actions for effective teaching with technology through readings and interactive media; technological knowledge was developed through digital application tasks where new technical tools were used in authentic learning experiences, and social interactions were promoted through collaborative online discussions.

The design of the *Quest* was also informed by the three design elements discussed in the previous section (Kapp, 2012). *Game mechanics* included a pointsification scheme of ten badges and a *TPACK Quest* manifesto which guided game play. Grade points were awarded for each task, and badges were earned for completing a mission or series of tasks. The *Quest* badge missions included topics such as gamification, TPACK, digital citizenship, problem-based learning, and Web 2.0 learning objects. Acquiring ten of these badges resulted in earning the *Cup of Wisdom*. Four additional badges were awarded for a digital portfolio, professional conduct, completion of in-class activities, and attendance and punctuality. Completion of the latter tasks resulted in the *Professional Order of the TPACK Teacher* award. The *Quest* also incorporated a limited narrative introduced by an avatar in a video to provide a purpose and overall context for the activities. *Game aesthetics*, such as the colorful interface of the *Quest*, the consistent navigational elements, and a variety of

interactive learning objects throughout the online environment, promoted ease of participation. *Game thinking* was employed in three ways to design the tasks. First, tasks incorporated a variety of activities, with varying levels of difficulty and complexity. Secondly, tasks incorporated an element of challenge, or problem solving, around pedagogical thinking. Lastly, the element of fun was incorporated by challenging ‘players’ to use Web 2.0 tools in ways teachers used the tools within classroom instruction. For example, to earn the *Digital Teacher* badge, ‘players’ examined videos about flipped and blended learning outside of class, then used Cube4Teachers (<http://www.cubeforteachers.com/>), an online repository of teacher-selected resources, to select an activity. The task was to write an introductory paragraph introducing the activity, and insert that text into the newspaper headline generator at Fodey.com. The resulting image (newspaper clipping with their inserted text) was submitted to earn their badge. In class, the resources gathered through the experience were evaluated and top resources identified for the collaborative resource board, resulted in points towards the *Collaborator* badge.

The Research Study

A design-based research study (Anderson & Shattuck, 2012) was conducted over two years to investigate how gamification influenced teacher candidates’ TPACK, or understanding of how to teach with technology (cognitive), collaborate and use technology (skill), and professionalism (attitude). Teacher candidates were enrolled in six different sections of the technology methods course in which the *Quest* was implemented. Data were collected from 133 teacher candidates and data sources were online surveys, written reflections, and artifacts in digital portfolios. Data sources also included email interviews with the five different instructors teaching the technology methods course, researcher field notes from weekly instructor meetings, and instructors’ records. Instructors met weekly to discuss any issues or concerns with the course. Any changes

to the *Quest*, or the process for playing the *Quest*, were implemented into all sections. All data were analysed and coded by two independent raters with 100 percent unity of agreement.

The findings from online surveys and written reflections revealed that the gamified learning environment motivated teacher candidates to complete tasks (Figg & Jaipal-Jamani, 2015) and is consistent with other findings on gamification (Lee & Hammer, 2011; Hamari, Koivisto, & Sarsa, 2014). All written artifacts by preservice teachers were subjected to content analysis for evidence of TPACK learning gains. Results showed that the gamified blended learning environment stimulated cognitive development (progress of learning) through the visual tracking of learning, promoted self-regulated learning (at their own pace), and facilitated risk taking with technology (Figg & Jaipal-Jamani, 2015). Results were corroborated by survey data. For example, in response to the statement, “I feel I gained more knowledge about TPACK from playing the TPACK Quest than I would have from reading a textbook on the topic,” more than two thirds of students agreed or strongly agreed with the statement.

In this case, a blended learning environment enabled the instructors to teach existing curriculum using face-to-face pedagogical practices while incorporating the benefits of an online gamified learning environment. However, there is no doubt that instructors need technical skills to be able to create the digital interfaces and learning objects that comprise the learning environment. Instructors also need to have pedagogical expertise to be able to design a series of game tasks that are varied in depth, complexity, and difficulty while scaffolding the learning process (Figg & Jaipal-Jamani, 2015; Deterding, 2011).

FUTURE RESEARCH DIRECTIONS

Studies of gamification have focused on investigating how gamification influences motivation, engagement, and enjoyment of students during learning (Hamari et al., 2014). The case reported

in this chapter is one of the few examples in higher education that provides empirical evidence of the effectiveness of gamification to achieve learning goals related to a course with a large sample of students. There is a need for additional empirical studies to examine gamification and learning outcomes in different higher education disciplinary contexts. As well, studies show that the appropriateness of the context being gamified and the characteristics of students play a role in how effective the gamification strategy is (Hamari et al., 2014). As such, a future area of research could be to examine how context and characteristics of students influence the blended gamified learning environment. Another area for future research is investigating how instructor knowledge base to design blended gamified learning environments develops. This is important because instructors need knowledge in two areas, creating blended learning environments and using game design elements within instruction.

CONCLUSION

The barriers encountered by instructors when trying to implement digital games such as video games and simulations, often hinder the widespread use of games as an instructional strategy. This chapter presented current perspectives on how the affordances of games, such as choice, feedback, autonomy, and freedom to fail, can be incorporated into a gamified blended learning environment in higher education. Gamification is another digital game-based strategy that utilizes game mechanics, game aesthetics, and game thinking to achieve the motivational and engagement benefits of digital games within the constraints of curricular content that has to be taught. Furthermore, incorporating gamification into blended learning environments provides instructors with opportunities to use traditional face-to-face instruction thereby scaffolding faculty adoption of unfamiliar instructional strategies. A gamification strategy provides many of the benefits derived from games, without requiring instructors

to have the knowledge and experience to create or play games. Hence, gamification in blended learning environments is a promising instructional strategy that draws motivational and engagement aspects from game-based learning, and provides learners with opportunities for cognitive, skill, and attitude development.

REFERENCES

- Anderson, T., & Shattuck, J. (2012). Design-based research a decade of progress in education research? *Educational Researcher*, 41(1), 16–25. doi:10.3102/0013189X11428813
- Boyle, E. A., Hainey, T., Connolly, T. M., Gray, G., Earp, J., Ott, M., & Pereira, J. et al. (2016). An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games. *Computers & Education*, 94, 178–192. doi:10.1016/j.compedu.2015.11.003
- Connolly, T. M., Boyle, E. A., MacArthur, E., Hainey, T., & Boyle, J. M. (2012). A systematic literature review of empirical evidence on computer games and serious games. *Computers & Education*, 59(2), 661–686. doi:10.1016/j.compedu.2012.03.004
- de-Marcos, L., Domínguez, A., Saenz-de-Navarrete, J., & Pagés, C. (2014). An empirical study comparing gamification and social networking on e-learning. *Computers & Education*, 75, 82–91. doi:10.1016/j.compedu.2014.01.012
- Deterding, S. (2011, January 24). *Meaningful play: Getting gamification right* [Video File]. Retrieved from <https://youtu.be/7ZGCPap7GkY>
- Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining gamification. In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments* (pp. 9–15). New York: ACM. doi:10.1145/2181037.2181040

- Dicheva, D., Dichev, C., Agre, G., & Angelova, G. (2015). Gamification in education: A systematic mapping study. *Journal of Educational Technology & Society*, 18(3), 1–14.
- Domínguez, A., Saenz-De-Navarrete, J., De-Marcos, L., Fernández-Sanz, L., Pagés, C., & Martínez-Herráiz, J. J. (2013). Gamifying learning experiences: Practical implications and outcomes. *Computers & Education*, 63, 380–392. doi:10.1016/j.compedu.2012.12.020
- Egenfeldt-Nielsen, S. (2007). Third generation educational use of computer games. *Journal of Educational Multimedia and Hypermedia*, 16(3), 263.
- Faria, A. J., & Wellington, W. J. (2004). A survey of simulation game users, former-users, and never-users. *Simulation & Gaming*, 35(2), 178–207. doi:10.1177/1046878104263543
- Federation of American Scientists. (2006). *Summit on educational games: Harnessing the power of video games for learning*. Retrieved from http://fas.org/programs/itp/policy_and_publications/summit/Summit on Educational Games.pdf
- Garrison, D. R., & Vaughan, N. D. (2008). *Blended learning in higher education: Framework, principles, and guidelines*. San Francisco, CA: John Wiley & Sons, Inc.
- Gee, J. P. (2003). What video games have to teach us about learning and literacy. *Computers in Entertainment*, 1(1), 1–4.
- Gee, J. P. (2013). Games for learning. *Educational Horizons*, 91(4), 16–20.
- Graham, C. R. (2013). Emerging practice and research in blended learning. In M. G. Moore (Ed.), *Handbook of distance education* (3rd ed.; pp. 333–350). New York, NY: Routledge. doi:10.4324/9780203803738.ch21
- Granic, I., Lobel, A., & Engels, R. C. (2014). The benefits of playing video games. *The American Psychologist*, 69(1), 66–78. doi:10.1037/a0034857 PMID:24295515
- Halverson, L. R., Graham, C. R., Spring, K. J., Drysdale, J. S., & Henrie, C. R. (2014). A thematic analysis of the most highly cited scholarship in the first decade of blended learning research. *The Internet and Higher Education*, 20, 20–34. doi:10.1016/j.iheduc.2013.09.004
- Hamari, J., Koivisto, J., & Sarsa, H. (2014, January). Does gamification work?—a literature review of empirical studies on gamification. In *2014 47th Hawaii International Conference on System Sciences* (pp. 3025–3034). IEEE.
- Johnson, L., Adams Becker, S., Estrada, V., & Freeman, A. (2013). *NMC Horizon Report: 2015 Higher Education Edition*. Austin, TX: The New Media Consortium.
- Johnson, L., Adams Becker, S., Estrada, V., & Freeman, A. (2015). *NMC Horizon Report: 2015 Higher Education Edition*. Austin, TX: The New Media Consortium.
- Kapp, K. M. (2012). *The gamification of learning and instruction: game-based methods and strategies for training and education*. San Francisco, CA: John Wiley & Sons.
- Kiili, K. (2005). Digital game-based learning: Towards an experiential gaming model. *The Internet and Higher Education*, 8(1), 13–24. doi:10.1016/j.iheduc.2004.12.001
- Lean, J., Moizer, J., Towler, M., & Abbey, C. (2006). Simulations and games use and barriers in higher education. *Active Learning in Higher Education*, 7(3), 227–242. doi:10.1177/1469787406069056
- Lee, J. J., & Hammer, J. (2011). Gamification in education: What, how, why bother? *Academic Exchange Quarterly*, 15(2), 1–5.

- Marton, F., & Saljo, R. (1976). On qualitative differences in learning: I Outcome and process. *The British Journal of Educational Psychology*, 46(1), 4–11. doi:10.1111/j.2044-8279.1976.tb02980.x
- Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, 108(6), 1017–1054. doi:10.1111/j.1467-9620.2006.00684.x
- Ritzhaupt, A. D., Gunter, E., & Jones, J. G. (2010). Survey of commercial off-the-shelf video games: Benefits and barriers in formal educational settings. *International Journal of Instructional Technology and Distance Learning*, 7(5), 45–55.
- Robertson, M. (2012). *Can't play, won't play. Hide and Seek: Inventing new kinds of play*. Retrieved from <http://hideandseek.net/2010/10/06/cant-play-wont-play/>
- Robson, K., Plangger, K., Kietzmann, J. H., McCarthy, I., & Pitt, L. (2015). Is it all a game? Understanding the principles of gamification. *Business Horizons*, 58(4), 411–420. doi:10.1016/j.bushor.2015.03.006
- Ryan, R. M., Rigby, C. S., & Przybylski, A. (2006). The motivational pull of video games: A self-determination theory approach. *Motivation and Emotion*, 30(4), 344–360. doi:10.1007/s11031-006-9051-8
- Salen, K., & Zimmerman, E. (2004). *Rules of play: Game design fundamentals*. Cambridge, MA: MIT Press.
- Seaborn, K., & Fels, D. I. (2015). Gamification in theory and action: A survey. *International Journal of Human-Computer Studies*, 74, 14–31. doi:10.1016/j.ijhcs.2014.09.006
- Shaffer, D. W., Squire, K. R., Havelson, R., & Gee, J. P. (2005). Video Games and the Future of Learning. Wisconsin Center for Education Research, School of Education, University of Wisconsin–Madison. Retrieved from <http://www.wcer.wisc.edu/>
- Stott, A., & Neustaedter, C. (2013). *Analysis of Gamification in Education*. Technical Report 2013-0422-01. Connections Lab, Simon Fraser University.
- Taradi, S. K., Taradi, M., Radic, K., & Pokrajac, N. (2005). Blending problem-based learning with Web technology positively impacts student learning outcomes in acid–base physiology. *Advances in Physiology Education*, 29, 35–39.
- Trilling, B., & Fadel, C. (2009). *21st Century Skills: Learning for life in our times*. San Francisco, CA: Jossey-Bass.
- Tsai, C. W. (2014). A quasi-experimental study of a blended course integrated with refined web-mediated pedagogy of collaborative learning and self-regulated learning. *Interactive Learning Environments*, 22(6), 737–751. doi:10.1080/10494820.2012.745422
- Van Eck, R. (2015, November/December). Digital Game-Based Learning: Still restless, after all these years. *EDUCAUSE Review*, 13–28. Retrieved from <http://er.educause.edu/articles/2015/10/digital-game-based-learning-still-restless-after-all-these-years>
- Vaughan, N. D., Cleveland-Innes, M., & Garrison, D. R. (2013). *Teaching in blended learning environments: Creating and sustaining communities of inquiry*. Edmonton, Canada: AU Press, Athabasca University.
- Vogel, J. J., Vogel, D. S., Cannon-Bowers, J., Bowers, C. A., Muse, K., & Wright, M. (2006). Computer gaming and interactive simulations for learning: A meta-analysis. *Journal of Educational Computing Research*, 34(3), 229–243. doi:10.2190/FLHV-K4WA-WPVQ-H0YM
- Weber, R., & Shaw, P. (2009). Player types and quality perceptions: A social cognitive theory based model to predict video game playing. *International Journal of Gaming and Computer-Mediated Simulations*, 1(1), 66–89. doi:10.4018/jgcms.2009010105

Werbach, K., & Hunter, D. (2012). *For the win: How game thinking can revolutionize your business*. Philadelphia, PA: Wharton Digital Press.

Wu, J. H., Tennyson, R. D., & Hsia, T. L. (2010). A study of student satisfaction in a blended e-learning system environment. *Computers & Education*, 55(1), 155–164. doi:10.1016/j.compedu.2009.12.012

Figg, C., & Jaipal-Jamani, K. (2015). Using gamification strategies to engage teacher candidates as they develop TPACK knowledge. M. Niess & H. Willows-Giles (Eds.), *Handbook of Research on Teacher Education in the Digital Age* (pp.663-690). Hershey, PA: IGI Global Publications.

Jaipal, K., & Figg, C. (2009). Using video games in science instruction: Pedagogical, social, and concept-related aspects. *Canadian Journal of Science. Mathematics and Technology Education*, 9(2), 117–134.

ADDITIONAL READINGS

Barata, G., Gama, S., Jorge, J., & Gonçalves, D. (2013, October). Improving participation and learning with gamification. In *Proceedings of the First International Conference on gameful design, research, and applications* (pp. 10-17). ACM.

Cheong, C., Filippou, J., & Cheong, F. (2013). Understanding student perceptions of game elements to develop gamified systems for learning. In *Proceedings of the Pacific Asia Conference on Information Systems 2013* (Paper 202). Retrieved from <http://aisel.aisnet.org/pacis2013/202>

Koster, R. (2013). *Theory of fun for game design*. Sebastopol, CA: O'Reilly Media, Inc.

Picciano, A. G. (2006). Blended learning: Implications for growth and access. *Journal of Asynchronous Learning Networks*, 10(3), 95–102.

KEY TERMS AND DEFINITIONS

Blended Learning: An instructional practice that involves using both online and traditional face-to face learning experiences where a part of the face-to-face instruction is replaced by online activities completed outside the classroom.

Digital Games: An interactive system on a digital device that is rule-based, played by one or more players, with variable or quantifiable outcomes that are ranked.

Digital Game-Based Learning (DGBL): Digital games created for learning purposes characterized by learning that occurs by interaction of the player with the mechanics of the game.

Gamification: One form of DGBL that uses game-based elements such as mechanics, aesthetics, and game thinking in non-game contexts aimed at engaging people, motivating action, enhancing learning, and solving problems (Borgos et al, 2014).

Chemistry Learning Through Designing Digital Games

Kamisah Osman

The National University of Malaysia, Malaysia

Ah-Nam Lay

Institute of Teacher Education – Sultan Abdul Halim, Malaysia

INTRODUCTION

As the labour market of the 21st century become increasingly emphasize on skilled human capital which are highly knowledgeable and innovative, STEM (Science, Technology, Engineering and Mathematics) education has slowly been making its way into classrooms. Efforts have been undertaken to improve the quality of STEM education in order to produce STEM literate students, i.e. students who are capable of identifying, applying, and integrating the STEM concept to understand complex problems and generate innovation to solve the problems (Chew, Noraini, Leong & Mohd Fadzil, 2013).

STEM literate students must have mastered the knowledge of science, particularly chemistry because chemistry is essential as the basic of most of the fields of science, technology and engineering (Balaban & Klein, 2006). Indeed, chemistry is often called the central science (Brown, LeMay, Bursten, Murphy & Woodward, 2011; Chang, 2007). Chemistry is also the foundation for innovation, scientific literacy and most notably problem solving in connection with sustainable development (Risch, 2010). Using chemistry knowledge, for instance, one can make judicious decisions to design new materials that are environmentally safe to solve various problems in everyday life. In the 21st century, chemistry will continue to play a leading role in providing solutions to ensure sustainable development.

Apart from knowledge, innovation and problem solving in the 21st century requires a new range of skills known as 21st century skills. For example, innovation and problem solving in today's world is driven by the formation of networks with multiple parties including experts and researchers with related interests as well as consumers and customers. The 21st century skills enable one to communicate and collaborate effectively with various parties.

Clearly, students in the 21st century need to become proficient in both chemistry knowledge and the 21st century skills. Therefore, the authors have initiated an innovation instructional strategy to support the acquisition of chemical concepts and the 21st century skills. A module known as MyKimDG has been developed as a mechanism for accomplishing the desired goals. The purpose of this article is to present conceptual framework of MyKimDG and demonstrate a brief lesson in MyKimDG to the teaching and learning of a specific chemistry unit.

BACKGROUND

Chemistry is usually perceived as a difficult and unpopular subject due to the abstract nature of chemical concepts. Previous studies on students' conception have revealed that students have many alternatives conception in chemistry. While the literature is replete with studies and papers, which

investigate students' understanding of chemical concepts and suggest potential remedies, fewer studies focus simultaneously on improving conceptual understanding and developing the 21st century skills. Hence, educators should be encouraged to design innovative and effective learning strategies to enhance both students' conceptual understanding and their 21st century skills. In this case, a change in chemistry teaching and learning approaches is critical. This is especially more crucial when educating today's students who are 'digital natives' (Prensky, 2001). The teaching and learning approaches must befit the needs of these digital natives and subsequently achieve the desired aspiration.

One approach suggested by researchers to educate the digital native generation is the integration of digital games in the teaching and learning processes as digital game is a medium favoured by students. Nowadays, the integration of digital games in learning or digital game-based learning (DGBL) is gaining popularity parallel with their popular reputation among students (Kamisah & Nurul Aini, 2013). Many studies have reported that DGBL can provide positive impact on students' learning. In general, the studies on DGBL were carried out through two approaches, namely (1) student as game consumer or player, and (2) student as game designer.

In the first approach, the students were involved in playing commercial digital games in the market or educational digital games developed by educators. However, there are many obstacles to implementing the student as game consumer approach. For instance, the contents of commercial digital games are inaccurate or incomplete (Van Eck 2006) and the development of professional educational digital games is time-consuming (Hwang et al. 2013). In addition, many gamers do not play educational digital games as they find these games not compelling (Pivec 2009). This happens because educational digital games are designed by academics who do not really understand the art, science and culture of digital game design (Van Eck 2006). As a result, the product has failed

dismally as a game. Prensky (2008) also raised this issue and states '...the students had no input into its creation, and the stuff came out cute to the adults, but boring to the kids'. According to Prensky (2008), students even told straight forwardly: 'Don't try to use our technology, you'll only look stupid.'

One alternative of DGBL approach that has been proposed by some scholars (such as Kafai, 1996; Papert, 1998; Jung & Park, 2009; Kamisah & Nurul Aini, 2013) is for students to design their own digital games. Many studies have reported that this approach provide opportunities for students to explore ideas according to their own interests (Kafai & Ching, 1996); acquire knowledge of programming (Kafai, Ching & Marshall, 1997); as well as become active participants and problem solvers, engage in social interaction by sharing their designs and helping each other, and take ownership of their own learning (Baytak & Land 2010).

In addition, Vos, van der Meijden and Denessen (2011) has reported that the student as game designer approach is a better way to increase student motivation and deep learning compared to the student as game consumer approach. Lim (2008) and Prensky (2008) also recognized the potential of this approach in improving student motivation and engagement. According to Lim (2008), when students are given the autonomy to take responsibility for their own learning and co-design learning experiences with teachers and other students, they are more inclined to engage in their own learning process. Hence, one way to do so is by allowing students to become designers of their own digital games based on their own interpretation of the school curriculum. For Prensky (2008), students will be motivated if they are allowed to do something extraordinary in learning at school and gain recognition for producing digital games. Therefore, the authors have initiated an innovation to take advantage of the student as game designer approach to support the acquisition of chemical concepts and the 21st century skills.

CONCEPTUAL FRAMEWORK OF MYKIMDG

Principles derived from learning theories play an important role in guiding MyKimDG development. Two important theories in learning and education that incorporate into MyKimDG development are constructivism and constructionism.

According to constructivist theory of learning, learner is knowledge builder. Learner does not receive knowledge passively, but he/she interprets the knowledge received and then modifies the knowledge in a form that is acceptable to him/her. In other words, individual learner actively constructs new knowledge pursuant to his/her existing knowledge. In addition, the process of knowledge construction can be improved through social interaction and discovery. Vygotsky (1978) believed that interaction between learner and teacher or more skillful peers will provide scaffolding to learner within the Zone of Proximal Development to construct new knowledge. However, no interaction would be beneficial if the new information is presented to students traditionally. Instead, students should be given the opportunity to explore or discover the new knowledge. Bruner (1966) believed that learning and problem solving emerged out of exploration of new knowledge.

In addition to the constructivist theory, constructionist theory of learning asserts that the construction of new knowledge happens felicitously in a context where students are consciously involved in the production of external and sharable artefacts (Papert 1991). This theory goes beyond the idea of learning-by-doing as indicated by Papert (1999a) that 'I have adapted the word constructionism to refer to everything that has to do with making things and especially to do with learning by making, an idea that includes but goes far beyond the idea of learning by doing.' Indeed, Papertian constructionism challenges the learner applying the knowledge being explored to construct more complex ideas or larger theory. This theory emphasizes the role of design (making,

building or programming) (Kafai & Resnick, 1996) and external objects (Egenfeldt-Nielsen, 2006) in facilitating the knowledge construction. The constructionist theory of learning also values the diversity of learners and social aspects of learning (Kafai & Resnick, 1996). In the constructionist learning environment, the designers or learners create artefacts which are significant to themselves based on their interests, learning styles and their experience, and share their artefacts as well as the artefacts' designing process with peers. In this process, peers act as collaborators, coaches, audiences and co-constructors of knowledge.

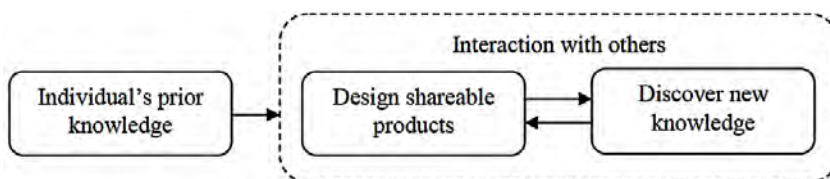
Computers play a role in the constructionist learning theory. Computers can be used as a building material (Papert, 1999a). According to Papert and Franz (1988), a computer is a 'material to be messed about with'. Learning occurs when learners are 'messaging about' with the computer. The introduction of computers is also able to change the context of learning (Papert, 1991). Computers can serve as a convivial tool (Falbel, 1991). The willingness of learners to learn will increase because they can use the computer in building artefacts (Papert, 1991).

Both constructivist and constructionist theories imply that learning depends on the learners themselves and learning can be enhanced through social discourse and discovery. Figure 1 illustrates the interconnections among the learning theories. Additionally, constructionist theory suggests that learning can be further enhanced if learners are involved in collaborative artefact designing projects using ICT as construction material.

Learning Approach

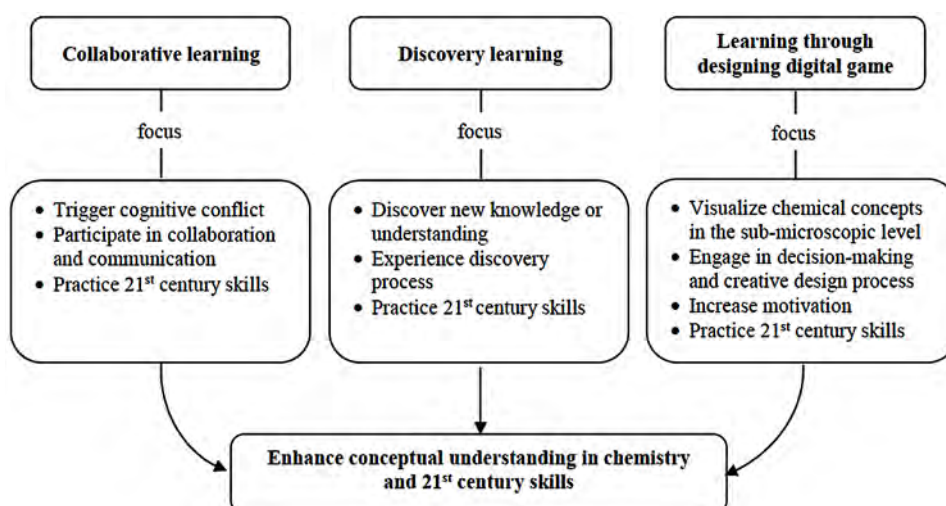
Based on constructivist and constructionist theories, learning approaches such as collaborative learning, discovery learning and learning through designing digital game (student as game designer) are integrated in MyKimDG. Figure 2 summarizes the focus of each learning approach in manoeuvring learning activities in MyKimDG.

Figure 1. Interconnections among the constructivist and constructionist learning theories



- Collaborative Learning:** Activities in MyKimDG are designed so that students engage in discussion, share and exchange ideas in groups. Through this approach, triggering of cognitive conflict and restructuring of ideas will occur when students share their ideas from their own perspective. It also improves students' 21st century skills such as collaboration, communication and interpersonal skill because students are able to practice in the real world.
- Discovery Learning:** Students are guided towards exploring chemical concepts. Students will gain deeper understanding when they are given opportunities to discover or construct new concepts or understanding for themselves. It also lets students experiencing the discovery activities, and stimulating their own thinking. As students embark on the discovery process, teacher reminds them of the important of the process in learning. If they can perceive the values of the process, they will be motivated to learn chemistry. In this approach, students are empowered to take responsibility for their own learning and practice the 21st century skills in real situations.
- Learning Through Designing Digital Game:** In MyKimDG, students are involved in designing PowerPoint games related to chemical concepts. They discuss in groups and apply the concepts learned to design PowerPoint games. With this, students can visualize the concepts in the sub-microscopic level. This approach can also create or learning environment that allows students to make decisions about the design of the desired PowerPoint game, apply ICT and collaborate in a social context.

Figure 2. Focus of learning approaches in MyKimDG



PowerPoint game is selected as Microsoft PowerPoint software is available at all schools and the use of the software does not involve additional cost and complicated programming languages. The only technical skill that students need to master to design PowerPoint games is how to create custom animations. In addition, existing PowerPoint game templates are available online and can be modified by students to help them progressively master the game designing skills. This strategy is parallel with the development phases proposed by Rieber, Barbour, Thomas and Rauscher (2008). However, students are also encouraged to use other software like *Game Maker* and programming languages such as *Java*, *Logo* and *Scratch* if they are skilled in the software.

When students carry out their digital game designing project, they are guided to move through the creative design spiral (Rusk *et al.*, 2009) in order to assist them develop new ideas. Students are also given the autonomy to choose their own game design, plan and carry out the project based on the group's consensus. The students are also told that the PowerPoint game will be used to help their peers who face difficulty in learning the chapter. It is expected that this strategy will improve students' perceived competency, autonomy and relatedness, and hence increase their motivation in chemistry.

The learning through designing approach aims to deepen students' conceptual understanding in chemistry as cognitive conflict may be triggered during activities and hence, new understanding may discover. As the same time, it provides a platform for students to develop their 21st century skills.

Instructional Model and Strategy

Over the past decade or so, there has been a significant effort to improve conceptual understanding. Studies have revealed that mastery of science concepts will be enhanced if students become aware of their misconception. To help students realize their misconception and replaced it with

scientifically acceptable concept (i.e. conceptual change), cognitive conflict strategy has been proposed by scholars such as Piaget (1977) and Posner, Strike, Hewson and Gertzog (1982). Therefore, the BSCS 5E Instructional Model (Bybee *et al.*, 2006) designed to facilitate conceptual change is applied in MyKimDG.

To help students understand the chemical concepts, students are guided to explain macroscopic experience at the sub-microscopic and symbolic levels. It is known that conceptual understanding in chemistry involves making use of three main representations or levels. The triplet relationship is the key model in chemical education (Gilbert & Treagust 2009).

In MyKimDG, the phases of the BSCS 5E Instructional Model and Creative Design Spiral have been modified and standardized. The resultant phases are Inquiry, Discover, Produce, Communicate and Review (IDPCR). Table 1 shows the IDPCR phases in MyKimDG and related phases of the BSCS 5E Instructional Model and Creative Design Spiral.

During implementation of MyKimDG, students are guided to experience and realise the IDPCR phases. As the process is done repeatedly, new ideas are always generated and students' 21st century skills such as inventive thinking skills are developed. Students are expected to practice the process in everyday life and in the workplace. Apart from that, it is expected that the acronym IDPCR can help students remember the five im-

Table 1. Phases in MyKimDG and related phases of the BSCS 5E Instructional Model and Creative Design Spiral

MyKimDG	BSCS 5E Instructional Model	Creative Design Spiral
Inquiry	Engage	Imagine
Discover	Explore	Experiment
Produce	Elaborate	Create
Communicate	Explain	Share
Review	Evaluate	Reflect

portant clusters of 21st century skills, i.e. Inventive thinking, Digital-age literacy, high Productivity, effective Communication and spiritual values (*nilai Rohani*). The five clusters of 21st century skills have been identified by Kamisah and Neelavany (2010).

Table 2 shows the outline of instructional activities in MyKimDG.

IMPLEMENTATION OF MYKIMDG

In the following section, the authors present a brief lesson in MyKimDG to the teaching and learning of a specific unit which involved precipitation reaction (see Figure 3). The lesson is designed to

created learning environment that allows students to work together to learn and discover ideas or concepts. Activities are designed to engage students in self-assessing their ideas, communicating their ideas and making decisions based on the group's consensus. To assist students understand why precipitation reaction is used in the preparation of insoluble salt, they are engaged in discovery activity (see the Inquiry phase). Afterwards, they are given opportunity to plan and conduct investigation based on the conclusion made (see the Discover phase).

To extend students' understanding about the observed phenomenon (i.e. precipitation reaction), they are given tasks to design digital games using ICT to teach their peers who face problems

Table 2. Outline of instructional activities in MyKimDG

Phase	Purpose	Activity
Inquiry <i>Predict, ask, hypothesize, identify problem, brainstorm</i>	1. Arouse students' interest 2. Access students' prior knowledge 3. Elicit students' misconceptions 4. Clarify and exchange current conceptions	1. Teacher shows discrepant events. 2. Students make observations and explain the phenomena at the sub-microscopic and symbolic levels. 3. Students discuss in groups and compare their ideas with their peers.
Discover <i>Investigate, experiment, explore</i>	1. Expose to conflicting situations 2. Modify current conceptions and develop new conceptions 3. Provide opportunities for students to demonstrate their conceptual understanding, and skills	1. Students perform hands-on and minds-on activities in groups. 2. Students are encouraged to engage in discussions and information seeking using ICT. 3. Students generate explanation of the observed phenomenon. 4. Students practise the skills needed in an experiment or activity. 5. Students are asked to communicate in groups and report back with their findings.
Produce <i>Create, construct, invent, build, design, tinker, elaborate</i>	1. Challenge and deepen students' conceptual understanding and skills 2. Provide additional time and experiences that contribute to the generation of new understanding	1. Students apply their new ideas by conducting additional activities 2. Students perform additional tasks that are more complex and involve higher order thinking skills (HOTS). 3. Students carry out open-ended projects. 4. Students create digital games.
Communicate <i>Explain, share, discuss with peers, ask an expert, defend</i>	1. Provide opportunities for students to share their new understanding and skills 2. Provide opportunities for students to exchange their new understanding	1. Students report back with their new ideas and skills. 2. Students also listen to input from peers and defend their ideas. Peer's input may guide them towards deeper level of understanding. 3. Students compare their ideas with the teacher's explanations.
Review <i>Check, evaluate, reflect, improve, repair</i>	1. Students assess their understanding, skills and competencies 2. Students think creatively for the purpose of improvement 3. Teachers evaluate student progress toward achieving the learning outcomes	1. Students reflect upon the extent to which their understanding, abilities and competencies have changed. 2. Students improve their ideas or skills based on reflection or input from peers. 3. Teacher conducts a test to determine the level of understanding of each student.

in the learning of the chemical concept (see the Produce phase). At this phase, students are engaged in designing *PowerPoint* games to represent the dissolution and precipitation reactions at the sub-microscopic level. First, each student develop an 'expressed model' to represent his/her own visual perception (or mental model) about the reactions. An individual's mental model is inaccessible to

others but the expressed models can be used for discussion (Gilbert, 2005). Eventually, a final model is developed based on group's consensus. Figure 4 shows example of dissolution model created by students. Through this activity, students gain knowledge about what occur between particles during reactions, hence increase students' conceptual understanding. Additionally, students

Figure 3. A brief lesson in MyKimDG

Phase 1: Inquiry

1. Teacher demonstrates two reactions that may be used to prepare lead(II) sulphate:

	Reaction	Observation	Chemical equation
A	Lead(II) nitrate solution + sodium sulphate solution		
B	Excess solid lead(II) carbonate + dilute sulphuric acid		

2. Students record the observations and write the chemical equations involved.
3. Students describe how to obtain lead(II) sulphate from the mixture in Reaction A and B.
 - (a) Draw the set-up of the apparatus is involved.
 - (b) In your opinion, which reaction is more appropriate to prepare insoluble salts such as lead(II) sulphate? Explain your answer.
4. Students make a conclusion about the appropriate reaction to prepare insoluble salts.
5. Students share their findings with other groups.
6. Students are asked to explain the strategy used, i.e. inquiry-discovery.

Phase 2: Discover

1. Students plan experiments to prepare lead(II) iodide and silver chloride in group.
 - (a) Discuss the materials needed to prepare lead (II) iodide and silver chloride.
 - (b) Write the chemical equations involved.
 - (c) Plan the procedures for experiment by constructing flowchart.
2. Students carry out experiment to prepare lead(II) iodide and silver chloride.
3. Students generate explanation of each phenomenon.
4. Students are asked to report back with their findings.

Phase 3: Produce

1. Students play a game related to the precipitation reactions involved in the preparation of insoluble salts.
2. Students are asked to differentiate between a good game and a bad game.
3. Students are asked to improve the game to make it more educational and entertaining following phases of IDPCR, in order to help their peers who face difficulty in learning the concept.
 - 3.1 Inquiry: Students convey his/her idea about the chemical concept and design of game through model (i.e., diagram, drawing, physical replica, etc.). Students select a promising model from their brainstorming session.
 - 3.2 Discover: Students create their designs using *PowerPoint*. Students are encouraged to test frequently and think critically about their designs, and rebuild as needed.
 - 3.3 Produce: Students produce their *PowerPoint* games based on improvements suggested through testing.
 - 3.4 Communicate: Students share their designs and *PowerPoint* games and get input from other groups.
 - 3.5 Review: Students describe the key strengths and weaknesses of their designs and *PowerPoint* games. Students create their own *PowerPoint* game in groups that incorporates the best aspects of all the designs.
4. Students are told that they may commercial their innovative product to benefit financially.
5. Students are reminded to apply 21st century skills during the project.

Phase 4: Communicate

1. Students share their digital games with other science or chemistry educators.
2. Students improve their digital games.

Phase 5: Review

1. Students plan and carry out experiments to prepare lead(II) chromate and barium sulphate in group.
2. Students write the chemical and ionic equation involved.
3. Students reflect upon the extent to which their understanding, abilities and competencies have changed.

need to carefully plan, utilize time and 21st century tools and resources toward the goal. At the end of project, they are also asked to improve and produce higher quality *PowerPoint* games that incorporates the best aspects of other groups' designs. The strategy is able to increase students' 21st century skills because students are able to immerse themselves in the real-world practice.

Students are also engaged in design or product justification (see the Produce and Communication phases). Students listened to input from peers and teacher, and defended their ideas. Input from peer and teacher may triggered cognitive conflict and result in reconstruction of existing ideas. Such support or scaffolding deepen students' understanding. At the end of lesson, students are assessed to determine their level of understanding. Students also reflect on how the IDPCR phases have contributed to their own development.

CONCLUSION

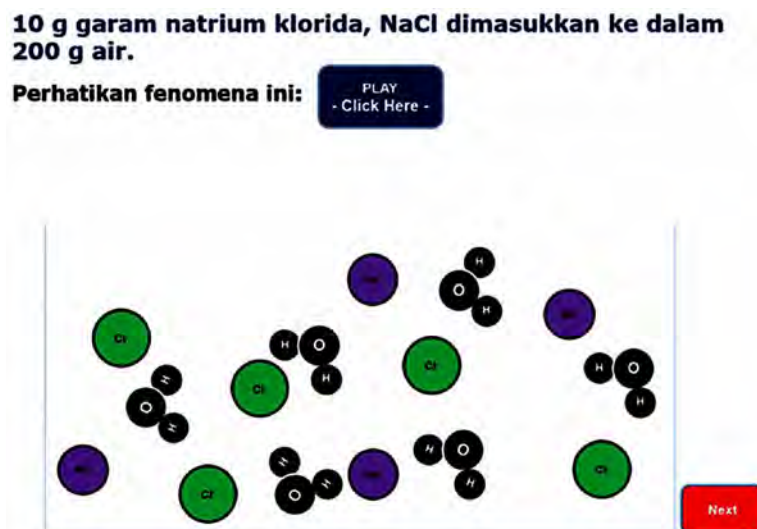
Students can be assisted to improve their conceptual understanding in chemistry and develop their 21st century skills. In this study, the authors have tried to provide an innovative chemistry

instructional strategy by integrating collaborative learning, discovery learning and learning through designing digital games. The learning strategy will create supportive learning environment for student to learn chemistry meaningfully. Most importantly, MyKimDG allows students to practice the 21st century skills in real situations. In conclusion, the implementation of MyKimDG can lift students' achievement in chemistry and their 21st century skills.

FUTURE RESEARCH DIRECTIONS

Quality assessment is essential in measuring the effectiveness of MyKimDG. Assessments enable teachers to evaluate learning while it is occurring. However, assessing of 21st century skills learning can be quite challenging. For instance, when faced with a collaborative task, teacher need to decide how to assign credit to each member of the group, as well as how to account for differences across groups that may bias a given student's performance (Binkley et al., 2012). Hence, further study need to be carried out to develop ways or specific rubrics to assess the 21st century skills. Assessment rubrics offer clear guidelines or expectations to

Figure 4. Example of dissolution model created by students



students on how the expectations are being met as well as the acceptable level of achievement. The assessment results inform teachers and students about their progress. It can also be leveraged to inform the revision of the MyKimDG module being developed.

REFERENCES

- Balaban, A. T., & Klein, D. J. (2006). Is chemistry The Central Science? How are different sciences related? Co-citations, reductionism, emergence, and posets. *Scientometrics*, 69(3), 615–637. doi:10.1007/s11192-006-0173-2
- Baytak, A., & Land, S. M. (2010). A case study of educational game design by kids and for kids. *Procedia: Social and Behavioral Sciences*, 2(2), 5242–5246. doi:10.1016/j.sbspro.2010.03.853
- Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., Miller-Ricci, M., & Rumble, M. (2012). Defining twenty-first century skills. In P. Griffin, B. McGaw, & E. Care (Eds.), *Assessment and teaching of 21st century skills* (pp. 17–66). Dordrecht: Springer. doi:10.1007/978-94-007-2324-5_2
- Brown, T. L., LeMay, H. E., Bursten, B. E., Murphy, C., & Woodward, P. (2011). *Chemistry: The central science* (12th ed.). Prentice Hall.
- Bruner, J. S. (1966). *Toward a theory of instruction*. Cambridge, MA: Harvard University Press.
- Bybee, R. W., Taylor, J. A., Gardner, A., Van Scotter, P., Powell, J., Westbrook, A., & Landes, N. (2006). *The BSCS 5E instructional model: Origins and effectiveness*. Colorado Springs, CO: Biological Sciences Curriculum Studies.
- Chang, R. (2007). *Chemistry* (9th ed.). New York, NY: McGraw-Hill.
- Chew, C. M., Noraini, I., Leong, K. E., & Daud. (2013). Secondary school assessment practices in science, technology, engineering and mathematics (STEM) related subjects. *Journal of Mathematics Education*, 6(2), 58–69.
- Egenfeldt-Nielsen, S. (2006). Overview of research on the educational use of video games. *Digital Kompetanse*, 1, 184–213.
- Falbel, A. (1991). The computer as a convivial tool. In I. Harel & S. Papert (Eds.), *Constructionism* (pp. 29–37). Norwood, NJ: Ablex Publishing Corporation.
- Gilbert, J. K. (2005). Visualization: A metacognitive skill in science and science education. In J. K. Gilbert (Ed.), *Visualization in science education* (pp. 9–27). Dordrecht: Springer Netherlands. doi:10.1007/1-4020-3613-2_2
- Gilbert, J. K., & Treagust, D. F. (2009). Introduction: Macro, submicro and symbolic representations and the relationship between them: Key models in chemical education. In J. K. Gilbert & D. Treagust (Eds.), *Multiple representations in chemical education* (pp. 1–8). Dordrecht: Springer Netherlands. doi:10.1007/978-1-4020-8872-8_1
- Hwang, G.-J., Sung, H.-Y., Hung, C.-M., Yang, L.-H., & Huang, I. (2013). A knowledge engineering approach to developing educational computer games for improving students differentiating knowledge. *British Journal of Educational Technology*, 44(2), 183–196. doi:10.1111/j.1467-8535.2012.01285.x
- Jung, J. Y., & Park, H. (2009). Learning by doing via game making. In D. Gibson & Y. Baek (Eds.), *Digital simulations for improving education: Learning through artificial teaching environments* (pp. 394–406). Hershey, PA: Information Science Reference. doi:10.4018/978-1-60566-322-7.ch021

- Kafai, Y. B. (1996). Learning design by making games: Children's development of design strategies in the creation of complex computational artifact. In Y. Kafai & M. Resnick (Eds.), *Constructionism in practice: Designing, thinking, and learning in a digital world* (pp. 71–96). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Kafai, Y. B., & Ching, C. C. (1996). Meaningful contexts for mathematical learning: The potential of game making activities. In *Proceedings of the 1996 International Conference on Learning Sciences* (pp. 164–171). Academic Press.
- Kafai, Y. B., Ching, C. C., & Marshall, S. (1997). Children as designers of educational multimedia software. *Computers & Education*, 29(2-3), 117–126. doi:10.1016/S0360-1315(97)00036-5
- Kafai, Y. B., & Resnick, M. (1996). Introduction. In Y. Kafai & M. Resnick (Eds.), *Constructionism in practice: Designing, thinking, and learning in a digital world* (pp. 1–8). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Lim, C. P. (2008). Spirit of the game: Empowering students as designers in schools? *British Journal of Educational Technology*, 39(6), 996–1003. doi:10.1111/j.1467-8535.2008.00823_1.x
- Osman, K., & Bakar, N. A. (2013). Teachers and students as game designers: Designing games for classroom integration. In S. de Freitas, M. Ott, M. M. Popescu, & I. Stanescu (Eds.), *New pedagogical approaches in game enhanced learning: Curriculum integration* (pp. 102–113). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-3950-8.ch006
- Osman, K., & Marimuthu, N. (2010). Setting new learning targets for the 21st century science education in Malaysia. *Procedia: Social and Behavioral Sciences*, 2(2), 3737–3741. doi:10.1016/j.sbspro.2010.03.581
- Papert, S. (1991). Situating constructionism. In I. Harel & S. Papert (Eds.), *Constructionism* (pp. 1–28). Norwood, NJ: Ablex Publishing Corporation.
- Papert, S. (1996). A word for learning. In Y. Kafai & M. Resnick (Eds.), *Constructionism in practice: Designing, thinking, and learning in a digital world* (pp. 9–24). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Papert, S. (1998). Does easy do it? Children, games, and learning. *Game Developer Magazine*, 88.
- Papert, S. (1999a). *Eight big ideas behind the constructionist learning lab*. Retrieved September 30, 2015, from <http://stager.org/articles/8bigideas.pdf>
- Papert, S. (1999b). Introduction: What is Logo? Who needs it? In C. Fonseca, G. Kozberg, M. Tempel, S. Soprunov, E. Yakovleva, H. Regini, ... D. Cavallo (Eds.), *Logo philosophy and implementation* (pp. iv–xvi). Logo Computer System Inc.
- Papert, S., & Franz, G. (1988). Computer as material: Messing about with time. *Teachers College Record*, 89(3), 408–417.
- Piaget, J. (1977). *The development of thought: Equilibration of cognitive structures*. New York, NY: Viking Press.
- Pivec, P. (2009). *Game-based learning or Game-based teaching?* British Educational Communications and Technology Agency (BECTA). Retrieved September 30, 2015, from <http://dera.ioe.ac.uk/id/eprint/1509>
- Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of scientific conception: Toward a theory of conceptual change. *Science Education*, 66(2), 211–227. doi:10.1002/sce.3730660207
- Prensky, M. (2001). Digital natives, digital immigrants. *On the Horizon*, 9(5), 1–6. doi:10.1108/10748120110424816

- Prensky, M. (2008). Students as designers and creators of educational computer games: Who else? *British Journal of Educational Technology*, 39(6), 1004–1019. doi:10.1111/j.1467-8535.2008.00823_2.x
- Rieber, L., Barbour, M., Thomas, G., & Rauscher, D. (2008). Learning by designing homemade powerpoint games. In C. T. Miller (Ed.), *Games: Purpose and potential in education* (pp. 23–42). New York, NY: Springer.
- Risch, B. (2010). *Teaching chemistry around the world*. Munster: Waxmann.
- Rusk, N., Resnick, M., & Cooke, S. (2009). Origins and guiding principles of the Computer Clubhouse. In Y. B. Kafai, K. A. Peppler, & R. N. Chapman (Eds.), *The Computer Clubhouse: Constructionism and creativity in youth communities* (pp. 17–25). New York, NY: Teachers College.
- Van Eck, R. (2006). Digital game-based learning: It's not just the digital natives who are restless. *EDUCAUSE Review*, 41(2), 16–18, 20, 22, 24, 26, 28, 30.
- Vos, N., van der Meijden, H., & Denessen, E. (2011). Effects of constructing versus playing an educational game on student motivation and deep learning strategy use. *Computers & Education*, 56(1), 127–137. doi:10.1016/j.compedu.2010.08.013
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Brennan, K., & Resnick, M. (2013). Imagining, creating, playing, sharing, reflecting: How online community supports young people as designers of interactive media. In C. Mouza & N. Lavigne (Eds.), *Emerging technologies for the classroom: A learning sciences perspective* (pp. 253–268). New York: Springer. doi:10.1007/978-1-4614-4696-5_17
- Carbonaro, M., Szafron, D., Cutumisu, M., & Schaeffer, J. (2010). Computer-game construction: A gender-neutral attractor to Computing Science. *Computers & Education*, 55(3), 1098–1111. doi:10.1016/j.compedu.2010.05.007
- Kafai, Y. B. (2006). Playing and making games for learning: Instructionist and constructionist perspectives for game studies. *Games and Culture*, 1(1), 36–40. doi:10.1177/1555412005281767
- Ke, F. (2014). An implementation of design-based learning through creating educational computer games: A case study on mathematics learning during design and computing. *Computers & Education*, 73, 26–39. doi:10.1016/j.compedu.2013.12.010
- Li, Q. (2010). Digital game building: Learning in a participatory culture. *Educational Research*, 52(4), 427–443. doi:10.1080/00131881.2010.524752
- Martinez, S. L., & Stager, G. (2013). *Invent to learn: Making, tinkering, and engineering in the classroom*. Torrance, CA: Constructing Modern Knowledge Press.
- Peppler, K., & Kafai, Y. (2009). Making games, art, and animations with Scratch. In Y. B. Kafai, K. A. Peppler, & R. N. Chapman (Eds.), *The Computer Clubhouse: Constructionism and creativity in youth communities* (pp. 47–57). New York, NY: Teachers College.

ADDITIONAL READING

Annetta, L. A., Cheng, M.-T., & Holmes, S. (2010). Assessing twenty-first century skills through a teacher created video game for high school biology students. *Research in Science & Technological Education*, 28(2), 101–114. doi:10.1080/02635141003748358

KEY TERMS AND DEFINITIONS

21st Century Skills: A set of skills that students need to develop in order to succeed in the 21st century.

Constructivism: A learning theory that suggests that individual learner actively constructs new knowledge pursuant to his/her existing knowledge.

Constructionism: A learning theory that suggests that construction of new knowledge happen best through designing external and sharable artefacts.

Digital Game: Any game played using electronic device, either online or stand-alone.

Digital Game-Based Learning: An instructional approach that integrates digital games in learning.

Learning Through Designing: An instructional approach that integrates design activities in learning.

MyKimDG: A module that incorporates collaborative learning, discovery learning and learning through designing digital game in chemistry learning.

Clinical Use of Video Games

Ben Tran

Alliant International University, USA

INTRODUCTION

Video game play is the fastest growing form of entertainment in the world and many adolescents play video games for hours every day. For example, a nationally representative study of video game play among adolescents in the United States showed that 97% of adolescents aged 12 to 17 years play computer, web, and portable or console video games (Lenhart, Kahne, Middaugh, Macgill, Evans, & Vitak, 2008). In terms of frequency, 31% of adolescents play video games every day and another 21% play video games 3 to 5 days a week. Similarly, Gentile (2009) conducted a large survey study in the United States and found that 88% of youth aged 8 to 18 years play video games and the average amount of time spent playing video games per week is 13.2 hours. Furthermore, in the United States, 91% of children between the ages of 2 and 17 play video games (Granic, Lobel, & Engels, 2014), and a nationally representative study of U.S. teenagers found that up to 99% of boys and 94% of girls play video games (Lenhart et al., 2008).

A video game, according to Wikipedia (2016), is an electronic game that involves human interaction with a user interface to generate visual feedback on a video device such as a television screen or computer monitor. The word video in video game traditionally referred to a raster display device, but in the 2000s, it implies any type of display device that can produce two- or three-dimensional images. The electronic systems used to play video games are known as platforms and these platforms range from large mainframe computers to small handheld computing devices (Wikipedia, 2016). Hence, in the United States alone, video games brought in over \$25 billion in

2010, more than doubling Hollywood's 2012 box office sales of \$10.8 billion in the United States and Canada (Motion Picture Association of American, 2012). Despite the extreme popularity of video games among adolescents, however, researchers in the fields of developmental and social psychology examining video game have focused mainly on the association between video game use and negative outcomes, while research on positive outcomes is more limited. Video games, and the usage of video games, in adolescent therapy and psychotherapy is anything but ubiquitous. Hence, the purpose of this chapter is on video games and their usages in adolescent therapy and psychotherapy. This chapter will cover a brief and condensed history in the usage of video games in relations to clinical usage (and not on the historical development of video games and the video game industry per se), video games in psychotherapy, and the different types of video games and their usages in psychotherapy.

BACKGROUND

According to Ceranoglu (2010), the first video game materialized on an oscilloscope screen in 1958 (Kent, 2001) featuring a game of simulated tennis that amused visitors to Brookhaven National Laboratory. Thereafter, video games have become a major part of pop culture and the entertainment medium of choice for millions of people (Gettler, 2008; Poole, 2000). However, for many, video games were first created in the 1970s and since then have grown into a multibillion-dollar industry: the annual U.S. retail sales of video games reached more than \$9.9 billion in 2004 alone (Greitemeyer & Osswald, 2010; Sestir &

DOI: 10.4018/978-1-5225-2255-3.ch284

Bartholow, 2010). According to Greitemeyer and Osswald (2010), large-scale surveys show that 70% of homes with children ages 2 to 17 years have computers and 68% have video game equipment (Woodard & Gridina, 2000). Eighty-seven percent of children play video games regularly (Walsh, Gentile, Gieske, Walsh, & Chasco, 2003). Children ages 2 to 7 years spent an average of 3 to 5 hours a week playing video games (Gentile & Walsh, 2002), while 8th and 9th-grade students average 9 hours per week (Gentile, Lynch, Linder, & Walsh, 2004).

Video Games in Psychotherapy

A review of the available literature on video game use in psychotherapy takes a dynamic perspective in regards to psychotherapy, although it should be noted that the issues addressed here can be applied to many forms of psychotherapy. A literature search was conducted, according to Ceranoglu (2010), on Medline and PsycInfo with keywords *video games*, *psychotherapy*, *computer games*, and *child and adolescent*, and relevant manuscripts were also identified through citations in the articles identified during the primary search. In so doing, video game, is defined as a game that employs electronics to create an interactive system that includes a user interface to generate a visual feedback on a video device (Wolf, 2002), and found their way into the clinical care of youth in mental health care as well as other fields, and academic interest in clinical use of video games is increasing steadily.

Review of a database of academic manuscripts reveals that 1,121 of 1,474 total reports on video games (76%) were published in the past decade. Reported clinical uses of video games include psychoeducation in chronic disease management to increase treatment adherence (Yoon & Godwin, 2007) and physical therapy and rehabilitation following traumatic brain injury (Jannink, van der Wilden, Navis, Visser, Gussinklo, & Ijzerman, 2008). Video games also serve as valuable adjuncts in pain management during medical procedures

(Das, Grimmer, Sparnon, McRae, & Thomas, 2005; Gold, Kim, Kant, Joseph, & Rizzo, 2006), induction of anesthesia (Patel, Schieble, Davidson, Tran, Schoenberg, Delphin, & Bennett, 2006), or cancer chemotherapy (Kato, Cole, Bradlyn, & Pollock, 2008; Redd, Jacobsen, Die-Trill, Dermatis, McEvoy, & Holland, 1987).

Video Games and Their Usages in Psychotherapy

The usage of video games that have received some attention of video game researchers is that of therapeutic sessions (Annema, Verstraete, Abeele, Desmet, & Geerts, 2013). According to Annema et al. (2013), video games can help motivate patients, develop skills and serve as a distractor in pain management. Video games have been used in physiotherapy, occupational therapy, and psychotherapy (Burdea, 2003; Griffiths, 2003). As such, when considering the use of video games in therapy, a distinction can be made between games specifically made for therapeutic purposes and games made for the general public but that are used in therapy as well. An example of the last category, dating back almost two decades, is the use of the *Super Mario Brothers* and (*The Legend of*) *Zelda* games in addition to the regular psychotherapy program to assess and assist various abilities of children, such as problem solving strategies and means of dealing with success and failure (Gardner, 1991). Hence, literature survey revealed that different kinds of software are used in different types of clinical practice and research. These included serious games, commercial video games, computer programs developed for use in assessing cognitive abilities, and virtual reality used in psychiatric symptom research and training of health care professionals. Literature was also diverse in regard to delivery format of the games, such as computers, specifically designed hardware, handheld devices, and gaming consoles. With that said, Wilkinson, Ang, and Goh's (2008) article, is one of few articles that outline the video game revolution.

First Usage of Video Games

Specific video games are also designed for use in mental health care. The repeatability aspect of video games, as compared with other media forms, is useful in delivering manual-behavioral interventions such as those involved in cognitive-behavioral therapy (USAB, Holzinger, & Gesellschaft, 2007). Recently, the first video game developed to support cognitive-behavioral therapy by offering attractive electronic homework assignments and rehearsing basic psychoeducational parts of treatment has been reported in the literature (Brezinka, 2007, 2008). Video games were also found useful in group therapy for youth in distress, and have been noted to facilitate change in the moral developmental stage of adolescents involved in such therapy (Sherer, 1994). Another game is still in the testing phase for use in a solution-focused intervention for adolescents and is available to mental health professionals participating in studies of its efficacy. Therapists observe the game play to provide a structure to sessions, help build an effective patient-therapist relationship, and improve patient engagement in the therapeutic process (Coyle, Doherty, & Sharry, 2009). Games to enhance social skills training for children with developmental disorders also exist (Mineo, Ziegler, Gill, & Salkin, 2009). These programs feature virtual environments where the player controls as *avatar*, an actor within this environment, using a keyboard, mouse, or a gamepad. Through interactions with the virtual environment, the player may rehearse learned and problem-solving skills.

Some researchers from a psychology/psychotherapy background have developed their own games for use in therapy (Clark and Schoech, 1984; Oakley, 1994) while others have surveyed the use of computers in the area (Resnick & Sherer, 1994; Griffiths, 1997). Thereafter, in 1984, a psychotherapeutic text based game entitled *Adventures of Lost Loch* was developed for use in therapy by adolescents with low impulse control (Clark & Schoech, 1984). Hence, clients

were more cooperative with their therapists, with whom they developed effective therapeutic relationships and their session attendance rate greatly improved (Coyle, Matthews, Sharry, Nisbet, & Doherty, 2005).

Other games have incorporated subject matters particularly relevant to therapy. For example, SMACK (Oakley, 1994) deals with the decisions and consequences related to drugs. Furthermore, there has been increasing use of biofeedback-based video games for the treatment of anxiety disorders and attention problems. At NASA's Langley Research Center, Alan Pope has developed methods for using off-the-shelf Nintendo and Playstation games in combination with electroencephalogram (EEG) biofeedback, to train children with attention deficit disorder (ADD), attention deficit hyperactivity disorder (ADHD), and hyperactivity disorders. Results concluded that the inclusion of games in normal biofeedback treatments increased the therapeutic effect on ADD symptoms. Both children and parents rated as significantly higher enjoyment of coming to video game based sessions. Children found the sessions more inherently motivating and remained more focused on tasks (Pope & Paison, 2001).

Second Usage of Video Games

Although the 1980s saw a variety of controlled experiments on the effectiveness of video games in therapy, mental health research on video games in the 1990s came to be dominated by their adverse effects, and only in the 2000s does the argument for video game therapy reappear with frequency. In the 1980s Larose, Gagnon, Ferland, and Pepin (1989) reported significantly improved spatial abilities of adolescents with minimal brain damage or attention problems by means of a modified Atari game. During the following decade video games were increasingly studied as medical and psychosocial hazards (Griffiths, 1996) or addictive (Griffiths & Hunt, 1998). Recent reviews of the positive advances in video game therapy (Griffiths, 2004; Saloni-Pasternak & Gelfond,

2005) consistently respond to this preconception by opening on a defensive note. Despite this diversion of attention, video game therapy has been applied to a wide range of mental health concerns, all of which might now be extended by online treatment.

- **Aggression:** Given the trend to view video game violence as a factor in aggression, it is interesting to note that handful of researchers who have opted to explore an alternative viewpoint whereby video games might play an active role in curbing aggressive behavior. One team has pursued the long-term, development of a multimedia program called SMART Talk (Bosworth, Espelage, DuBay, & Daytner, Karageorge, 2000), which incorporates video games and simulations as part of a computer-based intervention that significantly diminishes middle-school students' beliefs that are supportive of violence and increases their intentions to use non-violent strategies.
- **Anxiety Disorders:** Inquiries into 2D video games for anxiety treatment can be found alongside a variety of graded exposure therapies using virtual reality. Sharry, McDermott, and Condrón (2003) propose *Relax to Win*, a biofeedback-based 2D game for the treatment of children with general anxiety problems. With *Relax to Win*, two on-screen dragons would race each other, and the more relaxed the players are, the faster the dragon would run, as measured by the players' galvanic skin response. Video games also prove realistic enough to generate successful graded exposure trails against phobias of spiders (Bouchard, Cote, Saint-Jacques, Robillard, & Renaud, 2006), heights or enclosed spaces (Emmelkamp, Krijn, Hulsbosch, De Vries, Schuemie, & Van Der Mast, 2002; Robillard, Bouchard, Fournier, & Renaud, 2003), and fear of driving after an accident

can be comparably reduced by virtual reality driving games, on their 2D equivalents (Walshe, Lewis, Kim, O'Sullivan, & Wiederhold, 2003).

- **Attention Deficit Hyperactivity Disorder:** Video game therapy for children with ADHD has elicited a relatively large amount of scholarly attention, because many children who do not inhibit their hyperactivity in other contexts will do so when playing intrinsically motivating video games. Lawrence, Houghton, Tannock, Douglas, Durkin, and Whiting (2002) highlighted that 6- to 12-year-old boys with ADHD perform beneath a normally developing control group when playing a cognitively demanding adventure video game, an even more so on a route task outside the laboratory, but perform equally well on a motor-skill targeting game that does not involve high working memory or distractor loads. This engagement makes video games ripe for therapeutic applications. Furthermore, there is some evidence that video games can contribute to reducing ADHD symptoms. The first therapeutic video game for children with ADHD was by Pope and Bogart (1996), whose Extended Attention Span Training (EAST) system modifies a NASA program that assesses the engagement of pilots using automated flight management systems. Pope and Palsson (2001) further develop this NASA patent as an ADHD intervention that decreases players' control over off-the-shelf PlayStation games when their EEG reports lower attention, measured as higher theta-to-beta wave ratio.
- **Autism:** Therapeutic computer games are of special interest for autism, since rules-based environments present a safe-appealing vehicle for interventions to improve socialization. More recently, it has been demonstrated that 13- to 18-year-olds with autistic spectrum disorder interact with

virtual environments on a par with control groups, suggesting a potential for virtual environments to serve as a medium for education about social conventions (Parsons, Mitchell, & Leonard, 2004, 2005). Furthermore, several researchers have worked on therapeutic video games for autistic children. Tanaka, Klaiman, Koenig, and Schultz (2005) are developing *Let's Face It!*, a suite of games designed to teach ability-appropriate distinctions between faces and objects, and recognition and labelling of facial expressions (Tanaka et al., 2005). Whalen, Liden, Ingersoll, Dallaire, and Liden (2006) developed a rigorously tested computer-assisted instructional program called *TeachTown*, which uses a suite of game-like tests and professionally designed visual reinforcements to make a demonstrated positive impact on receptive language, social understanding, self-help, attention memory, auditory processing and early academic skills for children with autism and other developmental delays.

- **Personality and Psychotic Disorders:** Scattered efforts have been made to bring video games to bear on the most challenging personality and psychotic disorders. On one end of the spectrum are Sieswerda, Arntz, and Wolfis (2005), who employ the simplest of worm and tennis games not for any therapeutic content, but to tease out responses involving potential differences in dichotomous evaluations among subjects with border-line personality disorder, cluster or antisocial personality disorders, and a control group with no diagnosed personality disorder. On the other end is Shrimpton and Hurwirth's (2005) account detailing the development of an elaborate adventure game to educate young people who have experienced their first psychotic episode. An interdisciplinary panel of experts is convened to study the prototype, and they demand substantial redevelopment-

mental in order to make the intervention an effective aid in young people's recovery from psychosis. Studies of video games applied to schizophrenia include Crookes and Moran (2003), who prepared a simple joystick-controlled game in order to assess how quickly subjects of different ages and genders can find cheese hidden on a 4x4 grid according to certain classical conditioning rules that healthy adults naturally filter out, but that schizophrenics do not, and Da Costa and De Catvalho (2004), who established that a group of medicated schizophrenics respond positively to completing cognitive tasks in virtual reality, suggesting a new medium for therapeutic interventions.

- **Children and the Elderly:** It is evident from the foregoing that children and adolescent have been the primary targets of therapeutic video games. While adults are comfortable with direct face-to-face dialogue, many children struggle to express themselves with words alone and use of therapeutic channels such as video games and other means provide children with an avenue for indirect communication (Coyle, Doherty, Matthews, & Sharry, 2007; Coyle, Matthews, Sharry, Nisbet, & Doherty, 2005). Several comprehensive overviews of the literature for this entire age range have appeared (Goh, Ang, & Tan, 2008; Griffiths, 2003; Saloni-Pasternak & Gelfond, 2005), and beyond these, several reports on subgroups can be found. Kokish (1994) presents the first recommendations on materials for computer play with preschoolers, and Aymard (2002) discusses a computer drawing game for projective use, to stimulate catharsis and affective expression in preschoolers and school-age children. Reaching at-risk youth by means of video games is also a recurring theme. For example, Resnick (1995) reports on *BUSTED*, a computer-

ized game to promote reflection on anti-social behavior in young offenders, and Sherer (1994) reports a controlled study of the effects of a moral development game on youth in distress. Dominic Interactive is an example of a DSM-IV-based diagnostic tool for children aged 6-12, in which a series of yes/no questions aim to give children greater scope for self-expression via a format akin to a video game. It has undergone validity testing in the U.S. (Valla, Bergeron, Saint-Georges, & Gaudet, 1997) and other countries (Villa, Kovess, Chan-Chee, Berthiaume, Vantalon, Piquet, Gras-Vincendon, Martin, & Alles-Jardel, 2002). The elderly, on the other hand, represent another demographic that has received its share of attention from therapeutic video game developers. Counting for Goblins is a computerized version of the counting Span task, which older subjects found more interesting than the offline equivalent, and which features automatic measurement of accuracy and response times (Barnes, Yaffe, Belfor, Jagust, DeCarli, Reed, & Kramer, 2006; Ryan, 1994).

FUTURE RESEARCH DIRECTIONS

When it comes to video games and their usage in psychotherapy in the future, only technology, and the imagination of software programmers and innovative therapists/psychologists can tell. Approximately a decade ago, Berker, Brinkman, and Deardorff (1995) tested a computer intervention for adolescent children of divorce that involved a trip to the lab to play a therapeutic video games, mail-out a post-test one week later, and a 10-minute follow-up telephone interview one week after that. Many practitioners will prefer the immediate warmth of the traditional face-to-face encounter for its associations and its proven benefits, but as we shall see, certain researchers have begun to explore such alternative virtual spaces as well.

The entire community stands to gain from this growing wealth of therapeutic options.

CONCLUSION

Nearly three decades after the first wave in change due to the home computer explosion, we can see that the mental health community's early concerns about low computer literacy and poor computer communication have turned out not to pose the expected barriers to therapeutic use. While the immense appeal of video games went on and caused a backlash among scholars who saw harm to their players, pioneering researchers have not harnessed the intrinsic motivation that they elicit to help treat ADHD, anxiety, and other psychiatric conditions. A second wave in change arrived with the advent of the internet, causing computers to be seen less as potential replacements for the therapist and more as conduits for therapeutic interactions (Wilkinson et al., 2008). The research and clinical potential for combining video games and the communicative possibilities of the internet are immense.

Future research into online video game therapy for mental health concerns might focus on two broad types of game: simple society games such as cards or chess, online versions of which are freely available, and elaborate online worlds such as massively multiplayer online role-playing games (MMORPGs), whose worldwide membership numbers is the millions. Both genres might be used for assessment and training purposes, and with an online chat component, both provide a limitless platform for social interaction that could be incorporated as part of the intervention program. Society games present the advantage of being accessible and enjoyable to players of all ages, while online worlds offer a unique opportunity for narrative content and immersive remote interaction with therapists and fellow patients. Congruent to Coyle et al. (2007), more collaborative efforts among therapists, engineers,

and computer software programmers to make such innovations more widely used.

REFERENCES

- Annema, J. H., Verstraete, M., Abeele, V. V., Desmet, S., & Geerts, D. (2013). Video games in therapy: A therapists perspective. *International Journal of Arts and Technology*, 6(1), 106–122. doi:10.1504/IJART.2013.050695
- Aymard, L. L. (2002). Funny face: Shareware for child counseling and play therapy. *Journal of Technology in Human Services*, 20(1), 11–29. doi:10.1300/J017v20n01_05
- Barker, J., Brinkman, L., & Deardorff, M. (1995). Computer intervention for adolescent children of divorce. *Journal of Divorce & Remarriage*, 23(3/4), 197–213. doi:10.1300/J087v23n03_13
- Barnes, D. E., Yaffe, K., Belfor, N., Jagust, W. J., DeCarli, C., Reed, B. R., & Kramer, J. H. (2006). Computerized-based cognitive therapy for mild cognitive impairment: Results of a pilot, randomized, controlled trial. *Alzheimer Disease and Associated Disorders*, 23(3), 205–210. doi:10.1097/WAD.0b013e31819c6137 PMID:19812460
- Bosworth, K., Espelage, D., DuBay, T., Daytner, G., & Karageorge, K. (2000). Preliminary evaluation of a multimedia violence prevention program for adolescents. *American Journal of Health Behavior*, 24(4), 268–280. doi:10.5993/AJHB.24.4.3
- Bouchard, S., Cote, S., Saint-Jacques, J., Robillard, G., & Renaud, P. (2006). Effectiveness of virtual reality exposure in the treatment of arachnophobia using 3D games. *Technology and Health Care*, 14(1), 19–27. PMID:16556961
- Brezinka, V. (2007). Treasure hunt—A psychotherapeutic game to support cognitive-behavioral treatment of children. *Verhaltenstherapie*, 17, 191–194. doi:10.1159/000106750
- Brezinka, V. (2008). Treasure hunt—A serious game to support psychotherapeutic treatment of children. *Studies in Health Technology and Informatics*, 136, 71–76. PMID:18487710
- Burdea, G. C. (2003). Virtual rehabilitation—Benefits and challenges. *Methods of Information in Medicine*, 42(5), 519–523. PMID:14654886
- Ceranoglu, T. A. (2010). Video games in psychotherapy. *American Psychological Association*, 14(2), 141–146.
- Clark, B., & Schoech, D. (1984). A computer-assisted therapeutic game for adolescents: Initial development and comments. In D. Marc & E. Schwartz (Eds.), *Using computers in clinical practice: Psychotherapy and mental health applications* (pp. 335–353). New York: Haworth Press.
- Coyle, D., Doherty, G., & Sharry, J. (2007). Computers in talk-based mental health interventions. *Interacting with Computers*, 19(4), 545–562. doi:10.1016/j.intcom.2007.02.001
- Coyle, D., Doherty, G., & Sharry, J. (2009). An evaluation of a solution-focused computer game in adolescent interventions. *Clinical Child Psychology and Psychiatry*, 14(3), 345–360. doi:10.1177/1359104508100884 PMID:19515752
- Coyle, D., Matthews, M., Sharry, J., Nisbet, A., & Doherty, G. (2005). Personal investigator: A therapeutic 3D game for adolescent psychotherapy. *International Journal of Interactive Technology and Smart Education*, 2(2), 73–88. doi:10.1108/17415650580000034
- Crookes, A. E., & Moran, P. M. (2003). An investigation into age and gender differences in human kamin blocking, using a computerized task. *Developmental Neuropsychology*, 24(1), 461–477. doi:10.1207/S15326942DN2401_03 PMID:12850754

- Da Costa, R. M. E. M., & De Carvelho, L. A. V. (2004). The acceptance of virtual reality devices for cognitive rehabilitation: A report of positive results with schizophrenia. *Computer Methods and Programs in Biomedicine*, 73(3), 173–182. doi:10.1016/S0169-2607(03)00066-X PMID:14980398
- Das, D. A., Grimmer, K. A., Sparnon, A. L., McRae, S. E., & Thomas, B. H. (2005). The efficacy of playing a virtual reality game in modulating pain for children with acute burn injuries: A randomized controlled trial. *BMC Pediatrics*, 5(1), 1471–2431. doi:10.1186/1471-2431-5-1 PMID:15745448
- Emmelkamp, P. M. G., Krijn, M., Hulsbosch, A. M., De Vries, S., Schuemie, M. J., & Van Der Mast, C. A. P. G. (2002). Virtual reality treatment versus exposure in vivo: A comparative evaluation in acrophobia. *Behaviour Research and Therapy*, 40(5), 509–516. doi:10.1016/S0005-7967(01)00023-7 PMID:12038644
- Gardner, J. E. (1991). Can the Mario Bros. help? Nintendo games as an adjunct in psychotherapy with children. *Psychotherapy (Chicago, Ill.)*, 28(4), 667–670. doi:10.1037/0033-3204.28.4.667
- Gentile, D. A. (2009). Pathological video-game use among youth ages 8 to 18: A national study. *Psychological Science*, 20(5), 594–602. doi:10.1111/j.1467-9280.2009.02340.x PMID:19476590
- Gentile, D. A., Lynch, P. J., Linder, J. R., & Walsh, D. A. (2004). The effects of violent video games habits on adolescent hostility, aggressive behaviors, and school performance. *Journal of Adolescence*, 27(1), 5–22. doi:10.1016/j.adolescence.2003.10.002 PMID:15013257
- Gentile, D. A., & Walsh, D. A. (2002). A normative study of family media habits. *Journal of Applied Developmental Psychology*, 23(2), 157–178. doi:10.1016/S0193-3973(02)00102-8
- Gettler, J. (2008). *The first video game? Before “pong,” there was “tennis for two”*. Retrieved on February 16, 2015. Available at <http://www.bnl.gov/about/history/firstvideo.php>
- Goh, D. H., Ang, R. P., & Tan, H. C. (2008). Strategies for designing effective psychotherapeutic gaming interventions for children and adolescents. *Computers in Human Behavior*, 24(5), 2217–2235. doi:10.1016/j.chb.2007.10.007
- Gold, J. I., Kim, S. H., Kant, A. J., Joseph, M. H., & Rizzo, A. S. (2006). Effectiveness of virtual reality for pediatric pain distraction during IV placement. *Cyberpsychology & Behavior*, 9(2), 207–212. doi:10.1089/cpb.2006.9.207 PMID:16640481
- Granic, I., Lobel, A., & Engels, R. C. M. E. (2014). The benefits of playing video games. *The American Psychologist*, 69(1), 66–78. doi:10.1037/a0034857 PMID:24295515
- Greitemeyer, T., & Osswald, S. (2010). Effects of prosocial video games on prosocial behavior. *Journal of Personality and Social Psychology*, 98(2), 211–221. doi:10.1037/a0016997 PMID:20085396
- Griffiths, M. D. (1997). Video games and clinical practice: Issues, uses and treatments. *The British Journal of Clinical Psychology*, 36(4), 639–641. doi:10.1111/j.2044-8260.1997.tb01271.x PMID:9403157
- Griffiths, M. D. (2003). The therapeutic use of videogames in childhood and adolescence. *Clinical Child Psychology and Psychiatry*, 8(4), 547–554. doi:10.1177/13591045030084012
- Griffiths, M. D. (2004). Can videogames be good for your health? *Journal of Health Psychology*, 9(3), 339–344. doi:10.1177/1359105304042344 PMID:15117533
- Griffiths, M. D., & Hunt, N. (1998). Dependence on computer game playing by adolescents. *Psychological Reports*, 82(2), 475–480. doi:10.2466/pr0.1998.82.2.475 PMID:9621722

- Holzinger, A., & Gesellschaft, O. C. (2007, November). HCI and usability for medicine and health care. *Proceedings of the Third Symposium of the Workgroup Human-Computer iNteraction and Usability Engineering of the Austrian Computer Society*.
- Jannink, M. J., van der Wilden, G. J., Navis, D. W., Visser, G., Gussinklo, J., & Ijzerman, M. (2008). A low-cost video game applied for training of upper extremity function in children with cerebral palsy: A pilot study. *Cyberpsychology & Behavior*, 11(1), 27–32. doi:10.1089/cpb.2007.0014 PMID:18275309
- Kato, P. M., Cole, S. W., Bradlyn, A. S., & Pollock, B. H. (2008). A video games improves behavioral outcomes in adolescents and young adults with cancer: A randomized trial. *Pediatrics*, 122(2), 305–317.
- Kent, S. L. (2001). *The ultimate history of video games: From pong to pokemon and beyond. The story behind the craze that touched out lives and changed the world*. New York: Three Rivers Press.
- Kokish, R. (1994). Experiences using a PC in play therapy with children. *Computers in Human Services*, 11(1/2), 141–150.
- Larose, S., Gagnon, S., Ferland, C., & Pepin, M. (1989). Psychology of computers: XIV. Cognitive rehabilitation through computer games. *Perceptual and Motor Skills*, 69(1/3), 851–858. doi:10.2466/pms.1989.69.3.851 PMID:2608401
- Lawrence, V., Houghton, S., Tannock, R., Douglas, G., Durkin, K., & Whiting, K. (2002). ADHD outside the laboratory: Boys executive function performance on tasks in videogame play and on a visit to the zoo. *Journal of Abnormal Child Psychology*, 30(5), 447–462. doi:10.1023/A:1019812829706 PMID:12403149
- Lenhart, A., Kahn, J., Middaugh, E., Macgill, A. R., Evans, C., & Vitak, J. (2008). *Teens, video games, and civics*. Washington, DC: Pew Internet and American Life Project.
- Mineo, B. A., Ziegler, W., Gill, S., & Salkin, D. (2009). Engagement with electronic screen media among students with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 39(1), 172–187. doi:10.1007/s10803-008-0616-0 PMID:18626761
- Motion Picture Association of America. (2012). *Theatrical market statistics 2012*. Retrieved on February 16th, 2015. Available at <http://www.mpa.org/wp-content/uploads/2014/03/2012-Theatrical-Market-Statistics-Report.pdf>
- Oakley, C. (1994). SMACK: A computer driven game for at-risk teens. *Computers in Human Services*, 11(1), 97–99.
- Parsons, S., Mitchell, P., & Leonard, A. (2004). The use and understanding of virtual environments by adolescents with autistic spectrum disorders. *Journal of Autism and Developmental Disorders*, 34(4), 449–466. doi:10.1023/B:JADD.0000037421.98517.8d PMID:15449520
- Parsons, S., Mitchell, P., & Leonard, A. (2005). Do adolescents with autism spectrum disorders adhere to social conventions in virtual environments? *Autism*, 9(1), 95–117. doi:10.1177/1362361305049032 PMID:15618265
- Patel, A., Schieble, T., Davidson, M., Tran, M. C., Schoenberg, C., Delphin, E., & Bennett, H. (2006). Distraction with a hand-held video game reduces pediatric preoperative anxiety. *Paediatric Anaesthesia*, 16(10), 1019–1027. doi:10.1111/j.1460-9592.2006.01914.x PMID:16972829
- Poole, S. (2000). *Trigger happy: Videogames and entertainment revolution*. New York: Arcade.
- Pope, A. T., & Bogart, E. H. (1996). Extended attention span training system: Video game neurotherapy for attention deficit disorder. *Child Study Journal*, 26(1), 39–51.
- Pope, A. T., & Paison, O. S. (2001). *Helping video games 'rewire our minds*. NASA Langley Research Centre.

- Redd, W. H., Jacobsen, P. B., Die-Trill, M., Dermatis, H., McEvoy, M., & Holland, J. C. (1987). Cognitive/attentional distraction in the control of conditioned nausea in pediatric cancer patients receiving chemotherapy. *Journal of Consulting and Clinical Psychology*, 55(3), 391–395. doi:10.1037/0022-006X.55.3.391 PMID:3597954
- Resnick, H. (1995). Electronic technology and rehabilitation: A computerized simulation game for youthful offenders. *Computers in Human Services*, 11(1/2), 61–67.
- Resnick, H., & Sherer, M. (1994). Computer games in the human services—A review. *Computers in Human Services*, 11(1/2), 17–29.
- Robillard, G., Bouchard, S., Fournier, T., & Renaud, P. (2003). Anxiety and presence during VR immersion: A comparative study of the reactions of phobia and non-phobic participants in therapeutic virtual environments derived from computer games. *Cyberpsychology & Behavior*, 6(5), 467–476. doi:10.1089/109493103769710497 PMID:14583122
- Ryan, E. B. (1994). Memory for goblins: A computer game for assessing and training working memory skill. *Computers in Human Services*, 11(1/2), 213–217.
- Salonius-Pasternak, D. E., & Gelfond, H. S. (2005). The next level of research on electronic play: Potential benefits and contextual influences for children and adolescents. *Human Technology*, 1(1), 5–22. doi:10.17011/ht/urn.2005123
- Sestir, M. A., & Bartholow, B. D. (2010). Violent and nonviolent video games produce opposing effects on aggressive and prosocial outcomes. *Journal of Experimental Social Psychology*, 46(6), 934–942. doi:10.1016/j.jesp.2010.06.005
- Sharry, J., McDermott, M., & Condrón, J. (2003). Relax to win: Treating children with anxiety problems with a biofeedback video games. *Eisteach*, 2(25), 22–26.
- Sherer, M. (1994). The effect of computerized simulation games on the moral development of youth in distress. *Computers in Human Services*, 11(1/2), 81–95.
- Shrimpton, B., & Hurworth, R. (2005). Adventures in evaluation: Reviewing a CD-ROM based adventure game designed for young people recovering from psychosis. *Journal of Educational Multimedia and Hypermedia*, 14(3), 273–290.
- Sieswerda, S., Arntz, A., & Wolfis, M. (2005). Evaluations of emotional noninterpersonal situations b patients with borderline personality disorder. *Journal of Behavior Therapy and Experimental Psychiatry*, 36(3), 209–225. doi:10.1016/j.jbtep.2005.05.004 PMID:15950931
- Tanaka, J., Klaiman, C., Koenig, K., & Schultz, R. T. (2005, May). *Plasticity of the neural mechanisms underlying face processing in children with ASD: behavioral improvements following perceptual training on faces*. Poster presented at the International Meeting for Autism Research, Boston, MA.
- Valla, J. P., Bergerson, L., Saint-Georges, M., & Gaudet, N. (1997). Reliability of the Dominic-R: A young child mental health questionnaire combining visual and auditory stimuli. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 38(6), 717–724. doi:10.1111/j.1469-7610.1997.tb01698.x PMID:9315981
- Valla, J. P., Kovess, V., Chan-Chee, C., Berthiaume, C., Vantalón, V., Piquet, C., & Alles-Jardel, M. et al. (2002). A French study of the Dominic interactive. *Social Psychiatry and Psychiatric Epidemiology*, 37(9), 441–448. doi:10.1007/s00127-002-0575-2 PMID:12242621
- Walsh, D., Gentile, D., Gieske, J., Walsh, M., & Chasco, E. (2003). *Eight annual video games report card*. Minneapolis, MN: National Institute on Media and the Family.

Walshe, D. G., Lewis, E. J., Kim, S. I., OSullivan, K., & Wiederhold, B. K. (2003). Exploring the use of computer games and virtual reality in exposure therapy for fear of driving following a motor vehicle accident. *Cyberpsychology & Behavior*, 6(3), 329–334. doi:10.1089/109493103322011641 PMID:12855091

Whalen, C., Liden, L., Ingersoll, B., Dallaire, E., & Liden, S. (2006). Behavioral improvements associated with computer-assisted instruction for children with developmental disabilities. *Journal of Speech and Language Pathology—Applied. The Behavior Analyst*, 1(1), 11–26.

Wikipedia. (2016). *Video game*. Retrieved on April 19, 2016, available at https://en.wikipedia.org/wiki/Video_game

Wilkinson, N., Ang, R. P., & Goh, D. H. (2008). Online video games therapy for mental health concerns: A review. *The International Journal of Social Psychiatry*, 54(4), 370–382. doi:10.1177/0020764008091659 PMID:18720897

Wolf, M. J. P. (2002). *The medium of the video game*. Austin, TX: University of Texas Press.

Woodard, E. H., IV, & Gridina, N. (2000). Media in the home 2000: The fifth annual survey of parents and children (Survey Series No. 7). Philadelphia, PA: University of Pennsylvania' Annenberg Public Policy Center.

Yoon, S. L., & Godwin, A. (2007). Enhancing self-management in children with sickle cell disease through playing a CD-ROM educational game: A pilot study. *Pediatric Nursing*, 33(1), 60–63. PMID:17411004

ADDITIONAL READING

Allen, D. H. (1984). The use of computer fantasy games in child therapy. In M. D. Schwartz (Ed.), *Using computers in clinical practice: Psychotherapy and mental health application* (pp. 329–334). Hawthorn Books, Inc.

Cook, J., & Hyde, M. H. (2013). *But it's just a game*. National Center for Youth Issues.

Corey, G. (2012). *Theory and practice of counseling and psychotherapy* (9th ed.). Cengage Learning.

Dandeneau, S. D., & Baldwin, M. W. (2004). The inhibition of socially rejecting information among people with high versus low self-esteem: The role of attentional bias and the effects of bias reduction training. *Journal of Social and Clinical Psychology*, 23(4), 584–603. doi:10.1521/jscp.23.4.584.40306

Gingerich, W. J., & Eisengart, S. (2000). Solution-focused brief therapy: A review of the outcome research. *Family Process*, 39(4), 477–498. doi:10.1111/j.1545-5300.2000.39408.x PMID:11143600

Gould, M. S., Munfakh, J. L., Lubell, K., Kleinman, M., & Parker, S. (2002). Seeking help from the internet during adolescence. *Journal of the American Academy of Child and Adolescent Psychiatry*, 41(10), 1182–1189. doi:10.1097/00004583-200210000-00007 PMID:12364839

Ivey, A. E., Ivey, M. B., & Zalaquett, C. P. (2013). *Intentional interviewing and counseling: Facilitating client development in a multicultural society* (8th ed.). Cengage Learning.

Langlois, M. (2013). *Reset: Psychotherapy & video games* (2nd ed.). Online Therapy Institute.

McGonigal, J. (2011). *Reality is broken: Why games make us better and how they can change the world*. Penguin Group.

Münsterberg, H. (2011). *Psychotherapy*. New York: Moffat, Yard and Company.

Offer, D., Howard, K. I., Schonert, K. A., & Ostrov, E. (1991). To whom do adolescents turn for help? Differences between disturbed and nondisturbed adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry*, 30(4), 623–630. doi:10.1097/00004583-199107000-00015 PMID:1890097

Robertson, J. (2000). *The effectiveness of a virtual role-play environment as a preparation activity for story writing*. Retrieved on February 26, 2015. Available at <https://www.era.lib.ed.ac.uk/bitstream/handle/1842/589/2000-judyr.pdf?sequence=2&isAllowed=y>

Robertson, J., & Oberlander, J. (2002). Ghost-writer: Educational drama and presence in a virtual environment. *Journal of Computer Mediated Communication*, 8(1), 0-0.

Rochlen, A. B. (2006). *Applying counseling theories: An online, case-based approach*. Pearson.

Ryokai, K., & Cassell, J. (1999). Computer support for children's collaborative fantasy play and storytelling. In *Proceedings of the 1999 conference on Computer support for collaborative learning (CSCL '99)*, Stanford University, Palo Alto, California. doi:10.3115/1150240.1150303

Seligman, L., & Reichenberg, L. W. (2014). *Selecting effective treatments: A comprehensive systematic guide to treating mental disorders* (4th ed.). Wiley.

Sharry, J. (2001). *Becoming a solution detective: Identifying your clients' strengths in practical brief therapy*. London: Brief Therapy Press.

Sharry, J. (2004). *Counselling children, adolescents and families*. London: Sage.

Squire, K. (2003). Video Games In Education. *International Journal of Intelligent Simulations and Gaming*, 2(1), 49–62.

Sue, D. W., & Sue, D. (2012). *Counseling the culturally diverse: Theory and practice* (6th ed.). Wiley.

Wigner, J. (1994). Narrative completion in the treatment of trauma. *Psychotherapy (Chicago, Ill.)*, 31(3), 415–423. doi:10.1037/0033-3204.31.3.415

Zimmerman, E., & Tekinbas, K. S. (2003). *Rules of play: Game design fundamentals*. The MIT Press.

KEY TERMS AND DEFINITIONS

Aggression: Is overt, often harmful, social interaction with the intention of inflicting damage or other unpleasantness upon another individual.

Anxiety Disorders: Are a category of mental disorders characterized by feelings of anxiety and fear, where anxiety is a worry about future events and fear is a reaction to current events.

Attention Deficit Hyperactivity Disorder (ADHD): Is a developmental neuropsychiatric disorder in which there are significant problems with executive functions that cause attention deficits, hyperactivity, or impulsiveness which is not appropriate for a person's age.

Autism: Is a neurodevelopmental disorder characterized by impaired social interaction, verbal and non-verbal communication, and restricted and repetitive behavior.

Commercial Video Games: Games that, in their original license, were not considered free-ware, but were re-released at a later date with a freeware license, sometimes as publicity for a forthcoming sequel or compilation release.

Game: A voluntary activity structured by rules, with a defined outcomes or other quantifiable feedback that facilitates reliable comparisons of in-player performances.

Personality Disorders: A class of mental disorders characterized by enduring maladaptive patterns of behavior, cognition, and inner experience, exhibited across many contexts and deviating markedly from those accepted by the individual's culture.

Psychosis: Refers to an abnormal condition of the mind, and is a generic psychiatric term for a

mental state often described as involving a “loss of contact with reality”.

Psychotherapy: Is the treatment of a person’s problems by (typically) conversing with another person.

Serious Game (or Applied Game): Is a game designed for a primary purpose other than pure entertainment.

Therapy: Is the attempted remediation of a health problem, usually following a diagnosis.

Video Game: Games that are designed for players to actively engage with their systems

and for these systems to, in turn, react to players’ agentive behaviors.

Leveraging the Arduino Platform to Develop Information Technology Devices



Diego Reforgiato Recupero

University of Cagliari, Italy

Valentino Artizzu

University of Cagliari, Italy

Francesca Cella

University of Cagliari, Italy

Alessandro Cotza

University of Cagliari, Italy

Davide Curcio

University of Cagliari, Italy

Giorgio Amedeo Iengo

University of Cagliari, Italy

Riccardo Macis

University of Cagliari, Italy

Andrea Marras

University of Cagliari, Italy

Simone Picci

University of Cagliari, Italy

Michael Planu

University of Cagliari, Italy

Riccardo Scasseddu

University of Cagliari, Italy

INTRODUCTION

Arduino is a famous board, which incorporates serial communication interfaces, including universal serial bus (USB), and an integrated development environment (IDE) based on Processing, a programming language that supports C and C++. It consists of a microcontroller with several other components that provide easy interconnections with other devices.

An increasing number of developers utilizes Arduino for projects related to information technology and its community is growing and expanding in different domains, areas and countries creating several opportunities both in research and within the market sector. It is not surprising that several start-ups have initiated their business basing their services on Arduino. If the reader browses Kickstarter¹, the world's largest funding platform for creative projects, he would notice that there have

DOI: 10.4018/978-1-5225-2255-3.ch285

been a number of successful projects revolving around the Arduino platform. Moreover, there are also many devices that are Arduino's compatible that represent information technology products where companies and individuals can exploit for their business.

A lot of information is today exchanged within the domain of home automation. Home automation is what makes houses getting smart giving the users the possibility to monitor it using an interface on their computer, tablet or smartphone, or even panels mounted around the house. More in detail, home automation involves the control and automation of lighting, heating, ventilation, air conditioning, appliances and security. With a simple push of a button or a voice command it is possible to control items around the house, setting up a lamp to turn on or off according to certain events, switching on or off particular appliances depending on specific actions, etc. It is even possible to check the operating status and power consumption of any appliances through the web when they are connected to the Internet. As there are very few world-wide accepted industry standards and the fact that the smart home space is heavily fragmented, today, thanks to the availability of tools such as Arduino, it is possible to easily design, build and use smart appliances self made. Sometimes, it is even possible to disassemble a certain device to see how it is made or getting its core and build around a desired sub-device that can provide functions not present within the original hardware. One example that we provide in this chapter is represented by a joystick used for video games that has been taken apart to see how it was designed and an Arduino board has been embedded to provide a mechanism that counts all the times a button is pressed or a certain position of the controller has been pushed. In fact, the signals and data that the sensors on top of Arduino collect can be easily stored in a database for further processing and analytics.

Arduino and its components have been studied during the class of Computer Architecture for the degree in Computer Science at the University of

Cagliari in 2016. At the end of the class, seven groups of students have chosen to carry out an information system prototype on top of Arduino and show their methodology and possible obstacles they encountered. Within their projects, students had to integrate Arduino with sensors of different kind, extract data from sensors and perform some operations on them, such as analytics or store them in a database for further processing. Information science and technology includes several software technologies but with the widespread of the Internet of Things, it is possible to collect data of several kind from the every object equipped with sensors. One of the goals was to let the young researchers understand the vast amount of data that is possible to gather using sensors embedded in the Arduino board.

The first two projects presented in this chapter integrate Arduino with temperature sensors, a LCD and a button that gives the option to switch among the different information the system reads from its sensors (temperature, humidity level, date and time). Data read from the sensors are stored using NoSQL approach (we have adopted Spark) and ready for visualization and quick analytics.

One more project has been focused on a security appliance to be used in houses: a photocell detects a light with certain intensity. If the intensity is greater (lower) than a fixed threshold, a relay is activated to turn on (off) an electric appliance. A shift register has been used to both leave free some pins of the Arduino board and to modularize the project. Two push buttons with a rotary encoder have also been adopted to easily change the threshold levels. Read values can be stored in a database.

A fourth project exploits Arduino's flexibility to create a joystick game controller within a case similar to the old fashion coin-operated entertainment machines typically installed in restaurants, bars and amusement arcades. The joystick is provided with USB connector and plug-and-play capabilities so that it can be recognized on the fly by a normal pc. Similarly, another project integrates Arduino with an analogic joystick game

controller and 8 different led lights. Each different movement of the joystick turns on one of the lights. The rationale behind this project is to replace led lights with smart appliances in a house so that them can be controlled with a joystick.

A sixth project uses Arduino for domotics. It allows an electric appliance to be turn on or off through a web interface.

A final project represents a sound intensity detector that can be embedded in any system where it is critical to perform any action after a certain sound intensity has been produced.

For all the projects it is straightforward to embed a module to store the read data from the sensors in a database for further processing, and analytics (e.g. through data mining techniques).

More info about the background, design, building process, usage and application of the developed smart prototypes will be detailed in the next remaining sections of the chapter.

Following the open source best practices, documentation and codes of these projects have been made online for free downloading and sharing in order to further contribute to the advancement and widespread usage of the Arduino platform. In some cases, videos have been taken so that the reader can have a deeper insight of the underlying project².

BACKGROUND

Arduino was born in Italy in 2004 when a Colombian student created the development platform Wiring as his Master's thesis project whose goal was to create a low cost tool for a wide range of users for digital projects.

The first Arduino was introduced in 2005 with the goal of providing an easy way, innovative and low-cost for a wide range of users for developing devices that could interact with their environment (such as robots, thermostats, motion detectors, etc.). In 2011 over 300 thousand official Arduinos were produced and in 2013 over 700 thousand official boards were at disposal of users.

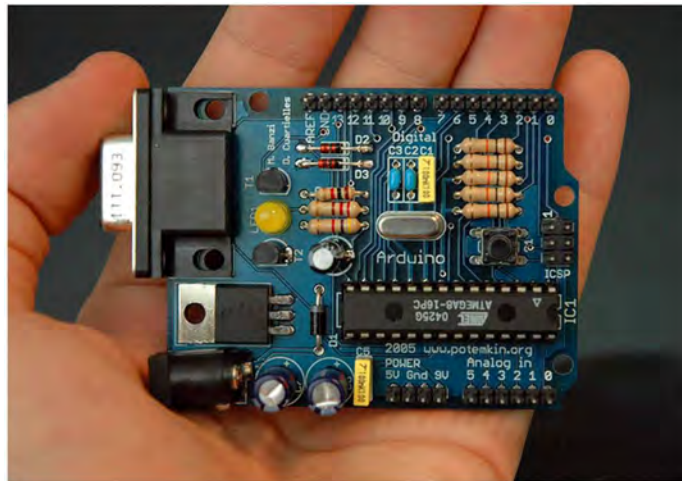
The Arduino board is provided with an Atmel 8, 16 or 32 bit AVR microcontroller with other components for the integration in other circuits. Arduino's connectors allow the connection between the CPU board to a variety of add-on modules called shields. One handy feature of Arduino's microcontroller is the boot loader it is provided that simplifies the upload of programs on the chip flash memory. Figure 1 shows an Arduino board. As far as the software is concerned, the Arduino Integrated Development Environment (IDE) supports the programming languages C and C++. Every Arduino program consists of two functions that are compiled and linked into an executable cyclic program, *setup()* and *loop()*.

The setup function is used for initialization and runs once at the start of the program whereas the loop function is called continuously until the board is turned off.

Arduino is employed in a number of projects within several disciplines and it is often used and mentioned in research papers.

One domain that is experiencing to a quick widespread of the utilization of the Arduino framework is the home control and automation. Authors in Piyare, R., (2013) presented a novel architecture for low cost and flexible home control and monitoring system using Arduino. RESTful based web services have been used for the communication between the user and the home devices. The system allows any Android based smart phone with WiFi, 3G or 4G to access the system. To note that the proposed system does not require a dedicated server PC as many other current systems but it is hosted within Arduino itself and offers a novel communication protocol. One of the directions that the authors of the mentioned paper wanted to head was the Internet of Things (IoT). It is defined as a global network that links physical objects using network communications, web applications and cloud computing technologies. Within the IoT, devices can communicate, access, store and retrieve information from/to the Internet, interact with users and create smart and pervasive environment. Arduino can easily

Figure 1. An Arduino board with an RS-232 serial communication interface (upper left) and an Atmel ATmega8 microcontroller (black, lower right)³



read data from all kinds of sensors and is well suited for IoT for its simplicity and extensibility. Authors in Doukas, C., (2012) provided details and information needed to design and create IoT applications using the Arduino platform. They described cloud computing concepts, open platforms used to store sensor data on the cloud, how to connect Arduino with an Android phone and store/retrieve data through the Internet, and how to reprogram the Arduino microcontroller remotely through the cloud. One more example of adoption of the Arduino platform for home automation is represented by the work in David, N., A. Chima, A. Ugochukwu, & E. Obinna, (2015) where the authors presented a flexible home control and environmental monitoring system. A micro web server is hosted with the Arduino and Bluetooth technology is used for the communication between the remote user and the home devices. A similar work related to Arduino for home automation that takes into account security of the home providing a safeguard from possible intruders is represented by Chattoraj, S., (2015) where several sensors (temperature, humidity, smoke) are used and controlled by Arduino that acts as a master controller.

Another domain where the adoption of Arduino is rapidly increasing is the robotics. Authors in Warren, J., J. Adams, & H. Molle, (2011) de-

scribe Arduino to control a variety of different robots and provide detailed instructions on how to build them (e.g. GPA-enabled robot, robotic lawn mower, fighting bot, etc.). They mention and show different motors used in robotics and interfaced with Arduino.

More recently, authors in Araujo, A., D. Portugal, M. S. Couceiro, & R. P. Rocha, (2013) presented a full integration of compact educational mobile robotic platforms built around Arduino controller board in the Robotic Operating System (ROS). Even though most of robotic platforms provide open source software, there are important constraints that limit their diffusion and wide spread along the utilizers: (i) they require a slow learning curve and (ii) the hardware has limited capabilities. This was the rationale behind the development of a new robotic platform on top of Arduino. One more example within the robotic domain is represented by Broccia, G., M. Livesu, & R. Scateni, (2011). Here the authors proposed a cost-effective four-wheeled surveillance robot built using Arduino and a smartphone running Android. The hardware of the smartphone has been used for the robot purposes through the API offered by Android (e.g. GPS, video-camera, etc.). The robot could be remotely controlled with a PC connected to Internet.

Arduino has also been used in Jena, S. P., S. Aman, & R. Das, (2015) for data acquisition of green house environment. Multiple sensors were used (to measure things such as temperature, humidity, CO2 gas, soil moisture), connected to the Arduino with the goal of achieving an enhancement of growth in green house.

Arduino has been employed to develop game devices and video games as well. One pretty common example is represented by the famous Pong videogame that has been designed using the Arduino board. There are different versions of it Alberti, D., (2013), Arduino. TFT Pong. A different and more ambitious project is Arduboy. Arduboy (2015) is a game system powered by Arduino that had enormous success to raise funds in Kickstarter. Arduboy is a miniature open-source game 8-bit computer that contains all of the instructions, graphics, and sound to produce the game played on the device. Users can reprogram the Arduboy and change the games on it.

DEVELOPED PROTOTYPES

In this section we will describe in detail each information system prototype developed by each group of students for computer architecture, class held at

the first year of the degree of computer science at the University of Cagliari. It is impressive to note that thanks to the simplicity and flexibility of the Arduino platform, freshmen were able to design and create systems that can easily be installed in an environment to collect data of several kinds. They might even get commercialized and/or exploited for further opportunities.

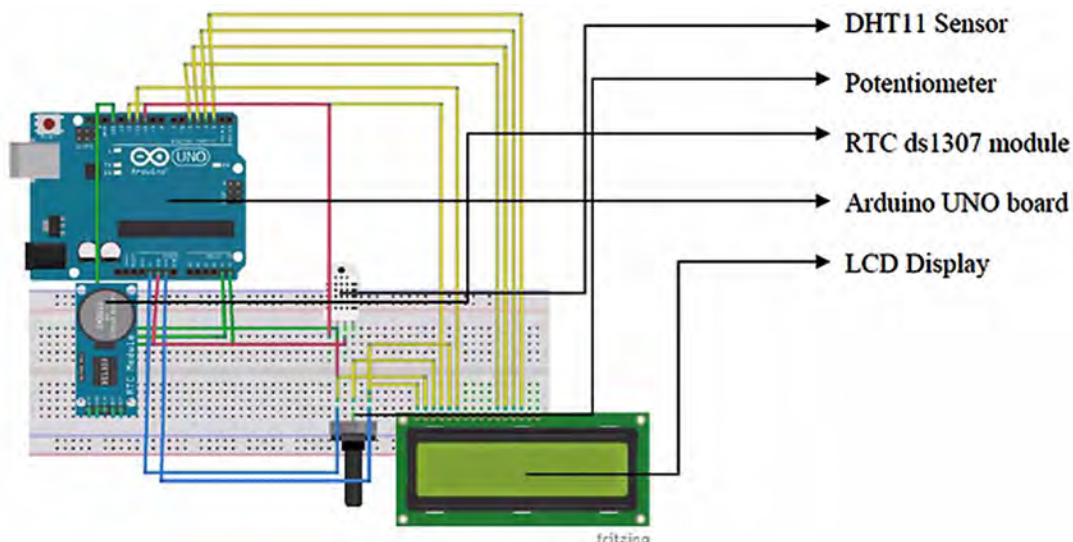
Measuring Light Intensity, Temperature and Humidity With Arduino

Two systems have been built working with these sensors. The first includes a display showing the values of light intensity, temperature and humidity. It also uses the following components:

- Arduino board UNO,
- 16x9 LCD display,
- Humidity and temperature sensor DHT11,
- RTC module for data and time, and
- A potentiometer.

Figure 2 shows the schema of the system for all the connections. Software has been divided into three parts: (i) initialization of Arduino and

Figure 2. Connections of LCD display, DHT11 sensor, potentiometer, RTC module, and Arduino UNO board using Fritzing



inclusion of all the needed libraries; (ii) check of each component when they are on; (iii) print of information on the LCD display.

The other related project does not use the LCD display but sends the value read out of the sensors to a Web server that stores them in a NoSQL database and shows the results online in a dedicated web page. For such a project we have used:

- Arduino Uno,
- Arduino Ethernet Shield,
- DHT11 sensor,
- A resistor of 10 k Ω ,
- A photo-resistor,
- An Ethernet cable.

As shown in Figure 3, and as already discussed previously, digital and analogical inputs of Arduino have been connected with the sensors. Light intensity has been shown as a 10bits number (value between 0 and 1023). For the temperature and humidity we have adopted a digital DHT11 sensor. An Ethernet board compatible with Arduino has been used in order to send read data

to a PHP page that processes GET requests and stores results in a remote database.

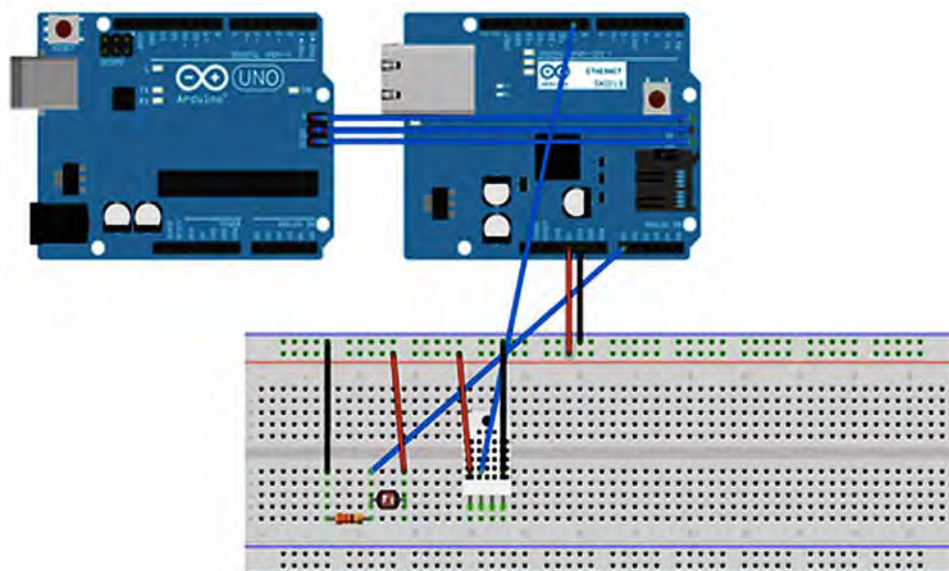
Smart Management of Electric Appliances Using LCD Display

Components that have been used for the construction of this prototype are the following:

- Arduino Uno;
- 74HC595 Shift register;
- 16x2 LCD display, wired through the use of:
 - One NPN transistor;
 - A trimmer;
- Photo resistor;
- Relay;
- 2x Pushbuttons;

Arduino sends data to be visualized on the LCD display to the shift register through the pins *INPUT (14)*, *CLOCK (11)* and *REFRESH OUTPUT (12)*. The shift register performs a translation from serial to parallel and sends data through four of the eight pins of the LCD display.

Figure 3. The schema of the light intensity, temperature and humidity sensors linked to Arduino and the Arduino Ethernet board using Fritzing



Moreover, it uses one pin of the NPN transistor to securely control the backlight of the LCD display. The reader is invited to check Figure 4 for the schema of the LCD display. Pins directly used by the shift register to send data to the LCD display are: *RS*, *E*, *R/W*, *D4-D7*, and *K* for controlling the backlight. The shift register is not essential but it has been employed to reduce the number of pins directly used on the Arduino. Without it, the Arduino would have been connected directly to the LCD using eight pins (as mentioned above). After describing the procedure how the data are sent and shown on the LCD display, let us move on with the other components. Figure 5 shows the schema of the used shift register.

The photo resistor is a light-controlled variable resistor. The resistance of a photoresistor decreases with increasing incident light intensity; in other words, it exhibits photoconductivity. It is connected to the analogic pin of Arduino with a push down resistor. Arduino reads through the pin connected to the photo resistor a value ranging from 0 to 1023, which indicates the intensity

of the light in the environment. A value closer to 0 means absence of light whereas a value closer to 1023 indicates the highest intensity. A value between 600 and 800 indicates acceptable indoor light. Just to get an idea on real settings, solar light has a value closer to 950. A default value of 650 is used as threshold of the system. When the light has an intensity value greater or equal than the threshold the system is activated, otherwise it is turned off.

The two push buttons are connected in a pull down configuration to two interrupt pins of the Arduino board. They start two interrupt routines whose purpose is to increase or decrease the threshold value for the intensity of light that enables the system.

Finally, a 5V relay is connected to a digital pin of Arduino and it is enabled when the intensity of light is below the threshold. The relay is connected to a lamp or any other electric appliance having a power in compliance with that supported by the relay.

Figure 4. The 16x2 LCD display and its pins employed for the prototype

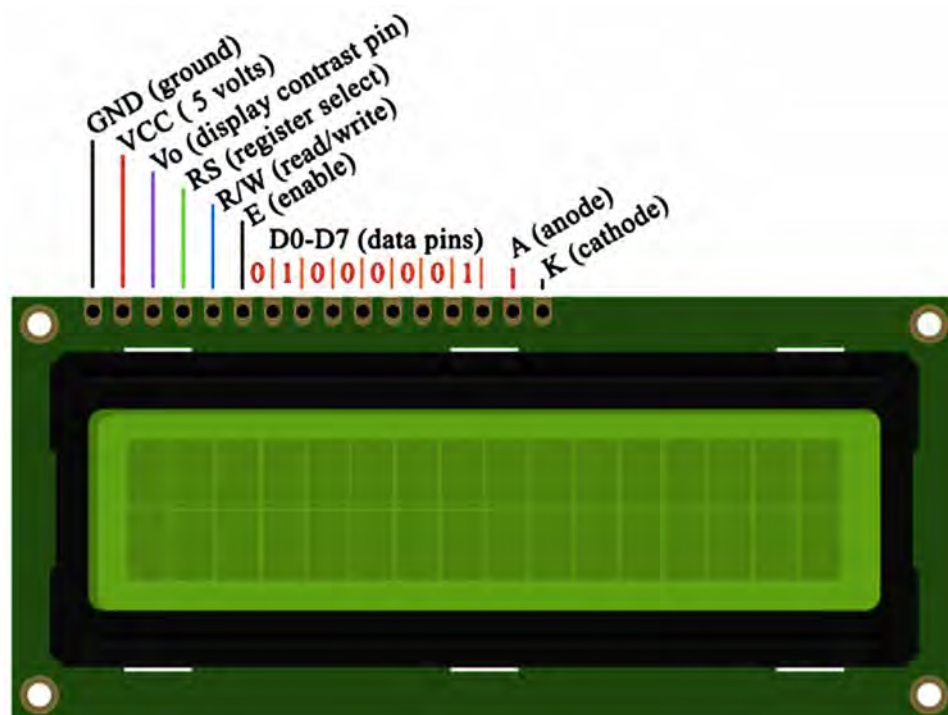
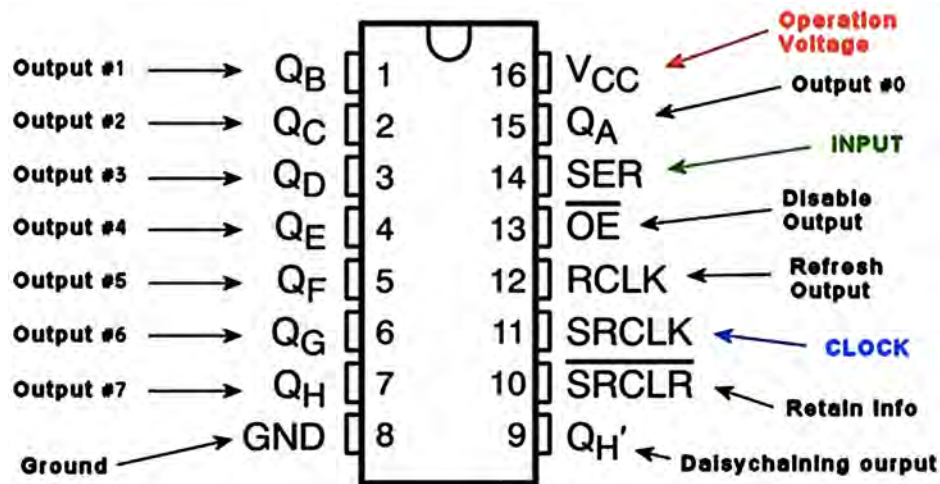


Figure 5. 595 Shift register and its schema



As far as the software is concerned, the LiquidCrystal595 Arduino Library has been employed for driving the LCD instead of the classic LiquidCrystal because of the presence of the Shift Register, this library also permits to manage the backlight of the LCD without any additional configuration.

Figure 6 shows the assembled prototype.

It is important to note that the schema presented in this subsection does not change at all if we replace the photo resistor with any other sensor (e.g. temperature, humidity, etc.) in order to control the electric appliance connected to the relay according to the temperature or humidity of the environment.

An Arcade Game Controller

The second prototype we describe consists in an arcade game controller developed to play video games in any PC. Components that we have employed for the design process were the following:

- Arduino Leonardo Microcontroller;
- Components from the Arduino Starter Kit;
- BreadBoard;
- 4 semi-transparent arcade game buttons with mechanical switch (bought on eBay);
- 4 leds (2 red from the Arduino Starter Kit and 2 white bought separately);

- A home-made wooden case we built.

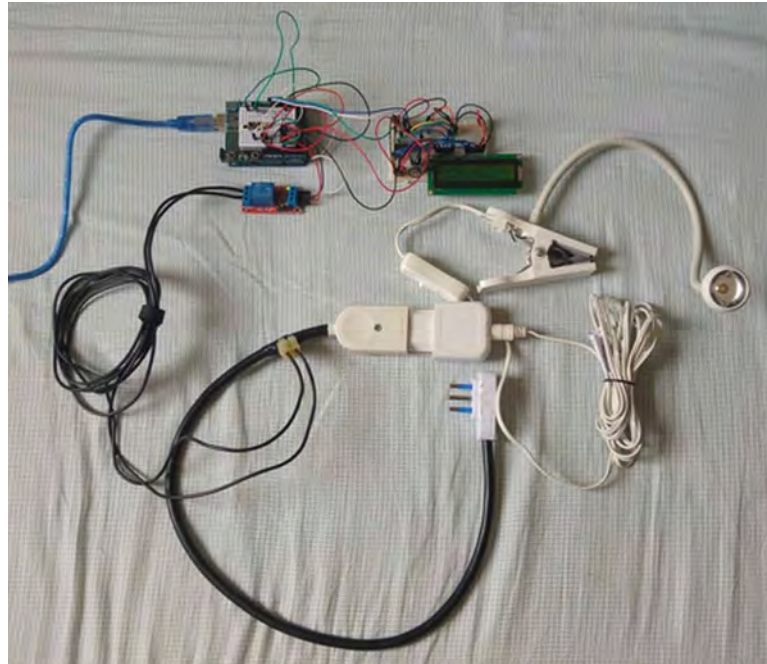
The core of the game controller lies on Arduino itself. The raw input signals are acquired from the buttons and sent to the Arduino board, then translated into keyboard signals and sent to the pc through the USB cable.

Each of the 4 buttons has 2 connectors for the switch and 2 connectors to light up the led inside each button. Whenever a button is pressed, its led is turned on. Each button is connected to the breadboard to allow a more precise circuit mapping. No additional resistors are necessary, only the built-in pull-up resistors inside the breadboard.

The Arduino board has been programmed to emulate a common plug & play USB keyboard. We have leveraged the *keyboard.h* library, in order to be compatible with every computer and OS (we have tested it on a Windows, Linux and Mac machine). We made sure the ghosting effect⁴ does not occur, and therefore all the 4 buttons can be pressed and recognized at the same time without any issues.

The code initializes (in the *setup* function) the status of the 4 pins related to the four buttons, setting them as inputs and 4 pins as output to light up the led lights. The *loop* function is executed every few milliseconds to accurately acquire any push of each button. This function checks at every

Figure 6. The assembled prototype for smart management of electric appliances

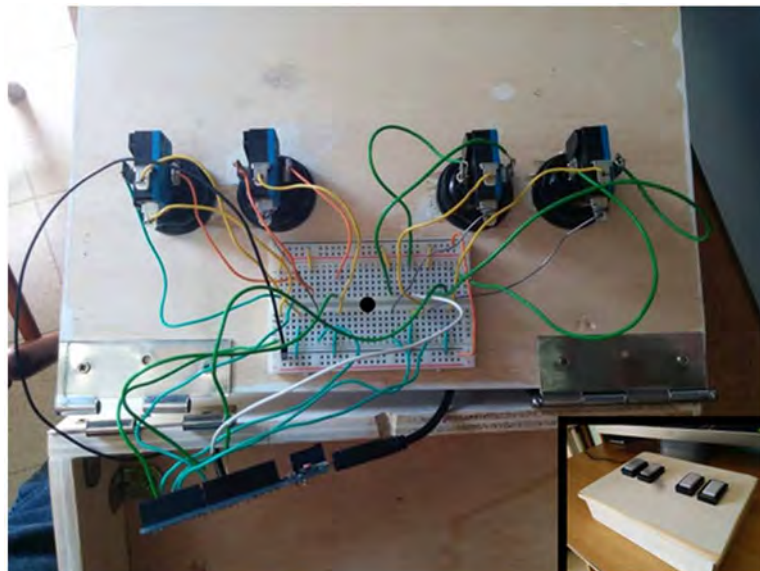


G

loop if a button is pressed; if this happens, the main board sends to the PC the corresponding keyboard signal. In the meantime the main board also sets to “high” the status of the pin associated with the button, lighting up the led inside

the corresponding button. As long as the button is pressed the board continues to send the same signal, and keeps the led on. Figure 7 shows the case and the assembled components.

Figure 7. The case (bottom right) and the assembled components of the arcade game controller prototype



A Sound Intensity Detector

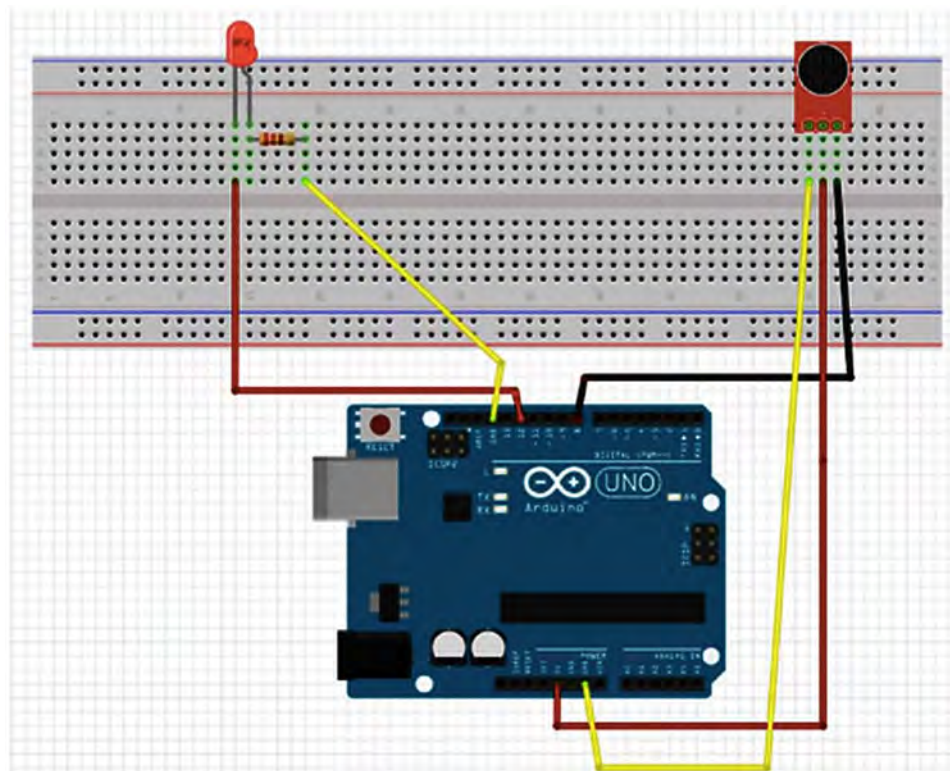
In this third prototype, Arduino has been connected with an acoustic sensor to detect sounds whose intensity is higher than a chosen threshold. If it is then a led is turned on. There are several applications of such a schema to be used in homes: e.g. baby crying detector, dog bark detector, etc. The adopted acoustic sensor has a microphone, seven resistors (one of them is variable), a LM939 module and two led lights. It is possible to set the sound intensity level to detect by tightening a screw of the embedded potentiometer. When the microphone detects a sound intensity value higher than the threshold level, a HIGH signal will be produced on the DO pin. Otherwise, a LOW signal will be propagated. Figure 8 shows the schema of the prototype (Arduino's connections with the acoustic sensor, the LED, and the resistor). More in detail, the acoustic sensor is connected to Arduino through three pins: DO, digital exit, to the pin 8.

Vcc and GND connected, respectively, to the 5V and GND pins. Once the system is connected to either USB or to a power supply it is active and ready to listen. If a sound with intensity greater than the threshold is detected, the LED is turned on for a second. For the next five seconds, the system is in pause and, right after, it comes back to active and ready to listen.

Management of Electric Appliances Through Web Services

This project aims at turning on and off electric appliances through a simple web interface. The extra modules we have used are: Arduino Ethernet Shield and the Relay module. The former allows linking an Arduino board with a PC or an Internet router therefore creating a web server in order to manage all the digital ports of Arduino. The latter includes two relays connected to a digital port of Arduino that will command to turn on or off the

Figure 8. The schema of the sound intensity detector using Fritzing⁵



electrical appliances connected to one endpoints of the relay. Other tools we have used are:

- 512MB MicroSD,
- Ethernet cable,
- 2 x relay,
- lamp,
- paper box.

A web page has been developed within the microSD of the Ethernet Shield board and other code for the Arduino main functions lets the web server wait the requests on the IP address indi-

cated in the configuration. Figure 9 shows the components used for such a project.

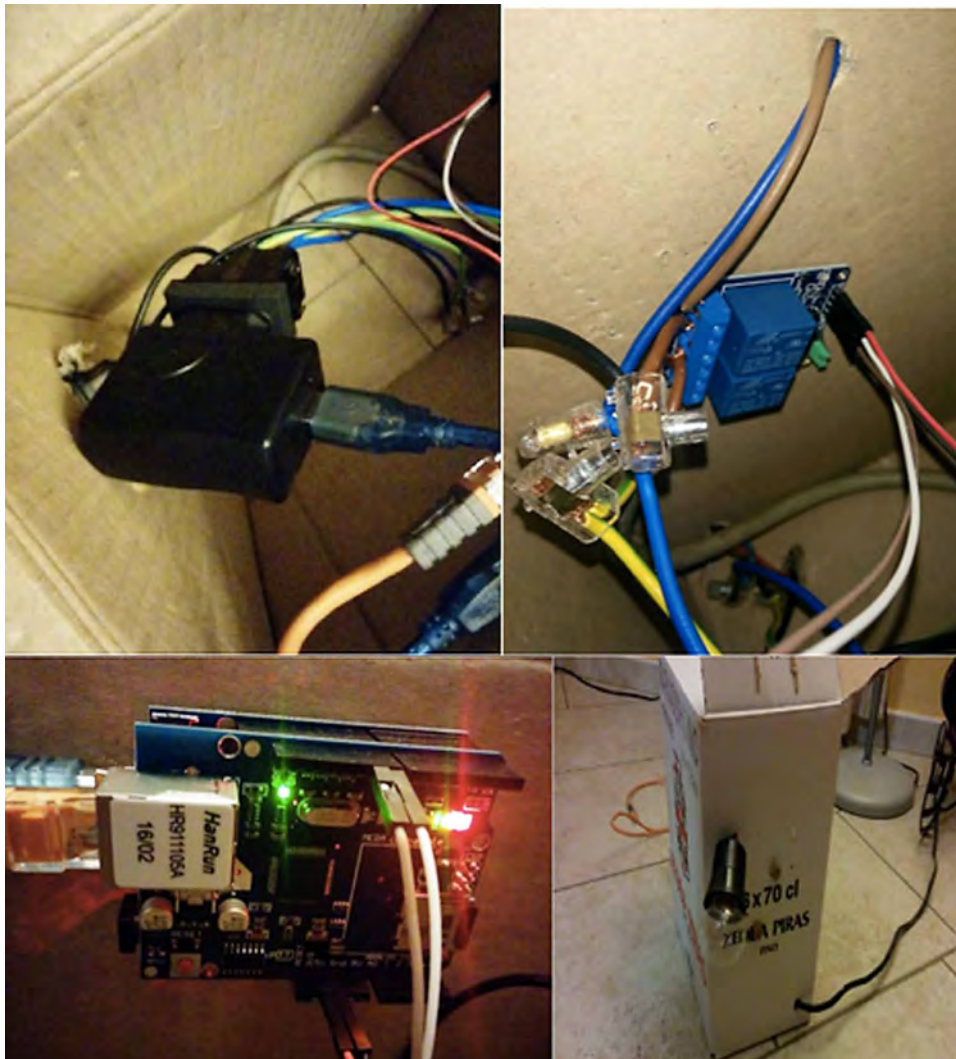
G

Analog Stick with Arduino

Components we have used for such a prototype are:

- Arduino UNO board,
- Breadboard,
- Analog stick,
- Eight 220 Ω resistors,
- Eight led lights (4 red and 4 blue), and
- A power bank as power supply.

Figure 9. Components used for the project



The rationale of the software developed on top of Arduino is to detect the x-y position of the analog stick and turning on the correspondent led light. Instead of the led lights, electric appliances might be connected and turned on or off using the stick. Figure 10 shows the connections of all the components used for this prototype where the reader may observe the top left red led light on when the user moves the stick toward the corresponding direction.

FUTURE RESEARCH DIRECTIONS

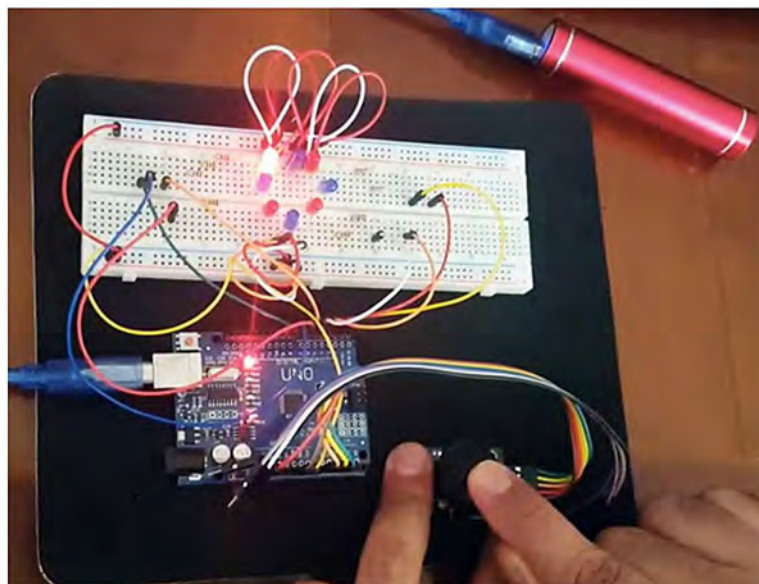
There are plenty of ideas to be implemented that stimulate us. For example, there are several smart components that can be designed, programmed and used in homes and connected with Internet and managed through smart phone applications whose data can be gathered for plenty of analysis. Definitely, the most stimulating of them is within the robotics domain. In fact, as future directions where we are headed we would like to develop on top of Arduino a robotic platform provided with several sensors (temperature, camera, humidity, acoustic), wheels to move on a flat surface, arms

to pick up objects, move and perform general. One advanced version will have a humanoid shape and will be provided of legs instead of the wheels. We want to also provide the robot with a basic artificial intelligence so that it can perform basic actions depending on the context. Last but not least, we want to come up with a set of API to access to all the sensors and functionalities of the robot so that everyone can build the same robot (we will publish the manual and data sheets of each component we will be using) and program software that can interface with all the robot capabilities. We would like to keep the project open source and follow its best practices.

CONCLUSION

Arduino is a platform that easily allows the development and integration of hardware in any domain. It can be used to easily collect data and information from any sensor and store them in cluster for further (big data) processing. In this chapter we have described a set of applications and prototypes developed on top of Arduino for different domains where data can be extracted

Figure 10. Breadboard and all the components for the analog stick built with Arduino



and analyzed for further processing. The ease of design and development of so many sensors and opportunity to extract data paves the way for start-up/spin-off opportunity creation that sell as business core products developed with Arduino or even consulting and training for the use of Arduino and its tools. We aim at creating such a start-up with skills in creating and positioning sensors, extract and store data in databases and perform analytics on them.

REFERENCES

Alberti, D. (2013). *Building Pong videogame with Arduino*. Retrieved from <http://www.danielealberti.it/2013/08/arduino-pong.html>

Araujo, A., Portugal, D., Couceiro, M. S., & Rocha, R. P. (2013). Integrating Arduino-based Educational Mobile Robots in ROS. *Proceedings of the 13th International Conference on Mobile Robots and Competitions*. doi:10.1109/Robotica.2013.6623520

Arduboy. (2015). Retrieved from <https://www.arduboy.com/>

Arduino. (n.d.). *TFT Pong*. Retrieved from <https://www.arduino.cc/en/Tutorial/TFTPong>

Broccia, G., Livesu, M., & Scateni, R. (2011). *Gesture Interaction for Robot Motion Control*. *Eurographics Italian Chapter Conference*.

Chattoraj, S. (2015). Smart Home Automation based on different sensors and Arduino as the master controller. *International Journal of Scientific and Research Publications*, 5(10).

David, N., Chima, A., Ugochukwu, A., & Obinna, E. (2015). Design of a Home Automation System Using Arduino. *International Journal of Scientific and Engineering Research*, 6(6).

Doukas, C. (2012). *Building Internet of Things with the Arduino*. CreateSpace.

Jena, S. P., S. Aman, & R. Das, (2015). Computerized Green House Data Acquisition System using Arduino with LabVIEW. *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, 4(4).

Piyare, R., (2013). Internet of Things: Ubiquitous Home Control and Monitoring System using Android based Smart Phone. *International Journal of Internet of Things*. DOI:10.5923/j.ijit.20130201.02

Warren, J., Adams, J., & Molle, H. (2011). *Arduino Robotis*. Apress. doi:10.1007/978-1-4302-3184-4

ADDITIONAL READING

20 Unbelievable Arduino Projects. (2016). <http://www.instructables.com/id/20-Unbelievable-Arduino-Projects/>

Annapurna, L. B., Mounika, K., & Chakradhara Chary, K., Roohi Afroz. (2015). Smart Security System using Arduino and Wireless Communication. *International Journal of Engineering Innovations and Research*, 4(2).

Balasubramanian, R., York, Z., Dorran, M., Biswas, A., Girgin, T., & Sankaralingam, K. (2014). *Hands-on Introduction to Computer Science at the Freshman Level*. SIGCSE '14 Proceedings of the 45th ACM technical symposium on Computer science education, 235-240. doi:10.1145/2538862.2538889

Dtto, a 3D printed modular robot with an Arduino brain (2016). <http://www.geek.com/news/dtto-is-a-3d-printed-modular-robot-with-an-arduino-brain-1658796/>

Kumar, S. (2014). Ubiquitous Smart Home System Using Android Application. [IJNCN]. *International Journal of Computer Networks & Communications*, 6(1), 33-43. doi:10.5121/ijncn.2014.6103

Lay, A., & Leung, S. Y. (2013). Mobile Bluetooth-Based Game Development Using Arduino on Android Platform. *Applied Mechanics and Materials*, 427(429), 2192–2196. doi:10.4028/www.scientific.net/AMM.427-429.2192

Lay, A., & Leung, S. Y. (2014). *Ubiquitous Mobile Game Development Using Arduino on Android Platform*. Chapter Ubiquitous Information Technologies and Applications. *Lecture Notes in Electrical Engineering*, 280, 271–276. doi:10.1007/978-3-642-41671-2_35

Mowad, M. A. E., Fathy, A., & Hafez, A. (2014). Smart Home Automated Control System Using Android Application and Microcontroller. *International Journal of Scientific & Engineering Research*, 5(5).

KEY TERMS AND DEFINITIONS

Arduino Board: Open source board provided with Atmel 8, 16 or 32 bit AVR microcontroller with other components used to program for development and embedding in other systems.

Arduboy: A credit card sized game system developed using Arduino whose cost is \$39. The original project was funded through Kickstarter.

Breadboard: It is a solderless device for prototyping of electronics and test circuit designs. Most electronic components in electronic circuits can be interconnected by inserting their leads or terminals into the holes and then making connections through wires where appropriate. The breadboard has strips of metal underneath the board and connect the holes on the top of the board.

Fritzing: Open source hardware initiative that makes electronics accessible as creative material for everyone.

Ghosting: Effect appearing in some models of keyboards when some key combinations pressed together cause some keypresses to disappear, although they are physically pressed.

Internet of Things: A development of the Internet where common objects have network connection and send and receive data.

Kickstarter: The world's largest crowdfunding platform for creative projects developed by an American corporation based in Brooklyn, New York.

NoSQL Approaches: Provide a mechanism for storage and retrieval of data without using a predefined schema as that present in relational databases. NoSQL indicates non relational.

Restful Web Services: Software architectural style of the World Wide Web that specifies constraints that if applied to a web service induce desirable properties. Data and functionality are resources accessed using Uniform Resource Identifiers.

ENDNOTES

- ¹ <https://www.kickstarter.com/>
- ² Prototypes, their descriptions and video showing how they work can be found online.
- ³ Taken from Wikipedia, <http://www.wikipedia.com>
- ⁴ <https://www.microsoft.com/appliedsciences/antighostingexplained.mspx>
- ⁵ <http://fritzing.org/home/>

Educational Serious Games Design



Ilias Karasavvidis

University of Thessaly, Greece

INTRODUCTION

Initially, digital games targeted mainly entertainment. The idea of combining fun and learning led to edutainment, the pairing of entertainment with education. The latest trend, serious games, marginalize entertainment and brings education into the spotlight. Serious games are games not exclusively designed for fun, serving non-entertainment goals in many diverse fields such as military, government, corporate, health-care, and education (Michael & Chen, 2006).

Digital games have become a defining phenomenon of contemporary culture. Over the last two decades, the educational interest in digital games has skyrocketed. This interest has taken two main forms. First, game-based learning that is entertainment-driven. This trend involves the use of commercial games for learning. Second, education-driven game-based learning that is currently manifested in trends such as gamification and serious games. The former refers to the application of game design elements to educational settings (Deterding, Khaled, Nacke, & Dixon, 2011; Kapp, 2012). The latter refers to the ad hoc development of games that bring education into the spotlight without excluding entertainment. Serious games are games not exclusively designed for fun, serving non-entertainment goals in many diverse fields such as military, government, corporate, health-care, and education (Michael & Chen, 2006).

Game design is an inherently interdisciplinary endeavor, involving experts from various disciplines such as graphic, audio, product, and interaction design, programming, animation, writing, and content area expertise (Salen, 2007). The game industry has developed highly sophisticated

narrative, artistic, and technical methodologies for creating engaging and immersive games (e.g. Salen & Zimmerman, 2004; Schell, 2014; Adams, 2010). Currently, the field of digital game design is mature, being in a paradigm state (Kuhn, 1996).

As the emphasis has gradually shifted to serious games, however, new requirements emerged. Compared to traditional digital game design, the main complication that emerges in the case of Serious Game Design (SGD) is that learning has priority over entertainment. Consequently, in addition to all other types of expertise required for digital game design, SGD necessitates professionals whose expertise is related to learning. Such professionals include educators, content experts, and learning sciences professionals in general. To design effective serious games, game design professionals would need to collaborate with learning sciences professionals (Charsky, 2010; Lim et al., 2014; El Mawas, 2014). Such a collaboration, however, might not be directly possible because a common vocabulary is missing (Arnab et al., 2014). As it has been stressed, the main limitation characterizing the field of serious game design pertains to the disconnect between established game development models and the design of learning (Arnab et al, 2014; Bellotti, Berta, De Gloria, D'ursi & Fiore, 2012; Van Staalduinen & de Freitas, 2011). To address this limitation, Moreno-Ger et al., (2014) argued that serious game development methodologies are needed that will eventually help systematize the creation of games. To bridge the game design - educational design gap, a number of serious game design models have been advanced over the past few years. To date, there has been no systematic review of such SGD models. Consequently, the extent to which these models address the major

DOI: 10.4018/978-1-5225-2255-3.ch286

challenges the field of SGD faces is unknown. The present work has two main objectives. First, it presents the state of the art in SGD by briefly introducing 13 models. Second, it acknowledges contribution these models make and identifies 5 principal challenges that remain open for the field of SGD. The chapter is concluded with an examination of the extent to which every model meets the design challenges and an outline of the road map ahead.

BACKGROUND: SERIOUS GAME DESIGN MODELS – STATE OF THE ART

In this section 13 SGD models that have been advanced in recent years are briefly introduced. An overview of the models is given in table 1.

While all models aim at SGD, they have different origins and constitute different solutions to the problem of design. For convenience, the SGD models are presented in chronological order.

Table 1. An overview of SGD models

Model	Sources	Features
Experiential Gaming Model (EGM)	Kiili (2005)	experiential learning influence; links gameplay with experiential learning
4 Dimensional Framework (4DF)	De Freitas & Oliver (2006); De Freitas & Jarvis (2009)	provides 4 dimensions for game design: learner, pedagogy, representation, and context
Game Object Model (GOM), GOM II	Amory & Seagram (2003); Amory (2007)	object-oriented programming influence; game components are described in terms of abstract (pedagogical) and concrete (design) interfaces
Document-Oriented Design and Development for Experiential Learning (DODDEL)	McMahon (2009)	extension of a generic ADDIE model comprising of the following stages: situation analysis, design proposal, design documentation, production documentation, prototype, development, and implementation
Is	Annetta (2010)	encapsulated model comprised of 6 elements: identity, immersion, interactivity, increasing complexity, informed teaching, and instructional
Game-Based Learning framework (GBLF)	Van Staalduinen & de Freitas (2011)	4DF extension with 25 game elements that improve memory and learning
Design Patterns Framework (DPF)	Kelle, Klemke & Specht (2011)	4 step procedure for mapping game design patterns on teaching and learning functions
Game Discourse Analysis (GDA)	Wouters, Oostendorp, Boonekamp & Spek (2011)	uses (a) information flow (resources needed in the game) and (b) game discourse (presentation of resources in the game) to inform game design
Six Facets Framework (SFF)	Marne, Wisdom, Huynh-Kim-Bang & Labat (2012)	6 design elements: pedagogical objectives, domain simulation, interactions with the simulation, problems and progression, decorum, conditions of use
Architecture for Representations, Games, Interactions, and Learning among Experts (ARGILE)	El Mawas (2014)	design methodology that employs Web 2.0 practices for SGs
Learning Mechanics – Game Mechanics (LM-GM)	Arnab et al., (2014); Lim et al., (2013); Lim et al., (2014)	associates ludic elements (game mechanics) with pedagogy (learning mechanics)
Cognitive Behavioral Game Design Model (CBGD)	Starks (2014)	Social Cognitive theory and Flow theory influences; combines (a) social cognitive elements and (b) multiple intelligences elements to promote flow
Activity Theory-based Model of Serious Games (ATMSG)	(Carvalho et al., 2015)	Activity Theory influence; extension of the GM-LM model that distinguishes 3 types of activity: (a) learning, (b) gaming, and (c) instructional; each activity type is further represented in terms of actions

Kiili, (2005) proposed the Experiential Gaming Model (EGM) for designing SGs. The EGM is based on Kolb's experiential learning cycle and the corresponding four stages of experiential learning. The starting point for the development of this model was the observation that the creation of educational games often fails because the emphasis on educational dimensions of games has displaced fun, resulting in unengaging games. The main idea underlying the EGM is that the link of gameplay with experiential learning will facilitate the state of flow and, eventually, lead to learning.

De Freitas & Oliver (2006) (also de Freitas & Jarvis, 2009) advanced the 4 Dimensional Framework (4DF) for the design of SGs. The framework comprises 4 main dimensions which can be employed to guide game design: learner, pedagogy, representation, and context. While this framework was initially conceptualized for classifying educational games, the authors have eventually adapted it for game design.

The Game Object Model (GOM) for SG design was proposed by Amory & Seagram (2003). The GOM attempts to describe the relationship between game elements and pedagogical dimensions of learning. The model draws on object-oriented programming (OOP) concepts such as classes, encapsulation, and inheritance. GOM II (Amory, 2007) is the updated version, aimed to enrich the initial GOM. The GOM considers educational games as consisting of components (objects) each of which is described in terms of abstract and concrete interfaces. Abstract interfaces refer to pedagogical and theoretical constructs (conceptualization phase) while concrete interfaces refer to design elements (design phase).

McMahon (2009) introduced the Document-Oriented Design and Development for Experiential Learning (DODDEL) model. DODDEL was developed for the purposes of an undergraduate course on game design and is an extension of the generic ADDIE model (i.e. Analyse, Design, Develop, Implement, and Evaluate – ADDIE). The DODDEL comprises several discrete design and development stages: situation analysis,

design proposal, design documentation, production documentation, prototype, development, and implementation.

Annetta (2010) proposed Is, a framework for serious educational game design that comprises six elements: identity, immersion, interactivity, increasing complexity, informed teaching, and instructional (all start with an I). In this framework the elements are nested, so the top element (instructional) encapsulates all others down the stack.

The Design Patterns framework (DPF) has been advanced by Kelle, Klemke & Specht (2011). This work constitutes more of a framework rather than a concrete development model. The DPF emerged in response to the need of mapping game design patterns onto educational methods. The authors draw on the idea of Design Patterns as the building blocks of a game and propose a 4 step procedure for mapping design patterns on teaching and learning functions.

The Game-Based Learning framework (GBLF) (Van Staaldin & de Freitas, 2011) was conceptualized as an extension of the aforementioned 4DF. The authors reviewed the literature and identified 25 game elements that improve memory and learning. They have then integrated these game elements into the four dimensions of the 4DF. The resulting framework explicitly incorporates learning, instruction, and assessment as core elements.

Wouters, Oostendorp, Boonekamp & Spek (2011) introduced the Game Discourse Analysis model (GDA). The authors introduce the concepts of information flow and game discourse. The former describes the information resources needed in a game while the latter pertains to how the resources can be presented in the game. The GDA model is proposed as a method for supporting designers in the game making process with respect to information flow and game discourse. The model comprises 3 main components that center on the information flow: (a) building blocks, which describe the information resources required to materialize a task in the game, (b) standardized description of the information flow, and (c) manipulation, which denotes the creation

of a game discourse through the information flow. This model utilizes constructs from cognitive psychology, such as cognitive task analysis (CTA), to explicitly inform game design.

Marne et al., (2012) developed the Six Facet Framework (SFF), a conceptual framework that allows all game design stakeholders (game experts and pedagogical experts) to speak a common language and work together in designing games. The main drive behind this framework is the communication problem between the two types of experts that are involved in SG design. The framework includes the following elements (facets) of design: (a) pedagogical objectives, (b) domain simulation, (c) interactions with the simulation, (d) problems and progression, (e) decorum, and (f) conditions of use. The authors developed a design pattern library consisting of 42 design patterns within the SFF. This model represents a systematic attempt to bridge formal game design and pedagogy design. While it is principally conceived as a means to foster interdisciplinary collaboration, the SFF is also helpful as a general purpose design guide due to the comprehensive library of design patterns.

The ARGILE model (Architecture for Representations, Games, Interactions, and Learning among Experts) (El Mawas, 2014) is mainly a methodology for the design of participatory and knowledge-intensive SGs. ARGILE constitutes a participatory architecture for the co-design of games by designers, experts, and players. This model attempts to extend Web 2.0 practices to game design practices through co-design.

The Learning Mechanics-Game Mechanics Model (LM-GM) for SG design has been described in a number of recent publications by a large international team (Arnab et al., 2014; Lim et al., 2013). The model was explicitly conceptualized so as to bridge the most challenging gap in the field: the SG design and educational design gap. The LM-GM model provides a means to relate pedagogy to ludic elements. The core elements of the LM-GM model are Serious Game Mechanics (SGMs) and the Mechanics of Learning (LM). The LM-GM model maps game mechanics onto learning mechanics in considerable detail. Interest-

ingly, Lim et al., (2014) have extended the LM-GM model by integrating narrative elements through Narrative Serious Game Mechanics (NSGM).

Starks (2014) introduced the Cognitive Behavioral Game Design Model (CBGD). The main argument behind CBGD is that game design should make more systematic use of established psychological theory. Consequently, CBGD draws on (a) Bandura's Social Cognitive Theory (SCT), (b) Gardner's Theory of Multiple Intelligences (MI) and (c) the enjoyment process. The CBGD model involves the use of social cognitive elements (i.e. knowledge, goals, outcomes, encouragement, and barriers) through the mechanisms of multiple intelligences elements (e.g. graphics, narrative, space, sound etc), so as to promote enjoyment in terms of engagement, challenge, flow, persistence, and mastery.

Lastly, The Activity Theory-based Model of Serious Games (ATMSG) (Carvalho et al., 2015) is an extension of the GM-LM model. It targets the problem of how concrete game components need to be structured to foster learning. Based on Activity Theory, the model adopts the concept of activity and its hierarchical structure and organization. Consequently, the model proposes 3 main types of activity: (a) gaming activity, (b) learning activity, and (c) instructional activity. Each activity type consists of separate actions. As Carvalho et al., (2015) argue, actions can be represented as triangles, each depicting how a subject achieves a goal using a tool. Game components are defined as actions forming subject-tool-goal triangles. Drawing upon this definition, the authors introduce a serious games taxonomy made up of three types of components: gaming, learning, and instructional ones.

FURTHER RESEARCH DIRECTIONS: PRINCIPAL CHALLENGES FOR SERIOUS GAME DESIGN

Overall, these models are clear evidence of the solid progress made over the last decade in terms of conceptualizing SGD. While each model has

it own unique merits, all the SGD models make important contributions to the field. As a whole, the models represent a rich and diverse set of solutions to the problem of SGD, helping piece the whole puzzle together. Despite progress, however, important issues remain unresolved. The core problem is how to bring two different professional communities together. More specifically, to design effective serious games, the expertise of educators, content experts, and learning experts will need to be utilized on the top of all other experts who are typically involved in game design. To advance as a field, SGD will need to facilitate the collaboration among experts from different disciplines. In practical terms, this involves bridging the gap between game design and learning design. Considering the state of the art in SGD as reflected in the models, we argue that the road map for advancing the field requires the community to address the five principal challenges outlined in Table 2.

The first challenge is how to integrate content into games. Learning is about academic content, so the design of serious games requires ways to integrate content into the games. Academic content cannot be embedded into games in the ways it is done in textbooks, i.e. in expository form. Content needs to be adapted, and embedded in the game in specific ways. Preserving fun while promoting learning seamlessly, requires intrinsic content integration into the game (Sanchez, 2011; Malone, 1981). The seamless content integration into games entails that distinguishing the learning

from the fun becomes difficult (Charsky, 2010). Preliminary evidence suggests that intrinsic integration into content might be essential for learning from games (Habgood & Ainsworth, 2011; Dickey, 2011).

The second major challenge is how to take into consideration students' ideas and conceptions of content when adapting and integrating content into serious games. For instance, as the Science education literature suggests, students have intuitive ideas about most physical phenomena (Driver, Asoko, Leach, Scott, & Mortimer, 1994; Smith, Disessa, & Roschelle 1994). Such ideas are often not in agreement with the established scientific viewpoint. As research has indicated, such intuitive conceptions influence learning and cannot be ignored (Duit, Treagust, 2003; Vosniadou, Ioannides, Dimitrakopoulou, Papademetriou, 2001).

The third challenge involves determining how to utilize instructional design theory to inform SGD. Mayer (2011) argues that to design serious games, we need to combine two fields: the science of instruction and the science of learning. As he puts it, the challenge is to find the proper mix between game and instructional features. One of the requirements mentioned is a science of instruction. The genuine challenge, however, is to translate the principles outlined in Mayer (2011; 2014) into actual game design methods since a direct translation might not be possible.

The fourth challenge is to examine how people learn (e.g. Bransford, Brown & Cocking, 2000) and utilize this body of knowledge for the purposes of SGD. Engineering learning as opposed to engineering entertainment is what differentiates games for entertainment from serious games. Therefore, learning theory and research will need to be translated into ludic approaches so as to be effectively integrated into games. Inversely, ludic approaches will need to be combined with learning theory and research. While it has been illustrated that game designers intuitively employ learning principles when creating games (Gee, 2003), a more systematic approach will have to be followed.

Table 2. Principal design challenges for the SGD field

N	Design Challenges
1	Integrate content
2	Address student ideas
3	Engineer instruction
4	Engineer learning
5	Conceptualize learning

As an extension of the previous one, the final challenge is related to the conceptualization of learning per se. While some SGD models have turned to instructional design theory for solutions, it should be noted that instruction is merely a means to an end. The end goal of all instruction is learning: instruction should simply be instrumental in facilitating it. As opposed to looking at the field of instructional design for inspiration, learning will need to be approached in a more principled way. We argue that effective SGD calls for both an empirically and theoretically grounded conceptualization of learning. Regarding the former, SGD needs to be founded on solid principles of learning. Mayer (2011) stressed the need for a science of learning for game design, specifically advocating the use of Multimedia Learning Theory (MLT) principles for informing SGD. We would extend this argument to include empirical research from the learning sciences at large. Regarding the latter, what is required is a concrete, coherent, and comprehensive conceptualization of learning for the purposes of SGD. Such an all-encompassing conceptualization will need to account for learning in terms of content, context, social interaction, and mediational artifacts. We argue that such a conceptualization will provide a much-needed framework for bridging traditional game elements with learning theory.

The extent to which the SGD models meet each of the 5 principal design challenges is presented in Table 3.

Several interesting conclusions can be drawn from the table. First, none of the models addresses all 5 challenges. However, six models (4DF, GBLF, SFF, LM-GM, CDBGD, and ATMSG) appear to be more comprehensive than others considering that they simultaneously meet several challenges. Second, the degree to which each SGD model addresses a specific challenge varies considerably: some of the models might fully address a challenge while others only tangentially. For instance, both the EGM and the ATMSG models adopt certain notions of learning but the latter approaches learning in a more principled manner. Third, no

Table 3. How SGD models meet the design challenges

Design Challenge Met	Serious Games Design Models
4	Experiential Gaming Model (EGM)
1, 3, 4	4 Dimensional Framework (4DF)
3, 4	Game Object Model (GOM), GOM II
1, 4	Document-Oriented Design and Development for Experiential Learning (DODDEL)
3, 4	Is
1, 3, 4	Game-Based Learning framework (GBLF)
3, 4	Design Patterns Framework (DPF)
3, 4	Game Discourse Analysis (GDA)
1, 3, 4	Six Facets Framework (SFF)
3, 4	Architecture for Representations, Games, Interactions, and Learning among Experts (ARGILE)
1, 3, 4	Learning Mechanics – Game Mechanics (LM-GM)
1, 3, 4	Cognitive Behavioral Game Design Model (CDBGD)
1, 3, 4	Activity Theory-based Model of Serious Games (ATMSG)

uniformity should be implied in how the different SGD models approach the same design problem. The models overcome a specific design challenge in different ways depending on their origin, goal, and overall approach. Models such as the EGM approach the issue of academic content integration in different terms compared to the LM-GM model. Fourth, the majority of the models draw rely heavily on psychology (for learning) and education (for instruction). Finally, the SGD models examined have placed virtually no importance on student ideas (2nd challenge) and conceptualization of learning (5th challenge). None of the models explicitly approaches the issue of previous student knowledge and how it might facilitate or hinder learning from the game. Similarly, all models fall short in terms of adopting a concrete, coherent, and comprehensive conceptualization of learning for ludic purposes.

CONCLUSION

As noted in the introductory section, the field of game design is in a paradigm state. A decade ago Gunter, Kenny and Vick, (2006) argued that a new game design paradigm is required for the field of SGD. Despite the considerable progress made, this conclusion still holds as the field of SGD is currently in a pre-paradigm state (Kuhn, 1996). It is imperative that the game design and the learning sciences communities join forces to bridge the game design and learning design gap. To advance the field to a paradigm state, the two communities will have to rise up to the five principal design challenges.

REFERENCES

- Adams, E. (2013). *Fundamentals of game design*. Berkeley, CA: Pearson Education.
- Amory, A. (2007). Game object model version II: A theoretical framework for educational game development. *Educational Technology Research and Development*, 55(1), 51–77. doi:10.1007/s11423-006-9001-x
- Amory, A., & Seagram, R. (2003). Educational game models: Conceptualization and evaluation. *South African Journal of Higher Education*, 17(2), 206–217.
- Annetta, L. A. (2010). The Is have it: A framework for serious educational game design. *Review of General Psychology*, 14(2), 105–112. doi:10.1037/a0018985
- Arnab, S., Lim, T., Carvalho, M. B., Bellotti, F., de Freitas, S., Louchart, S., & De Gloria, A. et al. (2014). Mapping learning and game mechanics for serious games analysis. *British Journal of Educational Technology*. doi:10.1111/bjet.12113
- Bellotti, F., Berta, R., De Gloria, A., Dursi, A., & Fiore, V. (2012). A serious game model for cultural heritage. *Journal on Computing and Cultural Heritage*, 5(4), 17. doi:10.1145/2399180.2399185
- Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn: Brain, mind, experience, and school (extended edition)*. National Academy Press.
- Carvalho, M. B., Bellotti, F., Berta, R., De Gloria, A., Sedano, C. I., Hauge, J. B., & Rauterberg, M. et al. (2015). An activity theory-based model for serious games analysis and conceptual design. *Computers & Education*, 87, 166–181. doi:10.1016/j.compedu.2015.03.023
- Charsky, D. (2010). From edutainment to serious games: A change in the use of game characteristics. *Games and Culture*, 5(2), 177–198. doi:10.1177/1555412009354727
- de Freitas, S., & Jarvis, S. (2009). Towards a development approach to serious games. In T. Connolly, M. Stansfield, & L. Boyle (Eds.), *Games-based learning advancements for multi-sensory human computer interfaces: techniques and effective practices* (pp. 215–231). London: Information Science Reference. doi:10.4018/978-1-60566-360-9.ch013
- de Freitas, S., & Oliver, M. (2006). How can exploratory learning with games and simulations within the curriculum be most effectively evaluated? *Computers & Education*, 46(3), 249–264. doi:10.1016/j.compedu.2005.11.007
- Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: defining gamification. In *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments* (pp. 9–15). ACM. doi:10.1145/2181037.2181040
- Dickey, M. D. (2011). Murder on Grimm Isle: The Design of a Game-Based Environment. In S. De Freitas, & P. Maharg, (Eds.), *Digital games and learning* (pp. 129–151). Continuum.
- Driver, R., Asoko, H., Leach, J., Scott, P., & Mortimer, E. (1994). Constructing scientific knowledge in the classroom. *Educational Researcher*, 23(7), 5–12. doi:10.3102/0013189X023007005

- Duit, R., & Treagust, D. F. (2003). Conceptual change: A powerful framework for improving science teaching and learning. *International Journal of Science Education*, 25(6), 671–688. doi:10.1080/09500690305016
- ElMawas, N. (2014). Designing learning scenarios for serious games with ARGILE. *Knowledge Management & E-Learning: An International Journal*, 6(3), 227–249.
- Gee, J. P. (2003). *What video games have to teach us about language and literacy*. Palgrave Macmillan.
- Gunter, G. A., Kenny, R. F., & Vick, E. H. (2006). A case for a formal design paradigm for serious games. *The Journal of the International Digital Media and Arts Association*, 3(1), 93–105.
- Habgood, M. J., & Ainsworth, S. E. (2011). Motivating children to learn effectively: Exploring the value of intrinsic integration in educational games. *Journal of the Learning Sciences*, 20(2), 169–206. doi:10.1080/10508406.2010.508029
- Kapp, K. M. (2012). *The gamification of learning and instruction: game-based methods and strategies for training and education*. John Wiley & Sons.
- Kelle, S., Klemke, R., & Specht, M. (2011). Design patterns for learning games. *International Journal of Technology Enhanced Learning*, 3(6), 555–569. doi:10.1504/IJTEL.2011.045452
- Kiili, K. (2005). Digital game-based learning: Towards an experiential gaming model. *The Internet and Higher Education*, 8(1), 13–24. doi:10.1016/j.iheduc.2004.12.001
- Kuhn, T. S. (1996). *The structure of scientific revolutions* (3rd ed.). University of Chicago press. doi:10.7208/chicago/9780226458106.001.0001
- Lim, T., Louchart, S., Suttie, N., Hauge, J. B., Stanescu, I. A., Ortiz, I. M.,... Berta, R. (2014). Narrative Serious Game Mechanics (NSGM)—Insights into the Narrative-Pedagogical Mechanism. In *Games for Training, Education, Health and Sports* (pp. 23–34). Springer International Publishing.
- Lim, T., Louchart, S., Suttie, N., Ritchie, J. M., Aylett, R. S., Stănescu, I. A., & Moreno-Ger, P. et al. (2013). Strategies for effective digital games development and implementation. In Y. Baek & N. Whetton (Eds.), *Cases on digital game-based learning: methods, models, and strategies* (pp. 168–198). Hershey, PA: Information Science Reference; doi:10.4018/978-1-4666-2848-9.ch010
- Malone, T. W. (1981). Toward a theory of intrinsically motivating instruction. *Cognitive Science*, 4(4), 333–369. doi:10.1207/s15516709cog0504_2
- Marne, B., Wisdom, J., Huynh-Kim-Bang, B., & Labat, J. M. (2012). *The six facets of serious game design: a methodology enhanced by our design pattern library*. In *21st Century Learning for 21st Century Skills* (pp. 208–221). Springer Berlin Heidelberg.
- Mayer, R. E. (2011). Multimedia learning and games. In S. Tobias & J.D. Fletcher (Eds.), *Computer Games and Instruction* (pp. 281–305). Information Age Publishing.
- Mayer, R. E. (2014). Multimedia instruction. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of Research on Educational Communications and Technology* (4th ed.; pp. 385–399). Springer. doi:10.1007/978-1-4614-3185-5_31
- McMahon, M. T. J. (2009). The DODDEL Model: A Flexible Document-Oriented Model for the design of Serious Games. In T. Connolly, M. Stansfield, & L. Boyle (Eds.), *Games-Based Learning Advancements for Multi-Sensory Human Computer Interfaces: Techniques and effective approaches*. Hershey, PA: Information Science Reference. doi:10.4018/978-1-60566-360-9.ch007

Michael, D., & Chen, S. (2006). *Serious Games: Games That Educate, Train and Inform*. Boston: Thomson.

Moreno-Ger, P., Martinez-Ortiz, I., Freire, M., Manero, B., & Fernandez-Manjon, B. (2014, October). Serious games: A journey from research to application. In *Frontiers in Education Conference (FIE), 2014 IEEE* (pp. 1-4). IEEE. doi:10.1109/FIE.2014.7044052

Salen, K. (2007). Gaming literacies: A game design study in action. *Journal of Educational Multimedia and Hypermedia*, 16(3), 301–322.

Salen, K., & Zimmerman, E. (2004). *Rules of play: Game design fundamentals*. MIT Press.

Sanchez, E. (2011). *Key criteria for Game Design. A Framework. MEET Project*. European Commission.

Schell, J. (2014). *The Art of Game Design. A Book of Lenses*. Elsevier. doi:10.1201/b17723

Sicart, M. (2008). Defining game mechanics. *Game Studies*, 8(2), 1–14.

Smith, J. P. III, Disessa, A. A., & Roschelle, J. (1994). Misconceptions reconceived: A constructivist analysis of knowledge in transition. *Journal of the Learning Sciences*, 3(2), 115–163. doi:10.1207/s15327809jls0302_1

Starks, K. (2014). Cognitive behavioral game design: A unified model for designing serious games. *Frontiers in Psychology*, 5. PMID:24550858

Van Staalduinen, J. P., & de Freitas, S. (2011). A Game-Based Learning Framework: Linking Game Design and Learning. In M. S. Khine (Ed.), *Learning to play: exploring the future of education with video games* (pp. 29–53). Peter Lang.

Vosniadou, S., Ioannides, C., Dimitrakopoulou, A., & Papademetriou, E. (2001). Designing learning environments to promote conceptual change in science. *Learning and Instruction*, 11(4), 381–419. doi:10.1016/S0959-4752(00)00038-4

Wouters, P., Oostendorp, H. V., Boonekamp, R., & Spek, E. V. D. (2011). The role of Game Discourse Analysis and curiosity in creating engaging and effective serious games by implementing a back story and foreshadowing. *Interacting with Computers*, 23(4), 329–336. doi:10.1016/j.intcom.2011.05.001

KEY TERMS AND DEFINITIONS

Digital Game: A game whose system is (partly or wholly) implemented in a digital device such as a game console, a personal computer, a smart-phone or tablet.

Game: A system in which participants compete to achieve a clearly identifiable result. The system is made up of interrelated components whose relations are rule-governed.

Game Design: The art and science involved in making all the necessary decisions for creating a game.

Ludic Elements: Game components that refer to player engagement and represent the entertainment dimension of a game.

Pedagogical Elements: Game components refer to the curricular, instructional and learning dimensions of a serious game.

Serious Game: A game that is not exclusively designed for fun, serving primarily non-entertainment goals in various fields such as education, health-care, corporate etc.

Serious Game Design: The art and science of making all the necessary decisions for creating a game that does not foreground entertainment.

Serious Game Design Models: Conceptualizations that aim to inform the decision-making processes involved in creating a serious game.

Exposure to Video Games and Decision Making

Giuseppe Curcio

University of L'Aquila, Italy

Sara Peracchia

University of L'Aquila, Italy

INTRODUCTION

In the last decade, the playing of video games (VGs) has become very popular among people. Video games represent a pervasive leisure activity beginning in middle childhood and continuing through adulthood (Gentile et al., 2004; Kubitzki, 2005). Population based surveys indicate that average gaming time ranges between 7 and 13 hs per week in both children and adolescents (Gentile and Anderson, 2003), and this value may underestimate the prevalence of use in some population segments. This high level of VGs consumption highlights the relevance for a clearer understanding of the potential influences of video game experience on human behaviour and cognition (Bioulac et al., 2008; Green and Bavelier, 2006).

Indeed, after several dramatic and murderous shoot-outs happened mainly in schools and colleges, game research focused mostly on the impact of aggressive shooter games on aggression-related cognitions, affects, and behaviors (Anderson and Bushman, 2001). The studies present in the literature showed different risks of excessive exposure to VGs: increase in aggression (Anderson et al., 2010), emergence of attention problem (Swing et al., 2010) and hyperactivity (Gupta et al., 1994), poor academic performance (Rideout et al., 2007), possible addiction (King et al., 2011), mood troubles as depression and anxiety (Mentzoni et

al., 2011), reduction of empathy (Bartholomew et al., 2005), impairment of social behavior (Gentile et al., 2011), reduction of sleep time, quality and efficiency (Weaver et al., 2010; King et al., 2013).

However, exposure to VGs can not be regarded only as a negative experience. A great body of literature has revealed that action video game players, compared to non video game players, can develop broader cognitive benefits from extensive playing. These benefits include visual acuity (Green and Bavelier, 2007; Wu and Spance, 2013; Granic et al., 2014), attention flexibility (Green and Bavelier, 2003; Cain et al., 2012), stimulus-response mapping (Clark et al., 1987; Castel et al., 2005), encoding speed (Wilms et al., 2013), and executive functioning (Strobach et al., 2012). Extensive experience playing action VGs can even affect memory for the stimuli presented in a very short period (e.g., iconic memory and visual working memory), resulting in better accuracy (Boot et al., 2008; Blacker and Curby, 2013), higher precision (Sungur and Boduroglu, 2012) and more efficient strategy in retrieving information (Clark et al., 2011).

At the present, only a few studies have investigated the potential effects of video game exposure on decision making. The aim of the present chapter is to describe this relation, reviewing published studies and discussing possible implications for future research.

BACKGROUND

1. Exposure to Video Game and Decision Making

The study of the VGs effects on decision-making is a new research field in psychology, with a limited number of published studies; nevertheless, it can offer important clues for understanding risks and potentialities.

Past research has demonstrated that VGs experience can influence cognition and emotion (West and Bailey, 2013). More specifically, the prolonged exposition to VG is associated with decreased use of proactive cognitive control (Kronenberger et al., 2005; Mathews et al., 2005; Bailey et al., 2010), changes in feeling and expressing both positive and negative affects (Bartholow et al., 2006; Kirsh and Mounts, 2007; Bailey et al., 2011). Since it is well known that the efficacy of decision making is modulated by emotion, executive/cognitive control, and by presence of chemical and behavioural addiction (Weber and Johnson, 2009; Figner and Weber, 2011), one could expect that VGs experience could have a detrimental effect on the efficacy of this complex process.

It has been demonstrated that exposure to racing VGs can influence real-world decision making related to driving behavior (Fischer et al., 2009; Beullens et al., 2011). There is evidence, however, that certain types of VGs may have differential effects on cognitive control, a set of abilities that allow the individual to maintain goal-directed information processing (Basak et al., 2008; Bailey et al., 2010). For example, in a study focused on individual differences (Bailey et al., 2010) it has been reported that experience with First Person Shooter (FPS) video games was correlated with a reduction in proactive control (active, sustained maintenance of goal-relevant information) and was not correlated with reactive control (just-in-time mobilization of control after the conflict is detected; Braver, 2012). Further-

more, Swing (2012) demonstrated that 10 hs of FPS experience resulted in a reduction in the use of proactive control in a training study. These findings may indicate that FPS gamers may be more likely to make their decisions immediately rather than after thoughtful deliberation, a tendency that could indicate a preference for immediate rewards rather than long-term assessment of the risks and benefits (Bailey et al., 2013).

In contrast to FPS games, strategy VGs may promote an increase in careful planning and executive control of behavior. Basak and coworkers (2008) demonstrated that more than 23 hs of training with strategy VGs improved task-switching and working memory abilities, that are cognitive processes relevant for the efficacy of decision making.

A study proposed by Bailey and colleagues (2013) examined the relationship between two of the most popular genres of video games (FPS and strategy) and decision making. The main purpose was to provide a more comprehensive understanding of how VGs experience is related to risky decision making. By measuring both behavioral and self-reported impulsivity, it emerged that pathological gaming and playing FPS games were positively associated with a greater impulsivity. This association sounds very interesting, supporting the evidence that this form of gaming is also associated with a reduction in the use of proactive cognitive control (Bailey, 2009; Bailey et al., 2010; Swing, 2012). Taken together these results may indicate that playing FPS games and pathological gaming are associated with an increase in impulsive behavior that results from a decrease in the use of proactive cognitive control of behavior (Bailey et al., 2013). Also, the association between gaming and risky decisions was sensitive to game genre: the number of hours spent to playing VGs and to be a FPS gamer negatively correlated with the tendency to select low risk options. These findings provide a clear evidence that gaming time, pathology, and game genre

(i.e., FPS) influence the individual's selection of risky options, and that such a behavior continues in spite of its detrimental effect on performance.

In contrast to FPS type, strategy VGs were not strongly related to increased risk-taking. One explanation for the differential influence of game genre is that there could be social repercussions for making impulsive decisions in a strategy game, where a successful performance requires cooperation within a team (Bailey et al., 2013). It is important to note, in fact, that both strategy and FPS games are associated with pathological gaming, but the consequences of impulsivity and risk-taking are not the same for the two genres. This may be due to the structure of the gaming environment or to player's goals within the different genres.

It has also provided an preliminary evidence of a negative association between playing violent VGs and prosocial behavior (Anderson & Bushman, 2001; Ballard & Lineberger, 1999; Wiegman & VanSchie, 1998). According to the General Aggression Model (GAM; Anderson & Bushman, 2002), pro- and anti- social behaviors may be affected by the activation of aggression-related patterns that are stored in memory after exposure to VGs. Such patterns can alter the way in which subsequent information is processed, so that even neutral or ambiguous cues may be interpreted as threatening or aggressive. Consequently, the activation of aggression-related patterns may alter appraisal and decision-making processes, ultimately leading to more antisocial (and, conversely, fewer prosocial) behaviors. Based on this idea, Sheese and Graziano (2005) carried out an experiment aiming at investigating the effect of VGs violence on cooperative decision making. Participants were divided in pair and randomly assigned to play either a violent or a nonviolent version of a famous VG. Following the VG exposure, participants were separated and had the opportunity to choose to cooperate with their partner for mutual gain, withdraw from the interaction, or exploit their partner for their own benefit. Results showed that playing violent

VGs affects subsequent decisions to cooperate or compete with other individuals in high-stakes situations, undermining prosocial and altruistic motivation by promoting competitive behavior in deliberate decision making (Sheese and Graziano, 2005). More specifically, participants to the violent condition were significantly more likely to choose to exploit their partners compared to those of the non-violent condition. Moreover, they were more convinced that their partners would exploit or distrust them. Instead, participants in both conditions generally expected that their partners would trust them and cooperate with them. However, participants in the violent condition were more likely to exploit than trust and choose to defect compared to those in the non-violent condition. These findings suggest that playing violent VGs may undermine prosocial motivation and promote exploitive behavior in social interactions.

Recent studies examined the influence of VG on self-perception, a process termed *identity simulation* (Fischer et al., 2010; Hull et al., 2012). In these studies, it was investigated how adoption of moral frameworks through identity simulation can impact on subsequent behaviors. By means of the *identity simulation*, players embrace the characters' decisions and behaviors as their own and a game player can adopt the traits' and attitudes of the controlled character, and this happens also outside virtual reality (Kaufman et al., 2012). Thus, moral decisions made while videogaming, could influence players' views of themselves and their morality, and this self-perception could impact real moral behaviors. The study of Ellithorpe and coworkers (2015), examined how people make and interpret moral behaviors in VGs, and how gaming behaviors influence real-world behavior. Participants made a moral decision in a VG, followed by two behavioral tasks. They were instructed to take on a deontological (save as many lives as possible) or utilitarian (win against the enemy, at any cost) mind-set during gameplay. The game asked participants to choose between saving one person or allowing him to suffer for the greater good, in this case, aiding the war ef-

fort. Immediately after, they played a game with another participant (actually a computer, in a task called Noise Blast; Bartholow et al., 2006; Bushman, 1995): whoever was faster would be rewarded with money, whoever was slower would hear an uncomfortable blast of noise. Participants set the noise and reward levels for their “partner,” and their “partner” set levels for them. In the post-test participants indicated the decision they perceived as more moral. Results showed that players tended to see the deontological option as the more “moral”. More specifically, participants who had implemented the deontological choice in the game may have felt more strongly that they had the moral high ground, so they were more likely to retaliate against another participant who had blasted them with noise by reducing the reward the other participant received. On the other hand, participants who had implemented the utilitarian choice shifted their view of this behavior to be slightly more moral, these participants were less likely to retaliate by reducing reward levels. Although it is believed that participants internalized the moral mind-set from the game, participants may have felt the need to justify their in game actions. Importantly, either explanation involves the character’s behavior being internalized, and the participant acting in a manner that is consistent with moral license.

These results indicate that people may see identity simulation through moral behavior in video games as a sufficient way to fulfill their moral quota (Ellithorpe et al., 2015). If mediated behavior acts as a sufficient substitute for actual behavior, then moral license processes may occur more frequently than previously expected. Thus, the adoption of different mind-sets can impact players’ real-world behavior. These changes provide evidence for moral license (Mikhail, 2007) as a theoretically important perspective in VG and moral behavior research, with identity simulation as one possible mechanism (Fischer et al., 2010; Hull et al., 2012). The effects of violent VG with simulation of moral behaviors is therefore not cut-and-dried, but instead a complex process the

outcomes of which depend on gameplay experiences and perceptions.

2. Pathological Gaming and Decision Making

Pathological gaming in adolescents and young adults is an emerging problem in developed societies, being a direct consequence of the rapid escalation of technological advances.

Some studies have shown that pathological gaming might overlap other behavioural and substance addictions. In fact, in healthy volunteers playing to VGs is usually associated with greater ventral striatal presynaptic dopamine release, suggesting that videogaming is potentially rewarding or motivating in itself (Koepp et al., 1998). Adolescents that can be described as frequent VGs players present greater volume in left striatal grey matter, a region that also had greater activity during loss feedback and that negatively correlate with deliberation time at the Cambridge Gamble Task (Kuhn et al., 2011). Following a 6 week extended gaming exposure of healthy volunteers, an increased orbitofrontal and anterior cingulate activity has been observed, suggesting that VGs playing can act as a reinforce and associated cues can become conditioned reinforcers (Han et al., 2010). Similarly, subjects with pathological gaming have a greater cognitive bias and cue reactivity towards game-related images, with greater medial prefrontal and anterior cingulated cortex activity (van Holst et al., 2012; Zhou et al., 2012; Lorenz et al., 2013). Pathological gaming is also associated with greater impulsivity on the Barratt’s Impulsiveness Scale and greater perseveration on the Wisconsin Card Sorting Test along with increased volume in thalamus and inferior temporal and occipital gyrus (Han et al., 2012). Moreover, pathological gamers further demonstrated an impaired motor response inhibition as assessed by Go/No Go tasks, along with decreased error-related negativity (Littel et al., 2012). These results reveal that some mechanisms of behavioral and substances addiction are common to both addiction and gambling.

Another feature that we can observe in both substance/behavioural addictions and game dependence is the impulsivity, that can be differentiated into decision and motor impulsivity (Robbins et al., 2012; Perry and Carrol, 2008). Decision impulsivity includes impulsive choice (a preference for an immediate smaller reward over a larger delayed reward; Kirby et al., 1999) and reflection impulsivity (the tendency to gather and appraise information prior to making a decision; Clark et al., 2006). Motor impulsivity includes motor response inhibition (Aron et al., 2003) and premature responding, or also anticipatory responding (Voon et al., 2013).

On the basis of these evidences, Irvine and colleagues (2013) investigated the different subtypes of decisional and motor impulsivity in a well-defined pathological gaming cohort, hypothesizing that VGs would be associated with greater decision impulsivity, both in terms of delay discounting and reflection impulsivity (Irvine et al., 2013). Results showed that pathological gaming is actually associated with greater decision impulsivity, with less evidence sampled prior to a decision and greater impulsive choice. Increased reflection impulsivity (e.g. sampling less evidence or opening fewer boxes prior to making a decision) had negative consequences with pathological gamers obtaining significantly fewer points, possibly mediated via the numeric increase in number of sampling errors (Irvine et al., 2013). From these results, it could be concluded that pathological gamers might be less sensitive to the introduction of cost or penalties or may be more impaired in the integration of decision cost in optimizing final outcomes; this suggests that pathological gamers may be more likely to respond to instrumental reward feedback and less to negative costs in decision making (Irvine et al., 2013).

Also, according to some researchers, pathological gambling players might suffer from more discrete cognitive biases of probability estimation and choices that are not specific to pathological gambling, but are exacerbated by the disorder. In fact, cognitive biases influence decision-making,

resulting in an increased risk willingness when winning probabilities are seen as high, and in risk aversion when winning probabilities are seen as low. Goodie (2003, 2005) and Lakey and co-workers (2006, 2007) observed that pathological gamblers were more confident in their decisions about a general knowledge task, even though they had no greater competence or performance level than non-gambling controls. Problem and pathological gamblers were also more likely to accept lower probability gambles, which suggests a higher degree of cognitive bias.

However it would appear that the condition is not created in experienced poker players which have a significantly lower bias in estimating the winning probability of hands and played hands with a significantly higher average winning probability compared with inexperienced players (Linnet et al. 2010). Linnet and colleagues (2012) reported that pathological gambling poker players had a larger error margin of probability estimation, played hands with lower winning probability, and had poorer differentiation of winning probability between played and folded hands compared to experienced poker players. Pathological gambling poker players and inexperienced poker players did not differ on these measures. These results suggest that cognitive biases in pathological gambling poker players cannot simply be explained as a problem of decision-making, that pathological gambling poker players have intact probability estimation, but are drawn toward risky gambles (Linnet et al., 2012).

FUTURE RESEARCH DIRECTION

On the basis of this brief analysis of the limited existing literature, we could say that the exposure to VGs may results in changes of individuals' decision making.

The results reported and discussed in this chapter necessarily require further investigation, studying these topics in more detail and with greater and balanced sample.

A well known limit, indeed, is related to the fact that almost all investigation have been carried out on men. Moreover, it would also be appropriate to conduct longitudinal studies, in order to verify if the observed results can be intended as permanent or if they are only acute and limited in time. Furthermore, this field of research would strongly gain from new investigations aimed at clarifying whether decision making is equally influenced in pathological gamblers and “healthy” VGs players. Also, additional research should be oriented on possible differential effects of different genres of VGs on decision making: it is, in fact, arguable that action and non-action VGs might differentially affect cognitive processes, an interesting issue not still directly tested. As a possible improvement, also differential effects between different kinds of games (strategy, sport, life simulation, etc) merit to be investigated.

In conclusion, this brief review underlines the relevance of new research to better understand how VGs in general and gambling in particular can influence decision making and for clarifying the mechanisms underlying such changes.

CONCLUSION

The study of the VGs exposure on decision making is a new field of research, currently characterized by few investigations. With this mini-review, we aimed at providing different considerations about the role of VGs on this higher order cognitive process.

Video games seem to influence impulsive decisions, both in subject with pathological gambling (Bailey et al., 2013) and healthy sample (Irvine et al., 2013). Experienced or pathological players tend to decide more impulsively, prefer smaller and immediate rewards and spend less time to analyze elements available before making a choice. Gamblers showed a lower sensitivity to negative reward or punishment (Bailey et al., 2013) that should discourage these behaviors.

Bailey and colleagues (2013) were the first to make a distinction between FPS and strategic video games, showing that strategic video game players are led likely to take risky decision and are more susceptible to negative feedback compared FPS players. These results, together with other studies (Fischer et al., 2009; Beullens et al., 2011) have highlighted the need to begin to distinguish various types of video games and focus on different effects that they can induce.

Several studies have shown that VGs have an important role in the moral decision, in particular the work by Ellithorpe and colleagues (2013) showed that players can identify themselves with the character's choices who plays in the VGs and that this choice influence the subject' behaviour out of the virtual world as a result of moral license (Merritt et al., 2010; Monin e Miller; 2001). This is an important result because prove that decisions taken in virtual context can modify the choice that we will take in the everyday life, well outside the game.

Finally, regarding the influence of violent VGs on competitive or cooperative behavior it has been seen that violent content can evoke the same type of behaviour towards a partner and lead to the decision to compete instead cooperate though without affecting the trust and the relationship that the subject puts in partner (Sheese e Graziano, 2005).

REFERENCES

- Anderson, C. A., & Bushman, B. J. (2001). Effects of violent video games on aggressive behavior, aggressive cognition, aggressive affect, physiological arousal, and prosocial behavior: A meta-analytic review of the scientific literature. *Psychological Science*, 12(5), 353–359. doi:10.1111/1467-9280.00366 PMID:11554666
- Anderson, C. A., & Bushman, B. J. (2002). Human aggression. *Annual Review of Psychology*, 53(1), 27–51. doi:10.1146/annurev.psych.53.100901.135231 PMID:11752478

- Anderson, C. A., Shibuya, A., Ihori, N., Swing, E. L., Bushman, B. J., Sakamoto, A., & Saleem, M. et al. (2010). Violent video game effects on aggression, empathy, and prosocial behavior in Eastern and Western countries: A meta-analytic review. *Psychological Bulletin*, 136(2), 151–173. doi:10.1037/a0018251 PMID:20192553
- Aron, A. R., Fletcher, P. C., Bullmore, E. T., Sahakian, B. J., & Robbins, T. W. (2003). Stop signal inhibition disrupted by damage to right inferior frontal gyrus in humans. *Nature Neuroscience*, 6(2), 115–116. doi:10.1038/nn1003 PMID:12536210
- Bailey, K. M. (2009). *Individual differences in video game experience: Cognitive control, affective processing, and visuospatial processing* (Unpublished Master's Thesis). Iowa State University.
- Bailey, K. M., West, R., & Anderson, C. A. (2010). A negative association between video game experience and proactive cognitive control. *Psychophysiology*, 47(1), 34–42. doi:10.1111/j.1469-8986.2009.00925.x PMID:19818048
- Bailey, K. M., West, R., & Anderson, C. A. (2011). The association between chronic exposure to video game violence and affective picture processing: An ERP study. *Cognitive, Affective & Behavioral Neuroscience*, 11(2), 259–276. doi:10.3758/s13415-011-0029-y PMID:21461985
- Bailey, K. M., West, R., & Kuffel, J. (2013). What would my avatar do? Gaming, pathology, and risky decision making. *Frontiers in Psychology*, 10(4).
- Ballard, M. E., & Lineberger, R. (1999). Video game violence and confederate gender: Effects on reward and punishment given by college males. *Sex Roles*, 41(7/8), 541–558. doi:10.1023/A:1018843304606
- Bartholomew, B. D., Bushman, B. J., & Sestir, M. A. (2006). Chronic violent video game exposure and desensitization to violence: Behavioral and event-related brain potential data. *Journal of Experimental Social Psychology*, 42(4), 532–539. doi:10.1016/j.jesp.2005.08.006
- Bartholomew, B. D., Sestir, M. A., & Davis, E. B. (2005). Correlates and consequences of exposure to video game violence: Hostile personality, empathy, and aggressive behavior. *Psychological Bulletin*, 31, 1573–1586. PMID:16207775
- Basak, C., Boot, W. R., Voss, M. W., & Kramer, A. F. (2008). Can training in a real-time strategy video game attenuate cognitive decline in older adults. *Psychology and Aging*, 23(4), 765–777. doi:10.1037/a0013494 PMID:19140648
- Beullens, M. A., Roe, K., & Van den Bulck, J. (2011). Excellent gamer, excellent driver? The impact of adolescents video game playing on driving behavior: A two-wave panel study. *Accident; Analysis and Prevention*, 43(1), 58–65. doi:10.1016/j.aap.2010.07.011 PMID:21094297
- Bioulac, S., Arfi, L., & Bouvard, M. P. (2008). Attention deficit/hyperactivity disorder and video games: A comparative study of hyperactive and control children. *European Psychiatry*, 23(2), 134–141. doi:10.1016/j.eurpsy.2007.11.002 PMID:18206354
- Blackler, K. J., & Curby, K. M. (2013). Enhanced visual short-term memory in action video game players. *Attention Perception Psychology*, 75(6), 1128–1136. doi:10.3758/s13414-013-0487-0 PMID:23709068
- Boot, W. R., Kramer, A. F., Simons, D. J., Fabiani, M., & Gratton, G. (2008). The effects of video game playing on attention, memory, and executive control. *Acta Psychologica*, 129(3), 387–398. doi:10.1016/j.actpsy.2008.09.005 PMID:18929349
- Braver, T. S. (2012). The variable nature of cognitive control: A dual mechanisms framework. *Trends Cognitive Neurosciences*, 16(2), 106–113. doi:10.1016/j.tics.2011.12.010 PMID:22245618
- Bushman, B. J. (1995). Moderating role of trait aggressiveness in the effects of violent media on aggression. *Journal of Personality and Social Psychology*, 69(5), 950–960. doi:10.1037/0022-3514.69.5.950 PMID:7473040

- Cain, M. S., Landau, A. N., & Shimamura, A. P. (2012). Action video game experience reduces the cost of switching tasks. *Attention Perception Psychology*, 74(4), 641–647. doi:10.3758/s13414-012-0284-1 PMID:22415446
- Castel, A. D., Pratt, J., & Drummond, E. (2005). The effects of action video game experience on the time course of inhibition of return and the efficiency of visual search. *Acta Psychologica*, 119(2), 217–230. doi:10.1016/j.actpsy.2005.02.004 PMID:15877981
- Clark, J. E., Lanphear, A. K., & Riddick, C. C. (1987). The effects of video game playing on the response selection processing of elderly adults. *Journal of Gerontology*, 42(1), 82–85. doi:10.1093/geronj/42.1.82 PMID:3794204
- Clark, K., Fleck, M. S., & Mitroff, S. R. (2011). Enhanced change detection performance reveals improved strategy use in avid action video game players. *Acta Psychologica*, 136(1), 67–72. doi:10.1016/j.actpsy.2010.10.003 PMID:21062660
- Clark, L., Robbins, T. W., Ersche, K. D., & Sahakian, B. J. (2006). Reflection impulsivity in current and former substance users. *Biological Psychiatry*, 60(5), 515–522. doi:10.1016/j.biopsych.2005.11.007 PMID:16448627
- Ellithorpe, M. E., Cruz, C., Velez, J. A., Ewoldsen, D. R., & Bogert, A. K. (2015). Moral License in Video Games: When Being Right Can Mean Doing Wrong. *Cyberpsychology, Behavior, and Social Networking*, 18(4), 203–207. doi:10.1089/cyber.2014.0599 PMID:25803312
- Figner, B., & Weber, E. U. (2011). Who takes risks when and why? Determinants of risk taking. *Current Directions in Psychological Science*, 20(4), 211–216. doi:10.1177/0963721411415790
- Fischer, P., Grietmeyer, T., Morton, T., Kastenmuller, A., Postmes, T., & Frey, D. et al.. (2009). The racing-game effect: Why do video racing games increase risk-taking inclinations. *Psychological Bulletin*, 35, 1395–1409. PMID:19596767
- Fischer, P., Kastenmuller, A., & Grietmeyer, T. (2010). Media violence and the self: The impact of personalized gaming characters in aggressive video games on aggressive behavior. *Journal of Experimental Social Psychology*, 46(1), 192–195. doi:10.1016/j.jesp.2009.06.010
- Gentile, D. A., & Anderson, C. A. (2003). Violent video games: The newest media violence hazard. In D. Gentile (Ed.), *Media violence and children* (pp. 131–152). Westport, CT: Praeger.
- Gentile, D. A., Coyne, S. M., & Walsh, D. A. (2011). Media violence, physical aggression and relational aggression in school age children: A short-term longitudinal study. *Aggressive Behavior*, 37(2), 193–206. doi:10.1002/ab.20380 PMID:21274855
- Gentile, D. A., Lynch, P. J., Linder, J. R., & Walsh, D. A. (2004). The effects of violent video game habits on adolescent hostility, aggressive behaviors, and school performance. *Journal of Adolescence*, 27(1), 5–22. doi:10.1016/j.adolescence.2003.10.002 PMID:15013257
- Goodie, A. S. (2003). The effects of control on betting: Paradoxical betting on items of high confidence with low value. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 29(4), 598–610. doi:10.1037/0278-7393.29.4.598 PMID:12924861
- Goodie, A. S. (2005). The role of perceived control and overconfidence in pathological gambling. *Journal of Gambling Studies*, 21(4), 481–502. doi:10.1007/s10899-005-5559-1 PMID:16311878
- Granic, I., Lobel, A., & Engels, R. C. M. E. (2014). The benefits of playing video games. *The American Psychologist*, 69(1), 66–78. doi:10.1037/a0034857 PMID:24295515
- Green, C. S., & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature*, 423(6939), 534–537. doi:10.1038/nature01647 PMID:12774121

- Green, C. S., & Bavelier, D. (2006). The cognitive neuroscience of video games. In L. Humphreys & P. Messaris (Eds.), *Digital media: Transformations in human communication* (pp. 211–223). New York: Peter Lang.
- Green, C. S., & Bavelier, D. (2007). Action-video-game experience alters the spatial resolution of vision. *Psychological Science*, 18(1), 88–94. doi:10.1111/j.1467-9280.2007.01853.x PMID:17362383
- Gupta, R. K., Saini, D. P., Acharya, U., & Miglani, N. (1994). Impact of television on children. *Indian Journal of Pediatrics*, 61(2), 153–159. doi:10.1007/BF02843606 PMID:7927612
- Han, D. H., Kim, Y. S., Lee, Y. S., Min, K. J., & Renshaw, P. F. (2010). Changes in Cue-Induced, Prefrontal Cortex Activity with Video-Game Play. *Cyberpsychology, Behavior, and Social Networking*, 13(6), 655–661. doi:10.1089/cyber.2009.0327 PMID:21142990
- Han, D. H., Lyoo, I. K., & Renshaw, P. F. (2012). Differential regional gray matter volumes in patients with on-line game addiction and professional gamers. *Journal of Psychiatric Research*, 46(4), 507–515. doi:10.1016/j.jpsychires.2012.01.004 PMID:22277302
- Hull, J. G., Draghici, A. M., & Sargent, J. D. (2012). A longitudinal study of risk-glorifying video games and reckless driving. *Psychology of Popular Media Culture*, 1(4), 244–253. doi:10.1037/a0029510
- Irvine M.A., Worbe Y., Bolton S., Harrison N.A., Bullmore E.T., Voon V., (2013). Impaired Decisional Impulsivity in Pathological Videogamers. *Plos One*, 16(8).
- Kaufman, G. F., & Libby, L. K. (2012). Changing beliefs and behavior through experience-taking. *Journal of Personality and Social Psychology*, 103(1), 1–19. doi:10.1037/a0027525 PMID:22448888
- King, D., Gradisar, M., Drummond, A., Lovato, N., Wessel, J., Mici, G., & Delfabbro, P. et al. (2013). The impact of prolonged violent video-gaming on adolescent sleep: An experimental study. *Journal of Sleep Research*, 22(2), 137–143. doi:10.1111/j.1365-2869.2012.01060.x PMID:23137332
- King, D. L., Delfabbro, P. H., Griffiths, M. D., & Gradisar, M. (2011). Assessing clinical trials of Internet addiction treatment: A systematic review and consort evaluation. *Clinical Psychology Review*, 31(7), 1110–1116. doi:10.1016/j.cpr.2011.06.009 PMID:21820990
- Kirby, K. N., Petry, N. M., & Bickel, W. K. (1999). Heroin addicts have higher discount rates for delayed rewards than non-drug-using controls. *Journal of Experimental Psychology. General*, 128(1), 78–87. doi:10.1037/0096-3445.128.1.78 PMID:10100392
- Kirsh, S. J., & Mounts, J. R. W. (2007). Violent video game play impacts facial emotion recognition. *Aggressive Behavior*, 33(4), 353–358. doi:10.1002/ab.20191 PMID:17593563
- Koepp, M. J., Gunn, R. N., Lawrence, A. D., Cunningham, V. J., & Dagher, A. et al. (1998). Evidence for striatal dopamine release during a video game. *Nature*, 393(6682), 266–268. doi:10.1038/30498 PMID:9607763
- Kronenberger, W. G., Matthews, V. P., Dunn, D. W., Wang, Y., Wood, E. A., & Giauque, A. L. et al. (2005). Media violence exposure and executive functioning in aggressive and control adolescents. *Journal of Clinical Psychology*, 61(6), 725–737. doi:10.1002/jclp.20022 PMID:15468343
- Kubitzki, J. (2005). Zur Problematik von Video-Rennspielen [The problem of video racing games]. *Zeitschrift fuer Verkehrssicherheit*, 3, 135–138.
- Kuhn, S., Romanowski, A., Schilling, C., Lorenz, R., Morsen, C., Seiferth, N., & Gallinat, J. et al. (2011). The neural basis of video gaming. *Translational Psychiatry*, 1(11), e53. doi:10.1038/tp.2011.53 PMID:22833208

- Lakey, C. E., Goodie, A. S., & Campbell, W. K. (2006). Frequent card playing and pathological gambling: The utility of the georgia gambling task and iowa gambling task for predicting pathology. *Journal of Gambling Studies*, 23(3), 285–297. doi:10.1007/s10899-006-9034-4 PMID:17171543
- Lakey, C. E., Goodie, A. S., Lance, C. E., Stinchfield, R., & Winters, K. C. (2007). Examining DSM-IV criteria for pathological gambling: Psychometric properties and evidence from cognitive biases. *Journal of Gambling Studies*, 23(4), 479–498. doi:10.1007/s10899-007-9063-7 PMID:17453325
- Linnet, J., Frøslev, M., Ramsgaard, S., Gebauer, L., Mouridsen, K., & Wohlert, K. (2012). Impaired Probability Estimation and Decision Making in Pathological Gambling Poker Players. *Journal of Gambling Studies*, 28(1), 113–122. doi:10.1007/s10899-011-9244-2 PMID:21484601
- Linnet, J., Gebauer, L., Shaffer, H. J., Mouridsen, K., & Møller, A. (2010). Experienced poker players differ in estimation bias and decision bias from inexperienced poker players. *Journal of Gambling Issues*, 10, 86–100. doi:10.4309/jgi.2010.24.6
- Littel, M., van den Berg, I., Luijten, M., van Rooij, A. J., Keemink, L., & Franken, I. H. A. (2012). Error processing and response inhibition in excessive computer game players: An event-related potential study. *Addiction Biology*, 17(5), 934–947. doi:10.1111/j.1369-1600.2012.00467.x PMID:22734609
- Lorenz, R. C., Kruger, J. K., Neumann, B., Schott, B. H., Kaufmann, C., Heinz, A., & Wüstenberg, T. (2013). Cue reactivity and its inhibition in pathological computer game players. *Addiction Biology*, 18(1), 134–146. doi:10.1111/j.1369-1600.2012.00491.x PMID:22970898
- Mathews, V. P., Kronenberger, W. G., Wang, Y., Lurito, J. T., Lowe, M. J., & Dunn, D. W. (2005). Media violence exposure and frontal lobe activation measured by functional magnetic resonance imaging in aggressive and non aggressive adolescents. *Journal of Computer Assisted Tomography*, 29(3), 287–292. doi:10.1097/01.rct.0000162822.46958.33 PMID:15891492
- Mentzoni, R. A., Brunborg, G. S., Molde, H., Myrseth, H., Skouvrøe, K. J. M., Hetland, J., & Pallesen, S. (2011). Problematic video game use: Estimated prevalence and associations with mental and physical health. *Cyberpsychology, Behavior, and Social Networking*, 14(10), 591–596. doi:10.1089/cyber.2010.0260 PMID:21342010
- Merritt, A. C., Effron, D. A., & Monin, B. (2010). Moral self-licensing: When being good frees us to be bad. *Social and Personality Psychology Compass*, 4(5), 344–357. doi:10.1111/j.1751-9004.2010.00263.x
- Mikhail, J. (2007). Universal moral grammar: Theory, evidence, and future. *Trends in Cognitive Sciences*, 11(4), 143–152. doi:10.1016/j.tics.2006.12.007 PMID:17329147
- Monin, B., & Miller, D. T. (2001). Moral credentials and the expression of prejudice. *Journal of Personality and Social Psychology*, 81(1), 33–43. doi:10.1037/0022-3514.81.1.33 PMID:11474723
- Perry, J. L., & Carrol, M. E. (2008). The role of impulsive behavior in drug abuse. *Psychopharmacology*, 200(1), 1–26. doi:10.1007/s00213-008-1173-0 PMID:18600315
- Rideout, V., Roberts, D.F., & Foehr, U.G. (2007). *Generation M: media in the lives of 8–18 year-olds*. Kaiser Family Foundation.
- Robbins, T. W., Gillan, C. M., Smith, D. G., de Wit, S., & Ersche, K. D. (2012). Neurocognitive endophenotypes of impulsivity and compulsivity: Towards dimensional psychiatry. *Trends in Cognitive Sciences*, 16(1), 81–91. doi:10.1016/j.tics.2011.11.009 PMID:22155014

- Sheese, B. E., & Graziano, W. G. (2005). Deciding to Defect. The Effects of Video-Game Violence on Cooperative Behavior. *Psychological Science*, 16(5), 354–357. doi:10.1111/j.0956-7976.2005.01539.x PMID:15869692
- Strobach, T., Frensch, P. A., & Schubert, T. (2012). Video game practice optimizes executive control skills in dual-task and task switching situations. *Acta Psychologica*, 140(1), 13–24. doi:10.1016/j.actpsy.2012.02.001 PMID:22426427
- Sungur, H., & Boduroglu, A. (2012). Action video game players form more detailed representation of objects. *Acta Psychologica*, 139(2), 327–334. doi:10.1016/j.actpsy.2011.12.002 PMID:22266223
- Swing, E. L. (2012). *Plugged in: The Effects of Electronic Media Use on Attention Problems, Cognitive Control, Visual Attention, and Aggression* (Unpublished doctoral dissertation). Iowa State University, Ames, IA.
- Swing, E. L., Gentile, D. A., Anderson, C. A., & Walsh, D. A. (2010). Television and video game exposure and the development of attention problems. *Pediatrics*, 126(2), 214–221. doi:10.1542/peds.2009-1508 PMID:20603258
- van Holst, R. J., Lemmens, J. S., Valkenburg, P. M., Peter, J., Veltman, D. J., & Goudriaan, A. E. (2012). Attentional Bias and Disinhibition Toward Gaming Cues Are Related to Problem Gaming in Male Adolescents. *The Journal of Adolescent Health*, 50(6), 541–546. doi:10.1016/j.jadohealth.2011.07.006 PMID:22626478
- Voon, V., Derbyshire, K., Lange, I., Worbe, Y., & Irvine, M. A. (2013). Measuring ‘waiting’ impulsivity in substance addictions and binge eating disorder in a novel analogue of the rodent serial reaction time task. *Biological Psychiatry*. PMID:23790224
- Weaver, E., Gradisar, M., Dohnt, H., Lovato, N., & Douglas, P. (2010). The effect of pre-sleep video-game playing on adolescent sleep. *Journal of Clinical Sleep Medicine*, 6, 184–189. PMID:20411697
- Weber, E. U., & Johnson, E. J. (2009). Decisions under uncertainty: psychological, economic, and neuroeconomic explanations of risk preference. In *Neuroeconomics: Decision Making and the Brain* (pp. 127–144). London: Academic Press. doi:10.1016/B978-0-12-374176-9.00010-5
- West, R. & Bailey, K. (2013). Video games and attention. In *The Oxford Handbook of Media Psychology*. New York, NY: Oxford University Press.
- Wiegman, O., & VanSchie, E. G. M. (1998). Video game playing and its relations with aggressive and prosocial behavior. *The British Journal of Social Psychology*, 37(3), 367–378. doi:10.1111/j.2044-8309.1998.tb01177.x PMID:9738313
- Wilms, I. L., Petersen, A., & Vangkilde, S. (2013). Intensive video gaming improves encoding speed to visual short-term memory in young male adults. *Acta Psychologica*, 142(1), 108–118. doi:10.1016/j.actpsy.2012.11.003 PMID:23261420
- Wu, S., & Spence, I. (2013). Playing shooter and driving videogames improves top-down guidance in visual search. *Attention, Perception & Psychophysics*, 75(4), 673–686. doi:10.3758/s13414-013-0440-2 PMID:23460295
- Zhou, Z. H., Yuan, G. Z., & Yao, J. J. (2012). Cognitive Biases toward Internet Game-Related Pictures and Executive Deficits in Individuals with an Internet Game Addiction. *PLoS ONE*, 7. PMID:23155434

ADDITIONAL READING

Anderson, C. A., Shibuya, A., Ihori, N., Swing, E. L., Bushman, B. J., Sakamoto, A., & Saleem, M. et al. (2010). Violent video game effects on aggression, empathy, and prosocial behavior in eastern and western countries: A meta-analytic review. *Psychological Bulletin*, 136(2), 151–173. doi:10.1037/a0018251 PMID:20192553

Anguera, J. A., Boccanfuso, J., Rintoul, J. L., Al-Hashimi, O., Faraji, O., Janowich, J., & Gazzaley, A. et al. (2013). Video game training enhances cognitive control in older adults. *Nature*, 501(7465), 97–101. doi:10.1038/nature12486 PMID:24005416

Ballesteros S., Prieto A., Mayas J., Toril P., Pita C., León L.P., Reales J.M. and Waterworth J., (2014). Brain training with non-action video games enhances aspects of cognition in older adults: a randomized controlled trial. *Frontiers in Aging Neuroscience*, 14;6:277.

Cardoso-Leite, P., & Bavelier, D. (2014). Video game play, attention, and learning: How to shape the development of attention and influence learning? *Current Opinion in Neurology*, 27(2), 185–191. doi:10.1097/WCO.0000000000000077 PMID:24553464

Exelmans, L., Custers, K., & Van den Bulck, J. (2015). Violent Video Games and Delinquent Behavior in Adolescents: A Risk Factor Perspective. *Aggressive Behavior*, 2, 1–13. PMID:25732796

Gentile, D. A., Li, D., Khoo, A., Prot, S., & Anderson, C. A. (2014). Mediators and moderators of long-term effects of violent video games on aggressive behavior: Practice, thinking, and action. *JAMA Pediatrics*, 168(5), 450–457. doi:10.1001/jamapediatrics.2014.63 PMID:24663396

Kovess-Masfety, V., Keyes, K., Hamilton, A., Hanson, G., Bitfoi, A., Golitz, D., & Pez, O. et al. (2016). Is time spent playing video games associated with mental health, cognitive and social skills in young children? *Social Psychiatry and Psychiatric Epidemiology*, 51(3), 349–357; Epub ahead of print. doi:10.1007/s00127-016-1179-6 PMID:26846228

Latham A.J, Patston L.L., Tippet L.J., (2013). The virtual brain: 30 years of video-game play and cognitive abilities. *Frontiers of Psychology*, 13;4:629.

Lorenz R.C., Gleich T., Gallinat J. and Kühn S., (2015). Video game training and the reward system. *Frontiers in Aging Neuroscience*, 5;9:40.

Prot, S., McDonald, K. A., Anderson, C. A., & Gentile, D. A. (2012). Video games: Good, bad, or other? *Pediatric Clinics of North America*, 59(3), 647–658. doi:10.1016/j.pcl.2012.03.016 PMID:22643171

Steenbergen L., Sellaro R., Stock A.K., Beste C., Colzato L.S. (2015). Action Video Gaming and Cognitive Control: Playing First Person Shooter Games Is Associated with Improved Action Cascading but Not Inhibition. *PLoS One*, 10;10(12).

Swing, E. L., Gentile, D. A., Anderson, C. A., & Walsh, D. A. (2010). Television and video game exposure and the development of attention problems. *Pediatrics*, 126(2), 214–221. doi:10.1542/peds.2009-1508 PMID:20603258

KEY TERMS AND DEFINITIONS

Adolescence: The transitional period between puberty and adulthood in human development, terminating legally when the age of majority is reached.

Behavior: The actions by which an organism interacts with surrounding environment.

Cognition: The mental process of knowing, including several aspects as, for example, awareness, perception, reasoning, memory and judgment.

Decision-Making: The thought process of selecting a logical choice from the available options.

Executive Control: A set of cognitive processes (as attentional control, or working memory) that are necessary for the cognitive control of behavior.

Gambling: Recreational activity that consist in gamble or bet money on the outcome of a game. Gambling can became an addictive habit.

Pathological Gambling: Persistent and recurrent problematic gambling behavior leading to clinically significant impairment or distress, as indicated by the individual itself or by friends and relatives.

Video Game: Electronic device that allows to interact with the projected images on an screen. This apparatus allows to simulate real or unreal situations, on which the players intervenes using a keyboard or a joystick.

Learning With Games and Digital Stories in Visual Programming



Wilfred W. F. Lau

The Chinese University of Hong Kong, China

INTRODUCTION

This paper traces the recent development and the use of games and digital stories for engaging students in learning in visual programming environments (Lau & Yuen, 2015).

Games have long been used to arouse and sustain students' learning interest. Wu and Wang (2012) contended that as students modify or develop a game by using a game development framework (GDF), they can learn different skills and concepts in computer science (CS) and software engineering (SE). They labeled this learning experience as game development-based learning (GDBL), and showed that it consisted of "four elements (course aim, pedagogical theory support, GDF resource pool, and impact factor), two methods (learning by creating and learning by modifying games), and six steps in the teaching process and two subjects (students and teachers)" (p. 16). More recently, Wang and Wu (2015) reviewed 66 articles related to game development and CS/SE education published between 2004 and 2012 and found that the number of articles published on this topic had increased steadily from 2004 to 2009. On average, 12 articles were published each year from 2009 to 2012. Game development was mainly used in colleges and universities (81%) but also in high schools (9%) and middle schools (9%). In CS, game development was adopted primarily to teach introductory courses and programming (77%).

Psomos and Kordaki (2012a, 2012b) advocated the practice of educational digital storytelling (EDS), which denotes the intersection of education, storytelling, and digital technology, to help

students achieve the six cognitive objectives of the revised Bloom's taxonomy (Bloom, Mesia, & Krathwohl, 1964) and acquire various literacy skills (Robin, 2006). McWilliam (2009) surveyed 300 online digital storytelling programs and found that 123 were provided by educational institutions. Of these 123 programs, 55 were hosted in K–12 schools; 42 in universities; and 26 in colleges or institutes. In schools, digital storytelling was mainly used to engage students in learning and to enhance student print and media literacy, whereas in universities and colleges, it was either embedded in student-teacher training programs or formed part of multimedia and design courses. A recent review by Gregori-Signes (2014) showed that EDS has been used extensively in different subjects and contexts at the primary, secondary, and tertiary levels as well as for teacher preparation programs.

In the following sections, I first discuss the theoretical basis and educational benefits of GDBL and EDS in general. I then present empirical evidence to support the positive effects of these two pedagogical approaches on student learning outcomes. Subsequently, I provide insights into future research directions regarding learning with games and digital stories in visual programming and then conclude the paper.

BACKGROUND

According to Dempsey, Lucassen, Haynes, and Casey (1996), computer games are rule-guided, artificial, and technologically rendered scenarios that involve one or more players and have specific

goals, constraints, payoffs, and consequences. Many people are enthusiastic about playing games and express high hopes for its positive impacts on learning. Connolly, Boyle, MacArthur, Hainey, and Boyle (2012) concluded that playing computer games, in general, was associated with numerous perceptual, cognitive, behavioral, affective, and motivational impacts and outcomes. In particular, the most prominent effects were found in knowledge acquisition/content understanding and affective and motivational outcomes. McClarty et al. (2012) identified five potential benefits of using digital games in education: 1. Games are built on sound learning principles. 2. Games provide personalized learning opportunities. 3. Games provide more engagement for learners. 4. Games teach 21st century skills. 5. Games provide an environment for authentic and relevant assessment (pp. 6–7).

Furthermore, Werner, Denner, and Campe (2014) argued that designing and programming a game can be regarded as an ill-structured design problem (Jonassen, 2000) that requires students to define the goal, decide how to reach that goal, and evaluate the solution. Because most games involve problem-solving tasks that are dynamic, time dependent, and complex, game development is often understood as a complex problem-solving process that draws on an individual's abilities to formulate complex problems, design systems, and understand human behavior (Denner, Werner, Campe, & Ortiz, 2014). Thus, it is anticipated that such a practice can help improve student problem-solving skills and higher-order thinking abilities.

Digital storytelling refers to “the art of combining narrative with digital media such as images, sound, and video to create a short story” (Dreon, Kerper, & Landis, 2011). The Center for Digital Storytelling in Berkeley, California, identified seven crucial elements of digital storytelling: point of view, a dramatic question, emotional content, the gift of your voice, the

power of the soundtrack, economy, and pacing (<http://digitalstorytelling.coe.uh.edu/page.cfm?id=27&cid=27&sublinkid=31>). Robin (2008) asserted that digital storytelling enables users to combine the traditional processes of creative story writing with various types of multimedia, which results in an electronic artifact that is accessible through both local computers and the Internet. Constructionism and the narrative paradigm are the two basic theories that support the use of digital storytelling in education (Wang & Zhan, 2010). In constructionism, students create external and sharable objects (learning by making) through active interaction and engagement in the learning process. The narrative paradigm posits that meaningful learning occurs as students attempt to interpret actions, words, and deeds from a story into something relevant to their experiences.

Digital storytelling benefits student learning in numerous ways. Using digital storytelling in the classroom helps both general education students and those with learning difficulties become more motivated to practice traditional writing. Digital storytelling can be an effective method for engaging students, from kindergarten through college, in student-centered activities mediated by technology. Digital storytelling empowers students to express themselves through multimodal communication techniques and to develop multiliteracies such as digital, global, technology, visual, and information literacy (Robin, 2008). It helps students to understand more clearly the subject matter and to improve their literacy skills.

Porter (2005) found several advantages of practicing digital storytelling: helping students to improve their multiple literacy skills, increasing their content learning, engaging students in learning across the curriculum in schools, fostering their mastery of many 21st century skills, and meeting technology standards. Digital storytelling can be integrated across a range of subject areas to enhance student learning outcomes. Curriculum

integration enables students to establish connections across content areas and to communicate their experiences to a wider audience.

EMPIRICAL EVALUATION

Hwang, Hung, and Chen (2014) found that a peer assessment-based game development approach enhanced students' learning achievement, learning motivation, problem-solving skills, and their perceptions of the use of educational computer games compared with those who were instructed with a conventional game development approach. The peer assessment-based approach also helped students engage in deep learning. Akcaoglu and Koehler (2014) showed that students attending the Game-Design and Learning program outperformed their counterparts who did not attend the program in several problem-solving tasks including system analysis and design, decision making, and troubleshooting. This provides empirical evidence that the program can improve student problem-solving abilities. In their review, Wang and Wu (2015) also reported the positive effects of using game development to improve student motivation and engagement.

However, Chu and Hung (2015) indicated that although the game-based development approach was effective in promoting student problem-solving skills, students learning with this approach showed no difference in academic achievement or motivation from others who learned by playing a digital game. This calls for further research to determine the contextual factors that influence the success and effectiveness of this learning approach.

Burke and Kafai (2012) demonstrated the efficacy of a 7-week writing workshop in helping students to learn the basics of both programming and storytelling, which was reflected in the digital stories they produced (products), the debugging and revising they performed (processes), and their overall perceptions of the workshop at its end. Students also realized that programming and

writing were interrelated composition processes. Fields, Kafai, Strommer, Wolf, and Seiner (2014) explored the effects of collaborative and interactive storytelling activities in a Scratch online community on youth learning of computer programming in two aspects: (1) how this learning experience supports creative expression in collaborative digital stories in both media and code; and (2) how feedback from commentators influences the artistic and computational qualities of stories undergoing revision. They found that the participants could make story designs that were more creative and develop code that was more sophisticated for their final submission after receiving constructive feedback from the Scratch online community.

FUTURE RESEARCH DIRECTIONS

It is believed that GDBL and EDS will remain frequently used approaches through which students learn computer programming in visual environments. Future research may continue to investigate how computational thinking (CT) is developed in visual programming languages. According to Brennan and Resnick (2012), CT involves problem solving by using CS concepts and entails three key dimensions: computational concepts, computational practices, and computational perspectives. Some empirical evidence suggests that visual programming helps increase K–12 student performance in the three dimensions of CT (Lye & Koh, 2014). In addition, Chang (2014) showed the effectiveness of using Alice and Scratch in corrective instruction, raising university students' playfulness and enjoyment levels, and improving their learning performance.

Another research direction concerns the viability of using visual programming to enhance CT, which may promote science, technology, engineering, and mathematics (STEM) learning among K–12 students. CT has been identified as the core component in all STEM disciplines (Grover & Pea, 2013) and there has been an ongoing effort to integrate CT into STEM curricula. For

example, Weintrop et al. (2014) proposed a CT-STEM skills taxonomy framework that consists of four major categories of relevant skills: data and information skills, modeling and simulation skills, computational problem-solving skills, and systems thinking skills. However, little is known about how CT can be taught in STEM subjects in practice and, more important, whether and how visual programming may affect STEM learning through the development of CT. Additional studies exploring these issues are necessary.

CONCLUSION

This paper reports on recent development of using GDBL and EDS to engage students in learning in visual programming environments. The empirical findings support the positive effects of these two learning approaches on a range of student learning outcomes. Because many available visual programming tools are free of charge and provide a low-floor, high-ceiling learning environment, teachers should encourage students to venture into the programming world with these tools. Such practice is beneficial to student learning both within the CS discipline and across disciplines. In particular, more research should be devoted to investigating how visual programming may enhance student CT and STEM learning.

REFERENCES

- Akcaoglu, M., & Koehler, M. J. (2014). Cognitive outcomes from the game-design and learning (GDL) after-school program. *Computers & Education*, 75(1), 72–81. doi:10.1016/j.compedu.2014.02.003
- Bloom, B. S., Mesia, B. B., & Krathwohl, D. R. (1964). *Taxonomy of Educational Objectives* (vols. 1-2). New York: McKay Publisher.
- Brennan, K., & Resnick, M. (2012). *New frameworks for studying and assessing the development of computational thinking*. Paper presented at the 2012 Annual Meeting of the American Educational Research Association, Vancouver, Canada.
- Burke, Q., & Kafai, Y. B. (2012). The writers' workshop for youth programmers: Digital storytelling with scratch in middle school classrooms. *Proceedings of the 43rd ACM Technical Symposium on Computer Science Education* (pp. 433-438). New York, NY: ACM.
- Chang, C.-K. (2014). Effects of using Alice and Scratch in an introductory programming course for corrective instruction. *Journal of Educational Computing Research*, 51(2), 185–204. doi:10.2190/EC.51.2.c
- Chu, H.-C., & Hung, C.-M. (2015). Effects of the digital game-development approach on elementary school students learning motivation, problem solving, and learning achievement. *International Journal of Distance Education Technologies*, 13(1), 87–102. doi:10.4018/ijdet.2015010105
- Connolly, T. M., Boyle, E. A., MacArthur, E., Hainey, T., & Boyle, J. M. (2012). A systematic literature review of empirical evidence on computer games and serious games. *Computers & Education*, 59(2), 661–686. doi:10.1016/j.compedu.2012.03.004
- Dempsey, J., Lucassen, B. A., Haynes, L. L., & Casey, M. S. (1996). *Instructional applications of computer games*. Paper presented at the Annual Meeting of the American Educational Research Association, New York, NY.
- Denner, J., Werner, L., Campe, S., & Ortiz, E. (2014). Using game mechanics to measure what students learn from programming games. *International Journal of Game-Based Learning*, 4(3), 13–22. doi:10.4018/ijgbl.2014070102

- Dreon, O., Kerper, R. M., & Landis, J. (2011). Digital storytelling: A tool for teaching and learning in the YouTube generation. *Middle School Journal*, 42(5), 4–9. doi:10.1080/00940771.2011.11461777
- Fields, D. A., Kafai, Y. B., Strommer, A., Wolf, E., & Seiner, B. (2014). Interactive storytelling for promoting creative expression in media and coding in youth online collaboratives in Scratch. *Proceedings of Constructionism*.
- Gregori-Signes, C. (2014). Digital storytelling and multimodal literacy in education. *PORTA LINGUARUM*, 22, 237–250.
- Grover, S., & Pea, R. (2013). Computational thinking in K-12: A review of the state of the field. *Educational Researcher*, 42(1), 38–43. doi:10.3102/0013189X12463051
- Hwang, G. J., Hung, C. M., & Chen, N. S. (2014). Improving learning achievements, motivations and problem-solving skills through peer assessment-based game development approach. *Educational Technology Research and Development*, 62(2), 129–145. doi:10.1007/s11423-013-9320-7
- Jonassen, D. H. (2000). Toward a design theory of problem solving. *Educational Technology Research and Development*, 48(4), 63–85. doi:10.1007/BF02300500
- Lau, W. W. F., & Yuen, A. H. K. (2015). Teaching in visual programming environments. In M. Khosrow-Pour (Ed.), *Encyclopedia of Information Science and Technology* (3rd ed.; pp. 2600–2608). Hershey, PA: Idea Group Inc. doi:10.4018/978-1-4666-5888-2.ch253
- Lye, S. Y., & Koh, J. H. L. (2014). Review on teaching and learning of computational thinking through programming: What is next for K-12? *Computers in Human Behavior*, 41, 51–61. doi:10.1016/j.chb.2014.09.012
- McClarty, K. L., Orr, A., Frey, P. M., Dolan, R. P., Vassileva, V., & McVay, A. (2012). *A literature review of gaming in education: Research report*. Upper Saddle River, NJ: Pearson.
- McWilliam, K. (2009). The global diffusion of a community media practice: Digital storytelling online. In J. Hartley & K. McWilliam (Eds.), *Story circle: Digital storytelling around the world* (pp. 37–75). Oxford, UK: Wiley-Blackwell. doi:10.1002/9781444310580.ch3
- Porter, B. (2005). *Digitales: The art of telling digital stories*. Denver, CO: Bernajean Porter Consulting.
- Psomos, P., & Kordaki, M. (2012a). Pedagogical analysis of educational digital storytelling environments of the last five years. *Procedia: Social and Behavioral Sciences*, 46, 1213–1218. doi:10.1016/j.sbspro.2012.05.277
- Psomos, P., & Kordaki, M. (2012b). A supporting framework for the creation of digital stories and learning programming by the students within Kodu, Scratch and Storytelling Alice. In P. Resta (Ed.), *Proceedings of Society for Information Technology & Teacher Education International Conference* (pp. 1419–1424). Chesapeake, VA: AACE.
- Robin, B. (2008). Digital storytelling: A powerful technology tool for the 21st century classroom. *Theory into Practice*, 47(3), 220–228. doi:10.1080/00405840802153916
- Robin, B. R. (2006). *The educational uses of digital storytelling*. Retrieved October 2, 2015 from <http://digitalliteracyintheclassroom.pbworks.com/f/Educ-Uses-DS.pdf>
- Wang, A. I., & Wu, B. (2015). Use of game development in computer science and software engineering education. In K. M. L. Cooper & W. Scacchi (Eds.), *Computer Games and Software Engineering* (pp. 31–58). Boca Raton, FL: CRC Press, Taylor & Francis. doi:10.1201/b18453-4

- Wang, S., & Zhan, H. (2010). Enhanced teaching and learning with digital storytelling. *International Journal of Information and Communication Technology Education*, 6(2), 76–87. doi:10.4018/jicte.2010040107
- Weintrop, D., Beheshti, E., Horn, M. S., Orton, K., Jona, K., Trouille, L., & Wilensky, U. (2014). *Defining computational thinking for science, technology, engineering, and math*. Paper presented at the 2014 Annual Meeting of the American Educational Research Association, Philadelphia, PA.
- Werner, L., Denner, J., & Campe, S. (2014). Using computer game programming to teach computational thinking skills. In S. Karen (Ed.), *Learning, Education and Games* (pp. 37–53). Pittsburgh, PA: ETC Press.
- Wu, B., & Wang, A. I. (2012). A guideline for game development-based learning: A literature review. *International Journal of Computer Games Technology*.
- Hung, W., & Van Eck, R. (2010). Aligning problem solving and gameplay: A model for future research and design. In R. Van Eck (Ed.), *Interdisciplinary models and tools for serious games. Emerging concepts and future directions* (pp. 227–263). Hershey, PA: IGI Global. doi:10.4018/978-1-61520-719-0.ch010
- Jeon, J., Kim, K., & Jung, S. (2012). A study on the game programming education based on educational game engine at school. *Journal of Education & Learning*, 1(2), 282–287. doi:10.5539/jel.v1n2p282
- Johnson, C. (2014). *'I liked it but it made you think too much': A case study of computer game authoring in the Key Stage 3 ICT curriculum* (Unpublished doctoral thesis). University of East Anglia, UK.
- Kafai, Y. B., & Burke, Q. (2014). *Connected code: Why children need to learn programming*. Cambridge, MA: The MIT Press.
- Kärkkäinen, K., & Vincent-Lancrin, S. (2013). Sparking innovation in STEM education with technology and collaboration: A case study of the HP catalyst initiative (OECD Education Working Paper No. 91). OECD Publishing.
- Kelleher, C. (2009). Supporting storytelling in a programming environment for middle school children. In I. A. Iurgel, N. Zagalo, & P. Petta (Eds.), *Proceedings of the 2nd Joint International Conference on Interactive Digital Storytelling: Interactive Storytelling (ICIDS '09)* (pp. 1–4). Berlin, Heidelberg: Springer-Verlag. doi:10.1007/978-3-642-10643-9_1
- Kordaki, M., & Psomos, P. (2014). An adaptive educational digital storytelling environment focusing on students' misconceptions. In L. G. Chova, A. L. Martínez, & I. C. Torres (Eds.), *Proceedings of 8th International Technology, Education and Development Conference (INTED 2014)* (pp. 6634–6641). Valencia, Spain.

ADDITIONAL READING

- Chang, C. K., & Biswas, G. (2011). Design engaging environment to foster computational thinking. *Proceedings of the World Conference on Educational Multimedia, Hypermedia and Telecommunications* (Vol. 2011, No. 1, pp. 2898–2902).
- Curzon, P., Dorling, M., Ng, T., Selby, C., & Woollard, J. (2014). *Developing computational thinking in the classroom: A framework*. Retrieved October 2, 2015 from <http://community.computingatschool.org.uk/files/3517/original.pdf>
- Dalal, A. C. (2008). Digital storytelling as a gateway to computer science. *Journal of the Research Center for Educational Technology*, 4(2), 124–137.

- Lee, I., Martin, F., Denner, J., Coulter, B., Allan, W., Erickson, J., & Werner, L. et al. (2011). Computational thinking for youth in practice. *ACM Inroads*, 2(1), 32–37. doi:10.1145/1929887.1929902
- Mannila, L., Dagiene, V., Demo, B., Grgurina, N., Mirolo, C., & Rolandsson, L. ...Settle, A. (2014). Computational thinking in K-9 education. In *Proceedings of the Working Group Reports of the 2014 on Innovation & Technology in Computer Science Education Conference (ITiCSE-WGR' 14)* (pp.1–29). New York, NY: ACM.
- Moreno, J. (2012). Digital competition game to improve programming skills. *Journal of Educational Technology & Society*, 15(3), 288–297.
- Peppler, K. A., Santo, R., Gresalfi, M., Tekinbas, K. S., & Sweeney, L. B. (2014). *Script changers: Digital storytelling with Scratch*. Cambridge, MA: The MIT Press.
- Prensky, M. (2008). Students as designers and creators of educational computer games: Who else? *British Journal of Educational Technology*, 39(6), 1004–1019. doi:10.1111/j.1467-8535.2008.00823_2.x
- Psomos, P., & Kordaki, M. (2015a). A novel educational digital storytelling tool focusing on student misconceptions. In *Proceedings of 6th World Conference on Educational Sciences* [University of Malta, Malta.]. *Procedia: Social and Behavioral Sciences*, 191, 82–86. doi:10.1016/j.sbspro.2015.04.476
- Psomos, P., & Kordaki, M. (2015b). Digital storytelling pedagogical evaluation star: Views of teachers. In L. G. Chova, A. L. Martínez, & I. C. Torres (Eds), *Proceedings of 7th International Conference on Education and New Learning Technologies (EDULEARN2015)* (pp. 1466–1473). Barcelona, Spain.
- Renda, C., Robin, B., Otañez, M., & Susono, H. (2013). New Frontiers in digital storytelling: Technological advances and changing critical teacher education approaches. In R. McBride & M. Searson (Eds.), *Proceedings of Society for Information Technology & Teacher Education International Conference 2013* (pp. 1706–1711). Chesapeake, VA: Association for the Advancement of Computing in Education (AACE).
- Robin, B. R., & McNeil, S. G. (2012). What educators should know about teaching digital storytelling. *Digital Education Review*, 22, 37–51.
- Schifter, C. C. (2013). Games in learning, design, and motivation. In M. Murphy, S. Redding, & J. Twyman (Eds.), *Handbook on innovations in learning* (pp. 149–164). Philadelphia, PA: Center on Innovations in Learning, Temple University; Charlotte, NC: Information Age Publishing. Retrieved October 2, 2015 from <http://www.centeril.org/>
- Smith, K. A., Douglas, T. C., & Cox, M. F. (2009). Supportive teaching and learning strategies in STEM education. In R. G. Baldwin (Ed.), *New directions for teaching and learning: Improving the climate for undergraduate teaching in STEM fields* (Volume 2009, Issue 117) (pp. 19–32). San Francisco, CA: Jossey-Bass. doi:10.1002/tl.341
- Vasilopoulos, I. V. (2014). *The design, development and evaluation of a visual programming tool for novice programmers: psychological and pedagogical effects of introductory programming tools on programming knowledge of Greek students* (Unpublished doctoral thesis). Teesside University, UK.
- Voogt, J., Fisser, P., Good, J., Mishra, P., & Yadav, A. (2015). Computational thinking in compulsory education: Towards an agenda for research and practice. *Education and Information Technologies*, 20(4), 715–728. doi:10.1007/s10639-015-9412-6

Werner, L., Denner, J., & Campe, S. (2014). Children programming games: A strategy for measuring computational learning. *ACM Transactions on Computing Education*, 14(4), Article 24.

Wing, J. M. (2006). Computational thinking. *Communications of the ACM*, 49(3), 33–35. doi:10.1145/1118178.1118215

Wong, S. Y., Tee, W. J., Nagappan, S. D., & Sien, K.-S. (2015). A survey analysis: Students' interest in simplification game development tools for creative learning. In S.F. Tang & L. Logonnathan (Eds.), *Taylor's 7th Teaching and Learning Conference 2014 Proceedings* (pp. 359-369). Singapore: Springer Science+Business Media.

KEY TERMS AND DEFINITIONS

Visual Programming Environments: They are computer software that support the use of visual expressions such as graphics, drawings, animations, and icons in the programming process.

Game Development Framework: It encompasses the toolkits used to develop or modify games such as the game engine, game editors,

simulation platforms, or any integrated development environment.

Game Development-Based Learning: This learning approach requires students to modify or develop a game on the basis of a game development framework. Students are expected to acquire different computer science and software engineering skills and concepts in the learning process.

Digital Storytelling: It refers to the combination of a narrative with multimedia elements such as text, images, sound, and video for creating a short story.

Educational Digital Storytelling: A blend of education with digital storytelling, educational digital storytelling aims to help students to attain the six cognitive objectives of the revised Bloom's taxonomy and develop various literacy skills.

Computational Thinking: It involves solving problems by employing concepts fundamental to computer science.

Science, Technology, Engineering, and Mathematics: It is increasingly important for any government to improve STEM education in order to maintain its competitiveness in science and technology development.

The Process Model of Gameplay to Understand Digital Gaming Outcomes



Linda K. Kaye
Edge Hill University, UK

INTRODUCTION

Digital games have been the focus of psychological research for a number of decades, yet there remains substantial concern and debate about the potentially harmful effects of violent digital games on children and young people (DeLisi, Vaughn, Gentile, Anderson & Shook, 2013). These concerns primarily relate to the amount of violent content in particular types of digital games, given that evidence has suggested exposure to violent content in games is related to increases in aggressive attitudes and behaviour (Anderson et al., 2010), and reduced sensitivity to real life violence (Bartholow, Bushman, & Sestir, 2006).

While some researchers claim that there is conclusive evidence of a link between violent game exposure and aggressive cognition, affect and behaviour (Anderson et al., 2010), others have criticised the theoretical and methodological basis of such claims (Adachi & Willoughby, 2011a; Elson & Ferguson, 2014; Ferguson, 2007). This suggests a need to reconsider the way in which outcomes of (violent) gaming are studied. This is the key focus of the current chapter. This includes a critical consideration of the socio-cognitive models which are typically used within this area, and questions the extent to which they can effectively represent the range of potential outcomes of playing digital games. Following this, other key factors are reviewed, and presented as a reason to reconsider the theoretical underpinnings of this research field. Here, a Process Model of Gameplay is presented as a solution to further understand these issues. Specifically, this model

aims to provide a framework through which to understand a variety of factors and the diversity of gaming experiences, and their combined role on gaming outcomes. In a practical sense, this can inform future psychological research to adopt a more holistic approach when measuring digital gaming outcomes, particularly in controlling for the extent of factors which are influential in this regard.

Background: Existing Theoretical Models

Existing theoretical models explaining the influence of violent game content on aggression outcomes include the General Aggression Model (GAM: Anderson & Bushman, 2002; Anderson & Huesmann, 2003), and the General Learning Model (GLM: Buckley & Anderson, 2006). These consolidate existing socio-cognitive models to explain the effects of media violence on aggression-related outcomes. The underlying principle of these models is that engaging with violent digital games interacts with an individual's internal state, trait and situational factors to influence appraisal and subsequent behaviour (DeWall & Anderson, 2011). The long-term effects of such processes are said to occur through the development of knowledge structures via learning processes which create a repeated pattern of responses, increasing the accessibility of violent "scripts" over time (Barlett & Anderson, 2013). Beyond this, the GLM more specifically explains how exposure to *any* media content can "teach" a behavioural response. This includes the role of prosocial media content

on teaching prosocial behaviours (Greitemeyer & Osswald, 2009), suggesting digital games can be effective “teachers” when exposing participants to specific types of content (Gentile & Gentile, 2008), and thus highlighting their use within educational contexts (Prensky, 2001). However, given that many games include violent content, this has caused substantial concern in particular academic communities, and prompted much research to investigate the way in which repeated exposure to game violence may be harmful through the way in which aggressive scripts are learnt and applied in the real world.

Although some studies have provided support for these models, other studies have not. Specifically, some researchers have criticised the restrictive nature of these models (e.g., Ferguson, 2009). One such criticism is that they imply “passive modelling” in which individuals exposed to violent media will be more likely to engage in real world violent behaviour, regardless of other key influential factors (e.g., family violence, trait aggression) (Ferguson, 2009). Relevant evidence here is the modelling which has been shown through objectification of females within many digital games (Burgess, Stermer & Burgess, 2007), and the consequences this has on real-world prejudices and beliefs towards women (Beck, Boys, Rose & Beck, 2012; Dill & Thrill, 2007). Similarly, particularly in relation to aggressive effects, existing models exclude the role of genetic predispositions and innate motivational systems towards violent behaviour which are believed to explain a substantial proportion of the variance in real-life violent behaviour (Eley, Lichtenstein, & Moffitt, 2003; Ferguson, Rueda, Cruz, Ferguson, Ritz, & Smith, 2008). Thus, greater specificity is needed within theory on how such factors function in this process (Ferguson & Dyck, 2012).

Additionally, existing models are not entirely clear in their capacity to test the influence of a range of other factors which have also been suggested to interact upon gaming outcomes (Krahé & Möller, 2004; Markey & Scherer, 2009). For example, emotional regulation has been found as a

mediator of the of violent gaming—aggression link (Unsworth, Devilly, & Ward, 2007). This suggests that gaming may serve a mood management function for some individuals (Unsworth et al., 2007). Indeed, mood management theory (Zillman, 1988) suggests that players may be motivated to play as a means of maximising exposure to positive stimuli, as a way of enhancing mood (Bowman & Tamborini, 2015). Conversely, this can result in games being successful in alleviating negative states (Ferguson & Rueda, 2010). This suggests a positive function of (violent) gameplay which existing theoretical models are unable to address. Understanding gaming processes are therefore key, highlighting the importance of understanding *changes* in experiences through gaming in order to fully examine the influence of gameplay processes on the outcomes of the activity. In response to this, the Process Model of Gameplay is presented (see Figure 1). This acknowledges the influence of the dynamic and varied nature of gaming processes upon the range of outcomes of the activity.

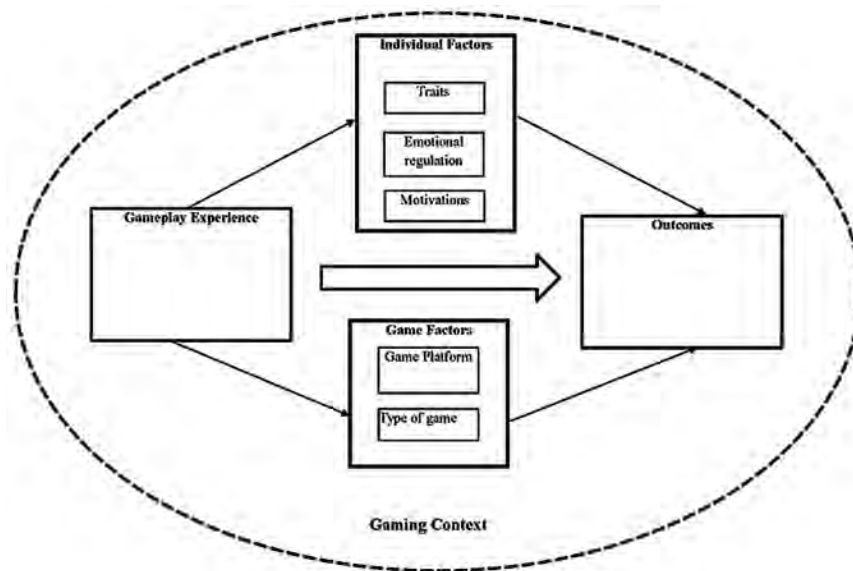
Future Research Directions: The Process Model

The Process Model can provide a basis through which to underpin research which aims to explore the various outcomes of gaming, particularly in how the factors identified here are influential in this regard. The merit of this model is that it highlights the range of factors and potential interactions which underpin the dynamic and varied nature of gaming experiences. This can therefore provide a basis from which to research a variety of different mechanisms which underpin gaming outcomes. The following sections identify key themes which are integrated within this process-based model to understanding digital gaming outcomes.

Type of Games and Platforms

In addition to individual factors (e.g., traits, emotional regulation) previously outlined, potential

Figure 1.



variations in gaming experiences may also depend on factors such as game genre and platform (Limpingos, Schmierbach, Kegerise & Dardis, 2011). For example, previous studies have identified reduced physiological and emotional responses when playing games on handheld compared to TV-based consoles (Ivory & Magee, 2009). Additionally, enhanced responses to violent content has been found when playing on immersive games technology compared to traditional platforms (Persky & Blascovich, 2007, 2008). This suggests that different gaming consoles have a role in determining the outcomes of the activity. Different game genres have also been found to have differential physiological effects, with greater arousal being identified when playing driving games compared to first-person shooter games (Goodson & Pearson, 2009). This suggests the importance of integrating these factors into models of gaming outcomes. Games are structurally complex, and consist of a range of different dimensions (Bryce & Kaye, 2011). Indeed, previous research shows how specific game dimensions (e.g., identification with game characters, pace of action) are influential upon the derived experiences of the activity (Elson, Breuer, Van Looy, Kneer, & Quandt, 2015). Therefore, understanding the complexities of

digital games and acknowledging this within the theoretical underpinnings of digital game research is a key requirement. Unfortunately, theories such as GAM or GLM fail to account for the way in which variables and experiential factors may be influential in the way in which individuals respond to games, and how these may vary as a result of the interaction of the individual and the games technology (Nacke et al., 2009). In light of this, the Process Model includes these as factors underpinning digital gaming outcomes. Here, “game platform” and “type of game” are two identified sub-factors included within the model, in an attempt to provide a more interactionist perspective for modelling gaming outcomes.

Knowledge and Engagement

Another issue with previous models is the lack of acknowledgement of the knowledge of players, and their experiences associated with moral engagement with virtual violence. Specifically in relation to player knowledge, evidence shows differences between players and non-players in cognitive defence mechanism responses to shooter games (Kneer, Munko, Glock & Bente, 2012). Additionally, other research demonstrates differences

between individuals in their gaming experiences (Cotton, Mayes, Jentsch & Sims, 2001; Cox, Cairns, Shah & Carroll, 2012), and responses to priming of aggression-related concepts (Glock & Kneer, 2009). In respect of moral engagement in violence, research shows these individual differences to be influential to the emotional responses associated with gameplay (Young & Whitty, 2010, 2011). The ability of some games to offer opportunities for players to engage in violent acts and violate real-world social norms in the virtual world can provide intense positive emotional experiences (Bertozzi, 2008; Hartmann & Vorderer, 2010), suggesting that the consequences of violent gameplay are not necessarily solely and exclusively negative in nature. In light of this, the Process Model acknowledges “player traits” in respect of the individual variations in responses to specific game content, and how these hold a combined influence on the nature of gaming outcomes. In this way, the Process Model can provide a more holistic perspective of gaming outcomes, in respect of player experience—game context interactions, which previous models fail to include.

Contextual Factors

A further key limitation of existing theoretical models is their inability to account for the variety of social contexts in which games are experienced. That is, existing models do not account for differences in experiences between immediate gaming contexts in which gamers may be playing alone, compared to with others, and whether or not this in the physical presence of others or via online gameplay. Evidence suggests that different social contexts of gameplay result in differential experiences (de Kort, IJsselsteijn, Poels, 2007). For example, social opportunities with others can enhance emotional gaming experiences (Kaye & Bryce, 2012), and result in enhanced positive mood following gameplay (Kaye & Bryce, 2014). Other research has compared levels of arousal, engagement and mood between different gameplay contexts and found that outcomes are more strongly enhanced in the physical presence of others during

gameplay relative to virtual presence (Gajadhar, de Kort & IJsselsteijn, 2009). Similarly, different types of gameplay (i.e., competitive versus cooperative play) do not hold equivalent outcomes (Lim & Lee, 2009), resulting in limitations of existing models which cannot account for these distinctions. The importance of acknowledging social contexts of gaming is particularly pertinent given the increasing popularity of social gaming (Information Solutions Group, 2011). It makes conceptual sense therefore to integrate these constructs into frameworks designed to theorise on gaming impacts.

As well as immediate gameplay contexts, there are wider socio-cultural contexts of interest here. That is, values of social systems may be largely determined by cultural conception and transformation of social representations (Moscovici, 2000). Social representation of identity (e.g., gamer identity) for example is one factor which has been found to be related to different psychological outcomes of digital gaming (Kaye, 2015), suggesting these wider social contexts and constructs play a role in this process. Similarly, when considering gaming within the context of gamers’ everyday lives, it is possible to see how it operates as a part of gamers’ social identity formation (Kaye, 2014). This highlights the importance of understanding these socio-contextual influences on gaming outcomes. Thus, the Process Model addresses the limitations of existing theoretical gaming models by acknowledging gaming context (in respect of immediate and wider socio-cultural context) within this framework. Specifically, this framework models the player—game experience interaction through a contextual lens, given that the outcomes associated with these interactions are expected to vary considerably as a result of contextual variations.

A Leisure Framework

The fact that gaming is a highly popular leisure activity suggests that it facilitates a range of positive experiences and outcomes for players. Most people engage in leisure to experience enjoyment

and freedom from everyday life (Harper, 1986). Specifically in relation to gaming, the derived enjoyment is a key intrinsic motivation for continued engagement (Ryan, Rigby & Przybylski, 2006), in similar ways to that of other leisure (Harper, 1986). That is, gaming provides opportunities for psychological need satisfaction, through enhancing perceptions of efficacy, skill growth and relatedness with others (Przybylski, Deci, Rigby & Ryan, 2014; Reinecke, Klatt & Kramer, 2011; Ryan & Deci, 2000).

A leisure perspective presents a useful approach through which to examine the motivational and experiential processes associated with gaming. In turn, this explains the potential psychological benefits, including; enhanced well-being (Ryan et al., 2006), positive mood (Hull, 1990), and recovery (Collins & Cox, 2014; Reinecke, 2009). Therefore, when viewing gaming through the lens of enjoyable leisure, it should hold similar benefits as other leisure activities, highlighting the utility of this framework for theorising on the more positive outcomes of gaming. Specifically, this perspective assumes a broader role of the activity itself; a key feature underpinning the Process Model presented here. That is, it can provide a way of examining the *processes* involved in gameplay, to better understand how these experiences operate for promoting a range of (positive) gaming outcomes. This extends the focus from solely concerning the content of games, and instead suggests them to operate in equivalent ways to that of other leisure. This can extend beyond the realms of immediate and short-term effects, and can offer insight into the more enduring, broader outcomes of this activity (Mayra, 2007).

An operational example of the Process Model relates to flow theory (Csikszentmihalyi, 1975). This theory proposes that individuals experience enjoyment during an intrinsically motivated activity, characterised by total concentration and lack of self-awareness (Csikszentmihalyi, 1975). Flow, in the context of gaming can provide an insight into motivations for participation, the associated positive experiences, and their combined influ-

ence on the nature of gaming outcomes. This is supported by research highlighting the importance of understanding player experiences, particularly in relation to the goal-directed nature of play, and how this might provide a more useful framework than a focus on game content, as in previous models (Oswald, Prorock & Murphy, 2014). As outlined in the Process Model, enjoyable and intrinsically motivated gaming experiences, are likely to largely result in positive emotional outcomes.

Considering gaming as enjoyable leisure provides a useful framework through which to examine the impact of experiential aspects of the activity, which may better account for a more diverse range of outcomes. Specifically in relation to flow for example, the Process Model can underpin its function in explaining continued motivation for specific types of games, highlighting the way in which gaming motivations and experiences are best understood in combination, as an effective way of predicting gaming outcomes. Therefore, the possibilities of future research avenues are fruitful, particularly in relation to examining the impacts which occur as a result of the *processes* of gameplay. On a practical level, research findings within this perspective can offer game industry representatives better scope for applying these approaches to player testing protocols, and indeed an informed evidence-base through which to effectively market their games in light of positive player experiences.

Although existing theory has largely focused on the role of violent game content on aggression-related outcomes, this represents only a small part of the overall gaming experience. Games are complex, providing diverse and variable experiences, indicating the utility of the Process Model in better capturing the dynamic and idiosyncratic nature of the player—game interaction. It is intended that the Process Model may initiate a research agenda which acknowledges the complexities of digital games, and the processes afforded to the activity itself, in order to expand beyond the “negative effects perspective”.

CONCLUSION

This chapter has provided a critical review of the theoretical underpinnings of digital gaming research. It has highlighted the limitations, and presented the Process Model of Gameplay as a more holistic framework for these issues. Specifically, this model presents a dynamic process framework for better examining digital gaming outcomes, through a more comprehensive interactionalist account of the relevant factors. In this way, this model may serve as a vehicle for modeling further research approaches. Although initial support is available for the range of influences included within the Process Model of Gameplay, future research is required which examines the validity and efficacy of the proposed framework. In particular, additional work which accounts for the dynamic process of gaming and how this impacts upon diverse gaming outcomes is greatly warranted. Such a focus would entail a move towards methods such as Experience Sampling Methodology, to gain players' accounts of their "real-world" gaming experiences (Kaye, Monk & Hamlin, in press). Indeed, this calls for research in this area to shift from the traditional "global, broad" perspective, towards a more idiographic one, if a more comprehensive understanding of the range and diversity of outcomes is to be established. This would provide a valuable contribution to the theoretical underpinnings for explaining how and why different individuals respond differently to particular digital games.

Overall, this chapter has argued for a greater focus on understanding the dynamics of gameplay as the direction for future research in this area, and presented a Process Model in response to this. Although a large body of research has examined the influence of violent game content on aggression-related outcomes, this constitutes only one aspect of the overall digital gaming experience. Understanding the complexity of digital games, gaming process and associated outcomes

can only be achieved by adopting a more holistic approach to the study of the activity. In this way, it is hoped that the Process Model of Gameplay may be a suitable starting point, through which to initiate this research agenda.

REFERENCES

- Adachi, P. J. C., & Willoughby, T. (2011b). The effect of video game competition and violence on aggressive behavior: Which characteristic has the greatest influence? *Psychology of Violence*, 1(4), 259–274. doi:10.1037/a0024908
- Adachi, T., & Willoughby, T. (2011a). The effect of violent video games on aggression. Is it more than just the violence? *Aggression and Violent Behavior*, 16(1), 55–62. doi:10.1016/j.avb.2010.12.002
- Anderson, C. A., & Bushman, B. J. (2002). Human Aggression. *Annual Review of Psychology*, 53(1), 27–51. doi:10.1146/annurev.psych.53.100901.135231 PMID:11752478
- Anderson, C. A., & Huesmann, L. R. (2003). Human Aggression: A Social-Cognitive View. In M. A. Hogg & J. Cooper (Eds.), *The Sage Handbook of Social Psychology* (pp. 296–323). Thousand Oaks, CA: Sage Publications Inc.
- Anderson, C. A., Shibuya, A., Ihori, N., Swing, E. L., Bushman, B. J., Sakamoto, A., & Saleem, M. et al. (2010). Violent video game effects on aggression, empathy, and prosocial behavior in Eastern and Western countries: A meta-analytic review. *Psychological Bulletin*, 136(2), 151–173. doi:10.1037/a0018251 PMID:20192553
- Barlett, C. P., & Anderson, C. A. (2013). Examining media effects: The General Aggression and General Learning Models. In E. Scharrer (Ed.), *Media Effects/Media Psychology* (pp. 1–20). Blackwell-Wiley.

- Bartholow, B. D., Bushman, B., & Sestir, M. A. (2006). Chronic violent video game exposure and desensitization to violence: Behavioral and event-related brain potential data. *Journal of Experimental Social Psychology*, 42(4), 532–539. doi:10.1016/j.jesp.2005.08.006
- Beck, V. S., Boys, S., Rose, C., & Beck, E. (2012). Violence against women in video games: A prequel or sequel to rape myth acceptance? *Journal of Interpersonal Violence*, 27(15), 3016–3031. doi:10.1177/0886260512441078 PMID:22550147
- Bertozzi, E. (2008). “I am shocked, shocked!”: Breaking taboos in digital gameplay. *Loading...*, 3(1).
- Bowman, N. D., & Tamborini, R. (2015). In the mood to game: Selective exposure and mood management processes in computer game play. *New Media & Society*, 17(3), 375–393. doi:10.1177/1461444813504274
- Bryce, J., & Kaye, L. K. (2011). Computer and Videogames. In G. Brewer (Ed.), *Media Psychology* (pp. 101–114). London: Palgrave Macmillan. doi:10.1007/978-0-230-35720-4_7
- Buckley, K. E., & Anderson, C. A. (2006). A theoretical model of the effects and consequences of playing video games. In P. Vorderer & J. Bryant (Eds.), *Playing video games: Motives, responses, and consequences* (pp. 363–378). Mahwah, NJ: Lawrence Erlbaum Associates.
- Burgess, M. C., Stermer, S. P., & Burgess, S. R. (2007). Sex, lies, and video games: The portrayal of male and female characters on video game covers. *Sex Roles*, 57(5-6), 419–433. doi:10.1007/s11199-007-9250-0
- Bushman, B. J., & Anderson, C. A. (2002). Violent video games and hostile expectations: A test of the General Aggression Model. *Personality and Social Psychology Bulletin*, 28(12), 1679–1686. doi:10.1177/014616702237649
- Collins, E., & Cox, A. (2014). Switch on to games: Can digital games aid post-work recovery? *International Journal of Human-Computer Studies*, 72(8-9), 654–662. doi:10.1016/j.ijhcs.2013.12.006
- Cotton, J., Mayes, D., Jentsch, F., & Sims, V. (2001). The relationship between video game characteristics and player ability. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 45(13), 945–947. doi:10.1177/154193120104501310
- Cox, A., Cairns, P., Shah, P., & Carroll, M. (2012). Not doing but thinking: The role of challenge in the gaming experience. In *Proceedings of the SIGCHI Conference on Human Factors* (pp. 79–88). ACM. doi:10.1145/2207676.2207689
- Csikszentmihalyi, M. (1975). *Beyond boredom and anxiety: Experiencing flow in work and play*. San Francisco: Jossey-Bass Publishers.
- De Kort, Y. A. W., IJsselstein, W. A., & Poels, K. (2007). Digital games as social presence technology: Development of the Social Presence in Gaming Questionnaire (SPGQ). *Proceedings of PRESENCE 2007: The 10th International Workshop on Presence*, 195–203.
- DeLisi, M., Vaughn, M. G., Gentile, D. A., Anderson, C. A., & Shook, J. (2013). Violent video games, delinquency, and youth violence: New evidence. *Youth Violence and Juvenile Justice*, 11(2), 132–142. doi:10.1177/1541204012460874
- DeWall, C. N., & Anderson, C. A. (2011). The General Aggression Model. In P. Shaver & M. Mikulincer (Eds.), *Human Aggression and Violence: Causes, manifestations and consequences* (pp. 15–33). Washington, DC: APA. doi:10.1037/12346-001
- Dill, K. E., & Thrill, K. P. (2007). Video game characters and the socialization of gender roles: Young peoples perceptions mirror sexist media depictions. *Sex Roles*, 57(11-12), 851–864. doi:10.1007/s11199-007-9278-1

- Eley, T., Lichtenstein, P., & Moffitt, T. (2003). A longitudinal behavioral genetic analysis of the etiology of aggressive and non-aggressive antisocial behavior. *Development and Psychopathology*, 15(02), 383–402. doi:10.1017/S095457940300021X PMID:12931834
- Elson, M., Breuer, J., Van Looy, J., Kneer, J., & Quandt, T. (2015). Comparing apples and oranges? Evidence for pace of action as a confound in research on violence in digital games. *Psychology of Popular Media Culture*, 4(2), 112–125. doi:10.1037/ppm0000010
- Elson, M., & Ferguson, C. J. (2014). Twenty-five years of research on violence in digital games and aggression: Empirical evidence, perspectives and a debate gone astray. *European Psychologist*, 19(1), 33–46. doi:10.1027/1016-9040/a000147
- Ferguson, C. J. (2007). Evidence of publication bias in video game violence effects literature: A meta-analytic review. *Aggression and Violent Behavior*, 12(4), 470–482. doi:10.1016/j.avb.2007.01.001
- Ferguson, C. J. (2009). Media violence effects: Confirmed truth or just another X-File? *Journal of Forensic Psychology Practice*, 9(2), 103–126. doi:10.1080/15228930802572059
- Ferguson, C. J., & Dyck, D. (2012). Paradigm change in aggression research: The time has come to retire the General Aggression Model. *Aggression and Violent Behavior*, 17(3), 220–228. doi:10.1016/j.avb.2012.02.007
- Ferguson, C. J., & Rueda, S. M. (2010). The Hitman Study: Violent video game exposure effects on aggressive behaviour, hostile feelings, and depression. *European Psychologist*, 15(2), 99–108. doi:10.1027/1016-9040/a000010
- Ferguson, C. J., Rueda, S. M., Cruz, A. M., Ferguson, D. E., Fritz, S., & Smith, S. M. (2008). Violent video games and aggression: Causal relationship or byproduct of family violence and intrinsic violence motivation? *Criminal Justice and Behavior*, 35(3), 311–332. doi:10.1177/0093854807311719
- Gajadhar, B. J., de Kort, Y. A. W., & IJsselstein, W. A. (2009). Rules of engagement: Influence of co-player presence on player involvement in digital games. *International Journal of Gaming and Computer-Mediated Simulations*, 1(3), 14–27. doi:10.4018/jgcms.2009070102
- Gentile, D. A., & Gentile, J. R. (2008). Violent video games as exemplary teachers: A conceptual analysis. *Journal of Youth and Adolescence*, 9(2), 127–141. doi:10.1007/s10964-007-9206-2
- Glock, S., & Kneer, J. (2009). Game Over? The impact of knowledge about violent digital games on the activation of aggression-related concepts. *Journal of Media Psychology: Theories, Methods and Applications*, 21(4), 151–160.
- Goodson, S., & Pearson, S. (2009). Driving beats violence for emotional gaming. *New Scientist*, 2703, 15.
- Greitemeyer, T., & Osswald, S. (2010). Effects of prosocial video games on prosocial behavior. *Journal of Personality and Social Psychology*, 98(2), 211–221. doi:10.1037/a0016997 PMID:20085396
- Harper, W. (1986). Freedom in the experience of leisure. *Leisure Sciences: An Interdisciplinary Journal*, 8(2), 115–130. doi:10.1080/01490408609513064
- Hartmann, T., & Vorderer, P. (2010). Its okay to shoot a character: Moral disengagement in violent video games. *Journal of Communication*, 60(1), 94–119. doi:10.1111/j.1460-2466.2009.01459.x
- Hull, R. B. (1990). Mood as a product of leisure: Causes and consequences. *Journal of Leisure Research*, 22(2), 99–111.
- Information Solutions Group. (2011). *PopCap Social Gaming Research*. Retrieved August 23rd, 2012 from http://www.infosolutionsgroup.com/pdfs/2011_PopCap_Social_Gaming_Research_Results.pdf

- Ivory, J. D., & Magee, R. G. (2009). You cant take it with you? Effects of handheld portable media consoles on physiological and psychology responses to video game and movie content. *Cyberpsychology & Behavior*, 12(3), 291–297. doi:10.1089/cpb.2008.0279 PMID:19445637
- Kaye, L. K. (2014). Football Manager as a persuasive game for social identity formation. In D. Ruggiero (Ed.), *Cases on Societal Effects of Persuasive Games* (pp. 1–17). IGI Global. doi:10.4018/978-1-4666-6206-3.ch001
- Kaye, L. K. (May, 2015). *Social identity as a predictor of self-esteem and psychological well-being in a sample of digital gamers*. Paper presented at the British Psychological Society Annual Conference 2015, ACC, Liverpool, UK.
- Kaye, L. K., & Bryce, J. (2012). Putting the fun factor into gaming: The influence of social contexts on experiences of playing digital games. *International Journal of Internet Science*, 7(1), 23–37.
- Kaye, L. K., & Bryce, J. (2014). Go with the Flow: The experience and affective outcomes of solo versus social gameplay. *Journal of Gaming and Virtual Worlds*, 6(1), 49–60. doi:10.1386/jgvw.6.1.49_1
- Kaye, L. K., Monk, R. L., & Hamlin, I. (in press). “Feeling appy?”: Using app-based methodology to explore contextual effects on real-time cognitions, affect and behaviours. In C. Costa & J. Condie (Eds.), *Doing research in and on the digital: Research methods across fields of inquiry*. Routledge.
- Kneer, J., Munko, D., Glock, S., & Bente, G. (2012). Defending the doomed: Implicit strategies concerning protection of First-Person shooter games. *Cyberpsychology, Behavior, and Social Networking*, 15(5), 251–256. doi:10.1089/cyber.2011.0583 PMID:22515170
- Krahé, B., & Möller, I. (2004). Playing violent electronic games, hostile attribution style, and aggressive-related norms in German adolescents. *Journal of Adolescence*, 27(1), 53–69. doi:10.1016/j.adolescence.2003.10.006 PMID:15013260
- Lim, S., & Lee, J. R. (2009). When playing together feels different: Effects of task types and social contexts on physiological arousal in multiplayer online gaming contexts. *Cyberpsychology & Behavior*, 12(1), 59–61. doi:10.1089/cpb.2008.0054 PMID:19006459
- Limperos, A. M., Schmierbach, M. G., Kegerise, A. D., & Dardis, F. E. (2011). Gaming across different consoles: Exploring the influence of control scheme on game-player enjoyment. *Cyberpsychology, Behavior, and Social Networking*, 14(6), 345–350. doi:10.1089/cyber.2010.0146 PMID:21198364
- Markey, P. M., & Scherer, K. (2009). An examination of psychoticism and motion capture as moderators of the effects of violent video games. *Computers in Human Behavior*, 25(2), 407–411. doi:10.1016/j.chb.2008.10.001
- Mayra, F. (2007). The contextual game experience: On the socio-cultural contexts for meaning in digital play. *Proceedings of Situated Play, DiGRA 2007 Conference*.
- Moscovici, S. (2000). *Social representations: Explorations in social psychology*. Cambridge, UK: Polity Press.
- Nacke, L. E., Niesenhaus, J., Poels, K., Drachen, A., Korhonen, H. J., IJsselstein, W. A., & de Kort, Y. A. W. et al. (2009). Playability and player experience research. *Proceedings of DiGRA 2009: Breaking new ground: Innovation in Games, Play, Practice and Theory*.
- Oswald, C. A., Prorock, C., & Murphy, S. M. (2014). The perceived meaning of the video game experience: An exploratory study. *Psychology of Popular Media Culture*, 3(2), 110–126. doi:10.1037/a0033828
- Persky, S., & Blascovich, J. (2007). Immersive virtual environments versus traditional platforms: Effects of violent and nonviolent video game play. *Media Psychology*, 10(1), 135–156.

- Persky, S., & Blascovich, J. (2008). Immersive virtual video game play and presence: Influences on aggressive feelings and behaviour. *Presence (Cambridge, Mass.)*, 7(1), 57–72. doi:10.1162/pres.17.1.57
- Prensky, M. (2001). *Digital game-based learning*. New York: McGraw-Hill.
- Przybylski, A. K., Deci, E. L., Rigby, C. S., & Ryan, R. M. (2014). Competence-impeding electronic games and players aggressive feelings, thoughts and behaviors. *Journal of Personality and Social Psychology*, 106(3), 441–457. doi:10.1037/a0034820 PMID:24377357
- Ravaja, N. (2009). The psychophysiology of digital gaming: The effect of a non co-located opponent. *Media Psychology*, 12(3), 268–294. doi:10.1080/15213260903052240
- Reinecke, L. (2009). Games and recovery: The use of video and computer games to recuperate from stress and strain. *Journal of Media Psychology*, 21(3), 126–142. doi:10.1027/1864-1105.21.3.126
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development and well-being. *The American Psychologist*, 55(1), 68–78. doi:10.1037/0003-066X.55.1.68 PMID:11392867
- Ryan, R. M., Rigby, C. A., & Przybylski, A. (2006). The motivational pull of video games: A Self-Determination Theory approach. *Motivation and Emotion*, 30(4), 344–360. doi:10.1007/s11031-006-9051-8
- Swing, E. L., & Anderson, C. A. (2008). How and what do video games teach? In T. Willoughby & E. Wood (Eds.), *Children's Learning in a Digital World* (pp. 64–84). Oxford, UK: Blackwell.
- Unsworth, G., Devilly, G. J., & Ward, T. (2007). The effect of playing violent video games on adolescents: Should parents be quaking in their boots? *Psychology, Crime & Law*, 13(4), 383–394. doi:10.1080/10683160601060655
- Young, G., & Whitty, M. T. (2010). Games without frontiers: On the moral and psychological implications of violating taboos within multi-player virtual spaces. *Computers in Human Behavior*, 26(6), 1228–1236. doi:10.1016/j.chb.2010.03.023
- Young, G., & Whitty, M. T. (2011). Should gamespace be a taboo-free zone? Moral and psychological implications for single player video games. *Theory & Psychology*, 21(6), 802–820. doi:10.1177/0959354310378926
- Zillman, D. (1988). Mood management through communication choices. *The American Behavioral Scientist*, 31(3), 327–334. doi:10.1177/000276488031003005

KEY TERMS AND DEFINITIONS

Digital Games: Used here as a generic term to include all electronic games which can be played in arcades, on game consoles, hand-held consoles, PCs, and over the Internet.

Emotional Regulation: One's ability to utilise emotional stability.

Genetic Predisposition: The extent to which an individual is biologically determined to think or behave in a certain way.

Intrinsic Motivation: Undertaking a task or activity for an internal or inherent sense of reward rather than an external reward (e.g., money).

Social Identity: The extent to which an individual defines themselves by their affiliation to a social group (e.g., gamer).

Serious Games Advancing the Technology of Engaging Information



Peter A. Smith

University of Central Florida, USA

Clint Bowers

University of Central Florida, USA

INTRODUCTION

The term Serious Games is an umbrella term that refers to any games that have goals other than pure entertainment. The term grew in popularity in the early 2000s when the Foresight and Governance Project at the Woodrow Wilson International Center for Scholars founded the Serious Games Initiative (SGI). The SGI was founded to pursue the goal of helping to organize and accelerate the adoption of computer games for non-entertainment purposes. This included exploring new applications for games in education, training, recruitment, and beyond. At this time many researchers were beginning to understand that games could have positive effects outside of pure entertainment. In Raph Koster's book, *A Theory of Fun for Game Design* (Koster, 2005) he described the motivating factor of fun in all games as the act of learning. James Paul Gee a well respected games researcher best known for his book, "What video games have to teach us about learning and literacy," focuses on the idea that all good video games exhibit thirty-six learning principles supported by literature in learning science and cognition research (Gee, 2007). While Serious Games are not based solely on the idea that games can teach, the principles behind good game design actually support learning. As a result, the research has shown that Serious Games are not just another media for learning through a passive act of absorbing material, but are a technology for engaging with information.

Games researchers are now moving from exploring if games can teach to how games teach. The caveat is that not all games teach but that all good games teach. Leaving a simple truth, it is hard to make a good game, no less a good game that is also educational. The real challenge is getting the people with the right design abilities to make these types of games and establish best practices and quantify what actually makes games as educational systems work. Efforts to move in that direction must begin with establishing terms and defining a framework for what goes into games for learning as formal systems.

BACKGROUND

Before the more modern notion of Serious Games took hold, the military made many attempts at using video games for training. The earliest being in 1980 when the Army commissioned Atari to build the Atari Bradley Trainer (P. Smith, In Press). This game was a modified version of the popular vector graphics based game *Battlezone*, also published in 1980. Only 2 Atari Bradley Trainers were ever built and shown at a trade show. It is unknown why the Army never deployed the game, but it was never actually used by soldiers.

Another military project was started by 1984, this time by the Navy, to use a video game to teach Morse Code (Driskell & Dwyer, 1984). This project also only made it through the pro-

totyping phase. The military's view of games at the time was that they were not serious enough for military training, though the problem seemed to be one of vocabulary only. This is illustrated by the Marines common use of games under the name, Tactical Decision-making Simulations (TDS) since development of the game Marine Doom in 1996 (P. Smith, 2005). Marine Doom is a modification (mod) of the popular first person shooter game Doom, and was created by the Marine Corps Modeling and Simulation Management Office (MCMSMO) developed for the training of Marine fire teams.

This prejudice against video games didn't carry over to the common practice of table top War Gaming, or the use of Flight Simulator Software on PC's, which were sold as games to the rest of the world. The military did not seem completely ready to embrace games for training until after DARPA created DARWARS Ambush, a mod to the game Operation Flashpoint, which was followed up by the Army creating TRADOC Capabilities Manager for Gaming (TCM Gaming) and deploying Virtual Battle Space 2 (VBS2) as one of many official Army Games in 2008. However this prejudice persisted after Serious Games were well established outside of the department of defense. (R. Smith, 2009)

Paralleling the emergence of games in the military is the development of the ill fated Edutainment market. In the early 1980s Edutainment games became an incredibly popular trend. These games, such as "Where in the World is Carmen Sandiego," "The Oregon Trail," "Reader Rabbit," "Math blaster," among many others flooded the market with games that contained some level of educational content. Mizuko Ito described it as a time where the developers were empowered with a "sense that they were creating possibilities for learning that freed it from the institutional constraints of schooling." (Ito, 2006).

Edutainment games succeeded in capturing an audience, and establishing itself as an accepted

part of the games industry, however, they never quite got established as a credible form of education. Ito, suggests that the reason behind this is that, "edutainment embodies the challenges which reformers face in creating new genres of representation and practice..." (Ito, 2006). However, the answer is much simpler. In general the games did not achieve the dual goals of being good educational platforms while also being good games. Edutainment, along with many of the other past attempts to develop learning games, have largely been deemed failures. A sentiment best stated by Michael Zyda, the Director of the Game Pipe Lab at USC, "The game industry has already witnessed the failure of edutainment, an awkward combination of educational software lightly sprinkled with game-like interfaces and cute dialog. This failure shows that story must come first and that research must focus on combining instruction with story creation and the game development process." (Zyda, 2005)

Clark C Apt's book, *Serious Games*, was published in 1970 and represents the first recorded use of the term Serious Games (Apt, 1970). The term Serious Games was not, however, an instant success. In the 30 years that followed, serious games had a few false starts on the road to becoming a main stream part of the non-entertainment world, the most dramatic of these being in both the education and training arenas.

Clark C. Apt defined Serious Games as games that "have an explicit and carefully thought-out educational purpose and are not intended to be played primarily for amusement" (Apt, 1970). Apt wrote these words over thirty years before the founding of the SGI but his words are still relevant and extremely close to the current definition that most game scholars adhere to for serious games. The one inconsistency of his definition is that serious games have evolved to include more applications than just education. Serious games are commonly defined as some derivation of a game designed for a primary purpose other than

pure entertainment. This definition is purposefully open ended in order to allow for the diverse backgrounds of various serious game practitioners.

Mike Zyda, the Director of GamePipe at USC, defined serious games as: “a mental contest, played with a computer in accordance with specific rules that uses entertainment to further government or corporate training, education, health, public policy, and strategic communication objectives” (Zyda, 2005). His particular definition met his vision of what a serious game could be, but others whose application of serious games do not fit into the categories defined were still searching for a definition. Further still, many industries utilize gaming technology but do not explicitly create games with the technology, yet have aligned themselves with the serious games movement.

In an effort to move towards a more open ended understanding of Serious Games, the term became an umbrella term that encompassed the efforts un-

derway, bringing all the groups working on Serious Games under one unified vision. The Taxonomy of Serious Games (Table 1.) was presented at the Serious Games Summit (SGS) held at the 2008 Game Developers Conference (GDC). It defines the current categories of games that have been developed by the Serious Games industry including, Games for Health, Advergames, Games for Training, Games for Education, Games for Science and Research, Games for Production, and Games as Work. It further cross references them with the industries that currently use Serious Games. Further slides show the amount of development in each category, illustrating that most of the work in the Serious Games Space was being done for education and training in both schools and the military. (Sawyer, Smith, P., 2008)

The success of the Taxonomy as a reference for where the market had evolved did not, unfortunately, meet the needs of everyone in the

Table 1. The Taxonomy of Serious Games

	Games for Health	Adver-Games	Games for Training	Games for Education	Games for Science Research	Production	Games as Work
Government & NGO	Public Health Education & Mass Casualty Response	Political Games	Employee Training	Inform Public	Data Collection / Planning	Strategic & Policy Planning	Public Diplomacy, Opinion Research
Defense	Rehabilitation & Wellness	Recruitment & Propaganda	Soldier/ Support Training	School House Education	Wargames / planning	War planning & weapons research	Command & Control
Healthcare	Cybertherapy / Exergaming	Public Health Policy & Social Awareness Campaigns	Training Games for Health Professionals	Games for Patient Education and Disease Management	Visualization & Epidemiology	Biotech manufacturing & design	Public Health Response Planning & Logistics
Marketing & Communication	Advertising Treatment	Advertising, marketing with games, product placement	Product Use	Product Information	Opinion Research	Machinima	Opinion Research
Education	Inform about diseases/risks	Social Issue Games	Train teachers / Train workforce skills	Learning	Computer Science & Recruitment	P2P Learning Constructivism Documentary?	Teaching Distance Learning
Corporate	Employee Health Information & Wellness	Customer Education & Awareness	Employee Training	Continuing Education & Certification	Advertising / visualization	Strategic Planning	Command & Control
Industry	Occupational Safety	Sales & Recruitment	Employee Training	Workforce Education	Process Optimization Simulation	Nano/Bio-tech Design	Command & Control

community. Debate on the appropriate categories continues. While criticisms can be made of final categorization, the model persisted and it is commonly understood that Serious Games cannot be defined as a single type of game or a particular field.

UNDERSTANDING SERIOUS GAMES THROUGH UNDERSTANDING GAMES

Despite the strides Serious Games have taken in recent years they are considered subpar as games by many due to the lack of an established design methodology, general development guidelines, and underprepared designers. Jacob Habgood of The University of Nottingham's Learning Science Research Institute labeled them to be "Chocolate-Covered Broccoli" due to their poor marriage of games and learning (Habgood, 2005). His suggestion to solve this is a tighter integration between game mechanics and the learning content or what he terms as Intrinsic Integration (Habgood, Ainsworth, & Benford, 2005). This is a sentiment shared by NavAir's Dr. Robert Hays in his game based research literature review (Hays, 2005).

Both Habgood and Hays understood that for good learning outcomes to occur gaming characteristics or features needed to support instructional objectives completely. This of course means that learning games need to go beyond the "lightly sprinkled... game-like interfaces and cute dialog" Zyda (2005) used as a charge against edutainment. They need game mechanics that support both gameplay and learning at the same time. Game mechanics are, "mechanisms through which players make meaningful choices and arrive at a meaningful play experience" (Salen & Zimmerman, 2004). In a book written by Adams and Dormans (2012), mechanics are referred to as feedback loops and some are specifically used to design the internal economy of games to make them challenging. These economies are a basic

exchange of resources between player and system and can create deadlocks and mutual dependencies.

In order to map these mechanics to learning outcomes it is important to understand what features of games support what types of mechanics. Further in order to insure those games support learning it is important to have an understanding of what features of games support what learning outcomes. By mapping these features against each other, an understanding of how game mechanics map to learning outcomes can be gained.

Features of Entertainment Games

Before being able to determine the features of games that lead to better learning, it is important to first identify the features that fundamentally define a game. Unfortunately, there isn't one agreed upon definition that everyone in the game industry uses. Further, the definitions that are used seem to vary widely.

Many definitions are far too simple to describe all games well. One of the most oft mentioned definitions is Sid Meier's declaration that, "A game is a series of interesting decisions." (Bateman, 2008). While this might be true of good strategy games, the type of games Meier is known for, this does not include simple twitch or rhythm games, where the player is tasked with maintaining good timing but is limited by the decisions they can make. The definition cited by Jane McGonigal, a well-known proponent of gamification, in her book, *Reality is Broken*, is one by the philosopher Bernard Suits. He stated, "Playing a game is the voluntary attempt to overcome unnecessary obstacles," (McGonigal, 2011; Suits, 1978). This definition, while open ended enough to justify gamification as a legitimate type of game, does not provide enough details to even meet the features McGonigal suggests in her book, which include: goals, rules, feedback, and voluntary participation (McGonigal, 2011). Though this set of features is already flawed as many educational games are compulsory to courses.

The game designer whose games are most often cited when the question of what makes a game versus a simulation is Will Wright. His games include *SimCity*, *The Sims*, and *Spore*, but he doesn't consider them games; he describes them as toys. "People call me a game designer, but I really like to think of these things more as toys," (Wright, 2007).

One of the earliest game scholars, Johan Huizinga, defined games in his book, *Homo Ludinus*, as "...a free activity standing quite consciously outside ordinary life as being not serious, but at the same time absorbing the player intensely and utterly... according to fixed rules and in an orderly manner," (Huizinga, 1949). Jesper Juul provided a definition that attempted to encompass the various views one could take on games. In doing so, he categorized definition in the categories of: the game as a formal system, the player and the game, the game and the rest of the world, and other (Juul, 2003).

A good definition for those interested in Serious Games would consider a game as a formal system that can be applied to learning. As such, the definition of games used moving forward here will be Katie Salen and Eric Zimmerman's definition provided in their book *Rules of Play*. Salen and Zimmerman, like many other game researchers, developed their definition through a thorough analysis of various definitions. In particular, they used a total of eight leading definitions that were suitably feature rich. Three of these definitions have already been discussed; all of them are by leading games researchers or designers themselves. By carefully comparing these definitions and analyzing their meaning, they settled upon their definition and in doing so they have provided a workable framework for the features that make a game. (Salen & Zimmerman, 2004)

Using this framework as a guide, Salen and Zimmerman defined a game as: "A game is a system in which players engage in an artificial conflict, defined by rules, that results in a quantifiable outcome." (Salen & Zimmerman, 2004). Using this definition, a game can be broken up

into the following set of core features: System, Players, Conflict, Rules/Goals, Outcomes, and the Artificial.

Features of Serious Games

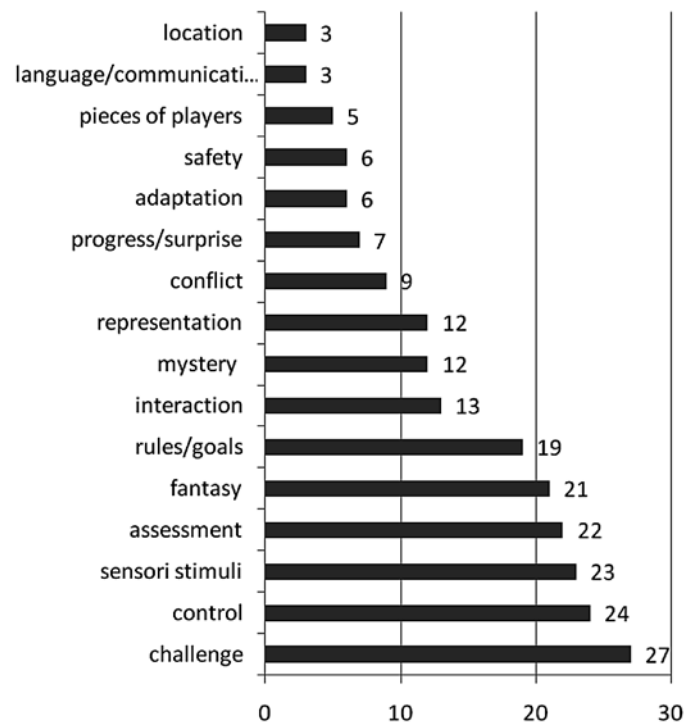
Learning games at their core can and should be considered games, and as games they should exhibit the same features of games that define games themselves. Therefore determining a list of features that defines learning games is a redundant process. The interesting task is in determining the features of games that support learning.

In a comprehensive review of over 41 papers performed in 2009, researchers identified over a dozen features of games that would support learning (Wilson et al., 2009). Upon closer inspection of the data provided, some features overlap with each other. Others are along considered features of games by a miniscule fraction of the 41 researchers. As seen in Figure 1, over 27 researchers found challenge to be an important feature for learning in games, while only 3 suggest location is important. For this research, only features agreed upon by 10 or more researchers will be considered agreed upon features. Further, overlapping features will be combined.

In particular, the feature of interaction, or the ability for the player to interact with the game, will be combined with control. Control is the ability for the player to maintain control of the flow of the game, an activity accomplished through interaction.

Another overarching feature of games is their aesthetic feel. This feature is represented by mystery, fantasy, representation, and sensory stimuli. The Aesthetics determine if a game provides a mystery to unravel, if the game is fantasy-based or based in realistic representation of reality, and they are responsible for the type and form of sensory stimuli provided to the player. Once repeated features are consolidated and fringe features are removed we are left with the following list: Interaction, Challenge, Rules/Goals, Assessment, and Aesthetics.

Figure 1. A Comparison of Game Features Supporting Learning



Complementary Features

The list of features that game designers have determined to define games share a remarkable level of overlap with the list of features learning games researchers determined support learning.

Interaction / Players

Interaction is a key feature of games. Games are to be played by players, not observed or reported on. Interaction is sometimes defined by the players themselves, “Players interact with the system of a game in order to experience the play of the game” (Salen & Zimmerman, 2004). Other times it is defined through the type of hardware, such as a game controller, or mouse and keyboard combinations. Interaction can occur in many ways, and through many mechanisms. The interaction and control of a game is often dependent on the core game mechanic of physics where it is “used to test the player’s dexterity, timing and accuracy,”

(Adams and Dormans, 2012). Thus, the most efficient controller to use would be one with the fastest and most reliable feedback between player and game objects.

Challenge / Conflict

“All games embody a contest of powers. The contest can take many forms, from cooperation to competition, from solo conflict with a game system to multiplayer social conflict. Conflict is central to games” (Salen & Zimmerman, 2004). While Salen & Zimmerman prefer to call it conflict, they have captured the essence of what this research will refer to as challenge. Challenge can be cooperative, competitive, or individualistic. It can also be a combination of any of the three. For example team versus team challenge has competition with inter team competition. Highly competitive E-sports are prime examples of how games can implement the concept of competition.

Rules/Goals / Rules

Wilson coupled the terms rules and goals into a single feature, while Salen & Zimmerman refer to only rules. “Rules provide the structure out of which play emerges, by delimiting what the player can and can-not do” (Salen & Zimmerman, 2004). They instead associate goals with outcomes. Goals in particular are difficult to separate from other features, and are possibly their own feature.

Assessment / Outcomes

“Games have a quantifiable goal or outcome. At the conclusion of a game, a player has either won, lost, or received some kind of numerical score” (Salen & Zimmerman, 2004). Games must provide feedback to the player in the form of an assessment or outcome. While the assessment might not be explicit to the player in all games, the game must provide the appropriate outcome based on the performance of the player in the game.

Aesthetics / Artificial

“Games maintain a boundary from so-called “real life” in both time and space” (Salen & Zimmerman, 2004). Games provide aesthetics to the player. This might be fantasy or reality based. It might mean a deep storyline filled with interesting plot twists or it might mean a simple song and falling blocks. While the Aesthetic is important, it can vary widely between games and can be manipulated using different gaming mediums, such as 2D vs. 3D, and Virtual Reality

Disparate Features

All five of the core features identified by learning games researchers have direct definitional overlap with features that game designers identified. The one feature game designers identified that learning games researchers did not is the concept of a system.

Salen & Zimmerman use a definition of system that is taken from Stephen W. Littlejohn, and include 4 defining elements. They are objects, attributes, internal relationships, and environment (Littlejohn & Foss, 2007; Salen & Zimmerman, 2004). Using this definition, Salen & Zimmerman further define systems as simulations (2004). The system is the core structure of a game. It determines how the environment works and what types of objects can operate within it. When other features are applied to it the game becomes fully formed. This is important when considering the game versus simulation debate. It is easy to imagine that a game is a simulation with others gaming features added to it. This, however, is not the case.

In his 2010 keynote address to the GameTech Conference, Will Wright defined his particular types of games as toys. These toys, SimCity, The Sims, Spore, among others, are commonly used to frame the argument between what is a game and what is a simulation. Will Wright further stated that his toys exist as a constraint of freeform play, and if they were constrained more they could be considered games. He did not want to constrain them with preformed goals, outcomes, or challenges (Wright, 2010). Taking a similar approach to the space of modeling and simulation one could consider a constructive simulation as a constraint of the space, but a simulation still provides a level of freeform use that makes it similar to how Will Wright refers to toys in the space of play. Simulations need a facilitator to add goals, outcomes, and challenges to the exercise. Further constraining a simulation by introducing gaming features may in fact result in a Serious Game.

FUTURE RESEARCH DIRECTIONS

Future research in Serious Games should focus on how the design of Serious Games can be used to improve learning outcomes. While research has focused on various features of Serious Games new developers have emerged that are building very

personal and informational games that combine the power of well-designed good games and a personal passion for the topic that the game covers. These games are often made for entertainment purposes and happen to be very effective serious games. These include Dys4ia, That Dragon Cancer, Tampon Run, and countless others, that cover incredibly complex and sensitive topics in new ways. The future of Serious Games may be found in the design of personal stories that explore both education and game design from much more personal view.

CONCLUSION

Current research in Serious Games has focused on exploring how the features of Serious Games effect learning outcomes. Serious Games have been found to outshine other forms of educational technology in engagement, motivation, goal orientation, and many other categories, but only when properly designed. It is important to determine if narrative can improve learning outcomes, or time on task, and this work is happening. With an understanding of how gaming features work, serious game developers can build best practices and develop even more effective and engaging games. More importantly more game designers are exploring the power of Serious Games and more great examples exist every day.

REFERENCES

- Adams, E., & Dormans, J. (2012). *Game mechanics: advanced game design*. New Riders.
- Apt, C. C. (1970). *Serious Games*. Academic Press.
- Bateman, C. (2008). *A Game Isn't a Series of Interesting Decisions*. Retrieved from http://onlyagame.typepad.com/only_a_game/2008/07/a-game-isnt-a-series-of-interesting-decisions.html
- Csikszentmihalyi, M. (1991). *Flow: The psychology of optimal experience*. Harper Perennial.
- Driskell, J. E., & Dwyer, D. J. (1984). Microcomputer Videogame Based Training. *Educational Technology*, 24(2).
- Gee, J. P. (2007). *What video games have to teach us about learning and literacy*. Palgrave Macmillan.
- Habgood, J. (2005). *Zombie Division: Intrinsic integration in digital learning games*. Cognitive Science Research Paper-University of Sussex CSRP.
- Habgood, M., Ainsworth, S., & Benford, S. (2005). Endogenous fantasy and learning in digital games. *Simulation & Gaming*, 36(4), 483–498. doi:10.1177/1046878105282276
- Hays, R. T. (2005). *The effectiveness of instructional games: A literature review and discussion*. DTIC Document.
- Huizinga, J. (1949). *Homo ludens: A study of the play-element in culture*. Taylor & Francis.
- Ito, M. (2006). Engineering Play: Childrens software and the cultural politics of edutainment. *Discourse (Abingdon)*, 27(2), 139–160. doi:10.1080/01596300600676003
- Koster, R. (2005). *A Theory Of Fun For Game Design*. Paraglyph press.
- McGonigal, J. (2011). *Reality is broken: Why games make us better and how they can change the world*. Penguin.
- Salen, K., & Zimmerman, E. (2004). *Rules of play: Game design fundamentals*. The MIT Press.
- Sawyer, B. (2007). *Ten Myths About Serious Games*. Retrieved from http://www.escapistmagazine.com/articles/view/issues/issue_121/2575-Ten-Myths-About-Serious-Games
- Sawyer, B., & Smith, P. (2008). *Serious games taxonomy*. Academic Press.
- SeriousGames.org. (2011). *About The Serious Games Initiative*. Retrieved from www.serious-games.org/about.html

Smith, P. (2005). Serious Game: Opportunities for Game Developers in Interactive Simulation and Training. In F. Lamaree (Ed.), *Secrets of the Games Business* (2nd ed.). Hingham, MA: Charles River Media, Inc.

Smith, P. (In Press). Battlezone: The Rise of Serious Games. In I. Bogost & M. Bittanti (Eds.), *Ludologica Retro*. Milan, Italy: Ludologica, Videogames d'Autore.

Smith, R. (2009). *A History of Serious Games*. Paper presented at the I/ITSEC 2009, Orlando, FL.

Suits, B. (1978). *2005: The Grasshopper: Games, Life and Utopia*. Toronto: University of Toronto Press.

Wilson, K. A., Bedwell, W. L., Lazzara, E. H., Salas, E., Burke, C. S., Estock, J. L., & Conkey, C. et al. (2009). Relationships between game attributes and learning outcomes. *Simulation & Gaming*, 40(2), 217–266. doi:10.1177/1046878108321866

Wright, W. (2007). *Will Wright makes toys that make worlds*. TED2007. Retrieved from http://www.ted.com/talks/will_wright_makes_toys_that_make_worlds.html

Wright, W. (2010). *SimCity Designer Presents GameTech 2010 Keynote*. Retrieved from <http://science.dodlive.mil/2010/05/07/simcity-designer-presents-gametech-2010-keynote-video/>

Zyda, M. (2005). From visual simulation to virtual reality to games. *Computer*, 38(9), 25–32. doi:10.1109/MC.2005.297

ADDITIONAL READING

Alhadeff, E. (2012). Serious Games Market: Digging up the best of Serious Games Fast Growing Market. [Blog] Retrieved from <http://serious-gamesmarket.blogspot.com>

Arnab, S., Lim, T., Carvalho, M. B., Bellotti, F., Freitas, S., Louchart, S., & De Gloria, A. (2015). Mapping learning and game mechanics for serious games analysis. *British Journal of Educational Technology*, 46(2), 391–411. doi:10.1111/bjet.12113

Becker, K. (2010). Why Educational Games are still boring... Retrieved from <http://minkhollow.ca/beckerblog/2010/06/08/why-educational-games-are-still-boring/> Game Developers Conference (2013). Retrieved online from <http://www.gdconf.com>

Conkey, C., Raybourn, E., & Smith, P. A. 2012. Serious Games Design and Development. I/ITSEC 2012. Orlando FL.

Delora, M. (2015) Level Up Report. Retrieved Online from <http://www.levelupreport.com>

Game Tech Conference. (2013). Retrieved online from <http://gametechconference.com>

Games for Change. (2015). Retrieved online from <http://gamesforchange.com>

Glass Lab. (2015). Retrieved online from <http://www.glasslabgames.org>

I/ITSEC Conference. (2013). Retrieved online from <http://www.iitsec.org>

McDowell, P., Cannon-Bowers, J., & Prensky, M. (2005). The Role of Pedagogy and Educational Design in Serious Games. Paper presented at the Serious Games Summit DC 2005, DC.

McNamara, J., Smith, P. A., Smith, B., & Gritton, K. 2012. Serious Games Showcase & Challenge Distilled. I/ITSEC 2012. Orlando FL.

Miller, C. H. (2004). Digital storytelling: a creator's guide to interactive entertainment: Focal Pr.

Nealen, A., & Boxerman, E. (2010). *Minimalist Game Design: Growing OSMOS*. Independent Games Summit.

Prensky, M. (2000). Digital Game Based Learning: McGraw-Hill Companies.

- Quick, J., & Atkinson, R. (2011). A Data-Driven Taxonomy of Undergraduate Student Video Game Enjoyment.
- Serious Games Association. (2015). Retrieved online from <http://www.seriousgamesassociation.com>
- Serious Games Research. (2015). Retrieved online from <http://www.seriousgamesresearch.com>
- Serious Games Showcase & Challenge. (2015). Retrieved online from <http://www.sgschallenge.com>
- Serious Games Society. (2015). Retrieved online from <http://www.seriousgamesociety.org>
- SGames conference. (2015). Retrieved online from <http://sgamesconf.org/>
- Singer, M. J., & Knerr, B. W. (2010). Evaluation of a Game-Based Simulation during Distributed Exercises: DTIC Document.
- Smith, P. A. 2012. Introduction to Serious Games. GameTech 2012. Orlando, FL.
- Smith, P. A., & McNamara, J. 2012. Top 10 Lessons Learned from the Serious Games Showcase & Challenge. GameTech 2012. Orlando, FL.
- Smith, P. A., & Sanchez, A. (2010). Mini Games to save the world. In C. Bowers & J. A. Cannon Bowers Serious Games. In J. Cannon-Bowers & C. Bowers (Eds.), *Serious Game Design and Development: Technologies for training and learning* (pp. 1-12). Hershey: IGI Global.
- Squire, K. (2004). *Replaying history: Learning world history through playing Civilization III*.
- Swink, S. (2008). *Game Feel: a game designer's guide to virtual sensation*. Morgan Kaufmann.
- Toppo, G. (2015). *The Game Believes in You: How Digital Play Can Make Our Kids Smarter*. St. Martin's Press.
- Vault, G. D. C. (2013). Retrieved online from <http://www.gdcvault.com>
- Virtual Heroes. (2015). Retrieved online from <http://www.virtualheroes.com/>
- Vogel, J. J., Vogel, D. S., Cannon-Bowers, J., Bowers, C. A., Muse, K., & Wright, M. (2006). Computer gaming and interactive simulations for learning: A meta-analysis. *Journal of Educational Computing Research*, 34(3), 229–243. doi:10.2190/FLHV-K4WA-WPVQ-H0YM
- Wexler, S., Aldrich, A., Johannigman, J., Oehlert, M., Quinn, C., & Barneveld, A. (2007). *Immersive learning simulations: The demand for, and demands of, simulations, scenarios, and serious games*. Santa Rosa, CA: The eLearning Guild.
- Willoughby, T., Adachi, P. J. C., & Good, M. (2011). A longitudinal study of the association between violent video game play and aggression among adolescents.
- Woodman, M. D. (2006). *Cognitive training transfer using a personal computer-based game: A close quarters battle case study*. Florida: University of Central Florida Orlando.

KEY TERMS AND DEFINITIONS

Edutainment: A name given to games developed during a briefly successful attempt to popularize educational games in the 80s.

Game: A game is a system in which players engage in an artificial conflict, defined by rules, that results in a quantifiable outcome” (Salen & Zimmerman, 2004).

Game Feature: An underlying component that defines what a game is.

Games For Change: Games with the primary goal of social change.

Learning Games: Games with the primary goal of education or training.

Serious Games: An umbrella term referring to any game developed for a non-entertainment based purpose.

Tactical Decision-Making Simulation: A term used by the marines to denote a Serious Game.

Towards Modelling Effective Educational Games Using Multi-Domain Framework



Mifrah Ahmad

Universiti Teknologi PETRONAS, Malaysia

Lukman Ab Rahim

Universiti Teknologi PETRONAS, Malaysia

Kamisah Osman

Universiti Kebangsaan Malaysia, Malaysia

Noreen Izza Arshad

Universiti Teknologi PETRONAS, Malaysia

INTRODUCTION

The use of game-based learning (GBL) as a common medium for educational deliverance, as opposed to pure entertainment, has gained immense popularity in recent years. Main attention has recently been diverted towards the impact of gaining knowledge, engaging, and motivating learners through playing educational games (Azadegan et al., 2014; Romero, 2015). Game-based learning (GBL) is an act of appropriate game mechanics, scenario recreation, and problem oriented learning processes to ensure learning objective is accomplished (Poulsen, 2011). Designers and developers need to enhance the educational tools by integrating game domains and elements to maximize the tools' effectiveness, hence, increase learning outcomes, level of engagement and motivation. To understand the link between multiple domains and elements of GBL, there is a need to fit multiple widely known instructional approaches with experts such as game designers, developers, educators, and software engineers.

For GBL development process to be effective and reliable, it is vital to provide emphasis on involving theoretical foundations with game rules, and fun with authentic learning for lean-

ers, thus, providing an precise base for learners to gain knowledge (Hays, 2005; Kebritchi & Hirumi, 2008). The innovative learning approach derived from EG possess educational values or even different kinds of software applications that compiles into knowledgeable aspects such as teaching enhancement, assessments and evaluation of learners (Tang & Hanneghan, 2010). Nevertheless, game technologies are specifically used for accessibility of simulated and modelling environments and visualization.

Educational games are designed through a process of modelling, depending on diverse criteria such as how one element can relate to another element or types of domains used in EG. When the game developers initiate the development of EG, challenges occurs in terms of planning from gathered requirements, verifying, and cross-check to ensure several possibilities occurrence. This is due to a mixture of pedagogical, educational, and fun elements needed to be collaborated accurately. Furthermore, an effectually designed pre-model unfolding the relationships amongst game elements, and their domains is highly recommended. Therefore, a modelling technique is needed to conceptualise the elements/components in the EG and provide a holistic idea based on

how domains and their related components can be modelled to provide a promising modelling and developing process.

The objective of this chapter is to study the relationships among EG elements in order to provide game developers, software engineers, and game designers; a medium of understanding connections, interrelations, and interactions between game elements and game domains. Furthermore, they should be able to map the relationships to model out an effective educational games during developing process.

BACKGROUND

This section mainly explains how the evaluation and modelling strategies adopted or proposed in designing of EG to provide a unified modelling techniques during development process of EG.

EG Evaluation and Modelling Languages

GBL evaluation is carried out to ensure design quality (Di Loreto & Gouaïch, 2010), identify usage (Djelil, Sanchez, Albouy-Kissi, Lavest, & Albouy-Kissi, 2014), and verify instructional outcomes (Casey, Baghaei, & Nand, 2014). However, the complex nature of GBL is not well-structured (Djelil et al., 2014). Certainly, the evaluation has to undergo a definitive number of steps to measure variables and provide appropriate analysis of these variables. Although not all evaluations provide coherent results (Annetta et al., 2013; Von Wangenheim, Thiry, & Kochanski, 2009), a few existing methods for data collection and analysis have proven their efficiency and reliability in practice. Even though, the evaluation of EG is time consuming and complex; it is the only reliable aspect to verify the goals and detect malfunctions throughout the EG. Therefore, a learning game ought to be evaluated prior to being used as a learning material (Djelil et al., 2014). A recent study implements a six-phase methodol-

ogy, HEXA-GBL, for designing and evaluating GBL activities from learner centered perspective (Romero, 2015). The phases are: 1) game design activity, 2) learning objectives definition, 3) the learner-centered need analysis, and the definition of the game modalities, 4) mechanics and rules, 5) the play activity evaluation from the learning outcomes, assessment and feedback, and 6) learners' gaming and learning experience during the GBL activity. While overcoming the barriers of GBL costs and focusing on game mechanics, HEXA-GBL also prioritized the educators to operationalize with adapting processes.

Modelling processes can be a powerful 'tool' to boost student's perceptive activities, hence, refining their scientific conceptual minds (de Jong & van Joolingen, 2008). A recent and a wide range of discussion of a potential value using model-driven engineering (MDE) approaches for EG has been recorded (Dormans, 2012; France & Rumpe, 2007). Furthermore, UML-based game specifications such as Statecharts (Sauer & Engels, 2001); offer a rigorous state machine foundation, which may be difficult to use for some stakeholders (e.g., game designers). With respect to GBL, MDE can provide an environment for domain experts to produce EG via modelling without upsetting the sophistication of game development process. Moreover, it provides an increased productivity value, portability among different platforms, an easier software maintenance, reliability of mapping from model to code, and minimized modelling error (Kelly & Tolvanen, 2008).

Game designers experienced in using Storyboards prefer using tabular Use Case (UC) method (MDE) due to the similarity of presentation (Truong, Hayes, & Abowd, 2006); as it is understood, well-established, maintainable, and straightforward to define. A recent research used MDE to integrate elements of game design, pedagogical content, and software engineering methodologies to provide a storyboard with textual description of the learning objectives and game play to user interfaces. Due to iterative development process, it transforms the model into UML UC model (visual

and tabular based specifications (Stevens, 2002)), then, it converts it into formal model which can then be executable (Cooper, Nasr, & Longstreet, 2014). Although Cooper et.al, provides various modelling techniques are used in this approach, the need of a complete new solution is not evident.

In game development field, the use of software frameworks and tools are usual practice among professional (commercial) game developers. Another modelling technique, namely, Model Driven Architecture (MDA) (Miller & Mukerji, 2001), has also been implemented in Domain-Driven Software Development Framework (Agrawal, Karsai, & Lédeczi, 2003) and Modelling Turnpike (Wada, Suzuki, & Oba, 2005). These frameworks aid software architects to develop their own MDE to suit a particular domain (security, content repurposing, software testing, and pervasive computing). Although current tools improve the productivity of the development team while providing maximum control and flexibility to artistically craft the game software, the production pipeline is still very reliant on specialist artists and programmers. This therefore presents a practical solution to assist non-technical domain experts in the production of serious games.

Educational Game Design Frameworks and Models

There has been quite a number of models and frameworks proposed by researchers in developing, modelling and designing EG. Game Object Model II (GOM II) describes relationships containing various elements of EG interfaces by using Object Oriented Programming to provide a constructive way of achieving theoretical and pedagogical goals in designing EG (Amory, 2007). However, GOM II does not represent how the relationship between game elements develops over time, and its diagram can become complex and difficult to understand. Another model, Mechanics-Game Mechanics (LM-GM) model provides a graphical representation of game flow to establish relationships among components which

translates pedagogical learning mechanics into game mechanics (Arnab et al., 2015). Although it identifies abstract patterns and predefined elements to be replicated across EG, it does not expose the connection with the educational objectives to allow sustainable mechanics.

A recent model proposed with seven steps to identify attributes with specific roles and steps to each expert including cognitive and pedagogical experts, storyboard writers, artistic directors, actors, graphic designers, sound managers, etc. (Marfisi-Schottman, George, & Tarpin-Bernard, 2010). However, it is sequential based and they do not easily fit into an iterative design. Another game-based learning framework works as an evaluation tool for designers by adopting instructional and cognitive approaches in designing process to map out existing games (van Staaldunin & de Freitas, 2010). This approach is beneficial for repurposing game content and providing scalable immersive game content, however, more work is needed to assess the validating of game-design strategies such as links between components and consistency.

When model-driven approaches (MDA) are used in developing process of EG, there are benefits and challenges at all aspects (Sauer & Engels, 2001). A Model-Driven Game Development Approach (Reyno & Cubel, 2008) introduces a framework which uses the selection of UML diagrams to gather requirements into Class Diagrams which is then extended to model the relationships between different game entities and their behavior through the game actions. Another framework based on MDA (Altunbay, Metin, & Çetinkaya, 2009) uses UML Class Diagrams as the modelling language to represent the game model with the combination of languages such as Meta-model with UML Class Diagram and Domain-specific language. All the three frameworks based on MDA have similarities in terms of modelling but they differ in terms of how non-technical domain expert can deal with the modelling languages used. With regards to that concern, MDA based framework (Tang & Hanneghan, 2010) designed

to help non-technical domain experts in producing EG. Although the gap of non-technical experts is covered and the creation of game models can be represented, a challenge of having defined and explicit model while maintaining the consistency is an overlapping issue.

Another way to design EG is by providing a language using Design Patterns based on the Six Facets of Serious Game Design, namely, pedagogical objectives, domain simulations, interactions with the simulation, problems and progressions, decorum, and conditions of use (Marne, Wisdom, Huynh-Kim-Bang, & Labat, 2012). This way, the game designers and teachers can easily relate to join the purpose of developing EG as they will communicate their ideas thoroughly. In contrast, the Educational Computer Game Design Model was proposed with game design components as Learn and Play, so that a balanced integration can construct good EG design (Osman & Bakar, 2012). Another author provided a conceptual framework for adapting collaborative multiplayer games by adopting the concept of multiplayer games and gameplay design principles (Sinkewicz, 2015). First, it is built on an existing model of players to provide insights of the audience, then, it develops a typology of gameplay themes to help designers with conceptualizing actions on the screen. Finally, it provides a framework with five main game design components, namely, learning objectives, story, 3D world, gameplay, and evaluation.

Recently, a framework is proposed for a blended learning environment to enhance learning outcomes and to provide a valuable strategy to facilitate both lecturers and students to obtain better educational knowledge (Jing, Yue, & Murugesan, 2015). Although it has proven a positive learning outcome, there is a need of more research work that can validate the learning-driven game design strategy. In contrast, an Activity Theory-based Model of Serious Games is proposed to provide a way to reason about the relationships between serious game components and the educational goals of the game it includes gaming activity, learning activity, and the instructional activity

(Arnab et al., 2015). The model explicitly accounts for the distinction between designer and instructor, therefore, the clarification of the role of the teacher/instructor in the game is needed. The game sequence representation follows the unified modeling language (UML) activity diagrams notation, which uses the shapes connected by arrows to represent the flow of the activities (Kühn, 2015).

MAIN FOCUS OF THE CHAPTER

Issues, Controversies, Problems

The gap between game design and educational design has become evident that to optimize the learning from serious games, pedagogical experts should be actively involved in the development process (Arnab et al., 2015). The experts involved in a development process is highlighted many times. It is a challenged and substantial to ensure that collaboration between game-based learning experts, theoretical experts, game designers teachers, learners, developers, and non-technical domain experts occurs. The sophistication in terms of popularity and budget of creating EG by scratch demands high level of production process is also challenging to researchers and game developers. Due to lack of infrastructure capabilities including hardware and software requirements, theoretical aspects in terms of pedagogical and instructional aspects are also very crucial in designing and development process. The game industry needs to insure that integrating an instructor in an EG can provide a useful feedback to users at a run-time significance (Sinkewicz, 2015).

Modelling and developing effective EG needs to have a clear understanding of what the desired learning outcomes are and to establish a transparent mapping amongst (Hainey, Connolly, Stansfield, & Boyle, 2011). To understand that, research methods and statistics are broad-ranging, complex, and interconnected with logical and scientific reasoning, understanding different representations of data, data analysis, interpreta-

tion of results and evaluation skills (Boyle et al., 2014). It is barely the case where domains such as game play, game environment, subject-matter, and learning theories interaction has been discussed in order to provide a holistic view of a particular game element. Mostly, the literature uses vaguely spread game elements that cover large number of relationships and domains.

The GBL and its diverse dimensions and characteristics makes it difficult to evaluate and model. However, evaluation remains the only way to verify if the educational targets are being achieved and to detect any functional vulnerability inside the learning game. If the gaps between EG community and industry, and between academia and industry is bridged, the game development efforts could benefit all communities, including industry (Arnab et al., 2015; Azadegan et al., 2014). Although evaluation process can be costly, there should be a pre-modelling strategy which can allow developer or designer to model out the elements and related elements, as well as domains that should be involved in EG actively.

SOLUTIONS AND RECOMMENDATIONS

Initially, a multi-domain framework was proposed as a result of systematic literature review (SLR) (Ahmad, Ab Rahim, Arshad, & Mehat, 2016). In SLR, there were fifteen (15) existing frameworks and models, game design elements/components,

domains, learning theories and subject-matter aspects were understood (Ahmad, Rahim, & Arshad, 2015). Systematic literature review (SLR) is a method adopted for characterizing frameworks; hence, five (5) steps proposed by (Khan, Kunz, Kleijnen, & Antes, 2003) has assisted in proposing the multi-domain framework are briefly shown in Figure 1. Figure 2 shows the multi-domain framework, which is composed of four major domains: (1) Game environment (GE), (2) Game play (GP), (3) Learning Theories (LT), and (4) Subject-matter (SM). Each domain is composed of game components categorized according to the domain's definition.

With reference to the proposed framework, this paper discusses implicit relationships which are not supported by literature and they are not explicitly mentioned. Therefore, the selected twenty-seven (27) relationships are mapped upon the four domains to be validated by game-based learning experts. The Table 1 shows the implicit twenty-seven (27) relationships statements with highlighted game components from framework.

The extraction of relationships was obtained by using the NVIVO software to 'code' the text in literature that contain the EG elements and domains. Using the coding technique through the paragraphs analyzed; provided discern sentences that explain the implicit relationships between elements and domains.

In validation process, the authors conducted eight (8) qualitative interviews with game-based learning experts (GBLe) who are highly experi-

Figure 1. Five major steps to provide an effective systematic literature review



Figure 2. The Multi-domain framework for modelling educational games



enced with this area of study. The interviews were transcribed and coded into NVIVO software to understand the emergence of new “nodes” in each relationship. This article discusses all four (4) domains as the major themes and its relationships as sub-themes. Each relationship has its components and those components are the existing “nodes” (highlighted as **bold**).

Each relationship is discussed by applying interpretive hermeneutic approach of qualitative study. Interpretive hermeneutic method is associated to interpretivism philosophical perspective of qualitative study, allowing understanding of in-depth experience of experts and self-reflexive for a constructive interpretation (Kafle, 2011).

Results and Discussions

There were 216 feedbacks collected from eight GBLe during the interview session. The interpre-

tations are presented with interpretive paradigm expectation, combining components suggested by GBLe and author’s understanding towards the relevancy with domains of framework. Hence, this self-reflection aspect is the goal of hermeneutic method and interpretivism paradigm approach. Due to limitation of word count, the authors have discussed three relationships to provide the reader with an understanding of interpretations.

- **Relationship: 16.** GBLe highly agreed that interactive learning environment should have the components, namely, instructional messages, pedagogy and story, and less load towards learner’s memory. Furthermore, multimedia environment elements such as sound, graphics and animation should also be integrated. When this occurs, there are fifteen (15) learning theories suggested to be integrated within

Table 1. The explicit relationships on the four domains and their elements.

#	Relationship Statement
1.	Flow Experience includes sense of control over activity , time distortion , increase learning , increased exploratory behavior , positive effect , and acceptance of technology .
2.	Players can be allowed to change game play to improve their performance and reach goals .
3.	Combining game play and narrative context provides cognitive resources , immersion , engagement , and perception patterns of narrative .
4.	If the difficulty level of a task in a game increases step-wise , the player forms knowledge and strategies , hence, reduces frustration .
5.	Restorative narrative structure include story realistic environment , sounds , and back story flashbacks where character or player talks to game directly.
6.	A game should have pedagogical foundations and proven educational practices such as fun and engagement to meet expectations of producers and consumers.
7.	Well-designed games expose learners to complex tasks , predesigned narrative content , story and has the ability to interact with others.
8.	Educational games have new concepts and logical learning progressions to provide set of skills and knowledge .
9.	Social capital has structural dimensions which has the patterns of social interactions between actors . Social learning involves observation , imitation , and modeling .
10.	The content of the games reflects the problem . Content transfer outside the games tends to be limited and low level.
11.	Educational games have explicit knowledge and conversation for learners.
12.	Simulated environments are integral part of learning, supports intellectual development , and it has recursive cycle learning .
13.	Multi-player role playing games support problem-based learning and visualization of their actions to solve the challenges.
14.	Games should have emotive environment , role models and challenges .
15.	Educational games should be gender-inclusive .
16.	Interactive learning environment can be evolves instructional messages , pedagogy and story to minimize chance of overloading the learner's cognitive system .
17.	Cognitive apprenticeship involves authentic practices . Authentic practice is a combination of activity and social interaction .
18.	Learner's construct learning in EG with the aid of visual and verbal representation occurring parallel to their memory .
19.	Behaviorism allows the change in behavior of player and change in environment shape to encourage learners in understanding concepts. Attribution theory allows observing and explaining the cause of event changed.
20.	Social constructivist approach in multi-player games involves interaction between players and social construction of knowledge .
21.	Cognitive load theory guides the design of multimedia learning environment and it has instructional messages embedded to help avoid overloading memory of learning.
22.	Cognitive load optimization can be achieved with less irrelevant multimedia elements .
23.	Learners experiencing bad usability in task flow , eventually sacrifice attention and cognitive activity due to limitation of working memory capacity .
24.	Scaffolding can be combined with graphical presentations to lower player's cognitive load .
25.	Problem solving skill is also known as discovery learning , increases in and across games but it is difficult to use it outside the game.
26.	To ensure the player's behavior does not change in the gaming world, adaptive features should be transparent.
27.	Game play , game environment , subject-matter and learning theories are important in educational games .

this environment, i.e. motivational learning (behaviorism, collaborative learning, cooperative learning, guidance theory, intrinsically rewarded, social constructivism), cognitive, cognitive apprenticeship, constructivist learning, contextual thinking, experiential learning, problem solving, bloom's taxonomy, creative thinking, innovating thinking, and critical thinking. Since it plays a role in cognitive, the player's thinking and memory is also related to this concept. Moreover, it involves cognitive (knowledge), affective (attitudinal) and psychomotor (skills) domains in taxonomy, which works concurrently with respect to learner's playing games. Finally, this relationship is not amended and it involves components from GE, and GP domain.

- **Relationship: 25.** GBLe greatly supported with new suggestions to it. The problem-solving methods are implemented in EG as assignments or activities; which can be applied in real life as the learners adapt innovative approaches, logical situations, and appropriate knowledge selection to overcome the problem outside. Two emerging nodes were the way the knowledge is transferred inside and outside of EG. 'Transfer' is defined as the process of knowledge/skills acquired in one context in a new or varied context. Transfer is divided into a few types in which the knowledge context can be changed. Low road transfer (also known as reflexive transfer), involves the triggering of well-designed routines by stimulus conditions similar to learning context. High road transfer (mindful transfer) involves a deliberate amount of effortful abstraction and searches for connections. Moreover, another expert mentions about CPS factors (Creative Problem Solving); a mental state, problem reframing, multiple idea facilitation, and inducing change of perspective. The relationship is not amended and it involves the components from SM and LT domain.

Verifying the framework domain selection, one question involved all four domains. To ensure correct adoption of those domains, validation was scaled from 1 to 10 (1 as lowest and 10 as highest). The importance to collaborate game plays (GP), game environment (GE), subject-matter (SM) and learning theories (LT) in educational games.

Relationship statement: 27. *"I would give 10, because this is what makes an ordinary game different as compared to the educational game. Because in EG, we have certain learning outcomes that we need to achieve and in order to achieve the learning outcomes, we need to make sure that the learning environment has been taken care of in such a way. For example, the interactive natures of the game and also certain learning theories have been inserted in the game. And then, of course we need to have the expertise of subject-matter expert because we need to tailor the game, you know, for each and every subject. Okay, therefore I think this is very important all together and this must be considered in the educational games."* (Game-based Learning Expert's Opinion)

FUTURE RESEARCH DIRECTIONS

Modeling language for the behaviors of educational game is to our knowledge not available. Another field that may be of benefit to educational game development is model-based testing (MBT). MBT is a method to improve the testing process of software by referring to the models for guidance (Hemmati, Arcuri, & Briand, 2013; Timmer, Brinksma, & Stoelinga, 2011). Verifying effectiveness of the game has been explored many times (Omar & Jaafar, 2010) and the most common method is a combination of acceptance testing and quantitative/qualitative research methods. Other methods such as unit testing, white-box testing and verification using models are yet to be well known.

Further research can also be performed on using Model-Driven Engineering (MDE) (Tang & Hanneghan, 2011) in development of EGs; where,

1) modelling languages can be used at different level of modelling from requirements to detailed design, 2) model for traceability from requirements to code, and 3) verification of educational games effectiveness from models.

CONCLUSION

This chapter has discussed various types of existing frameworks and models which aid in developing EGs. Due to lack in explicit relationships amongst game components and vaguely spread elements, it is difficult to examine the importance and relevancy with other game domains. Therefore, this chapter proposes a multi-domain framework and twenty-seven explicit relationships involving four major domains, game play, game environment, subject-matter, and learning theories. To verify those relationships, interviews with game-based learning experts were collected. Due to their expertise in areas of teaching, game play, environments, and theoretical aspects of EG, it was an informative data collection. The interviews were interpreted using interpretivism paradigm and hermeneutics method is used to illustrate each relationship in context with the four domains. Finally, this chapters provides an insight on how the modelling of an existing EG or a new EG can be done by understanding different elements and their relations with other elements and domains of game to thoroughly understand the theoretical and practical concepts of EG before implementation.

REFERENCES

- Agrawal, A., Karsai, G., & Lédeczi, Á. (2003). *An end-to-end domain-driven software development framework*. Paper presented at the Companion of the 18th annual ACM SIGPLAN conference on Object-oriented programming, systems, languages, and applications. doi:10.1145/949344.949347
- Ahmad, M., Ab Rahim, L., Arshad, N. I., & Mehat, M. (2016). *A Multi-domain Framework for Modelling Educational Games: Towards the Development of Effective Educational Games*. Paper presented at the 7th International Conference on University Learning and Teaching (InCULT 2014). doi:10.1007/978-981-287-664-5_34
- Ahmad, M., Rahim, L. A., & Arshad, N. I. (2015). an analysis of educational games design frameworks from software engineering perspective. *Journal of Information & Communication Technology, 14*.
- Altunbay, D., Metin, M. G., & Çetinkaya, M. (2009). *Model-driven Approach for Board Game Development*. Paper presented at the First Turkish Symposium of Model-Driven Software Development (TMODELS), Ankara, Turkey.
- Amory, A. (2007). Game object model version II: A theoretical framework for educational game development. *Educational Technology Research and Development, 55*(1), 51–77. doi:10.1007/s11423-006-9001-x
- Annetta, L. A., Holmes, S. Y., Vallett, D., Fee, M., Cheng, R., & Lamb, R. (2013). *Cognitive aspects of creativity: Science learning through serious educational games. Teaching creatively and teaching creativity* (pp. 53–62). Springer. doi:10.1007/978-1-4614-5185-3_4
- Arnab, S., Lim, T., Carvalho, M. B., Bellotti, F., Freitas, S., Louchart, S., & De Gloria, A. et al. (2015). Mapping learning and game mechanics for serious games analysis. *British Journal of Educational Technology, 46*(2), 391–411. doi:10.1111/bjet.12113
- Azadegan, A., Hauge, J. B., Bellotti, F., Berta, R., Bidarra, R., Hartevelde, C., & Stanescu, I. A. (2014). *The move beyond edutainment: Have we learnt our lessons from entertainment games? Games and Learning Alliance* (pp. 77–89). Springer.

- Boyle, E. A., MacArthur, E. W., Connolly, T. M., Hainey, T., Manea, M., Kärki, A., & Van Rosmalen, P. (2014). A narrative literature review of games, animations and simulations to teach research methods and statistics. *Computers & Education*, 74, 1–14. doi:10.1016/j.compedu.2014.01.004
- Casey, J., Baghaei, N., & Nand, K. (2014). *Improving learning outcomes by designing engaging educational tools*. Paper presented at the Frontiers in Education Conference (FIE). doi:10.1109/FIE.2014.7044042
- Cooper, K. M., Nasr, E. S., & Longstreet, C. S. (2014). *Towards model-driven requirements engineering for serious educational games: Informal, semi-formal, and formal models*. In *Requirements Engineering: Foundation for Software Quality* (pp. 17–22). Springer.
- de Jong, T., & van Joolingen, W. R. (2008). Model-facilitated learning. In *Handbook of research on educational communications and technology* (pp. 457–468). Academic Press.
- Di Loreto, I., & Gouaïch, A. (2010). *Social casual games success is not so casual*. Academic Press.
- Djelil, F., Sanchez, E., Albouy-Kissi, B., Lavest, J.-M., & Albouy-Kissi, A. (2014). *Towards a Learning Game Evaluation Methodology in a Training Context: A Literature Review*. Paper presented at the European Conference on Games Based Learning.
- Dormans, J. (2012). *The effectiveness and efficiency of model driven game design*. In *Entertainment Computing-ICEC 2012* (pp. 542–548). Springer. doi:10.1007/978-3-642-33542-6_71
- France, R., & Rumpe, B. (2007). *Model-driven development of complex software: A research roadmap*. Paper presented at the 2007 Future of Software Engineering. doi:10.1109/FOSE.2007.14
- Hainey, T., Connolly, T. M., Stansfield, M., & Boyle, E. A. (2011). Evaluation of a game to teach requirements collection and analysis in software engineering at tertiary education level. *Computers & Education*, 56(1), 21–35. doi:10.1016/j.compedu.2010.09.008
- Jing, T. W., Yue, W. S., & Murugesan, R. (2015). *Learning Outcome Enhancement via Serious Game: Implementing Game-Based Learning Framework in Blended Learning Environment*. Paper presented at the IT Convergence and Security (ICITCS), 2015 5th International Conference on.
- Kafle, N. P. (2011). *Hermeneutic phenomenological research method simplified*. Academic Press.
- Kebritchi, M., & Hirumi, A. (2008). Examining the pedagogical foundations of modern educational computer games. *Computers & Education*, 51(4), 1729–1743. doi:10.1016/j.compedu.2008.05.004
- Kelly, S., & Tolvanen, J.-P. (2008). *Domain-specific modeling: enabling full code generation*. John Wiley & Sons. doi:10.1002/9780470249260
- Khan, K. S., Kunz, R., Kleijnen, J., & Antes, G. (2003). Five steps to conducting a systematic review. *Journal of the Royal Society of Medicine*, 96(3), 118–121. doi:10.1258/jrsm.96.3.118 PMID:12612111
- Kühn, M. (2015). *Applying Patterns when Generating Code: a Model-based Design Approach*. Paper presented at the Multimedia, Interaction, Design and Innovation. doi:10.1145/2814464.2814474
- Marfisi-Schottman, I., George, S., & Tarpin-Bernard, F. (2010). *Tools and methods for efficiently designing serious games*. Paper presented at the 4th European Conference on Games Based Learning ECGBL.
- Marne, B., Wisdom, J., Huynh-Kim-Bang, B., & Labat, J.-M. (2012). *The six facets of serious game design: a methodology enhanced by our design pattern library*. In *21st Century Learning for 21st Century Skills* (pp. 208–221). Springer.

Miller, J., & Mukerji, J. (2001). *Model driven architecture (mda)*. Object Management Group, Draft Specification ormsc/2001-07-01.

Osman, K., & Bakar, N. A. (2012). Educational Computer Games for Malaysian Classrooms: Issues and Challenges. *Asian Social Science*, 8(11), 75. doi:10.5539/ass.v8n11p75

Reyno, E. M., & Cubel, J. Á. C. (2008). *Model Driven Game Development: 2D Platform Game Prototyping*. Paper presented at the GAMEON.

Romero, M. (2015). Work, Games and Lifelong Learning in the 21st Century. *Procedia: Social and Behavioral Sciences*, 174, 115–121. doi:10.1016/j.sbspro.2015.01.634

Sauer, S., & Engels, G. (2001). *UML-based behavior specification of interactive multimedia applications*. Paper presented at the Human-Centric Computing Languages and Environments. doi:10.1109/HCC.2001.995271

Sinkewicz, N. D. (2015). *A framework for designing serious games*. Academic Press.

Stevens, P. (2002). On the interpretation of binary associations in the Unified Modelling Language. *Software & Systems Modeling*, 1(1), 68–79. doi:10.1007/s10270-002-0002-x

Tang, S., & Hanneghan, M. (2010). *A model-driven framework to support development of serious games for game-based learning*. Paper presented at the Developments in E-systems Engineering (DESE). doi:10.1109/DeSE.2010.23

Truong, K. N., Hayes, G. R., & Abowd, G. D. (2006). *Storyboarding: an empirical determination of best practices and effective guidelines*. Paper presented at the 6th conference on Designing Interactive systems. doi:10.1145/1142405.1142410

van Staalduinen, J.-P., & de Freitas, S. (2010). *A game-based learning framework: Linking game design and learning outcomes*. Learning to Play: Exploring the Future of Education with Video Games. Peter Lang Publishers.

Von Wangenheim, C. G., Thiry, M., & Kochanski, D. (2009). Empirical evaluation of an educational game on software measurement. *Empirical Software Engineering*, 14(4), 418–452. doi:10.1007/s10664-008-9092-6

Wada, H., Suzuki, J., & Oba, K. (2005). *Modeling turnpike: a model-driven framework for domain-specific software development*. Paper presented at the Companion to the 20th annual ACM SIGPLAN conference on Object-oriented programming, systems, languages, and applications. doi:10.1145/1094855.1094897

KEY TERMS AND DEFINITIONS

Bloom's Taxonomy: it is known as a multi-tiered model of classifying thinking according to six cognitive levels of complexity.

Game Environment: A dimension which collaborates game rules, objectives, subject, and theoretical aspects together as a whole to provide an interactive flow of activity.

Game Play: This is where the rules and regulations of EG.

Hermeneutic Circle: The iterative process of interpretation, reflective writing, and reading to provide the details of the whole phenomenon discussed.

Hermeneutics Method: An interpretive and concentrated research on historical meanings of experience and their developmental and cumulative effects on individual and social levels. Iteratively taking parts of interdependent meaning and the whole of their formation to achieve understanding.

Interpretivism: The belief in multiple observations of reality, subjective and socially constructed through language, consciousness and shared meaning.

Model-Driven Engineering: A software development approach, focusing on creation of models that represents the system-under-study (SUS) and subsequent generation of fully-working software artefacts from these models.

Subject-Matter: The teacher/learner and the subject embedded in the EG.

Category G

Gender and Diversity

Computing Technologies and Science Fiction Cinema



Rocío Carrasco-Carrasco
University of Huelva, Spain

INTRODUCTION

The intersections of the human body with the latest computing technologies have opened up numerous debates on what it means to be human in our technologically mediated societies. Classical dichotomies body/mind, human/machine, natural/artificial seem to be dissolving in contemporary Western societies, and the limits of the “organic” body become difficult to establish. The result is a new ontology of the body. Practices such as cloning, reproductive technologies, robotics and implants, among others, have become familiar facets of our global societies that place the human body in an assemblage with technology. Moreover, the world is dominated by technological images that have become part of our daily life and that have created new spaces for representation (and eventually for living), such as the virtual world or cyberspace.

The present paper privileges the body as being key to understanding our intricate relationship with the latest advances in computing and robotics in contemporary societies. In doing so, it critically entrenches with a specific movement called “new materialism”, whereby matter is not perceived as fixed or passive but rather as a dynamic and shifting process. Specifically, I will stress the importance of approaching popular representations of the so-called “body in transit”, as this shifting idea of corporeality reflects contemporary anxieties and interests fueled by the relationship between physical bodies, computing technologies and gender representation. For this purpose, this work will focus on the notion of the fluid body

or “body in transit” as represented in US popular Sci-Fi cinema to contend that this posthuman figuration is still informed by gendered practices and dominant structures of power, despite its hybrid nature.

Computing and media technologies are everywhere and extend to the human body, affecting the way gender has been traditionally understood. As some feminist research has highlighted, technology is affected by gender relations. Technology in general has been traditionally considered as a sign of men’s power and masculinity defined in terms of technological capabilities. Nevertheless, current discourses have provided new definitions of technology, of gender identity and of what being human means. This inevitably challenges traditional power associations between men and technology. As Barbara Becker argues, the “difference between natural and artificial, real and virtual, material and immaterial phenomena is not an ontological one, but changes according to technological improvements and methods of communication” (Becker, 2000, p. 361). In the same way, definitions of gender also change with time, affected by technological developments.

Cybernetics, as a set of media technologies, offers grounds from where to analyze gender in contemporary contexts. Cyberspace has offered numerous possibilities for the redefinition of the human body outside traditional boundaries, suggesting a liberation of socio-cultural constraints. This is precisely the concern of many feminist theories that aim at deconstructing the human subject from binary polarization, implying the dissolution of sexualized identities in cyberspace.

Specifically, the discipline called “cyberfeminism” sees cyberspace as a gender-neutral site that enables women to communicate and act outside the constraints of male-dominated physical realms. Sadie Plant and many other cyberfeminists offer optimistic—sometimes utopian—views of the relationship between women and technology in the virtual age. In her essay “On the Matrix: Cyberfeminist Simulations,” Plant argues that virtual worlds “undermine both the world-view and the material reality of two thousands years of patriarchal control” (Plant, 2000, p. 265).

Yet, contrary to cyberfeminist postulates about the neutrality of cyberspace for gender relations, popular discourses like Sci-Fi cinema normally rely on this distinction when depicting the interaction between the body and computing technologies, inevitably adopting gender dualisms. Cyberspace is constructed by existing social, cultural and economic structures, and gender stereotypes and sexed body descriptions are normally employed in order to suggest authenticity to these visual texts. This is not to say, however, that popular discourses do not offer fresh instances of flesh and computing technologies. As argued here, contemporary US cinema has overcome old visualizations of static and gendered cyborg figures as represented by figures like *The Terminator* or *Robocop*, and offers instead images of fluid subjectivities as embodied by characters that live in transit between reality and virtuality, as shown in movies like *The Matrix* (1999), *eXistenZ* (1999), *Avatar* (2009), *Surrogates* (2009), *Inception* (2010), *Tron: Legacy* (2010) or *Ex Machina* (2015), among others. This idea of fluidity is achieved by a series of visual and narrative strategies that partly help overcome the obstacles posed by cultural products that, due to their conditions of production, cannot break away from the hegemonic assumptions of gender identities. It is precisely the paradoxical nature of cinematic bodies in transit what I intend to emphasize here, especially since they articulate contemporary cultural and gender concerns.

BACKGROUND

In cultures where the organic “natural” body is gradually disappearing and giving way to fruitful instances of flesh and machine, the current outburst of debates on corporeality becomes, if less, paradoxical. The body remains at the core of many contemporary analyses, proving to be a useful tool for examining culture and gender. As Arthur Kroker contends, “[w]hile the triumph of mass media, particularly television, may portend a future of pure simulation, the overriding cultural reality is that the image machine is itself haunted by memories of the body” (Kroker, 2012, p. 1). The body becomes, then, an archive from where to analyze contemporary society. Yet, the new body that emerges—also known as the posthuman body—is perceived as a contested notion linked to fluidity, hybridity and complexity. Kroker talks about “body drift” to refer to the fact that we no longer inhabit a body in any meaningful sense of the term but rather “occupy a multiplicity of bodies—imaginary, sexualized, disciplined, gendered, laboring, technologically augmented bodies” (p. 2). The idea of the fluctuating, fluid or hybrid body has been theorized by scholars like Donna Haraway, Katherine Hayles, Judith Butler and Rosi Braidotti, among others. These writings open up a new tradition of critical feminism that addresses the complexity of our bodies in technologized cultures.

One of the most striking examples of a hybrid body is that proposed by Donna Haraway in her material-semiotic approach “Cyborg Manifesto” (1985). As she sets out to defend, the cyborg is about transgressed boundaries, potent fusions and dangerous possibilities. The popularized cyborg figure has greatly stimulated many insights into the gender/power relations with technology, leading to consider the possibilities that technoscience offers women. Indeed, as she defines it, the cyborg is a fictional hybrid of machine and organism, “a creature in a post-gender world

[which] has no truck with bisexuality, pre-oedipal symbiosis, unalienated labor, or other seductions to organic wholeness through a final appropriation of all the powers of the parts into a higher unity” (Haraway, 1991, p. 150). The technological world frees women’s representations, in a sense, from patriarchal domination. Along with Haraway, postmodern feminists like Judith Butler have sought to deconstruct binary systems by reconceptualizing gender as unstable and performative. She considers corporeality as an act of imitation. Butler affirms that “the regulatory norms of ‘sex’ work in a performative fashion to constitute the materiality of bodies and, more specifically, to materialize the body’s sex, to materialize sexual difference in the service of the consolidation of the heterosexual imperative” (Butler 1993, p. 2). Sex, then, becomes one of the norms that “qualifies a body for life within the domain of cultural intelligibility” (1993, p. 2). In this sense, and very much like the cyborg figure proposed by Haraway, the performative body “has no ontological status apart from the various acts of identities which constitute its reality” (Butler 1993, p. 228), suggesting the fluidity of our identities.

There is a strand of feminist practices, called “material feminism” or “new materialism” whose writings become of special interest for the argument of this paper. For Hayles, Braidotti and other materialist theorists, the body becomes fundamental for analyzing contemporary human-machine interactions. Subjectivity and the body are decentered as distinct units, troubling the information/material separation. While technology and biology interact with other, the body stands as a useful tool from where to examine complex relationships with media and computing technologies. Accordingly, cyborgs are not only about people with machine legs or arms. As Braidotti affirms, cyborgs are no longer the space of Sci-Fi but they define our social framework at all levels: from medicine, to telecommunications, finance and modern warfare (2002, p. 17). She uses a multi-layered definition of the cyborg, which includes today fragile bodies that interface with computer technologies at posthuman levels.

Moreover, the new technologies offer prosthetic extensions to our bodies in the form of mobile phones, Internet networks and the like, which open new ways of challenging the classical humanist mode of representation (Braidotti, 2002, p. 18).

In a similar vein, Hayles argues that cyborgs indicate a culture in which information has got under our skin. Bodies are forms of information interacting with other forms of information. In *My Mother was a Computer*, Hayles talks about the ineludible interaction between subjectivity and materiality:

[intelligent machines] are neither objects to dominate nor subjects threatening to dominate me. Rather they are embodied entities instantiating processes that interact with the processes that I instantiate as an embodied human subject. The experience of interacting with them changes me incrementally, so the person who emerges from the encounter is not exactly the same person who began it. (Hayles, 2005, p. 243)

Accordingly, the difference between human and technology has become difficult to distinguish (2005, p. 228) and the (virtual) body that emerges speaks the language of 21st century cybernetics.

In offering a new definition of the body, these theories refuse the post-apocalyptic nihilism postulated by theorists like Jean Baudrillard (1994), who advocated for the dissolution of the material body in hypermedia societies where the real and the imaginary collapse into each other. Corporeal feminism is also against the movement known as “transhumanism” which envisions the possibility of enhancing human potential by means of science and technology. Instead, Hayles, Braidotti and the “new materialisms” provide us with detailed analyses of the impact of computation on embodiment and subjectivity, aiming to move toward a materialist reconfiguration of the posthuman, or an embodied posthumanism. Within these writings, the posthuman body becomes an icon of the 21st century cybernetic culture. In other words, these works present a return to the body, yet a hybrid or fluid body that troubles the information/material

separation, a boundless body, a body in transit. Specifically, the posthumanist mode is articulated by recent critics like Cary Wolfe (2010), Francesca Ferrando (2013), Patricia MacCormack (2012), or by Rosi Braidotti (2013). In her latest book, *The Posthuman* (2013), Braidotti deals with the posthuman predicament as a way of understanding our plural identities in contemporary mediated societies. The aim of the posthumanist thinker is, according to Braidotti, to go beyond traditional humanist limitations and embrace the risks that “becoming-other-than-human” entails. Throughout her work, Braidotti recognizes that the figure of the posthuman is ambiguous and caught up in projections of desire or fantasies of domination and disembodiment. Yet, in a similar way as Haraway’s cyborg, the posthuman as articulated by Braidotti can be used as a liberating force that overcomes the negativity of contemporary practices that reduce sexualized, racialized and naturalized “others” to disposable bodies (Braidotti, 2013, p. 15).

In short, the body in transit between reality and cyber-reality becomes fundamental for analyzing contemporary human-machine interactions, as I will attempt to illustrate in the next section.

BODIES IN TRANSIT: THE VIRTUAL REALITY FILM

Due to the ever-increasing presence of science and technology in our daily lives, the themes, topics and ideologies traditionally associated to Sci-Fi have been imported to popular culture. Accordingly, the creation of spaces such as virtual reality, the Internet or cyberspace, has allowed for challenging representations of the human body, sometimes conceived as pure information. While a common agreement among feminists is that Sci-Fi offers in many ways a traditional vision of gender relationships, and women are still relegated to a secondary position in such texts, contemporary movies showing cyberspace particularly offer opportunities for understanding the complex relationship between humanity and information technologies. Wrapped in hypermedia

societies, subjectivities seem to be disturbed and unable to distinguish reality from simulation, affecting the way the body is culturally understood in these films.

Many contemporary Sci-Fi films follow cyberpunk tradition of the 1980s and present cyberworlds in which the traditional body is contested. The movement known as cyberpunk stresses the superiority of the mind over the body, suggesting the elimination of the material body. Indeed, the term cyberspace is deeply entrenched with this literary movement whose main representative is William Gibson. It first appeared in *Neuromancer* (1984) to refer to “a virtual hallucination experienced daily by billions of legitimate operators” (Gibson, 1984, p. 69). For most cyberpunk fiction, the body is not a biological essence but it is regarded as pure mind. The Gibsonian concept of cyberspace has begun to be felt as tangible and representative of our social reality, which seems to be located somewhere between fiction and reality. In this sense, it is a problematic and ambiguous domain. On the one hand, cyberspace has opened numerous possibilities for the definition of the human body outside traditional boundaries, suggesting a liberation of socio-cultural constraints. On the other hand, and since such space is constructed by existing structures, stereotyped images and descriptions of bodies are normally used. Even in cybernetic contexts, where the blurring of frontiers is taken to an extreme, gender inequalities are to be found. Dani Cavallaro addresses these issues when he affirms that cyberpunk’s approach to gender roles is highly ambiguous, especially if we consider some “conflicting interpretations of the gender and sexual attributes of some of cyberpunk’s best-known characters” (Cavallaro, 2000, p. 121). Cavallaro is referring to characters like Molly in *Neuromancer*, whose gender connotations are ambivalent and it is up to the reader to decide if she can be considered as an image of the liberated woman or as a stereotype.

With the help of the latest special effects, Sci-Fi is symbolic of the attitudes of society to technological advances. Specifically, by focusing on the trope of the body in transit, contemporary Sci-Fi films

provide the ideal forum from where to articulate cultural anxieties regarding the fluid and complex relationship between humanity and technology. In this sense, this cinematic genre can be considered as an important cultural referent that foregrounds questions of gender and identity at work in contemporary society. There has been, especially in the last decade in the USA, a proliferation of films dealing with cyberspace in which men and women undergo virtual experiences. These films often rely on the disorientation of the senses caused by the interaction between body and technology at the end of the millennium. This disorientation further suggests that we have lost control of our material bodies and of the world around us, urging us to find new instances of representational regimes for the intrusion of technology in our bodies. Many films of the end of the 1990s and the beginning of the 21st century like *Virtuosity*, *The 13th Floor*, *The Matrix*, *Avatar*, *Tron: Legacy* or *Disconnect* offer significant representations of the digital media in the form of the Internet, virtual reality, computer games, digital databases, or social networks. This ability to simulate our experience of reality ends up in the confusion between reality and fiction, a main motif in films dealing with cyberspace, in which characters—and spectators by extension—are frequently unable to perceive what is “real.” Films dealing with hypermedia operate between two spaces (the real and the simulated) that sometimes merge, which affects the way gender is constructed in them.

The idea that the real self becomes a simulation in cyberspace opens great possibilities for the depiction of challenging body figurations, suggested by films that deal with ontological issues, like *Vanilla Sky* (2001), *Fight Club* (1999), *Dark City* (1998) or, more famously, *The Matrix* trilogy (1999; 2003; 2003) and *Inception* (2010). In these films, the fluid interaction of the body with cyberspace complicates issues of gender depiction, partly because spectators—like characters—are sometimes unable to distinguish the reality from the virtuality of the worlds that are displayed. The ambiguity posed by the hybridity of these bodies due to their contact with digital media is

explored in the cyberthriller *Strange Days* (1995), whose protagonist, ex-cop Lenny, accidentally enters a virtual experience that disturbs him and introduces him in a dark web of murder and blackmail. In *Hackers* (1995) their protagonists have a quasi-erotic but troubling relationship with their laptops and personal computers, which become extensions of their bodies. More problematic is, however, the brain expansion implanted into a mentally retarded man in *The Lawnmower Man* (1992) or the massive load of information that the protagonist of *Johnny Mnemonic* (1995) stores in his brain, entailing rape, murder and killing. The earlier *Videodrome* (1983) suggests the manipulative effect of video images literally inserted in the male body, which causes hallucinations to the protagonist. This trope persists in more recent films like *Disconnect* (2012), which shows how people experience the negative sides of today’s wired and globalized world. In this sense, these films coalesce with Christine Ross’s comments on the mutability of the body at the end of the millennium, partly due to the fast-expanding integration of technologies of information into everyday life (Ross, 2001, p. 28).

As suggested before, Hayles’s or Braidotti’s ideas of embodied virtuality may well be utilized here as a way of understanding the complex ways in which the body is interpenetrated with information patterns, suggesting reconfigurations of traditional body markers. Moreover, this confusion talks about a new—yet disturbing—relationship with computing technologies, while hinting at the problematics of new gender representations. Indeed, male cyborgs appearing in many films from the end of the 20th and beginning of the 21st century become key figures for the exploration of new values ascribed to manhood. This idea is illustrated, for instance, in the film *eXistenZ* (1999) whose protagonist, Ted, departs from conventional representations of male leading figures on screen in the sense that his constant worries over the ways the virtual game affect the materiality of his body (he needs to have his body drilled in order to enter cyberspace) cause destabilization and puzzlement. This situation suggests the contemporary crisis

of masculinity at work in US society at the turn of the century, precisely when new definitions of masculinity appeared in the public arena. Likewise, *The Matrix* trilogy proposes a new model of masculinity: Neo's androgynous and malleable body and his troubled subjectivity do not match conventional portrayals of the action hero. The virtual hero is called Neo, and this encourages spectators to expect a rather innovative personality and/or behavior, or at least something new concerning his physical aspect. The innovation is easily perceived in his looks, which confuses viewers at first with his gender identity. He can be also said to be the embodiment of the "New Man." The idea of playing with conventions is also explored in *Tron: Legacy* whose female hero, Quorra, adopts traits traditionally associated to masculinity, only to show that she is a strong, resourceful and accomplished female warrior.

By subverting earlier pessimistic feminist approaches that stressed the "masculinization" of media technologies, contemporary Sci-Fi offers a fresh relationship between women and digital technologies. Significantly, an outburst of films showing virtual realities and cyberspace appeared in the late 1990s, a time when there was a remarkable shift in power from men to women in technological development, as cyberfeminists noted. In these films, women are seen as creators of computing technologies, feeling comfortable when interacting with the new technologies, as it is shown in *eXistenZ*, *The 13th Floor* or *Inception*, to give some remarkable examples. In this sense, the Sci-Fi genre holds the power to subvert the sad reality of women's marginalization in present day technologized societies. As Ivana Milojevic argues when analysing visions of the future and virtual reality technologies: "women's images and selves are being created and valorized in the minds of adolescent net surfers", the net being "a place for the gathering of sexual harassers and paedophiles". Milojevic foresees a future world "where women will no longer be needed at all, creating the womanless real world and women-filled virtual world" (Milojevic, 2010, p.39). However, in these films, cyberspace is presented

as an innovative site where women can enjoy of an intimate and fruitful relationship with computing technologies.

In spite of women's visible engagement with various media technologies, mainstream virtual reality films do find real obstacles when representing alternative subjectivities and new body politics, what echoes at existing socio-material power inequities. The main obstacle may well derive from the fact that mainstream cinema presents digital worlds that are meant to simulate a still gender-biased reality. In order to confer meaning to the visual text, screened bodies need a referent and, while the reality in the film may be constructed as very different from our current reality, it is still a simulation of it. Precisely, in *Metamorphosis*, Braidotti uses film to illustrate the intensification of the gender gap and the polarization between the sexes in virtual contexts.

This inability to abandon old ideas of the body and to evolve in terms of gender depiction is shown in *Avatar*, whose protagonist, Jake, is a disabled ex-soldier who jacks into a vigorous alien-hunter body, and whose supremacy remains unquestionable throughout the film. The female hybrid character, Neytiri, is a brave and spiritually strong being, although nothing is known about her outside her sentimental relationship with the hero. Avatars are still constructed around gender stereotypes in movies. Women are simply ignored in *Tron* (1982), considered the first movie to visualize the virtual reality, or in *Virtuosity* (1999), where the only digital image of women (called Sheila 3.2) serves the mere purpose of sexual lure. The character of Lara Croft in the film adaptation of the popular video game *Tomb Raider* (2001) is but another more example of the fetishization of the female virtual body.

Mediated or virtual bodies appearing in many contemporary movies have paradoxically been designed according to Western standards of beauty, which problematizes their claims for a liberating body image. The film *Surrogates* (2009) shows how most humans live in isolation and use fit, healthy and attractive humanoid robots through which they interact with society,

while they preserve the integrity of their bodies in the comfort of their own homes. While real humans are depicted as possessors of fat and unfit bodies, surrogates are beautiful, healthy beings nicely dressed and visually appealing, reaffirming patriarchal assumptions about sex and gender. When dealing with contemporary societies, Susan Bordo states that “our bodies are trained, shaped and impressed with the prevailing forms of (...) masculinity and femininity” (Bordo, 1993, p. 91). By regimes of dieting, makeup, exercise, dress, cosmetic surgery, women, and increasingly men, try to sculpt their bodies into shapes which reflect the dominant societal norms. In *Surrogates*, bodies are disciplined to correspond to a social ideal, they remain invulnerable and do not suffer the pains of cosmetic surgery, dieting, or the destructive effects of natural disasters and other threats of extinction. Surrogates function as body prostheses that perform pre-established patterns or norms of beauty and success, while preserving an invisible, decadent and ageing human race. Another film showing this problematic is *The Hunger Games* (2012) where the main female protagonist, Katniss, plays with the idea of beauty as a social construct to suggest the artificiality of gender.

Some feminist researchers have pointed out the paradoxical nature of the material body as inscribed in cyberworlds. For instance, Ann Balsamo (1996) argues that the body in virtual reality does not erase gender or race markers (1996, pp. 125-127). The risk of falling into body-essentialism has been also noted by theorists on new technologies (Becker, 2000; Boler, 2007), who affirm that users normally rely on stereotyped images of the body while searching for new concepts of identity in virtual worlds. Van Loon has underlined the importance of embodiment in certain media contexts and contends that it should be understood as an “effect” of our mediated being, and as such, susceptible to transformation (2008, p. 100). When referring to virtual reality, Van Loon opts, then, for analyzing embodiment in close conjunction with disembodiment as both variations of the virtual (2008, p. 127). In addition, new media theories provide evidence of the impossibility of totally

transcending the body in cyberspace. Murphie and Potts argue that virtual reality systems rely very often on the perceptual mechanisms of the body in order to work (Murphie & Potts, 2003, p. 119): “we feel we are in virtual space only because our senses tell us we are—through our eyes, our hands and our ears. Even our haptic senses pick up the internal arrangements of our bodies, subject to things like movement or gravity to tell us we are in VR” (2003, p. 120).

The idea of simulated worlds where the real merges with the imaginary does not apply to gender representation in mainstream virtual reality films. Gender is, ironically, paralyzed in the hyper-technological panorama where everything seems to mingle with reality. This may be a consequence of the need to have a referent in a world where our identities are fragmented as a result of our close interface with the newest media technology. The body in transit is still the repository of marginalized gendered and sexed practices.

FUTURE RESEARCH DIRECTIONS

Body imagery in mainstream Sci-Fi films is still subjected to hegemonic definitions of masculinity and femininity, and show up the inability to offer gender free portrayals. As a popular discourse, Sci-Fi mirrors effectively cultural anxieties concerning the posthuman subjectivity as articulated by “new materialists”. A research on how the posthuman body is described and/or depicted in other contemporary discourses, such as video games, written Sci-Fi or *auteur* cinema could supplement this work and explore the intricacies of pioneering works by Braidotti and Hayles.

CONCLUSION

Accepting that our bodies have been significantly altered due to technology may not always be comforting, but can mean a solution to end up with the oppressive old rules of the gendered body. As I have attempted to illustrate, the Sci-Fi proves to

be aware of the paradoxes of body representation on screen. Popular Sci-Fi showing virtual environments do not totally undermine dominant gender values and, consequently, the body is never totally “invisible” in them. Even the most intrusive users of technology—as happens in films like *The Matrix* or *Surrogates*—seem to be stimulated by the patriarchal demands on the physical body. As a result, the post-gender alternative proposed by some researchers does not find a real place in this popular genre. In this sense, cinematic bodies in transit cannot be regarded as embodiments of a positive posthuman subjectivity, or as advocates for the politics of affirmation as described by Braidotti and other feminist materialists. The idea that different is positive is not fully articulated in these bodies, and they still follow established and oppressive conventions. Braidotti makes a call for posthuman thinking: “these non-profit experiments with contemporary subjectivity actualize the virtual possibilities of an expanded, relational self that functions in a nature-culture continuum and is technologically mediated” (2013, p. 61). Unfortunately, this is not the spirit of contemporary capitalism, a system that promotes uniformity and conformism to the dominant ideology.

Along with this conclusion emerges another one, less pessimistic. While the Sci-Fi film suggests the importance of keeping the materiality of our bodies, it also demonstrates that the interaction between media technologies and our subjectivities is possible and that new ways of understanding the reality of our bodies need to be found. While Hayles, Haraway, Plant and other feminist theorists have proposed ways in which the virtual subject can be redefined, it is still necessary to give attention to the ways popular culture faces the reality that our bodies interfere with technology. After all, most of the films commented here show hybrid and fluid bodies that live in transit reality and virtuality, that are vulnerable, and that project our inner conflicts with technology. Popular Sci-Fi needs to find a space between hegemonic discourses and the “space-off” that Sci-Fi allows, in relation to the representation of the posthuman body.

ACKNOWLEDGMENT

The author wishes to acknowledge the funding provided by the Spanish Ministry of Economy and Competitiveness (Research Project “Bodies in Transit”, ref. FFI2013-47789-C2-1-P) and the European Regional Development Fund for the writing of this essay.

REFERENCES

- Balsamo, A. (1996). *Technologies of the gendered body: Reading cyborg women*. Durham, NC: Duke University Press.
- Becker, B. (2000). Cyborgs, agents, and transhumanists: Crossing traditional borders of body and identity in the context of new technology. *Leonardo*, 33(5), 361–365. doi:10.1162/002409400552838
- Bordo, S. (1993). *Unbearable weight. Feminism, western culture and the body*. Berkeley, CA: University of California Press.
- Braidotti, R. (2002). *Metamorphoses: Towards a materialist theory of becoming*. Cambridge, UK: Polity Press.
- Braidotti, R. (2013). *The posthuman*. Cambridge, UK: Polity Press.
- Butler, J. (1993). *Bodies that matter: On the discursive limits of “sex”*. New York: Routledge.
- Cavallaro, D. (2000). *Cyberpunk and cyberculture: Science fiction and the work of william gibson*. London: Continuum International Publishing.
- Gibson, W. (1984). *Neuromancer*. New York: Ace Books.
- Haraway, D. J. (1991). *Simians, cyborgs and women: The reinvention of nature*. New York: Routledge.
- Hayles, N. K. (2005). *My mother was a computer*. Chicago: University of Chicago Press; doi:10.7208/chicago/9780226321493.001.0001

Kroker, A. (2012). *Body drift: Butler, Hayles, Haraway*. Minneapolis, MN: University of Minnesota Press. doi:10.5749/minnesota/9780816679157.001.0001

Milojevic, I., & Inayatullah, S. (2010). Feminist critiques and visions of the future. *Futures Research Quarterly*, 14(1), 35–46.

Murphie, A., & Potts, J. (2003). *Culture and technology*. New York: Palgrave. doi:10.1007/978-1-137-08938-0

Plant, S. (2000). On the matrix: Cyberfeminist simulations. In G. Kirkup, L. Janes, F. Hovenden, & K. Woodward (Eds.), *The Gendered Cyborg: A Reader* (pp. 265–275). New York: Routledge.

Ross, C. (2001). The insufficiency of the performative: Video art at the turn of the millennium. *Art Journal*, 60(1), 28–33. doi:10.1080/00043249.2001.10792047

Van Loon, J. (2008). *Media Technology. Critical Perspectives*. New York: Open University Press.

ADDITIONAL READING

Baudrillard, J. (2009). Introduction: The murder of the sign, jean baudrillard; interviewed by sylvere lotringer, columbia university. In S. Cohen & R. L. Rutsky (Eds.), *Consumption in an age of information* (pp. 9–12). Berg Publishers.

Boler, M. (2007). Hypes, hopes and actualities: New digital cartesianism and bodies in cyberspace. *New Media & Society*, 9(1), 139–168. doi:10.1177/1461444807067586

Constable, C. (2006). Baudrillard reloaded: Interrelating philosophy and film via The Matrix trilogy. *Screen*, 47(2), 233–249. doi:10.1093/screen/hjl018

de Souza e Silva, A. (2003). From simulations to hybrid space: How nomadic technologies change the real. *Technoetic Arts: an international journal of speculative research*, 1(3), 209–221.

Ferrando, F. (2013). Posthumanism, transhumanism, antihumanism, metahumanism, and new materialisms differences and relations. *Existenz*, 8(2), 26–32.

Green, N. (2001). Strange yet stylish headgear: virtual reality consumption and the construction of gender. In E. Green & A. Adam (Eds.), *Virtual gender: Technology, consumption and identity* (pp. 150–172). London: Routledge.

Halberstam, J., & Livingston, I. (Eds.). (1995). *Posthuman Bodies*. Bloomington: Indiana UP.

Hayles, K. N., & Gessler, N. (2004). The slipstream of mixed reality: Unstable ontologies and semiotic markers in The Thirteenth Floor, Dark City, and Mulholland Drive. *PMLA*, 119(3), 482–499. doi:10.1632/003081204X20541

Hayles, N. K. (1999). *How we became posthuman: Virtual bodies in cybernetics, literature and informatics*. Chicago: University of Chicago Press; doi:10.7208/chicago/9780226321394.001.0001

Herbrechter, S. (2006). The posthuman subject in The Matrix. In M. Diocaretz & S. Herbrechter (Eds.), *The Matrix in theory* (pp. 249–290). Amsterdam: Rodopi Editions.

Herbrechter, S. (2013). *Posthumanism: A critical analysis*. London: Bloomsbury Publishing.

Lauretis, T. (2003). Becoming Inorganic. *Critical Inquiry*, 29(4), 547–570. doi:10.1086/377720

MacCormack, P. (2012). *Post-Human Ethics*. Ashgate.

Scott, A., Semmens, L., & Willoughby, L. (2001). Women and the Internet: The natural history of a research project. In E. Green & A. Adam (Eds.), *Virtual gender: Technology, consumption and identity* (pp. 1–27). London: Routledge.

Springer, C. (1998). The Pleasure of the Interface. In P. Hopkins (Ed.), *Sex/machine: Readings in culture, gender, and technology* (pp. 265–275). Bloomington: Indiana University Press.

Springer, C. (1999). Psycho-Cybernetics in films of the 1990s. In A. Kuhn (Ed.), *Alienzone: Cultural theory and contemporary science fiction cinema* (pp. 203–218). London: Verso.

Wajcman, J. (2004). *Technofeminism*. Cambridge: Polity Press.

White, M. (2001). Visual pleasure in textual places: gazing in multi-user object-oriented worlds. In E. Green & A. Adam (Eds.), *Virtual gender: Technology, consumption and identity* (pp. 124–149). London: Routledge.

Wolfe, C. (2010). *What is Posthumanism?* Minneapolis: University of Minnesota Press.

KEY TERMS AND DEFINITIONS

Cinematic Cyborg: An image provided by films (mainly science fiction) in which flesh and machine combine, representing the relationship between material bodies and media technologies.

Cyberfeminism: Discipline within feminism that sees cyberspace and virtual reality as neutral realms in terms of gender. This school of thought

visions a society beyond gendered bodies where women can communicate and act outside the restrictions imposed by patriarchal societies.

Gender Identity: A person's sense of oneself as male, female, both or neither, resulting from culture and other external and/or internal factors.

Material Feminism: A discipline that considers the physical body as fundamental for the articulation of gender and for the analysis of contemporary human-media interactions. Key authors are Katherine Hayles or Rossi Braidotti.

Posthuman Body: A contested notion that stands as an icon of the fluidity, hybridity and complexity of corporealities in contemporary technologically-driven societies.

Virtual Body: An inorganic entity with human shape made of information and placed in cyberspace.

Virtual Reality Film: A visual text that shows, and normally relies on, cyberspace in the form of computer games, the Internet, digital databases and the like. It allows for a visualization of the complex relationship between humanity and information technologies. Examples may include *The Matrix*, *eXistenZ*, *Avatar*, or *Tron: Legacy*.

Gender Differences in Advertising Engagement Using the Case of Facebooks Ads



Eva Lahuerta-Otero

University of Salamanca, Spain

Rebeca Cordero-Gutiérrez

University of Salamanca, Spain

INTRODUCTION

In the last decades, the use of 2.0 technologies has dramatically changed the landscape of a globalized world. Social networks have reached a predominant role within the communication channels as users can get instant access to updated information. Social network profiles grow every second. According to Social Media Today (2015), nearly half of the global population (3,175 billion people) is an active Internet user. Social media have become increasingly fashionable with 2,206 billion active users, 87.2% out of which using mobile platforms. This means 12 new active mobile social users join these platforms every second, meaning one million per day! Therefore, billions of users and companies, with multiple social accounts, are constantly connected.

Regarding the social network concept, the literature provides a variety of definitions of this notion. According to Boy and Ellison (2007), virtual social networks or social network sites are a service based on a web platform that allows people to create their one public or semi-public profile within a limited system. It also allows the possibility of creating contact lists and visualizing the lists of other friends. Other authors also state that social network allows an information sharing and relationship generation by means of interaction with members with common interests (Bigné *et al.*, 2010; based on Preece, 2000; Wiertz & De Ruyter, 2007).

From a set of different social network categories, this study focuses on the horizontal network (ONTSI, 2011), where the use is not restricted to a group of users with a particular interest (gastroonomy, cooking or travelling social networks, for example). Users join horizontal social networks, Facebook in this case, with communication or entertainment purposes.

The web now is a participatory community where not only users, but also companies can co-create, share and modify content. According to Porter Novelli (2012), Spain is the third worldwide country in terms of active users (77%) and 83% of Spaniards are willing to follow a brand on a social network, which is the highest rate in Europe. Consequently, companies and brands invest on social networks to promote their products and to improve customer engagement (Camarero & Cabezudo, 2014). Thus, these platforms emerge as important marketing and communication tools to influence customers. By analyzing users' comments on social networks, firms can discover their tastes, wishes and needs. Therefore, firms can better understand users' behaviors and they can assess satisfaction levels within the purchase process (Kozinets, 2002). By knowing customers better, companies can design specific advertising campaigns addressing the needs of the digital users. These commercial tools provide time and cost savings (Malhotra and Peterson, 2001) and they have a strong personalization power in terms of

DOI: 10.4018/978-1-5225-2255-3.ch293

promotions for individual consumers (Ailawadi *et al.*, 2009; Kannan & Kopalle, 2001)

Social media marketing is a core part of any firms' marketing strategy. With global, connected customers, companies are aware of the need of generating leads that increase purchase intentions by means of social media. Online customer loyalty is hard to get, so marketers are implementing multichannel marketing strategies to generate advertising impacts on any means available. Although traditional media (TV, radio, press) remains important, digital ad spending will reach 30% of total advertising expenditures in 2016 (eMarketer, 2013). Therefore, academics and practitioners need to know how to target customers effectively.

Although Facebook ads platform is a powerful and widely used marketing tool, there is little research about the effectiveness of social marketing campaigns depending on gender. Thus, the purpose of this study is to determine if there are significant differences on Facebook ads' performance according to gender. The article describes the most important metrics regarding social media advertising and, by means of different case studies, it assesses which gender group has greater engagement and performance.

BACKGROUND

Not only do men and women communicate differently, but they think, feel, perceive, react, respond, love, need, and appreciate differently (Gray, 2009, p5).

The last several decades have brought many changes, including but not exclusively the technological ones. The role of women in society has evolved internationally. Around 70% of all working-age women work outside of the home (Meaning, 2013) and they represent the majority of professional workers in many countries, including Spain (Sheehan, 2013). Not only have women changed their working roles, but they have also increased their influential power as consumers.

Women have the decision or influence power in 85% of all consumer goods (Marketing to Women Quick Facts, 2011). However, even if men and women are more equal today, they respond to completely different stimuli when viewing and evaluating advertising messages (Popcorn & Marigold, 2000).

Moreover, women and men use different parts of their brains when processing information. Whereas men processing distributes across the entire brain, women's processing focuses on the frontal lobes, where humans also process multi-tasking (Hotchkiss, 2008).

Since advertisers know men and women process information differently, they have historically relied on stereotypes to establish links with consumers so they can process the information correctly. It is efficient to use gendered stereotypes in advertising because consumers immediately associate the brand, product or service within a sexual category (Sheehan, 2013). However, in terms of social media use, men and women exhibit similar rates: In America, 68% of women and 62% of men use social network sites (Perrin, 2015). Nevertheless, women seem to be more active as they like and share much more content in social media, compared to men (53% versus 34% in 2013) (Casas, 2014).

Gender is not a mere biological concept. It refers to a physiological phenomenon regarding the learned behaviors in relation to gender and men and women's attitudes (Gerrig & Zimbardo, 2002). Men and women tend to have different attitudes and behaviors not only as a result of genetics, but also because of their socialization experiences (San Martín & Jiménez, 2011). Women are usually more involved in shopping activities and they engage more in social media (Cordero-Gutiérrez & Santos-Requejo, 2016). Mazman & Usluel (2011) also observe that gender helps to explain differences in the usage patterns of social networks.

One of these stereotypes has to do with women and computer-related studies and jobs. Regarding attitudes towards computers, women usually exhibit less positive attitudes, as they have higher

levels of computer anxiety (Comber *et al.*, 1997; Kirkpatrick & Cuban, 1998). Women usually like and use computers less than men (Brosnan, 1998; Collis, 1985; Durndell *et al.*, 1987; Meredith *et al.*, 1998) and literature suggests that computer-related careers and information technologies are male stereotyped (Durndell & Lightbody, 1993).

In terms of social media usage, men prefer it to for doing business or dating, whereas women prefer to stay in touch with family and friends, to share content and to entertain themselves. Women tend to have a stronger positive attitude toward brands, and they follow brands more frequently than men do in order to get promotions and discounts. (Simmons Connect, 2013). In terms of time spent on social media, women statistically spent more time (3 hours on average) compared to men (2 hours) (Sánchez-Valle & Frutos-Torres, 2012).

According to FinancesOnline.com (2014) women dominate several areas of social media:

- There are more women in top social sites: Facebook, Twitter, Tumblr, Instagram, and Pinterest
- Women use social media more often daily: 30% of online U.S. women check their profiles several times in a day versus 26% of online U.S. men
- Women interact with brands more often: 55% of online U.S. women access deals compared to 36% of online U.S. men
- 58% of those who consume news in social media are women
- More women use smartphones and tablets to access social media
- Women drive the trend towards visually oriented social sites such as Instagram, Tumblr, and Pinterest, which are the three fastest growing social media networks today

Regarding Internet advertising, some authors believe men prefer web ads to traditional media because digital ads provide interactive features they enjoy (Bezjian-Avery *et al.*, 1998). Conse-

quently, women are less likely to take an action on paid digital ads, since they ignore social media ads more often than men do (Simmons Connect, 2013).

Finally, when referring to computer-related courses and masters, men and women also exhibit different behaviors. Women are usually less likely to be attracted to computer courses and computer-related careers, although this phenomenon is slowly changing (Li & Kirkup, 1993).

Despite the fact that women finish their university studies more than men (40 vs. 23%) in the OECD, only 5% of girls in the OECD are willing to study an engineering or computers degree, compared to 18% of male teenagers (Pisa in focus, 2014). In Spain, women represent the 59% of all university degrees, but only 21% of the total graduates in computer-related degrees are women. (Llaneras, 2015)

Consequently, and given the fact that social media is getting bigger with no signs of stop (Lahuerta-Otero & Cordero-Gutiérrez, 2016a, 2016b) the following case studies aim to determine if gender is a significant condition for users when reacting to social media advertising.

METHODOLOGY

With 1.04 billion daily active users, Facebook is the most used social network in the world. It accounts for the 23.39% of the worldwide traffic (Shareaholic, 2014) and every 60 seconds 510,000 comments are posted, 293,000 statuses are updated, and 136,000 photos are uploaded in this social network (The social skinny, 2012).

This research uses the case study method to analyze the results obtained by a research group from the University of Salamanca, Spain using Facebook ads platform to promote their activities.

This research presents the results of several advertising campaigns done by the research group across the year 2015. We have selected campaigns with different goals, according to the possibilities of Facebook Ads. Some of the campaigns want to increase the likes of the fan page, whereas

others want to increase traffic to the website of the research group. More specifically, ads intend to drive users to the website so students and prospective students could see the different master degrees and courses available.

These two different campaigns enables us to study specific aspects of social media advertising, as well as users' behavior in these media. First, we asked Facebook Ads to split the data by gender. Then, we statistically treated the reports in order to analyze the data. Finally, we divided the different performance metrics into three separate groups so the results could be useful for both academics and practitioners. The groups of metrics are:

- **Reach and Reaction:** This set of metrics represents the social and organic reach obtained in the campaigns, as well as the reactions obtained (actions and people who reacted to the ad).
- **Engagement:** These metrics refer to the performance of the ads on the different campaigns.
- **Cost:** The last group of metrics measures several advertising costs in order to reflect the investment alternatives of the university research center used in this case study.

RESULTS

This study presents the results obtained in several social media advertising campaigns from a research center group of the University of Salamanca (Spain) dedicated to teaching courses, degrees, masters and seminars in the computer and information technologies field. It is important to note that campaigns differ on their objectives, so there are remarkable differences that we need to consider when analyzing their performance.

Those campaigns whose primary goal is increasing the likes on a fan page require a greater engagement of the user, because they will receive regular information about the organization once they click on the like button. However, companies

where the objective is to increase website clicks imply a lower level of commitment, as users can control the actions they want to perform within the page and the content they decide to watch.

Once made this distinction, we will analyze the groups of metrics previously described in the methodology section.

With regard to reach and reaction metrics, we observe that likes campaigns get lower reach. This is because users need a stronger implication with the fan page as mentioned above. This means that organizations need a greater advertising effort (also in terms of budget, as like campaigns are usually more expensive on Facebook). It also means that results in terms of reach can be lower, compared to other campaigns with the similar budget but different objectives. As these characteristics of Facebook ads campaign can influence the chances of students' reactions, since the numbers will be lower in absolute terms, we will compare the results in relative terms.

About the first set of metrics, reach and reactions taken by users, we can see how women exhibit a larger number of reactions on likes campaigns (see Figures 1 and 2). Therefore, they interact more with the page, even though the initial ad reach is lower for this gender group. In relative terms, taking into account the number of people who react on each gender, 15.44% of women took an action when they saw the click on the website ad, compared to 14.17% of men. On the contrary, where the goal is to increase page likes, 3.08% of women interacted with the ad, compared to 3.15% of men. This may be due to the fact that men in general pay higher attention to technological content so they are willing to like the page in order to get the latest news and updates. However, women tend to use social media also with academic purposes, so ads become more relevant to them. We observe that, not only do women interact more with social ads, but they also tend to repeat more a specific action (54.89% of women vs 50.80% of men). As mentioned in the literature review, women need to be more confident about their actions in the digital world (they exhibit higher levels of



Figure 1. Reach and reaction (website clicks campaign)

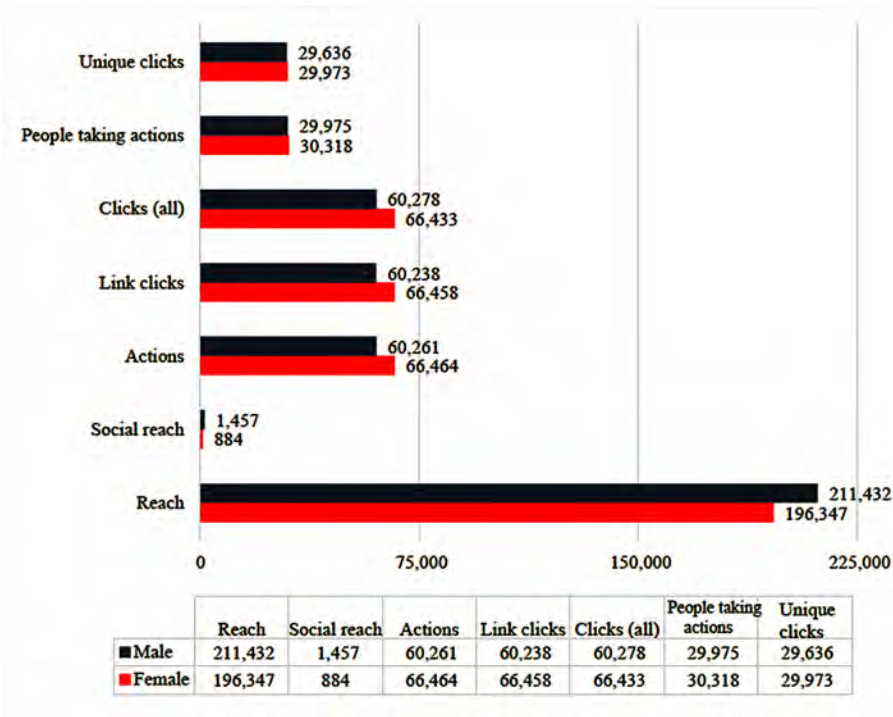
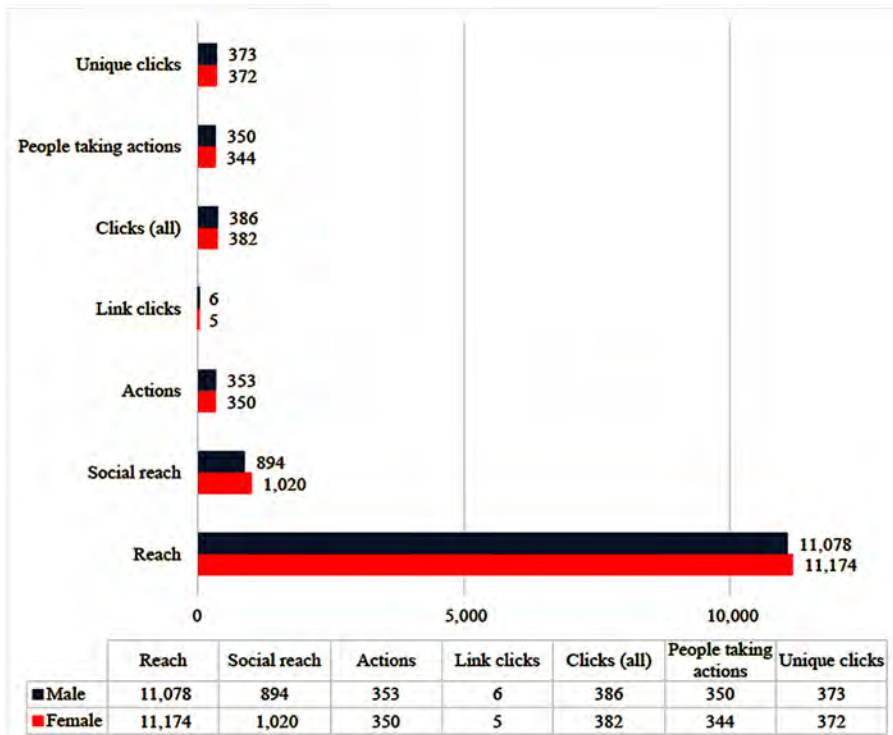


Figure 2. Reach and reaction (likes campaign)



computer anxiety), so they click again to confirm the same information about the educational offers the research center shows in the ad.

With regard to engagement (Figures 3 and 4), we measure the number of clicks or results obtained related to the number of people reached or the number of times the ad was shown. We observe that the results are in line with what was mentioned before. Women exhibit greater engagement levels in clicking on an ad with academic content. These results are logical, since the feminine gender need to reflect more before taking a decision, especially if that implies high expenditures. Women click more in this type of ads to inform themselves and, even if the topic is not of initial relevance to them, this tendency is changing. On the contrary, where campaigns aim to increase fan page likes, ads are more efficient in men. This is because men tend to show higher interest on ICT and computer-related topics, as previously mentioned. We need to highlight an exception because, if we look at the proportion of results, women performed a larger number of actions compared to the times the ad was shown. This may be due to their curiosity in discovering the content of the ad, even though this fact does not turn into a likes increase.

Lastly, concerning cost, we can observe that there also exist gender differences. If ads try to drive traffic to the website, the cost of clicks is lower for women. We can also see how the cost of ads is less than a euro cent per click, so the campaigns perform effectively (see Figures 5 and 6). However, where the aim is to increase page likes, we observe that cost is sensibly lower for men. The results are consistent with all the results previously explained. In the end, a stronger reaction and higher ad efficiency imply lower costs. If we focus on the type of ad, website clicks campaigns are cheaper. This is due to the fact that like campaigns, as previously stated, require a stronger user commitment as already explained. This complex action requires an active user engagement, and consequently, a higher cost.

FUTURE RESEARCH DIRECTIONS

The present research suffers from several limitations we want to deal with in future studies. First, even if the quantitative results seem to indicate the campaign's differences across gender, they only refer to a single data source. Moreover, we cannot deny that many social network users are reluctant

Figure 3. Engagement (website clicks campaign)

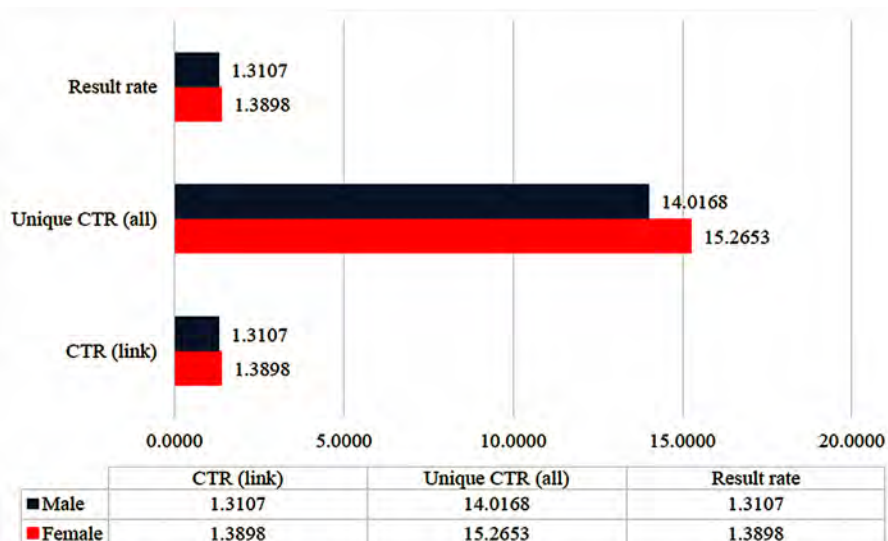
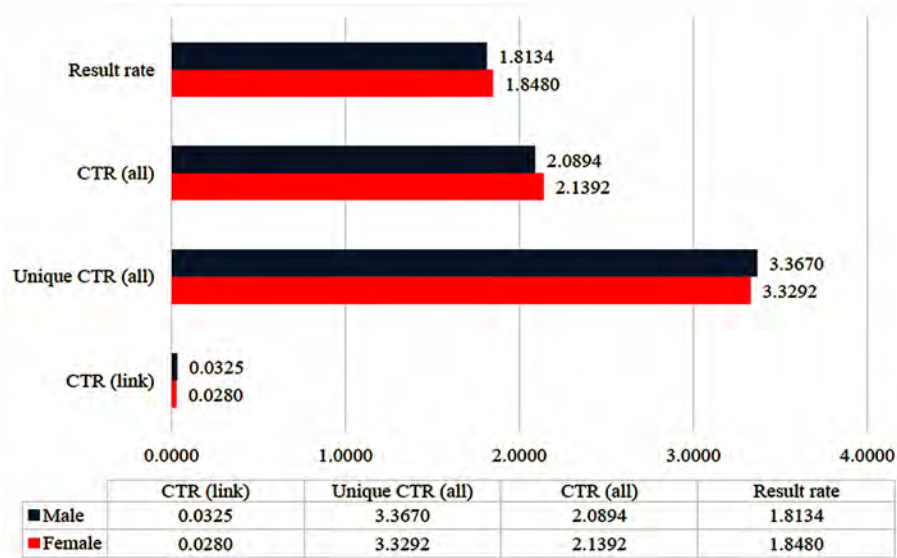


Figure 4. Engagement (likes campaign)



to advertising, so they ignore the ads they are exposed to. Furthermore, this case study belongs to a research group from a public University with only two types of campaigns. Moreover, advertising campaigns were designed following certain age and location criteria (not in terms of gender) so these variables could also affect results. Lastly, the research group works with limited budget, which

could also constraint results. Future research will intend to enlarge the sample and perform several analyses on diverse teaching centers across time, to see which campaigns become more efficient in terms of gender. Additional research could also include alternative variables (such as gender, country of origin or field of studies) to test if differences across groups prevail.

Figure 5. Cost (website clicks campaign)

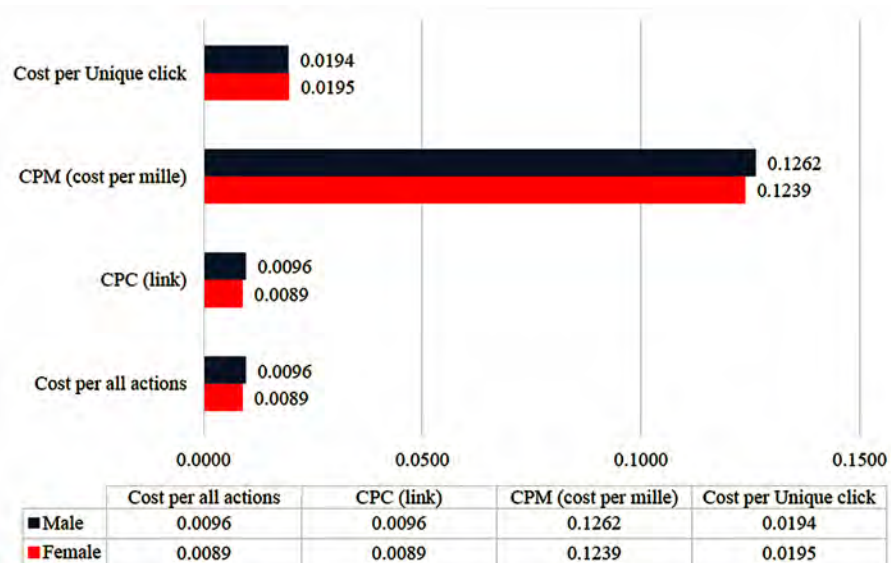
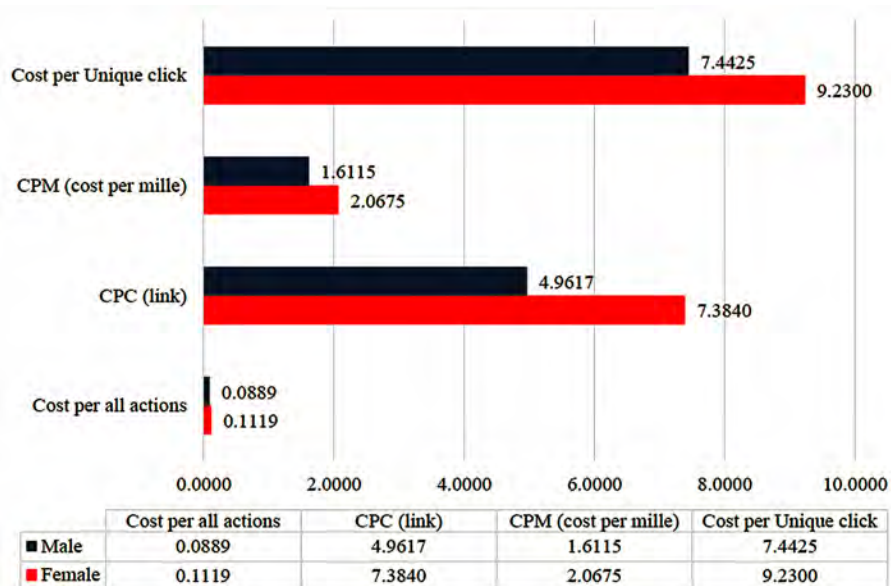


Figure 6. Cost (likes campaign)



CONCLUSION

As we have seen, social networks and their development constitute a powerful and efficient advertising media. Previous studies have proved the efficacy of advertising in horizontal networks to promote academic-related content (Lahuerta-Otero and Cordero-Gutiérrez, 2016b). However, the present study goes a step further and tries to identify behavioral differences of users depending on gender.

The billions of daily active users engaged in social media constitute an excellent source of prospective students. Many organizations are already present in virtual environments so competition increases every day. Therefore, firms and brands need to find alternative, efficient ways to communicate with current and prospective students.

After analyzing gender differences, we can conclude that women and men process ads differently. This is also the tendency of the case study. After concluding the campaigns, a total of 88 people enrolled in the promoted masters, where 31.8% were women. On the fan page of the case

study, the proportion of women is bigger (45% of women compared to 54% men). We can observe that, even though women reacted more to ads, organizations need to make further efforts in order to convince them to enroll in computer-related superior studies.

Accordingly, based on their genetic and sociological differences, women need more time to process ad information. However, they exhibit a higher degree of engagement if the objective is to increase traffic to the website. They might need to click on the same ad several times, but they are willing to analyze the information as they can use it for their academic purposes. On the contrary, men are more straightforward and they prefer to like the page so they can receive news and information directly on their Facebook wall. Consequently, advertisers should customize ads depending on gender, as this will increase their effectiveness. Bearing in mind the restrictions of the Facebook Ads platform (number of characters on the title and text), organizations should focus on attracting men and women by using different, segmented campaigns.

REFERENCES

- Ailawadi, K. L., Beauchamp, J. P., Donthu, N., Gauri, D. K., & Shankar, V. (2009). Communication and Promotion Decisions in Retailing: A Review and Directions for Future Research. *Journal of Retailing*, 85(1), 42–55. doi:10.1016/j.jretai.2008.11.002
- Bezjian-Avery, A., Calder, B., & Lacobucci, D. (1998). New media interactive advertising vs traditional advertising. *Journal of Advertising Research*, 38(4), 23–32.
- Bigné, E., Currás, R., Ruiz, C., & Sanz, S. (2010). *Factores determinantes del comportamiento del consumidor en las redes sociales*. Un análisis de los internautas españoles, XXII Congreso Nacional de Marketing. AEMARK.
- Boyd, D., & Ellison, N. (2007). Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication*, 13(1), 210–230. doi:10.1111/j.1083-6101.2007.00393.x
- Brosnan, M., & Lee, W. (1998). A cross-cultural comparison of gender differences in computer attitudes and anxieties: The United Kingdom and Hong Kong. *Computers in Human Behavior*, 14(4), 559–577. doi:10.1016/S0747-5632(98)00024-7
- Camarero, C., & Cabezudo, R. (2014). De «me gusta» a «quiero colaborar»: determinantes de la intención de los fans de marca de participar en cocreación. In *Experiencia y comportamiento del cliente en un entorno multicanal: claves de éxito para fabricantes y detallistas* (pp. 273–289). Oviedo: Cátedra Fundación Ramón Areces de Distribución Comercial.
- Casas, J. (2014). *Tendencias en redes sociales 2014: datos, infográficos y estadísticas*. Retrieved October 23, 2015, from <http://postcron.com/es/blog/todas-las-tendencias-en-redes-sociales-2014-infograficos-estadisticas-y-tips/>
- Collis, B. (1985). Reflections on inequities in computer education: Do the rich get richer? *Education and Computing*, 1(3), 179–186. doi:10.1016/S0167-9287(85)91519-5
- Comber, C., Colley, A., Hargreaves, D. J., & Dorn, L. (1997). The effects of age, gender and computer experience upon computer attitudes. *Educational Research*, 39(2), 123–134. doi:10.1080/0013188970390201
- Cordero Gutiérrez, R. & Santos Requejo, L. (2016). Intention to participate in online commercial experiments by social network's users. Differences in gender and age. *Management Research Review*, 39(4).
- Durndell, A., & Lightbody, P. (1993). Gender and computing: Change over time? *Computers & Education*, 21(4), 16–25. doi:10.1016/0360-1315(93)90036-I
- Durndell, A., Macleod, H., & Siann, G. (1987). A survey of attitudes to, knowledge about and experience of computers. *Computers & Education*, 11(3), 167–175. doi:10.1016/0360-1315(87)90051-0
- eMarketer. (2013). *US digital ad spending 2011-2017*. Author.
- Facebook Ads. (2015). *Easy and effective Facebook Adverts*. Retrieved March, 1, 2016, from <https://www.facebook.com/business/help/510910008975690>
- FinanzasOnline.com. (2014). *Who Runs the Social Media World: Men or Women?* [infographic]. Retrieved March, 10, 2015 from: <https://www.brandwatch.com/2014/03/social-media-and-women/>
- Gerrig, R. J., & Zimbardo, P. G. (2002). *Glossary of psychological terms, Psychology and life* (16th ed.). Boston: Allyn and Bacon.
- Gray, J. (2009). *Men are from Mars, women are from Venus: Practical guide for improving communication*. Zondervan.

- Hootsuite. (2016). *The Social Media Glossary: 207 Essential Definitions*. Retrieved January 27, 2016, from <https://blog.hootsuite.com/the-2015-social-media-glossary-207-essential-definitions/>
- Hotchkiss, G. (2008). *Human Hardware: Men And Women*. Retrieved February 29, 2016, from <http://searchengineland.com/human-hardware-men-and-women-13614>
- Kannan, P. K., & Kopalle, P. (2001). Dynamic pricing on the Internet: Importance and implications for consumer behavior. *International Journal of Electronic Commerce*, 5(3), 63–68.
- Kirkpatrick, H., & Cuban, L. (1998). Should we be worried. What the research says about gender differences in access, use, attitudes, and achievement with computers. *Educational Technology*, (July–August), 56–61.
- Kozinets, R. V. (2002). The Field Behind the Screen: Using Netnography for Marketing Research in Online Communities. *JMR, Journal of Marketing Research*, 39(1), 61–72. doi:10.1509/jmkr.39.1.61.18935
- Lahuerta-Otero, E., & Cordero Gutiérrez, R. (2016a). Advertising techniques in the digital world: The new era of social communications. In *Encyclopedia of e-Commerce development implementation, and management*. Western Illinois University. doi:10.4018/978-1-4666-9787-4.ch083
- Lahuerta-Otero, E., & Cordero Gutiérrez, R. (2016b). How to promote teaching centers by using online advertising: A case study. In *Encyclopedia of e-Commerce development implementation, and management, Edition*. Western Illinois University. doi:10.4018/978-1-4666-9787-4.ch020
- Li, N., & Kirkup, G. (2007). Gender and cultural differences in Internet use: A study of China and the UK. *Computers & Education*, 48(2), 301–317. doi:10.1016/j.compedu.2005.01.007
- Llaneras, K. (2015). *Género, educación y la brecha de las carreras técnicas*. Retrieved January 11, 2016, from <http://politikon.es/2015/02/17/genero-educacion-y-la-brecha-de-las-carreras-tecnicas/>
- Malhotra, N. K., & Peterson, M. (2001). Marketing research in the new millennium: Emerging issues and trends. *Marketing Intelligence & Planning*, 19(4), 216–235. doi:10.1108/EUM0000000005560
- Marketing to women quick facts. (2011). Retrieved January 3, 2016, from: <http://she-conomy.com/facts-on-women>
- Mazman, S. G., & Usluel, Y. K. (2011). Gender Differences in Using Social Networks, Turkish Online. *Journal of Educational Technology-TOJET*, 10(2), 133–139.
- Meaning, H. G. S. (2013). Gender and Advertising. *Controversies in Contemporary Advertising*, 89.
- Meredith, A., Helen, J., & Woodcock, L. (1998). Women can do I.T. – but why don't they? In *Conference Proceeding: Winds of change – women and culture of universities* (pp. 762–766). Sydney, Australia: University of Technology.
- ONTSI. (2011). *Las Redes Sociales en Internet*. Retrieved October 15, 2015, from <http://www.ontsi.red.es/ontsi/es/estudios-informes/estudio-sobre-el-conocimiento-y-uso-de-las-redes-sociales-en-espa%C3%B1>
- Perrin, A. (2015). *Social Media Usage: 2005-2015*, Retrieved January 20, 2016, from <http://www.pewinternet.org/2015/10/08/social-net-working-usage-2005-2015/>
- Pisa in focus. (2014). *¿A qué tipo de carreras aspiran los chicos y las chicas?* Retrieved January 27, 2016, from <http://www.oecd.org/pisa/pisaproducts/pisainfocus/PISA%20in%20Focus-n%C2%B014%20ESP.pdf>

Popcorn, F., & Marigold, L. (2000). *Evolution: The eight truths of marketing to women*. New York: Hyperion.

Porter Novelli. (2012). *Estudio Social Media consumer*.

Preece, J. (2000). *Online communities: Designing usability, supporting sociability*. Chichester, UK: Wiley.

San Martín, S., & Jiménez, N. H. (2011). Online buying perceptions in Spain: Can gender make a difference? *Electronic Markets*, 21(4), 267–681. doi:10.1007/s12525-011-0074-y

Sánchez-Valle, M., & Frutos-Torres, B. (2012). *Adolescent in social media Incarnating the Self through Digital Social Networks*. Retrieved November 19, 2015, from <https://telos.fundaciontelefonica.com/url-direct/pdf-generator?tipoContenido=articuloTelos&idContenido=2012071612040003&idioma=es>

Shareaholic. (2014). *Facebook and Pinterest are the King and Queen of Social Referrals*. Retrieved December 27, 2015, from <https://blog.shareaholic.com/social-media-traffic-trends-04-2014/>

Sheehan, K. B. (2013). *Controversies in contemporary advertising*. Sage Publications.

Simmons Connect. (2013). *Miles apart: how women and men use social media and mobile*. Retrieved November 21, 2016, from <http://www.marketingdirecto.com/wp-content/uploads/2015/09/smartphone-social-media-usage-men-vs-women-infographic.jpg>

Social Media Today. (2015). *10 Amazing Social Media Growth Stats From 2015*. Retrieved January 4, 2016 from <http://www.socialmediatoday.com/social-networks/kadie-regan/2015-08-10/10-amazing-social-media-growth-stats-2015>

The social skinny. (2012). *100 social media statistics for 2012*. Retrieved December 12, 2015, from <http://thesocialskinny.com/100-social-media-statistics-for-2012/>

Wiertz, C., & De Ruyter, K. (2007). Beyond the Call of Duty: Why Customers Participate in Firm-Hosted Online Communities. *Organization Studies*, 28(3), 349–378. doi:10.1177/0170840607076003

ADDITIONAL READING

Nielsen (2014). The digital consumer report. Retrieved March 12, 2016 from: <http://es.slideshare.net/tinhanhvy/the-digital-consumer-report-2014-nielsen>

Pew Research Center. (2012). The Demographics of Social Media Users. Retrieved March 12, 2016 from: <http://www.pewinternet.org/2013/02/14/the-demographics-of-social-media-users-2012/>

The Huffington post (2016). Women and Social Media. Retrieved March 13, 2016 from: <http://www.huffingtonpost.com/news/women-and-social-media/>

KEY TERMS AND DEFINITIONS

Engagement: Refers to the acts of talking to, messaging or otherwise interacting with other people on social networks. This broad term encompasses a several different types of actions on social media, from commenting on Facebook posts to participating in Twitter chats (Hootsuite, 2016).

Result Rate: The number of results the ad received divided by the number of impressions.

Social Media Marketing: The use of social media by marketers to increase brand awareness, identify key audiences, generate leads, and build meaningful relationships with customers. Social media marketing should be well coordinated with social customer service, community management, and social selling activities to create seamless relationships with customers across their life cycle (Hootsuite, 2016).

Social Network: A service based on a web platform that allows people to create their one

public or semi-public profile within a limited system. It also allows the possibility of creating contact lists and visualizing the lists of other friends (Boyd & Ellison, 2007).

Social Reach: The number of people the ad was serve to with social information. For example: if three people see an ad two times each that says a friend likes your page, it counts as three social reaches.

Unique Clicks: The total number of unique people who have clicked on the ad. For example, if three people click on the same ad four times, it counts as three unique clicks.

Unique CTR: The number of people who clicked on the ad divided by the number of people you reached. For example, if you received twenty unique clicks and the ad was served to 1,000 unique people, your unique click-through-rate (CTR) would be 2%.

The Gender Dimension in Urban Air Quality



Theodora Slini

Aristotle University of Thessaloniki, Greece

Fotini-Niovi Pavlidou

Aristotle University of Thessaloniki, Greece

INTRODUCTION

In the light of several reports (IPCC, 2001; UNDP, 2007), it is evident that human contribution in global warming and climate change can be attributed differently between genders. Within the developed societies female's contribution tends to be less than male's on average, due to different social roles but also to different environmental consciousness and behaviour. E.g. women have different consumption habits (use of public transport, walking or cycling trips) and are more likely to support greenhouse gas emissions reduction policies related (support of recycling and energy efficiency initiatives). In Europe, direct and indirect energy consumption tends to be higher among male than female inhabitants a fact that is independent of income and age, approximately 39% higher in Germany and 22% higher in Sweden (genderSTE, 2014). Worldwide, it is declared that money paid to females is more mainly consumed on family needs, e.g. food and clothing, while money paid to males is more likely to be consumed on leisure, energy-intensive goods that cause high emissions. Thus, a greater proportion of an average man's carbon footprint is due to leisure than an average woman's (Druckman et al., 2012). Last but not least, women are more sensitive to extreme weather conditions, such as floods and heatwaves and more likely to experience fuel poverty due to income inequalities (Fouillet et al., 2007).

At the same time, the global economic system is male dominated and while women have an increasingly significant role in the economic system that led to global warming, they are underrepresented in the decision making, industries and organisations focused in the environmental (climate change, transport and energy) sector. Female often surpass the male participation in voluntary environmental campaigning actions, accounting for approximately two thirds in Europe (WEN, 2010). The average proportion of women in national ministries responsible for the environment, transport and energy, by level of authority in the EU-27 is extremely low, reaching the 25.6% in 2012 according to data generated from the European Institute for Gender Equality. The aim of the current study is to review the current conditions in Europe and Greece and identify the gaps, if any, of women's involvement in climate change and global warming decision making.

BACKGROUND

Gender Priorities in the European Union

Climate change affects both women's and men's living conditions, welfare and wellbeing, however due to gender roles, women do not affect the environment in the same way as men, and in many

countries women's access to resources, and hence their opportunities to manage conditions and adapt are quite limited.

At the same time, environmental policies are characterized by lack of sensitivity to women's different economic and social status and needs, having as a result women to be directly and disproportionately affected from environmental degradation. While consumption and lifestyle patterns still differ between two genders, with women to consume less and being more environmentally conscious, "women are clearly under-represented in environmental negotiations, budget deliberations and decisions on achieving a green, sustainable economy" (EC, 2012). According to 'Gender aspects of the economic downturn and financial crisis European Parliament resolution of 17 June 2010 on gender aspects of the economic downturn and financial crisis (2009/2204(INI))', the European Parliament "urges the need to encourage women in local entrepreneurial initiatives in green economy". The European Parliament (2012) resolution of 11 September 2012 'on the role of women in the green economy (2012/2035(INI))' "calls on the Commission and the Member State to introduce gender equality into all environmental policy areas" and introduce gender equality into all environmental policy areas, and at all levels of economic decision-making. On 14 November 2012 the European Commission published its proposal for a Gender Diversity Directive for improving gender balance on company boards, obliging companies with less than 40% of women non-executive directors to make significant efforts to make appointments in the next seven years to reach this target.

Furthermore, the Opinion of the European Economic and Social Committee on 'The gender dimension in the Europe 2020 Strategy' 2013/C 76/02 (EU Legislation, 2013), the Commission is "highlighting the essential role played by women in sustainable development, while "women can have a key influence on decision-making concerning the environment, particularly with regard to climate change policy. This is a new opportunity

for women, who can play a key role and improve their personal and financial situation by getting involved with the new and emerging green economy, which is a crucial sector for development and job creation". After all "women's activism was critical in getting the conversation started," according to historian David Stradling, author of "Smokestacks and Progressives: Environmentalists, Engineers and Air Quality in America, 1881-1951."

In response to these requirements a gender perspective is a critical assumption towards a sustainable environment and a green economy in terms of the ecosystem, consumption, food, growth, transport, energy and citizen welfare.

Gender Priorities in Greece

According to the National Strategic Reference Framework (NSRF) for 2014-2020, the main tools for the implementation of gender policy will be integration of equal opportunities between men and women in all institutions, policies and actions (gender mainstreaming) and the assumption of certain specific activities in the following areas:

- Equal participation women in the labor market,
- The participation of women outdoor activities aimed at local development,
- Promotion of inclusion of women, prevention and control of female poverty and all forms of gender-based violence,
- Mainstreaming of gender issues in social protection and health,
- Support family,
- Equal participation of women in public life and processes of political, technological, social and economic decisions
- Combat discrimination based on sex and gender stereotypes and
- The integration of gender equality in public policies, monitoring and evaluation.

However, the Greek Legislation does not include the gender perspective as a critical parameter

through its development and implementation and in many cases it is characterized as neutral or negative, according to a report by the General Secretariat for Gender Equality (2013). For example,

1. The Law 4067/2012 (OG A79/9-3-2012) the “New Construction Regulation” that includes a number of provisions on accessibility of buildings and built environment. Regarding the motivation for environmental enhancement and the improvement of quality of life in densely populated urban areas, which gender neutral with no comment on specific needs.
2. The master plan for urban planning and environmental protection of the greater Thessaloniki area (L. 1561/1985) refers to the daily needs of citizens ignoring the gender matters. Special requirements due to diverse living habits and routines as well as the different ways in which environment is perceived and used between men and women should be considered for the optimum urban planning and city upgrading.

The most recent L. 3653/2008 on Research and Technology establishes the minimum proportion of male/female participation in national committees of research and development, that is at least 1/3 of each gender, as long the candidates are equally qualified. However, in practise the rule rarely applies.

MAIN FOCUS OF THE ARTICLE

Gender, Air Quality, and Global Warming

Despite it is widely accepted that balanced participation of men and women in decision-making is a prerequisite for improving the operation of democracy and society, the ongoing under-representation of women in political and economic decision-making reflects a basic democratic deficit

in Europe and the broader international context. At the same time it is important to highlight that the female population is more vulnerable in effects of environmental degradation and poor air quality (UN, 2014).

There are several studies that prove that poor air pollution conditions can generate negative effects in the female population in both rural and urban sites (Xu, Hu, Ha, & Roth, 2013; Huisman et al., 2005; Oiamo & Luginaah, 2013; Künzli et al., 2010). The proportion of women and young children is mainly affected as these groups of population spend most of their time indoors where solid fuels, fire places and cooking stoves are in extensive use. According to WHO (2014), around 3 billion people cook and heat their homes using open fireplaces and stoves using various burning materials. These materials are not always the appropriate ones, causing the death of over 4 million people prematurely from illness due to the poor indoor air quality. The mortality of premature deaths among children under 5 attributed to household air pollution is also extremely high.

Additionally according to The European environment — state and outlook 2010 (EEA, 2010), vulnerable population groups (e.g. pregnant women) are more threatened by the bio-accumulation of organic persistent pollutants such as mercury. Chen et al. (2005) and Miller et al. (2007), the associations of PM_{2.5} and cardiovascular morbidity and mortality were greater in women above 60 and at the same time Xu et al. (2013) states that air pollution (referring to air pollutants such as NO₂, SO₂, PM_{2.5}, O₃ and CO) increases risk for hypertension in pregnant women during early pregnancy and the full gestational period.

Ambient air pollution has stronger effects on females partly due to autoimmune disorders. Females were more likely to report cardinal symptoms after controlling for income, age and chronic diseases (Oiamo & Luginaah, 2013). As Bakke et al. (2007) suggests female gender has reported health symptoms more often than did men and complained more about physical but not psychosocial factors gender, psychosocial,

and physical environment factors were related to symptoms and perceived indoor climate. Female gender seems to be highly associated with the sick building syndrome and the symptom of atopia (Magnavita, 2014).

Several studies have linked air pollution exposure to autism, including the Volk et al. (2011) study in *Environmental Health Perspectives* that looked at children in Los Angeles who lived near freeways and the Hallmayer et al. (2011) study that analysed traffic-related air pollution exposure data among pregnant women. The main conclusion of the studies was that a large proportion of the variance in liability can be explained by shared environmental factors in addition to moderate genetic heritability.

In general, studies suggest that health responses to air pollution may differ between women and men, however it is still unclear, whether observed modification is a result of sex-linked biological differences or gender differences in activity patterns or exposure measurement accuracy (Clougherty, 2010). Cautious consideration of gender effects and examination in environmental epidemiology may provide critical information on other social factors that can affect population response to air pollution.

In addition, in countries that face the severity of current financial crisis, structural changes occur in European members states as well, including migration and changing of household structures due to energy poverty. For example, in Greece, an exponential increase of wood and pellet employment is monitored primarily for heating purposes (Slini, Giama & Papadopoulos, 2014; Santamouris et al., 2013), posing major concern especially for women health (Figure 1), based on the results of a survey conducted in Northern Greece by the Department of Mechanical Engineering, Aristotle University of Thessaloniki (AUTH).

In Greece, due to the economic situation the consumption of oil products, which are the main fuel that is used by the household sector, decreased already by 25.7% between 2007 and 2010. What is impressive is the exponential increase in the

use of wood and pellets, a trend that has begun to have its first implications in matters of air quality, both indoor and outdoor, but also in an enhanced deforestation, with major field studies being currently carried out on those matters (Slini et al., 2014). Towards this direction many countries in Europe are adopting policies and effective precautionary measures in order to reduce the exposure of vulnerable population groups in toxic environments (EEA, 2015).

At the same time, research on gender-related disparities are mainly focused on energy consumption profiles and transport use. As pointed out at the Report on Gender Equality and Climate Change (EIGE, 2012a), the most significant gendered differences in energy consumption were monitored in Greece and Sweden, with men energy consumption reaching a proportion of 39% and 22% more than women, respectively. Moreover, it is suggested that women use the public transport or other sustainable means of travel more often than men.

Gender Balance in Environmental Decision Making

As global warming is a multi-fold issue and the abatement policy (e.g. GHG emissions, legislation, and limit values) requires an interdisciplinary approach conducted within environmental authorities. Thus, the gender distribution in authorities responsible for climate change adaptation/mitigation measures is substantial.

According to EU data, concerning environmental governing positions in European States, such as senior and junior ministers in environment, energy and transport at national governments, the women participation was slightly increased from 20% to 27% during the last three years (2012-14), a proportion that is still significantly lower than male proportion (Figure 2). In Greece during the same period there was no female representation.

However, the national administration positions show a slightly improved distribution with the female proportion increasing from 33% to

Figure 1. The average annual fuel and electricity consumption for the years 2011–2012 by type: (a) oil in litres, (b) natural gas in Nm³, (c) electricity in kilowatt-hours and (d) wood in kilograms

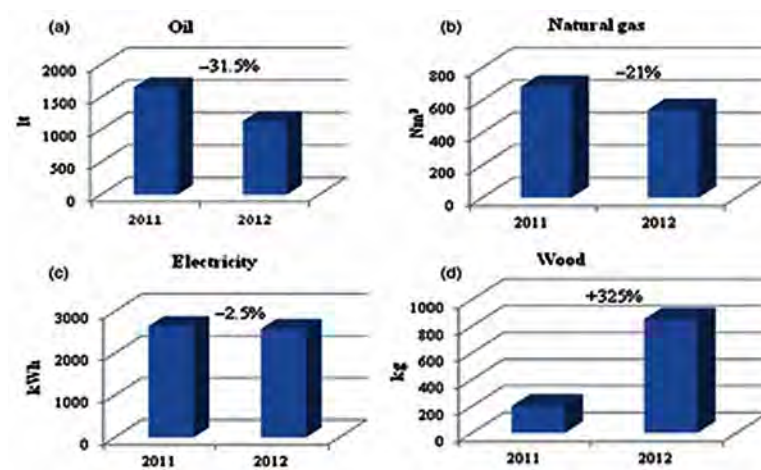
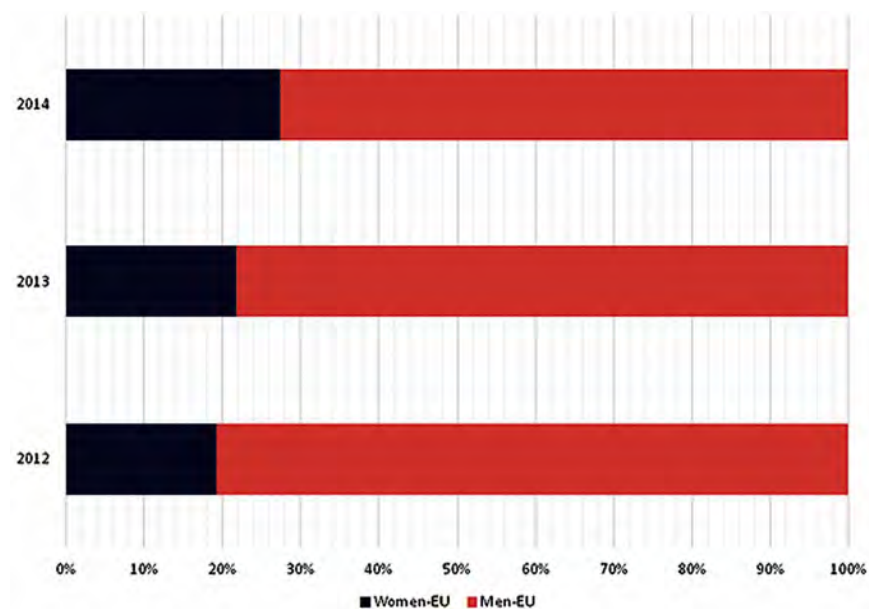


Figure 2. The proportion of men and women in environmental ministers (senior and junior) in European Member States for the period 2012–14



36% between 2012 and 2014. In Greece the data are considerably more balanced compared to the European figures, though decreased from 48% to 45% from 2012 to 2014 (Figure 3).

It is noteworthy to mention that in a broader approach the representation of women in boards of large listed companies is also discouraging,

with major gaps and little progress on women representation and career evolution, as presented in figure 4, in both the European Member States and Greece. More recently, and specifically in October 2014, Greece stands for the proportion of 8.9% of women in highly ranked position with a European average of 20.2% (EC, 2015).

Figure 3. The gender distribution of environmental national administrators (a) in the European Member States and (b) in Greece for the period 2012-14

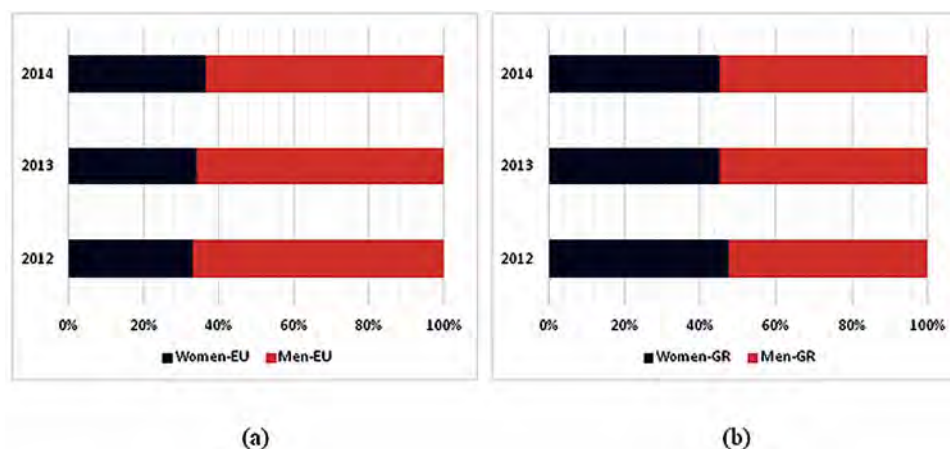
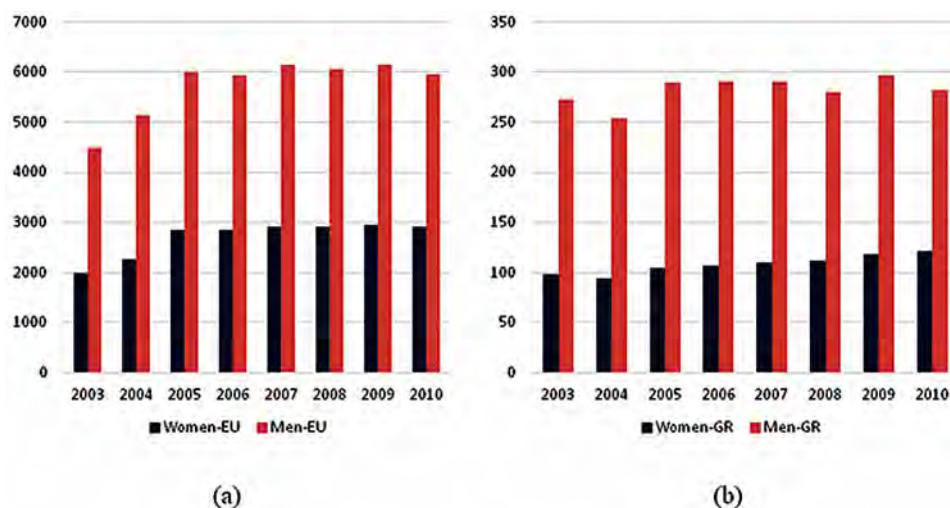


Figure 4. Representation of women and men (absolute numbers) on boards of high listed companies (a) in the European Member States and (b) in Greece for the period 2003-2010



SOLUTIONS AND RECOMMENDATIONS

It is broadly acknowledged that there is still a gap in legislation providing for non-discrimination against women and gender equality in the areas of social security, education and employment (EU Legislation, 2016a). At the same time, the gender dimension has not been taken into consideration in the current and planned initiatives and policies aimed at exiting the crisis (EU Legislation, 2016b),

despite the fact that women play a critical role in driving economic development.

The analysis and evaluation of the current situation is reflected in the development of initiatives and the adoption of strategic plans at national or global levels that are mainly focused in gender mainstreaming in education and training, the labour market, the economic and political decision making. The promotion of women empowerment, the establishment of fair and transparent systems and mechanisms, the enhancement of social and

economic opportunities, the support of inclusive procedures and initiatives for both genders in all community levels and the promotion of reliable and gender-responsive institutions that will support inclusive, participatory and sustainable development processes can foster the positive dynamic of gender equality and balance.

Focusing in gender and the environment, a number of specific, good-practices as well as the development of partnership and collaboration between sustainable development actors and authorities can raise progressive synergy for analyzing the current gender-environment status and promoting novel strategies and operations.

Having the aforementioned facts under consideration, the European Commission has established a set of indicators in order to determine and evaluate the women's participation in decision making on climate change (EIGE, 2012b), that is the following:

- Proportion of women in climate change decision-making bodies at the national level in EU Member State
- Proportion of women in climate change decision-making bodies at the EU level
- Proportion of women in climate change decision-making bodies at the international level
- Proportion of women tertiary graduates of total graduates in natural sciences and technologies at the EU and Member State level.

At a global level, the development and collection of this kind of data will support the study of gender-climate change issues and may provide the benchmark for the development of new indicators about the correlation between gender, environment and social welfare and justice (UNFCCC, 2015). The establishment of “gendered” capacity building through education, training and mentoring programs and action should be at the top agenda of authorities and stakeholders, in order to ensure the women's engagement in technology and further in information science.

Thus, the issue of gender equality and women empowerment in all policy fields and sustainable development sections will enhance the growth potential of a balanced, fair society.

FUTURE RESEARCH DIRECTIONS

Aiming at the support of women, future research directions and solid steps include the development and adoption of strategies for integrating gender perspective into policy-making process, the awareness' raising initiatives concerning the significance of gender perspective in science and innovation. In this frame, the investment in equal representation can be achieved by the support of structural changes in education institutions, boards of directors and councils via processes that highlight the significance of gender analysis. Additional data, both qualitative and quantitative, thorough and systematic analysis and evaluation are required regarding the gender perspectives, not only in environment, climate change, global warming, between and across other research fields and disciplines and innovation. Gender equality and balance need also to be addressed when developing national strategies, institutional strategies, or international research initiatives.

Last but not least, fostering the society participation in gender issues can support the strengthen policy -oriented research on gender equality challenges, leading eventually in societal prosperity, poverty reduction and sustainable development.

CONCLUSION

The current study unveils the relations and links between gender and research and development in cross cutting issues of environmental engineering and science. It is proven that there are dissimilarities between males and females regarding environmental attitudes, contributions and capabilities of adaptation in the mitigation measures as well as the socio-economic impacts of climate change policies. As little progress is achieved, it comes

without saying that gender perspective is a critical assumption towards a sustainable environment, technology and a green economy in terms of the ecosystem, consumption, food, growth, transport, energy and citizen welfare. As the engendering of decision making on climate change will reinforce and will improve the efficiency of the suggested policies, the systematic research and monitoring of gender equality data and development of awareness initiatives are on the top of the agenda of policy recommendations. Quantitative targets and strong commitment on gender equality on science and technology by national governments, international institutes and authorities are critical areas of concern towards sustainable social development.

REFERENCES

- Bakke, J. V., Moen, B. E., Wieslander, G., & Norbäck, D. (2007). Gender and the physical and psychosocial work environments are related to indoor air symptoms. *Journal of Occupational and Environmental Medicine*, 49(6), 641–650. doi:10.1097/JOM.0b013e31806e5fa0 PMID:17563607
- Chen, H. L., Knutsen, S. F., Shavlik, D., Beeson, W. L., Petersen, F., Ghamsary, M., & Abbey, D. (2005). The Association between Fatal Coronary Heart Disease and Ambient Particulate Air Pollution: Are Females at Greater Risk? *Environmental Health Perspectives*, 113(12), 1723–1729. doi:10.1289/ehp.8190 PMID:16330354
- Clougherty, J. E. (2010). A Growing Role for Gender Analysis in Air Pollution Epidemiology. *Environmental Health Perspectives*, 118(2), 167–176. doi:10.1289/ehp.0900994 PMID:20123621
- EC - European Commission. (2012). *Special Eurobarometer 376 “Women in decision-making positions”*. Conducted by TNS Opinion & Social at the request of Directorate-General Justice, March.
- EC - European Commission. (2013). *National Factsheet “Gender balance in Boards”, country: Greece, January*. EC.
- EC- European Commission. (2014). *Report on equality between women and men 2014*. EC.
- EC- European Commission. (2015). *Factsheet March 2015: Gender balance on corporate boards - Europe is cracking the glass ceiling*. EC.
- EEA- European Environment Agency. (2010). *The European environment — state and outlook 2010 (SOER 2010)*. EEA.
- EEA- European Environment Agency. (2015). *The European environment — state and outlook 2015 (SOER 2015)*. EEA.
- EIGE – European Institute for Gender Equality. (2012a). *Gender Equality and Climate Change*. EIGE.
- EIGE - European Institute for Gender Equality. (2012b). *Review of the Implementation in the EU of area K of the Beijing Platform for Action: Women and the Environment Gender Equality and Climate Change*. EIGE.
- EU Legislation. (2013). *Opinion of the European Economic and Social Committee on ‘The gender dimension in the Europe 2020 Strategy’ (own-initiative opinion) 2013/C 76/02*. Available at: <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52012IE1734&qid=1410262659773>
- EU Legislation. (2016a). *European Parliament resolution of 12 March 2013 on the impact of the economic crisis on gender equality and women’s rights (2012/2301(INI))*. Available at: http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.C_.2016.036.01.0043.01.ENG&toc=OJ:C:2016:036:TOC
- EU Legislation. (2016b). *European Parliament resolution of 12 March 2013 on eliminating gender stereotypes in the EU (2012/2116(INI))*. Available at: http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.C_.2016.036.01.0018.01.ENG&toc=OJ:C:2016:036:TOC

European Parliament. (2012). *Resolution of 11 September 2012 on the role of women in the green economy* (2012/2035(INI)). Accessed on October, 26, 2015.

genderSTE: a policy-driven targeted network funded by COST (European Cooperation in Science and Technology). (n.d.). Available at: <http://www.genderste.eu/index.php>

General Secretariat. (2013). *Gender Impact Assessment of 100 Laws and Regulations*. Available at: <http://www.genderimpact.gr/>

Hallmayer, J., Cleveland, S., Torres, A., Phillips, J., Cohen, B., Torigoe, T., & Risch, N. et al. (2011). Genetic Heritability and Shared Environmental Factors Among Twin Pairs With Autism. *Archives of General Psychiatry*, 68(11), 1095–1102. doi:10.1001/archgenpsychiatry.2011.76 PMID:21727249

Huisman, M., Kunst, A. E., Bopp, M., Borgan, J. K., Borrell, C., Costa, G., & Mackenbach, J. P. et al. (2005). Educational inequalities in cause-specific mortality in middle-aged and older men and women in eight western European populations. *Lancet*, 365(9458), 493–500. doi:10.1016/S0140-6736(05)70273-7 PMID:15705459

Künzli, N., Jerrett, M., Garcia-Esteban, R., Basagana, X., Beckermann, B., Gilliland, F., & Mack, W. J. et al. (2010). Ambient air pollution and the progression of atherosclerosis in adults. *PLoS ONE*, 5(2), e9096. doi:10.1371/journal.pone.0009096 PMID:20161713

Magnavita, N. (2014). Work-related symptoms in indoor environments: A puzzling problem for the occupational physician. *International Archives of Occupational and Environmental Health*. doi:10.1007/s00420-014-0952-7 PMID:24917077

Miller, K. A., Siscovick, D. S., Sheppard, L., Shepherd, K., Sullivan, J. H., Anderson, G. L., & Kaufman, J. D. (2007). Long-Term Exposure to Air Pollution and Incidence of Cardiovascular Events in Women. *The New England Journal of Medicine*, 356(5), 447–458. doi:10.1056/NEJMoa054409 PMID:17267905

Oiamo, T. H., & Luginaah, I. N. (2013). Extricating Sex and Gender in Air Pollution Research: A Community-Based Study on Cardinal Symptoms of Exposure. *International Journal of Environmental Research and Public Health*, 10(9), 3801–3817. doi:10.3390/ijerph10093801 PMID:23975108

Santamouris, M., Paravantis, J. A., Founda, D., Kolokotsa, D., Michalakakou, P., Papadopoulos, A. M., & Servou, E. et al. (2013). Financial Crisis and Energy Consumption: A Household Survey in Greece. *Energy and Building*, 65, 477–487. doi:10.1016/j.enbuild.2013.06.024

Slini, T., Giama, E., & Papadopoulos, A. M. (2014). The impact of economic recession on domestic energy consumption. *International Journal of Sustainable Energy*. doi:10.1080/14786451.2014.882335

UN - United Nations. (2014). *Prototype Global Sustainable Development Report*. New York: United Nations Department of Economic and Social Affairs, Division for Sustainable Development.

UNFCCC - United Nations Framework Convention on Climate Change (2015). *Report on gender composition*. FCCC/CP/2015/6.

Volk, H. E., Hertz-Picciotto, I., Delwiche, L., Lurmann, F., & McConnell, R. (2011). Residential Proximity to Freeways and Autism in the CHARGE Study. *Environmental Health Perspectives*, 119(6), 873–877. doi:10.1289/ehp.1002835 PMID:21156395

WHO - World Health Organisation (2014). *Household air pollution and health*. Fact sheet N°292, March. WHO.

Xu, X., Hu, H., Ha, S., & Roth, J. (2013). Ambient air pollution and hypertensive disorder of pregnancy. *Journal of Epidemiology and Community Health*, 68(1), 13–20. doi:10.1136/jech-2013-202902 PMID:24022815

ADDITIONAL READING

Alston, M. (2013). Women and adaptation. *Wiley Interdisciplinary Reviews: Climate Change*, 4(5), 351–358. doi:10.1002/wcc.232

Alston, M. (2014). Gender mainstreaming and climate change. *Womens Studies International Forum*, 47, 287–294. doi:10.1016/j.wsif.2013.01.016

Clougherty, J. E. (2010). A growing role for gender analysis in air pollution epidemiology. *Environmental Health Perspectives*, 118(2), 167–176. doi:10.1289/ehp.0900994 PMID:20123621

Guxens, M., Ghassabian, A., Gong, T., Garcia-Esteban, R., Porta, D., Giorgis-Allemand, L., & Sunyer, J. et al. (2016). Air Pollution Exposure during Pregnancy and Childhood Autistic Traits in Four European Population-Based Cohort Studies: The ESCAPE Project. *Environmental Health Perspectives*, 1, 133–140. doi:10.1289/ehp.1408483 PMID:26068947

Kassinis, G., Panayiotou, A., Dimou, A., & Katsifarakis, G. (2016). Gender and Environmental Sustainability: A Longitudinal Analysis. *Corporate Social Responsibility and Environmental Management*, 23(6), 399–412. doi:10.1002/csr.1386

Nelson, V., Meadows, K., Cannon, T., Morton, J., & Martin, A. (2002). Uncertain predictions, invisible impacts, and the need to mainstream gender in climate change adaptations. *Gender and Development*, 10(2), 51–59. doi:10.1080/13552070215911

O'Connor, P. (2014). *Management and gender in higher education*. Manchester University Press. doi:10.7228/manchester/9780719083587.001.0001

Pope, D., Diaz, E., Smith-Sivertsen, T., Lie, R. T., Bakke, P., Balme, J. R., & Bruce, N. G. et al. (2015). Exposure to Household Air Pollution from Wood Combustion and Association with Respiratory Symptoms and Lung Function in Nonsmoking Women: Results from the RESPIRE Trial, Guatemala. *Environmental Health Perspectives*, 123(4), 285–292. PMID:25398189

Wolf, K., Stafoggia, M., Cesaroni, G., Andersen, Z. J., Beelen, R., Galassi, C., & Forastiere, F. et al. (2015). Long-term Exposure to Particulate Matter Constituents and the Incidence of Coronary Events in 11 European Cohorts. *Epidemiology (Cambridge, Mass.)*, 26(4), 565–574. doi:10.1097/EDE.0000000000000300 PMID:25978793

KEY TERMS AND DEFINITIONS

Air Pollution: The increased concentration of pollutants in the atmosphere with negative impacts in human health and the natural environment, which can be generated by both natural causes and human activities. The increased imbalance of harmful gases in the atmosphere can trigger the global warming.

Climate Change: The change of the long-term average weather conditions and broadly refers to changes in our planet, such as the rise of the sea levels, the shrinkage of mountain glaciers, the ice melt in the Arctic.

Gender Mainstreaming: The inclusion of the gender aspect in structure, policy and strategic planning of organisations, authorities and stakeholders, at all levels.

Global Warming: The rise of the average global temperature mainly compared to the pre-industrial era.

Energy Poverty: The limited access to energy services that is vital for society well-being and sustainable development.

Environmental Management: The management of the use and conservation of the environment, the natural resources, the protection of habitats and species and the control of hazards.

Women Empowerment: The reinforcement of the social, economic, political and legal involvement of women in the workplace, marketplace and community.

Gender Equality: The equal treatment between genders in legislation, in labour market, earnings, education, decision making.

G

How Exclusive Work Climates Create Barriers for Women in IS&T

Katelyn R. Reynoldson

Old Dominion University, USA

Debra A. Major

Old Dominion University, USA

INTRODUCTION

Organizational climate consists of employees' shared perceptions of rewarded, supported and expected behaviors as well as perceptions of organizational events, practices, policies, and procedures (Schneider, Wheeler, & Cox, 1992). Climate can have significant effects on employee job satisfaction, performance, work attitudes, and motivation (Parker et al., 2003). When an organization's climate is *inclusive*, organizational members feel equally welcomed, respected and valued. Moreover, employees may experience increases in job satisfaction and organizational and career commitment (Major & Morganson, 2009). *Exclusive climates* are linked with turnover, reduced organizational commitment, decreased job satisfaction (Major, Davis, Sanchez-Hucles, Germano, & Mann, 2006) and feelings of anxiety and emotional stress (Brimhall, Lizano, & Barak, 2014).

Unfortunately, women in information sciences and technology (IS&T) frequently experience exclusive work climates (Walton, Logel, Peach, Spencer, & Zanna, 2015). This is due, in part, to the underrepresentation of women in the field. Relative to men, fewer women enter higher education with a science, technology, engineering, or mathematics (STEM) major, while even fewer pursue STEM careers (Hughes, 2014). Currently, women comprise 26 percent of those employed in computer and mathematical occupations and only 27 percent of those employed as computer and information systems managers (United States

Department of Labor, 2014). This article details some of the barriers women encounter in an exclusive IS&T work climate, including stereotypes, stereotype threat, the motherhood penalty, work-family conflict, and lack of mentoring.

BACKGROUND

While women's participation rates in male dominated fields, such as medicine, law, or accounting, are increasing, women remain underrepresented in IS&T (Diekmann, Clark, Johnston, Brown, & Steinberg, 2011; Walton et al., 2015). This shortage of women is not unique to the United States; in fact, many countries around the world encounter a similar dearth (Jackson, Hillard, & Schneider, 2014). Furthermore, women tend to leave STEM fields for other professions at high rates (Glass, Sassler, Levitte, & Michelmore, 2013).

Women are significantly underrepresented in higher-level managerial roles in the industry (Servon & Visser, 2011; Wentling & Thomas, 2009), stemming from a '*glass ceiling*,' which women encounter when they have greater difficulty climbing the corporate ladder due to subtle barriers that are not as prevalent for men (Adya, 2008; Major & Morganson, 2009). Other barriers women may experience arise from the nature of IS&T work (i.e., long hours, travel, on-call status; Ahuja, 2002; Armstrong, Riemenschneider, Allen, & Reid, 2007) and the IST&T work climate (i.e., male dominated, exclusive; Wentling &

Thomas, 2009). The IS&T climate can be “chilly” for women as they can feel out of place and less compatible with the stereotypically masculine environment of the field (Blickenstaff, 2005). As part of the male dominated climate in IS&T, women may experience exclusion from informal networks within the field, which negatively affects work opportunities, such as job placement and advancement (Adya, 2008) and self-confidence (Wentling & Thomas, 2009).

STEREOTYPES

Gender stereotypes are characterized as either descriptive (i.e., beliefs about characteristics men and women *do* hold) or prescriptive (i.e., beliefs about characteristics men and women *should* hold; Heilman, 2012; Koch, D’Mello, & Schneider, 2014). Both types can incite discrimination against women in IS&T, although in different ways. *Descriptive gender stereotypes* may lead to discrimination against women in IS&T fields because of a perceived lack of fit between typically feminine characteristics that women do hold and the characteristics of a stereotypically masculine field. *Prescriptive gender stereotypes* may lead to discrimination against women in IS&T because working in a traditionally male field violates traditional feminine gender norms. Because of a perceived lack of fit, due in part to prescriptive gender stereotypes, women may be viewed as less competent than men in male dominated fields (Amanatullah & Tinsley, 2013; Ceci & Williams, 2007). When women assert their competence by exhibiting agentic characteristics, however, this may result in *backlash*, negative reactions elicited when gender norms are perceived to be violated (Amanatullah & Tinsley, 2013). If competency is explicitly established, women may still be seen as less likeable and more hostile than their male counterparts (Amanatullah & Tinsley, 2013; Ceci & Williams, 2007). Whereas women’s competence tends to be questioned, men’s competency is often taken for granted (Heilman, 2012; Koch, et al., 2014).

The IS&T field itself can serve as a barrier to women because of its social construction as a masculine environment, often characterized as anti-social, individualistic, and competitive (Wentling & Thomas, 2009). To the extent that women are stereotypically viewed as more communal (e.g., concerned for others, deferential, emotionally sensitive; Heilman, 2012), and IS&T is a masculine environment that seemingly requires agentic traits (e.g., achievement-orientation, autonomy, analytical; Heilman, 2012) to succeed, women are likely to be perceived as a ‘poor fit’. This supposed incompatibility often results in lowered expectations for women’s success as well as perceptions of non-conforming women as deviant and less likeable (Koch et al., 2014).

For women that do value communal goals, stereotypes that IS&T is adverse to collaboration and teamwork can lead to disinterest. Diekman et al. (2011) found that female interest in IS&T was bolstered when the field was portrayed as conducive to communal goals. In addition, the IS&T workspace can be off-putting for women when it emits cues of incompatibility with the field. For example, the presence of stereotypical computer science objects (e.g., Star Trek memorabilia, comics) in an IS&T workplace can create signals that reinforce women’s feelings of exclusion, whereas neutral decorations (e.g., nature posters, art) can help foster feelings of inclusion (Cheryan, Plaut, Davies, & Steele, 2009).

STEREOTYPE THREAT

Stereotypes are learned from a young age and are so ingrained in us that they can unconsciously affect an individual’s perceptions of others (Rudman & Phelan, 2015). Stereotypes can be positive (e.g., Asian individuals are better at math) or negative (e.g., women perform worse than men in mathematics; Ambady, Shih, Kim, & Pittinsky, 2001; Franceschini, Galli, Chiesi, & Primi, 2014). When a group is stereotyped to perform negatively in a certain situation, that group often exhibits less than optimal performance due to concerns of confirm-

ing said stereotype. This effect, known as *stereotype threat*, impairs performance because some of the cognitive resources used for performance are instead preoccupied with stress, monitoring performance, and regulating negative thoughts and behaviors (Appel, Kronberger, & Aronson, 2011). For example, women knowledgeable of the stereotype that men outperform women in math may succumb to stereotype threat and underperform on a given math test. Stereotype threat sometimes leads to disidentification from the stereotyped domain (Appel & Kronberger, 2012) suggesting that negative stereotypes relating to IS&T may lead women to leave the field entirely (Streets, Major, & Morganson, 2015). Stereotype threat can be activated explicitly (i.e., mentioning the stereotype before the test) or implicitly (i.e., ask the test taker to record their gender before the test).

Stereotype threat is a concern for women in IS&T because they are stereotypically viewed as less competent in the field, relative to men (Appel & Kronberger, 2012; Appel et al., 2011; Logel, Walton, Spencer, Iserman, von Hippel, & Bell, 2009). However, the presence of stereotypes does not affect all women the same way. Franceschini et al. (2014) found that women were more vulnerable to stereotype threat if they strongly associated men with mathematics. Subtle sexism can also elicit stereotype threat in women. Logel et al. (2009) suggested that people use interpersonal interactions as a way to determine how their social identities are viewed in a given context. In male dominated environments, women may pick up subtle hints that they are at risk of being seen through a negative gender stereotype, thus activating stereotype threat.

MOTHERHOOD PENALTY

Women are underrepresented in STEM but mothers are even more rare (Glass et al., 2013). Family obligations combined with maintaining a career in STEM can prove to be somewhat problematic for women. In fact, getting married while working

in STEM is negatively associated with retention, unless the significant other also holds a STEM position (Glass et al., 2013). Mothers in the workplace may experience various negative outcomes such as lowered perceptions of competence and commitment, workplace incivility, slower career advancement, and lower wages and is known as the motherhood penalty (Correll, Benard, & Paik, 2007). Correll et al. (2007) suggested that this penalty arises from perceived incompatibility between what cultural norms dictate as a mother's role and the ideal worker's role. Fathers do not experience the same perceptions of incompatibility between being a father and a worker (Correll et al., 2007; Heilman & Okimoto, 2008).

After becoming a mother, women are viewed as less competent but warmer, however, warmth is not a valued trait in the traditional IS&T workplace, and thus no benefits are reaped from this gain. On the contrary, men maintain high levels of perceived competence and are viewed as warmer when they become fathers (Cuddy, Fiske, & Glick, 2004). Moreover, mothers are rated as less committed to the workplace than childless women (Armstrong et al., 2007; Roldan, Soe, & Yakura, 2004) whereas fathers are viewed as more committed to paid work, relative to childless men. This is likely because mothers are still expected to take care of children and family matters more than men are (Ayre, Mills, & Gill, 2014; Berdahl & Moon, 2013; Ceci, Williams, & Barnett, 2009). Given that women are already viewed as less competent than men in the IS&T field, additional decreases in perceived competency as mothers can be especially harmful.

Miner, Pesonen, Smittick, Seigel, and Clark (2014) found that mothers experience *workplace incivility*, or low-intensity workplace behavior with ambiguous intent to be rude or discourteous that violates workplace norms of respect for others (i.e., condescending comments, exclusion from professional camaraderie). Specifically, they found that mothers of three or more children experienced more workplace incivility, relative to those with less than three children and childless

mothers. Miner et al. (2014) found that experiencing incivility was associated with decreased job satisfaction, increased turnover intentions, and increased depressive symptoms. However, Miner et al. (2014) found that childless women were more strongly affected by these negative outcomes than mothers, suggesting that children can actually help to mitigate the effects associated with incivility. This is especially germane to mothers in IS&T as Miner et al. (2014) assessed incivility experienced within a male dominated field.

Career advancement is also an issue for mothers in IS&T. Heilman and Okimoto (2008) found that when applying for the same high-level management position, mothers were less likely to be considered for advancement, relative to other male and female applicants. Cuddy et al. (2004) found that mothers are less likely to be hired, promoted, and educated further, relative to childless women. Correll et al. (2007) submitted real application materials to over 600 jobs, identifying equally qualified male and female applicants as a parent or childless. They found that childless women were twice as likely to get a call back, relative to mothers, while fathers experienced no such disadvantage. Even when mothers are hired, there is a larger wage gap between childless women and mothers, relative to the gap between men and women, with mothers earning the least (Crittendon, 2002). Childless women climb the corporate ladder at slower rates than men (Ceci & Williams, 2007) and it seems as though mothers tend to fall in their wake.

WORK-FAMILY CONFLICT

The nature of IS&T work, which is characterized by long hours, travel, on-call status, etc., can lead to work-family conflict for both mothers and fathers. Amstad and colleagues (2011) found work-family conflict to be negatively correlated with well-being (e.g., general stress, work-related stress, health problems) and job performance. Moreover,

work-family conflict can result in turnover (Ayre et al., 2014; Servon & Vissor, 2011). Given that women are often expected to be caretakers and family focused (Ayre et al., 2014; Berdahl & Moon, 2013), women may experience more conflict than men. Armstrong et al. (2007) suggest that women experience an increase in work-family conflict due to extreme job requirements, such as travel, which manifests as increased stress. Furthermore, some mothers may experience loss of respect and professional credibility if they are unable to uphold expectations of working long hours (Ayre et al., 2014).

Family-friendly provisions (i.e., flexitime, part-time employment, paid maternity leave) can prove beneficial for parents; however, utilization sometimes results in penalties such as reductions in status (Ayre et al., 2014) or perceptions of decreased job commitment (Glass et al., 2013). Mothers that accept part-time work may be penalized by not being considered for key roles, challenging work, and promotions (Ayre et al., 2014). However, there are signs that the landscape may be beginning to change for women in IS&T. Recently, Armstrong and Riemenschneider (2014) found that work-family conflict is becoming less of a barrier to the persistence of women in information technology (IT). They suggest this change may stem from the increasing availability of telecommuting as well as changing gender roles, such as fathers accepting more familial responsibilities.

LACK OF MENTORING

Mentoring occurs when an inexperienced employee receives professional and personal guidance from a seasoned employee to assist in advancing the protégé's career (Ghosh & Reio, 2013). It can prove beneficial for both parties through increased job satisfaction, organizational commitment, job performance, and career success for mentors (Ghosh & Reio, 2013) as well as increased job satisfaction (Underhill, 2006), job performance,

career advancement (Ghosh & Reio, 2013), and increased socialization (Adya, 2008) for protégés. Mentoring is prevalent in two forms: formal and informal. Formal mentoring occurs when the organization pairs a mentor with a protégé while informal mentoring arises when a mentor and protégé meet and naturally form a mentoring relationship. Research suggests that the latter process typically yields better outcomes (Eby, Allen, Evans, Ng, & DuBois, 2008; Underhill, 2006).

Although women in the IT profession value mentoring (Quesenberry & Trauth, 2012), they often lack mentoring opportunities (Adya, 2008; Wentling & Thomas, 2009). One reason for this is a limited number of women in the field to act as mentors, formally or informally (Diekmann et al., 2011; Walton et al., 2015), especially higher up the corporate ladder (Armstrong & Riemenschneider, 2014; Bilmoria, Lord, & Marinelli, 2014). Further, mentors tend to have a preference for same-sex protégés (Ahuja, 2002), which disadvantages women and makes it especially unlikely that they will develop informal mentoring relationships in IS&T. Major and Morganson (2009) suggest that both men and women may be reluctant to engage in a male-mentor/female-protégé relationship because, compared to same-sex mentoring pairs, there is a heightened probability that sexual insinuations will be made about the nature of the relationship.

FUTURE RESEARCH DIRECTIONS

Substantial research has documented the features of exclusive IS&T work climates and the disadvantages associated with them, especially for women. Future research should focus on discerning effective strategies for building inclusive IS&T climate. It is important to recognize that the aspects of the IS&T work climate that are most important and impactful may differ by gender. For instance, in a recent study Major, Morganson, and Bolen (2013) found that for women in

IS&T, family-friendly work practices and policies that emphasize opportunities for growth and professional development are most important in creating a sense of commitment to the IT field. However for men, experienced stress on the job had the greatest impact on commitment to the field. Additional research is needed to develop a better understanding of the most fruitful avenues for developing inclusive IS&T work climates across gender.

Future research is also needed to explore interventions for confronting the obstacles faced by women in exclusive climates. The most effective strategies are likely to be two-pronged in that they both equip women to deal with the obstacles they encounter and actively work to eliminate obstacles from the work environment. For instance, training to inoculate and prepare women for the stereotypes they may encounter could be coupled with diversity training aimed at confronting and eliminating such stereotypes. The ultimate goal is to create an inclusive climate that facilitates the satisfaction and performance of all members of the IS&T workplace.

CONCLUSION

Shapiro and Sax (2011) assert that the underrepresentation of women in STEM has economic consequences for women as well as our country. Building an inclusive IS&T climate (i.e., reducing stereotypical appearance of computer science work environment, implementing family-friendly policies) is a vital way to attract and retain women interested in the field to avoid said economic consequences. Building inclusion in the workplace is important given the benefits associated with inclusive work climates, such as increased job satisfaction and organizational and job commitment (Major & Morganson, 2009) and the positive outcomes of diversity, such as increased job satisfaction, feelings of inclusion, and creativity (Brimhall et al., 2014).

REFERENCES

- Adya, M. P. (2008). Women at work: Differences in IT career experiences and perceptions between South Asian and American women. *Human Resource Management, 47*(3), 601–635. doi:10.1002/hrm.20234
- Ahuja, M. K. (2002). Women in the information technology profession: A literature review, synthesis and research agenda. *European Journal of Information Systems, 11*(1), 20–34. doi:10.1057/palgrave.ejis.3000417
- Amanatullah, E. T., & Tinsley, C. H. (2013). Punishing female negotiators for asserting too much... or not enough: Exploring why advocacy moderates backlash against assertive female negotiators. *Organizational Behavior and Human Decision Processes, 120*(1), 110–122. doi:10.1016/j.obhdp.2012.03.006
- Ambady, N., Shih, M., Kim, A., & Pittinsky, T. L. (2001). Stereotype susceptibility in children: Effects of identity activation on quantitative performance. *Psychological Science, 12*(5), 385–390. doi:10.1111/1467-9280.00371 PMID:11554671
- Amstad, F. T., Meier, L. L., Fasel, U., Elfering, A., & Semmer, N. K. (2011). A meta-analysis of work–family conflict and various outcomes with a special emphasis on cross-domain versus matching-domain relations. *Journal of Occupational Health Psychology, 16*(2), 151–169. doi:10.1037/a0022170 PMID:21280939
- Appel, M., & Kronberger, N. (2012). Stereotypes and the achievement gap: Stereotype threat prior to test taking. *Educational Psychology Review, 24*(4), 609–635. doi:10.1007/s10648-012-9200-4
- Appel, M., Kronberger, N., & Aronson, J. (2011). Stereotype threat impairs ability building: Effects on test preparation among women in science and technology. *European Journal of Social Psychology, 41*(7), 904–913. doi:10.1002/ejsp.835
- Armstrong, D. J., & Riemenschneider, C. K. (2014, May). The barriers facing women in the information technology profession: An exploratory investigation of Ahuja's model. In *Proceedings of the 52nd ACM conference on Computers and people research* (pp. 85–96). ACM. doi:10.1145/2599990.2600006
- Armstrong, D. J., Riemenschneider, C. K., Allen, M. W., & Reid, M. F. (2007). Advancement, voluntary turnover and women in IT: A cognitive study of work–family conflict. *Information & Management, 44*(2), 142–153. doi:10.1016/j.im.2006.11.005
- Ayre, M., Mills, J., & Gill, J. (2014). Family issues for women engineers. In M. Ayre & L. Lord (Eds.), *Women in STEM careers: International perspectives on increasing workforce participation, advancement and leadership* (pp. 79–100). Northampton, MA: Edward Elgar Publishing, Inc.
- Berdahl, J. L., & Moon, S. H. (2013). Workplace mistreatment of middle class workers based on sex, parenthood, and caregiving. *The Journal of Social Issues, 69*(2), 341–366. doi:10.1111/josi.12018
- Bilmoria, D., Lord, L., & Marinelli, M. (2014). An introduction to women in STEM careers: International perspectives on increasing workforce participation, advancement and leadership. In M. Ayre & L. Lord (Eds.), *Women in STEM careers: International perspectives on increasing workforce participation, advancement and leadership* (pp. 3–15). Northampton, MA: Edward Elgar Publishing, Inc. doi:10.4337/9781781954072.00009
- Blickenstaff, J. C. (2005). Women and science careers: Leaky pipeline or gender filter? *Gender and Education, 17*(4), 369–386. doi:10.1080/09540250500145072
- Brimhall, K. C., Lizano, E. L., & Barak, M. E. M. (2014). The mediating role of inclusion: A longitudinal study of the effects of leader–member exchange and diversity climate on job satisfaction and intention to leave among child welfare workers. *Children and Youth Services Review, 40*, 79–88. doi:10.1016/j.childyouth.2014.03.003

- Ceci, S. J., & Williams, W. M. (2007). *Why aren't more women in science?: Top researchers debate the answer*. Washington, DC: American Psychological Association. doi:10.1037/11546-000
- Ceci, S. J., Williams, W. M., & Barnett, S. M. (2009). Womens underrepresentation in science: Sociocultural and biological considerations. *Psychological Bulletin*, 135(2), 218–261. doi:10.1037/a0014412 PMID:19254079
- Cheryan, S., Plaut, V. C., Davies, P. G., & Steele, C. M. (2009). Ambient belonging: How stereotypical cues impact gender participation in computer science. *Journal of Personality and Social Psychology*, 97(6), 1045–1060. doi:10.1037/a0016239 PMID:19968418
- Correll, S. J., Benard, S., & Paik, I. (2007). Getting a job: Is there a motherhood penalty? *American Journal of Sociology*, 112(5), 1297–1339. doi:10.1086/511799
- Crittenden, A. (2002). *The price of motherhood: Why the most important job in the world is still the least valued*. Macmillan.
- Cuddy, A. J., Fiske, S. T., & Glick, P. (2004). When professionals become mothers, warmth doesnt cut the ice. *The Journal of Social Issues*, 60(4), 701–718. doi:10.1111/j.0022-4537.2004.00381.x
- Diekmann, A. B., Clark, E. K., Johnston, A. M., Brown, E. R., & Steinberg, M. (2011). Malleability in communal goals and beliefs influences attraction to stem careers: Evidence for a goal congruity perspective. *Journal of Personality and Social Psychology*, 101(5), 902–918. doi:10.1037/a0025199 PMID:21859224
- Eby, L. T., Allen, T. D., Evans, S. C., Ng, T., & DuBois, D. L. (2008). Does mentoring matter? A multidisciplinary meta-analysis comparing mentored and non-mentored individuals. *Journal of Vocational Behavior*, 72(2), 254–267. doi:10.1016/j.jvb.2007.04.005 PMID:19343074
- Franceschini, G., Galli, S., Chiesi, F., & Primi, C. (2014). Implicit gender–math stereotype and womens susceptibility to stereotype threat and stereotype lift. *Learning and Individual Differences*, 32, 273–277. doi:10.1016/j.lindif.2014.03.020
- Ghosh, R., & Reio, T. G. Jr. (2013). Career benefits associated with mentoring for mentors: A meta-analysis. *Journal of Vocational Behavior*, 83(1), 106–116. doi:10.1016/j.jvb.2013.03.011
- Glass, J. L., Sassler, S., Levitte, Y., & Michelmore, K. M. (2013). Whats so special about STEM? A comparison of womens retention in STEM and professional occupations. *Social Forces*, 92(2), 723–756. doi:10.1093/sf/sot092 PMID:25554713
- Heilman, M. E. (2012). Gender stereotypes and workplace bias. *Research in Organizational Behavior*, 32, 113–135. doi:10.1016/j.riob.2012.11.003
- Heilman, M. E., & Okimoto, T. G. (2008). Motherhood: A potential source of bias in employment decisions. *The Journal of Applied Psychology*, 93(1), 189–198. doi:10.1037/0021-9010.93.1.189 PMID:18211144
- Hughes, R. (2014). The evolution of the chilly climate for women in science. In B. Irby, B. Polnick, & J. Koch (Eds.), *Girls and women in STEM: A never ending story* (pp. 71–92). Charlotte, NC: Information Age Publishing.
- Jackson, S. M., Hillard, A. L., & Schneider, T. R. (2014). Using implicit bias training to improve attitudes toward women in STEM. *Social Psychology of Education*, 17(3), 419–438. doi:10.1007/s11218-014-9259-5
- Koch, A. J., D'Mello, S. D., & Sackett, P. R. (2014). A meta-analysis of gender stereotypes and bias in experimental simulations of employment decision making. *The Journal of Applied Psychology*, 100(1), 128–161. doi:10.1037/a0036734 PMID:24865576

- Logel, C., Walton, G. M., Spencer, S. J., Iserman, E. C., von Hippel, W., & Bell, A. E. (2009). Interacting with sexist men triggers social identity threat among female engineers. [Correction]. *Journal of Personality and Social Psychology*, 96(6), 1089–1103. doi:10.1037/a0015703 PMID:19469589
- Major, D. A., Davis, D. D., Sanchez-Hucles, J., Germano, L. M., & Mann, J. (2006). IT workplace climate for opportunity and inclusion. *Encyclopedia of Gender and Information Technology*, 2, 856-862.
- Major, D. A., & Morganson, V. L. (2009). An inclusive IS&T work climate. In M. Khosrow-Pour (Ed.), *Encyclopedia of Information Science and Technology* (2nd ed., pp. 1899–1905). Hershey, PA: Information Science Reference; doi:10.4018/978-1-60566-026-4.ch299
- Major, D. A., Morganson, V. J., & Bolen, H. M. (2013). Predictors of occupational and organizational commitment in information technology: Exploring gender differences and similarities. *Journal of Business and Psychology*, 28(3), 301–314. doi:10.1007/s10869-012-9282-5
- Miner, K. N., Pesonen, A. D., Smittick, A. L., Seigel, M. L., & Clark, E. K. (2014). Does being a mom help or hurt? Workplace incivility as a function of motherhood status. *Journal of Occupational Health Psychology*, 19(1), 60–73. doi:10.1037/a0034936 PMID:24447221
- Parker, C. P., Baltes, B. B., Young, S. A., Huff, J. W., Altmann, R. A., Lacost, H. A., & Roberts, J. E. (2003). Relationships between psychological climate perceptions and work outcomes: A meta-analytic review. *Journal of Organizational Behavior*, 24(4), 389–416. doi:10.1002/job.198
- Quesenberry, J. L., & Trauth, E. M. (2012). The (dis) placement of women in the IT workforce: An investigation of individual career values and organisational interventions. *Information Systems Journal*, 22(6), 457–473. doi:10.1111/j.1365-2575.2012.00416.x
- Roldan, M., Soe, L., & Yakura, E. K. (2004, April). Perceptions of chilly IT organizational contexts and their effect on the retention and promotion of women in IT. In *Proceedings of the 2004 SIGMIS conference on Computer personnel research: Careers, culture, and ethics in a networked environment* (pp. 108-113). ACM. doi:10.1145/982372.982399
- Rudman, L. A., & Phelan, J. E. (2015). The effect of priming gender roles on women's implicit gender beliefs and career aspirations. *Social Psychology*, 41(3), 192–202. doi:10.1027/1864-9335/a000027
- Schneider, B., Wheeler, J. K., & Cox, J. F. (1992). A passion for service: Using content analysis to explicate service climate themes. *The Journal of Applied Psychology*, 77(5), 705–716. doi:10.1037/0021-9010.77.5.705
- Servon, L. J., & Visser, M. A. (2011). Progress hindered: The retention and advancement of women in science, engineering and technology careers. *Human Resource Management Journal*, 21(3), 272–284. doi:10.1111/j.1748-8583.2010.00152.x
- Shapiro, C. A., & Sax, L. J. (2011). Major selection and persistence for women in STEM. *New Directions for Institutional Research*, 152(152), 5–18. doi:10.1002/ir.404
- Streets, V. N., Major, D. A., & Morganson, V. J. (2015). Building inclusive IS&T work climates for women and men. In M. Khosrow-Pour (Ed.), *Encyclopedia of Information Science and Technology* (3rd ed.; pp. 753-761). Hershey, PA: Information Science Reference. doi:10.4018/978-1-4666-5888-2.ch072
- Underhill, C. M. (2006). The effectiveness of mentoring programs in corporate settings: A meta-analytical review of the literature. *Journal of Vocational Behavior*, 68(2), 292–307. doi:10.1016/j.jvb.2005.05.003
- United States Department of Labor. (2014). *Computer and Information Technology Occupations*. Retrieved from: http://www.dol.gov/wb/stats/Computer_information_technology_2014.htm

Walton, G. M., Logel, C., Peach, J. M., Spencer, S. J., & Zanna, M. P. (2015). Two brief interventions to mitigate a chilly climate transform women's experience, relationships, and achievement in engineering. *Journal of Educational Psychology*, 107(2), 468–485. doi:10.1037/a0037461

Wentling, R. M., & Thomas, S. (2009). Workplace culture that hinders and assists the career development of women in information technology. *Information Technology, Learning and Performance Journal*, 25, 25–42.

ADDITIONAL READING

Bidwell, A. (2015). STEM gender stereotypes common across the world. *U.S. News & World Report*. Retrieved from <http://www.usnews.com/news/blogs/data-mine/2015/05/18/stem-gender-stereotypes-common-across-the-world>

Boysen, G. A. (2013). Confronting math stereotypes in the classroom: Its effect on female college students' sexism and perceptions of confronters. *Sex Roles*, 69(5-6), 297–307. doi:10.1007/s11199-013-0287-y

Burgess, D., & Borgida, E. (1999). Who women are, who women should be: Descriptive and prescriptive gender stereotyping in sex discrimination. *Psychology, Public Policy, and Law*, 5(3), 665–692. doi:10.1037/1076-8971.5.3.665

Cannady, M. A., Greenwald, E., & Harris, K. N. (2014). Problematizing the STEM pipeline metaphor: Is the STEM pipeline metaphor serving our students and the STEM workforce? *Science Education*, 98(3), 443–460. doi:10.1002/sce.21108

Ceci, S. J., Williams, W. M., Sumner, R. A., & DeFraigne, W. C. (2011). Do subtle cues about belongingness constrain women's career choices? *Psychological Inquiry*, 22(4), 255–258. doi:10.1080/1047840X.2011.619112 PMID:23136463

Charleston, L., Adserias, R., Lang, N., & Jackson, J. (2014). Intersectionality and STEM: The role of race and gender in the academic pursuits of African American women in STEM. *Journal of Progressive Policy & Practice*, 2, 273–293.

Dasgupta, N. (2011). Ingroup experts and peers as social vaccines who inoculate the self-concept: The stereotype inoculation model. *Psychological Inquiry*, 22(4), 231–246. doi:10.1080/1047840X.2011.607313

Davies, P. G., Spencer, S. J., Quinn, D. M., & Gerhardstein, R. (2002). Consuming images: How television commercials that elicit stereotype threat can restrain women academically and professionally. *Personality and Social Psychology Bulletin*, 28(12), 1615–1628. doi:10.1177/014616702237644

Dodson, L. (2013). Stereotyping low-wage mothers who have work and family conflicts. *The Journal of Social Issues*, 69(2), 257–278. doi:10.1111/josi.12014

El-Bahey, R., & Zeid, A. (2013, October). Women in computing: A case study about Kuwait. In *Frontiers in Education Conference*, 2013 IEEE (pp. 1871–1877). IEEE. doi:10.1109/FIE.2013.6685160

Heilman, M. E., & Okimoto, T. G. (2007). Why are women penalized for success at male tasks?: The implied communality deficit. *The Journal of Applied Psychology*, 92(1), 81–92. doi:10.1037/0021-9010.92.1.81 PMID:17227153

Holmes, J. (2005). Leadership talk: How do leaders do mentoring, and is gender relevant? *Journal of Pragmatics*, 37(11), 1779–1800. doi:10.1016/j.pragma.2005.02.013

Mamman, A., Kamoche, K., & Bakuwa, R. (2012). Diversity, organizational commitment and organizational citizenship behavior: An organizing framework. *Human Resource Management Review*, 22(4), 285–302. doi:10.1016/j.hrmr.2011.12.003

- McDonald, M. L., & Westphal, J. D. (2013). Access denied: Low mentoring of women and minority first-time directors and its negative effects on appointments to additional boards. *Academy of Management Journal*, 56(4), 1169–1198. doi:10.5465/amj.2011.0230
- Miller, D. I., & Wai, J. (2015). The bachelors to Ph. D. STEM pipeline no longer leaks more women than men: A 30-year analysis. *Frontiers in Psychology*, 6, 1–10. doi:10.3389/fpsyg.2015.00037 PMID:25741293
- Moen, P. (2015). An institutional/organizational turn getting to work–life quality and gender equality. *Work and Occupations*, 42(2), 174–182. doi:10.1177/0730888414568085
- Norris, M. (2015). STEM education: Closing the gap in leadership for women. Retrieved from: <http://sodexoinsights.com/stem-education-closing-the-gap-in-leadership-for-women/>
- Reid, M. F., Allen, M. W., Armstrong, D. J., & Riemenschneider, C. (2010). Perspectives on challenges facing women in IS: The cognitive gender gap. *European Journal of Information Systems*, 19(5), 526–539. doi:10.1057/ejis.2010.30
- Rudman, L. A., Moss-Racusin, C. A., Phelan, J. E., & Nauts, S. (2012). Status incongruity and backlash effects: Defending the gender hierarchy motivates prejudice against female leaders. *Journal of Experimental Social Psychology*, 48(1), 165–179. doi:10.1016/j.jesp.2011.10.008
- Shapiro, J. R., & Williams, A. M. (2012). The role of stereotype threats in undermining girls and women's performance and interest in STEM fields. *Sex Roles*, 66(3-4), 175–183. doi:10.1007/s11199-011-0051-0
- Spencer, S. J., Logel, C., & Davies, P. G. (2015). Stereotype threat. *Annual Review of Psychology*, 67, 14.1-14.23.
- Sturm, R. E., Taylor, S. N., Atwater, L. E., & Braddy, P. W. (2014). Leader self-awareness: An examination and implications of women's underprediction. *Journal of Organizational Behavior*, 35(5), 657–677. doi:10.1002/job.1915
- Trauth, E. M. (2002). Odd girl out: An individual differences perspective on women in the IT profession. *Information Technology & People*, 15(2), 98–118. doi:10.1108/09593840210430552
- Vanderbroeck, P. (2010). The traps that keep women from reaching the top and how to avoid them. *Journal of Management Development*, 29(9), 764–770. doi:10.1108/02621711011072478
- Walton, G. M., Murphy, M. C., & Ryan, A. M. (2015). Stereotype threat in organizations: Implications for equity and performance. *Annual Review of Organizational Psychology and Organizational Behavior*, 2(1), 523–550. doi:10.1146/annurev-orgpsych-032414-111322

KEY TERMS AND DEFINITIONS

Backlash: Negative reaction against women whose behavior violates gender norms (Amanatullah & Tinsley, 2013).

Descriptive Gender Stereotypes: Beliefs about characteristics women *do* hold (i.e., women *are* nurturing and soft-spoken; Heilman, 2012).

Exclusive Climate: When organizational members do not feel equally welcome, respected and valued.

Glass Ceiling: Subtle barriers that exist to prevent women from rising to the top of the corporate ladder in the same way that men do.

Inclusive Climate: When organizational members feel equally welcome, respected and valued.

Prescriptive Gender Stereotypes: Beliefs about characteristics women *should* hold (i.e., women *should be* nurturing and soft-spoken; Heilman, 2012).

Stereotype Threat: When a negatively stereotyped group's performance is hindered due to concern of confirming that stereotype.

Workplace Incivility: Low-intensity workplace behavior with ambiguous intent to be rude

or discourteous that violates workplace norms of respect for others (i.e., condescending comments, exclusion from professional camaraderie; Miner et al., 2014).

Women and IT in Lilongwe



Alice Violet Nyamundundu
Skyway University, Malawi

INTRODUCTION

Africa is like one family on the issue of gender. Women are deemed inferior and men are labeled superior in all arenas including education, world of work and even in health.. Both parties have internalized their positions as ‘natural’. The mentality is rooted in the African culture and socialization. Some women and men are now accepting a ‘wake up call’ but at a very slow rate. Women in Lilongwe and Africa in general are lagging behind in IT and society in general. This paper focuses on gender as in ‘women’ since they are the most disadvantaged in IT. However, gender challenges can be overcome.

BACKGROUND

When the word gender is mentioned it seems to refer to women only simply because women are the most affected between the two. According to Unesco 2004 gender differences are ‘manmade differences’ and can change and vary from region to region. Both men and women are born with a ‘clean slate’ on the issues of gender; it is society which construct them. Sex is a ‘biological difference’ (Mbiti 1997). The physical differences are interpreted differently and at most to the disadvantage of the other. One’s sex at birth will either give power or poverty. According to the African culture, contrary to the western culture, the fact that one is born a boy means intelligence, IT skills, fame, riches and high positions. On the contrary, being born a girl means inferiority, submission, poverty, lack of IT skills and second class position. There is need for a powerful mental rewiring for women to be emancipated from this hard shell.

Conversely women on their part have internalized the inferiority complex to the extent of enduring abuses, protecting the abusers and dying under the hands of their abusers. According to Haralambos et al (2004) socialization is ‘the process of learning roles, norms and values appropriate to the people of that particular society’. The process enhances the inferior gender roles of women and the superior gender roles of men. If girls display the so called ‘masculine behavior’ they are discouraged. Social negative sanctions can be applied to them, sometimes as names for example ‘sissy’ for boys and ‘tomboys’ for girls (Haralambos et al 2004). This mentality discourages women to embark on acquiring IT skills.

In Africa and Asia, it seems being born a girl is a ‘curse’. Female sex is traditionally perceived as inferior to male sex. In Chinese families with a feudal tradition, girls are ranked as low as pigs and dogs (China’s Ministry of culture 1984, Unesco 2004). Since socialization is a strong process most girls end up lacking confidence in IT, fight for their rights or to better themselves.

In families, men become tough with their wives in marriage as tradition considers tenderness ‘unmasculine’ (Soetan 2001). Women are viewed as mothers and wives and not technical experts. Women are considered ‘stay at homes’ (Fields 2004) although sometimes educated. If men try to help their wives they are ridiculed by friends, relatives or even their parents. This African perspective has brought detrimental effects to women in Lilongwe in all areas including IT, education in general, world of work, politics and in families. All this is reflected in the Malawian society where legal marriage age of girls was only thirteen years until recently in 2015 when a bill was passed to push the legal age of majority to 18

DOI: 10.4018/978-1-5225-2255-3.ch296

years. The responsible authorities are now trying to review the marriage age of girls from thirteen years to eighteen years. They take Gandhi's view that 'men can never be a woman's equal in the spirit of selfless service with which nature has endowed her'(www.google). Globally, women should be viewed as hard workers who deserve to be rewarded and women should view themselves as such too. According to Schaefer (2010) "there are no innate or biologically determined gender roles for men and women," society constructs them hence it is the same society that can correct the anomalies to achieve gender balance in IT. Women need support to showcase their IT abilities (Sadc protocol 2013). Unless there is mental rewiring on both sexes, women will remain IT illiterate, poor, abused and in low and non profitable positions both in Lilongwe and Africa in general.

LITERATURE REVIEW

A lot of research has been carried out on the issue of gender and quite a number of publications have been produced (Unesco 2004, Beijing conference report 1995, Malawi Sadc Protocol 2013, Schaefer 2010, Torrington et al 2010). However, there seems to be a deficit of literature in terms of women and computer skills in Lilongwe. Women are still discriminated on the basis of sex and not abilities. Discrimination should be based on a 'fair criteria' and a 'level playing field' (Torrington et al 2010). Men and women should compete on equal terms in IT.

Women dominate low 'paying jobs' (Carroll et al 2009). They need to step up and go up the ladder. One way is by engaging in IT skills from infancy just like men. However this is restricted by socialization and culture which view "IT as complicated and not meant for women" yet both have the same capabilities. Women are only expected to use simple IT skills like messaging and what's up and not internet browsing or any technical IT skills. To date some still view women's oppression as inevitable (Sengupta 2004).

According to sociology discipline, functionalists maintain that gender differentiation has contributed to overall stability. Talcott Parsons and Bale (1950) were of the view that women should take the "emotionally supportive role and men the instrumental role" and the two should complement each other. On the contrary the two are conflicting. Men's instrumental roles are highly rewarded in terms of "money and prestige"(Schaefer 2010). Women's emotional support is highly abused. This imbalance has led to the imbalance in IT, education in general, and world of work and in politics. Women are stereotyped and viewed as robotic persons. This negative mentality has led to very few women in Malawi to advance professionally. About 20% of MPS and 26% of Top civil servants are women in the world (Human rights commission 2007). Literature is showing that something is being done to address the issue of gender in Africa but more effort is needed less women will continue to suffer. The mental gap should be closed first in order to completely close the IT gap.

Steps Taken to Close the Gender Gap

Quite a number of activities ranging from conferences to signing of documents have been done to address the gender imbalance but as mentioned in the literature review more is needed. In 2013 Malawi Sadc Protocol Barometer to address the gender gap was launched, that is applauded. In early 2014 the president of Kenya Uhuru Kenyatta signed into law the marriage bill which states that marriage partners have equal rights and obligations at the time of their marriage. Several policies have also been put in place for example the Zimbabwe gender policy 2013-2017 which advocates for equality between women and men in leadership positions, in business, education and all other sectors. In addition a woman's university was established in Zimbabwe, Marondera Suburb second to the Sudanese University for Muslim women(<http://www.arts.uwa.edu.au/>).

In Zimbabwe again the domestic bill was signed which also protects men and women from abuse.

Equal employment act was also introduced in many countries in Africa to cater for both sexes in recruitment and employment (Unesco 2004).

FACTORS INFLUENCING GENDER IMBALANCE

Educations

Although 90% of Africa have attained primary education very few proceeded to secondary education let alone tertiary education. Among those who do not proceed to secondary or tertiary education, women have the highest rate of dropouts. According to Afro barometer 2013, only 11% of women have post secondary school. Most parents probably from the age of 45 years and above were strongly nurtured in the African culture where the girl child is not respected. This has detrimental effects; the girl child can end up not respecting her education, herself and even drop out of school. As a result it is estimated that there are about 563 million women against 313million men who are illiterate in Africa (Unesco 2000).

Teachers themselves are playing a larger role in the lack of education of girls. Most teachers both male and female encourage the 'boy child' more than the 'girl child' because they were socialized as such themselves. This brings IT reluctance in women and girls to the extent of accepting inferior positions. Teachers are highly valued by students and viewed as noble. Hence if the 'noble ones' are portraying the idea of inferiority then the progress of the girl child education is hampered. Girls have reported that the "discrimination by teachers is blatant" (UNESCO 2004). Girls are discouraged from taking math's and science because they are considered 'virile' subjects (UNESCO 1997). Automatically girls feel unqualified for the so called 'masculine subjects' namely IT. Parents should begin to believe in their children's intellectual capacities.

In countries like Swaziland being married to a king is a 'fabulous thing' to most parents and not education. Gender imbalance in education in Africa is still an eyesore to the extent that the target pegged by United Nation of reaching gender parity by 2015 could not be attained. The disparities are still large!(SADC Protocol Malawi 2013).

Childcare

Less shared child care responsibilities in Africa results in imbalances between men and women. Women are at the receiving end of disadvantages. There is no equitable distribution of labour in homes. Most women think that it is their sole right and responsibility to look after children and homes. Therefore they do not bother themselves with IT skills.

Marriage

In Africa, many women enter into forced marriages or 'encouraged marriages'. Socialization makes most girls and most women not to think of self development but to think of their husbands or marriage per se! In Ivory Coast child marriages are rampant. Girls are given into marriage to men 3 times their age (Soetan 2001). The people in the community condone the practice. To them, clan relationships are strengthened and girls are protected from an immoral life. This mentality label girls and women as inferior, immoral and without technical knowhow. In Malawi statistics show fewer girls than boys (Nation 2014) reaching secondary school. Most of them rush for marriage just after primary education. This reflects the mindset of the girl child due to socialization. They rush for marriage before acquiring necessary IT skills to run businesses. The problem of forced marriages is also seen beyond Africa. In Pakistan, women do not have the freedom to marry for love. Their rights are abused. They are supposed to marry a man chosen by the family. If they deviate they die in 'honor killing'.(Nation 2014) Women are not robotic creatures who can be controlled by other

people. They need the freewill to make decisions in all sectors of life including IT.

Health

Men and women are not at par in terms of health in Africa. Women are most disadvantaged. In Zimbabwe for example Hiv prevalence is at 7.8% high in women compared to 3.6% of men (Zimbabwe Gender Policy 2013-2017). Women sometimes are not free to make decisions about their health or family planning. In some religions, children and women are dying because they are not allowed to seek medical treatment, on the contrary their spouse and leaders stealthily and quietly seek medical treatment. Women need permission from their husbands to go to hospital since they are the once with the financial muscle. This makes the “male factor very important” (Soetan 2001).

In Nigeria there is Hausa proverb that reflects the mindset of people on women. This proverb says that a “healthy girl belongs to a husband, a sick woman belongs to her parents” (Soetan 2001) justifying the neglect of women in issues of health. This ideology is rampant across Africa and it reflects the unfair treatment of women. Weak health contribute to poor education, there is no way a woman can venture into advanced IT studies whilst having poor health. Most of the women end up dropping out of school. So the issue of gender imbalance and gender bias still needs to be addressed to a larger extent.

Role Models

Africa let alone Malawi lack role models especially in IT. Children learn fast when they have someone to imitate, the same applies to girls and women. They need someone they would look up to, a model who can motivate them to excel. In Africa such female people are scarce!

Stereotypes

African women are stereotyped by the society as people who always need male support and leader-

ship. Sex stereotyping persists in many forms and in all countries regardless of the level of development (Michel 1985). Basically women are trained in how to prepare food, how to behave towards men and how to look after children (Mbiti 1997). This constant portrayal of women and men occupying certain roles according to socially constructed gender division of labour (Zimbabwe Gender policy 2013-2017) is hampering the progress in closing the gender gap in IT studies. In the world of work, women are viewed as people who should be committed to their families and not technical jobs. This type of thinking has land women in inferior positions. They are viewed as “flowers of the world” (UNESCO 2001). To African society they can only flourish in the home and make their families happy yet women can be excellent engineers, architects, IT specialists, doctors the list is endless! As men can be excellent child minders, women can also be excellent managers and IT specialists in corporate organizations.

METHODOLOGY

For this qualitative document analysis research, I used the content analysis and the rubrics analysis approaches to identify themes and classify information from the documents. To achieve this, I consulted several documents in order to ensure trustworthiness of the findings. In addition to the document research, a 10 questions questionnaire was distributed to 20 female students at Skyway University in Lilongwe Malawi. This was done to affirm the real situation of women and IT at the moment even in higher levels of education.

FINDINGS

Both women and men do not encourage women leadership in politics. This was seen in Malawi Presidential elections 2014. Both women and men were castigating women leadership and were voting against the woman president. Women are stereotyped as poor leaders regardless of their

intellectual capacities and honest. On the contrary men are preferred on leadership regardless of their iron fists in ruling and several acts of corruption.

Socialization mostly by parents is leading to women oppression, abuse and stereotypes. Very few women proceed to secondary school for example in Malawi 50% of primary school children are girls but only 45% reach secondary school whilst only 40% reach tertiary education(SADC 2013, Gender Protocol Barometer). The reasons for drop out are certified as early marriages, arranged or forced marriages, poverty, distance to school, priority of boys over girls, socialization in general, lack of role models and stereotyping.

In Pakistan a woman was stoned to death in a so called “honor killing”(The Nation Malawi 2014) because she married a man of her choice. All this illustrating the harsh realities faced by women across the globe.

In this 21st century men are still shunning child care and view it as a ‘woman’s job’. In Africa may be only 2% of men will take child care responsibility let alone paternity leave. A man is ridiculed by peers, relatives and parents for cleaning after the child. The myth is, if a man does such things he is ‘bewitched by his wife’. Fear of ridicule and victimization prevent most men from helping in child care and domestic responsibilities. IT is considered a ‘masculine profession’ therefore most women shun it. Men need to be ‘man enough’ and take care of children and domestic responsibilities in order to relieve their wives. This will help women to find more time to advance themselves in technical professions such as IT.

Some laws in Africa are still discriminating against women leadership as seen in the Lesotho law of chieftaincy. A woman was denied the right to her father’s inheritance in chieftainship both by society and the law due to the fact that she is a woman. This intimidates other women and instill low self esteem in them so much that they confine themselves to the so called “women jobs and businesses”. There by keeping the IT gender gap wide.

According to Torrington et al (2010) women are beginning to enter some previously male dominated occupations but there remains a high degree of “subtle discrimination.” Women are still dominating low paying jobs for example care work(Carroll et al 2009) and the number of women as leaders is still pathetic!

According to the Equality and Human Rights Commission Report (2007), 20% of MPs and 26% of top civil servants are women and pay differentials between men and women have increased. The research has also discovered that there are no innate biologically determined gender roles for men and women. Both of them can be socialized to do any of the so called ‘masculine’ or ‘feminine jobs’. This can be revealed by a United States example where more women than men earned doctoral degrees (Bureau of the census 2008). Therefore women can also be managers, IT specialists, political leaders and supervisors respectively.

I also discovered that the treatment of girls across Africa and worldwide intimidate them and give them a low self esteem, for example the abduction of 200 girls in Nigeria in March 2014 by the organization called Boko Haram and the threat to sell them for prostitution instill fear in them. The girls would wonder “what the future hold for them, would they ever have the opportunity to become educated professionals”(2004).

According to United Nations(2001) women can be dismissed from the job or denied employment because they are pregnant. This treatment can be effected by other women or by men. There is stigma attached with pregnancy. Society view pregnant women as non performers.

In Malawi three quarters of girls who go to primary school do not proceed to secondary schools they drop out either to get married or to work as housemaids or start small fatcook¹(makandase) businesses which do not prosper at all. Therefore, IT becomes a night mare to them.

Below are the questions asked to 20 respondents at Skyway University and the responses given?

Table 1 shows the ages of the respondents at Skyway University. 95% of the female students

Table 1. Age

AGE	FREQUENCY	PERCENTAGE
18-24	19	95%
25-30	1	5%
31-40	-	0%
41-50	-	0%
Over 50	-	0%
Total	20	100%

asked are within the range of 18-24 years of age. These are the very people who are supposed to be more involved in IT issues since they are in a tertiary institution and that is why the researcher concentrated on this age.

Table 2 highlights the programs being undertaken by the respondents. 75% of the respondents are female students doing diplomas, whereas 25% of the students are doing Degrees. This information reveals that very few females continue with tertiary education to the upper levels that is degree level.

The research shows that 90% of the students use IT equipment for different activities. The below tables will clarify which ones and for what activities.

Table 4 shows the number of IT gadgets which are at women's disposal. As shown in the table

Table 3. Do you make use of any IT equipment?

RESPONSE	FREQUENCY	PERCENTAGE
Yes	18	90%
No	2	10%
Total	20%	100%

Table 4. Which IT equipment do you use?

EQUIPMENT	FREQUENCY	PERCENTAGE
Cell phones	18	90%
Tablets	1	5%
Laptops	3	15%
Desktops	1	5%
other	0	0%

Table 2. The programs in which the respondents are involved in

PROGRAM	FREQUENCY	PERCENTAGE
Degree	15	75%
Diploma	5	25%
Total	20	100%

4 90% have cell phones of different models and they use them to perform different tasks. 5% use tablets which is a more advanced technology than cell phones, 15% use laptops which is a much higher percentage than that of tablets. Another 5% use Desktops and no other technological gadget is used by the students except those mentioned above. This shows women are still lagging behind in technology.

Table 5 above shows that most of the girls use the social media at most and less on academic issues. The percentages are high on what's up and face book and lower on internet browsing which is shown by the 30% only. That makes one wonder whether students are aware of their technology in detail. Moreover Skyway University as an institution of higher learning, I expected to see internet browsing at a higher rate due to a lot of research needed at this level. However the reverse is true. 0% use twitter and the researcher assume that this is due to lack of IT skills.

Table 6 shows where women put much of their concentration, that is on what's up, where most gossips take place and issues of less importance take place. This fulfills what is mentioned by Bray (2007) that "men actively engage with machines,

Table 5. What do you mainly access when using these equipments?

FACILITIES	FREQUENCY	PERCENTAGE
Face book	7	35%
What's up	13	65%
Twitter	0	0%
Internet Browsing	6	30%
Other	0	0%

Table 6. how often do you access the previous mentioned facilities?

TIME PERIOD	FREQUENCY	PERCENTAGE	FACILITIES
Once a week	1	5%	Internet browsing
Everyday	14	70%	what's up
Twice a week	3	15%	Face book
Once a month	2	10%	Face book
Very rarely	3	15%	Internet browsing

making, using and loving them, women may have to use machines in the work place or in the home but they neither love nor seek to understand them. However this research is there to encourage women and girls in engaging in serious IT issues which allows them to research widely and get acquainted with the global world.

According to Table 7, three quarters of the student's skills ranges from good to excellent which shows that women are capable of venturing into IT issues even becoming IT experts. However, they do not give themselves time to get involved with IT equipment since they consider it a masculine profession.

Table 8 reveals that most students at tertiary level have received training of some sought and only a few do not manage to get the training. This shows that women and girls need to be assisted a lot in changing their mindset about IT. They should refrain from thinking that is a man's area.

Table 9 show that most of the girls did not receive formal training. 55% of the girls learned

Table 8. Did you receive any training?

RESPONSE	FREQUENCY	PERCENTAGE
Yes	17	85%
No	3	15%
Total	20	100%

Table 7. What is the level of your skills?

LEVEL OF SKILLS	FREQUENCY	PERCENTAGE
Poor	2	10%
Good	3	15%
Better	13	65%
Excellent	2	10%

IT skills from friends, 40% from college and only 10% from home. This leaves a lot to be desired. Not getting formal training can paralyze the performance of girls in IT. They lose self confidence when they face challenging tasks and it leads to low self esteem. This also assist in explaining why very few girls are engaged in internet browsing. Therefore their research will not be comprehensive due to lack of IT skills.

Table 10 shows that most women and girls view culture as their drawback. 80% agree that culture is blocking their way to technology. Most households care less on girls' technical skills than boy's technical skills. Their belief is that girls will get married and will be supported by their husbands therefore to them there is no need to equip them with IT skills. Some of the girls also do not make an effort when they get the chance

Table 9. Where did you get the training from?

PLACE	FREQUENCY	PERCENTAGE
Friends	11	55%
Office	-	0%
Home	2	10%
College	8	40%

Table 10. Do you believe that culture contributed to some of the challenges faced by women in IT?

RESPONSE	FREQUENCY	PERCENTAGE
Yes	16	80%
No	4	20%
Total	20	100%

due to socialization; they consider IT as men's job. Although gender empowerment is encouraged in the contemporary world, most households in Malawi are still clinging to the traditional culture which looks down upon women. This leads to many girls becoming IT illiterate. Three quarters of the population in Malawi stays in the villages and only a quarter in urban areas. This alone hampers progress of girls in acquiring IT skills.

CONCLUSION

By and large, women and gender in Africa is a cause for concern. The discussion has shown that society is contributing to a large extent to the gender gap in IT. Due to socialization girls grow with an inferiority complex and boys with a superiority complex. Being deemed inferior by society does not take away women's intelligence. Women can also be IT specialists, managers, engineers, doctors, pilots, architects and the like. Conversely, the women are internalizing the oppression and suppression due to socialization. They accept inferior positions and do not aspire for IT positions.

FUTURE RESEARCH DIRECTIONS

Gender awareness programs should be showcased to community through dramas, meetings and posters. Gender education can be delivered to parents and children in the community through media and gender educators. Media should play a vital role in applauding women's talents and their capabilities than stereotyping them. Parents should desist from socializing their children differently whether with their language, the things they buy for them, the schools they sent them to, the skills they train them into. More policies against sex, gender discrimination should be put in place.

Government of Malawi is also encouraged to make it possible for the girls both in villages

and urban areas to take up IT subjects in schools especially from primary schools. They can seek donor funding to provide electricity and IT gadgets in schools or even starting resource centers to help both boys and girls to learn IT from infancy. If this happens then it will be possible for Norris (2016) suggestion that "colleges should be equipped with innovative teaching pedagogy and new frameworks for quality online programs" to be designed and used by colleges and universities in Malawi.

Africa should organize its own conferences on gender and set its goals to close the gender gap. A lot of publicity on the advantages of gender equality should be done. Awareness should be at societal disposal showcasing why it is at its best interests to empower women and girls.

The curriculum in schools from infants should include gender and IT lessons. Schools should be encouraged to conduct gender workshops to help the girl child to have self esteem and aspire high. Gender sensitive curriculum should be put in place.

REFERENCES

- Bray, F. (2007). *Gender and Technology*. Massachusetts Institute of Technology.
- Carroll, M., Smith, M., Oliver, G., & Sung, S. (2009). Recruitment and Retention in frontline services: The case of child care. *Human Resource Journal*, 19(1), 59–74. doi:10.1111/j.1748-8583.2008.00076.x
- Fleenor, H. G., & Hodhod, R. (2016). *Assessment of learning and Technology: Computer Science education*. Columbus State University. doi:10.4018/978-1-4666-9577-1.ch003
- Haralambos, M., & Holborn, M. (2004). *Sociology themes and Perspectives*. Collins.
- Hochschild, A. R. (1990). the second shift: Employed women are putting in another day of work home. *The Utne Reader*, 66–73.

Hochschild, A. R. (2005). *Commercialization of ultimate life: Notes from home and work*. Berkeley, CA: University of California Press.

Kool, A. P. (2001). *Literacy from empowering women farmers*. Paris: Unesco.

Mbiti, J. S. (1997). *African Religions and Philosophy*. London: Heinemann.

Norris, S. E. (2016). *Designing online MBA Programs to promote Transformative Learning and knowledge creation through project based learning using the job characteristic model*. Spring Abor University. doi:10.4018/978-1-4666-9577-1.ch001

Osili, U., & Long, B. (2008). Does female schooling reduce fertility? Evidence from Nigeria. *Journal of Development Economics*, 87(1), 57–75. doi:10.1016/j.jdeveco.2007.10.003

Patton, M. O. (2002). *Qualitative research and evaluation methods*. London: Sage Publication.

Robbins, S. P., & Judge, T. A. (2013). *Organizational behavior* (5th ed.). Pearson Educational Limited.

Sengupta, S. (2004). For Iraq girls, changing land narrow lines. *New York Times*.

Soetan, R. O. (2001). *Culture, Gender and Development*. Awolomo University.

Torrington, D., Hall, L., Taylor, S., & Atkinson, C. (2013). *Human Resources Management*. New York: McGraw Hill.

UNESCO. (2001a). *Changing times, changing attitudes, Portrayal of women and men*. Paris: UNESCO.

UNESCO. (2001b). *Monitoring report on education for All*. Paris: UNESCO.

UNESCO. (2004). *Gender Sensitivity Manual*. Paris: UNESCO.

ADDITIONAL READING

Dakar framework for action. (2000). Unesco.

Mawanza, J. (2010) Dissertation on factors inhibiting growth of female owned businesses in Zimbabwe with special reference to Bulawayo. NUST. Bulawayo.

Schaefer, R. (2010). *Sociology*. New York: McGraw Hill.

KEY TERMS AND DEFINITIONS

Corporate Organizations: Large organizations with a surmountable number of employees.

Discrimination: Unfair treatment.

Gender Empowerment: More consideration and support given to a particular sex.

Gender Equality: Same treatment of women and men.

Gender Oppression: Denial in terms of privileges and advancement to a certain group of people.

Gender Sensitive: Awareness in terms of interests of both sexes.

Gender: Socially constructed roles.

Gender Stereotype: Portrayal of women and men as occupying certain positions in society.

Socialization: Upbringing instructions, norms and values from society.

ENDNOTE

¹ Makandase is the Chichewa name for fat-cooks.

Category G

Geographic Information Systems

Application of Geospatial Mashups in Web GIS for Tourism Development



Somnath Chaudhuri

Maldives National University, Maldives

Nilanjan Ray

Adamas University, India

INTRODUCTION

Tourism is defined by the World Tourism Organisation (NSCB, 2004), as the act travelling for the purpose of recreation and the provision of services for this act. Currently, tourism industry is one of the fastest growing industries all over the world. This smokeless industry is basically a kind of service industry, as it renders service to the tourists and all other supporting industries related to tourism like, hotel industry, transport industry etc. This business involves many socio-economic activities like promotion and advertising tourist spots and destinations, providing effective transport facility, fooding-lodging, entertainment etc. At the same time when the tourism industry is flourishing it helps in socio-economic development of those tourist destinations. It also helps in strengthening the economical status of the country by earning foreign currencies without exporting national wealth. So, it is obvious that if this industry becomes more effective and efficient, it will definitely be the major source of revenue and will take a leading role in the overall economic development of the nation. Information and Communication Technology (ICT) can lead tourism to emerge as a new mantra for alternative economic development (Buhalis, 1998). Information Technology breaks the geographical boundaries so it is shared to the global audiences. Information Technology integrates between tourism product and requirement of the tourists. Due to changes in tourists or visitors behavior, the tourism market

is becoming more segmented with each potential tourist belonging to a number of market segments (Cheng et al., 2002). Tourist operators need to be aware of these changes and be equipped to respond or better still, take a proactive approach. Technological revolution during 1990s brought with it new opportunities and challenges for the tourism industries. Technology has become fundamental to the ability of the global tourism industry to operate effectively and competitively. Information technology is being rapidly diffused throughout the tourism industry and that no player will escape from information technologies impacts. These technological innovations started in the 1970s when the main airlines set up CRSs (Computerized Reservation Systems), with the strategic aim of building a global distribution network for their products. Connecting travel agencies to the CRSs set off a process of distribution automation involving an ever-increasing number of tour operators, carriers, and car hire firms, individual hotels, hotel chains, and other hospitality firms. Geographical Information System, an ICT tool has been extensively used for tourism promotion and management. It was in use for GIS data design and collection, database design management and application of tourism analysis and problem solving. Currently, Internet has become the inseparable part of the Information and communication Technology. Online technologies within the tourism industry have significantly impacted on communications, transactions and relationships between the various industry operators and with

DOI: 10.4018/978-1-5225-2255-3.ch297

the customer, as well as between regulators and operators. The Internet provides many advantages in the tourism industry (Ray et al., 2014). The GIS technicians and researchers started research on how to share the GIS features online, rather than using it as a standalone system. In the year 1993, the Web GIS started evolving rapidly. The online static maps slowly changed to interactive dynamic maps over the World Wide Web. This is the first step of Web GIS. The greatest advantage was to get rid of traditional desktop GIS, its installation and data sharing hazard. Today's web user can create content on the web both collaboratively and individually, allowing for a personalized web experience through wikis, blogs, podcasts, photo sharing, and other technologies. GIS and mapping applications have both benefited from and contributed to these trends, collectively called "Web 2.0" (Pierce et al., 2009). This provides the concept of Geospatial Mashups, especially Map Mashups. The latest trend in the field of geospatial science and technology in Web GIS, is *Geospatial Mashups*. Integration of multiple data layers from multiple sources, is one of the most common and effective functional requirements of Web GIS applications. On the Web GIS context, a Mashup is the process of merging multiple sources of data, both spatial and non-spatial, into a single integrated spatial display. It is about extracting spatial data from a non-spatial source and combining with other spatial data and finally displaying it on a map. This research paper discusses the basic architecture of the Geospatial Mashups in Web GIS and its application in visual impact analysis and strategic management in tourism.

Objectives of the Study

This present study seeks to the application of Web-Based Geographical Information System (GIS), an ICT tool for tourism management and promotion particularly through internet, with a future plan to develop this type of promotion by implementing GIS tools for tourism. In the context of tourism management this present study penetrates the

usage of Geospatial Mashups, a spatial technical tool of Web GIS. It disseminates maximum level of information for tourism promotion in a collaborative manner. This paper examines current development in Web GIS with the implementation of Geospatial Mashup technologies, such as Google Map in the context of map Mashups, and presents a classification of map Mashups and their application in tourism management and promotion. Geospatial Mashup has great potential to facilitate and widen the rapid development of the future web mapping technology in Web GIS in the context of tourism development.

Background and Literature Survey

Geographic Information System is one of the most popular ICT tools for capturing, storing, retrieving, manipulating, mapping and analyzing spatial and non-spatial geographical data in the digital format. GIS is the information system that provides functions including visual 3D presentations about any geographical locations, advanced analysis of digital geospatial information by processing them in an integrated manner.

GIS technology integrates common database operations, such as storage of data, retrieval through query and converting those data to information through statistical analysis (see Table 1). GIS manages region-based information and provides tools for depiction and analysis of various statistics, including population density, economic development, transport facility, types of vegetation etc. GIS helps us to store the detailed information of any region in the databases and maps to create dynamic displays. Additionally, it provides tools to convert and display raw data in the form of 3D maps, run any query, and overlay those databases in ways which is not possible with traditional spreadsheets. These special-efficient as well as effective abilities distinguish GIS from other information systems, and make it one of the most effective ICT tool to a wide range of public and private enterprises for elucidation of events, predicting outcomes, and scheduling

strategies. The United States Geological Survey (USGS) defined “GIS as a computer hardware and software system designed to collect, manage, analyze and display geographically (spatially) referenced data”.

If we want to define GIS covering all its features then it can be categorised as shown in Table 1.

Basically, it can be stated that GIS can use any information that includes location or region. Now, the location of a particular region can be expressed in various manner. It can be done through address, pin code, latitude-longitude etc. GIS can store, compare and analyse different types of information. The system can include information about the land, such as the location of water bodies, different varieties of vegetation, different kinds of soil, contours, transport facilities like roads, railway tracks, subways etc. It can include data about the density of population, economic standard, education level etc. GIS can also include information about the different buildings in the region like sites of factories, farms, and schools, roads, and electric power lines. Now days the GPS provide the satellite images of the shopping malls, cinema halls, hotels, restaurants, banks and even small retail stores also.

Now the question is, how this vast information is stored in this system. Actually, GIS can store data, not the information. Data is stored in the system and it is retrieved and analysed to convert it into relevant information. Data in different forms

and formats can be entered into GIS. Putting information into GIS is called data capture. GIS being a digital system can capture the digital data directly. Data that are already in digital form, such as images taken by satellites and most tables, can simply be uploaded into GIS. An example of this kind is data collected by satellites that show land use—the location of farms, towns, or forests. GIS can also include data in table form, such as population information. But any other non-digital or analog data should be digitised and then punched into GIS database; viz.

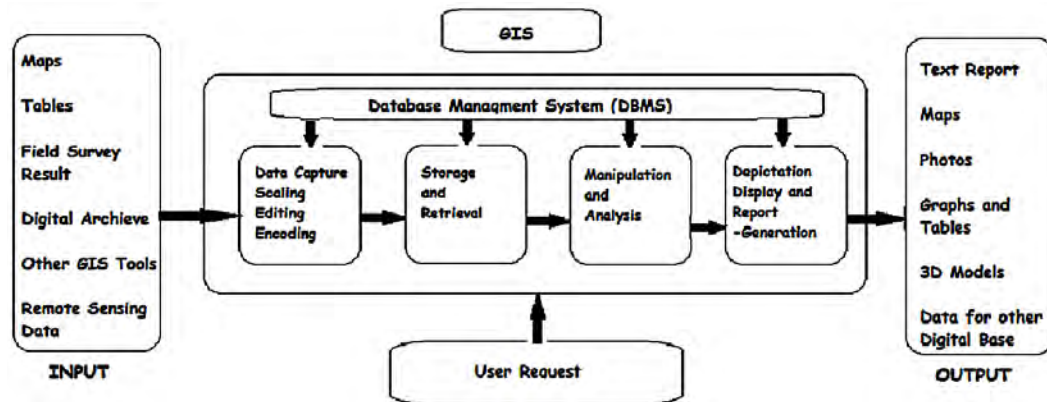
Maps must be scanned or converted into digital data. While punching the data in the database, GIS must make the information from all the various maps and resources align, so that they fit together. This is one of the most vital steps of data storage, this is necessary as the maps collected from different sources may be of different scales. A scale is the virtual relationship between the distance on a map and the real distance on the Earth. GIS integrates all the information from different sources in such a way that it will have all the same scale and no ambiguity during analysis. GIS technology allows all these different types of information, no matter their source or original format, to be overlaid on top of one another on a single map. This map or sometimes statistical reports or virtual models are generated as outputs just on a single click of the mouse of the users’ computer. The complete process of GIS function is depicted in Figure 1.

Table 1. Features and analytical functions of GIS

Features of GIS	Related Definition	Related Analytical Functions
A Process	A system for capturing, storing, checking, manipulating, analyzing and displaying data which are spatially referenced to the earth (DoE, 197: 132)	Presentation and thematic mapping Data Query Spatial Query Database
A Toolbox	Containing tools for collecting, storing, retrieving, transforming and displaying spatial data (Burrough, 1986: 6)	Integration Route finding
A Database	Of spatially referenced entities (Smith et al., 1987)	Point in polygon analysis Overlays
An Application	Cadastral information system, marketing information system, planning information system, etc.	Buffering Visualization and 3D modelling
A Decision Support System	Integrating spatial data within a problem solving environment (Cowem, 1988)	

Source: Bahaire and White (1999, p. 161)

Figure 1. Functionalities of GIS



From the basic features we can redefine GIS as:

- **Geographical**
 - It contains data and concepts which are concerned with spatial distributions and inventory mapping
- **Information**
 - It implies some notion of storage, retrieval and analyses of data, usually as an aid to decision-making.
- **System**
 - It involves the sequence of inputs, processes and outputs (Chaudhuri et al., 2015).

Tourism Classification and Problems

Tourism is viewed as travel for recreation, pleasure or instruction, often in organized groups. It could be visits to attractions, city breaks, and trips for business meetings, sports events, concerts or visits to friends and relatives. Tourism is becoming an integral part of our lives and culture (Balogun et al., 2010). According to Ayeni (2004) tourism can be categorized as follows:

- **Cultural Tourism**
 - It includes religious and national festivals, art galleries, historical monuments, arts and crafts, museum etc.

- **Ecological Tourism**
 - It includes geomorphological, geophysical and geological objects like waterfalls, mountains, sea-beach, desert, national parks etc.
- **Modern Tourism**
 - It includes modern notable engineering structures or architecture, amusement parks, travel and accommodation facilities.

The tourists get information about the popular tourist sites and then travel in search of cultural, ecological or modern tourism aspect. There are many locations or even countries which are sated with numerous ecologically or culturally attractive and significant tourist sites. But many of those locations are barely explored. Fajuyigbe and Balogun (2007) itemized those problems as follows:

- It is a time consuming, hectic and expensive work to updating existing graphical tourist guides and maps.
- Lack of digital online information for tourism amenities and destinations. There is hardly any comprehensive information over the internet.
- Inadequate technology-driven approach for sustainable tourism
- Lack of motivation for strategic marketing techniques and management

All the tourists are not tech-savvy to install a standalone GIS application in their individual computer system and do research about the tourist spots. Rather, it will be much simple if they can collect information over the internet. That leads towards the transition from GIS to Web GIS.

Tourism Using Web-Enabled GIS

Internet has become an inseparable part of the Information and communication Technology. The GIS technicians and researchers started research on how to share the GIS features online, rather than using it as a standalone system. In the year 1993, “Web 2.0”, the Web GIS started evolving rapidly. The online static maps slowly changed to interactive dynamic maps over the World Wide Web and Internet. This is the first step of Web GIS. The greatest advantage was to get rid of traditional desktop GIS, its installation and data sharing hazard. Web GIS is any Geographic Information System that uses Web technologies. In a narrower definition, Web GIS is any GIS that uses Web technology to communicate between two components (Fu et al., 2011). Traditionally, geographic information systems were realized as monolithic and platform-dependent applications (Wong et al., 2002). The major concern is to use GIS as a tool for communication between different interest groups like planners, decision makers, and the public. The rapid growth of the internet provides highly customized, accessible and interactive source of public information and is changing the ways that people capture and manipulate spatial information (Balogun et al., 2010). In 2004, Information Technology Planning Management and Development of Goa, India was developed by Dr. P. K. Pandley and Ruma Chakraborty to advice the Government in developing and managing tourism in Goa (Pandey et al., 2004). Ghana Geographical Information System (GGIS) was developed by Longmatey et al. (2002). It provides GIS in tourism management, marketing and promotion in Ghana. It is one of the principal GIT tool in Ghana tourism market.

Fajuyigbe et al. (2007) developed a web-enabled GIS for the management of tourism in Oyo state, Nigeria. The basic architecture of this application has two sections,

- Server-end and
- Client-end.

The client can access the geospatial information over the internet and the server will process and customize the query for execution. But it doesn't incorporate the Virtual tourism features. An implementation of Web-enabled GIS as a potential tool in sustainable tourism in western Nigeria was done by Fajuyigbe and Balogun (2007). This was an advanced version of Fajuyigbe et al. (2007)'s system. It was used as a potential tool for projecting data and analysis of critical information for the efficient management and promotion of tourism in most of the states in western Nigeria. Having reviewed the above cited works the requirement for integration of multiple datasets from different independent sources in Web GIS applications, has become prominent and inevitable.

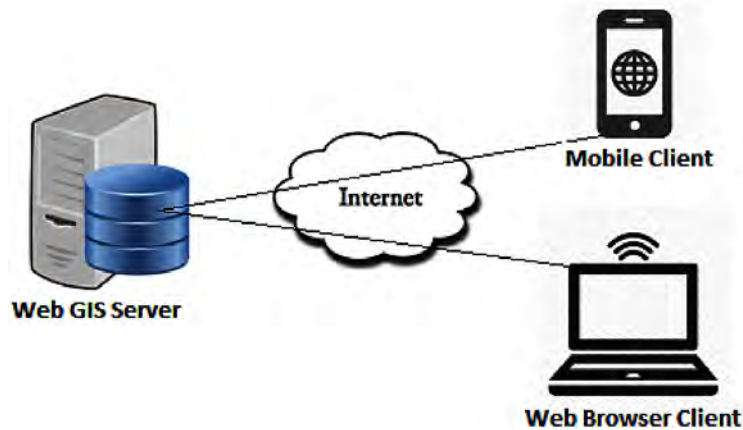
Web GIS Architecture

The “Web GIS” became a synonym for Web information systems that provides a functionality of geographic information systems on the Web through HTTP and HTML (Shanzhen et al., 2001).

The simplest architecture of Web GIS should have at least two components, a Web-application Server and one or more Clients. Client can be as simple as a Web browser or a mobile client. The GIS service providing server can be connected by the client using Internet (HTTP). This dynamic connection is interactive in nature. The basic architecture is represented as in Figure 2.

The simple client side architecture allows the users to carry out some data manipulation, data customization and analysis locally on their own machines. The greatest advantage of this model is its low overhead on the server and high interactivity at the client side. Most of the computing

Figure 2. Basic Architecture of Web GIS



tasks are performed by the client end, as a result, less workload and computation on the server side. With direct access to the database, the capability if client computer can be maximized to match the traditional client/server application and reserves the platform-dependent characteristics at client side concurrently (Ben et al., 1999). On the other side the server side architecture allows clients to submit query for data and analysis to a server. The server executes the query, retrieve the result from the database and process the result. The spatial and tabular data remain on the server. This architecture provides cross-platform capability at the client side. No special plug-in or application software is required for this interactive full-duplex communication. But though the architecture is providing a two-way communication without installing any third party application software, it has a severe drawback in terms of its speed. It is slow because each time a request is made, a new HTML document containing a graph, or map has to be generated and transferred to the client end. The next section will explain how Web GIS works as a web application over the internet. The working principal of Web GIS is basically a query request and response cycle. Suppose the client is accessing the GIS application over the internet using any web browser or from a mobile application. The client is searching for a specific location. The client posts the query in the application or in the browser interface. The Web Server will receive the

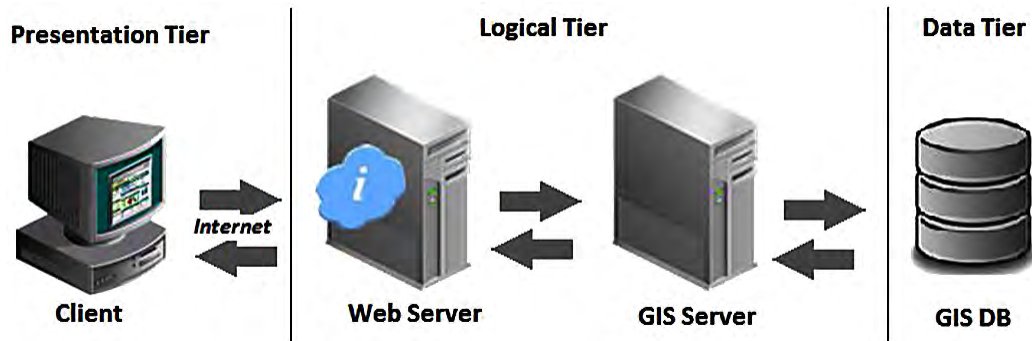
query. In the next step the request will be forwarded from the Web Server to the GIS Server. The query will be executed by the GIS Server and required data will be retrieved from the GIS database. The result can be a simple map or, a satellite image or, a Mashup map, which depends upon the request from the client. The result is then forwarded to that Web Server from where the GIS Server has received the query and finally forwarded to the client. The browser interface or the mobile application will display the result. Based upon the data processing in the Web GIS application the complete workflow has been categorized into three sections, namely (Figure 3):

- Presentation tier,
- Logical tier and
- Data tier.

This complete Web GIS workflow is depicted as follows (see Figure 3).

The Presentation tier is the Web browser interface or the mobile application interface from where the client can send a request. Using the same interface the client can view the result. The web server will display the customized result in the presentation tier in the form of graph or, chart or, may be tabular data. Web server forwards the request query to the GIS server in the Logical tier. The key processing of the request and the query analysis is performed in this tier. SQL query is

Figure 3. Web GIS Workflow



executed by the GIS server and it retrieves the data from the GIS database. After the query is executed the result is forwarded from the GIS server to the Web server. The principal role of the Data tier is to maintain normalised data for effective query processing.

Geospatial Mashups

The three tier Web GIS system is definitely an easier and user-friendly application to get access to geospatial data over the internet just on a click. But the effectiveness and reliability of a Web GIS system is determined by the standard of the GIS server it uses. The quality and standard of the Web GIS application is dependent on the functionality, scalability and ability to be customized of the GIS server. The GIS database is the backend support for any Web GIS application. The quality of the data in the GIS database, determines the standard of the result generated by the Web GIS application. Professional applications definitely need the up-to-date, reliable and high-quality data set. Slowly, it was becoming obvious that to maintain a quality and up-to-date Web GIS database it needs a collaborative contribution. There are clients who are accessing the system for information, but at the same time there might be some other end-users who can update the system with recent information about their locality. The first type of clients are just sending query in the system and accessing the information. They don't have any idea what is going on at the backend of the system, they don't

need to know about the backend processes. But the Web GIS system administrator slowly realised the need of another group of authenticated end-users who can perform some geo-spatial processing tasks. Web GIS initiated from a basic assumption that the global geo information cannot be captured, saved and customized under a single organizational unit. The second group of authenticated end users can provide the up-to-date information and can be accessed globally. This concept of collaborative contribution for the system leads to the new avenue of Geospatial science and technology called Geospatial Mashups. It is the modern technology of providing web mapping or GIS services globally. Integration of multiple data layers or datasets, often from multiple sources, is one of the most common functional requirements of GIS applications (Fu et al., 2011).

The term "Mashup" is derived from pop music, where the musicians create new songs by remixing multiple sound tracks together. Today's web user can create content on the web both collaboratively and individually, allowing for a personalized web experience through wikis, blogs, podcasts, photo sharing, and other technologies. GIS and mapping applications have both benefited from and contributed to these trends, collectively called "Web 2.0" (Pierce et al., 2009). According to OGC (2002), Web Feature Service is one of the most important web data service. It allows a client to retrieve heterogeneous geospatial data from multiple geospatial data servers. Using HTTP as the distributed computing platform it allows fun-

damental database Query operations like, Select, Update, Insert, Delete etc. The counterpart Web Map Service is defined by OGC (2004) as the service that is capable of creating and projecting maps that come concurrently from different heterogeneous sources in any of the following standard image format, SVG, PNG, GIF, and JPEG. WMS consists of three basic operations, namely,

- *GetCapabilities*,
- *GetMap* and
- *GetFeatureInfo*.

These three operations lead to a new feature of remixing data in the web called as the Mashup feature of the Web GIS. In Web GIS, geomashup is being used. A geomashup is a Mashup where at least one of the contents or functions is georeferenced. Geospatial Mashups integrate multiple data sources based on common geographic locations. Mashups carry the overlay functions. In an overlay function, information from different

heterogeneous sources are merged together and represented as a single informative Mashup layer. It includes topological overlay which restructures into a single vector dataset and a graphic overlay which superimpose images or maps on top of other. Most of today's online Mashups are graphic overlays. The simple architecture of Web GIS consists of a server end and client or browser end. The Mashup operation can be implemented in both ends. If the Mashup operation is performed in the server end, it is called a Server-side Mashup and if on the client end it is called as Browser-side Mashup. In the Server-side Mashup architecture, the Mashup web server sends requests to different web services, receives the responses and merges the results. Figure 4 depicts the Server-side Mashup architecture.

In the Browser-side Mashup architecture, the web browser sends requests to different services, receives the responses and displays the composite result. Figure 5 depicts the Server-side Mashup architecture.

Figure 4. Server-side Mashup Architecture

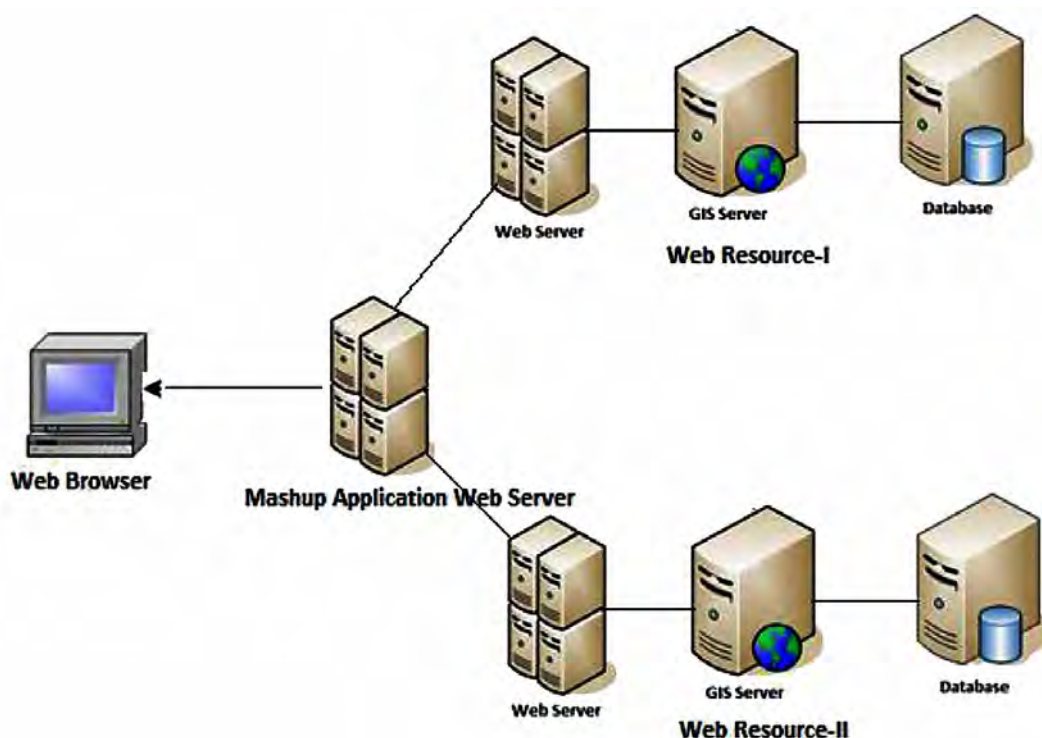
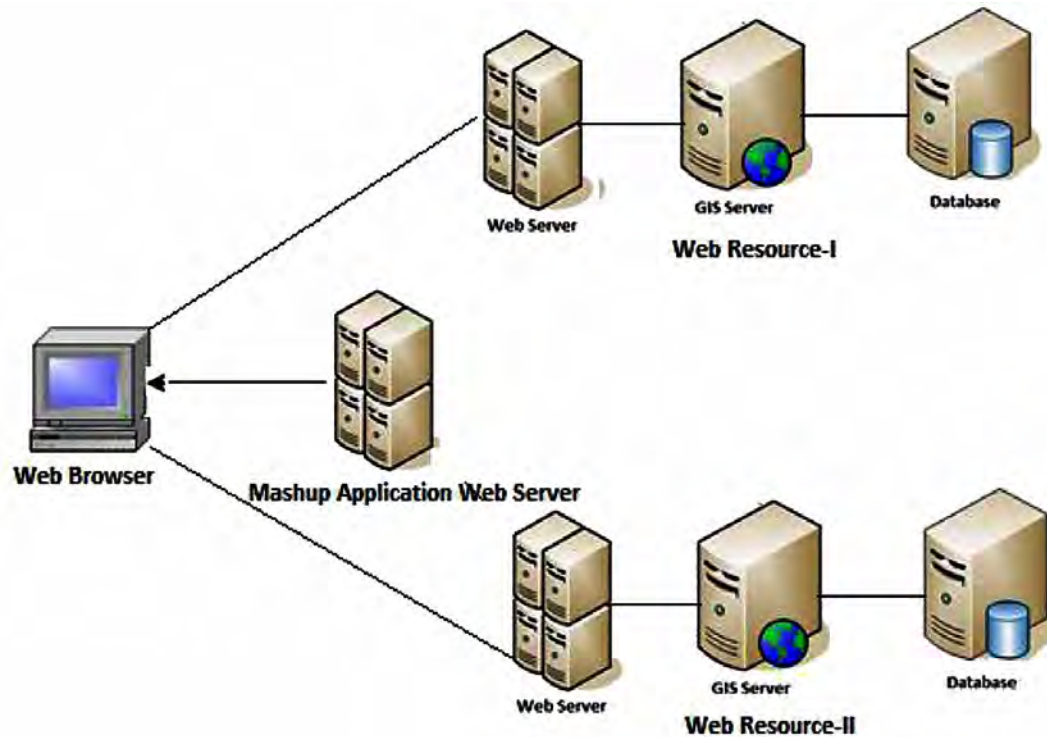


Figure 5. Browser-side Mashup Architecture



Both the architecture has some advantages and disadvantages. The advantage of Server-side Mashups is that the server has more powerful hardware and software than the browser, but these Mashups also require specialized programming tools and complex-server side programming, which includes more manual work during development and deployment. Because of this complex programming and tedious work, the original Server-side Mashups were mainly limited to professional programmers. But most of the today's Mashups are Browser-side. It uses mainly lightweight programming; as a result it has become easy to develop and professional GIS companies have adopted this approach as an easy and fast way to develop Web applications. ArcGIS by ESRI is one such example.

This recent trend of Geospatial Mashup has significant impact on Web GIS. The next section will highlight on the design pattern of Mashups in Web GIS. The common design pattern consists

of three components: Basemaps, Operational Layers, and Tools.

- Basemaps
 - It provides the initial framework for any geomashup. It is the frame of reference like, the aerial imagery maps often from the web services of Google, Yahoo, ArcGIS Online etc.
- Operational Layers
 - It is also known as the thematic layers, usually embedded on the top of a basemaps, like Café, building names, street names, bus stops etc.
- Tools
 - It can execute logical and analytical functions.

Mashups can potentially combine any type of contents and functions over the web, regardless of whether a formal programming interface is

available. The vast majority of contents over the web are HTML pages and photos that don't have formal APIs, but they contain a huge amount of valuable information that can be scraped and geospatially tagged or referenced and then reused to build new and value added applications. Virtually, the whole web can be remixed, which opens the door to unlimited value-added Mashup applications (Fu et al., 2011).

Application of Geospatial Mashup in Tourism

The ability to examine data in its spatial context is extremely important. The geocoding process takes a simple street address and matches it to a database containing address ranges for every link in the street network. In this way customer lists may be converted into simple dot maps, or coupled with census data and displayed as maps of market penetration (Goodchild et al., 1984). This research paper concentrates mainly on the application of Geospatial Mashups in tourism sec-

tor. With respect to the tourism industry the map Mashup can be classified into four broad sections:

- Informative Mashup
- Participatory Mashup
- Collaborative Mashup
- Enterprise Mashup

Check the Google map illustrated in Figure 6.

The initial map was developed without the detail information about the stadium, the tourist lodge, the hotels, the schools, colleges, banks and the historical and religious architectures like Nandalal temple, Chinnamasta mandir etc. Initial aerial imagery map was only with the land structure and the roads, that is the Basemaps. Other than the road and basic framework the basemaps include the water-bodies like the Lal Bandh or other basemaps' objects mention in blue in the above Google Mashup map. Next in the Operational Layers other end users have implemented Participatory or Collaborative Mashups. Some end users are providing the information of the

Figure 6. Mashup in Web GIS

Source: www.google.com/maps/



historical place of interest like the temples. Some other client has supplied the information about hotels, restaurants and banks. Few others may have contributed in providing information about the local school, colleges and health centers. There is high probability of receiving the data from different clients in heterogeneous form. When all these heterogeneous data is compiled and remixed using WFS and WMS service, it results into this detailed map, containing every in and out about the location. Thus, this Mashup service is providing detailed information about a location in a collaborative manner. The basic information are required by a potential tourist to visit any tourist location. Initially, the tourists have to depend on the information provided by the travel agents about a location. The agents used to provide the information over telephonic conversation or sharing documents over email. The information may not be up-to-date and the tourists have to solely trust the agents. The tourists could have viewed the aerial imagery maps over the internet, but those maps don't provide detail information about the road, hotels, place of interest, transport facilities available etc. But now with the detailed geospatial Mashup services over any Web GIS applications, the tourists can collect every in and out detail information about their place of choice. They have to just browse the internet, run the online Web GIS application and run a query. They will receive detailed information about the location. Another greatest advantage in this Participatory or Collaborative Mashups is that, the globally accessible geomashups are always up-to-date; they contain the most recent information about the location. Even the tourists while visiting those tourist sites if they found something is not matching with the Mashup information over the internet, they can customize those published information with the updated one. This will be beneficial for other potential tourists. These are the benefits of updating information of those locations which are already established as tourist spots. Now, if some entrepreneur or travel agency wants to project a new location as a tourist spot, they can perform

initial research work to analyse whether their local area can be a potential tourist location or not. First and foremost they have to find out whether that particular location has any Cultural, Ecological or Modern tourism component or not. If it is there they can update that information using the Mashup option in Web GIS application. Next, they have to update other facilities like, transport, hotels, café, health service etc. even internet facilities available or not. That means, the locations which can be potential tourist spots can slowly come into limelight just by some simple initiatives taken by locals. One such online application provided by Google is <http://www.editor.giscloud.com/> where the end users can update and create a map online contributing detailed information about their locality or any other place of interest. They don't need any Government intervention. In this way the Goespatial Mashups are making hundreds of new Cultural, Ecological or Modern tourist locations popular. This is effecting overall economic development of the location through tourism industry. On the other hand, in today's competitive marketplace, leading companies are analyzing and listening what customers really care about. Traditional forms of marketing are no longer as effective as they used to be in the past. This is where Web GIS comes in and allowing the companies to visualize where potential customers are located by analyzing demographic, psychographic, purchasing, and spending characteristics for accurate customer segmentation and helping the companies to find more like them (GIS for Marketing: Where Strategy Meets Opportunity, 2010). The tourism sector is implementing the Web GIS Mashup application as a promotional and marketing tool.

Challenges

Mashups provide different opportunities for sharing data collaboratively and using information in new applications and systems. It has become an integral design pattern for Web GIS applications. As more and more functions and applications

become available over the Web and more policies and technologies favour the use of Mashups, this branch of geospatial science and technology will grow exponentially in future and will contribute in tourism industry. Like all other technologies Mashup also has few challenges and demerits. New research works are going on to address the following drawbacks of Mashups:

- **Quality and Reliability of Information**
 - Using the Participatory and Collaborative Mashups anyone can publish contents over the internet. No one is authenticating the genuineness of the uploaded information. Quality can vary enormously, and if questionable sources are implicated in a chain of services, the results of uncertainty or error can be propagated over the internet. This can produce misleading and even wrong information (Goodchild et al., 2010).
- **Copyright and Ownership**
 - While the spirit of open exchange is still the main characteristics of the web, many websites post terms of use and retain their copyright. Copyright and terms of use can get complex when a Mashup uses multiple web resources, or even more complex when a Mashup uses information generated by another Mashup. It is important for Mashup developers to comply with provider's terms of use to avert future disputes and to understand the impacts of copyright on their application (Fu et al., 2011).
- **Security**
 - Enterprise Mashups can involve confidential information sharing over the internet, making security a key consideration.
- **Technical Challenges**
 - Technical challenges in dealing with map Mashups are mainly related to

technical incompatibility, data integration and quality assurance, interfaces and functionality, level of programming skills required, enterprise integration, etc (Li et al., 2008).

CONCLUSION

Impact of ICT in tourism is very vital. For instance, a guide requires historical and cultural information about the destinations, tour agency requires information on transformation, reservations, accommodation in Hotel, and for marketing, a tourist requires information from Cell Phone Application (SMS), Internet, Wireless (Wi-Fi), VOIP, GPS, GIS, Digital Radio, On demand application and so on. Role of Information Technology in tourism industry cannot be underestimated because it has vital contributions. Once the tourism information or data is in the system (computer) this may be transmitted to the proposed tourists through different channels viz. FAX, GPS, Website etc. One of these informative IT tools is Geospatial Information Technology (GIT). GIS is an information system that provides functions including visual presentation about the tourism destination, advanced analyses, etc., of digital geospatial information by processing it in an integrated manner on electronic maps of the tourist site. The benefits of Web GIS, especially Geospatial Mashups far outweigh its limitations and challenges. Web GIS disseminates maximum level of information to the potential tourists, at the same time allows the travel agents or organizations to promote an unknown location having adequate features to be a popular historical, ecological or modern tourist spot. The potential tourists can analyze demographic, economic, psychographic, market characteristics of the location just on a click. Thus, providing support for strategic planning, Collaborative information sharing, online marketing techniques, improving efficiency in tourism management and attaining competitive locational and business advantage for the tourism agencies. The future research scope

of this present paper depicts the application of Web GIS and Geospatial Mashups and how this modern GIT system can be used well as an analytical and strategic tool in tourism management and promotion.

Followings are the domain of future research in Web GIS and Geospatial Mashups in context to tourism business.

1. Web GIS Applications in tourism resource inventories
2. Web GIS application to location suitability under conflicting demands
3. Monitor and control tourism activities using Web GIS
4. Application of Web GIS in visual impact analysis in tourism research
5. Application of Web GIS community involvement and participation in tourism research
6. Web GIS uses as a decision supporting tools in tourism research
7. Application of Geospatial Mashups in tourism promotion
8. Limitations in Geospatial Mashups and their solution

REFERENCES

- Ayeni, O. O., Saka, D. N., & Ikwuemesi, G. (2004). Developing a Multimedia GIS Database for Tourism Industry in Nigeria. *ISPRS 2004: 20th International Society for Photogrammetry and Remote Sensing (ISPRS) Congress: Geo-Imagery Bridging Continents*.
- Balogun, V. F., Thompson, A. F., & Sarumi, O. A. (2010). A 3D Geo-Spatial Virtual Reality System for Virtual Tourism. *The Pacific Journal of Science and Technology*, 11(2), 601–609.
- Buhalis, D. (1998). Strategic use of information technologies in the tourism industry. *Tourism Management*, 19(5), 409–421. doi:10.1016/S0261-5177(98)00038-7
- Butler, R. (1993). Alternative Tourism: The Thin Edge of the Wedge. In V. Smith & W. Eadington (Eds.), *Tourism Alternatives* (pp. 31–46). Chichester, UK: Wiley.
- Chaudhuri, S., & Ray, N. (2015). Application of Web-Based Geographical Information System (GIS). In *Tourism Development, Encyclopedia of Information Science and Technology*. IGI Global.
- Cheng, C., & Piccoli, G. (2002). Web-Based Training in the Hospitality Industry: A Conceptual Definition, Taxonomy and Preliminary Investigation. *International Journal of Hospitality Information Technology*, 2(2), 19–33. doi:10.3727/153373402803617737
- Davin, K. (1997). Effects of Computer Support on Group Decisions. *Journal of Hospitality & Tourism Research (Washington, D.C.)*, 21(2), 44–57. doi:10.1177/109634809702100204
- Fajuyigbe, O., & Balogun, V. F. (2007). Web-Enabled GIS: A Potential Tool for Sustaining Tourism in (Western) Nigeria. *Nigeria. Computers & Society*, (June): 15–22.
- Fajuyigbe, O., Balogun, V. F., & Odembe, O. M. (2007). Web-Based Geographical Information System (GIS) for Tourism in Oyo State, Nigeria. *Information Technology Journal*, 6(5), 613–622. doi:10.3923/itj.2007.613.622
- Fesenmaier, D. R., Vogt, C. A., & Stewart, W. P. (1993). Investigating the influence of welcome center information on travel behavior. *Journal of Travel Research*, 31(3), 47–52. doi:10.1177/004728759303100308
- GIS for Marketing: Where Strategy Meets Opportunity. (2010). Retrieved on June 20, 2015, from <http://www.esri.com/marketing>
- Goodchild, M. F. (1984). ILACS: A Location-Allocation Model for Retail Site Selection. *Journal of Retailing*, 60, 84–100.

- ICT & Tourism USAID from the American People. (n.d.). *International Journal of Knowledge and Learning*, 3(6), 628-648. doi:10.1504/IJKL.2007.016837
- Bahaire, T., & Elliott-White, M. (1999). The Application of Geographical Information Systems GIS in Sustainable Tourism Planning: A Review. *International Journal of Sustainable Tourism*, 7(2), 159–174.
- Jafar Jafari. (2003). Encyclopedia of Tourism (p. 308). Academic Press.
- Li, S., & Gong, J. (2008). MASHUP: A New Way of Providing Web Mapping/GIS Services, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 37(B4), 639–648.
- Longmatey, D., Samuel, A., & Prah, B. K. (2002). *Management and Promotion of Tourism in Ghana: A GIS Approach from African Tourism Websites*. Retrieved on May 20, 2015. <http://www.african-tourism.com/ghana.htm>
- Pandey, O. K., & Chakraborty, R. (2007). *GIS: Information technology for planning management and development of GOA tourism master plan in next millennium*. Retrieved on May 20, 2015. <http://www.gisdevelopment.net/application>
- Pierce, T., Dobson, G., & Phillips, M. (2009). GIS Mashups on the Web: Using Google, ESRI, and Open Source to Develop Weather and Climate Applications for the Web. *Geospatial Infrastructure Solutions Conference*, Tampa FL
- Pinde, F., & Sun, J. (2011). *Web GIS, Principles and Applications*. ESRI Press.
- Rahaman, M. (2010). Application of GIS in EcoTourism Development: A Case Study in Sundarbans, Bangladesh. *International Journal of Entrepreneurial Behaviour And Research*, 10(4), 247-59.
- Ray, N., Das, D. K., Ghosh, S., & Banerjee, A. (2011). A Study on Application of Information Technology on Tourism Development, IEEE Proceeding Catalog No: CFP1141H-ART ISBN978-1-4244-6579- China
- Ray, N., Das, D. K., & Chaudhuri, S. (2014). Tourism Promotion Through Web. In *Strategic Marketing in Fragile Economic Condition*. IGI Global.
- Shanker, S. (2008). *ICT and Tourism: Challenges and Opportunities, Conference on Tourism in India – Challenges Ahead*. IIMK.
- Shanzhen, Y., Lizhu, Z., Chunxiao, X., Qilun, L., & Yong, Z. (2001). Semantic and Interoperable WebGIS. *Proceedings of the Second International Conference on Web Information Systems Engineering*, 2, 42-47. doi:10.1109/WISE.2001.996702
- Sigala, M. (2008). *Web 2.0, social marketing strategies and distribution channels for city destinations*. Academic Press.
- Wong, S. H., Swartz, S. L., & Sarkar, D. (2002, April). A Middleware Architecture for Open and Interoperable GISs. *IEEE MultiMedia*, 9(2), 62–76. doi:10.1109/93.998065

ADDITIONAL READING

- Buhalis, D. (1998). Strategic use of information technologies in the tourism industry. *Tourism Management*, 19(5), 409–421. doi:10.1016/S0261-5177(98)00038-7
- Chervenak, L. (1993). Hotel Technology at the start of the New Millennium. *Hospitality Research Journal*, 17(1), 113–120. doi:10.1177/109634809301700110

Cho, Y., & Fesenmaier, D. R. (2001). A new paradigm for tourism and electronic commerce: Experience marketing using the virtual tour. In E. Laws & D. Buhalis (Eds.), *tourism distribution channels: practices, issues and transformation* (pp. 351–370). Thomson.

Collins, C., Buhalis, D., & Peters, M. (2003). Enhancing SMTEs Business Performance Through the Internet and U-learning Platforms. *Education + Training*, 45(9), 483–494. doi:10.1108/00400910310508874

Fesenmaier, D. R. (1994). Traveler use of visitor information centers: Implications for development in Illinois. *Journal of Travel Research*, 33(1), 44–50. doi:10.1177/004728759403300106

Klein, S. (2002). Web impact on the distribution structure for flight tickets. In K. W. Wöber, A. J. Frew, & M. Hitz (Eds.), *Information and Communication Technologies in Tourism 2002* (pp. 219–228). New York: Springer-Verlag Wien. doi:10.1007/978-3-7091-6132-6_22

O'Brien, D., & Fitzgerald, B. (2006). Mashups, remixes and copyright law. *Internet Law Bulletin*, 9(2), 17–19.

Werthner, H., & Klein, S. (1999). ICT and the changing landscape of global tourism Distribution. *Electronic Markets*, 9(4), 256–262. doi:10.1080/101967899358941

KEY TERMS AND DEFINITIONS

CRS: Computer Reservation Systems is a computerized system used to store and retrieves information and carry out transactions related to the information or booking of airline ticket, railway ticket, hotel booking etc. CRS has made the online booking and collecting information much easier, faster and effective.

Geospatial Mashups: Mashup is picking data or functionality from different sources (both homogeneous and heterogeneous) on the web and

combining them within the browser to present to the user. On the Web GIS context, a Mashup is the process of merging multiple sources of data, both spatial and non-spatial, into a single integrated spatial display. It is about extracting spatial data from a non-spatial source and combining with other spatial data and finally displaying it on a map.

GIS: Geographic Information System is one of the most popular tools of GIT. It is the modern technology for capturing, storing, retrieving, manipulating, mapping and analyzing spatial and non-spatial geographical data in the digital format. GIS is the information system that provides functions including visual 3D presentations about any geographical locations, advanced analysis etc. of digital geospatial information by processing them in an integrated manner.

GIT: Geospatial Information Technology is a collection information communication tools used to capture, store, transform, manipulate, analyze, and produce information related to the surface of the Earth. This data may exist as maps, graphs, 3D virtual models, tables, and/or lists. Example: GIS, GPS etc.

GPS: Global Positioning System is one of the most popular tools of GIT. GPS is a radio navigation system that allows land, sea, and airborne users to determine their exact location, velocity, and time 24 hours a day, in all weather conditions, anywhere in the world. The GPS is being used in science to provide data that has never been available before in the quantity and degree of accuracy that the GPS makes possible.

ICT: Information and Communication Technologies, it refers to technologies that provide information through telecommunications. ICT and Information Technology (IT) are more or less similar, but ICT focuses primarily on communication technologies. This includes the Internet, wireless networks, cell phones, satellite systems and other communication mediums.

Tourism: Tourism is a social, cultural and economic phenomenon related to the movement of people to places outside their usual place of residence. Tourism comprises the activities of

persons travelling to and staying in places outside their usual environment for not more than one consecutive year for leisure, business and other purposes not related to the exercise of an activity remunerated from within the place visited.

Tourist: Someone who travels for pleasure and recreation and exchange of culture.

V. R. Technology (V. R.): Virtual Reality Technology is a computer-simulated environment that can simulate virtual reality, i.e. an artificial environment created by software and presented to the user in such a way that the user accepts it

as a real environment. In VR, highly visual 3D environments are created by using CAD software and graphics hardware. VR has its application in different sectors viz. Military training, Architecture, Game, Medical Procedures etc.

VOIP: Voice over Internet Protocol is a collection of technology for the full duplex real-time transmission of voice communications and multimedia sessions over the internet. VoIP offers a substantial cost savings over traditional long distance telephone calls.

Archaeological GIS for Land Use in South Etruria Urban Revolution in IX–VIII Centuries B.C.



Giuliano Pelfer

University of Florence, Italy

INTRODUCTION

Urban Revolution in Ancient Etruria Between IX–VIII Centuries B.C. Origin of the Etruria Protocities

Archaeological discussion about protourban centers growth in middle Tyrrhenian and in Ancient Etruria regions outlined the original and revolutionary features of such process, called Villanovan Revolution for resuming the deep transformations occurring in a few decades.

New largest protourban settlements grew through the concentration, fusion and absorption of the earlier scattered sites on the plateaux of South Etruria, extended for many hectares and well defended (Bartoloni, 2002; Bietti Sestieri, 1996; Carandini, 2003; Mandolesi, 1999; Pacciarelli, 2000; Peroni, 1994, 1996).

Geographic position of plateaux and agrarian quality of adjacent soils were the key factors for settlement position selection. Tens of scattered older villages were abandoned with simultaneous people transfer, called synoecism, to the plateaux from the late final Bronze age to the first Iron age. The process features, the plateaux selected and their area extension would prove existence of some urban planning. New settlements extension, until to about 200 hectares, would testify the great dimension of the new communities born from synoecism (Peroni, 1969, 1994) and a gradual reduction of settlements number with significant increase of their population size and surface extension (Carandini, 2003; Pacciarelli, 2000).

Analysis Objectives

Although many aspects of protourban centers growth, like the foundation of Rome and its organization in *curiae* are investigated, relevant criticism remains about the same processes in South Etruria. In fact, some issues about the strong changes in settlement patterns observed in this period have to be yet clarified and deepened (see Background section - Bartoloni, 2002; Pacciarelli, 2000; Peroni, 1994).

The present research has the goal to explain the main factors that favored such historical process as well the features of the settlement strategies.

Use of GIS is the most powerful technology introduced to archaeology since the introduction of carbon 14 dating and seems the most suitable for specific information.

Most widespread use of this technology is for Cultural Resource Management, data visualization, excavations but also for prediction of archaeological site locations.

This paper focuses on the use of GIS for archaeological predictive modeling of ancient land use, by critically applying this new technology and exploring its theoretical and analytical implications. Archaeological data are point like geographical and temporal data with more or less large uncertainty. Work of the archaeologists is to connect all the point like data related to a geographic region and to an historical period using a qualitative and quantitative based narration that fit in the best way with available data and with the theory and model.

DOI: 10.4018/978-1-5225-2255-3.ch298

An archaeological GIS is designed with the aim to record the existing and reconstructed data on a database, to visualize data by thematic maps and to use such data for advanced statistical and spatial analysis. Archaeological GIS is realized installing GRASS GIS on OS Linux with interface to PostgreSQL database, with its extension PostGIS for geographic information and to R package for statistical and geostatistical analysis.

BACKGROUND

Archaeological Theories on Protourban Centers Genesis

Archaeologists suggested two approaches to protourban centers origin: the first approach underlined a continuity between protourban centers and earlier settlements through gradual development of settled areas, between the late final Bronze age and the beginning phases of first Iron age (Pacciarelli, 2000); the second approach suggested an ungradual transition from the scattered villages on the territory to the protourban centers with an extension 30 times larger than the villages of preurban phase (Peroni, 1994, 1996). Surface surveys of last fifty years showed an uniform distribution of the villanovan pottery remains, attesting a protourban phase after 1,000 B.C., with occupation of large areas, because of transformations in social and economical structures (Pacciarelli, 2000).

Villanovan Urban Revolution During X-IX Centuries B.C.

Studies of Peroni and Rittatore Vonwiller in the 1960s (as cited in Pacciarelli, 2000, pp. 11-12) outlined the close relationship of Urban Revolution in South Etruria with social, political and economic transformations starting from 1,000 B.C. Changes concerned the techniques of agrarian production, the work organization and the social relationships.

A new economic organization of the agrarian property is based on the private ownership of land by division into lots for households (Mandolesi, 1999; Pacciarelli, 2000; Peroni, 1994).

In this context the nuclear family emerged as the basic social cell of the protourban centers, replacing the older multi familiar clans.

Social hierarchies reflected kinship and rank relationships, as shown from the funerary data since the X century B.C. (Bartoloni, 2002, 2003; Iaia, 1999; Pacciarelli, 2000; Peroni, 1996).

Urban Revolution developed between XII-VIII centuries B.C. in the Mediterranean area (Moscato, 2001) and along Mediterranean sea coasts, favored by the emergence of a global Mediterranean market area. New situation produced huge changes: villages in a few tenths of years disappeared and protourban centers grew.

Collapses of the great eastern territorial empires and of the earlier palatial states, also due to the invasions by the People of the Sea, promoted the propulsive role of Phoenicians in the new Mediterranean global market between XII-VIII centuries B.C.

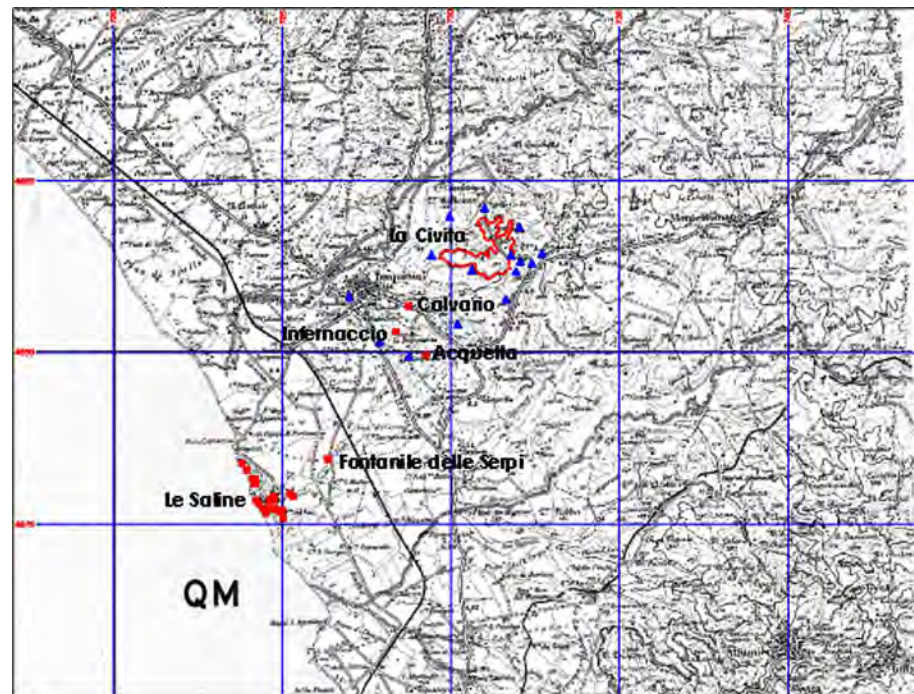
Such role of Phoenicians has been recognized in the last decades (Braudel, 2002; Giardino, 1998; Guidi, 1998; Liverani, 1988; Moscato, 2001; Pacciarelli, 2000; Trigger, Kemp, O'Connor, & Lloyd, 2000). Because of Phoenicians maritime activities, autonomous urban centers developed in the Etruria, middle Tyrrhenian and in central Aegean areas.

Case Study: Tarquinia

The Tarquinia settlement is an important case study on the genesis of protourban centers (see Figure 1.)

The surface surveys testified the uniform distribution of the villanovan pottery in all area, showing a full plateaux extension of the protourban center. Data show that the main center covered about 120 hectares on main plateaux of La Civita, including the oldest final Bronze age

Figure 1. The Tarquinia protourban center between final Bronze age and first Iron age



La Castellina and Cretoncini sites (Di Gennaro, 1986; Hencken, 1968; Mandolesi, 1999; Pacciarelli, 2000; Peroni, 1996).

The protourban center geographical structure is organized in three settlement subsystems: La Civita and the Calvario, Infernaccio, Acquetta villages on Monterozzi hill, their corresponding cemeteries, and the satellite coastal settlements of Fontanile delle Serpi and le Saline. The oldest settlement system developed on La Civita and on Castellina and Cretoncini sites with their cemeteries (Bietti Sestieri, 1996; Hencken, 1968; Iaia, 1999; Linington, 1982; Mandolesi, 1999; Pacciarelli, 2000; Peroni, 1996).

Funerary data and cemeteries distribution fit well with all the approaches about the protourban genesis and for this reason such information cannot be used for supporting neither theory (Carandini, 2003; Mandolesi, 1999).

The social evolution of the protourban center, linked to familiar subgroups, likewise to the organization of the future *curiae* and *gentes* in the archaic Rome (Giardino, 1998; Mandolesi, 1999), extended the settlement system to the Monterozzi

hill as shown by the cemetery at Le Rose near to the Calvario village (Carandini, 2003; Moscati, 2001; Pacciarelli, 2000). Monterozzi hill becomes important for its favorable position for coastal plain visibility and control.

MAIN FOCUS

Agrarian Production in Final Bronze Age

Ancient sources and archaeological data recognized key economic role of agriculture and sheep farming in the Etruria villanovan communities (Colonna, 2000; Pallottino, 1937; Sassatelli, 2009). The increasing demand of cereals characterized exchanges intensification during Urban Revolution. From late Bronze age, the agrarian activities prevailed on other ones, as proved by small amount of wild paleozoological remains. Changes in agrarian activities are due to major economic and social transformations as well to larger diffusion of bronze working tools rather

than to great innovations in the farming technology (Ampolo, 1980; Peroni, 1994; Sassatelli, 2009). In fact, heavy soils plowing required a larger work and time amount supplied by motivated people and by new work organization, before the iron plowshare developed (Forni, 2002; Peroni, 1994; Van Joolen, 2003).

The Main Crops

Ancient sources testify that selected cereals like the hulled wheats as *triticum monococcum* and *triticum dicoccum* (spelt) and the naked wheats as *triticum aestivum* and *triticum compactum*, were used in agriculture until to the VIII century B.C. (Bartoloni, 2002; Bietti Sestieri, 2002; Sassatelli, 2009). Sources clarify which were the more diffused crops and the suitable requirements for soil features (Columella, II, 9, 5-9; Varro, I, 9; Cato, 6,1).

All the variables linked to cultivated crops and agrarian techniques are used for understanding and modeling the productive process. The grazing and the breeding were also important activities in villanovan economy (Bedini 1997; Sassatelli, 2009).

This is testified by paleozoological remains in many protohistorical sites.

The breeding could be practiced both around plateaux and on coastal plain, due to proximity of rivers and lagoons as water resources satisfying animals daily needs.

Bovines were utilized for agrarian work and for increasing land productivity by their manure, but not for food, because religious tradition forbidden to eat their meats (Ampolo, 1980).

Instead, sheep and goats grazed on free terrains as well in the coastal plain marshy areas and were used in particular for milk and the secondary products, as wool and skins.

Territory Control and Viewshed

An optimal position respect to available resources and to communication ways could also explain the settlements location choice. The resources avail-

ability in surrounding territory was crucial factor for protourban centers growth, like its control by the main settlement.

The viewshed analysis can outline these factors and is also useful for defining the cultural and cognitive landscape (Van Leusen, 2002).

Issues and Problems

Archaeological land evaluation estimates ancient land uses of the studied territories.

Soil scientists developed land evaluation methods for estimating alternative soil uses.

The outcome is a limited number of alternative land uses for the different situations, according to geomorphological framework that allows to classify the landscape in different Landscape Systems, identified as homogeneous areas for their physical, geomorphological and ecological features.

The main goal of such classification is to distinguish such territorial areas with their specific features for favoring the agrarian soils classification with the suitability categories for different crops.

These methods are based on reorganization and interpretation of inventories of soil, vegetation, climate and other aspects of land, for comparing alternative land uses, and their requirements with land resources.

This approach, used in physical geography, has then to be replaced by models built on protohistorical context based on ethnographic, archaeological and historical sources.

SOLUTIONS AND RECOMMENDATIONS

Methodology

Tables and maps for the ancient land use are created by many and complex steps.

All the thematic maps needed for land use analysis are defined.

The list includes the Digital Elevation Model (DEM), the topographic maps, the geological and

hydrological maps, the land cover maps, the soils maps, the land use maps and any other useful map. Data available were generally not digitized and georeferenced.

Available maps displaying the present or the historical situation used different datums and coordinate systems. Any map had to be represented by using the same datum and the same coordinate system by suitable mathematical transformations.

Data finally have been stored in PostgreSQL database with PostGIS extension.

The same group of thematic maps, as in defined list, was built following an historical order, that starts from the present and includes historical periods for which data useful for analysis, although not complete, were available. Any group of thematic maps is characterized by the parameter of time indicating the historical period referred to the group of maps itself.

A time series of thematic maps has been built to infer the information needed for ancient soils maps. The maps information from historical sources and from archaeological remains, going back in time, is incomplete or missing.

GIS based model for reconstructing past soils maps and for watershed analysis has been built. The model used a geostatistical approach to the analysis of the time series of thematic maps, to predict the ancient soils maps: an approach similar

in some extent to the one used to predict the future soils maps (Rossiter, 1994; Van Joolen, 2003).

The evaluation of ancient soils suitability to main used crops required further information on ancient crops, on used techniques, on work organization due to the large scale changes in agrarian activities. For establishing the soils suitability to the main used crops, the physiographic and pedological features of Land Use Requirements (LUR), as drain categories, typology, depth, slope, and so forth, have to be related with the needed Land Use Qualities (LUQ), as moisture, workability, nutrients, and so forth, that represent the soil capacity to satisfy the proper requirements. This is shown in Table 1.

The latter allow to define the specific Land Use Types (LUT).

From the previous tables other ones can be extracted for the specific land use where the needed requirements correspond to the pertaining features according to the general table itself.

LUR have to be measured and related to the LUQ, classified in Severity Levels and Degrees of Limitation, that are the categories of LUR for any LUQ.

From this information and from the land features, three main suitability categories for Severity Levels and Degrees of Limitations for any specific

Table 1. Relationships between LUQ and LUR for soils crop suitability

Land Use Qualities	Land Use Requirements
Moisture	Drain categories
	Soil suitable depth
	Field capacity
Nutrients	Soil typology
	Soil thickness
Workability	Slope
	Stones or rocks presence
	Poor drainage
	Soil texture
Erosion risk	
Rooting conditions	Stages
	Soils structure and texture

land use, are defined: suitable, partially suitable, not suitable (Cremaschi, 2000).

For such information look at Table 2 and Table 3.

Territory maps for any soil feature, as slope, soil typology, available moisture, texture, and so forth, can be built by GIS. The maps obtained with this method are superimposed.

The suitability category of a soil is defined by its more unfavorable feature, whatever is the assigned value to the other features.

Main crops suitability are shown in Table 4 and Table 5.

Table 2. LUT for spelt

Land Use Type (LUT)		
Spelt cultivation on medium-sized plots		
Land Use Quality (LUQ)	Land Use Requirements (LUR)	Suitability categories
Moisture	Soil depth	
	• Less than 30 cm • More than 30 cm	• Not suitable = 3 • Suitable = 1
	Drain categories	
	• Poorly drained • Drained • Very drained	• Suitable = 1 • Suitable = 1 • Not suitable = 3
Nutrients	Soil typologies	
	• Luvisols, Vertisols, Fluvisols • Arenosols, Regosols, Planosols • Leptosols • All other types	• Suitable = 1 • Not suitable = 3 • Suitable = 1
	Lime content in the soil	
	• Very calcareous • Calcareous • Non-calcareous	• Not suitable = 3 • Partially suitable = 2 • Suitable = 1
Workability	Soil texture	
	• Sandy soils • Loamy soils • Clayey soils	• Not suitable = 3 • Suitable = 1 • Suitable = 1
	Slope class	
	• 0-13% • 13-55% • >55%	• Suitable = 1 • Partially suitable = 2 • Not suitable = 3
	Stoniness class	
	• 0-3% • 3-90% • >90%	• Suitable = 1 • Partially suitable = 2 • Not suitable = 3
	Rockiness class	
	• 0-10% • 10-50% • 50-100%	• Suitable = 1 • Partially suitable = 2 • Not suitable = 3

Table 3. LUT for other cereals

Land Use Type (LUT)		
Other cereals cultivation on medium-sized plots		
Land Use Quality (LUQ)	Land Use Requirements (LUR)	Suitability categories
Moisture	Soil depth	
	• Less than 30 cm	• Not suitable = 3
	• More than 30 cm	• Suitable = 1
	Drain categories	
	• Poorly drained	• Not suitable = 3
	• Drained	• Partially suitable = 2
	• Very drained	• Suitable = 1
Nutrients	Soil typologies	
	• Luvisols, Vertisols, Fluvisols	• Suitable = 1
	• Arenosols, Regosols, Planosols	• Not suitable = 3
	• Leptosols	• Partially suitable = 2
	• All other types	• Partially suitable = 2
	Lime content in the soil	
	• Very calcareous	• Partially suitable = 2
	• Calcareous	• Suitable = 1
	• Non-calcareous	• Not suitable = 3
Workability	Soil texture	
	• Sandy soils	• Not suitable = 3
	• Loamy soils	• Partially suitable = 2
	• Clayey soils	• Suitable = 1
	Slope class	
	• 0-13%	• Suitable = 1
	• 13-55%	• Partially suitable = 2
	• >55%	• Not suitable = 3

Table 4. Land suitability for spelt

Spelt	Suitable	Partially suitable	Not suitable
Soil requirements	At least 30 cm thickness		< 30 cm thickness
	Clayey soils, loamy clayey soils, loamy soils	Loamy soils	Sandy soils
	Moderately drained (wet) or drained	Marshy	Very drained (dry)
	Firmly structured		Loose
	Marginally fertile or fertile	Fertile	Very fertile
	Non-calcareous	Calcareous	Very calcareous
		Medium salinity	
Cultivation requirements	Rotation system of one to two years		
	No manuring		
	No irrigation		

Table 5. Land suitability for other cereals

Other cereals	Suitable	Partially suitable	Not suitable
Soil requirements	Deep soils		< 30 cm thickness
	Clayey soils	Loamy soils	Sandy soils
	Drained or very drained	Drained	Poorly drained
	Firmly structured		Loose
	Fertile or very fertile		Unfertile
	Calcareous (Marls)	Calcareous	Non-calcareous
	Plains	Hill slopes	
	Relatively warm		Cold
	Open and elevated position		
Cultivation requirements	No clear indications		

Data Analysis and Results for Land Use

The analysis provides an archaeological evaluation of land use to define a potential adaptability of soils of Tarquinia area to various types of land uses and to the recognized major cereal crops, like spelt or cereals that demand requirements other than those of spelt.

Data available, reconstructed or estimated by environmental, historical and archaeological information are used. Land requirements needed to cultivate specific crops are calculated by some predictive models developed to infer the ancient soils adaptability.

Models used a set of synchronic thematic maps from present to past for any single chronological step. Ancient sources, the studies carried out so far and any added element are used to complement the analysis on soils and their requirements, up to build tables and maps on LUR, LUQ, LUT (Bedini, 1997; Boerma, 1989; Cremaschi, 1992; Hunt, Malone, Sevink, & Stoddart 1990; Karmersmans, Loving, & Voorrips 1985; Mallegni, Rottoli, & Bruni 2001; Peroni, 1994; Van Joolen, 2003; Van Leusen, 2002; Renfrew & Bahn 1995; Rottoli, 1997).

The soils requirements suited to investigated crops are shown in the maps, in different colors. The analysis shows that the plateaux of La Civita

are placed at the center of a large area, including the coastal plain, the river valleys and the surrounding land, suitable without restrictions or partially suitable to the spelt specific culture as the coastal plain and the river valleys, while is partially suitable for other cereals as the belt surrounding the plains and the hills of the interior geographic system, including the Monti della Tolfa. Figure 2 and Figure 3 show the results.

Data Analysis and Results for Viewshed

The visibility analysis established which parts of the territory are visible and from which points, and which positions were strategic for territory control with its territorial hierarchies and its cultural and cognitive landscapes (Van Leusen, 2002; Wheatley, 1995).

The identification of the points used for visibility analysis into settlements is the first important step. In fact, settlements and urban centers were not surrounded by wall with watchtowers. The intervisibility between the three Monterozzi villages and the Pian della Regina on La Civita, emerged as cumulative viewshed of the territory visible from these settlements, can explain the settlements location on the plateau highest point. Also viewshed analysis confirmed central position of the protourban center located on the main

Figure 2. Map of soils suitable, partially suitable and not suitable for spelt cultivation

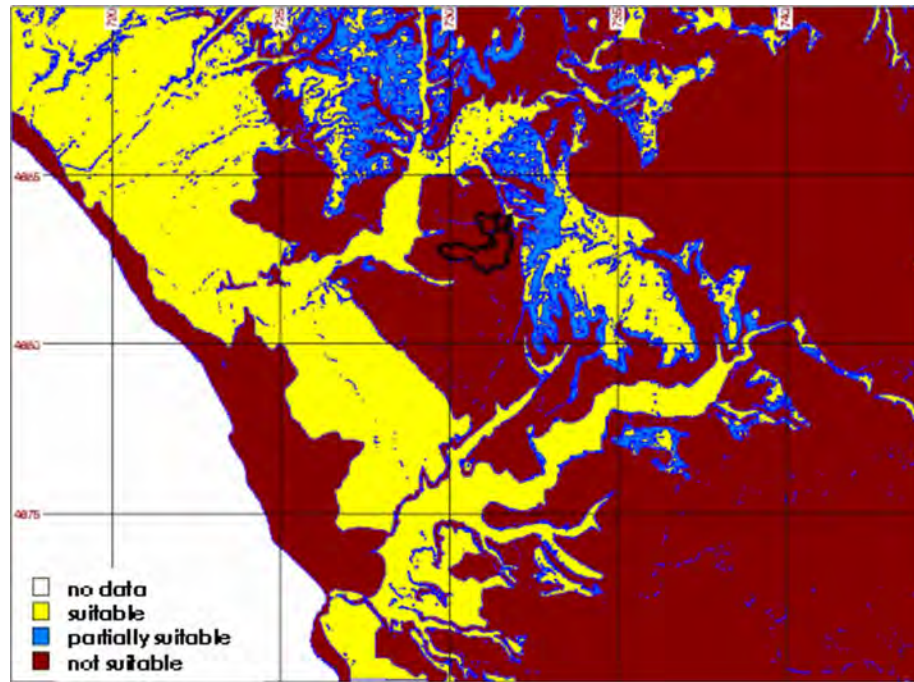
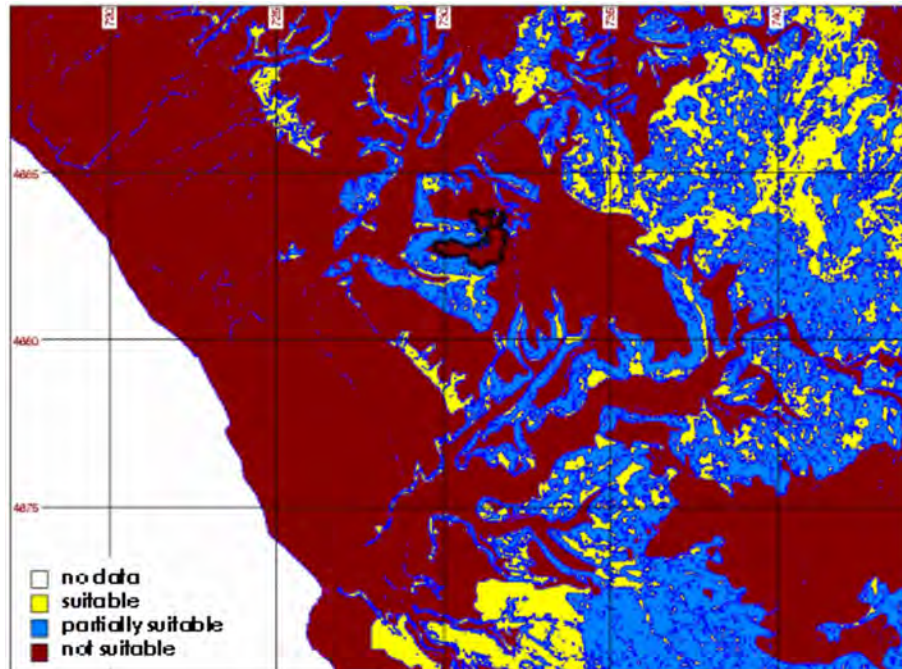


Figure 3. Map of soils suitable, partially suitable and not suitable for other cereals cultivation



plateaux of La Civita and demonstrated that from first Iron age settlements on Monterozzi hill it was possible to control all the coastal plain, the Mignone and Marta river valleys, the La Civita plateau (Mandolesi, 1999).

The viewshed analysis demonstrated also that the settlement location on La Civita plateau is further justified by its visual control on important areas like Marta valley, coastal plain, roads, Saline harbour (see Figure 4).

FUTURE RESEARCH DIRECTIONS

Archaeological GIS and Multi Agent Based Models (MABM) From Models to Archaeological Data

Although GIS based model used for reconstruction of ancient land use integrate well in a spatiotemporal uniform framework the available data, its approach based on analysis and interpretation on main important patterns for going back to the protourban processes, is static.

Integration of MABM simulations with archaeological GIS can allow to overcome GIS model limitations to better explain the dynamics of protourban processes.

So achievement of an investigation level more precise using computer simulations of dynamical evolution of the archaeological landscape is possible.

Tarquinia is an emblematic case study for South Etruria protocities origin, due to complexity of the settlement pattern starting from medium Bronze age, when small settlements coexist with large ones located on plateaux (Mandolesi, 1999) (See Figure 5).

The Multi Agent Based Models – MABM – analysis seems the best way for analyzing Tarquinia protocity dynamics and development between X-VIII centuries B.C., for discriminating between different hypotheses.

MABM simulations are necessary if the analysis purposes are a better interpretation of protourban processes dynamics, including paleo-agriculture and paleoproduction (Kohler, 2005). MABM simulations start from archaeological

Figure 4. Cumulative viewshed between La Civita, the coastal plain and the villages on the Monterozzi hill

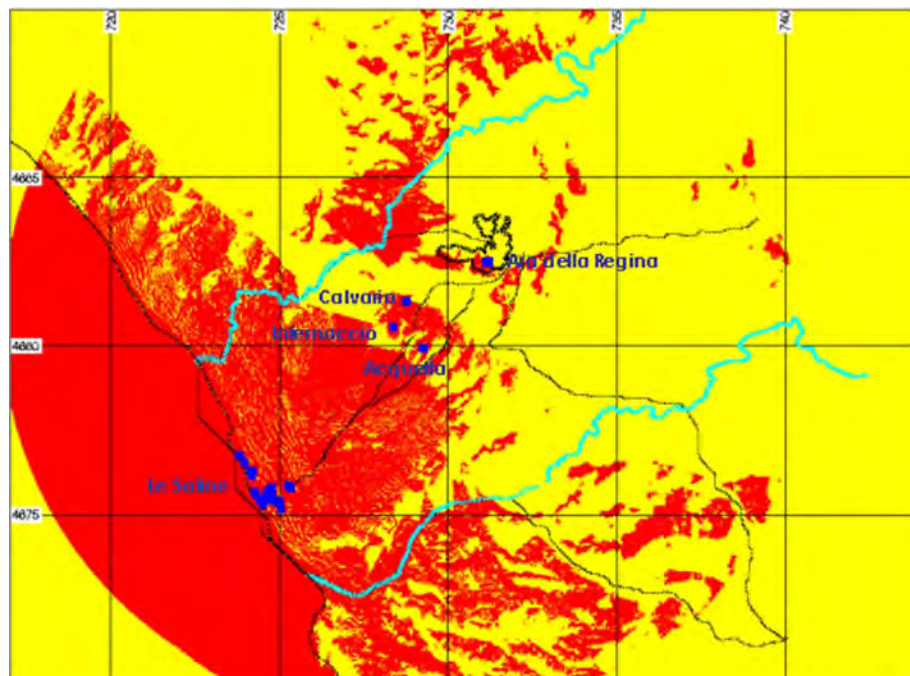
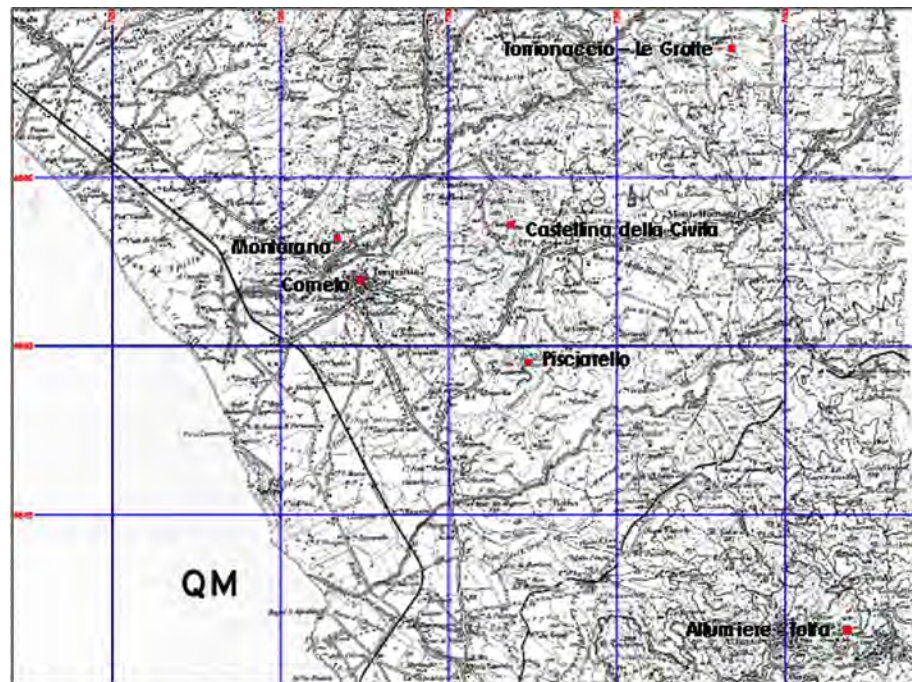


Figure 5. Preurban settlements of the older Bronze ages in the Tarquinia territory



and anthropological theories and models and its results are compared with archaeological records in an approach top to bottom.

The archaeological GIS integrated with MABM develops into a testbed laboratory for validation of models on social and cultural processes dynamics (Parisi, 2001).

CONCLUSION

Land Use Results Discussion

Also heavy clay soils of the Marta and Mignone rivers valleys, partially suitable for cereals, became cultivable only after deforestation, drainage and tillage.

Agricultural productivity was improved by intensive use of available technologies.

Agricultural practices and production increased for incentives caused by endogenous and exogenous factors that affected paleoproductivity indexes of various wheats (Ampolo, 1980).

These factors were a consequence of protourban center development.

In fact, the hypothesis that the people displacement on the plateaux, from the final Bronze age, was due to the availability of large adjacent amount of land suitable to cereal cultivations, is confirmed. Other factors affecting the settlements choice were favorable geographical and environmental conditions, due to large homogeneous districts of plateaux, to position near to sea and to two rivers, accessible but also well protected, and to well accessible woods and pastures.

For instance, the sites settled from the final Bronze age on the coastal plain like Fontanile delle Serpi in central position of land suitable to spelt cultivation and to grazing, meet similar criteria.

Viewshed Results Discussion

Viewshed analysis is very important for archaeological interpretation of protourban process and is relevant for many discussed aspects.

Viewshed analysis showed an easy control of the territory by the protourban center, but also by the settlements on Monterozzi hill. Control was very important during the deforestation of the coastal plain and river valleys. In fact, the agrarian practices of igniculture were used for obtaining soils suitable for agrarian activities in large areas. The control of fire diffusion during these practices was certainly very important to avoid dangerous consequences.

Other coastal and landing sites could play similar control of coast.

Visibility analysis helps to predict new sites not yet discovered.

In first Iron age, Saline settlement was the coastal plain control site, as before were the Monterozzi villages. Throughout Iron age, the Saline settlement controlled waterfront and Monterozzi villages will be abandoned. When Tarquinia later developed into City State, Monterozzi hill became, with Arcatelle necropolis, the Tarquinia larger burial area, integrating La Civita main town with Ara della Regina temple and coastal Gravisca and Le Saline settlements.

Overall Discussion

The major reasons for populating plateaux were the need of land accessible and suitable to cultivate cereals and of areas for building households. In late final Bronze age motivations for making cultivable the fertile lands represented by the wetlands of the coastal plain and river valleys can be explained by social and economical changes.

In fact, the Fontanile delle Serpi site of final Bronze age, located in the alluvial coastal plain on the edge of an ancient lagoon, was created in relation with intensive spelt cultivation on the coastal alluvial soils, and the Saline first Iron age was important site for the exploitation of the coastal plain resources. Agrarian surplus production is linked with growing demand due to increasing exchanges and to demographic development in the villanovan Tarquinia between X- VIII centuries B.C. The subsistence economy

of early villages located on the hills and river terraces, based on collective land ownership, was replaced by agricultural soils intensive exploitation to a greater extent, connected with new land ownership by households from final Bronze age (Bartoloni, 2002).

The protourban centers growth is a dynamic process carried forward from transformations in social structure and in production organization, compared with preurban clans organization. Villanovan society organized in households landowners reacted to market expansion in Mediterranean area. Demand of commodities like cereals and wheat increased for demography growth and for development of Mediterranean global market.

The settlement evolution shows the transition from resources management in catchment area plenty for the subsistence preurban villages to resources growth linked to the increasing scale of economic activities for emerging protourban centers.

The problem of protourban centers is how to produce a larger amount of agrarian products on a wider territory with more productive methods. For such reason new social organization linked to intensive agrarian production is based on private agrarian ownership. This new structure made possible a more effective work organization and a more rational resource management by deployment and concentration of the productive forces and ways of production on larger scale.

This pattern is reflected from the household clusters divided in *curiae*, according to the future model of the archaic Rome, using cooperation and coordination for managing a growing economy (Carandini, 2003).

In this context, plateaux work as possible territorial attractors: that is, places where households converged to favor common activities in agrarian works as tillage with igniculture and economical exchanges, both for external and internal markets in protourban centers, where material and human resources were concentrated.

With the synoecism, whatever bonds, internal and external to the villages, could favor people

concentration on same area, for supplying new economical requests and transformations. Demographic growth was an important factor for the production increase but also is the consequence of such increase: so that demographic growth has to be intended as cause and effect of the villanovan society changes.

These conclusions show the relevance of the GIS and of its results on ancient land uses, for selecting between the theories on the origin of the protourban centers (Bartoloni, 2003).

REFERENCES

- Ampolo, C. (1980). La formazione della città nel Lazio. Le condizioni materiali della produzione. Agricoltura e paesaggio agrario. *Dialoghi di Archeologia*, 1(2), 15–46.
- Bartoloni, G. (2002). *La civiltà villanoviana all'inizio della storia etrusca*. Roma, Italy: Carocci.
- Bartoloni, G. (2003). *Le società dell'Italia primitiva. Lo studio delle necropoli e la nascita delle aristocrazie*. Roma, Italy: Carocci.
- Bedini, E. (1997). I resti faunistici. In M. Bonghi Jovino & C. Chiaramonte Treré (Eds.), *Tarquinia. Testimonianze archeologiche e ricostruzione storica: scavi sistematici dell'abitato. Campagne 1982-1988* (Vol. 1-2, pp. 103–141). Roma, Italy: Tarchna.
- Bietti Sestieri, A. M. (1996). *Protostoria. Teoria e pratica*. Roma, Italy: La Nuova Italia Scientifica.
- Bietti Sestieri, A. M. (2002). L'agricoltura in Italia nell'età dei Metalli. In G. Forni & A. Marcone (Eds.), *Storia dell'agricoltura italiana. L'età antica. La preistoria* (Vol. 1-2, pp. 205–217). Firenze, Italy: Polistampa.
- Boerma, J. A. K. (1989). Land evaluation in prehistoric perspective: some observations. In Haex, O.C.M., Curvers, H.H., & Akkermans, P.M.M.G. (Eds.), *To the Euphrates and beyond* (pp. 17–28). Rotterdam, The Netherlands: CRC Press.
- Braudel, F. (2002). *Il Mediterraneo. Lo spazio la storia gli uomini le tradizioni*. Milano, Italy: Bompiani.
- Carandini, A. (2003). *La nascita di Roma. Dei, Lari, eroi e uomini all'alba di una civiltà* (Vol. 1-2). Torino, Italy: Einaudi.
- Colonna, G. (2000). I caratteri originali della civiltà etrusca. In M. Torelli (Ed.), *Gli Etruschi, Catalogo della mostra, Venezia, Palazzo Grassi* (pp. 25–41). Milano, Italy: Bompiani.
- Cremaschi, M. (1992). Insediamenti e sussistenza nella media età del Bronzo, Italia settentrionale. *Rassegna di Archeologia*, 10, 145–188.
- Cremaschi, M. (2000). *Manuale di geoarcheologia*. Roma-Bari, Italy: Laterza.
- Di Gennaro, F. (1986). *Forme d'insediamento fra Tevere e Fiora dal Bronzo Finale al principio dell'età del Ferro*. Firenze, Italy: Istituto di Studi Etruschi e Italici.
- Forni, G. (2002). L'agricoltura: coltivazione ed allevamento. Genesi, evoluzione, contesto. In G. Forni & A. Marcone (Eds.), *Storia dell'agricoltura italiana. L'età antica. La preistoria* (pp. 7–157). Firenze, Italy: Polistampa.
- Giardino, C. (1998). *I metalli nel mondo antico. Introduzione all'archeometallurgia*. Roma-Bari, Italy: Laterza.
- Guidi, A. (1998). The emergence of the state in Central and Northern Italy. *Acta Archaeologica*, 69, 139–161.
- Hencken, H. (1968). *Tarquinia, Villanovans and Early Etruscans. American School of Prehistoric Research Bulletin*, 23.
- Hunt, C., Malone, C., Sevink, J., & Stoddart, S. (1990). Environment, soils and early agriculture in Apennine Central Italy. *World Archaeology*, 22(1), 34–44. doi:10.1080/00438243.1990.9980127

- Iaia, C. (1999). *Simbolismo funerario e ideologia alle origini di una civiltà urbana. Forme rituali nelle sepolture "villanoviane" a Tarquinia e Vulci, e nel loro entroterra. Grandi contesti e problemi della protostoria italiana*. Firenze, Italy: All'Insegna del Giglio.
- Kamermans, H., Loving, S. H., & Voorrips, A. (1985). Changing patterns of prehistoric landuse in the Agro Pontino. In C. Malone & S. Stoddart (Eds.), *Papers in Italian Archaeology, IV, Part I, The Human Landscape* (pp. 53-68). Oxford, UK: BAR International Series 243.
- Kohler, T. A., Gumerman, G. J., & Reynolds, R. G. (2005). Simulating Ancient Societies. *Scientific American*, 293(1), 77-84. doi:10.1038/scientificamerican0705-76 PMID:16008305
- Linington, R. E. (1982). Il villaggio protostorico nella località Calvario sui Monterozzi a Tarquinia. In *Studi in onore di Ferrante Rittatore Vonwiller* (pp. 245-256). Como, Italy: Società archeologica comense.
- Liverani, M. (1988). *Antico Oriente. Storia società economia*. Roma-Bari, Italy: Laterza.
- Mallegni, F., Rottoli, M., & Bruni, S. (2001). I dati delle discipline consociate. In A.M. Moretti Sgubini (Ed.), *Tarquinia etrusca. Una nuova storia, Catalogo della mostra. Tarquinia, Museo Archeologico Nazionale* (pp. 59-61). Roma, Italy: L'Erma di Bretschneider.
- Mandolesi, A. (1999). *La "prima" Tarquinia. L'insediamento protostorico sulla Civita e nel territorio circostante. Grandi contesti e problemi della Protostoria italiana*. Firenze, Italy: All'Insegna del Giglio.
- Moscatti, S. (2001). *Civiltà del mare. I fondamenti della storia mediterranea*. Napoli, Italy: Liguori.
- Pacciarelli, M. (2000). *Dal villaggio alla città. La svolta protourbana del 1000 a.C. nell'Italia tirrenica. Grandi contesti e problemi della Protostoria Italiana*. Firenze, Italy: All'Insegna del Giglio.
- Pallottino, M. (1937). Tarquinia. *Monumenti Antichi della R. Accademia Nazionale dei Lincei*. 36.
- Parisi, D. (2001). *Simulazioni. La realtà rifatta nel computer*. Bologna, Italy: Il Mulino.
- Peroni, R. (1969). Per uno studio dell'economia di scambio in Italia nel quadro dell'ambiente culturale dei secoli intorno al 1000 a.C. *La Parola del Passato*, 125, 134-160.
- Peroni, R. (1994). *Introduzione alla protostoria italiana*. Roma-Bari, Italy: Laterza.
- Peroni, R. (1996). *L'Italia alle soglie della storia*. Roma-Bari, Italy: Laterza.
- Renfrew, C., & Bahn, P. (1995). *Archeologia. Teorie, Metodi, Pratica*. Bologna, Italy: Zanichelli.
- Rossiter, D. G. (1994). *Land Evaluation*. Cornell University, College of Agriculture & Life Sciences, Department of Soil, Crop & Atmospheric Sciences. Retrieved August 1994, from <http://www.css.cornell.edu/faculty/dgr2/docs/Scas494/s494ch1.pdf>
- Rottoli, M. (1997). I resti vegetali. In M. Bonghi Jovino & C. Chiaramonte Treré (Eds.), *Tarquinia. Testimonianze archeologiche e ricostruzione storica: scavi sistematici dell'abitato. Campagne 1982-1988* (Vol. 1-2, pp. 92-96). Roma, Italy: Tarchna.
- Sassatelli, G. (2009). Gli Etruschi e la loro economia. In V. Castronovo (Ed.), *Storia dell'economia mondiale. Dall'antichità al Medioevo. Dal Neolitico agli albori del I millennio* (Vol. 1-2, pp. 99-116). Roma-Bari, Italy: Laterza.
- Trigger, B. G., Kemp, B. J., O'Connor, D., & Lloyd, A. B. (2000). *Storia sociale dell'antico Egitto*. Roma-Bari, Italy: Laterza.
- VanJoolen, E. (2003). *Archaeological land evaluation. A reconstruction of the suitability of ancient landscapes for various land uses in Italy focused on the first millennium B.C.* (Doctoral dissertation). Retrieved from <http://dissertations.ub.rug.nl/faculties/arts/2003/e.van.joolen/>

Van Leusen, M. (2002). *Pattern to process. Methodological investigations into the formation and interpretation of spatial patterns in archaeological landscapes* (Doctoral dissertation). Retrieved from <http://dissertations.ub.rug.nl/faculties/arts/2002/p.m.van.leusen>

Wheatley, D. (1995). Cumulative viewshed analysis: a GIS-based method for investigating intervisibility, and its archaeological application. In G. Lock & Z. Stancic (Eds.), *Archaeology and Geographical Information Systems: A European Perspective* (pp. 171–186). London: Taylor & Francis.

Conolly, J. (2006). *Geographical Information Systems in Archaeology*. Cambridge, NY: Cambridge University Press. doi:10.1017/CBO9780511807459

Harvey, F. (2016). *A Primer of GIS. Fundamental Geographic and Cartographic Concepts* (2nd ed.). New York, NY: Guilford Press.

Langran, G. (1993). *Time in Geographic Information Systems*. Washington, DC: Taylor & Francis.

Wurzer, G., Kowarik, K., & Reschreiter, H. (Eds.). (2014). *Agent-based Modeling and Simulation in Archaeology*. Charm, Switzerland: Springer.

ADDITIONAL READING

Allen, M. S., Green, S. W., & Zubrow, E. B. W. (Eds.). (1990). *Interpreting Space: GIS and Archaeology*. London, England: Taylor & Francis.

Bivand, R., Pebesma, E., & Gómez-Rubio, V. (2008). *Applied Spatial Data Analysis with R*. New York, NY: Springer.

Christiansen, J. H., & Altaweel, M. R. (2006) Understanding Ancient Societies: A New Approach Using Agent-Based Holistic Modeling. *Structure and Dynamics: eJournal of Anthropological and Related Sciences Vol. 1, No. 2, Article 7*. Retrieved from <http://repositories.cdlib.org/imbs/socdyn/sdeas/vol1/iss2/art7>

KEY TERMS AND DEFINITIONS

Historical Ecology: Ecology of the historical natural environment.

Landscape Archaeology: Archaeology of ancient landscapes.

Multi-Agent Based Models: Simulation with multi-agent based models for the analysis of ancient social systems dynamics.

Paleoagriculture: Ancient agrarian techniques and crops.

Protourban Center: Urban settlement before the City State.

Synocism: Collective population transfer from one site to another.

Territory Control: Ways and geographical places for territory control.

Viewshed: Territory area visible by selected geographical points.

Exploring Tourism Cluster in the Peripheral Mountain Area Based on GIS Mapping

Ya-Hui Hsueh

National Taichung University of Education, Taiwan

Huey-Wen Chuang

National Taichung University of Education, Taiwan

Wan-Chiang Hsieh

National Taichung Girls' Senior High School, Taiwan

INTRODUCTION

Increasingly, developing tourism and making linkages with other economic sectors are strategically as a tool for regional economic development. In this article we will briefly resume and discuss the main results in the field of linkage between tourism and agriculture. In particular, we will survey the features of research area field data based on extensions of the linkage between tourism and agriculture on the several small agricultural settlements in Taiwan, which are the most relevant for mainstream application development. Finally, we will survey the currently available implementations of tourist spots by specified their locations on environmental conditions. Tourism in peripheral area is a relatively vulnerable to environment because of limited resource and disaster possibility. However, tourism provides alternative employment choice and greater economic inflexibility, so in peripheral area tourism development is still welcomed based on economic benefits. Clarifying the locational feature of the tourist spots is to realize the condition of tourism cluster, especially focus on the influencing factors of terrain conditions and accessibility. This research aims to explore the tourist attractions of a peripheral mountain area in the central Taiwan, analyze the important influenc-

ing locational factors of tourism cluster based on GPS (Global Positioning System) data of a set of tourist spots by processing GIS (Geographical Information System or Geographic Information System) mapping, and finally clarify what impacts the tourism cluster to the local environment by DTM (Digital Terrain Model) data.

BACKGROUND

The potential for creating synergistic relationships between tourism and agriculture has been widely recognized by development planners, policy makers. Commonly, economic leakage is the main factor for why the linkage of tourism and agriculture to promote local economic development in peripheral regions. Food is an essential component of tourism and also represents a significant part of tourism expenditure, and creating and strengthening the linkages between tourism and local food production sectors can provide a proximate market. The concept of “farm-to-fork” demonstrates the linkages between the dimensions of sustainable agriculture, sustainable cuisine and tourism by increasing demand for local products. Farm-to-fork concept also can lead to a range of related direct and indirect tourism activities such

as food festivals, farm visits, factory tours and souvenir food merchandise, thus further enhancing the benefits to the local (Berno, 2011).

Torres (2002) explores the linkages between tourism and agriculture in the Yucatan Peninsula, observes that the principal force driving hotel purchasing hotel food differences by tourist nationality and type of tourist based on tourist food consumption and preferences. So, in the Yucatan Peninsula Mexican foods, tropical fruits and organic produce are identified as this study area significant potential tourist food for linking tourism and local agriculture. Seaton (1999) examines tourism attraction in a peripheral region, identified critical success factors to small scale sustainable development, indicated that peripheral areas, distance from core areas with sparse populations and low GDP economic structure can motivate visitors to through some kind of special attraction such as book town for retailing. Gardiner and Scott (2014) investigates successful tourism cluster on the Gold Coast, Australia to develop the youth tourism market, through joint promotion and product development as an attractive destination. They propose that niche tourism clusters are often used to improve competitiveness to achieve economic advantages, through use of concepts of strategic alliances and networks. Cluster is therefore a fundamental factor to creating successful tourism industry within a destination zone.

Food supply chains of tourism accommodation providers in the coastal region of KwaZulu-Natal, South Africa, make the pro-poor tourism build of linkages between tourism and agriculture as a whole, and that revealing significant implications of tourist food consumption on destinations. Most research examining tourism and agriculture linkages has focused on hotel food procurement patterns while failing to address the main driving force of hotel purchasing tourist food based on tourist consumption and preferences (Pillay & Rogerson, 2013). The role of tasting room in the direct marketing of southwest Michigan wines educates visitors about Michigan wines, a way to differentiate the Michigan wines, and offers

spectacular views of fields, orchards is varied according to the different locations of the wineries. Commonly, these wineries all provide supporting services of cellar door sales and direct shipping for promoting local wine (Che & Wargenau, 2011).

The linkages between tourism and related sectors vary widely from nation to nation, region to region. There are numerous successful examples of strategic alliances between tourism and the agricultural related sectors in Taiwan, especially the Jhong-Liao Township, in Nantou County of central Taiwan. Jhong-Liao Township composes of several small agricultural settlements with a population of about 16000, is located on the elevation from 200 meters to 1264 meters. Jhong-Liao is not the most famous destination in Taiwan, but it is a popular mystery destination for its distant from urban area-Taichung City. During Ching Dynasty it was a stopping place of wildness space along a transport route from costal areas to mountainous areas of central Taiwan rather than a tourist spot. During Japanese Occupation Period (1895-1945) it was a high productive place of agricultural activity, planted in plenty of banana, so the reputation for Jhong-Liao is also called "Banana Mountain". 20 years ago, a visitor to the Jhong-Liao seemed to be a lonely trip into an empty mountain area, until 1999 there were almost no tourist spots in Jhong-Liao. It's famous for as a tourist attraction was traced to 1999, when an earthquake disaster dramatically destroyed here, an atmosphere of "Ghost Town" spread out in Taiwan. Since 1999 many tourists visited here to see the destroyed landscapes after earthquake, and the number of tourist spots in Jhong-Liao has increased to 70, so now it is transformed to one of most popular destination of Nantou County.

In recent years, the linkage between tourism and agriculture for environmental sustainability has been discussed as the development of alternative tourist food, which was perceived to be traditional and local agrarian products, appeals to the visitor's desire for authenticity within the holiday experience (Hall, 2005; Michael, 2007; Sims, 2009). Thus, tasting local food performs as an authentic

tourism experience that transforms traditional agriculture products to create cuisine culture of the destination. The significant implications of tourist food consumption on destinations have received a growing research interest recently (Telfer & Wall, 2000; Torres, 2002, 2003; Torres & Momsen, 2004, 2011; Pillay & Rogerson, 2013), and this means the linkage between tourism and agriculture has play important role in tourism development because tourism seems to help farmers to overcome economic growth constraints, upgrade and face new competitions in distant markets. Exploring the tourist consumption preferences is critical to the analysis of linkages between tourism and agriculture for creating either backward linkages to local agriculture or increased economic growth. Tourist consumption and demand for food vary greatly according to a number of factors, including tourist preference and type of tourist. Local food may be particularly popular with tourists because of its consideration of “iconic” products that capture the “typical” nature of a particular place. Local food is popular because it is associated with a host of value, and souvenir purchase is making local food as a symbol of place and culture (Sims, 2009).

GIS MAPPING WAY

The concept of tourism cluster has attracted attentions during the past decades, both as descriptive of an increasingly important strategy and an effective tool to enhance economic development in peripheral regions. This research used empirical data to explore the linkages of tourism, agriculture and religion for different implications both in terms of relevant evidence and the scope for promotional policies. In order to realize how many tourist spots clustering in this study area, this research locates tourist spots based on GPS data to collect data, and then specified all the points to raster layer of TWD 97 coordination system on GIS processing system. With elevation analysis, slope analysis, point density analysis, buffer analysis and overlay

analysis of GIS mapping way, this research aims to distinguish the significant association between the core clustering area of tourist spots and the influential factors such as environmental ones and accessibility. Finally, we conduct in-depth interviews of the holders of organic farmers, agricultural officers in government and the travelers visiting to Jhong-Liao, and defined there are 72 the total number of tourist spots, including:

- 38 religious temples,
- 14 organic leisure farms,
- 7 artificial landscapes,
- 6 natural sceneries,
- 5 mountain tracks, and
- 2 historical heritages.

Terrain Analysis of GIS Mapping

Hsueh (2013a) used terrain analysis of GIS mapping to specify a set of points based on the accommodations of a mountain area to compare the different percentages of accommodations on the different terrain conditions. Hsueh (2013b) also used terrain analysis of GIS mapping to explore the significant clustering area and locations of mushroom huts based on different terrain conditions in an agricultural mountain area. Hsueh (2015) analyzed the neighborhood of clustering area of a set of points of tea firms located on a plat form terrain by using terrain analysis of GIS mapping combined of related layers to process multiple buffer analysis to define the significant core area of tea firms clustering. This research aims to process terrain analysis to display the elevation and slope conditions of tourism cluster of tourist spots in the Jhong-Liao area. Jhong-Liao is a productive area of organic agricultural products in the central Taiwan. There are 14 organic farms in Jhong-Liao, thus Jhong-Liao has the advantage of tourist foods for its available for a lot of local organic foods, the most popular and favorite foods in Taiwan. As the representative agricultural products of oranges, longyans, bananas, it is obvious that Jhong-Liao has a lot of organic agricultural products, which

have converted to tourist foods to meet tourist consumption and preferences recently.

With many farmers switching to organic products of fruits, consolidation occurred in the organic agriculture as the number of Farm Alliance of Jhong-Liao increases gradually. It was not until 1999 that the He-Xing Organic Cultural Association created the “Original and Organic Cuisine” (OOC) whose purpose is to start to control the agricultural productive environment as well as ecology conservation and making health preserving cuisine as tourist foods based on local products. The linkages between tourism and agriculture become more apparent when we analyze what kinds of tourist foods and experiences tourists are seeking during their holiday. By telling the story of “vegetable dyes” in Jhong-Liao, it is possible to use the long scarf DIY activity of making tourist’s souvenir for authenticity to encourage the tourism development based on local agricultural products that will boost environmental sustainability and economic benefits. The experience of vegetable dyes with a particular focus on tourist attraction in Jhong-Liao is obtained from different local plant sources—roots, berries, bark, and leaves. Typically, the dye material is put in a pot of water and then the textiles tied with special techniques to be dyed are added to the pot, which is heated and stirred until the color is transferred. Without use of chemicals, the color of the vegetable dye is brilliant, permanent and with creative diversified local patterns, which is due to the different local plant sources such as:

- Betel nuts, Litchi, Longyan, and Tungoil Tree creating into light red shades;
- Cape Jasmine, Common Garcinia, kunyit, and onion into yellow shades;
- Chinese ixora, Subcostate Crape Myrtle, Kassod tree into green shades;
- Litchi, Longyan, rhus chinensis, and onion into brown shades.

By the interactions of cultivating technology of organic agricultural products through those

professional associations’ promotions, the tourist food reputation of Jhong-Liao thus enhanced. With their members concentrated in the same area, those professional associations are easy to share and offer professional knowledge to help the farmers in Jhong-Liao to establishment organic brand and organic certification. The tourism cluster of tourist spots strengthened the ties between tourism and agriculture providing for high quality and productive network of tourist food. There are a lot of agriculture related and supporting leisure farms co-located within the Jhong-Liao area to take an advantage based on close working relationships, and the cluster of leisure farms offers an empirical research of how tourism related sectors create competitive advantages. Organic farm holders, for instance, interact regularly with agricultural associations on tourist foods processing technologies and learn about how to promote tourist foods to visitors.

The notion of a new form of niche tourism has gone further than ever before. No longer is improved health on holiday or an escape from the routine work and the movement to a place with a cleaner climate, the rise of niche tourism has become the central theme of tourism in an active rather than a passive sense. The tourism literatures suggest there are different types in tourist attractions based on various characteristics of tourist spots. The important characteristics identified in the literature are tourist origin (Telfer & Wall, 1996; Telfer, 2001) and type of tourist (Sharkey & Momsen, 1995; Momsen, 1998; Telfer & Wall, 1996). Types of tourists range from the highly organized conservative mass tourist to the more flexible independent tourist. Most visitors who have the travel motivations to Jhong-Liao is triggered by religion, namely these visitors are flexible independent pilgrims by religion tourist attractions.

These diversity characteristics of the tourist spots locations enable comparisons based on elevation and slope conditions whether ‘suitable’ or ‘unsuitable’ tourist spots. Table 1 and Figure 1 show that all the total 72 tourist spots are near to

ivers and the popular elevation of tourist spots is located on an elevation < 300 meter accounting for 66% (48 tourist spots), 32% (23 tourist spots), while tourist spots on the area proximity to elevation > 600 meter is almost none. There are 27 tourist spots of 37% with a slope < 5%, 28 tourist spots of 39% with a slope 5%~15%, and 17 tourist spots of 24% with a slope 15%~30%, respectively (Table 2, Figure 2). Obviously, a flat slope is suitable for the management of tourist spots, beneficial to reach through an unfrequented mountain track for travelers, especially in a peripheral mountain area. Based on GIS mapping, we can realize terrain for tourist spots location is not only the main influencing factor in a peripheral mountain area, but also access to waters is becoming the important factor.

A cluster is a geographically proximate group of interconnected companies and associated institutions in a particular field, linked by commonalities and complementarities (Porter, 1990, p. 199). The local associations playing important role on Jhong-Liao's tourism cluster, such as Long-Yan-Lin Welfare Association, promotes the tourism image of North Jhong-Liao through the marketing network agriculture production through "organic brand" of the leisure farms; Farm Alliance of Jhong-Liao establishes banana market through the cooperation of banana farmers to promote Jhong-Liao's banana brand. Although the linkages of tourism and agriculture can result in higher levels of economic growth instead of becoming poor agrarian producers, it also contributes significantly to environmental impacts. There are 18 tourist spots of 25% with a slope < 5%, 31 tourist spots of 43% with a slope 5%~15%, and 23 tourist

spots of 32% with a slope 15%~30%, respectively (Table 3, Figure 3).

Apparently, several tourist spots located on unsuitable terrain condition because of its easy suffered natural disasters if there is heavy storm rain or typhoon. In Jhong-Liao, the high density clustering areas of tourist spots is apparent on the location with steep slope of slope > 5%, where is inclined to have debris flow disaster and slope land slide or collapse. Environmental control and resources conservation are essential missions for Jhong-Liao sustainable development on the tourism clustering process. The high-density clustering areas of tourist spots is apparent on the location with steep slope of slope > 5%, where is inclined to have debris flow disaster and slope land slide or collapse. Thus, slope factor may be an essential influencing factor for tourism development in peripheral mountain area.

Point Density Analysis of GIS Mapping

He and Gebhardt (2012) explore the spatial clustering of creative industries based on point density analysis of GIS mapping by locating the sites of creative industry parks to visualize the core area of creative industries cluster in Shanghai. To explore the tourism cluster of Jhong-Liao, this research also used point density analysis of GIS mapping to locate a set of points of the total 72 tourist spots, specified all the points of identified tourist spots' locations of TWD 97 coordination system. The obtained density surface of tourist spots distribution of Jhong-Liao by GIS mapping is essential to define the concentration center of tourism cluster.

Table 1. Number of tourist spots on different elevation condition

Elevation (Meters)	< 300	300 ~ 600	600 ~ 900	900 ~ 1200	1200 ~ 1500
Number of tourist spots	48	23	1	0	0
(Percentage)	66%	32%	2%	0%	0%

Table 2. Number of tourist spots on different slope condition

Slope	< 5%	5% ~ 15%	15% ~ 30%	30% ~ 40%	40% ~ 55%
Number of tourist spots	27	28	17	0	0
(Percentage)	37%	39%	24%	0%	0%

Figure 1. Spatial distribution of tourist spots -based on elevation condition

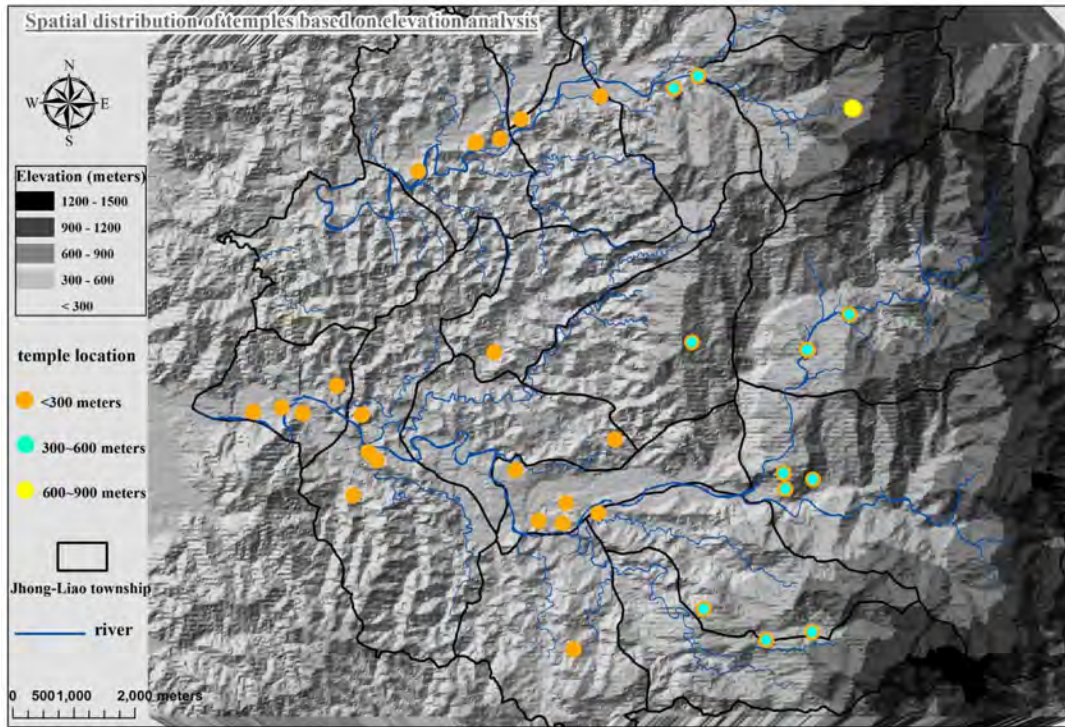


Figure 2. Spatial distribution of tourist spots - based on slope condition

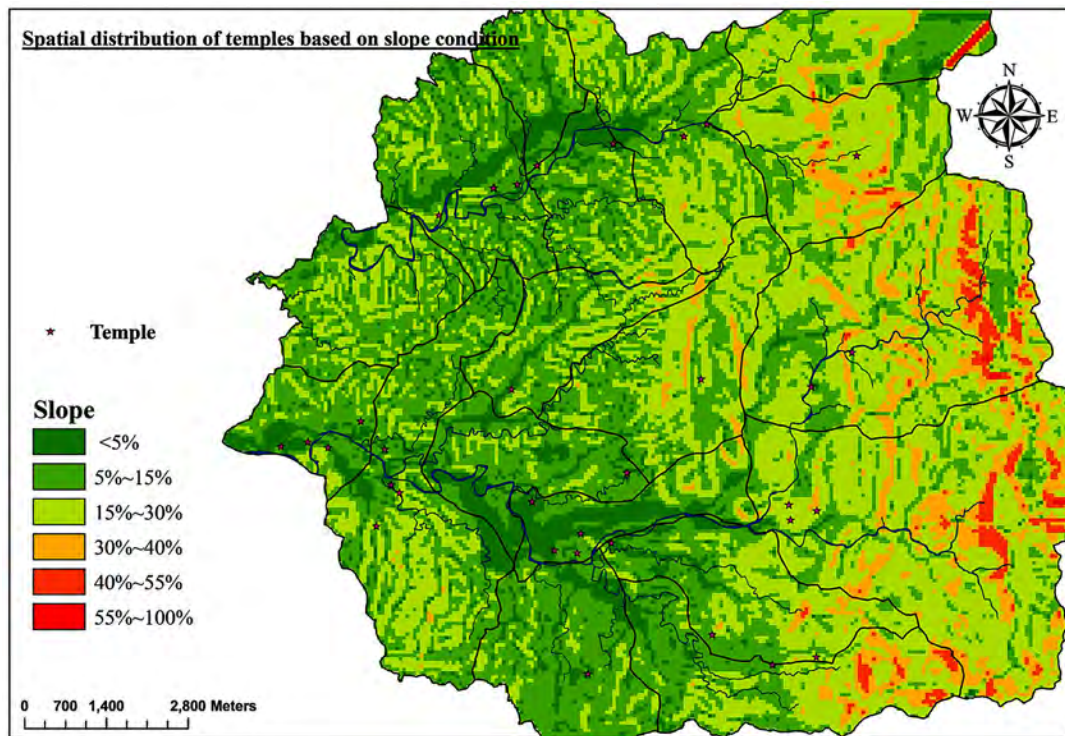


Table 3. Number of tourist spots based on slope condition

Slope	< 5%	5% ~ 15%	15% ~ 30%	30% ~ 40%	40% ~ 55%
Number of tourist spots	18	31	23	0	0
(Percentage)	25%	43%	32%	0%	0%

The calculated method of point density value for comparing the extents of clustering of a raster surface is dependent on the estimating cell size and searching radius. On a point density surface, individual cell value is calculated by the specific point that falls within the searching radius. This research used 300 meter x 300 meter of cell size and 500 meters of searching radius to process the quantity of points per cell of each searching radius. By using point density analysis, we calculated the density value to explore the spatial concentration trends of tourist spots, and define the most concentrated zones of point locations of tourist spots, namely to distinguish the significant clusters of spatial distribution of tourist spots. With the point density surface of locations of the tourist spots, many layers such as DTM terrain raster, the transport raster, and the settlement raster can be overlaid together, and then the influencing factors of suitable locations accounting for tourism cluster render to be demonstrated.

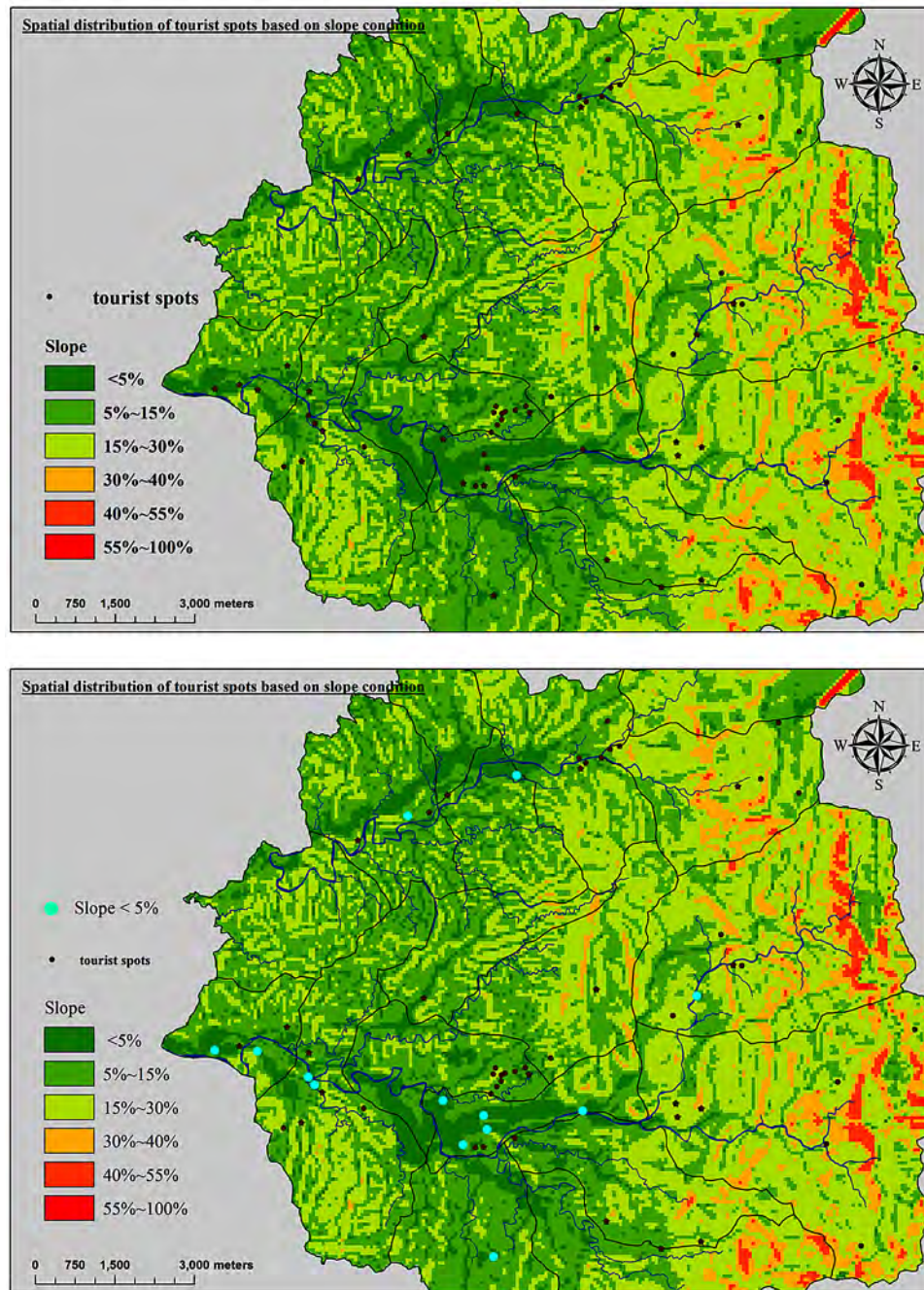
In peripheral mountain area continuous monitoring on negative effects of tourism should be carry out and vulnerable sites should either be protected or visitors should be guided away from them. Micro-clusters are visible in some forms of tourism, where groups of tourist spots co-located and interact to deliver particular kinds of outputs, revealing that cluster is not confined solely to metropolitan areas but occurs also in peripheral mountain area. Hsueh (2012) used Porter's concept of cluster co-location for the example of tourism cluster by GIS mapping to locate a set of points of tourist spots along a route, clarified the core-periphery relationship among the tourist spots due to accessibility, and illustrated the core area

of tourism clustering due to the increasing tourist attraction of destinations drawing more and more visitors, while the relative peripheral area is undergoing marginalized through the spatial competition process. The conception of tourism cluster creating tourist attraction is also demonstrated by the antiques industry clustered in a mountain area as a trip generator and supporting sustainable form of economic growth (Michael, 2002).

Based on the point density analysis (searching radius 500 meter, cell value 300x300 meters), obvious cluster cores of tourist spots are identified (Figure 4). The areas with the most high point density value of between 10.18591615~12.73239517 are located on the west northern part and the west southern part of Jhong-Liao. Figure 4 also shows that accessibility is also an influential factor on the locations of tourist spots clustering. Apparently, the west southern part of Jhong-Liao, with the highest point density value, is not only due to the geographical proximity to urban area-Nantou City for providing more tourism related services but also convenient to leave Jhong-Liao through Route 139 to back home. Obviously, the representative tourism cluster districts, in the west southern part and west northern part of Jhong-Liao, agglomerate along the main linkages on the Route 139 and Route 22 reveals that accessibility still has a decisive role in terms of location selection.

Geographical proximity characterizes tourism cluster as collaborative network with strong competitiveness and significant opportunities to economic development, so the collaborative networks were especially important for Jhong-Liao's tourism development. Establishing a stable network of producers and generating external economies are also the premises of the local associations for promoting the tourism development for sustainability. Why tourist spots cluster along the main route – Route 22 and Route 139 of Jhong-Liao? Along a route is suitable for tourist spots because of visitors to access a tourist spot and the related supporting services of community, especially in a peripheral mountain area. Apparently, tourist spots also clustered on the neighborhood of settle-

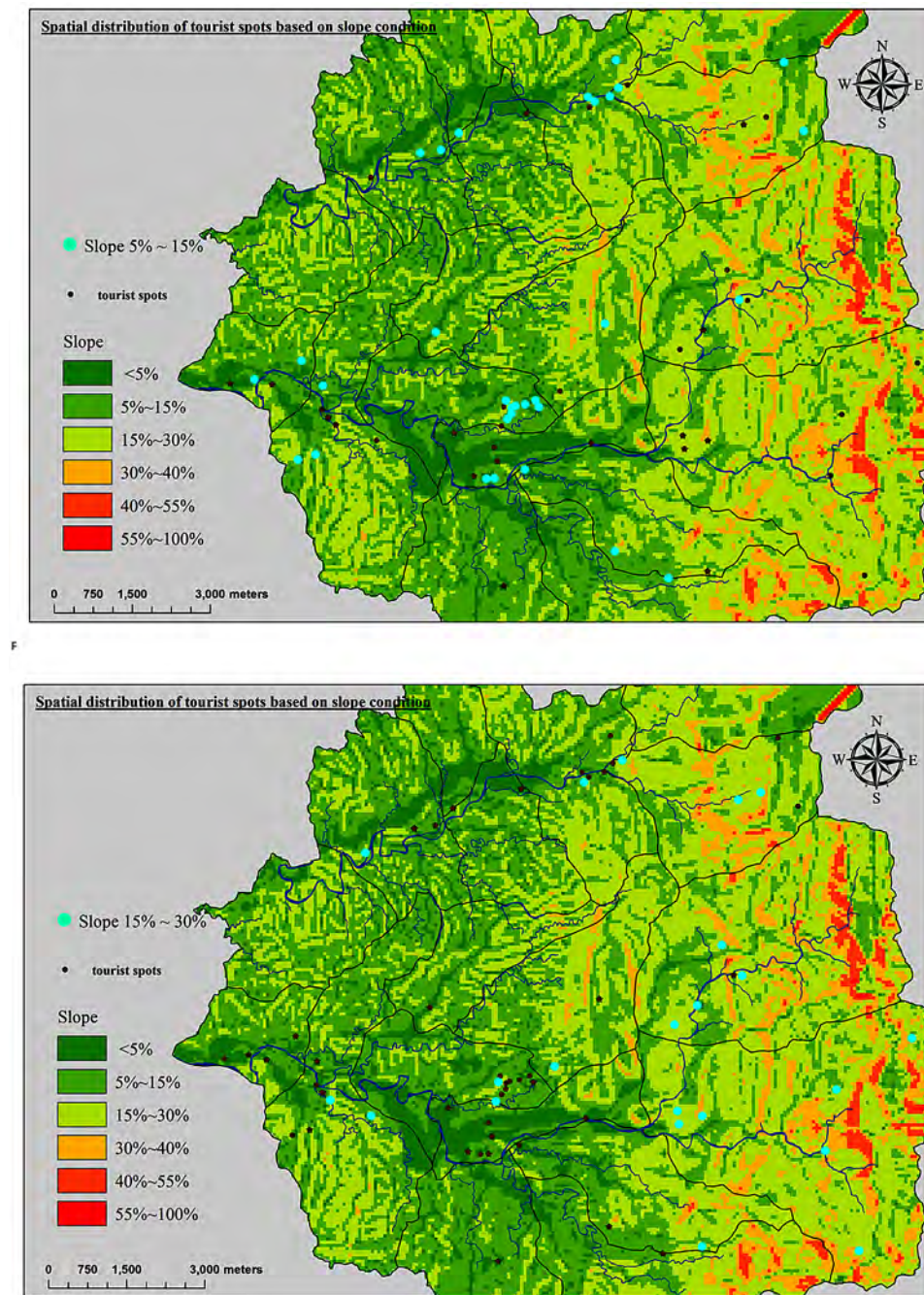
Figure 3a. Spatial distributions of the tourist spots based on slope condition



ments because settlement serve as the function of community by providing related supporting services, especially in a peripheral mountain area with confined activity space. There is considerable evidence that it is efficient for tourist spots next to community to access tourism services based on

point density analysis by GIS mapping. So, access to community may be the basic factor for Jhong-Liao's tourism cluster. There are several tourism related networks in the community between tourism associations and agriculture activities. The professional tourism associations tourism of

Figure 3b. Spatial distributions of the tourist spots based on slope condition



Jhong-Liao were established from 1999, such as Long-Yan-Lin Welfare Association, Dab-Hand Vegetable Dye Workshop, Xi-Di-Yao Agriculture Learning Garden, He-Xing Organic Cultural Association, Kan-Ding Cinnamon, Making Wine Specialized Garden, and Long-An Medicinal

Botanical Garden. One of the reasons for those professional associations playing important role on the process of tourism cluster is their creative knowledge on shaping tourist attractions and related support service networks between tourism and agriculture for segment market of niche tour-

ism. For example, there are common facilities for vegetable drying, fruit-picking tools shared by the small workshops or leisure farms or education providing by professional associations, which also offers agriculture information to members.

ment efforts in this direction by GIS mapping are required in the next further research.

G

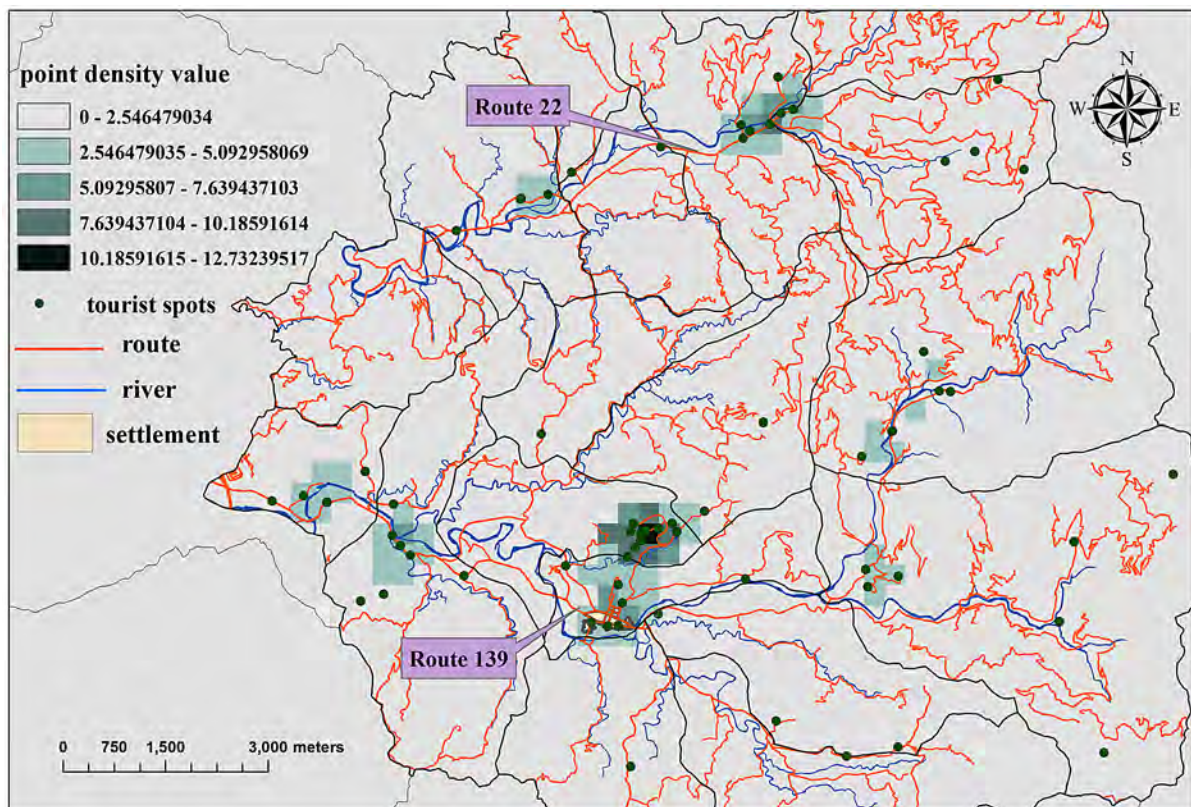
FUTURE RESEARCH DIRECTIONS

Future research should follow the direction of a location extension of spatial data base and analysis way in GIS mapping with advanced aspects in the linkage of tourism and agriculture implementations, including, for instance, which location is suitable for developing tourism based on agriculture products. Implementation of the tourist foods and their marketing integration with government sectors have not performed yet, so further pulling forces for the tourist foods marketing and develop-

CONCLUSION

In this article, we reviewed some of the most prominent theoretical results and practical achievements of research in the GIS mapping field. Although the ways of GIS mapping applied to the linkage between tourism and agriculture have been studied, we focused on the terrain analysis including elevation condition and slope condition and point density analysis, discussed their main features and surveyed their emerging implementations on tourist spots in the study area. Based on GIS mapping way, we realize the promoting of tourist attractions is not only by tourism cluster of tourist spots but also by linkages between tourism and

Figure 4. Point density analysis of tourist spots



related and supporting sectors such as agriculture. Linkage between tourism and agriculture makes a destination easier to shape its destination image, establish tourism activity network, and segment market. Most related literature reviews about tourism cluster focus on cluster benefits as the key elements of analyzing optimal location, but this research focuses on environmental impacts of tourism cluster for small scale tourist spots in a peripheral mountain area, reassesses the role of micro tourism cluster, like the leisure farms distinguished by organic agriculture and religious tourist spots, as tourist attractions of peripheral tourism. Instead of focusing on the benefits or patterns of tourism cluster, this research explores the environmental impacts of tourism cluster in a fragile mountain area, examines terrain factors of elevation and slope by GIS mapping way to realize how tourism cluster is influenced and further emphasize tourism cluster promoting peripheral area economic.

REFERENCES

- Berno, T. (2011). Sustainability on a Plate: Linking Agriculture and Food in the Fiji Islands Tourism Industry. In R. Torres, & J. Momsen (Eds.), *Tourism and Agriculture: New Geographies of Production and Rural Restructuring*, (pp. 87-103). London: Routledge.
- Botterill, D., Owen, R., Emmanuel, L. N., Foster, N. T., Gale, T., Nelson, C., & Selby, M. (2002). Perceptions from the Periphery: The Experience of Wales. In F. Brown & D. Hall (Eds.), *Tourism in the Peripheral Areas*, (pp. 7-38). Clevedon: Channel View Publications.
- Buhalis, D. (1999). Tourism on the Greek Islands: Issues of Peripherality, Competitiveness and Development. *International Journal of Tourism Research*, 1(5), 341–358. doi:10.1002/(SICI)1522-1970(199909/10)1:5<341::AID-JTR201>3.0.CO;2-0
- Chaperon, S., & Bramwell, B. (2013). Dependency and Agency in Peripheral Tourism Development. *Annals of Tourism Research*, 40(1), 132–154. doi:10.1016/j.annals.2012.08.003
- Che, D., & Wargenau, A. (2011). Visiting Winery Tasting Rooms: Venues for Education, Differentiation and Direct Marketing. In R. Torres, & J. Momsen (Eds.), *Tourism and Agriculture: New Geographies of Production and Rural Restructuring*, (pp. 192-204). London: Routledge.
- Connell, J. (2006). Medical Tourism: Sea, Sun, Sand and Surgery. *Tourism Management*, 27(6), 1093–1100. doi:10.1016/j.tourman.2005.11.005
- Eastman, C. M. (1995). Tourism in Kenya and the Marginalization of Swahili. *Annals of Tourism Research*, 22(1), 172–185. doi:10.1016/0160-7383(94)00069-5
- Gardiner, S., & Scott, N. (2014). Successful Tourism Clusters: Passion in Paradise. *Annals of Tourism Research*, 46, 163–184. doi:10.1016/j.annals.2014.01.004
- Getz, D. (1992). Tourism Planning and Destination Life Cycle. *Annals of Tourism Research*, 19(4), 752–770. doi:10.1016/0160-7383(92)90065-W
- Hall, C. M. (2005). Rural Wine and Food Tourism Cluster Network Development. In D. Hall, I. Kirkpatrick & M. Mitchell (Eds.), *Rural Tourism and Sustainable Business*, (pp. 149-164). Clevedon: Channel View Publications.
- He, J. L., & Gebhardt, H. (2012). Socio-Culture Embeddedness of Spatial Agglomeration of Creative Industries in Shanghai. *Sociology Study*, 2(8), 605–614.
- Hsueh, Y. H. (2012). The Influence of Transportation Revolution to Local Tourism Development-Example for Puli Area of Taiwan. *Sociology Study*, 2(8), 615–625.

- Hsueh, Y. H. (2013a). Exploring the Clustering Location of Accommodation Units through the Tourism Development in the Cing-Jing Area of Taiwan. *International Journal of Basic & Applied Sciences*, 13(4), 34–43.
- Hsueh, Y. H. (2013b). Mushroom huts diffusion analysis based on GPS data. In Y. H. Hsueh (Ed.), *Global positioning systems: Signal structure, applications and sources of error and biases*, (pp. 115–124). New York: Nova Science Publishers, Inc.
- Hsueh, Y. H., & Chen, C. H. (2015). Clustering of Tea Processing Industry Based on GIS Analysis. *International Journal of Agricultural Sciences and Natural Resources*, 2(3), 58–66.
- Marshall, J. (2001). Women and Strangers: Issues of Marginalization in Seasonal Tourism. *Tourism Geographies*, 3(2), 165–186. doi:10.1080/14616680010030266
- Michael, E. (2002). Antiques and Tourism in Australia. *Tourism Management*, 23(2), 117–125. doi:10.1016/S0261-5177(01)00053-X
- Michael, E. J. (2007). Micro-Clusters: Antiques, Retailing and Business Practice. In E. J. Michael (Ed.), *Micro-Clusters and Networks: The Growth of Tourism*, (pp. 63–78). London: Elsevier Publications.
- Momsen, J. (1998). Caribbean Tourism and Agriculture: New Linkages in the Global Era? In T. Klak (Ed.), *Globalization and Neoliberalism: The Caribbean Context*. Oxford, UK: Rowman & Littleman Publishers, Inc.
- Pillay, M., & Rogerson, C. M. (2013). Agriculture-Tourism Linkages and Pro-Poor Impacts: The Accommodation Sector of Urban Coastal KwaZulu-Natal, South Africa. *Applied Geography (Sevenoaks, England)*, 36, 49–58. doi:10.1016/j.apgeog.2012.06.005
- Prideaux, B. (2002). Building Visitor Attractions in Peripheral Areas: Can Uniqueness Overcome Isolation to Produce Viability? *International Journal of Tourism Research*, 4(5), 379–389. doi:10.1002/jtr.387
- Seaton, A. V. (1999). Book Towns as Tourism Developments in Peripheral Areas. *International Journal of Tourism Research*, 1(5), 389–399. doi:10.1002/(SICI)1522-1970(199909/10)1:5<389::AID-JTR204>3.0.CO;2-0
- Sharkey, D. A., & Momsen, J. (1995). Tourism in Dominica, West Indies: Problems and Prospects. *Caribbean Geography*, 6(1), 40–51.
- Sims, R. (2009). Food, Place and Authenticity: Local Food and the Sustainable Tourism Experience. *Journal of Sustainable Tourism*, 17(3), 321–336. doi:10.1080/09669580802359293
- Sindiga, I. (1996). International Tourism in Kenya and the Marginalization of the Swahili. *Tourism Management*, 17(6), 425–432. doi:10.1016/0261-5177(96)00051-9
- Telfer, D. J. (2001). Strategic Alliances along the Niagara Wine Route. *Tourism Management*, 22(1), 21–30. doi:10.1016/S0261-5177(00)00033-9
- Telfer, D. J., & Wall, G. (1996). Linkages between Tourism and Food Production. *Annals of Tourism Research*, 23(3), 635–653. doi:10.1016/0160-7383(95)00087-9
- Telfer, D. J., & Wall, G. (2000). Strengthening Backward Economic Linkages: Local Food Purchasing by Three Indonesian Hotels. *Tourism Geographies*, 2(4), 421–447. doi:10.1080/146166800750035521
- Torres, R. (2003). Linkages between Tourism and Agriculture in Mexico. *Annals of Tourism Research*, 3(3), 546–566. doi:10.1016/S0160-7383(02)00103-2

Torres, R., & Momsen, J. H. (2004). Challenges and Potential for Linking Tourism and Agriculture to Achieve Pro-Poor Tourism Objectives. *Progress in Development Studies*, 4(4), 294–318. doi:10.1191/1464993404ps092oa

Weaver, D. B. (1998). Peripheries of the Periphery Tourism in Tobago and Barbuda. *Annals of Tourism Research*, 25(2), 292–313. doi:10.1016/S0160-7383(97)00094-7

ADDITIONAL READING

Carvera, S., Comberb, A., McMorran, R., & Nuttera, S. (2012). A GIS Model for Mapping Spatial Patterns and Distribution of Wild Land in Scotland. *Landscape and Urban Planning*, 104(3-4), 395–409. doi:10.1016/j.landurbplan.2011.11.016

Gregoret, C., Degetto, M., & Boreggio, M. (2016). GIS-Based Cell Model for Simulating Debris Flow Runout on a Fan. *Journal of Hydrology (Amsterdam)*, 534, 326–340. doi:10.1016/j.jhydrol.2015.12.054

Haworth, B., Bruce, E., & Iveson, K. (2013). Spatio-Temporal Analysis of Graffiti Occurrence in an Inner-City Urban Environment. *Applied Geography (Sevenoaks, England)*, 38, 53–63. doi:10.1016/j.apgeog.2012.10.002

Kumarn, S., & Bansal, V. K. (2016). A GIS-Based Methodology for Safe Site Selection of a Building in a Hilly Region. *Frontiers of Architectural Research*, 5(1), 39–51. doi:10.1016/j.foar.2016.01.001

Lacher, R. G., & Nepal, S. K. (2010). From Leakages to Linkages: Local-Level Strategies for Capturing Tourism Revenue in Northern Thailand. *Tourism Geographies*, 12(1), 77–99. doi:10.1080/14616680903493654

Lee, S. H., Choi, J. Y., Yoo, S. H., & Oh, Y. G. (2013). Evaluating Spatial Centrality for Integrated Tourism Management in Rural Areas Using GIS and Network Analysis. *Tourism Management*, 34, 14–24. doi:10.1016/j.tourman.2012.03.005

Scaini, C., Felpetob, A., Martí, J., & Carniel, R. (2014). A GIS-Based Methodology for the Estimation of Potential Volcanic Damage and Its Application to Tenerife Island, Spain. *Journal of Volcanology and Geothermal Research*, 278–279, 40–58. doi:10.1016/j.jvolgeores.2014.04.005

Taramelli, A., Valentini, E., & Sterlacchini, S. (2015). A GIS-Based Approach for Hurricane Hazard and Vulnerability Assessment in the Cayman Islands. *Ocean and Coastal Management*, 108, 116–130. doi:10.1016/j.ocecoaman.2014.07.021

Zhai, S., Xu, X., Yang, L., Zhou, M., Zhang, L., & Qiu, B. (2015). Mapping the Popularity of Urban Restaurants Using Social Media Data. *Applied Geography (Sevenoaks, England)*, 63, 113–120. doi:10.1016/j.apgeog.2015.06.006

KEY TERMS AND DEFINITIONS

DTM: A digital representation of ground-surface topography or terrain represents the bare ground surface without any objects like plants and buildings. DTM is a topographic model of the bare earth-terrain relief that can be manipulated by computer programs. Vegetation, buildings and other man-made artificial features are removed digitally-leaving the underlying terrain. The data files contain the spatial elevation data of the terrain in a digital format which usually presented as a rectangular grid, and can further be transformed to visualize slope, aspect and viewshed conditions of the terrain.

Elevation Analysis: An analysis way of terrain based on GIS mapping. Elevation analysis can perform certain analytical tasks quickly and easily, without having to collect or update an authoritative set of base data. Elevation analysis is essential for many GIS applications example for natural resource management, conservation, agriculture, transportation, risk management.

GIS Mapping: A method of digital mapping and a technology that offers different tools in which we produce and use the maps to manage our communities and industries. GIS creates intelligent super maps through which sophisticated planning and analysis can be performed based on the transformation of spatial locations of real-world features and visualize the spatial relationships among them.

GIS: A system designed to capture, store, manipulate, analyze, manage, and present all types of spatial or geographical data. GIS tools allow users to create interactive queries, analyze spatial information, edit data in maps, and present the results of all these operations. GIS can specify locations or extents in the coordinate system recorded as dates/times of occurrence, and x, y, and z coordinates representing, longitude, latitude, and elevation by using location as the key index variable.

GPS: A satellite-based navigation system made up of a network of 24 satellites placed into orbit by the U.S. Department of Defense. GPS was originally intended for military applications, but in the 1980s, the government made the system available for civilian use. GPS works in any weather conditions, anywhere in the world, anytime in a day.

Point Density Analysis: A tool of GIS mapping that can calculate the density of point features around each output raster cell. A neighborhood is defined around each raster cell center, and the number of points that fall within the neighborhood is totaled and divided by the area of the neighborhood. Possible uses include finding density of disease cases, crime cases, or firms clustering. For example, one address might actually represent a crime case, or some crime cases might be weighted more heavily than others in determining crime rates. Increasing the radius will not greatly change the calculated density values. Although more points will fall inside the larger neighborhood, this number will be divided by a larger area when calculating density. The main effect of a larger radius is that density is calculated considering a larger number of points, which can be outside from the raster cell.

Slope Analysis: An analysis way of terrain based on GIS mapping. Slope analysis calculates the maximum rate of change between each cell and its neighbors, and the maximum change in elevation over the distance between the cell and its eight neighbors. Every cell in the output raster has a slope value. The lower the slope value, the flatter the terrain, while the higher the slope value, the steeper the terrain. In GIS mapping, slope analysis is very important for suitability analysis and predicting potential hazards. Analyzing the terrain slope of a given location plays an important role in the research of flood disaster, site planning, and conservation.

Geographic Information System (GIS) Modeling Analysis and the Effects of Spatial Distribution and Environmental Factors on Breast Cancer Incidence

Akram Gasmelseed

University of Science and Technology, Sudan

Ali H. Alharbi

Qassim University, Saudi Arabia

INTRODUCTION

Breast cancer is a major health issue in all countries affecting thousands of women (Tazzite et al., 2013; Dube & Gupta, 2015). So far its cause(s) are unknown and the national and international strategies to reduce its morbidity and mortality levels are based on early detection of cancer through screening and treatment according to clinical guidelines. Thus, knowledge of which women are at risk and why they are at risk is therefore essential component of disease prevention and screening. Researchers from the International Agency for Research on Cancer (IARC) and the World Health Organization (WHO) reports that globally breast cancer might contribute to the greatest burden on women's health when compared to other cancer sites (World Health Organization, n.d.). In 2015, an estimated 231,840 new cases of invasive breast cancer are expected to be diagnosed in women in the United States, along with 60,290 new cases of non-invasive (in situ) breast cancer (Siegel et al., 2015). However, all locations are not equal for breast cancer risk and thus support a major role of the geography in breast carcinogenesis (Akram & Nanna, 2003).

The purpose of this work is to provide a more detailed analysis of the breast cancer distribution in the United States by comparing the spatial distribution of breast cancer cases against

physical environmental factors using Geographic Information System (GIS) (Figure 1). Further, it gives background information to the GIS and its applications in health-related research.

BACKGROUND

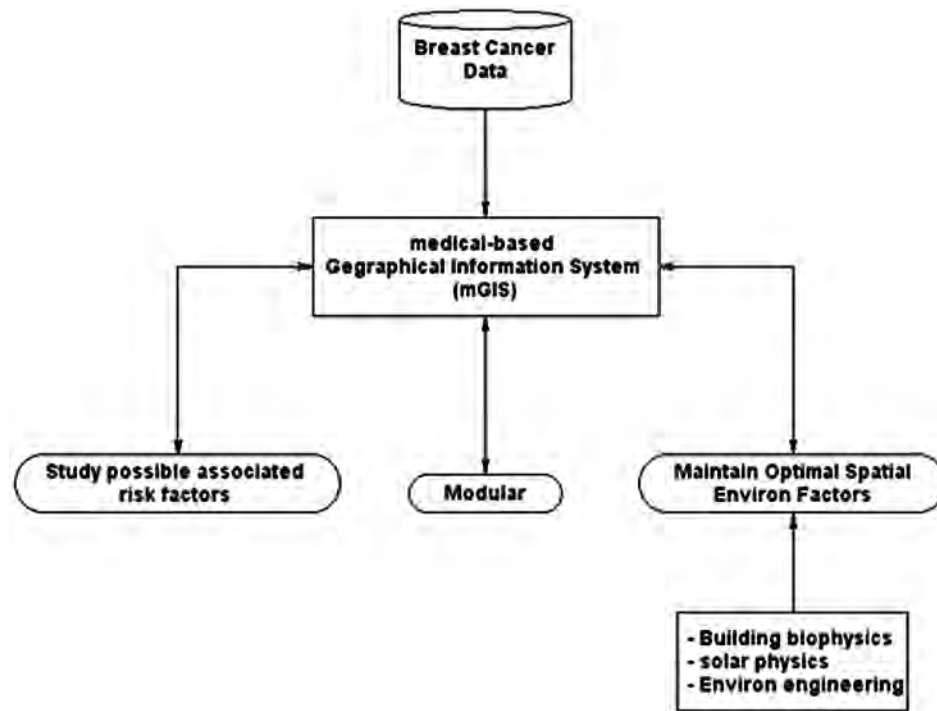
Breast Cancer Facts/ Spatial-Based Patterns

Previous reports have shown that the Northeast United States has a 16% higher breast cancer mortality rate than the rest of the country (Kulldorff et al., 1997). The probability of breast cancer risk is not equal for all locations which indicate that geography plays a very important role in the etiology of breast cancer.

There are geographic patterns of high cases of breast cancer, and the analysis of these patterns is very important in the formulation of hypotheses about risks and focus investment more effectively in research and intervention on the most significant areas (Laden et al., 1997).

In general, breast cancer incidence rates have continued to increase since 1980, although the rate of increase slowed down in the 1990s, compared to the 1980s (American Cancer Society, 2004). Furthermore, during the more recent time period, breast cancer incidence rates have increased only

Figure 1. medical-based GIS



in those aged 50 and over. The mortality rates declined by around 1.4% per year during 1989-1995 and by 3.2% afterwards, with the largest decreases in younger women in both whites and African Americans. These decreases are probably due to the result of both earlier detection and improved treatment. Clearly, the ultimate cause of breast cancer is unknown (Roche, 1998), but several risk factors appear to play a role.

Previous literatures have shown that breast cancer mortality and incidence rates vary geographically according to the different regions of the United States (Devesa et al., 1999; Joseph et al., 2004). Generally, the disease is most common in North America and Western Europe, account for about one in four female cancers in these regions, while in the Far East (China and Japan) it is very much rare (Le et al., 2002). Furthermore, the disease rates among Asian-Americans are lower than those of U.S. whites but considerably higher than rates prevailing in Asia. Thus, it is suspected that migration to the US brings about a change in endocrine function among Asian women, although

reasons for this change remain obscure (Wu et al., 1996). In fact, the geographic patterns of cancer around the world and within countries have provided important clues to the causes of cancer (Robert, 1996). The highest incidence rates of all are found in Hawaii, where a rate of 93.9 per 100,000 female populations has been reported, and in US white women. The incidence rises with age from about age 30. Moreover, there are ethnic variations, such as a high incidence in Israeli Jews compared with non-Jews in Israel. It is more common in single women, in higher social classes, and in urban rather than rural areas. On the other hand, about 1% of cases occur in males. Mortality has increased less rapidly than incidence, but breast cancer is most common cause of cancer death.

What is Geographic Information System (GIS)?

Maps are well-known to be one of the most used visual tools in our life today. They have very long tradition and they have been used from ancient

time. They are very useful tools to visually show a place and its features. However, maps, at least in their traditional form, have limited capabilities in the amount and type of information they can depict. In addition to this, the capability of human processing of huge information is limited. Gathering information from different sources, integrate, analysis, and interpret them is a complex and time-consuming task.

Data visualization is one of the areas that have been improved due to the adoption of new technology and visualization frameworks and systems. Advancement in information technologies has affected each aspect of our daily life. Data integration, processing and visualization are no exception. Today, intelligent computerized tools are available to help users enter and process information effectively and efficiently. The capability of computer systems in processing and make sense of data is improving dramatically in the last few years. Data visualization is the branch of information processing that deals with finding ways and methods to present data using various visualization techniques in order to make it easy for the users to interpret and make sense of the data.

One of the general and most comprehensive definitions of GIS is the one provided by ESRI (Environmental System Research Institute) (GIS dictionary, ESRI, n.d.), in which GIS is described as:

An integrated collection of computer software and data used to view and manage information about geographic places, analyze spatial relationships, and model spatial processes. A GIS provides a framework for gathering and organizing spatial data and related information so that it can be displayed and analyzed.

This definition stresses the fact that GIS is more than just the technology behind this system, but it is actually a comprehensive framework to support decision making by the integration of spatial data and other factors using visualization techniques.

(GIS) can be defined as:

A collection of science and technology tools used to manage geographic relationships and integrate information. GIS helps us analyze spatially-referenced data and make well-informed decisions based on the association between the data and the geography.

Applications of GIS in Health Science

Applications of GIS in health science have increased dramatically in the last few years with many innovative approaches have emerged. Understating the relationship between health and contextual factors, such as sociodemographic, environmental and political variables, can be complex. GIS can facilitate the integration and presentation of these contextual variables to gain insight into their influence on health. This section reviews the research related to the use of GIS to improve health care.

A previous study conducted a review that focuses on the applications of GIS in public health and health promotion (Nykiforuk & Flaman, 2011). The final review was composed of 621 journal articles and book chapters reporting health-related applications of GIS. According to the review, GIS applications in public health can be classified into four main themes: disease surveillance, risk analysis, health access and planning, and community health profiling. Moreover, some other reviewed literature on the applications of GIS in health science research (Lyseen, et al., 2014), and present a conceptual framework that can be used to better categorize and understand research in this area, and thus put emphasis on areas that need further research. The sample size of the study consists of a total of 865 articles pulled out from the research databases Scopus and Web of Science. By applying ground-theory data analysis method, the study presents a framework that comprises four conceptual domains to represent GIS applications in health science. These domains are:

- Spatial analysis of diseases,
- Spatial analysis of health service planning,
- Public health, and
- Health technology.

Disease Surveillance

One of the most well-known applications of using GIS in public research is for disease surveillance, which is the integration and tracking of data related to the incidence, prevalence, and the spread of the disease in the community. Epidemiology seek to answer basic questions such as when and where are cases of an infectious disease currently occupying, can we predict the next occurrence of the disease, and how can we prevent the spread of the disease? GIS can help increase the speed and occurrence of finding answers to epidemiology these questions by integrating spatial information on the disease with other types of geographic information to track and map the spread of the disease (Eisen & Eisen, 2014).

Research in this area can be classified into two interrelated categories: disease mapping and disease modelling. Disease modeling is the extension of GIS capabilities to support the decision making process. It can be applied to the data to (a) predict the outbreak of a disease, (b) identify Planning of Healthcare Services Literature reported differ-

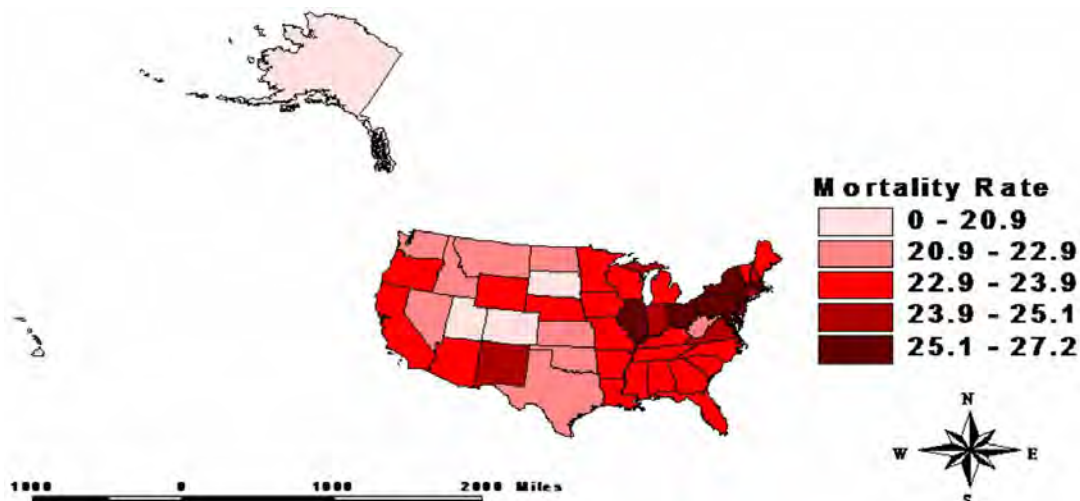
ent ways of applying GIS technology to support decision making in the planning and distribution of healthcare service.

Strategic planning is required to achieve optimal and demand-based distribution of healthcare centers. An example, GIS was used to assess hospital distributions in Seoul, the capital of South Korea (Lee & Moon, 2014). The study proved that GIS can be used to identify and examine the factors that influence health service distribution.

In developing countries, planning and delivery of healthcare services can be critical. Brijnath and De Souza (Brijnath & De Souza, 2012) discussed different ways in which GIS can be utilized to enhance health service planning and delivery for infectious diseases in low-income countries. according to the paper GIS be used to:

1. Get insight into the optimized locations for distribution healthcare centers,
2. Provide real-time surveillance system for the control of diseases,
3. Serve as an evidence-based accountability tool between funding agencies and healthcare provides, and finally
4. Transform data into accessible format to make it easy for stakeholders to make informed decisions.

Figure 2. Female breast cancer mortality (1994-98) rate by state



METHODOLOGY

In this research we obtained mortality estimates (Figure 2) in 50 states of the United States from the National Center of Health Statistics (NCHS) for the years 1994-98 (National Cancer Institute, n.d.). We excluded those states for which the corresponding data was not available.

The environmental factors used in the analysis were limited to seven spatial environmental factors collected from NSTATE, LLC (United States Geography, n.d.). In this work Regression analysis have been used to detect and evaluate the statistical significance between the mortality rates and the spatial environmental factors.

RESULTS

Preliminary results show an inverse relationship between the death rates and mean elevation ($T = -6.25$; $p = 0.000$) and positive relationship with the average temperature ($T = 3.07$; $p = 0.004$). The high F-value 21.71 and low p-value indicate there is a statistically significant difference between mortality rates and the selected spatial environmental factors. The differences are certified as statistically significant and are probably not due to sampling variability only. All the relevant finding related to the explanatory power of the independent variables, in the goodness of fit of all possible models can be enhanced by looking at the best-fit models with the significant (at the 5% level, $p \leq 0.05$) explanatory underlying variables, for each of the mortality/incidence pattern groups according to the spatial environmental factors (Table 1 and Table 2).

DISCUSSION

Effect of Average Temperature on Breast Cancer

Temperatures cause a variety of physiological changes, e.g. of blood composition, blood pres-

Table 1. Stepwise multiple regression, explanatory variables best-fit model according to mortality rates, 1991-95, 1992-96, 1993-97 and 1994-98.

Step	1	2	3	4
Constant	21.14	21.10	21.39	25.27
Ave. Temp	0.048	0.069	0.067	0.069
<i>T-Value</i>	3.42	6.20	6.09	6.27
<i>P-Value</i>	0.001	0.000	0.000	0.000
Elevation		-0.00080	-0.00075	-0.00067
<i>T-Value</i>		-11.24	-10.29	-8.14
<i>P-Value</i>		0.000	0.000	0.000
Land Area			-0.00000	-0.00000
<i>T-Value</i>			-2.53	-3.00
<i>P-Value</i>			0.012	0.003
Low Ave. Temp				-0.046
<i>T-Value</i>				-2.06
<i>P-Value</i>				0.041
<i>S</i>	2.13	1.65	1.63	1.61
<i>R-Sq</i>	5.81	43.54	45.40	46.61
<i>R-Sq (adj)</i>	5.32	42.95	44.53	45.47

sure (Stout & Crawford, 1991) and circadian rhythms (Reiter, 1991), which in turn are believed to contribute to an increased mortality/incidence of breast cancer (Kliukiene et al., 2001). Furthermore, published laboratory studies have reported that melatonin might have cancer-preventive activities (Coleman & Reiter, 1992). Moreover, melatonin has been observed to exert potent inhibition on cancer growth by acting as natural anti-angiogenic molecule, with a following opposition or angiogenesis-dependent cancer proliferation (Lissoni et al., 2001). This has been demonstrated in certain human breast cancer cell lines such as MCF-7, with additional in-vivo effects on breast oncogenesis in various rat models. Melatonin (N-acetyl-5-methoxytryptamine) is a hormone secreted by the pineal gland of vertebrates, which controls several physiological functions associated with circadian rhythm (Becker-Andre et al., 1994). Temperature is a major regulator of circadian rhythms. Circadian rhythms control many physiological activities. Synchronization of biological

Table 2. Stepwise multiple regression, explanatory variables best-fit model ($p \leq 0.05$) according to incidence rates, 1990-94, 1991-95, 1993-97 and 1994-98.

Step	1	2	3	4
Constant	105.2	182.8	185.2	182.7
Ave. Temp	0.026	0.055	0.130	0.171
T-Value	0.40	1.00	2.34	3.07
P-Value	0.690	0.317	0.020	0.003
Elevation	-0.00128	0.00033	0.00050	0.00073
T-Value	-3.29	0.84	1.34	1.98
P-Value	0.001	0.402	0.183	0.050
Low Ave. Temp		-0.93	-1.04	-1.05
T-Value		-7.86	-8.94	-9.30
P-Value		0.000	0.000	0.000
Water/Land Ratio			13.7	27.2
T-Value			3.98	4.88
P-Value			0.000	0.000
Water Area				-0.00014
T-Value				-3.03
P-Value				0.003
S	8.75	7.37	7.03	6.84
R-Sq	6.79	34.25	40.64	44.15
R-Sq (adj)	5.53	32.92	39.02	42.24

clocks to environmental time is adaptive and important for physiological homeostasis and for the proper timing of species-specific behaviors. A number of investigations have been conducted to examine the effects of temperature and light on melatonin rhythms (Underwood & Calaban, 1987). Based on our analysis, we proposed a second order equation ($y = a + bx + cx^2$) to describe the correlation between the average temperature and mortality rates as this model seems to be the one that best describes the relationship between the two parameters. Hence, based on this model, the estimated temperature (T_e) values at which the mortality rates are low is in the range $49 < T_e < 11^\circ\text{C}$.

To further examine the idea that temperature can influence breast cancer rates, we tested the relationship (if any) between mortality rates and geographic coordinates: latitude and longitude.

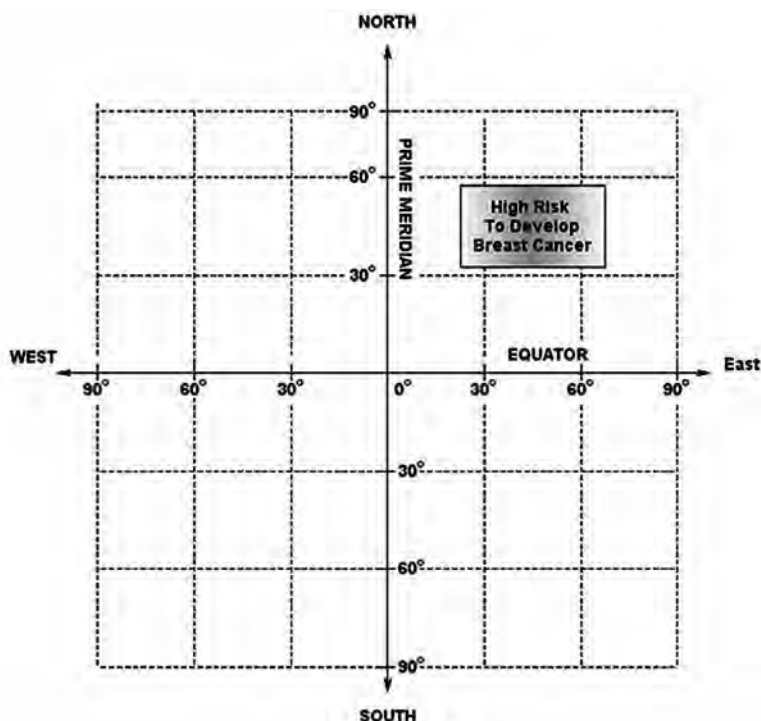
Clearly, the principal controls of temperature variations are latitude and elevation (altitude). The latitude for a given point on the earth has an effect on temperature; the farther away from the Equator (0°) a point is the cooler average temperature that point is likely to have. We obtained worldwide breast cancer mortality statistics from WHO Cancer Mortality Database, WHO/IARC (WHO Cancer Mortality Database, n.d.). Then, we investigated the possible correlation between the geographic coordinates of the countries and the age-specific rates (ASR) of breast mortality. The result revealed that there is existing negative relationship between latitude and mortality rates, i.e. by increasing latitude the death rate increases. Variations in latitude lead to variations in the temperature and sunlight intensity and hence for disturbance of melatonin rhythms. A previous report, investigators showed evidence for a link between cosmic ray intensity and cancer due to variations in the latitude (Juckett & Rosenberg, 1997). Moreover, it was demonstrated that light-induced melatonin suppression in humans is sensitive to short wavelength light (420 – 480 nm; $I_{\max}$ @ 460 nm) (Skene, 2003).

Jet Lag and Circadian Rhythm

Investigators reported that “Jet Lag” is caused from air travel through changing time zones, and hence leading to disturbance of melatonin rhythms (Samel & Wegmann, 1997). The main but not only cause of jet lag is crossing time zones. Usually going east is worse than going west (Delagrangé & Guardiola-Lemaitre, 1997). Finally, the obtained negative relationship between mortality data and elevation is clear in the sense that reduced temperature occur at higher elevation where there is less water vapor to trap and hold heat (Figure 3).

Further, we found a large, significant relationship between women at ages 30-74 and the

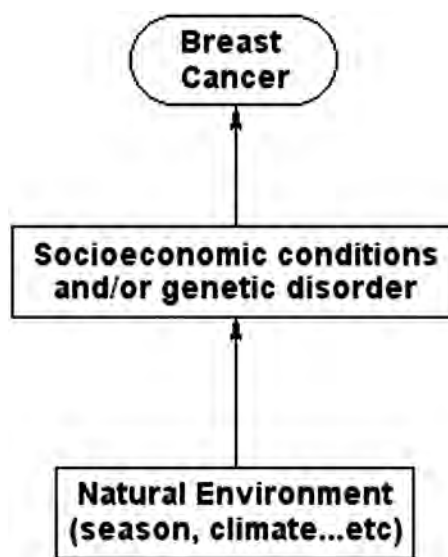
Figure 3. Influence of “jet lag” on breast cancer development



geographic latitude. For women aged 65 years and older, breast cancer mortality is 26% higher in New England than in the south, while incidence is only 3% higher. Mortality rates from breast cancer are approximately 25% higher for women in the northeastern United States than for women in the South or West. This can be depicted by the correlation with the geographic longitude, see Figure 3. Investigators reported that “Jet Lag” is caused from air travel through changing time zones, and hence leading to disturbance of melatonin rhythms.

The exposure to some spatial environmental factors seems likely to have a major impact on the overall trends in breast cancer rates. The possibility to develop cancer is existing naturally through the environment factors. Unbalanced in these nature factors and/or other existing risk factors, e.g. socioeconomic conditions will increase the chance for cancer development (Figure 4).

Figure 4. Role of spatial environmental factors in breast cancer development



FUTURE RESEARCH DIRECTIONS

The results presented in this work need to be elaborated based on the latest advancement in information technology. Hence, for future research, our goal is to incorporate cutting-edge information technology services to increase our understanding of variations in the breast cancer disease and geographical factors, to describe the disease burden, and to obtain guidance on the action to be taken.

Cloud computing is one of the cutting-edge technologies that may be feasible in the future. Thus, the integration of GIS with cloud computing technology can provide new insights into breast cancer research by providing shared information through internet.

Moreover, adopting such approach could maintain a reliable data as well as device and location independency.

CONCLUSION

We have combined data and techniques from the areas of environmental physics and cancer epidemiology to explore the hypothesis that spatial environmental factors can induce changes in melatonin rhythms, which can lead to increases in breast cancer rates. The results are interpreted as evidence that temperatures can modulate cancer via changes induced in melatonin rhythms.

Results show that the exposure to some spatial environmental factors seems likely to have a major impact on the overall trends in breast cancer rates. Moreover, the possibility to develop cancer is existing naturally through the environment factors.

Furthermore, this study provides valuable reference information for clinicians and health administrators, as well as baseline for more detailed studies of patient survival for individual environmental/geographical sites. The earlier the cancer is detected, the smaller the chance that it already has spread beyond the limited primary focus. This

implies that the extend of intervention needed is lesser and the prognosis improved.

Finally, map-based exploration of georeferenced health statistics will lead to a better understanding of health/environment interaction. By utilizing geographical information system technology we can enhance our ability to manage, estimate and predict breast cancer distribution.

REFERENCES

- Akram, A., & Nanna, S. (2003). *Medical-based geographical information system (mGIS): diagnosis versus healing*. Global Health Network Supercourse Project. Retrieved October 2, 2015, from <http://supercourse.isciii.es/lecture/lec10671/index.htm>
- American Cancer Society. (2004). *Breast Cancer Facts and Figures 2003-2004*. Atlanta, GA: American Cancer Society. Retrieved July 28, 2005, from <https://www.cancer.org/acs/groups/content/@nho/documents/document/caff2003brfp-wsecuredpdf.pdf>
- Becker-Andre, M., Wiesenberger, I., Schaeren-Wiemers, N., Andre, E., Missbach, M., Saurat, J. H., & Carlberg, C. (1994). Pineal gland hormone melatonin binds and activates an orphan of the nuclear receptor superfamily. *The Journal of Biological Chemistry*, 269(46), 28531–28534. PMID:7961794
- Brijnath, B., & De Souza, D. K. (2012). Four Ways Geographic Information Systems Can Help to Enhance Health Service Planning and Delivery for Infectious Diseases in Low-Income Countries. *Journal of Health Care for the Poor and Underserved*, 23(4), 1410–1420. doi:10.1353/hpu.2012.0146 PMID:23698658
- Cancer (Fact sheet N 297). World Health Organization. (n.d.). Retrieved from <http://www.who.int/mediacentre/factsheets/fs297/>

- Coleman, M. P., & Reiter, R. J. (1992). Breast cancer, blindness and melatonin. *European Journal of Cancer*, 28(2-3), 501–503. doi:10.1016/S0959-8049(05)80087-5 PMID:1591073
- Delagrang, P., & Guardiola-Lemaitre, B. (1997). Melatonin, its receptors, and relationships with biological rhythm disorders. *Clinical Neuropharmacology*, 20(6), 482–510. doi:10.1097/00002826-199712000-00002 PMID:9403224
- Devesa, S. S., Grauman, D. J., Blot, W. J., Pennello, G. A., Hoover, R. N., & Fraumeni, J. F. (1999). *Atlas of cancer mortality in the United States, 1950-94*. Bethesda, MD: National Institutes of Health, National Cancer Institute.
- Dictionary, G. I. S. ESRI. (n.d.). Retrieved from <http://support.esri.com/>
- Dubey, A. K., Gupta, U., & Jain, S. (2015). Breast cancer statistics and prediction methodology: A systematic review and analysis. *Asian Pacific Journal of Cancer Prevention*, 16(10), 4237–4245. doi:10.7314/APJCP.2015.16.10.4237 PMID:26028079
- Eisen, R. J., & Eisen, L. (2014). Use of Geographic Information Systems in Infectious Disease Surveillance. In N. M. M’ikanatha & J. K. Iskander (Eds.), *Concepts and Methods in Infectious Disease Surveillance*. Oxford, UK: John Wiley & Sons, Ltd. doi:10.1002/9781118928646.ch20
- Joseph, S. T., DeChello, L. M., Kulldorff, M., Gregorio, D. I., Gershman, S., & Mrosczyk, M. (2004). The geographic distribution of breast cancer incidence in Massachusetts 1988 to 1997, adjusted for covariates. *International Journal of Health Geographics*, 3(1), 17. doi:10.1186/1476-072X-3-17 PMID:15291960
- Juckett, D. A., & Rosenberg, B. (1997). Time series analysis supporting the hypothesis that enhanced cosmic radiation during germ cell formation can increase breast cancer mortality in germ cell cohorts. *International Journal of Biometeorology*, 40(4), 206–221. doi:10.1007/s004840050043 PMID:9225597
- Kliukiene, J., Tynes, T., & Andersen, A. (2001). Risk of breast cancer among Norwegian women with visual impairment. *British Journal of Cancer*, 84(3), 397–399. doi:10.1054/bjoc.2000.1617 PMID:11161406
- Kulldorff, M., Feuer, E. J., Miller, B. A., & Freedman, L. S. (1997). Breast cancer clusters in the northeast United States: A geographic analysis. *American Journal of Epidemiology*, 146(2), 161–170. doi:10.1093/oxfordjournals.aje.a009247 PMID:9230778
- Laden, F., Spiegelman, D., Neas, L. M., Colditz, G. A., Hankinson, S. E., Manson, J. E., & Hunter, D. J. et al. (1997). Geographic variation in breast cancer incidence rates in a cohort of U.S. women. *Journal of the National Cancer Institute*, 89(18), 1373–1378. doi:10.1093/jnci/89.18.1373 PMID:9308708
- Le, G. M., Gomez, S. L., Clarke, C. A., Glaser, S. L., & West, D. W. (2002). Cancer incidence patterns among Vietnamese in the United States and Ha Noi, Vietnam. *International Journal of Cancer*, 102(4), 412–417. doi:10.1002/ijc.10725 PMID:12402312
- Lee, K.-S., & Moon, K.-J. (2014). Hospital distribution in a metropolitan city: Assessment by a geographical information system grid modelling approach. *Geospatial Health*, 8(2), 537–544. doi:10.4081/gh.2014.43 PMID:24893031
- Lissoni, P., Rovelli, F., Malugani, F., Bucovec, R., Conti, A., & Maestroni, G. J. (2001). Anti-angiogenic activity of melatonin in advanced cancer patients. *Neuroendocrinology Letters*, 22(1), 45–47. PMID:11335879
- Lyseen, A. K., Nohr, C., Sorensen, E. M., Gudes, O., Geraghty, E. M., Shaw, N. T., & Bivona-Tellez, C. (2014). A Review and framework for categorizing current research and development in health related geographical information systems (GIS) studies. *Yearbook of Medical Informatics*, 9(1), 110–124. doi:10.15265/IY-2014-0008 PMID:25123730

National Cancer Institute. (n.d.). *SEER*Stat version 5.3.1*. Bethesda, MD: National Cancer Institute. Retrieved from <http://www.seer.cancer.gov/seerstat>

Nykiforuk, C., & Flaman, L. (2011). Geographic information systems (GIS) for health promotion and public health: A review. *Health Promotion Practice*, 12(1), 63–73. doi:10.1177/1524839909334624 PMID:19546198

Reiter, R. J. (1991). Pineal melatonin: Cell biology of its synthesis and its physiological interactions. *Endocrine Reviews*, 12(2), 151–180. doi:10.1210/edrv-12-2-151 PMID:1649044

Robert, N. P. (1996). *Cancer wars: How politics shapes what we know and don't know about cancer*. New York: Basic Books.

Roche, L. M. (1998). *Breast cancer in New Jersey, 1979-1995*. New Jersey Dept. of Health and Senior Services, Cancer Surveillance Program, Office of Cancer Epidemiology.

Samel, A., & Wegmann, H. M. (1997). Bright light: A countermeasure for jet lag? *Chronobiology International*, 14(2), 173–183. doi:10.3109/07420529709001154 PMID:9095377

Siegel, R. L., Miller, K. D., & Jemal, A. (2015). Cancer statistics, 2015. *CA: a Cancer Journal for Clinicians*, 65(1), 5–29. doi:10.3322/caac.21254 PMID:25559415

Skene, D. J. (2003). Optimization of light and melatonin to phase-shift human circadian rhythms. *Journal of Neuroendocrinology*, 15(4), 438–441. doi:10.1046/j.1365-2826.2003.01006.x PMID:12622847

Stout, R. W., & Crawford, V. (1991). Seasonal variations in fibrinogen concentration among elderly people. *Lancet*, 338(8758), 9–13. doi:10.1016/0140-6736(91)90004-9 PMID:1676131

Tazzite, A., Jouhadi, H., Saiss, K., Benider, A., & Nadifi, S. (2013). Relationship between family history of breast cancer and clinicopathological features in Moroccan patients. *Ethiopian Journal of Health Sciences*, 23(2), 150–157. PMID:23950631

Underwood, H., & Calaban, M. (1987). Pineal melatonin rhythms in the lizard *Anolis carolinensis*: I. Response to light and temperature cycles. *Journal of Biological Rhythms*, 2(3), 179–193. doi:10.1177/074873048700200302 PMID:2979659

United States Geography. (n.d.). Retrieved from <http://www.netstate.com>

WHO Cancer Mortality Database. (n.d.). World Health Organization. Retrieved from <http://www-dep.iarc.fr/WHODb/WHODb.htm>

Wu, A. H., Ziegler, R. G., Horn-Ross, P. L., Nomura, A. M., West, D. W., Kolonel, L. N., & Pike, M. C. et al. (1996). Tofu and risk of breast cancer in Asian-Americans. *Cancer Epidemiology, Biomarkers & Prevention*, 5(11), 901–906. PMID:8922298

ADDITIONAL READING

Barrett, R. K., & Takahashi, J. S. (1995). Temperature compensation and temperature entrainment of the chick pineal cell circadian clock. *The Journal of Neuroscience*, 15(8), 5681–5692. PMID:7643210

Barrett, R. K., & Takahashi, J. S. (1997). Lability of circadian pacemaker amplitude in chick pineal cells: A temperature-dependent process. *Journal of Biological Rhythms*, 12(4), 309–318. doi:10.1177/074873049701200403 PMID:9438879

- Boulos, Z., Macchi, M. M., Sturchler, M. P., Stewart, K. T., Brainard, G. C., Suhner, A., & Steffen, R. et al. (2002). Light visor treatment for jet lag after westward travel across six time zones. *Aviation, Space, and Environmental Medicine*, 73(10), 953–963. PMID:12398256
- Choi, W. S., Parker, B. A., Pierce, J. P., & Greenberg, E. R. (1996). Regional differences in the incidence and treatment of carcinoma in situ of the breast. *Cancer Epidemiology, Biomarkers & Prevention*, 5(4), 317–320. PMID:8722225
- Garcia, E., Hurley, S., Nelson, D. O., Hertz, A., & Reynolds, P. (2015). Hazardous air pollutants and breast cancer risk in California teachers: A cohort study. *Environmental Health*, 14(1), 14. doi:10.1186/1476-069X-14-14 PMID:25636809
- Garfinkel, L., Boring, C. C., & Heath, C. W. Jr. (1994). Changing trends: An overview of breast cancer incidence and mortality. *Cancer*, 74(1Suppl), 222–227. doi:10.1002/cnrcr.2820741304 PMID:8004590
- Keegan, T. H., Chang, E. T., John, E. M., Horn-Ross, P. L., Wrensch, M. R., Glaser, S. L., & Clarke, C. A. (2007). Recent changes in breast cancer incidence and risk factor prevalence in San Francisco Bay area and California women: 1988 to 2004. *Breast Cancer Research*, 9(5), R62. doi:10.1186/bcr1768 PMID:20210979
- Lacey, J. V., Devesa, S. S., & Brinton, L. A. (2002). Recent trends in breast cancer incidence and mortality. *Environmental and Molecular Mutagenesis*, 39(2-3), 82–88. doi:10.1002/em.10062 PMID:11921173
- Medina-Leendertz, S., Valero, N., Chacin-Bonilla, L., Anez, F., Giraldoth, D., Arias, J., & Bonilla, E. et al. (2001). High intensity light increases olfactory bulb melatonin in Venezuelan equine encephalitis virus infection. *Neurochemical Research*, 26(3), 231–234. doi:10.1023/A:1010964500370 PMID:11495546
- Moyer, R. W., Firth, B. T., & Kennaway, D. J. (1995). Effect of constant temperatures, darkness and light on the secretion of melatonin by pineal explants and retinas in the gecko *Christinus marmoratus*. *Brain Research*, 675(1-2), 345–348. doi:10.1016/0006-8993(95)00073-Y PMID:7796151
- Nair, I., & Morgan, M. G. (1990). Electromagnetic fields: The jury's still out. 1. Biological effects. *IEEE Spectrum*, 27(8), 23–27. doi:10.1109/6.58428
- Rea, M. S., Bullough, J. D., & Figueiro, M. G. (2002). Phototransduction for human melatonin suppression. *Journal of Pineal Research*, 32(4), 209–213. doi:10.1034/j.1600-079X.2002.01881.x PMID:11982788
- Reiter, R. J., & Richardson, B. A. (1992). Magnetic field effects on pineal indoleamine metabolism and possible biological consequences. *The FASEB Journal*, 6(6), 2283–2287. PMID:1544540
- Reynolds, P., Hurley, S., Goldberg, D. E., Anton-Culver, H., Bernstein, L., Deapen, D., & Ziogas, A. et al. (2004). Regional variations in breast cancer among california teachers. *Epidemiology (Cambridge, Mass.)*, 15(6), 746–754. doi:10.1097/01.ede.0000134863.45834.50 PMID:15475725
- Sturgeon, S. R., Schairer, C., Gail, M., McAdams, M., Brinton, L. A., & Hoover, R. N. (1995). Geographic variation in mortality from breast cancer among white women in the United States. *Journal of the National Cancer*, 87(24), 1846–1853. doi:10.1093/jnci/87.24.1846 PMID:7494228
- Tosini, G., & Fukuhara, C. (2003). Photic and circadian regulation of retinal melatonin in mammals. *Journal of Neuroendocrinology*, 15(4), 364–369. doi:10.1046/j.1365-2826.2003.00973.x PMID:12622835
- Wang, R. J. (1975). Lethal effect of daylight fluorescent light on human cells in tissue-culture medium. *Photochemistry and Photobiology*, 21(5), 373–375. doi:10.1111/j.1751-1097.1975.tb06688.x PMID:1208665

Wright, K. P. Jr, Hughes, R. J., Kronauer, R. E., Dijk, D. J., & Czeisler, C. A. (2001). Intrinsic near-24-h pacemaker period determines limits of circadian entrainment to a weak synchronizer in humans. *Proceedings of the National Academy of Sciences of the United States of America*, 98(24), 14027–14032. doi:10.1073/pnas.201530198 PMID:11717461

KEY TERMS AND DEFINITIONS

Breast Cancer: It is tumor growth that starts in the cells of the breast. The disease occurs almost entirely in women, and it can affect men as well.

Circadian Rhythm: A description of physiological and behavioral patterns altered within 24 hours.

Geographic Information System (mGIS): It is defined as computer-established information system that is particularly designed to view and manage information about geographic places, analyze spatial relationships, and model spatial processes.

Incidence and Mortality Rates: Are defined as how many people get and die from breast cancer respectively.

Jet Lag: It is a medical condition which results as consequence of alterations to circadian rhythms due to long hours of travel.

Medical-Based Geographic Information System (mGIS): It is defined as GIS applications and the related medical data for monitoring and analyzing medical risk factors associated with the spatial environmental factors.

Regression Analysis: It is a statistical tool for estimating the relationships between variables.

Geographic Information Systems

Paula Remoaldo

University of Minho, Portugal

Vitor P. Ribeiro

University of Minho, Portugal

Hélder Silva Lopes

University of Minho, Portugal

Sara Catarina Gomes Silva

University of Minho, Portugal

INTRODUCTION

One of the main challenges of the 21st century are caused by the large amount of geospatial information through a GIS. Throughout time there have been many attempts to define Geographic Information Systems (GIS). Yet there is no consensus on define it and restrict it to one definition is limited. In the acronym - Geographic Information Systems - the geographic refers to the Earth's surface and near-surface, therefore, all human production and activity, and non-human are possible spatialization in GIS.

GIS is recognized as an analytical and decision-making tool with many uses in different fields. Likewise it is used in many industries plus commercial, education or government. It is powerful for

- Land administration,
- Statistical mapping,
- Transport,
- Network and environment management,
- Remote sensing images,
- Water/waste management,
- Maintenance and management of public lighting,
- Regional and urban planning,
- Tourism planning,
- Healthcare planning, and in
- Crime and security management.

In broad terms GIS is a special class of information systems that keep track not only of events, activities, and things, but also their location. Computerization has opened a vast new potential in the way people communicate, analyze our surroundings and take decisions. The available data represent layers of the real world that can be stored, processed and presented later to answer future needs (Bernhardsen, 2002).

In the process of acquisition, processing and spatial representation there is the involvement of a multiplicity of inputs and outputs that can be managed on databases, which invariably seek analytical and graphical spatial embodiments. In the graphical display, vector or raster elements can be chosen, depending on the degree of specificity of the database and the type of expected results.

These databases can be collected at different scales and using a plurality of data types, including population census, aerial photography or satellite imagery. It allows to address multiple operating phases of the planning management process in a multiscale perspective with the challenge to meet more effective and efficient solutions. Due to this, nowadays it is frequently used as a spatial decision support system (SDSS) (Crossland, 2005).

Well-designed GIS should be able to provide a good computer system, because traditional GIS are intended to users operating on local servers. Traditionally GIS includes hardware and software. The hardware are the physical parts of the

DOI: 10.4018/978-1-5225-2255-3.ch301

computer itself and associated peripherals (e.g., plotters and printers); and the software is interoperable, supporting the many data formats (in the infrastructure life cycle) and implementation may be custom-designed for an organization.

Even so a GIS can have two types of groups typically called as “GIS carries” and “GIS users”, which are respectively responsible for the management and analysis. The heart of GIS technology is the ability to conduct spatial analysis, overlay data and integrate other solution and systems. Geoprocessing operations facilitate to link or merge data, spatial characteristics of data; search for particular characteristics or features in an area, update data quickly and cheaply and model data assess alternatives (maps, graphs, address lists, reports and summary statistics) tailored to meet particular needs.

Nonetheless GIS feature a number of operational advantages and have allowed the proliferation of new fields of endeavor in open access systems across multiple forms of acquisition, management, interpretation and spatial information analysis. This can be seen in the first item of the present paper where the background and GIS starting point is explored. The main goal of this paper is to underwrite the concept of GIS evolution and to identify new paths to accommodate recent scientific approaches with extensive range of application possibilities.

THEORETICAL BACKGROUND: THE STARTING POINT

GIS is the advent of a new stage of cartography. The evolution of this type of system is relatively recent, between the 50s and 60s of the XX century, but knowledge and technology have grown rapidly recently. The emergence of technological systems with computerized cartographic application arose from the need of the resolution of certain military problems and public administration domains. Many contributors and diverse influences concerning concepts and principles, data and issues

of spatial infrastructure, software vendors, application areas, allowed a cohesive growth (Figure 1). GIS organizational structure is as diverse well as the multitude of roots from which it originated multitude of proprietary and public domain GIS software packages (Hendriks, 2005). Nowadays, applicability of this type of systems widened for commercial, non-profit and academic areas.

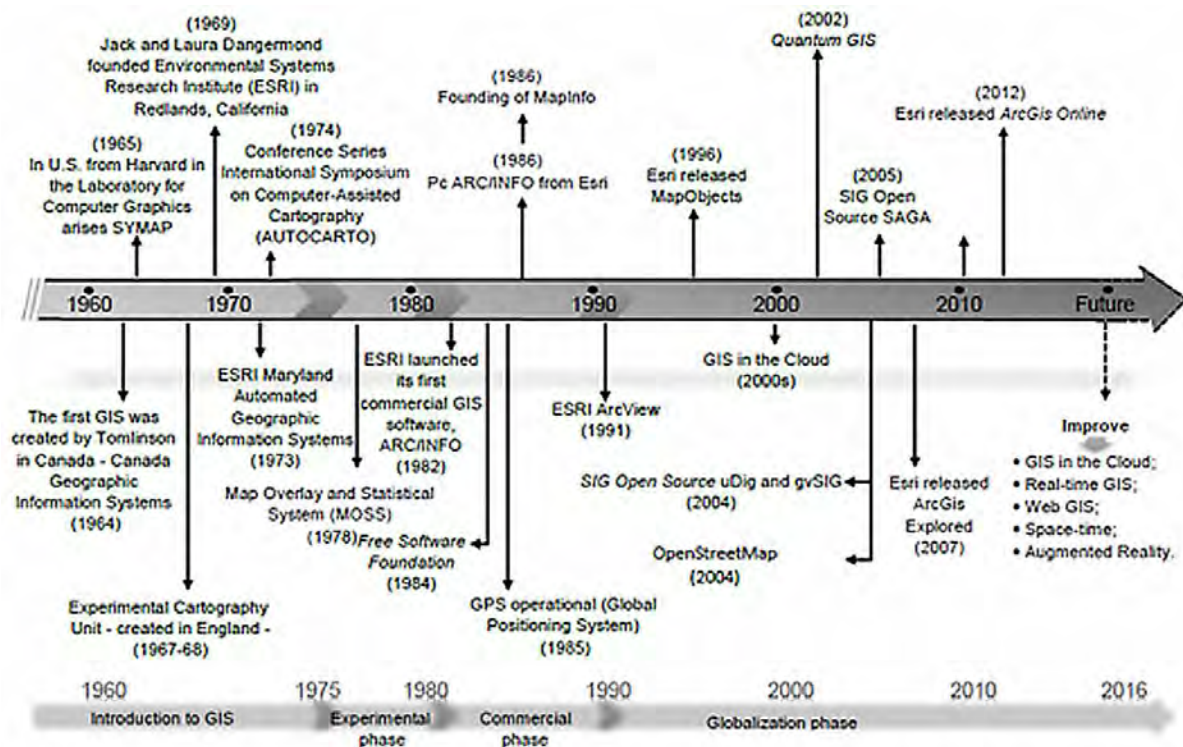
The mid 1960s witnessed the initial development of GIS in combining spatially referenced data, spatial data models and visualization. The actual roots of GIS are complex and difficult to determine (Miller & Goodchild, 2015).

Most authorities cite the Canada Geographic Information System (CGIS), designed around 1964, with a project led by Tomlinson (Bruno & Giannikos, 2015; Mordechai et al., 2008; Tomlinson, 1967). The objective was to obtain means for summaries and tabulations of areas of land from the Canada Land Inventory. For the registration of this lands, was made a massive federal-provincial effort to assess the utilization and potential of the Canadian land base. CGIS arose from the need to answer the challenges to measure accurately the areas of irregular geographic patches of homogeneous utilization and to overlay/compare different themes (Goodchild, 2006).

The period of the 1970s was characterized by rapid evolution, and ability of computer mapping automatic using data format and the solution of a wide range of technical issues. In the 1980s, democratization of access to computer allowed expand use of GIS. These innovations led to the first commercial viability of GIS, started to become popular as a standard computer application in government departments, universities, and private corporations. Accordingly, the ability to select, sort, extract, classify and display geographic data on the basis of complex topological and statistical criteria was available to users (Goodchild, 2006; Pourabbas, 2014).

The 1990s saw map analysis and modeling advances in GIS, and these systems became real management information tools as computing power increased. During this decade, the Open

Figure 1. Timeline of major GIS events, Source: Authors' own elaboration



GIS Consortium (OGC), aimed at developing publicly available geoprocessing specific actions, was founded. OGC is an international industry consortium, including government agencies and universities.

In 2000, the advent of web 2.0 and, more recently, the Web 3.0 (the semantic web) allow the Open Source GIS growth. Open source is a software that allows code source access to be open and free to distribute and modify. Nowadays it is possible for programmers to add new functions very rapidly and cheaply due to recent GIS advances. Mobile and internet devices, cloud computing, NoSQL databases, Semantic Web, Web services offer new ways of accessing, analyzing, and elaborating geospatial information in both real-world and virtual spaces, both for open source and commercial GIS (Pourabbas, 2014).

Despite the progress made during the twenty-first century, there are numerous challenges still lay ahead for geospatial sciences in various fields as shown in Table 1 (Yue et al., 2013).

DATA ANALYSIS AND TREATMENT SYSTEMS

Common citizens are constantly interacting with spatial dimension. Thereby, a GIS can be defined as a computer system operated by people, which comprises different aspects to be efficiently oper-

Table 1. The challenges for geospatial sciences for types of data intensity

Intensity	Description
Data	Collect of a multitude of data from space by day and accumulation at a similarly high rate.
Processing	Intensive modes of processing information in different spatiotemporal spectra.
Competition	Action of a multitude of end users accessing parallel to information (reception of a large number of users simultaneously are possible, because of development of the several services (e.g., Google Maps and Bing Maps).
Spatiotemporal	Set of spatial and temporal dimension. It can be distinguished into two types of information – dynamic or static.

Source: (Adapted from Yue et al., 2013.)

ated. Firstly, a GIS have the hardware, software and data components. Those multidimensional components can be articulated to give us the basis to develop spatial analysis. However human interaction is crucial to develop a conceptual model approach, plan, operate and analyze the information. Nowadays a GIS can be very helpful to different enterprises size, organizations and persons where geographic patterns can be modelled and predicted. Spatial/Geographical data is representing real world through layers or objects where spatial position is crucial. Typically geographical data have descriptive or spatial information (Faiz & Krichen, 2012). It can be used to represent discrete data, typically through a vector-based representation (points, line polygons), or as a continuous data through a cell-based or raster mode that uses a matrix representation.

The most interesting part of a GIS project is the Spatial analysis (Heywood et al., 2011). It is related with the capability to visual analysis of maps and imagery, computational analysis of geographic patterns, finding optimum routes, site selection, and advanced predictive modeling (ESRI, 2013:6). The world is complex, but the exponential growth of technologies, such as Global Position Systems (GPS), real time-sensors or GIS made possible there simplification (ESRI, 2013). GIS are efficient tools for recording, exploiting, analyzing and displaying geographical data that can be applied in:

- Transportation,
- Health, environmental,
- Urbanism,
- Political activities,
- Water/waste management,
- Geomarketing,
- Security,
- Tourism,
- Viticulture/enology,
- Education or
- Crime.

This broad kind of applications covering private and public sectors are growing exponentially. Open-source and proprietary software development have been contributing to this development due to its recent growth and attention.

Aside the raster and vector in spatial modelling more recently three and four-dimensional data is being investigated (Lin et al., 2013). Virtual Geographic Environments (VGEs) interest are growing faster in last years. It is characterized as being a bridge between the three scientific requirements of Geographic Information Science: multi-dimensional visualization, dynamic phenomenon simulation, and public participation (Lin et al., 2013).

GIS TRENDS

Currently, the key trends that face in GIS concerns are geospatial web, GIS in the cloud, space-time GIS, augmented reality and real-time GIS. Web-based in GIS combines information systems and geographic web technology (Chakraborty et al., 2015). Web GIS is a paradigm shift from a model based on national governments as major players in the production of geospatial data to one based on the collaboration between citizens and private institutions (Goodchild et al., 2007; Grossner et al., 2008). The enhanced participation of different actors in the generation of geospatial data, it is increasingly difficult to distinguish the producers and users (Budhathoki & Nedovic-Budic, 2008), mainly due to the free software and open source tools availability (Crampton, 2009). Due to the opening of the GIS world to a multitude of actors involved in the mapping data and Web GIS tools, refers to ads to meet end-user needs (Elwood, 2009). GIS based on the web are accessible not just from a computer but also on different multiple devices, including laptops, smartphones or tablets (Chakraborty, 2015).

The opening of the source code and the use of free software aims to contribute to greater

openness to collective voluntary participation in the use, study and modification of the software (Chakraborty et al., 2015). In the beginning of the 21st century started an innovative and leading geolocation-based service of crowdsourcing at a massive scale named as Open Street Map project (OSM). This project improves Volunteered Geographical Information' and aims to create a free digital map of the world. Those collaborative platforms are empowering citizens to create a global patchwork of geographic information (Goodchild, 2007; Mordechai Haklay, 2010). The international non for profit Open Geospatial Consortium was founded in 1994. It is an international voluntary organization that led the standards development process for geospatial and location services (Haklay et al., 2008). Among the three most relevant standards OGC include: Web Feature Service (WFS), Web Map Service (WMS) and Web Coverage Service (WCS) (Giuliani et al., 2016; Parker & Dominguez, 2015).

This new forms of utilisation of GIS on the web environment using a distributed and asynchronous requires a client-server architecture (C/S). This is characterized by a client request of a service, such as mapping, decision analysis, data processing or storage data while the server provide the service (Mekonnen & Gorsevski, 2015). Web GIS exits benefited by providing the best agents solutions to problems that traditional GIS (Chang & Park, 2006). More recently Web-GIS has become to a cloud GIS, based in model of "Software as a Service" (SaaS) (Kerski, 2015).

Cloud computing are increasingly widespread and make possible to run cloud applications in a shared data center accessed by internet. The emergence of GIS in the cloud solved the problems associated to the increase of precision and spatial-temporal scope of information. In general, there are an accumulation of multitude data records and this dataset has varied at a daily rate (Hey, 2012) and allows network access to a set of configurable data (servers, storage, applications and services) (Yang et al., 2011). The recent

emergence cloud GIS provides the ability to build a GIS service enabled for use in the cloud and can be made to scale up or down according to user needs. GIS cloud is equipped with new models of maintenance and use of geospatial data for a variety of users and to solve computing problems (Yu et al., 2014). This service provides users the ability to act in the manner of 'pay-as-you-go'. This mechanism of action has been a dream for several decades and has recently become a reality (Armbrust et al., 2010).

The cloud GIS has several characteristics namely: (i) it is not necessary to provide a software installation; (ii) not to use the computer's internal space for data storage; (iii) enable a collaborative action between different actors (flexibility); (iv) adapting services to demand and actual charge; (v) enhancing greater interoperability between various source code; (vi) decrease the time taken by decision-makers to implement processes of deliberation and (vii) the implementation of the entire system in a top-down scale (Armbrust et al., 2010; Blower, 2010; Yu et al., 2014). Cloud computing is a powerful technology that enables greater profitability in energy consumption and economic resources (Buyya et al., 2009; Lee & Chen, 2010; Marston et al., 2011). It perform massive-scale and complex computing and eliminate the need of maintain software, hardware or dedicated space (Assunção et al., 2015; Hashem et al., 2015).

Associated to this is the massive growth of generated data scale and volume which is challenging and time demanding tasks to ensure data processing, analysis and store (Hashem et al., 2015). Concerning to this Big Data concept is emerging and specified the four Vs, namely,

- Volume of data,
- Variety of data collectors,
- Velocity of data transfers, and
- Value process of discovering huge hidden values from large datasets (Gantz & Reinsel, 2011; Hashem et al., 2015).

The advancement of these technologies can enable the construction of spatial data infrastructures (SDI) and cyberinfrastructures (Schäffer et al., 2010; Yang et al., 2010). Some public cloud computing platforms are already available, including Microsoft Azure, Google App Engine or Amazon EC2. In any case a cloud can be public or private. The public cloud is available to the public while the private cloud is only used within an organization.

Cloud computing includes multiple domains, such as energy and mineral sciences, weather, traffic and simulation management systems, landscape ecology, water management, disaster management or human and environmental health (Yang et al., 2013). According to these authors the main obstacles to the success of cloud GIS are associated with policy, management, acquisition and operational requirements. Forward-looking multiple threads are identified:

- Cumulative advances for interdisciplinary approach, abreast of progress in geoscience and digital earth;
- Cloud interoperability based on standards (OGC, OGF, NIST, ISO, IEEE) and through a systemic architecture;
- Integration of innovative interaction systems for viewing and access;
- Real-time simulation for decision support;
- Security levels set to a deployed computer with a platform with distribution of certain information and the collaboration by integrating multiple platforms, and in view of the scope of science for the citizen, technology to crowdsourcing, dynamic events and challenges of education.

Another type of evolution in GIS is linked of framework of Hägerstrand (space-time model). In this context, the space-time studies the individual patterns, considering the various constraints in a particular spatial-temporal environment (Hägerstrand, 1970; Hägerstrand, 1989). There have been a number of efforts to ensure the incorpora-

tion of concepts in a GIS (e.g., Goodchild, 2013; Miller, 1991; Neutens et al., 2007; Shaw & Yu, 2009). This system presents a three dimensional orthogonal structure that consists of the union between two spatial dimensions and a temporal dimension. The spatial dimensions, structured 2D scale, represent the location of individuals, while the dimension of time represents the timing of the individual movement in a spatiotemporal system (Miller, 2004).

Several variables can represent the characteristics of the daily activity of an individual: location, time, duration, sequence and frequency of the type of activities (Ren & Kwan, 2009). It must be associated with at least one activity. Distinguished two types of activities: the movable and stationary activity. Mobile activity refers to a local motion toward another, while the stationary activity leads to a fixed location. Representation in Hägerstrand system is done in two ways: by vertical line segment when it comes to a stationary activity and a sloping straight line when there is movement toward a certain place (Chen et al., 2011).

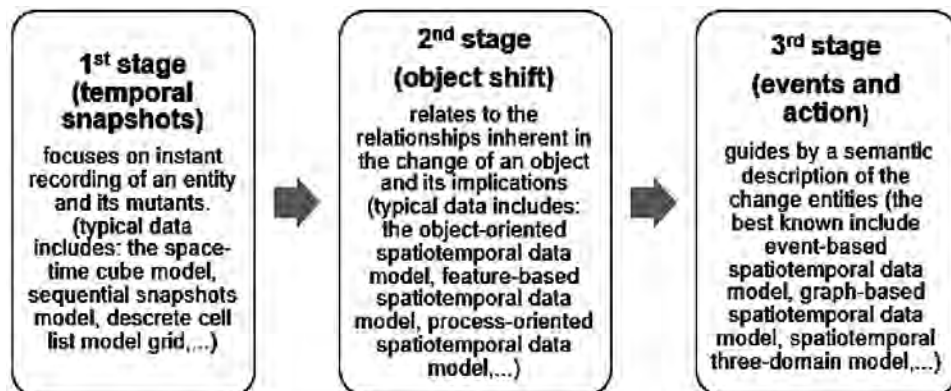
In recent years, attempts have been incremented to store and manage the activities of an individual based on their spatial and temporal characteristics (Chen et al., 2013; Wang & Cheng, 2001). This type of analysis has been mainly used in studies that assess individual accessibility (e.g., Delafontaine et al., 2011; Kwan, 1998; Miller, 1999; Neutens, 2015; Neutens et al., 2007). These studies demonstrate that the considerations of space-time contribute to the presentation of more complex models and real human activities (Shaw & Yu, 2009).

Real-Time GIS model was assumed as a new paradigm of information science to capture the real characteristics of human undertakings (Hey, 2012) and transforming historical changed data to real-time data (Gong et al., 2015). The authors divided the model into three stages: temporal snapshots (1st stage), object-shift (2nd stage) and events and action (3rd stage) (Figure 2).

Real-Time GIS analyses have also sought to incorporate collaborative functions. These types

Figure 2. Three stages of real-time models

Source: (Adapted from Gong et al., 2015).



of tools can be differentiated into several types, namely:

- The same time - same place;
- Same time - different location;
- Different time - same place;
- Different time - different location (Sun & Li, 2015).

The Real-Time Collaborative Geographic Information System (RCGIS) enables analysis of interactions in agile and flexible systems, equipped with collaborative principles.

The unprecedented growth of geographically referenced information combined with recent digital augmentation reality (AR) of places growth will become increasingly important in the future (Graham et al., 2013). AR had its recognition in 1992, when Caudell and Mizzel developed works for Boeing and designed a digital transport display in the head, so as to enable a framework of airplane schemes (Yew et al., 2016). Yet, the concept of AR is much older. This was used during the period of World War II with a project developed with the presentation of information on the windshield of the camera. In this statement, there must be a relationship between reality and the information made available in digital media. To this end, there must be a technological device (smartphone or other wireless equipment), track-

ing and computer software. AR is summarized in three distinct properties:

1. Combination of real and virtual objects in the real world;
2. Run interactively in real time; and
3. Registration of real and virtual objects and their connection (Azuma, 1997).

AR is receiving several applications for PC, smartphones, tablet and other devices and will be increasing in the future.

These advances in various types of GIS technologies can create significant digital divides, disadvantaging the poor, ethnic and racial minorities, residents in rural areas, the inhabitants of the global South. The own readiness for open apps coding creates exclusions face to those who can only view and contribute to these features (Elwood, 2009).

FUTURE RESEARCH DIRECTIONS

Information technology and Geographic Science are growing abreast and rapidly. Developments of GIS technology and applications must grow behind the scope of Big Data, Cloud GIS, Real-time GIS and Augmented Reality challenges. For future research, a more rigorous approach should

be implemented and guided by technology interoperability, integrated multidisciplinary approach, security, how to benefit and integrate GIS database with citizen-collected data and deeply understand how to collect and analyse real time data. The advent of data needed will be exceeded by the abundance of real time data arising the challenges on how to detach different sources to canvass their quality and include them in spatial analysis. For this more deeply and integrated analysis models are required and at the same time more spread applications will be required to provide the spatial information by multiple technological devices.

CONCLUSION

The expansion and advances of GIS technologies created conditions to proliferate different approaches to work in multiple areas, such as geography, cartography, remote sensing, image processing, education or environmental sciences. This section presents the evolution of GIS from its conception to the present. Moreover, it shows the main challenges posed by multiple skills acquired by the GIS in recent decades, namely presenting a shift from traditionally confined public planning areas to multi collaborate users, from desktop to the web and from real to virtual and augmented reality. In fact, maps have always been used for the removal of political borders, but today the networking capabilities generated conditions for GIS intelligence statement. The development of the web has supported the new challenges associated with GIS, particularly in areas connected with augmented reality, the real-time GIS information, GIS space-time or Cloud GIS. Associated to this arise Big Data challenges and problems to the Geographic Science (GS). The cumulative advances in the relationship between GIS and Web can hence contribute to the expansion of information generated and treated, interoperability between servers and users and networking. Geographic information will reinforce their position in our day's life and for GS those trends will catch much more research than ever.

REFERENCES

- Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A., & Zaharia, M. et al. (2010). A view of cloud computing. *Communications of the ACM*, 53(4), 50–58. doi:10.1145/1721654.1721672
- Assunção, M. D., Calheiros, R. N., Bianchi, S., Netto, M. A. S., & Buyya, R. (2015). Big Data computing and clouds: Trends and future directions. *Journal of Parallel and Distributed Computing*, 79–80, 3–15. doi:10.1016/j.jpdc.2014.08.003
- Azuma, R. T. (1997). A survey of augmented reality. *Presence (Cambridge, Mass.)*, 6(4), 355–385. doi:10.1162/pres.1997.6.4.355
- Bernhardsen, T. (2002). *Geographic information systems: an introduction* (3rd ed.). New York: John Wiley & Sons.
- Blower, J. D. (2010). *GIS in the cloud: implementing a web map service on Google App Engine*. Paper presented at the 1st International Conference and Exhibition on Computing for Geospatial Research & Application. doi:10.1145/1823854.1823893
- Bruno, G., & Giannikos, I. (2015). *Location and GIS. In Location Science* (pp. 509–536). Springer.
- Budhathoki, N. R., & Nedovic-Budic, Z. (2008). Reconceptualizing the role of the user of spatial data infrastructure. *GeoJournal*, 72(3-4), 149–160. doi:10.1007/s10708-008-9189-x
- Buyya, R., Yeo, C. S., Venugopal, S., Broberg, J., & Brandic, I. (2009). Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility. *Future Generation Computer Systems*, 25(6), 599–616. doi:10.1016/j.future.2008.12.001
- Chakraborty, D., Sarkar, D., Agarwal, S., Dutta, D., & Sharma, J. R. (2015). Web Based GIS Application using Open Source Software for Sharing Geospatial Data. *International Journal of Advanced Remote Sensing and GIS*, 4(1), 1224–1228.

- Chang, Y. S., & Park, H. D. (2006). XML Web Service-based development model for Internet GIS applications. *International Journal of Geographical Information Science*, 20(4), 371–399. doi:10.1080/13658810600607857
- Chen, B. Y., Li, Q., Wang, D., Shaw, S.-L., Lam, W. H., Yuan, H., & Fang, Z. (2013). Reliable space–time prisms under travel time uncertainty. *Annals of the Association of American Geographers*, 103(6), 1502–1521. doi:10.1080/00045608.2013.834236
- Chen, J., Shaw, S.-L., Yu, H., Lu, F., Chai, Y., & Jia, Q. (2011). Exploratory data analysis of activity diary data: A space–time GIS approach. *Journal of Transport Geography*, 19(3), 394–404. doi:10.1016/j.jtrangeo.2010.11.002
- Crampton, J. W. (2009). Cartography: Maps 2.0. *Progress in Human Geography*, 33(1), 91–100. doi:10.1177/0309132508094074
- Crossland, M. D. (2005). Geographical Information Systems as Decision Tools. In M. Khosrow-Pour (Ed.), *Encyclopedia of Information Science and Technology* (1st ed.; Vol. 2, pp. 1274–1277). Idea Group Reference. doi:10.4018/978-1-59140-553-5.ch224
- Delafontaine, M., Neutens, T., Schwanen, T., & Van de Weghe, N. (2011). The impact of opening hours on the equity of individual space–time accessibility. *Computers, Environment and Urban Systems*, 35(4), 276–288. doi:10.1016/j.compen-vurbsys.2011.02.005
- Elwood, S. (2009). Integrating participatory action research and GIS education: Negotiating methodologies, politics and technologies. *Journal of Geography in Higher Education*, 33(1), 51–65. doi:10.1080/03098260802276565
- ESRI. (2013). *The language of spatial analysis*. New York: ESRI Press.
- Faiz, S., & Krichen, S. (2012). *Geographical information systems and spatial optimization*. CRC Press.
- Gantz, J., & Reinsel, D. (2011). Extracting value from chaos. *IDC iview*, 1142, 1–12.
- Giuliani, G., Guigoz, Y., Lacroix, P., Ray, N., & Lehmann, A. (2016). Facilitating the production of ISO-compliant metadata of geospatial datasets. *International Journal of Applied Earth Observation and Geoinformation*, 44, 239–243. doi:10.1016/j.jag.2015.08.010
- Gong, J., Geng, J., & Chen, Z. (2015). Real-time GIS data model and sensor web service platform for environmental data management. *International Journal of Health Geographics*, 14(1), 1. doi:10.1186/1476-072X-14-2 PMID:25572659
- Goodchild, M. (2006). Geographic Information Systems. In S. Aitken & G. Valentine (Eds.), *Approaches to Human Geography* (Vol. 251–262). SAGE Publications. doi:10.4135/9781446215432.n24
- Goodchild, M. (2007). Citizens as sensors: The world of volunteered geography. *GeoJournal*, 69(4), 211–221. doi:10.1007/s10708-007-9111-y
- Goodchild, M. F. (2013). Prospects for a space–time GIS: Space–time integration in geography and GIScience. *Annals of the Association of American Geographers*, 103(5), 1072–1077. doi:10.1080/00045608.2013.792175
- Goodchild, M. F., Yuan, M., & Cova, T. J. (2007). Towards a general theory of geographic representation in GIS. *International Journal of Geographical Information Science*, 21(3), 239–260. doi:10.1080/13658810600965271
- Graham, M., Zook, M., & Boulton, A. (2013). Augmented reality in urban places: Contested content and the duplicity of code. *Transactions of the Institute of British Geographers*, 38(3), 464–479. doi:10.1111/j.1475-5661.2012.00539.x
- Grossner, K. E., Goodchild, M. F., & Clarke, K. C. (2008). Defining a digital earth system. *Transactions in GIS*, 12(1), 145–160. doi:10.1111/j.1467-9671.2008.01090.x

- Hägerstraand, T. (1970). What about people in regional science? *Papers in Regional Science*, 24(1), 7–24. doi:10.1111/j.1435-5597.1970.tb01464.x
- Hägerstrand, T. (1989). *Reflections on “what about people in regional science?”*. Paper presented at the Regional Science Association. doi:10.1111/j.1435-5597.1989.tb01166.x
- Haklay, M. (2010). How Good is Volunteered Geographical Information? A Comparative Study of OpenStreetMap and Ordnance Survey Datasets. *Environment and Planning. B, Planning & Design*, 37(4), 682–703. doi:10.1068/b35097
- Haklay, M., Singleton, A., & Parker, C. (2008). Web mapping 2.0: The neogeography of the GeoWeb. *Geography Compass*, 2(6), 2011–2039. doi:10.1111/j.1749-8198.2008.00167.x
- Haklay, M., & Zafiri, A. (2008). Usability engineering for GIS: Learning from a screenshot. *The Cartographic Journal*, 45(2), 87–97. doi:10.1179/174327708X305085
- Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Ullah Khan, S. (2015). The rise of big data on cloud computing: Review and open research issues. *Information Systems*, 47, 98–115. doi:10.1016/j.is.2014.07.006
- Hendriks, P. H. (2005). Space opera-GIS basics. In M. Khosrouw-Pour (Ed.), *Encyclopedia of Information Science and Technology* (Vol. 5, pp. 2571–2575). Idea Group Reference. doi:10.4018/978-1-59140-553-5.ch456
- Hey, T. (2012). *The Fourth Paradigm—Data-Intensive Scientific Discovery*. In *E-Science and Information Management* (pp. 1–1). Springer.
- Heywood, I., Cornelius, S., & Carver, S. (2011). *An Introduction to geographical Information systems* (4th ed.). Essex, UK: Pearson Education.
- Kerski, J. J. (2015). Geo-awareness, Geo-enablement, Geotechnologies, Citizen Science, and Storytelling: Geography on the World Stage. *Geography Compass*, 9(1), 14–26. doi:10.1111/gec3.12193
- Kwan, M.-P. (1998). Space-time and integral measures of individual accessibility: A comparative analysis using a point-based framework. *Geographical Analysis*, 30(3), 191–216. doi:10.1111/j.1538-4632.1998.tb00396.x
- Lee, Y.-T., & Chen, K.-T. (2010). *Is server consolidation beneficial to MMORPG? A case study of World of Warcraft*. Paper presented at the Cloud Computing (CLOUD), 2010 IEEE 3rd International Conference. doi:10.1109/CLOUD.2010.57
- Lin, H., Chen, M., Lu, G., Zhu, Q., Gong, J., You, X., & Hu, M. et al. (2013). Virtual Geographic Environments (VGEs): A New Generation of Geographic Analysis Tool. *Earth-Science Reviews*, 126, 74–84. doi:10.1016/j.earscirev.2013.08.001
- Marston, S., Li, Z., Bandyopadhyay, S., Zhang, J., & Ghalsasi, A. (2011). Cloud computing—The business perspective. *Decision Support Systems*, 51(1), 176–189. doi:10.1016/j.dss.2010.12.006
- Mekonnen, A. D., & Gorsevski, P. V. (2015). A web-based participatory GIS (PGIS) for offshore wind farm suitability within Lake Erie, Ohio. *Renewable & Sustainable Energy Reviews*, 41, 162–177. doi:10.1016/j.rser.2014.08.030
- Miller, H. J. (1991). Modelling accessibility using space-time prism concepts within geographical information systems. *International Journal of Geographical Information System*, 5(3), 287–301. doi:10.1080/02693799108927856
- Miller, H. J. (1999). Measuring space-time accessibility benefits within transportation networks: Basic theory and computational procedures. *Geographical Analysis*, 31(1), 1–26. doi:10.1111/j.1538-4632.1999.tb00408.x
- Miller, H. J. (2004). Toblers first law and spatial analysis. *Annals of the Association of American Geographers*, 94(2), 284–289. doi:10.1111/j.1467-8306.2004.09402005.x
- Miller, H. J., & Goodchild, M. F. (2015). Data-driven geography. *GeoJournal*, 80(4), 449–461. doi:10.1007/s10708-014-9602-6

- Neutens, T. (2015). Accessibility, equity and health care: Review and research directions for transport geographers. *Journal of Transport Geography*, 43, 14–27. doi:10.1016/j.jtrangeo.2014.12.006
- Neutens, T., Witlox, F., Van De Weghe, N., & De Maeyer, P. (2007). Space–time opportunities for multiple agents: A constraint-based approach. *International Journal of Geographical Information Science*, 21(10), 1061–1076. doi:10.1080/13658810601169873
- Parker, R., & Dominguez, J. (2015). Features to Look for in a GIS Viewer. *GeoInformatics*, 18(4), 24.
- Pourabbas, E. (2014). *Geographical Information Systems: Trends and Technologies*. Boca Raton, FL: CRC Press. doi:10.1201/b16871
- Ren, F., & Kwan, M.-P. (2009). The impact of the Internet on human activity–travel patterns: Analysis of gender differences using multi-group structural equation models. *Journal of Transport Geography*, 17(6), 440–450. doi:10.1016/j.jtrangeo.2008.11.003
- Schäffer, B., Baranski, B., & Foerster, T. (2010). Towards spatial data infrastructures in the clouds. In *Geospatial Thinking* (pp. 399–418). Springer.
- Shaw, S.-L., & Yu, H. (2009). A GIS-based time-geographic approach of studying individual activities and interactions in a hybrid physical–virtual space. *Journal of Transport Geography*, 17(2), 141–149. doi:10.1016/j.jtrangeo.2008.11.012
- Sun, Y., & Li, S. (2015). Real-time collaborative GIS: A technological review. *ISPRS Journal of Photogrammetry and Remote Sensing*.
- Tomlinson, R. (1967). *An introduction to the geo-information system of the Canada Land Inventory*. Ottawa, Department of Forestry and Rural Development, ARDA.
- Wang, D., & Cheng, T. (2001). A spatio-temporal data model for activity-based transport demand modelling. *International Journal of Geographical Information Science*, 15(6), 561–585. doi:10.1080/13658810110046934
- Yang, C., Goodchild, M., Huang, Q., Nebert, D., Raskin, R., Xu, Y., & Fay, D. et al. (2011). Spatial cloud computing: How can the geospatial sciences use and help shape cloud computing? *International Journal of Digital Earth*, 4(4), 305–329. doi:10.1080/17538947.2011.587547
- Yang, C., Raskin, R., Goodchild, M., & Gahegan, M. (2010). Geospatial cyberinfrastructure: Past, present and future. *Computers, Environment and Urban Systems*, 34(4), 264–277. doi:10.1016/j.compenvurbsys.2010.04.001
- Yang, C., Xu, Y., & Nebert, D. (2013). Redefining the possibility of digital Earth and geosciences with spatial cloud computing. *International Journal of Digital Earth*, 6(4), 297–312. doi:10.1080/17538947.2013.769783
- Yew, A., Ong, S., & Nee, A. (2016). Towards a griddable distributed manufacturing system with augmented reality interfaces. *Robotics and Computer-integrated Manufacturing*, 39, 43–55. doi:10.1016/j.rcim.2015.12.002
- Yu, G., Zhang, S., Yu, Q., Fan, Y., Zeng, Q., Wu, L., & Zhao, P. (2014). Assessing ecological security at the watershed scale based on RS/GIS: A case study from the Hanjiang River Basin. *Stochastic Environmental Research and Risk Assessment*, 28(2), 307–318. doi:10.1007/s00477-013-0750-x
- Yu, M., Fu, P., Zhou, N., & Xia, J. (2014). ArcGIS in the cloud. In Z. Li, C. Xu, & K. Liu (Eds.), *Spatial Cloud Computing. A practical approach*. Boca Raton, FL: CRC Press.
- Yue, P., Zhou, H., Gong, J., & Hu, L. (2013). Geoprocessing in Cloud Computing platforms—a comparative analysis. *International Journal of Digital Earth*, 6(4), 404–425. doi:10.1080/17538947.2012.748847

ADDITIONAL READING

Baker, T. R., Battersby, S., Bednarz, S. W., Bodzin, A. M., Kolvoord, B., Moore, S., & Uttal, D. et al. (2015). A research agenda for geospatial technologies and learning. *The Journal of Geography*, 114(3), 118–130. doi:10.1080/00221341.2014.950684

Bhat, M. A., Shah, R. M., & Ahmad, B. (2011). Cloud Computing: A solution to Geographical Information Systems (GIS). *International Journal on Computer Science and Engineering*, 3(2), 594–600.

Bodenhamer, D. J. (2013). Beyond GIS: Geospatial technologies and the future of history. In *History and GIS* (pp. 1–13). Netherlands: Springer. doi:10.1007/978-94-007-5009-8_1

Chakraborty, D., Sarkar, D., Agarwal, S., Dutta, D., & Sharma, J. R. (2015). Web Based GIS Application using Open Source Software for Sharing Geospatial Data. *International Journal of Advanced Remote Sensing and GIS*, 4(1), 1224–1228.

Chen, N., Zhang, X., & Wang, C. (2015). Integrated open geospatial web service enabled cyber-physical information infrastructure for precision agriculture monitoring. *Computers and Electronics in Agriculture*, 111, 78–91. doi:10.1016/j.compag.2014.12.009

Crampton, J. W. (2009). Cartography: Maps 2.0. *Progress in Human Geography*, 33(1), 91–100. doi:10.1177/0309132508094074

Crampton, J. W., Graham, M., Poorthuis, A., Shelton, T., Stephens, M., Wilson, M. W., & Zook, M. (2013). Beyond the geotag: Situating ‘big data’ and leveraging the potential of the geoweb. *Cartography and Geographic Information Science*, 40(2), 130–139. doi:10.1080/15230406.2013.777137

Gitis, V., Derendyaev, A., & Weinstock, A. (2016). Web-based GIS Technologies for Monitoring and Analysis of Spatio-Temporal Processes. *International Journal of Web Information Systems*, 12(1), 102–124. doi:10.1108/IJWIS-10-2015-0032

Hedley, M. L., Templin, M. A., Czajkowski, K., & Czerniak, C. (2013). The use of geospatial technologies instruction within a student/teacher/scientist partnership: Increasing students geospatial skills and atmospheric concept knowledge. *Journal of Geoscience Education*, 61(1), 161–169. doi:10.5408/11-237.1

Yang, C., Wong, D. W., Yang, R., Kafatos, M., & Li, Q. (2005). Performance-improving techniques in web-based GIS. *International Journal of Geographical Information Science*, 19(3), 319–342. doi:10.1080/13658810412331280202

KEY TERMS AND DEFINITIONS

Augmented Reality: Augmented Reality (AR) into GIS assures the link between the perception of user and the relationship with the real world. The real world is represented with 2D and 3D virtual information. The computer augments the actual landscape with additional information that can be supported by inserting fields based on GIS applications.

Crowdsourcing: Crowdsourcing is an act of performing a GIS task by a user on a voluntary basis for a set of users. This type of action is based on a bottom-up approach. It is associated with the creation of data through a group dynamic. The crowdsourced data collection is carried out using portable devices (GPS, PC, mobile phones) and the data is synchronized in the central database, accessible and shareable, based on services and maps on the web.

GIS in the Cloud: GIS in the cloud or cloud GIS bases on the integrated web systems. The data generates maps to support the analysis and optimization of real time operations. The cloud integration helps the organization of complex workflows and maintaining extensive geodatabases.

GIS: GIS is a system that permits to visualize, analyze, display and understand the relationships between spatial phenomena. Nowadays, GIS is capable to transform large numbers of data, to

analyze and transform momentarily alternate data and generate charts, graphs, summary and descriptive statistics. Among the main key elements to a noble GIS it is noted: computer hardware and software, operational context (people and organizations) and internet service.

Real-Time: Real-time is a term often used to describe the time of execution of a task. This tool helps the users to obtain a frequently monitoring and more efficiently. GIS technology enables the sharing of a series of real-time data. Among the main features is visualization, analysis and understanding of phenomena in reduced timescale.

Pace-Time: Space-time is suggested in Hägerstrand time-geographic framework. Presents and analyses the individual activities in time and

space dimensions. In GIS environment results in the spatial representation of the dimension x and y and temporal dimensions of time in hours, minutes or seconds. The space-time patch is used for the implementation of the daily trajectory of the individual in time and space.

Web-Based GIS: Web-based GIS is based on a type of distributed information. This set of technological services is part of a communication structure between the GIS server and the client. Their relationship is expressed through URLs (created by the server) and HTTP (for the customer). Spatial data access, advanced mapping and spatial analysis are the most common type of analysis options in Web-based GIS.

Geospatial Influence in Science Mapping



Carlos Granell-Canut

Universitat Jaume I of Castellón, Spain

Estefanía Aguilar-Moreno

Universitat Jaume I of Castellón, Spain

INTRODUCTION

Sharing and making research data publicly available are increasingly getting attention to academia and research policy agenda. According to a 2014 European Commission guidelines for the re-use of datasets (OJEU C/240-1), geospatial data is of most importance. It raises an invaluable opportunity for libraries to play a dominant role in the not-so-distant-future for managing large collections of open (geospatial) research data. However, the geospatial dimension goes beyond data itself and embraces a wide range of spatial analysis and techniques (Smith et al., 2015). In particular, mapping and visualization techniques of geospatial data may provide endless opportunities to libraries and information science researchers in the sense of exploring the most of large open research datasets from a new perspective. With exceptions, librarians and information science professionals miss an overall perception of the possibilities that geospatial data and tools may bring them to geographically explore, analyze, and mapping research datasets and, especially, science related data. The question we pose in this chapter is whether or not geospatial technologies and mapping techniques have a role in the know-how of librarians and how these technologies and techniques may influence science mapping. Furthermore the chapter aims at drawing attention to the opportunities that the geospatial dimension applied to science related data can bring to the field of science mapping.

In what follows, the chapter briefly distinguishes the notion of mapping between the Geospatial

Information Science (GIScience) and Librarianship and Information Science (LIS). Afterwards, an overview about recent initiatives and research work relative to (geospatial) mapping of science is presented. Based on these examples, opportunities and challenges of applying geospatial technology to science mapping are discussed. Finally, based on relevant while evolving geospatial technologies, next steps for increasing up the influence of geospatial technology in science mapping are pointed out.

BACKGROUND

Concepts and Terminology

Science mapping, bibliographic mapping, or mapping scientific bibliography is often defined as a visual representation of how scientific disciplines and fields, authors and institutions, and scientific and technical documents and articles are related to each other (Cobo et al., 2011; Small, 1999). In order to define the aim and scope of the present chapter it is paramount to first clarify what actually science and mapping mean in the expression ‘science mapping’.

Firstly, the focus on ‘science’ in the term science mapping refers to all the data and metadata generated during the gathering and compilation of scientific bibliographic information such as: authors, article titles, source, citations, affiliation, and related scientific data (Chen, 2013). This data will be afterwards processed, analyzed and visualized using different scientometric techniques

(Garechana et al., 2012). It is worth noting that bibliography data is only one potential working area to which geospatial technology and GIScience can be applied. Other areas within LIS, such as geospatial segmentation of patrons for marketing purpose and indoor position inside library buildings for space management and for enhancing navigation of patrons, are also of interest to GIScience (Aguilar & Granell, 2013; Granell & Aguilar 2013; Scaramozzino et al., 2014; Aguilar & Granell, 2015).

The second, and most important, clarification refers to the term ‘mapping’, which may have different connotations from distinct fields and disciplines. As the chapter mixes ideas from two distinct fields or disciplines, namely LIS and GIScience, it is worth delimiting early the scope of ‘mapping’. As introduced earlier, science mapping or bibliometric maps of science, under the lens of the Librarianship discipline, is meant to visually represent bibliographic data relative to science. Noyons (2004) provides a clear definition of the resulting science maps in which “the items are positioned in relation to each other in such a way that the ones which are cognitively related to each other are positioned in each other’s vicinity, whilst the ones that are not or hardly related are distant from each other”. For example, we can have citation mappings, as visualizations of citation networks from scientific documents, or author mappings, as the analysis and visualization of collaboration (joint articles) among scientists. In general, as these mappings explore datasets relative to scientific activities and results, they are altogether referred to as science mapping.

In the current literature, though, the term “mapping” often refers to “record in detail the spatial distribution of (something)” (Oxford definition’s map - verb) (English Oxford Living Dictionaries, 2016), which refers to spatially arranging data over an area. For example, network diagrams using force-directed layout are common visualizations for science mapping (Boyack et al., 2015). Despite these visualizations of science maps evoke items displayed on a geographical map, such a

meaning of mapping has nothing to do with a geographic map or displaying data on a map. From the GIScience perspective, though, (geospatial) mapping implies to explicitly project data on a (geographical) map. For doing so, data must be georeferenced, in other words, data must contain a clear reference to a position or location, in order to be spatially displayed in a map. Furthermore, the main difference is that science mapping in LIS usually refers to visualizations that do not necessarily include or rely on a (geographic) map. Network graphs are typical examples of visually arranged bibliographic data because they emphasize the connectivity of data, i.e. network graphs provide an easy way to quickly grasp how items of data (e.g. authors, publications, citations, institutions, etc.) are connected. Furthermore, in this chapter, the term science mapping refers to ways of visually representing bibliographic data or science-related data, but paying special attention to technologies, concepts and analysis techniques from GIScience for creating such visualizations. Indeed, as we outline later on, the geospatial influence in the science mapping literature is still anecdotal.

Before going into the literature review, it is important to briefly introduce the concepts of geospatial analysis and visualization which are used extensively later on. Geospatial analysis is concerned with statistical and analytical techniques to process geospatial data, paying particular attention to topological, geometric, or geographic properties of the data. The typology of data (geospatial) calls for slightly different types of analytics than those found in the literature (Chen et al., 2012). In short, there exist two grand categories of spatial analysis techniques according to the dichotomy of geospatial data: vector-based or raster-based data. In the case of vector-based data, typical and basic operations are map overlay (combining two or more maps or map layers), and buffering (identifying regions of a map within a specified criteria (distance, time, etc.) of one or more geographic entities or features, such as buildings, streets, or town). Raster-based data, though, is more often used in disciplines such as environmental sciences

and remote sensing other than LIS. Nevertheless, analysis techniques for raster-based data still might be helpful for surface analysis. These raster-based techniques typically involve a range of actions (such as filtering and algebraic operations) applied to the grid cells (i.e. an image is often divided in regular individual units or cells that altogether form a grid) of a single raster layer (image) or multiple layers. These actions involve processing one or more raster layers according to some statistical rules leading to a new raster layer. For example a surface analysis computes each cell value by interpolating values from neighboring cells. Visualization thus complements geospatial analysis in that it centers on the use, creation and manipulation of images and maps to effectively transmit spatial information to end users (Slocum et al., 2008; Dodge et al., 2008).

Related Work

Science mapping is a sort of analytical process composed of well-defined tasks that include but are not limited to data access to sources, data preparation, computation of similarities measurement, analysis and visualization methods. Cobo et al. (2011) reviewed in detail each step of the science mapping process and provided pointers to further related work and literature for interested readers. Cobo et al. (2011) also compared various software tools that are aimed to perform mapping science analysis. The authors identified only three tools (out of nine tools) – Science of Science (Sc2) Tool, CiteSpace, and VantagePoint – that to some extent supported some basic types of geospatial analysis and visualizations. What is relevant to this chapter is whether (or not) geospatial analysis and visualization techniques are used in science mapping, which could make a big difference in yielding novel insights during the visualization and interpretation of such science maps.

Geospatial techniques commonly used are merely geocoding and geospatial thematic maps. While geocoding is the task of attaching a location and/or position to a piece of data to turn it

in spatial data, geospatial thematic maps result from visualizing one or more spatial datasets over a world or base map (e.g. Figure 1). Scimago Institution Ranking web site provides data visualizations of bibliometric indicators such as international collaboration and normalization impact of worldwide research in government and higher education institutions. Figures 1 and 2 show some Scimago available visualizations, ranging from simple geospatial mapping visualizations (e.g. choropleth maps, i.e. a sort of variation of thematic maps) to more traditional, spatial representation in tabular form like distribution charts and scatterplots.

Based on this Scimago data set, researchers have explored the distinct geospatial signatures among countries and institutions attending to indicators like international collaboration and leadership (Maganote et al., 2014). The authors used scatterplots to explain how geography matters in finding unique patterns derived from bibliometric indicators taking into account distinct geographic regions to show how international collaborations behave differently from regions to regions. Similarly, Pan et al. (2012) studied the geography of worldwide citation networks using a data set other than Scimago data. The article tries to validate the Tobler's first law of geography – “Everything is related to everything else, but near things are more related than distant things” (Tobler, 1970) – using citations networks and collaborations between scientists, without the authors explicitly mentioned it. Their hypothesis is that scientists tend to interact more frequently with other scientists who work in nearby areas forming clusters based on geographical proximity. In a similar vein, Chen (2006) described the CiteSpace toolset (<http://cluster.cis.drexel.edu/~cchen/citespace/>) for visualizing and analyzing trends and patterns in scientific data, especially to detect cluster and citations hotspots (Chen et al., 2010).

It is worth noting the recent work conducted by the Semantic Web journal (SWJ) by IOS Press to set up a modular, linked data-driven web portal for data relative to the journal activities, scientific

Figure 1. Choropleth map to categorize countries in function of the number of institutions from the Scimago Institution Ranking (Source: Scimago Institutions Rankings, n.d.).

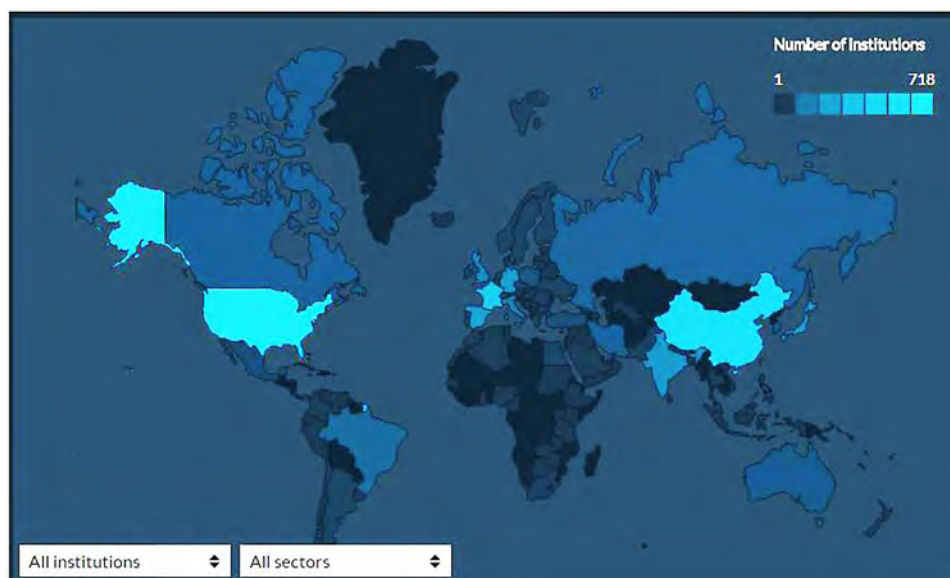
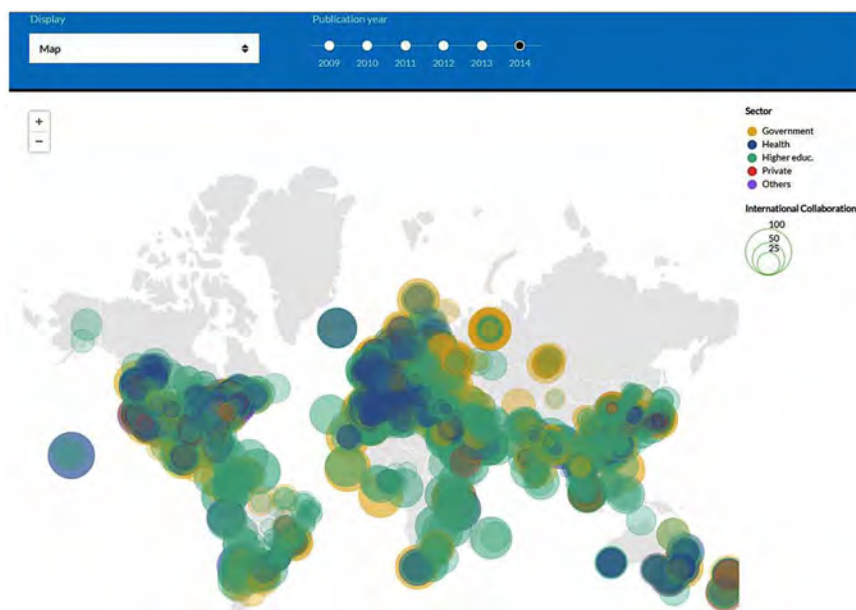


Figure 2. Map display for international collaboration indicator (Source: <http://bit.ly/1RdW7IH>)



papers, authors, editors, and reviewers (Hu et al., 2013). Apart from the fact that the web portal is semantically-enabled, i.e. all data is linked and encoded using semantic data formats and exposed through an SPARQL endpoint, what is relevant in the context of the chapter is the inclusion of

spatio-temporal visualization to better understand how authors, co-authors, and authors citations are geographically related to each other in order to for example detect hot or densely connected regions in science (Bornmann & Waltman, 2011). Figure 3 shows some types of visualizations provided

by the SWJ Web Portal. Left part of Figure 3 appears a cartogram that takes a location and resizes (distorts) the geographic area according to a value of indicator and/or variable being analyzed. Each nation's size gives a qualitative indicator of the geographical distribution of contributing authors to the SWJ journal. Cartogram are widely recognized as geographically-influenced data visualizations in information science (Olmeda-Gómez, 2014).

In the particular case of scientific publications, bibliographic reference management tools are useful to help users collect, categorize, and annotate their references of interest, which is a usual service offered by libraries. Also, cloud-based reference management services that incorporate social aspects (sharing, ranking or scoring references) like Mendelev or ReadCube are gaining momentum over the past years. Even traditional desktop solutions like Endnote have recently launched its counterpart cloud-based tools. A remarkable research done to let users browse, navigate and query a collection of scientific datasets through spatial and time dimensions is TimeBibliography (Siabato et al, 2013; Siabato et al., 2014). It is a

dynamic web tool that hosts bibliography data relative to temporal GIS topics. As depicted in Figure 4, it provides novel ways to browse an online repository through dynamic and innovative visualizations that attempt to exploit temporal and spatial attributes for each bibliography metadata record, amongst others. For example diverse types of timeliness visualizations to explore long and short periods of time are integrated into the user interface (upper part of Figure 4).

Finally, Field and O'Brien (2010) introduced the spiral spatio-temporal timeline visualization of geo-located tweets, where size of symbols larger denote recent tweets, the same color identify common words, and the location of individual markers in spiral denotes spatial association with the real origin. This experimental Twitter Map visualizations that use connected graphs may be a starting point to enable more sophisticated and informative network-based visualizations. Again space and time turn out to be key contextual information to better understand the dynamics of a connected network of people (or citations, papers, tweets, etc.)

Figure 3. Linked data portal for the Semantic Web Journal (Source: Linked Data Portal for the Semantic Web Journal, 2010)

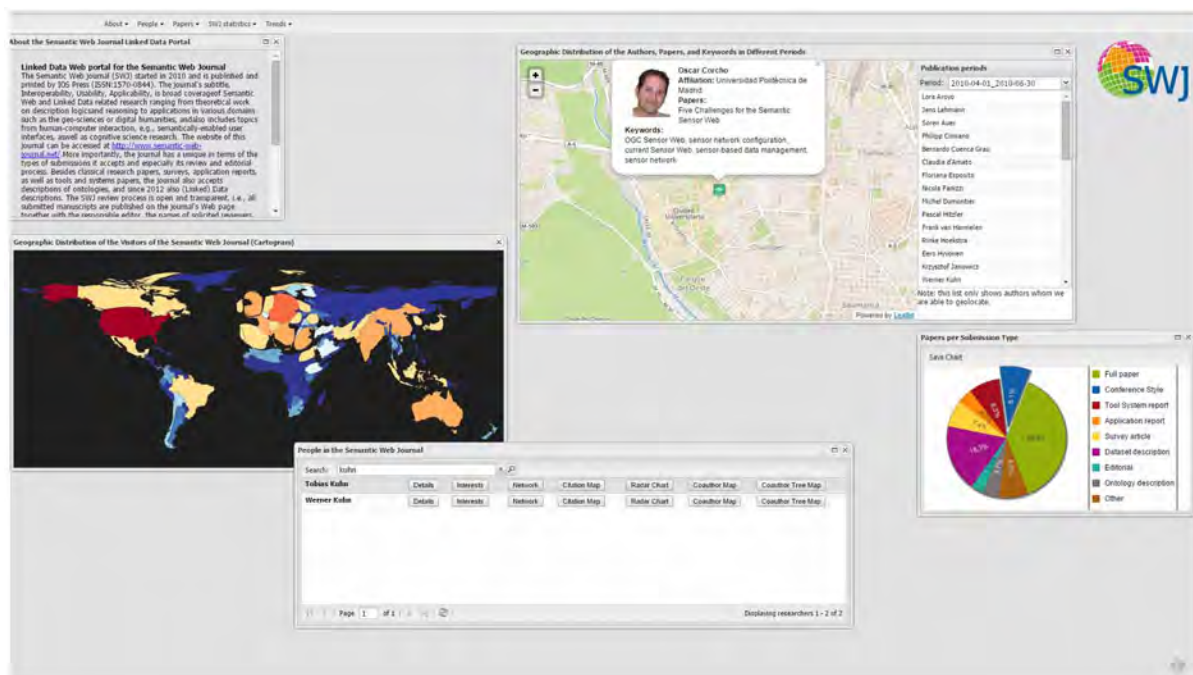


Figure 4. TimeBibliography web tool (Siabato et al., 2014) (Source: TimeBibliography, n.d.)



GEOSPATIAL INFLUENCE IN SCIENCE MAPPING

Despite the wide use of geographic terminology such as mapping or maps in bibliometric and scientometric, geospatial analysis and visualization methods have been largely unexplored in the literature. The brief overview described earlier presents interesting but incipient examples that attempt to include the geospatial dimension in mapping scientific bibliography data. It seems clear that there is a niche for applying geospatial aspects in science mapping, and it needs further exploration. The process of science mapping could take advantage of a wide range of geospatial analysis and visualization techniques so as to reveal new patterns and insights. Therefore, to suggest new areas where geospatial aspects can leverage science mapping to a new level, we rely on the concept of cybermetrics. Orduña-Malea and Aguilló (2015)

provided a holistic view of the academic field of cybermetrics. The authors identified three main working areas within cybermetrics:

- Descriptive cybermetrics,
- Instrumental cybermetrics, and
- Applied cybermetrics.

For the sake of clarity in the exposition, this classification is taken in order to come up with unexplored uses of geospatial related techniques in science mapping.

Descriptive cybermetrics refers mainly to the definition and modelling of cybermetrics indicators. It is well beyond the scope of the chapter to discuss existing indicators, their classification and scope in the concept of cybermetrics or science mapping. What is somehow surprising is that, despite recent studies highlight the importance and relevance of geography in science mapping

(Pan et al., 2012), geospatial characteristics has still poorly influence when defining bibliographic or cybermetrics indicators. Such geographically-aware indicators may give a completely new perspective of the underlying data being analyzed and visualized, and thus could serve for multiple purposes, ranging from professional career and project evaluation processes to policy making processes, as well as to even estimate the geospatial influence of an individual researcher, a scientific institution, or groups of researchers and institutions. In particular, the geospatial influence of an individual scientists has been recently explored by Zhan et al. (2014) for example. The authors devised the GeoSI index, which computes the number of countries or regions where researchers have cited a scholar's work during a given period of time, to get an estimation of the geographic coverage of the scientific influence of an individual scholar. Through the GeoSI index, the authors also examined how citations of a scholar have expanded to different countries over time, leading to the geographical influence or penetration of the research topics of one's work.

As instrumental cybermetrics is concerned with the methodological steps to obtain and process data to conveniently assess the feasibility of cybermetrics indicators, which is out of the scope of the chapter, the focus is shifted to the third and last area, applied cybermetrics, which in turn relies on both descriptive and instrumental cybermetrics.

Applied cybermetrics refers to the application of indicators (modelled by descriptive cybermetrics) combined with data analysis and visualization techniques, and altogether applied to concrete scenarios, use cases, and application domains. Scientific bibliography in general has been traditionally a recurrent domain where applied cybermetrics has been extensively used and explored. In the related literature concerning applied cybermetrics and academia, geographic network diagrams are used to represent patterns over a base map (Hu et al., 2013). Such geographic network diagrams combine network analysis methods (Boyack et al., 2005), com-

monly used for links analysis being for example co-authorship or citations those links, with geospatial visualizations by projecting these linked networks into a geographic map. This permits the detection of strong and weak geospatial connections among geographical areas, regions and nations. Unfortunately, geospatial visualizations in science mapping are practically constrained to geographic network diagrams, cartograms, and geospatial thematic maps. A similar deficit occurs in geospatial analysis, since geocoding is the most geospatial method used (for preprocessing data) in science mapping projects. There exist a great deal of untapped spatial analysis methods for the mapping science community (Smith et al., 2015), which for example combined with link analysis, network analysis, or data mining techniques, could most likely bring rich spatial contexts to science mapping in form of innovative geospatial data visualizations and explorations.

Apart from the lack of spatial techniques for data analysis and visualization, those few science maps that exploit spatial dimensions in bibliographic data do not consider interactive and dynamic feature in data visualizations. This prevents researchers, patrons, governors, and administrators from analyzing, understanding, and visualizing scientific maps and networks from a spatio-temporal perspective, which in the end pursues to improve the communication of information through maps.

FUTURE RESEARCH DIRECTIONS

Interactive, dynamic, and spatio-temporal visualizations are extremely useful in science mapping. Nevertheless preparing and processing geospatial data to fit the purpose of science mapping projects are time-consuming tasks. The value would not remain only in the set of references or bibliographic data, but in the availability of right tools, as spatio-temporal visualizations based on maps or other kind of charts, to enable visual analysis to inform users on the temporal, spatial, and topi-

cal variations of a given scientific topical area. In such a way, such tools could benefit researchers and scientists, as well as other collectives, as Noyons (2004) argued, such as funding agencies (to prepare research calls and evaluate research results, for example) and research policy makers (to establish topics priority and research plans in the long term).

The availability and efficiency of software tools depend strongly on the rapid changing pace of technology. Groenendyk (2013) provided an overview of emerging data geospatial visualization technologies which looked futuristic at first glance but which are partially being deployed in real projects. Holographic imaging and maps, 2-D city models, and 3-D printing were pointed by Groenendyk as key technologies for map libraries. Could existing 2-D network visualization for science mapping turn into immersive graph visualizations? Or even into physical 3-D models produced by 3D printers? Current technologies already open new opportunities for creative and innovative data visualizations for science mapping.

CONCLUSION

This chapter have presented some science mapping initiatives and research works that are being slightly influenced by geospatial technology. We argues that this trend have just started, as most of literature research works are exploratory research projects that merely grasp the surface of what geospatial technology might offer to science mapping analysis and visualization. As Field and O'Brien (2010) noted, "extracting and using the spatial and temporal dimensions reveal a contextual richness needed to truly represent and spatially organize meaningful information in a map". Location is key context, and such spatial context should be truly embedded in analytical and visualization processes for science mapping projects.

REFERENCES

- Aguilar-Moreno, E., & Granell-Canut, C. (2013). Sistemas de información geográfica para unidades de información [Geographic information systems for informatio units]. *Profesional de la Información*, 22(1), 80–86. doi:10.3145/epi.2013.ene.11
- Aguilar-Moreno, E., & Granell-Canut, C. (2015). *Geobibliotecas* [Geolibraries]. Barcelona: Editorial UOC.
- Bibliometrics. (2016). Retrieved from <https://en.wikipedia.org/wiki/Bibliometrics>
- Bornmann, L., & Waltman, L. (2011). The detection of hot regions in the geography of science – A visualization approach by using density maps. *Journal of Informetrics*, 5(4), 547–553. doi:10.1016/j.joi.2011.04.006
- Boyack, K. W., Klavans, R., & Börner, K. (2005). Mapping the backbone of science. *Scientometrics*, 64(3), 351–374. doi:10.1007/s11192-005-0255-6
- Cartogram. (2016). Retrieved from <https://en.wikipedia.org/wiki/Cartogram>
- Chen, C. (2006). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology*, 57(3), 359–377. doi:10.1002/asi.20317
- Chen, C. (2013). The Structure and Dynamics of Scientific Knowledge. In C. Chen (Ed.), *Mapping Scientific Frontiers: The Quest for Knowledge Visualization* (2nd ed.; pp. 163–199). London: Springer. doi:10.1007/978-1-4471-5128-9_5
- Chen, C., Ibekwe-SanJuan, F., & Hou, J. (2010). The structure and dynamics of co-citation clusters: A multiple-perspective co-citation analysis. *Journal of the American Society for Information Science and Technology*, 61(7), 1386–1409. doi:10.1002/asi.21309

- Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *Management Information Systems Quarterly*, 36(4), 1165–1188.
- Chloropleth Map. (2016). Retrieved from https://en.wikipedia.org/wiki/Choropleth_map
- Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2011). Science mapping software tools: Review, analysis, and cooperative study among tools. *Journal of the American Society for Information Science and Technology*, 62(7), 1382–1402. doi:10.1002/asi.21525
- de Smith, M., Goodchild, M. F., & Longley, P. (2015). *Geospatial Analysis: a comprehensive guide to principles, techniques and software tools* (5th ed.). Leicester, UK: Troubador Publishing.
- Dodge, M., McDerby, M., & Turner, T. (2008). *Geographic Visualization: Concepts, Tools and Applications*. Chichester, West Sussex, England: Wiley. doi:10.1002/9780470987643
- Edmilton, J. T., Maganote, M. S. A., & Schulz, P. A. (2014). Visualization of ranking data: Geographical signatures in international collaboration, leadership and research impact. *Journal of Informetrics*, 8(3), 642–649. doi:10.1016/j.joi.2014.05.005
- English Oxford Living Dictionaries. (2016). *Map*. Retrieved from <https://en.oxforddictionaries.com/definition/map>
- Field, K., & O'Brien, J. (2010). Cartoblography: Experiments in using and organizing the spatial context of micro-blogging. *Transactions in GIS*, 14(s1), 5–23. doi:10.1111/j.1467-9671.2010.01210.x
- Garechana, G., Río, R., Cilleruelo, E., & Gavilanes, J. (2012). Visualizing the scientific landscape using maps of science. In S. P. Sethi et al. (Eds.), *Industrial Engineering: Innovative Networks* (pp. 103–112). London: Springer. doi:10.1007/978-1-4471-2321-7_12
- Geographic information system. (2016). Retrieved from https://en.wikipedia.org/wiki/Geographic_information_system
- Granell-Canut, C., & Aguilar-Moreno, E. (2013). Se busca geobibliotecario: Los datos geográficos entran en la biblioteca [Geolibrarian wanted: geographic data come to the library]. *Profesional de la Información*, 22(6), 569–575. doi:10.3145/epi.2013.nov.10
- Groenendyk, M. (2013). Emerging data visualization technologies for map and geography libraries: 3-D printing, holographic imaging, 3-d city models, and 3-d model-based animations. *Journal of Map & Geography Libraries: Advances in Geospatial Information. Collections & Archives*, 9(3), 220–238.
- Hu, Y., Krzysztof, J., McKenzie, G., Sengupta, K., & Hitzler, P. (2013). A Linked-data-driven and semantically-enabled journal portal for scientometrics. *LNCS*, 8219, 114–129. doi:10.1007/978-3-642-41338-4_8
- Linked Data Portal for the Semantic Web Journal. (2010). Retrieved from <http://semantic-web-journal.com/SWJPortal/>
- Noyons, E. C. (2004). Science maps within a science policy context. In H. F. Moed et al. (Eds.), *Handbook of Quantitative Science and Technology Research* (pp. 237–255). New York: Kluwer.
- OJEU C. 240-1, 24/07/2014 (2014). *Commission notice: Guidelines on recommended standard licences, datasets and charging for the reuse of documents*. Retrieved April 10, 2015, from [http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52014XC0724\(01\)&from=EN](http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52014XC0724(01)&from=EN)
- Olmeda-Gómez, C. (2014). Visualización de información [Information visualization]. *El Profesional de la Información*, 23(3), 213–219. doi:10.3145/epi.2014.may.01
- Orduña-Malea, E., & Aguilló, I. F. (2014). *Cibermetría: midiendo el espacio red*. Barcelona: Editorial UOC.

Pan, R. K., Kaski, K., & Fortunato, S. (2012). World citation and collaboration networks: Uncovering the role of geography in science. *Scientific Reports*, 2, 902. doi:10.1038/srep00902 PMID:23198092

Scaramozzino, J., White, R., Essic, J., Fullington, L. A., Mistry, H., Henley, A., & Olivares, M. (2014). Map room to data and GIS services: Five university libraries evolving to meet campus needs and changing technologies. *Journal of Map & Geography Libraries*, 10(6), 6–47. doi:10.1080/15420353.2014.893943

Scientometrics. (2016). Retrieved from <https://en.wikipedia.org/wiki/scientometrics>

Scimago Institutions Rankings. (n.d.). Retrieved from <http://www.scimagoir.com/>

Siabato, W., Claramunt, C., & Bernabé-Poveda, M. A. (2013). An interactive bibliography on temporal GIS. *Journal of Spatial Information Science*, 7, 99–101.

Siabato, W., Claramunt, C., Manso-Callejo, M. Á., & Bernabé-Poveda, M. Á. TimeBliography. (2014, December). A dynamic and online bibliography on temporal GIS. *Transactions in GIS*, 18(6), 799–816. doi:10.1111/tgis.12080

Slocum, T. A., McMaster, R. B., Kessler, F. C., & Howard, H. H. (2008). *Thematic Cartography and Geovisualization* (3rd ed.). Pearson Prentice Hall.

Small, H. (1999). Visualizing science by citation mapping. *Journal of the American Society for Information Science*, 50(9), 799–813. doi:10.1002/(SICI)1097-4571(1999)50:9<799::AID-ASI9>3.0.CO;2-G

Spatial Analytics. (2016). Retrieved from https://en.wikipedia.org/wiki/spatial_analytics

TimeBliography. (n.d.). Retrieved from <http://spaceandtime.wsiabato.info/tGIS.html>

Tobler, W. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46(2), 234–240. doi:10.2307/143141

Zhan, F. B., Gong, X., & Xingjian, L. (2014). Mark on the globe: A quest for scientific bases of geographic information and its international influence. *International Journal of Geographical Information Science*, 28(6), 1233–1245. doi:10.1080/13658816.2013.847186

ADDITIONAL READING

Börner, K. (2010). *Atlas of Science: Visualizing What We Know*. MIT Press.

Börner, K. (2015). *Atlas of Knowledge: Anyone Can Map*. MIT Press.

Chen, C. (2013). *Mapping Scientific Frontiers: The Quest for Knowledge Visualization* (2nd ed.). London: Springer. doi:10.1007/978-1-4471-5128-9

Goodchild, M. F. (2010). Twenty years of progress: GIScience in 2010. *Journal of Spatial Information Science*, 1, <http://www.josis.org/index.php/josis/article/viewArticle/32>

KEY TERMS AND DEFINITIONS

Applied Cybermetrics: Applied cybermetrics refers to the application of indicators (modelled by descriptive cybermetrics) combined with data analysis and visualization techniques, and altogether applied to concrete scenarios, use cases, and application domains. Further information is available in Orduña-Malea and Aguilló (2014).

Bibliometrics: “Bibliometrics is statistical analysis of written publications, such as books or articles.” Bibliometric methods are frequently used to provide quantitative analysis of academic literature or for evaluating budgetary spending. “Citation analysis is a commonly used bibliometric method” (source: Bibliometrics, 2016).

Cartograms: A cartogram is a map in which some thematic mapping variable – such as travel time, population, or socio-economic variables – is substituted for land area or distance. The geometry

or space of the map is distorted in order to convey the information of this alternate variable (source: Cartogram, 2016).

Choropleth Maps: A choropleth map is a thematic map in which areas are shaded or patterned in proportion to the measurement of the statistical variable being displayed on the map, such as population density or per-capita income. The choropleth map provides an easy way to visualize how a measurement varies across a geographic area or it shows the level of variability within a region (source: Choropleth Map, 2016).

Geospatial Mapping: Geospatial mapping is a type of spatial analysis techniques that typically employs software capable of rendering maps processing spatial data, and applying analytical methods to terrestrial or geographic datasets, including the use of geographic information systems.

GIScience: Geographic information science (GIScience) is the scientific discipline that studies data structures and computational techniques

to capture, represent, process, and analyze geographic information. Goodchild (2010) summarized its core interests, including spatial analysis, visualization, and the representation of uncertainty (source: Geographical Information Systems).

Scientometric: Scientometrics is the study of measuring and analyzing science, technology and innovation. Major research issues include the measurement of impact, reference sets of articles to investigate the impact of journals and institutes, understanding of scientific citations, mapping scientific fields and the production of indicators for use in policy and management contexts (source: Scientometrics).

Spatial Analysis: Spatial analysis or geospatial analysis includes approaches to applying statistical analysis and other analytic techniques to data which has a geographical or spatial aspect, most notably in the analysis of entities using their topological, geometric, or geographic properties (source: Spatial Analysis, 2016).

Parallel Development of Three Major Space Technology Systems and Human Side of Information Reference Services as an Essential Complementary Method

Joyce Gosata Maphanyane

University of Botswana, Botswana

INTRODUCTION

Ever since the first human footprints were made more than 1,9 million years ago; the Homo erectus Era; ‘upright man’; socio-economic-Historical Development came into being. When humans could walk; it left their other limbs, the hands free to do other things. This Era was followed by the Stone Age; the man used stone tools to defend their territories and to hunt for food. This culminated to the Bronze Age and the Iron Age. These were the human development eras when man could actually work with minerals to make tools. These people could farm and rear domestic animals. The primitive smelting industries grew into the Industrial Revolution; The Industrial Revolution was a period from the 18th to the 19th century. The major changes in agriculture, manufacturing, mining, transport, technologies had a profound effect on the socio-economic and cultural conditions. This was followed by the Information Age; the Era we are in now. It is a shift from traditional industry that the previous era brought through industrialization. It is an economy based on the manipulation of information. The Information Age has allowed rapid global communications and networking to shape modern society. It is a digital world of Information Communication Technology (ICT). It is also commonly known as the Computer Age or Information Era, and it is an idea that the current age will be characterized by the ability of individuals to transfer information freely, and to have instant access to knowledge that would have

been difficult or impossible to find previously. The idea is linked to the concept of a Digital Age or Digital Revolution. This chapter elaborates upon Geographical Information Systems (GIS) and Remote Sensing; the highest echelon of the ICT world. It looks at their development in studying our home, the Earth and its systems. In particular it looks at, compares and contrasts the four globally used systems; these are:

- **GOES:** The Geosynchronous Orbiting Environmental Satellite;
- **LANDSAT,**
- **SPOT:** Satellite Pour l’Observation de la Terre and
- The WorldView.

The first three has transacted five generation and the last system, the WorldView is the newest and most fast developing satellite system. These Earth Resources Data capture satellite systems are compared in their longevity - temporal resolution and age factor which gives results as:

- **LANDSAT1 (ARTS 1) - LANDSAT 8** range July 1972 to February 1993;
- **GOES 1 - GOES 15,** Range October 1975 to March 2010;
- **SPOT1 – SPOT 7** range February 1986 to June 2014 and
- **WorldView 1 – WorldView 3,** range September 2007 to August 2014.

DOI: 10.4018/978-1-5225-2255-3.ch303

The comparison criteria were made based on spectral resolution (Figure 5 – WorldView 3); spatial resolution (Figure 6 – WorldView 3) and radiometric resolution (Figure 7 – SPOT). Apart from that; the human side of information reference services in the form of Traditional Ecological Knowledge (TEK) is discussed as it is an essential complementary entity in GIS and remote sensing endeavours. In order to be useful information, the remote sensing data need human input in the form of referencing coordinates system, data interpretation using the visual variables of position, shape, size, texture, tone, orientation, and motion. These can then be analyzed and used for modeling the environment, disaster preparedness and decision making. The 3 Dimensional characteristics of satellite, digital elevation models (DEM) (Figure 7 – SPOT DEM) are also investigated.

The Chapter is also about the essence of data sources specific for geo-spatial science (Bossler, Jensen, McMaster, & Rizos, 2002) information for land cover mapping. The use of geospatial science techniques provides opportunities and challenges in many aspects of life including for land cover, forestry for climatic change measurements and agricultural engineering which is vital for food security (Opara, 2003)

First, it elaborated about developments of digital world in remote sensing and geographical information system (GIS) as modern day techniques for Earth surface monitoring (Mather & Koch, 2011). Satellites whose data has widespread international applications for the past four decades are discussed in details and compared. Second, it puts emphasis on the importance of the human side of information reference services that is essentially the brain behind machine based data. The human thinking is applied in the interpretation of this machine data and it acts as knowledge substitute where the remote sensing and GIS data are inadequate or unavailable. The 2012 study (Maphanyane, 2012) on the reconstruction of historical landscape for the investigation of land cover changes in the human side of information reference services based method on Ramotswa,

Botswana case, had sufficiently proved that it can be applied to fill in the gap where modern technology is inadequate or unavailable (Figure 1 and Table 1). For emphasis, graphical examples are given, where the aerospace based Ramotswa digital data; its analytical techniques and their resultant information are compared to the same area data, but those that had been derived from the systems based on the human side of information reference services (Figure 2). This too has shown that the human side of information reference services was complementary.

BACKGROUND

The essence of this study hinges on technological developments of robust data collection tools. Remote sensing is one such. It shows the importance of human side of information reference services and the role it plays still, as it adds meaning and time independent data sources, even in the wake of superior tools of remote sensing.

Curiosity of Humankind: Their Cognitive Ability and Feature Naming

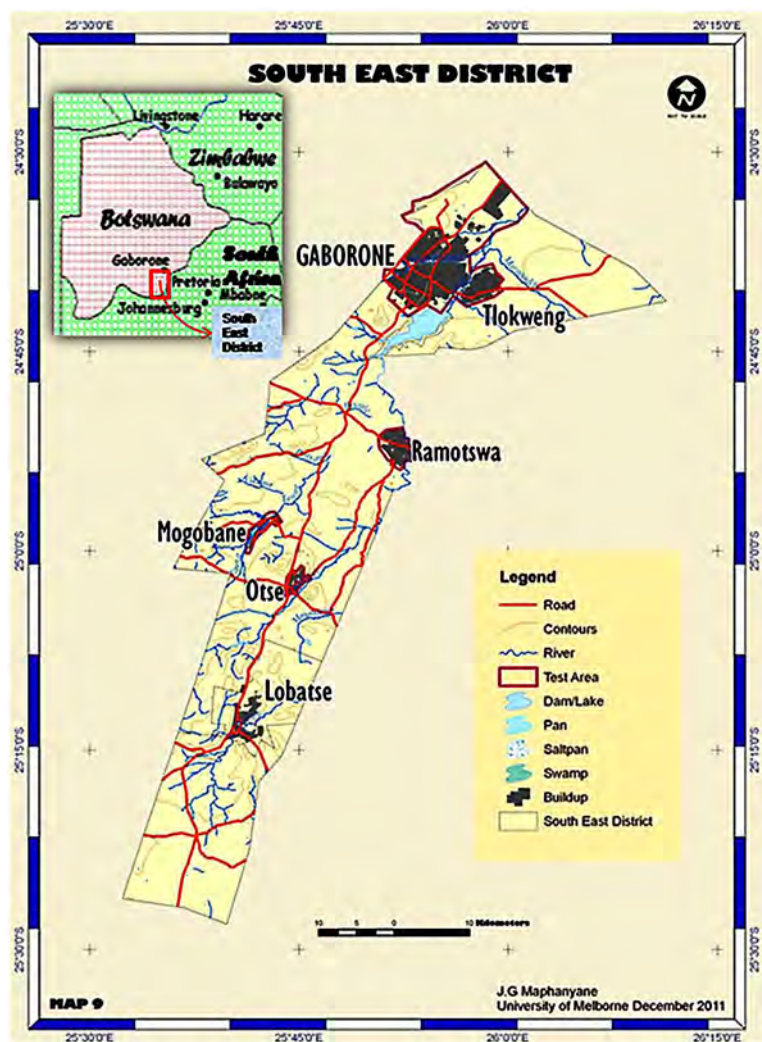
Humankind interest in the distribution of everything that exists on and beyond the Earth itself was triggered by what they could see (Benedict, 2002). The evidence to their cognate abilities is the byzantine names they gave to the phenomena. The Batswana were captivated by all that was around them. Some evidence in that are meaningful names they gave to celestial bodies (Kitauro & Enßlin, 2008) They also employed the skies - *legodimo* as direction tools and identification of time and its subdivisions using daily and seasonal rhythms. They put what seemed to be a scattered litter of sky furniture into distinguishable bodies with meaning. The obvious ones been the *Sun - Letsatsi*, it was used to distinguish between night and day – *Bosigo* le *Motshegare* and the stars – *Dinaledi*. *Mars* was also identified and named *Mphatalatsane*, the red

Morning Star and the *Pointer*, the All-Night-Long Star – *Kgogamasigo* was observed as well. *Venus* – *Kopadilalelo*, the evening star - has long being observed by this people as the second brightest body after the Moon - *Kgwedi* - in the evening skies, but different from the stars as it was always in the same planetary position. The moon is the celestial body most used by the Batswana. The Motswana female-body functions, the menstruation cycle, conception, child-birth and well-being are believed to be dependent on the cadence of the moon. Also, theories about the eclipse of the Sun and the Moon, and the Comets were also devised. For example, the visit of a Comet was a bad omen,

and it might signify something bad, or it might only foreshadow the death of some great chief.

Also, the humankind cognitive abilities are revealed by the work of the Egyptians who had created the precise terrestrial map of the three stars of Orion's belt in the ancient times as early as 10,500 BC; the era of Leo as depicted by the Sphinx of Giza. In this symbolic map, the Milky way is replicated onto the ground by the position of the River Nile and the positions of the three pyramids of Giza have been plotted against the three belt stars of the Orion constellation as such the great pyramid of Khufu occupying the position of Al Nitak; the second pyramid of Khafre tak-

Figure 1. Botswana: South East District



ing the place of Al Nilam and the third pyramid, Menkaure, representing the smallest star which is at an offset position (Hancock, 1995). Modern horoscope astrology is associated with the interaction of Indian and Hellenistic cultures in the Indo-Greek period (Pingree, 1981; Dhavale, 1984) and their eras had been named as stars thus:

- Aquarius,
- Pisces,
- Aries,
- Taurus,
- Gemini,
- Cancer,
- Leo,
- Virgo,
- Libra,
- Scorpio,
- Sagittarius, and
- Capricorn.

Remote Sensing Essentials

These human visual (Sensor) abilities as mentioned above have been amplified by scientific technological developments of remote sensing. Remote sensing is the science and art of obtaining information about an object, area or phenomenon through the analysis of data acquired by a device that is not in contact with the object, area, or phe-

nomenon under investigation (Lillesand & Kiefer, 2004; Jensen, 2005; Schowengerdt, 2007; Mather & Koch, 2011; NASA, 2013). This is done by sensing and recording reflected or emitted energy and processing, analyzing, that into information which is then applied for informed decision making. This is possible because different substances like the Earth's surface phenomena absorbs, transmits and reflects electromagnetic waves energy from the source e.g. the Sun in ways that are diagnostic to the type and condition of the reflecting material. Then this reflected energy could be captured by various satellite sensors. Then, information could be extraction from the resultant images for use. One other thing that is crucial to the fundamentals of remote sensing are the distance (altitude) between the observer and the phenomena been captured and the type of the sensor.

Phenomena under study could be observed from sensors on board platforms anchored at different altitudes, determining the instantaneous field of view and the scale of the data, the spatial resolution. This means that a phenomenon can be viewed from high altitudes in its entirety at a small scale and/or only a part of it can be viewed in detail from lower altitudes at a large scale. It is always beneficial to have a closer view of phenomena; but also seeing them at small scale can be beneficial too. Also, remote sensing allows for revisits of the same place at regular intervals, this count for the temporal resolutions and makes the study of trends possible. Furthermore different types of sensors with various characteristics are possible. That is the sensor used determines the electromagnetic waves window to be captured, its spectral resolutions. And furthermore still, the level of clarity in which the phenomenon is distinguishable from its surrounding, the radiometric resolutions could be measured. That is remote sensing sees what the human eye could perceive through the visible electromagnetic waves; and far beyond that, what the eye could not see; in the gamma ray, x-ray, ultraviolet light, infrared light, microwaves and radio waves. Remote sensing gives an option of studying the scientific nature

Table 1. Botswana: south east district data, comparison of the two methods for the investigation of land cover change.

Analysis of the Results on How Reconstruction of Historical Landscape Compares to Remote Sensing Method (Maphanyane 2012a)				
No.	Test Area	1973	1990	2002
1	Gaborone/Tlokweng	66.94%	66.13%	77.42%
2	Ramotswa	80.00%	71.79%	71.88%
3	Otse	65.27%	65.87%	67.07%
4	Mogobane	77.88%	70.72%	67.31%

Source: (Maphanyane, 2012a)

of phenomena. Even those phenomena that are hidden from normal human view like heat (thermal) and embedded moisture content in plants' leaves could be studied from this type of data as various wavelengths beyond the visible and these could later be viewed through false colour image creation by sophisticated computer software program. In the past fifty years great advances in remote sensing had been achieved by placing the sky-telescopic view, in space above the Earth's atmosphere; the Astronomical Remote Sensing (Kitauro & Enßlin, 2008). At these very high altitudes the focus is on capture gamma, x-ray, ultra-violet, visible, infra-red, microwaves and radio-waves spectrum emitted by the stars, galaxies, interstellar or intergalactic gases, and the void called "space" for their study. Space flight and exploration have led to new technologies that extend the capabilities of astronomers. For instance it is now possible to recreate and study the beginning of the Universe; the BIG BANG! And that the universe is expanding and is finite in both time and space (Jones & Lambourne, 2004). The study of detailed minute fabric of the matter – microscopic view, the Atomic Remote Sensing, which operates at particle or microscopic level, had always being possible in the field of chemistry (specific example (Linga, Kumar, & Englezos, 2007)), physics, biology, and geology and health sciences. Such devices include the microscopes, probes, electro and magnetic survey instruments, spectroscopes, scanners and x-rays. For instance, the absorption spectroscopy that uses infra-red and ultra violet and the UV light sterilisers are examples such devices (Nichol, Shaker, & Nichol, 2006). This article concentrates on the study of the Earth surface and its systems, that incorporates electromagnetic waves capture at altitudes of the ground, atmospheric and space view levels, the Earth Resources Remote Sensing. These what makes it possible for the environment to be studied in quick, frequent and more accurate ways. But, still remote sensing is not the panacea to everything; there are some problems. These normally call for human information reference services input. The four major problems related to

the Earth surface and its systems remote sensing, more specifically the space view data capture and analysis issues are that:

- Although, the remote sensing data capture goes beyond the human visual capabilities, it is machine based, it lacks human mind of reason, hence it needs:
 - To be calibrated and interpreted against the real-world phenomena for it to be meaningful and useful.
 - Trained people to interpret and analyze it.
- Besides all that, it is limited by its time of first data capture. Although it began in 1939, in the form of photogrammetry, the world view dedicated coverage by satellite imagery has only been available since 1972. Consequently, for research that spans periods earlier than this period, remote sensing data is non-existent. Then, since land cover change is a slow process, it is necessary to access evidence about the land cover conditions of past eras, world coverage data from 40, 50, 60, 100 years ago and beyond!
- Furthermore, its data is not automatically available everywhere. Besides, adverse weather conditions renders satellite images useless for the remote sensing analysis of land cover.

So, although remote sensing is useful; it still requires the human side of information reference services for it to be useable and also to fill up the gap were it is inadequate or unavailable. The method for the investigation of land cover change and trends by the reconstruction of historical landscapes could be used to alleviate the problems (Maphanyane, 2012).

Human Based Reference Services in Brief

The reconstruction of historical landscapes could be used as an alternative for the investigation of

land cover change. First, this method uses the elements of traditional ecological knowledge (TEK) of oral history, songs, poems, praise singing and relics. Second, it employs written records in the form of explorers' diaries (Livingstone, 1857; Rey, 1988), the government reports and development plans, the land laws and acts, verdict on major court cases on land disputes and land commissions concerning land uses. Third, it also utilizes the archival records which consist of land surveyors' field sketches, maps, photographs and aerial photographs (Shoemaker & Shoemaker, 1987). And fourth, yet again it uses the scientific endeavours like, archaeological research findings, fossil, sediments, dating such as ^{14}C carbon, ^{210}Pb lead, regional industrial pollutants and regionally significant biostratigraphic markers and the beta - globins, in particular, Mitochondrial deoxyribonucleic acid as research tools (Stringer, 2001).

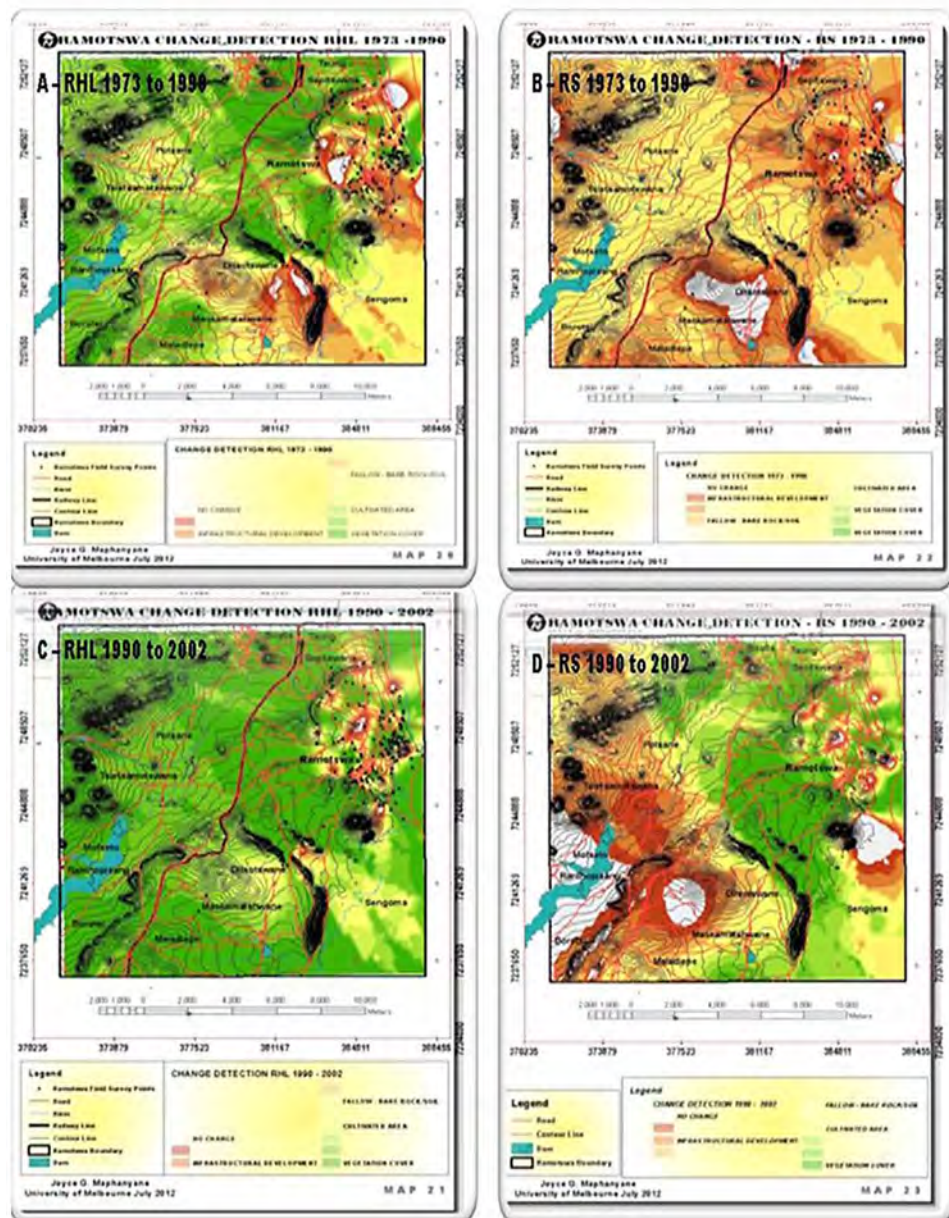
These elements of reconstruction of historical landscape taken ensemble, enables studies that reach ages that goes far beyond dating epochs known today. Since this approach does not rely on technology in its original data capture phase, it can be applied anywhere and long time after land cover changes has taken place (Maphanyane, 2012).

The concept, the reconstruction of historical landscapes, in the context of this communiqué is defined as a method of building and mapping past landscapes by finding out how communities lived in a given historic period in a particular geographical area. This is achieved by identifying, reading and analyzing marks made at the time and left behind from certain period. This brings in a human element in the study. People's life and experiences, the knowledge that an individual has about his/her sphere of influence. It is acquired and registered in their mind through their experience of the area and it is continuously updated when new developments come up (Maphanyane, 2012). Maphanyane (2012) has further reported that each of these eight elements could be useful to the investigation of land cover change to a varying degree: That archaeological findings

and pre-historic sites for instance, gave useful background knowledge of the areas studied, and at such promoted a better understanding of the areas, and later helped in the analysis of the data (Ortner, Butler, Cafarella, & Milligan, 2001). That, Written records which included informal data like travellers diaries; are unique as the authors wrote these, true to themselves as they never knew that one day their work would be used for such a purpose. And other information came from formal data like government development plans. Such documents formed sources of continuous information, which are essential for a land cover change investigation. That the elements comprising of songs, lyrics and poems (Lightfoot, 1995) were difficult to get, hence sporadically available; but, once stumbled upon could prove to be useful where upon an entire history of a group of people could be embedded in one poem or the geography of the area could be sung in a lyric which would be remembered long after the particular landscape of an area had long been eroded. The elements of place names are essential, as they gave insights into past landscapes and the history of the earlier inhabitants of an area (Atteh, 1989; Mabona, 2004). And that the element of traditional ecological knowledge sourced from oral history (with examples coming from various literature - (Brokesha, Warren, & Werner, 1980; Davis, Back, & MacLean, 1992; Leffler & Brent, 1992; Delgamuukw v. British Columbia, 3 S.C.R. 1010, 1997; Russell, 1998); the cultural norms (Berkes, Colding, & Folke, 2000; Davidson-Hunt & Berkes, 2003; Donovan & Puri, 2004); questionnaires, interviews and field (Livingstone, 1857; Rey, 1988) gave the most complete evidence data for the land cover change investigation.

The human based reference services data sources approach deepened on the understanding of current environmental issues, including global warming, desertification, loss of biodiversity, sustainable resource use and sustainability of the biosphere (Russell, 1998).

Figure 2. Graphical comparison of Ramotswa, Botswana land cover change detection by reconstruction of historical landscape (ROHL – A, C) to that by remote sensing method (RS – B, D) for two periods - 1973 to 1990 and 1990 to 2002 based on ArcGIS Software Geo-spatial Statistics Analyst Module (Maphanyane, 2012)



MAIN FOCUS

The focus is on two distinction methods for land cover change investigations. First, the importance and developments of remote sensing as the geo-spatial data source are discussed. To

this end, analysis and comparison (Figure 3) of three satellites systems whose data had reached the most widespread global usage and that had stood the test of time that spans four decades are made. These are the Geo-synchronized Operational Environmental Satellite (GOES), for

international environmental disaster monitoring and weather focusing; the Landsat Systems, which is applied for land cover mapping in green house gas (GHG) carbon loads monitoring projects for the world climate change monitoring and mitigation endeavours; and the SPOT (the Satellite Pour l'Observation de la Terre), that provided the yearly images for the entire planet in Google Earth, "One World, One Year" project applications. The four trans-generational levels developments in spatial, temporal, spectral and radiometric resolutions of these systems have been explored. Second, yet again, comparison is made between these technical sources and the human service sources (Figure 2). That is another focus that brought in the highlights from the human side of information reference services as an essential complementary method that added meaning to remote sensing and that functioned as an alternative to fill-in the gap when these digital data are inadquent or unavailable.

The Advancement in GOES System

The basic element of GOES satellite is that they are geosynchronous and geostationary. They have the ability of remaining permanently in the same area of the sky, as viewed from a particular location on Earth, and so remaining permanently within view of a given ground station. Also, they have special ability that their paths follow a circular geosynchronous orbit directly above the Earth's equator. GOES came into operation in May 1974 then the first satellite SMS 1 was launched. Its program capture data in two levels; from instruments carried by platforms orbiting the Earth at altitudes of about 35 790 km and also from ground stationed data acquisition elements such as river and rain gauges, seismometers, tide gauges, buoys, ships and automatic weather stations.

Its constellation forms part of the global network of meteorological satellites (Leese, Noar, & Pastre, 1989), spaced at approximately 70° longitude intervals around the Earth in order to provide near-global data coverage on weather forecasting, severe storm tracking, and meteorological

research. The launches of these satellites are staggered to keep at least two operational satellites on orbit at all times. To date seventeen spacecraft has been put into orbit, the latest of these being GOES 15, launched in March 2010 and operates as GOES West located at 135°W over the Pacific Ocean. Other two satellites still in operational, are GOES 12 providing Latitude coverage for South America and GOES 13 operates as GOES east, and is currently located at 75°W. It provides most of the U.S. weather information. GOES R, S, T and U are scheduled for launch in 2016, 2017, 2019 and 2024 respectively. The proposed instrument package for the GOES R-series initially included: the Advanced Baseline Imager (ABI); the Hyperspectral Environmental Suite (HES); the Space Environment In-Situ Suite (SEISS), which includes two Magnetospheric Particle Sensors (MPS-HI and MPS-LO), an Energetic Heavy Ion Sensor (EHIS), and a Solar and Galactic Proton Sensor (SGPS); the Solar Imaging Suite (SIS), which includes the Solar Ultraviolet Imager (SUVI), the Solar X-Ray Sensor (XRS), and the Extreme Ultraviolet Sensor (EUVS); the Geostationary Lightning Mapper (GLM); and the Magnetometer (Hill, 2008).

GOES satellites are given an alphabetical number before launch and only designated a numerical number after launch. One launch failed, GOES G, and GOES 14 is in on-orbit storage, whereas GOES 3 and GOES 7 had been re-purposed to communications satellites duties, the former as a relay for Amundsen-Scott South Pole research station and the later by Peacesat. And among the seventeen, ten satellites have been decommissioned and boosted to graveyard orbit.

GOES is SMS 1 and 2 and GOES 1, 2 and 3 first generation make-ups, were sustained from 1974 to 1986 are characterized by latest models are the imager, a multichannel instrument that senses infrared radiant energy and visible reflected solar energy from the Earth's surface and atmosphere and the sounder that provides data for vertical atmospheric temperature and moisture profiles, surface and cloud top temperature, and ozone

distribution (Gunter, 1996). They carried the visible infrared spin-scan radiometer (VISSR), the solar x-ray imager (SXI), and space environment monitoring (SEM) instruments as well. Additionally, GOES carried meteorological data collection and transmission system that relayed processed data from central weather facilities to automatic picture transmission (APT) equipped regional stations and collected and retransmitted data from remotely located ground data acquisition stations (LaViolette & Diachok, 1974; Gunter, 1996). Five satellites, GOES 4, 5, 6, G and 7 with VISSR formed the second generation series that run from 1980 to 1995. They incorporated visible and infrared atmospheric sounder (VAS) another improvement from the earlier series that obtained atmospheric circulation patterns from frame-to-frame movement onto selected clouds at different altitudes and that captured the wind direction and its speed. The VAS added a vital third dimension to the imager; it improved the accuracy for weather prediction data (Gunter, 1996). From this system, GOES 7 had a signal transponding ability and it could locate ships and planes in distress (Doody & Stephan, 1993). The third generation includes GOES 8, 9, 10, 11 and 12 run from 1994 and is still in operational as GOES 12; the last of these series launched in 2001 is still in service; the rest have been decommissioned. These spacecraft payloads are new-generation Imager and a Sounder. The improvements on the SEM include the magnetometer, energetic-particle sensor, high-energy proton and alpha-particle detector, and a solar X-ray sensor and are used for in-situ surveying of the near-earth space environment. They provided 24-hour monitoring and measurement of dynamic weather events in real-time, and are the first to deliver simultaneous independent imaging and sounding from geostationary orbit (Gunter, 1996; SSL, 2013). The new body-stabilized spacecraft design enabled the primary sensors to “stare” at Earth and thus frequently image clouds, monitor Earth’s surface. Atmospheric phenomena were tracked, ensuring real-time coverage of short-lived dynamic events, such as severe local storms

and tropical hurricanes and cyclones. Temporal resolution of 15 minutes and the flexibility; for alteration of normal schedules that provided necessary coverage during significant weather or other events were made. The sensors have an imager with five channels/bands, so specific weather trouble spots were monitored to assist in improved short-term forecasting. The primary sensor is the *Advanced Very High Resolution Radiometer* (AVHRR). The fourth generation included GOES 13, 14, 15 launched between 2006 and 2010. This series, like GOES 12, have a *sun-pointed Extreme Ultraviolet Sensor* (EUVS). In addition to that their spacecrafts payload include the *Advanced Baseline Imager* (ABI); the *Space Environment In-Situ Suite* (SEISS), which includes two *Magnetospheric Particle Sensors* (MPS-HI and MPS-LO), an *Energetic Heavy Ion Sensor* (EHIS), and a *Solar and Galactic Proton Sensor* (SGPS); the *Solar Imaging Suite* (SIS), which includes the *Solar Ultraviolet Imager* (SUVI), the *Solar X-Ray Sensor* (XRS), and the *Geostationary Lightning Mapper* (GLM); and the *Magnetometer* (Segundo, 2010) (Figure 4).

AVHRR has lower spatial resolution (ranging from 0.9 km to 4 km) than other typical land observations sensors for example the Landsat and SPOT program, but its data is useful for study of phenomena which cover the entire globe. Advantages: Frequent coverage- very high temporal resolution, wide coverage covers whole hemisphere at one time, and that it is multi-spectral. So it is used extensively for monitoring regional, small-scale phenomena, including mapping of sea surface temperature, and natural vegetation and crop conditions (NASA - NOAA, 1992; Segundo, 2010).

Land Resources Satellites

The weather satellites are not optimized for sensing detailed mapping of the land surface. The data they bring back is just too coarse.

In 1965, director of the U.S. Geological Survey (USGS), William Pecora, proposed the idea

Figure 3. Landsat (LN) and SPOT satellites 1-4 generations panchromatic (GP) and multispectral (GM) resolutions

G

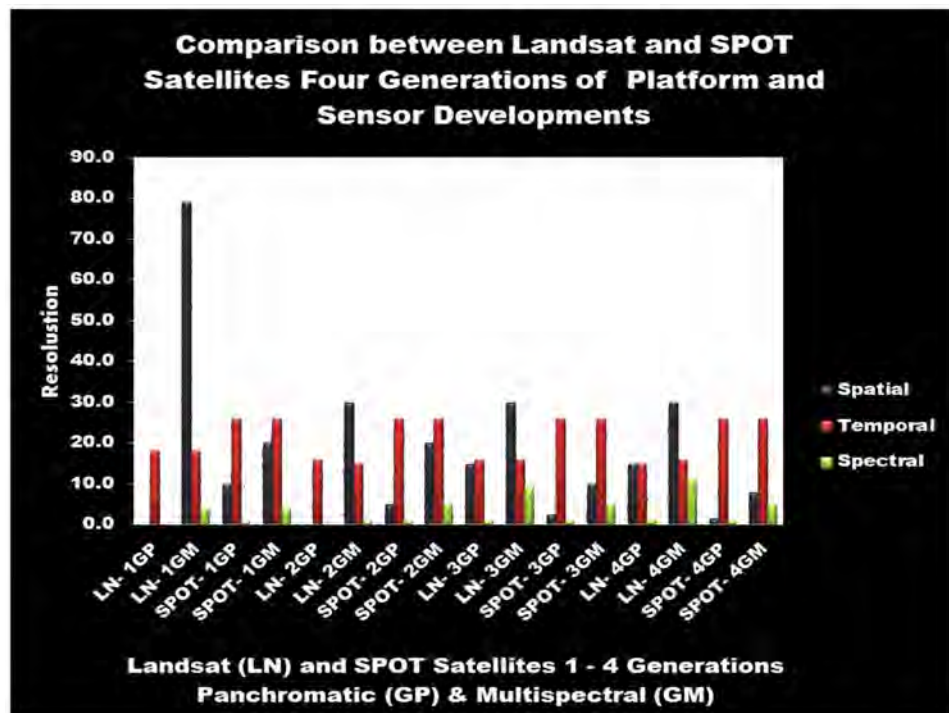
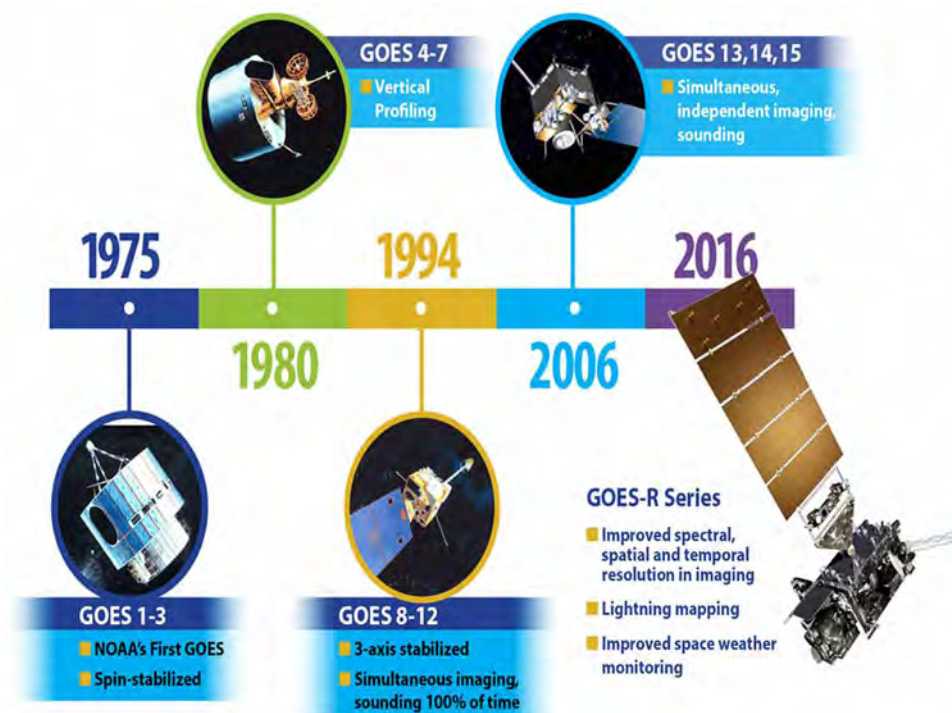


Figure 4. History of GOES Satellite System - GOES 1 - GOES 15, Range October 1975 to March 2010 (Source Graham, 2016).



of a remote sensing satellite program to gather facts about the natural resources of the Earth. This was opposed by the Bureau of Budget for obvious reasons and the Department of Defense who feared that a civilian program would compromise the secrecy of their reconnaissance missions. Also, there were geo-political concerns about photographing foreign countries without permission. Eventually, after these huddles had been overcome, the initial idea culminated to the launching of Earth resources technology satellite (ERTS), which was designed to specifically to monitor the Earth's surface. In 1986, SPOT 1, a French, Landsat-like satellite was launched and it broke the United States of America spacecrafts land resource censorship monopoly.

Landsat Satellite System

ERTS-1, later renamed Landsat 1 was launched by NASA in July 1972. So far eight satellites have been launched in this system, five have been decommissioned, one, Landsat 6 was lost and two are still in service Landsat 7 and 8. The program is jointly managed by the United States Geological Survey (USGS) and NASA. The Landsat archival and latest data have been used in studies for monitoring the dynamic changes caused by both natural and manmade processes for informed decision making and resource management.

The first generation in this series is Landsat 1, 2 and 3 which were in orbit from July 1972 when Landsat 1 was launched to March 1983 when Landsat 3 was terminated. The spacecraft carried the camera system, with the Return Beam Vidicon (RBV), and the Multispectral Scanner (MSS). They orbited in near-polar, sun-synchronous at 900 km altitude (NASA - NOAA, 1992), and 18 days revisits, a much coarser temporal resolution than its predecessor, GOES. It had four multispectral bands; two bands in visible, green and red; and the other two in near infrared. The spatial resolutions were 60mx80m for Landsat 1 and 2; and 38mx38m for Landsat 3. The System's second generation; Landsat 4 and 5 started in July 1982, when Landsat

4 was launched and ended in December 2012 when Landsat 5 was terminated (the longest-operating Earth observation satellite). They have near-polar, sun-synchronous orbits hinged at 705 km altitude with 16 days revisits (NASA - NOAA, 1992). They carried two sensors, the MSS and the Thematic Mapper (TM). The TM sensor captured data in seven spectral bands in visible: blue, green, red' and the infrared: near-infrared, mid-infrared and thermal infrared. Bands 1-5 and 7 had a spatial resolution of 30mx30m; whereas, Band 6, was 120mx120m. Landsat 6 and 7 formed the third generation and it run from April 1999 when Landsat 7 was launched as Landsat 6 had been lost; to date. It had a relapse from the faulty scan line corrector for six weeks in May 2003. Then, the scientific community reverted back to using Landsat 5 imagery. Landsat 7 Like its predecessor was launched in near-polar, sun-synchronous orbits at 705 km altitude with 16 days revisits as well. It has the Enhanced Thematic Mapper Plus (ETM+) that replicated the TM characteristics; with an additional 15m spatial resolution panchromatic band and two thermal: Bands 6-1 and 6-2 with improved spatial resolutions of 60mx60m. Other major development is the on-board, full aperture, 5% absolute radiometric calibration, which rendered it the most accurately calibrated Earth-observing satellite and the on-board data recorder (SSL, 2013). The future of Landsat Satellites Systems lays on the fourth generation series, Landsat 8 launched in February 2013 with orbits and revisits similar to its predecessor (NASA, 2013), (USGS, 2013). It is the Landsat Data Continuity Mission (LDCM) generation pledged to ensure the continued acquisition and availability of Landsat-like data beyond the duration of the current Landsat missions. Landsat 8 major developments are the Operational Land Imager (OLI) with nine spectral bands, one panchromatic at with 15m resolution, four visible bands in green, blue, yellow and red; four infrared bands in one in near infrared, two short wave infrared and another one cirrus infrared; all these have 30m resolution. The OLI has improved signal-to-noise ratio compared

to past Landsat instruments. It also has a special Thermal Infrared Sensor (TIRS) with two bands at 100m resolution that measures land surface temperatures.

Initially, the costs of the images of this satellite series were prohibitive as they kept on rising from US\$ 600, to US\$ 3700 reaching a whopping price of US\$ 4400 per scene. As a result of this cost recovery mandate its use dwindled. The Landsat coverage standards also languished as many observations were missed because there was no buyer. As a result of this setback, the true scientific mission based on collects as much global data as possible for future scientific study motto failed, and as data was lost forever (Landsat5, n.d.). Eventually, in January 2009 all Landsat data were made free and its use immediately increased 60-fold (Satellite Imaging Cooperation, 2001).

SPOT Satellite System

SPOT System was initiated by the French Space Agency, Centre National d'Études Spatiales

(CNES), the Internationale Géographe Nationale (IGN), and Space Manufacturers; this was in collaboration with the Belgian scientific, technical and cultural services (SSTC) and the Swedish National Space Board (SNSB). SPOT is composed of both spacecrafts and a network of direct receiving stations handling images acquired by the satellites. It has been designed to improve the knowledge and management of the Earth by exploring the Earth's resources, detecting and forecasting phenomena involving climatology and oceanography, and monitoring human activities and natural phenomena (Satellite Imaging Cooperation, 2001).

SPOT's onboard sensors can point across the satellite track, providing a revisit capability of one to four days depending on latitude. The satellites are anchored at 832 km latitude and fly over any point on Earth within 26 days. The satellites capture both panchromatic and multispectral digital imagery from visible and infrared electromagnetic wave portion of the spectrum with a spatial resolution of 10m and 20 m respectively.

Figure 5. WorldView 3; Very High Spectral Image - Madrid, Spain (Source: Satellite Imaging Cooperation, 2001).



Figure 6. WorldView 3; Very High Spatial Resolution Image – Cape Town, South Africa (Source: Satellite Imaging Cooperation, 2001).



SPOT system has stereo pair imagery which is vital for applications that call for 3D terrain modeling Astrium. (2013). The first generation of SPOT satellite is SPOT 1, 2 and 3 that started in February 1986 up to November 1997. They had 4 spectral bands in one panchromatic, two visible: green and red and in infra red: near infra red. The second generation is composed of SPOT 4 launched in March 1998 to July 2013. It has five spectral bands, one in panchromatic, three in visible: blue, green and red and one infra red. SPOT 5 is the third generation satellite, launched in May 2002 and had five spectral bands It gave higher resolution of 2.5 to 5 m in panchromatic mode, 10 m in multispectral mode: green and red and 20 m on infrared: near infra red and short wave And the fourth generation of these, SPOT 6 was launched in September 2012 and SPOT 7 which was 30 June 2014. This 2 satellite form a constellation at 90° and at 694 km altitude and has five spectral bands. It has very high resolutions of 1.5 meters in panchromatic mode, 6 m in multispectral: blue, green and red mode and 20 m on short wave infrared: near infra red. This constellation offers a wide imaging swath, which up to 60 km x 120 km and like all SPOT satellites 3D capabilities (Figure 7). It is essentially used for the creation of 3-Dimensional data models – Digital terrain/Elevation Models – (DEM/DTM). The only three SPOT satellites which are still in orbit are SPOT 5, 6 and 7.

Satellites with Special Unique Applications

Other types of satellite systems worth mentioning here are those with unique capabilities which are indeed complementary to aerial spatial data capture efforts, although, in essence, their longevity and widespread use falls far short from the three explained above. The Indian Remote Sensing Satellite (IRS) satellite series the features mimic both the Landsat sensors and those of SPOT sensors; and has two-channel Wide Field Sensor (WiFS) as well with a coarse resolution that is equivalent

Figure 7. SPOT – DEM of Los Pelambres Mine, Chile (Source – Airbus Defense Space, 2013)



to GOES. The new Very High Resolution Remote Sensing Satellites (VHRRS) formats combine the spatial resolution equivalent to aerial photographs with the temporal, spectral and radiometric resolutions, which by far surpasses those possible with those three studied here. Examples are:

- WorldView (3.1m, 29 bands) (Figure 5 – Very High Spectral & Figure 6 - Spatial Resolution),
- GeoEye (0.41m, 4 Bands),
- QuickBird (0.61m, 4 bands),
- IKONOS (0.82m, 5 bands), and
- ASTER (15m, 14 bands).

The Radio Detection and Ranging (RADAR) satellites, known as the all weather, are unique, as they use the microwave energy which relatively has long wavelengths that allows these systems to “see” through clouds, smoke, and some vegetation. They generate their own energy, Active Sensors; so they can be operated day or night. All other satellite sensors, discussed above utilize the energy from the Sun; and are Passive Sensors.

Data from the Two Sources Interchangeability Problems

The main fundamental reasons for the differences between technical and human services sources are in the fabric of the origins of the data, when it happened, when and how it was preserved before it was eventually collected and used for the investigation of land cover change. The resolution of the data is paramount to its usefulness in both methods. It depends on the spatial, the thematic, temporal and the model resolutions. The spatial resolution is the smallest discernible unit in the geographical space whereas the thematic resolution addresses the level of classification aggregation used to show distributions of patterns in an area; the temporal resolution is how often data updates are made; and the model resolution gives the level of details incorporated in assessing any form of phenomena in an area.

- The basic tenet is that remote sensing and GIS and the method for the reconstruction of historical knowledge measured totally different things. For these two methods to be used interchangeably for the investigation of land cover change for a particular geographical area there is a need to reconcile what each of them measures.
- The challenge is to make them measure the same thing. So, to that end; a land cover class specification commensurate to the two methods was needed to be created.
- Besides all that, care must be taken when remote sensing cross generational data is used in the land cover change and trends investigations. This is because the remote sensing platforms and sensors have improved tremendously over the years. Consequently, land cover data from consecutive time period captured by satellite from different generation could show differences pertaining to the differences in visual abilities, due to the improved tools with better clarity that might be mistakenly taken to be the change the landscape.

- So, for that reason, for efficient land cover change investigation employment of trans-generation satellite data has to be taken with caution and some form of standardization and a form of calibration of data from the different generation satellite sources is necessary. The class separability was made within the particular Landsat sensor and across different sensors.

For instance classes which had very little separability in MSS sensor were found to have huge separability in TM and ETM+. It was difficult to attribute these differences to land cover change alone as there is technology development for the better in satellite sensor sensitivity at play as well.

SOLUTIONS AND RECOMMENDATIONS

The land cover is dynamic but at the same time the technology for measuring it keeps on changing as well, there are many improvements. As a consequence, there is no uniform continuous data from which class separability alone could be determined throughout the years.

For these reasons, the percentage of how much class separability is across different land cover classes by different sensors could not be uniformly assessed because of the improved progressive difference in the quality of data itself. Maybe apparent reduction of data quality can be made by applying a statistical ratio to all the data to make it uniform so that its separability throughout the satellite capture years could be determined.

FUTURE RESEARCH DIRECTIONS

The basic tenet for future research would have to be based on the main finding of this research which is that remote sensing and GIS and the proposed new method for the reconstruction of historical knowledge measured totally different things. For these two methods to be used interchangeably

for the investigation of land cover change for a particular geographical area there is a need to reconcile what each of them measures. The challenge is to make them measure the same thing, a good research topic for the future.

CONCLUSION

GOES great improvements from the earliest satellites to the present ones with more bands, so specific to weather trouble spots were monitored to assist in improved short-term forecasting. These had seen improvements on the satellite systems hardware themselves with better flexibility and more automatic ground stationed data acquisition elements such as river and rain gauges, seismometers, tide gauges, buoys, and ships weather stations.

Ever since the first satellites were launched in 1972 for Landsat System and 1986 for SPOT Satellites there has been tremendous amount of developments in platform and sensor for both system. These developments are directed towards a certain goal for each system. Landsat developments are more onto the refinement of spectral resolution variable through the improvements of the two sensors from mere 4 bands of the earliest systems to 7, 9 and 12 for the 2nd, 3rd and 4th generation. Landsat system also developments were in the improvements in the infra red more especially in thermal bands which had improved in both its spatial and spectral resolution. Also, generally these satellites have seen a slight improvement in spatial resolution and with the addition of panchromatic whereas its temporal resolutions have been maintained throughout (Figure 3). SPOT boost the very important characteristics of seeing stereo, therefore having 3-dimensions landscape creation. This also gave it increased revisits of one to four days. Although its spectral resolution are much lower 5 as compared to 12 of those of Landsat 5; it has had superior improvements in spatial resolution going from 10m, 5m, 2.5m and 1.5m for panchromatic and 20m, 20m, 10m and 8m for Multispectral from 1st, 2nd, 3rd and 4th

generation satellites respectively. Generally, SPOT maintained its temporal and spectral resolutions with slight improvements on its infra red spectra but huge strides were made in the spatial resolutions improvements (Figure 3).

The results of these show that the developments of GOES, Landsat and SPOT are not in competition of one another, but are complementary as each system has maintained its niche in geo-spatial space-view (Figure 3). Other systems like the IRS, VHRRS and RADAR satellites are becoming available and these are added to increased use of digital data.

Remote sensing digital data have shown tremendous development, but on the other hand, need for the human side of information reference services for the interpretation of these more sophisticated digital data increased. It has also been shown that the accuracy of these reconstruction to historical landscape are commensurate with that of remote sensing although work need to be done to harmonize these two inherently diverse data (Figure 2 and Table 1).

REFERENCES

- AIRBUS-Group. (2013). *Airbus Group Annual Review*. Toulouse: Astrium.
- Atteh, O. D. (1989). *Indigenous local knowledge as key to local-level development: Possibilities, constraints and planning issues in the context of Africa*. Arusha, Tanzania: Conference Paper.
- Benedict, B. M. (2002). *Curiosity: A cultural history of early modern inquiry*. Chicago: University of Chicago Press.
- Berkes, F., Colding, J., & Folke, C. (2000). Rediscovery of traditional ecological knowledge as adaptive management. *Rediscovery of the Ecological Applications*, 10, 251–262.
- Bossler, J. D., Jensen, J. R., McMaster, R. B., & Rigos, C. (2002). *Manual of geospatial science and technology*. New York: Taylor and Francis.

- Brokesha, D., Warren, D., & Werner, O. (1980). *Indigenous Knowledge Systems and Development*. Lanham, MD: University Press of America.
- Davidson-Hunt, I., & Berkes, F. (2003). *Learning as You Journey: Anishinaabe Perception of Social-ecological Environments and Adaptive Learning*. Retrieved December 2015, from <http://www.consecol.org/vol8/iss1/art5/>
- Davis, C., Back, K., & MacLean, K. (1992). Understanding Oral History. In K. Phyllis & J. Brent (Eds.), *Public History Readings* (pp. 264–269). Malabar, FL: Krieger.
- Delgamuukw v. British Columbia, 3 S.C.R. 1010, 23799 (Supreme Court Judgments December 11, 1997).
- Dhavalé, D. G. (1984). The Yavanajataka of Sphujidhvaja by David Pingree. *The YavAnnals of the Bhandarkar Oriental Research Institute*, 65(1/4), 266–267.
- Donovan, D., & Puri, R. (2004). *Ecology and Society*. Retrieved December 26, 2015, from Vol 9 No. 3 Art 3: <http://www.ecologyandsociety.org/vol9/iss3/art3/>
- Doody, D., & Stephan, G. (1993). *NASA - Jet Propulsion Laboratory*. Retrieved December 26, 2015, from Basics of Space Flight: <https://solar-system.nasa.gov/basics/editorial.php>
- Graham, W. (2016, November 19). *Atlas V successfully launches with GOES-R advanced weather satellite*. Retrieved from <https://www.nasaspaceflight.com/2016/11/atlas-v-launch-goes-r-advanced-weather-satellite/>
- Gunter, D. K. (1996). *Gunter's Space Page*. Retrieved August 14, 2013, from GOES: http://space.skyrocket.de/doc_sdat/goes-d.htm
- Hancock, G. (1995). *The fingerprints of god: A quest for the beginning and end*. London: William Heinemann Ltd.
- Hill, S. (2008). *GOES-R Solar and Space Environment Data Products: Benefiting Users*. Retrieved December 26, 2015, from GOES-R Solar and Space Environment Data Product Geostationary Operational Environmental Satellite Program (GOES): www.weather.gov/satellite
- Jensen, J. R. (2005). *Introduction to digital image processing: A remote sensing perspective*. Upper Saddle River, NJ: Prentice Hall.
- Jones, M. H., & Lambourne, R. J. (2004). *An Introduction to Galaxies and Cosmology*. Cambridge, UK: Cambridge University Press.
- Kitaura, F. S., & Enßlin, T. A. (2008). Bayesian reconstruction of the cosmological large-scale structure: Methodology, inverse algorithms and numerical optimization. *Monthly Notices of the Royal Astronomical Society*, 389(2), 497–544. doi:10.1111/j.1365-2966.2008.13341.x
- Landsat5. (n.d.). Retrieved from <http://landsat.gsfc.nasa.gov/landsat-5/>
- LaViolette, P. E., & Diachok, O. I. (1974). The Use of APT satellite imagery in a subarctic airborne oceanographic survey. *Arctic*, 27(4), 249–324. doi:10.14430/arctic2888
- Leese, J. A., Noar, P. F., & Pastre, C. (1989). Operational continuity of the global meteorological satellite network. *Space Policy*, 5(1), 12–24. doi:10.1016/0265-9646(89)90025-8
- Leffler, P. K., & Brent, J. (1992). *Public history readings*. Malabar, FL: Krieger Publishing Company.
- Lightfoot, K. G. (1995). Culture Contact Studies: Redefining the relationship between prehistoric and historical archaeology. *American Antiquity*, 60(2), 199–217. doi:10.2307/282137
- Lillesand, T. M., & Kiefer, R. W. (2004). *Remote sensing and image interpretation*. Chichester, UK: John Wiley & Sons.

- Linga, P., Kumar, R., & Englezos, P. (2007). Gas hydrate formation from hydrogen/carbon dioxide and nitrogen/carbon dioxide gas mixtures. *Chemical Engineering Science*, 62(16), 4268–4276. doi:10.1016/j.ces.2007.04.033
- Livingstone, D. (1857). *The missionary travels and researches in south africa*. London: Royal Geographical Society.
- Mabona, M. (2004). *Diviners and prophets among the Xhosa (1593-1856): A study in Xhosa cultural history*. Münster: Lit Verlag.
- Maphanyane, J. G. (2012, December 1). *The investigation of land cover change and trends through the reconstruction of historical landscape* (PhD Thesis). Melbourne, Victoria, Australia: The University of Melbourne.
- Mather, P., & Koch, M. (2011). Computer processing of remotely-sensed images: An introduction. Wiley-Blackwell. doi:10.1002/9780470666517
- NASA. (2013, July 28). *Landsat*. Retrieved August 22, 2013, from Landsat 8: http://www.nasa.gov/mission_pages/landsat/spacecraft/index.html
- NASA - NOAA. (1992, August 28). *GOES - R - SERIE*. Retrieved September 06, 2013, from GOES-R Spacecraft Overview: <http://www.goes-r.gov/spacesegment/overview.html>
- Nichol, J. E., Shaker, A., & Nichol, J. E. (2006). Application of high-resolution stereo satellite images to detailed landslide hazard assessment. *Geomorphology*, 76(1-2), 68–75. doi:10.1016/j.geomorph.2005.10.001
- Opara, L. U. (2003). Traceability in agriculture and food supply chain: a review of basic concepts, technological implications, and future prospects. *Agriculture and Environment*, 1(1), 101–106.
- Ortner, D. J., Butler, W., Cafarella, J., & Miligan, L. (2001). Evidence of probable scurvy in subadults from archeological sites in North America. *American Journal of Physical Anthropology*, 114(4), 343–351. doi:10.1002/ajpa.1046 PMID:11275963
- Rey, C. (1988). *Monarch of all I survey: Bechuanaland diaries 1929 - 1937*. Gaborone: Botswana Society: Lilian Barber Press Inc.
- Russell, E. (1998). *People and the land through time: Linking ecology and history*. New Haven, CT: Yale University Press.
- Satellite Imaging Cooperation. (2001). *Satellites*. Retrieved August 22, 2013, from Satallite Data: <http://www.satimagingcorp.com/satellite-sensors/spot-5.html>
- Schowengerdt, R. A. (2007). *Remote sensing: models and methods for image processing*. Burlington, MA: Academic Press.
- Segundo, E. (2010, September 01). *Boeing*. Retrieved December 26, 2015, from GOES-15 Weather Satellite: <http://boeing.mediaroom.com/2010-09-01-Boeing-built-GOES-15-Weather-Satellite-Enters-Service-for-NASA-NOAA>
- Shoemaker, H. G., & Shoemaker, E. M. (1987). *In the Footsteps of John Wesley Powell*. Boulder, CO: Johnson Books and Powell Society.
- SSL. (2013). *SPACE System Lorel*. Retrieved August 22, 2013, from GOES I: <http://www.ssloral.com/html/satexp/goes.html>
- Stringer, C. (2001). Modern Human Origins: Progress and Prospects. *The Royal Society*, 357, 563–579. PMID:12028792
- USGS. (2013, August 29). *Landsat Mission*. Retrieved September 22, 2013, from Landsat 8: http://landsat.usgs.gov/about_ldcm.php

KEY TERMS AND DEFINITIONS

Archaeology: The importance of archaeology lies in the fact that it seeks to learn about culture from the fragmentary remains of the products of human activity (Deetz, 1992).

Biostratigraphic Markers: Biostratigraphy is a sub-discipline of sedimentary geology that relies on the physical zonation of biota, both in

time and space, in order to establish the relative stratigraphic position (i.e. older, younger, same age) of sedimentary rocks between different geographic localities.

Classification: The process of classification identified and assigned each pixel of all channels of the multi-spectral images to a particular class or theme based on the statistical characteristics of the pixel brightness values known as spectral signatures.

Electromagnetic Waves Energy: The electromagnetic (EM) spectrum describes the range of wavelengths of energy that can be recorded using remote sensing and it can be broadly divided (by increasing wavelength) into Gamma rays, X-rays, Ultra-violet light, Visible light, Infrared light, Microwave and Radio waves.

Geographical Information Systems (GIS): Geographical Information Systems (GIS) is defined as a technique which people employ using computers and specific computer software and hardware to locate physical features and describe their characteristics and condition on a raster based or vector co-ordinate based digital map by interrogating attribute data and to engage in spatial analysis, and this is a sound basis for informed decision making (Burrough & McDonnell, 1998).

Geo Spatial Science (GSS): Geospatial Science is a discipline that focuses on using information technology to understand people, places, and processes of the earth. Spatial analysis of human and physically variables is fundamental to the discipline. Remote Sensing, Geographic Information Systems, and Global Positioning Systems technologies are commonly used as measurement, observation, and analysis tools for this (Radford University, 2014).

Geo-Stationary: Geostationary satellite Platforms are platforms that revolve at speeds which match the rotation of the Earth so they seem stationary, relative to the Earth's surface. This allows the satellites to observe and collect information continuously over specific areas.

Geo-Synchronized and Sun-Synchronous: The platforms are designed to follow a north-south

orbit and in conjunction with the Earth's west-east rotation, they are able to capture the entire Earth's surface over a certain period of time. These satellite platforms orbits are often sun-synchronous: - They are able to cover each particular area on the Earth's surface at a constant local time of day. This ensures consistent illumination conditions for images of the same area that have been acquired at different times.

Land Cover Change: Land cover change forms a fundamental part of sustainable resource management. It is actually the main indicator of natural resource use.

Multi-Spectral: A multispectral image is one that captures image data at specific frequencies across the electromagnetic spectrum, the visible light range, and infrared. Spectral imaging sees beyond what human eye can perceive.

Panchromatic: Black and white aerial photograph or satellite images.

Platform: In order for the sensor to capture and record reflected or emitted energy from a target it must reside on a stable platform which can be on the ground (hand held camera), within the atmosphere (airplanes), and in space (the space shuttle or satellite).

Radiometric Resolution: The radiometric resolution of an imaging system describes its ability to discriminate very slight differences in energy.

Reconstruction of Historical Landscape (ROHL): *Reconstruction of historical landscapes* is defined as a method of building and mapping past landscapes by finding out how communities lived in a given historic period in a particular geographical area. This is achieved by identifying, reading and analysing marks made at the time and left behind from certain period. A historical landscape basically depicts a type of cultural landscape that contains, within a specific geographical area, both natural and human-made features that typify connected human activities, past events or patterns of physical development (Maphanyane, 2012).

Remote Sensing: Remote Sensing is the art, science, and technology of obtaining reliable information about physical objects and the environment through the processes of recording, measuring, and interpreting photographic images and patterns of electromagnetic radiant energy and other phenomena (Lillesand & Kiefer, 2004).

Sensor: Sensors are devices that functions like a human eye. They detect the reflected or emitted electromagnetic radiation from natural sources (Passive Remote Sensing) or detect responses from objects which are irradiated from artificially generated energy sources such as radar (Active Sensors).

Spatial Resolution: Spatial Resolution, Pixel Size, and Scale. With remote sensing instruments, the distance between the target being imaged and the platform, plays a large role in determining the detail of information obtained and the total area imaged by the sensor.

Spectral Resolution: Different classes of features and details in an image can often be

distinguished by comparing their responses over distinct electromagnetic wavelength ranges. Spectral resolution describes the ability of a sensor to define fine wavelength intervals.

Temporal Resolution: Temporal resolution: Other than: Spatial, Spectral and radiometric resolution; there is temporal resolution. This is the time it takes to image the exact same area at the same viewing angle a second time: Revisit Time.

Traditional Ecological Knowledge (TEK): The traditional ecological knowledge as a cumulative body of knowledge and beliefs, handed down through generations by cultural transmission, about the relationship of living beings with one another and with their environment. These are gained through the teaching of family history and cultural values. They have been recorded for thousands of years by storytelling, praise-singing, and songs as oral history, and are passed on to and learned by the descendants through the recitation of the narrative at events and during ceremonies (Berkes, Colding, & Folke, 2000).

Use of GIS and Remote Sensing for Landslide Susceptibility Mapping



Arzu Erener

Kocaeli University, Turkey

Gulcan Sarp

Suleyman Demirel University, Turkey

Sebnem Duzgun

Middle East Technical University, Turkey

INTRODUCTION

In recent years, geographical information systems (GISs) and Remote Sensing (RS) have proven to be common tools adopted for various studies in different scientific disciplines. GIS provides as a set of tools for the input, storage, retrieval, manipulation, management, modeling, analysis and output of spatial data. RS, on the other hand, offers earth observation data for thematic maps related to spatial studies. The use of GIS and RS data for landslide susceptibility mapping are demonstrated by three different landslide susceptibility maps with five different variables (Normalized Difference Vegetation Index (NDVI), Topographic Wetness Index (TWI), slope; lineament density and distance to roads). The comparison of the generated final susceptibility maps with historical landslide locations is given with important factors affecting the accuracy of susceptibility map. The accuracy analysis of the final susceptibility maps for various weighting strategies is performed. The results indicate that assignment of weights to the slope parameter impacts the accuracy in the high susceptible zones.

BACKGROUND

Landslides are among the most common natural hazards and are the most damaging, leading to

substantial economic, human, and environmental losses throughout the world. The quantitative assessment of landslide hazards for a large area is critical for mitigation of the associated risks. They are often triggered by natural phenomena and/or human activity, such as earthquakes, precipitation, erosion, deforestation etc. and are difficult to predict. One of the greatest limiting factors in predicting and mapping landslide activity is the lack of understanding of scale-dependent processes, such as erosion, weathering, and fracturing (Glenn et al., 2006). Such maps normally aim at providing a document that depicts the likelihood or possibility of new movements occurring in an area, and therefore helping to reduce future damages. To express the potential for occurrence of landslides in a quantitative manner, maps must incorporate the concept of probability, which is an assessment of the relative frequency of occurrence (Ohlmacher & Davis, 2003). Susceptibility expresses the likelihood that a landslide will occur in an area on the basis of the local terrain conditions (Soeters & Van Westen, 1996); return period or annual probability of occurrence is not considered. The main difference between susceptibility and hazard is therefore that the latter considers the temporal factor, by estimating the probability of occurrence of the phenomenon within a specified period of time (Varnes, 1984) whereas the former considers the likelihood of landslide occurrence. There are many studies in the literature about use

DOI: 10.4018/978-1-5225-2255-3.ch304

of GIS to evaluate landslide susceptibility (e.g. Gokceoglu et al. 2005; Akgun et al. 2011; Akgun 2012; Pradhan et al. 2013; Kavzoglu et al. 2013)

MAIN FOCUS

The purpose of this study is to apply the grid based GIS techniques for landslide susceptibility mapping using five different factors including Normalized Difference Vegetation Index (NDVI), Topographic Wetness Index (TWI), slope, lineament density, and distance to roads. The scope includes the preparation of landslide susceptibility map to identify highly susceptible areas and, the accuracy assessment related to the obtained maps.

The susceptibility assessment methodology is demonstrated for More and Romsdal region in Norway (Erener & Duzgun, 2010). The study area occupies approximately 606.755 km² in the west part of Norway. The upper left coordinates on 112707,770408 m - 6952112,603469 m and lower right coordinates 6929466,479194 m -144909,272731 m respectively (Figure 1).

Method of Study

An empirical approach is used to map and evaluate landslide susceptibility. In this approach a

grid based GIS is adopted to construct a landslide susceptibility map. Five layers of data with 30 × 30 m resolution grid were superimposed to create the landslide susceptibility map. Slope is given the most emphasis, followed by, TWI, NDVI, lineament density and distance to roads. A numerical rating system is applied and each of the five factors is grouped into three categories, and each category is assigned a value between 1 and 3, with 1 being least susceptible and 3 most susceptible to landslides. Based on their relative importance to slope instability in the study area, the five factors are assigned weights between 0.0 and 1.0 (collectively adding to 1.0). A raster-based GIS is used to overlay the five layers with 30 × 30 m resolution and calculate a Landslide Susceptibility Index (LSI) for each individual cell. The final map shows areas of low, medium, and high landslide susceptibility. The method of the study is shown in Figure 2.

Scale and properties of the data used in the study for susceptibility mapping is given in Table 1. Before beginning the processes, the water areas are masked from the layers. Vector based 50 meters contour interval topographic maps, provided from the Geological Survey of Norway (NGU), are the main data used in the study (Figure 3a). DEM data are generated from maps with a cell size of 30 m using triangular irregular network. The

Figure 1. Study region (parts of figure adopted from NGI)

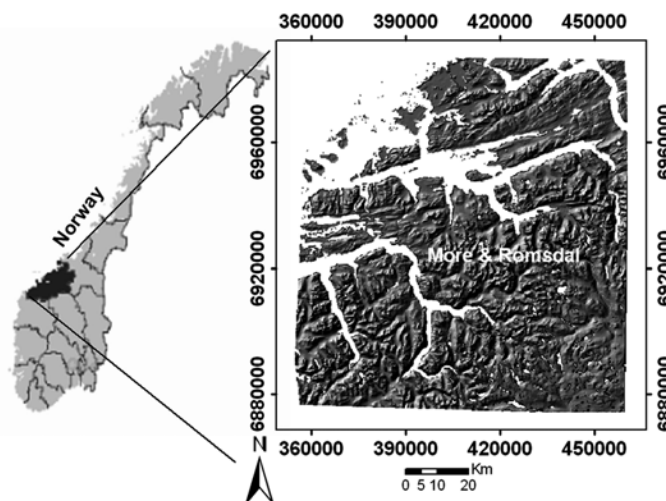


Figure 2. The method of the study

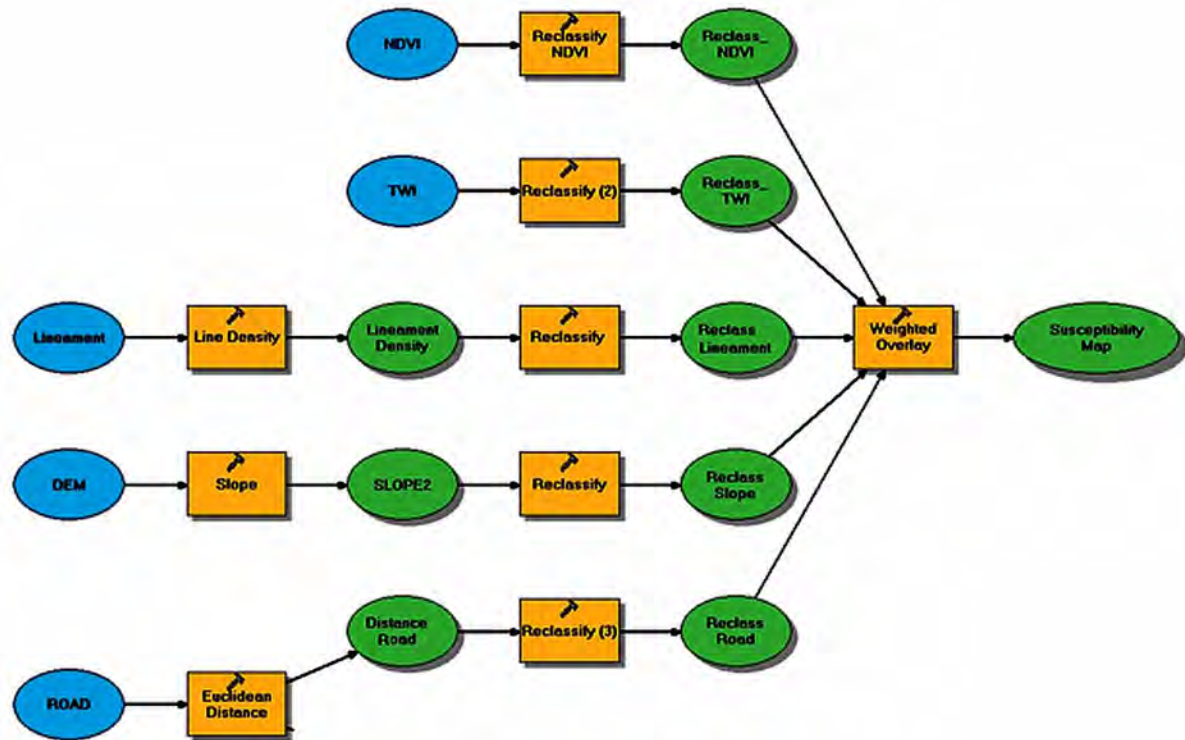


Table 1. Data scale and properties

Classification	Sub-Classification	GIS Data Type	Scale or Grid Size
Geological Hazard Map	Landslide	Point Coverage	1/500 000
Basic Map	Elevation	Line Coverage	1/100000
	Slope	GRID	30*30
Derived Maps	Topo. Wetness Index	GRID	30*30
	Lineament	Line Coverage	1/100 000
	Vegetation Index	GRID	1/100 000

result of the generated DEM includes the highest and the lowest elevation 1807.46 m-13.7282 m, respectively (Figure 3b).

The vertical accuracy of the DEM data are tested with root-mean-square error (RMSE) statistic. For this test some principles suggested by USGS are used. According to USGS National Mapping Program Technical Instructions, a representative sampling of test points is used to verify the accuracy of any category of DEM. A minimum

number of 28 test points per DEM is required (20 interior points and 8 edge points) (USGS).

The vertical RMSE is defined as:

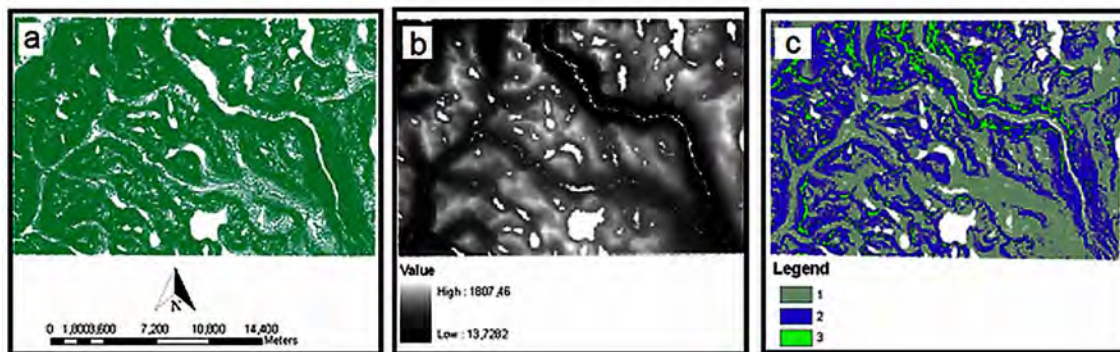
$$RMSE = \sqrt{\frac{\sum (z_i - z_t)^2}{n}}$$

z_i = interpolated DEM elevation of a test point

z_t = True elevation of a test point

n = number of test points.

Figure 3. Contour map (a), DEM data of the area (b) categorized slope map of the area (0-25=1; 25-50=2 and 50-84=3)



Accuracy is computed by a comparison of interpolated elevations in the DEM with corresponding measured elevations from the topographic map. The total vertical RMSE is found 2.30 meter which is quite admissible within the accuracy requirements of the study.

Slope, which measure of the steepness of an area on the Earth's surface, is the most important factor to affect the landslide occurrence (Lee & Min 2001). The steeper the slope, the more likely it is to slide. The slope map is derived from the DEM of the study area. The slope layer has values ranging from 0° to 84.55° . The slope image is then classified into three slope angle classes (0-25=low, 25-50= moderate and 50-84.55=high), and each class is assigned a value between 1 and 3, with 1 being least susceptible and 3 most susceptible to landslides. The result is shown in Figure 3c.

The TWI has been used extensively to describe the effect of topography on the location and size of saturated source areas of runoff generation. TWI is a function of both the slope and the upstream contributing area per unit width orthogonal to the flow direction. The calculation of the wetness index is based on logical ideas of down slope water movement and accumulation of water at the base of slopes and in depressions or swales where there is convergence of flow (Kokkila, 2002). TWI is calculated using the following expression:

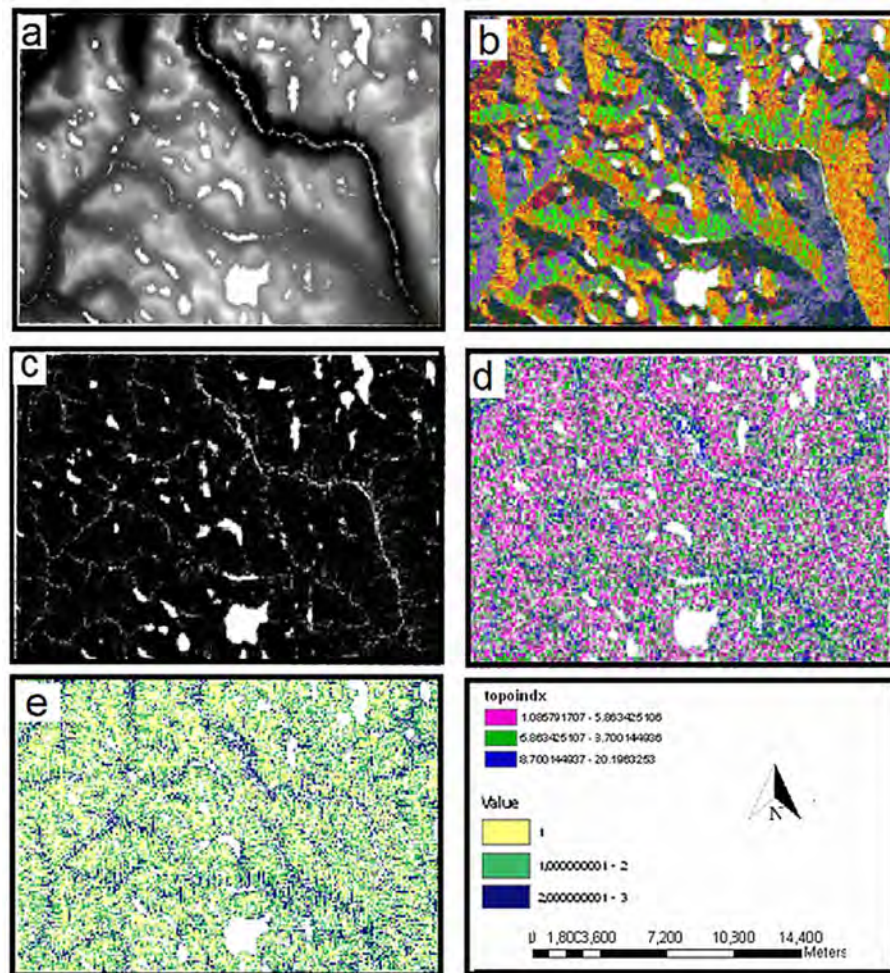
$$TWI = (As/\tan \alpha)$$

where, As is the specific catchments area and α is slope angle. For TWI fill sinks operation (Figure 4a), flow direction (Figure 4b), flow accumulation (Figure 4c) operations are applied to the DEM data. Resultant map of the TWI is given in the Figure 4d.

The result of the TWI is categorized into three classes by using natural breaks algorithm. This is done by seeking to minimize each class's average deviation from the class mean, while maximizing each class's deviation from the means of the other groups. In other words, the method seeks to reduce the variance within classes and maximize the variance between classes (Jenks, 1967). TWI categorization was performed by considering the high slope areas doesn't keep rain and has less susceptibility however hollow area keeps more water has high topographic index showing high susceptibility to landslide. For this reason, low index values between 1.08-5.86 areas are given a score (1) showing low susceptibility; medium index values between 5.86–8.70 are given a score (2) showing medium susceptibility and, finally high index values between 8.70-20 are assigned a score (3) showing high susceptibility. The result is shown in Figure 4e.

The lineament density of the area has a significant influence on the occurrence of landslides. Lineament density is also used in landslide hazard assessment (Atkinson & Massari 1998; Sarkar & Kanungo 2004; Suzen & Doyuran 2004) as it is

Figure 4. Topographic wetness index generation steps: fill sinks operation 4 (a), Flow direction (b), flow accumulation (c) wetness index is (d) and reclassified topographic index (e).

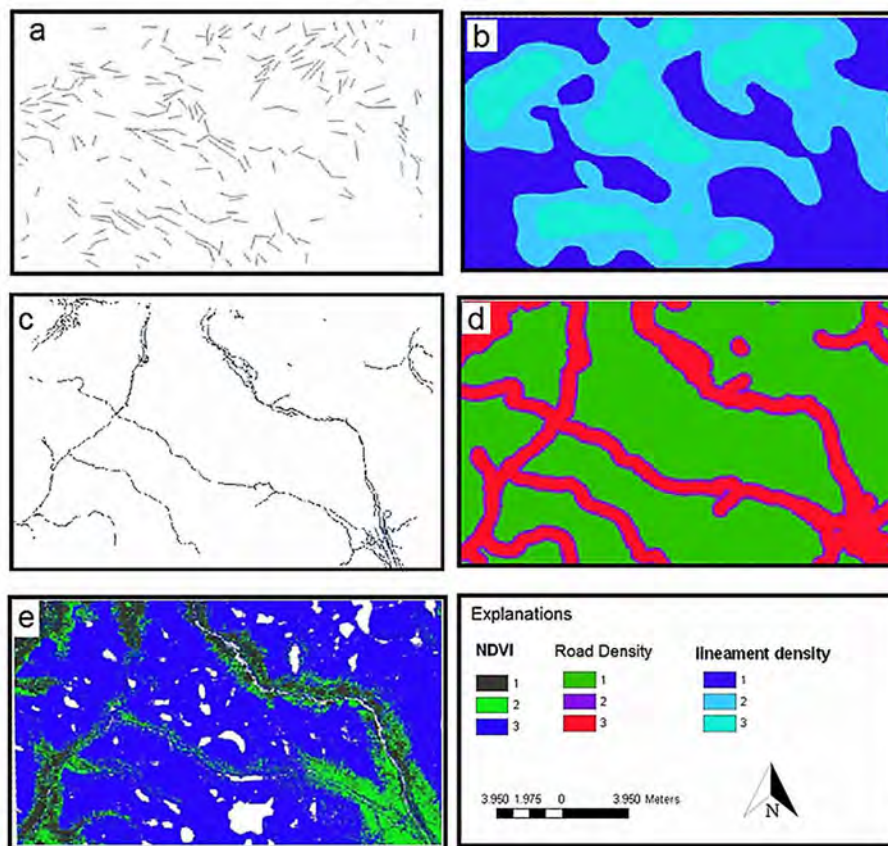


generally considered the probability of landslides occurring is greater in highly fractured areas compared to those with a lower fracture density. In the study lineaments are extracted from Landsat ETM image by using line option of PCI Geomatica software. Band 7 with a spatial resolution 30x30 meter is selected for the extraction of the lineaments considering the purpose of this study; since this band is useful for discrimination of lineaments and other geological features such as mineral and rock types and is also sensitive to vegetation moisture content (Sabins, 1996). The results of the automatically extracted lineaments are given in Figure 5a. Spatial density of the linea-

ments of the area is determined and this density again categorized into three classes. Low dense lineament between 0-0.29 is considered as least susceptible areas and scored with (1); Medium dense lineament between 0-0.63 is scored with (2), density of lineament higher than 0.63 being most susceptible to landslides is scored with (3) which result is shown in Figure 5b.

The external factors such as road construction also contribute to occurrence of the landslide. For that reason, the road network of the study area also is taken into consideration as a triggering factor. The road map of the study area is provided from NGU in a vector format (Figure 5c). The highest

Figure 5. Lineament map of the study area (a); Lineament density map (b); Road network of the study area (c); Buffered and Reclassified road network of the area (d), Vegetation Indices (e)



susceptible areas closest to road in 200m buffer area are given high susceptibility score (3). Road between 200-400m buffer area given medium susceptibility score (2), Road buffer >400 m is given low susceptibility score as shown in (Figure 5d).

The NDVI is produced using 4th and 3rd bands of Landsat ETM data the water areas are excluded from the map by using masking operation. The resultant map is grouped into three class and densely vegetated areas ranges with NDVI values between 0.33-0.98 and given score of 1; Medium dense areas with NDVI values between 0.12-0.33 are scored as 2; Low dense areas with NDVI values less than 0.012 scored as 3. Showing 1 as least susceptible and 3 as most susceptible to landslides, the results are illustrated in Figure 5e.

SOLUTIONS AND RECOMMENDATIONS

Final Map Generation

In order to generate landslide susceptibility map; a numerical rating system is applied to five factors that contribute to landslide occurrence: slope, TWI, NDVI, lineament density and distance to roads. Each of the five factors is grouped into three categories, and each category is assigned a value between 1 and 3, with 1 being least susceptible and 3 being the most susceptible to landslides. Scoring for slope, NDVI, and lineament density is based upon previous studies, which showed that higher slopes, less vegetation, and denser lineaments increase landslide susceptibility (Sarkar et al., 1995; Pachauri et al., 1998). The landslide

casual parameters are weighted by means of direct, pair-wise, and rank ordering comparison and the output is a composite index map (Castellanos and Van Westen 2007). Based on their relative importance to landslide occurrence in the study area, the five factors are assigned weights between 0.0 and 1.0 (collectively adding to 1.0). Slope is assigned the highest weight, followed by TWI, NDVI, lineament density and distance to roads. Different weights are assigned to five different layers as a result of weighing operation each weights produced different landslide susceptibility map. In order to define most suitable landslide susceptibility map of the area the accuracy of the generated map is tested with the previously defined landslide data which is shown in the accuracy assessment section in Table 3

Performance Evaluation of Final Susceptibility Maps

The accuracy of the three final landslide susceptibility maps are evaluated by using the GIS overlay analyses. For this operation a buffer zone of 30 m. is assigned to the previously defined real landslide data. 30m buffer is selected due to 1

mm shift at the location of the landslide locations on the map, which should always be considered because this amount is almost thickness of the pen used for producing landslide locations on the topographic map.

Accuracy of the results largely depends on the selection of the buffer amount. A small buffer will result in matching few landslide locations. In this study, different values for the buffer are tested. Larger buffer sizes intersect with nearly all highly susceptible zones. For this reason the amount of buffer is kept minimum, which is equal to the pixel size of the maps. Three different weighting is applied in this study in order to see the effects of the weights to the final map (Table 2). The results show that: generally, the medium susceptible areas show the highest accuracy as map 1 (Figure 6c) gives 70% (Table 3), map 2 (Figure 7c) and map 3 (Figure 8c) gives 95% accuracy (Table 3). The highest susceptibility accuracy is obtained in the map 2, it may be due to a weight factor of 40%, which is assigned to slope. It can be concluded that the weighting factors assigned to map 2 gives the better estimate of susceptible regions in the study region.

Figure 6. Results of final map 1 (a); High risky areas intersect with landslide points (b); Results of final map 3; Medium risky areas intersect with four landslide point s(c), and Low risky areas intersect with four landslide points (d)

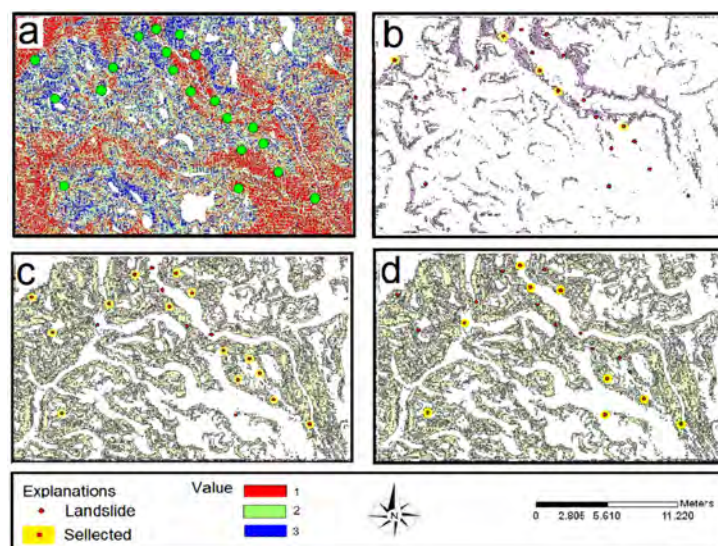


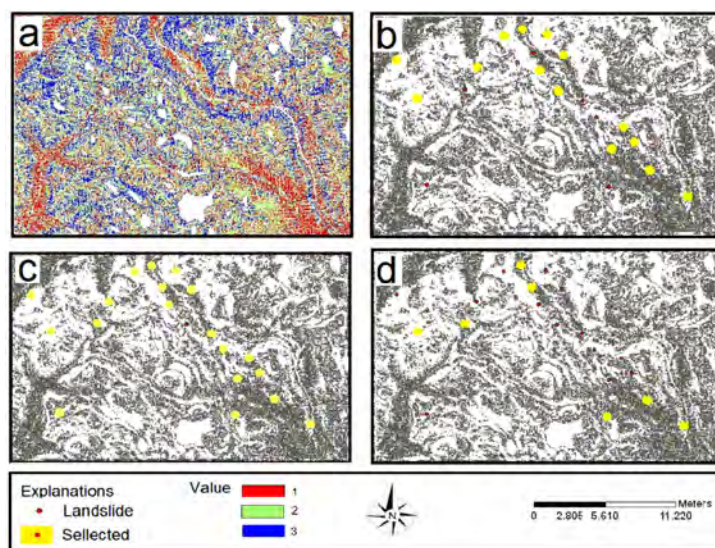
Table 2. Weighting factor used for map 1

Data Layers	Slope Gradient	Topographic Index	Vegetation Index (NDVI)	Lineament Density	Road Networks
Weighting factors of map1	0,3	0,25	0,2	0,15	0,1
Weighting factors of map2	0,4	0,3	0,15	0,1	0,05
Weighting factors of map3	0,35	0,15	0,15	0,25	0,10

Table 3. Accuracy of map 1, 2, and 3

	Total Landslide Points	Matching Points With Most Susceptible Areas	Matching Points With Medium Susceptible Areas	Matching Points With Less Susceptible Areas
Map 1	20	5	14	9
Accuracy		40%	70%	45%
Map 2	20	14	19	7
Accuracy		70%	95%	35%
Map 3	20	12	19	9
Accuracy		60%	95%	45%

Figure 7. Results of final map 2 (a); High risky areas intersect with landslide points (b); Results of final map 3; Medium risky areas intersect with four landslide point s(c), and Low risky areas intersect with four landslide points (d)



The proportions of susceptibility zones of the whole region are calculated. High risk zone is reached the maximum value (97.01 km²) in the Map 2. Minimum High risk zone value (41.26 km²) is in the Map1 as shown in the Table 4.

FUTURE RESEARCH DIRECTIONS

In this study it is aimed to focus on use of GIS and RS data for landslide susceptibility mapping. In order to discuss the performance of the study,

Figure 8. Result of Map 3 (a); High risky areas intersect with landslide points (b); Results of final map 3; Medium risky areas intersect with four landslide points (c), and Low risky areas intersect with four landslide points (d)

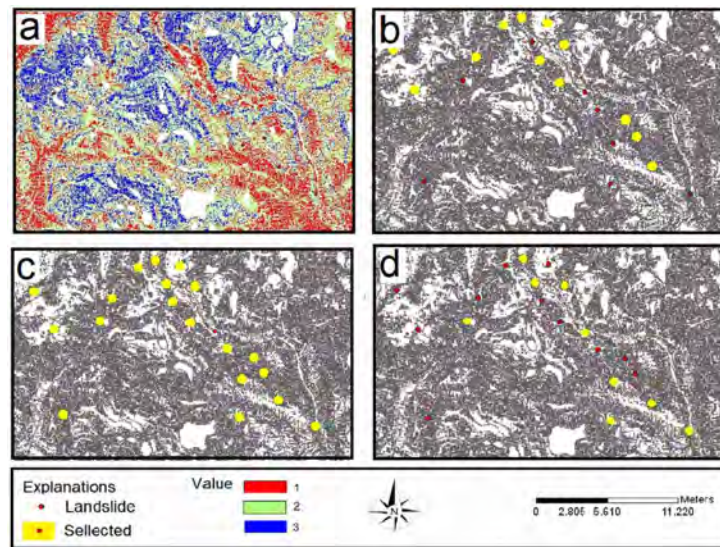


Table 4. Proportions of susceptibility zones to the whole region

	Area km ²			Proportion%		
	Low	Medium	High	Low	Medium	High
Map1	89.88	215.67	41.26	14.81	35.55	6.80
Map2	75.16	166.39	97.01	12.39	27.42	15.99
Map3	96.85	170.57	87.46	15.96	28.11	14.415
Total Region	606.76	606.76	606.76			

three different weighting factors were evaluated. The weights were obtained from a single landslide expert. The study may be extended in the future by preparing a survey to different landslide experts in order to obtain the weights. This may provide the researcher how the expert knowledge affects the performance of the susceptibility mapping (Wang, 2008). Another issue is the factors handled in the study. In this study five different factors are used in order to obtain the susceptibility maps. It is worth to study the influence of inclusion of different factors to the susceptibility mapping in the future. One other issue may be to evaluate the quality of the factors. Factors may involve uncertainty and

their propagation in susceptibility mapping should be analyzed in order to determine uncertainty in the obtained maps in a systematic way.

CONCLUSION

In the present study, landslide susceptibility mapping using combination of various factors responsible for landslide susceptibility is presented. Each factor has relative importance to probable landslide activity. A reliable and accurate susceptibility map depends on the inclusion and proper determination of the role of these parameters.

The accuracy of the generated maps can change according to the buffer size applied on the previously occurred landslide test points. Large buffer size will provide higher accuracy so it is a critical value and subjective process. The result of generated three different landslide maps according to five different variables assigned with different weight factors show that different weights result in different susceptibility maps. The comparison of the generated final susceptibility maps with previously occurring landslide points shows that the most important factor affecting the accuracy is the slope of the study area. This signifies that the weight assigned to the slope generate the highest accuracy in the high risk zones. Results of this study suggest that GIS and RS can effectively be used to compile data and overlay several data layers relevant to landslide susceptibility mapping.

REFERENCES

- Akgun, A., Kincal, C., & Pradhan, B. (2012). Application of remote sensing data and GIS for landslide risk assessment as an environmental threat to Izmir city (west Turkey). *Environmental Monitoring and Assessment*, 184(9), 5453–5470. doi:10.1007/s10661-011-2352-8 PMID:21915598
- Akgun, A., Sezer, E. A., Nefeslioglu, H. A., Gokceoglu, C., & Pradhan, B. (2011). An easy-to-use MATLAB program (MamLand) for the assessment of landslide susceptibility using a Mamdani fuzzy algorithm. *Computers & Geosciences*, 38(1), 23–34. doi:10.1016/j.cageo.2011.04.012
- Atkinson, P. M., & Massari, R. (1998). Generalised linear modelling of susceptibility to landsliding in the Central Apennines, Italy. *Computers & Geosciences*, 24(4), 373–385. doi:10.1016/S0098-3004(97)00117-9
- Castellanos, E., & Van Westen, C. J. (2007). Generation of a landslide risk index map for Cuba using spatial multi-criteria evaluation. *Landslide*, 4(4), 311–325. doi:10.1007/s10346-007-0087-y
- Glenn, N. F., Streutker, D. R., Chadwick, D. J., Thackray, G. D., & Dorsch, S. J. (2006). Analysis of LiDAR-derived topographic information for characterizing and differentiating landslide morphology and activity. *Geomorphology*, 73(1-2), 131–148. doi:10.1016/j.geomorph.2005.07.006
- Gokceoglu, C., Sonmez, H., Nefeslioglu, H. A., Duman, T. Y., & Can, T. (2005). The 17 March 2005 Kuzulu landslide (Sivas, Turkey) and landslide susceptibility map of its near vicinity. *Eng. Geol. B.*, 81(1), 65–83. doi:10.1016/j.eng-geo.2005.07.011
- Gonzalez, A., Ramon, J., Cendrero, A., & Fabirri, A. (2003). Validation of Landslide Susceptibility Maps; Examples and Applications from a Case Study in Northern Spain. *J. Natural Hazards*, 30(3), 437–449. doi:10.1023/B:NHAZ.00000007201.80743.fc
- Guzzetti, F., Paola, R., Mauro, C., Mirco, G., & Ardizzone, F. (2005). Probabilistic landslide hazard assessment at the basin scale. *Geomorphology*, 72(1-4), 272–299. doi:10.1016/j.geomorph.2005.06.002
- Jenks, George, F. (1967). The Data Model Concept in Statistical Mapping. *International Yearbook of Cartography*, 7, 186–190.
- Kavzoglu, T., Sahin, E. K., & Colkesen, I. (2013). Landslide susceptibility mapping using GIS-based multi-criteria decision analysis, support vector machines, and logistic regression. *Landslides*, 11(3), 425–439. doi:10.1007/s10346-013-0391-7
- Kokkila, M. (2002). *Digital elevation models (DEM) in trafficability analysis*. ECOWOOD.
- Lee, S., & Min, K. (2001). Statistical analysis of landslide susceptibility at Yongin, Korea. *Environ Geol.*, 40(9), 1095–1113. doi:10.1007/s002540100310
- NGI. (n.d.). Retrieved March 1, 2016, from <http://www.ngi.no/en/Geohazards/Research/Stability-of-rock-slopes/>

Ohlmacher, G. C., & Davis, J. C. (2003). Using multiple logistic regression and GIS technology to predict landslide hazard in northeast Kansas, USA. *Engineering Geology*, 69(3-4), 331–343. doi:10.1016/S0013-7952(03)00069-3

Pachauri, A. K., Gupta, P. V., & Chander, R. (1998). Landslide zoning in a part of the Garhwal Himalayas. *Environ Geol.*, 36(3-4), 325–334. doi:10.1007/s002540050348

Pradhan, B. (2013). A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. *Computers & Geosciences*, 51, 350–365. doi:10.1016/j.cageo.2012.08.023

Sabins, F. F. (1996). *Remote Sensing: Principles and Interpretation* (3rd ed.). W. H. Freeman and Company, New York Press.

Sarkar, S., & Kanungo, D. P. (2004). An integrated approach for landslide susceptibility mapping using remote sensing and GIS. *Photogrammetric Engineering and Remote Sensing*, 70(5), 617–625. doi:10.14358/PERS.70.5.617

Sarkar, S., Kanungo, D. P., & Mehrotra, G. S. (1995). Landslide hazard zonation: A case study in Garhwal Himalaya, India. *Mountain Research and Development*, 15(4), 301–309. doi:10.2307/3673806

Soeters, R., & Van Westen, C. J. (1996). Slope instability recognition, analysis, and zonation. In A. K. Turner & R. L. Shuster (Eds.), *Landslides Investigation and mitigation*, Transp. Res. Board, Special Report 247, Natural Acad (pp. 129–177). Washington, DC: Press.

Suzen, M. L., & Doyuran, V. (2004). Data driven bivariate landslide susceptibility assessment using geographical information systems: A method and application to Asarsuyu catchments, Turkey. *Engineering Geology*, 71(3-4), 303–321. doi:10.1016/S0013-7952(03)00143-1

USGS. (n.d.). Retrieved March 1, 2016, from <http://rockyweb.cr.usgs.gov/nmpstds/acrodoks/dem/3DEM0897.PDF>

Vames, D. J. (1984). Landslide hazard zonation: A review of principles and practice, Unesco. Paris. *Natural Hazards*, 3, 63.

Wachal, D. J., & Hudak, P. F. (2000). Mapping landslide susceptibility in Travis County, Texas, USA. *GeoJournal*, 51(3), 245–253. doi:10.1023/A:1017524604463

Wang, R. (2008). *An expert knowledge-based approach to landslide susceptibility mapping using GIS and fuzzy logic* (PhD Dissertation). University of Wisconsin-Madison, UMI Number: 3314311.

ADDITIONAL READING

Eker, A. M., Dikmen, M., Cambazoğlu, S., Düzgün, H. S. B., & Akgün, H. (2015). Evaluation and comparison of landslide susceptibility mapping methods: A case study for the Ulus district, Bartın, Northern Turkey. *International Journal of Geographical Information Science*, 29(1), 132–158. doi:10.1080/13658816.2014.953164

Erener, A., & Düzgün, H. S. B. (2010). Improvement of statistical landslide susceptibility mapping by using spatial and global regression methods in the case of More and Romsdal (Norway). *Landslides*, 7(1), 55–68. doi:10.1007/s10346-009-0188-x

Erener, A., & Düzgün, H. S. B. (2012). Landslide susceptibility assessment: What are the effects of mapping unit and mapping method? *Environ Earth Sci.*, 66(3), 859–877. doi:10.1007/s12665-011-1297-0

Erener, A., & Düzgün, H. S. B. (2013). A regional scale quantitative risk assessment for landslides: Case of Kumluca watershed in Bartın, Turkey. *Landslides*, 10(1), 55–73. doi:10.1007/s10346-012-0317-9

Erener, A., Mutlu, A., & Düzgün, H. S. (2016). A comparative study for landslide susceptibility mapping using GIS-based multi-criteria decision analysis (MCDA), logistic regression (LR) and association rule mining (ARM). *Engineering Geology*, 203, 45–55. doi:10.1016/j.eng-geo.2015.09.007

Sassa, K., Canuti, P., & Yin, Y. (2014). *Landslide Science for a Safer Geoenvironment*, (Methods of Landslide Studies), New York, Springer, pp.2. Michael, E.A. and Samanta S. (2016). Landslide vulnerability mapping (LVM) using weighted linear combination (WLC) model through remote sensing and GIS techniques. *Modeling Earth Systems and Environment*, 2, 88.

Wang, R. (2008). *An expert knowledge-based approach to landslide susceptibility mapping using GIS and fuzzy logic*. Doctor of philosophy, at the university of Wisconsin-Madison, UMI Number: 3314311, (157)

KEY TERMS AND DEFINITIONS

DEM: Digital Elevation Model is the 3D representation of a terrain's surface.

GIS: Geographic Information Systems is a technology that is used to capture, store, manipulate, analyze, manage, and present spatial data.

Landslide Susceptibility: Landslide susceptibility is investigating the spatial likelihood of occurrence of landslides by correlating principal factors with landslide inventory data.

NDVI: Normalized Difference Vegetation Index is a most well-known index to detect vegetation and their condition in an area by using bands of remote sensing data.

RS: Remote sensing is the acquisition of information from earth without making physical contact with the object by using satellite- or aircraft-based sensor technologies.

TWI: Topographic Wetness Index describes the effect of topography on the location and size of saturated source areas of runoff generation.

Weighting Factors: Weighting Factors are estimated impact of values indicating relative importance.

Category G

Government and Law

Accessibility in E-Government

Christian Sonnenberg

Florida Institute of Technology, USA

INTRODUCTION

Access to government-created digital content has come a long way since the early days of the Web. The growth of Internet usage coupled with advances in technology and reduced costs has made the use of Electronic Government (E-Government) resources and websites a primary means of content access for citizens in the United States. According to the latest census and national records, over 86% of the US population has access to the Internet with a continued growth rate of 7% from the previous year. (Internet Users by Country, 2014) However, as illustrated by the Healthcare.gov website launch in 2013, technological, accessibility, and usability issues can seriously hinder and cause detrimental effects for user relying on these services. How content is presented and delivered on the Web makes an impact on how effective and helpful it is, but even more so for users with disabilities. This paper will cover the methods and standards of digital government content, the compliance with accessibility guidelines for disabled users, current challenges, and possible avenues of future delivery methods.

Web accessibility as defined by the World Wide Web Consortium (W3C) is the means by which anyone regardless of physical or cognitive disability can use and operate a website (W3C Introduction to Web Accessibility, 2005). People with disabilities or normal aging considerations find it difficult if not impossible to use technology that nondisabled individuals could use freely. For example, a blind user visiting a website must rely on screen-reading technology to interpret the site while a nondisabled user can browse it without any additional assistance. In order to achieve accessibility in their websites, a number of rules

and guidelines have been developed by the federal government. In 1998, congress amended the Rehabilitation Act of 1973¹ with Section 508 to require federal agencies to make electronic and information technology accessible to people with disabilities. Section 508 was enacted to eliminate barriers in electronic and information technology by requiring that disabled users have access to government information that is comparable to the access available to others without disabilities (www.section508.gov).

While accessibility focuses on the ability for all users, regardless of disability, to interact with content, another attribute is also important. Usability is how effectively, efficiently and satisfactorily a user can interact with a user interface (Chou & Hsiao, 2007). A focus on usability implies that a site is designed for easier access of content and information, which affects all users. The International Organization for Standardization (ISO) interprets usability as effectiveness, efficiency and satisfaction with which the user achieves specific goals in the specified context of use (ISO, 1998). Although not as strictly defined or required like Section 508, The U.S. Web Design Standards were developed as the U.S. government's very own set of common components and designs for websites. It's structured to make things easier for government site developers, while raising the bar on what users expect from their digital experience. Many of these standards are built upon the existing section 508 standards in hopes of taking them one step further.

Since the adoption of Section 508, compliance has been slow, but steady. However, technology advances quickly, and with it comes new challenges and concerns that were not considered in the original requirements. The advent of the

DOI: 10.4018/978-1-5225-2255-3.ch305

mobile revolution has highlighted an accessibility dilemma with web sites that were never intended for use on small devices. With global mobile phone use at an all-time high, developers are racing to adapt content to fit these new screens. It is important to understand the current criteria put down in Section 508 to define accessibility standards, what constitutes compliance, and what updates need to be applied to accommodate for the advent of mobile usage. Furthermore, advances in mobile technologies and the shift to mobile app usage have raised additional questions, such as how the platform can become an interactive, dynamic process rather than the traditional passive distribution of content. Generating greater user engagement that translates into information access, service utilization, and participation in government decision helps both empower the user and provide valuable feedback to the government. Usability and accessibility will present challenges to citizens' acceptance and adoption of more advanced services and will influence their day-to-day interactions with e-government websites (Clemmensen & Katre, 2012).

BACKGROUND

The design and accessibility of government websites today is driven by a particular set of criteria known as Section 508. This amendment, which went into effect in June 2001, requires all federal agencies to comply with accessibility standards administered by the Architectural and Transportation Barriers Compliance Board (referred to as the Access Board)². These standards ensure that electronic and information technology is accessible to disabled persons to the extent it does not pose an "undue burden" on an agency. When Section 508 went into effect, federal agencies could no longer procure noncompliant electronic and information technology (Charles, 2001). Therefore, vendors who supply hardware, software, web sites, and other information technologies, have to ensure

compliance with Section 508 in order to be eligible for government contracts. The original standards were organized into three areas including: accessibility of operation and information, compatibility with peripheral devices, and documentation and services associated with electronic and information technology. Although mobile devices and tablets are not specifically listed, the inclusion of "web sites" as information technology necessitates that they be accessible regardless of the target output, whether it is desktop or mobile.

The U.S. Access Board's Section 508 web site summarizes each section of the technology standards as they apply to users with disabilities. These standards focus on assistive technologies (such as screen reader devices for blind users) and alternative technologies (e.g., keyboard navigation instead of mouse navigation) that allow access to those with disabilities. However, not all standards can be broadly applied to every device that accesses the Web. Established in 2000, the original Section 508 standards have become antiquated with current technology. While the standards were intended for traditional desktop websites, they were not considered for mobile sites. Since 2006, the Access Board has attempted to update these guidelines to account for new technologies (2006). On February 18, 2015, the Board released a proposed rule updating the 508 Standards and the 255 Guidelines. The Board held public hearings on the proposed rule in San Diego, Washington, DC and Salt Lake City. However, as of November of 2015 no official changes have been enacted (U.S. Access Board, 2015).

WEB ACCESSIBILITY

It is important to understand the history that drives Section 508 and the changes being put into place by the current proposal. The W3C, as the standards setting body for the Web, provided support to the Web Accessibility Initiative (WAI) in the development of Web Content Accessibility

Guidelines (WCAG) (Chisolm, Vanderheiden, & Jacobs, 2001). The original 1.0 version of these guidelines forms the foundation for the Section 508 Web accessibility standard. The WAI guidelines for Web content accessibility focus on making online information accessible to those with disabilities. This includes text, images, links, audio, and other elements that compose a Web page or application. One such example of these guidelines is as follows:

1194.22 (a) A text equivalent for every non-text element shall be provided. Pictures, graphs, and other elements on a Web page not in electronically readable form are supplemented with a text description to be read by screen reader software and Braille displays.

Such guidelines clearly present the issue at hand, such as users with vision impairment, followed by the solution, text alternatives for screen readers. Such rules that are not device or context specific can be safely applied to mobile sites because the technology exists to interpret text descriptions of images regardless of platform. However, other rules have not aged as well. For example, the following rule stipulates the inclusion of plug-in sources:

1924.22 (m) When a web page requires that an applet, plug-in or other application be present on the client system to interpret page content, the page must provide a link to a plug-in or applet that complies with §1194.21(a) through (l).

While this is fine for a desktop environment, a mobile device is typically limited in the plug-ins available. Even if a link is provided, it may not be operational with said device and the user will be unable to retrieve the document. Apple devices, for example, will not run content that requires a Flash plug-in. An alternative guideline might stipulate the need for external documents to be made available in formats accessible by the target device.

For a full list of the WAI guidelines, please see the WCAG 1.0 guidelines available at <http://www.w3.org/TR/WCAG10/>.

WEB USABILITY

While the Section 508 guidelines focus primarily on accessibility rules, they also affect usability. Evidence from some studies suggests that e-government websites can benefit from a high level of usability in at least two ways. First, an e-government website provides a first impression of a government and its online services. Therefore, to foster good will and active participation with constituents, an e-government should aim for user-friendliness (Baker, 2009). Second, usability improves users' performance, making interactions more efficient, as well as increasing citizens' satisfaction with e-government (Verdegem & Verleye, 2009). One study identified a number of usability issues on e-government websites such as "links show different colors", "difficulty in using online help", and "categories without a logical order". This study suggests that usability issues can potentially hamper the adoption rate of such sites (Huang & Benyoucef, 2014).

Although the Section 508 guidelines do have tangential influence on usability, they do not address it specifically. The U.S. Web Design Standards were created as a collaboration between 18F and the U.S. Digital Service (USDS) to fill in the Section 508 gaps, under the guidance of an advisory board from the CFPB, FDA, DVA, SSA, IRS, and the Department of Education. They incorporate specific UI components on an HTML foundation, utilizing a technique called "progressive enhancement" to deliver the same core content regardless of target platforms. Included in these guidelines is a visual style guide that defines typography and color recommendations that are 508 compliant and designed for readability. They also incorporate common UI components and patterns, a collection of foundational interface elements for all U.S. government sites. (U.S. Web Design Standards, 2015)

MOBILE ACCESSIBILITY FACTORS

The WAI guidelines that Section 508 relies on were originally intended to be broad enough to cover any viewing device. However, these guidelines were put in place before mobile devices became a major consumer of Web content. Mobile devices present a number of issues and concerns that may not be adequately addressed by the current Section 508 standards.

One of the most important factors affecting handheld devices is screen size. Mobile devices are mobile because they are small in form factor, but only a limited amount of information can be displayed at any given time. Studies show that comprehension of handheld content is dramatically lower than desktop counterparts (Singh, Sumeeth & Miller, 2011). Input is another factor as touch screens have issues regarding precision and manipulation. Hands are not well adapted to handle small screens where the space between buttons is minuscule (Siegenthaler et al., 2012). These factors become magnified for users with a disability such as poor vision or cognitive limitations. For example, accuracy demands are much higher on smaller screens, forcing users to precisely select buttons that might normally be bigger on a desktop interface. As alternatives are often limited in the mobile environment, it is very important that mobile sites be designed to limit or remove problematic content.

Ideally, mobile web sites are ones that have been specifically crafted for mobile devices. This allows developers the chance to reengineer a site to better align with the context of mobile usage by eliminating extraneous content. In practice, an agency may find it difficult to spend the manpower necessary to make a separate mobile site, in addition to the original desktop site, simultaneously ensuring that both comply with Section 508. On the other hand, mobile-specific web sites are generally more compliant with Section 508 because they have removed so much content in the first place.

The simplicity of a mobile site lends itself well to accessibility compliance, but it requires addi-

tional effort on the part of the agency. A number of adaptation mechanisms have been developed over the years to try and alleviate the development burden. Segmentation, summarization, and filtering are popular techniques used to take existing desktop sites and make them usable on handheld devices. One such mechanism invokes a method of analyzing site metrics to identify important content and displaying it first (Sonnenberg, 2013). These metrics consider such factors as semantics, location, context, and user preferences and then restructure the entire site to provide enhanced usability to each particular user. These metrics could be further extended to consider additional disability requirements, allowing a certain degree of accessibility compliance.

LOCALIZATION AND PERSONALIZATION

One of the most useful attributes of mobile devices is their access to position and location information. This allows the device to tailor its usage to the context of the user. For example, GPS information can be used to determine how close a person is to the nearest government facility with the services they need. This also allows for dynamic personalization of the user experience, incorporating context and location to create a unique experience to cater to specific constituents (Ekelin & Eriksén, 2015).

As an accessibility feature, localization can be used as a means of redefining the user experience within the context of a disabled user. Therefore, it is worth considering incorporating such features when applicable into the Section 508 guidelines. If location information is available, a user with disabilities could be provided with audio or visual cues to highlight their surroundings and get their bearings. Furthermore, the ability to collect location information gives e-government content providers more information about their constituents' behavior and focus. This in turn allows them to build more robust and meaningful websites.

WCAG 2.0 UPDATES

Though there are methods of applying Section 508 to mobile web sites, the main issue is that Section 508 is simply not designed for them. The WCAG 1.0 guidelines were established in 1999. Since then the Web has undergone tremendous changes in the methods of content display and presentation. In 2008, W3C created the new WCAG 2.0 guidelines to accommodate these new developments (Reid, Snow-Weaver, 2008). As of November 2013, federal agencies and government websites are still only required to conform to Section 508 guidelines, itself a representation of WCAG 1.0 principles. However, these guidelines are currently being updated by the Telecommunications and Electronic and Information Technology Advisory Committee (TEITAC) to conform to the new WCAG 2.0 standards (U.S. Access Board, 2015). While the debate is still ongoing, it is important to understand the critical differences between WCAG 1.0 and 2.0 and how that affects Section 508.

The WCAG 2.0 standards were developed to reduce the device dependency of WCAG 1.0 and present technology-independent guidelines. This is important as the Web is now accessed from a myriad number of devices. These standards apply more broadly to technology in general and were designed to be applied to future devices as well (Alonso et al., 2010). WCAG 2.0 is also designed to be more precisely tested and evaluated. It allows for easier automation in testing and integration into design specification.

The evaluation of conformance metrics of WCAG 2.0 differ from 1.0 in the newly created “success criteria.” Originally, 1.0 utilized checkpoints to verify conformance. The new 2.0 standards present three layers of conformance criteria ranked A (lowest), AA, and AAA (highest) (Keith, Floratos, & Whitney, 2012). The highest levels of conformance are typically the most difficult to adopt, but provide the highest degree of assured accessibility. These layers of conformance allow for a greater degree of flexibility in evaluation.

The Section 508 update is slated to implement levels A and AA requirements from the WCAG 2.0 guidelines, which is considered in line with the global standard. Section 508 will also broaden the definition of information technology to more readily handle the advancements of technology in the future (U.S. Access Board, 2013). The W3C have also come up with a number of best practices regarding mobile specific web sites. While not all of these recommendations are necessary, it is planned to incorporate these guidelines into the Section 508 updates. For a more detailed look at these guidelines please visit the W3C site: <http://www.w3.org/TR/mobile-bp/>.

WEB ACCESSIBILITY COMPLIANCE

Since the adoption of Section 508, overall compliance has been sporadic. Stowers (2002) conducted a study on 148 federal Web sites and found that only 13.5% of the sites met the accessibility standards of Section 508. In the years following the release of Section 508, improvements were made in the area of compliance, specifically with the introduction of support tools such as the W3C Markup Validator Suite. These tools provide automated feedback on noncompliance errors associated with the Web content accessibility guidelines. W3C also provides a markup tool specifically designed for mobile devices (W3C, 2013).

However, in recent years, compliance rates have drastically dropped down to and below the original adoption rates. One study by Olalere and Lazar (2011) tested 100 popular US Government websites and discovered that 90% of home pages were in breach of Section 508. The issues discovered on these websites were varied, largely relating to a lack of sufficient alternative text for images and a lack of captions on video. These studies also reveal a lack of compliance across sites regardless of device, be it desktop or mobile.

The overall drop in compliance can be attributed to a number of different factors. New

browsers, devices, and platforms increase the chance of failure, as websites are ill-prepared for translation to these permutation (Duchateau, Boulay, & Burger, 2010). The growth in the number of websites is another contributing factor. Since 1999, the number of total websites has to over 4.37 billion indexed pages as of 2015 (World Wide Web Size, 2015). Lastly, as stated previously, Section 508 has not been formally updated in years and agencies may feel their limitations stifle usage in the current Web environment.

Although the current situation may appear alarming, the U.S. Government is refreshing Section 508 to make it more closely aligned with version 2.0 of the WCAG (U.S. Access Board, 2015). Studies also show that small number of rules cause a disproportionate number of failures in websites (Reid, Snow-Weaver, 2008). Improvements in compliance metrics could be achieved by prioritizing these in the new WCAG 2.0 layered approach. Once WCAG 2.0 principles are incorporated into Section 508, it is likely to lead to accessibility improvements across all e-government websites.

MOBILE ACCESSIBILITY INITIATIVE

In an effort to promote and deliver mobile access to government employees and constituents, the U.S. General Services Administration (GSA) has launched an initiative called Mobile Gov (GSA Mobile Government Toolkit, 2015). One of Mobile Gov's primary objectives is to provide services and information to the public regardless of how they access it. GSA created HowTo.gov to support agencies in developing mobility presences, whether it is a web site or application, and provides procedures to maintain Section 508 compliance.

A number of applications have arisen from this initiative. One such example is the Transportation Security Administration (TSA) "My TSA" app designed specifically for mobile devices. This app provides answers to common questions such as what items are allowed on flights, security

wait times, and other regulations. The "My TSA" mobile application is completely accessible for people with disabilities and meets the full U.S. Section 508 compliance standards. While this is an application, as opposed to a web site, the lessons learned can be applied to both domains.

MOBILE QR INTEGRATION

One particular method of interaction has been gaining traction as a method of improving e-government accessibility and usability. Quick Response (QR) codes are a type of matrix barcode (or two-dimensional code) that are much faster than traditional UPC barcodes. These codes can be scanned and read by a QR-Code reader, software that is installed on most mobile devices (Mobile Barcodes.com, 2012). QR codes provide a cheap, easy, and secure method to transmit information in a "push" format to individuals. QR codes, when appropriately used, can serve as a way of enhancing the scope and scale of government services and there is a growing push for mobile barcodes in the government sector (ICMA, 2010). Since QR codes are easy to read and generate, they promote real-time data sharing of government content. They also make data entry and data management easy from an administrative perspective.

QR codes have interesting possibilities for users with disabilities. One particular use for QR codes in e-government is the way pointing systems in national parks. While navigating park trails, visitors might get lost and not be able to get a cell signal. QR codes on trail markers can help hikers get their bearings in the events of disorientation or a GPS failure. The trail markers have QR codes that have coordinates to a map already contained within the parks application on the device. (Lorenzi et al., 2014) This may tremendously help users with cognitive disabilities.

QR codes were established as an ISO (ISO/IEC18004) standard and are capable of handling many types of data, such as numeric and alphabetic characters, symbols, binary, and control codes. QR

code data capacity varies based on data type but they can encode a website URL, video links, social media links, geo-coordinates, email addresses and messages, purchase links, and many others. Data can even be restored even if the symbol is partially dirty or damaged (Lorenzi et al., 2014). QR codes are capable of omni-directional, high speed reading, making them a very robust tool to empower mobile users. Perhaps the most important benefit of QR codes is that they exist as an alternative for devices that may be limited in interaction. For example, blind users can “read” the surroundings with such a device and not worry about interactions with the screen of the device. Therefore, it is worth considering incorporating the QR code and specific standards in their use into the upcoming Section 508 updates.

FUTURE COMPLIANCE

Compliance with Section 508 is improving with the advent of resources and tools provided by government, nonprofit organizations, academic institutions, and commercial vendors. The WAI guidelines provide important information about each accessibility barrier and potential solutions to it (2006). The University of Melbourne has provided a checklist for Level AA compliance with WCAG 2.0, which may prove useful when Section 508 formally adopts the standard (2015).

However, accessibility compliance has not reached its full potential. As compliance rates stagnate, the standards become out of date with each passing year. A number of 508 guidelines are too rigid or difficult to interpret in today’s environment. New technologies and languages that were not considered at the foundation of Section 508 need to be considered more closely as well. For example, QR codes and GPS localization provide alternative solutions to mobile interactions and should be highlighted to inform developers and agencies.

Research has suggested that a rating system for “no,” “partial,” or “full” compliance, as well

as, an identification component for whom the site is accessible, would be beneficial to the Section 508 guidelines (Jaegar, 2004). This trend is finally seeing fruition with the new WCAG 2.0 Layer approach. Government sites will be categorized by low or high compliance according to how strict they adhere to each layer. With amendments that accommodate these issues, Section 508 should see a much greater adoption rate in the future.

CONCLUSION

Section 508 has certainly made headway in bridging the digital divide, especially for those with disabilities. It has provided the structure for building e-government sites properly and accessible to all citizens. However, automated technologies and checker systems alone will not improve compliance with Section 508. These tools can only guide the developer through an accessibility assessment and may not uncover barriers related to meaningful content. In addition, the growth and advancement of mobile technology and new design techniques has caused Section 508 to become outdated.

Significant changes must be made to accommodate the growth in technology and the mobile environment. Section 508 is in the process of being updated, incorporating a number of features from WCAG 2.0 to counter the weaknesses of its current iteration. These improvements, coupled with initiatives to foster proper mobile web design, will provide agencies with a much better solution towards improving accessibility across all government web sites.

REFERENCES

- W3C. (2015). *W3C mobileOK Checker*. Retrieved on November 26, 2015, from <http://validator.w3.org/mobile/>
- W3C. (2015). *Web Accessibility Evaluation Tools*. Retrieved on November 26, 2015, from <http://www.w3.org/WAI/ER/tools/complete>

- W3C Introduction to Web Accessibility. (2005). *W3C: Web Accessibility initiative*. Retrieved on January 8, 2013 from <http://www.w3.org/WAI/intro/accessibility.php>
- Alonso, F., Fuertes, J., González, A., & Martínez, L. (2010). Evaluating conformance to WCAG 2.0: open challenges. In *Proceedings of the 12th international conference on Computers helping people with special needs: Part I (ICCHP'10)* (pp. 417-424). Berlin: Springer-Verlag. doi:10.1007/978-3-642-14097-6_67
- Baker, D. L. (2009). Advancing e-government performance in the United States through enhanced usability benchmarks. *Government Information Quarterly*, 26(1), 82–88. doi:10.1016/j.giq.2008.01.004
- Charles, S. (2001). *Section 508 compliance—Not a simple yes or no*. Immix Group. Retrieved on November 14, 2005, from http://www.immixgroup.com/pslibrary/eu_display.cfm?ID=91
- Chisholm, W., Vanderheiden, G., & Jacobs, I. (1999). *Accessibility Guidelines (WCAG) 1.0*. W3C. Retrieved on November 18, 2013, from <http://www.w3.org/TR/WCAG10/>
- Chisolm, W., Vanderheiden, G., & Jacobs, I. (2001). Web content accessibility guidelines. *Interactions of the ACM*, 8(4), 35–54. doi:10.1145/379537.379550
- Clemmensen, T., & Katre, D. (2012). *Adapting e-gov usability evaluation to cultural contexts. Usability in government systems*. Boston: Morgan Kaufmann.
- Duchateau, S., Boulay, D., & Burger, D. (2010). Assessing WCAG 2.0 conformance in practice. In *Proceedings of the 12th international conference on Computers helping people with special needs: Part I (ICCHP'10)* (pp. 408-412). Berlin: Springer-Verlag. doi:10.1007/978-3-642-14097-6_65
- Ekelin, A., & Eriksén, S. (2015). Citizen-Driven Design: Leveraging Participatory Design of E-Government 2.0 Through Local and Global Collaborations. In *Case Studies in e-Government 2.0* (pp. 67-85). Springer International Publishing.
- Electronic and Information Access Advisory Committee (EI-TAAC). (1999). *Electronic and information access advisory committee recommendations for accessibility standards: Electronic and information technology*. Retrieved on October 9, 2005, from <http://www.access-board.gov/sec508/com-mrept/eitaac.pdf>
- Federal Register. (2006). Telecommunications Act Accessibility Guidelines. *Federal Register*, 71(74). Retrieved on Jan 6, 2014, from <http://www.gpo.gov/fdsys/pkg/FR-2006-04-18/html/E6-5761.htm>
- Federal Register. (2015). *Proposed Information and Communication Technology (ICT) Standards and Guidelines*. Retrieved on November 26, 2015 from <http://www.access-board.gov/guidelines-and-standards/communications-and-it/about-the-ict-refresh/proposed-rule>
- GSA Mobile Government Toolkit. (2015). Retrieved on November 24, 2015 from <http://www.gsa.gov/portal/content/302057>
- Huang, Z., & Benyoucef, M. (2014). Usability and credibility of e-government websites. *Government Information Quarterly*, 31(4), 584–595. doi:10.1016/j.giq.2014.07.002
- ICMA Knowledge Network. (2010). *Redefining Government Communication with QR-Codes*. Retrieved on November 24, 2015, from http://icma.org/en/icma/knowledge_network/documents/kn/Document/301861/Redefining_Government_Communication_with_QRCodes
- Internet Users by Country. (2014). Retrieved on November 25, 2015 from <http://www.internetlivestats.com/internet-users-by-country/>

- ISO. (1998). *ISO 9241-11:1998, ergonomic requirements for office work with visual display terminals (VDTs) - Part 11: Guidance on usability*. Retrieved on November 24, 2015 from: <http://www.it.uu.se/edu/course/homepage/acsd/vt09/ISO9241part11.pdf>
- Jaeger, P. T. (2004). Beyond Section 508: The spectrum of legal requirements for accessible e-government web sites in the United States. *Journal of Government Information*, 30(4), 518–533. doi:10.1016/S1352-0237(04)00057-7
- Keith, S., Floratos, N., & Whitney, G. (2012). Certification or conformance: making a successful commitment to WCAG 2.0. *Proceedings of the International Cross-Disciplinary Conference on Web Accessibility (W4A '12)*. New York, NY: ACM doi:10.1145/2207016.2207029
- Lorenzi, D., Vaidya, J., Chun, S., Shafiq, B., & Atluri, V. (2014). Enhancing the government service experience through QR codes on mobile platforms. *Government Information Quarterly*, 31(1), 6–16. doi:10.1016/j.giq.2013.05.025
- Matthews, W. (2002). One year and counting: Section 508. *Federal Computer Week*. Retrieved on June 24, 2006, from <http://www.fcw.com/article77009-06-24-02-Print>
- Mobile Barcodes.com. (201). *QR-code readers*. Retrieved on November 24, 2015 from <http://www.mobile-barcodes.com/qr-code-software/>
- Olalere, A., & Lazar, J. (2011). Accessibility of U.S. federal government home pages: Section 508 compliance and site accessibility statements. *Government Information Quarterly*, 28(3), 303–309. doi:10.1016/j.giq.2011.02.002
- Reed, M. A. T. (2003, August 11). Agencies still on the learning curve. *Federal Computer Week*. Retrieved on October 22, 2005, from <http://www.fcw.com/article80520-08-11-03-Print>
- Reid, L. G., & Snow-Weaver, A. (2008). WCAG 2.0: a web accessibility standard for the evolving web. *Proceedings of the 2008 international cross-disciplinary conference on Web accessibility (W4A '08)* (pp. 109-115). New York: ACM Press. doi:10.1145/1368044.1368069
- Siebra, C., Anjos, M., Florentin, F., Gouveia, T., Correia, W., Penha, M., & Santos, A. L. et al. (2015, August). Accessibility Devices for Mobile Interfaces Extensions: A Survey. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct* (pp. 644-651). ACM. doi:10.1145/2786567.2793683
- Siegenthaler, E., Bochud, Y., Wurtz, P., Schmid, L., & Bergamin, P. (2012). The effects of touch screen technology on the usability of e-reading devices. *Journal of Usability Studies*, 7(3), 94–104.
- Singh, R., Sumeeth, M., & Miller, J. (2011). Evaluating the readability of privacy Policies in mobile environments. *International Journal of Mobile Human Computer Interaction*, 3(1), 55–78. doi:10.4018/jmhci.2011010104
- Size of the World Wide Web. (2015). *Word Wide Web Size: Daily estimated size of the world wide web*. Retrieved on November 25, 2015, from <http://www.worldwidewebsite.com/>
- Sonnenberg, C. (2013). *An Autonomous Framework for Mobile Web Content Adaptation and Prioritization* (Ph.D. Dissertation). Florida Institute of Technology, Melbourne, FL.
- Stowers, G. N. L. (2002). *The state of Federal Websites: The pursuit of excellence*. Pricewaterhouse Cooper Endowment for the Business of Government report. Retrieved on June 24, 2006, from <http://www.businessofgovernment.org/pdfs/StowersReport0802.pdf>
- University of Melbourne. (2015). *WCAG Checklist – Web Accessibility*. Retrieved on November 20, 2015, from <http://www.unimelb.edu.au/accessibility/training/wcagchecklist.html>

U.S. Access Board. (1999). *Questions & Answers about Section 508 of the Rehabilitation Act Amendments of 1998*. Retrieved on June 24, 2006 from, <http://www.access-board.gov/sec508/FAQ.htm#14>

U.S. Access Board. (2015). *Provisions considered for update in Section 508*. Retrieved from <http://www.access-board.gov/508.htm>

Youngblood, N. E., & Mackiewicz, J. (2012). A usability analysis of municipal government website home pages in Alabama. *Government Information Quarterly*, 29(4), 582–588. doi:10.1016/j.giq.2011.12.010

ADDITIONAL READING

A Guide to Disability Rights Laws. (n.d.). Civil Rights Division, Disability Rights Section. U.S. Department of Justice. Retrieved from <http://www.ada.gov/cguide.htm>

Section 508 Standards for Electronic and Information Technology. (n.d.) *U.S. Access Board*. Retrieved from <http://www.access-board.gov/guidelines-and-standards/communications-and-it/about-the-section-508-standards/section-508-standards>

KEY TERMS AND DEFINITIONS

Assistive Technology: Equipment, device, or other product that assists a disabled user in performing tasks that otherwise would be difficult or not possible to accomplish.

Color Deficiency: Color deficient vision results in an inability to distinguish certain colors and shades when compared to normal vision.

Digital Divide: The digital divide is the gap between those who have access to electronic and information technology and those who do not.

Electronic and Information Technology: Information technology and any equipment or interconnected system or subsystem of equipment used in the creation, conversion, or duplication of data or information (www.access-board.gov).

Mobile Device: Any handheld technology that allows the user to operate the device in transit. This includes such items as smartphones and tablets.

Motor Disabilities: Physical impairments that can impede movement, coordination, or sensation. They can include weakness and lack of muscle control.

Quick Response (QR): Matrix barcodes used by mobile device cameras to interpret localized information.

Screen Reader: Speech synthesis software used by a vision-impaired person to read aloud what is displayed on a computer screen.

Section 508: Amendment to the 1973 Rehabilitation Act requiring federal agencies to make electronic and information technology accessible to people with disabilities.

Web Accessibility: Web accessibility means that a person, regardless of disabilities, is able to use Web technology without encountering any barriers.

ENDNOTES

- ¹ Rehabilitation Act of 1973, *as amended by* § 508, 29 U.S.C. § 794(d) (1998).
- ² Architectural and Transportation Barriers Compliance Board is an independent federal agency, established by Section 502 of the 1973 Rehabilitation Act. Its primary function is to promote accessibility for individuals with disabilities.

The Adoption and Transformation of Capability Maturity Models in Government

Terry F. Buss

Carnegie Mellon University, Australia

INTRODUCTION

The Information Technology (IT) field has long been criticized as having problems with the quality of goods and services it produces. Gartner in its annual “Assessment of IT Practices” observes that about three-fourths of software projects “failed,” judging on cost, reliability, usability and timeliness.

Organizations, concerned about quality, stimulated the creation of models, frameworks or approaches that would “improve the quality” of the goods and services produced. Carnegie Mellon University’s (CMU) Capability Maturity Model (CMM), along with numerous subsequent competing and complimentary approaches, was launched to fill this need. CMM was developed for the US Department of Defense (DOD), making it a major quality improvement tool in government.

CMM would be eventually adopted by many governments around the globe in whole or in part. Analysts began applying CMM not only to IT operations and software development, but also in everything from risk through financial to innovation management. CMM soon led to a variety of *ad hoc*, derivative, hybrid, imitations or customizations by various governments, professional associations, vendors, researchers, and consulting firms, all based on capability maturity models.

The diffusion and modification of CMM seems similar to Xerox and photocopying. Xerox perfected photocopying technology and then dominated the market to such an extent that a Xerox became synonymous with a photocopy.

There are now hundreds of capacity maturity models around, but many, maybe most, are very different from CMU’s.

CMM has exploded in recent years as a quality improvement methodology. But there is much disagreement among practitioners, researchers, and theorists in the fields of IT, public management, and business over what works, what does not and why.

This entry looks at the variety of CMM government applications, asking: What is the basic methodology underlying CMM; how did other quality improvement initiatives generally and in government contribute to CMM’s development; what models currently compete in the CMM space; how are CMMs used in government; what problems, issues and controversies surround CMM as an approach and how can these be corrected; and what might constitute a future research agenda in the field?

BACKGROUND

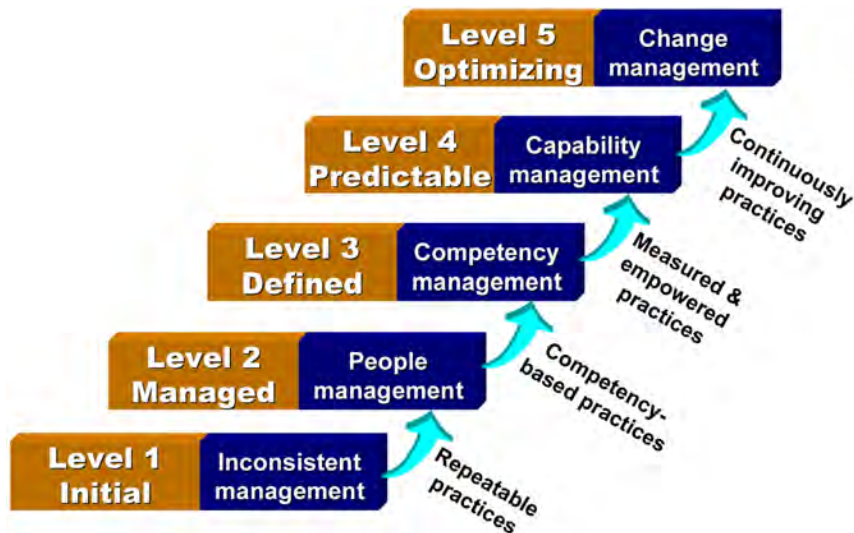
The CMM Methodology

Most, if not all, CMM frameworks have the same basic methodology. This is how the CMU’s Software Engineering Institute’s CMM model works in the context of human resource management (P-CMM) (Curtis, Hefley, Miller, 2009).

The model’s purpose is to “help organizations characterize the maturity of their workforce practices, establish a program of continuous workforce development, set priorities for improvement

Figure 1. The five maturity levels of the P-CMM¹

Source: Curtis, Hefley, Miller (2009).



actions, integrate workforce development with process improvement, and establish a culture of excellence”(p.8).

The model “consists of five maturity levels, or evolutionary stages, through which an organization’s workforce practices and processes evolve. At each maturity level, a new system of practices is added to those implemented at earlier levels. Each overlay of practices raises the level of sophistication through which the organization develops its workforce.”

“Each maturity level consists of three to seven process areas, identifying a cluster of related practices that, when performed collectively, achieve a set of goals considered important for enhancing workforce capability. Each process area organizes a set of interrelated practices in a critical area of workforce management, such as staffing, compensation, or workgroup development” (p.45). Process areas are linked to specific measurable goals to be achieved for each maturity level.

The model is grounded in standard process improvement practices common to most high performing, successful organizations. The model is best-practices based. These serve as a reference model against which an organization’s process

improvements are benchmarked. “Workforce practices are standard organizational processes that can be improved continuously through the same methods that have been used to improve other business processes.” This is accomplished by means of a “standard appraisal method for process improvement” (SCAMPI) (p.9). The model is evidence-based.

The model incorporates common best practices, such as:

... competency modeling, 360° performance reviews, Web-enabled learning, knowledge management, team building, cool space, participatory decision making, incentive-based pay, mentoring, meeting management, and empowered work (p.26).

The Quality Improvement Movement

The “quality improvement movement” began after World War II with the work of Edwards Deming (“statistical quality control”) and Joseph Juran (“managing quality”). They sought to improve quality in basic manufacturing as a way to increase customer satisfaction. Total Quality Management (TQM), inspired by Deming and Juran, came into

its own in 1985, arising out of initiatives by the US Navy. TQM eventually was adopted across the federal government in the 1990s.

Now TQM generally refers to a wide variety of quality improvement models, frameworks or methods that are evidence-based, relying heavily on data gathering, sampling and statistical analysis, on the one hand, and management principles and best practices on the other, all in the context of continuous improvement.

Formed in 1946-7, International Standards Organization (ISO) (www.iso.org) is a membership organization of 163 country government standardization boards in fields ranging from risk management through health to climate change. Standards are based on consensus and are voluntary for compliance. One standard on quality management, ISO 9000, is widely used.

Six Sigma originated from a quality improvement initiative at Motorola Corporation, then General Electric in 1986. Six Sigma is

... a disciplined, data-driven approach and methodology for eliminating defects (driving toward six standard deviations away from the mean and the nearest specification limit) in any process – from manufacturing to transactional and from product to service (www.isixsigma.com).

The Baldrige Award for Performance (<http://www.nist.gov/baldrige/>), beginning in 1988, gave the quality improvement movement a boost by annually recognizing high-performing organizations in the US.

Another movement is Lean Thinking or Manufacturing, popularized by James Womack and Daniel Jones in the 1990s. Their work on the success of Toyota Motor Company revealed how the company reduced waste in manufacturing. The US Environmental Protection Agency frequently uses Lean in its quality improvement efforts (Lean Government Initiative, 2014).

The quality movement intersected with a “sister” movement in the public sector, “Reinventing Government.” David Osborne, in the 1980s,

sought to improve efficiency and effectiveness in government calling for it to be run more like a business. This eventually morphed into “The New Public Management” (NPM) in the 1990s, where it remains the dominant theory of public administration globally.

CMU’s Capability Maturity Model Evolution

Early efforts at quality improvement proved effective in basic manufacturing, but as IT became more prevalent, they proved less effective or relevant in achieving quality: enter CMM.

CMM originated in the seminal work on “stages of growth” conceptualized by Richard Nolan (1973). Nolan’s work in turn inspired Phillip Crosby’s (1979) “quality management maturity grid,” a tool similar to CMM. Watts Humphrey, the father of CMM, crystalized his thinking on CMM while an executive at IBM, and then at CMU’s Software Engineering Institute (SEI) in the early 1980s (Humphrey, 1989).

Unlike other frameworks, CMM is premised on the notion that organizations must pass through sequential stages of maturity (rather than seeking “continuous improvement”) before becoming fully capable. CMM defined and codified what it means to be “capable” in these different stages of maturity. CMM also developed metrics to assess progress.

In 1984, the US Department of Defense (DOD), recognizing CMM’s utility as a framework for software engineering, established SEI as a Federally Funded Research and Development Center (FFRDC) at CMU (www.sei.cmu.edu). DOD intended that SEI “address pervasive and significant problems that impede the ability of organizations to acquire, build, and evolve software-intensive systems predictably on time, within expected cost, and with expected functionality.” US government agencies commonly require a favorable CMM appraisal as a requisite for awarding contracts.

In 1991, the first edition of the CMM guidance was released. CMM is “based on actual industry

practices, reflects the best practices of the industry and reflects the needs of individuals performing software process improvement and process appraisals (measurement and benchmarking)” (www.sei.cmu.edu).

CMM soon spawned a variety of frameworks not only by SEI, but also by numerous others, including software development, procurement, network development, security, human resource (people), portfolio management, project management, and just about every aspect of management. (This entry uses “CMM” to survey the field. CMM, which is a registered trademark of SEI-CMU, has become a generic label for most capability maturity models.)

In 2002, CMM was expanded into a suite of products, Capability Maturity Model Integration (CMMI), including: product and service development, service establishment and management, and product and service acquisition. Interestingly, CMMI no longer mentions “software,” reflecting the extent to which it has become viewed as a more general management tool.

In 2010, SEI spun off its CMM activities into a separate entity, the CMMI Institute (www.cmmi-institute.com). The Institute now offers training, professional certifications, organizational assessments, and professional publications. DOD, having accomplished its goals for CMM, still employs SEI as a FFRDC.

Other Capability Maturity Models

CMU’s CMM directly or indirectly stimulated other organizations to include CMM in their frameworks. Depending on one’s definition, there are hundreds of CMM models (see <http://ibpi.org/standard/isoiec-15504/> for a catalog of websites). These are the most widely used.

ISACA is an independent, nonprofit specializing in information systems governance, security, audit and assurance (www.isaca.org). Developed in 2000, its signature framework is CoBit (Control Objectives for Information and Related Technol-

ogy) which addresses “strategic level and focused on high-level IT management processes.” ISACA reports that CoBit is not related to CMMI.

In 2008, the International Standards Organization (ISO) and the International Electro Technical Commission (IEC) (<http://www.iec.ch/>) released ISO/IEC 15504, as a follow-on to its 1993 standard SPICE: Software Process Improvement and Capability Determination. 15504 is not a true CMM but does include “a continuous representation in which organizational maturity is implicit in the processes.” 15504 is, however, a “metamodel that harmonizes several best practice frameworks.”

The IT Capability Maturity Framework (IT-CMF), developed in 2004 by the Innovation Value Institute (<https://ivi.nuim.ie/it-cmf/>), reduces complexity by systematically addressing IT organizational capability gaps, thereby leading to optimized value and innovation.” IT-CMF has been applied to public hospital management.

AXELOS in 2008 developed the Portfolio, Program, and Project Management Maturity Model (P3M3) which has subsequently been “mandated by the Australian government for use in projects – breaks down into seven perspectives: organizational governance, management control, benefits management, risk management, resource management, financial management and stakeholder management.” The UK also applied this model in a variety of agencies including the National Health Service.

AXELOS (www.axelos.com/itil) hosts the Information Technology Infrastructure Library (ITIL) which “use(s) IT as a tool to facilitate business change, transformation and growth” in: Service Strategy, Service Design, Service Transition, Service Operation, Continual Service Improvement (AXELOS, 2013).

Recently, Six Sigma has added a Six Sigma Maturity Model (SSMM) to its portfolio, one that quantifies the evolutionary stages organizations pass through as they implement Six Sigma (<http://www.isixsigma.com/>).

CMM in Government

CMM government applications are legion, and growing. A google search on “CMM government” yields 541,000 hits. Here are some illustrative applications.

The Government IT Space

Acquisitions is a perpetual problem—especially for large systems—in every phase of procurement: developing specifications, bidding, selecting vendors, monitoring vendor performance, controlling, managing investments, and changing requirements. US government IT mandates an array of CMM quality control tools (GAO, 2004, 2015a).

Enterprise architecture (EA), implemented in the 1980s, has lagged behind initial quality improvement expectations. The US employs a variety of CMM frameworks to improve EA adoption and effectiveness (GAO, 2010).

Open data—releasing agency data to the public—has become a global movement across governments. The Open Data Center Alliance publishes CMM standards for those public agencies wanting to improve the quality of data released to the public (ODCA, 2013).

Cloud computing, a way to reduce costs, improve performance, and promote quality, is becoming widely used in government. CISCO Systems developed a Cloud Capability Maturity Model (CCMM) for use by UK local and national governments.

Big data, business intelligence, and analytics offer government powerful new tools to gather, store, analyze and utilize the massive amount of data gathered. The Innovation Value Institute (2013) offers CMMs dedicated to information management; while Accenture offers governments CMM in business intelligence and analytics at the Centers for Disease Control, Federal Drug Administration, and Centers for Medicare and Medicaid Management.

The Green Grid, a consortium of industry and government, offers a Data Center Maturity Model

to help make data centers more efficient (www.thegreengrid.org).

Cyber security is perhaps the hottest issue facing government IT. In 2012, the Center for Strategic and International Studies recommended that US adopt a cybersecurity CMM government-wide (Reeder, et al., 2012). In 2013, the Rand Corporation conducted a cyber defense CMM analysis of the European Defense Agency.

E-government has transformed from a system exclusively intended to improve transactions to one where citizens interact with government through collaborative technologies and social media (Ibribarren, 2013). E-government CMM models are widely applied in government, Gartner’s Government Maturity Model is a popular one.

With the explosion of mobile technology in government, IBM developed the Enterprise Mobile Progressive Maturity Model to move away from “old fashioned” IT strategies to digital strategies.

The State of Oregon, adopted Gartner’s Maturity Model for Web Content, to help cities create state-of-the-art *websites*.

Tying these IT advances together is Forrester Consulting’s Technology Management Leadership Maturity Model. While KPMG employs an Information Technology Asset Management (ITAM) Capability Maturity Models to help federal agencies manage their technology assets.

The Government Non-IT Space

Non-IT CMMs are common in government in a variety of areas (NAO, 2010). Human capital management for IT workforces rely heavily on CMU’s People-CMM (Curtis, Hefley, Miller, 2009). The Balanced Scorecard methodology widely used in US government has been converted into a CMM and applied to strategic and performance management (BSI, 2015). Risk management is perhaps the most common of non-IT CMM applications used by government particularly in the finance area (Cienfuegos, 2013). New Zealand’s Office of Auditor General uses CMM in its process im-

provement program (OAG, 2012). The Australian Public Service Commission, in 2009, launched a National Innovation System—*Empowering Change*—mandating CMM models to encourage implementation and development. And the Malaysian Public Works Department employs its own CMM.

Some applications are rather surprising. The Canadian Human Rights Commission (<http://www.hrmm-mmdp.ca/eng/about-model>) developed The Human Rights Maturity Model to help employers create “a self-sustaining human rights culture in the workplace.” Lancaster University, UK, developed a Digital Forensics Comprehensive CMM (DF-C2M2) to help investigators raise the competency of their technical workforce.

City planning agencies often use CMMs. Digital Dublin (Ireland) developed the Digital City Maturity Model. The Scottish City Alliance utilizes a Smart City Maturity Model. The Transportation Research Board at the National Academies of Science makes available its Institutional CMM to US cities attempting to improve transportation operations. IBM markets its Intelligent Transportation Maturity Model to cities. And Deloitte Consultants conducted a maturity analysis, *Closing the Infrastructure Gap: The Role of Public Private Partnerships*, of governments around the world.

ARTICLE FOCUS

Major Issues, Controversies, Problems

The issues below apply more or less to all CMM models. There are hundreds of references assessing CMM (e.g., Morgan, 2007; Greiner, 2007; Betz, 2011; Van Wangenheim, 2010; Cienfuegos, 2013; McSweeney, 2013). Below is a distillation of the controversies and problems identified.

CMM Not Delivering on Promises

With IT and public management and policy failures—many guided by CMM approaches continuing to plague government—critics wonder whether these schemes are worth it (Ashkenas, 2013). The Government Accountability Office (GAO) maintains a “High Risk List” representing dysfunctional programs that threaten US government operations. Perpetually on that list are “government-wide IT” (GAO, 2015a), “Defense Department programs including acquisitions and operations,” and “strategic human capital management” (GAO, 2015b). Just for IT alone, the US government spends \$80 billion annually and still cannot get acquisitions and investment management right, even with the widespread use of CMM and other CIO tools. The US is not alone; most governments have similar problems. Australia recently reorganized its IT agency into two separate departments because of inefficiency and ineffectiveness, even though it employs the Portfolio, Program and Project Management Maturity Model (P3M3) to assure quality.

CMM Frameworks Lack Theoretical Underpinnings

CMM is grounded in expert opinion, peer review, best practices, benchmarking, learning trial and error, knowledge management, and performance assessments. It is not, however, based on scientifically validated theories of management. Proponents of CMM do not generally think this is an issue because the methods are self-correcting over time through careful assessment. Critics believe that CMM’s failings, when they occur, may be the result of a lack of theoretical rigor: CMM cannot demonstrate whether a particular improvement was the result of one of its principles or something else.

All CMM Frameworks Are Not Equally Credible

SEI-CMU and other products like ISO and Six Sigma are highly credible because they have rigorous methodologies, have been vetted by experts and have undergone extensive testing. Many other frameworks are lacking, but may appear to be credible because they exploit, whether intentionally or unintentionally, established CMM brands. This is a problem in government applications in that policy makers tend to borrow a lot from one another. Canada, UK, NZ and Australia all share frameworks, the “Capability Review,” for example. The Capability Review appears on the surface to be like CMM, but has nowhere near the rigor. The Australians, realizing this, have a “whole-of-government” approach that combines Capability Review and Capability Maturity (APSC, 2012).

CMM Assumes that Methods Apply to All Organizations Regardless of Function, Structure, Process and Context

CMM may not be universally applicable across organizations. Government organizations range wildly in terms of size, workforce unionization and characteristics, IT adoption, function, regulation, and so on. Is it possible that all government organizations would find the same CMM to be appropriate? On the surface, it seems not. Some government organizations are unconcerned with efficiency, effectiveness or quality. They focus, instead, on policy and political outcomes.

CMM Does Not Describe or Advocate Processes, It Offers Principles Good Processes Must Satisfy

CMM is a reference model against which organizations assess the capabilities of their processes. CMM does not advocate for specific processes,

neither does it describe them. This causes confusion in that organizations can know where they stand on the maturity level scale and where they need to be at full maturity, but may not know how to get there given the processes they have. This is problematic if processes do not support maturity levels: strategic alignment of missions, goals and strategies can easily come out of alignment when applying CMM.

Organizations Are Dynamic, While CMM Is Static

Organizations, even those in government, can be so dynamic that CMM may be unable to improve quality. CMM requires a stable, predictable, static organizational environment that can be fully (and perhaps laboriously documented). Take the US Department of Homeland Security (DHS), for example. In response to the terrorist attacks of September 11, 2001, the George W. Bush Administration created DHS by merging 22 disparate agencies, employing some 220,000 people. Each agency dragged along with it all of their legacy systems, nearly all of which lacked interoperability. Many were guided by CMM. Now in its fourteenth year, the reorganization continues and the system remains problematic. Proponents suggest that no quality improvement system could have addressed the DHS creation and subsequent reorganizations.

Further, Governments Do Not Have Years to Achieve Maturity Levels

Governments do not have the luxury of long time horizons in which to achieve some optimum level of capability maturity. CMM tends to be adopted in the public sector just after administrations change hands. Newly-elected or constituted governments typically throw out everything their predecessors have done and pursue their own initiatives.

CMM May Stifle Innovation, Rather Than Facilitate It

Critics raise the question, can organizations managed under CMM be truly innovative? Likely not. Phillips (2011) in *Relentless Innovation* convincingly argues that these systems are premised on the need for stability, predictability and risk avoidance. Hence, they have difficulty innovating. Time will tell whether innovation maturity models are able to overcome organizational barriers to innovation (Esterhuizen, Schutte, & du Toit, 2011).

CMM Maturity Levels Seem Subjective or Arbitrary

CMM separates itself from other frameworks by imposing change and improvements through maturity stages. Maturity stages seem somewhat arbitrary with respect to what is included in each level. It seems possible to swap items across the stages. Some suggest that stages are merely a rank ordering. It seems possible to skip through earlier stages fairly quickly. It's possible simply to "cherry pick" items from each stage and not go through the process. Advocates are adamant that the stages are meaningful and must be judiciously pursued.

In the End, CMM Is About Compliance

There is debate about what constitutes a best practice under CMM. CMM typically lists hundreds of practices and measurements that must be followed to attain maturity levels. Critics point out that these are really an exercise in compliance or "box ticking."

The Value Add of CMM to the Bottom Line Is Unclear

Critics ask whether the cost of undertaking CMM is worth it. The cost of conducting a CMM assessment is high for many organizations, especially small ones. But the cost of reengineering

an organization over a period of several years as it passes through various levels of maturity may be prohibitive. But is the cost worth the benefit? In many cases it may be impossible to determine whether a reengineered organization can even demonstrate that CMM made any contribution at all.

The Plethora of CMM Models Leads to Inefficiency and Confusion

The inventory of CMM frameworks is large. Some frameworks supplement or compliment others, but there are gaps. Audiences intended as beneficiaries are quite disparate. Organizations with many disparate needs may find that implementing different models can be problematic, not to mention inefficient. For example, ISO/IEC 15504 processes have no counterpart in CMML. Further, CMML focuses on maturity levels, while 15504 concerns "a continuous representation" of "implied" maturity stages.

SOLUTIONS AND RECOMMENDATIONS

There are a myriad of studies addressing issues above. But, like CMM itself, they are scattered all over the map, many not available on the Internet or even publicly available. What is required is a systematic program to address these issues.

Create an Organization to Catalog CMM

There are a few heroic ongoing efforts to catalog CMM applications, studies, and methods. The International Best Practice Institute, for example, provides a spreadsheet of websites for various models (<http://ibpi.org/about/frameworks/>). But there needs to be greater, more comprehensive effort in collecting a larger number of items for analysis.

In addition to collecting materials, there should be an effort to crosswalk or “map” competing models and frameworks to derive a picture of what is covered and what is omitted in each (e.g., Wendler, 2012). ISO/IEC 15504 does this to some extent. But much of the published work consists of comparisons of two or three methods (e.g., Doyle, 2009), not the entire collection of approaches.

Vet CMM Products

Catalogues and comparisons are a start, but these do little to establish credibility of the approaches. Just as *Consumer Reports* tests consumer products, there ought to be an organization to assess, vet and certify CMMs. The EU’s Open Source Observatory offers a possible model (<https://joinup.ec.europa.eu/community/osor/description>). European governments deposit open source software applications in a platform that allows participants to share software, content, studies, best practices and opinions.

Fund Scholars to Scientifically Study CMM’s Theoretical and Empirical Underpinnings

If there were not so many problems in managing IT organizations and others, it wouldn’t make much difference if CMM were grounded on expertise, benchmarking and assessments. But since there are problems, it might make sense to examine the underpinnings more systematically. Some believe that such an assessment might find hidden problems in CMM methods. Just as DOD originally set up SEI to develop CMM, a government organization (like the National Science Foundation, National Academy of Sciences, or the National Institute of Standards and Technology) or private foundation might facilitate scientific testing.

FUTURE RESEARCH DIRECTIONS

CMM represents a fruitful area of research not only for IT but also for public management. At present, CMM assessments and management studies are not scientific. Putting the two together affords an opportunity for both: CMM clearly spells out in great detail how to improve management and offers an empirical test bed for study; management studies have a lot of theory floating around, but limited opportunities for empirically testing: a perfect match.

CONCLUSION

Proliferation of quality improvement models and the problems associated with them should not lead one to conclude that CMM should be abandoned. What this means is that the field is maturing: there is more and more interest among researchers and practitioners. Competition is good. Furthermore, earlier efforts were effective, but perhaps limited. Now, CMM is being applied to nearly everything. CMM offers the best hope of advancing the quality improvement field. Now is the time to get serious about more carefully studying CMM as a discipline. Inevitably CMM will be replaced as the “science” of management advances, but that comes later.

REFERENCES

Ashkenas, R. (2013). Why continuous improvement may need to be discontinued. *Forbes*, 24(July). Retrieved from <http://www.forbes.com/sites/ronashkenas/2013/07/24/why-continuous-improvement-may-need-to-be-discontinued/>

Australia Public Service Commission (APSC). (2012). Organizational capability. In *State of the Service Report, 2012*. Canberra: APSC. Retrieved March 2015 http://www.apsc.gov.au/__data/assets/pdf_file/0019/29233/SOSR-2012-13-chapter-10.pdf

AXELOS. (2013). *ITIL® Maturity Model 2013*. AXELOS. Retrieved March 2015 https://www.axelos.com/Corporate/media/Files/Misc%20Qualification%20Docs/ITIL_Maturity_Model_v1_2W.pdf

Balanced Scorecard Institute (BSI). (2015). *Strategic Management Maturity Model*. Cary, NC: BSI. Retrieved March 2015 <http://www.balancedscorecard.org/portals/0/pdf/bscistrategicmanagementmaturitymodel.pdf>

Betz, C. (2011). ITIL®, COBIT®, and CMMI®: ongoing confusion of process and function. *BPTrends*. Retrieved March 2015 <http://www.bptrends.com/publicationfiles/10-04-2011-ART-Ongoing%20Confusion%20of%20Process%20and%20Function-Betz-Final.pdf>

Cienfuegos, I. (2013). *Developing a risk management maturity model*. University of Twente, Department of Public Administration. Retrieved from <http://www.utwente.nl/bms/pa/uploads/Cienfuegos%20-%20Developing%20a%20Risk%20Maturity%20Model.pdf>

Controller-Auditor General (OAG). (2012). *Reviewing financial management in central government*. Wellington, New Zealand: Government of New Zealand. Retrieved March 2015 <http://www.oag.govt.nz/2012/financial-management/docs/financial-management.pdf>

Crosby, P. B. (1979). *Quality is free*. New York: New American Library.

Curtis, B., Hefley, B., & Miller, S. (2009). *People Capability Maturity Model (P-CMM), Version 2.0*. Pittsburgh, PA: Software Engineering Institute, Carnegie Mellon University. Retrieved from <http://www.sei.cmu.edu/reports/09tr003.pdf>

Doyle, K. (2009). *CMMI, ITIL & ISO 20000: A mutually supportive relationship (power point)*. Pittsburgh, PA: SEI. Retrieved March 2015 http://resources.sei.cmu.edu/asset_files/Presentation/2009_017_001_22742.pdf

Esterhuizen, D., Schutte, C., & du Toit, A. (2011). Enhancing innovation capability maturity through knowledge conversion. *Acta Commerciali*, 231-239. Retrieved March 2015 <http://www.actacommercii.co.za/index.php/acta/article/viewFile/162/162>

Government Accountability Organization (GAO). (2004). Information technology investment management: A framework for assessment and improving process maturity. Washington, DC: GAO, GAO-04-394G.

Government Accountability Organization (GAO). (2010). A framework for assessing and improving enterprise architecture management. Washington, DC: GAO, GAO-10-846G.

Government Accountability Organization (GAO). (2015a). *Improving the management of IT acquisitions and operations*. Washington, DC: GAO, High Risk List.

Government Accountability Organization (GAO). (2015b). *Strategic human capital management*. Washington, DC: GAO, High Risk List.

Greiner, L. (2007). Capability Maturity Model Integration (CMMI) definition and solutions. *CIO Magazine*. Retrieved March 2015 <http://www.cio.com/article/2437864/process-improvement/capability-maturity-model-integration--cmmi--definition-and-solutions.html>

Humphrey, W. (1989). *Managing the software process*. Reading, MA: Addison-Wesley.

Innovation Value Institute (IVI). (2013). *Information management and big data*. IVI website. Retrieved March 2015 <http://ivi.nuim.ie/sites/ivi.nuim.ie/files/publications/IVI%20Exec%20Briefing%20-%20Big%20Data%20%2B%20EIM%20v1.0.pdf>

- International Standards Organization (ISO). (2015). *ISO Management System Standards*. Website. Retrieved March 2015 <http://www.iso.org/iso/home/standards/management-standards/mss-list.htm>
- Iribarren, M. (2013). Capability maturity framework for e-government: a multi-dimensional model and assessing tool. Washington, DC: Inter-American Development Bank. Retrieved from <http://web.stanford.edu/~gvaldesu/articles/CapabilityMaturityFrameworkForEgov.pdf>
- Kneuper, R. (2009). *CMMI: improving software development and engineering processes*. New York: Rocky Nook.
- Lean Government Initiative. (2014). *Lean government methods guide*. Washington, DC: Environmental Protection Agency. Retrieved March 2015 <http://www.epa.gov/lean/government/pdf/lean-methods-guide.pdf>
- McSweeney, A. (2013). *Review of data management maturity models*. SlideShare. Retrieved March 2015 at <http://www.slideshare.net/alanmc-sweeney/review-of-data-management-maturity-models>
- Mitchell, S., & Switzer, C. (2012). *GRC capability model, version 2.1*. Open Compliance & Ethics Group. Retrieved March 2015 <http://www.oceg.org/resources/grc-capability-model-red-book/>
- Morgan, P. (2007). Process improvement: Is it a lottery? *Mind and Tools*, 15(1), 4-12. Retrieved March 2015 <http://www.methodsandtools.com/PDF/mt200701.pdf>
- National Audit Office (NAO). (2010). *Maturity of process management in central government*. London: UK Government, NAO. Retrieved March 2015 http://www.nao.org.uk/wp-content/uploads/2013/03/Process_Management.pdf
- Nolan, R. L. (1973). Managing the computer resource: A stage hypothesis. *Communications of the ACM*, 16(7), 399-405. doi:10.1145/362280.362284
- Open Data Center Alliance (ODCA). (2013). *Open data center alliance usage model: cloud maturity model rev. 2.0*. Retrieved March 2015 http://www.opendatacenteralliance.org/docs/Cloud_Maturity_Model_Rev_2.0.pdf
- Phillips, J. (2011). *Relentless innovation*. New York: McGraw-Hill.
- Reeder, F., Chenok, D., Evans, K., Lewis, J., & Paller, A. (2012). Updating US federal cybersecurity policy and guidance. Washington, DC: Center for Strategic & International Studies. Retrieved from http://csis.org/files/publication/121019_Reeder_A130_Web.pdf
- Von Wangenheim, C. (2010). Systematic literature review of Software Process Capability/Maturity Models. *Proceedings of International Conference on Software Process Improvement and Capability Determination (SPICE)*. Retrieved March 2015 http://www.inf.ufsc.br/~gresse/download/SPICE2010_Systematic_Literature_vf.pdf
- Wendler, R. (2012). Maturity of maturity model research: A systematic mapping study. *Information and Software Technology*, 54(12), 1317-1339. doi:10.1016/j.infsof.2012.07.007

ADDITIONAL READING

- Asia and Pacific Center. (2001). *E-government Capability Maturity Models*. New York: UN. Retrieved July 2015 file:///C:/Users/Pr0/Desktop/CMM%20Model%20%20FINAL.PDF
- Department of Energy. (2014). *Cybersecurity CMM*. Washington, DC: DOE. Retrieved July 2015 http://energy.gov/sites/prod/files/2014/03/f13/C2M2-v1-1_cor.pdf
- Meyer, J. (2014). *Institutional maturity models*. Washington, DC: Millennium Challenge Corporation; <http://www.worldbank.org/content/dam/Worldbank/Event/Meyer.pdf>, Retrieved July 2015.

Paulk, M. (2009). *A history of the Capability Maturity Model for Software*. <http://www.asq.org/software> Retrieved March 2015 <http://citeseerx.ist.psu.edu/viewdoc/download?rep=rep1&type=pdf&doi=10.1.1.216.199>

UN Development Program. (2015). CMM for program and projects. New York UNDP. Retrieved July 2015 http://ppmtoolkit.undp.org/4b_Maturity_Concept.cfm

University of Canberra. (2011). A critical assessment of P3M3 in Australia. Canberra: University of Canberra. Retrieved July 2015 http://www.governanceinstitute.edu.au/magma/media/upload/media/529_P3M3-Anzsig-Insight.pdf

URISA. (2013). GIS Capability Maturity Model. London: Urban and Regional Information System Association. Retrieved July 2015 [http://www.urisa.org/clientuploads/directory/GMI/GISCMM-Final201309\(Endorsed%20for%20Publication\).pdf](http://www.urisa.org/clientuploads/directory/GMI/GISCMM-Final201309(Endorsed%20for%20Publication).pdf)

KEY TERMS AND DEFINITIONS

Capability Maturity Model (CMM): A model that specifies various evolutionary or ma-

turity stages an organization must pass through in order to become increasingly capable in achieving efficiency, effectiveness and economy in its operations.

Maturity Stages: Clearly delineated stages that represent increasingly greater levels of competency or capability.

Model: An abstraction or representation which captures the essence of an organization, process, or operation to improve performance.

Practice: An activity, performance or exercise intended to become more proficient.

Process Improvement: Improving operations to increase efficiency, effectiveness and economy.

Reference Model: A model based on best practices and research that serves a comparator against which to evaluate a model under assessment.

ENDNOTE

- ¹ Content has been provided from Technical Report CMU/SEI-2009-TR-003, "People Capability Maturity Model (P-CMM®), Version 2.0, Second Edition," Copyright 2009 by Carnegie Mellon University, with permission from its CMMI Institute.

Bridging Between Cyber Politics and Collective Dynamics of Social Movement

Kazuhiko Shibuya
ROIS, Japan

INTRODUCTION

This era of social media extends the bounds of traditional politics to ‘Cyber-Politics’ (Hill & Hughes, 1998; Jordan, 2001; Choucri, 2012). The advent of social media summons *the collective dynamics of democracy of the citizens, by the citizens, and for the citizens*. Such patterns using social media can readily alter the form of social movements, allowing their mutual interconnection and shaping the enclaves of networked clustering. Social media offer a new paradigm of democracy that encourages engagement and participation in both cyber and actual political actions for ordinary citizens.

Nevertheless, little is known about co-occurrence and linkages between cyber and real world actions by numerous participants. These are not limited to voluntary support for elections, donations, public discussion, and other collaborations, whether online or not. Especially, many observers have noted that ordinary citizens are apt to cope with social media as a bridge linking virtual and actual political activities.

Consequently, this issue should be investigated with open questions related to the following points.

1. **Social and Legal Background:** Social institutional matters related to legitimation crises caused by social movements
2. **Social Media and Its Relative Background:** Co-occurrence and Linkages of collective dynamics between cyber and actual political actions
3. **System Thinking and Simulation:** Enlargement of participants in social movement
4. **Beyond Borders:** Systemic Risks from Local to International Affairs

BACKGROUND

Certainly many reports have described that social media easily raise social movements or anti-autocratic revolutions by ordinary citizens. Those media have eagerly inspired ordinary citizens to participate in vigorous discussions as well as actual social movements (Casilli & Tubaro, 2012; Choudhary et al., 2012). It seems readily apparent that participants indeed have shared further motivations and common goals.

In Table 1, although the total population of participants could not be estimated accurately in each case, the entries exemplify serial movements: ‘Arab Spring’ (e.g. Jasmine and Egyptian revolutions), ‘Occupy Wall Street’ movement, the umbrella revolution in Hong Kong, and other recent events (e.g., a series of protests against the Charlie Hebdo shooting at France, and a series of protests against policy-making on the national security related bills in Japan and other nations). Those who participated expressed their desires for democratic institutions, freedom of public opinion, socioeconomic chances, and opposition against controversial political issues. They also made other demonstrative appeals.

Table 1. A part of recent movements

Events	Date	Nations	Estimated Max of Participants	Main Intention and Goal of Participants	Final Results
Jasmine Revolution	Dec. 17, 2010-Jan. 14, 2011	Tunisia	100,000?	Political Claims?	Revolution
Egyptian Revolution	Jan. 25, 2011-Feb.11, 2011	Egypt	2,000,000? (Tahrir square, Cairo)	Political Claims?	Revolution
'Occupy Wall Street'	Sep. 17, 2011	U.S.A.	15,000? (Foley Square, New York)	Socioeconomic claims and occupying area	Failure and Dissolution
The Umbrella Revolution	Sep.26,-Dec. 15, 2014	Hong Kong (China)	100,000? (Nearby the Admiralty)	Political Claims on Freedom for democracy and occupying area	Failure and Dissolution
Protests against the Charlie Hebdo shooting	Jan. 7, 2015	France	35,000? (Paris)	Protests for Public Opinion	Temporal Mass Meeting
Greece Economic Crisis 2015	Jun. -Aug. 2015	Greece	30,000? (Opposition Side, Atene)	Socioeconomic reasons	Temporal Mass Meeting
Opposition against the policy making on the national security related bills	Aug. 30, 2015	Japan	120,000? (Around the National Diet, Tokyo)	Political Claims and oppositions against controversial issues	Temporal Mass Meeting and oppositions
Catalan Nationalism for Independence	Sep. 11, 2015	Spain	1,000,000? (Català)	Political Claims for independence	Ongoing matter

Especially, the outcomes of the 'Arab Spring' became the trigger of similar actions. Furthermore, in Arab cases, each political regime was overturned by numerous participating citizens using social media. Subsequently, the results spurred regional and global strain (Boening, 2014; Sadiki, 2015). Democratic progress in Tunisia was awarded a Nobel Prize in Peace at 2015, but their activities require more efforts.

Sociologically, those phenomena might be categorized as social movements and expressions of the collective dynamics of behavior by numerous citizens (Le Bonn, 1931a, 1931b; Smelser, 1962). The movements realized social influence and catastrophic collapse of regimes caused by the crowds. Here, the categories of those actions by citizens include opinion expressions, group processes, identification, emotional attitudes, fads, mobs, criminal riots, demonstrations, and politically assertive agitations as well as social change against ancient regimes caused by large groups of citizens in uprisings and revolution (Olson, 1971; Brown, 2000; Rohlinger & Snow, 2003).

Legitimation Crisis Caused by Social Movement

A question arises: as exemplified by the case of 'Arab Spring', did those movements and revolutionary outcomes caused by mass popular movements have any legal background to subvert their government?

Principally, liberal democracy requires the freedom of expression for ordinary citizens and their deliberations in community (Mill, 2012). However, democracy has its own specific standards and procedures to deal with somewhat controversial policies (Johnson, 2008). Consequently, social scientists intensively regard all forms of collective action as problems of coordination and synchronization between stakeholders and actors of various types (Dogan, 1992; Klandermans & Roggeband, 2010). Democratic procedures can only legitimize fundamentals on ruling governance and sovereignty.

Certainly some theorists assert that people always retain the right to resist illegal policies

by the government (Rousseau, 2009). But such illegal policies signal a forfeiture of legitimacy in democracy. It is called a *legitimation crisis* (Habermas, 1976), and it denotes a breach of trust in national governance, institutions, and leaders. It sometimes causes danger of disrupting national governance.

Consequently, *legitimacy* is indeed an important fundamental in democratic society. Now again, the democracy and its legitimation fundamentals should be reconsidered. As its lexical meaning, legitimacy is ‘*in accordance with rules*’ in politics (Oxford English Dictionary). It is the nexus of political power, institutions, trust, and legal procedures including social justice and equality (Rawls, 1971; Nozick, 1974; Jost & Major, 2001; Fukuyama, 2014).

To date, and traditionally, legitimacy has been explored in many theoretical discussions. Originally, Aristotle, an extremely famous and historical ancient Greek philosopher, discussed theories of legitimacy based on the politics and ethics (e.g. *Politics*, *Nicomachean Ethics*). Today’s political democracy is strongly founded in these backgrounds and traditions.

As renowned philosophers, Rousseau (2009) critiqued the foundation of social contracts and constitution for the nation and citizens. And Locke (1988) argued the concept of liberal independence from natural law and slavery conditions in his famous representative works. He differentiated humans as three types: a nation, a citizen, and collective people. Their discussions clearly addressed citizens’ rights and obligations for governing national power.

However, the sociologist Weber (1958) categorized legitimacy as traditional, charismatic, or legal-rational. Then he broadened it for discussions of economic and politically legitimized powers, coercions, and dominance in the social relations and structure. Social power determines various social relationships, procedures, institutions, and the distribution of social capital (Lipset, 1959).

In addition, it is necessary to remain mindful of the acceptance of political regime by citizens (It equals with the concept on sovereign in democratic society). According to Anheier & Toepler (2010), “*the legitimacy was dealing with the relationship between the nature of acceptance of political regimes and their capacity to exercise power and authority*”. That is to say, it shall be consensually accepted by citizens, and it is grounded on democratic civil society. Democratic regimes are often approved by majority rule and consensual deliberations. It is necessary to assess legitimacy itself as well as democratic procedures (Ariely, 2015).

Regarding these concerns, Rousseau (2009) emphasized the general will of all citizens and the geometric balance between the governor, the sovereign, and citizens. He asserted that saving the ideal democracy requires that a tyrant be purged from public governance. Within a term of office, a governor is delegated to govern the nations by citizens as the general will in democratic society.

Generally, social movements and opinion assertions by collective actions of citizens against their governments in accord with rules are not considered to be beyond the grounds of democratic action. However, it seems that the final result of ‘uprising’ and ‘revolution’ caused by them could not be evaluated by the extrajudicial standard even though each has the right of resistance against tyranny of their government. When citizens face such conditions, social contracts as well as legitimacy in democratic meanings are often reconsidered.

Finally, to summarize, the concept of legitimacy in democracy consists of at least three factors. When these factors did not meet each standard, it might cause a legitimation crisis.

1. Legal-Rationality
2. Accordance with rules and social contract between the governor and a sovereign
3. Acceptance of political regimes by consensual sovereignty

Co-Occurrence and Linkages of Collective Dynamics Between Cyber and Democratic Movements

An important matter is to unveil the co-occurrence and linkages of the dynamics of mass participants in both cyber and political actions. That is the problem of intention, timing, and relations within social media usage and participation. *When and where do participatory citizens submit their opinions and collaborate with others using social media?* Differences between social media usages before participation, after participation or without social media in movements, and/or real-time social media usage in participation should all be noted.

These are certainly open questions posed for *computational social science* (Lazer et al., 2009). For example, they require understanding of social media usage patterns, structural organization patterns in human relations among leaders and followers, wireless clustered networking, geospatial mobility of collective participants using smartphones, population estimation and enlargement of participants in movements, and linkages between the results of public opinion and voting behavior.

What it observes online and offline were underlying not only in sentimental but somewhat organized responses by numerous citizens against governmental policies during the 'Arab Spring' (Choudhary et al., 2012). Alternatively, in the case of Hong Kong, participatory citizens had adopted wireless technologies (e.g. 'fire-chat' as a wireless communicable applications). Using those applications, they were able to construct ad hoc networking dynamically for social movements.

Then, researchers must consider the unconventional norms of cyber politics and demonstrations stemming from the influence of social media (Malinick et al., 2013; Agarwal et al., 2013; Rhue & Sundararajan, 2014; Bello & Rolfe, 2014; Golbeck & Hansen, 2014; Gonzalez-Bailon & Wang, 2016). Furthermore, as described above, it is now the era of social media and a ubiquitous computing society (Shibuya, 2004). Ordinary citizens can mutually collaborate easily using smart phones and mobile

devices. Attention should be devoted to their networked dynamics mechanisms. Consequently, researchers should unveil collective dynamics of cyber networking and actual behaviors in social movements for additional understanding.

Next, what are factors to engage citizens in both cyber and political actions? Even if citizens are able to use social media and ubiquitous services, they usually feel some trouble about participation. Namely it is too natural to think that vigorous discussion online has no direct reasons to activate and invoke citizens for actual efforts and revolutionary uprisings.

However, their activities often can easily produce bridges between cyber and offline differences (Bond et al., 2012). Some reasons why many ordinary citizens participate in those movements are considered to be social influence, opinion-group polarization and self-organization of networking such as opinion leaderships opinion-group polarization, and self-organization of networking such as opinion leaderships (Katz & Lazarsfeld, 1955; Huckfeldt et al., 2014), identification of group beliefs, peer-to-peer interactions, like-minded group formation and amplifications of motivations among members (Sunstein, 2001). Past reports have described that their mentalities are ruled under group minds (Le Bonn, 1931a, 1931b; Smelser, 1962; Rohlinger & Snow, 2003; Hedstrom & Bearman, 2009).

Many researchers have articulated that, using social media and internet communications, citizens have obtained tools for globally collaborating with like-minded partners at will (Aral et al., 2009). They can proactively choose information, web links, and communication partners. Computational algorithms can also be designed to offer personalized materials for each person. These systems learn from their web queries and records through collaborative filtering and machine learning. Moreover, each citizen can see and know only a part of the worlds. It is just stereotyped information contact for the actualities. As result, opinion-group polarization, group stereotyping, ideological expressions, are spreading information

of hateful opinions are now frequently observable (Sunstein, 2001; Bakshy et al., 2015).

Enlargement of Participants in Social Movement

Physicist Newton said “*I can calculate the motion of heavenly bodies but not the madness of people*”. As social constraints, it remains not to examine empirically dynamics on expansion of massive participants in social movement. But both empirical and simulation studies have been reported recently (Downs, 1957; Shibuya, 2004; Hedstrom&Bearman, 2009; Braha, 2012; Epstein, 2014). And some cases were identified as decentralized networking patterns without core agitating leaders (Castells, 2012; Malinick et al., 2013). According to interview studies conducted in the aftermath of the ‘Occupy Wall Street’ movement (Manilov, 2013), one participant reminisced “*What started as a couple of hundred people in a park with no plan has turned into a decentralized network of activities*”. Furthermore, it should

be exactly distinguished between onlookers and hard-core demonstrators whether social media usage or not (Bryan et al., 2014).

As one of the latter simulation studies, for example, the author presents an example of system modeling and its simulations on those movements and the legitimization crisis (Figure 1; Shibuya, 2016). Various elements and interactions as a system encompass the complexity of crises of legitimacy. Furthermore, it depicts hypothesized variables and cybernetic systems such as types of participants in movements (e.g., ordinary citizens, claimants, and hard core protesters), increasing total populations of participants, forming enclaves of participants, reinforcing a group entities and movements, and amplifying systemic risk of legitimization crises and disruptions of governmental systems. These factors can be reinforced recursively by other factors such as mobility dynamics of human behavior in spatial geometry (Rossi et al., 2015; Blondel et al., 2015), social media usage, and percolation criteria in networked clustering among participatory citizens.

Figure 1. System modeling on legitimization crisis caused by a series of social movements

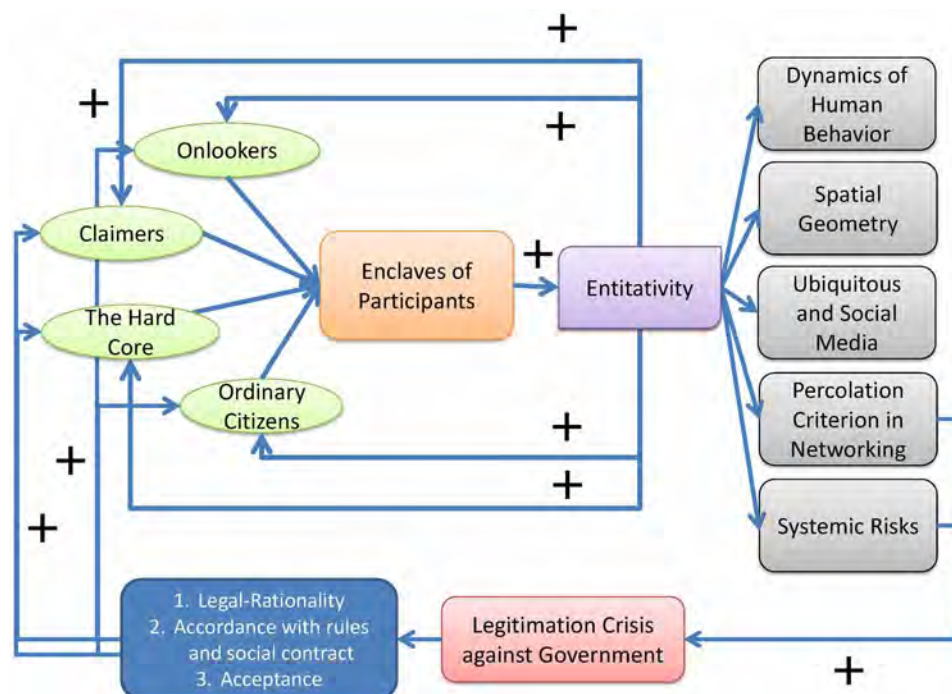


Figure 1 depicts the progress of expansion by a self-reinforcing feedback loop in system modeling on a legitimization crisis caused by a series of social movements. Furthermore, it is similar to social phenomena such as *cyber-cascading* in politics (Sunstein, 2001), the spiral of silence (Noelle-Neumann, 1984), the band-wagon effect, political dynamics and networked clusters in social sciences.

As Sunstein has noted, these phenomena are caused by a ‘critical mass’ of people as participating citizens. Consequently, the need exists for quantitative means to estimate probabilistically the threshold of such a critical mass in advance. However, it is possible to estimate the criterion. Standing on percolation theory, a threshold of critical mass is strictly defined in the percolation principle (O’Sullivan&Perry, 2013). Furthermore, a percolation model is applicable to some of those social phenomena (Broadbent&Hammersley, 1957; Hoshen&Kopelman, 1976; Sahimi, 1994; Newman&Ziff, 2001; Newman&Park, 2003; Malarz&Galam, 2005). Sociologically, the critical mass of social movements is definable as a threshold model (MacCoun, 2012). Especially, empirical works have examined renowned theories and serial examinations proposed by Granovetter (1978). Recently, sociologically advanced investigations of collective dynamics in regular and random networks have been reexamined by cascading and percolation standpoints (Watts, 2002). He specifically examined random finite networks. Then he devoted some attention to systemic risk and cascading influences on fragile networking structures, designating it as a ‘Percolating vulnerable cluster’.

Systemic Risks from Local to International Affairs

In the ‘Arab Spring’, the results of citizens’ actions crossed unintentionally beyond international and geopolitical borders. They shed light on legitimization crises, conflict resolution (Deutsch & Coleman, 2000), and systemic risk in complexity

and network science (Helbing, 2013; Smith et al. 2014; Cranmer et al., 2015). That is because those risks raised further contagion and diffusion of disturbing factors beyond boundaries in the global networking era. The aftermath of ‘Arab Spring’ diffused around other nations and caused a domino cascade of national disruptions (Boening, 2014; Sadiki, 2015). After those movements and uprisings achieved their initial goals, and their ancient régime was reassigned to a renewed government in the name of citizens.

The new regime in Egypt has been struggling because no legitimate authority has been accepted by civilians since the ‘Arab Spring’. Furthermore, the direct and indirect results of revolutionary actions certainly caused severe damage against their daily life and common wealth.

Moreover, more severe international affairs related to masses of emigrants seeking asylum and refugees from Middle-East regions to the EU. Those populations, estimated as nearly half a million people (Migration policy centre, 2012), eagerly communicate with others using smart-phones and social media. Small exceptions made for them paved the ways to a rush into the door of EU. Calls for opposition demonstrations against immigrants’ policy were made in the EU member states, and antagonists described inequities and aggressions committed against them.

FUTURE RESEARCH DIRECTIONS

Social scientists explored appropriate methodologies to assess real-time dynamics of human behaviors. Their efforts can be improved by analyzing large amounts of data using social media and ubiquitous computational services (Shibuya, 2004). It is also regarded as big data for additional analysis in political studies (Golbeck & Hansen, 2014). As noted earlier, this research style is realized as *computational social science* (Lazer et al., 2009). It has been applied by computational advancements such as visualization, data mining, text mining, network analysis, sentiment analysis,

modeling, and simulations. Furthermore, geospatial information systems (GIS) can be appended to the list above. These backgrounds have been examining collective behavior patterns of ordinary citizens using mobile devices (Gonzalez, Hidalgo & Barabasi, 2008; Hedstrom & Bearman, 2009; Newman, 2010; Bond et al., 2012; Evangelopoulos & Visnescu, 2012; Doer, Fouz & Friedrich, 2012; Agarwal et al., 2013). Therefore, it is not surprising that many researchers will be striving to elucidate these mechanisms.

CONCLUSION

The author argued that the advent of social media summoned *the collective dynamics of democracy of the citizens, by the citizens, and for the citizens*. Those patterns readily transformed the form of social movements to mutual interconnection and shaping of the enclaves of networked clustering. Social media have offered a new paradigm of democracy such as engagement and participation for ordinary citizens.

As Fukuyama (1992) forecasted “Is history directional, and is there reason to think that there will be a universal evolution in the direction of liberal democracy?”, the ‘Arab Spring’ and similar events have been repeated ever since civilizations have existed. Furthermore, is the utilitarian’s slogan “The greatest happiness of the greatest number” always correct in democracy? Then stakeholders must tackle unavoidable questions at each occasion. They should choose solutions through deliberative and collaborative activities.

REFERENCES

- Agarwal, N., Wigand, R. T., & Lim, M. (2013). *Online Collective Action: Dynamics of the Crowd in Social Media*. Springer.
- Anheier, H. K., & Toepler, S. (2010). *International Encyclopedia of Civil Society*. Springer. doi:10.1007/978-0-387-93996-4
- Aral, S., Muchnik, L., & Sundararajan, A. (2009). Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences of the United States of America*, 106(51), 21544–21549. doi:10.1073/pnas.0908800106 PMID:20007780
- Ariely, G. (2015). Democracy-Assessment in Cross-National Surveys: A Critical Examination of How People Evaluate Their Regime. *Social Indicators Research*, 121(3), 621–635. doi:10.1007/s11205-014-0666-y
- Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348(6239), 1130–1132. doi:10.1126/science.aaa1160 PMID:25953820
- Bello, J., & Rolfe, M. (2014). Is influence mightier than selection? Forging agreement in political discussion networks during a campaign. *Social Networks*, 36, 134–146. doi:10.1016/j.socnet.2013.06.001
- Blondel, V. D. (2015). A survey of results on mobile phone datasets analysis. *EPJ Data Science*, 4(10). doi:10.1140/epjds/s13688-015-0046-0
- Boening, A. (2014). *The Arab Spring: Re-Balancing the Greater Euro-Mediterranean?* Springer. doi:10.1007/978-3-319-04606-8
- Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D. I., Marlow, C., Settle, J. E., & Fowler, J. H. (2012). A 61-million-Person Experiment in social influence and political mobilization. *Nature*, 489(7415), 295–298. doi:10.1038/nature11421 PMID:22972300
- Braha, D. (2012). Global Civil Unrest: Contagion, Self-Organization, and Prediction. *PLoS ONE*, 7(10), e48596. doi:10.1371/journal.pone.0048596 PMID:23119067
- Broadbent, S. R., & Hammersley, J. M. (1957) Percolation Processes. *Mathematical Proceedings of the Cambridge Philosophical Society*. doi:10.1017/S0305004100032680

- Brown, R. (2000). *Group Processes* (2nd ed.). Blackwell.
- Bryan, R. (2014). Social media in the 2011 Egyptian uprising. *The British Journal of Sociology*, 65(2), 266–292. doi:10.1111/1468-4446.12080 PMID:24798232
- Casilli, A. A., & Tubaro, R. (2012). Social Media Censorship in Times of Political Unrest - A Social Simulation Experiment with the UK Riots. *Bulletin de Methodologie Sociologique*, 115(1), 5–20. doi:10.1177/0759106312445697
- Castells, M. (2012). Networks of Outrage and Hope: Social Movements in the Internet Age. *Polity*.
- Choucri, N. (2012). *Cyberpolitics in International Relations*. The MIT Press.
- Choudhary, A., Hendrix, W., Lee, K., Palseltia, D., & Liao, W.-K. (2012, May). Social media evolution of the Egyptian revolution. *Communications of the ACM*, 55(5), 74–80. doi:10.1145/2160718.2160736
- Cranmer, S. J., Menninga, E. J., & Mucha, P. J. (2015). Kantian fractionalization predicts the conflict propensity of the international system. *Proceedings of the National Academy of Sciences of the United States of America*, 112(38), 11812–11816. doi:10.1073/pnas.1509423112 PMID:26338977
- Deutsch, M., & Coleman, P. T. (2000). *The Handbook of Conflict Resolution Theory and Practice*. Jossey-Bass Publisher.
- Doer, B., Fouz, M., & Friedrich, T. (2012). Why Rumors spread so quickly in social networks. *Communications of the ACM*, 55(6), 70–75. doi:10.1145/2184319.2184338
- Dogan, M. (1992). Conceptions of Legitimacy. In M. Hawkesworth & M. Kogan (Eds.), *Encyclopedia of Government and Politics*. Routledge.
- Downs, A. (1957). An Economic Theory of Political Action in a Democracy. *Journal of Political Economy*, 65(2), 135–150. doi:10.1086/257897
- Epstein, J. M. (2014). *Agent_Zero: Toward Neurocognitive Foundations for Generative Social Science*. Princeton University Press.
- Evangelopoulos, N., & Visnesu, L. (2012). Text-Mining the Voice of the People. *Communications of the ACM*, 55(2), 62–69. doi:10.1145/2076450.2076467
- Fukuyama, F. (1992). *The End of History and the Last Man*. Free Press.
- Fukuyama, F. (2014). *Political Order and Political Decay*. Farrar Straus & Giroux.
- Golbeck, J., & Hansen, D. (2014). A method for computing political preference among Twitter followers. *Social Networks*, 36, 177–184. doi:10.1016/j.socnet.2013.07.004
- Gonzalez, M. C., Hidalgo, C. A., & Barabasi, A. (2008). Understanding individual human mobility patterns. *Nature*, 453(7196), 779–782. doi:10.1038/nature06958 PMID:18528393
- Gonzalez-Bailon, S., & Wang, N. (2016). Networked disconnect: The Anatomy of protest campaigns in social media. *Social Networks*, 44, 95–104. doi:10.1016/j.socnet.2015.07.003
- Granovetter, M. S. (1978). Threshold Models of Collective Behavior. *American Journal of Sociology*, 78(6), 1420–1443. doi:10.1086/226707
- Habermas, J. (1976). *Legitimation Crisis*. Heine-mann.
- Hedstrom, P., & Bearman, P. (2009). *The Oxford Handbook of Analytical Sociology*. Oxford University Press.
- Helbing, D. (2012). *Social Self-Organization*. Springer. doi:10.1007/978-3-642-24004-1

- Helbing, D. (2013). Globally networked risks and how to respond. *Nature*, 497(7447), 51–59. doi:10.1038/nature12047 PMID:23636396
- Hill, K. A., & Hughes, J. E. (1998). *Cyberpolitics: Citizen Activism in the Age of the Internet*. Rowman & Littlefield Publishers.
- Hoshen, J., & Kopelman, R. (1976). Percolation and cluster distribution. I. Cluster multiple labeling technique and critical concentration algorithm. *Physical Review B: Condensed Matter and Materials Physics*, 14(8), 3438–3445. doi:10.1103/PhysRevB.14.3438
- Huckfeldt, R., Pietryka, M. T., & Reilly, J. (2014). Noise, bias, and expertise in political communication networks. *Social Networks*, 36, 110–121. doi:10.1016/j.socnet.2013.02.003
- Johnson, G. F. (2008). *Deliberative Democracy for the Future: the case of nuclear waste management in Canada*. University of Toronto Press. doi:10.3138/9781442687837
- Jordan, T. (2001). Language and libertarianism: The politics of cyberculture and the culture of cyberpolitics. *The Sociological Review*, 49(Issue.1), 1–17. doi:10.1111/1467-954X.00241
- Jost, J. T., & Major, B. (2001). *The Psychology of Legitimacy Emerging Perspectives on ideology, justice, and Intergroup relations*. Cambridge University Press.
- Katz, E., & Lazarsfeld, P. F. (1955). *Personal Influence: The Part played by People in the Flow of Mass Communication*. Free Press.
- Klandermans, B., & Roggeband, C. (2010). *Handbook of Social Movements Across Disciplines*. Springer. doi:10.1007/978-0-387-70960-4
- Le Bonn, G. (1931a). *The Crowd: Study of the Popular mind*. iBook Version.
- Le Bonn, G. (1931b). *The Psychology of Revolution*. iBook Version.
- Lipset, S. M. (1959). Some Social Requisites of Democracy: Economic Development and Political Legitimacy. *The American Political Science Review*, 53(1), 69–105. doi:10.2307/1951731
- Locke, J. (1988). *Two Treatises of Government*. Cambridge University Press. doi:10.1017/CBO9780511810268
- MacCoun, R. J. (2012). The Burden of Social Proof: Shared Thresholds and Social Influence. *Psychological Review*, 119(2), 345–372. doi:10.1037/a0027121 PMID:22352358
- Malarz, K., & Galam, M. (2005). Square-lattice site percolation at increasing ranges of neighbor bonds. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, 71(1), 016125. doi:10.1103/PhysRevE.71.016125 PMID:15697676
- Malinick, T. E., Tindall, D. B., & Diani, M. (2013). Network Centrality and social movement media coverage: A two-mode network analytic approach. *Social Networks*, 35(2), 148–158. doi:10.1016/j.socnet.2011.10.005
- Manilov, M. (2013). Occupy at one Year: Growing the Roots of a Movement. *The Sociological Quarterly*, 54(2), 159–228. doi:10.1111/tsq.12022
- Migration Policy Centre. (2012). *The European Response to the Syrian Refugee Crisis, What Next?* Retrieved from <http://www.migrationpolicycentre.eu/docs/MPC%202012%2014.pdf>
- Newman, M. E. J. (2010). *Network: An Introduction*. Oxford University Press. doi:10.1093/acprof:oso/9780199206650.001.0001
- Newman, M. E. J., & Park, J. (2003). Why social networks are different from other types of networks. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, 68(3), 036122. doi:10.1103/PhysRevE.68.036122 PMID:14524847
- Newman, M. E. J., & Ziff, R. M. (2001). Fast Monte Carlo algorithm for site or bond percolation. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, 64(1), 016706. doi:10.1103/PhysRevE.64.016706 PMID:11461441

- Noelle-Neumann, E. (1984). *The spiral of silence*. University of Chicago Press.
- Nozick, R. (1974). *Anarchy, State, and Utopia*. Basic Books.
- O'Sullivan, D., & Perry, G.L.W. (2013). *Spatial Simulation: Exploring Pattern and Process*. Wiley-Blackwell.
- Olson, M. (1971). *The Logic of Collective Action: Public Goods and The Theory of Groups*. Harvard University Press.
- Rawls, J. (1971). *A Theory of Justice*. Harvard University Press.
- Rhue, L., & Sundararajan, A. (2014). Digital access, political networks and the diffusion of democracy. *Social Networks*, 36, 40–53. doi:10.1016/j.socnet.2012.06.007
- Rohlinger, D. A., & Snow, D. A. (2003). Social Psychological Perspectives on Crowds and Social Movement. In J. Delamater (Ed.), *Handbook of Social Psychology*. Kluwer Academic Publishing.
- Rossi, L. (2015). Spatio-temporal techniques for user identification by means of GPS mobility data. *EPJ Data Science*, 4(11). doi:10.1140/edjds/s13688-015-0049-x
- Rousseau, J. J. (2009). *Of The Social Contract, Or Principles of Political Right*. Regnery Pub.
- Sadiki, L. (2015). *Routledge Handbook of the Arab Spring: Rethinking Democratization*. Routledge.
- Sahimi, M. (1994). *Applications of Percolation Theory*. Taylor & Francis.
- Shibuya, K. (2004). A Framework of Multi-Agent Based Modeling, Simulation and Computational Assistance in an Ubiquitous Environment. *Simulation*, 80(7-8), 367–380. doi:10.1177/003754970404046740
- Shibuya, K. (2016). (in press). Why is such criterion in Democracy?: Thinking about the Majority Rule and Legitimacy. *Proceedings of the Meeting of the German-Japanese Society for Social Science*, 13.
- Smelser, N. J. (1962). *Theory of Collective Behavior*. The Free Press.
- Smith, J. M., Halgin, D. S., Kidwell-Lopez, V., Labianca, G., Brass, D. J., & Borgatti, S. P. (2014). Power in politically charged networks. *Social Networks*, 36, 162–176. doi:10.1016/j.socnet.2013.04.007
- Sunstein, C. R. (2001). *Republic.com*. Princeton University Press.
- Watts, D. (2002). A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences of the United States of America*, 99(9), 5766–5771. doi:10.1073/pnas.082090499 PMID:16578874
- Weber, M. (1958). The three types of legitimate rule. *Berkeley Publications in Society and Institutions*, 4(1), 1–11.

ADDITIONAL READING

- Agarwal, N., Wigand, R. T., & Lim, M. (2013). *Online Collective Action: Dynamics of the Crowd in Social Media*. Springer.
- Choucri, N. (2012). *Cyberpolitics in International Relations*. The MIT Press.
- Habermas, J. (1976). *Legitimation Crisis*. Heine-mann.
- Hedstrom, P., & Bearman, P. (2009). *The Oxford Handbook of Analytical Sociology*. Oxford University Press.
- Helbing, D. (2013). Globally networked risks and how to respond. *Nature*, 497(7447), 51–59. doi:10.1038/nature12047 PMID:23636396

Hill, K. A., & Hughes, J. E. (1998). *Cyberpolitics: Citizen Activism in the Age of the Internet*. Rowman & Littlefield Publishers.

Sadiki, L. (2015). *Routledge Handbook of the Arab Spring: Rethinking Democratization*. Routledge.

Smelser, N. J. (1962). *Theory of Collective Behavior*. The Free Press.

Sunstein, C. R. (2001). *Republic.com*. Princeton University Press.

KEY TERMS AND DEFINITIONS

Acceptance of Political Regimes by Consensual Sovereignty: It shall be consensually accepted by citizens, and it is grounded on democratic civil society (Anheier & Toepler, 2010). Citizen equals with the concept on sovereign in democratic society.

Accordance with Rules and Social Contract between the Governor and a Sovereign: These are a lexical definition (Oxford English Dictionary) and social theories (e.g. Rousseau and others).

Cyber Politics: Political actions online have flourished since the advent of social media. Using big-data accumulated in Twitter and SNS,

real-time dynamics of users' opinion and their political attitudes have been unveiled.

Legal-Rationality: Sociologist Weber (1958) categorized legitimacy as three types such as traditional, charismatic and legal-rational.

Legitimacy: In political definitions, the concept of legitimacy in democracy consists of at least three factors.

Legitimation Crisis: It means the forfeiture of legitimacy in society (Habermas, 1976), and it denotes a breach of trust in national governance.

Percolation: Percolation can be observed widely as various diffusion and self-organized clustering patterns in a geometric field. Percolation was explored in mathematics, statistical physics and chemistry, computational complexity sciences and social sciences (Broadbent & Hammersley, 1957; Hoshen & Kopelman, 1976; Sahimi, 1994; Newman & Ziff, 2001; Newman & Park, 2003; Malarz & Galam, 2005; O'Sullivan & Perry, 2013). Percolation in clustered networks can be categorized to two types as sites (spatial adjacent nodes are occupied or not) and bonds (interconnected edges or not). In addition, each network pattern has its own specific critical point (threshold) for phase transitions, which indicates the borderline between clustered organization and disruption of networks in a spatiotemporal field.

Community Broadband Networks and the Opportunity for E-Government Services

G

Idongesit Williams

Aalborg University, Denmark

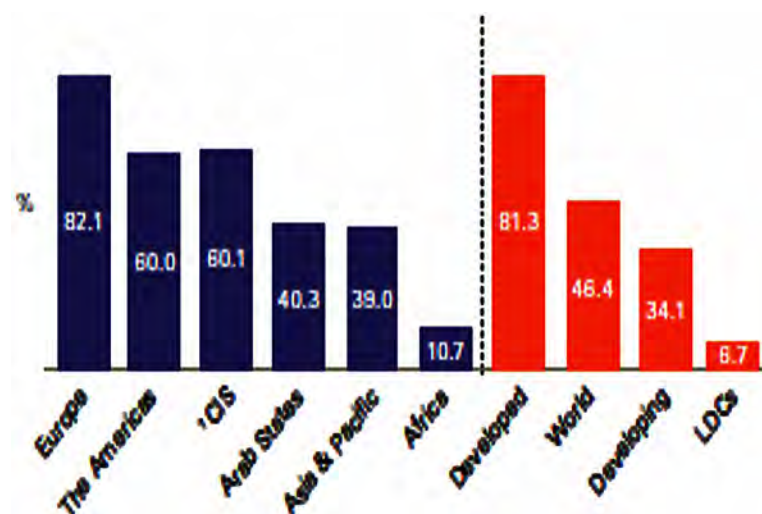
INTRODUCTION

This paper discusses how communities in underserved areas in developed and developing countries can develop Broadband connectivity in order to access e-government services. Community Based Broadband Mobilization (CBNM) models developed by Williams (2015) are used as explanatory tools for the discussion. In this article, a community refers to a social unit bound by a common purpose or characteristics. Communities include villages, suburban areas, academic groups, professional organizations, neighborhoods, non-profit organizations, etc., In this article, the context of “the community” is rural areas. Communities have developed fixed and wireless Broadband networks in developed and developing countries (see (Salemink & Bosworth, 2014; Williams, 2015)).

They do so by forming Community Broadband Networks (CBN). The CBNs are communities that facilitate the development of Broadband networks in their communities.

The existence of CBNs presents an opportunity for facilitating Broadband connectivity in underserved (rural) areas. It also provides the opportunity to facilitate e-government services in underserved areas. This is because, the existence of Broadband network connectivity presents the possibility for the delivery of the e-government services in such areas (Zambrano & Seward, 2013). Hence, CBNs provided the boost for the implementation of e-government infrastructure and services in rural areas in many countries. They aid in reducing e-government network infrastructure deficiency in many countries. This deficiency exists (ITU, 2015; Zambrano & Seward, 2013;

*Figure 1. Percentage of households with internet access
ITU, 2015.*



DOI: 10.4018/978-1-5225-2255-3.ch308

Wachira & Arlikatti, 2010; UN, 2014). The ITU version of the Internet Infrastructure deficiency is represented in the figure below. Countries in developing countries suffer more from the infrastructure deficiency than those in the developed countries.

However, it is not every community that has the desire to facilitate CBNs. Residents of rural communities are inundated with competing economic and social needs. These needs eclipse their desire to facilitate Broadband infrastructure. The question now is, how can the public sector mobilize communities to facilitate Broadband infrastructure for the purpose of e-government service delivery?

To answer this question, this article adopts the CBNM models by Williams (2015). These models explain why communities facilitate CBNs. It presents an idea of the conducive conditions that should exist before communities in developed and developing countries can facilitate CBNs. In both models, the priority is placed on the level of usefulness of the service to the community. If the proposed service has the potential to enhance an important aspect of their lives, then - for them - it is a service worth having. Based on the premise of these models, this article explains that e-government services are services worth having access to. Hence, if people in communities are made to understand their need for e-government services, the desire to facilitate a cost effective Broadband service will be prompted among them. These explanations are accompanied by simulated examples.

Based on this exercise, this article reveals that CBNs are important but ignored players in the facilitation of e-government infrastructure and service delivery. Hence, efforts should be made to see how CBNs can be incorporated to deliver network infrastructure needed for e-government services, to underserved areas. In the upcoming sections, the concept of e-government will be discussed. Here the connectivity problem will be highlighted. This will be followed by a discussion on the concept of CBNs. A link will be created

between two concepts using the CBNM model by hypothetically simulating the relationship between e-government services and CBN.

BACKGROUND

In this section, the concept of e-Government and how the connectivity factors impede e-Government service delivery will be discussed. The major themes in this section are: what is E-government and why it is necessary? How is e-government facilitated and what is the connectivity problem?

What Is E-Government and Why Is It Necessary?

E-government is the delivery of government services to citizens via the Internet (Kettani & Moulin, 2014). National governments globally are keen on facilitating an information society via e-government (UN, 2014). The primary beneficiaries of the information society are citizens and businesses. E-government enhances Government-Citizen and Government-Business relationships. The success of these relationships is hinged on “efficiency” and “trust” resulting from the level of transparency in the delivery of government services. The quest to achieve efficiency and transparency in the delivery of government service makes e-government necessary.

Previously transparency, in the delivery of government services, was enhanced by the use of old media. Such media platforms included television, radio, newspapers, notice boards at government offices and local community information outposts. The disadvantage of the old media platforms was the difficulty in creating efficient accessible archives to citizens. Retrieving vital data from the archives was a challenge. The advent of the mainframe computer, minicomputer and later Personal Computers (PCs) in the 20th century provided an efficient way of archiving information for government agencies (Campbell-Kelly & Aspray, 1996; Norris & Kraemer, 1996). This

was the bridge between the old media platform and new media platform. Government agencies gradually transferred relevant archived data needed for Government-citizen and Government - Business interaction into the computers. But computers, then, were only used for interactions between Government - Employee and Government-Government relationships. In this scenario, the supporting technologies were Local Area Networks, telephones, fax and telegraph services (Norris & Kraemer, 1996). Citizens and businesses had to rely on old media (radio and TV), telephones, fax and telegraph services in dealing with governments.

However, the efficient delivery of government services was impeded by bureaucracy (Ho, 2002). Citizens and businesses battled with the command structure of the government agency to get things done. The processes involved in the delivery of public services, such as procurement, required some level of scrutiny. This entailed the creation of various oversight positions for government personnel. This was necessary to ensure value for money for the public sector in the procurement process. However, bureaucracy resulted in high transaction costs for the public sector and for the citizens and businesses as well.

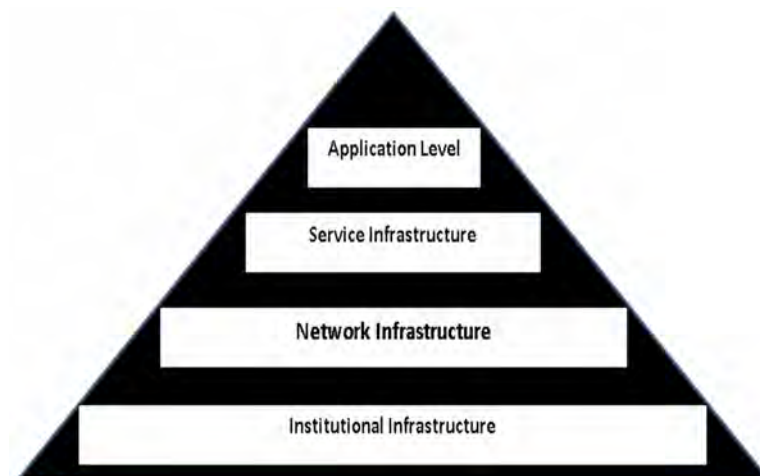
Today, technological advancements provides the opportunity for the public sector to deliver their services in an efficient way and in a more transparent manner. The interactive features of new media has enhanced transparency in the governance process (Calista & Melitski, 2007). New media platforms are enabled via the technological convergence of the Broadcast, telecommunications and electronic communication platforms (Williams & Falch, 2012). This convergence is facilitated via Broadband networks (Calista & Melitski, 2007).

This technological evolution has minimized bureaucracy. Fewer staffs are now needed by government agencies. Soft copies of relevant information stored on PCs earlier be accessed electronically by citizens and businesses today. The citizen and business owners do not necessarily have to visit the government agency. Rather, they have remote access to Government services at their convenience. Hence the move towards efficiency and more transparency in government service delivery via e-government platforms is assured.

How Is E-Government Facilitated?

E-government facilitation can be identified in four levels. These include, the Service infrastructure,

Figure 2. The four levels of e-government facilitation to the citizen Williams, 2016.



Network infrastructure level, the institutional infrastructure and the application level.

As seen in Figure 2, the upper layers rely on the existence of the lower layer. The emphasis of this article is on the network infrastructure level. The argument is, if this infrastructure exists in underserved communities, then people in such communities will have access to the services provided at the application level. However, a brief overview of the other levels are as follows:

1. **The Institutional Infrastructure:** They coordinate the network infrastructure, the service infrastructure and the applications needed to facilitate e-government implementation. The institutional infrastructure includes the existence of regulatory bodies with the requisite legal mandate to coordinate the implementation of e-government in a given country or region (Zambrano & Seward, 2013). Internationally, UN agencies promote e-government implementation (See (UN, 2014)). They partner with national e-government agencies and International Development Partners to facilitate national e-government ecosystem.
2. **The Network Infrastructure:** This is the telecommunication or Internet

Infrastructure, transporting the e-government services to and from the end user (See (Zambrano & Seward, 2013)). Broadband is the preferred communications technology by Governments today. This can be seen in the case of Sweden, Denmark, Germany and South Korea, etc (OECD, 2008). Broadband is adopted because of the enhanced data rates, capacity and Quality of Service it provides.

3. **The Service Infrastructure Includes:** These consist of infrastructure elements used by the Government agency(ies) to support the delivery of the e-government applications. These include software, hardware, data center, personnel and requisite policies mandating the specific government service to deliver the service. The service infrastructure may be a coordinated service infrastructure or a service infrastructure developed in silos.
4. **The Application Level:** This is the layer where e-government applications are accessed by the citizens, governments, businesses and employees (Chairunisa, Malik, & Rahman, 2016).

In each of these levels, there are the demand and supply actors that aid the e-government facilitation process. This reflects in Table 1.

Table 1. Supply and demand actors in each layer of e-government implementation

	Supply Actors	Demand Actors
Institutional infrastructure	• National agencies mandated to facilitate e-government services in the country • *The law making entities	• Citizens • Businesses • Employees • Government
Network infrastructure	• Broadband Network infrastructure manufacturers • The e-government infrastructure operator • The law making entities, • Commercial Broadband Network operators	• Citizens • Businesses • Employees • Government
Service Infrastructure	• The software and hardware vendors, • The supplier of connectivity	Sectoral government agencies
Application level	Sectoral government agencies	• Citizens • Businesses • Employees

Williams, 2016.

*These law making entities are the legislative bodies that provide the regulatory framework for the facilitation of e-government initiatives in a country.

However, at the network infrastructure level, which is of interest to article, CBNs currently are not recognized by e-government implementation agencies. This is because communities are seen as consumers. Studies indicate that the e-government network infrastructure is often provided by commercial operators. The commercial operator in many cases operates under a Public Private Partnership (PPP) agreement (UN, 2014). But can communities really help? Before discussing this question, the next subsection explains why there is a connectivity problem.

Why Is There a Connectivity Problem?

The connectivity problem as mentioned earlier in this article was linked to the existence of Broadband Access gaps in different countries. This is a supply-side problem at the network infrastructure level. The Universal Access of Broadband connectivity in any region or country presents the opportunity for Government institutions to deliver their services online. This is why, network connectivity is seen as a fundamental problem. Hence the argument, is, if CBNs can solve this problem, why not facilitate them? Arguments for the facilitating these networks will be explained in the next section.

SOLUTIONS AND RECOMMENDATIONS

In this section the concept of CBNs and the CBNM models are discussed. The themes in this section are: What are CBNs? And why does CBNs exist? This question introduces the CBNM models.

What Are Community Broadband Networks?

CBNs are Broadband Network facilitated by a community. Their area of operations could be a neighborhood in a city or a rural area. These

networks are mostly wireless networks. However, in recent times, communities have embarked on facilitating expensive high capacity networks (Salemink & Bosworth, 2014; Williams, 2015). Community networks have been in existence since the early part of the 20th century. Fixed-line telephony infrastructure in rural areas have been facilitated by local cooperatives mostly in developed countries (Finkelievich & Kisilevsky, 2005). This tradition still continues today with the facilitation of Broadband infrastructure (see (Salemink & Bosworth, 2014)). This is evident in communities and neighborhoods in the global north and south respectively. Some affluent communities in the global north have facilitated expensive networks such as Fiber to the home. This can be seen in Sweden and in the Netherlands (Lindskog & Johansson, 2005). However, in less affluent communities in the global north and south, Wi-Fi over fiber connectivity is used to deliver Broadband connectivity. Examples can be found in the United States, Argentina, India, South Africa, Philippines, Canada, and in Europe (Williams, 2015). In some cases, these communities facilitate the network either with the aid of public financing, joint public ownership or in most cases self-effort (Kakekaspan, O'Donnell, Beaton, Walmark, & Gibson, 2014). The advantage of CBNs is the fact that they facilitate Broadband networks in areas where commercial operators shy away from (Williams, 2015). They manage these networks by forming cooperatives governed by democratic values.

Why Do Community Broadband Networks Exist?

This question is answered in a study conducted by Williams (2015). The study was conducted to understand the ways these networks can be financed and organized in rural areas in developed and developing countries. As part of the study, there were investigations in to why rural communities in developed and developing countries facilitate these networks (Williams, 2015). He studied three

Figure 3. Community Broadband Network Mobilization (CBNM) model for the developed countries
Source: Williams, 2015.



cases each in developed and developing countries. The developed countries were namely, USA, Sweden and Denmark. The developing countries were namely, Ghana, South Africa and India. That aspect of the research produced, 2 models named the Community Based Network Mobilization Models (CBNM). These models indicated that, if the causal factors of the CBNM models exist or are promoted, rural communities in developed and developing countries will facilitate Broadband networks in their areas.

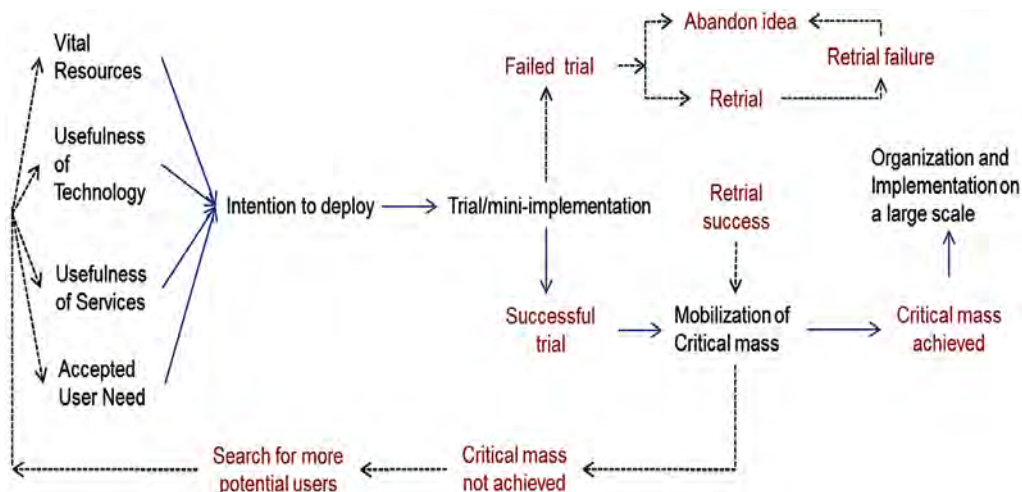
The first model was the CBNM model for the developed countries. This is seen in the figure below.

The first model explains thus: If the citizens of communities in rural areas of a developed country can perceive the usefulness of a technology, the deployment possibilities and they possess the vital resources needed to facilitate the technol-

ogy, they will mobilize themselves and deploy the technology. The findings of the study suggest that, people in rural areas in the developed countries already had an idea of services, the technology can deliver. This was because, they were already using the services. Hence they already knew the usefulness of the service. What they cared about was, how they could have access to a better Quality of Service at a cost they could afford. The vital resources concept in the study implied any resource needed to facilitate the network. These could be intrinsic resources such as technical knowledge, innovative thinking to extrinsic resources such as financial resources, human resources (mostly volunteers), technical resources, such as technical equipment etc..

The second model was the CBNM model for developing countries. This is seen in the figure below.

Figure 4. Community Broadband Network Mobilization (CBNM) model for developing countries
Source: Williams, 2015



The second model explains thus: If the citizens of communities in rural areas of developing countries can perceive the usefulness of the service, the usefulness of a technology, the place of the technology in their daily lives (accepted user need) and they possess the vital resources to facilitate the technology, they will mobilize themselves and deploy the technology.

Based on the study, it was realized that the citizens of developed countries were better off financially than their counterparts in developing countries. The citizens in the developed country were conversant with the value of the service to their lives. Hence, when they realized that they could facilitate the technology that will deliver the service, they went for it.

However, in developing countries, many had no idea about Broadband services. Hence, they needed demonstrations to enable them see the usefulness of the service. But seeing the usefulness of the service was not enough. They had other competing needs in their lives. Hence the demonstrations had to prove that using the Broadband service was a better substitute to their existing solutions. Williams (2015) mentioned VoIP as a cheaper substitute to voice telephony and YouTube as a better entertainment substitute to the limited channels provided by the local television stations. The study further revealed that this insight into substitute possibilities did aid the would-be user see the usefulness of the technology. Hence, if they have the vital resources, they will facilitate the network. The vital resources concept is similar to that of the developed country cases. However, as seen in the CBNM model for developing countries, the lack of financial resources led the initiators to use the demonstration as a way of mobilizing users to deploy the network. Once the population achieved critical mass needed to raise an economy, they formed organizations and deployed on a broad scale. That is why the loop exists in the framework.

HYPOTHETICAL RELATIONSHIP BETWEEN E-GOVIRONMENT SERVICES AND COMMUNITY BROADBAND NETWORKS

Having explained, the concept of CBN and introduce the CBNM models, how can community networks aid in facilitating e-government, hence extending connectivity? In this section e-government service inspirations that could trigger communities to adopt these models are explored. The examples here are hypothetical.

1. Hypothetical Relationship From the Developed Country Perspective:

- a. **Service Inspiration:** In Denmark and Sweden citizens can submit their tax returns online (Williams, Kwofie, & Sidii, 2016). In Germany, citizens have access to an online “tax calculator”. Here citizens can have an idea of their tax rates (Steuer, 2016). In Estonia and Switzerland, voters participate in election via online voting from the comfort of their homes (Evoting Estonia, 2016; Evoting Switzerland, 2016). In some states in the United States, such as Texas and Colorado, people can renew their drivers license online (TDPS, 2016). These are examples of government services provided online in some developed countries. In most developed countries, municipalities also have various services by which they provide online (OECD, 2008).
- b. **Perceived Usefulness of the Technology to the Services:** The examples cited above are services needed by those living in commercially viable areas as well as non-commercially viable areas. No one wants to pay more tax than he or she is supposed to. No one enjoys queuing all day or

having to drive to a polling station to vote. Also, having your drivers license delivered to your mailbox is of great convenience. If these services are to be accessed by the citizens and businesses, then there is a need for access to a high capacity Internet or telecommunications network in their premises. Hence, Broadband networks delivering at least 256 kbps to the citizen are of importance. Therefore, these needs can serve as a trigger that can facilitate the CBNM model. This is an example of the “perceived usefulness” for a the technology (Broadband) for accessing e-government services.

- c. **Deployment Possibilities:** As explained by Williams (2015), the deployment possibilities consist of gateway possibilities to the community, the access connectivity to the homes and the ability to facilitate the technology technically and economically. The study explains that the communities, driven by the need for the service, conducted a search for deployment possibilities. In facilitating CBN aimed at facilitating the delivery of the above mentioned e-government services to its resident, there are technical deployment possibilities.

- i. **Community Gateway**

Connectivity Possibilities: In a lot of developed countries, there are different Fixed and Wireless Broadband gateway possibilities. The fixed gateway possibilities include fiber optics and ADSL networks passing close to the communities (Williams, 2015). This is because central and middle level governments (municipalities inclusive) in the United States and in the EU have invested in the development of

National Broadband backhaul projects (Feijoo, Gomez-Barroso, & Bohlin, 2011). Most of these infrastructures are in close proximity to rural areas and underserved areas. The second possibility is the rapid deployment of 3G and 4G wireless Networks by Mobile network operators in western cities with close proximity to the community (Williams, 2015). Hence, gateway connectivity can be achieved via a wired or wireless link, depending on the economic ability of the community.

- ii. **Community Access**

Connectivity Possibilities: A low cost Wireless Local Area Network, such as Wi-Fi can be deployed as an access network to the homes in the community. Recently, Wi-Fi standards have improved. In 2014, the 802.11ac (so called 5G WiFi) was developed. It can provide a throughput of up to 6Ghz (Aboul-Magd, Kim, Wentink, Yang, & Stacey, 2013). The Wi-Fi spectrum in most countries operating in the 2.4GHz to 5MHz frequency band is unlicensed. As seen in the study by Williams (2015), it was interesting to note that the cost of deploying Wi-Fi can be reduced, by the fabrication of antennas and the use of rooftops as base stations for the antennas. Hence there is the possibility of deploying such networks in developed countries to aid the extension of e- government services.

- d. **Vital Resources:** The vital resources here in the developed country context will include, financial resources of the individual members of the community,

financial resources they can harness from outside, human resources (technical and non-technical volunteers), non-technical resources, technical resources and self-will. Williams (2015) explains that self-determination was critical to the communities as they had to overcome regulatory odds, technical odds, and financial odds to fulfil their dreams. Examples of non-technical resources include, the use of hills, houses and elevated plane as natural base stations. There could be other non-technical resources needed depending on the technology to be deployed by the community and the terrain of the community.

This example indicates how the CBNM model for the developed countries can be used to deploy hypothetical e-government solutions by CBNs in the developed countries. The important factor here that the e-government service serves as a driver.

2. Hypothetical Relationship From the Developed Country Perspective:

- a. **Services Inspiration:** In some developing countries, there is the possibility of facilitating public health, election and local market-management online. Though few of such initiatives exist, the supply of e-government services is still limited (UN, 2014).
- b. **Usefulness of the Service:** The services mentioned are useful. This is because, the market is where food to nourish the body is bought. Without the market, most people without farms will starve. The health center is where ailments are treated. If they do not exist, the mortality rate will be high. People in developing countries care about whom they elect as leaders. Election turnout in most developing countries is high (Pintor, Gratschew, & Sullivan, 2002).
- c. **Usefulness of the Technology:** The technology solution needed to deliver health, election and market management as e-government services, is similar to that of the developed country. Williams (2015) recommends a Wi-Fi network over 3G or 4G connectivity for CBNs in developing countries. Wi-Fi is recommended because it provides Broadband connectivity and the antenna parts can be fabricated by the people. Broadband will provide a better Quality of Service delivery for the services. Hence, it makes the technology useful. However, the people may not abandon competing daily needs to facilitate the network, by just knowing the usefulness the technology.
- d. **Accepted User Need:** A small network created for a single case such as e voting, e health or market management. The Williams (2015) study, explains that such an initiative will lead people to try the service for themselves. In this manner, they could see the convenience in service delivery and will whet their appetite for the service. However, it is the existence of the vital resources that leads the community to attempt implementation.
- e. **Vital Resources:** This concept is same as that of the developed countries. However, here, Williams (2015) explains that retirees, educated youths that could not succeed in the city and urban-rural migrants are great non-technical human resources. This is because they can learn about the technology and teach the others why the service is useful and how they can facilitate it. In rural areas in developing countries, there are scrap yards. In these yards lay materials that can be recycled by the locals and used as antennas. There are many other vital resources that would exist. This will vary from country to country.

This example indicates how the CBNM model for the developing countries can be used to deploy hypothetical e-government solutions by CBNs in the developed countries. The important factor here again is that the e-government service serves as a driver.

FUTURE RESEARCH DIRECTIONS

Research into CBNs centers on how communities develop and use ICT Infrastructure. ICT infrastructure and service delivery via CBNs has been investigated in sociology and Information Science respectively.

(Carrol & Rosson, 2007). The CBNM models serves as analytical frameworks for researchers in both fields. The model will aid researchers investigate the facilitation and delivery of an ICT services in a particular location.

CONCLUSION

CBNM models adopted by the CBNs in the Williams (2015) study have the potential to facilitate Broadband infrastructure to aid the delivery of e-government services. However, an enabling environment is needed to facilitate the CBNs. The factors that would create an enabling environment includes:

1. **The Identification of User Needs:** The e-government implementation agency has to identify the e-government services that people in such communities cannot do without.
2. **Awareness Creation:** There are communities whose residents may not ignorant of the usefulness of the e-government services to them. Capacity building initiatives will be relevant here. Such initiatives will create the awareness of the usefulness of the e-

government services to the community. They will identify their need for the services. This will lead to the desire to have access to the service.

3. **The Removal of Regulatory Barriers:** The Williams (2015) study proposes the removal of regulatory barriers for CBNs. This is possible if national e-government implementing agencies, recognize CBN as partners in the e-infrastructure delivery process. CBNs hold the key towards enhancing access connectivity to their locality. The fact that such communities have the potential to facilitate broadband connectivity, should be an encouragement via requisite policies.
4. **Public Policies Aimed at Sustaining CBNs:** The sustenance of CBNs should be of concern to policy makers. Some initiatives to ensure sustainability includes:
 - a. **Regulating the Cost of Interconnecting the Community Networks to the Gateway:** This can be done either via the public subsidization of the cost of interconnecting to the gateway for the CBNs. This will reduce the monthly or yearly operational cost of running the network.
 - b. **Public Financial Incentives:** This could be subsidies, loans or grants. This will subsidize the CBN's capital expenditure greatly.

In conclusion, this article does not imply that CBNs are either the best or only way by which Broadband connectivity can be extended to rural areas. Rather, this article implies that, since communities have the ability to facilitate such networks, then it is important to recognize them by developing e-government services that is of use to them. If this is done, and the communities are made to see that they can facilitate the network to serve the need. They will do so.

REFERENCES

- Aboul-Magd, O., Kim, J., Wentink, M., Yang, D., & Stacey, R. (2013). *Status of Project IEEE 802.11ac*. IEEE.
- Calista, D. J., & Melitski, J. (2007). E-Government and E-governance: Converging constructs of Public Sector Information and Communication Technologies. *Public Administration Quarterly*, 31(1/2), 87–120.
- Campbell-Kelly, M., & Aspray, W. (1996). *A History of the Information Machine*. Basic Books.
- Chairunisa, F., Malik, T., & Rahman, A. (2016). Contextual-participation E-government Implementation Model: A case study of the Office of Demographic and citizen data records of Makassar. *Mediterranean Journal of Social Sciences*, 7(1), 345–353.
- Evoting Estonia. (2016). *Independent Report on Evoting in Estonia*. Retrieved 03 17, 2016, from Estonia E voting: <https://estoniaevoting.org/>
- Evoting Switzerland. (2016). *Electronic voting in Switzerland*. Retrieved 03 17, 2016, from <https://www.ch.ch/en/elections2015/electronic-voting/electronic-voting-in-switzerland/>
- Feijoo, C., Gomez-Barroso, J., & Bohlin, E. (2011). Public Support for the deployment of Next Generation Access Networks -Common themes, methodological caveats and further research. *Telecommunications Policy*, 2011(9-10), 791–793. doi:10.1016/j.telpol.2011.08.002
- Finkelievich, G., & Kisilevsky, G. (2005). Community Democratization of Telecommunications Community Cooperatives in Argentina: The case of TELPIN. *The Journal of Community Informatics*, 1(3).
- Ho, A. T. (2002). Reinventing Local Governments and the E-government Initiative. *Public Administration Review*, 62(4), 434–444. doi:10.1111/0033-3352.00197
- ITU. (2015). *ICT Facts and Figures*. ITU.
- Kakekaspan, M., O'Donnell, S., Beaton, B., Walmark, B., & Gibson, K. (2014). The First Mile Approach to community services in Fort Severn First Nation. *The Journal of Community Informatics*, 10(2).
- Kettani, D., & Moulin, B. (2014). *E-Government for Good Governance in Developing Countries: Empirical Evidence from the eFez Project*. London: Anthem Press.
- Lindskog, H., & Johansson, M. (2005). Broadband: A municipal Information Platform: Swedish experience. *International Journal of Technology Management*, 31(1/2), 47–62. doi:10.1504/IJTM.2005.006622
- Norris, D. F., & Kraemer, K. L. (1996). Main-frame and PC Computing in American Cities: Myths and realities. *American Society for Public Administration*, 56(6), 568–576.
- OECD. (2008). Broadband Growth and Policies in OECD countries. OECD Ministerial Meeting on the future of the internet economy. Seoul, Korea: OECD.
- Pintor, R. L., Gratschew, M., & Sullivan, K. (2002). *Voter Turnout Rates from a comparative perspective*. Voter Turnout.
- Salemink, K., & Bosworth, G. (2014). *Investigating community-led broadband initiatives as a model for neo-endogenous development*. Lincoln University UK.
- Steuer. (2016). *German Income tax calculator*. Retrieved 3 17, 2016, from Parmentier: <http://www.parmentier.de/steuer/incometax.htm>
- TDPS. (2016). *Texas Department of Public Safety*. Retrieved 03 17, 2016, from Texas Department of Public Safety: <https://www.txdps.state.tx.us/DriverLicense/>
- UN. (2014). *United Nation E-Government Survey 2014*. New York: United Nations.

Wachira, D., & Arlikatti, S. (2010). Challenges of Effective E-governance: Problems of Transparency, Infrastructure and Connectivity in Kenya. In C. G. Riddick (Ed.), *Comparative E-government* (pp. 259-273). Springer.

Williams, I. (2015). *Analysis of Public Private Interplay Frameworks in the Development of Rural Telecommunications Infrastructure: A Multi-Case Study*. Aalborg: Aalborg University Press.

Williams, I., Adjin, D. M., & Tsivor, K. (2014). The Impact of PPIs in the Development of Broadband and the way Forward: Ghana's case. *Communications and Strategies*, (91), 81 - 102.

Williams, I., & Falch, M. (2012). Public Private Partnerships and Next Generation Networks. In A. M. Hadjiantonis & B. Stiller (Eds.), *Telecommunication Economics, LNCS 7216* (pp. 100-107). Berlin: Springer Berlin Heidelberg. doi:10.1007/978-3-642-30382-1_14

Williams, I., Kwofie, B., & Sidii, F. S. (2016). Public Demand Aggregation as a means of Bridging the ICT Gender Divide. In J. Wilson & N. D. Gapsiso (Eds.), *Overcoming Gender Inequalities through Technology Integration* (pp. 123-143). Hershey, PA: IGI. doi:10.4018/978-1-4666-9773-7.ch006

Zambrano, R., & Seward, R. K. (2013). *From Connectivity to Service Delivery*. UNDP.

ADDITIONAL READING

Acilar, A. (2011). Exploring the aspects of digital divide in a developing country. *issues in Informing Science and Information Technology*, 8(2011), 231-244.

Annan, N. K., Ofori-Dwumfou, G. O., & Kwofie, B. (2014). Mobile ICT and Education Delivery. In K. E. Skouby & I. Williams (Eds.), *The African Mobile Story*. Aalborg: River Publishers.

Ayeni, A. (2002). *Rural Telephony as a means of poverty Alleviation*. UniIlorin Publications.

Baemish, A. (1995). *Communities On-line: Community Based Computer Networks*. Massachusetts: Massachusetts Institute of Technology.

Bernhard, C. I. (2014). Local e-Government in Sweden - Municipal Contact Centre Implementation with Focus on Public Administrators and Citizens. *The Journal of Community Informatics*.

Igari, N. (2014). How to successfully promote ICT usage: A comparative analysis of Denmark and Japan. *Telematics and Informatics*, 31(1), 115-125. doi:10.1016/j.tele.2012.10.001

West, D. (2007). *Global E-Government, 2007*. Inside Politics.

Williams, I. (2014). *Legislative Framework Required for Africa's M-economy*. UK: Technology Banker.

KEY TERMS AND DEFINITIONS

Broadband: A telecom network whose data rate is at least 256kbps.

Community Broadband Networks: A community that facilitates Broadband networks.

Community: A social unit bound by a common purpose or characteristics.

Convergence: The delivery of telecoms and broadcast services on one service platform.

E-Applications: Web applications providing access to e-government services.

E-Government: The delivery of government services to citizens via the Internet.

E-Services: Government services delivered online.

New Media: The delivery of broadcast and interactive media platforms over the Internet.

Critical Success Factors in E-Democracy Implementation



Aderonke A. Oni
Covenant University, Nigeria

Adekunle O. Okunoye
Xavier University, USA

INTRODUCTION

The use of Information and Communication Technology (ICT) in enhancing citizens' political participation has been identified as a solution to the problems of representative democracy, particularly, the disconnection between representatives and citizens and the decline of political interest amongst the populace (OECD, 2003b; Polat, 2005; Backhouse, 2007; Kang and Dugdale, 2010). One of the key issues in e-democracy development is to acquire an e-democracy system that considerably meet the needs of the citizens to participate in the democratic process and the government's needs to provide citizens with adequate participation channels (Funikul and Chutimaskul, 2009).

A successful e-democracy implementation should target developing a system that will meet the government's needs and provide citizens with adequate participation channels. There have been several cases of e-participation projects initiated by different actors around the world but the issue of sustainability and citizens acceptance for online public participation remains a difficult task (Panagiotopoulos and Al-Debei 2012; Saebo et al., 2008). As noted by Wouters (2008), examples of good practice are extremely rare in e-democracy implementation. Most e-democracy implementation has experienced mixed success (Coleman and Norris 2005; Blackhouse 2007), some have failed to meet up with the demand of the dedicated advocates (Oni et al., 2014) while some are battling with lot of replication of efforts

within countries driving full fledged e-participation across their governmental bodies. Several scholars have also worked on having a lasting e-democracy implementation, some notable efforts are found in Black and Noble (2001), Clift (2004), Local E-Democracy National Project (2006), Funikul and Chutimaskul (2009) and e-democracy project planning route map. Ensuring success in e-democracy implementation requires harmonizing the technological, economical, political, legal, cultural issues pertaining to e-democracy. As noted by Kotsiopoulos (2009) barriers to greater online citizen engagement in policy-making are not technological but cultural, organisational and constitutional.

This article will provide an in-depth description of key issues to consider in making participatory e-democracy user-friendly, effective and deliver the expected outcomes. Important issues pertinent to the success of e-democracy project were explored. These are technological, social and political issues that make for good success in e-democracy implementation. These are beyond the traditional concerns of the digital divide.

BACKGROUND

Democracy is a system of government in which the power is vested on the people and the will of the qualified majority rules. Chi-Ha (1977) further describes democracy as an ideology opposed to silence, a system that respects a free logos and

freedom of speech and encourages the cacophony of dissent. Combining the words “Electronic” and “Democracy” according to Caldow (2004) means the use of ICT tools to facilitate, improve and extend democratic activities. Cliff (2000), e-democracy is considered as the use of information and communication technologies and strategies by democratic actors (governments, elected officials, the media, political organizations, citizen/voters) within political and governance processes of local communities, nations and on the international stage. E-democracy is, therefore, anything that governments do to facilitate greater participation in government and enhance effective governance using digital or electronic means (Colman and Norris, 2005). Hye, Jong and Hae (2008) gave a more explicit definition as they defined e-democracy as the use of cyberspace and mobile technologies to enhance effective governance. Macintosh (2004) referred to e-democracy as the use of information and communication technologies to engage citizens, support the democratic decision-making processes and to strengthen representative democracy. From these definitions, three major elements of e-democracy can be identified which are people, technology and political process or culture.

The people are the democratic actors including governments, elected officials, the media, political organizations, citizen, civil society groups, etc. Technology includes the various ICT infrastructure and e-participation tools used to facility online political participation such as e-consultation, e-petition, e-panels, etc. Political culture is the distinguishing beliefs, values, attitude, habits, and behavioural patterns that characterized a political community. It is the common perception of the right and obligations of citizens and rules for participating in political process.

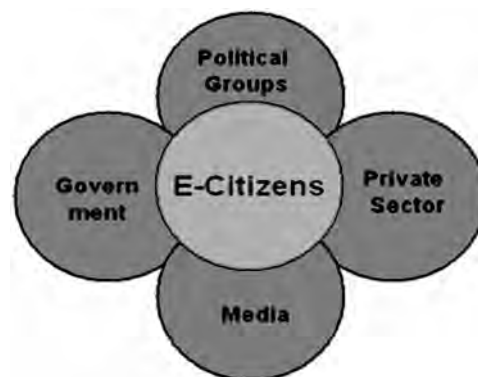
According to Caldow (2004), e-democracy has two sides: the tactical side and the strategic side. The tactical side deals with the use of information technology to advance communication and promote access to information while the strategic side considers how government can use digital media to actively engage citizens and advance

its public policies to the global community. This includes the use of e-democracy initiatives to engage citizens in democratic processes such as policy deliberation, information enquiry, petition against corrupt practice, and expression of opinions through forums (Coleman and Gotze, 2001; Funikul and Chutimaskul, 2009).

One of the key issues in e-democracy development is to acquire an e-democracy system that considerably meet the needs of the citizens to participate in the democratic process and the government's needs to provide citizens with adequate participation channels (Funikul and Chutimaskul, 2009). A successful e-democracy implementation should target developing a system that will meet the government's needs and provide citizens with adequate participation channels. The key characteristics of e-democracy are better service with the appropriate access time, reasonable cost of utilizing suitable ICT, responsiveness of government in listening, and support of citizens' participation (Funilkul and Chutimaskul, 2009; Blumler and Coleman, 2001).

Over the last decade, academics have been exploring ways to effectively utilize technological tools to support democratic process. Literature reveals different models of e-democracy that can help governments and developers structure e-democracy implementation. Among such include modes from the following bodies of knowledge:

Figure 1. E-democracy conceptual model
Clift, 2003.



Clift (2003), proposed a conceptual model of e-democracy as depicted in Figure 1. There are five elements that constitute the model: government, e-citizens, media, political groups, and media.

According to Clift (2003), the government provide extensive access to information and interact electronically with citizens. The political groups run online advocacy campaigns and online political parties' campaign, and the media and portal/search sites play a crucial role in providing news and online navigation. The "Private Sector" is commercially driven connectivity, software, and technology. In the model citizens are those who experience e-democracy as a whole (Clift, 2003). E-Citizens are individuals who use ICTs to participate in democratic process. This participation includes interaction with social groups, government agencies, media and private sectors. It also involves the ability to create and disseminate information, and demand for a more open and democratic society and/or organize social action in form of e-petition, political debates and policy dialogues.

Institute for Electronic Government (IEG) also proposed a four quadrant e-democracy model, which advanced/strategies the definition and implementation of e-democracy (Caldow, 2004). The model explores how government entity (representatives, political parties, legislative bodies, national and international organizations, etc) can successfully interpret and respond to the digital world and explore technology to advance influence (Caldow, 2004). The IEG model (Caldow, 2004) can help government determine its current state of e-democracy implementation and identify technological initiatives to undertake so as to move to the next level. The vertical axis of the model measures the degree of engagement while the horizontal axis measures influence.

The first quadrant discussed the early stage of e-democracy implementation which is characterized by passive and one-way asynchronous communication. E-democracy implementation at quadrant two of IEG model can help to achieve

two-way communication but is still asynchronous in nature (Caldow, 2004). The third quadrant describes advanced political entities collaboration. Although communication is still asynchronous, collaboration begins to emerge. Activities at this stage mostly concern political players and electoral process. Some of the activities include recruiting online volunteers, online fund-raising, campaigning, communication with media and constituencies. The fourth quadrant tagged 'Domestic Citizen Engagement' presents the most sophisticated implementation of e-democracy, which should be strategic, interactive, synchronous and global in nature. The crust of this quadrant is the advocacy for early integration of public input into policy making cycle so that public interests can influence policy outcomes. E-democracy implementation at this stage supports the Hansard Society's argument for the development of mechanisms for promoting public deliberation, embedding it within the constitutional process and demonstrating real links between public input and policy outcomes (Coleman and Hall, 2002; Caldow, 2004).

Macintosh (2004) identified key dimensions that are needed to characterize online political engagement initiatives. The dimensions include; level of engagement, political actors, technologies, rules of engagement, duration and sustainability, accessibility, evaluation and outcomes and critical factors for success. The critical factors for success dimension according to Macintosh (2004) is for researchers to capture political, legal, cultural economic or technological issues that stand out to make e-participation a success. Kotsiopoulos (2009) also argued that barriers to greater online citizen engagement in policy-making are cultural, organisational and constitutional and not technological.

In the next section, this article discusses in depth those various issues governments and researchers implementing e-democracy should be considering for a sustainable implementation.

CRITICAL SUCCESS FACTORS IN E-DEMOCRACY

In this section, we present a discussion on a diverse set of political, legal, cultural, economic or technological issues that governments and researcher need to consider in e-democracy implementation. Success factors in e-democracy implementation is divided into three (3) categories: Technological, Social-Cultural, and Political.

Political Factors

E-democracy is not all about technology; technology is just an enabler Kotsiopoulos (2009). This sub section highlights the importance of political and constitutional issues in ensuring successful and sustainable e-democracy implementation

Political Will: E-democracy will succeed where there is political power, will and leadership to make it work effectively by introducing the constitutional support and structural changes needed to integrate citizens' input into public decision making. This will include structural changes, constitutional amendment or procedural reform to support governments' effort on e-engagement initiatives.

Constitutional Change and Support: Democracy as a system of government is heavily influenced by institutional context and human agencies. In same vein online democratic initiatives need to have constitutional backing to gain acceptance across all level of government and enforce implementation. Integrating technology drive initiatives democratic processes usually requires structural changes and procedural reform involving development of e-democracy policy, formalization of the status of e-democratic tools (Millard et al., 2012) by integrating them into the decision making processes (Oni et al., 2014). Legal protection of the individual's rights of citizens to express themselves without fear, articulation of e-democracy policies and goals with national development programs and budget and legislative frameworks that guide implementation across all

levels of government. Clift (2003) said that "if e-democracy is not part of what is evaluated or budgeted, then the administrative and resource priorities within agency e-government efforts will not likely address the e-democracy responsibilities of governments".

Social-Cultural Factors

- **Publicity and Promotion:** Presently the most prominent promotional services in e-democracy platforms are link to social media site such as Facebook, Twitter, Flickr and Really Simple Syndication (RSS). This is good but limited to recruiting only active social media users. E-democracy requires diverse means of advertisement and publicity. Establishing e-participation platforms does not automatically imply that citizens will use them. A combination of off-line and online promotional platforms will help give citizens quick awareness. Off-line promotional avenues, such as bill boards, press releases and television and radio adverts should clearly give the electronic web platforms publicity (Macintosh, 2004). The online promotional routes such as web Ads, "tell a friend" postcards and clickable logos advertising the e-participation channels on related websites should complement the off-line modes.
- **Integration of Online and Offline Political Engagement Channel:** Technological tools should be made to align with established best practice in the democratic culture. Technology itself is not a solution to disengagement but has potential to ameliorate democratic apathy through adequate implementation. Integration of online engagements channels with traditional, "off-line" tools for citizens' participation in policy-making is important to derive the full benefit of ICT investment in democracy (Kotsiopoulos, 2009).

- **Online Feedback:** The essence of e-participation is to widen the level and volume of citizens' contribution in public decision making. Seeking and accumulating the opinion of people is not enough, it is important for them to have knowledge of what has become of their contribution and effort in participation. Appropriate feedback on citizens' input at any level of e-engagement can contribute to the overall transparency, accountability and openness of government and provide more motivation to participate (OECD, 2003). The feedback processes need to be timely, comprehensive, personalised (if need be), secure and credible. Providing appropriate feedback mechanism as part of the e-Engagement system will provide citizens with a clear idea of what they can expect in terms of government feedback to their input, as well as the degree to which their input is considered in the development of government policy (OECD, 2003). Feedback can be provided both during and after the consultation.
- **Resource and Investment:** E-democracy often requires huge financial investment and does yield little or no financial return. It is important to understand the financial implications of using ICTs to support political participation right from planning stage.
- **Moderation:** Moderation is important to effectiveness of online democratic participation. This involves determining the role(s) of each group of democratic actors and putting in place comprehensive rules to guide the actions of each individual in the system. The facilitators/ moderators must be seen as a reliable, non-partisan source of legitimacy to the whole process and establish measures to safeguard the integrity of participants.

Technological Factors

G

Beyond consideration for appropriate online tools to use to engage citizens in online public deliberation, there are other key technological issues to consider about online public participation user friendly, successful and sustainable. Technological issues such as accessibility, usability, reliability, security, authentication, interoperability need to be given due consideration right from the planning stage of e-democracy for acceptance and sustainability.

Accessibility: There is general accessibility guideline for web sites which include design for inclusion (www.W3.org/WAI) Accessibility in online engagement also includes having access to ICT tools and being able to get specific piece of government information. Accessibility issues ranges from 'digital divide' to 'design for all' (Coleman and Gotze, 2001). These must be critically considered when designing e-participation system. In as much as one may not be able to include everyone, e-participation systems designer need to be careful not to deliberately or by omission exclude any group of citizens. Government led initiatives also need to have policy guideline relating to accessibility so as to create checks and balances for designers.

Usability and Utility: Usability has to do with how easy the users find online engagement web user interface while utility refers to the system's design functionality. Usability and utility are equally important for the survival of online engagement systems as citizens need to find it useful and easy to use. Taking the role of ICT in governance and political participation beyond information provision to active participation both at grass root and national level will increase usefulness of the system. On the average, online engagement system should satisfy four basic usability criteria which are efficiency, intuitive, support and satisfaction (Coleman and Gotze, 2001).

1. **Efficiency:** This involves giving users a clear understanding of their work on the site, being able to keyboard to perform task on the site and fast response rate for good flow.
2. **Intuitive:** This can be described as the interactive behavior of online engagement channels. That is, consistence and behavior of every part of the system. This is important to helps users to easily establish proficiency each time they logon.
3. **Support:** Human being are liable to make mistakes, a good system design should help users to discover their error early and to easily undo. Online engagement channels should include intelligence to provide advice/reference materials.
4. **Satisfaction:** Users need to feel in control and experience some pleasantness while using e-participation channels.
5. **Content Clarity and Easy Navigation:** This is another usability criteria that e-participation systems must satisfy. This is very important because of the enormity of information on e-participation websites. If the website is difficult such that users cannot easily locate the right information or appropriate function, they will leave the web site no matter how usefulness it is. If the homepage fails to clearly state the various functions that users can perform on the site, people leave. It therefore connote that content of each page should be kept simple, clear and description of the function of the page. Involving users in system test early in the design process will to avoid difficulty after in content clarity and after deployment.
6. **Security:** This involves guide against threats such as hackers, worms and virus. Every on-line activity is prone to these attacks but the risk of getting hit by whatever kind of attack varies based on implementation platform and security mechanisms built into the system (Coleman and Gotze, 2001). E-engagement system are prone to more deliberate attacks ranging from terrorism to simple criminality (Coleman and Gotze, 2001) because of the various kinds of political activism that are involved in policy deliberation. Security and privacy issues to consider in e-participation begin with ensuring privacy of users' data. Electronic security requirements in e-participation systems include: protection of users' private data such as phone number and residential address, secure user's navigation, software protection and document security. As suggested by Parajuli (2009), government web sites should contain privacy statements as a proof of government responsiveness to protect citizens' information privacy. It is important to implement security and privacy measures in e-participation systems so that users can trust the system (Zissis et al., 2009). However, ensuring security should not cause trade off in transparency.
7. **Interoperability:** Electronic democracy entails the use of various channels of communication ranging from simple e-mails to complex online deliberation tools such as e-forum, e-petition, e-Panels etc (OECD, 2003; Zissis et al., 2009). To avoid replication of efforts among government bodies implementing e-democracy, it is important to establish efficient communication among software modules and web servers for resource sharing. Through interoperability, government agencies information and citizens' data can become a central reference for all online democratic processes within across the country. As suggested by Coleman and Gotze (2001) various online engagement channels running different platforms can speak to each other in a technical sense using XML-RPC to share resources among web servers and applications.
8. **Personalization and Customization:** Capturing and storing data on users' preference to provide recommendation and customization for future visit has been a viable tool to assist and retain online shoppers in e-commerce. In e-participation, making

users feel connected to the system through mechanisms such as personalization and customization will further appeal to their sense of engagement. This also involves defining various levels of membership which in turn determines the kind of conversations that each user can be involved. Giving users opportunity to set profiles and preferences the way they like. Also, given users subscription options such as notification of event and link to external social media site are attractive way of popularizing political discussion to promote political participation.

CONCLUSION

Active engagement of citizen in democratic process through ICT can reduce the disengagement between the representative and the electorate however, researchers, practitioners and government should have adequately consider the various of factors that will impact the success of e-democracy right the planning stage. This research article presents a collection of importance issues related technological factors, social factors and political factors in e-Democracy development and best practices in its implementation. It suggests that government agencies and researchers should take a more inclusive approach to evaluating e-democracy implementation. They are key issues to be considered to make participatory e-democracy user-friendly, effective and deliver the expected outcomes. These are technological, social and political, issues that make for good success in e-democracy implementation. These are beyond the traditional concerns of the digital divide.

REFERENCES

- Backhouse, J. (2007). e-democracy in Australia: The Challenge of Evolving a Successful Model. *The Electronic Journal of E-Government*, 5(2), 107–116.
- Black, A., & Noble, P. (2001). *E-Democracy around the World: A Survey for the Bertelsmann Foundation*. Charleston, SC: Phil Noble and Associate.
- Caldow, J. (2004). *E-democracy: Putting Down Global Roots*. *Institute of Electronic Government*. IBM.
- Clift, S. (2003). *E-Democracy*. E-Governance and Public Net-Work.
- Clift, S. L. (2004). *E-government and Democracy: Representation and Citizen Engagement in the Information Age*. Retrieved Nov., 2010, www.publicus.net
- Coleman, S., & Gotze, J. (2001). *Bowling Together: Online Public Engagement in Policy Deliberation*. Hansard Society.
- Coleman, S., Hall, N., & Howell, M. (2002). *Hearing Voices: The Experience of Online Public Consultations and Discussion in UK Governance*. London: Hansard Society.
- Coleman, S., & Norris, D. F. (2005). *A new agenda for e-democracy*. Oxford Internet Institute Forum Discussion Paper No. 4.
- Funilkul, S., & Chutimaskul, W. (2009). The Framework for Sustainable eDemocracy Development. *Transforming People, Process and Policy*, 3(1), 16–31. doi:10.1108/17506160910940713
- Hye, K., Jong, K. & Hae, L. (2008). A Study on the Public Sphere of Mobile Media. *Journal of Korea Information and Communications Society*, 31(4B).
- Kang, H., & Dugdale, A. (2010). A Comparative Study of e-Civil Participation in Australia and South Korea: The Case Study of GetUp! in Australia and the 2008 Candlelight Protest in South Korea. *Proceedings of 6th International Conference on e-government*.
- Kotsiopoulos, I. (2009). *Bringing Together and Accelerating e-government Research in the EU*. *ICT for Government and Public Service Unit*. European Commission.

Macintosh, A. (2004). Characterizing E-Participation in Policy-Making. *Proceedings of the 37th Hawaii International Conference on System Sciences*. Computer Society Press.

Millard, P., Millard, K., Adams, C., & McMillan, S. (2012). *Transforming Government through e-Participation: Challenges for eDemocracy*. European Conference on e-Government (ECEG) 2012, Barcelona, Spain.

OECD. (2003a). *Engaging Citizens Online for Better Policy-making*. Retrieved Dec., 2010, www.oecd.org/dataoecd/62/23/2501856.pdf

OECD. (2003b). *Promise and Problems of E-Democracy: Challenges of Online Citizen Engagement*. Retrieved Dec., 2010, from www.oecd.org/dataoecd/9/11/35176328.pdf

Parajuli, J. (2007). A Content Analysis of Selected Government Web Sites: A Case Study of Nepal. *The Electronic Journal of E-Government*, 5(1), 87–94. Available online at www.ejeg.com

Polat, R. K. (2005). The Internet and democratic local governance: The context of Britain. *The International Information & Library Review*, 37(2), 87–97. doi:10.1080/10572317.2005.10762669

Web Accessibility Initiative (WAI). (n.d.). Retrieved from <http://www.w3.org/WAI/>

Zissis, D., Lekkas, D., & Papadopoulou, A. E. (2009). Competent Electronic Participation Channels in Electronic Democracy. *Electronic Journal of E-Government*, 7(2), 195–208.

KEY TERMS AND DEFINITIONS

E-Engagement: The use of information and communication technologies (ICTs) in supporting information, consultation and participation.

Interoperability: The capacity to move data between different software applications.

Online Consultation (E-Consultation): A form of electronic deliberation or exchange between government and citizens using the Internet.

Personalization: The process of tailoring web pages to individual users' characteristics or preferences.

XML-RPC: A set of implementations that allow software running on disparate operating systems, running in different environments to make procedure calls over the Internet (<http://xmlrpc.scripting.com/defaulty.html>).

E-Activism Development and Growth



John G. McNutt

University of Delaware, USA

Lauri Goldkind

Fordham University, USA

INTRODUCTION

The use of sophisticated technology to promote social change has developed over the past three decades from tentative beginnings to an expected part of the arsenal of movement organizations and advocacy groups. The development of practical politics throughout the world has made greater use of ever more sophisticated technologies. This article will discuss the nature of e-activism, the development of electronic social change activities, the organizational and practice issues, the research base and the potential future developments in the field.

BACKGROUND

For the purpose of this review E-activism is defined as the use of technology tools by activists for addressing policy issues and social problems. E-activism is also called Cyberactivism (McCaughey & Ayers, 2003), Cyberadvocacy (Bennett & Fielding, 1999), Electronic Advocacy (West & Francis, 1996; McNutt & Boland, 1999), Cyberprotest (Van De Donk, Loader, Nixon & Rucht, 2004), Liberation Technology (Diamond, 2010) and Digitally Enhanced Social Change (Earl & Kimport, 2011). The important components of the practice are that it is technology enhanced, issue oriented and used by activists for policy change. E-activism as a strategy itself is issue neutral, rather it is a constellation of tools which may be applied to any social issue and similarly, it is also value and morality neutral, meaning champions

on either side of an issue might employ the same strategy or tool to achieve radically different ends.

E-activism is strongly related to other concepts such as Cyber campaigning and Electronic Democracy, but there are important differences. Partisan political campaigning refers to efforts to change office holders, while E-activism looks at changing issues or problems. E-democracy (also e-participation and civic technology) often refers to the part of e-government that encourages citizen participation and involvement. The dividing line between these activities is often indistinct. The growth of civic technology has further complicated these already faint distinctions.

The techniques that e-activism uses to address issues or problems are often combined with more traditional methods used by advocacy groups and interest organizations. These traditional methods include community organizing, lobbying, administrative advocacy, petition campaigns, lawsuits and so forth. While less visible than these intervention tools, social change efforts have always been dependent on research and information gathering activities. Within the traditional advocacy arena, there is a well-established toolset for these activities that can be enhanced or replaced by technology tools.

Activists can combine community organizing, demonstrations, lobbying and electoral strategies with e-mail campaigns, mobile notifications using push technologies or short message services (SMS), social media efforts and sophisticated data analysis. Campaigns can also be waged completely online. This creates a situation where one may have online only efforts (pure e-activism), hybrid efforts

DOI: 10.4018/978-1-5225-2255-3.ch310

using a mix of technology tools and traditional social change tools and finally, efforts which are nearly completely traditional with small amounts of embedded technology.

E-Activism is used by a wide range of organizations in a variety of situations. These include traditional advocacy organizations, social movement organizations, political organizations and other types of associations and organizations. There is some evidence that the growth of technology and its capacities to reduce transaction costs have promoted the growth of virtual advocacy organizations and leaderless organizations (Earl & Kimport, 2011; Brainard, Boland & McNutt, 2012). Recent experience with the U.S Tea Parties, Occupy Wall Street, BlackLivesMatter and the Arab Spring Demonstrations appears to support this idea. In any case, technology is moving many social change organizations away from those described in the political science literature on interest groups or the sociological literature on social movement organizations. One illustration is the role of socialization of movement actors. Some of the activists who use these technology enhanced tools are amateurs while many are highly skilled political operatives. Conventional wisdom in social movements was that people worked their way into leadership positions through long hours of work at lower levels. They then became qualified to lead movement groups and organizations. This is also reflected in the political participation literature in discussion about civic skills and the value of associations (Verba, Schlozman & Brady, 1995—See also Smith, Schlozman, Verba & Brady, 2009). The work of Earl and her colleagues demonstrates that technology can change this dynamic in important ways, creating the opportunity for new activists to conduct campaigns without previous experience (Earl & Kimport, 2011; Earl, 2007; Schussman & Earl, 2004).

Technology changes organizations in important ways. It makes them flatter and often changes the economics of productions. In terms of social change organizations, it can minimize the need for

the bricks and mortar facilities that older organizations found essential and makes distributed work possible. This creates issues for theories, such as resource mobilization theory in sociology, that assume that these facilities are essential and the activities (such as fundraising) to support them are crucial. The rise of the so called sharing economy may push this further as activists become familiar with technology led collaboration. Scholars will have to reconcile these issues as knowledge building goes forth and new paradigms for organizing work are coming on line based on the ability of connecting workers to consumers via technology without a centralized or organizing core.

Increasingly technology is essential to E-activism, but it frequently requires a set of techniques to make the technology useful in political situations. While there are tech tools specifically written or developed for political applications, more often, activists use technology developed for another reason. It then becomes the task of a thoughtful person to adapt the technology to the new use. Sometimes this means modifying the technology in some fashion but usually it means changing the way it is used. This might be thought of as a new technology in its own right.

The Evolution of E-Activism

Many people think that technology in activism evolved in the past few years. While it is true that the growth of this practice grew quickly in the recent past, there were efforts in the 1980s that blended technology to social change activities (Downing, Fasano, Friedland, McCollough, Mizrahi & Shapiro, 1991; Schuler, 1991; 1996). Most of the technology that was used during this early period would be considered primitive by the standard of today's cutting edge efforts. These included Bulletin Boards, newsgroups, e-mail and early mapping systems. The overwhelming majority of the technology used was developed for some other purpose. It should be noted that many potential users did not have access to the

Internet or other technologies at this point and those that did found a very different situation than we have today.

The growth of technology, along with the accumulated experience of the social change community, led to more capable and robust efforts. They used the emerging World Wide Web, sophisticated e-mail efforts and the beginnings of online fundraising and online petitions. This phase also saw the development of advocacy oriented technology. In the related area of political campaigning, major political parties started using technology in earnest. While it paralleled their face to face and mass media strategies, technology had made a foothold in the land of partisan political campaigning. This technology augured well with the managed politics that placed control of campaigns in the hands of political professionals and consultants. It allowed for a high degree of message control, a touchstone of both electoral and issue advocacy at the time.

In the early part of the last decade there was change afoot. New technology, often called Web 2.0 or Social Media (Germany, 2006; Bryant, 2006; Madden & Fox, 2006) began to develop a larger following in both society at large and the political system. These techniques promote user generated content, the development of collective intelligence, networking and a high degree of interactivity. They were a poor fit with message control and professionalized politics. There was some experimenting, however, with new forms of campaigns. The Howard Dean Campaign in 2004 experimented with a variety of Web 2.0 tools including Blogging, Meetup and computer gaming (Trippi, 2004; Cornfield, 2004; Teachout & Streeter, 2008). While Dean eventually lost, his campaign demonstrated how these new technologies could be used. In 2008, Barak Obama took what Dean had learned and developed a campaign that ended in victory. His technology actively involved supporters in his campaign, rejecting the logic of a more managed campaign. Other campaigns throughout the world moved toward

this approach (Davis, 2005; 2010). These developments in cyber campaigning were complemented by a similar evolution in issue advocacy and social movements. It is fair to say that diffusion of techniques from various sectors is frequent and quick. Specialist in digital campaigning are emerging in almost all arenas.

The wide scale introduction of wireless technology joined with new software developments and innovative ways to use the technology have changed the landscape of activism. The emergence of Occupy Wall Street (Chen & Pirolli, 2012; Gaby & Caren, 2012) and the revolutions in Northern Africa and West Asia, often called Arab Spring (Lim, 2012; Hamdy & Gomaa, 2012; Tufec & Wilson, 2012), attest to the power of e-activism. These face to face movements were fueled by the power of technology.

Technology continues to evolve and activists continue to make better use of the tools that they have. The use of Big Data, predictive analytics and Data Mining, applied to selecting potential supporters and contributors, represents a major thrust. The further development of mobile technologies and the integration of social media approaches top mobile technology promises even larger gains. Finally, in organizations all around the world dedicated men and women are building new ways to use technology in the service of social change.

How E-Activism Works

The practice of e-activism varies from setting to setting, but there are commonalities that tend to be reflected across the board. In general, e-activism practice can be reflected as a series of four processes: Advocacy Research, Informing the Public, Coordinating and Organizing and finally, Applying Pressure. These four processes play out in different patterns within an e-activism effort. This is not to imply that there is a top down process but only that activities tend to coalesce around a series of activities.

- **Advocacy Research:** Activism may be fueled by commitment and range but it also runs on information. This includes information on policies and problems, information on opponents (Opposition Research), information on potential supporters or voters and a variety of other data needs. This information has always been collected on behalf of activist causes, often by hand. The growth of technology makes gathering all of this information easier to collect, store and analyze. In recent years, the ability of activists to use sophisticated data analysis has taken this process to a new level. Data mining, network analysis and sophisticated statistical analysis techniques, combined with advanced data visualization and mapping techniques are available to many social change groups that could not have dreamed of such capacities a few years ago. Databases developed for marketing purposes bring new sources of information about voters and potential supporters. People who are likely to support a cause can be identified using a variety of new tools. What this means is that activists can now enjoy an enriched information environment that can support the other three processes in a substantial fashion.
- **Informing the Public:** This process includes both making the public aware of a problem or issue and educating the public and supporters about the nature of the concern. The growth of technology means that activists can design a wide variety of message using multiple sources (such as video, text, voice, pictures and so forth) and targeted to individual recipient characteristics. The information developed by the advocacy information function can be readily applied to both message design and selection of recipients.

At a basic level, e-mail, websites, social media (particularly Social networking sites, blogging and

microblogging), discussion lists and so forth are frequently used (McNutt & Barlow, 2010). Also used are online videos (such as YouTube) and static pictures. The Invisible Children's video Kony2012, for example, targeted an African Warlord and drew a substantial audience. Online mapping is useful where there is a spatial issue. This last technology has been especially useful to environmental advocates who can use it to display proximity to natural hazards. Also useful is e-learning technologies such as Moodle and teleconferencing systems.

- **Organizing/Coordinating:** The activity needed to focus social action can be time and effort intensive, which limits the size of the activity. In addition, the impact of distance is to make the process both harder and less effective. These challenges are serious for domestic activism and represent major barriers for international activism. One might consider these the transaction costs of organizing.

Technology allows us to reduce these costs (Earl & Kimport, 2011) and extend capacity and reach. Early on, newsgroups and bulletin boards provided this technology help to activists. E-mail and websites were also helpful and the range of social media technologies added new dimensions (see Guo & Saxton, 2012). The impact of emerging mobile technologies is just now being felt and the potential is enormous. Some of those developments were seen in the streets of Egypt and Libya. Facebook and Twitter, combined with mobile technology, were certainly critical in organizing the face to face demonstrations. Involving people is central to activism, either in traditional campaigns or in cyberspace.

- **Applying Pressure:** Finally, technology can be used to pressure decision makers. Early approaches included e-mail campaigns, on-line petitions and Internet fax efforts. Later efforts combined social me-

dia and more sophisticated comprehensive advocacy suites. Disruptive efforts include hacking, client side denial of services attacks and theft of confidential data. While early campaigns were combined with face to face efforts there have been Internet only campaigns (Such as Move On's Million Mouse March against the US Invasion of Iraq). As the practice of e-activism developed, so did the e-democracy element of e-government. Decision makers began to use the Internet actively in both their personal campaigns and in engaging constituents. These two forces tax government technology. A 2005 report from the Congressional Management Foundation found that Congressional Offices in the United States were encountering ever more pressure online and were struggling to keep up with the level of activity, which can only have accelerated in the past decade.

These four processes play out in e-activism efforts in a variety of ways. As the context and problem change, different processes become dominant. The capacities that they signify represent a major move forward for the practice of activism. Not every activists organization can or will use these new capacities. Every organization, however, can make improvements in what information they gather and how they use and interpret that material.

Concerns and Controversies

There are a number of challenges that face e-activism as a practice. These include the digital divide, confusion with other activities, conflict with traditional approaches to social change, There is also the question of how these techniques are used and to what effect.

The Digital Divide: In the 1990s and early 2000s the digital divide most often referred domestically to groups of individuals without adequate resources for engaging with internet technologies (Ebo, 1998;). With the rise of

digital technologies embedded in telephones and national outreach and implementation of wireless in public spaces as well as educational institutions the newest iteration of the digital divide focuses on organizations. Nonprofits have often lagged behind the corporate and government sectors in using information and communication technology effectively (McNutt, 2008; Cordes & Rafter, 2007). This situation may be more pronounced in organizations that are small and underfunded, such as advocacy organizations. In some cases, it is clear that they do not have this capacity or do not utilize the technology adequately, and it is unclear what supports exist to help the sector to catch up and/or adequately implement such technologies. Collectively this set of circumstances is referred to by organizational scholars as the organizational digital divide (McNutt, 2008). The individual digital divide, coupled with the organization digital divide can limit the potential of activism to engage people and achieve meaningful results.

- **Confusion With Other Activities:** Over the years, there have abortive attempts to define e-activism as part of alternative journalism or part of virtual volunteering. These efforts, while well meaning, are conceptually and practically flawed. They seize on small parts of the overall phenomena and ignore the rest. At the same time, there are those who feel, for a variety of reasons, change efforts conducted online cannot be considered to lead to tangible change.

While blogging might be considered journalism and online video might fit into earlier conceptions of alternative media, it would be very problematic to consider the lobbyists, community organizers, campaign consultants and so forth that use these technologies in social change efforts as journalists. Much of the effort has little to do with informing the public as more traditional alternative social movement media once had.

Equally, while there are virtual volunteering opportunities in social change campaigns, many of the practitioners of E-activism technology are not volunteers. Engaging the public to participate in their government is not volunteering, nor is voting or jury duty.

Conflict with Traditional Social Change Advocates: Some supporters of traditional methodologies have argued that “real change only happens off-line” (see Kohn, 2008). Social critic Malcolm Gladwell (2010) made a similar argument, disparaging those who use online as opposed to offline activism. This is more of an ideological argument, as opposed to an empirical one. Earl and Kimport (2011) point out that empirical evidence to support this contention is generally lacking. There is little research to support the contention that tradition efforts are better. Realistically, many traditional activists have embraced the new technology.

While not really a controversy, it ought to be noted that advocacy groups working on pro-social change efforts are not the only ones to benefit from this new technology. Hate groups, terrorist organizations and criminal cartels can use the same e-activism technology in the same way as more positive organizations. Astroturf, deceptive efforts to mimic public participation, can also use this technology.

FUTURE RESEARCH DIRECTIONS

There is a substantial research base on e-activism and its related concepts. In addition, a growing body of research aimed at understanding how people interact with technology in a variety of settings is being developed. Social scientists in a variety of disciplines including political science, communications, sociology, public administration, social work, law and anthropology are working to help us understand these new phenomena. Some areas where new research and theory building include:

1. **Intervention Research:** More substantial research on the effectiveness of specific interventions. While there is a growing body of intervention research, so using experimental designs, more are needed. These new studies ought to reflect different target populations, problem areas and different levels of practitioner abilities.
2. **Research on the Interaction Between Governments and Advocacy Organizations in Cyberspace:** The interaction between government, politics and advocacy organizations in an online environment ought to be studied as a dynamic entity. This must go further than studies which look at acceptance and satisfaction of e-government efforts. This would allow more sophisticated modeling of online campaigns.
3. More research that seeks to use the findings of the existing research base to inform e-activism. Researchers in many disciplines have amassed a substantial pool of research on how people and groups engage in online environments. While most of this research does not directly involve e-activism, many of the findings could be applied to designing e-activism interventions if they could be synthesized.

CONCLUSION

We live in a world with a considerable number of threats to social justice and human dignity. The efforts of those who seek to redress these wrongs have always involved difficult tasks with long odds for success. As an effective new tool, technology can lighten those burdens and make success more likely.

E-activism has evolved from an experimental practice to a major set of tools in the arena of social change. The practice today bears little resemblance to what once was. The potential for future development is almost unlimited.

REFERENCES

- Bennett, D., & Fielding, P. (1999). *The net effect: how cyber-advocacy is changing the political landscape*. Merrifield, VA: E-Advocates Press.
- Brainard, L., Boland, K., & McNutt, J. G. (2012). *The Advent of Technology Enhanced Leaderless Transnational Social Movement Organizations: Implications for Transnational Advocacy*. Paper read at the 2012 ARNOVA Meeting, Indianapolis, IN.
- Bryant, A. (2006). Wiki and the Agora: Its organizing Jim, but not as we know it. *Development in Practice*, 16(6), 559–569. doi:10.1080/0961520600958165
- Chen, J., & Pirolli, P. (2012). Why you are more engaged: Factors influencing twitter engagement in Occupy Wall Street. *Proceedings of the sixth International AAAI Conference on weblogs and social media*, 423–426.
- Congressional Management Foundation. (2005). *Communicating with Congress: How Capitol Hill is coping with the surge in citizen advocacy*. Washington, DC: Author.
- Cornfield, M. (2004). *Politics moves online*. Washington, DC: Brookings.
- Cortes, M., & Rafter, K. (Eds.). (2007). *Nonprofits and Technology: Emerging Research for Usable Knowledge*. Chicago, IL: Lyceum Books.
- Davis, R. (2005). *Politics On-line: Blogs, chat rooms and discussion groups in American democracy*. New York: Routledge.
- Davis, R. (2010). *Typing politics: The Role of Blogs in American Politics*. New York: Oxford University Press.
- Diamond, L. (2010). Liberation Technology. *Journal of Democracy*, 21(3), 69–83. doi:10.1353/jod.0.0190
- Earl, J. (2006). Pursuing social change on-line: The Use of four protest tactics on the Internet. *Social Science Computer Review*, 24(3), 362–377. doi:10.1177/0894439305284627
- Earl, J. (2007). Leading tasks in a leaderless movement: The Case of strategic voting. *The American Behavioral Scientist*, 50(10), 1327–1349. doi:10.1177/0002764207300161
- Earl, J., & Kimport, K. (2011). *Digitally enabled social change*. Cambridge, MA: MIT Press. doi:10.7551/mitpress/9780262015103.001.0001
- Ebo, B. (Ed.). (1998). *The Cyberghetto or Cybertopia: Race, class, gender & marginalization in Cyberspace*. New York: Praeger.
- Fisher, M. (2011, March 26). In Tunisia, Act of One Fruit Vendor Unleashes, Wave of Revolution Through Arab World. *Washington Post*. Retrieved from http://www.washingtonpost.com/world/in-tunisia-act-of-one-fruit-vendor-sparks-wave-of-revolution-through-arab-world/2011/03/16/AFjfsueB_story.html
- Gaby, S., & Caren, N. (2012). Occupy Online: How Cute Old Men and Malcolm X Recruited 400,000 US Users to OWS on Facebook. *Social Movement Studies*, 11(3-4), 367–374. doi:10.1080/14742837.2012.708858
- Germany, J. B. (Ed.). (2006). Person to person to person: Harnessing the political power of on-line social networks and user generated content. Washington, DC: The Institute for politics, democracy and the Internet, George Washington University.
- Gladwell, M. (2010). Small change. *The New Yorker*, 4(2010), 42–49.
- Guo, C., & Saxton, G. (2012). Tweeting Social Change: How Social Media Are Changing Non-profit Advocacy. *Nonprofit and Voluntary Sector Quarterly*.

- Hamdy, H., & Gomaa, E. (2012). Framing the Egyptian Uprising in Arabic Language Newspapers and Social Media. *Journal of Communication*, 62(2), 195–211. doi:10.1111/j.1460-2466.2012.01637.x
- Kohn, S. (2008, June 30). Real change happens off-line: Millennials need to be activists face to face. *Christian Science Monitor*. Retrieved from <http://www.csmonitor.com/Commentary/Opinion/2008/0630/p09s01-coop.html>
- Lang, A. S., & Levitsky, D. (Eds.). (2012). *Dreaming in public: Building the occupy Movement*. Oxford, UK: New Internationalist Publications.
- Lim, M. (2012). Clicks, Cabs, and Coffee Houses: Social Media and Oppositional Movements in Egypt 2004–2011. *Journal of Communication*, 62(2), 231–248. doi:10.1111/j.1460-2466.2012.01628.x
- Madden, M., & Fox, S. (2006). *Riding the waves of "Web 2.0". backgrounder*. Washington, DC: Pew Internet and American Life Project. Retrieved from <http://www.pewinternet.org>
- McCaughey, M., & Ayers, M. D. (Eds.). (2003). *Cyberactivism: Online activism in theory and practice*. London: Routledge.
- McNutt, J. (2008). Advocacy organizations and the organizational digital divide. *Currents: Scholarship in the Human Services*, 7(2).
- McNutt, J. G., & Barlow, J. (2012). A Longitudinal Study of Political Technology Use by Nonprofit Child Advocacy Organizations. In A. Manoharan & M. Holtzer (Eds.), *E-Governance and Civic Engagement: Factors and Determinants of E-Democracy*. IGI Books. doi:10.4018/978-1-61350-083-5.ch020
- McNutt, J. G., & Boland, K. M. (1999). Electronic Advocacy by Non-Profit Organizations in Social Welfare Policy. *Nonprofit and Voluntary Sector Quarterly*, 28(4), 432–451. doi:10.1177/0899764099284004
- McNutt, J. G., & Menon, G. M. (2008). The rise of cyberactivism: Implications for the future of advocacy in the human services. *Families in Society*, 89(1), 33–38. doi:10.1606/1044-3894.3706
- National Telecommunications and Information Administration. (2004). *A nation online: Entering the broadband age*. Washington, DC: United States Department of Commerce.
- Norris, P. (2001). *Digital divide*. New York: Cambridge University Press. doi:10.1017/CBO9781139164887
- Orcutt, M. (2011, November 9). How Occupy Wall Street Occupied Twitter, Too. *Technology Review*. Retrieved from <http://www.technologyreview.com/view/426079/how-occupy-wall-street-occupied-twitter-too/>
- Rainie, L., Purcell, K., & Smith, A. (2011). *The social side of the internet*. Pew Internet and American Life Project.
- Schuler, D. (1996). *New community networks: Wired for change*. Reading, MA: Addison-Wesley.
- Schussman, A., & Earl, J. (2004). From barricades to firewalls? Strategic voting and social movement leadership in an Internet age. *Sociological Inquiry*, 74(4), 439–463. doi:10.1111/j.1475-682X.2004.00100.x
- Smith, A., Schlozman, K., Verba, S., & Brady, H. (2009). *The Internet and Civic Engagement*. Washington, DC: PEW Internet and American Life Projects. Retrieved from <http://www.pewinternet.org/Reports/2009/15--The-Internet-and-Civic-Engagement.aspx>
- Teachout, Z., & Streeter, T. (Eds.). (2008). *Mousepads, shoeleather and hope*. Boulder, CO: Paradigm Publishers.
- Trippi, J. (2004). *The revolution will not be televised: Democracy, the Internet and the overthrow of everything*. New York: Reagan Book/Harper Collins.

Tufec, Z., & Wilson, C. (2012). Social Media and the Decision to Participate in Political Protest: Observations from Tahrir Square. *Journal of Communication*, 62(2), 363–379. doi:10.1111/j.1460-2466.2012.01629.x

Van De Donk, W., Loader, B. D., Nixon, P. G., & Rucht, D. (Eds.). (2004). *Cyberprotest: New media, citizens and social movements*. Routledge.

Verba, S., Schlozman, K. L., & Brady, H. E. (1995). *Voice and equality: Civic voluntarism in American democracy*. Cambridge, MA: Harvard University Press.

West, D. M., & Francis, R. (1996). Electronic advocacy: Interest groups and public policy making. *PS: Political Science & Politics*, 29(2), 25–29.

ADDITIONAL READING

Brainard, L. A., & Siplon, P. D. (2004). Toward nonprofit organization reform in the voluntary spirit: Lessons from the Internet. *Nonprofit and Voluntary Sector Quarterly*, 33(3), 435–457. doi:10.1177/0899764004266021

Castells, M. (2012). *Networks of Outrage and Hope*. New York: Polity Publishers.

Goldkind, L. (2014). E-advocacy in human services: The impact of organizational conditions and characteristics on electronic advocacy activities among nonprofits. *Journal of Policy Practice*, 13(4), 300–315. doi:10.1080/15588742.2014.929073

Hick, S., & McNutt, J. G. (2002). *Advocacy, activism, and the Internet: Community organization and social policy*. Lyceum Books, Incorporated.

Kanter, B., & Fine, A. (2010). *The Networked Nonprofit: Connecting with social media to drive change*. Wiley.com.

McNutt, J. G. (2006). Building evidence-based advocacy in cyberspace: A social work imperative for the new millennium. *Journal of Evidence-Based Social Work*, 3(3-4), 91–102. doi:10.1300/J394v03n03_07

McNutt, J. G., Curtis, K., O'Boyle, T., & Fox, S. (2010). Coffee or Tea? An Examination of On-line Organizing Techniques of the Tea Party and Coffee Party Movements. Presented at ARNOVA's 2010 Conference in Alexandria, VA, November 18-20, 2010

Obar, J. A., Zube, P., & Lampe, C. (2012). Advocacy 2.0: An analysis of how advocacy groups in the United States perceive and use social media as tools for facilitating civic engagement and collective action. *Journal of information policy*, 2, 1-25.

Shirky, C. (2008). *Here Comes Everybody: The Power of Organizing without organizations*. New York: Penguin.

Shirky, C. (2011). The Political Power of Social Media. *Foreign Affairs*, 00157120, Jan/Feb 90 (1)

Song, F. W. (2010). Theorizing Web 2.0. *Information Communication and Society*, 13(2), 249–275. doi:10.1080/13691180902914610

Suárez, D. F. (2009). Nonprofit advocacy and civic engagement on the Internet. *Administration & Society*, 41(3), 267–289. doi:10.1177/0095399709332297

KEY TERMS AND DEFINITIONS

Activism: An effort to promote social change through a variety of traditional and electronic techniques.

Astroturf: Deceptive efforts to simulate popular support or involvement in the interest of a policy goal or effort.

Civic Engagement: Or civic participation has been defined as the individual and collective actions designed to identify and address issues of public concerns. It is the rights of the people, citizens, to define the public good, determine the policies by which they will seek the good, and reform or replace institutions that do not serve that good.

Social Change: An alteration in the social order or fabric of society in the US indicating a change

in social policy at the city, state or federal level. Social change may include changes in organizations, social institutions, the governance process or social relations.

Web 2.0: Refers to a set of technology tools that emerged in the first decade of the 21st century and are characterized by interactivity, pooling of collective intelligence, the Internet as platform and the promotion of user generated content. Web 2.0 is also referred to a Social Media.

E-Government Service Adoption and the Impact of Privacy and Trust



Mehree Iqbal

North South University, Bangladesh

Nabila Nisha

North South University, Bangladesh

Afrin Rifat

North South University, Bangladesh

INTRODUCTION

The importance of privacy and trust for e-Government services has been evidenced in numerous past studies, but there is a paucity of research regarding the effects of these factors on the acceptance and use of e-Government services in a developing economy. This study fills into the void by aiming to conceptualize the proposed constructs of privacy and trust to examine the factors that can influence future use intentions for e-Government services in the context of Bangladesh. To pursue this purpose, the unified theory of acceptance and use of technology (UTAUT) model has been used. Findings will provide further insights into understanding and managing current and potential users of government services via the online platform. This study can also assist various government authorities to consider the idea of providing appropriate

e-Government services to aid the urban people for communicating and conducting transactions in a timely fashion.

BACKGROUND

In recent times, e-Government services are a breakthrough of communication and transaction between the government and citizens. The communication of e-Government broadly has three

categories, namely Government to Citizen (G2C), Government to Business (G2B) and Government to Government (G2G) services. Among these, it is basically the G2C services which most of the governments from developed and developing countries are focusing upon today.

G2C services relates to the ability of government and citizen to communicate information to each other in an efficient and electronic manner. Its sole purpose is to provide information and assistance to citizens including links to renew a passport, download visa forms, file tax returns, etc (Schaupp, Carter & Hobbs, 2009). In fact, G2C has enabled many governments to become more responsive to the needs of its citizens, ultimately resulting in less corruption, increased transparency, greater convenience, revenue growth and cost reductions (Suki & Ramayah, 2010).

In Bangladesh, a task force of information and communication technology (ICT) was formed in 2001 in order to introduce e-Government services. However, due to lack of proper planning, e-Government services for G2C could not be implemented on a large scale. Today, the present government of the country has again initiated a priority-based project regarding

e-Government services in order to promote their vision of a 'Digital Bangladesh'. Under 'Digital Bangladesh' the aim is to increase momentum in efficiency and technological enhancement in government system and its performance (Islam

& Khair, 2012). For this reason, the government has started pilot e-Government projects in various sectors of Bangladesh and it is being implemented under the Ministry of Planning. Also, the government is arranging substantial foreign co-operation in terms of financial assistance and technical collaboration for realization of e-Government services in the country at a national scale (Alam, 2012).

Despite being a developing country, Bangladesh has come to a fast realization that ICT is necessary for promoting economic growth as well as good governance. The primary goal of

e-Government projects in Bangladesh is therefore to establish a relationship between government officials and citizens by providing greater access to government information and services from all parts of the country. Moreover, increasing government accountability by making its operations more transparent can reduce the opportunities for corruption and support sustainable development goals by providing business, rural and traditionally underserved communities with information, opportunities and communications capabilities (Hassan, 2013). Due to this growing phenomenon of e-Government services, Bangladesh even managed to attain the 148th position in world e-Government ranking compared to its 2012 ranking of the 150th position (UN, 2014).

Although e-Government services is potentially a route to the provision of better services delivered to citizens at a lower cost, there is still a low level of acceptance of such services in many developing countries, particularly Bangladesh. Since the success of such services largely depends on citizens' perceptions of the offered services, it is important to explore the factors that may influence the adoption of e-Government services by the citizens of Bangladesh.

ISSUES, CONTROVERSIES, AND PROBLEMS

Today, information technology is the driving force for reform and change in the globalized

world and it is necessary that institutions at all levels employ this trend to become more effective and accessible. The issue of e-Government thus gained momentum since government is an important instrument of public administration that is used to render services for the citizens. Use of technologies in delivering public services can improve the capacities of government institutions as well as the quality of life of citizens by redefining the relationship between citizens and their government. Initially, this may seem as another option for communicating with citizens; however, in the face of rising demands from demographic, economic, social and global trends,

e-Government no longer appears to be a matter of choice, but a necessity for any country.

Many developed countries today have achieved significant progress in implementing e-Government services, but it is particularly a challenge for the developing nations to make government services easily available and accessible over the electronic network to their citizens. This is because of social issues like lack of awareness or Internet access, IT literacy and language barriers, along with financial and political constraints or technical complexities like the need of a strong infrastructure for information technology (Chowdhury & Satter, 2012). All these factors largely impede the implementation of e-Government services in developing countries.

Moreover, there are very real concerns about turning over personal information to the government. While the government may seem like a benevolent organization, it is possible that threats can come from external sources, like professional hackers, criminal or terrorist organizations, intelligence and investigation agencies, etc who may misuse citizens' information for financial or other gains (Alam, 2012). Citizens therefore need to be assured that this information will be treated in an extremely secure environment, so that the information they have listed with the government cannot make them the victims of reprisals by disaffected government workers.

Many citizens may also be reluctant to adopt e-Government services in a developing country

due to a lack of trust in the security of online transactions and concerns regarding the use of information submitted electronically. These security concerns are not without merit. In light of the inherent uncertainty of using an open technological infrastructure like Internet, citizens want assurance that their online interaction with the government is secure and reliable. While

e-Government has the potential to improve government transparency, responsiveness, and accountability, the services will only be adopted over the electronic network if citizens deem them trustworthy (Belanger & Carter, 2008).

Following Tranfield, Denyer & Smart (2003), incorporation of a good systematic review into any research makes it easier for people to understand by synthesizing extensive evidences from prior studies. This chapter has therefore applied an evidence-based approach to enhance the legitimacy and authority of results that were found in this study. For instance, this research has found significant evidences in favor of lack of trust and privacy as biggest challenges in the path of adoption of e-Government services. Impressive bodies of academic research like Alsaghier, Ford, Nguyen & Hexel (2009), Mahadeo (2009), Chan, Thong & Tam (2010), Featherman, Miyazaki & Sprott (2010), Hussein, Mohamed, Ahlan & Mahmud (2010), Alzahrani and Goodwin (2012) and Mamta (2012) have evidently argued that the lack of trust in the service or service provider and concerns over privacy and security of personal information are the key reasons behind citizens' hesitancy to adopt e-Government services. However, very few studies have focused on these two dimensions in particular from the perspective of a developing country. This study thus attempts to significantly contribute to the literature by conceptualizing a proposed research model to investigate the influencing factors for e-Government services in the context of Bangladesh.

RESEARCH MODEL AND HYPOTHESES

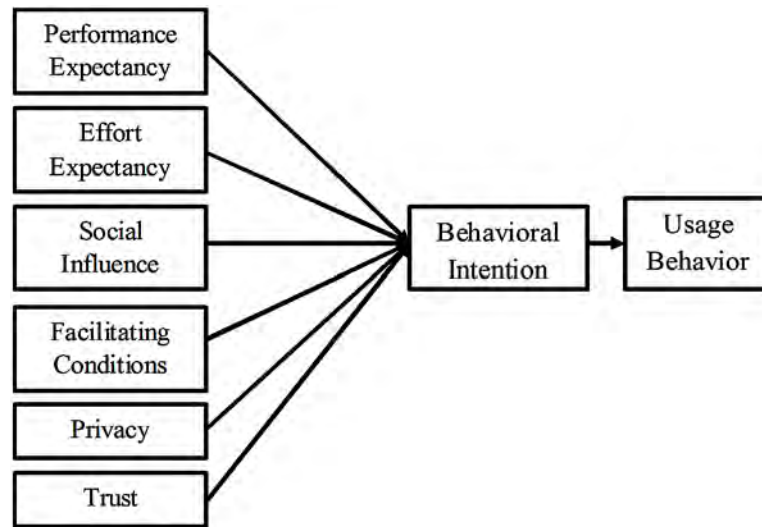
The purpose of this chapter is to determine the factors that can explain and predict citizens' intention to use e-Government services in Bangladesh. To conduct this study, the UTAUT model has been used. Venkatesh, Morris, Davis & Davis (2003) developed the UTAUT model to be used extensively as a baseline to investigate a variety of technologies. The original model comprises of four primary factors: performance expectancy, effort expectancy, social influence and facilitating conditions. However, given the number of technology devices, applications, and services that are targeted at consumers in recent times, Venkatesh, Thong & Xin (2012) introduced a revised UTAUT2 model to identify factors that can influence consumer adoption of technologies (Stofega & Llamas, 2009).

This study therefore chose the original UTAUT and UTAUT2 model as a theoretical foundation to develop a proposed research model. In addition to the original four key constructs, two additional constructs, drawn from previous literature of e-Government services, has been incorporated to make a significant theoretical contribution to the user context of the conceptual model. The proposed model is presented in Figure 1. In addition, all the variables hypothesized in this study and their probable relationships towards citizens' acceptance and use of e-Government services has been discussed next.

Performance Expectancy

Performance expectancy has often been proved as the strongest predictor of the intention of adopting technology-based services (Venkatesh et al., 2003). For this chapter, performance expectancy has been measured by the perceptions of using e-

Figure 1. The proposed research structure



Government services in terms of benefits like saving time, effort, improving the quality of services or facilitating communication with government authorities (Schaupp et al., 2009). Several prior studies by researchers like Alzahrani and Goodwin (2012) and Imarah, Zwain & Al-Hakim (2013) have proved that this construct has considerable influence on individual intention of adopting e-Government services. As a result, the following hypothesis has been proposed:

H1: Performance expectancy significantly affects individual intention to use e-Government services.

Effort Expectancy

Effort expectancy is considered to be directly related with the ease of using a particular technology. According to Venkatesh et al. (2003), effort-oriented constructs like perceived ease of use, complexity and ease of use act as more salient factors under this construct. Past studies have examined direct and indirect effects of effort expectancy on the adoption of e-Government services and have reported mixed results. While Alzahrani and Goodwin (2012) stated that effort

expectancy does not really influence individual intention to use e-Government services, several other researchers like Mcleod, Pippin & Masion (2009) and Imarah et al. (2013) proved the existence of considerable impact of this construct. Therefore, the following hypothesis has been proposed:

H2: Effort expectancy significantly affects individual intention to use e-Government services.

Social Influence

Social influence has been acknowledged as a significant construct in many prior studies by researchers like Venkatesh and Davis (2000) and Lopez-Nicolas, Molina-Castillo & Bouwman (2008). The idea behind social influence is that even if people are not themselves favorable towards a particular technology, they may choose to adopt it because people who are important to them think they should use that technology. This construct has already been used by Imarah et al. (2013) and Barati and Bakhshayesh (2015), where its influence on adoption of e-Government services was supported. This study thus hypothesizes that:

H3: Social influence significantly affects individual intention to use e-Government services.

Facilitating Conditions

Many prior empirical studies have proved that facilitating conditions, in terms of resources and technology, influence citizens' intention to adopt e-Government services. Studies by Chan et al. (2010) and Imarah et al. (2013) claimed that all things being equal, an individual with a lower level of facilitating conditions will have lower intention to use e-Government services. On the other hand, few studies like Carter and Schaupp (2008) found facilitating conditions to be an insignificant determinant of individual intention of adopting e-Government services. Hence, the following hypothesis has been developed:

H4: Facilitating conditions significantly affects individual intention to use e-Government services.

Privacy

Adoption of any kind of e-Government services is severely influenced by the security and privacy concerns of citizens (Alzahrani & Goodwin, 2012). An individual who uses e-Government services relies on the system privacy and expects that no personal information will be disclosed to unauthorized people or authorities (Mcleod et al., 2009). Several empirical studies like Featherman et al. (2010) and Alzahrani and Goodwin (2012) have argued that higher the risk of breaching privacy, lower the willingness will be to adopt e-Government services, particularly in a developing country. This study thus poses the following hypothesis:

H5: Privacy significantly affects individual intention to use e-Government services.

Trust

According to El-Wajeeh, Galal-Edeen & Mokhtar (2014), trust of the citizens in the government providing electronic services and the trust they possess in electronic network and information technology are both essential for adopting e-Government services. Besides, prior experience of an individual with e-Government services also plays a major role in creating this trust. Many empirical studies like Hussein et al. (2010) and Mamta (2012) have found significant positive association between trust and individual intention to adopt e-Government services. Thus, the following hypothesis has been conceived:

H6: Trust significantly affects individual intention to use e-Government services.

Behavioral Intention

Behavioral intention is the most important indicator of the ultimate adoption decision for a technology-based service (Nisha, Iqbal, Rifat & Idrish, 2015). For e-Government services, behavioral intention can be defined as the strength of the prospective users' intention to use or to support the adoption of e-Government innovations. Mahadeo (2009) and Imarah et al. (2013) further found significant positive correlation between behavioral intention and actual usage of e-Government services. As such, this study hypothesizes that:

H7: Behavioral intention significantly affects individual usage behavior of using e-Government services.

DATA ANALYSIS AND DISCUSSION

Data for this study has been collected using a structured questionnaire on a sample of 1200

respondents, who mostly represent the urban population of Bangladesh. Respondents were selected on the basis of probability sampling and a stratified random sampling method and after a three-week survey, 966 completed and usable responses were obtained from the survey.

The quantitative survey contained 26 statements in order to evaluate the constructs of the proposed model. Scales for performance expectancy, effort expectancy, social influence, facilitating conditions and behavioral intention were adapted from Mahadeo (2009), Schaupp et al. (2009), Hussein et al. (2010) and Imarah et al. (2013). Alternatively, scales for privacy and trust were based on Chan et al. (2010), Featherman et al. (2010) and Alzahrani and Goodwin (2012). All these items were measured using a five-point Likert scale, ranging from “strongly disagree” to “strongly agree” and were adapted from previous studies on e-Government services - with minor changes in wording to tailor them to the context of Bangladesh. This ensured the content validity of the questionnaire used to assess each constructs depicted in Figure 1.

The conceptual research model was tested using the structural equation modeling (SEM) facilitates of SmartPLS. SEM was applied to investigate the underlying relationships between the range and quality of e-Government services, focusing particularly into urban citizens’ behavioral intentions. SEM has the ability to isolate observational error from measurement of latent variables like the proposed constructs of privacy and trust in this study. Thus, the use of SEM for evaluating e-Government service adoption is the unique contribution of this chapter, since to the best of the authors’ knowledge SEM has not been used for any e-Government research in the context of Bangladesh.

As such, through partial least squares (PLS) method, first the measurement model was examined to assess the reliability and validity of all variables and then, the structural model was analyzed to examine the relationships hypothesized in the research model.

MEASUREMENT MODEL ANALYSIS

Table 1 and Table 2 present the measurement results regarding the reliability, validity, correlations and factor loadings. Composite reliabilities of constructs ranged between 0.818 and 1.000, which exceeds the stated 0.7 cut-off value (Nunnally and Bernstein, 1994). In all cases, average variance extracted (AVE) was greater than 0.5 and greater than each square correlations as well, which confirms both convergent and discriminant validity of the model (Fornell and Larcker, 1981). In addition, the internal consistency reliabilities (ICRs) of multi-item scales modeled with reflective indicators were observed to be 0.75 or greater, suggesting adequate reliability. The pattern of loadings and cross-loadings even supported internal consistency and discriminant validity, with few exceptions: one item from social influence and facilitating conditions, and two items from privacy and trust were deleted due to their low loadings and high cross-loadings.

STRUCTURAL MODEL ANALYSIS

Figure 2 presents the path coefficients and significance levels in the structural model of this study. The constructs of trust (0.301, $p < 0.05$), facilitating conditions (0.277, $p < 0.05$) and effort expectancy (0.198, $p < 0.05$) were observed to have significant and positive paths to behavioral intention of users for e-Government services, in their order of influencing strength.

However, the constructs of performance expectancy (0.141, $p > 0.05$), social influence (0.122, $p > 0.05$) and privacy (0.009, $p > 0.05$) reported an insignificant path towards the individual behavior of using e-Government services. Therefore, all hypotheses (except H1, H3 and H5) dealing with behavioral intention to use e-Government services are supported. Subsequently, the hypothesized relationship between behavioral intention and usage (0.321, $p < 0.05$) is found to be statistically significant, thereby supporting hypothesis H7.

Table 1. Factor loadings, composite reliability and AVEs

Constructs	Items	Factor Loadings	Composite Reliability	AVE
Performance Expectancy	PE1	0.807	0.895	0.688
	PE2	0.847		
	PE3	0.850		
	PE4	0.761		
Effort Expectancy	EE1	0.822	0.818	0.699
	EE2	0.834		
	EE3	0.840		
	EE4	0.802		
Social Influence	SI1	0.867	0.891	0.762
	SI2	0.876		
	SI3	0.823		
Facilitating Conditions	FC1	0.776	0.846	0.660
	FC2	0.877		
	FC3	0.818		
Privacy	PC2	1.000	1.000	1.000
Trust	T1	0.915	0.851	0.695
	T2	0.737		
Behavioral Intention	BI1	0.868	0.865	0.738
	BI2	0.873		
	BI3	0.823		

Table 2. Measurement model estimations

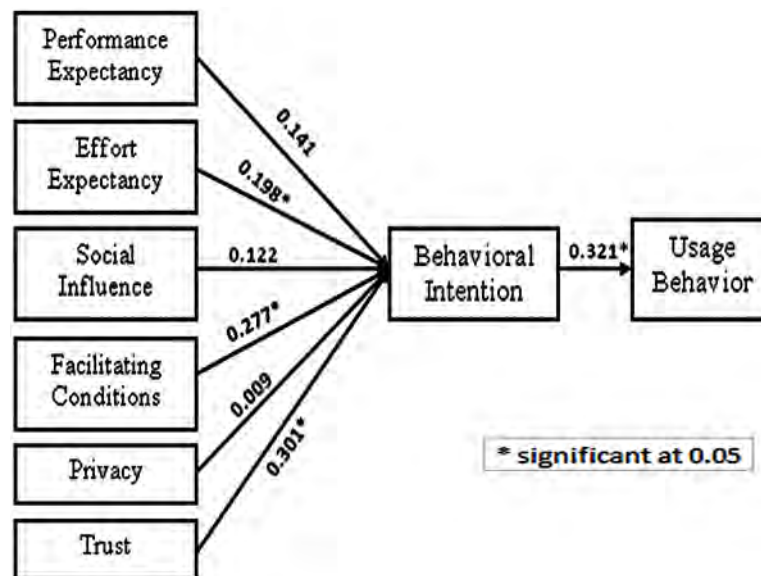
	ICRs	BI	EE	FC	P	PE	SI	T
BI	0.792	0.846						
EE	0.810	0.223	0.832					
FC	0.853	0.136	0.134	0.975				
P	0.781	0.254	0.306	0.534	Single-Item Construct			
PE	0.865	0.219	0.435	0.577	0.444	0.995		
SI	0.835	0.247	0.237	0.457	0.129	0.183	0.860	
T	0.802	0.273	0.164	0.413	0.023	0.625	0.139	0.923

CONCLUSION

The primary objective of this chapter is to provide a conceptual model that determines the influencing factors of citizens' intention towards the use of e-Government services. Lack of studies in the area of G2C adoptions is a major reason why such study

is required to obtain a clear picture of the current status of e-Government implementation particularly in developing countries. Thus, the findings are expected to make a significant contribution to the emerging literature on e-Government adoption by employing new variables like privacy and trust into the well-accepted UTAUT model. Results

Figure 2. Results of structural equation modeling



reveal that factors like trust, facilitating conditions and effort expectancy play a deterministic role in persuading citizens towards the adoption of

e-Government services in Bangladesh. These antecedents of intentions are perceptual in nature and they can be influenced by appropriate promotional campaigns through fairs and seminars, demonstrations, ministry-funded projects, various online portals, etc. Furthermore, the research model could be used as a reference for studies to cater different forms of e-Government technology.

SOLUTIONS AND RECOMMENDATIONS

This study has revealed quite a few understandings towards the possibilities of potential usage of e-Government services. Constructs such as trust, facilitating conditions and effort expectancy saliently influence behavioral intention, which directly stimulates the usage of e-Government services. Results here imply that citizens are more likely to trust the service if they are convinced with the credibility of the service and the service provider. For e-Government services, trust in the

government and Internet is therefore critical in order to mitigate any uncertainty or vulnerability in the information exchange. Moreover, as e-Government services are mostly going to depend on the extensive use of the electronic network, technical aspects like computers and information systems needs to be stable and reliable to convince citizens for adopting such services. Findings further state that less effort input by citizens while availing e-Government services are more likely to show positive attitude in embracing the service in Bangladesh.

Conversely, performance expectancy, social influence and privacy are identified as insignificant factors of behavioral intention for the citizens of Bangladesh. This implies that citizens may simply choose to reduce their effort for any government services and may trade-off their privacy concerns for this convenience. On the other hand, citizens are not expecting much relative advantage from e-Government services and this may be due to the initial stage of the implementation of such services in Bangladesh. Additionally, positive or negative evaluations from familiar people are less likely to influence the citizens of Bangladesh to consider the adoption of e-Government services.

The inclined reasons behind such findings are mostly shaped by the local culture and infrastructure of the country. As such, important implications for practice are discussed next. To attain trust and increase citizens' adoption of e-Government services, the government needs to increase awareness regarding these new offerings. For this purpose, the government may arrange fairs or seminars related to ICT, where they can demonstrate the electronic government services available to the citizens. Moreover, if local information booths can be established in every ward of the capital city, then it would help citizens to gather information about e-Government services and hence, pave a way for establishing trust. Additionally, it is important that the government ensures faster and cheaper internet connections, computer facilities at low cost and skilled human resources to train people for using electronic networks. This will enhance the facilitating conditions for promoting e-Government services in Bangladesh. In fact, the government has already taken few initiatives by encouraging computer literacy for both urban and rural people alike. However, the government can commence more ministry-funded projects and outsource the training or facilitating services to registered ICT organizations. They can even establish an online platform where all government services will fall under one umbrella. For instance, there will be a collection of portals related to citizens' advice bureau, applications and renewals for passport, national identification card or driving license, utility bill payments, metropolitan police, city corporation tax, and government announcements. 24/7 update, call centers, video clips on 'how to do' issues in both the native and English language, along with mobile applications may also be developed to facilitate e-Government services in Bangladesh. Besides, demonstrations through short films in public places or advertisements over television channels or radio platforms can be utilized to educate people and enhance trust for e-Government services. Demonstrations should be

such that it clearly reflects the benefits and ease of use of the e-Government platform. Not only that, competencies of the government officials and their assurance to help with the electronic network needs to be highlighted as well, so that it helps to motivate citizens to use e-Government services in Bangladesh.

A common culture in developing countries is to visit the physical place in order to avail any service. Bangladesh is no different and for the betterment of the country, the government needs to mandate the adoption of the electronic network for any government related services. For this, lower fee for electronic services compared to manual paper-based services can be promoted and all government-related information must only be available online. This will prompt citizens to consider the adoption of e-Government services in Bangladesh.

LIMITATIONS AND FUTURE RESEARCH DIRECTIONS

Like most empirical research, this study has several limitations. First, the conclusions drawn from this study are based solely on the urban population of Bangladesh. Future research can examine a wider respondent base by considering the rural population. Primary research methods like face-to-face interviews can be used to gather reliable data in this regard. Moreover, this study has mainly employed the UTAUT model but a longitudinal study can be adopted in future works. In fact, use of moderators like age, gender and voluntariness in the research model may provide a better insight into the perceptions of e-Government service users. Also, additional constructs like perceived self-efficacy, perceived risk, personal innovativeness, etc., can be added in the research model to enhance understanding of individual users' intention to adopt e-Government services in Bangladesh.

REFERENCES

- Alam, M. J. (2012). E-governance in Bangladesh: Present problems and possible suggestions for future development. *International Journal of Applied Information Systems*, 4(8), 21–25. doi:10.5120/ijais12-450804
- Alsaghier, H., Ford, M., Nguyen, A., & Hexel, R. (2009). Conceptualizing citizen's trust in e-Government: Application of Q methodology. *Electronic Journal of E-Government*, 7(4), 295–310.
- Alzahrani, M. E., & Goodwin, R. D. (2012). Towards a UTAUT-based model for the study of e-Government citizen acceptance in Saudi Arabia. *World Academy of Science. Engineering and Technology*, 6, 4–23.
- Barati, A., & Bakhshayesh, S. (2015). Electronic tax system and the facing challenges (Case Study: Kermanshah Province Tax Payers). *Indian Journal of Fundamental and Applied Life Science*, 5(1), 480–497.
- Belanger, F., & Carter, L. (2008). Trust and risk in e-government adoption. *The Journal of Strategic Information Systems*, 17(2), 165–176. doi:10.1016/j.jsis.2007.12.002
- Carter, L., & Schaupp, L. C. (2008). Efficacy and Acceptance in E-File Adoption. *Americas Conference on Information Systems (AMCIS) Proceedings*.
- Chan, F. K. Y., Thong, J., & Tam, K. Y. (2010). Modelling citizen satisfaction with mandatory adoption of an e-Government technology. *Journal of the Association for Information Systems*, 11(10), 519–549.
- Chowdhury, M. M. H., & Satter, A. K. M. Z. (2012). The role of e-Governance in creating Digital Bangladesh. *International Journal of Advances in Science and Technology*, 4(6), 24–33.
- El-Wajeeh, M., Galal-Edeen, G., & Mokhtar, H. (2014). Technology acceptance model for mobile health systems. *IOSR Journal of Mobile Computing and Acceptance*, 1(1), 21–33.
- Featherman, M. S., Miyazaki, A. D., & Sprott, D. E. (2010). Reducing online privacy risk to facilitate e-service adoption: The influence of perceived ease of use and corporate credibility. *Journal of Services Marketing*, 24(3), 219–229. doi:10.1108/08876041011040622
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *JMR, Journal of Marketing Research*, 18(1), 39–50. doi:10.2307/3151312
- Hassan, M. R. (2013). E-governance and E-government in Bangladesh: Performance, challenges and remedies. *Asian Journal of Applied Science and Engineering*, 2(2), 111–117.
- Hussein, R., Mohamed, N., Ahlan, A., & Mahmud, M. (2010). An integrated model on online tax adoption in Malaysia. *Proceedings of the European, Mediterranean and Middle Eastern Conference on Information Systems: Global Information Systems Challenges in Management*, 225–248.
- Imarah, A. A. T. A., Zwain, A. A. A., & Al-Hakim, L. A. Y. (2013). The adoption of e-government services in the Iraqi higher education context: An application of the UTAUT Model in the University of Kufa. *Journal of Information Engineering and Applications*, 3(10), 77–84.
- Islam, M. J., & Khair, R. (2012). Preparation of e-Government in Bangladesh: An exploratory analysis. *JU Journal of Information Technology*, 1, 19–26.
- Lopez-Nicolas, C., Molina-Castillo, F. J., & Bouwman, H. (2008). An assessment of advanced mobile services acceptance: Contributions from TAM and diffusion theory models. *Information & Management*, 45(6), 359–364. doi:10.1016/j.im.2008.05.001

- Mahadeo, J. D. (2009). Towards an understanding of the factors influencing the acceptance and diffusion of e-Government services. *Electronic Journal of E-Government*, 7(4), 391–402.
- Mamta, B. (2012). Tax payers' perception towards e-file adoption: An empirical investigation. *Ganpat University-Faculty of Management Studies Journal of Management and Research*, 5, 1–16.
- McLeod, A., Pippin, S., & Masion, R. (2009). Individual taxpayer intention to use tax preparation software: Examining experience, trust, and perceived risk. *Journal of Information Science and Technology*, 6(1), 25–44.
- Nisha, N., Iqbal, M., Rifat, A., & Idrish, S. (2015). Mobile health services: A new paradigm for health care systems. *International Journal of Asian Business and Information Management*, 6(1), 1–18.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric Theory*. New York: McGraw-Hill.
- Schaupp, L. C., Carter, L., & Hobbs, J. (2009). E-file adoption: A study of U.S. taxpayers' intentions. *Proceedings of the 42nd Hawaii International Conference on System Sciences*.
- Stofega, W., & Llamas, R. T. (2009). *Worldwide mobile phone 2009-2013 forecast update*. IDC Document Number 217209. Framingham, MA: IDC.
- Suki, N. M., & Ramayah, T. (2010). User acceptance of the e-Government services in Malaysia: Structural Equation Modeling approach. *Interdisciplinary Journal of Information, Knowledge, and Management*, 5, 395–413.
- Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British Journal of Management*, 14(3), 207–222. doi:10.1111/1467-8551.00375
- UN. (2014). *E-government Survey 2014: E-government For the Future We Want*. New York: United Nations.
- Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science*, 46(2), 186–204. doi:10.1287/mnsc.46.2.186.11926
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *Management Information Systems Quarterly*, 27(3), 425–478.
- Venkatesh, V., Thong, J. Y. L., & Xin, X. (2012). Consumer acceptance and use of information technology: Extending the Unified Theory of Acceptance and Use of Technology. *Management Information Systems Quarterly*, 36(1), 157–178.

ADDITIONAL READING

- Al-adawi, Z., Yousafzai, S., & Pallister, J. (2005). Conceptual model of citizen adoption of e-Government. *Proceedings of the Second International Conference on Innovations in Information Technology*, 1-10.
- Hassan, M. M. (2015). E-government success and failure: A case study of Bangladesh Police. *Daffodil International University Journal of Science and Technology*, 10(1-2), 61–67.
- Karim, M. R. (2015). E-Government in service delivery and citizens satisfaction: A case study on public sector in Bangladesh. *International Journal of Managing Public Sector Information and Communication Technologies*, 6(2), 49–60. doi:10.5121/ijmpict.2015.6205
- Kumar, V., Mukerji, B., Butt, I., & Persaud, A. (2007). Factors for successful e-Government adoption: A conceptual framework. *The Electronic Journal of E-Government*, 5(1), 63–76.
- Rahman, M. M., & Rajon, S. A. A. (2012). An effective framework for implementing electronic governance in developing countries: Bangladesh perspective. *International Journal of Computer and Information Technology*, 3(1), 20–29.

Uddin, G. (2012). E-governance of Bangladesh: Present scenario, expectation, ultimate target and recommendation. *International Journal of Scientific & Engineering Research*, 3(11), 1–20.

Warkentin, M., Gefen, D., Pavlou, P. A., & Rose, G. M. (2002). Encouraging citizen adoption of e-Government by building trust. *Electronic Markets*, 12(3), 157–162. doi:10.1080/101967802320245929

Zambrano, R., & Seward, R. K. (2013). *From connectivity to service delivery: Case studies in e-governance*. United Nations Development Program.

KEY TERMS AND DEFINITIONS

Behavioral Intention: Individual intention to use a particular technology that directly affects actual usage.

Effort Expectancy: Degree of ease associated with the use of a technology.

e-Government: Information system used to conduct government transactions and communicate with citizens via Internet.

Facilitating Conditions: User perception that resources and support are available to perform a particular activity.

Performance Expectancy: Degree to which using a technology provides benefits to the user in performing a particular activity.

Privacy: A belief that personal information entered into an information system will remain private and secure.

Social Influence: User perception that important others like friends or family believe they should adopt a particular technology.

Trust: Confidence in the online environment and the electronic services provided by Government.

Mastering Electronic Government in the Digital Age



Kijpokin Kasemsap

Suan Sunandha Rajabhat University, Thailand

INTRODUCTION

The ultimate goal of electronic government (e-government) is to offer the increased portfolio of public services to citizens in a cost-effective manner. Most governments make tremendous efforts to deliver the online services to citizens (Roy, Chartier, Crete, & Poulin, 2015). The operation of information and communications technology (ICT) has been the major development of e-government in the past decade (Reddick & Anthopoulos, 2014). ICT has altered public administration by transforming the internal processes and external interactions (Meijer & Bekkers, 2015). E-government services must be redesigned to ensure that the benefits of ICT systems are completely employed (Kasemsap, 2016). The most significant role of ICT is to drive the organizational innovation through information systems and solve the crucial problems that the government cannot solve on its own (Sindelar, Mintz, & Hughes, 2010).

E-government has emerged as an effective method of delivering government services to citizens (Weerakkody, Dwivedi, & Kurunananda, 2009). The diffusion of e-government is an international phenomenon (Carter & Weerakkody, 2008). As an integral part of administration modernization (Stier, 2015), e-government is one of the most important ways to bridge the digital platform in developing countries (Venkatesh, Sykes, & Venkatraman, 2014), acts as an effective exploration of government innovation (Wu & Guo, 2015), and can improve the government performance and create the new public value for citizens and businesses (Wang, 2014). The success of e-government system lies with its cost savings

in implementation, adoption, benefits provided to the recipients of the system, and associated risks in operating the system (Weerakkody, Irani, Lee, Osman, & Hindi, 2015).

This article aims to bridge the gap in the literature on the thorough literature consolidation of e-government. The extant literature of e-government provides a contribution to practitioners and researchers by describing the multifaceted applications of e-government to appeal to the different segments of e-government in order to maximize the public sector impact of e-government in the digital age.

BACKGROUND

Regarding e-government, ICT is an effective instrument for reducing the role of bureaucracy in government organizations (Cordella & Tempini, 2015). There has been a social evolution on the Internet recognized as the Web 2.0 (Waters, Burnett, Lamm, & Lucas, 2009). Web 2.0 is characterized by enabling and encouraging participation through open applications and through services with rights granted to use content in the new and exciting contexts (Chadwick, 2009). The adoption of ICT in public sector organizations has been associated with the e-government reform programs aiming at reducing the inefficiencies generated by the bureaucratic burden (Osborne & Plastrik, 1997). The levels of human and technological development of a country are the driving forces of e-government (Siau & Long, 2009).

Global interest in e-government has produced a wide range of internal and external evaluations of national performance in service delivery (Taylor,

Marshall, & Amiri, 2010). E-government standard describes how governments work, share information, and deliver services to the internal and external stakeholders (Sun, Ku, & Shih, 2015). ICT artifacts are recognized as the linear catalysts of transformation of public sector organizations and structures (West, 2004). ICT diffusion leads to transaction integration, process reengineering, and administrative transformation, toward creating the citizen-centric government (Zhang, Meng, Guo, Yin, & Luo, 2015). Governments' investments in public sector information systems are correlated with organizational transformations designed to enhance the policy effectiveness (Gil-Garcia & Pardo, 2005).

CHALLENGES AND IMPLICATIONS OF ELECTRONIC GOVERNMENT IN THE DIGITAL AGE

The overview of e-government; the adoption of e-government; the digital era governance (DEG) and new public management (NPM); and the significance of e-government in the digital age are described in this article.

Overview of Electronic Government

Electronic government (e-government) refers to the use of ICT tools and applications to enhance the government transparency and accountability in the public administration by improving the public services delivery, access to information, and public governance (Chatfield & Alhujran, 2009). Transparency is recognized as a key value for trustworthy governments (Grimmelikhuijsen, Porumbescu, Hong, & Im, 2013). The scope of transparency is required to be carefully managed for the effective e-government in public sector organizations (Bannister & Connolly, 2011). The characteristics of good public governance include the improved transparency and accountability (Al-Hujran, Al-Debei, Chatfield, & Migdadi, 2015).

The promise of greater government transparency and accountability is the essential perspective for developing countries to take part in the e-government projects (Chen, Jubilado, Capistrano, & Yen, 2015).

E-government projects can be technically complicated, and involve many customers, engineers, and regulatory authorities (Li, 2009). Many governments worldwide are still experiencing the practical problem of the low-level adoption of e-government services by citizens (Rana & Dwivedi, 2015). The e-government adoption-related problem needs the urgent research attention since the success of e-government is highly dependent upon the citizens' adoption and the use of e-government services (Ozkan & Kanat, 2011). One of the main factors of the success of e-government is the development government websites (Rana, Dwivedi, Williams, & Weerakkody, 2015). Most extensive e-government measurements are tailored for the front-end website evaluation from the perspectives of citizens and businesses (Fan & Luo, 2014).

Adoption of Electronic Government

The citizen's adoption of e-government services is an important issue for the success of e-government initiatives (Carter & Belanger, 2005). Prior research classified the e-government adoption literature into two streams (Reddick, 2005). The first stream studies the e-government adoption from the supply-side perspective, which reflects factors that are related to the supplier of public services (Al-Hujran et al., 2015). The supply-side perspective explores the factors that affect the government organizations' adoption and the implementation of e-government services (Li & Feeney, 2014). Examples of these factors concerning supply-side perspective include organizational characteristics (e.g., size, red tape, culture, and top management support), information technology (IT) infrastructure, financial resources, and skilled personnel.

The second stream studies the e-government adoption from the demand-side perspective which focuses on the customers of public services (i.e., citizens) (Al-Hujran et al., 2015). Many scholars have utilized a number of theoretical frameworks to examine the citizens' adoption of e-government services provided by the government (Lin, Fofanah, & Liang, 2011). Examples of these factors concerning demand-side perspective include trust, risk, security, usefulness, attitude, quality, satisfaction, and computer experience. The implementation of digital reforms and the management of e-government programs should be promoted within an effective administrative environment (Stier, 2015). Administrations need the sufficient funding and the qualified staff in order to handle the challenges of digital reforms concerning e-government (Norris & Moon, 2005).

Digital Era Governance and New Public Management

The Internet and the Web 2.0 have revolutionized the public management in advanced industrial countries (Margetts & Dunleavy, 2013). The toolkit for public management reform has shifted away from the new public management (NPM) approach emphasizing fragmentation and competition toward the digital era governance (DEG) perspective, focusing on reintegrating services, providing holistic services for citizens, and implementing digital changes in administration (Margetts & Dunleavy, 2013). NPM reform effectively facilitates the transition of the governance regime from strategic partnership to executive dominance in e-government (Lee, 2012).

NPM and the utilization of complex network structures spanning multiple tiers of government can save governments money (Eggers & Goldsmith, 2008). Handling digital change favorably requires the new macro-theory of public sector development, culture, and characteristic patterns of organizational governance in e-government (Dunleavy, Margetts, Bastow, & Tinkler, 2005).

DEG appearance potentially works with the perspectives of business and civil society changes, within an overall pathway of social modernization that has greatly changed with the development of the Internet and online social processes (Goldfinch & Wallis, 2009).

Significance of Electronic Government in the Digital Age

E-government employs ICT to provide citizens with government-related information about public services (Lee, Chang, & Berry, 2011). Regarding political outcomes, e-government programs can reduce corruption (Bertot, Jaeger, & Grimes, 2010), modernize bureaucracy (Ahn & Bretschneider, 2011), and improve trust in local government (Tolbert & Mossberger, 2006). There is a positive relationship between democracy and e-government (Gulati & Yates, 2011). E-government has been considered as an effective way to increase the citizen's trust in government and improve the citizen's evaluation of government (Wu & Guo, 2015). Democratic politicians often apply e-government to increase their electoral chances by improving the public service provision and by engaging with citizens (Bussell, 2011). Bureaucratic agents can be bypassed by centralizing the parts of their competencies within the national e-government portals, thus restraining the opportunity structures for power misuses (Andersen, 2009).

Public administrations must make a greater effort to implement e-government systems that allow the improved communication with the general public, thus achieving a greater degree of participation (Willoughby, Gomez, & Lozano, 2010). The increasing number of governments are formulating the effective action plans to move service delivery to the Web 2.0, to enhance the government-related information to citizens, and to improve public sector workplaces (Sarantis, Smithson, Charalabidis, & Askounis, 2010). Governments and public bodies have been fostering

the development of e-government services during the last decade, promoting better administrative services through digital channels (Fernández-i-Marín, 2011). In the phase of development of the information society, governments are playing an important role in order to enable the citizens to engage in the different administrative processes in an effective manner (Garcia, Oliva, Belleboni, & de la Cruz, 2011). One of the most important stages of establishment of the information society is the development and implementation of e-government as a system of efficient interaction among government authorities, citizens, and businesses (Zubareva & Byelov, 2015).

Benefits of e-government stem from the relatively low cost and high availability (Kopackova, Michalek, & Cejna, 2010). The outcomes of e-government are related to the adoption of e-government, government capacity, and institutional characteristics (Moon & Norris, 2005). E-government application saves the e-government users' money in the form of faster, easier and more convenient service, better quality and reduced turnaround times, and in some cases a reduction in the direct cost for the e-government services. E-government users can gain the enhanced service levels represented by the improved service quality, reduced turnaround times, improved access to services, and availability of new e-government services.

E-government performance is positively correlated with government effectiveness (Kim, 2007), as it takes a certain level of administrative capabilities to create and maintain the sophisticated e-government framework (Norris & Moon, 2005). As e-government evolves into the transactions stage, governments must grapple with how to encourage the development of e-transactions (Chen & Thurmaier, 2008). The utilization of ICT can facilitate the relationships among citizens, businesses, and government branches (Sun et al., 2015). The benefits of applying ICT in the public sector organizations are concerned with the enhanced transparency, accountability, and trust in

government (Xu, 2012). E-government becomes part of managerial strategies, and promotes the domestic generation of intellectual property (Cook & Horobin, 2006).

FUTURE RESEARCH DIRECTIONS

The classification of the extant literature in the domains of e-government will provide the potential opportunities for future research. E-procurement in the public sector is globally developing in the digital age. E-procurement projects are part of the country's larger e-government efforts to favorably serve its citizens and businesses in the digital economy. E-procurement has various benefits (e.g., cost reduction, process efficiency, spending controls, and compliance) for the public sector organizations. Practitioners and researchers should acknowledge the applicability of a more multidisciplinary approach toward research activities in implementing e-procurement value chain. Cloud computing is the modern pattern of digital computing technology in which virtualized resources are provided as the services over the Internet (Kasemsap, 2015). The adoption of cloud computing in e-government will be the beneficial topic for future research direction.

CONCLUSION

This article highlighted the overview of e-government; the adoption of e-government; the DEG and NPM; and the significance of e-government in the digital age. E-government is the use of ICT to improve the activities of public sector organizations. E-government can open new opportunities for city and local governments to engage in governance by requiring the reforms of underlying working processes. The use of e-government can advance the local democracy by improving the access to information and deepening the citizens' participation in the policy-making process. E-government

offers a path to sustain with the civil society and the private sector to design the effective services and tools to execute the government policies. By improving cities' ability to provide services, achieve policy goals, and increase transparency, e-government encourages the greater trust, participation, and engagement of citizens.

Mastering e-government helps public sector employees and e-government users improve their insights into the systematic e-government systems. E-government enables the efficient provision of services, the enhanced citizen's trust, participation, and ability to participate in the process of governance. The application of ICT to improve the e-government services has been a primary driver for e-government activity. E-government can help accomplish the reform agenda in the public sector organizations. When aligned with modernization goals, implementing e-government can help public administrations focus on the additional changes needed to meet service delivery and good governance concerns. Through citizen engagement, e-government can improve the overall trust relationship between government and public administrations.

The benefits of e-government include the enhanced efficiency, improved services, effective accessibility of public services, and increased accountability in the public sector organizations. However, there are many barriers to using e-government services, such as poor search capabilities, incomplete information, difficulty finding the e-government services via the Internet, and concerns about security and privacy. The major disadvantage concerning e-government is the lack of equality in public access to the Internet, reliability of information on the Web 2.0, and hidden agendas of government groups. Although large amount of money is spent on the development and implementation of e-government, the outcomes and effects of trial Internet-based governments are often difficult to gauge or unsatisfactory. Mastering e-government has the potential to enhance public sector performance and achieve strategic goals in the digital age.

REFERENCES

- Ahn, M. J., & Bretschneider, S. (2011). Politics of e-government: E-government and the political control of bureaucracy. *Public Administration Review*, 71(3), 414–424. doi:10.1111/j.1540-6210.2011.02225.x
- Al-Hujran, O., Al-Debei, M. M., Chatfield, A., & Migdadi, M. (2015). The imperative of influencing citizen attitude toward e-government adoption and use. *Computers in Human Behavior*, 53, 189–203. doi:10.1016/j.chb.2015.06.025
- Andersen, T. B. (2009). E-government as an anti-corruption strategy. *Information Economics and Policy*, 21(3), 201–210. doi:10.1016/j.infoecopol.2008.11.003
- Bannister, F., & Connolly, R. (2011). The trouble with transparency: A critical review of openness in e-government. *Policy & Internet*, 3(1), 1–30. doi:10.2202/1944-2866.1076
- Bertot, J. C., Jaeger, P. T., & Grimes, J. M. (2010). Using ICTs to create a culture of transparency: E-government and social media as openness and anti-corruption tools for societies. *Government Information Quarterly*, 27(3), 264–271. doi:10.1016/j.giq.2010.03.001
- Bussell, J. (2011). Explaining cross-national variation in government adoption of new technologies. *International Studies Quarterly*, 55(1), 267–280. doi:10.1111/j.1468-2478.2010.00644.x
- Carter, L., & Belanger, F. (2005). The utilization of e-government services: Citizen trust, innovation and acceptance factors. *Information Systems Journal*, 15(1), 5–25. doi:10.1111/j.1365-2575.2005.00183.x
- Carter, L., & Weerakkody, V. (2008). E-government adoption: A cultural comparison. *Information Systems Frontiers*, 10(4), 473–482. doi:10.1007/s10796-008-9103-6

- Chadwick, A. (2009). Web 2.0: New challenges for the study of e-democracy in an era of informational exuberance. *I/S: A Journal of Law and Policy for the Information Society*, 5(1), 9–41.
- Chatfield, A. T., & Alhujran, O. (2009). A cross-country comparative analysis of e-government service delivery among Arab countries. *Information Technology for Development*, 15(3), 151–170. doi:10.1002/itdj.20124
- Chen, J. V., Jubilado, R. J. M., Capistrano, E. P. S., & Yen, D. C. (2015). Factors affecting online tax filing: An application of the IS success model and trust theory. *Computers in Human Behavior*, 43, 251–262. doi:10.1016/j.chb.2014.11.017
- Chen, Y. C., & Thurmaier, K. (2008). Advancing e-government: Financing challenges and opportunities. *Public Administration Review*, 68(3), 537–548. doi:10.1111/j.1540-6210.2008.00889.x
- Cook, I., & Horobin, G. (2006). Implementing eGovernment without promoting dependence: Open source software in developing countries in Southeast Asia. *Public Administration and Development*, 26(4), 279–289. doi:10.1002/pad.403
- Cordella, A., & Tempini, N. (2015). E-government and organizational change: Reappraising the role of ICT and bureaucracy in public service delivery. *Government Information Quarterly*, 32(3), 279–286. doi:10.1016/j.giq.2015.03.005
- Dunleavy, P., Margetts, H., Bastow, S., & Tinkler, J. (2005). New public management is dead. Long live digital-era governance. *Journal of Public Administration: Research and Theory*, 16(3), 467–494. doi:10.1093/jopart/mui057
- Eggers, W., & Goldsmith, S. (2008). *Government by network: The new public management imperative*. Boston, MA: Deloitte Research/Ash Institute, Harvard University.
- Fan, B., & Luo, J. (2014). Benchmarking scale of e-government stage in Chinese municipalities from government chief information officers' perspective. *Information Systems and e-Business Management*, 12(2), 259–284.
- Fernández-i-Marín, X. (2011). The impact of e-government promotion in Europe: Internet dependence and critical mass. *Policy & Internet*, 3(4), 1–29. doi:10.2202/1944-2866.1093
- Garcia, S. S., Oliva, A. G., Belleboni, E. P., & de la Cruz, I. P. (2011). Solving identity delegation problem in the e-government environment. *International Journal of Information Security*, 10(6), 351–372. doi:10.1007/s10207-011-0140-7
- Gil-Garcia, J. R., & Pardo, T. A. (2005). E-government success factors: Mapping practical tools to theoretical foundations. *Government Information Quarterly*, 22(2), 187–216. doi:10.1016/j.giq.2005.02.001
- Goldfinch, S., & Wallis, J. (2009). *International handbook of public management reform*. Cheltenham, UK: Edward Elgar Publishing. doi:10.4337/9781849802031
- Grimmelikhuijsen, S., Porumbescu, G., Hong, B., & Im, T. (2013). The effect of transparency on trust in government: A cross-national comparative experiment. *Public Administration Review*, 73(4), 575–586. doi:10.1111/puar.12047
- Gulati, G. J., & Yates, D. J. (2011). Strategy, competition and investment: Explaining the global divide in e-government implementation with policy variables. *Electronic Government: An International Journal*, 8(2/3), 124–143. doi:10.1504/EG.2011.039833
- Kasemsap, K. (2015). The role of cloud computing adoption in global business. In V. Chang, R. Walters, & G. Wills (Eds.), *Delivery and adoption of cloud computing services in contemporary organizations* (pp. 26–55). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-8210-8.ch002

- Kasemsap, K. (2016). The roles of business process modeling and business process reengineering in e-government. In J. Martins & A. Molnar (Eds.), *Handbook of research on innovations in information retrieval, analysis, and management* (pp. 401–430). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-8833-9.ch015
- Kim, C. K. (2007). A cross-national analysis of globale-government. *Public Organization Review*, 7(4), 317–329. doi:10.1007/s11115-007-0040-5
- Kopackova, H., Michalek, K., & Cejna, K. (2010). Accessibility and findability of local e-government websites in the Czech Republic. *Universal Access in the Information Society*, 9(1), 51–61. doi:10.1007/s10209-009-0159-y
- Lee, C. P., Chang, K., & Berry, F. S. (2011). Testing the development and diffusion of e-government and e-democracy: A global perspective. *Public Administration Review*, 71(3), 444–454. doi:10.1111/j.1540-6210.2011.02228.x
- Lee, E. W. Y. (2012). The new public management reform of state-funded social service non-profit organizations and the changing politics of welfare in Hong Kong. *International Review of Administrative Sciences*, 78(3), 537–553. doi:10.1177/0020852312444855
- Li, M. H., & Feeney, M. K. (2014). Adoption of electronic technologies in local US governments distinguishing between e-services and communication technologies. *American Review of Public Administration*, 44(1), 75–91. doi:10.1177/0275074012460910
- Li, Z. (2009). How e-government affects the organisational structure of Chinese government. *AI & Society*, 23(1), 123–130. doi:10.1007/s00146-007-0167-5
- Lin, F., Fofanah, S., & Liang, D. (2011). Assessing citizen adoption of e-government initiatives in Gambia: A validation of the technology acceptance model in information systems success. *Government Information Quarterly*, 28(2), 271–279. doi:10.1016/j.giq.2010.09.004
- Margetts, H., & Dunleavy, P. (2013). The second wave of digital-era governance: A quasi-paradigm for government on the Web. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(1987), 1–17. doi:10.1098/rsta.2012.0382 PMID:23419851
- Meijer, A., & Bekkers, V. (2015). A metatheory of e-government: Creating some order in a fragmented research field. *Government Information Quarterly*, 32(3), 237–245. doi:10.1016/j.giq.2015.04.006
- Moon, M. J., & Norris, D. F. (2005). Does managerial orientation matter? The adoption of reinventing government and e-government at the municipal level. *Information Systems Journal*, 15(1), 3–60. doi:10.1111/j.1365-2575.2005.00185.x
- Norris, D. F., & Moon, M. J. (2005). Advancing e-government at the grassroots: Tortoise or hare? *Public Administration Review*, 65(1), 64–75. doi:10.1111/j.1540-6210.2005.00431.x
- Osborne, D., & Plastrik, P. (1997). *Banishing bureaucracy: The five strategies for reinventing government*. Reading, MA: Addison-Wesley.
- Ozkan, S., & Kanat, I. E. (2011). E-government adoption model based on theory of planned behavior: Empirical validation. *Government Information Quarterly*, 28(4), 503–513. doi:10.1016/j.giq.2010.10.007
- Rana, N. P., & Dwivedi, Y. K. (2015). Citizen's adoption of an e-government system: Validating extended social cognitive theory (SCT). *Government Information Quarterly*, 32(2), 172–181. doi:10.1016/j.giq.2015.02.002
- Rana, N. P., Dwivedi, Y. K., Williams, M. D., & Weerakkody, V. (2015). Investigating success of an e-government initiative: Validation of an integrated IS success model. *Information Systems Frontiers*, 17(1), 127–142. doi:10.1007/s10796-014-9504-7

- Reddick, C. G. (2005). Citizen interaction with e-government: From the streets to servers? *Government Information Quarterly*, 22(1), 38–57. doi:10.1016/j.giq.2004.10.003
- Reddick, C. G., & Anthopoulos, L. (2014). Interactions with e-government, new digital media, and traditional channel choices: Citizen-initiated factors. *Transforming Government: People, Process and Policy*, 8(3), 398–419. doi:10.1108/TG-01-2014-0001
- Roy, M. C., Chartier, A., Crete, J., & Poulin, D. (2015). Factors influencing e-government use in non-urban areas. *Electronic Commerce Research*, 15(3), 349–363. doi:10.1007/s10660-015-9193-4
- Sarantis, D., Smithson, S., Charalabidis, Y., & Askounis, D. (2010). A critical assessment of project management methods with respect to electronic government implementation challenges. *Systemic Practice and Action Research*, 23(4), 301–321. doi:10.1007/s11213-009-9161-9
- Siau, K., & Long, Y. (2009). Factors impacting e-government development. *Journal of Computer Information Systems*, 50(1), 98–107.
- Sindelar, J., Mintz, D., & Hughes, T. (2010). The past is prologue: The Obama technology agenda. *Public Management*, 38(4), 24–27.
- Stier, S. (2015). Political determinants of e-government performance revisited: Comparing democracies and autocracies. *Government Information Quarterly*, 32(3), 270–278. doi:10.1016/j.giq.2015.05.004
- Sun, P. L., Ku, C. Y., & Shih, D. H. (2015). An implementation framework for e-government 2.0. *Telematics and Informatics*, 32(3), 504–520. doi:10.1016/j.tele.2014.12.003
- Taylor, W. J., Marshall, S., & Amiri, S. (2010). The e-evolution of the i-society in the delivery of e-government. *AI & Society*, 25(3), 359–368. doi:10.1007/s00146-010-0288-0
- Tolbert, C. J., & Mossberger, K. (2006). The effects of e-government on trust and confidence in government. *Public Administration Review*, 66(3), 354–369. doi:10.1111/j.1540-6210.2006.00594.x
- Venkatesh, V., Sykes, T. A., & Venkatraman, S. (2014). Understanding e-government portal use in rural India: Role of demographic and personality characteristics. *Information Systems Journal*, 24(3), 249–269. doi:10.1111/isj.12008
- Wang, C. (2014). Antecedents and consequences of perceived value in mobile government continuance use: An empirical research in China. *Computers in Human Behavior*, 34, 140–147. doi:10.1016/j.chb.2014.01.034
- Waters, R. D., Burnett, E., Lamm, A., & Lucas, J. (2009). Engaging stakeholders through social networking: How nonprofit organizations are using Facebook. *Public Relations Review*, 35(2), 102–106. doi:10.1016/j.pubrev.2009.01.006
- Weerakkody, V., Dwivedi, Y. K., & Kurunandana, A. (2009). Implementing e-government in Sri Lanka: Lessons from the UK. *Information Technology for Development*, 15(3), 171–192. doi:10.1002/itdj.20122
- Weerakkody, V., Irani, Z., Lee, H., Osman, I., & Hindi, N. (2015). E-government implementation: A bird's eye view of issues relating to costs, opportunities, benefits and risks. *Information Systems Frontiers*, 17(4), 889–915. doi:10.1007/s10796-013-9472-3
- West, D. (2004). E-government and the transformation of service delivery and citizen attitudes. *Public Administration Review*, 64(1), 15–27. doi:10.1111/j.1540-6210.2004.00343.x
- Willoughby, M., Gomez, H. G., & Lozano, M. A. F. (2010). Making e-government attractive. *Service Business*, 4(1), 49–62. doi:10.1007/s11628-009-0083-5

Wu, J., & Guo, D. (2015). Measuring e-government performance of provincial government website in China with slacks-based efficiency measurement. *Technological Forecasting and Social Change*, 96, 25–31. doi:10.1016/j.techfore.2015.01.007

Xu, H. (2012). Information technology, public administration, and citizen participation: The impacts of e-government on political and administrative processes. *Public Administration Review*, 72(6), 915–920.

Zhang, N., Meng, Q., Guo, X., Yin, C., & Luo, H. (2015). Key e-government issues in China: An empirical study based on the orientation-maturity framework. *Electronic Commerce Research*, 15(3), 407–425. doi:10.1007/s10660-015-9190-7

Zubareva, O. O., & Byelov, S. V. (2015). Electronic government system of Ukraine and a method for increasing its security. *Cybernetics and Systems Analysis*, 51(3), 481–488. doi:10.1007/s10559-015-9739-4

ADDITIONAL READING

Abu-Shanab, E., & Al-Azzam, A. (2012). Trust dimensions and the adoption of e-government in Jordan. *International Journal of Information Communication Technologies and Human Development*, 4(1), 39–51. doi:10.4018/jicthd.2012010103

Alawneh, A., Al-Refai, H., & Batiha, K. (2013). Measuring user satisfaction from e-government services: Lessons from Jordan. *Government Information Quarterly*, 30(3), 277–288. doi:10.1016/j.giq.2013.03.001

Alghamdi, I. A., Goodwin, R., & Rampersad, G. (2011). E-government readiness assessment for government organizations in developing countries. *Computer and Information Science*, 4(3), 3–17. doi:10.5539/cis.v4n3p3

Alomari, M., Woods, P., & Sandhu, K. (2012). Predictors for e-government adoption in Jordan: Deployment of an empirical evaluation based on a citizen-centric approach. *Information Technology & People*, 25(2), 207–234. doi:10.1108/09593841211232712

Aloudat, A., Michael, K., Chen, X., & Al-Debei, M. M. (2014). Social acceptance of location-based mobile government services for emergency management. *Telematics and Informatics*, 31(1), 153–171. doi:10.1016/j.tele.2013.02.002

Armstrong, C. L. (2011). Providing a clearer view: An examination of transparency on local government websites. *Government Information Quarterly*, 28(1), 11–16. doi:10.1016/j.giq.2010.07.006

Bekkers, V. (2012). Why does e-government looks as it looks? Looking beyond the explanatory emptiness of the e-government concept. *Information Polity: The International Journal of Government Democracy in the Information Age*, 17(3/4), 329–342.

Belanger, F., & Carter, L. (2012). Digitizing government interactions with constituents: An historical review of e-government research in information systems. *Journal of the Association for Information Systems*, 13(5), 363–394.

Chatfield, A. T., & Alanazi, J. M. (2015). Collaborative governance matters to e-government interoperability: An analysis of citizen-centric integrated interoperable e-government implementation in Saudi Arabia. *International Journal of Public Administration in the Digital Age*, 2(3), 24–44. doi:10.4018/ijpada.2015070102

Dwivedi, Y. K., Weerakkody, V., & Janssen, M. (2012). Moving towards maturity: Challenges to successful e-government implementation and diffusion. *ACM SIGMIS Database*, 42(4), 11–22. doi:10.1145/2096140.2096142

- Hall, D., Grimaldi, E., Gunter, H. M., Møller, J., Serpieri, R., & Skedsmo, G. (2015). Educational reform and modernisation in Europe: The role of national contexts in mediating the new public management. *European Educational Research Journal*, 14(6), 487–507. doi:10.1177/1474904115615357
- Hung, S. Y., Chang, C. M., & Kuo, S. R. (2013). User acceptance of mobile e-government services: An empirical study. *Government Information Quarterly*, 30(1), 33–44. doi:10.1016/j.giq.2012.07.008
- Joseph, R. C. (2013). A structured analysis of e-government studies: Trends and opportunities. *Government Information Quarterly*, 30(4), 435–440. doi:10.1016/j.giq.2013.05.006
- Karkin, N., & Janssen, M. (2014). Evaluating websites from a public value perspective: A review of Turkish local government websites. *International Journal of Information Management*, 34(3), 351–363. doi:10.1016/j.ijinfomgt.2013.11.004
- Karunasena, K., & Deng, H. (2012). Critical factors for evaluating the public value of e-government in Sri Lanka. *Government Information Quarterly*, 29(1), 76–84. doi:10.1016/j.giq.2011.04.005
- Khalil, O. E. (2011). E-government readiness: Does national culture matter? *Government Information Quarterly*, 28(3), 388–399. doi:10.1016/j.giq.2010.06.011
- Liang, S. W., & Lu, H. P. (2013). Adoption of e-government services: An empirical study of the online tax filing system in Taiwan. *Online Information Review*, 37(3), 424–442. doi:10.1108/OIR-01-2012-0004
- Luna-Reyes, L. F., & Gil-García, J. R. (2011). Using institutional theory and dynamic simulation to understand complex e-government phenomena. *Government Information Quarterly*, 28(3), 329–345. doi:10.1016/j.giq.2010.08.007
- Napitupulu, D. (2014). The critical success factors study for e-government implementation. *International Journal of Computers and Applications*, 89(16), 23–32. doi:10.5120/15716-4588
- Norris, D. F., & Reddick, C. G. (2013). Local e-government in the United States: Transformation or incremental change? *Public Administration Review*, 73(1), 165–175. doi:10.1111/j.1540-6210.2012.02647.x
- Olasina, G., & Mutula, S. (2015). The influence of national culture on the performance expectancy of e-parliament adoption. *Behaviour & Information Technology*, 34(5), 492–505. doi:10.1080/0144929X.2014.1003326
- Rufin, R., Medina, C., & Sánchez, J. C. (2012). Moderating factors in adopting local e-government in Spain. *Local Government Studies*, 38(3), 367–385. doi:10.1080/03003930.2011.636035
- Shyu, S. H. P., & Huang, J. H. (2011). Elucidating usage of e-government learning: A perspective of the extended technology acceptance model. *Government Information Quarterly*, 28(4), 491–502. doi:10.1016/j.giq.2011.04.002
- Sultan, A., AlArfaj, K. A., & AlKutbi, G. A. (2012). Analytic hierarchy process for the success of e-government. *Business Strategy Series*, 13(6), 295–306. doi:10.1108/17515631211286146
- Susanto, T. D., & Goodwin, R. (2013). User acceptance of SMS-based e-government services: Differences between adopters and non-adopters. *Government Information Quarterly*, 30(4), 486–497. doi:10.1016/j.giq.2013.05.010
- Tzikopoulos, A., Manouselis, N., Kastrantas, K., & Costopoulou, C. (2012). An online information system to support blended training of rural SMEs on e-government. *Program: Electronic Library and Information Systems*, 46(1), 123–143. doi:10.1108/00330331211204593

Voutinioti, A. (2013). Determinants of user adoption of e-government services in Greece and the role of citizen service centres. *Procedia Technology*, 8, 238–244. doi:10.1016/j.protcy.2013.11.033

Zhang, H., Xu, X. L., & Xiao, J. Y. (2014). Diffusion of e-government: A literature review and directions for future directions. *Government Information Quarterly*, 31(4), 631–636. doi:10.1016/j.giq.2013.10.013

Zhao, F. (2013). An empirical study of cultural dimensions and e-government development: Implications of the findings and strategies. *Behaviour & Information Technology*, 32(3), 294–306. doi:10.1080/0144929X.2011.644580

KEY TERMS AND DEFINITIONS

Accountability: The obligation of an individual or organization to account for its activities, accept responsibility for them, and to reveal the results in a transparent manner.

Citizen: The person who is entitled to enjoy all the legal rights and privileges granted by a state to the people comprising its constituency.

Electronic Government: The use of information and communications technology (ICT) to improve the activities of public sector organizations.

Government: The act or process of governing, especially the control and administration of public policy in a political unit.

Information Technology: A broad term that includes the development, installation, and use of anything to do with computing and telecommunications.

Public Sector: The part of the economy that is controlled or funded by the government.

Public Service: The service provided or supported by a government or its agencies.

Transparency: The lack of hidden agendas and conditions, accompanied by the availability of full information required for collaboration, cooperation, and collective decision making.

A Model for Connected E-Government in the Digital Age

Qiuyan Fan

Western Sydney University, Australia

INTRODUCTION

Dawes (2008, p.86) defines e-government as ‘the use of information and communication technologies to support public services, government administration, democratic processes and relationships among citizens, civil society, the private sector and the state.’ In the simplest of terms, e-government refers to electronic government or the use of information and communications technologies (ICTs) in the management and delivery of public information and services at all levels of government agencies (Edmiston, 2003). E-government projects can potentially enhance information sharing, aggregation and reuse, and reducing the costs of back-end office operations. In the digital age, governments are using digital technologies and media to enhance delivered services and public engagement (Reddick and Anthopoulos, 2014).

The UN e-government survey states ‘governments around the world are moving forward in e-government development in an effort to enhance public information and service delivery and to improve the efficiency and productivity of government processes and systems.’ (United Nations, 2008, p.23). One of the emerging changes in structures of governance is the growth of connected governance structures. A connected government model provides an efficient and effective way of conducting government business transactions with citizens and businesses and within governments themselves. For the purposes of this study, the term ‘connected e-government’ means thinking and acting in ways that tie the levels, units, and agencies of government together for common

purposes. These purposes include service to citizens, business and other sectors and support for government operations.

The Australian Commonwealth Government established a new service agenda to adopt a whole of government approach to realising responsive government in its 2006 e-government strategy (AGIMO, 2006). To move forward towards the vision of a connected and responsive government, local government plays a crucial role in key areas of service provision of particular importance to local communities. In fact, community participation at the local level is often higher than at a national level in Australia (Shackleton, Fisher & Dawson, 2006).

The study reviews e-government development at the local level in Australia and proposes a connected e-government model that aims to increase the quality of government services and improve the effectiveness and efficiencies of local government operations. This research attempts to provide a framework for understanding how connected e-government at the local level can help achieve this end.

BACKGROUND

While Australian e-government initiatives have received longstanding international recognition (United Nations, 2008), Australian local government lags behind in terms of showing signs of preparedness to move into the next stage of service provision in comparison with the UK local government initiatives in e-government (Mckeown, Teicher & Dow, 2004). As Sarikas

and Weerakkody pointed out, ‘many local governments are lagging behind the national expectations for e-government implementation due to various political, organisational and technical challenges’ (Sarikas & Weerakkody, 2007, p.155).

As the international research suggests, 80% of citizen to government transactions take place with local, not central government (Socitim & I&DeA, 2002). However, the government websites at the local level are typically not as well developed as those at the federal level. Edmiston (2003) conducts two surveys of 2600 municipal and county governments and finds that although the vast majority of local governments have established Web sites, very little had been done to integrate e-government into their daily affairs because of marketing, privacy and funding barriers.

Local city council’s web sites are a prominent product of e-government initiatives. Although the vast majority of the local councils in Australia maintain web sites and provide a good source of information for citizens, most of the council sites are still relatively basic. The wider benefits of G2B (government to business) and G2C (government to citizen) interactions still remain largely unrealized. The literature indicates that e-government falls short for interactions between government, business and citizens (Reddick, 2009). Few have made substantial progress in integrating e-government into their business processes (Mckeown, Teicher & Dow, 2004). A survey of the state of e-government in the Australian local government sector shows only 6% of e-government initiatives focus on the inter-municipal or inter-governmental activities and more than half of the governance professionals surveyed were unable to identify electronic government activities undertaken in other organisations (The E-governance Team, 2004). The research findings indicate that understanding of the connected government approach in the local government sector remains basic and lacks depths (The E-governance Team, 2004).

A recent research on the e-local government development in the great western Sydney (GWS) region shows that the local governments in the

GWS in Australia have not developed truly sophisticated e-government services (Fan, 2011). The research findings suggest that the majority of the local government websites are primarily informational and they provide one way communication of information from government to citizens and none of those local websites have established full portal capacity through which residents can navigate to needed information and services within governments or across governments (Fan, 2011). The researcher concludes that even the leading councils investigated are less than half way to reach their full online service potential and most of the local councils still have a long way to go to achieve connected government (Fan, 2011). A more recent study on the topic shows that most local governments in the region are using mobile platforms and web 2.0 and social media tools to enhance interactivity, transparency, ubiquitous connectivity and openness of local governments but in general the concept of multichannel e-government and the use of web 2.0 and social media tools to create innovative government services are still in their infancy at the local level (Sun et al 2015)

A FRAMEWORK FOR CONNECTED E-GOVERNMENT

How could e-government evolve from a simple website into a fully integrated one? The existing e-government literature considers e-government development an evolutionary phenomenon and has provided various models of e-government maturity or stage models (Layne & Lee 2001, Moon, 2002). For instance, Layne and Lee (2001) suggest a four-stage growth model for e-government starting with establishing an online presence, moving to interactions, transactions and transformation. These stage models offer the advantage of simplicity and can provide guidance for e-government development. However, Shackleton, Fisher & Dawson (2006) argued that the existing e-government models did not ‘truly

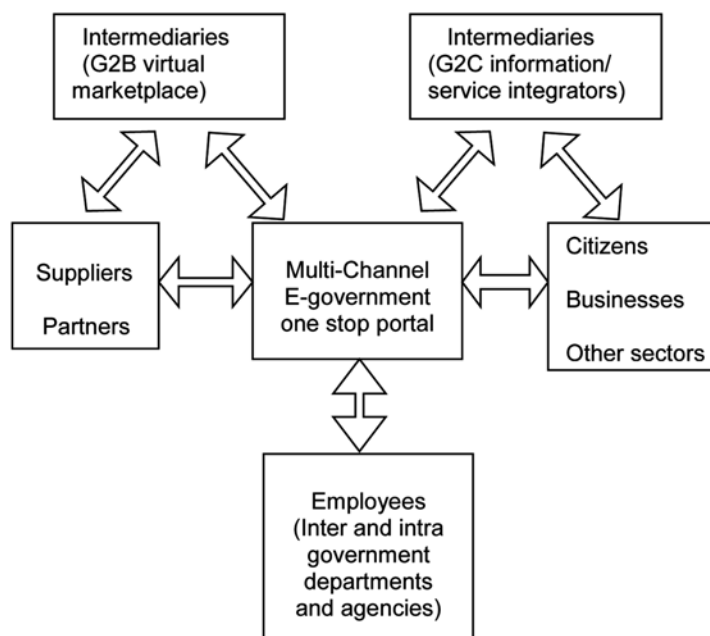
reflect how local governments are implementing successful online web services'. Brown suggests that maturation models play some role but linear e-government progresses sequentially from lower levels to higher ones are less helpful at the local level (Brown, 2007 cited in Dawes, 2008).

Klievink and Janssen (2009) found 'the development and implementation of the final stage of e-government is a major and complex task'. Many local governments do not have the organisational capacity to take the next steps of creating a transaction-enabled citizen-centred e-government. On the other hand, if local government organisations have the resources and capabilities needed to reach the highest growth stage, they do not have to progress through stages of maturity (Klievink & Janssen, 2009). In shifting from a basic e-government to networked e-governance, a more effective e-government development model is clearly required. This research attempts to develop a framework for connected e-government that aims to improve the desired operations that the local government websites are meant to fulfil and efficiency and effectiveness of government operations and to facilitate communication and the coordination

of authorities at different levels of government, within and between organizations. Underlying the concept of connected e-government fits well with a service orientation approach to bring high value information and services to citizens and to improve efficiency and effectiveness of government back end operations. Drawing on the five SOA¹ (service oriented architecture) entry points developed by Carter (2007), the proposed framework identifies the key areas that will need to be addressed by local authorities if they are to harness the full potential of connected e-government.

Connected e-government involves government interaction with citizens (G2C), businesses (G2B), intermediaries and other sectors such as other governments (G2G) and non-government organisations (NGOs). Figure 1 shows the key components of conceptual connected e-government at the local level. It illustrates the key types of stakeholder (suppliers, partners, intermediaries, citizens, businesses and other users as well as employees) interacting with a local council's website through the back-end integration of previously separate systems.

Figure 1. Conceptual connected e-government model



The following sections will discuss the key elements of the framework.

User Centric E-Government Portal

Connected e-government requires a user-centric e-government one stop portal where local authorities are able to bring together previously separate information points and office information. Most local councils in Australia started their e-government initiatives with a focus on providing information and services to the citizens while service delivery platforms remained separate and parallel across various government agencies. Achieving a fully integrated e-government service requires significant reengineering of both front-end and back-end systems and an understanding of associated transformation capabilities and organisational changes.

On the demand side, connected e-government aims to help citizens and other government service recipients to access government information and services more easily. It has become clear that simply putting these traditional government structures online does not meet citizens' expectations. Jaeger and Bertot (2010) state 'It is important for e-government research to focus on the issues of how e-government is meeting the needs of users and the ways in which it is possible to improve user-centred e-government'. It is essential to understand the needs of users of e-government services and resources. In the case of local government, local authorities should engage users in evaluation activities that assess e-government service functionality, usability and accessibility (Editorial, 2006). As Jans and Verdegem point out, e-government will not achieve its goals without a good understanding of the diversity of user groups and their specific needs and capacities of using government online services (Jans & Verdegem, 2006). The user centric e-government portal is best supported through an integrated multichannel approach as there are increasing user demands to access government services from anywhere, anytime through different channels.

Multi-Channel Government Service Delivery and Integration

G

As digital technologies mature, the web portal, mobile and social media channel are becoming main drivers to reach out to a wider user base. Therefore an integrated multichannel approach is essential to greater citizen satisfaction in their interaction with the government in the next stage of e-government development. The importance of multichannel approach to e-government delivery can be illustrated through the view that taking the full advantage of coordinated channels such as web portal, mobile app, mobile portal and social media tools and services will result in government information and services delivery in a cost-efficient and effective manner and a good customer experience (United Nations, 2014). Local governments are aware of the advantages of providing information and services through various channels but generally lack experience and knowledge about implementing social media channels and mobile platforms. The various factors and choices have to be considered carefully while establishing a multichannel approach. One important consideration is to map each digital channel option to meet the needs and expectations of citizens and users. Other factors including cost-effective delivery of e-government services through right channel selection and integration, data security and privacy and management and marketing of the digital channels to reach out to the target user groups should be considered when building a multichannel e-government strategy (Charfield and Alanazi 2015).

The Role of Intermediaries in the Provision of E-Government Services

With increasing levels of government information and services moving online, many local governments do not have organisational capacity needed to interact with citizens and businesses electronically, particularly in areas involving cross government function and complex transactional

services. As Atkinson suggests, 'governments can move beyond engaging with the private sector as e-government vendors and instead empower third party, for profit and non-profit organisations as partners in the provision of e-government services (Atkinson, 2006). However, the potential of intermediaries in the provision of government information and services online has been largely ignored by local authorities despite the fact that they play a key role in e-business and e-commerce.

As shown in Figure 1, G2C third party service integrators have the potential to develop complex online applications to integrate a range of government information and services and to offer transactional based e-government services. Atkinson (2006) argues 'the most cost-effective approach may be to have a third party develop one application rather than have each individual government reinvent the wheel'. On the supply side, the connected e-government model offers an opportunity for governments to reduce time and overhead costs to purchase goods and services as it enables governments to use e-procurement and online group purchasing systems to buy large amounts of goods and services directly from suppliers or a G2B exchange marketplace (Turban et al, 2008). G2B marketplaces are of vital importance as they not only provide an opportunity for local councils to procure supplies at lower prices due to greater pricing transparency and competition but also offer a new channel for suppliers or manufacturers to sell their products. It would be particularly helpful for SMEs who traditionally do not have much chance of doing business with governments.

E-Government Business Process Integration

A one stop portal would require considerable demand for more interdepartmental working within local councils and partnerships between councils and other local agencies. 'A useful way of breaking down the complexity of requirements for a connected e-government system is to con-

sider, in general terms, what government business processes the system must support'. (Jones, et al, 2000) The UN e-government survey identifies a number of business processes required for connected e-government, including:

- The internal operations of an organization that support core processes and are not accessible or visible to the general public.
- Horizontal integration among government departments at same level.
- Vertical integration among local, state and federal government agencies.
- connections among multi-stakeholders (government, private sector, academic institutions, NGOs, and civil society) (United Nations, 2008).

In order to provide a transactional service for businesses or citizens, two or more government organizations need to connect their business processes to allow multiple databases to interact. It is not possible to offer citizens or businesses access to the application, submit information and receive the service without some kind of integration of back-end government systems (Pagano & Cook, 2004). The e-government business process integration not only provides the functionalities, data and shared services needed to enable e-government one-stop portal but also help streamline and improve processes across inter and intra government departments and agencies.

Information Integration

Given that e-government services extend across different organisational boundaries and infrastructures, integrating information and resources stored in these disparate systems is a critical component in the design and implementation of connected e-government (Lakshmi, 2006). Most governments acquire data and information in multiple formats and from multiple sources. 'Integrating and sharing information across traditional government boundaries involves complex interactions among

and with technical and organizational processes' (Pardo, 2007). The key data need to flow between the portal and the back-end systems. The information entered by users on the portal needs to be updated dynamically from back-end database and made available to all other systems within the authority that require the data (Beynon-Davies & Martin, 2004). Information integration offer governments an advantage to share information across systems, across agency boundaries and across levels of government.

To achieve information integration, local authorities need to understand information sources, relationships and business processes and contexts. This understanding can help overcome problems related to the existence of multiple platforms and database, new work processes and organizational changes.

Reusable Services

Governments at all levels have made investments in a variety of systems and developed duplicate functions and services that are available to end-users such as the payment of councils rates and the issuance of a certificate. It is essential to be able to reuse existing services and applications in developing connected e-government. SOA is based on open standards, which makes it possible to create reusable services (Carter, 2007).

The true value of reuse is not only in eliminating duplicate development and maintenances but also supporting standardized services and business processes that can be shared among government agencies and different departments of the government.

Connected IT Architecture

Although connectivity has always been a requirement for e-government, SOA connectivity brings new levels of flexible linkages between government services and throughout IT environment, typically through enterprise service bus (ESB) (Carter, 2007). SOA connectivity creates a highly

connected and interdependent environment where all existing applications and business processes are flowing through existing and new infrastructure without changing them or with minimal rework (Carter, 2007).

FUTURE RESEARCH DIRECTIONS

Further research into connected e-government at the local level is important as local authorities increasingly face the challenge of responding to community demand for more responsive, efficient and effective government. Local governments are also becoming important as an administrative arm for the implementation of federal policy at a local level and the other levels of government are increasingly rely on using the physical and human resources of local government to deliver programs and services for them. Mckeown, Teicher and Dow pointed out that 'Deficiencies in the lack of functionality at the grass-roots level of local government will be a significant impediment to the larger context of aims such as the potential for 'joined-up-government' (Mckeown, Teicher & Dow, 2004). The proposed framework is a conceptual model on which future research might be based. However, empirical study should be carried out to understand the complexity of connected e-government and to increase the success of implementing connected e-government. Further research can help public managers and policy makers shape future direction and identify the capabilities and changes that are needed to realize the full potential of connected e-government.

CONCLUSION

E-government is an opportunity that no government should ignore. Government around the world are moving from basic government online to connected e-government. However, most local governments in Australia appear to be slow to advance their efforts to create functionally connected

and citizen centred web sites. The current state of e-local governments can be best viewed as little more than moving existing public services online. Most of local councils in Australia started their e-government initiatives with a focus on providing information and services to the citizens while service delivery platforms remained separate and parallel across various government agencies. Little of discussion and research work about the next step of connected e-government development in Australia has been focusing on the local level.

A framework for developing more connected and responsive e-government at the local level is of paramount importance. To move forward and develop connected e-government, local councils need to adopt a user-centric focus for the development of e-government services and resources and realise that connected government requires government business process integration and information integration. Reusable services and connected IT architecture are essential characteristics of connected e-government. The proposed framework links to third party efforts (intermediaries), which provides more effective way of developing a more connected e-government by potentially breaking down bureaucratic barriers. The physical and the digital worlds are merging, in business, government, and society. People are becoming accustomed to, and indeed are demanding, instant access to information, to financial transactions, and to government. They are also becoming increasingly accustomed to using a range of transactional channels by means of increasingly portable and mobile devices. No longer are they tied to a desktop computer to conduct online business, whether it is of an official or a social nature. Accordingly, increasingly they are demanding access to local government via the various digital channels. The use of social media and mobile platforms is becoming an increasingly important aspect of e-government and a means of engaging with citizens and delivering improved service. Local governments need to develop an integrated multichannel service delivery approach to connected e-government.

Local governments play an important role in moving forward towards the vision of a connected and responsive e-government. It is crucially important for governments to understand the potential and impact of advanced e-government services at the local level as local authorities increasingly face the challenge of responding to community demand for more responsive, efficient and effective government. This research is significant as it attempts to develop an innovative connected e-government model that aims to enable local governments migrate to a higher level but much empirical research needs to be done to ensure the success of implementation of more connected and responsive e-government at the local level in the future.

REFERENCES

- AGIMO (Australian Government Information Management Office). (2006). *The 2006 e-Government Strategy, Responsive Government: A New Service Agenda*. Accessed on 18 March 2009, <http://www.finance.gov.au/publications/2006-e-government-strategy/index.html>
- Atkinson, R. D. (2006). "Turbo Government": A bold new vision for e-government. Access on 1 March 2010, <http://www.innovationpolicy.org>
- Bertot, J. C., & Jaeger, P. T. (2006). Editorial. (2006). User-centred e-government: Challenges and benefits for government web sites. *Government Information Quarterly*, 23(2), 163–168. doi:10.1016/j.giq.2006.02.001
- Beynon-Davies, P. & Martin, S. (2004). Electronic local government and the modernisation agenda: progress and prospects for public service improvement, *Local Government Studies*, 30(2), 214–229.
- Brown, M. (2007). Understanding e-government benefits: An examination of leading-edge local governments. *American Review of Public Administration*, 37(2), 178–197. doi:10.1177/0275074006291635

- Carter, S. (2007). *The new language of business SOA & Web 2.0*. Upper Saddle River, NJ: IBM Press Pearson.
- Chatfield, A. T., & Alanazi, J. M. (2015). Collaborative governance matters to e-government interoperability: An analysis of citizen-centric integrated interoperable e-government implementation in Saudi Arabia. *International Journal of Public Administration in the Digital Age*, 2(3), 24–44. doi:10.4018/ijpada.2015070102
- Dawes, S. (2008). The evaluation and continuing challenges of e-governance. *Public Administration Review*, 86–102.
- Edmison, K. D. (2003). State and local e-government: Prospects and challenges. *American Review of Public Administration*, 33(1), 20–45. doi:10.1177/0275074002250255
- Fan, Q. (2011). An evaluation and analysis of e-government development by local authorities in Australia. *International Journal of Public Administration*, 34(14), 922–934. doi:10.1080/01900692.2011.615550
- Jaeger, P. T., & Bertot, J. C. (2010). Designing, implementing, and evaluating user-centered and citizen-centered e-government. *International Journal of Electronic Government Research*, 6(2), 1–17. doi:10.4018/jegr.2010040101
- Jans, G., & Verdegem, P. (2006). *Dashgov: developing an e-government dashboard*. Access on 1 April 2010, <http://soc.kuleuven.be/com/mediac/cuo/admin/upload/DashGov%20-%20Developing%20an%20eGovenment%20dashboard.pdf>
- Jones, S., Wilikens, M., Morris, P. & Masera, M. (2000, December). Trust requirements in e-business. *Communications of the ACM*.
- Klievink, B., & Janssen, M. (2009). Realizing joined-up government – Dynamic capabilities and stage models for transformation. *Government Information Quarterly*, 26(2), 275–284. doi:10.1016/j.giq.2008.12.007
- Lakshmi, S. L. (2006). Knowledge management for government to government (G2G) process coordination. *Electronic Government: an International Journal*, 3(1), 18–35. doi:10.1504/EG.2006.008490
- Layne, K., & Lee, J. (2001). Developing fully functional e-government: A four-stage model. *Government Information Quarterly*, 18(2), 122–136. doi:10.1016/S0740-624X(01)00066-1
- Mckeown, T., Teicher, J., & Dow, N. (2004). *Evaluating the value: Comparing local government initiatives in E-Gov in the UK and Australia*. Accessed on 20 March 2009, <http://www.buseco.monash.edu.au/mgt/research/working-papers/workingpapers04pdf.html>
- Moon, M. J. (2002). The evolution of e-government among municipalities: Rhetoric or reality? *Public Administration Review*, 62(4), 424–433. doi:10.1111/0033-3352.00196
- Pagano, C. M., & Cook, M. E. (2004). *The New York State-Local Internet Gateway Prototype Project: Current Practice Research*. Accessed 18 March 2010, http://www.ctg.albany.edu/publications/reports/current_practice
- Pardo, T. A. (2007). Interorganizational information integration: A key enabler for digital government. *Government Information Quarterly*, 10.
- Reddick, C. G. (2009). Factors that explain the perceived effectiveness of e-government: A survey of United States City Government Information Technology Directors. *International Journal of Electronic Government Research*, 5(2), 1–15. doi:10.4018/jegr.2009040101
- Reddick, C. G., & Anthopoulos, L. (2014). Interactions with e-government, new digital media, and traditional channel choices: Citizen-initiated factors. *Transforming Government: People, Process and Policy*, 8(3), 398–419.

Sarikas, O.D., & Weerakkody, V. (2007). Realising integrated e-government services: a UK local government perspective. *Transforming Government: People, Process and Policy*, 1(2), 153–173.

Shackleton, P., Fisher, J., & Dawson, L. (2006). E-government services in the local government context: An Australian case study. *Business Process Management Journal*, 12(1), 88–100. doi:10.1108/14637150610643788

Socitim & I&DeA. (2002). *Local e-government now: A worldwide view*. London: Improvement & Development Agency, Society of Information Technology Management.

Sun, P. L., Ku, C. Y., & Shih, D. H. (2015). An implementation framework for e-government 2.0. *Telematics and Informatics*, 32(3), 504–520. doi:10.1016/j.tele.2014.12.003

The E-governance Team. (2004). *E-governance: Taking the next step at the local level*. Accessed 19 November 2008, http://www.wellington.vic.gov.au/e-governance/project/E-Governance_Project_Proposal.pdf

Turban, E. F., Chung, M., Lee, J. K., & King, D. (2008). *Electronic Commerce: A Managerial Perspective*. Englewood Cliffs, NJ: Prentice Hall.

United Nations. (2008). *UN e-government survey 2008: From e-government to connected government*. Accessed on 1 December 2008, <http://www.google.com.au/search?hl=en&q=UN+2008+survey+e-government&btnG=Google+Search&meta=>

United Nations. (2014). *United Nations E-government Survey 2014*. Accessed on 1 Oct 2015, <http://unpan3.un.org/egovkb/en-us/Reports/UN-E-Government-Survey-2014>

ADDITIONAL READING

Cordella, A., & Tempini, N. (2015). E-government and organizational change: Reappraising the role of ICT and bureaucracy in public service delivery. *Government Information Quarterly*, 32(3), 279–286. doi:10.1016/j.giq.2015.03.005

Jaeger, P. T., Geene, N. N., Bertot, J. C., Perkins, N., & Wahl, E. E. (2012). The co-evolution of e-government and public libraries: Technologies, access, education, and partnerships. *Library & Information Science Research*, 34(4), 271–281. doi:10.1016/j.lisr.2012.06.003

Janssen, M., & Esteve, E. (2013). Lean government and platform-based governance—Doing more with less. *Government Information Quarterly*, 30, S1–S8. doi:10.1016/j.giq.2012.11.003

Knepper, R., & Chen, Y.-C. (2006). ‘Digital government development. In M. Khosrow-Pour (Ed.), *Encyclopaedia of E-commerce, E-government and Mobile Commerce* (pp. 203–209). USA: Information Resources Management Association. doi:10.4018/978-1-59140-799-7.ch034

Nichols, J., & Chen, A. (2006). Coordination of a Service Oriented Architecture. In M. Khosrow-Pour (Ed.), *Encyclopaedia of E-commerce, E-government and Mobile Commerce* (pp. 157–162). USA: Information Resources Management Association. doi:10.4018/978-1-59140-799-7.ch026

Roy, M. C., Chartier, A., Crete, J., & Poulin, D. (2015). Factors influencing e-government use in non-urban areas. *Electronic Commerce Research*, 15(3), 349–363. doi:10.1007/s10660-015-9193-4

Sandoval-Almazan, R., & Gil-Garcia, J. R. (2012). Are government internet portals evolving towards more interaction, participation. *Government Information Quarterly*, 29, S72–S81. doi:10.1016/j.giq.2011.09.004

Weerakkody, V., El-Haddadeh, R., Sabol, T., Ghoneim, A., & Dzupka, P. (2012). Weerakkody, v., El-Haddadeh, R., Sabol, T., Ghoneim, A. & Dzupka, P. (2012). E-government implementation strategies in developed and transition economies: A comparative study. *International Journal of Information Management*, 32(1), 66–74. doi:10.1016/j.ijinfomgt.2011.10.005

Youngblood, N. E., & Mickiewicz, J. (2012). A usability analysis of municipal government website home pages in Alabama. *Government Information Quarterly*, 29(4), 582–588. doi:10.1016/j.giq.2011.12.010

KEY TERMS AND DEFINITIONS

Connected E-Government: Thinking and acting in ways that tie the levels, units, and agencies of government together for common purposes including service to citizens, business and other sectors and support for government operations.

E-Government: Electronic government or the use of information and communications technologies (ICTs) in the management and delivery of public information and services at all levels of government agencies.

Local Governments: In Australia (also known as local councils) are established by state and territory governments to handle local community needs.

ENDNOTE

- ¹ The IBM SOA community defines service oriented architecture (SOA) as ‘business-driven IT architectural approach that supports integrating your business as linked, repeatable tasks or services’.

Presidential Elections Web 2.0

Ramona Sue McNeal

University of Northern Iowa, USA

Lisa Dotterweich Bryan

Upper Iowa University, USA

INTRODUCTION

An important question for any candidate is, “how does he/she reach potential voters?” One method is through making personal appeals known as the “ground war.” This campaign strategy, in the past, involved both phone calls and house-to-house canvassing. This method had taken a backseat to the “air wars” or using mass media such as television or radio to reach voters. Recently, candidates have started reconsidering the ground war. One reason campaigns are reviving this approach to campaigning is that the electorate has become more polarized. Because the political center is shrinking, it no longer seems logical to make broad appeals to the electorate. Instead, reaching out to the base with targeted messages is becoming a central campaign strategy. For example, until relatively recently modern presidential campaigns had spending 70-75% of their war chest on the “air war.” In 2008, however, the Obama campaign only spent 50% of campaign funds on mass media and instead elected to focus more money on the “ground war” (Hershey, 2013).

The ground war has become easier with the widespread use of cell phone. A Pew Internet & American Life Project Survey (2012) found that during the 2012 presidential campaign, approximately 88% of registered voters own a cell phone and made extensive use of it for political activities. The survey also found that smartphone owners were particularly apt to use their cell phone for political activities including fact checking political statements and taking part in political discussion on social network sites. A recent Pew Research

Center study found almost two-thirds of those surveyed reported owning a smartphone with twenty-three percent of smartphone owners reporting that they used their phone to donate money to a political or charitable cause (2015). The fact that citizens are using their cell phones for activities ranging from fact checking candidate statements in real time to presenting their own personal views on social network sites has not been lost on candidates. During the 2012 election, candidates began adopting a number of strategies that used cell phones to reach the electorate including integrating mobile apps into their campaign strategy for connecting with the public. Finding avenues for reaching out to constituents through mobile apps have continued with the 2016 presidential election. One example is Snapchat, a mobile app that allows for the sharing of photos and videos. Presidential primary hopefuls including Senator Rand Paul (R), Governor Scott Walker (R) and Governor John Kasich (R) each experimented using the app to release campaign ads (Roth, July 30, 2015).

The widespread use of cell phones and an increasingly polarized electorate are two pieces of the puzzle that help explain why candidates are allocating more of their resources to the ground war. The final puzzle piece is social media. When the 2008 Obama campaign was redirecting its resources to more targeted messaging, the Internet was moving from Web 1.0 to Web 2.0 with the development of Facebook (2004), YouTube (2005), and Twitter (2006). Although his use of social media for communicating with potential voters was considered innovative in 2008, it

became common place by the 2012 presidential election. For example, in 2012 President Obama announced his reelection bid with a tweet and a YouTube video while former governor Mitt Romney announced he was forming an exploratory committee by posting it on Facebook, in a tweet and in a video posted on his website. Currently there are a number of narratives coming out of the 2016 presidential primaries. One of the more dominant themes is the importance of social media. This can be attributed, in part, to Donald Trump. While candidates in previous elections used social media as a secondary tool for communicating with the electorate, Donald Trump has been using Twitter as a direct means of communication with voters. Others candidates have followed his lead by increasing the use of social media including Facebook and Twitter as part of their campaign (Lang, April 5, 2016).

What do we know about the effect of these trends on voter turnout? More Americans are now turning to social media for campaign news and smartphones are playing an important role in this phenomenon. These changes in voter behavior have not gone unnoticed by candidates who are adopting “ground war” strategies that utilize the widespread usage on social media and adoption of smartphones. Do the changes in how candidates communicate with the electorate encourage voter turnout? To examine this question, this chapter presents an examination of recent presidential elections and the role social media in mobilize voters.

BACKGROUND

Early researchers (e.g. Davis et al., 2009) predicted that the Internet would change political campaigns in four important ways. The first is through allowing the campaign to transmit information to others. Websites could be used to distribute information that would have previously been distributed through campaign flyers and brochures. This could make it easier for voters to obtain political information. Secondly, the Inter-

net can be used to gather information including background information on opponents. This can help campaign staffers better prepare attack ads against their candidate’s opponents as well as better prepare for attacks directed at their own campaign. The Internet could also increase the ability of the public to interact with the campaign. For example, a candidate can use blogs or social network sites such as Facebook to both communicate with supporters and allow them a means of voicing their opinions. By provided an avenue for voters to interact with a candidate’s campaign and provide input, the campaign can get a better sense of what issues are important to voters and also allow supporters to feel like they are part of the candidate’s team. This could increase loyalty to the candidate. Finally, the Internet might be utilized to mobilize or encourage political participation by citizens. Campaigns could use a number of strategies to encourage participation including sending email and text messages with reminders to register and vote.

Despite these predictions, candidates were slow to include the Internet as part of their overall campaign strategy. The 2004 presidential election was the first to use the Internet in a significant manner for generating online donations, mobilize voters and encourage volunteer efforts. One notable change in election strategy during this period was the advancement of the “ground war”. Because the 2004 election was expected to be highly competitive, both parties worked to get-out-the-vote (GOTV) or contact voters either face-to-face or over the phone. The 2004 presidential election forced campaigns to address a glaring weakness in how they used the Internet. While citizens were now able to obtain political information easier, as well as donate money and volunteer to help campaign, they had chosen to seek out these activities. Campaigns created websites that facilitated these actions but did not use the Internet in a meaningful way to reach out to supporters.

The dynamics of the 2004 election forced campaigns to tap into the potential of the Inter-

net that they had largely ignored. Advances in technology existed allowing parties to combine commercial data sets, Census information and voter registration records to help predict voters, party identifiers and independents that might lean more toward one party or another. This allowed parties to send volunteers into the fields supplied with voter contact information. Aiding in these GOTV efforts was the Help America Vote Act of 2002, which required states to maintain electronic copies of which citizens voted in previous federal elections (Nickerson & Rogers, 2014). The Republican Party made better use of these resources with the creation of a national Republican voter database named *Voter Vault*. They used a series of field experiments during the 2002 congressional elections in devising a 72-hour strategy for the 2004 presidential election. Volunteers and staff members were trained for a final 72-hour push during the election. They went door-to-door in competitive states in a well-organized drive to get voters to the polls. Although the Kerry campaign also had a canvassing effort, it was less organized and primarily conducted by outside groups such as unions (Hershey, 2013).

PRESIDENTIAL CAMPAIGNS AND WEB 2.0

The ground game reemerged as a result of the tightness of the 2004 election and the success of the Republican GOTV effort would make it a blueprint for running future presidential campaigns. Candidates in 2008 focused on the ground war and online donations. The recent development of online social network services would play an important role. Although all of the major candidates had incorporated web-based tools into their election campaign, it was then Senator Barack Obama whose team made the greatest use of online tools. His campaign used Facebook and other social network sites such as Digg, Flickr, LinkedIn and MiGente. It even paid to have billboards placed within popular computer games containing ads for

Obama. In addition, his campaign made extensive use of YouTube to post campaign ads and encouraged supporters to develop and post their own (Bimber, 2014). Campaigns used websites to sign up volunteers and solicit donations. Supporters who supplied their email address also received additional requests for donations, volunteering efforts and reminders to vote. The Obama campaign also collected cell phone numbers from individuals who then received between 5 and 25 messages a month. Facebook was an important avenue for communication between candidates and supporters. It was also a means of expanding its messages to more individuals. Individuals who “liked” a candidate’s Facebook page were asked to forward campaign messages to others on their contact list (Hershey, 2013). Finally, Web 2.0 was instrumental in expanding the ground war. In addition, to using email, social network sites and text-messages to remind voters to turn out, campaigns increased their canvassing efforts.

Rather than waiting until the last few weeks of the election to put in place canvassing efforts, the Obama team set up over 700 field offices to direct canvassing operations. Volunteers and paid staff went door-to-door contacting members of the electorate armed with background information on each person being contacted. This permitted for targeted messages to be used when talking to individual voters. A database constructed by the Obama team was used to determine which households to canvass based on whether they were predicted to be members of his base or swing voters. The McCain campaign had a similar strategy, but it was on a smaller scale. One important development resulting from this ground game was that individuals normally ignored by election campaigns were being contacted and asked to vote (Bimber, 2014). This was an encouraging sign for those who hoped that the Internet could have a positive impact on voter turnout.

With the spread of smartphones, politicians found ways to incorporate smartphones into their campaigns through mobile apps during the 2012 election. Political candidates began using them

for newsletter registration, micro-donation, “a payment or donation a person has charged to their mobile device bill,” and polling (Pessin, 2011). Both presidential candidates, President Barack Obama (D) and former Governor Mitt Romney (R), made smartphone apps available to voters during their campaigns in 2012. Those who downloaded the “Obama for America” app had access to registered Democrats near their address, the dates and locations for Obama-related events, and the ability to donate or volunteer for the campaign. Those who downloaded the “Mitt’s VP” app were promised to be among the first to know who Romney’s vice presidential nominee was. Both apps were free to download for iPhone and Android users (Cline, 2012).

In addition to adding smartphones to their list of online election tools, candidates updated strategies used during the 2008 election. For the Obama campaign that meant improving upon its existing canvassing database. The updated database, *Narwhal*, combined Census data, precinct level data, commercial datasets and information gathered through canvassing, phone calls, etc. (Nickerson & Rogers, 2014). The Romney campaign also created a dataset “*Orca*” to help with GOTV efforts. Both candidates aggressively incorporated social network tools into their campaigns including Facebook, YouTube, Hulu and Twitter. In addition, these tools were used in new ways. For example, the Obama campaign sent supporters lists of their Facebook friends in swing states and asked them to contact these friends and remind them to vote (Bimber, 2014). Despite the efforts of both candidates, voter turnout dropped in 2012. This was due, in part, to lack of enthusiasm by the voters. Because of issues including the economy, voters were not as hopeful as they were in 2008. Another reason is that the candidates adjusted their campaign to target those most likely to turn out. Databases like *Narwhal* can be a double-edged sword. They can be used by candidates to launch more inclusive GOTV drives of the past or can be used to target voters to canvass based on likelihood of volunteering, donating and voting.

In 2012, turnout was expected to be lower and campaigns adjusted their strategy based on their ability to profile voters (Bimber, 2014).

THE 2016 SOCIAL MEDIA ELECTION

During the 2016 presidential primary, candidates including Hillary Clinton, Bernie Sanders, and Donald Trump continued to experiment with strategies developed during the 2008 and 2012 presidential elections including offering free app downloads to Apple and Android smartphone owners. Those who downloaded the Bernie Sanders 2016 app through Apple were able to read news articles, campaign updates, Facebook posts, and tweets in addition to the candidate’s positions on issues. Users could communicate with Sanders through pictures, text messages, voice messages, and video (Group, Inc., 2016a). Hillary Clinton, who did not utilize digital technology to the extent Obama did during the 2008 presidential campaign (Vega, 2015), has now been connecting to her followers through the Hillary Clinton 2016 app which has similar features to the Sanders’ app (Groupe, Inc., 2016b). The Trump 2016 app included similar features along with several others. Those who download this app can donate money, sign up to volunteer for the campaign, and subscribe to push notifications to find out where “the Donald” will next appear and what he is doing. Users can also read Donald Trump quotes of the day and participate in polls to see if the campaign is on track for winning the presidency (Mercer, 2016).

In addition to their extensive use of mobile apps, Hillary Clinton, Bernie Sanders, and Donald Trump have also used Twitter to communicate with supporters. According to the website Statista (2016), Clinton has 6.46 million followers, Sanders 2.36 million followers, and Trump has 8.56 million followers as of May 31, 2016. These candidates have also used tweets to address comments made by opponents and the media. Trump, in particular, has gained attention for his usage of Twitter. He has used it mainly to thank voters in

various states, but has also used it to attack other candidates and journalists. For example, he called Clinton crooked, and criticized Fox News journalist Megyn Kelly after she asked him about comments he had previously made during a presidential debate in August of 2015. He called her a bimbo twice in tweets and stated that he liked her show “The Kelly File” better without her on it (Zaru, 2015; Twitter, 2016).

The 2016 presidential primary has proven to be a paradox. On the one hand, presidential hopefuls have used the 2008 and 2012 as a textbook on how to wage a social media election, albeit with more advanced mobile apps and more widespread cellphone usage. On the other hand, they are using social media in a way that previous campaigns have shied away from. Davis et al. (2009) argued that the Internet would change political campaigns in four ways and until the 2016 presidential primaries it appeared that only three would happen to any great extent. The one prediction that had not materialized was that the Internet would increase the ability of the public to interact with the campaign. While it was argued that using the Internet to increase communication with supporters could lead to greater candidate loyalty, campaigns have been hesitant to take advantage of this capability. It conflicts with the desire to maintain control over the campaign message. Candidates have been using the Internet to mobilize voters in the several presidential cycles but they have been less willing to use its potential to engage citizens (Stromer-Galley, 2014). As a consequence, social media use in previous elections has been underutilized and structured in a manner that limited true interaction. This has changed with Donald Trump’s unprecedented use of Twitter as a campaign tool. Through posting tweets that are controversial, he has been able to dominate the news cycle. Other candidates have altered their social media strategy in order to garner media coverage.

It is still early in the 2016 presidential election. It is not yet known how this change in social media usage will impact turnout in the general election. In

addition, it is not known how the two major parties will work to mobilize voters in November. There are early signs that mobilization strategies might look like the 2008 presidential election. During this election, individuals who were normally ignored by candidates were courted by candidates during the primaries. Two candidates (Bernie Sanders and Donald Trump), in particular, have reached out to unlikely voters in attempts to obtain party nominations. Like other election years, who to target might have been a strategic decision. There has been an anti-establishment mood among voters during this election cycle. Being an outside candidate in a field of establishment hopefuls (particularly during the Republican primaries) may have had a strategic advantage given the public mood. It is still too early to know if this will foreshadow mobilization efforts during the general election.

SOLUTIONS AND RECOMMENDATIONS

Early on there were optimistic predictions that the Internet could increase voter turnout. It was argued that the Internet could be used to mobilize voters. In addition, social media could be used to increase the ability of the public to interact with candidates (Davis et al., 2009). Since the 2004 presidential election, campaigns have been using the Internet to mobilize voters but they have been doing so in a strategic manner. Whether they reach out to a general audience or target likely voters has depended on the dynamics of the election. The 2016 presidential primary is our first example of candidates using social media in more than a secondary tool for interacting with voters and we do not yet know how this change in social media usage will impact turnout. Nevertheless, neither of these campaign activities appears to be the answer for significantly improving voter turnout.

How do we get more voters to the polls? The answer may be in making elections more competitive. Research (Cox & Munger, 1989) finds

that individuals are more likely to feel their vote can impact the election when races are close. Furthermore, close elections are associated with more information and media coverage. Finally, both parties and interest groups tend to increase their GOTV efforts and contact more voters when elections are closer (Key, 1949). This is consistent with the findings from the 2004, 2008 and 2012 elections. The campaigns adjusted their ground game based on expected turnout in the election. To achieve closer elections may require changes in campaign finance laws and rules for redistricting. For example, state legislatures could be replaced with nonpartisan redistricting committees to achieve more balanced voting districts.

FUTURE RESEARCH DIRECTIONS

The focus of this study was the relationship between social media and presidential elections. After several presidential cycles, trends have emerged in how campaigns are taking advantage the widespread adoption of cell phones and social media. Campaigns are using these advances in telecommunication technology to mobilize voters but they are doing so strategically. On the other hand, the 2016 presidential primaries were the first where presidential candidates used social media to interact with voters in more than a superficial manner. It is possible that this may lead to greater turnout in November. Future research should be directed at determining the impact of social media use on voter turnout in the 2016 general election. In addition, research should explore whether social media usage expands in future elections or if the 2016 presidential election is an anomaly. Candidates have been hesitant in previous elections to use social media to interact fully with voters. Depending on how the 2016 presidential election plays out, candidates might try to return social media usage back to the more structured model of earlier election. It is also possible that now that the “genie has been left out of the bottle” they cannot return to the controlled social media usage of previous elections.

CONCLUSION

This chapter began with the question of how has social media changed presidential campaigns. In particular, has social media helped to increase voter turnout? An examination of elections following advances in social media and the widespread adoption of smartphones has shown that they have helped to bring back the “ground war.” Candidates are spending more of their war chest on making personal appeals to the electorate and less on mass media or the “air wars.” At first glance, using the Internet to facilitated GOTV drives appears to hold promise for increasing voter turnout. This was true in 2008; there was an increased voter turnout along with an effort made by campaigns to contact those individuals who are normally overlooked. This was not however the same story for the 2012 presidential election. The election was characterized by less enthusiasm among voters and campaign strategies reflected this. Voters were contacted based on predictions of how likely they were to vote, donate and volunteer. While it is true the ground game can increase voter turnout, the ultimate goal of a campaign is to get its candidate elected. Depending on the dynamics of an election, campaigns may opt for an inclusive ground game or one that is more narrowly targeted.

Although the return to the “ground war” and its influence on voter turnout has been well documented, there are other ways that social media can impact turnout that are less understood. Candidates have historically been slow to incorporate the Internet into their campaigns in part because of fear that it can lead to lose of control over the campaign message. This is particularly true when using social media in a manner that could allow interaction between the candidate and supporters. By allowing supporters to voice their opinions through social media, campaigns may increase voters’ feeling of being a member of the team and increase allegiance to the candidate. Until the 2016 presidential primaries, candidates had structured social media use in a manner that gave an illusion of interaction. True interaction was deemed too risky. This changed as a result of the surprising

success of presidential hopeful Donald Trump during the Republican primaries. This success has been in part because of his strong social media presence. In order to compete, other candidates have left their comfort zones and increasingly taken part in less structured social media use. It is too early to predict the consequences of this new era in social media campaigning. Turnout in the 2016 general election may give us a better sense of whether social media is part of remedy for lagging voter engagement.

REFERENCES

- Bimber, B. (2014). Digital media in the Obama campaigns of 2008 and 2012: Adaptation to the personal political communication environment. *Journal of Information Technology & Politics*, 11(2), 130–150. doi:10.1080/19331681.2014.895691
- Cline, S. (2012). A political app for that. *U.S. News Digital Weekly*, 4(32), 25.
- Davis, R., Baumgartner, J., Francia, P., & Morris, J. (2009). The Internet in U.S. Election campaigns. In A. Chadwick & P. Howard (Eds.), *Routledge handbook of Internet politics* (pp. 13–24). New York: Routledge.
- Groupe, Inc. (2016a). *Sanders 2016, mobile application software*. Retrieved May 31, 2016, from <http://itunes.apple.com>
- Groupe, Inc. (2016b). *Clinton 2016, mobile application software*. Retrieved May 31, 2016, from <http://itunes.apple.com>
- Hershey, R. (2013). *Party politics in America* (15th ed.). New York: Pearson.
- Key, V. O. Jr. (1949). *Southern politics: In the state and nation*. New York: Knopf.
- Lang, M. (2016). *2016 Presidential media circus, Is social media the cause?* Retrieved May 7, 2016, from <http://www.govtech.com/social/2016-Presidential-Election-Circus-Is-Social-Media-the-Cause.html>
- Mercer, S. (2016). *Trump'16, mobile application software*. Retrieved May 31, 2016, from <http://itunes.apple.com>
- Nickerson, D., & Rogers, T. (2014). Political campaigns and big data. *The Journal of Economic Perspectives*, 28(2), 51–73. doi:10.1257/jep.28.2.51
- Pessin, A. (2011). Smartphone mobile apps as key campaign drivers for 2012. *Campaigns & Elections*, 32(304), 11–12.
- Pew Internet & American Life Project. (2012). *The state of the 2012 election — mobile politics*. Retrieved November 19, 2012, from <http://pewinternet.org/Reports/2012/Election-2012-Mobile.aspx>
- Pew Internet & American Life Project. (2013). *Cell phone activities 2013*. Retrieved March 12, 2014, from <http://www.pewinternet.org/2013/09/19/additional-demographic-analysis/>
- Pew Research Center. (2015). *The smartphone difference*. Retrieved May 31, 2016, from <http://pewinternet.org/2015/04/01/us-smartphone-use-in-2015/>
- Roth, S. (2015). *Candidates connecting with supporters, 10 seconds at a time: Where else can you see Rand Paul with a chainsaw?* Retrieved July 30, 2015, from http://www.huffingtonpost.com/entry/snapchat-election-2016_55ba5983e4b0af35367a8973
- Statista. (2016). *Number of Twitter followers of 2016 U.S. presidential candidates, as of May 31, 2016*. Retrieved June 5, 2016, from <http://www.statista.com/statistics/509579/twitter-followers-of-2016-us-presidential-candidates/>

Stromer-Galley, J., & Baker, A. (2004). Joy and sorrow of Interactivity on the campaign trail: Blogs in the primary campaign of Howard Dean. In A. Williams & J. Tedesco (Eds.), *The Internet election: Perspectives on the web campaign 2004* (pp. 111-131). New York: Rowman & Littlefield Publishers, Inc.

Stromer-Galley, J. (2014). *Presidential Campaigning in the Internet Age*. New York: Oxford University Press. doi:10.1093/acprof:oso/9780199731930.001.0001

Twitter. (2016). Retrieved June 5, 2016 from <https://twitter.com/realDonaldTrump/media>

Vega, T. (2016). *The next political battleground: Your phone*. Retrieved May 31, 2016, from <http://www.cnn.com/2015/05/29/politics/2016-presidential-campaigns-mobile-technology/>

Zaru, D. (2015). *Donald Trump's new Twitter tirade against Megyn Kelly*. Retrieved June 5, 2016, from <http://www.cnn.com/2015/08/25/politics/donald-trump-megyn-kelly-tweets/>

ADDITIONAL READING

Abramowitz, A. (2010). *The disappearing center: Engaged citizens, polarization, and American democracy*. New Haven: Yale University Press.

Alvarez, R., & Hall, T. (2004). *Point, click and vote: The future of Internet voting*. Washington, D.C.: Brookings Institute Press.

Bannon, D. (2005). Relationship marketing and the political process. *Journal of Political Marketing*, 4(2/3), 85-102.

Berinsky, A. (2005). The perverse consequences of electoral reform in the United States. *American Politics Research*, 33(4), 471-491. doi:10.1177/1532673X04269419

Bimber, B., & Davis, R. (2003). *Campaigning online: The Internet in the U.S. elections*. New York: Oxford University Press.

Bode, L. (2012). Facebooking it to the polls: A study in online networking and political behavior. *Journal of Information Technology & Politics*, 9(4), 352-369. doi:10.1080/19331681.2012.709045

Bode, L., Vraga, E., Borah, P., & Shah, D. (2014). A new space for political behavior: Political social networking and its democratic consequences. *Journal of Computer-Mediated Communication*, 19(3), 414-429. doi:10.1111/jcc4.12048

Carlisle, J., & Patton, R. (2013). Is social media changing how we understand political engagement? An analysis of Facebook and the 2008 presidential election. *Political Research Quarterly*, 66(3), 883-895. doi:10.1177/1065912913482758

Carpenter, C. (2010). The Obamamachine: Techno-politics 2.0. *Journal of Information Technology & Politics*, 7(2/3), 216-225. doi:10.1080/19331681003765887

Davis, R., & Owen, D. (1998). *New media and American politics*. New York: Oxford University Press.

Garrett, R. (2009). Echo chambers online?: Politically motivated selective exposure among Internet users. *Journal of Computer-Mediated Communication*, 14(2), 265-285. doi:10.1111/j.1083-6101.2009.01440.x

Gerodimos, R. (2012). Online civic attitudes and the limits of civic consumerism. *Information Communication and Society*, 15(2), 217-245. doi:10.1080/1369118X.2011.572983

Hetherington, M. (2001). Resurgent mass partisanship: The role of elite mass polarization. *The American Political Science Review*, 95(3), 619-631. doi:10.1017/S0003055401003045

Hindman, M. (2005). The real lessons of Howard Dean: Reflections of the first digital campaign. *Perspectives on Politics*, 3(1), 121-128. doi:10.1017/S1537592705050115

Howard, P. (2006). *New media campaigns and the managed citizen*. Cambridge: Cambridge University Press.

Klofstad, C. (2007). Talk leads to recruitment: How discussion about politics and current events increase civic participation. *Political Research Quarterly*, 60(2), 180–191. doi:10.1177/1065912907301708

Macafee, T. (2013). Some of these things are not like the others: Examining motivations and political predispositions among political Facebook activity. *Computers in Human Behavior*, 29(6), 2766–2775. doi:10.1016/j.chb.2013.07.019

Perlmutter, D. (2008). *Blogwars: The new political battleground*. New York: Oxford University Press.

Rash, W. Jr. (1997). *Politics on the net: Wiring the political process*. New York: W.H. Freeman.

Robertson, S., Vatrappu, R., & Medina, R. (2010). Off the wall political discourse: Facebook use in the 2008 U.S. presidential election. *Information Polity*, 15(1/2), 11–31.

Schwartz, E. (1996). *Netactivism: How citizens use the Internet*. Sebastopol, CA: Songline Studios.

Stimson, K. (2013). Voting goes mobile. *Capitol Ideas Magazine*, 56(4), 32–33.

Stromer-Galley, J. (2000). On-line interaction and why candidates avoid it. *Journal of Communication*, 50(4), 111–132. doi:10.1111/j.1460-2466.2000.tb02865.x

Towner, T. (2013). All political participation is socially networked? New media and the 2012 election. *Social Science Computer Review*, 31(5), 527–541. doi:10.1177/0894439313489656

Zúñiga, G. G., Jung, N., & Valenzuela, S. (2012). Social media use for news and individuals social capital, civic engagement and political participation. *Journal of Computer-Mediated Communication*, 17(3), 319–336. doi:10.1111/j.1083-6101.2012.01574.x

KEY TERMS AND DEFINITIONS

Air Wars: Using mass media such as television or radio to reach voters.

Facebook: An online social network service.

Ground War: More personalized strategies for mobilizing voters that includes activities such as door-to-door canvassing and phone calls.

Mobile App: A software application created to run on mobile devices.

Smartphone: A cell phone with advanced capabilities.

Snapchat: A mobile app that allows for the sharing of photos and videos that disappear within a short period of time.

Social Media: Electronic communication sources that allow users to network with friends to share a wide variety of content.

Twitter: A microblogging service.

Project Management in Government



Shauneen Furlong

University of Ottawa, Canada & John Moores Liverpool University, UK

INTRODUCTION

The case for 21st century government success continues unabated; impatient stakeholders are more demanding; people, process, and content are profoundly impacted; opportunity is rampant - but so is risk and complexity; still, transformational eGovernment looks to leadership, integration, change, and innovation through enhanced project management.

Enhanced project management has made progress in advancing eGovernment through a holistic manner so that project activities are fully integrated with on-going operational activities: all with an emphasis on measurable results and outcomes.

This paper builds upon previous publications that there is another dimension to the transformational eGovernment navigation tool-kit that coalesces with enhanced project management and creates a multi-dimensional approach to transformation; it is existential change leadership that focuses on human mindset behaviour along with critical leadership, integration, change, and innovation.

eGovernment, and especially transformational eGovernment, progress remains slow and halting and shackled to time-honoured approaches to project management, especially in the information communication technology (ICT) domain. eGovernment is the traditional, transactional, and service focused improvement on Government operations through the application of ICTs, whereas transformational eGovernment encompasses the reform and modernization of the business process reengineering opportunities and enterprise-wide reform, as well as what and how the government

achieves its mandate. Ineffective project management is one of most significant reasons for transformational eGovernment failure (Aikens, 2012b; Misuraca, 2009).

There are a number of reasons and examples for transformational eGovernment project failures, including the lack of capacity to manage unanticipated transparent and concealed organizational opposition; the inability to effectively and precisely identify current, changing, disparate, and conflicting key information requirements; and lack of insight into the obstacles in obtaining parochially coveted information. These are in addition to a review of literature (developed and developing countries) that outlines the most common issues and problems that cause eGovernment failure to be cultural barriers, infrastructure, resources, socio-economic barriers, security and privacy, and e-integration (Zhao, 2012). Previously, in 2007, a list representing 99 retrospectives in 74 organizations revealed the 36 most common mistakes in information technology (IT) management, with poor estimation/scheduling, ineffective stakeholder management, and insufficient risk and planning leading the list (Nelson, 2007).

Enhanced project management focuses on project integration and change leadership to respond to these mistakes and to specific eGovernment uniqueness because of the broad and disparate goal requirements, the multidimensional and conflicting policy and delivery responsibilities, the dearth of knowledge transfer due to the lack of a “whole of Government” culture, and the complex and competing interest and security concerns of disparate stakeholders (Sarantis, Smithson, Charalabidis, & Askounis, 2010).

BACKGROUND

The eGovernment literature introduces a paradigm that examines the digital government evolution and includes a discussion whether digitization, when applied to internal structures and external relationships, changes them or not (Janowski, 2015). This is critical in determining expectations for change and the degree and impact from the application of technology. This paper builds upon previous research and is concerned with the transformational outcome and thereby attempts to assess the effectiveness of the digitization to transform structures and relationships.

eGovernment studies note that stakeholder disappointment is reported as a root problem that causes many unsuccessful projects (Eslerod, 2016). Other authors and papers as recent as 2016 also state that the management literature falls short in analyzing the impact of project management. Specifically the literature does not make the linkages and connections between stakeholders and the information needed to assist project managers in analyzing and prioritizing the challenges that confront them (Van Offenbeek, 2016).

In addition to this author's research about a compendium of ten eGovernment synergistic challenges and barriers that could be mitigated through enhanced project management, this paper builds upon other research, including the following ten principal challenges concerning eGovernment project management: human resources, work milieu, relations within and across organizational boundaries, project failure impact, goals definition, project dimensions, planning horizon, best practices, legal and regulatory issues, and politics driven nature (Sarantis, Smithson, Charalabidis, & Askounis, 2010).

Internationally, there has been a high and critical failure rate related to IT solutions (Aikins, 2012b; Fraser, 2006). More recently the failure in IT solutions that was the bane of transactional processing is now appearing in eGovernment initiatives (Heeks, 2008; Arif, 2008; Janowski, Estevez, & Ojo, 2007; Aikins, 2012b). eGovern-

ment failures are often hushed up (Heeks, 2008) and as Misuraca (2009) points out, the majority of eGovernment projects are failures as high as 70-80% and are not meeting the breakthrough expectations. Failures are costly; as per Irani, Al-Sebie, & Elliman (2006), the United Kingdom Parliamentary Office of Science and Technology reported that cancelled or over-budgeted eGovernment projects were greater than 1.5 billion British pounds. It is a mistake to automatically link IT investment with fiscal prudence even if eGovernment advocates regularly assert that IT investment in restructuring government programs and practices makes programs more cost-effective and affordable (Longford, 2001).

There are a number of reasons for the lack of transformational eGovernment success, including unanticipated organizational opposition, difficulties in communicating requirements, and obstacles to obtaining information from different government departments and agencies (Kamal, Weerakkody & Irani, 2011). However, there is some support for the belief that one of the most significant reasons for transformational eGovernment failure is ineffective project management (Aikins, 2012b; Misuraca, 2009). The literature refers to the dearth of peer-reviewed information on the effective role of project management, including front-end or early stages of decision making in project management (Samset 2016), and its impact on transformational eGovernment project success even though there are non-peer-reviewed business publications and country audits (British Computer Society, 2004; Fraser, 2006) that identify ineffective project management as an important cause of ICT failure.

The literature review finding is that ineffective project management is a leading cause of eGovernment failure. And the reason for this ineffectiveness is a result of the use of the traditional project management methodologies that do not meet the demands of transformational eGovernment for results, accountability, and problem solving.

This paper supports the importance of project management as a holistic discipline as opposed

to an examination of the specific criteria and iterative approaches used in developing project management methodologies. Project management science is described in the North American focused PMBOK (PMI 2008; PMI 2013) and the European focused PRINCE2 (2009). The fundamental scientific codification sets out project management concepts in terms of project processes and project knowledge areas. By cycling through the processes and knowledge areas, project management spans the full array of information that must be merged to develop an eGovernment project. PMBOK 2013, the 5th edition, describes the process groups to be initiating, planning, executing, monitoring and controlling, and closing. The knowledge areas are integration, scope, time, cost, quality, human resources, communications, risk, stakeholders, and procurement.

There is also abounding evidence that creeping incremental resourcing systems cannot keep pace with the escalating resource demand brought about by radical innovation (Lettice & Thomond, 2008). So enhanced project management must develop more effective processes, techniques, and tools in the key project areas of risk, change, and stakeholder management to reduce the need for increased resource investment in disruptive radical innovation.

The examination of traditional project management methods is also important to address project management effectiveness in today's complex environment. Generally, the project management methodology follows a linear process model (Sommer, Dukovska-Popovska, & Steger-Jensen, 2013) even though this is blamed in product development circles for low performance and market failure. The alternative to a linear model is an integrated approach which is more holistic and includes other domains that involve a high degree of uncertainty, such as project governance, organizational structure, activity dependencies, the need for information exchange, and human resource problems in change impacts.

Managing all of these domains requires working simultaneously and in an integrated way called

the "rugby approach" (Sommer, Dukovska-Popovska, & Steger-Jensen, 2013). Understanding and supporting the project's contribution in fulfilling the organization's mandate - as opposed to merely focusing on meeting the traditional time, budget, and performance goals - is also becoming much more critical (Patanakul & Shenhar, 2012). It has been said that project management methodologies have helped to change the course of world history (Baptista, Santos, & Pascoa, 2016) and the need for constant innovation, creation of new products, and the continuous improvement of production services make it necessary for organizations to work together via the project management lens.

There are many reasons cited for project management failure and many of them are attributed to one or more breakdowns in the traditional project management systems approach (Aikins, 2012c). But when a project finally meets key stakeholder (user) requirements, many other project shortcomings are overlooked such as cost overruns, late schedules, and scope creep. However, in the author's opinion, transformational eGovernment project management must result in success by ensuring that project management evolves from a system activity approach to a system results approach that must meet key stakeholder expectations. This all starts with identifying an interrelated set of transformational eGovernment project barriers, challenges, and change complexities. This paper is focused on enhancing the project management process in order to upgrade the traditional systems activities approach and support the project results orientation.

Of the reasons cited for project management failure, many of them are attributed to one or more breakdowns in the traditional project management systems approach (Aikins, 2012b). Transformational eGovernment project management must result in success by ensuring that project management evolves from a system activity approach to a system results approach that starts with identifying an interrelated set of transformational eGovernment project barriers, challenges, and change requirements. This research is focused on

informationally enhancing the project management process in order to upgrade the traditional systems activities approach and support the project results orientation.

To address the difficulties currently experienced specifically in eGovernment projects, it can be argued that modern project management growth that began in the 1960s (Kerzner, 2001) now needs to be radically accelerated: become less process bound, more results driven, and change oriented.

NEED FOR IMPROVED eGOVERNMENT PROJECT MANAGEMENT METHODOLOGIES

Transformational eGovernment project management has adopted the broad project management components; such as standardized frameworks, governance, certifications, and qualifications that were developed in the past and in different industries. And academics, researchers, and practitioners continue to compile knowledge, skills, tools, and techniques that form the basis of the science of project management. This science has been encoded into project methodologies or guides that are generally accepted by theoreticians and practitioners and they are heavily promoted by government decision-makers and contracting authorities. But most practitioners, and even some academics, perceive and act as though the prime purpose of project management methodologies is to manage administration and compliance rather than support and guidance (Wells, 2012).

In transformational eGovernment, the project management systems approach is not enough. Instead, in transformational eGovernment, project management must discover the interrelated sets of challenges, barriers, and changing conditions that impede transformational eGovernment project success and respond to and cope with them from a “results achieved” perspective. The project management systems approach must become a basic entry level to the transformational eGovernment project management regime. Now, project

results and change management must become the project drivers that are measured by the effective management of objectives, stakeholders, clients, technical and subject matter experts, resources, and functional support services (Kerzner, 2001). Change control is fundamental to good project management (Whyte, 2016).

Managing investment resourcing for disruptive radical change innovation in the project management domain relies heavily on interactive, dynamic project scheduling and monitoring (Jüngen & Kowalczyk, 1995). So in the radical disruptive change arena project planning, execution, and control is searching for frameworks and methods to develop systems that can enhance, or even replace, the human role in resource allocation; activity scheduling; cost and time trade-offs; and analysis of planned and actual delivery discrepancies.

In addition, if examined within a holistic view, there are a number of additional elements that impact project management and eGovernment. Disruptive radical innovation is often cited as the way forward to viable and competitive business, not-for-profit, and government organizational progress - particularly in technological areas (Lettice & Thomond, 2008). However, disruptive radical innovation is accompanied by a significant up-front resource requirement, and it adds management complexity especially when the disruption includes demand for continuous quality in management, processes, and products which impacts the application and design of effective project management methodologies. Adequate resource investment and effective project management are germane to beneficial disruptive innovation and they are considered critical to overcoming management resistance and opposition.

Thus, disruptive radical innovation requires a project management shift in resource allocation models since prevailing resource allocation models inherently mitigate against innovation in such ways as: inappropriate business standards and prioritization approaches, stifling profit margin targets, short-term focuses on existing customers and current technology, distorted incentive

and compensation schemes, and non-sustaining competitive positions (Lettice & Thomond, 2008).

To address the difficulties currently experienced specifically in eGovernment projects, it can be argued that modern project management growth that began in the 1960s (Kerzner, 2001) now needs to be radically accelerated; become less process bound, more results driven, and change oriented. Transformational eGovernment project management could take on the functions and features of other management professions similar to the example of accounting and finance. By comparison, accounting is subsidiary to finance, and project processes are subsidiary to project results. Results should far outweigh processes.

The current more popular international project management methodologies, PRINCE2 and PMBOK, along with a litany of others, do not meet the needs of eGovernment. eGovernment failure is disappointing and much research has been dedicated to examine why and if project management could be the culprit. Clearly, project management plays a significant role. This paper introduces the proposition that current parlance and management culture accepts that the science of project management is only enacted once a project has been identified.

Inexplicably, common practice does not acknowledge that it is the application of the science of project management to any operational endeavour that creates a specific project. The project is born through the application of the project management principles by bringing rigour, discipline, and specificity to a challenging and complex, though often vague, operational endeavour and objective.

In addition, the science of project management expressed through popular methodologies does not address nor assist with the synergistic compendium of ten barriers and challenges to international eGovernment success recently studied (Furlong, 2011; Furlong, 2012, Furlong, 2015) and outlined in the Appendix. And further, the discipline of change management is not one of the knowledge areas in the generally accepted project management methodologies, even though

change management is evolving from an occupation into a true profession and managing projects is developing into a real profession (Silvius & Batenburg, 2009). Change management impacts project management by incorporating a more effective integration framework that is anchored on:

- Developing strategic and operational programs that are based upon a strong linkage among ideas, concepts, action, and feedback;
- Adapting new business processes, subject matter expertise, and support technology;
- Aligning professional project technical skills with broader based management skills in culture, policies, and human resources;
- Ensuring that project management research is focused on the improvement of evidence-based best practices; and
- Exploiting new tools, techniques, and processes that have been well conceived, researched, and tested (Söderlund & Maylor, 2012).

A characteristic of existential change leadership is that the emphasis must be on the human side of the enterprise: people management over task management, and people orientation over task execution. Business transformation is temporary so existential change leadership, the human side of enterprise, solves project problems related to lack of trust in project start-up, untried project experiences and routines, new project organizational role ambiguity, and un-ingrained project commitment (Tyssen, Wald, & Spieth, 2014).

Lastly, one of the key characteristics of existential change management is that it bridges the gap between the more standard project management linear processes and the more amorphous network processes related to disruptive radical project changes. The network processes are able to cope with the creation of a project through progressive collaboration over time as well as instantaneous disruptive innovation. The existential

change management approach includes a flexible project organization structure that develops the needed project processes, tools, and techniques to overcome the misalignment of mainstream project practices with the needs of disruptive innovation and radical change (Robertson, Galliers, Oswick, & Scarbrough, 2011).

In world-wide project management methodologies and other more parochial ones, methodologies and time-honored project management processes associated with project integration, scope, schedule, cost, quality, resource, communications, risk, and procurement are often used as safeguards that split management and control of the projects and thereby water down accountability for project development and implementation success. In no case is there a domain within the methodologies that directly and specifically provides the transformational eGovernment project manager with the tools to cope with the intrinsic problems that impede transformational eGovernment (Furlong, 2011; Furlong, 2012).

The objective of this paper is to revamp project management methodologies by transforming them from process-bound mechanisms to a problem and results oriented instrument. Akin to the early medical profession's emphasis on procedure (e.g., in order to preclude the adage that the operation was a success but the patient died), the project management profession has too long suffered from the use of project management methodologies that focus on procedures and processes instead of those that focus on results and accountability.

Futuristic transformational eGovernment project management methodologies must contribute to the successful management of transformational eGovernment projects. They must move well beyond the generic body of knowledge that is generally recognized as good practice. Instead, the enhanced methodologies must address the very caveat that existing methodologies hedge against; that is, the embracement of the responsibility for management to obtain successful results for transformational eGovernment projects.

The enhanced methodologies must reach above and beyond the goals to provide professional project management certification; standardization of processes, skills, tools, and techniques; and ethical codes of conduct. They must address a higher objective. The required methodology must enhance the science of transformational eGovernment project management so that the project team can be held accountable for results achieved. The project manager and team are not stewards of the administrative procedures; they are responsible for project success and outcomes.

Various key transformational eGovernment organizations (United Nations, 2010; West, 2007) have completed studies that identify the causes of transformational eGovernment project failure and, just as importantly, they have identified, described, and analyzed the reasons for project development and implementation success. Current project management methodologies unwittingly allow the project manager to escape his accountability by retreating behind the mantra: "we're on budget; we're on time; you changed the requirements." These processes protect him at the expense of project results-oriented success as opposed to process-oriented success.

Current transformational eGovernment project management is built on the broad project management components such as standardized frameworks, governance, certifications, and qualifications; and administrative tools and techniques that were developed for a different time and for different industries, some of which no longer exist or are radically changed like agriculture, auto making and mining coal. These project components are focused on the techniques and science of project methodology as opposed to the successful delivery of project products, services, and results; and on the solving of problems and barriers to project success.

For example, current project methodologies related to project cost management involve estimating, budgeting, and controlling costs. But these processes can lead to the creation, control, and

analysis of an array of data points that can overwhelm and even misinform the project manager, the project team, and key project stakeholders. Other project management domains such as project quality management include process elements such as cost benefit analysis, control charts, bench marking, statistical sampling, and flowcharting.

The point is that ineffective project management is one of most significant reasons for transformational eGovernment failure (Aikens, 2012b; Misuraca, 2009), and that the focus on project methodology instead of project results is the root cause of ineffective project management. Methodology trumps results.

Time-honored project management administrative processes such as scope, schedule, cost, quality, resources, communications, risk, and procurement often tend to split the management and control of the projects and dilutes the accountability for project success.

But enhanced project management redirects the definition of projects (particularly ICT work) towards results. It means existing in a futuristic milieu of complexity and uncertainty wherein it is the application of the science of project management to any endeavor that creates a project, not the project start-up definition. Futuristic projects will be created, they will not be defined; they will be created by evolution, unintended consequences, and responsive iteration that solves problems and produces project results. Futuristic projects will encompass a significant component of existential change leadership to cope with the behavior and mind-set uncertainty that permeates projects in an ever-faster changing environment.

In eGovernment project management enhancements, management of eGovernment projects would focus on project problems rather than methodological processes. The enhanced project management solution would provide the tools, techniques, and mind-sets to account for the impact of the holistic, synergistic challenges and barriers that surround and influence eGovernment projects.

SOLUTIONS AND RECOMMENDATIONS

G

In addition to proposing enhancements to project management that include technological support to address the compendium of ten challenges, along with a deliberate intentional focus on results vs. administration, this paper proposes the inclusion of existential leadership to address the human side of the enterprise. The critical aspects of existential change leadership are the content of change, and the decisions and actions that will produce the change results and outcomes, content, people, and process (Anderson & Ackerson Anderson, 2010). And conscious existential change leadership should be tightly linked to enhanced project management to produce a unified, integrated eGovernment transformation strategy. In transformational eGovernment there are well-known challenges and barriers to enhanced project management. But in existential change leadership there are internal and external drivers and resistants that are more subtle, difficult to perceive or understand, and often insidious.

Enhanced project management should become a pragmatic business discipline that is anchored on a strategic, conceptual, theoretical model and an integrated, tactical, and operational management framework. And conscious existential change leadership deals with planned and revolutionary change which is the antithesis of organizational continuity and stabilization. Its discipline is visioning, behaviour changing, and mind-set altering; all of which is based on a set of theoretical principles and a change leadership road map.

Both disciplines center around portfolio agendas; planning and execution processes; standard structures and practices for designing, implementing and monitoring project and change efforts; project and change centres of excellence; subject matter experts; and strategic management offices. Both are equipped with specific proven theoretical and methodological frameworks to

guide practitioners and stakeholders. And both rely heavily on conscious and enlightened leadership to achieve transformational results.

Integration of conscious existential change leadership and enhanced project management is paramount because each discipline brings a specialized perspective to the common essence and key elements of revolutionary change initiatives and transformational project initiatives; they are the common essentials of content, people, and processes. Collaboration, integration and transparency is another growing ingredient in modern enhanced project management that bridges the gap between academic project management perspectives and in-situ project practice; the bridge (gap) is often framed in terms of strategy and execution (Söderlund & Maylor, 2012).

A recommendation for project management integration centers around strategy and execution which is about developing and realizing ideas, business and technology with respect to both product and process, skill sets and attitudes as regards toolsets and people management, research and practice in rigor and relevance, and exploration and exploitation of new and existing knowledge.

It is recommended that the amalgamation of these two disciplines - along with leadership, innovation, change and integration - will effectively respond to the inexorable acceleration demands of transformational eGovernment. Specifically, by overcoming the constraints and risks that have impeded eGovernment progress and by capitalizing on eGovernment opportunities that have remained so elusive.

Enhanced project management will resolve the compendium of ten operational and technical challenges (Appendix); conscious existential change management will address market, business, organizational and cultural imperatives, and leader and employees' mindsets and behaviour.

Transformational eGovernment project management and change management disciplines should include a tactical and operational framework comprised of a project roadmap (Ackerman Anderson & Anderson, 2010) and a set of targeted

compendium of eGovernment challenges and barriers.

The enhanced project management framework remains prescriptive insofar as it recommends activities that should be performed and results that should be achieved; all with a comprehensive specificity and a high level of in-depth exactitude. The basis for this type of framework is that the objective of enhanced project management is predicated on a fully committed project sponsor; defined and articulated stakeholder requirements; multi-level management support; capacity to effectively plan, execute, and monitor project activities; and fully defined project results and outcomes.

FUTURE RESEARCH DIRECTIONS

The objective thus defined enables the integration of the framework, tools and techniques to create the project initiation, a project charter, and a preliminary project scope, schedule, and cost scenario. In this way the project content is defined, an implementation plan is developed, and project assumptions and constraints are determined.

This research has shown that transformational eGovernment is not living up to promises made; progress is stalling and project failure rates are high.

Further research could investigate:

- Why the project management methodologies, originally designed to address the industrial and manufacturing age, do not adequately respond to the needs of today's eGovernment initiatives;
- Why project management seems keen to adopt new factors to achieve success (such as methodologies, tools, knowledge and skills), but continues to measure or judge project management using tried and failed criteria (Atkinson, 1999);
- How methodologies have to be revamped from an administrative compliance methodology to a results-based accountability methodology;

- How project management could be evolved to a state where it can become a critical force in eGovernment solutions;
- How technology could facilitate radical changes to organizational arrangements, reengineered business processes, and become a more client-focused human resource factor; and finally
- How the use of an informationally enhanced project management methodology, along with existential change leadership, could potentially address some of these issues and highlight the need for technological support within the project management discipline.

Thus, the next step in research is to examine a multi-pronged approach, including enhanced project management and existential change leadership to respond to the challenges and barriers that have long impeded transformational eGovernment progress and the accountability vacuum for the elusive transformational breakthrough results.

CONCLUSION

Transformational eGovernment project management must accelerate a more holistic approach in project management. The systems-based approach incorporated in the generic project management methodologies is not enough for transformational eGovernment. Instead, in transformational eGovernment, project management must become more results focused and less prescription focused. It must discover any interrelated and holistic sets of challenges, barriers, and change issues that impede transformational eGovernment project success; it must respond to and cope with them from a “results achieved” perspective. The project management systems approach must devolve to a basic entry level to the transformational eGovernment project management regime, and project results and outcomes must be the project drivers that are measured by the effective management

of objectives, stakeholders, clients, technical and subject matter experts, resources, and functional support services (Kerzner, 2001).

When an organization introduces a change within a project or initiative, that change needs to be effectively managed on both the technical side and the people side. A technical side focus ensures that the change is developed, designed, and delivered effectively. The discipline of enhanced project management provides the structure, processes, and tools to make this happen. A people side focus ensures that the change is embraced, adopted, and utilized by the employees who have to do their jobs differently as a result of the project. The discipline of existential change leadership would provide the structure, processes, and tools to make this happen.

Each discipline addresses the key result areas in its particular domain. Enhanced project management confronts the generic compendium of ten barriers and challenges to transformational projects along with leadership, innovation, change, and integration. Existential change leadership responds to the leadership principles and common mistakes that are endemic to transformational change initiatives.

Enhanced project management and existential change leadership would both aim to increase the likelihood that projects and initiatives deliver the intended results and outcomes. While each discipline can function independently, the most effective approach is to integrate existential change leadership and enhanced project management to create a unified approach to implementing change on the technical front and people side front. The interfaces between project management and change management overlap, and they are certainly interdependent when it comes to successfully delivering value by supporting the strategic initiatives of the business.

Project management is key when it comes to initiating, planning, executing, and monitoring the project activities and deliverables. They ensure a strong solution design backed up with detailed project plans.

Existential change leadership would prepare the organization for the change impact, manage the transition from how we do things today to how they will be done tomorrow, and put special efforts into reinforcing and anchoring the change into the everyday work and life of the organization.

Critical tools in project management are the project charter, business case, budget estimations, work breakdown structure, resource allocations, scheduling, and tracking. Key tools in existential change management are organizational assessments, stakeholder mapping and interventions, communication and coaching plans, training programs, sponsorship road maps, and reinforcement activities.

With the impact of leadership, change, innovation and integration, what remains is the challenge to build an enhanced project management methodology. Accelerating the design of a fully developed enhanced project management approach for the digital age would harmonize both the project and change management disciplines, put results ahead of administration, promote the benefits of innovative change, and deliver eGovernment and ICT project success.

REFERENCES

- Ackerman Anderson, L., & Anderson, D. (2010). *The Change Leader's Roadmap: How to Navigate your Organization's Transformation*. Wiley Imprint.
- Aikins, S. K. (2012a). *Foreword, Managing E-Government Projects: Concepts, Issues and Practices*. Hershey, PA: IGI Global. doi:10.4018/978-1-4666-0086-7
- Aikins, S. K. (2012b). *Preface, Managing E-Government Projects: Concepts, Issues and Practices*. Hershey, PA: IGI Global. doi:10.4018/978-1-4666-0086-7
- Aikins, S. K. (2012c). *Improving E-Government Project Management: Best Practices and Critical Success Factors, Managing E-Government Projects: Concepts, Issues, and Best Practices*. Hershey, PA: IGI Global. doi:10.4018/978-1-4666-0086-7
- Aikins, S. K. (Ed.). (2012d). *Managing E-Government Projects: Concepts, Issues, and Best Practices*. Hershey, PA: IGI Global. doi:10.4018/978-1-4666-0086-7
- Anderson, D., & Ackerman Anderson, L. (2010). *Beyond Change Management: How to Achieve Breakthrough Results through Conscious Change Leadership*. Wiley Imprint, USA.
- Arif, M. (2008). Customer Orientation in eGovernment Project Management: A Case Study. *The Electronic Journal of E-Government*, 6(1), 1–10.
- Atkinson, R. (1999). Project management: Cost, time and quality, two best guesses and a phenomenon, its time to accept other success criteria. *International Journal of Project Management*, 17(6), 337–342. doi:10.1016/S0263-7863(98)00069-6
- Baptista, A., Santos, F., & Pascoa, J. (2016, March). Project Management Methodologies as Main Tool for Current Challenges in Global Economy Driving Historical Changes. *Journal of Advanced Management Science*, 4(2).
- British Computer Society (BCS). (2004, August 26). *Parliamentary Report on IT Project Waste Management of IT Projects: Making IT Deliver for Department of Work and Pension Customers*. Retrieved from <http://www.bcs.org/content/conWebDoc/1762>
- Eslerod, P., Huemann, M., & Ringhofer, C. (2015, December). 2016, January). Stakeholder Inclusiveness Enriching Project Management with General Stakeholder Theory. *Project Management Journal*, 46(6), 42–53. doi:10.1002/pmj.21546

- Fraser, S. (2006, November). Report of the Auditor General of Canada to the House of Commons. In *Large Information Technology Projects*. Retrieved from http://www.oag-bvg.gc.ca/internet/English/parl_oag_200611_03_e_14971.html
- Furlong, S. (2011). *Transformational eGovernment Success through Enhanced Project Management* (Ph.D. Thesis). Liverpool.
- Furlong, S. (2012). *Project Management: An eGovernment Driver? Handbook of Research on E-Government in Emerging Economies: Adoption, E-Participation, and Legal Frameworks*. IGI Global. Available from <http://www.igi-global.com>
- Furlong, S. (2015). International Challenges to Transformational Government: Enhanced Project Management Identifies Need for Existential Change. *eJournal of eDemocracy and Open Government*, 7(2).
- Heeks, R. (2008). *eGovernment for Development – Success and Failure in eGovernment Projects*. eGovernment for Development Information Exchange, coordinated by the University of Manchester's Institute for Development Policy and Management. Retrieved from <http://www.egov4dev.org/success/sfrates.shtml>
- Irani, Z., Al-Sebie, A., & Elliman, T. (2006). Transaction Stage of e-Government Systems: Identification of its Location & Importance. *Proceedings of the 39th Hawaii International Conference on System Sciences*. doi:10.1109/HICSS.2006.507
- Janowski, T. (2015). Digital government evolution: From transformation to contextualization. *Government Information Quarterly*, 32(3), 221–236. doi:10.1016/j.giq.2015.07.001
- Janowski, T., Estevez, E., & Ojo, A. (2007, April). *A Project Framework for e-Government*. United Nations International Institute for Software Technology, UNU-IIST Report No. 359.
- Jüngen, F. J., & Kowalczyk, W. (1995). An intelligent interactive Project Management Support System. *European Journal of Operational Research*, 84(1), 60–81. doi:10.1016/0377-2217(94)00318-7
- Kamal, M., Weerakkody, V., & Irani, Z. (2011). Analyzing the role of stakeholders in the adoption of technology integration solutions in UK local government: An exploratory study. *Government Information Quarterly*, 28(2), 200–210. doi:10.1016/j.giq.2010.08.003
- Kerzner, H. (2001). *Project Management: A Systems Approach to Planning, Scheduling and Controlling* (7th ed.). John Wiley and Sons Inc.
- Lettice, F., & Thomond, P. (2008). Allocating resources to disruptive innovation projects: Challenging mental models and overcoming management resistance. *International Journal of Technology Management*, 44(1/2), 140. doi:10.1504/IJTM.2008.020702
- Longford, G. (2001). *Rethinking E-Government: Dilemmas of Public Service, Citizenship and Democracy in the Digital Age*. Presentation to the Canadian Political Science Association Annual General Meeting, Quebec, Canada.
- Misuraca, G. (2009, April). e-Government 2015: Exploring m-government scenarios, between ICT-driven experiments and citizen-centric implications. *Technology Analysis and Strategic Management*, 21(3), 407–424. doi:10.1080/09537320902750871
- Nelson, R. R. (2007, June). IT Project Management: Infamous Failures, Classic Markets, and Best Practices. *MIS Quarterly Executive*, 6(2), 67–78.
- Patanakul, P., & Shenhar, A. J. (2012, February). What Project Strategy Really Is: The Fundamental Building Block in Strategic Project Management. *Project Management Journal*, 43(1), 4–20. doi:10.1002/pmj.20282

- PRINCE2 Pocketbook. (2009). *TSO Information and Publishing Solutions*. Norwich, UK: Author.
- Project Management Institute (PMI). Inc. (2008). *A Guide to the Project Management Body of Knowledge* (4th ed.). Newtown Square, PA: Author.
- Project Management Institute (PMI). Inc. (2013). *A Guide to the Project Management Body of Knowledge* (5th ed.). Newtown Square, PA: Author.
- Robertson, M., Galliers, R., Oswick, C., & Scarbrough, H. (2011). Managing Interactive Innovation: From Project Management to Process Mobilization. *Proceedings of the 44th Hawaii International Conference on System Sciences*. doi:10.1109/HICSS.2011.296
- Samset, K., & Volden, G. H. (2016, February). Front-end definition of projects: Ten paradoxes and some reflections regarding project management and project governance. *International Journal of Project Management*, 34(2), 297–313. doi:10.1016/j.ijproman.2015.01.014
- Sarantis, D., Smithson, S., Charalabidis, Y., & Askounis, D. (2010). *A Critical Assessment of Project Management Methods with Respect to Electronic Government Implementation Challenges*. In *System Practice Action Research* (pp. 301–321). Springer.
- Silvius, A. J., & Batenburg, R. (2009, January). Future development of Project Management Competencies. *Proceedings of the 42nd Hawaii International Conference on System Sciences*.
- Söderlund, J., & Maylor, H. (2012). Project management scholarship: Relevance, impact and five integrative challenges for business and management schools. *International Journal of Project Management*, 30(6), 686–696. doi:10.1016/j.ijproman.2012.03.007
- Sommer, A. F., Dukovska-Popovska, I., & Steger-Jensen, K. (2013, October 29). Barriers toward integrated product development – Challenges from a holistic project management perspective. *International Journal of Project Management*.
- Tyssen, A. K., Wald, A., & Spieth, P. (2014). The challenge of transactional and transformational leadership in projects. *International Journal of Project Management*, 32(3), 365–375. doi:10.1016/j.ijproman.2013.05.010
- United Nations. (2010). *E-Government Survey 2010 – Leveraging e-government at a time of financial and economic crisis*. United Nations Department of Economic and Social Affairs, United Nations. Retrieved from http://www2.unpan.org/egovkb/global_reports/10report.htm
- Van Offenbeek, M. A. G., & Vos, J. F. J. (2016, January). An integrative framework for managing project issues across stakeholder groups. *International Journal of Project Management*, 34(1), 44–57. doi:10.1016/j.ijproman.2015.09.006
- Wells, H. (2012). How Effective are Project Management Methodologies? An Explorative Evaluation of Their Benefits in Practice. *Project Management Journal*, 43(6), 43–58. doi:10.1002/pmj.21302
- West, D. (2007, August). *Global E-Government*. Brown University.
- Whyte, J., Stasis, A., & Lindkvist, C. (2016, February). Managing change in the delivery of complex projects: Configuration management, asset information and big data. *International Journal of Project Management*, 34(2), 339–351. doi:10.1016/j.ijproman.2015.02.006
- Zhao, F., Scavarda, A., & Waxin, M. (2012). Key issues and challenges in e-government development: An integrative case study of the number one eCity in the Arab world. *Information Technology & People*, 25(4), 395–422. doi:10.1108/09593841211278794

ADDITIONAL READING

Dawes, S. S. (2009). Governance in the digital age: A research and action framework for an uncertain future. *Government Information Quarterly*, 26(2), 257–264. doi:10.1016/j.giq.2008.12.003

Furlong, S., & Al-Karaghoul, W. (2011, January). *Delivering Professional Projects: The Effectiveness of Project Management in Transformational Initiatives*, Emerald Publishing.

Marchewka, J. (2015). *Information Technology Project Management Fifth Edition: Providing Measurable Organizational Value*, John Wiley & Sons, Inc., USA.

Oxford Institute, eGovernment Unit, European Commission. (2007, December). *Breaking Barriers to eGovernment – Overcoming obstacles to improving European public services*.

Roy, J. (2006). *E-Government in Canada: Transformation for the Digital Age*. University of Ottawa Press.

Schwester, R. (2009). Examining the Barriers to e-Government Adoption, *Electronic Journal of E-Government*, 7(1), 113–122.

Sharif, A., & Irani, Z. (2010). The logistics of information management within an eGovernment context, *Journal of Enterprise Information Management*. Emerald Group Publisher, 23(6), 694–723.

United Nations. (2014). *E-Government Survey 2014 – E-Government for the Future We Want*, United Nations Department of Economic and Social Affairs, United Nations, New York. Retrieved from https://publicadministration.un.org/egovkb/Portals/egovkb/Documents/un/2014-Survey/E-Gov_Complete_Survey-2014.pdf

Weerakkody, V. (2009). *Handbook of Research on ICT Enabled Transformational Government: A Global Perspective*, Information Science Reference. New York: Brunel University. doi:10.4018/978-1-60566-390-6

Weerakkody, V., & Reddick, C. (2012). *Public Sector Transformation through E-Government: Experiences from Europe and North America*, (2013). New York, USA: Routledge. doi:10.4018/978-1-4666-1776-6

KEY TERMS AND DEFINITIONS

Change Management: Refers to any approach to transitioning individuals, teams, and organizations using methods intended to re-direct the use of resources, business process, budget allocations, or other modes of operation that significantly reshape a company or organization.

eGovernment: The utilization of Information Technology (IT), Information and Communication Technologies (ICTs), and other web-based telecommunication technologies to improve and/or enhance the efficiency and effectiveness of service delivery in the public sector.

Innovation: Defined as a new idea, device, or method. Innovation is the application of better solutions that meet new requirements, unarticulated needs, or existing market needs. The term innovation can be defined as something original and more effective and, as a consequence, new, that breaks into the market or society.

Leadership: A research area and a practical skill encompassing the ability of an individual or organization to lead or guide other individuals, teams, or entire organizations.

Project Management: The discipline of initiating, planning, executing, controlling, and closing the work of a team to achieve specific goals and meet specific success criteria. A project is a temporary endeavor designed to produce a unique product, service, or result with a defined beginning and end (usually time-constrained, and often constrained by funding or deliverables) undertaken to meet unique goals and objectives, typically to bring about beneficial change or added value.

Transformational Government: The use of computer-based information and communications technologies (ICT) to change the way governments

work. The term is commonly used to describe a government reform strategy which attempts to radically change the way people understand government, especially those working within government.

Transformational Leadership: A style of leadership where a leader works with subordinates to identify needed change, creating a vision to guide the change through inspiration, and executing the change in tandem with committed members of a group.

APPENDIX

G

The following summarizes the author's compendium of ten international challenges and barriers that prohibit eGovernment transformation, and offers enhancements to the project management process to address these limitations in an eGovernment environment.

1. *Requirement to manage diverse and conflicting stakeholder interests within a governance framework.* Stakeholder interests are usually conflicting because eGovernment applications are usually developed with one or more departments and central agencies. Each of these departments and agencies has a unique legislative mandate, accountability regime, culture, history and background, and more recently security requirements.
2. *Challenge to continuously adapt to and blend technology, people, and processes.* Today's system environment is more organic than it was in the past; previously, system solutions were applied to a corporate service environment. Today's systems are at the core of company performance, not on the periphery. They are significantly affected by evolving priorities and circumstances, and are more integrated with the operational environment including technological developments, the capacity of the resource experts, and constantly changing and evolving business processes.
3. *Outdated business models that reward traditional applications.* Most business models do not recognize that collaborative and unprecedented solutions do not meet the criteria for performance measurement targets, accurate costing and resource utilization, and work plan deliverables whose solutions are not known until they are negotiated and well into the implementation stage. Promises of cost and resource reductions along with improved efficiency and effectiveness gains the funder's attention more than promises of transformation and innovation.
4. *System development models affected by political realities and a new relationship with the private sector.* Most system development models do not recognize the "stop and start" reality of projects affected by political cycles and funding priorities, and the need for system development fragments to be reused instead of continuously "starting over." Though cancelling projects is generally due to changing systems objectives, it is critical to recognize the waste of precious resources and time, and the inability to recover and reuse these efforts. However, public service has been impacted significantly through private sector contracting and outsourcing arrangements. The integration of private and public sector resources is now mandatory.
5. *Lack of access to lessons learned and a body of knowledge for government wide projects.* Project managers are designing and implementing system solutions that are often unprecedented and government wide, and yet they have no facility to access the knowledge nor benefit from the experience gained from other project managers in similar circumstances. The problem is that there is no way to harness previous experience and no demand to conduct and access lessons learned.
6. *Promises of interoperability, integration, and cost and resource savings.* The eGovernment environment is predicated upon a collaborative and partnership-based environment that requires sharing both work and accountability responsibilities, and it is usually argued (and ultimately funded) under a banner of promised cost savings and resource reductions.
7. *Proliferation of information and the challenge to judiciously access and manage information.* The information age exacerbates project management because of the massive and exponentially produced data that must be sorted out to effectively implement system solutions. The interconnectedness of information and system requirements is so overwhelming that projects suffer from the weight of information. Mining through this data to retrieve the relevant information produces a "spin and

churn” that can be non-productive; and this, along with the lack of authoritative control to wind through the layers of information, can derail the project.

8. *Lack of a comprehensive holistic approach to project management as the driving force.* Project management often plays the role of arbitrator, as it is often the agent that brings the disparate parties together to deliver a solution that was not driven by either party. This is usually the case with citizen centric applications as they cross the program interests of each of the contributing organizations. Project management needs to drive the solution to change the business processes of the affected departments and turn the solution into a government wide enterprise.
9. *Limited access to vital subject matter expertise.* Within governments, knowledge is either so vastly spread or not available that it is difficult for the project manager to understand the implications of systems design. The knowledgeable personnel are difficult to locate and approach given hierarchical and organizational limitations, and are frequently reassigned and no longer accessible.
10. *Organizational environment not presupposed to enterprise-wide transformation.* Departments do not necessarily act as units of a government enterprise; they are vertically based with individual objectives and resource reward mechanisms. Accountability of each department is to its Minister and senior officials, and to the government acts for which it was created.

Technology and Terror



Maximiliano Emanuel Korstanje
University of Palermo, Argentina

Geoffrey Skoll
SUNY at Buffalo, USA

INTRODUCTION

Our founding parents envisaged a world where progress, rationality, and technology played a vital role in building a better place to live. They never imagined the effects of 9/11 nor the financial market and stock crisis in 2008. The rise of uncertainty as a main cultural value of contemporary society raised the question of how much technology facilitated the evolution towards a more pacific, fairer, and safer world. We have witnessed how 9/11 strengthened a process of securitization where high technology was used to surveille citizens, accompanied by ethical dilemmas as illustrated by the Edward Snowden case. David Lyon (2003) points to the public spaces of airports, city squares, and restaurants which are monitored by digital cameras and biometric technology. Securitization has reinforced authorities' trust in technology while terrorist attacks continue across the globe. Maximiliano Korstanje (2014) has argued that English-speaking cultures and technology were inevitably entwined. Technology facilitates prediction of events in the future, and thereby encouraged an excess of trust in technological solutions to social conflicts and problems. This is why the concept of risk is an essential element of Anglo culture (Korstanje 2014; 2015): namely its association with capitalism and the desire to predict outcomes to assure profitability. Much of the connection between technology and English society goes back to the creation of modern science by Issac Newton (1643-1727), which coincided with Britain's leadership in capitalist development and global imperialism.

The advent of terrorism as justification of states' social control and growth of repressive apparatuses has led to ethical dilemmas in purported democracies. After Edward Snowden's revelations, citizens understood the limits of democracy as well as the darkest side of state terrorism. The rational nation state, far from enhancing the well-being of its citizens, manipulates the fear instilled by terrorism to increase social control and subordinate private life to the government. This chapter explores ethical issues of electronic surveillance, which is ostensibly applied to thwart terrorism, but which has the effect of undermining democracy in the United States and the developed world. The main thesis is that the apparent conflict of security versus surveillance is what keeps terrorism alive. In the first section, we discuss how the use of electronic cybernetics produces a dissociation between morality and action. This leads to question to what extent digital technology can prevent disasters. In the twentieth and twenty-first centuries, if not before, science has been instrumentalized to protect the interests of the status quo and to try to control the market. Instead of understanding facts as they are or save lives, science became employed and deployed to reducing risks and losses of wealth. The excess of information produced by capitalism obscures reality for decision makers, and contributes to a permanent state of emergency (Mueller and Stewart 2016).

Robert Boguslaw (1965) noted rise of a new utopian class, composed of aficionados of technology and high-tech design. This technophilia fed the maximization of profits over ethics. Today's

DOI: 10.4018/978-1-5225-2255-3.ch316

alienation, he said, derives from the powerlessness to accept probability as a mainstream cultural value for society. He opined that reliance on technology increasingly promotes a moral indifference so long as their behavior coincides with profitability. Decades later, Jean Baudrillard and Paul Virilio made similar arguments.

For Baudrillard, technology plays a crucial role in configuring a hypothetical scenario rooted in the future, but with disciplinary practices. He argued that this state of affairs was accelerated by 9/11, which promoted a parallel reality where pseudo-events prevail. Baudrillard referred to Steven Spielberg's movie, *Minority Report*, as depicting his thesis about the convergence of the future with the present. In the movie certain people are clairvoyant about future crimes, thereby allowing for neutralizing criminals before they can commit offenses. The movie raises the question of how a crime can be punished before its commission. Baudrillard points to the War on Terror and US counter-terrorism strategies as using similar methods, but which in reality legitimate US geopolitical control of the World. As Baudrillard construes current policies, fear paves the way toward increasing paranoia which is a product of the multiplication of information and the hegemony of object-sign (Baudrillard, 2006).

In Baudrillard's view, the 9/11 attacks represented the clash of triumphant globalization at war with itself and unfolded a fourth world war: the first put an end to European Supremacy and to the era of colonialism; the second put an end to Nazism; and the third to Communism. Each one brought us progressively closer to the single world order of today, which is now nearing its end, everywhere opposed, everywhere grappling with hostile forces. This is a war of fractal complexity, waged worldwide against rebellious singularities that, in the manner of antibodies, mount a resistance in every cell. (Kellner, 2005: 3)

For his part, Paul Virilio warned about the dictatorship of virtuality as a new mechanism of

generating commitment and exclusion. He agrees that by means of the imposition of stereotypes mass media moulds citizenship. Public opinion is shaped according to economic interests of informational chains controlled by owners of media. They are able to link disconnected events in any part of the world, which are then disseminated to global audiences in minutes. One result is the diminution and devalorizing of human face to face communication and interaction. Consequently, the distinction between what is or is not real is decoupled from being there and social agreement, a point noted by many in the field of social psychology. Sharing similar norms, people construct symbolic archetypes which together make up a social imaginary. In contrast, Mass-mediated interaction short circuits this social process, creating a dictatorship of oculo-centrism: a signification of the image wherein the interpretations of images are standardized according to the dictates of the media. This mediated tele-interaction reduces the independence of citizens for comprehending their environment (Virilio, 1996). As Nicholas Carr (2011) says, the possibility digital technology is making us more stupid than other generations. The excess of stimulation and the velocity digital narratives to gather and read information undermines the possibility of understanding what is being processed. This means that we handle more information but with lessened capacity to understand events. As we will discuss in next section, technology seems not to be part of solution, but the problem.

BACKGROUND

From Disasters to Terrorism

Ulrich Beck's (2006) popularization of risk analysis put forth the argument that accidents under some conditions were the result of an inadequate manipulation of technology. Using the Chernobyl nuclear meltdown as exemplar, Beck argued that humanity had entered a new era in which tech-

nological advance surpassed threats from nature. According to Beck it is the shift from natural to technological threats that marks post-modernity. Despite, or maybe because of, Beck's a-historicism and anti-Marxian sociology, the popularity of his work reflects a recent recognition of threats that arise from new technologies developed in the latter half of the twentieth century. Although, technological advances are aimed at making this world a safer place by mitigating and controlling risk, the fact is that they contribute to creating new risks. Of course this is not new. For example, at the turn of the twentieth century growing horse manure in urban areas presented a major public health hazard, until internal combustion engines replaced horsepower, which then led to air pollution from burning fossil fuel. Ecological concerns are perhaps a point where the paradox of technology is most vividly appreciated.

The goal of this chapter aims at exploring the connection between technology and risk. In doing so, the discussion between Cass Sunstein and Anthony Giddens illustrates what might be called the establishment framing of the argument. Sunstein has been part of US President Obama's cabinet, and Giddens helped construct the third way program of Tony Blair and New Labour in Britain. While Sunstein argues that fears are determined by cognitive shortcuts, Giddens says that risk seems to be a result of technology. Giddens questions confidence in technology, although he praises its benefits. Sunstein is convinced that risk is a product of human ignorance and inaccuracy in the decision making. Two views, two alternatives are juxtaposed in an ongoing debate.

The Sunstein-Giddens debate draws the boundaries of policy analysis within a world capitalist framework. That is, their debate stays safely within the ethos and assumptions of the currently prevailing political economic system that dominates the globe. Sunstein is a legal scholar and professor in the Law School the University of Chicago. He is part of the clique there that adheres to the so-called law and economics framework derived from the Chicago School of Economics. This Chicago

School has been led by such world luminaries as Gary Becker, Milton Friedman, and others whose ideas owe much to the Austrian school of Frederick Hayek. They are anti-Marxist, anti-Keynesian, and avowedly neoclassical. In practice their ideas formed the basis for the neoliberalism of Western hegemony in the late twentieth century. A hallmark of their thought and its legal and public policy applications is the central figure of the rational actor. The rational actor is a heuristic to allow various econometric formulae to have some reference to the real world. The rational actor is the *homo economicus* who always acts to ensure the greatest economic advantage to him or herself as an individual. All theories flow from this assumption which is markedly individualistic and assumes a kind of cognitive functioning rarely if ever found among real people.

Anthony Giddens takes a social analytic approach that is neither individualistic nor based on the assumption of blind social forces and structures that operate without human agency. Giddens' approach combines individual agency with social structure. He does not assume a rational actor, but sees a dialectic between the effects of socialization (social structures, institutions, and the like) and the ways people act. Most relevant to the Sunstein-Giddens debate and the present essay is that Giddens says that technology is both a consequence and cause of human behaviour as they shape each other. Their debate comes down to one between the neoliberals and the Keynesian. It ignores the far more far reaching and radical critiques offered by a number of authors reviewed in this article.

In what follows a number of analysts are reviewed. None is directly involved in the Sunstein-Giddens debate. Nonetheless, each analyst presents a different aspect of technological evolution, its consequences, and its relationship to human behaviour and decision making. Technology plays a pivotal role in organizing not only behaviour but also the society itself. Undoubtedly, the technical advances blurred the connection between time and space, facilitating many things for people. Among

the benefits of technology applied to health for example, we have:

- Lights and electricity created a real revolution in temporal-spatial displacements.
- Life expectancy has been extended.
- The techniques of education have been radically altered providing new resources.
- Risk, disasters, and other dangers may be mitigated by means of technology.

G. Amar (2011) argues that the evolution of technology has made life safer in many senses. The current meaning of mobility seems to be something else than a technique. This exhibits a spirit and kind of social bond that connects self with territory. Technology may be not only positive, but allows re-discovering the principle of “religance”. This neologism refers to the anthropological sense of place. The principle of religance that circumscribes the subject to the community may create new technologies, more sustainable for ecology that improves our quality of life. From this perspective, Amar argues that innovation would play a pivotal role in the industry of mobility worldwide. In contrast to the prevailing French literature, Amar is convinced that there are two ways of moving. If we evaluate the problem of mobility in terms of space-time criterion, we need to conclude that technology has made life faster, but not safer. Rather, Amar adds, there is surfacing a new manner of transport, where people are experiencing the “time-substance” to fabricate sentiment about visited spaces. This new type of mobility follows recreational goals determining long-standing and satisfactory experiences (Amar, 2011).

Technology can serve as a mechanism of mitigating, forecasting, and preventing disasters. In opposition to this, Paul Virilio (1996) says that technology acquires a negative tendency because it expands not only the process of alienation but blurs the boundaries of heritage and nationhood. He notes that mass media is framing and controlling the sense of reality. Today it is in vain to question the veracity of news, what is important

for an audience is hyper-reality. Human perception now can see only events that never have happened; they are rooted in the future. Showing a natural tendency to communicate with others, human beings adapt their behaviour to specific environs. Events that are geographically dispersed are broadcast on the same screen synchronized in seconds. The acceleration of mobility triggered an inevitable confusion between present and future. As a result of this, technology leads to a decline of trust and social bonds. Unlike Amar, Virilio thinks the technology eliminates the natural barriers that prevent risks. Before the twentieth century, cities were built as refuges that marked the ends and beginnings of civilization. Whatever else its nature, the city was bounded by the wall, but in the digital world, where all cities are cloned and globalized, any event may trigger a real disaster (Virilio, 2007; 2010).

In his study of the new social self, *One Dimensional Man*, first published 1964, Herbert Marcuse argued that technological societies are irretrievably flawed. At the same time that technology advances, liberty is being sacrificed. The dependency of human beings on the newest technique not only paves the way for the advent of a new ideology, but also depersonalizes the workers in favour of capital (Marcuse, 1991). Marcuse’s criticism echoed that of Aldous Huxley, whose 1931 *Brave New World* depicted a fictional dystopia where people are controlled by various technologies, chiefly genetic manipulation. In this valuable novel, Huxley introduces readers to a discussion where society is defined as a set of embodiments based on abstract ideas. This novel represents an acid criticism of a technologically determined world and overpopulation. Given the demographic decontrol in urban areas, Huxley envisaged that technology would be efficient in controlling human beings as machines. As a result of this, democracies would be converted to governments bound to totalitarianism (Huxley, 2006).

In this respect, Darin Barney presents a model to understand the role of technology and its effects on democracy and political life. He says that

although technology leads humankind beyond ethical questions, it is important not to lose sight of the fact that it should be defined as a political construction, whose ends are based on imposing a specific discussion (Barney, 2007). The problem of technology is that any question is answered before being formulated. Proponents of technology have criticized this view by saying it transcends the boundaries of culture and ethnocentrism. Technology in a globalized world permits changes, and different postures directly to a species of trans-humanism. This would entail a more democratic and fairer society (Hughes, 2004). In this vein, the British journalist Guy Sorman admits that risk detractors are more interested in preserving the status quo than in exploring the benefits of development. Technology for Sorman is not only positive, but also emancipatory from many points of view. Enemies of technology are reluctant to progress. In lieu of accepting this, they prefer to confuse public opinion by inventing risks that do not exist (Sorman, 2002). Generally, academics have agreed that given some circumstances, technology may engender some risks, which if not duly evaluated, lead to future states of emergency. Of this point, likely, Chernobyl is one of the most vivid examples. Next, we will consider to what extent reflection and knowledge generate panic in public opinion. The sentiment of panic paralyzes the natural barriers to strategic risk management. Unless otherwise resolved, minor risk creates serious disasters by its cascade effects.

Technology in Knowledge-Oriented Societies

A historiography of evolution of science merits deepening the question of technology. Although many historical waves of critics have discussed the issue, A. Cuevas distinguishes three important schools: a) the hierarchal model proposed by econometrics, b) non-hierarchal models proposed by sociology, and c) a mixture of agent and system recently developed by the theorists of complexity. While each stage is characterized by focusing on

the diverse forces technology exerts on human behaviour, the fact is that the connection between technology and science depends on its application. Technology is based on two significant elements: artefacts and techniques (Cuevas, 2005). To expand the current understanding of the issue it is important to conduct interdisciplinary research. Every discipline develops a particular definition of technology. It is important to remember that those groups whose access to the monopoly of technique have a substantial advantage that allows them to dominate others who are relegated to the currently prevailing technology. The problem is that technologies may be harmful whenever they do not support common well-being. More often than not, imposition of new technologies risks a Huxleyan style dystopia. Steven Benko admits that technology is being accepted as a way of rationalizing the altering of conditions of existence for humankind. He argues that it may be used in gaining more knowledge but it should be accompanied by ethics to achieve some commonly shared goals. Therefore, social science should focus on the evolution of the ethic of technology in lieu of the technology of science (Benko, 2005). The understanding of the possibilities or hazards of technological revolutions depends on the ethics of technologies' application. In this vein, M. Scott Ruse replies that it is one thing to use technology in the accompaniment of human evolution, and quite another to confront what he calls hegemony of techno-humankind that has accelerated after the advent of post-modernity. The misuses of technology are determined by political goals, or social pathologies. What is important to remember here is the connection between technology, risk, and mobility. Ruse argues that technological revolution needs philosophical scrutiny because the nature of man is at stake. The state, the economy, and the ontological perception of the world penetrate by means of technology. However, in recent centuries the global trends show a hegemony that transformed technology into a totalitarian force. The resulting alienation is counter-productive for societies as it virtualizes space to become the

primary human habitat in a great air-conditioned system. James Farris notes that regardless of the technology, people ultimately have to recognize their groundedness in the everyday and their connection with the material realities of earth.

The extension of technology into the infinite reaches of cyberspace is both exhilarating and wondrously productive in advancing the reach of intellectual and technical power. The accompanying dilemmas and ethical questions are bound to relate back to practical terms of our existence on planet Earth. Our interventions in the sphere of ecology are a critical test of our integrity. The simple action of sorting out the household garbage can remind us of the chain of being extending along our technological story. (Farris 2005: 7)

Virilio anticipated the problem when he cautioned in his *University of Disaster*, that advances in technologies of communication and transportation have created new forms of displacement in time and space. Today people have access to any geographical point of this planet in hours. The time of waiting or travelling has changed forever. Travellers can now move with little genuine contact in the visited lands. Global transportation and communication technologies have made a new kind of real time in which synchronization of watches is neither necessary nor relevant. In addition, citizens have been transformed into consumers. History has been emptied into a fragmentation of events, dispersed globally, and broadcast repeatedly. The attack on New York's World Trade towers in 2011 remains the iconic case in point (Virilio, 2010).

Turning to the generation of knowledge, Virilio says that what had characterized the labour of the university has been dispersed to delocalized territories. Based on an ongoing future that never makes room for the presentiment of disaster, knowledge announces the eschatology of neurosis. In plain words, Virilio argues that everything happens at the same time in hyper-reality without a logical sequence. The world stage is represented

outside the planet, in an exo-earth. The notion of science, as an all encompassing instrument based on rational understanding, has changed. It has turned into an exo-science that promotes the simultaneous globalization of fear, whilst biology and astronomy are eclipsed by the eternal present. Virilio emphasizes the mea culpa of science for its failure to create an ethic of life. Based on the belief that global warming is not reversible in the short run, science explores issues from the perspective of homeland security understood as capital maximization. Big corporations, banks, and the capitalist elite call on climatologists and geographers to design catastrophe simulation software that provides some information about where the next disaster will hit. As Naomi Klein (2007) points out, much money can be made by predicting the next disaster. A new profession is rising: the "economic-disaster-modelling-geek". This expert seems to be more interested in finding and eliminating the risks to businesses, or finding ways to profit from such risks, than in protecting the environment. The philosophy of the science is today determined by the logic of digital screens. The simulation of future that characterizes the digital world has replaced the daily life (Virilio, 2010).

The Excesses of Technology

Zygmunt Bauman has emphasized the hedonist consumption that leads modernity. With a different approach, Anthony Giddens examines the experts who frame and deal with risks. Unlike other times, today's consumers are familiar with the product they buy but this point is secondary as that knowledge is suffering from a process of reflexivity. What characterizes social life in late modernity seems to be the complexity of capitalism that re-structures the connection of institutions along with their individual actors. Where pre-modern societies saw in witchcraft a valid instrument to predict the future, the science of actuary or risk statistics not only validates the policies of the state,

but also represents reality for civil society. “[M]odernity’s reflexivity refers to the susceptibility of most aspects of social activity, and material relation with nature, to chronic revision in the light of a new information or knowledge. Such information or knowledge is not incidental to modern institutions, but constitutive of them ... because many possibilities of reflection about reflexivity exist in modern social conditions” (Giddens, 1991: 20). Modern reflexivity has the capacity to create many realities depending on the purchasing power of consumers.

In a recent book dedicated to ecology, Giddens acknowledges that technology could make a better world, but this does not happen. Instead, technology creates threats to public health. While some risks are monitored and controlled, other, more globalized ones, work as a runaway train (Giddens, 2011). It is important to mention that accidents, like runaway trains, are based on randomness. Accidents could have been other things. Although capitalism has constructed a rational basis of control over almost all human interactions, the fact is that its immanent contradictions lead to the collapse of the whole system. Contingency, uncertainty, and randomness permeate the world system of capitalism. Charles Perrow (1999) uses a risk analysis approach used by insurance companies. He says that the timing of actual accidents is based on the sad reality that a similar event could be repeated at least six times. We are daily facing serious, imperceptible risks that randomly do not materialize in an accident. The circles of control tend to be petrified so as not to monitor these minor risks, which sooner or later cause the disaster (Perrow, 1999). The technical perspective not only ignores this reality but also thinks, erroneously, that risk may be controlled by technological machinery. Risk analysis is central not just to the insurance industry. It lies at the very heart of capitalism. It is the foundation of cost-benefit analysis which is merely a way to calculate whether an endeavour is worthy of

investment, whether expected returns are worth the amount of capital risked.

Robert Castel (2006) examines how individual liberties were tied to private property. Based on the hegemony of capital, modern societies regarded safety as a mechanism to reduce the impacts of illness, poverty, disasters, and aging. Modernity brought a degree of social well-being. At the same time, security gained through technological advance in industrialized societies corresponds with an inflation of perceived risk. Paradoxically, far from feeling more secure, people are frightened by the new, such as scientific reports that discover new causes of cancer. Castel identifies the feeling of insecurity as an obsession with protection. This collective risk aversion corresponds with individualism, atomization, and decreasing dependence on traditional social bonds. Towns and cities were protected in medieval times by the symbolic hegemony of the Catholic Church, but with the growth of market-based relations, secularization replaced religious subordination bringing with it more autonomy. In effect, the transition to modernity resulted in more autonomy but less security, something belaboured by the social theorists of the late nineteenth and early twentieth century—from Emile Durkheim, to Georg Simmel, Ferdinand Tönnies, and Max Weber. The vulnerability of citizenry, as it is conceived today, comes with the decline of feudalism and its replacement by industrial capitalism. In preindustrial times, the degree of violence and hazards of disease in daily life led to low life expectancy. With modernity, people experienced notable increased longevity but concomitant with weaker social bonds. Kai Erikson (1994) reminds us that risks not only derive from the decline of trust, but by the excess of technology whenever protocols of operation are not regulated. The theory of complexity raises the levels of uncertainty in systems and thereby exponentially increases the need for cybernetic controls. The more complex the system, the more regulation is needed to maintain the system, and therefore the more need for cybernetics.

Terrorism and the Problem of Surveillance

The attacks of 9/11 shocked the world. The use of civilian aircraft as weapons combined with the targets—Wall Street, the Pentagon, and a third target that might have been either the White House or the Capitol in Washington DC seemed unimaginable. The Tourism industry, which had been a major symbolic expression of Western supremacy, plummeted. The denominated perpetrators, presumably nineteen jihadists, seemingly planned the attacks according to the rational steps of a management guidebook, or so the story cooked up by US authorities portrayed the incident. From that moment on, Western governments construed terrorism as a global threat which emanated from the Muslim world, especially that part of it with vast oil deposits (Sandlers & Enders, 2004; Steiner & Al-Hamarneh; Korstanje & Clayton 2012; Korstanje, Tzanelli & Clayton, 2014; Tarlow 2014). The effects of terrorism on tourism industry went from the suspension of many flights and the rise of unemployment to deployment of technological boondoggles like full body scanners at US airports. Henceforth, nobody can feel safe anywhere (Korstanje 2015). The so-called scourge of terrorism brought forth a phony debate that counter posed individual freedoms and public security. Individual rights were legally abrogated amidst a massive deployment of communication technology to spy on Americans and other citizens in the Western World. WikiLeaks and Snowden's revelations showed citizens that their governments used terrorism and security issues to curtail any semblance of privacy and democracy (Altheide 2014; Skoll 2014).

David Lyon and Zygmunt Bauman (2013) note that 9/11 did not create the logic of liquid surveillance; it accelerated the conditions of its reproduction. Employing the term *adiahorization*, as dissociation between action and ethic fields, Bauman adds that the introduction of technology originally was aimed at mitigating some major risks. However, it has paved the way for ethically

devoid policies and practices. Counter-terror violence increasingly became controlled by machines, digital instruments manipulated by automatons. Any error, any mistake at time of calculating an attack, was labelled as collateral damages. In this vein, Bauman and Lyon cite what Hannah Arendt wrote in her report on the trial of Adolph Eichmann (1963) that she called the banality of evil, which is the bureaucratization of instrumental reason. The main thesis of Lyon and Bauman's *Liquid Surveillance: A Conversation* (2013) is that modern polities have adopted surveillance technologies for two reasons. First, they are used to control undesired guests, a sort of extended passport control system. Secondly, they ostensibly protect the citizens of the polity. Undesired guests, those who cannot pay for the amenities of the polity are marked and pushed to the peripheries of the city. Think here of the banlieues of Paris or the favelas of Rio de Janeiro. Lyon and Bauman imply a weakening of the state as its primary role has become facilitation of global commerce, or globalization. By introducing surveillance technology state control apparatuses try to mitigate citizens' exposure to threats made possible by globalized commerce, communication, and travel, whether from infectious diseases, globe-trotting terrorists, or transnational organized crime. The quest for order that characterizes the human existence is mitigated by the need for change.

In this respect, Rebecca Fiske (2016) uses the notion of the state of exception as expostulated by Carl Schmitt (1985) and critiqued by Giorgio Agamben (2005). Fiske starts with the premise that law originated from the authority of divinity, which intervenes in human affairs at the moment of moral emergencies. Schmitt's concept of sovereignty is that of a gatekeeper. The state of exception opens the doors for paradoxical situations. Fiske notes that after 9/11 the United States violated the human rights of its citizen in the name of security and well-being—to say nothing of its war crimes and crimes against humanity in foreign land. The “global terrorist became an internal and external threat, and the war on terror became so complex

and pervasive that the definition of an immediate threat originating outside the border changed considerably” (Fiske 2016: 11). The challenge to what was left of democratic institutions in the United States rationalized by 9/11 was excused by positing the false dichotomy between liberty and security as addressed by Hartzel and Gerde (2016). The proposition that security demands obedience leads to conditions under which information presumably garnered to protect people is used to control and oppress them. The Nazi Gestapo operated on the same principle.

The Excess of Information

The intersection of terrorism and the digital media was assessed in a book edited by Mahmoud Eid (2014). Introducing the term, ‘terroredia’, Eid expresses the connection between terrorism and media coverage. Many terrorists are fluent in English, the global hegemonic language, and can disseminate their messages to world-wide audiences, and they are able to use digital technologies such as websites, Facebook and other social networks. Eid’s edited volume explores the mutual dependency of terrorism and media. In one of the chapters Samuel Winch comments on the use of media by the ruling class for domestic control and international influence.

In terms of media coverage of protest movements, we could expect ruling class interests to include maintaining the status quo, and therefore, efforts to marginalize dissent and dissenters, attempts to make them seem deviant and strange. Likewise, American Middle East foreign policy has long been criticized for the tendency to support corrupt autocratic dictators friendly to elite capitalists (particularly oil companies). (222)

The war on terror, post 9/11, attempted to defeat an invisible enemy, and consequently terrorism engulfed postmodern politics. The US obsession with terrorism became a virtual *raison d’état*, as its

tactics in counter-terrorism create more terrorism to feed back as an unending atmosphere of fear.

As Mahmoud Eid puts it, we are educated to imagine terrorism is a criminal act while media are a positive phenomenon. But both sides help each other, simply because the media facilitate the terrorist goals, terror attacks provide substantial content and topics debate for journalism. Understanding terrorism as a way of communicating a violent message, Eid’s volume contains considerable criticism of an ostensibly free-value media. The mediatization of terrorism corresponds with a tactic beneficial for both terrorists and governments. Terrorism and media cohabit a swamp in which each produces the oxygen for the other to breathe.

Meanwhile, in order for both to survive, terrorists seek to garner public attention and the media seek to find stories to sell. In a sense, both parties target wide-ranging audiences (although for different purposes); hence, they interact in a highly toxic relationship that involves a process of exchange necessary for their survival. Acts of terrorism provide media stories that result in more broadcasts, press texts, and digital data bytes, while the media coverage brings public attention to terrorists—the oxygen necessary for their existent. (Eid 2014: 24)

Vincent Casaregola (2016) addresses the role of surveillance. He argues that more than a passive instrument, surveillance awakens our deeper fears, producing a psychological climate of self-limitation. A genre of films uses surveillance as the main theme. Based on an exploration of how these archetypes were formed, Casaregola writes that ordinary citizens were gradually pressed to accept control and strict surveillance as a normal rule simply because it enhances security. Instead of seeing government invasions of privacy as real violations, cinema has indoctrinated audiences to ignore the limitations imposed by the security state. Since the beginning of the twentieth century,

the factory system, among other things, employed surveillance as part of a toolkit to exert control over workers and exploit them. Later, ruling elites introduced the same to government apparatuses. Various state bureaucracies manipulated surveillance to gain or keep dominance over the masses. In this vein, recent movies portrayed a futurist realm where the decline of liberty and autonomy leads to the rise of dictatorial governments. In movies such as *Minority Report* (2002), or *Gattaca* (1997), the plot present humans not as weary workers, but as passive consumers, who voluntarily sacrifice their liberties for entertainment. Movies have often served as propaganda vehicles in which enemies are demonized like the Nazis during the Second World War and Communists in the Cold War. Now terrorists and organizations like al Qaeda and ISIS serve as the antagonists who reveal the vulnerability of the United States that reinforces the culture of surveillance.

It is clear that our concerns about observing and being observed continue to proper our interest in political thrillers in which surveillance, lawful or not, plays a major role. We identify with the spies and law enforcement officials whose assiduous but authorized surveillance preserves our own safety, yet we suspect that not all such individuals or organizations are careful or caring enough not to make serious errors. We also doubt that they would have the moral courage to admit such errors subsequently, allowing the innocent to suffer rather than leaving themselves open to criticism or even prosecution. Additionally, we fear that government may be neither benevolent nor benign in tis surveillance but corrupt and Machiavellian, maintaining control over us rather than keeping us secure. (Casaregola, 2016: 72)

From Counterfeit Politics to the Needs for Plots

Politics today consists in the need to introduce secret plots as a valid source of legitimacy to impose policies that otherwise would never be accepted

by voters. Such so-called counterfeit politics or conspiracy theories are typically interpreted as a pathological form of politics, which is based on paranoia and can lead to chaos. In contrast David Kelman has taken an opposite tack in his *Counterfeit Politics, Secret Plots and Conspiracy Narratives in the Americas* (2012). Kelman is basically treating the problem of ideology as a key factor of secrecy. The idea of secrecy rests on the legitimacy of silence which produces two contrasting circuits, official and unofficial. It presents a disestablishing event that intersects both stories. The credibility of one story is linked to the secret that allows the discovering of the other. In these terms, the theory of conspiracy, Kelman argues, does not threaten legitimate politics as is commonly thought. Crossing the cultures of Latin and Anglo-America, Kelman's says that conspiracy reveals a problem, its secret is necessary to articulate specific policies that otherwise would be rejected. He defines conspiracy as a narrative with political effects. Counter intuitively, Kelman maintains that conspiracy does not flourish in dictatorships, but in the core of liberal democracies. Far from being a social pathology, conspiracy is a key factor of politics that represents a valid effort to unravel an obscured story.

Across all modern democracies, politics needs to include a conspirational fable to legitimate an official one-sided and authoritative discourse. Kelman provides an all-encompassing model to understand the coming and goings of politics in the Americas: "politics is not based on an ideology decided in advance, but it is rather constituted through a specific type of narrative that is often called conspiracy theory. This type of theory is always a machination, that is, a narrative mechanism that secretes, as it were, ideological labels such as the right or the left" (p. 8). The success of this process depends on its ability to pose two contrasting groups, we and they, to impose an "antagonist relation" of competence. Therefore, every conspiracy narrative alludes to a double structure, where the visible story is continuously eroded by a secret one.

In explaining the political struggle as an unlimited game, rather than a hierarchical line of power, Kelman suggests that conspiracy seems not to be a symptom of decline or crisis, but the necessary condition for one discourse to gain hegemony. Politics, in his terms, can be defined as an illusory state of emergency where the sense of community (we) is opposed to others, who are the enemies (they). Conspiracy narratives, Kelman adds, are always placed in a near future, never in present reality. "Politics occurs when one discourse is being undermined and ultimately replaced by other contrasting voice".

Kelman provides a fresh alternative to reading modern politics, but his argument is based on ideology instead of politics. Rather, our argument rests on the idea that conspiracy closes the doors for interrogating ideology. Starting from the premise that political narratives, whatever their nature may be, explain an undesired situation to keep the status quo, it is almost impossible for conspiracy to lead to a radical change.

S. Murillo (2008) has explained that ideology is not an illusory story told to gain legitimacy, but the human fantasy to control death. So, we are ideological animals in our vulnerability. Alvin Gouldner (1976) offers a more political viewpoint without existentialist baggage.

Ideologies, then, are belief systems distinguished by the centrality of their concern for What Is and by their world-referencing "reports." Ideologies are essentially public doctrines offering publicly scrutable evidence and reasoning on their behalf; they are never offered as secret doctrines... Ideologies are intended to be believed in by those affirming them publicly and by all men, because they are "true," and they thus have a universal character... With the waning of traditionalism, there is now an increasing struggle over "ideas." This means a greater struggle over which definitions of social reality (or reports) and which moral rules (or commands) are dominant. Social struggle in part takes the form of contention about What Is and what should be done about it. (33-34)

The ideology fabricated and transmitted to the world says that United States is a mega-political power, whose hegemony is global. This power, from its inception, attempted to expand egalitarian values (democracy for example) to all countries to strengthen their collective well beings. But if the United States is this benevolent and powerful all seeing eye, why was 9/11 not prevented? A tentative answer would be that US officials were familiar with the possibilities of an attack. The theory of conspiracy exhibits the other pole of ideology. Any conspiracy poses valid efforts to make controllable what in nature is uncontrollable. Whenever the reality principle overrides the fiction, the theory of conspiracy responds with the questions that ideology keeps open. The hermeneutic circle of conspiracy works as a mechanism to validate the official discourse, silencing those voices which might bring a radical shift for the system. It takes part as a significant element of ideology, not politics. Last but not least, in psychology, conspiracy and paranoia are adaptive mechanisms of a frustrated ego. Paranoia surfaces to explain the failures of ego to reach goals in the life, which leads to the rejection of the reality principle. The meaning of safety, where ideology works, opens an ongoing quest which never ends. Inasmuch we look for safety, we feel less safe (Skoll & Korstanje, 2012).

DISCUSSION

J. Robert Oppenheimer was the scientific director of the Manhattan Project, the US program to develop the first nuclear weapons. The success of the program resulted in US President Truman authorizing the atomic bombings of the cities Hiroshima and Nagasaki in Japan, eventually killing hundreds of thousands. Oppenheimer speaking in a 1965 television broadcast about the moments following the first successful test of an atomic bomb, the Trinity test: "We knew the world would not be the same. A few people laughed, a few people cried. Most people were silent. I remembered the

line from the Hindu scripture, the *Bhagavad-Gita*. Vishnu is trying to persuade the Prince that he should do his duty, and, to impress him, takes on his multi-armed form and says, 'Now I am become Death, the destroyer of worlds.' I suppose we all thought that, one way or another."

Technology opens ethical dilemmas by making possible what had not hitherto been possible. What is possible is time bound and culture bound, and therefore dependent on the prevailing patterns of social relations, or social structure if you will. And here is where ideology plays a singular role. To paraphrase Karl Marx (1932) from *The German Ideology* which he wrote in 1845-6, the hegemonic ideology of a given time and place is the ideology, the belief system in Gouldner's terms, of the ruling class. The ruling class's view of the world is that which justifies their position and power. Their position and power derive from the sum total of all material social relations, which themselves are based on how the available forces of production are used to reproduce social relations. Technology is another word for the forces of production. Inevitably the development, deployment, and use of the forces of production will fit with the prevailing ideology.

What Oppenheimer recognized was his social role, the part he played, in developing a new force of production, or in its immediate use, force of destruction. Oppenheimer took responsibility for what he did. In contrast, Hannah Arendt in her conception of the banality of evil pointed to instances where people did not take responsibility, and relied on the dual mechanisms of bureaucratic procedures and technology to deny their participation. Those mechanisms of denial and the prevailing ideology—that of Nazism in Adolph Eichmann's case—decouple individual subjects from their actions. Just as Eichmann could say, in effect, that he was just doing his job by employing railroad timetables to move people around greater Germany. The timetables and rail movements were his immediate technology. They served to separate him from what he did so as to preserve his self conception as a moral

man. Technologies in association with prevailing ideologies operate similarly today. Consider for instance US personnel who operate military drones that murder people half a world from where they sit at their controls. Eichmann could think that what he did was justified because the people he arranged to transport people, mostly Jews from Nazi-conquered Europe, presented a threat to the German people. Drone operators can think they are protecting the American people from terrorists. It comes down to the same thing.

The crucial ethical question is about responsibility. Oppenheimer took it; Eichmann did not. The difference, pace Bauman, does not lie in a new kind of liquidity. The twenty-first century is no more liquid than those that preceded it, at least not where ethics matter. Today's drone operators can rationalize what they do as fighting terrorism. They work in the main state apparatus of repression and violence in the world: the US military-intelligence complex. Among its personnel, and among the US population at large, the ideology of a war on terror is part of the product of the umbrella state ideological apparatus most of which relies for its dissemination on the mass media. Today's mass media are faster than those in the past, and they are accessible to a greater portion of humanity. For example, a copy of *The Times* of London in the nineteenth century could take weeks to reach the functionaries in Britain's far lung empire. Today's drone operators can get their ideological messages from the *New York Times* almost instantaneously even while they are murdering people in Afghanistan. British imperial ideology was no less effective; however, than US ideology is today.

If there is a difference introduced by global electronic communication, it is that today more people interact at a distance. In 1876 in London a far greater proportion of social interactions were face to face than today where people use Facebook, Twitter, and various social media to carry on their social lives. To converse with a Facebook friend, one need only consult one's Smartphone, whereas in 1876 London, people had to go out into the

street, to coffee houses or pubs. This does produce a difference, but it is not one of liquidity. It is that today proportionately more social interaction occurs through a single channel of communication that imposes strictures on human perception. With face to face interaction in contrast, many channels operate simultaneously—including so-called body language, use of interpersonal space, odours, dress, demeanour, and so on. Today's social interactions lack such fullness, and therefore social relations are impoverished by reliance on contemporary mediation through electronics. By impoverishing interactions and social relations, more people are more susceptible to control by ideological state apparatuses than in the past.

Another difference that marks the early twenty-first century is that of an increasing mismatch between the forces of production and the relations of production. Relations of production follow a late nineteenth to twentieth century model. Those relations conform to monopoly capitalism and national economies, and still retain the character of Fordist mode of production. Fordist production is spatially centralized, and depends on a stationary workforce. Focus on electronic communication, computers, and the like misses the central changes and contradictions. Today production is increasingly decentralized in which component parts of products, raw materials, and capital machinery can be spread across a dozen or more locations. Automobile production offers a paradigmatic example. Auto parts might be crafted in 20 different countries, using raw materials from another 20 locales, with final assemblage not holding the determining role in manufacture. The most mobile part of the process is finance in which money speeds around the world via electronic transfers. Coordination is supplied by electronic cybernetics. Production, including finance, is mobile while workforces remain stationary, and often work is performed at some spatial remove from the manufacturing machines—that is, machine operators and the machines may be separated somewhat similarly to drone operators mentioned above. With increasing reliance on electronic cybernetics which replaces

mechanical cybernetics—for example, electronic sensors replacing mercury thermostats—fewer and fewer workers are needed to produce goods and services that make up the global commodity market. A consequence is the growing wealth and income gap, a phenomenon that has been a global phenomenon for at least the last 20 years. Where under Fordism capital required a reserve army of labour, today such a workforce is merely redundant. The potential for revolution becomes more unmistakable daily. To contain it, both the repressive and ideological apparatuses of the state work to keep the lid on. Being dazzled by speed, speed of communication, and technological virtuosity, misses the fundamental social changes currently underway.

CONCLUSION

In the abstract and in and of themselves, science and technology are morally neutral endeavours. The discovery of nuclear energy, for instance, merely furthered our knowledge of the physical universe. But science and technology do not remain abstract for long. As soon as they enter into the forces of production they become subject to the requirements of the current mode of production and the structure of social relations that are associated with it. Moreover, the prevailing ideology sets the truth conditions for all deployments of technology and even whether scientific discoveries are acknowledged. Historical examples abound, but to mention just a few, many of Nikola Tesla's discoveries and inventions have been ignored if not actively suppressed. Television was technically feasible in the 1920s and was available in a small region in the 1930s, but it did not become commercially viable until after the Second World War when the radio broadcast companies decided to embrace it. Mobile phones, which simply use the principle of point-to-point radio transmission were technologically possible in the first decade of the twentieth century, but it was only the US Telecommunications Act of 1996 (PL 104-104,

110 Stat. 56) that made them a reality. Penicillin was discovered in 1928, but it was not used as an antibiotic until 1942. It is not technology that makes for a better world; it is people. Only when humanity takes responsibility and works together to make a better world for everyone, will such a world be possible.

FUTURE RESEARCH DIRECTIONS

Current technology for the protection of people is not enough. It never will be until humanity takes into its own hands the welfare of all people. Future research must be humanistic in the broadest and deepest sense of the word. Until we force the technical to serve all human kind, it will forever serve only the interests of the privileged few.

REFERENCES

- Agamben, G. (2005). *State of Exception*. Chicago: University of Chicago Press. doi:10.1215/9780822386735-013
- Altheide, D. L. (2014). The Triumph of Fear: Connecting the Dots about Whistle blowers and Surveillance. *International Journal of Cyber Warfare & Terrorism*, 4(1), 1–7. doi:10.4018/ijcwt.2014010101
- Amar, G. (2011). *Homo Mobilis. The new era of mobility*. Buenos Aires: La Crujia.
- Arendt, H. (1963). *Eichmann in Jerusalem: a report on the banality of evil*. Viking Press.
- Barney, D. (2007). *One Nation Under Google*. Toronto: House of Anansi Press.
- Baudrillard, J. (2006). Virtuality and Events: the hell of power. *Baudrillard Studies*, 3(2). Available at <http://www.ubishops.ca/BaudrillardStudies/>
- Bauman, Z. (2008). *Liquid Fears contemporary society and its liquid fears*. Buenos Aires: Paidos Editorial.
- Beck, U. (2006). *Risk Society, towards a new modernity*. Buenos Aires: Paidos.
- Benko, S. (2005). Ethics, Technology, and post human communities. *Essays in Philosophy*, 6(1).
- Boguslaw, R. (1965). *The New Utopians: a study of system design and social change*. Englewood Cliffs, NJ: Prentice-Hall.
- Carr, N. (2011). *Is google making more stupid?*. In *The Digital Divide* (pp. 63–75). New York: Jeremy Tarcher.
- Casagerola, V. (2016). Who Screens Security? Cultures of surveillance in film. In R. Cropf & T. Bagwell (Eds.), *Ethical Issues and Citizens Rights in the era of Digital Government Surveillance* (pp. 57–75). Hershey, PA: IGI Global.
- Castel, R. (2006). *The Social Insecurity: what does it mean to be protected?* Buenos Aires: El Manantial.
- Cuevas, A. (2005). The many Faces of Science and Technology relationships. *Essays in Philosophy*, 6(1).
- Eid, M. (2014). *Exchange Terrorism Oxygen for Media Airwaves: the age of terroredia*. Hershey, PA: IGI Global. doi:10.4018/978-1-4666-5776-2
- Erikson, K. (1994). *A New Species of Trouble: explorations in disasters, trauma and community*.: Norton.
- Farris, J. (2005). Philosophy regarding technology. *Essays in Philosophy*, 6(1).
- Fiske, R. (2016). The borders of Corruption: living in the state of exemption. In R. Cropf & T. Bagwell (Eds.), *Ethical Issues and Citizens Rights in the era of Digital Government Surveillance* (pp. 1–15). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-9905-2.ch001
- Giddens, A. (1991). *Modernity and Self-Identity. Self and society in The late modern age*. Stanford University Press.

- Giddens, A. (2011). *The Politics of Climate Change*. Cambridge, UK: Polity Press.
- Hartzel, K., & Gerde, V. (2016). Using Duality theory to reframe E-Government Challenges. In R. Cropf & T. Bagwell (Eds.), *Ethical Issues and Citizens Rights in the era of Digital Government Surveillance* (pp. 35–56). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-9905-2.ch003
- Hugues, J. H. (2004). *Citizen Cyborg. Why democratic societies must respond to the redesigned human of the future*. Cambridge: West-view Press.
- Huxley, A. (2006). *Brave New World*. London: Harper Perennial Modern Classics.
- Kellner, D. (2005). Baudrillard, Globalization and Terrorism: some comments in recent adventures of the Image and Spectable on the occasion of Baudrillard's 75th birthday. Available at <http://www.ubishops.ca/BaudrillardStudies/>
- Kelman, D. (2012). *Counterfeit Politics, Secret Plots and Conspiracy Narratives in the Americas*. Bucknell University Press.
- Klein, N. (2007). *The Shock Doctrine: The Rise of Disaster Capitalism*. New York: Metropolitan Books.
- Korstanje, M. E. (2014). Why Risk-Research is More Prominent in English Speaking Countries in the Digital Society. *International Journal of Cyber Warfare & Terrorism*, 4(1), 8–18. doi:10.4018/ijcw.2014010102
- Korstanje, M. E. (2015). *A Difficult World, examining the roots of Capitalism*. New York: Nova Science.
- Korstanje, M. E., & Clayton, A. (2012). Tourism and terrorism: Conflicts and commonalities. *Worldwide Hospitality and Tourism Themes*, 4(1), 8–25. doi:10.1108/17554211211198552
- Korstanje, M. E., Tzanelli, R., & Clayton, A. (2014). Brazilian World cup 2014: Terrorism, tourism, and social conflict. *Event Management*, 18(4), 487–491. doi:10.3727/152599514X14143427352391
- Lyon, D. (2003). Technology vs terrorism: Circuits of city surveillance since September 11th. *International Journal of Urban and Regional Research*, 27(3), 666–678. doi:10.1111/1468-2427.00473
- Lyon, D., & Bauman, Z. (2013). *Liquid surveillance: A conversation*. Cambridge, UK: Polity.
- Marcuse, H. (1991). *One Dimensional Man. Studies in ideology of advanced industry society*. London: Routledge.
- Marx, K. (1932). *The German Ideology*. Retrieved from <https://www.marxists.org/archive/marx/works/1845/german-ideology/>
- Mueller, J., & Stewart, M. G. (2016). *Chasing Ghosts: The Policing of Terrorism*. New York: Oxford University Press.
- Murillo, S. (2008). *Colonizar el Dolor: la interpelación ideológica del Banco Mundial en América Latina: el caso argentino desde Blumberg a Cromañón*. Buenos Aires: CLACSO.
- Oppenheimer, J. R. (1965). *I am Become Death*. Retrieved from http://www.englishclub.com/ref/esl/Quotes/Death/I_am_become_death_the_destroyer_of_worlds._2620.htm
- Perrow, C. (1999). *Normal Accidents. Living with high risk Technologies*. Princeton, NJ: Princeton University Press.
- Ruse, M. S. (2005). Technology and the Evolution of the Human, from Bergson to the philosophy of Technology. *Essays in Philosophy*, 6(1).
- Sandler, T., & Enders, W. (2004). An economic perspective on transnational terrorism. *European Journal of Political Economy*, 20(2), 301–316. doi:10.1016/j.ejpoleco.2003.12.007

Schmitt, C. (1985). *Political Theology: Four Chapters on the Concept of Sovereignty*. Cambridge, MA: MIT Press.

Skoll, G., & Korstanje, M. (2012). Risks, Totems, and Fetishes in Marx and Freud. *Sincronía: A Journal for the Humanities and Social Sciences*, 17(61), 1–19.

Skoll, G. R. (2014). Stealing Consciousness: Using Cybernetics for Controlling Populations. *International Journal of Cyber Warfare & Terrorism*, 4(1), 27–35. doi:10.4018/ijcwt.2014010104

Sorman, G. (2002). *The Progress and its enemies*. Buenos Aires: Emece.

Steiner, C., & Al-Hamarneh, A. (2005). Islamic tourism: Rethinking the strategies of tourism development in the Arab world after September 11, 2001. *Comparative Studies of South Asia, Africa and the Middle East*, 24(1), 173–182.

Sunstein, C. (2002). *Risk and Reason*. New York: Cambridge University Press.

Tarlow, P. (2014). *Tourism Security: Strategies for Effectively Managing Travel Risk and Safety*. Oxford: Elsevier.

Virilio, P. (1996). *El Arte del Motor: aceleración y realidad*. Buenos Aires: Ediciones el Manantial.

Virilio, P. (2007). *Ciudad Pánico: el afuera comienza aquí*. Buenos Aires: Libros el Zorzal.

Virilio, P. (2010). *The University of Disaster*. Oxford, UK: Polity Press.

ADDITIONAL READING

Alderman, E., & Kennedy, C. (1997). *The right to privacy*. New York: Vintage.

Altheide, D. L. (2004). Consuming terrorism. *Symbolic Interaction*, 27(3), 289–308. doi:10.1525/si.2004.27.3.289

Altheide, D. L. (2006). Terrorism and the Politics of Fear. *Cultural studies critical methodologies*, 6(4), 415–439.

Altheide, D. L. (2009). *Terror post 9/11 and the media* (Vol. 4). New York: Peter Lang.

Bazan, E. B. (2002). *The Foreign Intelligence Surveillance Act*. Nova Publishers.

García-Peñalvo, F. J., González-González, J. C., & Murray, M. (2012). Myelvin: A web-based informal learning platform for languages practice. *International Journal of Knowledge Society Research*, 3(1), 26–39. doi:10.4018/jksr.2012010103

Howie, L. (2007). The terrorism threat and managing workplaces. *Disaster Prevention and Management: An International Journal*, 16(1), 70–78. doi:10.1108/09653560710729820

Korstanje, M. E. (2013). Preemption and terrorism. When the future governs. *Cultura (Asociación de Docentes de la Universidad de San Martín de Porres)*, 10(1), 167–184.

Korstanje, M. E. (2015). The spirit of Terrorism: Tourism, Unionization and Terrorism. *Pasos: Revista de turismo y patrimonio cultural*, 13(1), 239–250.

Korstanje, M. E., & Clayton, A. (2012). Tourism and terrorism: Conflicts and commonalities. *Worldwide Hospitality and Tourism Themes*, 4(1), 8–25. doi:10.1108/17554211211198552

Korstanje, M. E., Tzanelli, R., & Clayton, A. (2014). Brazilian World cup 2014: Terrorism, tourism, and social conflict. *Event Management*, 18(4), 487–491. doi:10.3727/152599514X14143427352391

Kraft, E., & Wang, J. (2010). An exploratory study of the cyberbullying and cyberstalking experiences and factors related to victimization of students at a public liberal arts college. *International Journal of Technoethics*, 1(4), 74–91. doi:10.4018/jte.2010100106

Lyon, D. (2002). Surveillance Studies: Understanding visibility, mobility and the phenetic fix. *Surveillance & Society*, 1(1), 1–7.

Lyon, D. (2003). *Surveillance after september 11* (Vol. 11). Cambridge: Polity.

Lyon, D. (2004). Globalizing Surveillance Comparative and Sociological Perspectives. *International Sociology*, 19(2), 135–149. doi:10.1177/0268580904042897

Lyon, D. (2007). *Surveillance studies: An overview*. Cambridge: Polity.

Lyon, D., & Zureik, E. (Eds.). (1996). *Computers, surveillance, and privacy*. Minneapolis: U of Minnesota Press.

Mason, R. O. (1986). Four ethical issues of the information age. *Management Information Systems Quarterly*, 10(1), 5–12. doi:10.2307/248873

OBrien, M., Tzanelli, R., Yar, M., & Penna, S. (2005). The Spectacle of Fearsome Acts: Crime in the Melting P (l) ot in Gangs of New York. *Critical Criminology*, 13(1), 17–35. doi:10.1007/s10612-004-6111-9

Palanisamy, R., & Mukerji, B. (2012). Security and Privacy issues in e-Government. IGI Global, 236–248.

Reitz, J. C. (2006). E-government. *The American Journal of Comparative Law*, 54, 733–754.

Shaw, L. E. (2004). USA Patriot Act of 2001, the Intelligence Reform and Terrorism Prevention Act of 2004, and the False Dichotomy between Protecting National Security and Preserving Grand Jury Secrecy, The. *Seton Hall Law Review*, 35, 495.

Tzanelli, R. (2004). Giving gifts (and then taking them back): Identity, reciprocity and symbolic power in the context of Athens 2004. *Journal for Cultural Research*, 8(4), 425–446. doi:10.1080/1479758042000310117

Tzanelli, R. (2006). Capitalizing on value: Towards a sociological understanding of kidnapping. *Sociology*, 40(5), 929–947. doi:10.1177/0038038506067516

Wade, L. (2003). Terrorism and the internet: Resistance in the information age. *Knowledge, Technology & Policy*, 16(1), 104–127. doi:10.1007/s12130-003-1018-4

KEY TERMS AND DEFINITIONS

Digital Surveillance: The use of digital technology to control others, citizens, and the life of a nation.

Disaster: It represents A disruption of functioning of a society that leads to economic or material losses, affecting the tolerance of society to react. Disasters appear when the possibility to answer towards risk are vulnerated.

Ideology: A cognitive framework of truth conditions.

Nuclear Power: It is The use of nuclear energy to produce heat which is channelled steam turbines to get electricity. Nuclear power has escalated a tension between US and Soviet Union during Cold War.

Plot: Secret narrative that evokes a theory of conspiracy which is hidden from the public.

Privacy: Right of peoples and their group to keep their own information from others.

Risk Perception: It is The perception of any potential threat or danger which can affect the well-functioning of society, groups or people.

Terrorism: v. Violent coercion of other people.

Users Behavioral Intention Towards eGovernment in an African Developing Country

Ayankunle A. Taiwo

Texas A&M University – Commerce, USA

INTRODUCTION

Sequel to successful implementation of electronic commerce in the private sector, huge amounts has been invested in e-Government globally. Nevertheless, evaluating results of the investments using numerous metrics shows disappointment in most developing countries. With impending failures of e-Government initiatives in developing nations (Maumbe, Owei & Alexander, 2008), it is important to note that implementation of e-Government without redesigning the processes is tantamount to waste of public funds and resources. Having targeted users in mind brings about efficient and effective e-government services, not bureaucracy. Importantly, the accomplishment of e-Government initiatives hinges on rethinking and redesigning processes that the government operates. Also, as critical as technologies is to e-Government, its' optimal results is not exclusively dependent on technological innovation, but the combination of technology and end users. Furthermore, channels of distributing government solutions ought not to make the difference, but high quality of services through the channels.

Presently, public sector of many developed nations utilizes the power of the Internet as an outlet for public service provision. Flavin, Guinaliu, & Torres (2006) iterated that arrival of internet came with substantial advantages for all stakeholders concerned. After a successful e-commerce exploits by the private sector, government of many countries sort after provision of public services over the internet. e-Government comes with enormous

gains to the governments, including reduction in delivery cost of public services with a wider reach to the people. Furthermore, citizens can have access to varieties of solutions because operations are no more time bound by office hours. Although lots of funds have been spent on building e-Government systems, studies have shown that citizen may not use the systems for many reasons (Olasina, 2012). The most important stakeholders (citizens) in e-Government are often left behind while crucial decisions are made. Interestingly, the greater impact is on the citizens, because their behavioral pattern will have to change. In addition, there have always been resistant to change, thus citizens acceptance of e-Government can be complicated, because it affects human behavioral patterns (Meuter et al., 2002). Therefore, reviewing factors that influences human behavior towards technology cannot be overemphasized. The objective of this paper is to investigate behavioral intention towards the use of eGovernment service in Nigeria, a developing country.

BACKGROUND

The National Information Technology Development Agency (NITDA) was initiated by the Nigerian government as a passage for eGovernment to bring government closer to its citizens and facilitate national strategies for the spread of e-society / e-government (NITDA, 2001; Obasanjo, 2003, 2004). The expectation of the eGovernment initiatives is to use Information & Com-

munication Technologies (ICT) for breaking “barriers of hierarchical traditions, secrecy, and bureaucracies” (Obasanjo, 2004; p1) associated with public services. Numbers of eGovernment initiatives were rolled out to prepare citizens for the change towards eGovernment, encourage citizens to use eGovernment applications and build trust in citizens to reduce future resistance (Ifinedo, 2007). The implementation of eGovernment was also foreseen as a way to curtail the effect of corrupt practices as eGovernment initiatives were employed in India to combat corruption (Ifinedo, 2007). Also, e-government initiatives were expected to improve culture, transparency and accountability in the country.

Recent studies shows that the implementation and the use of eGovernment in Nigeria is less than desirable while some of the challenges facing eGovernment initiatives are yet to be resolved (Ifinedo, 2005; 2007). A research survey conducted by Awoleye, Oluwaranti, Siyanbola & Adagunodo (2008) revealed that the use of eGovernment application is low and 20% of the respondents out rightly lack interest in visiting government websites. The current state of eGovernment in Nigeria is beyond preliminary stages. There are several eGovernment services available across the country at local, state and federal level, many of which are not patronised due to workers negative attitude to change and citizens’ negative perceptions about eGovernment (Olasina, 2012). Some of these services could be found on the website for government services “services.gov.ng”. Nigerian immigration eservices are one of the surviving and successful eGovernment initiatives in Nigeria. The targeted population in this study, undergraduate and post-graduate students, are individuals who have used online services provided by the Nigerian government for immigration (immigration.gov.ng) purposes. Chete et al., (2015) investigated the use of SMS based eGovernment services in Lagos Nigeria and suggested that mass media is the most effective way to show citizens the benefits of using eGov-

ernments services. In addition, there are limited empirical studies on eGovernment in Sub Sahara African nations (Maumbe et al., 2008). Lastly, Ifinedo (2007) suggested the investigation of the perceptions of Nigerians regarding the eGovernment acceptance in Nigeria.

LITERATURE REVIEW

Unified Theory of Acceptance and Use of Technology (UTAUT)

The acceptance of information technology has been carefully investigated dynamically for the last four decades. Looking through the technology acceptance literature, many studies proposed models and examined them to show vividly more than forty percent variance in a persons’ intention to adopt technological a innovation (Davis et al., 1989; Venkatesh & Morris, 2000). Venkatesh et al (2003) made a concrete examination of eight (8) prominent models and derived a unified concept of acceptance and use of technology (UTAUT) which could explain around 70 percent of variance in intention. The eight (8) technology adoption models studied consist of the theory of reasoned action (TRA), the technology acceptance model (TAM), the motivational model (MM), the theory of planned behavior (TPB), a model combining TAM and TPB (C-TAM-TPB), the model of PC utilization (MPCU), the innovation diffusion theory (IDT), and social cognitive theory (SCT).

Performance expectancy, effort expectancy, and social influence of the UTAUT model show significant immediate determinants of behavioral intention while facilitating condition is a determinant of usage behavior (Venkatesh et al., 2003; Taiwo & Downe 2012). These constructs have usually been validated in lots of empirical reports for being essential factors that influences users’ acceptance of a system (Guo & Barns, 2007; Iahad & Rahim, 2011; Taiwo et al., 2014).

Relationship Between Performance Expectancy and Behavioural Intention

In many previous studies, it has been found that a positive relationship exist between performance expectancy and behavioral intention towards using a particular technology. Goh and Yoon (2011) found performance expectancy as a predictor of behavioural intention in an empirical study to investigate facilitators of virtual world acceptance. In addition, Venkatesh, Sykes and Zhang (2011) investigated the acceptance of electronic medical record (EMR) system by medical doctors, and found that performance expectancy significantly predicts behavioural intention to accept and use the EMR system. Besides, Foon and Fah (2011) found performance expectancy to be significant in predicting users' intention to accept and use Internet banking. Nawaz & Thelijagoda (2015) investigated citizens' behaviour towards eGovernment services in Sri Lanka and found performance expectancy to be a significant predictor of intention to use eGovernment services. Gupta, Singh & Bhaskar (2016) investigated citizens' adoption of eGovernment services at New Delhi Municipal Council (NDMC), results suggests that performance expectancy is significant factor that predicts intention to use eGovernment services. As such it is hypothesized that:

H1: Performance expectancy has a direct positive relationship with behavioural intention.

Relationship Between Effort Expectancy and Behavioural Intention

It is found that significant positive relationship exist between effort expectancy and behavioural Intention. Venkatesh, Sykes and Zhang (2011) investigated the acceptance of electronic medical record (EMR) system by medical doctors and found that effort expectancy significantly predicts behavioural intention to accept and use the EMR

system. Also, Wang and Shih (2009) found effort expectancy as a significant predictor of behavioural intention to accept the use of information kiosks in Taiwan. Niehaves and Plattfaut (2010) found effort expectancy to be a predictor of intention to use Internet amongst elderly citizen in Europe. Furthermore, Chiu, Fang and Tseng (2010) also found effort expectancy to significantly predict consumers' intention to accept the use of a kiosk technology. Nawaz & Thelijagoda (2015) investigated citizens' behaviour towards eGovernment services in Sri Lanka and found effort expectancy to be a significant predictor of intention to use eGovernment services. Furthermore, Gupta et al., (2016) explored intention to adopt eGovernment services at a local Municipal Council and found that effort expectancy have a significant positive relationship with intention to use eGovernment services. As such it is hypothesized that:

H2: Effort expectancy has a direct positive relationship with behavioural intention.

Relationship Between Social Influence and Behavioural Intention

Previous researchers confirmed significant relationship between social influence and intention to adopt. Maldonado et al (2009) employed only social influence amongst the three UTAUT predictor of intention alongside with some other cognitive variables, and found out that social influence significantly predict students' intention to use an educational portal. Also, Foon and Fah (2011) found social influence to be significant in predicting users' intention to accept and use Internet banking. Wang and Shih (2009) investigated the acceptance and use information kiosks using the UTAUT and found out that social influence significantly predict behavioural intention towards the use of the information kiosk. In a study to know what factor would predict citizens' use of Internet services, Niehaves & Plattfaut (2010) found effort expectancy to be a predictor of elderly citizens' intention to use Internet in Europe. Nawaz &

Thelijjagoda (2015) investigated behavioural intention to use eGovernment services in Sri Lanka and found social influence to have a significant influence on intention to use eGovernment services. Gupta et al., (2016) investigated citizens' adoption of eGovernment services at New Delhi Municipal Council (NDMC), results suggests that social influence have a significant influence on intention to use eGovernment services. As such it is hypothesized that:

H3: Social Influence has a direct positive relationship with behavioural intention.

Relationship Between Facilitating Conditions and Behavioural Intention

Facilitating conditions are the degree to which people believes that there are organizational and technical infrastructures are available to support the use of a system (Venkatesh et al., 2003). Venkatesh, Sykes and Zhang (2011) investigated the acceptance of electronic medical record (EMR) system by medical doctors and found that facilitating condition significantly predicts behavioural intention to accept and use the EMR system. Similarly, Foon & Fah (2011) found facilitating condition to be significant in predicting users' intention to accept and use Internet banking.

Similarly, Niehaves & Plattfaut (2010) found facilitating condition to significantly influence intention to use Internet amongst elderly citizen in Europe. Similarly, Chiu, Fang and Tseng (2010) also found facilitating conditions to significantly predict consumers' intention to accept the use of a kiosk technology. As such it is hypothesized that:

H4: Facilitating conditions have a direct positive relationship with behavioural intention.

Web Trust

Mayer et al (1995) gave a simple definition of Trust as the belief that other party will behave in

a socially accountable way. Therefore, the trusted will fulfill the trusting party's expectations without taking benefit of its vulnerability. Taking a cue from McKnight & Chervany (2002), perception in e-Government in relation to trust can be described as a combination of beliefs that make it easy for a citizen to willingly become susceptible to egovernment after taking careful thoughts into its characteristics before usage. According to Rexha et al (2003), absence of trust is one of the main reasons why certain people are still reluctant to conduct financial transactions online or give away personal information over the internet.

Importantly, citizens are more likely to use web services offered by government agencies with very good reputation. For that reason, citizen's confidence in the e-government provider is crucial to e-government acceptance. Trust and trust related issues on electronic services have been investigated widely in e-government research (Carter&Belanger; Welch, Hinnat& Moon, 2005, Taiwo, Downe & Loke 2014). Several studies have used the web trust by McKnight et al (2002) to investigate users' intention (Metehan, & Yasemin, 2011).

Relationship Between Disposition to Trust and Behavioural Intention

It has been found that there exist a positive relationship between disposition to trust and Behavioural Intention. McKnight, Chevany and Kacmar (2002) applied the web trust model to investigate the human trust on the internet. The construct disposition to trust was validated as reliable alongside trust intentions and was found to have a significant influence on personal innovativeness of the individual. Colesca (2009) investigated the importance of trust in eGovernment acceptance and found disposition to trust (propensity to trust) as an important factor that predicts intention to accept a technology. Taiwo et al (2014) investigated the effect of citizens' disposition to trust on behavioural intention to adopt eGovernment services and found disposition

to trust as a significant predictor of intention. As such it is hypothesized that:

H5: Disposition to Trust has a direct positive relationship with behavioural intention.

Relationship Between Institution Based Trust and Behavioural Intention

In relation to institution based trust, it often used as trust of Internet or Technology (Belanger et al., 2008; Teo et al., 2009). Trust in Internet for government services (eGovernment) comprises of the conventional view of trust as a specific entity (government) and the dependability of the Institution that enables technology (trust of the Internet) (Carter and Belanger, 2005; Belanger and Carter, 2008). Colesca (2009) investigated the importance of trust in eGovernment acceptance and found institution based trust (trust in technology) as an important factor that predicts intention to accept a technology. Belanger and Carter (2008) investigated the influence of trust and risk in eGovernment adoption and found Institution based trust (trust of Internet) to be an important factor in determining acceptance and adoption of eGovernment services. Furthermore, in an empirical study to uncover factors that could significantly influence the decision of citizens of developing countries to accept eGovernment services, Taiwo et al (2014) found institution based trust to have a significant positive effect on behavioural intention to adopt eGovernment services. Also, Gupta et al., (2016) found trust of technology to have a significant relationship with intention to use eGovernment services at a local municipal council in India. As such it is hypothesized that:

H6: Institution based trust has a direct positive relationship with behavioural intention.

Relationship Between Trusting Belief and Behavioural Intention

Trust belief is a construct that has gained wide attention of IS researchers on users behavioural intention on the Internet. The construct has also been adapted in several technology adoption studies (Chen, Liu, Song and Qian 2008; Cheng, Liu, Qian 2008).

Citizens with high Trusting belief in a government agency are more likely to accept eGovernment services. Dashti, Benbasat and Burton-Jones (2009) investigated the role of trusting belief in Government, and found trusting belief (trust in eGovernment) to be an important predictor of intention to adopt eGovernment. Gupta et al., (2016) found Trust in Government (trusting belief) to have a significant relationship with intention to use eGovernment services at a local Municipal Council in India. Using the above discussion and other studies that have applied the construct in the field of eGovernment trust (Belanger and Carter 2008; Teo et al 2009, Taiwo et al 2012; Taiwo et al., 2014). As such it is hypothesized that:

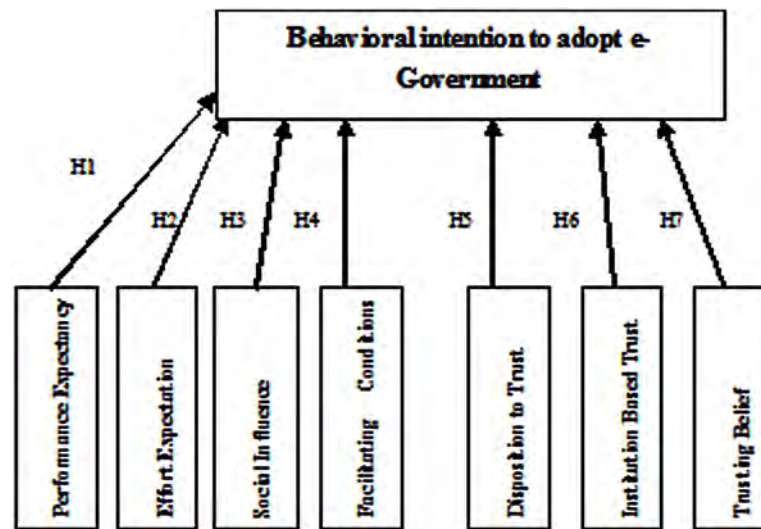
H7: Trusting Belief has a direct positive relationship with behavioural intention.

METHODOLOGY

Data Collection

Quantitative research method was employed in this study to investigate the intention to use eGovernment services in Nigeria. Therefore, using a face to face approach, a total of 500 questionnaires were randomly distributed among undergraduate and postgraduate students in four Nigerian Universities (University of Ilorin, Federal University of Technology Minna, Ladoke Akintola University of Technology and Ajayi Crowther University). The rationale behind the selection of student popula-

Figure 1. Research framework



tion is that; they represent hopeful future and early adopter of eGovernment services in the country. The questionnaire consists of ten (10) sections. The first section contains screening questions regarding the use of Nigerian immigration website and eGovernment in general. The second section collected demographic data. While the remaining sections collected data on the UTAUT and Web Trust dimensions. Out of 500 subjects, 310 usable questionnaires were received giving a response rate of sixty two percent (62%).

Measures

Using the literature as a guide we integrated dimensions that could possibly reveal the factors that affect citizens' behavioral intention and developed structural model to see the influence of the dimensions on behavioural intention. Our research model (see Fig.1) shows the proposed model which integrates the Performance Expectancy, Effort Expectancy, Social Influence and Facilitating Condition with Web trust dimensions; Disposition to Trust, Institution based Trust and Trusting Belief to investigate how the combination can provide a more a better understanding towards citizens' intension to use the identified e-Government service.

DATA ANALYSIS AND RESULTS

Preliminary Statistics

The gender distribution of the study subjects was forty four percent (44%) females and fifty six percent (56%) males respectively. Respondents between the ages of 18-30 formed the largest age group ninety percent (90%) with a total of 189. Sixty percent (60%) of the respondents had communicated with the government online before. The cronbach alphas of all the constructs are above the recommended value of 0.7 except for peer influence at 0.639, which is considered satisfactory for this study (see Table 1).

Furthermore, Pearson's correlation between the observed constructs, findings revealed that the there are moderate correlation between the constructs. Table 2 shows the correlation between the variables.

Discriminate validity was assessed by calculating Composite reliabilities in the measurement model which ranged from 0.6900 to 0.8217 (see Table 3). Nunnally& Bernstein, (1994) recommended cutoff of 0.70, but social influence construct has a (CR = 0.6900), but its value is close to 0.70, therefore it was retained for further analysis. Convergent and discriminant validity was

Table 1. Constructs reliability and validation table (N=310)

Variables	Total Item	Chrobach Alpha	Fit Indices						
			χ^2	df	χ^2/df	GFI	CFI	RMSEA	p-Value
Performance Expectation (PE)	4	0.707	1.45	2	0.725	0.99	0.99	0.000	0.48
Effort Expectancy (EE)	5	0.700	5.56	5	1.112	0.99	0.99	0.023	0.35
Social Influence (SI)	4	0.639	1.24	2	0.62	0.99	0.99	0.000	0.54
Facilitating Conditions (FC)	4	0.774	0.68	2	0.34	0.99	0.99	0.000	0.71
Disposition to Trust (DTT)	5	0.793	5.05	5	1.01	0.99	0.99	0.007	0.41
Institution-Based Trust (IBT)	5	0.709	6.08	5	1.216	0.99	0.99	0.032	0.30
Trusting Belief (TB)	4	0.705	2.29	2	1.145	0.99	0.99	0.026	0.32
Behavioral Intention(BI)	3	0.702	0	0	0	1	1	0.00	1

assessed in terms of average variance extracted (AVE). All constructs have an AVE greater than 0.5 as recommended by (Fornell & Larcker, 1981), except social influence construct has a (AVE = 0.4905), but its value is close to 0.50, therefore it was retained for further analysis (see Table 3).

Structural Model

A structural equation modeling (SEM) was used to analyze the data using Listrel 8.52. By categoriz-

ing and combining all the predictors found in the literature, a structural model was developed and tested to see the effects of all endogenous variables on behavioral intention. The study revealed that Performance expectancy, effort expectancy and facilitating condition are significantly related to behavioural intention with standardized coefficients of ($\beta=0.21, p<0.05$), ($\beta=0.23, p<0.05$) and ($\beta=0.15, p<0.05$) respectively. Social influence does not have a significant influence on behavioural intention with a standardized coefficient of

Table 2. Correlation matrix, component reliability, AVE and roots of the AVEs (shown as diagonal elements) (n=310)

	Composite Reliability	AVE	1	2	3	4	5	6	7	8
BI	0.7040	0.5275	.726							
PE	0.8217	0.5120	.673	.716						
EE	0.7046	0.5090	.681	.639	.713					
SI	0.6900	0.4905	.687	.564	.604	.700				
FC	0.7260	0.5155	.494	.469	.610	.465	.717			
DTT	0.8019	0.5532	.591	.419	.558	.544	.430	.744		
IBT	0.7538	0.5008	.647	.601	.636	.615	.538	.611	.708	
TB	0.7273	0.5132	.737	.697	.662	.677	.531	.583	.735	.716

**Correlation is significant at the 0.01 level (2-tailed) Average variance extracted (AVE) = (sum of squared standardized loading)/[(sum of squared standardized loadings) + (sum of indicator measurement error)]. Composite reliability = (sum of standardized loadings)²/[(sum of standardized loadings)² + (sum of indicator measurement error)]. Indicator measurement error can be calculated as 1 - (standardized loading)².

Figure 2. Structural model

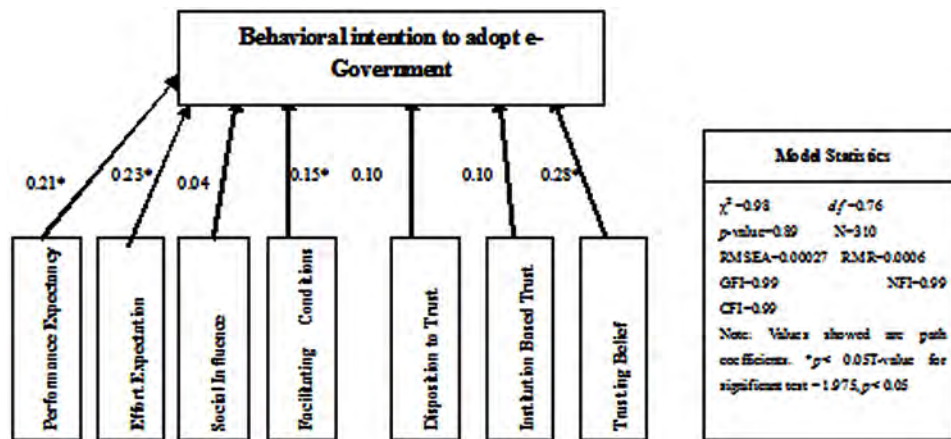


Table 3. Hypothesis testing results (path analysis)

Hypothesis	Path	β	T	Comment
H1	PE $\rightarrow$ BI	0.21	3.96	Supported
H2	EE $\rightarrow$ BI	0.23	4.98	Supported
H3	SI $\rightarrow$ BI	0.04	0.74	Not Supported
H4	FC $\rightarrow$ BI	0.15	3.27	Supported
H5	DTT $\rightarrow$ BI	0.10	1.19	Not Supported
H6	IBT $\rightarrow$ BI	0.10	1.67	Not Supported
H7	TB $\rightarrow$ BI	0.28	5.73	Supported

($\beta=0.04$, $p<0.05$). Amongst the Web Trust dimensions, only trusting belief have positive significant influence on behavioural intention with a standardized coefficients of ($\beta=0.28$, $p<0.05$). Disposition based trust, Institution based trust does not have significant influence on behavioural intention with a standardized coefficients of ($\beta=0.10$, $p<0.05$) and ($\beta=0.10$, $p<0.05$) respectively. Furthermore, the goodness of the theoretical model was established by the variance explained (R^2) of the independent variables on behavioral intention as 0.71, which implies that the model accounts for 71 percent of variance in intention to use eGovernment. Therefore, H1, H2, H4 and H7 are accepted. While H3, H5, H6 were not accepted (See Figure 2 and Table 3). Furthermore, with the exception of the p-value all other fit indices are above the recommended values (Hair et al 2006; Suki & Ramayah, 2010).

DISCUSSION

The purpose of this study is to examine behavioral intention to use eGovernment services in Nigeria using the UTAUT dimensions; Performance Expectation, Effort Expectancy, Social Influence and Facilitation Conditions. And three dimensions of the Web Trust Model; Disposition to Trust, Institution Based Trust and Trusting Belief. The outcome of the study revealed that Performance Expectancy has a significant relationship towards behavioural intention. This is in line with previous studies by Niehaves & Plattfaul (2010), Venkatesh et al (2011), Nawaz & Thelijagoda (2015), Gupta, et al., (2016).

Effort Expectancy has a significant relationship towards behavioural intention. This is consistent with previous studies such as Niehaves & Plattfaul

(2010) and Venkatesh et al (2011). This suggests that people will use eGovernment services when they perceive that it will assist them to gain efficiency receiving services from the Government conveniently.

Facilitating Condition has a significant relationship towards behavioural intention. This is consistent with previous studies such as Dashti,et al., (2009)and Nawaz & Thelijjagoda (2015). This suggests that people will use eGovernment services when they perceive that it will be easy for them to use.

Trusting Belief has a significant effect on behavioural intention. This is in line with the studies of Dashti,et al., (2009), Taiwo et al., 2014 and Gupta, et al., (2016). This suggests that people will use eGovernment services when they have confidence in the government to fulfill its obligation to its citizens.

In this study, Social Influence, Disposition to Trust and Institution based Trust do not have significant influence on behavioural intention. This could be as a result of the high level of education of respondents employed in this study.

The proposed structural model shows that Performance Expectation, Effort Expectancy, Facilitation Conditions and Trusting Belief can explain 71% of the variance in behavioural intention to use eGovernment. This shows that the model has relatively good predictive power on behavioural intention to use eGovernment.

LIMITATION AND SUGGESTIONS FOR FUTURE RESEARCH

This study is not without some limitation. First moderating effects in the previous UTAUT models were not explored. Second, most of our respondents are working students and it would be interesting to see the outcome of other samples. Future works could explore the effect of moderators and other dimensions on intention.

CONCLUSION

Towards a successful acceptance of eGovernment, it is expedient for the government to focus on actual issues that surrounds the use of electronic services provided to the people. Such as working on trust related strategies before and after introducing e-services to the people. Improving on such strategy can improve positive perception of the people towards using eGovernment. As efforts are currently on ground towards accountability and transparency in governance, such efforts should be intensified to increase citizens' belief in the government that they will fulfill their electoral promises.

The outcome of this study revealed that trust in relevant government agencies and institution can be of great importance in creating positive perception for citizens to accept e-Government initiatives. Trust has been an important composite of any good relationship as people will promote the services with words of mouth (Ongori, 2009). Most importantly, this study shows that trusting belief in government agencies is a strong predictor of citizens' intention to adopt eGovernment. Therefore, Nigerian government must endeavour to increase trusting relations with its citizens. Because when there is high level of trust in the government, perception of citizens towards e-Government will be positive and the likelihood of acceptance will increase.

REFERENCES

- Awolaye, M., Oluwaranti, A., Siyanbola, W., & Adagunodo, R. (2008). Assessment of eGovernance Resource Use in South-Western Nigeria. In *Proceedings of the 2nd international conference on Theory and practice of electronic governance*. New York, NY: ACM.
- Belanger, F., & Carter, L. (2008). Trust and risk in e-government adoption. *The Journal of Strategic Information Systems*, 17(2), 165–176. doi:10.1016/j.jsis.2007.12.002

- Carter, L., & Belanger, F. (2005). The Utilization of E-Government Services: Citizen Trust, Innovation and Acceptance Factors. *Information Systems Journal*, 15(1), 2–25. doi:10.1111/j.1365-2575.2005.00183.x
- Cheng, D., Liu, G., & Qian, C. (2008). On Determinants of User Acceptance of Internet Banking: A Theoretical Framework and Empirical Study. *Advanced Management of Information for Globalized Enterprises, AMIGE, IEEE Symposium*, 1, 1-5.
- Cheng, D., Liu, G., Qian, C., & Song, Y.-F. (2008). Customer acceptance of internet banking: Integrating trust and quality with UTAUT model. *Service Operations, Logistics and Informatics*, 1, 383–388.
- Chete, F. O., Oyemade, D., Abere, R., Chiemeké, S. C., & Ima-Omasogie, I. (2015). Citizens Adoption Of SMS Based E-Government Services In Lagos State, Nigeria. *Journal of Emerging Trends in Computing and Information Sciences*, 3(4), 654–660.
- Chiu, C.-M., & Wang, E. T. G. (2008). Understanding web-based learning continuance intention: The role of subjective task value. *Information & Management*, 45(3), 194–201. doi:10.1016/j.im.2008.02.003
- Chiu, Y.-T. (2010). Early versus potential adopters Exploring the antecedents of use information in the context of retail service innovations. *International Journal of Retail & Distribution Management*, 38(6), 443–459. doi:10.1108/09590551011045357
- Colesca, S. E. (2009). Increasing E-Trust: A Solution To Minimize Risk In E-Government Adoption. *Journal of Applied Quantitative Methods*, 4(1), 31–44.
- Dashti, A., Benbasat, I., & Burton-Jones, A. (2009). Developing Trust Reciprocity In Electronic government: The Role Of Felt Trust. In *Proceeding at European Mediterranean Conference on Information Systems*. Izmir, Turkey: Brunel University.
- Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management Science*, 35(8), 982–1002. doi:10.1287/mnsc.35.8.982
- Flavián, C., Guinalíu, M., & Torres, E. (2006). How bricks-and mortar attributes affect online banking adoption. *International Journal of Bank Marketing*, 24(6), 406–423. doi:10.1108/02652320610701735
- Foon, Y. S., & Fah, B. C. Y. (2011). Internet banking adoption in Kuala Lumpur: An application of UTAUT model. *International Journal of Business and Management*, 6(4), 161–167.
- Goh, S., & Yoon, T. (2011). If you build it will they come? An empirical investigation of facilitators and inhibitors of hedonic Virtual world acceptance. In *Proceedings of the 44th Hawaii International conference on system sciences*. Kauai, HI: ACM. doi:10.1109/HICSS.2011.233
- Guo, Y., & Barnes, S. (2007). Why people buy virtual items in virtual worlds with real money. *The Data Base for Advances in Information Systems*, 38(4), 69–76. doi:10.1145/1314234.1314247
- Gupta, K. P., Singh, S., & Bhaskar, P. (2016). Citizen adoption of e-government: a literature review and conceptual framework. *Electronic Government, an International Journal*, 12(2), 160–185.
- Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (2006). *Multivariate data analysis* (6th ed.). Upper Saddle River, NJ: Prentice-Hall.
- Ifinedo, P. (2005). Measuring Africa's e-readiness in the global networked economy: A nine-country data analysis. *International Journal of Education and Development Using ICT*, 1(1), 53–71.
- Ifinedo, P. (2007). Moving Towards E-Government in a Developing Society: Glimpses of the Problems. In L. Al-Hakim (Ed.), *Global E-government: Theory, Applications and Benchmarking* (pp. 148–166). Hershey, PA: Idea Group Inc. doi:10.4018/978-1-59904-027-1.ch009

- Maldonado, U. P. T., Khan, G. F., Moon, J., & Rho, J. J. (2009). E-learning motivation, Students' acceptance/use of educational portal in developing countries: A Case of Peru. In *Proceedings of the 4th International conference on computer sciences and the convergence information Technology*. Seoul, Korea: IEEE. doi:10.1109/ICCIT.2009.77
- Maumbe, B. M., Owei, V., & Alexander, H. (2008). Questioning the pace and pathway of e-government development in Africa: A case study of South Africa's Cape Gateway project. *Government Information Quarterly*, 25(4), 757–777. doi:10.1016/j.giq.2007.08.007
- Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An Integrative Model Of Organizational Trust. *Academy of Management Review*, 20(3), 709–734.
- McKnight, D. H., & Chervany, N. L. (2002). What Trust Means In Ecommerce Customer Relationships: An interdisciplinary conceptual typology. *International Journal of Electronic Commerce*, 6(2), 35–60.
- McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and Validating Trust Measures for Ecommerce: An Integrative Typology. *Information Systems Research*, 13(3), 334–359. doi:10.1287/isre.13.3.334.81
- Metehan, T., & Yasemin, Z. A. (2011). The Effect of Web Vendor Trust On Turkish Online Shoppers Buying Behavior. *Australian Journal of Business and Management Research*, 1(6), 87–96.
- Meuter, M. L., Ostron, A. L., Roundtree, R. I., & Bitner, M. J. (2002). Self-service technologies: Understanding customer satisfaction with technology-based service encounters. *Journal of Marketing*, 64(3), 50–64. doi:10.1509/jmkg.64.3.50.18024
- Nawaz, S. S., & Thelijjagoda, S. (2015). Sri Lankan Citizens' Use Behaviour towards E-Government Services. *European Journal of Business and Management*, 7(8), 219–231.
- Niehaves, B., & Plattfaut, R. (2010). What is the issue with internet acceptance amongst elderly citizens? In *Proceedings of the 9th International Conference on E-Government*. Lausanne, Switzerland: Springer.
- NITDA. (2001). *NITDA begins strategic action plan for the implementation of Nigeria IT policy*. Retrieved May 16, 2015, from <http://www.nitda.gov.ng>
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw Hill.
- Obasanjo, O. (2003). *The president of Nigeria's speech delivered at the World Summit on the Information Society*. Geneva, Switzerland: WSIS.
- Obasanjo, O. (2004). *Speech by His Excellency, President Olusegun Obasanjo at the Stakeholders Conference on National eGovernment Strategies and Implementation for Nigeria*. Retrieved Dec 16, 2015, from http://www.nigeriafirst.org/objspeeches/2004/stakeholders_conf.html
- Olasina, G., (2012). A review of e-government services in Nigeria. *Library and Information Science in Developing Countries Contemporary Issues*, 205–221.
- Ongori, H. (2009). Role of Information Communication Technologies Adoption in SMES: Evidence from Botswana. *Research Journal of Information Technology*, 1(2), 79–85. doi:10.3923/rjit.2009.79.85
- Oye, N. D., Iahad, N. A., & Rahim, N. Z. (2011). An Application of the UTAUT Model for Understanding Acceptance and Use of ICT by Nigerian University Academicians. *International Journal of Information Communication Technologies and Human Development*, 3(4), 1–16. doi:10.4018/jicthd.2011100101
- Rexha, N., John, R. P., & Shang, A. S. (2003). The impact of the relational plan on adoption of electronic banking. *Journal of Services Marketing*, 17(1), 53–67. doi:10.1108/08876040310461273

Suki, M. N., & Ramayah, T. (2010). User Acceptance of the E-Government Services in Malaysia: Structural Equation Modelling Approach. *Interdisciplinary Journal of Information, Knowledge, and Management*, 5, 395–413.

Taiwo, A. A., & Downe, A. G. (2012). The theory of user acceptance and use of Technology: A Meta-Analysis. *Journal of Theoretical and Applied Information Technology*, 49, 48–58.

Taiwo, A. A., Downe, A. G., & Loke, S.-P. (2014). Behavioural Intention Towards eGovernment In Malaysia: A Structural Equation Modelling Approach. *International Journal of Electronic Government Research*, 10(2), 8–21. doi:10.4018/ijegr.2014040102

Taiwo, A. A., Downe, A. G., & Mahmood, A. K. (2012) User Acceptance of eGovernment: Integrating risk and trust dimensions with UTAUT model. In *Proceedings at the International Conference on Computer & Information Science*. Kuala Lumpur, Malaysia: IEEE.

Teo, T. S. H., Srivastava, S. C., & Jiang, L. (2009). Trust and Electronic Government Success: An empirical study. *Journal of Management Information Systems*, 25(3), 99–131. doi:10.2753/MIS0742-1222250303

Venkatesh, V., & Morris, M. G. (2000). Why don't men ever stop to ask for directions? Gender, social influence, and their role in technology acceptance and usage behavior. *Management Information Systems Quarterly*, 24(1), 115–139. doi:10.2307/3250981

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *Management Information Systems Quarterly*, 27, 425–478.

Venkatesh, V., Skyes, T. A., & Zhang, X. (2011). Just what the doctors ordered': A revised UTAUT for EMR system adoption and use by Doctors. In *Proceedings of the 44th Hawaii International conference on System Sciences*. IEEE. doi:10.1109/HICSS.2011.1

Wang, Y.-S., & Shih, Y.-W. (2009). Why do people use information Kiosks? A validation of the Unified theory of Acceptance and use of Technology. *Government Information Quarterly*, 26(1), 158–165. doi:10.1016/j.giq.2008.07.001

Welch, E. W., Hinnant, C. C., & Moon, M. J. (2005). Linking Citizen Satisfaction with E-Government and Trust in Government. *Journal of Public Administration: Research and Theory*, 15(3), 371–391. doi:10.1093/jopart/mui021

ADDITIONAL READING

Bouaziz, F. (2008). Public Administration Presence on the Web: A Cultural Explanation" The Electronic. *Journal of E-Government*, 6(1), 11–22.

Bwalya, K. J., & Healy, M. (2010). Harnessing e-Government Adoption in the SADC Region: A Conceptual Underpinning. *Electronic. Journal of E-Government*, 8(1), 23–32.

Dode, R. O. (2007). Prospects of e-government implementation in Nigeria. In *Proceedings of the 1st international conference on Theory and practice of electronic governance (ICEGOV '07)*, New York, NY, ACM. doi:10.1145/1328057.1328137

Heeks, R. (2002). e-Government in Africa: Promise and practice. *Journal of Information Polity*, 7(2), 97–114.

Relyea, H. C. (2002). E-gov: Introduction and Overview. *Government Information Quarterly*, 19(1), 9–35. doi:10.1016/S0740-624X(01)00096-X

KEY TERMS AND DEFINITIONS

Behavioural Intention (BI): The degree to which people are willing to perform a specific behaviour towards the use of technology.

Disposition to Trust (DTT): An individual's tendency to be willing to depend on others.

Effort Expectancy (EE): The degree of ease associated with use of the computer systems.

Facilitating Condition (FC): The degree to which an individual believes that organizational and technical infrastructures are available to support the use of the system.

Institution Based Trust (IBT): The level to which an individual believes structural conditions are good enough to support his or her success.

Performance Expectancy (PE): The degree to which an individual believes that using a computerised system will assist him or her in increasing his or her job performance.

Social Influence (SI): The degree to which an individual sees that important others believe he should use the new computer system or technology.

Structural Equation Modeling (SEM): Class of methodologies that seeks to represent hypothesis about summary statistics derived from empirical measurements in terms of smaller numbers of “structural” parameters defined by a hypothesized underlying model.

Trusting Belief (TB): An individual’s confidence that the trade partner shall fulfill his or her transactional obligations as expected by the member.

UTAUT: Unified theory of acceptance and use of technology.

Young People, Civic Participation, and the Internet



Fadi Hirzalla

Erasmus University Rotterdam, The Netherlands

Shakuntala Banaji

LSE, UK

INTRODUCTION

This chapter reviews the emerging body of literature on the online civic participation of young people. Following swiftly on the advent of the internet in the 1990s, this literature emerged and expanded rapidly in the context of the ways in which Western policy makers seem to have construed young people as generally not properly motivated or skilled to participate in civic activities. Many scientific studies have also portrayed young people as inactive compared to older age cohorts or previous generations of young people. This depiction of apathy and disaffection has been particularly emphasized in regard to more traditional civic or political activities, such as voting and membership of political parties (Banaji & Buckingham, 2013).

It is against this backdrop that the widespread ‘mediatization’ of contemporary young people’s lives has prompted media and political science scholars to question whether and how new media technologies, particularly the internet, might reinvigorate civic participation amongst young people. The attribution of civic potential to internet activities constitutes the newest episode in old scholarly discussions about what civic participation envelops. The next section outlines the contours of these discussions, as well as the development of the literature on young people’s civic participation online. The section thereafter

discusses the (sometimes conflicting) forms of knowledge and insights that have been produced within four strands of empirical research on young people’s online civic participation. The last two sections consist of a discussion of directions for further research and a conclusion.

BACKGROUND

The online activities to which civic potential has been attributed differ in nature and in form. Aside from visiting informational sites, recent forms of online participation have been described as more ‘interactive’ and include activity on online petition platforms, forums, blogs, photo sharing sites such as Instagram, video sharing platforms such as YouTube, micro-blogs such as Twitter, and social network sites such as Facebook. Studies that have investigated people’s participation in such activities have been conducted, as outlined below, in the context of wider discussions about the activities that should be considered as civic in the first place. Therewith impacting on any conclusion drawn about the manner and extent of young people’s civic participation, these discussions have formed a major thrust behind the literature on online civic activity of youth. The second part of this section discusses the historical development of this literature about young people’s online civic participation.

What is the Civic?

People's participation in activities within the formal political arena has received much attention within political science research (Van Deth, 2011). Voting and activism for a political party are among the most prominent of these activities. Studies that have focused on participation in such activities have generally concluded that there is some sort of democratic 'deficit' or 'crisis' among young people in particular. Such research has, for instance, concluded that contemporary young people, compared to older age cohorts or previous generations of youth, vote less and are less knowledgeable about formal political processes (Bakker & De Vreese, 2011).

However, an increasing number of scholars have objected to the notion that non-participation in these traditional activities signifies a problem *per se*. These scholars have asserted that, for young people, such activities are often not appealing in the context of their circumstances and everyday life. Some have maintained, for instance, that young people are not attracted to politicians' communication style and that they lack the feeling that participation in traditional political activities is effective (Banaji & Buckingham, 2013a).

For these reasons, young people are presumed, by some, to have embraced various forms of issue-based participation outside the formal political arena. Some of these activities are considered as 'traditional', such as participation in demonstrations and signing petitions. Other activities are considered as more 'individualistic' or 'creative', and mainly include activities related to consumption, such as buying organic, eco-friendly and fair trade products, or the activities on the internet mentioned above. While there are evident differences between these activities, most of them have in common that they are, or can be, conducted outside formal organizational structures. These activities, in this sense, are not only 'extra-parliamentary', but also 'extra-institutional' (Hirzalla & Van Zoonen, 2010).

Online Civic Participation

The differences in the activities on which scholars have focused signify that the very meaning of civic participation has changed and been contested historically. The study of internet-based civic participation has also been through different stages. Shortly following the advent of the internet, studies began theorizing the medium's potential impact on society and democracy. These studies often foresaw extreme interpersonal and socio-political changes, either utopian or dystopian by nature. On the one hand, studies depicted the individuality of internet applications as a sure recipe for socio-political tragedy, with high levels of internet use expected to result in, among other things, diminishing levels of collective civic action, face-to-face interactivity, civil debate, or the scrutiny of political authority (e.g., Street, 1992). On the other hand, there were studies that predicted that citizens will effectively use the internet to contest political or economic authority or benefit from direct democracy (e.g., Negroponte, 1995).

Early studies continued to revolve around theoretical speculation on these issues, but were also increasingly prescriptive in nature, addressing how websites should be designed by NGOs or governmental agencies if they are to successfully improve people's attention for civic issues. Recommendations for 'best practice' in this respect concern, among other things, the organization of the personnel that produce websites and the preferred modes of communications online. A list with such recommendations comes from Stephen Coleman (2008), who prescribes, among other things, total freedom of expression; dialogic links with those in authority; and clear agreements about what involvement will achieve.

Studies that were conducted in the last decade or so, however, were increasingly based on empirical research, inquiring into whether, how and why the internet might promote civic participation. The next section provides a concise overview of the key knowledge and insights that have been

produced over the years by the four main strands of empirical research.

FOUR STRANDS OF EMPIRICAL RESEARCH

Interview-Based Studies of the Production of Civic Internet Applications

Relatively few qualitative, interview-based case studies have focused on the production of internet applications with an apparent civic potential. The overall goal of these studies has generally been to explore how the backgrounds, circumstances and experiences of website producers relate to their website producing activities, reflecting on how the communicative effectiveness of websites is affected by potential discrepancies between the internet-mediated messages that website producers intend to communicate (the meanings ‘encoded’ in internet content) and how internet users interpret these messages (the ‘decoded’ meanings); how the production of content online depends on producers’ resources in terms of, among other things, finances and personnel; and how the functionalities, subject matters, interface design and structure, and mode of address of websites are intertwined with the worldviews and institutional goals of the web producers (e.g., Banaji & Buckingham, 2013b; Branigan & Mitsis, 2014).

Several studies have also begun to explore questions about networking and political connectivity within cyberspace from the perspective of activist producers. Kevin Gillan’s (2009) study of the exploitation of online civic networks by anti-war activists, for instance, suggests clear parallels between the general aims and forms of these activist groups and the designs and purposes of their online campaign materials and strategies.

Text-Based Studies of Civic Content Online



A second research branch is composed of qualitative case studies of civic content online. These studies have generally attempted to illustrate how internet applications *do* have a civic potential. While not based on analyses of *who* uses web applications, a portion of this research does focus on youth sites in particular. The best-known example of this type of study is the work of Kathryn Montgomery and her colleagues (2004) in the US civic sphere online. Based on their exploration of 300 sites, they found that the American-based, youth-oriented internet offers an abundance of informative, communicative and participatory civic potential.

Other studies focused on sites that are not aimed at young people, but produced general claims about the civicness of the internet. Especially the self-representations and discussions on forums and e-democracy projects, it is maintained, resemble a ‘virtual sphere’, a concept derived from Jürgen Habermas’ (1962/1989) thesis on the ‘public sphere’. In this sphere, discussants exchange critical and reasoned arguments about socio-political matters. Evidence suggesting the existence of a complex, stratified online public sphere where debate of such matters can take place is not merely explored within websites that explicitly aim to engage citizens with civic issues. It is also claimed that such debate is conducted on the sites of entertainment programs, such as the British reality TV series *Wife Swap* (Graham, 2012). Furthermore, public spheres have also been found in authoritarian contexts, where open and critical debate is less likely to happen in traditional media (Rauchfleisch & Schäfer, 2015).

However, aspects of these claims have been contested. Some questions raised relate to the accessibility of this online public sphere to all

citizens regardless of location, class, age and other demographic features (e.g., Banaji & Buckingham, 2009). Other studies have queried the civicness of the discussions taking place, even drawing attention to propagandist or inflammatory networks of the far right in the online public sphere (e.g., Harindranath & Khorana, 2013). Scholars have also warned that undemocratic powers use social networks sites as surveillance tools and propaganda channels, and that the role of social networks in the contestation of such powers can be easily overestimated (e.g., Morozov, 2011).

Survey-Based Research on Civic Internet Use

Survey-based studies, which form a third strand of research, have generally yielded less optimistic conclusions about the internet's civic potential. These studies have mainly aimed to investigate patterns in the circumstances, attitudes and skills of people that use the internet for civic purposes (e.g., Oyedemi, 2015) and their ability to understand civic content online (e.g., Bowyer & Kahne, 2015); the material or attitudinal factors that influence or are influenced by civic internet use (e.g., Lee, Shah & McLeod, 2013); and, to a lesser extent, how civic internet use is related to civic participation offline or duration and forms of traditional media use (e.g., Vissers & Stolle, 2014).

While various methodological and theoretical issues within these themes remain contested, some findings are gaining consensus. For instance, the internet seems to be still primarily used for recreation and social communicative activities, although civic and political purposes are gaining ground (Banaji & Cammaerts, 2014). Other studies have indicated that, while some 'digital divides' in internet access *intensity* may be gradually diminishing in developed countries, they are not in developing countries, and there are also still discrepancies in internet use *forms*, with groups of (young) people equipped with relatively low levels of material or attitudinal resources (such as media literacy) using the internet less for civic action

than groups that possess more of such resources (Oyedemi, 2015).

Further, youth and others who are civically active online are often also civically active offline. Survey-based studies, therefore, have often acknowledged that the internet has a potential to promote civic participation, but that this potential has yet to be fully realized (Hirzalla & Van Zoonen, 2010). This is not to say, however, that the internet has no purpose at all on top of civic activity offline. People who would normally not participate offline may use the internet in civic ways. Hence, while for most people the internet seems to play a reinforcing role (i.e., supporting existing patterns in offline participation), it can also mobilize new people into digital forms of civic participation (Vissers & Stolle, 2014).

Interview-Based Research on Civic Internet Use

A fourth strand of research has also focused on the young users of internet applications, but has paid more attention to the role of context and culture, and has relied on qualitative methods, mainly in-depth interviews and focus groups, sometimes in combination with ethnographic, participant observation methods. Some of these studies have attempted to clarify how youth are civically active and engaged, yet in their own ways and on their own terms, which are allegedly misunderstood in survey-based research (e.g., Bennett, 2007). Other studies have been oriented at assessing how young people may consider internet technologies as a valuable resource for civic participation (e.g., Thorson, 2014; Vromen, Xenos & Loader, 2015). A third group of studies in this category attempted to yield detailed insights into the aforementioned 'digital divide' problem, mainly in relation to young people's ability to understand internet technologies, and social factors that may influence young people's access of the internet (e.g., Bobkowski & Smith, 2013; Linne, 2014). Based on in-depth interviews and participant observation among Latino immigrant families

in the American context, Lisa Tripp (2011), for instance, found that young people can be clearly restricted in their opportunities to go online due to the attitudes of their parents as well as policies of their schools. Lastly, some studies have looked into how young people experience political marketing which NGOs or governments conduct to engage and mobilize them. Loader, Vromen and Xenos (2015), for example, found on the basis of a qualitative assessment of discussions online by young people that:

the effectiveness of social media use in mobilizing support is likely to depend upon the receptiveness of politicians and political celebrities to the expectations and performative demands of young networking citizens. (...) The overwhelming majority of our young citizens were cautiously open to the prospect of politicians and political discussion being an aspect of their social networking, provided that politicians used social media appropriately. The clear implication being that social media could indeed be an effective channel for politicians to connect with young citizens, but that in order to do so they would have to develop more authentic digital personas (pp. 15-16).

The Roots of Incongruence

In sum, the literature about online civic participation among young people is widely varied, with studies focusing on different research questions, and assessing these questions through different methods. In this light, it should not be surprising that extant research provides different and, in some respects, conflicting cues about the internet's civic potential for young people. Here, it is possible to distinguish between at least two interrelated analytic axes along which the knowledge and insights produced in extant research tend to conflict.

First, across the four strands of research, there are differences between studies that have focused on the civic *potential* of the internet on the one hand, and studies that have focused on the actual *realization* of the internet's civic potential on the

other. Compared to the former group of studies, the latter group is generally more pessimistic about the role of the internet in young people's civic life.

Second, within extant research, there are often differences in the conclusions between survey-based studies that focus on *patterns* in civic internet use on the one hand, and text-based studies that focus on the characteristics of civic *manifestations* online (i.e., information, self-representations and discussions) on the other. Compared to pattern-oriented research, studies that focus on online manifestations have concluded far more frequently that the internet's civic potential is obvious and valuable for internet users.

It should be noted, however, that studies on different sides of either one of these axes do not yield *empirical* conclusions that are necessarily incongruous. Rather, scholars' *theoretical* inferences from empirical evidence tend to vary along these axes, with potential- and manifestation-focused studies most often generating optimistic inferences, and realization- and pattern-focused studies generally yielding more pessimistic inferences.

Underlying most of the differences between the research strands and transcending discrepancies in research focus are more fundamental issues concerning scholars' epistemological approaches. On the one hand, survey-based studies have generally taken a more 'positivistic' approach by, among other things, generalizing the role of perception, meaning and context. Based on deductively constructed and quantifiable units of analysis, such as the extent to which people report engaging in civic participation, such studies construct generalizing claims through statistical techniques. On the other hand, qualitative studies, conducted within a 'post-positivistic', 'interpretative' tradition, presume that such claims cannot reveal the whole story since, for instance, closed-question surveys may disaccord with the intricate ways in which people make sense of the world. Such studies, therefore, turn to building theory on the basis of inductive explorations of the perspectives and experiences of those involved in producing and using civic spaces online.

FUTURE RESEARCH DIRECTIONS

This section discusses a number of directions for further research. These suggestions call primarily for more pressing context-specific and people-centered research approaches, which may stimulate scholars to further review some of the incongruent or even conflicting forms of knowledge and insights produced over the years, and yield novel and more rigorous understandings of the role of the internet in young people's civic lives.

First of all, while qualitative research has partly focused on online civic participation, survey-based studies have distinguished rather persistently between online and offline civic participation. New insights might be gained, however, when people's perspectives on participation are taken as an analytic point of departure, rather than their activities. Such an approach might underpin explorations of how activities generally, regardless of their online or offline form, can be straightforwardly embedded in people's motivations and interests.

Taking young people's perspectives more centrally into account might also motivate scholars to move beyond the 'democratic' narratives on which their empirical analyses and theoretical inferences may be based. Most notably, many studies in the field have assumed that young people's participation is motivated by civic attitudes (e.g., a belief that changing or contributing to society is possible) and/or civic goals (e.g., a desire to change or contribute to society). However, social and personal factors, such as a desire to make new friends or earn a living, may also play a role but have hardly been explored by extant research.

A third research innovation pertains to the focus of survey-based studies on *general* patterns in online civic participation. People might use particular internet applications for civic purposes at particular moments only, when events occur that are experienced as 'special' or highly relevant for their personal lives. Investigating these instances of temporary civic internet is bound to alleviate the tensions between studies that focus on general patterns in people's civic participation online,

and studies that focused on online manifestations of such behaviors. More specifically, evaluating who participates in the online civic discussions and self-representations that are celebrated by text-based studies will reveal whether these online manifestations are generated by a wide public of any age group (or rather only by those people who are civically active and engaged on a structural level), and, in so doing, whether the rather pessimistic conclusions about general patterns in civic internet use can be nuanced with more sanguine observations of civic internet use in particular contexts.

The relevance of context-specific and people-centered approaches becomes furthermore evident from the few studies conducted to date that contrasted aspects of the foregoing survey and textual data with interview-based research. Although studies have been expanded significantly and methods refined over the past decade to include such triangulation of data, there remains much one-dimensional scholarship. Triangulation, however, can reveal the complex interactions of technological, social, historical and economic factors in shaping the manner and extent to which internet applications are used for civic purposes.

Triangulation will also be highly relevant in the newest chapter of new media research, which revolves around 'Big Data'. In the context of media and communication science, Big Data refers to relatively large amounts of internet content that can be collected, analyzed and reported automatically or semi-automatically. The innovative software programs that accomplish these things provide, among other things, bigger samples (if not populations) of data than generally can be collected manually, and they enable large-sample content analyses that would be too cumbersome and time consuming to do manually. With these bigger samples and analyses on larger scale, it is possible to produce fuller and more detailed insights into the textual, sentimental and networked nature of civic content online. These text-based insights, however, reveal little (if any) information about the producers of civic content. Furthermore, the

results yielded by automated content analyses are generally confined and quantified, rather than thick and specific. Therefore, to interpret the results of Big Data analysis properly, it will be important to triangulate such analysis with further qualitative interview- and text-based research (Hirzalla, Müller & Van Zoonen, 2013).

CONCLUSION

This article discussed the socio-political and theoretical background of the literature on the question as to whether and how the internet may promote young people's internet-based civic participation; which main strands of empirical research on that issue have emerged over the years; and in what ways and for what reasons different forms of knowledge that have been produced by extant research seem to be intrinsically incompatible. The suggested directions for future research should not only alleviate those tensions between the divergent forms of knowledge, but also alter the rather standardized narratives in the literature about youth's online civic participation, which, as a partial artifact of the differences in scholars' research focus and methods, have all too often been constructed either as an elitist exercise that normalizes offline business as usual, or as a panacea for old socio-political inequalities. More nuanced narratives may lie somewhere in the middle, and are yet to be fully developed by this young but lively field of literature.

REFERENCES

- Bakker, T. P., & De Vreese, C. H. (2011). Good news for the future? Young people, internet use, and political participation. *Communication Research*, 38(4), 451–470. doi:10.1177/0093650210381738
- Banaji, S., & Buckingham, D. (2009). The civic sell: Young people, the internet and ethical consumption. *Information Communication and Society*, 12(6), 1–27.
- Banaji, S., & Buckingham, D. (2013a). *The civic web: Young people, the internet and civic participation*. MIT Press.
- Banaji, S., & Buckingham, D. (2013b). Creating the civic web: Exploring the perspectives of web producers in Europe and Turkey. In T. Olsson (Ed.), *Producing the internet: Critical perspectives of social media* (pp. 221–238). Gothenburg: Nordicom.
- Banaji, S., & Cammaerts, B. (2014). Citizens of Nowhere Land: Youth and news consumption in Europe. *Journalism Studies*. doi: 2014.89034010.1080/1461670X
- Bennett, W. L. (Ed.). (2007). *Civic life online: Learning how digital media can engage youth*. Cambridge, MA: MIT Press.
- Bobkowski, P., & Smith, J. (2013). Social media divide: Characteristics of emerging adults who do not use social network sites. *Media Culture & Society*, 35(6), 771–781. doi:10.1177/0163443713491517
- Bowyer, B. T., Kahne, J. E., & Middaugh, E. (2015). Youth comprehension of political messages in YouTube videos. *New Media & Society*. doi:10.1177/1461444815611593
- Branigan, E., & Mitsis, A. (2014). Reach for generation Y: Using celebrity endorsement to communicate about nonprofit causes with young people in Australia. *International Journal of Nonprofit and Voluntary Sector Marketing*, 19(4), 314–321. doi:10.1002/nvsm.1503
- Coleman, S. (2008). Doing it for themselves: Management versus autonomy in youth e-citizenship. In L. Bennett (Ed.), *Civic life online: Learning how digital media can engage youth* (pp. 189–206). Cambridge, MA: MIT Press.
- Gillan, K. (2009). The UK anti-war movement online. *Information Communication and Society*, 12(1), 25–43. doi:10.1080/13691180802158532

- Graham, T. (2012). Beyond political communicative spaces: Talking politics on the Wife Swap discussion forum. *Journal of Information Technology & Politics*, 9(1), 31–45. doi:10.1080/19331681.2012.635961
- Habermas, J. (1962). *The structural transformation of the public sphere*. Cambridge, MA: MIT Press.
- Harindranath, R., & Khorana, S. (2013). Civil society movements and the ‘Twittering classes’ in the postcolony: An Indian case study. *Journal of Southeast Asian Studies*, 37(1), 60–71.
- Hirzalla, F., Müller, F., & Van Zoonen, L. (2013). How funny can Islam controversies be? Comedians defending their faiths on YouTube. *Television & New Media*, 14(1), 46–61. doi:10.1177/1527476412453948
- Hirzalla, F., & Van Zoonen, L. (2010). Beyond the online/offline divide: How youth’s online and offline civic activities converge. *Social Science Computer Review*, 10385538. doi:10.1177/08944393
- Lee, N.-J., Shah, D., & McLeod, J. M. (2013). Processes of political socialization: A communication mediation approach to youth civic engagement. *Communication Research*, 40(5), 669–697. doi:10.1177/0093650212436712
- Linne, J. (2014). Adolescents from low-income sectors: The challenge of studying in a time of digital environments. *International Journal of Adolescence and Youth*, 19(4), 434–443. doi:10.1080/02673843.2014.942792 PMID:25364088
- Loader, B., Vromen, A., & Xenos, M. (2015). Performing for the young networked citizens: Celebrity politics, social networking, and the political engagement of young people. *Media Culture & Society*. doi:10.1177/0163443715608261
- Montgomery, K., Gottlieb-Robles, B., & Larson, G. O. (2004). *Youth as e-citizens: Engaging the digital generation*. Washington, DC: Center for Social Media.
- Morozov, E. (2011). *The net delusion: How not to liberate the world*. London: Allen Lane.
- Negroponte, N. (1995). *Being digital*. London: Coronet.
- Oyedemi, T. (2015). Participation, citizenship and internet use among South African Youth. *Telematics and Informatics*, 32(1), 11–22. doi:10.1016/j.tele.2014.08.002
- Rauchfleisch, A., & Schäfer, M. S. (2015). Multiple public spheres of Weibo: A typology of forms and potentials of online public spheres in China. *Information Communication and Society*, 18(2), 139–155. doi:10.1080/1369118X.2014.940364
- Street, J. (1992). *Politics and technology*. New York, NY: Guildford Press. doi:10.1007/978-1-349-22274-2
- Thorson, K. (2014). Facing an uncertain reception: Young citizens and political interaction on Facebook. *Information Communication and Society*, 17(2), 203–216. doi:10.1080/1369118X.2013.862563
- Tripp, L. (2011). The computer is not for you to be looking around, it is for schoolwork: Challenges for digital inclusion as Latino immigrant families negotiate childrens access to the internet. *New Media & Society*, 13(4), 552–567. doi:10.1177/1461444810375293
- Van Deth, J. (2011). *Studying political participation: Towards a theory of everything?* Retrieved August 22, 2011, from <http://www.essex.ac.uk/ecpr/events/jointsessions/paperarchive/grenoble/ws3/deth.pdf>
- Vissers, S., & Stolle, D. (2014). The internet and new modes of political participation: Online versus offline participation. *Information Communication and Society*, 17(8), 937–955. doi:10.1080/1369118X.2013.867356

Vromen, A., Xenos, M. A., & Loader, B. (2015). Young people, social media and connective action: From organisational maintenance to everyday political talk. *Journal of Youth Studies*, 18(1), 80–100. doi:10.1080/13676261.2014.933198

ADDITIONAL READING

Banaji, S. (2011). Framing young citizens: Explicit invitation and implicit exclusion on youth civic websites. *Language and Intercultural Communication*, 11(2), 126–141. doi:10.1080/14708477.2011.556738

Banaji, S., & Buckingham, D. (2010). Young people, the internet, and civic participation: An overview of key findings from the CivicWeb project. *International Journal of Learning and Media*, 2(1), 15–24. doi:10.1162/ijlm_a_00038

Bennett, W. L., Wells, C., & Freelon, D. (2011). Communicating civic engagement: Contrasting models of citizenship in the youth web sphere. *Journal of Communication*, 61(5), 835–856. doi:10.1111/j.1460-2466.2011.01588.x

Cammaerts, B., Bruter, M., Banaji, S., Harrison, S., & Anstead, N. (2014). The myth of youth apathy: Young Europeans critical attitudes toward democratic life. *The American Behavioral Scientist*, 58(5), 645–664. doi:10.1177/0002764213515992

Chadwick, A., & Howard, P. (2009). *The handbook of internet politics*. London: Routledge.

Coleman, S., & Freelon, D. (Eds.). (2015). *Handbook of digital politics*. Abingdon: Edward Elgar Publishing. doi:10.4337/9781782548768

Dahlgren, P. (2005). The internet, public spheres, and political communication: Dispersion and deliberation. *Political Communication*, 22(2), 147–162. doi:10.1080/10584600590933160

Delli Carpini, M. X. (2000). Gen.com: Youth, civic engagement, and the new information environment. *Political Communication*, 17(4), 341–349. doi:10.1080/10584600050178942

Freelon, D., Wells, C., & Bennett, W. L. (2013). Participation in the youth civic web: Assessing user activity levels in web sites presenting two civic styles. *Journal of Information Technology & Politics*, 10(3), 293–309. doi:10.1080/19331681.2013.792309

Gerodimos, R. (2011). Online youth civic attitudes and the limits of civic consumerism. *Information Communication and Society*. doi:10.1080/1369118X.2011.572983

Hargittai, E. (2010). Digital natives? Variation in internet skills and uses among members of the Net Generation. *Sociological Inquiry*, 80(1), 92–113. doi:10.1111/j.1475-682X.2009.00317.x

Hooghe, M. (2014). Defining political participation: How to pinpoint an elusive target? *Acta Politica*, 49, 338–341.

Livingstone, S. (2009). *Children and the internet: Great expectations, challenging realities*. Cambridge: Polity.

Loader, B. D. (Ed.). (2007). *Young citizens in the digital age: Political engagement, young people and new media*. London: Routledge.

Norris, P. (2002). *Democratic phoenix: Political activism worldwide*. New York, NY: Cambridge University Press. doi:10.1017/CBO9780511610073

Olsson, T., & Dahlgren, P. (Eds.). (2010). *Young people, ICTs and democracy: Theories, policies, identities and websites*. Göteborg: Nordicom.

Papacharissi, Z. (2010). *A private sphere: Democracy in a digital age*. Cambridge: Polity Press.

Putnam, R. (2000). *Bowling Alone: The collapse and revival of American Community*. New York, NY: Simon & Schuster. doi:10.1145/358916.361990

Quiring, O., & Schweiger, W. (2008). Interactivity: A review of the concept and a framework for analysis. *Communications*, 33(2), 147–167. doi:10.1515/COMMUN.2008.009

Schudson, M. (1998). *The good citizen: A history of American civic life*. New York, NY: Free Press.

Theocharis, T. (2015). The conceptualization of digitally networked participation. *Social Media + Society*, doi:10.1177/2056305115610140

Thomas, M. (2011). *Deconstructing digital natives: Young people, technology and the new literacies*. New York, NY: Routledge.

Van Dijk, J. (2005). *The deepening divide in equality in the information society*. London: Sage Publications.

Vromen, A. (2011). Constructing Australian youth online. *Information Communication and Society*, 14(7), 959–980. doi:10.1080/1369118X.2010.549236

Warschauer, M. (2004). *Technology and social inclusion: Rethinking the digital divide*. London: MIT Press.

KEY TERMS AND DEFINITIONS

Citizenship: Status referring to citizens' formal duties and rights, as well as informal norms regarding their behavioral and attitudinal commitments to social and political life in a democracy.

Civic Engagement: Attitudinal commitment of citizens to norms and processes deemed political or social by nature.

Civic Participation: Participation in activities deemed political or social by nature.

Democracy: A society formally and culturally organized to enable eligible citizens to participate equally in decision-making processes.

Political Engagement: Attitudinal commitment of citizens to norms and processes that are deemed parliamentary political by nature.

Political Participation: Participation in activities deemed parliamentary political by nature.

Young People: People between 12 and 30 years old.

Category H

Healthcare Administration

Electronic Health Record (EHR) Diffusion and an Examination of Physician Resistance

Kristen MacIver

Northern Michigan University, USA

Madison N. Ngafeeson

Northern Michigan University, USA

INTRODUCTION

Since the enactment of the United States Health Information Technology for Economic and Clinical Health (HITECH) Act in 2009, there has been substantial progress in the diffusion of Electronic Health Record (EHR) systems in medical clinics throughout the United States. Many physicians, however, continue to resist adopting EHR technology despite several recognized long-term benefits and available government sponsored financial incentives. The objective of this article is analyze physician resistance to EHR technology adoption. Previous researches indicate that physicians have been reluctant to adopt EHR systems due to high implementation and maintenance cost besides the uncertainties associated with EHR's return on investment. Apart from financial cost, physicians also seem to encounter challenges in adopting EHRs due to lack of technical skills, time, and technical support. Workflow disruption, loss of autonomy, confidentiality issues, and interoperability are major challenges hindering the adoption of EHRs (Ajami & Arab-Chadegani, 2013; Lorenzi, Kouroubali, Detmer, & Bloomrosen, 2009). Moreover, an annual report prepared by the United States Department of Health and Human Services (2014) shows that, loss of productivity, lack of adequate training and incompatibility of EHRs with practice needs are the top reasons why physicians resist the adoption of EHR systems. With the awareness of the major physician barriers

to the adoption of EHR technology, stakeholders and policy makers can address barriers and pursue actions to mitigate or reduce physician resistance to achieve nationwide diffusion targets and pursue initiatives to digitize all patient records.

Although physicians are hesitant to adopt EHR technology, there are several benefits of EHR systems that make the technology worth adopting. Evidence from healthcare research and systematic reviews show that the benefits of EHRs continue to grow and include: enhancing the accurate collection, storage and sharing of medical information of patients with authorized personnel to support informed decision making, prevention of medical errors, availability of data for clinical research and analysis (United States Department of Health and Human Services, 2014). The Institute of Medicine and other stakeholders in the healthcare field suggest that the wide-scale adoption of EHR systems could be pivotal for improving patient safety and health care quality and could reduce the costs of providing ambulatory care (Ajami & Bagheri-Tadi, 2013). It is generally accepted that EHRs could also improve efficiency, portability and research capabilities (Pipersburgh, 2011).

Given the potential benefits of EHRs, physicians could be helped to overcome their barriers to adopting EHR technology. For example, physicians can be more productive or efficient and effective if their routine tasks are automated. With the input of physicians, automated technologies could be produced to handle repeated tasks and

DOI: 10.4018/978-1-5225-2255-3.ch319

daunting processes. Physicians will then be able to spend more time with patients and workflow interruptions will be eliminated as well. This research illuminates the challenges encountered by physicians in adopting EHRs, and suggests how future research could tackle these.

BACKGROUND

An EHR can be defined as a digitally stored record of an individual's healthcare information used for supporting the continuity of care, education, and research (Ajami & Arab-Chadegani, 2013). The Health Information Technology for Clinical Health (HITECH) Act, which was enacted as a part of the American Recovery and Reinvestment Act (ARRA) of 2009, served as a turning point in the evolution of EHR technology in the United States. The HITECH Act was signed into law with the explicit intention of accelerating the adoption of Electronic Health Records (EHRs) by physicians, and since its enactment, the program has distributed billions of dollars to physicians for adopting certified EHRs through the meaningful use (MU) program (Mennemeyer, Menachemi, Rahurkar, & Ford, 2016). One research study noted that through December 2015, the federal government had made \$13 billion worth of EHR incentive payments to physicians and other eligible professionals who

met stage 1 of meaningful use (Cohen, 2016). Overall, the stimulus act included a requirement that grant recipients achieve all meaningful-use requirements, including interoperability, by the end of 2017 (Sandler, 2016).

In acute care hospitals, there has been substantial growth in the adoption of basic and certified EHR systems (Charles, Gabriel, & Furukawa, 2014). In 2014, 3 out of 4 hospitals had adopted at least a basic EHR system which represents an increase of 27% from 2013 (Charles et al., 2014). Figure 1 delineates the growth in the percent of non-federal acute care hospitals with adoption of at least a basic EHR system and possession of a certified EHR system from 2008 to 2013 (Charles et al., 2014).

EHR adoption has also been on the rise for rural practices which has reversed initial trends (Hirsch, 2015). A new study by the Journal of the American Medical Informatics Association (JAMIA) revealed that EHR adoption in rural areas was 56 percent compared to 49 percent in urban areas, and the study concluded that 27 states have adoption rates statistically higher in rural areas than in urban areas (Hirsch, 2015). The USA is not alone in its efforts of EHR implementation and use. Figure 2 shows notable efforts by other organizations in key developing countries.

Despite the progression of the technology and broad agreements on the benefits of EHR

Figure 1. EHR system adoption for non-federal acute care hospitals in the U.S.
Charles et al., 2014.

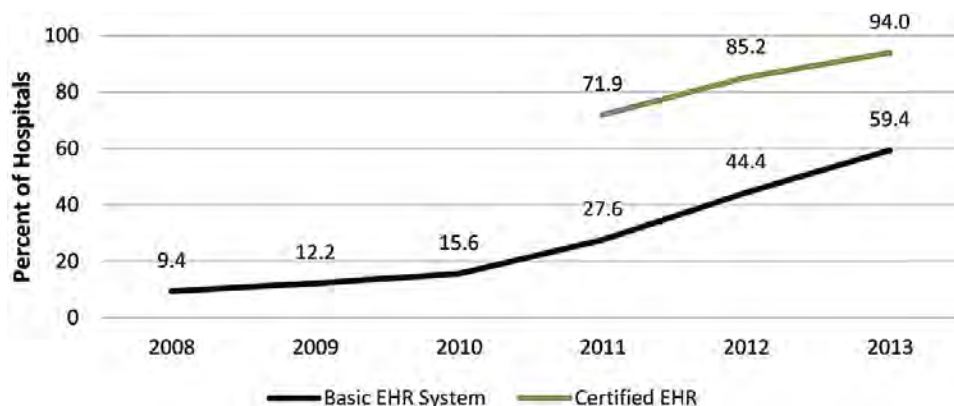
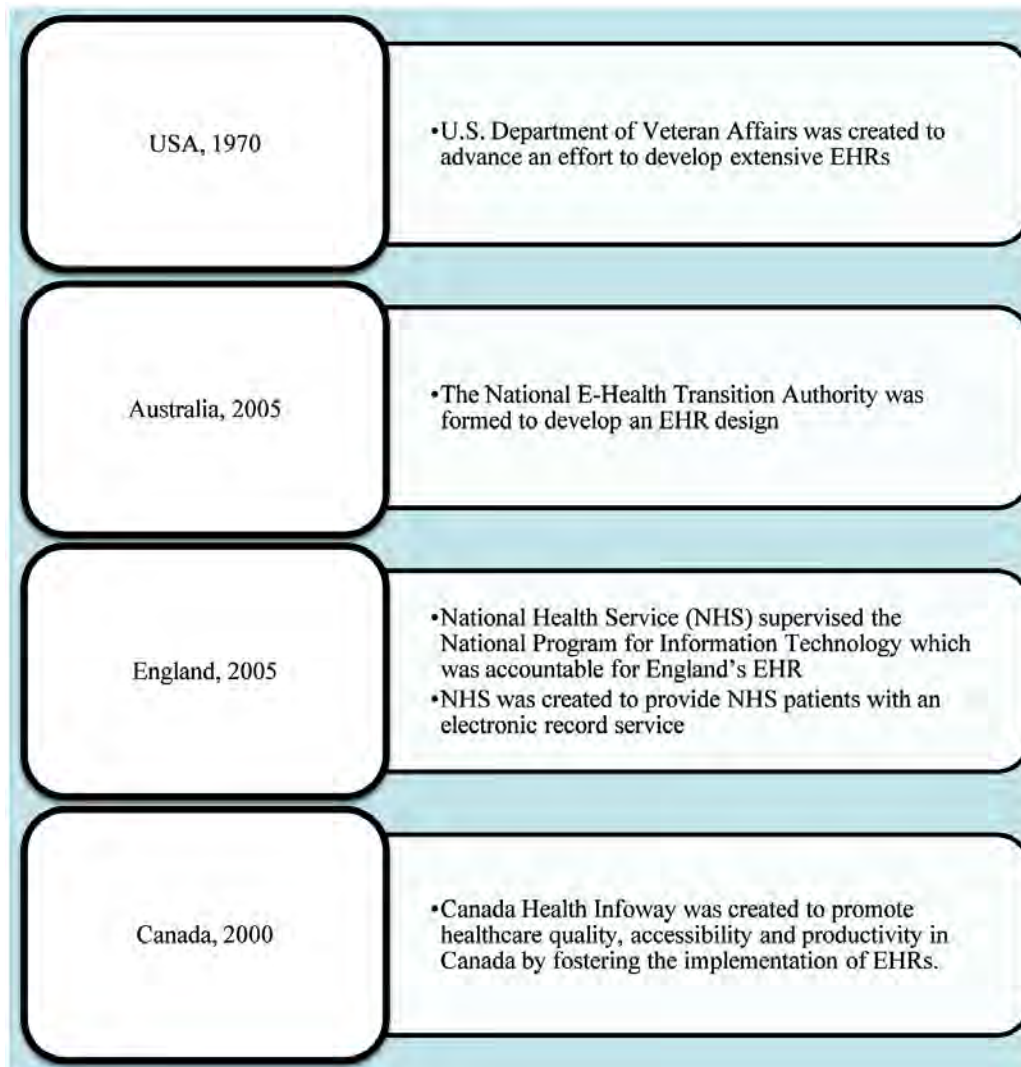


Figure 2. EHR initiation in the USA, Canada, England, and Australia
Aminpour, Sadoughi, & Ahamdi, 2014.



technology, many physicians in the United States remain hesitant to adopt. Research shows that that the diffusion and physician acceptance of EHR technology may be hindered by a variety of intrinsic and extrinsic barriers (Paré et al., 2014). The purpose of this study is to explore the most prominent perceived challenges and barriers that physicians face in adopting EHR technology. Specifically, this study is intended to answer the question, “Why do physicians resist the adoption of EHR technology?”

This study utilizes a systematic review of literature on physicians’ resistance to the adoption of EHRs. The literature on physicians’ resistance and perceived barriers to the adoption of EHRs and EHR technology was completed through a research database search engine. The search employed various combinations of keywords such as electronic health record, electronic medical record, barriers, physicians, resistance, and adoption to search for previous research and literature related to this study. Journal articles used in this study were based on relevance to physician bar-

riers to the adoption of EHR technology although many of the detailed barriers to adoption of EHR technology are applicable and relevant to other user groups. Research databases leveraged in this study include journal articles and digital articles or reports. Approximately thirty-three articles were reviewed for this study and twenty-seven of the articles met or supported the specific criteria of this study which focused on physician barriers to adoption of EHRs. Moreover, approximately 67 percent of all the reviewed journal articles were cross-referenced and analyzed.

PHYSICIAN BARRIERS TO ADOPTION OF EHRs

In most clinical settings, physicians' orders trigger most determinants of healthcare quality and costs (Zaroukian, 2011). Although other users, groups, or factors may influence the adoption of EHR technology in a clinical setting, this study focuses on physician resistance since efforts to achieve health information technology (HIT) transformation will not succeed unless the physicians support the desired improvements in quality and value of new technology (Zaroukian, 2011).

The professional journals reviewed in this study maintained many similarities in categorizing or sub-categorizing physician barriers to adoption of EHR technology. The most common barriers identified in this study are outlined in the following categories: cost restraints, technical constraints and system limitations, time constraints and workflow disruptions, physiological constraints, social influences, and privacy and security concerns. A review of twenty-seven articles from 2004 to 2016, found in twenty-five databases showed that there are six major categories of physician barriers to EHR adoption as follows:

Cost Restraints

The short-term and long-term financial investments for an EHR system remains a significant

barrier to EHR system adoption for practices without large IT budgets (Ajami & Bagheri-Tadi, 2013). Financial issues including adoption and implementation costs, ongoing maintenance costs, loss of revenue associated with temporary loss of productivity, and declines in revenue may deter physicians from adopting an EHR system (Menachemi & Collum, 2011). Given the infancy of EHR systems and its use in the United States, there is not substantial evidence on the long-term return on investment (ROI) of EHR systems. Thus, studies reveal that physicians worry that their practices will face substantial financial risk without a promise of return (Boonstra & Broekhuis, 2010). Additionally, research shows that many financial benefits of an EHR do not accrue to the provider but rather to the third-party payers in the form of errors averted and improved efficiencies, which generally translate to reduced claim payments which can be another concern for physicians (Menachemi & Collum, 2011).

Technical Constraints and System Limitations

The successful implementation of EHR technology requires users and suppliers to have enhanced technological capabilities. Most users, even highly regarded and leading healthcare organizations, consider EHRs to be challenging to use because of the multiplicity of screens, options and navigational aids (Miller & Sim, 2004). Research has concluded that physicians generally have insufficient technical knowledge and skills to deal with EHRs which results in resistance (Boonstra & Broekhuis, 2010). The skills needed to communicate with a patient, analyze medical relevance, complete interventions as well as type notes simultaneously would require a significant level of concentration, typing skills, and familiarity with the application's user interface. These multiple requirements and additional tasks resulting from the use of EHR technology could be seemingly overwhelming to physicians (Ajami & Bagheri-Tadi, 2013). To foster efficiency and ease of use

for physicians, many vendors of EHR technology have worked to improve usability of their systems, but many vendors doubt that even the most significant upgrades to their systems, such as voice recognition, will dramatically simplify EHR usage (Miller & Sim, 2004). For physicians, a lack of technical skills paired with complex usability requirements is likely to create further resistance to their acceptance or adoption of EHR technology.

Given the complexity of EHR systems, vendor technical support and troubleshooting are crucial factors in the successful implementation and continuation of EHR systems. Many physicians who have implemented EHR systems complain of poor service, follow-up, and training from contracted vendors. These complaints have created a negative perception for non-users who are considering the technology (Boonstra & Broekhuis, 2010). Poor vendor reputation is a barrier to adoption as many physicians are concerned that vendors are not qualified to provide the proper service or will go out of business and disappear from the market, leading to a lack of technical support and financial loss (Ajami & Bagheri-Tadi, 2013).

Another major physician barrier to the adoption of EHR technology is the concern about system capabilities and limitations. Many physicians report that they cannot find an EHR system that meets their specific needs or requirements (Boonstra & Broekhuis, 2010). Results of a study indicated that approximately 54 percent of physician participants who did not have access to an EHR system, could not find a system to meet their specific needs (DesRoches et al., 2008). Despite the progression of EHR system technology and a willingness of many EHR vendors to tailor clinic services and functionalities, customization typically comes at a premium and may raise concerns of system stability and longevity. These factors can deepen financial concerns and be viewed as another risk that may add to physician resistance for EHR system adoption.

Another technical barrier to adoption is EHR system reliability. Reliability, in this context, is described as the dependability of the EHR tech-

nology systems (Randeree, 2007). In leveraging EHR systems, reliability is notably important for a system dealing with patient information, and many physicians are concerned about losing access to patient records if computers crash, viruses attack or during events of power failure (Boonstra & Broekhuis, 2010). Research reveals that physicians are also hesitant to adopt EHR technology out of fear of the possibility of record loss due to technical deficiencies which could translate into financial loss and be costly for a clinic (Boonstra & Broekhuis, 2010).

Given the immaturity of EHR technology, EHR systems must be adapted or altered to connect with other devices that complement the EHR system. However, many physicians have reported problems of interconnectivity leading to a negative perception of the transfer capabilities of current EHR technology (Boonstra & Broekhuis, 2010). The majority of EHR systems are not compatible with existing practice systems; consequently, physicians may be hesitant to replace fully functional systems with a fully integrated EHR system (Boonstra & Broekhuis, 2010). Additionally, research shows that there are hundreds of varying types of EHR software programs currently in use, and inconsistent data standards have prevented existing systems from being capable of transferring data (Valdes, Kibbe, Tolleson, Kunik, & Petersen, 2004). Manual data transfer to a new, fully integrated EHR system could lead to a loss in productivity or could increase costs for a clinic if third-party data transfer is required. Physicians may be resistant to adopt EHR technology because several EHR systems lack data transfer capabilities, which is a potential source of inefficiency and high cost.

Time Constraints and Workflow Disruptions

EHR systems are complex and require extensive research, training, and adaptations to workflow to be successful. Physician involvement during system initiation and selection is important

because participation in the assessment and implementation of EHR systems will provide an opportunity for physicians' information needs to be communicated, considered or addressed during the implementation process (Lorenzi, Kouroubali, Detmer, & Bloomrosen, 2009). Additionally, several research studies have found that successful EHR system implementation is usually driven by a core team that includes physician champions. Physician champions are imperative to the successful implementation of a new EHR system as they typically drive change and motivate staff to buy-in to the new system. Despite the need for physician involvement in a core team or in the EHR selection processes, perceived time constraints serve as a significant physician barrier to EHR system adoption. During initiation stages, research uncovers that many physicians are not willing to invest time in EHR system selection and procurement as they believe that they should spend their time and efforts with patients (Paré et al., 2014).

Physicians have also expressed concerns over the potential time constraints that would result from the implementation of an EHR system which also serves as a barrier to EHR system adoption (Boonstra & Broekhuis, 2010; Paré et al., 2014). It is accepted that physicians will need to spend a considerable amount of time, effort, and training learning how to use an EHR system. Unfortunately, many physicians believe that learning an EHR system would slow their workflow and increase their workload besides thinking that they would not have the time needed to effectively learn an EHR system given the demands and pressures of delivering office-based care (Boonstra & Broekhuis, 2010). Although utilizing an EHR system could increase efficiency for physicians in the long term, the complexity of EHR systems paired with a seemingly daunting training timeline creates resistance to EHR adoption for physicians. This resistance occurs especially if physicians do not support or believe that EHR systems will increase their efficiency.

With regard to concerns over post-implementation time loss, many physicians believe that

EHR systems will interfere with the doctor-patient relationship (Boonstra & Broekhuis, 2010). This negative perception might have come from some physicians who have used EHR systems and are reporting that sometimes they stop using the technology during interactions with the patient because hunting for menus and buttons disrupts the clinical encounter (Ajami & Bagheri-Tadi, 2013). Therefore, physicians may be led to believe that they will need to spend more time with each patient due to EHR system usage (Paré et al., 2014). Taking extra time to use an EHR and not being compensated for taking on more work also suggests why physicians may hesitate to adopt EHR system (Ajami & Bagheri-Tadi, 2013).

Physiological Constraints

Personal resistance and physician skepticism towards EHR systems also serves as a barrier to EHR system adoption. Studies have found that more than half, approximately 58 percent, of physicians without an EHR doubt whether or not the technology can really improve patient care or clinical outcomes, and other research studies argue that those physicians who are unwilling to use such a system are skeptical about claims that EHRs can successfully improve the quality of medical practices (Boonstra & Broekhuis, 2010). Physicians' negative perceptions are generally derived from social influences; and to compound the barriers that social influences may create, people are generally afraid of change which can further slow the diffusion of technology (Boonstra & Broekhuis, 2010).

Professional autonomy, which can be described as an individual's control over the environment, processes, procedures, or content of their work, is important in the working practices of physicians (Boonstra & Broekhuis, 2010). In regard to professional autonomy, some physicians are concerned that using an EHR platform may result in a loss of their control of patient information and working processes since data will be shared with and assessed by others (Boonstra & Broekhuis, 2010).

Other research findings reveal that physicians are concerned that policymakers, insurers, and administrators will use EHRs as a proxy mechanism to influence, or dictate how they practice medicine (Ford, Menachemi, Peterson, & Huerta, 2009). Threats to professional autonomy are an important factor in EHR acceptance or resistance because privileges associated with professional autonomy directly link to social value systems, status, and economic outcome. Therefore, physicians are likely to support elements that foster their autonomy and resist those that may threaten their professional autonomy (Walter & Lopez, 2008). A survey of physicians found that 68 percent agreed or strongly agreed that having clinical freedom was essential in practicing medicine and that constraints to clinical freedom should be fought (Walter & Lopez, 2008).

Social Influences

Lack of support from external entities is another barrier to physician adoption of EHR technology. One study noted that one-third to one-half of physicians reported that their decision-making was affected by local and regional organizations that were not active in the EHR debate (Boonstra & Broekhuis, 2010). Internally, many physicians also experience lack of support from other colleagues or their management team which may demotivate them to pursue EHRs for their clinic. Overall, lack of internal and external support may create further physician hesitation to EHR adoption as research suggests that support from either within a physician's practice or from reliable third parties is essential to overcome knowledge barriers and to configure EHR applications in ways that facilitate improved clinical practices (Davidson & Heslinga, 2007).

Privacy and Security Concerns

Many physicians' perceptions of EHRs is that they are not secure, and that patient information may become accessible to unauthorized entities which

can translate into legal problems and be costly for a clinic (Boonstra & Broekhuis, 2010). Even among physicians who use EHRs, most believe that there are more security and confidentiality risks involved with EHRs than with paper records (Boonstra & Broekhuis, 2010). There have been many reports of accidental loss or theft of sensitive clinical data as EHRs can be vulnerable to hackers, viruses, and worms (Fernández-Alemán, Señor, Lozoya, & Toval, 2013).

With EHR systems, it is also difficult to maintain privacy as administrative staff could access information without the patient's explicit consent (Fernández-Alemán et al., 2013). Although patient privacy and confidentiality have been a major topic of discussion through the progression of EHR technology in the healthcare field, one study finds that physicians may harbor more concerns for patient privacy and confidentiality than the patients themselves (Simon et al., 2007). Reports of internal and external data breaches have heightened physicians' concerns over patient privacy and confidentiality and may serve as another potential detriment to their acceptance and adoption of EHR technology.

IMPLICATIONS FOR RESEARCH AND PRACTICE

This review of literature provides a broad range of the perceived physician barriers to adopting an EHR platform. All reviewed research suggests that the diffusion and physician acceptance of EHR technology may be hindered by a variety and/or mixture of intrinsic and extrinsic barriers. The barrier categories identified are not all inclusive, but outline the most common physician barriers to the adoption of EHR technology as discovered through the analysis and systematic review of multiple articles and journals related to physician resistance to adoption of EHRs. With regard to EHR system adoption, the most common barriers encountered by physicians include: cost restraints, technical constraints and system limitations, time

constraints and workflow disruptions, physiological constraints, social influences, and privacy and security concerns.

Widespread adoption of EHR technology, applied correctly, could benefit clinics and improve healthcare in the United States (Kumar & Aldrich, 2010). The detailed physician barriers uncovered in this research, however, could continue to slow the diffusion of the technology and negatively impact the adoption of EHR systems by healthcare organizations and/or doctors who were already reluctant to adopt. As EHR technology progresses, those organizations who choose not to adopt will face a technological lag relative to their competitors which may be unattractive to patients seeking a modern healthcare experience. Apart from not being able to leverage the benefits of an EHR system, physicians who choose not to adopt an EHR platform may face financial penalties, in the form of Medicare payment cuts, for failing to meet federal electronic health record incentive-payment program standards (Conn, 2014). Research shows that there was a 1 percent penalty in Medicare reimbursements for physicians who failed to achieve meaningful use stage 2 in 2014, and that penalties will escalate to 2 percent in 2016 and 3 percent in 2017 (Conn, 2013).

Overall, failure to adopt an EHR platform could also hinder data sharing and data collection that is needed for research and analysis. Specifically, EHRs offer access to longitudinal data that could be used to predict future outcomes or diagnoses, providing opportunities for personalized decision making for a given patient (Wu, Roy, & Stewart, 2010). Physicians who choose not to adopt an EHR platform will not have the ability to efficiently contribute to data sets or benefit from the analysis of the data collected through EHRs. Therefore, the single most important implication of this research is that, the evidence-based information about the major barriers to physician resistance in adopting EHRs will be useful in formulating solutions to overcome these barriers. Equally important, more concrete solutions to the barriers highlighted in

this study may be explored in further research and analysis.

H

CONCLUSION

Despite accepted long-term benefits and available government sponsored financial incentives for the adoption of an EHR system, complete diffusion of EHR technology has not been realized in the United States as the technology continues to meet resistance from physicians. Overall, successful EHR adoption requires extensive research, investment, training, and adaptation to workflow to improve patient health and safety. The synopsis of physician barriers outlined in this study show that stakeholders and groups driving EHR system adoption need to be aware of the issues surrounding resistance to the system. By understanding and detecting physician barriers that are preventing the widespread adoption of EHR technology, stakeholders can work to mitigate resistance.

Although it is accepted that physicians' intrinsic barriers and social influences may be difficult to overcome, positive advancement in EHR technology usability and functionality could foster the creation of a stronger reputation for the technology thus reducing or removing negative perceptions. In consideration of financial constraints, the imbalance of cost versus return will continue to be a major factor and drive physician resistance until further research is provided on the potential ROI of the technology. The adoption of EHRs should be viewed as an ongoing process that will continuously need to evolve through system upgrades, interoperability, and continued training of staff (Mack et al., 2016). Overall, stakeholders must recognize that the "one size fits all" approach to EHR system technology and implementation is impractical and that, to meet targets of EHR system implementation to achieve positive healthcare outcomes, more work, time, education for physicians and evolution in the technology will be required.

REFERENCES

- Ajami, S., & Arab-Chadegani, R. (2013). Barriers to implement Electronic Health Records (EHRs). *Materia Socio-Medica*, 25(3), 213–215. doi:10.5455/msm.2013.25.213-215 PMID:24167440
- Ajami, S., & Bagheri-Tadi, T. (2013). Barriers for Adopting Electronic Health Records (EHRs) by Physicians. *Acta Informatica Medica*, 21(2), 129–134. doi:10.5455/aim.2013.21.129-134 PMID:24058254
- Aminpour, F., Sadoughi, F., & Ahamdi, M. (2014). Utilization of open source electronic health record around the world: A systematic review. *Journal of Research in Medical Sciences : The Official Journal of Isfahan University of Medical Sciences*, 19(1), 57–64. PMID:24672566
- Boonstra, A., & Broekhuis, M. (2010). Barriers to the acceptance of electronic medical records by physicians from systematic review to taxonomy and interventions. *BMC Health Services Research*, 10(1), 231. doi:10.1186/1472-6963-10-231 PMID:20691097
- Charles, D., Gabriel, M., & Furukawa, M. (2014, May). *Adoption of Electronic Health Record Systems among U.S. Non-federal Acute Care Hospitals: 2008-2013*. Retrieved from <https://www.healthit.gov/sites/default/files/oncdatabrief16.pdf>
- Cohen, M. F. (2016). Impact of the HITECH Financial Incentives on EHR Adoption in Small, Physician-Owned Practices. *International Journal of Medical Informatics*, 94, 143–154. doi:10.1016/j.ijmedinf.2016.06.017 PMID:27573322
- Conn, J. (2013, December 23). *Taking the EHR penalty: More doc offices may opt out*. Retrieved June 29, 2016, from http://ezpolson.nmu.edu:5749/ps/i.do?p=HRCA&u=lom_nmichu&iid=GALEIA353990156&v=2.1&it=r&sid=summon&userGroup=lom_nmichu&authCount=1
- Conn, J. (2014, December 22). *Providers decry imposing EHR penalties in 2015 - ProQuest*. Retrieved June 29, 2016, from <http://ezpolson.nmu.edu:6744/docview/1640351429?pq-origsite=summon&http://ezpolson.nmu.edu:5000/login?url=accountid=2745>
- Davidson, E., & Heslinga, D. (2007). Bridging the IT Adoption Gap for Small Physician Practices: An Action Research Study on Electronic Health Records. *Information Systems Management*, 24(1), 15–28. doi:10.1080/10580530601036786
- DesRoches, C. M., Campbell, E. G., Rao, S. R., Donelan, K., Ferris, T. G., Jha, A., & Blumenthal, D. et al. (2008). Electronic Health Records in Ambulatory Care — A National Survey of Physicians. *The New England Journal of Medicine*, 359(1), 50–60. doi:10.1056/NEJMs0802005 PMID:18565855
- Fernández-Alemán, J. L., Señor, I. C., Lozoya, P. Á. O., & Toval, A. (2013). Security and privacy in electronic health records: A systematic literature review. *Journal of Biomedical Informatics*, 46(3), 541–562. doi:10.1016/j.jbi.2012.12.003 PMID:23305810
- Ford, E. W., Menachemi, N., Peterson, L. T., & Huerta, T. R. (2009). Resistance Is Futile: But It Is Slowing the Pace of EHR Adoption Nonetheless. *Journal of the American Medical Informatics Association*, 16(3), 274–281. doi:10.1197/jamia.M3042 PMID:19261931
- Hirsch, M. D. (2015). EHR adoption grows for rural doc practices, reversing trend. In FierceEMR. Questex Media Group LLC. Retrieved from <http://search.proquest.com/docview/1654847342/citation/D9B0CFBDF136414DPQ/1>
- Kumar, S., & Aldrich, K. (2010). Overcoming barriers to electronic medical record (EMR) implementation in the US healthcare system: A comparative study. *Health Informatics Journal*, 16(4), 306–318. doi:10.1177/1460458210380523 PMID:21216809

- Lorenzi, N. M., Kouroubali, A., Detmer, D. E., & Bloomrosen, M. (2009). How to successfully select and implement electronic health records (EHR) in small ambulatory practice settings. *BMC Medical Informatics and Decision Making*, 9(1), 15. doi:10.1186/1472-6947-9-15 PMID:19236705
- Mack, D., Zhang, S., Douglas, M., Sow, C., Strothers, H., & Rust, G. (2016). Disparities in Primary Care EHR Adoption Rates. *Journal of Health Care for the Poor and Underserved*, 27(1), 327–338. doi:10.1353/hpu.2016.0016 PMID:27587942
- Menachemi, N., & Collum, T. (2011). Benefits and drawbacks of electronic health record systems. *Risk Management and Healthcare Policy*, 47–55. <http://doi.org/10.2147/RMHP.S12985>
- Mennemeyer, S. T., Menachemi, N., Rahrurkar, S., & Ford, E. W. (2016). Impact of the HITECH Act on physicians adoption of electronic health records. *Journal of the American Medical Informatics Association*, 23(2), 375–379. doi:10.1093/jamia/ocv103 PMID:26228764
- Miller, R., & Sim, I. (2004). Physicians Use of Electronic Medical Records: Barriers and Solutions. *Health Affairs*, 23(2), 116–126. doi:10.1377/hlthaff.23.2.116 PMID:15046136
- Paré, G., Raymond, L., de Guinea, A. O., Pobanzaou, P., Trudel, M.-C., Marsan, J., & Micheneau, T. (2014). Barriers to organizational adoption of EMR systems in family physician practices: A mixed-methods study in Canada. *International Journal of Medical Informatics*, 83(8), 548–558. doi:10.1016/j.ijmedinf.2014.06.003 PMID:24969270
- Pipersburgh, J. (2011). The Push to Increase the Use of EHR Technology by Hospitals and Physicians in the United States through the HITECH Act and the Medicare Incentive Program. *Journal of Health Care Finance*, 13(3), 60–84. PMID:22372032
- Randeree, E. (2007). Exploring Physician Adoption of EMRs: A Multi-Case Analysis. *Journal of Medical Systems*, 31(6), 489–496. doi:10.1007/s10916-007-9089-5 PMID:18041282
- Sandler, M. (2016, February 29). *Penalty fears help drive EHR investment*. Retrieved July 9, 2016, from http://ezpolson.nmu.edu:5749/ps/i.do?p=ITOF&u=lom_nmichu&id=GALE|A444941000&v=2.1&it=r&sid=summon&userGroup=lom_nmichu&authCount=1
- Simon, S., Kaushal, R., Cleary, P., Jenter, C., Volk, L., Orav, J., & Bates, D. et al. (2007). Physicians and electronic health records: A statewide survey. *Archives of Internal Medicine*, 167(5), 507–512. doi:10.1001/archinte.167.5.507 PMID:17353500
- United States Department of Health and Human Services. (2014). *Report to Congress: Update on the Adoption of Health Information Technology and Related Efforts to Facilitate the Electronic Use and Exchange of Health Information*. Retrieved November 2, 2016, from https://www.healthit.gov/sites/default/files/rtc_adoption_and_exchange9302014.pdf
- Valdes, I., Kibbe, D., Tolleson, G., Kunik, M., & Petersen, L. (2004). Barriers to proliferation of electronic medical records. *Journal of Innovation in Health Informatics*, 12(1), 3–9. doi:10.14236/jhi.v12i1.102 PMID:15140347
- Walter, Z., & Lopez, M. S. (2008). Physician acceptance of information technologies: Role of perceived threat to professional autonomy. *Decision Support Systems*, 46(1), 206–215. doi:10.1016/j.dss.2008.06.004
- Wu, J., Roy, J., & Stewart, W. (2010). Prediction Modeling Using EHR Data: Challenges, Strategies, and a Comparison of Machine Learning Approaches. *Medical Care*, 48(6), S106–S113. doi:10.1097/MLR.0b013e3181de9e17 PMID:20473190

Zaroukian, M. (2011). Driving on Main Street: The Road to Widespread Physician EHR Use. *Frontiers of Health Services Management*, 28(1), 35–41. PMID:21961384

ADDITIONAL READING

Dent, E. B., & Goldberg, S. G. (1999). Challenging resistance to change. *The Journal of Applied Behavioral Science*, 35(1), 25–41. doi:10.1177/0021886399351003

Jones, G. R. (2010). *Organizational theory, design, and change*. Upper Saddle River: Pearson.

Piderit, S. K. (2000). Rethinking resistance and recognizing ambivalence: A multidimensional view of attitudes toward an organizational change. *Academy of Management Review*, 25(4), 783–794.

Waddell, D., & Sohal, A. S. (1998). Resistance: A constructive tool for change management. *Management Decision*, 36(8), 543–548. doi:10.1108/00251749810232628

KEY TERMS AND DEFINITIONS

EHR Adoption Barrier: An obstacle that restricts, prevents, or discourages the adoption of EHRs.

EHR Diffusion: The adoption or spread of EHRs.

EHR Resistance: The opposition to accepting EHR technology.

EHR Technology Adoption: The selection and use of EHR technology.

Electronic Health Records (EHR): A digital record of an individual's healthcare that is collected over their lifetime.

Healthcare Stakeholders: An individual or group with interest in the progression and technological advancement of healthcare.

Meaningful Use: The U.S. government mandate of 2010 requiring an eligible provider to use certified technology to meet specified government benchmark outcomes.

Internet of Things Applications for Healthcare



Ljubica Diković

Business Technical College, Serbia

INTRODUCTION

According to the US National Intelligence Council, there are six technologies with potential impacts on the US interests out to 2025 (the US National Intelligence Council, 2009):

- Biogerontechnology as a technology related to the biological aging processes;
- Energy Storage Materials;
- Biofuels and Bio-based Chemicals;
- Clean Coal Technologies;
- Service Robotics;
- The Internet of things.

The great potential offered by the Internet of Things technology enables their wide applications in many areas of society, which would significantly increase and improve the quality of their functioning. By equipping various environments, i.e. domains, even with devices with primitive intelligence and modest communication capabilities, the communication of these entities with each other would be possible, with an aim to ensure data management. Such systems can be widely used in the following areas:

- Healthcare Domain;
- Smart Environment Domain;
- Personal and Social Domain;
- Transport and Logistics.

BACKGROUND

The Internet of Things (IoT) refers to wireless networks between objects (things). ‘Things’, i.e. objects, become entities with virtual properties which operate and communicate in smart spaces using intelligent interfaces.

Also, the “Internet of Things” is the general idea of things, especially everyday objects that are readable, recognizable, locatable, addressable, and controllable via the Internet - either via Radio Frequency Identification (RFID), Bluetooth, Wi-Fi, telephonic data services, wide-area network, or other means (the US National Intelligence Council, 2009).

In their research paper, Atzori et al. (2010) state that the Internet of Things can be realized in three paradigms: internet-oriented (middleware), things-oriented (sensors) and semantic-oriented (knowledge).

Over the last 20 years, continuous changes in the healthcare domain have taken place, caused by the wide use of information and communication technologies in the medical field. IoT plays a significant role in the broad range of healthcare applications which could be grouped as follows (Atzori et al. (2010)):

- Tracking of Objects and People (Staff and Patients);
- Identification and Authentication of People;
- Automatic Data Collection and Sensing.

The rapid growth of IoT has resulted in a massive growth of data generated by these devices and sensors put on the Internet. The physical-cyber-social big data consist of these IoT data, complemented by the relevant Web-based and social data (Sheth, 2016).

The Internet of Things has been identified as one of the emerging technologies in the IT field. The market adoption of IoT has been forecast to take 5–10 years (Gubbi et al., 2013).

The popularity of different paradigms varies with time. The web search popularity regarding the term IoT in industry, as measured by the Google search trends (n.d.) during the last 10 years, is shown as the red line for Germany and as the blue line for United States in Figure 1. As can be seen, the search volume is consistently increasing, and according to the Google search forecast, the trend is likely to continue. Average interest over time for United States is 22 and for Germany are 13. Numbers represent search interest relative to the highest point on the chart.

E-health in Serbia has been the subject of some studies. It was concluded that information and communication technologies are rarely implemented (Milenkovic et al., 2012).

IoT is a vision which refers to the humanization of technology.

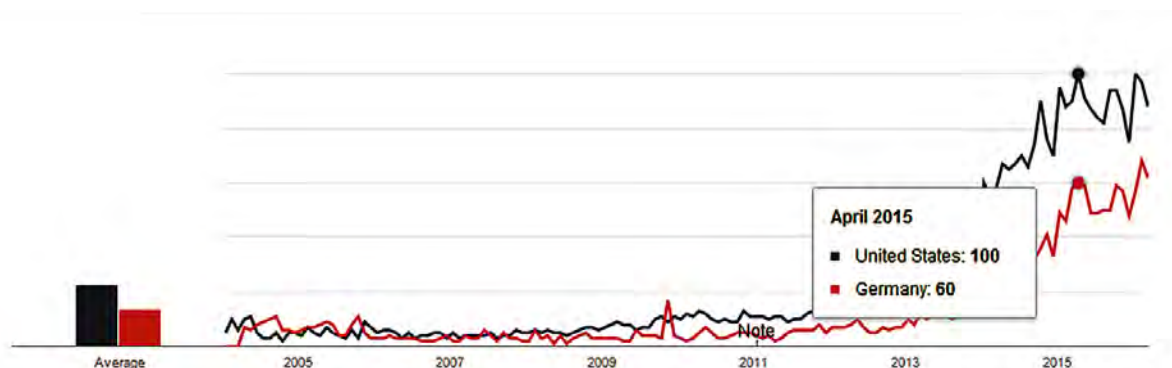
IoT IN HEALTHCARE DOMAIN

Medical sensors are devices that measure a number of physical, chemical, or biological parameters and then transmit or report these data. Some sensors are designed to work outside the body, and others are implanted in the body.

In the healthcare domain, the Internet of Things in its essence covers the following (Li et al., 2009):

- **Body Area Network (BAN):** The thermometers, smart t-shirts, smart devices and sensors for health, paper-based home pregnancy tests, etc. supporting personal medical treatment and healthcare.
- **Wireless Body Area Network (WBAN):** Supporting remote medical treatment and healthcare.
- **Local Area Network (LAN):** The wireless access-based remote patient monitoring system, the smart devices as a hospital interface, pulse oximeter also known as the blood-oxygen monitor, etc.
- **Wide Area Network (WAN):** Telemedicine solutions, distance medicine, etc.
- **Very Wide Area Network (VWAN):** The smart healthcare solutions as e-health services everywhere, no longer tied to physical locations.

Figure 1. Google search trends since 2005 for the term Internet of Things



Tracking involves memorizing the positions in real time (e.g. patient flow monitoring) in order to improve work processes in hospitals, but can also refer to the general tracking of motion (e.g. the access to certain rooms, warehouses or inventory material tracking such as blood samples and other materials necessary for the hospital functioning). Identification of patients includes activities, actions and tasks aimed at reducing incidents harmful to patients (e.g. wrong drug, dose, time, and procedures), the maintenance and updating of the complete medical archives in the digital form. Identification and authentication include protocols and procedures regarding the work of medical staff, which promote morals and ethics of employees in order to increase patient safety. The automatic data collection and transfer involves the automation of various processes as well as the management of medical supplies. This function is supported by the integration of the Radio Frequency Identification (RFID) technology and other health information and technologies. RFID is a method of storing and downloading data remotely, using devices carrying tags. From the RFID group's point of view, the Internet of Things is the worldwide network of interconnected objects uniquely addressable based on the standard communication protocols (Gubbi et al., 2013). Sensor devices allow more successful diagnosis of patients in real time, provide various solutions in the field of telemedicine, warnings about the patient's condition and so on. They offer numerous opportunities that would provide, for example, a wireless access-based remote patient monitoring system, which would be designed so as to accompany the patient everywhere in real time, thus enabling continuous monitoring of the patient's bio-signals (Niyato et al., 2009).

CONCEPT OF M-HEALTH

As mobile technology becomes more affordable and widespread all over the world, using

the mobile-enabled technology when providing healthcare services has great potential to improve the health of millions.

Mobile Health, mHealth is a term used for medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, etc.

MHealth is a relatively new concept, but its main advantage is that it enables healthcare professionals to collect data about their patients at any time. The data is processed, analyzed and wirelessly communicated to a gateway device over a cloud. (Paschou et al., 2013; Paschou et al., 2015; Istepanian et al., 2004).

Considering that MHealth, as a part of eHealth, can be defined as the synergistic solution of mobile computing, a medical sensor, and communications healthcare technologies, whose common goal is wireless monitoring of the psycho-physical state of health or remote patient monitoring. In the Republic of Serbia, electronic health cards are a pioneering venture aimed at promoting mobile healthcare services (Milenkovic et al., 2012).

An MHealth solution usually consists of a unit designed for patients, which is composed of sensors for medical measurements and accessories, as well as a communication device for 2G and 3G communication, which forwards the data to the health system via the cloud. Because of their miniature dimensions, high performance and low cost/price, different sensors, which can be easily wirelessly connected, are used for the measurement of chemical, biological and physical parameters in the body. Also, there are a lot of web-based platforms that provide additional opportunities for the expansion of social networks for different purposes (e.g. health education of patients worldwide). Even more, the network of sensors can be connected via patient's mobile phone to a server that is located in a medical institution. The data collected by the sensor network can be periodically or in real time sent to the server, which approaches the patient's doctor suggesting corrective actions via e-mail or SMS. The program

can be set so as to, if the sensor encounters some irregularities, immediately notify a doctor, who could react momentarily.

Thanks to this approach, patients with chronic diseases can be under continuous remote supervision at home. In addition to the increased safety of the patient, it represents significant savings of hospital resources, and hospital beds remain available for other patients who need them. Also, this solution helps hospitals manage beds efficiently and consequently decreases hospital costs. Hospitalized patients whose health status requires full attention can be constantly monitored using IoT-driven monitoring (sensors).

The market division is as follows:

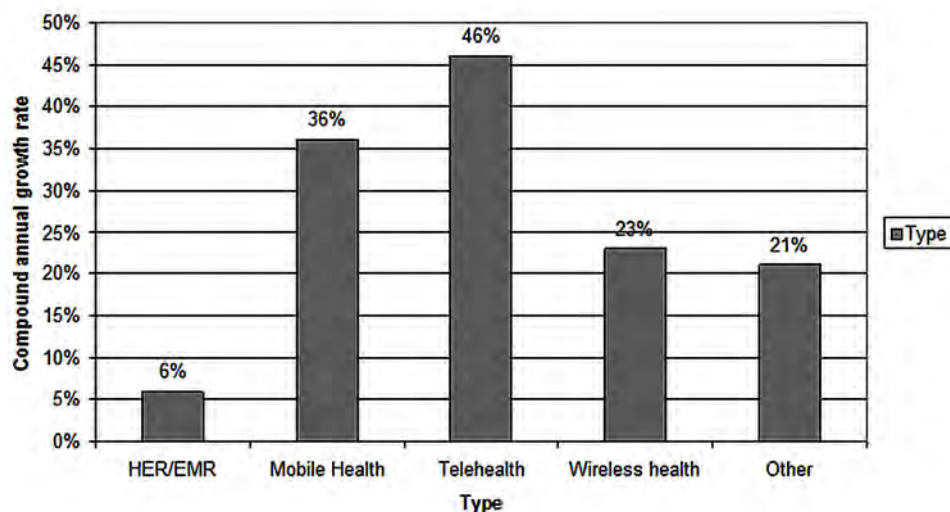
- Preventive care tools which help people stay healthy (e.g. Behavioral change, Community communication, Education).
- Curative care tools as new efficient ways to treat patients (e.g. Remote monitoring, Assisted living, Tele-consultations).
- Productivity tools which improve efficiency and availability (e.g. Electronic health records, Professional communication).

TELEHEALTH

Telehealth (telemedicine) is a health delivery system which refers to the exchange of medical information between two sites via electronic communications in order to improve a patient's health status. Telehealth includes a growing range of applications and services using two-way video telephony, wireless tools and networks, emails, smart devices, Internet and cloud services and infrastructures, and other forms of telecommunications technology. Clinical telemedicine practice supported by video conferencing, medical connectivity through m-health, remote patient monitoring, disease management, home treatment, continuing medical education, medical call centers, are all considered part of telemedicine and telehealth. Health professionals providing telehealth services should have the continuing education, training, licensure, in order to acquire the necessary knowledge and competencies needed to provide safe and high-quality telehealth services.

Figure 2 shows the projected compound annual growth rate (CAGR) for the global digital health market from 2013 to 2020, by segment. During this

Figure 2. Projected CAGR for the global digital health market from 2013 to 2020, by segment. Source: (Statistic, n.d.).



period, the telehealth market's compound annual growth rate is expected to be around 46 percent. The digital health market is expected to reach 233.3 billion US dollars by 2020 (Statistic, n.d.).

TECHNICAL REQUIREMENTS FOR REAL-TIME INTERACTIVE SESSIONS

The quality of a video healthcare interactive session primarily depends on the characteristics of the network connection between the conferencing sites. The H.323 standard specifies the components, protocols, and procedures for the transmission of real-time audio, video, and data communications over IP (Internet Protocol), i.e. packet-based, networks. In the case of H.323, a high-quality conference, which implies excellent audio and video, needs from 768 kbps (kilobits/second) to 2 Mbps (megabits/second) in each of the downlinks and uplink directions.

In the case of "Point to point video conferencing", two H.323 video conferencing systems communicate to each other over the Internet. One station initiates the call, and the other either accepts or rejects it. Once accepted, encoded and compressed audio and video flow between the two stations. In the case of "Multipoint video conferencing" more than two sites communicate with each other. In the H.320/ISDN world, a Multipoint Control Unit (MCU) was used. This allowed three or more user stations to connect to the MCU. In the case of the voice switching mode, the MCU made a decision about which site was currently talking. Smartphones and tablets (such as iPhones and iPads) do support live conferences; however, 3G/4G/LTE/WiFi connections are not sufficiently reliable for participating in online sessions. It is highly recommended to use a computer with a wired internet connection. In the case of high-quality (HQ) resolution, 640x360@15, 30 fps (frame per second) is recommended, but in the case of full High Definition resolution (FHD), 1920x1080 30 fps is recommended. Graphical

cards must be compatible with DirectX 10.0, from 512MB (min. NVidia 8600 GS, Radeon HD 2600 XT, Intel HD Graphics 2000) to 2GB (min. NVidia GT640, AMD Radeon HD 6670, Intel HD Graphics 4000).

Minimum system requirements for conference hosts are: Windows XP 32 bit or Windows 7 (32bit or 64bit); Core2 Quad, Intel i3, i5 or i7 equivalent CPU; 4 GB of RAM; 100 MB of available HDD space (additional space recommended for conference recording features); 100 MB NIC or higher; High speed broadband Internet access (at least 1 Mbps upstream and 4 Mbps downstream); DirectX 9.1 or later compatible webcam; DirectSound compatible audio input and headset or speakers (Video Conferencing System Requirements (n.d.), Web & Multiparty Video Conferencing System Requirements (n.d.), VCI-Next Generation Video Conferencing (n.d.)).

The video conference software should be able to adapt to changing bandwidth environments without losing the connection. Health organizations should have the appropriate redundant systems in a place that ensures availability of the data transmission infrastructure for critical connectivity.

DIGITAL TRANSFORMATIONS IN MEDICINE

mHealth has the potential for use in the following areas:

- Education and training of healthcare workers;
- Remote data collection and remote monitoring;
- Remote diagnostic and treatment support.

The main advantages brought by the concept of mHealth imply the increase in the quality of health care while creating new opportunities on the one hand, and reduced costs on the other hand. About 75% of patients suffer from chronic diseases (high blood pressure, diabetes, heart

disease, cancer, etc.). Using mHealth solutions, the biomarkers of these chronic patients can be monitored ensuring the quality management of their diseases in order to improve the quality of their life. These technologies exist and have been adopted, and their examples are the wireless Halter monitors, weight scales, blood pressure cuffs and glucometers (Malvey et al., 2014).

In order to be cost-effective and commercially-sustainable for the mass-market, mobile health services should be standardized. One of the problems is that today there is not an organization in the world which deals with the standards in the field of mHealth. For example, the International Standards Organization (ISO) has developed standards (IEEE 11073) for the transmission of blood pressure in a basic binary data format between two low-level devices, which can be transferred into a human-readable format using health messaging standards, such as HL7 (Engineering in Medicine and Biology Society, n.d.).

Some of the mHealth categories (Applications and Devices) are Nutrition Application, Glucometers, Biometrics devices, Sleep Trackers Applications and devices, Tobacco Cessation Applications and Programs, Medication Adherence Devices and so on (Validic, n.d.).

Some examples of the use of medical sensors are as follows: the sensors can help detect, and thus avoid, about 5000.00 known viruses that cause serious illness around the world. Biochemical sensors can continuously scan a billion miles long chain of Deoxyribonucleic acid (DNA) and look for changes and potential diseases. Taking into account the fact that the human heart throughout a life pumps about 15 million gallons of blood, cardiac sensors or heart rhythm sensors can provide information about how efficiently a heart works. The human brain has about 100 billion neurons. Gum sensors have the preventive effect on the development of other diseases in the organism because people who suffer from a gum disease represent 50% of those who are predisposed to suffer brain and heart attacks. New sensors can detect a high level of blood sugar without draw-

ing blood. By using an ultra-sensitive sensor to detect the presence of tumors and early abnormal cell growth, cancer can be recognized at an early stage and thus effectively treated. Biosensors may provide insights into disease processes that are hard to detect directly, such as dysfunctions in brain chemistry that are thought to play an important role in many mental disorders (Health Sensing Infographic, 2014; Sensors, n.d.).

The next generation of IoT quantified tracking devices falls into the category of wearable electronics and multi-sensor platforms. These products include smart watches, wristband sensors, wearable sensor patches, artificial reality-augmented glasses, brain-computer interfaces, wearable body metric textiles.

Other important categories include smartphone applications and their enhancements, and environmental monitoring and home automation sensors.

For example, Hexoskin products give insight into the physical training, sleep, and personal daily activities:

- Track heart rate and view real-time ECG with great precision.
- Heart rate variability (HRV) is an essential tool to measure stress and training fatigue to avoid overtraining.
- Breathing rate in order to control own breathing, increase performance and reduce stress.
- Breathing volume in order to control lung capacity for each of daily activities and avoid hyperventilation and apnea.
- Sleep tracking device in order to measure resting heart rate, breathing, and sleep positions.

CONCLUSION

The advances in science, technology, engineering, and medicine will allow the use of sophisticated sensor technologies in the further medical research that could drastically improve the quality

of life of people, particularly people with serious disabilities. Also, technological advancement in the healthcare area revolves around the following outcomes: improved access to care, increased care quality and reduced care costs. At the end, we can conclude that the technological The Internet of Things is potentially disruptive because global society, global economy and governments are unprepared for a possible future when the Internet nodes will be able to remotely control, locate, and monitor everyday things. Having in mind that everyday objects become information security risks, the IoT could distribute those risks far more widely than the Internet has to date.

FUTURE RESEARCH DIRECTIONS

Generally, the next Internet revolution will be the interconnection between everyday existing objects in order to create a smart grid and intelligent environment. The future application of technology in health care will lead to the creation of an entirely new level of personalized, digital health care, where everyone is responsible for monitoring their own health and the quality of their own life. Further research is aimed at improving the existing sensors through increasing their capabilities and enhancing their efficiency. In the coming period, IoT is expected to play the key roles in all the aspects of modern medical treatment and health management - prevention, diagnosis, disease monitoring, treatment monitoring.

REFERENCES

- Atzori, L., Iera, A., & Morabito, G. (2010). The Internet of Things: A survey. *Computer Networks*, 54(15), 2787–2805. doi:10.1016/j.comnet.2010.05.010
- Biometric shirts for performance improvement and sleep tracking. (n.d.). Retrieved March 28, 2016, from <http://www.hexoskin.com/>
- Global digital health CAGR by segment 2013-2020 | Statistic. (n.d.). Retrieved March 27, 2016, from <http://www.statista.com/statistics/387875/forecast-cagr-of-worldwide-digital-health-market-by-segment/>
- Google Trends. (n.d.). Retrieved March 27, 2016, from <https://www.google.com/trends/>
- Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. *Future Generation Computer Systems*, 29(7), 1645–1660. doi:10.1016/j.future.2013.01.010
- Health Sensing Infographic. (2014). Retrieved March 28, 2016, from <http://sensing.xprize.org/about/health-sensing-infographic>
- Istepanian, R., Jovanov, E., & Zhang, Y. (2004). Guest Editorial Introduction to the Special Section on M-Health: Beyond Seamless Mobility and Global Wireless Health-Care Connectivity. *IEEE Trans. Inform. Technol. Biomed.*, 8(4), 405–414. doi:10.1109/TITB.2004.840019
- Li, H., Takahashi, T., Toyoda, M., Mori, Y., & Kohno, R. (2009). Wireless Body Area Network Combined with Satellite Communication for Remote Medical and Healthcare Applications. *Wireless Pers Commun Wireless Personal Communications*, 51(4), 697–709. doi:10.1007/s11277-009-9765-5
- Malvey, D., & Slovensky, D. J. (2014). From Telemedicine to Telehealth to eHealth: Where Does mHealth Fit? *MHealth*, 19-43.
- MHealth Ecosystem | Validic. (n.d.). Retrieved March 27, 2016, from <http://validic.com/developers/mhealth-ecosystem>
- Milenkovic, D., Jovanovic-Milenkovic, M., Vujin, V., Aleksic, A., & Radojicic, Z. (2012). Electronic health system: Development and implementation into the health system of the Republic of Serbia. *Military Medical and Pharmaceutical Journal of Serbia VSP*, 69(10), 880–890. PMID:23155610

Nefsis Online. (n.d.). *Web & Multiparty Video Conferencing System Requirements*. Retrieved March 27, 2016, from <http://www.nefsis.com/Support-Video-Software/conferencing-system-requirements.html>

Niyato, D., Hossain, E., & Camorlinga, S. (2009). Remote patient monitoring service using heterogeneous wireless access networks: Architecture and optimization. *IEEE J. Select. Areas Commun. IEEE Journal on Selected Areas in Communications*, 27(4), 412–423. doi:10.1109/JSAC.2009.090506

Paschou, M., Papadimitriou, C., Nodarakis, N., Korezelidis, K., Sakkopoulos, E., & Tsakalidis, A. (2015). Enhanced healthcare personnel rostering solution using mobile technologies. *Journal of Systems and Software*, 100, 44–53. doi:10.1016/j.jss.2014.10.015

Paschou, M., Sakkopoulos, E., Sourla, E., & Tsakalidis, A. (2013). Health Internet of Things: Metrics and methods for efficient data transfer. *Simulation Modelling Practice and Theory*, 34, 186–199. doi:10.1016/j.simpat.2012.08.002

Sensors. (n.d.). Retrieved March 28, 2016, from <https://www.nibib.nih.gov/science-education/science-topics/sensors>

Sheth, A. (2016). Internet of Things to Smart IoT Through Semantic, Cognitive, and Perceptual Computing. *IEEE Intell. Syst. IEEE Intelligent Systems*, 31(2), 108–112. doi:10.1109/MIS.2016.34

VCI-Next Generation Video Conferencing. Blog. (n.d.). Retrieved March 27, 2016, from <http://www.vci.no/portal/Blog.aspx>

Video Conferencing System Requirements. (n.d.). Retrieved March 27, 2016, from <http://trueconf.com/support/system-requirements.html>

Your Global Connection to the Biomedical Engineering Community - Engineering in Medicine and Biology Society. (n.d.). Retrieved March 27, 2016, from <http://www.embs.org/>

ADDITIONAL READING

Appari, A., & Johnson, M. E. (2010). Information security and privacy in healthcare: Current state of research. *International Journal of Internet and Enterprise Management IJIEM*, 6(4), 279. doi:10.1504/IJIEM.2010.035624

IoT Reference Design Kit. (2015). *For Future Smart Connected Life and Business Collaborative Internet of Things (C-IOT)*, 209–223.

Kim, J. T. (2014). Privacy and Security Issues for Healthcare System with Embedded RFID System on Internet of Things.

Maksimovic, M., Vujovic, V., & Perisic, B. (2015). A custom Internet of Things healthcare system. *2015 10th Iberian Conference on Information Systems and Technologies (CISTI)*.

Mukherjee, S., Dolui, K., & Datta, S. K. (2014). Patient health management system using e-health monitoring architecture. *2014 IEEE International Advance Computing Conference (IACC)*. doi:10.1109/IAdCC.2014.6779357

Po, S. N., Dagang, G., Hapipi, M. D., & Hock, F. T. (2006). MEMSWear - Biomonitoring - Incorporating sensors into smart shirt for wireless sentinel medical detection and alarm. *J. Phys.: Conf. Ser. Journal of Physics: Conference Series*, 34, 1068–1072. doi:10.1088/1742-6596/34/1/176

Villar, R., Beltrame, T., & Hughson, R. L. (2015). Validation of the Hexoskin wearable vest during lying, sitting, standing, and walking activities. *Appl. Physiol. Nutr. Metab. Applied Physiology, Nutrition, and Metabolism*, 40(10), 1019–1024. doi:10.1139/apnm-2015-0140 PMID:26360814

Xiao, Z. R., Lv, B. G., Wang, X., & Zhao, Y. J. (2013). A Healthcare Service System Based on Internet of Things. *AMR Advanced Materials Research*, 774–776, 1903–1907. doi:10.4028/www.scientific.net/AMR.774-776.1903

KEY TERMS AND DEFINITIONS

Body Area Network: A network that can collect information about an individual's health, fitness, and energy expenditure.

Health Care Monitoring: The medical applications can be of two types: wearable and implanted. Wearable devices are used on the body surface of a human or just at a close proximity of the user. The implantable medical devices are those that are inserted into the human body.

Health Informatics (Healthcare Information Systems): Software solutions for appointment scheduling, patients' data management, work schedule management and other administrative tasks regarding health.

Internet of Things (IoT): A new concept that has emerged thanks to the development of new technologies, especially the Internet Technologies and Wireless Sensor Networks that together enable the existence of the health system at the global level.

Mobile Health (mHealth): A collection of mobile technologies as tools and platforms for health research and healthcare provision.

Sensor Network: A network that consists of multiple detection stations called sensor nodes, each of which is small, lightweight and portable.

Sensor Node: A node in a sensor network that is capable of performing some processing, gathering sensory information and communicating with other nodes connected to the network. The main components of a sensor node are a microcontroller, transceiver, external memory, power source and one or more sensors.

Telemedicine: A physical and psychological diagnosis and treatment at a distance, including telemonitoring of patients' functions.

Wireless Sensor Networks (WSN): A network that consists of spatially distributed autonomous sensors used to monitor physical or environmental conditions, and to forward their data jointly to the main location through the network.

Maintenance Policies Optimization of Medical Equipment in a Health Care Organization

Juan Ignacio Roig

University of Castilla-La Mancha, Spain

Andrés Gómez

University of Castilla-La Mancha, Spain

Isabel Romero

University of Castilla-La Mancha, Spain

María Carmen Carnero

University of Castilla-La Mancha, Spain & University of Lisbon, Portugal

INTRODUCTION

For a long period of time, maintenance has only focused on repairing breakdowns occurring in the machines as a result of its production. Maintenance policies has not been extensively analyzed and implemented in the actual companies; this aspect has been taken more into account in manufacturing companies than in service ones, where the maintenance was considered a department without influence on the service final quality (Gómez, Ruiz de la Hermosa, & Carnero, 2009). The concept of maintenance has evolved to be considered as a productive activity of the organization, since the correct operation of the equipment ensures the availability of production.

In a healthcare organization, the electromedical service is the department responsible for maintaining the entire healthcare equipment and specific facilities that cover the center, as well as its management through inventory control. It is essential that all medical equipment is properly maintained to ensure the highest level of availability and reliability. It is also important to have a maintenance strategy in which the maintenance policies are defined in order to keep track of medical equipment and to give priority to those required on critical

operations (Jamshidi, Abbasgholizadeh Rahimi, Ait-kadi, & Ruiz, 2015).

The aim of this chapter is to minimize the corrective breakdowns produced in the electro-medical equipment of a healthcare organization. To this, an optimization plan of maintenance policies will be developed in the equipment that needs it. The objective pursued is to increase scheduled preventive maintenance shutdowns carried out by Electromedical Department and to reduce the corrective maintenance. This plan includes the creation of preventive maintenance datasheets in which the activities undertaken, the periodicity, the estimated time to do the activity, etc. are identified.

The data used in this research has been obtained from the Computerized Maintenance Management System (CMMS) owned by the healthcare organization. Different types of rankings have been made in the stocks to prioritize items by family of equipment and by average work orders (WOs) by equipment generated. According to this analysis, the maintenance tasks are optimized to preserve the satisfactory working conditions and the performance of equipment in the healthcare organization optimizing available resources. All of this is intended to anticipate the maintenance

DOI: 10.4018/978-1-5225-2255-3.ch321

activity to breakdowns and to prevent the care quality from diminishing by increasing the patient waiting time by unavailability of medical equipment in the treatments and diagnostic tests.

The structure of this chapter is as follows. Next section draws a literature review where the concept of overall maintenance and maintenance of electromedical equipment in particular is detailed. Also, the ABC analysis is defined. Then, a general description of the healthcare center in which the study is carried out, an example of instruction datasheet and the analysis undertaken is done. The results obtained are shown below. Later, the conclusion of the research is presented. Finally, the references, additional readings and key terms used are shown.

BACKGROUND

For a long time, maintenance has been understood as part of the work that encompassed only the repairing breakdowns occurring in the machines and therefore, the employer included it as an economic burden to bear. However, this concept has progressively evolved over time until nowadays that is considered as a productive activity of the organization. Therefore, the role of maintenance has gained great interest and importance. Improving maintenance in a company is seen as an investment that will have a positive impact on quality and availability of the product as well as on productivity of the organization (Alrabghi & Tiwari, 2016).

It is common to consider in literature (see Dekker, 1996; Scarf, 1997; Eti, Ogaji, & Probert, 2006; and Hamdi, Oweis, Zraiq, & Sammour, 2012), that maintenance is the function responsible for ensuring availability of equipment and machines of a company; that is, it is associated with a set of actions that aim at preserving the assets (equipment and facilities) of the company during the period of useful life of these. Based on the criteria developed by cited authors above, the definition of maintenance is set as the range

of activities performed to a system, equipment or component to ensure that it continues to perform desired functions within an operational context given at the lowest possible cost. The main objective is to preserve its function, good operating conditions, to optimize performance and to increase the useful life, by ensuring optimal investment of resources.

Maintenance management includes all activities that determine the objectives, strategies, maintenance responsibilities and the implementation of all these. There are several significant factors to take into account to carry out an adequate management such as the importance and criticality equipment, the cost, the reliability and security, the environmental impact, etc. (Carnero & Gómez, 2016).

Maintenance department should have a software called Computerized Maintenance Management System (CMMS) to manage the maintenance of all equipment and facilities of a company, which incorporates a database and facilitates maintenance management with scheduling tasks to be performed by the maintenance department. This way, it will be possible to have a real time control of the whole system. This tool helps plan and manage the necessary functions to achieve an effective maintenance (Gulati & Smith, 2009).

Three basic kinds of maintenance policies are defined depending on the moment in time in which they are made, the purpose for which they are implemented and the resources used:

1. **Corrective Maintenance:** Performed activities to repair equipment when a failure appears (Carnero, 2009).
2. **Preventive Maintenance:** Scheduled actions which are performed while the system is running in order to keep it available and improve their conditions to avoid unpredicted failures (Doostparast, Kolahan, & Doostparast, 2014).
3. **Predictive Maintenance:** Actions that predicts system failure before it happens. The

purpose is to mitigate the failures looking for the underlying causes, and thus reduce the cost incurred by corrective maintenance (Faiz & Edirisinghe, 2009).

Particularly, medical equipment is essential in the prevention, diagnosis, and treatment of illness, as well as patient rehabilitation in hospitals (Noor, Magray, & Chawla, 2016). Electromedical equipment maintenance in a healthcare organization includes all equipment and high level of technology facilities in any field in bioengineering (Gómez, 2013) for instance: capturing and processing images of high level of digital resolution techniques, diagnosis and treatment with radioactive isotopes, conventional and laser surgical techniques, obtainment analytical techniques by robotized chains, cryogenic systems, precision mechanics, microscopy, with highly advanced components and advanced control systems.

Maintenance management in hospitals is much more complex than other manufacturing organizations, for it undertakes medical and health responsibilities that deal with the patient's life (Sadaghiyani, 1998). Moreover, there are a large amount of medical devices with high technology. For these reasons, it must be incorporated an excellent management tool to classify all equipment effectively.

The ABC analysis allows the classification of items according to their importance. It is based on the application of the Pareto principle, also known as 80-20 rule according to which, about 80% of the effect come from 20% of the main causes (Koch, 1998). This technique is the most commonly used in inventory management to categorize a large number of articles into three ordered and predefined groups: group A includes the very important items, group B contains the moderately significant pieces and category C includes the relatively insignificant items (Douissa & Jabeur, 2016).

This tool can be used in a lots of applications. If applied to the maintenance of electromedical equipment, machines can be divided into 3 groups:

- **Group A:** It includes approximately 10-20% of total machines, but accounts for 60-80% of the total maintenance cost.
- **Group B:** It constitutes about 30-40% of equipment and account for 20-30% of the total maintenance cost.
- **Group C:** A numerous groups of machines around 50-60% represent only 5-10% of the total maintenance cost.

The really important are groups A and C since they provide information about a few medical devices that are significant and a large number of machines that are relatively insignificant. The percentage may vary depending on the case. The objective of this technique is to manage in an effective way a maintenance management program by determining which equipment is more important to keep available.

HOSPITAL DESCRIPTION

Medical equipment of a new generation hospital is analyzed. Equipment and facilities of all branches of bioengineering such as different capturing and processing images techniques, diagnosis and treatment with radioactive isotopes, conventional and laser surgical techniques, obtainment analytical techniques, cryogenic systems, precision mechanics, microscopy and generally, equipment with latest generation components are included.

The hospital has the following services (Gómez, 2013): 540 hospital beds, 17 oncohaematological outpatient care stations, 28 dialysis stations, surgical and obstetrical wing, Intensive Care Unit (ICU), emergency department, nuclear medicine, radiation therapy, physical therapy and 119 out-

patient visits. It also has support services such as laboratories, blood bank and hospital pharmacy.

PREVENTIVE MAINTENANCE DATASHEETS

In order to reduce the corrective maintenance, preventive maintenance datasheets are created, in which the activities undertaken, the periodicity, the estimated time to do the activity, etc. are identified. Using preventive maintenance datasheets, apart from meeting the recommended safety parameters for each family of electromedical equipment, aims to reduce corrective maintenance. Such approach shows the most optimal and suitable for health care center facilities because if it is a serious breakdown to stop a service for a long time, it may be a decrease in care quality by increasing the patient waiting time. Moreover, it is cost-effective to replace a worn component of a machine when it is detected in a preventive maintenance review than having to repair if a catastrophic failure occurs, since these often induce failures in other components of medical equipment.

Preventive Maintenance Datasheets of Electrosurgical Equipment

H

An example of the content included in a preventive maintenance datasheets of electrosurgical electro-medical equipment developed is showed below. The datasheet consists of the following sections:

1. General test
2. Leakage test
3. Power test
4. Frequency test
5. Electrical safety

Also, the measurement devices used, the name and signature of the technician and the date are included in all preventive maintenance datasheets.

All instruction datasheets have been made following the specific electromedical equipment regulation. When the appropriate preventive maintenance has been performed, each maintenance technician must fill in the datasheet. Figure 1 and Figure 2 shown the electrosurgical datasheet template.

Figure 1. Section 1 and 2 of electrosurgical datasheet

Source: (Own elaboration, 2016)

ELECTROMEDICAL EQUIPMENT ANNUAL MAINTENANCE					
SERVICE	Maintenance department	OT number			
EQUIPMENT	Electrosurgical	EQUIPMENT			
					Good Bad

1 GENERAL TEST

1.1 Power, network cable and plug

1.2 Switches, buttons, auxiliary control and foot control

1.3 Acoustic and light indicator

1.4 Sheets and accessories condition

2 LEAKAGE TEST

OPEN CONFIG. (F)

I (mA)

REFERENCE

I (mA)

☐☐
☐☐
☐☐
☐☐

☐☐
☐☐
☐☐
☐☐

2.2 Leakage measurement

≤ 150 mA

OPEN

CLOSE

2.1 Leakage measurement

≤ 150 mA

OPEN

CLOSE

≤ 100 mA

☐☐
☐☐
☐☐
☐☐

☐☐
☐☐
☐☐
☐☐

Figure 2. Section 3, 4 and 5 of electrosurgical datasheet

Source: (Own elaboration, 2016)

3 POWER TEST (According to UNE-EN 60601-2-2)

Measuring range $\leq \pm 20\%$

MODE		$R_C = 200 \Omega$			$R_C = 500 \Omega$			$R_C = 750 \Omega$			$R_C = \infty$		
		Min	Med	Max	Min	Med	Max	Min	Med	Max	Min	Med	Max
PWR													
I (mA)													

4 FREQUENCY TEST

5 ELECTRICAL SAFETY

MEASUREMENT DEVICES USED	Calibration Date

NAME AND SIGNATURE OF THE TECHNICIAN

DATE:

ELECTROMEDICAL EQUIPMENT ANALYSIS

In order to establish a ranking according to the impact regarding the preventive maintenance work orders that are produced, two analyses have been made that encompass 187 electromedical equipment belong to the healthcare center. Two rankings are established under Pareto principle are:

1. By work orders by family of equipment.
2. By average work orders by equipment.

According to Pareto analysis, Table 1 and Table 2 show the results included in A category of first and second ranking respectively. All equipment is not shown because of a lack of space.

A simply method of classification is considered to get successful results. The key of this arrangement is that while there is not used a creative contribution, it is a tool that works efficiently. In other

words, it consists in ensuring that electromedical equipment are ranked in order of importance in a suitable manner.

SOLUTIONS AND RECOMENDATIONS

The following results have been achieved in the ABC analysis. In the first ranking, the result obtained is that 20.32% of the family of medical equipment accounts for 68.71% of work orders whereas in the second case, 20.32% of the medical equipment accounts for 47.01% of work orders.

The first analysis is classified according to the highest work orders irrespective of included equipment, whereas in the second, they are taken into consideration. In other words, in the first study, equipment that has many WOs and a large number of equipment are at the top for not taking into account the number of equipment while after

Table 1. ABC by family of medical equipment

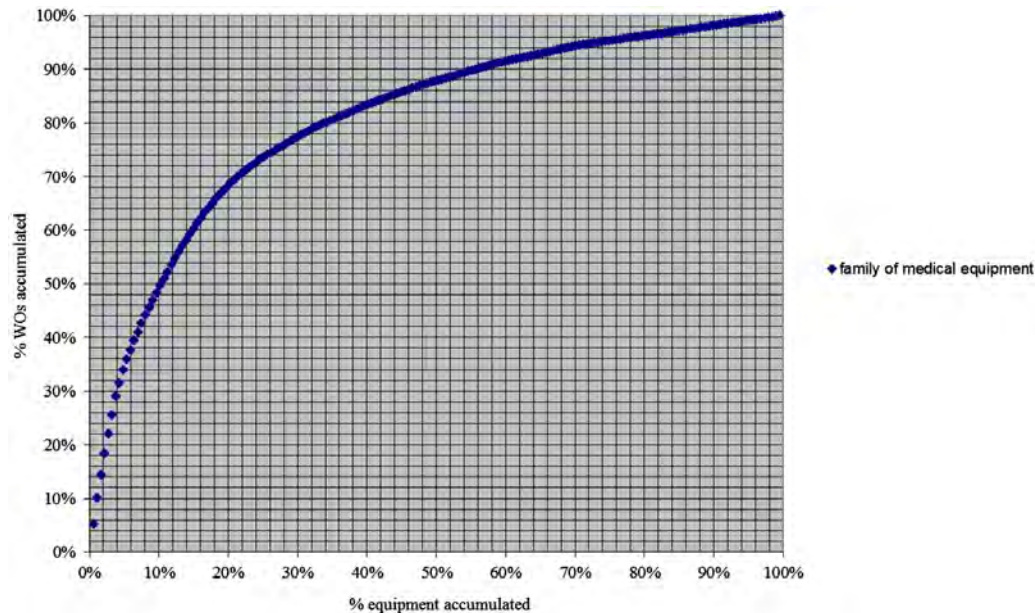
Code	Equipment Description	Number of Devices	% System Accumulated	N° WO Family	% Wo Accumulated by Family
1	Auto-analyzer	2	0.53%	103	5.22%
2	Hematologic analyzer	5	1.07%	97	10.13%
3	ECG2	27	1.60%	85	14.43%
4	Blood pressure monitor with pulse oximetry	20	2.14%	79	18.43%
5	Anesthetic respirator	16	2.67%	71	22.03%
6	Pulse oximetry	23	3.21%	70	25.57%
7	ECG1	5	3.74%	69	29.06%
8	Volumetric respirator	16	4.28%	50	31.59%
9	Perfusion pump	22	4.81%	48	34.03%
10	Multi-parameter monitor with SpO2, NIBP, PI (Touch Screen)	10	5.35%	38	35.95%
11	Ultrasound scanner	12	5.88%	35	37.72%
12	Could light source	6	6.42%	34	39.44%
13	Portable RX Equipment	2	6.95%	33	41.11%
14	Pyxis	8	7.49%	32	42.73%
15	Ontological microscope (NT)	10	8.02%	30	44.25%
16	Electroencephalograph	2	8.56%	28	45.67%
17	Autoclave	3	9.09%	27	47.04%
18	Washing machine for endoscopy tubes N° 3	3	9.63%	27	48.41%
19	Autoclave miniclave	5	10.16%	26	49.72%
20	Engine traumatology	4	10.70%	25	50.99%
21	Electrosurgical	7	11.23%	25	52.25%
22	Surgical table	7	11.76%	25	53.52%
23	Neonatal incubator	8	12.30%	25	54.78%
24	Multi-parameter monitor with SpO2	7	12.83%	23	55.95%
25	Muti-energy linear accelerator	1	13.37%	22	57.06%
26	Multi-parameter monitor	9	13.90%	22	58.18%
27	Electronic sphygmomanometer	7	14.44%	21	59.24%
28	Slip lamp	5	14.97%	21	60.30%
29	Ophthalmic chart projector	4	15.51%	20	61.32%
30	Blood pressure monitor with pulse oximetry and temperature	7	16.04%	20	62.33%
31	Applanation Tonometry	6	16.58%	19	63.29%
32	Coagulometer	3	17.11%	18	64.20%
33	RX emergency room	1	17.65%	18	65.11%
34	Equipment washing machine	2	18.18%	16	65.92%
35	Cardiotocograph	5	18.72%	15	66.68%
36	Packaging machine solid forms	1	19.25%	14	67.39%
37	Central monitors	4	19.79%	14	68.10%
38	Freezer	6	20.32%	14	68.81%

Table 2. ABC by average work orders by medical equipment

Code	Equipment Description	Number of Devices	% System Accumulated	Average WO by Equipment Family	% WO Accumulated by Family
1	Auto-analyzer	2	0.53%	51.50	6.39%
25	Muti-energy linear accelerator	1	1.07%	22.00	9.12%
2	Blood pressure monitor	5	1.60%	19.40	11.53%
33	RX emergency room	1	2.14%	18.00	13.76%
13	Portable RX Equipment	2	2.67%	16.50	15.81%
16	Electroencephalograph	2	3.21%	14.00	17.54%
36	Packaging machine solid forms	1	3.74%	14.00	19.28%
7	ECG1	5	4.28%	13.80	20.99%
42	Incubator for anaerobic	1	4.81%	11.00	22.36%
43	Steam Autoclave 80 l	1	5.35%	11.00	23.72%
45	Automated glass coverslipper	1	5.88%	10.00	24.96%
46	Dual-head gamma camera (N° Sist DPN00002)	1	6.42%	10.00	26.20%
17	Autoclave	3	6.95%	9.00	27.32%
18	Washing machine for endoscopy tubes N° 3	3	7.49%	9.00	28.44%
47	Multi-parameter monitor with SpO2	1	8.02%	9.00	29.55%
34	Equipment washing machine	2	8.56%	8.00	30.55%
49	RX generator	1	9.09%	8.00	31.54%
51	Formaldehyde autoclave	1	9.63%	8.00	32.53%
52	SCTL receptor microbiology	1	10.16%	8.00	33.52%
58	Multi-parameter monitor with SpO2, PNI	1	10.70%	7.00	34.39%
59	Cryostats	1	11.23%	7.00	35.26%
20	Engine traumatology	4	11.76%	6.25	36.04%
32	Coagulometer	3	12.30%	6.00	36.78%
39	Water purifier	2	12.83%	6.00	37.52%
64	Multi-parameter monitor with SpO2	1	13.37%	6.00	38.27%
65	Digital mammography and acquisition workstation	1	13.90%	6.00	39.01%
66	Engine maxillofacial surgery	1	14.44%	6.00	39.76%
67	PET-CT (N° Sist. DPN00012)	1	14.97%	6.00	40.50%
68	Immunochemical system	1	15.51%	6.00	41.25%
69	Arthroscopy control unit	1	16.04%	6.00	41.99%
12	Could light source	6	16.58%	5.67	42.69%
19	Autoclave miniclave	5	17.11%	5.20	43.34%
29	Ophthalmic chart projector	4	17.65%	5.00	43.96%
74	Multi-parameter monitor with SpO2	1	18.18%	5.00	44.58%
75	Ortopantomography	1	18.72%	5.00	45.20%
76	spectrophotometer	1	19.25%	5.00	45.82%
77	Robot for analytical.	1	19.79%	5.00	46.44%
78	Tower digestive endoscopy	1	20.32%	5.00	47.06%

Figure 3. ABC graph by family of medical equipment

Source: (Own elaboration, 2016)



making the second analysis, the ranking position decreased because WOs by equipment are reduced. Figures 3 and 4 illustrate this.

In line with this classification, the first 38 families of medical equipment that set up Pareto which produce the most WOs have been examined in order to optimize the Preventive Maintenance Plan as well as to achieve the objective of diminishing breakdowns, that is to say, smaller corrective work orders.

The following stages have been carried out:

Step 1: Identification of breakdowns:

- Number.
- Cause (failure mode).

Step 2: Establishment of overall conclusions taking into account several aspects:

- Repetitive failures linked to badly use on the part of the user. Design training activities for those responsible for the services involved. Training action plan.

- Repetitive failures due to normal use. Identification of tasks and the frequency to be implemented in the range of preventive maintenance to be created for reducing the repetition of this failure mode.

Some examples are shown below.

Neonatal Incubator

Device whose function is to create an environmental with humidity, temperature and oxygen concentration suitable for growth of neonates.

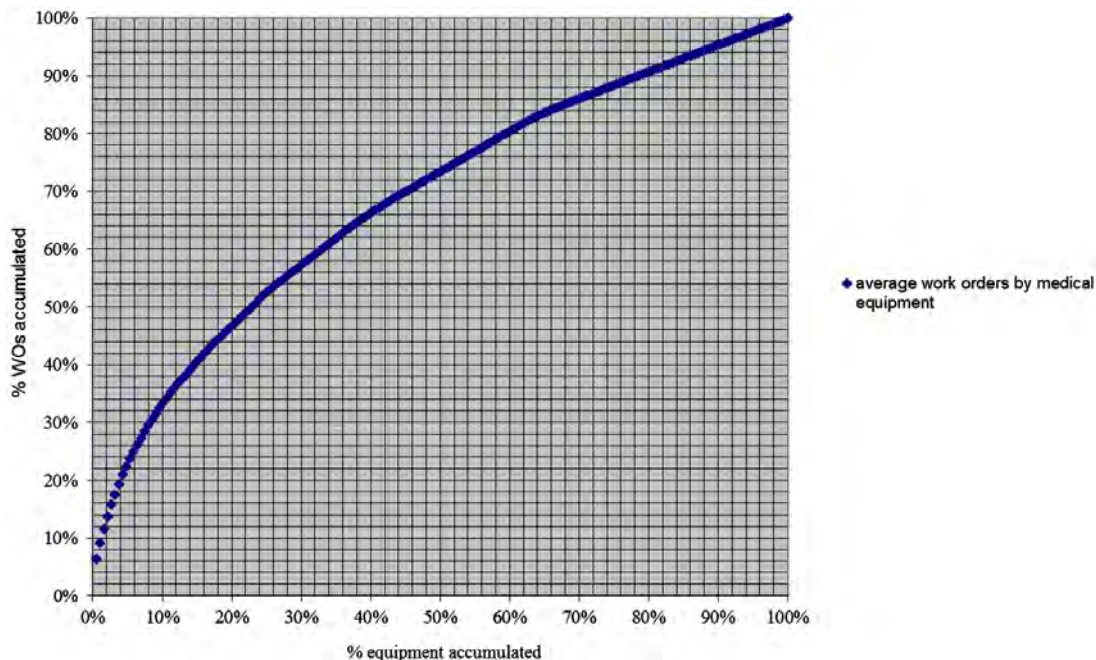
Step 1: Identification of breakdowns: fan breakdowns.

- **Number:** 5 WOs are produced (20% of total failures).
- **Cause - Misplacement of the Fan:** Damping spring adjustment or bearings lubrication.

Step 2: Established findings:

Figure 4. ABC graph by average work orders by medical equipment

Source: (Own elaboration, 2016)



- Training guide or video to install correctly the fan.
- Preventive maintenance (operation to include in the existing range):
 - **Visual Review:** Determine if there are problems that can apparently be detected and the lack or the loss of accessories that are part of equipment as signs of short or damages.
 - **Cleaning:** Clean at least some parts and equipment surfaces, using suitable cleaning compounds for each of them.
 - **Test Runs:** Check that the equipment are completely operational within the specific limits with in the implementation of the software auto-test.

Electrosurgical

Electrosurgical device based on electronic technology be able to produce a given set of high

frequency electromagnetic waves for the support of remove or cut soft tissue.

Step 1: Identification of breakdowns: damage in the bipolar forceps cable and failures due to performance reviews.

- **Number:** 6 WOs are produced (24% of total failures) and 7 WOs (28% of total failures) respectively.
- **Cause:** Part failure and misuse.

Step 2: Established findings:

- Implementation of a daily protocol review by the user to avoid performance reviews in which no action is taken.
- Preventive maintenance:
 - **External Cleaning:** To keep equipment clean.
 - **Connector and Electronic Boards:** Dirt accumulation on the boards produce breakdowns so they must be kept clean. Ensure that all connector are properly connected.

Multi-Parameter Monitor

Equipment for the simultaneous display of different biological patient parameter such as electrocardiogram, arrhythmia analysis, cardiac output, blood pressure, breathing, blood oxygen saturation, etc.

Step 1: Identification of breakdowns: failures in non-invasive blood pressure cuffs and in the sensors.

- **Number:** 4 WOs are produced (18.18% of total failures).
- **Cause:** Break of sensors or cuffs and imbalance of parameters. They have no review protocol. Failures due to the user.

Step 2: Established finding:

- Implementation of a daily protocol review by the user to avoid performance reviews in which no action is taken.

RX Emergency Room

Equipment for a non-invasive medical test used to diagnose and treat medical conditions. Imaging with x-rays implicates exposing a part of the body to a small amount of ionizing radiation to produce images of the inside of the body.

Step 1: Identification of breakdowns: Breakdowns on the grid to filter the radiation produced (Bucky).

- **Number:** 6 WOs are produced (33.33% of total failures).
- **Cause:** Inappropriate way of working with the equipment producing an imbalance in the Bucky drive mechanism

Step 2: Established finding:

- Implementing a training program about the way of working with the system.

- Analyze failures next year to assess whether problems in the Bucky continue to exist. Thus, the origin must be identified and the problem solved.

FUTURE RESEARCH DIRECTIONS

According to this analysis, apart from achieve an increasing patient satisfaction by decreasing waiting times, a continuous improvement in the control and management of maintenance will be accomplished. In future research, it should be checked that benefits have been obtained and for the following years, similar studies will be performed in order to update the examination and to implement a continuous improvement process. Also, due to the medical technological renewal current rate, it has become necessary to update analysis with new family of medical equipment that are being introduced into hospital.

In short, all of this is intended to optimize maintenance policies of electromedical equipment belonging to the hospital. As a result, it is expected to get a reduction of corrective activities, just as substantial improvement in the effective use of resources by electromedical department.

CONCLUSION

An ABC analysis of electromedical equipment in a healthcare organization has been performed to introduce a preventive maintenance ranges in the equipment that need it. Daily maintenance tasks that minimize unscheduled shutdowns have been generated. This will greatly improve available resources in electromedical department. A healthcare organization has large amount of maintenances requirements for the electromedical facilities because of the equipment are used directly with patients. For this, it is important improvement maintenance policies carried out by the workers. This straightforward investigation

will help the maintenance department in order to achieve this aim.

The maintenance control in a healthcare center is essential to carry out maintenance tasks optimizing resources without negatively affecting care quality. The issue of optimizing maintenance policies in healthcare organizations is not well known, thus, there are almost non-existent contributions. Applying this new analysis, in addition to achieving increased patient satisfaction, an improved management and control of maintenance will be accomplished.

ACKNOWLEDGMENT

This research was supported by the Junta de Comunidades de Castilla-La Mancha and the European Regional Development Fund under Grant number PPII-2014-013-P.

REFERENCES

- Alrabghi, A., & Tiwari, A. (2016). A novel framework for simulation-based optimisation of maintenance systems. *International Journal of Simulation Modelling*, 15(1), 16–28. doi:10.2507/IJSIMM15(1)2.316
- Arques, J. L. (2009). *Ingeniería y gestión del mantenimiento en el sector ferroviario*. España: Ediciones Díaz de Santos.
- Carnero, M. C. (2009). Evaluating a Maintenance Department in a Service Company. *International Journal of Mathematical Models and Methods In Applied Sciences*, 3(3), 230–237.
- Carnero, M. C., & Gómez, A. (2016). A multicriteria decision making approach applied to improving maintenance policies in healthcare organizations. *BMC Medical Informatics and Decision Making*, 16–47. PMID:27108234
- Dekker, R. (1996). Applications of maintenance optimization models: A review and analysis. *Reliability Engineering & System Safety*, 51(3), 229–240. doi:10.1016/0951-8320(95)00076-3
- Doostparast, M., Kolahan, F., & Doostparast, M. (2014). A reliability-based approach to optimize preventive maintenance scheduling for coherent systems. *Reliability Engineering & System Safety*, 126, 98–106. doi:10.1016/j.ress.2014.01.010
- Douissa, M. R., & Jabeur, K. (2016). A New Model for Multi-criteria ABC Inventory Classification: PROAFTN Method. *Procedia Computer Science*, 96, 550–559. doi:10.1016/j.procs.2016.08.233
- Eti, M. C., Ogaji, S. O., & Probert, S. D. (2006). Reducing the cost of preventive maintenance (PM) through adopting a proactive reliability-focused culture. *Applied Energy*, 83(11), 1235–1248. doi:10.1016/j.apenergy.2006.01.002
- Faiz, R., & Edirisinghe, E. (2009). Decision Making for Predictive Maintenance in Asset Information Management. *Interdisciplinary Journal of Information. Knowledge and Management*, 4, 23–36.
- Gómez, A. (2013). *Metodología para la optimización de la gestión y del mantenimiento de activos en un centro sanitario*. Ciudad Real: Universidad de Castilla-La Mancha.
- Gómez, A., Ruiz de la Hermosa, R., & Carnero, M. C. (2009). Desarrollo procedimental del mantenimiento industrial en un hospital. *Dirección y Organización*, 37, 83–90.
- Gulati, R., & Smith, R. (2009). *Maintenance and Reliability Best Practices*. New York: Industrial Press.
- Hamdi, N., Oweis, R., Zraiq, H. A., & Sammour, D. A. (2012). An intelligent healthcare management system: A new Approach in work-order prioritization for medical equipment maintenance requests. Springer. *Journal of Medical Systems*, 36(2), 557–567. doi:10.1007/s10916-010-9501-4 PMID:20703695

Ishikawa, K. (1994). *Introducción al control de calidad* (J. Nicolau & M.M. Gonzalbes, trans.). Madrid: Ediciones Díaz de Santos, S.A. (Original work published 1989)

Jamshidi, A., Abbasgholizadeh Rahimi, S., Aitkadi, D., & Ruiz, A. (2015). A comprehensive fuzzy risk-based maintenance framework for prioritization of medical devices. *Applied Soft Computing*, 32, 322–334. doi:10.1016/j.asoc.2015.03.054

Koch, R. (1998). *The 80/20 Principle: The secret of achieving more with less*. London: Nicholas Brealey Publishing.

Noor, M. M., Magray, I. A., & Chawla, S. (2016). Integration of Healthcare System with its Experts for Improving the Life Expectancy of MEDical Decives: A Review. *International Journal of Scientific Research in Science and Technology*, 2(2), 223–231.

Sadaghiyani, E. (1998). *Hospital's Organization and Management* (1st ed.). Tehran: Jahanrayane Publications.

Scarf, P. (1997). On the application of mathematical models in maintenance. *European Journal of Operational Research*, 99(3), 493–506. doi:10.1016/S0377-2217(96)00316-5

ADDITIONAL READING

Al-Zubaidi, H., & Christer, A. H. (1997). Maintenance manpower modelling for a hospital building complex. *European Journal of Operational Research*, 99(3), 603–618. doi:10.1016/S0377-2217(96)00324-4

Gómez, A., & Carnero, M. C. (2011). Selection of a Computerised Maintenance Management System: A case study in a regional health service. *Production Planning and Control*, 22(4), 426–436. doi:10.1080/09537287.2010.500455

Iung, B., Véron, M., Suhner, M. C., & Muller A. (2005). Integration of maintenance strategies into prognosis process to decision-making aid on system operation. *CIRP Annals - Manufacturing Technology*, 54(1), 5–8.

Liljegren, E. (2006). Usability in medical technology context assessment of methods for usability evaluation of medical equipment. *International Journal of Industrial Ergonomics*, 36(4), 345–352. doi:10.1016/j.ergon.2005.10.004

Mobley, R. K. (2004). *Maintenance fundamentals*. Burlington, MA: Elsevier Butterworth-Heinemann.

Moore, W. J., & Starr, A. G. (2006). An intelligent maintenance system for continuous cost-based prioritisation of maintenance activities. *Computers in Industry*, 57(6), 595–606. doi:10.1016/j.compind.2006.02.008

Shafiee, M. (2015). Maintenance strategy selection problem: An MCDM overview. *Journal of Quality in Maintenance Engineering*, 21(4), 378–402. doi:10.1108/JQME-09-2013-0063

Wang, H. (2002). A survey of maintenance policies of deteriorating systems. *European Journal of Operational Research*, 139(3), 469–489. doi:10.1016/S0377-2217(01)00197-7

KEY TERMS AND DEFINITIONS

ABC Analysis: Rating method based on Pareto principle. In maintenance of equipment, allows to identify items depending on generated impact and divide them into three group to establish an order of priorities when managing maintenance.

Computerized Maintenance Management System (CMMS): A software package which manages resources of maintenance department, incorporating a database continuously updated. It allows to schedule preventive, corrective and predictive maintenance tasks.

Electromedical Department: Technical service department whose activity includes total maintenance of medical equipment in a healthcare organization and specific facilities which provide coverage the same, as well as management through the inventory unit responsible for updating existing equipment.

Electromedical Equipment: Equipment and facilities belonging to a healthcare organization.

Healthcare Organization: Center that provides health services such as diagnosis of diseases, surgical operations and treatment and recovery of patients. Also, research and teaching assignments can be performed.

Maintenance Datasheet: Document which specifies preventive maintenance activities that are required to minimize corrective maintenance of electromedical equipment to comply with safety parameters.

Maintenance: A set of actions responsible for ensuring machines and production equipment availability in a company.

Work Orders (WOs): Generated document when a failure of any equipment occurs in order to achieve maintenance tasks to solve the problem.

The Optimal Workforce Staffing Solutions With Random Patient Demand in Healthcare Settings



Alexander Kolker
GE Healthcare, USA

INTRODUCTION

The current demand for more efficient use of healthcare resources, cost reduction, and patient safety improvement stimulates developing the novel operations management solutions. Workforce management has emerged as the one with the biggest impact on cost efficiency and quality of care in the USA. Indeed, labor cost typically absorbs about 54% of total hospital's operating revenue (Herman, 2013). Therefore, an accurate assessment of the required staffing that matches the highly variable patient demand is an integral part of the hospital's budgeting and planning process. Healthcare administrators must accomplish multiple clinical and quality goals while simultaneously developing realistic staffing plans and budgets. Random fluctuations of patient demand present staffing planning challenges to many hospitals. There is a growing trend for using data analytics to address these challenges.

The objective of this chapter is providing an overview and examples of application of the data analytics methodology called the "newsvendor" framework. This methodology helps to determine the optimal staffing solutions for the specified time periods for hospital units with randomly fluctuating daily patient census.

BACKGROUND

The newsvendor model is the widely used analytic model in which the optimal inventory level is determined for a specified time period. Historically, it originates from the problem in which a newsvendor has to decide on the optimal stocking quantity of the newspaper (a single product) to be ordered from the publisher in some defined ahead time period; hence it is called the newsvendor problem. If too many issues are ordered there will be some financial loss due to unsold inventory. If not enough issues are ordered there will also be some financial loss due to unmet customer demand. The problem is determining the optimal quantity order that will minimize the total financial loss due to both over- and understock during some time period.

The newsvendor framework has been widely applied to problems in which decisions should be made on the fixed supply level with an uncertain (random) demand. Such problems are often occurring in supply chain management, retail, transportation, manufacturing, banking, and many other industries (Choi, 2012; Arikan, 2011; Porteus, 2002). Motivated by the importance of various practical applications of the newsvendor model, the entire special issue of the Decision Sciences journal (Chen et al., 2016), and the review paper (Qin et al., 2011) were dedicated to its novel advances and applications.

At the same time, the use of the newsvendor framework was rather limited in healthcare management for planning and budgeting the hospital units' staffing while patient demand is uncertain.

For example, in the handbook of newsvendor problems, which is the first handbook dedicated exclusively to the state of the art in this area (Choi, 2012), the optimal nursing staffing problem with uncertain patient demand was not presented at all. However, this is a fruitful area of application of the newsvendor framework. The long-term nursing staffing plans should be developed on the annual basis. The medium-term staffing plans should usually be developed for a 4-6 weeks period, and be posted 1-2 weeks before the start of the planned period. Because of inevitable occurrences of unforeseen deviations from the planned staffing level, some short-term staffing adjustments should be made shortly before each shift to determine whether overtime, pooled or agency nurses are needed, or if the unit is overstaffed and some nurses are not currently needed. There is a cost associated with flexing staff up or down, along with issues of staff dissatisfaction with the erratic unpredictable schedules. There is an empirical evidence that the frequent staffing adjustments costs are accumulated to significant amounts that were not previously budgeted for. The optimal staffing level determined by the newsvendor model minimizes these accumulated costs, thus making nursing staffing plans and budgets more realistic.

One of a few publications that mention the use of the newsvendor framework for determining the optimal nursing staffing level is Hopp and Lovejoy (2013). These authors included the newsvendor framework as one of the management principles: in a single time period (month, quarter, or year) with uncertain staffing demand, the staffing level which corresponds to the minimal (optimal) possible total cost of under- and overstaffing is given by the solution, S , of equation

$$F(S) = Cu / (Cu + Co) \quad (1)$$

where $F(S)$ is the cumulative distribution function of staffing demand, Cu and Co are the hourly costs of under- and overstaffing per nurse, respectively. Derivation of Equation 1 is given in Appendix.

These authors assumed that staffing demand in any single time period is normally distributed with the known mean, m , and standard deviation, σ . Therefore, the optimal staffing level S is given by $S = m + Z * \sigma$, where Z is the inverse standard cumulative normal distribution function, i.e. $Z = \Phi^{-1}[Cu / (Cu + Co)]$.

At the same time, the assumption of normal staffing demand distribution is rarely true (if ever). Staffing demand is usually a skewed distribution function that poorly fits any theoretical distribution. This is pointed out by Davis et al. (2014) in which the newsvendor framework is applied for analyzing nursing staffing under demand uncertainty. The authors propose the use of distributional robust formulations of the newsvendor problem suggested by Perakis et al. (2008). In this formulation, it is sufficient to know only the moments of the demand distribution. However, the closed-form solutions to these models are not available. The authors propose to solve those using linear programming for sub-problems.

One important condition for application of the classic newsvendor framework is that the staffing distribution function must be stable and applicable both for the current and for the future time periods. This requires forecasting the demand which is itself a separate challenging problem. Usually, the demand forecast generated, for example, by a time series analysis can be reasonably accurate only for about a week or so. Beyond that time horizon, the forecasting errors are similar to that coming simply from the historical demand distributions (Davis et al., 2014). It is quite restrictive to assume that the future demand distribution is well known. Therefore, there have been some efforts to relax this assumption. One direction has been the nonparametric (data-driven) approach. In this approach, independent past demand data are used to estimate the minimal number of ob-

servations required for the estimated solution to be near-optimal. Rudin et al. (2013) applied this approach to estimate the optimal nursing staffing in a hospital emergency room. The authors investigated the newsvendor problem with past demand observations and a number of features to predict the future demand, such as the day of the week, time of the day and the number of past demands. This setting is similar to He et al. (2012) who considered staffing in a hospital operating room with two features (number and type of cases) to predict the required OR time.

However, selecting the right features is not easy and requires multiple models' cross-validations for testing the predictive power of the chosen features. Regardless of the sophistication of the prediction technique, if a sudden rare one-time event occurs such as mass casualty or a large epidemic outbreak that could not be previously accounted for, such an event would result in significant bias in the optimal staffing level.

There are several software products commercially available that facilitate daily staffing planning and management, such as McKesson ANSOS One-Staff, Kronos Workforce Scheduler, API/GE Healthcare Staffing & Scheduling, Cerner Clairvia Care Value Management Scheduling, among others. None of them offer predictive data analytics capability that foretells the optimum staffing level and the costs of deviation from that level that the newsvendor framework is providing.

Note that no popular nowadays machine learning algorithms will improve the solution proven as mathematically optimal. The term 'optimal' means that the staffing solution provides the minimal possible total cost of under- and overstaffing occurrences compared to any other staffing values within these time periods.

In the next sections, examples are provided in which the optimal workforce staffing solutions are determined using the newsvendor framework. In these examples, empirical staffing cumulative distribution functions (ECDF) are used, and Equation 1 is solved numerically assuming that the staffing demand is a random sample from a stable

but unknown distribution. Numerical solutions are produced using a computational procedure that was built in the excel spreadsheet. In contrast to the previous research, no best fit theoretical staffing distribution functions, or distribution moments or any other special information on the staffing distribution is needed.

MAIN FOCUS

Method

Implementation of the newsvendor framework includes the following steps: (i) variable patient census is converted into the corresponding variable staffing demand for a specified time period, (ii) the staffing demand is converted into the staffing empirical cumulative distribution function (ECDF), (iii) the cost ratio is calculated $r = Cu / (Cu + Co)$ using the hourly costs of under- (Cu) and overstaffing (Co) per nurse, and (iv) steps (ii) and (iii) are combined to solve Equation 1 numerically using a computational procedure described below.

Conversion Variable Patient Census Into Variable Staffing Demand

Conversion the variable patient census into variable nursing staffing can be performed using: 1) established patient-to-nurse ratio (PNR) that takes or does not take into account the patient acuity levels, or 2) hours per patient day benchmark standards or the calculated hours of care from the patient classification system (PCS), or 3) a staffing grid indicating for each census range the number of each caregiver skill for each shift.

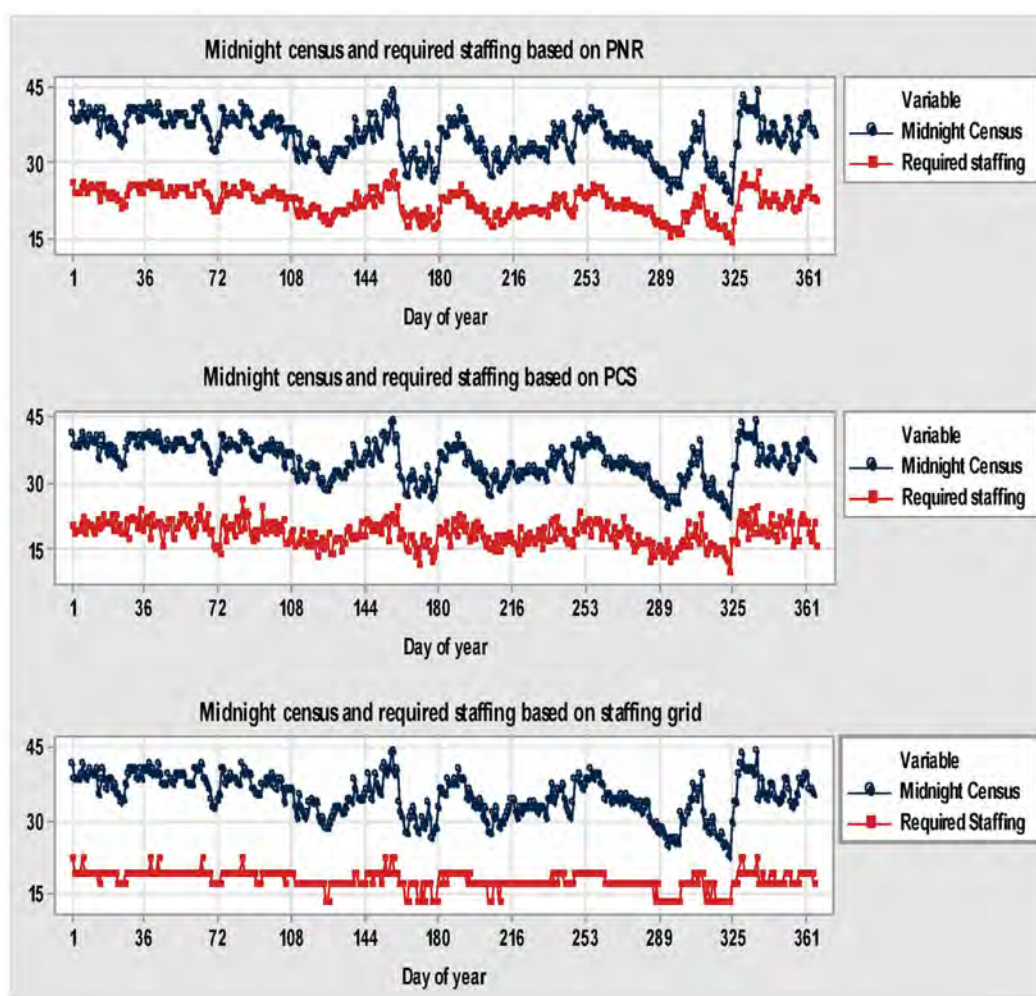
1. **The Use of Patient-To-Nurse Ratio (PNR):** To illustrate, daily midnight census data for a medical-surgical unit is presented on Figure 1 for the entire year to account for some seasonal census variation. Three levels of PNRs were established in this unit. Typically,

28% of patients with the high acuity level require PNR 1:1; 67% of patients with the medium acuity level require PNR 2:1, and 5% of patients with low acuity level require PNR 3:1 (These percentages vary seasonally but are kept fixed here for simplicity. Other nursing units can have different acuity percentages and PNRs). Each census value was multiplied by the weighted sum of inverted PNR values to convert it to the corresponding staffing, i.e. $\text{staffing} = \text{census} * (0.28/1 + 0.67/2 + 0.05/3)$. Results are presented on Figure 1 (upper panel). Note

that dividing each census value by the average $\text{PNR} = (0.28*1 + 0.67*2 + 0.05*3) = 1.77$ is not correct because dividing by the average is a non-linear operation that produces a biased staffing value. (Dividing census by PNR would be correct if PNR is an exact number, not the average).

2. **The Use of Patient Classification System (PCS):** In general, PCS is a system that quantifies categories of care in order to estimate the required nursing hours for direct patient care (Malloch et al., 2013). PCS-generated hours of care for each patient are summed

Figure 1. Annual daily census and the corresponding staffing



up, and the sum is then divided by the shift length hours. This gives the required staffing for the particular period. An example of the variable staffing generated by the PCS is presented, for comparison, on Figure 1 for the same daily midnight census used for the above PNR example (middle panel).

3. **The Use of the Staffing Grid:** A staffing grid reflects the perceived nursing workload by the nursing management. It is usually budget target based, and serve as a baseline guide. The skills and competency of the staff is also a consideration. An example of a simplified staffing grid for a medical-surgical unit is presented in Table 1.

An example of the variable staffing generated by this grid is presented, for the comparison, on Figure 1 for the same daily midnight census used for the above examples (bottom panel). Notice that the required staffing generated by the grid is less sensitive to census variation because the staffing levels stay constant within each patient census range.

Conversion of Staffing Demand Into the Empirical Cumulative Distribution Function (ECDF)

If a random set of n data points $S_1, S_2, \dots, S_n$ contains m data points that are less than or equal to some value S_k , then ECDF $F_k(S)$ is defined as the fraction m/n . The ECDF is calculated using the rule:

$$F_k(S) = \begin{cases} 0, & \text{if } S_k < S_{min} \\ m/n, & \text{if } S_{min} \leq S_k \leq S_{max}, k = 1, 2, \dots \\ 1, & \text{if } S_k > S_{max} \end{cases}$$

To illustrate, consider required staffing values presented on Figure 1 (upper panel). They range from $S_{min}=14$ to $S_{max}=28$. The sample size is $n = 364$ data points. The ECDF for these values is presented on Figure 2. The ECDF is not usually a smooth analytic function. In general, ECDF poorly fits the common theoretical distributions.

The Hourly Costs of Under- and Overstaffing Per Nurse and the Cost Ratio

If too few nurses are planned and staffed vs. the actual nursing demand (understaffing case), then the additional nurses can be called on from the internal float pool at no extra cost (if a trained nurse is available), or from the off-duty nursing pool, or a staffing agency at a premium above the base pay rate.

To illustrate, suppose that a trained nurse is available in the internal pool 50% of time. Hence, a nurse from a staffing agency should be called on 50% of time. The premium above the base pay is 150%. Thus, the additional cost per called-on nurse in excess to the base pay rate

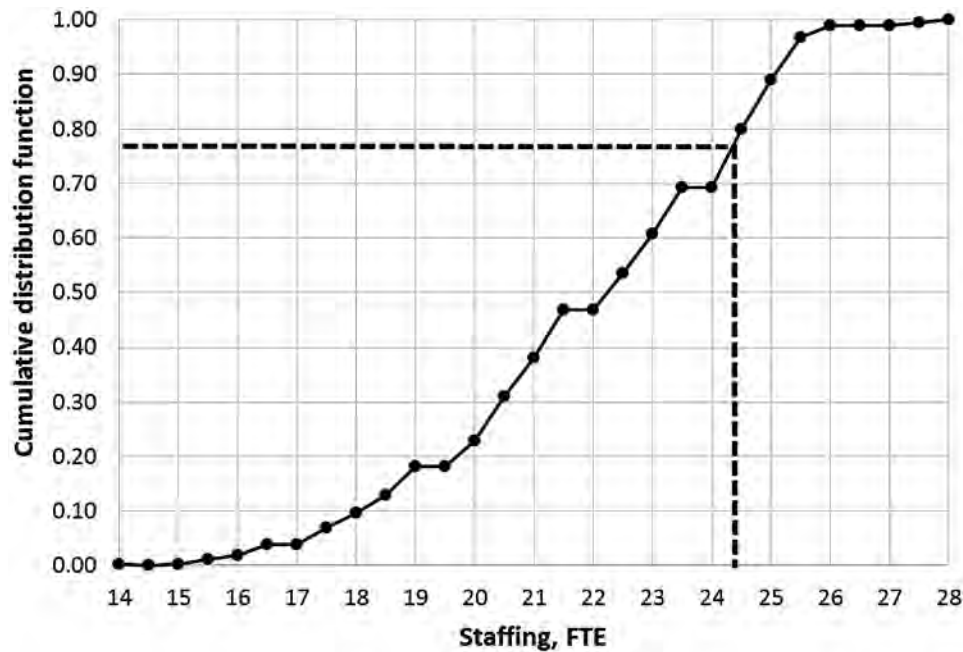
$$P(\$/hour) \text{ is } Cu = (1-0.5) * 1.5 * P(\$/hour) = 0.75 * P$$

If too many nurses are staffed (overstaffing), then the extra nurses can be called-off either into the internal float nursing pool or offered to take paid/unpaid vacation or put on call. If the nurse is put on call then she/he is not usually get payed (in majority USA hospitals). This makes an impression that calling-off a nurse does not cost anything (except the case for contractual obligation to pay

Table 1. A simplified staffing grid for the Medical-Surgical unit

Patient Census Range	Nursing Staffing
<24	10
24-28	13
29-35	17
36-40	19
41-44	21
>44	24

Figure 2. ECDF for variable staffing: The dotted lines illustrate the graphical solution of Equation 1



for putting on call, if any). Hence, the unit always has a financial incentive to staff to the maximum. While this might be true for the unit's base payroll, there is a fixed cost for the called-off nurse. KPMG reported (KPMG, 2011) that the average nursing hourly costs in the USA is about $P=\$45/\text{hour}$. This includes the base payroll cost of $\$35/\text{hour}$, and the total of insurance, recruiting, training and other costs per nurse of about $\$10\text{-}11/\text{hour}$, on average. Therefore, even if the called-off nurse is not get paid through the unit's payroll, she/he still costs the hospital about 22%-24% of the base pay, i.e. $0.24 \cdot P(\$/\text{hour})$, but brings no patient care value. Using this fixed cost of overstaffing, the cost ratio is $r=0.75/(0.75+0.24)=0.76$. Notice that the base pay rate is cancelled out in this ratio.

In most cases, the understaffing costs more than the overstaffing. Therefore, usually $r>0.5$. In rare cases of the same under- and overstaffing costs, $r=0.5$.

Note that an approach was suggested in which the cost parameters C_o and C_u were imputed from the observed actual staffing rather than treated as

the model inputs (Olivares et al., 2008). However, this approach assumes that the actual staffing was already close to the optimal. Otherwise, the imputed cost parameters will be significantly biased.

Computational Procedure for Solving Equation 1

Solution of Equation 1 is the staffing value, S , for which $F(S)$ is equal to r . It is illustrated graphically on Figure 2 by dotted lines for $r=0.76$. The optimal staffing solution is 24.5. This value can be fractional because, in general, staffing is expressed by FTE (full-time-equivalent) which is not necessarily a whole number. One FTE is usually 40 hours per week. For staffing software, Equation 1 should be solved numerically. A simple computational procedure for doing so is the following:

1. Make the FTE bins from minimal to maximal staffing values with the increment 0.25 FTE. (From a practical standpoint it is usually enough to get the staffing solution accuracy

+/- 0.125 FTE. Hence, 0.25 FTE staffing increment is usually suffice).

2. For each incremented FTE value S test if the corresponding ECDF value $F(S) < r$. If this condition is TRUE then go to the next value of $F(S)$, and so on.
3. 3. If $F(S) \geq r$ then STOP. The optimal staffing will approximately be the mean value

$$S_{opt} = [\text{FTE (previous step bin)} + \text{FTE (current step bin)}] / 2.$$

This procedure can easily be implemented in the Excel spreadsheet, as well as coded in R, Python, Microsoft.NET, or in any other coding language.

SOLUTIONS AND RECOMMENDATIONS

In this section, the optimal nursing staffing solutions for some nursing units are presented. The particular hospitals are not named here because some data used in the examples are considered proprietary.

Example 1. Annual and Monthly Optimal Nursing Staffing

The daily variable required nursing staffing for the medical-surgical unit is presented on Figure 3 (panel A), as well as the corresponding ECDF (panel B).

Figure 3. The daily nursing staffing demand for the medical-surgical unit (panel A) and the corresponding ECDF (panel B)

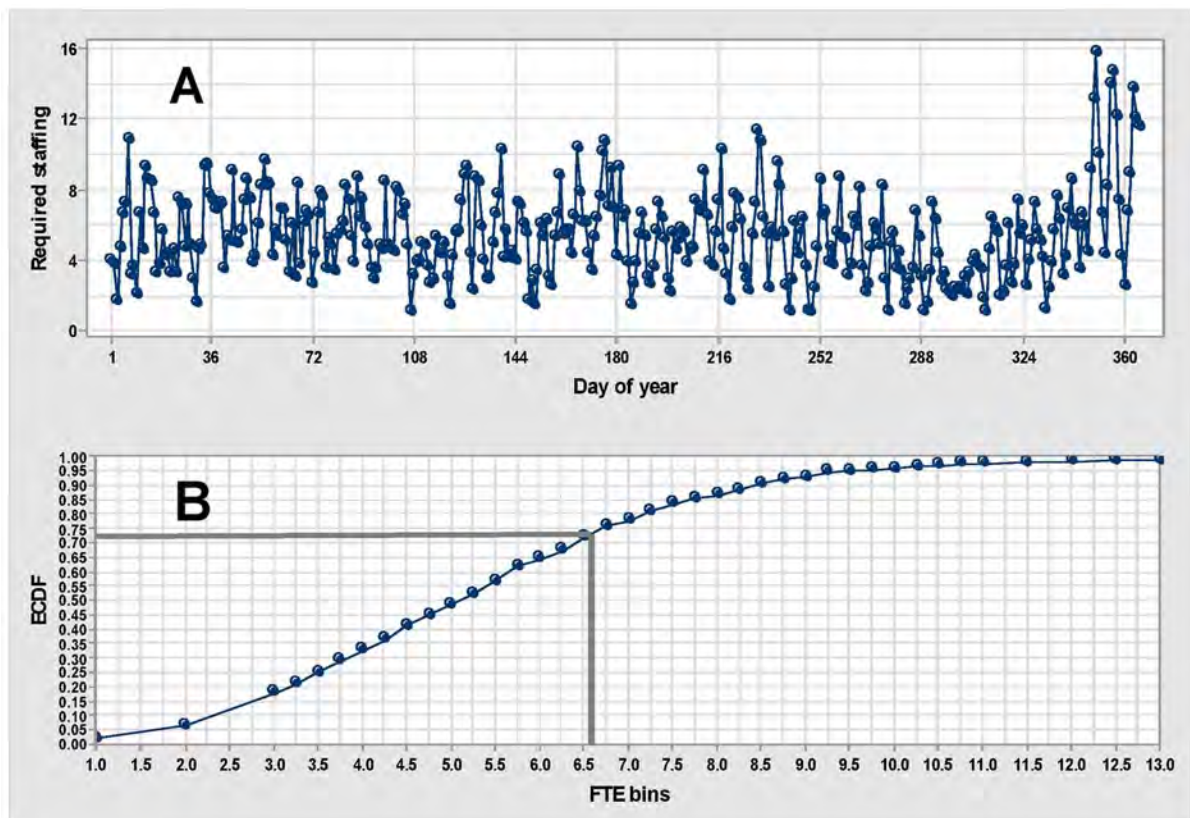


Table 2. FTE bins, ECDF, and the optimal annual nursing staffing level

Staffing FTE Bins	ECDF, $F(S)$	Optimal Staffing, FTE
6.25	0.674	
6.50	0.715	
6.75	0.759	6.625
7.00	0.775	
7.25	0.811	

Using the unit's available financial data, the extra cost of understaffing was assumed \$30/hour per nurse, while the overstaffing per nurse was assumed being the fixed rate \$11/hour. This makes $r=0.73$. The staffing bins and ECDF values are presented in Table 2 (abbreviated to save space). It is seen that $r=0.73$ lies in the interval between ECDF values 0.715 and 0.759. Thus, the annual optimal nursing staffing level is the mean of the corresponding FTE bins, i.e. $(6.5+6.75)/2=6.625$ FTE, as shown in the third column in Table 2. The graphical solution is indicated by the dotted lines on Figure 3 (panel B).

It was previously pointed out that the news vendor's optimal staffing level always corresponds to the minimal possible total cost of under- and overstaffing. Thus, any other heuristically adjusted (higher or lower) staffing level would inevitably

result in a higher total cost of deviations from that level. To illustrate, let us compare two costs of deviation using (i) the optimal staffing value from Table 2 and (ii) the suboptimal average staffing for the same period. The average value is frequently used as a basis in traditional staffing planning.

The total annual cost of deviations of the daily required staffing FTE from the optimal 6.6 FTE (above or below this level) was calculated as the sum of these deviations multiplied by the corresponding hourly costs and by the shift duration 8 hours. The result was \$94,979.

The average annual staffing for data presented on Figure 3 (panel A) was 5.3 FTE. The total annual cost of deviations for this suboptimal level calculated similarly was \$115,012. Thus, the optimal staffing level resulted in about 17% lower costs compared to the average one.

It was mentioned earlier that staffing plans are frequently developed for shorter time periods such as monthly or every 4-6 weeks. To illustrate, the month-to-month optimal staffing levels were calculated using the same computational procedure. Results are presented on Figure 4, along with the annual optimal staffing level (flat solid line) for the comparison.

It is seen a significant month-to-month variation of the optimal FTE vs. the annual one. The bump of the optimal FTE for December reflects

Figure 4. The optimal monthly staffing and the annual optimal staffing level (flat solid line)

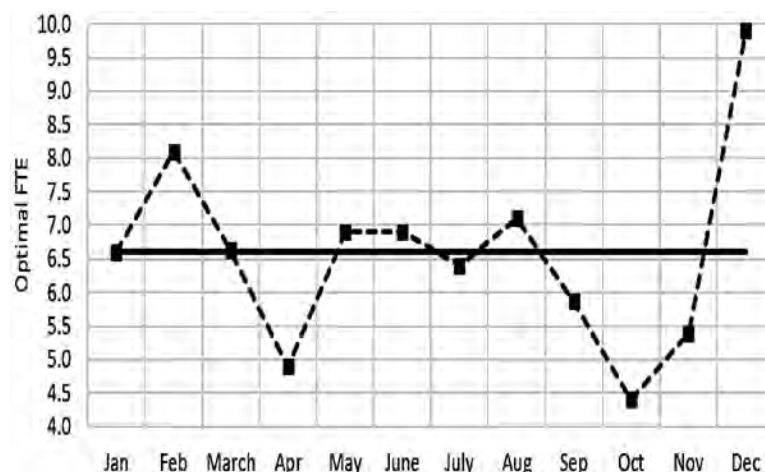


Table 3. Monthly optimal and average FTE and the total costs of under- and overstaffing

Month	Optimal FTE	Monthly Average FTE	Total Costs of Under- and Overstaffing		
			For Optimal FTE	For Average FTE	Cost Saving, %
Jan	6.6	5.1	\$ 7,766	\$ 9,697	20%
Feb	8.1	6.5	\$ 5,200	\$ 7,486	31%
Mar	6.6	5.7	\$ 4,988	\$ 7,094	30%
Apr	4.9	4.6	\$ 5,093	\$ 5,893	14%
May	6.9	5.3	\$ 7,870	\$ 9,887	20%
Jun	6.9	6.2	\$ 7,165	\$ 8,167	12%
Jul	6.4	5.0	\$ 5,104	\$ 7,232	29%
Aug	7.1	5.5	\$ 8,924	\$ 10,670	16%
Sep	5.9	4.7	\$ 5,545	\$ 7,477	26%
Oct	4.4	3.4	\$ 6,511	\$ 7,195	10%
Nov	5.4	4.0	\$ 5,087	\$ 6,749	25%
Dec	9.9	8.0	\$ 13,625	\$ 15,351	11%

the corresponding staffing demand peak for this time period (Figure 3, panel A), apparently due to increased patient census related to typical end-of-year seasonal flu epidemics. Thus, the shorter term month-to-month optimal staffing is typically better accounts for seasonal patient census and required staffing variations.

The total costs of under- and overstaffing for the monthly optimal staffing FTE were compared to that for the average staffing FTE in Table 3.

Example 2. The Optimal Staffing for Caregivers' Skills Mix

Previous examples demonstrated the optimal staffing calculation only for one category of caregivers-the nurses. However most nursing units include other staff skills mix, such as nursing assistants (NA), clinical nurse specialists (CNS), nurse practitioners (NP), among others.

Let's consider, as an example, a nursing unit that is staffed with three caregivers' skills mix: nursing assistants (NA), registered nurses (RN) and certified nurse specialists (CNS). In this unit a typical share of workload by the skills mix is 28% for NA, 57% for RN and 15% for CNS. Standard

hours of care per patient day (HPPD) for various levels of patient acuity are presented in Table 4.

Patient data for this unit are presented in Figure 5 (only three weeks' worth of data is included to save space).

The typical cost of under- and overstaffing for RN in this unit was similar to the one used in the previous example. However, the costs for NA and CNS were not readily available. It was assumed that the costs of under- and overstaffing for NA were about the same (\$13-\$15/hour) resulting in the NA cost ratio $r=0.5$. The understaffing costs for CNS was assumed twice the cost of overstaffing

Table 4. Standard hours of care per patient for various levels of patient acuity

Acuity Level	Standard Hours of Care
Level 03	2.8
Level 04	3.4
Level 05	4.0
Level 06	4.6
Level 07	5.5
Level 08	6.6
Admissions	0.5
Discharges	0.5

resulting in $r=0.67$. At the same time, the calculated optimal staffing values for these caregiver categories turned out not too sensitive to the r values. Due to the particular staffing demand and, hence, the shape of ECDF, the NA optimal staffing is not changed with r values in the range $r=0.43-0.56$. CNS optimal staffing is not changed with r values in even wider range $r=0.56-0.75$. Thus, some variations of the poorly known costs were not too critical for the calculated optimal staffing values presented in Table 5, which also includes the total costs of under- and overstaffing for three categories of caregivers both for the optimal FTE and for the suboptimal average FTE. It is seen that the costs of the former is much lower than that for the latter, similar to results demonstrated in Example 1.

In summary, these examples demonstrated that the optimal staffing levels vary with the time periods for which they are calculated (annual,

monthly, or any other). They also demonstrated that the mathematically proved minimal total costs of under- and overstaffing for the optimal staffing values are indeed less than that for any suboptimal staffing levels such as the average. Another goal of these examples was demonstrating the use of ECDF and a simple computational procedure for numerical solving the newsvendor Equation 1. It is recommended that the newsvendor framework is applied to justify a staffing plan whenever it is being developed and budgeted for.

FUTURE RESEARCH DIRECTIONS

The author believes that the following problems merit further research.

Determining the optimal staffing for the future time periods with some projected gradual non-random trend in demand. This trend should be

Figure 5. The number of patients at various levels of acuity, admissions & discharges (ADT) and the daily required staffing FTE for three categories of caregivers NA, RN, and CNS

Date	Level 03	Level 04	Level 05	Level 06	Level 07	Admissions & Discharges (ADT)	Total hours of care	NA share of total hours of care (28 %)	RN share of total hours of care (57%)	CNS share of total hours of care (15%)	NA staffing, FTE	RN staffing, FTE	CNS staffing, FTE
1/1/12			1	16	3		93.9	26.8	53.7	13.4	3.3	6.7	1.7
1/2/12			1	17	4	3	107.0	30.5	61.2	15.3	3.8	7.6	1.9
1/3/12			3	11	7	6	107.3	30.6	61.3	15.3	3.8	7.7	1.9
1/4/12			1	21	5	7	134.9	38.5	77.1	19.3	4.8	9.6	2.4
1/5/12	2			22	3	11	133.9	38.2	76.5	19.1	4.8	9.6	2.4
1/6/12				23	5	4	137.0	39.1	78.3	19.6	4.9	9.8	2.4
1/7/12				23	4	7	134.5	38.3	76.8	19.2	4.8	9.6	2.4
1/8/12		1		18	4	7	115.0	32.8	65.7	16.4	4.1	8.2	2.1
1/9/12				18	4	3	107.6	30.7	61.5	15.4	3.8	7.7	1.9
1/10/12				14	9	2	116.0	33.1	66.3	16.6	4.1	8.3	2.1
1/11/12				23	6	7	145.5	41.5	83.2	20.8	5.2	10.4	2.6
1/12/12				20	3	1	109.2	31.1	62.4	15.6	3.9	7.8	2.0
1/13/12			1	21	4	7	129.3	36.9	73.9	18.5	4.6	9.2	2.3
1/14/12			1	20	4	2	119.8	34.1	68.4	17.1	4.3	8.6	2.1
1/15/12				17	3		94.4	26.9	54.0	13.5	3.4	6.7	1.7
1/16/12			5	11	1	5	81.1	23.1	46.3	11.6	2.9	5.8	1.4
1/17/12			1	25	1	5	129.0	36.8	73.7	18.4	4.6	9.2	2.3
1/18/12			1	26	2	8	142.1	40.5	81.2	20.3	5.1	10.2	2.5
1/19/12		1		28	3	9	157.2	44.8	89.8	22.5	5.6	11.2	2.8
1/20/12		1		14	5	7	102.3	29.2	58.4	14.6	3.6	7.3	1.8
1/21/12				14	7	4	106.9	30.5	61.1	15.3	3.8	7.6	1.9

Table 5. Optimal and average staffing FTE for three caregiver categories and the total costs of under- and overstaffing

Caregiver Category	Optimal FTE	Average FTE	Total Cost of Under- and Overstaffing	
			for Optimal FTE	for Average FTE
Nursing Assistant (NA)	4.4	4.3	\$20,705	\$25,681
Registered Nurse (RN)	9.9	8.6	\$49,701	\$75,245
Certified Nursing Specialist (CNS)	2.4	2.2	\$8,414	\$15,146

incorporated into the current ECDF to account for some shift in the future staffing demand distribution.

A related problem is incorporating sudden demand spikes into the prior staffing demand distribution, e.g., caused by a mass casualty event or an infection outbreak. One approach for doing this could be the use of the Bayesian framework.

One more problem is related to the true costs of occurrences of under- and overstaffing. The called-on float or agency nurses are usually new to the unit, thus they are less productive than the regular staff nurses. Similarly, calling-off nurses by putting them on call or floating into the pool after scheduling up for work usually negatively impacts their morale and productivity. These non-monetary factors could result in underestimating the true costs of under- and overstaffing. It seems that this problem was not adequately addressed in prior research.

Also, the classic newsvendor framework assumes that the costs of under- and overstaffing are constant within the time period. However, these costs could be interrelated with staffing demand. Incorporating this interrelation into the newsvendor framework will further advance its area of application in healthcare settings.

CONCLUSION

The author believes that this chapter convincingly demonstrates the power and versatility of the newsvendor framework to determine the optimal

staffing levels in healthcare settings. In fact, the newsvendor framework is suitable for analysis of various operations management problems. It is also demonstrated the advantages of the ECDF and a simple computational procedure for numerical solving the newsvendor Equation 1. Using this approach, there is no need anymore for developing the best fit theoretical demand distribution or moments of the distribution suggested in earlier research. The author expects that more newsvendor model based operations management applications will emerge in healthcare settings, along with other appropriate data analytics methodologies.

REFERENCES

- Arikan, E. (2011). *Single Period Inventory Control and Pricing: An Empirical and Analytical Study of a Generalized Model*. Peter Publishing, Inc.
- Chen, R. R., Cheng, T. C. E., Choi, T. M., & Wang, Y. (2016). Novel Advances in Applications of the Newsvendor Model. *Decision Sciences*, 47(1), 8–10. doi:10.1111/deci.12215
- Choi, T. M. (Ed.). (2012). *Handbook of Newsvendor Problems: Models, Extensions and Applications*. Springer.
- Davis, A., Mehrotra, S., Holl, J., Daskin, M., (2014). Nurse staffing under demand uncertainty to reduce cost and enhance patient safety. *Asia-Pacific Journal of Operational Research*, 31(1).

He, B., Dexter, F., Macario, A., & Zenios, S. (2011). The timing of staffing decisions in hospital operating rooms: Incorporating workload heterogeneity into the newsvendor problem. *Manufacturing & Service Operations Management*, 14(1), 99–114. doi:10.1287/msom.1110.0350

Herman, B. (2013). Ten statistics on hospital labor costs as a percentage of operating revenue. *Becker's Hospital Review: Dec. 10*. Retrieved from <http://www.beckershospitalreview.com/racs/-icd-9/-icd-10/10-statistics-on-hospital-labor-costs-as-a-percentage-of-operating-revenue.html>

Hopp, W., & Lovejoy, W. (2013). *Hospital Operations: Principles of high efficiency health care*. New York: FT Press.

KPMG's. (2011). *US Hospital Nursing Labor Costs Study (2011)*. Retrieved from http://www.natho.org/pdfs/KPMG_2011_Nursing_Labor-CostStudy.pdf

Malloch, K., & Meisel, M. (2013). Patient Classification System: State of the Science. *Nurse Leader*, 11(6), 35–37. doi:10.1016/j.mnl.2013.09.008

Olivares, M., Terwiesch, C., & Cassorla, L. (2008). Structural Estimation of the Newsvendor Model: An Application to Reserving Operating Room Time. *Management Science*, 54(1), 41–45. doi:10.1287/mnsc.1070.0756

Perakis, G., & Roels, G. (2008). Regret in the Newsvendor Model with Partial Information. *Operations Research*, 56(1), 188–203. doi:10.1287/opre.1070.0486

Porteus, E. (2002). *Foundations of Stochastic Inventory Theory*. Stanford University Press.

Qin, Y., Wang, R., Vakharia, A. J., Chen, Y., & Seref, M. M. H. (2011). The newsvendor problem: Review and directions for future research. *European Journal of Operational Research*, 213(2), 361–374. doi:10.1016/j.ejor.2010.11.024

Rudin, C., & Vahn, G-Y. (2013). *The Big Data Newsvendor: Practical Insights from Machine Learning Analysis*. MIT Sloan School Working Paper 5036-13. Retrieved from MIT DSpace Electronic Paper Collection: <http://hdl.handle.net/1721.1/81412>

ADDITIONAL READING

Halman, N., Orlin, J., & Simch-Levi, D. (2012). Approximating the Nonlinear Newsvendor and Single-Item Stochastic Lot-Sizing Problems When Data Is Given by an Oracle. *Operations Research*, 60(2), 429–446. doi:10.1287/opre.1110.1031

Khanra, A., Soman, C., & Bandyopadhyay, T. (2014). Sensitivity Analysis of the Newsvendor Model. *European Journal of Operational Research*, 239(2), 403–412. doi:10.1016/j.ejor.2014.05.039

Lehtonen, J.-M., Torrkki, P., Peltokorpi, A., & Moilanen, T. (2013). Increasing operating room productivity by duration categories and a newsvendor model. *International Journal of Health Care Quality Assurance*, 26(2), 80–92. doi:10.1108/09526861311297307 PMID:23534145

Levi, R., Perakis, G., & Uichanco, J. (2015). The Data-Driven Newsvendor Problem: New Bounds and Insights. *Operations Research*, 63(6), 1294–1306. doi:10.1287/opre.2015.1422

Scweitzer, M., & Cachon, G. (2000). Decision Bias in the Newsvendor Problem with a Known Demand Distribution: Experimental Evidence. *Management Science*, 46(3), 404–420. doi:10.1287/mnsc.46.3.404.12070

KEY TERMS AND DEFINITIONS

Data Analytics: The discipline and practice of using various quantitative methods to aid in solving business or engineering problems.

Empirical Cumulative Distribution Function (ECDF): A fraction of data-points that are less than or equal to some predetermined values within a set of random numbers.

Newsvendor Framework: An analytical model that provides the optimal inventory or staffing level (in terms of the lowest financial loss due to overage and underage) within a specified single time period.

Optimal Staffing: The staffing level that provides the lowest possible total costs of occurrences of under- and overstaffing compared to any other staffing value within a specified time period.

Overstaffing (Overage): Staffing level that is higher than the actually required.

Patient Classification System: The system that quantifies categories of care in order to objectively estimate the required nursing hours for direct patient care.

Understaffing (Underage): Staffing level that is lower than the actually required.

H

APPENDIX

Let the distribution density function of demand, D , be $f(D)$.

Then the average loss due to overage, i.e. for staffing $S > D$ is:

$$C_o \int_0^S (S - D) f(D) dD,$$

where C_o is the unit cost of overage.

The average loss due to underage, i.e. for staffing $S < D$ (unmet demand) is:

$$C_u \int_S^{D_{max}} (D - S) f(D) dD$$

where C_u is the unit cost of underage.

Total loss function is:

$$L(S) = C_o \int_0^S (S - D) f(D) dD + C_u \int_S^{D_{max}} (D - S) f(D) dD$$

It is minimized with respect to staffing, S , by taking the derivative and making it equal to 0.

$$\frac{dL}{dS} = C_o \int_0^S f(D) dD - C_u \int_S^{D_{max}} f(D) dD = 0$$

Here, the derivative of $L(S)$ with respect to S is taken using the general rule:

$$\text{If } f(p) = \int_{a(p)}^{b(p)} f(x, p) dx, \text{ then}$$

$$\frac{df}{dp} = \int_a^b \frac{\partial f(x, p)}{\partial p} dx + f(b, p) \frac{db}{dp} - f(a, p) \frac{da}{dp}$$

The 2-nd derivative is

$$\frac{d^2L}{dS^2} = C_o f(S) + C_u f(S) > 0 \text{ hence this is minimum.}$$

Thus, $C_o F(S) - C_u (1 - F(S)) = 0$. Consequently, $F(S) = C_u / (C_u + C_o)$.

The same result is produced if the demand distribution is a discrete function rather than the continuous one.

Using Technology to Reduce a Healthcare Disparity



Nilmini Wickramasinghe

Epworth HealthCare, Australia & Deakin University, Australia

INTRODUCTION

Healthcare costs globally continue to rise and the US in particular is struggling to contain healthcare costs (Wickramasinghe et al., 2015, 2012). This has led most OECD countries to focus on healthcare reform coupled with greater investment in IS/IT to facilitate superior healthcare delivery (Wickramasinghe et al., 2015). At the same time, the federal government has affected policy to emphasize meaningful use of such technology in healthcare (Wickramasinghe et al., 2015; “Meaningful Use,” 2012). Hence, it is now prudent to develop appropriate technology solutions that not only comply with this requirement but also facilitate superior healthcare delivery to ensue.

An area that can particularly benefit from the application of technology solutions is that of equal experience and trying to provide equality of care and access to all citizens. A key area within healthcare disparities relates to access of language services in healthcare or more specifically supporting limited English proficient patients (LEP patients). In particular, English language proficiency should not impinge on access to- and quality of- service for healthcare. Improving access to language services in healthcare has been an ongoing issue that continues to be at the forefront of various healthcare agendas ((Au et al., 2009; Barrett et al., 2008; Chen et al., 2007).

BACKGROUND

Recent discussions on healthcare disparities (Gibbons, 2011) all note the significant potential benefit

technology can make in trying to provide an equal experience to all Americans. Sadly, while there are many good points about the US healthcare system, there also exists a considerable racial and ethnic disparity in the delivery of healthcare across the US (Gibbons, 2011). The underlying root causes for these disparities are all amenable to interventions using IS/IT (information systems/information technology). The thesis of this paper is that technology is well suited to assist is that of limited English proficiency (LEP).

More than 23 million Americans today have limited English proficiency, which in turn has a negative impact on their ability to receive and comprehend appropriate healthcare delivery (Youdelman, 2008; Flores et al., 2008). Integral to the delivery of care is communication between doctor and patient; however, language barriers typically lead to problems such as delay or denial of services, issues with medication management, and underutilization of preventative services (Green et al., 2005; Jacobs et al., 2004; Ghandi et al., 2000; Karliner et al., 2004). The literature suggests that the quality of communication between Although Title VI of the Civil Rights Act 1964 has always required that entities receiving federal funds provide language services to those with LEP, the law has not often been enforced in healthcare settings (Jacobs et al., 2006). However, awareness of the need to provide language services in healthcare has increased in recent years (Gibbons, 2011).

Current Problem

In 2001 The Institute of Medicine has published two key reports “To err is Human” and “Cross-

DOI: 10.4018/978-1-5225-2255-3.ch323

ing the Quality Chasm”. Taken together these reports highlight that patient safety should be one of the essential components of high quality healthcare and that patients should not be harmed by the care that is intended to help them. These profound statements have had far reaching impacts to policy reform and efforts to address patient safety and quality of care delivered today. The role of language barriers and their impact on adverse events is thus now also receiving heightened attention. Especially given that research is consistently highlighting that adverse events affect LEP patients disproportionately more and result in serious consequences to the patient.

Approximately 57 Million people or more than 20 percent of the US population speak a language other than English and this figure is growing while approximately 8.6 percent of the population is defined as LEP. Thus at least 8.6 percent of the US population is at risk for adverse events because of barriers associated with language issues. This adds further cost pressures to an already strained healthcare system

To address this problem a technology mediated solution is proffered to provide multi-lingual support at in-take and registration for LEP patients.

Development of the Conceptual Model

Web 2.0 technology and cloud computing which provides and facilitates anytime anywhere access to (Troshani et al., 2011; Svantesson & Clark, 2010; Mell & Grance, 2010; Gilbert, 2010; Amazon, 2011; Armbrust et al., 2010), affords us the possibility to leverage these technology benefits to design and develop an appropriate solution to address the problem – a portal that support multi language real time translation at intake and registration. Figure 1 provides the conceptual model.

From the above conceptual model in Figure 1, it is possible to model the intake process into critical steps. All these 13 steps of care must be addressed in the technology solution and/or suite of solutions if the solution is to provide the necessary assistance for LEP patients. Taken together these steps traverse the healthcare encounter at intake.

Figure 1. Conceptual model

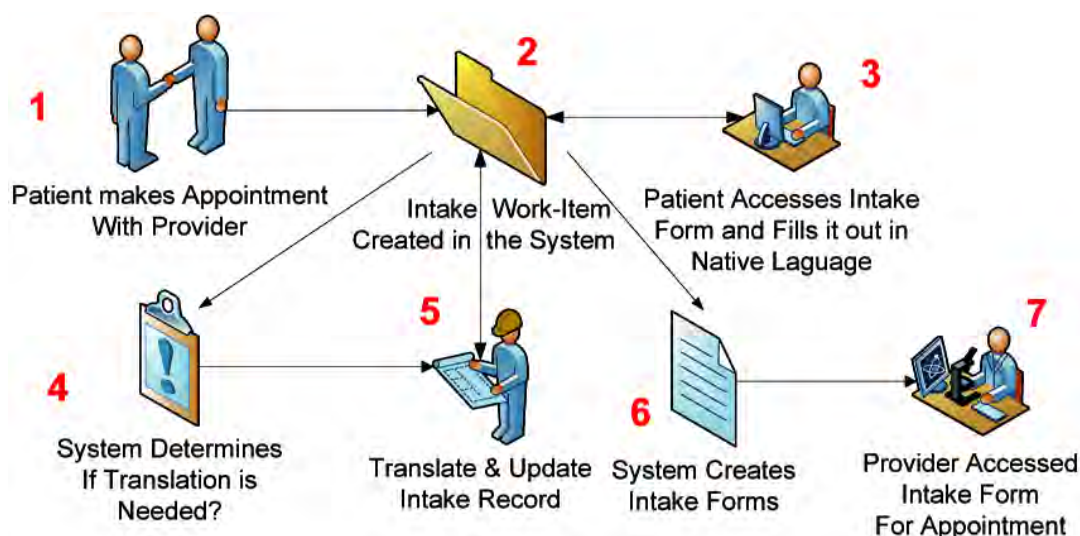


Table 1. Study activity

Actor	Before	During Study
Spanish Speaking Patients	Handed an English language intake form and were asked to fill it out to the best of their ability Often had a family member fill out the form, forms were often difficult to read and were incomplete If no family member was present a Spanish speaking staff administrator was taken off their job and asked to serve as interpreter and fill out the form with patient and increased the cost of serving an LEP patient Completed Forms were often difficult to read and of marginal value to physician provider Cultural behaviors limited the ability to gather information, Older Hispanics, men and women are very reluctant to volunteer information Completed form given to administration desk Completed form was placed in patient intake folder and was available to physician, was not available for other purposes Wait times varied as to availability of	Asked to participate in beta study, if agreeable directed to intake kiosk Offered Spanish language form which was translated into English per study description by medically competent interpreter. In some cases family member filled out form in other cases interpreter filled out form. No problems observed in using kiosk technology Form presentation was clear and understandable by patients Completed form given to administration desk Completed form in electronic format was available for other purposes such as Average duration of form completion action was not a factor
English Speaking Patients	Given paper form Filled out form to best of ability Very often completed forms were not easy to read and of minimal value to physician Completed form was placed in patient intake folder and available to physician, was not available for other purposes	Given option of using electronic intake kiosk with drop down menus and input boxes Faster intake form completion Intake form presentation was clear and understandable by patient Completed form was placed in patient intake folder and available to physician and available for other purposes
Physician	Clarity of intake form depended on writing and communication skill of patient	Electronic format provided easy to read and use intake form
Administration	Begins w patient request for appoint and scheduling of appointment, and request for completion of paper forms prior to actual appoint, on appointment date Patients rarely completed intake forms in advance of appointment If patient did not complete intake form prior to visit a blank form was presented to patient at time of visit If patient was a Spanish speaker a family member or staff would assist with intake completion If no family member available administration staff acted as interpreter. Completed form was used presented to administrator	Administration asked English or Spanish speaking patients if they would be interested in completing an intake form electronically and in native language. Once form was complete either handwritten or electronic it was returned to the administration and placed in the intake folder for clinical visit with physician.

PILOT STUDY

The developed solution was then tested in a small pilot study to assess its reliability and feasibility. A 2 arm control study was designed where both English proficient patients and Spanish proficient patients were randomly selected at a leading Mid-West Multi-specialty healthcare facility. Before the study commenced appropriate ethical clearances were obtained. Table 1 and 2 detail the results from the two week pilot trial as well as the activities that were observed in the study.

Results

The study qualitatively measured the interactions between the patients and the electronic kiosk, the value of a clear and concise intake summary for the physician at the beginning of the medical encounter and the ability of the administrative team in the department to take advantage of the completed electronic intake form for other purposes. There was no attempt to integrate the intake form with electronic medical records. Table 3 summarizes the key findings.

Table 2. Control: English speaking patients

Actor	Event	Outcome
English Speaking Patient	English speaking patients are offered option of traditional paper intake form or completing the form at kiosk with touch screen and drop down menu capabilities.	The Department of Urology serves an aged skewed population. As a general rule the younger the patient the more likely they were to opt for the electronic format. 80% of the patients offered the electronic option chose for to use that. Several patients commented that the electronic version was much faster to use and more consistent with a modern healthcare setting. Other inquired if the electronic form was transferable to other departments. For those patients that chose the electronic version patient satisfaction ranged on the scale of 4 or 5 with 5 being the highest on a 1 to 5 scale.
Spanish Speaking Patient	Spanish speaking patients are offered option of the traditional paper intake form or completing the form at kiosk with touch screen and drop down menu capabilities. The hand written intake form provided by the administration was not translated into Spanish, consequently each Spanish speaking patient opted for the electronic version of the form offered in their native language.	In each case at the electronic kiosk, translator services were incorporated to facilitate the completion of the intake form. Spanish speaking patients were afforded the opportunity to complete the intake in their language while the administration and physician were provided an English language version of the intake form. In each case customer satisfaction was rated a 5 on a 1 to 5 scale. Multiple copies of the completed intake form given to each patient for potential use in other departments.
Physician and Patient	English and Spanish speaking patients were queued to meet with the physician in a business as usual manner.	Physicians who had access to the electronic forms regardless of language spoken by the patient rated the clarity of the information on the electronic form higher than the hand written form. Also, physicians stated that the quantity of the information was greater and offered a more clear picture of the patient medical, family and social status. Physician satisfaction with the electronic form over the handwritten form was ranked higher because the electronic form was easier and faster to read. See appendix xx for sample of handwritten and electronic form.
Administration and Patient	Once the intake forms were completed (English and Spanish), they were handed to the front desk for administration, placed in the patient intake folder and placed in queue for the physician. The form is reviewed by the physician. It is generally not used beyond physician review for other purposes.	After the physician encounter patients were often asked to complete other forms very similar in content to the original intake form. Form redundancy is a major detractor to patient satisfaction and high cost component to administrative efficiency. No other forms have a language assistance component. The impact of this is obvious i.e. when a dietary requirement is needed before a test can be conducted and this requirement is not met on the day of the scheduled procedure because the instructions were not understandable by the non-English speaking patient.

FUTURE RESEARCH DIRECTIONS

Two areas are especially critical when considering future developments of the technology mediated solution. Specifically, one area is to assess the health literacy of the patients no matter if they are LEP or proficient English language speakers. The other key direction involves developing and enlarging the languages offered to include Chinese, polish, Farci and German as these are the foreign languages most in prevalent within the US.

CONCLUSION

The US healthcare provider setting is working to deliver the best possible patient care. It must also meet cost of service objectives and provide physicians the best possible tools to meet patient's expectations. Technology in its many applications will continually be employed and adapted to accomplish these goals. Technology that facilitates communication between the patient and the provider (physician and administration) will have an important role. All healthcare interactions begin

Table 3. Summary of findings

Actor	Event	Outcome
English Speaking Patient	English speaking patients are offered option of traditional paper intake form or completing the form at kiosk with touch screen and drop down menu capabilities.	The Department of Urology serves an aged skewed population. As a general rule the younger the patient the more likely they were to opt for the electronic format. 80% of the patients offered the electronic option chose for to use that. Several patients commented that the electronic version was much faster to use and more consistent with a modern healthcare setting. Other inquired if the electronic form was transferable to other departments. For those patients that chose the electronic version patient satisfaction ranged on the scale of 4 or 5 with 5 being the highest on a 1 to 5 scale.
Spanish Speaking Patient	Spanish speaking patients are offered option of the traditional paper intake form or completing the form at kiosk with touch screen and drop down menu capabilities. The hand written intake form provided by the administration was not translated into Spanish, consequently each Spanish speaking patient opted for the electronic version of the form offered in their native language.	In each case at the electronic kiosk, translator services were incorporated to facilitate the completion of the intake form. Spanish speaking patients were afforded the opportunity to complete the intake in their language while the administration and physician were provided an English language version of the intake form. In each case customer satisfaction was rated a 5 on a 1 to 5 scale. Multiple copies of the completed intake form given to each patient for potential use in other departments.
Physician and Patient	English and Spanish speaking patients were queued to meet with the physician in a business as usual manner.	Physicians who had access to the electronic forms regardless of language spoken by the patient rated the clarity of the information on the electronic form higher than the hand written form. Also, physicians stated that the quantity of the information was greater and offered a more clear picture of the patient medical, family and social status. Physician satisfaction with the electronic form over the handwritten form was ranked higher because the electronic form was easier and faster to read. See appendix xx for sample of handwritten and electronic form.
Administration and Patient	Once the intake forms were completed (English and Spanish), they were handed to the front desk for administration, placed in the patient intake folder and placed in queue for the physician. The form is reviewed by the physician. It is generally not used beyond physician review for other purposes.	After the physician encounter patients were often asked to complete other forms very similar in content to the original intake form. Form redundancy is a major detractor to patient satisfaction and high cost component to administrative efficiency. No other forms have a language assistance component. The impact of this is obvious i.e. when a dietary requirements is needed before a test can be conducted and this requirement is not met on the day of the scheduled procedure because the instructions were not understandable by the non-English speaking patient.

and end with communication between the provider and patient. This is a determining factor in successful health outcomes. With greater clarity and availability of patient information there is a greater probability of meeting patient need and managing the cost of delivering service. For example, it is well documented that language barriers (communication barriers) between patients and providers are a leading cause of healthcare disparities. Communication barriers are not limited to the interactions between the patient and physician

but also include the patient's interactions with the physician supporting healthcare infrastructure.

Communication Barriers Impact

Patient - Physician Communication

For example,

1. The inability for physicians to understand a patient's need and thus increase the potential

for medical errors and the associated higher cost of service.

2. Poor communications between the physician and patient extends the healthcare encounter and creates delays in physician access.

Patient – Administration Communication

For example,

1. The ability for the patient to follow with procedural or pharmacy instructions,
2. The ability for patients to provide or understand billing or insurance information, and
3. The ability to follow up with next steps in a treatment protocol which leads to higher cost of service

This study represents a first step to understand the impact of using an electronic intake format for gathering patient information from both English speaking and non-English speaking (LEP) patients. Healthcare disparities plague the US healthcare system and language proficiency is one of these key barriers. Technology does have the potential to reduce this barrier effectively and efficiently so that superior care delivery might ensue. The proffered solution described in the preceding holds the key to potentially addressing this challenge.

REFERENCES

- Amazon. (2011). *Announcing Amazon SimpleDB in the EU Region*. Available: <http://aws.amazon.com/about-aws/whats-new/2009/09/22/announcing-amazon-simplydb-in-the-eu-region/>
- Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., & Konwinski, A. et al.. (2010). A view of cloud computing. *Communications of the ACM*, 53(4), 50–58. doi:10.1145/1721654.1721672
- Au, M., Taylor, E., & Gold, M. (2009). *Improving Access to Language Services in Health Care. Policy Brief Mathematica*. Policy Research Inc. White paper.
- Barrett, S., Dyer, C., & Westpheling, K. (2008). *Language Access: Understanding the Barriers and Challenges in Primary care Settings*. McLean, VA: The Association of Clinicians for the Underserved.
- Chen, A., Youdelman, M., & Brooks, J. (2007). The Legal Framework for Language Access in Healthcare Settings: Title VI and Beyond. *Journal of General Internal Medicine*, 22(S2), 362–367. doi:10.1007/s11606-007-0366-2 PMID:17957427
- DrTango Inc. (2011). *A Comparison of DMHC and CDISB853 Requiriementns*. Available online at <http://www.sb853.com/pdf/regulations-Table.pdf>
- Dutta, S., & Mia, I. (2011). *The Global Information Technology Report 2010–2011*. Geneva: World Economic Forum and INSEAD.
- Etro, F. (2010). The economic consequences of the diffusion of cloud computing. In S. Dutta & I. Mia (Eds.), *The Global Information Technology Report 2009-2010*. Geneva: World Economic Forum and INSEAD.
- Flores, G. (2005). The Impact of Medical Interpreter Services on the Quality of Health Care: A Systematic Review. *Medical Care Research and Review*, 62(3), 255–299. doi:10.1177/1077558705275416 PMID:15894705
- Flores, G., Torres, S., Holmes, L., Salas-Lopez, D., Youdelman, M., & Tomany-Kprman, S. (2008). access to Hospital Interpreter Services for Linted English Proficnet Patietsns in New Jersey: A Statewide Evaluation. *Journal of Helathcare for the Poor and the Underserved*, 19(2), 391–415. doi:10.1353/hpu.0.0007 PMID:18469412

Gandhi, T., Burstin, H., Cook, E., Puopolo, A., Haas, J., Brennan, T., & Bates, D. (2000). Drug complications in outpatients. *Journal of General Internal Medicine*, 20(3), 149–154. doi:10.1046/j.1525-1497.2000.04199.x PMID:10718894

Gibbons, M. (2011). Use of Health Information technology Among Racial and Ethnic Underserved communities. *Perspectives in Health Information Management*, 8(Winter). Available at <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3035830/>

Gilbert, F. (2010). Cloud service contracts may be fluffy: selected legal issues to consider before taking off. *Journal of Internet Law*, 14(6), 17–30.

Green, A., Ngo-Metzger, Q., Legedza, A., Mas-sagli, M., Philips, R., & Iezzoni, L. (2005). Interpreter Services Language Concordance and Health Care Quality. *Journal of General Internal Medicine*, 20, 1050–1056.

Institute of Medicine. (2003). *Unequal treatment: confronting racial and ethnic disparities in health care*. Washington, DC: The National academic Press.

Interpreter Services Work Group Report to Minnesota Legislature. (2008). *Interpreter Services for Patients in Medical and dental Settings*. Available on line at <http://www.health.state.mn.us/dics/ppc/hci/ISWGreport08.pdf>

Jacobs, E., Shepard, D. S., Suaya, J. A., & Stone, E.-L. (2004). Overcoming Language Barriers in Health care. *American Journal of Public Health*, 94(5), 866–869. doi:10.2105/AJPH.94.5.866 PMID:15117713

Julisch, K. & Hall, M. (2010). Security and control in the cloud. *Information Security Journal: A Global Perspective*, 19(6), 299–309.

Kallick, D. (2007). *Working for a better Life*. New York: Fiscal Policy Institute.

Karliner, L., Perez-Stable, E., & Gildengorin, V. (2004). The Language Divide. *Journal of General Internal Medicine*, 19(2), 175–183. doi:10.1111/j.1525-1497.2004.30268.x PMID:15009797

Meaningful Use. (2012). Retrieved from <http://www.healthit.gov/policy-researchers-implementers/meaningful-use>

Mell, P., & Grance, T. (2010). The NIST definition of cloud computing. *Communications of the ACM*, 53(6), 50.

National Committee for Quality Assurance. (2008). *Standards aimed at Improving Health care for Diverse Groups*. Available online at <http://www.ncqa.org/tabid/916Default.aspx>

Svantesson, D., & Clarke, R. (2010). Privacy and consumer risks in cloud computing. *Computer Law & Security Report*, 26(4), 391–397. doi:10.1016/j.clsr.2010.05.005

Troshani, I., Wickramasinghe, N., & Rampersad, G. (2011). *On cloud Nine? An Integrative Risk Management Framework for Cloud Computing*. Bled eConference.

Wickramasinghe, N., Bali, R., Kirn, S., & Reima, S. (2012). *Creating Sustainable e-health Solutions*. New York: Springer. doi:10.1007/978-1-4614-1536-7

ADDITIONAL READING

Wickramasinghe, N., Bali, R., Mann, R., Baskaran, V., Immonen, A., Naguib, R., & Marshall, I. et al. (2011). Knowledge-based issues for aid agencies in crisis scenarios: Evolving from impediments to trust. *International Journal of Information Systems for Crisis Response and Management*.

KEY TERMS AND DEFINITIONS

Cloud-Based Computing (or Cloud Computing): When computing resources such as software and hardware are delivered as a service over a network.

Healthcare Delivery: The network, structure and stakeholders involved in administering medical care to individuals.

Human Interpreter: Person who performs language translation.

Language Translation: The act of converting the meaning into a different language.

LEP-Limited English Proficiency: Individuals for whom English is not their native language and thus have difficulty communicating and understanding in English.

Meaningful Use: Federal government stipulation for technology solutions in healthcare.

Real-Time Translation: Translation that happens instantaneously at the point of action.

Category H

Health Information Systems

Challenges and Implications of Health Literacy in Global Health Care

Kijpokin Kasemsap

Suan Sunandha Rajabhat University, Thailand

INTRODUCTION

The specific level of health literacy is required for patients to fully understand health information and services to make effective decisions about their health care, including decisions about screening and treatment (Koay, Schofield, & Jefford, 2012). Being able to employ a certain degree of control is recognized as a precondition for the active patient participation and is often discussed concerning health literacy (Nutbeam, 2008). Patients' perceived control over their care is indicated by their perceived ability to perform the self-care activities in the home environment, to interact effectively with health care providers, and to organize care at the right moment (Coulter, 2012). Older patients and patients who are less educated about their health care condition are more likely to have the lower health literacy, visit their health care provider more frequently, and perceive more difficulties in exerting control over their care than the highly educated younger patients (Dahlke, Curtis, Federman, & Wolf, 2014).

This article aims to bridge the gap in the literature on the thorough literature consolidation of health literacy. The extensive literature of health literacy provides a contribution to practitioners and researchers by describing the challenges and implications of health literacy in order to promote the health literacy in global health care.

BACKGROUND

Health literacy is a complicated perspective that depends on the individuals' ability to communi-

cate and the demands posed by society and health care system (Baker, 2006). Lack of specific skills and knowledge associates with health literacy has been shown to negatively affect the people's understanding and use of information provided by health professionals (Rubinelli, Schulz, & Nakamoto, 2009). Effective health care strategies can be used by health care providers to address the serious health care-related problem, including the effective health communication, development of health education materials, professional education, and development of health care community partnerships (Corrarino, 2013). Developing educational materials aimed at individuals with low health literacy, as well as training health care providers on how to effectively communicate with individuals with limited literacy, should be promoted (Ojinnaka et al., 2015). Lupattelli et al. (2014) indicated that clinicians should take time to inquire into their patients' ability to understand health information, perception, and beliefs regarding health literacy.

THEORY AND APPLICATIONS OF HEALTH LITERACY

This section emphasizes the perspectives on health literacy; trends and issues with health literacy; and the challenges and implications of health literacy in global health care.

Perspectives on Health Literacy

Health literacy encompasses several abilities including word recognition, reading comprehension,

DOI: 10.4018/978-1-5225-2255-3.ch324

communication skills, and conceptual knowledge (Macek et al., 2010). The components of literacy include reading, writing, verbal communication, numeracy, and conceptual knowledge (Nielsen-Bohlman, Panzer, & Kindig, 2004). Federman et al. (2009) stated that memory and verbal communication fluency are strongly associated with health literacy. Effective communication among health professions is a necessary component of health care, as no single profession can adequately respond to the complexity of health problems that patients may possess (Barr, 2002).

While health literacy is a complex concept that includes many components, print prose and print document literacy are two essential health literacy skills that help patients understand the written health information (Baker, 2006). Written health information can be found in various areas of health, and includes medical instructions, medication information, disease information, admission forms, informed consent materials, and other examples (Hadden, 2015).

Trends and Issues With Health Literacy

The public health response should include seeking out the new strategies for health systems to promote the public's health literacy, while working with the educational system to better equip the younger generations with the health care-related knowledge and skills necessary to navigate health care (Parker, Wolf, & Kirsch, 2008). In the medical contexts, low health literacy is recognized as a barrier to the health care and treatment adherence, requiring that providers and health care community-based organizations adapt to become more accessible and user-friendly (Liechty, 2011). In the public health contexts, health literacy is recognized as an opportunity for education and empowerment, a capacity-building challenge to enable individuals to exercise more agency in their lives and better utilize the health information and services (Nutbeam, 2000). Public health literacy encompasses skills in evaluating public health information and

the ability to apply it in ways that affect small groups and communities rather than individual's health alone (Freedman et al., 2009).

van der Heide et al. (2015) indicated that in chronic care, patients are expected to fulfill an active role in the care of their condition by fulfilling the care tasks in their home environment. This active patient role is considered as important for maintaining the best possible state of health (Coulter, 2012) and reducing the burden on health care (Cramm & Nieboer, 2012). Not all adults with chronic conditions are able to fulfill this role (Kawi, 2014), either because they are not interested or because they lack the knowledge (Swenson et al., 2004). Patients with higher health literacy are found to take a more active role with respect to their care (Fransen, von Wagner, & Essink-Bot, 2012) and to make less use of health care services than those with lower health literacy (Berkman, Sheridan, Donahue, Halpern, & Crotty, 2011).

An evaluation of government websites from a health literacy perspective has recognized the need for the improvement of readability, user-friendly content, and cultural tailoring (Neuhauser, Rothschild, & Rodriguez, 2007). Cultural competency and health literacy are directly related to health care (Ingram, 2012). Cultural competency is an ongoing process in which a health professional continuously strives to achieve the ability to effectively work within the cultural context of the patient and is a process where there is room for improvement (Campinha-Bacote, 2002). Incorporation of cultural competency into health care education for pharmacy, medical, and nursing students has proven to be effective (Durand, Abel, Silva, & Desilets, 2012) and should be continued in order to increase the awareness of future health professionals.

Challenges and Implications of Health Literacy in Global Health Care

Sentell et al. (2014) stated that primary care providers and facilities should consider health literacy at both community and individual levels. Providing

advice and guidance to individuals regarding both their mental well-being and physical well-being improves the service users' self-esteem and increases their ability to manage their own health (Lennox, Bain, & Rey-Conde, 2007). Hemingway et al. (2015) indicated that people with a learning disability or serious mental illness significantly experience the complex health inequalities that cannot be only understood from the perspective of their condition. Health and social care staff providing services for people with a learning disability increasingly need the education and training in the physical health interventions (Manthorpe & Martineau, 2010).

Adults with inadequate health literacy skills have less ability to identify their medications (Kripalani et al., 2006). Regarding medication adherence, clinic efforts to improve the patient's access to the affordable medications include helping patients with filling out prescription assistance program paperwork, prescribing generic medications, providing samples, and encouraging effective patient education (Alton, March, Mallery, & Fiandt, 2015). Kairuz et al. (2015) stated that low health literacy has the important consequences for health status, medication adherence, and utilization of health services.

Low health literacy is expected to be associated with medication non-adherence (Ostini & Kairuz, 2014). Lower levels of health literacy are associated with a wide range of health-related outcomes, including higher mortality, worse general health, poor health care access, more hospitalizations, and greater use of emergency care, higher health care costs (Eichler, Wieser, & Brugger, 2009), greater difficulty participating in the shared decision making, and a worse medication adherence, and self-management in general (Easton, Entwistle, & Williams, 2010).

Poor health literacy is associated with poor physical health care and mental health care (Wolf, Gazmararian, & Baker, 2005), less use of flu vaccination (Scott, Gazmararian, Williams, & Baker, 2002), and increased mortality (Sudore et al., 2006). Public mental health efforts are

needed to raise the caregiver's mental health literacy (Mendenhall & Frauenholtz, 2015). Mental health literacy is an important factor in the early detection and prompt treatment for the mental, emotional, and behavioral disorders among young people (Olsson & Kennedy, 2010). Poor mental health literacy is associated with the lower rates of help-seeking and service use, as well as societal stigma and discriminatory behavior (Evans-Lacko, Brohan, Mojtabai, & Thornicroft, 2012).

Dunn et al. (2009) indicated that tracking mental health literacy within populations is important as it has been shown to be related to the individual help-seeking as well as the provision of support for those with a mental disorder. Mental health disorders have detrimental effects on well-being, functioning, and development in adolescence (O'Connell, Boat, & Warner, 2009). A recent review of help-seeking for mental disorders considers lack of knowledge regarding mental health as the major obstacle preventing young people from seeking the health-related assistance (Gulliver, Griffiths, & Christensen, 2010).

Regarding mental health literacy, both public stigma and self-stigma have a broad range of negative ramifications for those with mental illness, remarkably social exclusion, and reduced treatment-seeking (Patel et al., 2010). Negative attitudes toward mental illness are commonly endorsed by adolescents (Reavley & Jorm, 2011). High levels of these stigmatizing attitudes among adolescents are associated with the reduced intentions to seek the professional assistance and a lower likelihood of viewing such health-related support as helpful (Yap, Reavley, & Jorm, 2013), and effectively influence the first aid actions toward the distressed peers (Yap & Jorm, 2011).

Numerous studies have emphasized the importance of an adequate level of health literacy to the disease management process (Cho, Lee, Arozullah, & Crittenden, 2008). The inadequate health literacy is a potentially modifiable determinant of poor health outcomes in people with chronic disease (DeWalt, Berkman, Sheridan, Lohr, & Pignone, 2004). Poor reading skills are

associated with an increased use of emergency care, a decreased use of preventive services, and poor health maintenance (Easton et al., 2010). The connection between health literacy and health behavior is likely to be intermediated by knowledge and self-efficacy (Osborn, Paasche-Orlow, Bailey, & Wolf, 2011).

Among adults, low health literacy has been associated with poorer health status and poorer health outcomes, such as the lower medical adherence, lower awareness of medical condition, longer hospital stays, and greater use of the emergency department (Nielsen-Bohlman et al., 2004). Children of parents with lower literacy skills are more likely to have the worse health outcomes (Sanders, Shaw, Guez, Baur, & Rudd, 2009). Low health literacy among parents has been found to be associated with the poor child's health outcomes (DeWalt, Dilling, Rosenthal, & Pignone, 2007).

FUTURE RESEARCH DIRECTIONS

The classification of the extensive literature in the domains of health literacy will provide the potential opportunities for future research. Health literacy is the degree to which individuals have the capacity to obtain, process, and understand the basic health information needed to make appropriate health decisions. Training in health literacy, the use of simple language, and the effective communication is essential for people working in health information and services toward improving the quality of care and reducing the medical error in the health care industry.

The effective communities of practice (CoPs) facilitate the social interactions and encourage the members' willingness to share knowledge and ideas in the workplace (Kasemsap, 2016a). Social media tools are open to anyone, whereas reaching the traditional media often requires a lot of money and a good network of media industry contacts (Kasemsap, 2016b). Total quality management (TQM) practices in quality performance has the

potential to improve organizational performance (Kasemsap, 2015). Promoting health literacy in the health care industry through health care-related training program, CoPs, social media, and TQM activities will be the valuable topic for future research direction.

CONCLUSION

This article highlighted the perspectives on health literacy; trends and issues with health literacy; and the challenges and implications of health literacy in global health care. Health literacy is very important in the health care industry. Health literacy concerns the knowledge and skills of patient to meet the complex demands of health care perspectives in the health care industry. Providing effective patient information means acknowledging, understanding, and overcoming barriers to health literacy that physicians, health professionals, and patients might experience. Health literacy affects health behavior and the use of health services, thus affecting health outcomes and health costs in the health care organizations. The benefits of health literacy improvement include improved communication, greater adherence to treatment, greater ability to engage in self-care, improved health status, greater health care efficiency, and cost savings to the health care systems.

REFERENCES

- Alton, S., March, A. L., Mallery, L., & Fiandt, K. (2015). Medication adherence in a nurse practitioner managed clinic for indigent patients. *Journal of the American Association of Nurse Practitioners*, 27(8), 433–440. doi:10.1002/2327-6924.12211 PMID:25682757
- Baker, D. W. (2006). The meaning and the measure of health literacy. *Journal of General Internal Medicine*, 21(8), 878–883. doi:10.1111/j.1525-1497.2006.00540.x PMID:16881951

- Barr, H. (2002). *Interprofessional education: Today, yesterday and tomorrow*. London, UK: LTSN Health Sciences and Practice.
- Berkman, N. D., Sheridan, S. L., Donahue, K. E., Halpern, D. J., & Crotty, K. (2011). Low health literacy and health outcomes: An updated systematic review. *Annals of Internal Medicine*, 155(2), 97–107. doi:10.7326/0003-4819-155-2-201107190-00005 PMID:21768583
- Campinha-Bacote, J. (2002). The process of cultural competence in the delivery of healthcare services: A model of care. *Journal of Transcultural Nursing*, 13(3), 181–184. doi:10.1177/10459602013003003 PMID:12113146
- Cho, Y. I., Lee, S. Y. D., Arozullah, A. M., & Crittenden, K. S. (2008). Effects of health literacy on health status and health service utilization amongst the elderly. *Social Science & Medicine*, 66(8), 1809–1816. doi:10.1016/j.socscimed.2008.01.003 PMID:18295949
- Corrarino, J. E. (2013). Health literacy and women's health: Challenges and opportunities. *Journal of Midwifery & Women's Health*, 58(3), 257–264. doi:10.1111/jmwh.12018 PMID:23631442
- Coulter, A. (2012). Patient engagement: What works? *The Journal of Ambulatory Care Management*, 35(2), 80–89. doi:10.1097/JAC.0b013e318249e0fd PMID:22415281
- Cramm, J. M., & Nieboer, A. P. (2012). Self-management abilities, physical health and depressive symptoms among patients with cardiovascular diseases, chronic obstructive pulmonary disease, and diabetes. *Patient Education and Counseling*, 87(3), 411–415. doi:10.1016/j.pec.2011.12.006 PMID:22222024
- Dahlke, A. R., Curtis, L. M., Federman, A. D., & Wolf, M. S. (2014). The mini mental status exam as a surrogate measure of health literacy. *Journal of General Internal Medicine*, 29(4), 615–620. doi:10.1007/s11606-013-2712-x PMID:24327311
- DeWalt, D. A., Berkman, N. D., Sheridan, S., Lohr, K. N., & Pignone, M. P. (2004). Literacy and health outcomes: A systematic review of the literature. *Journal of General Internal Medicine*, 19(12), 1228–1239. doi:10.1111/j.1525-1497.2004.40153.x PMID:15610334
- DeWalt, D. A., Dilling, M. H., Rosenthal, M. S., & Pignone, M. P. (2007). Low parental literacy is associated with worse asthma care measures in children. *Ambulatory Pediatrics*, 77(1), 25–31. doi:10.1016/j.ambp.2006.10.001 PMID:17261479
- Dunn, K. I., Goldney, R. D., Grande, E. D., & Taylor, A. (2009). Quantification and examination of depression-related mental health literacy. *Journal of Evaluation in Clinical Practice*, 15(4), 650–653. doi:10.1111/j.1365-2753.2008.01067.x PMID:19522721
- Durand, C., Abel, C., Silva, M., & Desilets, A. (2012). An elective course in cultural competence. *Currents in Pharmacy Teaching and Learning*, 4(2), 102–108. doi:10.1016/j.cptl.2012.01.002
- Easton, P., Entwistle, V. A., & Williams, B. (2010). Health in the “hidden population” of people with low literacy: A systematic review of the literature. *BMC Public Health*, 10(1), 459–468. doi:10.1186/1471-2458-10-459 PMID:20687946
- Eichler, K., Wieser, S., & Brugger, U. (2009). The costs of limited health literacy: A systematic review. *International Journal of Public Health*, 54(5), 313–324. doi:10.1007/s00038-009-0058-2 PMID:19644651
- Evans-Lacko, S., Brohan, E., Mojtabai, R., & Thornicroft, G. (2012). Association between public views of mental illness and self-stigma among individuals with mental illness in 14 European countries. *Psychological Medicine*, 42(8), 1741–1752. doi:10.1017/S0033291711002558 PMID:22085422

- Federman, A. D., Sano, M., Wolf, M. S., Siu, A. L., & Halm, E. A. (2009). Health literacy and cognitive performance in older adults. *Journal of the American Geriatrics Society*, 57(8), 1475–1480. doi:10.1111/j.1532-5415.2009.02347.x PMID:19515101
- Fransen, M. P., von Wagner, C., & Essink-Bot, M. L. (2012). Diabetes self-management in patients with low health literacy: Ordering findings from literature in a health literacy framework. *Patient Education and Counseling*, 88(1), 44–53. doi:10.1016/j.pec.2011.11.015 PMID:22196986
- Freedman, D. A., Bess, K. D., Tucker, H. A., Boyd, D. L., Tuchman, A. M., & Wallston, K. A. (2009). Public health literacy defined. *American Journal of Preventive Medicine*, 36(5), 446–451. doi:10.1016/j.amepre.2009.02.001 PMID:19362698
- Gulliver, A., Griffiths, K. M., & Christensen, H. (2010). Perceived barriers and facilitators to mental health help-seeking in young people: A systematic review. *BMC Psychiatry*, 10(1), 113–121. doi:10.1186/1471-244X-10-113 PMID:21192795
- Hadden, K. B. (2015). Health literacy training for health professions students. *Patient Education and Counseling*, 98(7), 918–920. doi:10.1016/j.pec.2015.03.016 PMID:25850755
- Hemingway, S., Stephenson, J., Trotter, F., Clifton, A., & Holdich, P. (2015). Increasing the health literacy of learning disability and mental health nurses in physical care skills: A pre and post-test evaluation of a workshop on diabetes care. *Nurse Education in Practice*, 15(1), 30–37. doi:10.1016/j.nepr.2014.08.003 PMID:25547647
- Ingram, R. R. (2012). Using Campinha-Bacote's process of cultural competence model to examine the relationship between health literacy and cultural competence. *Journal of Advanced Nursing*, 68(3), 695–704. doi:10.1111/j.1365-2648.2011.05822.x PMID:21895740
- Kairuz, T. E., Bellamy, K. M., Lord, E., Ostini, R., & Emmerton, L. M. (2015). Health literacy among consumers in community pharmacy: Perceptions of pharmacy staff. *Health Expectations*, 18(5), 1041–1051. doi:10.1111/hex.12077 PMID:23710827
- Kasemsap, K. (2015). The role of total quality management practices on quality performance. In A. Moutzoglou, A. Kastania, & S. Archonidakis (Eds.), *Laboratory management information systems: Current requirements and future perspectives* (pp. 1–31). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-6320-6.ch001
- Kasemsap, K. (2016a). Utilizing communities of practice to facilitate knowledge sharing in the digital age. In S. Buckley, G. Majewski, & A. Giannakopoulos (Eds.), *Organizational knowledge facilitation through communities of practice in emerging markets* (pp. 198–224). Hershey, PA: IGI Global. doi:10.4018/978-1-5225-0013-1.ch011
- Kasemsap, K. (2016b). Utilizing social media in modern business. In I. Lee (Ed.), *Encyclopedia of e-commerce development, implementation, and management* (pp. 2171–2182). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-9787-4.ch156
- Kawi, J. (2014). Predictors of self-management for chronic low back pain. *Applied Nursing Research*, 27(4), 206–212. doi:10.1016/j.apnr.2014.02.003 PMID:24636072
- Koay, K., Schofield, P., & Jefford, M. (2012). Importance of health literacy in oncology. *Asia Pacific Journal of Clinical Oncology*, 8(1), 14–23. doi:10.1111/j.1743-7563.2012.01522.x PMID:22369440
- Kripalani, S., Henderson, L. E., Chiu, E. Y., Robertson, R., Kolm, P., & Jacobson, T. A. (2006). Predictors of medication self-management skill in a low-literacy population. *Journal of General Internal Medicine*, 21(8), 852–856. doi:10.1111/j.1525-1497.2006.00536.x PMID:16881946

- Lennox, N., Bain, C., Rey-Conde, T., Purdie, D., Bush, R., & Pandeya, N. (2007). Effects of a comprehensive health assessment programme for Australian adults with intellectual disability: A cluster randomized trial. *International Journal of Epidemiology*, 36(1), 139–146. doi:10.1093/ije/dyl254 PMID:17218326
- Liechty, J. M. (2011). Health literacy: Critical opportunities for social work leadership in health care and research. *Health & Social Work*, 36(2), 99–107. doi:10.1093/hsw/36.2.99 PMID:21661299
- Lupattelli, A., Picinardi, M., Einarson, A., & Nordeng, H. (2014). Health literacy and its association with perception of teratogenic risks and health behavior during pregnancy. *Patient Education and Counseling*, 96(2), 171–178. doi:10.1016/j.pec.2014.04.014 PMID:24862909
- Macek, M. D., Haynes, D., Wells, W., Bauer-Leffler, S., Cotten, P. A., & Parker, R. M. (2010). Measuring conceptual health knowledge in the context of oral health literacy: Preliminary results. *Journal of Public Health Dentistry*, 70(3), 197–204. doi:10.1111/j.1752-7325.2010.00165.x PMID:20337901
- Manthorpe, J., & Martineau, S. (2010). Followers or leaders? What is the role for social care practitioners in annual health checks for adults with learning disabilities? *Journal of Intellectual Disabilities*, 14(1), 53–66. doi:10.1177/1744629510361811 PMID:20630927
- Mendenhall, A. N., & Frauenholtz, S. (2015). Predictors of mental health literacy among parents of youth diagnosed with mood disorders. *Child & Family Social Work*, 20(3), 300–309. doi:10.1111/cfs.12078
- Neuhauser, L., Rothschild, R., & Rodriguez, F. M. (2007). MyPyramid.gov: Assessment of literacy, cultural and linguistic factors in the USDA food pyramid web site. *Journal of Nutrition Education and Behavior*, 39(4), 219–225. doi:10.1016/j.jneb.2007.03.005 PMID:17606248
- Nielsen-Bohlman, L., Panzer, A. M., & Kindig, D. A. (2004). *Health literacy: A prescription to end confusion*. Washington, DC: National Academies Press.
- Nutbeam, D. (2000). Health literacy as a public health goal: A challenge for contemporary health education and communication strategies into the 21st century. *Health Promotion International*, 15(3), 259–267. doi:10.1093/heapro/15.3.259
- Nutbeam, D. (2008). The evolving concept of health literacy. *Social Science & Medicine*, 67(12), 2072–2078. doi:10.1016/j.socscimed.2008.09.050 PMID:18952344
- O’Connell, M. E., Boat, T., & Warner, K. (2009). *Preventing mental, emotional, and behavioral disorders among young people: Progress and possibilities*. Washington, DC: National Academies Press.
- Ojinnaka, C. O., Bolin, J. N., McClellan, D. A., Helduser, J. W., Nash, P., & Ory, M. G. (2015). The role of health literacy and communication habits on previous colorectal cancer screening among low-income and uninsured patients. *Preventive Medicine Reports*, 2, 158–163. doi:10.1016/j.pmedr.2015.02.009 PMID:26844065
- Olsson, D. P., & Kennedy, M. G. (2010). Mental health literacy among young people in a small US town: Recognition of disorders and hypothetical helping responses. *Early Intervention in Psychiatry*, 4(4), 291–298. doi:10.1111/j.1751-7893.2010.00196.x PMID:20977685
- Osborn, C. Y., Paasche-Orlow, M. K., Bailey, S. C., & Wolf, M. S. (2011). The mechanisms linking health literacy to behavior and health status. *American Journal of Health Behavior*, 35(1), 118–128. doi:10.5993/AJHB.35.1.11 PMID:20950164
- Ostini, R., & Kairuz, T. (2014). Investigating the association between health literacy and non-adherence. *International Journal of Clinical Pharmacology, Therapy and Toxicology*, 36(1), 36–44. PMID:24293337

- Parker, R. M., Wolf, M. S., & Kirsch, I. (2008). Preparing for an epidemic of limited health literacy: Weathering the perfect storm. *Journal of General Internal Medicine*, 23(8), 1273–1276. doi:10.1007/s11606-008-0621-1 PMID:18452047
- Patel, V., Maj, M., Flisher, A. J., Silva, M. J., Koschorke, M., & Prince, M. (2010). Reducing the treatment gap for mental disorders: A WPA survey. *World Psychiatry; Official Journal of the World Psychiatric Association (WPA)*, 9(3), 169–176. doi:10.1002/j.2051-5545.2010.tb00305.x PMID:20975864
- Reavley, N. J., & Jorm, A. F. (2011). Stigmatizing attitudes towards people with mental disorders: Findings from an Australian national survey of mental health literacy and stigma. *The Australian and New Zealand Journal of Psychiatry*, 45(12), 1086–1093. doi:10.3109/00048674.2011.621061 PMID:22023236
- Rubinelli, S., Schulz, P. J., & Nakamoto, K. (2009). Health literacy beyond knowledge and behaviour: Letting the patient be a patient. *International Journal of Public Health*, 54(5), 307–311. doi:10.1007/s00038-009-0052-8 PMID:19641846
- Sanders, M. L., Shaw, J. S., Guez, G., Baur, C., & Rudd, R. (2009). Health literacy and child health promotion: Implications for research, clinical care, and public policy. *Pediatrics*, 124(s3), S306–S314. doi:10.1542/peds.2009-1162G PMID:19861485
- Scott, T. L., Gazmararian, J. A., Williams, M. V., & Baker, D. W. (2002). Health literacy and preventive health care use among Medicare enrollees in a managed care organization. *Medical Care*, 40(5), 395–404. doi:10.1097/00005650-200205000-00005 PMID:11961474
- Sentell, T., Zhang, W., Davis, J., Baker, K. K., & Braun, K. L. (2014). The influence of community and individual health literacy on self-reported health status. *Journal of General Internal Medicine*, 29(2), 298–304. doi:10.1007/s11606-013-2638-3 PMID:24096723
- Sudore, R., Metha, A., Simonsick, A., Harris, T. B., Newman, A. B., & Satterfield, S. (2006). Limited literacy in older people and disparities in health and healthcare access. *Journal of the American Geriatrics Society*, 54(5), 770–776. doi:10.1111/j.1532-5415.2006.00691.x PMID:16696742
- Swenson, S. L., Buell, S., Zettler, P., White, M., Ruston, D. C., & Lo, B. (2004). Patient-centered communication: Do patients really prefer it? *Journal of General Internal Medicine*, 19(11), 1069–1079. doi:10.1111/j.1525-1497.2004.30384.x PMID:15566435
- van der Heide, I., Heijmans, M., Heijmans, A. J., Heijmans, E., & Rademakers, J. (2015). Functional, interactive and critical health literacy: Varying relationships with control over care and number of GP visits. *Patient Education and Counseling*, 98(8), 998–1004. doi:10.1016/j.pec.2015.04.006 PMID:25956070
- Wolf, M. S., Gazmararian, J. A., & Baker, D. W. (2005). Health literacy and functional health status among older adults. *Archives of Internal Medicine*, 165(17), 1946–1952. doi:10.1001/archinte.165.17.1946 PMID:16186463
- Yap, M. B. H., & Jorm, A. F. (2011). The influence of stigma on first aid actions taken by young people for mental health problems in a close friend or family member: Findings from an Australian national survey of youth. *Journal of Affective Disorders*, 134(1), 473–477. doi:10.1016/j.jad.2011.05.039 PMID:21658776
- Yap, M. B. H., Reavley, N. J., & Jorm, A. F. (2013). Associations between stigma and help-seeking intentions and beliefs: Findings from an Australian national survey of young people. *Psychiatry Research*, 210(3), 1154–1160. doi:10.1016/j.psychres.2013.08.029 PMID:24011848

ADDITIONAL READING

- Burke, M. T., Kapojos, J., Sammartino, C., & Gray, N. A. (2014). Kidney disease health literacy among new patients referred to a nephrology outpatient clinic. *Internal Medicine Journal*, 44(11), 1080–1086. doi:10.1111/imj.12519 PMID:24989694
- Cailor, S. M., & Chen, A. M. H. (2015). Immediate and longitudinal effects of incorporating health literacy and cultural competency into a yearlong pharmacy curriculum. *Currents in Pharmacy Teaching and Learning*, 7(3), 292–301. doi:10.1016/j.cptl.2014.12.005
- Cameron, K. A. (2013). Advancing equity in clinical preventive services: The role of health communication. *Journal of Communication*, 63(1), 31–50. doi:10.1111/jcom.12005
- Caplan, L., Wolfe, F., Michaud, K., Quinzanos, I., & Hirsh, J. M. (2014). Strong association of health literacy with functional status among rheumatoid arthritis patients: A cross-sectional study. *Arthritis Care and Research*, 66(4), 508–514. doi:10.1002/acr.22165 PMID:24023051
- Chan, H. K., Hassali, M. A., Lim, C. J., & Saleem, F. (2015). Exploring health literacy and difficulty in comprehending pediatric medication labels among caregivers in Malaysia: A pilot study. *Journal of Pharmaceutical Health Services Research*, 6(3), 165–168. doi:10.1111/jphs.12102
- Cheng, C., & Dunn, M. (2015). Health literacy and the Internet: A study on the readability of Australian online health information. *Australian and New Zealand Journal of Public Health*, 39(4), 309–314. doi:10.1111/1753-6405.12341 PMID:25716142
- Devraj, R., Borrego, M., Vilay, A. M., Gordon, E. J., Pailden, J., & Horowitz, B. (2015). Relationship between health literacy and kidney function. *Nephrology (Carlton, Vic.)*, 20(5), 360–367. doi:10.1111/nep.12425 PMID:25661456
- Fry-Bowers, E. K., Maliski, S., Lewis, M. A., Macabasco-O’Connell, A., & DiMatteo, R. (2014). The association of health literacy, social support, self-efficacy and interpersonal interactions with health care providers in low-income Latina mothers. *Journal of Pediatric Nursing*, 29(4), 309–320. doi:10.1016/j.pedn.2014.01.006 PMID:24503164
- Garcia-Marcinkiewicz, A. G., Long, T. R., Danielson, D. R., & Rose, S. H. (2014). Health literacy and anesthesia: Patient’s knowledge of anesthesiologist roles and information desired in the preoperative visit. *Journal of Clinical Anesthesia*, 26(5), 375–382. doi:10.1016/j.jclinane.2014.01.013 PMID:25086485
- Gong, A. T., & Furnham, A. (2014). Mental health literacy: Public knowledge and beliefs about mental disorders in mainland China. *PsyCh Journal*, 3(2), 144–158. doi:10.1002/pchj.55 PMID:26271766
- Lee, J., Stucky, B., Rozier, G., Lee, S. Y., & Zeldin, L. P. (2013). Oral health literacy assessment: Development of an oral health literacy instrument for Spanish speakers. *Journal of Public Health Dentistry*, 73(1), 1–8. doi:10.1111/jphd.12000 PMID:23215757
- Leung, A. Y., Lou, V. W., Cheung, M. K., Chan, S. S., & Chi, I. (2013). Development and validation of Chinese health literacy scale for diabetes. *Journal of Clinical Nursing*, 22(15/16), 2090–2099. doi:10.1111/jocn.12018 PMID:23186320
- Mackert, M., Champlin, S. E., Holton, A., Muñoz, I. I., & Damásio, M. J. (2014). eHealth and health literacy: A research methodology review. *Journal of Computer-Mediated Communication*, 19(3), 516–528. doi:10.1111/jcc4.12044
- Marshall, J. M., & Dunstan, D. A. (2013). Mental health literacy of Australian rural adolescents: An analysis using vignettes and short films. *Australian Psychologist*, 48(2), 119–127. doi:10.1111/j.1742-9544.2011.00048.x

- McNaughton, C., Rothman, R., Storrow, A., & Wallston, K. (2014). Measuring numeracy and health literacy in the emergency department. *Academic Emergency Medicine*, 21(8), 944–945. doi:10.1111/acem.12426 PMID:25156278
- Moore, K., Smith, B. J., & Reilly, K. (2013). Community understanding of the preventability of major health conditions as a measure of health literacy. *The Australian Journal of Rural Health*, 21(1), 35–40. doi:10.1111/ajr.12005 PMID:23384135
- Naigaga, M. D., Guttersrud, Ø., & Pettersen, K. S. (2015). Measuring maternal health literacy in adolescents attending antenatal care in a developing country: The impact of selected demographic characteristics. *Journal of Clinical Nursing*, 24(17/18), 2402–2409. doi:10.1111/jocn.12796 PMID:25763773
- Patel, B. (2015). Communicating across cultures: Proceedings of a workshop to assess health literacy and cross-cultural communication skills. *Journal of Pharmacy Practice and Research*, 45(1), 49–56. doi:10.1002/jppr.1062
- Perry, E. L. (2014). Health literacy in adolescents: An integrative review. *Journal for Specialists in Pediatric Nursing*, 19(3), 210–218. doi:10.1111/jspn.12072 PMID:24612548
- Prins, E., & Mooney, A. (2014). Literacy and health disparities. *New Directions for Adult and Continuing Education*, 2014(142), 25–35. doi:10.1002/ace.20092
- Rubenstein, E. L. (2013). Crowdsourcing health literacy: The case of an online community. *Proceedings of the American Society of Information Science and Technology*, 50(1), 1–5. doi:10.1002/meet.14505001119
- Shin, W. K., Braun, T. M., & Inglehart, M. R. (2014). Parent's dental anxiety and oral health literacy: Effects on parents' and children's oral health-related experiences. *Journal of Public Health Dentistry*, 74(3), 195–201. doi:10.1111/jphd.12046 PMID:24329522
- Srinivasan, D. (2014). The impact of trust in e-government on electronic health literacy. *World Medical & Health Policy*, 6(1), 22–38. doi:10.1002/wmh3.86
- Stikes, R., Arterberry, K., & Logsdon, M. C. (2015). A nurse leadership project to improve health literacy on a maternal-infant unit. *Journal of Obstetric, Gynecologic, and Neonatal Nursing*, 44(5), 665–676. doi:10.1111/1552-6909.12742 PMID:26295935
- Wang, J., Thombs, B. D., & Schmid, M. R. (2014). The Swiss health literacy survey: Development and psychometric properties of a multidimensional instrument to assess competencies for health. *Health Expectations*, 17(3), 396–417. doi:10.1111/j.1369-7625.2012.00766.x PMID:22390287
- Wei, Y., Hayden, J. A., Kutcher, S., Zygmunt, A., & McGrath, P. (2013). The effectiveness of school mental health literacy programs to address knowledge, attitudes and help seeking among youth. *Early Intervention in Psychiatry*, 7(2), 109–121. doi:10.1111/eip.12010 PMID:23343220
- Yap, M. B. H., Mackinnon, A., Reavley, N. J., & Jorm, A. F. (2014). The measurement properties of stigmatizing attitudes towards mental disorders: Results from two community surveys. *International Journal of Methods in Psychiatric Research*, 23(1), 49–61. doi:10.1002/mpr.1433 PMID:24550065

KEY TERMS AND DEFINITIONS

Communication: The two-way process of reaching mutual understanding, in which participant not only exchange information, news, ideas, and feelings, but also create and share meaning.

Health Care: The act of taking the preventative or necessary medical procedures to improve a person's well-being.

Information: The data that is organized for a specific purpose and presented within a context that gives it meaning and relevance.

Knowledge: The understanding of a circumstance gained through experience.

Literacy: The ability to read and write.

Patient: A person who is receiving medical care, or who is cared for by a physician or health professional, when necessary.

Skill: The ability acquired through the deliberate, systematic, and sustained effort to smoothly and adaptively execute the complex activities.

Understanding: The knowledge about a situation or about how something works.

Cyber Behaviors in Seeking Health Information

H

Xiaojun (Jenny) Yuan
University of Albany, USA

José A. Pino
Universidad de Chile, Chile

INTRODUCTION

Warner and Procaccino (2004) defined health information seeking (HIS) as “the process of seeking information needed for personal decision-making related to health and medical issues.” HIS constitutes an important part of people’s information seeking behavior. In analyzing a Pew Research Center survey, Kennedy and Funk (2015) reported that about 66% online adults show interests on health and medicine topics, while 37% of online adults claim that “health and medicine” is one of the topics they deem as most interesting. Kennedy and Funk (2015) also noted that research on measuring public understanding about science and technology usually relates science and technology to the health and medical domain.

Consider the following scenario in your ordinary daily life (Yuan & White, 2012). Shown in Figure 1.

In this scenario, people need to 1. Deal with a need of seeking information on headache and brain tumor, which could be a complex task for people with limited medical knowledge. 2. Overcome potential stress caused by knowing about others’ stressful events (e.g. deaths). 3. Share information with others who have similar concerns. Each of these three items is explained in further detail below.

1. In completing complex tasks, people need guidance to go through the required steps towards completing the tasks, instead of a simple ranked list (Joachims et al., 1997). The information foraging theory developed by Pirolli and Card (1999) provided a theoretical foundation for this situation. This theory indicated that information seekers rely on cues left by previous visitors to identify *patches* of information, and then use them to

Figure 1.

Scenario: You have been bothered by a headache for a while. Your primary doctor suggested that you take a series of tests, but the results revealed nothing serious. One of your friends died of a brain tumor several years ago. You are worried about your own situation and would like to explore the issue in more detail. Specifically, you want to learn more about the types of headache, the corresponding symptoms, the causes, and the remedies for each type. You want to find websites with useful information and share your findings with others by posting as many Web links as you feel are necessary to adequately cover the topic on a social networking site such as Facebook.

Task: Please find as many Web links as you feel are necessary to adequately cover the topic and you believe can be useful to people in a similar situation. Copy and paste the Web links to the answer sheet. In the answer sheet, please answer each question. Also, create a sequence of Web links, comprising all links you selected or some sub-set, arranged so as to be useful to others.

DOI: 10.4018/978-1-5225-2255-3.ch325

solve their information problems. Building on the previous findings and the information foraging theory, in a HIS context, Yuan and White (2012) studied the impact of domain knowledge on information seeking behavior and found that expert-generated sequences of web pages are more useful than those of novice users.

2. According to Folkman (1984), stress is “a relationship between the person and the environment that is appraised by the person as taxing or exceeding his or her resources and as endangering his or her well-being.” Studying information behavior models, Wilson (1997) reported that “stress arises in many situations that are considerably less than life-threatening but may endanger one’s “well-being” or, perhaps, self-image, in other ways.” Wilson also argued that stress arose in HIS because “the emotional impact of life-threatening diseases or operations is very obvious.” Current research indicates that people’s awareness and stress level are increased when they are provided information about stressful events associated with other people’s lives (Hampton et al., 2015; Smith and Rose, 2011).
3. How to design a system that can help users learn, understand and use during the information seeking process has been challenging. Twidale et al. (1997) suggested that “a truly user-centred system must acknowledge and support collaborative interactions between all users.” They proved that collaboration can actually help improve users’ learning and understanding of the systems. A well designed system should be able to help people get needed information and share it with those who have the same information need. People who are seeking the same or related information may probably also benefit from collaboration by division of labor.

Yuan & Dumas (2012) reviewed the recent literature in the field of information science and

human computer interaction, and proposed Collaborative Information Seeking (CIS) as one of the promising directions. CIS was defined as “a research area about studying how people work in collaborative groups for information seeking, gathering and sharing, and also about how to build systems for supporting such activities.” (Yuan & Dumas, 2012). Collaboration may be an advantageous process for individuals faced with a task, as Grey (1989) identifies it: “a process through which parties who see different aspects of a problem can constructively explore their differences and search for solutions that go beyond their own limited vision of what is possible”. Karunakaran et al. (2013) proposed that collaborative information behavior involved three phases, including problem formulation, collaborative information seeking, and information use.

The field of CIS has been covered in main international conferences such as the ACM International Conference on Information and Knowledge Management (CIKM), the annual European Conference on Information Retrieval (ECIR), the ACM Special Interest Group on Information Retrieval (SIGIR), the ACM Conference on Computer-Supported Cooperative Work and Social Computing (CSCW), the ACM SIGCHI conference on Human Factors in Computing Systems (SIGCHI), the ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR), and the Annual Meeting of the Association for Information Science and Technology (ASIST). Relevant journals are Journal of Information Processing and Management (IP&M) and the Journal of the Association for Information Science and Technology (JASIST). A special issue of IEEE Computer in 2014 was also dedicated to CIS. Finally, a related book has been released (Hansen et al., 2015).

This paper discusses HIS behavior in general, and then followed with collaborative HIS behavior in terms of the current status, the representative researchers and their work, challenges and future directions. All aspects of collaborative HIS constitute a wide field and thus, this paper focuses

on the HIS social dimension and explores how collaborative HIS could contribute to promoting people's positive health behavior. Afterwards, collaborative information seeking systems are introduced, and future directions and conclusion are discussed.

BACKGROUND

Research has shown that HIS behavior varies depending on user's information needs and situations. People tend to use different strategies and resources to look for various types of information to satisfy their information needs. Johnson and Meischke (1991) reported that when searching about cancer, people tried to find factual information on how to prevent, detect and treat cancer. This study indicated that people may also look for information to help them emotionally cope with their health problems. In a research on HIS behavior of medical experts versus novice users, Yuan & White (2012) found that medical experts utilized different seeking strategies and selected different information resources than those of novice users; and medical experts selected more relevant and useful URLs than novice users. Warner and Procaccino (2004) indicated that people's HIS process may involve multiple channels, including formal ones (e.g. physicians and other health professionals, health organizations, librarians, and media), informal channels (e.g. friends and relatives), and hybrid channels (e.g. formal and informal Internet health resources, and interactive sites including online chat groups, facebook, and twitter).

Researchers have studied factors affecting people's HIS behavior. These factors include technology use and social media (Smith and Rose, 2011), age (Connell and Crawford, 1988; Watkins et al., 2014), gender (Connell and Crawford, 1988; Warner & Procaccino, 2004), motivation and emotion (Kassulke et al., 1993; Wilson, 1997).

Connell and Crawford (1988) proved that women believed they collected more health information than men from different sources. They

related this finding to the mother nature of women: they are care-givers and "lay health care providers". This finding was confirmed by research on a national cancer information service (BACUP) by Slevin et al. (1988), in which women contributed 80% use of the service. Connell and Crawford (1988) also found that age plays a critical role in the amount of health information urban residents received from a variety of sources. Specifically, older men received much fewer information than younger men; and older women received slightly fewer information than younger women. In terms of an exploratory study of online HIS behavior in Hong Kong, Yan (2010) found that digital divide is obvious based on age and education. To address the information need of older adults and to balance the gap between their knowledge on health need and recent technologies, it could be helpful to design appropriate mobile apps for older adults. Watkins et al. (2014) found that it is important to help older adults change their health behavior in a positive way, such as increasing their daily fruit and vegetable intake by supporting their search for useful information with touch devices like Apple iPads. They generated 14 usability heuristics for older adults and iPad apps and further conducted a heuristic evaluation study to identify the existing usability issues of the current iPad apps.

It has been debated whether the use of technologies negatively affected or positively promoted people's health behavior in general. Hampton et al. (2015) examined whether the use of social media, mobile phones and Internet can potentially lead to higher levels of stress. Results indicated that (1) use of Internet and social media does not lead to higher levels of stress. In fact, women who use social media frequently perceived lower levels of stress; (2) in a situation that people are aware of other people suffering stressful events, their own stress increases. This situation is called "cost of caring." In particular, in comparison to men, women are more sensitive to the stressful events of their closest friends and family (Hampton et al., 2015). Social interaction is an important factor to be considered in this "cost of caring" situation.

Thus, we must take into account the social dimension in HIS process. Radecki and Jaccard (1995) demonstrated that user perception of a friend's knowledge has an impact on their perception of their own knowledge. The importance of social dimension in HIS behavior has also been addressed by other researchers in the field. Kassulke et al. (1993) noted that emotional barriers, defined as "a construct consisting of questions relating to an inability to make decisions about health and to take advantage of existing health services", could dramatically limit people's access to health services. The information seeking behavior model of Wilson (1997) pointed out the relation between emotion, motivation and effectiveness of information seeking processes, and the critical role of social interaction in people's HIS behavior.

There exist evidence to support the claim that the social dimension of HIS behavior demands more research on collaborative HIS behavior. For example, the cost of caring situation (Smith and Rose, 2011), in which the link between the level of stress to HIS behavior was considered by taking into account the social dimension in HIS. Yuan & White (2012) compared the usefulness and effectiveness of search trails (the sequence of webpages in the order of usefulness) generated by medical experts and novice users, and found that the trails of experts were more effective and useful than those of novice users. In their study, the task scenarios (mentioned in the Introduction) are about finding information for someone they know who was diagnosed with serious illness. The findings indicated that under such a "cost of caring" situation, people tended to go to trusted websites (e.g., www.nih.gov) to search for information, and they rely on opinions of experts. Results also indicated that domain experts used different strategies in selecting resources and structuring search trails than those of novice users. In a study relevant to information visualization and presentation techniques, Yuan et al. (2015) investigated how different information presentation styles affected consumers' preference and decision-making in purchasing product online. They found that con-

sumers prefer some specific presentation styles in their purchase decision making in an opinion summarization task environment. These research efforts created opportunities to integrate or apply research in general HIS to the field of CIS.

In sum, research has been paying attention to the importance of incorporating collaboration in information seeking processes and designing information retrieval systems to satisfy people's health information needs. Next, CIS in health in general will be discussed, followed by system and user interface design of CIS systems.

COLLABORATIVE INFORMATION SEEKING IN HEALTH

Traditionally, information seeking is a solo activity, but with technology evolution, the ways people seek information changed. In many cases, people need to share information and/or learn from others. As Twidale and Nichols (1996) noted, "The use of library resources is often stereotyped as a solitary activity, with hardly any mention in the substantial library science and information retrieval literature of the social aspects of information systems." They believed users would more effectively learn and use information retrieval systems if such systems can support collaboration. Twidale et al. (1997) proposed that "a truly user-centred system must acknowledge and support collaborative interactions between users." Morris (2007) found that introducing support for collaborative search could improve the features of exploratory search, i.e., coverage, confidence, exposure, and productivity. These studies demonstrate the importance of considering collaboration in the information seeking process and system design/development.

Research revealed the importance of CIS in people's daily work. Based on a survey of academic researchers' collaborative information seeking behavior, Spence et al. (2005) found that researchers employed various tools (e.g. e-mail, phone, and video-conferencing) to collaborate during information seeking processes. In this

study, researchers felt that they were motivated to collaborate in seeking information mainly because of their lack of expertise. More importantly, results indicated that CIS is generally successful and more useful than seeking information individually. In the health and medical domain, research on CIS process and its affecting factors have become increasingly important. In a study of patent office workers, Hansen and Järvelin (2005) reported that researchers should reframe the hypothesis that information retrieval performance is just individual-based. Within the context of hospital care, Gorman et al. (2000) studied how members of an intensive care team collaborated together to seek and share information. They discussed the importance of putting different information sources together to answer team members' questions. Based on examining the radiological request form in a study of intra-departmental coordination at a hospital, Symon et al. (1996) reported that the form used to schedule radiology examinations can only provide incomplete information. Thus a number of workarounds by the staff were needed in order to identify the needed information. This research concentrated on information needs and information records, as well as helping the medical staff find needed information. Reddy and Dourish (2002) investigated how work rhythms played the role in team members' CIS activities in an intensive care unit. Understanding work rhythms can enable team members know when information is needed. Subsequently, team members could collaborate with each other for needed information at appropriate timing. As Grudin (1994) pointed out, "Many expensive failures in developing and marketing software that is designed to support groups are not due to technical problems. They result from not understanding the unique demands this class of software imposes on developers and users." The above mentioned research made it possible to develop effective CIS systems by understanding such demands.

On the technical side, effort has been made to develop collaborative information systems (CIS) to better support interaction and collaboration dur-

ing people's information seeking process. Some representative systems are introduced below. These systems were designed with different focus and approaches, including support for collaboration (1) at the user interface level (Twidale et al., 1995); (2) at the system-mediated level (Pickens et al., 2008); (3) on awareness (Morris & Horvitz, 2007; Shah, 2010); and (4) on communication (Krishnappa, 2005).

In terms of system design for collaboration, Twidale and Nichols (1996) claimed that a visualization of the search process that can be changed and communicated by the real users must be developed. The "Ariadne" system was developed by Twidale et al. (1995) for supporting the collaborative learning of database browsing techniques. In this system, a visualization of the search process was designed to "facilitate complex browsing processes in collaboration." Such a visualization is composed of thumbnails of screens presented as play cards, representing *command-output* pairs. Each play card can be expanded to display the contents. Pickens et al. (2008) proposed a CIS system designed using algorithmic mediation for synchronous collaborative search and communication activities. In this system, people can search together to solve a common information problem. MUSE, a collaborative information seeking and retrieval prototype (Krishnappa, 2005) was designed with the focus on communication. After evaluating MUSE, Krishnappa suggested that the collaborative features (e.g. the chat function) improved collaboration between teams in the information seeking and retrieval process. In particular, they found that using the chat function enabled users to understand both the search process and the findings.

As one of the important features in the design of CIS systems, awareness has been considered from various perspectives. Schmidt (2002) suggested that awareness should be considered "not as a separate entity, but as someone's being aware of some particular occurrence." Morris and Horvitz (2007) developed a system called SearchTogether that incorporates awareness using various methods,

including per-user query histories (it shows each team member's screen name, photo, and queries in the "QueryAwareness" region); and display of page-specific metadata. Query awareness was identified as an important feature in collaborative searching by allowing team members to share their query terms and to learn various query formulation techniques from each other. Coagmento (Shah, 2010) is a system exploring awareness in various interface designs. The major difference between SearchTogether and Coagmento is that the former supports searching and sharing, but the latter also supports organizing and synthesizing information. Regarding awareness, besides query and results awareness, the latter also considered contextual and workspace awareness.

Challenges exist in the design and evaluation of CIS systems, and in health information system research in a broader context. After examining the publication of health information systems research (HISR) in 17 IS journals since 1985, Chiasson and Davidson (2004) claimed that "healthcare poses significant challenges to IS theory, but, by taking on these challenges, HISR researchers can contribute to this multidisciplinary field as well as expand the body of IS knowledge." Furthermore, they suggested that the health care industry "poses important social challenges and intriguing research possibilities for researchers interested in the development and use of information systems and technologies." Baeza-Yates & Pino (1997, 2006) took the initiative to formally evaluate Collaborative Information Retrieval proposing metrics to do so. The evaluation of Collaborative Information Retrieval/Seeking is still a challenging endeavor. Soulier et al. (2015) mention this occurs because there is a variety of confounding factors such as the multi-user context, the exploratory aspect of the search through multi-session search activities, the multiplicity of relevance factors, the individual vs. collective value of relevance, and the search interfaces supporting the collaborative interactions. A framework to evaluate CIS has been proposed by Antunes et al. (2012).

FUTURE RESEARCH DIRECTIONS

Collaborative Health Information Seeking is an area facing technical challenges common to CIS in general, as mentioned above, like the evaluation one. However, the area will greatly benefit from advances in social networks, since it presents many characteristics associated to them. It is yet to be investigated how people in the scenario mentioned in the Introduction can get a good result belonging to communities with varying degrees of knowledge on health-related subjects, using advanced systems either controlled by the community members or system-driven, accessed from any place at any time, using crowdsourcing, distributed search strategies and ontologies.

Generally speaking, future directions in CIS include, but not limited to the following aspects.

- It is important to comprehensively understand in which situations people need to collaboratively seek for health information, including topics of interest, level of detail, and stages of information seeking. Based on such understanding, how people interact and collaborate to satisfy their health information needs should be investigated, e.g., the relationship between tasks (including task complexity and task types), supporting techniques, and stages of information processes. Furthermore, CIS systems should be designed to respond to the findings of these investigations, in order to better serve people's HIS.
- With the development and use of mobile devices and applications, mobile health behavior is becoming an important research area. Kassab and Yuan (2012) found that searching for information is the main motivation for people to access the internet resources using mobile devices. Taking into account the importance of HIS (Kennedy and Funk, 2015) and the fact that 44% active users accessing Facebook using mobile

devices (Dan, 2011), research on people's collaborative health information seeking using mobile devices would be an important area to grow. Meanwhile, the progress of speech recognition technology in a mobile environment creates opportunities to incorporate a spoken language interface to collaborative information seeking systems for the purpose of improving interaction and collaboration during people's HIS process. Previous findings (Begany, Sa, & Yuan, 2015; Yuan & Sa, 2016) on spoken language interface may be able to apply to the CIS interface.

- Research on how to apply methodologies and findings of health ICT (information and communication technologies) to collaborative information seeking would be a promising direction. Begany, Mirzaei and Yuan (2014) started a mobile diary study project to investigate health information behaviors of ICT users. They aim at understanding the use of ICTs (e.g. computers, laptops, tablets, smartphones, cellphones, and health wearables) for HIS and provide practical implications for the design of future health information systems, applications and devices (including health wearables).
- As Shah (2015) pointed out, it is meaningful to look at the different visualization methods of CIS interfaces, and find out which kind is useful for CIS users. Future research should also address the need of evaluation of CIS systems and setting up standards for such evaluation.

CONCLUSION

Seeking health information is an important component of Cyber behavior. CIS has become part of people's daily lives with the advancement of technologies, devices and tools provided in the

digital society. CIS became increasingly important because of the nature of the task, and the phenomenal change on the way people interact and share information through social networks. The development of technologies and social networks pose challenges and opportunities to researchers and practitioners interested in CIS to deeply explore relevant collaboration activities, and design and implement CIS systems to support such activities. More studies involving a large and diverse group of users should be conducted to find out real information needs in the situation requesting collaboration and sharing. System design and evaluation should take into account the complexity of tasks, the information needs and situations.

REFERENCES

- Antunes, P., Herskovic, V., Ochoa, S., & Pino, J. A. (2012). Structuring dimensions for collaborative systems evaluation. *ACM Computing Surveys*, 44(2), 1–28. doi:10.1145/2089125.2089128
- Baeza-Yates, R., & Pino, J. A. (1997). A First Step to Formally Evaluate Collaborative Work. *ACM International Conference on Supporting Group Work (GROUP'97)*, 55–60. doi:10.1145/266838.266860
- Baeza-Yates, R., Pino, J.A. (2006). Towards Formal Evaluation of Collaborative Work and its application to Information Retrieval. *Information Research*, 11(4).
- Begany, G., Mirzaei, K., & Yuan, X.-J. (2014). *Investigating health information behavior: A mobile diary study*. Workshop on “Designing the Future of Mobile Healthcare Support” in Mobile HCI 2014, Toronto, Canada.
- Begany, G., Sa, N., & Yuan, X.-J. (2015). Factors affecting user perception of a spoken language vs. textual search interface: A content analysis. *Journal of Interacting with Computers*. doi:10.1093/iwc/iwv029

- Chiasson, M. W., & Davidson, E. (2004). Pushing the contextual envelope: Developing and diffusing IS theory for health information systems research. *Information and Organization*, 14(3), 155–188. doi:10.1016/j.infoandorg.2004.02.001
- Connell, C. M., & Crawford, C. O. (1988). How people obtain their health information: A survey in two Pennsylvania counties. *Public Health Reports*, 103, 189–195. PMID:3128837
- Dan (blogger). (2011). *Digital stats: Facebook has more than 750m active users (full stats)*. Retrieved November 30, 2012 from <http://digital-stats.blogspot.com/2011/09/facebook-has-morethan-750m-active.html>
- Folkman, S. (1984). Personal control and stress and coping processes: A theoretical analysis. *Journal of Personality and Social Psychology*, 46(4), 839–852. doi:10.1037/0022-3514.46.4.839 PMID:6737195
- Gorman, P. N., Ash, J., Lavelle, M., Lyman, J., Delcambre, L., Maier, D., & Bowers, S. et al. (2000). Bundles in the Wild: Managing Information to Solve Problems and Maintain Situation Awareness. *Library Trends*, 49(2), 266–289.
- Grey, B. (1989). Collaborating: finding common ground for multiparty problems. Jossey-Bass.
- Grudin, J. (1994). Groupware and social dynamics: Eight challenges for developers. *Communications of the ACM*, 37(1), 92–105. doi:10.1145/175222.175230
- Hampton, K. N., Rainie, L., Lu, W., Shin, I., & Purcell, K. (2015). Social Media and the Cost of Caring. Washington, DC: Pew Research Center. Retrieved from http://www.pewinternet.org/files/2015/01/PI_Social-media-and-stress_0115151.pdf
- Hansen, P., & Järvelin, K. (2005). Collaborative information retrieval in an information-intensive domain. *Information Processing & Management*, 41(5), 1101–1119. doi:10.1016/j.ipm.2004.04.016
- Hansen, P., Shah, C., & Klas, C.-P. (Eds.). (2015). *Collaborative Information Seeking: Best Practices, New Domains and New Thoughts*. Springer. doi:10.1007/978-3-319-18988-8
- Joachims, T., Freitag, D., Mitchell, T. (1997). WebWatcher: A tour guide for the world wide web. *IJCAI*, 770-775.
- Johnson, J. D., & Meischke, H. (1991). Women's preferences for cancer information from specific communication channels. *The American Behavioral Scientist*, 34(6), 742–755. doi:10.1177/0002764291034006010
- Kassab, D., Yuan, X.-J. (2012). Understanding the information needs and search behaviors of mobile users. *Journal of Information Research*, 17(4).
- Karunakaran, A., Reddy, M. C., & Spence, P. R. (2013). Toward a model of collaborative information behavior in organizations. *Journal of the American Society for Information Science and Technology*, 64(12), 2437–2451. doi:10.1002/asi.22943
- Kassulke, D., Stenner, D. K., Coory, M., & Ring, I. (1993). Information seeking behaviour and sources of health information: Associations with risk factors status in an analysis of three Queensland electorates. *Australian Journal of Public Health*, 17(1), 51–57. doi:10.1111/j.1753-6405.1993.tb00104.x PMID:8357895
- Kennedy, B., & Funk, C. (2015). Public Interest in Science and Health Linked to Gender, Age and Personality. Washington, DC: Pew Research Center. Retrieved from <http://www.pewinternet.org/2015/12/11/public-interest-in-science-and-health-linked-to-gender-age-and-personal>
- Krishnappa, R. (2005). *Multi-user search engine (MUSE): Supporting collaborative information seeking and retrieval* (Master's thesis). University of Missouri-Rolla, MO.

- Morris, M. R. (2007). Interfaces for collaborative exploratory Web search: Motivations and directions for multi-user design. *ACM SIGCHI 2007 Workshop on Exploratory Search and HCI: Designing and Evaluating Interfaces to Support Exploratory Search Interaction*, 9–12.
- Morris, M. R., & Horvitz, E. (2007). SearchTogether: An interface for collaborative Web search. *The 20th ACM Symposium on User Interface Software and Technology*, 3–10.
- Pickens, J., Golovchinsky, G., Shah, C., Qvarfordt, P., & Back, M. (2008). Algorithmic mediation for collaborative exploratory search. *The 31st ACM SIGIR Conference on Research and Development in Information Retrieval*, 315–322.
- Pirolli, P., & Card, S. K. (1999). Information foraging. *Psychological Review*, 106(4), 643–675. doi:10.1037/0033-295X.106.4.643
- Radecki, C. M., & Jaccard, J. (1995). Perceptions of knowledge, actual knowledge, and information search behaviour. *Journal of Experimental Social Psychology*, 31(2), 107–138. doi:10.1006/jesp.1995.1006
- Reddy, M., Dourish, P. (2002). A finger on the pulse: temporal rhythms and information seeking in medical work. *2002 ACM Conference on Computer Supported Cooperative Work*, 344–353.
- Schmidt, K. (2002). The problem with awareness: Introductory remarks on awareness in CSCW. *Computer Supported Cooperative Work*, 11(3), 285–298. doi:10.1023/A:1021272909573
- Shah, C. (2010). Coagmento—A collaborative information seeking, synthesis and sense-making framework. *Integrated demo at the CSCW 2010 Workshop on Collaborative Information Seeking*.
- Shah, C. (2015). Collaborative Information Seeking: From ‘What?’ and ‘Why?’ to ‘How?’ and ‘So What?’. In *Collaborative Information Seeking* (pp. 3–16). Springer International Publishing.
- Smith, R. L., & Rose, A. J. (2011). The “cost of caring” in youths’ friendships: Considering associations among social perspective taking, co-rumination, and empathetic distress. *Developmental Psychology*, 47(6), 1792.
- Slevin, M. L., Terry, Y., Hallett, N., Jefferies, S., Launder, S., Plant, R., & McElwain, T. et al. (1988). BACUP—the first two years: Evaluation of a national cancer information service. *British Medical Journal*, 297, 669–672. doi:10.1136/bmj.297.6649.669 PMID:3140928
- Smith, R. L., & Rose, A. J. (2011). The cost of caring in youths friendships: Considering associations among social perspective taking, co-rumination, and empathetic distress. *Developmental Psychology*, 47(6), 1792–1803. doi:10.1037/a0025309 PMID:21895362
- Soulier, L., Tamine, L., Sakai, T., Azzopardi, L., & Pickens, J. (2015). Report on the First International Workshop on the Evaluation on Collaborative Information Seeking and Retrieval (ECol’2015). *Workshop in 24th CIKM*, Melbourne, Australia.
- Spence, P. R., Reddy, M., & Hall, R. H. (2005). A Survey of Collaborative Information Seeking Practices of Academic Researchers. *2005 International ACM SIGGROUP Conference on Supporting Group Work*. doi:10.1145/1099203.1099216
- Symon, G., Long, K., & Ellis, J. (1996). The Coordination of Work Activities: Cooperation and Conflict in a Hospital Context. *Computer Supported Cooperative Work*, 5(1), 1–31. doi:10.1007/BF00141934
- Twidale, M. B., Nichols, D. M., & Paice, C. D. (1995). Supporting collaborative learning during information searching. In *Proceedings of Computer Supported Collaborative Learning* (pp. 367–374). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

- Twidale, M. B., & Nichols, D. M. (1996). Collaborative browsing and visualisation of the search process. *Proceedings of Aslib*, 48, 177–182. Retrieved from <http://www.comp.lancs.ac.uk/computing/research/cseg/projects/ariadne/docs/elvira96.html>
- Twidale, M. B., Nichols, D. M., & Paice, C. D. (1997). Browsing is a collaborative process. *Information Processing & Management*, 33(6), 761–783. doi:10.1016/S0306-4573(97)00040-X
- Warner, D., & Procaccino, J. D. (2004). Toward wellness: Women seeking health information. *Journal of the American Society for Information Science and Technology*, 55(8), 709–730. doi:10.1002/asi.20016
- Watkins, I., Kules, B., Yuan, X.-J., & Xie, B. (2014). Heuristic evaluation of healthy eating apps for older adults. *Journal of Consumer Health on the Internet*, 18(2), 105–127. doi:10.1080/15398285.2014.902267
- Wilson, T. D. (1997). Information behaviour: An interdisciplinary perspective. *Information Processing & Management*, 33(4), 551–572. doi:10.1016/S0306-4573(97)00028-9
- Yan, Y. Y. (2010). Online health information seeking behavior in Hong Kong: An exploratory study. *Journal of Medical Systems*, 34(2), 147–153. doi:10.1007/s10916-008-9226-9 PMID:20433053
- Yuan, X.-J., & Dumas, C. (2012). Cyber behaviors in seeking information. *Encyclopedia of Cyber Behavior*. IGI Global.
- Yuan, X.-J., & Sa, N. (2016). User Query Behaviour in Different Task Types in a Spoken Language vs. Textual Interface: A Wizard of Oz experiment. *The Proceeding of ISIC 2016, the Information Behaviour Conference*.
- Yuan, X.-J., Sa, N., Begany, G., & Yang, H. (2015). What users prefer and why: A user study on effective presentation styles of opinion summarization. Chapter Human-Computer Interaction – INTERACT 2015. *Lecture Notes in Computer Science*, 9297, 249–264. doi:10.1007/978-3-319-22668-2_20
- Yuan, X.-J., & White, R. (2012). Building the trail best traveled: Effects of domain knowledge on web search trailblazing. *ACM SIGCHI Conference on Human Factors in Computing Systems (CHI 2012)*. doi:10.1145/2207676.2208312

ADDITIONAL READING

- Alarcon, R. A., Guerrero, L., Ochoa, S., & Pino, J. A. (2005). “Context in collaborative mobile scenarios”. CONTEXT Workshop on Collaboration and Technology. CEUR Proceedings, 5-8.
- Antunes, P., Herskovic, V., Ochoa, S., & Pino, J. A. (2014). Reviewing the quality of awareness support in collaborative applications. *Journal of Systems and Software*, 89, 146–169. doi:10.1016/j.jss.2013.11.1078
- Borges, M., Brézillon, P., Pino, J. A., & Pomerol, J. (2004). “Groupware system design and the context concept”. *International Conference on Computer supported Cooperative Work in Design*, 45-54.
- Shah, C. (2014). Collaborative information seeking. *Journal of the Association for Information Science and Technology*, 65(2), 215–236. doi:10.1002/asi.22977
- Yuan, X.-J., & Belkin, N. J. (2014). Applying an information-seeking dialogue model in an interactive information retrieval system. *The Journal of Documentation*, 70(5), 829–855. doi:10.1108/JD-06-2013-0079

KEY TERMS AND DEFINITIONS

Collaborative Health Information Seeking: People seek for health information collaboratively.

Collaborative Information Seeking System: An information system supporting collaboration and sharing among users during their information seeking processes.

Cost of Caring: People get stressful when they know of stressful events occurred in other people's lives.

Health Information Seeking (HIS) Behavior: Human information behavior related to seeking for health information to satisfy human information needs.

Health information Systems Research (HISR): The multidisciplinary research relevant to the design, construction and use of information systems and technologies in a healthcare setting.

H

Information Systems and Technology Projects in Healthcare Organisations

Jorge Gomes

ISEG, Universidade de Lisboa, Portugal

Mário José Batista Romão

ISEG, Universidade de Lisboa, Portugal

INTRODUCTION

The challenges facing healthcare organizations require more comprehensive and integrated solutions and efficient resource management as a means of eliminating inefficiencies and of achieving promised benefits. In academic literature, information systems and technology (IS/IT) have been recognized as being an organizational capability that can lead to competitive advantage and better performance (Bharadwaj 2000; Kohli & Devaraj 2003). Organizations recognize project management as being a fundamental tool for the development of initiatives which lead to the implementation of the organizational strategies (Crawford, 2005; Hodgson, 2002). One way that the effectiveness of IS/IT project management capability has been assessed is through the use of maturity models, with the underlying assumption that higher levels of project management maturity imply a higher effectiveness of project management capability (Kwak & Ibbs 2002; Sonnekus & Labuschagne, 2004).

Our research focuses on the combination of the project management and maturity models approaches as a means of strengthening the final results of IS/IT projects in the healthcare sector. It is the authors' belief that this combination of approaches enhances not only the success of projects, but also the realization of the expected

benefits. It is also important to emphasize that, by taking advantage of the specific features of each of these approaches, their structure will certainly increase the effectiveness of IS/IT projects in the health sector, by enhancing both the confidence of sponsors and investors, and also the achievement of the promise benefits.

The maturity models approach provides a framework which helps organizations increase their capability to deliver projects on schedule, within budget, and according to the desired technical performance (Levin & Skulmoski, 2000). Projects are temporary achievements that are used to solve various types of tasks of variable size, and are applicable in a very broad range of business sectors (Maylor, 2001). Project management coordinates skills and organizational knowledge and follows the progress of a set of pre-established activities in order to achieve objectives (Kronbichel, 2009). Many organizations fail to review whether the planned benefits of IS/IT projects have been achieved, or not, as they do not possess sufficient resources to undertake such a benefit review, and, moreover, they are constantly under pressure to deliver other projects (Bennington & Baccarini, 2004). Benefits management identifies goals and benefits by combining organizational changes and investments in IS/IT, and also by showing the way to achieve them (Gomes, Romão & Caldeira, 2013; Ward & Daniel, 2006).

BACKGROUND

Literature Review

Whilst there is general agreement that IS/IT does indeed contribute to adding business value, there is uncertainty as to how these contributions were really obtained (Melville, Kraemer & Gurbaxani, 2004; Devaraj & Kholi, 2003). Although many studies have focused on the consequences of IS/IT investments, fewer studies have examined factors which impact the capability of IS/IT (Devaraj & Kholi, 2003).

Project Management Institute (PMI) (2012) define project as a limited effort in time, which is undertaken to create a product, service or a result. The essence of project management is to support the implementation of these temporal initiatives under the framework of an organization's competitive strategy, in order to successfully deliver a particular outcome (Milosevic, 2003; Senhar & Dvir, 2007). Project management is thus a set of management activities which is required to ensure that projects which are defined, planned and monitored, go on to achieve agreed objectives and benefits (Deveraj & Kohli, 2003). Kerzner (2009) highlights the importance of project management in the planning and control of organizations' resources, helping to achieve, not only short-term goals, but also broader, temporal objectives.

It appears that determining whether a project is a success, or not, is far more complex. Success is perceived differently by the different stakeholders involved in the projects (Freeman & Beale, 1992). The differences in success criteria definition should reflect the different interests and points of view, which leads us to conclude that project success is a multidimensional criterion (Freeman & Beale, 1992; Pinto & Mantel, 1990). Success criteria known as the 'iron triangle' have been criticized for their exclusive focus on the project management process, to the detriment of including the vision and goals of the different stakeholders (Baccarini, 1999; Bannerman, 2008).

This classic approach remains the most widely used measure of project success and its main value is in offering a simple, direct measure of performance of the project, but it neglects whether the deliverables fulfilled the objectives of the project (Bannerman, 2008).

The improvement in the success of projects results from increased maturity and organizational competence (Sergeant et al., 2010). Higher levels of maturity will, in most cases, lead to improved project outcomes (Nieto-Rodrigues & Evrard, 2004). Projects which have multiple stakeholders, with different perspectives about the purpose of the project, usually have different expectations as to what the project should achieve (Lim & Mohamed, 1999). Since the success of a project depends on the perceptions of a large number of stakeholders, "*absolute success*" probably never exists in project management, but only "*perceived success*" (Baker, Murphy & Fisher, 1988). In the academic literature, we found examples of projects that have successfully completed the criteria of the "iron triangle", but resulted in disappointing business experiences (Shenhar et al., 2005). On the other hand, initiatives that did not meet the constraints of cost and time later proved to be successful (Pinto & Slevin, 1988). The understanding of the concept of project success has evolved over recent decades, and a gradual understanding is now emerging that project success requires a broader and more comprehensive definition.

Aaltonen et al. (2008), state that the key issue in project management is managing the relationship between the project and its stakeholders. There is little evidence to suggest that process capability improvement result in improved project success, although a few studies are promising in this respect (Mullaly, 2006; Lee & Anderson, 2006).

Maturity models have become an important evaluation tool for measuring the internal and external capabilities of organizations. Maturity models describe the development of an entity over time (Klimko, 2001) and they represent a structured collection of elements which highlight

the characteristics of effective processes at different stages of development (Pullen, 2007). The maturity models approach provides a theoretical framework for improving the business outcome of an organization, for assessing their strengths and weaknesses, and also for allowing comparisons with industry best practices and also by benchmarking with similar organizations (Ibbs & Kwak, 2000). Several studies focus on the recognition of the benefits of investment in project management skills in organizations and others have discussed the issue of the correlation between level of maturity and the performance of projects (Ibbs & Kwak, 2000; Ibbs & Reginato, 2002; Mullaly, 2006). In an IS/IT discipline, maturity is considered to be a measure for evaluating an organization's capabilities (Rosemann & Bruin, 2005). Measuring the maturity of organizations is a difficult and somewhat subjective task, as such an audit process focuses mainly on individuals' tasks (Andersen & Jessen, 2003). Working with different types of projects within an organization requires standard models, in order to deliver successful future projects repeatedly and as a means of improving the quality of future projects and of gaining knowledge and learning from past successes and mistakes. The assessment of maturity typically involves variations along five developmental stages (Jugdev & Thomas, 2002).

These assessment procedures help organizations to understand where they have been, where they are, and what processes they need to implement, in order to continue their implementation of management methodologies. The underlying assumption in the maturity models is that there is a relationship between higher levels of maturity and improved organizational performance (Ibbs & Kwak, 2000; Lee & Anderson, 2006) and that this enhances project success.

The Healthcare Sector

Today, healthcare organizations are increasing focusing on the need for investment in IS/IT, with the goal of achieving the minimum level of benefits that these projects can attain. The study

of the success or failure of these initiatives has become vitally important for the performance of these organizations (Rahimi & Vimarlund, 2007). Since the 1990's, the health sector has sought to improve its effectiveness and efficiency by adopting IS/IT to increase quality levels (Raghupathi & Tan, 1999). The use of IS/IT in health sector is recognized as being a major factor which has improved clinical practices and supportive care (McDonald et al., 1998). The use of these systems provides an important support for specialized services and it increases the efficiency, quality and safety of patient care, and also reduces medical errors (Low & Chen, 2012). It is not always investments in IS/IT that result in efficiency and effectiveness gains, and thus it becomes essential to evaluate the factors that limit performance, and to identify opportunities for enhancing their use.

IS/IT has the potential to dramatically change the way individuals or society see the healthcare sector, and also to provide tremendous opportunities for supporting professionals, and for improving effectiveness and efficiency in the health sector (Ammenwerth et al., 2006).

There is a growing consensus that organizational factors are far more critical for the successful implementation of IS/IT, than technical considerations (Markus et al., 2000). Achieving successful change is much easier if all stakeholders are committed, and the earlier this commitment is achieved, the smoother is the path to a successful outcome (Bradley, 2006).

We live in times where healthcare providers generate significant amounts of personal data about patients and the major obstacle to the management of this increasing volume of information is the difficulty, or inability, of sharing information across systems and between organizations (Grimson et al., 2000). Medical information needed for clinical decision making has increased dramatically, however the accessibility of health data is still poor, resulting in inappropriate decisions and sometimes in medical errors (Tierney, 2001).

The use of informatics tools has been developed to increase the accessibility and management of

medical information (Bleich, Beckley & Horowitz, 1985), with the aim of supporting medical decision, of increasing the coordination between different health care providers, and of promoting the use of guidelines, thereby improving the global quality of care (Pringle, 1988; Shiffmann et al., 1999). However, in addition to providing new capabilities, new technologies also impact the technical, social, organizational, economic, cultural, and political dimensions of work in new and different ways (Anderson & Aydin, 1994).

Observations of new technology implementations have shown that a change in technology alters roles, strategies, and paths to failure (Sarter, Woods & Billings, 1997). In recognizing this, the Institute of Medicine of the USA recommends examining new technologies for avoiding threats to safety and redesigning them to prevent undesirable accidents (Kohn, Corrigan & Donaldson, 2000).

There is a widespread feeling that a significant proportion of initiatives in IS/IT healthcare have failed (Heeks & Davies, 1999). Studies have identified high failure rates in IS/IT projects in various sectors, including that of healthcare, particularly in hospitals (Kaplan & Harris-Salamone, 2009; Wears & Berg, 2005). The results of the implementation of IS/IT projects in healthcare have revealed a waste of financial resources in acquiring large sized systems, which are totally ineffective (Heeks, 2006). In various aspects, these implementations are different from other projects, in other industries. The key main differences were related to the environment, the diversity of systems and the devices that need to work, together with the challenge of integration and interoperability which is required to meet the expectations of different stakeholder groups regarding that which constitutes project success (Abouzahra, 2011). Healthcare projects are a complex undertaking, which depend largely on the quality of existing information (Bose, 2003).

Proper training is a major determinant for success in the adoption of IS/IT by health professionals, and it has a great influence on the integration of technologies in clinical practice (Allen et al.,

2000). The effectiveness of interventions aimed at the integration of IS/IT applications in the practices of health professionals tends to be influenced by several factors, which are related to individuals, professional groups, organizational and contextual characteristics, and the nature of the intervention per se (Grol et al., 2007; Aarts et al., 2004). One of the most critical factors recognized by the academic literature is resistance to change by healthcare professionals, particularly amongst doctors (Lapointe & Rivard, 2006; Phansalker et al., 2008).

The complexity of systems, together with organizational diversity and the volume of investment required, as well as failure in adopting IS/IT, is all largely justified by the way that IS/IT is implemented, and by the need to identify best practices and to act on a number of critical factors, in order to reduce the chance of failure (Olson & Zhao, 2007). Cooke-Davis & Arzymanow (2003) identify organizational culture as exerting a positive influence on the development of superior project management practices.

FUTURE RESEARCH DIRECTIONS

Healthcare organizations require interdisciplinary cooperation and coordination. It needs to be highlighted that insufficient communication and missing information are among the major issues that have contributed to unintentional injury caused by medical mismanagement (Lenz & Reichert, 2007).

IS/IT processes have the potential to significantly reduce the rate of these adverse events by providing relevant information in real time to all who need it (McDonald et al., 1984). An important challenge for the future is to seek for a *real clinical integration* of systems. Clinical integration between providers and hospitals has historically been a goal which is continually sought, but rarely achieved. It will become crucial that the design of future applications be integrated easier into existing systems, through open communication interface (Geissbuhler et al., 2001).

The *intersection of healthcare and social media* represents another promising field of IS/IT research. Anderson and Agarwal (2010) demonstrate that individuals are more willing to share personal healthcare information if they think that it could help others. Social media platforms could provide aggregation through the mechanisms of information filtering and knowledge synthesis (Kane et al, 2009).

Evidence-based medicine (EBM) is a form of medicine which aims to optimize decision-making by emphasizing the use of evidence from well-designed and well-conducted research. EBM produces quantitative research, especially from randomized controlled trials (RCTs). A randomized controlled trial (Chalmers et al., 1981) is a type of scientific experiment whereby those studied are randomly allocated one or other of the different treatments under study. EBM has been gaining increased attention among researchers (Carey, 2006) as a tool for addressing concerns about healthcare costs and quality.

Another emerging topic for knowledge discovery arises from the use of digital technology to enable new kinds of *mathematical healthcare modelling and simulations* (Lumpkin, 2007). Simulation models are distinguished from other types of conceptual models by the fact that they include simulated objects, such as people, which correspond to real objects, one-to-one. Development of such a model requires creating a population of simulated individuals who all have experienced important events which occur in real life with real people, and who respond to interventions in the same way as real people (Schlessinger & Eddy, 2002).

Another research trend concerns the use of the combined knowledge about a person for predicting treatment response and this can thereby improve that person's health (Redekop & Mladsi, 2013). This concept of *personalized medicine* allows for an earlier and more precise diagnosis, cheaper and

more effective treatments, and the minimization of side effects of treatment (Glaser et al. 2008). Personalized medicine uses genomic medicine to take advantage of a molecular understanding of disease, in order to optimize preventive health care strategies and drug therapies whilst people are still well, or at the earliest stages of disease (Ginsburg & Willard, 2009). Another term which has emerged in recent years is *personalized healthcare*, which extends the previous concept to address questions related to treatment monitoring and disease surveillance.

CONCLUSION

Although project management emerges as being one of the main approaches used by organizations, there is no strong evidence in the literature of the success of the implementation of any of the available approaches in the health sector. This study attempts to show that organizational maturity contributes to the success of a project, through the systematic application of project management best practices.

The combination of the project management and the organizational process maturity approaches provides a more efficient and useful framework for supporting decision-making. The approach proposed in this study tries to prove that it is easier to implement best practices in project management in organizations which have higher levels of organizational maturity, and that the correct combination between investments in IS/IT and management practices can also be a positive influence, leading to successful projects.

We have a strong conviction that higher levels of organizational maturity have a positive influence on the success of projects implemented by organizations particularly in cases where good project management practices are in evidence.

REFERENCES

- Aaltonen, K., Jaakko, K., & Tuomas, O. (2008). Stakeholders salience in global projects. *International Journal of Project Management*, 26(5), 509–516. doi:10.1016/j.ijproman.2008.05.004
- Abouzghara, M. (2011). Causes of failure in Healthcare IT projects. In *Proceedings of 3rd International Conference on Advance Management Science*, 19. Singapore: IACSIT Press.
- Allen, M. J., Kaufmann, D. M., Barret, A., Paterson, G., Sargeant, J., & McLeod, R. (2000). Self-reported effects of computer workshops on physicians computer use. *The Journal of Continuing Education in the Health Professions*, 20(1), 20–26. doi:10.1002/chp.1340200105 PMID:11232068
- Ammenwerth, E., Iller, C., & Mahler, C. (2006). IT-adoption and the interaction of task, technology and individuals: A fit framework and a case study. *BMC Medical Informatics and Decision Making*, 6(3). doi:10.1186/1472-6947-6-3 PMID:16401336
- Anderson, C. L., & Agarwal, R. (2010). Practicing Safe Computing: A Multimedia Empirical Examination of Home Computer User Security Behavioral Intentions. *MIS Quarterly*, 34(3), 613–643.
- Anderson, E. S., & Jessen, S. A. (2003). Project maturity in organizations. *International Journal of Project Management*, 21(6), 457–461. doi:10.1016/S0263-7863(02)00088-1
- Anderson, J. G., & Aydin, C. E. (1994). Overview: Theoretical perspectives and methodologies for the evaluation of health care information systems. In J. G. Anderson, C. E. Aydin & S. J. Jay (Eds.), *Evaluating health care information systems: Methods and applications*, (pp. 5–29). Thousand Oaks, CA: Sage Publications, Inc.
- Baccarini, D. (1999). The logical framework method for defining project success. *Project Management Journal*, 30(4), 25–32.
- Baker, B. N., Murphy, D. C., & Fisher, D. (1988). Factors affecting project success. In D. I. Cleland & W. R. King (Eds.), *Project Management Handbook* (2nd ed.; pp. 902–919). New York: Van Nostrand Reinhold.
- Bannerman, P. (2008). Defining project success: a Multilevel Framework. *Proceedings of Project Management Institute Research Conference*.
- Bennington, P., & Baccarini, D. (2004). Project benefits management in IT projects – An Australian perspective. *Project Management Journal*, 35(2), 20–30.
- Bharadwaj, A. S. (2000). A Resource Based Perspective on Information Technology and Firm Performance; An EMprical Investigation. *Management Information Systems Quarterly*, 24(1), 169–196. doi:10.2307/3250983
- Bleich, H. L., Beckley, R. F., Horowitz, G. L., Jackson, J. D., Moody, E. S., Franklin, C., & Slack, W. V. et al. (1985). Clinical computing in a teaching hospital. *The New England Journal of Medicine*, 312(12), 756–764. doi:10.1056/NEJM198503213121205 PMID:3838364
- Bose, R. (2003). Knowledge management-enabled health care management: Capabilities, infrastructure, and decision-making. *Expert Systems with Applications*, 24(1), 59–71. doi:10.1016/S0957-4174(02)00083-0
- Bradley, G. (2006). *Benefit Realization Management: A Practical Guide for Achieving Benefits through Change*. Gower Publishing, Ltd.
- Carey, J. (2006, May 29). Medical Guesswork. *Business Week*, 72–79.
- Chalmers, T. C., Smith, H. Jr, Blackburn, B., Silverman, B., Schroeder, B., Reitman, D., & Ambroz, A. (1981). A method for assessing the quality of a randomized control trial. *Controlled Clinical Trials*, 2(1), 31–49. doi:10.1016/0197-2456(81)90056-8 PMID:7261638

- Cooke-Davies, T., & Arzymanow, A. (2003). The Maturity of Project Management in Different Industries: An Investigation into Variations Between Project Management Models. *International Journal of Project Management*, 21(6), 471–478. doi:10.1016/S0263-7863(02)00084-4
- Crawford, L. (2005). Senior management perceptions of project management competence. *International Journal of Project Management*, 23(1), 7–16. doi:10.1016/j.ijproman.2004.06.005
- Devaraj, S., & Kohli, R. (2003). Performance Impacts of Information Technology: Is Actual Usage the Missing Link? *Management Science*, 49(3), 273–289. doi:10.1287/mnsc.49.3.273.12736
- Freeman, M., & Beale, P. (1992). Measuring Project Success. *Project Management Journal*, 23(1), 8–17.
- Geissbuhler, A., Lovis, C., Lamb, A., & Spahni, S. (2001). Experience with an XML/HTTP-based federative approach to develop a hospital-wide clinical information system. In *Proceedings of the International Medical Informatics Conference, Medinfo 2001*, (pp. 735-739). London, UK: Academic Press.
- Ginsburg, G. S., & Willard, H. F. (2009). Genomic and Personalized Medicine: Foundations and Applications. *Translational Research; the Journal of Laboratory and Clinical Medicine*, 154(6), 277–287. doi:10.1016/j.trsl.2009.09.005 PMID:19931193
- Glaser, J., Henley, D. E., Downing, G., Brinner, K. M., & Brinner, . (2008). Advancing personalized healthcare through health information technology: An update from the American health information communitys personalized health workgroup. *Journal of the American Medical Informatics Association*, 15(4), 391–394. doi:10.1197/jamia.M2718 PMID:18436899
- Gomes, J., Romão, M., & Caldeira, M. (2013). The Benefits Management and Balanced Scorecard Strategy Map: How They Match. *International Journal of IT/Business Alignment and Governance*, 4(1), 44–54. doi:10.4018/jitbag.2013010104
- Grimson, J., Grimson, W., & Hasselbring, W. (2000). The SI challenge in health care. *Communications of the ACM*, 43(6), 49–55. doi:10.1145/336460.336474
- Grol, G., Bosch, M. C., Hulscher, M., Eccles, M. P., & Wensing, M. (2007). Planning and Studying Improvement in Patient Care: The use of theoretical perspectives. *The Milbank Quarterly*, 83(1), 93–138. doi:10.1111/j.1468-0009.2007.00478.x PMID:17319808
- Heeks, R. (2006). Health Information Systems: Failure, success and improvisation. *International Journal of Medical Informatics*, 75(2), 125–137. doi:10.1016/j.ijmedinf.2005.07.024 PMID:16112893
- Heeks, R. B., & Davies, A. (1999). Different approaches to information age reform. In R. B. Heeks (Ed.), *Reinventing Government in the Information Age: International Practice in IT-enabled Public Sector Reform*. London, UK: Routledge. doi:10.4324/9780203204962
- Hodgson, D. (2002). Disciplining the professional: The case of project management. *Journal of Management Studies*, 39(6), 803–821. doi:10.1111/1467-6486.00312
- Ibbs, W. C., & Kwak, Y. H. (2000). Assessing project management maturity. *Project Management Journal*, 31(1), 32–43.
- Ibbs, W. C., & Reginato, J. (2002). *Quantifying the value of Project Management*. Newton Square, PA: Project Management Institute.
- Jugdev, K., & Thomas, J. (2002). Project Management Maturity Models: The Silver Bullets of Competitive Advantage? *Project Management Journal*, 33(4), 4–14.

- Kane, G. C., Fichman, R. G., Gallagher, J., & Glaser, J. (2009). Community Relations 2.0. *Harvard Business Review*, 87(11), 45–50. PMID:19891388
- Kaplan, B., & Harris-Salamone, K. D. (2009). Health IT success and failure: Recommendations from literature and an AMIA workshop. *Journal of the American Medical Informatics Association*, 16(3), 291–299. doi:10.1197/jamia.M2997 PMID:19261935
- Kerzner, H. (2009). *Project management: A systems approach to planning, scheduling, and controlling*. New York: John Wiley & Sons.
- Klimko, G. (2001). Knowledge management and maturity models: building common understanding. *Proceedings of the 2nd European Conference on Knowledge Management*, 269–278.
- Kohli, R., & Devaraj, S. (2003). Measuring Information Technology Payoff: A Meta-Analysis of Structural Variables in Firm-Level Empirical Research. *Information Systems Research*, 14(2), 127–145. doi:10.1287/isre.14.2.127.16019
- Kohn, L. T., Corrigan, J. M., & Donaldson, M. S. (2000). *To err is human: building a safer health system*. Washington, DC: National Academy Press.
- Kronbichel, S. A., Ostermann, H., & Staudinger, R. (2009). A review of critical success factors for ERP-projects. *The Open Information Systems Journal*, 3(1), 14–25. doi:10.2174/1874133900903010014
- Kwak, Y. H., & Ibbs, W. (2002). Project management process maturity (PM)² model. *Journal of Management Engineering*, 18(3), 150–155. doi:10.1061/(ASCE)0742-597X(2002)18:3(150)
- Lapointe, L., & Rivard, S. (2006). Getting physicians to accept new information technology: Insights from case studies. *Canadian Medical Association*, 174(11), 1573–1578. doi:10.1503/cmaj.050281 PMID:16717265
- Lee, L. S., & Anderson, R. M. (2006). An exploratory investigation of the antecedents of the IT project management capability. *E-Service Journal*, 5(1), 27–42. doi:10.2979/ESJ.2006.5.1.27
- Lenz, R., & Reichert, M. (2007). IT support for healthcare processes: Premises, challenges, perspectives. *Data & Knowledge Engineering*, 61(1), 39–58. doi:10.1016/j.datak.2006.04.007
- Levin, G., & Skulmoski, G. (2000). The project management maturity. *ESI Horizons*, 2(3), 1–7.
- Lim, C. S., & Mohamed, M. Z. (1999). Criteria of project success. *International Journal of Project Management*, 17(4), 243–248. doi:10.1016/S0263-7863(98)00040-4
- Low, C., & Chen, Y. H. (2012). Criteria for the evaluation of a cloud-based hospital information system outsourcing provider. *Journal of Medical Systems*, 36(6), 3543–3553. doi:10.1007/s10916-012-9829-z PMID:22366976
- Lumpkin, J. R. (2007). Archimedes: A bold step into the future. *Health Affairs*, 26(2), w137–w139. doi:10.1377/hlthaff.26.2.w137 PMID:17259195
- Markus, M. L., Axline, S., Petrie, D., & Tanis, C. (2000). Learning from adopters experiences with ERP: Problems encountered and success achieved. *Journal of Information Technology*, 15(4), 245–265. doi:10.1080/02683960010008944
- Maylor, H. (2001). Beyond the Gantt chart: Project management moving on. *European Management Journal*, 19(1), 92–100. doi:10.1016/S0263-2373(00)00074-8
- McDonald, C. J., Hui, S. L., Smith, D. M., Tierney, W. M., Cohen, S. J., Weinberger, M., & McCabe, G. P. (1984). Reminders to physicians from an introspective computer medical record. *Annals of Internal Medicine*, 100(1), 130–138. doi:10.7326/0003-4819-100-1-130 PMID:6691639

- McDonald, C. J., Overhage, J. M., Dexter, P. R., Blevins, L., Meeks-Johnson, J., Suico, J. G., & Schadow, G. et al. (1998). Canopy computing: Using the web in clinical practice. *Journal of the American Medical Association*, 280(15), 1325–1329. doi:10.1001/jama.280.15.1325 PMID:9794311
- Melville, N., Kraemer, K., & Gurbaxani, V. (2004). Information Technology and Organizational Performance: An Integrative Model of IT Business Value. *Management Information Systems Quarterly*, 28(2), 283–322.
- Milosevic, D. (2003). *Project management toolbox: tools and techniques for the practicing project manager*. Hoboken, NJ: John Wiley & Sons.
- Mullaly, M. (2006). Longitudinal Analysis of Project Management Maturity. *Project Management Journal*, 36(3), 62–73.
- Nieto-Rodriguez, A., & Evrard, D. (2004). *Boosting Business Performance through Programme and Project Management: A first global survey on the current state of Project Management Maturity in Organizations across the World*. London: PriceWaterhouseCoopers.
- Olson, D. L., & Zhao, F. (2007). CIO perspectives of critical success factors in ERP upgrade projects. *Enterprise Information Systems*, 1(1), 129–138. doi:10.1080/17517570601088364
- Phansalker, S., Weir, C. R., Morris, A. H., & Warner, H. R. (2008). Clinicians' perceptions about use of computerized protocols: A multi-center study. *International Journal of Medical Association*, 77(3), 184–193. doi:10.1016/j.ijmedinf.2007.02.002 PMID:17409016
- Pinto, J. K., & Mantel, S. J. (1990). The causes of project failure. *IEEE Transactions on Engineering Management*, 37(4), 269–276. doi:10.1109/17.62322
- Pinto, J. K., & Slevin, D. P. (1988). Project success: Definitions and measurement techniques. *Project Management Journal*, 19(1), 67–72.
- Pringle, M. (1988). Using computers to take patient histories. *BMJ (Clinical Research Ed.)*, 297(6650), 697–698. doi:10.1136/bmj.297.6650.697 PMID:3147730
- Project Management Institute. (2012). *PMI Lexicon of Project Management Terms, version 2.0*. Newtown Square, Pennsylvania: Project Management Institute Inc.
- Pullen, W. (2007). A public sector HPT maturity model. *Performance Improvement*, 46(4), 9–15. doi:10.1002/pfi.119
- Raghupathi, W., & Tan, J. (1999). Strategic use of information technology in healthcare: A state-of-the-art. *Topics in Health Information Management*, 1(1), 1–15. PMID:10539419
- Rahimi, B., & Vimarlund, V. (2007). Methods to evaluate health information systems in health care settings: A literature review. *Journal of Medical Systems*, 31(5), 397–432. doi:10.1007/s10916-007-9082-z PMID:17918694
- Redekop, W. K., & Mladi, D. (2013). The Faces of Personalised Medicine: A Framework for Understanding its Meanings and Scope. *Value in Health*, 16(6), S4–S9. doi:10.1016/j.jval.2013.06.005 PMID:24034312
- Rosemann, M., & de Bruin, T. (2005). Towards a business process management maturity model. *Proceedings of the ECIS2005, 13th European Conference on Information Systems*.
- Sarter, N. B., Woods, D. D., & Billings, C. E. (1997). Automation surprises. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (2nd ed.; pp. 1926–1943). New York: John Wiley and Sons.
- Schlessinger, L., & Eddy, D. M. (2002). Archimedes: A new model for simulating health care systems—the mathematical formulation. *Journal of Biomedical Informatics*, 35(1), 37–50. doi:10.1016/S1532-0464(02)00006-0 PMID:12415725

Sergeant, R., Hatcher, C., Trigunaryah, B., Coffey, V., & Kraatz. (2010). Creating value in Project Management using PINCE 2. Research Project of Queensland University of Technology, Australia and Office of Government Commerce, UK.

Shenhart, A. J., & Dvir, D. (2007). *Reinventing Project Management: The Diamond Approach to Successful Growth and Innovation*. Boston: Harvard Business School Press.

Shenhart, A. J., Dvir, D., Guth, W., Lechler, T., Panatakul, P., & Poli, M. (2005). Project strategy: The missing link. *Proceedings of Academy of Management Annual Meeting*.

Shiffman, R., Liaw, Y., Brandt, C., & Corb, G. (1999). Computer-based guideline implementation systems: A systematic review of functionality and effectiveness. *Journal of the American Medical Informatics Association*, 6(2), 104–114. doi:10.1136/jamia.1999.0060104 PMID:10094063

Sonnekus, R., & Labuschagne, L. (2004). Establishing the relationship between IT project management maturity and IT project success in a South African context. *Proceedings of the 2004 PMSA International Conference Global Knowledge for Project Management Professionals*, 183–192.

Tierney, W. (2001). Improving clinical decisions and outcomes with information: A review. *International Journal of Medical Informatics*, 62(1), 1–9. doi:10.1016/S1386-5056(01)00127-7 PMID:11340002

Ward, J., & Daniel, E. (2006). *Benefits Management: Delivering Value from IS and IT Investments*. Chichester, UK: John Wiley & Sons.

Wears, R. L., & Berg, M. (2005). Computer technology and clinical works: Still waiting for Godot. *Journal of the American Medical Association*, 293(10), 1261–1263. doi:10.1001/jama.293.10.1261 PMID:15755949

ADDITIONAL READING

Aarts, J., Doorewaard, H., & Berg, M. (2004). Understanding implementation: The case of a computerized physician order entry system in a large Dutch University Medical Center. *Journal of the American Medical Informatics Association*, 11(3), 207–216. doi:10.1197/jamia.M1372 PMID:14764612

Flower, J. (2013). *Healthcare Beyond Reform: Doing It Right for Half the Cost*. Boca Raton: CRC Press, Taylor & Francis Group.

Hodge, J. G. (2003). Health Information Privacy and Public Health. *The Journal of Law, Medicine & Ethics*, 31(4), 663–671. doi:10.1111/j.1748-720X.2003.tb00133.x PMID:14968668

Hyrkäs, K., & Lehti, K. (2003). Continuous quality improvement through team supervision, supported by continuous self-monitoring of work and systematic patient feedback. *Journal of Nursing Management*, 11(2), 208–226. PMID:12694365

Low, C., & Chen, Y. H. (2012). Criteria for the evaluation of a cloud-based hospital information system outsourcing provider. *Journal of Medical Systems*, 36(6), 3543–3553. doi:10.1007/s10916-012-9829-z PMID:22366976

KEY TERMS AND DEFINITIONS

Benefits Management: Benefits Management is an integral part of managing change and aims to increase the success of the measurable and significant benefits of investments in IS/IT for an organization.

Critical Success Factors: Critical success factors are elements which are vital for a project to be successful.

Maturity Models: A theoretical framework that helps organizations to increase their ability to deliver projects by assessing strengths and weaknesses, good sector practices and benchmarking with similar organizations.

Project Management: Project management is the application of processes, methods, knowledge, skills and experience to achieve the project objectives.

Project Management Success: Project management success is measured mostly against cost, time and quality.

Project Success: Project success is measured against the overall objectives of the project.

Project Success Criteria: Project success criteria are the standards by which the project will be judged at completion.

Peer-to-Peer Health-Related Online Support Groups



Neil S. Coulson

University of Nottingham, UK

INTRODUCTION

In recent years more and more individuals living with long-term conditions are turning to the internet for information, advice and support. In particular, there has been a rapid expansion of the number of online support groups (also known as 'online support communities') and this rise in popularity is arguably related to many of the unique characteristics inherent within this form of communication. Researchers have begun to examine the nature of online support groups (e.g., Wright & Bell, 2003), their role in the lives of those living with illness (e.g., Mo & Coulson, 2010) as well as the activities which are taking place online (e.g., Malik & Coulson, 2010). There is much qualitative and cross-sectional evidence (e.g., Coulson, 2005) describing their impact on individuals but as yet there are few well-designed studies which have clearly tested the effects of peer to peer online communication and support. This entry will consider the current state of knowledge with regards key aspects of this growing online phenomenon.

BACKGROUND

Unique Characteristics of Online Support Communities

Online support groups present several unique communication characteristics, which can include anonymity, asynchronous text-based communication and the ability to transcend geographical and temporal barriers (White & Dorman, 2001;

Joinson, 2003; Coulson & Knibb, 2007). These unique characteristics also give rise to a number of unique advantages and disadvantages for individuals choosing to seek support through the medium of online groups.

Potential Advantages

A number of studies (e.g., Ferguson, 1996; Buchanan & Coulson, 2007) have documented the potential advantages of online support groups over more traditional face-to-face sources of support (e.g., face-to-face support groups).

Access Support at Any Time

In contrast to a face-to-face group where members typically meet once a week or less frequently, online groups are available 24 hours a day, 7 days a week (Malik & Coulson, 2008) thereby providing opportunities for peer support at any time of the day or night. As a consequence, this introduces considerable flexibility for individuals in terms of when they wish to access support and this flexibility has been shown to be very helpful for those who have work, family or educational commitments (Coulson & Knibb, 2007).

Asynchronous Communication

The asynchronous nature of many online support groups means that there is no pressure or commitment to engage in real-time communication with other members (Coulson & Knibb, 2007). Instead, members can take their time to read messages and reflect upon their feelings while considering if and

DOI: 10.4018/978-1-5225-2255-3.ch327

how to reply, allowing members to control their level of involvement in a way that is not possible with traditional face-to-face communication (Malik & Coulson, 2008). Indeed, it has been argued that asynchronous communication reduces the pressure associated with real-time communication thereby allowing participants time to carefully construct their messages before contributing to the group (Joinson, 2003).

Anonymity

Online support groups also offer a degree of anonymity that would not be possible in face-to-face communication. This may encourage certain individuals, particularly those patients who feel stigmatised because of their condition, to openly discuss their experiences without fear of a negative reaction.

Group Composition

Since participation in online support groups is not restricted by geographical, physical or spatial barriers, members can potentially access a larger and more heterogeneous mix of people with ease (White & Dorman, 2001). This is likely to be of particular benefit to patients suffering from chronic health conditions, as these individuals, may experience difficulties attending a regular face-to-face meeting due to limitations arising from their condition such as problems with mobility or treatment side effects. The Internet thus offers a novel opportunity for individuals to communicate with similar others in the comfort of their own home.

The ability to reach people from geographically diverse locations also increases the chances of finding others with similar experiences (White & Dorman, 2001). This is particularly helpful for individuals with rare conditions who may be unable to locate people with the same problem in their geographical area. At

the same time, due to their diversity, online support groups are advantageous because they can offer participants a wide variety of different perspectives, viewpoints and experiences on issues related to their condition (Wright, 1999; Wright, 2000). Walther and Boyd (2002) argue that for this reason online communication provides numerous opportunities for people with health concerns to seek support from 'weak tie' relationships. 'Weak ties' refer to relationships between people who might communicate on a regular basis but are not necessarily close. Prior to the advent of the Internet, weak tie networks typically incorporated neighbours, service providers or individuals known through other associations. However, the recent growth in the popularity of the Internet has led to a larger and more heterogeneous pool of individuals that could potentially develop into weak tie networks for people with health concerns.

Granovetter (1982) proposed that since weak ties provide access to a large pool of individuals with a range of different characteristics they are extremely useful sources of information. In addition, since weak tie relationships tend to exist outside of traditional social networks they provide a useful alternative for the discussion of taboo topics, which people may feel reluctant or uncomfortable discussing with their close relations. Online support groups, therefore, can potentially increase the chances of individuals forming weak tie relationships that offer access to diverse sources of information and advice that would not be available from close tie relationships. Furthermore, the absence of visual cues in the online environment and the opportunity to remain entirely anonymous may further facilitate the discussion of taboo or highly sensitive topics. Indeed, a frequently mentioned advantage of health-related online support groups is the perception that members attach less stigma to one's illness, thus removing the fear of being judged negatively (Walther & Boyd, 2001; Wright, 2000).

Adjunct Support

Online support groups can be a useful adjunct to support that is received through face-to-face support groups and other traditional sources (White & Dorman, 2001). There might be occasions when an individual is too ill to attend a face-to-face group, making online support particularly beneficial. Likewise, some individuals might have higher support needs than can be fulfilled by face-to-face groups, and online groups provide a way for them to receive additional support between face-to-face meetings.

Potential Disadvantages

Despite the many potential advantages of peer to peer online support groups, there exists a number of disadvantages (Finfgeld, 2000; White & Dorman, 2001), some of which appear to arise as a result of the computer-mediated nature of the communication.

Absence of Social Cues and Lack of Face-To-Face Contact

Within online support groups social cues including facial expressions, tone of voice or body language are not apparent and this can sometimes present unique challenges to group members (Coulson & Knibb, 2007). As a consequence, it is not uncommon for misunderstandings to arise when members misinterpret what has been posted by others, or feelings of awkwardness when members make incorrect assumptions about other members (Attard & Coulson, 2012). In order to compensate for these limitations, group members may employ emoticons or excessive punctuation in their communications (van Uden-Kraan, Drossaert, Taal, Lebrun et al., 2008a). In addition, the lack of physical proximity may potentially limit the extent to which a meaningful face-to-face relationship can develop between members as well as precluding the ability to offer a hug or touch someone (Coulson, 2013).

Anonymity

The anonymity inherent to the Internet could lead to deindividuation. Deindividuation is described by Festinger, Pepitone, and Newcomb (1952) as a state of lowered self-awareness where a person temporarily loses their individual sense of identity and becomes immersed into the identity of a larger group. Some authors argue that due to this deindividuation Internet behaviour will inevitably be characterised by increased instances of disinhibited communication such as aggressive and hostile exchanges often referred to as flaming.

Incorrect or Misleading Advice

In a number of studies, concerns have been identified by online support group members around the challenges of assessing the credibility and trustworthiness of information exchanged (Sillence, 2010). For example, it might be the case that some of the messages posted to an online support group may not in fact be from genuine patients living with a particular long-term condition. For example, it may be that an individual makes a series of postings to an online support group in order to advocate a specific treatment. In such a scenario, it may be that this individual is actually an employee of a drug company or some other organization who has a vested interest in the particular product being promoted (Coulson, 2013).

When an individual member within an online support group posts a message offering information or advice, it is not uncommon to see other members also replying to share their experiences and these responses may be helpful to the individual in terms of determining the credibility of the original information or advice. It has been argued that incorrect or misleading information may be more of a problem within online support groups with low levels of activity (Hwang et al., 2007), but broadly speaking it appears that the actual number of messages which contain incorrect or misleading information is low (Esquivel, Meric-Bernstam & Bernstam, 2006).

Delayed Feedback

Another disadvantage associated with online support group is delayed feedback (Wright, 2002). Although participants can log in and post messages at virtually any time of day, in comparison to face-to-face groups there may be a considerable time lag before the individual receives a response from other participants.

Individual Differences in Online Support Group Use

Socio-Demographic and Psychosocial Characteristics

Empirical data from Internet user surveys suggests that individuals who seek health-related information and support online tend to be white, female, well-educated, of a younger age and in employment (Eysenbach, 2003; Im et al., 2007). There is some evidence to indicate that these factors are also associated with participation in online support groups (Owen et al., 2010). For example, in a recent survey of cancer sufferers, Im et al. (2007) found that patients accessing online groups were typically middle-aged, well-educated, female and middle class. Similarly, Hoybye et al. (2010) found that cancer patients participating in online support groups belonged to higher socioeconomic groups compared to non-participants. Another recent survey of Internet use among individuals with HIV/AIDS, also revealed that frequent users of online support groups were more likely to be female, younger and single when compared to infrequent or non-users of online support groups (Mo & Coulson, 2010).

The association between patient's demographic characteristics and health-related Internet use may reflect barriers to Internet access among certain groups of people. For instance, individuals who are educated and in employment, may be more likely to have easier access to Internet resources compared to those who are unemployed with a lower educational record of attainment. Shaw

et al. (2006) examined whether demographic variables would predict participation in online support communities when barriers to Internet access were removed. A group of breast cancer patients were therefore provided with free Internet services, computer hardware and IT training. As hypothesised, the study found that demographic differences in participation were diminished. Nevertheless, the factors that predicted higher levels of participation were found to include having fewer breast cancer-related worries, a positive relationship with doctors and higher levels of social/family well-being. This implies that a range of psychosocial characteristics may influence participation in online groups.

However, there is comparatively little research specifically examining the relationship between patient's psychological characteristics and health-related Internet use, particularly the use of online support groups. Furthermore, the findings from the few studies (e.g., Fogel et al., 2002) that have addressed psychosocial factors are mixed. While some studies suggest that individuals turning to online groups report high levels of social support from offline relationships (Fogel et al., 2002), others have found that individuals seek online support when offline support is perceived as inadequate. For example, in a study comparing women with breast cancer who use the Internet for support and women with breast cancer who do not, Winefield et al. (2003) found that women using the Internet were significantly less satisfied with support from their family and were more psychologically distressed than non-users. Conversely, an Internet use survey conducted by Ybarra & Suman (2006) reported that those individuals who did not use the Internet for health-related purposes were more likely to report being unhappy or lonely. However, recent surveys by van Uden-Kraan et al. (2009) and Hoybye et al. (2010) found no relationship between psychological well-being and health-related Internet use among arthritis, cancer and fibromyalgia sufferers.

Since most studies employ cross-sectional designs it is difficult to determine causality. The

exact nature of the relationship between social support, distress and participation in online support groups is therefore unclear.

Level of Interaction

To date, little is known about the reasons why some members become active contributors of online groups, while other members are less involved. There are however several ideas which have emerged in the literature. For example, Kollock (1999) proposes three key motives to explain why some individuals contribute to an online group. Firstly, anticipated reciprocity, where an individual is motivated to offer help and support in the expectation that they will also receive the same support and information when required. Secondly, for a sense of efficacy, where members will contribute useful information or support to the other members in the belief that they are having a valuable impact on the group, which may help to promote their own self-efficacy. Thirdly, to gain increased recognition and help build up a reputation within the online group.

Another approach for understanding participation in online group is through the 'uses and gratification' model (Blumler & Katz, 1974). The basic premise of the model is that people are motivated to use mass communication in order to gratify a range of needs or goals. Thus, when considering level of participation in online groups, the theory would suggest that members are motivated to be passive or active participants depending on which activity offers the best means of meeting their needs. For example, an individual who wishes to obtain information whilst also maintaining anonymity, may find gratification in 'lurking' (Nonnecke & Preece, 2000). In comparison, individuals who wish to connect with other participants and develop new friendships will be motivated to take a more active role in the online group. Although the theory offers a useful way for examining participation in online groups, it has not yet been extensively applied to understand participation.

Posting Messages or 'Lurking'

Individuals who access online groups but do not participate are referred to as 'lurkers'. Lurking describes the act of 'reading messages posted by others on electronic spaces, without also posting one's own messages or in any way signalling one's vicarious observation' (Walther & Boyd, 2002, p168). It is estimated that 'lurkers' make up the largest group of individuals using online groups (Burnett, 2000).

Despite the prevalence of lurking, little is known about the online experiences of 'lurkers' and their motives for visiting online groups. This is in part due to the difficulty of identifying and accessing the views of individuals who choose to remain entirely anonymous. In an attempt to redress this, Preece, Nonnecke and Andrews (2004) surveyed 219 'lurkers' from 375 MSN bulletin boards to explore their reasons for not posting. The most frequently selected reason for lurking was 'just reading/browsing is enough' followed by 'still learning about the group', 'shy about posting', 'nothing to offer' and 'no requirement to post.' It was observed that many individuals lurked while familiarising themselves with the dynamics and norms of the group. Others commented that by not posting their views they felt they were being helpful (e.g., not pretending to be an expert).

Attention has also focussed on investigating differences between the attitudes of 'lurkers' and posters. Research findings indicate that posters perceive a greater sense of group membership in comparison to 'lurkers', who appear to be less positive about the benefits of group participation (Nonnecke, Andrews & Preece, 2006; Okleshen & Grossbart, 1998; Preece, Nonnecke & Andrews, 2004). It has been suggested that this might arise from the fact that members who do not actively participate in online discussions will often feel like distant outsiders who are simply observing a separate group of people interacting.

Some authors (e.g., Walther & Boyd, 2002) argue that individuals reading messages within health-related support groups are, on the contrary,

likely to feel empathy with the stories they read due to their shared experiences. As a result, these 'lurkers' may perceive a stronger sense of membership and connection to the group (Preece, 1999). Walther and Boyd (2002) propose that patients lurking in online support groups will also derive similar benefits to that of active participants. They argue that since patients can obtain information, support and validation of their feelings of stigma through reading other people's messages and without having to contribute anything themselves, lurking may be as beneficial as active participation. Lurking can also function as a helpful way for patients to find answers to frequently asked questions without overloading the group with a repetition of earlier queries.

However, there remains a scarcity of empirical work exploring the phenomenon of lurking in the context of health-related support groups. Recently, van Uden-Kraan et al. (2008b) investigated self-reported differences in the benefits associated with participation between 'lurkers' and posters in various online patient support groups. Results indicated that both 'lurkers' and posters gained a range of benefits from the online groups including feeling better informed, enhanced self-esteem, feeling more confident with physicians, and improved acceptance of the disease. Thus, offering support for Walther and Boyd's argument. However, in line with findings from other types of online groups, the study found that 'lurkers' were significantly less satisfied with the online group compared to posters and had poorer social and mental well-being. This suggests that although individuals can benefit from reading messages, individuals who lurk may be less satisfied with their online experiences than active participants and have poorer psychological well-being.

The Nature of Online Communication

What Do Members of Online Support Groups Discuss?

Research which has considered what members talk about within online support groups has provided

either a broad overview of discussion areas, such as symptoms, diagnoses, treatments and medications (Finn, 1999; Haker, Laubler & Rossler, 2005; Mursch & Behnke-Mursch, 2003; Perron, 2002; van Uden-Kraan et al., 2008a) or a richer insight into the specific health-related issues. A number of broad, recurring themes emerge from these studies. First, they demonstrate the importance of being able to communicate with others who can understand them and have been through similar experiences. Second, they demonstrate how people benefit from having a safe environment where they can vent and express their true feelings without the risk of hurting others or damaging social relationships. Third, they highlight the frustrations caused by misconceptions about what each illness is and how it affects sufferers. Fourth, the information, social support and advice shared within these groups can have an important impact on members by helping them to feel less alone and develop more effective coping strategies to overcome the issues unique to each illness.

Understanding the Therapeutic Processes Involved in Online Support Group Communication

There is a growing body of research which has attempted to understand the potential therapeutic processes at play within online support groups. These include: self-help mechanisms, social support and self-disclosure.

1. **Self-Help Mechanisms:** A small number of research studies have shown that online support groups can provide many of the therapeutic exchanges that occur in face-to-face self-help and support groups (Finn & Lavitt, 1994; Finn, 1999; Haker et al., 2005; Perron, 2002; Salem, Bogat, & Reid, 1997; Schielein, Schmid, Dobmeier, & Spiesl, 2008; van Uden-kraan et al., 2008a; Weinberg, Uken, Schmale, & Adamek, 1995). For example, Weinberg et al. (1995) investigated the presence of therapeutic factors in an online support group for women with breast cancer.

The study was guided by a range of factors which are thought to influence the process of change and recovery among patients in group therapy – universality, altruism, instillation of hope, imparting information, catharsis, group cohesiveness, interpersonal learning, imitative behaviour, developing social skills, existential factors and the corrective recapitulation of the primary family experience. Results revealed that women perceived the instillation of hope, group cohesion, and universality to be the most prevalent therapeutic factors. These factors were also moderately correlated with the perceived helpfulness of the group. Perron (2002) coined the term ‘self-help mechanisms’ to describe the parts of messages that facilitate the development of supportive or helping relationships among members of a group. Their analysis of communication in an online community for caregivers of people with a mental illness found that similar to face-to-face groups, members of the online group employed a range of self-help mechanisms. Specifically, self-disclosure and the provision of information or advice were found to be a central focus of many of the messages posted to the online group. Similar findings have been observed in studies examining self-help mechanisms within other illness support groups, for example, infertility (Malik & Coulson, 2010).

2. **Social Support:** A larger number of studies (e.g., Braithwaite et al., 1999; Coulson, Buchanan & Aubeeluck, 2007) have focussed on examining the type and nature of social support exchanged within online groups. Some of these studies employed content analyses as a means of quantifying the prominence of different categories of social support (e.g., Coulson, Buchanan, & Aubeeluck, 2007). One study analysed messages posted to an online disability support group for the presence of 5 categories of social support – informational, emotional, network, esteem

and tangible aid (Braithwaite et al., 1999). Of the 5 categories it was observed that messages exchanging emotional support, followed by informational support were most prominent, whilst tangible assistance was the least common category. The primary focus of these support messages appeared to relate to life functions affected by disability. For instance, informational support typically involved sharing advice and suggestions to help members tend to their self-care and house-keeping needs, while network support aimed to reduce limitations on mobility and access to others. In an analysis of messages posted to an Irritable Bowel Syndrome bulletin board, Coulson (2005) also noted that tangible assistance was the least visible category of social support, while the exchange of information concerned with symptom interpretation, illness management and interactions with health-care providers were the most common forms of support. Overall, research suggests that informational and emotional support appear to be the most prominent exchanges in health-related online support groups. Alongside the prevalence of emotional and informational support, a number of authors also identify features of online support that appear to be relatively unique to this format. For example, humor, personal narratives/journals and poetry have all been found to play an important role in supportive exchanges taking place in various online groups (Braithwaite et al., 1999; Hwang et al., 2010; White & Dorman, 2001). Indeed, Braithwaite et al. (1999) found that many members of a disability community joked about their own problems as a method of conveying empathy and in some instances used poetry to express emotional support.

3. **Self-Disclosure:** Along the development of strong interpersonal relationships, evidence suggests that Internet behaviour is also characterised by high levels of self-disclosure. Self-disclosure can be defined as the act

of revealing personal information to others. Computer-mediated discussions are thought to contain higher levels of disclosure compared to face-to-face discussions. This phenomenon is also apparent within online support groups (Barak & Gluck-Ofri, 2007). One factor that might explain the high levels of self-disclosure observed in these online groups is the anonymity associated with computer-mediated environments. Online support groups provide individuals with the opportunity to keep their identities entirely anonymous, this may encourage certain individuals to disclose information that they may otherwise feel uncomfortable discussing with others due to the possibility of a negative reaction. As Caplan and Turner (2007) point out, the anonymity offered by these groups may help to 'establish a safe and secure environment in which participants are willing to discuss deeply personal and upsetting matters' (p. 989). Tichon and Shapiro (2003) argue that self-disclosure plays an important role in the elicitation and provision of support. In their examination of messages posted to a group listserv for siblings of people with special needs and learning difficulties, self-disclosure was used in three distinct ways. Firstly, self-disclosure was used by participants to elicit support and start a relationship with other members, for example members would often share concerns and problems with the group to generate a discussion about ways to cope with their problems. Second, self-disclosure was used by support providers to demonstrate that successful coping is possible through disclosing similar experiences where they had successfully coped. Finally, self-disclosure was used to share social companionship with other members.

4. **Psychosocial Benefits of Participation:**

To date, there has been a lack of outcome research (Eysenbach et al., 2004) evaluating the efficacy of health-related online support

groups. Most quantitative studies in the literature focus on professionally developed online support groups or comprehensive computer support programmes. Comprehensive support programmes typically incorporate a wide array of interactive features including question and answer functions, decision-making and problem-solving aids, encyclopedia reference materials and individually tailored behaviour change support. Two of the most extensively researched comprehensive computer support systems for chronic health conditions are the Comprehensive Health Enhancement Support System (CHESS) and ComputerLink. The basic format of these interventions have been tailored to meet the needs of individuals suffering from various health conditions such as Breast cancer, Prostate cancer, Alzheimer's disease, Asthma and HIV and AIDS, and for patients from different socio-demographic backgrounds. Randomised controlled trials and observational studies assessing the impact of these computer systems indicate that usage may be associated with a variety of positive outcomes including improved emotional well-being and quality of life, more efficient use of health care systems and improved ability to cope (e.g., Gustafson et al., 2005; Gustafson et al., 2008). In addition, a number of studies note that the communication features of the support systems appeared to be the most heavily accessed functions (Gustafson et al., 1994). This suggests that participants may particularly value and benefit from computer support interventions that offer the opportunity to interact with other patients. While these studies show that computer-mediated patient support interventions can have beneficial effects for individuals, they tell us very little about the effectiveness of the millions of "natural" online support groups found on the Internet. Most studies evaluate complex interventions with a range of different components mak-

ing it difficult to distinguish the stand alone effect of online communication and peer-to-peer support. However, in a systematic review of the effects of health-related online groups, Eysenbach et al. (2004) found only six studies that evaluated pure peer-to-peer online communication. Most of these studies contained relatively small samples sizes and no comparison group, thus limiting the generalisability of the findings. The review therefore concluded that there was a lack of robust evidence for the psychological and health benefits of peer-to-peer online groups. Despite this, there is a growing body of qualitative literature which suggests that use of an online support groups can be associated with a range of positive psychosocial benefits for patients, particularly in the context of self-management behaviours (Allen et al., 2016).

5. **Empowerment:** More recent evidence suggests that participation in online support groups can have an empowering effect on patients suffering from physical and mental health problems (e.g., van Uden-Kraan et al., 2008c). The term empowerment describes a process through which people gain greater mastery and control over decisions and actions affecting their lives. In the context of health care, empowerment can reflect an increased confidence in one's ability to manage their condition, the ability to access appropriate health information and greater control over health-related decision-making and interactions with healthcare providers. The exact nature of patient empowerment however is complex; various factors can have an empowering effect including a patient's skills and knowledge, trust, locus of control and a reduction in feelings of hopelessness, isolation and oppression (van Uden-Kraan et al., 2008c). Different patients may also have different experiences and perspectives regarding empowerment. Furthermore, empowerment not only reflects an individual's

sense of control and efficacy but can also occur at the group and community level. There is also a distinction between empowering processes (i.e. the process through which individuals take control of their life) and empowering outcomes (i.e. the psychosocial consequences of empowerment). The majority of research documenting empowerment in online support groups has focused on exploring the empowering processes that take place in messages exchanged between participants (Finn, 1999; Perron, 2002; Sharf, 1997). For instance, in a discourse analysis of communication in an online breast cancer mailing list, Sharf (1997) reports that the online group allowed women to express feelings of grief and depression surrounding their experience of breast cancer whilst simultaneously offering a unique source of information and emotional support. This not only resulted in enhanced decision-making but also helped the women meet new challenges associated with breast cancer and transform their feelings of negativity into optimism.

CONCLUSION

Relatively little work has been conducted to examine the extent to which the empowering processes that occur online lead to empowering outcomes. Some qualitative studies have addressed this issue through exploring empowerment from the perspective of the patient. Buchanan and Coulson (2007) found that dentally anxious individuals accessing an online support group felt that the guidance, encouragement and shared experience offered by the online community empowered them to try to conquer their own fears. The support available from the group also led to a growth in confidence about disclosing dental anxiety to dental professionals. In another study, van Uden-Kraan et al. (2008c) asked participants using online breast cancer, arthritis and fibromyalgia support

groups about the full range of both empowering and disempowering processes that occur online. Results revealed that empowering processes included emotional support, information exchange, recognition, sharing experiences and helping others. These processes appeared to bring about a number of empowering outcomes namely feeling better informed, a greater sense of confidence about their treatment and interactions with health care professionals, improved acceptance of the disease, increased optimism and self-esteem and greater social well-being.

FUTURE RESEARCH DIRECTIONS

Over the past decade there has been a rapid growth in the number of studies which have considered peer-to-peer health-related online support groups. Despite this, there has been a notable lack of well-conducted randomized-controlled trials therefore there is an urgent need to redress this situation, particularly in the context of long-term physical health conditions. Beyond this, however, there is a need to more fully understand predictors of engagement with online support groups, factors impacting on levels of engagement as well as their linkage with relevant clinical outcomes. More broadly, if we are to fully understand and realize the potential of peer-to-peer health-related online support groups then a multi-disciplinary research perspective must be adopted.

REFERENCES

- Allen, C., Vassilev, I., Kennedy, A., & Rogers, A. (2016). Long-term condition self-management support in online communities: A meta-synthesis of qualitative papers. *Journal of Medical Internet Research*, 18(3), e61. doi:10.2196/jmir.5260 PMID:26965990
- Attard, A., & Coulson, N. S. (2012). A thematic analysis of patient communication in Parkinsons disease online support group discussion forums. *Computers in Human Behavior*, 28(2), 500–506. doi:10.1016/j.chb.2011.10.022
- Barak, A., & Gluck-Ofri, O. (2007). Degree and reciprocity of self-disclosure in online forums. *CyberPsychology & Behaviour*, 10(3), 407–417. doi:10.1089/cpb.2006.9938 PMID:17594265
- Blumler, J. G., & Katz, E. (1974). *The uses of mass communications: Current perspectives on gratifications research*. Beverly Hills, CA: Sage.
- Braithwaite, D. O., Waldron, V. R., & Finn, J. (1999). Communication of social support in computer-mediated groups for people with disabilities. *Health Communication*, 11(2), 123–151. doi:10.1207/s15327027hc1102_2 PMID:16370973
- Buchanan, H., & Coulson, N. S. (2007). Accessing dental anxiety online support groups: An exploratory qualitative study of motives and experiences. *Patient Education and Counseling*, 66(3), 263–269. doi:10.1016/j.pec.2006.12.011 PMID:17320336
- Burnett, G. (2000). Information exchange in virtual communities: A typology. *Information Research*, 5(4). Retrieved from <http://informationr.net/ir/5-4/paper82>
- Caplan, S. E., & Turner, J. S. (2007). Bringing theory to research on computer-mediated comforting communication. *Computers in Human Behavior*, 23(2), 985–998. doi:10.1016/j.chb.2005.08.003
- Coulson, N. S. (2005). Receiving social support online: An analysis of a computer-mediated support group for individuals living with irritable bowel syndrome. *Cyberpsychology & Behavior*, 8(6), 580–584. doi:10.1089/cpb.2005.8.580 PMID:16332169

- Coulson, N. S. (2013). How do online patient support communities affect the experience of inflammatory bowel disease? An online survey. *JRSM Open*, 4(8). doi:10.1177/2042533313478004 PMID:24040493
- Coulson, N. S., Buchanan, H., & Aubeeluck, A. (2007). Social support in cyberspace: A content analysis of communication within a Huntingtons disease online support group. *Patient Education and Counseling*, 68(2), 173–178. doi:10.1016/j.pec.2007.06.002 PMID:17629440
- Coulson, N. S., & Knibb, R. C. (2007). Coping with food allergy: Exploring the role of the online support group. *Cyberpsychology & Behavior*, 10(1), 145–148. doi:10.1089/cpb.2006.9978 PMID:17305463
- Esquivel, A., Meric-Bernstam, F., & Bernstam, E. V. (2006). Accuracy and self correction of information received from an Internet breast cancer list: Content analysis. *British Medical Journal*, 332(7547), 939–942. doi:10.1136/bmj.38753.524201.7C PMID:16513686
- Eysenbach, G. (2003). The impact of the Internet on cancer outcomes. *Cancer Journal for Clinicians*, 53(6), 356–371. doi:10.3322/can-jclin.53.6.356 PMID:15224975
- Eysenbach, G., Powell, J., Englesakis, M., Rizo, C., & Stern, A. (2004). Health related virtual communities and electronic support group: Systematic review of the effects of online peer to peer interactions. *British Medical Journal*, 328(7449), 1166–1171. doi:10.1136/bmj.328.7449.1166 PMID:15142921
- Ferguson, T. (1996). *Health online: How to find health information, support groups and self-help communities in cyberspace*. Reading, MA: Perseus.
- Festinger, L., Pepitone, A., & Newcomb, T. (1952). Some consequences of de-individuation in a group. *Journal of Abnormal and Social Psychology*, 47(2), 382–389. doi:10.1037/h0057906 PMID:14937978
- Finfgeld, D. L. (2000). Therapeutic groups online: The good, the bad, and the unknown. *Issues in Mental Health Nursing*, 21(3), 241–255. doi:10.1080/016128400248068 PMID:11075065
- Finn, J. (1999). An exploration of helping processes in an online self-help group focusing on issues of disability. *Health & Social Work*, 24(3), 220–231. doi:10.1093/hsw/24.3.220 PMID:10505283
- Finn, J., & Lavitt, M. (1994). Computer-based self-help/mutual aid groups for sexual abuse survivors. *Social Work with Groups*, 17(1-2), 21–46. doi:10.1300/J009v17n01_03
- Fogel, J., Albert, S. M., Schnabel, F., Ditkoff, B. A., & Neugut, A. I. (2002). Internet use and social support in women with breast cancer. *Journal of Health Psychology*, 21(4), 398–404. doi:10.1037/0278-6133.21.4.398 PMID:12090683
- Granovetter, M. S. (1982). The strength of weak ties: A network theory revisited. In P. V. Marsden & N. Lin (Eds.), *Social structure and network analysis* (pp. 105–130). Newbury Park, CA: Sage. doi:10.2307/202051
- Gustafson, D. H., Hawkins, R., McTavish, F., Pingree, S., & Chan, C. L. (1994). The use and impact of a computer-based support system for people living with AIDS and HIV infection. *Proceedings of the Annual Symposium on Computer Application in Medical Care*, 604–608.
- Gustafson, D. H., Hawkins, R., McTavish, F., Pingree, S., Chih Chen, W., & Volrathongchai, K. (2008). Internet-based interactive support for cancer patients: Are integrated systems better? *Journal of Communication*, 58(2), 238–257. doi:10.1111/j.1460-2466.2008.00383.x PMID:21804645
- Gustafson, D. H., Mctavish, F. M., Stengle, W., Ballard, D., Hawkins, R., Shaw, B. R., & Landucci, G. et al. (2005). Use and impact of eHealth system by low-income women with breast cancer. *Journal of Health Communication*, 10(sup1), 195–218. doi:10.1080/10810730500263257 PMID:16377608

- Haker, H., Lauber, C., & Rossler, W. (2005). Internet forums: A self-help approach for individuals with schizophrenia? *Acta Psychiatrica Scandinavica*, 112(6), 474–477. doi:10.1111/j.1600-0447.2005.00662.x PMID:16279878
- Hoybye, M. T., Dalton, S. O., Christenson, J., Ross, L., Kuhn, K. G., & Johansen, C. (2010). Social and psychological determinants of participation in Internet-based cancer support groups. *Supportive Care in Cancer*, 18(5), 553–560. doi:10.1007/s00520-009-0683-6 PMID:19579037
- Hwang, K. O., Farheen, K., Johnson, C. W., Thomas, E. J., Barnes, A. S., & Bernstam, E. V. (2007). Quality of Weight Loss Advice on Internet Forums. *The American Journal of Medicine*, 120(7), 604–609.e602. doi:10.1016/j.amjmed.2007.04.017 PMID:17602934
- Hwang, K. O., Ottenbacher, A. J., Green, A. P., Cannon-Diehl, M. R., Richardson, O., Bernstam, E. V., & Thomas, E. J. (2010). Social support in an Internet weight loss community. *International Journal of Medical Informatics*, 79(1), 5–13. doi:10.1016/j.ijmedinf.2009.10.003 PMID:19945338
- Im, E. O., Chee, W., Lim, H. J., Liu, Y., Guevara, E., & Kim, K. S. (2007). Patients attitudes toward Internet cancer support groups. *Oncology Nursing Forum*, 34(3), 705–712. doi:10.1188/07.ONF.705-712 PMID:17573329
- Joinson, A. (2003). *Understanding the psychology of Internet behaviour: Virtual worlds, real lives*. New York, NY: Palgrave Macmillan.
- Kollock, P. (1999). The economies of online cooperation: Gifts and public goods in cyberspace. In M. Smith & P. Kollock (Eds.), *Communities in cyberspace* (pp. 220–239). London, UK: Routledge.
- Malik, S., & Coulson, N. S. (2010). Examining self-help mechanisms in a web-based infertility support community. *Patient Education and Counseling*, 81(2), 315–318. doi:10.1016/j.pec.2010.01.007 PMID:20144521
- Malik, S. H., & Coulson, N. S. (2008). Computer-mediated infertility support groups: An exploratory study of online experiences. *Patient Education and Counseling*, 73(1), 105–113. doi:10.1016/j.pec.2008.05.024 PMID:18639409
- Mo, P. K. H., & Coulson, N. S. (2010). Living with HIV/AIDS and use of online support groups. *Journal of Health Psychology*, 15(3), 339–350. doi:10.1177/1359105309348808 PMID:20348355
- Mursch, K., & Behnke-Mursch, J. (2003). Internet-based interaction among brain tumour patients. Analysis of a medical mailing list. *Zentralblatt fur Neurochirurgie*, 64(02), 71–75. doi:10.1055/s-2003-40375 PMID:12838475
- Nonnecke, B., Andrews, D., & Preece, J. (2006). Non-public and public online community participation: Needs, attitudes and behavior. *Electronic Commerce Research*, 6(1), 7–20. doi:10.1007/s10660-006-5985-x
- Nonnecke, B., & Preece, J. (2000). *Lurker demographics: Counting the silent*. Paper presented at the ACM CHI 2000 Conference on Human Factors in Computer Systems, The Hague, The Netherlands.
- Okleshen, C., & Grossbart, S. (1998). Usenet groups, virtual community and consumer behaviors. *Advances in Consumer Research*. Association for Consumer Research (U. S.), 25, 276–282.
- Owen, J. E., Boxley, L., Goldstein, M. S., Lee, J. H., Breen, N., & Rowland, J. H. (2010). Use of health-related online support groups: Population data from the California health interview survey complementary and alternative medicine study. *Journal of Computer-Mediated Communication*, 15(3), 427–446. doi:10.1111/j.1083-6101.2010.01501.x
- Perron, B. (2002). Online support for caregivers of people with a mental illness. *Psychiatric Rehabilitation Journal*, 26(1), 70–77. doi:10.2975/26.2002.70.77 PMID:12171285

- Preece, J., Nonnecke, B., & Andrews, D. (2004). The top five reasons for lurking: Improving community experiences for everyone. *Computers in Human Behavior*, 20(2), 201–223. doi:10.1016/j.chb.2003.10.015
- Salem, A. D., Bogat, G. A., & Reid, C. (1997). Manual help goes on-line. *Journal of Community Psychology*, 25(2), 189–207. doi:10.1002/(SICI)1520-6629(199703)25:2<189::AID-JCOP7>3.0.CO;2-T
- Schielein, T., Schmid, R., Dobemeier, M., & Spiessl, H. (2008). Self-help from the cyberspace?—An analysis of self-help forums for patients with bipolar affective disorders. *Psychiatrische Praxis*, 35, 28–32. doi:10.1055/s-2006-952041 PMID:17546531
- Sharf, B. F. (1997). Communicating breast cancer on-line: Support and empowerment on the Internet. *Women & Health*, 26(1), 65–84. doi:10.1300/J013v26n01_05 PMID:9311100
- Shaw, B. R., Hawkins, R., McTavish, F., Pingree, S., & Gustafson, D. H. (2006). Effects of insightful disclosure within computer mediated support groups on women with breast cancer. *Health Communication*, 19(2), 133–142. doi:10.1207/s15327027hc1902_5 PMID:16548704
- Sillence, E. (2010). Seeking out very like-minded others: Exploring trust and advice issues in an online health support group. *International Journal of Web Based Communities*, 6(4), 376–394. doi:10.1504/IJWBC.2010.035840
- Tichon, J. G., & Shapiro, M. (2003). The process of sharing social support in cyberspace. *Cyberpsychology & Behavior*, 6(2), 161–170. doi:10.1089/109493103321640356 PMID:12804028
- van Uden-Kraan, C. F., Drossaert, C. H. C., Taal, E., Lebrun, C. E. I., Drossaers-Bakker, K. W., Smit, W. M., & van de Laar, M. A. F. J. et al. (2008a). Coping with somatic illnesses in online support groups: Do the feared disadvantages actually occur? *Computers in Human Behavior*, 24(2), 309–324. doi:10.1016/j.chb.2007.01.014
- van Uden-Kraan, C. F., Drossaert, C. H. C., Taal, E., Seydel, E. R., & van de Laar, M. A. F. J. (2009a). Participation in online patient support groups endorses patients empowerment. *Patient Education and Counseling*, 74(1), 61–69. doi:10.1016/j.pec.2008.07.044 PMID:18778909
- van Uden-Kraan, C. F., Drossaert, C. H. C., Taal, E., Seydel, E. R., & van de Laar, M. A. F. J. (2008b). Self-reported differences in empowerment between lurkers and posters in online patient support groups. *Journal of Medical Internet Research*, 10(2), e18. doi:10.2196/jmir.992 PMID:18653442
- van Uden-Kraan, C. F., Drossaert, C. H. C., Taal, E., Shaw, B. R., Seydel, E. R., & van de Laar, M. A. F. J. (2008c). Empowering processes and outcomes of participation in online support groups for patients with breast cancer, arthritis, or fibromyalgia. *Qualitative Health Research*, 18(3), 405–417. doi:10.1177/1049732307313429 PMID:18235163
- Walther, J. B., & Boyd, S. (2002). Attraction to computer-mediated social support. In C. A. Lin & D. J. Atkin (Eds.), *Communication, technology and society: Audience adoption and uses* (pp. 153–188). Hampton Press.
- Weinberg, N., Uken, J. S., Schmale, J., & Adamek, M. (1995). Therapeutic factors: Their presence in computer-mediated support group. *Social Work with Groups*, 18(4), 57–69. doi:10.1300/J009v18n04_05

White, M., & Dorman, S. M. (2001). Receiving social support online: Implications for health education. *Health Education Research*, 16(6), 693–707. doi:10.1093/her/16.6.693 PMID:11780708

Winefield, H. R., Coventry, B. J., Pradhan, M., Harvey, E., & Lambert, V. (2003). A comparison of women with breast cancer who do and do not seek support from the Internet. *Australian Journal of Psychology*, 55(1), 30–34. doi:10.1080/00049530412331312844

Wright, K. B. (1999). Computer-mediated support groups: An examination of relationships among social support, perceived stress, and coping strategies. *Communication Quarterly*, 47(4), 402–414. doi:10.1080/01463379909385570

Wright, K. B. (2000). The communication of social support within an on-line community for older adults: A qualitative analysis of the Senior-Net community. *Qualitative Research Reports in Communication*, 1, 33–43.

Wright, K. B. (2002). Social support within an on-line cancer community: An assessment of emotional support, perceptions of advantages and disadvantages, and motives for using the community from a communication perspective. *Journal of Applied Communication Research*, 30(3), 195–209. doi:10.1080/00909880216586

Wright, K. B., Bell, S. B., Wright, K. B., & Bell, S. B. (2003). Health-related Support Groups on the Internet: Linking Empirical Findings to Social Support and Computer-mediated Communication Theory. *Journal of Health Psychology*, 8(1), 39–54. doi:10.1177/1359105303008001429 PMID:22113899

Ybarra, M., & Suman, M. (2006). Help seeking behavior and the Internet: A national survey. *International Journal of Medical Informatics*, 75(1), 29–41. doi:10.1016/j.ijmedinf.2005.07.029 PMID:16129659

ADDITIONAL READING

Bartlett, K., & Coulson, N. S. (2011). An investigation into the empowerment effects of using online support groups and how this affects doctor/patient communication. *Patient Education and Counseling*, 83(1), 113–119. doi:10.1016/j.pec.2010.05.029 PMID:20599338

Bunde, M., Suls, J., Martin, R., & Barnett, K. (2006). Hystersisters online: Social support and social comparison among hysterectomy patients on the Internet. *Annals of Behavioral Medicine*, 31(3), 271–278. doi:10.1207/s15324796abm3103_9 PMID:16700641

Eysenbach, G., & Till, J. E. (2001). Ethical issues in qualitative research on Internet communities. *British Medical Journal*, 323(7321), 1103–1105. doi:10.1136/bmj.323.7321.1103 PMID:11701577

Malik, S., & Coulson, N. S. (2010). Examining self-help mechanisms in a web-based infertility support community. *Patient Education and Counseling*, 81(2), 315–318. doi:10.1016/j.pec.2010.01.007 PMID:20144521

Owen, J. E., Boxley, L., Goldstein, M. S., Lee, J. H., Breen, N., & Rowland, J. H. (2010). Use of health-related online support groups: Population data from the California health interview survey complementary and alternative medicine study. *Journal of Computer-Mediated Communication*, 15(3), 427–446. doi:10.1111/j.1083-6101.2010.01501.x

KEY TERMS AND DEFINITIONS

Deindividuation: When a person has a lower sense of self-awareness and 'loses' themselves within a larger group.

Lurking: When an individual does not actively post messages to a discussion forum but rather passively reads messages.

Online Support Group: A collection of individuals who come together online to engage in mutual support in the context of health and illness.

Self-Help Mechanisms: The therapeutic processes that may help oneself to address a challenge, crisis or difficult.

Social Support: The perception and actuality that one is cared for, has resources available from others and is part of a support network.

H

Software Evaluation From the Perspective of Patients and Healthcare Professionals

Rui Pedro Charters Lopes Rijo

Polytechnic Institute of Leiria, Portugal

Domingos Alves

University of São Paulo, Brazil

INTRODUCTION

Healthcare software evaluation is a complex process. Specifically, in the health information systems, focusing on the patients' health and on the healthcare professionals' motivation is particularly important.

Doctors, nurses and other healthcare professionals use software that indirectly affects the patients. Does software improve the patients' health, their satisfaction, or the healthcare professionals' commitment/job satisfaction? How can the impact of an information system be measured from the perspective of the patients, the doctors, the nurses or the supporting staff? Some relevant efforts have been made in the last years to measure healthcare software impact. Nevertheless, the decision to extend a study to different fields may lead to many difficulties as far as its conclusions are concerned. By identifying the research questions and the most relevant works, as well as indicating the open research issue, this article is a revision of the literature on the subject. This work may be expected to be useful to all those wishing to contribute by their research in this field. The article provides a historical perspective of the field, followed by the main research works, their goals, instruments and key results. A thorough discussion of the results, advances and challenges will also be developed, finishing the manuscript with fruitful research directions.

It is believed that this article will be of great value for those who need a solid insight in the

field and want to contribute for advances in the research questions identified.

The next section introduces the concepts related with the evaluation of the software, patients' satisfaction evaluation, and the healthcare professionals' satisfaction. The following section presents the dimensions of evaluation, tools and approaches used, ending with a discussion of issues and challenges of the assessment of patients' and healthcare professionals' satisfaction. Solutions and recommendations are preceding the discussion of future and emerging trends. Finally, the last section provides discussion of the overall coverage of the chapter and concluding remarks.

BACKGROUND

According to the World Health Organization (World Health Organization, n.d.), e-health can be defined as the providing of services and resources per electronic means either for health professionals, health consumers or for health systems management. Software plays in this way a central role in the healthcare systems because it can interact with the patients or doctors, or being embedded software in medical instrumentation, and other healthcare-oriented life-critical systems. Software is becoming more pervasive in all facets of medical device design and development. As transitions from hardware to software controls occur, there is a growing need for formalized software assurance processes (Cooper & Pauley, 2006).

Table 1. Challenges encountered at the intersection of healthcare services, information systems and evaluation technologies (Alalwany, 2010)

The Research Field	The Challenges Encountered
Healthcare services	• Healthcare services are characterized by having many stakeholders who are working in different disciplines and pursuing different goals. • Healthcare services are dictated by complex regulations. • The medical knowledge is an enormous and dynamic field. • The main aspects of medical knowledge require an interactive environment to be transferred or practiced.
Information Systems	• The multiple perspectives involved. • The difficulties of quantifying benefits. • The nature of information systems investments is changing both in terms of technological capability and the benefits they can deliver, as well as in terms of diffusion in most aspects of society. • Consider the social and technical context of use.
Evaluation Methodologies	• The limited experience of using methods. • The unfamiliarity with evaluation techniques. • The difficulty in interpreting results.

The use of the modern information technology in health offers tremendous opportunities: 1) to reduce clinical errors, e.g., medication errors, diagnostic errors; 2) to support health care professionals, e.g., availability of timely, up-to-date patient information; 3) to increase the efficiency of care, e.g., less waiting times for patients; 4) or even to improve the quality of patient care (Elske Ammenwerth, Gräber, Herrmann, Bürkle, & König, 2003)(Walker et al., 2005). According to Alalwany (2010), e-health supports disease allows people to better manage their own health and provides more accessible and consistent healthcare services, besides improving the efficiency of healthcare systems. Despite the opportunities, e-health presents some challenges namely considering that modern information technology systems are costly (Drouin, Hediger, & Henke, 2008)(Deloitte Touche Tohmatsu Limited, 2015) and that their failures may cause negative effects on both patients and staff. Adopting e-health applications is also highly complex. Finally, the privacy and security concerns in e-health applications are other key challenges. Therefore, a rigorous evaluation of IT in health care is recommended (Brender et al., 2013) and of great importance for decision makers and users (Kaplan & Shaw, 2004). Evaluation is likewise of some importance considering the high rate of failed IT projects in public sector where 35 percent of IT government projects have

been classified as total failures, and 50 percent as partial failures (Heeks, 2006).

As for evaluation it can be defined as the decisive assessment of defined objects, based on a set of criteria, to solve a given problem (Elske Ammenwerth et al., 2003). It is challenging evaluating software in health. The complexity and difficulty lie in the challenges faced at the intersection of three areas all well-known for their complexity (Hamid & Sarmad, 2010): healthcare services, information systems, and evaluation methodologies. Table 1 sums up the challenges encountered at the intersection of healthcare services, information systems and evaluation methodologies (Alalwany, 2010).

Ammenwerth *et al.* (2003) identify the typical problems in evaluation of IT in health care considering the complexity of the evaluation object, the evaluation project, and the motivation to perform evaluation.

E-health evaluation involved many stakeholders, the most important of them being the users (Gustafson & Wyatt, 2004). Therefore, assessing e-health from the users' perspective should address all the key factors that influence the users' acceptance of the new adopted technologies including the risks and benefits associated with the design and implementation of the e-health initiative in specific contexts. Two types of users, direct and indirect users, may be distinguished. In the first

domain, patients that use e-health applications and health care professionals that use healthcare software will be considered. In the second domain, indirect users are the patients that do not use directly the applications but are indirectly affected by the use of the systems in, among other examples, hospitals, home healthcare, or health centers. Therefore, it is important direct and indirect evaluations, which is why the next section will focus on the evaluation, dimensions to evaluate and tools and techniques.

EVALUTION, DIMENSIONS, TOOLS AND APPROACHES

Evaluation can occur at different phases of a software project: prior to the implementation of the e-health software, during and after the implementation. When evaluate at the pré-implementation, the main reason stands for maintain accountability for expenditure of resources and/or evaluation for decision-making. At the implementation, the rationale for doing evaluation is to assess the user's reaction to system or service and/or evaluation for decision making. Finally, at the post implementation, the intention is for performance evaluation and/or the impact evaluation to assess the overall social and technical impact of the system on users and organizations and/or usability and acceptability evaluations. Alawany (2010) proposes an evaluation roadmap for e-health evaluation with the main elements: 1) the rationale of e-health evaluation; 2) the time frames of evaluation; 3) the stakeholders; 4) the criteria; 5) the methods of e-health evaluation; 6) dealing with the ethical issues; 7) determination of how to interpret and present the evaluation outcomes.

Taking into consideration the current state of knowledge on this question, assessment by the users, either patients or healthcare professionals, occurs at a post implementation level. Hence, the focus of the present article is to focus on the post-implementation evaluation of the satisfaction of the patients and health professionals. It is

therefore essential to determine which dimensions are evaluated and what instruments are used.

What Really Matters for Patients?

What really matters for patients is the starting point in order to determine what should be evaluate when considering a software evaluation. One can consider the perspective from either the tele-healthcare, the software engineering or the healthcare institutions. The analysis of these perspectives may help to understand what must be performed by software in order to meet, or even increase, patients' satisfaction, either directly or indirectly.

Regarding the patients' satisfaction (Williams, May, & Esmail, 2001), the aspects of satisfaction measured by existing studies in tele-healthcare are: 1) professional-patient interaction; 2) patient's experience/feeling; 3) overall satisfaction; 4) technical aspects; 5) preferences for telemedicine or face-to-face consultation; 6) convenience; 7) administrative aspects; and 8) physical environment.

From the software engineering perspective, the approaches to measure the usability of a software using questionnaires as the Computer System Usability Questionnaire (CSUQ) and the Post-Study System Usability Questionnaire (PSSUQ) (Lewis, 1995) the aspects measured are: 1) the overall system usability (OVERALL); 2) the system usability (SYSUSE); 3) the quality of the information (INFOQUAL); 4) the quality of the interface (INTERQUAL).

Existing instruments to measure patients' satisfaction at healthcare institutions such as the Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) (National Quality Forum, 2013) or the Patient Satisfaction Questionnaire Short Form (PSQ-18) (Marshall & Hays, 1994) use both similar and different dimensions. The Centers for Medicare & Medicaid Services (CMS), along with the Agency for Healthcare Research and Quality (AHRQ), developed the HCAHPS Survey, also known as Hospital CAHPS®, to provide a standardized survey

Table 2. Scotland's Patient Experience Programme: Patient Priorities for Inpatient Care ((Reeves & Bruster, 2009), from 2,213 patients)

10 Most Important Items	10 Least Important Items
• A clean ward • Staff cleaning their hands before touching patients • Being treated quickly in an emergency • Getting the best treatment for my condition • Doctors knowing enough about my condition and treatment • Clear explanations about what will happen during an operation or procedure • Being told the risks and benefits of any treatment in a way I can understand • Clear explanations of my condition or treatment • Being treated with dignity and respect • Being told how my operation or procedure has gone in a way I can understand	• Being given help to eat my meals when I need it • Being given accurate information about ward routines • Not having to wait around in hospital once I have been told I can go home • Being given help to arrange transport home • Having a choice about which hospital I go to • Not having to pay too much to make phone calls • My religious beliefs being respected • Being able to watch television without having to pay for it • Being able to get an interpreter • Having access to food when I am hungry (not just at mealtimes)

instrument and data collection methodology for measuring patients' perspectives on hospital care. HCAHPS has composite measures, individual items, and global items. Composite measures are communication with nurses, communication with doctors, responsiveness of hospital staff, pain management, communication about medicines, discharge information, and care transition. Individual items concern the cleanliness and the quietness of hospital environment. Finally, global items relate with the overall hospital rating and whether the patient would recommend the hospital. PSQ-18, in its turn, presents separate scores for each of seven different subscales: 1) general satisfaction; 2) technical quality; 3) interpersonal manner; 4) communication; 5) financial aspects; 6) time spent with doctor; 7) and accessibility and convenience.

Cornwell & Robert (2011) in the study of "*What matters to patients'?*", found consistency both across conditions and in relation to the various sectors. Generic themes were:

- Good information provision;
- Having confidence in health professionals;
- Awareness and understanding of specific health condition;
- The right treatment from the right staff at the right time;

- Continuity of care;
- Being treated as a person;
- Partnership with professionals.

Table 2 lists the ten most and least important items for the patients identified Scotland's survey. The findings of the survey broadly support the argument that a generic framework of 'what matters' to patients could be applied across conditions and sectors.

Main Approaches Used for Patients' Satisfaction Assessment

Knowing the key dimensions of patients' satisfaction, as well as the methods used for their assessment, is of great importance.

Regarding e-health (Williams et al., 2001) the majority of studies are descriptive, regarding one service only, some comparing telemedicine services with standard healthcare using the same sample of patients. Only 20% of the studies included an independent control group against which telemedicine was compared. The research was mainly quantitative, with some studies including at least some qualitative component. Tables 3 sums up the methods used for the assessment from the perspective of the healthcare organizations.

Table 3. Methods including tools/surveys for the assessment of patients' satisfaction. Adapted from (Williams et al., 2001).

Author, year	Methods (incl. Tools/Surveys)
(Boyd, 2007)	• Postal questionnaire consisting of 82 statements using a 5-point scale. 54 of the statements were based on core items from the 2006 inpatient survey, 12 from the question bank for the inpatient surveys and 16 were created at the request of the DH. • Cognitive interviews with 8 ethnically diverse volunteers with recent experience of inpatient stays in England were conducted to test the questionnaire.
(Bridges, Flatley, & Meyer, 2010)	• Systematic review of worldwide grey and published literature written in English between January 1999 and June 2008 identified from databases: CINAHL, Medline, British Nursing Index, EMBASE Psychiatry, International Bibliography of the Social Sciences, PsychINFO, and AgeInfo. • 42 primary studies and 1 systematic review met the inclusion criteria.
(Coulter, 2006)	• Expert opinion
(Davies, 2005)	• Qualitative interviews with senior health professionals and managers and a review of the literature
(The NHS Confederation, 2010)	• Seven organisational case studies

Assess Health Care Professional Satisfaction

Doctors and nurses are the main users of software health applications. Studies consider these users as software users but don't measure the possible relation between the use of the applications and the impact on job satisfaction. The only existing job satisfaction scale specifically aimed at community nurses defined was Ellenbecker's Home Healthcare Nurses' Job Satisfaction (HHNJS) scale. The 2001 edition was composed of eight subscales (21 items, 5-point Likert scale): general feeling of overall job satisfaction, interaction, stress, professionalism, task requirements, autonomy, organizational policy and quality of care delivered. Nine subscales were presented in the 2005 (Ellenbecker & Byleckie, 2005) edition (32 items, 5-point Likert scale): autonomy/independence, professional growth, group cohesion/peers, group cohesion/physician, characteristics of the organization, stress/work load, autonomy/flexibility in work scheduling, autonomy/control of work activities and salary/benefits/perception of opportunities elsewhere. So far, no scale to assess doctors' professional satisfaction has been known.

Users' Acceptance

According to (Ullah, Fiedler, & Wac, 2012), the most important determinants for successful e-health implementations are shown in Table 4. As can be seen, user acceptance is the most important factor for any healthcare service implementation

Technology acceptance model was designed to be used in the context of Information Systems for the purpose of predicting information technology acceptance and usage in the work environment. The final conceptualization of TAM model (Davis, 1989) includes two core constructs which are defined as follows: 1) perceived usefulness, i.e., "the

Table 4. Determinants for successful e-Health implementation (Ullah et al., 2012)

Determinants	Influence in Implementation (%)
User acceptance	37%
Technology deployment factors	29%
Leaving behind the organization	13%
Policy and Legislation	11%
Financing factors	10%

degree to which a person believes that using a particular system would enhance his or her job performance”, 2) perceived ease of use: “the degree to which a person believes that using a particular system would be free of effort”. TAM therefore identifies two factors: the perceived ease of use (PEOU) and perceived usefulness (PU) that determine individuals’ attitude (ATT), which influences individuals’ behavioural intention. The Technology Acceptance Model (TAM) is one technology acceptance theory that can be used to understand perceived barriers and benefits of using technology in health care settings. TAM can help predict acceptance of health information technology among users. The systematic review by Holden & Karsh (2010) shows over twenty studies that applied TAM to health care settings by presenting the views of physicians and non-physicians, among which nurses, pharmacists, and patients. The study found that TAM can forecast accurately the portion of Health IT users that are accepting the system. In 2000, Venkatesh and Davis (Venkatesh & Davis, 2000) developed TAM2 in order to capture the social influence (SN) that influences individuals to positively or negatively accept such IT. In 2003, Venkatesh et al. formulated a new model called the Unified Theory of Acceptance and Use of Technology (UTAUT) (Venkatesh, Morris, Davis, & Davis, 2003) that was extended in 2012 to UTAUT2 (Venkatesh, Viswanath and Thong, James Y. L. and Xu, 2012). UTAUT introduces three constructs, which are the main determinants of intention to use an information technology. These are defined as performance expectancy, effort expectancy and social influence. UTAUT2 incorporates three constructs into UTAUT: hedonic motivation, price value, and habit. Compared to UTAUT, the extensions proposed in UTAUT2 produced a substantial improvement in the variance explained in behavioral intention and technology use.

Existing Studies to Assess the Impact of Software in the Patients and Healthcare Professionals

H

A nursing record system is the record of care planned and/or given to individual patients/clients by qualified nurses or other caregivers under the direction of a qualified nurse (Currell & Urquhart, 2003). The goal of the study was to assess the effects of nursing record systems on nursing practice and patient outcomes. Eight trials involving 1497 people were included. In three studies of client held records, there were no overall positive or negative effects, although some administrative benefits through fewer missing notes were suggested. A paediatric pain management sheet study showed a positive effect on the children’s pain intensity. A computerised nursing care planning study showed a negative effect on documented nursing care planning, although two other computerised nursing information studies showed an increase in recording but no change in patient outcomes. Care planning took longer with these computerised systems, but the numbers of patients and nurses included in these studies was small. A controlled before-and-after study of two paper nursing record systems showed improvement in meeting documentation standards. No evidence was found of effects on practice attributable to changes in record systems. The number of studies on this topic, albeit of poor methodological quality, indicates that this is an important area of concern for nursing practice (Currell & Urquhart, 2003). However, none of the studies produced results that could be interpreted as evidence of a change in practice that resulted from a change of record system. The studies indicate that nurses believe that there is a link between their clinical practice and the nursing record, beyond its obvious use as a note keeping system. The client held records studies suggest altered and improved relationships between clients and health professionals, and the pain management study sought to demonstrate a direct effect on patient care.

Currell, Urquhart, Wainwright, & Lewis (2010) developed a study to assess the effects of telemedicine as an alternative to face-to-face patient care. Establishing systems for patient care using telecommunications technologies is feasible, but there is little evidence of clinical benefits. The studies provided variable and inconclusive results for other outcomes such as psychological measures, and no analyzable data about the cost effectiveness of telemedicine systems. The review demonstrates the need for further research and the fact that it is feasible to carry out randomized trials of telemedicine applications. Policy makers should be cautious about recommending increased use and investment in unevaluated technologies.

Sinha & Shetty (2015) show the relevance of the attitude of users, namely doctors, in the impact of health information technology. Health information technology (HIT) equips healthcare professionals with the required information and tools for making quality decisions in patient care, but it is always advisable to assess their attitude before its actual implementation. A cross-sectional survey was carried out among 140 doctors of an Ayurvedic center of Southern India. A validated questionnaire consisting of 18 questions based on a 5-point Likert scale was administered to the participants after receiving their due consent. About 75–80% of the respondents concurred that the HIT application, such as electronic health record, has the potentials to reduce the duplication of documentation work, is easy and has an instant processing and real-time access to patient information. They also felt the need of such application to report the patient data to local and national health institutions. A total of 85% of them mentioned that these applications can make the collection and accessibility of patient data much easier compared with paper-based records, whereas 87.4% of them claimed telemedicine as a platform for multidisciplinary collaborative research and patient care. Even though most of the respondents agreed about the role of HIT in improving the quality of health

care, there were many who held no opinion about HIT, including privacy and security of patient data. The need of proper awareness and training program is identified to make them aware about the HIT and its application in patient care, education and research.

Main Approaches Used for Professionals' Satisfaction Assessment

The types of study designs are (Currell & Urquhart, 2003) (Williams et al., 2001) (E Ammenwerth & Keizer, 2005; E. Ammenwerth et al., 2004; Brender et al., 2013): 1) randomized controlled trials; 2) controlled clinical trials; 3) interrupted time series analyses; 4) and controlled before and after studies.

Issues, Controversies, Problems

Studies indicate that many of the main barriers to healthcare information systems (specifically referencing electronic health records) are not technical, but other issues including stakeholder issues such as risk tolerance, physicians' resistance related to time concerns, fears about privacy, the number of sellers vendors on the marketplace and their transience (Walker et al., 2005). Each of the aforementioned barriers is rooted in the associated vested issues in health technologies held by stakeholders. Stahl and Shaw (2011) indicate that success with health information systems needs to bring together various stakeholders and their work practices. Ultimately, Patient Centered Care (PCC) success is determined by the quality of the interactions between patients and clinicians (Epstein, Fiscella, Lesser, & Stange, 2010). A fundamental objective of the PCC model is therefore to integrate information technology applications and infrastructures in every link of the care chain in the belief that Web portals, shared patient records, electronic consultation systems, and online data access for patients facilitate relationships between

professionals and patients by providing sufficient information, patient engagement and mutual feedback (Vikkelsø, 2010). In this context, assessing the impact of the e-health software in patients and healthcare professionals is critical. Despite a large amount of published evaluation studies, many authors report problems during evaluation, such as (E Ammenwerth & Keizer, 2005):

- Unclear, conflicting or changing evaluation goals during the study;
- Large efforts needed for the preparation and execution of the study;
- Complex and sometimes contradictory results;
- Dependence of the evaluation results on the motivation and expectation of the users;
- Uncertainty as to whether results can be generalized to other environments.

Reviews of patient satisfaction research in health care highlight several methodological problems (Williams et al., 2001), including the use of short quantitative questionnaires, lack of standardization, and confounding due to a range of social psychological biases, including self-selection bias, social desirability response bias, and acquiescent replies. Answers to patient satisfaction surveys conducted at healthcare delivery sites are undoubtedly influenced to some degree by the patient's concerns about the impact that any negative feedback may have on the services that are available to them. The results of patient satisfaction studies are often influenced by the method of data collection. Other factors that influenced patient satisfaction included the choice of response format for rating satisfaction, order of placement of satisfaction items within the overall questionnaire, and the distinction between open-ended and closed or dichotomous questions. These findings suggest that "patient satisfaction" is influenced, at least in part, by the nature of the methodology used to measure it.

SOLUTIONS AND RECOMMENDATIONS

H

E-health software satisfaction assessment is difficult, time consuming and resource demanding. Existing approaches are designed for a very specific scenario, so the existing results are difficult to generalize and do not allow a comparison between software applications. One possible solution is the development of a global validated tool and method for each sectorial health service allowing in this way the benchmarking of the applications. A validated instrument and their application method leading to a unique scale of satisfaction impact assessment could be used in decision-making and impact estimation of a commercial e-health product, the development of new products, and as the backbone of regulatory work. Finally, as stated in this work, these instruments must encompass a holistic approach for measuring the satisfaction either of the patients or of the healthcare professionals.

FUTURE RESEARCH DIRECTIONS

Due to the importance of the evaluation of the software from the perspective of patients and healthcare professionals many fruitful research directions exist and the academic and non-academic community is invited to develop efforts for contributions in the field. This work shows the need to develop standard and holistic tools and methods to perform this assessment at least at the health service level. It also, shows the need of tools and methods easy and faster to apply. The heavy investments in e-health request the existence of an impact scale in the patients and professionals that would allow the comparison between software applications. Moreover, the presented studies are time consumer, requiring a lot of availability either of the research teams either of the patients and health professionals. The existing instruments neither take into account many dimensions of

patients' satisfaction nor consider jobs satisfaction in its equation. Finally, there's a lack of tools and methods to perform satisfaction assessment prior to the implementation of the e-health software and during the implementation.

CONCLUSION

The paper provides a thorough discussion about the relevance of the evaluation of the software from the perspective of the patient and healthcare professionals. It holistically presents the concept of satisfaction of the patients and the professionals, with dimensions and methods both measured. It demonstrates the need to develop new efforts in this research area due to the existing fragilities. The paper may be a helpful tool for those who need a solid support from where to start or to continue their work.

REFERENCES

- Alalwany, H. (2010). *Cross disciplinary evaluation framework for e-health services*. Brunel University. Retrieved from <http://bura.brunel.ac.uk/handle/2438/8216>
- Ammenwerth, E., Brender, J., Nykanen, P., Prokosch, H.-U., Rigby, M., & Talmon, J. (2004). HIS-EVAL workshop participants. Visions and strategies to improve evaluation of health information systems: Reflections and lessons based on the HIS-EVAL workshop in Innsbruck. *International Journal of Medical Informatics*, 73(6), 479–491. doi:10.1016/j.ijmedinf.2004.04.004 PMID:15171977
- Ammenwerth, E., & De Keizer, N. (2005). An Inventory of Evaluation Studies of Information Technology in Health Care: Trends in evaluation research 1982–2002. *Methods of Information in Medicine*, 44, 44–56. Retrieved from <http://www.lina-schwab.de/Publikationen/z30.pdf> PMID:15778794
- Ammenwerth, E., Gräber, S., Herrmann, G., Bürkle, T., & König, J. (2003). Evaluation of health information systems—problems and challenges. *International Journal of Medical Informatics*, 71(2-3), 125–135. doi:10.1016/S1386-5056(03)00131-X PMID:14519405
- Bevan, N. (1995). Measuring usability as quality of use. *Software Quality Journal*, 4(2), 115–130. doi:10.1007/BF00402715
- Boyd, J. (2007). *The 2006 inpatients importance study*. Academic Press.
- Brender, J., Talmon, J., de Keizer, N., Nykänen, P., Rigby, M., & Ammenwerth, E. (2013). STARE-HI – Statement on Reporting of Evaluation Studies in Health Informatics. *Applied Clinical Informatics*, 4(3), 331–358. doi:10.4338/ACI-2013-04-RA-0024 PMID:24155788
- Bridges, J., Flatley, M., & Meyer, J. (2010). Older peoples and relatives experiences in acute care settings: Systematic review and synthesis of qualitative studies. *International Journal of Nursing Studies*, 47(1), 89–107. doi:10.1016/j.ijnurstu.2009.09.009 PMID:19854441
- Cooper, J. G., & Pauley, K. A. (2006). Healthcare software assurance. *AMIA... Annual Symposium Proceedings / AMIA Symposium*, 166–70. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/17238324>
- Cornwell, J., & Robert, G. (2011). *What matters to patients? Developing the evidence base for measuring and improving patient experience. Project Report for the Department of Health and NHS Institute for Innovation & Improvement*. Retrieved from http://www.institute.nhs.uk/images/Patient_Experience/Final Project Report pdf doc january 2012.pdf
- Coulter, A. (2006). Can patients assess the quality of health care? *BMJ (Clinical Research Ed.)*, 333(7557), 1–2. doi:10.1136/bmj.333.7557.1 PMID:16809674

- Currell, R., & Urquhart, C. (2003). Nursing record systems: effects on nursing practice and health care outcomes. In R. Currell (Ed.), *The Cochrane Database of Systematic Reviews*. Chichester, UK: John Wiley & Sons, Ltd.; doi:10.1002/14651858.CD002099
- Currell, R., Urquhart, C., Wainwright, P., & Lewis, R. (2010). Telemedicine versus face to face patient care: effects on professional practice and health care outcomes. *Cochrane Database of Systematic Reviews*, (2), CD002098. doi:10.1002/14651858.CD002098
- Davies, E. (2005). Hearing the patients voice? Factors affecting the use of patient survey data in quality improvement. *Quality & Safety in Health Care*, 14(6), 428–432. doi:10.1136/qshc.2004.012955 PMID:16326789
- Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *Management Information Systems Quarterly*, 13(3), 319. doi:10.2307/249008
- Deloitte Touche Tohmatsu Limited. (2015). *2015 Global health care outlook. Common goals, competing priorities*. Author.
- Drouin, J. P., Hediger, V., & Henke, N. (2008). *Health care costs: A market-based view*. The Mackinsey Quarterly.
- Ellenbecker, C. H., & Byleckie, J. J. (2005). Home Healthcare Nurses Job Satisfaction Scale: Refinement and psychometric testing. *Journal of Advanced Nursing*, 52(1), 70–78. doi:10.1111/j.1365-2648.2005.03556.x PMID:16149983
- Epstein, R. M., Fiscella, K., Lesser, C. S., & Stange, K. C. (2010). Why the nation needs a policy push on patient-centered health care. *Health Affairs (Project Hope)*, 29(8), 1489–1495. doi:10.1377/hlthaff.2009.0888 PMID:20679652
- Gustafson, D. H., & Wyatt, J. C. (2004). Evaluation of ehealth systems and services. *BMJ (Clinical Research Ed.)*, 328(7449), 1150. doi:10.1136/bmj.328.7449.1150 PMID:15142895
- Hamid, A., & Sarmad, A. (2010). Towards an Evaluation Framework for E-health Services. In *Handbook of Research on Advances in Health Informatics and Electronic Healthcare Applications* (pp. 1–16). IGI Global. doi:10.4018/978-1-60566-030-1.ch001
- Heeks, R. (2006). Health information systems: Failure, success and improvisation. *International Journal of Medical Informatics*, 75(2), 125–137. doi:10.1016/j.ijmedinf.2005.07.024 PMID:16112893
- Holden, R. J., & Karsh, B.-T. (2010). The Technology Acceptance Model: Its past and its future in health care. *Journal of Biomedical Informatics*, 43(1), 159–172. doi:10.1016/j.jbi.2009.07.002 PMID:19615467
- IEEE P730™/D8 Draft Standard for Software Quality Assurance Processes. (2012). IEEE Computer Society/Software & Systems Engineering Standards (C/S2ESC) Committee of the IEEE Computer Society. Retrieved from <https://www.yumpu.com/en/document/view/9939479/ieee-p730tm-d8-draft-standard-for-software-quality-assurance-ISO/IEC>
- Kaplan, B., & Shaw, N. T. (2004). Future directions in evaluation research: People, organizational, and social issues. *Methods of Information in Medicine*, 43(3), 215–231. doi:10.1267/METH04030215 PMID:15227551
- Lewis, J. R. (1995). IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. *International Journal of Human-Computer Interaction*, 7(1), 57–78. doi:10.1080/10447319509526110

- Marshall, G., & Hays, R. (1994). The Patient Satisfaction Questionnaire Short-Form (PSQ-18). Santa Monica, CA: Rand Corporation. Retrieved from <http://www.rand.org/content/dam/rand/pubs/papers/2006/P7865.pdf>
- Meyer, J. P., Kam, C., Goldenberg, I., & Bremner, N. L. (2013). Organizational commitment in the military: Application of a profile approach. *Military Psychology*, 25(4), 381–401. doi:10.1037/mil0000007
- National Quality Forum. (2013). *Hospital Consumer Assessment of Healthcare Providers and Systems (H-CAHPS)*. Retrieved August 13, 2015, from <http://www.hcahpsonline.org/home.aspx>
- Reeves, R., & Bruster, S. (2009). *Better Together: Scotland's Patient Experience Programme: Patient Priorities for Inpatient Care Report No. 5/2009*. Retrieved from <http://www.gov.scot/resource/doc/278973/0083963.pdf>
- Shaw, M. C., & Stahl, B. C. (2011). On Quality and Communication: The Relevance of Critical Theory to Health Informatics. *Journal of the Association for Information Systems*, 12(3), 255–273.
- Sinha, R., & Shetty, S. (2015). Attitude of Ayurveda doctors toward the impact of health information technology. *Journal of Ayurveda and Integrative Medicine*, 6(2), 89. doi:10.4103/0975-9476.146559 PMID:26166998
- The N. H. S. Confederation. (2010). *Feeling better? Improving patient experience in hospital The voice of NHS leadership*. Retrieved from http://www.nhsconfed.org/~media/Confederation/Files/Publications/Documents/Feeling_better_Improving_patient_experience_in_hospital_Report.pdf
- Ullah, M., Fiedler, M., & Wac, K. (2012). On the ambiguity of Quality of Service and Quality of Experience requirements for eHealth services. In *2012 6th International Symposium on Medical Information and Communication Technology (ISMICT)* (pp. 1–4). IEEE. doi:10.1109/ISMICT.2012.6203030
- van Dick, R., & Ullrich, J. (2013). Identifikation und Commitment. In W. Sarges (Ed.), *Management Diagnostik* (4th ed.; pp. 349–354). Göttingen: Hogrefe.
- Venkatesh, V., & Davis, F. D. (2000). A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. *Management Science*, 46(2), 186–204. doi:10.1287/mnsc.46.2.186.11926
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User Acceptance of Information Technology: Toward a Unified View. *Management Information Systems Quarterly*, 27(3).
- Venkatesh, V., Thong, J. Y. L., & Xu, X. (2012). Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology. *Management Information Systems Quarterly*, 36(1), 157–178.
- Vikkelsø, S. (2010). Mobilizing information infrastructure, shaping patient-centred care. *International Journal of Public Sector Management*, 23(4), 340–352. doi:10.1108/09513551011047233
- Walker, J., Pan, E., Johnston, D., Adler-Milstein, J., Bates, D. W., & Middleton, B. (2005). The value of health care information exchange and interoperability. *Health Affairs (Project Hope), Suppl Web*, W5–10–W5–18. doi:10.1377/hlthaff.w5.10
- Williams, T. L., May, C. R., & Esmail, A. (2001). Limitations of Patient Satisfaction Studies in Telehealthcare: A Systematic Review of the Literature. *Telemedicine Journal and e-Health*, 7(4), 293–316. doi:10.1089/15305620152814700 PMID:11886667
- World Health Organization. (n.d.). *World Health Organization. E-Health definition*. Retrieved March 28, 2016, from <http://www.who.int/trade/glossary/story021/en/>

KEY TERMS AND DEFINITIONS

e-Health: According to the World Health Organization (World Health Organization, n.d.), e-health is: “*the transfer of health resources and health care by electronic means. It encompasses three main areas: 1) the delivery of health information, for health professionals and health consumers, through the Internet and telecommunications; 2) the use of the power the power of IT and e-commerce to improve public health services, e.g. through the education and training of health workers; 3) the use of e-commerce and e-business practices in health systems management*”.

Hospital Consumer Assessment of Health-care Providers and Systems (HCAHPS): A standardized survey instrument and data collection methodology for measuring patients’ perspectives on hospital care. The HCAHPS survey contains 21 patient perspectives on care and patient rating items that encompass nine key topics: communication with doctors, communication with nurses, responsiveness of hospital staff, pain management, communication about medicines, discharge information, cleanliness of the hospital environment, quietness of the hospital environment, and transition of care. The survey also includes four screener questions and seven demographic items, which are used for adjusting the mix of patients across hospitals and for analytical purposes. The survey is 32 questions in length (National Quality Forum, 2013).

Organizational Commitment: Organizational commitment refers to the extent to which the employees of an organization see themselves as belonging to the organization (or parts of it) and feel attached to it (Meyer, Kam, Goldenberg, & Bremner, 2013)(van Dick & Ullrich, 2013).

Patient Satisfaction Questionnaire Short Form (PSQ-18): PSQ-18, Likert scale questionnaire proposes seven dimensions of patient satisfaction directed toward their healthcare professionals. These are 1) general satisfaction, 2) technical quality, 3) interpersonal manner, 4) communication, 5) financial aspects, 6) time spent with doctor, and 7) accessibility and convenience. Each domain is tested through different related questions, which is of substantial benefit when one aims to identify a particular area to improve on. The general satisfaction has strong correlation with the other domains and thus it is important to improve in all (Marshall & Hays, 1994).

Software Quality: The “*capability of software product to satisfy stated and implied needs under specified conditions*” (“ISO/IEC 25010:2011 Systems and software engineering -- Systems and software Quality Requirements and Evaluation (SQuaRE) -- System and software quality models,” 2011) and as “*the degree to which a software product meets established requirements; however, quality depends upon the degree to which those established requirements accurately represent stakeholder needs, wants, and expectations*” (IEEE P730™/D8 Draft Standard for Software Quality Assurance Processes, 2012).

Usability: Usability relates with the concept of ease of use and the ability to use a product for its intended purpose (Bevan, 1995).

User Acceptance: For technology to improve productivity in organizations first users’ must accept the technology and use it. Models explaining the factors and predicting user intention for using technology are of great importance for health information systems development (Venkatesh et al., 2003).

Technology Design and Routes for Tool Appropriation in Medical Practices

Manuel Santos-Trigo
Cinvestav-IPN, Mexico

Ernesto Suaste
Cinvestav-IPN, Mexico

Paola Figuerola
Cinvestav-IPN, Mexico

INTRODUCTION

Nowadays it is recognized that the developments and use of digital technologies are changing the way in which people communicate, get access to a variety of information, solve problems, and discuss solutions. These technologies open up novel avenues for people to share practical and specialized knowledge from a variety of domains, and interest or goals. Tools can be physical or virtual artifacts and they influence human cognition and act as mediators between the phenomenological and conceptual world (Santos-Trigo, et al. 2016). In health areas, experts or professionals such as physicians, nurses or physiotherapists rely on social media (YouTube channels, blogs, wikis, or specialized webs) to share knowledge and practical experiences. Similarly, patients look for information on diseases and health advice and they often interact, via Internet, with experts to monitor their health and to interpret lab results. An important feature of the Web or an online platform is that it allows the participation of experts and users with a variety of experience or knowledge to build a source of accumulated knowledge in constant scrutiny. Denecke (2015) identifies three main groups as users of medical social media with different aims and purposes: patients, healthcare professionals and researchers. Thus, through the medical web, patients interact with health professionals and other patients to get

information on diseases, treatments, prevention or ways for operating tools for diseases. Healthcare professionals interact with other colleagues and patients to discuss latest results or to identify and share experiences and best practices. Researchers interact with other researchers to discuss treatment results and information on treatment outcomes and information of disease development and progression. How is a medical web system constructed? What criteria are used to select the information to be included in the online system? ¿How do patients process and use the information they consult? Available medical data in online platforms might vary and increase in volume and themes, then data visualization becomes important for physicians and patients to overview and interpret changes in health status. Likewise, physical tools designed by engineers and other professionals (magnetic resonance imaging, catheters, etc.) are also transforming medical practices such as surgeries and disease detection and treatments. Thus, it is important to analyze the extent to which users of both physical and virtual artifacts get engaged in activities to transform those artifacts into an instrument to solve problems. To this end, we focus on the importance for users to develop cognitive schemata that shape their appropriation of the tool affordances in problem solving environments (Santos-Trigo, et al, 2015).

The design and use of technological artifacts in medical practices involve the participation of

several experts' communities including engineers, scientists, and medical doctors. Thus, to delve into what happens to artifacts (designed by engineers) when they enter into medical practices implies not only to uncover what engineering and medical practices entail, but also to address how both agendas could converge or incorporate common goals. Regarding the engineering as a discipline, the National Research Council (NRC) (2009) pointed out its problem solving approach to design and create human-made products under certain conditions or constraints. Some of those constraints involve taking into account principles that encompass science and scientific laws, budget restrictions, available materials, sustainability, ergonomics, and ethical issues.

Progress in science and engineering goes hand in hand, science advances often depend on tools developed by engineers and reciprocally scientific knowledge guides and permeates engineering designs. Nowadays, it is not unusual to find engineers working with medical doctors and researchers in the design of artifacts that aim to improve human health.

Engineering design is a purposeful activity bound by specifications and constraints and an eminently collaborative enterprise. It involves an interactive process in which the design is tested and modified and it often offers several solutions to a particular problem (NRC, 2009). In general, the design and the construction of artifacts are tasks that require an expertise in science of materials, control, ergonomics and biomedicine. And there is set of standards that any medical artifact needs to fulfill in order to be used within the medical community. However, as Béguin (2003) pointed out the design of artifacts does not finish when the tool or object fulfills material and technical requirements; it should include how users transform the artifact into an instrument. Then, how does the medical community develop the needed expertise to use those artifacts efficiently in medical practice? This question becomes important to identify and delve into a research area that examines ways in which medical doctors transform

and artifact (physical device) into an instrument to solve problems. Then, what information and actions are important to characterize the process to transform an artifact into an instrument? Hadolt, Hörbst & Müller-Rockstroh (2012) cited a four-phase model (Hahn, 2004) that includes appropriation, objectification, incorporation, and conversion activities. The authors stress that the incorporation of artifacts into practices depends on social, cultural, and economic conditions.

It is important to analyze the extent to which medical doctors construct cognitive schemata that explain what we called their appropriation process of an artifact. In this process, it is recognized that tools shape and are shaped by the users' actions. Trouche (2004) pointed out the importance of considering the instrument as an extension of the body that becomes an organ formed by the artifact itself and by a psychological part that helps the user to mobilize the artifact to carry out activities and solve problems. Béguin (2003) emphasizes the difference between an artifact and an instrument or a problem-solving tool. The latter is made up by the artifact and the user's social and private cognitive schemata. The artifact characteristics that include ergonomics and constraints and the cognitive schemata developed by the user during the activities are important for the transformation of the artifact into a problem-solving tool or instrument to solve problems.

Artigue (2002) pointed out that users need to get involved in an appropriation process to transform an artifact or physical device into an instrument for specific use. This process leads users to gradually construct personal schemata or to appropriate pre-existing social schemata to appreciate its potentialities and use them in problem solving situations. Thus, relations between users and objects are shaped by ways in which a community of practice acts in problem solving environments. Trouche (2004) also mentioned that the development of the user's psychological component could be categorized in terms of three related functions: A *pragmatic* function where the subject achieves a particular goal, a *heuristic* func-

tion in which the subject visualizes and pursues an action plan; and an *epistemic* function where the subject comprehends and makes sense of what is being achieved.

BACKGROUND AND RATIONALE

Delving into the ways that medical doctors develop an expertise in the use of artifacts in their practice implies examining not only what activities they engaged in order to develop such expertise; but also reflecting on common issues that permeate or characterize two fields: the engineering design and medical practice. In this context, we stress the interdependence between scientific principles that support the design of tools and those activities involved in using those tools in problem solving environments.

Throughout the chapter, it is recognized that performances and advances of humankind go hand in hand with developments and use of technology. Thus, the use of tools not only influence cultural and science developments (Trouche, 2004); but also some tools facilitate the maintenance of physical wellness and humans comfort. Koszalka et al. (2014) pointed out that the use of technology goes beyond developing an expertise in its practice, it also includes that subjects develop a set of abilities for accessing relevant medical information, constructing an effective communication among colleagues and patients, and supporting the subjects participation in virtual practices environments. Thus, the use of tools has been shaping the developments and practices in medicine and in all disciplines. As a consequence, tasks in the design field and ways in which users appropriate the tools become relevant not only to incorporate the tools in different practicing domains, but also to provide feedback to improve the tools design.

Are there common principles that characterize the design and development of particular tools? How do engineers design artifacts? How are designed technologies or tools used in medical practices? To what extent does the design process

influence or shape ways in which medical doctors use the tool in their practices? The discussion of these questions provides basis to comprehend and frame ways in which medical doctors develop an expertise in the use of the tools. In a broader context, the relation between science and technology is a theme that should be examined from diverse angles or perspectives. In terms of goals and methods, engineering design, as an area of study, can be compared with processes involved in scientific developments, both processes rely on or are conceptualized as problem solving approaches (Santos-Trigo, 2010).

Trouche (2004) pointed out there are cultural biases or differences between Western and Eastern cultures regarding the value of physical and manual work vs. cognitive or intellectual work. While Eastern cultures recognize a dialectic interaction between mind and body, the West tends to privilege intellectual or mind activities over physical or manual activities. Recently, science and technology communities recognize that scientific inquiry or science development, technology, engineering, and mathematics are fields in constant interaction and play an important role in society developments. Trouche cites a Francis Bacon' passage that seems to recognize the importance of both intellectual and manual activities: "Human hand and intelligence, alone, are powerless: what gives them power are tools and assistants provided by culture" (p. 283). That is, both science and engineering or technology activities are fields that depend on each other and as a consequence, multidisciplinary approaches that incorporate principles and methods from several realms seem essential to identify, frame, and solve society problems. In particular, information regarding how scientific and technology developments influence not only the building of disciplinary knowledge or practices; but also the users' appropriation process of those developments becomes an important component of the academic agenda. Similarly, Ritella and Hakkarainen (2012) pointed out that in order to fully understand the role of artifacts in disciplinary practices, it is important to think

of them as cognitive prostheses that enhance and expand human creativity and ways to solve practical problems.

DESIGN AND USERS CHALLENGES

Béguin (2003) underlines that the design of an artifact, factory, or a vehicle is both an individual and collective process based on two principles: a) *differentiation* in which tasks involved in the design are distributed among the team members and b) *interdependence* that recognizes the need to ask for the participation of several specialists to work on a product or artifact. Likewise, the NRC (2009) states that engineering design demands that engineers work in collaboration with teams and clients. That is, the design process can be conceptualized as an ongoing activity in which the designers' plans and knowledge are key to achieve the task; but also during the same process some unexpected goals and challenges need to be attended or addressed. Indeed, although there is no formula to characterize the interactive process involved in engineering design, it is accepted that one initial step is the identification of a problem that eventually needs to be explicitly formulated. This problem becomes a platform to generate ideas and actions to solve it. Next, is the evaluation of potential solutions in which models and prototypes are constructed and evaluated in terms of the extent to which the solutions meet standards and constraints; an optimization process. A salient feature of engineering design activities is that goals and actions take place within a set of specifications and constraints and involve the consideration of multiple solutions. Thus, important steps that characterize the engineering tasks design include: the identification of the problem and objectives, the definition of goals and constraints, the search and gathering of key information, the identification of possible design solutions, contrasting and analyzing the pertinence of solutions, building and implementing the design, testing and evaluating such design and repeating all steps as necessary (NRC, 2009, p. 83).

It is observed that the aforementioned design process steps do not include, at least explicitly, the importance of taking into account the appropriation process in which users engage in order to use "efficiently" the product generated by the design. Thus, the challenge of communities involved in artifact designs is to extend the design process to include information regarding the subject's appropriation process of the artifact. That is, the interactive cycle that characterizes the process of designing artifacts should also include information regarding how the artifact becomes a tool for the client or user. In addition, feedback provided by users to designers plays an important role in improving and designing new artifacts. In the appropriation process of artifacts, the analysis of experts' behaviors and actions shown through dialogue-conducting strategies play a crucial role for subjects to transform an artifact into an instrument to solve problems (Folcher, 2003).

SCOPE OF THE ARTICLE: A FOCUS ON ARTIFACT APPROPRIATION

The analysis of the process involved in the subjects' appropriation of tools requires the documenting of ways in which a community of practices (Perkins, 2009) behaves in problem solving environments. In this context, it becomes important to analyze and document group or subjects' interactions within a community formed by expert such as cardiologists, mechanical and biomedical engineers, technicians, and nurses during the process of transforming a particular artifact into a problem-solving tool. The user's tool appropriation in medical practice often takes place within a community of practice formed by experts and specialists including anesthesiologists, nurses, medical experts, technicians, etc. where novices doctors participate gradually in several activities that involve the coordination of physical skills required while inserting of a catheter or stent and the corresponding interpretation information available through monitor displays.

There are different frameworks to delve into the subjects' appropriation process of particular tool; however, recently a *bricolage* perspective that involves relying on more than one framework has been used to explain how a particular community develops knowledge and expertise in problem solving activities. For example, the construct *community of practice* was originally developed as a learning theory that fosters self-empowerment and professional development and it has evolved to become a framework for improving organization's competitiveness (Li, et al. 2009). In this context, a community of practice framework can be extended to explicitly take into account theoretical constructs used in cognition and artifact studies (Verillon & Rabardel, 1995). These studies explicitly recognize and address the importance for subjects to transform a technological artifact into a problem-solving tool. Key in this approach is the recognition that the subject's appropriation of tools for problem solving activities involves an adaption and transformation of the artifact known as an instrumentation process and the development of cognitive schemata or instrumentalization process. Those schemata are explained in terms of subjects' cognitive systems associated with the use of artifacts (Rabardel, 1995). Ritella and Hakkarainen (2012) relied on these two concepts, instrumentation and instrumentalization to explain a dialectic relationship or mutual shaping between people and tools during the appropriation process. Cognitive behaviors are expressed by subjects' use of language and explicit use of strategies to coordinate information provided during the use of the tool. In addition, the artifact appropriation process takes place through subjects' legitimate peripheral participation in an ongoing social practice. Lave (1991) recognizes that an active participation as member of a community of practice shapes new members' identities and help them internalize community values and actions. Thus, as a part of a community of practice the subject develop and internalize knowledge to become a member of that community.

SOLUTIONS, RESULTS, AND REFLECTION

Currently, many research projects emphasize the importance of formulating and approaching scientific and technological problems through the convergence and collaboration of multiple disciplines. For example, Stewart (2011) introduces what he calls the sixth revolution in biology, the first five include: the microscope, classification of living creatures, the theory of evolution, the discovery of the gene, and the structure of DNA, in which he proposes to wide the scope of ways of thinking of the discipline to include mathematical approaches as the sixth revolution to frame and work on biological problems. "Mathematics provides a new point of view, addressing not just the ingredients for life, but the processes that use those ingredients" (p. 18). Similarly, a task of designing artifacts requires the participation of experts to develop both the technical and ergonomic components, but also to incorporate information regarding how an artifact is used in actual practices. Recently, in the USA there is an increasing interest for medical school reform that includes the use of technologies to help enhance and standardize the medical curriculum. For instance, scholars in medical educations recognize the importance to carry out multi and interdisciplinary studies to analyze ways to incorporate educational technologies in medical curricula (Koszalka, Epling, and Reece-Barnes, 2014). In particular, it becomes relevant to characterize medical doctors' cognitive and metacognitive behaviors that appear during their processes of appropriation of technological tools. That characterization includes the subject's development of schemata, mental models, and a language to communicate and eventually carry out the community's goal. Figuerola (2014) reviewed different studies that focused on individual's tools appropriation in several fields including the use of digital tools in mathematical practices. In addition, she examined medical practices involving the use several artifacts to introduce catheters or

stents where newcomers' doctors were guided to develop an expertise in those practices. In her analysis, she found that during the subjects' appropriation process of artifacts there are goals to meet at different stages that are monitored within the community. The stages involve initial recognition of a community, direct involvement in particular actions or routine activities, and the management of time to carry out and regulate the development of the activities (Dousay, Igoche and Branch, 2014). Thus, research results consistently show that there are three crucial phases to distinguish the subject processes to develop an expertise in the use of artifacts.

Initial familiarization within the community's goals and carrying out peripheral activities. At this stage, the medical doctor, who is developing an expertise to use a tool, gets engaged into a set of routine tasks that includes checking proper disposition of the surgery equipment, functioning of the auxiliary equipment and in general terms he/she observes the role played by each team member during the entire process of using a particular tool (catheter, for instance). In addition, the subject carries out pragmatic and routine tasks to help other team members and listen to others and to observe actions that medical expert exhibit during the entire process of using the tool. It is important to mention that the expert continuously calls the attention of the team members to pinpoint critical events that occur during the use of the tool. At this stage, they develop cognitive skills to initially visualize the complexity involved in using the tool to carry out some particular tasks and recognize the need to work and interact with others (Lavigne and Mouza, 2013).

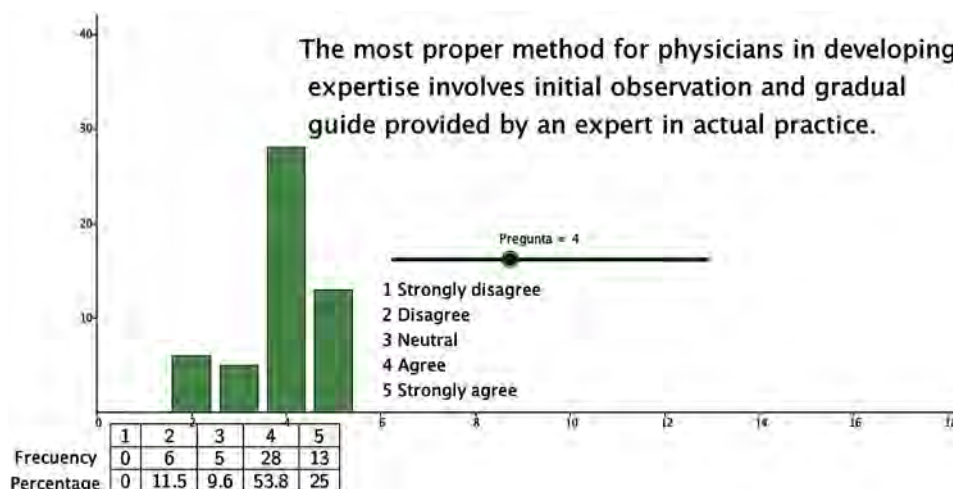
Gradual use of the tool. At this stage, the subject is gradually asked to participate into a series of tasks that includes the actual use of the artifact, in particular stages of the surgery involving the initial introduction of a catheter or stent and moving the artifact to detect a particular spot that needs attention. In general terms, the subject begins to develop and show a heuristic approach to take decisions that define and anticipate actions

related to the use of the device. For example, it is evident that in this phase, that even when the subject receives some explicit guidance to carry out particular tasks, the expert scaffolding activities to direct such guidance eventually get reduced or faded. Van Merriënboer, and de Bruin (2014) emphasize that people develop an expertise or learn something by being engaged directly in activities of doing, exploring, observing others, solving problems, analyzing information, repeating, etc. Subjects not only gradually get involved in tasks but it is evident the complexity of the tasks they participate increase gradually within the community.

Full participation that involves the subject's response to new events that appear during the actual use of the tool. In this phase, there is evidence that the subject has developed strategies and skills that help him/her to not only carry out all procedures to use the main artifact and auxiliary instruments; but also he/she exhibit a full coordination of emerging information displayed during the intervention process. That is, at this stage, the expert role is reduced to basically monitor the subject behavior and actions. It is important to observe that in addition to the subject full control of the process of using the tool, the subject shows a clear physical coordination between his/her hand and the manipulation of the artifact. That is, the artifact becomes an extension of his/her hand. Hahn (2004) mentions that as a result of an appropriation, understood this as the process by which objects change (not materially), the object or artifact is no longer what it used to be; but it becomes an extension of the subject's body and cognition.

Figuerola (2014) asked 52 practicing physicians working on different areas in public and private institutions about their experiences in the use of several medical artifacts and about how they developed the skills and knowledge to incorporate them in their professional practices. The questionnaire included 9 statements addressing issues regarding the importance of theoretical knowledge and the artifact; the role of observing

Figure 1. Relation between experts and residents in medical practices

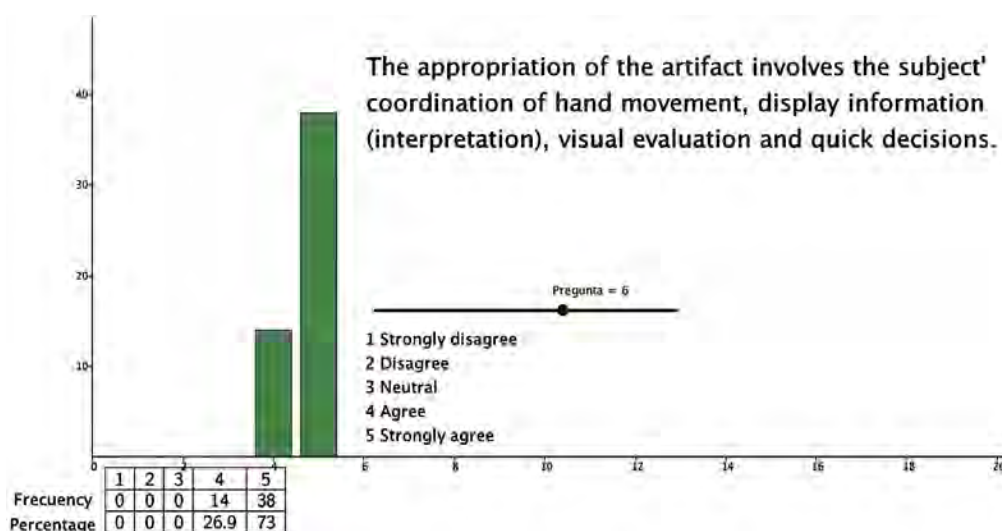


experts and gradually being involved in artifact use; the development of a specialized language to understand and explain decisions that emerge during the use of a particular artifact; and effective coordination between visual, hand movement, monitor information and the use of the artifact, and the recognition of patterns involved in actual practice and patient status. Figure 1 and 2 show frequency, percentages, and a bar graph of the answers that the 52 participants gave to two statements.

FUTURE RESEARCH DIRECTIONS

In this contribution we present a research area that deals with the importance of extending the process of designing and producing medical artifacts to include or take into account information regarding the transformation of an artifact into an instrument or problem solving tool. That is, it has been argued that the design of an artifact should not only involve the contribution of experts in several fields addressing technical or material issues; it

Figure 2. The coordination cognitive and body movements



should also include the participation of research groups that explicitly focus on the analysis of what the subjects' process of tool appropriation entails. The initial review provides a rationale to focus on analyzing cognitive behaviors that users exhibit during their appropriation process of the artifact. A challenge that emerges in this perspective is to develop the language and communication strategies to productively interact and communicate among the members of those diverse communities. Schoenfeld (2011) proposes three constructs to explain in detail what problem solvers do on a moment-by-moment basis while engaging in a problem solving process: the problem solver's resources, goals, and orientations. He suggests that these constructs offers practitioners, and problem solvers in different domains tools for reflecting on their practicing decisions. In his book, he uses the constructs to analyse and predict the behaviours of mathematics and science teachers and a medical doctor. It is also important to recognize that artifacts are developed within a cultural context where communities of practices reside and work, then what happens when the artifacts or products are used in environments that not necessarily reflect those cultural values? This could be a research question that designers and practicing communities should frame and pursue. In this context, the analysis of the subjects' appropriation process of the artifact should also include information regarding cultural values and traditions associated with environments where the device will be used. Finally, it is important to examine extant frameworks that are used to analyze ways in which subjects transform an artifact into an instrument and discuss the extent to which a *bricolage* perspective can be used to identify basis to characterize what is common in the subject's process of appropriation of different types of artifacts and what elements depend on the use of a particular device.

CONCLUSION

This contribution provides important information for both medical communities and biomedical engineers to discuss the process and ways to guide novice doctors to fully learn the use of the tool, and to improve the design of the tool. The distinction between an artefact and an instrument or tool is key to comprehend how the subject, a medical doctor in this case, develops knowledge about the artifact itself, its uses to transform such artifact into an instrument or problem solving tool. In particular, it became important to foster a full collaboration between both communities in designing and developing an expertise in using artifacts in medical practices. That is, it becomes important to consider users in the design of an artifact and also to involve them in the process by which users transform the artifact into an instrument. Design-in-use captures the idea that design and use is a continuous process that never ends. In addition, thinking of users' appropriation process of artifacts as an essential component of the artifact design implies to take into account information regarding how that appropriation takes place and could be achieved and monitored. This consideration or integration could be of benefit for both communities, engineering and medical communities, in terms of addressing and solving users' appropriation issues through a continual collaboration.

ACKNOWLEDGMENT

The authors thank the support received from project Conacyt-168543 during the preparation of this chapter.

REFERENCES

- Artigue, M. (2002). Learning mathematics in a CAS environment: The genesis of a reflection about instrumentation and dialectics between technical and conceptual work. *International Journal of Computers for Mathematical Learning*, 7(3), 245–247. doi:10.1023/A:1022103903080
- Béguin, P. (2003). Design as a mutual learning process between users and designers. *Interacting with Computers*, 14(5), 709–730. doi:10.1016/S0953-5438(03)00060-2
- Denecke, K. (2015). Health web science. Social media data for healthcare. New York: Springer Science+Business Media. doi:10.1007/978-3-319-20582-3
- Dousay, T. A., Igoche, D., & Branch, R. M. (2012). Self-regulated Learning as a Foundational Principle for a Successful Strategy in Teaching Educational Research Methods to Doctor of Philosophy Students. In M. Orey (Ed.), *Educational Media and Technology Yearbook* (Vol. 36, pp. 23-36). New York: Springer Science+Business Media. doi:10.1007/978-1-4614-1305-9_3
- Figuerola, R. P. (2014). *Technology and appropriation of tools in medical practices* (Doctoral dissertation). Cinvestav-IPN.
- Folcher, V. (2003). Appropriating artifacts as instruments: When design-for-use meets design-in-use. *Interacting with Computers*, 15(5), 647–663. doi:10.1016/S0953-5438(03)00057-2
- Hadolt, V., Hörbst, V., & Müller-Rockstroh, B. (2012). Biomedical techniques in context: On the appropriation of biomedical procedures and artifacts. *Medical Anthropology*, 31, 179–195. doi:10.1080/01459740.2011.636410
- Hahn, H. P. (2004). Global goods and the process of appropriation. In P. Probst & G. Spittler (Eds.), *Between Resistance and Expansion: Dimensions of Local Vitality in Africa* (pp. 213–231). Münster, Germany: Lit Verlag.
- Koszalka, T. A., Epling, J. W., & Reece-Barnes, J. L. (2014). Synthesis of recent literature on educational technology in medical curricula. In J. M. Spector, M. D. Merrill, J. Elen, & M. M. Bishop (Eds.), *Handbook of Research on Educational Communications and Technology* (pp. 855-866). New York: Springer Science+Business Media. doi:10.1007/978-1-4614-3185-5_69
- Lave, J. (1991). Situating learning in communities of practice. In L. Resnick, J. Levine, & S. Teasley (Eds.), *Perspective on socially shared cognition* (pp. 63–82). Washington, DC: American Psychology Association; doi:10.1037/10096-003
- Lavigne, N. C., & Mouza, C. (2013). Epilogue: Designing and integrating emerging technologies for learning, collaboration, reflection, and creativity. In C. Mouza & N. Lavigne (Eds.), *Emerging technologies for the classroom* (pp. 269–288). Springer; doi:10.1007/978-1-4614-4696-5_18
- Li, L. C., Grimshaw, J. M., Nielsen, C., Judd, M., Coyte, P. C., & Graham, I. D. (2009). Evolution of Wenger's concept of community of practice. *Implementation Science*, 4(11).
- National Research Council (NRC). (2009). *Engineering in K-12 education*. Washington, DC: The National Academic Press.
- Perkins, D. N. (2009). *Making Learning Whole*. San Francisco, CA: Jossey-Bass.
- Rabardel, P. (1995). *Les homes & les technologies. Approche cognitive des instruments contemporains*. Paris: Armand Colin Éditeur.

Ritella, G., & Hakkarainen, K. (2012). Instrumental genesis in technology-mediated learning: From double stimulation to expansive knowledge. *Computer-Supported Collaborative Learning*, 7(2), 239–258. doi:10.1007/s11412-012-9144-1

Santos-Trigo, M. (2010). A mathematical problem-solving approach to identify and explore instructional routes based on the use of computational tools. In J. Yamamoto, J. Kush, R. Lombard, & J. Hertzog (Eds.), *Technology implementation and teacher education: Reflective models* (pp. 296–313). Hershey, PA: IGI Global; doi:10.4018/978-1-61520-897-5.ch017

Santos-Trigo, M., Reyes-Martínez, I., & Ortega-Moreno, F. (2015). Fostering and supporting the coordinated use of digital technologies in mathematics learning. *Int. J. Learning Technology*, 10(3), 251–270. doi:10.1504/IJLT.2015.072360

Schoenfeld, A. H. (2011). *How we think. A theory of goal-oriented decision making and its educational application*. New York: Routledge.

Stewart, I. (2011). *The mathematics of life*. New York: Basic Books.

Trouche, L. (2004). Magaging the complexity of human/machine interactions in computerized learning environments: Guiding students command process through instrumental orchestrations. *International Journal of Computers for Mathematical Learning*, 9(3), 281–307. doi:10.1007/s10758-004-3468-5

Van Merriënboer, J. J. G., & de Bruin, A. B. H. (2014). Research paradigms and perspectives on learning. In J.M. Spector, M.D. Merrill, J. Elen, & M. M. Bishop (Eds.), *Handbook of Research on Educational Communications and Technology* (pp. 21-29). New York: Springer Science+Business Media. doi:10.1007/978-1-4614-3185-5_2

Verillon, P., & Rabardel, P. (1995). Cognitions and artifacts: A contribution to the study of thought in relation to instrument activity. *European Journal of Psychology of Education*, 10(1), 181–185. doi:10.1007/BF03172796

ADDITIONAL READING

Azevedo, R., & Aleven, V. (Eds.). (2013). *International handbook of metacognition and learning technologies*. NY: Springer; doi:10.1007/978-1-4419-5546-3

Cox, A. (2005). What are communities of practice? A comparative review of four seminal works. *Journal of Information Science*, 31(6), 527–540. doi:10.1177/0165551505057016

Folcher, V., & Rabardel, P. (2004). Hommes-Artefacts-Activiive review of four seminal works. ught in relation to instrument activity.itnd process through <http://ergoserv.psy.univ-paris8.fr/>

Freund, L. E., & Spohrer, J. C. (2012). The human side of service engineering. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 23(1), 2–10. doi:10.1002/hfm.20521

Hammond, J., Richard, J. K., & Craig, M. H. (2001). Distributed collaboration for engineering design: A review and reappraisal. *Human Factors and Ergonomics Design in Manufacturing*, 11(1), 35–52. doi:10.1002/1520-6564(200124)11:1<35::AID-HFM3>3.0.CO;2-E

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. New York: Cambridge University Press; doi:10.1017/CBO9780511815355

Luengo, V., Aboulafia, A., Blavier, A., Shorten, G., Vadcad, L., & Zottmann, J. (2009). Novel technology for learning in medicine. In N. Balacheff, S. Ludvigsen, T. de Jong, A. Lazonder, & S. Barnes (Eds.), *Technology-Enhanced Learning* (pp. 105–120). NY: Springer; doi:10.1007/978-1-4020-9827-7_7

National Academy of Engineering. (2011). *Frontiers of engineering*. Washington, DC: The National Academic Press.

National Academy of Engineering. (2012). *Making things. 21st Century manufacturing design*. Washington, DC: The National Academic Press.

Riva, G., & Mantovani, F. (2012). From the body to the tools and back: A general framework for presence in mediated interactions. *Interacting with Computers*, 24(4), 203–210. doi:10.1016/j.intcom.2012.04.007

Spector, J. M., Merrill, M. D., Elen, J., & Bishop, M. M. (Eds.). (2014). *Handbook of Research on Educational Communications and Technology* (pp. 21–29). NY: Springer Science & Business Media. doi:10.1007/978-1-4614-3185-5

KEY TERMS AND DEFINITIONS

Artifact: A material or physical object designed to sustain and improve human activities.

Catheter: A medical artifact that is inserted in a patient body to treat a disease or to perform a surgical operation.

Cognitive Scheme: The subject's internalisation of actions to use the tool in practice and to anticipate and respond to emerging actions and to comprehend what the user is doing with the tool.

Community of Practice: A group of people or special interest group involved in activities to promote individual and collective learning. Problem solving activities are key in learning communities.

Engineering Design: A goal oriented activity to create and design human-made products under constraint such as nature and science principles, time, money, sustainability, materials and ergonomics.

Instrument or Tool: An Artifact becomes a tool or an instrument when its user develops cognitive schemata to incorporate them into its social or discipline practices.

Stent: A medical artifact that is inserted into a natural conduit in a patient body to prevent or to treat a disease.

Tool Appropriation: The process by which the user transforms an artifact into an instrument. Thus, the tool or instrument becomes a user's functional organ formed by an artifact and a psychological component.

Trends in Health Care Information Technology and Informatics



T. Ray Ruffin

*University of Phoenix, USA & Colorado Technical University, USA & Grand Canyon University, USA
& Ashford University, USA, & North Carolina Wesleyan College, USA*

Donna Patterson Hawkins

University of Phoenix, USA

INTRODUCTION

It was not until 1994 that the United States (U.S.) health care industry established information systems capable of handling a universal delivery system. These Information Technology Systems (ITS) operated along enterprise and system boundaries in the Health Care Delivery System (HDS). ITS have become fragmented by the proprietary business benefits of large vendors that want to control patient information (Accenture, 2001). Practical tools, especially computers, continue to be created and rapidly placed in industry with the ability of organizations to accept, accommodate, and even embrace technology moving at a varied pace (McHaney, n.d.). The health care industry has been one of the unhurried organizations to embrace the computer revolution in regards to patient care. Health organizations have been using computers for years in business departments. Health care comprises the use and management of a profusion of information that must be collected, managed, reviewed, processed, and mined (McHaney, n.d.). Technology changes rapidly and maintaining the status quo in health care actually means falling behind; and health organizations cannot afford to do that in a technology-driven world (Rutsky, 1999).

The implications facing health organizations are driven by the substantial pressure to implement Health Information Technology (HIT) systems that have “certified” Electronic Health Records (EHRs) applications and that fulfill the federal

government’s definition of “meaningful use” or risk significant financial penalties in the near future (Ford, Menachemi, Huerta, & Yu, 2010). To offset this pressure and cost, every hospital in the U.S. is eligible for a minimum of \$2 million incentive with larger hospitals eligible for more funds to purchase and incorporate EHRs (Bau, 2011). With this in mind, health organizations such as larger hospitals located in urban areas, and teaching hospitals are implementing EHRs at a quicker rate (Jha et al, 2009). In 2014, research found approximately 38 percent possibly eligible hospitals achieved meaningful EHR implementation by the end of 2012 (Diana et al, 2014). The Health Information Technology for Economic and Clinical Health Act (HITECH) of 2009 is driving fundamental market and industry changes that health organizations need to be strategically poised to deal with and implement. This chapter investigated the trends in health care information technology and informatics related to patient care. The chapter will consist of the following sections: background; issues controversies, and problems; solutions and recommendation; and future research directions in health care.

BACKGROUND

Health care expenditures have amplified melodramatically during the past 50 years, mutually in total terms and as a share of Gross Domestic Product (GDP) (Center for Medicare and Medicaid

DOI: 10.4018/978-1-5225-2255-3.ch330

Services [CMS], n.d.). Expenditures in the U.S. health care segment computed over \$2.7 trillion in 2011, increasing from up the \$698.3 billion expended in 1980, increasing by a factor of 3.9. Health care expenditures in 2011 attributed for 17.9 percent of GDP, resulting in doubling of the shares from 1980 (CMS, n.d.). These expenditures have led to Quality Management for Health Care System to make available a structure to aid health organizations in communicating, monitoring, and incessantly advancing the whole HDS (James, n.d.). The vision for the Center for Medicare & Medicaid Services (CMS) Quality Strategy is to optimize health outcomes by leading clinical quality improvement and health system transformation. This has resulted in and gives indication back to the very suggestion that a systematized system to achieve high quality care can be a front-runner to lowering health care costs (James, n.d.).

The Patient Protection and Affordable Care Act (PPACA) entails 10 separate legislative Titles, with more than a few major goals. As it relates to quality, the goal is to increase health-care value, quality, and efficiency, while dropping lavish expenditures and creating the health-care system more accountable to a varied patient population (Rosenbaum, 2011). Also, the Act invests in the development of a multi-payer National Quality Strategy to create multi-payer quality and efficiency procedures to stimulate worth procuring, better safety, and all-embracing health information across public and private insurers (Public Law 111-148, 2010). In the long run, this will build on the HITECH Act, which was ratified into law in 2009 as measure of the American Recovery and Reinvestment Act (ARRA) (Public Law 111-5, 2009).

Trends in Health Care Information Technology and Informatics Issues, Controversies, Problems

Today's digital technology plays a significant role, permitting the storage and rapid retrieval of patient records and other important information. At the same time, patients expect their sensitive per-

sonal information to be handled appropriately, to ensure accuracy and confidentiality (Hall, 2014). Privacy and security is a humongous concern for a hospital's network infrastructure. Implementing security systems to prevent data breaches and leaks, keep Patient Health Information (PHI) secure, and managing the secure transmission of electronic medical record data are all major issues in the industry. Management must continue addressing ethical and legal issues with regard to control of information. In multifaceted health care organizations, management is realizing efficiency goals assist with meeting more intricate goals (Prior, 2006).

Due to the economic crisis of 2008, the American Recovery and Reinvestment Act (ARRA) of 2009 directed nearly \$150 billion in new funds to the health care industry. According to Health care IT News (2015), it included \$87 million for Medicaid, \$24.7 billion to subsidize private health insurance for people who lose or have lost their jobs, \$19.2 billion for health information technology, and \$10 billion for the National Institutes of Health (NIH). Also, the Act provided \$650 million to support prevention and wellness activities targeting obesity, smoking, and other risk factors for chronic diseases. Subsequently, it obligated \$500 million for health professions training programs to include \$300 million for revitalizing the National Health Service Corps (NHSC) (Health care IT News, 2015). This will allow promoting the foundation for performance reporting on a system-wide source so that patients can more readily get information about their own health care and how their health-care providers perform.

ARRA funding was designated to modernize the health care system by promoting and expanding the adoption of HIT. The federal government stated achieving this goal would reduce its health costs by more than \$12 billion over the next 10 years (Sandlot Solutions, 2014). ARRA targeted funds at implementing a national infrastructure for health information technology. ARRA allocates \$19.2 billion to increase the federal coordination of

that effort and the use of electronic health records (Portfolio Media, Inc., 2009). Politicians are still arguing over its efficacy. Opposition continues to state that say the stimulus was too big or too small and the government should have allocated those billions of dollars more prudently (Gura, 2014).

According to Blumenthal (2010), the HITECH Act of 2009 is an attempt to improve the performance of health care systems through the use of technology, such as EHR's. The HITECH Act is focused on health organizations using certified EHRs. The HITECH Act was touted to expand the federal government ability to establish a national electronic patient records system by 2014 (APA Practice Organization, 2014). This allows the Department of Health and Human Services (HHS) to have the authority to institute programs for improving health care quality, safety and efficiency through the advancement of HIT (HealthIT.gov, 2014). The HITECH Act provides for enforcement and penalties for violation of privacy and security standards, including permitting enforcement through state attorneys general (APA Practice Organization, 2014).

Patient implications include reduced cost and better quality of care. The HITECH Act of 2009 supports health care quality improvement and cost reductions. This falls in line with the PPACA. An example is the required functions of EHRs being aligned "to requirements for emerging models of health care delivery improvement, including "patient-centered medical homes," a term for health care providers who emphasize partnerships between patients, their physicians, and Accountable Care Organizations" (Bau, 2011, p 1). The Act has the ability to involve patients and health consumers in their own self-care, even rallying entire communities into more patient self-awareness. In addition, PPACA has come up with a Duals MediPlan, where only some regions and specific states have the luxury to take advantage of these benefits. The Dual-Medi services provide in-home care and other services that Medicare and any other plan will not cover (Senior Quote, 2014).

The PPACA launched the Institute for Comparative Clinical Effectiveness Research to support the type of research essential to identifying the most appropriate and efficient means of delivering health care for diverse patient populations (Public Law 111-148, 2010). All the way through, these ingenuities foster improved quality and materials that underline hard work to accumulate statistics about health and health-care inequalities. This permits the country to enhance measure advancement, for the populace as a total and subpopulations with prominent possibility for reduced health consequences (Rosenbaum, 2011). Quantities of health care assistances are mandated by either federal law, state law, or both. Between the federal government and the states, there are in the neighborhood of 2,000 health insurance mandates (Bihari, 2014).

According to Dingley, Daugherty, Derieg, & Pershing (2008) effective communication among health care professionals is challenging due to a number of interrelated dynamics:

- Health care is difficult and volatile, with professionals from an assortment of disciplines involved in providing care at various times, often dispersed over several locations, creating spatial gaps with limited opportunities for regular synchronous interaction (Zwarenstein & Reeves, 2002).
- Health care providers often have their own disciplinary view of what the patient needs, with each provider arranging the activities in which they act independently (Zwarenstein & Reeves, 2002).
- Health organizations have traditionally had a hierarchical organizational structure, with substantial power distances between physicians and other health care professionals. This commonly leads to a culture of inhibition and restraint in communication, rather than a sense of open, safe communication (psychological safety).

Table 1. Benefits of using information and communication technologies in the health care sector. from Implementation and Integration of Regional Health care Data Networks in the Hellenic National Health Service, by Lampsas, Vidalis, Papanikolaou, & Vagelatos (2002)

Health Care Players	Benefits
Administration Units	• Policy development and decision-making are strongly supported by effective and on-time information gathering and distribution. • Easier adaptation to eEurope challenges. • Supply control; better budget monitoring. • Overall improvement in the way citizens are served.
Hospitals	• Increased efficiency in communication between hospitals, administration units, social security services, careers, physicians, and citizens. • Personnel familiarization with information technologies through Internet-access operations. • Patient-record traffic support. • Reinforcement of the need to build health care information systems (HCISs) and local networks in hospitals. • Utilization of the developed Intranets. • Better information services for the citizens. • Advanced telematics services (i.e. telemedicine applications in difficult-to-reach regions).
Health Care Personnel	• Meets the increased need for telecommunications not only for medical, but also for compensation reasons. • Participation in care chains and relevant coordination. • Physicians' collaboration. • Patients' history data retrieval. • Continuing education services; familiarization with new technologies through special training programs. • Interaction with patients to provide advice or prescriptions.
Citizens	• Use of the Internet for health-related information retrieval. • Information and communication technologies will increase interest in citizens' health-issues management. • Creation of the appropriate infrastructure for future provision of special health services for specific population groups (eg, in-house services for older people or patients with long-lasting attendance and nursing needs).

- Differences in education and training among occupations in the health care field often result in different communication styles and methods that further complicate the scenario and render communications ineffective.
- Teamwork and effective communication are crucial for safe patient care; however, the educational curricula for most health care professions focus primarily on individual technical skills, neglecting teamwork and communication skills.

Table 1 is factors that lead to the requirement to implement health-authorities interconnection infrastructure. Furthermore, these factors may be seen as the targets that are and will be accomplished (Lampsas, Vidalis, Papanikolaou, & Vagelatos, 2002).

SOLUTIONS AND RECOMMENDATIONS

Technology in the form of software and hardware has given the health care industry the opportunity to transform services. Regulations and technologies have given patients control over their care and the autonomy to choose innovative physicians able to deliver personal service. Continuous adherence to government requirements and insurance carrier demands requires software, hardware, and related components to ensure success. Health care organizations need to safeguard patient information with secure storage of all compliance-related documentation. This includes ensuring the ability to control access, add, or modify records from patient information to billing and reconciliation of data. Health Insurance Portability and Accountability Act (HIPA) and ARRA affect the various health care organizations.

Table 2. Economic impact for the U.S. health care industry resulting from Communications Inefficiencies Sector from *The Imprivata Report on the Economic Impact of Inefficient Communications in Health care*, by Ponemon Institute (2014 p. 3).

Workflows	Extrapolated Annual Cost of Inefficient Time per Hospital	Number of Registered Hospitals ²	Extrapolated Annual Impact for the Industry (U.S. \$ Billions)
Patient Admissions	\$727,957	6.409	\$4.67
Emergency Response	\$265,254	6.409	\$1.70
Patient Transfers	\$753,755	6.409	\$4.83
Total	\$1,746,966	6.409	\$11.20

¹United States Bureau of Labor Statistics, Occupational Employment Statistics May 2012.

²American Hospital Association May 2014 Update on the number of registered U.S. hospitals

Prohibiting the disclosure of records without the written consent of the individuals is essential unless one of the twelve disclosure exceptions listed in the Act applies. The Privacy Act of 1974 created the System of Records Notices (SORNs). This moved onto the Privacy Impact Assessments (PIAs), the “E-Government Act of 2002 that required government agencies to assess the impact on privacy for systems that collect personally identifiable information in PIAs” (U.S. Department of Health & Human Services [HHS], n.d., para 6.). Prohibiting the disclosure of records without the written consent of the individuals is essential unless one of the twelve disclosure exceptions listed in the PPACA applies. There needs to be a thorough look at integrated health care systems and installation of computer networks by health organizations.

It is mandatory on Information Technology (IT) to meet provider demand for more modern and efficient communication technologies, while maintaining patient privacy and complying with security and regulatory requirements,” as concluded by the team of researchers (Ponemon Institute, 2014). Overpoweringly, respondents of the Ponemon Institute study agree that substantial time is wasted and it is now up to hospitals to embrace technology by coming up with appropriate and compliant approaches. Table 2 summarizes the annual productivity and economic impact of inefficiency in communication and collaboration.

Ponemon Institute (2014) study found three areas of opportunity that can be improved with communication and collaboration:

- Patient Admission: Admitting one patient takes about 51 minutes, which an average of 33 minutes (65 percent) is wasted due to inefficient communications. This translates into an annual loss of about \$728,000 per U.S. hospital.
- Emergency Response Coordination: Coordinating an emergency response team takes an average of 93 minutes per patient. Of this time, an average of 40 minutes (43 percent) is wasted due to inefficient communications. This equates to an annual loss of more than \$265,000 per U.S. hospital.
- Patient Transfer: Transferring a patient to another facility or home care/hospice takes about 56 minutes, if all approvals are given in a timely manner, which an average of 35 minutes (63 percent) is wasted due to inefficient communications. The total annual cost of this waste is about \$754,000 per U.S. hospital (p. 2).

FUTURE RESEARCH DIRECTIONS

At the national level, most nations must reconsider their e-health strategies, according to the new

trends and requirements, to adopt the developed standards, and introduce certification procedures for use of all Information Technology and Communication (ITC) applications in health care (Mihalas, 2014). Potential medical applications in public health and in the “ever-lasting health care reform” determined the creation or development of ITC compartments. The ITC elaborates or develops national/regional strategies for implementation of IT in health care/medical activities (Mihalas, 2014). In European countries, the old trend in biomedical/health informatics was to analyze the courses presented to the students for medical students/computer science students or future Health care Informatics Professionals (HIPs) (Mihalas & Lungeanu, 1997). The new trend is numerous European countries introducing doctoral level degrees in medical informatics. This includes a universal approach to medical informatics courses and specialized medical/health informatics schools being at both undergraduate and master levels (Hasman, Mantas, & Zarubina, 2014). Continually, IT applications refer to ‘big data’, which is the enormous amount of data collected that requires applicable infrastructure such as cloud computing, as an impression on countless ITC applications in health care. This includes the national health care plans and strategies (Shimrat, 2009).

During research studies mobile devices were shown to improve communication between doctors and nurses on inpatient wards (Ozdalga, Ozdalga, & Ahuja, 2012). A study by Wallace, Clark, and White (2012) of medical school health care providers and students revealed that over 80 percent of those who responded described using mobile devices to communicate with colleagues about patient care via e-mail, telephone, and text messages. According to Mickan, Tilson, Atherton, et al. (2013), doctors have reported that the use of a mobile device for retrieving information from drug databases led to more efficient decision-making and patient care. They discovered that mobile devices have the ability to increase quality of patient documentation through fewer errors, more

complete records, rapid access to new information, and improved workflow patterns (Mickan, Tilson, Atherton, et al. 2013).

There should be further research on smart devices and their capabilities in communication, assessing, and documenting patient information. It is important to look at the future and the use of medical or health care systems applications becoming more useful in the prevention and management of chronic health conditions, such as diabetes, obesity, and heart diseases. Electronic health applications are electronic tools that utilize information technology to store, manage and share health information (Tang, Ash, Bates, Overhage, & Sands, 2006). One of the most widely known electronic health applications is EHRs, which is an official health record that is digitized for patients to be shared across multiple health care organizations. Mobile device hardware and applications should continue getting better and more user friendly, bringing additional and enhanced benefits to clinical practice (Mickan, Tilson, Atherton, et al. 2013).

CONCLUSION

This chapter investigated the trends in health care information technology and informatics related to patient care. The chapter consisted of the following sections: background; issues controversies, and problems; solutions and recommendation; and future research directions in health care. The term smart devices will be used interchangeably with mobile devices such as smart phones, tablets and iPads; each commonly used by the general public which includes health care providers. Not all health care providers are eager to adopt mobile devices even though they have the potential to increase access to point-of-care tools that has been proven to support better clinical decision-making and improved patient outcomes (Divali, Camosso-Stefinovic, & Baker, 2013).

Information and time management has been the most frequent uses for mobile devices. Popular information management applications, such as Evernote and Notability, allow users to write or dictate notes, record audio, store photographs, and organize material into categories within a searchable electronic database (O'Neill, Holmer, Greenberg, & Meara, 2013). This allows for better documentation of patient records. Research has indicated that mobile devices improve the completeness and accuracy of patient documentation (Payne, Wharrad, & Watts, 2012). Mobile devices were judged higher than documentation using paper records (Mickan, Tilson, Atherton, et al. 2013). Ultimately, their study found a significant increase in the average rate of electronic prescribing, from 52 percent to 64 percent ($P = 0.03$) with the use of mobile devices (Mickan, Tilson, Atherton, et al. 2013).

REFERENCES

- Accenture. (2001). *Centralized information system to improve nation's health care delivery*. Retrieved from: <http://www.atp.nist.gov/ea/sp950-3/accenture.pdf>
- APA Practice Organization. (2014). *The Health Information Technology for Economic and Clinical Health Act (HITECH Act)*. Retrieved from <http://www.apapracticecentral.org/advocacy/technology/hitech-act.aspx>
- Arsoniadis, E. G., & Melton, G. B. (2016). Leveraging the Electronic Health Record for Research and Quality Improvement: Current Strengths and Future Challenges. *Seminars in Colon and Rectal Surgery*, 27(2), 102–110. doi:10.1053/j.scrs.2016.01.009
- Bau, I. (2011). Connected for health: The potential of health information and communications technologies to reduce health care disparities. *National Civic Review*, 100(3), 15–18. doi:10.1002/ncr.20064
- Bihari, M. (2014). *Mandated Benefits - Understanding Mandated Health Insurance Benefits: Health Benefit Mandates Are Controversial*. Retrieved from: http://healthinsurance.about.com/od/reform/a/mandated_benefits_overview.htm
- Blumenthal, D. (2010). Launching HITECH. *The New England Journal of Medicine*, 362–385. PMID:20042745
- Center for Medicare & Medicaid Services. (2015). *EHR incentive programs*. The Official Web Site for the Medicare and Medicaid Electronic Health Records (EHR) Incentive Programs. Retrieved from: <http://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/index.html?redirect=/EHRIncentivePrograms/>
- Center for Medicare & Medicaid Services. (n. d.). *CMS Quality Strategy: About the CMS Quality Strategy*. Retrieved from: <http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityInitiativesGenInfo/CMS-Quality-Strategy.html>
- Community Health Access Network. (2002). *Integrated health center network information technology applications*. Retrieved from: http://www.providersedge.com/ehdocs/ehr_articles/CHAN_EMR_White_Paper.pdf
- Diana, M. L., Harle, C. A., Huerta, T. R., Ford, E. W., & Menachemi, N. (2014). Hospital Characteristics Associated With Achievement of Meaningful Use. *Journal of Healthcare Management*, 59(4), 272–284. PMID:25154125
- Dingley, C., Daugherty, K., Derieg, K., & Persing, R. (2008). Improving patient safety through provider communication strategy enhancements. In *Advances in Patient Safety: New Directions and Alternative Approaches* (Vol. 3). Rockville, MD: Agency for Health care Research and Quality (US). Retrieved from: <http://www.ncbi.nlm.nih.gov/books/NBK43663>

- Divali, P., Camosso-Stefinovic, J., & Baker, R. (2013). Use of personal digital assistants in clinical decision making by health care professionals: A systematic review. *Healthcare Informatics*, 19(1), 16–28. doi:10.1177/1460458212446761 PMID:23486823
- Facts, O. (n.d.). *Obamacare Bill: Full PPACA & Related Laws*. Retrieved from: <http://obamacarefacts.com/obamacarebill/>
- Ford, E. W., Menachemi, N., Huerta, T. R., & Yu, F. (2010). Hospital IT Adoption Strategies Associated with Implementation Success: Implications for Achieving Meaningful Use. *Journal of Healthcare Management*, 55(3), 175–188. PMID:20565034
- Gura, D. (2014). *How well did the stimulus work?* Market Place. Retrieved from <http://www.marketplace.org/topics/economy/how-well-did-stimulus-work>
- Hall, A. (2014). *Security trends in cloud computing part 2: Addressing data challenges in health care*. Retrieved from: [http://blogs.microsoft.com/cybertrust/2014/05/01/security-trends-in-cloud-computing-part-2-addressing-data-challenges-in-health care/](http://blogs.microsoft.com/cybertrust/2014/05/01/security-trends-in-cloud-computing-part-2-addressing-data-challenges-in-health-care/)
- Hasman A, Mantas J, Zarubina T. (2014). An abridged history of medical informatics education in Europe. *Acta Inform Med*, (1), 25–36. doi. 10.5455/aim.2014.22.25-36
- HealthIT.gov. (2014). *Health IT legislation. HITEACH Act*. Retrieved from <http://www.healthit.gov/policy-researchers-implementers/health-it-legislation>
- Health care Information and Management Systems Society. (2008). *Health care integration and connectivity: Results of a survey by the Enterprise Information Systems Steering Committee*. Retrieved from: <http://www.himss.org/files/himssorg/content/files/integrationconnectivitysurvey.pdf>
- James, B. (n.d.). *Quality Management for Health Care Delivery*. The Hospital Research and Educational Trust of the American Hospital Association. Retrieved from: <http://www.ghdonline.org/uploads/QMMP.pdf>
- Jha, A. K., DesRoches, C. M., Campbell, E. G., Donelan, K., Rao, S. R., Ferris, T. G., & Blumenthal, D. et al. (2009). Use of electronic health records in U.S. hospitals. *The New England Journal of Medicine*, 360(16), 1628–1638. doi:10.1056/NEJMs0900592 PMID:19321858
- Joseph, S., Sow, M., Furukawa, M. F., Posnack, S., & Chaffee, M. (2014). Policy. HITECH Spurs EHR vendor competition and innovation, resulting in increased adoption. *The American Journal of Managed Care*, 20(9), 734–740. PMID:25365748
- Lampsas, P., Vidalis, I., Papanikolaou, C., & Vagelatos, A. (2002). Implementation and integration of regional health care data networks. *Hellenic National Health Service Journal Med Internet Res*, 4(3), e20. doi:10.2196/jmir.4.3.e20 PMID:12554551
- McHaney, D. (n.d.). *Information management and technology*. Burlington, MA: Jones and Bartlett Publishers, LLC.
- Mickan, S., Tilson, J. K., Atherton, H., Roberts, N. W., & Heneghan, C. (2013). Evidence of effectiveness of health care professionals using handheld computers; a scoping review of systematic reviews. *Journal of Medical Internet Research*, 15(10), e212. doi:10.2196/jmir.2530 PMID:24165786
- Mihalas, G. (2014). Evolution of trends in European Medical Informatics. *Acta Informatica Medica*, 22(1), 37–43. doi:10.5455/aim.2014.22.37-43 PMID:24648618
- Mihalas, G., & Lungeanu, D. (1997). *Yearbook of Medical Informatics*. Strategy for Medical Informatics Education at the University of Medicine and Pharmacy in Timisoara.

- Moodley, A., Mangino, J., & Goff, D. (2013). Review of infectious diseases applications for iPhone/iPad and Android: From pocket to patient. *Clinical Infectious Diseases*, 57(8), 1145–1154. doi:10.1093/cid/cit455 PMID:23839999
- Mosa, A. S., Yoo, I., & Sheets, L. (2012). A systematic review of health care apps for smartphones. *BMC Medical Informatics and Decision Making*, 12(1), 67. doi:10.1186/1472-6947-12-67 PMID:22781312
- National Institutes of Health. Health & Human Services. (2014). *Glossary of Frequently Encountered Terms in Health Economics*. Retrieved from <https://www.nlm.nih.gov/nichsr/edu/healthecon/glossary.html>
- O'Neill K.M., Holmer H, Greenberg S.L., Meara J.G. (2013). Applying surgical apps: Smartphone and tablet apps prove useful in clinical practice. *Bulletin of the American College of Surgeons*, 98(11), 10–18.
- Ozdalga, E., Ozdalga, A., & Ahuja, N. (2012). The Smartphone in Medicine: A Review of Current and Potential Use Among Physicians and Students. *Journal of Medical Internet Research*, 14(5), e128. doi:10.2196/jmir.1994 PMID:23017375
- Payne, K. B., Wharrad, H., & Watts, K. (2012). Smartphone and medical related app use among medical students and junior doctors in the United Kingdom (UK): A regional survey. *BMC Medical Informatics and Decision Making*, 12(1), 121. doi:10.1186/1472-6947-12-121 PMID:23110712
- Ponemon Institute. (2014). *The Imprivata report on the economic impact of inefficient communications in health care*. Retrieved from: <http://www.imprivata.com/imprivata-report-economic-impact-inefficient-communications-health-care>.
- Portfolio Media, Inc. (2009). *ARRA impacts on health care*. Retrieved from: <http://www.oppenheimer.com/uploadedFiles/News/ARRA%20Impacts%20On%20Health%20Care.pdf>
- Prior, D. (2006). Efficiency and total quality management in health care organizations: A dynamic frontier approach. *Annals of Operations Research*, 145(1), 281–299. doi:10.1007/s10479-006-0035-6
- Public Law 111-148. (2010, March 23). *The Patient Protection and Affordable Care Act (PPACA)*. Retrieved from: <http://www.gpo.gov/fdsys/pkg/PLAW-111publ148/pdf/PLAW-111publ148.pdf>
- Public Law 111-5. (2009, February 17). *The American Recovery and Reinvestment Act (ARRA)*. Retrieved from: <http://www.gpo.gov/fdsys/pkg/PLAW-111publ148/pdf/PLAW-111publ148.pdf>
- Quote, S. (2014). *Medi-Medi Dual Eligibility*. Retrieved from: <http://seniorquote.com/medi-medi-dual-eligibility/>
- Rosenbaum, S. (2011). The Patient Protection and Affordable Care Act: Implications for Public Health Policy and Practice. *Public Health Reports*, 126(1), 130–135. PMID:21337939
- Rutsky, L. (1999). Techno-cultural interaction and the fear of information. *Style (DeKalb, IL)*, 3, 267.
- Shimrat, O. (2009). *Cloud computing and Health-care Bad weather or Sunny Forecast? Technology Matters*. Retrieved from: https://www.himss.org/files/HIMSSorg/content/files/Code%2093_Shimrat_CloudComputingandHealthcare_2009.pdf
- Solutions, S. (2014). *ARRA Stimulus Package*. Retrieved from: <http://www.sandlotsolutions.com/resources/stimulus-package>
- Tang, P. C., Ash, J. S., Bates, D. W., Overhage, J. M., & Sands, D. Z. (2006). Personal health records: Definitions, benefits, and strategies for overcoming barriers to adoption. *Journal of the American Medical Informatics Association*, 13(2), 121–126. doi:10.1197/jamia.M2025 PMID:16357345
- U.S. Department of Health & Human Services. (n.d.). Retrieved from: <http://www.hhs.gov/foia/privacy/>

Wallace, S., Clark, M., & White, J. (2012). Its on my iPhone: Attitudes to the use of mobile computing devices in medical education, a mixed-methods study. *BMJ Open*, 2(4), e001099. doi:10.1136/bmjopen-2012-001099 PMID:22923627

Zwarenstein, M., & Reeves, S. (2002). Working together but apart: Barriers and routes to nurse-physician collaboration. *Journal Quality Improvement*, 28(5), 242–247. doi:10.1016/S1070-3241(02)28024-4 PMID:12053458

ADDITIONAL READING

Arsoniadis, E. G., & Melton, G. B. (2016). Leveraging the Electronic Health Record for Research and Quality Improvement: Current Strengths and Future Challenges. *Seminars in Colon and Rectal Surgery*, 27(2), 102–110. doi:10.1053/j.scrs.2016.01.009

Baker, L., Wagner, T. H., Singer, S., & Bundorf, M. K. (2003). Use of the Internet and email for health care information: Results from a national survey. *Journal of the American Medical Association*, 289, 2400–2406. doi:10.1001/jama.289.18.2400 PMID:12746364

Blumenthal, D. (2009). Stimulating the Adoption of Health Information Technology. *The New England Journal of Medicine*, 360(15), 1477–1479. doi:10.1056/NEJMp0901592 PMID:19321856

Gagnon, M., Simonyan, D., Gaston Godin, K., Labrecque, M., Ouimet, M., & Rousseau, M. (2016). Factors influencing electronic health record adoption by physicians: A multilevel analysis. *International Journal of Information Management*, 36(3), 258–270. doi:10.1016/j.ijinfomgt.2015.12.002

Hesse, B. W., Nelson, D. E., Kreps, G. L., Croyle, R. T., Arora, N. K., Rimer, B. L., & Viswanath, K. (2005). Trust and Sources of Health Information: The Impact of the Internet and Its Implications for Health Care Providers: Findings From the First Health Information National Trends Survey. *American Medical Association's. Journal of Internal Medicine*, 165(22), 2618–2624.

Huang, Y., & Gramopadhye, A. (2014). Systematic engineering tools for describing and improving medication administration processes at rural healthcare facilities. *Applied Ergonomics*, 45(6), 1712–1724. doi:10.1016/j.apergo.2014.06.003 PMID:25024094

Marsolo, K. (2013). Informatics and operations get integrated. *Journal of the American Medical Informatics Association*, 20(1), 122–124. doi:10.1136/amiajnl-2012-001194 PMID:22940670

Meyer, R., & Degoulet, P. (2010). Choosing the right amount of healthcare information technologies investments. *International Journal of Medical Informatics*, 79(4), 225–231. doi:10.1016/j.ijmedinf.2010.01.001 PMID:20122867

KEY TERMS AND DEFINITIONS

American Recovery and Reinvestment Act (ARRA): Had three immediate goals: to create new jobs and save existing ones, to spur economic activity and invest in long-term growth and to foster more accountability and transparency in government spending (Blumenthal, 2010).

Clinical Effectiveness: Application of interventions which have been presented to be effective to appropriate patients to improve patients' outcomes and value for the use of resources (National Institutes of Health, Health & Human Services, 2014).

Gross Domestic Product (GDP): Represents the monetary value of all goods and services produced within a nation's geographic borders over a specified period of time.

Health Information for Economic and Clinical Health (HITEACH) Act of 2009: Requires certified Electronic Health Records (EHR's). This requires providers to commit continued use of EHR technology over a period of years. There are no measurable thresholds and programmatic timelines (Joseph, Sow, Furukawa, Posnack, & Chaffee, 2014). However, the certification process for EHR vendors has reduced provider uncertainty

by establishing an unbiased validation service to ensure that certain functionality is available. Providers have to meet certain criteria of “meaningful use”. To an extent that has put the market and not the federal government in determining the vendors that will meet the provider needs with the financial incentives (Joseph, et al., 2014).

Health Insurance Portability and Accountability Act (HIPA): Act created by the U.S Congress in 1996 that amends both the Employee Retirement Income Security Act (ERISA) and the Public Health Service Act (PHSA) in an effort to protect individuals covered by health insurance and to set standards for the storage and privacy of personal medical data.

Patient Protection and Affordable Care Act (PPACA): The health care interrelated portions of the Health Care and Education Reconciliation Act of 2010 and the Student Aid and Fiscal

Responsibility Act were also affirmed into law end to end with the PPACA. These Acts seek to overhaul the current healthcare system. “Obamacare” H.R. 3590, or Affordable Care Act (ACA) for short turned out to be the largest renovation to healthcare reform since Medicare/Medicaid which was signed into law in 1965 (Obamacare Facts, n. d.).

Privacy Act of 1974: This act “protects records that can be retrieved by personal identifiers such as a name, social security number, or other identifying number or symbol” (HHS, n. d.).

System of Records Notices (SORNs): “Required health organizations to create and maintain SORNs. “A system of records consists of any item, collection, or grouping of information about an individual, where those records can be retrieved by the name of the individual or by some other type of identifier unique to the individual” (HHS, n. d.).

User Resistance to Health Information Technology

Madison N. Ngafeeson

Northern Michigan University, USA

INTRODUCTION

The vision to use information technology (IT) in healthcare to improve outcomes has been adopted by many a nation, including, the United States. If effectively implemented and efficiently leveraged, these technologies will greatly lower healthcare costs, improve safety concerns, and elevate the quality of healthcare (Blumenthal & Tavenner, 2010). No doubt, several industrialized nations are making significant investments in healthcare costing a substantial portion of their national gross domestic product (OECD, 2011). Recently, these investments have been in the area of health IT. Socio-economic factors such as aging population, increasing need for better healthcare, and rising health care costs have pressured governments to consider controlling costs while improving care (Romanow, Cho, & Straub, 2012).

Nevertheless, health IT implementation and adoption has met with its fair share of challenges. Reports show that physicians, nurses and other healthcare professionals continue to resist the use of technology within healthcare (Laumer et al., 2016; Dinev et al., 2016; Petrakaki et al., 2016). Researchers have proposed models to explain the *how* and *why* of user resistance to IT, and till date, there still is a lack of an overarching paradigm through which to examine this phenomenon. Hirschheim and Newman (1988) noted that resistance is a complex phenomenon which defies simple explanation and analysis; thereby requiring well-accepted theories or paradigms encompassing the full range of variables associated with an individual user's resistance of IT (Martinko,

Henry & Zmud, 1996). If not for anything else, governments and the healthcare sector need a return on investment on these systems that have been implemented. Understanding and mitigating user resistance is a clear step forward in improving sector-wide adoption.

This paper surveys extant literature on resistance, synthesizes its theories, and provide a paradigmatic lens for leveraging this knowledge in the practice of healthcare IT implementation. It is hoped that this study will not only increase theory-based understanding of the subject, but that it will provide managerial guidance for change managers and project leaders who live the effects of user resistance in practice. The rest of this paper is structured thus: a methodological as well as contextual background of the paper is set forth, the theories of resistance are then synthesized and analyzed, and the implications of the study are discussed.

BACKGROUND

In order to delimit the scope of this research, to synthesize and to analyze extant literature, we use Cooper's (1988) taxonomy of literature reviews. This model (Cooper, 1988) proposes a taxonomy of reviews based on six categories, namely: the focus, goal, perspective, coverage, organization, and audience of the literature. Given the objective of this study, the literature was reviewed with varying emphases based on Cooper's taxonomy. To achieve this, the review focused on different theories and models of resistance, and highlight the unique perspectives that these theories offer about

the phenomenon. First, we discuss a contextual background in the United States healthcare reform that has triggered nation-wide organizational change leading to resistance.

The Meaningful Use Reform and Change

On July 13, 2010, the United States administration rolled out a five-year transition plan for the U.S. healthcare industry to move from paper health records system to electronic health records. Over a ten-year period, the U.S. government plans to invest about \$70 billion to help the healthcare sector in this transition process (DHHS Press Release, 2010). This governmental mandate has been named, the Meaningful Use policy. It is arguably the greatest organizational change trigger in the U.S. healthcare system to date. Summarily, this reform program set the standards for health IT definitions, guidelines, and implementation.

The introduction of HIT into the healthcare industry has been associated with distinctive changes. Lorenzi et al. (2000) describes possible changes that typify healthcare organizational change namely:

- Operational changes. These affect the way the ongoing operations of the business are carried out, such as the automation of a particular area.
- Strategic changes. These occur in the strategic business direction, e.g., moving from an inpatient to outpatient focus.
- Cultural changes. These affect the basic organizational philosophies by which the business is conducted, e.g. implementing continuous quality improvement system in a clinical environment.
- Political changes. These involve staffing that occur primarily for political reasons of various types, e.g. those that occur at top patronage job levels in government agencies.

THEORIES AND CONCEPTUALIZATIONS OF RESISTANCE TO IT IN RESEARCH

H

User resistance to information technology is defined as a user's insistence to not use new IT. It is seen as a user's attempt to minimize his or her outputs while attempting to maximize and increase others' inputs (Joshi, 1991). This resistance, according to Markus, (1983) may range from subtler manifestations of resistance as passivity, misuse, low levels of use, and lack of use to harmful use (Marakas and Hornik, 1996; Martinko et al., 1996; Selander and Henfridsson, 2012). Marakas and Hornik (1996) discuss a form of resistance in which behaviors take "the form of overt cooperation and acceptance of the proposed system combined with covert resistance and likely sabotage of the implementation effort" (p. 208). Lapointe and Rivard (2005), on the other hand, mention the more overt side of resistance with scenarios where users "delivered an ultimatum, demanding that the system be withdrawn" (p. 477). In this study, resistance refers to covert or overt behaviors that oppose change towards the use of- or avoidance of an information system manifested as reactance, distrust, scrutiny or inertia (Ngafeeson, 2013).

The Interaction Theory

Markus's (1983) pioneering research in user resistance to information systems (IS) is arguably the departure point for most IS resistance research. Her seminal work basically responded to the question as to *why* people resist IS. Three types of theories are proposed: first, people are said to resist technology because of internal factors inherent in those who resist or their organizations. Second, resistance stems from external factors emerging from the introduced system. Lastly, resistance may also stem from the interaction of both internal and external factors. Markus termed the latter, the "interaction theory." While the theory of the interaction between the subject

of resistance and the object of resistance was not entirely new, Markus's work discusses this theory within an IT context.

Arguing from a group level and organizational level perspective, Markus (1983) demonstrates the importance of the interaction theory in understanding user resistance to information technology. Markus established a model which showed the interplay of power within the organization following the implementation of a new system. According to Markus, the interaction of an individual's characteristics with the specific system design features and the organizational context will determine user resistance. Markus then describes and evaluates these sets theories based on the underlying assumptions and tests them with data from a case study. Results showed that the interaction theory was a superior theory for explaining resistance for implementers than just from the technology or from the individual/organizational perspective.

Markus contends that the interaction theory is important in analyzing, diagnosing and finding solutions to organizational resistance to change introduced by information technology. This conceptualization, however, is limited in two significant ways: it views resistance from purely a political sense, and is likely to be more applicable for systems that are implemented across departments (Joshi, 1991).

The Equity Implementation Model (EIM)

Another significant theoretical model, the equity implementation model (EIM), was posited by Joshi (1991). Derived from the equity theory, the EIM attempted to explain resistance to change from a fairness perspective. The EIM proposes that individuals attempt to evaluate most changes. The changes considered as favorable are generally welcomed, while those considered as unfavorable are likely to be resisted. In a similar way, the equity theory suggests that in an exchange relationship, individuals are constantly concerned about inputs, outputs and fairness of the given exchange.

Hence, individuals are continuously comparing themselves with referent others and peers to assess if net gains are same. When these persons, therefore, experience a net loss in equity, the equity theory suggests that the people are likely to become distressed. This distress caused by inequity then leads the concerned individual to resist the change by lowering his or her inputs while trying to minimize the outputs of others. Building on the equity theory, Joshi (1991) proposed a three-level analysis of equity namely: a comparison of self with self, a comparison of self with employer, and a comparison of self with co-workers. According to the EIM, users will resist use, if inequity was perceived at any of these three levels, following the introduction of a new system.

In summary, EIM postulates that people analyze equity at three major levels, namely: at *self*, *organizational*, and *peer* levels. At level one (*self*) equity evaluation, individuals compare their inputs with outputs and calculate their net gains. If the new system brings with it net gains in equity such as less tension and better working conditions, the individual is more likely to embrace the change. On the contrary, if the system brings the individual a net loss in equity such as loss of seniority or increased workload, then the individual is likely to oppose change. At the second level of equity (*organizational*) evaluation, an individual compares equity net gains, in much the same way as at the first level: only this time, the individual compares their personal net gains to that of the organization. If the net gains are positive, change is welcome; but if the individual perceives that the organization's net gains are more than theirs, resistance is likely to result. A third level of analysis (*peer* level) is also proposed in which an individual compares their net gains with those of their peers and referent others after the introduction of the new system. Here, the user asks questions like: "Does the new system treat each user equitably? Or, does it increase others' equity while lowering those of others?" If the assessment is that personal equity is increased, change is received; otherwise, it is resisted.

A possible critique for using the EIM as a stand-alone model for explaining resistance is that inputs and outputs are difficult to name and measure. This is especially so because these inputs and outputs could be tangible as well as intangible—the measure of which might be as challenging as impossible. As Joshi (1991), himself, states: “the nature of changes may also make it difficult for users to make objective assessment.” This suggests that, a need for a model that would be sensitive to other types of changes is a welcome necessity.

The Attributional Model of Reactions to IT (AMRIT)

Martinko, Henry and Zmud (1996) begin their theory-building process on the premise that, hitherto, there had not been any “well accepted theory or paradigm encompassing the full range of variables associated with an individual user’s resistance of IT” (p. 313). Drawing from extant research in social psychology and organizational behavior, Martinko et al. (1996) proposed an attributional model to explain IT resistance (AMRIT). Based on previous literature that utilized the attribution theory and learned helplessness, Martinko and associates develop a model that they suggest could serve as a “basis for a more comprehensive theory” of worker resistance to the introduction of new technology in the workplace. According to the attribution theory, an individual’s behaviors are a function of his or her beliefs about the possible outcomes. More specifically, it looks at the perceptions of what causes these outcomes. The AMRIT posits that individuals’ attributions are informed by their experience with successes and failures. It suggests that individuals interpret their behaviors in terms of the causes. Consequently, their behaviors in the future are constructed from the expectancy that comes from these attributions. In other words, success or failure with prior implementation of a technology may cause implementers to develop

certain attributions to why this happened. These attributions then create certain expectations with regard to the technology’s efficacy, leading to behaviors or feelings about the system that can be either negative or positive. Like many of the other models before this one, the Martinko et al.’s (1996) model shed more light on the nature of IS user resistance; however, this conceptualization ended at a purely conceptual level.

The Lapointe and Rivard (2005) Model (L-R)

Lapointe and Rivard (2005) examined group resistance dynamics at three levels namely: pre-implementation, during implementation, and post-implementation of an IS. The L-R model was developed to explain a multilevel dimension of resistance to IT implementation; arguing from the perspective that previous models had only looked at resistance at one level of the organization instead of examining different levels of the organization. Building on existing research, the L-R model, has proven to be an important model for a closer examination and extension. In summary, the L-R framework posited that resistance to an information system results from perceived threats which in turn evolve from a complex interplay of political and interpersonal/group factors resulting from people’s interaction with an IS. Lapointe and Rivard (2005) termed these “interaction effects”, initial conditions. Simply put, when an individual or an organization interact with a newly installed technology in the workplace, resistance targeting either the technology itself, the advocates of the technology or the significance of the technology may result. The L-R model also posited that resistance to information technology is mediated through *perceived threats*. The L-R model serves as a generic process model that provides a good departure point for more empirical investigation. For one thing, the L-R model incorporates previous frameworks and deals with user resistance from a multilevel and process perspective.

The Status Quo Bias Theory

The status quo bias perspective was proposed by Kim and Kankanhalli (2009). It is one of the few theoretical models of resistance to information technology that has been proposed and empirically tested. Kim and Kankanhalli (2009) contended that people resist an information system due to their desire to stay in their current state or *status quo*. In other words, people exhibit a bias towards maintaining their “old habits” when confronted with the choice of change (see Hsieh, 2015).

In the context of technology-enabled change, therefore, the *status quo* bias perspective holds that people may resist the implementation of new technology, for the sole purpose that they are comfortable with the current state. Because of this desire to maintain the *status quo*, they judge the decision to change as a cost associated with this change, rather than the benefits thereof. By testing their model within the context of a new enterprise system, Kim and Kankanhalli (2009), found that the cost of switching to a new system from an old one caused user resistance. The attempt to empirically test user resistance of IS was probably the hallmark of Kim and Kankanhalli’s (2009) work.

IT Conflict-Resistance Theory (IT-CRT)

Situating their research within a pre-implementation context, Meissonier and Houzé (2010) propose the IT conflict-resistance theory and use the framework as a basis for a two-year action research project. The research suggests that pre-implementation resistance to IT could be used as a gauge to assess possible post-implementation concerns. Resistance in this sense, therefore, is seen as positive tool for technology acceptance in the post-implementation phase.

Like previous works, Meissonier and Houzé (2010) highlighted socio-political factors as well as task-oriented factors as salient determinants of conflict and resistance. However, being an action

research in itself, the conceptual model proposed, suggests a more heuristic and practical approach to the topic than a testable concept.

Cynicism Theory

Selander and Henfridsson (2012) conceptualized *cynicism* as resistance. Building on the Lapointe and Rivard (2005) process framework, Selander and Henfridsson (2012) propose three dimensions of user cynicism as the precursors of resistance. They argue that cynicism is a remote form of resistance that is not mediated through *perceived threats*, but is rather manifested as a form of *passive resistance*. They also suggest that cynicism, being a passive resistance in itself, could escalate to other forms of resistance. The three dimensions of cynicism conceptualized were *cognitive distance*, *negative affect*, and *seeing through espoused claims*. *Cognitive distancing* here refers to the way employees of an organization alienate themselves from the management’s policies and stance. *Negative affect* on the other hand is defined to involve negative emotions towards managerial behavior including defeatism, betrayal, and disillusionment. *Seeing through espoused claims* supposes that the employee possesses knowledge of how things could better be done and is aware of managerial inconsistencies and repudiates these assertions.

In sum, Selander and Henfridsson (2012) contend that these three dimensions of user cynicism constitute a form of *passive resistance*. Using this framework, they built upon the Lapointe and Rivard (2005) framework. Their work however deferred from the Lapointe and Rivard model in that it completely ignored *perceived threats* as mediator between *initial conditions* and *resistance*. Like other models before it, it remained at a purely conceptual level: suggesting three antecedents of user resistance, but going no further.

Alternative Theories of Resistance

The following section discusses theories that either have been used to explain user resistance,

but have not been leveraged in IS literature, or theories that inform our knowledge of user resistance, albeit not fully. Some of these theories include: the technology acceptance model, the psychological reactance theory, and the cognitive dissonance theory.

The Technology Acceptance Model (TAM)

TAM has been used to predict user acceptance of technology in different IS studies (Venkatesh & Davis, 2000; Mathieson, 1991). TAM is based in the *theory of reasoned action* (Fishbein & Ajzen, 1975) and is now widely regarded as the standard model for examining *technology acceptance*. Because of the failure by employees to accept and use technology, organizations as well as technology vendors see *user acceptance* of technology as crucial to organizational survival. The TAM in its simplest form suggested that *perceived usefulness* and *perceived ease of use* would predict *user acceptance* of an information technology. In other words, Davis (1989) posited that the acceptance or rejection of a technology is a function of how individuals perceived that the technology would be either useful or how easy it could be to use the given technology.

The Psychological Reactance Theory (PRT)

Another alternative explanation of resistance is the PRT. Brehm's PRT is one of the most widely accepted resistance theories in social psychology. It has also been widely used to explain resistance and persuasion—its antithesis. PRT is built around the notion of “free behaviors” and “freedoms” It assumes a “conservative stance on people's desire for change” (Brehm & Brehm, 1981, p. 25). The PRT is based on the assumption that individuals generally believe that they have specific *behavioral freedoms*. When these freedoms are threatened,

individuals are aroused by the motivation to reassert their freedoms. The *psychological reactance theory* assumes that people's behaviors are motivated by the desire to protect their “freedom” to carry out a particular behavior in a particular context. This “protection” of freedoms is then manifested as resistant behaviors.

The Cognitive Dissonance Theory (CDT)

This theory holds that, “when an individual holds two or more elements of knowledge that are relevant to each other but inconsistent with one another, a state of discomfort is created” (Harmon-Jones et al., 2010). The resulting discomforting state is called *dissonance*. Because dissonance originates from the conflicting views or beliefs about one's self, the term *cognitive dissonance* is used to collectively describe the concept (Festinger, 1957; p. 9). According to the dissonance theory, this inconsistency or “lack-of-fit” of cognitions motivates the individual to be involved in a psychological effort to reduce the inconsistency between the cognitions. It is this effort to reduce inconsistency that leads to *resistance*.

According to the *cognitive dissonance theory*, user resistance can be conceptualized as follows. When an individual's intention or action to reduce inconsistency is to rationalize or support his present state of cognition: such that “new knowledge” is considered as dissonant with the individual's present cognition, the consequent behavior can be described as resistance. Simply put, resistance is an implicit or explicit intension that results to a behavior that opposes change towards a particular “new” attitude or behavior. Notwithstanding, the CDT has shortcomings. Harmon-Jones and Harmon-Jones (2002) argued that Festinger's (1957) theory stopped short of explaining *why* individuals do not like to be in a state of cognitive inconsistency.

IMPLICATIONS AND FUTURE RESEARCH

This study examined user resistance to information systems from several theoretical angles within the wider context organizational change. The claims and assumptions of relevant extant theories and models were specifically discussed. The results of the synthesis of resistance to information technology literature can be summarized as follows:

- Resistance is a complex phenomenon: requiring different theorizations for a more comprehensive understanding.
- Resistance attitudes should be studied from an unbiased perspective: that is, not regarding resistance as either positive or negative; rather, letting the context define its nature.
- Resistance attitudes may last through pre-implementation, implementation, and post-implementation phases of an information system.
- Resistance behaviors could originate from a single individual as well as groups of individuals.
- Resistance is a result of the dynamic interaction between people, technology and the processes and procedures within an organizational environment.

In general, this research showed that there is a lack of an overarching paradigm of user resistance to information systems in spite of the burgeoning alternative theoretical explanations of user resistance to IT that have been proposed. Second, there is a noticeable lack of empirical investigations. Most of the research so far has been dominated by case studies, action research and other qualitative methodologies. Consequently, many of the proposed models are yet to be validated. Third, not many studies have specifically looked into user resistance in the context of health information technology. Lastly, with the several theories

and perspectives already proposed, there still is a lack of a comprehensive theoretical framework that can be tested.

Theoretical and Practical Implications of Research

This work offers implications and contributions to theory. First, the study extends the body of literature by combining social psychological theories, information systems theories and a policy framework to explain the concept of resistance. It broadens our understanding about user resistance by introducing social psychological theories that are often not leveraged in information systems to explain resistance. As has been argued by other researchers before, resistance needs to be addressed from a more comprehensive theoretical standpoint. Future research can incorporate these theoretical perspectives to better understand resistance to IT.

The study also has implications for practice. In a healthcare environment fraught with technology-induced organizational change, change managers' understanding of the importance of resistance is critical. Knowles and Linn (2004, p. 9) have suggested that understanding resistance, would shed light about the role of persuasion—its antithesis. Persuasion is seen as a very important antidote to resistance. Persuasion skills are expected, if not required of change managers' portfolio of abilities. This study provides change managers with empirically tested knowledge for decision-making.

Overall, this study points to the critical role of organizational support in the change management process. Management would need to provide sensitization, guidance, training and opportunity for employee feedback before, during and after the implementation of a new system. Just as some researchers have recommended, top management support in organization-wide implementation is critical (DeLone, 1988). Reducing threats and mitigating resistance will clearly require the need for management support.

Suggestions for Future Research

Future research could use this study as an important departure point to answer some important questions: (1.) Do specific *threats* lead to specific dimensions of user resistance? (2.) Are there more types of perceived threats that are salient to IT user resistance? (3.) Are there any other theories that can inform our knowledge on possible *initial conditions* yet to be uncovered? These are some areas in which future research could take. Additionally, finding which type of *threats* and *initial conditions* contribute to resistance during the different stages of implementation could be particularly insightful. For example, if individual factors were found to be particularly associated with user resistance during the pre-implementation phase, then, change management programs could be designed to target causative initial conditions.

CONCLUSION

The introduction of electronic health records into the United States healthcare system is expected to increase legibility, reduce medical errors, shrink costs and boost the overall quality of healthcare. Early field reports about this transition reveal that healthcare professionals including physicians and nurses are resisting the full use of these systems. This study proposed key theories from social psychology, and information systems to provide a multi-perspective look at the concept of resistance to information technology. These theories, together, explain and expound on resistance in a way that can increase our understanding. The application of these findings has the potential to improve the chances of successful IT implementation within the healthcare industry in particular.

REFERENCES

- Bhattacharjee, A., & Hikmet, N. (2007). Physicians resistance toward healthcare information technology: A theoretical model and empirical test. *European Journal of Information Systems*, 16(6), 725–737. doi:10.1057/palgrave.ejis.3000717
- Blumenthal, D., & Tavenner, M. (2010). The Meaningful Use Regulation for electronic health records. *The New England Journal of Medicine*, 363(6), 501–504. doi:10.1056/NEJMp1006114 PMID:20647183
- Brehm, J. W., & Brehm, S. S. (1981). *Psychological reactance: A theory of freedom and control*. San Diego, CA: Academic Press.
- Cooper, H. M. (1988). Organizing knowledge synthesis: A taxonomy of literature reviews. *Knowledge in Society*, 1(1), 104–126. doi:10.1007/BF03177550
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technologies. *Management Information Systems Quarterly*, 13(3), 319–340. doi:10.2307/249008
- DHHS. Press Release 07/13/2010. Retrieved on 10/30/2010 from <http://www.hhs.gov/news/press/2010pres/07/20100713a.html>
- Dinev, T., Albano, V., Xu, H., D'Atri, A., & Hart, P. (2016). Individuals' Attitudes Towards Electronic Health Records: A Privacy Calculus Perspective. In *Advances in Healthcare Informatics and Analytics* (pp. 19-50). Springer International Publishing.
- Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford, CA: Stanford University Press.
- Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention and behavior: An introduction to theory and research*. Reading, MA: Addison-Wesley.

- Harmon-Jones, E., Amodio, D. M., & Harmon-Jones, C. H. (2010). Action-based model of dissonance on cognitive conflict and attitude change. In *The Psychology of Attitude and Change* (pp. 163–181). New York: Psychology Press.
- Harmon-Jones, E., & Harmon-Jones, C. (2002). Testing the action-based model of cognitive dissonance: The effect of action-orientation on post-decisional attitudes. *Personality and Social Psychology Bulletin*, 28(6), 711–723. doi:10.1177/0146167202289001
- Hirschheim, R., & Newman, M. (1988). Information systems and user resistance: Theory and practice. *The Computer Journal*, 31(5), 1–11. doi:10.1093/comjnl/31.5.398
- Hsieh, P. J. (2015). Healthcare professionals use of health clouds: Integrating technology acceptance and status quo bias perspectives. *International Journal of Medical Informatics*, 84(7), 512–523. doi:10.1016/j.ijmedinf.2015.03.004 PMID:25842155
- Joshi, K. (1991). A model of users perspective on change: The case of Information systems technology implementation. *Management Information Systems Quarterly*, 15(2), 229–242. doi:10.2307/249384
- Kim, H. W., & Kankanhalli, A. (2009). Investigating user resistance to information systems implementation: A status quo bias perspective. *Management Information Systems Quarterly*, 33(3), 567–582.
- Knowles, E. S., & Linn, J. A. (2004). Alpha and Omega Strategies for Change. *Resistance and Persuasion*, 117.
- Lapointe, L., & Rivard, S. (2005). A multilevel model of resistance to information technology implementation. *Management Information Systems Quarterly*, 29(3), 461–491.
- Laumer, S., Maier, C., Eckhardt, A., & Weitzel, T. (2016). Work routines as an object of resistance during information systems implementations: Theoretical foundation and empirical evidence. *European Journal of Information Systems*, 25(4), 317–343. doi:10.1057/ejis.2016.1
- Lorenzi, N. M., & Riley, R. T. (2000). Managing Change: An Overview. *Journal of American Informatics Association*, 7(2), 116–124. doi:10.1136/jamia.2000.0070116 PMID:10730594
- Marakas, G. M., & Hornik, S. (1996). Passive resistance misuse: Overt support and covert recalcitrance in IS implementation. *European Journal of Information Systems*, 5(3), 208–219. doi:10.1057/ejis.1996.26
- Markus, M. L. (1983). Power, politics, and MIS implementation. *Communications of the ACM*, 26(6), 430–444. doi:10.1145/358141.358148
- Martinko, M. J., Henry, J. W., & Zmud, R. W. (1996). An attributional explanation of individual resistance to the introduction of information technologies in the workplace. *Behaviour & Information Technology*, 15(5), 313–330. doi:10.1080/014492996120085a
- Mathieson, K. (1991). Predicting user intentions: Comparing the technology acceptance model with the theory of planned behavior. *Information Systems Research*, 2(3), 173–191. doi:10.1287/isre.2.3.173
- Meissonier, R., & Houzé, E. (2010). Toward an IT Conflict-Resistance Theory: Action research during IT pre-implementation. *European Journal of Information Systems*, 19(5), 540–561. doi:10.1057/ejis.2010.35
- Ngafeeson, M. N. (2013). *Understanding User Resistance to Information Technology: Toward A Comprehensive Model in Health Information Technology* (Doctoral Dissertation). Proquest/UMI Manuscript No. 10530.

Petrakaki, D., Klecun, E., & Cornford, T. (2016). Changes in healthcare professional work afforded by technology: The introduction of a national electronic patient record in an English hospital. *Organization*, 23(2), 206–226. doi:10.1177/1350508414545907

Romanow, D., Cho, S., & Straub, D. (2012). Editor's comments: riding the wave: past trends and future directions for health IT research. *Management Information Systems Quarterly*, 36(3), III–A18.

Selander, L., & Henfridsson, O. (2012). Cynicism as user resistance in IT implementation. *Information Systems Journal*, 22(4), 289–312. doi:10.1111/j.1365-2575.2011.00386.x

Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science*, 46(2), 186–204. doi:10.1287/mnsc.46.2.186.11926

ADDITIONAL READING

Armenakis, A. A., & Bedeian, A. G. (1999). Organizational change: A review of theory and research in the 1990s. *Journal of Management*, 25(3), 293–315. doi:10.1177/014920639902500303

Dent, E. B., & Goldberg, S. G. (1999). Challenging resistance to change. *The Journal of Applied Behavioral Science*, 35(1), 25–41. doi:10.1177/0021886399351003

Jones, G. R. (2010). *Organizational theory, design, and change*. Upper Saddle River, NJ: Pearson.

Waddell, D., & Sohal, A. S. (1998). Resistance: A constructive tool for change management. *Management Decision*, 36(8), 543–548. doi:10.1108/00251749810232628

KEY TERMS AND DEFINITIONS

Cognitive Dissonance Theory: The theory suggests that when there exists a “lack of fit” of cognitions, a state of discomfort called *dissonance* is aroused, triggering a reaction to reduce it.

Cynicism Theory: Proposes a passive form of resistance that is not manifested through perceived threats.

Interaction Theory: Proposed by Markus (1983), it proposes that IT resistance stems from the interaction between people, the technology in question, and the environment surrounding the technology.

IT User Resistance: The overt or covert attitude or behavior towards an IT system resulting in avoidance, underuse or outright nonuse of an IT system.

Meaningful Use: The U.S. government mandate of 2010 requiring an eligible provider to use certified technology to meet specified government benchmark outcomes.

Psychological Reactance Theory: This theory posits that when individual freedoms are threatened, and uncomfortable emotional state is created that necessitates a return to “normal”, which may manifest as resistance.

Status Quo Bias Theory: Posits that, people, who are comfortable with the current state of affairs, may resist the change that will ensue following the implementation of a new system.

Category H

Higher Education

The Effect of Innovative Communication Technologies in Higher Education



Stavros Kiriakidis
University of Crete, Greece

Efstathios Kefallonitis
State University of New York at Oswego, USA

Androniki Kavoura
Technological Educational Institute of Athens, Greece

INTRODUCTION

The aim of this chapter is to: (1) present a review of some innovative communication technologies in higher education and how these are implemented for organizational, marketing and learning processes, (2) illustrate how Information and Communication Technologies (ICTs) may drive the introduction and use of plethora of inventive activities in education.

Current trends in higher education institutions (HEIs) are associated with supplying the student with those learning methods that can lead to skills and contribute to his/her creative and individual thought (Cuenca et al., 2015). At the same time, teaching instructors need to be aware of the know-how in the communication and information technology sectors that can be further employed in different fields of research. In parallel the use of new technologies in the delivery of courses offers HEIs competitive differential advantages.

Financial cutbacks and limited government support for institutions of higher education, led to: (a) either a withering effect for them that shrink or merge in order to survive or (b) find innovative ways to survive and attract new students and audiences as potential users while making con-

nections with the community and firms (Bowen, 2013). The implementation of innovative communication technologies with the use of ICTs via computers, mobile phones, telephone lines to access and manage electronic data on the web allows users to actively participate in the learning process, provides lifelong learning opportunities and offers a competitive advantage to HEIs that have not fully implemented the possibilities on offer. HEIs may further adopt ICTs to communicate and promote their services on offer to attract existing and potential audiences but also to initiate technologies in order to provide the most updated tools for their learning process. Virtual and online learning processes cannot substitute for the physical experience, nonetheless, they may contribute to its enhancement.

For innovation to take place in higher education, the organizational culture of institutions of higher education and leadership should support such initiatives (Zhu, 2015).

This chapter presents an overview of the technological advances that enhance communication and strengthen innovation in higher education. A review of literature along with the use of popular business sources provides an as-is on the topic.

BACKGROUND-THE USE OF NEW TECHNOLOGIES AS A MARKETING TOOL IN HIGHER EDUCATION

The Organizational Perspective

The flexibility in using new technologies in HEIs is reflecting the socio-economic changes affecting the needs of the student population across the world (Narduzzi & Campbell, 2015). The characteristics of students are continuously evolving reflecting expectations for increased mobility, distant-learning, internationalization of the curricula and the introduction of new disciplines of study. New technologies such as learning management systems (LMSs) and web-based learning programs (BlackBoard Learn, Moodle, Brightspace by Desire2Learn(D2L), etc.) that extend to applications available on handheld devices such as BlackBoard Learn-Mobile Learn, Collaborate, Podcasts and iTunes U, or user-generated content (Wikipedia, Youtube, etc.) address some of the needs of learners. These applications offer some access to the functions available in the full-version web-based platforms. The mobile applications serve as enhancers rather than a mobile equivalent to the LMSs platforms.

The LMSs provide institutions with the advantage of offering credit-bearing instruction online. Access to these is course-specific and requires the use of a username and password available through an academic course provider. Using LMSs, schools enhance existing courses, offer hybrid options and/or introduce entirely online classes and programs of study.

Blackboard Learn in particular is a learning management system that offers both the capability of offering educational courses online and managing a learner's account through the same platform. It becomes a useful tool for online learners or students' on the go. Blackboard is a paid service.

Moodle is an unrestricted and open-source software that was introduced as a means to enhance courses or used a tool for hybrid courses taught partially in a classroom setting and partially on-

line. This is now the world's second most popular learning management system after Blackboard.

Brightspace by D2L, is another popular learning management system, providing an alternative to Blackboard and Moodle.

A number of free non-credit-bearing applications are used to enhance and compliment courses. iTunes U is such an application available on apple's iTunes store. Universities and schools across the world are able to create school-specific applications and content available to a worldwide audience.

Wikipedia and YouTube are both free worldwide user-generated platforms. Wikipedia is one of the world's largest user-generated content platforms referred to as an encyclopaedia. This cannot be considered as a scholarly source of reference due to its content-generation principle.

YouTube is a video-sharing platform that allows individuals and organizations to upload and share video clips, music clips, etc. Often academic establishments use YouTube to share lectures, seminars, product demonstrations or training videos.

Marketing and public relations departments of HEIs in charge of promoting and communicating to the public the institutions and services available are implementing new technologies and the use of internet in order to cater for existing and potential students' needs who are adept in internet applications (Siamagka & Christodoulides, 2016). In that way, people become aware of the methods and techniques on offer by HEIs and this may attract more potential students in the institutions that provide and promote innovative services and methods related to higher education. The tools could be implemented to promote the educational institutions and their activities. Both national and international students can be attracted while online learning can enhance collaborations between institutions.

In parallel, new technologies allow HEIs to cater for:

1. Accreditation needs (State, country, or by independent agencies – AACSB for Business schools, etc.).
2. Marketing and promotional activities (internally to current students, externally - alumni, prospective students, etc.). This strengthens the institutional brand and provides tangible advantages (Khanna, Jacob, & Yadav, 2014).
3. Promoting and enhancing healthy competition among institutions.
4. Societal changes and new trends by presenting an image of care to the student population.

The use of new technologies by HEIs gives education establishments a differential advantage (see Table 1). It also creates open, mobile, flexible and accessible educational experiences that cater to the needs of the learner on-the-go. In particular, new technologies are used as an enhancement or extension of courses that target a wider audience of learners. In addition, they provide a virtual learning environment and help with the enhancement of courses and training with digital media and make it easier to provide learner feedback regarding tracking, reporting and managing a learner's account.

The User's Perspective

Higher education aims to provide competencies to students associated with their ability to communicate and handle different scenarios throughout their lives by using creative tools while working as part of a group (Cuenca et al., 2015).

HEIs must continue to experiment and share best practices in providing students with high quality learning experiences that are globally engaging, intellectually rigorous, and personally rewarding. This is part of a growing global trend that recognizes new technologies in their enhancing knowledge availability and retention (Bulfin, Pangrazio, & Selwyn, 2014) (see Table 2).

Innovation competency is associated with the ability of students to make self assessments via learning methods acquired that will help them to

Table 1. Advantages and disadvantages of the use of innovative technologies in HEIs from an organizational perspective

Advantages	Disadvantages
Marketable advantages (Competitive advantage in competition)	Variability and complexity of technological advances
Flexibility & accessible education on-the-go	Time invested to learn and use
Enhancement or introduction of new courses	Compatibility with other forms of technology and existing infrastructure
Administrative ease (managing a learner's profile)	Applicability and sustainability issues

Table 2. Learning advantages and disadvantages of new technologies from a user's perspective

Advantages	Disadvantages
Enhancement of creativity and innovation for the user	Increased workload
Flexibility when and how to participate in the educational programs	Social and communication skills may become less cultivated. The dynamic result of teamwork may be significantly reduced
Encouragement of knowledge retention (Enthusiasm)	The individual needs to be adept with innovative tools in order to employ them
Ease of self-assessment and monitoring of progress	The personal contact and interaction is missing
Development of personal and professional skills (Confidence)	The individual becomes socially isolated

communicate and interact in multicultural and multidisciplinary environments (Cuenca et al., 2015).

Researchers have highlighted the need of equipping students with the skills to be employable. These may include adapting academic curricula to emphasize skills, introduce and new instructional methods, use information technology, service learning, career planning, students' logbooks, extra-curricular activities, work-based education and other (Asonitou, 2015; Blackwell et al., 2001).

United States of America (USA) students are part time or even full time while raising children

and attending classes and thus, educational programs need to take into account their audiences' needs (Knowles, Holton III, & Swanson, 2014). In addition, there is pressure for embracing further minorities (López, 2015), different races (Hiraldo, 2015) and sustainable academic program development (Adomßent et. al, 2014; Godemann, Bebbington, Herzig, & Moon, 2014).

When educators use new technologies such as handheld computing devices (tablets) as a method of instruction, they create a notion of enthusiasm and hype in class (Nguyen, Barton, & Nguyen, 2015). This boosts student confidence and strengthens their willingness to embrace new learning technologies. The adoption of new technologies in HEIs is also faced with some skepticism by some (Muller, 2014), as there are fears that it may interfere with the student workload (Gregory & Lodge, 2015) (see Table 2).

Based on the fact that academic programs are not geographically defined anymore, initiatives have developed so that both teaching staff and students can benefit. In Australia for example, emphasis has been put on teaching staff's professional development where through open line courses, professors share knowledge with another all over the world in regard to educational design (Mirriahi et al., 2015). Staff interacts, improves digital literacy and shares knowledge.

The tools presented in the next sections may be to value for both students and teaching staff but also to the society and community where universities are located based on the innovation that is transmitted there. Academics may implement social media to approach students. This has taken place in many cases with a successful engagement between student staff communication of ideas (Sessa, 2014).

Technology Transfer

Cooperation between universities and/or firms is often associated with patenting and licensing of academic inventions and their connection to

industrial innovation; this brings forth the technology transfer element (Mowery, Ielson, Sampat, & Ziedonis, 2004). Among the technology transfer mechanisms patents, copyrights, trademarks, licensing agreements between the university and private firms, university-based start-ups but also technology parks are incorporated (Link, Siegel, & Wright, 2015). In regard to US Universities, the Bayh-Dole Act allowed the use of University laboratories to implement technological advancements that can be further useful to communities and firms (Link et al., 2015; Mowery et al., 2004). Since there are financial cutbacks, universities have an alternative way to find sources through research and technology (Siegel & Wright, 2015). There is room for further growth regarding technology transfer in HEIs pertaining entrepreneurship opportunities (Lackéus & Middleton, 2015). The necessity for technology transfer could bring together academic expertise and streamline this with industry needs (Wright, 2014).

Medical research centers of Universities are known to excel at technology transfer among past and current trainees and scientists. As an example, Emory University in Atlanta, USA has developed tools and invested in infrastructure that enables technology transfer. The development of the Emory Cardiac Toolbox software is an example of this approach.

Innovation is created within the technological parks of the Universities. It is typical to note that this is usually the result of teamwork that can be developed as part of the generic competencies of those involved.

The teamwork that is created within the technological parks of the Universities is a miniature of the real world. These students will work in the market, in companies and they need to handle skills and competencies that have learnt while participating in higher education's environment. That is why, project team work may result in technological innovation and participants in technology transfer, need to find ways and skills to cooperate with each other as a significant at-

tribute of their behavior that leads them to work for technological innovation (Vick, Nagano, Author Vitae, & Popadiuk, Author Vitae 2015).

It has been found that the workforce that is consisted of similar high educational background, illustrates a positive innovation performance (Subramanian, Choi, Lee, & Hang, 2015). We should also take into consideration that when many people work in a project or in an office as those specially designed for technology transfer, conflicts may arise. Teamwork and sharing of responsibilities may prove to be of help for this sector of research.

Win-win solutions from business communication and social psychology theories where all the participating parts in the technology transfer, both the higher education institution and the management of the business that are involved in the technology transfer (O'Kane, Mangematin, Geoghegan, & Fitzgerald, 2015).

Libraries can also be a valuable partner in universities for implementing innovation and transfer knowledge. Hoppenfeld and Malafi (2015) argue that there is an initiation of cooperation with centers of entrepreneurship and two public and academic libraries in Texas; they promote their resources and databases for this reason to the public but also to students interested in engaging with entrepreneurial activity.

Experiential Learning and ICTs

Experiential learning is associated with the uncover of subjective experiences at a setting where the learner has direct contact with the phenomenon under examination; thus, learners bring into perspective the content of what is presented in the curriculum via the courses and they do not have to create ideas from scratch (Bublitz & Philipich, 2015).

Other areas that can initiate and employ knowledge transfer exist within HEIs that include experience-based education, internships and cooperative learning experiences. In these cases, students benefit from working with an outside

organization, applying theory and processes learned as well as acquiring business skills in practice. Some HEIs require student to complete such a program during their third year of undergraduate studies. This enables students to bring back to class professional skills and excel at the remaining duration of their studies.

Mobile and technology-based learning, digital game-based-learning, innovative e-learning strategies or tools in web information, the use of tablet PCs (TPCs) in classroom instruction can contribute significantly so that technology is employed for the student's active participation in learning activities (Hung et al., 2015)

Internship programs or work-placements have been recognized as one of the most efficient methods to bridge the gap between education and employment requirements (Asonitou, 2015).

The knowledge acquired while in university based on the theory gained from the study program, can be combined with practice through projects associated with a working environment. This enables students to get used to teamwork and to develop skills that are necessary in order to be competitive in the market (Sulas & Baynar, 2015).

In order for surgical educators to be able to communicate and exchange ideas between peers WIGGIO is an online networking and collaboration tool that brings people with similar professional and scientific interests together and had been implemented at the Montreal General Hospital Canada. Petrucci et al. (2015) found that in fundamental laparoscopic surgery, WIGGIO in fact, improves manual skills based on simulator experience and training but also with the use of video since members of the groups interact with each other and exchange ideas about scientific papers that they could read while post online important events on the online calendar so that all members can be informed about them.

Another experiential learning example takes place in the Georgia Institute of Technology that organizes the Georgia Tech (GT) Convergence Innovation Competition (Georgia Tech, 2015). The Georgia Institute of Technology initiates the GT

journey that calls participants to develop mobile, web and desktop applications with the contribution of resources of the camp and community, industry and academic support. All these support student innovation as a living lab (Sanders et al., 2015). The Convergence Innovation Competition (CIC) is a semester long event that motivates students to use laboratories, professors; guidance, industry's guidance and campus resources in a lab space in order to create innovative products and experiences. This form of a living lab can be associated with experiential learning and at the same time, allows for technology transfer.

Among the best practices presented in experiential learning, is the application of theory to practice via for example, the student hands on experience with the run free clinic that was created in association with the University of Michigan in Flint while direct effects to the community emerge (Kruger, Kruger, & Suzuki, 2015).

Learning experiences are congruent with learning styles that in turn influence learning styles while depending on whether the theory base plays a significant and important role, the development of useful devices and processes are implemented or not (Kolb, 2014).

The use of innovative tools can be further used to teach students in higher education how create competitive electronic portfolios with the use of infographics and multimedia tools that can employ sight and sound and create stories for the owners (Haverkamp & Vogt, 2015).

Distance Learning Education and ICTs

Distance education has employed classes in virtual world learning that offers participants learning and emotional experiences and skills (Reinsmith-Jones, Kibbe, Crayton, & Campbell, 2015).

The use of digital mobile devices as a technique of ICT innovation found that when employed in Spanish universities, improved the way of teaching in distance learning but also contributed to generic competencies (Sevillano-García & Vázquez-Cano,

2015). The use of tablets and smartphones in higher education contributes to the development of oral and writing skills that allow people to effectively communicate with each other, the development of teamwork where team responsibilities are allocated with teammates, to develop interpersonal effectiveness where people mingle with different people (Hogg & Vaughan, 2008). Mobile devices are thus, used not for learning reasons but to be able to connect online and be able to collaborate with other people.

Lifelong and Continuing Education and ICTs

The meaning of lifelong learning (Falk, 2014) is continuously evolving considering the accessibility, availability to all and depth of knowledge today's learners have (Head, Van Hoeck, & Garson, 2015). Lifelong learning also accounts for community wellbeing among adult learners (Merriam & Kee, 2014). Current trends suggest that lifelong learning:

(a) helps learners achieve the most up-to-date education, (b) is cost effective, (c) is inclusive and eliminates elimination of learners, (d) supports the creation of a sustainable support system, for learners, (e) can be self-drive and assessed, (f) can accommodate life-work balance (Hanemann, 2015; Neimeyer & Taylor, 2014). The creation of teamwork projects allows individuals to communicate within the work they participate, having the maturity and responsibility to act responsibly.

People can be trained and learn throughout their lives at the right and convenient time and place for them, are motivated to continue to gain knowledge at their own pace, and thus, are able to reach their goals in the community or business environments (Martin & Pear, 2015). Since school in the traditional sense of participating in class for some years when one is a student at a younger age cannot equip learners with all the skills they require throughout their lifetimes, lifelong learning and the implementation of new technologies are adaptable to be convenient for learners, are

personalized and unobtrusive (Zolinas, 2015). Lifelong learning may contribute to smooth group behavior and decision making since the individual will acquire continuous knowledge, resources and skills that may contribute to self-fulfilment and better evaluation of the social world and interpersonal relations (Franzoi, 2012).

Open access courses take place at the higher institution on broader topics and these can be attended by the community that may be interested and benefit from such lectures and open discussions. New technologies allow for such process to take place via web based technologies.

Coursera, a formerly non-credit bearing for-profit massive open online courses (MOOCs) provider has been gradually offering courses for academic credit. It will be interesting to see how this learning platform provider will shape online and distant education in the years to come in comparison to BlackBoard, Moodle and Brightspace by D2L.

Executive Education in HEIs and Innovation with ICTs

Corporate-level learning, often promoted as corporate university, provides personalized reliable education for the benefit of employees and the company (Stephens & Margey, 2015).

A special focus on executive education programs has been given on public administration programs in Egypt (Wafa, 2015).

The working knowledge can be combined with academic knowledge where executive learners aim to understand their workplace as this was the case of executive education in Ireland. Waller and Fawcett (2015) argued for the executive education as an integrator among the whole university.

Combining high quality research from universities and practice consists of a challenge for universities that aim to promote entrepreneurial education that is incorporated in their curricula either within other program or as a distinct program (Vanevenhoven & Drago, 2015).

A custom executive education program charter can be posed as a key task to complete at the outset of a new custom executive education provider/client relationship (Haskins & Clawson, 2015).

Executive education can be either credit bearing or non-credit bearing. This means that learners may seek credit or non-credit as part of a degree or certification program. By attending courses at their own pace and choice that may contribute to their career or personal advancement, individuals get accustomed to teamwork and cultivate their communication skills more since they return more mature to attend classes within higher education. This is how social psychology's courses could approach the graduate mature student in combination with innovative communication technologies.

International Cooperations and Web-Based Technologies in HEIs

Another innovative approach in higher education is the international branch campuses (IBC) that is an organizational model for higher education where borrowing and importing of practices from the cooperating institutions is the norm, creating a cluster model. IBC represents international collaboration and is associated with an institution that is located in another country from the one that initiated IBC and awards degrees in the host country through some physical presence there (Crist, 2015). New technologies substitute for and cover the physical absence by means of communication, gathering and transferring information, attending programs.

Around the Persian Gulf, United Arab Emirates have introduced the first IBC and supports the largest of them by far anywhere in the world based on official reports and studies presented by Crist (2015).

Individuals may cooperate with higher levels in the pyramid such as nations, firms, teams in order to be able to bring innovation and knowledge for the benefit of all (Andersson, Dasi, Mudambi, & Pedersen, 2015).

RECOMMENDATIONS FOR FURTHER RESEARCH

Flexibility is the key term that should be adopted and further searched in regard to higher education that cares for its people and the communities within which it operates, thus, creating a brand.

As HEIs embrace an increasing number of new technologies we will continue to witness change in terms of delivery, access and variety of knowledge.

In addition, the closer interaction of HEIs with industry provides an opportunity for practical applications and executive education.

The axes in higher education in regard to innovation should be threefold, students, teaching staff, implementation of inspiring policies. It is clear that the student community, educators and administrators need to embrace further the technological opportunities available. HEIs should adapt to the changing socio-economical changes and the student needs. This translates in identifying and using the best possible technologies in delivering quality and sustainable learning experiences.

Technology transfer is usually associated with copyright, patenting and licensing. There are opportunities for further collaborations among HEIs and the corporate world that could prove beneficial. Furthermore, University Technological parks and research centers could become entities that will bridge gaps between academic and corporate learning.

The growth of experiential learning options available to students, provides an option for further use of technology as well as the basis for developing professional skills. In an inter-connected world, distance learning offers educational and emotional experiences to learners. Executive and corporate education is another area that has benefited from the use of new technologies in the higher education industry. This sector has benefited from collaborations among HEIs and corporations for the benefit of learners.

Based on each person's unique abilities and competencies, the team within which the person is related to, will more easily develop competencies and skills. HEIs may well cater to develop such competencies, capabilities and skills that may be further employed by students in effective communication and teamwork and in social interaction.

CONCLUSION

Innovation in science and technology within higher education, may develop healthy competition with the innovative tools that it employs. Innovative tools consist of the promotional communication strategies of the institutions that are online advertised on their website and on social media. It also illustrates that the institution does not stay stable, it develops and goes with the current wave adopting new information and communication technologies. In that way, a virtuous circle is created since innovation attracts more attendants such as alumni, postgraduates, undergraduates, funding from big companies. Learning processes in physical environments can be harmonically used with online learning processes.

REFERENCES

- Adomßent, M., Fischer, D., Godemann, J., Herzig, C., Otte, I., Rieckmann, M., & Timm, J. (2014). Emerging areas in research on higher education for sustainable development—management education, sustainable consumption and perspectives from Central and Eastern Europe. *Journal of Cleaner Production*, 62, 1–7. doi:10.1016/j.jclepro.2013.09.045
- Andersson, U., Dasi, A., Mudambi, R., & Pedersen, T. (2015). (in press). Technology, innovation and knowledge: The importance of ideas and international connectivity. *Journal of World Business*. doi:10.1016/j.jwb.2015.08.017

- Asonitou, S. (2015). Employability Skills in Higher Education and the Case of Greece. *Procedia: Social and Behavioral Sciences*, 175, 283–290. doi:10.1016/j.sbspro.2015.01.1202
- Blackwell, A., Bowes, L., Harvey, L., Hesketh, A. J., & Knight, P. T. (2001). Transforming work experience in higher education. *British Educational Research Journal*, 27(3), 269–285. doi:10.1080/01411920120048304
- Bowen, W. (2013). *Higher Education in the Digital Age*. Princeton, NJ: Princeton University Press. doi:10.1515/9781400847204
- Bublitz, B., & Philipich, R. (2015). An example of the use of research methods and findings as an experiential learning exercise in an accounting theory course. *Journal of Instructional Pedagogies*, 16, 1–11.
- Bulfin, S., Pangrazio, L., & Selwyn, N. (2014). Making MOOCs: The construction of a new digital higher education within news media discourse. *The International Review of Research in Open and Distributed Learning*, 15(5), 290–305. doi:10.19173/irrodl.v15i5.1856
- Christ, J. (2015). Innovation in a Small State: Qatar and the IBC Cluster Model of Higher Education. *The Muslim World*, 105(1), 103–115.
- Cuenca, L., Fernández-Diego, M., Gordo, M., Ruiz, L., Alemany, M., & Ortiz, A. (2015). Measuring Competencies in Higher Education. The Case of Innovation Competence. In M. Peris-Ortiz & M. Merigó Lindahl (Eds.), *Sustainable Learning in Higher Education. Developing Competencies for the Global Marketplace* (pp. 131–142). Springer. doi:10.1007/978-3-319-10804-9_10
- Dunne, E., Bennett, N., & Carre, C. (2000). Skill Development in Higher Education and Employment. In F. Coffield (Ed.), *Differing visions of a learning society: Research Findings* (Vol. 1, pp. 105–138). The Policy press.
- Falk, J. (2014). *Lifelong Learning. Encyclopaedia of Science Education*. Springer.
- Franzoi, S. (2012). *Social Psychology* (6th ed.). Redding, CA: BVT.
- Georgia Tech. (2015). *Turn Ideas to Reality*. Retrieved December 20 December, 2015, from <http://cic.gatech.edu/>
- Godemann, J., Bebbington, J., Herzig, C., & Moon, J. (2014). Higher Education and Sustainable Development: Exploring possibilities for organisational change. *Accounting, Auditing & Accountability Journal*, 27(2), 218–233. doi:10.1108/AAAJ-12-2013-1553
- Gregory, M. S. J., & Lodge, J. M. (2015). Academic workload: The silent barrier to the implementation of technology-enhanced learning strategies in higher education. *Distance Education*, 1–21.
- Hanemann, U. (2015). Lifelong literacy: Some trends and issues in conceptualising and operationalising literacy from a lifelong learning perspective. *International Review of Education*, 61(3), 295–326. doi:10.1007/s11159-015-9490-0
- Haskins, M., & Clawson, J. (2015). Custom executive education program charters: A beneficial task and useful template. *Development and Learning in Organizations: An International Journal*, 29(1), 7–10.
- Head, A. J., Van Hoeck, M., & Garson, D. S. (2015). Lifelong learning in the digital age: A analysis of recent research on participation. *First Monday*, 20(2). doi:10.5210/fm.v20i2.5857
- Hiraldo, P. (2015). The role of critical race theory in higher education. *The Vermont Connection*, 31(1), 53–59.
- Hogg, M., & Vaughan, G. M. (2008). *Social Psychology* (5th ed.). London: Prentice Hall.

- Hoppenfeld, J., & Malafi, E. (2015). Engaging with entrepreneurs in academic and public libraries. *RSR. Reference Services Review*, 43(3), 379–399. doi:10.1108/RSR-02-2015-0011
- Hung, C. Y., Sun, J. C., & Yu, P. T. (2015). The benefits of a challenge: Student motivation and flow experience in tablet-PC-game-based learning. *Interactive Learning Environments*, 23(2), 172–190. doi:10.1080/10494820.2014.997248
- Kavoura, A., & Andersson, T. (2016). Applying Delphi Method for Strategic Design of Social Entrepreneurship. *Library Review*, 65(3), 185–205. doi:10.1108/LR-06-2015-0062
- Khanna, M., Jacob, I., & Yadav, N. (2014). Identifying and analyzing touchpoints for building a higher education brand. *Journal of Marketing for Higher Education*, 24(1), 122–143. doi:10.1080/08841241.2014.920460
- Knowles, M. S., Holton, E. F. III, & Swanson, R. A. (2014). *The adult learner: The definitive classic in adult education and human resource development*.: Routledge.
- Kolb, D. (2014). *Experiential Learning. Experiences as the source of learning and development* (2nd ed.). Pearson Education.
- Kruger, J., Kruger, D., & Suzuki, R. (2015). Assessing the effectiveness of experiential learning in a student run- free clinic. *Pedagogy in health promotion. The scholarship of teaching and learning*, 11, 1-4.
- Lackéus, M., & Middleton, K. (2015). Venture Creation Programs: bridging entrepreneurship education and technology transfer. *Education+ Training*, 57(1), 48-73.
- Link, A., Siegel, D., & Wright, M. (2015). *The Chicago Handbook of University Technology Transfer and Academic Entrepreneurship*. Chicago: The University of Chicago Press. doi:10.7208/chicago/9780226178486.001.0001
- López, L. S. (2015). An Analysis of the Integration of Service Learning in Undergraduate Spanish for Specific Purposes Programs in Higher Education in the United States. *Cuadernos de ALDEEU*, 28(1), 155–170.
- M. (2015). Beyond Academic Evidence: Innovative Uses of Technology Within e-Portfolios in a Doctor of Nursing Practice Program. *Journal of Professional Nursing*, 31(4), 284-289.
- Martin, G., & Pear, J. (2015). *Behavior Modification. What it is and how to do it* (10th ed.). New York: Routledge.
- Merriam, S. B., & Kee, Y. (2014). Promoting community wellbeing: The case for lifelong learning for older adults. *Adult Education Quarterly*, 64(2), 128–144. doi:10.1177/0741713613513633
- Mirriahi, N., Alonzo, D., McIntyre, S., Kligyte, G., & Fox, B. (2015). Blended learning innovations: Leadership and change in one Australian institution, *International Journal of Education and Development using Information and Communication. Technology (Elmsford, N.Y.)*, 11(1). Retrieved from <https://www.questia.com/library/journal/1P3-3688346831/blended-learning-innovations-leadership-and-change>
- Mowery, D., Ielson, R., Sampat, B., & Ziedonis, A. (2015). *Technology transfer, Ivory Tower and Industrial Innovation: University-Industry Technology Transfer Before and After the Bayh-Dole Act*. Stanford, CA: Stanford University Press.
- Muller, J. (2014). The future of knowledge and skills in science and technology higher education. *Higher Education*, 1–8.
- Narduzzi, J. L., & Campbell, J. D. (2015). The More Things Change: Reflections on the State of Marketing in Continuing Higher Education. In D. W. Shannon & R. Wiltenburg (Eds.), *Centennial Conversations: Essential Essays in Professional, Continuing, and Online Education* (pp. 309–317). Washington, DC: UPCEA.

- Neimeyer, G., & Taylor, M. (2014). 14 Ten Trends in Lifelong Learning and Continuing Professional Development. In W.B. Johnson & N.J. Kaslow (Eds.), *The Oxford handbook of education and training in professional psychology* (pp. 214-234). Oxford, UK: Oxford University Press.
- Nguyen, L., Barton, S. M., & Nguyen, L. T. (2015). iPads in higher education-Hype and hope. *British Journal of Educational Technology*, 46(1), 190–203. doi:10.1111/bjet.12137
- OKane, C., Mangematin, V., Geoghegan, W., & Fitzgerald, C. (2015). University technology transfer offices: The search for identity to build legitimacy. *Research Policy*, 44(2), 421–437. doi:10.1016/j.respol.2014.08.003
- Petrucchi, A., Kavena, P., Lebedeva, E., Feldman, L., Fried, G., & Vassiliou, M. (2015). You Have a Message! Social Networking as a Motivator for FLS Training. *Journal of Surgical Education*, 72(3), 542–548. doi:10.1016/j.jsurg.2014.10.007 PMID:25496795
- Rae, D., & Wang, C. (2015). *Entrepreneurial Learning: New Perspectives in Research, Education and Practice*. London: Routledge.
- Reinsmith-Jones, K., Kibbe, S., Traci Crayton, T., & Campbell, E. (2015). Use of Second Life in Social Work Education: Virtual World Experiences and Their Effect on Students. *Journal of Social Work Education*, 51(1), 90–108.
- Sanders, M., Clark, R., Davidson, B., & Jayaraman, S. (2015). GT Journey: The Importance of Accessible Rich Data Sources to Enable Innovation. In M. Kurosu (Ed.), *Human-Computer Interaction: Users and Contexts*, 9171 (pp. 82–91). Springer. doi:10.1007/978-3-319-21006-3_9
- Sessa, W. L. (2014). Targeting Millennials Social Media Strategies with Higher Education. *College and University*, 90(1), 2–11.
- Sevillano-García, M. L., & Vázquez-Cano, E. (2015). The Impact of Digital Mobile Devices in Higher Education. *Journal of Educational Technology & Society*, 18(1), 106–118.
- Siamagka, N.-T., & Christodoulides, G. (2016). Social Media in Higher Education: An Investigation into UG Marketing Education in the UK. In *Marketing Challenges in a Turbulent Business Environment. Developments in Marketing Science: Proceedings of the Academy of Marketing Science* (pp. 19-23). Springer. doi:10.1007/978-3-319-19428-8_6
- Siegel, D., & Wright, M. (2015). University Technology Transfer Offers, Licensing and Start-ups. In *The Chicago Handbook of University Technology Transfer and Academic Entrepreneurship* (pp. 1–40). Chicago: The University of Chicago Press. doi:10.7208/chicago/9780226178486.003.0001
- Stephens, S., & Margey, M. (2015). Action learning and executive education: Achieving credible personal, practitioner and organisational learning. *Action Learning Research and Practice*, 12(1), 37–51. doi:10.1080/14767333.2014.993592
- Subramanian, A., Choi, Y. R., Lee, S.-H., & Hang, C.-C. (2015). Linking technological and educational level diversities to innovation performance. *The Journal of Technology Transfer*, 1–23.
- Sulas, T., & Banyar, M. (2015). Innovative Marketing as a Tool for Building a Positive Image of an Institution of Higher Education and Increasing the Competitiveness of its Graduates – Analysis of the Functional Use of Projects of the Department of Marketing Communications at the Faculty of Multimedia Communications of TBU in Zlín and their Potential for Integration into the Teaching Process. *Procedia: Social and Behavioral Sciences*, 175, 146–153. doi:10.1016/j.sbspro.2015.01.1185

Vanevenhoven, J., & Drago, W. (2015). The structure and scope of entrepreneurship programs in higher education around the world. In D. Rae & C. Wang (Eds.), *Entrepreneurial Learning: New Perspectives in Research, Education and Practice*. London: Routledge.

Vick, T., Nagano, M. S., & Popadiuk, S. (2015). Information culture and its influences in knowledge creation: Evidence from university teams engaged in collaborative innovation projects. *International Journal of Information Management*, 35(3), 292–298. doi:10.1016/j.ijinfomgt.2015.01.010

Wafa, D. (2015). Arab Spring impact on executive education in Egypt. *International Journal of Educational Management*, 29(5), 596–608. doi:10.1108/IJEM-11-2012-0125

Waller, M., & Fawcett, S. (2015). Executive Education is an Implementation of Supply Chain Management in Collegiate Schools of Business. *Journal of Business Logistics*, 36(2), 135–138. doi:10.1111/jbl.12085

Wright, M. (2014). Academic entrepreneurship, technology transfer and society: Where next? *The Journal of Technology Transfer*, 39(3), 322–334. doi:10.1007/s10961-012-9286-3

Zhu, C. (2015). Organisational culture and technology-enhanced innovation in higher education. *Technology, Pedagogy and Education*, 24(1), 65–79. doi:10.1080/1475939X.2013.822414

Zolinas, R. (2015). *Life learning and mlearning*. Retrieved December 18, 2015, from <https://prezi.com/ji3tjyka4vhi/lifelong-learning-and-mlearning/>

ADDITIONAL READING

Altbach, P. (2014). International Higher Education: Vol. 1. *An Encyclopedia*. Routledge.

Avrilev, N., & Efimova, I. (2014). Global university rankings as indicators of the implementation of the integration process and competitive tool in the context of globalization of higher education. *Life Science Journal*, 11(10), 648–652.

KEY TERMS AND DEFINITIONS

Executive Education: Education that aims at mainly providing non-credit education. This can be tailored to the needs of people working in the business environment and/ or aim to acquire further knowledge on a specific topic of interest.

Experiential Learning: Learning that takes place through hands-on experience.

Innovation in Higher Education: Use of innovative tools and methods by academic institutions.

New Technologies as a Marketing Tool: Innovative web based devices that can promote products and services within a specific framework.

Social Skills: Competencies to interact successfully with other people. Innovative communication technologies may contribute to this successful interaction.

Teamwork: Cooperation among a group of people by employing skills and competencies in order to accomplish a specific goal.

Technology Transfer: The process of transferring skills, technologies, competencies that have been created based on research and can be further utilized to create products and services for the use of many people.

Experiences of Implementing a Large-Scale Blended, Flipped Learning Project



Hazel Owen

Ethos Consultancy NZ, New Zealand

Nicola Dunham

Massey University, New Zealand

INTRODUCTION

Many schools, tertiary institutions, and other organisations are involved in the design of eLearning experiences, but, it is questionable whether the investment results in more engaged, knowledgeable, skilled learners (Poulova, & Simonova, 2015). Two key prevalences that influence effectiveness have been identified. The first is a set of beliefs, often unquestioned (Byham, 2007), that tend to shape overall expectations of what eLearning experiences might comprise. The second is a tendency to embark on large-scale, ‘monolithic’ eLearning developments (JISC, 2005). The term monolithic, in the context of software and eLearning, refers to developments that do not have separate components and are part of the same architecture (Tanenbaum, 2014). It is sometimes used as a pejorative term, referring to the fact that it is slow, if not impossible, to change anything after rollout without starting over again. Therefore, monolithic deliverables in education are problematic, in part because there is an up-front cost before any learning value is realised.

In this chapter, experiences of implementing a large blended-learning project at a tertiary institution in Aotearoa New Zealand are discussed. Tertiary education in Aotearoa New Zealand covers all post-secondary education (from certificate to PhD level) and is akin to the term Higher Education in other countries. The project (instigated at the beginning of 2012) was based on a phased rollout, with each subsequent

stage being informed by the ones before. Steps included needs analysis, design, development, prototyping, refinement, and deployment. The discussion focuses on personal reflections on the two initial phases of the project from three different perspectives. The overall aim is to share contextualised experiences, to add to the knowledge base on blended learning, and to provide some general, practical recommendations.

BACKGROUND

This section provides an overview of the context and scope of the project. The project was driven by initiatives to improve student learning experiences; enhance interdisciplinary education; introduce a common semester; and develop a new Masters level course and a Professional Doctorate (Owen, & Dunham, 2015). One caveat was that technology itself would do nothing to enhance learning and teaching; as such, sound pedagogical theory and eLearning principles (Weidert, 2012) needed to be driving forces. It was also recognised that the change would impact professional identity, and therefore require flexible management and responsive processes.

The three participants in the study (two of whom are also the authors of this chapter) worked closely together on the project, and have been involved for many years with facilitation, teacher professional development (PD), and curriculum design. One has been working within the eLearn-

ing field for almost 15 years. As participants they undertook key roles, and were able to provide perspectives on, and insights into, motivations, challenges, and PD needs.

The remit was to design, develop and implement a blended, flipped approach to interdisciplinary learning. As such, the project team worked with subject matter experts, writers, and curriculum editors to redesign curricula. The design included discovery-orientated tasks that learners were encouraged to engage with prior to attending regular facilitated synchronous sessions (either face-to-face or in a webinar). Design and facilitation were flipped to focus on the learners and encourage them to find and create their own resources, as well as engage with, guide and 'teach' their peers. Prior learning could be extended by facilitators and peers - especially important in courses where students may already be working and have years of experience. Interdisciplinary learning was integrated by encouraging learners to explore from multiple practice and discipline perspectives.

As part of the change process a new centre was established in May 2014. The centre ensured a reasonable amount of autonomy, so that while the 'what' was directed by the executive team, the 'how' was more for the project team to decide. As such, they were able to, for example, trial an agile collaborative approach to writing and development. The course writing and design process, built around iterative cycles, shifted from being the sole responsibility of individual lecturers to become a team-based, collaborative and transparent approach at all levels: from writing and design, to the facilitation of the courses (Figure 1).

LITERATURE REVIEW

Two aspects are key to the design approach and implementation, and these are explored before moving into a more detailed discussion of the associated research and results.

Iterative Approach and Agility

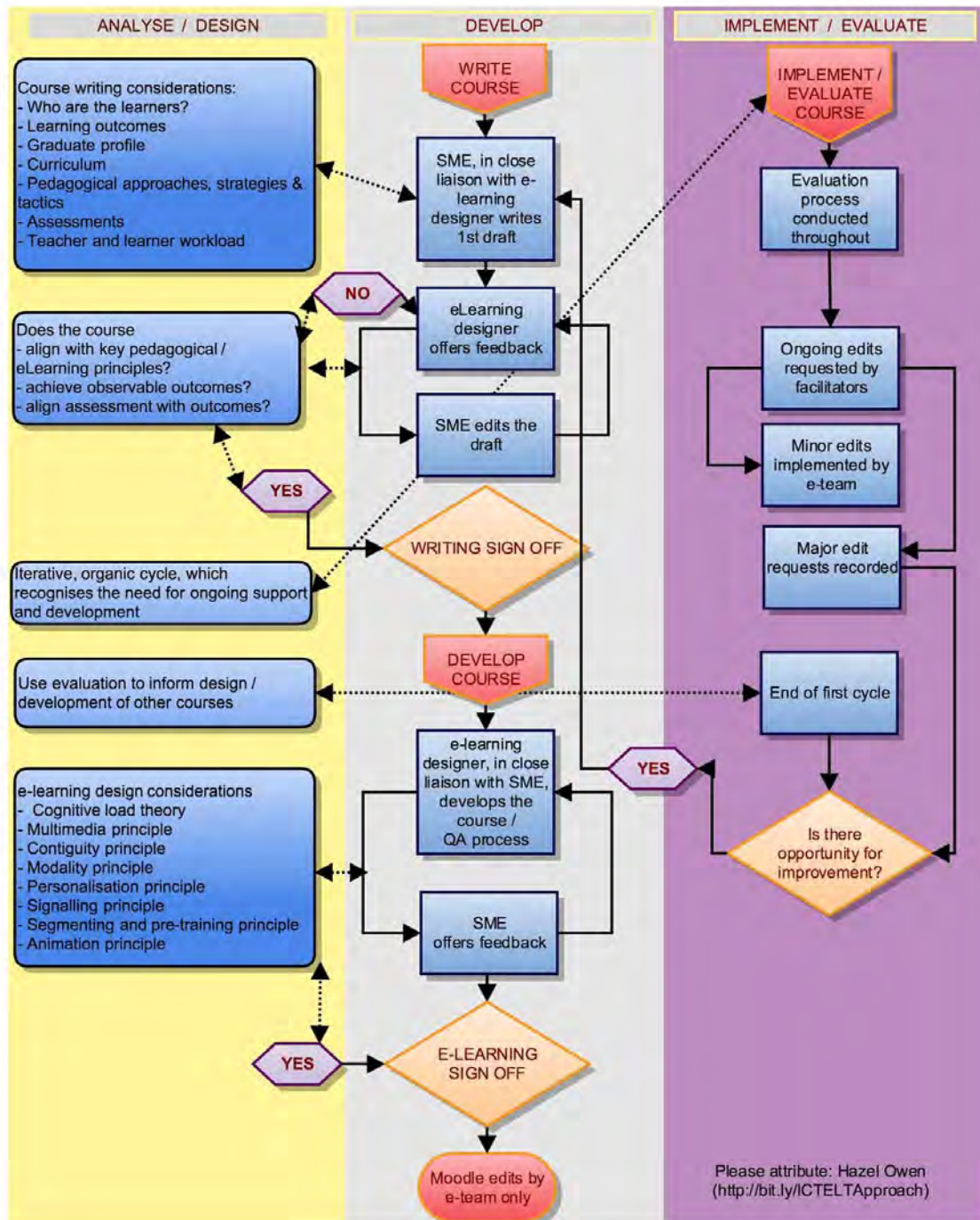
Russell (2006, para. 16) advises that "because e-learning projects are... chaotic, it is usually best to implement a project management approach that is built for chaos." A release-based, iterative approach can address some issues that arise from chaos, especially when underpinned by an agile framework. In the context of the project referred to in this chapter, a brief comparison of the iterative approach with an alternative monolithic approach, illustrates some of the benefits (Russell, 2006):

- The iterative approach enabled a project plan to be developed for each two- to three-month phase, compared with a monolithic approach where a project plan would have been necessary for the whole initiative, which could extend over several years.
- The project started small, and had increased transparency and resiliency (Haikin, 2013).
- Each iteration was designed, developed, tested, changed in response to feedback, and rolled-out to learners. In contrast, a monolithic approach would have required the design, development, testing, and roll-out of the whole thing to learners. The iterative approach enabled learning activities and experiences to be evaluated (Hudson, 2012) and changed as needed.
- The institution was able to undertake iterations for as long as either 1) budgeted and signed off for; or 2) built into the ongoing design of programmes such that courses remained current.
- There was an ability to respond to issues as they arose and this saved "resources and, ultimately, help[ed] deliver a successful project on time and within budget" (Rouse, 2011, para. 2).

Identity and Openness to Change

In simplistic terms, there are two types of change in education: cyclical and transformative. Cyclical

Figure 1. The writing, design and development process



change is ongoing, and institutions initiate it or respond to it as it happens, occasionally returning to the stage they were at previously (Toynbee, 1947). In contrast, transformative change is ho-

listic, disruptive, often uncomfortable, happens at all levels, and is irreversible (Wallace, 1970). In other words, once an institution has undergone transformative change it will not be able to return

to the way it was before. Transformative change, however, requires the people and culture within an institution to work with ambiguity, as well as to question their beliefs, professional identity and practice. Not everyone will want to be part of this type of change, especially when it re-frames existing notions of formal learning. As such, there can be an uneasy dichotomy where the current culture of an institution sits alongside innovations happening within it.

Mindset theory has been associated with organisational culture and change, with a fixed mindset being associated with a culture of genius, and a growth mindset being associated with a culture of development (Murphy & Dweck, 2010). A genius culture endorses the belief that intelligence is fixed, while a development-culture views intelligence as malleable and expandable. When a person possesses a fixed mindset, they:

- Tend to avoid challenges,
- Deflect feedback,
- Withdraw when they meet obstacles,
- See little point in putting in extra effort, and
- Are threatened by other people's success (Dweck, 2012).

In contrast, people who possess a growth mindset tend:

- To see failure as part of the process (Dweck, 2000),
- Actively seek challenges, perceiving them as something that, if they work hard work to overcome them, will help them to grow intellectually (Dweck, 2012),
- See feedback as something that will help them develop, and celebrate the success of other people.

However, Dweck's research suggests that people can, with support, shift from a fixed to a growth mindset (Dweck, 2012).

Transformative change can be associated with learning goals and growth mindsets, whereas a genius culture may value performance goals over learning goals (Murphy & Dweck, 2010). Cyclical change, with the risk of returning to previous ways of doing things, aligns with a culture of genius in which maintaining standards may impede the undertaking of challenges. These concepts have implications for the type of support offered to people within a project, as well as recognition that some practitioners may not be change-ready, and will need time to move into a space where they feel comfortable and confident with taking risks.

METHOD

This research draws on principles of phenomenology to explore 'lived experience' from the perspectives of people involved (Welman & Kruger, 1999), through the use of reflective accounts. Reflective accounts are an established means of conducting practitioner-focused research within the field of education (Campbell, McNamara & Gilroy, 2004), and are a means by which lived experiences can be attributed meaning (Schutz, 1997). The researchers do not claim that they are in the possession of 'truth' (Griffiths, 2009), but rather they are approaching this experience of knowing as one of potential transformation within the context of organisational change.

The three participants in the project, from a stance of identity and integrity (Palmer, 2007), incorporated the principle of personal situated inquiry into writing reflective accounts about their experiences (Samaras, 2011). The researchers then carried out an independent cross-analysis of each other's reflections. A synthesis of codes and themes from the three reflections were collated and form the heart of this chapter. This approach of independent and cross-analysis is in keeping with aforementioned principles of personal situated inquiry and is a transparent and systematic research process (Samaras, 2011). To preserve

some semblance of anonymity, quotes used in the discussion of the analysed reflections are attributed to participant 1, 2, or 3.

DISCUSSION OF EMERGENT THEMES

The following section is organised around themes emerging from analysis of the reflective accounts, with verbatim quotes incorporated into the discussion. Themes include:

- Working with dilemmas and ambiguity,
- Shifting identities and practices, and
- Iterative and agile approaches.

Working with Dilemmas and Ambiguity

The need for strong support, decision-making and leadership to appropriately and consistently manage dilemmas plays a significant part in an effective change process. Within this project, however, Participant 2 identified “an uneasy dichotomy between the support... from senior management, and the sense of urgency to see a return on investment,” resulting in the participant feeling “torn between ensuring a really rigorous quality assurance and piloting process, and meeting the expectations to deliver.” Participant 3 also observed that “if a new teaching and learning paradigm is introduced at a time when an organisation is undergoing major financial restructuring ... the potential... benefits to learners are in danger of being overshadowed.” Indeed one source of ambiguity was that broader institutional interests could both support, and undermine, the project.

It was also realised that current institutional systems and structures were not compatible with the new approaches, creating issues with day-to-day logistics. For instance, there were “gaps in existing orientation systems ... [which] left both students and the teaching team exposed”

(Participant 1). These concerns led to the team being mindful about ensuring the ideology and philosophy behind developing the blended, flipped approach was not lost in the everyday and the administrative.

While institutional support for the project was evidenced through ongoing resourcing there was a “gap in understanding between how much resource ... [had been originally] estimated for an initiative of this scope, and how much was actually required” (Participant 2). This resulted in a “sense of a job done not quite well enough; of not quite enough support offered,” which resulted in writers “writing under incredible pressure, of curriculum editors who [wanted] ... to develop professional relationships further, and of eLearning designers and builders working flat out” (Participant 2). Furthermore, Participant 3 identified that “staffing was not conducive to the level of support the literature was telling me would ensure the best outcome for student experience and success.” The institution responded, but slowly, not having the nimbleness to change processes and support in response to rapidly emerging requirements.

Time was seen as both a positive and a negative. At a project-management level, complex interdependencies, including resourcing, constantly impacted the iterative process. For example, an iteration would be timeboxed, but then because “the process of recruiting and seconding writers was painfully slow... timelines ... slipped, and then slipped again” (Participant 2). The three participants, also, were eager to meet the demands of the project, but needed to balance their personal lives and ongoing professional aspirations (two, for example, were studying for their doctorates). On the other hand, the iterative approach provided time for all project staff to understand writing for the blended, flipped environment, and to acclimatise to the high degree of collaboration and transparency. Also, the courses could be designed and written such that they were ‘good enough’, knowing that improvements could be made in future iterations, informed by feedback and experience.

Shifting Identities and Practices

On a practical level the attitudes and professional practices of the educators involved needed to shift (Bonk & Cummings, 1998). The underpinning principles and approaches used in the project had led to re-framing roles for lecturers (to collaborative writers and facilitators), eLearning designers (to mentors and guides), and students (to active, self-defining learners). The writing approach, for instance, required what Participant 1 described as “a meeting of minds.” For this to work well, there was a “need for a team to be built not only on skills and experience, but also on relationships, complementary... beliefs, and an openness to true collaboration” (Participant 2). No longer were academics working in isolation; rather teams of writers and developers liaised closely with multiple stakeholders - “a less than usual practice within academic course development” (Participant 3). However, although providing benefits, the approach did create, for some project staff, challenges around identity - especially when roles were ill-defined.

Project team members worked hard to ensure effective professional development support, and to recognise opportunities. Participant 2, for instance, explained that writers were immersed in the process “while being mentored and guided by a Curriculum Editor.” Writers and facilitators supported each other in a more collegial and ‘open’ space, in ways that were responsive to the evolving nature of the project. However, “there were some writers... who felt quite vulnerable and found the collaborative writing process threatening” (Participant 2). Further complications became apparent when content was the focus of an existing course, whereby the writer often tended toward “giving students ... information (knowledge) as opposed to ... engaging them more with the process of learning” (Participant 1). As such, the the whole team had to strive to keep “co-constructivism at the heart of what we are designing and not slipping back into didactic habits” (Participant

1). Part of the issue was that the re-framed roles challenged existing understandings of what an academic practitioner is and does. Participant 3 explained that “On the one hand I wanted to teach responsively and reach out to Māori students to engage them in relation-based, face-to-face learning On the other hand I had an obligation to the methodological approach... being trialled,” and found herself “struggling to ... know the new terrain being created, as a teacher.”

Working collaboratively was defined not only by skills and knowledge, but relationships that supported educators to embrace change. Exposing raw thinking to peers as part of the writing process was enlightening, inspiring and unnerving. For some, for instance, it encouraged them to question existing perceptions of self as ‘knower’ of content and pedagogy, and explore their own knowledge and mindsets. In addition, working collaboratively and transparently meant that the academic silos could not continue to function.

Considering these factors in light of Dweck’s mindsets:

- Participant 1 identified herself as “a bit of a change-magnet” who was ready to “be part of something bigger.”
- Participant 2 felt “intrigued” by the proposed changes.
- Participant 3 was more ambivalent and saw the project as something where “the risk is unknown and the stakes are quite high.”

While all three acknowledged multiple challenges, it appeared that their willingness to take risks, along with a well-developed resilience, and a creative attitude to problem-solving (all features of a growth mindset), helped them relish change to the point where Participant 1 shared: “I was really feeling a sense of flying, of freshness and a chance to give things a go.” Without having growth mindset dispositions it is unlikely that these participants would have become, and stayed, involved in the project.

Iterative and Agile Approaches

The iterative approach had both benefits and drawbacks. One of the major benefits was that, in Phase 1, only initial requirements were required. These were used to specify, design and develop a prototype course, which stakeholders evaluated, and their feedback was applied to the other three initial courses required in Phase 1. In turn, findings from the roll-out of the four courses to learners helped the team identify and avoid the replication of errors; something that proved “invaluable... as we moved into Phase 2 and 3” (Participant 2). On a more micro-scale, course writers found the opportunity to edit a course multiple times was positive, although limited experience made it difficult for some to know what ‘good enough’ was and when to stop, thereby creating slippage. Such benefits were constrained by a wider context. As alluded to earlier, the project centre was “struggling within the larger overall organisation’s existing structures and formal procedures” (Participant 2). The structures and procedures were essential for the overall institution, but sometimes undermined “the dynamic capabilities required for an iterative cycle of rapid innovation” (Participant 2).

In spite of the challenges, toward the second-half of semester 1, 2014, it was observed that “students...were showing signs of engaging and succeeding. The final ... triumph ... [was] our innovative integrated assessment. At last we could see the learning that had occurred and just what the students could achieve” (Participant 1). In addition, Participant 2 felt “that we were developing courses that [were] ... accessible,... culturally responsive, [and]... ‘look and feel’ different to the usual blended learning experience.”

In relation to the project itself the benefits have been numerous; the three participants appeared to have sufficient confidence to embrace ambiguity, and to continue to draw on strategies to address challenges. However, the long-term impact on project personnel remains a concern and questions have been raised about sustainability: while “there

is still a great enthusiasm for the project ... there are a lot of tired faces” (Participant 2).



SOLUTIONS AND RECOMMENDATIONS

In this section key learnings have been identified and offered as recommendations for organisations implementing eLearning projects, which include:

- Roles and responsibilities,
- Project management,
- Ownership and shared understandings, and
- Review processes.

Roles and Responsibilities

When involved in a period of change, a project using an iterative approach provides time for practitioners - and departments - to shift and develop their identity, as well as to build liaisons and shared understandings. A complementary aspect is the opportunity for a project team to identify risks around roles and responsibilities early and put in place robust mitigations. Associated considerations are:

- Agreements are made around roles and responsibilities within the wider institution.
- There is widespread understanding that innovative projects include ‘doing things differently’.
- Institutional processes are unlikely to be responsive enough to keep up with rapid change.
- Every project team will have members with different levels of expertise and experience; parity of input and output is important for a well-functioning team (Owen & Allardice, 2014).

Innovation does not happen without tension and indeed tension provides some of the energy that

drives innovation. In part it is fostered by critical engagement which, by its very nature, causes dissonance. So, being innovative is not about removing the tensions, but about being aware and mindful of them, thus creating an environment where people can thrive and grow without being overwhelmed or over-stressed - which can lead to a reversion to original ways of being and doing.

Project Management

Project management is an essential aspect of an iterative, large-scale eLearning project - and something that is sometimes overlooked or not appropriately resourced. A project manager will, amongst other things:

- Organise, manage and allocate resources (including writers / subject matter experts, curriculum editors, and eLearning designer / developers),
- Set up timelines for development, and timebox each iteration,
- Help an institution complete an iteration within the agreed scope, while also meeting quality requirements,
- Identify risks and issues, as well as possible mitigations,
- Provide for PD, time release, and other incentives,
- Analyse results of evaluations to inform and modify a project on an ongoing basis, and
- Communicate progress regularly to all stakeholders.

‘Ownership’ and Shared Understandings

Ownership is achieved in supportive groups where there is a good level of consensus around underpinning principles and approaches. Where shared understandings are not present a coherent approach to project implementation is not possible. Construction of meaning may “result

from individual critical reflection but ideas are generated and knowledge constructed through the collaborative and confirmatory process of sustained dialogue within a critical community” (Garrison & Archer, 2000, p. 91). As such, it is recommended that project participants should:

- Be actively involved from the inception of a project,
- Have input into choosing underpinning educational philosophies,
- Identify their own professional needs, as well as those of their learners (Yoo, Huang, & Wenhao, 2013),
- Contribute to the design and prototype/pilot, and
- Take time and space to develop professionally until they are ready to be involved.

Review Processes

Finally, it is recommended that a longitudinal review study be conducted during the implementation of an eLearning project (Owen & Allardice, 2014). The review needs to be systematic, iterative, meaningful and non-intrusive. Results can inform following iterations around effective practices, recognise changing needs and technologies, and evaluate the efficacy of the approach. Comparative studies of effective practices and academic research are also desirable. Students must feel a sense of ownership, and this can be encouraged through fostering a tight-knit learning community that actively seeks and applies their feedback (Owen & Allardice, 2014).

FUTURE RESEARCH DIRECTIONS

The flipped blended model ensured that students were actively involved in their learning as co-constructors of knowledge, as opposed to passive recipients. Early results indicate that students were engaged, aware of metacognitive strategies they were employing, and were enjoying the flexibility

the blended, flipped design enabled. However, as yet the evidence is mainly anecdotal, and student success and retention will be the focus of a future research study.

CONCLUSION

This chapter is not about an end point; rather it is a snapshot of initial steps. The aim over the next phases is to continue to “emphasize the little things” (Flyvbjerg, 2001, p. 133), thereby providing an ongoing account of the minutiae of day-to-day undertakings that are part of implementing a flipped, blended, interdisciplinary initiative in one “remote corner of the universe” (Flyvbjerg, 1998, p. 1). It is hoped that the level of detail might help identify nuances that may otherwise be missed, and yet that can fundamentally influence the success, or otherwise, of an initiative.

Shifting mindsets has been a major part of the experience for the project participants. The highly collaborative and iterative approach enabled them to challenge their perspectives of intelligence - both their own and their prospective learners, as they worked through the process of developing and facilitating a design where active learning experiences were the driver, and content took a back seat. The change process required a willingness and ability to move out of predefined roles into a dynamic environment where previous ways of working were overridden, and highly responsive alternatives adopted.

The shift to a collaborative, transparent, agile approach to eLearning development has not been simple. The transformative change experienced was uncomfortable, disruptive, happened at all levels, was irreversible, and caused dissonance. The authors suggested that innovation does not happen without tension and indeed it was tension that provided some of the energy that drove this innovation. So, being innovative was not about removing tensions, but about being aware and mindful, while providing an environment that

encouraged people to move from fixed to growth mindsets. It was about thriving and surviving; about feeling inspired, supported, and valued.

REFERENCES

- Bonk, C., & Cummings, J. (1998). A dozen recommendations for placing the student at the center of Web-based learning. *Educational Media International*, 35(2), 82–89. doi:10.1080/0952398980350205
- Byham, W. (2007). *70: The new 50*. Bridgeville, PA: Development Dimensions International Press.
- Dweck, C. S. (2000). *Self-Theories: Their role in motivation, personality, and development*. Hove, UK: Taylor & Francis Ltd.
- Dweck, C. S. (2006). *Mindset: The new psychology of success*. New York, NY: Random House.
- Dweck, C. S. (2012). *Mindset*. London: Robinson.
- Flyvbjerg, B. (1998). *Rationality and power: Democracy in practice*. Chicago: University of Chicago Press.
- Flyvbjerg, B. (2001). *Making social science matter why social inquiry fails and how it can succeed again*. Oxford, UK: Cambridge University Press. doi:10.1017/CBO9780511810503
- Garrison, D., & Archer, W. (2000). *A transactional perspective on teaching-learning: A framework for adult and higher education*. Oxford, UK: Pergamon.
- Griffiths, M. (2009). *Using personal narrative and other stories in educational research: Issues of validity and truthfulness*. Paper presented at the First International Conference on Educational Research for Development, College of Education. Retrieved from http://www.research.ed.ac.uk/portal/files/15324289/Using_personal_narrative_and_other_stories_in_educational_research.pdf

- Haikin, M. (2013). *Reflections on applying iterative and incremental software development methodologies (Agile, RAD etc.) to aid and development work in developing countries*. Retrieved from <https://matthaikin.files.wordpress.com/2013/03/agile-blarticle-part-11.pdf>
- Hudson, T. (2012). *Create a “Lean” learning environment*. Retrieved from <http://www.clomedia.com/articles/how-to-create-a-lean-learning-environment>
- Joint Information Services Committee. (2005). *JISC infoNet - SFC e-Learning toolkit*. Retrieved from <http://tools.jiscinfonet.ac.uk/downloads/implementing-elearning/implementing-elearning.pdf>
- Owen, H., & Allardice, R. (2014). Managing the implementation of blended E-learning initiatives with the unconverted in a climate of institutionally driven change. In J. H. Westover & J. P. Westover (Eds.), *Engaging Hybrid and Blended Learning in Higher Education* (pp. 50–68). Champaign, IL: Common Ground Publishing LLC.
- Owen, H., & Dunham, N. (2015). Reflections on the use of Iterative, agile and collaborative approaches for blended flipped learning development. *Education Sciences*, 5(2), 85–103. doi:10.3390/educsci5020085
- Palmer, P. J. (2007). *The courage to teach: Exploring the inner landscape of a teacher's life* (10th ann. ed.). San Francisco, CA: Jossey-Bass.
- Poulova, P., & Simonova, I. (2015). *Blended learning in practice: Tracking students' performance*. Paper presented at the European Conference on e-Learning. Retrieved from <http://search.proquest.com/openview/ec1d7298600f0b6b1b4c19db5382868d/1?pq-origsite=gscholar>
- Rouse, M. (2011). *Agile project management is an iterative approach to planning and guiding project processes*. Retrieved from <http://searchcio.techtarget.com/definition/Agile-project-management>
- Russell, L. (2006). *Project management and e-learning: More is worse*. Retrieved from http://www.astd.org/LC/2006/0806_russell.htm
- Samaras, A. P. (2011). *Self-study teacher research: Improving your practice through collaborative inquiry*. Thousand Oaks, CA: SAGE Publications.
- Tanenbaum, A. (2014). *Modern operating systems* (4th ed.). Prentice Hall.
- Toynbee, A. (1947). *A study of history: Abridgement of volumes I to VI*. Oxford, UK: Oxford University Press.
- Wallace, F. (1970). *Culture and personality*. New York, NY: Random House.
- Weidert, J. (2012). *The mobile college community: A study of adult learners' adoption and use of digital technologies on the campuses of Florida's community colleges* (Doctorate of Education). Santa Barbara, CA: Fielding Graduate University. Retrieved from <http://gradworks.umi.com/34/94/3494860.html>
- Welman, J., & Kruger, S. (1999). *Research methodology for the business and administrative sciences*. Johannesburg, South Africa: International Thompson.
- Yoo, S. J., Huang, W. D., & Wenhao, D. (2013). Engaging online adult learners in higher education: Motivational factors impacted by gender, age, and prior experiences. *Journal of Continuing Higher Education*, 61(3), 151–164. doi:10.1080/07377363.2013.836823

ADDITIONAL READING

- Armitage, S., & O'Leary, R. (2003). *A guide for learning technologists*. York, UK: Learning and Teaching Support Network.

Collis, B., & Van der Wende, M. (2002). *Models of Technology and Change in Higher Education: An international comparative survey on the current and future use of ICT in Higher Education*. Enschede: University of Twente.

Salmon, G. (2002). *E-tivities. The key to active online learning*. London, UK: Kogan Page.

Ziskovsky, B., & Ziskovsky, J. (2007). Doing more with less - Going lean in education. Retrieved from <http://www.leaneducation.com/whitepaper-DoingMoreWithLess.pdf>

KEY TERMS AND DEFINITIONS

Agile: A flexible approach, where the focus is small-scale but functioning releases of a course / product. These releases are evaluated, and improvements integrated.

Blended Learning: Learning facilitated through a variety of modes of delivery, styles of learning, and transparent communication (on- and off-campus).

Change Management: Approaches, strategies and frameworks used to support the people-side of change to achieve an institution's required outcomes.

Flipped Learning: Shifting the focus of learning from the facilitator to the learner, partly by reversing what happens inside and outside of a formal learning session.

Iterative: Approach for arriving at desired result by repeating a cycle of processes.

Mindset: Set of attitudes held by someone, which in the context of this chapter relate specifically to intelligence and learning.

Project Management: Methodical approach to planning and guiding a project.

A Flipped Learning Approach to University EFL Courses

Yasushige Ishikawa

Kyoto University of Foreign Studies, Japan

Reiko Akahane-Yamada

Advanced Telecommunications Research Institute International (ATR), Japan

Craig Smith

Kyoto University of Foreign Studies, Japan

Masayuki Murakami

Kyoto University of Foreign Studies, Japan

Mutsumi Kondo

Kyoto University of Foreign Studies, Japan

Misato Kitamura

Advanced Telecommunications Research Institute International (ATR), Japan

Yasushi Tsubota

Kyoto Institute of Technology, Japan

Masatake Dantsuji

Kyoto University, Japan

BACKGROUND

Defining Flipped Learning

BL is a combination of face-to-face delivery and online delivery of learning materials and activities (Osguthorpe & Graham, 2003). Teachers interested in BL are searching for ways to make use of the rapidly expanding number of online easily-accessible learning resources. The increase in the use of technology to connect learning environments inside and outside the classroom has recently accelerated due to two developments in educational resources: the free online access to university courses via software, e.g. iTunes U, and websites such as Coursera (<https://www.coursera.org/>); and the sophisticated communication capability of mobile devices such as smart phones and tablet computers.

A promising response to these developments is the FL approach to the BL teaching methodology (Stuntz, 2013; Bishop & Verleger, 2013) which reverses the conventional patterns of classroom learning. Hamdan, McKnight, McKnight, & Arfstrom (2013) define the differences between FL practices and distance learning and BL courses by explaining that if the use of computers and online content does not alter conventional patterns of direct instruction in teacher-centered classrooms, it is not FL. In FL courses students are provided with outside-of-class online learning materials conventionally presented in class by the teacher. Classroom time is used for students to seek advice from the teacher and to help each other as they complete tasks which are usually done as outside-of-class assignments (Lage, Platt, & Treglia, 2000).

DOI: 10.4018/978-1-5225-2255-3.ch334

Yarbro, Arfstrom, McKnight, & McKnight (2014) define FL as “a pedagogical approach in which direct instruction moves from the group learning space to the individual learning space, and the resulting group space is transformed into a dynamic, interactive learning environment where the educator guides students as they apply concepts and engage creatively in the subject matter” (p.5). FL facilitates active collaborative learning during class time by allowing teachers to respond to individual differences in the comprehension of course content. At the same time students are given opportunities to find learning methods and materials that suit their own learning styles (Lage, Platt, & Treglia, 2000) through engagement in project-based learning activities which include small-group discussion and problem-solving activities. Thus, FL has the greatest chance of success with small-sized classes that make peer interaction manageable and allow teachers to take on a coaching role.

The rationale of FL, the expectations for student participation in their own learning, and the role of the teacher should be explained and demonstrated to students in the early stages of a course.

Flipped Learning for Foreign Language Learning

Recent studies offer encouragement that an FL approach to language teaching should be further investigated through a classroom-based action research methodology. Stuntz (2013) reported that students in a FL CALL EFL course needed instruction and practice in the use of communication and study media such as Gmail and Google Docs to complete outside-of-class assignments. Improvements in these skills allowed for effective use of class time to discuss outside-of-class learning tasks with both their peers and the instructor. The collaboration resulted in higher quality task products. A Learning Management System platform can guide students through the outside-of-class online learning tasks (Sung, 2015). Student satisfaction can enhance motivation when FL

course learning task products are shared with the class and with members of a broader community via YouTube® (Leis, Cooke, & Tohei, 2015). Interactive communication with an international community can be achieved by engaging EFL students in video conferences with students in other countries (Kuhn & Hoffstaedter, 2015). Access to authentic language-use opportunities of this nature in FL courses may decrease in-class performance anxiety among students (Egbert, Herman, & Chang, 2014).

As teachers and students learn how to apply the on-going developments in educational infrastructure, they will discover effective combinations of what can be done best outside, and inside, of classrooms. The potential of FL rests on an assumption that students will use the learning materials before class, so that classroom time can be devoted to problem-solving tasks and analytical examinations of the learning materials. It is crucially important to sustain student engagement in outside-of-class online learning tasks.

Sustaining Student Engagement in Flipped Learning

As with any innovation in education initial mistrust must be overcome. In FL the doubts center on the online element of learning, especially if the outside-of-class work-load increases, and if the students believe that learning is best in face-to-face lecture courses. Moreover, online learning components can be perceived as risky by parents, teachers, and learners if there appears to be no immediate and significant improvement in achievement (Hess, 2006; Dixon, Osment, & Panke, 2009). Parsons (2011) reports that a feeling of being overwhelmed at having to adapt to a new learning environment and discover new roles without proper support can result in a poor understanding of FL course purposes. Thus, students may not stay involved long enough in the outside-of-class course components until a critical mass of satisfaction and achievement is reached.

Although Internet communication is a ubiquitous feature of the lives of today's learners, it does not mean that they can employ it successfully for learning. A loss of motivation due to unfamiliarity with their learning tools can be compounded when students feel overwhelmed cognitively by the array of resources (Barker, 2006). Vaughan (2007) found that students who felt overwhelmed by the technology and by an overabundance of resources did not develop such successful self-regulated learning (SRL) skills. Thus, if learners do not already have SRL skills, there may be a consequent early loss in motivation soon after an FL course begins. Although there is some conflicting evidence whether FL can foster the development of SRL skills (Cho, Demei, & Laffey, 2010; Ishikawa et al., 2015), SRL competence may be a necessary prerequisite for success in FL courses, as is the case in Distance Learning (Andrade & Bunker, 2009; Zhao & Johnson, 2012). Furthermore, Ng, Seeshing Yeung, & Yuk Hung Hon (2006) found that in the case of foreign language learners, those with a higher level of language proficiency, and confidence, could handle unfamiliar learning styles better than lower level learners. In addition, lower level learners may not enjoy the multiple FL interaction channels, and so, feedback may be less effective.

Hartnett, George, & Dron (2010) found that learners felt that online interaction for collaborative work was less productive than meeting face-to-face in terms of saving time and receiving more effective feedback. Online mentoring may be a solution. Azevedo, Moos, Greene, Winters, & Cromley (2008) found that in a hyper-media environment, students demonstrated better performance when they received their goals from a teacher rather than from online learning tools. Thus, a combination of e-mentoring and in-class advice from other students and a teacher may be important consideration in FL course design.

The Transformation to Flipped Learning

Conventional in-class EFL methodology typically consists of three phases: input, intake, and output activities. Students are exposed to comprehensive input of an English text in order that they may have an opportunity to learn about a language feature that is new to them. If they understand the new feature of the language and it becomes part of their language knowledge, intake has occurred (Ishikawa, Kondo, & Smith, 2010). After engaging in learning tasks which are intended to allow intake to occur, students conduct output activities to practice using the new language feature. At the end of the class, the teacher usually assigns homework as an individual outside-of-class learning task to reinforce the in-class learning. The grading of the homework assignment by the teacher confirms whether the intended teaching aims have been achieved.

In an FL course, as the input component, students work on learning tasks online outside of class. The teacher supports students' learning by e-mentoring communication, such as by the sending of timely needs-based messages to the students via social networking systems. E-mentoring can be defined as similar to traditional mentoring but accomplished through online communication.

In class, after the students reflect on their outside-of-class activities, they collaborate in small groups on intake and output learning tasks which the teacher facilitates within a collaborative learning environment. At the end of the class the students reflect on what they did in class. E-mentoring may establish a computer mediated, mutually beneficial, relationship between a mentor and a learner when the mentor advises and encourages the learner by modeling effective learning behavior in ways that are boundary-less, egalitarian, and qualitatively different from traditional face-to-face mentoring (Bierema & Merriam, 2002). E-mentoring can be used to support

SRL skill development by developing a strong sense of self-efficacy, and by building relationships with peers and teachers that contribute to effective learning (Norton, 2005; Chang, 2004; McKenzie & Ozkan, 2006).

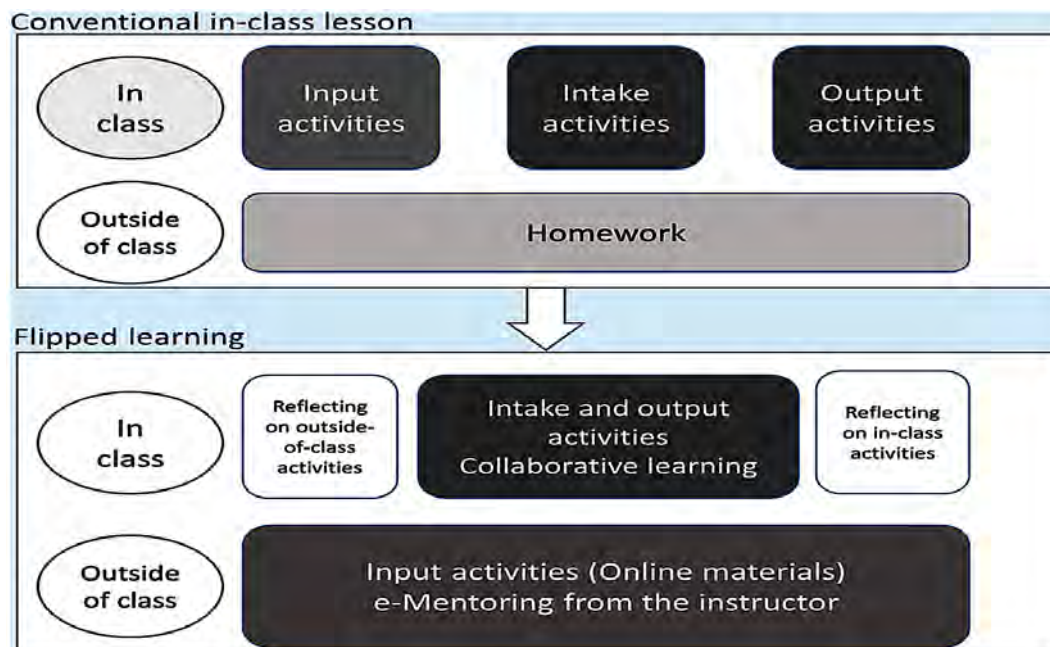
The following conceptual diagram was organized according to the principles of the phases of the continuous academic learning cycle described in the work of Schunk & Zimmerman (1998) which was further developed by Cleary & Zimmerman (2004). The three phases of the cycle make up a process which facilitates the transfer of a sense of ownership of the learning from the teacher to the student. The first phase of the cycle is *forethought*, intended to be self-motivating, in which students analyze their tasks, set goals, and plan strategies to achieve their goals. The second phase is termed *performance control*. In this phase self-instruction is based on self-determined focuses of attention in order to plan learning task strategies. Self-observation of the processes and products of learning are intended to achieve a sense of self

control. The third phase, intended to restart the cycle is *self-reflection* in which students make self-judgments by evaluating their own learning and attributing the various results to consequences of their own needs and their own actions (Kondo et al., 2012). The aim of the third phase is to achieve a feeling of self-satisfaction that may lead to the development of adaptability as the students learn how to deal with unexpected, as well as expected, consequences of their learning. Figure 1 shows a conceptual diagram of the transformation of a conventional in-class lesson to an FL lesson.

A Definition of Flipped Learning

FL, in this description of the transformation process from conventional EFL methodology to a flipped learning approach, is a form of BL in which students complete EFL course TOEIC study materials outside of class online and receive personalized problem-solving guidance from the teacher in class.

Figure 1. Conceptual diagram of the transformation of a conventional in-class lesson to flipped learning



The Flipped Learning Course Design

In order to help students improve their TOEIC scores, an FL course was designed and implemented at a university in Japan. Tasks were presented in the course in which classroom teaching methods and online delivery of course content and instruction were combined.

Each session had three phases: 1) pre-class online completion of TOEIC course materials; 2) individualized problem-solving instruction in class; and 3) post-class online completion of self-assessment and reflection learning tasks. Outside of class, the www-based courseware, ATR CALL BRIX was used.

In class the students first discussed their weaknesses and strengths in English listening and reading skills with their peers. After the students learned test-taking strategies to improve their TOEIC scores from the instructor, they completed quizzes given by the instructor and checked with their peers whether they had overcome their weaknesses. In an outside-of-class post-class session, the students carried out reviews of the in-class lesson with the learning materials contained in the ATR CALL BRIX online learning management system. Figure 2 outlines the learning tasks in one session of an FL lesson.

ATR CALL BRIX

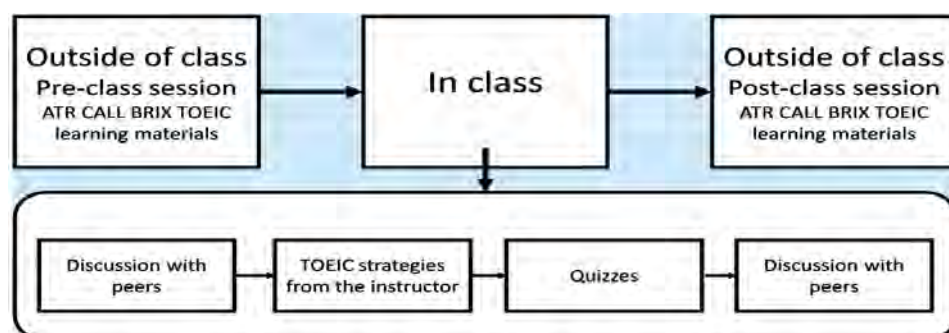
The www-based courseware, ATR CALL BRIX (<http://www.atr-lt.jp/products/brix/index.html>),

includes a learning management system (LMS). The LMS provides a variety of learning materials designed to prepare students for the TOEIC Test. Seven different functions were featured on the LMS: 1) study logs, 2) feedback on the achievement rates of student-set goals, 3) records of the frequency of the use of the materials, 4) a record of time spent on learning, 5) a continuous update of the students' average score on the TOEIC learning tasks, 6) an evaluation of students' weak points and advice for further learning, and 7) students' rankings in comparison with other students in the course (Ishikawa et al., 2014).

E-Mentoring: A Self-Evaluation System

A student self-evaluation system was integrated in the courseware that was intended to contribute to the development of students' SRL attitudes, skills and behavior, and thus, sustain student use of the learning materials. The system combined, as part of the outside-of-class course routine, e-mentoring with weekly self-evaluations by means of the LMS. The students in the experimental FL group were placed in three sections—high, mid and low (mean $\pm$ 0.5SD) – according to their scores of the TOEIC Test which the students took at the beginning of the course. An e-mentoring team of one teacher and a teaching assistant sent different needs-based messages of advice and encouragement to the students every week by means of a social networking system called LINE (<http://line>).

Figure 2. Procedure of activities in one session



me/en/). The messages were varied according to how successfully the students had completed the learning materials (Ishikawa et al., 2016).

VALIDATION OF THE STUDY

Participants

551 first-year students who enrolled in a TOEIC course participated in this study. 348 students were included in an experimental FL group and 203 students in a control group. The two groups were placed in separate classes.

Methods

For the FL experimental group, the course design described above, was implemented. In the control group, students followed the same course design but they used neither the web-based courseware, ATR CALL BRIX, nor the self-evaluation system. All their learning materials were paper-based. Pre- and post-TOEIC Tests were conducted before the course started and at the end of the second semester in January 2015. The experimental-group students took the TOEIC Test again at the end of the first semester in July 2014, in addition to the

pre- and post-TOEIC Test. Students in the FL group completed course evaluations at the end of the second semester.

Results and Discussion

TOEIC Testing

The increase in TOEIC scores between the mid and low sections of the FL experimental group and the control group was compared. The high section of the experimental group was excluded because the average scores for the pre-test in both experimental and control groups should be comparable. The increase of the scores in the experimental group was 151.38 and that in the control group was 54.04. The scores of the post-test in the two groups was compared and there was a significant difference ($p < .001$; $d = .56$) as is shown in Table 1.

Furthermore, students who took the TOEIC Test in April and July, 2014, and January, 2015 in each of the three sections of the FL experimental group improved their TOEIC scores by approximately 150 points from April, 2014 to January, 2015 (see Table 2). This is an indication the students in all three sections of the FL course sustained their study efforts throughout the full

Table 1. Results of pre- and post-TOEIC testing

	Experimental			Control			<i>t</i>	<i>d</i>
	<i>Mean</i>	<i>SD</i>	<i>n</i>	<i>Mean</i>	<i>SD</i>	<i>n</i>		
Pre-TOEIC (April)	345.97	66.82	243	361.03	89.14	203	2.00 *	0.14
Post-TOEIC (January)	497.35	107.77	243	415.07	107.20	203	8.05 ***	0.56

* $p < .05$, *** $p < .001$

Table 2. Results of TOEIC testing of the three sections in the FL experimental group

	April		July		January		<i>n</i>
	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>	
High	511.95	61.21	603.38	81.75	652.38	86.43	105
Intermediate	393.89	28.71	507.17	71.91	543.44	96.14	122
Low	286.12	43.65	409.58	79.17	443.50	96.35	107

academic year at a level at least equal to that of the students in the conventional class since they used the same basic TOEIC learning materials.

Post Course Student Evaluation

The course evaluation for the experimental FL group conducted at the end of the second semester consisted of 20 questions (see Table 3). Four open-ended questions asked about the students' feelings about the course and the in-class activities, about the problems that they may have had in using the LMS learning materials, and particularly about the messages sent by the e-mentor team. A 4-point Likert Scale was used for the responses in order to adequately allow for the expression of a range of students' feelings about the course. The ratings of 2 and 1, respectively, corresponded to disagree, and strongly disagree; and the ratings of 3 and 4, respectively, corresponded to agree, and strongly agree. The rate of reliability of the twenty 4-point Likert Scale questions was high ($\alpha = .91$).

The FL students were convinced that the in-class learning tasks were useful in strengthening their TOEIC listening and reading skills. Typical comments were similar to the following student's response: "I enjoyed the in-class activities. I was

able to learn TOEIC test-taking strategies which improved my TOEIC scores."

As for the use of the outside-of-class learning materials in the LMS of the ATR CALL BRIX, the majority of the students felt that the learning tasks improved their TOEIC scores. However, some of the students did not feel that they were able to solve problems that they had faced while using the LMS learning materials: "I don't know why but I often had problems in using the learning materials." There were other negative reactions to the use of the LMS: for example, looking at the display screen hurt their eyes; and they found problems in using the learning materials at home.

The advice of the e-mentor team seemed to be a key factor in the students' sustained FL study efforts. In response to the open-ended question many students commented, "The encouragement from the teacher helped me continue with the outside-of-class activities. I really think so." Some of the students believed that the FL approach to study suited their own learning styles.

The results of the evaluations showed that the FL students sustained their study efforts both in the in-class activities and in the online outside-of-class learning tasks. Furthermore, there was a suggestion that the use of a FL approach in EFL teaching would be sustainable.

Table 3. Results of the questions in the FL students' course evaluation

	Item	Mean	SD
1	The in-class activities helped improve my total TOEIC score.	3.26	0.79
2	The in-class activities helped improve my TOEIC listening score.	3.24	0.81
3	The in-class activities helped improve my TOEIC reading score.	3.17	0.84
4	The in-class activities helped improve my vocabulary.	2.90	0.84
5	The in-class activities helped improve my English listening skills.	3.19	0.82
6	The in-class activities helped improve my English reading skills.	3.00	0.85
7	In the in-class activities, I was able to solve problems that I found in the outside-of-class activities.	2.76	0.84
8	I was interested in the in-class activities.	2.82	0.89
9	I want to continue the in-class activities.	2.80	0.93
10	I used the learning materials in the LMS of the ATR CALL BRIX.	3.38	0.82
11	The learning materials in the LMS of the ATR CALL BRIX helped improve my total TOEIC score.	2.91	0.90
12	The learning materials in the LMS of the ATR CALL BRIX helped improve my TOEIC listening score.	3.05	0.86
13	The learning materials in the LMS of the ATR CALL BRIX helped improve my TOEIC reading score.	2.77	0.93
14	The learning materials in the LMS of the ATR CALL BRIX helped improve my vocabulary.	2.88	0.91
15	The learning materials in the LMS of the ATR CALL BRIX helped improve my English listening skills.	3.00	0.89
16	The learning materials in the LMS of the ATR CALL BRIX helped improve my English reading skills.	2.81	0.94
17	I was interested in using the learning materials in the LMS of the ATR CALL BRIX.	2.50	0.98
18	I want to continue using the learning materials in the LMS of the ATR CALL BRIX.	2.49	1.05
19	I often found some problems using the learning materials in the LMS of the ATR CALL BRIX.	2.51	1.03
20	The flipped learning lessons suited my own learning style.	2.67	0.87
<i>n</i> = 331			

FUTURE RESEARCH DIRECTIONS

Until recently, it was assumed that university students would mainly use laptop and desktop computers to access online learning materials at school and elsewhere. However, the use of a variety of mobile devices with online access, such as mobile phones and tablet computers, is now the preferred means of information gathering and student-to-student online communication in some university communities in Japan (Mynavi, 2015). Thus, further research should be conducted to investigate the most productive, and sustainable, uses of mobile devices for FL.

CONCLUSION

The FL approach to TOEIC test preparation, described in this chapter, helped improve students' scores on the TOEIC Test in comparison with the gains made by students in a conventional course. SRL skills at the mid- and low-English proficiency levels appeared to be a factor in sustaining students' efforts which resulted in greater gains in TOEIC scores. According to the review of BL literature referred to previously, students at mid and low English-proficiency levels may not only have weaker SRL skills, but they also cannot usually be expected to develop SRL skills through the use of online learning materials. Thus, the design and implementation of this FL course may be a useful guide to the design of other FL courses.

FL rests on an assumption that students will use learning task materials before class, so that classroom time can be devoted to interaction between students and their teacher that will support the outside-of-class study through collaborative problem-solving tasks and analytical examinations of the learning materials. Thus, the findings reported in this chapter are important.

REFERENCES

- Andrade, M. S., & Bunker, E. L. (2009). A model for self-regulated distance language learning. *Distance Education*, 30(1), 47–61. doi:10.1080/01587910902845956
- Azevedo, R., Moos, D. C., Greene, J. A., Winters, F. I., & Cromley, J. G. (2008). Why is externally-facilitated regulated learning more effective than self-regulated learning with hypermedia? *Educational Technology Research and Development*, 56(1), 45–72. doi:10.1007/s11423-007-9067-0
- Barker, P. (2006). Motivating Learners – Can Blended Learning Help? In T. Reeves, & S. Yamashita (Eds.), *Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 2006* (pp. 1849-1854). Chesapeake, VA: Association for the Advancement of Computing in Education.
- Bierema, L. L., & Merriam, S. B. (2002). E-mentoring: Using computer mediated communication to enhance the mentoring process. *Innovative Higher Education*, 26(3), 211–227. doi:10.1023/A:1017921023103
- Bishop, J. L., & Verleger, M. L. (2013). The flipped classroom: A survey of the research. *120th ASEE Annual Conference & Exposition*, 1-18. Retrieved from www.asee.org
- Chang, S. (2004). The roles of mentors in electronic learning environments. *AACE Journal*, 12(3), 331–342.
- Cho, M. H., Demei, S., & Laffey, J. (2010). Relationships between self-regulation and social experiences in asynchronous online learning environments. *Journal of Interactive Learning Research*, 21(3), 297–316.
- Cleary, T. J., & Zimmerman, B. J. (2004). Self-regulation empowerment program: A school-based program to enhance self-regulated and self-motivated cycles of student learning. *Psychology in the Schools*, 41(5), 537–550. doi:10.1002/pits.10177

- Dixon, S., Osment, M., & Panke, S. (2009). Comparing effectiveness of traditional versus blended teaching methods: Efforts to meet the demands of students in a blend 2.0. In G. Siemens, & C. Fulford (Eds.), *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2009* (pp. 2765-2774). Chesapeake, VA: Association for the Advancement of Computing in Education.
- Egbert, J., Herman, D., & Chang, A. (2014). To flip or not to flip? That's not the question: Exploring flipped instruction in technology supported language learning environment. *International Journal of Computer-Assisted Language Learning*, 4(2), 1–10. doi:10.4018/ijcallt.2014040101
- Hamdan, N., McKnight, P., McKnight, K., & Arfstrom, K. M. (2013). *A review of flipped learning*. Retrieved from <http://www.flippedlearning.org/review>
- Hartnett, M., St. George, A., & Dron, J. (2010). Factors undermining motivation in place-based blended learning. *Journal of Open. Flexible and Distance Learning*, 15(1), 1–16.
- Hess, H. W. (2006). Beyond the rhetoric of the 'autonomous learner': Combining e-learning and the classroom. *Electronic Journal of Foreign Language Teaching*, 3(1), 102–120.
- Ishikawa, Y., Akahane-Yamada, R., Kitamura, M., Smith, C., Tsubota, Y., & Dantsuji, M. (2016). Student self-evaluation system: Sustaining outside-of-class CALL activities in a university EFL blended learning course. In A. Gimeno-Sanz, M. Levy, F. Blin, & D. Barr (Eds.), *WorldCALL: Sustainability and computer-assisted language learning* (pp. 190–208). London, UK: Bloomsbury Publishing.
- Ishikawa, Y., Akahane-Yamada, R., Kondo, M., Smith, C., Tsubota, Y., & Dantsuji, M. (2014). An interoperable ICT educational application for TOEIC preparatory study. In M. Khosrow-Pour (Ed.), *Encyclopedia of information science and technology* (3rd ed.; pp. 2433–2444). Hershey, PA: Information Science Reference.
- Ishikawa, Y., Kondo, M., & Smith, C. (2010). Design and implementation issues of interoperable educational application: An ICT application for primary school English education in Japan. In F. Lazarinis, S. Green, & E. Pearson (Eds.), *Developing and utilizing e-learning applications* (pp. 100–124). Hershey, PA: Information Science Reference. doi:10.4018/978-1-61692-791-2.ch006
- Kondo, M., Ishikawa, Y., Smith, C., Sakamoto, K., Shimomura, H., & Wada, N. (2012). Mobile assisted language learning in university EFL courses in Japan: Developing attitudes and skills for self-regulated learning. *ReCALL*, 24(2), 169–187. doi:10.1017/S0958344012000055
- Kuhn, K., & Hoffstaedter, P. (2015). Flipping intercultural communication practice: Opportunities and challenges for the foreign language classroom. In J. Colpaert, A. Aerts, M. Oberhofer, M. Gutiérrez-Colón Plana (Eds.), *Proceedings of Seventeenth International CALL Conference Task design and CALL* (pp. 338-345). Antwerp, Belgium: Universiteit Antwerpen.
- Lage, M., Platt, G. L., & Treglia, M. (2000). Inverting the classroom: A gateway to creating an inclusive environment. *The Journal of Economic Education*, 31(1), 30–43. doi:10.1080/00220480009596759
- Leis, A., Cooke, S., & Tohei, A. (2015). The effects of flipped classrooms on English composition writing in an EFL environment. *International Journal of Computer-Assisted Language Learning*, 5(4), 37–51. doi:10.4018/IJCALLT.2015100103
- McKenzie, B. K., & Ozkan, B. C. (2006). Building an effective peer to peer distance mentoring program. In T. Reeves, & S. Yamashita (Eds.), *Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 2006* (pp. 740-745). Chesapeake, VA: Association for the Advancement of Computing in Education.

Mynavi. (2015). *2016 nen sotsu daigakusei no life style report* [Survey report on the life style of university students who will graduate in 2016]. Retrieved from http://saponet.mynavi.jp/enq_gakusei/lifestyle/data/lifestyle_2016.pdf

Ng, C., Seeshing Yeung, A., & Yuk Hung Hon, R. (2006). Does online language learning diminish interaction between student and teacher? *Educational Media International*, 43(3), 219–232. doi:10.1080/09523980600641429

Norton, P. (2005). Scaffolding online learning: The ART of mentoring. In C. Crawford, R. Carlsen, I. Gibson, K. McFerrin, J. Price, R. Weber, & D. A. Willis (Eds.), *Proceedings of Society for Information Technology & Teacher Education International Conference 2005* (pp. 2025–2032). Chesapeake, VA: Association for the Advancement of Computing in Education.

Osguthorpe, R. T., & Graham, C. R. (2003). Blended learning environments: Definitions and directions. *The Quarterly Review of Distance Education*, 4(3), 227–233.

Parsons, D. (Ed.). (2011). *Combining e-learning and m-learning: New applications of blended educational resources*. Hershey, PA: Information Science Reference. doi:10.4018/978-1-60960-481-3

Schunk, D. H., & Zimmerman, B. J. (Eds.). (1998). *Self-regulated learning: From teaching to self-reflective practice*. New York: The Guilford Press.

Stuntz, D. F. (2013). Flipped classrooms and CALL sustainability: A rationale for the development of flipped classrooms for sustainable CALL. *WorldCALL 2013: Global Perspectives on Computer-Assisted Language Learning*, 323–326.

Sung, K. (2015). A case study on a flipped classroom in an EFL content course. *Multimedia-Assisted Language Learning*, 18(2), 159–187.

Vaughan, N. (2007). Perspectives on blended learning in higher education. *International Journal on E-Learning*, 6(1), 81–94.

Yarbro, J., Arfstrom, K. M., McKnight, K., & McKnight, P. (2014). *Extension of a review of flipped learning*. Retrieved from http://researchnetwork.pearson.com/wp-content/uploads/613_A023_FlippedLearning_2014_JUNE_SinglePage_f.pdf

Zhao, P., & Johnson, G. (2012). A theoretical framework of self-regulated learning with web-based technologies. *Proceedings of Global TIME, 2012*, 163–168.

ADDITIONAL READING

Baker, J. W. (2000). The “classroom flip”: Using web course management tools to become the guide by the side. In J. A. Chambers (Ed.), *Selected papers from the 11th international conference on college teaching and learning* (pp. 9–17). Jacksonville, FL: Florida Community College at Jacksonville.

Bergmann, J., & Sams, A. (2012). *Flip your classroom: Reach every student in every class every day*. Washington, DC: International Society for Technology in Education.

Berrett, D. (2012). How ‘flipping’ the classroom can improve the traditional lecture. *The Chronicle of Higher Education*, 58(21), 16–18.

Chen, Y., Wang, Y., Kinshuk, N., & Chen, N. (2014). Is FLIP enough? Or should we use the FLIPPED model instead? *Computers & Education*, 79, 16–27. doi:10.1016/j.compedu.2014.07.004

Cheon, J., Lee, S., Crooks, S. M., & Song, J. (2012). An investigation of mobile learning readiness in higher education based on the theory of planned behaviour. *Computers & Education*, 59(3), 1054–1064. doi:10.1016/j.compedu.2012.04.015

Clark, K. R. (2015). The effects of the flipped model of instruction on student engagement and performance in the secondary mathematics classroom. *The Journal of Educators Online*, 12(1), 91–114. doi:10.9743/JEO.2015.1.5

Cockrum, T. (2014). *Flipping your English class to reach all learners: Strategies and lesson plans*. New York: Routledge.

Dörnyei, Z., & Kubanyiova, M. (2014). *Motivating learners, motivating teachers: Building vision in the language classroom*. Cambridge, UK: Cambridge University Press.

Dörnyei, Z., & Ryan, S. (2015). *The psychology of the language learner revisited*. New York: Routledge.

Enfield, J. (2013). Looking at the impact of the flipped classroom model of instruction on undergraduate multimedia students at CSUN. *Techtrends: Linking Research & Practice To Improve Learning*, 57(6), 14–27. doi:10.1007/s11528-013-0698-1

Freeman Herreid, C., & Schiller, N. (2013). Case studies and the flipped classroom. *Journal of College Science Education*., 42(5), 62–66.

Fulton, K. (2012). Upside down and inside out: Flip your classroom to improve student learning. *Learning and Leading with Technology*, 39(8), 12–17.

Kalk, D., Kinney, L., Liu, M., & Orr, G. (2012). Web 2.0 and its use in higher education from 2007-2009: A review of literature. *International Journal on E-Learning*, 11(2), 153–179.

Kim, M. K., Kim, S. M., Khera, O., & Getman, J. (2014). The experience of three flipped classrooms in an urban university: An exploration of design principles. *The Internet and Higher Education*, 22, 37–50. doi:10.1016/j.iheduc.2014.04.003

Moraros, J., Islam, A., Yu, S., Banow, R., & Schindelka, B. (2015). Flipping for success: Evaluating the effectiveness of a novel teaching approach in a graduate level setting. *BMC Medical Education*, 15(1), 27. doi:10.1186/s12909-015-0317-2 PMID:25884508

Scheg, A. G. (Ed.). (2015). *Implementation and critical assessment of the flipped classroom experience*. Hershey, PA: Information Science Reference; doi:10.4018/978-1-4666-7464-6

Slomanson, W. R. (2014). Blended learning: A flipped classroom experiment. *Journal of Legal Education*, 64(1), 93–102.

Strayer, J. F. (2012). How learning in an inverted classroom influences cooperation, innovation and task orientation. *Learning Environments Research*, 15(2), 171–193. doi:10.1007/s10984-012-9108-4

Vaughan, M. (2014). Flipping the learning: An investigation into the use of the flipped classroom model in an introductory teaching course. *Education Research and Perspectives*, 41(1), 25–41.

KEY TERMS AND DEFINITIONS

Blended Learning: Blended learning in this chapter is defined as a combination of in-class activities with outside-of-class activities.

E-Mentoring: E-Mentoring may establish a computer mediated, mutually beneficial relationship between a mentor and a learner if the mentor advises and encourages the learner by modeling effective learning behavior in ways that are boundary-less, egalitarian, and qualitatively different from traditional face-to-face mentoring.

Flipped Learning: Flipped learning in this chapter is defined as a form of blended learning in which students complete the EFL course TOEIC study materials outside of class online and receive personalized problem-solving guidance from the teacher in class.

Self-Regulated Learning: Self-regulated learning is a set of proactive processes that students use to acquire academic skills, such as setting goals, selecting and deploying strategies, and self-monitoring one's own effectiveness.

A Framework for Profiling Prospective Students in Higher Education



Santhosh Kumar Lakkaraju
Dakota State University, USA

Deb Tech
Dakota State University, USA

Shuyuan Deng
Dakota State University, USA

INTRODUCTION

Prospective student acquisition is a prominent issue in higher education marketing. Noel-Levitz (2012) estimated that higher education institutions are losing as high as 75% of the prospects after receiving an inquiry. Another study reported that 80% of the students who decide to apply to a program were influenced by the post-inquiry communications they had received from the higher education institutions (Aarinen, 2012). This chapter attempts to study the underlying concepts from literature and design a framework to extract prospective student profiles and further extend a discussion on how these profiles can be used to address the prospect engagement.

BACKGROUND

In general, the consumer buying decision-making process consists of five different phases that drive potential shoppers throughout their purchase process. Kotler & Armstrong (2006) defined the five consumer buying decision-making phases as Need Recognition, Information Search, Evaluation of Alternatives, Making a Decision and Post Purchase Behavior. The higher education institutions tackle these consumer buying decision-making phases through the four phases of the admission funnel. The admission funnel primarily consists of the

awareness, inquiry, and application and admissions phases as shown in Figure 1.

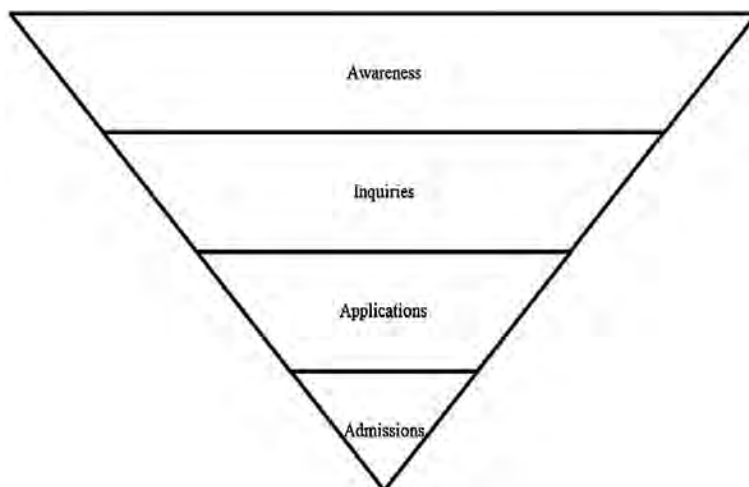
The awareness phase involves different marketing techniques the institutions rely on to reach out to prospects. The awareness phase will address the need recognition and information search phases by providing relevant information on the institutional websites. During the inquiry phase, a prospect tends to look for potential information on the institutional website and makes an inquiry by filling out the inquiry form. The institution responds to those inquiries by sending out different kinds of communications to prospects. These communications play a critical role in helping prospects to make a decision to apply or not.

The inquiry phase primarily targets the evaluation of alternatives and making a decision phases.

In the application phase, the institutions receive an application from the prospect. The cycle ends with the admission phase where the prospect would receive a decision on the application from the Institution.

To better market themselves and increase their student population, higher education institutions are employing different techniques like online (or) pay per click marketing and print media to increase their local and global presence and, social media to increase their social presence as well as brand awareness. Because of this exponentially growing educational market and varying prospective student behavior, institutions are receiving a

Figure 1. Admission Funnel



large number of inquiries from prospects about a specific program than the actual enrollments into that program (Hemsley-Brown & Oplatka, 2006; Moogan, 2011; Morris, 2009).

Earlier studies investigated several key decision-making variables of a prospective student (Aarinen, 2012; Moogan, Baron, & Harris, 1999; Moogan, 2011; Schäfer & Kummer, 2013), and some studies investigated the current student demographics to predict prospective student enrollment (Desjardins, 2002; Goenner & Pauls, 2006; Tareef & Balas, 2009). Other research studies developed predictive models using prospective student geo-demographic information collected through the online inquiries and estimated the prospective student enrollment rates (Goenner & Pauls, 2006; Michael, 1990; Morris, 2009). However, most of the online inquiries that educational institutions receive are incomplete, which will eventually provide incorrect predictions (Dupaul, 2010).

Moogan (2011) articulated that due to the lack of awareness about the kind of information a prospective student might be interested in receiving during the decision-making period, many educational institutions are losing potential prospects. The prospects tend to look for information before and after making the inquiry. The information accessed before making an inquiry is considered

as pre-inquiry navigational behavior and the information accessed after making an inquiry is considered as post-inquiry navigational behavior.

In general, profiling can be defined as the recording and analysis of an individual's psychological and behavioral characteristics (Nicoletti, Schiaffino, & Godoy, 2013). Building prospective student profiles is a complex task, as prospects do not usually give away explicit information about their interests (Catherine Bounsaythip, 2001; Srivastava, Cooley, Deshpande, & Tan, 2000). Therefore, the prospective student interests must be mined implicitly from the web server logs. Constructing accurate and comprehensive customer profiles play a key role in target marketing and enhanced customer engagement (Adomavicius & Alexander, 2001; Crossley, Kings, & Scott, 2003; Nicoletti et al., 2013).

Constructing prospective student profiles begin with collecting the prospect's information from various sources like online inquiry forms, campus visits, information brochures, educational fairs and job fairs. From a prospective student perspective, the general educational purchase process can be described in four phases: general interest in higher education, research for a specific institution or program of interest, decision to apply for one or more schools and finally, making a decision to enroll in a specific program (Goenner & Pauls, 2006).

From a higher education institution perspective, the general educational sales and marketing funnel or the admission funnel shown in Figure 1 has four distinct stages: awareness, inquiries, applications and admissions (Kotler & Armstrong, 2006; Nicolescu, 2009; Noel-Levitz, 2012; Oplatka & Hemsley-Brown, 2004). A prospect browses through different pages within an institutional website leaving trails of navigational information that can be mined from the server logs. This navigational behavior will be used in the extraction of prospective student profiles. Profiling prospective students based on their priorities would help in channeling a prospect to specific communications and increase satisfaction (Bhate & Pasha, 2014). It is proposed that prospective students' priorities can be identified from the prospect's existing browsing activities using pre-inquiry and post-inquiry navigational information.

Desjardins (2002) implemented an analytical strategy to assist higher education institutional marketing efforts. Desjardins (2002) applied a conceptual model based on the human capital theory that considered variables like current student demographics, admitted years, enrolled programs, application forms etc. Desjardins (2002) attempted to fit a statistical model by considering the historical data of admitted students and tried to accurately predict enrollment. Following this work, Goenner & Pauls (2006) proposed a model to predict the enrollment decisions of prospective students based on their inquiries. Goenner & Pauls (2006) combined the prospective student demographics with US census data and proposed that the prospects from a specific geographic region behave in a specific pattern.

Goenner & Pauls (2006) & Desjardins (2002) predicted a prospect's enrollment decision and then suggested specific marketing communications channels for prospective students, current students, and alumni. Moogan, (2011) specified that customer specific information in the communications chain might improve the retention rates as well as the brand image of the institution.

He also articulated that the prospects are to be considered as valuable customers and complete effective communications are to be exchanged between the prospects. Such relationship needs to be established from the inquiry phase till his/her graduation from the program. According to Moogan (2011), most of the existing research was conducted before or in the early stages of the evolution of online marketing and do not reflect the current marketing technologies to match the information needs of the students.

This chapter attempts to generate a dialog of ongoing higher education marketing efforts and the contribute to the institutional advancement by making use of analytics.

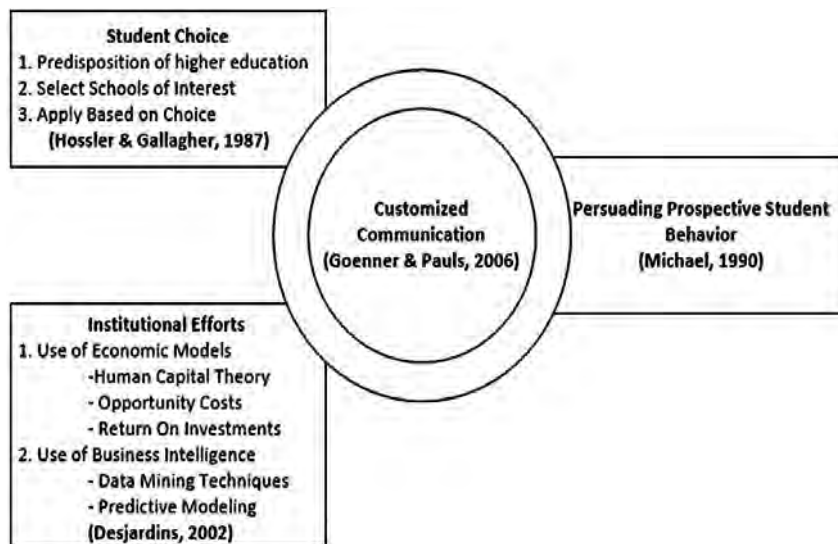
MAIN FOCUS OF THE CHAPTER

This chapter focuses on a comprehensive review of existing higher education marketing literature and extracts a theoretical model. This theoretical model is further used in formulating the framework to profile prospective students.

THEORETICAL MODEL

The theoretical model is derived from the works of Desjardins, (2002); Goenner & Pauls, (2006); Hossler & Gallagher, (1987); Michael, (1990). Hossler & Gallagher, (1987) articulated that a student's choice of an institution depends on the predisposition of pursuing higher education, selecting schools of interest and apply to the institutions based on choice. Desjardins, (2002) articulated that higher education institutions tend to make use of economic models and business intelligence models in promoting marketing methods and reporting. Goenner & Pauls, (2006) developed a model that made use of the prospect demographic and financial information extracted from the applications and predicted the enrollment numbers. Michael, (1990) emphasized on different factors that influence the prospects behavior from choosing one university over the other.

Figure 2. Theoretical model derived from literature



The theoretical model shown in Figure2 illustrates the interdependency of the student choice in researching for potential schools in the market as well as the institutional efforts in capturing the prospective student's attention through their communications. This prospective student information can be further mined and used to personalize and customize the communications that have a direct impact on persuading prospective students. According to Oinas-kukkonen & Harjumaa, (2009), information tailored to the potential needs of a prospect or to the interests and personality of a prospect will be more persuasive. From the theoretical model, it is clear that student choice, institutional efforts and the communications that a prospect receives will persuade a prospective student's decision-making behavior.

PROSPECT DECISION-MAKING FACTORS

There are different factors that contribute to a prospect's decision in selecting a graduate program in an educational institution. Extensive research has been done in identifying different decision-making factors of a prospective student (Moogan et al., 1999; Moogan, 2011; Sheppard, 2013).

Moogan (2011) analyzed the decision-making criteria of new undergraduates enrolling in terms of marketing techniques employed throughout the decision-making period.

Sheppard (2013) investigated different factors that influence prospective students in decision-making and the aggregated analysis is provided in the following table. A survey instrument was developed by Sheppard, (2013) that addressed six different aspects namely: external influences, education and career goals, information gathering, university financial aid, program characteristics and university characteristics. Noel-Levitz (2012) conducted a survey to extract different factors that influence graduate student college choice. The results varied from online versus regional prospects. Different key prospective student decision-making factors identified from the literature are articulated in Table 1.

The decision-making factors in table 1 have provided this study a foundational reference. The following section provides an overview on the translation of the decision-making factors into prospect profiles based on their contextual relevance.

Table 1. Key prospective student decision-making factors

Author	Decision-Making Factors
(Aarinen, 2012; María Cubillo, Sánchez, & Cerviño, 2006)	International recognition, suitability, reputation, specialization, quality of the program, courses, future earnings, future job or career opportunities, admission requirements, language requirements, educational facilities, fee, financial aid, City image, institution size.
(Moogan, 2011)	Teaching quality, course content, university reputation, research quality, faculty reputation, accreditation, facilities, student life, career prospects, entry dates, open day, the cost of living, accommodation, friends and family opinion, teacher's opinion, distance from home.
(Morris, 2009)	Electronic catalog, electronic application, inquiry forms, financial aid forms, course registration, email correspondence are some of the key decision-making factors
(Sheppard, 2013)	Program availability, career goals, income, credentials, personal development, flexibility of class scheduling, location, cost of attendance, reputation

FRAMEWORK

Different decision-making factors identified from literature (Aarinen, 2012; María Cubillo et al., 2006; Moogan, 2011; Morris, 2009; Sheppard, 2013) spoke about the factors consisting of specific contextual information. These decision-making factors are categorized into five different prospect profiles namely: price, program, future employment, institutional image, and the environment. This categorization is done based on their contextual relevance as shown in the framework below in Figure 3.

From Table 1, the factors cost of attendance, financial aid availability and cost of living provide information related to price.

The factors availability of the program, online, part-time, distance, location, and flexibility of class schedules provide program related information to the prospects.

The factors career advancements and goals, course content, future jobs, future earnings and on-campus employment provide a prospect with future employment related information.

The factors institutional Reputation, teaching quality, faculty expertise and reputation, research quality, quality of the program and institutional provide the prospect with an institutional image related information.

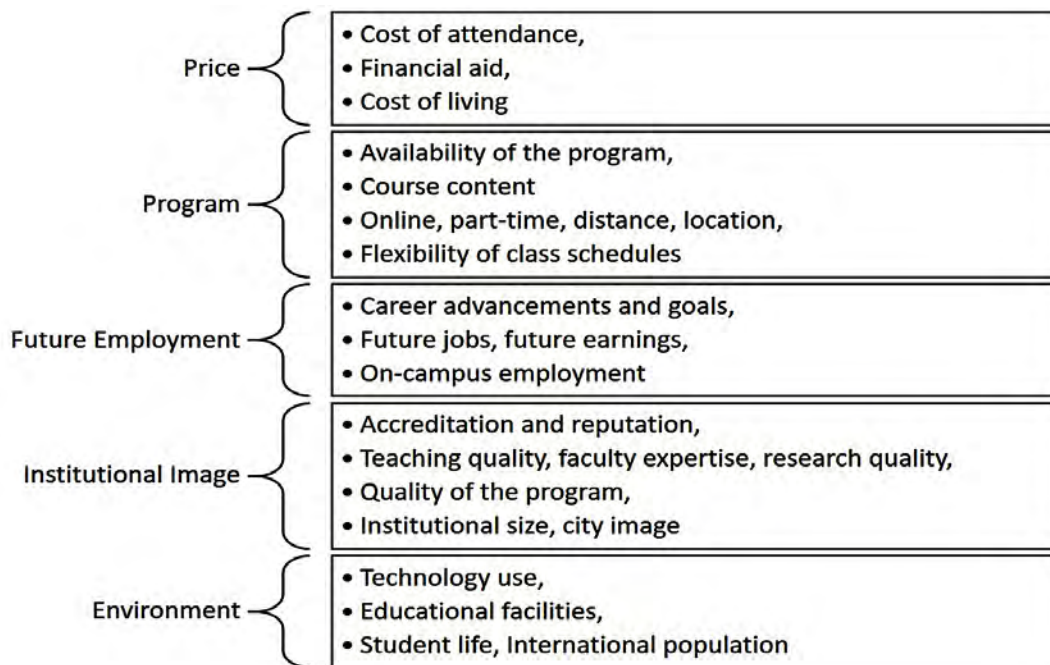
The factors technology use, educational facilities, and student life provide the prospect with the environment related information.

Institutions provide specific information addressing the decision-making factors on their websites. Based on these prospect profiles the web pages within the institutional website are categorized into different content profiles. Pages that address a specific context are tagged with prospect profile names based on the context as followed: Price, Program, Future employment, Institutional Image and Environment.

IMPORTANCE OF PROSPECT PROFILES

This section provides an overview of crucial decision-making factors and their importance in this research. Extensive study has been conducted on identifying different influential factors that impact a prospect's choice of an institution or program. Ivy & Naude, (2004) introduced a 5P model where the 5P's stands for product, price, place, promotion and people. In a detailed sense, the product represents the program a prospect is going to invest his money on; price represents the cost of the program, place deals with the environment or location of the institution, promotion targets the future employment and people deals with the student life and institutional image. Filip, (2012) proposed a 7P model and included processes and physical facilities to the existing 5Ps. Processes refer to the way the enrollment system, teaching and learning habits, social and sports activities are established within the institution. Physical facili-

Figure 3. Framework representing the prospect profiles



ties address the institutional equipment, technical infrastructure etc.

Although it is up to an institution to target specific factors in reaching out to prospects, a research study conducted by Noel-Levitz, (2012) clearly articulated that cost, financial aid, academic reputation, institutional size, future employment and campus location are crucial prospect decision-making factors. Higher education institutions usually structure their program-related information targeting these specific decision-making factors on their websites.

SOLUTIONS AND FUTURE RESEARCH DIRECTIONS

This framework can be used along with analytical models to optimize prospect communications, provide timely and relevant information to the prospects. This framework can act as a foundational reference for accurately profile prospects. The prospect profiling framework is designed by relying on the existing literature.

With the increasing use of social media by the prospects, institutions are trying to expand their reputation and social presence by investing valuable resources in social media. There might be some unknown variables that may act as decision-making factors for prospects in choosing one institution over the other. In the future, this study can be extended in extracting prospect preferences and decision-making factors on institutional social profiles.

CONCLUSION

This chapter provided an overview of the existing higher education marketing literature and extracted different prospect decision-making factors. These decision-making factors are further translated into prospect profiles based on their contextual relevance. The prospect profiles from the framework can be used in the institutional marketing strategies to optimize the prospect communications, identify prospect behavior over the institutional website and optimize the web pages accordingly.

ACKNOWLEDGMENT

We would like to sincerely thank all the reviewers for their inputs and suggestions. The chapter is intended for audience from higher education marketing. Different profiles extracted from the literature intend to bring a conceptual awareness in the marketing departments to find different ways to target their prospects and address their needs. The chapter has been modified in accordance with the reviewers suggestions and we also appreciate the reviewers for the time spent on this chapter.

REFERENCES

- Aarinen, A. (2012). *University Websites As Facilitators Of International Student Decision Making*. Academic Press.
- Adomavicius, G., & Alexander, T. (2001). Using Data Mining Methods to Build Customer Profiles. *IEEE Computer*, 74–82.
- Bhate, D. V., & Pasha, M. Y. (2014). Analyzing Target Customer Behavior Using Data Mining Techniques for E- Commerce Data. *International Journal of Innovative Research in Computer Science & Technology*, 2(1), 16–19.
- Catherine Bounsaythip, E. R.-R. (2001). *Overview of Data Mining for Customer Behavior Modeling*. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.22.3279>
- Crossley, M., Kings, N. J., & Scott, J. R. (2003). Profiles — Analysis and Behaviour. *BT Technology Journal*, 21(1), 56–66. doi:10.1023/A:1022404310934
- Desjardins, S. L. (2002). An Analytic Strategy To Assist Institutional Recruitment And Marketing Efforts. *Research in Higher Education*, 43(5), 531–553. doi:10.1023/A:1020162014548
- Dupaul, S. (2010). *Understanding The Behaviors of Stealth Applicants In The College Search Process*. Academic Press.
- Filip, A. (2012). Marketing theory applicability in higher education. *Procedia: Social and Behavioral Sciences*, 46, 912–916. doi:10.1016/j.sbspro.2012.05.223
- Goenner, C. F., & Pauls, K. (2006). A Predictive Model of Inquiry to Enrollment. *Research in Higher Education*, 47(8), 935–956. doi:10.1007/s11162-006-9021-8
- Hemsley-Brown, J., & Oplatka, I. (2006). Universities in a Competitive Global Marketplace: A Systematic Review of The Literature on Higher Education Marketing. *International Journal of Public Sector Management*, 19(4), 316–338. doi:10.1108/09513550610669176
- Hossler, D., & Gallagher, K. S. (1987). Studying student college choice: A three-phase model and the implications for policymakers. *College and University*, 62(3), 207–221.
- Ivy, J., & Naude, P. (2004). Succeeding in the MBA Marketplace: Identifying the Underlying Factors. *Journal of Higher Education Policy and Management*, 26(3), 401–417. doi:10.1080/1360080042000290249
- Kotler, P., & Armstrong, G. (2006). *Principles of Marketing*. Academic Press.
- María Cubillo, J., Sánchez, J., & Cerviño, J. (2006). International students decision-making process. *International Journal of Educational Management*, 20(2), 101–115. doi:10.1108/09513540610646091
- Michael, B. P. (1990). College Choice: Understanding Student Enrollment Behavior. ASHE - ERIC Higher Education Report No.6.

- Moogan, Y. J. (2011). Can a Higher Education Institutions Marketing Strategy Improve the Student-Institution Match? *International Journal of Educational Management*, 25(6), 570–589. doi:10.1108/09513541111159068
- Moogan, Y. J., Baron, S., & Harris, K. (1999). Decision-Making Behaviour of Potential Higher Education Students. *Higher Education Quarterly*, 53(3), 211–228. doi:10.1111/1468-2273.00127
- Morris, C. G. (2009). *Prospective Student Interaction and Engagement with the College : Linking Timeliness to Enrollment Likelihood*. Academic Press.
- Nicolescu, L. (2009). Applying Marketing to Higher Education: Scope and Limits. *Management & Marketing*, 4(2), 35–44.
- Nicoletti, M., Schiaffino, S., & Godoy, D. (2013). Mining interests for user profiling in electronic conversations. *Expert Systems with Applications*, 40(2), 638–645. doi:10.1016/j.eswa.2012.07.075
- Noel-Levitz. (2012). *2012 Marketing and Student Recruitment Practices for Master's-Level Graduate Programs*. Academic Press.
- Oinas-kukkonen, H., & Harjumaa, M. (2009). Persuasive Systems Design: Key Issues, Process Model, and System Features. *Communications of the Association for Information Systems*, 24(1), 1–18.
- Oplatka, I., & Hemsley-Brown, J. (2004). The Research on School Marketing: Current Issues and Future Directions. *Journal of Educational Administration*, 42(3), 375–400. doi:10.1108/09578230410534685
- Schäfer, K., & Kummer, T.-F. (2013). Determining the performance of website-based relationship marketing. *Expert Systems with Applications*, 40(18), 7571–7578. doi:10.1016/j.eswa.2013.07.051
- Sheppard, J. S. (2013). *Factors That Influence College Choice: Decisions of Graduate Students*. Academic Press.
- Srivastava, J., Cooley, R., Deshpande, M., & Tan, P. (2000). Web Usage Mining : Discovery and Applications of Usage Patterns from Web Data. *SIGKDD Explorations*, 1(2), 12–23. doi:10.1145/846183.846188
- Tareef, F., & Balas, A. (2009). *Marketing In Higher Education Institutes: Using An Internal Marketing Strategy And Innovation Models*. Academic Press.

ADDITIONAL READING

- Ali, D. M. M. (2013). The Role of Data Mining In Education Sector. *International Journal of Computer Science and Mobile Computing*, 2(4), 374–383.
- Bhate, D. V., & Pasha, M. Y. (2014). Analyzing Target Customer Behavior Using Data Mining Techniques for E- Commerce Data. *International Journal of Innovative Research in Computer Science & Technology*, 2(1), 16–19.
- Corley, J. K., Jourdan, Z., & Ingram, W. R. (2013). Internet marketing: A content analysis of the research. *Electronic Markets*, 23(3), 177–204. doi:10.1007/s12525-012-0118-y
- Husin, H. S. (2013). News recommendation based on Web usage and Web content mining. In *2013 IEEE 29th International Conference on Data Engineering Workshops (ICDEW)* (pp. 326–329). Ieee. <http://doi.org/> doi:10.1109/ICDEW.2013.6547475
- Nicoletti, M., Schiaffino, S., & Godoy, D. (2013). Mining interests for user profiling in electronic conversations. *Expert Systems with Applications*, 40(2), 638–645. doi:10.1016/j.eswa.2012.07.075
- Schäfer, K., & Kummer, T.-F. (2013). Determining the performance of website-based relationship marketing. *Expert Systems with Applications*, 40(18), 7571–7578. doi:10.1016/j.eswa.2013.07.051

Vellingiri, J., & Pandian, S. C. (2015). A Survey on Web Usage Mining. *International Research Journal of Engineering and Technology*, 02(1), 6–12.

Vesset, D., Morris, H. D., Eastwood, M., Woo, B., Villars, R. L., Bozman, J. S., Villars, R. L. (2012). Worldwide Big Data Technology and Services 2012 – 2015 Forecast. *IDC Analyze the Future*, 1(March).

Zarrella, D. (2014). *The science of Email Marketing*.

KEY TERMS AND DEFINITIONS

Analytics: The discovery, interpretation and communication of meaningful patterns in data.

Conversion: A *marketing* tactic that encourages a customer to take a specific action. In electronic commerce, *conversion marketing* is the phrase used to often describe the act of *converting* a customer who browses your site to a paying customer.

E-Marketing: Also known as internet marketing, web marketing, or digital marketing. They refer to advertising and marketing efforts that make use of web and email to drive traffic and make sales.

Profiling: The recording and analysis of a person's psychological and behavioral characteristics, so as to assess or predict their capabilities in a certain sphere or to assist in identifying a particular subgroup of people.

Prospects: A person regarded as likely to become a potential customer.

Importance of Information Literacy

Lidia Sanchez-Ruiz

University of Cantabria, Spain

Beatriz Blanco

University of Cantabria, Spain

INTRODUCTION

During the last decade an educational change has taken place in many European countries. The European Union has promoted several reforms in the university education system in order to create a common European Higher Education Area (EHEA).

The changes introduced aimed to adopt a comparable degree system (Sanchez-Ruiz, Pérez-Pérez, Blanco-Rojo, & Serrano-Bedia, 2013), being one of their main objectives the adoption of a competency-based education system.

Among the wide range of competences that students may acquire, the relevance of information literacy has increased during the last decade as we are said to live in an information society.

But, what does information society mean? As Webster (2006) said this concept has been extensively studied from different perspectives (technological, economic, occupational, spatial and cultural). Whatever the perspective, *they all share the conviction that quantitative changes in information are bringing into being a qualitatively new sort of social system, the information society* (Webster, 2006).

In this new scenario, information becomes a key aspect in everybody's professional and personal lives. Thus, a new concept appears: information literacy. *It entails the learning of the skills, competences, knowledge and value to access, use and communicate information in any of its forms, in order to produce competent professionals trained in the routines of identifying, evaluating and recording information sources appropriately and*

with the knowledge to process and produce their own information (Pinto, Doucet, & Fernández-Ramos, 2007).

Despite information literacy is not always considered a core competence, many organisations require information literacy skills on their job offers (Klusek & Bornstein, 2008). Therefore, every university student should acquire and develop information literacy skills during their undergraduate studies. For instance, among the whole range of information literacy skills, students should: be aware of the goals for which they are going to use information, know how to use available technologies to organise and store information, be able to assimilate the key information of documents or know how to synthesise and represent essential information properly (Pinto et al., 2007).

Taking this into consideration, as a first step, students should be aware of the relevance of information literacy skills. Once they are aware of its importance, they are able to improve and develop them. Thus, the aim of this study is to analyse whether university students are aware of the importance of information literacy. Additionally, we will analyse whether their awareness varies or not after attending an information training course.

Once the objectives of this study have been stated, the structure of the rest of the paper will be as follows. First, in the next section (Background) the basic concepts of this study (European Higher Education Area, Information literacy and the IL-HUMASS survey) are explained. After this, the following section includes information about the methodology used and, after that, the results obtained in this study are presented. Next, some

recommendations based in the results are made. Finally, future research suggestions are presented together with the conclusion of this research.

BACKGROUND

The European Higher Education Area

The European Higher Education Area (EHEA) is aimed to adopt a comparable degree system from a common system of credits (ECTS) based on two main levels –undergraduate and postgraduate– harmonizing its duration and combining the balance between basic education, cross-cutting skills, specific knowledge and discipline skills, and professional competence (Sanchez-Ruiz, Pérez-Pérez, Blanco-Rojo, & Serrano-Bedia, 2013).

Although the intention of convergence had already been expressed in the Sorbonne Declaration (1998), it was due to the Bologna Declaration that the process was accelerated. Specifically, the main objectives of the Bologna Declaration were (ANECA 2005, pp.27-28):

- Adoption of a model of degrees based on two levels: undergraduate and graduate.
- Creation of the European Credit Transfer System (ECTS).
- Establishment of a comprehensible and comparable degree system.
- Promoting mobility in Europe.
- Improving the quality of institutions.
- Promoting education and training throughout the whole career.
- Adopting a competence-based education system.

According to an extensive review of literature, companies require four characteristics from their employees: content of study, social skills, methodological skills and participation skills (Marzo-Navarro, Pedraja-Iglesias, & Rivera-Torres 2009). Before the implementation of the EHEA, universities primarily focused on the first one, the content

of study. However, later the development of other skills has been encouraged from the university.

Competence is understood as “the combination of knowledge, skills (intellectual, manual, social...), attitudes and values that will enable the graduate to successfully tackle the problem solving or intervention in an issue in an academic, professional or social context” (MEC 2006).

Two types of competences may be distinguished: specific and generic. Generic competences are those attributes that a graduate must have regardless of their degree; while specific competences are defined as observable behaviours that are directly related to the use of concepts, theories and skills of the degree (Martinez-Caro, & Cegarra-Navarro 2012, p.10).

There is a wide range of competences that students are supposed to acquire during their degrees (organising, solving, managing...). As above-stated, depending on the field of study, specific competences vary. However, generic competences are common to all the degrees. Among them, this study is focused in the information literacy competence that will be described in the following section.

Information Literacy

Although the concept of information literacy come into use in 1974, it has achieved its current popularity due to a proliferation of information available (Gradstein, 2002). Due to this fact, the concept of information literacy has been widely studied along literature from different perspectives and fields (Zabel, 2004).

Among all the existing definitions, the American Library Association definition has become widely accepted (Bherens, 1994). According to it (American Library Association, 1989), in order to *be information literate, a person must be able to recognize when information is needed and have the ability to locate, evaluate, and use effectively the needed information. Ultimately, information literate people are those who have learned how to learn. They know how to learn because they*

know how knowledge is organized, how to find information, and how to use information in such a way that others can learn from them. They are people prepared for lifelong learning, because they can always find the information needed for any task or decision at hand.

Taking this definition into consideration, Rader (1991) defined the following characteristics of an information literate person “*To be information literate means to do the following:*

- Survive and be successful in an information/technology environment;
- Lead productive, healthy, and satisfying lives in a democratic society;
- Effectively deal with rapidly changing environments;
- Ensure a better future for the next generation;
- Find appropriate information for personal and professional problem solving; and
- Have writing and computer proficiencies.

Therefore, according to Rader (1991), information-literate people know how to be lifelong learners in an information society.

More recently, Weber and Johnston (2003) defined information literacy as the adoption of appropriate information behaviour to identify, through whatever channel or medium, information well fitted to information needs, leading to wise and ethical use of information in society. Similarly, Pinto (2010) defined information literacy as the set of literacies or competencies that an informed citizen needs in order to participate judiciously and actively in an information society.

Regardless of whatever definition, there is a sufficient consensus about the relevance of information literacy as an educational goal (Owusu-Ansah, E.K., 2004). In fact, Burnheim (1992) already defined information literacy as a core competence as it is needed to learn and develop other competences. Additionally, regarding the above-mentioned classification, information literacy could be described as a generic competence

due to, no matter the degree, every student should learn and develop it.

As a result of the relevance of the concept, this study is aimed, firstly, to analyse whether university students are aware of the importance of information literacy. And, secondly, we will analyse whether their awareness varies or not after attending an information training course.

The IL-HUMASS Questionnaire

Among literature, different measurement and self-assessment tools have been created in order to measure information literacy development. Among them, in this study the IL-HUMASS survey is going to be used. It is *a comprehensive and user-friendly survey containing an exhaustive set of variables related to information literacy for the specific target population of the higher education in the humanities and social sciences* (Pinto, 2010). The election of this survey is due to the fact that this tool was specifically developed to be used among Spanish and Portuguese university students (Pinto, 2010).

According to this survey, information literacy skills are divided in four main factors: information search, information evaluation, information processing, and information communication and dissemination. In Table 1, the complete list of items which integrate each of the five factors are included.

In order to fulfil the objective of this research, this survey was conducted among higher education students. Thus, as it can be seen in Table 2, the scope of this study is limited to the Autonomous Community of Cantabria (a region in the North of Spain). Specifically, this survey was conducted among First Year university students from the Business and Management degree who attended an information training course. Each student had to value in a Likert scale (1 to 9) whether they consider each of the skills included in the survey important or not for their professional future.

The survey was conducted in March 2015 when the information training course began, and

Table 1. IL-HUMASS survey (Pinto, 2010)

Competence	Abilities	Code
Information search	Using printed sources of information (books, papers,...)	B.1
	Entering and using automatic catalogues	B.2
	Consulting and using electronic sources of primary information (journals...)	B.3
	Using electronic sources of secondary information (databases...)	B.4
	Knowing the terminology of your subject	B.5
	Searching for and retrieving internet information (advanced searches, directories...)	B.6
	Using informal electronic sources of information (blogs, discussion lists, etc.)	B.7
	Knowing information-search strategies (descriptors, Boolean operators, etc.)	B.8
Information evaluation	Assessing the quality of information resources	E.1
	Recognizing the author's ideas within the text	E.2
	Knowing the typology of scientific information sources (thesis, proceedings, etc.)	E.3
	Determining whether an information resource is updated	E.4
	Knowing the most relevant authors and institutions within your subject area	E.5
Information processing	Schematizing and abstracting information	T.1
	Recognizing text structure	T.2
	Using database managers (Access, MySWL,...)	T.3
	Using bibliographic reference managers (Endnote, Reference Manager...)	T.4
	Handling statistical programs and spreadsheets (SPSS, Excel...)	T.5
	Installing computer programs	T.6
Information communication and dissemination	Communicating in public	C.1
	Communicating in other languages	C.2
	Writing a document (report, academic work...)	C.3
	Knowing the code of ethics in your academic/professional field	C.4
	Knowing the laws on the use of information and intellectual property	C.5
	Creating academic presentations (Powerpoint...)	C.6
	Disseminating information on the internet (web, blogs...)	C.7

Table 2. Research technical record

Characteristics	Survey
Population	First Year university students from the Business and Management degree (University of Cantabria)
Population size	21 students
Geographical scope	Autonomous Community of Cantabria (region in the North of Spain)
Unit of analysis	Student
Period	March 2015-June 2015

in May 2015 when the course finished. The course is about searching and processing information strategies. The main learning outcomes of this basic course were: knowing the basic structure of a text; knowing the basic information sources that are available to students (books, specialised databases...); knowing the existing referencing rules; knowing the basics of several tools for data processing (Excel spreadsheet, Word processor, Powerpoint, posters...).

Finally, 21 valid surveys were collected. It represents around 60% of the students who were enrolled in the course, but nearly 90% of those who

usually attended it. At this point, authors would like to highlight that Rasch Measurement Theory is able to obtain valid and reliable results from small samples so, sample size, is not a problem in this case.

After collecting the data, Rasch Measurement Theory was used for data treatment. Due to the use of this methodology is still not widespread, the next section explains the basis of this method.

RASCH MEASUREMENT THEORY

In the Social Sciences field and, specifically, in the Business and Management area, there are many realities that cannot be directly measured. Thus, measurement is usually done indirectly by measuring a group of items that, in theory, integrate the construct or reality that we are interested in. In those cases, it is very common to use Likert scales to value those items. The scores obtained from the Likert scales cannot be considered measurements due to they are ordinal scores and, in order to consider them a measurement they should have an additive structure, a characteristic that only interval variables have.

Rasch Measurement Theory, initially developed by George Rasch (1960), solves this problem by transforming ordinal variables into interval variables. Rasch Measurement Theory is based in three principles: unidimensionality (a construct is unidimensional when all the items are referred to the same construct or latent variable so they can be located in the same lineal construct with the subjects); invariance (the results are independent from the samples of subjects and items used); and additivity.

When it comes to transform the ordinal variables into interval variables, the Theory is based in the following statement: *“persons who are more able/more developed have a greater likelihood of correctly answering all the items in the observation schedule. And, easier items are more likely to be answered/reached correctly by all persons”* ((Rasch 1960) in (Bond & Fox 2007, p.28)).

Then, the mathematical model is derived from the logistic function that relates the increasing probability of response to all items with the persons' ability (Gonzalez Montesinos, 2008).

From this statement, a new concept is defined: distance. It is the difference between the ability of the person (β) and the difficulty of the item (δ). Three situations may happen depending on the result of this difference:

- Si $\beta - \delta > 0$: the probability of the subject answering correctly to the item is more than 50%
- Si $\beta - \delta < 0$: the probability of the subject answering correctly to the item is less than 50%
- Si $\beta - \delta = 0$: the probability of the subject answering correctly to the item equals 50%

The mathematical expression of the model is derived from this idea and it can be consulted in any of the handbooks about the methodology (see (Bond, Fox 2007; Alagumalai, Curtis, & Hnugi 2005; von Davier, & Carstensen, 2007)).

Among its many applications we highlight the following: analysis of the global feasibility and reliability of the construct; analysis of the individual feasibility and reliability of the items and subjects; unidimensionality analysis; questionnaire scale analysis; and ranking of the items and/or persons (Sanchez & Blanco, 2012).

With regard to the data treatment, several software packages are available (Facets, Winsteps...). In this study the 3.75 Winsteps software is used (Linacre, 2012a).

RESULTS: UNIVERSITY STUDENTS' PERSPECTIVE

Firstly, the construct information literacy was validated, this is, the dimensionality, validity and reliability of the measures were globally and individually checked

After having checked the reliability and validity conditions, the abilities which integrated each of the competences were ordered based on the importance given by students.

Figure 1 shows the results obtained in the first survey, that is, the survey conducted at the beginning of the course. The vertical line represents the construct and the units are expressed in logits (interval units).

The crosses in the left of the vertical line represent the students' positions. On the other side, the codes in the right represent the items position. In each side, the M represents the average measure, the S represents one time the standard deviation distance, and finally T represents two times the standard deviation distance.

In general, due to the fact that the average measure for students (1.41 logits) is higher than the position of the average measure for the items (0.00 logits), it could be concluded that students do not consider information abilities important. Specifically, the most important item, according to students' perceptions, is the item C.2 "Communicating in other languages". In the opposite side, the least important item is E.3. "Knowing the typology of scientific information sources (thesis, proceedings, etc.)".

Once the course was finished, the survey was conducted again. Results are shown in Figure 2, whose interpretation follows the same rules explained in Figure 1. Although the specific results will be stated later, it is interesting to highlight that, comparing the position of the average measures, students consider information abilities more important after the course than at the beginning of the course.

In order to clarify the results obtained in Figures 1 and 2, the classification of information abilities based on students' perceptions are summarised in Table 3. Thus the first and second columns describe the code and the name of the abilities; the third column describes the position of each ability according to students' perceptions at the beginning of the course

(Figure 1), being the most important ability the one at the bottom as it was above mentioned; equally, the fourth column refers to the position of each ability according to students' perceptions at the end of the course (Figure 2), again being the most important one the item located at the bottom; finally, the fifth column compares the results of the two previous columns. Thus, if the position of an item gets worse means that students considered the ability more important at the beginning than at the end of the course; whereas if the item goes better, students value better the ability after the course than at the beginning. As it can be seen, abilities have been ordered based on students' perceptions at the beginning of the course.

Figure 1. Results from the initial survey

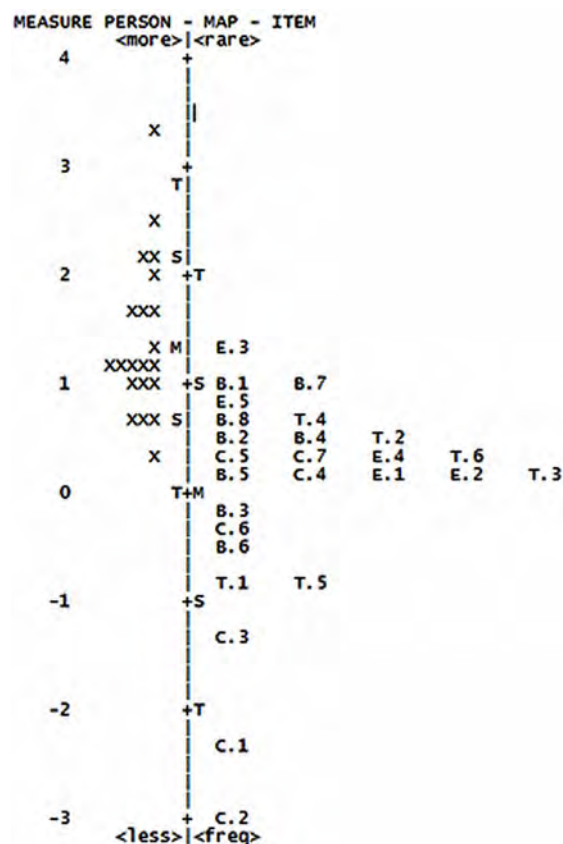
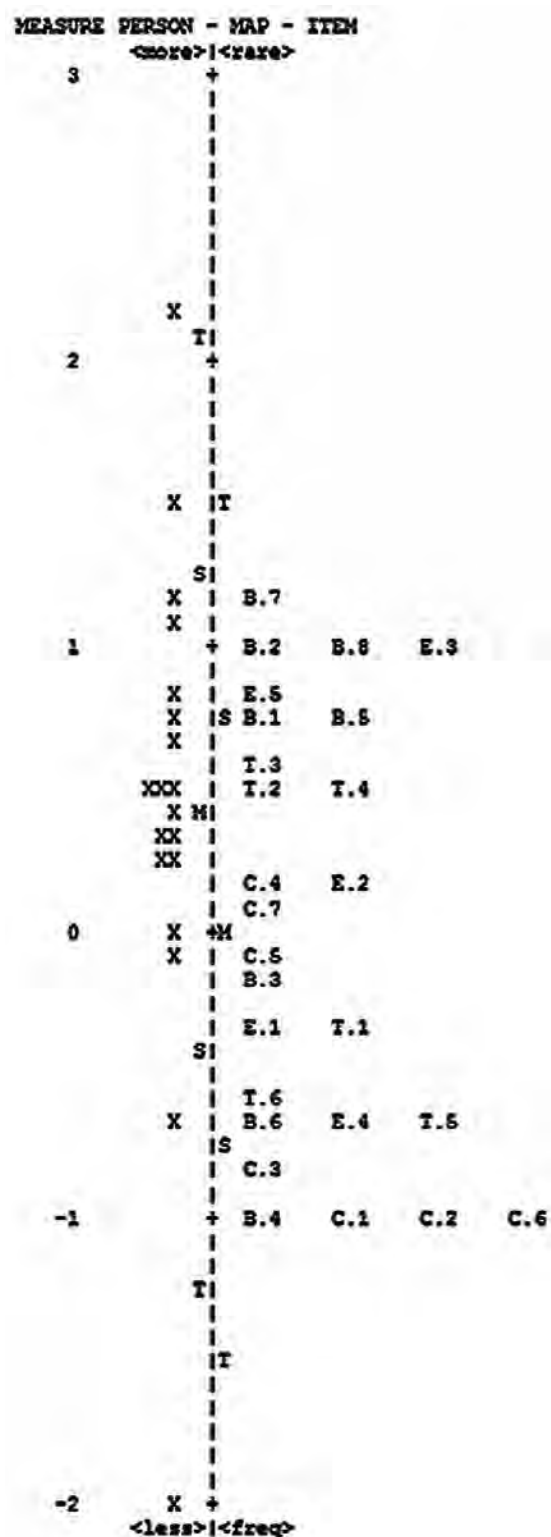


Figure 2. Results from the final survey



SOLUTIONS AND RECOMMENDATIONS

The objectives of this study were:

1. Analyzing whether university students are aware of the importance of information literacy
2. Analysing whether their awareness varies after attending an information training course

According to results, students do not consider information abilities important. This can be concluded from Figure 1, due to fact that the average measure for students (1.41 logits) is higher than the position of the average measure for the items (0.00 logits). However, it must be highlighted students increased the importance given after the information training course, as the difference between the average measures is reduced (average measure for items: 0 logits; average measure for students: 0.44 logits).

Additionally to these results, this research offers a classification of information abilities based on students' perceptions. Thus, writing and oral communication skills (C.1, C.2 and C.3) are the most important ones for them, being their position stable before and after the course. On the other side, knowing the typology of scientific information sources (E.3) is one of the least important abilities, although its position improves after the course.

FUTURE RESEARCH DIRECTIONS

Despite this study has been rigorously developed, the authors are aware of some of its limitations. For instance, authors consider that the scope of the study should be increased. Thus, in the future, the survey could be conducted among university students from different degrees, not only for the business and management degree. In the long term, the study should be conducted among students

Table 3. Results: Abilities position in the initial and in the final survey

Code	Abilities	Position According to Initial Survey	Position According to Final Survey	Evolution
C.2	Communicating in other languages	1	2	Get worse
C.1	Communicating in public	2	3	Get worse
C.3	Writing a document (report, academic work...)	3	5	Get worse
T.5	Handling statistical programs and spreadsheets (SPSS, Excel...)	4	6	Get worse
T.1	Schematizing and abstracting information	5	10	Get worse
B.6	Searching for and retrieving internet information (advanced searches, directories...)	6	8	Get worse
C.6	Creating academic presentations (Powerpoint...)	7	1	Get better
B.3	Consulting and using electronic sources of primary information (journals...)	8	12	Get worse
T.3	Using database managers (Access, MySWL,...)	9	19	Get worse
E.1	Assessing the quality of information resources	10	11	Get worse
B.5	Knowing the terminology of your subject	11	20	Get worse
C.4	Knowing the code of ethics in your academic/ professional field	12	15	Get worse
E.2	Recognizing the author's ideas within the text	13	16	Get worse
E.4	Determining whether an information resource is updated	14	7	Get better
C.7	Disseminating information on the internet (web, blogs....)	15	14	Get better
C.5	Knowing the laws on the use of information and intellectual property	16	13	Get better
T.6	Installing computer programs	17	9	Get better
T.2	Recognizing text structure	19	18	Get better
B.4	Using electronic sources of secondary information (databases...)	20	4	Get better
B.2	Entering and using automatic catalogues	21	25	Get better
B.8	Knowing information-search strategies (descriptors, Boolean operators, etc.)	22	24	Get better
T.4	Using bibliographic reference managers (Endnote, Reference Manager...)	23	17	Get better
E.5	Knowing the most relevant authors and institutions within your subject area	24	22	Get better
B.7	Using informal electronic sources of information (blogs, discussion lists, etc.)	25	26	Get worse
B.1	Using printed sources of information (books, papers,...)	26	21	Get better
E.3	Knowing the typology of scientific information sources (thesis, proceedings, etc.)	27	23	Get better

from other universities, even from other countries. Thus, cultural differences could be identified and analysed.

It would be also rewarding to analyse whether there are differences among first year and final year students.

Additionally, it would be interesting to analyze the self-efficacy that each student considers they have in each of the information abilities. By doing this, better training courses could be designed as they would better fit students' needs.

CONCLUSION

The concept of information literacy has been widely studied, especially during the last decade. Therefore, along literature many authors highlight the need of developing information abilities and skills in order to become an information-literate citizen. Before analysing whether students have these information abilities, it seems appropriate to analyse whether they are aware of their importance. If students are not aware of the relevance of information competences, they are not going to work in order to improve them.

Taking this into consideration, this study analysed the importance given by university students to a complete list of information abilities, concluding that students are not enough aware of their importance.

It was also concluded that, thanks to the information training course, the students' awareness was increased. This result highlights the need of developing transversal information courses that complement the specific modules of each degree. As it was said before, information literacy is a generic competence needed in every degree to complement the specific training of each professional branch. Sometimes these transversal courses do not receive the attention they deserve and the focus is on the specific modules. However, as it was also highlighted information literacy is a core competence needed to develop other competences. Therefore, some information courses

should be designed and included during the first undergraduate levels. These courses should be aimed at offering students information tools and techniques so they can be more efficient when learning other topics.

Overall, the authors consider that this research offers new sights on information literacy as it evaluates the awareness of students regarding information abilities. Nonetheless, due to the limited scope of the study, this study could be defined as exploratory and it would be interesting to replicate it in other universities and/or countries.

REFERENCES

- Alagumalai, S., Curtis, D. D., & Hnugi, N. (2005). *Applied Rasch Measurement: A Book of Examples*. Springer.
- American Library Association. (1989). *Final Report*. Chicago: American Library Association.
- ANECA. (2005). *Libro Blanco: Título de Grado en Economía y en Empresa*. Agencia Nacional de Evaluación de la Calidad y Acreditación.
- Behrens, S. (1994). A conceptual analysis and historical overview of information literacy. *College & Research Libraries*, (July): 320.
- Bond, F. G., & Fox, C. M. (2007). *Applying the Rasch Model. Fundamental Measurement in the Human Sciences* (2nd ed.). Lawrence Erlbaum Associates, Inc.
- Burnhein, R. (1992). Information literacy—A core competency. *Australian Academic & Research Libraries*, 23(4), 188–196. doi:10.1080/00048623.1992.10754796
- Grafstein, A. (2002). A discipline-based approach to information literacy. *Journal of Academic Librarianship*, 28(4), 197–204. doi:10.1016/S0099-1333(02)00283-5
- Klusek, L., & Bornstein, J. (2008). Information Literacy Skills for Business Careers. *Journal of Business & Finance Librarianship*, 11(4).

- Linacre, J. M. (2012a). Winsteps® Rasch measurement computer program User's Guide. Beaverton, OR: Winsteps.com
- Linacre, J. M. (2012b). Winsteps® Rasch measurement computer program. Beaverton, OR: Winsteps.com
- Martínez-Caro, E., & Cegarra-Navarro, J. G. (2012). El desarrollo de competencias transversales en dirección de operaciones mediante proyectos de emprendimiento. *Working Papers on Operations Management*, 3(2), 9-19.
- Marzo-Navarro, M., Pedraja-Iglesias, M., & Rivera-Torres, P. (2009). Curricular profile of university graduates versus business demands. Is there a fit or mismatch in Spain?. *Education + Training*, 51(1), 56-69.
- MEC. (2006). *Borrador de propuesta Directrices para la Elaboración de Títulos Universitarios de Grado y Máster*. Madrid: MEC.
- Owusu-Ansah, E. K. (2004). Information literacy and higher education. Placing the academic library in the center of a comprehensive solution. *Journal of Academic Librarianship*, 30(1), 3-16. doi:10.1016/j.jal.2003.11.002
- Pinto, M. (2010). Design of the IL-HUMASS survey on information literacy in higher education: A self-assessment approach. *Journal of Information Science*, 36(1), 86-103. doi:10.1177/0165551509351198
- Pinto, M., Doucet, A., & Fernández-Ramos, A. (2007). The role of information competences and skills in learning to abstract. *Journal of Information Science*, XX(X), 1-17.
- Rader, H. (1991). Information literacy: A revolution in the library. *RQ*, 31, 25-30.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Chicago: The University of Chicago Press.
- Sanchez, L., & Blanco, B. (2012). El Modelo de Rasch en Dirección de Operaciones. *Working Papers on Operations Management*, 3(2), 35-47.
- Sanchez-Ruiz, L., Pérez-Pérez, M., Blanco-Rojo, B., & Serrano-Bedia, A. M. (2013). The Operations Management Area in the EHEA: Skills developed. University of Cantabria case Study. *Proceedings of ICERI2013*, 6787-6794.
- Von Davier, M., & Carstensen, C. H. (2007). *Multivariate and Mixture Distribution Rasch Models. Extensions and Applications*. New York: Springer Science + Business Media, LLC. doi:10.1007/978-0-387-49839-3
- Webber, S., & Johnston, B. (2003). *Information literacy: Definitions and models*. Academic Press.
- Webster, F. (2006). *Theories of the Information Society* (3rd ed.). Taylor & Francis e-Library.
- Zabel, D. (2004). A reaction to Information literacy and higher education. *Journal of Academic Librarianship*, 30(1), 17-21. doi:10.1016/j.jal.2003.11.003

ADDITIONAL READING

- Mutch, A. (1997). Information literacy: An exploration. *International Journal of Information Management*, 17(5), 377-386. doi:10.1016/S0268-4012(97)00017-0
- Saranto, K., & Hovenga, E. J. S. (2004). Information literacy-what is it about? Literature review of the concept and the context. *International Journal of Medical Informatics*, 73, 503-513. doi:10.1016/j.ijmedinf.2004.03.002 PMID:15171979
- Villa, A., & Poblete, M. (2008). *Aprendizaje basado en competencias. Una propuesta para la evaluación de las competencias genéricas*. 2ª edn. Bilbao, Spain: Ediciones Mensajero.

KEY TERMS AND DEFINITIONS

Competence: The combination of knowledge, skills (intellectual, manual, social...), attitudes and values that will enable the graduate to successfully tackle the problem solving or intervention in an issue in an academic, professional or social context (MEC 2006).

EHEA: European Higher Education Area.

IL-HUMASS Questionnaire: Self-assessment tool created to measure information literacy development among university students.

Information Communication and Dissemination: One of the factors that integrate the information literacy competence. It is about being able of transferring (orally or written) all the information, including the use of digital media.

Information evaluation: One of the factors that integrate the information literacy competence. It is referred to the ability of analyzing whether a piece of information is reliable, update.

Information Literacy: It entails the learning of the skills, competences, knowledge and value to access, use and communicate information in any of its forms, in order to produce competent professionals trained in the routines of identifying, evaluating and recording information sources appropriately and with the knowledge to process and produce their own information (Pinto et al., 2007).

Information Processing: One of the factors that integrate the information literacy competence. It is about being able of properly treating the information. Not only is it about understanding and using it, but it is also about correctly citing and referencing the sources used.

Information Search: One of the factors that integrate the information literacy competence. It includes all the skills and knowledge needed to find reliable information: searching techniques, information source.

An Integrated Electronic IQA System for HEI



Teay Shawyun

King Saud University, Saudi Arabia

INTRODUCTION

Regardless of what academics believe or like, with the depletion of public funding and increased competition, the reality is that higher education institutions' (HEI) future is more business oriented potentially affecting educational quality (Haworth & Conrad, 1997, Bowden & Marton, 1998). This over-commercialization and internalization to achieve individual "economic" needs rather than meeting the public needs and concerns is an issue that affects strong and sustained academic performance excellence. Yin, *et al.* (2002) noted that the education policy needs to include economic viewpoints that highlighted the needs for the institution to change the internal educational structures and systems to meet different educational purposes and aspirations by identifying, procuring and allocating appropriate resources for inputs that enhances the efficiency of internal processes of the system and its sub-systems to meet the short-term and/or long-term education needs. Conti (2006) also emphasized the need of understanding the quality management from the systems perspective by extending the quality management concepts of economic transactions to social relations that ultimately creates value to the stakeholders.

HEIs have a responsibility to the society to develop the future societal human capital through its educational value that they propose to the stakeholders through its internal value creation processes. The central issue is what and how these internal processes are aligned to create this educational value proposed to the stakeholders.

BACKGROUND OF STUDY

External Quality Assurance (EQA) as represented by any accreditation requirements has been prescribed voluntarily or mandated where all HEIs subscribe to or are mandatorily coerced into (Wells, 2014). While EQA hypes the protection of stakeholders' values in learning outcomes and competencies, the accompanying and complementary Internal Quality Assurance (IQA) struggles to keep up with the pace of EQA progress and requirements (Kettunen, 2012). Systems and mechanisms in accomplishing and achievement EQA requirements falters or are sidetracked due to the mundane IQA with volumes of documents, reports, statistics and evidences requirements of quality measures of processes and results underlying education value (Prikulis, Raugvargers, & Rusakova, 2011).

In attempting to resolve this issue, this chapter explores the key components of quality, information and planning underpinning education excellence to align the integration of the 3 main IQA core systems of quality management (QM), information management (IM) and planning management (PM) as these encompass most aspects of the creation and delivery of the educational value of HEIs. The proposed integrated eIQA is strategically and tactically aligned top-down and bottom-up where quality management is aligned with planning management via information management leading to informed decisions affecting quality management at all levels of the institution, administrative units, colleges and programs. To illustrate these QM-IM-PM linkages via its key

integrated e-modules, a case study of a leading university in the Middle East demonstrates this strategic integrated eIQA system.

CONCEPTUALIZATION OF AN INTEGRATED eIQA IN HEI

Arcaro (1995) stated that “Quality is creating an environment where educators, parents, government officials, community representatives, and business leaders work together to provide students with the resources they need to meet current and future academic, business, and societal needs”. This would mean that comprehensive systems and mechanisms to enrich and enhance these quality environments, albeit an integrated one or hybrids (Marsh, 1995), are an imperative for improving education through quality management (Tribus, 1993). These policies and procedures include faculty and personnel selection, motivation, development, engagement and retention; learner entry, guidance and student service and support systems; management information systems; student assessment; management of developing, delivering and evaluating learning programs and resources; availability and allocation of financial, administrative and physical resources; governance & administration; planning, monitoring and management of all educational performance measurement and management. All these allude to the ISO 8402 definition of a quality management system (QMS) as “the organizational structure, processes, procedures and resources needed to implement quality management”.

Goldberg and Cole (2002) identified three levels of education quality management application as 1) the management processes of HEI; 2) teaching quality to students who are both customers and workers in the educational system; with 3) highest level of quality in learning to achieve the desired results. These meant that educators must question their core teaching and learning processes, teaching pedagogy & strategies, and assessment methods. They contended that the traditional “institutionalized rationale” be replaced

by a systemic quality approach as espoused by Deming (1994) that “a system is a network of interdependent components that work together to try to accomplish the aim of the system.” The change to a systems perspective where the change is more systemic, the more the institution embodies changes in behaviors, culture, and structure, strategies, philosophies, processes, people, systems and mechanisms, the more lasting the change will be which is in line with Deming’s seventh principle resemblance of Senge’s (2006) systems thinking.

These systemic and the systematic processes where activities interact to achieve defined goals resulting in the quality of the product and services, constituting the QMS, is defined by DTI (2015) as “A set of coordinated activities to direct and control an organization in order to continually improve the effectiveness and efficiency of its performance”. This QMS enables an institution to provide consistency and coherence in the interactions and integrations of all education activities and processes of the institution’s goals and objectives achievements of its policies and strategies. This begins with the identification of stakeholders’ (students & parents, faculty & staff, regulatory bodies & society) requirements resulting with their satisfaction at every transaction interface. The QMS can be visualized as a “wedge” that both clasps the improvements, innovations and achievements along the quality journey, and prevents good practices from slipping back. Rouse & Putterill (2003) identified the QMS performance triplet as: performance measures, performance analysis and performance evaluation that corresponds to Altman’s (1979) three component of evaluation, data analysis and performance measures underpinning the QMS architecture.

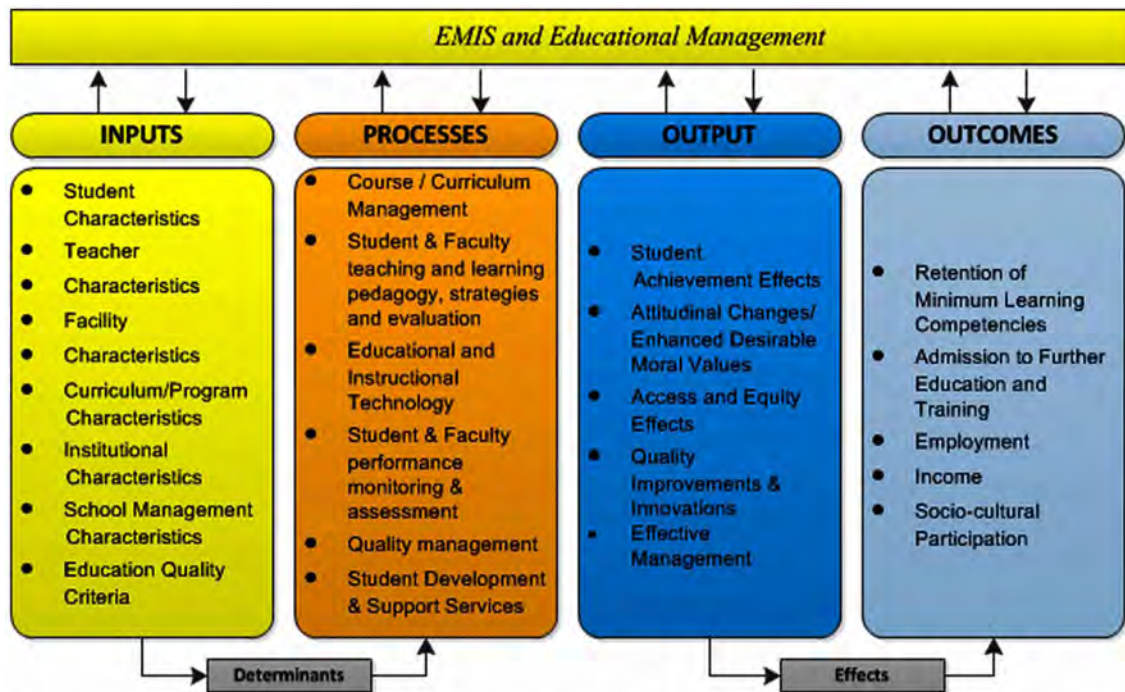
An effective and efficient integrated QMS supports strategic directions with tactical implementations meeting stakeholders’ requirements and expectations; bringing about efficient and effective educational processes to create and deliver on educational value; reducing wastages and lower costs through activity based costing of the educational costs in deliverance of education values; and ultimately increasing morale and

motivations to strive for performance excellence and morals towards conscientious education benefiting the students. Hence, designing and building the integrated QMS includes the QMS infrastructure, the philosophies, plans, processes, procedures, people, practices and its deployment with performance measured and assessed bringing about continuous improvements and informed decision making paving the way towards performance excellence. The crux is the focus of improvement efforts in a holistic systemic and systematic approach to education QM. On the electronic dimension, the UN advocated the EMIS (Electronic Management Information System) that sought for the synergy and integration in terms of eight main aspects or dimensions: needs of producers and users; data; information handling; storage of data; retrieval of data; data analysis; computer and manual procedures; networking among users. This includes the data warehouse, the data marts and BI (Business Intelligence) tools, and its mire of networks. Arbogast (2009)

advocated that the “QMS be electronic, linked, or related so that the piece parts can be analyzed; be dynamic, when change is required and capable of being analyzed and changed quickly; maintains history from a corrective action management perspective, and is a real-time managed system published electronically to the total organization that adds value and providing speed and agility on a long-term basis”.

This calls for great efforts to be dedicated to the rationalization and strengthening of existing structures and processes and to improvement in the coordination of information flows and organizational commitment and an investment in real time online applications and database systems. All activities in terms of IPOO (Inputs, Processes, Outputs and Outcomes) in the handling of educational management information together with their performance measure and indicators are to be placed and coordinated within the same EMIS framework from-all-to-all internal and external sources and stakeholders (Figure 1) that includes

Figure 1.



Adapted from: Villanueva, C.C., 2003, Education Management Information System (EMIS) and the formulation of Education For All (EFA) Plan of Action, 2002-2015, U N E S C O Almaty Cluster Office and the Ministry of Education of Tajikistan

an eDRMS (electronic Document and Records Management System) (Queensland Government, 2010).

The bottom line is that the QMS, and eDRMS must be integrated electronically into an efficient and effective HEI eMIS quality and planning system of which this paper aims to discuss and demonstrate through the case study.

CASE STUDY OF KSU ELECTRONIC INTEGRATION OF QMS-IMS-PMS

The case study of King Saud University (KSU), a ranking university and leading university in the Middle East, will be used as the main illustration of the integrated eIQA system. A key question for QM is what takes place in between the 5/7 years accreditation cycle to sustain planning and continuous improvements. This is addressed through internal audits and assessments in between the accreditation cycles to assure continuous improvements and innovations. This can be an annual or bi-annual exercise to provide for time for planning

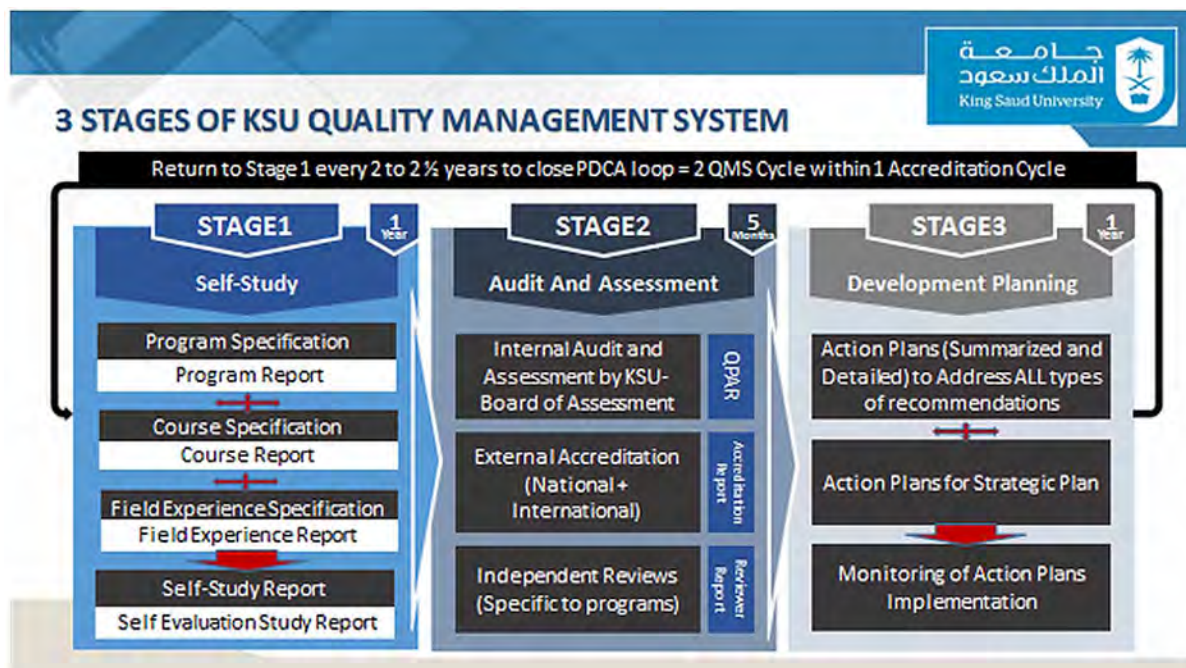
of actions with evidence of implementation bringing to fruits these plans which are monitored and measured for performance.

To manage KSU's IQA, a 3 staged KSU-QMS approach was developed (Figure 2) comprising of:

Stage 1 Self-Study: This comprises the tedious and mundane documentation required for QM and accreditation, which is the bane of equating quality to just documentations. The electronic part of the Course Specifications (CS) & Course Reports (CR) and Field Experience Specifications (FES) & Field Experience Report (FER) done on a semester basis culminates in the annual Program Report (PR) based on the Program Specification (PS). All these supports the development of the program Self-Study Report (SSR) and Self Evaluation Study Report (SESR) required for accreditation.

Stage 2 Internal Audit and Assessment (IAA): The bi-annual monitoring cycle and IAA of the program are done by an independent internally appointed Board of Assessors

Figure 2. Three Stages of KSU Quality Management System



(BOA) before the programs go for their every 5/7 years mandatory national accreditation. The main monitoring normally takes place at the core of the educational processes which is represented by the colleges and their program offerings. Its aim is to ensure that the periods in between the accreditation cycle strengthen or sustain improvements.

Stage 3 Developmental Planning: The aim of eIQA is to ensure continuous improvements with evidence substantiating these improvements meeting the vision, mission and goals as committed to the stakeholders and as planned in the strategic plan of the college or programs. The recommendations of the accreditation and IAA exercises leading to action plans and projects planned for addressing these recommendations are the key link to the strategic plan as these will address all the QMS and accreditation criteria which are in line with the overall strategic plan. This will avoid redundant planning and inconsistencies or incoherence across the quality-planning duo. As such, the bi-annual monitoring process of the developmental planning is aimed at capturing and closing the quality feedback loop to ensure that the quality drive is maintained and sustained through continuous improvements and actions from one accreditation cycle to another.

All these 3 stages are supported by the SID (Statistics, Information and Documents) system component in the KSU Data warehouse (Figure 3) of the “ITQAN” System part of the eIQA. The imperative is a set of corresponding statistics, information or documents that supports the assessment that is evidence-based. Key evidences are needed to substantiate and support the accomplishment and achievement of quality actions that had been planned, implemented, monitored for accomplishment and measured for achievements with comparisons to the previous years and to the industry benchmark. The same logic applies to key performance indicators (KPI) which are based

on the statistics & data. In effect, the institution/ college/programs/administrative units need to set up a system or mechanism to collate and analyze the statistics and information.

H

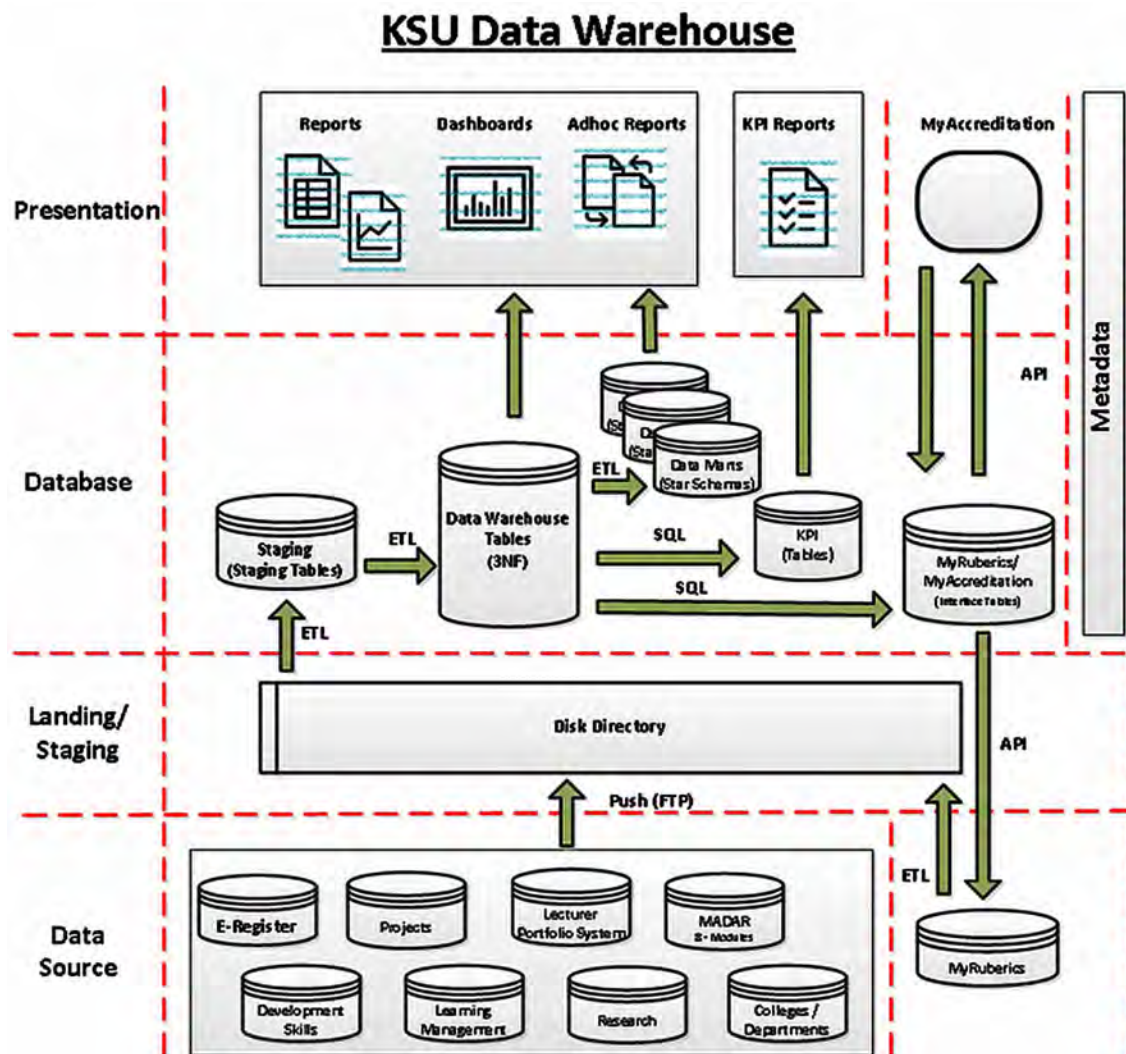
KSU ITQAN SYSTEM OF THE INTEGRATED eIQA SYSTEM

Figure 4 shows the re-conceptualized and re-configured integration of sub-systems components working in tandem and holistically of the KSU eIQA. These interwoven parts show key data sources inputs from various systems of surveys, performance metrics, eRegister, MADAR-HR, finance, learning resources, facilities & infrastructure, research and community services into the SID data mart. All these are used by the faculty to create their CS & CR and FES & FER, program chairs to create their PS & PR, program quality committee to develop their SSR & SESR and the Internal Benchmarking, all of which feeds into the IAA and accreditation components. These will culminate in the recommendations of assessors and Quality Performance and Assessment Report (QPAR) which are input to the developmental planning component linking the strategic plan's action plans and projects. These leads to consistent and coherent planning based on quality management inputs via its integrated data marts in the DWH, electronic information networks culminating in informed decision making efficiencies and effectiveness.

The integrated eIQA (Figure 5) linking and supporting the QM-IM-PM dimensions in the ITQAN system are composed of the following modules:

- **eSurvey Module:** The minimum 7 surveys of Course Satisfaction / Student Experience / Program Satisfaction / Faculty & Staff Satisfaction / Alumni & Employment Market Satisfaction, conducted on a semester and annual basis measuring stake-

Figure 3. KSU Data Warehouse (DWH) in the ITQAN System



holders' satisfaction, are key data inputs to the performance metrics module.

- **ePerformance Metrics Module:** This module handles all computation, aggregation and compilations of the 17 Strategic KPIs and 56 QMS KPIs inclusive of the 33 NCAAA (National Commission of Academic Accreditation and Assessment) KPIs measuring planning and quality performance and benchmarked in the benchmarking module where internal & external benchmarks are compared.

- **eQMS-Accreditation Module:** The core of the eQMS's QM is based on the multifarious NCAAA forms and templates requirements to be uploaded online to the Accreditation Management System (AMS) of NCAAA. In addition, this is supported by Curriculum Mapping Module where the Student Learning Outcomes, the teaching methodologies and students' assessment are mapped according to Program Objectives and Courses & Program Learning Outcomes (CLO/PLO) within the CLO/PLO Matrix. These include the key

Figure 4. KSU eIQA (internal quality assurance) conceptual mapping

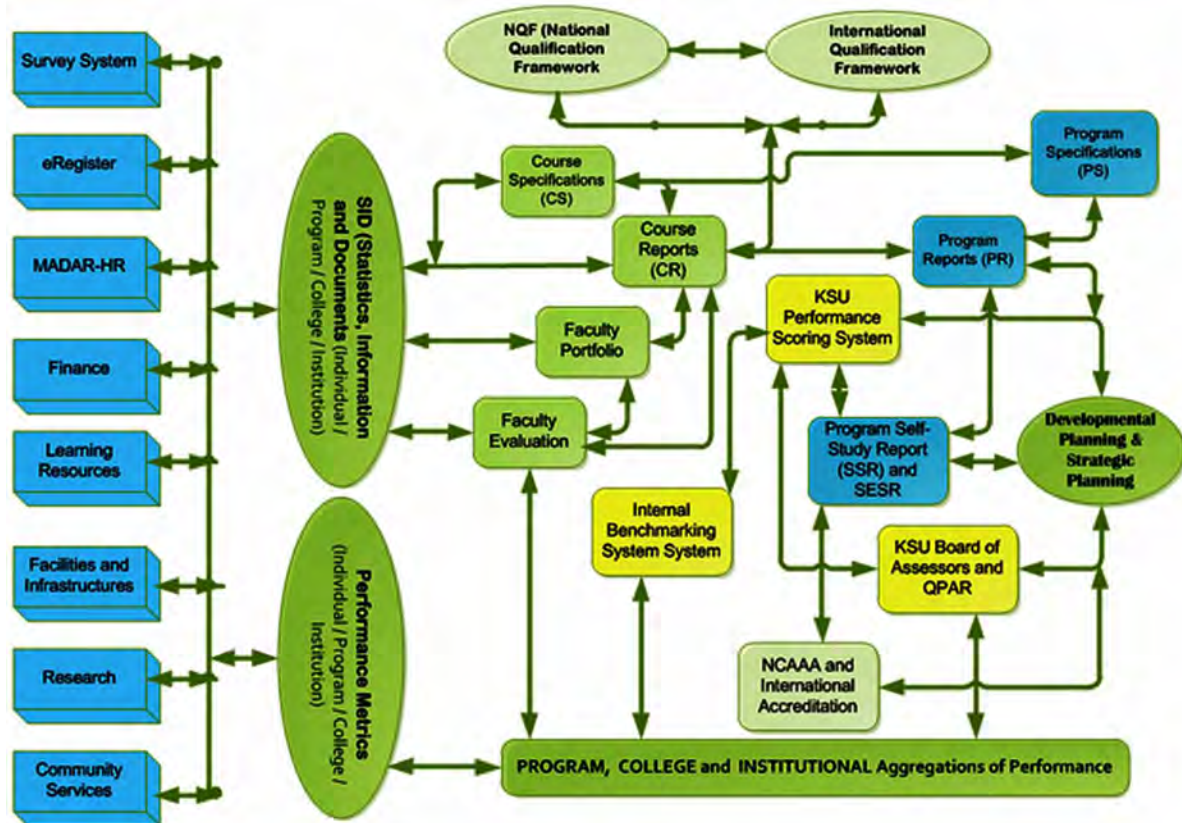
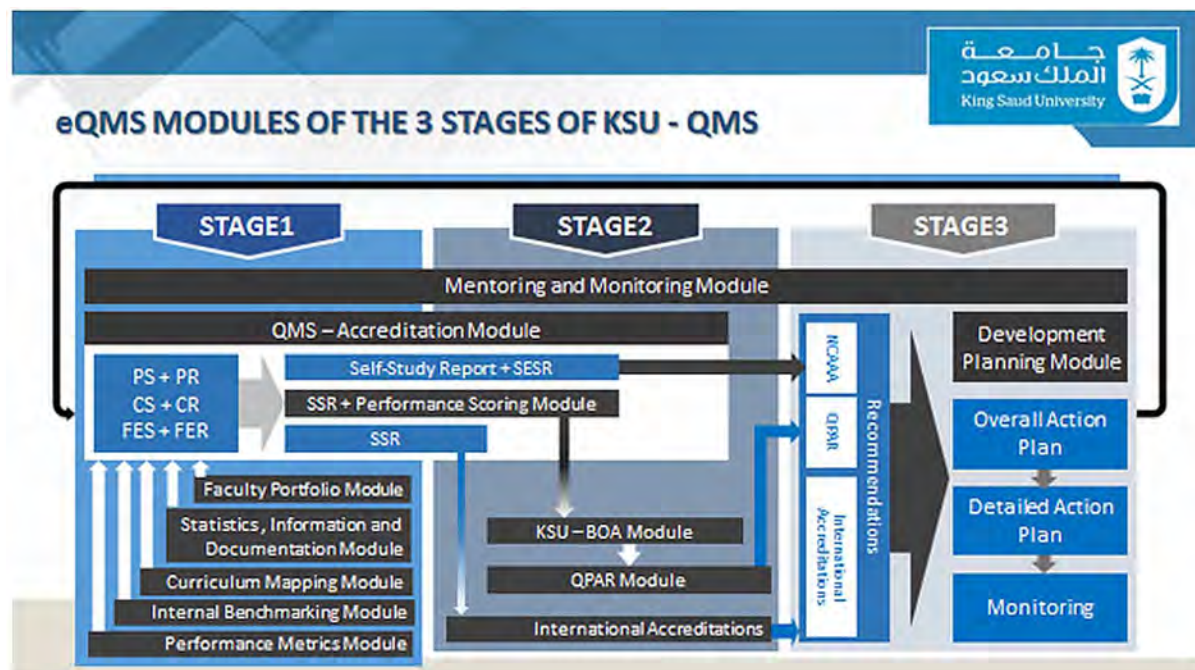


Figure 5. Integrated and Electronic conceptualization of KSU eIQA and its modules



modules used by all the colleges, programs and faculty members to manage their CS & CR, PS & PR, FES & FER and SSR & SESR which are monitored for performance in key dashboards aggregated from the course levels up to the institution level.

- **eKSU-BOA Module:** The computerization help lighten the loads of the BOA and the programs in the IAA exercises of the 350 undergraduate and post graduate programs by the BOA once every 2 and half years for each program that requires substantial time and efforts of performance analysis, identifying strengths and areas for improvements, site visits issues based on the hefty evidence needed of the programs audit and assessment.
- **ePerformances Scoring Module:** The performance of the programs is based on the Performance scoring system with integration into the QPAR.
- **ePlanning's:** two sub-modules of:
 - **Planning Module:** where the strategic plan of the university's vision, mission, goals, objectives are defined and cascaded to the colleges' & administrative units' vision, mission, goals, objectives for alignment and linkages with the program objectives and action planning and the 4 perspectives of the Balanced Scorecard (BSC) of the KSU's 9 objectives. This will result in action plans and projects linked to the quality action plans and objectives subscribing to the same goals and objectives.
 - **Monitoring and Mentoring Module:** where KSU assigns mentors to the colleges to support them in the PM-QM, while at the same time monitoring the progress of the planning and quality action plans' accomplishment and achievements.
- **eFaculty & eStudent Portfolios Module:** To avoid duplications of data entry to 4 different systems of research, faculty portfolio

and evaluation, promotion and "FAC" web page, all the common data pertaining to the faculty work, responsibilities and performance are entered through a singular portal in this module. These are then channeled to each of the 4 systems for their specific usage, and this will cut down on the same data entry and requirement of 4 different systems. The efaculty or eStudent portfolios have a key purpose of formative and summative assessment of the student and his/her performance (Thompson & Cobb-Reiley, 2003; Barrett, 2004) for personal reflection and learning management. They have benefits of the tracking and documentation of teaching and learning experiences; encouraging the reconnection between learning process and outputs & outcomes; institutionalizing collaboration, reflection, and discussion. The outcomes of the student assessment and performance are integrated and customized to individual student's needs and requirements. The aim is for each student to integrate his/her study, social and extra curricula and personal requirements to better customize his/her formative and summative academic, social and personal workspace and work life.

- **eSID Module:** This is the main data mart for data, information statistics, documents of the ITQAN application of the different modules described above supported by the KSU DWH and data marts for key applications utilization.

SAMPLES OF KEY MODULES SCREENSHOTS OF ITQAN APPLICATION

Basically, the dashboard is one of the most important feature in the ITQAN system used for performance monitoring allowing drill-downs from the highest levels of institution to the program levels. Since the two main ITQAN's nuclei are

QM-IM-PM, there are 3 main sets of dashboards of 1) eQMS-Accreditation; 2) ePerformance Metrics and 3) eDevelopmental Planning & BSC. The dashboard (Figure 6) shows the key category of the university/colleges/programs degree of accomplishment in %, of the key areas of accreditation/program/course managements of the selected college/program (Figure 7) down to the level of all courses in each program (Figure 8). This allows the college/program or central eIQa management unit to have oversight and overview of the performance of all the colleges down to the levels of the individual courses accomplishment of the key CS & CR (Figure 9); FES & FER and the PS & PR; and ultimately the SSR and SESR progress towards IAA & accreditation (Figure 10).

A required QM best practice and accreditation for the curriculum mapping (Figure 11) is the definition & design of PLOs (Figure 12) and CLOs, the mapping of the CLO to the PLO based on the program objectives in the PLO/CLO mapping matrix Figure 13. At the course level, the assessment methods for the CLO for each of the student is shown, all of which are reported

as part of the course student assessment rubric. There is also a facility for the assessment of all the students' performance in a course (Figure 14) leading to the reporting of the competency index for that course/student (Figure 15) and performance metrics (Figure 16).

IMPLICATIONS OF THE INTEGRATED eIQa QUALITY-INFORMATION-PLANNING SYSTEMS

In the development of the integrated eIQa, it is considered a complete business process redesign that combine separate education processes into a singular core value process (Ashworth, 1999; Childe *et al.*, 1994; CIM-OSA Committee, 1989) cutting across different owners and domains as opposed to fragmented collections of individuals and small scale systems (Tovey, 1991; Mason, 1993), with the output directed to the internal and external stakeholders based on their strategic needs and satisfactions (Temporal, 1990; Bolt, 1993; Burach *et al.*, 1997).

Figure 6. eQMS-Accreditation Dashboard for overall Institution and Colleges

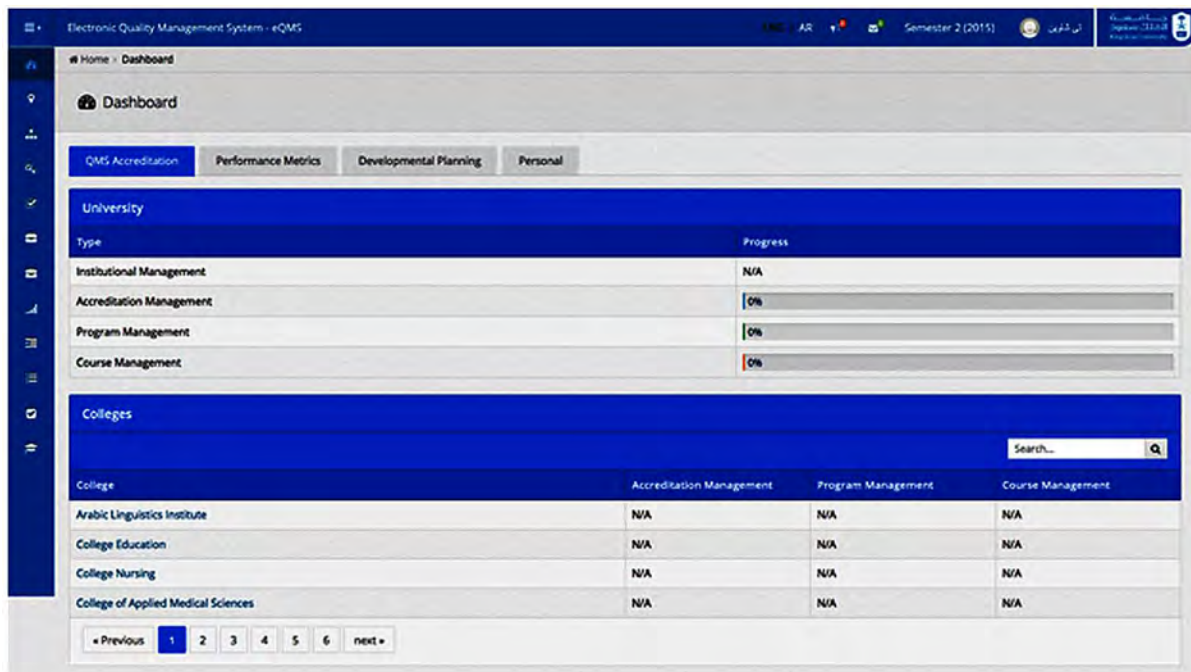


Figure 7. eQMS-Accreditation Dashboard drill-down to program

Colleges			
Search...			
College	Accreditation Management	Program Management	Course Management
College of Applied Studies and Community Service	N/A	N/A	N/A
College of Architecture and Planning	N/A	N/A	N/A
College of arts	N/A	N/A	N/A
College of Business Administration	N/A	N/A	N/A
Previous 1 2 3 4 5 6 next			
Programs			
Search...			
Program	Accreditation Management	Program Management	Course Management
Bachelor of Accounting	0%	0%	0%
Bachelor of programming and Database	0%	0%	0%
Bachelor of Marketing	0%	0%	0%
Bachelor of banking	0%	0%	0%
Previous 1 2 next			

Figure 8. eQMS-Accreditation Dashboard for details of a program

Programs

Search...

Program	Accreditation Management	Program Management	Course Management
Bachelor of Accounting	0%	0%	0%
Bachelor of programming and Database	0%	0%	0%
Bachelor of Marketing	0%	0%	0%
Bachelor of banking	0%	0%	0%

[Previous](#)
[1](#)
[2](#)
[next](#)

Selected Program

Program : Bachelor of banking

Program Management	Progress
Program Specifications	0%
Annual Program Report	0%
Course Management	Progress
Course Specification	0%
Course Report	0%
Field Experience Specification	0%
Field Experience Report	0%

Course Management

Search...

Course	Type	Progress
Business Process Management	Course Specifications	N/A
	Course Report	N/A
Capstone Project-I	Course Specifications	N/A
	Course Report	N/A
Capstone Project-II	Course Specifications	N/A
	Course Report	N/A
DATA COMMUNICATIONS AND COMPUTER NETWORKS	Course Specifications	N/A
	Course Report	N/A

[Previous](#)
[1](#)
[2](#)
[3](#)
[4](#)
[5](#)
[6](#)
[7](#)
[next](#)

Figure 9. Course specifications and course reports dashboard



As noted by Franco-Santos *et al.*, (2007), the two key characteristics of a strategic performance management system (SPMS) of “performance measures” and “supporting infrastructure” and the 3 key processes of: “information provision”, “measure design and selection” and “data capture” were incorporated into the architectural design of the integrated eIQa linking the quality-information-planning dimensions (Teay, 2012). The integrated eIQa, part of a SPMS provides measures of performance, progress and opportunities to refine or improve on the “what, why and how” mechanisms.

FUTURE RESEARCH DIRECTIONS

While this chapter discusses a basic integrated eIQa infrastructure that supports the QM-IM-PM in any HEI, there are vast potentials in enhancing this key enabler in:

- Predictive modeling of the vast data to support education predictive analytics (AAEA, 2014 and Tagtow, 2014) and Institutional Research Intelligence (Djunaidi, 2012) that goes beyond the typical institutional research and SPMS dashboards for HEIs.

Figure 10. eSelf-Study Report Dashboard

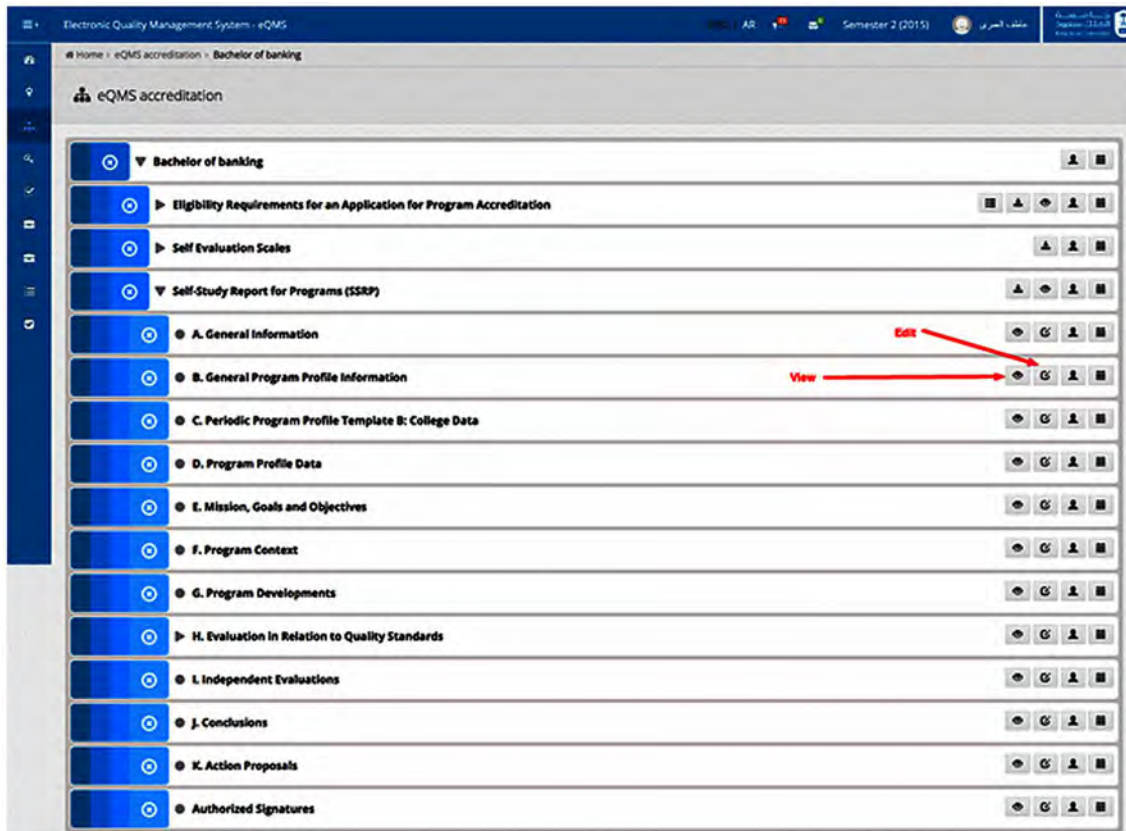


Figure 11. eCurriculum mapping of the objectives, leaning outcomes & assessment

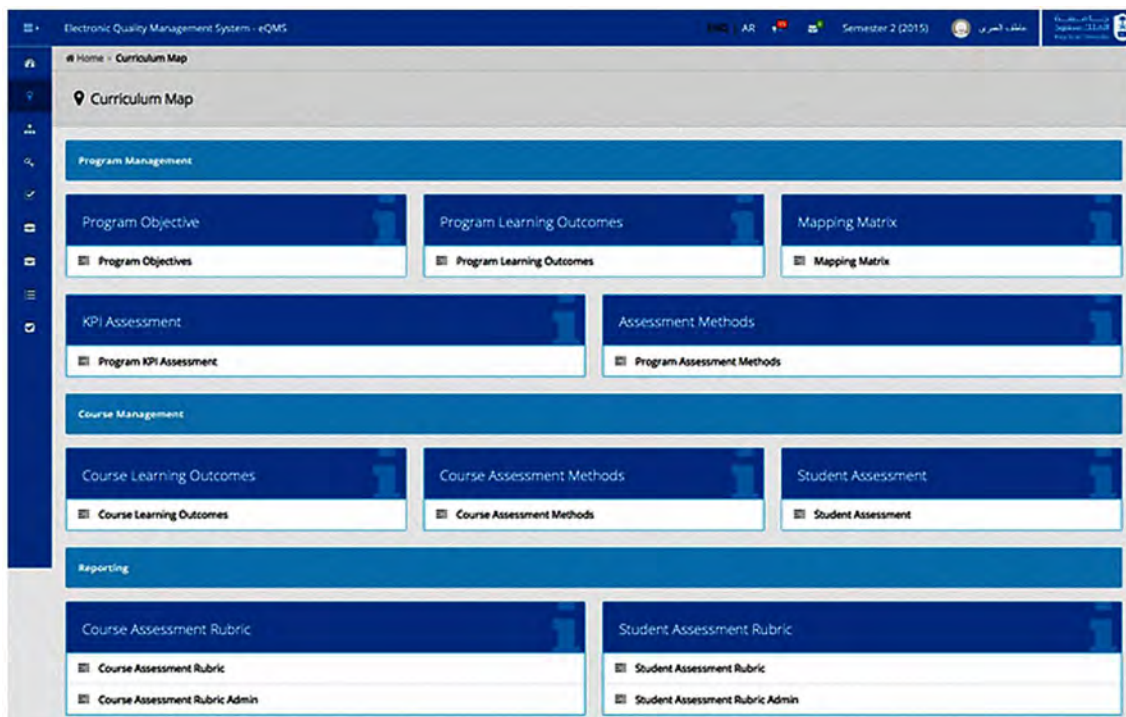


Figure 12. eProgram learning outcomes

Electronic Quality Management System - eQMS

Home > Curriculum Mapping > Program Learning Outcomes

Program Learning Outcome

Set the program learning outcomes (PLOs) → Update

#	Title	Actions
1	Knowledge	
1.1	Describe the contents and scope of the subject.	✓ Edit Delete
2	Cognitive Skills	
2.1	Identify information, formulate questions, refine questions through answers and identify issue(s).	✓ Edit Delete
3	Critical & Analytical Skills	
4	Interpersonal & Entrepreneurial Skills	
5	Professional Ethics, Morals & Values	
6	Information Literacy	
7	Communication Skills	
8	Psychomotor Skills	

Figure 13. ePrograms and eCourse learning outcomes mapping matrix

Electronic Quality Management System - eQMS

Home > Curriculum Mapping > Mapping Matrix

Program Learning Outcome Mapping Matrix

Set the PLO mapping matrix with the courses (the matrix should be for 2000 level - 2000 years)

Level 2 | **Level 3** | Level 4

#	Program Learning Outcome	IS 370	IS 335	IS 351	IS 230	IS 362	IS 324	IS 224
1	Knowledge							
1.1	Describe the contents and scope of the subject.	N/A	N/A	N/A	Proficient	N/A	N/A	N/A
2	Cognitive Skills							
2.1	Identify information, formulate questions, refine questions through answers and identify issue(s).	N/A	N/A	N/A	N/A	N/A	N/A	N/A
3	Critical & Analytical Skills							
4	Interpersonal & Entrepreneurial Skills							
5	Professional Ethics, Morals & Values							
6	Information Literacy							
7	Communication Skills							
8	Psychomotor Skills							

Introduction
Proficient
Advanced

Save

Figure 14. eStudents assessment for a sample course

Electronic Quality Management System - eQMS

Home > Students Assessment

Students Assessment

1. Select the course

2. Select the Assessment Method

3. Put the marks for each student

Courses

INFORMATION SYSTEMS ANALYSIS & DESIGN - 22246

Assessment

Quiz

Student	Mark	Knowledge	Cognitive Skills
Student 1			
Student 2			
Student 3			
Student 4			
Student 5			
Student 6			
Student 7			

Figure 15. eCompetency index of the assessment rubric for a course

Electronic Quality Management System - eQMS

Home > Curriculum Mapping > Assessment Rubric

Assessment Rubric

Select the Course

IS 240 INFORMATION SYSTEMS ANALYSIS & DESIGN Assessment Rubric

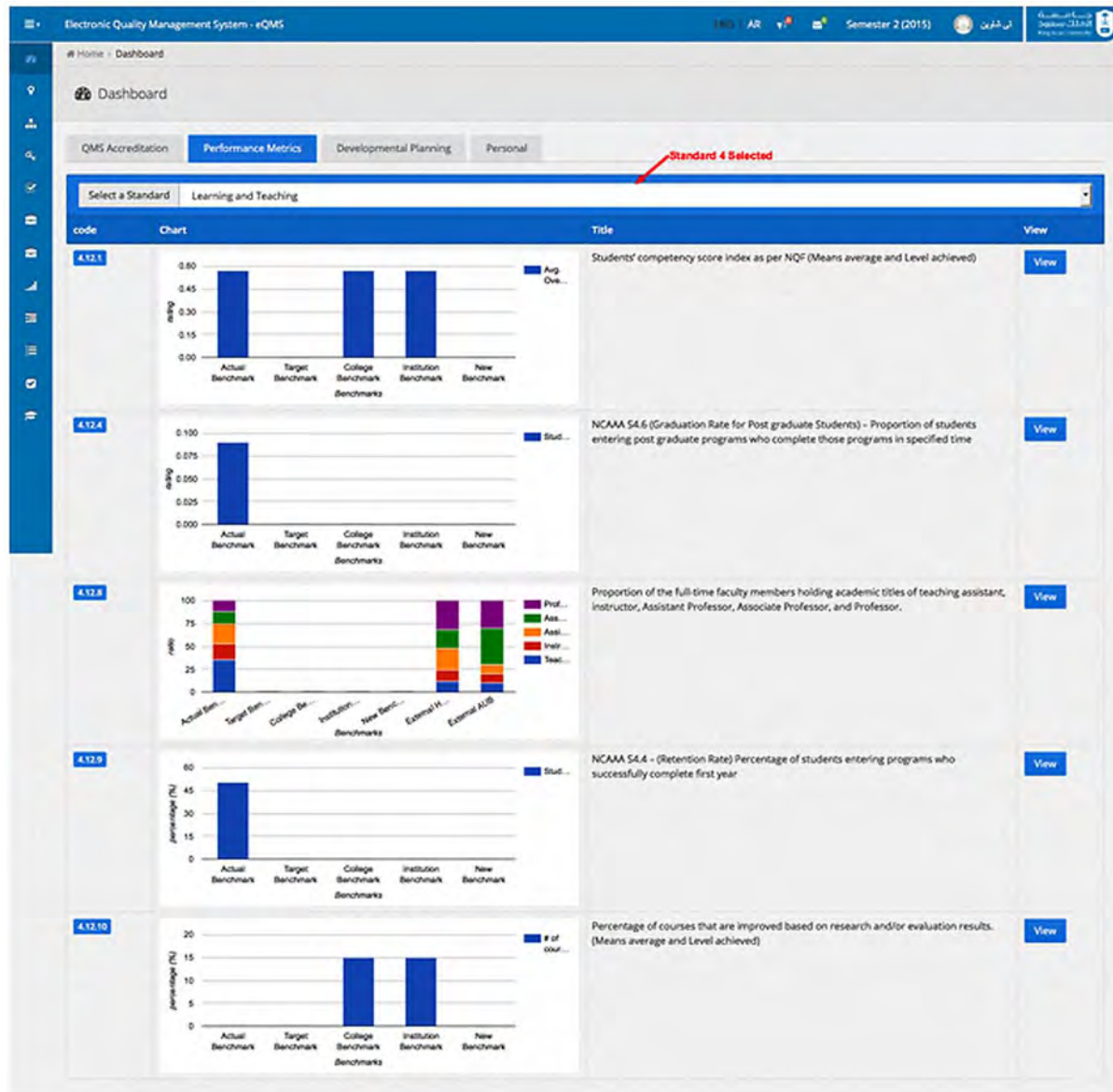
Print PDF Print PNG

Course title	Course Code	Competencies Dimension	Quiz	Examination - Final	Overall Band Performance
INFORMATION SYSTEMS ANALYSIS & DESIGN	IS 240	Knowledge	0.4	0.23	0.32
		Cognitive Skills	0.24	0.12	0.18
Overall Band Performance based on SLOs			0.32	0.17	0.25

Courses

INFORMATION SYSTEMS ANALYSIS & DESIGN-22246

Figure 16. ePerformance metrics dashboard



- Linkages directly to the national EQA for simplifying and supplementing international accreditation agencies requirements which can be extracted and modified from the core institution's QM-IM-PM.
- Drill downs to the individual student performance management and support focused on a more student-centric approach, with potentials of linking to faculty performance evaluation and management based on its teaching-learning-research database in the system.
- Improved overall SPMS of the HEI, colleges, programs and stakeholders.

CONCLUSION

This paper and the illustrated case suggest that the fundamental “piece-meals” flaws that had

always been obstacles in effective application of ICT for HEIs' educational aims and mission could be reconciled by:

- Bringing about an integrative approach of the strategic and education management with IS/IT as the enabler for QM-IM-PM integration, that are aligned to bring about the convergence of 3 very different and potentially unrelated disciplines. This could be done through the expert knowledge of the champion or change agent who must have knowledge and skills of the strategic needs of the HEI, the operationalization of the technicalities of ICT and details of the tedious and overwhelming quality management and accreditation requirements.
- Moving from the macro organizational strategic needs to the micro level operational or functional processes requirement of a new mind-set that calls on the capability and capacity of all the individuals and the organization as one to move in the same strategic direction for holistic aligned accomplishments and achievements. The key factor is in the operationalization of the micro aspect of the macro overview that underlies the precarious of internal processes and the human capital re-engineering and re-discovery.
- Linking and integrating strategically and tactically the core value addition processes of teaching-learning-research to the QM-IM-PM trio with ICT as the supporting enablers through a set of strategic and operational plans that forms the integration and electronic triangularization of the quality-information-planning approach as expounded.

In conclusion, even-though this paper emphasized on the strategic integration of the quality-information-planning trilogy, this does not mean that other concepts and frameworks are irrelevant. On the contrary, as noted by Andersen *et al.*

(2006), the quality, information and planning management, or for that matter, all aspects of the HEIs' commitment of educational value to society must be approached from a holistic perspective through sets of appropriate plethora of tools & techniques, systems & mechanisms depending on the situational and/or institutional requirements meeting and going beyond stakeholders' needs.

REFERENCES

- Association of American Education Analytics (AAEA), Center for Education Analytics (CEA) and Institutional Research Intelligence (IRI). (2014). Transforming US Colleges and Education System to be more Competitive in the New World Economy. *International Journal of E-Adoption*, 1(1).
- Andersen, B., Henriksen, B., & Aarseth, W. (2006). Holistic performance management: An integrated framework. *International Journal of Productivity and Performance Management*, 55(1), 61–78. doi:10.1108/17410400610635507
- Altman, S. (1979). Performance monitoring systems for public managers. *Public Administration Review*, 39(1), 31–35. doi:10.2307/3110375
- Arbogast, S. (2009). *Electronic Quality Management Systems Are worth the Money*. Retrieved from <http://www.qualitydigest.com/inside/quality-insider-article/electronic-quality-management-systems-are-worth-money.html>
- Ashworth, G. (1999). *Delivering Shareholder Value through Integrated Performance Management*. Prentice Hall.
- Barrett, H. C. (2004). Differentiating Electronic Portfolios and Online Assessment Management Systems. *Proceedings of SITE Conference*.
- Bolt, J. (1993). Achieving the CEO's agenda; education for executives. *Management Review*, 44–49.

- Bowden, J., & Marton, F. (1998). *The University of Learning – beyond quality and Competence in Higher Education*. London: Kogan Page.
- Burach, E. (1997). The new management development paradigm. *Human Resource Planning*, 20(1), 14–21.
- Chide, S. J., Maull, R. S., & Bennett, J. (1994). Frameworks for understanding business re-engineering. *International Journal of Operations & Production Management*, 14(12), 22–34. doi:10.1108/01443579410072346
- CIM-OSA Standards Committee. (1989). *CIM-OSA Reference Architecture*. Brussels: AMICE ESPRIT.
- Conti, T. (2006). Quality thinking and systems thinking. *The TQM Magazine*, 3(3), 297–308. doi:10.1108/09544780610660013
- De Bruyn, P. P., Erasmus, M., Janson, C. A., Mentz, P. J., Steyn, S. C., Theron, A. M. C.,... Xaba, M. I. (2007). *Schools as Organizations* (3rd ed.). Pretoria, South Africa: Van Schaik Publishers.
- Deming, W. E. (1994). *The new economics for industry, government, and education* (2nd ed.). Cambridge, MA: Massachusetts Institute of Technology.
- Department of Trade and Industry. (2015). *Quality Management System*. Retrieved from www.dti.gov.uk/quality/qms
- Djunaidi, H. (2012). Institutional Research Intelligence (IRI): Go Beyond Reporting. *Proceedings of South Central SAS Users Group Annual Conference*.
- Education Labor Relations Council (ELRC). (2003). *Policy Handbook for Educators*. Pretoria: Universal Print Group.
- FASSET. (2000). *Implementing Quality Management Systems, EU/DoL/GTZ P7 Task Team: Quality Management System for Setas and ETQAs*. Retrieved from www.saqqa.org.za
- Franco-Santos, M., Kennerly, M., Michelli, P., Martinez, V., Mason, S., Marr, B., & Neely, A. et al. (2007). Towards a definition of a business performance measurement system. *International Journal of Operations & Production Management*, 27(3), 784–801. doi:10.1108/01443570710763778
- Goldberg, J. S., & Cole, B. R. (2002). Quality Management in Education: Building Excellence and Equity in Student Performance. *Quality Management Journal ASQ*, 9(4), 8–22.
- Haworth, J. G., & Conrad, C. F. (1996). *Emblems of Quality in Higher Education*. London: Allyn and Bacon.
- Mason, A. (1993). *Management training in Medium sized UK Business organizations*. London: Harbridge Consulting Group.
- Marsh, J. (1995). Special report: Quality in education on the move. *Quality Digest Magazine*, 15(9), 26–32.
- Queensland Government. (2010). *Guideline for the planning of an electronic Document and Records Management System (eDRMS)*. Queensland State Archives, *The State of Queensland*, Department of Public Works.
- Rouse, P., & Putterill, M. (2003). *Measuring the Effects of Local Government Reform: A New Zealand Highway Maintenance Application*. Auckland: Department of Accounting and Finance, The University of Auckland.
- Senge, P. (2006). *The fifth discipline: The art and practice of the learning organization* (rev. ed.). New York, NY: Doubleday.
- Tagtow, M. (2014). *Predictive Analytics in Higher Education*. EKS&H. Retrieved from <http://www.infolinkconsulting.com/wp-content/uploads/2014/09/EKSH-Predictive-Analytics-in-Higher-Education-Workshop-Sept-23-2014.pdf>
- Teay, S. (2012). *Strategic Performance Management System: An Integrated Framework* (International Edition). Bangkok: AU Digital Press.

Temporal, P. (1990). Linking management development to the corporate future—the role of the professional. *Journal of Management Development*, 9(5), 7–15. doi:10.1108/02621719010000263

Thompson, S., & Cobb-Reiley, L. (2003). *Using Electronic Portfolios to Assess Learning at the University of Denver*. Paper presented at 2003 AAHE Assessment Conference, Seattle, WA.

Tovey, L. (1991). *Management Training and Development in Large UK Business Organizations*. London: Harbridge Consulting Group.

Tribus, M. (1993). Why not education: Quality management in education. *Journal for Quality and Participation*, 16(1), 12–21.

Villanueva, C. C. (2003). *Education Management Information System (EMIS) and the formulation of Education For All (EFA) Plan of Action, 2002-2015*. UNESCO Almaty Cluster Office and the Ministry of Education of Tajikistan.

Yin, C. C., Kwok, H. N., & Mo, C. M. (2002). Economic considerations in education policy making: A simplified framework. *International Journal of Educational Management*, 16(1), 18–39.

ADDITIONAL READING

High Level Group on the Modernization of Higher Education. (2013). *Report to the European Commission on Improving the quality of teaching and learning in Europe's higher education institutions*. Luxembourg: Publications Office of European Union.

Kettunen, J. (2012). External and internal quality audits in higher education. *The TQM Journal*, 24(6), 518–528. doi:10.1108/17542731211270089

Mishra, S. (2006). *Quality Assurance in Higher Education: An Introduction*. National Assessment and Accreditation Council (NAAC), Karnataka, India and Commonwealth of Learning. Vancouver, BC, Canada: COL.

Prikulis, A., Raugvargers, A., & Rusakova, A. (2011). *Identifying Barriers in Promoting the European Standards and Guidelines for Quality Assurance at Institutional Level*. *Internal Quality Assurance Systems Cross-country comparative study*. Lifelong Learning Programme of European Union.

Wells, P. J. (2014). *The DNA of a Converging Diversity: Regional Approaches to Quality Assurance in Higher Education*. Washington, DC: Council for Higher Education Accreditation.

KEY TERMS AND DEFINITIONS

eDRMS (electronic Document and Records Management System): This contains the secure and systematic management of unstructured or semi-structured data, statistics, information and documents; a reduction in redundancy and duplication of information; a reduced risk of not being able to retrieve information when required; improved security, thereby reducing the risk of unauthorized access; greater ability to discover and re-use information; better control of statistics, information, data & document versions; and a reduction in the response time for information requests. This includes institutional information and metadata management strategies, standards, policies and procedures, effective data management culture of understanding and appreciation of the value of making, keeping and managing full and accurate statistics, data, information and documents, standard operating procedures including roles and responsibilities and security models (Queensland Government, 2010).

eIQA (electronic IQA): A collation of modular sub-systems which act inter-dependently for educational outputs that cover inputs, processes, output & outcomes (IPOO) modules inter-linked through the data warehouse (DWH), its data marts and use of Business Intelligence (BI) tools for analysis and ad-hoc reporting in assuring educational quality and performance excellence.

eMIS (Electronic Management Information System): This is “an organized group of information and documentation services that collects, stores, processes analyzes and disseminates information for educational planning and management. It is (1) a collection of component parts that include inputs processes, outputs and feed backs that are integrated to achieve a specific objective; (2) a system for managing a large body of data and information that can be readily retrieved, processed, analyzed, and made available for use and dissemination; (3) a tool that uses systems theory and developments in computerization to create a comprehensive approach to management and use of vast quantities of information in the education system” (Villanueva, 2003).

IQMS (Integrated Quality Management System): This is an integration of the performance and quality management for 3 key education purposes of 1) Development Appraisal; 2) Performance Management; 3) Whole-School Evaluation (De Bruyn, et al., 2007; ELRC, 2003).

QMS (Quality Management System): A QMS is a documented set of policies and procedures that provide assurance to the customer of the product and service levels expected which is: systems-based; people-based; process-based with emphasis on prevention (FASSET, 2000) and continuous improvements in all spheres to ensure and sustain efficiencies and effectiveness of educational quality and value.

H

International Students in Online Courses

María Ángeles Rodríguez Manzanares

Memorial University of Newfoundland, Canada

INTRODUCTION

Postsecondary institutions are focusing their efforts on recruiting higher numbers of international students. At the same time, they are increasing opportunities for students in general to complete courses online. We find ourselves, therefore, “at the intersection of two trends in postsecondary education” (Rodríguez Manzanares, 2012, p. 1), that is, a growth in online learning (including online courses) and an increase in numbers of postsecondary international students. However, to date, few studies have been conducted specifically on international students learning in online courses (Murphy & Rodríguez Manzanares, 2014; Rodríguez Manzanares, 2012).

The present article focuses on the little research that has so far been conducted on international students in online courses. In order to allow for a focus on recent research, only reports of studies of international students in online courses published after 2000 were included. Some of them (e.g., Rodríguez Manzanares, 2012; Sadykova, 2014; Zhang & Kenny, 2010) related to research that has been reported on elsewhere (see Murphy & Rodríguez Manzanares, 2014; Sadykova, 2013; Zhang, 2007). In terms of the studies’ research paradigms, Sadykova’s study relied on mixed methods and Liu’s (2008) and Sheu’s (2005) survey research fit within a quantitative research paradigm, while the other studies included in this article fit within a qualitative research paradigm.

In the studies that were selected for inclusion, the international students which were the focus of research were taking or had completed online courses or programs. Not all of the studies’ participants were completing their programs entirely online, as there were participants in the

studies who were enrolled in on-campus courses at least for part of their studies (whether taking them concurrently with online courses or not).

Researchers have different perceptions of the meaning of some terms used to describe different learning environments, such as is the case with “online learning,” “distance learning,” and “e-Learning” (see Moore, Dickson-Deane, & Galyen, 2010). For the purpose of this article, the focus was on online courses and an online course was defined as one being fully delivered using online technologies, without any face-to-face components being part of its design. For this reason, only studies in which international students completed courses designed to be completed entirely online were considered. Excluded were studies (e.g., Chen, Bennett, & Maton, 2008; Wang, 2006) that referred to most of the instruction (such as a percentage of it) being online.

The term “international student” is used interchangeably with “foreign student” (Abdullah, Aziz, & Ibrahim, 2014) in some contexts, but, in other contexts, it is used to refer to a subset of international students, in the sense that foreign students would comprise those “who are not citizens of the countr[y in which they are learning], but may be long-term residents or were born in that country” (Organisation for Economic Cooperation and Development, 2014, p. 352). To avoid terminological confusion, for the purpose of this article only the term “international students” is used. They are defined as those “enrolled in a university in a foreign country, where they do not have permanent residence” (Mehtap-Smadi & Hashemipour, 2011, p. 418). In order to ensure that the focus of the studies included in the present article was on students who would in fact have been considered “international students” in

DOI: 10.4018/978-1-5225-2255-3.ch338

their institutions, only studies indicating in their title that research was conducted in relation to “international students,” using that specific term, were selected for inclusion.

BACKGROUND

The increasing numbers of postsecondary international students in different countries attest to the efforts that postsecondary institutions are currently placing on recruiting more international students. For example, in the United States, the number of international students grew 72% between 2000 and 2014 (Institute of International Education, 2014) and, in Canada, international student enrollment (including students both at the postsecondary and other levels of education) grew 84% between 2003 and 2013 (Canadian Bureau for International Education, 2015). In the period from 2000 to 2012, as reported by the Organisation for Economic Co-operation and Development (2014), “the number of foreign tertiary students enrolled worldwide more than doubled, with an average annual growth rate of almost 7%” (p. 343). As indicated in a British Council report, demand for international student places in Australia, Canada, New Zealand, the United Kingdom, and the United States alone is predicted to increase from one million places to approximately 2.6 million places by 2020 (Bohm et al., 2004).

Alongside the trend of increasing international student enrollments at the postsecondary level, another current trend in postsecondary settings relates to increasing online course offerings (Murphy & Rodríguez Manzanares, 2014; Rodríguez Manzanares, 2012). For example, the EDUCAUSE Center for Analysis and Research published results of a survey in 2014 (Dahlstrom & Bichsel, 2014) which showed that almost half of surveyed students (47% exactly) reported having taken a completely online course the previous year. Data were collected from more than 75,000 students in 213 institutions in the United States and 15 other countries.

The present article aims to contribute to understanding postsecondary international students’ learning in online courses by focusing on the little research that has so far been conducted on international students in that specific learning setting. The next section reviews some of the findings of the studies of international students in online courses that were selected for inclusion in the article, following the selection criteria outlined in the Introduction above.

STUDIES OF INTERNATIONAL STUDENTS IN ONLINE COURSES

Students’ Reasons for Learning Online

Some of the insights gained from the studies of international students in online courses related to students’ reasons for taking courses in that format. This information was sometimes elicited directly from students, for example by including survey items (e.g., Liu, 2008; Sheu, 2005) or interview questions (e.g., Rodríguez Manzanares, 2012) in the studies on why they enrolled in online courses.

The main reason reported for taking courses online was time flexibility (e.g., Liu, 2008; Sheu, 2005; Rodríguez Manzanares, 2012) and also place flexibility in addition to time flexibility (e.g., Sheu, 2005; Rodríguez Manzanares, 2012). In Liu’s survey research, international students indicated that they enrolled in online courses, among other reasons, because of their advisor’s recommendation and the opportunity to work at one’s own pace (as indicated by 63% and 45.7% of the surveyed students, respectively). In terms of other findings, more than half of the students disagreed that they had enrolled in an online course because it was less expensive than a face-to-face course. Almost half disagreed that they had enrolled in online courses because they could provide the same quality as face-to-face ones.

In Tan et al.’s (2010) study, which relied on in-depth interviews with graduate international

students, participants reported avoiding taking more than one online course at the same time or per semester because of the perceived language- and culture-related challenges of learning in these courses. Wang and Reeves (2007) indicated that, sometimes, as was the case with the international students in their study, there are courses that are required and are offered only online. Finally, Sheu's (2005) study identified a reason for enrolling in online courses related to the fact that the face-to-face version of the course that students wanted to take was full.

Asynchronous Online Learning: Advantages and Disadvantages

Some studies identified the advantages (as perceived by students) of asynchronous learning in online courses for international students. Some of these were highlighted in relation to the time independence afforded by online courses (e.g., Liu, 2008; Park, 2006; Rodríguez Manzanares, 2012). In Liu's research, 67.4% of participants responded "agree" or "strongly agree" in relation to a survey item regarding whether the schedules of their online courses were flexible. As explained in Park's study of Asian students, students can "control time, place, and pace in their learning" (p. 67) in online courses. The time delay which is characteristic of text-based asynchronous learning in particular was identified as an advantage for international students in online courses, especially for English as a Second Language (ESL) speakers. In Park's study and others (e.g., Rodríguez Manzanares, 2012; Tan et al., 2010), the benefits of learning asynchronously in online courses for ESL speakers were highlighted in relation to asynchronous online discussions. Park referred to students having the opportunity to "[have] more time to think and write for their discussions" (p. 69). Tan et al. reported on international students' positive perceptions that online courses "promoted writing with care for an audience, providing opportunity for editing and revising correspondence before posting it..." (p. 12). Rodríguez Manzanares

also reported on students being able to organise their thoughts in advance when crafting messages, work in their own time, and concentrate on writing accurately in English, such as by relying on online dictionaries.

In Rodríguez Manzanares' (2012) research, other advantages of asynchronous learning in general that could "help students compensate for missing linguistic skills" (p. 175) included, in terms of reading, the opportunity to receive organized instructor feedback in writing. Some students also used online resources that they looked up in their first language to help them understand some of the concepts they were studying. Another benefit reported by international students in the same study and elsewhere (e.g., Park, 2006) was having course content and online discussions available to review at one's leisure. Asynchronous learning provided opportunities for self-directed learning for some of the students in Rodríguez Manzanares' study, as they took advantage of the opportunity to have "any-time access to online resources by means of independent online searching, outside of the learning management system" (p. 186). In fact, engaging in self-directed learning might be important for international students in online courses, as highlighted in Park's study. All of the study's participants were Asian and some of them had difficulties to "motivat[e] themselves without teachers pushing them to study[, as] they were accustomed to strong direction from teachers" (p. 96).

The studies also identified some disadvantages of online asynchronous learning for international students in online courses. Jusung and Park (2003), Park (2006), and Zhang and Kenny (2010), among others, identified disadvantages that related specifically to online discussions. Jusung and Park, who focused exclusively on online discussions in their study, relied on data from more than 1,000 online postings, whereas Park conducted interviews and Zhang and Kenny relied on an online survey, online observation, and email and telephone interviews. The disadvantages of online discussions for international students were highlighted

in relation to English language proficiency issues. For example, Park commented on students feeling the pressure to write well written, long messages in English. Zhang and Kenny reported on some of their findings as follows: “Non-native speakers of English were not very confident about their English proficiency and spent considerably more time than their English-speaking peers reading and composing messages [and] preferred to read others’ postings first rather than initiate a message....” (p. 28). JuSung and Park also reported on international students not initiating discussion but focusing on replying to others, one of the reasons being difficulty catching up on reading in discussions with numerous postings.

Besides students’ English language proficiency, other “cultural conditions” (Zhang & Kenny, 2010, p. 28) could affect their participation in online courses and, more specifically, in asynchronous discussions. Students in Zhang and Kenny’s study who had less experience in North American learning settings “tended to think that other course members were more knowledgeable” (p. 28), which diminished their confidence in their ability to express themselves in English. Some students were concerned about delay in communication, as they anxiously waited for replies “assum[ing] that if their posting did not get a response, it might mean they did not express themselves clearly or that their message was interpreted as offensive” (p. 28). As with some Asian students who may not consider it appropriate to challenge the ideas of others, some students in their study were “reluctant to argue with peers in a public forum if they did not agree with somebody’s opinion” (p. 27). Sheu (2005) also reported on most international students in her study, who were from an Asian cultural background, being “uncomfortable... communicating in a public environment such as a discussion board or chat room” (p. 93). Park (2006) and Jusung and Park (2003) found that international students received fewer responses from domestic students. Jusung and Park in addition reported on responses showing “excessive sympathy” (p. 9). The international students in

their study also relied on more frequent use of “powerless language” (p. 7), as illustrated in their use of question tags, hedges, and disclaimers.

Synchronous Online Learning: Advantages and Disadvantages

Students in some of the studies (e.g., Park, 2006; Rodríguez Manzanares, 2012) believed that they could have benefitted from synchronous sessions, had they been available in their online courses. In Park’s study, participants suggested including office hours (whether face-to-face or synchronously online) instead of using email, which meant having to wait for a reply. Although synchronous voice- or text-based sessions might “provide the immediacy that would otherwise be lacking with asynchronous interaction” (Rodríguez Manzanares, 2012, p. 182), they can also present disadvantages related to access to the sessions because of students’ time zones as well as responsibilities such as work-related ones (e.g., Liu et al., 2010; Zhang & Kenny, 2010).

Wang and Reeves’ (2007) study, which relied on interviews and observations to investigate international students in an online synchronous course, identified additional challenges related to these students’ participation in synchronous sessions. Students “could not listen to the lecture and read... instant messages at the same time[, as they] struggled to deal with multiple information channels....” (p. 353); however, it was not possible to ascertain to what extent this challenge might have been caused by language barriers. The students were also nervous “because they had to speak about their ideas online, because they were afraid of making language mistakes, and because of their fears that other classmates might not understand them” (pp. 347-348). Native English speakers spoke too fast for them to be able to follow during the synchronous sessions. Other findings of the study related to students feeling that, compared to face-to-face classes, it was difficult to socialize and establish friendships in online courses because of lack of non-verbal cues.

Participants in Wang and Reeves' (2007) study reported on developing friendships among each other when working in a group project. However, all of them believed that students in an online synchronous course would still need some face-to-face interaction. The lack of visual cues, according to the students, made it harder for them to capture their classmates' ideas, whereas being able to see others would have helped "understand different accents and guess what they were saying" (p. 348). On the other hand, the study's observations suggested that "none of [the participants] adopted strategies to minimize distractions during class" (p. 347), as they engaged simultaneously in activities such as web surfing and attending to personal phone calls, emails, or text messages.

Contrasting Findings

There were contrasting findings in some cases among individuals in the same study, as different students had different perceptions regarding online courses. For example, the findings of Tan et al.'s (2010) research in relation to lack of visual cues in online courses illustrate this point: "While some participants saw the lack of face-to-face contact... as a liability, others perceived it to be a benefit, [as] they were able to keep pace with communication better and comprehend more... than they would have in... face-to-face courses" (p. 13).

In some cases, there were contrasting findings among studies. For example, unfamiliar references, as Zhang and Kenny (2010) noted, could affect the participation levels of international students in online courses. In their study, "students with strong English language proficiencies and Western cultural backgrounds tended to dominate the discussion forums" (p. 29). In contrast, Park's (2006) study highlighted that international students, because of the delay associated with asynchronous interaction, could look for context when encountering unfamiliar references in their courses. Text-based interaction was for that reason "less threatening" (p. 101) for them.

The graduate international students in Tan et al.'s (2010) research were dissatisfied in relation to issues related to cultural inclusivity in their online courses, "holding the perception that very few instructors and peers pay attention to such issues" (p. 12). However, in Liu et al.'s (2010) study, which was also conducted with graduate international students, participants "noted different culture patterns in online collaboration and communication but indicated that these differences did not affect their learning in a negative way" (p. 183). These different patterns related, for example, to a collectivistic orientation, which is what the international students were accustomed to, vs. an individualistic orientation. The participants' perception regarding different cultural patterns may have related to the fact that they were Master of Business Administration (MBA) students who worked in large global companies. Their experiences working with international customers may have made them skilled in dealing with cultural difference, which, in relation to what they encountered when learning online related to, for example: different assessment styles in the United States and in their countries (process-oriented vs. examination-oriented styles and a focus on application vs. memorization); different rules of conduct, such as those regarding citation of others' work in written assignments; and differences in instruction and interaction related to a focus on conversing vs. lecturing, on induction (through reliance on cases in the MBA courses) vs. deduction, and on less structured vs. more structured instruction.

There were differences among some studies in relation to the importance that international students ascribed to socializing in online courses. Rodríguez Manzanares (2012) and Sadykova (2014) found that some international students may want to establish friendships with domestic students and participate in cultural exchange. Sadykova's mixed methods study, which included survey and interview data as well as data from a case study of one student, found that establishing personal friendships with domestic students might

be a matter of necessity for international students in online courses. She also found that, in online courses, these students may benefit from close ties with domestic students in order to help them learn about academic cultural conventions in the host institution and to “satisfy [their] interest in the host culture” (p. 43). The case study in particular “demonstrated the significance of social relationships that students, especially those who come from collectivistic high-context cultures, may seek in any format of learning – online or traditional” (p. 43). However, in Zhang and Kenny’s (2010) study of graduate international students, participants did not consider social interaction necessary. Some of them thought a space for socializing in online courses was “somewhat important,” but they “did not perceive it as critical for their learning” (p. 28).

Recommendations for Practice Identified in the Studies

The studies of international students in online courses identified recommendations for practice that referred not only to teaching but also to course design as well as support services. In some cases, recommendations were suggested by participants themselves (e.g., Park, 2006; Rodríguez Manzanares, 2012). Rodríguez Manzanares (2012) discussed in her study some of the recommendations for practice that have been identified in her research and other research on international students in online courses, which included adding some face-to-face interaction, whenever possible, and adding online synchronous interaction (e.g., for office hours) as well as asynchronous video or voice recordings, if these are not already part of the learning environment. Additional recommendations related to enhancing language-related services and supports, such as, for example, by offering support with writing skills for off-campus international students registered in online courses. In addition, instructors could support international students and students in general in online courses through provision of timely, extensive

feedback. These types of support could in turn help enhance teaching presence, which might be important for international students who are used to receiving direction from instructors. At the same time, instructors and course designers could focus on designing courses and providing course-related supports that facilitate more self-directed, independent student learning. In this regard, they could provide carefully designed opportunities for peer support and collaboration. In terms of issues related to cultural inclusivity in online courses, designers and instructors could focus on including global topics and cases in online courses, having some choice for students regarding topics for assignments, and providing information regarding cultural references which might be unfamiliar for international students. Finally, international students who want to establish friendships or participate in cultural exchange in general might benefit from the provision of opportunities in online courses for students to seek social interaction.

FUTURE RESEARCH DIRECTIONS

Studies of postsecondary international students learning in online courses have tended to include students as participants. Future studies may include also other individuals, such as online course designers and other learning specialists as well as course instructors and administrative and support personnel. In addition, most of the studies fitted within a qualitative research paradigm, except for Liu’s (2008) and Sheu’s (2005) studies, which fitted within a quantitative research paradigm, and Sadykova’s (2014), which relied on mixed methods. With the numbers of international students increasing at the postsecondary level and online course offerings also increasing, there might be growing interest among researchers in conducting further research on international students in online courses, which may result in more studies being conducted which fit within quantitative or mixed methods paradigms.

There were contrasting findings in some cases among individuals participating in the same study, as different students within the same study could have different perceptions regarding online courses (e.g., Tan et al., 2010), as outlined above. As also discussed above, there were contrasting findings among studies, for example in relation to the importance of socializing for international students in online courses (e.g., Sadykova, 2014; Zhang & Kenny, 2010). These types of contrasting findings emphasize the importance of paying attention both to the context of individual participants and of their learning setting in future studies of international students in online courses. Regarding individual students, detailed descriptions might be needed of their personal characteristics, including their cultural backgrounds and previous learning experiences in general as well as in relation to technology. As regards the learning setting, detailed descriptions could be provided of the types of tools used in the students' courses and of what tools or supports students may have relied on outside of their courses to help them learn. In-depth case studies of online courses or programs in which international students participate might help provide further insights into their learning.

CONCLUSION

For several decades, postsecondary institutions have focused on internationalization, particularly those in developed countries (Arabkheradmand et al., 2015). Arabkheradmand et al. have referred to a view of internationalization at this educational level as “a necessity, not a mandatory or even beneficiary but rather an existential requirement for every educational center in the twenty-first century” (p. 1). The two educational trends that were the focus of this article, an increase in numbers of postsecondary international students and an increase in online course offerings (with their potential for global reach), provide an illustration of some of the intersecting facets of international-

ization in postsecondary institutions. As the two trends continue, there might be greater interest among researchers in conducting more studies of postsecondary international students in online courses, which might help alleviate the current paucity of research in the area.

REFERENCES

- Abdullah, D., Aziz, M. I. A., & Ibrahim, A. L. M. (2014). A research into international student-related research: (Re)Visualising our stand? *Higher Education*, 67(3), 235–253. doi:10.1007/s10734-013-9647-3
- Arabkheradmand, A., Shabani, E. A., Zand-Moghadam, A., Bahrami, H. S., Derakhshesh, A., & Golkhandan, A. R. (2015). *An introduction to the internationalization of Higher Education: Essential topics*. Lanham, MD: University Press of America.
- Bohm, A., Follari, M., Hewett, A., Jones, S., Kemp, N., Meares, D., & Van Canter, K. (2004). Vision 2020. Forecasting international student mobility: A UK perspective. London: British Council. Retrieved from <http://www.britishcouncil.org/education/ihe/knowledge-centre/transnational-education/vision-2020>
- Canadian Bureau for International Education. (2015). *About IE: Facts and figures*. Retrieved from <http://www.cbie.ca/about-ie/facts-and-figures/>
- Chen, R. T. H., Bennett, S., & Maton, K. (2008). The adaptation of Chinese international students to online flexible learning: Two case studies. *Distance Education*, 29(3), 307–323. doi:10.1080/01587910802395821
- Dahlstrom, E., & Bichsel, J. (2014). ECAR study of undergraduate students and information technology, 2014. Louisville, CO: EDUCAUSE Center for Analysis and Research. Retrieved from <http://net.educause.edu/ir/library/pdf/ss14/ERS1406.pdf>

- Institute of International Education. (2014). *Press release open doors 2014: International students in the United States and study abroad by American students are at all-time high*. Retrieved from <http://www.iie.org/Who-We-Are/News-and-Events/Press-Center/Press-Releases/2014/2014-11-17-Open-Doors-Data>
- JuSung, J., & Park, J. H. (2003). *Power relations within online discussion context: Based on adult international students' perspective and their participation in the learning context*. Paper presented at the Forty-fourth Annual Meeting of the Adult Education Research Conference, San Francisco, CA. Retrieved from <http://eric.ed.gov/?id=ED480496>
- Liu, C.-H. (2008). *A study to analyze the attitudes toward online learning from international students at University of Arkansas* (Doctoral dissertation). Available from ProQuest dissertation and theses database. (Document ID 3341243)
- Liu, X., Liu, S., Lee, S.-H., & Magjuka, R. J. (2010). Cultural differences in online learning: International student perceptions. *Journal of Educational Technology & Society*, 13(3), 177–188. Retrieved from http://www.ifets.info/journals/13_3/16.pdf
- Mehtap-Smadi, S., & Hashemipour, M. (2011). In pursuit of an international education destination: Reflections from a university in a small island state. *Journal of Studies in International Education*, 17(2), 409–428. doi:10.1177/1028315310365539
- Moore, J. L., Dickson-Deane, C., & Galyen, K. (2010). e-Learning, online learning, and distance learning environments: Are they the same? *The Internet and Higher Education*, 14(2), 129–135. doi:10.1016/j.iheduc.2010.10.001
- Murphy, E., & Rodríguez Manzanares, M. A. (2014). *An Activity Theory perspective on technology in Higher Education*. Hershey, PA: IGI Global. doi:10.4018/978-1-4666-4590-5
- Organisation for Economic Co-operation and Development. (2014). *Education at a Glance 2014: OECD Indicators*. Paris: OECD Publishing. doi:10.1787/eag-2014-en
- Park, M. K. (2006). *The lived experience of Asian international students in online learning environments in higher education* (Doctoral dissertation). Retrieved from <https://mospace.umsystem.edu/xmlui/bitstream/handle/10355/4488/research.pdf?sequence=3>
- Rodríguez Manzanares, M. A. (2012). *Portraits of the activity systems of postsecondary international students in online learning: From tensions to transformations in activity* (Doctoral dissertation). Available from ProQuest dissertation and theses database. (Document ID 1038448572)
- Sadykova, G. (2013). Learning in new online cultures: East meets West. In C. Meskill (Ed.), *Online teaching and learning: Sociocultural perspectives* (pp. 21–38). London: Bloomsbury.
- Sadykova, G. (2014). Mediating knowledge through peer-to-peer interaction in a multicultural online learning environment: A case study of international students in the US. *International Review of Research in Open and Distance Learning*, 15(3). doi:10.19173/irrodl.v15i3.1629
- Sheu, Y. (2005). International students' perceptions toward online learning in American higher education. Unpublished Ed. D., University of South Dakota.

Tan, F., Nabb, L., Aagard, S., & Kim, K. (2010). International ESL graduate student perceptions of online learning in the context of second language acquisition and culturally responsive facilitation. *Adult Learning, 21*(1-2), 9–14. doi:10.1177/104515951002100102

Wang, C.-M., & Reeves, T. C. (2007). Cyberspace and online education: The influences of global cyberculture on international students. In K. S. Amant (Ed.), *Linguistic and cultural online communication issues in the global age* (pp. 239–252). Hershey, PA: Idea Group, Inc. doi:10.4018/978-1-59904-213-8.ch015

Wang, H. (2006). *How cultural values shape Chinese students' online learning experience in American universities* (Doctoral dissertation). Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.138.7437&rep=rep1&type=pdf>

Zhang, Z. (2007). *An investigation of the preferred learning practices of culturally diverse students in an online distance education environment* (Doctoral dissertation). Available from ProQuest dissertation and theses database. (Document ID 304706570)

Zhang, Z., & Kenny, R. (2010). Learning in an online distance education course: Experiences of three international students. *International Review of Research in Open and Distance Learning, 11*(1). doi:10.19173/irrodl.v11i1.775

ADDITIONAL READING

Amant, K. (2005). Distance education in a global age: A perspective for internationalizing online learning communities. *ACM SIGGROUP Bulletin, 25*(1), 12–19. doi:10.1145/1067699.1067702

Amirault, R. J., & Visser, Y. L. (2010). *The impact of e-learning programs on the internationalization of the university*. New York: Nova Science Publishers.

Bates, T. (2001). International distance education: Cultural and ethical issues. *Distance Education, 22*(1), 122–136. doi:10.1080/0158791010220107

Catterick, D. (2007). Do the philosophical foundations of online learning disadvantage non-Western students? In A. Edmundson (Ed.), *Globalized e-learning cultural challenges* (pp. 116–129). Hershey, PA: Information Science Publishing; doi:10.4018/978-1-59904-301-2.ch007

Dunn, P., & Marinetti, A. (2007). Beyond localization: Effective learning strategies for cross-cultural e-learning. In A. Edmundson (Ed.), *Globalized e-learning cultural challenges* (pp. 255–266). Hershey, PA: Information Science Publishing; doi:10.4018/978-1-59904-301-2.ch015

Eberle, J. H., & Childress, M. D. (2007). Universal design for culturally-diverse online learning. In A. Edmundson (Ed.), *Globalized e-learning cultural challenges* (pp. 239–254). Hershey, PA: Information Science Publishing; doi:10.4018/978-1-59904-301-2.ch014

Gunawardena, C. N., & LaPointe, D. (2007). Cultural dynamics of online learning. In M. G. Moore (Ed.), *Handbook of distance learning* (2nd ed.; pp. 593–607). Mahwah, NJ: Lawrence Erlbaum Associates.

Mason, R. (2007). Internationalizing education. In M. G. Moore (Ed.), *Handbook of distance education* (2nd ed.; pp. 586–591). Mahwah, NJ: Lawrence Erlbaum Associates.

Pincas, A. (2001). Culture, cognition and communication in global education. *Distance Education, 22*(1), 30–51. doi:10.1080/0158791010220103

Seufert, S. (2002). Trends and future developments: Cultural perspectives of online education. In H. H. Adelsberger, B. Collis, & J. M. Pawlowski (Eds.), *Handbook on information technologies for education and training* (pp. 121–135). Berlin: Springer.

Uzuner, S. (2009). Questions of culture in distance learning: A research review. *The International Review of Research in Open and Distributed Learning*, 10(3). doi:10.19173/irrodl.v10i3.690

Wang, C., & Reeves, T. C. (2007). The meaning of culture in online education: Implications for teaching. In A. Edmundson (Ed.), *Globalized e-learning cultural challenges* (pp. 1–17). Hershey, PA: Information Science Publishing; doi:10.4018/978-1-59904-301-2.ch001

KEY TERMS AND DEFINITIONS

Foreign Students: Used synonymously with the term international students in some contexts, whereas, in other contexts, foreign students are considered to be a subset of international students.

Higher Education: See Postsecondary Education.

International Students: As traditionally defined, international students are those who cross borders in order to further their education.

Online Course: A course delivered using online technologies and, for the purpose of this article, a course delivered fully online, without face-to-face components being part of its design.

Online Asynchronous Learning: Learning delivered using technologies allowing for asynchronous (not in real time) interaction.

Online Synchronous Learning: Learning delivered using technologies allowing for synchronous (in real time) interaction.

Postsecondary Education: Education beyond the high school (secondary) educational level.

Tertiary Education: See Postsecondary Education.

IT Solutions Supporting the Management of Higher Education Institutions in Poland

Elżbieta Janczyk-Strzała

Wroclaw School of Banking, Poland

INTRODUCTION

Operating as they are on the contemporary difficult and dynamic market, higher education institutions (HEIs) should strive to ensure suitable conditions for a continuous furthering of their educational mission. Demographic changes and competition force the authorities of HEIs to effectively manage their resources and potential. It is vital then that HEIs, through the use of sophisticated information technology, draw from the opportunities offered by multi-dimensional reporting tailored to the requirements of specific customers. The present paper presents various IT solutions that can assist HEIs by supporting effective management, administration and organization of their work. It is a result of literature studies, empirical research and own experience of the author. This publication fills a gap as a source of information about IT solutions and their application in such specific entities as higher education institutions. It shows the great role and importance of information for HEI management purposes.

BACKGROUND

The beginnings of higher education go back to antiquity when the Athenaeum was founded in Rome by Emperor Hadrian as well as the university of Alexandria. However, proper multi-departmental higher education institutions appeared only in medieval Western Europe. Among them were the University of Bologna (approx. 1088), University of Oxford (approx. 1167), University of Paris (approx. 1170), University of Cambridge

(approx. 1209) and University of Padua (1220). (*Selekcyjna funkcja...*). The first Polish university was the Kraków Academy founded in 1364. It was the second oldest university in Central Europe, after Charles University in Prague (1348) (see: www.uj.edu.pl). At the beginning of the Second Polish Republic, after the period of partitions of Poland, there were five universities (in Kraków, Vilnius, Lviv, Warsaw and Poznań), two technical universities (in Lviv and Warsaw), Szkoła Główna Gospodarstwa Wiejskiego (Warsaw University of Life Sciences), Mining Academy in Kraków and the Academy of Veterinary Medicine (Lviv) (Jaczewski, 1987, pp.206-210). WWII brought severe losses to Polish science. By the decision of the occupier, higher education was to be eliminated. In 1945-1989, during the period of People's Poland (PRL), first under Stalin and then under the socialist regime, higher education was under a strong political indoctrination. Management of higher education institutions was partly in the hands of the central administration (Thieme, 2009, p.229). In 1989 the number of HEIs went up to 97 while the number of students reached 378,000. They were only state-owned institutions with only one exception for the Catholic University of Lublin (KUL) which is a non-public university whose founding body is the church.

Since 1990s, higher education has been undergoing numerous transformations. The Higher Education Act enacted by the Polish Sejm on 12 September 1990 provided legal framework for the development of non-public education in Poland. The next law of 26 June 1996 on higher vocational schools regulated the activity of schools offering only vocational education and awarding bachelor

Table 1. Number of students in Poland in 1999-2011

Academic Year	All Students	Students Number Indicator (1990/1991=100)	Academic Year	All students	Students Number Indicator (1990/1991=100)
1990/1991	390409	100	2001/2002	1718747	440
1991/1992	414609	106	2002/2003	1800548	461
1992/1993	481273	123	2003/2004	1858680	476
1993/1994	568702	146	2004/2005	1930917	495
1994/1995	666712	171	2005/2006	1953832	500
1995/1996	779907	200	2006/2007	1941445	497
1996/1997	917939	235	2007/2008	1937404	496
1997/1998	1082657	277	2008/2009	1927762	494
1998/1999	1265347	324	2009/2010	1900014	487
1999/2000	1421277	364	2010/2011	1841251	472
2000/2001	1584804	406	2011/2012	1764060	452

Source: *Szkoły wyższe i ich finanse 1999-2011 [HEIs and their finances 1999-2011]*, Warsaw.

or engineer degrees. The Law on Higher Education systematised tertiary education and introduced a division into first- and second-cycle studies. Additionally, some institutions offer individual programmes of studies or regular studies in a foreign language.

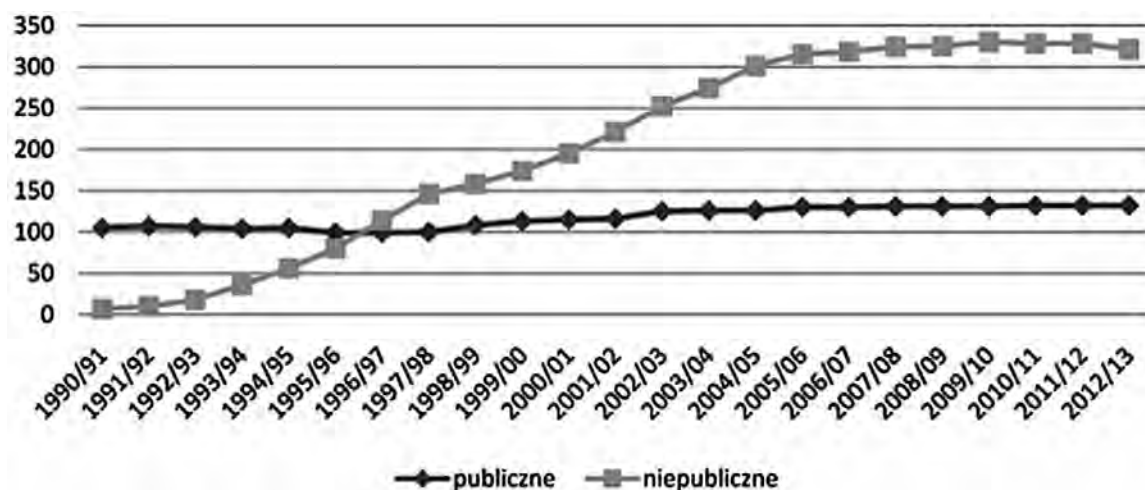
From the academic year 1990/1991 the number of students was growing until the year 2005/2006 when it reached its peak at 1.95 million. Since that time, the number of students has been on a gradual

decline. The above changes can be attributed to the decreasing population aged 19-24, to Poland's EU accession which offered Polish prospective students the opportunity to study abroad, and are also probably a consequence of amendments to the Law on Higher Education that introduced tuition fees for pursuing a second and other degrees (Report "Higher Education...", 2013).

As forecasted by the Ministry of Science and Higher Education, in the following years the above

Figure 1. Number of public and non-public HEIs in Poland

Source: *Szkoły wyższe i ich finanse w 2012 [HEIs and their finances in 2012]*, Warsaw, 2013, p.31



number will continue to fall until 2025-2026 when it will reach 1.26 million. In the next years it is expected to grow steadily until 2030-2035 when it will reach a level recorded at the turn of the 21st century.

FEATURES OF IT TOOLS SUPPORTING HEI MANAGEMENT

Operating as they are on the contemporary difficult and dynamic market, higher education institutions should strive to provide suitable conditions for the fulfilment of their educational mission. Various IT tools may turn out to be helpful, supporting the process of managing the higher education institution, by facilitating the gathering, processing and transfer of information. Therefore, the process of building a given IT tool is of paramount importance. It should have the following features (Vollmuth, 2000, pp.84-85; Janczyk-Strzała, 2008, p.41; Janczyk-Strzała, 2013, pp.166-167):

- **Currency:** Present information relevant for the reality at hand.
- **Economy:** Present only the necessary data.
- **Objectivity:** Making sure that no “window dressing” takes place.
- **Consistency and Stability:** Ensuring data comparability thanks to predefined and fixed reporting system.
- **Clarity:** Use of unequivocal terms and phrases (without undue “theoretising”), numerical data and discussions supported by suitable graphics.
- **Single Source of Information:** Ensuring data consistency.
- **Recipient Focus:** Adjusting of the scope and degree of detail of the data to the information needs of individual recipients.

The choice of a suitable IT tool supporting planning, control and coordination determines the effectiveness of the entire HEI management process. For the above reasons, the manner of

presenting information for various levels of HEI management should be different (Figure 2).

TYPES OF IT TOOLS SUPPORTING HEI MANAGEMENT

An information system must be adjusted to the situation of a HEI, to its needs and capabilities. Its purpose is to ensure proper evaluation of the HEI's position (through an analysis of the processes taking part in it) as well as of the process of implementation of its strategy (in the context of opportunities and threats). For the above reasons, HEIs should use the following *IT tools supporting their operation*:

- Universal office software packages (mainly spreadsheets, word processors and database management software),
- Dedicated software,
- Integrated HEI management systems,
- Company websites based on website building technology.

The features of above IT tools are presented in Table 2.

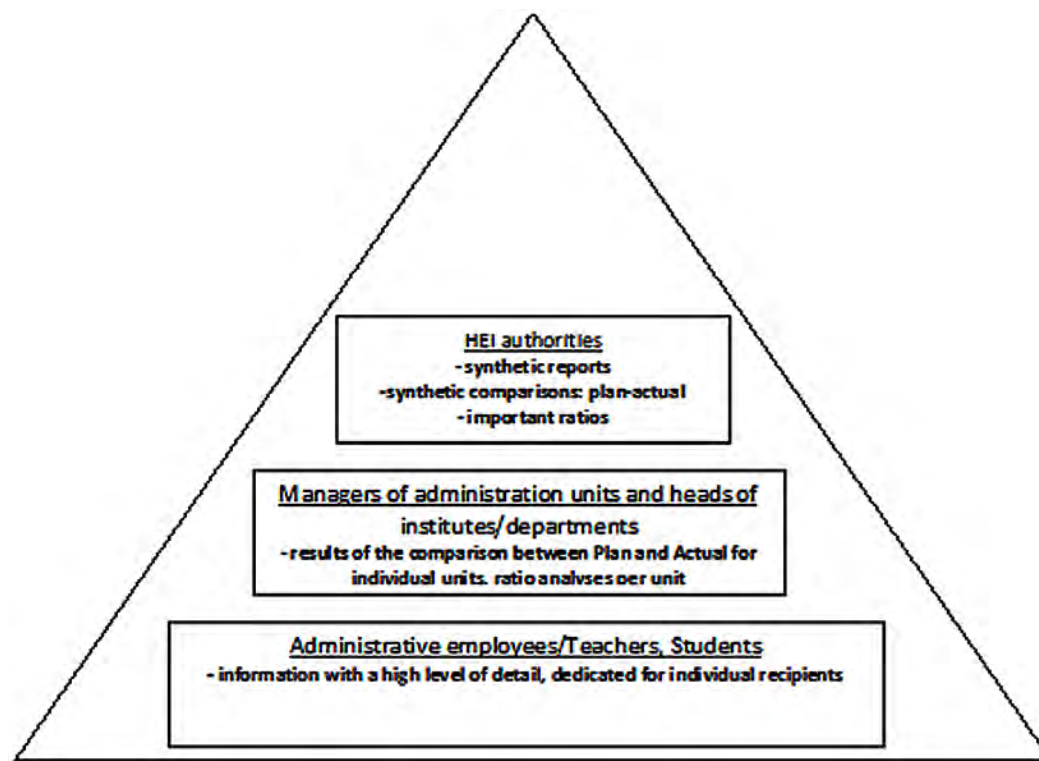
Users of the above tools must remember that the IT systems supporting HEI management must meet the requirements of the law (among others general statutory requirements, regulations applicable to higher education institutions or EU legislation in force in Poland). They should also be adapted to the needs, operating conditions and capabilities of the HEI in question.

EXAMPLES OF IT SOLUTIONS FOR HIGHER EDUCATION INSTITUTIONS

For the purposes of this paper, an analysis has been made of the various IT tools supporting the operation of educational establishments in Poland. The research was conducted in February-April 2015. The research results show the possibilities

Figure 2. Management level and type of information presented

Source: Own data.



offered by the analysed solutions used to enhance internal procedures and to support various business processes and the performance of statutory duties. The IT technology presented below is used in the management of individual educational institutions (HEIs), but some of them also serve to provide information for the management of the national education system as a whole.

HEI Management Systems

The research showed that a half of Polish HEIs use in-house developed computer programmes. They serve individual areas of the institution's activity (e.g. bursary, teaching, research, etc.) and require coordination and integration through the use of suitable interfaces enabling the flow of information between them. Some HEIs still use manual transfer of data between systems.

If a HEI does not want or cannot develop a dedicated software in-house it may use solutions available on the market, such as:

- Bazus (admissions, student registry, virtual dean's office, class planning, student ID or dormitory management),
- HMS Solutions (supports the work of the dean's office, bursary, HR and payroll),
- Sokrates (admissions, records and accounting of credits/examinations, reporting and statistics),
- Uczelnia.XP (supports the admission process, teaching, accounting of teaching hours and management of dormitories),
- USOS (supports the management of courses of study).

Table 2. Features of IT tools supporting HEI management

Type	Properties	Advantages	Disadvantages
Office package	• universal • easily available and popular	• the most cost effective solution available on the market • easy to use (does not require any professional IT training), • small financial outlays (costs of licence and implementation of the software), • precisely and quickly adjustable to the current information needs of the HEI • highly flexible	• limited possibilities of processing a large number of data • limited capabilities for multidimensional analysis
Dedicated software	adjustable to the needs and requirements of the HEI	highly adjustable to the specific features of the sector, environment and size of the institution	• requires considerable financial and time outlays • risk of dependence on a single contractor
Integrated Management System	• modular architecture • flexibility - range of functionalities may be adjusted through using basic and additional functions • open architecture - individual modules are self-sufficient but may also be connected to external databases • modularity - possibility to introduce new applications individually (e.g. starting with Accounting and then adding Reporting), • central database - shared by all the modules and enabling the flow of information between them • adjustable to the specific user requirements (through suitable parameter settings) • technologically advanced	• enables cost control – through full control over information and HEI activities by using various system functionalities • increased productivity - easy to use and intuitive user interface enabling efficient work • controllable, systematized and continuous data recording • offers immediate data analysis for any time period • global reach – e.g. multi-currency (the transactions performed can be recorded in multiple currencies) • revenue increase – immediate access to information from various parts of the system leads to quick responding of the HEI to the emerging opportunities and threats, better customer service and to improved cooperation with other parties	• high implementation costs • dedicated mainly for large and medium-sized institutions • benefits grow only after several modules have been implemented
Company website	modularity and independent operation of individual components	• scalability • quickly extendable functionalities • adaptable to user needs (website layout and functions can be configured) • dynamic content management (easy and quick content updating) • security (state-of-the-art data security features)	very expensive (price exceeds even IMS)

Source: own study based on: Janczyk-Strzała, 2008, pp.43-44.

Dedicated software is adjusted to the specific features of the education sector and to the needs and requirements of individual HEIs. To learn more about dedicated systems, see: *Raport końcowy "Modele zarządzania uczelniami w Polsce"* [Final Report "HEI Management Models in Poland"], 2011, pp. 240-241.

Some larger institutions offering MA programmes have implemented integrated systems. They are advanced tools supporting the process of non-public HEI management. Their implementation is not an easy task, as it requires the involvement of many people from various areas of the institution, as well as significant financial

outlays (e.g. for licence purchases, for the implementation, employee training and purchase of the necessary IT equipment). Additionally, their implementation requires organisational changes within the HEI involving for instance organisational structure, work organisation in individual departments or the flow of information between them. Although the implementation of the above systems is not a fast process, once successfully completed, it may contribute to optimising internal processes as well as the processes taking part in the close environment of the institution. The systems in question are comprehensive, have flexible structure and functionalities, and offer technologically advanced solutions supporting the management of various areas of the HEI. Their architecture is usually modular, with each module responsible for supporting a different function of the institution (e.g. accounting, controlling, HR, project or investment management). On the Polish education market, systems of the above class have been mainly implemented at public HEIs (e.g. Impuls 5, Microsoft Dynamics, SAP, SIMPLE.ERP, Teta Constellation), but recent years also witnessed a couple of examples of ERP system implementations at non-public schools – such as for example SAP or SIMPLE.ERP.

E-Learning Platforms/Systems

When offering e-learning under the distance-learning system, HEIs may use a number of various tools, starting from email, through a website enabling access to digital study materials, to integrated IT systems which are becoming more and more popular (also referred to as e-learning platforms) supporting the study process at its individual stages. The above systems comprise:

- **Authoring Systems (AT):** Original tools used for the preparation of e-learning courses.
- **Learning Management Systems (LMS):** Referred to as LMS platform or e-learning platform; a system enabling course ad-

ministration and content presentation. The system enables course launch, registration of participants and participant progress tracking.

- **Assessment Systems (AS):** Assessment systems used for designing, drafting and servicing of electronic tests/examinations.
- **Learning Content Management System (LCMS):** Apart from the functions of an LMS, it has modules used for creating teaching content (Caniëls, Smeets-Verstraeten, Bosch, 2007).
- **Course Management Systems (CMS):** Systems used in the management of training processes offering automation of management, administration and reporting of e-learning results.
- **Virtual Learning Environments (VLE):** Platform enabling uploading of teaching content in the form of courses for a closed user group with a possibility to extend the system to include new accounts and user groups. It is composed mainly of the curriculum, timetable and teaching materials (e.g. animations, video).
- In Poland, the dominant educational platform is Moodle whose popularity is chiefly owed to the free GNU General Public Licence that enables to avoid the additional costs of purchasing the system (Zieliński, 2007). There are also other systems such as ILIAS which, like Moodle, is based on the GNU GPL licence.

The following are possible applications of learning platforms:

- Ready-made commercial learning platforms used for the design of training sessions (e.g. IBM's Lotus LearningSpace platform – used by the Warsaw University of Technology).
- **Original Solutions Designed and Created by Individual HEIs:** They can incorporate individual elements of ready-

made systems or such elements may be built from scratch (e.g. e-sgh platform at SGH - Warsaw School of Economics, EDUX at the Polish-Japanese Academy of Information Technology).

- **Rented Learning Platforms:** Used for specific training sessions when the institution does not wish to purchase the entire platform. Such rentals are offered for instance by Alatus or Oracle (for more on the subject, see: Nojszewski D., Platformy...).

First attempts can be observed at pooling the efforts of several institutions to exchange experience and at building a common learning platform. Among the examples is the Warsaw University of Technology with its Wirtualna Politechnika project (<http://www.okno.pw.edu.pl/>).

Electronic Documents Flow

To optimise the costs of clerical services (streamline the flow of incoming and outgoing mail) and to reduce the time of documents flow, an increasing number of large HEIs decide to introduce electronic documents flow. The above relates primarily to multi-department HEIs, which use the above method to minimise the risk of loss of documents, including in particular financial documents. Thus, among the immediate results of the implementation is increased security of the institution and a significant improvement in the efficiency of its internal processes. Electronic documents flow is a scalable system, so it can be tailored to the specific features and needs of any HEI. Additionally, a document flow system may be integrated with other systems implemented at the HEI, such as ERP, proprietary systems, website, etc.

Polish HEIs may choose from among a dozen or so products. Among the examples of systems used by higher education institutions are:

- **DocuSafe®:** Used by the Wrocław Medical University, Strzemiński Academy of Fine

Arts Łódź and Cardinal Stefan Wyszyński University in Warsaw, Maritime University of Szczecin, Casimir the Great University in Bydgoszcz or Poznań University of Medical Sciences

- **SIMPLE.EOD (by SIMPLE):** Used, among others, by the University of Łódź and the Częstochowa University of Technology
- **CONTMAN DIRECTOR (by Opt Team):** Implemented at the Kazimierz Pułaski University of Technology and Humanities in Radom
- **OfficeObjects@Document Manager (by Rodan Systems):** Warsaw University of Technology
- **Dokumenty (by BETASOFT)**
- **EOD (by PARTNERS In PROGRESS)**
- **OPTiRCM (by OPTeam S.A.)**
- **Comarch Workflow (by Comarch)**

It should be stressed that the significance attached to electronic flow of documents has reached a level that made it a subject of the 6th Forum of IT University Management held on 25 March 2015 which included presentations on electronic documents flow, automation of business processes or document archiving (for more on the forum see: <http://docusafe.pl/author/admin/>).

Plagiarism Detection Software

In order to ensure originality of diploma theses, HEIs have implemented verification mechanisms including supervisor guidance during the seminar, student's declarations of originality of their diploma projects, thesis reviews and diploma examinations. Moreover, since last year, HEIs have had a statutory duty to implement plagiarism detection measures using specialist software. The above software detects similarities between the student's work and earlier papers published on the Internet or stored in the systems.

Plagiat.pl is the leading plagiarism detection software on the Polish education market. Ac-

cess to the software was purchased by 185 out of 441 public and private HEIs (according to 2014 data <http://prawo.rp.pl/artykul/1080058.html?print=tak&p=0>, retrieved on 23 April 2015). The software enables to generate a similarities report providing the following information:

- “The so-called similarity index, showing what percentage of the verified text is identical with texts in the database or on the Internet.
- List of the longest fragments identified by the system as identical with those in the database or on the Internet.
- List of sources in which fragments of texts are identical with fragments of the submitted text.
- Text of the submitted document with marked fragments identified as identical with fragments of texts from the database and from the Internet.
- All fragments allegedly copied from a specific document marked with a specific colour – which facilitates the assessment of their type” (according to data given on the website informing about the system https://www.plagiat.pl/webplagiat/main.action?menu=important_info#subject, retrieved on 23 April 2015).

Each HEI sets its own rules for interpreting the report as well as for other areas of system operation. Usually the above report is generated by dean’s office employees, and where a high percentage of copied text is detected its results are transferred to the thesis supervisor and/or to a special committee.

Apart from Plagiat.pl, there are other systems available on the Polish market, of which the following should be mentioned:

- **OSA Open Plagiarism Detection**
System: Developed by the Interuniversity Computerization Centre (MUCI), composed

of universities from, among others, Warsaw, Wrocław, Kraków, Katowice, Poznań and Lublin, in collaboration with the Institute of Fundamentals of Informatics of the Polish Academy of Sciences. Unlike other systems, this one does not use directly the papers in their original form but it uses data structures storing fragmentary information about the texts – the so-called incidence vectors and maps (<http://osaweb.pl/index.php>, retrieved on 23 April 2015).

- **SowiDocs:** Developed by the Gdańsk University of Technology. It enables searching similar documents, searches identical fragments (copy-paste) and similar fragments (copy-edit-paste). It is a part of the comprehensive system of intellectual property protection (SOWI) and may be extended to include a system for diploma thesis archiving – Digital Document Repository (RDC), a platform for the student-teacher cooperation and a platform offering access to diploma theses (<https://sowi.pg.gda.pl/index.php/pl/content/ap-usage>, retrieved on 23 April 2015).

The OSA and SOWI systems were presented during the European Anti Plagiarism Forum EFA 2014, the first Polish conference devoted to plagiarism detection solutions for HEIs hosted by the European Forum of Law and Education on 24 October 2014. Conference participants also learned about two other systems used in other countries – Turnitin (popular in the UK) and about the Sweden’s Urkund.

It should also be stressed that there are free programmes available on the Internet for detecting plagiarism of intranet content; they can be used by students at the stage of verifying their Bachelor’s/ Master’s theses (e.g. Antyplagiat.net.pl). Reviewers who often check papers on similar topics, also have a possibility of verifying whether a given piece of text is repeated in other files saved on their computer (e.g. using AntiPlagiarist).

Other Examples of Modern IT Technology Dedicated for HEIs

Other examples of modern IT solutions that more and more frequently support the operation of contemporary higher education institutions include:

- **Digital Libraries:** Offering online access to digital publications as well as to digitised hard copy publications,
- **Digital/Virtual Dean's Office:** Among other places, the functionality is implemented at the Pontifical University of John Paul II in Kraków, University of Economics in Katowice, Lodz University of Technology, Warsaw School of Economics, Cracow University of Economics, School of Administration in Bielsko-Biała or the University of Dąbrowa Górnicza. It enables the use of network resources supporting the school's teaching process,
- **Electronic Student ID Card:** Introduced in the regulation of the Minister of Science and Higher Education of 2 November 2006 on course of study documentation (Dz. U. of 2006 No. 224, item 1634). It is in the form of a chip card. The card enables personalization of the ID, as it may be associated with a banking card or with an electronic key (access to laboratories), etc. It may also be used as a public transport pass (under the URBANCARD system).
- **Electronic ID of HEI Employee:** May be issued to administration employees, teachers and academics. It offers a number of functions, depending on the holder's rights – such as for example user identification, logging into workstations, work time tracking, access to buildings/premises/parking lots or to library resources,
- **Electronic Admissions:** Enabling online registration by filling out a special admissions form,
- **Information Kiosks:** Multimedia computers available to students 24/7 (also beyond

working hours of dean's offices or other administrative departments) placed at various locations on university premises; offered mainly by private HEIs,

- IT tools supporting staff management and offering, among others, employee assessments or evaluation of research output.

INFORMATIZATION OF HIGHER EDUCATION INSTITUTIONS IN THE CONTEXT OF REPORTING FOR THE MINISTRY OF SCIENCE AND HIGHER EDUCATION

The Ministry of Science and Higher Education has created a higher education information system called POL-on. It is an integrated system composed of a number of modules. The following are among the most important subject areas of the POL-on system:

- **Basic Details of Academic Units and Higher Education Institutions:** The register shows the structure of the higher education and science system. The data collected are publically available and are gathered for information purposes. The register has a functionality that enables access to information about quality assurance system in place at a chosen HEI or organisational unit.
- **Register of Rights Granted to Organisational Units of Higher Education Institutions and to Research Centres to Award Academic Degrees:** The register offers information about units having rights to award the degree of doctor and doctor habilitated in individual areas and disciplines of arts and science.
- **Statement of Rights to Offer Programmes of Study by Field of Study, Level or educational Profile:** It is an informational list presenting all the rights to offer programmes of study in a specified field, of

- a specified level and educational profile that have been granted to HE providers.
- **Nationwide Register of Students:** It includes the most important information about students enrolled in individual courses (both full-time and part-time), including: personal details, number of ECTS points and information about using any financial aid programmes. The purpose of the register is to monitor the financial aid offered and to manage tuition fees.
 - **Register of Research Employees and Academic Teachers:** The register is kept to verify if a given HE unit meets the requirements for minimum academic staff complement at individual departments, as well as to keep track of other jobs held by employees as well as to verify the HEI's rights to award academic titles and degrees.
 - **HEI Reporting:** The purpose of gathering the data is to streamline the reporting process and to remove the burden from the institution by replacing the financial statements filed with the Central Statistical Office (GUS), MNiSW (Ministry of Science and Higher Education) and to the Ministry of Finance with a single standard report.
 - **List of Science Promotion Events:** The purpose of the list is to gather data about awards and distinctions granted to individual institutions or their employees. The list also includes information about scientific conferences hosted or co-hosted by a given unit.
 - **Registers of Real Estate, Infrastructure, Laboratories and Apparatus:** A generally available database gathering information about research resources of research units open to a broad group of prospective users, both from the academic and industry sector.
 - **List of Patents and Achievements:** Includes information about patents registered, invention applications, rights of protection and copyrights, as well as about implemented products or research results.
 - **Higher Education Unit Survey:** A report that must be filed by all units applying for subsidies from the central government for their statutory activity. On the basis of the survey, a parametric evaluation of the current activity of the unit is calculated and a resulting research unit category is assigned.
 - **CKdsST Electoral System:** The system was used for the conducting of Poland's first fully electronic elections to the Central Commission for Academic Titles and Degrees (<http://polon.nauka.gov.pl/system>, retrieved on 24 April 2015).
- HEIs (as well as research centres) have been reporting to POL-on since autumn 2011 and the scope of this reporting is going to be extended in the coming years. The following modules are planned: statistical module for the Central Statistical Office (GUS), repository of Bachelor's/ Master's theses and doctoral dissertations and a register of diplomas issued by individual institutions.

FUTURE RESEARCH DIRECTIONS

The study discusses the importance of IT tools for HEIs and shows, based on research results, an examples of various IT tools supporting the operation of educational establishments in Poland. Further future improvements of the IT solutions impact of the polish HEIs shall be expected and they may have a positive effect on the Polish education market. Further statistical analysis of the related issues will be one of our future works.

CONCLUSION

These days, given volatile market conditions, competition and the coming population decline, higher education institutions should not only strive to increase the quality of their services and ensure rational management of their resources, but they should be equally focused on investing in state-of-the-art IT solutions and technology that may support and optimize their management. Through the use of suitable solutions and IT tools a HEI can support its management and adjust itself better to internal and external conditions. It can also keep and even improve its competitive position on the education services market.

REFERENCES

- Act of 27 July 2005 Law on Higher Education, Dz.U. of 2005, No. 164, item 1365
- Caniëls, M. C. J., Smeets-Verstraeten, A. H. J., & Bosch, H. M. J. van den (2007). The Challenges of Educating People to Lead in a Challenging World. *Educational Innovation in Economics and Business*, 10.
- Du Vall, M. (2011). *Report Modele zarządzania uczelniami w Polsce*. Kraków: Uniwersytet Jagielloński, Centrum Badań nad Szkolnictwem Wyższym.
- Higher Education Act of 12 September 1990*. Dz.U. of 1990 No. 65, item 385, as amended.
- Hołdyńska, E. (n.d.). *Nowy bat dla plagiatorów*. Retrieved April 23, 2015, from <http://prawo.rp.pl/artykul/1080058.html?print=tak&p=0>
- Jaczewski, B. (1987). Życie naukowe w Polsce odrodzonej. In B. Jaczewski (Ed.), *Życie naukowe w drugiej połowie XIX i w XX wieku*. Organizacje i instytucje (pp. 206-210). Zakład Naukowy im. Ossolińskich, Wydawnictwo PAN.
- Janczyk-Strzała, E. (2008). *Controlling w przedsiębiorstwach produkcyjnych*. Warszawa: CeDeWu.
- Janczyk-Strzała, E. (2013). *Controlling we wspomaganiu zarządzania uczelniami publicznymi i niepublicznymi w Polsce*. Warszawa: CeDeWu.
- Nojszewski, D. (2003). *Platformy e-learningowe w polskich instytucjach edukacyjnych*. Retrieved April 24, 2015 from <http://www.e-mentor.edu.pl/artykul/index/numer/2/id/20>
- Raport "Szkolnictwo wyższe w Polsce", Ministerstwo Nauki i Szkolnictwa Wyższego. (2013). Retrieved April 24, 2015 https://www.nauka.gov.pl/g2/oryginal/2013_07/0695136d37bd577c8ab03acc5c59a1f6.pdf
- Selekcyjna funkcja szkolnictwa wyższego w krajach Europy Zachodniej. (n.d.). Retrieved February 11, 2012 from http://www.sciaga.pl/tekst/46291-47-selekcyjna_funkcja_szkolnictwa_wyzszego_w_krajach_europy_zachodniej
- Szkoły wyższe i ich finanse w 2012r. (2013). GUS.
- Thieme J.K. (2009). *Szkolnictwo wyższe. Wyzwania XXI wieku*. Warszawa, wyd. Difin.
- Vollmuth, H. J. (2000). *Controlling. Planowanie, kontrola, kierowanie*. Warszawa: Agencja Wydawnicza Placet.
- Web of Internetowy System Antyplagiatowy. (n.d.). Retrieved April 23, 2015, from https://www.plagiat.pl/webplagiat/main.action?menu=important_info#subject
- Web of Ośrodek Kształcenia na odległość. (n.d.). Retrieved April 21, 2015, from <http://www.okno.pw.edu.pl/>
- Web of Otwarty System Antyplagiatowy OSA. (n.d.). Retrieved April 23, 2015, from <http://osaweb.pl/index.php>
- Web of POLON. (n.d.). Retrieved April 24, 2015, from <http://polon.nauka.gov.pl/system>

Web of SowiDocs. (n.d.). Retrieved April 23, 2015, from <https://sowi.pg.gda.pl/index.php/pl/content/appusage>

Web of Uniwersytet Jagielloński w Krakowie. (n.d.). Retrieved April 23, 2015 from www.uj.edu.pl

Zieliński, Z. E. (2007). *Przegląd wybranych systemów i narzędzi e-learning*. Zeszyty Naukowe SCENO, 4.

ADDITIONAL READING

Clark, B. R. (1986). *The Higher Education System: Academic Organization in Cross-national Perspective*. California, PA: University of California Press.

Conole, G., & Panagiota Alevizou, P. (2010). *A literature review of the use of Web 2.0 tools in Higher Education. A report commissioned by the Higher Education Academy*. Retrieved January 19, 2016 from <https://core.ac.uk/download/files/86/5162.pdf>

Diagnoza stanu szkolnictwa wyższego w Polsce. (2009). Edited by Ernst&Young.

Ryńca, R., & Matuszak, M. (2010). Uwagi w sprawie zarządzania strategicznego w szkołach wyższych. In A. Kaleta, K. Moszkowicz (Ed.), *Zarządzanie strategiczne w praktyce i teorii*. Wrocław, PA: UE we Wrocławiu.

KEY TERMS AND DEFINITIONS

E-Learning Platforms: Platforms/systems used in teaching with the use of the Internet and of computer networks.

Electronic Documents Flow: IT system used for managing the flow of work and documents.

Gross Enrolment Ratio: Ratio between the number of students enrolled in tertiary education and the entire population of the age nominally assigned to tertiary education, expressed as a percentage.

Informatization: Process aimed at improving information systems, control systems and other systems through the use of computer equipment.

Integrated IT System: IT system supporting management, with modular or integral architecture.

Knowledge Networks in Higher Education

Filipa M. Ribeiro

University of Porto, Portugal

INTRODUCTION

The theme of knowledge creation is approached by focusing on the link between creation of knowledge and the relations among actors. This article aims at conceptualizing and discussing knowledge networks in the field of higher education.

BACKGROUND

The role of social networks for the creation of knowledge has been studied outside the educational field, highlighting the crucial role of formal and informal networks in organizational learning by stimulating new knowledge and new practices (Ahuja, 2000; McGrath and Krackhardt, 2003). However, less is known about the role of social networks in the field of education. In fact, up to this point there is only one book published on social networks and education (Daly, 2010).

A social network is a collection of individuals (commonly called actors) and an enumeration of the relations (or ties) among such individuals (Kindermann, 2008). The term social network is depicted from Barnes' work (1954), when he used it to designate the social relationships found in a community in Bremmes, Norway. Since then, the term has been associated to many different types of relations among many different types of individuals. Contemporary networks, unlike local communities, are not only centered on place-based affiliation, but more based on niche cultural affiliations and knowledge communities. These new ways of sharing culture and knowledge have broad implications on the relations between production and consumption and the traditional sources of authority for culture and knowledge. Standards

are continuously being reshaped as networks have become the dominant cultural logic (Varnelis, 2008). "Today, network culture succeeds postmodernism. It does so in a more subtle way. No new 'ism' has emerged: that would lay claim to the familiar territory of manifestos, symposia, definite museum exhibits, and so on" (Varnelis, 2008, p. 149). As it happens in other spheres, universities are made of networked actors¹ and, thus, the cultures that emerge are varied.

In this networked society, the creation and production of knowledge and expertise rises the likelihood that current knowledge will be retained and multiplied in new knowledge and practices. Recent educational studies stressed the importance of strong social networks among teachers for the spread and depth of policy, reform, innovation and change implementation (Coburn and Russel, 2008; Moolenaar, Daly and Slegers, forthcoming; Penuel, Frank and Krause, 2007, Brown and Duguid, 2000; Chiffoleau, 2005; Carre et al., 1989).

Forman and Markus (2005), Drejer and Jorgensen (Drejer & Jorgensen, 2005), and Hkopic et al. (2002), have studied knowledge creation and the role of collaboration. They identified the need for further research on social network characteristics relating to the creation of knowledge in a collaborative research environment. Also Drejer and Jorgensen (2005), and Hkopic et al. (2002), have observed the need for further research integrating the domains of social networking and knowledge creation. These researchers recognized that although collaboration and interdisciplinary research are often recommended, there is still a lack of empirical or theoretical research that validates the role of network sociology in the context of knowledge creation. Forman and Markus (2005) also recognize the value of an area of further

DOI: 10.4018/978-1-5225-2255-3.ch340

empirical quantitative exploration of their own existing qualitative research on this subject.

Moolenaar and Sleegers (2010) tried to find out more exactly to what extent the characteristics of teacher's social networks affect schools' innovative climate, when this is mediated by trust. The authors used a whole network approach, i.e., they focused on specific network characteristics, such as density, reciprocity and centralization, of the social network of the school team as a whole. Findings suggested that the density of the network related to work discussions was significantly associated with school's innovative climate and trust.

Diane Crane (1972) developed a seminal work in trying to understand where the knowledge learnt at universities come from. Who is responsible? Who should wield it? The author argues that the problem of the relationship between the internal structure of a particular cultural institution and the cultural products developed and accepted within has been neglected by the sociology of knowledge. The tendency to view social groups as abstract entities rather than as collections of individuals whose modes of interaction can be precisely observed was probably responsible for this gap. This task requires, as Diana Crane already pointed out back in 1972 (Crane, 1972), the analysis of the development of belief systems of these groups as well as sociometric analysis of the relationships between their members, of the relations between such groups and of the relations of such groups to the larger social structure. The development of social network analysis has been giving a relevant contribution to fill the gap identified by Crane.

In fact, the subsequent development of network theory represented an important contribution to deal with the issue of knowledge as it combined what is intuitively known with a growing body of network research suggesting that relationships within a system matter in enacting change, flows, diffusion strategies, implying both formal and informal networks of social relations that create nets of understandings, influence, and knowledge prior to, during and after any implementation of a change strategy (Daly, 2010).

Gamble and Blackwell (2001) defined knowledge as "a fluid mix of framed experience, values, contextual information, expert insight, and grounded intuition that provides an environment and framework for evaluating and incorporating new experiences and information. It originates and is applied in the mind of the knowers. In organizations it often becomes embedded not only in documents or repositories, but also in organizational routines, practices and norms". Embeddedness in social networks has been considered as a major cause of scientific achievement and scientists' behaviour (Crona & Parker, 2011; Gilsing, Nooteboom, Vanhaverbeke, Duysters, & Oord, 2008). The literature about academic knowledge focuses particularly on how actors' embeddedness within larger structures of co-authorship networks and collaborations in patents and projects is related to individual knowledge outcomes. Less is known about what properties of these networks affect knowledge creation. Knowledge creation studies have mainly focused on the influence of the networks on the efficacy and efficiency by which individuals transfer and apply knowledge (Botero & Cuartas, 2012; Cross, Parker, Prisak, & Borgatti, 2001; Fritsch & Kauffeld-Monz, 2008), but not on how they create new knowledge. This literature provides mixed evidence about the role of relational properties in knowledge creation and dissemination.

Knowledge creation on the other hand, and particularly social knowledge creation, refers to the knowledge that is increasingly created through interactions among the different members of a specific network (or a series of them). It is not only about input/output but rather about the relationship that the individual has with knowledge itself and how that relation is spread through his/her social relationships. Conceptually, creation is linked with creativity and it is known that creative people think and expresses themselves in a relational way (Quintás, 2003). Knowledge creation typically has the forms of ideas, practices, research papers and inventions (Phelps, Heidl, & Wadhwa, 2012), but also, we add, the capacity to create new

knowledge through thematic interrelations which, most times, are created according to the values and empathy among individuals and between individuals and their ideals. In turn, knowledge networks are usually defined as a set of actors who are repositories of knowledge and who create, transfer and adopt knowledge (Phelps et al., 2012). The social connections among these nodes are seen as channels and/or conduits of information and knowledge (Owen-Smith & Powell, 2004). These two definitions emphasise node and tie properties for knowledge creation but they also evidence a lack of understanding on the content of those networks. That is why, we argue that a definition of knowledge network encompasses both the network properties but also the network epistemics. Thus, *knowledge networks* can be defined as epistemic conduits of knowledge.

Networks function as an appropriate device to explore the processes of creation and adoption of knowledge by academics in higher education institutions (HEIs), and how it can be operationalised with the concept of epistemic authority and the analysis of knowledge networks. The claim that underlies this entry is that emergent processes of knowledge creation – in terms of epistemic states – are highly shaped by the social and knowledge networks in which academics are engaged. The primary focus of this approach to knowledge networks will be on knowledge creation. Thus, instead of focusing on the vehicles of distribution of knowledge and scientific outputs (Goldman, 1999), the emphasis will be on the role of knowledge networks – seen as epistemic conduits. This means social networks are the epistemic place where emergent processes of knowledge creation occur.

The claim that networks are epistemic conduits regarding the emergent processes of creation of knowledge also finds support on DiMaggio when the author points out that “intellectual and creative progress is characterized by strong positive network externalities, causing temporal and spatial concentration of significant movements”

and “institutional change often unleashes rapid network elaboration and an explosion of creativity” (DiMaggio, 2011). Also Burt (Burt, 2004) argues that unique occupancy of brokerage positions (‘structural holes’) triggers a wider exposition to novel ideas. Such boundary-spanners, in turn, generate more ideas. Thus, as well as Bourdieu views fields as sites of strategic competition between actors with different levels of cultural capital and economic resources, networks can be linked to epistemic changes in the knowledge created by academics. There are already several studies that support this claim. DeNooy (2003) argued that networks, which are themselves shaped by author’s social origins, form literary genres that double as author identities, which in turn influence social and aesthetic hierarchies. Also Collin’s (1989, 1998, 2000) asserts about how networks drive intellectual movements and he identifies four aspects of how that occurs: 1) intellectual and creative progress is characterized by strong positive network externalities that enhance those movements; 2) the most distinguished thinkers are densely connected to eminent peers; 3) reputational contagion; 4) movement leaders create new ties to extend their influence outward. Crossley (2009) identified similar themes in his analysis of networks that produced Britain’s punk music scene.

In sum, DiMaggio (2011) has argued that cultural products are produced by networks of collaborating creative professionals and organizations; that group and individual identity emerges from relations among people and many types of culture; and that meaning itself emerges from relations among symbols or other elements of texts, broadly defined. However, as the author said:

Longstanding habits of mind – the reluctance of many students of culture, in the social sciences as well as the humanities, to employ formal methods and the radical structuralism explicit in formative contributions to modern SNA – stood in the way of these developments (DiMaggio, 2011:296).

By claiming that knowledge networks of academics are epistemic conduits of their knowledge creation, the notion of epistemology of networks is central to much cultural and knowledge research. Networks are epistemic conduits by making knowledge interpretable and changeable. Thus, SNA can be used in this research to explore how knowledge emerges from the structure of relations among academics as knowledge creators.

Issues and Controversies

There is mixed evidence about the role of relational properties in knowledge creation and diffusion. The literature has stressed the importance of two related relational properties for the diffusion of knowledge in organizations: similarity and tie strength. With regard to the first, social or ethnic similarity (Baycan-Levent & Nijkam, 2010; Mügge, 2011; Rienties, Beausaert, Grohnert, Niemantsverdriet, & Kommers, 2012; Rienties, Hélot, & Jindal-Snape, 2013), similarity in personality and status, and geographic, institutional and organizational proximity (Whitley, 2008) have all been associated with knowledge outcomes, although the associations are ambiguous. On the one hand, similarity increases the ease of collaboration as it increases trust and reciprocity, but on the other hand, similarity lowers the possibilities of new knowledge creation. Furthermore, it is not known whether network similarity affects the motivation of the researcher to innovate and create new knowledge. The lack of understanding in this realm is due to the overemphasis of the role of the collective knowledge and social co-operation, but also on the lack of empirical studies with a deeper focus on the personal and interpersonal level of knowledge networks.

Second, studies have focused on the role of tie strength in knowledge creation and transfer. There is consistent evidence that strong ties (in terms of affinity, tie duration, and contact frequency) are more efficient for the transfer of knowledge by increasing trust and reciprocity norms, and that

they are more efficient in the transfer of certain types of knowledge, particularly complex, tacit knowledge and private knowledge (see Phelps et al., 2012 for a review). Up until now however, there is little agreement on the relative importance of weak and strong ties for knowledge creation. The discussion dates back to Granovetter (1973), who argued that weak ties are more useful for accessing diverse information, while strong ties are better for accessing influence. This distinction between information and influence may provide a clue to long-standing controversies about the relative efficacy of strong and weak network ties. Ruef (2002 and 2003) suggests that, in many cases, strong ties do not provide significant new information, so it isn't helpful to be embedded in clusters of strong ties. Most likely, it is a mixture of strong and weak ties that is necessary for knowledge creation. McFadyen et al. (2009) empirically demonstrate the effects of egonetwork density and average tie strength but also the interaction between these two factors on knowledge creation. They conclude that while strong ties do not provide diverse knowledge resources per se, they are efficient for capturing and using the diverse knowledge made available through sparse ego networks. Sosa (2007) found that strong ties that channel diverse knowledge also play a contributory role in the generation of creative ideas. A third way has been sustained by some authors who defend a complementary view between Burt's social capital theory (1992) and Coleman's theory (1988). Rost (2011), for instance, argues that weak network architectures (structural holes or a peripheral network positions) leverage the strength of strong ties in the creation of innovation.

In short, there is conflicting evidence about whether and under which circumstances tie strength and similarity affect individual knowledge creation (see Phelps et al., 2012).

In order to further explore under which conditions network ties are associated with knowledge creation and transfer, we argue that more attention should be paid to 1) knowledge creation itself;

for instance most literature on knowledge creation and diffusion is based on coauthorship and collaboration networks, but to what extent are these processes important for the creativity and relational creation of knowledge?; 2) the personal level, i.e., the multiple personal networks in which academics are embedded that shape the ways and reasons why they investigate what they investigate and how they actually create knowledge in a relational way. To conceive this, we combine a quantitative personal network approach with a qualitative approach. Our research on knowledge networks stems from a first premise that personal network research has been overlooked in the investigation into knowledge networks in general and even more in knowledge networks in the sector of higher education. The local network that is relevant for knowledge creation is likely larger than collaboration partners alone, and a personal network approach, based on multiple name generators, allows delineating the relevant network. A second premise claims that a deeper qualitative approach brings out benefits for the understanding of the creation and evolution of social networks in research communities. Although both approaches – structural and qualitative – are not new in this research area, to our knowledge, they have seldomly been combined. Therefore, phenomena like knowledge creation have been focused either by actor's attributes or by relational attributes, but separately. There is, thus, a need to integrate the two approaches, which we propose can be achieved by a mixed methods design that combines a structural and a content approach.

To resume, our research aims at going a step further in disambiguating the causal relationships and the content elucidation of knowledge networks, which implications for a deeper understanding of factors that may lead or not to a wider diversity in higher education institutions. In order to do so, this paper locates itself on the interpersonal level of knowledge creation and analyses the influence of tie strength and nodal similarity on knowledge creation.

CONCLUSION

A growing body of empirical research shows that social relationships and the networks these relationships play an important role in explaining the processes of creation, diffusion, absorption, and use of knowledge. Phelps et al. made an excellent review of that body of research.

First, personal network research has been overlooked in the investigation into knowledge networks in the sector of higher education - our empirical work shows a high correspondence between the personal knowledge networks and a permanent structural aspect of knowledge networks that is based on the role played by values on knowledge creation. It is to this personal networks that the respondents stick to no matter what the conjectural fluctuation (academic position shifts, criteria defined by journals, etc.) in their trajectory is. However, this is more due to the personal values attributed to specific ties than to dyadic similarity in other dimensions. Thus, scientific and personal values tend to be more appreciated as drivers for strongest ties than collaboration or expertise.

Secondly, the combination of a quantitative approach with a qualitative one proved to be suitable in the elucidation of the content and context of relations and of the link between actor's attributes and knowledge creation within a network.

FUTURE RESEARCH DIRECTIONS

Research on knowledge networks is still giving the first steps, thus there are plenty of research hypothesis. Knowledge networks research is concerned with how the nature of actors' social embeddedness influences their creation, transfer, and adoption of knowledge. Social embeddedness is, however, only one form or source of embeddedness (Dacin, Ventresca, & Beal, 1999). Other sources include informal cultural systems of meaning and formalized political-legal institutions (Dacin et al., 1999; Zukin & DiMaggio, 1990).

Regarding the network epistemics, one challenge worth exploring is how to leverage knowledge from other networks or from different networks. Our proposal intends to move a step further in understanding the links between network structure, perceptions, and actions in a dynamic field of interaction such as knowledge creation at universities.

REFERENCES

- Ahuja, M., & Carley, K. (1999). Network Structure in Virtual Organizations. *Organization Science*, 10(6), 741–747. doi:10.1287/orsc.10.6.741
- Apple, M. (2000). *Official Knowledge - democratic education in a conservative age*. London: Routledge.
- Barnes, J. A. (1954). Class and Committees in a Norwegian Island Parish. *Human Relations*, 7.
- Bourdieu, P. (2004). *Science of Science and Reflexivity*. Chicago: University of Chicago Press.
- Burt, R. L. (2004). structural holes and good ideas. *American Journal of Sociology*, 110(2), 349–399. doi:10.1086/421787
- Coleman, J. S. (1988). Social Capital in the Creation of Human Capital. *American Journal of Sociology*, 94, 95–120. doi:10.1086/228943
- Crane, D. (1972). *Invisible colleges: diffusion of knowledge in scientific communities*. Chicago: Chicago University Press.
- Crossley, N. (2009). The Man Whose Web Expanded: Network Dynamics in Manchester's Post-Punk Music Scene 1976–1980. *Poetics*, 37(1), 24–49. doi:10.1016/j.poetic.2008.10.002
- Daly, A. J. (2010). *Social Network Theory and educational Change*. Harvard University Press.
- De Grande, P. (2013). Aportes de Norbert Elias, Erving Goffman y Pierre Bourdieu al estudio de las redes personales. *Andamios*, 10(22), 237–258.
- De Nooy, W. (2003). Fields and networks: correspondence analysis and social network analysis in the framework of field theory. *Poetics*, 31(5), 305–327. doi:10.1016/S0304-422X(03)00035-4
- DiMaggio, P. (2011). Cultural networks. In Scott, J.; Carrington, P. J. (2011). *The Sage handbook of social networks analysis* (pp. 286–300). London: Sage.
- Drejer, I., & Jorgensen, B. H. (2005). The dynamic creation of knowledge: Analysing public-private collaborations. *Technovation*, 25(2), 83–94. doi:10.1016/S0166-4972(03)00075-0
- Etzkowitz, H., & Leydesdorff, L. (1997). *Universities and the Global Knowledge Economy: A Triple Helix of University-Industry-Government Relations*. London: Cassell.
- Forman, J., & Markus, M. L. (2005). Research on collaboration, business communication, and technology: Reflections on an interdisciplinary academic collaboration. *Journal of Business Communication*, 24(1), 78–102. doi:10.1177/0021943604271958
- Gallon, R., & Galés, N. (2015). Higher Education and globalization. In Ribeiro, Filipa M, Politis, Yurgos, Culum, Bojana (2015). *New Voices in Higher Education Research and Scholarship*. IGI Global.
- Goldman, A. (1999). *Knowledge in a Social World*. Oxford, UK: Clarendon Press. doi:10.1093/0198238207.001.0001
- Hadjimichalis, C. (1994). Global-Local conflicts: examples from southern Europe. In A. Amin & N. Thrift (Eds.), *Globalisation, Institutions and Regional Development in Europe*. Oxford, UK: Oxford University Press.
- Hlupic, V., Pouloudi, A., & Rzevski, G. (2002). Towards an Integrated Approach to Knowledge Management: Hard, Soft, and Abstract Issues. *Knowledge and Process Management*, 9(2), 90–102. doi:10.1002/kpm.134

- Johri, N., Ramage, D., McFarland, D., & Jurafsky, D. (2011). *A Study of Academic Collaborations in Computational Linguistics using a Latent Mixture of Authors Model*. Paper presented at the Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities.
- Kilduff, M., & Krackhardt, D. (2008). *Interpersonal networks in organizations*. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511753749
- Kindermann, T. (2008). Can we make causal inferences about the influence of children's naturally-existing social networks on their school motivation? In T. Little, N. Card, & J. Selig (Eds.), *Interdependent data in developmental psychology*. Mahwah, NJ: Erlbaum.
- Lin, N., Fu, Y., & Hsung, R. M. (2001). The position generator: measurement techniques for investigations of social capital. In N. Lin, K. Cook, & R. S. Burt (Eds.), *Social Capital: theory and research* (pp. 57–81). Hawthorne: Aldine de Gruyter.
- McGrath, C., & Krackhardt, D. (2003). Network conditions for organizational change. *The Journal of Applied Behavioral Science*, 39(3), 324–336. doi:10.1177/0021886303258267
- Moolenaar, N. M., & Sleegers, P. J. C. (2010). Social Networks, Trust and Innovation - The role of relationships in supporting an innovative climate in dutch schools. In A. J. Daly (Ed.), *Social Network Theory and educational change*. Cambridge, MA: Harvard Education Press.
- Muller, J. (2009). Forms of knowledge and curriculum coherence. *Journal of Education and Work*, 22(3), 205–226. doi:10.1080/13639080902957905
- Putnam, R. (1995). Bowling Alone: Americas Declining Social Capital. *Journal of Democracy*, 6(1), 65–78. doi:10.1353/jod.1995.0002
- Rawlings, C. M., & McFarland, D. A. (2011). Influence flows in the academy: Using affiliation networks to assess peer effects among researchers. *Social Science Research*, 40(3), 1001–1017. doi:10.1016/j.ssresearch.2010.10.002
- Santos, B. S. (2004). *A universidade do século XXI: para uma reforma democrática e emancipatória da universidade*. São Paulo Cortez Editora.
- Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. In B. Smith (Ed.), *Liberal education in a knowledge society* (pp. 76–98). Chicago: Open Court.
- Varnelis, K. (2008). *Networked Publics*. Cambridge, UK: Cambridge University Press. doi:10.7551/mitpress/9780262220859.001.0001
- Young, M. (1999). Multifocal educational policy research: Toward a method for enhancing traditional educational policy studies. *American Educational Research Journal*, 36(4), 677–714. doi:10.3102/00028312036004677

ADDITIONAL READING

- Daly, A. J. (2010). *Social Network Theory and educational Change*. Harvard Harvard University Press.
- Peters, M. A. (2013). *Education, Science and Knowledge capitalism- Creativity and the promise of openness*. New York: Lang Publishing. doi:10.3726/978-1-4539-0989-8
- Ribeiro, F. M., Politis, Y., & Culum, B. (2015). *New Voices in Higher Education Research and Scholarship*. Hershey, PA: IGI Global. doi:10.4018/978-1-4666-7244-4

KEY TERMS AND DEFINITIONS

Knowledge Networks: Usually knowledge network is the term given to different types of team or social networks and communities that are recognized to add significant value to the creation, dissemination and application of knowledge. In the scope of our research on processes of knowledge creation is a conceptual and structural device that reflects how individuals deal with problems, situations, and make sense of phenomena; they are the epistemic conduits by which circulates the know-how (and know-why) that individuals call on to accomplish their work.

Knowledge: A fluid mix of framed experience, values, contextual information, expert insight, and grounded intuition that provides an environment and framework for evaluating and incorporating new experiences and information. It originates and is applied in the mind of the knowers. In organizations it often becomes embedded not only in documents or repositories, but also in organizational routines, practices and norms (Gamble and Blackwell, 2001).

Networks Epistemics: Deals with the relationship between social structures of personal knowledge networks and the epistemic choices individuals make regarding the types of knowledge they privilege. Epistemic authority relies upon the

deliverance of experiences provided by the knowledge Networks in which researchers manifest in epistemic states that occur inside network ties.

Social Networks: Channels and/or conduits of information and knowledge. SN research concentrates mainly on 2 types of social networks that reflect different contents or resources flowing through ties. Instrumental social networks are conduits for the circulation of information and resources. These are the most studied. Then there are the expressive networks reflect patterns of more affect-laden relationships. By focusing on epistemics, we also consider meaning, social and personal context, offering the theoretical and methodological lenses through which SNT and sociology tries to answer questions related to the creation and epistemic authority of knowledge, in our case in higher education.

ENDNOTES

- ¹ According to Boyd (2007), there are four properties that separate unmediated publics from networked publics: persistence (they always aim a certain type of prosperity); searchability (identity is established through relations, texts, search and discovery); replicability and invisible audiences.

Quality Online Learning in Higher Education

Deborah G. Wooldridge

Bowling Green State University, USA

Sandra Poirier

Middle Tennessee State University, USA

Julia M. Matuga

Bowling Green State University, USA

INTRODUCTION

Traditional education systems alone, despite the essential role they have played and will continue to play in learning, are simply not capable of serving the world's growing and changing needs. The knowledge explosion, driven by the power of the network to connect people and spread ideas, has changed the very nature of learning. We must innovate and develop new modes of learning, both formal and informal, that meet the demands of the knowledge-driven economy in this Information Age. This chapter begins by identifying the technological changes that are affecting all societies and how these changes will specifically impact postsecondary education. The topic of course delivery within this chapter is viewed as a cultural issue that permeates processes from the design of an online course to the evaluation of an online course. This chapter will examine and review key components of, and tools for designing high impact online courses that support student learning and provide suggestions for faculty teaching online courses to assist in creating high quality online courses that supports teaching and, consequently, facilitates opportunities for student learning. Suggestions for conducting course evaluation and a feedback loops for continual improvement of the online learning and teaching will be addressed.

BACKGROUND

Technology broadly defined has been transforming human life in one way or another for thousands of years (Jerald, 2009). However, beginning in the 1990's, this technological change came at an exponentially faster rate due to factors such as increased competition in a global economy, automation, workplace change and policies increasing personal responsibility. As the world's labor markets evolve in the digital economy, we cannot predict what specific jobs will exist in the future, however what is clear is the shift from print to digital is a profound transition in how human beings learn (Pearson Learning, 2014). Currently, there are 84 million students enrolled in higher education worldwide. According to Ryan Craig (2014), "Global demand for higher education is forecasted to reach 160 million by 2025 – if online learning captures even half of this growth, there would be 40 million students requiring online education."

The advent of the personal computer, the Internet and the electronic delivery of information have transformed the world from a manufacturing, physically-based economy to an electronic, knowledge-based economy. Whereas the resources of the physically-based economy are coal, oil and steel, the resources of the new, knowledge-

based economy are brainpower and the ability to acquire, deliver and process information effectively. Ryan Craig (2015) in his new book titled *College Disrupted: The Great Unbundling of College Education*, has argued that “technology may bring more change to teaching and learning than college leaders anticipated.” Online learning will center the learning around students rather than the classroom, tailoring education to the needs and abilities of individual learners, and making life-long learning a practical reality for all (Balanko, 2002).

The global economic crisis and especially the unemployment of youth have prompted the urgency to develop educational systems that are aligned with the needs of the society it serves. Statistics from the United Nations indicate that one half of the global population is currently under the age of 25 years. The Organization of Economic Co-operation and Development (OECD, 2012) has examined this young population from its 33 member nations and concluded that 39 million or one in four 16-29 year olds were neither employed nor enrolled in some type of education or training program.

Those countries that invest in a 21st century education benefit immediately by transforming an outdated system to a more sustainable approach. Educators worldwide must develop challenging and relevant learning environments to prepare the future workforce of tomorrow (Beetham & Sharpe, 2013). Using digital education to connect students anywhere at any time has been touted as a viable option especially where access to post-secondary education is limited (Hosie & Schibeci, 2005). The Internet will “democratize” knowledge, increasing access, lowering the cost and improving the quality. (Moe, 2000). There should be no doubt that online learning is vital to all disciplines involved in education in the 21st century (Ternus, M. et al, 2007).

By 2020, there will be 55 million job openings in the United States. Sixty-five percent will require some postsecondary education. Our current system will fail to produce those skilled workers,

falling short by 5 million postsecondary credentials (Pearson Education, 2014). Ryan Craig (2015), also indicated in his book that the future of post-secondary education online degree programs will focus on “customizing course offerings where there is a true bottom line return for the majority of students in terms of graduation, employment, and wages. “A one-size-fits- all approach, Craig explains, “is no longer viable for the majority of students.” Packaged courses and degrees need to be transformed to smaller units, i.e., modules and micro-degrees allowing students more flexibility in their program of study.

A module is a set of independent units that can be used to construct an online course. An experiment at Massachusetts Institute of Technology is being conducted where students will develop an online course from parts they have assembled themselves online (Craig, 2015). This approach would allow students to retake any module where the student possibly struggled and not require the student to retake the entire course. Another approach is a micro-degree which offers online courses in high demand content areas in the job market. Generally, the online courses, which are an outgrowth of a MOOC, are shorter and allow the student to work at their own pace. A certificate is provided at the completion of the program.

It may first be helpful to highlight three myths and misconceptions about online teaching and learning. These myths/misconceptions are held by students, faculty, and administrators and influence any discussion about the quality of online course delivery (e.g., White, n.d.):

1. Online teaching and learning is ‘worse’ (or ‘better’) for meeting student learning outcomes than face-to-face courses.
2. Online teaching and learning is easier and more convenient for students and faculty than face-to-face courses.
3. Online teaching and learning is less interactive for both student and faculty than face-to-face courses.

All of these statements are based upon the premise that there are no special affordances or constraints of the environment, either online or face-to-face, when it comes to teaching or learning—that we are comparing apples to apples. It has been argued elsewhere that this is not the case, that there are many fundamental differences that the educational environment both affords or constraints (Anderson, 2004; Matuga, 2001, 2005 & 2007) and that establishing a dichotomistic relationship does not adequately reflect the complexity of teaching or learning within either environment. In essence, learning and teaching within online environments is fundamentally *different* than learning and teaching in face-to-face environments. One is not comparing apples to apples, but more like apples to oranges (Matuga, Wooldridge, & Poirier, 2011).

A useful concept to use as a framework, one that more adequately reflects the complexity of online teaching and learning, would be to view both through the lens of a cultural system. There are many definitions of culture and descriptions of what constitute a cultural system. LeVine (1984), for example, defined culture as “a shared organization of ideas that includes the intellectual, moral, and aesthetic standards prevalent in a community and the meanings of communicative actions” (p. 67). Others have claimed that these organizations of ideas and meanings derived from actions are not static and that culture should be thought of as systems that may be more complicated and organic collections of cognitive functions, practices, and meaning (D’Andrade, 2001; Giddens, 1984; Kitayama, 2002). Online teaching and learning may be viewed as cultural systems in that understandings and meanings are socially shared within online environments (Courtney, 2001; Melhinger & Powers, 2002). There are also cultural practices and customs within these environments that may be linked, in various ways, to the values and beliefs of larger cultural systems, like face-to-face educational environments (Courtney, 2001; LeVine, 1984).

A more traditional view of cultural systems, for example, is based upon the idea that they contain within them, nested systems that are interdependent to the functioning of the system as a whole (White, 1975). This view holds that technological (or physical subsystem), social, and psychological factors guide a multitude of functions and influence the behaviors of individuals that are participating in cultural communities (Kityama, 2002; White, 1975). For example, Kitayama (2002) stated that “each person’s psychological processes and structures are organized through the active effort to coordinate his or her behaviors with the pertinent cultural systems of practices and public meanings” (p. 92). While this may be viewed as a valid preposition, this view does imply that there are somewhat distinct sub-systems that guide or organize psychological processes and practices. It may be quite common to reduce discussions regarding online teaching and learning to cultural sub-systems, like technological ones, for example, because the impact of technology may be seen as more explicit within online teaching and learning environments.

This position is arguable in light of contemporary pedagogical theory, however, which would hold that psychological and social factors need to be explored in conjunction with technological ones. Perhaps one of the most critical characteristic of a cultural system is that they support the development and transmission of meaning and understanding within and between participants. Rosaldo (1984) stated that “we must appreciate the ways in which such understandings grow, not from an “inner” essence relatively independent of the social world, but from experience in a world of meanings, images, and social bonds, in which all persons are inevitable involved” (p. 139). In the case of meaning making and understanding within online teaching and learning environments, the importance of viewing the psychological in conjunction with social and technological dimensions of cultural systems is implied. These are important points that frame the conversation of

designing high impact online courses, providing suggestions for faculty teaching online courses, describing critical administrative support of faculty, and conducting course evaluation and feedback loops for continual improvement of online learning and teaching.

Designing High Impact Online Courses

Viewing online teaching and learning as a cultural system provides a framework to describe and understand ‘high impact’ online courses. Within education, ‘high impact’ refers to educational experiences that are meaningful, require student action and participation, and that contribute to the life-long learning of the student (Kuh, 2008). It is important to note that the examples illustrating high impact practices within online courses in this chapter do not represent an exhaustive or comprehensive list, but serve as important points of reference for discussion within this chapter. Two such high impact practices that assist in the design of online courses are pedagogical alignment and meaning making.

Pedagogical Alignment

Pedagogical alignment, also called systematic instructional design (Gagne, Briggs, & Wagner, 1992), entails aligning instructional variables to provide the fundamental framework for online cultural systems, conveying meanings to community participants, and defining cultural activity. Instructional design is currently in the midst of a paradigm shift towards a more situated view of design activity within cultural systems (Anderson, 2004; Derry & Lesgold, 1996; Mayes & de Freitas, 2013). When designing high impact online courses, pedagogical alignment involves the optimum use of a wide array of instructional features including, but not limited to, instructional goals, instructional strategies, and assessment measures and evaluation practices that support teaching and learning (see Matuga, 2005).

The issue of pedagogical alignment in an online course is an important one, for several reasons. While in a face-to-face environment, instructional design and alignment may occur in conjunction with the other during instruction, alignment in an online course, in contrast, is often completed *a priori*. In fact, a clear, comprehensive, and logical course structure may be one of the primary factors which students use to judge whether or not an online course is an effective one. Online course alignment is further complicated by the affordances and constraints stemming from technology and the social and psychological composition of community participants. Each instructional variable is influenced by the affordances and constraints that influence technological, social, and psychological subsystems. For example, discussions regarding effective online course alignment must also incorporate pedagogical and technological expertise of students and teachers in the design process (Chang, Kurcz, El-Bishouty, & Graf, (2015).

There may be several methods in which to assess whether or not pedagogical alignment was effective, more often than not, discussions regarding the effectiveness of online activities are reduced to formal, summative assessment practices such as tests, quizzes, projects, or portfolios. Pedagogical decisions regarding the selection of formal, summative assessments within online courses are important, especially when designing instructional strategies (see Duffy & Cunningham, 2001). However, it is equally important to plan for the manner in which formative assessments and other evaluative information will be utilized to inform other educational practices within the online environment. Teacher and student self-reflection, as a formative, informal assessment to inform learning is a potentially powerful tool for evaluating pedagogical effectiveness.

Another tool to assess pedagogical alignment is peer evaluation of the course itself. There may be many mechanisms and tools that faculty and administration can use that carry out the peer evaluation of online courses. One is a tiered

method at the local level asking more experienced faculty or if available, online instructional designers within the institution to review and constructively evaluate the pedagogical alignment of a particular course. Another method may be to request the course be formally evaluated by peers through organizations such as Quality Matters. Quality Matters is a peer review system of online courses utilizing a rubric that examines pedagogical alignment. While there are costs involved in the later, the former would take some due diligence on the part of the faculty members involved but would one step towards assuring quality in online course design.

Meaning Making

Cultural meaning making is a complex activity tied to the cultural systems in which they are created and shared. One concept that is interesting to explore and is of particular importance to the development of quality online courses is the concept of intersubjectivity. Intersubjectivity is a term associated with Vygotskian theory of cognitive development and refers to the shared cultural understanding between two people (Rogoff, 1990; Rosaldo, 1999; Wertsch, 1985). If there is no shared understanding between two people, for example a teacher and a student, then attempts to communicate, create meaning, and establish understanding are somewhat fruitless endeavors. If the teacher is unable to understand the misconceptions, questions, or understanding of his or her students, then that teacher will not be able to address the student needs and learning will be hampered. It is in this manner that the constraints of an online environment may present challenges to establishment of intersubjectivity (Anderson, 2004). There is the suggestion, that like culture, some meanings may be more explicit due to the “pragmatics of social life and their history for a given society” but not all meanings may be “reduced to its explicit or implicit dimensions” (Le Vine, 1984, p. 77). Social referencing (i.e., gesture, gaze, and other nonverbal communication cues)

which is critical for establishing intersubjectivity (Rogoff, 1990) in a face-to-face environment is not impossible to establish in an online environment, it does, however, take a more concentrated effort in high impact online courses.

The culture of online teaching and learning may share similar characteristics with traditional, face-to-face teaching and learning and there may also be new territory for future investigations. The critical investigation of how teachers and learners understand, navigate, and utilize the culture of online learning environments to become competent participants is critical to the future of designing high impact courses. As Gardner (1984) has stated, “the human being, who, starting from a state of total ignorance about his or her particular culture, must within a decade or two acquire sufficient competence so that he or she can carry out productive work and interact effectively with other individuals to achieve valued ends” (p. 261). The enculturation of teachers and students within online teaching and learning environments requires learning how to use psychological tools within similar, yet dramatically different cultural systems, systems that have different environmental affordances and constraints.

Suggestions for Faculty Teaching Online Courses

There are many resources, books and websites, to assist faculty when teaching high quality online courses (e.g., Cooper, 2015; Eison, 2010; Ko & Rossen, 2001; Koszalka & Ganesan, 2004; Masoumi & Lindström, 2012; Nedungadi, & Raman, 2012). There is also assistance to help faculty effectively utilize discussion boards, online tools, blogs, wikis, open source programs, iPads, iPods, the iGoogle suite, cell phones, etc. within their online courses. Returning to the framework of viewing online courses as cultural systems, two inter-related important considerations are outlined within this chapter for faculty teaching high quality online courses: rituals of participation and co-regulation.

Rituals of Participation

One of the ways in which individuals navigate technological, social, and psychological systems are what Courtney (2001) called “rituals of participation” (p. 236). These rituals for participation encompass the norms and behaviour for participating in cultural activities. Learning how to “do school,” or mastering implicit and explicit academic and social knowledge needed to be successful in school, is an important ritual of participation required for effective teaching and learning (Westby, 1997). This issue is critically important due to the alarming drop rates from online courses by students (Diaz, 2002). Two popular explanations of why online students fail to complete online courses seem to be individual (demographic and/or learning style) or performance (low) differences (Diaz, 2002). However, researchers have suggested that the issue of student preparedness for online classes may be more complicated and include a variety of factors including student, situational, and educational factors (Gibson, 1998).

There may be cross cultural interference between how students conceptualize “doing school” within face-to-face environments and how this concept is challenged and, by necessity, altered within online teaching and learning environments. For example, one affordance within online environments is the ability, on the part of the faculty, to “hear” what every student thinks about a particular subject by requiring all students to post on a discussion board, for example. This, of course, requires that each student contributes to the conversation in a concrete and physical way that reflects what was read, in this there is simply a lot of individual accountability and challenges what a student may view of ‘doing school’. Many students, it could be argued, have learned what it means to “do school” by either engaging in discussion or gaining enough inference from what others are discussing in class to participate adequately, or simply keep quiet during discussion.

There may be several strategies to help online students assimilate to “doing school” online. Per-

haps the most utilized strategy is to give student a pre-assessment evaluating certain skills that are needed to be a successful online student. Many universities now utilize some sort of instrument that asks students a variety of questions about their learning habits to find a “goodness of fit.” For example, the University of Georgia uses the Readiness for Education at a Distance Indicator (READI @ <http://goml.readi.info/>) which is a self-assessment in which the student evaluates him- or herself on a variety of indicators like: life factors, personal attributes, learning styles, reading rate and recall, technical competency, technical knowledge, and typing speed and accuracy. There may be other ways to address the issue of student readiness within online courses, including scaffolding student learning at the beginning of the class and providing assistance with establishing regular activities and other strategies mentioned in this chapter. Another potential powerful strategy is assigning student-student pairs or each student to a small group and requiring they participate in activities in which they form a bond or rely upon each other to complete course tasks. For example, you may ask groups to define expectations about group standards of behaviour or something a little more fun like identifying a group name, wiki, or mascot. All of these activities are socially binding acts that may help students from dropping out if they feel they have more connections and support for learning in online environments.

Co-Regulation

Navigating online teaching and learning, or becoming an efficient online teacher and/or online learner, requires that an individual is able to adequately use processes, strategies, and responses to plan and monitor his or her participation in pedagogical activity (Zimmerman, 2001, 1994). This concept, known as self-regulation, reflects those processes that occur at an individual level that play an important role in student academic achievement (Zimmerman, 1994). In many cases, online students are ill-prepared for online courses

or drop out of online courses due to their inability to regulate (i.e., plan or monitor) their own learning. Ironically, this is also related to the convenience myths/misconceptions of online teaching and learning—a teacher or student need not go to a bricks-and-mortar classroom at a prescribed time, one can learn anytime and anyplace. In reality, it is very difficult for some students to be self-regulated enough to complete a course online.

McCaslin and Hickey (2001) proposed, however, that co-regulation is a more appropriate concept when discussing regulation of the teaching and learning process within socio-cultural contexts from a Vygotskian perspective (Courtney, 2001). In reference to this position, Zimmerman (1986) stated that “self-regulation is not an idiosyncratic product of the child’s own discovery experiences; but rather, it is a culturally transmitted method for optimizing and controlling learning events” (p. 311). The processes of regulating teaching and learning does not fall simply on the isolated individual, but is shared between and among students and teachers. In this sense, the self-regulatory functions of the student are influenced by others within the socio-cultural environment (i.e., peers and the teacher), just as others’ self-regulatory functions are influenced by that individual student.

Another aspect of co-regulated learning is monitoring strategies. Self-monitoring strategies refer “to students’ efforts to observe themselves as they evaluate information about specific personal processes or actions that affect their learning and achievement in school” (Zimmerman & Paulsen, 1995, p. 14). Students who have effective self-monitoring strategies are able to evaluate their own progress towards an established goal, making appropriate strategy changes as they proceed to regulate their learning effectively (Zimmerman & Paulsen, 1995).

The ability to effectively monitor one’s own learning processes are also dependent upon a wide array of technical, social, and psychological variables (see Zimmerman & Paulsen, 1995). Let us briefly discuss self-monitoring of the writing process as an example. Writing the “old way,” in

other words before word processing was pretty rudimentary and in retrospect seems almost primitive. An individual first had to prepare quite thoroughly before writing, often generating a detailed outline and/or completing a rough draft in long hand, before moving to type the paper on a typewriter. There was a certain diligence required on the part of the individual for if a mistake was made, it could necessitate starting the entire process all over (unless your professor did not mind a lot of liquid paper). The process described was necessary because there were constraints associated with the technological tool that was utilized to write (i.e., typewriters...an electric one if you were lucky). The process, itself also served as a self-regulatory strategy that monitored the writing process. The invention and utilization of word processors has, however, altered the procedural script of writing used by students and, essentially, made obsolete a monitoring strategy used for writing.

There are three primary suggestions for faculty when teaching high quality online courses. First, it is important for faculty to be explicit about how to “do school” online and ways to assist in the co-regulation of student learning in your course. One promising practice in online courses is to keep the patterns of behavior or interactions similar throughout the duration of the course. For example, always ‘open’ online course discussions on Mondays and students are expected to have their points posted by Friday. Second, faculty should plan for activities to help students identify and address the ways in which there may be differences in how to “do school” in their online course. Faculty should also assist students in how they plan to work their online course requirement into their calendar; this is also referred to as an ‘orienting activity’ (see Olgren, 1998). Third, faculty should provide guidelines for activities and assessments that address both issues, rituals of participation and co-regulation, to help scaffold these skills throughout the duration of the course. Scaffolding simply means that the faculty member should provide a lot of assistance with these activities at the beginning of the course

and as the course progresses; the faculty member slowly withdraws that help as the student gains confidence and masters course material.

Administrative Support of Faculty

Quality online course delivery is contingent upon administrative support of faculty. Additionally “administrators need to understand their faculty population if they are to support faculty participation in [online teaching and learning]” (Schifter, 2004, p. 25). However, according to Jorge Gaytan (2009) while all valued online teaching and learning there was little agreement among deans, vice presidents for academic affairs, and administrators of distance learning at campuses regarding the organizational structures that would support online teaching and learning. It could be that online teaching and learning as a field has, traditionally, been focused upon the pedagogy of teaching and learning that takes place within online learning environments and contributes to institutional culture. Two important factors that face administrative support of faculty and delivering quality online courses and programs are adequate support for planning and faculty workload.

Adequate Support for Planning

Osika (2006) warns that there are many administrative issues that need to be addressed within and across institutions to provide support for online students and faculty beyond the borders of the virtual learning environment. The issue of adequate faculty support, however, is perhaps the most examined administrative issue in distance learning literature (see Ko & Rossen, 2001). Levy (2003) outlined critical factors for administrators to take into account when planning for online courses and/or programs; she proposed that planning be systematic and strategic. Levy (2003) stated that “the challenge to colleges in the 21st century is not to decide why they should have an online distance learning program, but to decide how to design and implement such a program” (p. 3).

Effective and innovative leadership plays an important role in the development and delivery of online course and programs (see Latchem & Hanna, 2001). Beaudoin (2003) stated that “any focused consideration of the dimension of leadership and its impact on the growth and apparent success of distance education at literally hundreds of institutions worldwide” has been largely absent from the literature (p. 3). There appear to be some disconnect, however, between the roles that faculty and administrators play in the development of quality online courses and programs and this disconnect is reflected recent studies on the topic. One of the first steps for administrators to support online students, faculty, and programs would be to understand the demands on online faculty when designing and planning for online courses and programs. Administrators should clarify, in discussions with their faculty, any misconceptions about developing and offering a high quality online course/program. For example, Gaytan (2009) found that there was a disconnect between administrators’ rhetoric and practice and that there was still, on the part of administrators, “an emphasis is on cost savings, remaining competitive, and delivery of information as opposed to instructional quality” (p. 69).

Howell, Williams, and Lindsay (2003) stated that there is a need to pose difficult questions about online programs, but a need to address those questions “from an informed perspective” (p. 1). As Gaytan (2009) found, there were many instances in which administrators and faculty differed in important ways when discussing online teaching and learning. For example, “while online education coordinators and faculty [in his study] were thinking about the ways to improve the quality of online education, academic administrators had other priorities such as being able to remain competitive” (Gaytan, 2009, p. 69). This illustrates a common assumption, on the part of institutional administrators, that online courses will address the need to service more students for the same costs, often by increasing the number of students within online courses (Concieção &

Baldor, 2009). As stated in the beginning of the chapter, there are also myths and misconceptions about online teaching and learning that may be held by both administrators and faculty. Assumptions, myths, and misconceptions need to be examined and discussed by administrators and faculty members within the context and characteristics (i.e., pedagogical alignment) of a particular course or program. This is a critical first step during the design and planning stage of online courses and programs. Administrative support of faculty for high quality online courses needs, however, to extend beyond support to encompass adequate planning (Osika, 2006).

Faculty Workload

One issue that has received some attention in the online teaching and learning literature has been that of faculty workload. Lehmann and Chamberlin (2009), for example, illustrate the number and variety of hats that online faculty wear when teaching a high quality online course: (1) teacher, (2) facilitator, (3) instructional technologist, (4) course designer, (5) writer, editor, and proofreader, (6) counselor and mediator, and (7) advisor and registrar. Wearing more hats, however, may not necessarily result in increased compensation for the online faculty member. For example, in Schifter's (2004) research surveying administrators about online teaching and learning issues, both found that there was little consistency with faculty compensation. Additionally, faculty may not be prepared to wear many of those hats and professional development, mentoring programs, or providing other models to those faculties may be critical (Blythe, 2001).

The complications and added workload for the online faculty member is debated in the research literature. Some researchers have found that more students in an online course increases workload due to more student-teacher interactions and/or more individual feedback on assessments, particularly if a pedagogical alignment is used for the course that is more student-centered (Matuga, 2005). Other

researchers, however, have found that placing additional students within an online course may not increase faculty workload substantially, especially if the online faculty member has previous online teaching experience or the course's pedagogical alignment is more teacher-centered and does not require a lot of student-teacher interaction and student monitoring (Anderson & Avery, 2008; DiBiase & Rademacher, 2005; Matuga, 2005). It is very important, however, to realize that many of the studies investigating faculty workload were comparison studies with face-to-face faculty. As stated in the introduction of this chapter, this premise may be flawed as we are not comparing apples-to-apples therefore, investigating faculty workload from a comparison perspective may not adequately reflect what an online faculty *does*.

Just as it may be more accurate to examine online teaching and learning as a cultural system, it may be necessary to redefine the roles, responsibilities, and what it means to be an online faculty member. The supposition that online faculty is treated differently at educational institutions, and this may be a threat to academic quality, has been a recurring theme in the research literature (Concieção & Baldor, 2009). Schifter (2004) found that faculty workload for those teaching online courses were not to consistent within and across institutions (Schifter, 2004). This has given rise to a trend at educational institutions of hiring online adjunct faculty to teach online courses (Puzziferro & Shelton, 2009). On the one hand, while this group of faculty may have highly specialized skills to teach online there may still be a perceived threat to the online quality of courses and programs. Regardless, the rise of the online adjunct faculty illustrates the need for an examination of knowledge and skills required to be an effective online faculty member and support from administration for quality online course delivery. It also necessitates the transformation of how we assess and evaluate effective online faculty, especially if we look at online learning environments as complex, cultural systems. A cornerstone of support for online faculty from

an administrative perspective should include support for the systematic assessment and evaluation of online courses with an eye toward continual improvement.

FUTURE RESEARCH DIRECTIONS

The last issue explored in this chapter, which of continual improvement, encompasses the various aspects for all the topics features highlighted in this chapter: pedagogical alignment, meaning making, rituals of participation, co-regulation, administrative support for faculty planning courses/programs, and faculty workload. The relationships between and among these different factors illuminate themselves to be reciprocal in nature through the process of continual improvement. It is in this sense that online teaching and learning contributes to the notion of cultural systems discussed at the beginning of the chapter.

Continual improvement is the act of reflecting on the effectiveness of pedagogical alignment within the context of the constraints and affordances of the online teaching and learning environment. Perhaps the most important component for continual improvement, for example, is pedagogical alignment which, as stated earlier, requires that instructional variables like student learning outcomes, learning activities, and assessment and evaluation practices, that support student success and learning, be selected with care during the planning process (Matuga, 2005). Pedagogical alignment provides a blueprint for continual improvement. However, continual improvement also mandates the re-visitation of pedagogical alignment during and after the course has been taught with the goal of perfecting the course with an eye towards assuring quality.

Sims and Jones (2002) proposed a three-phased model for continuous improvement in online classes that emphasizes the importance of pedagogical alignment and instructional design. They proposed that there be a “Pre-Delivery” phase in which the focus is on the functionality of the

newly designed course be peer evaluated. This could be by utilizing some of the peer-evaluation methods mentioned in this chapter. The second phase is that of an initial delivery, or enhancement stage, and the last stage is an ongoing delivery or course maintenance phase in which the course is modified based upon feedback from students and peers. This approach emphasizes a team approach to course design, evaluation, and redesign. This is also called within socio-cultural literature creating ‘communities of practice’. Sims and Jones (2002) outline that building shared understanding, establishing rituals of participation, and maintaining communication are all critical components of their continuous improvement model. Continuous improvement is by its definition, a dynamic process that examines those shared ideas of what is quality online teaching and learning and is essential for assuring quality.

CONCLUSION

As a review of the online teaching and learning research by Tallen-Runnels and colleagues (2006) illustrated, online teaching and learning can be and has been a field typically organized by following four categories: course environment, learners’ outcomes, learners’ characteristics, and institutional and administrative factors. There are two issues when designing online courses, according to Ternus et al, (2002), “The first is the quality of teaching tool and the second is the quality of learning that takes place. In this chapter we have presented a foundation for viewing some of these topics through the lens of cultural systems.”

Three common myths/misconceptions regarding online teaching and learning were presented at the beginning of this chapter to help frame the discussion presented. One last myth or misconception that was not highlighted in this paper is that online teaching and learning is less expensive than face-to-face teaching. Divorcing this myth from the comparison (apples to apples)

metaphor, this is a topic that was touched upon when the subject of administrative support of faculty was discussed, but not fully examined. Colleges and universities are pressured into developing online courses and programs, in essence, to meet economic shortfalls. However, as Levy (2003) stated “the challenge to colleges in the 21st century is not to decide why they should have an online distance learning program, but to decide how to design and implement such a program” (p. 3). The subject or relationship between how much a course/program cost and quality of learning within that course/program remains a very complex and difficult topic to frame within the topic of assuring quality.

The issue of assuring quality in online courses and programs is a multifaceted one that requires we look at the online teaching and learning environment with a fundamentally different lens, that as a complex, cultural system with unique affordances and constraints. While an exhaustive list is not presented here, important factors that influence the design of high impact online courses, make suggestions for faculty who are teaching online classes that ensure quality, impact administrative support of online faculty, and offer a potential model for continual improvement. All of these factors are aligned with the central premise of this chapter that assuring quality in online course delivery is a complicated communal activity focusing on the alignment of sound instructional features, meaning-making, working together with students to develop new ways ‘to do school’ online, recognition of faculty work, and continual improvement.

REFERENCES

- Anderson, K. M., & Avery, M. D. (2008). Faculty teaching time: A comparison of web-based and face-to-face graduate nursing courses. *International Journal of Nursing Education Scholarship*, 5(1), 1–12. doi:10.2202/1548-923X.1539 PMID:18241197
- Anderson, T. (2004). Toward a theory of online learning. In T. Anderson & F. Elloumi (Eds.), *Theory and Practice of Online Learning* (pp. 33–60). Athabasca, Canada: Athabasca University. Retrieved from http://cde.athabascau.ca/online_book/index.html
- Balanko, S. L. (2002). *Review and Resources: Online Education Implementation and Evaluation*. University of Washington Office of Educational Assessment, OEA Report 02-11. Retrieved from <https://www.washington.edu/assessment/files/2015/01/OEAReport0211.pdf>
- Beaudoin, M. F. (2003). Distance education leadership for the new century. *Online Journal of Distance Learning Administration*, 6(2).
- Beetham, H., & Sharpe, R. (2013). *Rethinking pedagogy for a digital age: Designing for 21st century learning* (2nd ed.). New York: Routledge.
- Blythe, S. (2001). Designing online courses: User-centered practices. *Computers and Composition*, 18(4), 329–346. doi:10.1016/S8755-4615(01)00066-4
- Carlson, S. (2012). The Future of Colleges May Lie, Literally in Students’ Hands. *The Chronicle of Higher Education*, (February), 5.
- Chang, T. W., Kurcz, J., El-Bishouty, M. M., & Graf, S. (2015). Roadmap for adaptive and personalized learning. In *Ubiquitous Learning Environments and Technologies* (pp. 1-12). Berlin, Germany: Springer.
- Conceição, S. C. O., & Baldor, M. J. (2009). *Faculty workload for online instruction: Individual barriers and institutional challenges*. Midwest Research-to-Practice Conference in Adult, Continuing, and Community Education, Chicago, IL.
- Cooper, K. (2015). Breaking the mold. *Diverse Issues in Higher Education*, 32(7), 16–17.
- Courtney, S. (2001). Technology and culture of teaching and learning. In D. Lieberman & C. Wehlberg (Eds.), *To Improve the Academy* (Vol. 19, pp. 232–249). Bolton, MA: Anker Publishing.

- Craig, R. (2015). *College Disrupted: The Great Unbundling of Higher Education*. New York, NY: Palgrave Macmillan Press.
- DAndrade, R. (2001). A cognitivists view of the unit debate in cultural anthropology. *Cross-Cultural Research*, 35(2), 242–257. doi:10.1177/106939710103500208
- Derry, S., & Lesgold, A. (1996). Toward a situated social practice model for instructional design. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of Educational Psychology* (pp. 787–806). New York: Simon & Schuster Macmillan.
- Diaz, D. P. (2002). Online drop rates revisited. *The Technology Source Archives*. University of North Carolina. Retrieved from http://technologysource.org/article/online_drop_rates_revisited/
- DiBiase, D., & Rademacher, H. (2005). Scaling up: Faculty workload, class size, and student satisfaction in a distance learning course on Geographic Information Sciences. *Journal of Geography in Higher Education*, 29(1), 139–158. doi:10.1080/03098260500030520
- Duffy, T. M., & Cunningham, D. J. (2001). Constructivism: Implications for the design and delivery of instruction. In D. H. Jonassen (Ed.), *Handbook of research for educational communications and technology* (pp. 170–198). Mahwah, NJ: Erlbaum.
- Eison, J. (2010). *Using active learning instructional strategies to create excitement and enhance learning*. Retrieved on September 24, 2015 from <http://www.cte.cornell.edu/documents/presentations/Eisen-Handout.pdf>
- Gagne, R. M., Briggs, L. J., & Wagner, W. W. (1992). *Principles of instructional design*. Fort Worth, TX: Harcourt Brace Publishers.
- Gardner, H. (1984). The development of competence in culturally defined domains: A preliminary framework. In R. A. Shweder & R. A. LeVine (Eds.), *Culture theory: Essay on mind, self, and emotion* (pp. 257–275). New York: Cambridge University Press.
- Gaytan, J. (2009). Analyzing online education through the lens of institutional theory and practice: The need for research-based and –validated frameworks for planning, designing, delivering, and assessing online instruction. *Delta Pi Epsilon Journal*, 51(2), 62–75.
- Gibson, C. C. (1998). The distance learner's academic self-concept. In C. Gibson (Ed.), *Distance learners in higher education: Institutional responses for quality outcomes* (pp. 65–76). Madison, WI: Atwood.
- Giddens, A. (1984). *The constitution of society*. Oxford, UK: Polity Press.
- Hosie, P., & Schibeci, R. (2005). Checklist and context-bound evaluations of online learning in higher education. *British Journal of Educational Technology*, 36(5), 881–895. doi:10.1111/j.1467-8535.2005.00513.x
- Howell, S. L., Williams, P. B., & Lindsay, N. K. (2003). Thirty-two trends affecting distance education: An informed foundation for strategic planning. *Online Journal of Distance Learning Administration*, 6(3).
- Jerald, C. D. (2009). *Defining a 21st Century Education*. Center for Public Education. Retrieved from http://www.cfsd16.org/public/_century/pdf/Defininga21stCenturyEducationJerald_2009.pdf
- Kitayama, S. (2002). Culture and basic psychological processes—toward a system view of culture: Comment on Oyserman et al. *Psychological Bulletin*, 128(1), 89–96. doi:10.1037/0033-2909.128.1.89 PMID:11843550

- Ko, S., & Rossen, S. (2001). *Teaching online: A practical guide*. Boston: Houghton Mifflin.
- Koh, C. (2015). Understanding and facilitating learning for the net generation and twenty-first-century learners through motivation, leadership and curriculum design. In C. Koh (Ed.), *Motivation, Leadership And Curriculum Design: Engaging The Net Generation and 21st Century Learners* (pp. 1–10). Singapore: Springer. doi:10.1007/978-981-287-230-2_1
- Koszalka, T. A., & Ganesan, R. (2004). Designing online courses: A taxonomy to guide strategic use of features available in course management systems (CMS) in distance education. *Distance Education*, 25(2), 243–256. doi:10.1080/0158791042000262111
- Kuh, G. D. (2008). *High-impact educational practices: What they are, who has access to them, and why they matter*. Washington, DC: Association of American Colleges and Universities.
- Latchem, C., & Hanna, D. E. (2001). Leadership in open and flexible learning. In C. Lathem & D. E. Hanna (Eds.), *Leadership for 21st Century Learning: Global Perspectives from Educational Innovators* (pp. 53–62). London: Kogan Page Limited.
- Lehmann, K., & Chamberlin, L. (2009). *Making the move to elearning: Putting your course online*. Lanham, MD: Rowman & Littlefield.
- LeVine, R. A. (1984). Properties of culture: Ethnographic view. In R. A. Shweder & R. A. LeVine (Eds.), *Culture theory: Essay on mind, self, and emotion* (pp. 67–87). New York: Cambridge University Press.
- Levy, S. (2003). Six factors to consider when planning online distance learning programs in higher education. *Online Journal of Distance Learning Administration*, 6(1).
- Masoumi, D., & Lindström, B. (2012). Quality in e-learning: A framework for promoting and assuring quality in virtual institutions. *Journal of Computer Assisted Learning*, 28(1), 27–41. doi:10.1111/j.1365-2729.2011.00440.x
- Matuga, J., Wooldridge, D. G., & Poirier, S. (2011). Assuring Quality in Online Course Delivery. *International Journal of Adult Vocational Education and Technology*, 2(1), 36–49. doi:10.4018/javet.2011010104
- Matuga, J. M. (2001). Electronic pedagogical practice: The art and science of teaching and learning on-line [Special issue]. *Journal of Educational Technology & Society*, 4(3), 77–84.
- Matuga, J. M. (2005). The role of assessment and evaluation in context: Pedagogical alignment in online courses. In D. D. Williams, S. L. Howell, & M. Hricko (Eds.), *Online Assessment, Measurement and Evaluation* (pp. 316–330). Hershey, PA: Information Science Publishing.
- Matuga, J. M. (2007). Self-regulation and online learning: Theoretical issues and practical challenges to support life-long learning. In Y. Inoue (Ed.), *Online Education for Lifelong Learning* (pp. 146–168). London: Information Science Publishing. doi:10.4018/978-1-59904-319-7.ch007
- Mayes, T., & de Freitas, S. (2013). Technology-enhanced learning: The role of theory. In H. Beetham & R. Sharpe (Eds.), *Rethinking pedagogy for a digital age: Designing for 21st century learning* (2nd ed., pp. 17–30). New York: Routledge.
- McCaslin, M., & Hickey, D. T. (2001). Self-regulated learning and academic achievement: A Vygotskian view. In B. J. Zimmerman & D. H. Schunk (Eds.), *Self-regulated learning and academic achievement: Theoretical perspectives* (pp. 227–252). Mahwah, NJ: Erlbaum.

- Mehlinger, H. D., & Powers, S. M. (2002). *Technology and teacher education: A guide for educators and policymakers*. Boston, MA: Houghton Mifflin Company.
- Moe, M. T., & Blodget, H. (2000). *The knowledge web: Part 1. People power: Fuel for the new economy*. New York: Merrill Lynch.
- Nedungadi, P., & Raman, R. (2012). A new approach to personalization: Integrating e-learning and m-learning. *Educational Technology Research and Development*, 60(4), 659–678. doi:10.1007/s11423-012-9250-9
- Olgren, C. H. (1998). Improving learning outcomes: The effects of learning strategies and motivation. In C. C. Gibson (Ed.), *Distance learners in higher education: Institutional responses for quality outcomes* (pp. 77–95). Madison, WI: Atwood Publishing.
- Organization of Economic Co-operation and Development. (2012). *Better Skills, Better Jobs, Better Lives: A Strategic Approach to Skills Policies*. OECD Publishing. Retrieved from http://www.keepeek.com/Digital-Asset-Management/oecd/education/betterskills-better-jobs-better-lives_9789264177338-en#page4
- Osika, E. (2006). The concentric support model: A model for planning and evaluation of distance learning programs. *Online Journal of Distance Learning Administration*, 9(3). Retrieved from <http://www.westga.edu/~distance/ojdla/fall93/osika93.htm>
- Pearson Education. (2014). *The Shift to Digital Learning: Key Indicators of Success*. Retrieved from <https://www.pearsoned.com/wp-content/uploads/ShiftToDigitalInfographic.pdf>
- Puzziferro, M., & Shelton, K. (2009). Supporting online faculty – Revisiting the seven principles (A few years later). *Online Journal of Distance Learning Administration*, 12(3). Retrieved from <http://www.westga.edu/~distance/ojdla/fall123/puzziferro123.html>
- Rogoff, B. (1990). *Apprenticeship in thinking: Cognitive development in social context*. New York: Oxford University Press.
- Rosaldo, M. Z. (1984). Toward an anthropology on self and feeling. In R. A. Shweder & R. A. LeVine (Eds.), *Culture theory: Essay on mind, self, and emotion* (pp. 137–157). New York: Cambridge University Press.
- Schifter, C. (2004). Faculty participation in distance education programs: Practices and plans. In D. Monolescu, C. C. Schifter, & L. Greenwood (Eds.), *The distance education evolution: Issues and case studies* (pp. 22–39). Hershey, PA: Information Science Publishing. doi:10.4018/978-1-59140-120-9.ch002
- Sims, R., & Jones, D. (December, 2002). *Continuous improvement through shared understanding: Reconceptualizing instructional design for online learning*. ASCILITE 2002. Auckland, New Zealand.
- Tallent-Runnels, M. K., Thomas, J. A., Lan, W. Y., Cooper, S., Ahern, T. C., Shaw, S. M., & Liu, X. (2006). Teaching courses online: A review of the research. *Review of Educational Research*, 76(1), 93–135. doi:10.3102/00346543076001093
- Ternus, M. P., Palmer, K. L., & Faulk, D. R. (2007). Benchmarking quality in online teaching and learning: A rubric for course construction and evaluation. *The Journal of Effective Teaching*, 7(2), 51–67.

Wertsch, J. V. (1985). *Vygotsky and the social formation of mind*. Cambridge, MA: Harvard University Press.

Westby, C. (1997). There's more to passing than knowing the answers. *Language, Speech, and Hearing Services in Schools*, 28(3), 274–286. doi:10.1044/0161-1461.2803.274

White, L. A. (1975). *The concept of cultural systems: A key to understanding tribes and nations*. New York: Columbia University Press.

White, R. (n.d.). *Four myths about online learning*. Retrieved from <http://www.learnnc.org/lp/pages/2720>

Zimmerman, B. J. (1986). Becoming a self-regulated learner: Which are the key subprocesses? *Contemporary Educational Psychology*, 11(4), 307–313. doi:10.1016/0361-476X(86)90027-5

Zimmerman, B. J. (1994). Dimensions of academic self-regulation: A conceptual framework for education. In D. H. Schunk & B. J. Zimmerman (Eds.), *Self-regulation of learning and performance: Issues and educational implications* (pp. 3–20). Hillsdale, NJ: LEA.

Zimmerman, B. J. (2001). Theories of self-regulated learning and academic achievement: An overview and analysis. In B. J. Zimmerman & D. H. Schunk (Eds.), *Self-regulated learning and academic achievement: Theoretical perspectives* (pp. 1–38). Mahwah, NJ: Erlbaum.

Zimmerman, B. J., & Paulsen, A. S. (1995). Self-monitoring during collegiate studying: An invaluable tool for academic self-regulation. In P. R. Pintrich (Ed.), *Understanding self-regulated learning* (pp. 13–27). San Francisco, CA: Jossey-Bass. doi:10.1002/tl.37219956305

KEY TERMS AND DEFINITIONS

Assuring Quality: Online course delivery focusing on the alignment of sound instructional features, meaning-making, working together with students to develop new ways ‘to do school’ on-line, recognition of faculty work, and continual improvement.

Continuous Improvement: A dynamic process that examines those shared ideas of what is quality online teaching and learning and is essential for assuring quality.

Continual Improvement: The act of reflecting on the effectiveness of pedagogical alignment within the context of the constraints and affordances of the online teaching and learning environment.

Cultural Systems: Within online environments refers to understandings and meanings that are socially shared.

Evaluation: Includes examining the content, processes, impact and outcomes of on-line courses in order to improve the course quality.

High Impact Practices: Refers to educational experiences that are meaningful, require student action and participation, and that contribute to the life-long learning of the student.

On-line Teaching and Learning: Faculty-delivered instruction via the Internet or distance learning.

Self-Regulation: Refers to those processes that occur at an individual level that play an important role in student academic achievement.

A Study on Extensive Reading in Higher Education



Diana Presadă

Petroleum-Gas University of Ploiesti, Romania

Mihaela Badea

Petroleum-Gas University of Ploiesti, Romania

INTRODUCTION

There is a widely spread consensus among researchers and practising teachers that extensive reading programs are an important tool for teaching English as a second language. But, in spite of its tremendous contribution to the development of students' language skills, the difficulties entailed by the implementation and monitoring of such programs diminish teachers' enthusiasm about integrating them in the instructional process, especially in higher education.

As to university students, they tend to limit their reading activities to the topics of the syllabus, perceiving reading as a compulsory task that can ensure school success and not as a pleasant free-time activity. In other words, students are more interested in extrinsic rewards rather than in being intrinsically motivated readers. As Cambria & Guthrie (2010: 17) put it, 'students who read only for the reward of money, a grade or a future job are not the best readers', their major purpose being the reward itself and not learning. On the other hand, reading for enjoyment is a necessary but not sufficient condition to increase students' motivation for extensive reading. As the same authors point out (2010: 16), motivation includes 'the values, beliefs and behaviours surrounding reading for an individual. Some productive values and beliefs may lead to excitement, yet other values may lead to determine hard work.' This

determination, the two authors explain, consists in three factors that propel students to read, that is, interest, dedication and confidence, meaning the combination between reading for pleasure and a planned effort to reach the intended success.

Using the internet as a source for reading may help teachers develop students' intrinsic motivation within a well-organized reading program. It is a fact that students resort to the internet on a regular basis with various purposes: to communicate with their friends, to look for information they are interested in, to enjoy themselves etc. Whatever the reason, they will always turn to the internet as a first choice. Tempted by its tremendous advantages (rapidity, accessibility, huge stock of information etc.), most learners prefer doing their research more on the web and less in a traditional library. Under such circumstances, changing the internet into a tool for developing students' reading skills cannot but contribute to the success of an extensive reading program.

Taking into consideration the power of the internet to motivate students to read, the paper aims to find and evaluate students' attitude towards extensive reading. It also assesses the impact on students of a pilot program based on web resources analysing the results, identifying the difficulties involved by such a program, and trying to find solutions that may help the implementation of an extensive reading program into the curriculum of philological students in the future.

BACKGROUND

The general framework of the research on the implementation of an extensive reading program in higher education was provided by two categories of studies, their principles being applicable to reading no matter the medium, printed material or web resources.

The first category includes theories dealing with the effectiveness and positive effects of extensive reading on both cognitive and affective levels (Horst, 2005; Farrell, 2009; Nakanishi, 2015), whose synthesis of benefits is suggested by Nakanishi (2015: 9): ‘the freedom students have to choose books, the degree of autonomy enjoyed by the students, and the motivation to continue reading’. Moreover, the academic environment seems to be ideal for the application of an extensive reading program because, as Nakanishi (2015: 10) highlights, ‘the effect of extensive reading increases with older participants (...), who tend to learn language explicitly drawing on their analytical skills.’ The second category refers to the extensive reading activities described by Bamford & Day (2004) who offer a wide variety of useful ideas for an extensive reading program, including complex activities by means of which ‘teachers must take into account the effect of a class activity and a reading material not just on students’ ability to read, but on students’ self-images as readers, and on their feeling toward reading itself.’ (Day & Bamford, 1998: 166).

Generally, extensive reading of printed material or online resources implies the same cognitive processes as intensive reading does, presupposing an interactive model that consists in ‘constructing meaning from text through the use of both bottom-up and top-down strategies and skills’ (Şahin et al., 2014: 65, Nuttall, 1996). In other words, to successfully accomplish a reading activity, students need three types of competences at work: background knowledge to help them understand the possible meaning of the text, information about the distinguishing features of the reading texts (fictional, non-fictional etc.) and a good command of the target language system. In the

case of philology students, their ‘familiarity with the context’ and content is supplemented by their awareness of ‘literary devices, such as metaphors and flashbacks’. (Jacobs & Farrell, 2012: 15). In brief, as Nuttall shows (1996: 127, 128), extensive reading ‘can help readers enter the circle of growth’, which she calls ‘the virtual circle of the good reader’, on the grounds that it implies a multilateral and interrelated improvement connected with the reading speed, the capacity to understand ideas and a strong feeling of satisfaction while discovering the fictional world of the text.

Besides the characteristics of the reading process in general, the differences between intensive and extensive reading should also be considered when starting an extensive reading program. As shown in Table 1, what fundamentally separates extensive from intensive reading is that the former requires written material within students’ language competence, these texts being accessible, easy and enjoyable. Furthermore, students are stimulated to read what they want in their own time and as much as they can.

Table 1. Intensive vs. extensive reading

Intensive Reading	Extensive Reading
Reading for comprehension	Reading for pleasure
Students have to read	Students are motivated to read
100% comprehension of the text	Less than 100% comprehension
Word for word reading	Inferring the general meaning
Study of individual words and structures	Understanding sentences and discourse rules
Aid of the teacher in reading	Autonomous, independent reading
Reading above level	Reading according to level
Use of various comprehension strategies	Use of already acquired overall comprehension strategies
Use of dictionary for unfamiliar words and structures	Use of dictionary for vital words only
Short texts	Long texts
Reading done in classroom	Reading done outside the classroom
Teachers choose the reading material	Students select what to read

Due to the particular nature of extensive reading, it is essential that students get familiar with various techniques for inferring meaning and with the idea to accept overall comprehension instead of word-for-word understanding, specific to intensive reading. To achieve these goals, regular exercises done during the intensive reading and semi-extensive reading activities as a preparatory stage are necessary because they may form a solid basis for the development of the students' skills within an extensive reading program.

According to numerous researchers (Anderson, 2009; Grabe, 2009; Day, 2013), extensive reading represents a great opportunity for students to use the target language structures in various situational contexts and develop their vocabulary knowledge in view of the fact that they can 'meet the same lexis repeatedly in communicative contexts' (Nakanishi, 2015: 10, Grabe & Stoller, 2013). As to philology students, an extensive reading program would help them become proficient readers, an indispensable quality for their future career. Being proficient readers means meeting five parameters, as indicated by Coady & Huckin (1997: 228): 'immersion in meaningful texts, incidental language learning, integration of oral and written language, focus on meaning rather than form and the fostering of high intrinsic motivation'.

In such a context, the implementation of an extensive reading program based on web resources can help the educational process in higher education to shift from the traditional focus on the teacher to student centred methods 'by giving students more control of their own learning, for example, by students choosing what they will read; by students giving their own interpretation of what they read and bringing their own experiences, hopes and beliefs to the text and by students putting their own stamp on what they do while and after reading.' (Jacobs & Farrell, 2012: 3). Thus, extensive reading may pave the way for a qualitative educational change at academic level.

METHODOLOGY AND METHODS

To make students value reading outside the classroom, an extensive online reading program would be beneficial to adult learners, especially students in Philology, for the reason that it suits the cognitive processes required in an academic environment.

Piloting this program offered the possibility to see how intensive reading harmonises with extensive reading within a controlled approach whose final aim was to reinforce and develop students' English language skills.

Starting from the above-mentioned objectives, the following research questions were kept in mind:

- RQ1:** What are the major gains of an extensive reading program adopted at academic level?
- RQ2:** What drawbacks can such a program bring about in the educational process?
- RQ3:** To what extent can extensive online reading motivate students to become proficient readers?

PARTICIPANTS AND PROCEDURE

To put into practice an extensive reading program based on web resources in the University of Ploiesti, instructors constituted 2 groups of learners who volunteered to take part in the research. 64 undergraduate Philology students in the second year were involved, 38 specialising in Romanian and English and 26 in English and French. The program was based on volunteering because the Philology curriculum includes a variety of reading materials for the studied subjects and an extensive reading program would have meant a supplementary effort that not all students were willing to make.

The research consisted of 3 major stages with different targets: to examine students' reading experience and establish the list of online reading materials; to implement the extensive reading

program; to evaluate its effectiveness and set new strategies for the future.

During the initial stage a survey was administered to discover the participants' experiences, habits and attitudes towards reading in English. They had to answer several closed items to check their reading preferences and several open items focused on the reading time in English, their favourite writers, the difficulties encountered when reading and their perception of their own reading skills. It should be noted that the survey was very useful for orienting the reading list included in the program because it showed the kinds of texts mostly enjoyed. But to establish a coherent and efficient reading approach, it was necessary to identify the reading level of the students by means of a reading assessment test which was conducted at the end of the first semester of the academic year 2013-2014.

The results of the surveys enabled the preparation of the second stage that consisted in making the list of the online reading material. As freedom of choice is the basic criterion of a successful extensive reading program, students were encouraged to express their preferences in point of writers and types of reading texts. After identifying the students' reading tastes and levels, instructors had to establish a cohesive list of texts. It was necessary to do so because, as some sceptical researchers mentioned by Jacobs and Farrell (2012: 16) state, abusing the power of choice or adopting an 'anything goes attitude' to reading may be counter-productive.

The list contained non-fictional material (magazines and newspapers) and fictional prose (novels and short stories). The magazines and newspapers showed students' interests in topics like food, fashion, travelling, science, films etc., whereas the fictional texts, especially prose, referred to classical and modern writers (Emily Brönte, John Fowles, Vladimir Nabokov etc.) reflecting their desire not only to expand their knowledge about the issues studied in courses, but also to read other authors corresponding to their tastes and curiosity.

As students had different levels of competence in reading, the list also included graded fictional materials although some researchers think that 'the use of graded readers in EFL is controversial because some believe that they are not authentic' (Day et al., 2013). Being of real value as they encourage students to read according to their own level, graded materials were intended for intermediate and upper-intermediate learners.

In addition to students' preferences, several basic principles discussed in literature (Nuttall, 1982, Hill, 1997, Bamford & Day, 2004) were taken into consideration when making the reading list. The reading ease criterion (texts within students' reading level) guided the choice of texts whereas the variety of topics and genres, facilitated by the huge possibilities of the internet, could suit all learners' needs. To stimulate students' interest in reading, as Nuttall points out (1982, 171), texts 'should not smell of the classroom', their value consisting not in the number, but in the purpose for which it was established. Even if the reading list included a limited number of titles, especially at the same language level, as Hill (1997) notes, this may turn into an advantage because 'students will soon read similar titles and they can then share their experiences together' if encouraged during the monitoring activity held by the teacher.

To persuade students to participate in the program a semi-extensive reading activity was performed in the first week of the second semester, during which the whole class had to read the same online material which was a short story corresponding to the average reading level of the class and to their interests. The familiarization stage being achieved, students were invited to enrol in a program dedicated to reading internet-based resources outside the classroom. The students that signed up (64) were given the weekly reading target: three one-page newspaper or magazine articles and one book. Setting the reading target is an important step to be taken because, as the majority of the researchers claim, no progress can be achieved in the absence of a well-defined number of reading materials. This target should be

Table 2. Sample vocabulary log

Student's Name:	Title of the Book/Article:
New word/expression:	Definition:
Example sentence:	
Synonym:	Antonym:

understood 'in terms of the level and the number of books' (Hill, 1997) to be read each week as long as the program lasts. It should be mentioned that, to avoid wasting time when searching for the online material, students were provided with numerous links to access the texts included in the reading list.

To monitor their reading activity, a series of tasks were also conceived for the participants to complete by the end of each week. The tasks were gradual ranging from simple to medium and complex requirements. The first category comprised poster reports and story webs which students had to display on the walls in the classroom, focusing on characters, relationships and actions. The second one included vocabulary and plot logs for which the following vocabulary pattern was suggested as in Table 2:

The most complex and creative tasks such as book/article reports/reviews and 'The book/article and me' reports belonged to the third category. The 'me report' activity aimed at students' personal responses, as they had to explain their impression of what they had read and what new things they had learnt.

As concerns the post-implementation stage of the extensive online reading program, this was done in the last week of the second semester and consisted in a report on the advantages and disadvantages of the program, indicating its impact on the students and their attitude towards reading outside the classroom.

RESULTS AND DISCUSSIONS

The implementation of the extensive online reading program was oriented by the initial survey

administered to the students. According to the findings, a percentage of 57.69 students spend each week between 0 and 2 hours reading for pleasure in English (printed or electronic material), whereas 38.46% of them indicated between 3 and 5 hours. The surprising fact is that only 3.85% of the students dedicate more than 10 hours a week to reading in their free time.

The relatively high percentage of students who spend a short period of time reading for enjoyment is quite worrying because their philological profile requires not only imposed but also free reading as a way to improve language acquisition. Students' attitude towards reading may derive from the fact that they tend to consider extra-curriculum requirements to be time-consuming and useless. This demonstrates that they base their motivation on extrinsic rewards such as grades, which, as Cambria & Guthrie (2010: 17) explain, 'represent their quality as a student and as a reader. Being a high achiever is a symbol of how they are doing. A high grade is an icon of success and these students strive to feel successful.' Such a view of reading is also supported by their answers to the questions meant to find if they enjoy reading in English, the difficulties they encounter and what their favourite writers are. Almost two thirds of the respondents are fond of reading in English, but only the writers they study. As concerns the difficulties they have, most of them refer to vocabulary problems which hinder overall comprehension, this revealing that students are accustomed with intensive and not with extensive reading.

Another objective of the survey was to discover students' preferences regarding the reading text types. Almost 60% of the students like reading fiction, which is in accordance with the answers to the previous questions reflecting their interest in literature as part of the syllabus. As to their favourite genres, 73.08% mentioned prose and only 15.38% drama, respectively 11.54% poetry. These results may be explained by their perception of prose as more interesting and reader-friendly than drama and especially poetry. With respect to the non-literary texts preferred by the

respondents, 30.43% chose science books while 69.57% ticked magazines and newspapers. A possible justification for this high percentage is that reading newspapers and magazines makes students feel connected to real life because the various subject areas they find here respond to their needs and interests.

Interesting viewpoints were also shown by students' opinions of reading in English as a way of improving language abilities. Although all of them gave a favourable answer, they confuse extensive reading skills with the ability to speak, translate and pronounce correctly. Nevertheless, some students (38.63%) mentioned consolidation of vocabulary and grammatical structures among the benefits of reading in English. This percentage together with the one of participants who spend 3 to 5 hours a week reading extensively shows awareness of the advantages of reading outside the classroom.

In connection with students' English reading abilities, more than half believe that they are not good readers, the most frequent reasons being linked to poor vocabulary and a loaded timetable that does not allow reading outside the classroom. The overloaded bibliography they are required to study for regular classes does not give them time to read for pleasure. However, they admit that reading plays an important part in their lives, emphasising its informative role and the enrichment of their cultural background.

As to the students' final reports, their highly positive opinions on the program are worth analysing. First, students stated that they liked reading online materials outside the classroom considering it as useful as the compulsory reading syllabus. Second, they pointed out that they could read independently, enjoying the freedom to choose the most interesting articles and books. They also appreciated that they were able to read whenever and wherever they wanted to and at their individual reading speed. Choosing materials in accordance with their level eliminated the vocabulary difficulties that might have hindered the rhythm of the reading. The third aspect students indicated was the

chance to suit the reading materials to their levels, which made reading easier and more captivating than their former reading style. At the same time, inferring meaning from context instead of using the dictionary was recognized as a great step forward towards fluent reading. Briefly speaking, participants' answers show that they appreciate the benefits of extensive reading, benefits which, in Nuttall's terms, mean 'speed, enjoyment and comprehension' (1996: 128). Learners' opinions reflect a new attitude to reading, overcoming their former belief that school success depends only on reading what teachers indicate. Such a prejudice might be due to the authoritarian university teaching style according to which the teacher is the main decision factor in the learning process, irrespective of the subject.

Another advantage was that, by reading a large variety of online texts, participants had the opportunity to come across a great number of new words in different contexts, which led to a significant development of their vocabulary. In other words, by 'multiple exposures to new words' students can 'learn a large amount of new words' (Grabe, Stoller, 2013: 53). Finally, 'by reading the same patterns of letters, words, combinations of words again and again students process them more quickly and accurately and thus develop a sight vocabulary (words that are recognized automatically)' (Day et al., 2013). Students' perception somehow contradicts the findings reported by Tudor and Hafiz (1989, 1990) in their studies who reached the conclusion that extensive reading cannot be regarded as the best way of improving vocabulary.

In addition, learners pointed out that extensive reading helped them reinforce and develop structures already studied as well as other language abilities, their opinions reflecting what researchers mean by 'spread of effect', that is, development of interconnected skills such as 'writing, speaking and control over syntax' (Day and Bamford, 1998: 33). Although the cognitive mechanisms linking reading to other skills are not completely understood, as Hoey (2005) and Nuttall (1996)

state, it stands to reason that the more students are exposed to the new vocabulary and grammatical structures, the more competent they become in speaking and writing.

Finally, students admitted that extensive reading enabled them to develop their general knowledge about the world, to make connections with other fields of interest and to adopt new perspectives on life. This demonstrates the passage from the former extrinsic to intrinsic motivation which, in Farrell's terms, means that they internalised the idea that extensive reading is a way to 'develop themselves both personally and professionally' (Farrell, 2009: 88).

SOLUTIONS AND RECOMMENDATIONS

Students' reports also contained several disadvantages of the extensive online reading program, which should be considered when implementing it on a large-scale. Thus, they complained of the lack of time for extra-reading, mentioning their loaded timetable as well as the amount of compulsory reading as part of the syllabus. They also incriminated the fact that some links were accessed with difficulty or they wasted too much time when searching through different databases.

The dysfunctions of the pilot program as suggested by students will be of great help in improving the implementation of a future one. A possible way to solve the problem of the overloaded reading tasks could consist in a rigorous selection of the material and in harmonising the compulsory reading with the free time activities. This could be achieved if the extensive reading program becomes part of the curriculum within a reading course based on the principle of alternating intensive reading with extensive reading activities. Besides reading in quantity and regularly, the key to success is to provide students with time for doing it. According to some researchers, 'Reluctant readers simply do not read outside of the classroom, so if we truly believe that students must interact

with the text regularly, then our schedules need to be altered to provide time for that during the school day.' (Jacobs, Farrell, 2012: 2). As to the difficulties encountered when accessing the online material, a good idea would be to precede the reading activities by a web-based tutorial to help students search on web catalogues and through databases more efficiently.

FUTURE RESEARCH DIRECTIONS

Although the pilot extensive reading program based on web resources had a very good impact on students, its implementation on a large-scale implies a larger team of teachers that will have to re-analyse the basic findings of the present program in order to replicate it.

As motivating students for reading by means of every resource available, including the internet, is an important objective of any language teacher, the study can develop into a larger analysis of the role of technology in developing reading skills and in making their reading experience most productive.

Another challenging perspective of research would be the implementation of an extensive reading program taking into account the combination between printed materials and internet-based resources. Such a complex program may offer both students and teachers new educational opportunities to explore. Moreover, teachers can conduct specific research to compare the role of printed materials and internet-based resources in developing students' reading skills.

CONCLUSION

The implementation of an extensive online reading program at academic level proved to be beneficial for the development of the skills required by the specificity of philological studies. The success of the study in stimulating students to read for pleasure was demonstrated by the feedback provided in the final stage. It is worth mentioning that, in

spite of an overloaded timetable, the majority of the students volunteered to take part in an activity that was different from their daily routine, even if that required additional effort and work. Thanks to their openness and motivation, the implementation of the program unfolded without serious problems and obstacles. The monitoring process revealed that they got more and more seriously involved in doing their reading-related tasks, meeting deadlines and sharing opinions on the books they read.

One of the greatest gains of extensive reading is the formation of a lifetime habit according to which reading is perceived as a personal experience and interaction with the reading material, irrespective of its kind, printed or electronic, corresponding to varied purposes people may have in life such as to entertain, to get information, to relax etc. In addition, extensive reading focuses the teaching-learning process on students, enhancing the responsibility for their personal growth.

In conclusion, the successful design, development and implementation of an extensive reading program in higher education may contribute to the creation of a positive environment that will lead to the development of the learning and teaching process in EFL.

REFERENCES

- Anderson, N. J. (2009). Active reading: The research base for a pedagogical approach in the reading classroom. In N. J. Anderson & Z. Han (Eds.), *Second Language Reading Research and Instruction: Crossing the Boundaries*. Ann Arbor, MI: University of Michigan Press. doi:10.3998/mpub.339190
- Bamford, J., & Day, R. (Eds.). (2004). *Extensive Reading Activities for Teaching Language*. Cambridge, UK: Cambridge University Press.
- Cambria, J., & Guthrie, J. T. (2010). Motivating and Engaging Students in Reading. *The NERA Journal*, 46(1), 16–29.
- Coady, J., & Huckin, T. (1997). *Second Language Vocabulary Acquisition: A Rationale for Pedagogy*. Cambridge, UK: Cambridge University Press.
- Day, R. (2013). *Bringing Extensive Reading into the Classroom*. Oxford, UK: Oxford University Press.
- Day, R., & Bamford, J. (1998). *Extensive Reading in the Second Language Classroom*. Cambridge, UK: Cambridge University Press.
- Farrell, T. S. C. (2009). *Teaching Reading to English Language Learners*. Corwin Press.
- Grabe, W. (2009). *Reading in a Second Language: Moving from Theory to Practice*. Cambridge, UK: Cambridge University Press.
- Grabe, W., & Stoller, F. L. (2013). *Teaching and Researching: Reading* (2nd ed.). New York: Routledge.
- Hafiz, F., & Tudor, I. (1990). Graded Readers as an Input Medium in L2 Learning. *System*, 18(1), 31–42. doi:10.1016/0346-251X(90)90026-2
- Hill, D. (1997). Setting up an Extensive Reading Program. Practical Tips. *The Language Teacher*. Retrieved from <http://www.jalt-publications.org/tlt/articles/2135-setting-extensive-reading-program-practical-tips>
- Hoey, M. (2005). *Lexical Priming. A New Theory of Words and Language*. New York: Routledge.
- Horst, M. (2005). Learning L2 vocabulary through extensive reading: A measurement study. *Canadian Modern Language Review*, 61, 355–382. Retrieved from <http://utpjournals.metapress.com/content/96035r8253436173/fulltext.pdf>
- Jacobs, G. M., & Farrell, T. S. C. (2012). *Teachers Sourcebook for Extensive Reading*. Charlotte, NC: IAP- Information Age Publishing Inc.
- Nakanishi, T. (2015). A meta-analysis of extensive reading research. *TESOL Quarterly*, 49(1), 6–37. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1002/tesq.157/epdf>

Nuttall, C. (1982). *Teaching Reading Skills in a Foreign Language* (3rd ed.). Oxford, UK: Macmillan.

Şahin, S. (2014). *ÖABT İngilizce*. Ankara: Pegem Akademi.

Tudor, I., & Hafiz, F. (1989). Extensive Reading as a means of Input to L2 Learning. *Journal of Research in Reading*, 12(2), 164–178. doi:10.1111/j.1467-9817.1989.tb00164.x

ADDITIONAL READING

Bell, T. (2001). Extensive reading: Speed and comprehension. *The Reading Matrix*, 1(1), 1–13.

Birch, M. B. (2015). *English L2 Reading. Getting to the bottom*. NY: Routledge.

Davis, C. (1995). Extensive reading: An expensive extravagance? *ELT Journal*, 49(4), 329–336. doi:10.1093/elt/49.4.329

Han, Z., & Anderson, N. J. (2009). *Second language reading research and instruction: Crossing the boundaries*. Ann Arbor: University of Michigan Press.

Koda, K. (2004). *Insights into second language reading*. New York, NY: Cambridge University Press.

Nation, I. S. P. (2009). *Teaching ESL/EFL reading and writing*. New York, NY: Routledge.

Robb, T. N., & Susser, B. (1989). Extensive reading vs. skills building in an EFL context. *Reading in a Foreign Language*, 5, 239–251.

KEY TERMS AND DEFINITIONS

Extensive Reading: Reading in quantity, outside the classroom, to get a general understanding of a text.

Fictional Texts: Types of texts including poems, short stories, novels and plays.

Graded Reading: Reading according to the students' language level.

Intensive Reading: Reading in the classroom in order to get a detailed understanding of a text.

Language Motivation: Various factors, such as willingness, desires and attitudes, determining students to learn a second language.

Non-Fictional Texts: Types of texts including newspapers, magazines, textbooks etc.

Online Library: List of links to download reading material.

Reading Level: The level at which a reader understands a text.

Technology Policies and Practices in Higher Education

Kelly McKenna

Colorado State University, USA

INTRODUCTION

Teacher preparation classrooms are filled with digital learners, and as educators, we encourage ideas of integrating technology into their future content and pedagogy. The U.S. Department of Education Enhancing Education Through Technology Act of 2001 requires the integration of technology into elementary and secondary education, so teacher educators should be preparing their preservice teachers for integration in their future content and pedagogy. As educators we know teachers teach how they were taught, so effective use of available technology needs to be modeled for preservice teachers in order to permeate their future teaching of our youth. So, what classroom technology policies and practices are education classes actually modeling for their learners? For optimal learning and to effectively prepare preservice teachers for teaching in a digital world, teacher educators should be encouraging them to utilize the devices at their disposal.

BACKGROUND

Technology is an integral component of life as we know it, and many students have mobile technology devices available to them. To meet learners where they are, educators must discontinue the use of outdated teaching methods and embrace a digital world (Rossing, 2012). But, out of fear that students may not be focusing on the instruction, educators are banning the use of smart devices in their classrooms. Smart phones are used by indi-

viduals in the workplace for a variety of purposes, so “why do we prohibit students from using these devices in the place where they do their own daily work?” (Hill, 2011, p. 22). Students should be encouraged to utilize the devices at their disposal to improve class participation, investigation, activities, and creativity. The real advantage of using smart phones in education comes when they are no longer supplemental, but essential components in the learning process (Hill, 2011). By utilizing the devices that are a part of their everyday lives, students have the ability to continue gathering information and gaining knowledge outside of school. Integrating technology that is utilized by today’s learners’ in their everyday lives has the potential to revolutionize formal education (Peluso, 2012).

Technology integration in education refers to the utilization of technology to promote teaching and learning. By meeting students where they are in respect to technology, educators can create more learner-centered instruction, solving the challenges of students’ needs and desires to learn differently (McCaffrey, 2011). The inclusion of technology in teaching and learning modifies the current teaching paradigm and empowers students through hands-on learning. By integrating technology into pedagogy, educators can engage students like never before and cultivate deep, meaningful learning (McCaffery, 2011). When mobile devices are integrated into education, students’ learning can be extended beyond the boundaries of the classroom; “students with these devices can go any place and anywhere to get their information, learning from the palms of their hands” (Hill,

2011, p. 24). But, for many teachers, technology use in the classroom is an intimidating prospect; the ability to effectively integrate technology into the existing structure of teaching is no easy task.

Teachers are the key to effective technology integration; in order to transform their teaching paradigms it is imperative preservice teachers learn how to integrate technology (Bitner & Bitner, 2002). Instructing preservice teachers on effective technology integration can lead to teacher buy-in and the establishment of technology enriched curriculum (Hogue, 2013). Although preservice teachers are often digital learners and may be comfortable working with technology, they are frequently not taught how to integrate it into their pedagogy. Often “models of teaching based on their own experiences as students do not include the integration of technology into instruction” (Rosenfeld & Martinez-Pons, 2005, p. 146). Unfortunately, teachers have limited access to examples of effective technology integration after which to pattern their teaching (Bitner & Bitner, 2002). Rosenfeld and Martinez-Pons posit that teacher education programs often “focus on how to use technology rather than on how to teach with technology and integrate it into everyday teaching” (2005, p. 146). By observing their peers integrating technology and incorporating tools, teachers can be encouraged to apply the practices to their pedagogy (Hogue, 2013).

Technology education is a requisite for today’s learners to be prepared for leadership in a digital world. In spite of the Department of Education’s mandate to integrate technology into K-12 education and the International Society for Technology in Education’s (ISTE) emphasis on infiltrating teacher preparation programs in order to accomplish effective incorporation of technology into teaching and learning (ISTE, n.d.), teacher preparation continues to be overlooked (Bitner & Bitner, 2002). Although preservice teachers may be comfortable working with technology, models for integrating it into their pedagogy is lacking. Effective preservice teaching prototypes of technology integration need to be modeled in order to

succeed in achieving successful implementation in K-12 education.

H

MAIN FOCUS OF THE ARTICLE

The purpose of this study was to reveal the technology policies and practices being modeled for preservice teachers in their teacher preparation curriculum. The research sought to answer the following research questions:

- What classroom technology policies are education classes modeling for their learners?
- Are teacher educators encouraging the use of smart devices for learning in their teacher preparation courses?
- Are teacher educators effectively demonstrating personal technology integration to preservice teachers?
- What technology integration practices are educators modeling in their teacher preparation courses?

The study was comprised of faculty who are instructors of teacher preparation courses at a teaching intensive university in the Rocky Mountain region. The sample consisting of 26 teacher preparation instructors was purposeful, as the intent of the study was to gather information about how teacher educators are modeling technology utilization and policies for their preservice teachers.

Syllabi for teacher preparation courses are collected and archived each semester. The researcher was granted access to the Fall 2014 syllabi for teacher education courses for the purpose of this study. In total, 84 syllabi were reviewed for this study. Also, a 13 question survey was sent to all 26 faculty of teacher preparation courses via their university email. The survey consisted of primarily open-ended questions regarding technology policies and technology integration in teacher preparation courses (see Appendix). Some sample questions were:

- If a policy regarding students' use of technology during class time has been communicated, what are the details of the policy?
- Do you integrate technology into your teaching?
- What is the best way to teach preservice teachers about TPACK (Technological, Pedagogical, and Content Knowledge)?

The initial data was collected by reviewing archived syllabi for teacher preparation courses. Following review of the syllabi from all teacher preparation courses, a questionnaire was sent out to all instructors of these courses via SurveyMonkey.

Descriptive analysis was employed to analyze data from the archived syllabi. The syllabi were reviewed for the presence of explicitly stated technology policies regarding personal technology use during class time. The qualitative data collected from the open-ended questions in the survey were analyzed and coded, and themes from the open-ended questions were examined. The analysis was comprised of an iterative process through which patterns in the data emerged (Creswell, 2014). Following coding of the responses and emergence of themes, the larger meaning, or interpretation, of the qualitative data was conceptualized (Creswell, 2013). Credibility issues were addressed by using triangulation, and the incorporation of multiple methods of data collection for comparison bolstered the internal validity and reliability of the research (Merriam, 1998). Analysis of the data showed the findings made sense through consistency and dependability, thus strengthening the reliability of the study (Merriam, 1998). The process utilized in this research allowed for a holistic analysis of the data.

FINDINGS AND DISCUSSION

There are faculty in the School of Teacher Education who, based on the data, appear to be interested in technology integration. They recognize the need

for preservice teachers to learn how to infuse their pedagogy with technology enhanced teaching as divulged in the literature. Some are even modeling technology integration, "there are a lot of us... actively modeling the infusion of technology into [our] courses. It's particularly important in Geography, with all of the cool geospatial tools now available". However, the majority of faculty seem to struggle utilizing technology in their own teaching and allowing students to practice mobile technology in their classes. As one faculty response indicated, "continuing professional development in the inclusion of technology and pedagogy would be helpful." Data collected from syllabi and surveys illustrated the lack of technology integration modeled in teacher preparation courses and the deficiency in faculty knowledge regarding technology enhanced teaching and learning.

Syllabi from the teacher education courses offered in the fall of 2014 were collected for the research. Any reference made regarding the use of personal technology use during class was noted. Of the 84 syllabi reviewed, 38 (45%) included some policy regarding the use of personal technology (including: cell phones, smart phones, iPods, and pagers), and 46 (55%) made no reference. The education courses were in Culture Linguistics Diverse Education, Elementary Education Early Childhood, Elementary Education, Foundations of Education, and Reading.

For the most part, the use of cell/smart phones was prohibited in courses that included a technology policy. In some cases, no electronic devices were permitted to be used during class time. And, many syllabi required that cell/smart phones not be visible, but put away. Only one syllabus acknowledged the use of a smart phone for learning purposes:

- "Using a cell phone or text messaging during class is the ultimate demonstration of disrespect for your fellow classmates and instructor. However you are invited to use them for classroom and learning purposes."

The majority of policies stated that the use of cell/smart phones during class was either unacceptable or not tolerated. In addition, ten syllabi included a penalty in the students' grade if cell/smart phones were used in class:

- "Put cell phone away during class time and during the tutoring time—cell phones that are visible will result in loss of points."
- "The following criteria will be used for earning the full 4 points for each class session: Cell phones are put away and visible only during break times."
- "Cell Phone Policy: Please turn cell phones off upon entering the classroom and leave them off until the class is completed. Using a cell phone or text messaging during class is the ultimate demonstration of disrespect for your classmates and instructor. Violation of this policy will result in loss of participation points."

Included in the syllabi containing a technology policy were three that banned the use of any electronic devices (including laptops and computers), a representative sample was:

- "Electronic and other etiquette:... Please do not use personal electronic devices such as laptops, netbooks, smartphones, iPods, cell phones, etc., in this class, or you risk losing points from your professionalism & participation grade."

Eight included verbiage that required cell phones not be visible during class time:

- "You should keep your portable electronic devices in your backpack or purse during class. Your personal electronic devices should not be on your desks."

The survey consisting of 13 questions was sent to 26 teacher education instructors at the

start of the spring semester and was completed by 17 individuals. Half of the 12 instructors who do not include a technology policy in their syllabi stated that they have communicated a policy in their classes. For the most part these policies, not included in syllabi, prohibited the use of cell/smart phones for texting or talking during class, but allowed students to have them. One instructor encouraged the use of these devices:

- "I have asked students to keep their phones handy so they can 'hitchhike' or look up something about anything that is going on in class. I want them to see google as a central classroom tool during discussions".

One unique response by an instructor stated that if he/she sees a student on their phone, he/she will:

- "share with them the research that shows their phones have control over them, they don't choose when to look. That means put it where you cannot see or hear it".

Question six of the survey asked instructors to explain the reason for their policy or lack of one. One instructor who does not have a policy regarding technology in the class shared the following:

- "Are students surfing during class? Yes – Periodically, but I do not consider this to be such an issue that I need to create a policy. Finally if I do see students surfing then I have to ask myself, 'What can I do differently to better engage students in the learning process?' I see that as a more appropriate response as opposed to saying you cannot do this and failing to take any personal ownership in contributing to the situation."

Two instructors referenced the current digital age when answering this same question:

- “It’s the 21st century and people have multiple ways of learning. I’m trying to make use of technology as an asset, not a deterrent. I really don’t understand faculty who don’t allow computers for note-taking in their classes. In 2015??”
- I think we should be moving toward integrating and inviting mobile tech right into the classroom. The more we assume it is a distraction, the more we are teaching in the previous century (or even the one before that). Larry Cuban’s book ‘How Teachers Taught’ shows clearly how our profession is unreasonably conservative about what counts as appropriate classroom behavior, and specifically what counts as ‘attention’. The research on classroom attentiveness, on-task behaviors, and so on suggest we should be rolling with the current technology instead of trying to push back against it.”

Technology integration was also covered in the survey. All of the respondents stated that they incorporate technology into their teaching. PowerPoint, videos and BlackBoard (the learning management system employed by the university) were the most common technology tools listed for how instructors integrate technology. However, 10 of the 15 respondents to the question, describe your knowledge of Technological, Pedagogical and Content Knowledge (TPACK), were unfamiliar with it and only five of 14 had ideas for how to teach their preservice teachers about TPACK.

- “By using the tools that are most used and popular. If we show them how the most popular tools can be involved in learning and teaching, we can claim relevance.”
- “Good practices should be included in all college level instruction. Not just one time in a separate class. Since that’s impossible, then all education related coursework

should integrate technology effectively and for the sole purpose of improving student learning. If preservice teachers are trained in how students learn, then in their education courses and methods courses as they learn methods for teaching their topic, then technology would be part of this training so that they learn how to evaluate how the technology is helping them teach their students. Technology changes so rapidly that teaching specific uses, is only somewhat helpful in the very near future. Technology is just another tool for teaching and should be an integral part of the coursework. Research has demonstrated that teaching skills out of context, is difficult if not impossible to transfer.”

When asked about the SAMR Model (Substitution, Augmentation, Modification, Redefinition), only two instructors were familiar with it and three were able to explain where they felt their teaching with technology examples fell on the scale.

- “I would say mostly augmentation... I use YouTube for example, to make my face-to-face feedback asynchronous (not really face to face, I guess), but it is a powerful way to reach students on their own time.”
- “R-Redefine. There are several things that I can do in my classes that were impossible 10 years ago.”

With few exceptions faculty at this teaching intensive university are failing to establish the relationship between technology integration and teacher preparation that is established throughout the literature. As one faculty noted, “we as university instructors are overwhelmed with administrative technology applications, when we should be focused on teaching applications, both in what we teach and in how we teach”. Fortunately there are those exceptions who feel that “I always have more to learn!”

FURTHER RESEARCH DIRECTIONS

In spite of the emphasis on combining technology into content and pedagogy and the availability of smart devices, teacher preparation programs are failing to model integration for their students. This study focused on one teaching intensive university. Further research exploring additional teacher education programs and their policies and practices may be beneficial in discovering how prevalent this problem may be. In addition, further research may identify a program effectively modeling smart device integration into the curriculum. An exemplar might further cement the need for showcasing the integration of smart devices in teaching and learning and its impact for preservice teachers' ultimate adoption.

CONCLUSION

The purpose of this study was to discover the personal technology policies and technology practices currently being modeled in teacher preparation courses. The data collected highlighted the technology policies being modeled for preservice teachers (research question one). The syllabi and surveys collected for this research established that not only are teacher preparation courses failing to incorporate personal technology, they are going so far as to restrict its use in the classroom. Only a minimal number of instructors at this teaching intensive university are encouraging their preservice teachers to utilize smart devices in the classroom (research question two). A few instructors are familiar with successful technology practices and are modeling smart device integration in their courses (research question three). Technology is an integral component of life as we know it and many portable electronic devices now have access to a plethora of information via the internet or various applications. This opportunity for learning should be recognized and utilized. Currently minimal technology (PowerPoint, videos, Black-

Board) is being demonstrated in the curriculum preservice teachers undertake (research question four). The integration of technology is one of the standards in K-12 education where our preservice teachers intend to educate, so preservice teachers need to have successful examples of technology integration after which to model their own teaching. Modeling effective technology policies and practices are imperative for preservice teachers to learn the practice of technology integration and later assimilate it into their classrooms (Bitner & Bitner, 2002) (research question five). This study aimed to highlight the comprehensive influence on digital learners when preservice teachers are successfully prepared for the allowance and incorporation of the technology readily available to their future students.

REFERENCES

- Bitner, N., & Bitner, J. (2002). Integrating technology into the classroom: Eight keys to success. *Journal of Technology and Teacher Education*, 10(1), 95–100.
- Conn, C. (2012). Managing and maximizing a class set of iPads. *Learning and Leading with Technology*, 39(8), 32–33.
- Creswell, J. W. (2013). *Qualitative inquiry and research design: Choosing among five approaches* (3rd ed.). Los Angeles, CA: Sage.
- Creswell, J. W. (2014). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Upper Saddle River, NJ: Prentice Hall.
- Hill, R. A. (2011). Mobile digital devices. *Teacher Librarian*, 39(1), 22–26.
- Hogue, R. H. (2013). Considerations for a professional development program to support iPads in higher education teaching. *Ubiquitous Learning: An International Journal*, 5, 25–35.

International Society for Technology in Education. (n.d.). *Essential conditions*. Retrieved from <http://www.iste.org/standards/essential-conditions>

McCaffrey, M. (2011, February). Why mobile is a must. *THE Journal*. Retrieved from http://thejournal.com/articles/2011/02/08/why-mobile-is-a-must.aspx?sc_lang=en

Merriam, S. B. (1998). *Qualitative research and case study applications in education*. San Francisco: Jossey-Bass.

Peluso, D. C. C. (2012). The fast paced iPad revolution: Can educators stay up to date and relevant about these ubiquitous devices. *British Journal of Educational Technology*, 43(4), 125–127. doi:10.1111/j.1467-8535.2012.01310.x

Rosenfeld, B., & Martinez-Pons, M. (2005). Promoting classroom technology use. *The Quarterly Review of Distance Education*, 2(6), 145–153.

Rossing, J. P. (2012). Mobile technology and liberal education. *Liberal Education*, (Winter), 68–72.

U.S. Department of Education. (2001). *Enhancing education through technology act of 2001*. Retrieved from <http://www2.ed.gov/policy/elsec/leg/esea02/pg34.html>

ADDITIONAL READING

Al-Awidi, H., & Alghazo, I. (2012). The effect of student teaching experience on pre-service elementary teachers self-efficacy beliefs for technology integration in the UAE. *Educational Technology Research and Development*, 60(5), 923–941. doi:10.1007/s11423-012-9239-4

Albion, P., & Ertmer, P. (2002). Beyond the foundations: The role of vision and belief in teachers preparation for integration of technology. *Tech-Trends*, 46(5), 34–38. doi:10.1007/BF02818306

Bitner, N., & Bitner, J. (2002). Integrating technology into the classroom: Eight keys to success. *Journal of Technology and Teacher Education*, 10(1), 95–100.

El-Hussein, M. M., & Cronje, J. C. (2010). Defining mobile learning in the higher education landscape. *Journal of Educational Technology & Society*, 13(3), 12–21.

Hoppe, H., Joiner, R., Milrad, M., & Sharples, M. (2003). Guest editorial: Wireless and mobile technologies in education. *Journal of Computer Assisted Learning*, 19(3), 255–259. doi:10.1046/j.0266-4909.2003.00027.x

Iqbal, S., & Qureshi, I. (2012). M-Learning adoption: A perspective from a developing country. *International Review of Research in Open and Distance Learning*, 13(3), 147–164. doi:10.19173/irrodl.v13i3.1152

Koehler, M., & Mishra, P. (2005). Teachers learning technology by design. *Journal of Computing in Teacher Education*, 21(3), 94–102.

Lee, Y., & Lee, J. (2004). Enhancing pre-service teachers' self-efficacy beliefs for technology integration through lesson planning practices. *Computers & Education*, 7(3), 121–128.

Rosenfeld, B., & Martinez-Pons, M. (2005). Promoting classroom technology use. *The Quarterly Review of Distance Education*, 6(2), 145–153.

KEY TERMS AND DEFINITIONS

Learning Management System (LMS): The software system used for the delivery of courses online.

Mobile Learning: The use of personal electronic devices for the purposes of education. These devices include: handheld devices, tablets, and smart phones, or any other devices that allow for the learner to be mobile.

Preservice Teacher: Students who are currently in an educational program studying to become teachers. These students have not yet met the requirements needed to be certified teachers.

Smart Phone: A cellular phone containing many of the elements available on a computer. Often these devices have access to the internet.

Teacher Preparation Program: An approved course of study that meets the requirements necessary for certification.

Technological, Pedagogical, and Content Knowledge (TPACK): The framework required for effective instruction. TPACK is the integration of a teachers knowledge in the areas of technology, pedagogy, and content.

Technology Integration: Technology integration in education refers to the utilization of technology to promote teaching and learning.

H

APPENDIX

Questionnaire

1. List the programs in which you teach (EDF, ET, ELED, etc.).
2. Do you have an explicitly stated policy regarding students' use of technology during class time?
3. If you do have an explicitly stated policy regarding students' use of technology during class time, what does it state and where can it be found?
4. If you do not have an explicitly stated policy, have you ever communicated a policy regarding students' use of technology during class time?
5. If a policy regarding students' use of technology during class time has been communicated what are the details of the policy?
6. Please explain the reason for your policy or the lack of one.
7. Do you integrate technology into your teaching?
8. If yes, how do you integrate technology?
9. Describe your knowledge of TPACK (Technological, Pedagogical and Content Knowledge)?
10. What is the best way to teach pre-service teachers about TPACK?
11. Are you familiar with the SAMR Model (Substitution, Augmentation, Modification, Redefinition Model)? Please explain.
12. If you integrate technology into your teaching where on the SAMR Model do your lessons typically fall? Please explain.
13. Is there any additional information regarding technology policies and instructional modeling you would like to share?

The University–Industry Collaboration



Marcello Fernandes Chedid
University of Aveiro, Portugal

Leonor Teixeira
University of Aveiro, Portugal

INTRODUCTION

The collaborations between academia and industry - University-Industry Collaboration (UIC) - may occur according to different formats (multiple types) and recently have increased based on the third mission of the universities – knowledge transfer between university and external actors. This relationship offers advantages to both entities, addressing global challenges to their mutual benefit as well as benefits to society. Both university and industry recognize the potential of UIC relationship. Nevertheless, this relationship is complex and often appear threatening to both the university and industry through value and goals conflicts. The major reason for this complexity is that the collaboration between partners with different models of organization and culture needs a considerable management effort in order to be successful.

Despite the relevance of the theme, the studies in this area neither explain the various complexities associated with this relationship, nor present recommendations of improvement for the process (Santoro & Bierly, 2006).

In order to achieve success in this relationship, it is important the understanding of three drivers which are part of UIC, i.e., the motivation for collaboration, the channel of interaction and outcome and benefits of collaboration. Their understanding allows to mitigate barriers, overcome the differences, create an environment of trust and commitment (Kauppila, Mursula, Harkonen, &

Kujala, 2015) and consequently achieve success of the UIC.

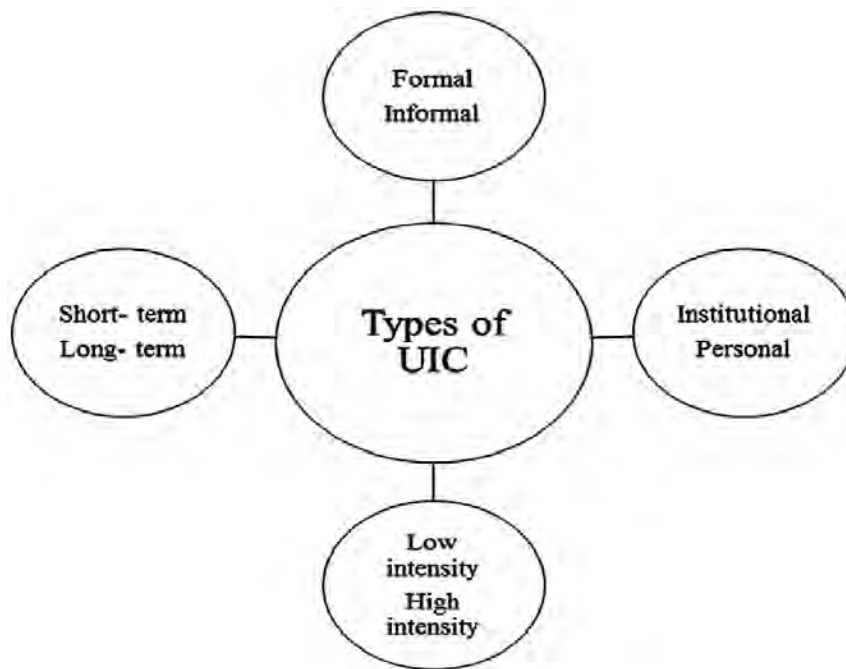
This chapter aims to address these three main drivers of this relationship based on literature review.

BACKGROUND: THE UNIVERSITY - INDUSTRY COLLABORATION

Since the end of last century, factors such as the globalization, the growing competition and the rapid technological advances have transformed the complex business environment with impact on life cycles of processes, products and services (Kauppila et al., 2015; Mendes, Nunes, & Sequeira, 2012). This scenario forced the companies to find partners to face the new challenges, representing the University-Industry Collaboration (UIC) relationship a key resource for innovation (Lee, 2000), for promotion of technological change (Cohen, Nelson, & Walsh, 2002; Freitas, Geuna, & Rossi, 2013; Lee, 2000; Mansfield & Lee, 1996) and for promotion of higher productivity and greater economic growth (Freitas et al., 2013).

For universities, this relationship also became important, as with better awareness of the business value of its work and its research, universities have shown more interest in the marketing of their products (Santoro & Bierly, 2006). So, in addition, to contributing to the better training of their students, the UIC can provide to universities access to expertise that they do not have and

Figure 1. Types of University - industry collaboration



that is only possible with direct experience with companies (Ankrah & AL-Tabbaa, 2015).

In fact, both the university, and the industry recognize the potential of UIC relationship. Nevertheless, this relationship is complex and often appears threatening to both the university and industry through value and goal conflicts. The key challenge is the understanding of the organizational form of the other partner. As soon as each institution understands the needs of the other, a large number of opportunities will exist (Sherwood, Robinson, & Butts, 2011; Wallin, Isaksson, Larsson, & Elfström, 2014; Wright, 2008).

Universities and industries have different objectives, focus and ways of working, which represents some barriers to the UIC (Sherwood et al., 2011). One of the barriers faced in this relationship is the difference of views with respect to the deadline for execution of works. Universities have a long-term vision, while industries work with a short-term vision. The time frames are different (Pertuzé, Calder, Greitzer, & Lucas, 2010). Another important barrier highlighted by some

authors refers to existing divergence between what is developed by the researchers in universities and the real needs or expectations of the industries (Franco & Haase, 2015) which sometimes are completely disconnected or opposed to seeking industries (Arza, 2010).

According to Santoro and Bierly (2006), academic researchers have not adequately studied many of the complexities associated with this relationship and thus have not been able to provide insightful recommendations to improve the process. Franco and Haase (2015) complement with the information that a great number of investigations is concentrated on the academic side of UIC and attention is mostly paid to the individual researchers.

Regarding the type of collaboration between university and industry, it can be established according to different approaches, having different types of classification (Figure 1).

Generally, UIC is associated with the level of involvement of organizations and type of resources that are used, and the relationship include components such as problem solving, technology

development, ideas testing or knowledge generation (Perkmann & Walsh, 2009).

According to Nilsson et al. (2010), the different types of collaborations, sometimes occur simultaneously and in other cases one type is dependent on another.

In this sense, UIC relationships can be classified as formal or informal (Hagedoorn, Link, & Vonortas, 2000; Nilsson et al., 2010; Polt, Rammer, Gassler, Schibany, & Schartinger, 2001), and as short-term or long-term (Bruneel, D'Este, & Salter, 2010; Wallin et al., 2014). On the basis of its character the informal relationships are not much discussed or known (Hagedoorn et al., 2000), and usually do not require a formalized contract. This type of UIC relationship is, normally, a short-term relationship. On the other hand, the formal relationships require formalized agreements and generally are classified as a long-term relationship (Bruneel et al., 2010) and therefore have a better chance for success (Lee, 2000). Verheugen and Potocnik (2005) suggests that a long-term relationship demonstrate commitment, leading to good results, mutual understanding and respect.

UIC can be low intensity or high intensity (Perkmann & Walsh, 2007, 2009), being differentiated according to the intensity of contacts and activities demanded by the relationship. Finally UIC can be institutional or personal (Freitas et al., 2013; Verheugen & Potocnik, 2005). The type of institutional relationship involves the structure of the university, while the personal mode refers to the direct contract between firms and researchers at the university.

THE MAIN DRIVERS OF UNIVERSITY - INDUSTRY COLLABORATION RELATIONSHIP

In order to achieve success in UIC relationship, it is important the understanding of its three main drivers, i.e., the motivation for collaboration, the channel of interaction and the outcome and benefits of the collaboration. The understanding of

these drivers and their proper articulation, allows to mitigate barriers, overcomes the differences, creates an environment of trust and commitment (Kauppila et al., 2015) and consequently achieve the desired success of the UIC.

According to the several literature reviewed, the Table 1 presents a complete information identified concerning to these three drivers – motivations, channels and outcomes – in the context of UIC. It is also identified the information of which is the focus (actor) of the articles, i.e., on the universities (U), industries (I) or both (U/I).

Motivations of Interaction

The motivations are internal and external factors that stimulate to attain a goal, they are different for universities and industries (Ankrah & AL-Tabbaa, 2015) and it play an important role in the achievement of results (Lee, 2000), as well as in determining the type of channel to be used in the UIC.

Universities and industries seek through the UIC, to have access to resources, skills and abilities that are currently limited to one of the entities (Wallin et al., 2014; Wright, 2008). The characteristics of the university and industry, experience in this kind of relationship and areas of interest are factors that differentiate the establishment of motivation. Whereas the motivations for participation in UIC are different for universities and industries (Ankrah & AL-Tabbaa, 2015), understanding the real meaning of each motivation is one of the important drivers for the success of the relationship. (Guimón, 2013) complements stating that successful UIC needs to support the missions and motivations of each partner.

Regarding the motivation for universities collaborating with industries, the literature presents different aspects that may encourage universities to seek collaboration with industries. Based on the knowledge transfer between university and external actors known as third mission of universities, the most important one is the creation, transfer and sharing of knowledge. However, beyond the

Table 1. List of articles identified in the literature mentioning one or more drives

Article	Actors	Motivations	Channels	Outcomes
Ankrah et al. (2013)	U/I	x	-	x
Ankrah & AL-Tabbaa (2015)	U/I	x	x	x
Arza (2010)	U/I	x	x	x
Barnes et al. (2002)	U/I	-	x	x
Bekkers & Freitas (2008)	U/I	-	x	-
Freitas et al (2013)	I	-	x	-
Bruneel et al. (2010)	U	-	x	-
Cohen et al. (2002)	I	-	x	x
D'Este & Patel (2007)	U	-	x	-
D'Este & Perkmann (2011)	U	x	x	x
De Fuentes & Dutrénit (2012)	U/I	x	x	x
Dooley & Kirk (2007)	U/I	-	x	x
Dutrénit & Arza (2010)	U/I	-	x	x
Franco & Haase (2015)	U	x	x	-
Grimpe & Hussinger (2008)	I	-	x	-
Guimón (2013)	U	x	x	-
Hagedoorn et al. (2000)	I	x	-	x
Kneller et al. (2014)	U	-	x	x
Lee (2000)	U/I	x	-	x
Muscio & Vallanti (2014)	U	x	x	-
Nilsson et al. (2010)	U	x	x	-
Plewa et al. (2013)	U/I	-	x	-
Perkmann & Walsh (2007)	U	-	x	-
Perkmann & Walsh (2009)	U	x	x	x
Perkmann et al. (2013)	U		x	
Philbin (2008)	U/I	x	x	x
Polt et al. (2001)	U/I	x	x	-
Wallin et al. (2014)	I	x	x	x

third mission, the universities look for ways to improve the training of its students, offering job opportunities to these students and to have access to expertise that they do not have and also to benefit from direct experience in companies (Ankrah & AL-Tabbaa, 2015; Lee, 2000). D'Este and Perkmann (2011) and Muscio and Vallanti (2014) also claim that, in recent years, due to cuts in government funding for research, universities have been more involved with industries, especially to get support for their research activities.

Arza (2010) summarizes and classifies the motivations of universities to collaborate with industries in two types: (i) intellectuals, regarding the exchange of information, education, ideas for new research, new publications and the consequent increase of academic productivity; and (ii) economics, relating to the funding of research.

Regarding the motivation for industries collaborating with universities, the available literature, although wide, is rather directed to developed countries. Probably, because in these countries the UIC has greater experience.

Lee (2000) lists a set of industries' motivations that appear frequently in the literature and support that the motivations for which the industries seek collaboration with university are many and complex, as industries have their own agendas for which they are willing to commit corporate resources.

Systematic review by Ankrah and AL-Tabbaa (2015) shows motivations categorized according to Oliver's six critical determinants: necessity, reciprocity, efficiency, stability legitimacy and asymmetry.

Arza (2010) classifies the motivations of industries to interact with universities in two types: (i) passive, using the results of the interaction for increased operational efficiency; and, (ii) proactive, exploiting the resources and expertise available in universities to develop innovative activities in industries.

Channels of Interaction

Given that the UIC can promote gains for both parties involved in the relationship, it must go through an interface or tangible communication processes (Feng, Zhang, Du, & Wang, 2015). The channels are the systems of interaction existent between the partners that allows the establishment of a relationship.

The authors in the subject consider a variety of channels of interaction may be set up (Nilsson et al., 2010; Sherwood et al., 2011), sometimes they occur simultaneously and in other cases one mechanism is dependent on another, and the emphasis in some channels depends on the motivation of each actor (Franco & Haase, 2015). Dutrénit and Arza (2010) and Arza (2010) add that some channels of interaction are more effective than others in order to achieve certain kinds of benefits. Despite the channel of interaction playing an important role of addressing motivation, it is often possible to observe that, regardless of the motivation, the well-known channels are the first to be addressed (Barnes, Pashby, & Gibbons, 2002).

With regard to literature about channel of interaction, some studies focus on industry (Cohen et al., 2002; Freitas et al., 2013; Grimpe & Hussinger, 2008); other focus on university and industry perspectives (Ankrah & AL-Tabbaa, 2015; Arza, 2010; Bekkers & Freitas, 2008; De Fuentes & Dutrénit, 2012; Dutrénit & Arza, 2010; Polt et al., 2001) and the majority of UIC surveys are concentrated in universities.

Regarding the channel, an important point to emphasize is that university has been treated as a single unit (Bruneel et al., 2010) when in fact it is a complex organization. The report Global Connect (2007) conceptualizes it very well when affirms that university is a permeable system with multiple points of UIC. In other words, there are many doors to the university through which university partners can enter through to develop collaborative relationships. Although the relevance of the subject, most of the studies present in the literature are conducted in the perspective of the individual researcher.

Regarding to types of channels, there are multiple channels that can be used during the UIC. In this work, six groups to classify the interaction channels are identified: (i) information—including publications, conferences, informal contacts, others; (ii) R&D projects—including contract R&D, consulting, and joint R&D; (iii) licenses and patents; (iv) business—including joint or cooperative ventures, purchase of prototypes developed at science, creation of physical facilities, university spin-offs; (v) training—including supervision of PhDs and Masters theses, training of employees of enterprises, students working as trainees; and, (vi) human resources—including personnel mobility, hiring of recent graduates. Table 2 presents compilation of the channels, based on the classification described above, that are cited in the reviewed articles that address this driver. It is also presented the information of which the actors are involved in the studies (U – University and/or I - Industry).

Table 2. Channels of interaction and their actors across different studies

Article	Actors	Information	R&D Projects	Licenses & Patents	Business	Training	Human Resources
Ankrah & AL-Tabbaa (2015)	U/I	x	x	x	x	x	x
Franco & Haase (2015)	U	x	x	x	x	x	-
Kneller et al. (2014)	U	-	x	x	x	x	x
Wallin et al. (2014)	I	x	x	-	x	x	-
Muscio & Vallanti (2014)	U	x	x	-	x	x	-
Freitas et al (2013)	I	-	x	-	-	-	x
Guimón (2013)	U	x	x	x	x	x	x
De Fuentes & Dutrénit (2012)	U/I	x	x	x	-	x	x
D'Este & Perkmann (2011)	U	-	x	x	x	-	-
Bruneel et al. (2010)	U	-	x	-	x	x	-
Dutrénit & Arza (2010)	U/I	x	x	x	x	x	x
Arza (2010)	U/I	x	x	x	x	x	-
Nilsson et al. (2010)	U	x	x	x	x	x	x
Perkmann & Walsh (2009)	U	-	x	-	-	x	-
Bekkers & Freitas (2008)	U/I	x	x	x	x	x	x
Grimpe & Hussinger (2008)	I	x	x	x	x	-	-
D'Este & Patel (2007)	U	x	x	-	x	x	-
Dooley & Kirk (2007)	U/I	x	x	x	-	-	-
Perkmann & Walsh (2007)	U	x	x	x	x	x	-
Cohen et al. (2002)	I	x	x	x	x	-	x
Polt et al. (2001)	U/I	x	x	x	x	x	x

Outcomes of Interaction

According to Ankrah and AL-Tabbaa (2015), UIC is based on the assumptions that the benefits are greater than the (social) cost or risks associated with interaction. The capacity to generate significant outcomes, some expected and others unexpected (Lee, 2000), to the partners and society (Franco & Haase, 2015), has been one of the main factors that has contributed to the expansion of the use of UIC. The literature presents relevant research on this topic.

Regarding the main outcomes promoted by the interaction, they differ between the partners of the relationship, and are strongly associated with initial motivation for interaction and with interaction channel used for this purpose (Arza, 2010). The recent systematic review by Ankrah

and AL-Tabbaa (2015) identifies an extensive list of UIC outcomes realized by universities and industries.

According to Lee (2000), the most significant outcome gathered by industries is an increased access to new university research and discoveries, and the most significant outcome by universities members is complementing their own academic research by securing funds for graduate students and lab equipment, and by seeking insights into their own research.

However, these considerable potential benefits are often not recognized in practice (Barnes et al., 2002; Dooley & Kirk, 2007). The major reason is that collaborations between partners with different models of organization and different culture need considerable management effort in order to be successful.

Table 3. Threats and opportunities to university across different studies

Threats	Opportunities
Greater involvement with industry can corrupt academic research and teaching (De Fuentes & Dutrénit, 2012)	Close collaboration facilitates interactive learning which in turn indirectly benefits scientific production by generating new ideas and motivating new research projects (Arza, 2010; Perkmann & Walsh, 2009)
Shift from basic research towards more applied topics and less academic freedom (P. D'Este & Patel, 2007; Perkmann & Walsh, 2009)	Applied projects offer more learning opportunities during via highly interdependent interaction with industry (Perkmann & Walsh, 2009)
Can reduce the openness of communications among academic research and put restriction on publishing (De Fuentes & Dutrénit, 2012)	Collaborative engagement may benefit academic's research activities by establishing relationships with knowledge users (Pablo D'Este & Perkmann, 2011)
The knowledge becoming obsolete because of withholding the publishing of their results until patenting has taken place (Ankrah et al., 2013)	Researches learn to identify the most valuable research from the pool of their inventions and, therefore, become better at raising private funds (Arza, 2010)
Risks that UICs pose to the integrity of academic investigation, which could undermine public faith in universities and research involving UICs (Ankrah et al., 2013)	

For the UIC to be successful and reach mutual benefits, it is important that the channels are correctly identified in each level of the UIC (Ankrah & AL-Tabbaa, 2015), and having a clear definition of the desired and expected goals by the partners (Barnes et al., 2002). Lee (2000) complements this reasoning, suggesting that the sustainability of the relationship is assigned to a mutual benefit.

Outcomes for Universities

Several authors categorize universities benefits in two relevant outcomes: the intellectual outcomes and the economic outcomes (Arza, 2010; De Fuentes & Dutrénit, 2012; Dutrénit & Arza, 2010). In a recent survey made by Ankrah et al. (2013), social aspect is also focused and is classified in the lowest level, both for universities and industries. At the academy, the most cited social benefit is "interesting and provides personal satisfaction".

The literature point out acquisition of funds for research and lab equipment, and creation of business opportunities as main economics outcomes for universities (Ankrah & AL-Tabbaa, 2015; Ankrah et al., 2013; De Fuentes & Dutrénit, 2012; Lee, 2000). It is also possible to add the opportunity of sharing of equipment and instruments as

another economic benefit (Ankrah & AL-Tabbaa, 2015; De Fuentes & Dutrénit, 2012).

As intellectual outcomes, authors refer gain of insights for further research and collaborative projects, knowledge sharing and opportunity to expose students and university to practical problems and to state-of-the-art technology (Ankrah et al., 2013; De Fuentes & Dutrénit, 2012; Lee, 2000).

It is important to point out that some authors present arguments that the UIC relationship is beneficial and advantageous to universities and there are those who consider harmful and threatening (Perkmann & Walsh, 2009). In the literature examined for this chapter it is possible to observe the conflict between the statements of different authors (Table 3).

Outcomes for Industries

Industry's partnership with universities may result in benefits to the industry to ensure competitive advantage and productivity gains with impact on financial performance (Ankrah et al., 2013; Pertuzé et al., 2010).

According to Ankrah et al. (2013), these collaborations give industries access to diverse resources, sometimes at prices lower than market rates, thus

enabling industry to reduce their overall costs, especially those relating to knowledge creation such as research and development. Another important point is that technological knowledge produced in the universities is the result of a dynamic development based on discussion of previous research results, including detailed documentation of trial-and-error events, which is sometimes difficult for industry to develop internally (Barnes et al., 2002; Grimpe & Hussinger, 2008).

In general, these studies categorize outcomes for industries as: (i) research and development (R&D) – acquisition of complementary or substitute R&D, innovation and new patents and processes; (ii) non R&D – technology transfer, acquisition of solution to solve production problem, acquisition of knowledge and access to qualify human resources; and (iii) quality – quality control and test of products and processes (Arza, 2010; De Fuentes & Dutrénit, 2012; Lee, 2000).

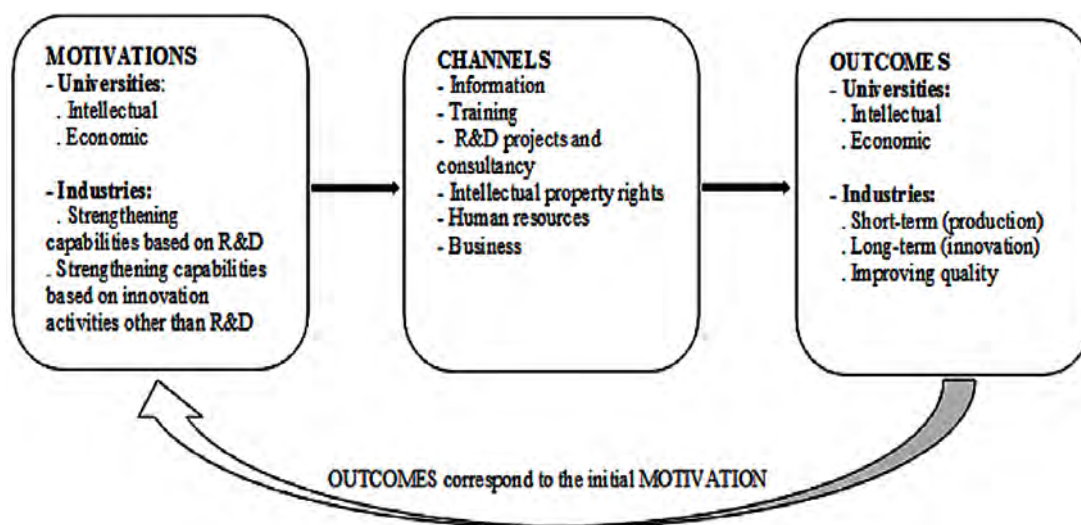
As shown above, social benefits are also classified at the lowest level by industry, being “enhanced image and reputation/credibility with the industrial community” the most cited benefit (Ankrah et al., 2013).

The UIC is based on the assumption that the benefits are greater than the costs and

risks associated with the investment required (Ankrah & AL-Tabbaa, 2015). However, in general, greater investment in resources is done by industry. Then, from the company’s perspective, is expected tangible benefits that include an appropriate return on investment through the value created from the technology generated (Philbin, 2008). However, in the literature reviewed for this chapter, it is possible to identify risks that offer some threat to the industry such as: (i) fundamental differences in the relative priorities, perspectives and time horizons of university and industry (Barnes et al., 2002); (ii), gap between the knowledge produced by university researchers and what is used in practice (Arza, 2010; Franco & Haase, 2015); (iii) the loss of control of vital technology and information leakage about the firm’s new technologies (Ankrah et al., 2013); (iv) other risks cited, like as financial and market risks, risk of incompetent academics in the technology transfer process, risk of incomplete transfer, and risk of non-performance of the technology (Ankrah et al., 2013).

The Figure 2, based on the literature reviewed, presents a resume of the three drivers discussed above.

Figure 2. Motivations, channels of interaction, outcomes of UIC



FUTURE RESEARCH DIRECTIONS

Some directions of future research in this topic could be associated with the need to understand better the UIC relationship. At first, the authors consider that although the UIC be a relationship between two organizations with fundamental differences, it is possible to identify in the literature that the major part of studies emphasizes the impact of the university in the industry. Probably caused by the fact that the authors of research articles working in universities. As an example, is mentioned the article (Cohen et al., 2002), reference too many other articles, which does not refer any impact of industry on university (Feng et al., 2015). It is expected that future researches are needed to examine the contributions and impacts of the relationship in each of the partners. Secondly, despite of universities are complex organizations with diversity and conflicts, many studies in UIC have treated the university as a single unit (Brueneel et al., 2010) and mostly into the individual researcher. Such differences are considered as

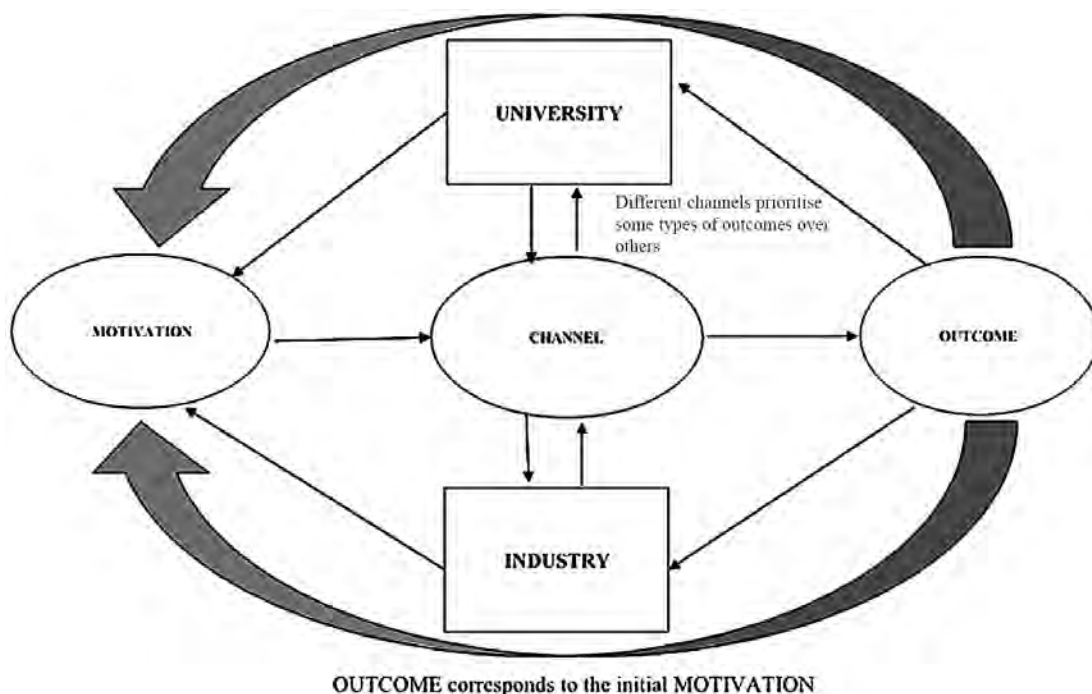
major obstacle to successful UIC (Barnes et al., 2002). The development of a systematic procedure of governance and management of UIC relationship is essential to its success.

Finally, it is possible identify others areas for further research: (i) development of a performance measurement system for UIC and, (ii) the shift in modern economies from manufacturing to services sector become this sector an important area for future researches.

CONCLUSION

The aim of this chapter was to approach and understand the three main drivers of the UIC relationship – motivation, channel of interaction and outcome – that the authors consider that may lead this relationship to success and achieve mutual benefits and therefore promote the establishment of a long-term relationship. In order to achieve this goal, the literature was revised and consequently the main motivations to promote

Figure 3. UIC framework



the interactions, the different channels of interaction and the expected outcomes were identified. During the review, it was possible to perceive the key role of the channels for the relationship. The proper identification and definition of channels of interaction allow the motivations are properly addressed and consequently the outcome achieved. Even if the motivation is very well outlined, the incorrect identification of the channels will lead to failure in achieving the goals and the failure of the relationship, undermining future actions of partnership. The authors also recognize, considering the differences of culture and organizational models between the partners involved in the UIC, that it is critical for the success of this relationship the establishment of a governance model for relations of partnership between universities and industries. The industry has the opportunity to support this aspect for its large experience in developing and implementing projects. The figure 3 presents a framework to the understanding of the UIC relationship, based on the three main drivers described above.

ACKNOWLEDGMENT

The authors would like to thank the editor and the anonymous reviewers for their constructive comments and suggestions which contributed to the enhancement of this chapter.

In a general manner, the relevant suggestions and comments have been incorporated in this reviewed version.

REFERENCES

- Ankrah, S., & AL-Tabbaa, O. (2015). Universities–industry collaboration: A systematic review. *Scandinavian Journal of Management*, 31(3), 387–408. doi:10.1016/j.scaman.2015.02.003
- Ankrah, S., Burgess, T. F., Grimshaw, P., & Shaw, N. E. (2013). Asking both university and industry actors about their engagement in knowledge transfer: What single-group studies of motives omit. *Technovation*, 33(2-3), 50–65. doi:10.1016/j.technovation.2012.11.001
- Arza, V. (2010). Channels, benefits and risks of public–private interactions for knowledge transfer: Conceptual framework inspired by Latin America. *Science & Public Policy*, 37(7), 473–484. doi:10.3152/030234210X511990
- Barnes, T., Pashby, I. A. N., & Gibbons, A. (2002). Industry Interaction : A Multi-case Evaluation of Collaborative R & D Projects. *European Management Journal*, 20(3), 272–285. doi:10.1016/S0263-2373(02)00044-0
- Bekkers, R., & Freitas, I. M. B. (2008). Analysing knowledge transfer channels between universities and industry: To what degree do sectors also matter? *Research Policy*, 37(10), 1837–1853. doi:10.1016/j.respol.2008.07.007
- Bruneel, J., DEste, P., & Salter, A. (2010). Investigating the factors that diminish the barriers to university–industry collaboration. *Research Policy*, 39(7), 858–868. doi:10.1016/j.respol.2010.03.006
- Cohen, W. M., Nelson, R. R., & Walsh, J. P. (2002). Links and Impacts: The Influence of Public Research on Industrial R&D. *Management Science*, 48(1), 1–23. doi:10.1287/mnsc.48.1.1.14273
- De Fuentes, C., & Dutrénit, G. (2012). Best channels of academia–industry interaction for long-term benefit. *Research Policy*, 41(9), 1666–1682. doi:10.1016/j.respol.2012.03.026
- DEste, P., & Patel, P. (2007). University–industry linkages in the UK: What are the factors underlying the variety of interactions with industry? *Research Policy*, 36(9), 1295–1313. doi:10.1016/j.respol.2007.05.002

- DEste, P., & Perkmann, M. (2011). Why do academics engage with industry? The entrepreneurial university and individual motivations. *The Journal of Technology Transfer*, 36(3), 316–339. doi:10.1007/s10961-010-9153-z
- Dooley, L., & Kirk, D. (2007). University-industry collaboration: Grafting the entrepreneurial paradigm onto academic structures. *European Journal of Innovation Management*, 10(3), 316–332. doi:10.1108/14601060710776734
- Dutrénit, G., & Arza, V. (2010). Channels and benefits of interactions between public research organisations and industry: Comparing four Latin American countries. *Science & Public Policy*, 37(7), 541–553. doi:10.3152/030234210X512043
- Feng, F., Zhang, L., Du, Y., & Wang, W. (2015). Visualization and quantitative study in bibliographic databases: A case in the field of university–industry cooperation. *Journal of Informetrics*, 9(1), 118–134. doi:10.1016/j.joi.2014.11.009
- Franco, M., & Haase, H. (2015). University–industry cooperation: Researchers motivations and interaction channels. *Journal of Engineering and Technology Management*, 36, 41–51. doi:10.1016/j.jengtecman.2015.05.002
- Freitas, I. M. B., Geuna, A., & Rossi, F. (2013). Finding the right partners: Institutional and personal modes of governance of university–industry interactions. *Research Policy*, 42(1), 50–62. doi:10.1016/j.respol.2012.06.007
- Grimpe, C., & Hussinger, K. (2008). *Discussion Paper N° 08-080: Formal and Informal Technology Transfer from Academia to Industry: Complementarity Effects and Innovation Performance. Mannheim: ZEW – Centre for European Economic Research*. Retrieved from <ftp://ftp.zew.de/pub/zew-docs/dp/dp08080.pdf>
- Guimón, J. (2013). Promoting University-Industry Collaboration in Developing Countries. *The Innovation Policy Platform*, 1–11.
- Hagedoorn, J., Link, A. N., & Vonortas, N. S. (2000). Research partnerships. *Research Policy*, 29(4), 567–586. doi:10.1016/S0048-7333(99)00090-6
- Kauppila, O., Mursula, A., Harkonen, J., & Kujala, J. (2015). Evaluating university–industry collaboration: The European Foundation of Quality Management excellence model-based evaluation of university–industry collaboration. *Tertiary Education and Management*, 21(3), 229–244. doi:10.1080/13583883.2015.1045550
- Kneller, R., Mongeon, M., Cope, J., Garner, C., & Ternouth, P. (2014). Industry-university collaborations in Canada, Japan, the UK and USA - With emphasis on publication freedom and managing the intellectual property lock-up problem. *PLoS ONE*, 9(3), e90302. doi:10.1371/journal.pone.0090302 PMID:24632805
- Lee, Y. S. (2000). The Sustainability of University-Industry Research Collaboration: An Empirical Assessment. *The Journal of Technology Transfer*, 25(2), 111–133. doi:10.1023/A:1007895322042
- Mansfield, E., & Lee, J.-Y. (1996). The Modern University: Contributor to Industrial Innovation and Recipient of Industrial R & D Support. *Research Policy*, 25(7), 1047–1058. doi:10.1016/S0048-7333(96)00893-1
- Mendes, L., Nunes, A., & Sequeira, T. (2012). Influence of human capital on firms' and nations' performance. A bi-disciplinary literature review. *Argumenta Oeconomica*, 29(2), 5–32.
- Muscio, A., & Vallanti, G. (2014). Perceived Obstacles to University–Industry Collaboration: Results from a Qualitative Survey of Italian Academic Departments. *Industry and Innovation*, 21(5), 410–429. doi:10.1080/13662716.2014.969935
- Nilsson, A. S., Rickne, A., & Bengtsson, L. (2010). Transfer of academic research: Uncovering the grey zone. *The Journal of Technology Transfer*, 35(6), 617–636. doi:10.1007/s10961-009-9124-4

- Perkmann, M., Tartari, V., McKelvey, M., Autio, E., Broström, A., DEste, P., & Sobrero, M. et al. (2013). Academic engagement and commercialisation: A review of the literature on university-industry relations. *Research Policy*, 42(2), 423–442. doi:10.1016/j.respol.2012.09.007
- Perkmann, M., & Walsh, K. (2007). University-industry relationships and open innovation: Towards a research agenda. *International Journal of Management Reviews*, 9(4), 259–280. doi:10.1111/j.1468-2370.2007.00225.x
- Perkmann, M., & Walsh, K. (2009). The two faces of collaboration: Impacts of university-industry relations on public research. *Industrial and Corporate Change*, 18(6), 1033–1065. doi:10.1093/icc/dtp015
- Pertuzé, J. A., Calder, E. S., Greitzer, E. M., & Lucas, W. A. (2010). Best Practices for Industry-University Collaboration. *MIT Sloan Management Review*, 51(4), 83–90.
- Philbin, S. (2008). Process model for university-industry research collaboration. *European Journal of Innovation Management*, 11(4), 488–521. doi:10.1108/14601060810911138
- Plewa, C., Korff, N., Johnson, C., MacPherson, G., Baaken, T., & Rampersad, G. C. (2013). The evolution of university-industry linkages - A framework. *Journal of Engineering and Technology Management*, 30(1), 21–44. doi:10.1016/j.jengtecman.2012.11.005
- Polt, W., Rammer, C., Gassler, H., Schibany, A., & Schartinger, D. (2001). Benchmarking industry-science relations: The role of framework conditions. *Science & Public Policy*, 28(4), 247–258. doi:10.3152/147154301781781453
- Santoro, M. D., & Bierly, P. E. (2006). Facilitators of Knowledge Transfer in University-Industry Collaborations: A Knowledge-Based Perspective. *IEEE Transactions on Engineering Management*, 53(4), 495–507. doi:10.1109/TEM.2006.883707
- Sherwood, A. L., Robinson, D. F., & Butts, S. B. (2011). Four stages to more successful knowledge transfer between universities and industry entrepreneurs. *Entrepreneurial Executive*, 16, 9–23.
- Verheugen, G., & Potocnik, J. (2005). *Responsible Partnering-A Guide to Better Practices for Collaborative Research and Knowledge Transfer Between Science and Industry*. European Commission, DG Enterprise & Industry.
- Wallin, J., Isaksson, O., Larsson, A., & Elfström, B.-O. (2014). Bridging the gap between university and industry: Three mechanisms for innovation efficiency. *International Journal of Innovation and Technology Management*, 11(01), 1440005. doi:10.1142/S0219877014400057
- Wright, R. (2008). How to Get the Most From University Relationships. *MIT Sloan Management Review*, 49(3), 75–80.

ADDITIONAL READING

- Etzkowitz, H., Webster, A., Gebhardt, C., & Terra, B. R. C. (2000). The future of the university and the university of the future: Evolution of ivory tower to entrepreneurial paradigm. *Research Policy*, 29(2), 313–330. doi:10.1016/S0048-7333(99)00069-4
- Huang, M.-H., & Chen, D.-Z. (2016). How can academic innovation performance in university-industry collaboration be improved? *Technological Forecasting and Social Change*, 1–6. doi:10.1016/j.techfore.2016.03.024
- Iskanius, P., & Pochola, I. (2016). Leveraging communities of practice in university-industry collaboration: A case study on Arctic research. *International Journal of Business Innovation and Research*, 10(2/3), 283–299. doi:10.1504/IJBIR.2016.074830
- Mansfield, E. (1998). Academic research and industrial innovation: An update of empirical findings. *Research Policy*, 26(7-8), 773–776. doi:10.1016/S0048-7333(97)00043-7

Perkmann, M., Tartari, V., McKelvey, M., Autio, E., Broström, A., DEste, P., & Sobrero, M. (2013). Academic engagement and commercialisation: A review of the literature on university-industry relations. *Research Policy*, 42(2), 423–442. doi:10.1016/j.respol.2012.09.007

Rast, S., Tourani, A., & Aslan, A. S. (2015). Effect of Organizational Factors on University-Industry Collaboration: A Conceptual Model. *International Journal of Business and Management*, 10(6), 188–198. doi:10.5539/ijbm.v10n6p188

Steinmo, M. (2015). Collaboration for Innovation: A Case Study on How Social Capital Mitigates Collaborative Challenges in University–Industry Research Alliances. *Industry and Innovation*, 22(7), 597–624. doi:10.1080/13662716.2015.1105127

Teixeira, A., & Mota, L. (2012). A bibliometric portrait of the evolution, scientific roots and influence of the literature on university-industry links. *Scientometrics*, 93(3), 719–743. doi:10.1007/s11192-012-0823-5

KEY TERMS AND DEFINITIONS



Channel of Interaction: Mechanisms of interaction existent between the partners of a relationship.

Governance: The establishment of policies, and continuous monitoring of their proper implementation to ensuring effectiveness.

Informal Collaboration: Mechanism that does not involve any contractual relationship between the university and the industry.

Motivation: Internal and external factors that stimulate to attain a goal.

Outcome: Consequence or conclusion after a period of time, which can be one result, multiple results, or no results.

Types of UIC: Organization forms of UIC relationship.

University-Industry Collaboration (UIC): Bi-directional relationship between university and industry entities, established to enable the diffusion of creativity, ideas, skills and people with the aim of creating mutual value over time.

Using Communities of Inquiry Online to Perform Tasks of Higher Order Learning

Ramon Tirado-Morueta
University of Huelva, Spain

Pablo Maraver-López
University of Huelva, Spain

Ángel Hernando-Gómez
University of Huelva, Spain

INTRODUCTION

Given the growth of online collaborative learning in higher education and the proliferation of technologies that facilitate it, it is critical to gain a better understanding of the mechanisms that promote quality. In this sense, the community of inquiry (CoI) model (Garrison, Anderson, & Archer, 2000) represents one of those mechanisms. This model is useful to describe the development of online collaborative learning, taking into account the processes of instructional dialogue that can facilitate successful online learning (e.g., Akyol & Garrison, 2011; Szeto, 2015). The CoI model explains the online learning as a by-product of collaborative work between active participants in a learning community characterized by proper orchestration between three components: the online pedagogical environment (teaching presence); social/referee support (social presence); and the cyclical process of interaction that allows meaningful learning within the student community (cognitive presence).

The overall objective of the study presented in this article is to use the CoI model to deepen understanding of the mechanisms that facilitate successful online learning when students tackle learning tasks of a higher order. There is evidence

showing that the nature of the learning task modulates the profile of the learning process; however, we consider that there is insufficient evidence to obtain patterns to monitor group relations in similar contexts.

For this purpose, authors have combined two techniques of analysis: quantitative content analysis (QCA); and social network analysis (SNA). QCA is a process involving searching text according to a defined scheme, and is used to identify frequencies. The CoI model provides the category system used in this study.

The authors have also used SNA because of its ability to explain the nature of group relations, based on the flow of information and communication found in the interactions of participants. SNA has been used by several researchers to improve understandings of individual and group online learning dimensions. However, few of these studies have used a comprehensive conceptual framework that considers the core elements of the CoI model.

Therefore, with the support of QCA and SNA, and using as the conceptual framework of the CoI model, the specific objectives of this study is to determine the influence of group social structure and centrality of coordinators on social and cognitive activity for the different types of collaborative learning task.

BACKGROUND

To achieve the stated objective, this section describes the conceptual frameworks and the background used in this empirical study.

Community of Inquiry (CoI) Framework

During the last fifteen years, many researchers have both studied and analyzed educational experiences according to the CoI model (e.g., Akyol & Garrison, 2011; Szeto, 2015; Author). The CoI model focuses on learning processes from a collaborative constructivist point of view. The model also assumes that learning in online environments occurs through the interaction of three core elements: social presence, cognitive presence, and teaching presence. These elements work together to support deep and meaningful online inquiry and learning.

The first element, social presence, is defined as the ability of learners to project themselves socially and affectively into a community of inquiry (Rourke, Anderson, Garrison, & Archer, 1999). Social presence is divided into three categories of affective, interactive, and cohesive presence, which reflect a supportive context for emotional expression, open communication and group cohesion for task resolution respectively. Social presence, a factor critical to face-to-face teaching, is challenging to facilitate in online learning environments.

However, social presence is essential to collaborative learning experience and is an essential element in establishing cognitive presence. Cognitive presence refers to the extent to which online learners can construct and validate meanings based on communication and thinking (Garrison et al., 2000, 2001).

The CoI model categorizes cognitive presence into four phases: a triggering event (an issue is identified for inquiry); exploration (exploring the issue through discussion and critical reflection); integration (constructing meaning from the ideas

developed through exploration); and resolution (applying new knowledge into a real world context), with specific descriptors for each phase.

The third element of the CoI model is teaching presence. It consists of two general functions: (1) the design of the educational experience; and (2) facilitation among the teacher and the students. It is the responsibility of the teacher to design and integrate both cognitive and social presence for educational purposes through scaffolding, modeling, and/or coaching.

Interest in understanding the processes of learning through the CoI framework, and its relationship with learning outcomes of higher order is based on the study by Marton and Saljo (1976), in studying different learning strategies used by students and their different results. Marton (1988) indicated that lesson learned, and how it is learned are two inseparable aspects of learning. Therefore, the process and the result are intimately associated. In this sense, Akyol and Garrison (2008) conducted a study through content analysis, in order to show how the CoI framework progresses during the learning process, finding that cognitive and teaching presence are determined primarily by the task and not by time.

The results from research on the CoI framework, have shown the importance of considering the disciplines, and contents/coursework as study variables. However, there are few studies which deepen the processes and outcomes (related) learning

Social Network Analysis

In general, a social network can be defined as a group of entities that collaborate or compete. Mathematically, this is a graph, in which each participant is represented by a node. The relationships between nodes are represented as links between the corresponding nodes.

Social network analysis (SNA) provides useful information and quantitative indicators of the quality of the learning process as a way to effectively analyze the process of co-construction of

knowledge in a CoI (e.g., Shea, Hayes, Uzuner et al., 2013). It also allows for description, analysis and visualization of the patterns of interaction and social structures of the online groups formed by students, offering a multitude of network indicators to evaluate the process and quality of the knowledge construction of students from a social perspective (Nistor et al., 2014).

In this study, the authors will focus on some of the variables most used in SNA, for both individuals and groups: centrality degree, centralization, and cohesion.

Centrality degree is the measure of the interaction considering the directionality of the sending or receiving of messages. The idea is that if a subject is central in their group then they will be the most popular and get the most attention. Centrality is an important measure because previous research has found that it correlates with positive learning outcomes. Centrality can be of two types: centrality in-degree (in) and centrality out-degree (out). Centrality (in) is indicator of network prestige, and centrality (out) is indicator of influence. Prestige measures the number of incoming responses directed to a student's discussion post, and represents the degree to which other students seek out that student for interaction (deLaat, et al., 2007). Students with high prestige are notable because their opinions may be considered more important than others in the network. Students with high influence are in contact with many other students because they initiate a large number of discussion posts to others (see Shea et al., 2013). Some authors found that participants who occupy central positions in the formal network are more likely to put forward ideas and suggestions on teamwork.

Network centralization refers to the extent to which a structure of relationships is centered on a student. Centralization of the network can be of two types: centralization in-degree (in) and centralization out-degree (out). Centralization (in) refers to the extent to which a group is focused on a subject, that is, who most group members directed their interventions towards. Centralization (out) refers to the extent to which a group

is focused on a subject, that is, who among the group engages in most of the interventions. A positive relationship has been identified between cohesion and centralization (in) of the forum and these associations' direct effects on knowledge construction in communities of inquiry (e.g., Tirado, Hernando & Aguaded, 2012). However, the centralization (out) has also shown to be inversely related (Thormann, Gable, Fidalgo & Blakeslee, 2013), to categories (Newman, Webb, & Cochrane, 1995) of critical thinking.

Cohesion describes the general level of linkage among the actors in the network (Scott, 2000). Cohesion is the forces holding the individuals within their groupings (Yang & Tang, 2003). In some cases, high levels of cohesion and centralization of the group can be good predictors of teamwork, achievement of the results of the learning task (Yang & Tang, 2003), and critical thinking phases of Gunawardena and colleagues' model (Aviv, Erlich, Ravid & Geva, 2003).

However, these advances show the overall performance of the model. Although the task (outcome) learning is generally considered in most of these studies are still few studies that integrate the type of learning task analysis among the variables.

Therefore, given the relevance of the learning task in defining the learning process, it is necessary to conduct studies that consider it as a central element in the analysis of model performance.

Bloom's Taxonomy

The current study focuses on required tasks in an online— and self-regulated—learning environment. Specifically, the authors wondered if the type of academic task proposed in a CoI would determine the frequency of the social and cognitive actions of a group of online learners. Bloom's taxonomy, a well-known and widely used schema, was used to better understand the systematic classifications and processes of thinking and learning in an online learning environment. In the current study, the authors considered three higher order

thinking levels from the revised Bloom's taxonomy (Anderson & Krathwohl, 2001, pp. 67–68): (a) *Analyzing* (breaking material into constituent parts and determining how the parts relate to one another and to an overall structure or purpose by differentiating, organizing, and attributing); (b) *Evaluating* (making judgments based on criteria and standards by checking and critiquing); and (c) *Creating* (putting elements together to form a coherent or functional whole, and reorganizing elements into a new pattern or structure by generating, planning, or producing).

The research question is: Using SNA metrics and QCA, how does social structure influence the level of group social and cognitive presence in each type of learning task? In addition, how does the level of intervention [centrality (out)] and popularity [centrality (in)] of the coordinators influence the social and cognitive presence of the group?

METHOD

During three academic years (2010-11, 2011-12 and 2012-13), a total of 206 university students from different degree areas and from 9 universities across Andalusia, Spain participated in the current study. Each of these students participated in a common online course. The first year, 71 students participated in the course; the second year, 77 participated; and, the last year, 64 participated.

Experimental Setting

The present study took place at the Virtual Campus of Andalusia, Spain where students from 9 public universities from the Andalusian region of Spain participated in a common online course through Moodle LMS. The online course is called, "Intervention on risk behaviors". This course is offered since 2008 to the present. It can enroll up to 10 students from each of the public universities of Andalusia, and from different academic areas (social, engineering, experimental, and health). The study was based on group discussions among the participating students. A total of 96 discussion forums were conducted: 8 groups per three academic years for each of four group activities (including the first socialization activity which was not included in the analysis).

Moodle discussion forums were employed to facilitate content delivery and to promote higher order thinking skills among the students. The required tasks were based on the skills that are present in the top three categories of the revision of Bloom's taxonomy. The course design was inspired by the model of Salmon (2013), because provides a staged, practical approach to teaching and learning online. In accordance with this model, was following steps:

In Step 1, 8 working groups were established, each one into a discussion forum where students were given directions for how to proceed by the tutor. In step 2, the first actions were performed, in order for students to socialize, exchange mes-

Table 1. Online educational activities*

Task	Description	Supporting Resources
Analyzing case study (ACS)	Analyzing, in a collaborative way, the real world case of a student from a secondary school.	Case description, scheme and help for the case analysis.
Evaluating websites (EW)	Evaluating web resources which can be used to provide information about risky behaviors.	List of risky behaviors and example of the data sheet.
Creating a WebQuest (CWQ)	Creating a WebQuest that can serve as a tool for information-training for drug prevention.	Web sources and examples.

*Based on Bloom's taxonomy

sages, interact and learn. In the step 3, three types of tasks were presented (Table 1) with different levels of cognitive demand that required participants in each group to share information, begin to engage in group work, and to share the same learning goal. In step 4, the learners focused on knowledge development and discussion activities.

To conduct the assigned tasks, each of the eight working groups were provided a private forum to work together, and a file space for uploading or downloading documents. They were offered theoretical and practical resources that also served as reference material and support for the implementation of the tasks. Two different moderators were selected by each group to coordinate each task. One instructor monitored the interactions, becoming active only in case of necessity.

Measures and Procedures

Data were collected using the messages in the forums created by each group, using the instruments developed by Anderson et al. (2001), Garrison et al. (2000, 2001) and Rourke, Anderson, Garrison, and Archer (1999) as references for measuring social presence and cognitive presence. For the QCA, the units of analysis were the *units of meaning*, not the specific messages (Rourke et al., 2001). Depending on the semantic sense used, several units of meaning could be conveyed in each message. A total of 13,501 units of meaning were coded into the qualitative analysis program AtlasTi v.6. Finally, data was imported to SPSS v.21 in order to perform quantitative analysis.

In accordance with the principles of systematicness, objectiveness, and reliability of QCA indicated by Rourke and Anderson (2002), three coders classified the messages using the CoI model (Appendix 1 and 2). The researchers followed the categorization of the CoI model and coded the messages into cognitive or social presence, including indicator notes for each message. The statistical package macro (KALPHA) by Hayes & Krippendorff (2007) was used for the calculation of the interrater and intra-rater reliability coef-

ficient Krippendorff's alpha. The global output from the macro Krippendorff's ordinal α is 0.766, a modest degree of interrater reliability. The data was then imported into SPSS v.21 in order to conduct quantitative analysis. In the numerical matrix of imported data, each indicator becomes a quantitative variable that reflects its frequency for each participant.

In order to perform the SNA, the authors considered individual and group variables associated with the structure of the groups (Appendix 3). The authors identified the network properties as density, cohesion and centralization index and used them to explain the structure of interpersonal communication. The data were entered into UCINET 6 for Windows (Borgatti, Everett, & Freeman, 2002), where rates of the different variables were obtained for SNA. Finally, data was imported to SPSS v.21 in order to conduct quantitative analysis.

The statistical analysis of data was performed using measurements of QCA and SNA. The linear regression analysis was used. With this analysis the influence on social and cognitive presence of the social structure of the group (density and centralization [out/in]) and the level of intervention/popularity (centrality [out/in]) of group coordinators was identified.

RESULTS

Based on the analysis of linear regressions, the following results were obtained (Table 2).

The results regarding the influence of the group structure on the social and cognitive presence of the group were as follows:

1. The density of the group directly influenced all categories of social presence and cognitive presence in the three learning tasks; however, the strength of this influence depended on the type of learning task. In this sense, density had more influence in the categories of social presence than in

Table 2. Linear regression

ACS		Social Presence (Group)			Cognitive Presence (Group)			
		Aff.	Int.	Coh.	Act.	Exp.	Inte.	Res.
Coordinator	Centrality (out)	.63***	.46*	.47*	.75***	.59**	.46**	.32
	Centrality (in)	-.01	.11	.01	-.22	.03	.30	-.16
	<i>F</i>	5.60**	4.03*	2.87*	8.39***	5.31**	5.70**	1.01
	<i>R</i> ²	.37	.30	.23	.41	.36	.37	.09
Group	Density	.54***	.71***	.57***	.46***	.23***	.20**	.29***
	Centralization (out)	-.35***	-.43***	-.35***	-.05	-.29***	-.12	-.27***
	Centralization (in)	.14**	.28***	.11*	-.42***	-.04	-.15*	.19**
	<i>F</i>	41.98***	143.51***	46.58***	43.60***	10.16***	5.65**	13.36***
	<i>R</i> ²	.38	.67	.40	.38	.13	.06	.16
<i>EW</i>		Aff.	Int.	Coh.	Act.	Exp.	Inte.	Res.
Coordinator	Centrality (out)	.70***	.56**	.29	.57**	.20	.00	-.21
	Centrality (in)	-.05	.17	.27	-.28	.40*	.16	-.10
	<i>F</i>	8.07***	8.28***	3.89*	2.87*	4.69**	.16	1.03
	<i>R</i> ²	.43	.44	.27	.21	.31	.02	.09
Group	Density	.30***	.42***	.50***	.21***	.57***	.32***	-.02
	Centralization (out)	-.05	-.13*	-.13**	-.17**	-.00	-.15*	-.00
	Centralization (in)	.35***	.29***	.52***	.27***	.16**	.09	-.17
	<i>F</i>	23.53***	30.28***	112.03***	11.88***	44.74***	10.08***	.34
	<i>R</i> ²	.25	.31	.62	.15	.39	.13	.03
<i>CWQ</i>		Aff.	Int.	Coh.	Act.	Exp.	Inte.	Res.
Coordinator	Centrality (out)	.40	.69***	.35*	.31	.77***	-.24	-.32
	Centrality (in)	-.05	.07	.26	.15	-.05	.17	.09
	<i>F</i>	3.33*	12.05***	4.44**	3.50*	13.03	.62	1.27
	<i>R</i> ²	.24	.50	.27	.23	.52	.05	.09
Group	Density	.61***	.63***	.38***	.16**	.66***	.06	.04
	Centralization (out)	.00	-.00	.09	.19**	-.22***	-.01	-.31***
	Centralization (in)	.07	.00	.06	.40***	-.08	-.29***	-.17*
	<i>F</i>	43.56***	45.19***	15.02***	22.81***	49.53***	6.37***	9.54***
	<i>R</i> ²	.39	.40	.18	.25	.42	.07	.12

* $p < .05$, ** $p < .01$, *** $p < .001$

the categories of cognitive presence in the ACS task. Instead, density more strongly influenced exploration, both in the EW task ($\beta[\text{Exp}] = .57$, $p < .001$), and the CWQ task ($\beta[\text{Exp}] = .66$, $p < .001$).

- Moreover, centralization (out) usually inversely influenced all categories of social and cognitive presence, with a more negative influence on the ACS task than on the rest of the learning tasks.

3. However, the influence of centralization (in) on the categories of social and cognitive presence was usually variable, depending on the type of learning task. Its influence on the categories of social presence was direct and significant in the ACS and EW tasks, but it was not significant in the CWQ task. Also, in the EW task, its influence was also significant in the activation ($\beta[\text{Act}]=.27$, $p<.001$) and exploration ($\beta[\text{Exp}]=.16$, $p<.01$) categories. However, while there was a direct effect of centralization (in) on the activation category of the CWQ task ($\beta[\text{Act}]=.40$, $p<.001$), its effect on the same category was reversed in the ACS task ($\beta[\text{Act}]=-.42$, $p<.001$).

The results regarding the influence of the coordinators on the social and cognitive presence of the group are as follows:

1. The centrality (out) of coordinators often had a direct influence on the three types of task, although the nature of the learning task modulated its effect on the social and cognitive presence categories.
2. In ACS, the influence of the centrality (out) of coordinators was positive and significant in almost all categories of social presence ($\beta[\text{Aff}]=.63$, $p<.001$; $\beta[\text{Int}]=.46$, $p<.05$; $\beta[\text{Exp}]=.47$, $p<.05$) and cognitive presence ($\beta[\text{Act}]=.75$, $p<.001$; $\beta[\text{Exp}]=.59$, $p<.01$; $\beta[\text{Inte}]=.46$, $p<.01$).
3. However, in the EW task, there was a positive and significant influence on the social presence categories of affective and interaction ($\beta[\text{Aff}]=.70$, $p<.001$; $\beta[\text{Int}]=.56$, $p<.01$) and on the cognitive category of activation ($\beta[\text{Act}]=.57$, $p<.01$).
4. Moreover, in the CWQ task, the influence was positive and significant on the interaction and cohesion categories of social presence ($\beta[\text{Int}]=.69$, $p<.001$; $\beta[\text{Coh}]=.35$, $p<.05$) and on the cognitive category of exploration ($\beta[\text{Exp}]=.77$, $p<.001$).

5. A significant influence regarding the centrality (in) (prestige) of coordinators was not seen in any of the three learning tasks.

DISCUSSION AND EDUCATIONAL IMPLICATIONS

In a CoI in which the learning task is ACS, the factor that facilitates social presence in the group is a dense social presence and a small centralized network. Moreover, the coordinator occupies a prominent role, triggering the problem definition, the exchange, and integration of ideas, and the proposed solutions; however, it is especially necessary for other members of the group to be active in suggesting ideas and providing data relevant to the resolution of the case.

In a CoI in which the learning task is EW, the search for consensus and the definition of evaluation criteria is a precondition and the fundamental condition to be solved. This first concern is usually led by the group coordinators with the involvement of other group members. In this case, the forum was a whole communication space with a very important consultation and resolution role, allowing coordinators and other group members to address problems.

In a CoI in which the learning task is CWQ—the less structured task—the density of group interactions is a very important factor, as is the actions of coordinators in searching for consensus and the exchange of proposals and suggestions. The prestige of the coordinators is also important so that group members will consult them.

FUTURE RESEARCH DIRECTIONS

One limitation of this case study is that it focused on three learning tasks that attempt to represent the top three levels of Bloom's taxonomy (analysis, evaluation and creation). Given that it is difficult to find practical examples suitable for Higher Education, which represent the exact taxonomic levels,

in later studies it will be appropriate to specify learning objectives in each of the task areas and to conduct more modular analysis.

Considering that cognitive presence includes self-reflection and co-construction of knowledge, for a better understanding of cognitive activity in a community of inquiry it would seem appropriate to incorporate an analysis of meta-cognition (e.g., Bjork, Dunlosk, & Kornell, 2013) especially in online self-regulated learning environments (e.g., Garrison & Akyol, 2013; Kovanović et al., 2015).

CONCLUSION

Understanding the mechanisms that favor successful online learning experiences in higher education remains a prime target. Although many pedagogical models designed to exploit the virtues of technology in higher education have been developed, they provide insufficient empirical evidence to guide educators and students in developing online learning processes.

In this study, using methods such as QCA and SNA, it has been found that the nature of the learning task determines the social structure of the group (cohesion and centralization) as well as the centrality of the coordinators on the social and cognitive actions of students.

REFERENCES

- Akyol, Z., & Garrison, D. R. (2008). The development of a community of inquiry over time in an online course: Understanding the progression and integration of social, cognitive and teaching presence. *Journal of Asynchronous Learning Networks*, 12(3–4), 3–22. Retrieved from http://olc.onlinelearningconsortium.org/sites/default/files/v12n3_akyol_0.pdf
- Akyol, Z., & Garrison, D. R. (2011). Assessing metacognition in an online community of inquiry. *The Internet and Higher Education*, 14(3), 183–190. doi:10.1016/j.iheduc.2011.01.005
- Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching and assessing: A revision of Bloom's Taxonomy of educational objectives. Complete edition*. New York: Longman.
- Garrison, D. R., Anderson, T., & Archer, W. (2000). Critical inquiry in a text-based environment: Computer conferencing in higher education. *The Internet and Higher Education*, 2(2-3), 87–105. doi:10.1016/S1096-7516(00)00016-6
- Garrison, D. R., Anderson, T., & Archer, W. (2001). Critical thinking, cognitive presence, and computer conferencing in distance education. *American Journal of Distance Education*, 15(1), 7–23. doi:10.1080/08923640109527071
- Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures*, 1(1), 77–89. doi:10.1080/19312450709336664
- Marton, F. (1988). Describing and improving learning. In *Learning strategies and learning styles* (pp. 53-82). Springer US. doi:10.1007/978-1-4899-2118-5_3
- Marton, F., & Säljö, R. (1976). On qualitative differences in learning: I—Outcome and process. *The British Journal of Educational Psychology*, 46(1), 4–11. doi:10.1111/j.2044-8279.1976.tb02980.x
- Nistor, N., Baltes, B., Dascălu, M., Mihăilă, D., Smeaton, G., & Trăuşan-Matu, Ş. (2014). Participation in virtual academic communities of practice under the influence of technology acceptance and community factors. A learning analytics application. *Computers in Human Behavior*, 34, 339–344. doi:10.1016/j.chb.2013.10.051

Rourke, L., & Anderson, T. (2002). Using peer teams to lead online discussions. *Journal of Interactive Media in Education*, 1(1), 1–21. doi:10.5334/2002-1

Rourke, L., Anderson, T., Garrison, D. R., & Archer, W. (1999). Assessing social presence in asynchronous text-based computer conferencing. *International Journal of E-Learning & Distance Education*, 14(2), 50–71. Retrieved from <http://www.ijede.ca/index.php/jde/article/view/153/341>

Salmon, G. (2013). *E-tivities: The key to active online learning*. Routledge.

Shea, P., Hayes, S., Uzuner, S., Vickers, J., Bidjerano, T., Gozza-Cohen, M., Jian, S., Pickett, A., Wilde, J. and Tseng, C. (2013). Online Learner Self-Regulation: Learning Presence Viewed through Quantitative Content and Social Network Analysis. *The International Review of Research in Distance Learning*, 14(3), 427–461. doi:10.19173/irrodl.v14i3.1466

Szeto, E. (2015). Community of Inquiry as an instructional approach: What effects of teaching, social and cognitive presences are there in blended synchronous learning and teaching? *Computers & Education*, 81, 191–201. doi:10.1016/j.compedu.2014.10.015

Thormann, J., Gable, S., Fidalgo, P. S., & Blakeslee, G. (2013). Interaction, Critical Thinking, and Social Network Analysis (SNA) in Online Courses. *The International Review of Research in Distance Learning*, 14(3), 294–318. doi:10.19173/irrodl.v14i3.1306

Tirado, R., Hernando, Á., & Aguaded, J. I. (2012). The effect of centralization and cohesion on the social construction of knowledge in discussion forums. *Interactive Learning Environments*, 23(3), 293–316. doi:10.1080/10494820.2012.745437

ADDITIONAL READING

Garrison, D. R. (2011). *E-Learning in the 21st century: A framework for research and practice* (2nd ed.). London: Routledge/Falmer.

Koh, J. H. L., Herring, S. C., & Hew, K. F. (2010). Project-based learning and student knowledge construction during asynchronous online discussion. *The Internet and Higher Education*, 13(4), 284–291. doi:10.1016/j.iheduc.2010.09.003

Liu, C. J., & Yang, S. C. (2014). Using the Community of Inquiry Model to Investigate Students Knowledge Construction in Asynchronous Online Discussions. *Journal of Educational Computing Research*, 51(3), 327–354. doi:10.2190/EC.51.3.d

Richardson, J. C., & Ice, P. (2010). Investigating students level of critical thinking across instructional strategies in online discussions. *The Internet and Higher Education*, 13(1-2), 52–59. doi:10.1016/j.iheduc.2009.10.009

Swan, K., Garrison, D. R., & Richardson, J. (2009). A constructivist approach to online learning: the community of inquiry framework. In C. R. Payne (Ed.), *Information technology and constructivism in higher education: progressive learning frameworks*, 43–57. Hershey, PA: IGI Global. doi:10.4018/978-1-60566-654-9.ch004

Venkataramani, V., Zhou, L., Wang, M., Liao, H., & Shi, J. (2016). Social networks and employee voice: The influence of team members and team leaders social network positions on employee voice. *Organizational Behavior and Human Decision Processes*, 132, 37–38. doi:10.1016/j.obhdp.2015.12.001

Wasserman, S., & Faust, K. (2007). *Social Network Analysis*. New York: Cambridge University Press.

KEY TERMS AND DEFINITIONS

Centrality: The measure of the interaction considering the directionality of the sending or receiving of messages.

Cognitive Presence: Refers to the extent to which online learners can construct and validate meanings based on communication and thinking.

Cohesion: Describes the general level of linkage among the actors in the network.

Network Centralization: Refers to the extent to which a network is centered on a subject.

Social Presence: The ability of learners to project themselves socially and affectively into a community of inquiry.

Teaching Presence: The design of the educational experience and facilitation among the teacher and the students.

H

APPENDIX 1

Table 3. Social presence coding scheme

Category	Indicator	Examples
Affective	Expressions of emotions	"This is a little confusing at first..." (JAL_T1_GC_2013)
	Use of humor	"Put a picture in your profile for people to know how cute you are:D" (IRP_T1_GH_2012)
	Self-disclosure	"I have had personal problems" (FMA_T1_GE_2012)
	Apology	"I am sorry, I am late =S" (SPA_T1_GA_2012)
Interactive	Referring explicitly to others' messages	"As our colleague said ..." (PC_T3_GB_2012)
	Asking questions	"How can we know who are the members of our group?" (CME_T1_GC_2012)
	Expressing appreciation	"Thank you very much for the link" (JPT_T3_GA_2013)
	Expressing agreement	"Your thoughts seems very good, I agree with you" (MPF_T3_GG_2013)
	Expressing disagreement	"I understand your position, but I do not agree" (PZV_T3_GC_2013)
	Inviting, suggesting	"What you think if tomorrow we send the activity ..." (BLM_T2_GD_2012)
Cohesive	(Vocatives) Addressing peers by name	"Thanks Ana for offering" (FMA_T1_GF_2012)
	Addressing the group as we, us, our, group	"We have done an excellent job" (TCF_T3_GC_2012)
	Communicating solely for social function	"I'm glad you're feeling better" (MVM_T3_GH_2013)

APPENDIX 2

Table 4. Cognitive presence coding scheme

Category	Indicator	Examples
Triggering	Recognizing the problem	"We must follow some certain guidelines to do the task" (PRM_T3_GC_2013)
Exploration	Divergence of ideas	"We need to reach an agreement" (IRP_T1_GH_2012)
	Exchanging ideas	"Ask about their responsibility, as a result of the problems at school" (LAT_T1_GE_2013)
	Suggestions for consideration	"I'm sharing a very good link, when I saw this video I thought about the issue that concerns us right now" (JMH_T1_GF_2013)
Integration	Convergence among group members	"Just as my colleague, I totally agree that there are thousands of cases" (MDE_T1_GG_2012)
	Connecting ideas, inference, synthesis	"In summary, we have the following tasks..." (VDC_T3_GH_2013)
	Creating solutions	"For example, a possible solution would be to propose extra-curricular activities" (ZMJ_T1_GD_2012)
Resolution	Application to real world	"It is a good idea for people to become aware of what they really know or think they know" (FSY_T2_GB_2012)

APPENDIX 3



Table 5. Metrics for social network analysis

Factor	Measures
Degree Centrality (out/in)	Is the number of incoming and outgoing ties, and indicates how well an actor is connected within the overall network. Valued by: • Centrality (out) is the number of messages sent by the student. • Centrality (in) is the number of responses to messages sent by the student.
Centralization (out/in)	Its measurement is comprised of an overall value between 0-100. A star network (approximately 100 value), for example, would indicate that there was a central subject that everyone was linked to without any other connections among them. Therefore, it is assumed that all communication is channeled by a central actor. On the other hand, a fully connected network (approximately 0 value) means there are no central actors, but all communication is distributed to all members of the group (all linked actors).
Cohesion	The cohesion is measured by the density index (Berkowitz (1982). $Density = 2a / n(n-1)$, where a = the real number of interactions and n = the number of participants in the network. It is calculated by dividing the number of real relationships between the number of possible relationships and multiplying by 100.

Category H

High Performance Computing

Cost Evaluation of Synchronization Algorithms for Multicore Architectures

H

Masoud Hemmatpour
Politecnico di Torino, Italy

Renato Ferrero
Politecnico di Torino, Italy

Filippo Gandino
Politecnico di Torino, Italy

Bartolomeo Montrucchio
Politecnico di Torino, Italy

Maurizio Rebaudengo
Politecnico di Torino, Italy

INTRODUCTION

One of the major issues in modern computer architecture is multicore design. Programmers have been urged to design innovative algorithms by exploiting multicore facilities. Synchronization, i.e., the technique adopted for coordinating threads or processes to have appropriate execution order, is one of the main issues in programming on a multicore processor. In the literature, many synchronization techniques based on hardware and software have been proposed (Petrović, 2014; Yoo, 2013; McKenney, 1998). Modern computers provide special hardware instructions that allow to test and modify the content of a word atomically (e.g., the *cmpxchg* instruction of Intel) which can be used for synchronization of threads (Valois, 1995; Gao, 2007). Software techniques can synchronize threads without any dependency on hardware instructions (McKenney, 1998; Mellor-Crummey, 1991). One important aspect of a synchronization algorithm is its performance, which is evaluated in terms of overhead. In this study, the term cost is used to address an overhead of synchronization algorithm. The state-of-the-art approaches strive to increase the performance by

reducing the cost of the synchronization. To the best of the authors knowledge a study to analyze the possible costs of synchronization mechanisms is absent. So, this chapter investigates the costs in the main steps of a synchronization mechanism. Moreover, since memory access is one of the most important costs in synchronization mechanisms, a discrete time Markov chain model of memory access cost is presented to evaluate the memory access overhead.

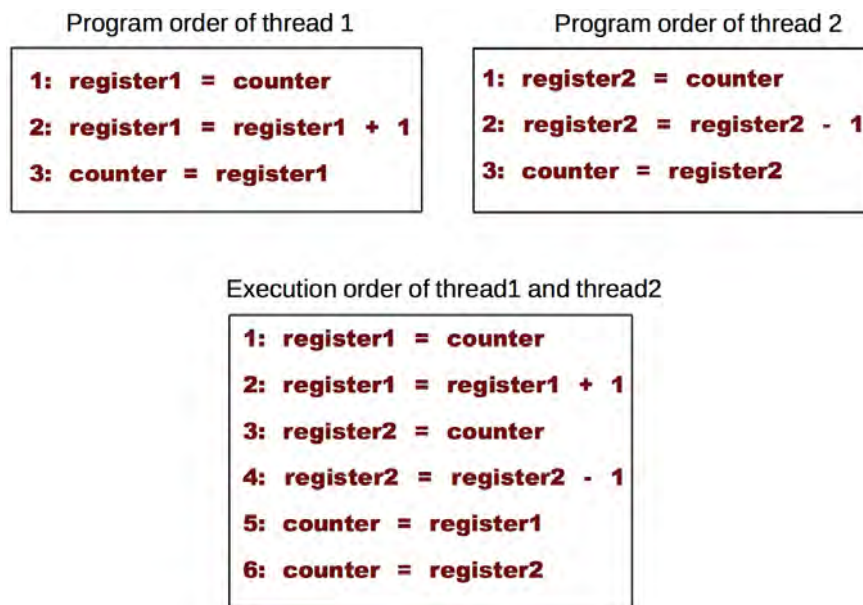
The remainder of this chapter is organized as follows. The primitives of the main synchronization algorithms are described in Section 2. Then, a theoretical evaluation of each cost and experimental results are presented in Section 3. Finally, some conclusions are described in Section 4.

BACKGROUND

When threads are working simultaneously on a shared object, their synchronization should be managed properly, otherwise the instructions of different threads interleave on the shared object in a wrong way. For example, Figure 1 shows the program order of two threads that are working on

DOI: 10.4018/978-1-5225-2255-3.ch346

Figure 1. Incorrect execution of the instructions order



the shared object counter (Silberschatz, 2006). Since one thread is incrementing the counter and another one is decrementing it, at the end, the counter is expected to have the initial value. However, as Figure 1 illustrates, there is a possible execution order of instructions that leads to an incorrect result.

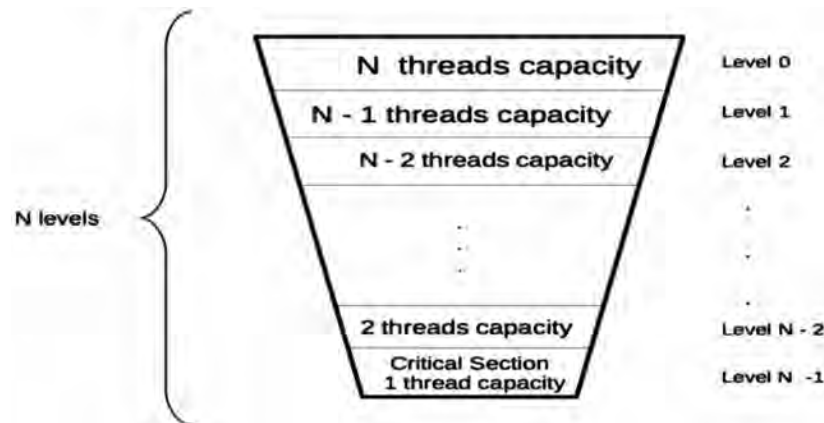
Synchronization mechanisms are used to avoid the problematic interleaving instructions. The part of the code that accesses to the shared object is called critical section. The critical section should be protected by synchronization primitives to avoid concurrent access to the shared object:

- Ticket lock is a synchronization mechanism to guarantee fairness execution to all the threads. Ticket lock is a kind of spin-lock that keeps checking to see if a lock is available to acquire the lock. Ticket locks are conceptually two bytes, one indicating the current head of the queue, and the other one indicating the current tail. The lock is acquired by atomically noting the tail and incrementing it by one (thus adding it to the queue and noting the position), then

waiting until the head becomes equal to the initial value of the tail (Piggin, 2008).

- Filter lock is a synchronization algorithm that works for multiple threads. The filter lock creates $N - 1$ waiting lists, called levels. A thread must pass through all the levels before entering the critical section. As shown in Figure 2, at level 0 at most N threads are waiting. In level 1, at most $N - 1$ threads are waiting. The procedure continues till level $N - 1$, that corresponds to the critical section, in which only one thread is enabled to enter. Each level should satisfy two properties: first, among the threads that are trying to enter into a level, at least one thread succeeds. Second, at least one thread is blocked in the level (Herlihy, 2006).
- Readers -writer lock (rwlock) is intended a solution for solving the readers-writers problem, where a resource is shared among readers and writers. The readers-writers problem happens when a reader starts to read the first item of the shared resource, then stores it in a local variable.

Figure 2. Filter lock structure



Subsequently, a writer updates the shared resource. Afterwards, the same reader reads the second item of the shared resource. Obviously, in this case the reader reads inconsistent data. Rwlock is a solution for solving this problem through postponing a writer till all the readers complete their operations.

- Read-Copy Update (RCU) significantly reduces the overhead in Readers-writer lock (Desnoyers, 2012). The key point is to safely read data even when they are being modified by an update. Coherence between readers is kept by maintaining multiple versions of the same data and ensuring that they are not freed up until all pre-existing read-side critical sections are completed. There are two types of RCU: the general-purpose RCU that uses a global variable to synchronize the readers and the writer, and the signal-based RCU that uses Portable Operating System Interface (POSIX) signals from the write side for synchronization.
- Test_and_set instruction atomically stores '1' in a memory address, and returns its old content. This instruction can be used for implementing a synchronization mechanism: the critical section is protected by `__sync_lock_test_and_set()` and `__sync_lock_release()` statements that use test_and_set instruction inside (Herlihy, 2006).

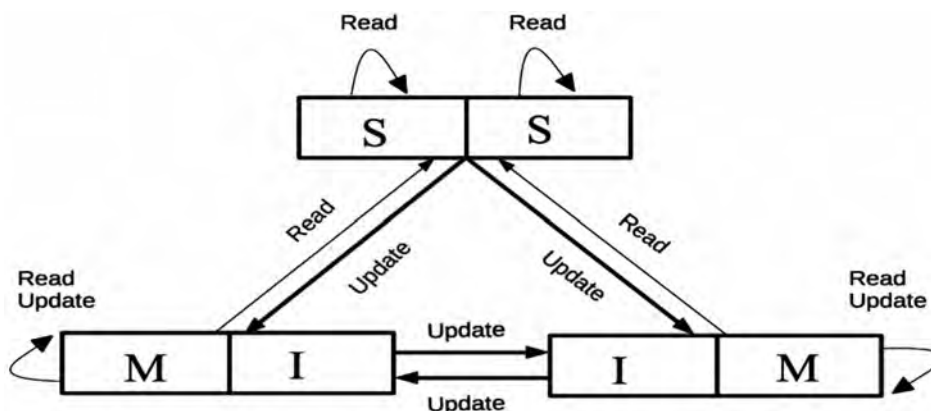
COST OF SYNCHRONIZATION ALGORITHM

In this section, different costs of synchronization procedures are described. Furthermore, experimental results of each cost are presented by exploiting above-mentioned synchronization algorithms on a node of cluster (Opteron 6276 2.3 GHz CPU with 16 cores and running CentOS 6.3 Operating System). Experiments are compiled with gcc 4.4.7. The benchmarks are implemented with empty critical section except the experiments in Section 3.1 that read and write a single shared variable. An empty critical section is selected to eliminate the impact of the critical section on the performance.

MEMORY ACCESS

Moving a synchronization algorithm from a uniprocessor to a multiprocessor environment can improve the performance, but the improvement is not proportional to the number of CPUs due to the cost of communication among processes (or threads). The process communication in a uniprocessor environment has a lower cost due to the lower cache misses (memory access) to retrieve the shared object and flags to keep synchronization in the processor. Instead in a multiprocessor environment, objects and locks are shared among

Figure 3. Cache state



different CPUs. So, updating a shared object in a CPU causes the invalidation of the data cache in the other CPUs. So, the number of cache misses has a big influence on the performance of the synchronization algorithm. In this study, memory access cost is the number of cache misses in a synchronization mechanism. Cache misses are strictly related to data locality, which can be temporal or spatial. Spatial locality means that near addresses are accessed in the near future, so, a block of addresses are fetched to the cache. Temporal locality means that recent memory can be reused in the near future, so, it should be kept in cache memory. In general, there are different techniques to improve data locality. For example, data prefetching technique imports probable data earlier than access time to reduce cache misses. Flat combining (Hendler, 2010), asymmetric multicore (Suleman, 2009) and critical section migrating (Lozi, 2012) delegate the critical section execution to a server thread that improves the cache locality. Cache compression technique compresses recent data in the cache to reuse them in the near future (Alameldeen, 2004). Furthermore, a compiler can optimize the code to improve cache locality (Carr, 1994). In multicore architecture, each core has its own cache.

If an object is shared among some cores, the coherence of their caches should be considered. For example, if a data is shared in the cache of two

cores, and one core updates its value, the data on the other cache is not valid anymore. The cache coherence protocol takes care of this problem in modern CPU architectures. If the data of the cache are not valid, the cache coherence protocol should retrieve the fresh copy of the data from the first accessible memory. The operation of retrieving the fresh data has a cost on the performance of the program. Based on the validity of the data in the cache, different classifications of the cache states are proposed in literature (Sorin, 2011). MESI protocol (Papamarcos, 1984) is a standard cache coherence protocol in modern computer architectures. It classifies the state of the cache to Modified, Exclusive, Shared and Invalid (MESI).

Figure 3 shows the state transition of two caches on different CPUs: each box shows the state of the cache on both CPUs. Initially, an object is shared between two CPUs so the cache states of two CPUs are Shared (S). If a core updates its cache, it modifies its state to Modified (M). By means of a message exchanged with the other caches it notifies other cores that their copies of data are Invalid (I). If a core updates or reads its cache in the Modified state, the state does not change. On the contrary, if it reads its cache in the Invalid state, the state goes back to the Shared state. If a cache with the Invalid state updates its cache, it should retrieve the fresh copy of the data and update it. So, the cache state changes to the Shared state. Also,

the cache that has the copy of the data changes the state to Shared. The thickest lines represent expensive operations in cache coherence protocol.

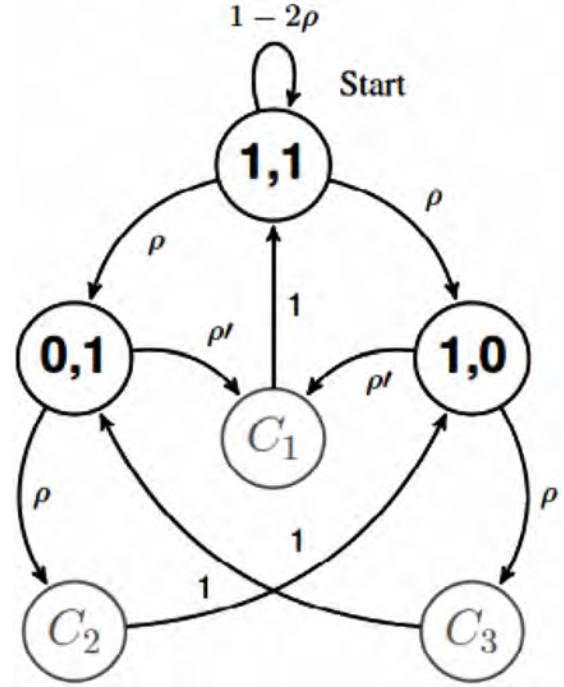
A model is proposed to evaluate the impact of the expensive operations of the cache coherence protocol. The shared object is written and read by both cores at the same rate (ρ is the update probability and ρ' is the read probability). Since update and read of the shared object are discrete time operations, the state of the cache can be modeled by a Discrete Time Markov Chain (DTMC). Figure 4 shows the proposed DTMC model of the cache states of two cores. Each state of the model shows the cache state in the core: 1 means that the state of the cache for the shared object is valid (Modified, Exclusive, Shared), and 0 means that the cache state is invalid (Invalid). Initially, the state of the cache is valid. If a cache updates the shared object, its status remains 1 but the status of the other cache becomes 0. In both (0, 1) and (1, 0) states, if the core with state 0 reads the shared object, the cache coherent protocol should retrieve the fresh copy of the data and the states become (1,1). Moreover, in both states (0, 1) and (1, 0) it might happen that the core with invalid cache state decides to update the shared object. First, it should read the latest data and update the shared object. As Figure 4 shows, when C1, C2 and C3 states are visited, the cost of the cache coherence protocol should be paid.

The average cost of the model in n operations is computed as following:

$$E[\text{Cost}] = \frac{n}{\mu_{c1}} \times E[C_1] + \frac{n}{\mu_{c2}} \times E[C_2] + \frac{n}{\mu_{c3}} \times E[C_3]$$

where $E[C_1]$, $E[C_2]$ and $E[C_3]$ indicate the average cost of cache coherence protocol in associated states. μ_{ci} indicates the expected required operations to go back to state C_i leaving the C_i state. μ_{ci} is computed as:

Figure 4. Markov model



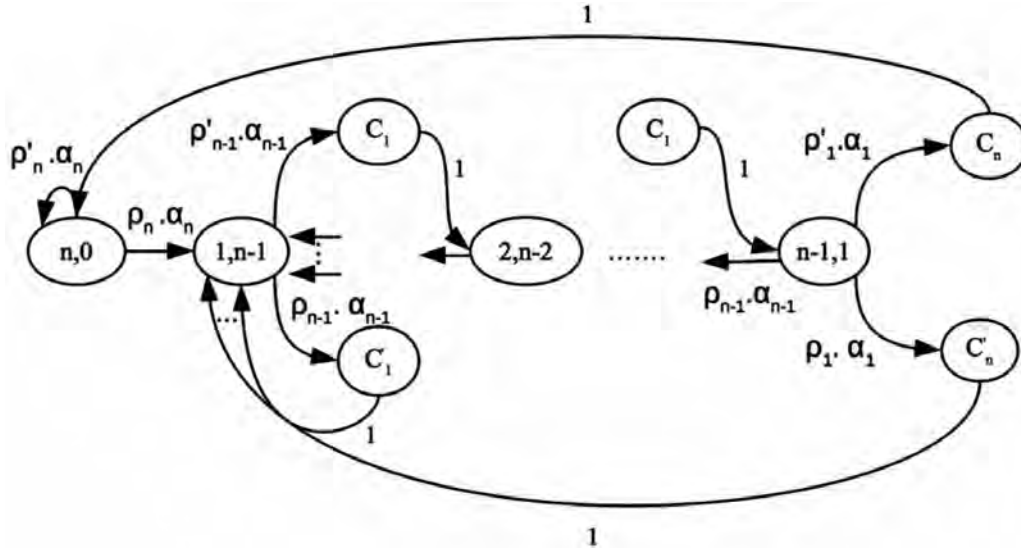
$$\mu_{ci} = \frac{1}{\pi_{iC_i}}$$

where π_i is the stationary distribution of state C_i . Stationary distribution indicates the measures of the distribution in a long process. After computation of μ_{ci} and in case that $E[C_1] = E[C_2] = E[C_3] = C$, the formula will be:

$$E[\text{Total Cost}] = n \times C \times \left[\frac{(4\rho - 3\rho^2)(1 - \rho)}{2\rho^3(3\rho + 1)} \right]$$

As the formula shows, the cost of the cache coherence protocol in n operations has direct relation with ρ (update probability). So, the final formula can be written as a function of ρ . In the following, the function $f(\rho, n)$ shows the average of the cache coherence protocol cost to retrieve a fresh copy of data in n operations:

Figure 5. Generalized model of cache miss cost



$$f(\rho, n) = n \times C \times \left[\frac{(4\rho - 3\rho^2)(1 - \rho)}{2\rho^3(3\rho + 1)} \right]$$

As described before, there are different techniques to attenuate the number of cache misses. Let assume that $g(n)$ indicates the cost of the cache miss attenuation technique in n operations. The technique is feasible if:

$$g(n) < f(\rho, n)$$

A generalized Markov model is presented in Figure 5. Each state (n, m) represents the number of valid cache states (n) and invalid cache states (m). It is assumed that initially all the states are valid $(n, 0)$, and there are n cores in the system. ρ_n' specifies the probability that n cores read their own caches without writing it. ρ_n specifies that at least one of the n cores tries to update its own cache. Moreover, α_n specifies the probability of accessing the bus among n cores.

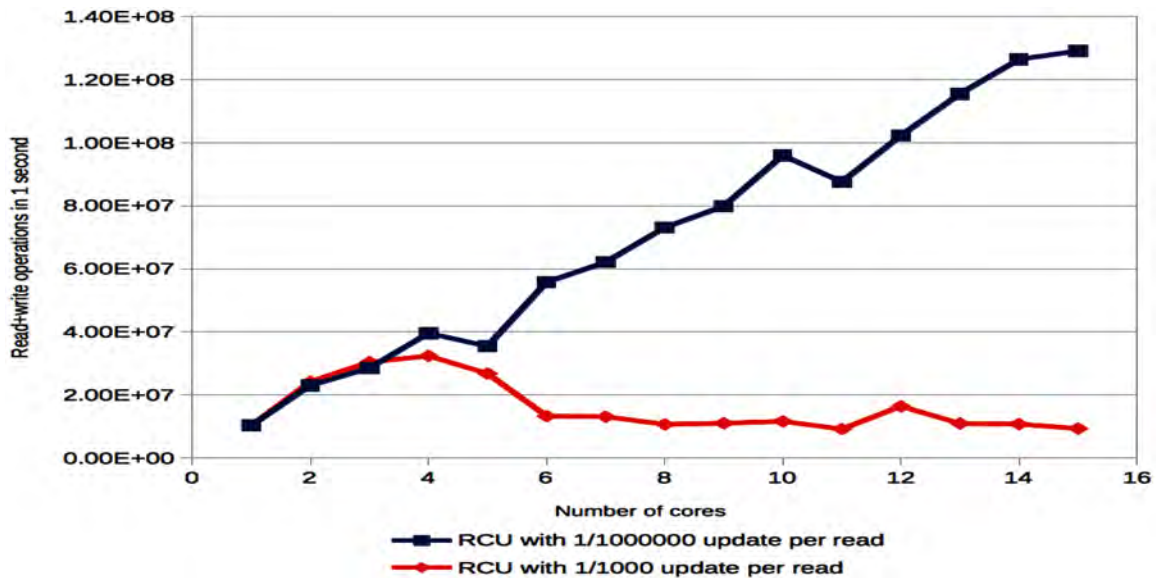
To verify the effect of the memory access on the performance of the synchronization mechanisms, some experiments with RCU technique on different rates of update have been performed. The RCU technique was selected because it is a

non-blocking algorithm and it shows more clearly the cost of the memory access effect in critical section. The experiments have been executed with 1 update out of 1,000 and 1,000,000 reads. Figure 6 shows that the performance of RCU with low update rate is higher. As discussed, on high speed update environment there are higher cache misses and consequently higher memory accesses to retrieve the fresh copy of the data, determining a significant impact on the performance.

SPINNING

In a synchronization mechanism, spinning is the act of querying (or in some cases modifying) an object in memory, and waiting for its content. Usually, it is a busy operation that wastes CPU cycles. When the desired content is achieved, the thread can enter into the critical section. This operation is called lock acquiring. High contention for acquiring a lock generates intensive bus traffic. Moreover, the contention for acquiring the lock increases when the thread that acquired the lock spends a significant CPU time in the critical section. Also, it should be considered that the spinning algorithm on uniprocessor architecture is not

Figure 6. RCU with different ratio of update/read



efficient since a thread will completely waste its time slice spinning for a content change that cannot happen until the releasing thread is scheduled. High contention for acquiring the lock generates bus traffic, and should be properly managed. In this study, spinning cost address the waste CPU cycles of threads to enter critical section.

Exponential backoff is a frequently used method for bus traffic attenuation (Herlihy, 2006; Mellor-Crummey, 1991; Anderson, 1990; Petrovic, 2015). It proposes a delay to access memory. One possible implementation of the exponential backoff technique is shown in Figure 7 (Herlihy, 2006), where *MinDelay* represents the minimum delay, *MaxDelay* represents the maximum delay and *Limit* shows the current delay. First, *Limit* is initialized to *MinDelay*. Then, backoff is set to a random number between 0 and *Limit*. If acquiring the lock was unsuccessful, the limit should be doubled and again backoff is set to a random number between 0 and *Limit*. This procedure can be continued until *Limit* becomes lower than *MaxDelay*.

Intel adds *monitor* and *mwait* instructions to improve performance (Boggs, 2004). The *monitor* instruction specifies a range of addresses to monitor events in the memory range. The *mwait*

instruction is used to indicate that the software thread is waiting for an event. A waiting thread can enter an optimized state until the occurrence of desired events. So, it avoids busy waiting and saves power consumption.

Another technique to implement spinning in an efficient way is scheduling another process or thread when a thread is waiting without performing useful operation. The cost of context switch should be considered. For example, if a thread spins for

Figure 7. A possible implementation of the backoff technique

```

Limit= MinDelay;
Backoff= random(0, Limit);
while(1){
    if (!try_acquire_lock()){
        sleep(Backoff);
        if (Double * Limit< MaxDelay)
            Limit=Double * Limit;
        Backoff= random(0, Limit);
    }else
        break;
}

```

few CPU cycles to acquire the lock, scheduling another process is not efficient because the cost of context switch is higher. So, this technique is useful when threads spin for long time.

A non-blocking algorithm can safely access to shared data without the use of traditional synchronization primitives. There are two types of non-blocking thread synchronization algorithms: lock-free and wait-free. In lock-free systems, any thread may be blocked for some time by other threads. Lock-free algorithms increase the overall throughput of a system by occasionally increasing the latency of a transaction. By contrast, wait-free algorithms ensure that no thread can be blocked by another thread. A read operation of RCU is an example of a wait-free system. An experiment with two different synchronization algorithms, i.e., RCU and readers-writer lock, has been performed in order to verify the effect of spinning on the performance of synchronization. Since RCU is a non-blocking algorithm, the performance of RCU is significantly higher than the readers-writer lock, as shown in Figure 8.

SYSTEM CALL AND HARDWARE SUPPORT INSTRUCTIONS

Some synchronization mechanisms require system calls or hardware instructions to acquire a lock, but the performance is penalized because system calls and hardware instructions are heavy operations. On the contrary, user-level locks such as RCU, ticket lock, filter lock, etc. should be used to optimize the performance. Test-and-set synchronization, which uses test-and-set atomic instruction, and ticket lock, which is a user level synchronization algorithm, are compared in order to verify the impact of hardware instructions. As shown in Figure 9, ticket lock outperforms the test-and-set due to lighter instructions.

BARRIER

Barriers are a hardware/software support to correctly implement a synchronization mechanism. They avoid reordering of store and load operations

Figure 8. RCU synchronization method vs rwlock

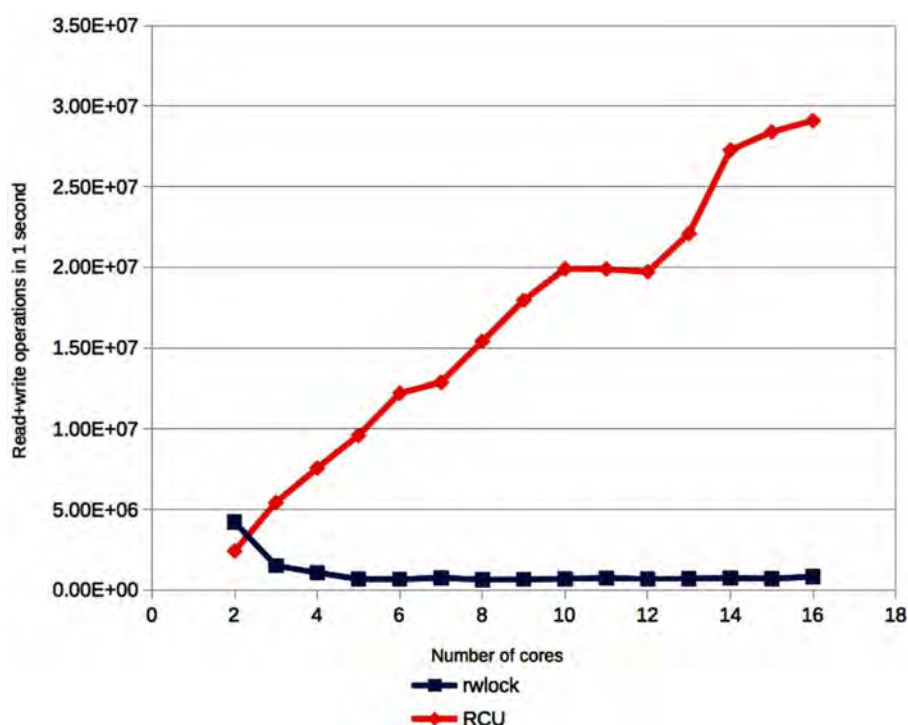
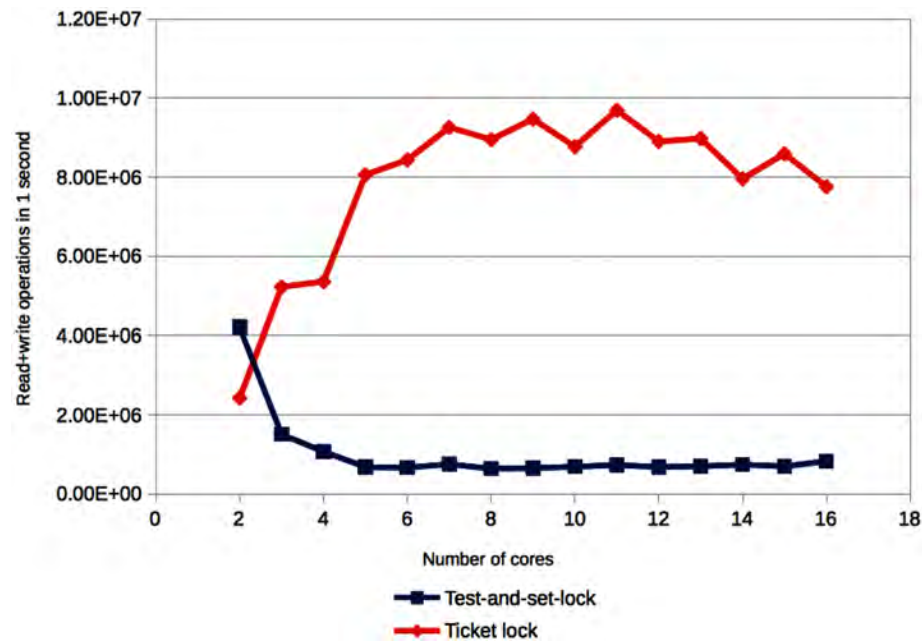


Figure 9. Ticket lock vs test-and-set lock



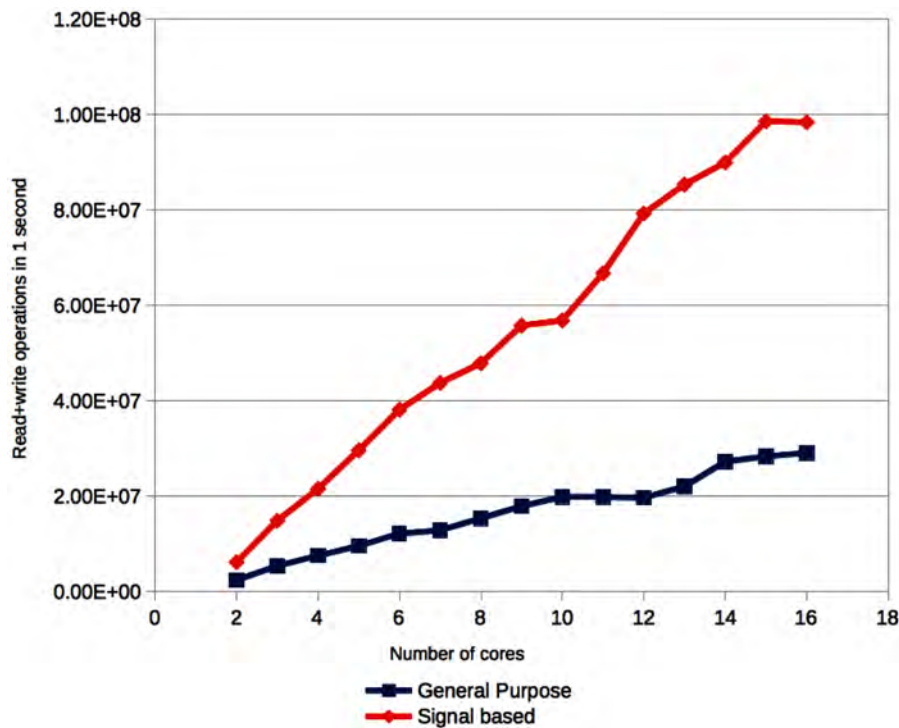
in multi-thread programming in parallel environment. In this study, barrier cost is the overhead of the barrier instruction in a synchronization algorithm. A memory consistency model specifies the behavior of multithreaded programs executing with shared memory (Sorin, 2011). There are different consistency models like Sequential Consistency (SC) and Total Store Order (TSO). In a SC model, instructions are not allowed to reorder neither by compiler nor CPU. Instead, in TSO if a Load instruction is followed by a Store instruction in the program order, this order is not guaranteed in execution. So, in TSO model, programmers (or compilers) can prevent the reordering of mentioned instructions through a barrier instruction between Store and Load. There are two types of barriers. A compiler barrier prohibits any code motion optimization; it is strictly a compiler directive, and it does not generate any code. A full memory barrier avoids reordering instruction not only in compiling time but also in execution time. Whereas full memory barrier is stricter, it has higher impact on the application performance. A possible technique to avoid full memory barrier is promoting compiler barrier

on demand when the writer needs to update the shared object (Desnoyers, 2012). Signal based RCU is a kind of synchronization that exploits compiler barrier, while general purpose RCU uses full memory barrier. An experiment with empty critical section has been performed to observe the performance of both algorithms. As Figure 10 shows, signal-based RCU outperforms general purpose, thanks to compiler barrier.

MESSAGE PASSING

Message passing is a communication technique among threads in multi-threaded programming and it is appropriate for synchronization mechanisms. Message passing cost is the overhead of communication messages among the threads. There are different methods to send messages between threads such as socket, pipe, and queue. Since the cost of sending and receiving message is high, CPU designers intend to provide hardware message passing among the cores. Recently, synchronization methods exploit hardware message passing in their mechanisms (Petrović, 2014). CPU

Figure 10. Signal based RCU vs General purpose RCU



manufacturers (e.g., Single- Chip Cloud by Intel and TILEGX by Tilera) provide a mesh between cores which enable programmers to send/receive messages in a fast way.

of all the approaches reduces and converges together.

CRITICAL SECTION LENGTH

The size of critical section has direct impact on the performance of the synchronization algorithm. Studies have proven that performance of different synchronization mechanisms converge with large critical section size (Desnoyers, 2012; Petrovic, 2015). To observe the effect of long critical section, an experiment with different sizes of critical section has been performed on three different synchronization mechanisms. Figure 11 and Figure 12 show the results with a loop of 1,000 and 100,000 iterations in the critical section, respectively. When the size of the critical section increases, the performance

FUTURE RESEARCH DIRECTIONS

The performance of a synchronization mechanism depends not only on the algorithm, but also on how hardware facilities are exploited. It is likely that the future direction in enhancing the performance of the synchronization mechanism regards exploiting new hardware facilities in sophisticated non-blocking algorithms. Message passing, atomic broadcast, prefetching techniques, and atomic operations can be used in modern processor for an advanced synchronization mechanism. The evaluation of the synchronization costs given in this study can help future research in designing a novel software/hardware synchronization algorithm.



Figure 11. 1000 loop iterations

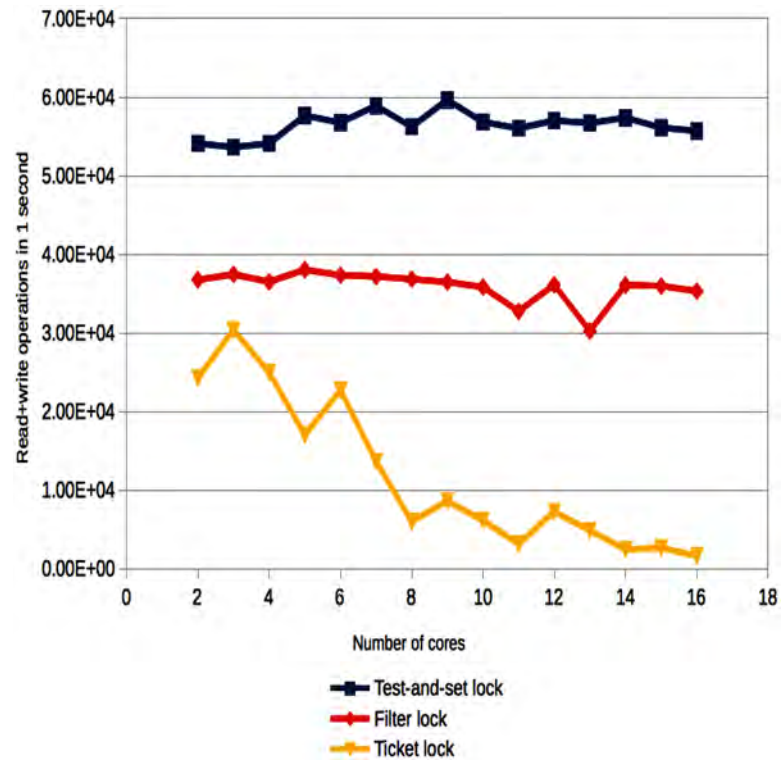
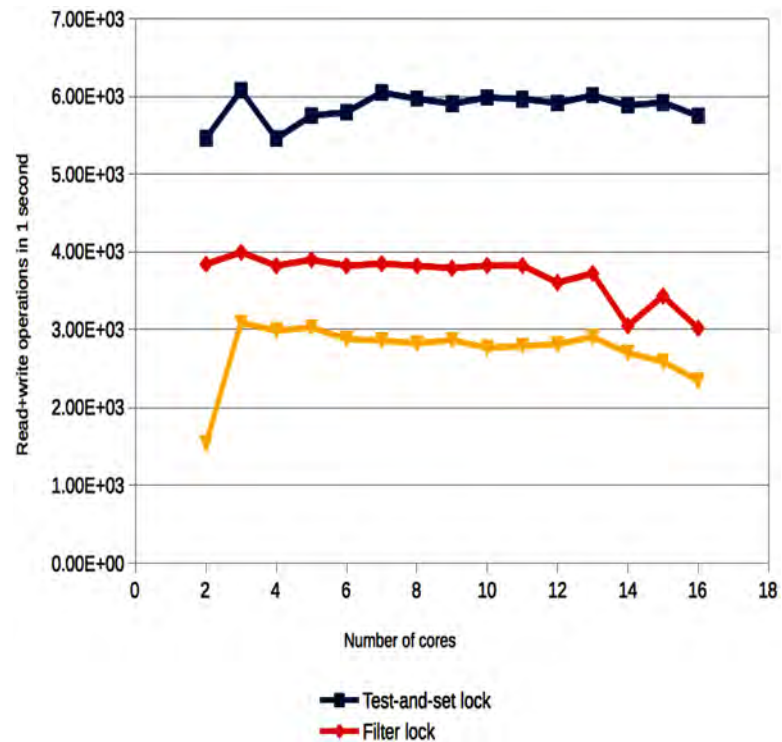


Figure 12. 100,000 loop iterations



CONCLUSION

This study has investigated the costs of synchronization algorithms in parallel environment. The costs of memory access, spinning, system call and hardware support instruction, barrier, message passing and the size of critical section length on the performance of the state-of-the-art approaches have been reviewed and experimental results are presented to support the evaluation. Some good practices are deduced as follows:

- The number of update operations should be reduced because it has a direct impact on the performance of the application.
- If a cache locality technique is used to reduce the number of cache misses, the cost of keeping data locality should not exceed the cost of cache misses.
- A synchronization method without system call or heavy instruction and a non blocking synchronization algorithm lead to a better performance.
- If a synchronization method in shared memory environment is required, choosing a technique with hardware message passing support can increase the performance.
- Since heavy critical section decrease and converge the performance of different synchronization techniques, the critical section length should be reduced as much as possible.

ACKNOWLEDGMENT

The authors thank the High Performance Computing (HPC) project at the Politecnico di Torino (<http://www.hpc.polito.it>) for the computational resources.

REFERENCES

- Alameldeen, A. R. (2004). Adaptive cache compression or high-performance processors. *IEEE Symposium on Computer Architecture*.
- Anderson, T. E. (1990). The performance of spin lock alternatives for shared-memory multiprocessors. *IEEE Transactions on Parallel and Distributed Systems*, 1(1), 6–16. doi:10.1109/71.80120
- Boggs, D. B. (2004). The Microarchitecture of the Intel Pentium 4 Processor on 90nm Technology. *Intel Technology Journal*, 7-23.
- Carr, S. M. (1994). *Compiler optimizations for improving data locality. In Architectural support for programming languages and operating systems* (pp. 252–262). ACM.
- Desnoyers, M. M., McKenney, P. E., Stern, A. S., Dagenais, M. R., & Walpole, J. (2012). User-level implementations of read-copy update. *IEEE Transactions on Parallel and Distributed Systems*, 23(2), 375–382. doi:10.1109/TPDS.2011.159
- Gao, H., & Hesselink, W. H. (2007). A general lock-free algorithm using compare-and-swap. *Information and Computation*, 205(2), 225–241. doi:10.1016/j.ic.2006.10.003
- Hendler, D. I. (2010). Flat combining and the synchronization-parallelism tradeoff. Symposium on Parallelism in Algorithms and Architectures, 355-364.
- Herlihy, M. (2006). *The art of multiprocessor programming*. Morgan Kaufmann. doi:10.1145/1146381.1146382
- Lozi, J. P. (2012). Remote core locking: migrating critical section execution to improve the performance of multithreaded application. *USENIX Annual Technical Conference*, 65-76.

McKenney, P. E. (1998). Read-copy update: Using execution history to solve concurrency problems. *Parallel and Distributed Computing and Systems*, 509-518.

Mellor-Crummey, J. M., & Scott, M. L. (1991). Algorithms for scalable synchronization on shared-memory multiprocessors. *ACM Transactions on Computer Systems*, 9(1), 21-65. doi:10.1145/103727.103729

Papamarcos, M. S. (1984). A low-overhead coherence solution for multiprocessors with private cache memories. In *Computer architecture* (pp. 348-354). ISCA.

Petrović, D. R. (2014). *Leveraging Hardware message passing for efficient thread message passing*. ACM SIGPLAN Notices.

Petrovic, D. T. (2015). on the performance of delegation over cache-coherent shared memory. *International Conference on Distributed Computing and Networking ACM*, (pp. 17:1--17:10). doi:10.1145/2684464.2684476

Piggin, N. (2008). Retrieved 07 17, 2016, from <http://git.kernel.org/cgiit/linux/kernel/git/torvalds/linux.git/commit/?id=314cdebfd1fd0a7acf3780e9628465b77ea6a836>

Silberschatz, A. G. (2006). *Operating System Principles*. John Wiley and Sons.

Sorin, D. J. (2011). *A primer on memory consistency and cache coherence*. Synthesis Lectures on Computer Architecture.

Suleman, M. A. (2009). Accelerating critical section execution with asymmetric multicore architectures. *ACM SIGARCH Computer Architecture News*, 253-264.

Valois, J. D. (1995). Lock-free linked lists using compare-and-swap. *ACM Symposium Principles of Distributed Computing*, 214-222.

Yoo, R. M. (2013). Performance evaluation of Intel transactional synchronization extensions for high performance computing. *Conference for High Performance Computing, Networking, Storage and Analysis*, 19:1-19:11.

ADDITIONAL READING

Busaba, F. Y., Cain, H. W., III, Gschwind, M. K., Michael, M. M., Salapura, V., Schwarz, E. M., & Shum, C. L. K. (2016). *U.S. Patent No. 9,329,890*. Washington, DC: U.S. Patent and Trademark Office.

Dan, A., Meshman, Y., Vechev, M., & Yahav, E. (2016). Effective abstractions for verification under relaxed memory models. *Computer Languages, Systems & Structures*.

David, T., Guerraoui, R., & Trigonakis, V. (2013, November). Everything you always wanted to know about synchronization but were afraid to ask. In *Proceedings of the Twenty-Fourth ACM Symposium on Operating Systems Principles* (pp. 33-48). ACM. doi:10.1145/2517349.2522714

Dechev, D., Pirkelbauer, P., & Stroustrup, B. (2010, May). Understanding and effectively preventing the ABA problem in descriptor-based lock-free designs. In *2010 13th IEEE International Symposium on Object/Component/Service-Oriented Real-Time Distributed Computing* (pp. 185-192). IEEE. doi:10.1109/ISORC.2010.10

Gramoli, V., Guerraoui, R., & Trigonakis, V. (2012, April). TM 2 C: a software transactional memory for many-cores. In *Proceedings of the 7th ACM european conference on Computer Systems* (pp. 351-364). ACM.

Kuznetsov, P., & Ravi, S. (2011, December). On the cost of concurrency in transactional memory. In *International Conference On Principles Of Distributed Systems* (pp. 112-127). Springer Berlin Heidelberg. doi:10.1007/978-3-642-25873-2_9

- Matveev, A., Shavit, N., Felber, P., & Marlier, P. (2015, October). Read-log-update: a lightweight synchronization mechanism for concurrent programming. In *Proceedings of the 25th Symposium on Operating Systems Principles* (pp. 168-183). ACM. doi:10.1145/2815400.2815406
- McKenney, P. E. (2011). Is parallel programming hard, and, if so, what can you do about it?. *Linux Technology Center, IBM Beaverton*.
- McKenney, P. E., & Slingwine, J. D. (1998, October). Read-copy update: Using execution history to solve concurrency problems. In *Parallel and Distributed Computing and Systems* (pp. 509-518).
- Menon, P., Rabl, T., Sadoghi, M., & Jacobsen, H. A. (2014, November). Optimizing key-value stores for hybrid storage architectures. In *Proceedings of 24th Annual International Conference on Computer Science and Software Engineering* (pp. 355-358). IBM Corp.
- Michael, M. M. (2004). Hazard pointers: Safe memory reclamation for lock-free objects. *IEEE Transactions on Parallel and Distributed Systems*, 15(6), 491–504. doi:10.1109/TPDS.2004.8
- Petrović, D., Shahmirzadi, O., Ropars, T., & Schiper, A. (2012, June). High-performance RMA-based Broadcast on the Intel SCC. In *Proceedings of the twenty-fourth annual ACM symposium on Parallelism in algorithms and architectures* (pp. 121-130). ACM. doi:10.1145/2312005.2312029
- Petrovic, D., Shahmirzadi, O., Ropars, T., & Schiper, A. (2012). Asynchronous broadcast on the intel scc using interrupts. In *5th Many-core Applications Research Community (MARC) Symposium* (No. EPFL-CONF-177890).
- Rabl, T., Gómez-Villamor, S., Sadoghi, M., Muntés-Mulero, V., Jacobsen, H. A., & Mankovskii, S. (2012). Solving big data challenges for enterprise application performance management. *Proceedings of the VLDB Endowment*, 5(12), 1724–1735. doi:10.14778/2367502.2367512
- Rabl, T., Zhang, K., Sadoghi, M., Pandey, N. K., Nigam, A., Wang, C., & Jacobsen, H. A. (2012, July). Solving manufacturing equipment monitoring through efficient complex event processing: DEBS grand challenge. In *Proceedings of the 6th ACM International Conference on Distributed Event-Based Systems* (pp. 335-340). ACM. doi:10.1145/2335484.2335520
- Ropars, T., Martsinkevich, T. V., Guermouche, A., Schiper, A., & Cappello, F. (2013, November). SPBC: Leveraging the characteristics of MPI HPC applications for scalable checkpointing. In *2013 SC-International Conference for High Performance Computing, Networking, Storage and Analysis (SC)* (pp. 1-12). IEEE.
- Ropars, T., & Morin, C. (2010, August). Improving message logging protocols scalability through distributed event logging. In *European Conference on Parallel Processing* (pp. 511-522). Springer Berlin Heidelberg. doi:10.1007/978-3-642-15277-1_49
- Ropars, T., & Morin, C. (2011). Active optimistic and distributed message logging for message-passing applications. *Concurrency and Computation*, 23(17), 2167–2178. doi:10.1002/cpe.1775
- Sadoghi, M., Canim, M., Bhattacharjee, B., Nagel, F., & Ross, K. A. (2014). Reducing database locking contention through multi-version concurrency. *Proceedings of the VLDB Endowment*, 7(13), 1331–1342. doi:10.14778/2733004.2733006
- Sadoghi, M., Singh, H., & Jacobsen, H. A. (2011, June). Towards highly parallel event processing through reconfigurable hardware. In *Proceedings of the Seventh International Workshop on Data Management on New Hardware* (pp. 27-32). ACM. doi:10.1145/1995441.1995445
- Saraswat, V. A., Jagadeesan, R., Michael, M., & Von Praun, C. (2007, March). A theory of memory models. In *Proceedings of the 12th ACM SIGPLAN symposium on Principles and practice of parallel programming* (pp. 161-172). ACM.

Shalev, O., & Shavit, N. (2006). Split-ordered lists: Lock-free extensible hash tables. [JACM]. *Journal of the ACM*, 53(3), 379–405. doi:10.1145/1147954.1147958

Triplett, J., McKenney, P. E., & Walpole, J. (2010). Scalable concurrent hash tables via relativistic programming. *Operating Systems Review*, 44(3), 102–109. doi:10.1145/1842733.1842750

Wang, Y., Meng, X., Zhang, L., & Tan, J. (2014, November). C-hint: An effective and reliable cache management for rdma-accelerated key-value stores. In *Proceedings of the ACM Symposium on Cloud Computing* (pp. 1-13). ACM. doi:10.1145/2670979.2671002

Wang, Y., Zhang, L., Tan, J., Li, M., Gao, Y., Guerin, X., & Meng, S. et al. (2015, November). HydraDB: a resilient RDMA-driven key-value middleware for in-memory cluster computing. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis* (p. 22). ACM. doi:10.1145/2807591.2807614

KEY TERMS AND DEFINITIONS



Critical Section: A part of the code that access to the shared object.

Memory Barrier: An operation to avoid the reordering of instructions.

Multicore Architecture: A processor with two or more cores, i.e., independent processing units.

Performance: Number of operations in the unit of measure.

Race Condition: Attempt to read and write a shared object, by more than one thread with an undefined behavior.

Spinning: The act of querying (or in some cases modifying) an object, and waiting till desired content is achieved, before entering into the critical section.

Synchronization: A technique for coordinating threads or processes to have appropriate execution order.

The Future of High-Performance Computing (HPC)

Herbert Cornelius

Intel Corporation EMEA, Germany

INTRODUCTION

High-Performance Computing (HPC) is used to address and solve the world's most complex computational problems. For decades, HPC has established itself as an essential tool for discoveries, innovations and new insights in science, research and development, engineering and business across a wide range of application areas in academia and industry. It has become an integral part of the scientific method – the third leg along with theory and experiment.

Today, High-Performance Computing is also well recognized to be of strategic and economic value – HPC matters and is transforming industries (Osseyran & Giles, 2015).

High Performance Computing enables scientists and engineers to solve complex and large science, engineering, and business problems using advanced algorithms and applications that require very high compute capabilities, fast memory and storage, high bandwidth and low latency throughput, high fidelity visualization, and enhanced networking.

Today, the IT industry is being transformed by cloud, big data, social media, artificial intelligence, and “Internet of Things” technologies and business models. All of these trends require advanced computational simulation models and powerful highly scalable systems. Hence, sophisticated HPC capabilities are critical to the organizations and companies that want to establish and enhance leadership positions in their respective areas.

Some industry verticals and application areas where HPC is used are as follows:

- Manufacturing, Computer Aided Engineering (CAE)
- Automotive Industry
- Aerospace Industry
- Weather Forecast and Climate Research
- Energy, Oil & Gas Industry, Geophysics
- Life-Science and Bio-Informatics (Genomics)
- Government Research Laboratories
- Universities (Academics), Machine Intelligence, Machine/Deep Learning, Artificial Intelligence (AI)
- Astrophysics, High-Energy Physics, Computational Chemistry, Material Science
- Financial Services Industry (FSI)
- Digital Content Creation (DCC)
- Defense
- Security and Intelligence

BACKGROUND

After its initial years of proprietary computer systems in the 1970/1980's, HPC has evolved with industry standards that democratized Supercomputing, making advanced computing available to more users and wider application segments.

Today's modern HPC solutions are utilizing high-performance server compute nodes connected with high performance fabrics connected to high-performance storage systems, mainly deployed on distributed cluster architectures running on Linux based operating systems with up to tens of thousands of processors. For specific workloads and applications with the need for large

coherent shared memory capacity (terabytes of data: TB), more specialized solutions and systems are used based on cc:NUMA (Cache-Coherent Non-Uniform Memory Access) architectures. For example, the SGI UV system supports up to 256 CPU sockets and up to 64TB of cache-coherent shared memory in a single system.

While in the past chip designs used to be limited by space and the number of transistors available, now power consumption is becoming the main constraint for High-Performance Computing. With several new emerging technologies there will be multiple opportunities to address some of the ongoing challenges in HPC such as balanced architectures, energy efficiency, density, reliability, resiliency, sustainability, and last but not least ease-of-use.

As stated in a recent article (Bader, 2016), the trend in high-performance computing (HPC) to adopt more accelerators such as graphics processing units (GPUs), field-programmable gate arrays (FPGAs), and coprocessors has led to significant heterogeneity in computation and memory subsystems. Application developers typically employ a hierarchical message passing interface (MPI) programming model across the cluster's compute nodes for internode communication, and an intranode model (normally based on shared memory) such as OpenMP for the CPUs or an accelerator-specific library (usually based on an off-load model) such as compute unified device architecture (CUDA) or open computing

language (OpenCL) for accelerator devices within each compute node of a cluster.

The following sections will discuss some of the new emerging technologies that are being developed for all parts of HPC to enable balanced solutions regarding data processing (compute), memory, storage, fabric and I/O, as well as parallel software and programming aspects for HPC.

Processing

A modern HPC solution can be characterized by its functional high level components consisting of the following parts as shown in Figure 1:

- **Compute:** Advanced calculations with data in fast memory performing mathematical and logical operations.
- **Processing:** Including data movements, data packing/unpacking, compression/decompression and encryption/decryption of data.
- **Visualization:** High fidelity visualization of results and data, often in-situ.
- **Storage:** fast and high capacity permanent data storage.
- **Fabric:** High performance interconnect to connect all parts of a HPC solution together.
- **Networking:** To provide access to other systems and traditional IT infrastructures.

Figure 1. High level view of HPC solution components

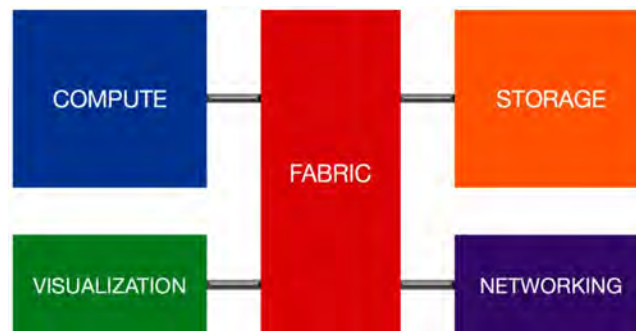
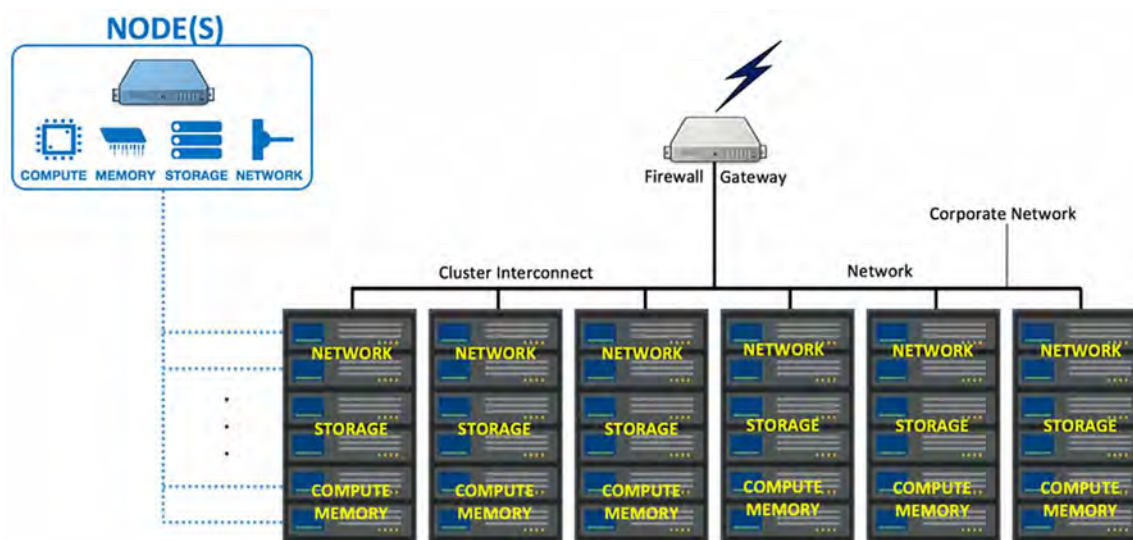


Figure 2. Schematic overview of a HPC Cluster solution



The main HPC solution architecture today is cluster based, consisting of a certain number of shared memory nodes, each with several processors/cores for compute, memory, local storage and networking devices as shown in Figure 2. The cluster nodes are connected together via a high-performance HPC fabric and can be used for a variety of tasks like (parallel) computing, visualization, I/O, storage, networking, login/interactivity and management.

Future HPC systems are expected to continue to deploy enhanced clustering architectures and technologies to achieve scalable high performance in an energy efficient way.

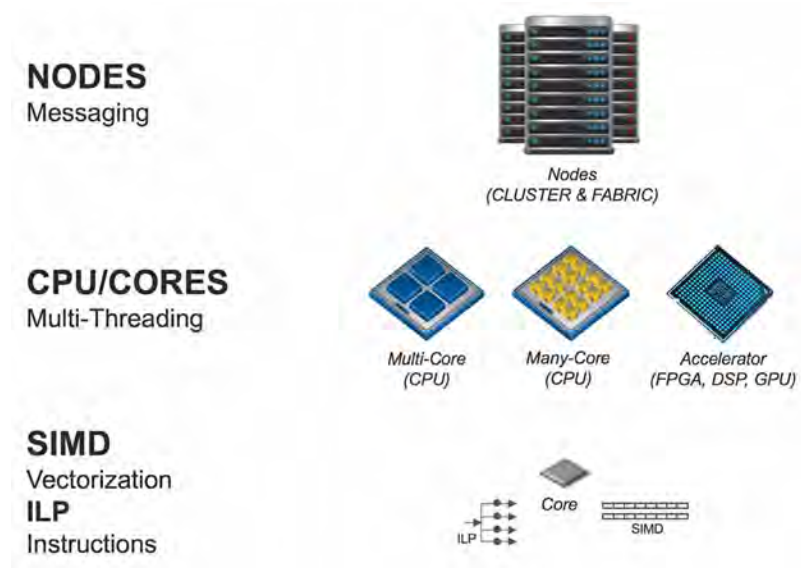
As it is recognized today, the common theme in the industry for high performance and energy efficiency is to utilize more and more parallelism on all computing platforms and at all levels of integration. Thus, in general, a modern HPC solution deploys several levels of parallelism concurrently as shown in Figure 3:

- **Clustering:** Shared memory nodes connected via a high-performance fabric.
- **Multi-Threading:** Across all processors and cores in a shared memory cluster node.

- **Acceleration:** Heterogeneous accelerators (DSP, GPU, FPGA) connected to cluster nodes.
- **Vectorization:** Single Instruction Multiple Data (SIMD) processing within each core.
- **Parallel Instructions:** Instruction Level Parallelism (ILP) inside each core.

On the processor side, multi-core CPUs are being used across the industry. In general, a multi-core processor consists of a number of cores on a single die (chip) with an on-chip interconnect connecting all the cores, memory and I/O-subsystem together, providing a cache coherent shared memory environment. The number of cores inside a multi-core processor have evolved over time to 10-20, and this trend is predicted to increase only moderately in the future. Multi-core processors use a relatively large number of transistors for each core, so those cores are often called “big cores”. Examples of “big core” multi-core processors are Intel Xeon, AMD Opteron, IBM POWER, and SPARC based CPUs, while “small core” examples are ARM architecture based CPUs, normally featuring a lower number of cores with somewhat less performance per core.

Figure 3. Levels of parallelism in modern HPC solutions

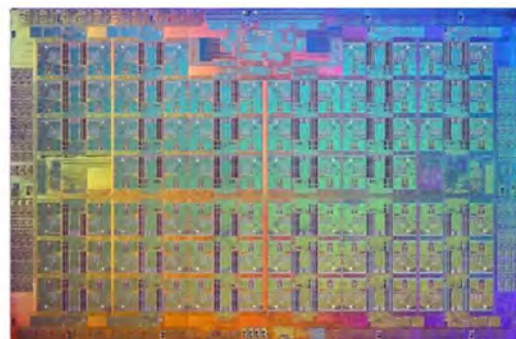
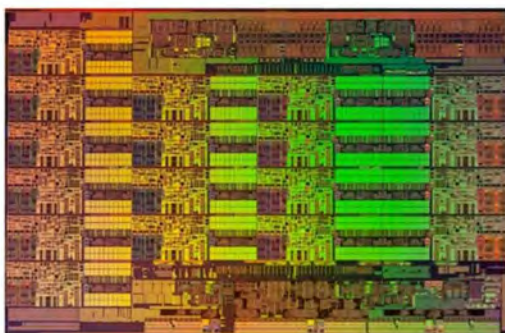


A design goal for big core multi-core processors is to achieve high-performance and powerful capabilities for each single core (thread). If the single thread performance of processors continues to increase, then applications will continue to execute faster. However, an important factor in single thread performance is the processor's clock frequency, and that has plateaued in recent years. This is because the chip's power requirement increases with the cube of the clock frequency ($Power \propto Frequency^3$), and the frequency of today's processors are limited by this equation.

In recent years the industry has extended the concept of multi-core to many-core processors

by integrating more and more "small" cores onto a single chip. Though each core utilizes fewer transistors and runs at a lower frequency, but the aggregated processing power of such a large number of cores exceeds the processing power of a multi-core processor of similar size and power consumption. An example of a general purpose many-core processor is the Intel Xeon Phi 2nd generation (code name Knights Landing) with up to 72 cache coherent cores connected by a 2D-mesh on-chip interconnect. Examples of special purpose many-core processors are GPUs and DSPs with large numbers of simple specialized execution units.

Figure 4. Multi-core processor chip (left) and many-core chip (right). Images not to scale (Source: Intel)



Actual chip die photos of a general purpose multi-core and many-core processor are shown in Figure 4.

The trend to more and more cores per processor chip will continue, mainly driven by power constraints and improvements of the manufacturing process which continues to allow the integration of more and more transistors into smaller and smaller chips. The state-of-the-art volume logic process technology today is 14nm with 3D-gate and FinFET technologies moving to 10nm soon, and the industry projection is that it should be possible to go down to about 5nm structures in the future (Waldrop, 2016).

In order to increase performance and keep power consumption low, new core architectures and additional capabilities are being developed and implemented. Besides using more cores, one of the technologies in this area for data level parallelism is Single Instruction Multiple Data (SIMD). This performs the same operation on multiple data elements simultaneously. Examples for SIMD implementations are the Intel SIMD instruction set extensions (MMX, SSE, AVX, AVX2, AVX-512), AMD 3DNow!, IBM AltiVec, SPARC VIS, and as well as ARM's NEON and the recently announced SVE (Scalable Vector Extension) technology for ARMv8-A architecture based processors. It is projected that SIMD technologies will be enhanced and continue to be an important part of future HPC focused processors.

In addition the industry continues to explore on-going improvements in the area of cache technologies to keep data close to the cores for fast access, out-of-order execution and parallel superscalar execution to perform as many instructions and operations concurrently as possible (Instruction Level Parallelism: ILP).

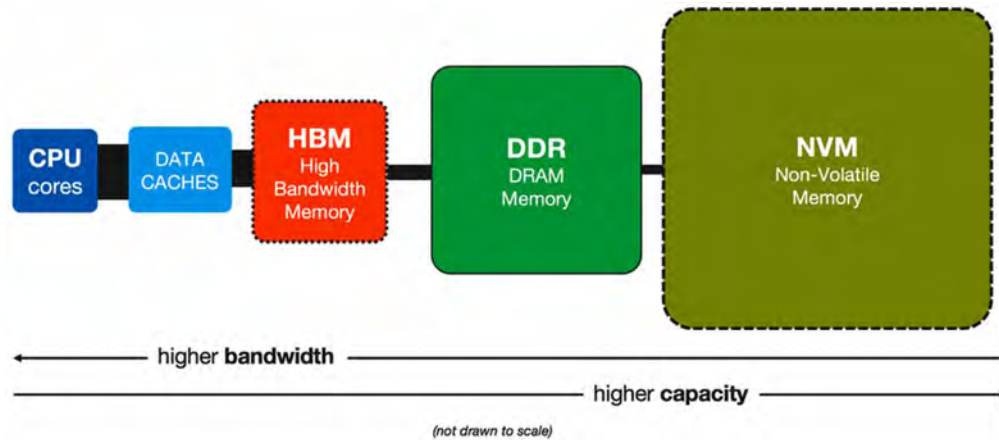
Because of the ever-increasing demand for more performance in HPC, people have always considered additional acceleration technologies to augment general purpose computing solutions. For specific and suitable workloads and algorithms, accelerators can indeed deliver higher performance with lower power, increasing the performance/watt

ratio – which means higher energy efficiency. Such accelerators are usually provided by PCI Express (PCIe) add-in cards installed in some of the compute or I/O cluster nodes. Traditional accelerators include Digital Signal Processors (DSPs) and Graphics Processing Units (GPUs), each with its own programming model and programming language. With advancements in technology, Field-Programmable Gate Arrays (FPGAs) are emerging with high energy efficiency and high performance to go beyond the capabilities of GPU and DSP acceleration. For example, Altera's 14nm Stratix 10 FPGA is capable of 10TFLOPS peak performance (single-precision IEEE-754) and delivering up to ~100GFLOPS (FP32) performance per watt, and future FPGA products from Xilinx are also expected to deliver higher performance and increased energy efficiency.

In the past, FPGAs have been difficult to program, but now they use modern programming models like OpenCL (and potentially OpenMP in the future) allowing them to be used more easily, harnessing the full performance potential of a flexible and freely re-programmable accelerator. It is projected that it will be possible to integrate FPGAs more tightly with general-purpose processors to eliminate the PCIe bottleneck of the FPGA add-in cards to provide more features and more capabilities such as shared virtual memory between CPUs and FPGAs. Microsoft's project Catapult is a recent example of how modern FPGAs can be integrated to accelerate particular computational tasks (Wired, 2016).

For some more specific workloads custom ASICs – explicitly designed to fit the specific needs – might be a reasonable solution for high performance and energy efficiency, but they are not reconfigurable devices that can be programmed and adopted to evolving requirements. An example of such a solution is Google's Tensor Processing Unit (TPU), a custom ASIC built specifically for its TensorFlow machine learning framework, and the planned specific ASIC for deep learning by Intel utilizing its acquired Nervana technology.

Figure 5. Illustration of an evolving new memory hierarchy (not drawn to scale)



This trend is also outlined in a recent IEEE article (Moore, 2016) about other types of hardware accelerators. “We have to improve performance by improving energy efficiency. The only way to do that is to move some software to hardware. The challenge is to figure out which software is used frequently enough that we could justify implementing it in hardware,” he (Yan Solihin, NC State) says. “There is a sweet spot.”

Other new evolving technologies and designs being researched are Quantum Computing (D-Wave), Cognitive and Neuromorphic Computing (IBM and its TrueNorth chip), and HPE’s “The Machine” project utilizing a memory centric system architecture approach rather than a more traditional compute centric design point.

It is clear that parallelism and heterogeneity will play a major role in future HPC solutions, for both performance and energy efficiency reasons. See (Zahran, 2017) for a good overview on this subject.

Memory, Storage and I/O

It is well recognized that memory performance is evolving at a slower pace than compute performance. This increasing unbalance poses an ongoing challenge for balanced HPC solutions. There is an urgent need to develop new memory technologies and architectures to allow higher

performance and eventually also higher memory capacities. Fortunately, several new memory technologies are being developed and will become available in the near future to overcome the respective roadblocks.

While traditional DDR memory like DDR4 will continue to be present in HPC solutions for some time, it is not clear yet if and how a potential next generation of DDR memory (likely to be called DDR5?) will look like and what its future timeline might be.

Modern memory hierarchies already utilize several levels of data caching to mitigate the performance differences between processing and memory, some implementations go up to 4 levels of private/shared caches. In addition sometimes additional memory buffers are used to speed-up DDR based memory performance for both latency and bandwidth.

It is envisioned that in the future the single monolithic DDR memory subsystem will be complemented - or even be replaced over time – by a combination of somewhat lower capacity but much higher bandwidth memory closer integrated to the CPU and much larger capacity but lower bandwidth non-volatile memory (NVM); see Figure 5.

For the high-bandwidth memory part several approaches are being developed by the industry like Hybrid Memory Cube (HMC), High-Band-

width Memory (HBM), and Multi-Chip DRAM (MCDRAM: 3D stacked DDR memory). A high-bandwidth memory could be used as explicit managed memory or as transparent cache to DDR/NVM memory, its initial capacities are projected to be a few gigabytes (GB) with hundreds of gigabytes per second bandwidth.

The new non-volatile memory 3D XPoint technology, jointly developed by Intel and Micron, will allow to use it for non-volatile memory DIMMs (byte addressable) as well as in SSDs for I/O block devices (see below) for fast storage (Figure 6). A key feature of this technology is higher performance in general and specifically a dramatically improved latency and endurance compared to today's NAND (Flash) based NVM technologies, up to 1000 times higher on the technology level. It will allow users to do things with storage that are fundamentally different from what have been done before. Other new technologies under development in this area are Hybrid Memory Cube (HMC) by Micron, Memristor driven by Hewlett-Packard Enterprise, High Bandwidth Memory (HBM) by Samsung and SK Hynix, Phase Change Memory (PCM) by IBM and 3D-ReRAM by Toshiba/Western Digital. Those technologies are also sometimes referred to as Storage Class Memory (SCM) because they can be used for both memory (byte-addressable load/store) and for storage (block-I/O read/write) solutions.

Going beyond and away from traditional DDR memory, there is also the idea of what is called

“Universal Memory”, a single layer of persistent (non-volatile) memory that can provide much increased memory capacities with low latencies, high bandwidth and enough performance to support the emerging memory and data driven computing environments. All before mentioned NVM technologies are potential candidates for new server memory solutions in the future.

Also on the storage side there is a need to improve performance to have balanced HPC systems. Using PCI Express 3.0 based I/O interfaces, the delivered performance is reasonable good depending on how many I/O lanes are utilized. PCI Express 3.0's 8 GT/s bit rate delivers up to 985 MB/s per x1 lane. In HPC it is very common to use x4, x8 or x16 PCIe 3.0 based I/O devices such as accelerators, SSDs and fabric. The integration of PCIe lanes directly into the processor also helped speeding up I/O. The industry is currently working on the next generation PCIe 4.0 specification for a 16 GT/s bit rate per lane, thus doubling the bandwidth of PCI Express 3.0 and maintaining backward compatibility. The final industry standard specifications of PCI Express 4.0 are estimated to be available in the 2017 timeframe.

The benefits of Solid State Drives (SSDs) over Hard Disk Drives (HDDs) for storage are already proven for multiple reasons such as much higher performance (higher bandwidth, lower latency, more IOPS), energy efficiency and better robustness due to missing moving parts of an HDD. SSDs typically use flash memory based on

Figure 6. Illustration of 3D XPoint (left) and 3D-NAND technology (right). Images not to scale (Source: Intel and Micron).

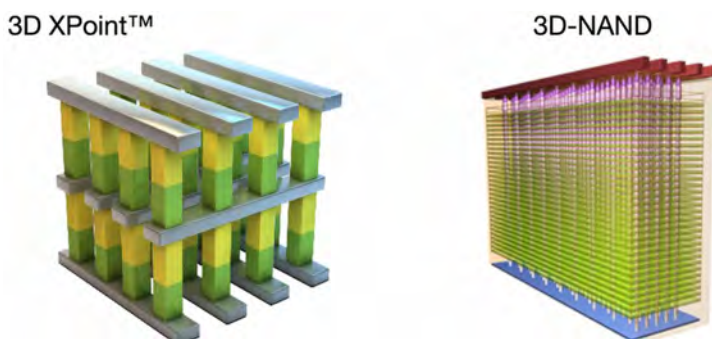
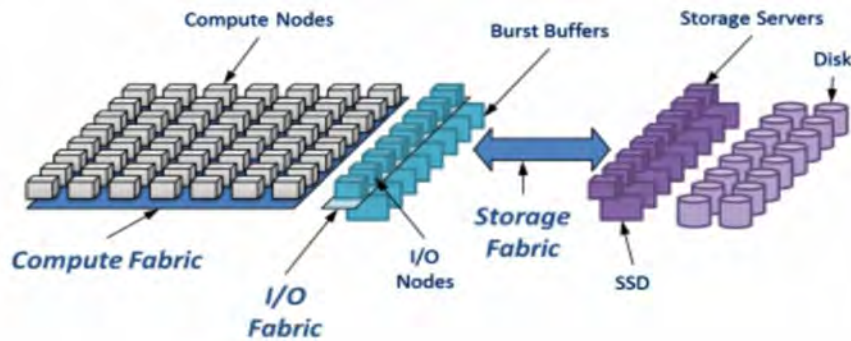


Figure 7. Schematic overview of a Burst Buffer architecture concept



NAND technologies to permanently store data (non-volatile, persistent). While today's SSD capacities can go up to 16TB using 2D/3D-NAND technologies and being well on par with HDD capacity, emerging newer 3D-NAND technologies (Figure 6) will enable SSDs to increase its capacities dramatically, and potentially overtake HDDs in capacity in the future. To get additional performance especially for PCIe based SSDs, the Non-Volatile Memory Express (NVMe) protocol was introduced a few years ago, improving both random and sequential performance by reducing latency, higher levels of parallelism, lower overhead and optimized control flow.

Furthermore, from a HPC system architecture point of view the concept of an I/O and storage Burst Buffer is emerging, initially around the widely used high performance global parallel file system Lustre. The idea of a Burst Buffer is to use it as a layer of (fast) non-volatile NVM storage that is located between the processors' memory and the parallel HPC file system, serving to accelerate I/O and storage performance of an HPC solution. The overall concept of a Burst Buffer is shown in Figure 7.

To further improve the I/O performance of server attached devices, there are several recent new industry initiatives being formed to define improved interfaces.

One of them is OpenCAPI, an open-standard and high-speed interface for connecting external devices to server systems. According to current

plans, OpenCAPI is targeted to deliver up to 25 gigabits per second per lane, which makes it three times faster compared to PCIe Gen3 and one and a half times faster than PCIe Gen4 (planned to be introduced during 2017). Furthermore, OpenCAPI is designed to allow attached devices to directly access system memory the same way as a respective CPU.

Another initiative in the area of interconnect fabrics is the Gen-Z industry consortium with the following goal: "An open systems Interconnect designed to provide memory semantic access to data and devices via direct-attached, switched or fabric topologies." (genzconsortium.org).

HPC Interconnect Fabric

The hardware "glue" of modern HPC cluster architecture solutions is the Interconnect Fabric which connects all the components together. It is a scalable, end-to-end low latency, high bandwidth and high message rate throughput fabric to allow fast communication and data exchange between all the nodes in a cluster. Originally clusters used Ethernet based interconnects. Current HPC cluster solutions often deploy an InfiniBand based fabric (QDR, FDR or EDR) or a higher performance Ethernet implementation like 10GbE. With the new Omni-Path Architecture Intel has recently introduced an advanced HPC cluster fabric based on the Cray Aries interconnect protocol to overcome some of the inherent limitations of

InfiniBand for performance, reliability and scaling. Today, all high-end fabrics (EDR InfiniBand, Intel Omni-Path and 100Gb Ethernet) are utilizing 100Gbs links and are projected to double the bandwidth with the respective future generations' implementation in the future, e.g. Mellanox HDR InfiniBand at 200 Gbit/s, and larger switch sizes. Such new technologies will likely be utilizing Software Defined Networking (SDN) functionalities, allowing flexible system configurations and fabric topologies like hypercube, dragonfly, and multi-dimensional tori. Additional features of a modern HPC fabric include (hardware) RDMA, MPI tag-matching, off-load and virtualization support functionalities. The OpenFabrics Alliance (OFA), an open source-based organization, provides the standard fabric APIs for HPC fabrics.

For high-performance and mid- to long-range communication and I/O in the data center, Silicon Photonics (SiPh) technology based solutions are about to become available in the market. SiPh will be extremely beneficial for architecting very large HPC clusters deploying 10s-100s thousands processors with a fast optical interconnect fabric for low end-to-end latency and high throughput and bandwidth. Silicon Photonics is combining two very different technologies - electronics and photonics - in the same silicon chip and has been a high hurdle to overcome in the past. Optical (light based) interconnects will address the bandwidth bottleneck inherent in the (copper) wires and can take full advantage of the improvements in processor speed and capabilities. In addition, SiPh is also very energy efficient. Thus, the combination of Silicon Photonics with new optical cables (e.g. Corning Clearcurve) enable massive data rates of up to 1.6 terabits per second at lengths up to 300 meters. It extends high speed connectivity from 3 meters to 2 kilometers, facilitates higher density arrays, enables integration of communications into silicon and efficiently scales line rate speeds and throughput – all in a single integrated chip (see Figure 8 for illustration).

Today, SiPh transceivers are available as an alternative to the conventional VCSEL technol-

ogy. In the future, SiPh will enable higher levels of functionality to be integrated into the optical interface. The initial bandwidth of SiPh based networking is supposed to be 100Gbit/s with plans to increase it to 400Gbit/s in the future. Companies actively working today on Silicon Photonics technologies for example are Intel and IBM.

Emerging HPC Workloads and Delivery Models

Traditional HPC workloads are evolving and growing, becoming more complex, and requiring more performance. However, HPC workloads are changing as well - not that traditional HPC is disappearing, but there are new workloads being added. One of the fastest emerging areas for HPC is High-Performance Data Analytics (HPDA): Big Data (Analytics) using HPC. This is due to the increasing need to extract intelligence (both scientific and business) and value out of massive data sets based on complex data analytics technologies (neural networks, deep/machine learning, data clustering, predictive modeling, data categorization, etc.) requiring data intensive computing. The projected evolution for data analytics is going to cognitive computing and further on to neuromorphic computing in the future: turning data into actionable insights.

Beyond the traditional applications of HPC, new application areas are emerging based on the

Figure 8. Silicon Photonics prototype (Source: Intel)



Internet of Things (IoT). The IoT is a huge source of data, and extracting value from this data requires HPC-like techniques to capture, store, filter, process and analyze at a scale which is similar to some of the largest HPC science projects such as the Large Hadron Collider (LHC) at CERN and the planned Square Kilometer Array (SKA).

We can already see that HPC and Big Data Analytics are merging together using advanced technologies such as machine learning and deep learning. Because of the sheer amount of data and the compute power required to process the data, one can benefit from HPC technologies (hardware and software) and HPC solutions in the future to handle these increasingly complex data processing environments efficiently. The importance of IoT can be seen in the recent Forbes article (Forbes, 2016) that “IDC estimates the direct Internet of Things (IoT) market will grow to more than \$1.7 trillion by 2020.”

All of the above new areas for HPC will enhance scientific, engineering and business (compute and simulation) capabilities, especially when combined with advanced Artificial Intelligence (AI). However, with this also comes the need for more consideration of ethical constraints when applying emerging Artificial Intelligence technologies. Current AI is used to augment human capabilities, and support people to do things better. All these new technologies should be developed, deployed and adopted in a responsible, ethical and enduring way.

One of the more traditional access mechanism to HPC capabilities is Grid Computing, it combines computer resources from different domains – often geographically dispersed and opportunistically – to spread tasks across multiple computers interconnected by a communication network. This approach is mainly used for the type of “farming” workloads where the distributed tasks are independent of each other, do not communicate tightly together and latency is not important. An example of such a Grid is the Worldwide LHC Computing Grid (WLCG) project, a global collaboration of more than 170 computing centers

world-wide in 42 countries, linking up national and international grid infrastructures.

Delivery mechanisms for HPC are also changing: HPC in the Cloud is an emerging trend. It includes different implementations from private to hybrid to public cloud environments and infrastructures; see Figure 9 for illustration. As with many new technologies and solutions, there are both challenges and opportunities. On the opportunity side, HPC in the (public) Cloud enables more people and companies to benefit from HPC to gain more and better insights and knowledge, as well as to start using HPC in new ways. This is particularly true for public Clouds being used by small to medium sized manufacturers who do not have access to HPC today or have only temporary needs for it. In some sense it actually helps democratizing HPC. HPC in the Cloud can also be used for ad-hoc peak demand “offload” to augment in-house HPC capacity. There are already companies offering specialized HPC in the (public) Cloud solutions like the company Rescale, but also traditional Cloud hosting companies like Amazon Web Services (AWS) and Microsoft Azure provide certain HPC focused Cloud instances. Also in this area cloud brokering services and marketplaces are emerging, for example UberCloud and Cycle Computing. Some of the challenges for public Cloud usage are considerations around the following topics, but they are not limited to HPC: data movement volume, bandwidth and delays; security and compliance concerns; data protection, privacy and security; risks and liability concerns; legal regulations; and certainly service level agreements (SLA), predictability, business continuity and total cost of ownership (TCO) – especially for transferring data in and out of a Cloud environment.

On the other hand, private HPC Cloud infrastructures (often implemented on-premise) allows organizations to transform, complement or extend their HPC environment to a more elastic resource pool. A private HPC Cloud solution allows organizations to keep all their data under their own control and follow any required policies. In the

future these solutions may be implemented on top of a Software Defined Infrastructure (SDI), offering the potential to optimize overall HPC compute infrastructures, reduce TCO, achieve more flexibility, increase productivity, and gain higher operational and business efficiencies.

In this context, virtualization is gaining more momentum for HPC workloads in form of Linux Containers (e.g. Docker), especially targeted for high-performance environments and applications like the Singularity (Lawrence Berkeley National Laboratory, LBL) and Shifter (National Energy Research Scientific Computing Center, NERSC) projects.

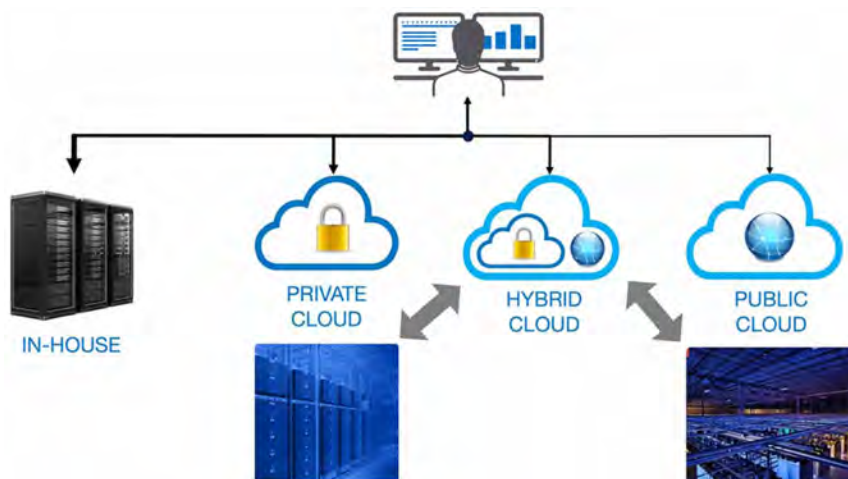
HPC Software

With all the new parallel technologies being implemented in hardware, there is an increasing need for Software Modernization – enabling software to take full advantage of all available parallel performance capabilities in the hardware in a scalable manner. The goal here is to develop and implement software and applications in a way that is portable, re-usable and sustainable in the future, while at the same time deliver high performance and scalability to utilize ongoing improvements in hardware. The ideal result would be that when hardware

advances the software doesn't have to change to take advantage of increased performance. This can be done by using industry standard higher level programming models and languages, software tools and techniques, and optimized libraries for the respective architectures. Examples of this are vectorization (using SIMD), shared memory multi-threading (task parallelism) via OpenMP, heterogeneous parallel acceleration via OpenCL and explicit message passing via MPI or Distributed/Virtual Global Shared Memory interfaces like partitioned global address space (PGAS). In addition, new parallel programming models and languages are being explored on all levels for ease-of-use, scalability and high-performance. A comprehensive overview of modern parallel programming models and languages for parallel computing is provided by Balaji (2015).

Because of the ability of HPC solutions to crunch a lot of data fast and supporting the rapidly growing “Big Data” use cases, there is an increasing need for high-fidelity visualization of very large data sets (often in-situ). For this, open source Software Defined Visualization (SDVis) based frameworks are emerging to allow better and faster high-quality (in 3D) immersive visualization of large and complex data sets not possible with standard graphics hardware (see Figure 10). Be-

Figure 9. Emerging HPC delivery models via Cloud infrastructures



cause such solutions are implemented in software, they can be optimized to fully utilize the compute power and capabilities of modern HPC solutions.

Open Source software is pervasive and drives innovation in the industry faster. At the annual Supercomputing 2016 conference, the HPC community announced the OpenHPC initiative, a collaborative, community effort to provide system software building blocks for HPC systems freely available for open source distribution. This industry activity will help making the deployment, management and operation of (Linux based) HPC solutions quicker, easier and more stable, and improve interoperability.

FUTURE RESEARCH DIRECTIONS

As already mentioned before, there are several ongoing challenges in HPC that need to be solved. They include aspects such as balanced architectures at scale, extreme scalability, high energy efficiency, improved density, higher reliability and resiliency, as well as ease-of-use, sustainable software and scalable parallel programming. Intensive research and development is already ongoing to address all of it – and with challenges there are always opportunities.

The areas listed before are of specific interest when looking at the next frontier in HPC: ExaScale computing. It was envisioned that an EFLOPS (ExaFlops: 10^{18} floating-point operations per second - one billion billion operations

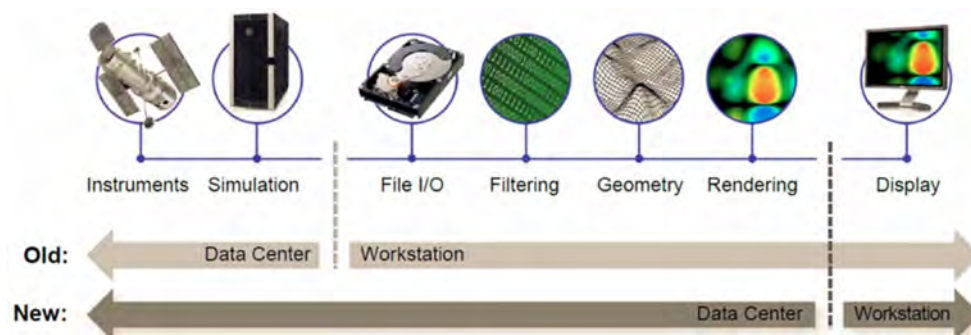
of 64-bit double precision) capable system could potentially be built sometime in the 2020s within a power envelope of about 20MW (Hsu, 2016) but not more than 30MW. Recent evaluations of technologies have pushed this goal out more likely to 2021-2023 for delivery of such an ExaScale system, and first machines might feature more novel system architectures. A real ExaScale system is expected to deliver true capabilities and must be suitable for a wide range of different applications and application areas. For a meaningful ExaScale system one will need to archive fifty times more sustained (production) application performance than is available in 2016/2017 – at about the same power consumption level.

It should be noted that theoretically an EFLOPS system could be assembled and operated today in 2016 – but only at ridiculous costs and effort. We will also see much tighter co-design approaches of hardware and software for future HPC solutions, similar to what is already happening in all parts of the IT industry, because software and applications will run most efficiently on hardware that is tuned for it.

CONCLUSION

High-Performance Computing matters and its future is pretty bright. New and exciting technologies are being developed and implemented to address the challenges ahead for HPC, including the way HPC solutions are being architected and new

Figure 10. Emerging Software Defined Visualization: Workflow moving into the data center



emerging HPC use cases. In general, advances are always driven from an application perspective – and going forward the application span is broadening. HPC’s next frontier, EFLOPS (ExaFlops: 10^{18}) computing in a ~20MW power envelope, is certainly of interest from a pure performance and capability point of view. However, maybe even more important are the technologies that are being developed to enable EFLOPS computing, as they can potentially be applied and scaled down to smaller HPC solutions to allow powerful and energy efficient PFLOPS (PetaFlops: 10^{15}) systems consuming only ~20KW, and eventually TFLOPS (TeraFlops: 10^{12}) implementations within a ~20W power envelope.

1.000.000.000.000.000.000 operations per second - imagine what such HPC based solutions could do to advance mankind!

REFERENCES

Bader, D. A. (2016). Evolving MPI+X Toward Exascale, Computing Edge, IEEE. *Computers & Society*, (September), 5.

Balaji, P. (2015). *Programming Models for Parallel Computing*. Cambridge, MA: MIT Press.

Forbes. (2016). *The Internet Of Things And Machine Learning*. Retrieved from www.forbes.com/sites/moorinsights/2016/03/16/the-internet-of-things-and-machine-learning

Hsu, J. (2016). Three Paths to Exascale Supercomputing. *IEEE Spectrum*, (Jan), 12–13. doi:10.1109/MSPEC.2016.7524158

Moore, S. K. (2016). *Breaking the Multicore Bottleneck*. Retrieved from <http://spectrum.ieee.org/semiconductors/processors/breaking-the-multicore-bottleneck>

Osseyran, A., & Giles, M. (2015). *Industrial Applications of High-Performance Computing: Best Global Practices*. Boca Raton, FL: CRC Press Taylor & Francis Group.

Waldrop, M. (2016). More than Moore. *Nature*, 530(11), 144–147.

Wired. (2016). *Microsoft Bets Its Future on a Reprogrammable Computer Chip*. Retrieved from <https://www.wired.com/2016/09/microsoft-bets-future-chip-reprogram-fly>

Zahran, M. (2017). Heterogeneous Computing: Here to Stay. *Communications of the ACM*, 60(3), 42–45.

ADDITIONAL READING

Berry, M. W., Gallivan, K. A., Gallopoulos, E., Grama, A., Philippe, B., Saad, Y., & Saied, F. (2012). *High-Performance Scientific Computing: Algorithms and Applications* (2012th ed.). Springer. doi:10.1007/978-1-4471-2437-5

Dowd, K., & Severance, C. (1998). *High Performance Computing (RISC Architectures, Optimization & Benchmarks)* (2nd ed.). O’Reilly Media.

Eijkhout, V. (2015). *Introduction to High Performance Scientific Computing* (2nd ed.).

Reinders, J., & Jeffers, J. (2015). *High Performance Parallelism Pearls Volume Two: Multicore and Many-core Programming Approaches* (1st ed.). Morgan Kaufmann, Elsevier.

KEY TERMS AND DEFINITIONS

Cluster: A computer cluster consists of a set of loosely or tightly connected computers (nodes) that work together so that, in many respects, they can be viewed as a single system, controlled by software.

CPU: Central Processing Unit, a general purpose integrated circuit chip, the brains of a computer where most calculations take place.

DDR: Double Data Rate, a type of volatile DRAM memory using both the falling and rising edges of the clock signal.

DDR4: Latest generation of DRAM memory technology.

DRAM: Dynamic Random Access Memory, a memory chip that depends upon an applied voltage to keep the stored data.

DSP: Digital Signal Processing, refers to various techniques for improving the accuracy and reliability of digital communications, also used for acceleration of regular structured simple parallel computational tasks.

FPGA: Field-Programmable Gate Array, an integrated circuit designed to be configured and re-programmed by a customer or a designer after manufacturing.

FLOPS: Floating-Point Operations per Second, a performance unit for computer processing capabilities.

GFLOPS: Gigaflops, 10^9 floating-point operations per second, a compute performance unit.

GPU: Graphics Processing Unit, a single-chip special purpose processor primarily used to manage and boost the performance of video and graphics, also used for acceleration of regular structured simple parallel computational tasks.

HDD: Hard Disk Drive, a data storage device used for storing and retrieving digital information using rapidly rotating disks (platters) coated with magnetic material.

HPC: High-Performance Computing, solving the world's hardest computational problems.

ILP: Instruction Level Parallelism, a mechanism to execute several processor instructions concurrently for higher instruction throughput and higher performance.

IOPS: Input/Output Operations per second, a data I/O performance unit.

IoT: Internet of Things, a network of physical objects that feature an IP address for internet connectivity, and the communication that occurs between these objects and other Internet-enabled devices and systems.

KW: Kilo Watt, one thousand watt unit of power.

MPI: Message Passing Interface, a library specification for message-passing between cluster nodes to program parallel applications.

MW: Mega Watt, one million watt unit of power.

NAND: Negative-AND logic gates based flash memory type of non-volatile storage technology that does not require power to retain data.

NVM: Non-Volatile Memory, a type of persistent computer memory that can retrieve stored information even after having been power cycled (turned off and back on).

OpenCL: Open Computing Language, a programming framework for writing software that execute across heterogeneous computing platforms.

OpenMP: Open Multi-Processing, an application programming interface that supports multi-platform shared memory multiprocessing programming used to program parallel multi-threaded software.

PFLOPS: Petaflops, 10^{15} floating-point operations per second, a compute performance unit.

PCI: Peripheral Component Interconnect, a standard for connecting computers and their peripherals.

SDI: Software Defined Infrastructure, a computing infrastructure entirely under the control of software with no operator or human intervention.

SIMD: Single Instruction Multiple Data, a technology to perform the same operation on multiple data points simultaneously to exploit data level parallelism.

SSD: Solid-State Disk, a data storage device containing non-volatile memory technology.

TFLOPS: Teraflops, 10^{12} floating-point operations per second, a compute performance unit.

High-Performance Reconfigurable Computing

Mário Pereira Vestias

Instituto Politécnico de Lisboa, Portugal

INTRODUCTION

High-Performance Reconfigurable Computing (HPRC) systems integrate reconfigurable technology in the computing architecture to improve performance. System designers and engineers of HPRC systems are aware of the potential speed up that can be achieved by integrating Field Programmable Gate Arrays (FPGAs) or other reconfigurable devices in their multiprocessing systems. Besides performance, reconfigurable hardware devices also achieve lower power consumption compared to General-Purpose Processors (GPP), a major advantage given the high power consumption of existing HPC machines. Better performance and lower power consumption could be achieved using Application Specific Integrated Circuit (ASIC) technology. However, ASICs are not reconfigurable, turning them application specific. Reconfigurable logic becomes a major advantage when hardware flexibility permits to speed up whatever the application with the same hardware module. The first and most common devices utilized for reconfigurable computing are fine-grained FPGAs with a large hardware flexibility. To reduce the performance and area overhead associated with the reconfigurability, coarse-grained reconfigurable solutions have been proposed as a way to achieve better performance and lower power consumption. In this chapter we will provide a description of reconfigurable hardware for high performance computing.

BACKGROUND

High-Performance Reconfigurable Computing (HPRC) is a computing paradigm that combines

reconfigurable-based processing (e.g., FPGA technology) with general purpose computing systems, whether single general purpose processors or parallel processors. The idea is to introduce hardware flexibility to accelerate computationally intensive tasks and therefore to achieve higher performance computing compared to platforms without reconfigurable hardware. These systems not only potentially improve the performance relative to non-reconfigurable general-purpose computing systems but also reduce the energy consumption. Saves of up to four orders of magnitude in both metrics are reported for some compute intensive applications (El-Ghazawi, 2008). Almost all HPC vendors provide HPRC solutions materializing their beliefs in the capacities of reconfigurable computing as accelerators for high-performance computing.

The first commercial reconfigurable computing platform for high-performance computing was the Algotronix CHS2x4 (Algotronix) consisting of an array of 1024 processors and 8 FPGAs with 1024 programmable cells each. This architecture was followed by many other reconfigurable proposals for HPRC.

A major concern in the design of HPRC systems is how reconfigurable computing is connected to the non-reconfigurable computing side, whether general-purpose computing or dedicated computing (e.g. general purpose processors or ASICs). A few alternatives exist with different expected performances, cost and flexibility. The typical approach is to consider reconfigurable systems, generally FPGAs, mounted in a board that is connected to the main system using some serial bus to operate as a coprocessor. The approach has a

relative low cost but has a severe limitation from the serial communication between the host and the coprocessors whose bandwidth determines that for the architecture to be computationally efficient the computation to I/O ratio must be high. To improve the co-processing solution, a few architectures have implemented direct point-to-point connections among the co-processors. This speeds-up the communications between reconfigurable co-processors but keeps the communication bottleneck between the host system and the co-processing system. Some works have refined the communication between the host and the FPGA co-processors through a dedicated network interface. A well-known example of this architecture is Cray XD1 (Cray Inc., 2006). To speed-up even more the communication between the host CPU and the reconfigurable units, all units can be connected to a single communication network, like a shared memory system. In this case, all processing units see each other as part of a unique architecture with access to shared memories (SGI RASC RC100 blade from Silicon Graphics).

Another dimension in the architectural design of HPRC systems is the granularity of the reconfigurable platform. The granularity of a reconfigurable system is the smallest reconfigurable functional unit. In terms of granularity they tend to be defined as fine-grained or coarse-grained. Raw FPGAs are fine-grained since they can be

reconfigured at the bit level. Coarse-grain architectures consist of arrays of coarser units, like arithmetic logic units that are reconfigurable at the word level. Coarse-grained architectures are more amenable to design and reconfigure but are less flexible than fine-grained architectures. When the application to run in the reconfigurable architecture can be bit-level optimized then fine-grained architectures are usually more adequate than coarse-grained architectures achieving faster solutions. On the other side, coarse-grained architectures are faster and more efficient when running word-level applications.

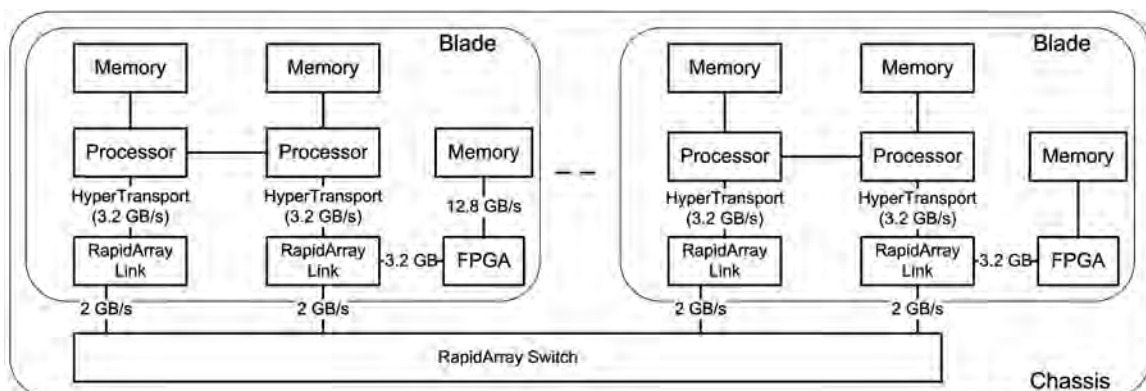
Formally, we define fine-grained HPRC as those platforms that use fine-grained reconfigurable logic and coarse-grained HPRC as those systems that try to improve the performance and power consumption of reconfigurable systems by increasing the granularity at which operations are computed.

HIGH-PERFORMANCE RECONFIGURABLE COMPUTERS

In this section HPRC, examples of fine and coarse-grained reconfigurable computing architectures are described in the context of high performance computing.

Fine-grained high-performance reconfigurable computers consist of basically one or more FPGAs

Figure 1. Cray XD1 chassis



for hardware acceleration combined with general-purpose systems. Well-known fine-grained HPRC includes Cray XD1 (Cray Inc., 2006), SRC MAPstation (SRC Computers), Starbridge Hypercomputer (Starbridge Hypercomputers), SGI RASC RC100 blade (Silicon Graphics), Novo-G (George, Lam, & Stitt, 2011), Maxwell (Baxter, 2007), Convey HC-1 (Brewer, 2010), (Bakos, 2010) and Convey HC-2 (Convey Computer, 2011).

Cray XD1 consists of multiple general-purpose processors combined with multiple FPGAs in a set of chassis. Each Cray XD1 chassis consist of up to six compute blades and a blade has two AMD Opteron 200 processors and one Xilinx Virtex-II Pro FPGA. Therefore, a single chassis consists of twelve processors with up to 8 GBytes of memory for each processor and six FPGAs (see figure 1). Each processor has 1 or 2 RapidArray links connected to a non-blocking RapidArray fabric switch. The switch provides up to 96 GB/s of total bandwidth.

The application must be manually partitioned between hardware running on the FPGA and software running on the Opteron processors.

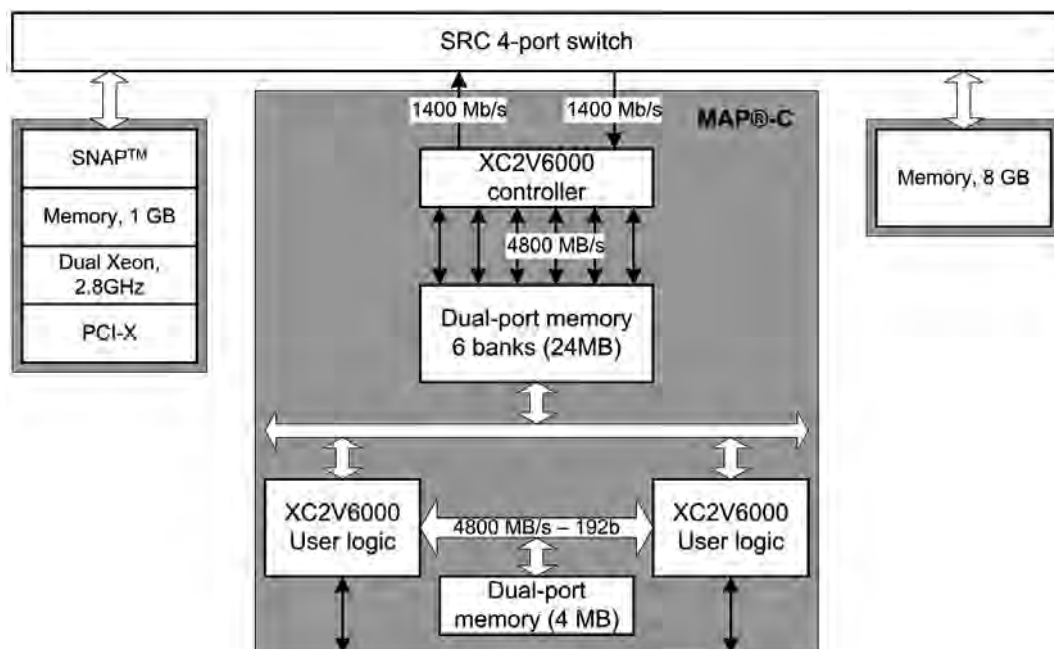
The SRC MAPstation has a board with a dual Xeon processor, a MAP®-C processor and a shared memory of 8 GByte connected with a SRC 4-port switch with a transmission capacity of 1.4 Gbyte/s (see figure 2).

The reconfigurable block, the MAP®-C module, contains one FPGA that controls the access to the dual-port memory with six banks and two Virtex-2 6000 FPGAs for user logic configuration. The control FPGA provides an interface between the SRC 4-port switch and the MAP-C module. The dual-port memory has six banks each with 4 Mbytes 64 bits wide. All six banks are available to the user FPGAs but not at the same time. Transfers between both user logic FPGAs can be operated over one dual-port memory with 4 Mbytes.

The Starbridge Hypercomputer consists of a PC that controls a set of independent processing modules consisting of a Virtex-2 6000 FPGA and 2 Gbytes of local memory. The communication between the PC and the FPGAs (hypercomputer) is through a 64 bits PCI bus running at 66/100 MHz.

The SGI RASC RC100 blade consists of two Virtex-4 LX 200 FPGAs, each with 40 Mbytes

Figure 2. SRC MAPstation



of SRAM. The blade connects to the SGI ALTix 4700 host system through a NUMalink interconnect capable to achieve 6.4 Gbytes per second of bandwidth.

Maxwell is a general-purpose FPGA super-computer. It comprises five IBM BladeCentre chassis. Four of the BladeCentres have seven IBM Intel Xeon blades and the fifth has four. Each blade has a 2.8 GHz Intel Xeon with 1 GByte of main memory and two FPGAs connected through a PCI-X. Each FPGA has access to 512 or 1024 Mbytes of external memory and four Rocket IO connectors delivering each 3.125 Gbps. All 64 FPGAs are connected with a 2D torus interconnection network.

Novo-G is a supercomputer architecture consisting of 24 computing nodes. A node has a Linux server with an Intel quad-core Xeon processor. These nodes communicate via Gigabit Ethernet and 20 Gbps InfiniBand. Each node has also two PROCStar-III boards each containing four Startix-III E260 FPGAs from Altera. This gives a total of 192 FPGAs. Each FPGA has 4.25 Gbytes of dedicated memory and adjacent FPGAs can communicate via a bidirectional bus at rates of 25.6 Gbps.

Convey HC-1 consists of a 2.13 GHz dual-core Intel Xeon with a memory capacity of 128 GBytes and a 1066 MHz front side bus to connect to a coprocessor board. The coprocessor board has four Virtex-5 LX330 FPGAs with 64 GBytes of memory. These four FPGAs connect to eight memory controllers through a full crossbar. The FPGAs are also connected in a ring configuration with a bandwidth of 668 Mbps, full-duplex. Recently, Convey has launched the newest HC-2 (and HC-2ex) server architecture. The new architecture uses the recent Xeon X5600 processor and Xeon E5-2600. Versions with 4, 6 and 8-cores were implemented. The co-processor board has kept the Virtex-5 FPGAs of HC-1 but upgraded to Virtex-6 FPGAs for the more performant HC-2ex architecture. The memory capacity of the main board was upgraded to 192 GBytes of memory. Both HC-1 and HC-2 are not machines for floating-

point executions. The FPGA coprocessors are designed to accelerate data-intensive applications like genome sequence alignment.

Fine-grained high performance reconfigurable computing architectures all use FPGAs to implement reconfigurable logic. They basically differ in how the FPGAs are connected to the general-purpose processors which determines the bandwidth available for the communication between the general-purpose system and the reconfigurable system. Recently, System-on-Chip FPGAs have been introduced (ZYNQ, 2011), (Altera SoC FPGAs, 2012) which integrates an ARM processor with reconfigurable logic connected through high-speed ports. This is a tight integrated of hardware and reconfigurable software that permits the design of very efficient hardware/software solutions for high-performance computing and high-performance embedded computing. Recently (2015), Intel Corp. agreed to buy Altera Corp. to provide heterogeneous solutions based on general-purpose CPUs and reconfigurable logic FPGAs.

FPGAs work at bit-level granularity. Therefore, to support multi-bit operations many resources are needed degrading the area efficiency of a computing system based on FPGAs. Also, there will be a high routing overhead which will considerably reduce the performance of the implemented operations. FPGAs also require large amounts of configuration data increasing very much the configuration time and the power dissipation. To overcome such disadvantages, coarse-grained reconfigurable architectures consider atomic units in the form of 4, 8, 16 or even higher bits wide arithmetic oriented processing units. This permits to achieve greater efficiency executing word-level operations improving the area, the configuration time, the performance and the power consumption compared to fine-grained reconfigurable architectures.

Several coarse-grained reconfigurable architectures (CGRA) were proposed during the last decade. DP-FPGA (Cherepacha & Lewis, 1996) is a coarse-grained architecture oriented to the design of datapaths. It consists of four-bit datapath

width units with a structure similar to that of an FPGA. KressArray-1 (Kress, 1996) is an array of reconfigurable datapath units working in a data-driven fashion consisting of a generalization of the systolic array. Rapid (Ebeling et al., 1996) is a linear array of datapath units designed to work mainly in a computational pipeline optimized to repetitive computation intensive tasks. The units consist of an integer multiplier, three ALUs, general-purpose registers and local memories. MATRIX (Mirsky & DeHon, 1996) is an array of identical 8-bit elements containing each 256 bytes, an 8-bit ALU, a multiply unit and control logic distributed in a 2D-mesh architecture. Raw Machine (Waingold et al., 1997), an array of tiles consisting of instruction and data memory, an ALU, registers, configurable logic and a programmable switch. KressArray-3 (Hartenstein et al., 1999) a mesh of reconfigurable datapath units (rDPU) connected to the four nearest neighbors with bidirectional links with a datapath width of 32 bits. Each PE can be configured to run integer operators. MorphoSys (Lu, 1999) is a more elaborate architecture with a core processor, a context memory, a frame buffer, a DMA controller and an array of 8×8 reconfigurable cells. Each reconfigurable cell contains an ALU with multiplier, a shifter, and a register file. The ALU can perform a multiply-accumulate operation. PipeRench (Schmit, 2002), consists of a linear array of processing elements for stream computing in a pipelined fashion. Each processing element contains a bank of registers and an ALU implementing bitwise functions, addition and subtraction with eight bits wide.

ADRES (Meiet et al., 2003) is an heterogeneous architecture with a tightly coupled VLIW (Very-Long Instruction Word) processor and a coarse-grained reconfigurable matrix. Each reconfigurable cell consists basically of a register file and a functional unit.

MORA (Lanuzza et al., 2007) consists of a 2D array of configurable cells organized in a 4×4 quadrants connected with a reconfigurable network. The reconfigurable cell contains a dual-port 256 bytes SRAM, an 8-bit ALU and a control

unit. The ALU contains a set of registers, two 8×4 multipliers, and a 16-bits adder. WPPA (Kissler et al., 2006) (Weakly Programmable Processor Arrays) – consists of an array of weakly programmable processing elements (WPPE) with limited instruction memory and optimized control. The instruction set is parameterizable at compilation time. Each WPPE can be parameterized to contain several functional units like adders, multipliers, shifters, and other modules for logical operations. It is also possible to include specific functional units implementing user-defined functions (e.g., a Fast Fourier Transformation). FPOA (Field Programmable Object Array) is a reconfigurable array proposed by MathStar (www.mathstar.com) The array consists of small ALUs organized in a regular structure similar to an FPGA. The device also contains memory and MAC blocks for signal processing The XPP architecture (PACT-XPP) is another coarse-grained architecture containing a 2D array of programmable elements, capable to perform typical arithmetic operations, and a number of internal RAMs. BiRC (Bilkent Reconfigurable Computer) (Atak & Atalar, 2012) is a statically reconfigurable architecture consisting of three types of processing elements, namely an arithmetic logic unit, a memory and a multiplier, displaced in columns and connected to four neighbors.

The above coarse-grain architectures improve the performance of integer operations, but have little advantage over fine-grain reconfigurable logic when floating-point operations are required. Recent works have proposed coarse-grain reconfigurable architectures with dedicated support for floating-point operations. For example, FloRA (Lee et al., 2009) is a coarse-grain reconfigurable architecture that performs both integer and floating-point operations. The system contains a RISC processor, a DMA controller, an external memory interface and the reconfigurable computing module. The RISC processor controls all the components in the platform and is also used to execute control-intensive pieces of software. The reconfigurable module contains an array of

processing elements (PE), an execution control unit to monitor the status of the reconfigurable module, a configuration cache and a frame buffer. The PEs are arranged in an array of 8×8 connected to neighbors and contain an integer ALU and 16-bit registers. PEs in the same row share a multiplier, a divider and a square root logic. Floating-point operations are implemented by grouping two PEs. One PE calculates the mantissa and the other calculates the exponent. However, the architecture only implements a custom floating-point format with 8 bits for the exponent and 15 bits for the mantissa.

Another approach to support floating-point reconfigurable computing was proposed in (Ho et al., 2009). The floating-point FPGA (FPFPGA) is a fine-grained FPGA with dedicated columns with reconfigurable coarse-grained units containing fixed-function floating-point adders and multipliers and general-purpose bit blocks all connected with unidirectional buses. The floating-point units support both 32 and 64-bits floating-point format.

The new generation 10 FPGA from Altera (2015) provides hardened floating-point blocks. These new devices integrate coarse-grain arithmetic blocks with fine-grained logic in a single device providing up to 10 TFLOPs of peak performance in a single device that can be configured at fine granularity.

More recently, some new directions for coarse grain reconfigurable arrays were proposed. EGRA (Expression-Grain Reconfigurable Architecture) (Ansaloni, Bonzini, & Pozzi, 2011) consists of an array of cells where each cell is a group of arithmetic logic units. They call it coarse grain cell RAC (Reconfigurable ALU Cluster). Inside a RAC, ALUs are organized in rows connected by switchboxes. The inputs to RACs come from neighbor RACs, from the RAC itself and from constants. The number and functionality of each ALU can be customized by the designer. The cell configurations are stored in a context memory and can be loaded at runtime. The EGRA array is a mesh of cells that can be RACs, memories or multipliers. As most of the CGRA architectures,

EGRA does not have support for floating-point operations. SYSCORE (Patel, McGettrick, Bleakley, & Chris, 2011) is an CGRA architecture for low power, real time processing of biomedical systems. The architecture proposed consists of an array of Configurable Functional Units (CFU) and RoundAbout Interconnect (RAI) units. Data is injected into the array using two input Direct Memory Access (DMA) units and the output data is driven by an output DMA. The CFU has a computation unit supporting MULT-ADD, MULT-SUB and compare operations. The bit width of all units and registers is 24 bits. CFU are connected to neighbor units. Since the CFU are tailored to specific target applications, good performance and energy consumptions were achieved. In (Silva, Fernandes, Véstias, & Neto, 2012) a new architecture for a coarse-grained reconfigurable array targeted to linear algebra problems. The reconfiguration memories are implemented using magnetic tunneling junctions. These storage elements provide for non-volatility and for a very effective implementation of multi-context planes. The proposed architecture is organized as a 2-dimensional mesh of double precision floating-point execution units. The execution units are run-time reconfigurable. Their configurations define the operation to be executed and the data flow intra and inter execution units. The array is organized as a 2D mesh of reconfigurable execution units (REU). The arithmetic unit of the REU has a fused multiply-accumulate operator and a reciprocal/division operator. This Configurable Floating-point Arithmetic Unit (CFAU) currently implements the following 64-bit floating-point operations: addition, subtraction, multiplication, reciprocal, division, fused multiply-add and fused multiply-accumulate. As referred, the reconfigurable memory is implemented with magnetic memory cells and currently supports two planes of context, which can be exchanged within a clock cycle.

All these works provide reconfigurable platforms to execute word-level operations faster than when executed in fine-grained reconfigurable FPGA platforms. All solutions try to innovate the

design of reconfigurable coarse-grain solutions. However, most of them have failed to compete with fine-grained FPGAs (Coudert, 2009). The failure is mainly due to commercial reasons. Innovative start-ups succeeded in a very specific niche of the market but soon are acquired by a stronger company to complement their portfolio of products. Also, these new technologies require some particular way of programming without any standard behind it, making it very hard to be adopted in a standard based industry. All these problems are serious obstacles to find a well-established HPRC system based on reconfigurable coarse-grained architectures.

FUTURE RESEARCH DIRECTIONS

High-performance reconfigurable computing platforms have already shown their potential to speed-up specific computation intensive applications with orders of magnitude improvement in performance, size and cost over traditional HPC systems. HPRC platforms are based on fine-grained reconfigurability which for word-level applications are less efficient. Most works on coarse-grained reconfigurability improve the performance, area and power consumption compared to fine-grained reconfigurability by considering word-level reconfigurable operations. However, coarser-grained architectures are less flexible and become more application oriented. We are not expecting to design a reconfigurable platform very efficient to all type of applications. But even considering specific application domains, finding the most efficient reconfigurable architecture is still an open research problem.

To the present, reconfigurable computing for high performance is generically based on fine-grained reconfigurable devices, because these are available in the form of FPGAs, while coarse-grained architectures have had a very difficult way to find a place in the market.

Whether fine or coarse-grained, there are many challenges to guarantee the widespread adoption

of reconfigurable computing systems in HPC. A very first requirement of HPC is floating-point processing. New HPRC must support the efficient execution of floating-point operations. Up to date, state-of-the-art FPGAs have the potential to implement hundreds of double precision floating-point units, but with low efficiency compared to hardwired solutions. The recent FPGAs from Altera with hardwired floating-point units have brought FPGAs again to the high-performance computing scenario. Some coarse-grained architectures already target floating-point operations. Should FPGAs keep the integer aspects and leave floating-point issues for coarse-grained platforms is an open issue. FPGAs allow also the utilization of dynamic/partial reconfiguration permitting the utilization of virtual hardware, that is, hardware that is configured at runtime due to a lack of space for hardware implementation of the whole hardware architecture.

Other major challenges are under intensive research, including the programmability of the platform and the memory bandwidth improvement. Programmability is a very important aspect since users already have a lot of code and do not want to apply great changes to the code in order to take advantage of the reconfigurable platform. How can the code be ported to the new platform without major investments to rewrite the code? This is a major requirement that the providers of the reconfigurable platforms must answer. Good programming environments already exist, but are very architecture specific. It is important to address portability and maintainability to guarantee the investment on code development. Memory bandwidth is also an important aspect of parallel computing machines and is frequently the main bottleneck of parallel architectures. Adding more processing power into a platform must be followed by more memory bandwidth. Memory architectures must be improved as well as memory bandwidth. Bigger reconfigurable devices allow the use of more memory banks to access external memory and also come with more internal memory that can be used to cache

data inside the reconfigurable device. More has to be done, not only by memory manufacturers, but also by manufacturers of reconfigure devices by, for example, integrating dedicated memory controllers.

FPGA devices are the leading technology for designing HPRC systems. Recent FPGAs manufactured with the top integrated circuit technologies, Ultrascale+ from Xilinx (Virtex) and Stratix 10 from Altera (Altera, 2015), have capacity to achieve TFLOPs (Tera Floating-Point Operations) with access to fast memory interfaces. New SoC FPGAs from Xilinx and Altera already contains a state-of-the-art processor that can be tightly-coupled to reconfigurable logic. Reconfigurable logic in future devices may have a variety of architectures, granularities, etc. It is imperative to keep the research on reconfigurable architectures to look for the most appropriate granularities for HPRC.

New reconfigurable architectures, whether fine or coarse, have to be explored. Fine-grained reconfigurable computing has been preferred mainly due to its availability but there is no evidence that these will be the reconfigurable platforms in the future. Heterogeneous reconfigurable devices with a mix of fine and grain logic should also be explored, like the recent Altera 10 FPGAs. Another research direction is the tight integration of multi-core CPU and FPGA. These heterogeneous architectures integrate the advantages of both devices into a single platform.

CONCLUSION

The article describes the concept of high-performance reconfigurable computing and presents several high-performance reconfigurable computers. Commercial HPRC platforms are based on fine-grained reconfigurable devices (FPGAs). However, coarse-grained reconfigurable computing architectures have shown better efficiency in terms of area, performance, and power consumption. Therefore, it would be expected to see HPRC

platforms based on coarse-grained reconfigurable architectures. However, coarse-grained architectures have failed to compete in the market with fine-grained architectures for several reasons, but mostly due to commercial reasons. So, there have been huge difficulties to have commercial HPRC systems based on coarse-grain reconfigurable computing.

Many application domains will greatly benefit from coarse-grained architectures. For example, many scientific computing applications are very computationally intensive requiring floating-point operations. A coarse-grained architecture with support for floating-point operations would achieve orders of magnitude better performance compared to a fine-grained architecture. Hence, it is expected that coarse-grained architectures become dominant in the HPRC area. Recent FPGAs achieve Teraflops of peak performance due not only to improvements in integrated circuit technology, but also the integration of coarse-grain arithmetic blocks. This may be the way of FPGAs in high-performance computing, that is, devices with both fine and coarse grain floating-point units whose computing datapath can be optimized for the target application. So, the research about high-performance reconfigurable computing architectures is an actual topic of research.

REFERENCES

- Algotronix. (n.d.). Retrieved from http://www.algotronix.com/index_flash.html
- Altera. (2015). Retrieved from <https://www.altera.com/products/fpga/stratix-series/stratix-10/overview.html>
- Altera SoC FPGAs. (2012). Retrieved from <https://www.altera.com/products/soc/overview.html>
- Ansaloni, G., Bonzini, P., & Pozzi, L. (2011). EGRA: A Coarse Grained Reconfigurable Architectural Template. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 19(6), 1062–1074.

- Atak, O., & Atalar, A. (2012). BilRC: An Execution Triggered Coarse Grained Reconfigurable Architecture. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 1-10.
- Bakos, J. (2010). High-Performance Heterogeneous Computing with the Convey HC-1. *Computing in Science & Engineering*, 12(6), 80–87. doi:10.1109/MCSE.2010.135
- Baxter, R., Booth, S., Bull, M., Cawood, G., Perry, J., Parsons, M., & Genest, G. et al. (2007). Maxwell - a 64 FPGA Supercomputer. *Second NASA/ESA Conference on Adaptive Hardware and Systems*, 287-294. doi:10.1109/AHS.2007.71
- Brewer, T. M. (2010). Instruction Set Innovations for the Convey HC-1 Computer. *IEEE Micro*, 30(2), 70–79. doi:10.1109/MM.2010.36
- Cherepacha, D., & Lewis, D. (1996). DP-FPGA: An FPGA Architecture Optimized for Datapaths, *Proceedings of VLSI Design*. doi:10.1155/1996/95942
- Convey Computer. (2011). *The Convey HC-2™ Computer – Architectural Overview*. Convey White Paper.
- Coudert, O. (2009). *Why FPGA startups keep failing*. Retrieved from www.ocoudert.com/blog/2009/09/15/why-fpga-startups-keep-failing
- Cray Inc. (2006). *Cray XD1 Supercomputer*. Retrieved from <http://www.cray.com/products/xd1>
- Ebeling, C., Cronquist, D., & Franklin, P. (1996). RaPiD - Reconfigurable Pipelined Datapath. *Proceedings of International Workshop on Field Programmable Logic and Applications*.
- El-Ghazawi, T., El-Araby, E., Huang, M., Gaj, K., Kindratenko, V., & Buell, D. (2008). The promise of high-performance reconfigurable computing. *IEEE Computer*, 41(2), 69–76. doi:10.1109/MC.2008.65
- FHPCA. (n.d.). Retrieved from www.fhpca.org
- George, A., Lam, H., & Stitt, G. (2011). Novo-G: At the Forefront of Scalable Reconfigurable Supercomputing. *Computing in Science & Engineering*, 13(1), 82–86. doi:10.1109/MCSE.2011.11
- Hartenstein, R., Herz, M., Hoffmann, T., & Nageldinger, U. (1999). Mapping Applications onto reconfigurable KressArrays. *Proceedings of International Workshop on Field Programmable Logic and Applications*, 385-390. doi:10.1007/978-3-540-48302-1_42
- Ho, C., Yu, C., Leong, P., Luk, W., & Wilton, S. (2009). Floating-Point FPGA. Architecture and Modeling. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 17(12), 1709–1718.
- Kissler, D., Hannig, F., Kupriyanov, A., & Teich, J. (2006). A Highly Parameterizable Parallel Processor Array Architecture, *Proceedings of IEEE International Conference on Field Programmable Technology*, 105–112. doi:10.1109/FPT.2006.270293
- Kress, R. (1996). *A Fast Reconfigurable ALU for Xputers* (Doctoral Thesis). Department of Computer Science, University of Kaiserslautern, Germany.
- Lanuzza, M., Perri, S., & Corsonello, P. (2007). MORA: A New Coarse-Grain Reconfigurable Array for High Throughput Multimedia Processing, *Proceedings of SAMOS*, 159–168. doi:10.1007/978-3-540-73625-7_18
- Lee, D., Jo, M., Han, K., & Choi, K. (2009). FloRA: Coarse-Grained Reconfigurable Architecture with Floating-Point Operation Capability. *International Conference on Field-Programmable Technology*, 376-379. doi:10.1109/FPT.2009.5377609
- Lu, G., Singh, H., Lee, M., Bagherzadeh, N., Kurdahi, F., & Filho, E. M. C. (1999). The MorphoSys Parallel Reconfigurable System. *Proceedings of Euro-Par*, 99I, 727–734. doi:10.1007/3-540-48311-X_102

Mei, B., Vernalde, S., Verkest, D., De Man, H., & Lauwereins, R. (2003). ADRES: An architecture with tightly coupled VLIW processor and coarse-grained reconfigurable matrix. *Field-Programmable Logic and Applications*, 2778, 61–70. doi:10.1007/978-3-540-45234-8_7

Mirsky, E., & DeHon, A. (1996). MATRIX: A Reconfigurable Computing Architecture with Configurable Instruction Distribution and Deployable Resources. *Proceedings of the IEEE Workshop on FPGAs for Custom Computing Machines*, 157–166. doi:10.1109/FPGA.1996.564808

PACT-XPP PACT-XPP. (n.d.). Retrieved from <http://www.pactxpp.com/>

Patel, K., & McGettrick, S. (2011). SYSCORE: A Coarse Grained Reconfigurable Array Architecture for Low Energy Biosignal Processing. *IEEE 19th Annual International Symposium on Field-Programmable Custom Computing Machines*, 109–112.

Schmit, H., Whelihan, D., Tsai, A., Moe, M., Levine, B., & Taylor, R. (2002). PipeRench: A virtualized programmable datapath in 0.18 micron technology. *Proceedings of the IEEE Conference on Custom Integrated Circuits*, 63–66. doi:10.1109/CICC.2002.1012767

Silicon Graphics. (n.d.). *SgiR rasct rc100 blade, dramatic application speed-up with next generation reconfigurable compute technology*. Retrieved from <http://www.sgi.com>

Silva, V., Fernandes, J., Véstias, M., & Neto, H. (2012). A High-Performance Reconfigurable Computing architecture using a magnetic configuration memory. *International Conference on Reconfigurable Computing and FPGAs*, 1–6. doi:10.1109/ReConFig.2012.6416756

SRC Computers Inc. (n.d.). *Starbridge Hypercomputers*. Retrieved from <http://www.starbridgesystems.com/products/hardware.html>

Stratix. (n.d.). Retrieved from <http://www.altera.com/devices/fpga/stratix-fpgas/stratix-v/stxv-index.jsp>

Ultrascale. (n.d.). Retrieved from <http://www.xilinx.com/products/technology/ultrascale.html>

Waingold, E., Taylor, M., Srikrishna, D., Sarkar, V., Lee, W., Lee, V., & Agarwal, A. et al. (1997). Baring It All to Software: Raw Machines. *IEEE Computer*, 30(9), 86–93. doi:10.1109/2.612254

ZYNQ. (2011). Retrieved from <http://www.xilinx.com/content/xilinx/en/products/silicon-devices/soc/zynq-7000.html>

ADDITIONAL READING

Alle, M., Varadarajan, K., Fell, A., Reddy, R., Joseph, N., Das, S., Biswas, P., Chetia, J., Rao, A., Nandy, S., & Narayan, R. (2009). Redefine: Runtime reconfigurable polymorphic ASIC. *ACM Transactions Embedded Computing Systems*, vol. 9, pp. 11:1–11:48.

Baxter, R. (2007). High-performance reconfigurable computing - the view from Edinburgh. *Proceedings of the second NASA/ESA Conference on Adaptive Hardware and Systems*, pp. 273–279. doi:10.1109/AHS.2007.61

Benkrid, K. (2008). High performance reconfigurable computing: From applications to hardware. *International Journal of Computer Science*, 35(1).

Brunelli, C., Garzia, F., & Nurmi, J. (2008). A Coarse-Grain Reconfigurable Architecture for Multimedia Applications Featuring Subword Computation Capabilities. *Journal of Real-Time Image Processing*, Springer-Verlag, 3(1-2), 21–32. doi:10.1007/s11554-008-0071-3

Buell, D., El-Ghazawi, T., Gaj, K., & Kindratenko, V. (2007). High-Performance Reconfigurable Computing. *IEEE Computer*, March, pp. 23–27.

Center for High Performance Reconfigurable Computing - <http://www.chrec.org/>

Chiu, M. & Herbordt, M. (2010). Molecular Dynamics Simulations on High-Performance Reconfigurable Computing Systems. *ACM Transactions on Reconfigurable Technology Systems*, 3(4), n° 23.

Compton, K., & Hauck, S. (2002). Reconfigurable computing: A survey of systems and software. *ACM Computing Surveys*, 34(2), 171–210. doi:10.1145/508352.508353

De Sutter, B., Raghavan, P., & Lambrechts, A. (2010). Coarse-grained reconfigurable array architectures. In S. Bhattacharyya, E. Deprettere, R. Leupers, & J. Takala (Eds.), *Handbook of Signal Processing Systems* (pp. 449–484). Springer-Verlag. doi:10.1007/978-1-4419-6345-1_17

El-Araby, E., Gonzalez, I., & El-Ghazawi, T. (2009). Exploiting Partial Runtime Reconfiguration for High-Performance Reconfigurable Computing. *ACM Transactions on Reconfigurable Technology and Systems*, Vol. 1 Issue 4.

El-Ghazawi, T., El-Araby, E., Huang, N., Gaj, K., Kindratenko, V., & Buell, D. (2008). The promise of high-performance reconfigurable computing. *IEEE Computer*, 41(2), 69–76. doi:10.1109/MC.2008.65

Hartenstein, R. (2001). A decade of reconfigurable computing: a visionary retrospective. *Proceedings Conference on Design, Automation and Test in Europe*, pp. 642–649. doi:10.1109/DATE.2001.915091

Iqbal, M., Azam, F., Awan, U., & Hammad, S. (2011). Performance enhancement techniques for modern reconfigurable computing systems. *International Journal of Computers and Applications*, 27(09), 33–38. doi:10.5120/3327-4577

Iqbal, M., & Khan, S. (2012). Exploring High Performance Processing Through Reconfigurable Computing Systems. *Journal of Basic and Applied Scientific Research*, 2(10), 10185–10202.

Iqbal, M., Khannum, A., Iqbal, S., & Asif, M. (2010). Emerging requirements of reconfigurable computing systems for performance enhancement. *International Journal on Computer Science and Engineering*, 2(5), 1572–1579.

Kim, C., Chung, M., Cho, Y., Konijnenburg, M., Ryu, S., & Kim, J. (2012) ULP-SRP: Ultra low power Samsung Reconfigurable Processor for biomedical applications. *International Conference on Field-Programmable Technology (FPT)*, pp.329-334. doi:10.1109/FPT.2012.6412157

Liang, C., & Huang, X. (2008). *Smartcell: A power-efficient reconfigurable architecture for data streaming applications* (pp. 257–262). Workshop Signal Processing Systems.

Liang, S., Yin, S., Liu, L., Guo, Y., Wei, S. (2015). A Coarse-Grained Reconfigurable Architecture for Compute-Intensive Mapreduce Acceleration. *Computer Architecture Letters*, vol.PP, no.99, pp.1-1.

NSF center for High-performance Reconfigurable Computing.

Rakossy, Z., Naphade, T., & Chattopadhyay, A. (2012). Design and analysis of layered coarse-grained reconfigurable architecture. *International Conference on Reconfigurable Computing and FPGAs*, pp.1-6. doi:10.1109/ReConFig.2012.6416736

Shami, M., & Hemani, A. (2012). Classification of massively parallel computer architectures. *International Symposium on Parallel and Distributed Processing Workshops and Phd Forum*, pp. 341-351.

Smith, M., & Peterson, G. (2012). Optimization of Shared High-Performance Reconfigurable Computing Resources. *ACM Transactions on Embedded Computing Systems*, 11(2), 1–22. doi:10.1145/2220336.2220348

Todman, T., Constantinides, G., Wilton, S., Mencer, O., Luk, W., & Cheung, P. (2005). Reconfigurable computing: Architectures and design methods. *Proceedings of IEE Computers and Digital Techniques*, 152(02), 193–207. doi:10.1049/ip-cdt:20045086

Vanderbauwhede, W., & Benkrid, K. (Eds.). (2013). *High-Performance Computing Using FP-GAs*. Springer. doi:10.1007/978-1-4614-1791-0

Vassiliadis, S., & Soudris, D. (Eds.). (2007). *Fine- and Coarse-Grain Reconfigurable Computing*. Springer. doi:10.1007/978-1-4020-6505-7

Voros, N., Nikolaos, R., Rosti, A., & Hübner, M. (Eds.). (2009). *Dynamic System Reconfiguration in Heterogeneous Platforms - The MORPHEUS Approach*. Springer Verlag. doi:10.1007/978-90-481-2427-5

Williams, J., Massie, C., George, A., Richardson, J., Gosrani, K., & Lam, H. (2011). Characterization of Fixed and Reconfigurable Multi-Core Devices for Application Acceleration. *ACM Transactions on Reconfigurable Technology and Systems*, 3(4), n° 19.

Zhuo, L. (2007). *High-performance Linear Algebra on Reconfigurable Computing Systems*. ProQuest.

KEY TERMS AND DEFINITIONS

Coarse-Grained Granularity: Refers to the granular size of reconfigurable architectures consisting of arrays of units reconfigurable at the word level. A typical example is an arithmetic unit.

Fine-Grained Granularity: Refers to the granular size of the reconfigurable hardware at the bit level. Fine-grained granularity is typically associated to field-programmable gate arrays.

Granularity: Refers to the relative size of the reconfigurable elements in a reconfigurable hardware architecture.

High-Performance Computing (HPC): Refers to the use of parallel processing techniques to execute computationally demanding algorithms.

High-Performance Reconfigurable Computing (HPRC): A computing paradigm that combines reconfigurable-based processing (e.g., FPGA technology) with general purpose computing systems, whether single general purpose processors or parallel processors.

Reconfigurable Computing (RC): A computing paradigm that uses reconfigurable hardware for computing purposes.

Reconfigurable Hardware: A hardware structure whose logic elements and their interconnections can be reconfigured to implement a particular logic circuit.

Category H

Hospitality, Travel, and Tourism Management

Augmented Reality for Tourist Destination Image Formation



Azizul Hassan

Cardiff Metropolitan University, UK

INTRODUCTION

Destination image in recent years is getting hugely interacted with innovative technology application (Lakshmi, & Ganesan, 2010). Tourism is granted as having attachments with destination image when, destination image beholds a significant position. In the simplest understanding, destination image is referred as the expression of an individual's or a group's all imagination, prejudices and objective knowledge about a particular location. In the given context of tourism, this study outlines the roles of AR in destination image formation when, the technology is seen as innovative. This is true that AR has turned into a buzzing word in terms of its uniqueness. AR has already been applied in many areas of human knowledge and business industries creating significant economic and non-economic benefits for both stakeholders and beneficiaries. Still, the benefit generating capacities of AR has not been fully exploited. When, AR and the term 'Gimmick' are almost intertwined, the application of AR in many industries including tourism can hardly be generalised as gimmick. AR does not only serve promotional or marketing activities but also can possibly expanded to destination image formation. This is particularly evident that AR application plays important role in destination image formation (Dadwal & Hassan, 2015). The recent popularity of AR is largely indebted to the technological advancements as wearable or hand-held computing devices and Smartphone. Also on practical ground, Smartphones have played crucial roles for both introducing and popularising AR (Azim & Hassan, 2013).

BACKGROUND

Tourist Destination Image Formation: The exact meaning of tourist destination image is a bit difficult and knotty. So far, this term has been used in diverse perspectives relating destination image. Such image is normally projected by tourism promoters publicly, a destination's stereotype image or individual led destination images. Echtner and Ritchie (1991) noted that, a major part of the definitions of tourist destination image in earlier studies is obsolete. This means that a comprehensive research to define tourist destination image becomes essential. As defined 'image is one of those terms that won't go away ¼ a term with vague and shifting meanings' (Pearce, 1988: 162). Still, the actual meaning of 'image' centres at the core of such complexities. The term has been applied in as considerable number of knowledge disciplines including psychology, behavioural geography and marketing. In psychology, the notion refers to a type of visual representation. In behavioural geography, this turns as more comprehensive as associated with beliefs, values, knowledge, impressions and emotions. However, in marketing, the term points to the attribution underlying image and attach image to consumer behaviour.

One of the accepted definitions of tourist destination image is offered by Crompton (1979:18), 'the sum of beliefs, ideas and impressions that a person has of a destination'. This definition clearly involves individuals when, many other definitions concedes that images can be shared by groups of peoples. From tourism marketing perspective, this is essential to understand the common image facts

with a particular group's other members. Such considerations support market segmentations by facilitating marketing strategy formulations. Another commonly accepted definition of tourist destination image is offered by Lawson and Baud Bovy (1977) as, such image is the outlining of all objective knowledge, imaginations, prejudice, impressions and emotional thoughts that a group of people or an individual can have about a particular destination.

Augmented Reality (AR) and Destination Image Formation: AR technology certainly supports in destination image formation. The technology is viewed as a reliable source for providing necessary photographic and typographic information about a tourist destination (Hassan & Jung, in press). The technology affects tourists' decision by using virtual spaces. A number of platforms as Layar and few others offer the destination marketing organisations (DMO) to publish contents with geo-referencing, tourist attraction descriptions, accommodations, restaurants supported by required information and data about that particular destination (Hassan, 2013). However, such AR technology based contents cannot be widely available to adopt the general users.

Augmented reality (AR) operations rely on computer graphics, computing, sensor and wireless technologies (Hassan & Rahimi, in press). Conventional AR devices as head-mounted displays (HMDs) can be of many types depending on their applications and usability. In a modern GPS supported Smartphone, AR applications can also cover outdoor use. This becomes existent when, tourists normally point the device towards physical objects in a real environment. They are then become able to see the added virtual information on the camera view topped by virtual annotations. The amount and type of contents and information can differ ranging from images, texts, videos or symbols of diverse landmark types.

AR offers benefits to consumers and the tourism society by offering valid destination choice for potential tourists (Hassan & Ramkissoon, in press). This technology allows gathering tourism

industry tailored information to empower tourism consumers. They can then become able to meet certain demands highlighting their preferences and needs (Fritz et al., 2005). The potentials of AR application require proper attention from all concerned parties when, the market potentials become massive. The growing popularity of AR has expanded to move from conventional computing to Smartphone devices. AR as a technology can incorporate with Google AdSense, search engines or even act as a powerful medium of electronic word-of-mouth (Hassan & Donatella, in press; Hassan & Dadwal, in press & Hassan, 2015).

MAIN FOCUS OF THE ARTICLE

Issues, Controversies, Problems

The aim of this study is to explore AR as an innovative technology application for destination image formation. Also, the study presents some global examples of AR application for destination image formation relating them in the Bangladesh context as an example of emerging tourism economy. The two separate study contexts are the United Kingdom tourism market as pioneering AR application while, the Bangladesh scenario symbolizes a tourism market that readily available to accept an experimental technology.

The sources of travel information as used by the domestic tourists were investigated by Nolan (1976) in this country. The recommendation or advice of friends and family members were found as the most frequent as used as travel information sources. These sources were followed by commercial tourist information and guidebooks and promotional publications. Considering the credibility of the sources of travel information, guidebooks were the highest rated. The government services by the State and suggestions of friends and family members were considered as the most informative. Nolan (1976) also determined the travel information sources' 'objectivity' by inquiring respondents for rating their biasness.

Outcomes as inferred from the results recognised an overall bias in travel brochures as the travel information communication. In the most recent times, communication or recommendations are commonly based on technology when, AR as an innovative technology can have obvious validity.

Solutions and Recommendations

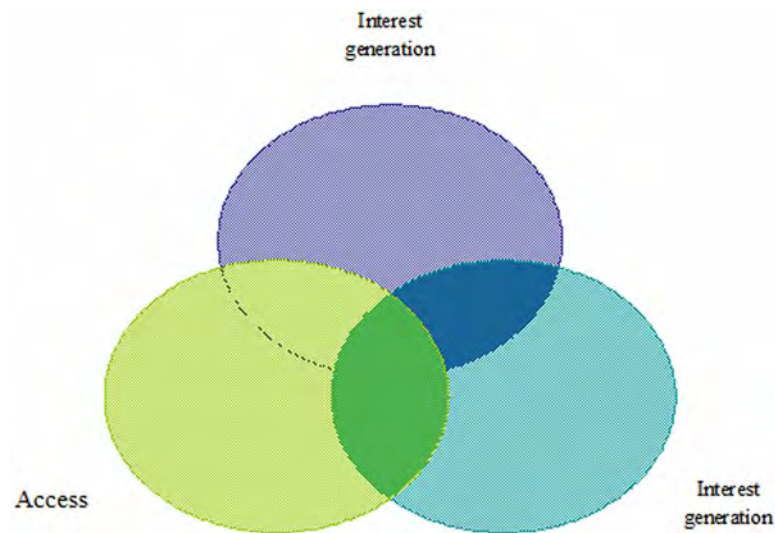
AR and Destination Image in Bangladesh: Bangladesh is a South Asian country having enormous potentials for tourism. The country has enormous tourism resources including the UNESCO World Heritage Sites. A particular tourism destination after enlisted in this list certainly attracts more attention from tourists all over the world (Hassan & Rahman, in press; Hassan & Iankova, 2012). Bangladesh has diverse tourism beauties ranging from naturally to culturally important attractions. The Historic Mosque City of Bagerhat, the Sundarbans and the Ruins of the Buddhist Vihara at Paharpur are some major tourist destinations in the country having the UNESCO WHS accolade. Unfortunately, very limited number of researchers is conducted to explore the manifold touristic approaches of these sites. A series of researches has been conducted to outline manifold aspects of these WHS in Bangladesh facilitating marketing and promotional propensities as: in the study titled, 'Tour on an Imagined Heritage Trail Set in the Mosque City of Bagerhat, Bangladesh: Cogitation for Market Potentials' Hassan (2014a) defines a novel heritage trail in this historic city. Later, he explained its market potentials based on theoretical explanations. Then in the paper titled, 'Package Eco-tour' as Special Interest Tourism Product-Bangladesh Perspective', Hassan (2012a) identifies a relatively new tourism product having a better fit in a delicate and pristine natural setting of the Sundarbans. Also, Rahman and Hassan (in press) critically outline sustainability issues while promoting tourism in the Sundarbans. In addition in the study as titled, 'Macromarketing Perspective in Promoting Tourism: The Case of the Buddhist Vihara at Paharpur' Hassan and Rahman (2015)

suggested the experimental application of macromarketing outlining its affectivity in the given perspective. The country also have untapped medical tourism potentials when, the Algerian scenario also renders some of similarities (Hassan, et al., 2015; Bouziane & Hassan, 2015).

Still, the country does not necessarily have a set of national policies that should benefit tourism (Hassan & Burns, 2014). In tourism, the potential capacities of AR have not been widely exploited. In tourism destinations of Bangladesh, AR is still viewed as remaining at its very infancy stage requiring further development and capacity exploitations. Still, the capacities of destination image formation by AR need to be explored and more researched. These obviously lead to modify the definition of AR in its applied meaning (Olsson et al., 2012). The application of AR technology on Smartphone devices has been naturally adopted by the tourists. However in Bangladesh, the availability of Smartphones is rather limited meaning that the use of AR in this country limits within defined affluent social classes. Still, present status of this specific technology addresses a series of issues to be implemented in a country like Bangladesh having an emerging tourism economy. Tourists' accounts of experiences are important in destination image formation as stated by citing the example of South Korea (Rahman et al., 2013). In Bangladesh, this technology needs to offer pleasant and purposeful tourist experiences.

Innovative marketing approaches are going to lead the future of tourism (Hassan, 2012b; Hassan, 2012c). The incorporation of effective technologies and expert human resources can appear as rewarding in such context (Hassan, 2014b). The impact and effects of AR in destination image formation is significant and expected to get enormous attention in coming years in Bangladesh. However, Bangladesh does not appear to be reluctant to ensure a solid position for this. This means that, AR technology can necessarily attract and motivate tourists to get engaged with a specific tourist destination in Bangladesh. The full

Figure 1. AR for Tourism Destination Image Formation



potentials of AR technology need to be exploited in this country. This should help in destination image formation and benefit the tourism industry very closely and effectively. The effective and proper application of AR should be attached with positive user experience in tourist destination image formation. Over decades, Bangladesh has found non-governmental organisations (NGOs) in development initiatives (Hassan & Ahmed, 2013a). The acceptance of AR for tourism destination image formation becomes important in Bangladesh. An effective use of AR for tourist destination image formation relies on its successful design and implementation. The Bangladesh experiences of non-governmental organisations (NGOs) seemingly appear as a success so far for sustainable development (Hassan & Forhad, 2013b). These positive experiences can suggest including NGOs in the tourism destination image formation supported by AR application. Relying on explanations on potential tourist bases, three basic functionalities of AR are: first, to create interesting image of a tourist destination; second, to form accessible image of a tourist destination and third, to create a playful image of a tourist destination resulting more tourist engagement. The functionalities are as in Figure 1.

FUTURE RESEARCH DIRECTIONS

This is evidenced that, AR is expected to be the key technology in the near future and should be necessarily attended by the tourism industry. Empirical evidences could have better support the arguments made in this research meaning that this was the basic lack of this study. Future researchers can necessarily research on practical user views to outline the concept in clearer and meaningful ways.

CONCLUSION

This study aimed to outline the application of AR as a valid technology to both support and promote destination image formation. The study explores that AR supports in destination image formation. For the academic researchers, this study offers a basic and conceptual underpinning that outlines the potentials and applications of AR technology for destination image formation. This study suggests that the outcome of AR application in tourism destination formation can be positive indicating that the technology can offer good perceived knowledge prior to make a visit. Still

this has also been found that, the application of AR should be based on effective user engagements in a given perspective. Developers and researchers are continuously improving the standards of this technology creating more potentials of AR that has not been exploited. This study establishes that, AR can redefine a tourist destination having interesting, accessible and playfulness images allowing more tourist engagements. In addition, this study offers an overview of AR technology application in tourism destination image formation in an emerging tourism market economy as Bangladesh. In this country, there are opportunities to place this technology into use when, a good destination image can necessarily have positive impacts on tourist's decision making. So, the study proposes to involve innovative technology as AR in destination image formation of countries having emerging tourism economies as Bangladesh. The application of AR is diverse in many areas and particularly in emerging tourism markets as Bangladesh.

REFERENCES

- Al-Kwafi, O. S. (2015). The Impact of Destination Images on Tourists Decision Making: A Technological Exploratory Study Using Fmri. *Journal of Hospitality and Tourism Technology*, 6(2), 174–194. doi:10.1108/JHTT-06-2015-0024
- Azim, R., & Hassan, A. (2013). Impact Analysis of Wireless and Mobile Technology on Business Management Strategies. *Journal of Information and Knowledge Management*, 2(2), 141–150.
- Bouziane, F., & Hassan, A. (2015). Showcasing the Untapped Medical Tourism Potentials of Algeria. *Tourism Spectrum*, 1(2), 38–47.
- Crompton, J. L. (1979). An Assessment of the Image of Mexico as a Vacation Destination and the Influence of Geographical Location upon that Image. *Journal of Travel Research*, 17(4), 18–23. doi:10.1177/004728757901700404
- Dadwal, S., & Hassan, A. (2015). The Augmented Reality Marketing: A Merger of Marketing and Technology in Tourism. In N. Ray (Ed.), *Emerging innovative marketing strategies in the tourism industry* (pp. 78–96). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-8699-1.ch005
- Echtner, C. M., & Brent Ritchie, J. R. (1991). *The Measurement of Tourism Destination Image*. Calgary: University of Calgary.
- Fritz, F., Susperregui, A., & Linaza, M. T. (2005). *Enhancing cultural tourism experiences with augmented reality technologies*. Retrieved from: <http://bit.ly/1OJe1Mo>
- Hassan, A. (2012a). Package Eco-tour' as Special Interest Tourism Product-Bangladesh Perspective. *Developing Country Studies*, 2(1), 1–8.
- Hassan, A. (2012b). Key Components for an Effective Marketing Planning: A Conceptual Analysis. *International Journal of Management & Development Studies*, 2(1), 68–70.
- Hassan, A. (2012c). Rationalization of Business Planning Through the Current Dynamics of Tourism. *International Journal of Management & Development Studies*, 2(1), 61–63.
- Hassan, A. (2013). Perspective Analysis and Implications of Visitor Management - Experiences from the Whitechapel Gallery, London. *Anatolia: An International Journal of Tourism and Hospitality Research*. DOI: 10.1080/13032917.2013.797916
- Hassan, A. (2014a). Tour on an Imagined Heritage Trail Set in the Mosque City of Bagerhat, Bangladesh: Cogitation for Market Potentials. In S. P. Bansal, S. Walia, & S. A. Rizwan (Eds.), *TOURISM: Present and Future Perspectives* (pp. 30–44). New Delhi: Kanishka Publishers.
- Hassan, A. (2014b). Revising the Five - Fold Framework in Human Resource Management Practices – Insights from a Small Scale Travel Agent. *Tourism Analysis*, 19(6), 799–805. doi:10.3727/108354214X14146848084400

- Hassan, A. (2015). The Customization of Electronic Word of Mouth: An Industry Tailored Application for Tourism Promotion. In S. Rathore & A. Panwar (Eds.), *Capturing, Analyzing and Managing Word-of-Mouth in the Digital Marketplace* (pp. 61–75). Hershey, PA: IGI Global.
- Hassan, A., & Ahmed, F. (2013a). The Role of NGOs in Preventing Unemployment in Bangladesh. *Kuwait Chapter of Arabian Journal of Business and Management Review*, 2(9), 1–12. doi:10.12816/0001227
- Hassan, A., Ahmed, M. U., & Shoeb-Ur-Rahman, M. (2015). The Development, Nature, and Impact of Medical Tourism in Bangladesh. In M. Cooper, K. Vafadari, & M. Hieda (Eds.), *Current Issues and Emerging Trends in Medical Tourism* (pp. 296–311). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-8574-1.ch021
- Hassan, A., & Burns, P. (2014). Tourism Policies of Bangladesh – A Contextual Analysis. *Tourism Planning & Development*. DOI: 10.1080/21568316.2013.874366
- Hassan, A., & Dadwal, S. (in press). Search Engine Marketing – An Outlining of Conceptualization and Strategic Application. In W. Ozuem & G. Bowen (Eds.), *Competitive Social Media Marketing Strategies*. Hershey, PA: IGI Global.
- Hassan, A., & Donatella, P. S. (in press). Google AdSense as a Mobile Technology in Education. In J. L. Holland (Ed.), *Handbook of Research on Wearable and Mobile Technologies in Education*. Hershey, PA: IGI Global. doi:10.4018/978-1-5225-0069-8.ch011
- Hassan, A., & Forhad, A. (2013b). The Role of NGOs in the Sustainable Development of Bangladesh. *Present Environment & Sustainable Development*, 7(2), 59–72.
- Hassan, A., & Iankova, K. (2012). Strategies and Challenges of Tourist Facilities Management in the World Heritage Site: Case of the Maritime Greenwich, London. *Tourism Analysis*, 17(6), 791–803. doi:10.3727/108354212X13531051127348
- Hassan, A., & Jung, T. (in press). Augmented Reality as an Emerging Application in Tourism Education. In D. H. Choi, A. Dailey-Hebert, & J. S. Estes (Eds.), *Emerging Tools and Applications of Virtual Reality in Education*. Hershey, PA: IGI Global. doi:10.4018/978-1-4666-9837-6.ch008
- Hassan, A., & Rahimi, R. (in press). Consuming ‘Innovation’: Feature Analysis of Augmented Reality in Digital Tourism Marketing. In P. Nikolous & I. Bregoli (Eds.), *Global Dynamics in Travel, Tourism, and Hospitality*. Hershey, PA: IGI Global.
- Hassan, A., & Rahman, M. (2015). Macromarketing Perspective in Promoting Tourism: The Case of the Buddhist Vihara at Paharpur. *Tourism Spectrum*, 1(2), 13–19.
- Hassan, A., & Rahman, M. (in press). World Heritage Site as a Label in Branding a Place. *Journal of Cultural Heritage Management and Sustainable Development*.
- Hassan, A., & Ramkissoon, H. (in press). Augmented Reality for Visitor Experiences. In J. N. Albrecht (Ed.), *Visitor Management*. Oxfordshire, UK: CABI.
- Lakshmi, R., & Ganesan, P. (2010). Does self-service technology impact destination image and customer delight? *Worldwide Hospitality and Tourism Themes*, 2(5), 520–553. doi:10.1108/17554211011090139
- Lawson, F., & Baud-Bovy, M. (1977). *Tourism and Recreational Development*. London: Architectural Press.
- Nolan, D. (1976). Tourists Use and Evaluation of Travel Information. *Journal of Travel Research*, 14, 6–8. doi:10.1177/004728757601400302
- Olsson, T., Kärkkäinen, T., Lagerstam, E., & Ventä-Olkkonen, L. (2012). User evaluation of mobile augmented reality scenarios. *Journal of Ambient Intelligence and Smart Environments*, 4(1), 29–47.

Pearce, P. L. (1988). *The Ulysses Factor: Evaluating Visitors in Tourist Settings*. New York: Springer Verlag. doi:10.1007/978-1-4612-3924-6

Rahman, M., & Hassan, A. (in press). Sustainable Tourism Practices and Development: The Case of the Sundarbans. *Tourism Dimensions*.

Rahman, M., Rasheed, S., Chowdhury, D., & Hassan, A. (2013). Sense of a Place: Tourist's Account of the Political History and Society of South Korea. *Review of Enterprise and Management Science*, 1(1), 51–59.

ADDITIONAL READING

Chon, K. S. (1990). The Role of Destination Image in Tourism: A Review and Discussion. *The Tourist Review*, 2(2), 2–9. doi:10.1108/eb058040

Chon, K. S. (1992). The Role of Destination Image in Tourism: An Extension. *The Tourist Review*, 2(1), 2–7. doi:10.1108/eb058086

Olsson, T., Kärkkäinen, T., Lagerstam, E., & Ventä-Olkkonen, L. (2012). User Evaluation of Mobile Augmented Reality Scenarios. *Journal of Ambient Intelligence and Smart Environments*, 4(1), 29–47.

Yovcheva, Z., Buhalis, D., & Gatzidis, C. (2012). Overview of Smartphone Augmented Reality Applications for Tourism. *e-Review of Tourism Research (eRTR)*, 10(2), 63–66.



KEY TERMS AND DEFINITIONS

Augmented Reality: This technology is regarded as the later stage of Virtual Reality (VR). Following this technology, a human as a user interacts with a computer generated environment that augments reality.

Emerging Tourism Economy: The economy of a country where, the contribution of tourism is getting a gradual importance.

Innovation: Innovation refers to a method, idea, or product that is considered as new.

Internet: Internet is an electronic network connecting billions of computers globally.

Technology Adoption: The adoption or use of a technology.

Tourist Destination Image Formation: Tourist destination is meant as an attitudinal or mental construct development based on impressions.

Tourist Perception: The ideas of a tourist prior to actual visit.

Destination @–Branding of Ten European Capitals Through the Institutional Stems and Commercial Logos

Elena Bocci

Sapienza University of Rome, Italy

Annamaria Silvana de Rosa

Sapienza University of Rome, Italy

Laura Dryjanska

Sapienza University of Rome, Italy

INTRODUCTION

Starting from a broad research programme on *Place-identity and Social Representations of European Capitals in first visitors of six different nationalities* begun by de Rosa in the 1990s (de Rosa, 1995; 1997; 2013b) - later developed along multiple interrelated research lines based on “*field studies*” and “*media studies*”, inspired by a multi-method *modelling approach* to social representations (de Rosa, 2013a; 2013c) - this contribution represents an integrative work concerning a study on the “*Destination@-branding*” (Morrison & Anderson, 2002) of ten European Capitals through communication via their institutional stems and commercial logos.

The research line based on the “*media studies*” (de Rosa, Bocci & Picone, 2012; de Rosa & Bocci, 2014) focuses on the comparative analyses carried out between:

- The “*City@-brand identity*” (Aaker & Joachimsthaler, 2002) created by the marketers (assumed as vehicle of “expert knowledge”) through the institutional tourist websites of ten European Capitals (Rome, London, Paris, Helsinki, Vienna, Warsaw, Berlin, Madrid, Brussels and Lisbon) -examining their usability, interactivity and contents.

- The “*City@-brand image*” (Keller, 1998; Cai, 2002) perceived through spontaneous conversations and experience exchanges among members of the Social Networks, like *Facebook* and *Yahoo Answer* and forum discussions like *TripAdvisor*, assumed as vehicle of the “common sense knowledge”.

In accordance with the model of “*destination branding*” (Cai, 2002) composed by three interrelated components: *brand identity*, *brand image* and *brand element mix* (name, logo, sign, design, symbol, slogan...), this contribution focuses on the institutional stems and commercial logos as symbolic tools and cultural artefacts created in different historical periods in order to contribute to the “*distinctiveness*” of the different cities.

Therefore, the aim is to compare the iconic structural elements of the brands (ancient and modern stems and logos) of ten historical European capitals, which play a determinant role in the narration of urban history.

The research also compares the social representations evoked by brands (stems and logos) of the ten European Capitals among potential first-visitors.

THEORETICAL BACKGROUND

Destination branding constitutes a way to communicate a destination's unique identity by differentiating a destination from its competitors (Morrison & Anderson, 2002).

In the model of destination branding proposed by Cai (2002) -organized around brand identity, brand image and brand element mix- the process starts choosing one or more brand elements -identifying the destination- and goes on with the formation of "brand associations" (attributes, affective and attitudes components of an image -Gartner, 1993; Keller, 1998-) driven by brand identity.

Moving beyond the molecular studies interested in identifying the cognitive and evaluative factors in perception, purely focused on the processes of categorization, encoding, storage and retrieval of information in memory, this chapter captures the multi-dimensionality of the theory

of social representations (Moscovici, 1961/1976; Jodelet, 1989; de Rosa, 2013a, 2016).

MAIN FOCUS: THE BRANDS

Modelling Approach Research Design

Given the relevance assigned to the iconic-imaginary dimension to social representations by the "modelling approach", the brands have been studied by using an appropriate research design as described in the Figure 1.

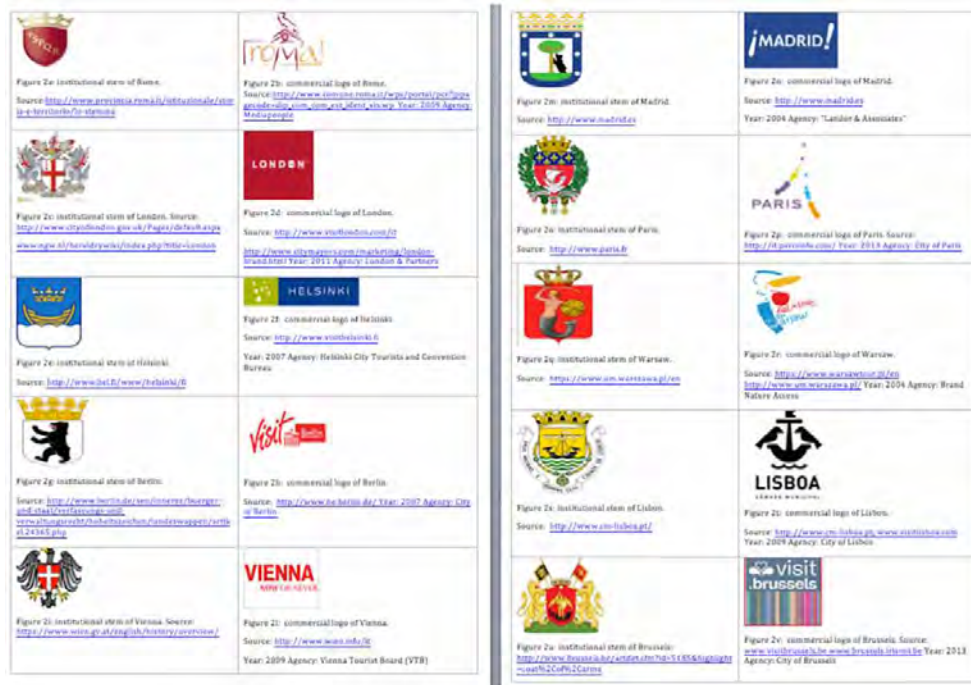
The "*modelling approach*", developed by de Rosa (2013a, 2014) is a paradigmatic option specific to the research field inspired by the Social Representation theory. It is aimed to grasp its core value as a unifying meta-theory of the social sciences, by operationalizing the investigation about any object of this supra-disciplinary field

Figure 1.

Table 1. The main features of the research design

Objective	Methodology	Questions	Hypotheses	Main results
The aim is to compare the iconic structural elements of the brands (ancient stems and modern logos) of ten historical European capitals, which play a determinant role in the narration of urban history.	In order to identify the social representations of the brands (stems and logos) we used: 1) a grid has been created ad hoc to identify the structural (distinctive) elements of the brands. (structured tool)	1) Which are the structural elements of the stems and logos?	We expect to identify discrepancies and/or contrasts of meaning between institutional stems and commercial logos in terms of structure of the elements that characterize these brands since, even though both are aimed at distinctiveness and recognition: • stems have historic roots, which often date back to the origins of the city; • logos were created in modern times, for well defined marketing purposes.	The stems origins from heraldry and evoke the prestige of the cities through rich combinations of different kind of elements: royal, military, religious or symbolic/mythological and natural. In many cases the history of the city contains elements of legends or religious beliefs, often recalling virtues of the Saints. These elements, present in the stems, narrate the history since the cities' foundation. On the commercial logos complex figures disappear completely, leaving space to modern elements: essential lines, abstract signs and circles innovating the brand and referring to socio-recreational, emotional, architectural and design aspects. These logos aim "to sell the city product", starting from the assumption that places can be branded just like products to maximize their attractiveness and their enjoyment by visitors.
The research compares the social representations evoked by brands (stems and logos) of the ten European Capitals among potential first-visitors.	2) the "associative network" using as iconic stimuli both the stems and the logos of each of the ten Capitals, able to determine the structure, contents and polarity of the Social Representations of stems and logos (de Rosa, 2002, 2003, 2005) (projective tool)	2) Which are the main differences between the social representations evoked by stems and logos? (Strengths and weaknesses; attitudes etc.) 3. Concerning commercial logos, is it possible to improve them and how?	We expect to identify discrepancies and/or contrasts of meaning between institutional stems and commercial logos concerning attitudinal components of their social representations since, even though both are aimed at distinctiveness and recognition: • stems have historic roots, which often date back to the origins of the city; • logos were created in modern times, for well defined marketing purposes. • We expect that it is possible to improve some logos by taking into account social representations of the cities.	Referring to the evaluative component of the social representations, the calculation of the indexes of polarities allowed us to detect a general more negative attitude towards institutional stems than towards commercial logos. In particular, the stems with the lower index of polarity are those of Berlin and Warsaw, while the institutional stem of Lisbon is the only case where the institutional stem has a more positive evaluation than the commercial logo. The reason has been found in the contents of the different brands: the richness of details -that take us back in time to the colonialism- and the predominance of yellow are the strengths of the stem; while, having taken over the central part of the emblem, using only the black colour and clean lines constitute leads to more negative evaluation of the commercial logo. It is possible to improve some logos. i.e. while the commercial logo of Rome is not among the best ones in terms of approval rating, the idea of an increased focus on a city with which one can co-establish a communication able to evoke a more intimate relational dimension should reinforce -with a positive effect- the trend already registered for brands like Paris. To embody the emotional-relational dimension seems right also for other capitals, such as Helsinki, London and Vienna.

Figure 2.



in multi-methodological research designs. These require to be fully justified and adequately complex depending on the multi-theoretical perspective adopted and the variety of constructs selected (representations, identity, social memory, emotions, etc.) the diverse techniques of data collection (structured and projective, textual and figurative, verbal or behavioural, etc.) the choice of the data analysis strategies and the expected results.

Research Focus

In this contribution the following brands (institutional stems and commercial logos) have been considered:

Research Instruments and Data Analysis Strategies

Structural and projective techniques have been created or developed, considering as iconic stimuli both the stems and the logos of the ten Capitals.

1. The grid is a structured tool, in format of a table, realized ad hoc for this research line and aimed to detect the presence/absence of the following distinctive components of the brands:
 - a. Name of the capital and texts in national languages;
 - b. Slogan;
 - c. Acronyms and texts in Latin language
 - d. Royal elements;
 - e. Elements of military origin;
 - f. Religious or symbolic/mythological elements;
 - g. Natural elements;
 - h. Abstract graphic elements.

The grid has been applied to all the stems and logos in order to identify the structural main elements of the different brands.

2. The “associative network” (de Rosa, 2002; 2003) is a tool of a projective nature; there-

Figure 3.



fore it is less subject to the phenomenon of social desirability, if compared to structured tools like for example a questionnaire. It enables respondents to specify the structure, content and polarity of a semantic field by themselves.

The tool requires first to associate words with stimulus and then to establish connections and branching patterns between the elicited words that are written around the brand.

The associative network requires people to attribute a particular polarity to each word (positive, negative or neutral) to describe its connotations. This allows to detect the evaluative component of the representations. A polarity index calculates the positive, negative, or neutral connotations of the free associations evoked by each of the iconic stimulus used in this study. This index, which varies from +1 to -1 is calculated using the following formula:

Polarity Index (P) = (n° of positive words – n° of negative words)/total n° of associated words.

A second “neutrality” control index, which also varies between -1 and +1, is also calculated.

Neutrality Index (N) = (n° of neutral words – (n° of positive words + n° of negative words))/total n° of associated words.

Sample iconic stimulus are presented in Figure 3.

Through the software SPAD (Lebart, Morineau & Beque, 1989) applied to the associative network, it is possible to reconstruct the structure and content of the representational fields associated with the iconic stimuli (procedures

Talex-contingency tables- and *Corbit*-analysis of latent dimensions, including as active variables indexes and words evoked-).

For each of the iconic stimuli the first five factors were extracted, whose cumulative percentage of variance explained more than 80% of the total.

Among the indicators useful in the interpretation of the results, the following are taken into account:

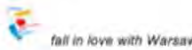
1. The factorial coordinates of each word on the first 5 factors which establish their position in the axes, in terms of distance from the origin and positioning on the positive or negative side;
2. The absolute contribution, which constitutes the part of the total inertia of the factor explained by each variable (Ercolani, Areni & Mannetti, 1990);
3. The relative contribution, or cosine squared, which assesses the contribution the factor provides to the explanation of the variability of each modality.

The analysis makes it possible to identify hidden dimensions (factors) that are subtended to the data and summarize the relations between original variables.

The aim is that of rendering of simpler interpretation the whole range of information through a synthesis, so that newly identified factors could represent a good approximation of the starting data-matrix. Due to limited space, this chapter takes into account the factorial interpretation, leaving out further elaborations of geometric-structural nature.

Destination @-Branding of Ten European Capitals

Table 1. Elements of congruence/difference between the brands of the various capitals under scrutiny

Brands	Institutional stems		Commercial logos	
Name of the capitals and texts in national languages	Absent		BERLIN BRUSSELS HELSINKI LISBON LONDON MADRID PARIS ROME VIENNA WARSAW 	
Slogan	Absent		VIENNA  WARSAW 	
Acronyms and texts in Latin language	LONDON  Domine Dirige Nos PARIS  Fluctuat nec mergitur ROME  S.P.Q.R.		Absent	
Royal elements (wreaths and crosses)			Absent	
Elements of military origin	LONDON  helmet and shield PARIS  medals WARSAW  shield and sword BRUSSELS  Shield		Absent	

Brands	Institutional stems		Commercial logos		Brands	
Religious or symbolic/ mythological elements	LONDON  winged dragon WARSAW  mermaid BRUSSELS  golden lions with red tongues		WARSAW  mermaid -also abstract graphic elements-		ROME  wolf and Traian's column	
Natural elements	LONDON  oak leaves BELGIUM  waves of the white sea PARIS  laurel wreath BERLIN  bear MADRID  bear and fruit tree LONDON  sea and black crows BRUSSELS  lawn ROME  Eagle		LONDON  river -also abstract graphic elements-		LISBOA  sea and black crows BRUSSELS  iris	
Abstract graphic elements	Absent		HELSINKI  the image of H consists of 7 circles metaphors of: 1) bridge between West and East; 2) the archipelago and the Baltic Sea; 3) a metropolis of a friendly size; 4) environmental concern and safety; 5) activism and culture; 6) architecture and design; 7) agreeable people.		PARIS  Eiffel Tower BERLIN  Brandenburg Gate	

The analysis was carried out on a group of 40 potential first visitors, experts in training in “Marketing and Communication” (Master level) at

the Faculty of Medicine and Psychology - University of Rome “La Sapienza”. They have followed the Laboratory of Web Marketing, focusing the

analysis of Social Representations of Historical Capital Cities.

Results of the Analysis Through the Grid

The analysis of the brands, made using the *ad hoc* designed grid, highlighted differences between stems and logos under scrutiny (Table 1).

The table allows us to answer to the following main question:

Question 1: Which are the structural elements of the stems and logos?

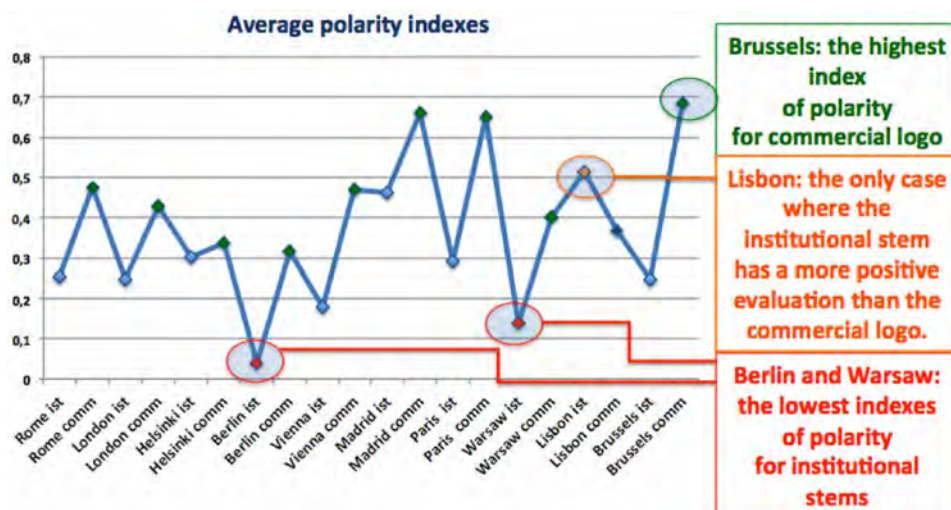
The stems origins from heraldry and evoke the prestige of the cities through rich combinations of different kind of elements: royal, military, religious or symbolic/mythological and natural. In many cases the history of the city contains elements of legends or religious beliefs, often recalling virtues of the Saints:

- The “armed mermaid” recalls a mythological figure prepared to protect the city of Warsaw and its inhabitants;

- The “dragons” refer to Saint George’s victory over one of them depicted in the stem of London;
- The “crows” in the stem of Lisbon lift up in the air the body of Saint Vincent, then recovered by his followers;
- The “burgundy color” of Rome recalls the legend of Ancile, one of twelve sacred shields kept in the Temple of Mars, which fell from heaven during the reign of Numa Pompilius to protect the city;
- The “Archangel Michael” pierces the black devil in the stem of Brussels.

These elements present in the stems narrate the history since the cities’ foundation. On the commercial logos complex figures disappear completely, leaving space to modern elements: essential lines, abstract signs and circles innovating the brand and referring to socio-recreational, emotional, architectural and design aspects. These logos aim “to sell the city product”, starting from the assumption that places can be branded just like products to maximize their attractiveness and their enjoyment by visitors (Ashworth, 1994; Anholt, 2003, 2010; Freire 2005), as summed up in Figure 4.

Figure 4.



Results of the Analysis Through the Associative Networks

The following figure helps us to answer the main question:

Question 2: Which are the main differences about the social representations evoked by stems and logos? (Strengths and weaknesses; attitudes etc.)

The figure compares the average polarity indexes calculated from the positive, negative and neutral assessments attributed by research participants to the associations evoked by the stems and logos of the ten historic European capitals.

First, the cases where the average polarity index of the representational field evoked by institutional stems is lower shall be examined. *Which social representations are evoked by these stems?*

In both cases the overwhelmingly important topic is the “war”.

Concerning the city of *Berlin* (average index of polarity institutional stem 0,040), next to the functional aspects of a city considered as “organized” (first factor, positive semi-axis, a.c.3.6 r.c.0.49) appears the historic-artistic-cultural dimension with terms such as “wall” (a.c.7.9 c.r. 0.78) and “museums” (a.c.3.1 r.c.0.29) on the negative semi-axis of the first factor; “Brandenburg Gate” (a.c.6.2 r.c.0.43) and “culture” (a.c.5.2 r.c.0.40) on the positive semi-axis of the second factor; the tight relationship of art/culture with the darkest period of recent history tends to be associated with a negative polarity index on the two factors

(negative semi-axis of the first factor a.c.11.8 r.c.0.26; positive semi-axis of the second factor a.c.11.9 r.c.0.16).

The binominal art/culture and the history of the XXth century can be seen even clearer in the social representations of stems, as demonstrated in the first two factors: the first factor–negative semi-axis- includes such terms as: “nazism” (a.c.17.8 r.c.0.89), “East Berlin” (a.c.11.1 r.c.0.68), “war” (a.c.10.3 r.c.0.88) and “wall” (a.c.5.1 r.c.0.37) and the second factor – negative semi-axis- includes such terms as: “wall” (a.c.8.7 r.c.0.50) and “violence” (a.c.4.1 r.c.0.30). In both factors active variables elicited coincide with a negative polarity index (negative semi-axis of the first factor a.c.14.2 r.c.0.33; negative semi-axis of the second factor a.c.11.2 r.c.0.51).

It is the second factor in particular that shows a split in the attitudinal component of social representations; while the negative polarity goes with the historical events of the last century, the positive polarity corresponds to recent cultural events (“Cinema Festival” a.c.10.2 r.c.0.60) and to ancient history of the city (“ancient” a.c.4.2 r.c.0.37).

References to the XXth century for the city of Berlin are in continuity with the city of Warsaw (average index of polarity institutional stem 0.1393) whose brands are full of events related to World War II. In reference to the commercial logo, the word “Auschwitz” occurs on the second factor of the positive semi-axis (a.c.15.2 r.c.0.53) next to “cold” (a.c.26.6 r.c.0.78) and “sad” (a.c.11.4 r.c.0.86) associated with a negative polarity index (a.c.70.6 r.c.0.96); on the positive semi-axis of the fourth factor (a.c.10.1 r.c.0.14), as well as on the negative semi-axis of the fifth factor (a.c.21.6 r.c.0.22).

The history of the XXth century also shows an impact in the social representations evoked by the stems.

The first factor in particular demonstrates a rift in the attitudinal component of social representations:

Figure 5.



- On the positive semi-axis the positive polarity (a.c.26.8 r.c.0.79) corresponds to general aspects such as “history” (a.c.5.4 r.c.0.29) and nature-related themes (“river” a.c.7.1 r.c.0.60);
- On the negative semi-axis negative polarity (a.c.43.1 r.c.0.83) corresponds to the references to the “difficult” (a.c.7.5 r.c.0.65) period of the Second World War that caused suffering and victims, with terms like “blood” (a.c.14.4 r.c.0.68) and “war” (a.c.10.6 r.c.0.63).

The word “Auschwitz” appears on the positive semi-axis of the second factor (a.c.45.9 r.c.0.73), together with “Poland” (a.c.19.0 r.c.0.82) and “antique” (a.c.14.3 r.c.0.40), contaminating time and space, and on the positive semi-axis of the third factor (a.c.11.5 r.c.0.14) together with “war” (a.c.6.4 r.c.0.23).

Next, the city of *Lisbon* (average index of polarity institutional stem 0.5135) is examined; the only case where the stem has a more positive evaluation than the logo.

What are the strengths of the institutional stem and what are the weaknesses of the commercial logo?

The richness of details -that take us back in time to the colonialism- and the predominance of yellow are the strengths of the stem; while, having taken over the central part of the emblem, using only the black colour and clean lines constitute leads to more negative evaluation of the commercial logo.

In spite of common elements:

- For the stem the positive social representations coincide with the references to “colonialism” presented on different factors: negative semi-axis of the third factor with keywords such as “wealth” (a.c.6.5 r.c.0.32), “history” (a.c.6.1 r.c.0.66), “power” (a.c.5.4 r.c.0.43) and “sea” (a.c.3.1 r.c.0.24); positive semi-axis of the fourth factor with the keywords “Middle Ages” (a.c.6.3 r.c.0.19)

and “boat” (a.c.7.3 r.c.0.37) and also on the negative semi-axis where are present “colonies” (a.c.9.1 r.c.0.39), “riches” (a.c.7.0 r.c.0.32), “trade” (a.c.6.0 r.c.0.37) and “discovery” (a.c.4.2 r.c.0.45); positive semi-axis of the fifth factor with the words “colonies” (a.c.11.7 r.c.0.34), “trade” (a.c.6.0 r.c.0.25), “discovery” (a.c.5.1 r.c.0.37), “boat” (a.c.4.8 r.c.0.16);

- For the logo the negative polarity index corresponds to the positive semi-axis of the second factor (a.c.46.2 r.c.0.72) and to the negative semi-axis of the third factor (a.c.8.5 r.c.0.9) to the negative emotion “sadness” (respectively a.c.34.1 r.c.0.73 on the second factor and a.c.6.1 r.c.0.9 on the third factor), also associated with the “black” colour on the second factor (a.c.12.5 r.c.0.47).

In the following section we present the analysis of the cases where the positive polarity index of the commercial logos are higher:

Which social representations are evoked by these logos?

Social representations associated with the logo of *Brussels* (average index of polarity for commercial logo 0.6855), rich in metaphors, are centered on the topic of “multi-culturalism”. Sometimes interpreted as a “bar code” (first factor negative semi-axis a.c.54.1 r.c.0.91; second factor negative semi-axis a.c.8.2 r.c.0.6), the colors bring to mind the idea of a “game” (first factor negative semi-axis a.c.24.1 r.c.0.56), in other cases they are a synonym of the “rainbow” of races-cultures (third factor positive semi-axis a.c.5.5 r.c.0.43; fifth factor negative semi-axis a.c.5.5 r.c.0.20), an invitation to “open up” (third factor positive semi-axis a.c.8.0 r.c.0.42) and to “love” (a.c.5.2 r.c.0.19). Love appears both on the positive semi-axis of the third factor and fourth factor (a.c.4.5 r.c.0.13), here also together with “multicultural” (a.c.3.2 r.c.0.21).

The content associated with the institutional stem of *Brussels* appears very different from the

one evoked by the commercial logo, concentrating on the historical/cultural and religious elements evoked by the legendary mythological figures represented.

After Brussels, also the commercial logos of Madrid and Paris are very positive. The positivity seems related to “nightlife” that cuts across different factors, referring to socio-recreational city through more keywords: the “night life” appears on the negative semi-axis of the second factor (a.c.6,8 r.c.0.46), together with “inviting” (a.c.5.5 r.c.0.39) and on the positive semi-axis of the fifth factor (a.c.6.2 r.c.0.25), together with “happiness” (a.c.14.7 r.c.0.63), “sea” (a.c.13.7 r.c.0.59) and “youth” (a.c.4.9 r.c.0.76); on the positive semi-axis of the third factor there is “movida” (a.c.3.3 r.c.0.20), together with the “sun” (a.c.4.9 r.c.0.66), “sea” (a.c.3.8 r.c.0.20) and “heat” (a.c.3.5 r.c.0.12).

Free associations related to the institutional stem of Madrid do not seem to refer to the characteristic “movida” (nightlife), remaining somewhat of a descriptive/evaluative dimension of the elements represented.

Also for Paris, as in case of Madrid, the relational dimension is enhanced in the commercial logo from a more intimate perspective, which leaves room for romance, as we can observe in the third factor extracted from Spad. On the positive semi-axis appear in fact words like “charming” (a.c.16.4 r.c.0.65), “romantic” (a.c.8.7 r.c.0.63), “magic” (a.c.7.3 r.c.0.56), “evening” (a.c.4.7 r.c.0.29), “lights” (a.c.3.2 r.c.0.11); on the negative semi-axis of the same factor the dimension of values is expressed through such terms as “love” (a.c.13.0 r.c.0.72) and “Christmas” (a.c.4.8 r.c.0.28). “Christmas” (a.c.20.0 r.c.0.48), “fascinating” (a.c.9.7 r.c.0.16) and “romantic” (a.c.3.5 r.c.0.11) is also repeated on the positive semi-axis of the fifth factor.

The relational/value dimension is also detected in reference to the institutional stem; on the first, second, fourth and fifth factor recur the words “light” and “romantic”, also repeated on the positive semi-axis of the second factor in correspondence with the word “love” (“lights”

a.c.5.4 r.c.0.26; “romantic” a.c.5.4 r.c.0.26; “love” a.c.4.3 r.c.0.51).

FURTHER RESULTS AND RECOMMENDATIONS

Recently “Roma Capitale” has modified her brand, from the old stem to and the old logo both with historical iconic (“crown”, “shield”, “wolf”) and textual symbols (“S.P.Q.R.”, “capital”) to the new one including a strong “relational” attachment and identification dimension (Rome & You)¹.

According to our results, while the commercial logo of Rome is not among the best ones in terms of approval rating registered by the polarity index, the idea of an increased focus on a city with which one can co-establish a communication that is not pre-defined and able to evoke a more intimate relational dimension should reinforce -with a positive effect- the trend already registered for the analyzed brand (similar to that of Paris, with keywords as “lights”, “romantic” and “love”).

To embody the emotional-relational dimension seems right also for other capitals, such as Helsinki, London and Vienna. In fact, despite that one of the seven points of the commercial logo of Helsinki refers to the friendliness of people, in the eyes of research participants the relational component is absent in the five factors extracted for each of the three capitals –for both brands-; moreover, concerning the emotional dimension

Figure 6.



Figure 7.



appears only the word “sad” (a.c.26.4 r.c.0.63) -for commercial logo- precisely referring to Helsinki, on the negative semi-axis of the third factor, together “communication” (a.c.6.5 r.c.0.83) and “modern” (a.c.4.7 r.c.0.53), marked by a negative polarity index (a.c.43.3 r.c.0.58).

In proposing recommendations, however, it should be noted that these changes often become the subject of a polemical social discourse that is increasingly divided over the Internet (especially websites, blogs and social networks) and can even last for years, as was the case for commercial logos used in this research related to London² and Rome³ or more recently during the presentation of the new logo that generated strong protests and complaints⁴:

The controversies triggered by the web show that: touching iconic symbols generates more controversies than replacing words, as de Rosa learned investigating the controversial representations provoked by the change of the logo of the Italian Communist Party (P.C.I.) (de Rosa & Farr, 2001; de Rosa, 2014).

The symbols seem to possess the power of materialising the deepest dimension of the figurative nucleus of the social representation. More specifically, some of the elements that can be included

under the label “brand element mix” possess, as Rouquette noted (1994), the following properties:

- They embody the emotive dimension of the central nucleus of a social representation -dimension defined by the author with the term “nexus”-;
- Merge denotation and connotation;
- Reify (objectify) a system of cadences and values;
- Are hard to modify through analysis and rational reflection;
- Constitute a system of communication relatively independent vis-à-vis that of the discourse. (...)

In the symbolic order, in fact, to act on the symbol of a person or of an object, means to act on the person or on the object itself. To strike the symbol means to strike directly the entity that it represents (de Rosa & Farr, 2001; de Rosa, 2014).

FUTURE RESEARCH DIRECTIONS

Given the continuous evolution of some of the commercial logos, like the logo of Rome discussed

above, the future research should include these new logos and compare them with the previous ones. Moreover, the results obtained by analyzing the institutional and commercial logos as elements of the “brand element mix” should be further integrated with those concerning the divergence/convergence of the City@-brand identity and City@-brand image, for a comprehensive view of the Destination@-branding of the ten European Capitals.

CONCLUSION

Following our hypotheses, the main differences between stems and logos emerged through the use of the ad hoc grid that showed the divergences in the structural elements of stems and brands and also through the calculation of the indexes of polarities that allowed us to detect a more positive attitude towards commercial logos than towards institutional stems.

On the contrary, shared contents between stems and logos have been found referring to:

- The “history” - the “Second World War” in the brands of Berlin and Warsaw and the “colonialism” in the brand of Lisbon -;
- The “emotional” and “value-related” dimensions in the brands of the romantic Paris and Rome.

Concerning commercial logos, especially the “emotional” and “value-related” dimensions affected in the opposite way (negative vs positive) the social representations:

- Negatively in the cases of Warsaw and Lisbon (sadness);
- Positively in the cases of Brussels, Paris and Rome (love) or Madrid (happiness).

Following the improved importance of the “emotional” and “value-related” dimensions (Preiti, 2013) the recent idea of a new commercial

brand for Rome - including a strong relational dimension - is in line with the communicative policies adopted by other capitals. The polemical social discourse that it has activated among citizens confirms the dynamic of resistance/acceptance in the social representations provoked by changes related to their symbolic and iconic dimensions and the interest for investigating the social and communicative phenomena related to the iconic narratives about cities evoked by the brands and the city destination@-branding” (Bocci, de Rosa, & Dryjanska, 2015).

Besides their descriptive value and possible function in guiding city’s institutional and tourist communication managers, the results have a historical value with respect to the evolution of the Institutional communication, assuming a particular interest for further comparative analyses (de Rosa & Bocci, 2013; de Rosa, Bocci & Picone, 2012) concerning the specific component of the brand element mix (Cai, 2002), here analytically described through structured and projective tools aimed at identifying both the structural (distinctive elements) of the brands and the structure, contents and polarities of the social representations associated to them.

The study is part of a more general attempt to deal with the complexity of the new communicative web scenario based on multi-channels, assumed to be influential for the genesis, transmission and negotiation of the social representations of historical Capital Cities. The policy makers should therefore take into account social representations and test innovative design ideas in order to verify that a new logo for example includes elements considered as essential in how people represent a city.

REFERENCES

Aaker, D., & Joachimsthaler, E. (2000). *Brand Leadership: Building Assets in an Information Economy*. New York: The Free Press.

- Anholt, S. (2003). Branding places and nations. In R. Clifton & J. Simmons (Eds.), *Brands & Branding* (pp. 213–226). Bloomberg Press.
- Anholt, S. (2010). Definitions of place branding: Working towards a resolution. *Place Branding and Public Diplomacy*, 6(1), 1–10. doi:10.1057/pb.2010.3
- Ashworth, G.J. (1994). Marketing of places: What are we doing? In G. Ave & F. Corsico (Eds.), *Marketing Urbano International Conference*. Torino, Italy: Edizioni Torino Incontra.
- Bocci, E., de Rosa, A. S., & Dryjanska, L. (2015). *Iconic Narratives of ten European Capitals evoked by the institutional stems and the commercial logos*. International conference organized by the Institute of Psychology, University of Pécs, in the framework of Cost Action IS1205 ‘Social psychological dynamics of historical representations in the enlarged European Union’, University of Pécs, Hungary. Retrieved from <http://costis1205.wix.com/home>
- Cai, A. (2002). Cooperative Branding for rural destinations. *Annals of Tourism Research*, 29(3), 720–742. doi:10.1016/S0160-7383(01)00080-9
- de Rosa, A. S. (1995). Psicologia del turismo: per una psicologia sociale dell’ambiente applicata. In B. Zani (Ed.), *Le dimensioni della psicologia sociale* (pp. 153–186). Bologna, Italy: Il Mulino.
- de Rosa, A. S. (1997). Turisti di sei nazionalità per la prima volta nella ‘città eterna’: ‘place identity’ e rappresentazioni sociali di Roma e del suo centro storico. In A. Nenci (Ed.), *Conoscere e rappresentare la città* (pp. 149–214). Padova, Italy: ed. Cedam.
- de Rosa, A. S. (2002). The “associative network”: A technique for detecting structure, contents, polarity and stereotyping indexes of the semantic field. *European Review of Applied Psychology*, 52(3/4), 181–200.
- de Rosa, A. S. (2003). Le réseau d’association. Une technique pour détecter la structure, les contenus, les indices de polarité, de neutralité et de stéréotypie du champ sémantique liés aux représentations sociales. In J.C. Abric (Ed.), *Méthodes d’étude des Représentations Sociales* (pp. 81–117). Ramonville Saint Agne: Erès.
- de Rosa, A. S. (Ed.). (2013a). *Social Representations in the “social arena”*. New York: Routledge.
- de Rosa, A. S. (2013b). Place-identity and Social Representations of historic capital cities: Rome through the eyes of first-visitors from six countries. In A. S. de Rosa (Ed.), *Social Representations in the “social arena”*. New York: Routledge.
- de Rosa, A. S. (2013c). Taking stock: a theory with more than half a century of history. In A.S. de Rosa (Ed.), *Social Representations in the “social arena”*. New York: Routledge.
- de Rosa, A. S. (2014). The role of the Iconic-Imaginary dimensions in the Modelling Approach to Social Representations. *Papers on Social Representations*, 23, 17.1–17.27. Available at <http://www.psych.lse.ac.uk/psr/>
- de Rosa, A. S. (2016). Mise en réseau scientifique et cartographie de la dissémination de la théorie des représentations sociales et son impact à l’ère de la culture bibliométrique. In G. Lo Monaco, S. Delouvée, & P. Rateaux (Eds.), *Les représentations sociales* (pp. 51–68). Brussels, Belgium: Editions de Boeck.
- de Rosa, A. S., & Bocci, E. (2014). Place @-Branding and European Capitals: “city visiting cards” via municipal websites, virtual tours of significant places flying with Google Earth, and conversational exchanges about city-places experienced/imagined via social networks. In A. Capoor & C. Kulshrestha (Eds.), *Dynamics of Competitive Advantage and Consumer Perception in Social Marketing* (pp. 126–168). Hershey, PA: IGI Global.

de Rosa, A. S., Bocci, E., & Picone, M. (2012). E-branding and institutional web sites: the “visiting card” of the Municipalities of Rome and Paris. In A. Kapoor & C. Kulshrestha (Eds.), *Branding and sustainable competitive advantage: building virtual presence* (pp. 207–247). Hershey, PA: IGI Global. doi:10.4018/978-1-61350-171-9.ch013

de Rosa, A. S., & Farr, R. (2001). Icon and symbol: Two sides of the coin in the Investigation of Social Representations. In F. Buschini & N. Kalampalikis (Eds.), *Penser la vie, le social, la nature. Mélanges en hommage à Serge Moscovici* (pp. 237–256). Paris, France: Les Editions de la Maison des Sciences de l’Homme.

Ercolani, A. P., Areni, A., & Mannetti, L. (1990). *La ricerca in psicologia*. Roma, Italy: Carocci.

Freire, J. R. (2005). Geo-branding, are we talking nonsense? A theoretical reflection on brands applied to places. *Place Branding*, 1(4), 347–362. doi:10.1057/palgrave.pb.5990033

Gartner, W. C. (1993). Image formation process. *Journal of Travel & Tourism Marketing*, 2(2/3), 191–215.

Jodelet, D. (1989). *Les représentations sociales*. Paris, France: Presses Universitaires de France.

Keller, K. L. (1998). *Strategic brand management: Building, measuring, and managing brand equity*. Prentice Hall.

Lebart, L., Morineau, A., & Bécue, M. (1989). *SPAD.T Système portable pour l’analyse des Données textuelles, Manuel de l’utilisateur*. Paris, France: Cisia.

Morrison, A., & Anderson, D. (2002). *Destination branding*. Available from: <http://www.macvb.org/internet/presentation/DestinationBrandingLO-zarks6-10-02.ppt>

Moscovici, S. (1961/1976). *La psychanalyse son image et son public. Etude sur la représentation sociale de la psychanalyse*. Paris, France: Presses Universitaires de France.

Preiti, A. N. (2013). *Roma, il turismo e la rivoluzione digitale. Come cambia l’industria dell’ospitalità*. Roma, Italy: Unindustria.

Rouquette, M. L. (1994). *Sur la connaissance des masses. Essai de psychologie politique*. Grenoble: Presses Universitaires de Grenoble.

ADDITIONAL READING

Baloglu, S., & McCleary, K. W. (1999). A model of destination image formation. *Annals of Tourism Research*, 26(4), 868–897. doi:10.1016/S0160-7383(99)00030-4

Florek, M., Insch, A., & Gnoth, J. (2006). City council websites as a means of place brand identity communication. *Place Branding*, 2(4), 276–296. doi:10.1057/palgrave.pb.6000036

Fuchs, C., & Schreier, M. (2011). Customer Empowerment in New Product Development. *Journal of Product Innovation Management*, 28(1), 17–32. doi:10.1111/j.1540-5885.2010.00778.x

Kapferer, J. (1997). *Strategic brand management*. U.K.: Kogan Page.

Pike, S. (2009). Destination brand positions of a competitive set of near home destination. *Tourism Management*, 30(6), 857–866. doi:10.1016/j.tourman.2008.12.007

Qu, H., Kim, L. H., & Im, H. H. (2011). A model of destination branding: Integrating the concept of the branding and destination image. *Tourism Management*, 32(3), 465–467. doi:10.1016/j.tourman.2010.03.014

KEY TERMS AND DEFINITIONS

Associative Network Technique: A tool developed by Annamaria Silvana de Rosa of projective nature used in order to study the Social

representations of iconic and verbal stimuli, by considering structure, contents and attitudes.

Destination@-Branding: Extension of the concept of destination-branding to the digital world of the Internet, including websites, social networks, portals, etc.

Destination-Branding: An articulation of different interrelated components: brand identity, brand image and brand element mix.

Index of Polarity: An index created by Annamaria Silvana de Rosa aimed at detecting the evaluative component of social representations, on the basis of the polarities (positive/negative) attributed to the words associated by the subjects to the stimuli.

Modelling Approach: A paradigmatic approach to Social Representations developed by Annamaria Silvana de Rosa that integrates different methods and techniques coherently with the articulation of different theoretical constructs and dimensions.

Nexus: The emotive dimension of the central nucleus of the social representations.

Social Representations: Naïve theories of common sense that express systems of values, convictions and norms of behavior, with a double function of organizing the perception of the world and serving as a shared system of social communication and interpersonal exchange.

H

ENDNOTES

- ¹ http://comune.roma.it/wps/portal/pcr?contentId=NEW807152&jp_pagecode=newsview.wp&ahew=contentId:jp_pagecode
- ² On the websites of Moving Brands and London and Partners.
- ³ <http://blog.adci.it/riflessioni/un-capitello-per-la-capitale-la-ciliegina-sul-logo-di-roma/>; <http://www.06blog.it/post/9340/roma-in-unimmagine-un-brand-per-la-citta-di-roma-il-nuovo-logo-di-roma>
- ⁴ http://www.tgcom24.mediaset.it/cronaca/roma-proteste-contro-il-nuovo-logo_2096177-201502a.shtml

The Effect of Social Media Networking in the Travel Industry

Androniki Kavoura

Technological Educational Institute of Athens, Greece

Efstathios Kefallonitis

State University of New York at Oswego, USA

INTRODUCTION

Many travelers have at least one account in one of the popular social media networking websites or applications (Facebook, Twitter, Flickr, Instagram). Airlines and travel agencies often encourage their customers to use their personal profile to: (a) either interact with the company directly and (b) participate in online discussions with other users about that company. The use of social-media for the purpose of interacting with travel industry companies and other travelers has created multiple online communities. The travel industry is using these online communities as a new advertising and communication channel to reach customers, thus serving both as a customer service and innovative marketing tool.

The multidisciplinary nature of this chapter identifies under-researched areas and sheds new light in the way that connected passengers search for travel-related information. The authors aim in exploring ways that connected passengers develop a sense of belonging in social-media networking websites.

The use of new technologies and new methods of communications in the travel industry has reinforced the need for dialogue among providers (airlines, travel agencies and so on) and users (passengers). Online communities serve as a networking platform that brings together positive and negative comments.

BACKGROUND OF RESEARCH: DEFINING SOCIAL MEDIA

Social media are Internet tools where people interact by creating user-generated content. This is an asynchronous channel of communication. There is perceived interactivity since interaction may not be necessary nor interpersonal. Communication may take place either via web-based technology or via mobile device applications (Carr & Hayes, 2015). Facebook users post comments and like other people's posts in order to communicate. In addition, Facebook users play online games, share personal experiences and interact by following other people's updates. This makes users feel a sense of belonging by sharing moments and keeping in touch with friends and relatives.

Flickr is a 'photo sharing community'. It consists of a user-generated online social network where users tell stories or share moments using photographs. The personal perspective of the user gets shared with the entire Internet community. This photo-sharing creates experiences that can be understood by larger groups of people (Dotan & Zaphiris, 2010). Photographs are used as a form of a visual vocabulary used along with hashtags. These are user-generated metadata, which organize the uploaded photos using hashtag as a means of coding. On Flickr a relationship exists between social affiliation (contacts) and hashtag vocabulary formation (Dotan & Zaphiris, 2010).

According to the Annual Report of Facebook, the global advertising revenue in 2015 is expected to reach 14.27 billion US dollars (“Facebook’s Advertising Revenue”, 2015). This demonstrates the tremendous advertising growth and a great opportunity. An increasing number of people spend more time online and less time using other digital media. The greater use of social media networks as part of marketing communications has the potential to outnumber traditional channels (Hutter, Hautz, Severin Dennhardt, & Füller, 2013). The success of social media marketing lays on the interactivity of the medium. This offers interaction among users and companies with the scope of increasing a company’s brand awareness.

There has been little research on the importance of interactivity between companies and consumers using new technologies. In particular ways that firms can communicate with their customers, learn from feedback received and engage with users in an interactive way.

THE ROLE OF CULTURE IN THE IMPLEMENTATION OF SOCIAL MEDIA USAGE

Social sensitivities need to be taken into consideration when online social networks and online communities are created. Customs, social norms and traditions sometimes dictate what is socially acceptable or not in an online community. Community is best understood as a sense of (or belief in) group commonality. Online communities are products of culture and reflect certain practices or acts exercised within that culture. Examples include ways of naming a community. Hofstede (1980) argued that culture influences a human group’s response to its environment. Culture may determine dominant issues within a community.

An example of cultural interference comes from the People’s Republic of China, where Facebook, one of the most popular social network websites, is currently banned (Riley, 2015). This is only accessible in Hong Kong (SAR). On the contrary,

local social networking websites such as renren.com are rather popular. Japan’s most popular social network website mixi is accessible to Japanese citizens with more than 27.1 million users. Mixi users are estimated to be approximately 5 million back in 2006, while Facebook users today are about 15 million (“About Mixi”, 2015; Ghedin, 2013).

In order for marketers to create strong online brand communities, they need to create or adopt a language or symbols that will help them connect better to their respective communities. This language will have to take into consideration cultural issues and sensitivities. Further research is needed to examine the significance attributed to the concept of a constructed language, along with its symbolism. Preliminary exploratory research in the hotel and tourism industry illustrated that members of an online community prefer using code-words, symbols and a commonly understood language (Kavoura & Tiago, 2016). Further research in this field will provide further insights.

ONLINE COMMUNITIES: PASSENGERS AND THE TRAVEL INDUSTRY

Users attracted to social networking websites have a certain degree of homogeneity. Usually they speak the same language and share similar interests in order to be incorporated in the social networking group. Gregorie, Salle, & Tripp (2015) pointed out that emphasis should be placed on the connected passenger and the respective online communities he/she belongs. Social media are changing the traditional communication norms between the customer and the company and user-generated content can in fact influence a company’s reputation.

The user-generated content of all users is available to the entire social media platform. This may influence others, meet popularity and create trends. Content creators often act as ambassadors of ideas or may represent a company and its products. Content creators are members of the

public or even well-known entities such as actors and athletes and so on. There are businesses that send posts to first-time users; recipients of these notes often post them on social media, increasing the brand's exposure (Lee, 2013). Research on the use of Facebook by airports, shows that the most influential posts are those that engage users and encourage online comments (Wattanacharoensil & Schuckert, 2015). User-generated communities on social media are generally more trusted than the corporate websites or blogs created by travel-related providers (Del Chiappa, 2013; Amaral, Tiago, & Tiago, 2014). This reveals that social-media users do not always trust the information that companies decide to share publicly. There have been cases that incidents of dissatisfied customers go viral on social-media platforms. Such an example was the case of a passenger's broken guitar when he flew United Airlines. The passenger created a YouTube video that attracted more than 15,343,972 views ("United Breaks Guitars", 2009).

Large amounts of useful data can be collected based on customers' points of view who act as co-creators in the marketing but also in the research and development process of the firm (Harrigan, Soutar, Choudhury, & Lowe, 2015). In fact, customers become the co-creators and companies can enhance their experiences by listening to their point of view (Prahanland & Ramaswamy, 2013). Thus, social media can be used as a means of enhancing Customer Relationship Management (CRM). Social media offers a platform to engage with the public on multiple levels and serves as a useful customer preference database (Harrigan et al., 2015).

Gunarathne, Huaxia, & Seidmann (2015) argue that firms and airlines are willing to respond to customers' comments and provide customer support in real time basis via the use of social media. An interesting finding reveals that some airlines give priority to customers with a higher number of online followers. A tool that helps marketers evaluate the social networking comments is reputation monitoring (Ntalianis, Kavoura, Tomaras,

& Drigas, 2015), besides websites such as Brand Watch, Hootsuite and others (Harrigan et al., 2015). Airlines can monitor the comments made with regard to their products and act accordingly.

Social media create online communities in the sense that online communication is used in a virtual environment that connects all those with similar interests (Sakas, Nasiopoulos & Kavoura, 2015). An airline may use social media to engage those customers who use social media intensely since a correlation has been found to exist based on research of an international airline; this in turn, may be used for the benefit of the airline's reputation (Dijkmans, Kerkhof, & Beukeboom, 2015). There are a number of different ways to complete travel reservations online. These include the use of online travel agents or airline websites. Both use a variety of ways to make their services accessible to the consumer such as websites, mobile applications, self-service kiosks and so on. This elevated experience and the variety of choice available are factors that influence consumer preference and choice (Holland, Jacobs, & Klein, 2015; Ntalianis et al., 2015; Xiang, Magnini, & Fesenmaier, 2015).

Reposting information on a variety of social media that feeds off from one single source is also a valid way to generate a unified voice and have presence in a variety of social media networks (Kroque, 2013).

Firms may invest a lot of money to attract customer engagement and promote a brand's name with the use of contests and competitions, (see for example, JetBlue's Airways contest in Sakas et al., 2015). Engaging the user is another way of communication that demonstrates care by answering questions, providing advice but also sharing life experiences online (Palsule, 2011; Kosarek, 2014). Such actions present a fine way of online interactivity by engaging in users' everyday lives. Twitter uses such an engagement in order to measure the number of retweets, responses and mentions. This allows an advertiser to have greater control over the budget and the outcome (Bershidsky, 2013).

Twitter engages users by employing replies and mentions. This is intended to make communication between users easier (“How to”, 2015). JetBlue Airways has almost 1.800.000 followers on Twitter (“JetBlue”, 2015a) and uses a personalized way to get in touch with users, aiming at specific people and offering a personal touch. It is typical that JetBlue Airways refers to the names of their customers on their personalized communication. As an example: “We’re lucky to have you as a customer [name of the customer]! As you know, we always welcome you on board with open arms! See you soon!” (“JetBlue”, 2015b).

Elaborating on stories is another practice that may bring a community together. Video clips that show the company’s interest in making a customer happy are increasing in popularity. Another example associated with the airline industry, was a marriage proposal that became viral. A JetBlue passenger wanted to propose to his girlfriend in-flight and the screens on the airplane loaded the video he had prepared (“Adam and Taryn”, 2013). It is not the narrative that is of significance but rather a story with a thematic development (van Dijck, 2013).

Overall, there are a number of opportunities regarding the use of social media in the travel industry:

Opportunities:

1. To create a unique personality and seamless experience for a traveler. This can be utilized best by using a variety of social media platforms such as Twitter, Instagram, Facebook etc.
2. To reach a wider range of users, since some travelers may show preference to one social media platform than another.
3. To maximize the impact of their social media by driving traffic to their main corporate site.

What can make or break a social media campaign is the way that engages the public and the response rate to customer posts.

THE CONNECTED PASSENGER, SOCIAL MEDIA APPLICATIONS AND CHALLENGES OF THE TRAVEL INDUSTRY

The connected passenger is a traveler that reads and interacts on user-generated websites and other web-based applications in order to find travel-related information regarding flights, hotels and so on. The development and growth of online communities has affected the modern-day traveler (Samson, Mehta, & Chandani, 2014). Online communities continue to develop a rather extensive wealth of travel-related information, such as reviews and provide advice for fellow travelers (TripAdvisor, Yahoo Travel, Seatguru).

Connected passengers use information from user-generated websites to:

1. Acquire current travel-related information from other travelers rather than rely on the material that the respective businesses decide to make public. Travel-related information may refer to airline services, hotel amenities, specialty restaurants.
2. Connect online with other travelers. This often generates code words that may form a language. This is common among online communities.

Modern-day passengers do not passively observe what the travel industry providers chose to present (Pang, Begam Binte Abul Hassan, & Chee Yang Chong, 2014). They would rather search for information themselves and try to verify details before completing an actual travel purchase. Passengers like to be aware of what is happening behind the scenes of a travel services provider, particularly what is willingly made public and whether companies are true to their promises. The online communities available on social media offer a platform to share such information among travellers (Bernabé-Moreno, Tejeda-Lorente, Porcel, Fujita, & Herrera-Viedma, 2015).

Figure 1. Schematic perception of airline communication challenges



Travellers show resistance to mass-produced products. The connected traveller shows preference to custom-made tailored options based on his/her needs (Kefallonitis & Sackett, 2004). Tailor-made travel options give travellers the feel of being special; a feeling that some consumers desire. Passengers look for uniqueness and are motivated to purchase offerings that will offer them this unique, special feeling (Rajović, & Bulatović, 2014).

Travel-related services begin and end with people. It is a transitional process from human to human (Perry, Damian, & Lagu, 2014). Taking an airline example: All members of crew, ranging from ground to flight-crew employees personify the brand image of the airline. The overall behaviour of the employees becomes part of the airline brand. The airlines with the most awards are the ones that invest the most in their people. This is important for the successful delivery of a likeable offering that will engage the consumer and satisfy his/her needs.

Kefallonitis (2015) identified problem areas that may influence an airline's profile. These areas may be associated with language barriers, message delivery methods and cultural sensitivities. Such challenges cause passenger confusion due to the complexity of the different communication channels and media used to get the message across. Gregorie, Salle, & Tripp (2015) argued that online venues provide the foreground for customers' complaints to take place.

Travel industry communications are no longer serving the practical purpose of simply informing passengers. Engaging the connected passenger in an emotional and experiential way is equally important through the use of brand characteristics

(Aziz, Kefallonitis, & Friedman, 2014). Social media has extended the virtual presence of travel-service providers not only through websites but also through a number of online platforms and applications (Yang, 2014). Social media presence is seen as an extension of a company's virtual self and accounts for corporate image and reputation (Dijkmans et al., 2015; Sodeman & Gibson, 2015). The connected traveller shows preference to travel brands that are more realistic and meet their immediate needs (Hoffman, Novak, & Li, 2015).

Facebook, Twitter and a numerous of mobile device applications have transformed how airlines, hotels and so on engage and interact with their audience. Yet, these new avenues of communication present their own challenges (Figure 1). The airline offering is often understated, commoditized or does not meet travelers' needs. Some media platforms used to promote airlines do not highlight differential advantages but advertise a generic overview of the airline product. The above tactics make passengers think that airlines do not pay attention to their needs and seem disinterested.

Nowadays travel service providers have little or no control over the ways that passengers express their satisfaction or dissatisfaction (Schweitzer, 2014). The concept of sharing information among connected passengers has a much greater potential (Gal-Tzur et al., 2014; Mutum & Ghazali, 2014).

Customizing the entire experiential process provides an answer to personalized service. Marketing campaigns of travel industry service providers should emphasize the service characteristics that loyal passengers prefer (Akamavi, Mohamed, Pellmann, & Xu, 2015; Curras-Perez & Sanchez-Garcia, 2015). Besides a pleasant travel experience, innovation, and comfort, airlines can

greatly impact passenger satisfaction by introducing unexpected “bonuses” that delight customers (Chow, 2015).

Often little things can have a great impact on passenger experience (Gault, Corsar, Edwards, Nelson, & Cottrill, 2014). As an example, the recognition of passenger’s loyalty by offering an upgraded meal service to someone flying in economy class over a long-haul flight would be memorable. Other examples utilized are: premium quality food and snacks in cooperation with premium local food/drink providers, onboard chefs, complimentary giveaways, and specialty events such as inflight live music performances. Positively surprising a passenger with something unexpected works as a catalyst in making a passenger happy and loyal. Equally important to the above is the positive interaction among consumers. Satisfied travellers will mention positive things about an offering online via the use of social media (Dessart, Veloutsou, & Morgan-Thomas, 2015), besides sharing this with another consumer, friend, colleague. Encouraging and rewarding positive offline/online consumer-to-consumer interactions would help create loyalty (Nadeem, 2015) and would be beneficial for any travel-related service provider (Pranter & Martin, 1991).

In an age when airlines are reducing passenger services, other product and service providers are asked to bridge the gaps. Credit cards offer travel insurance benefits, priority boarding and lounge access that enhance the airline experience. Membership programs offer door-to-door baggage handling. Air travel today is less about formal relationships between passengers and specific airline brands and more about multi-connected relationships.

Other innovative examples of social media applications and tools that excite passengers and make travel brands stand out in competition follow:

- A new trend in social media, encourages the simultaneous use of a number of applications. The social media awards Skifties for travel brands have been initiated where

a multi-platform approach on integrating social media is encouraged. In 2015, Turkish Airlines employed such an approach for its advertising campaign where key influencers on social media that their audience follows contributed to advertise the airline via social media platforms. Another participating travel brand in the social media awards, the “Visit Las Vegas” social media campaign streamlined a variety of applications that included Facebook, Instagram, Pinterest, Twitter and YouTube. The campaign orchestrated and communicated several social-media-only events. The organizers carefully targeted users and encouraged them to use interactive material such as videos, pictures and infographics to further engage others and increase traffic as was the case with Turkish Airlines social media winning advertising campaign.

- A popular social media application in the USA is the United Airlines application. It enables a United Airlines passenger to make or alter a flight reservation, check frequent flyer details and get real-time updates on flights or upcoming travel. It helps the check-in for a flight, check bags and locate airport club locations. In addition it serves as the airline’s onboard entertainment device. Passenger can use the app to stream movies and TV programs directly to their handheld device upon connecting to the aircraft’s wi-fi. Last it helps users get a transfer for their destination airport to their desired location using Uber.
- **Etihad:** The option to self check-in a hold-luggage bag using an airline app (ios/android).
- **Swiss Air Lines:** The option to check your hold baggage at home and print baggage labels using your own printer.
- **EVA Air and Hello Kitty Jets:** Taiwanese hospitality with a Hello Kitty® twist. One of the most powerful co-branding concepts in the airline industry. The way EVA Air

of Taiwan have associated Hello Kitty® brand to the airline, offers the airline an advantageous position. Passengers recognize this unique association and connect the Hello Kitty® brand to EVA Air.

- **Icelandair:** The airline prides itself in creating a distinctive Icelandic flying experience. All aspects of passenger service introduce the passenger to an aspect of Icelandic culture, life and traditions. The highlight of their experience is the cabin mood lighting that resembles the northern lights. A campaign that encourages Icelandair passenger to stop in Iceland for a brief period of time went viral on social media as #mystopover

Overall, there are a number of challenges regarding the use of social media in the travel industry:

Challenges:

1. The increasing number of social media channels and applications.
2. Privacy and security concerns regarding user profiles.
3. Staying up-to-date with changing norms and educating users and audiences who follow and engage with social media platforms.

FUTURE RESEARCH DIRECTIONS

Travel brands that hold accumulated equity, built through high advertising expenditures are more likely to be perceived by consumers of superior quality (Aaker, 1996; Guihong & Yu, 2014; Hwang & Hyun, 2014), and therefore satisfy the consumer, in this case the connected-passenger (Esu & Anyadighibe, 2014). Travellers use brand names and product attributes as a source of information about product performance (Hutchinson

and Alba, 1991; Kucukusta, Heung, & Hui, 2014). Communicating these attributes through the use of connected communities has the potential of greater online exposure and word-of-mouth.

This becomes of significant importance if we consider the drop in popularity of TripAdvisor website upon the announcement of teaming up with Priceline (Lorenzetti, 2015). This shows the value that user-generated websites have over corporate-generated information. The power and value of user-generated travel websites need to be explored further.

Travel-related brands and their offerings are less about names and identities and more about people. This is achieved with a view to consumer needs and wants and adjustment to core values, offering abilities and future plans of the company. After all, consumers prefer brands that are realistic (Johnson, 2002). In addition, the environment in which interactions between travel-related providers and connected-passengers need to be such that will generate exciting and therefore satisfaction (Zhu & Chen, 2015). The link between realistic travel brands and the level of satisfaction of the connected passenger needs to be examined.

The connected online communities have the potential of meeting some of the policies and needs of the transportation (Gal-Tzur et al., 2014) and tourism industries, with regards to engaging the customer and offering up-to-date solutions or answers (Misopoulos, Mitic, Kapoulas, & Karapiperis, 2014). There is some evidence on the role of social-media and the connected communities regarding the impact on strategic planning of travel-related companies but this needs to be researched further (Venkataraman & Das, 2014).

Kim & Law (2015) call for research on the role of smartphones in travel and tourism. Smartphones can be further used by firms and tourists to communicate and share information since they have not been used to the degree that they could be employed.

CONCLUSION

The chapter illustrated that connected passengers search for travel-related information and in doing so develop a sense of belonging in social-media networking websites reinforcing the need for dialogue between airline companies and passengers. There are opportunities for the connected passenger in the travel industry available via the user-generated content and the freedom that the passenger has to create material in the online communities and platforms as has been discussed.

Travel industry brands should go beyond the purpose of reflecting a certain ideology and advantages. Travel companies' purposes are to connect to passengers by demonstrating that they are close to them and create emotional ease and comfort. The values of quality and service are expected to be present and taken for granted. Service excellence is not an element of added value any longer. Travel firms create and support the creation of an emotional world around their brand by creating and projecting lifestyles. This lifestyle is a story that customers can associate with, react to or even something they can object to.

REFERENCES

- Aaker, D. A. (1996). Measuring Brand Equity across Products and Markets. *California Management Review*, 38(2), 102–120. doi:10.2307/41165845
- About Mixi. (2015). Retrieved October 1, 2015 from <http://mixi.co.jp/en/about/>
- Adam and Taryn get Engaged on JetBlue. (2013). Retrieved October 1, 2015 from <http://vimeo.com/68789673>
- Akamavi, R. K., Mohamed, E., Pellmann, K., & Xu, Y. (2015). Key determinants of passenger loyalty in the low-cost airline business. *Tourism Management*, 46, 528–545. doi:10.1016/j.tourman.2014.07.010
- Amaral, F., Tiago, T., & Tiago, F. (2014). User-generated content: Tourists' profiles on TripAdvisor. *International Journal of Strategic Innovative Marketing*, 1(3), 137–145.
- Aziz, N., Kefallonitis, E., & Friedman, B. A. (2014). Turkey as a Destination: A Study of Sensory Brand Associations. *Tourism Culture & Communication*, 14(2), 77–89. doi:10.3727/109830414X14133839512587
- Bernabé-Moreno, J., Tejeda-Lorente, A., Porcel, C., Fujita, H., & Herrera-Viedma, E. (2015). CARESOME: A system to enrich marketing customers acquisition and retention campaigns using social media information. *Knowledge-Based Systems*, 80, 163–179. doi:10.1016/j.knosys.2014.12.033
- Bershidsky, L. (2013). *Why Twitter Is a Better Bet Than Facebook*. Retrieved October 1, 2015, Retrieved from <http://www.bloombergvie.com/articles/2013-11-08/whytwitter-is-a-better-bet-than-facebook>
- Carr, C. T., & Hayes, R. A. (2015). Social media: Defining, developing, and divining. *Atlantic Journal of Communication*, 23(1), 46–65. doi:10.1080/15456870.2015.972282
- Chow, C. K. W. (2015). On-time performance, passenger expectations and satisfaction in the Chinese airline industry. *Journal of Air Transport Management*, 47, 39–47. doi:10.1016/j.jairtraman.2015.04.003
- Curras-Perez, R., & Sanchez-Garcia, I. (2015). Antecedents and Consequences of Consumer Commitment in Traditional and low-cost Airlines. *Journal of Travel & Tourism Marketing*, 1–13.
- Del Chiappa, G. (2013). Internet versus travel agencies: The perception of different groups of Italian online buyers. *Journal of Vacation Marketing*, 19(1), 55–66. doi:10.1177/1356766712466613

- Dessart, L., Veloutsou, C., & Morgan-Thomas, A. (2015). Consumer engagement in online brand communities: A social media perspective. *Journal of Product and Brand Management*, 24(1), 28–42. doi:10.1108/JPBM-06-2014-0635
- Dijkmans, C., Kerkhof, P., & Beukeboom, C. J. (2015). A stage to engage: Social media use and corporate reputation. *Tourism Management*, 47, 58–67. doi:10.1016/j.tourman.2014.09.005
- Dotan, A., & Zaphiris, P. (2010). A cross-cultural analysis of Flickr users from Peru, Israel, Iran, Taiwan and the UK. *International Journal of Web Based Communities*, 6(3), 284–302. doi:10.1504/IJWBC.2010.033753
- Esu, B. B., & Anyadighibe, J. A. (2014). Social Media Micromarketing and Customers Satisfaction of Domestic Airlines in Nigeria. *American Journal of Tourism Research*, 3(1), 9–21. doi:10.11634/216837861403412
- Facebook's Advertising Revenue Worldwide from 2012 to 2016. (2015). Retrieved October 1, 2015 from <http://www.statista.com/statistics/271258/facebook-advertising-revenue-worldwide/>
- Gal-Tzur, A., Grant-Muller, S. M., Kuflik, T., Minkov, E., Nocera, S., & Shoor, I. (2014). The potential of social media in delivering transport policy goals. *Transport Policy*, 32, 115–123. doi:10.1016/j.tranpol.2014.01.007
- Gault, P., Corsar, D., Edwards, P., Nelson, J. D., & Cottrill, C. (2014). You'll Never Ride Alone: The Role of Social Media in Supporting the Bus Passenger Experience. In *Ethnographic Praxis in Industry Conference Proceedings* (Vol. 1, pp. 199–212). American Anthropological Association: Wiley.
- Ghedin, G. (2013). *The Story of Mixi in Japan: The Role, the Fall and the Facebook Takeover*. Retrieved October 1, 2015 from <http://www.digitalintheround.com/japan-mixi-facebook/>
- Gregorie, Y., Salle, A., & Tripp, T. (2015). Managing social media crises with your customers: The good, the bad, and the ugly. *Business Horizons*, 58(2), 173–182. doi:10.1016/j.bushor.2014.11.001
- Guihong, Z., & Yu, F. (2014). Research on evaluation system of social media's advertising effect from the air traveller's perspective. In *Service Systems and Service Management (ICSSSM), 2014 11th International Conference on* (pp. 1–6). IEEE. doi:10.1109/ICSSSM.2014.6874029
- Gunarathne, P., Huaxia, R., & Seidmann, A. (2015). Customer service on social media: the effect of customer popularity and sentiment on airline response. In *Proceedings of the 48th Hawaii International Conference on System Sciences (HICSS)* (pp. 3288–3297). Kauai, HI: IEEE. doi:10.1109/HICSS.2015.397
- Harrigan, P., Soutar, G., Choudhury, M., & Lowe, M. (2015). Modeling CRM in a social media age. *Australasian Marketing Journal*, 23(1), 27–37. doi:10.1016/j.ausmj.2014.11.001
- Hoffman, D. L., Novak, T. P., & Li, Y. (2015). *Online Consumer Behavior. In The International Encyclopedia of Digital Communication and Society*. John Wiley and Sons.
- Hofstede, G. (1980). *Culture's Consequences: International Differences in Work-Related Values*. Beverly Hills, CA: Sage.
- Holland, C., Jacobs, J., & Klein, S. (2015). An analysis of consumer search patterns in the German airline market using panel data. In I. Tussyadiah & A. Inversini (Eds.), *Information and Communication Technologies in Tourism* (pp. 85–100). Springer. doi:10.1007/978-3-319-14343-9_7
- How to Use Twitter for Your Business. (2015). Retrieved October 1, 2015 from <https://business.twitter.com/best-practices>

- Hutchinson, J. W., & Alba, J. W. (1991). Ignoring Irrelevant Information: Situational Determinants of Consumer Learning. *The Journal of Consumer Research*, 18(3), 325–345. doi:10.1086/209263
- Hutter, K., Hautz, J., Severin Dennhardt, S., & Füller, J. (2013). The impact of user interactions in social media on brand awareness and purchase intention: The case of MINI on Facebook. *Journal of Product and Brand Management*, 22(5/6), 342–351. doi:10.1108/JPBM-05-2013-0299
- Hwang, J., & Hyun, S. S. (2014). First-class airline travellers' perception of luxury goods and its effect on loyalty formation. *Current Issues in Tourism*. Retrieved October 1, 2015 from <http://www.tandfonline.com/doi/abs/10.1080/13683500.2014.918941>
- JetBlue Twitter. (2015a). Retrieved October 1, 2015 from <https://twitter.com/JetBlue>
- JetBlue Twitter. (2015b). Retrieved October 1, 2015 from https://twitter.com/JetBlue/with_replies
- Johnson, M. (2002). *Problem Solved: A Primer in Design and Communication*. London: Faidon Press Ltd.
- Kavoura, A., & Tiago, T. (2016). Understanding online communities on social networks via the notion of imagined communities: The case of TripAdvisor. *International Journal of Web Based Communities*, 12(3), 238. doi:10.1504/IJWBC.2016.077759
- Kefallonitis, E. G. (2015). Airline Brand dissatisfaction: An overview. *International Journal of Strategic Innovative Marketing*, 2(1), 1–16.
- Kefallonitis, E. G., & Sackett, P. J. (2004). Brand and product integration for consumer recognition: a review. In D. McDonagh, P. Hekkert, J. Van Erp, & D. Gyi (Eds.), *Design and Emotion* (pp. 406–411). London: Taylor and Francis.
- Kim, H. H., & Law, R. (2015). Smartphones in Tourism and Hospitality Marketing: A Literature Review. *Journal of Travel & Tourism Marketing*, 32(6), 692–711. doi:10.1080/10548408.2014.943458
- Kosarek, J. (2014). Retrieved October 1, 2015 from <http://www.socialmediaexaminer.com/how-to-network-with-influential-people-using-twitter/>
- Krogue, K. (2013). *31 Twitter Tips: How to use Twitter Tools and Twitter Best Practices for Business*. Retrieved October 1, 2015, from <http://www.forbes.com/sites/kenkrogue/2013/08/30/31-twitter-tips-how-to-use-twitter-tools-and-twitter-best-practices-for-business/>
- Kucukusta, D., Heung, V. C., & Hui, S. (2014). Deploying Self-Service Technology in Luxury Hotel Brands: Perceptions of Business Travelers. *Journal of Travel & Tourism Marketing*, 31(1), 55–70. doi:10.1080/10548408.2014.861707
- Lee, A. (2013). *Digital Marketing Town Hall. Inspired marketing advice from 15 experts*. Extract.com. Retrieved October 1, 2015, from <http://www.exacttarget.com/blog/digital-marketing-town-hall-inspired-marketing-advice-from-15-experts/>
- Lorenzetti, L. (2015). *Why TripAdvisor's Stock Is Surging 22%*. Retrieved October 15, 2015, from <http://fortune.com/2015/10/14/tripadvisor-stock-gain-priceline-deal/>
- Misopoulos, F., Mitic, M., Kapoulas, A., & Karapiperis, C. (2014). Uncovering customer service experiences with Twitter: The case of airline industry. *Management Decision*, 52(4), 705–723. doi:10.1108/MD-03-2012-0235
- Mutum, D. S., & Ghazali, E. M. (2014). Case Study 14: Air Asia: Using Social Media to Reach Out to New Customers. In *Marketing Cases from Emerging Markets* (pp. 143–147). Springer Berlin Heidelberg.

- Nadeem, M. (2015). Customer Loyalty At Any Cost: Why Is Social Media Poorly Integrated With Marketing Strategy? *International Journal of Academic Research in Business and Social Sciences*, 5(2), 78–96. doi:10.6007/IJARBSS/v5-i2/1464
- Ntalianis, K., Kavoura, A., Tomaras, P., & Drigas, A. (2015). Non-Gatekeeping on Social Networks: A Reputation Monitoring Approach. *Journal of Tourism and Services*, 10(3), 19–44.
- Palsule, M. (2011). *Top 10 best practices of social media*. Retrieved October 1, 2015 from <http://www.skepticeek.com/socialweb/top-10-best-practices-of-social-media-experts-on-twitter/>
- Pang, A., Begam Binte Abul Hassan, N., & Chee Yang Chong, A. (2014). Negotiating crisis in the social media environment: Evolution of crises online, gaining credibility offline. *Corporate Communications: An International Journal*, 19(1), 96–118. doi:10.1108/CCIJ-09-2012-0064
- Perry, L. J., Damian, O., & Lagu, A. (2014). *Understanding the Value of Social Media at Airports for Customer Engagement* (No. Project A11-03, Topic S01-08). Academic Press.
- Prahaland, C. K., & Ramaswamy, V. (2013). *The Future of Competition: Co-creating unique value with customers*. Boston: Harvard Business School Press.
- Pranter, C. A., & Martin, C. L. (1991). Compatibility Management: Roles in Service Performances. *Journal of Services Marketing*, 5(2), 43–53. doi:10.1108/08876049110035530
- Rajović, G., & Bulatović, J. (2014). Hotels and Catering Industry With Special Reference To Anthropogenic Tourist Values: The Case Northeastern Montenegro. *Tactful Management Research Journal*, 2(6), 1–11.
- Riley, C. (2014). *Banned! 8 things you won't find in China*. Retrieved October 1, 2015 from <http://money.cnn.com/gallery/technology/2014/12/30/banned-china-8/3.html>
- Sakas, D., Nasiopoulos, D., & Kavoura, A. (2015). The development of Facebook's competitive advantage for brand awareness. *Procedia. Economics and Finance*, 24, 589–597.
- Samson, R., Mehta, M., & Chandani, A. (2014). Impact of Online Digital Communication on Customer Buying Decision. *Procedia Economics and Finance*, 11, 872–880. doi:10.1016/S2212-5671(14)00251-2
- Schweitzer, L. (2014). Planning and social media: A case study of public transit and stigma on Twitter. *Journal of the American Planning Association*, 80(3), 218–238. doi:10.1080/01944363.2014.980439
- Sodeman, W. A., & Gibson, L. A. (2015). Corporate Usage of Social Media and Social Networking Services in the USA. In *Multidisciplinary Social Networks Research* (pp. 264–278). Berlin: Springer. doi:10.1007/978-3-662-48319-0_21
- United Breaks Guitars. (2009). Retrieved October 1, 2015, from <https://www.youtube.com/watch?v=5YGc4zOqozo>
- van Dijck, J. (2013). You have one identity: Performing the self on Facebook and LinkedIn. *Media Culture & Society*, 35(2), 199–235. doi:10.1177/0163443712468605
- Wattanacharoensil, W., & Schuckert, M. (2015). How Global Airports Engage Social Media Users: A Study of Facebook use and its Role in Stakeholder Communication. *Journal of Travel & Tourism Marketing*, 32(6), 656–679. doi:10.1080/10548408.2014.955245
- Xiang, Z., Magnini, V., & Fesenmaier, D. (2015). Information technology and consumer behavior in travel and tourism: Insights from travel planning using the internet. *Journal of Retailing and Consumer Services*, 22, 244–249. doi:10.1016/j.jretconser.2014.08.005

Yang, H. P. S. (2014). Case Study 16: Cathay Pacific Airways Using LinkedIn for Brand Building. In D. Mutum, S. K. Roy, & E. Kipnis (Eds.), *Marketing Cases from Emerging Markets* (pp. 159–162). Berlin: Springer. doi:10.1007/978-3-642-36861-5_21

Zhu, Y. Q., & Chen, H. G. (2015). Social media and human need satisfaction: Implications for social media marketing. *Business Horizons*, 58(3), 335–345. doi:10.1016/j.bushor.2015.01.006

ADDITIONAL READING

De Vries, N. J., & Carlson, J. (2014). Examining the drivers and brand performance implications of customer engagement with brands in the social media environment. *Journal of Brand Management*, 21(6), 495–515.

Jung, S. Y., & Yoo, K. E. (2014). Passenger airline choice behavior for domestic short-haul travel in South Korea. *Journal of Air Transport Management*, 38, 43–47.

KEY TERMS AND DEFINITIONS



Connected Passenger: A passenger who uses Internet-based technology to acquire travel-related information directly through the service provider (airline, hotel) or indirectly through online-communities (Twitter, Facebook).

Multi-Platform Social Media Advertising Campaign: Simultaneous integration of internet-based innovative communication tools that bring users together and advertise companies' communication messages.

Online Communities: Online groups of connected individuals of similar interests, ideas, beliefs and habits.

Travel Industry: This is the industry that is responsible for the transportation of people and goods from one location to another. Travel may be defined as domestic or international and vary in duration. It can also be distinguished as business, recreational, or religious. Travel providers may include airlines, rail and coach operators among others.

Evaluative Dimensions of Urban Tourism in Capital Cities by First-Time Visitors

Annamaria Silvana de Rosa

Sapienza University of Rome, Italy

Laura Dryjanska

Sapienza University of Rome, Italy

Elena Bocci

Sapienza University of Rome, Italy

INTRODUCTION

Information science and technology have a tremendous impact concerning a place, given the fact that online resources, such as websites, portals and engine searches shape the potential and future visitors' expectations. Numerous factors converge when it comes to evaluating such a complex scenario, yet most often tourists are asked to rate their experience on a scale or a series of scales. Although such data appears as fairly easy to gather and analyze, it does not reflect various dimensions of the encounter with the city, including aspects mediated by technology. In order to provide a deeper, social psychological view of the evaluative dimensions of urban tourism, the theory of social representations (Moscovici, 2000) offers a sound epistemological foundation. It has a long tradition, after its birth more than a half a century ago (Moscovici, 1961), of focusing on how people perceive places and cities in particular (Milgram, 1984). However, the theoretical construct of social representation in tourism from a geographic perspective (considering geo-cultural differences between urban locations) has not yet been fully developed, in spite of the fact that various scholars have been applying it to guide empirical research (D'Hauteserre, 2010; Monterrubio & Androtis, 2014). For example: de Rosa and colleagues for decades have worked on social representations of European capital cities and place-identity

among first-visitors from different nationalities (de Rosa, 1995, 1997, 2013; de Rosa, Bocci, & Dryjanska, in press; de Rosa, Bocci & Picone, 2013; de Rosa, & D'Ambrosio, 2011). Moscardo (2011) has concentrated on social representations of tourism planning; Lai and colleagues have applied the construct of social representations of services in context of a national park (Lai, Hsu, & Nepal, 2013), while Dickinson and Dickinson (2006) considered the social representations of transport in tourism.

The strength of the theory of social representations applied to urban tourism lies in its interdisciplinary approach that bridges sociology and psychology, based on the premise that social reality is being continuously constructed (Wagner, 1996). Technology has certainly transformed urban tourism, not only concerning online sources of information, but also when it comes to using smartphones for photographs and video-conversations that add a new dimension to interpersonal virtual communication about the city. In such a dynamic scenario, social representations undergo transformations and as a consequence their evaluative components change.

The online tourism domain integrates a number of theoretical perspectives, including: the *industry perspective*; (2) the *symbolic representation perspective*; (3) the *travel behavior perspective*; and, (4) the *travel information search perspective* (Xiang et al., 2008). Their research demonstrated

that the representations change through technological interfaces, i.e., a search engine, reflecting the idiosyncratic nature of destinations and travelers' heterogeneous information needs when considering IT applied to urban tourism (Xiang, & Gretzel, 2010).

The aim of this chapter is to provide such a multi-dimensional outlook on the evaluation, including not only the internet and the collection of data electronically, etc., but especially the impact of the interactive sources as social media, rooted in the dialogical epistemology that permeates the theory of social representations (Markova, 2000). In particular, whether reality confirms the expectations of tourists concerning a place influences the outcome of this process. In order to explore concrete examples, this chapter concentrates on how tourists evaluate historic capital cities of London, Madrid and Warsaw, emphasizing the difference between initial technologically-mediated virtual expectations and actual first-hand experiences. As demonstrated by Tussyadiah and Fesenmaier (2009, p.37): modern technology, such as videos shared on YouTube.com (and potentially other such sites) can be regarded as a means of "transportation" to destinations, and "facilitating the sharing of such experiences through videos can be viewed as a process of opening or providing access to (realistic and imaginative) tourist experiences".

BACKGROUND

Social representations have been described by the founder of the theory, Serge Moscovici, as "system(s) of values, ideas and practices" (1973, p. xiii). A group of renowned scholars in the field has proposed the following definition to further clarify the term: a social representation is the ensemble of thoughts and feelings being expressed in verbal and overt behavior of actors which constitutes an object for a social group (Wagner, Duveen, Farr, Jovchelovitch, Lorenzi-Cioldi, Marková, & Rose, 1999).

A wide range of research tools, from structured to open has been implemented in this research in order to examine the interactions between expected results and methods, taking into account diverse constructs, such as place-identity, time, space and place-memory.

Methodology

Research Participants

The data has been collected in 2011-2013 in three capital cities: London, Madrid and Warsaw from the total of 420 participants from seven different EU and extra-EU countries (France, Germany, Italy, Poland, Spain, United Kingdom and United States). Each respondent, contacted using the snowball sampling technique, was visiting the city for the first time. On a voluntary basis, the participants filled out *questionnaires in their native language*, starting from *socio-demographic information*, reason of residence in their current location and the additional reasons for visiting the city apart tourism (study, work, etc.). It is important to notice that even the first-visitors, who had additional reasons other than tourism, actually spent their time dedicated to sightseeing and other typical tourist activities, in line with the contemporary trend that urban tourism is often interwoven with other activities, such as business meetings, conferences or studies. Concerning gender, two thirds of participants are women in case of each city. According to age, the research participants have been categorized in three groups: less than 26 y.o.; 26-40 y.o.; 41- 60 y.o. The average age was 27 for London, 26 for Madrid and 34 for Warsaw.

Multi-Method Research Tools

In line with the *modeling approach*, different tools are applied to detect the social representations and their evaluative components within the multi-method questionnaire - created according

to the goals of the investigation and validated in the first wave research on Rome (de Rosa, 1995, 1997, 2013b).

In this chapter, we refer to a selection of results first based on the tools relevant for the *informative and evaluative components of social representations* are considered. On the one hand, the Likert scales from 1 to 10 have been applied to assess the ‘*self-rated importance of sources of information about the city*’, differentiating between school/university, novels and poetry, movies, songs, Internet, newspapers and magazines, tourist guidebooks, documentaries and interpersonal communication, and specifying the details. In particular, modern information science makes it possible to consult multiple types of sources through technology available on smartphones, tablets and computers. On the other hand, comparing ‘*lists of adjectives that describe the city*’ makes it possible to obtain additional evaluative dimension of social representations.

Then, we discuss the results of the “*associative network*”, a graphic-projective technique developed by de Rosa (2002) to specifically study the structure, contents and polarity of the semantic field of specific social representations. It includes as stimulus word the “*city target of the tourist destination*” with no limit in the number of associations evoked at the imagined level and with the ramification between words and connections between trees of words in the semantic space. ‘*Associative networks*’ include the stimulus words ‘London’, ‘Madrid’ or ‘Warsaw’ (corresponding to the respective capital city destination of the groups of the first-time visitors interviewed). The value of such open technique is that it allows for the visitors to include their impressions directly related to information science and technology, including images from sources such as social media, the internet and the collection of data electronically, etc.

The data analysis using descriptive statistics and multivariate techniques, in particular lexical correspondences analysis with the software package SPAD-T allows relating the different

participants’ groups identified on the basis of the *illustrative variables* with the semantic space, positioning them on the representation field determined by the *active variables* of the evoked words through the associative networks.

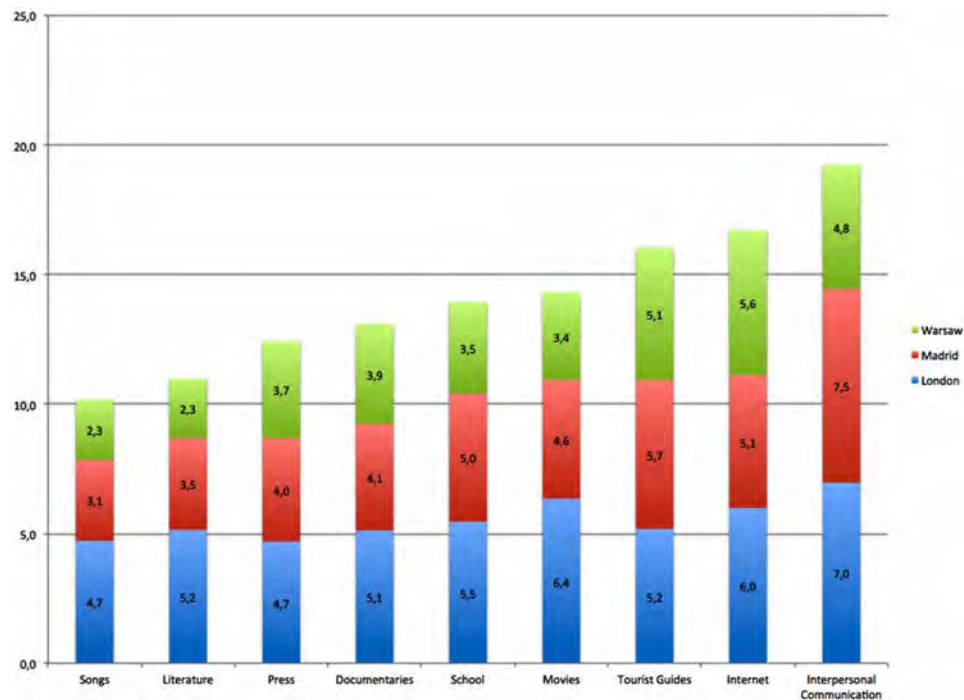
UNDERPINNING EVALUATION THROUGH SOCIAL REPRESENTATIONS

Knowledge about European capitals forms a part of cultural baggage in many parts of the world thanks to education, mass media and other forms of communication, mediated through modern technology. Paired with natural activities of gathering additional information prior to the trip, especially concerning practical details, social representations of these specific destinations appear as complex and multi-faced. For instance, a movie may provide positive impression of the city and actually induce tourism (Beeton, 2005) or, on the contrary, it may portray a place as grim and dangerous. Evaluation is influenced by such indirect perception, whether or not direct experience confirms it. In this case, Figure 1 demonstrates that on average the participants declared the highest level of knowledge about London, followed by Madrid and Warsaw.

In fact, how people perceive a place depends on their characteristics, as well as the level of information from different sources. For London and Madrid the highest importance was assigned to interpersonal communication, while for Warsaw – to Internet. The increasing relevance of electronic texts and images of tourist destinations easily found on the Internet may be treated both as a source of information for the researchers (Brunet, Bauer, De Lacy & Tshering, 2004) and as a target to be shaped in order to promote urban tourism.

The *associative networks* (de Rosa, 2002) offer insight to different attitudinal dimensions of the social representations including assessment, which is detected by the assignment of a positive, negative or neutral value to every elicited word. In this way, the evaluation of the word does not depend

Figure 1.



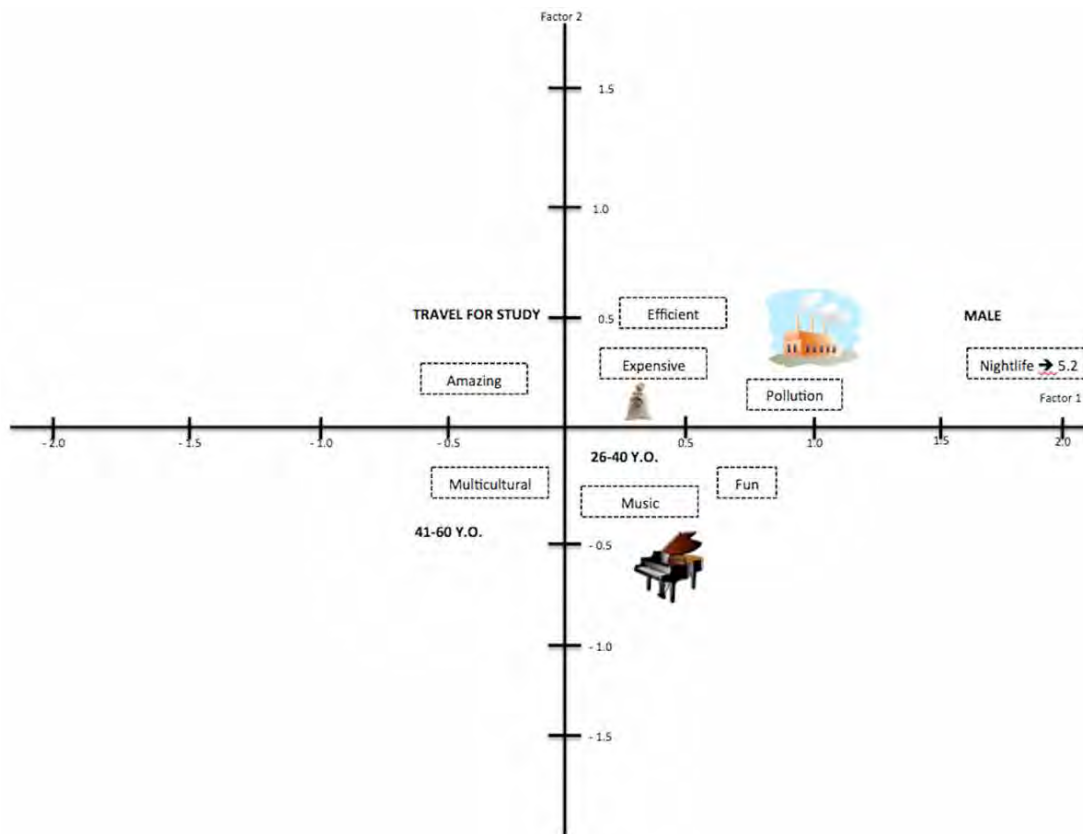
on the researcher but on participants themselves, who place a plus, minus or zero next each word. Based on averages, polarity index can be calculated per each associative network or for the city as a whole; in any case its value ranges from -1 to +1. Here the polarity index is applied in two ways: first the average values are considered for each city in order to get a general idea, and then - the polarity index embedded in the most significant factors for each city.

The average polarity indexes are: 0.502 for London, 0.599 for Madrid and 0.416 for Warsaw. The above figures do not represent a direct evaluation of the cities on a classic Likert attitudinal scale, but stem from the projective nature of associative networks detecting freely words evoked. Therefore, first-time visitors to all of the capitals predominantly elicit terms that they rate as positive with each one of the cities; while for Madrid the proportion of this positive rating is the highest, London comes on the second place and Warsaw – third. These results are due to the basic characteristics of each city that are

not subject to change, for example in realms of climate and history, as well as other factors that can be influenced in order to promote sustainable tourism. Weather conditions clearly favor Madrid, since participants rate as positive such words as ‘sun’ and ‘warm’, while they lower the polarity index of London (negative value of the word ‘rain’) and Warsaw (negative value of the word ‘cold’ and ‘freezing’). History has a strong influence on social representations of Warsaw that consist of phenomena related to World War II, destruction, tragic history of the Jews and communism, rated as negative. However, functional aspects that relate to sustainable tourism are quite frequently mentioned and, as opposed to the weather and history, may be further promoted (if positive) or changed (if negative).

Based on the associative networks, not only polarity indexes are helpful to understand how first-time visitors evaluate the city, but also the main factors offer some insight into indirect assessment, with the advantage of having the positive, negative or neutral significance assigned to each

Figure 2.



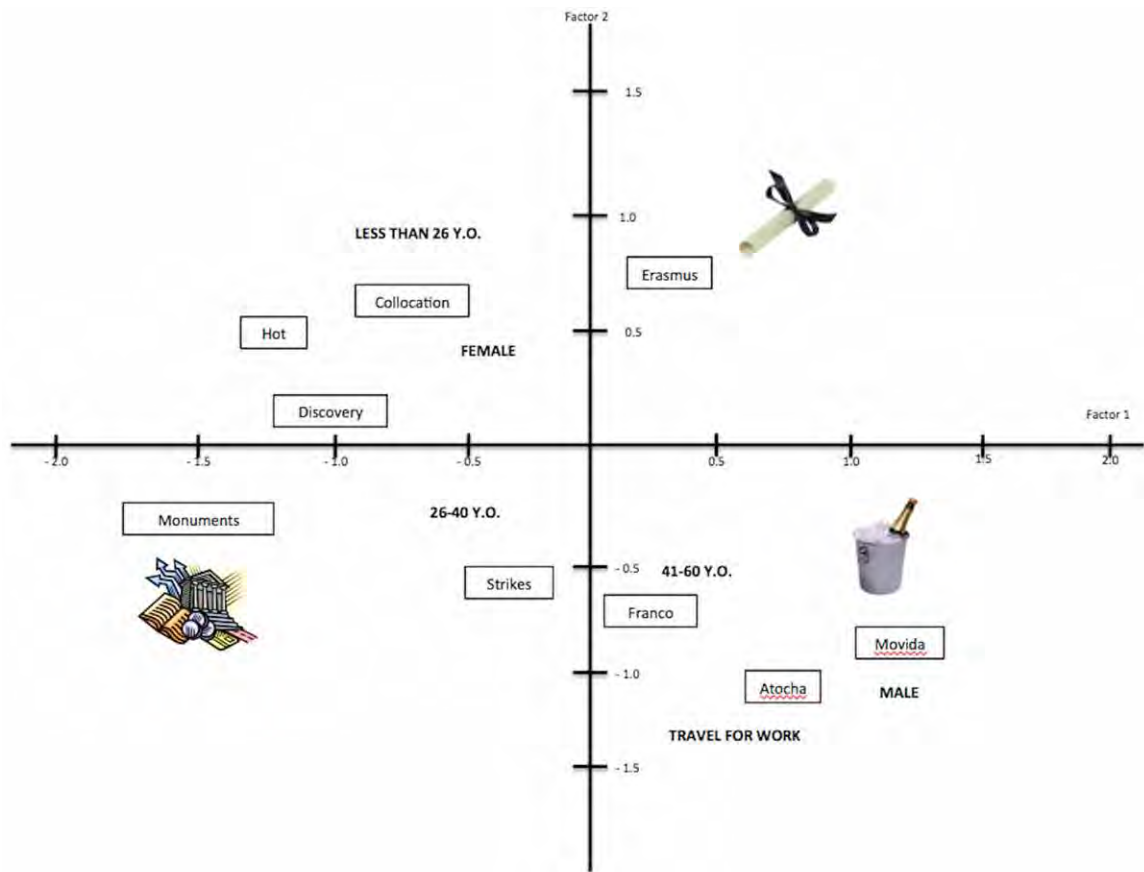
term by the participants themselves. Figures 2-4 represent the structural organization of the semantic field evoked by free associations. The factors were extracted by multi-dimensional analysis of the lexical correspondences of words evoked by the participants in the associative networks. The scale of the axis varies from 0 to 2 demonstrating different profiles of replies; Chi-square distances between profiles can be observed in ordinary physical (or Euclidean) space, by transforming the profiles before plotting, or by stretching the plotting axes by different amounts, so that a unit on each axis has a physical length inversely proportional to the square root of the corresponding element of the average profile (Greenacre, 1993).

Figures 2-4 are the graphic representations of the first most important factor and the major elements of the second one, which organize the

representational field detected via *associative network*.

The first factor is located exclusively on the positive horizontal semi-axis and consists of words 'nightlife', 'pollution', 'fun' and 'expensive'. The second factor contains numerous words on both semi-axes, to the point that makes a legible graphic representation of all of them impossible. For this reason only the four most significant words are displayed: 'efficient' and 'amazing' on positive vertical semi-axis and 'multicultural' and 'music' on negative vertical semi-axis. In social representations of London, evaluation focused on these elements plays an important role. The participants give high importance to 'pollution' in London, marking it as negative, yet typical for the city. Another outstanding characteristic of London is its economic perception as 'expensive',

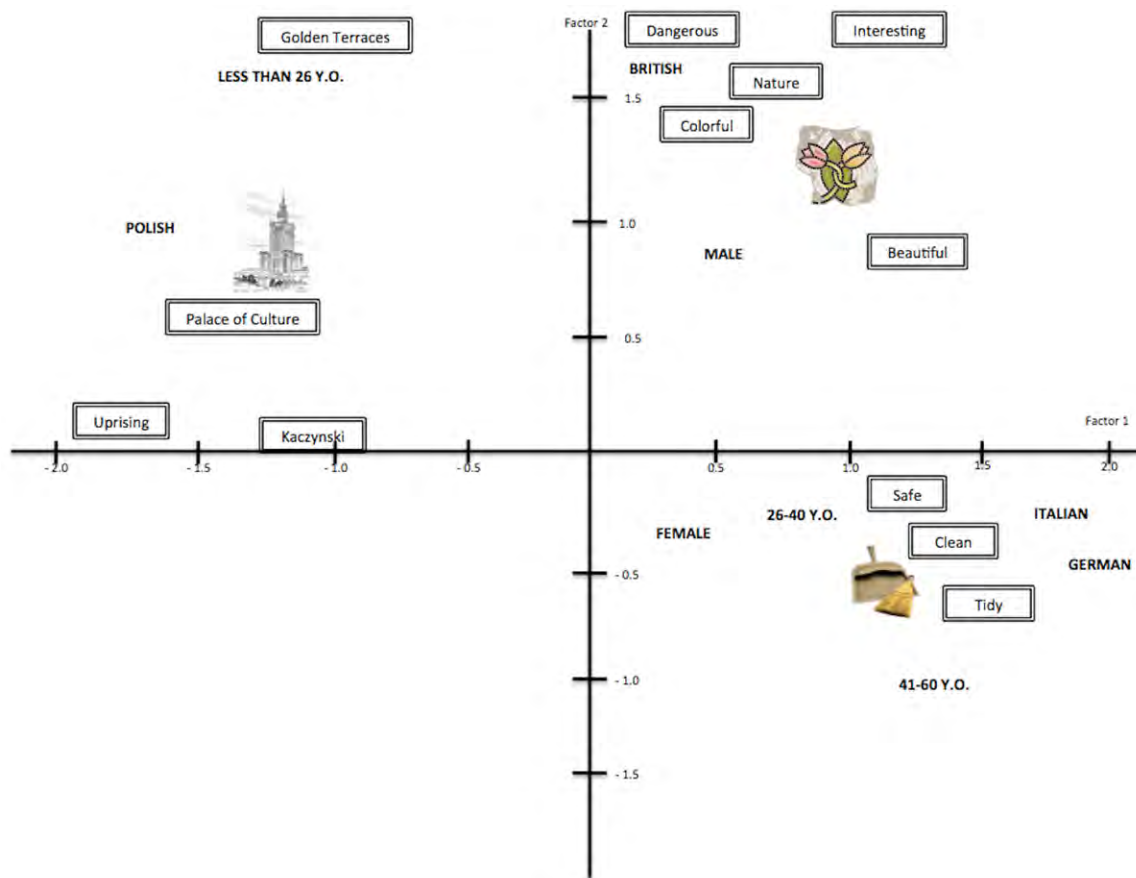
Figure 3.



seen as negative. Probably equitable access to the city seen as a common resource is made much more difficult when prices are high. On the other hand, the rest of main associations with London is rated as positive, especially in relation to two main areas, concerning respectively socio-recreational and functional dimensions: 'nightlife' and 'efficiency'. Clearly, while from the point of efficiency visitors and residents may have similar positive expectations as users of public transportation and infrastructure, nightlife too assumes a positive connotation, when occurring together with 'music' and 'fun'. Indeed it plays a significant role in case of urban tourism, as we may also observe in Figure 3 that graphically displays the first three words of each dimension of the first two most significant factors for Madrid.

The first three words of the positive semi-axis of the first factor overlap with those of the negative semi-axis of the second factor: 'Franco', 'Atocha' and 'movida'. It is interesting how each one of them actually forms a part of different aspects of reality, rated from negative through neutral to positive. Thus, evoking the social memory of dictatorship, participants assign to 'Franco' negative connotations, which is also true in case of 'strikes', considered as an efficiency problem that renders the city less visitor-friendly. Other aspects that also affect functionality, such as the main metro station 'Atocha' or Madrid's collocation in a hot climate tend to be evaluated in positive terms. Elements related to history, 'monuments' and 'discovery' receive positive assessment, appreciated by tourists. As for London with nightlife, in

Figure 4.



case of Madrid the research participants opt for ‘movida’ and ‘Erasmus’ (in line with their young age and the most common target destination for international student’s mobility). In general, both factors for Madrid appear as articulate and diverse, pointing to multiple aspects of the city’s life and history. In geographical terms, the first-visitors interviewed seem to appreciate the hot climate and collocation in Southern Europe. While visitors do not mention elements related to ecology of the city, they demonstrate knowledge of local customs and language by referring to certain phenomena directly in Spanish, which may evidence respect and friendly approach to residents.

The history of the city plays an important role also in case of Warsaw and seems familiar to the visitors, as graphically displayed in Figure 4, that shows the most significant words (‘Uprising’,

‘Kaczynski’ and ‘Palace of Culture’ – negative semi-axis; ‘safe’, ‘clean’ and ‘tidy’ – positive semi-axis) (that constitute the first factor, as well as all six words (‘dangerous’, ‘interesting’, ‘Golden Terraces’, ‘colorful’, ‘nature’ and ‘beautiful’) that constitute the entire second factor concentrated in the positive vertical semi-axis.

In fact, ‘Uprising’, ‘Kaczynski’ and ‘Palace of Culture’ all point to three different periods of history of Poland: World War II rated as negative, recent history as a democracy under President Kaczynski himself rated as positive although dead in controversially represented circumstances in Smolensk (Dryjanska, 2011) and communism rated as negative. On the other hand, the first three words in the positive semi-axis of the first factor, ‘safe’, ‘clean’ and ‘tidy’ clearly indicate functional aspects of life in the city evaluated in

positive terms. However, very high relevance of the negatively rated word 'dangerous' in the second factor demonstrates that the functionality of the city is not only seen in positive terms. 'Golden Terraces', a place with a highly functional character as a modern shopping center located at the central train station in Warsaw, tends to receive positive evaluations from first-visitors. The third major dimension identified relates more closely to environment perceived as positive, as evidenced by the presence of words 'nature', 'interesting', 'colorful' and 'beautiful'. Tourists seem to appreciate these aspects of the city, and it would be interesting to extend our study to know if these evaluations are in line with fast developing information technology and web-based preferences.

All the three graphs above presented show not only the contents and structure that organize the social representations of the three target cities, but also the different positioning of the "illustrative variables" in the representational field, that clearly shows the meaningful role played by the socio-demographic variables selected, including gender, age, nationalities and reason for the visit experience. An analytical reading of these results would allow more detailed interpretation of the anchoring of social representations into specific social groups. Here we limit to underline very briefly that males elicit more often socio-recreational places, like 'nightlife' in London and 'movida' in Madrid, while females emphasize functional characteristics of the city (for example Warsaw as 'tidy', 'clean' and 'safe'); domestic first-visitors (Yang & Wall, 2009) mention more precisely buildings significant for the history of their own country (for example the Polish visitors in Warsaw), while first-visitors from other countries are more attracted by functional dimensions and natural environment (for example British mention 'nature' in Warsaw, while Italian and German focus on the city's evaluation as 'tidy', 'clean' and 'safe').

In spite of the similarity, a change occurs in representations after the visit. The major change

that occurred in the social representation of Warsaw is interesting since in case of this city modern information and technology channels, such as Internet, has come to play the most important role. For example, among modern sources of information, travel blogs (Schmallegger, & Carson, 2008) virtual communities, wikis, social networks, collaborative tagging, and media files shared on sites like YouTube and Flickr (Gretzel, 2006) have started to change approached to information exchange about tourist destinations.

The top frequent adjectives used to evaluate the three different European cities and their historic centers tend to be quite similar. It is also interesting that especially for London and Madrid the visitors use adjectives that could apply to virtually any European capital, which enables us to easily compare them. The adjective with a clear positive evaluative component is 'beautiful', and in case of Madrid it has the highest frequency, especially after the visit. The data demonstrates that the experience of visiting the city tends to enrich social representations by adding to their complexity and in general has a positive influence on evaluation. Warsaw tends to be perceived in the most negative way, as demonstrated by the presence of such adjectives as 'sad' and 'dirty', although in some cases in contrast with opposite evaluations, like 'lively' and 'clean'.

DISCUSSION

Overall, the evaluation of the same city by the same person sometimes gives different results depending on the method and its structured versus projective nature. Whether a person is asked to use a scale, an adjective or a word to which he or she assigns a positive, neutral or negative value does matter, as evidenced by the 'modeling approach' to the study of social representations, which integrates different methods and techniques coherently with the articulation of different theoretical constructs (like in this study place-identity, attitudes, prior

knowledge and channels of information, etc.) and dimensions (informational, evaluative, emotional, imaginal/experiential, etc.).

Different dimensions of evaluation of the capital cities under scrutiny are taken into account by asking the first-visitors to evoke adjectives and by calculating polarity indexes of the semantic space projected in *associative networks* (de Rosa, 2002, 2003, 2005).

The evaluation of a city by first-time visitors is based on the interplay of expectations (the social representations of the city before the visit) and actual experience. Thus, the “city’s evaluation” after the visit already encompasses “prior knowledge”, rated on Likert scales from 1 to 10, for each source of information: school/university, novels and poetry, movies, songs, Internet, newspapers and magazines, tourist guidebooks, documentaries and interpersonal communication. It may also be expressed in a different way by freely using adjectives to describe and represent the city. Following the direct experience, associative networks envision social representations of the city, also providing information about its polarity index (to which degree it is seen in a positive, neutral or negative way) and the main factors that reveal the organization of its semantic space. Such integrative framework may be applied to analyze the evaluative aspects of urban tourism in a more profound, structured way, taking into account different aspects. It also allows the researchers to compare how first-time visitors assess different cities or how the evaluation of a given city changes over time.

To sum up, different preferences for top places in London, Madrid and Warsaw may be observed demonstrated by first-visitors from seven distinct nationalities, outlined in the Figure 5, in line with the cultural differences and styles of relation between tourists and their preferred places already detected in the first-wave research on Rome (de Rosa, 1997, 2013b).

Certain elements of the proposed integrative framework are already well known and applied in research on urban tourism, as for example the

Figure 5.



differentiation between pre- and post-visit images of the city (Ben-Dalia, Collins-Kreiner & Churchman, 2013). However, the strength of the modeling approach lies in its capacity to integrate in an organized rationalized and justified way various methods coherently with the multiple theoretical constructs that social representation pretends to articulate (compared to the fragmentation of the micro-paradigms in the mainstream social psychology tradition: Moscovici, 2000). A deep understanding of how visitors evaluate a city can be helpful when trying to form new models or attempting to predict future patterns, based on socially constructed understandings of risk and adventure in the context of urban tourism (Beedle, 2005).

CONCLUSION

The theory of social representations offers a holistic view of socially constructed reality, acknowledging that the perception of a destination is dynamic. Therefore, periodic monitoring and adjustment of decisions appear necessary, also in order to be aware of new niche markets, for example based on performing arts (Lim & Bendle, 2012) or interest in communist past of

a city. Eventual evaluation is a two-way street, based on expectations and previous knowledge, as well as the actual experience. Therefore, there is a clear need to intervene both during city-branding, when shaping the image of a city for example through its website or other available materials, and by investing in the infrastructure, services and preparation of local population. One has to be careful not to limit the decisions to the supply side of urban tourism that takes visitors for granted and tries to modify their behavior once at the destinations (Dolnicar, 2006). Exploring the evaluative dimension of social representations of cities helps to gain insight into visitors' expectations and preconceived ideas in the light of fast developing information technology.

Overall, the theory of social representations has much to offer in order to enrich the underlying socio-psychological perspective of research on urban tourism in a new light of technology.

FUTURE RESEARCH DIRECTIONS

Future research directions shall emphasize even more technology and new media as vital for urban tourism (see the additional readings below cited), integrating and comparing several interrelated research lines, as we have tried to develop combining the field study and the new media studies on *Place @-branding and European Capitals*, including analysis of 'city visiting cards' via municipal websites, comparison of preferred itineraries of past visitors with virtual tours by potential first-visitors flying with Google Earth, analysis of conversational exchanges about city-places experienced/imagined via social networks, institutional stems and online commercial logos of the cities (de Rosa, 2013b; de Rosa, Bocci & Picone, 2013; de Rosa & Bocci, 2014). Other possibilities that we are currently developing look at the urban destinations' evaluations via YouTube and travel forums such as Trip Advisor or looking at augmented reality tourist experiences.

These kind of the studies, theoretically-driven by the *modeling approach* to Social Representations, also taking into account the whole phases of the tourist life-cycle (pre-trip, on-site and after-trip-phase), may be in future research directions integrated with the new trend of data-driven studies, based on *big-data business intelligence* and *knowledge management* in the society of algorithms and machine-generated data also including the related critical questions (boyd & Crawford, 2012; Parks, 2014) and those on the impact of mobile *augmented reality* interfaces and user-centered designs on decision-making based on fast retrieval of information by tourists exposed to unfamiliar environments (Tussyadiah & Zach, 2012).

REFERENCES

- Beedle, P. (2005). The Adventure of Urban Tourism. *Journal of Travel & Tourism Marketing*, 18(3), 37–48. doi:10.1300/J073v18n03_04
- Beeton, S. (2005). *Film-Induced Tourism*. Clevendon, UK: Channel View Publications.
- Ben-Daliaa, S., Collins-Kreinerb, N., & Churchmana, A. (2013). Evaluation of an Urban Tourism Destination. *Tourism Geographies*, 15(2), 233–249. doi:10.1080/14616688.2011.647329
- Boyd, D., & Crawford, K. (2012). Critical Questions for big data. *Information Communication and Society*, 15(5), 662–679. doi:10.1080/1369118X.2012.678878
- Brunet, S., Bauer, J., De Lacy, T., & Tshering, K. (2001). Tourism Development in Bhutan: Tensions between Tradition and Modernity. *Journal of Sustainable Tourism*, 9(3), 243–263. doi:10.1080/09669580108667401
- de Rosa, A. S. (1994). From theory to meta-theory in S.R.: The lines of argument of a theoretical-methodological debate. *Social Sciences Information. Information Sur les Sciences Sociales*, 33(2), 273–304. doi:10.1177/053901894033002008

- de Rosa, A. S. (1995). Psicologia del turismo: per una psicologia sociale dell'ambiente applicata. In B. Zani (Ed.), *Le Dimensioni della Psicologia Sociale* (pp. 153–197). Bologna: Il Mulino.
- de Rosa, A. S. (1997). Turisti di sei nazionalità per la prima volta nella 'città eterna': 'place-identity' e rappresentazioni sociali di Roma e del suo centro storico. In A. Nenci (Ed.), *Conoscere e Rappresentare la Città* (pp. 149–214). Padua: Cedam.
- de Rosa, A. S. (2002). The 'associative network': A technique for detecting structure, contents, polarity and stereotyping indexes of the semantic fields. *European Review of Applied Psychology*, 52, 181–200.
- de Rosa, A. S. (2013a). *Social Representations in the 'Social Arena'*. London: Routledge.
- de Rosa, A. S. (2013b). Place-identity and social representations of historic capital cities: Rome through the eyes of first-visitors from six countries. In A. S. de Rosa (Ed.), *Social Representations in the "social arena"* (pp. 311–381). New York: Routledge.
- de Rosa, A.S. (2014). The role of the Iconic-Imaginary dimensions in the Modelling Approach to Social Representations. *Papers on Social Representations*, 23, 17.1-17.27.
- de Rosa, A. S., & Bocci, E. (2014). Place @-Branding and European Capitals: 'city visiting cards' via municipal websites, virtual tours of significant places flying with Google Earth, and conversational exchanges about city-places experienced/imagined via social networks. In A. Kapoor & C. Kulshrestha (Eds.), *Dynamics of Competitive Advantage and Consumer Perception in Social Marketing* (pp. 126–168). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-4430-4.ch006
- de Rosa, A. S., Bocci, E., & Dryjanska, L. (In press). Paris-paradis ou Paris-enfer? Les deux faces de la ville souterraine et de surface entre passé et présent, avec un regard vers le futur. In M. Dargentas & M. de Alba (Eds.), *Penser et agir dans les espaces et dans les villes*. Rennes, France: Presse Universitaire Rennes.
- de Rosa, A. S., Bocci, E., & Picone, M. (2013). E-branding and institutional websites: the 'visiting card' of the municipalities of Rome and Paris. In A. Kapoor & C. Kulshrestha (Eds.), *Branding and Sustainable Competitive Advantage: Building Virtual Presence* (pp. 207–247). Hershey, PA: IGI Global.
- de Rosa, A. S., & D'Ambrosio, M. (2011). *Universi semantici tra luoghi immaginati e luoghi esperti: first visitor italiani in sei capitali europee. Turismo e Psicologia*. Padua: Padova University Press.
- D'Hauterres, A. (2010). Representations of rurality: Is Foxwoods Casino Resort threatening the quality of life in southeastern Connecticut? *Tourism Geographies*, 3(4), 405–429. doi:10.1080/146166800110070496
- Dickinson, J. E., & Dickinson, J. A. (2006). Local Transport and Social Representations: Challenging the Assumptions for Sustainable Tourism. *Journal of Sustainable Tourism*, 14(2), 192–208. doi:10.1080/09669580608669052
- Dolnicar, S. (2006). Nature-conserving tourists: The need for a broader perspective. *Anatolia*, 17(2), 235–256. doi:10.1080/13032917.2006.9687188
- Greenacre, M. (1993). *Correspondence analysis in practice*. London: Academic Press.
- Gretzel, U. (2006). Consumer generated content—trends and implications for branding. *E-review of Tourism Research*, 4(3), 9–11.

- Lai, P.-H., Hsu, Y.-S., & Nepal, S. K. (2013). Representing the landscape of Yushan National Park. *Annals of Tourism Research*, 43, 37–57. doi:10.1016/j.annals.2013.03.004
- Lim, C. C., & Bendle, L. J. (2012). Arts tourism in Seoul: Tourist-orientated performing arts as a sustainable niche market. *Journal of Sustainable Tourism*, 20(5), 667–682. doi:10.1080/09669582.2011.636817
- Liu, Z. (2003). Sustainable Tourism Development: A Critique. *Journal of Sustainable Tourism*, 11(6), 459–475. doi:10.1080/09669580308667216
- Marková, I. (2000). Amédée or how to get rid of it: Social representations from a dialogical perspective. *Culture and Psychology*, 6, 419–460. doi:10.1177/1354067X0064002
- Milgram, S. (1984). Cities as social representations. In R. Farr & S. Moscovici (Eds.), *Social representations* (pp. 289–309). London: Cambridge University Press.
- Monterrubio, J. C., & Andriotis, K. (2014). Social representations and community attitudes towards spring breakers. *Tourism Geographies*, 16(2), 288–302. doi:10.1080/14616688.2014.889208
- Moscardo, G. (2011). Exploring social representations of tourism planning: Issues for governance. *Journal of Sustainable Tourism*, 19(4-5), 423–436. doi:10.1080/09669582.2011.558625
- Moscovici, S. (1961). *La psychanalyse, son image et son public*. Paris: Presses Universitaires de France.
- Moscovici, S. (1973). Introduction. In C. Herzlich (Ed.), *Health and illness. A social psychological analysis*. London: Academic Press.
- Moscovici, S. (2000). *Social representations*. Cambridge, UK: Polity.
- Parks, M. R. (2014). Bid Data in Communication Research: Its Contents and Discontents. *Journal of Communication*, 64(2), 355–360. doi:10.1111/jcom.12090
- Proshansky, H. M. (1978). The city and self-identity. *Environment and Behavior*, 10(2), 147–169. doi:10.1177/0013916578102002
- Schmallegger, D., & Carson, D. (2008). Blogs in tourism: Changing approaches to information exchange. *Journal of Vacation Marketing*, 14(2), 99–110. doi:10.1177/1356766707087519
- Tussyadiah, I. P., & Fesenmaier, D. R. (2009). Mediating tourist experiences: Access to places via shared videos. *Annals of Tourism Research*, 36(1), 24–40. doi:10.1016/j.annals.2008.10.001
- Tussyadiah, I., & Zach, F. (2012). The role of geo-based technology in place experiences. *Annals of Tourism Research*, 39(2), 789–800. doi:10.1016/j.annals.2011.10.003
- Wagner, W. (1996). Queries about social representation and construction. *Journal for the Theory of Social Behaviour*, 26(2), 95–120. doi:10.1111/j.1468-5914.1996.tb00524.x
- Wagner, W., Duveen, G., Farr, R., Jovchelovitch, S., Lorenzi-Cioldi, F., Marková, I., & Rose, D. (1999). Theory and Method of Social Representations. *Asian Journal of Social Psychology*, 2(1), 95–125. doi:10.1111/1467-839X.00028
- Xiang, Z., & Gretzel, U. (2010). Role of social media in online travel information search. *Tourism Management*, 31(2), 179–188. doi:10.1016/j.tourman.2009.02.016
- Xiang, Z., Wöber, K., & Fesenmaier, D. R. (2008). Representation of the online tourism domain in search engines. *Journal of Travel Research*, 47(2), 137–150. doi:10.1177/0047287508321193

Yang, L., & Wall, G. (2009). Authenticity in ethnic tourism: Domestic tourists perspectives. *Current Issues in Tourism*, 12(3), 235–254. doi:10.1080/13683500802406880

ADDITIONAL READING

Dippelreiter, B., Grün, C., Pöttler, M., Seidel, I., Berger, H., Dittenbach, M., & Pesenhofer, A. (2008). Online tourism communities on the path to Web 2.0: An evaluation. *Information Technology & Tourism*, 10(4), 329–353. doi:10.3727/109830508788403132

Palmer, A., & McCole, P. (2000). The role of electronic commerce in creating virtual tourism destination marketing organisations. *International Journal of Contemporary Hospitality Management*, 12(3), 198–204. doi:10.1108/09596110010320760

Wang, Y., Yu, Q., & Fesenmaier, D. R. (2002). Defining the virtual tourist community: Implications for tourism marketing. *Tourism Management*, 23(4), 407–417. doi:10.1016/S0261-5177(01)00093-0

Yovcheva, Z., Buhalis, D., Gatzidis, C., & van Elzakker, C. P. (2014). Empirical evaluation of smartphone augmented reality browsers in an urban tourism destination context. *International Journal of Mobile Human Computer Interaction*, 6(2), 10–31. doi:10.4018/ijmhci.2014040102

KEY TERMS AND DEFINITIONS

European Capitals: Capital cities in Europe, containing an area that can be identified as a historic center (sometimes called “Old Town”, like in the case of Warsaw).

First-Time Visitors: People who visit a given place (for example a city) for the first time in their lives (sometimes also called first-time visitors).

Geo-Cultural Context: Specific geographic location that takes into account the cultural characteristics of inhabitants, such as common language, history, customs or art; for example, insular, Mediterranean or North-European geo-cultural contexts.

Modeling Approach: A paradigmatic approach to social representations developed by A.S. de Rosa (2013a, 2014) that integrates different methods and techniques coherently with the articulation of different theoretical constructs and dimensions.

Place-Identity: A construct created by H.M. Proshansky (1978) that links both spatial and temporal dimensions of identity, reflecting social aspects of built and natural environment.

Social Representations: A complex construct created by Serge Moscovici (1961), which – according to de Rosa (1994) can be operationalized on three different levels: a) social representations as *phenomenon*; b) *theory* of social representations; c) *meta-theory* of social representations.

Urban Tourism: A form of tourism aimed at visiting a city while traveling for recreational purposes.

Fifty Shades of Dark Stories

H**Lea Kuznik***University of Maribor, Slovenia*

INTRODUCTION

Dark tourism is a special type of tourism, which involves visits to tourist attractions and destinations that are associated with death, suffering, disasters and tragedies venues.

Visiting dark tourist destinations in the world is the phenomenon of the twenty-first century, but also has a very long heritage. Number of visitors of war areas, scenes of accidents, tragedies, disasters, places connected with ghosts, paranormal activities, witches and witchhunt trials, cursed places, is rising steeply. Reasons and motives for the visit are varied. Dark tourism has been recognised as a distinctive tourism phenomenon of the twenty-first century, with increasingly significant numbers of visitors and tourists going to dark tourism attractions and sites, new dark tourism products and attractions emerging, and modern global communication media generating interest in dark tourism attractions, while at the same time affecting the image of destinations. The phenomenon of dark tourism has been examined in academia from the mid-1990s. Since then, study of this phenomenon has increased, and the scales of relevant studies have been enlarged.

Dark tourism in Slovenia is very poorly developed comparing to the world and it is mostly limited only on tourist sites connected with both wars. Therefore the theme is a novelty in Slovenia, as well as in Slovenian professional and scientific literature where is almost unknown.

The main purpose of this article is to explore the current situation of dark tourism and propose a typology of dark tourism in the world and in Slovenia, which should serve as a basis for further efforts in the design of a new dark tourism products based on the dark heritage in Slovenia

and other countries, as shown below in the case of witchcraft.

The research is based on in-depth analysis of literature and fieldwork in Slovenia which contains the method of unstructured interviews with curators in Slovene Ethnographic Museum, Museum of Ribnica and Celje Regional Museum as well as the method of observation with participation in a guided tour through the exhibition at Ljubljana Castle: The Barbarism of Torture.

BACKGROUND

The term dark tourism was coined by Foley and Lennon (1996: 198) to describe the attraction of visitors to tourism sites associated with death, disaster, and depravity. Other notable definitions of dark tourism include the act of travel to sites associated with death, suffering and the seemingly macabre (Stone, 2006: 146), and as visitations to places where tragedies or historically noteworthy death has occurred and that continue to impact our lives (Tarlow, 2005: 48). Scholars have further developed and applied alternative terminology in dealing with such travel and visitation, including thanatourism (Seaton, 1996), black spot tourism (Rojek, 1993), atrocity heritage tourism (Tunbridge & Ashworth, 1996), and morbid tourism (Blom, 2000). In a context similar to »dark tourism«, terms like »macabre tourism«, »tourism of mourning« and »dark heritage tourism« are also in use. Among these terms, dark tourism remains the most widely applied in academic research (Sharpley, 2009).

The concept of dark tourism is in contrast to marketing slogans that prefer the broader promotional aspect and call this type of tourism »historic

DOI: 10.4018/978-1-5225-2255-3.ch353

tourism«. Major encyclopedias of tourism identify »dark tourism« also as »thanatourism«, in which the core meaning of the term relates mostly to visits to the tombs, cemeteries and memorials of prominent people (Gosar, 2015a).

Although this is a newer type of specialized tourism, researchers can speak as one of the oldest types of tourism, because death is historically always attracted human inquisitiveness. Some kind of organized »thanato tourism« were already gladiator games in the Coliseum of ancient Rome (Gosar, 2015b). Popular festivals in the past have been a public hanging, beheading and burning of witches. Walking and paid visits to the battlefield at Waterloo in Belgium, the place of Napoleon's last battle between the English nobility had been ongoing since the time of the battle in 1815. Therefore the kind of dark tourism has a very long heritage.

Dark tourism relates to tourist travel, which interprets the heritage through tragedies and conflicts and is raising awareness of dark historical realities, or the heritage of it (Stone, 2013). The central research centre for dark tourism is located at the University of Central Lancashire, in England. The Institute for Dark Tourism Research (iDTR) is led by Dr. Philip Stone. According to iDTR, the main contours of dark tourism destinations are to be found in three groups of geographically expressed areas: destination of the death, burial, and/or the tragedies of celebrities, destinations of great battles and falling soldiers, destinations of collective suffering and death.

Visiting such sites can play a significant part in a tourist's experiences, and in turn, that there will most probably be anxiety about the development of these sites as tourist attractions (Ashworth & Hartmann, 2005; Ryan, 2007; Sharpley & Stone, 2009; Wilson, 2008). These concomitant aspects of dark tourism have indeed lead to concerns about the morality of commodifying death, disaster, and atrocity (Lennon & Foley, 2000; Seaton, 2009).

All the actions associated with the tourism trips that expose/define the places associated

with death, suffering and/or everything that is reminiscent of the grim period of mankind is to be related dark tourism (Stone, 2006). According to researchers of iDTR, dark tourism is a subcategory of the historic tourism, which includes the content of the material and intangible heritage, as both strengthen our historical memory.

Different Shades of Dark

Dark tourism has been also called place-specific tourism (Ashworth & Hartmann, 2005: 4). As such, several researchers have classified dark tourism sites according to their defining characteristics.

Miles (2002) proposed a darker-lighter tourism paradigm in which there remains a distinction between dark and darker tourism according to the greater or lesser extent of the macabre and the morose. In this way, the sites of the holocaust, for example, can be divided into dark and darker tourism when it comes to their authenticity and scope of interpretation. The United States Holocaust Memorial Museum in Washington D.C. is associated with death, and thus categorised a dark tourism site only, whereas the site of Auschwitz - Birkenau in Poland possesses a unique location authenticity as a former concentration and extermination camp, and thus site of darker tourism.

On the basis of the dark tourism paradigm of Miles (2002), Stone (2006) proposed a spectrum of dark tourism supply which classifies sites according to their perceived features, and from these, the degree or shade of darkness (darkest to lightest) with which they can be characterised. This spectrum has seven types of dark tourism suppliers, ranging from Dark Fun Factories as the lightest, to Dark Camps of Genocide as the darkest. A specific example of the lightest suppliers would be dungeon attractions, such as London Dungeon, or planned ventures such as Dracula Park in Romania. In contrast, examples of the darkest sites include genocide sites in Rwanda, Cambodia, or Kosovo, as well as holocaust sites such as Auschwitz - Birkenau.

REASONS FOR VISITING DARK TOURISM SITES

There are a number of reasons for travelling to dark tourism sites, which most simply can include curiosity, education, survivor guilt, remembrance, nostalgia, empathy, and horror (Ashworth & Hartmann, 2005; Baldwin & Sharpley, 2009; Garwood, 1996; Lennon & Foley, 2000; Miles, 2002; Smith, 1996). Each of these can be discussed separately as follows:

1. Curiosity

Many tourists are interested in the unusual and the unique, whether this be a natural phenomenon (e.g. Niagara Falls), an artistic or historical structure (e.g. the pyramids in Egypt), or spectacular events (e.g. a royal wedding). Importantly, the reasons why tourists are attracted to dark tourism sites derive, at least in part, from the same curiosity which motivates a visit to Niagara Falls. Visiting dark tourism sites is an out of the ordinary experience, and thus attractive for its uniqueness and as a means of satisfying human curiosity. So the main reason is the experience of the unusual.

2. Empathy

One of the reasons for visiting dark tourism sites may be empathy, which is an acceptable way of expressing a fascination with horror. Ashworth and Hartmann (2005) note that empathy relies upon the capacity of heritage consumers to identify with individual victims of the atrocity. While this identification is assumed to be more with the victims in question, it could equally conceivably be with perpetrators also. In many respects, the interpretation of dark tourism sites can be difficult and sensitive, given the message of the site as forwarded by exhibition curators can at times conflict with the understandings of visitors.

3. Horror

Horror is regarded as one of the key reasons for visiting dark tourism sites, and in particular, sites of atrocity. Ashworth and Hartmann (2005) note that there is a considerable amount of literature, folk stories, and more recently internet, film and television portrayals of scenes of horror which evoke emotions of fear and fascination in consumers. Relating atrocity as heritage at a site is thus as entertaining as any media depiction of a story, and for precisely the same reasons and with the same moral overtones. Such tourism products or examples are: Ghost Walks around sites of execution or murder (Ghost Tour of Prague), Murder Trails found in many cities like Jack the Ripper in London.

4. Education

In much tourism literature it has been claimed that one of the main motivations for travel is the gaining of knowledge, and the quest for authentic experiences. One of the core missions of cultural and heritage tourism in particular is to provide educational opportunities to visitors through guided tours and interpretation. Similarly, individual visits to dark tourism sites to gain knowledge, understanding, and educational opportunities, continue to have intrinsic educational value (Ashworth & Hartmann, 2005). Moreover, a number of sites emphasise the visitors educational expectations in terms of their capacity to learn from past mistakes, for example, the United States Holocaust Memorial Museum in Washington DC. In turn, many dark tourism attractions or sites are considered important destinations for school educational field trips, achieving education through experiential learning (Marcuse, 2005).

5. Nostalgia

Nostalgia can be broadly described as yearning for the past (Dann & Potter, 2001; Smith, 1996), or

as a wistful mood that an object, a scene, a smell or a strain of music evokes (Belk, 1990: 670). In addition however, it has also been recognised as a reason for travelling to dark tourism sites, although not perhaps a key or central motivation. In this respect Smith (1996) examined war tourism sites and concluded that old soldiers do go back to the battlefields, to revisit and remember the days of their youth.

6. Remembrance

Remembrance is a vital human activity connecting us to our past (Young, 1993). Remembrance helps people formulate an identity, allowing them to learn from past mistakes, and to go forward with a clear vision of the future. In the context of dark tourism, remembrance and memory are considered key elements in the importance of sites (Lennon & Foley, 2000; Walter, 2009; Young, 1993). For several sites associated with the holocaust and the Second World War in European cities, commemoration and remembrance are key reasons for their existence.

7. Survivor's guilt

One of the distinctive characteristics of dark tourism is the type of visitors such sites attract, which include survivors and victim's families returning to the scene of death or disaster. These types of visitors are particularly prevalent at sites associated with Second World War and the holocaust. For many survivors returning to the scene of death and atrocity can achieve a therapeutic effect by resolving grief, and can build understanding of how terrible things came to have happened. This can be very emotional experience (Braithwaite & Lee, 2006).

Such reasons for visiting dark tourism sites or attractions can in turn influence visitors on site experiences.

TYOLOGY OF DARK TOURISM

When analyzing a wide variety of dark tourism in the world, depending on the content researchers can give the following typology of dark tourism which can be divided into two levels:

1. Dark places in nature:
 - a. Disaster area tourism (visiting places of natural disaster after e.g. hurricane Katrina in New Orleans, tsunami in Thailand, earthquake in Nepal, sites of volcanic destruction e.g. Pompeii, Montserrat in the Caribbean...).
2. Dark places in connection with people:
 - a. Grave tourism (visiting famous cemeteries, graves of famous individuals, or grand mausoleums of some real cult personality, e.g. Paris Pere Lachaise Cemetery with Jim Morrison and Edith Piaf graves, Taj Mahal in Agra, Sedlec Ossuary in the Czech Republic, Pierce Brothers Westwood Village Memorial Park in Los Angeles with graves of Marilyn Monroe, Roy Orbison and other Hollywood stars).
 - b. War or battlefield tourism (visiting former war places e.g. Napoleon's battle of Waterloo in Belgium, USS Arizona Memorial in Hawaii, Cu Chi Tunnels in Vietnam, the Bridge over the River Kwai in Thailand...).
 - c. Holocaust tourism (visiting concentration camp, memorial sites, memorial museums, former ghettos and sites where the Nazi perpetrators planned it all e.g. The House of the Wannase Conference and Führerbunker in Berlin, United States Holocaust Museum in Washington D.C., Yad Vashem in Jeruzalem, Auschwitz in Poland...).

- d. Genocide tourism (visiting places of genocide e.g. Rwanda genocide, Toul Sleng Genocide Museum and The Killing Fields in Cambodia, the Srebrenica mass killing in Bosnia and Herzegovina...).
- e. Prison tourism (Alcatraz Island in the San Francisco Bay, Canon City in Colorado, Ottawa Jail Hostel, Devil's Island in French Guiana...).
- f. Communism tourism (visiting North Korea, the big four: Lenin, Mao, Ho Chi Minh and Kim).
- g. Cold war and iron curtain tourism (seeking out traces and remains of the Berlin Wall).
- h. Nuclear tourism (visiting sites of civil nuclear disaster e.g. Chernobyl in Ukraine, Fukushima in Japan, sites of nuclear testing e.g. Semipalatinsk in Kazakhstan and Nevada Test Site in USA, or missile silos e.g. Titan Missile Museum in Arizona, and also two places where atom bombs were actually used for real: Hiroshima and Nagasaki in Japan).
- i. Murderers and murderous places tourism (Jack the Ripper in London, Lee Harvey Oswald in Dallas – Visitors can see the sixth floor of the building from which he fired at US president John F. Kennedy – and the Dakota, apartment building in New York City where John Lennon was shot).
- j. Slum tourism (visiting impoverished areas e.g. slums in India, Brazil, Kenya...).
- k. Terrorist tourism (visiting areas e.g. Ground Zero in New York City, The Boston Marathon, the Bardo Museum in Tunis and tourist resort at Sousse in Tunisia...).
- l. Paranormal tourism (visiting crop circles, UFO sightings, the haunted house in Amityville, paranormal activities in Stanley Hotel in Colorado...).
- m. Witched tourism (visiting the city of Salem in Massachusetts...).
- n. Accident tourism (visiting places e.g. Paris tunnel Pont de l'Alma, where the British Princess Diana and Dodi Fayed died in a car accident).
- o. Icky medical tourism (visiting e.g. Josephinum, medical museum with anatomical wax models in Vienna, the Mütter Museum in Philadelphia, Meguro Parasitological Museum in Tokyo and also all the "Bodies Exhibitions").
- p. Dark amusement tourism (visiting amusement parks e.g. the Dungeons exhibitions in London, Dracula theme park, ghost tours in London, Prague, Chicago...).

H

Typology of Dark Tourism in Slovenia

Dark heritage is also the basis for the development of dark tourism in Slovenia. Military cemeteries and ossuaries, monuments and museums, the theater battles execution and solemn memorial events at the anniversaries are an integral part of the European cultural landscape and society. In Slovenia dark tourism destinations include:

- Cemeteries (Zale Cemetery in Ljubljana, Cemetery Pobrezje in Maribor, Roman Necropolis in Sempeter in Savinja Valley).
- Prisons and penitentiaries (Penitentiary at the Ljubljana Castle, Celica Jail Hostel).
- Museums (The Hospital Franja, Park of Military History in Pivka, Kobarid Museum, Idrija War Museum, War Museum in Logatec).
- Concentration camps (Ljubelj, Mauthausen).
- Memorial sites and monuments (Cerje Monument, Monument for the Battle of Drazgose, Monument to Pohorje Partisan Battalion in Osankarica, Teharje Memorial

- Park, The Memorial Church of the Holy Spirit in Javorca, Russian Chapel at Vrsic)
- Fortification systems (Vallo Alpino, Rupnik`s Line, Fortress Kluze).
 - Guided tours of military facilities (Rupnik`s Line, The Hospital Franja).
 - Remembrance paths – (The Walk of Peace, from Alps to the Adriatic, Circular Path of Military History in municipality of Pivka).
 - Shows (The show about First World War at Fortress Kluze performed by Dreizehn Dreizehn Society – 1313. It is a different way of discovering the everyday life of the soldiers on the Isonzo front. Visitors can see tooth extraction, reading the long-awaited love letter, fear, nursing wounded. Every soldier at the front experienced all that, irrespective of their nationality).
 - Performing battles (Rupnik`s Line battle in municipality of Ziri, »Liberation of Primorska 1945« battle in the Park of Military History in municipality of Pivka, both performed by Triglav Cultural and Historical Society).
 - Hiking along the trails of war memories (Memorial march »Along the Trail of the Cankar Battalion«, Along the Barbwire of the Occupied City in Ljubljana).
 - Post-war killings (Kren Cave mass grave, Kocevski Rog mass grave, The Barbara Pit).

The analysis shows that dark tourism in Slovenia is mainly associated with topics related to the First World War and the Second World War period, comprising conventional museum presentations, visiting places “in situ”, performances and thematic trails in conjunction with the two world wars.

ALONG THE TRAILS OF SLOVENIAN “WITCHES”

Beside heritage of war Slovenia can offer different opportunities for the development of dark tourism

on the basis of dark cultural heritage. One of the topics is witchcraft which has a long heritage in Slovenia and it is not yet included in tourism. Women accused of witchcraft were usually burned at the stake, like in four stories of Slovene “witches” described below. All of them could be effectively incorporated into a comprehensive and innovative tourism product.

1. The tragic love story of Veronika of Desenice and Frederick II., Count of Celje.

The first process under the accusation of witchcraft in Slovenia took place in 1427 in Celje against Veronika of Desenice. She was the second wife of Frederick II., Count of Celje. Little is known of her early life. It is believed the name Deseniska derives from the village of Desinic in Croatia, where Frederick also had extensive estates. Veronika was minor nobility and Frederick’s father Hermann II. greatly opposed to the marriage. The chronicles of the Counts of Celje suggest he had his son arrested and, while holding him prisoner, initiated a trial against Veronika accusing her of witchcraft. She was acquitted by the court, but despite this incarcerated in Ojstrica Castle near Tabor and murdered by being drowned. She was buried in Braslovce and a few years later Frederick arranged for her remains to be reburied at the Carthusian monastery at Jurkloster. Her grave was discovered in 2005.

2. The story of Anica at Ljubljana Castle: The Barbarism of Torture - an exhibition of torture devices from the 16th to the 18th century.

The exhibition offers a more detailed understanding of the course of criminal procedures of the time, from jail and interrogation to the execution of the sentence, with the aid of costumed characters – the Ljubljana executioner Hans and his victim Anica, who has been accused of witchcraft in a story that is based on an actual witchcraft trial that unfolded in Ljubljana at the end of the 17th century.

The exhibition of early modern torture devices, tools and requisites, which were used both to achieve a confession of guilt by means of torture and to administer punishment for minor and serious offences, is intended to shed light on the time of the proliferation of torture practices and cruel punishments, while also serving as a reminder that even today these practices are nowhere near being merely a residue of the distant past. Is today's supposedly civilised society any different at all?

3. The tragic love story of Agatha and Friderich Herberstein, count of Hrastovec Castle.

In the 16th century, a young count from Hrastovec, Friedrich Herberstein, fell in love with beautiful Agatha from the Stralek manor. Friedrich's mother opposed to their marriage, therefore the young couple got secretly married in the chapel of Stipler. So, Friedrich's mother accused Agatha of witchcraft. While Friedrich was at war, his mother tortured Agatha and made her throw a new-born baby into the burning stove. Then, she reported the fact to the judge who condemned Agatha to death. Her step-mother demanded that Agatha be beheaded. When Friedrich returned from the war, he could not find his beloved wife. In his grief, he planted a black cross, a symbol of unhappy love, in the ground where the tragedy happened. Today the black cross can be found in Lormanje in Lenart municipality. The story is described in the book "The Black Cross", by Ozbalt Ilaunig.

4. The last one on the stake: Marina Susarek in Ribnica.

The Process of Ribnica was a typical witch process, with the usual accusations - from flying on the Sabbath and making contract with the Devil, which was eagerly advocated by the judge, to accusations of causing damage by witchcraft, partly reported by the people from the village. It is also characteristic that it was a group process in which the accused were forced to inform against

their colleagues, which resulted in new victims. The process was held for at least two years, from 1700 to 1701, but it might have started earlier. The Process of Ribnica resulted in at least seven victims beside Marina Susarek (Kosir, 2001).

Marina Susarek was forty years old, mother of six children, married to the shoemaker. She was accused of witchcraft and put on the witchcraft chair to confess her guilt. After three hours of torture she admitted to be a witch. She was sentenced to death. They cut off her head with a sword. Her body was thrown on the stake (Kosir, 2001). Records of her trial are kept in the Archive of the Republic of Slovenia.

FUTURE RESEARCH DIRECTIONS

Number of visitors of places of disasters, tragedies and scenes of murder, witchcraft and cursed places is rising sharply. Such places visitors attended previously in history, but this were not called dark tourism.

Irresistibly attractive places of accidents were identified in the tourism industry as a promising market niche. Human inquisitiveness has become a source of income of the organizers of tourist trips, also locals are satisfied who earn some money because of their past disaster.

Dark tourism is in fact often linked to processes of selected memorialization of past violent events, and to a related production of specific geographies of memory. The question is to what extent it is in fact acceptable to market a tragic event, who should deal with destination images created by the media, to what extent tourist and visitor expectations created by the media can in fact be met, and how managers and curators of dark tourism sites can effectively communicate the message of the site to tourists and visitors.

Debates should also be focused on the modalities of the visits, on the technologies of display implemented to remember and visualize the past, on the ethical issues related to leisure activities in sensitive locations.

Dark tourism is the type of tourism that should have highly educational role, while provoking endless discussions on how the difficult past could and should be presented to the visitors.

CONCLUSION

Dark tourism is a growing phenomenon internationally that has gathered significant attention on the part of the academic literature in the past decade. Including forms of cultural and historical commemoration, but also visits to sites of horror and violence, dark tourism is becoming an important source of income for some destinations.

The main motive for visiting all those dark memorials is to understand the causes and consequences of various events in order what to do that would never happen again, so find out what happened and understand why this has occurred. Strong motivation for visiting is also to commemorate the victims of a particular historically important area.

Also in Slovenia there are many dark stories (not only dedicated to war) and themes that can be developed into dark tourist products, as shown in the case of witches and witchcraft, which was prepared by the atmosphere of fear, envy and resentment. Usually an unfortunate happening, like hail, or illness, supposedly provoked by witches, or accusations among neighbours, initiated the charges of witchcraft. On most occasions people were not able to explain the reasons for disasters, thus witchcraft was a handy explanation. If prayers wouldn't help, nor amulets, nor church bells, only the stake remained.

The witch processes are classified among the dark phenomena in the history of mankind. These processes were nothing else but legal murders, the accused were burned for the actions which they could not have possibly done. In spite of numerous researches a satisfactory and final explanation for these wide-spread events has not been found. A reasonable explanation would be indeed quite hard to conceive. The idea of »evil«, which the

world is trying to eradicate by »fire«, is an old one, constantly reoccurring, and not limited to the past. The timeless conspiracy theory can be easily recognized in it, and whoever dares oppose is in danger of being quite quickly classified as one of the alleged conspirators. Thus the witch processes point out the possible connections with the contemporary history, and open the questions regarding the impulses within ourselves.

REFERENCES

- Ashworth, G., & Hartmann, R. (2005). Introduction: managing atrocity for tourism. In G. Ashworth & R. Hartmann (Eds.), *Horror and human tragedy revisited: the management of sites of atrocities for tourism* (pp. 1–14). Sydney: Cognizant Communication Corporation.
- Baldwin, F., & Sharpley, R. (2009). Battlefield tourism: bringing organised violence. In R. Sharpley & P. R. Stone (Eds.), *The darker side of travel* (pp. 186–206). Bristol: Channel View.
- Belk, R. W. (1990). The role of possessions in constructing and maintaining a sense of past. *Advances in Consumer Research. Association for Consumer Research (U. S.)*, 17(1), 669–676.
- Blom, T. (2000). Morbid tourism - a postmodern market niche with an example from Althorp. *Norwegian Journal of Geography*, 54(1), 29–36.
- Braithwaite, D., & Lee, Y. L. (2006). *Dark tourism, hate and reconciliation: The Sandakan experience*. International Institute for Peace through Tourism.
- Dann, G. M. S., & Potter, R. B. (2001). Supplanting the planters: Hawking heritage in Barbados. *International Journal of Hospitality & Tourism Administration*, 2(3/4), 51–84. doi:10.1300/J149v02n03_03
- Foley, M., & Lennon, J. (1996). JFK and dark tourism: A fascination with assassination. *International Journal of Heritage Studies*, 2(4), 198–211. doi:10.1080/13527259608722175

- Garwood, A. (1996). The holocaust and the power of powerlessness: Survivor guilt an unhealed wound. *British Journal of Psychotherapy*, 13(2), 243–258. doi:10.1111/j.1752-0118.1996.tb00880.x
- Gosar, A. (2015a). The concept of dark tourism. In A. Gosar, M. Koderman, & M. Rodela (Eds.), *Post - WWI Destinations of Human Tragedies and Opportunities for Tourism Development*. University of Primorska.
- Gosar, A. (2015b). Dark tourism and or tanato tourism destinations of our historical memories. In A. Gosar, M. Koderman, & M. Rodela (Eds.), *Post – WWI Destinations of Human Tragedies and Opportunities for Tourism Development*. University of Primorska.
- Ilaunig, O. (1988). *The Black Cross*. Lenart: Klopotec.
- Kosir, M. (2001). Bloody hell against the witches. In F. Jaklic (Ed.), *The Last One on the Stake. Celje: Mohorjeva družba*.
- Lennon, J., & Foley, M. (2000). *Dark Tourism*. London: Continuum.
- Marcuse, H. (2005). Reshaping Dachau for visitors: 1933-2000. In G. Ashworth & R. Hartmann (Eds.), *Horror and human tragedy revisited: the management of sites of atrocities for tourism* (pp. 118–148). Sydney: Cognizant Communication Corporation.
- Miles, W. F. S. (2002). Auschwitz: Museum interpretation and darker tourism. *Annals of Tourism Research*, 29(4), 1175–1178. doi:10.1016/S0160-7383(02)00054-3
- Rojek, C. (1993). *Ways of escape*. Basingstoke, UK: Macmillan. doi:10.1057/9780230373402
- Ryan, C. (Ed.). (2007). *Battle tourism: history, place and interpretation*. Amsterdam: Elsevier Butterworth-Heinemann.
- Seaton, A. V. (1996). From thanatopsis to thanatourism: Guided by the dark. *International Journal of Heritage Studies*, 2(4), 234–244. doi:10.1080/13527259608722178
- Seaton, T. (2009). Purposeful Otherness: approaches to the management of thanatourism. In R. Sharpley & P. R. Stone (Eds.), *The darker side of travel* (pp. 75–108). Bristol: Channel View.
- Sharpley, R., & Stone, P. R. (Eds.). (2009). *The darker side of travel: the theory and practice of dark tourism*. Bristol: Channel View.
- Smith, V. L. (1996). War and its tourist attractions. In A. Pizam & Y. Mansfeld (Eds.), *Tourism, crime and international security issues* (pp. 247–264). New York: John Wiley & Sons.
- Stone, P. (2013). Dark Tourism, Heterotopias and Post-Apocalyptic Places: The Case of Chernobyl. In L. White & E. Frew (Eds.), *Dark Tourism and Place Identity* (pp. 29–46). Melbourne: Routledge.
- Stone, P. R. (2006). A dark tourism spectrum: Towards a typology of death and macabre related tourist sites, attractions and exhibitions. *Tourism*, 54(2), 145–160.
- Tarlow, P. E. (2005). Dark tourism: the appealing dark side of tourism and more. In M. Novelli (Ed.), *Niche tourism - contemporary issues, trends and cases* (pp. 47–58). Oxford, UK: Butterworth – Heinemann. doi:10.1016/B978-0-7506-6133-1.50012-3
- Tunbridge, J. E., & Ashworth, G. (1996). *Dissonant heritage: the management of the past as a resource in conflict*. New York: John Wiley & Sons.
- Walter, T. (2009). Dark tourism: mediating between the dead and the living. In R. Sharpley & P. R. Stone (Eds.), *The darker side of travel* (pp. 39–55). Bristol: Channel View.
- Wilson, J. Z. (2008). *Prison: cultural memory and dark tourism*. New York: Peter Lang.

Young, J. E. (1993). *The texture of memory: Holocaust memorials and meaning*. New Haven, CT: Yale University Press.

ADDITIONAL READING

Buda, D. M., & McIntosh, A. J. (2013). Dark tourism and voyeurism: Tourist arrested for »spying« in Iran. *International Journal of Culture, Tourism and Hospitality Research*, 7(3), 214–226.

Causevic, S., & Lynch, P. (2011). Phoenix tourism: Post-conflict tourism role. *Annals of Tourism Research*, 38(3), 780–800. doi:10.1016/j.annals.2010.12.004

Dann, G., & Seaton, A. V. (2008). Slavery, contested heritage and thanatourism. *International Journal of Hospitality & Tourism Administration*, 2(3), 1–29.

Frew, E., & White, L. (Eds.). (2011). *Tourism and National Identity: An International Perspective*. London, New York: Routledge.

Hartmann, R. (2014). Dark tourism, thanatourism, and dissonance in heritage tourism management: New directions in contemporary tourism research. *Journal of Heritage Tourism*, 9(2), 166–182. doi:10.1080/1743873X.2013.807266

Hughes, R. (2008). Dutiful tourism: Encountering the Cambodian genocide. *Asia Pacific Viewpoint*, 49(3), 318–330. doi:10.1111/j.1467-8373.2008.00380.x

Korstanje, M. (2011). Detaching the elementary forms of dark-tourism. *Anatolia: An International Journal of Tourism and Hospitality Research*, 22(3), 424–427. doi:10.1080/13032917.2011.620800

Korstanje, M., & Clayton, A. (2012). Tourism and terrorism: Conflicts and commonalities. *World-wide Hospitality and Tourism Themes*, 4(1), 8–25. doi:10.1108/17554211211198552

Miles, S. (2014). Battlefield sites as dark tourism attractions: An analysis of experience. *Journal of Heritage Tourism*, 9(2), 134–147. doi:10.1080/1743873X.2013.871017

Mowatt, R., & Chancellor, C. (2011). Visiting death and life: Dark tourism and slave castles. *Annals of Tourism Research*, 38(4), 1410–1434. doi:10.1016/j.annals.2011.03.012

Podoshen, J. (2013). Dark tourism motivations: Simulation, emotional contagion and topographic comparison. *Tourism Management*, 35, 263–271. doi:10.1016/j.tourman.2012.08.002

Rami, I. (2009). Alternative Tourism: Can the segregation wall in Bethlehem be a tourist attraction? *Tourism and Hospitality Planning & Development*, 6(3), 247–254. doi:10.1080/14790530903363381

Robinson, P., Heitmann, S., & Dieke, P. (Eds.). (2011). *Research Themes for Tourism*. London: CABI.

Sharpley, R., & Stone, P. R. (Eds.). (2010). *Tourist Experience: Contemporary Perspectives*. London, New York: Routledge.

Stone, P. R. (2011). Dark tourism and the cadaveric carnival: Mediating life and death narratives at Gunther von Hagens Body Worlds. *Current Issues in Tourism*, 14(7), 685–701. doi:10.1080/13683500.2011.563839

Stone, P. R. (2012). Dark tourism and significant other death: Towards a model of mortality mediation. *Annals of Tourism Research*, 39(3), 1565–1587. doi:10.1016/j.annals.2012.04.007

Stone, P. R. (2012). Dark tourism as »mortality capital«: the case of Ground Zero and the significant other dead. In R. Sharpley & P. R. Stone (Eds.), *Contemporary Tourist Experience: Concepts and Consequences* (pp. 71–94). Abington, UK: Routledge.

KEY TERMS AND DEFINITIONS

Cultural Heritage: The whole complex of distinctive material, spiritual and social achievements from the past that characterize a society or national identity.

Cultural Heritage Tourism: Visits by persons to experience, learn and appreciate the cultural legacy of people.

Dark Cultural Heritage: Cultural heritage that is associated with real and commodified sites of atrocity, death, disaster, human depravity, tragedy, human suffering, and sites of barbarism and genocide.

Dark Tourism: Special type of tourism, which involves visits to tourist attractions and destinations that are associated with death, suffering, disasters and tragedies venues.

Interpretation: All planned activities that communicate the importance and significance of what is seen and experienced by visitors.

Thanatourism: Travel to a location wholly, or partially, motivated by the desire for actual or symbolic encounters with death.

Tourism Product: Any product that is marketed by a country or an institution to attract the visitors and experience the said product.

H

Social Media as a Channel of Constructive Dialogue for Tourism Businesses

Marios D. Sotiriadis

University of South Africa (UNISA), South Africa

INTRODUCTION

Information and communication technologies (ICTs) are having a major impact on tourism in general. Ongoing developments in ICTs have significant changes in the way tourism-related information is distributed and on consumer behavior. The major impacts of the “digital revolution” on tourism have come through Social Media (SM) and mobile devices. According to the latest estimations the worldwide audience of SM has surpassed two billion users in 2015 (eMarketer, 2015). SM allow tourists to interact and share their views and experiences with potentially unrestricted virtual communities (Gretzel & Yoo, 2013; Leung et al., 2013; Sigala et al., 2012). This approach to meeting and communicating with people globally has revolutionized communications and consequently the marketing of tourism businesses (Munar & Jacobsen, 2014).

SM or Web 2.0 platforms include social networks (e.g. Facebook), blogs (e.g. Blogger), microblogs (e.g. Twitter), wikis (e.g. Wikipedia), content sharing platforms (e.g. AssociateContent.com) and text/audio/video sharing platforms (e.g. Flickr and YouTube). SM are presenting a host of new challenges as well as opportunities for tourism providers. One of their main functions is to establish an interactive channel of communication, which is mutually beneficial to both parties involved: it offers a medium for tourists to express their requirements, and gives tourism providers a tool to acquire customer feedback (Bilgihan et al., 2016). Whilst this represents a real challenge for tourism businesses, it also opens new channels of communication. It is therefore

imperative for tourism businesses to consider how SM is shaping business-to-consumer marketing communications and how may exploit their full potential (Benckendorff et al., 2014; Law et al., 2014; Sotiriadis & Van Zyl, 2013).

Within this context, the present article takes a marketing perspective to analyze the potential contribution of SM. The main aim is to examine and suggest the ways in which tourism businesses could take advantage of SM as a channel of constructive dialogue with customers. In sum, the paper suggests ways to take full advantage of SM in order to enhance listening to, and understanding of, their current and potential customers.

BACKGROUND AND LITERATURE REVIEW

The Internet has become the first-choice place for consumers to search for information on tourism destinations, as well as being the most important venue for marketing of tourism services (Benckendorff et al., 2014; Gretzel & Yoo, 2013; Law et al., 2014). The challenges and opportunities for tourism-related industries that arise from the digital environment are obvious in everyday business practice (Leung et al., 2013; Sigala et al., 2012).

During the last two decades scholars have shown an increasing interest in the growing role of SM in the tourism field (Law et al., 2014; Zeng & Gerritsen, 2014). The dominant topic investigated is the effectiveness of different digital marketing strategies and tools including SM (Law et al., 2014). The reason for this is that SM play a significant role in many aspects of tourism, especially

in decision-making behavior, tourism promotion and interaction with consumers (Bilgihan et al., 2016; Gretzel & Yoo, 2013; Law et al., 2014; Zeng & Gerritsen, 2014).

User-generated content (UGC) is one of the main activities of tourists on SM. They share their experiences, views and evaluations through online reviews. Examples of this behavior include posts, wikis, discussion forums, chats, tweets, podcasting, pins, and sharing of digital images, video and audio files (Munar & Jacobsen, 2014). The issue of implications of UGC has been examined by Ayeh et al. (2013), Boley et al. (2013), Wilson et al. (2012), Xiang and Gretzel (2010), and Ye et al. (2011). The topic of factors motivating tourists' involvement in SM was examined by Bronner and de Hoog (2011), and Munar and Jacobsen (2014). The study by Bilgihan et al (2016) explored the factors motivating consumers to share tourism information via SM. The study results show that the main factors of intention to share knowledge are belief in integrity and perceived ease of use. The issue of influence of SM on tourist behavior was examined by Parra-López et al. (2011) and Sotiriadis and Van Zyl (2013).

The present study is focusing on the impact of SM in the field of tourism business marketing and management. As argued by scholars, the changes of tourist behavior influence the approaches and tools that tourism businesses have to adopt and use in managing and marketing their offerings in the digital environment (Leung et al., 2013; Sigala, 2012; Sigala et al., 2012). Kim et al. (2015) suggested that online reviews in SM should be managed as a critical part of hotel marketing. Furthermore, tourism businesses make use of SM for various purposes, such as engagement, commitment, relationship building, and managing their reputation and brand (Dijkmans et al., 2015). The reason for the success of SM in business originates from the possibility to manage relationships with customers in a customized way (Calefato et al., 2015).

Literature also suggested that SM are an effective marketing channel that can be wisely used in integrated marketing communications of tourism services (see, for instance, Oz, 2015; Sotiriadis & Van Zyl, 2013). Additionally, SM can be used by tourism businesses to support and enhance knowledge management (KM) activities, as suggested by Sigala (2011), and Sigala and Chalkiti (2014). The mass collaboration and communication functionalities of the SM can support and enhance conversational, participatory and collaborative KM processes; i.e. searching, sharing and creating information and intelligence. Examples of internal KM processes include searching, storing and collecting information, and practices of organisational learning and internal social bonding. External KM processes include sharing, discussing and creating knowledge with others (Sigala & Chalkiti, 2014). Finally, the study by Xiang et al. (2015) pointed out the tremendous growth of SM and UGC on the Internet and demonstrated the utility of big data analytics to better understand important hospitality business issues.

The above literature review indicates that the field of SM is a challenging topic. As can be seen, the crucial issues are to adopt the appropriate approach and to use SM in the field of tourism efficiently in order to exploit the full potential of this social phenomenon. Savvy tourists are now 'unbundling' the whole planning, booking and sharing of their experiences (Morrison, 2013). SM are one of the tools at their disposal and they are taking full advantage of it.

The interaction of the main challenges in the tourism market with the digital environment and the impressive adoption and use of SM clearly indicates that tourism providers need to adopt new approaches in the field of communications, such as developing a multi-channel approach to communications, embracing the influence of SM, and engaging customers in a mutually beneficial dialogue. This article argues that SM could have a stronger contribution in the fields of service experience improvement and innovation in tourism.

SOCIAL MEDIA AND TOURISM SERVICES

The present article suggests that tourism providers have to examine the ways in which tourists are using SM in order to better communicate and interact with them. SM have become important social platforms for online communications giving people the opportunity to find and then communicate to others with similar interests (Ayeh et al., 2013). SM are powerful platforms that allow consumers to collaborate and contribute to developing, extending, rating, and commenting on tourism experiences. They have been recognized as innovative knowledge sharing networks by enabling consumers to connect, share, and interact with others (Oz, 2015). That is the reason why SM, as knowledge sharing platforms, are gaining attention for the tourism industry.

SM sharing and adoption is growing globally, according to a report by eMarketer, the number of SM users across the globe exceeds 2 billion users and more than 28% of people worldwide used SM regularly in 2015 (eMarketer, 2015). Business reports indicate that SM are well established as key sources of information for tourists before, during and after their trips. SM now play a major role in travel planning for many travelers around the world. These estimations are confirmed by the findings of Tripbarometer survey conducted by Ipsos and TripAdvisor (TripAdvisor, 2014). As for the reviews, it seems that customer feedback plays a key role in the travel cycle: 89% of global tourists say reviews are influential when choosing where to book. Global tourists' decisions of where to travel are based 65% on online travel reviews and 48% on personal recommendations.

The Global Trends Report (WTM, 2015) also highlighted the emerging trends in the global tourism industry. The key findings of this report include the following: (i) Mobile technology is currently transforming the tourism landscape in terms of bookings, customer service and consumer behavior; mobile bookings are expected to reach 35% of online tourism bookings by 2018;

and (ii) Destination services, personalization, mobile bookings and peer-to-peer platforms and services are expected to be the main disruptive forces in the tourism industry over the next five years. These developments represent a challenge for tourism providers which will need to build a flexible technological architecture to follow their customers from one device to another.

Given the fast-growing SM usage trend worldwide, tourism providers are embracing SM more than ever as part of their marketing strategies. SM strategies of tourism providers have been robust in recent years, and nearly all major brands and global chains are now present on SM platforms. However, those companies that have succeeded the most on SM are the ones using content drawn from their customers' and potential customers' behavior online (WTM, 2015). Within the context of a fragmenting marketplace, the crucial point is to make SM more successful in keeping people engaged (Bilgihan et al., 2016; Law et al., 2014).

Let us consider the example of Twitter that is one of the best-known SM and micro-blogs in Europe and America. This SM is a real-time information network, powered by people in nearly every country in the world in six languages, and is continually expanding (Twitter, 2015). Twitter is now one of the world's most popular micro-blogs and functions as a source of entertainment and leisure (Morrison, 2013). Many hospitality businesses are using Twitter to build their brand and connect more personally with their customers (Chan & Guillet, 2011). Some recent facts and figures to note about Twitter are (Bullas, 2015): In 2015 Twitter has 284 million active users at last count, 88% of Twitter users are on mobile, and they post 500 million tweets per day. The information that spreads through Twitter can help tourists make better choices and decisions.

A study by Sotiriadis and Van Zyl (2013) explored the use of online reviews by Twitter users and followers in their decision-making behavior about tourism services purchases and the factors influencing the use of information retrieved from this micro-blog. The study findings indicate that

three factors are very influential regarding the use of information about tourism services retrieved from Twitter, namely: (i) credibility of Twitter followers/users; (ii) degree of involvement/posting; and (iii) expertise and know-how of Twitter users/followers. This knowledge is very useful in tourism marketing. Tourism businesses are increasingly adopting SM marketing, which consists mainly of exploring opportunities provided by SM and implementing actions to respond positively to online users/members' reviews, suggestions and experience sharing. These valuable sources of feedback provide information that enhances customer relationships, determining their needs and requirements and contributes to improving service quality (Sotiriadis & Van Zyl, 2015).

It is believed that the true value of SM in general lies in the ability to really listen and respond to, to efficiently interact with customers. SM give tourism marketers access to a targeted audience and relevant conversations. How can SM be used by tourism providers? The next section deals with this question.

SUGGESTIONS AND RECOMMENDATIONS: SOCIAL MEDIA AS A CHANNEL OF CONSTRUCTIVE DIALOGUE BETWEEN TOURISM BUSINESSES AND CUSTOMERS

Most tourism businesses are now heavily engaged in using SM platforms to distribute information and communicate with others. The major activities of tourism providers in using SM include collecting UGC, displaying photography and videos, distributing topical news stories, emphasizing current events, encouraging electronic Word-of-Mouth (eWOM) recommendations, and getting feedback (Morrison, 2013).

This article focuses on the appropriate approach to, and adequate use of SM by tourism providers based on two key concepts, namely engagement and interaction/dialogue. Tourism businesses

should approach SM as a means of interactive communication with customers and a tool for customer feedback. Customer feedback is "the transmission of negative information (complaints) or positive information (compliments) to providers about the services used" (Saha & Theingi, 2009, p. 354). Hjalager and Nordin (2011) argue that the active exchange of ideas and resources with customers may facilitate and benefit the learning process, enhance innovation capacities and ultimately increase the competitiveness of companies. Customers could influence tourism providers to improve the experience service quality and to affect or improve a company's innovation performance. Hence, it is suggested that these are the two fields in which SM might make a contribution: (i) experience quality/performance adjustments and (ii) service innovation.

1. SM: a Medium of Feedback for Performance Adjustments and Improvement

The practice of listening to and learning from the customers when delivering a service is extensively used in tourism but not in a systematic way (Sigala et al., 2012). Posts and comments on SM provide input that can be useful for tourism providers in identifying areas in which adjustments of performance are required, product innovation might be developed or new offerings introduced. In a service experience, customers participate in the process of providing this experience and there is high customer contact throughout the service process. That is the reason why the customers' perceptions of the company and of its offerings must be managed well.

It is suggested that tourism providers should use SM a channel for listening to and for interactive communication with customers, by implementing specific strategies, such as (i) create a virtual environment that builds a virtual community by engaging tourists to share their knowledge; (ii) design SM communications to encourage tourists to share the nature of their experiences both during and after their experiences (Bilgihan et al., 2016; Kim et al. 2015; Xie & Zhang, 2014).

Attempts to interpret customer reactions in external and internal contexts are valid enough as a basis for everyday managerial adjustments. The importance of monitoring information on customer satisfaction as a measure of system performance, and the significance of people involvement and continual improvement in the service process, is well documented in the literature (see, for instance, Fitzsimmons & Fitzsimmons, 2008). Service management must enable the management to explore customers' ideas and suggestions for improvements. This approach enhances and supports the continual improvement of processes and, consequently improves customer satisfaction.

Data-mining from SM should be implementing in better understanding of consumer behavior, choices, concerns and determinants. Literature suggests some ways of effectively leveraging of UGC across SM platforms by using the adequate techniques / tools (see, for instance, Capriello et al., 2013; García-Pablos et al., 2016; Ghose et al., 2012; Xiang et al., 2015). Methods include approaches for mining consumer sentiment, such as manual content coding, corpus-based semantic method, and stance-shift analysis. These methods vary greatly in both process and output; however, from a methodological viewpoint, all three methods produce reliable results. As for the tools, they include machine learning techniques, text analysis, crowdsourcing, image classification, social geotagging, human annotations, and geomapping, as well as Natural Language Processing (NLP) tools to automatically process and manage textual customer reviews (e.g. to perform sentiment analysis).

Thus, listening carefully to customers can be a goldmine of information that can lead to improvements in products, services and processes, as suggested by Hjalager and Nordin (2011). Therefore, tourism providers might benefit from feedback and input coming from SM users.

2. SM: Knowledge Source for Customer-Driven Innovation

It is suggested that tourists may provide companies with input and knowledge that could be used to improve processes and develop new concepts. In simple terms, innovation means the introduction of new products/services, new production processes, managerial revisions and communication changes. Hjalager (2011, p. 127) defines innovation as “the creation of better or more effective products, processes, services, technologies, or ideas that are accepted by society, markets, and government”.

Listening to customers and relationships with them has always played a central role in tourism. However, the involvement of customers as informants often seems to be coincidental and unsystematic, and may also be contested. Hjalager and Nordin (2011) outline a series of methodologies/ approaches to customer-driven innovation. One of these approaches is blog mining. In fact, tourism blogging is not a new form of review. Interactive travel forums such as TripAdvisor.com and travelblog.org encourage tourists to recommend facilities to others (Lee et al., 2011).

The study by Volo (2010) assessed the value of tourists' blogs as a research data source by using them to evaluate the tourist experience and to assess their influence on the decision-making process of prospective tourists. The methods implemented were textual and image content analysis. On the whole, blogs constitute an immense resource providing insights that may be fed into innovation processes. Like other knowledge sources, blogging has its obvious drawbacks in terms of fostering innovation in tourism (Hjalager & Nordin, 2011).

Let us give some practical examples of performance adjustments and improvement resulting from customer input and knowledge that lead to new/better products and services in the field of wellness tourism and spas (Sotiriadis et al., 2016): (i) improvement of elements of architecture and design to create harmony and balance within the space; pairing built environment, treatment and adequate design allowing proper satisfaction of all senses; (ii) redesign of functionality and décor to add a sense of relaxation, health and

Table 1. Summary: Suggestions for potential contribution and uses of SM

Issues	Elements
Areas of potential contribution	Listening carefully to and learn from customers can be a goldmine of input and knowledge that could influence tourism providers (i) to improve the experience service quality, and (ii) to improve their innovation performance.
Approach and Policy	Integrated communications with customers: (i) Develop a multi-channel approach to communications; and (ii) SM as channel for interactive communication.
Strategies in acquiring customer feedback	- Engagement of customers in a mutually beneficial dialogue - Enhance participation and intention to share input and feedback through the use of SM. - Listening to customers and responding to tourists' comments and feedback - Use SM as a medium and source of feedback for performance improvement - Use SM as knowledge source for customer-driven innovation.
Operational management actions (Meaning implementing actions to:)	- Encourage tourists to relate effective service stories online. - Respond positively (taking advantage of positive postings and addressing negative comments) to online users' reviews, suggestions and experience sharing. - Analyze and interpret customer reactions for operational management adjustments. - Monitor customer feedback - Monitor customer satisfaction as a measure of system performance. - Involve customers systematically as informants in a process of continuous service improvement. - Focus more on lurking: Monitoring online reviews and experience sharing to understand tourists' needs and experiences - Use SM in listening on the online conversation; - Instigate engagement behavior / Use social gratifications to enhance tourists' participation and engagement on provider's website - Employ systematic approaches to exploit ideas and inspiration from customers - Mine blogs as a resource providing input into innovation processes - Use ideas and suggestions from customers in improving company's innovation performance.

Source: Adapted from Sotiriadis and Van Zyl, 2015, p. 173.

cleanliness to customers; more particularly offer aesthetic configuration, comfort and space; (iii) reprocessing/adequate design of treatments using innovative methods to rejuvenate body, soul and spirit; (iv) moving from a basic package of treatments to specialization, customized/thematic and signature treatments (e.g. signature chocolate treatments and spa packages); (v) offering new experience opportunities for a variety of treatments and adequate products (e.g. anti-ageing and result-oriented treatments); and (vi) use of SM platforms in communicating the highlight offers to the target market segments.

SM constitute electronic platforms for tourism businesses to interact with existing and potential tourists (Sotiriadis & Van Zyl, 2015). The following table is a summary of suggestions about the potential contribution and possible uses of SM by tourism providers. To benefit, they need to identify specific areas of contribution and develop effective strategies to optimize returns.

As was stressed earlier, customer knowledge and input should be seen as an opportunity, not a threat. It is quite clear that listening to tourists is vital to understanding and responding to their needs. It is worth pointing out that influencing the participation and the intention to share input and feedback through the use of SM requires tourism providers to influence consumers' beliefs and to differentiate themselves from their competitors. Scholars suggest developing SM platform that provides functional and practical benefits (Bilgihan et al., 2016).

FUTURE RESEARCH DIRECTIONS

This article provides some guidance on an approach to and possible uses of SM by tourism providers. It is believed that there is a need for more in-depth conceptual approaches and empirical investigation and studies to advance research

and knowledge in this area. Some suggestions for future research include exploring issues such as (i) the degree to which tourists are willing to participate in customer feedback and customer-driven innovation activities; and (ii) the feasibility of different SM within a general approach of interactive dialogue with customers. Another possible direction for future research is to examine the similarities and differences in engagement behavior of tourists using general SM and tourism-specific SM (for instance, Travelblog). Another research path could be the comparative analysis of micro-blog and tourism-related SM, such as interactive travel forums (e.g. TripAdvisor), and their use by both interest groups: tourists for inspiration and information purposes and tourism providers as complementary tools for marketing intelligence and innovation purposes.

CONCLUSION

For tourists, the Internet and SM have become an entertaining venue to gain knowledge and get first-hand reports of experiences. SM have opened up a new world of dialogue among people, including conversations about tourism services and destinations since they facilitate direct interactions between suppliers and tourists. It is therefore imperative for tourism providers to understand how they can be used for managerial and marketing purposes. By using SM effectively, tourism providers are communicating directly and efficiently with their current and potential customers.

Markets become conversations and tourism providers must engage others to have good conversations about their offerings. This involves marketing challenges for tourism businesses, mainly (i) elaborate and implement strategies to use these channels for dialogue and interaction purposes with consumers; and (ii) convince tourists to share their experience and knowledge via SM. Tourism providers who manage to create an

SM environment that builds a virtual community by engaging tourists to share their knowledge and insights will be more successful.

This article suggested two specific uses of SM by tourism providers, namely as a medium of feedback for performance improvements, and as a knowledge source for customer-driven innovation. It also pointed out some appropriate methods and suitable techniques/tools of effectively leveraging of UGC (reviews and experience sharing) across SM platforms. The latter must be approached and used as a tool and a resource to harvest valuable ideas and inspiration from customers.

However, tourism marketers must be aware of the limitations of every tool and method. How can tourism providers use SM and knowledge and input from customers? In practice, it will often be necessary to combine customers' inputs with other types of knowledge acquisition, as well as with other methods and tools. Customer feedback must be regarded as food for marketing, although it is usually a difficult management task to draw conclusions and come up with solutions. Therefore, this tool and its uses should be part of a marketing strategy and innovation management. In other words, the effective approach is to transform SM into a website dedicated to feedback; that is, websites allowing users to post, read, review, discuss and share experiences and opinions. Within a digital environment, the approach to marketing must be interactive. Tourism providers need to adopt a multi-channel approach to communications. The crucial issue is to communicate efficiently with tourists; a successful digital strategy will provide opportunities for suggestions and feedback.

Finally, it is worth emphasizing that SM are not panaceas; they are additional tools in the marketing toolbox. In order to improve their efficiency and effectiveness, SM must fit into integrated marketing communications plans and used as a channel for interactive communications, with the main aim to convert them in a medium for constructive dialogue that is mutually beneficial.

REFERENCES

- Ayeh, J. K., Au, N., & Law, R. (2013). Do we believe in TripAdvisor? Examining credibility perceptions and online travelers attitude toward using user-generated content. *Journal of Travel Research*, 52(4), 437–452. doi:10.1177/0047287512475217
- Benckendorff, P., Sheldon, P. J., & Fesenmaier, D. R. (2014). *Tourism information technology* (2nd ed.). London: UK CABI.
- Bilgihan, A., Barreda, A., Okumus, F., & Nusair, K. (2016). Consumer perception of knowledge-sharing in travel-related Online Social Networks. *Tourism Management*, 52(2), 287–296. doi:10.1016/j.tourman.2015.07.002
- Boley, B. B., Magnini, V. P., & Tuten, T. L. (2013). Social media picture posting and souvenir purchasing behavior: Some initial findings. *Tourism Management*, 37(1), 27–30. doi:10.1016/j.tourman.2012.11.020
- Bronner, F., & de Hoog, R. (2011). Vacationers and eWOM: Who posts, and why, where, and what? *Journal of Travel Research*, 50(1), 15–26. doi:10.1177/0047287509355324
- Bullas, J. (2015). *33 social media facts and statistics you should know in 2015*. Retrieved November 4, 2015, from <http://www.jeffbullas.com/2015/04/08/33-social-media-facts-and-statistics-you-should-know-in-2015/>
- Calefato, F., Lanubile, F., & Novielli, N. (2015). The role of social media in affective trust building in customer–supplier relationships. *Electronic Commerce Research*, 15(4), 453–482. doi:10.1007/s10660-015-9194-3
- Capriello, A., Mason, R. P., Davis, B., & Crotts, C. J. (2013). Farm tourism experiences in travel reviews: A cross-comparison of three alternative methods for data analysis. *Journal of Business Research*, 66(6), 778–785. doi:10.1016/j.jbusres.2011.09.018
- Chan, N. L., & Guillet, B. D. (2011). Investigation of social media marketing: How does the hotel industry in Hong Kong perform in marketing on social media websites? *Journal of Travel & Tourism Marketing*, 28(2), 345–368. doi:10.1080/10548408.2011.571571
- Dijkmans, C., Kerkhof, P., & Beukeboom, C. J. (2015). A stage to engage: Social media use and corporate reputation. *Tourism Management*, 47(1), 58–67. doi:10.1016/j.tourman.2014.09.005
- eMarketer. (2015). *Worldwide social network users: eMarketer's updated estimates for 2015. Report preview*. Retrieved November 5, 2015, from <http://www.emarketer.com/corporate/coverage#/results/1298?look>
- Fitzsimmons, J. A., & Fitzsimmons, M. J. (2008). *Service management: operations, strategy, information technology*. New York, NY: McGraw-Hill.
- García-Pablos, A., Cuadros, M., & Linaza, M. T. (2016). Automatic analysis of textual hotel reviews. *Information Technology & Tourism*, 16(1), 45–69. doi:10.1007/s40558-015-0047-7
- Ghose, A., Ipeirotsis, P. G., & Li, B. (2012). Designing ranking systems for hotels on travel search engines by mining user-generated and crowdsourced content. *Marketing Science*, 31(3), 493–520. doi:10.1287/mksc.1110.0700
- Gretzel, U., & Yoo, K.-H. (2013). Premises and promises of social media marketing in tourism. In S. McCabe (Ed.), *The Routledge handbook of tourism marketing* (pp. 491–504). New York: Routledge.
- Hjalager, A.-M. (2011). Strategic innovation in tourism business. In L. Moutinho (Ed.), *Strategic management in tourism* (2nd ed.; pp. 127–140). Oxfordshire, UK: CAB International. doi:10.1079/9781845935887.0127

- Hjalager, A.-M., & Nordin, S. (2011). User-driven innovation in tourism – A review of methodologies. *Journal of Quality Assurance in Hospitality & Tourism*, 12(4), 289–315. doi:10.1080/1528008X.2011.541837
- Kim, W. G., Lim, H., & Brymer, R. A. (2015). The effectiveness of managing social media on hotel performance. *International Journal of Hospitality Management*, 44(1), 165–171. doi:10.1016/j.ijhm.2014.10.014
- Law, R., Buhalis, D., & Cobanoglu, C. (2014). Progress on information and communication technologies in hospitality and tourism. *International Journal of Contemporary Hospitality Management*, 26(5), 727–750. doi:10.1108/IJCHM-08-2013-0367
- Lee, H. A., Law, R., & Murphy, J. (2011). Helpful reviewers in Tripadvisor, an online travel community. *Journal of Travel & Tourism Marketing*, 28(7), 675–688. doi:10.1080/10548408.2011.611739
- Leung, D., Law, R., van Hoof, H., & Buhalis, D. (2013). Social media in tourism and hospitality: A literature review. *Journal of Travel & Tourism Marketing*, 30(1–2), 3–22. doi:10.1080/10548408.2013.750919
- Morrison, A. M. (2013). *Marketing and managing tourism destinations*. New York, NY: Routledge.
- Munar, A. M., & Jacobsen, J. K. S. (2014). Motivations for sharing tourism experiences through social media. *Tourism Management*, 43(1), 46–54. doi:10.1016/j.tourman.2014.01.012
- No, E., & Kim, J. K. (2015). Comparing the attributes of online tourism information sources. *Computers in Human Behavior*, 50, 564–575. doi:10.1016/j.chb.2015.02.063
- Oz, M. (2015). Social media utilization of tourists for travel-related purposes. *International Journal of Contemporary Hospitality Management*, 27(5), 1003–1023. doi:10.1108/IJCHM-01-2014-0034
- Parra-López, E., Bulchand-Gidumal, J., Gutiérrez-Taño, D., & Díaz-Armas, R. (2011). Intentions to use social media in organizing and taking vacation trips. *Computers in Human Behavior*, 27(2), 640–654. doi:10.1016/j.chb.2010.05.022
- Saha, G. C., & Theingi, A. (2009). Service quality, satisfaction, and behavioural intentions: A study of low-cost airline carriers in Thailand. *Managing Service Quality*, 19(3), 350–372. doi:10.1108/09604520910955348
- Sigala, M. (2011). eCRM 2.0 applications and trends: The use and perceptions of Greek tourism firms of social networks and intelligence. *Computers in Human Behavior*, 27(2), 655–661. doi:10.1016/j.chb.2010.03.007
- Sigala, M. (2012). Social networks and customer involvement in new service development (NSD) the case of www.mystarbucksidea.com. *International Journal of Contemporary Hospitality Management*, 24(7), 966–990.
- Sigala, M., & Chalkiti, K. (2014). Investigating the exploitation of web 2.0 for knowledge management in the Greek tourism industry: An utilisation–importance analysis. *Computers in Human Behavior*, 30, 800–812. doi:10.1016/j.chb.2013.05.032
- Sigala, M., Christou, E., & Gretzel, U. (Eds.). (2012). *Social media in travel, tourism and hospitality: Theory, practice and cases*. London: Ashgate.
- Sotiriadis, M., & Van Zyl, C. (2013). Electronic word-of-mouth and online reviews in tourism services: The use of Twitter by tourists. *Electronic Commerce Research Journal*, 13(1), 103–124. doi:10.1007/s10660-013-9108-1
- Sotiriadis, M., & Van Zyl, C. (2015). Tourism services, micro-blogging and customer feedback: a tourism provider perspective. In J. N. Burkhalter & N. T. Wood (Eds.), *Maximizing commerce and marketing strategies through micro-blogging* (pp. 157–176). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-8408-9.ch007

Sotiriadis, M., Van Zyl, C. & Poole, C. (2016). Suggesting a framework for innovation management in the industry of wellness tourism and spas. *African Journal of Hospitality, Tourism and Leisure*, 5(4), 1-17.

TripAdvisor. (2014) *Research TripBarometer April 2014: Global Edition*. Retrieved April 30, 2014, from <http://www.tripadvisor.com/TripAdvisorInsights/n2200/tripbarometer-april-2014-global-edition>

Twitter. (2015). *Home page*. Retrieved November 4, 2015, from <http://www.twitter.com>

Volo, S. (2010). Bloggers reported tourist experiences: Their utility as a tourism data source and their effect on prospective tourists. *Journal of Vacation Marketing*, 16(4), 297-311. doi:10.1177/1356766710380884

Wilson, A., Murphy, H., & Cambra Fierro, J. (2012). Hospitality and travel: The nature and implications of user-generated content. *Cornell Hospitality Quarterly*, 53(2), 220-228. doi:10.1177/1938965512449317

World Travel Market. (2015). *The WTM global trends report 2015*. London: WTM & Euromonitor International.

Xiang, Z., & Gretzel, U. (2010). Role of social media in online information search. *Tourism Management*, 31(1), 179-188. doi:10.1016/j.tourman.2009.02.016

Xiang, Z., Schwartz, Z., Gerdes, J. H. Jr, & Uysal, M. (2015). What can big data and text analytics tell us about hotel guest experience and satisfaction? *International Journal of Hospitality Management*, 44(1), 120-130. doi:10.1016/j.ijhm.2014.10.013

Xie, L. K., & Zhang, Z. (2014). The business value of online consumer reviews and management response to hotel performance. *International Journal of Hospitality Management*, 43(1), 1-12.

Ye, Q., Law, R., Gu, B., & Chen, W. (2011). The influence of user-generated content on traveler behavior: An empirical investigation on the effects of e-word-of-mouth to hotel online bookings. *Computers in Human Behavior*, 27(2), 634-639. doi:10.1016/j.chb.2010.04.014

Zeng, B., & Gerritsen, R. (2014). What do we know about social media in tourism? A review. *Tourism Management Perspectives*, 10, 27-36. doi:10.1016/j.tmp.2014.01.001

ADDITIONAL READING

Benckendorff, P., Sheldon, P. J., & Fesenmaier, D. R. (2014). *Tourism information technology* (2nd ed.). London: UK CABI.

Bilgihan, A., Barreda, A., Okumus, F., & Nusair, K. (2016). Consumer perception of knowledge-sharing in travel-related Online Social Networks. *Tourism Management*, 52(2), 287-296. doi:10.1016/j.tourman.2015.07.002

Capriello, A., Mason, R. P., Davis, B., & Crotts, C. J. (2013). Farm tourism experiences in travel reviews: A cross-comparison of three alternative methods for data analysis. *Journal of Business Research*, 66(6), 778-785. doi:10.1016/j.jbusres.2011.09.018

García-Pablos, A., Cuadros, M., & Linaza, M. T. (2016). Automatic analysis of textual hotel reviews. *Information Technology & Tourism*, 16(1), 45-69. doi:10.1007/s40558-015-0047-7

Ghose, A., Ipeirotis, P. G., & Li, B. (2012). Designing ranking systems for hotels on travel search engines by mining user-generated and crowdsourced content. *Marketing Science*, 31(3), 493-520. doi:10.1287/mksc.1110.0700

Kabani, S. (2013). *The Zen of social media marketing: an easier way to build credibility, generate buzz, and increase revenue* (3rd ed.). New York, NY: BenBella Books.

Sigala, M., Christou, E., & Gretzel, U. (Eds.). (2012). *Social media in travel, tourism and hospitality: Theory, practice and cases*. Surrey, England: Ashgate.

Volo, S. (2010). Bloggers reported tourist experiences: Their utility as a tourism data source and their effect on prospective tourists. *Journal of Vacation Marketing*, 16(4), 297–311. doi:10.1177/1356766710380884

Wilson, A., Murphy, H., & Cambra Fierro, J. (2012). Hospitality and travel: The nature and implications of user-generated content. *Cornell Hospitality Quarterly*, 53(3), 220–228. doi:10.1177/1938965512449317

Zeng, B., & Gerritsen, R. (2014). What do we know about social media in tourism? A review. *Tourism Management Perspectives*, 10, 27–36. doi:10.1016/j.tmp.2014.01.001

KEY TERMS AND DEFINITIONS

Blogs: Blogs are websites on which people (bloggers) keep logs, share personal experiences and insights in a particular area, and interact with readers through the posting of comments.

Customer Feedback: The positive or negative information transmitted by customers to busi-

nesses about the quality of services provided and the degree of satisfaction.

Innovation: The introduction of new products/services or new production processes. It is defined as the creation of better or more effective products, processes, services or ideas that are accepted by the market.

Social Media: Any tool or service that uses the Internet to facilitate conversations; a group of internet-based applications that enable internet users from all over the world to interact, communicate and share ideas, experiences and information.

Social Media Marketing: A set of marketing techniques that use SM to create increases in brand awareness, achieve increased visitation and enable interaction, communication and collaboration of user-generated content.

Tourism Providers: All businesses offering tourism services and experiences to consumers when the latter are travelling and performing tourism activities. These businesses include transport companies, hospitality companies, travel agents and tour operators and other tourism-related companies directly involved in servicing tourists' needs.

Web 2.0: Describes websites that emphasize user-generated content, usability, and interoperability. A Web 2.0 site may allow users to interact and collaborate with each other in a SM dialogue as creators of user-generated content in a virtual community.

Usability Evaluation of Tourism Icons in India

H

Rajshree Tushar Akolkar

Zeal College of Engineering and Research, India

Ganesh D. Bhutkar

Vishwakarma Institute of Technology, India

INTRODUCTION

In India, there are twenty nine states and seven union territories and most of them have their official tourism websites. These websites are designed to provide relevant tourism information and services to the Public and visiting Tourists, who are Indian Citizens, Non-Resident Indians (NRI) or Foreigners. There are unique tourism icons used by these Indian tourism websites, which represent a diversified culture of the respective state of India. There are three major types of icons viz. simple icons, indices and symbols. An icon is a visual sign that looks like what it means. An index indicates about a particular object or process through indirect means. A symbol is a sign which is used for representation of objects or process through convention or standard (Agrawal & Bhutkar, 2012, 2015; Bhutkar et al., 2011; Cumbria, 2015; “Horton”, 1994; Specht, 2014; Kalsi et al., 2009; Shah, 2007; Withrow et al., 2000). The examples of icon, index and symbol in e-Governance websites are depicted in Table 1. An icon showing a house represents housing. An index for education is represented using a graduation cap with rolled degree and red ribbon, which is an indirect representation of education. A symbol for differently-abled services is depicted by a person in sitting position on chair-like object. All the icons in Table 1 are referred from a section – ‘Topic’ on Indian e-Governance Portal (“India Web Portal”, 2016).

The tourism icons used in these websites are of the type – index and each of these icons indicates about state tourism through indirect mean representing regional or cultural aspect(s) of the respective state of India (Bhutkar et al., 2011; Kergosien, 1991). In this chapter, the tourism icons available with these state tourism websites are studied and evaluated to propose a comprehensive methodology of icon evaluation.

Characteristics of Icons

The icons have about 40 characteristics representing several aspects incorporated in them (“Horton”, 1994; Specht, 2014). These icon characteristics include major characteristics such as –

- **Functional Analogy:** If you can’t show the object itself, you can show the object that performs an analogous function in more familiar area.
- **Animal Use:** Animals or birds are used as symbols.

Table 1. Examples icon, index and symbol in e-governance websites

Icon	Index	Symbol
		

- **Overlapping:** Allowing one object to block the view of another for conveying the proximity.
- **Size:** Size is a dimension or magnitude of any object.
- **Shape:** Shape is an outline of any object.
- **Clarity:** Clarity is defined as a clear, understandable, informative and easy association with the message.
- **Universal:** Globally used.
- **Labels Off:** Labels are not displayed or required (If symbol is not obvious or known to all readers then it must be labeled).
- **Functional Analogy:** If you can't show the object itself, you can show the object that performs an analogous function in more familiar area.
- **Color Indicator:** In color indication, each color denoted by background, border or text may have a different meaning.
- **Animal Use:** Animals or birds are used as symbols.
- **Overlapping:** Allowing one object to block the view of another for conveying the proximity.
- **Size:** Size is a dimension or magnitude of any object.
- **Shape:** Shape is an outline of any object.
- **Clarity:** Clarity is defined as a clear, understandable, informative and easy association with the message.

The icon characteristics are divided into two parts. A first part contains primary characteristics, which are required characteristics observed in most of the icons. The tourism icons have three required characteristics such as labels off, clarity and functional analogy. These characteristics are observed in most of the tourism icons. For example, a tourism icon for a state of 'Himachal Pradesh' is shown in Figure 1, which consists of Deodar Trees, Sun, Moon, Caves and Hills. This index does not have any label; which is a 'labels off' characteristics. This index is easily understandable and 'Clear'. It has a 'Functional

Figure 1. Tourism icon for the state of 'Himachal Pradesh'



analogy' in the form of natural entities such as Sun, Moon, Caves, Hills and Deodar trees representing a state of 'Himachal Pradesh' and natural life in Himalayan mountain ranges.

Ten tourism icons have been selected satisfying primary characteristics and are depicted in Table 2. The other part has secondary characteristics,

Table 2. Icons from state e-governance websites of India list of selected tourism

Icon	Icon Represented In State	Icon	Icon Represented In State
	Bihar		Nagaland
	Gujarat		Kerala
	Haryana		Rajasthan
	Himachal Pradesh		Jammu and Kashmir
	Sikkim		Tamil Nadu

which are optional and can be observed in a few of the icons. The tourism icons have many secondary characteristics such as animal use, overlapping, color indicator, shape and structural analogy. For example, a tourism icon for state of Gujarat has the 'Lion Face' which is the unique species ('animal use') found only in 'Gir Forest' in Gujarat. A characteristic like 'Overlapping' is present in a tourism icon for state of 'Jammu & Kashmir'. This index has the 'Shikara', which is a type of wooden boat normally found in 'Dal Lake', mountains and lake. A 'Color indicator' conveys the purpose (Specht, 2014). Many icons have used a green color to indicate trees or vegetation around, like tourism icons for 'Himachal Pradesh', 'Kerala' and 'Bihar' show tree(s) in green color.

Research Questions

This book chapter is an attempt to investigate two major research questions related with icon evaluation as follows:

1. Are the icons from tourism websites easily understandable to Public / Citizens?
2. How to evaluate and validate icons on tourism websites satisfactorily to fulfill their role in interaction with Public / Citizens?

BACKGROUND

An e-Governance is a big challenge and far big opportunity to bring services to all the Citizens for India, which is second-most populous and the largest democratic nation in the World. The most significant characteristic of any successful e-Governance website is its quality and accessibility (Agrawal & Bhutkar, 2015). The issue (appearance, cost and time) of integration of legacy systems also comes into the scene. The presence of e-Governance icons in general, makes these user interfaces and websites more user-friendly and easy to use in Human Computer Interaction (HCI) or usability perspective. Such iconic user

interfaces may be displayed for interactive usage to the users, who may be from weaker socio-economic and educational background. The appropriateness of these icons can be identified by usability evaluation of their designs. The user participation is always vital in such evaluations as e-Governance websites are used by Public i. e. people representing several categories such as Government Servants, Public Sector Employees, Private Sector Employees, Social Workers, Politicians, Academicians, Students as well as Foreign Nationals. The major usability issues with e-Governance icons include legibility, size and position of icons (Bhutkar et al., 2011). Icon legibility and size of most icons can be important issues during icon selection. These icons should be visible throughout the web pages visited for the necessary operations, which can be observed during the process of lexical analysis. These issues put icons at center of attention during usability research.

LITERATURE SURVEY ABOUT E-GOVERNANCE AND ICONS

Most of the papers studied in the literature survey section have tried to evaluate the e-Governance websites on the basis of maximum possible parameters pertaining to an aspect of visual design.

A research by Katre et al. (2004 & 2011) highlights the minimum essential parameters belonging to 8 different aspects of the state government websites in India. The overall result of the evaluation is considered as integrated feedback of Usefulness and Usability of the state web portal. The evaluation is only in terms of the presence or absence of the particular factors or applicability of the given parameter. The presence of proper visual design is recorded as (1) whereas its absence or poor design is recorded as (0). They have not rated the websites in terms of which one has better or worst visual design. The following are the usability aspects for e-Governance: Accessibility, Navigation, Visual Design, Readability, Informa-

tion Content, Textual Elements and Interactivity (Katre & Gupta, 2011).

Shah (2007) has discussed about what is e-Governance and structure of e-Governance in India. This study has analyzed the governance issues and challenges for the country. Withrow et al. (2000) has undergone a study of comparative usability evaluation for an e-Governance website. These studies have been instrumental in development of in-depth understanding about how to work with e-Governance websites.

Horton (1994) has discussed in his book - "The Icon" about fundamental of icons such as icon style, appearance, fitting, design and evaluation methods. It has described various components of icons viz. border, background, label, size, shape, synecdoche, clarity etc. This study has applied the appropriate component-based criteria learned for selecting suitable icons from e-Governance website for further semiotic analysis.

Bhutkar et al. (2011) has used the semiotic analysis combined with usability and ergonomic testing for evaluation of icons in medical user interfaces. In this paper, Usability Evaluation Methods (UEM) such as usability testing, user survey, lexical analysis, semiotic analysis and long distance visibility testing are used to evaluate the selected icons. This study have used lexical breakup for generating a data required for classification of icons related with medical user interfaces.

International language of ISO Graphical Symbols (2013) has defined, internationally accepted requirements for designs, colors, contents and shape of graphical symbols. The standardized colors and basic shapes for icons are important in helping the different types of signs, and messages they convey. This study has helped in understanding importance of identification as well as location of the icons.

Francesco (2011) has discussed about qualification of the tourist destination being worthy and the colors employed to indicate trails for outdoor activities. This study has been used in tourist world for recognizable and memorable iconography of the object and the use of related

images. It has been useful in identifying and selecting the appropriate tourism icons during this research work. Cumbria Country Council (2015) has explained benefits of symbols in simplifying the sign contents, size, use of colors, basic shapes to design and recognizing the icons using signs and symbols. It has integrated a tourism signage with strategic signage for towns and villages. This study has been beneficial in understanding and selecting the characteristics of icons.

There are many studies which elaborate on classifications of icons. Kergosien (1991) has discussed the semiotic analysis of medical imaging. As per this study, there are three types of signs viz. icon, index and symbol. Poovaiah (1995) has stated that the symbol design is governed by visual language which has various message representations such as - iconic, indexical and symbolic representations. This study is also useful for the classification of icons into icon, index and symbol.

RESEARCH METHODOLOGY

Following are the major steps involved in research work related with usability evaluation of tourism icons:

- Selection of Tourism Icons
- Usability Evaluation Methods (UEM)
 - Using Expert-based UEMs
 - Using User-based UEMs

These steps are discussed in details in this section.

Selection of Tourism Icons

In India, twenty nine states and seven union territories have their official tourism website. There is a distinct tourism icon representing each of these states or territories in India. These icons are extensively used by respective state tourism departments with their e-Governance Websites. Ten icons are selected based on their primary as well

as secondary characteristics as discussed earlier. The legibility of icons has been also a vital issue during icon selection. The selected ten icons are depicted in Table 2. The primary characteristics considered are labels off, clarity and functional analogy. These icons also have secondary characteristics such as shape, structural analogy, animal use, overlapping, interact icon and color indicator.

Most of the icons are rectangular in shape except the icons representing Kerala and Tamil Nadu. An icon of Himachal Pradesh has used background structure whereas Nagaland and Rajasthan have used hut and train structures respectively. The icons of Gujarat and Haryana have used as animals/birds and strips. An overlapping means a use of many related objects, which has a strong relation (“Horton”, 1994). The icon designs of Himachal Pradesh and Kerala depict a characteristic of overlapping.

Usability Evaluation of Tourism Icons

Usability evaluation of icons is carried out using selected UEMs. These methods include both expert-based UEMs as well as User-based UEMs (Agrawal & Bhutkar, 2012, 2015). These methods are discussed ahead.

Expert-Based UEMs

An expert-based UEM is any UEM, which involves usability expert examining the application or product and estimating its likely usability for a given user population. In expert-based methods, the users may not be employed and the basis for the evaluation lies in the interpretation and judgment of the expert evaluator. There is a considerable interest in such methods since they can produce results faster and are presumably cheaper than user-based UEMs. Two major expert-based UEMs used for usability evaluation of tourism icons are lexical analysis and semiotic analysis, which are discussed in next section.

Lexical Analysis

Linguistics has a term - ‘Lexical Analysis’ which is a process of converting a sequence of characters into a sequence of tokens. This study has used the concept for analysis of icons and performed a lexical breakup of icons during analysis. A lexical analysis has helped in classification of icons into mainly three categories: icon, index and symbol. A Table 3 has depicted the lexical breakup and categories of 10 tourism icons based on lexical analysis. This table shows that all tourism icons have indexical features in them.

Semiotic Analysis

A semiotics is a study of cultural sign processes, analogy, signification, communication, signs and symbols. Signs and relations are two key notions in semiotic analysis. An arbitrary and temporary separation is made between content and form to provide enough attention on the system of signs (“Cumbria County Council”, 2015; Francesco, 2011). Thus, a semiotic analysis is concerned with meaning, which came from relationships - in particular, the relationship among signs (“The International Language of ISO Graphical Symbols”, 2013). A semiotic analysis is closely related to the field of linguistics (De Souza, 2006). It is based on three major dimensions - namely semantics, syntactic and pragmatics. Semantics deals with a relationship of icon to its meaning. Syntactic deals with a relationship of icon to its form. Pragmatics deals with a relationship of icon to its users. The icons can be evaluated with these three dimensions by answering related questions.

Following are the standard questions for Semantic, Syntactic and Pragmatic dimensions (Agrawal & Bhutkar, 2012, 2015):

Questions for Semantic Dimension

- How well does the icon represent the message?

Table 3. Lexical breakups and classification of tourism icons based on lexical analysis

Sr. No	Tourism Icon (Index)	State associated with Represented Icon	Lexical Breakup	Remark
1		Bihar	Peepal (Bodhi) Tree	Peepal (Bodhi) tree is the tree under which Goutama Buddha attained his enlightenment. (ToI)
2		Gujarat	Lion Face + Border	Lion is unique species found only Gir Forest, Gujarat and no-where-else in Asia. (email)
3		Haryana	Peacock Animal + Back Ground	Peacocks are found in almost everywhere in Haryana. Suraj Kund in Lokarpur (Haryana) is a tourist place, which looks like a dancing peacock (Website) .
4		Himachal Pradesh	Deodar Trees + Hills + Moon + Sun + Cave	Deodar Tree – State Tree Hills – Roads are major mode of Transport and it is hilly Sun and Moon – Climate varies from hot and sub-humid to cold Cave – Rock cut caves (Website)
5		Sikkim	Tashi Tegye	The eight signs are called tashi tegye in taibetan. These auspicious signs are intimate with life and teaching of Buddha. (Website)
6		Nagaland	Ne Hut	A hut represents cottage industry such as spinning, weaving, woodwork, pottery and cane work. (Website)
7		Kerala	Coconut Tree + Mountain	Coconut tree
8		Rajasthan	Rath (Chariot) + Back Ground	The Royal Rajasthan on Wheels is a luxury tourist train run by Indian Railways, which follows a route through Rajasthan. (Website)
9		Jammu and Kashmir	Hills + House Boat + Back Ground	Shikara is a type of wooden boat which is normally found in Jammu and Kashmir, which is a hilly region. Shikaras and houseboats are used for transportation, fishing and seaweed harvesting. (email)
10		Tamil Nadu	Umbrella	Artistic work selected by TTDC Ltd. (TTDC Website an email)

- Has the icon already been accepted?
- Do people fail to understand the message that the icon denotes?

Questions for Syntactic Dimension

- How does the icon look?
- How well do the parts of icon relate to each other?
- Is the construction of the icon consistent in its use of orientation, format, scale, color and texture?

Questions for Pragmatic Dimension

- Can a person see the icon?
- What is the preferred position of icon?
- Does the icon remain visible throughout the operation?

For such categorical analysis, questionnaire method was used during the process of related user-based evaluation. These evaluations involving users are discussed in next section.

User-Based UEMs

Usability testing is used as User-based UEM in evaluation of icons. It involves a sample of users performing a set of pre-determined process or tasks. An aim of such testing is to examine the extent to which the application or product supports the intended users in their work (Agrawal & Bhutkar, 2012). Thus, users are asked to perform a set of tasks with the application or product. Depending on the primary focus of the evaluator, the users' success at completing the tasks and their performance may be recorded. After the tasks are completed, users are often asked to provide data on likes and dislikes through a survey or interview.

Following user-based methods are applied during usability evaluation of tourism icons:

1. User Survey
2. Test without Context
3. Test with Context

User Survey

User survey is a non-experimental descriptive research method. Surveys can be useful when a researcher needs to collect data on phenomena that cannot be directly observed (Kothari, 2004). Surveys are used extensively in library and information science to assess attitudes and characteristics of a wide range of subjects, from the quality of user-system interfaces to library user reading habits. A role was class of user such as primary or secondary user. The user preferences mainly included user's opinion about icons in terms of scale of goodness (1-5) as well as specific suggestions for improvement (Johns, 2010).

About 200 users were contacted in person or through email, but only 81 users responded with all three questionnaires answered completely. These 81 users participated in the usability testing as well as related user survey. These users are selected through convenient sampling method (Bhutkar et al., 2011; Kothari, 2004). Most of these users often deal with e-Governance Websites. The members such as Academicians, Social Workers and Students had responded in person and others had responded through an email. The frequency of users' responded under user type survey is listed in Table 4.

Table 4. User type wise frequency of e-governance users

User Type	Frequency
Central Government Servant	05
State Government Servant	02
Public Sector Employee	10
Private Sector Employee	23
Academician	11
Student	19
Politician	02
Social Worker	04
NGO Member	03
Others	02
Total	81

User survey mainly involved collection of data about user profiles and their preferences. User profile was an important document in evaluation of icons. It had vital information about e-Governance website users such as user type, age, state, contact, language skills etc. The state wise representation of users is shown in Table 5.

Table 6 shows the age range of users' and age wise frequency of users. In other state category of state wise users are from Uttarakhand, Uttar Pradesh, Andhra Pradesh, Karnataka, Jharkhand and Chandigarh as well as from a few countries such as USA, Saudi Arabia and Switzerland.

User Profiles

Following are the observations made based on study of user profiles of participant e-Governance users during usability testing and user survey.

Table 5. State wise frequency of e-governance users

States		Frequency
Andhra Pardesh		2
Bihar		2
Gujarat		3
Haryana		4
Himachal Pradesh		3
Jammu & Kashmir		1
Jharkhand		1
Kerala		1
Karnataka		2
Maharashtra		38
Orissa		2
Rajasthan		3
Tamil Nadu		4
Uttar Pradesh		5
Uttarakhand		2
West Bengal		2
Others	USA	2
	Saudi Arabia	1
	Switzerland	1
Total		81

- Many users were personally interviewed in Maharashtra, India. All Central Government Servants, State Government Servants, Public Sector Employees, Private Sector Employees, Academicians, Politicians and Non-Government Organization (NGO) Members responded through an email. Among Private Sector Employees, two users were working in USA, one worked in Saudi Arabia and one worked in Switzerland. Three users working in USA and Switzerland were NRIs and a user from Saudi Arabia was Foreigner.
- Participant users had proficiency in two to four languages. These languages mainly include English and Indian languages such as Assamese, Bengali, Gujarati, Hindi, Kannada, Kashmiri, Marathi, Nepali, Oriya, and Telugu. One user even had good knowledge of Arabic.
- The frequency of users calculated as User Type wise shown in Table 4, State wise shown in Table 5 and by Age wise shown in Table 6.

The actual User Profile entered by the user is shown in Figure 2.

Test Without Context

In this user-based test, only one icon is displayed at a time without any context or associated proposition like title, other icon or interface screen, as

Table 6. Age wise Frequency of e-Governance Users

Age Range	Frequence
Below 20	12
21-35	44
36-50	17
51-60	6
Above 60	2
Total	81

Figure 2. User Profiles

Icons from Indian e-Governance Website: User Profile		
No	Personal Information	Details
1	Name:	[REDACTED]
2	Designation:	RTI ACTIVIST
3	Organisation:	YASHADA, PUNE.
4*	User ID:	
5	Educational Qualification:	B.Tech (Computer Engineering)
6	Age:	24
7	City / Town / Village:	PUNE
8	State:	MAHARASHTRA
9	Address:	B-21, 17 Queen's Garden, Shivneri Govt. Quarters, Pune-4
10	Nationality:	Indian (Citizen / NRI / PIO) Foreigner (Country:)
11	Contact details:	Telephone: 020-26340014 Mobile: 9420462188 E-Mail: kedarjoshi128@gmail.com
12	Purpose of Accessing the Website:	
13*	Response:	In person / Through email / Other
14	Language Skills:	English / Hindi / (Any Indian Language: Marathi / Other)
15*	User Type:	Central Government Servant / State Government Servant / Public Sector Employee / Private Sector Employee / Academician / Student / Politician / Social Worker / NGO Member / Other (specify)
16	Signature:	[Signature]
17	Date:	20/04/2012

(* The items in bold will be filled by Interviewee (Researcher))

shown in Figure 3. Then, users are asked to evaluate it based on a related questionnaire.

The questionnaire for Test without Context in semiotic analysis (AIGA, 1981) is provided below:

Questionnaire for this test had following questions:

1. What does the icon represent?

Figure 3. Icon for Gujarat tourism using test without context

Icon No. 2



2. What are operations indicated by the icon?
3. Does the logical structure among objects shown, help to find the information one is looking for?
4. Any other suggestions/observations (e.g. color, color code, orientation etc.)?

Test With Context

In this user-based test, each icon is displayed along with context i. e. with related interface in part or completely with associated propositions, as shown in Figure 4. Then, users are asked to evaluate it based on a related questionnaire.

The questionnaire for Test with Context in semiotic analysis (AIGA, 1981) is provided below:

Questionnaire for this test had following questions:

1. What does the icon represent?
2. What are operations indicated by the icon?

Figure 4. Icon for Gujarat tourism using test with context



3. Does the logical structure among objects shown, help to find the information one is looking for?
 4. Is a size of the icon appropriate?
 5. Is a location of the icon proper? If no, then – What is the preferred location?
 6. Should the icon be visible throughout the operation?
 7. Scale of Goodness (1-5)-Likert Scale (Johns, 2010)
 8. Any other suggestions / observations (e. g. color, color code, orientation etc.)?
2. The icons which are passed both icon tests in usability testing are the ‘Well-Accepted’ icons.
 3. The icons which are failed in first test and passed in second test are ‘Well-Understood’ icons.
 4. The icons which did not pass both tests in usability testing are rejected or failed ones. These are ‘Difficult to Understand’ icons.
 5. Table 7 shows 2 out of 10 tourism icons that passed the first test and 5 tourism icons out of 10 icons that passed the second test.

RESULTS OF ICON EVALUATION USING SEMIOTIC ANALYSIS

Usability testing was the most important phase during this research work. A Table 7 shows a data about results of usability testing with tourism icons.

1. The each number indicates number of users who identified the corresponding icon in related tests namely, ‘Test without Context’ and ‘Test with Context’.

The observations and findings are explained with respect to three dimensions of semiotic analysis based on results of icon tests and user survey. The icons are divided into three category – ‘Well-Accepted’, ‘Well-Understood’ and ‘Difficult to understand’.

Semantic Dimension

Following are major observations based on semantic dimension of semiotic analysis of tourism icons under consideration:

Table 7. Results of usability testing with tourism icons

Sr. No.	Tourism Icon	State associated with Represented Icon	No. of Successful Users		Observations
			Test without Context	Test with Context	
1		Bihar	10	28	'Difficult to Understand'
2		Gujarat	64	69	'Well-Accepted'
3		Haryana	4	21	'Difficult to Understand'
4		Himachal Pradesh	22	72	'Well-Understood'
5		Sikkim	2	26	'Difficult to Understand'
6		Nagaland	6	23	'Difficult to Understand'
7		Kerala	24	35	'Difficult to Understand'
8		Rajasthan	17	55	'Well-Understood'
9		Jammu & Kashmir	42	61	'Well-Accepted'
10		Tamil Nadu	5	42	'Well-Understood'

1. The Tourism Icons: 'Gujarat' and 'Jammu & Kashmir' are the more popular icons which are identified by the maximum users correctly. 'Gujarat' icon has the 'Lion Face' which is the unique species found only in 'Gir Forest' in Gujarat and no-where-else in Asia. 'Jammu & Kashmir' icon has the 'Shikara' which is a type of wooden boat normally found in 'Dal Lake'. Out of 81 users, 64 had identified 'Gujarat' and 42 users
2. had identified 'Jammu & Kashmir' tourism icon in 'Test without Context'.
2. The Tourism Icon: 'Haryana' was not identified correctly because it shows only the bird 'Peacock'. Most of the users' opinion was that it represents 'National Bird of India', which is not the actual reason for its representation as depicted in Table 3. Only 4 out of 81 users identified this icon correctly.

Syntactic Dimension

Following are major observations based on syntactic dimension of semiotic analysis of e-Governance icons under consideration:

1. The tourism icon - 'Sikkim' was not identified correctly. The structure of this icon was not clear and understandable. Most of the users' opinion was that it is 'Rangoli' and one user even identified it as black magic. Only users which belong to user type - 'Central Government Employee' had given the correct answer.
2. The tourism icon - 'Nagaland' was not identified correctly because users did not identify this as a 'Ne Hut' which is crafted by Nagas – the residents of state of Nagaland.
3. The most of the users identified tourism icon - 'Rajasthan' as temple in 'Test without Context' as the structure in the icon looks like a temple. But when the Context of 'Rajasthan' was given, then the icon was properly identified as luxury tourist train by 55 users.
4. Few users suggested that the tourism icons - 'Nagaland' and 'Bihar' should have a better icon representation as these icons are very difficult to understand. Users have failed to understand the 'Ne Hut' which is crafted by Nagas from 'Nagaland' and 'Buddha Tree' which indicate the 'Bihar'. One user observed that the tourism icon - 'Kerala' was too abstract in the form of coconut tree and many of the users identified it as tourism icon for state of 'Goa'.

Pragmatic Dimension

Following are major observations based on pragmatic dimension of semiotic analysis of e-Governance icons under consideration:

1. Users felt that size of most icons was appropriate.

2. All users agreed that the all icons should be visible throughout the web pages visited for the improved understanding and comprehension.

FUTURE RESEARCH DIRECTIONS

Usability evaluation of icons is an interesting topic for usability researchers. The methodology proposed can be applied with selected icons from other domains with necessary alterations, if required.

CONCLUSION

During usability evaluation of tourism icons, it was observed that all of the selected icons have indexical features in them and therefore, they may not be easy to understand to users of e-Governance websites, who may be the users from Public or Citizens and also, from weaker socioeconomic or educational background.

It was also observed that the combination of different UEMs are effective in ensuring the overall communicability and usability of iconic user interfaces of e-Governance websites. These methods are as follows -

- Usability testing (comprehension by users).
- User survey (user preferences and feedback).
- Lexical analysis (concept by concept mapping of visual elements in the icon).
- Semiotic analysis (properties necessary for evolving a sign language out of icons).

These UEMs provide different insights pertaining to e-Governance icons and the icons are categorized as 'Well-Accepted', 'Well-Understood' and 'Difficult to Understand' icons.

ACKNOWLEDGMENT

We thank the project jury - Prof. M. V. Kulkarni. We also thank Mrs. P. A. Bailke and Mrs. Reshma Bhosale at VIT, Pune, India for their support to this research work.

REFERENCES

- Agaloglou, M., & Economides, A. (2009). Evaluating European Ministries' Websites. *International Journal of Public Information Systems*, 3, 147–177.
- Agrawal, R., & Bhutkar, G. (2012). Lexical Analysis of Selected Icons available with Indian e-Governance Website. *International Journal of Scientific & Engineering Research*, 3(5).
- Agrawal, R., & Bhutkar, G. (2015). *Usability Evaluation of Icons Related with e-Governance Website in India*. Pune, India: National Conference on Modeling, Optimization and Control.
- AIGA. (1981). *The American Institute of Arts, Symbols Signs*. New York: Hasting House Publishers.
- Bhutkar, G., Poovaiah, R., Katre, D., & Karmarkar, S. (2011). *Semiotic Analysis combined with Usability and Ergonomic Testing for Evaluation of Icons in Medical User Interface*. International Conference on Human Computer Interaction - IndiaHCI 2011. doi:10.1145/2407796.2407804
- Bihar Tourism. (n.d.). Retrieved March, 2016, from <https://www.bihartourism.gov.in/home.html>
- Cumbria County Council. (2015, July). *Cumbria Brown and White Tourism Signs*. Retrieved March, 2016, from <https://www.cumbria.gov.uk/elibrary/Content/Internet/544/40988142554.pdf>
- De Souza, C. S., Leita, C. F., & Prates, R. I. O. (2006). *The Semiotic Inspection Method*. Paper presented at Anais do IHC2006. doi:10.1145/1298023.1298044
- Frangialli. (2011). *Tourism Signs and Symbols*. World Tourism Organization.
- Gardner, J. (2007). Remote Web Usability Testing- Benefits over Traditional Methods. *International Journal of Public Information Systems*, 2(3), 63–72.
- Gujarat Tourism. (n.d.). Retrieved March, 2016, from <http://www.gujarattourism.com/>
- Haryana Tourism. (n.d.). Retrieved March, 2016, from: <http://www.haryanatourism.com>
- Himachal Pradesh Tourism. (n.d.). Retrieved March, 2016, from <http://www.himachaltourism.gov.in/>
- Horton, W. (1994). *The Icon Book: Visual Symbols for Computer Systems and Documentation*. New York: John Wiley and Sons.
- India Web Portal. (n.d.). Retrieved March 4, 2016, from <https://india.gov.in/>
- Jammu and Kashmir Tourism. (n.d.). Retrieved March, 2016, from <http://www.jktourism.org/>
- Johns, R. (2010). *Likert Items and Scales, Survey Question Bank: Methods Fact Sheet 1*. University of Strathclyde. Retrieved February, 2015, from URL: [http://survenet.ac.uk/sqb/data collection/likertfactsheet.pdf](http://survenet.ac.uk/sqb/data%20collection/likertfactsheet.pdf)
- Jones, S., Irani, Z., & Sharif, A. (2007). *E-Government Evaluation: Reflections on Three Organizational Case Studies*. 40th Hawaii International Conference on System Sciences. doi:10.1109/HICSS.2007.189
- Kalsi, N. S., Kiran, R., & Vaidya, S. C. (2009). Effective e-Governance for Good Governance in India. *International Review of Business Research Papers*.
- Katre, D. (2004). Uniface: Internet based Software for Remote Usability Testing of Icons. *ACM-SIGCHI-SI National Usability Conference Easy*.

Katre, D. (2011). *Visualization of Interface Metaphor for Software: An Engineering Approach* (Thesis). Retrieved from Universal Publisher's Dissertation Dot Com, ISBN: 1599423774.

Katre, D., & Gupta, M. (2011). Expert Usability Evaluation of 28 State Government Web Portals of India. *International Journal of Public Information Systems*, 3, 115–130.

Kerala Tourism. (n.d.). Retrieved March 4, 2016, from <http://ktil.in/>

Kergosien, Y. L. (1991, September). Generic Sign Systems in Medical Imaging. *IEEE Computer Graphics and Applications*, 11(5), 46–65. doi:10.1109/38.90567

Kothari, C. R. (2004). *Research Methodology: Methods and Techniques* (2nd ed.). New Age International (P) Ltd. New Delhi.

Nagaland Tourism. (n.d.). Retrieved March 4, 2016, from <http://tourismnagaland.com/>

Poovaiah, R. (1995). *Graphic Symbols for Environmental Signage: A Design Perspective*. *International Symposium on Public Graphics*.

Rajasthan Tourism. (n.d.). Retrieved November 10, 2015, from <http://www.rajasthantourism.gov.in>

Shah, M. (2007). E-Governance in India: Dreams or Reality? *International Journal of Education and Development using Information and Communication Technology*, 3(2), 125-137.

Sikkim Tourism. (n.d.). Retrieved March 4, 2016, from <http://www.sikkimtourism.gov.in/Webforms/General/Default.aspx>

Specht, J. (2014). *Architectural Tourism: Building for Urban Travel Destinations*. Springer Gabler. doi:10.1007/978-3-658-06024-4

Tamil Nadu Tourism. (n.d.). Retrieved March 4, 2016, from <http://www.tamilnadu tourism.org/>

The International Language of ISO Graphical Symbols. (2013). Retrieved March, 2016, from http://www.iso.org/iso/graphical-symbols_booklet.pdf

Withrow, J., Brinck, T., & Speredellozzi, A. (2000, December). *Comparative Usability Evaluation for an e-Government Portal*. Ann Arbor, MI: Diamond Bullet Design Report.

ADDITIONAL READING

Gardner, J. (2007). Remote Web Usability Testing- Benefits over Traditional Methods. *International Journal of Public Information Systems*, 2, 63–72.

Jones, S., Irani, Z., & Sharif, A. (2007). E-Government Evaluation: Reflections on Three Organizational Case Studies. *40th Hawaii International Conference on System Sciences*. doi:10.1109/HICSS.2007.189

Kurniawan, S. (1998). *Computer Icon Evaluation* (Doctoral dissertation). Hong Kong University of Science and Technology. 10.14711/thesis-b603843

KEY TERMS AND DEFINITIONS

E-Governance: Information and Communication Technology (ICT) that empowers the Government, its employees and citizens including women and weaker sections of society.

Lexical Analysis: The process of converting visual representations into a sequence of objects or tokens.

Semiotics: The study of signs and symbols.

Usability Testing: A technique used in user-centered design to evaluate a product or application by testing it with its intended users.

Virtual Tourism and Its Potential for Tourism Development in Sub-Saharan Africa



Paul Ankomah

North Carolina A&T State University, USA

Trent Larson

North Carolina A&T University, USA

INTRODUCTION

The rapid growth of ICT (Information and Communications Technology) has resulted in the development of innovative tools that could extend new opportunities for tourism destination marketers and potential tourists. The ICT-based tools provide the option of expanding the tourism product and tourism experience into the realm of Virtual Tourism (VT). VT has been defined as an ICT-based tool that can facilitate potential tourists' experiencing tourism attractions via the medium of Virtual Reality (VR) without visiting these tourist destinations. Since the tourism product is intangible and therefore cannot be pretested by the potential tourist before purchase, VT makes it possible to have a sense of the experience through immersion in VR. This chapter focuses on how the phenomenon of VT could be explored to realize its full potential, particularly by Sub-Saharan African countries. The chapter is divided into six sections. The first section reviews the literature on the concept of VR and also discusses VT in the context of VR. The section also examines the ICT components that support VR and consequently VT. The second section explains the reasons for the rise of VT and identifies some of the efforts and attempts (e.g. teleporter booths with 4D renditions of exotic destinations; virtual walks through historical sites) that have been made to develop VT as a part of the tourism industry.

The third section, examines some tourist attractions that could be marketed in the form of VT in Sub-Saharan Africa (SSA) to add value to the sub-region's existing physical tourist attractions. The section also identifies potential Sub-Saharan African VT markets. The fourth section identifies problems and provides suggestions to address problems that might scuttle these efforts at VT development. The fifth section recommends some areas for future research in VT and the sixth and final section provides some concluding remarks.

BACKGROUND

VT is an ICT-based tool that involves participant immersion and interaction (via visual graphics, sound, etc.) with the culture, history or other aspects of a tourist destination without physically traveling there (Ali & Frew, 2014). Sussman & Vanhegan (2000) indicated that VT is a convergence of human and computer interfaces to establish a 3D illusion of virtual (nonphysical) travel experiences. VT is also referred to as simply cybertourism (Prideaux, 2002). There are many technology-based activities assumed to belong to the realm of VT (e.g. panoramic photographs of a location). One of the fundamental characteristics of a VT experience is the level of interactivity by the person within the virtual environment. Creating a more realistic human experience within the

realm of VT requires a technological framework that allows the user to establish a presence in a virtual environment. This framework is known as VR.

VR is “a medium composed of interactive computer simulations that sense the participant’s position and actions and replace or augment the feedback to one or more senses, giving the feeling of being mentally immersed or present in the simulation (a virtual world)” (Sherman & Craig, 2003, p. 13). Pinho (as cited in Piovasan, Passerino & Pereira, 2012, p. 296) suggests that the essence of VR is captured by three fundamental purposes: immersion, interaction and involvement. Immersion refers to the degree to which the person is disengaged from the real world and perceives a connection to a virtual world synthesized by computer technology. The term interaction connotes the user’s ability to alter or reshape components of, or objects within a virtual environment. Involvement means that the user can navigate the virtual space (either actively or passively) to the degree desired.

In order for a VR application to meet its purposes and function properly, a VR system must be in place. A VR system consists of hardware—a computer/VR engine connected with input and output devices, e.g. keyboard, mouse, a helmet or head mounted display (HMD); software (application software and databases saturated with useful information) (Sherman & Craig, 2003). The market value of VR technology is projected to increase from \$980.4 million in 2014 to approximately \$1.66 billion by the year 2020 (MarketsandMarkets, 2015). According to IHS Technology, the demand for VR products will increase and approximately 7 million VR headsets will be purchased by consumers by the end of 2016 (Graham, 2016).

VR systems are classified by their technological sophistication, level of immersion in the virtual world and come in four varieties: 1) non-immersive systems include simple electronic 3D simulations viewed through a portal and displayed by a com-

puter monitor and manipulated by standard input devices (e.g. mouse, keyboard). 2) semi-immersive systems (or augmented reality systems) combine digital images of virtual environments displayed in devices (e.g. HMDs) together with objects from the real world. 3) projected systems—involves the creation of a physical space outfitted with display screens, video projectors and one or more users wearing input devices such as stereo glasses or haptic gloves. The computer-generated graphics displayed on the screens update or refresh in response to the user’s body movements as captured by the tracking devices imbedded in the glasses or gloves. An example of this would be Cave Automatic Virtual Environment (CAVE), a specific kind of multiple projection room that allows one or more users to experience virtual worlds without being fully immersed. 4) complete immersion systems rely on HMDs, tracking devices and haptic gloves to send feedback to the display system (Blackledge, Barrett, & Coyle, 2011). Some of the other terms used in place of VR include: virtual environment, artificial reality, virtual worlds, and artificial worlds (Bamodu & Ye, 2013).

Another closely related term vital to understanding VR is telepresence. Telepresence is the extent to which an individual feels present in a computer-generated, multi-sensory environment and possesses the capacity to remotely interact with or manipulate the form/content of the environment in some way (Steuer, 1993). Telepresence makes possible the use of remotely-operated devices such as robots, vehicles and drones that perform jobs in high-risk areas such as nuclear plants, law enforcement, military operations, and space exploration, etc. (Sofge, 2015). The telepresence concept has many applications including utility for VT.

Within the tourism industry, VR systems apply in several general areas such as planning and management, marketing, entertainment, education, accessibility and heritage preservation (Gutentag, 2010). The next section examines some of the reasons for the spread of VT.

VIRTUAL TOURISM

Reasons for Expansion and Development of VT Products

There are several possible reasons for the expansion of VT in the travel industry. With the rise of terrorism and groups such as ISIS operating in many countries, people are fearful of leaving the safety of their own country. VT provides opportunities to participate in tourism activities without being physically present in the location. Additionally, VT has the advantage of allowing individuals to experience the destination and make evaluation prior to the actual purchase and subsequent travel to the destination.

Another reason for VT offerings could be the cost of travel. With travel related expenses soaring, the overall cost of traveling extensive distances may be prohibitive. Also, VT could appeal to those afraid to travel or physically unable or unwilling to sit for 10-15 hours in an airplane to experience a distant location with all the associated discomfort and inconvenience of security check points, delayed or cancelled flights.

Additionally, the rapid development of technology has the potential to make VT more easily accessible to people who already have access to ICT. With the trend towards more home-based leisure (e.g. computer games, online entertainment) on the rise, VT activities are likely to become ubiquitous like videogames (Williams & Shaw, 2009). VT connects people with a gamut of tourism destinations without the need for money or resources or having to go to those places (Cheong, 1995). Through virtual channels, people can experience a variety of places and even jump from one continent to another in minutes or go back in time to see digital replicas of ancient civilizations or navigate newly-created fantasy worlds (Cheong, 1995).

VT could also be a way to travel when time is in short supply. There is also the convenience factor. Virtual tourists don't have to know a language or have to go through the trouble of finding

a location in an unfamiliar destination. They also don't need to worry about monetary exchange, food, weather, etc. VT is predictable and appeals to people who don't like surprises.

Visits to remote areas via virtual technology always occur under optimal conditions, i.e. with sufficient lightening, no crowds and without limits on the length of stay (Kaminsky, Arnold, & Benson, 2014). Furthermore, VT could serve as recollection for people who have been to a destination and want to reminisce over the experience and to see what changes have occurred if any, since their last visit (Hyun, Lee, & Hu, 2009).

Tourism providers have sought to utilize the VR framework to create useful applications for potential tourists to have easy access to destinations and these applications are pervading the travel industry. The remainder of the section examines how tourism providers are manipulating non-immersive, semi-immersive, and complete immersive systems of VR to complement existing tourism offerings and develop new VT products.

Non-immersive system applications for VT have been available for quite some time and appear in 2D and 3D formats. Such computer-generated renderings of historical sites allow visitors to engage with ancient buildings, other artifacts and virtual people (from that time period) in a digital world (Catalano, et al., 2011). YouVisit.com is an example of a company that specializes in the development of a variety of virtual 3D tours including parks, hotels, and cities in various parts of the world. Tourism websites that feature 3D renderings of natural and cultural elements of a destination to capture the essence of the place are more likely to result in more enjoyable visitor experiences (Huang, Backman, Backman, & Chang, 2016). Providing online VR simulations of resorts/facilities before travel begins is becoming a necessity for tourist destinations to maintain their relevance in the competitive global tourism market (Clayton, Ajagunna, & Pinnock, 2014).

Semi-immersive systems (also referred to as augmented reality) are those that combine virtual with real world elements of the tourism industry.

For example, the Marriott Corporation, as part of its *Travel Brilliantly* campaign, has developed teleporter booths that use VR to simulate a beach experience in Hawaii (Shu, 2014). Each booth features virtual video content displayed via HMDs and headphones that must be carefully synchronized with 4D content: the action of ocean waves simulated by misters, the warmth of breezes created by heat blowers and the sensation of tactile simulators on the booth floor to create the feeling of walking in sand (Shu, 2014).

HMDs and projection systems such as CAVEs are two commonly used applications of VR to attract tourists to theme parks in North America (Jelinski, 2013). At Disney Quest, HMDs are used in Aladdin's Magic Carpet Ride while CAVEs are part of Pirates of the Caribbean: Battle for Buccaneer Gold and Hercules in the Underworld attractions. The world's first exclusively VR theme park is slated to open in Utah in 2016; it will feature physical objects from the real world and simulated 4D effects of artificial wind, mists of water, and heat lamps (Metz, 2015).

Landmark Entertainment, a company that develops rides and other attractions at Universal Studios Hollywood, Paramount Parks and some Las Vegas hotels, is initiating a new VR project that includes the construction of L.I.V.E. Centers at shopping malls in China that would provide patrons opportunities to experience virtual animal parks, aquariums and museums through the use of HMDs and eventually, using the same technology, virtual theme parks accessible online from any location in the world (Roettgers, 2015). Besides semi-immersive systems there are complete immersive systems that do not rely on real objects from the physical world.

Complete immersive systems are creating VT offerings and these are helping to stimulate visits to tourism destinations in various parts of the world. In conjunction with the VisitScotland expo, a company called Augmented Visualisations has developed a 3D application that takes visitors on a virtual tour of local landmarks such as Arbroath

Abbey, Angus Glens and Glamis Castle to encourage them to visit the country (British Broadcasting Corporation, 2015). Another company in Scotland has undertaken to develop a realistic VR model of a ship (RSS Discovery) used by Captain Robert Scott to sail to Antarctica (British Broadcasting Corporation, 2014). Once the digital scanning of the original ship is complete, visitors will be able to virtually visit the bridge, climb the rigging and explore areas below the deck of the ship from the comfort of their own living room (British Broadcasting Corporation, 2014). Quantas has partnered with Samsung Electronics to develop an inflight entertainment service that relies on VR headsets to show tourists virtual images of attractions, destination accoutrements and other products offered by the airline (Quantas, 2015). Marriott has implemented another program to promote the use of HMDs by guests. This program is called *VRoom Service* and permits hotel patrons to check out VR headsets to view of virtual postcards containing highlights from trips to destinations in Rwanda, Chile and China (Graham, 2016).

VT is still in its infancy, however, tourism-related businesses in collaboration with ICT companies have created VT experiences on an experimental scale in their tourism promotion activities. For instance, Thomas Cook has teamed with Samsung and Visualise (VR film makers) to create a series of VT clips labeled 'Try Before you Fly' of retail shopping experience of the U.K., Germany and Belgium. Potential visitors to these countries can virtually visit their favorite retail stores prior to their actual visitation (Graham, 2016). Another example of VT on an experimental scale is provided by Matoke Tours which recently released a VT brochure featuring unique tour offerings (gorillas, safari balloon flights, safari boat rides, a gamedrive in Queen Elizabeth National Park) of Uganda (Graham, 2016). The purpose of virtual brochure application was to "convey the intensity and emotion of the travel experience before the journey has even started" (Graham, 2016).

Another example of a pre-trip immersive virtual tourist experience is found in British Columbia. Destination BC, marketed by the Crown corporation, offers a virtual boat ride along British Columbia's coast or a virtual hike through the Great Bear Rainforest (Meissner, 2015). The digital 3D video becomes animated through Oculus Rift headsets and headphones. Capturing the spectacular views for these unique tourism offerings required special cameras and the use of drone technology (Meissner, 2015). Landmark Entertainment is also currently exploring options for the creation of virtual theme parks accessible online through HMDs from any online location in the world on a daily basis (Roettgers, 2015). Each of the aforementioned VR systems can create unique VT products that complement actual trips to locations throughout the world.

All tourist attractions could be experienced through the medium of VT. Consequently, it is reasonable to assume that similarly all tourist attractions in the sub-region that people physically travel to visit could also be experienced through VT. The Sub-Saharan African VT offerings could be transmitted in all formats previously discussed. SSA has many categories of tourism products that could appeal to tourists from international markets as VT offerings. Some of these products could include safari tourism (viewing Africa's *Big Five* via various modes of transportation or on foot), natural products (birdwatching, tracking gorillas, sightseeing, hiking, etc.), resort tourism (beach tourism, lake tourism, watersports, etc.) (World Bank, 2009); heritage tourism (Yankholmes & McKercher, 2014); volunteer tourism (Tomazos, 2010); medical tourism (Connell, 2011; Ramirez de Arellano, 2011); and education tourism (Ankomah & Larson, 2000). Developing VT products from the tourism resources of SSA could complement existing tourism development efforts and could also generate new tourism experiences. The next section identifies the potential market for VT in SSA.

Possible Sub-Saharan Africa VT Markets

H

This part of the section describes the VT markets in SSA and then identifies some possible areas for development in the sub-region. According to the 2009 World Bank Report, the top five international markets for SSA are France, UK, US, Germany, and Portugal; identified emerging markets for the sub-region include Spain, Australia, India, China, and Russia (World Bank, 2009). The availability of VT options could possibly substantially increase the numbers of potential tourists from these areas who might want to either physically visit or remotely-experience the sub-region. These countries are also among the leading innovators and adopters of ICT.

Potential Problems Associated with Implementation of VT Strategies by Sub-Saharan Africa Countries

Potential problems that could derail or impede strategies for developing and implementing VT products in the sub-region may include the following:

1. **Weak Technological Infrastructure:** Many countries in SSA have a marginal technological infrastructure at best. Though cell phones have become commonplace in the sub-region, there remain many antiquated ICT systems in operation. For instance, many locations continue to rely upon dial-up modems to transmit data back and forth. Connecting sophisticated VR systems to outdated hardware could result in VT applications that do not function at all or that operate inefficiently. These kinds of difficulties could frustrate technological-savvy tourists wanting to engage in VT portrayals of tourist destinations in the sub-region.

2. **Erratic Supply of Electricity:** Electricity is not a reliably-available commodity in the sub-region and extended disruptions in service are common. This could present a problem for keeping essential computer hardware online and also lead to damage of other equipment needed for the development and maintenance of VT offerings. It makes no difference how modern the ICT equipment is if there is no consistent source of electricity, VT products will not be available.
3. **Lack of Foreign Exchange:** Due to a lack of foreign exchange, many Sub-Saharan African countries do not have the financial resources necessary to update obsolete computer equipment and to purchase new technology needed to enable VT activities.
4. **Constantly Evolving Technology:** Technological innovations come out so rapidly that countries in SSA already behind the adoption curve may never find sufficient resources to purchase upgrades and other equipment necessary to remain competitive with the developed world in terms of VT applications.
5. **Technology Skills Deficit:** Another problem facing countries in the sub-region is the lack of ICT skills in the areas of system analysts, programmers, maintenance experts and consultants (Edoho, 2013). With an insufficient number of people with expertise in computers and other technological skills, many VT projects could remain on hold indefinitely or only become available on a limited basis. Also, there may be tourism stakeholders that are reticent to adopt additional technologies because they are not educated about or familiar with how the new technologies operate.
6. **Underdeveloped Domestic Tourism Market:** For any strategy to implement VT to be sustainable over the long-haul there must be some buy-in by the locals in countries of the sub-region and this may be difficult to accomplish. Many local tourism

stakeholders may be reluctant to adopt the VT approach because they may feel that it is only a passing fad and a waste of scarce resources. Any successful attempt to make VT applications a significant part of tourism must also include development of the domestic tourism market, particularly inter and intra-regional products.

The next section identifies some solutions to these problems and also offers some recommendations that could increase the chances of making VT development efforts successful.

SOLUTIONS AND RECOMMENDATIONS

The previous section identified some of the major problems associated with the development of VT products in SSA. This section provides some solutions to these problems including:

1. **Strengthening Technological Infrastructure:** The technological infrastructure of the sub-region could be strengthened by these countries pooling their resources to purchase the necessary computer hardware and other devices that make VT feasible.
2. **Improving Supply of Electricity:** The problem of erratic supply of electricity could be addressed by pooling resources of African countries to purchase better grid network equipment to improve the efficiency of the delivery of electrical power to its citizens. Countries could also collaborate to develop solar power that could compensate for the shortage of electricity generated through other sources.
3. **Enhancing Foreign Exchange:** Sub-Saharan African countries could make up for the lack of foreign exchange by pooling their resources to buy much-needed ICT components. The countries could also find

areas where they can eliminate waste of scarce resources. One way to do that would be to eliminate political corruption and cronyism in government.

4. **Keeping Pace with Evolving Technology:** The only way for countries of the sub-region to meet the problem of rapidly changing technology is to pool resources to be able to afford the onslaught of technological innovations. These countries could also engage in research on emerging technologies to determine which ones will be the best long-term investments.
5. **Closing the Gap on Technology Skills Deficit:** The problem of the technology skills deficit could be addressed by dialoging with the diaspora from Sub-Saharan African countries and invite them to return to become a nucleus (similar to what China and India have done) to help train locals in how to use and maintain the computer components and other devices that comprise VT. Also, for those diasporas who wish to remain in their host countries, a virtual community could be established to transfer technology skills while on sabbatical or through distance education.
6. **Developing Domestic Tourism Market:** Sub-Saharan African countries could make the VT effort more sustainable in the long-term by continuing to promote and develop the domestic tourism market in the sub-region. This could be done by encouraging locals to participate in the creation and marketing of VT experiences. In so doing, they could create new forms of tourism through drone technology and this may help stimulate greater demand for domestic tourism and VT products.

In addition to solutions to the problems associated with developing VT products, this section also provides some recommendations that could make these virtual offerings more attractive and feasible. It is suggested that Sub-Saharan African

countries organize and house their VT products in one location on the internet. This could be the website of the country's NTO (National Tourism Organization) at the country level or other multinational organizations. Tourists looking for a VR experience could easily find and engage with VT resources.

FUTURE RESEARCH DIRECTIONS

There are several avenues that future research could take to advance the spread of VT products in SSA. Some these include:

1. **Determining VT Readiness:** Studies could be conducted to determine the degree of readiness of tourism stakeholders for accepting VT as a viable and additional option to traditional tourism offerings.
2. **Effectiveness of VT Products:** Research studies could be designed to discover whether the use of VT products prior to travel played a role in decisions to travel to SSA.
3. **Appropriate Areas for VT Development:** Researchers could sample the opinions of tourism stakeholder as to which kind of tourism attractions in the sub-region are best-suited for VT applications.
4. **Impact of VR on the Future of VT:** Research studies should focus on identifying the future direction of VR and its potential to impact VT.

CONCLUSION

Countries of SSA are looking for new ways to generate revenue to support their meager resources. Tourism is one of the areas that countries have been focusing on in their efforts to expand economic development resources, especially foreign currency. VT has the potential to increase total tourism revenue in the sub-region. VT also presents the potential to generate additional resources

by giving these countries the opportunity to offer different prices to potential tourists by charging different rates for non-immersive, semi-immersive or complete immersive experiences. The countries in the sub-region have an opportunity to display tourism artifacts (e.g. carvings, traditional clothing, trinkets, etc.) on the websites that might be of interest to non-visitors who may still want to own these artifacts as souvenirs. Additionally, lesser-known tourism destinations have the opportunity to gain global exposure for their tourism attractions via VT. Furthermore, by making efforts to be involved in the VT industry, these countries will ensure that they are not left behind when VT becomes the phenomenon of the future.

REFERENCES

- Ali, A., & Frew, A. (2014). ICT and sustainable tourism development: An innovative perspective. *Journal of Hospitality and Tourism Technology*, 5(1), 2–16. doi:10.1108/JHTT-12-2012-0034
- Ankomah, P., & Larson, T. (2000). Education tourism: A strategy to sustainable tourism development in Sub-Saharan Africa. *DPMN Bulletin*, 7(1), 19–24.
- British Broadcasting Corporation. (2015). *Virtual reality tour of Angus aims to drum up tourism*. Retrieved October 14, 2015 from <http://www.bbc.com/news/uk-scotland-tayside-central-32416220>
- British Broadcasting Corporation. (2014). *3D virtual reality tours of RRS Discovery developed*. Retrieved October 14, 2015 from <http://www.bbc.com/news/uk-scotland-tayside-central-29363987>
- Blackledge, J., Barrett, M., & Coyle, E. (2011). Using Virtual Reality to Enhance Electrical Safety and Design in the Built Environment. *ISAST Transactions on Computers and Intelligent Systems*, 3(1), 1–9.
- Bamodu, O., & Ye, X. (2013). Virtual reality and virtual reality system components. In F. Cui (Ed.), *Proceedings of the 2nd International Conference On Systems Engineering and Modeling (ICSEM-13)*, 921–924. doi:10.4028/www.scientific.net/AMR.765-767.1169
- Catalano, C., Mortara, M., Spagnuolo, M., & Falcidieno, B. (2010). Semantics and 3D media: Current issues and perspectives. *Computers & Graphics*, 35(11), 869–877.
- Chung, C., & Koo, C. (2015). The use of social media in travel information search. *Telematics and Informatics*, 32(2), 215–229. doi:10.1016/j.tele.2014.08.005
- Clayton, A., Ajagunna, I., & Pinnock, F. (2014). Tourism in a transforming world economy – the impacts of the brave new world: How might Jamaica achieve long-term sustainable development? *Worldwide Hospitality and Tourism Themes*, 6(2), 191–196. doi:10.1108/WHATT-01-2014-0007
- Connell, J. (2011). *Medical tourism*. Oxfordshire, UK: Wallingford CABI. doi:10.1079/9781845936600.0000
- Edoho, F. (2013). Information and communications technologies in the age of globalization: Challenges and opportunities for Africa. *African Journal of Economics and Management Studies*, 4(1), 9–33. doi:10.1108/20400701311303131
- Graham, L. (2016). Virtual reality devices could transform the tourism experience. Retrieved January 31, 2016 from <http://www.cnn.com/2016/01/08/virtual-reality-devices-could-transform-the-tourism-experience.html>
- Huang, Y., Backman, K., Backman, S., & Chang, L. (2016). Exploring the implications of virtual reality technology in tourism marketing: An integrated research framework. *International Journal of Tourism Research*, 18(2), 116–128. doi:10.1002/jtr.2038

- Hyun, M., Lee, S., & Hu, C. (2009). Mobile-mediated virtual experience in tourism: Concept, typology and applications. *Journal of Vacation Marketing*, 15(2), 149–164. doi:10.1177/1356766708100904
- Jelinsky, J. (2013). *Popularity of virtual reality immersion in theme park attractions of North America*. Retrieved February 25, 2015 from http://aquila.usm.edu/cgi/viewcontent.cgi?article=1105&context=honors_theses
- Kaminsky, J., Arnold, D., & Benson, A. (2013). Cultural heritage tourism: future drivers and their influence. In J. Kaminsky, D. Arnold, & A. Benson (Eds.), *Contemporary issues in cultural heritage tourism* (pp. 319–337). New York: Routledge.
- MarketsandMarkets. (2015). *Virtual reality market by technology, device, component, application, geography-Trend & forecasts to 2014-2020*. Retrieved March 5, 2016 from <https://www.reportbuyer.com/product/3086634/virtual-reality-market-by-technology-device-component-application-geography-trends-and-forecasts-to-2014-2020.html>
- Meissner, D. (2015). B.C. uses VR tech to offer virtual rides, hikes in tourism pitch. Retrieved March 15, 2016 from <http://o.canada.com/travel/b-c-uses-vr-tech-to-offer-virtual-rides-hikes-in-tourism-pitch>
- O'Connor, P. (2008). User-generated content and travel: A case study on TripAdvisor.com. In P. O'Connor, W. Höpken, & U. Gretzel (Eds.), *Information and communication technologies in tourism 2008* (pp. 47–58). Vienna, Austria: Springer Verlag. doi:10.1007/978-3-211-77280-5_5
- Sherman, W., & Craig, A. (2003). *Understanding virtual reality: Interface, application and design*. San Francisco, CA: Morgan Kaufmann.
- Pease, W., Rowe, M., & Cooper, M. (2007). *Information and communication technologies in support of the tourism industry*. Hershey, PA: Idea Group. doi:10.4018/978-1-59904-159-9
- Piovasan, S., Passerino, L., & Pereira, A. (2012). *Virtual reality as a tool in the education*. Retrieved from ERIC database. (ED542606)
- Qantas. (2015). *Qantas & Samsung unveil industry-first virtual reality experience for travelers*. Retrieved March 10, 2016 from <http://www.bbc.com/news/uk-scotland-tayside-central-29363987>
- Ramirez de Arellano, A. (2011). Medical tourism in the Caribbean. *Signs (Chicago, Ill.)*, 36(2), 289–297. doi:10.1086/655908 PMID:21114073
- Roettgers, J. (2015). *Landmark Entertainment looks to launch virtual reality theme parks*. Retrieved March 10, 2016 from <http://variety.com/2015/digital/news/landmark-entertainment-look-to-launch-virtual-reality-theme-parks-1201514371/>
- Tomazos, T. (2010). Volunteer tourism – an ambiguous marketing phenomenon. *Innovative Marketing*, 6(4), 42–47.
- Yankholmes, A., & McKercher, B. (2014). Re-thinking slavery heritage tourism. *Journal of Heritage Tourism*, 10(3), 233–247. doi:10.1080/1743873X.2014.988159
- Williams, A., & Shaw, G. (2009). Future play: Tourism, recreation and land use. *Land Use Policy*, 26S, S326–S335. doi:10.1016/j.landusepol.2009.10.003
- World Bank. (2009). *Sub Saharan Africa tourism industry research*. Washington, DC: Author.

ADDITIONAL READING

- Ali, A., & Frew, A. (2010). ICT – an innovative approach to sustainable tourism development. *Journal of Hospitality Application & Research*, 5(2), 28–58.
- Chuah, J., & Lok, B. (2012). Experience in using a smartphone as a virtual reality interaction device. *The International Journal of Virtual Reality*, 11(3), 25–31.

Minghetti, V., & Buhalis, D. (2009). Digital divide in tourism. *Journal of Travel Research*, 49(3), 267–281. doi:10.1177/0047287509346843

Mirk, D., & Hlavacs, H. (2014). Using drones for virtual tourism. In D. Reidsma, I. Choi, & R. Bargar (Eds.), *Intelligent technologies for interactive entertainment* (pp. 14–147). Chicago, IL: Springer. doi:10.1007/978-3-319-08189-2_21

Najafipour, A., Heidari, M., & Foroozanfar, M. (2014). Describing the virtual reality and virtual tourism community (applications and implications for tourism industry). *Kuwait Chapter of Arabian Journal of Business and Management*, 3(12a), 17–23.

Neuhofer, B., Buhalis, D., & Ladkin, A. (2015). Smart technologies for personalized experiences: A case study in the hospitality domain. *Electronic Markets*, 25(3), 243–254. doi:10.1007/s12525-015-0182-1

Tavakoli, R., & Mura, P. (2015). Journeys in Second Life-Iranian Muslim womens behavior in virtual tourism destinations. *Tourism Management*, 46, 398–407. doi:10.1016/j.tourman.2014.07.015

Young, M., Markham, F., Reis, A., & Higham, J. (2015). Flights of fantasy: A reformulation of the flyers dilemma. *Annals of Tourism Research*, 54, 1–15. doi:10.1016/j.annals.2015.05.015

KEY TERMS AND DEFINITIONS

Head Mounted Display (HMD): A digital display device worn over the head that makes viewing of 3D computer images possible and leads the user to semi-immersion or total immersion in a virtual environment.

ICT: Electronic devices (e.g., computers, cell phones) and the wired or wireless infrastructure necessary for communication and other unique services such as e-learning, social networking and videoconferencing.

Immersion: The level at which someone engaged in a virtual environment senses the real world outside of it.

Interactivity: The degree to which a user of a VR system is able to control the content and appearance of a simulated environment.

Virtual Community: A cooperative group of individuals from various geographical regions that come together via ICT to solve social, business or educational objectives.

Virtual Reality: Computer-generated simulations of real-world objects and locations (and sometimes imaginary ones) that provide stimulation to the senses of the user and lead to a sensation of being present in a replicated environment almost as if it were the original.

Virtual Tourism: An ICT-based tool that uses digital images and sensory feedback to simulate tourist attractions available at remote destinations.

Category H

Human-Computer Interaction

Affect–Sensitive Computer Systems

Nik Thompson

Curtin University, Australia

Tanya McGill

Murdoch University, Australia

David Murray

Murdoch University, Australia

INTRODUCTION

Affective computing is the broad domain encompassing all of the hardware, software and underlying theoretical models underpinning the development of affect sensitive computer systems. Such systems facilitate more intuitive, natural computer interfaces by enabling the communication of the user's emotional state. Despite rapid growth in recent years, affective computing is still an under-explored field, which holds promise to be a valuable direction for future software development. Human-computer interaction has traditionally been dominated by the information processing metaphor and as a result, interaction between the computer and the user is generally unidirectional and asymmetric. The next generation of computer interfaces aim to address this gap in communication and create interaction environments that support the motivational and affective goals of the user.

This chapter will introduce and elaborate on the field of affective computing. First the background and origins of the field will be discussed. Next the elements of affective computing and affective human-computer interaction will be discussed along with associated concerns and issues. Next, examples of the diverse range of affective computing applications in current and recent development will be provided. Finally, the chapter will present a discussion of future directions for this promising technology, followed by some concluding remarks.

BACKGROUND

Computer usage has traditionally been regarded as a rational activity in which emotions are not involved. This view, however, has been changing as the importance of emotions in all aspects of human thinking, activity and interaction is becoming more apparent. Human interactions do not just include those with other people, but also with their surroundings, including inanimate objects. One such object that has a big role in the day to day life of many people is the computer.

It is not uncommon for a person to spend more hours in a day interacting with a computer than face to face with other people. For this reason it is important to design computers that are user-friendly and easy to use (Preece et al., 1994). One important aspect of this drive towards user-friendliness is that the user should be able to use his or her natural way of interacting rather than having to learn new ways of working (Norman, 1988). The goal of improving the interaction between users and computers requires that emotions be taken into account in this interaction.

The field of HCI has greatly matured over the last several decades since the first conference on human factors in computing systems was held in the early 1980's. Since this time the emphasis within HCI has shifted from a focus on trained systems operators, to analyzing how technology influences the general user. To this end, there has been a substantial amount of attention devoted to

the concept of usability, as well as the role of the user in the development of successful interfaces. Usability is simply defined as “the extent to which a product can be used by specified users to achieve specific goals with effectiveness, efficiency, and satisfaction in a specified context of use” (International Organization for Standardization, 2010). This broad definition sets the stage for the fact that usability is a complex construct that can be influenced by a large number of external factors including context or environment.

In the early 80s, the role of a HCI specialist would be to evaluate interface components such as menus or terminology. As the field progressed, and the specialists came to realize the broader applicability of their work, new directions and specializations were created. The term “user-centered” is extensively used in the field of HCI (Karat & Karat, 2003) when describing approaches to building usable systems. For user-centered design, the main focus is that the needs of the user are used as a way to inform design (Vredenburg, Isensee, & Righi, 2001). This perspective is also sometimes referred to as human-centered design, or human centered computing (HCC). HCC broadly describes the methodology that would be applied to any field that uses computers in any form where users directly interact with them (Jaimes, Sebe, & Gatica-Perez, 2006). Thus HCC aims to integrate human sciences (such as cognitive and affective) into the existing body of computer science and HCI knowledge with a human focus throughout the lifecycle. HCC is said to incorporate social and cognitive sciences more closely than traditional HCI (Foley, 2006).

The recognition that interaction is not limited to simple interface modalities gives support to the development of new technologies. The ISO 9241 standard encapsulates this view in the following high level goal for user-centered design: “the design addresses the whole user experience” (International Organization for Standardization, 2010). This acknowledges that the HCC principles of cognitive and affective design are important

when developing usable software and systems. To this end, a successful user interface would have an understanding of what emotions are, how they can be identified and what the implications of various emotional patterns are for a given interaction situation.

The term “affective computing” was coined as long ago as 1997 by Rosalind Picard, who defined it as “computing that relates to, arises from, or deliberately influences emotions” (Picard, 1997a, p. x). This is the most comprehensive and widely used definition and is often cited. Picard, a pioneer in this field, reports that the initial response to the very concept of emotion-sensitive machines was somewhat lackluster (Picard, 2010), and it is interesting to observe the dramatic rise in interest from both developers and the research community in recent years.

Affective computer interfaces improve human-computer interaction by enabling the communication of the user’s emotional state. The growing interest in affective computing arises from findings in psychology and physiology which demonstrate the importance of emotional state in human behaviour (Partala & Surakka, 2004).

Emotion and cognition are linked and there is evidence of emotion influencing aspects of cognitive performance and decision making (Cytowic 1989; Eysenck et al. 2007). The interaction between affect and cognition is bi-directional, thus the underlying affective state of the individual will also influence the outcome of various cognitive processes. This, predictably, has far ranging implications. There is evidence that emotion has an impact on the speed at which information is processed (Öhman, 2001) and whether it is attended to (Anderson, 2001; Vuilleumier, 2001). Emotion also has a relation to motivation in that evaluations or feelings regarding the current situation will largely determine the action that is taken in response. Therefore, emotions are often precursors of motivations (Oatley, 1992). Memory is also impacted by emotional state, and again there are many mechanisms by which this

can occur (Reisberg & Hertel, 2003). Thus there are substantial potential benefits to be had from the development of user interfaces that support users' emotional, as well as cognitive processes, in their day to day work or life.

Currently, affective computing research is conducted in a large number of areas including education, autonomous agents, games and healthcare. There is also a rapidly growing body of published literature on the subject with 2010 marking the launch of the *IEEE Transactions on Affective Computing*. This is the first cross-disciplinary, international journal dedicated toward disseminating the results of research in areas such as theories of affective human-computer interaction systems, algorithms to detect and respond to emotions, and applications of affective computing.

AFFECTIVE HUMAN-COMPUTER INTERACTION

Although a computer will not actually experience emotions in the same way that a human would, the quality of interaction has been shown to improve even if the system *appears* to do so. Klein, Moon and Picard (2002) conducted a study in which users interacted with a computer system that was designed to deliberately elicit negative feelings of frustration. It did this by inserting random delays or periods of unresponsiveness to hinder the users from carrying out the goals of the study. The results demonstrated that if the computer system provided users with the ability to vent their frustrations (as a form of affect-support), users continued to interact with the frustrating system significantly longer than if no affect-support was provided. Empathic agents have also been successfully used in software to improve usability. Prendinger, Mayer, Mori and Ishizuka (2003) utilized an on-screen animated agent to provide empathic support to users who had been carrying out a frustrating computer interface task. Findings indicated that a character demonstrating empathy

may decrease users' levels of experienced stress. For these findings to be applicable for real-world affective interfaces, a key aspect of software functionality is the ability to detect the emotional state of the user.

Detection of Affective State

Mehrabian (1981) is often cited for his 7%-38%-55% rule of non-verbal communication which simply states that in human communication, 7% of the message is communicated by the words, 38% by tone of voice and the remaining 55% by body language. This is particularly relevant when we consider the expression of emotional content, or affect, which is largely non-verbal in nature. The use of subtle non-verbal communication methods is not only desirable but almost a necessity for a successful affect detection system to operate. Methods for inferring affective state are numerous but may be broadly categorized into a few areas, which are described below.

Self-Report Measures

A number of self-report measures of affect have been developed and used in research on mood and emotion; many of these share similar features but also have differences in the way that the responses are formatted and the way in which tests are conducted. Many of the most prominent self-report affective measures involve presenting lists of adjectives to the subjects, and obtaining a rating as to how appropriate or strong these particular emotions are (e.g. POMS (McNair, 1971) or PANAS (Watson, Clark, & Tellegan, 1988)). Self-report measures are a valuable instrument in the development and assessment of affective computer interfaces. However, as their use often requires the user to be interrupted or to recall a memory of an event, these have limited applicability for end-user applications which generally aim to produce the most natural interaction environment possible.

Observable Traits

The use of observations to infer the emotional state of an individual stems largely from the work of Ekman and Friesen (1978) who theorized a relationship between particular facial configurations and the underlying “basic” emotions present and later derived lists of facial expressions that would be used as markers for particular emotions. More recent work has called for other “non-basic” emotions to be considered within affect sensing systems, as different application domains may bring different emotional states into relevance (D’Mello & Calvo, 2013). Similar techniques may be applied to the observation of other features, such as the user’s posture or gestures. Observable traits have the benefit that an affective application may utilize this source of information with little or no user intervention. Subtle changes in facial expressions and posture occur without conscious effort on the part of the individual and may also reveal a deeper insight into the underlying affective state. One disadvantage of these methods is that the technical and implementation environment is often quite “noisy” and the success rate of automatically detecting affective state from natural expressions or gestures may be impaired.

Psychophysiology

Researchers have become increasingly aware that a critical component of emotion is physiological activity. According to some theories, if there is no physiological reaction there is no emotion (e.g. Schachter & Singer, 1962). It is theorized that every psychological event or affective state has some physiological referent (Cacioppo & Tassinary, 1990), therefore the issue is not so much of whether or not a physiological signal is present, but rather which aspects of emotion may be inferred from this signal. There are many advantages to this approach. Physiological signals are unconscious and do not carry any of the subjectivity of self-report measures, furthermore they bring about the potential for real time measurement with no need

to interrupt or otherwise distract the user. Finally, as technology advances, physiological sensors may be suitable for incorporating into existing physical interfaces to ensure a more natural interface which the user need not be constantly aware of. Recent developments have permitted the integration of physiological sensing into wearable form factors (e.g. Lanatà, Valenza, & Scilingo, 2012) and even enabled physiological sensing from existing commodity devices such as smartphones (Hernandez, McDuff, & Picard, 2015).

APPLICATIONS OF AFFECTIVE COMPUTING

Affective computing applications have potential uses in practically any situation where a human-computer interaction is taking place. Technology that can recognize and even express affect can provide insights into human-computer (and in some cases human-human) interactions. Measuring the stress or difficulty caused by a system may also allow developers to pinpoint problems, or simply allow the system to be improved by being able to respond in a more natural and realistic way. Such technologies have been successfully implemented in very diverse environments. These include robotic personas (Breazeal, 2003), learning companions (D’Mello, Lehman, & Graesser, 2011; Sarrafzadeh, Alexander, Dadgostar, Fan, & Bigdeli, 2008), games (Gilleade, Dix, & Allanson, 2005) or wearable computers (Picard, 1997b)..

Sociable humanoid robots present a novel and under explored area of human-machine interaction. Traditionally robots have been utilized for functional roles such as automation or inspections, in roles that require autonomy and minimization of human interaction. However, as robots may be used for any number of applications, including domestic use, the requirement for natural and usable human-machine (in this case robot) interaction, presents itself.

Humanoid robots may express affective states in a number of ways. Robotic characters, if human

like in appearance, are well suited to affective expression as they can communicate in ways which emulate the natural communication modalities of humans. This can include non-verbal communication such as gestures, facial expression or body positioning. One example is Kismet: a socially expressive anthropomorphic robot. Kismet perceives a variety of cues from visual and auditory channels and delivers affective information back to the human through gaze direction, facial expression, body posture, and vocalization (Breazeal, 2002). The robot possesses a computation model of emotion, in which affective state is sensed, used as part of the internal decision (action) making strategy, and communicated to the human.

As mentioned, animated agents may also be used in a similar way to communicate affective content and to put the user in a positive and constructive affective state to maximize enjoyment, entertainment, learning or productivity. Education is an area in which affective computing applications have shown promise, partly due to the increasing reliance on online and computer mediated learning and teaching strategies. Goleman (1995) reported that expert teachers are able to recognize emotional states of students, and respond appropriately to positively impact learning. Whilst the way in which this is accomplished is not well documented a key element involves the recognition of negative affect or states that are detrimental to learning and guiding the learner into a more positive and constructive state. Intelligent tutoring systems incorporating an emotional or affective model are known as affective tutoring systems. An affective tutoring system is thus any tutoring system that can adapt to perceived emotion. This may be to respond to any negative emotions being experienced by the learner, or to interact in a manner that is more natural and engaging for the learner.

A number of such affective tutoring systems exist, that respond to different types of input. For example, AutoTutor is an intelligent tutoring system that interacts with learners using natural language and helps them to construct explanations

in simulation environments (Graesser, McDaniel, & Jackson, 2007). It detects the learner's affective state using physiological and facial expression analysis and conversational cues. Easy with Eve (Alexander, Sarrafzadeh, & Hill, 2006; Sarrafzadeh, et al., 2008) is an affect sensitive mathematics tutor. Affect recognition is performed by video analysis to capture facial expression and gesture information from the user. These systems have also been shown to be effective and result in increased learning, however are still not as effective as a one-to-one human tutor.

A number of affective games have been developed to encourage the user to express their affective states and to dynamically respond and adapt to this form of input. For example, in SenToy, the user interacts with a doll to communicate one of six emotions through gestures (Paiva, 2003). Other games utilize biofeedback to guide the user into certain affective states, for example in the game "The Wild Divine" (2012) the player needs to achieve a state of relaxation to interact and progress within the game environment. Affective interaction and expression has been identified as a valuable direction of research to develop more engaging and realistic games (Hudlicka, 2009).

Wearable computers provide a rich and diverse ground for evaluating and implementing affective technologies. The close contact with the user enables easy communication of subtle non-verbal cues that may be valuable indicators of affective state. In some cases, the affect detection capabilities may even be used to improve the users' own abilities to perceive emotions in others, and thus improve human-human communication. For example "expression glasses", developed at MIT provide the wearer with feedback about the emotional expressions of others (Scheirer, Fernandez, & Picard, 1999). This technology may improve the quality of life for those with autism or other disorders that impair human-human communication (el Kaliouby, Picard, & Baron-Cohen, 2006). Existing devices that are in close contact with the user may also have potential to be used as affect sensing devices, for example a mouse may sense

the user's stress levels (Kirsch, 1997), or a car steering wheel may sense when the user is falling asleep or identify lapses in attention (Gusikhin, Filev, & Rychtycky, 2008).

ISSUES

The diverse nature of affective computing applications described in the previous section highlights that this technology is successful and adaptable to a wide range of situations with positive outcomes. However, as many of the implementations are disparate and “one-off” in nature, it is potentially difficult to transfer findings from one particular domain to a new application. Furthermore, as this is a new and emerging field, there is little evidence of “shared best practice” aside from the high level principles that have been established regarding inference of affective state.

Affective computing applications are often built in the same way as more traditional applications, with the affective functionality inserted into the program architecture wherever the developer considers it appropriate. Consequently, the current trend for ad-hoc development in affective computing is hampering progress. Allanson and Fairclough (2004) noted that research in the area was disparate and uneven, and it seems that little progress has been made since then. One goal that has been identified in the literature is that of “device-independence” – any successful solution to the issue must be capable of abstracting over multiple implementation environments which may have different outputs, manufacturers and operating requirements.

Due to the diverse nature of implementation, and the many methods by which affective information may be gathered, affective computer interface components often need to deal with many different types of data (all with different characteristics and requirements for processing). This often presents a complex signal processing task, which involves a number of stages from extraction of the raw signal to analysis and transformation of the data into a

computer input with well understood parameters. For this reason it may be necessary to add a layer of signal processing between the intelligent sensors and the interface to limit the complexity of the interaction techniques (Allanson, 2000).

It becomes apparent that the one commonality amongst affective computing applications is the extent to which developments are unique and tied to a particular implementation. Hamming (1969) stated that “a central problem in all of computer science is how we are able to get to the situation where we build on top of the work of others rather than redoing much of it in a trivially different way” (p.10) – an observation that is valid to this day. As affective applications are highly specialized and complex, there has to date been no discussion regarding the concept of reusing existing affective applications in new problem domains and situations. Furthermore, the potential for adding affect support as an additional layer above existing software has not been investigated thoroughly to date. Aist, Kort, Reilly, Mostow and Picard (2002) demonstrated the utility of adding emotional support to an existing tutoring system, and noted that this approach may be useful for future developments. Certainly, the ability to augment existing software with affect sensing capabilities could for the most part turn the entire operating system and all its application software and tools into an affective computing application. This would be a breakthrough for those who envision affective computing as being a part of the entire computing experience rather than the domain of a few isolated applications

FUTURE TRENDS

In a relatively new field such as affective computing, theories regarding emotion, tools, methods and software are constantly evolving and improving as our knowledge grows. The rapidly developing body of research in the field of affective computing gives a clear indication that affective computing is going to play a major role in the future of human-computer interaction.

The domain of affect sensing technologies is one which will directly be improved as advances in computer technology take place. The increased processing power and portability of modern computing devices makes advanced signal processing and affective pattern recognition, more feasible. Furthermore, as computers become ubiquitous and become integrated into vehicles, clothing and our surroundings, the opportunity for greater physical contact between user and machine increases and makes new input paradigms increasingly viable.

Education is an area in which applications of affective computing are highly applicable and substantial research has been carried out in this area. The fact that interaction with computers is a fundamental part of study in most disciplines renders this a prime candidate for affective computing developments. Endowing a computer with the ability to respond to affective state should enhance learning outcomes and have a positive impact on the user experience of e-learning. Furthermore with the increased dependence on online learning, such technology may be in even more demand as teachers no longer have access to students' non-verbal cues in classrooms (Crosby, Brent, Aschwanden, & Ikehara, 2001). Interest in the educational implications of affective computing is not limited to the academic research community. In 2012, industry analysts Gartner Research discussed how the field is on the rise in education. Whilst most of the affective tutoring systems are in the proof of concept stage, the advice given to education institutions is to track the progress and developments in the field and that those with a large online presence should immediately get involved. Affective computing is described as having "the potential to bring back a bit of the lost pedagogical aspect of in-classroom learning and increase the personalization of online learning" (Lowendahl, 2012, p. 15).

CONCLUSION

Affective computing facilitates more intuitive, natural computer interfaces by taking into account the emotional state of the user. As such, affective computing holds great promise for improving human-computer interaction. This chapter has provided an insight into the field of affective computing: covering the origins of the field, the underlying elements of affective computing, and highlighting issues in the field. Detailed examples of a diverse range of affective computing applications are provided and future directions have been identified.

REFERENCES

- Aist, G., Kort, B., Reilly, R., Mostow, J., & Picard, R. (2002). Experimentally augmenting an intelligent tutoring system with human-supplied capabilities: Adding human-provided emotional scaffolding to an automated reading tutor that listens. *Proceedings of the 4th IEEE International Conference on Multimodal Interfaces*, 483-490. doi:10.1109/ICMI.2002.1167044
- Alexander, S., Sarrafzadeh, A., & Hill, S. (2006). Easy with Eve: A functional affective tutoring system. In G. Rebolledo-Mendez & E. Martinez-Miron (Eds.), *Proceedings of Workshop on Motivational and Affective Issues in ITS. 8th International Conference on ITS* (pp. 38-45). Academic Press.
- Allanson, J. (2000). *Supporting the development of electrophysiologically interactive computer systems* (Unpublished doctoral dissertation). Lancaster University, Lancaster, UK.
- Allanson, J., & Fairclough, S. H. (2004). A research agenda for physiological computing. *Interacting with Computers*, 16(5), 857-878. doi:10.1016/j.intcom.2004.08.001

- Anderson, A. K., & Phelps, E. A. (2001). Lesions of the human amygdala impair enhanced perception of emotionally salient events. *Nature*, 411(6835), 305–309. doi:10.1038/35077083 PMID:11357132
- Breazeal, C. (2002). *Designing sociable robots*. Cambridge, MA: MIT Press.
- Breazeal, C. (2003). Emotion and sociable humanoid robots. *International Journal of Human-Computer Studies*, 59(1-2), 119–155. doi:10.1016/S1071-5819(03)00018-1
- Cacioppo, J. T., & Tassinary, L. G. (1990). Inferring psychological significance from physiological signals. *The American Psychologist*, 45(1), 16–28. doi:10.1037/0003-066X.45.1.16 PMID:2297166
- Crosby, M. E., Brent, A., Aschwanden, C., & Ikehara, C. (2001). Physiological data feedback for application in distance education. *Proceedings of the 2001 Workshop on Perceptive User Interfaces*, 1-5. doi:10.1145/971478.971496
- D'Mello, S., & Calvo, R. A. (2013). *Beyond the basic emotions: what should affective computing compute?* Paper presented at the CHI '13 Extended Abstracts on Human Factors in Computing Systems. doi:10.1145/2468356.2468751
- D'Mello, S., Lehman, B., & Graesser, A. (2011). A motivationally supportive affect-sensitive AutoTutor. In R. A. Calvo & S. K. D'Mello (Eds.), *New perspectives on affect and learning technologies* (Vol. 3, pp. 113–126). New York: Springer. doi:10.1007/978-1-4419-9625-1_9
- Ekman, P., & Friesen, W. (1978). *Facial action coding system: A technique for the measurement of facial movement*. Palo Alto, CA: Consulting Psychologists Press.
- el Kaliouby, R., Picard, R. W., & Baron-Cohen, S. (2006). Affective computing and autism. *Annals of the New York Academy of Sciences*, 1093(1), 228–248. doi:10.1196/annals.1382.016 PMID:17312261
- Foley, J. (2006). *HCC Education Digital Library*. Retrieved 5th June, 2013, from <http://hccedl.cc.gatech.edu/>
- Gilleade, K., Dix, A., & Allanson, J. (2005). Affective videogames and modes of affective gaming: Assist me, challenge me, emote me. *Proceedings of the DIGRA*.
- Goleman, D. (1995). *Emotional intelligence*. New York: Bantam Books.
- Graesser, A. C., McDaniel, B., & Jackson, G. T. (2007). Autotutor holds conversations with learners that are responsive to their cognitive and emotional states. *Educational Technology*, 47, 19–22.
- Gusikhin, O., Filev, D., & Rychtycky, N. (2008). Intelligent vehicle systems: Applications and new trends in control automation and robotics. In J. A. Cetto, J.-L. Ferrier, J. M. Costa dias Pereira & J. Filipe (Eds.), *Lecture Notes in Electrical Engineering* (Vol. 15, pp. 3-14). Berlin: Springer Heidelberg.
- Hamming, R. W. (1969). One mans view of computer science. *Journal of the ACM*, 16(1), 3–12. doi:10.1145/321495.321497
- Hernandez, J., McDuff, D. J., & Picard, R. W. (2015). Biophone: Physiology monitoring from peripheral smartphone motions. *Proceedings of the Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, 7180-7183.
- Hudlicka, E. (2009). Affective game engines motivation and requirements. *Proceedings of the 4th International Conference on Foundations of Digital Games*, 299-306.
- International Organization for Standardization. (2010). *ISO 9241-210:2010 Ergonomics of human-system interaction -- Part 210: Human-centred design for interactive systems*. Retrieved 2nd June 2012, from http://www.iso.org/iso/catalogue_detail.htm?csnumber=52075

- Jaimes, A., Sebe, N., & Gatica-Perez, D. (2006). Human-centered computing: a multimedia perspective. *Proceedings of the 14th annual ACM international conference on Multimedia*, 855-864. doi:10.1145/1180639.1180829
- Karat, J., & Karat, C. M. (2003). The evolution of user-centered focus in the human-computer interaction field. *IBM Systems Journal*, 42(4), 532–541. doi:10.1147/sj.424.0532
- Kirsch, D. (1997). *The Sentic Mouse: Developing a tool for measuring emotional valence*. Retrieved Nov 5th, 2012, from http://affect.media.mit.edu/projectpages/archived/projects/sentic_mouse.html
- Klein, J., Moon, Y., & Picard, R. W. (2002). This computer responds to user frustration: Theory, design and results. *Interacting with Computers*, 14(2), 119–140. doi:10.1016/S0953-5438(01)00053-4
- Lanatà, A., Valenza, G., & Scilingo, E. P. (2012). A novel EDA glove based on textile-integrated electrodes for affective computing. *Medical & Biological Engineering & Computing*, 50(11), 1163–1172. doi:10.1007/s11517-012-0921-9 PMID:22711069
- Lowendahl, J.-M. (2012). *Hype Cycle for Education*. Retrieved 1st September, 2012, from http://www.gartner.com/DisplayDocument?doc_cd=233974&ref=g_sitelink
- McNair, D., Lorr, M., & Dropplemen, L. (1971). *Edits manual: Profile of mood states*. San Diego, CA: Educational and Industrial Testing Services.
- Mehrabian, A. (1981). *Silent Messages: Implicit communication of emotions and attitudes*. Belmont, CA: Wadsworth.
- Norman, D. (1988). *The design of everyday things*. New York: Doubleday.
- Oatley, K. (1992). *Best laid schemes: The psychology of emotions*. Cambridge, UK: Cambridge University Press.
- Öhman, A., Flykt, A., & Esteves, F. (2001). Emotion drives attention: Detecting the snake in the grass. *Journal of Experimental Psychology*, 130(3), 466–478. doi:10.1037/0096-3445.130.3.466 PMID:11561921
- Paiva, A., Costa, M., Chaves, R., Piedade, M., Mourão, D., Sobral, D., & Bullock, A. et al. (2003). SenToy: An affective sympathetic interface. *International Journal of Human-Computer Studies*, 59(1-2), 227–235. doi:10.1016/S1071-5819(03)00048-X
- Partala, T., & Surakka, V. (2004). The effects of affective interventions in human–computer interaction. *Interacting with Computers*, 16(2), 295–309. doi:10.1016/j.intcom.2003.12.001
- Picard, R. W. (1997a). *Affective computing*. MIT Press. doi:10.1037/e526112012-054
- Picard, R. W. (2010). Affective Computing: From laughter to IEEE. *IEEE Transactions on Affective Computing*, 1(1), 11–17. doi:10.1109/T-AFFC.2010.10
- Picard, R. W., & Healey, J. (1997b). Affective wearables. *Personal Technologies*, 1(4), 231–240. doi:10.1007/BF01682026
- Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S., & Carey, T. (1994). *Human-computer interaction*. Addison-Wesley.
- Prendinger, H., Mayer, S., Mori, J., & Ishizuka, M. (2003). Persona effect revisited. Using bio-signals to measure and reflect the impact of character-based interfaces. *Proceedings of the Fourth International Working Conference on Intelligent Virtual Agents (IVA-03)*, 283-291.
- Reisberg, D., & Hertel, P. (2003). *Memory and emotion*. Oxford University Press.
- Sarrafzadeh, A., Alexander, S., Dadgostar, F., Fan, C., & Bigdeli, A. (2008). How do you know that I dont understand? A look at the future of intelligent tutoring systems. *Computers in Human Behavior*, 24(4), 1342–1363. doi:10.1016/j.chb.2007.07.008

Schachter, S., & Singer, J. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychological Review*, 69(5), 379–399. doi:10.1037/h0046234 PMID:14497895

Scheirer, J., Fernandez, R., & Picard, R. W. (1999). Expression glasses a wearable device for facial expression recognition. *Proceedings of the Computers in Human Interaction, CHI*, 99, 262–264.

Vredenburg, K., Isensee, S., & Righi, C. (2001). *User centered design: An integrated approach*. Upper Saddle River, NJ: Prentice Hall.

Vuilleumier, P., & Schwartz, S. (2001). Beware and be aware: Capture of spatial attention by fear-related stimuli in neglect. *Neuroreport*, 12(6), 1119–1122. doi:10.1097/00001756-200105080-00014 PMID:11338176

Watson, D., Clark, L. A., & Tellegan, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, 54(6), 1063–1070. doi:10.1037/0022-3514.54.6.1063 PMID:3397865

Wild Divine. (2012). *The Journey to Wild Divine*. Retrieved 1 Nov 2012, from <http://www.wilddivine.com>

ADDITIONAL READING

Allanson, J., & Fairclough, S. H. (2004). A research agenda for physiological computing. *Interacting with Computers*, 16(5), 857–878. doi:10.1016/j.intcom.2004.08.001

Ball, G., & Breeze, J. (2000). Emotion and personality in a conversational agent. In J. Cassel, J. Sullivan, S. Prevost, & E. Churchill (Eds.), *Embodied conversational agents* (pp. 189–219). Cambridge, MA: The MIT Press.

Baylor, A. L., & Rosenberg-Kima, R. B. (2006). Interface agents to alleviate online frustration. *Proceedings of the 7th International Conference of the Learning Sciences*, Bloomington, IN. 30–36

Beale, R., & Peter, C. (2008). The role of affect and emotion in HCI. In P. Christian & B. Russell (Eds.), *Affect and Emotion in Human-Computer Interaction* (pp. 1–11). Springer-Verlag. doi:10.1007/978-3-540-85099-1_1

Cacioppo, J. T., & Tassinary, L. G. (1990). Inferring psychological significance from physiological signals. *The American Psychologist*, 45(1), 16–28. doi:10.1037/0003-066X.45.1.16 PMID:2297166

Cacioppo, J. T., Tassinary, L. G., & Berntson, G. G. (Eds.). (2007). *Handbook of psychophysiology* (3rd ed.). New York: Cambridge University Press. doi:10.1017/CBO9780511546396

Coan, J. A., & Allen, J. J. B. (Eds.). (2007). *Handbook of emotion elicitation and assessment*. New York: Oxford University Press.

Fulda, J. S. (1998). Giving computers emotions - Why and how. *ACM SIGCAS Computers and Society*, 28(4), 30–31. doi:10.1145/308364.606115

Goleman, D. (1995). *Emotional intelligence*. New York: Bantam Books.

Klein, J., Moon, Y., & Picard, R. W. (2002). This computer responds to user frustration: Theory, design and results. *Interacting with Computers*, 14(2), 119–140. doi:10.1016/S0953-5438(01)00053-4

Marcus, A. (2003). The emotion commotion. *Interaction*, 10(6), 28–34. doi:10.1145/947226.947239

Mehrabian, A. (1981). *Silent Messages: Implicit communication of emotions and attitudes*. Belmont, CA: Wadsworth.

Norman, D. (1988). *The design of everyday things*. New York: Doubleday.

Partala, T., & Surakka, V. (2004). The effects of affective interventions in human-computer interaction. *Interacting with Computers*, 16(2), 295–309. doi:10.1016/j.intcom.2003.12.001

Picard, R. W. (1997). *Affective computing*. MIT Press. doi:10.1037/e526112012-054

Picard, R. W. (2010). Affective Computing: From Laughter to IEEE. *IEEE Transactions on Affective Computing*, 1(1), 11–17. doi:10.1109/T-AFFC.2010.10

Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S., & Carey, T. (1994). *Human-computer interaction*. Massachusetts: Addison-Wesley.

Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. New York: Cambridge University Press.

Riseberg, J., Klein, J., Fernandez, R., & Picard, R. W. (1998). Frustrating the user on purpose: using biosignals in a pilot study to detect the user's emotional state. *Proceedings of the CHI '98 Conference on Human Factors in Computing Systems*, Los Angeles, United States. 227-228 doi:10.1145/286498.286715

Sarrafzadeh, A., Alexander, S., Dadgostar, F., Fan, C., & Bigdeli, A. (2008). How do you know that I dont understand? A look at the future of intelligent tutoring systems. *Computers in Human Behavior*, 24(4), 1342–1363. doi:10.1016/j.chb.2007.07.008

Scherer, K. R., Banziger, T., & Roesch, E. (Eds.). (2010). *A blueprint for affective computing*. Oxford: Oxford University Press.

Sears, A., & Jacko, J. (2008). *The Human-Computer Interaction handbook*. New York: Lawrence Erlbaum Associates.

Serbedzija, N. B., & Fairclough, S. H. (2009). Biocybernetic loop: From awareness to evolution. *Proceedings of the IEEE Congress on Evolutionary Computation, CEC '09.*, 2063-2069. Trondheim, Norway.

Shneiderman, B. (2003). *Leonardo's laptop: Human needs and the new computing technologies*. Massachusetts: MIT Press.

Thompson, N., & McGill, T. (2012). Affective Tutoring Systems: Enhancing e-learning with the emotional awareness of a human tutor. *International Journal of Information and Communication Technology Education*, 8(4), 75–89. doi:10.4018/jicte.2012100107

KEY TERMS AND DEFINITIONS

Affective Computing: Defined by Picard (1997a) as “computing that relates to, arises from, or deliberately influences emotions”.

Affective Human-Computer Interaction: Affective HCI incorporates the communication of affective state information as an interface modality. This aims to enrich the quality of interaction and permit the user to employ more intuitive methods of communication.

Affective State: This term refers to the experience of feeling the underlying emotional state. The description often distinguishes between the more diffused longer term experiences (termed moods) and the more focused short term experiences (termed emotions).

Human-Computer Interaction (HCI): The study of how users interact with computer based devices. This includes techniques for assessing elements of the effectiveness or ease of use of an interface as well the development of more intuitive and natural interfaces.

Psychophysiology: Research suggests that all underlying affective states have some physiological manifestation that may be subtle, but potentially observable. The field of psychophysiology bridges the domains of psychology and physiology with the study of how these aspects of human experience interact.

User-Centered Design: A type of user interface and interaction design in which the main focus is that the needs of the user are used as a way to

inform design. This often involves a participatory or cooperative design approach in which designers and users work collaboratively.

Wearable Computers: Any portable, miniature devices that are computer based and worn by

the user as part of their clothing or accessories. Increasing miniaturization and widespread use of portable computers (including smartphones) makes this a viable and promising domain in which affective computer interfaces may be developed.

H

Computer-Assisted Indian Matrimonial Services

Robert Leslie Fisher

Independent Researcher, USA

INTRODUCTION

People seeking romantic relationships have advertised for partners on the internet in both developed countries (e.g. U.S.A.) and developing countries such as India since that first became possible twenty or more years ago. However, advertising patterns differ among societies with more advertisements in the United States by people looking for short term relationships (as evidenced by the success of such sites as Tinder)), and more advertisements in India from people looking for suitable spouses in a society where overwhelmingly people have arranged marriages (Krishnan 2012).

Thanks to its complex nature the matrimonial decision in India, which has great cultural importance, has stimulated the rise of a large thriving industry of computer assisted matchmaking services catering to the millions of Indians searching for suitable spouses. This phenomenon, in turn, has attracted the interest of a variety of disciplinarians, including information scientists and social scientists.

This chapter's focus is sociological: it concerns the way the computer has been harnessed to assist both an ancient custom of arranged marriage among Hindus and the special occupational group known as matchmakers. The principal argument is that computer assisted matchmaking, thanks to the computer's virtually unlimited memory, improves the chances that the marriage contracted by a couple utilizing the service will be durable and happy because it offers the possibility for fine grained analysis.

From a sociological standpoint, it is maintained here that Indian matchmaking and medical diag-

nosis are fundamentally similar phenomena in that both are the result of non-random, non-rational processes that can be efficiently analyzed using *the garbage can model* of Cohen, March and Olsen, as modified by Zeldenrust (1990) and Fisher (2005). The model itself is a kind of dynamic input-output model that emphasizes flows into and out of the "garbage can" and storage within the "garbage can." Since its development in the early 1970s the *garbage can model* has been applied to study "decision making in organizational anarchies" (1972) and government planning (Kingdon, 1984); and in the modified form employed here it has been used to study research problem choices of scientific research teams in the Netherlands (1990) and in the United States and Canada (2005); and to study how to improve medical diagnosis (2014).

Although this chapter focuses on India, some conclusions may apply to other societies where arranged marriages also occur: in Japan (see Blood, 1967), among European royalty from the Middle Ages onward, (see, e.g. Harris, 1989); and among such groups as the *haridim* (ultra pious Jews). Elsewhere, e.g. among the wealthiest families in the United States, elements of arranged marriage survive in such traditions as debutante balls whose purpose is to enhance the probability that the eligible children of the wealthy elite meet and fall in love with suitable mates, i.e. eligible singles from "good" families. However, as Krishnan points out, in India, the tradition of arranging marriages has survived the homogenization of culture of the modern era (exemplified by the popularity of "pizza, burgers, denim and rock music in India as in Europe and North America") with the result that "in most Indian communities" people

feel “the need to conform to traditional marital practices” (p.18).

This chapter first briefly considers the complexity of the marriage decision as a consequence of constraints that must be satisfied, especially as it bears on the variety of data to be collected and analyzed. In India, the first constraint is that the marriage ceremony be performed as required by the tenets of the Hindu religion. This can be hugely expensive. According to Yee (2008), “the ceremonies are very colorful, and celebrations of the nuptials may extend for several days.”

Another constraint is the cost in time spent on the search for a suitable mate. Neither the family of the person searching for a mate for their eligible offspring nor the prospective bride (or groom) want a long drawn out search. Therefore, people seeking the services of a matchmaker or a computer assisted matchmaking service, will insist on a relatively limited group of highly desirable spouses, perhaps no more than six from whom to select the spouse. Unsurprisingly, given India’s billion plus population the computer matchmaking services have accumulated data about millions of marriage eligible individuals to cater to this demand.

A third constraint is that the negotiations between families can be similar in complexity to those between Indian “business groups”—somewhat akin to conglomerates or loosely affiliated companies that share key personnel (e.g. senior managers or owners). The latter for instance engage in lengthy, complex negotiations arising from a desire to merge or coordinate in some manner (Khanna & Yafeh, 2015).. Given the constraints and the importance for families of both bride and groom of a suitable spouse for their progeny, if information scientists can adequately address the technical problems of collecting, storing, and analyzing data on tens of millions of potential mates, it stands to reason computer assistance would be popular among marriage minded Indians.

How can *the garbage can model* illuminate the dynamics of the computer assisted arranged marriage in India? The model’s critical feature

relates to arriving at decisions based on imperfect and incomplete information in a nonrandom non-rational process: selection of a mate as a key societal decision to assure the replacement of deceased members of society.

H

BACKGROUND

What are the essential features of the decision to marry in the Indian context? First, without any slight intended to the many varieties of relationships among people, this chapter defines marriage as the union of a man and a woman with the intention of procreating and rearing children.

Selecting a mate given this intention becomes a non-rational decision made under conditions of imperfect information because the sexual compatibility necessary to procreate is unknown until conception has occurred. Selecting a mate in this context is also a non-random decision. Whether the decision to marry is preceded by rational calculation of the decision maker(s) or “chemistry,” the choice is definitely not random or probabilistic (though perhaps it could be modeled stochastically). It is socially patterned or cultural in a broadly understood way because the criteria of choice reflect societal standards of what is desirable in a mate etc. Finally, except perhaps in isolated cases it is not the decision of only the individuals who are to be married but of the organizational unit—the family, or more accurately, the families (a collectivity of organizations)—of the people to be married.

Krishnan illustrates this critical point noting that in her own case her “grandparents’ alliance was arranged solely by the family patriarchs.” The marriage of Krishnan’s parents “was initiated by the family elders but was approved by both of the [people to be wed]” while Krishnan herself and her future husband “initiated their relationship themselves and later received approval from both sets of parents.

As Krishnan points out, the participation of all stakeholders in the decision to marry even

when the romantic relationship is initiated by the people to be married has persisted over a long time, thus suggesting that marriage, beyond replacing members of a society, has other societal functions in India as indicated by Prakasa and Rao (1979) cited in Krishnan:

1. Maintain the social stratification system.
2. Give parents control over family members.
3. Enhance the probability of preserving the ancestral line.
4. Allow for the consolidation and extension of family property.

One further pertinent observation: while the uncertainty inherent in the decision to marry is never eliminated, the importance to the organizational units (i.e., the families of the marriageable individuals) of this decision is a strong inducement for them to seek specialized assistance in identifying suitable mates and perhaps negotiating the best terms possible for acquiring these mates for their progeny. In ancient times and continuing into the present the families in India (and elsewhere) would turn to a matchmaker who would gather information about potential mates for consideration by the family elders. Nowadays, that same role performed by a machine assisted matchmaking service—available from many companies in India—allows users to upload a profile describing themselves and the desirable traits they are looking for in a marriage partner. Unlike *social network sites* the matchmaking services use a special software, typically proprietary, to match users to potential mates who have also uploaded their profiles and the traits they are looking for in a partner.

Krishnan reports that the “web-based” matrimonial services, (called “*matrimonial portals*”) became popular almost instantaneously. (Also see Goel [2006]; Jana [2000]; Pepper [2007]; Philip [2005]; Ramalingam and Nair [2006]; and Sharma [Ph.D. diss. 2006] referenced in Krishnan) She credits three forces for this: “cable television; several international companies that entered In-

dia as a result of India’s globalization policies in the 1990s” (also see Kulkarni [2005]); and “the prominent role of Indian software engineers ... in the Information Technology revolution and Silicon Valley.” Together they “deeply influenced people’s attitudes and behaviors and hastened the adoption of the Internet by “both Indians living overseas and in India” Krishnan says, citing Shrama (2006). Indeed, while the decision for most people to marry is not entirely in the hands of family elders (as in ancient India and perhaps among the most traditional segments of society in India today), their involvement in the decision makes the choice of mate *a social fact* (Durkheim, 1982) *requiring explanation by sociological theories*.

MAIN FOCUS

In the following two sections first the theoretical model is presented and then how it applies to the issue of computer assisted matrimonial services is demonstrated.

a. The Garbage Can Model

The *garbage can model* which addresses *choices* in conditions of uncertainty (and indeed as shown above the decision to marry fits that description) seems promising. Nonrandom, non-rational actions are basically *choices*, which cannot be addressed by *utilitarian* approaches as they require perfect *information*.

The *garbage can model* accommodates elements of uncertainty, mainly random and stochastic events or inputs, involving actions and elements that may lack rational basis. The decision elements, notably problems, solutions, participants, settings and choice situations, pour into the organizational garbage can haphazardly (Miller and Wilson [(2006)]).

Zeldenrust (1990) emphasized that a decision (in his case a research problem choice) made in this environment of imperfect information has among its characteristics, that it results from a fortuitous

confluence of factors. That decision is a highly contextual event, depending substantially on the patterns of flows in the several streams. Problems may have more solutions; or fewer solutions may apply to more problems; or problems and solutions may emerge, or disappear, or become refined or redefined over time (an evolutionary process). And he noted that adopted solutions may not even solve a problem at hand.

The process of reaching a decision creates an *evolutionary trajectory* which becomes modified by the addition of informational items that may reflect uncertainties, opinions, *cultural patterning*, and, can themselves be subject to a cluster of constraints which are legal, situational, political, social, and psychological. Where the *garbage can model* works a confluence brings a decision-- perhaps unforeseen-- and which may not solve the problem to which it is attached. Furthermore, it may reflect politics, a certain amount of bargaining, and even a partial rational analysis (Mohr [1976]). The *garbage can model* downplays these emphasizing instead the *processes of going, coming and waiting*. For this seemingly chaotic situation to exist an uncoupling of the means and ends of decisions making are needed (Miller and Wilson referencing Weick [1976]). The uncoupling of process streams permits solutions to seek problems and vice versa.

The key modification of the *garbage can model* needed to apply it to a match--a choice by two respective actors or more to unite an eligible man and woman in marriage--is in the basic assumptions that flows are loosely coupled or uncoupled. Zeldenrust says there may be a *systematic coupling* in investigated cases. One or more flows may be *stable* or *consistent*. On occasion, the flow of problems may consist of just one, stable, consistent problem. Some choices may 'constrain or enable subsequent choices' Thus, as Fisher and Fisher (2014) note there may be "more stability, logic, or consistency" in decisions than foreseen by Cohen, March, and Olsen.

Beyond any necessary relaxing of assumptions about the coupling of flows if a *garbage*

can model is applied to a particular situation it is also necessary to have: (1) "choice arenas" such as places where families can go to get professional assistance in the hunt for a mate (these places may exist only in cyberspace); and (2) streams independent of or only loosely coupled to each other. One stream must be a demand stream (e.g. for treatment of the illness in the case of medical diagnostics or lack of a suitable mate in the case of matchmaking); another might be a solutions stream (knowledge of disease sufficient for a treatment protocol to be possible or a data base sufficiently large and diverse of possible mates for the matchmaker).

Constraints, individual or clustered, have either their own streams *or* become mathematical boundary conditions. Zeldenrust lists two main properties of constraints: *robustness* (can a constraint be overcome and [if yes] at what cost?); and, *specificity* (how well defined are they?; and also level, with reference to quality criteria, "how high is the standard?").

Only if *all the streams link* (an evolutionary threshold criterion) will a match emerge. In the case of a marriage the model says it emerges only if all the constraints are satisfied *and* the solution indeed meets the demand. This assumes that the demand has neither disappeared nor changed. In the arranged marriage context a changed demand requires a new potential candidate for a match which becomes final only by meeting the demand and satisfying all the constraints. This final result may in fact be the first of the conjectured tentative choices, or a later one, or altogether different from the set originally considered by the matchmaker. It may change if another match maker has entered the picture and applied a different set of constraints and solution streams. In the medical context, Sadegh-Zadeh (1982) has discussed this issue at length noting that uncertainty permeates even a diagnostic statement as when a physician informs the patient of his opinion in carefully couched language, e.g. 'I believe that' Or 'I consider it possible that'

b. Characteristics of the Matchmaking Process

In the matchmaking process as defined here it is assumed that the two people to be joined in marriage are both legally adults. This is quite different from the betrothal of underage children—a situation in which matchmaking often occurred in ancient India and still may occur in many cultures. As Krishnan deftly puts it, “because people in ancient times married very young, it made sense for the parents to make the decision to choose their children’s life partner. The decisions of the elders also were considered significant because of the newly-wed couple’s lack of physical or psychological resources (Bhagat, 2002).

In the case of two adults of childbearing age, as Krishnan remarks, “most people insist on interacting, meeting, or even courting before marriage.” It is this situation that underpins the process of computer assisted matchmaking.

In matchmaking, the matchmaker has gathered information about possible mates from other families to share with a client. It is advantageous to both client and matchmaker to have a huge data base of prospective mates because someone seeking a mate is likelier to be encouraged to seek a mate if they believe the odds of finding a mate are favorable. And the larger the pool of marriageable members of the opposite sex, especially if a great deal of pertinent data exists for all members of the pool, the more favorable the odds will seem to the client. In India, as Krishnan notes, certain attributes are likely to appear in the files of the matchmaker services: “age, education, salary, physical characteristics (height, weight, complexion), religion, community, personality attributes, and interests.” This information not only is provided about the potential client and the desirable match for that individual. Therefore, when finding a match, the matchmaker will exclude from the pool of potential matches for the client anyone who desire characteristics the matchmaker’s client does not have. For instance, if the potential match requires a Brahmin and the

client is not, that potential match will be excluded from the client’s initial list.

Some critics of match making worry that match making limits social mobility because people of higher status will decline to meet and court people of a lower social status. But that concern while not entirely unfounded is probably exaggerated. The reason is that the initial pool of potential mates may be too small when the client has extremely demanding sets of criteria. (How many eligible billionaires who have Hollywood --or “Bollywood”-- good looks are there likely to be)? In India, according to Krishnan, “fastidious” clients may find their search for a mate fruitless.

But mistakes of the other sort—not being demanding enough—also are probably discouraged by the matchmaker. The client who is not demanding enough will find the initial pool perhaps too large and the client will probably not meet and court the best possible future mate. It is to the matchmaker’s benefit—and perhaps the client’s as well-- if the client pool is neither too large nor too small. Probably handing a list of four to six or seven potential matches to the client is ideal.

Returning to the analogy of the matchmaking as a form of diagnosis consider now a fictitious (but plausible) example.

Dr. Indira Singh is a professional woman of Indian ancestry who has worked for a number of years in Antarctica as a biological scientist studying survival of organisms in extreme environments. She is in her late twenties and decides she really would like to be a wife and mother. She also would prefer a husband of her ethnicity and social status etc. However, in her professional circle she has not met men that she feels are right for her. To help her, she contacts a relative in India who arranges to initiate a search for a suitable mate through a matrimonial service. She agrees to spend time in India meeting possible mates generated by the matchmaking service.

The matchmaking service provides her a list of half a dozen possible mates and she goes out on dates with them. She is mildly interested in one of them but he does not reciprocate her interest.

Discouraged she returns to the United States where she has a good job with a university. During her first weeks on campus she meets and becomes romantically involved with a Jewish physicist. He is fairly religious and while he is attracted to her he insists his wife must raise their children in his faith. After consulting friends and professional people she trusts she agrees that she will convert to Judaism and raise their children in her husband's faith. Once that issue is resolved he proposes and they marry.

The example should make clear that the process of arriving at the final outcome is one of "trial and error." Laor and Agassi (1990) in their work on medical diagnostics acknowledge a disjunction between the perception of diagnosis as an initial activity and the reality of the activity. They comment that while the diagnostic process is supposedly confined to initial clinical diagnostic encounters, in more elaborate cases the process includes laboratory examinations of samples and specimens, and that may further elaborate into patient hospitalization. At that point the process is usually no longer viewed as merely diagnostic by most people, because it has become more complicated as well as more costly. This has occurred in the given example with the evolution of Dr. Singh's findings about the relationships she encountered.

First, Dr. Singh had the feeling that "something is wrong" and she further concluded that her life was incomplete, and therefore she needed a mate. Second, she felt that available help would be effective for resolving the problem. In the Indian culture, Krishnan reports, the matchmaker role has been institutionalized for millennia and therefore for the specific situation in the example matchmakers are likewise granted permission to collect information and, based on the tentative diagnosis of the client's need, offer a "treatment plan" that consists of a list(s) of suitable mates for the client.

Third, the matchmaker help did not immediately prove effective. The lonely scientist found that the initial "treatment" --the first list of "potential

mates" from the professional matchmaker in India-- did not yield a suitable match. She needed to generate a new list of acceptable mates who were likely to find her attractive.

Finally, her search was successful, albeit without the matchmaker's assistance. To use the medical analogy a diagnosis—the correct one—far from being the first step in effective treatment was the *final step* after various failed treatments in response to earlier tentative diagnoses.

Can Matchmaking Be Improved?

Economists looking at computer assisted match-making have offered some ideas for improving it (Mims, 2016). One of their ideas, "create a market that is 'thick' {i.e.} a market that has a lot of people seeking to link up" is certainly worthwhile. The suggestion is especially germane to the American market, more so certainly than to the Indian market which has been the subject of most of the commentary in this paper because the use of a middleman is widely accepted in Indian culture but rather novel in the United States.

While a bigger data base clearly is advantageous, equally important in finding a suitable match is that the list of candidates is based on high quality data derived from a well designed questionnaire/interview protocol. The aim is a pool that is the result of a fine grained analysis.

Another interesting idea that may not prove practical has been suggested by "{Paul} Oyer, a Stanford University economist." He argues the key to avoiding 'romantic unemployment' is to use "cost-benefit analysis," a highly technical form of analysis that few people outside that discipline have the time and interest to master.

Krishnan (2016) told the author that in India, faced with the possibility of having to choose between someone who does not meet their standards and "romantic unemployment," dissatisfied clients will either ask the matrimonial service to generate a new list or go to a rival matrimonial service for help because "settling for a mate is now considered an unwanted outcome" among users

of Indian matrimonial services. She added that this extra expenditure of effort spent on searching for an ideal match may be a consequence of the matrimonial services' being too successful in selling the concept of the online match service "Online databases give the illusion of infinite mate choices" and therefore customers "believe that they do not have to settle." A cost-benefit analysis is a last resort among consumers "wanting only benefits" from their online search.

Mims clearly understands that American consumers of match services also are loath to doing a formal cost-benefit analysis in their search for a suitable mate. However, he thinks American consumers, will not be as willing as Indian matrimonial service clients to continue their search. In a recent study of 1000 couples who met on [an American dating site] Mims reports within two months of joining women in these couples were 20% less picky on average" while "men in these couples were on average 12% more picky."

Mims unfortunately oversimplifies Oyer's position. A more nuanced description of Oyer's views might be called "finding undiscovered value" in a potential mate" that allows one to achieve a match. And sometimes the added expense of this research is justified by the return on investment from a superior match.

Consider the following case of a young woman scientist who in using an American matchmaking service finds two possible mates. They have overall scores that make them a toss-up in regard to choosing between the candidate with undiscovered or hidden value and being flexible enough to settle.

One is a good looking fellow who works with his hands, makes a decent living, but likes to hang out with the boys on Friday nights. Invariably, this likable fellow gets drunk while out with the boys and then either sleeps most of the next day or is somewhat abusive to his mate after he returns home from a bout of drinking.

The other suitor is a well educated fellow, a bit dull but otherwise nice, who regrettably has a low paying job. However, he is ambitious to improve himself. Which of the two suitors would the woman pick? Neither of them is a star but this

chapter's argument would suggest that the ambitious if somewhat dull fellow is clearly preferable. He has unrecognized value (because he has ambitions to improve himself) whereas the good looking fellow who hangs out with the boys and drinks to excess is either overvalued or at least fully valued. His drinking threatens to undo a possible marriage because of his abusive behavior after he drinks and unwillingness to give up his Friday nights out with his friends. This fictitious case is loosely based on people known to the author either because they are in his circle or are related to people in his circle of acquaintances. However, the aim here is to have the reader not lose sight of the point that settling may not yield results equivalent to benefit-cost analysis. On the other hand, for young men/women to do even a crude benefit-cost analysis may require a level of judgment and maturity that they may lack simply because they have too little experience in living. To highlight the superior choice the matchmaking service may need to intervene tactfully to highlight the differences between two candidates that the client cannot decide between.

CONCLUSION

Computerization has been helpful in making arranged marriage in India more satisfactory for all stakeholders by expanding the range of mates from whom a marriageable individual can choose and by allowing more fine grained analysis of the assets and liabilities any potential mate brings to the union. Thus a more informed decision is possible for all stakeholders. The author does not see any merit in the argument that the innovative use of computers to help select potential mates is more inherently supportive of the caste system than the old way of selecting a suitable spouse. The wealth of information the computerized search can provide enhances the potential of identifying someone as a mate with offsetting desirable traits and "undiscovered" value compared to the traditional way of selecting a mate in an arranged marriage.

FUTURE RESEARCH DIRECTIONS

Social science has developed a number of middle range theories that have demonstrated applicability across the traditional specialties and disciplines within the field. Notable among these middle range theories are the “garbage can model of organizational behavior” and coalition theory as well as such social psychological theories as the theory of cognitive dissonance, the Hochbaum-Huntington model of political mobilization, reference group theory to name just a few. These theories are breaking down traditional disciplinary classifications and substituting new ones including chaos theory, organization studies and so on. The present paper is one of a planned series on how technological change is affecting traditional work such as medical diagnosis, decisions to marry, etc. and how new models of social behavior can improve both our understanding of these changes and their efficacy for enhancing the wellbeing of society and its members.

REFERENCES

- Blood, R. (1967). *Love Match and Arranged Marriage: A Tokyo-Detroit Comparison*. New York: The Free Press.
- Cohen, M., March, J., & Olsen, J. (1972). A Garbage Can Model of Organizational Choice. *Administrative Science Quarterly*, 17(1), 1–25. doi:10.2307/2392088
- Durkheim, E. (1964). *The Rules of Sociological Method* (8th ed.). New York: The Free Press.
- Fisher, R. (2005). *The Research Productivity of Scientists*. Lanham, MD: University Press of America.
- Fisher, R., & Fisher, J. (2014). Medical Diagnosis: A ‘Garbage Can’ Model Perspective. In J. Wang (Ed.), *Encyclopedia of Business Analytics and Optimization* (pp. 1162–1172). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-5202-6.ch107
- Goel, (2006, November 2). Online marriages are a runaway success. *Knight Rider Tribune Business News*, 1.
- Harris, B. J. (1989). Power, Profit, and Passion: Mary Tudor, Charles Brandon, and the Arranged Marriage in Early Tudor England. *Feminist Studies*, 15(1), 59–88. doi:10.2307/3177818
- Jana, R. (2000, August 17). Arranged marriages, minus the parents. *New York Times*.
- Khanna, T., & Yafeh, Y. (2015). Business groups in emerging markets: Paragons or parasites?. *Review of Economics & Institutions/Economia, Societa e Istituzioni*, 6(1).
- Kingdon, J. W. (1984). *Agendas, alternatives, and public policies*. Boston: Little Brown.
- Krishnan, A. (2012). e-Matchmaker, e-Matchmaker, Make Me a Match: Indian Matrimonial Services in the Internet Age. In *Computer-Mediated Communication across Cultures: International Interactions in Online Environments*. Hershey, PA.: Information Science Reference.
- Laor, N., & Agassi, J. (1990). *Diagnosis: Philosophical and Medical Perspectives*. Dordrecht, Netherlands: Kluwer Academic Publishers.
- Miller, S., & Wilson, D. (2006). Perspectives on Organizational Decision-Making. In *The Sage Handbook of Organization Studies*. London: Sage Publications.
- Mims, C. (2016, February 8). How Economists Would Fix Online Dating. *The Wall Street Journal*.
- Mohr, L. B. (1976). Organizations, Decisions, and Courts. *Law and Society Review*, 10, 621–642.
- Pepper, D. (2007, March). Matchmaking Indian-style. *Fortune*, 155(5), 14.
- Philip, A. (2005; October 17). Getting married the snappy, global way. *Indian Express*. Retrieved October 18, 2005, from <http://www.expressindia.com/fullstory.php?newsid=56735&pn=0>

Ramalingam, A., & Nair, S. (2005). Swayamvar in mouse mode: Matrimonials get an e-twist. *The Financial Express*. Retrieved September 1, 2005 from http://www.financialexpress.com/fe_full_story.php?content_id=82474

Saint Amant, K., & Kelsey, S. (Eds.). (2012). *Computer-Mediated Communications across Cultures: International Interactions in Online Environments*. Hershey, PA: Information Science Reference. doi:10.4018/978-1-60960-833-0

Sharma, A. (2006). *Girl seeks suitable boy: Indian marriage dot com* (Unpublished Ph.D. dissertation). University of Toronto, Canada. Retrieved January 22, 2010, from Dissertations & Theses: Full Text.

Weick, K. (1976). Education Organizations as Loosely Coupled Systems. *Administrative Science Quarterly*, 21(1), 1–19. doi:10.2307/2391875

Yee, A. (2008, May 17). Sari Night and henna parties. *The Financial Times*.

Zeldenrust, S. (1990). *Ambiguity, Choice and Control in Research* (Unpublished Ph. D. dissertation). Universiteit van Amsterdam, Amsterdam, The Netherlands.

ADDITIONAL READING

Amato, P. R. (2012). Institutional, Companionate, and Individualistic Marriages. In M. Garrison & E. S. Scott (Eds.), *Marriage at the Crossroads: Law* (pp. 107–124). Policy, and the Brave New World of Twenty-First-Century Families. doi:10.1017/CBO9781139087568.009

Islam, M. M. (2012). The practice of consanguineous marriage in Oman: Prevalence, trends, and determinants. *Journal of Biosocial Science*, 44(5), 571–594. doi:10.1017/S0021932012000016 PMID:22317781

Johnson, D. R., & Bachan, L. K. (2013). “What can we learn from studies based on small sample sizes? Comment on Regan, Lakhanpal, and Anguiano (2012)” “Relationship Outcomes in Indian-American Love-Based and Arranged Marriages. *Psychological Reports*, 113(1), 221–224. doi:10.2466/21.02.07.PR0.113x12z8 PMID:24340813

Regan, P. C., Lakhanpal, S., & Anguiano, C. (2012). Relationship Outcomes in Indian-American Love-Based and Arranged Marriages. *Psychological Reports*, 110(3), 915–924. doi:10.2466/21.02.07.PR0.110.3.915-924 PMID:22897093

KEY TERMS AND DEFINITIONS

Computer Assisted Diagnosis: A (tentative) conclusion or assumption which identifies the pathology represented by a patient’s (client’s) inquiry, particular set of symptoms, or complaints that has been reached with the assistance of a computer through the use of special (usually proprietary) software.

Confluence: As used in the “garbage can model” the intersection of various information streams at some critical point which enables one to proceed preferentially along a given informational evolutionary trajectory to the final diagnosis. Also referred to as a “linkage”.

Cultural Patterning: A behavior related to traditions, group social interactions and influences, and associated factors which do not necessarily have a scientific or nonrandom basis.

Diagnosis: A conclusion or assumption which identifies the pathology represented by a patient’s inquiry, particular set of symptoms, or complaints. As used in the Fisher=Zeldenrust version of the “Garbage Can Model” it refers to the final diagnosis of the problem as indicated by the fact the patient responded well to the treatment ordered based on the tentative diagnosis or the client was successfully matched to a potential marriage partner.

Evolutionary Trajectory: A vector which tracks the information inputs to the diagnostic process for a particular conjecture diagnosis until a final decision is made as to whether this particular conjecture diagnosis is accepted or rejected.

Garbage Can Model: A group of closely related theoretical models that analyze decisions under conditions of uncertainty. Originally developed by M. D. Cohen, J. G. March, and J. P. Olsen to analyze “organized anarchies”. These models make certain similar assumptions about the specific cases under consideration including that the problems and solutions are either “uncoupled” or “loosely coupled.”

Matrimonial Portals/Websites: Portals or websites designed for the sole purpose of finding a marriage partner. Similar to dating sites, these websites also feature users (clients) uploading their profiles onto a database and using various search criteria to find potential partners.

Social Networking Site: An online place where a user creates a profile and builds a personal network with other users. Although this can be used to find a marriage partner, it is more general and does not use a special proprietary software to match the user to other potential partners.

H

Creative Collaborative Virtual Environments

Luís Eustáquio

Universidade do Porto, Portugal

Catarina Carneiro de Sousa

Polytechnic Institute of Viseu, Portugal

INTRODUCTION

This article offers a definition of Creative Collaborative Virtual Environments (CCVEs), concerning models of collaborative and distributed creation in online virtual communities. Necessary affordances to enable a CCVE are described, and their importance is evidenced in the context of co-creation of content, using art practice as an example. This definition benefits a continued development and use of virtual worlds, as platforms for new collaborative models.

A CCVE is grounded on three key affordances: creation, collaboration, and distribution. These relate not only to the technical, but also to the social layers of virtual online communities.

Shared creativity and distributed authorship are approached as examples of specific dynamics rooted upon those three elements. Because the communities emerging from this type of creative flux provide fertile ground for the advancement of critical research on collaboration and creation in cyberspace, examples of multiple configurations are discussed, regarding the use of networks, technologies and participation frameworks.

Trough this definition, the authors propose to define models of collaborative and distributed creation in virtual online communities. Networked collectives of different practices and practitioners are discussed, across various virtual spaces, as examples of such models. Second Life (SL) is analyzed as a typical instance of a CCVE, since it

currently presents the most accessible and integral approximation to this concept.

The discussion supported by these observations ultimately demonstrates how the co-creation of new content and meaning takes place through collaborative practices in virtual worlds, and how Creative Collaborative Virtual Environments widen the gamut of communicative and creative agency in digital communities.

BACKGROUND

Virtual worlds are commonly referred to in literature as the Metaverse, a term coined by novelist Neal Stephenson in his seminal fiction *Snow Crash*. There, the Metaverse is an immersive virtual 3D world, where people interact through their digital manifestations, avatars. While the term has been broadly applied to the entire collective online space, it is specifically connected with simulated worlds in virtual 3D space. Spatiality is the most distinguishing feature of virtual worlds, as they provide an immersive experience where one moves across a (virtually) infinite, simulated world, rather than a two-dimensional metaphor of a desk with folders and a trash bin. Tom Boellstorff (2008) advances three fundamental properties of virtual worlds: they are places, inhabited by people, and enabled by networked technologies.

Virtual worlds are also often called Collaborative Virtual Environments (CVEs). Churchill,

Snowdon & Munro describe them as locations for action and interaction (Churchill, Snowdon, & Munro, 2001), virtual spaces where people can meet and interact with other people, agents and virtual objects. CVEs promote users from spectators to active participants in the Metaverse, able to engage each other and the virtual environment.

THE CREATIVE APPROACH TO CVE

A collaborative space enables dialogue and exchanges between users, but is not required to enable content creation at its core. On the other hand, a creative environment does not strictly require online collaborative features to afford creativity. As noted by Lévy (2001), the distinction between read-only and read/write virtual worlds is not an opposition. Many virtual environments are able to digest “offline” processes to some extent, importing or exporting content. Others may allow some degree of self-expression, through limited customization options. However, limited presets do not empower users to create or reinvent their own virtual world. To achieve this potential, users must be able to create, modify, transform and redistribute media assets that constitute the very fabric of the virtual world: notably, audiovisual components (including 3D data, if applicable) and program code.

Users must also be able to employ such tools according to their own policies and methodologies, regarding aspects such as creative process and media rights management. Virtual worlds can be complemented by other platforms, such as forums and social networks (Al-Jarrah & Pontelli, 2014; Ferguson, 2011; Kohler, Fueller, Matzler, & Stieger, 2011), but in-world resources for creation, communication and asset distribution, enable a CVE to become a platform for works of flow and process. CVEs also remediate specific features from other platforms. This includes live and relayed text messaging, video streaming and file sharing. They also introduce features unique to virtual worlds, such as live visual expressive-

ness through avatars, and shared 3D spaces. These features promote users from spectators to active participants, but they still don’t afford users creative control over the virtual world they inhabit.

CCVE AFFORDANCES

For a CVE to be considered a CCVE, it must afford creative input and action. For creativity to be considered collaborative, users must be able to distribute and modify in-world content. This requires built-in platform features for collaborative creation, modification and distribution of content, to a degree that empowers users to collectively shape the virtual world itself.

Necessary features to enable such affordances are described in the following sections.

Creation

The first key affordance is creation. For a virtual environment to be considered creative, users must be able to partake in the making of the world itself. The range of creative input may vary according to a platform’s specific features: a wider range of creative options increases creative possibilities. However, platforms easily become overly complex, making it difficult for average users to master these creative possibilities, effectively professionalizing the creative activity. A sound balance between creative possibility and tool complexity is hard to achieve, and this is often remedied with content upload. Complex content can be built externally, allowing the platform itself to remain more accessible. The ability to create content can thus be split between building within the CCVE platform, and/or uploading content built using external resources. Table 1 offers examples of features a CCVE might offer within this framework.

Collaboration

Another core affordance of CCVEs is the possibility of creative collaboration. Collaborative

Table 1. Affordances and related features for creation in CCVEs

Creation		
	Affordances	Features
Within the virtual world	Content creation	Create new 3D objects; map textures; edit landscape and environment; trigger sound reproduction; write new code to change behaviour of world, objects and avatars.
	Content modification	Tools to modify settings and properties of existing avatars, terrain, objects, audio and visual media, environment and dynamic behaviours.
Outside the virtual world	Import content	Allow user to upload data from local storage to virtual world server; accept input of standard file formats for text (including code), images, video, 3D objects, and sound.
	Export content	Allow user to transfer data from server inventory to local storage; produce output in standard formats for text, images, video, 3D objects and sound.

activities require actions of a user to be perceived by other participants, for there to be interaction between users. This kind of interaction implies negotiation. Churchill and Snowdon's (1998) research on negotiation processes advances important aspects required for successful collaborative actions in mediated environments. First, one needs to understand the transition between shared and individual actions. This requires explicit and tacit communication, as well as the ability to perceive what is being done and what has been done. Flexible and multiple viewpoints are also important in shared activities. A second aspect is multilingual text display. A shared context is probably the most important, requiring shared environment, shared artefacts, but also shared knowledge or shared understandings. Equally important is the awareness of others, the sense of shared activities and the ability to communicate about them, verbally and nonverbally. Churchill and Snowdon refer

studies which suggest facial expression, body posture and gesture carry 60% to 90% of information communicated. This is a known challenge faced by current existing virtual worlds: although there are built-in gestures, and it is possible to upload animations to move an avatar in intentional ways, both lack the spontaneity required for casual communication¹.

Table 2 articulates aspects highlighted by Churchill and Snowdon (1998), as co-creative affordances, associated with specific features required for a platform to enable them.

Distribution

Distribution is the third key affordance of CCVEs, extruding the circle of creation and collaboration to a sphere of action and meaning, defined by a multitude of points on stems connecting it to as many other spheres. This image represents the inbound and outbound nature of such connections.

Table 2. Affordances and related features for collaboration in CCVEs

Collaboration	
Affordances	Features
Communication	Channels for verbal communication, written or spoken, relayed or in real-time. Forms of nonverbal communication, such as facial and body expression.
Awareness of others	Perceive and observe relative position and actions of others in virtual space; feel the impact of other's actions.
Shared context	Ability to share real-time experience of the same location and environment.
Flexible viewpoints	Viewpoint control: travel, pan, tilt, rotate, zoom.

Table 3. Affordances and related features for collaboration in CCVEs

Distribution	
Affordances	Features
Exchange content with other users	Transfer and save files: tools to give/sell/lend, and store content.
Present content in shared space	Render objects; produce sound; performative action.
Peer property mode	Copy and save content from shared spaces; shared file storage.

Distribution applies to all kinds of artefacts and experiences in virtual worlds, multiplied by the number of users who experience them. This entails enabling users to: store and exchange files; render media and objects in common space, including performative action; save assets from the environment; share ownership of and access to media assets.

CCVEs thus extend participation beyond observation, allowing peer to peer collaboration. This unlocks a transformative potential not limited to technical features. Rather, it configures new programs of governance, modes of production and ownership regimes. Users are able to replace a “market exchange value” with a “use-value for a community of users”, creating a “peer property mode” fundamentally different from private or public property (Bauwens, 2006, p. 33) as it enables direct and collective governance.

Distribution also concerns authorship, as manifestations of creation in virtual worlds inevitably draw from labor in previous connections. Aggregations of multiple connections constitute scenarios of distributed creativity and authorship, converging towards Axel Bruns’ (2007) concept of produsage, which describes “projects which harness the creativity of a large range of participants to build on and extend upon an existing pool of artistic material” (Bruns, 2010, p. 1). Produsage is prevalent across online creative communities², where massively distributed bodies of work emerge through processes in which practitioners shift back and forth from producers to users, originating hybrid roles (Bruns & Schmidt, 2010, p. 3) within a creative process where artefacts and their

production cycles are open-ended, repositioning users as co-creators and active participants (Bishop, 2012). Table 3 presents a summary of affordances for creative distribution, and corresponding necessary features.

SHARED CREATIVITY

CCVE affordances lead to new creation/collaboration models, characterized by Catarina Carneiro de Sousa (2013, 2015) as shared creativity, a creative input distributed by several creators, processed in different timeframes, across different places and through different approaches, enabling a fluid creative stream. These variables work together in three distinct modes: collective creativity, collaborative creativity, and distributed creativity.

Collective Creativity

This mode draws from collective effort, asynchronously or through simultaneous participation. Each participant’s contribution is mostly indistinguishable, making all participants equally responsible and creditable. This modality does not exclude other significant aspects of collectivity, as it simply designates a workflow where all participants form a singular creative entity. In a cellular structure, an egalitarian basis provides greater chance of success when individuals waive their personal mark to benefit collective authorship (CAE, 2002). Collective creativity is then a process undertaken by a group of creative partners, in a profound level of intimacy.

Collaborative Creativity

As an “open-ended concept”, collaboration “becomes an umbrella term for the diverse working methods that require more than one participant” (Lind, 2007, p. 17). Collaborative creation does not strictly deal with the collaborative aspect, for it provides a rather wide frame of reference, regarding organization methods and creative process. Instead, collaborative creation designates a process in which participants create together, but each individual authorial mark can be variably discernible, as the borders of each contribution blend into a collectively authored outcome. This process often takes place as a conversation, where each participation is a creative response to earlier contributions.

Distributed Creativity

Distributed authorship was initially proposed in 1986, by digital art pioneer Roy Ascott. This concept emerged from *La Plissure du Texte: A Planetary Fairytale (LPDT)*, a collaborative, remote authoring project, launched by Ascott in 1983. Axel Bruns (2010) employed this concept to describe internet-augmented creative processes. Bruns specifically addresses projects comprising a considerable number of participants, contributing to a common pool of creative resources and artistic production. Here, however, participants act individually, taking their own steps along the creative process, configuring a modality of distributed creation (Bruns 2010, p. 1).

INSTANCES OF CCVE

The definition of CCVE refers exclusively to online virtual worlds, therefore excluding collaborative platforms such as text-based forums, and virtual environments such as single-player games. Not all online virtual worlds are implicitly creative and collaborative, and not all are networked in freeform configurations. Even when a virtual

environment offers creative and collaborative features, their design and operation may hinder in-world co-creation, and freeform collaboration. A CCVE thus requires more than the sum of these parts, to afford users actual control over modes of collaborative creation.

Virtual worlds like SL or OS grids offer the closest available approximation to the CCVE definition³, as they meet most requirements previously described. Both offer its users real-time in-world interaction, through avatars, in a spatial environment, inhabited by users, connected by a network. Together with tools to reshape every aspect of the virtual world from the ground up, including the ground itself, this allows synchronous co-creation. These features thus enable in-world creation, collaboration and distribution, making this collaborative virtual world also a creative one. These platforms embrace internal forms of collaboration, by providing tools for communication and sharing. Public and private chatrooms enable both real-time (conversational) and asynchronous (message log) verbal communication between users. Embedded automated translators, albeit somewhat limited, facilitate conversations between different languages. Every user can store various media types: images, texts, sounds, 3D objects and scripting code. These items can be copied and sent to other users, or placed on virtual space to be picked up by others. Collaboration using external resources is equally possible, by importing rich media into the virtual world. Through built-in features or alternative techniques, users can extract and store audiovisual results from the virtual world, for documentation or further work. Thus, tangible and virtual world are not separate networks but closely connected, heterogeneous nodes of a rhizome, enabling workflows from global community to virtual environment to physical museum, for example.

Actions in SL and OS grids are not limited to preset contents or programs of action. Beyond “inhabiting” and perusing virtual spaces and goods, residents can create, modify and distribute new artefacts, texts, topologies, and narratives. Not be-

ing limited to preset inventories and geographies, these platforms put the creation of impossible objects and dream environments within reach of any user. This combination of creative possibility and sharing ability configures a wider range of participatory aesthetical experiences than is possible in games such as *World of Warcraft* or *League of Legends*, as the ability to create all aspects of a virtual world from the ground up is exponentially more empowering than incremental customization or orchestrated theatricality. Jennifer Novak-Leonard and Alan Brown (2011) propose a framework for modalities of artistic participation which comprises five levels: Ambient Participation describes chance encounters with artworks; Observational Participation describes passive attendance of events; Curatorial Participation includes selecting and relating artworks; Interpretive Participation describes performative self-expression; and finally, Inventive Participation “engages the mind, body and spirit in an act of artistic creation that is unique and idiosyncratic” (Novak-Leonard & Brown, 2011). In publicly accessible CVEs, one can find users whose activity can easily fit most of these categories. However, SL (and OpenSim based equivalents) enables the full spectrum of modality in artistic participation, through its specific combination of features, which cannot be said of most virtual platforms it is usually compared to, most notably mainstream multiplayer games. Creation and participation in SL is enabled by the technical realm, but fulfilled in the social realm. While most games implement interconnectivity for the purpose of joint action towards preset goals, SL configures a program without totality, driven by autonomy and an openness toward alterity (Lévy, 2001).

Meta_Body

Currently virtual worlds such as SL and OS grids are home to a prolific community of artists and art-driven spaces (Sousa & Eustáquio, 2015). The authors of this article are active members of this community, developing their projects within the

virtual artworld. This discussion will now turn to one such project, as an example of how virtual world features drive shared creative processes.

Meta_Body is an ongoing collaborative project, set in motion in SL by Meilo Minotaur (Sameiro Oliveira Martins’ avatar) and Author 2⁴, spanning several virtual and tangible environments since 2011.

As a first step, eighteen avatars were created and distributed with full permissions, to be freely duplicated, modified and redistributed by any willing participant. Derived creations could also be freely distributed. Each avatar included a text note, inviting residents to participate with derivative artworks in a tangible art exhibition. 120 derived works were initially selected for an exhibition in Austria⁵, including virtual photographs⁶ and machinima⁷.

A second edition started in 2012, titled Meta_Body II. Residents were invited to produce new avatars, derived from the Meta_Body set, which were then redistributed, again with full permissions. They were now collectively authored, built from parts of the original avatars from the first edition. The project achieved increased exposure and embraced an even wider range of participations⁸, of which Figure 1 provides an example.

In both editions, a unique virtual installation was built from the ground up, where one could enjoy a shared aesthetical experience. Meta_Body II encompassed the creation of four expansive landscapes, each with its own aesthetic character and nature, related to the avatars distributed there. Author 1⁵ developed soundscapes that changed dynamically as one moved across virtual space, adding immersiveness to the experience.

Meta_Body offers an opportunity to observe various phenomena within CCVEs. Retrieving concepts of produsage and peer to peer collaboration, one such phenomena is artistic creation as a collaborative process, set in motion by a community, within the context of decentralized co-creation of content. While popular social networks are valuable to spread awareness of projects and call for participations, CCVEs add

Figure 1. To the right, an avatar from Meta_Body phase I: Godiva, by CapCat Ragu and Meilo Minotaur. To the left, an avatar from phase II: Appointment in the garden, by Simotron Aquila.



new variables and mechanisms to this participatory context. These stem from a combination of three key factors mentioned earlier: creation, collaboration and distribution.

Meta_Body translates these practice values into a shared creative process, which comprises three approaches previously described: collective creation, distributed creation and collaborative creation. Collective creation is the process used in the construction of original avatars by Meilo Minotaur and Author 2⁵. Distributed creation describes how most participants entered the creation flow, building upon previous work and sharing derived output for further reinterpretation. Finally, collaborative creation is the process used with Author 1⁵ in creating soundscapes for the Meta_Body II virtual installation: while developed independently by a single author, the sounds inhabit virtual spaces without physically altering them, but strongly affecting their perception.

These modalities of shared creativity represent the variable “open-endedness” of cultural and creative expression in online participatory environments, in collaboration as a collection

of methods for working together (Lind, 2007). Through these processes, “communicative practices offered online actualize, enact, and thereby reshape specific cultural ideals”(Langlois, 2013, p. 92). By selecting previous work to deconstruct, reinterpret, act and build upon, participants employ the platform’s communicative agency to strengthen or weaken connections in a semantic network, changing its shape and dynamics from within collective practice. Figure 2 presents a practical example.

FUTURE RESEARCH DIRECTIONS

New platforms and communities continue to emerge, industrial and academic research in the field is blooming, and multinationals compete to set the standard platform for virtual reality. As they mature and mutate, virtual worlds operate on rapidly changing software and hardware. Greater computational power enables finer detail and fluid motion. Increased bandwidth enables complex environments to be shared online.

Figure 2. To the left: You see my inside, avatar from Meta_Body phase I. To the right, the alpha_tribe reinterpretation.



Experimental hardware research pursues more intuitive and expressive interfaces for humans. These factors amount to a considerable potential for the future of virtual worlds, where accessibility, immersion and ubiquity become possible goals. Numerous avenues for important research lie in this future, as new interaction paradigms entail new aesthetic experience modalities, and new organizational structures in virtual communities. These are relevant to artistic practice, collective and individual, but also to research in areas such as games, psychology, sociology, pedagogy, health and economy. For virtual worlds to remain open, features to enable creative and collaborative affordances must be kept and expanded. The authors advance the concept of CCVE as a contribution toward such a future.

CONCLUSION

A Creative Collaborative Virtual Environment is a digital virtual environment, where a group of people can remotely interact and collaborate, asynchronously or in real time, in the creation, modification and distribution of media and content that shapes the virtual world itself. Expanded from Collaborative Virtual Environments (CVEs),

CCVEs exist in networked, computer-generated 3D spaces, simultaneously inhabited by various visitors (Boellstorf, 2008). Accessible as software platforms, they enable different levels of engagement and modalities of collaboration (Lind, 2007). CCVEs feature an enhanced creative dimension, through specific software features and affordances of creation, collaboration and distribution. These affordances enable rhizomatic formations of heterogeneous, non-hierarchical connections, accommodating variable origins and flows for creative practice (Deleuze & Guattari, 1987). This increased spectrum of communicative agency contributes to heightened senses of presence and copresence in virtual worlds, shaping new conditions for negotiation of action and meaning (Ijsselsteijn & Riva, 2003). Produsage (Bruns, 2007) enables various modes of collaboration as described by Sousa (2015), which demonstrate open-endedness in cultural and creative expression, through networks of decentralized communication and decision-making. Ultimately, CCVEs demonstrate how “networks of technologies, users, and social processes define and delineate specific modes of experiencing meaning” (Langlois, 2013, p. 103), by enabling co-creation of new content and meaning through collaborative practices in virtual worlds.

REFERENCES

- Al-Jarrah, A., & Pontelli, E. (2014). AliCe-ViL-lagE Alice as a Collaborative Virtual Learning Environment. *Frontiers in Education Conference* (pp. 1-9). Madrid: IEEE.
- Ascott, R. (2005). Distance Makes the Art Grow Further: Distributed Authorship and Telematic Textuality in La Plissure du Texte. In A. Chandler & N. Neumark (Eds.), *At a distance: precursors to art and activism on the Internet* (pp. 282–296). Cambridge, MA: The MIT Press.
- Bauwens, M. (2006). The Political Economy of Peer Production. *Post-Autistic Economics Review*, (37), 33-44.
- Bishop, C. (2012). *Artificial hells: participatory art and the politics of spectatorship*. London: Verso.
- Boellstorff, T. (2010). *Coming of Age in Second Life, An Anthropologist Explores the Virtually Human*. Princeton University Press.
- Bruns, A. (2007). *Produsage, Generation C, and Their Effects on the Democratic Process*. Boston: MIT.
- Bruns, A. (2010). Distributed Creativity: File-sharing and Produsage. In *Mashup Cultures* (pp. 24-37). Vienna: Springer.
- Bruns, A., & Schmidt, J.-H. (2010). Produsage: A Closer Look at Continuing Developments. *New Review of Hypermedia and Multimedia*, 17(1), 3–7. doi:10.1080/13614568.2011.563626
- CAE. (2002). Collective Cultural Action The Critical Art Ensemble. *Variant*, 2(15), 24–25.
- Churchill, E. F., & Snowdon, D. (1998). Collaborative virtual environments: An introductory review of issues and systems. *Virtual Reality (Waltham Cross)*, 3(3), 3–15. doi:10.1007/BF01409793
- Churchill, E. F., Snowdon, D. N., & Munro, A. J. (2001). Collaborative Virtual Environments: Digital Places and Spaces for Interaction for CSCW: An Introduction. In E. F. Churchill, D. N. Snowdon, & A. J. Munro (Eds.), *Collaborative Virtual Environments: Digital Places and Spaces for Interaction* (pp. 3–20). London: Springer. doi:10.1007/978-1-4471-0685-2
- Deleuze, G., & Guattari, F. (1987). *A thousand plateaus: capitalism and schizophrenia*. Minneapolis, MN: University of Minnesota Press.
- Ferguson, R. (2011). Meaningful learning and creativity in virtual worlds. *Thinking Skills and Creativity*, 6(3), 169–178. doi:10.1016/j.tsc.2011.07.001
- Gibson, J. J. (1986). *The ecological approach to visual perception*. Hillsdale, NJ: Erlbaum.
- Ijsselstein, W., & Riva, G. (2003). Being There: The experience of presence in mediated environments. In *Being There: Concepts, effects and measurement of user presence in synthetic environments* (pp. 3-16). Amsterdam: IOS Press.
- Langlois, G. (2013). Participatory Culture and the New Governance of Communication The Paradox of Participatory Media. *Television & New Media*, 14(2), 91–105. doi:10.1177/1527476411433519
- Lévy, P. (2001). *Cyberculture*. Minneapolis, MN: University of Minnesota Press.
- Lind, M. (2007). The Collaborative Turn. In *Taking The Matter Into Common Hands: On Contemporary Art and Collaborative Practices* (pp. 15-31). London: Black Dog Publishing.
- Norman, D. (1999). Affordance, Conventions, and Design. *Interaction*, 6(3), 38–43. doi:10.1145/301153.301168
- Novak-Leonard, J. L., & Brown, A. S. (2011). *Beyond attendance: A multi-modal understanding of arts participation*. Washington, DC: National Endowment for the Arts.

OpenSimulator. (2014, November 10). *OpenSimulator Main Page*. Retrieved Outubro 14, 2015, from OpenSimulator: http://opensimulator.org/wiki/Main_Page

Simonite, T. (2014, October 22). *The Quest to Put More Reality in Virtual Reality*. Retrieved October 14, 2015, from MIT Technology Review: <http://www.technologyreview.com/featured-story/531751/the-quest-to-put-more-reality-in-virtual-reality/>

Sousa, C. C. (2013). Meta_Body — A Project on Shared Avatar Creation. *International Conference in Illustration & Animation — CONFIA* (pp. 147-163). Porto: IPCA.

Sousa, C. C. (2015). Meta_Body — virtual corporality as a shared creative process. In D. Doyle (Ed.), *New Opportunities for Artistic Practice in Virtual Worlds*. Hershey, PA: IGI Global. doi:10.4018/978-1-4666-8384-6.ch009

Sousa, C. C., & Eustáquio, L. (2015). Art Practice in Collaborative Virtual Environments. In H. Baranhan, & S. S. Martins (Ed.), *Uncertain Spaces: Virtual Configurations in Contemporary Art and Museums* (pp. 211-240). Lisbon: Instituto de História da Arte, Faculdade de Ciências Sociais e Humanas – Universidade Nova de Lisboa; Associação do Instituto Superior Técnico para a Investigação e Desenvolvimento – Universidade de Lisboa; Gulbenkian Next Future Programme – Calouste Gulbenkian Foundation.

Xenakis, I., & Arnellos, A. (2013). The relation between interaction aesthetics and affordances. *Design Studies*, 34(1), 57–73. doi:10.1016/j.destud.2012.05.004

ADDITIONAL READING

Ayiter, E., & Dahlsveen, M. H. (2016). The Story of Volund: A translation from the oral to the visual. *Journal of Gaming and Virtual Worlds*, 8(8), 43–62. doi:10.1386/jgvw.8.1.43_1

Boellstorff, T. (2010). *Coming of Age in Second Life, An Anthropologist Explores the Virtually Human*. Nova Jersey. Princeton University Press.

Coleman, B. (2011). *Hello Avatar: Rise of the Networked Generation*. Cambridge, MA: MIT Press.

Doyle, D. (2015). *New Opportunities for Artistic Practice in Virtual Worlds*. Hershey: IGI Global. doi:10.4018/978-1-4666-8384-6

Sousa, C. C., & Eustáquio, L. (2015). Art Practice in Collaborative Virtual Environments. In H. Baranhan, & S. S. Martins (Ed.), *Uncertain Spaces: Virtual Configurations in Contemporary Art and Museums* (pp. 211-240). Lisbon: Universidade Nova de Lisboa; Calouste Gulbenkian Foundation.

Zabel, G. (2014). Through the Looking Glass: Philosophical Reflections on Art of Virtual Worlds. In M. Grimshaw (Ed.), *The Oxford Handbook of Virtuality* (pp. 407–419). New York: Oxford University Press. doi:10.1093/oxfordhb/9780199826162.013.018

KEY TERMS AND DEFINITIONS

Affordance: The concept of affordance is somewhat variable across scientific areas. Its mutation from verb to noun occurs in Gibson's Theory of Affordances (Gibson, 1986), referring to environmental actionable properties, but ultimately designating relationships between environments and actors therein. Affordances have been ascribed a functional role in active cognition, in an attempt to deconstruct our perception of such relationships, as when Donald Norman (1999) discusses real versus perceived affordances in interface design. Currently, affordances are strongly related with interaction processes, designating a range of action possibilities (Xenakis & Arnellos, 2013) emerging from engagement between actor and environment or artifact.

Distributed Authorship: Term coined by Roy Ascott (2005) to describe interactive and remote

authoring. The same term was later used by Axel Bruns (2010) to refer to projects in which a large number of participants contribute to a common pool of artistic material.

Feature: Designates specific characteristics and programs of action, found in software, artefacts and processes. In this sense, a feature is understood to exist as part of a virtual object, a software application, or a workflow, irrespective of its employment by an actor in a given environment.

OpenSim: Open-source multi-user 3D application server, used to create digital environments enabled by network technologies, experienced through the use of avatars that represent the user in the environment, on multiple protocols. This software enables developers to create and customize virtual worlds, using their preference of technologies. These worlds can be hyperlinked with each other, using the Hypergrid as an additional resource (OpenSimulator, 2014).

Participation: Describes a user's level of engagement in a specific platform. This notion is borrowed from contemporary art theory, analyzing recent transformations in the role of the audience, from simple viewer to co-producer or participant (Bishop, 2012).

Produsage: Term coined by Axel Bruns (2007) to describe a work process where participants shift from users to producers and vice versa, originating a hybrid role in between, the produser.

Second Life (SL): Digital, multi-user, 3D world, enabled by network technologies, experienced through the use of avatars that represent the user in the environment. This world is entirely created by its residents.

ENDNOTES

- ¹ Emerging virtual worlds and associated hardware (some offering public Alpha versions, e.g. High Fidelity) may mitigate this problem in the future, through the use of motion capture devices and facial tracking software (Simonite, 2014).
- ² These communities span various platforms outside of virtual worlds, such as flickr (<https://www.flickr.com>) and DeviantART (<http://www.deviantart.com>).
- ³ Other virtual worlds and new hardware are currently emerging (e.g. High Fidelity, as mentioned before) with the potential to meet the requirements described.
- ⁴ Avatar and real name will be provided after blind peer review.
- ⁵ The 6th edition of All My Independent Women, curated by Carla Cruz. Since then, Meta_Body has integrated several other "real-world" exhibitions and festivals.
- ⁶ Still images captured from virtual worlds.
- ⁷ Videos captured from real-time rendering of virtual worlds.
- ⁸ Twenty-two creators built twenty-six new avatars, with participations from renowned metaverse artists, designers, and new residents, in their first exploration of Second Life and avatar building. Furthermore, residents continue to produce derivative work in multiple platforms,, beyond the initiators' tracking ability. Some of this work can be found at the Meta_Body Flickr Group, now encompassing more than a thousand works and 376 members (https://www.flickr.com/groups/meta_body/pool/).

Cyberbullying



Gilberto Marzano

Rezekne Academy of Technologies, Latvia

INTRODUCTION

Information and communication technology (ICT) plays a central role in contemporary society. Computers are affecting human life in a great deal in that they are essential in all life dimensions. An increasing number of things run by computers, while massive amount of multimedia materials circulate through the Internet, especially by smartphones. New technologies have thoroughly penetrated into the everyday life since not only affect science and business, but also leisure and interpersonal relationships.

The difference between contemporary and past epochs lies in that, nowadays, the reality includes both physical and virtual things as well as living and artificial agents.

One of the chief impacts of computer is on communication which has made a great leap forward thanks to the advances in networking and internet technologies. Today, one can connect with friends and family around the world as well as can deal with business anywhere and anytime.

Despite these changes, human nature remains the same and the people's evil behavior didn't change. Paradoxically, thanks to the Internet, cheaters may have at hand more ways to cheat and perpetrators dispose of new powerful means for harming and harassing their victims.

Cyberbullying is a new, alarming, and evil phenomenon closely connected with relational changes that new technologies are causing in our society.

Essentially, it consists in using the Internet to harass, threaten, and harm individuals who are the weakest and most vulnerable. Victims of cyberbullying are mightily children and adolescents. In fact, young people are immersed in new digital

technologies and use them without knowing their implications.

In the Internet world, it is meaningless to say that "children should be treated like children": information technology put absolutely on the same level adults and children. The difference between children toys that imitate adult objects, e.g. a tricycle and a bicycle or a pedal car and a motor car, doesn't work on the Web. There isn't the *Internet for children* and the *Internet for adults*. Both adults and children use the same devices, tools, and ways of communicating and interacting.

It is universally agreed that no responsible adult would give the car keys to a child or an armed gun. Even though, this is what an adult does when puts in the hands of children a smartphone with an active Internet connection without give them any advice about the cyber threats.

This chapter aims at presenting cyberbullying as a direct consequence of the social metamorphosis that, induced by the diffusion of the new technologies, endangers our security and exposes us to new actual threats. The following paragraphs illustrate the background of this evil phenomenon, then, specific aspects, such as cyberbullying diffusion, the difference between cyberbullying and traditional bullying, and the characteristic of cyberbullying, are discussed. Finally, some new emerging phenomena related to the sexual sphere, the so-called revenge pornography and the homophobic cyberbullying, are highlighted.

Background: Cyberspace Threats and Cyber Violence

As the real world, cyberspace is full of real dangers and pitfalls: cyber violence can produce similar effects as the real-world violence (Finn, 2004;

DOI: 10.4018/978-1-5225-2255-3.ch360

Figure 1. Lifetime Cyberbullying Victimization Rates (nine different studies 2007-2015)

Source: Cyberbullying Research Center¹



Temple et al., 2016). Furthermore, what in real life is blamed, such as gambling and pornography, in cyberspace is easily affordable. Cyberspace relative anonymity gives people the opportunity of accessing places where, in real life, decency and shame usually prevent them from frequenting. Anonymity is one of the crucial aspects of cyberspace which offers opportunities and, at the same time, exposes users to risks and threats.

At the moment, cyber violence is difficult to define and, as a results, its prevalence rates are largely unknown (Peterson & Densley, 2016).

Many forms of cyber violence correspond to traditional forms of aggressive behavior. Perpetrators of cyber violence are not a uniform group, but they share the characteristic of exploiting digital media as a powerful weapon to inflict real damages to their victims. This is the case of *cyberbullying*. This term refers to bullying through new technologies, especially the Internet and smartphones.

Cyberbullying is the most controversial form of cyber violence (Corcoran, Guckin & Prentice, 2015) and there is a vast literature on this issue that addresses the phenomenon from different points of view: that of victims, perpetrators, and

bystanders, that of educational institutions, that of legal measures, that of prevention, and so on. Many sites have been created by parents of children who, after being victims of cyberbullying, committed suicide or by associations that to combat the cyber violence.

Figure 1 shows data from a survey conducted by the *Cyberbullying Research Center* directed by two influential researchers on cyberbullying, Sameer Hinduja (Florida Atlantic University) and Justin Patchin (University of Wisconsin-Eau Claire).

In 2013, the results of a three year multi-disciplinary project which involved researchers from 28 countries of the European Union have been published (Smith & Steffgen, 2013). This project traced a demarcation line from the previous research, often too oriented to psychological interpretations of cyberbullying.

Many of the first cyberbullying interpretations were based on media reports. For obvious reasons, media emphasized cyberbullying extreme cases, this is, which culminated in the suicide, or the attempted suicide, of victims. Media continue to underline the toxic and suicidal consequences of harassment resulting from new technologies. But, this is a wrong assumption. A few years ago, Marr

and Field (2001) documented and analyzed some cases of suicide due bullying, and coined the term *bullycide*. Authors recognized in the UK at least 16 cases per year of evident and diagnosed bullycide, claiming that bullycides were much more since many of them were not considered as such by investigative authorities.

In the next paragraphs we focus on some essential aspects of cyberbullying: diffusion, definition, and peculiar characteristics.

Cyberbullying Diffusion

Cyberbullying is wide spreading globally: affects over 50% of young people of school age.

However, a few years ago, Olweus, the pioneer in the field of bullying prevention and creator of the Olweus Bullying Prevention Program ² (Olweus & Limber, 2010), claimed that cyber-bullying was an overestimated phenomenon. He argued that media often exaggerated about the cyberbullying diffusion and the first investigations in this field were scientifically poor (Olweus, 2012).

Under certain aspects, the Olweus's criticism was correct. Since the 2010s, the most of literature flourished around cyberbullying was not always based on large samples or on reliable statistical data. Indeed, due to the novelty of the phenomenon, many studies were not based on a common definition of cyberbullying and many researchers were brought to extend the sphere cyberbullying including all sort of malicious behavior. Olweus well underlined the defects of the first investigations on cyberbullying, which studied this phenomenon out of the general context of the traditional bullying and created the image of cyberbullies as young introverted adolescents, expert in computing, who, taking advantage of the Internet anonymity, discharge their frustrations harassing classmates. Early research on cyberbullying (*online bullying*) was often based on media reports of cyberbullying cases and some studies were not theoretically and methodologically rigorous (Hinduja & Patchin, 2012).

However, the diffusion of cyberbullying cannot be underestimated and a reductionist attitude is dangerous. Cyberbullying is really an alarming phenomenon considering two distinctive aspects. The first concerns the fact that the Internet is the primary means of socialization for young people. Adults usually use the Internet for working, buying items, paying bills, and similar things, whilst young people use it for staying in touch and communicating among them. The second aspect concerns the nature of the Internet, namely that its contents are accessible to all, or at least accessible to a large number of users, are not controllable, and not easy to delete. The "right to be forgotten" is a concept discussed and put into practice in the European Union since 2006; it concerns also the Internet contents and is matter of a legal debate around the world (Bartolini, & Siry, 2016). A recent survey indicated that, in the USA, 9 in 10 Americans want some form of the right to be forgotten.³

Despite the people interest, it is not easy to apply the right to be forgotten on the Web due to many technical problems. As a consequence of this, Internet contents remain, at moment, globally accessible, uncontrollable, and permanent.

Table 1 shows the common types of harassment practiced by cyberbullies.

Cyberbullying Definitions

A debate is ongoing about the definition of cyberbullying and the differences between cyberbullying and traditional forms of bullying (Thomas, Connor & Scott, 2015; Menesini et al 2012).

Some of the most cited definitions of cyberbullying are: ⁴

An aggressive, intentional act carried out by a group or individual, using electronic forms of contact, repeatedly and over time against a victim who cannot easily defend him or herself. (Smith, Mahdavi, Carvalho, Fisher, Russell & Tippett, 2008, p. 376).

Table 1. Common cyberbullying harassments

Types of Harassment	Description
Name-calling	The use of language to defame, demean or degrade individuals or groups
Flaming	Posting hostile and insulting messages to intentionally harass somebody
Exposure	Posting or forwarding of personal communication, images or video which contain sensitive personal information or are detrimental to an individual
Excluding	Cruelly excluding someone from an online group
Denigration	Sending, posting, or publishing cruel rumors, gossip and untrue statements about an individual to intentionally damage their reputation or friendships
Impersonation	Impersonating somebody to make unpopular online comments or creating fake profiles to offer defaming information
Image and video dissemination	Images and video are disseminated to humiliate and disparage an individual
Happy slapping	A victim is physically attacked or embarrassed in person and video or pictures of the incident are recorded and diffused
Tagging	Inappropriate images and texts are tagged so that hurtful information appears on victim timeline
Cyberstalking	Cyberstalking includes threats of harm, intimidation, and/or offensive comments sent through Information and Communications Technology channels
Threats	Threatening a victim by cell calls or messages
Blackmail	Personal information are used to obtain ransom payments

[cyberbullying is when someone] repeatedly makes fun of another person online or repeatedly picks on another person through e-mail or text messages or when someone posts something online about another person that they don't like. (Hinduja & Patchin, 2012, p. 540)

Cyberbullying involves the use of information and communication technologies

to support deliberate, repeated, and hostile behavior by an individual or group, that is intended to harm others.⁵

The above mentioned European project of 2013, after a thorough examination of bullying and cyberbullying characteristics concluded:

[...] there is a large agreement within the literature about the definition of cyberbullying based on the criteria of intentionality and power imbalance [...] A lower level of agreement characterizes the criterion of repetition: qualitative studies underlined it is very relevant and literature on general aggression is likely to stress this criterion of specific of bullying and therefore of cyberbullying. (Menesini et al., 2013, p. 33)

Bullying vs. Cyberbullying

The Norwegian researcher Olweus (1993) claimed that *intentional aggressiveness*, *repetitively*, and *power imbalance* are the three key components bullying.

The bullies' behavior is aggressive and intentionally designed to hurt others or their relationship, or even to frighten them. Bullying behavior often occurs without an apparent provocation.

Beran and Li (2005), referring to cyberbullying, used the expression "old wine in a new bottle". They claim that cyberbullying is a form of bullying people by using new technologies that bullies handle just as a tool to attack and harass their victims. For certain aspects, this position was a reaction to the idea of those researchers who argued that cyberbullying was a specific and completely new phenomenon.

The contiguity between bullying and cyberbullying seems confirmed by the data of numerous investigations (Ybarra & Mitchell, 2004; Beran & Li, 2008; Kowalski, Limber & Agatston, 2008; Perren, Dooley, Shaw, & Cross, 2010; Sourander et al., 2010; Del Rey, Elipe & Ortega-Ruiz, 2012).

Data shows that there is an overlapping (44%) between online victims and victims of physical aggressions in the real world. In particular, either Beran and Li or Kowalski and Limber found that about 60% of victims of bullying are also victims of cyberbullying. However, if 40% of victims are only bullied online, this means that cyberbullying also discloses an autonomous and specific nature. This aspect is indicative of the complexity of the problem and should be carefully considered to understand cyberbullying and prevent and combat it. The fact that cyberbullying occurs not only as a pure extension of bullying but also as something specific suggests that different prevention strategies should be designed for the different contexts.

Data from a German study (Riebel, Jaeger & Fischer, 2009) confirms that over 80% of cyberbullies are bullies in real life.

But what really is cyber-bullying?

The literature contains many definitions of cyber-bullying (Smith, Cowie, Olafsson & Liefvooghe, 2002; Patchin & Hinduja, 2006; Ybarra, Mitchell, Wolak & Finkelhor, 2006; Ybarra & Mitchell, 2007; Dooley, Pyżalski & Cross, 2009; Ortega-Ruiz & Núñez, 2012; Heirman & Walrave, 2012; Wachs, Wolf & Pan, 2012; Palladino, Nocentini & Menesini, 2012; Paul, Smith & Blumberg, 2012; Vandebosch, Beirens, D'Haese, Wegge & Pabian, 2012; Rodríguez-Franco et al., 2012).

Many researchers share the definition of cyberbullying by Kowalski and Limber (2007) who consider cyberbullying as a particular form of bullying through new media and new digital communication devices, such as mobile phones, e-mail, the Internet, social networks, web pages, chats, and blogs. Cyberbullies are depicted how individuals who threaten, terrorize, defame, humiliate, offend their victims at a distance, and use images, movies, e-mail, messages, and comments on the Internet to harass and harm their victims.

Cyberbullying Characteristics

Although traditional bullying and cyberbullying share many common characteristics, some distinctive features have been identified (Dooley, Pyżalski & Cross, 2009; Smith et al., 2008). Two are very relevant: the audience of the traditional bullying is formed of small groups of peers, whilst that of cyberbullies can be very vast and uncontrolled; furthermore, cyberbullies may be little aware or even completely unaware of the consequences of their evil actions.

Furthermore, there are some other elements which are exclusive of cyberbullying. One is due to the nature of technology and is the lack of control over it. For example, once a digital harmful content (sms, text, images or video) has been sent or uploaded to the Internet, this can be downloaded, transmitted and distributed by anyone, not just by its original author. As a consequence of this, the flow of diffusion cannot in any way be arrested. The victim may also experience the same aggression repeatedly and without the direct intervention of the perpetrators: access and execution of the harmful digital content it can be repeated by the victim. Internet is over the control of both perpetrators and victims. Evil contents remains on the Web indefinitely and consequences are unpredictable.

Bullies can take advantage of some features of online communication (anonymity, asynchrony and accessibility) to harass their victims continuously and automatically (Valkenburg & Peter, 2011).

The same notion of *power imbalance* can assume different and more complex connotations (Law, Shapka, Hymel, Olson & Waterhouse, 2012). For example, it has been argued that cyberbullies have a deep understanding of technology and take advantage of their anonymity. The reality may be quite different: cyberbullies could be less technologically skilled than their victims and use the technology only as an extension of their hurtful

actions, for example, to document aggressions and humiliations through smartphone videocameras. Many traditional acts of bullying, once filmed and posted on YouTube, become acts of cyberbullying. It has been observed that anonymity provides to victims of traditional bullying the opportunity to revenge themselves online, attacking in the virtual world who torment them in the real world (Kowalski et al., 2008). An online survey of 473 students (König, Gollwitzer & Steffgen, 2010) investigated the phenomenon of revenge and retaliation as a motive to engage in acts of cyberbullying and showed that 149 cyberbullies were victims of traditional bullying.

Notwithstanding the obvious individual differences, it seems that cyberbullies tend to choose their victims among their former persecutors.

In Table 2, the most common effects of acts of cyberbullying on their victims are synthesized, referring them to the emotional, behavioral, and cognitive spheres.

The more one studies cyberbullying, the more one realizes that this is a multifaceted and complex phenomenon.

In a very simple way, cyberbullying discloses two different faces, due to the different use of communications technology. Online bullying cannot simply be deemed as an extension of the traditional

bullying. To continue to claim that it is “new wine in old bottles” (McGuckin, Cummins & Lewis, 2010) leaves in the shadow important aspects of the phenomenon. If it is true that the victims of cyberbullying often know their persecutors and many dynamics which tie persecutors to victims are the same as in traditional bullying, cyberbullying shows something new. This concerns the nature of virtual relations and how they are perceived and managed today, especially by young people. Cyberbullying is evil and dangerous not only because it involves the sphere of intimacy of individuals and their reputation, but also because it undermines the communicative dimension. Stop to be connected to avoid harassment is tantamount, in particular for young people, to close their relationships with the world.

On this purpose, we have to consider that mostly of victims of cyberbullying are teenagers of 12-13 years. The research attributed this to several factors: the transition to the secondary school, a greater degree of autonomy from parents, the availability of smartphones, the growing interest to socialize, and so on. There is an element that should not be undervalued: young people are not likely to report their persecutors. They are afraid that adults will prevent them the use of the Internet and smartphones, and sometimes are even ashamed to show what circulates about them.

Table 2. Cyberbullying consequences

Type of Consequence	Appearance
Emotional	• Anxiety symptoms • Depressive states • Sense of shame • Sense of abandonment and despair • Various somatizations
Behavioral	• Exclusion, marginalization • Sleeping and eating disorders • Posture problems • Aggressive responses • Self-harm • Substance abuse
Cognitive	• Attention disturbs • Learning and speech difficulties • Confusion • Poor academic results

FUTURE RESEARCH DIRECTIONS

Cyberbullying is a complex phenomenon and can be affected by numerous contextual factors that involve not only the personality of perpetrators, victims, bystanders, educators, and parents, but also their cultural sphere: cultural attitudes may influence cyberbullying, or at least its perception around the social context (Boronenko, Menshikov & Marzano, 2013).

Future research should explore this context more deeply and investigate the links between cyberbullying episodes, public reactions, and policy adoptions.

Furthermore, cyberbullying is a topic that encompasses several investigation perspective and a common effort is needed to overcome the current limitations of general reviews and the absence of large empirical data and analysis.

On this regard, a new interesting research area concerns the changes in the sexual behavior, especially among young people, and the related use of smartphones. (Espelage, 2016).

In the next paragraphs we highlight some new interesting research issues tied to two alarming emerging problems.

Revenge Porn

A new phenomenon which is so-called revenge pornography or revenge porn is recently spreading. It involves a perpetrator – typically an ex-partner from a previous romantic relationship – uploading nude or semi-nude images/videos of a person online without the consent of the person appearing in the content.

It has been observed that, despite recent media attention on revenge porn, relatively few academic studies have focused on this issue, and those that do concentrate mainly on the legal aspects and legal theories related to revenge porn cases.

Homophobic Cyberbullying

Many sites of gay associations denounce the existence of a homophobic cyberbullying, and argue on the spread of this phenomenon. Finding of a survey conducted at British schools in 2012 by the *Centre for Family Research* of the Cambridge University are broadly diffused. This data is also reported on the major networks engaged combating cyberbullying.

Despite the seriousness and topicality of the problem, few studies are available on discriminatory aspects of cyberbullying, including racism and homophobia. Scientific investigations on homophobic cyberbullying are rather limited and data doesn't allow an effective assessment of this phenomenon. For example, in Canada, there are

studies that show conflicting data. According to a recent study, 29.4% of young gays are victims of homophobic cyberbullying (Cénat, Blais, Hébert, Lavoie & Guerrier, 2015); while a previous survey reported that the victims of homophobic cyberbullying are about 61% (Blais, Gervais, & Hébert, 2014).

At the moment, it is not well defined what homophobic cyberbullying means, and researchers have not developed common methodologies of analysis.

Certainly, investigating homophobic component within the complex area of cyberbullying could contribute to a deeper understanding of this phenomenon. However, emotional generalizations induced by gay activists and media exploitation of some extreme cases of online homophobic harassment don't facilitate the scientific analysis.

CONCLUSION

Many advices about preventing cyberbullying concern training children and their parents in e-safety and implementing technological tools to counteract the cyberbullies' behaviour, such as blocking bullying behaviour online or creating panic buttons for cyber victims that they can push when feeling under threat (Marzano, 2015).

However, designing preventive strategies is not simple at all: effective preventing programs need the knowledge of new media and the understanding of their use by the different typologies of users, especially by adolescents and young people.

There is a general consensus that prevention through education is the better strategy against cyberbullying. All researchers and educators concord with the opinion that criminalization is not the right answer to combating cyberbullying, since its negative consequences outweigh the possible positive effects. Internet safety education has proven effective, and empowering educators with theoretical and practical tools is an extremely valid key.

Peer education programs and peer support models are also deemed useful. They are based on the assumption that peers learn from one another, having significant influence on each other. Conducts and behaviours are most likely to change when liked and trusted group members take the lead in changing the situation (Naylor & Cowie 1999; Turner & Shepherd, 1999; Maticka-Tyndale & Barnett 2010).

There are several curriculum based programs available to educators that strive to help in preventing cyberbullying (Snakenborg, Van Acker & Gable, 2011). Some examples of cyberbullying combating programs are also available on the Internet, while prevention programs focus at school level appear to be more efficient (Couvillon & Ilieva, 2011).

However, in fighting cyberbullying, we need a continuous monitoring of both technologies and their use (see the recent anonymous social networks' spread).

REFERENCES

- Bartolini, C., & Siry, L. (2016). The right to be forgotten in the light of the consent of the data subject. *Computer Law & Security Report*, 32(2), 218–237. doi:10.1016/j.clsr.2016.01.005
- Beran, T., & Li, Q. (2005). Cyber-harassment: A study of a new method for an old behavior. *Journal of Educational Computing Research*, 32(3), 265–277. doi:10.2190/8YQM-B04H-PG4D-BLLH
- Beran, T., & Li, Q. (2008). The relationship between cyberbullying and school bullying. *The Journal of Student Wellbeing*, 1(2), 16–33. doi:10.21913/JSW.v1i2.172
- Blais, M., Gervais, J., & Hébert, M. (2014). Internalized homophobia as a partial mediator between homophobic bullying and self-esteem among youths of sexual minorities in Quebec (Canada). *Ciencia & Saude Coletiva*, 19(3), 727–735. doi:10.1590/1413-81232014193.16082013 PMID:24714888
- Boronenko, V., Menshikov, V., & Marzano, G. (2013). Topicality of Cyberbullying Among Teenagers in Russia and Latvia. *Social Sciences Bulletin*, 1(16), 84–104.
- Cénat, J. M., Blais, M., Hébert, M., Lavoie, F., & Guerrier, M. (2015). Correlates of bullying in Quebec high school students: The vulnerability of sexual-minority youth. *Journal of Affective Disorders*, 183, 315–321. doi:10.1016/j.jad.2015.05.011 PMID:26047959
- Corcoran, L., Guckin, C. M., & Prentice, G. (2015). Cyberbullying or cyber aggression?: A review of existing definitions of cyber-based peer-to-peer aggression. *Societies*, 5(2), 245–255. doi:10.3390/soc5020245
- Couvillon, M. A., & Ilieva, V. (2011). Recommended practices: A review of schoolwide preventative programs and strategies on cyberbullying. *Preventing School Failure: Alternative Education for Children and Youth*, 55(2), 96–101. doi:10.1080/1045988X.2011.539461
- Del Rey, R., Elípe, P., & Ortega-Ruiz, R. (2012). Bullying and cyberbullying: Overlapping and predictive value of the co-occurrence. *Psicothema*, 24(4), 608–613. PMID:23079359
- Dooley, J. J., Pyżalski, J., & Cross, D. (2009). Cyberbullying versus face-to-face bullying: A theoretical and conceptual review. *Zeitschrift für Psychologie. The Journal of Psychology*, 217(4), 182–188.

- Espelage, D. L. (2016). Sexual orientation and gender identity in schools: A call for more research in school psychology—No more excuses. *Journal of School Psychology, 54*, 5–8. doi:10.1016/j.jsp.2015.11.002 PMID:26790698
- Finn, J. (2004). A survey of online harassment at a university campus. *Journal of Interpersonal Violence, 19*(4), 468–483. doi:10.1177/0886260503262083 PMID:15038885
- Heirman, W., & Walrave, M. (2012). Predicting adolescent perpetration in cyberbullying: An application of the theory of planned behavior. *Psicothema, 24*(4), 614–620. PMID:23079360
- Hinduja, S., & Patchin, J. W. (2012). Cyberbullying: Neither an epidemic nor a rarity. *European Journal of Developmental Psychology, 9*(5), 539–543. doi:10.1080/17405629.2012.706448
- König, A., Gollwitzer, M., & Steffgen, G. (2010). Cyberbullying as an Act of Revenge? *Australian Journal of Guidance & Counselling, 20*(02), 210–224. doi:10.1375/ajgc.20.2.210
- Kowalski, R. M., Limber, S. P., & Agatston, P. W. (2008). *Cyber bullying*. Malden, MA: Blackwell. doi:10.1002/9780470694176
- Law, D. M., Shapka, J. D., Hymel, S., Olson, B. F., & Waterhouse, T. (2012). The changing face of bullying: An empirical comparison between traditional and internet bullying and victimization. *Computers in Human Behavior, 28*(1), 226–232. doi:10.1016/j.chb.2011.09.004
- Marr, N., & Field, T. (2001). *Bullycide: Death at playtime*. Success Unlimited.
- Marzano, G. (2015). Cyberbullying Prevention: Some Preventing Tips. In Sahlim J. (ed) *Social Media and the Transformation of Interaction in Society*, IGI Global, 133–157.
- Maticka-Tyndale, E., & Barnett, J. P. (2010). Peer-led interventions to reduce HIV risk of youth: A review. *Evaluation and Program Planning, 33*(2), 98–112. doi:10.1016/j.evalprogplan.2009.07.001 PMID:19647874
- McGuckin, C., Cummins, P., & Lewis, C. A. (2010). New bottle: Old wine! School bullying among primary school pupils and the emergence of cyberbullying. *Australian Journal of Guidance & Counselling, 2*(20), 210–224.
- Menesini, E., Nocentini, A., Palladino, B. E., Sheithauer, H., Schultze-Krumbholz, A., & Frisé, A. et al. (2013). Definition of cyberbullying. In P. K. Smith & G. Steffgen (Eds.), *Cyberbullying through the new media: Findings from an international network* (pp. 3–36). Psychology Press.
- Naylor, P., & Cowie, H. (1999). The effectiveness of peer support systems in challenging school bullying: The perspectives and experiences of teachers and pupils. *Journal of Adolescence, 22*(4), 467–479. doi:10.1006/jado.1999.0241 PMID:10469511
- Olweus, D. (2012). Cyberbullying: An overrated phenomenon? *European Journal of Developmental Psychology, 9*(5), 520–538. doi:10.1080/17405629.2012.682358
- Olweus, D., & Limber, S. P. (2010). Olweus bullying prevention program. The handbook of bullying in schools: An international perspective, 377–401.
- Ortega-Ruiz, R., & Núñez, J. C. (2012). Bullying and cyberbullying: Research and intervention at school and social contexts. *Psicothema, 24*(4), 603–607. PMID:23079358
- Palladino, B. E., Nocentini, A., & Menesini, E. (2012). Online and offline peer led models against bullying and cyberbullying. *Psicothema, 24*(4), 634–639. PMID:23079363

- Patchin, J. W., & Hinduja, S. (2006). Bullies move beyond the schoolyard a preliminary look at cyberbullying. *Youth Violence and Juvenile Justice*, 4(2), 148–169. doi:10.1177/1541204006286288
- Paul, S., Smith, P. K., & Blumberg, H. H. (2012). Investigating legal aspects of cyberbullying. *Psicothema*, 24(4), 640–645. PMID:23079364
- Perren, S., Dooley, J., Shaw, T., & Cross, D. (2010). Bullying in school and cyberspace: Associations with depressive symptoms in Swiss and Australian adolescents. *Child and Adolescent Psychiatry and Mental Health*, 4(28), 1–10. PMID:21092266
- Peterson, J. K., & Densley, J. (2016). Is Social Media a Gang? Toward a Selection, Facilitation, or Enhancement Explanation of Cyber Violence. *College of Liberal Arts All Faculty Scholarship*. Retrieved from http://digitalcommons.hamline.edu/cgi/viewcontent.cgi?article=1004&context=cla_faculty
- Riebel, J. R. S. J., Jaeger, R. S., & Fischer, U. C. (2009). Cyberbullying in Germany—an exploration of prevalence, overlapping with real life bullying and coping strategies. *Psychology Science Quarterly*, 51(3), 298–314.
- Rodríguez-Franco, L., Borrego, J. L. C., Rodríguez-Díaz, F. J., Molleda, C. B., Pineda, C. E., Bellerín, M. Á. A., & Quevedo-Blasco, R. (2012). Labeling dating abuse: Undetected abuse among Spanish adolescents and young adults1. *International Journal of Clinical and Health Psychology*, 12(1), 55.
- Smith, P. K., Cowie, H., Olafsson, R. F., & Liefooghe, A. P. D. SMITH. (2002). Definitions of bullying: A comparison of terms used, and age and gender differences, in a fourteen-country international comparison. *Child Development*, 73(4), 1119–1133. doi:10.1111/1467-8624.00461 PMID:12146737
- Smith, P. K., Mahdavi, J., Carvalho, M., Fisher, S., Russell, S., & Tippett, N. (2008). Cyberbullying: Its nature and impact in secondary school pupils. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 49(4), 376–385. doi:10.1111/j.1469-7610.2007.01846.x PMID:18363945
- Smith, P. K., & Steffgen, G. (2013). *Cyberbullying through the new media: Findings from an international network*. Psychology Press.
- Snakenborg, J., Van Acker, R., & Gable, R. A. (2011). Cyberbullying: Prevention and intervention to protect our children and youth. *Preventing School Failure: Alternative Education for Children and Youth*, 55(2), 88–95. doi:10.1080/1045988X.2011.539454
- Sourander, A., Klomek, A. B., Ikonen, M., Lindroos, J., Luntamo, T., Koskelainen, M., & Helenius, H. et al. (2010). Psychosocial risk factors associated with cyberbullying among adolescents: A population-based study. *Archives of General Psychiatry*, 67(7), 720–728. doi:10.1001/archgenpsychiatry.2010.79 PMID:20603453
- Temple, J. R., Choi, H. J., Brem, M., Wolford-Clevenger, C., Stuart, G. L., Peskin, M. F., & Elmquist, J. (2016). The temporal association between traditional and cyber dating abuse among adolescents. *Journal of Youth and Adolescence*, 45(2), 340–349. doi:10.1007/s10964-015-0380-3 PMID:26525389
- Thomas, H. J., Connor, J. P., & Scott, J. G. (2015). Integrating traditional bullying and cyberbullying: Challenges of definition and measurement in adolescents—A review. *Educational Psychology Review*, 27(1), 135–152. doi:10.1007/s10648-014-9261-7
- Turner, G., & Shepherd, J. (1999). A method in search of a theory: Peer education and health promotion. *Health Education Research*, 14(2), 235–247. doi:10.1093/her/14.2.235 PMID:10387503

Valkenburg, P. M., & Peter, J. (2011). Online communication among adolescents: An integrated model of its attraction, opportunities, and risks. *The Journal of Adolescent Health, 48*(2), 121–127. doi:10.1016/j.jadohealth.2010.08.020 PMID:21257109

Vandebosch, H., Beirens, L., D’Haese, W., Wegge, D., & Pabian, S. (2012). Police actions with regard to cyberbullying: The Belgian case. *Psicothema, 24*(4), 646–652. PMID:23079365

Wachs, S., Wolf, K. D., & Pan, C. C. (2012). Cybergrooming: Risk factors, coping strategies and associations with cyberbullying. *Psicothema, 24*(4), 628–633. PMID:23079362

Ybarra, M. L., & Mitchell, K. J. (2004). Online aggressor/targets, aggressors, and targets: A comparison of associated youth characteristics. *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 45*(7), 1308–1316. doi:10.1111/j.1469-7610.2004.00328.x PMID:15335350

Ybarra, M. L., & Mitchell, K. J. (2007). Prevalence and frequency of Internet harassment instigation: Implications for adolescent health. *The Journal of Adolescent Health, 41*(2), 189–195. doi:10.1016/j.jadohealth.2007.03.005 PMID:17659224

ADDITIONAL READING

Navarro, R., Yubero, S., & Larrañaga, E. (2016). *Cyberbullying Across the Globe*. Springer. doi:10.1007/978-3-319-25552-1

Sahlim, J. (Ed.). (2015). *Social Media and the Transformation of Interaction in Society*. IGI Global. doi:10.4018/978-1-4666-8556-7

Shore, K. (2016). *The Bullying Prevention Book of Lists: A Practical, Easy-to-Use Guide for All School Staff*. National Professional Resources Inc. Dude Publishing.

Sigmund, T. P. (Ed.). (2016). *Handbook of Research on Individualism and Identity in the Globalized Digital Age*. IGI Global.

KEY TERMS AND DEFINITIONS

Anonymous Social Media: Social media that ensure members anonymity.

Bullicide: Suicide due bullying.

Bystanders: Persons that make decision to stay out of cyberbullies aggression and don’t intervene.

Cyberspace: The online world of computer networks and the Internet.

Cyberspace Relative Anonymity: The relative possibility to perform activities on the Web without being identified.

Homophobia: Is the irrational fear or hatred of homosexuality.

Revenge Porn: Is the publication of sexual explicit material portraying someone who has not consented for the image or video to be shared.

ENDNOTES

- ¹ <http://cyberbullying.org/summary-of-our-cyberbullying-research>; last accessed 19.06.2016.
- ² <http://www.violencepreventionworks.org/public/index.page>; last accessed 19.06.2016.
- ³ <http://beltway.bsgco.com/content/privacy-and-tech>; last accessed 19.06.2016.
- ⁴ See also Robert S. Tokunaga, 2010 (p. 278) for conceptual definitions of cyberbullying used in research.
- ⁵ Belsey, B., *Cyberbullying. An emerging threat to the “always on” generation*, http://www.cyberbullying.ca/pdf/Cyberbullying_Article_by_Bill_Belsey.pdf; last accessed 23.06.2016.

Defining and Conceptualizing Cyberbullying

Karin Spenser

Nottingham Trent University, UK

Lucy R. Betts

Nottingham Trent University, UK

INTRODUCTION

The word ‘bully’ dates back to the 16th century and the original meaning is somewhat different to that of today. Originally a term of endearment implying friendly admiration but over time the meaning of bullying has darkened to “a person who uses strength or influence to harm or intimidate those who are weaker” (Oxford English Dictionary, 1989). Until recently face-to-face bullying was seen by some as a normal and acceptable part of youth (Koo, 2007). The practice, tolerated within England’s public school system for decades, appears in both fiction and real life. In *Tom Brown’s School Days*, the character Flashman is seen bullying the younger Tom Brown; the line “‘very well then, let’s roast him’ [...]. One or two boys hesitate, but the rest join in” (Hughes, 1857, p. 188), being the most famous. At a similar time, the death of a 12 year old boy attending the King’s School in Cambridge, was made public. The boy’s death, in 1885, was attributed to experiences of bullying. Afterwards, a former student wrote to the editor of *The Times* newspaper saying “bullying, of the kind mentioned, constantly occurred during the seven years I was at the school” (Anon, 1885).

The turning point, in recognizing face-to-face bullying as a problem, came in the 1970s. At that time, a general concern regarding violent behavior emerged and the pioneering work of Olewus (1978) began to address young people’s experiences of bullying. Since then numerous studies have examined various forms of face-to-face bullying

(e.g., Hawker & Boulton, 2000). However, in the wake of societal and technological changes bullying has evolved (Hinduja & Patchin, 2008). Recent statistics report that 99% of all 12- to 14-year-olds in the UK are frequent internet users with young people spending increasing amounts of time using technology (Office for National Statistics, 2013). With the many benefits of this connectivity, such as enhanced educational attainment (Jackson, 2011) and social networks (Valkenburg & Peter, 2007), there are also risks. Cyberbullying has been identified as one of the greatest risks (Cross, Monks, Campbell, Spears, & Slee, 2011). Cyberbullying is “the intentional act of online/digital intimidation, embarrassment, or harassment.” (Mark & Ratliffe, 2011, p. 92).

BACKGROUND

Although cyberbullying is undoubtedly a by-product of the union of adolescent aggression and electronic communication; it is its propensity for growth which gives cause for concern for researchers and educational practitioners (Cassidy, Faucher, & Jackson, 2013). Further, empirical evidence reports that the impacts of cyberbullying include: distress (Li, 2010), loneliness (Sahin, 2012), depression (Tynes, Rose, & Williams, 2010), increased psychosomatic symptoms (Sourander et al., 2010), suicidal ideation (Hinduja & Patchin, 2010), and reduced academic performance (Smith et al., 2008).

DOI: 10.4018/978-1-5225-2255-3.ch361

Despite this attention, many questions remain unanswered with regard to the conceptual and theoretical similarities between face-to-face bullying and cyberbullying. It is widely accepted that definitions of face-to-face bullying include aspects of repetition, power imbalance, and intention (Olweus, 2013). There are three forms of face-to-face bullying: physical, verbal, and social (Rigby, 1997). Physical bullying is a ‘direct’ form of aggression that involves hitting, punching, kicking, or any other action that can inflict physical pain or harm. The power imbalance between the perpetrator and the target in physical bullying makes it difficult for the target to defend themselves and prevent the actions being repeated (Rigby, 2002).

Verbal bullying which includes making rude remarks, telling hurtful jokes about an individual, calling an individual hurtful names, and threatening an individual with the intent to intimidate or humiliate the target, or to provide ‘humor’ at the expense of the target (Crick et al., 2001) can be direct or indirect. The perpetrator can ‘directly’ bully by teasing, taunting, or mocking the target to their face, or ‘indirectly’ bully by doing the same thing behind the target’s back (Cole, Cornell, & Sheras, 2006).

Social bullying is aggressive behavior in the form of rumor spreading, backbiting, and social exclusion that results in harming an individual’s psychological state of mind and/or social connections. Often delivered in a covert manner social bullying allows the perpetrator to remain anonymous (Björkqvist, Österman, & Kaukiainen, 1992).

Issues, Controversies, Problems

The reported prevalence rates of cyberbullying vary from 6.5% (Jones, Mitchell, & Finkelhor, 2012) to 72% (Juvonen & Gross, 2008) with frequencies converging between 20 to 40% (Hinduja & Patchin, 2008). Typically, cyberbullying experiences peak around the age of 14 (Ortega et al., 2009), although those most at risk of being involved

in cyberbullying are adolescents and young adults (Ševčíková & Šmahel, 2009). Tokunaga (2010) proposed that a holistic view of cyberbullying should be adopted to include: aggression, hostile, and harmful acts via an electronic device. Clearly, not all acts encountered online may be intended or interpreted as cyberbullying; young people make the distinction between cyberbullying behaviors and banter (Spenser & Betts, 2014). Banter is an “interactional bonding game” (Dyrel, 2008, p. 246) that is interpreted as playful by participating individuals rather than having sinister undertones.

Willard (2007) suggested eight forms of cyberbullying: Flaming (angry and vulgar online messages); harassment (repeated sending of nasty and insulting messages to the victim); denigration (spreading of rumors and gossiping about a person online to damage his/her reputation or friendship); impersonation (causing someone to get into trouble or damaging someone’s reputation by pretending to be that person); outing (sharing secrets or humiliating information); trickery (convincing someone to share humiliating information, then making it available); exclusion (intentionally excluding someone from an online group in order to cause hurt); and cyber stalking (repeatedly harassing someone such that the person feels threatened or afraid). There are a number of high profile cases where the types of cyberbullying identified by Willard have been attributed as a contributing factor to young people’s suicide. For example, Ryan Halligan committed suicide after being duped by an individual who befriended him, humiliated him, and then sent threats via instant messenger which were subsequently circulated to the wider peer group (Moreno, 2011). Clearly parallels exist between face-to-face bullying and cyberbullying; however, cyberbullying has some unique characteristics. The chapter will now discuss seven facets of cyberbullying that are particularly pertinent to the definition and conceptualization of cyberbullying: anonymity, access, repetition, permanency, publicity, power, and identity and motivation.

Anonymity

Physical and verbal face-to-face bullying identify the perpetrator to the target. In contrast, cyberbullying allows the perpetrator to hide their identity through creating: anonymous screen names and web pages, false identities, and unidentified numbers. Indeed, Kowalski and Witte (2006) found that 74% of targets of cyberbullying did not know the identity of their attacker. More recently Spenser and Betts (2014) reported participants believed anonymity may actually empower and encourage individuals to engage in cyberbullying. Anonymity can have a disinhibiting effect on the perpetrator of cyberbullying, leading individuals who might not otherwise engage in face-to-face bullying behaviors to do so more comfortably online (Vandebosch & van Cleemput, 2009). In other words the lack of immediate consequences combined with the inability to see the targets' emotional reactions often means that self-regulatory processes, employed during face-to-face contact, are disregarded (Kowalski, 2009). From the point of view of the target, being unaware who is causing the harm can be very daunting and can cause negative emotions such as frustration, insecurity, and fear (Dooley, Pyzalski, & Cross, 2009). Consequently, future research should explore the role and effect of anonymity and the extent to which targets of cyberbullying know the identity of their perpetrator.

Access

For the most part, face-to-face bullying takes place at school; thus giving the target a chance to retreat to a 'safe' place after school hours. Conversely, cyberbullying can occur anywhere and at any time. Through technology, the cyberbully is able to extend their bullying beyond the school grounds and 'follow' their target home. Consequently, the target is potentially accessible 24 hours a day. Recent research by Betts and Spenser (2015) reports that teachers perceive the fear of receiving

a message at any time to have a greater impact on young people who are the target of cyberbullying than the content of the message. The only way to avoid these messages is for the target to 'dispose' of their electronic devices. However, this does not stop the perpetrator from posting messages for the online community to see (Kowalski, 2009). Further, young people are unlikely to give up their electronic devices because of the fear of missing out. Fear of missing out is characterized by the desire to remain continually connected to social media (Przybylski, Murayama, DeHann, & Gladwell, 2013). This fear is motivated by the belief that others are having 'rewarding experiences' that the individual is not part of and is exacerbated through social media when individuals learn that others have had such experiences. Therefore, when exploring cyberbullying researchers must ensure that young people's reports do not subsequently impact on their access to technology as a way to try to overcome the under-reporting of cyberbullying.

Repetition

Olweus (1993) argued that, as face-to-face bullying behaviors take many forms, it is the repetitive nature of the act that is crucial rather than the behavior *per se*. Repetition in face-to-face bullying serves to instill fear and cause psychological harm to a target although in cyberbullying it is more difficult to operationalize. Whilst repetition is clear when a perpetrator posts numerous messages (Slonje & Smith, 2008), it is not so clear when a perpetrator creates a single derogatory comment, which is then repeatedly accessed resulting in widespread ridicule and humiliation. Further, not all forms of cyberbullying are equal in terms of impact on the target (Smith et al., 2008). For example, receiving an offensive text message is not the same as someone posting an embarrassing or humiliating photograph or video of the target online. Therefore, it follows that some acts may not need to be repeated (or repeated as often) to induce the same or greater harm (Slonje & Smith,

2008) and, as such the frequency of repetition should be controlled for in future studies assessing cyberbullying.

Permanency

Whilst undoubtedly unpleasant, once performed the act of face-to-face bullying is over. People may have witnessed the act of face-to-face bullying and they may talk about it for a while, but it is likely to then be forgotten. However, with cyberbullying the act potentially never goes away. When internet users visit various web sites, they leave behind evidence of which sites they have visited. This collective, ongoing record of one's web activity is called the 'digital footprint' (Greysen, Kind, & Chretien, 2010). Often young people, who lack an awareness regarding privacy issues and post inappropriate messages, pictures, and videos, do not fully appreciate the fact that 'what goes online stays online' (Palfrey, Gasser, & Boyd, 2010). As a result a young person's future may be put into jeopardy by inexperienced and rash 'clicks of the mouse' (O'Keeffe & Clarke-Pearson, 2011). Consequently, longitudinal studies are required to determine the psychosocial effects of the relative permanency of cyberbullying episodes.

Publicity

It has been noted that one of the most distressing aspects of face-to-face bullying is the effect of the group (Crick, Grotpeter, & Bigbee, 2002). This aspect of face-to-face bullying may have similarities with cyberbullying in that when cyberbullying episodes are public the impact of the behaviour is likely to be more serious. However, the audience for cyberbullying is potentially limitless (Nocentini et al., 2010).

Aligned to the public nature of cyberbullying is the controllability of the situation. The ability to control what is posted on line by others is completely beyond the targets' reach. This lack of control is associated with feelings of helplessness, helpless reactions, and helpless coping strategies

(Spears, Slee, Owens, & Johnson, 2009), which are in turn associated with depressive symptoms (Machmutow, Perren, Sticca, & Alsaker, 2012). Consequently, the size and scope of the audience who witness cyberbullying episodes should be assessed by researchers to gain insight in to the impact of publicity.

Power

Power imbalance is a fundamental aspect of bullying with Olweus (1978) making reference to the 'weak' target who is both physically and psychologically weak. Power can be social, psychological, or physical in nature (Monks & Smith, 2006). However, assessing the power imbalance in cyber-based interactions is more complicated. One study undertaken by Vandebosch and van Cleemput (2008) with 10- to 18- year-olds found that those individuals who were considered to be more powerful in face-to-face interactions were sometimes targeted online due to the anonymity that technology affords. Consistent with this empirical evidence, Fauman (2008) suggested that the ability to remain anonymous may minimize the necessity for those that engage in cyberbullying to be more powerful than their targets. In addition, whilst the perpetrator of cyberbullying may be perceived as being more technically advanced than their target, in reality it does not require an advanced ability to take a picture using a mobile phone camera and send it to others. Similarly, posting a picture online or creating a fake social network site profile requires little skill (Smith et al., 2008). Therefore, when assessing cyberbullying, it is important to consider the real or imagined power dynamic between the perpetrator and target.

Identity and Motivation

As with most competitive situations, the social struggle amongst young people has a hierarchy. Face-to-face bullying is a component of this hierarchy. According to Olweus (2001), there are seven different levels within the bullying ladder:

those who want to bully and initiate the action, their followers or henchmen, supporters or passive bullies, passive supporters or possible bullies, disengaged onlookers, possible defenders, and defenders who dislike the action of bullying and help those that are target. However, the identity of the perpetrator of cyberbullying is less straight forward.

Aftab (2011) suggests that perpetrators of cyberbullying fall into five different groups. The ‘vengeful angels’ see themselves as righting wrongs or protecting themselves or others. The ‘power-hungry’ want to exert their authority and show that they can make others do what they want them to do such that they will continue to escalate their actions until they get a response. This group of cyberbullying perpetrators most closely resembles the perpetrators of face-to-face bullying. The ‘nerd’ targets individuals directly, keep their activities secret, and rarely appreciate the consequences of their actions. The ‘accidental cyber bully’ does not recognize that they are a perpetrator of cyberbullying and they do not lash out intentionally. ‘Mean girls’ are girls who engage in cyberbullying, have high social status and tend to target individuals to overcome boredom with the aim of damaging the targets’ social status. In addition to these groups, there is also evidence that perpetrators of cyberbullying have previously been the target of face-to-face bullying (Steffgen & König, 2009). Therefore, when examining cyberbullying researchers should examine the motives of those who act as perpetrators.

SOLUTIONS AND RECOMMENDATIONS

The characteristics of cyberbullying discussed so far in this chapter have implications for how cyberbullying is defined and operationalized to enable accurate and reliable data to be collected. Specifically, although anonymity, power, motivation, permanence of messages, access, and

publicity are recognized as distinct features of cyberbullying, one of the biggest challenges is how to capture these facets in tools designed to assess cyberbullying.

A recent review by Kowalski, Giumetti, Schroeder, and Lattanner (2014), argued that the definitions of cyberbullying comprised four distinct components: “(a) intentional aggressive behavior that is (b) carried out repeatedly, (c) occurs between a perpetrator and victim who are unequal in power, and (d) occurs through electronic technologies” (p. 1073). However, whilst these components relate to the facets of cyberbullying discussed in this chapter, it is not clear how these are captured in measures designed to assess cyberbullying. Therefore, the first recommendation is to present participants with these aspects of cyberbullying and ensure that they are all fully assessed in scales. Alternatively researchers could consider providing participants with a definition of cyberbullying or generating a ‘user’ defined conceptualization of cyberbullying. A final solution would be to ask participants to estimate the frequency with which they have experienced, or engaged in, a particular behavior to determine the repetition element of the definition. Participants could also be asked to report the perceived power dynamics and intention in the episode of cyberbullying. Developing appropriate strategies to assess cyberbullying will also enable researchers and practitioners to gain a clearer understanding of the prevalence rates of cyberbullying.

A second recommendation that arises from the literature discussed in this chapter relates to how cyberbullying can be distinguished from face-to-face bullying. In particular, it is important that when researchers design studies to assess cyberbullying that assessment tools distinguish between aspects of face-to-face experiences and cyberbullying. Researchers should also be mindful that young people may under-report their experiences of cyberbullying because of concerns associated with restricting access to technology. Together these recommendations will enable a more realistic appraisal of cyberbullying.

FUTURE RESEARCH DIRECTIONS

As alluded to in the section on solutions and recommendations, future research should consider how young people conceptualize cyberbullying. Taking a young person centered approach to the definition of cyberbullying, similar to Spenser and Betts (2014), will allow greater exploration in to how young people conceptualize behavior as cyberbullying and what the distinguishing features between cyberbullying and ‘banter’ are. Similarly, exploring teachers’, educational practitioners’, and parents’ understanding and conceptualization of cyberbullying similar to the work of Betts and Spenser (2015) will add to understanding in this area from alternative perspectives. Further, by gaining a greater appreciation of how young people and key stakeholders conceptualize cyberbullying this may go some way to address the current debate in the literature with regard to repetition, the power imbalance, and intention associated with cyberbullying.

In addition to developing and refining tools to assess cyberbullying, future research should also examine the extent to which the distinct facets of cyberbullying discussed in this chapter impact on young people’s psychosocial adjustment. Whilst there is growing evidence of the impact of cyberbullying on adjustment (e.g., Hinduja & Patchin, 2010) it may be that the different facets of cyberbullying have influence adjustment through different mechanisms. Therefore, future research should further explore this area.

CONCLUSION

In summary, cyberbullying differs from face-to-face bullying in a number of ways. Whilst these are not absolute differences they may affect other aspects such as motives for perpetration and impact on targets. Although there is some consensus in the

academic and practitioner literature that cyberbullying comprises a set of aggressive behaviors that are enacted via electronic media, relatively little is known about some aspects of cyberbullying. For example, there is limited research examining the motivations and goals of those who engage in cyberbullying and the long-term impact of experiencing cyberbullying either as a perpetrator or as a target. This in turn makes it difficult to develop interventions to address this behavior. With the increasing availability, use and reliance on electronic technology, the issues outlined here are going to become progressively more important and are clearly worthy of far greater understanding. There is a clear need for further in-depth research addressing issues of anonymity, access, repetition, permanency, publicity, power, identity, and motivation.

REFERENCES

- Aftab, P. (2011). *Wired safety: A parent’s guide to cyberbullying*. Available at https://www.wiredsafety.org/toolkitmedia/files/file/Parent_s_Articles/A_Parent_s_Guide_to_Cyberbullying_-_Extended.pdf
- Anon. (1862, August 6). What is bullying. *The Times*, Col. F, p. 8.
- Betts, L. R., & Spenser, K. A. (2015). “A large can of worms”: Teachers’ perceptions of young people’s experiences using technology. *International Journal of Cyber Behavior, Psychology and Learning*, 5, 15-29. doi: 10.4018/ijcbpl.2015040102
- Björkqvist, K., Österman, K., & Kaukiainen, A. (1992). The development of direct and indirect aggressive strategies in males and females. In K. Björkqvist & P. Niemela (Eds.), *Of mice and women: Aspects of female aggression* (pp. 51–64). San Diego, CA: Academic Press. doi:10.1016/B978-0-12-102590-8.50010-6

- Cassidy, W., Faucher, C., & Jackson, M. (2013). Cyberbullying among youth: A comprehensive review of current international research and its implications and application to policy and practice. *School Psychology International*, 34(6), 575–612. doi:10.1177/0143034313479697
- Cole, J. C. M., Cornell, D. G., & Sheras, P. (2006). Identification of school bullies by survey methods. *Professional School Counseling*, 9(4), 305–313. doi:10.5330/prsc.9.4.wh4n8n4051215334
- Crick, N. R., Grotpeter, J. K., & Bigbee, M. A. (2002). Relationally and physically aggressive children's intent attributions and feelings of distress for relational and instrumental peer provocations. *Child Development*, 73(4), 1134–1142. doi:10.1111/1467-8624.00462 PMID:12146738
- Crick, N. R., Nelson, D. A., Morales, J. R., Cullerton-Sen, C., Casas, J. F., & Hickman, S. E. (2001). Relational victimization in childhood and adolescence: I hurt you through the grapevine. In J. Juvonen & S. Graham (Eds.), *Peer harassment in school: The plight of the vulnerable and victimized* (pp. 196–214). New York: The Guilford Press.
- Cross, D., Monks, H., Campbell, M., Spears, B., & Slee, P. (2011). *School-based strategies to address cyberbullying*. Occasional Papers #119. Centre for Strategic Education: Victoria.
- Dooley, J. J., Pyzalski, J., & Cross, D. (2009). Cyberbullying versus face-to-face bullying: A theoretical and conceptual review. *Zeitschrift fur Psychologie mit Zeitschrift fur Angewandte Psychologie*, 217(4), 182–188. doi:10.1027/0044-3409.217.4.182
- Dynel, M. (2008). No aggression, only teasing: The pragmatics of teasing and banter. *Lodz Papers in Pragmatics*, 4(2), 241–261. doi:10.2478/v10016-008-0001-7
- Fauman, M. A. (2008). Cyber-bullying: Bullying in the digital age. *The American Journal of Psychiatry*, 165(6), 780–781. doi:10.1176/appi.ajp.2008.08020226
- Greysen, S. R., Kind, T., & Chretien, K. C. (2010). Online professionalism and the mirror of social media. *Journal of General Internal Medicine*, 25(11), 1227–1229. doi:10.1007/s11606-010-1447-1 PMID:20632121
- Hawker, D. S. J., & Boulton, M. J. (2000). Twenty years research on peer victimization and psychosocial maladjustment: A meta-analytic review of cross-sectional studies. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 41(4), 441–455. doi:10.1111/1469-7610.00629 PMID:10836674
- Hinduja, S., & Patchin, J. W. (2008). Cyberbullying: An exploratory analysis of factors related to offending and victimization. *Deviant Behavior*, 29(2), 129–156. doi:10.1080/01639620701457816
- Hinduja, S., & Patchin, J. W. (2010). Bullying, cyberbullying, and suicide. *Archives of Suicide Research*, 14(3), 206–221. doi:10.1080/13811118.2010.494133 PMID:20658375
- Hughes, T. (1857). *Tom Brown's schooldays*. London: Charles H. Kelly.
- Jackson, C. (2011). Your students love social media... and so can you. *Teaching Tolerance*, 39, 38–41.
- Jones, L. M., Mitchell, K. J., & Finkelhor, D. (2012). Trends in youth Internet victimization: Findings from three youth internet safety surveys 2000-2010. *The Journal of Adolescent Health*, 50(2), 179–186. doi:10.1016/j.jadohealth.2011.09.015 PMID:22265114
- Juvonen, J., & Gross, E. F. (2008). Extending the school grounds? – Bullying experiences in cyberspace. *The Journal of School Health*, 78(9), 496–505. doi:10.1111/j.1746-1561.2008.00335.x PMID:18786042
- Koo, H. (2007). A time line of the evolution of school bullying in differing social contexts. *Asia Pacific Education Review*, 8(1), 107–116. doi:10.1007/BF03025837

- Kowalski, R. M. (2009). *Bullying*. Salem Health: Psychology and Mental Health. Pasadena, CA: Salem Press.
- Kowalski, R. M., Giumetti, G. W., Schroeder, A. N., & Lattanner, M. R. (2014). Bulling in the digital age: A critical review and meta-analysis of cyberbullying research among youth. *Psychological Bulletin*, 140(4), 1073–1137. doi:10.1037/a0035618 PMID:24512111
- Kowalski, R. M., & Witte, J. (2006). *Youth internet survey*. Unpublished manuscript, Clemson University.
- Li, Q. (2010). Cyberbullying in high schools: A study of students behaviours and beliefs about this new phenomenon. *Journal of Aggression, Maltreatment & Trauma*, 19(4), 372–392. doi:10.1080/10926771003788979
- Machmutow, K., Perren, S., Sticca, F., & Alsaker, F. D. (2012). Peer victimisation and depressive symptoms: Can specific coping strategies buffer the negative impact of cybervictimisation? *Emotional & Behavioural Difficulties*, 17(3-4), 403–420. doi:10.1080/13632752.2012.704310
- Mark, L., & Ratliffe, K. T. (2011). Cyber worlds: New playgrounds for bullying. *Computers in the Schools*, 28(2), 92–116. doi:10.1080/07380569.2011.575753
- Monks, C., & Smith, P. K. (2006). Definitions of bullying: Age differences in understanding of the term, and the role of experience. *The British Journal of Developmental Psychology*, 24(4), 801–821. doi:10.1348/026151005X82352
- Moreno, G. (2011). Cases of victimization case 2: Ryan Halligan (Vermont, 2003). *Preventing School Failure*, 55(2), 78. doi:10.1080/1045988X.2011.560495
- Nocentini, A., Calmaestra, J. J., Schultze-Krumholz, A., Scheithauer, H., Ortega, R., & Menesini, E. (2010). Cyberbullying: Labels, behaviours and definition in three European countries. *Australian Journal of Guidance & Counselling*, 20(02), 129–142. doi:10.1375/ajgc.20.2.129
- Office for National Statistics. (2013). *Statistical bulletin: Internet Access Quarterly Update*. Retrieved from: <http://www.ons.gov.uk/ons/rel/rdit2/internet-access-quarterly-update/q3-2013/stb-ia-q3-2013.html>
- O'Keeffe, G. W., & Clarke-Pearson, K. (2011). Clinical report- the impact of social media on children, adolescents, and families. *Paediatrics*, 127(4), 800–804. doi:10.1542/peds.2011-0054
- Olweus, D. (1978). *Aggression in the schools: Bullies and whipping boys*. Washington, DC: Hemisphere (Wiley).
- Olweus, D. (1993). *Bullying at school: What we know and what we can do*. Cambridge, MA: Blackwell Publishers, Inc.
- Olweus, D. (2001). *Olweus' core program against bullying and antisocial behavior: A teacher handbook*. Bergen, Norway: Author.
- Olweus, D. (2013). School bullying: Development and some important challenges. *Annual Review of Clinical Psychology*, 9(1), 751–780. doi:10.1146/annurev-clinpsy-050212-185516 PMID:23297789
- Ortega, R., Elipe, P., Moran-Merchán, J. A., Genta, M. L., Brighi, A., Guraini, A., & Tippet, A. et al. (2009). The emotional impact of bullying and cyberbullying on victims: A European cross-national study. *Aggressive Behavior*, 38(5), 342–356. doi:10.1002/ab.21440 PMID:22782434
- Oxford English Dictionary. (1989). *Bully* (2nd ed.). Oxford, UK: Clarendon Press.
- Palfrey, J., Gasser, U., & Boyd, D. (2010, February). *Empowering parents and protecting children in an evolving media landscape*. Response to FCC notice of inquiry 09-94. The Berkman Centre for Internet and Society at Harvard University. Retrieved from: http://dash.harvard.edu/bitstream/handle/1/3745922/Palfrey_Gasser_boyd_response_to_FCC_NOI_09-94_Feb2010.pdf?sequence=2

- Przybylski, A. K., Murayama, K., DeHaan, C. R., & Gladwell, V. (2013). Motivational, emotional, and behavioural correlates of fear of missing out. *Computers in Human Behavior*, 29(4), 1814–1848. doi:10.1016/j.chb.2013.02.014
- Rigby, K. (1997). What children tell us about bullying in schools. *Children Australia*, 22(02), 18–28. doi:10.1017/S1035077200008178
- Rigby, K. (2002). *A meta-evaluation of methods and approaches to reducing bullying in pre-schools and early primary school in Australia*. Canberra: Commonwealth Attorney Generals Department.
- Sahin, M. (2012). The relationship between cyberbullying/cybervictimisation and loneliness among adolescents. *Children and Youth Services Review*, 3(4), 834–837. doi:10.1016/j.childyouth.2012.01.010
- Seiffge-Krenke, I., & Klessinger, N. (2000). Long-term effects of avoidant coping on adolescents depressive symptoms. *Journal of Youth and Adolescence*, 29(6), 617–630. doi:10.1023/A:1026440304695
- Ševčíková, A., & Šmahel, D. (2009). Online harassment and cyberbullying in the Czech republic: Comparison across age groups.. *The Journal of Psychology*, 217, 227–229.
- Slonje, R., & Smith, P. K. (2008). Cyberbullying: Another main type of bullying? *Scandinavian Journal of Psychology*, 49(2), 147–154. doi:10.1111/j.1467-9450.2007.00611.x PMID:18352984
- Smith, P. K., Mahdavi, J., Carvalho, M., Fisher, S., Russell, S., & Tippet, N. (2008). Cyberbullying: Its nature and impact in secondary school pupils. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 49(4), 376–385. doi:10.1111/j.1469-7610.2007.01846.x PMID:18363945
- Sourander, A., Klomek, A. B., Ikonen, M., Lindroos, J., Luntamo, T., Koskelainen, M., & Helenius, H. et al. (2010). Psychosocial risk factors associated with cyberbullying among adolescents. *Archives of General Psychiatry*, 67(7), 720–728. doi:10.1001/archgenpsychiatry.2010.79 PMID:20603453
- Spears, B., Slee, P., Owens, L., & Johnson, B. (2009). Behind the scenes and screens: Insights into the human dimension of covert and cyberbullying. *The Journal of Psychology*, 217, 189–196.
- Spenser, K. A., & Betts, L. R. (2014). “People think it’s a harmless joke when really it could be hurting someone”: Young people’s experiences of cyber bullying. Poster presented at the British Psychological Society Annual Conference, Birmingham, UK.
- Steffgen, G., & König, A. (2009). Cyber bullying: The role of traditional bullying and empathy. In B. Sapeo, L. Haddon, E. Mante-Meijer, L. Fortunati, T. Turk, & E. Loos (Eds.), *The good, the bad and the challenging. Conference Proceedings (Vol. 2, pp. 1041–1047)*. Brussels, Belgium: Cost Office.
- Tokunaga, R. S. (2010). Following you home from school: A critical review and synthesis of this research on cyberbullying victimization. *Computers in Human Behavior*, 26(3), 277–287. doi:10.1016/j.chb.2009.11.014
- Tynes, B. M., Rose, C. A., & Williams, D. R. (2010). The development and validation of the online victimization scale for adolescents. *Cyberpsychology (Brno)*, 4, 1–15.
- Valkenburg, P. M., & Peter, J. (2007). Preadolescents and adolescents online communication and their closeness to friends. *Developmental Psychology*, 43(2), 267–277. doi:10.1037/0012-1649.43.2.267 PMID:17352538

Vandebosch, H., & Van Cleemput, K. (2008). Defining cyberbullying: A qualitative research into the perceptions of youngsters. *Cyberpsychology & Behavior*, 11(4), 499–503. doi:10.1089/cpb.2007.0042 PMID:18721100

Willard, N. (2007). The authority and responsibility of school officials in responding to cyberbullying. *The Journal of Adolescent Health*, 41(6), 64–65. doi:10.1016/j.jadohealth.2007.08.013 PMID:18047949

ADDITIONAL READING

Bauman, S., & Bellmore, A. (2015). New directions in cyberbullying research. *Journal of School Violence*, 14(1), 1–10. doi:10.1080/15388220.2014.968281

Berne, S., Frisé, A., Schultze-Krumbholz, A., Scheithauer, H., Naruskov, K., Luik, P., & Zukauskienė, R. et al. (2013). Cyberbullying assessment instruments: A systematic review. *Aggression and Violent Behavior*, 18(2), 320–334. doi:10.1016/j.avb.2012.11.022

Betts, L. R., & Houston, J. E. (2012). The effects of cyberbullying on children's school adjustment. In J. Jia (Ed.), *Educational stages and interactive learning: From kindergarten to workplace training* (pp. 209–230). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-0137-6.ch013

Chisholm, J. F. (2014). Review of the status of cyberbullying and cyberbullying prevention. *Journal of Information Systems Education*, 25, 77–87.

Garaigordobil, M. (2015). Cyberbullying in adolescents and youth in the Basque country: Prevalence of cyber victims, cyber aggressors, and cyber observers. *Journal of Youth Studies*, 18(5), 569–582. doi:10.1080/13676261.2014.992324

Olweus, D. (2012). Cyberbullying: An overrated phenomenon? *European Journal of Developmental Psychology*, 9(5), 520–538. doi:10.1080/17405629.2012.682358

Olweus, D. (2012). Comments on cyberbullying: A rejoinder. *European Journal of Developmental Psychology*, 9(5), 559–568. doi:10.1080/17405629.2012.705086

Schultze-Krumbholz, A., Göbel, K., Scheithauer, H., Brighi, A., Guarini, A., Tsorbatzoudis, H., & Smith, P. K. et al. (2015). A comparison of classification approaches for cyberbullying and traditional bullying using data from six European countries. *Journal of School Violence*, 14(1), 47–65. doi:10.1080/15388220.2014.961067

Smith, P. K. (2014). *Understanding school bullying: Its nature & prevention strategies*. London: Sage. doi:10.4135/9781473906853

Thomas, H. J., Connor, P., & Scott, J. G. (2015). Integrating traditional bullying and cyberbullying: Challenges of definition and measurement in adolescents – a review. *Educational Psychology Review*, 27(1), 135–152. doi:10.1007/s10648-014-9261-7

KEY TERMS AND DEFINITIONS

Cyberbullying: An aggressive act or behavior targeted towards an individual via digital technology.

Digital Footprint: The record of activity that an individual leaves when using technology (often without their knowledge).

Face-To-Face Bullying: An aggressive act or behavior targeted towards an individual in the same physical environment.

Fear of Missing Out: The belief that others are having 'rewarding experiences' that the individual is not aware of.

Perpetrator: The person who engages in an act of face-to-face bullying or cyberbullying.

Social Disinhibition: The reduction in awareness of social conventions and normative behaviors.

Target: The person who an act of face-to-face bullying or cyberbullying is directed towards.

Developing Creativity and Learning Design by Information and Communication Technology (ICT) in Developing Contexts

Chunfang Zhou

Aalborg University, Denmark

Aparna Purushothaman

Aalborg University, Denmark

INTRODUCTION

This chapter has two aims: (1) to establish the link between creativity, learning, information ecology and community of practice that underpins the theoretical necessity of a contextual user-centered approach to learning design for ICT in developing contexts; and (2) to specifically discuss how a Human Computer Interaction for Development (HCI4D), based on learning design, can be applied to provide the practical instrument for building creative learning environments in developing contexts. Accordingly, the chapter will develop a new theoretical framework by using three prominent theories: creativity theory, information ecology, and the theory of communities of practice. This chapter also offers practical contributions to developmental scholars and project managers in the form of a vocabulary to address the process and learning issues in both formal and informal learning environments, and by opening up new ways of understanding creativity, learning and usages of ICT in a developing context.

BACKGROUND

Generally, the term creativity means to develop new and useful ideas (Amabile, 1996). The development of different perspectives in describing creativity has been traced from the focus of the 1950s to the 1970s on areas of personality, cogni-

tion and the stimulation of creative individuals, to the awareness in the 1980s and 1990s of the influence of environments and social contexts on the creativity of individuals, groups and organizations (Loveless, 2007). In addition to discussion of the characteristics of novelty, effectiveness and ethicality (Cropley, 2001), the current creativity discourse also encompasses (1) operating in the economic and political field, (2) acting as a possible vehicle for individual empowerment in institutions and organizations, and (3) its use in developing effective learning (Jeffrey & Craft, 2001). Creativity is, therefore, now discussed as “a good thing”, promoting both personal expression and enhancing opportunities to engage in the complexities of problem-solving in the economic and cultural landscape of the 21st century (Loveless, 2007).

The uses of information communication technology (ICT) to support creativity by learning design have been described, reviewed and theorized in a range of published work in recent years (Loveless et al., 2006). ICT can be seen as a set of tools, which can be chosen as and when they are appropriate to creative processes. It can be argued that the characteristics of ICT can also make a distinctive contribution to those processes, providing new tools, media and environments for learning to be creative and for learning through being creative (Loveless, 2007). In particular, the use-centered approach to learning design by ICT can be viewed as a potential strategy to promote

DOI: 10.4018/978-1-5225-2255-3.ch362

learners' creative thinking skills and to improve learning abilities (Purushothaman, 2013). As explored by the domestication theory, the arrival of ICT in homes has brought with it the mobilization of material resources, skills, cultural values and social competences and capabilities. The recent rise of social media is also having an influential impact on organization innovation. These applications have shifted the way that users seek information and create and connect knowledge (Loveless, 2007). In developing contexts, designing learning for creativity by ICT should be paid more attention than in developed contexts due to the complex technical, economic, social and cultural problems.

DEVELOPING CREATIVITY AND LEARNING DESIGN IN DEVELOPING CONTEXTS

Designing Creativity-Based Learning Environment by ICT

In the 21st century, we are moving towards the “creative society”. In so doing, success in the future - for individuals, for communities, for companies, for nations as a whole — will be based not on what we know or how much we know, but on our ability to think and act creatively. Meanwhile, the rapid development of technology, mainly as a result of the internet, such as ICT, has brought about an upsurge of technological tools which young people are appropriating in their everyday lives (Purushothaman & Zhou, 2014).

Wheeler, Waite, and Bromfield (2002) propose a model for the creative use of ICT in learning which includes dimensions of problem-solving, creative cognition and social interaction. The model shows that the three dimensions are independent but interactive and, in some cases, it may be difficult to distinguish between them. This means that creativity development would need to incorporate social interaction within the community of activity. Furthermore, according to

Loveless et al., (2006), the framework for creativity and ICT attempts to describe the interaction between three elements of creative practices with ICT: creative processes (for example, using imagination, fashioning, pursuing purpose and evaluating originality and value), the features of ICT (for example, provisionality, interactivity, capacity, range, speed, automatic functions and multimodality), and ICT capability as an expression of elements of higher order thinking-finding things out, developing ideas and making things happen, exchanging and sharing information and reviewing, modifying and evaluating work as it progresses through a breadth of study. Thus, it is important to note that it is not the access to digital resources, which “delivers” creativity, but the opportunities such access affords for interaction, participation, and the active demonstration of imagination, production, purpose, originality and value (Loveless, 2007).

Needs of Learning Design by ICT in Developing Contexts

Designing for learning to use a technology in a developing world scenario is often constrained because of complexity and the impact of socio-cultural factors that influence ICT usage. These factors include patriarchal gender roles, illiteracy, cultural beliefs and values. This chapter takes the position that in order for the successful acceptance and usage of technology in a developing context, greater focus has to be placed on the design elements of the technological usage in the context where it will be used. This is very important because, in most cases, the design dimensions that are successful in the western world are adapted without sufficient consideration of the social and local contextual factors in a developing scenario where they are then introduced and used. As development is a social phenomenon and ICT is a technical phenomenon, ICT design framed for developmental goals should analyze the intervention strategies based on socio-technical terms (Dearden & Rizvi, 2009). The technological

design elements of developed countries may not fit into the culture of developing countries (Dray, Siegel, & Kotzé, 2003), thus, more effort should be made in developing contexts.

THEORETICAL UNDERPINNINGS

A Sociocultural Perspective to Learning and Creativity

This chapter takes a socio-cultural perspective on learning and creativity. Socio-cultural approaches to the study of learning represent a move away from the decontextualized, individualistic approach of certain types of cognitive psychology towards more collective, social and participatory types of framing (Haggis, 2008). In this regard, we can say that learning is always learning-in-context; a further element is learners' engagement, through which learning happens. Furthermore, in the social learning system, new patterns can be acquired through direct experience or by observing the behaviors of others. The more rudimentary form of learning, rooted in direct experience, is largely governed by the rewarding and punishing consequences that follow any given action. People are repeatedly confronted with situations with which they must deal in one way or another. In other words, during the course of learning, people not only perform responses, they also observe the differential consequences that accompany their various actions. On the basis of this informative feedback they develop thoughts or hypotheses about the types of behavior most likely to succeed. These hypotheses then serve as guides for future actions (Bandura, 1971).

As noted, creativity can be regarded as a driver for learning process. Socio-cultural theory posits that creativity can be described as an interaction between characteristics in people and communities, creative processes, subject domains and the wider social and cultural context (Loveless et al., 2006). Csikszentmihalyi (1996) proposes that creativity arises from the interaction between the

“intelligence” of individuals, the domain or areas of human endeavor, disciplines, crafts or pursuits and the field, such as people, institutions, reward mechanisms and “knowledgeable others” through which the judgements of individual performances in society are made. The levels of achievement of originality for individuals, peer groups or within the domain, are evaluated within the field, whereas the judgement of value can relate to the individual's critical reflection as well as recognition of a unique contribution to the domain itself. Creativity can, therefore, be seen in the interaction between a person's thoughts and actions both as an individual and in communities, their knowledge and skills within a domain, and within mediating cultural tools, and a socio-cultural context which can encourage, evaluate and reward. This has important implications for thinking about creativity where the learning environment might either nurture or dismiss the development of creative individuals, groups and communities (Loveless et al., 2006).

Information Ecology and Communities of Practice (CoP): Theoretical Roots for Designing for Learning by ICT

Information ecology is a “system of people, practices, values and technologies in a particular local environment” (Nardi & O'Day, 1999, p. 49). “An information ecology deals with the local differences while still capturing the interrelationships among the social, economic and political context in which the technology is invented and used” (Nardi & O'Day, 1999, p. 47). According to this theory there is an element of uncertainty that the adoption of new technologies brings to the quality of life and work. It asserts that “people often distance themselves from a critical evaluation of technologies in their lives, as if these technologies were inevitable forces of nature rather than things we design and choose” (Nardi & O'Day, 1999, p. 14). The theoretical underpinning of “value and beliefs” can help with the understanding of how the intended value of the technology and the

beliefs of the users impact on the specific ICT. The “why” aspect of the theory will help us to understand why the specific ICT is needed for the target group.

To address the socio-cultural aspect of learning to use the technology in a developing context, the design uses Wenger’s (1998) social theory of Communities of Practice (CoP). CoP has its roots in the situated theory of learning, which is an alternative approach to the dominant cognitive perspectives on learning. “Communities of practice are formed by people who engage in a process of collective learning in a shared domain of human endeavor: a tribe learning to survive, a band of artists seeking new forms of expression, a group of engineers working on similar problems, a clique of pupils defining their identity in the school, a network of surgeons exploring novel techniques, a gathering of first-time managers helping each other cope ” (Wenger, 2006, p. 1). In CoP, learning is seen as a form of social participation. “Participation here refers not just to local events of engagement in certain activities with certain people, but to a more encompassing process of being active participants in the practices of social communities and constructing identities in relation to these communities” (Wenger, 1998, p. 4). Thus, learning is not just the accumulation of skills and information, but also a process of becoming.

A LEARNING DESIGN BY ICT: HUMAN COMPUTER INTERACTION FOR DEVELOPMENT (HCI4D)

What is HCI4D?

Human Computer Interaction for Development (HCI4D) focuses on the “relationship between humans and technology in the context of international development, ranging from lower-level interface design issues to higher-level social interactions” (Anokwa et al., 2009). The socio-cultural aspects

that influence the design and usage of technologies in developing countries are at the core of HCI4D (Winters & Toyama, 2009). This addresses questions of “how interactive products, applications, and systems can be appropriately designed to both address the distinctive needs of users in developing regions, and to cope with the difficult infrastructural contexts where these technologies must work” (Ho, Smyth, Kam, & Dearden, 2009, p. 1). Thus, fundamentally, HCI4D deals with designing ICT tools and technology that can be used by marginalized people to better their lives.

This chapter proposes a design that is driven by the social learning aspect of the efficient use of technology, an aspect that is rarely addressed in the HCI4D literature. Most of the HCI4D designs place emphasis on the product outcomes. There is a desperate need from the scholars that the digital divide can be discussed by addressing the issues that are central to HCI4D, such as how to improve the fit between technology, human needs and human context and how to design technology to facilitate human interaction with the technology (Dray et al., 2003). The focus on learning is mainly driven by the fact that “in the learning approach, outcomes come through the ongoing process of using ICT’s rather than from the product of ICT’s and process can be explicitly designed or shaped through understanding informal learning which links between actions and use and informal learning outcomes” (Foster, 2011, p. 2).

By taking HCI4D as an example, we propose a design for learning to use ICT that has design components that the scholars and designers can consider for the efficient acceptance and usage of ICT in a developing context. The design takes contextual factors as its backbone. The design also takes into consideration the theoretical underpinning of information ecologies, as proposed by Nardi & O’Day (1999), which can assist designers of ICT in a developing context to understand how people value the specific ICT, what perceptions they hold about ICT, and why the specific ICT is going to be useful for the intended users in a

developing context. A further design dimension that is an important aspect when seen from a digital divide perspective, is that the design should provide opportunities for identity work and should focus on the social and cultural practices of the context in which learning takes place. The design takes into consideration the socio-cultural learning theory developed by Wenger (1998) which will enable designers and scholars to design according to how the participants experience ICT usage and how they understand and perceive themselves as ICT users. This can be achieved by providing them with the scope to construct identities as ICT users which can, in turn, make learning to use ICT more effective for the participants. The learning design also incorporates appropriate training programs for the users of technology for the long-term sustainability of developmental efforts.

Contextual Factors

The first and most significant element when designing learning for any form of ICT in an HCI4D context is a thorough understanding of the contextual factors in which the ICT is to be used. “Sociocultural factors influence the perceived benefits of acquiring technology-related skills and availability of skills required to use a technology influences perceived ease of use and perceived usefulness of the technology and hence attitude towards its adoption” (Dholakia, Dholakia, & Kshetri, 2003, p. 14). Scholars note the significance of knowledge about the contextual design factors at the local level, such as sociocultural factors, technological factors, economic factors and users’ past ICT experiences, all of which are relevant for designing ICT in a developmental scenario (Parmar, 2009; Walton & Heeks, 2011; Winters & Toyama, 2009). We would emphasize that the specific background information of the intended users and the socio-cultural environment of the context where the ICT will be used should be thoroughly investigated.

Value and Beliefs About the ICT

How the intended users value the specific ICT that is introduced and what their beliefs are about the specific ICT are highly significant for acceptance and successful usage. Dearden & Rizvi (2009) state that whilst designing for ICT4D it is also important to examine what ICT might mean for the users in the context. “Values represent a manifestation of culture that signify espoused beliefs identifying what is important to a particular cultural group” (Leidner & Kayworth, 2006, p. 359). Studies show that cultural perceptions of technology influence the acceptance and future usage behavior of the users (Chen, Mashhadi, Ang, & Harkrider, 1999; Leidner & Kayworth, 2006). However, it is noted that there is a lack of attention to cultural beliefs and their impact on ICT adoption in developing countries (Albirini, 2006; Loch, Straub, & Kamel, 2003). Poor understanding of ideas, beliefs and values about how technology should be utilized in developing countries challenges the designers of technically advanced societies for successful implementation and usage of technologies (Straub, Loch, & Hill, 2001).

Values and beliefs about technology are the outcome of the socio-cultural structures that people belong to. Thus, cultural perceptions about the specific technology will significantly influence how they value and use the technology. People’s perceptions of ICT differ from culture to culture and the way in which they value ICT usage will also differ. If the intended users of technology are women, sufficient attention has to be given to the gender specific usage of the technology, how women value it, and what beliefs they have about the technology. Gendered roles are determined by cultural, social and economic factors and they differ within cultures and between countries (Moghaddam, 2010). These gendered roles, defined by cultural influences, impact on the usage of technology by women (Johnson, 2010; Primo, 2003; Purushothaman, 2013).

The concept of information ecologies developed by Nardi & O'Day (1999) can, in a developing context, provide insights for designers of ICT about how people value ICT and what perceptions they hold about the specific ICT that is introduced; thus, learning can bring effective results. The values and beliefs that participants hold about ICT have to be taken into consideration before the project begins and before they start to learn how to use the ICT. This is because if participants have negative attitudes and do not assign much value to ICT, before the training starts and they start to learn how to use the ICT, they have to be given a greater awareness of the potential benefits of the specific ICT and the positive changes it can bring to their lives. If participants do not believe in the technology and do not believe that using ICT can bring any value to their lives, making them learn how to use the technology and facilitating training and the scope for interaction and mutual engagement, and providing scope for constructing identities as ICT users (which is another design dimension addressed by this paper) may not be effective unless their negative beliefs about using the technology are changed.

Why the Specific Technology

A further important element that should be considered in the design is how the specific technology is going to add usefulness to the lives of the intended users. Nardi & O'Day (1999) state that the "technology that is in question should be evaluated to know what significance it holds and what it means for the users in settings" (p.69). There is usually a tendency to only consider the "how" question that deals with the practicality of technology, whereas "the 'know-how' questions do inform the design but it is equally important to ask the 'know-why' questions so that technology can be used productively" (Nardi & O'Day, 1999, p. 71). In an ICT4D context, consideration should be given to the possible ways in which they are going to use the specific ICT and how it can bring significance to their lives. This "why" question

can provide insight into the motivation for using a particular technology/tool, why the particular technology is best for the task at hand, why it fits well with the practices of the participants, and also, possibly, why it might not be fit for purpose.

Constructing Identity

The learning design should also provide the scope for constructing identities as ICT users. Since the design intervention in ICT4D is always situated (Dearden & Rizvi, 2009), and in CoP learning is viewed as negotiating new identities that are both situated and dynamic, a CoP-based learning framework can bring new ways of understanding learning and using any form of technological tool for development that is context situated and specific. Learning to use technology can be more effective when learning is conceptualized as a process of identity change rather than a process of gaining and transferring information through technology (Brosnan & Burgess, 2003; Mayes & Fowler, 2006). CoP can help in understanding how the users negotiate meaning, how they interpret the experiences of ICT usage and how they understand and see themselves as ICT users in this specific socio-cultural context.

Reducing Digital Divide and Training Programs

In developing ICT-based creative learning designs in developing contexts, challenges exist both in technology and in the social sciences. For example, reducing the digital divide requires improvements across all the dimensions of ICT (dubbed the 4C framework): computing, connectivity, content, and human capacity (Wang & Zhou, 2013):

- **Computing:** PCs are prohibitively expensive for most people, and shared access (e.g., community centers or cybercafés) becomes inevitable. PCs today are very difficult to use and even "experts" spend a lot of time maintaining their machines,

worrying about updates, security, compatibility of hardware, etc. As a complementary technology, non-PC devices are an important option, e.g., mobile phones.

- **Connectivity:** Whilst mobile technology is improving worldwide, it remains expensive, limited in rural areas, and poor at proving data connectivity.
- **Content:** Meaningful content is lacking in many languages and most content is not locally relevant. Today's systems tend to make people passive consumers of information instead of enabling the generation of local information. In addition, rich content demands multimedia (useful to overcome literacy issues), which, in turn, requires broadband connectivity.
- **(Human) Capacity:** Users need to be aware, literate, and innovative in order to harness the power of ICT. They should also be empowered to use ICT, both by society and by the state.

Meanwhile, training to use the specific ICT is very important for successful usage. Appropriate training programs should, thus, be an integral part of the learning design. The program should also be helpful for dealing with diverse issues in relation to creativity and learning in developing contexts. For example, in some areas women are not safe when they are alone and in public places at night (Wangmo, Violina, & Haque, 2004). This restricted mobility is a barrier to women's use of ICT in developing countries (Purushothaman, 2013). Therefore, if the intended users are women, the training programs need to be gender sensitive (Purushothaman, 2013).

FUTURE RESEARCH DIRECTIONS

As mentioned earlier, ICT usage does not occur in a vacuum but, rather, it occurs within social and cultural norms that also shape the digital divide between developed and developing con-

texts. ICT usage is based on policy and business models, especially regulation. Thus, in the long term, research on ICT must provide value and be sustainable from both a user and a provider perspective. Designers and developers of ICT should consider innovative ways to work with and develop understandings of the nature of digital technologies as a medium in creative process (Tongia, Subrahmanian, & Arunachalam, 2005). A creativity framework should underpin planning, practice and evaluation. The framework should recognize the interaction between individuals and communities, processes, domains and fields, and the characteristics of imagination, fashioning, pursuing purpose, originality and value judgments (Loveless, 2007). These elements stand as the core philosophy of user-centered approaches to learning design, for example, HCI4D, which offer great potential for bringing benefits to future development.

CONCLUSION

In developing contexts, learners should have opportunities to develop ICT skills and techniques in an authentic and challenging creative society. The information ecology approach, recommended in learning design, can bring new dimensions to the field of ICT4D by connecting the needs of participants in terms of skills for using ICT and in designing an intervention to match needs with the value that intended participants give to the ICT in question and what beliefs they have about the technology that is going to be introduced. CoP theory can offer developmental scholars and project managers a vocabulary for addressing the process and learning issues in informal learning environments and can open up new ways for them to understand how learning to use ICT based tools in a developmental context may occur. In short, this chapter offers both theoretical and practical contributions to areas such as creativity learning and to the future improvement of developmental efforts.

REFERENCES

- Albirini, A. (2006). Cultural perceptions: The missing element in the implementation of ICT in developing countries. *International Journal of Education and Development Using Information and Communication Technology*, 2(2), 49–65.
- Amabile, T. M. (1996). *Creativity in Context*. Boulder, CO: Westview Press.
- Anokwa, Y., Smyth, T. N., Ramachandran, D., Sherwani, J., Schwartzman, Y., Luk, R., & DeRenzi, B. (2009). Stories from the Field: Reflections on HCI4D Experiences. *Information Technologies & International Development*, 5(4), 101–116.
- Bandura, A. (1971). *Social Learning Theory*. New York: General Learning Press.
- Brosnan, K., & Burgess, R. C. (2003). Web based continuing professional development – a learning architecture approach. *Journal of Workplace Learning*, 15(1), 24–33. doi:10.1108/13665620310458794
- Chen, A. Y., Mashhadi, A., Ang, D., & Harkrider, N. (1999). Cultural Issues in the Design of Technology-Enhanced Learning Systems. *British Journal of Educational Technology*, 30(3), 217–230. doi:10.1111/1467-8535.00111
- Cropley, A. J. (2001). *Creativity in Education and Learning: A Guide for Teachers and Educators*. London: Kogan Page Limited.
- Csikszentmihalyi, M. (1996). *Creativity: Flow and the Psychology of Discovery and Invention*. New York: HarperCollins.
- Dearden, A., & Rizvi, H. (2009). *A deeply embedded sociotechnical strategy for designing ICT for development*. Retrieved February 14, 2013, from <http://shura.shu.ac.uk/2936/>
- Dholakia, R. R., Dholakia, N., & Kshetri, N. (2003). Gender & Internet Usage. In *The Internet Encyclopedia*. Hoboken, NJ: John Wiley & Sons.
- Dimaggio, P., Hargittai, E., Celeste, C., & Shafer, S. (2004). From Unequal Access to Differentiated Use: A Literature Review and Agenda for Research on Digital Inequality. *Social Inequality*, 355–400.
- Dray, S. M., Siegel, D. A., & Kotzé, P. (2003). Indras Net: HCI in the developing world. *Interaction*, 10(2), 28–37. doi:10.1145/637848.637860
- Foster, C. (2011). ICTs and informal learning in developing countries. Manchester, UK: University of Manchester, Institute for Development Policy and Management. Retrieved from http://www.sed.manchester.ac.uk/idpm/research/publications/wp/di/documents/di_wp46.pdf
- Hafkin, N. (2002). *Gender Issues in ICT Policy in Developing Countries: An Overview*. Seoul, Republic of Korea: Division for the Advancement of Women (DAW) - United Nations. Retrieved from www.un.org/womenwatch/daw/egm/ict2002/.../Paper-NHafkin.PDF
- Haggis, G. (2008). Beyond ‘mutual constitution’: looking at learning and context from the perspective of complexity theory. In R. Edwards, G. Biesta, & M. Thorpe (Eds.), *Rethinking Contexts for Learning and Teaching, Communities, Activities and Networks* (pp. 44–61). New York: Routledge.
- Ho, M. R., Smyth, T. N., Kam, M., & Dearden, A. (2009). Human-Computer Interaction for Development: The Past, Present, and Future. *Information Technologies & International Development*, 5(4), 1–18.
- Huyer, S., & Carr, M. (2002). Information and Communication Technologies: A Priority for Women. *Gender, Technology and Development*, 6(1), 85–100. doi:10.1177/097185240200600105
- Jeffrey, B., & Craft, A. (2001). The universalization of creativity. In A. Craft, B. Jeffrey, & M. Leibling (Eds.), *Creativity in Education*. London: Continuum.

- Johnson, V. (2010). Women and the Internet A Micro Study in Chennai, India. *Indian Journal of Gender Studies*, 17(1), 151–163. doi:10.1177/097152150901700107
- Leidner, D. E., & Kayworth, T. (2006). A Review of Culture in Information Systems Research: Toward a Theory of Information Technology Culture Conflict. *Management Information Systems Quarterly*, 30(2), 357–399.
- Loch, K. D., Straub, D. W., & Kamel, S. (2003). Diffusing the Internet in the Arab World: The Role of Social Norms and Technological Culturation. *IEEE Transactions on Engineering Management*, 50(1), 45–63. doi:10.1109/TEM.2002.808257
- Loveless, A. M. (2007). *Creativity, technology and learning – a review of recent literature* (No. 4 update). Retrieved May 2009, from www.futurlab.org.dk/litereviews
- Loveness, A., Burton, J., & Turvey, K. (2006). Developing conceptual frameworks for creativity, ICT and teacher education. *Thinking Skills and Creativity*, 1(1), 3–13. doi:10.1016/j.tsc.2005.07.001
- Mayes, T., & Fowler, C. (2006). Learners, Learning Literacy and the Pedagogy of e-Learning. In *Digital literacies for learning*. London: Facet.
- Moghaddam, G. G. (2010). Information technology and gender gap: Toward a global view. *The Electronic Library*, 28(5), 722–733. doi:10.1108/02640471011081997
- Nardi, B. A., & O'Day, V. L. (1999). *Information ecologies : using technology with heart*. Cambridge, Mass.: MIT Press.
- Parmar, V. (2009). A Multidisciplinary Approach to ICT Development. *Information Technologies and International Development*, 5(4), 89–96.
- Primo, N. (2003). *Gender Issues in the Information Society*. UNESCO. Retrieved from http://portal.unesco.org/ci/en/file_download.php/250561f24133814c18284feedc30bb5egender_issues.pdf
- Purushothaman, A. (2013). Empowering Women Through Learning To Use The Internet - An Ethnographic Action Research Project To Address The Second Order Digital Divide. Aalborg University. Retrieved from http://vbn.aau.dk/files/195244502/Aparna_Thesis_VBN_Upload.pdf
- Purushothaman, A., & Zhou, C. (2014). Change toward a Creative Society in Developing Contexts: Womens Barriers to Learning by Information and Communication Technology. *Gender, Technology and Development*, 18(3), 363–386. doi:10.1177/0971852414544008
- Straub, D., Loch, K. D., & Hill, C. E. (2001). Transfer of Information Technology to the Arab World. *Journal of Global Information Management*, 9(4), 6–28. doi:10.4018/jgim.2001100101
- Tongia, R., Subrahmanian, E., & Arunachalam, V. (2005). *Information and Communications Technology for Sustainable Development: Defining a Global Research Agenda*. Bangalore, India: Allied Publishers.
- Walton, M., & Heeks, R. (2011). *Can a Process Approach Improve ICT4D Project Success?* Development Informatics Group Institute for Development Policy and Management. Retrieved from <http://www.sed.manchester.ac.uk/idpm/research/publications/wp/di/index.htm>
- Wang, F., & Zhou, C. (2013). A Theoretical Study on Development of Information and Communication Technology (ICT)-Supported Education Systems. In J. Chen, X. Wang, L. Wang, J. Sun, & X. Meng (Eds.), *The Proceedings of 2013 10th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)* (Vol. 1, pp. 1080–1084). Red Hook, NY: IEEE Press.
- Wangmo, S., Violina, S., & Haque, M. (2004). *Trend and Status of Gender Perspectives in ICT Sector: Case Studies in Asia-Pacific Countries*. ITU. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.118.7888&rep=rep1&type=pdf>

Wenger, E. (1998). *Communities of practice : learning, meaning, and identity*. New York: Cambridge University Press. doi:10.1017/CBO9780511803932

Wenger, E. (2006). *Communities of practice - a brief introduction*. Retrieved from <http://www.ewenger.com/theory/>

Wheeler, S., Waite, S. J., & Bromfield, C. (2002). Promoting creative thinking through the use of ICT. *Journal of Computer Assisted Learning*, 18(3), 367–378. doi:10.1046/j.0266-4909.2002.00247.x

Winters, N., & Toyama, K. (2009). Human-Computer Interaction for Development: Mapping the Terrain. *Information Technologies & International Development*, 5(4), iii–viii.

ADDITIONAL READING

Awang, H., & Ramly, I. (2008). Creative thinking skill approach through Problem-Based Learning: Pedagogy and practice in the engineering classroom. *International Journal of Social Sciences*, 3(1), 18–23.

De Graaff, E., & Ravesteijn, W. (2001). Training complete engineers: Global enterprise and engineering education. *European Journal of Engineering Education*, 26(4), 419–427. doi:10.1080/03043790110068701

Delic, K. A., & Dum, R. (2006). On the emerging future of complexity sciences, *ACM Ubiquity*, 7.

Ekvall, G. (1999). Creative climate. In M. A. Runco & S. R. Pritzker (Eds.), *Encyclopedia of Creativity* (pp. 403–412). San Diego: Academic Press.

Eteläpelto, A., & Lahti, J. (2008). The resources and obstacles of creative collaboration in a long-term learning community. *Thinking Skills and Creativity*, 3(3), 226–240. doi:10.1016/j.tsc.2008.09.003

Hoegl, M., & Parboteeah, P. (2007). Creativity in innovative projects: How teamwork matters. *Journal of Engineering and Technology Management*, 24(1-2), 148–166. doi:10.1016/j.jengtecman.2007.01.008

Hughes, T. (2004). *Human-Built world: How to think about technology and culture*. Chicago: University of Chicago Press.

Illeris, K. (2002). *The three dimensions of learning*. Denmark: Roskilde University Press.

Kim, K. H. (2007). Exploring the interactions between Asian culture (Confucianism) and creativity. *The Journal of Creative Behavior*, 41(1), 28–53. doi:10.1002/j.2162-6057.2007.tb01280.x

Lubart, T. I. (1999). Creativity across cultures. In R. J. Sternberg (eds.), *Handbook of Creativity*. 339–350. New York: Cambridge University Press.

Mathisen, G. E., & Einarsen, S. (2009). A review of instruments assessing creative and innovative environments within organizations. *Creativity Research*, 16, 1–5.

Ng, A. K. (2001). *Why Asians are Less Creative than Westerners*. Singapore: Prentice-Hall.

Paulus, P. B. (2003). *Group creativity: Innovation through collaboration*. USA: Oxford University Press. doi:10.1093/acprof:oso/9780195147308.001.0001

Pool, R. (1997). *Beyond engineering: How society shapes technology*. New York: Oxford University Press.

Sawyer, R. K. (2003). Emergence in creativity and development. In R. K. Sawyer (Ed.), *Creativity and Development* (pp. 4–12). USA: Oxford University Press.

Shrader-Frechette, K., & Westra, L. (1997). *Technology and values*. Lanham: Rowman & Littlefield Publishers, Inc.

Sparrow, P., & Gaston, K. (1996). Generic climate maps: A strategic application of climate survey data? *Journal of Organizational Behavior*, 17(6), 679–698. doi:10.1002/(SICI)1099-1379(199611)17:6<679::AID-JOB786>3.0.CO;2-M

Verbeek, P. P. (2011). *Moralizing technology: Understanding and designing the morality of things*. Chicago: The University of Chicago Press. doi:10.7208/chicago/9780226852904.001.0001

Westbury, I., Hopmann, S., & Riquarts, K. (2000). *Teaching as a reflective practice: The german didaktik tradition*. Mahwah: Lawrence Erlbaum Associates, Publishers.

KEY TERMS AND DEFINITIONS

Creativity: Etymologically speaking, the term “creativity” means to generate new and useful ideas. The field of creativity was practically started from psychological studies. Today the field has seen an explosion of interest: creativity has been discussed much by the theories such as psychology, social psychology, cultural psychology, social culture and even philosophy.

Learning: Learning is the act of gaining new knowledge, behaviors, skills, or ability. It may be regarded as a process, rather than a collection of factual and procedural knowledge. Human learning may occur as part of education, professional development, or training.

Learning Design: It means to design a learning environment that is based on theories of learning and aims to facilitate learning process and improve learning quality.

Information Communication Technology (ICT): ICT are a set of tools those include computer, Internet, mobile phones and diverse communication devices that contributes to education system and social learning system.

Information Ecology: It is a theory that deals with the local differences while still capturing the interrelationships among the social, economic and political context in which the information technology is invented and used.

Human Computer Interaction for Development (HCI4D): It is a learning model that focuses on the relationship between humans and information technology in the context of international development, ranging from lower-level interface design issues to higher-level social interactions.

Communities of Practice: It is a theory rooted in the situated theory of learning that are formed by people who engage in a process of collective learning in a shared domain of human endeavor. In a community of practice, learning is seen as a form of social participation. Participation refers not just to local events of engagement in certain activities with certain people, but to a more encompassing process of being active participants in the practices of social communities and constructing identities in relation to the communities.

Existential Aspects of the Development E-Culture



Liudmila Vladimirovna Baeva
Astrakhan State University, Russia

INTRODUCTION

The e-culture first was mentioned at the end of the 1990s. According to the European tradition, e-culture was originally understood as a form of cultural heritage preservation and the opposition to the e-commerce. Later, the term was used for the notion of different objects having an electronic or other digital form. Nowadays, the “e-culture” is an interdisciplinary concept having connotations in Philosophy, Cultural Studies, Sociology, Political Science, Economics and of course in the field of information technologies.

In general, *E-culture* represents cumulative results of creative activity and communication of people under the conditions of the information technology implementation, which characterized with creating of free information space, a virtual form of expression, distant technology and content liberality.

Advanced high technologies represent a vector of the civilization development, mediating economic and communicative processes of the world formation. Under these conditions, the crisis of socio-cultural and individual identity escalates, “technogenic” values extrapolate on the sphere of the interpersonal interaction. The ideologeme of technocentrism devaluates the traditional axiological models, informative and technogenic directivity of social and anthropological dynamics and determines new types of interaction both on structural and spiritual levels of social being. Under the domination of new technogenic institutes, the dysfunction of cultural and spiritual regulators of the social development occurs, in particular, the traditional ethical and epistemologi-

cal mechanisms of the public conscience shifting to massive, irrational and virtual sphere are broken. The sphere of the social discourse, mediated by electronic mass media, separates from subjects, making human beings feel the quasi reality being, solitude and desolation in the “boundless” information space.

The peculiarity of the present day age became the situation of the uncontrolled technological development, that is unique in its own way and requires detailed study, as its consequences have already caused significant deformation of psychology, world view, values and society. One of the most striking phenomena of the digital age becomes the formation of new digital culture (e-culture). Although, the concept “e-culture” has been still developing, it is evident it cannot be compared as a phenomenon with anything that has ever existed what requires the necessity of its detailed study.

Methodological basis of the research is related to the set of categories and the principles of existential philosophy and philosophical anthropology. The theories by M.Heidegger, V.Frankl and N.Abbagnano, A. Meneghetti have influenced on the formation of the conception of existential axiology in different ways. The research of the structure and the essence of values by R. Hartman, R. Frondizi, S.O. Hansson, symbolic and logical expression of a value by G.Vernon, the correlation of meaning and significance of a value structure, the search for subject and object of values by K.Baier, the ethical content of values by R.Brumbaugh, vital and existential analysis of values - A. Maslow, Ph. Foot, E. Levinas, D. Vokey, the analysis of the priorities of postindus-

trial and information epochs by Y. Masuda, A. Giddens, M. Castells, B.J. Kallenberg, A. Toffler, P. Drucker, M. McLuhan etc. played an important role in our research.

The most significant for understanding e-culture and existential and axiological aspects of development research by A.M. Ronchi, R. Capurro, D. R. Raban, L. Floridi, etc. deal with the study of ethical and anthropological issues of the information space. These studies influenced greatly on the author's research representing the theoretical basis for the further development of the ideas. The study of human issues, communications, cultural values, social norms, threats to an information culture are made by such scientists as Baarda, R., Rocci, L., Zhou, L., Ding, L., Finin, T., Sartor, G., Ott, M., Pozzi, F., Cockton, G. etc.

A feature of my approach to the study of electronic culture is existential-axiological research methodology. This approach allows to identify the values and intrapersonal problems of human existence, which are formed under the impact of e-culture. E-culture is considered in the framework of this approach expanded as a sphere, created with the help of digital technology.

Information technologies has greatly influenced on the development of the specific culture – electronic, digital or virtual one – within recent decades. It generates technologically and qualitatively new phenomena, involving more and more spheres, such as science, art, social interaction, education, mass media, commerce and political system. E-culture cannot be assigned to material or spiritual culture as it has features of both of them. After the creation of “the second nature” – “the world of things” a human being in actual fact created “the third nature” – the world of virtual phenomena, that is a specific synthesis of the conciseness world and advanced information technologies. The study of e-culture became rapidly important for the science and the practice as the development of new possibilities; and at the same time certain threats for existing forms and expression ways of culture appeared. At the same time e-culture also generated new values which

became higher priority for human beings nowadays and determined human world view objectives.

The analysis of the specific character of e-culture should be started with the definition of this concept, as it is polysemantic and requires its content specification. E-culture or digital culture is first of all a new sphere of the human activity, associated with the creation of the electronic copies of spiritual and material objects as well as the creative work of the virtual objects of science, communication and art (Ronchi 2009). “Electronic” means the representation in a digital form. E-culture first was mentioned at the end of the 1990s. According to the European tradition, e-culture was originally understood as a form of cultural heritage preservation (Ronchi 2009) and also as some opposition to e-commerce. Later, the term was used for the notion of different objects having electronic or other digital form. Nowadays, “e-culture” is an interdisciplinary concept having connotations in Philosophy, Cultural Studies, Sociology, Political Science, Economics and of course in the field of information technologies. Its subjects and creators are scientists, programmers, artists, representatives of mass media and average users of information systems, creating electronic forms of self-representation and self-manifestation in the global network Internet by the means of technological facilities. The most important characteristics of e-culture is transparency, globality and availability for every user; besides every subject can become both an user and a creator of its phenomena, being enough free in creative work and not having strict limitations.

BACKGROUND

E-culture comprises the following phenomena: first of all electronic forms of modern communication (Internet, cellular communication and smartphone applications, social networking sites, virtual communities, chats, blogs, web-forums, and web-sites), electronic cultural heritage (on-line museums, galleries and exhibitions), on-line

education, electronic reconstruction of cities and objects of cultural heritage in their historical and space perspective, computer games including network games, electronic mass media (on-line magazines and newspapers), digital modern art (animation, photo, cinema, music and advertizing, created by means of advanced information technologies), electronic reference systems (archives, encyclopedias, dictionaries and libraries), and information programs (security, forms of data security, etc.).

The distinctive features of e-culture are a digital form, virtuality, freedom of access, openness for the members of “information community” (those who have electronic resources), remotability, real distance from a subject, activity in accessing electronic information, possibility to participate in developing of the information content from any point of “information space”, liberality, linguistic description, absence of strict rules and norms, dominance of visual things over conceptual ones, innovativeness, technocracy, and changes of a high speed.

There are two main structural types of e-culture expression: 1. electronic form of former (traditional, classical – here not electronic) cultural objects (e.g., on-line museums, libraries, exhibitions, etc.) 2. electronic cultural objects in form and essence (computer games, social networking sites, Internet, digital art pieces, etc.)

Each of the above-mentioned types has its own characteristics, although they are formally similar and intercross. Culture, which has electronic form and essence, is the continuation and the development of the former one, having become an electronic one in its expression nowadays (to maintain its value and competitiveness).

The real “live” culture is changing in the part of traditional form to compete successfully with new electronic one. E-culture significantly turned out to be the transformation of real artifacts, ideas and creatures of human beings into virtual forms. The area where e-culture became the continuation and “shell” for “classical” (“pre-informational”) one is more than the products of e-culture and is

beyond measure. Because of it, the value of this phenomenon and the former achievements cannot be compared. The outstanding feature of the form and content of this culture is evident and is showing the pronounced tendency for extension and development of this sphere. The main space of e-culture development is the Internet, just here all of its forms are concentrated. E-culture comprises such spheres as communication, leisure and education, influencing significantly on human world view. Under the conditions of e-culture, that has been actively developing within last three decades, a new generation has already grew up having its own values, needs and ethical principles which can become an active factor of social dynamics in their turn.

EXISTENTIAL CHALLENGE UNDER THE CONDITIONS OF E-CULTURE

Anyway the world of values is associated with individuals’ desire to stabilize and prolong their being in the material and spiritual dimensions. E-culture allows a human in that to get a new form of being that is connected with information technologies and can be seen and felt. Human beings find in the virtual (online) space their “other” being, thereby, at first sight they contribute to the solving of the problem of extremity and limitation of their biosocial possibilities. Nowadays, human beings are not already “classical” *Homo Sapiens*, but this is a greatly virtualized biosocioelectronic subjects, being active both in the sphere of their physical nature, social practice and as electronic and virtual subjects, having great technological super-possibilities (crossing space-time borders and overcoming of speeds) and being a part of the global Internet community, having a demand for the virtual interaction with a nickname, electronic account, achievements and rating. The virtual alternative of human beings of the e-culture favors the development of their character features and possibilities and transforms and sublimates their real activity into a virtual one.

Human beings of e-culture spend the most part of their life in the virtual space, where they work, play, get to know new people, intercommunicate and even bury themselves. In this space personalities transform into a virtual character – an image, created by them to logging onto the Internet, where they live in an extremely dynamic sphere, requiring them to apply best efforts and mobilize the strengths and possibilities not used before. The sphere of the e-culture is a kind of the “second” life, prototype and continuation of the reality and “live culture”, where human beings of the pre-digital age searched for the solving of their existential problems: being, death, irrationality, “not freedom” and solitude.

The existence of human beings, living in the virtual space, has significant distinctions to the reality: human beings can live and die here many times, choose their presence in the Net or ignore it, specify the conditions of their being, imagine their own life story and purposes. Under my proposed approach in some specified sense the conditions of e-culture allow to find *new solutions of the human existential problems*:

“Being-towards-death” A digital form of the being becomes one more version of the existence in the world and the prolongation of human existence till the moment of the storage of the information about a subject. Nowadays more often “to be” means for a human being to be online, to have a profile, web-site, blog, etc., in such a way the information reality becomes a new kind of personal manifestation and therefore overcoming of its extremity.

“Fallenness into the world” as an impossibility to choose the conditions of the birth and the existence: In the virtual sphere the existence is determined and modeled by subjects, that involves the possibility of subjects to choose the conditions of their existence.

“Transcendence into the world”: The virtuality becomes the possibility of an individual to entrance the external world at unprecedented level, when a human being can address to millions of people

and be heard and understood. The entrance of the external space is the logging onto the Internet, joining other virtual and real participants of the interaction.

“I and Other one”: information addressing from a person to another one is conducted in new forms of multi-sided reflection when the epistolary form of the communication and existence of diary - the self-analysis (blog) becomes available for a number of people.

“Freedom” The electronic form of being offers individuals a new life world, simulated or chosen from a lot of versions by themselves. Human beings are free in such a world; they can be or not to be (log onto or out the Net at their own wishes), save or change themselves (their status, names or essence); accept or reject the terms of one or another system (a game or a social networking site); communicate or be alone; follow rules and norms or violate them.

In terms of existence, the virtuality attracts that a human being feels maximum freedom of action here and “infinite” possibilities, which one cannot achieve in the world with natural, social, and ethical restrictions. However, the online being is not real. Does it contribute to the solution of real meaningfully-vital problems in this case or help a human being to abstract from them behind the illusion of the freedom of own online image in actual fact?

I assume that e-culture gives human beings not only illusions but also favors the finding of own essence and creation of own world when they are not satisfied with real (offline) life “here and now”. Human beings still have individual decision making process under the conditions of this new being from as they can dive into new “Das man”, become the mass with specified qualities or a creator and find freedom to control the information, time, and space.

It should be noted, the conditions of e-culture (the existence of human beings “here and now”), where they find their essence, form and also *new existential problems*, challenges, that they had to face for the first time:

“Irrationality of relationship ‘virtual I and virtual Other one’: In fact, this is a new problem definition of the relationship to Other one, dealing with the communication virtualization, missing of the factor of emotional and personal perception, responsibility, honesty, etc. Under the communication virtualization, the interaction with each its part turns into a game, quasi-reality, where individuals realize that, their partners can be “any one”, even an answering machine that makes honesty, frankness, and trusting absurd.

“Loss of the reality limits”: The issue, related to the double consciousness of a human being, living in two worlds; In this regard, the offline world becomes less important, but the online one, the value of which is growing, remains to be a simulacrum, depriving a human being of “ground”.

“Virtual objectivization”: The frontier between individual and community phenomena under e-culture gets less clear. A human being is being involved into the mass more and more and becomes a part of the global network mechanism, conducting the options as many other ones according to the standard developed by a moderator.

“Freedom of online choice”: The integration with their online images, technologization of human beings, results in multi-alternative and pluralistic models of behavior, the moral regulator of which are replaced with anarchy, “instructions” or “game rules”; Ethical choice is extremely hampered and transferred from the sphere of moral appraisal into the sphere of personal and moral self-evaluation in full.

“Solitude in the Net”: The issue of loss of the interpersonal communication, dipping into the Internet-space, where everybody exists alone with a machine; The alienation of human beings from traditional grounds: nature, family, school, friends, which play less significant role, turning a person into a biosociovirtual “monad”.

“Online not liberty”: Within information environment human beings find new liberties and dependences. The main one of them (from e-culture) is the striving to be all the time online, the fear of the reality and interpersonal communication, and loss of interest in the life.

Particularly, the urgent problems has turned to be related with the development of new not liberties of the information age. Within a given context the following examples can be given. Thus, the neurologist P. Whybrow compares computer dependence with drug addiction and notes that computers intensify the excitement and enhance depressions of users at the same time: “The computer is electronic cocaine for many people. Our brains are wired for finding immediate reward. With technology, novelty is the reward. You essentially become addicted to novelty” (Whybrow 2012).

The Internet-dependence represents with the development of human aspiration to be online, interaction with online partners and online game obsession, and as consequence the disregard to own health, offline social relationships, reduction of cognitive and creative activities, and intensification of reality perception form when one perceives the information and does not process it. The serious risks involve the transferring of values into online world, social networking sites, hindering offline communication and the ability to solve life problems and enhancing destructive thinking and self-destructive behavior. In China, South Korea and Taiwan the internet-dependence is recognized in medical science as a disease and the Internet dependence is considered as a nationwide crisis. In these countries tens of millions people are considered to be Internet-dependent.

It should be noted, Internet-dependence is both psychological and existential issue, as it makes human beings feel themselves incomplete and “disabled” offline. Gamers or active users of social networking sites can get the feeling of being fullness only online, the reality is not important for them.

The perspectives of these changes can result in the further tendency when a person turns into any “superhuman”, existing in offline and online worlds, and loses emotional and sensorial sphere, demand for reflection and meaningfully-life search. The existential problems and risks get new forms, breaking down the reality boundaries for

human beings and degrading their uniqueness at the same time.

The overcoming of the solitude and the insulation of personal existence involve the human possibility of communication and transcendence to other one through love, friendship, empathy, and sympathy. E-culture produces new possibilities for the communication of new global type and human interaction without boundaries which have never existed before. Online communication has united the milliards of people nowadays. According to some estimation over 2 billion people from every country and continent are online at the same time. Of course, this unity cannot be compared with real communication; however, it is impossible and even dangerous to ignore this fact. Neglecting of the role of online communication and online communities can result in the estrangement of parents from their children, facilitate the breaking off between online and offline relationships, contribute to the extension of crisis of human social identity. It should be pointed out that a large number of people consider the Internet is an important form of the communication and interaction.

The virtualization of the communication sphere and social interaction is probably an area of the highest risk under the information age that contributes to the changing of personal essence, the forms of experience transfer existing for millenniums, traditions, and values.

The shift from the personal communication to the communication mediated by information technologies becomes a new condition of the development of modern human beings, who overcome age-related, psychological and physiological complexes within the communication on the one hand and can get new phobias and fears on the other hand. Thus, modern psychologists diagnosed nomophobia as a social fear of the direct interpersonal communication caused by the habit to use the computer-aided or cellular communication (King 2013).

The interpersonal communication is more replaced with its online simulation, displacing such the most important relationship spheres as love

and friendship. Modern researchers note that the number of network romantic relations has been increased, and virtual relations (called “the Second Life”) sometimes can improve the mental state of persons, help to solve their real problems (Golbert 2011). The concept “friendship” is undergoing certain changes under e-culture, dealing with the loss of stability, social and financial support, and responsibility (Amichai-Hamburger 2012). The simulation of feelings in these spheres facilitates the loss of human existential grounds and changes moral and ethical milestones related to the understanding of Other one’s value.

In some specified sense, human beings live in two worlds, which they do not trust in full; offline communication remains to be important, although the value of virtual Other one is developing, the interaction with who has its rules. The logging onto Net to “meet” Other one, and transcending under these conditions turns out to be insignificant, senseless, as these relationships involve the simulation of feelings, trusting, understanding, and participation in Other one’s life.

INFORMATION FREEDOM

Freedom remains to be the most important existential value of a person. The UNESCO Code of Ethics for the Information Society (2011) was developed for the absolute guarantee of the rights and the freedom of a person in the information society. Has a human of the Information Age become freer in such a way?

On the one hand, new technologies make possible high-speed and high-precision processing of the information, discharging a human from the necessity to do routine activity. Thus, it seems, a person has more free time, but in fact a modern human spends this free time for new activities, involving (a) maintenance of information systems (development and installation of new software programs, fight against viruses, protection of data and failure recovery), (b) implementation of volume of work increased in conditions of

information boom, (c) uploading and downloading of data and their reprocessing, (d) search of necessary information and check of its validity, (e) processing of increased flow of documents, high tech innovations, etc.

Nowadays a modern specialist should have more competencies (language, intercultural, information and other competences), than last years and centuries, and much of free time is spent for the development of this knowledge and these skills. In its turn, it requires new methods of training, extension of professional skills, especially for brain workers, HR managers and workers, dealing with information processing (Hemingway & Gough, 2000). During the development of information facilities a person more often spends time to achieve a high level of high technologies. Human beings are developing together with them, in which case they are becoming unfreer in their activity.

What are the perspectives of freedom development in the Information Age? Information first of all has provided us with significant freedom for obtaining of information which should not be underestimated. That, what was earlier limited and was represented only for professionals, became available to the wide public. Politicians, businessmen and community leaders have to regard, that their acts have not already been confidential and evoke a broad public response at once. In this case the quality of information and messages, presented by mass media, has sharply become lower. Information is full of plagiarism, unverified and unconfirmed facts. There are almost no prohibitions and taboos for the modern society of information freedom. Alistair Duff emphasized the situation of “normative crisis”, that is characteristic for the information society, where traditional norms have not already been effective, but new principles have not formed yet (Duff 2008). It veils one of its main dangerous tendencies. Risks and instability, produced by achieved freedom, are growing and result in instability, chronic crisis and bifurcation of the system. A new factor, that can break fragile peace and agreements, can appear at any moment of such system development. The

maximum of freedom involves a high degree of responsibility and morality, though the last ones, as we have found out, are not valuable priorities of the new world.

Electronic credit cards, identification documents both are becoming opportunities for more freedom in the activity and produce new nonliberties for a person. As Simon Rogerson pointed out, human beings, not having such choice, cannot feel themselves comfortable in the modern world as nowadays their personalities exist in two levels: physical and electronic ones (Rogerson 1998).

Freedom for new generation became the main achievement and welfare, but during its realization, the issue of sense and purpose of its use arose more acutely. Freedom that has been understood as the end in itself for a long time became a value and a facility during its implementation to achieve something else. Now freedom represents a value as a constant selection opportunity, where it is important not so much direction and result of a choice, as the state of choosing. Freedom of choice is realized during consumption of goods, in communication with other people, as an opportunity of moving in space and a choice of leisure, education, etc. The availability of freedom does not still mean its high function by itself. Freedom can turn out to be not a form of self-development or moral development, but only a choice of one commodity from many ones. The freedom of the information society is an opportunity to achieve a specified purpose; however, technologies do not set these purposes.

FUTURE RESEARCH DIRECTIONS

Further research of this topic related to the study of such phenomena and the problems of virtual communication, ethics in the electronic culture, the problem of preservation of identity in the information world. The main current trend of the modern society encompasses the decline of the role of the traditional (real) and distant-traditional communication and the increase of the value of the

distant-nominal communication. As characteristic features of the distant-nominal communication, the special attention to the Ego demonstration, the continuity of the virtual communication, “radio effect”, weakening of emotional communication, formation of the need to keep the virtual communication turning sometimes into addiction, the simulation of real relations and decline in trust. As the result, the real life has been virtualized.

The ethics and etiquette of the virtual communication have still been “devised”. The external processes of the regulation of the ethical principles of the information society (In this case, it is necessary to recognize the leading role of UNESCO) and the self-organization projects (i.e., design of ethical standards in virtual communities), have taken place. Virtual communication does not change the people’s ideas of moral values but facilitates their attenuation and transformation into “relicts” of the traditional reality in view of the simulation and game character and of reduced society control. The strict arrangement of the virtual content and its subjection to the proposed codes inevitably will produce the aspiration to escape from the regulations therefore the censorship development will not resolve this problem. In this author’s view, the educational system and cultural life in the information world needs to contribute to the socialization of people and their adaptation to new technological conditions without losing sight of social attributes to be effective.

CONCLUSION

Thus, e-culture does not replace the original things and should not oust them completely. The former classical, “live” culture is passing in large part through the crisis dealing with the loosing of spiritual values in the age of consumption because of electronic culture. Its innovative form attracts a present-day’s person and in such a way continues developing. In such a case, it is important to understand that this form does not replace but

develop and upgrade cultural heritage allowing it to be available and preserved for descendants. E-culture is not only the duplication of “live” culture, but also new achievements done by the application of creative high technologies that provided radically new facilities and trends of the self-expression, cognition, and development. The electronic resources multiply the abilities of human beings and give them new liberties and new life world. In this regard, the transition from the real sphere of the interaction into the virtual one reduces the importance of the interpersonal communication, moral and ethical norms, and traditional life forms. The e-culture development in the existential contexts should be referred to the most important ones: the development of a human value of a new type – virtual existence; the development of e-culture, contributing to the solution of personal existential problems (death, solitude, not-liberty) within virtual “being here”; the technocratization and virtualization of human beings themselves, intergrowing with information systems and getting super-possibilities in exchange for the freedom in terms of existence; the deepening of existential vacuum under the virtualization of culture and human beings, coming from the loss of the reality boundaries, the virtualization of human beings’ essence, losing the real identity with the social world (socium) (and several its groups) in exchange for the identity with virtual groups, images, and Technologies; the reduction of the importance of the role of real interpersonal communication, being a traditional basis of personal being, shift to the communication, and mediated information space; the development of new ethics in the field of e-culture and e-communication, where axiological pluralism becomes a new challenge to the moral behavior of individuals.

The changes, that culture and the sphere of human being are undergoing nowadays, require to be studied in details as the direction of the further subject’s activity in the nature and the society, multiply increased by the facilities of high technologies.

REFERENCES

- Amichai-Hamburger, Y., Kingsbury, M., & Schneider, B. H. (2012). Friendship: An old concept with a new meaning? *Computers in Human Behavior*, 29, 33–39.
- Baarda, R., & Rocci, L. (2012). The Use and Abuse of Digital Democracy. *International Journal of Technoethics*, 3(3), 50–69. doi:10.4018/jte.2012070104
- Baeva, L. V. (2012). Existential Axiology. *Culture. International Journal of Philosophy of Culture and Axiology*, 9(1), 73–85.
- Brumbaugh, R. S. (1972). *Changes of Value Order and Choices in Time. In Value and Valuation. Axiological Studies in Honor of Robert S. Hartman. The University of Tennessee Press.*
- Capurro, R. (2006). Towards an Ontological Foundation of Information Ethics. *Ethics and Information Technology*, 8(4), 175–186. Retrieved from <http://www.capurro.de/oxford.html>
- Castells, M. (1997). *The End of the Millennium, the Information Age: Economy, Society and Culture: Vol. 3.* Oxford, UK: Blackwell.
- Cockton, G. (2004). *From quality in Use to Value in the World. In Proceedings of the CHI'04 Extended Abstracts on Human factors in Computing Systems (CHI'04).* New York: ACM Press. doi:10.1145/985921.986045
- Duff, A. (2008). The Normative Crisis of the Information Society. *Cyberpsychology (Brno)*, 2(1), 3. Retrieved from <http://cyberpsychology.eu/view.php?cisloclanku=2008051201&article=3>
- Ethical code of Information society. (n.d.). 36 general conference of UNESCO. Retrieved from <http://unesdoc.unesco.org/images/0021/002126/212696e.pdf>
- Gilbert, R. L., Murphy, N. A., & Ávalos, C. M. (2011). Communication Patterns and Satisfaction Levels in Three-Dimensional Versus Real-Life Intimate Relationships. *Cyberpsychology, Behavior, and Social Networking*, 14(10), 585–589. doi:10.1089/cyber.2010.0468 PMID:21381970
- Haddon, L. (2004). *Information and Communication Technologies in Everyday Life: A Concise Introduction and Research Guide.* Oxford, UK: Berg.
- Masuda, Y. (1983). *The Information Society as Post-Industrial Society.* Washington, DC: Academic Press.
- McLean, A. (2011). Ethical frontiers of ICT and elders: Cultural, pragmatic and ethical issues. *Ethics and Information Technology*, 13(4), 313–326. doi:10.1007/s10676-011-9276-4
- Rogerson, S. (1998). *Social Values in the Information Society.* In FTI Annual Report, Forum of Information Technology, Milan, Italy.
- Ronchi, A. M. (2009). *E-Culture.* New York: Springer-Verlag, LLC.
- Vokey, D. (2001). *Moral Discourse in Pluralistic World.* Notre Dame Press.
- Whybrow, P. (2012). *Manic Nation.* Retrieved from <http://www.psmag.com/health/manic-nation-dr-peter-whybrow-says-were-addicted-stress-42695>

ADDITIONAL READING

- Brey, P. (1999). The Ethics of Representation and Action in Virtual Reality. *Ethics and Information Technology*, 1(1), 5–14. doi:10.1023/A:1010069907461

Floridi, L. (1999). Information Ethics: On the Philosophical Foundations of Computer Ethics. *Ethics and Information Technology*, 1(1), 37–56. doi:10.1023/A:1010018611096

Hemingway, C. J., & Gough, T. G. (2000). The Value of Information Systems Teaching and Research in the Knowledge Society. *Informing Science*, 4 (3), 167–184

King, A. L. S., Valença, A. M., Silva, A. C. O., Baczynski, T., Carvalho, M. R., & Nardi, A. E. (2013). Nomophobia: Dependency on virtual environments or social phobia? *Computers in Human Behavior*, 29(1), 140–144. doi:10.1016/j.chb.2012.07.025

Ott, M., & Pozzi, F. (2011). Towards a New Era for Cultural Heritage Education: Discussing the Role of ICT. *Computers in Human Behavior*, 27(4), 1365–1371. doi:10.1016/j.chb.2010.07.031

Raban, D. R. (2009). Self-Presentation and the Value of Information in Q&A Websites. *Journal of the American Society for Information Science and Technology*, 60(12), 2465–2473. doi:10.1002/asi.21188

Sartor, G. (2012). Human Rights in the Information Society: Utopias, Dystopias and Human Values. Philosophical Dimensions of Human Rights. Ed. Claudio Corradetti., 2012, 293–307.

Zhou, L., Ding, L., & Finin, T. (2011). How is the Semantic Web evolving? A Dynamic Social Network Perspective. *Computers in Human Behavior*, 27(4), 1294–1302. doi:10.1016/j.chb.2010.07.024

KEY TERMS AND DEFINITIONS

Digital Information And Virtual Communication: Is the communication form between people by means of information technologies at remote distances with the following features: global character, the self-organizing network

form, the spontaneously formed diverse content, simultaneous participation of a great number of communication subjects, impersonal nature, possibilities of anonymous interaction, coexistence of real interpersonal communications, and artificial simulacrum (nominal) relations.

E-Culture Culture or Digital Culture: Is a new sphere of the human activity, associated with the creation of the electronic copies of spiritual and material objects as well as the creative work of the virtual objects of science, communication and art.

Ethics of Information Society: Ethical principles, norms, rules of behavior of people in the conditions of a virtual form of communication (the Internet, games, mobile communication, etc.), directed on formation of valid, humane, legal space for communication.

Transcendence Into the Virtual World: Is the process of entrance of identity in the information interaction in the use of digital technology, which creates a feeling part of the whole transpersonal. The entrance of the external space is the logging onto the Internet, joining other virtual and real participants of the interaction.

Types of e-Culture Expression: 1. electronic form of former (traditional, classical – here not electronic) cultural objects (e.g., on-line museums, libraries, exhibitions, etc.). 2. electronic cultural objects in form and essence (computer games, social networking sites, Internet, digital art pieces, etc.).

Virtual Objectivization: blurring the boundaries between personal and social phenomena in a virtual interaction that converts human values and outlook

Virtualization: Is the process when the real processes and phenomena are replaced by the virtual forms. On the one hand, it deals with the creation of digital analogs and simulations of the real culture objects and on the other hand, it deals with the creativity of the new cyberspace and its phenomena.

The Fundamentals of Human-Computer Interaction



Kijpokin Kasemsap

Suan Sunandha Rajabhat University, Thailand

INTRODUCTION

As modern technologies continue to develop, the ability of users to interface with new computer systems becomes a paramount concern (Information Resources Management Association, 2016). Human-computer interaction (HCI) is a field of research that develops the advanced activities of user-computer interaction (Anishchenko, Osinov, & Shaposhnikov, 2011). HCI has been the focus of attention for researchers in the past decade, with considerable work being done in the various modalities for communicating with the computer (Reddy & Basir, 2010). HCI studies play an important role in the design, implementation, and evaluation of a new generation of graphical user interfaces designed to support consumer behaviors and information needs (Servidio, Davies, & Hapeshi, 2016).

Humans are limited in their capacity to execute the information. This perspective has the important implications for the interaction design. Information is received and responses given via a number of input and output channels, such as visual channel, auditory channel, haptic channel, and movement. A primary goal of virtual environments is to support the efficient, powerful, and flexible HCI (Devyatkov & Alfimtsev, 2013). A computer system comprises various elements, each of which affects the user of the system. Input devices for interactive use, allowing text entry, drawing, and selection from the screen. Output display devices for interactive use: different types of screen mostly using some form of bitmap display; large displays and situated displays for shared and public use; and digital paper may be usable in the near future.

The interaction takes place within a social and organizational context that affects both user and system. Iterative design practices work to incorporate crucial customer feedback early in the design process to inform critical decisions which affect usability. Interaction models help practitioners and researchers understand what is going on in the interaction between user and system. Many interfaces, relying on body gestures (Liu & Jia, 2004), speech (Green & Eklundh, 2003), and brain computing (Mason & Birch, 2003), have been proposed or developed which assist humans to interact with robots in a more intelligent and natural activity. Various interaction techniques have been developed for interactive 3D environments (Jankowski & Hachet, 2015).

This article aims to bridge the gap in the literature on the thorough literature consolidation of HCI. The extensive literature of HCI provides a contribution to practitioners and researchers by explaining the fundamentals of HCI in order to maximize the impact of HCI in global operations.

BACKGROUND

HCI emerged as a distinct research discipline in the late 1970s and early 1980s when monitors and workstations became available and opened up the use of computers to non-engineers (Grudin, 1990). In the 1990s, network technology and mobile devices broadened the scope of research beyond the individual user and personal computers (Hollender, Hofmann, Deneke, & Schmitz, 2010). HCI should be designed to be effective, efficient, engaging, error tolerant, and easy to learn (Fall-

DOI: 10.4018/978-1-5225-2255-3.ch364

man, 2010). With the development of computer vision technology, researchers have studied a lot of HCI methods to replace the keyboard and mouse (Juan, 2012).

Major HCI research areas comprise theories and models of human behavior when interacting with information technology (IT), general or more specific guidelines or heuristics for the design and evaluation of IT, methods for the user-centered development of IT, and the development of new interaction paradigms (Preece, Sharp, & Rogers, 2002). Interaction models address the translations between what the user wants and what the system does. Ergonomics looks at the physical characteristics of the interaction and how these influence its effectiveness. The dialog between user and system is influenced by the style of the interface. Examples of effective strategies for building interactive systems provide paradigms for designing usable interactive systems. The evolution of these usability paradigms also provides a good perspective on the history of interactive computing.

Different HCI techniques and technologies with different design criteria have been developed for several decades (Rantanen, Niemenlehto, Verho, & Lekkala, 2010). Interaction design is about creating interventions in often complex situations using technology of many kinds including PC software, the web, and physical devices. Usability is key to the success of any interactive system—from commercial software to business-to-business (B2B) websites to the handheld devices (Rosson & Carroll, 2001). Promoting consistency in user interface and application design remains a prominent practical issue (Nielsen, 1989). Software engineering provides a means of understanding the structure of the design process, and that process can be assessed for its effectiveness in interactive system design. Designing for maximum usability is the goal of interactive systems design. Evaluation tests the usability, functionality and acceptability of an interactive system.

THEORY AND APPLICATIONS OF HUMAN-COMPUTER INTERACTION

This section emphasizes the overview of HCI; cognitive models, socio-organizational issues, and stakeholder requirements; HCI and hand gesture recognition; and the multifaceted applications of HCI.

Overview of Human-Computer Interaction

With the advancement of technologies, a computer system has become a very powerful machine which has been designed to make the human beings' tasks easier (Sharma & Verma, 2015). The power of computers begins to be exploited in subjective areas of human study like those related to human psychology in order to make the interaction between humans and computers more natural (Choubey & Singh, 2012). To achieve the effective HCI, the computer should be able to naturally interact with the user, similar to the way that HCI takes place (Sebe, Cohen, Cozman, Gevers, & Huang, 2005).

Defining what, how, when, and why to build computer resources with accessibility features is one of the main concerns of the HCI area (Peres, Boscarioli, Bidarra, & Fantinato, 2012). HCI pervasively exists in the individuals' daily lives (Yeo, Lee, & Lim, 2015). From an applied perspective, technological advances have changed the way that humans interact with their computers (Posard, 2014). The application of HCI principles are frequently successful in increasing user satisfaction and engagement with an interface (Wohl, Parush, Kim, & Warren, 2014). The correct functioning of interactive computer systems depends on both the faultless operation of the device and correct human actions (Curzon, Ruksenas, & Blandford, 2007).

More natural and comfortable HCI methods are the goal of several research efforts (Tang, Chen, Zheng, Han, & Li, 2015). Traditional methods

in gesture recognition are based on visual sense (Munoz-Salinas, Aguirre, & Garcia-Silvente, 2007), wearable devices or a combination of the two (Cheng, Xie, Bian, & Tao, 2012). Benefits of utilizing the visual sensing include the passive detection, non-contact, and freedom of movement (Tang et al., 2015). As a human body in motion becomes electrostatically charged through contact and friction charging (Cross, 1987), the resultant change of the electrostatic field enables an indirect measurement of body movement.

Many researchers measure electrostatic signals during walking with the measuring device on the body (Ficker, 2006) and utilize the non-contact method to measure the human body's electrostatic signal (Kurita, 2011), particularly the electrostatic signal generated from walking (Takiguchi, Wada, & Toyama, 2008). Evolutionary visual software analytics is the process for supporting software maintenance, with the active participation of users, through the understanding and comprehension of software evolution by means of visual analytics and HCI (Gonzalez-Torres, Garcia-Penalvo, & Theron, 2013). The visual representations of huge datasets supported by browsing and navigation capabilities are challenging due to the limited size of screens (Leung & Apperley, 1994).

The design of visual representations uses the linked views and interaction mechanisms to allow the exploration of details while keeping the context (Hornbaek & Hertzum, 2011) and allows the tracking of relationships to facilitate the interpretation of specific elements. Interaction design patterns has been designed for the repeatable solutions in interface and interaction design (Tidwell, 2011) and several researches has been conducted on visualization design and HCI. Radio frequency identification (RFID) tags have been used by several researchers as a key component for HCI in a computer-augmented environment (Nishi, Sato, & Koike, 2001). RFID solutions can be utilized to reduce the operating costs through decreasing labor costs, enhancing automation, improving tracking and tracing, and preventing the loss of materials (Kasemsap, 2015a).

Cognitive Models, Socio-Organizational Issues, and Stakeholder Requirements

A cognitive model is an approximation to animal cognitive processes (predominantly human) for the purposes of comprehension and prediction. Cognitive models can be developed within or without a cognitive architecture, though the two are not always easily distinguishable. Cognitive models tend to be focused on a single cognitive phenomenon or process (e.g., list learning), how two or more processes interact (e.g., visual search and decision making), or to make behavioral predictions for a specific task or tool (e.g., how instituting a new software package will affect productivity). Cognitive models represent users of interactive systems. Hierarchical models represent a user's task and goal structure. Linguistic models represent the user-system grammar. Physical and device models represent human motor skills. Cognitive architectures underlie all of these cognitive models.

There are several organizational issues that affect the acceptance of technology by users and that must be considered in system design: systems may not take into account conflict and power relationships; those who benefit may not do the work; and not everyone may use systems. HCI designers must identify specific stakeholder requirements within their organizational context. Socio-technical models capture both human and technical requirements. Soft systems methodology takes a broader view of human and organizational issues. Participatory design includes the user directly in the design process. Ethnographic methods study users in context, attempting to take an unbiased perspective.

Human-Computer Interaction and Hand Gesture Recognition

Various approaches to HCI have been proposed in the last few decades as an alternative to the classic input devices of keyboard and mouse (Kiliboz

& Gudukbay, 2015). The touch panel and voice control are the most new input devices for HCI (Hsu, 2015). Human hand recognition plays an important role in a wide range of applications ranging from sign language translators, gesture recognition, augmented reality, surveillance, and medical image processing to various HCI domains (Bilal, Akmeliawati, Shafie, & Salami, 2013). Hand gesture recognition is widely used in many applications, such as in computer games, machinery control (e.g., crane), and mouse replacement (Hasan & Abdul-Kareem, 2014).

HCI with hand gestures plays a significant role in these modalities because humans often rely on their hands in communication or to interact with their environment (Kılıboz & Gudukbay, 2015). Hand-gesture-based methods provide the natural way of interaction and communication (Pavlovic, Sharma, & Huang, 1997). Many studies evaluate the gesture-based interaction techniques, and propose the useful ways to increase their effectiveness (Rautaray & Agrawal, 2015). On the human side, system performance is limited by the cognitive and motor processes used for gesture formation (Rempel, Camilleri, & Lee, 2014). On the computer side, system performance is limited by the rate and fidelity of gesture capture and processing (Rempel et al., 2014).

Multifaceted Applications of Human-Computer Interaction

The computers are social actors (CASA) effect refers to the application of social rules when individuals interact with computers (Liang, Lee, & Jang, 2013). Speech emotion recognition (SER) represents one of the emerging fields in HCI (Ramakrishnan & El Emary, 2013). Emotion plays an important role in the interaction with users and computers (Park & Kim, 2015). Speech technology plays a central role in enhancing HCI, especially for small devices for which graphical user interface has obvious limitations (Deng & Yu, 2005).

Lee (2008) also found that people adhere to the recommendations from a male-voiced computer more than a female-voiced computer, reflecting gender stereotypes.

Regarding the proliferation of the low-cost vision sensors, embedded processors, and efficiency of wireless networks, a large amount of research has focused on the utilization of multiple sources of information for the analysis of behavior (Garcia-Rodriguez & Garcia-Chamizo, 2011). This perspective has the potential applications, such as human-computer interfaces to virtual reality, gaming, gesture-based control, presence detection, and event intelligence applications for human-centric applications, such as the detection of falls in care of older people (Garcia-Rodriguez & Garcia-Chamizo, 2011). Researchers have examined the role of mindlessness during HCI (Liang et al., 2013).

Nass and Moon (2000) identified four perspectives of general responses when individuals interact with computers: the overuse of human social categories (e.g., applying gender stereotypes to computers); the exhibition of overlearned social behaviors (e.g., reciprocity and politeness); the demonstration of premature cognitive commitments (e.g., specialist television perceived as having better content than a generalist television set); and the alteration of responses to a given computer personality (e.g., participants responded to computers when the communication style from the computer matched versus mismatched participants' personality). Many researchers examine the effects by altering the nature or content of computer feedback (Liang et al., 2013). Mindlessness occurs even when computer agents display more human-like cues (Lee, 2008). Recent research extends the CASA effect to the interaction with computer agents on websites (Karr-Wisniewski & Prietula, 2010). The cognitive deficit approach treats mindlessness as an inability to process information due to the presence of competing cognitive demands (Lee, 2010).

FUTURE RESEARCH DIRECTIONS

The classification of the extensive literature in the domains of HCI will provide the potential opportunities for future research. The Internet of Things (IoT) is the growing range of everyday objects acquiring connectivity, sensing abilities, and increased computing power. Cloud computing includes network access to storage, processing power, development platforms, and software (Kasemsap, 2015b). Wireless sensor networks (WSNs) consist of many small, resource-constrained sensor nodes together with several base stations. WSNs can combine the objective physical world and the logic information world together, toward enhancing the utilization of robots. The benefits of robots has increased their flexibility with being capable of performing a variety of tasks and applications. Regarding robotics, robots can allow for increased production and profit margin because they can complete tasks faster. HCI, the IOT, cloud computing, WSNs, and robotics are remain important issues for future research directions.

CONCLUSION

This article highlighted the overview of HCI; cognitive models, socio-organizational issues, and stakeholder requirements; HCI and hand gesture recognition; and the multifaceted applications of HCI. HCI is characterized as a dialogue or interchange between the human and the computer because the output of one serves as the input for the other in an exchange of actions and intentions. HCI is a sociotechnological discipline whose goal is to bring the power of computers and communication systems to people in ways and forms that are both accessible and useful in the effective manner. HCI involves the activities of humans using computers. A basic goal of HCI is to improve the interactions between users and computers by making computers more usable and receptive to the user's needs.

HCI encompasses not only ease of use, but also new interaction techniques for supporting user tasks, providing better access to information, and creating more powerful forms of communication. HCI involves input and output devices and the interaction techniques that use them; how information is presented and requested; how the computer's actions are controlled and monitored; all forms of help, documentation, and training; the tools used to design, build, test, and evaluate user interfaces; and the processes that developers follow when creating effective human-computer interfaces.

HCI addresses the design, evaluation, and implementation of interactive computing and computing-based systems for the benefit of human use. HCI plays an important role in identifying the environmental and social issues which can affect the use of systems, and providing techniques to ensure the design of the system will be usable, effective, and safe. HCI draws on computer science, computer and communications engineering, graphic design, management, psychology, and sociology as it tries to make computer and communications systems ever more usable in executing the tasks. HCI is an important consideration for any business that uses computers in their everyday operation.

REFERENCES

- Anishchenko, S. I., Osinov, B. A., & Shaposhnikov, D. G. (2011). Head pose estimation in solving human-computer interaction problems. *Pattern Recognition and Image Analysis*, 21(3), 446–449. doi:10.1134/S1054661811020052
- Bilal, S., Akmeliawati, R., Shafie, A. A., & Salami, M. J. E. (2013). Hidden Markov model for human to computer interaction: A study on human hand gesture recognition. *Artificial Intelligence Review*, 40(4), 495–516. doi:10.1007/s10462-011-9292-0

- Cheng, J., Xie, C., Bian, W., & Tao, D. (2012). Feature fusion for 3D hand gesture recognition by learning a shared hidden space. *Pattern Recognition Letters*, 33(4), 476–484. doi:10.1016/j.patrec.2010.12.009
- Choubey, A., & Singh, R. (2012). Neuro linguistic programming: Towards better understanding of human computer interaction. In *Computer engineering: Concepts, methodologies, tools and applications* (pp. 1733–1743). Hershey, PA: IGI Global. doi:10.4018/978-1-61350-456-7.ch708
- Cross, J. A. (1987). *Electrostatics: Principles, problems and applications*. Bristol, UK: Adam Hilger.
- Curzon, P., Ruksenas, R., & Blandford, A. (2007). An approach to formal verification of human-computer interaction. *Formal Aspects of Computing*, 19(4), 513–550. doi:10.1007/s00165-007-0035-6
- Deng, L., & Yu, D. (2005). A speech-centric perspective for human-computer interface: A case study. *Journal of VLSI Signal Processing Systems for Signal. Image and Video Technology*, 41(3), 255–269.
- Devyatkov, V., & Alfimtsev, A. (2013). Human-computer interaction in games using computer vision techniques. In *Image processing: Concepts, methodologies, tools, and applications* (pp. 1210–1231). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-3994-2.ch061
- Fallman, D. (2010). A different way of seeing: Albert Borgmann’s philosophy of technology and human-computer interaction. *AI & Society*, 25(1), 53–60. doi:10.1007/s00146-009-0234-1
- Ficker, T. (2006). Electrification of human body by walking. *Journal of Electrostatics*, 64(1), 10–16. doi:10.1016/j.elstat.2005.04.002
- Garcia-Rodriguez, J., & Garcia-Chamizo, J. M. (2011). Surveillance and human-computer interaction applications of self-growing models. *Applied Soft Computing*, 11(7), 4413–4431. doi:10.1016/j.asoc.2011.02.007
- Gonzalez-Torres, A., Garcia-Penalvo, F. J., & Theron, R. (2013). Human-computer interaction in evolutionary visual software analytics. *Computers in Human Behavior*, 29(2), 486–495. doi:10.1016/j.chb.2012.01.013
- Green, A., & Eklundh, K. S. (2003). Designing for learnability in human-robot communication. *IEEE Transactions on Industrial Electronics*, 50(4), 644–650. doi:10.1109/TIE.2003.814763
- Grudin, J. (1990). *The computer reaches out: The historical continuity of interface design*. Paper presented at the 8th Annual ACM Conference on Human Factors in Computing Systems (CHI 1990), Seattle, WA. doi:10.1145/97243.97284
- Hasan, H., & Abdul-Kareem, S. (2014). Human-computer interaction using vision-based hand gesture recognition systems: A survey. *Neural Computing & Applications*, 25(2), 251–261. doi:10.1007/s00521-013-1481-0
- Hollender, N., Hofmann, C., Deneke, M., & Schmitz, B. (2010). Integrating cognitive load theory and concepts of human-computer interaction. *Computers in Human Behavior*, 26(6), 1278–1288. doi:10.1016/j.chb.2010.05.031
- Hornbaek, K., & Hertzum, M. (2011). The notion of overview in information visualization. *International Journal of Human-Computer Studies*, 69(7), 509–525. doi:10.1016/j.ijhcs.2011.02.007
- Hsu, W. Y. (2015). Brain-computer interface: The next frontier of telemedicine in human-computer interaction. *Telematics and Informatics*, 32(1), 180–192. doi:10.1016/j.tele.2014.07.001
- Information Resources Management Association. (2016). Human-computer interaction: Concepts, methodologies, tools, and applications (Vols. 1–4). pp. 1–2208). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-8789-9
- Jankowski, J., & Hachet, M. (2015). Advances in interaction with 3D environments. *Computer Graphics Forum*, 34(1), 152–190. doi:10.1111/cgf.12466

- Juan, W. (2012). Human-computer interaction system based on GRBF & HMM. *Physics Procedia*, 24, 769–773. doi:10.1016/j.phpro.2012.02.114
- Karr-Wisniewski, P., & Prietula, M. (2010). CASA, WASA, and the dimensions of us. *Computers in Human Behavior*, 26(6), 1761–1771. doi:10.1016/j.chb.2010.07.003
- Kasemsap, K. (2015a). The role of radio frequency identification in modern libraries. In S. Thanuskodi (Ed.), *Handbook of research on inventive digital tools for collection management and development in modern libraries* (pp. 361–385). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-8178-1.ch021
- Kasemsap, K. (2015b). The role of cloud computing adoption in global business. In V. Chang, R. Walters, & G. Wills (Eds.), *Delivery and adoption of cloud computing services in contemporary organizations* (pp. 26–55). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-8210-8.ch002
- Kiliboz, N. C., & Gudukbay, U. (2015). A hand gesture recognition technique for human-computer interaction. *Journal of Visual Communication and Image Representation*, 28, 97–104. doi:10.1016/j.jvcir.2015.01.015
- Kurita, K. (2011). Human heartbeat measurement on the basis of current generated by electrostatic induction. *The Review of Scientific Instruments*, 82(2), 026105–026107. doi:10.1063/1.3541554 PMID:21361644
- Lee, E. J. (2008). Gender stereotyping of computers: Resource depletion or reduced attention? *Journal of Communication*, 58(2), 301–320. doi:10.1111/j.1460-2466.2008.00386.x
- Lee, E. J. (2010). What triggers social responses to flattering computers? Experimental tests of anthropomorphism and mindless explanations. *Communication Research*, 37(2), 191–214. doi:10.1177/0093650209356389
- Leung, Y. K., & Apperley, M. D. (1994). A review and taxonomy of distortion-oriented presentation techniques. *ACM Transactions on Computer-Human Interaction*, 1(2), 126–160. doi:10.1145/180171.180173
- Liang, Y., Lee, S. A., & Jang, J. W. (2013). Mindlessness and gaining compliance in computer-human interaction. *Computers in Human Behavior*, 29(4), 1572–1579. doi:10.1016/j.chb.2013.01.009
- Liu, Y., & Jia, Y. (2004). *A robust hand tracking for gesture-based interaction of wearable computers*. Paper presented at the 2004 8th IEEE International Symposium on Wearable Computers (ISWC 2004), Arlington, VA.
- Mason, S. G., & Birch, G. E. (2003). A general framework for brain-computer interface design. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 11(1), 70–85. doi:10.1109/TNSRE.2003.810426 PMID:12797728
- Munoz-Salinas, R., Aguirre, E., & Garcia-Silvente, M. (2007). People detection and tracking using stereo vision and color. *Image and Vision Computing*, 25(6), 995–1007. doi:10.1016/j.imavis.2006.07.012
- Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *The Journal of Social Issues*, 56(1), 81–103. doi:10.1111/0022-4537.00153
- Nielsen, J. (1989). *Coordinating user interfaces for consistency*. Boston, MA: Academic Press.
- Nishi, T., Sato, Y., & Koike, H. (2001). *SnapLink: Interactive object registration and recognition for augmented desk interface*. Paper presented at the IFIP Conference on Human-Computer Interaction (INTERACT 2001), Tokyo, Japan.
- Park, J. S., & Kim, J. H. (2015). Emotional information processing based on feature vector enhancement and selection for human-computer interaction via speech. *Telecommunication Systems*, 60(2), 201–213. doi:10.1007/s11235-015-0023-8

- Pavlovic, V. I., Sharma, R., & Huang, T. S. (1997). Visual interpretation of hand gestures for human-computer interaction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7), 677–695. doi:10.1109/34.598226
- Peres, S. M., Boscarioli, C., Bidarra, J., & Fantinato, M. (2012). Human-computer interaction and artificial intelligence: Multidisciplinarity aiming game accessibility. In *Computer engineering: Concepts, methodologies, tools and applications* (pp. 1–18). Hershey, PA: IGI Global. doi:10.4018/978-1-61350-456-7.ch101
- Posard, M. N. (2014). Status processes in human-computer interactions: Does gender matter? *Computers in Human Behavior*, 37(1), 189–195. doi:10.1016/j.chb.2014.04.025
- Preece, J., Sharp, H., & Rogers, Y. (2002). *Interaction design: Beyond human-computer interaction*. New York, NY: John Wiley & Sons.
- Ramakrishnan, S., & El Emary, I. M. M. (2013). Speech emotion recognition approaches in human computer interaction. *Telecommunication Systems*, 52(3), 1467–1478. doi:10.1007/s11235-011-9624-z
- Rantanen, V., Niemenlehto, P. H., Verho, J., & Lekkala, J. (2010). Capacitive facial movement detection for human-computer interaction to click by frowning and lifting eyebrows. *Medical & Biological Engineering & Computing*, 48(1), 39–47. doi:10.1007/s11517-009-0565-6 PMID:20016948
- Rautaray, S. S., & Agrawal, A. (2015). Vision based hand gesture recognition for human computer interaction: A survey. *Artificial Intelligence Review*, 43(1), 1–54. doi:10.1007/s10462-012-9356-9
- Reddy, B. S., & Basir, O. A. (2010). Concept-based evidential reasoning for multimodal fusion in human-computer interaction. *Applied Soft Computing*, 10(2), 567–577. doi:10.1016/j.asoc.2009.08.026
- Rempel, D., Camilleri, M. J., & Lee, D. L. (2014). The design of hand gestures for human-computer interaction: Lessons from sign language interpreters. *International Journal of Human-Computer Studies*, 72(10/11), 728–735. doi:10.1016/j.ijhcs.2014.05.003 PMID:26028955
- Rosson, M. B., & Carroll, J. M. (2001). *Usability engineering: Scenario-based development of human-computer interaction*. San Francisco, CA: Morgan Kaufmann Publishers.
- Sebe, N., Cohen, I., Cozman, F. G., Gevers, T., & Huang, T. S. (2005). Learning probabilistic classifiers for human-computer interaction applications. *Multimedia Systems*, 10(6), 484–498. doi:10.1007/s00530-005-0177-4
- Servidio, R., Davies, B., & Hapeshi, K. (2016). Human-computer interaction in consumer behaviour. In *Mobile computing and wireless networks: Concepts, methodologies, tools, and applications* (pp. 1530–1549). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-8751-6.ch068
- Sharma, R. P., & Verma, G. K. (2015). Human computer interaction using hand gesture. *Procedia Computer Science*, 54, 721–727. doi:10.1016/j.procs.2015.06.085
- Takiguchi, K., Wada, T., & Toyama, S. (2008). Rhythm pattern of sole through electrification of the human body when walking. *Journal of Advanced Mechanical Design, Systems and Manufacturing*, 2(3), 429–440. doi:10.1299/jamdsm.2.429
- Tang, K., Chen, X., Zheng, W., Han, Q., & Li, P. (2015). A non-contact technique using electrostatics to sense three-dimensional hand motion for human computer interaction. *Journal of Electrostatics*, 77, 101–109. doi:10.1016/j.elstat.2015.07.006
- Tidwell, J. (2011). *Designing interfaces*. Sebastopol, CA: O'Reilly Media.

Wohl, M. J. A., Parush, A., Kim, H. S., & Warren, K. (2014). Building it better: Applying human-computer interaction and persuasive system design principles to a monetary limit tool improves responsible gambling. *Computers in Human Behavior*, 37, 124–132. doi:10.1016/j.chb.2014.04.045

Yeo, H. S., Lee, B. G., & Lim, H. (2015). Hand tracking and gesture recognition system for human-computer interaction using low-cost hardware. *Multimedia Tools and Applications*, 74(8), 2687–2715. doi:10.1007/s11042-013-1501-1

ADDITIONAL READING

Ahlrichs, C., Iben, H., & Lawo, M. (2015). Context-aware mobile and wearable device interfaces. In K. Curran (Ed.), *Recent advances in ambient intelligence and context-aware computing* (pp. 308–321). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-7284-0.ch019

Anishchenko, S. I., Osinov, B. A., & Shaposhnikov, D. G. (2012). Assessment of human head pose in human-computer interaction. *Pattern Recognition and Image Analysis*, 22(4), 541–545. doi:10.1134/S1054661812040025

Biswas, P. (2014). A brief survey on user modeling in human computer interaction. In *Assistive technologies: Concepts, methodologies, tools, and applications* (pp. 102–119). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-4422-9.ch006

Blazica, B., Vladusic, D., & Mladenec, D. (2013). A personal perspective on photowork: Implicit human-computer interaction for photo collection management. *Personal and Ubiquitous Computing*, 17(8), 1787–1795. doi:10.1007/s00779-013-0650-6

Botha, A. (2016). Framework to enhance the mobile user experience of goal orientated interactions. In *Human-computer interaction: Concepts, methodologies, tools, and applications* (pp. 583–612). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-8789-9.ch026

Cheng, S., & Liu, Y. (2012). Eye-tracking based adaptive user interface: Implicit human-computer interaction for preference indication. *Journal on Multimodal User Interfaces*, 5(1), 77–84. doi:10.1007/s12193-011-0064-6

Courgeon, M., & Clavel, C. (2013). MARC: A framework that features emotion models for facial animation during human-computer interaction. *Journal on Multimodal User Interfaces*, 7(4), 311–319. doi:10.1007/s12193-013-0124-1

Gupta, P. K., & Singh, G. (2015). A novel human computer interaction aware algorithm to minimize energy consumption. *Wireless Personal Communications*, 81(2), 661–683. doi:10.1007/s11277-014-2151-y

Heimgärtner, R. (2016). Intercultural user interface design. In *Human-computer interaction: Concepts, methodologies, tools, and applications* (pp. 1–34). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-8789-9.ch001

Herrera-Acuna, R., Argyriou, V., & Velastin, S. A. (2015). A Kinect-based 3D hand-gesture interface for 3D databases. *Journal on Multimodal User Interfaces*, 9(2), 121–139. doi:10.1007/s12193-014-0173-0

Hickling, E. M., & Bowie, J. E. (2013). Applicability of human reliability assessment methods to human-computer interfaces. *Cognition Technology and Work*, 15(1), 19–27. doi:10.1007/s10111-012-0215-x

- Hung, V. C., & Gonzalez, A. J. (2013). Context-centric speech-based human-computer interaction. *International Journal of Intelligent Systems*, 28(10), 1010–1037. doi:10.1002/int.21614
- Jetter, H. C., Reiterer, H., & Geyer, F. (2014). Blended interaction: Understanding natural human-computer interaction in post-WIMP interactive spaces. *Personal and Ubiquitous Computing*, 18(5), 1139–1158. doi:10.1007/s00779-013-0725-4
- Kwon, B. C., Hur, I., & Yi, J. S. (2014). A review of web-based dietary interventions: From the human-computer interaction practitioners perspective. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 24(3), 241–261. doi:10.1002/hfm.20371
- Lu, G., Shark, L. K., Hall, G., & Zeshan, U. (2012). Immersive manipulation of virtual objects through glove-based hand gesture interaction. *Virtual Reality (Waltham Cross)*, 16(3), 243–252. doi:10.1007/s10055-011-0195-9
- Morikawa, C., & Lyons, M. J. (2014). Design and evaluation of vision-based head and face tracking interfaces for assistive input. In G. Kouroupetroglou (Ed.), *Assistive technologies and computer access for motor disabilities* (pp. 180–205). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-4438-0.ch007
- Nataraja, P., & Raju, G. T. (2013). Quantitative influence of HCI characteristics in a blended learning system. *Education and Information Technologies*, 18(4), 687–699. doi:10.1007/s10639-012-9191-2
- Nielsen, J., & Budiu, R. (2012). *Mobile usability*. Berkeley, CA: New Riders Press.
- Novak, D., Mihelj, M., & Munih, M. (2012). Dual-task performance in multimodal human-computer interaction: A psychophysiological perspective. *Multimedia Tools and Applications*, 56(3), 553–567. doi:10.1007/s11042-010-0619-7
- Poole, E. S. (2013). HCI and mobile health interventions. *Translational Behavioral Medicine*, 3(4), 402–405. doi:10.1007/s13142-013-0214-3 PMID:24294328
- Ren, G., & O'Neill, E. (2013). 3D selection with freehand gesture. *Computers & Graphics*, 37(3), 101–120. doi:10.1016/j.cag.2012.12.006
- Saadé, R. G., Kira, D., & Otrakji, C. A. (2012). Gender differences in interface type task analysis. *International Journal of Information Systems and Social Change*, 3(2), 1–23. doi:10.4018/jissc.2012040101
- Seffah, A., & Taleb, M. (2012). Tracing the evolution of HCI patterns as an interaction design tool. *Innovations in Systems and Software Engineering*, 8(2), 93–109. doi:10.1007/s11334-011-0178-8
- Siebert, I., Bock, R., & Wendemuth, A. (2014). Inter-rater reliability for emotion annotation in human-computer interaction: Comparison and methodological improvements. *Journal on Multimodal User Interfaces*, 8(1), 17–28. doi:10.1007/s12193-013-0129-9
- Su, K. W., & Liu, C. L. (2012). A mobile nursing information system based on human-computer interaction design for improving quality of nursing. *Journal of Medical Systems*, 36(3), 1139–1153. doi:10.1007/s10916-010-9576-y PMID:20827569
- Thompson, N., & McGill, T. (2015). Affective human-computer interaction. In M. Khosrow-Pour (Ed.), *Encyclopedia of information science and technology* (3rd ed., pp. 3712–3720). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-5888-2.ch364
- Tran, B. (2016). Extended assistive technology: The impact of interactivity of human-computer interfaces on independence, employment, and organizations. In *Special and gifted education: Concepts, methodologies, tools, and applications* (pp. 223–253). Hershey, PA: IGI Global. doi:10.4018/978-1-5225-0034-6.ch010

KEY TERMS AND DEFINITIONS

Communication: The exchange of thoughts, messages, or information, as by speech, signals, writing, or behavior.

Computer: The device that computes, especially a programmable electronic machine that performs the high-speed mathematical operations.

Gesture: The expressive movement of the body, or something that is done to show a feeling.

Information: The data that can be stored in and retrieved from a computer.

Information Technology: The development, installation, and implementation of computer systems and applications.

Interaction: The situation in which two or more objects act upon one another to produce a new effect.

Software: The programs and routines for a computer or the program material for an electronic device which make it run.

Software Engineering: The process of making, testing, and documenting computer programs.

Technology: The scientific method and material used to achieve a commercial or industrial objective.

H

Interface Trends in Human Interaction, the Internet of Things, and Big Data

William J. Gibbs

Duquesne University, USA

INTRODUCTION

The Web and other networked-based services are primary vehicles for news and information dissemination. In 2008, Kohut, (2008, p. 21) reported that, “Thirty-seven percent of people today, including more than half of Internet users, obtain news online whereas ten years ago only 13% of the public and 35% of Internet users went online for news.” More recently, Conaghan (2015, p. 1) noted that greater than 93% of men and 92% of women between the ages of 25-44 who went online in August 2015 engaged with newspaper content and half of this online audience use mobile devices (smartphones or tablets) to get news. A majority of major news websites are finding that most of their traffic comes from mobile devices rather than from desktop computers (Mitchell, 2015). Due in large part to mobile connectivity, people can obtain news instantaneously and become aware of worldwide events at any time of day or in any location around the world. It is estimated that there are over 7 billion mobile-cellular subscriptions (ICT, 2015), which enable people to access the news wirelessly on mobile devices, making news and information services portable, personalized, and participatory (Purcell, Rainie, Mitchell, Rosenstiel & Olmstead, 2010). The transformations taking place in news and information services magnify questions regarding the influence networked-based services have on newsreaders (Santana, Livingstone, & Cho, 2011). Researchers indicate that media are not solely transmitters of information, but they influence the process of thought (Carr, 2008; Purcell et al., 2010).

Online news content is frequently represented on digital displays as a highly dynamic interface characterized by a proliferation of media and interactivity that supersedes what is found in traditional informational sources such as newsprint or television news. Digital interfaces or points-of-contact through which people experience news and information services have never been so diverse or transformative. They present complex visual landscapes comprised of and supported by multimedia, communications, and networking technologies. Pervasive worldwide, they afford people an unprecedented degree of functionality and access to news, information services, and other people. The actions or ways in which users interact with modern interfaces are diverse and include behaviors such as swiping, scaling, dragging scrolling, hovering, and flipping (Sundar, Bellur, Oh, Xu & Jia1, 2014). Interfaces are a foundational technology that has helped instigate tectonic shifts in news and information consuming behavior, journalistic reporting, and news preparation and distribution, the impact of which is not fully understood.

In this chapter, I examine trends in today’s news-orientated interfaces and the impact of digital interfaces on news consumption. Digital interfaces will be differentiated from traditional informational sources such as newspapers and television news. Additionally, I will explore several major characteristics or trends germane to today’s news interfaces and their implications for how people consume news and, more generally, for how they transform information services: a) rapid innovation, b) interactivity, c) social, d) standardization, e) scale, f) media convergence and, g) the Internet of Things and Big Data.

DOI: 10.4018/978-1-5225-2255-3.ch365

TRADITIONAL MEDIA FORMS

Access to the news and one's understanding of it are influenced by interface elements. When people have contact with news, news organizations preplan or design the visual, auditory, conceptual, and functional aspects of that experience or point-of-contact. The manifestation of this design comprises an interface intended to help people access news and derive meaning from it. For example, when reading a newspaper, the printed document, type, content organization, headings, writing style, the proximity of page elements, the surrounding context, groupings and placement, and page numbering establish a context that guides readers' attention and provides them information about how to use the newspaper to glean information. One can imagine how readers would fare if a paper suddenly removed all headlines, page numbers, table of contents, and used disparate type. Some authors contend that the typographical design of newspaper makes reading easier and enhances comprehension relative to news content published on the web (Shafer, 2011). The inherent attributes of newsprint and television news media greatly influence how people access and comprehend news as well as how news is reported. Pippas, Walter, Endres, and Tabatcher (2009), for example, report research showing that content recall for television and radio news was lower than recall of textual information. Santana et al. (2011) found that news consumption varied based on the frequency, duration, and visual content of a news presentation. Access to a news story and learning from it is directly affected by many contextual features that make up the social and behavioral state of the environment in which a person acquires news (DeFleur, Davenport, Cronin, & DeFleur, 1992, p. 1011).

Characteristics of Television and Newsprint

Television broadcasts provide a vastly different news gathering experience compared to newsprint

and online news. People adapt their behavior to accommodate the differences. A person may view a TV news broadcast in a room with other people or while engaging in some other activity. Television is a passive medium wherein viewers watch and listen as content gets delivered to them. They have limited physical engagement with the display interface as they endeavor to get news. A distinct separation exists between news content and the television interface. Broadcasts present short video-based stories sequenced linearly within a specified timeframe at a fixed location for on-air viewing. One's access to and the sequencing of such stories is controlled by the news organization. People can record TV broadcasts for archival purposes. Once recorded, the broadcast can be controlled by the individual.

Newspapers, unlike TV and radio broadcasts and online news, are tangible, highly portable, and can be easily archived by the reader. Although, today's mobile devices, web interfaces, and associated networking technologies, increase the portability of TV and radio broadcasts as well as online news. However, unlike newspapers, online content, while recordable, often requires software and hardware for recording or downloading. As people acquire news, there is a high degree of physical interaction with the printed medium or interface. The interface and content display are malleable - as people touch it, the display of content changes. Newspapers afford in-depth reading as well as active information search. Readers control the access to information. They navigate the printed document linearly or non-linearly, depending on their preference. News content is presented primarily as text and it is usually more in-depth compared to broadcasts that use visual and auditory stimuli to underscore messages. With newsprint, people bring established attitudes about what to expect and knowledge of newsprint conventions. Additionally, newsprint has a sense of permanence unlike online news, which is transient in that it appears and disappears without notice (Santana et al., 2011). Similarly, TV news broadcast are temporary unless recorded, but they

do have a scheduled broadcast times so they do not appear without notice. Moreover, compared to online news, newsprint often gives readers more informational cues about the magnitude of a story, which may assist readers in story selection (Santana et al., 2011).

Newspaper and television news organizations excel at providing consumers sophisticated interfaces (points-of-contact) that, by-and-large, engender meaningful and efficient access to news. As news organizations evolved, interfaces have become customary with designs grounded firmly in established conventions or rules for type, placement, content organization, and layout, among others. Conventions for newsprint, for instance, date back hundreds of years. Consequently, when people experience or have contact with news, they find much commonality in how different news organizations report, broadcast, or print it. Moreover, the interface is familiar. For the most part, news consumers, with minimal mental or physical effort, can effectively access news content in newspapers or on television any given day from almost any news organization. From a consumer perspective, the adherence to convention and ensuing commonalities among interfaces (points-of-contact) are advantageous because they make news accessible and potentially more comprehensible.

CHARACTERISTICS OF ONLINE NEWS INTERFACES

The interfaces or points-of-contact in the digital realm are fundamentally different from those of traditional newsprint and television. Now, between the user of a device and the information or service he/she seeks is a computational or digital interface encompassing the hardware and software components that users see, hear, touch, or talk to as they interact. Digital interfaces are becoming a dominant means of access as people retrieve news on websites, news feeds, or podcasts using computers, phones, or a host of other devices.

Correspondingly, broadband, broadcasting, and information technologies enable news networks to deliver multimedia content (voice, imagery, video, and computing) and to support multimodal human-computer interactions (e.g., touch, speech) in a single interface across many different types of devices.

News interfaces have emerged as amalgamations of visual and information design, multimedia content, tools, and functionality that afford unparalleled connections among content, services, and people. For instance, reading news stories, one can easily link to related stories, post comments, or participate in social network services. The interfaces of all major online networks give access to services, such as Twitter and Facebook. During many recent political upheavals and crises, news interfaces gave people instantaneous, global access to breaking newsworthy information by way of a single display.

Transformations in news and information services brought about by rapid and pervasive technological advancements impact, among other things, news literacy, the types of news sources being used for news, reporting approaches, and how people read and experience news. Today's communication devices and interfaces are diverse, complex, and seemingly ever evolving. Compared to newsprint and television broadcast, digital interfaces are distinct in many ways, several of which are presented in Table 1. In the following sections, characteristics or trends related to news interfaces are presented.

Rapid Innovation

Innovation in the realm of digital interfaces occurs rapidly. As new organizations discern industry trends, their interfaces change often (unexpectedly) and they have given increased prominence to personalized and social services (e.g., discussing news stories, contributing news), seemingly without regard to the needs of users or the service quality they encounter at the interface level. It is not uncommon for the entire design of a news

Table 1. TV, newspaper, digital interface characteristics

	TV	Newsprint	Digital
Access to content	Linear	Nonlinear	Nonlinear
Control of content presentation	News Organization	Reader	Reader/User
Content type	Non-static: video, audio, images, graphics, text, animation	Static: Text, images, graphics	Non-static: video, audio, images, graphics, text, animation
Scale	Fixed, clear delineation of beginning/end	Fixed, clear delineation of beginning/end	Not fixed, no beginning/end, little sense of scale
Content updating	Daily; continuous for 24 hour news	Daily	Continuous; dynamic
Distribution medium	Television	Printed paper	Multiple electronic devices
Interactivity	Low	Low-Moderate	High
Portability	Moderately portable, with player or devices	Highly portable	Portable with mobile devices
Viewport variations	Moderate	Moderate	High
Content Adaptation to Viewport	Low	Moderate	High
Archiving by user	Moderate: User records	User stores paper	Moderate
Tangibility	Low	High	Low
Functionality	Low	Low	High
Affords socialization	Low, tangential socialization	Low	High, virtual socialization
Big Data and Analytics	Low	Low	High

website to change with new, sometimes arcane features, functions, and controls added. The extent to which website interfaces support or impede users' access to news content, while organizations keep pace with industry trends, is unknown. The impact on people, news consumption, the user experience, and trust in an organization is also largely unknown. How and why decisions to change interfaces or adopt news interface technologies are made is critical to understanding emergent interfaces and their impact on news distribution and consumption. For example, targeting advertisements and stories to users based on their interests is common and seemingly done in the interest of improving the user experience. However, these efforts often impede the experience one has at the interface. Advertisements consume the screen, taking control of the interface even briefly. When reading a story, it is not uncommon for content to shift vertically to accommodate ads as the pages refresh, interrupting reading as one's eyes traverse lines of text. Video ads auto start often resulting in multiple audio tracks playing simultaneously.

The evolution and maturation of interface technologies greatly influence news consumers' information seeking behavior. Additionally, as interfaces change, becoming more sophisticated and technically complex, particularly back-end technologies, news organizations are faced with many new questions concerning interface design, development, and implementation and who or what drives interface technology decisions, the news organization, journalists, editors, advertisers, technology or technologists.

Interactivity

Digital interfaces represent a highly dynamic point-of-contact characterized by a proliferation of motion and static media and interactivity that supersedes what is found in either traditional newsprint or on television. A significant issue posed by hypermedia environments such as the web is effectively dealing with complexity and not overwhelming users with information and

functions (Vaughan & Dillon, 2006). While highly dynamic, feature-rich interfaces are pervasive, they may prove too complex for users. A premise of the cognitive information processing model of cognition holds that humans have limited cognitive capacity. Complex tasks performed with visually intensive and dynamic interfaces may consume cognitive resources and cause cognitive overload, diminishing meaningful processing (Smith & Ragan, 2005). Opgenhaffen and d'Haenens (2011) point out that news content with features such as multimedia, interactivity, and hypertext may be overly complex and consume more cognitive resources than users have available, resulting in cognitive overload. They found that "cognitive load plays an important role during information processing and that the form of the medium can impact this mental load. (p. 20) Moreover, Santana et al. (2011) reported that users are disinclined to engage with multimedia tools, preferring first the text of a story instead. Berry (1999) found that in terms of comprehension and information recall, multimedia did not make a difference.

To effectively interact with today's interfaces, one must understand a host of interaction conventions such as the magnifying glass icon indicating search or what is colloquially known as the hamburger icon representing hidden menu items. Icons comprise a visual language that is often misinterpreted and used inconsistently across sites or applications. The placement of interaction objects and their proximity to one another is also meaningful, as location in a display conveys purpose and hierarchy. For instance, websites designed for desktop computers typically orient primary navigation horizontally at the top of the page with secondary navigation located immediately below or oriented vertically on the left side. People come to know that these regions of the display serve as navigation. However, on responsive sites, the interface adjusts to the size of the viewport. Due to the limited screen space on mobile devices, navigation regions may get hidden or oriented differently compared to when the site is viewed on a desktop computer. Thus,

the placement of navigation and content alter depending on the device being used. Given that many people use multiple devices simultaneously, such differences at the interface add complexity and potentially diminish the user experience. In addition to the aforementioned factors, navigational structures vary. The New York Times home page, for instance, presents a hierarchical structure from which users drill-down to access content. USA Today presents a flat structure with which user traverse content sections. Among other things, these structures impact the user's journey through content, they determine the number of interactions necessary to reach a target, and ultimately influence how easy or difficult it is to access information. Navigation structure issues become especially critical on mobile devices with limit display area. Moreover, actions such as swiping, scaling, dragging, and tapping that must be perform to execute functions or to navigate may not be apparent to some users and can cause unintentional responses from the interface. For example, when using a mobile device, it is not uncommon for a user to tap or swipe an image intending to see additional detail only to land on a different page or section of the site. The small viewport and minimal cues about interaction functionality often contribute to a less than positive user experience.

Social

Interfaces and associated technologies connect people and resources in unprecedented ways. They enable people to contribute to and participate in the news. People are no longer solely news consumers but contributors as well. CNN.com, for example, has a section of its site devoted to iReports where people can take part in news by uploading compelling and/or urgent stories. People interact by posting comments or participating in discussions with other readers or the author. The interfaces of all major online networks give access to services, such as Twitter and Facebook. By way of a single display, interfaces provided people instantaneous, global access to breaking

newsworthy information. The aforementioned have implications for how content is written, the expectations of journalists to participate in discussions, and the role of citizen journalists.

Standardization

Viewports through which people access news interfaces vary widely, impacting the design of interfaces and how and where people use them. Today, people interact with large devices, small devices, touch screens, keyboards, and devices with extraordinary web browsing capabilities, and those with limited capabilities (Kadlec, 2012). On each device, the visual characteristics and functionality of the digital interface may be distinct, impacting usability and the user's experience with the news. While television monitors and newspapers vary in size, the digital viewports through which people access online news range from large to extremely small. News organizations must adjust the content and content structure dynamically to deliver news to these devices. For instance, in a TV broadcast, the news organization encodes the media according to pre-define standards (e.g., NTSC, Digital-TV) which set, among other things, the dimensions of the content to be displayed. Viewers watch content on displays of various sizes. For newsprint, the organizations use established paper dimensions to format content accordingly. Content delivered on the web gets displayed on viewports of multiple orientations and sizes. Moreover, if organizations adhere to current trends in responsive design whereby the structure and style of content responds to the display, then they must develop sites and applications so that content structure and style respond or are accessible on all types of devices.

Coupled with dynamic content, animation, video, device portability, viewport size and orientation affects how people read and browse news. It impacts how news organizations design interfaces and structure content. Content presented on the 52 inch monitor must be re-designed for a 2.5 inch display.

Scale

Interfaces provide unique indications of scale and amount, compared to more traditional informational sources. Newspaper and television broadcast media clearly delimit the beginning and end of news content. A newspaper represents a certain amount of news coverage for a specific period of time. The number of pages and text density help readers discern the breadth and depth of coverage. Even during 24 hour news broadcasts, viewers can tell where segments begin and end. Coverage of topics is sequential. One topic gets covered before another, with little deviation. Program segmentation, commentary, and commercials help people delineate the amount and range of coverage. For the most part, these same cues are absent in the digital realm. Because digital interfaces offer unprecedented access to content, people, and services, people have almost infinite possibilities to explore news. One might follow a link embedded in a news story, traversing countless news organizations worldwide. Thus, using digital interfaces, it is much more difficult for one to intuit where news coverage begins and ends and the amount of coverage, which likely influences how people read news and how news gets distributed.

Media Convergence

The distribution of multimedia and rendering it within a single interface is an important innovation. It has contributed to media convergence of newsprint and television media blurring the distinctions between traditional newsprint and television-orientated new organizations. "Individuals can watch video on a newspaper site such as *usatoday.com*, or read news articles on television-oriented site such as *CNN.com*, as well as participate in blogs about various topics" (Gibbs, 2008, p. 130). As Opgenhaffen (2011) points out, a convergence of media, technologies, and services has altered how news content is presented and the way people consume it. Convergence has instigated new types

of interfaces that effect communication as well as information access and dissemination. Similarly, mobile devices represent a convergence of services and functionality. Phone, texting, camera, GPS, multimedia entertainment, games, email are integrated in devices and have important implications for how news is obtain and even generated.

At the same time, there is a preponderance of major TV and newspaper websites, each with distinct information and graphical layouts designed to underscore the news organization's television or newsprint tradition and reporting methods. These differences are discernible and can impact how people use sites. Gibbs and Bernas (2009) found greater across-user variability in the viewing of newspaper homepages compared to TV-oriented homepages, which may indicate that TV-oriented sites, with concentrations of links, provide more direct paths to content.

Interaction, Internet of Things and Big Data

Unlike traditional forms of media, today's interfaces are characterized by the potential to collect usage data as a result of a proliferation of connected devices. According to Chang, Dong and Sun (2014), the Internet of Things (IOT) was initially referenced by Kevin Ashton who coupled Radio Frequency Identification (RFID) with the Internet (see Ashton, 2009). The IOT is a network of devices connected to the Internet through RFID technology and other sensing devices that communicate (Chang, Dong & Sun, 2014; Magrassi & Berg, 2002). Potentially every object in the world can be equipped with sensors and connected to the Internet (Fleisch, 2010; Chen, 2013). Rapid and continuous technological development has resulted in computing that is ubiquitous. Computational, sensing and display technologies have become integrated into work and everyday public settings (Valkanova, Jorda, & Vande Moere, 2015, p. 4). Computers are in appliances, offices, vehicles and even being worn by people. Computation and interaction are finding their way in walls, furni-

ture – they are also disappearing, becoming less obtrusive and noticeable. Interfaces and human-computer interactions are virtually inescapable. As the International Telecommunication Union (2005) indicates,

...by embedding short-range mobile transceivers into a wide array of additional gadgets and everyday items, enabling new forms of communication between people and things, and between things themselves. A new dimension has been added to the world of information and communication technologies (ICTs): from anytime any place connectivity for anyone, we will now have connectivity for anything. (p. 2)

The capacity to monitor the interactions people have with devices and to record, often in real-time, the digital traces of those interactions is advantageous (Valkanova, Jorda, & Vande Moere, 2015). Gibbs and Mckendrick, (2015, p. xxiv) citing Goes (2014) indicate, "Analyst, researchers, businesses, and organizations can now inquire about, explain and potentially predict human behavior patterns and trends using data compiled from sensor-equipped devices and the *Internet of Things*." Large amounts of interaction data can be collected and analyzed regardless of the interaction modality, duration, frequency, or the locality in which the event occurred. Moreover, devices that communicate with each other provide for ubiquitous communication and computing, affording a host of context-awareness services (e.g., home automation, environment monitoring) as people move through the environment (Jara et al., 2014).

These enormous amounts of data, commonly referred to as Big Data, are unique because of the *volume* or amount of data that can be accumulated; the *velocity* or real-time analysis that is possible; the *variety* of data, from various devices and sources (e.g., social media, content, behavior) and *veracity* or uncertainty of data due to variation in quality (Chen, Chiang, & Storey, 2012; Waller & Fawcett, 2013; Goes, 2014; Taylor, Schroeder, & Meyer, 2014; Gibbs & McKendrick, 2015). Unlike

other media forms such as newspaper and television, users only have to interact with the digital interfaces and their interactions are recorded – with digital interfaces user actions beget data and make analysis possible. Conversely, to amass data about television viewing habits or newspaper reading habits, the news and information services organization typically must contact news consumers who provide data. Self-reporting about usage behaviors can be problematic since people may be unable to articulate or even understand their own behavior. However, interaction data (behavior) from digital interfaces get recorded as interactions occur providing a clear representation of human-computer interactions.

News and information services organizations use Big Data to analyze behavioral patterns and interaction that can help determine ways to attract users and ultimately how to bring more traffic to a news website. In addition, data analytics can help reveal the type of content that gets the attention of readers and the type of material that retain their attention (Hiers, 2015). Not only are data and analytics used to increase traffic and to direct advertising, increasingly they are used to understand and potentially refine the user experience with digital interfaces and associated content. By tracking interactions, news organizations can make predictions about the type of content someone may find interesting or location data may be used to deliver newsworthy content based on one's location. The possibilities for the IOT, data and analytics are only just beginning to be explored but they have great implications for information services.

Solutions and Recommendations

Examination of interface elements on news websites and applications, suggests, to a degree, a mass production services model, wherein services are viewed as commodities and user and product are separate in production processes. For instance, while engagement at the interface has a role in establishing relations with users, users of major

news sites often encounter unexpected changes to the interface that dramatically affect their loyalty to the site, and how they browse and access to information. Some sites modify the interface without notifying or consulting users, thus separating services (products) and users as well as restricting user input. As Parker and Heapy (2006, p. 28) noted, we are more likely to remember something about the company brand from our interactions with the company than from looking at some element of a design work. It may be that users are more likely to remember the news organization by how the interface supports their interactions and needs. This is particularly germane to online information service interfaces because they serve as the sole enabler of relationships between news consumers and information service providers.

When the interface of a news site changes significantly without notice, rapid innovation and technology trends may be the impetus rather than usability and user experience (UX) design. While many news organizations espouse UX principles, all too often interfaces alter appreciably leaving users surprised and confused. Adherence to UX precepts can likely afford users enhanced experiences at the interface, while at the same time building customer loyalty and commitment to the news organization.

UX design is concerned with the total effect felt by users as they interact with a system, device or product. It encompasses usability and usefulness as well as what the user feels internally (Hartson & Pyla, 2012, p. 19). Several precepts from UX and interface design are germane to the role of today's visually intensive and interactive news interfaces. For example, Davidson, Dove, and Weltz (1999) proposed seven high-level principles to help users formulate accurate mental models of an interface or product (Mathis, 2011): a) simplicity, b) familiarity, c) recognition, d) flexibility, e) feedback, f) safety, g) affordance. Similarly, Shneiderman (2010) provided eight rules of interface design: a) strive for consistency; b) cater to universal usability; c) provide informative feedback; d) design dialogs to yield closure; e) prevent errors; f) permit

easy reversal of actions; g) support internal locus of control; h) reduce short term memory load. As innovations occur and interfaces change, adherence to such UX principles becomes important, especially given that people tend not to engage with multimedia tools and spend limited amounts of time learning new controls or functionally on sites. In this regard, Big Data and analytics can prove useful. Tracking when users abandoned content, or what they clicked or touched or the sentiment of discussions provides valuable data that can help improve the user experience. The IOT and mobile devices enables information service organizations, by way of data, to more fully understand users and their behavior at the interface. These data and corresponding analytics must be used to inform interface design and to ultimately create a better user experience - so that all people can get news and information and be able to comprehend it.

FUTURE RESEARCH DIRECTIONS

Many news organizations today not only provide newsworthy stories but they also attempt to build dynamic communities by establishing connections among sets of information and a diversity of people. These communities or environments and supporting interfaces provide an invaluable source of data about human information-seeking behavior and preferences previously unavailable to news-gathering organizations. At the interface level, organizations are increasingly using innovative digital data collection and analytics to observe and understand trends in order to improve their online practices and the user experience. These data can help organizations monitor and better understand usage patterns, reader attention, interests, and loyalty. Analytics and digital data collection methods will continue to be used to inform decisions regarding news content, interface design, user experience, and news consumption, reporting, and distribution. They will likely pro-

vide additional insights regarding how the design of interfaces and interface elements support or impede users' access to information. User research, analytics and data will play an ever increasing role in online news and information services. There will be continued innovation and greater focus on methodologies for studying online news and information services and how these data can be utilized to refine interface designs and improve services overall.

CONCLUSION

In this chapter, I examined trends in today's news-orientated interfaces and the impact of digital interfaces on news consumption. I discussed how digital interfaces differ from traditional informational sources such as newspapers and television news. I also examined characteristics of today's news interfaces and their implications for how people consume news.

The nature and proliferation of today's interactive multimedia interfaces add complexity, impeding some user access to and comprehension of the news and/or reshaping how people access news. I suggested that UX principles provide a useful framework for examining and building new types of interfaces and that data and analytics can play an increasing important role in this endeavor. All too often it seems that technology innovation, not usability and user experience is the driving impetus for interface design decisions.

REFERENCES

- Arr, N. (2008). Is Google making us stupid - Why you can't read the way you used to? *Atlantic Monthly*, 302(1), 56-63.
- Ashton, K. (2009, June 22). That 'internet of things' thing: in the real world things matter more than ideas. *RFID Journal*. Retrieved from <http://www.rfidjournal.com/articles/view?4986>

- Berry, L. (1999). *Comprehension and recall of Internet news: A quantitative study of web page design*. Paper presented at the annual meeting of the Association for Education in Journalism and Mass Communication (AEJMC). New Orleans, LA.
- Chang, Y., Dong, X., & Sun, W. (2014). Influence of characteristics of the internet of things on consumer purchase intention. *Social Behavior and Personality*, 42(2), 321–330. doi:10.2224/sbp.2014.42.2.321
- Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *Management Information Systems Quarterly*, 36(4), 1165–1188.
- Chen, M. (2013). Towards smart city: M2M communications with software agent intelligence. *Multimedia Tools and Applications*, 67(1), 167–178. doi:10.1007/s11042-012-1013-4
- Conaghan, J. (2015). *Newspaper digital audience grew twice as fast as the internet in past 12 months*. Retrieved December 6, 2015 from Newspaper Association of America website: <http://www.naa.org/Trends-and-Numbers/Newspaper-Websites/Newspaper-Web-Audience.aspx>
- Davidson, M. J., Dove, L., & Weltz, J. (1999). *Mental models and usability*. Technical report, DePaul University, Chicago. Retrieved July 27, 2013 from <http://www.lauradove.info/reports/mental%20models.htm>
- DeFleur, M. L., Davenport, L., Cronin, M., & DeFleur, M. (1992). Audience recall of news stories presented by newspaper, computer, television and radio. *The Journalism Quarterly*, 69(4), 1010–1022. doi:10.1177/107769909206900419
- Fleisch, E. (2010). What is the internet of things? An economic perspective. *Economics, Management, and Financial Markets*, 5(2), 125–157.
- Gibbs, W. (2008). Examining Users on news provider web sites: A review of methodology. *Journal of Usability Studies*, 3(3), 129–148.
- Gibbs, W., & Bernas, R. S. (2009). Visual attention in newspaper versus tv-oriented news websites. *Journal of Usability Studies*, 4(4), 147–165.
- Gibbs, W. J., & McKendrick, J. (2015). *Contemporary research methods and data analytics in the news industry*. Hershey, PA: IGI Global; doi:10.4018/978-1-4666-8580-2
- Goes, P. B. (2014). Big data and is research. *Management Information Systems Quarterly*, 38(3), iii–viii.
- Hartson, R., & Pyla, P. S. (2012). *The UX book: Process and guidelines for ensuring a quality user experience*. Waltham, MA: Morgan Kaufman.
- Hiers, M. (2015). *Analytics and the newspaper industry: What matters and why?* Newspapers & Local Media, Publishers. Retrieved November 6 2015 from <http://blog.realmatch.com/news-publishers/analytics-newspaper-industry-matters/>
- International Telecommunication Union. (2005). *ITU Internet Report 2005: The Internet of Things*. Geneva: ITU. Retrieved December 6, 2015 from <https://www.itu.int/net/wsis/tunis/newsroom/stats/The-Internet-of-Things-2005.pdf>
- International Telecommunication Union. (2015). ITU ICT Facts and Figures –. *WORLD (Oakland, Calif.)*, 2015. Retrieved from <https://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFacts-Figures2015.pdf>
- Jara, A. J., Lopez, P., Fernandez, D., Castillo, J. F., Zamora, M. A., & Skarmeta, A. F. (2014). Mobile digcovery: Discovering and interacting with the world through the internet of things. *Personal and Ubiquitous Computing*, 18(2), 323–338. doi:10.1007/s00779-013-0648-0

- Kadlec, T. (2012). *Implementing responsive design*. Berkeley, CA: New Riders Publishing.
- Kohut, A. (2008). *Audience segments in a changing news environment: Key news audiences now blend online and traditional sources*. Pew Research Center Biennial News Consumption Survey. Washington, DC: The Pew Research Center for the People & Press.
- Margrassi, P., & Berg, T. (2002). *A world of smart objects: The role of auto-identification technologies*. Gartner Research Report R-17-2243.
- Mathis, L. (2011). *Designed for use: Create usable interfaces for applications and the web*. Pragmatic Programmers, LLC.
- Mitchell, A. (2015). *State of the news media 2015*. Retrieved November 6 2015 from Pew Research Center: Journalism & Media: <http://www.journalism.org/2015/04/29/state-of-the-news-media-2015/>
- Opgenhaffen, M. (2011). Multimedia, interactive and hypertextual features in divergent online news platforms: An exploratory study of Flemish online news. *First Monday*, 16(3). doi:10.5210/fm.v16i3.2826
- Opgenhaffen, M., & d'Haenens, L. (2011). The impact of online news features on learning from news: A knowledge experiment. *International Journal of Internet Science*, 6(1), 8–28.
- Parker, S., & Heapy, J. (2006). *The Journey to the interface: How public service design can connect users to reform*. London: Demos Magdalen House.
- Pipps, V., Walter, H., Endres, K., & Tabatcher, P. (2009). Information recall of Internet news: Does design make a difference? A pilot study. *Journal of Magazine and New Media Research*, 11(1). Retrieved from <http://aejmcmagazine.arizona.edu/Journal/Fall2009/Pipps.pdf>
- Purcell, K., Rainie, L., Mitchell, A., Rosenstiel, T., & Olmstead, K. (2010). *Pew Internet: Understanding the participatory new consumer*. Retrieved July 1, 2011 from <http://www.pewInternet.org/Reports/2010/Online-News.aspx>
- Santana, A., Livingstone, R., & Cho, Y. (2011). *Medium matters: Newsreaders' recall and engagement with online and print newspapers*. Paper presented at the annual meeting of the Association for Education in Journalism and Mass Communication. St. Louis, MO.
- Shafer, J. (2011). Print vs. online: The ways in which old-fashioned newspapers still trump online newspapers. *Slate*. Retrieved July 27, 2013 from http://www.slate.com/articles/news_and_politics/press_box/2011/08/print_vs_online.single.html
- Shneiderman, B. (2010). *Designing the user interface: Strategies for effective human-computer interaction*. 1995. *Information Seeking in Electronic Environments*. Boston, MA: Pearson.
- Smith, P. L., & Ragan, T. J. (2005). *Instruction design*. Hoboken, NJ: John Wiley & Sons, Inc.
- Sundar, S. S., Bellur, B., Oh, J., Xu, Q., & Jia, H. (2014). User experience of on-screen interaction techniques: An experimental investigation of clicking, sliding, zooming, hovering, dragging, and flipping. *Human-Computer Interaction*, 29(2), 109–152. doi:10.1080/07370024.2013.789347
- Taylor, L., Schroeder, R., & Meyer, E. (2014). Emerging practices and perspectives on big data analysis in economics: Bigger and better or more of the same? *Big Data & Society*, 1(2), 1–10. doi:10.1177/2053951714536877
- Valkanova, N., Jorda, S., & Vande Moere, A. (2015). Public visualization displays of citizen data: Design, impact and implications. *International Journal of Human-Computer Studies*, 81, 4–16. doi:10.1016/j.ijhcs.2015.02.005

Vaughan, M., & Dillon, A. (2006). Why structure and genre matter for users of digital information: A longitudinal experiment with readers of a web-based newspaper. *International Journal of Human-Computer Studies*, 64(6), 502–526. doi:10.1016/j.ijhcs.2005.11.002

Waller, M. A., & Fawcett, S. E. (2013). Data science, predictive analytics, and big data: A revolution that will transform supply chain design and management. *Journal of Business Logistics*, 34(2), 77–84. doi:10.1111/jbl.12010

ADDITIONAL READING

Byrne, M. B., John, B. E., Wehrle, N. S., & Crow, D. C. (1999). The tangled web we wove: taxonomy of WWW use. In *Proceedings of ACM CHI 99 Conference on Human Factors in Computing Systems* (Volume 1, pp. 544–551). doi:10.1145/302979.303154

Card, S., Pirolli, P., Van Der Wage, M., Morrison, J., Reeder, R. W., Schraedley, P. K., & Boshart, J. (2001). Information scent as a driver of web behavior graphs: Results of a protocol analysis method for Web usability. In *Proceedings of CHI 2001* (pp. 498–505). doi:10.1145/365024.365331

Carroll, J. M. (Ed.). (2002). *Human-computer interaction in the new millennium*. New York: Addison-Wesley.

Catledge, L. D., & Pitkow, J. E. (1995). Characterizing browsing strategies in the world-wide web. *Computer Networks and ISDN Systems*, 27(6), 1065–1073. doi:10.1016/0169-7552(95)00043-7

Chi, E. H., Pirolli, P., Chen, K., & Pitkow, J. (2001). Using information scent to model user information needs and actions and the Web. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Seattle, Washington, United States). CHI '01 (pp. 490–497). New York, NY: ACM. doi:10.1145/365024.365325

Cypher, A. (1986). The structure of users' activities. In D. A. Norman & S. W. Draper (Eds.), *User Centered System Design* (pp. 243–263). New Jersey: Lawrence Erlbaum Associates.

Halvorson, K. (2010). *Content Strategy for the Web*. Berkeley, California: New Riders Publishing.

Jacob, R. J. K., & Karn, K. S. (2003). Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. In J. Hyönä, R. Radach, & H. Deubel (Eds.), *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 573–605). Amsterdam: Elsevier; doi:10.1016/B978-044451020-4/50031-1

Lenhart, A. (2009). Adults and social network websites. Pew Internet & American life project. Retrieved July 1, 2009, from <http://www.pewInternet.org/Reports/2009/Adults-and-Social-Network-Websites.aspx>

Mayer, R. E., & Sims, V. K. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology*, 86(3), 389–401. doi:10.1037/0022-0663.86.3.389

Norman, D. (2002). *The Design of Everyday Things*. New York: Basic Books.

Ostrow, J. (2009). Amateur web reporters shame networks (p. A.13). Denver Post, Denver, Colorado: June 25. Retrieved July 3, 2009, from http://www.denverpost.com/headlines/ci_12683997

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372–422. PubMed doi:10.1037/0033-2909.124.3.372

Riley, K. C. M., Meilich, O., & Pierson, J. (1998). Community or colony: The case of online newspapers and the web. *JCMC*, 4(1), September. Retrieved, July 3, 2009, from <http://jcmc.indiana.edu/vol4/issue1/keough.html>

Sharp, H., Rogers, Y., & Preece, J. (2012). *Interaction Design: Beyond Human-Computer Interaction*. Hoboken, New Jersey: John Wiley & Sons, Inc.

Spiro, R. J., Feltovich, P. J., Jacobson, M. J., & Coulson, R. L. (1991). Cognitive flexibility, constructivism, and hypertext: Random access instruction for advanced knowledge acquisition in ill-structured domains. *Educational Technology*, 31, 24–33.

Tidwell, J. (2011). *Designing Interfaces*. Sebastopol, CA: O'Reilly.

Unger, R., & Chandler, C. (2009). *A project guide to UX design*. Berkeley, California: New Riders Publishing.

Usability Professionals' Association. (2008). What is usability? Retrieved July 15, 2009, from http://www.upassoc.org/usability_resources/about_usability/definitions_of_usability.html

Van Duyne, D., Landay, J., & Hon, J. (2007). *The Design of Sites*. New York: Addison-Wesley.

Ware, C. (2013). *Information visualization: Perception for design*. Waltham, MA: Morgan Kaufman.

White, R. W., & Drucker, S. M. (2007). Investigating behavioral variability in web search. In *Proceedings of the 16th international Conference on World Wide Web* (Banff, Alberta, Canada, May 08 - 12, 2007). WWW '07 (pp. 21-30.). New York, NY: ACM. doi:10.1145/1242572.1242576

KEY TERMS AND DEFINITIONS

Digital Interface: A digital interface is the medium through which humans interact with

computers. Interfaces represent an amalgamation of visual, auditory, and functional components that people see, hear, touch, or talk to as they interact with computers (digital devices).

Human-Computer Interaction and Learning: Refers to interactional processes between learners and computational devices and the extent to which that interaction supports or impedes learning.

Interface: A mechanism that mediates the conversation between a person and a product or system. An interface is a means to make the conversation between the human and machine or product easier.

Interface Design: The process of designing the visual, metaphorical, functional aspects of a product or system. The process of interface design begins with understanding people, the tasks they perform, and the goals they are trying to reach. Based on this information, interface designers create the medium to help users interact with products or systems so they may reach their goals.

Responsive Design: Designing web content in such a way that it responds to any device on which it is viewed. Working from the content structure, designers design websites so when they are viewed on varying sized displays, users can easily access the information content.

Usability: A process of determining how effectively and efficiently people are able to use products or systems created by designers. A core aspect of usability is having people use the product or interface and examine how easy or difficult it is to use. Based on these data, designers make refinements to the system.

User Experience Design: Concerned with the total effect or effects felt by users as they interact with a system, device or product. It encompasses usability and usefulness as well as what the user feels internally (Hartson & Pyla, p. 19, 2012).

Internet Addiction in Context



Petra Vondrackova

Charles University in Prague, Czech Republic

David Šmahel

Masaryk University – Brno, Czech Republic

INTRODUCTION

Internet addiction can be defined as the overuse of the Internet such that it leads to the impairment of an individual's psychological state (both mental and emotional), as well as their scholastic, occupational, and social interactions (Beard & Wolf, 2001). Young (1998) describes Internet addiction as any online-related compulsive behavior that completely dominates the addict's life, interferes with normal living, and causes severe stress to family, friends, loved ones, and one's work environment.

From a historical perspective, in 1996, the American clinical psychologist Kimberley Young (1998) published the results of a two-year study of Internet behavior and misuse, and was the first to place the phenomenon of Internet addiction in a clinical context. This study received widespread public attention, and popular and professional debate. Since the 1990s, Internet addiction has attracted the attention of professionals around the world, mainly from developing and developed countries with widespread Internet access. Some countries, such as the United States, China, South Korea, and Germany, even began developing the first specialized centers focused solely on this phenomenon (Block, 2008). Nevertheless, inconsistencies regarding the terminology still exist. Usually the term "Internet addiction" is used. However, some researches do not agree with using the term "Internet addiction" and use several others, such as "Internet addiction disorder" (Chou, Condron, & Belland, 2005), "compulsive Internet use" (Black, Belsare, & Schlosser, 1999),

"Internet pathological use" (Davis, 2001), "problematic Internet use" (Caplan & High, 2011), and "Internet dependency" (Wang, 2001), each of which reflects a slightly different conception of this behavior. In this text we will use the term "Internet addiction" as an umbrella term for all of these term modifications.

In this chapter, we will first present a short historical overview of the Internet addiction phenomenon and its place in the context of mental health. We will then introduce the contributions of major researchers who focused on defining its core components, designing measurement scales and diagnostic criteria. Furthermore we will focus on the main areas of research in this field: the major surveys regarding prevalence rates and the correlates of Internet addiction. In the last section, we will introduce basic approaches to the treatment of Internet addiction.

The term "Internet addiction" is very broad. It refers to different kinds of overuse of the Internet. Starcevic (2013) stated that the term "Internet addiction" should be abandoned because individuals do not usually get addicted to the Internet in general, but to specific online activities; therefore, being addicted to the Internet implies addiction to a delivery mechanism or to a medium for achieving something. He states that it would be more accurate and appropriate to refer to specific activities that are presumed to be addictive. Some of the most typical online activities related to Internet addiction include online gaming (typically but not only Massive Multiplayer Online Role-Playing Games - MMORPGs), excessive online communication (email, chat rooms, and social

DOI: 10.4018/978-1-5225-2255-3.ch366

networking), cybersex activity overload (visiting online pornographic sites and initiating cybersex relationships), and online gambling (betting via the Internet) (Subrahmanyam & Smahel, 2011). Due to limited space, this chapter does not focus specifically on any of these activities; it focuses on Internet addiction in general.

BACKGROUND

Internet Addiction and Diagnostic Mental Disorder Manuals

A majority of experts include Internet addiction together with pathological gambling, compulsive shopping, sex addiction, and eating disorders in the category of non-substance or behavioral addictions. Behavioral addictions display the basic elements of addictive behavior in relation to certain activities and have similarities with drug addiction, specifically with respect to the genetic (Potenza, 2006), neurobiological (Han, Kim, Lee, & Renshaw, 2012), personal (Potenza, 2006), and clinical characteristics (Grant, Brewer, Potenza, 2006). All of the above-mentioned addictions are included in the diagnostic manuals DSM-5 or the ICD-10 (both are explained in the Key Term and Definitions section), except for compulsive shopping and Internet addiction. Internet addiction was considered for inclusion in the official DSM-5 diagnoses, but, in the end, was left it out. However, Internet gaming disorder (the compulsive playing of online games) is classified in the DSM-5 appendix so as to encourage additional research (APA, 2013). According to Pies (2009), the main reasons against Internet addiction's inclusion in the DSM-5's list of mental disorders are that symptoms of Internet addiction are likely to be the symptoms of other disorders, such as depression or obsessive-compulsive disorders, and that creating a separate category would further expand an already fast-growing list of supposed "disorders" and thus undermine the public's trust in psychiatric

diagnosis. In 2018, the 11th revision of the ICD should be published, but the revision discussion was started in 2012. The ICD-11 Working Group on Obsessive-Compulsive and Related Disorders noted that Internet addiction is a heterogeneous condition, and that use of the Internet may, in fact, constitute a delivery system for various forms of impulse control dysfunctions (e.g., pathological game playing or pornography viewing). In their view, based on the limited current data, it would therefore seem premature to include it in the ICD-11 (Grant et al., 2014).

Diagnostics of Internet Addiction

In the last few years, there have been several proposals for diagnostic criteria for Internet addiction. An absence of a consensus on the nature of Internet addiction has resulted in various authors emphasizing different diagnostic criteria. In most cases, although some researchers have added or removed individual elements, most researchers have adapted the DSM pathological gambling criteria for identifying Internet addiction. The best-known diagnostic criteria of Internet addiction were proposed by Young (1998). To confirm the presence of Internet addiction, five or more of the following points must be present: (1) the person is preoccupied with the Internet, thinks about online activities, or anticipates the next session on the Internet; (2) the person needs to use the Internet for increasing amounts of time in order to achieve satisfaction; (3) the person has made unsuccessful efforts to control, cut back, or stop Internet use; (4) the person is restless, moody, depressed, or irritable when attempting to cut down or end Internet use; (5) the person stays online longer than originally intended; (6) the person has jeopardized or risked the loss of a significant relationship, job, or educational or career opportunity because of Internet use; (7) the person has lied to family members or others to conceal the extent of their involvement with the Internet; and (8) the person uses the Internet

as a way to escape from problems or to relieve a dysphoric mood (e.g., feelings of helplessness, anxiety, and depression).

Griffiths (2000) proposed the following six basic components of Internet addiction, where a high score in all six dimensions suggest that the person is addicted: (a) salience: When the particular online activity turns out to be the most important activity in the person's life and governs their thinking and behavior; (b) mood modification: a subjective experience influenced by the pursued online activity; (c) tolerance: the process whereby increasing amounts of the particular online activity are required to achieve the former effects; (d) withdrawal symptoms: unpleasant feelings, states, and/or physical effects that come about when the specific online activity is discontinued or suddenly limited; (e) conflict: disagreements between the addicts and those around them, or from within the individuals themselves associated with the online activity; and (f) relapse: a tendency for repeated decline into earlier usage patterns of the online activity.

Inspired by the aforementioned core components and diagnostic criteria, experts have constructed various methods for assessing Internet addiction. Available tools can be divided into two categories: scales measuring generalized Internet addiction and scales measuring specific online addictions, especially problematic online gaming, addiction to cybersex, and social networking sites (Weinstein, Feder, Rosenberg, & Dannon, 2014). The above-mentioned first effort to identify Internet addiction was made by Young (1998), who developed a brief diagnostic questionnaire based on the DSM-IV criteria of pathological gambling. There are several tools measuring generalized Internet addiction, such as CIAS - Chen Internet Addiction Scale (Ko, Yen, Chen, Chen, & Yen, 2005) and CIUS - Compulsive Internet Use Scale (Davis, Flett, & Besser, 2002). Different scales were developed for measuring online gaming addiction, like GAS - Game Addiction Scale (Lemmens, Valkenburg, & Peter, 2009) and, for Internet sex addiction, ISST - Internet Sex Screening Test

(Carnes, Green, & Carnes, 2010). Wölfling, Beutel, and Müller (2012) introduced a standardized diagnostic tool for assessing Internet addiction in clinical settings – the Assessment of Internet and Computer Game Addiction - Checklist (AICA-C), which is the only tool that has been tested in a clinical population.

Prevalence Rates of Internet Addiction

Many surveys focusing on the prevalence of Internet addiction have been carried out, resulting in prevalence rates that vary widely from study to study. The problem with these surveys is that various definitions of Internet addiction were employed, different assessment methods and an unambiguous recall period for the presented questions were used, and the number and character of the surveys' participants also differed. In order to determine prevalence, many studies used samples of high school and college students (Cao & Su, 2007; Zhang, Amos, & McDowell, 2008) and some used Internet-based surveys (Wang, 2001; Young, 1998). In studies where representative samples of the general population were selected, the prevalence rates range from 1% in Germany (Rumpf et al., 2014) to 3.4% in the Czech Republic (Šmahel, Vondráčková, Blinka, & Godoy-Etcheverry, 2009).

Regarding the prevalence of Internet addiction across countries in the world, this is a difficult comparison to make due to the different methodology used in individual studies. But it seems that the highest rates are in Asian countries, such as China, South Korea, and Taiwan, where estimates are about 6-15% (Lin, Ko, & Wu, 2011; Ni, Yan, Chen, & Liu, 2009), whereas in Europe and North America, the estimates are lower, specifically 1-5% (Durkee et al., 2012; Šmahel & Blinka, 2012; Villella et al., 2011). Cheng and Li (2014) conducted a meta-analysis of Internet addiction studies in 31 nations across seven world regions and their correlation with quality of life. The highest average prevalence was in the Middle East (studies from

Iran, Israel, Lebanon, and Turkey) with 10.9%. The lowest was in Northern and Western Europe (studies from Austria, Estonia, France, Germany, Ireland, Norway, Sweden, and the United Kingdom) with 2.6%. In this meta-analysis, the average prevalence for Asia (studies from China, Hong Kong, India, South Korea, and Taiwan) was 7.1%. A direct comparison between different studies is impossible, but it seems that the prevalence is higher in Eastern than in Western countries.

Correlates of Internet Addiction

Research shows that Internet addiction is correlated with several biopsychosocial factors and patterns of Internet use. These can be divided into psychopathological factors, personality factors, physiological factors, sociodemographic factors, environmental factors, and patterns of Internet use. The following variables were identified as risk factors for Internet addiction:

1. **Psychopathological Factors:** ADHD, depressive and anxiety disorders, social phobia (Alavi et al., 2012), substance use (Ko, Yen, Yen, Chen, & Chen, 2012), obsessive compulsive symptoms (Jang, Hwang, & Choi, 2008), and conduct behavior (Müller et al., 2015);
2. **Personality Factors:** Low self-esteem (Sariyska et al., 2014), hyperactivity and impulsivity (Wu et al., 2013), high novelty seeking and low reward dependence (Dalbudak et al., 2015), introversion, low conscientiousness, agreeableness, high neuroticism/low emotional stability (Kuss, Shorter, van Rooij, van de Mheen, & Griffiths, 2014), hostility (Alavi et al., 2012), a low level of self-control and self-regulation (Blachnio & Przepiorka, 2015), and low social skills (Caplan & High, 2011);
3. **Physiological Factors:** Stronger blood volume pulse and respiratory response and weaker peripheral temperature (Lu, Wang, &

Huang, 2010), and dysfunctional activation of the supplement motor area for response inhibitions (Chen et al., 2015);

4. **Sociodemographic Factors:** Being male (Ha & Hwang, 2014), family economic disadvantage (Shek & Yu, 2016), low age at initial use of the Internet (Wang et al., 2013), and having divorced parents (Li, Garland, & Howard, 2014);
5. **Environmental Factors:** Loneliness and stress (Alavi et al., 2012), family disharmony (Shek & Yu, 2012), child-parent conflict (Li et al., 2014), affiliation with peers who have lower levels of social acceptance, and youth situated in a class with higher levels of Internet addiction (Zhou & Fang, 2015); and
6. **Patterns of Internet Use:** A high number of hours spent online (Kuss et al., 2013), engagement in different video games (Donati, Chiesi, Ammannato, & Primi, 2015), and excessive weekend Internet use (Xu et al., 2012)

Most of the mentioned factors were researched in isolated studies, which is also a limitation for this knowledge.

Treatment of Internet Addiction

With the increasing popularization of Internet addiction, health professionals have begun to publish treatment approaches for this target group (Orzack, Voluse, Wolf, & Hennen, 2006; van Rooij, Zinn, Schoenmakers, & van de Mheen, 2012; Young, 2007). There is an agreement that the main treatment approach should be psychosocial, especially psychotherapy, which should be supplemented by pharmacotherapy in more serious cases of Internet addiction.

Many psychotherapeutic approaches can be used for the treatment of Internet addiction, as is the case in other mental disorders, and each of them stresses different aspects of the problem.

There are three main treatment approaches:

1. The Cognitive Behavior Approach.

Cognitive behavior therapy sees Internet addiction as a learned way of thinking and behavior which helps the individuals to cope with excessive stress and fulfill their needs. Clients in therapy are encouraged to focus on and identify the thoughts and behavior which lead to the addictive behavior and replace them with adaptive ones (Hall & Parsons, 2001). The treatment programs usually last from two weeks (Koo, Wati, Lee, & Oh, 2011) to three months (van Rooij et al., 2012). The main pioneer of CBT in the treatment of Internet addiction is Young, who formulated her experiences in the treatment model CBT-IA (Young, 2011).

2. The Psychodynamic Approach.

The majority of papers describing the use of the psychodynamic approach on Internet addiction have the character of case studies (Toronto, 2009; Vondráčková, 2011) and, so far, have not proven effective on a large number of clients. The basic assumption of the psychodynamic approach agrees on the fact that addictive behavior serves as compensation or the possibility to escape from/dissociate unpleasant states that the individual experiences in relation to the world (Toronto, 2009). Therefore, the therapist helps the client to become aware, accept, and integrate negative feelings and thoughts connected with these unpleasant states.

3. Family Interventions.

Young (1999) stresses the importance of family interventions, especially when the Internet addiction negatively influences the family relationships. The involvement of family members has two basic goals: the use of family resources and strength to change the addictive behavior and in the decreased influence of the addiction on the family and the individual (Kaufman & Yoshioka, 2004).

The assessment of the quality of the Internet addiction treatment is yet unclear. King, Delfabro, Griffiths, and Gradisar (2011) published a systematic review to assess the clinical trials of Internet addiction treatment and concluded that the current studies have several key limitations in methodology, such as in the diagnosis of Internet addiction, the lack of randomization and the adequate control of comparison groups, and insufficient information about sample and effect sizes.

CONCLUSION

This chapter presented up-to-date studies focused on the diagnostics, prevalence, correlates, and treatments of Internet addiction. We presented several tools for the identification of Internet addiction, although only one of them – AICA-C – was standardized in a clinical setting (Wölfling et al, 2012). The development of more standardized tools is needed for the future. The prevalence rates of Internet addiction were identified in representative studies between 1-3.4% among Western countries and 6-15% among Eastern countries, which is probably still much less than the anecdotal evidence presented by the media.

FUTURE RESEARCH DIRECTIONS

The prevalence studies on representative samples of the population are very rare. We have most of our knowledge about prevalence from adolescents and university students. This is also a goal for future studies. We presented several factors which are correlated with higher Internet addiction: psychopathological, personality, physiological, sociodemographic, environmental, and patterns of Internet use. Most of these factors were revealed in cross-sectional studies which are mostly unable to reveal casual relationships between variables: future studies require research with longitudinal perspectives. We introduced

three basic approaches to the treatment of Internet addiction: the cognitive-behavioral approach, the psychodynamic approach, and family interventions. But the treatment of Internet addiction and its research are both in their infancy and future studies should reveal proven short- and long-term treatment strategies for Internet addicts.

ACKNOWLEDGMENT

Petra Vondráčková and David Šmahel acknowledge the support of Charles University in Prague and its program PRVOUK P03/LF1/9 and the Faculty of Informatics, Masaryk University.

REFERENCES

- Alavi, S. S., Ferdosi, M., Jannatifard, F., Eslami, M., Alaghemandan, H., & Setare, M. (2012). Behavioral Addiction versus Substance Addiction: Correspondence of Psychiatric and Psychological Views. *International Journal of Preventive Medicine*, 3(4), 290–294. PMID:22624087
- American Psychiatric Association (APA). (2013). *Internet gaming disorder*. Retrieved July 3, 2014, from <http://www.dsm5.org/Documents/Internet%20Gaming%20Disorder%20Fact%20Sheet.pdf>
- Beard, K. W., & Wolf, E. M. (2001). Modification in the proposed diagnostic criteria for internet addiction. *Cyberpsychology & Behavior*, 4(3), 377–383. doi:10.1089/109493101300210286 PMID:11710263
- Blachnio, A., & Przepiorka, A. (2015). Dysfunction of Self-Regulation and Self-Control in Facebook Addiction. *The Psychiatric Quarterly*, 87(3), 493–500. doi:10.1007/s11126-015-9403-1 PMID:26589423
- Black, D. W., Belsare, G., & Schlosser, S. (1999). Clinical features, psychiatric comorbidity, and health-related quality of life in persons reporting compulsive computer use behavior. *The Journal of Clinical Psychiatry*, 60(12), 839–844. doi:10.4088/JCP.v60n1206 PMID:10665630
- Block, J. (2008). Issues for DSM-V: Internet addiction. *The American Journal of Psychiatry*, 165(3), 306–307. doi:10.1176/appi.ajp.2007.07101556 PMID:18316427
- Cao, F., & Su, L. (2007). Internet addiction among Chinese adolescents: Prevalence and psychological features. *Child: Care, Health and Development*, 33(3), 275–281. doi:10.1111/j.1365-2214.2006.00715.x PMID:17439441
- Caplan, S. E., & High, A. C. (2011). Online social interaction, psychosocial well-being and problematic Internet use. In K. S. Young & C. Nabuco de Abreu (Eds.), *Internet addiction: A handbook and guide to evaluation and treatment* (pp. 35–53). Hoboken, NJ: John Wiley & Sons.
- Carnes, P., Green, B., & Carnes, S. (2010). The same yet different: Refocusing the Sexual Addiction Screening Test (SAST) to reflect orientation and gender. *Sexual Addiction & Compulsivity*, 17(1), 7–30. doi:10.1080/10720161003604087
- Chen, C. Y., Huang, M. F., Yen, J. Y., Chen, C. S., Liu, G. C., Yen, C. F., & Ko, C. H. (2015). Brain correlates of response inhibition in Internet gaming disorder. *Psychiatry and Clinical Neurosciences*, 69(4), 201–209. doi:10.1111/pcn.12224 PMID:25047685
- Cheng, C., & Li, A. Y. (2014). Internet addiction prevalence and quality of (real) life: A meta-analysis of 31 nations across seven world regions. *Cyberpsychology, Behavior, and Social Networking*, 17(12), 755–760. doi:10.1089/cyber.2014.0317 PMID:25489876

- Chou, C., Condrón, L., & Belland, J. C. (2005). A review of the research on Internet addiction. *Educational Psychology Review*, 17(4), 363–388. doi:10.1007/s10648-005-8138-1
- Dalbudak, E., Evren, C., Aldemir, S., Taymur, I., Evren, B., & Topcu, M. (2015). The impact of sensation seeking on the relationship between attention deficit/hyperactivity symptoms and severity of Internet addiction risk. *Psychiatry Research*, 228(1), 156–161. doi:10.1016/j.psychres.2015.04.035 PMID:25962354
- Davis, R. A. (2001). A cognitive-behavioral model of pathological Internet use. *Computers in Human Behavior*, 17(2), 187–195. doi:10.1016/S0747-5632(00)00041-8
- Davis, R. A., Flett, G. L., & Besser, A. (2002). Validation of a new scale for measuring problematic Internet use: Implications for pre-employment screening. *Cyberpsychology & Behavior*, 5(4), 331–345. doi:10.1089/109493102760275581 PMID:12216698
- Donati, M. A., Chiesi, F., Ammannato, G., & Primi, C. (2015). Versatility and Addiction in Gaming: The Number of Video-Game Genres Played Is Associated with Pathological Gaming in Male Adolescents. *Cyberpsychology, Behavior, and Social Networking*, 18(2), 129–132. doi:10.1089/cyber.2014.0342 PMID:25684613
- Durkee, T., Kaess, M., Carli, V., Parzer, P., Wasserman, C., Floderus, B., & Wasserman, D. et al. (2012). Prevalence of pathological internet use among adolescents in Europe: Demographic and social factors. *Addiction (Abingdon, England)*, 107(12), 2210–2222. doi:10.1111/j.1360-0443.2012.03946.x PMID:22621402
- Grant, J. E., Atmaca, M., Fineberg, N. A., Fontenelle, L. F., Matsunaga, H., Janardhan Reddy, Y., & Veale, D. et al. (2014). Impulse control disorders and behavioural addictions in the ICD-11. *World Psychiatry; Official Journal of the World Psychiatric Association (WPA)*, 13(2), 125–127. doi:10.1002/wps.20115 PMID:24890056
- Griffiths, M. (2000). Does Internet and computer” addiction” exist? Some case study evidence. *Cyberpsychology & Behavior*, 3(2), 211–218. doi:10.1089/109493100316067
- Ha, Y.-M., & Hwang, W. J. (2014). Gender differences in internet addiction associated with psychological health indicators among adolescents using a national web-based survey. *International Journal of Mental Health and Addiction*, 12(5), 660–669. doi:10.1007/s11469-014-9500-7
- Hall, A. S., & Parsons, J. (2001). Internet addiction: College student case study using best practices in cognitive behavior therapy. *Journal of Mental Health Counseling*, 23(4), 312–327.
- Han, D. H., Kim, S. M., Lee, Y. S., & Renshaw, P. F. (2012). The effect of family therapy on the changes in the severity of on-line game play and brain activity in adolescents with on-line game addiction. *Psychiatry Research: Neuroimaging*, 202(2), 126–131. doi:10.1016/j.psychresns.2012.02.011 PMID:22698763
- Jang, K. S., Hwang, S. Y., & Choi, J. Y. (2008). Internet addiction and psychiatric symptoms among Korean adolescents. *The Journal of School Health*, 78(3), 165–171. doi:10.1111/j.1746-1561.2007.00279.x PMID:18307612

- Kaufman, E. M. D., & Yoshioka, M. R. (2004). *Substance Abuse Treatment and Family Therapy: A Treatment Improvement Protocol (TIP) 39*. Rockville Substance Abuse and Mental Health Services Administration.
- King, D. L., Delfabbro, P. P., Griffiths, M. D., & Gradisar, M. (2011). Assessing clinical trials of Internet addiction treatment: A systematic review and CONSORT evaluation. *Clinical Psychology Review, 31*(7), 1110–1116. doi:10.1016/j.cpr.2011.06.009 PMID:21820990
- Ko, C. H., Yen, J. Y., Chen, C. C., Chen, S. H., & Yen, C. F. (2005). Proposed diagnostic criteria of Internet addiction for adolescents. *The Journal of Nervous and Mental Disease, 193*(11), 728–733. doi:10.1097/01.nmd.0000185891.13719.54 PMID:16260926
- Ko, C. H., Yen, J. Y., Yen, C. F., Chen, C. S., & Chen, C. C. (2012). The association between Internet addiction and psychiatric disorder: A review of the literature. *European Psychiatry, 27*(1), 1–8. doi:10.1016/j.eurpsy.2010.04.011 PMID:22153731
- Koo, C., Wati, Y., Lee, C. C., & Oh, H. Y. (2011). Internet-Addicted Kids and South Korean Government Efforts: Boot-Camp Case. *Cyberpsychology, Behavior, and Social Networking, 14*(6), 391–394. doi:10.1089/cyber.2009.0331 PMID:21204693
- Kuss, D. J., Shorter, G. W., van Rooij, A. J., van de Mheen, D., & Griffiths, M. D. (2014). The Internet addiction components model and personality: Establishing construct validity via a nomological network. *Computers in Human Behavior, 39*, 312–321. doi:10.1016/j.chb.2014.07.031
- Lemmens, J. S., Valkenburg, P. M., & Peter, J. (2009). Development and validation of a game addiction scale for adolescents. *Media Psychology, 12*(1), 77–95. doi:10.1080/15213260802669458
- Li, W., Garland, E. L., & Howard, M. O. (2014). Family factors in Internet addiction among Chinese youth: A review of English-and Chinese-language studies. *Computers in Human Behavior, 31*, 393–411. doi:10.1016/j.chb.2013.11.004
- Lin, M. P., Ko, H. C., & Wu, J. Y. W. (2011). Prevalence and Psychosocial Risk Factors Associated with Internet Addiction in a Nationally Representative Sample of College Students in Taiwan. *Cyberpsychology, Behavior, and Social Networking, 14*(12), 741–746. doi:10.1089/cyber.2010.0574 PMID:21651418
- Lu, D. W., Wang, J. W., & Huang, A. C. (2010). Differentiation of Internet Addiction Risk Level Based on Autonomic Nervous Responses: The Internet-Addiction Hypothesis of Autonomic Activity. *Cyberpsychology, Behavior, and Social Networking, 13*(4), 371–378. doi:10.1089/cyber.2009.0254 PMID:20712495
- Müller, K., Janikian, M., Dreier, M., Wölfling, K., Beutel, M., Tzavara, C., & Tsitsika, A. et al. (2015). Regular gaming behavior and internet gaming disorder in European adolescents: Results from a cross-national representative survey of prevalence, predictors, and psychopathological correlates. *European Child & Adolescent Psychiatry, 24*(5), 565–574. doi:10.1007/s00787-014-0611-2 PMID:25189795
- Ni, X. L., Yan, H., Chen, S. L., & Liu, Z. W. (2009). Factors Influencing Internet Addiction in a Sample of Freshmen University Students in China. *Cyberpsychology & Behavior, 12*(3), 327–330. doi:10.1089/cpb.2008.0321 PMID:19445631
- Orzack, M. H., Voluse, A. C., Wolf, D., & Hennen, J. (2006). An ongoing study of group treatment for men involved in problematic Internet-enabled sexual behavior. *Cyberpsychology & Behavior, 9*(3), 348–360. doi:10.1089/cpb.2006.9.348 PMID:16780403

- Pies, R. (2009). Should DSM-V designate "Internet addiction" a mental disorder? *Psychiatry (Edgmont)*, 6(2), 31–37. PMID:19724746
- Potenza, M. N. (2006). Should addictive disorders include non-substance-related conditions? *Addiction (Abingdon, England)*, 101(1), 142–151. doi:10.1111/j.1360-0443.2006.01591.x PMID:16930171
- Rumpf, H.-J., Vermulst, A. A., Bischof, A., Kastirke, N., Guertler, D., Bischof, G., & Meyer, C. et al. (2014). Occurrence of internet addiction in a general population sample: A latent class analysis. *European Addiction Research*, 20(4), 159–166. doi:10.1159/000354321 PMID:24401314
- Sariyska, R., Reuter, M., Bey, K., Sha, P., Li, M., Chen, Y.-F., & Suárez-Rivillas, A. et al. (2014). Self-esteem, personality and Internet Addiction: A cross-cultural comparison study. *Personality and Individual Differences*, 61, 28–33. doi:10.1016/j.paid.2014.01.001
- Shek, D. T., & Yu, L. (2012). Internet addiction in Hong Kong adolescents: profiles and psychosocial correlates. *International Journal on Disability and Human Development*, 11(2), 133–142.
- Shek, D. T., & Yu, L. (2016). Adolescent internet addiction in Hong Kong: Prevalence, change, and correlates. *Journal of Pediatric and Adolescent Gynecology*, 29(1), S22–S30. doi:10.1016/j.jpag.2015.10.005 PMID:26461526
- Šmahel, D., & Blinka, L. (2012). Excessive internet use among European children. *Children, risk and safety on the internet: Research and policy challenges in comparative perspective*, 191–204.
- Starcevic, V. (2013). Is Internet addiction a useful concept? *The Australian and New Zealand Journal of Psychiatry*, 47(1), 16–19. doi:10.1177/0004867412461693 PMID:23293309
- Subrahmanyam, K., & Smahel, D. (2011). *Digital youth: The role of media in development*. Springer Science & Business Media. doi:10.1007/978-1-4419-6278-2
- Toronto, E. (2009). Time out of mind: Dissociation in the virtual world. *Psychoanalytic Psychology*, 26(2), 117–133. doi:10.1037/a0015485
- van Rooij, A. J., Zinn, M. F., Schoenmakers, T. M., & van de Mheen, D. (2012). Treating Internet Addiction With Cognitive-Behavioral Therapy: A Thematic Analysis of the Experiences of Therapists. *International Journal of Mental Health and Addiction*, 10(1), 69–82. doi:10.1007/s11469-010-9295-0
- Villella, C., Martinotti, G., Di Nicola, M., Cas-sano, M., La Torre, G., Gliubizzi, M. D., & Conte, G. et al. (2011). Behavioural Addictions in Adolescents and Young Adults: Results from a Prevalence Study. *Journal of Gambling Studies*, 27(2), 203–214. doi:10.1007/s10899-010-9206-0 PMID:20559694
- Vondráčková, P. (2011). Závislostní chování na internetu—případová studie. *Česká a Slovenská*, 107(5), 301–306.
- Wang, L., Luo, J., Bai, Y., Kong, J., Luo, J., Gao, W., & Sun, X. (2013). Internet addiction of adolescents in China: Prevalence, predictors, and association with well-being. *Addiction Research and Theory*, 21(1), 62–69. doi:10.3109/16066359.2012.690053
- Wang, W. (2001). Internet dependency and psychosocial maturity among college students. *International Journal of Human-Computer Studies*, 55(6), 919–938. doi:10.1006/ijhc.2001.0510
- Weinstein, A., Feder, L. C., Rosenberg, K. P., & Dannon, P. (2014). Internet addiction disorder: overview and controversies. In K. P. Rosenberg & L. Curtiss (Eds.), *Behavioral addictions: Criteria, evidence, and treatment* (pp. 99–118). New York, NY: Academic Press. doi:10.1016/B978-0-12-407724-9.00005-7

Wölfling, K., Beutel, M. E., & Müller, K. (2012). Construction of a Standardized Clinical Interview to Assess Internet addiction: First Findings Regarding the Usefulness of AICA-C. *Journal of Addiction Research & Therapy*, S6.

Wu, X. H., Chen, X. G., Han, J., Meng, H., Luo, J. H., Nydegger, L., & Wu, H. R. (2013). Prevalence and factors of addictive Internet use among adolescents in Wuhan, China: Interactions of parental relationship with age and hyperactivity–impulsivity. *PLoS ONE*, 8(4), e61782. doi:10.1371/journal.pone.0061782 PMID:23596525

Xu, J., Shen, L. X., Yan, C. H., Hu, H., Yang, F., Wang, L., & Shen, X. M. et al. (2012). Personal characteristics related to the risk of adolescent internet addiction: A survey in Shanghai, China. *BMC Public Health*, 12(1), 1106. doi:10.1186/1471-2458-12-1106 PMID:23259906

Young, K. S. (1998). Internet addiction: The emergence of a new clinical disorder. *Cyberpsychology & Behavior*, 1(3), 237–244. doi:10.1089/cpb.1998.1.237

Young, K. S. (1999). Internet addiction: symptoms, evaluation and treatment. In L. VandeCreek & T. Jackson (Eds.), *Innovations in Clinical Practice: A Source Book* (pp. 19–31). Sarasota, FL: Professional Resource Press.

Young, K. S. (2007). Cognitive behavior therapy with Internet addicts: Treatment outcomes and implications. *Cyberpsychology & Behavior*, 10(5), 671–679. doi:10.1089/cpb.2007.9971 PMID:17927535

Young, K. S. (2011). CBT-IA: The First Treatment Model for Internet Addiction. *Journal of Cognitive Psychotherapy*, 25(4), 304–312. doi:10.1891/0889-8391.25.4.304

Zhang, L. X., Amos, C., & McDowell, W. C. (2008). A Comparative Study of Internet Addiction between the United States and China. *Cyberpsychology & Behavior*, 11(6), 727–729. doi:10.1089/cpb.2008.0026 PMID:18991530

Zhou, N., & Fang, X.-y. (2015). Beyond peer contagion: Unique and interactive effects of multiple peer influences on Internet addiction among Chinese adolescents. *Computers in Human Behavior*, 50, 231–238. doi:10.1016/j.chb.2015.03.083

ADDITIONAL READING

Beard, K. W., & Wolf, E. M. (2001). Modification in the proposed diagnostic criteria for internet addiction. *Cyberpsychology & Behavior*, 4(3), 377–383. doi:10.1089/109493101300210286 PMID:11710263

Cheng, C., & Li, A. Y. (2014). Internet addiction prevalence and quality of (real) life: A meta-analysis of 31 nations across seven world regions. *Cyberpsychology, Behavior, and Social Networking*, 17(12), 755–760. doi:10.1089/cyber.2014.0317 PMID:25489876

Griffiths, M. (2000). Does Internet and computer” addiction” exist? Some case study evidence. *Cyberpsychology & Behavior*, 3(2), 211–218. doi:10.1089/109493100316067

King, D. L., Delfabbro, P. P., Griffiths, M. D., & Gradisar, M. (2011). Assessing clinical trials of Internet addiction treatment: A systematic review and CONSORT evaluation. *Clinical Psychology Review*, 31(7), 1110–1116. doi:10.1016/j.cpr.2011.06.009 PMID:21820990

Starcevic, V. (2013). Is Internet addiction a useful concept? *The Australian and New Zealand Journal of Psychiatry*, 47(1), 16–19. doi:10.1177/0004867412461693 PMID:23293309

Toronto, E. (2009). Time out of mind: Dissociation in the virtual world. *Psychoanalytic Psychology*, 26(2), 117–133. doi:10.1037/a0015485

Young, K. S. (1998). *Caught in the net: How to recognize the signs of internet addiction – and a winning strategy for recovery*. New York, NY: John Wiley & Sons.

Young, K. S., & Nabuco de Abreu, C. (Eds.). (2011). *Internet addiction: A handbook and guide to evaluation and treatment*. Hoboken, NJ: John Wiley & Sons.

KEY TERMS AND DEFINITIONS

Cybersex: Is a virtual sexual encounter between two or more people connected via the Internet.

Diagnostic and Statistical Manual of Mental Disorders: (DSM) provides a common language and standard criteria for the classification of mental disorders. It is published by the American Psychiatric Association.

International Classification of Diseases: (ICD) is a list of mental and behavioral disorders published by the World Health Organization.

Internet Addiction: Refers to the overuse of the Internet such that it leads to the impairment of an individual's mental or emotional psychological state, as well as their scholastic, occupational, and social interactions (Beard & Wolf, 2001).

Internet Addiction Treatment: Refers to treatment approaches for Internet addiction.

MMORPG: Is the acronym for massive multiplayer online role-playing games, which are online games where a player is able to play with other people all around the world.

H

Mediated Embodiment in New Communication Technologies

Laura Aymerich-Franch

CNRS-AIST JRL (Joint Robotics Laboratory), AIST, Japan

INTRODUCTION

Becoming someone else in a fantasy or digitally-created world is a daydreaming fantasy often explored in literature and cinema, especially in Cyberpunk. Stories about digital worlds in which characters are transported somewhere else and experience becoming someone new through avatar self-representations -such as in *The Matrix* (1999) or *Avatar* (2009)- create a sort of “meta-transportation” experience to the audience. Viewers are transported into a fantasy world and feel identified with a character which is in turn transported into a new world and transformed into a new character. The possibility of adopting a digital or robotic surrogate body, described in this kind of narratives, is closer to real than ever before thanks to new communication technologies such as virtual reality or last generation robots. As Dyson (2005) argues:

The ‘new’ of new media such as virtual reality, has been identified by its ability to transcend ‘old’ media that are based on seeing (film, television), with a state of immersion based on being. By being in, rather than looking at, virtual environments, the immersant is said to occupy the space and time, the here and now, the virtual present of a separate but ontologically ‘real’ space. In terms of screen culture, the discourse of representation undergoes a strange re-articulation: the ‘as if you are there’ of screen-based media is truncated to a ‘you are there’ – one is in cyberspace, not watching it, one is a navigator, not a viewer (p.86).

Traditional media such as cinema or television and literature are characterized by being passive experiences for the body. *Transportation* in these media is an experience of cognitive, emotional, and imagery involvement into a narrative (Green, Brock, and Kaufman, 2004) that can be best explained from a Cartesian dualistic approach of mind-body distinction. The mind is the true active protagonist in the process of transportation. The mind travels, the body stays. One step further, interactive television, classic videogames, or on-line games, actively involve users in the process of creating or deciding the narrative to increase the feeling of being transported. However, body participation is not a decisive feature in these experiences either.

The principal transformation concerning the experience of transportation in fully immersive systems compared to traditional media is the involvement of the body in the process. In such systems, media interface development is aimed at providing users with fully immersive experiences, with the ultimate goal of making the virtual, real. In these systems, the body plays a central role by becoming progressively embodied in the process. In *the Cyborg’s dilemma*, Biocca (1997) describes this pattern as *progressive embodiment*:

Each progressive step in the development of sensor and display technology moves telecommunication technology towards a tighter coupling of the body to the interface. The body is becoming present in both physical space and cyberspace. The interface is adapting to the body; the body is adapting to the interface (para.2).

DOI: 10.4018/978-1-5225-2255-3.ch367

While *transportation* in traditional media can be connected to Cartesian Dualism, in fully immersive systems, this experience needs to be framed from a body-subject approach (Merleau-Ponty, 1962) or that of embodied cognition (Shapiro, 2010), in which the body is understood as the medium that humans use for having a world (Merleau-Ponty, 1962). For embodied cognition theorists, as in fully immersive systems, embodiment plays a central role in structuring experience and cognition:

To say that cognition is embodied means that it arises from bodily interactions with the world. From this point of view, cognition depends on the kinds of experiences that come from having a body with particular perceptual and motor capacities that are inseparably linked and that together form the matrix within which memory, emotion, language, and all other aspects of life are meshed (Thelen, Schöner, Scheier, & Smith, 2001, p.1).

Body Ownership Illusions

The Rubber Hand Illusion (RHI) paradigm (Botvinick & Cohen, 1998) is considered as the earliest and most relevant precursor of the works on embodiment. On the RHI experiment, participants are seated with their left arm resting on a table. A screen is positioned beside the arm to hide it from the subject's view and a life-sized rubber model of a left hand and arm is placed on the table directly in front of the subject. The subject stays still, looking at the artificial limb. The rubber hand and the real hidden hand are stroked simultaneously with a paintbrush, either synchronously (illusion) or asynchronously (control condition). When stroked synchronously, participants report a feeling of ownership of the rubber hand and mislocalize the position of their real hand towards the fake hand (proprioceptive drift). Visuotactile correlation has been so far the most expanded method to induce the illusion of embodiment of a fake body or body part.

Recently, as an extension of the RHI paradigm, a series of studies have utilized full-body ownership illusions involving fake and virtual bodies in order to explore the mechanisms underlying self-consciousness and, arguably, as a way to demonstrate that the spatial unity between self and body can be disrupted (Ehrsson, 2007; Guterstam & Ehrsson, 2012; Lenggenhager et al., 2007). Specifically, full-body illusions have been used to displace the sense of self-location outside the bodily borders (Ehrsson, 2007; Guterstam & Ehrsson, 2012; Lenggenhager et al., 2007). Other works have looked into how mannequin bodies can be experienced as the own body (Petkova & Ehrsson, 2008) and how the characteristics of the embodied body influence perception (Van der Hoort, Guterstam, & Ehrsson, 2011). Generally, in these experiments, the illusion is achieved by visuo-tactile correlation and a manipulated visual perspective displayed on a head mounted display (HMD).

However, whereas tactile sensory stimulation plays a central role in classical experiments of body ownership transfer (Botvinick & Cohen, 1998), the existing works on mediated embodiment using avatars and humanoids have preferably used other methods to induce embodiment (Alimardani, Nishio, & Ishiguro, 2013; Aymerich-Franch, et al., 2015; González-Franco, et al., 2010; Maselli and Slater, 2013; Nishio, Watanabe, Ogawa, & Ishiguro, 2012; Slater et al., 2010; Watanabe, Nishio, Ogawa, & Ishiguro, 2011), as it will be explained in future sections.

Related Concepts in Media Studies

Self-presence and *Identification* are the most relevant concepts from the Communication Literature related to mediated embodiment. *Self-presence* is used in Virtual Reality to describe the state in which the virtual self is experienced as the actual self (Lee, 2004; Ratan, 2010). The construct of *self-presence* provides a theoretical framework for the study of how users connect to their mediated

self-representations (Aymerich-Franch, Karutz, & Bailenson, 2012). Self-presence, altogether with social and spatial presence, constitute the subdimensions of *Presence*, a broadly studied phenomenon in VR which grew out of the need to understand the effects of highly immersive media technologies (Ratan, 2010). Broadly speaking, *Presence* is understood as the feeling of *being there*, in a mediated environment (Schuemie, Van Der Straaten, Krijn, 2001).

Identification with characters is another widely studied process in the Communication Literature which can be related to embodiment. In studies of media reception, *identification* is described as a mechanism through which audience members adopt the perspective of a character and experience the events that the character goes through as if they were happening to them (Cohen, 2001, 2006). During this process, self-awareness is increasingly lost and temporary replaced with heightened emotional and cognitive connections with the character (Cohen, 2001, p.251). The construct is widely used to study character identification in films and television (Cohen, 2006; Eyal & Rubin, 2003; Hoffner, 1996; Hoffner & Buchanan, 2005; Moyer-Gusée et al., 2011). More recently, it has also been used to study identification in interactive media (Klimmt, Hefner, & Vorderer, 2009; Klimmt, Hefner, Vorderer, Roth, & Blake, 2010; Soto-Sanfiel, Aymerich-Franch, & Ribes, 2010), becoming closer to the concept of mediated embodiment. Most research on identification in media is based on identification with real human or human-looking characters (Cohen, 2006; Eyal & Rubin, 2003; Hoffner, 1996; Hoffner & Buchanan, 2005; Moyer-Gusée, Chung & Jain, 2011; Oatley, 1994). Some works (Aymerich-Franch, 2012) have also used this approach to analyze identification with non-anthropomorphic avatars.

EMBODIMENT IN NEW COMMUNICATION TECHNOLOGIES

Defining Mediated Embodiment

Embodiment is a polysemic word used in a wide range of disciplines, from Philosophy and Psychology to Cognitive Neuroscience and Robotics. Despite its importance in fields such as virtual reality, there have been very few attempts (de Vignemont, 2011; Kilteni, Groten, & Slater, 2012; Longo, et al., 2008) to conceptualize and operationalize the concept. Thus, it remains confusing to the research community what embodiment is on the context of new communication technologies, what its components are, and how it can be measured, studied, and experimentally manipulated.

De Vignemont (2011) distinguishes *embodiment* from *sense of embodiment* in that “embodiment corresponds to a specific type of information processing, whereas the sense of embodiment corresponds to the associated phenomenology” (p.84). In line with de Vignemont’s reflections, Kilteni, Groten, and Slater (2012) define the *sense of embodiment* toward a body as “the sense that emerges when its properties are processed as if they were the properties of one’s own biological body” (Kilteni, Groten, & Slater, 2012, p. 375). Also, for de Vignemont (2011), *embodiment* and *full embodiment* differ in the sense that “E is fully embodied if and only if all its properties are processed in the same way as the properties of one’s body” whereas “for being embodied, there is no need for all the properties of E to be taken into account by the corresponding type of bodily processing, it suffices that some are” (p.87).

Although there is no general agreement, embodiment can be considered to consist of three subcomponents: the sense of body ownership, the sense of self-location, and the sense of agency (Kilteni, Groten, & Slater, 2012; Longo, et al.,

2008). Body-ownership is the sense of owning a body. Self-location is a determinate volume in space in which one feels to be located, normally localized within the bodily boundaries (Blanke & Metzinger, 2009). Finally, agency is “global motor control, including the subjective experience of action, control, intention, motor selection and the conscious experience of will” (Blanke & Metzinger, 2009, p.7). However, more empirical and systematic research is needed to discern the relative contribution of each subcomponent to the overall experience of embodiment as well as to determine whether the three components are necessary conditions to experience the illusion.

To distinguish *embodiment* in the context of Communication Studies from acceptations coming from other disciplines I introduce it as *mediated embodiment*, which can be defined as *the technologically induced illusion of adopting an artificial body in which one perceives to be located*. The concept of *mediated embodiment* has previously been used in a similar manner, to indicate the degree to which the user’s body is connected to the telecommunication system or coupled to the interface (Biocca, 2002).

As the adjective *mediated* indicates, the core differentiation of mediated embodiment to other forms of embodiment resides in that the illusion is created using communication technologies. At the same time, the adjective *mediated* is more inclusive than other possibilities, such as virtual embodiment, because it embraces the experience of embodiment produced by any type of technology, not only virtual reality.

In the current state of the art, the most extended form of mediated embodiment is avatar embodiment in virtual reality. Another form of embodiment is embodiment in a robot. However, the research dealing with embodiment in a robot’s body is still very incipient.

Mechanisms Underlying Mediated Embodiment

H

Clark’s notion of the *negotiable body* is useful to understand why the experience of mediated embodiment is possible. How can an artificial body of an avatar or a robot be experienced as one’s own? Clark (2008) sustains that bodies are open to episodes of transformative restructuring in which new equipment can be incorporated into the thinking and acting systems that humans identify as their minds and bodies (p.31). Such approach may contribute to explain not only embodiment of external objects but also artificial body embodiment in the context of new communication technologies:

Humans are profoundly embodied agents: creatures for whom body, sensing, world, and technology are resources apt for recruitment in ways that yield a permeable and repeatedly reconfigurable agent/world boundary. For the profoundly embodied agent, the world is not something locked away behind the fixed veil of a certain skin-bag, a reasoning engine, and a primary sensory sheath. Rather, it is a resource apt for active recruitment and use, in ways that bring new forms of embodied intelligence into being (Clark, 2013, p.279-80).

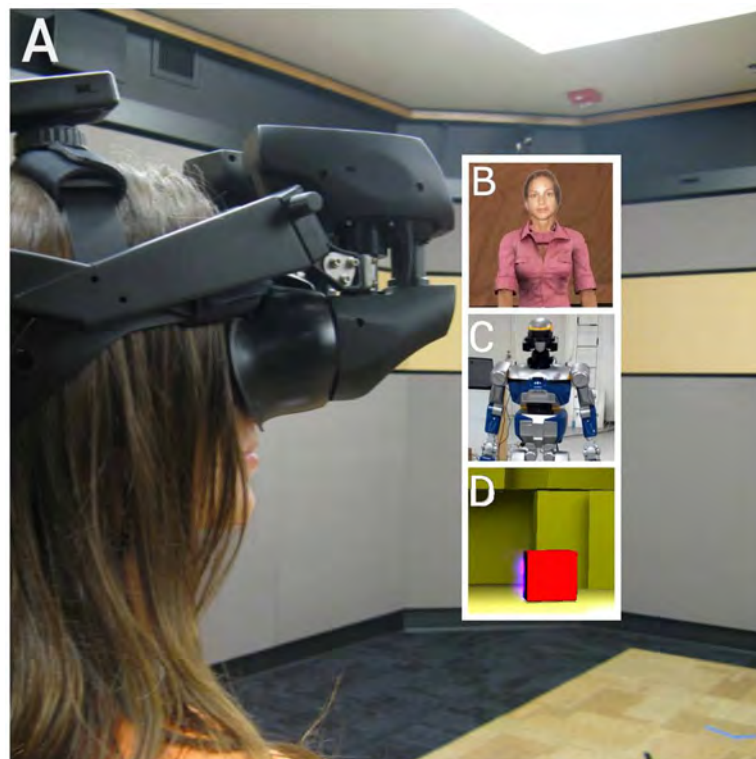
At a technological level, in virtual reality, the illusion of adopting a virtual body is achieved by providing sight and sound feedback to the user from the perspective of the avatar using an HMD (Figure 1A). Also, physical movements of users are tracked and synchronized with those of the avatar (Figure 1B) and spaces are rendered according to users’ movements to make interaction possible with the virtual world (Fox, Arena, & Bailenson, 2009). Sometimes, haptic feedback is implemented to improve the sense of embodiment (Stone, 2001). Generally, users are able to see the virtual limbs

of their avatars. In addition, full-body identification with the digital self-representation can be achieved by reflecting the avatar's appearance in mirrors or other surfaces so users gain knowledge of how they look like and are seen by other users in the virtual environment (Aymerich-Franch, Kizilcec, & Bailenson, 2014; González-Franco, Pérez-Marcos, Spanlang, & Slater, 2010).

Full-body embodiment of a humanoid robot (Figure 1C) can be achieved with the use of an HMD to display vision in 1PP from the artificial body and occlude participant's vision from the real world altogether with visuo-movement synchronization of the robot's body (Aymerich-Franch, et al., 2015). With 1PP, users are able to perceive themselves located behind the eyes of the artificial body and within its bodily borders. In turn, body move-

ment makes it possible to discern the embodied body from the surrounding space. However, while the literature dedicated to robot tele-operation is abundant, there are very few works in the domain of robots that focus attention in creating full-body experiences of robotic embodiment. Ishiguro and colleagues induced embodiment of an android arm by providing visual feedback of the arm combined with synchronization of the operator's hand and the robot's hand movement (Nishio et al., 2012, Watanabe et al., 2011). In a different study, they used a brain-computer interface to control the arm (Alimardani, Nishio, & Ishiguro, 2013). In Aymerich-Franch et al. (2015), the illusion of full-humanoid embodiment was achieved by providing first-person audio-visual feedback from the robot using an HMD and a headset, altogether with

Figure 1. Mediated embodiment (ME) process. Users wear a head-mounted display (A), which provides first person audiovisual perspective (1PP) from the artificial body and blocks vision from the real world. Users embody and control an avatar in virtual reality (B) or a robot in a physical environment (C) and experience its body and its surroundings in 1PP. Non-anthropomorphic avatars such as a cube (D), an animal, or an object may also be used for the purposes of embodiment.



robot movement and the reflection of the robot's body on a mirror.

However, multisensory correlations are not always a necessary condition in the creation of embodiment, since 1PP together with the artificial body moving have been shown to be enough to produce the illusion (Aymerich-Franch, et al., 2015). On this regard, while 1PP seems to contribute to the sense of self-location, body movement seems to contribute to the sense of body-ownership (anecdotal observations).

Other factors such as synchronizing gaze direction, head orientation, or body movement between the two bodies as well as providing different sorts of sensory feedback may contribute to increase the feeling of being embodied in both technologies (virtual reality and robots). Also, resemblance of the artificial body to a human body may improve embodiment of the surrogate body in some cases (Tsakiris & Haggard, 2005; Maselli & Slater, 2013). However, the feeling of being embodied can be also obtained in non-human looking surrogate bodies, such as non-human looking humanoid robots (Aymerich-Franch et al., 2015), or even in non-anthropomorphic avatars (Figure 1D), such as three-dimensional geometric shapes (Aymerich-Franch, 2012).

FUTURE RESEARCH DIRECTIONS

Mediated embodiment presents several potential applications. Artificial bodies might be used to explore remote or virtual places and people with mobility impairments might use devices such as brain-computer interfaces to control these artificial bodies and use them as surrogate bodies (Alimardani, Nishi, & Ishiguro, 2013; Tidoni, Gergondet, Kheddar, & Aglioti, 2014). These forms of embodiment are a growing phenomenon thanks to the democratization of virtual reality and humanoid robot technologies, which are now affordable to consumers at individual level.

In view of this expansion, researchers need to act in three directions. First, by finding a more

systematic approach to decompose embodiment that can clearly identify its subcomponents and its relative contribution in the whole experience of embodiment. Second, by elaborating reliable measures specifically designed for mediated embodiment. Measures that are commonly utilized in other forms of embodiment might not always be applicable to embodiment in the context of communication technologies. Finally, and most important, by examining the consequences of mediated embodiment in all the available systems in order to predict them before these technologies expand into the market.

Regarding the latest, it has already been demonstrated that embodying an avatar in virtual reality may cause effects on attitudes and behaviors. Among others, these effects have been analyzed in relation to racial bias (Groom, Bailenson, & Nass, 2009; Peck et al., 2013), prosocial behavior (Rosenberg, Baughman, & Bailenson, 2013), social anxiety (Aymerich-Franch, Karutz, & Bailenson, 2012; Aymerich-Franch, Kizilcec, & Bailenson, 2014), self-objectification (Fox, Bailenson, & Tricase, 2013), or saving behaviors (Hershfield et al., 2011). A particularly interesting phenomenon that identifies the consequences on attitudes and behavior of embodying avatars in virtual reality is *The Proteus Effect*, which describes how the appearance of the embodied avatar is able to modify the behavior of the user. This phenomenon is believed to occur because individuals associate certain traits of an avatar to specific behavioral stereotypes and expectations. Thus, when users believe that others will expect certain behaviors from them because of their avatars' appearance, they engage in those expected behaviors (Yee & Bailenson, 2007). On the other hand, the consequences of robotic embodiment on behavior have not been yet analyzed. Thus, it is important to carefully examine both the dangers and benefits of these effects to promote a positive use and implementation of embodiment technologies. Also, it is necessary to examine whether these effects found for avatar embodiment in virtual reality also occur in robotic embodiment.

CONCLUSION

This chapter introduces the concept of *mediated embodiment*, the technologically induced illusion of owning an artificial body in which one perceives to be located. Mediated embodiment refers to the experience of embodying a surrogate body produced by new communication technologies such as virtual reality or humanoid robots. During these experiences, users may adopt an avatar's body or a robot's body as their own. Most literature on embodiment has either focused on visuo-tactile correlations or movement synchronization to produce the illusion. However, in mediated embodiment, other characteristics such as first person perspective together with the movement of the surrogate body are also able to induce the illusion of embodiment with no need to apply multi-sensory correlations.

Embodiment is not an absolute state which either takes place or not. Users can feel a lower or higher degree of embodiment into an artificial body. When the experience of embodiment in the artificial body is successful, users feel identified with that body. They perceive and interpret the world through it. Users may even conform to the identity of the body they represent and behave accordingly (Yee and Bailenson, 2007).

Given the novelty of the phenomenon, the consequences of mediated embodiment are still unknown. At the moment, avatars in virtual reality and humanoid robots represent the maximum exponents of mediated embodiment. However, as new technologies come out, new forms of embodiment might appear to be included under this umbrella. If, as suggested above, the form of embodiment determines how humans perceive the world and act upon it, mediated embodiment might have important psychological consequences as well as be capable of inducing attitudinal and behavioral changes on the embodied users. These consequences might not at all be insignificant and deserve careful attention from the research community.

REFERENCES

- Alimardani, M., Nishio, S., & Ishiguro, H. (2013). Humanlike robot hands controlled by brain activity arouse illusion of ownership in operators. *Scientific Reports*, 3, 2396. doi:10.1038/srep02396 PMID:23928891
- Aymerich-Franch, L. (2012). Can we identify with a block? Identification with non-anthropomorphic avatars in virtual reality games. *Proceedings of the International Society for Presence Research Annual Conference*.
- Aymerich-Franch, L., Karutz, C., & Bailenson, J. N. (2012). Effects of Facial and Voice Similarity on Presence in a Public Speaking Virtual Environment. *Proceedings of the International Society for Presence Research Annual Conference*.
- Aymerich-Franch, L., Kizilcec, R. F., & Bailenson, J. N. (2014). The Relationship between Virtual Self Similarity and Social Anxiety. *Frontiers in Human Neuroscience*, 8(944). PMID:25477810
- Aymerich-Franch, L., Petit, D., Ganesh, G., & Kheddar, A. (2015). Embodiment of a humanoid robot is preserved during partial and delayed control. *2015 IEEE International Workshop on Advanced Robotics and its Social Impacts*. doi:10.1109/ARSO.2015.7428218
- Biocca, F. (1997). The Cyborg's Dilemma: Progressive Embodiment in Virtual Environments. *Journal of Computer-Mediated Communication*, 3(0). doi:10.1111/j.10836101.1997.tb00070.x
- Biocca, F. (2002). *The evolution of interactive media*. In M. C. Green, J. J. Strange, & T. C. Brock (Eds.), *Narrative impact. Social and cognitive foundations* (pp. 97–130). Taylor & Francis.
- Blanke, O., & Metzinger, T. (2009). Full-body illusions and minimal phenomenal selfhood. *Trends in Cognitive Sciences*, 13(1), 7–13. doi:10.1016/j.tics.2008.10.003 PMID:19058991

- Botvinick, M., & Cohen, J. (1998). Rubber hands feel touch that eyes see. *Nature*, 391(6669), 756. doi:10.1038/35784 PMID:9486643
- Clark, A. (2008). *Supersizing the mind: Embodiment, action, and cognitive extension*. Oxford University Press. doi:10.1093/acprof:oso/9780195333213.001.0001
- Clark, A. (2013). Re-Inventing Ourselves: The Plasticity of Embodiment, Sensing, and Mind. *Journal of Medicine and Philosophy: A Forum for Bioethics and Philosophy of Medicine*, 32(3), 263–282. doi:10.1002/9781118555927.ch11
- Cohen, J. (2001). Defining Identification: A theoretical look at the identification of audiences with media characters. *Mass Communication & Society*, 4(3), 245–264. doi:10.1207/S15327825MCS0403_01
- Cohen, J. (2006). Audience identification with media characters. In J. Bryant & P. Vorderer (Eds.), *Psychology of Entertainment* (pp. 183–197). Mahwah, NJ: Lawrence Erlbaum Associates.
- De Vignemont, F. (2011). Embodiment, ownership and disownership. *Consciousness and Cognition*, 20(1), 82–93. doi:10.1016/j.concog.2010.09.004 PMID:20943417
- Dyson, F. (2005). Wireless Affections: Embodiment and Emotions in New Media/Theory and Art. *Convergence (London)*, 11(4), 85–105. doi:10.1177/1354856505061056
- Ehrsson, H. H. (2007). The experimental induction of out-of-body experiences. *Science*, 317(5841), 1048. doi:10.1126/science.1142175 PMID:17717177
- Eyal, K., & Rubin, A. M. (2003). Viewer aggression and homophily, identification and parasocial relationships with television characters. *Journal of Broadcasting & Electronic Media*, 47(1), 77–98. doi:10.1207/s15506878jobem4701_5
- Fox, J., Arena, D., & Bailenson, J. N. (2009). Virtual reality: A survival guide for the social scientist. *Journal of Media Psychology*, 21(3), 95–113. doi:10.1027/1864-1105.21.3.95
- Fox, J., Bailenson, J. N., & Tricase, L. (2013). The embodiment of sexualized virtual selves: The Proteus effect and experiences of self-objectification via avatars. *Computers in Human Behavior*, 29(3), 930–938. doi:10.1016/j.chb.2012.12.027
- González-Franco, M., Pérez-Marcos, D., Spanlang, B., & Slater, M. (2010). The Contribution of Real-Time Mirror Reflections of Motor Actions on Virtual Body Ownership in an Immersive Virtual Environment, *IEEE. Virtual Reality (Waltham Cross)*, 111–114.
- Green, M. C., Brock, T. C., & Kaufman, G. F. (2004). Understanding Media Enjoyment: The Role of Transportation Into Narrative Worlds. *Communication Theory*, 14(4), 311–327. doi:10.1111/j.1468-2885.2004.tb00317.x
- Groom, V., Bailenson, J. N., & Nass, C. (2009). The influence of racial embodiment on racial bias in immersive virtual environments. *Social Influence*, 4(1), 1–18.
- Guterstam, A., & Ehrsson, H. H. (2012). Disowning ones seen real body during an out-of-body illusion. *Consciousness and Cognition*, 21(2), 1037–1042. doi:10.1016/j.concog.2012.01.018 PMID:22377139
- Hershfield, H. E., Goldstein, D. G., Sharpe, W. F., Fox, J., Yeykelis, L., Carstensen, L. L., & Bailenson, J. N. (2011). Increasing saving behavior through age-progressed renderings of the future self. *JMR, Journal of Marketing Research*, 48(SPL), S23–S37. doi:10.1509/jmkr.48.SPL.S23 PMID:24634544
- Hoffner, C. (1996). Childrens wishful identification and parasocial interaction with favorite television characters. *Journal of Broadcasting & Electronic Media*, 40(3), 389–402. doi:10.1080/08838159609364360

- Hoffner, C., & Buchanan, M. (2005). Young adults wishful identification with television characters: The role of perceived similarity and character attributes. *Media Psychology*, 7(4), 325–351. doi:10.1207/S1532785XMEP0704_2
- Kiltner, K., Groten, R., & Slater, M. (2012). The Sense of Embodiment in Virtual Reality. *Presence (Cambridge, Mass.)*, 21(4), 373–387. doi:10.1162/PRES_a_00124
- Klimmt, C., Hefner, D., & Vorderer, P. (2009). The video game experience as true identification: A theory of enjoyable alterations of players self perception. *Communication Theory*, 19(4), 351–373. doi:10.1111/j.1468-2885.2009.01347.x
- Klimmt, C., Hefner, D., Vorderer, P., Roth, C., & Blake, C. (2010). Identification with video game characters as automatic shift of self-perceptions. *Media Psychology*, 13(4), 323–338. doi:10.1080/15213269.2010.524911
- Lee, K. M. (2004). Presence, Explicated. *Communication Theory*, 14(1), 27–50. doi:10.1111/j.1468-2885.2004.tb00302.x
- Lenggenhager, B., Tadi, T., Metzinger, T., & Blanke, O. (2007). Video ergo sum: Manipulating bodily self-consciousness. *Science*, 317(5841), 1096–1099. doi:10.1126/science.1143439 PMID:17717189
- Longo, M. R., Schüür, F., Kammers, M. P., Tsakiris, M., & Haggard, P. (2008). What is embodiment? A psychometric approach. *Cognition*, 107(3), 978–998. doi:10.1016/j.cognition.2007.12.004 PMID:18262508
- Maselli, A., & Slater, M. (2013). The building blocks of the full body ownership illusion. *Frontiers in Human Neuroscience*, 7, 83. doi:10.3389/fnhum.2013.00083 PMID:23519597
- Maselli, A., & Slater, M. (2014). Sliding perspectives: Dissociating ownership from self-location during full body illusions in virtual reality. *Frontiers in Behavioral Neuroscience*, 8, 1–19. doi:10.3389/fnhum.2014.00693 PMID:24478648
- Merleau Ponty, M. (1962). *Phenomenology of Perception* (C. Smith, Trans.). Routledge & Kegan Paul.
- Moyer-Gusé, E., Chung, A. H., & Jain, P. (2011). Identification with characters and discussion of taboo topics after exposure to an entertainment narrative about sexual health. *Journal of Communication*, 61(3), 387–406. doi:10.1111/j.1460-2466.2011.01551.x
- Nishio, S., Watanabe, T., Ogawa, K., & Ishiguro, H. (2012). Body ownership transfer to Teleoperated Android Robot. *LNAI*, 7621, 398–407. doi:10.1007/978-3-642-34103-8_40
- Oatley, K. (1994). A taxonomy of the emotions of literary response and a theory of identification in fictional narrative. *Poetics*, 23(1–2), 53–74.
- Peck, T., Seinfeld, S., Aglioti, M., & Slater, M. (2013). Putting Yourself in the Skin of a Black Avatar Reduces Implicit Racial Bias. *Consciousness and Cognition*, 22(3), 779–787. doi:10.1016/j.concog.2013.04.016 PMID:23727712
- Petkova, V. I., & Ehrsson, H. H. (2008). If I were you: Perceptual illusion of body swapping. *PLoS ONE*, 3(12), e3832. doi:10.1371/journal.pone.0003832 PMID:19050755
- Ratan, R. (2010). *Self-Presence, Explicated*. Paper presented at the 60th Annual Conference of the International Communication Association, Singapore.
- Rosenberg, R. S., Baughman, S. L., & Bailenson, J. N. (2013). Virtual Superheroes: Using Superpowers in Virtual Reality to Encourage Prosocial Behavior. *PLoS ONE*, 8(1), 1–9. doi:10.1371/journal.pone.0055003 PMID:23383029
- Schuemie, M. J., Van Der Straaten, P., Krijn, M., & Van Der Mast, C. (2001). Research on presence in virtual reality: A survey. *Cyberpsychology & Behavior*, 4(2), 183–201. doi:10.1089/109493101300117884 PMID:11710246

Shapiro, L. (2010). *Embodied cognition*. Routledge.

Slater, M., Spanlang, B., Sanchez-Vives, M. V., & Blanke, O. (2010). First person experience of body transfer in virtual reality. *PLoS ONE*, 5(5), e10564. doi:10.1371/journal.pone.0010564 PMID:20485681

Soto-Sanfiel, M. T., Aymerich-Franch, L., & Ribes Guàrdia, F. X. (2010). Impacto de la interactividad en la identificación con los personajes de las ficciones. *Psicothema*, 22(4), 822–827. PMID:21044519

Stone, R. J. (2001). Haptic feedback: A brief history from telepresence to virtual reality. In *Haptic Human-Computer Interaction* (pp. 1–16). Springer Berlin Heidelberg. doi:10.1007/3-540-44589-7_1

Thelen, E., Schöner, G., Scheier, C., & Smith, L. B. (2001). The dynamics of embodiment: A field theory of infant perseverative reaching. *Behavioral and Brain Sciences*, 24(1), 1–86. doi:10.1017/S0140525X01003910 PMID:11515285

Tidoni, E., Gergondet, P., Kheddar, A., & Aglioti, S. M. (2014). Audio-visual feedback improves the BCI performance in the navigational control of a humanoid robot. *Frontiers in Neurorobotics*, 8. doi:10.3389/fnbot.2014.00020 PMID:24987350

Tsakiris, M., & Haggard, P. (2005). The rubber hand illusion revisited: Visuotactile integration and self-attribution. *Journal of Experimental Psychology. Human Perception and Performance*, 31(1), 80–91. doi:10.1037/0096-1523.31.1.80 PMID:15709864

Van der Hoort, B., Guterstam, A., & Ehrsson, H. H. (2011). Being Barbie: The size of one's own body determines the perceived size of the world. *PLoS ONE*, 6(5), e20195. doi:10.1371/journal.pone.0020195 PMID:21633503

Watanabe, T., Nishio, S., Ogawa, K., & Ishiguro, H. (2011). Body ownership transfer to android robot induced by teleoperation. *Trans. IEICE, J94*(1), 86–93.

Yee, N., & Bailenson, J. N. (2007). The Proteus Effect: Self transformations in virtual reality. *Human Communication Research*, 33, 271–290. doi:10.1111/j.1468-2958.2007.00299.x

ADDITIONAL READING

Alimardani, M., Nishio, S., & Ishiguro, H. (2011). Body ownership transfer to tele-operated android through mind controlling. *Proceedings of Human-Agent Interaction Symposium* (pp. I-2A-1). Kyoto Institute of Technology.

Aymerich-Franch, L., & Ganesh, G. (2015). The role of functionality in the body model for self-attribution. *Neuroscience Research*. PMID:26602981

Furlanetto, T., Bertone, C., & Becchio, C. (2013). The bilocated mind: New perspectives on self-localization and self-identification. *Frontiers in Human Neuroscience*, 7. PMID:23482653

Hoffmann, L., & Krämer, N. C. (2013). Investigating the effects of physical and virtual embodiment in task-oriented and conversational contexts. *International Journal of Human-Computer Studies*, 71(7), 763–774. doi:10.1016/j.ijhcs.2013.04.007

Slater, M., Perez-Marcos, D., Ehrsson, H. H., & Sanchez-Vives, M. V. (2008). Towards a digital body: The virtual arm illusion. *Frontiers in Human Neuroscience*, 2, 6. doi:10.3389/neuro.09.006.2008 PMID:18958207

Spanlang, B., Normand, J. M., Borland, D., Kiltien, K., Giannopoulos, E., Pomés, A., & Slater, M. et al. (2014). How to build an embodiment lab: Achieving body representation illusions in virtual reality. *Frontiers in Robotics and AI*, 1, 9. doi:10.3389/frobt.2014.00009

KEY TERMS AND DEFINITIONS

Agency: The sense of perceiving oneself as being the cause of the own actions.

Body Ownership: The sense of experiencing a body as the own body.

Embodiment: The illusion of adopting an artificial body as the own body, in which one feels to be located.

Identification (With Characters): The mechanism through which audience members adopt the perspective of a character and experience the events the character goes through as if they were really happening to them.

Mediated Embodiment: The technologically induced illusion of adopting an artificial body in which one perceives to be located, generated with new communication technologies.

Presence: The feeling of really being in an artificial environment.

Self-Location: The place in space in which one feels to be located and which is normally localized inside the physical body.

Self-Presence: A phenomenon taking place in virtual reality by which an avatar is being experienced as if it was the real self.

The Nature of Cyber Bullying Behaviours



Lucy R. Betts

Nottingham Trent University, UK

INTRODUCTION

The continued and rapidly increasing digitalisation of society perpetuates our reliance on technology. There are many benefits associated with our increasing access to, and use of, digital technology and online resources. For example, technology and online resources can be used to: Develop and maintain social connectedness (e.g., Chayko, 2014), promote social responsibility (e.g., Cassidy, Jackson, & Brown., 2009), enhance wellbeing (e.g., Hill, Betts, & Gardner, 2015) and innovation (e.g., Oldham & Da Silva, 2015), prevent cognitive decline (e.g., Slegers, van Boxtel, & Jolles, 2012), facilitate knowledge acquisition (e.g., Thorpe et al., 2015) and knowledge transfer (e.g., Erickson & Johnson, 2011), and complete day-to-day activities such as monitoring health behaviours (e.g., Banchs, & Scher, 2015). However, this increasing access to, and reliance on, technology is not without risks. One such risk is that technology can be used as a mechanism to engage in antisocial behaviour directed towards specific others. For example, through threatening emails and images, and spreading rumours technology can be used to intimidate and victimize others (Dehue, 2013).

Cyber bullying represents a specific form of aggressive behaviour directed towards an individual that takes place using digital means (Law, Shapka, Hymel, Olson, & Waterhouse, 2012). Cyber victimisation can be considered as the experiences of being the target of bullying behaviours. The current chapter will begin by exploring what acts constitute cyber bullying and the various forms that cyber bullying behaviour can take. The chapter will also explore why in-

dividuals engage in such behaviour. Finally, the chapter will make some recommendations that should be considered by researchers examining cyber bullying.

BACKGROUND

Interest in understanding victimisation experiences and bullying behaviours was initially prompted by Olweus' work in the 1970s and subsequently by the wealth of research evidence that has reported longitudinal relationships between experiences of bullying and wellbeing and adjustment (e.g., Fergusson, Boden, & Horwood, 2014; Ttofi, Bowes, Farrington, & Lösel, 2014; Wolke, Copeland, Angold, & Costello, 2013). Together, these studies have suggested that negative consequences may occur for those individuals who engage in bullying behaviour and those who experience victimisation.

Cyberbullying involves individuals using technology as a medium to bully others (Smith, 2009) and has been defined as "the use of the Internet or other digital communication devices to insult or threaten someone" (Juvonen & Gross, 2008, p. 498). Following a recent review and meta-analysis of existing cyber bullying literature, cyber bullying was defined as: "(a) intentional aggressive behaviour that is (b) carried out repeatedly, (c) occurs between a perpetrator and victim who are unequal in power, and (d) occurs through electronic technologies" (Kowalski, Giumetti, Schroeder, & Lattanner, 2014, p. 37). Understanding young people experiences of cyber bullying is important because it has been regarded as an "emerging international public health concern" (Nixon, 2014, p154).

DOI: 10.4018/978-1-5225-2255-3.ch368

FORMS OF CYBER BULLYING

Cyber bullying can occur in many forms and the variation, to some extent, represents the evolving nature of technology. Consequently, there is often little agreement among researchers, practitioners, and young people as to what constitutes cyber bullying. Some researchers, such as Mason (2008), have suggested that cyber bullying comprises both written and verbal acts which can be aligned to the more traditional face-to-face forms of bullying. Conversely, other researchers such as Tokunaga (2010) suggested that cyber bullying includes elements of aggressive, hostile, and harmful acts that are carried out through an electronic device. However, whilst different conceptualisations of cyber bullying have been proposed, when assessing cyber bullying behaviours and cyber victimisation experiences it is important to consider: (a) what technology individuals actually use and (b) how individuals use the technology. Therefore, it is likely that as new technologies emerge and current technologies evolve, new forms of cyber bullying will also continue to emerge and evolve (Slonje, Smith, & Frisé, 2013).

Parallels have also been drawn between cyber bullying and the various forms of face-to-face bullying. Mark and Ratliffe (2011) argued that cyber bullying is a form of relational bullying that uses technology, rather than face-to-face methods, as the medium to bully others. For example, technology can be used to victimise by calling others names, making threats, spreading rumours, disclosing another individual's private information, and purposefully socially isolating or excluding individuals. Similarly, Wang, Iannotti, and Luk (2012) argued that parallels could be drawn between face-to-face relational bullying and cyber bullying as both forms of bullying involve verbal bullying, social exclusion, and spreading rumours but not physical acts, although of course the medium through which these acts occur is different. However, whilst parallels have been drawn by some researchers between face-to-face bullying and cyber bullying it is clear that some

young people regard them as distinct entities. In addition to young people's perception that cyber bullying is distinct from face-to-face bullying, further evidence of this distinction is provided by research that has reported that the different forms of bullying have distinct consequences (e.g., Kubiszewski, Fontaine, Potard, & Auzoult, 2015; Pieschl, Kuhlmann, & Prosch, 2015). Some young people also regard cyber bullying as a distinct form of bullying that is perceived to be more serious than face-to-face bullying (Mishna, Saini, & Solomon, 2009). One potential explanation for why young people regard cyber bullying as more serious than face-to-face bullying is that whilst cyber bullying may occur because of something that happened at school (Cassidy et al., 2009), the accessibility of technology means that incidences of cyber bullying often extend beyond the school day. Consequently, compared to face-to-face forms of bullying, cyber bullying is regarded as more relentless in nature because the constant connectiveness of society means that it is harder for the target to avoid cyber bullying and, as such, the negative consequences are likely exacerbated (Davies, Randall, Ambrose, & Orand, 2014).

There is also little agreement amongst researchers as to whether cyber bullying represents an indirect or direct form of bullying. For example, Huang and Chou (2010) argue that because of the range and scope of technology available cyber bullying can be regarded as an indirect form of bullying. Conversely, Vandebosch and van Cleemput (2009) argue that cyber bullying comprises both direct and indirect forms. The direct forms of cyber bullying include physical (e.g., purposely sending a virus infected file), verbal (e.g., using the internet or mobile phone to insult or threaten), non-verbal (e.g., sending threatening or obscene pictures or illustrations), and social (e.g., excluding someone from a group online) acts. The indirect forms of cyber bullying can involve disclosing entrusted or private information (e.g., through an email), masquerading (e.g., deceiving someone by impersonating someone else), spreading gossip (e.g., using a mobile phone, email, or chat

facility), and taking part in voting on a defamatory polling website. In addition to the forms of cyber bullying outlined by Vandebosch and van Cleemput (2009), Calvete, Orue, Estévez, Villardón, and Padilla (2010) suggested a number of additional acts that constitute cyber bullying. In particular, intentional exclusion from an online group; spreading, rumours, gossip, or embarrassing comments about an individual on the internet; circulating the link of such comments on to others; hacking an individual's email account in order to send messages that would cause trouble for the victim; and filming someone whilst they were forced to do something humiliating or whilst they are being attacked.

Alongside the previously mentioned forms of cyber bullying, the use of images as a form of cyber bullying is increasing with webcams increasingly being used as a mechanism to victimise others. Mishna et al. (2009) reported that many of their participants regarded webcams as an integral part of cyber victimisation. In particular, participants reported that they were blackmailed in to performing various acts in front of webcams to prevent the bully from disclosing personal information about them. Similarly, Blakeney (2012) suggested that images could be used to victimise both in the form of inclusion where a hurtful or harmful image was sent directly to the victim or exclusion where a harmful image about the victim was sent to others, often without the victim's knowledge. Aligned to the issues around blackmail and image manipulation, cyber bullying can also comprise identity theft. In some cases of identity theft, the bully may purposefully adopt the identity of one of the victims' friends before carrying out the bullying episode to heighten the sense of fear in the victim (Mishna et al., 2009).

More recently, Chisholm (2014) proposed 11 behavioural acts that could be regarded as cyber bullying: (1) catfishing (i.e., tricking people to relationships by creating false profiles), (2) cheating, forming roving gangs, and blocking entry points in games, (3) spreading insults, images, or humiliating or threatening messages, (4)

flaming (i.e., adopting an argumentative style), (5) impersonating others, (6) slamming, (7) rating (i.e., taking over a device without the target's knowledge), (8) relational aggression, (9) sexting, (10) shock trolling (i.e., creating offensive posts to get a reaction from other users), and (11) stalking or threatening violence.

Together, there are various forms that cyber bullying can take which can encompass both direct and indirect means. However, the media through which these bullying episodes occur can also vary. For example, Beale and Hall (2007) suggested that cyber bullying can occur through: Instant messenger, social networking sites, email, small text messaging, websites, voting booths, chat rooms, and bash boards. It is likely that many of the forms of cyber bullying previously outlined in this chapter can take place through multiple means of technology. Consequently, the range of media that can be used to cyber bully has led to some arguing that researchers should examine both the content of the bullying episode and the media through which it occurred (e.g., Akbulut, Sahin, & Eristi, 2010).

Whilst much of the literature has focused on the use of online media to engage in bullying behaviours, both the increasing use of mobile telephones and the increasing functionality of many mobile telephones and smart telephones, means that such devices can also be used during cyber bullying episodes. For example, children report experiencing hurtful small text messages as a form of cyber victimisation (D'Antona, Kevorkian, & Russom, 2010; Marsh, McGee, Nada-Raja, & Williams, 2010). Although some researchers have argued that cyber bullying using mobile phones was distinct from cyber bullying using personal computers (e.g., Slonje & Smith, 2008; Wang, Iannotti, & Nansel, 2009), more recently Slonje et al. (2013) argued that the changing capabilities of smart telephones means that cyber bullying using mobile telephones and online devices are beginning to merge. Consequently, it is important to examine both the media used to cyber bully and also the content of the bullying episode.

Although cyber bullying can take many forms as discussed in the chapter, what young people experience may differ. For example, the most frequent forms of cyber bullying young people from Spain reported experiencing were: stealing passwords, frightening telephone calls, offensive or insulting messages, and slandering (Garaigordobil, 2015). Therefore, researchers and practitioners should examine the various forms of cyber bullying when examining the influence of cyber bullying as it may be that the different value attached to these various episodes may determine the psychological consequences for the individual.

FUTURE RESEARCH DIRECTIONS

One pertinent point with regard to the many definitions and characteristics of cyber bullying, is how those who experience cyber victimisation view the behaviour directed towards them. Specifically, some individuals who experience cyber victimisation may misinterpret the intended motives behind the behaviour as teasing and, as such, may not consider it cyberbullying (Dehue, Bolman, & Völlink, 2008). Moreover, behaviour that others regard as cyber bullying may not be interpreted as such because of the norms of the peer group that an individual belongs to (Dehue et al., 2008). In other words, such behaviour whilst meeting the definition of cyber bullying, may be regarded as normative and acceptable behaviour within the group. Together, these variations in how behaviour is interpreted may result in variations in the prevalence rates of cyber bullying. Consequently, researchers such as Hopkins, Taylor, Bowen, and Wood (2013) have argued that it is important that researchers and participants develop a shared understanding of the definition of cyber bullying behaviours and cyber victimisation experiences before data collection commences. However, it is important to acknowledge that young people may be relying on existing or taught knowledge when discussing cyber bullying (Pieschl et al., 2015) which may be different from the researchers' understanding.

The previous studies discussed in this chapter have highlighted a range of forms that cyber bullying can take and these are generally conceptualised as unwanted behaviours directed towards an individual or their friends (Akbulut et al., 2010). However, according to Akbulut et al. the extent to which an individual feels that they are a specific target of these unwanted behaviours is one mechanism for distinguishing between cyber victimisation and more general encounters in the digital environment. For example, encountering obscene images in web pages, obscene pop ups, spam, or propaganda were regarded as not constituting cyber victimisation because participants felt that these were not personally targeted towards them. Conversely, when these same images and messages were encountered through instant messenger systems, participants regarded them as cyber bullying because they were directly targeted at the individual. Therefore, in future studies researchers may need to recognise the content of the messages and also the media through which they are received as there appears to be context dependent aspects that lead to behaviour being interpreted as cyber bullying behaviours and cyber victimisation experiences. In support of this proposition Calvete et al. (2010) argue that when examining cyber bullying and cyber victimisation researchers and practitioners should consider both the content of the bullying episode and the modality used.

As previously noted how we engage with technology is changing and as more technology becomes available it is important that the conceptualisations of cyber bullying behaviour and cyber victimisation experiences continue to be refined by researchers and practitioners to reflect this. Moreover, by refining the definition and conceptualisation of cyber bullying it will ensure that as the new forms of cyber bullying emerge with the technology (Slonje et al., 2013) researchers are able to accurately assess individuals' cyber victimisation experiences. Therefore, researchers should consider adopting a bottom-up approach when devising definitions of cyber bullying behaviour

and cyber victimisation experiences so that they fully capture the participants' experiences and examine a range of behaviour across the media that young people report using as Mehari, Farrell, and Le (2014) advocate.

CONCLUSION

Whilst cyber bullying represents a relatively new phenomenon, there is an emerging line of research that has tried to assess individuals' experiences both from the bullying behaviours they engage in and their victimisation experiences. From this research it is clear that whilst cyber bullying can occur through a range of medium and take many forms, it is important to recognise that how the victims interpret the intent behind the episode depends on whether they regard it to be cyber victimisation. Further, the wide range of media that are used to victimise others and the evolving nature of technology should be taken in to consideration when researchers and practitioners examine cyber bullying to ensure that the measures are as accurate as possible. There is also emerging evidence of how cyber bullying differs from more traditional forms of bullying and again this is something that should be acknowledged by researchers and practitioners.

REFERENCES

- Akbulut, Y., Sahin, Y. L., & Eristi, B. (2010). Development of a scale to investigate cyber victimization among online social utility members. *Contemporary Educational Technology, 1*, 46–59.
- Banchs, J. E., & Scher, D. L. (2015). Emerging role of digital technology and remote monitoring in the care of cardiac patients. *The Medical Clinics of North America, 99*(4), 877–896. doi:10.1016/j.mcna.2015.02.013 PMID:26042888
- Beale, A. V., & Hall, K. R. (2007). Cyberbullying: What school administrators (and parents) can do. *Clearing House (Menasha, Wis.), 18*(1), 8–12. doi:10.3200/TCHS.81.1.8-12
- Blakeney, K. (2012). An instrument to measure traditional and cyber bullying in overseas schools. *International Schools Journal, 32*, 45–54.
- Calvete, E., Orue, I., Estévez, A., Villardón, L., & Padilla, P. (2010). Cyberbullying in adolescents: Modalities and aggressors profile. *Computers in Human Behavior, 26*(5), 1128–1135. doi:10.1016/j.chb.2010.03.017
- Cassidy, W., Jackson, M., & Brown, K. N. (2009). Sticks and stones can break my bones, but how pixels hurt me?: Students experiences with cyberbullying. *School Psychology International, 30*(4), 383–402. doi:10.1177/0143034309106948
- Chayko, M. (2014). Techno-social Life: The internet, digital technology, and social connectedness. *Social Compass, 8*(7), 976–991. doi:10.1111/soc4.12190
- Chisholm, J. F. (2014). Review of the status of cyberbullying and cyberbullying prevention. *Journal of Information Systems Education, 25*, 77–87.
- DAntona, R., Kevorkian, M., & Russom, A. (2010). Sexting, texting, cyberbullying and keeping youth safe online. *Journal of Social Sciences, 6*(4), 521–526. doi:10.3844/jssp.2010.523.528
- Davies, K., Randall, D. P., Ambrose, A., & Orand, M. (2014). I was bullied too: Stories of bullying and coping in an online community. *Information Communication and Society, 18*(4), 357–375. doi:10.1080/1369118X.2014.952657
- Dehue, F. (2013). Cyberbullying research: New perspectives and alternative methodologies. Introduction to the special issue. *Journal of Community & Applied Social Psychology, 23*(1), 1–6. doi:10.1002/casp.2139

- Dehue, F., Bolman, C., & Völlink, T. (2008). Cyberbullying: Youngsters experiences and parental perceptions. *Cyberpsychology & Behavior*, 11(2), 217–223. doi:10.1089/cpb.2007.0008 PMID:18422417
- Erickson, J., & Johnson, G. M. (2011). Internet use and psychological wellness during late adulthood. *Canadian Journal on Aging*, 30(02), 197–209. doi:10.1017/S0714980811000109 PMID:24650669
- Fergusson, D. M., Boden, J. M., & Horwood, L. J. (2014). Bullying in childhood, externalizing behaviors, and adult offending: Evidence from a 30-year study. *Journal of School Violence*, 13(1), 146–164. doi:10.1080/15388220.2013.840642
- Garaigordobil, M. (2015). Cyberbullying in adolescents and youth in the Basque country: Prevalence of cybervictims, cyberaggressors, and cyberobservers. *Journal of Youth Studies*, 18(5), 569–582. doi:10.1080/13676261.2014.992324
- Hopkins, L., Taylor, L., Bowen, E., & Wood, C. (2013). A qualitative study investigating adolescents understanding of aggression, bullying, and violence. *Children and Youth Services Review*, 35(4), 658–693. doi:10.1016/j.childyouth.2013.01.012
- Huang, Y.-Y., & Chou, C. (2010). An analysis of multiple factors of cyberbullying among junior high school students in Taiwan. *Computers in Human Behavior*, 26(6), 1581–1590. doi:10.1016/j.chb.2010.06.005
- Juvonen, J., & Gross, E. F. (2008). Extending the school grounds? Bullying experiences in cyberspace. *The Journal of School Health*, 78(9), 496–505. doi:10.1111/j.1746-1561.2008.00335.x PMID:18786042
- Kowalski, R. M., Giumetti, G. W., Schroeder, A. N., & Lattanner, M. R. (2014). Bulling in the digital age: A critical review and meta-analysis of cyberbullying research among youth. *Psychological Bulletin*, 140(4), 1073–1137. doi:10.1037/a0035618 PMID:24512111
- Kubiszewski, V., Fontaine, R., Potard, C., & Auzoult, L. (2015). Does cyberbullying overlap with school bullying when taking modality of involvement into account? *Computers in Human Behavior*, 43, 49–57. doi:10.1016/j.chb.2014.10.049
- Law, D., Shapka, J. D., Hymel, S., Olson, B., & Waterhouse, T. (2012). The changing face of bullying: An empirical comparison between traditional and internet bullying and victimization. *Computers in Human Behavior*, 28(1), 226–232. doi:10.1016/j.chb.2011.09.004
- Mark, L., & Ratliffe, K. T. (2011). Cyber worlds: New playgrounds for bullying. *Computers in the Schools*, 28(2), 92–116. doi:10.1080/07380569.2011.575753
- Marsh, L., McGee, R., Nada-Raja, S., & Williams, S. (2010). Brief report: Text bullying and traditional bullying among New Zealand secondary school students. *Journal of Adolescence*, 33(1), 237–240. doi:10.1016/j.adolescence.2009.06.001 PMID:19573903
- Mason, K. L. (2008). Cyberbullying: A preliminary assessment for school personnel. *Psychology in the Schools*, 45(4), 323–348. doi:10.1002/pits.20301
- Mehari, K. R., Farrell, A. D., & Le, A.-T. H. (2014). Cyberbullying among adolescents: Measures in search of a construct. *Psychology of Violence*, 4(4), 399–415. doi:10.1037/a0037521
- Mishna, F., Saini, M., & Solomon, S. (2009). Ongoing and online: Children and youths perceptions of cyber bullying. *Children and Youth Services Review*, 31(12), 1222–1228. doi:10.1016/j.childyouth.2009.05.004
- Nixon, C. L. (2014). Current perspectives: The impact of cyberbullying on adolescent health. *Adolescent Health. Medicine and Therapeutics*, 5, 143–158. doi:10.2147/AHMT.S36456 PMID:25177157

- Oldham, G. R., & Da Silva, N. (2015). The impact of digital technology on the generation and implementation of creative ideas in the workplace. *Computers in Human Behavior*, 42, 5–11. doi:10.1016/j.chb.2013.10.041
- Pieschl, S., Kuhlmann, C., & Prosch, T. (2015). Beware of publicity! Perceived distress of negative cyber incidents and implications for defining cyberbullying. *Journal of School Violence*, 14(1), 111–132. doi:10.1080/15388220.2014.971363
- Slegers, K., van Boxtel, M. P. J., & Jolles, J. (2012). Computer use in older adults: Determinants and the relationship with cognitive change over a 6 year episode. *Computers in Human Behavior*, 28(1), 1–10. doi:10.1016/j.chb.2011.08.003
- Slonje, R., & Smith, P. K. (2008). Cyberbullying: Another main type of bullying? *Scandinavian Journal of Psychology*, 49(2), 147–154. doi:10.1111/j.1467-9450.2007.00611.x PMID:18352984
- Slonje, R., Smith, P. K., & Frisén, A. (2013). The nature of cyberbullying, and strategies for prevention. *Computers in Human Behavior*, 29(1), 26–32. doi:10.1016/j.chb.2012.05.024
- Smith, P. (2009). Cyberbullying: Abusive relationships in cyberspace. *The Journal of Psychology*, 217, 180–181. doi:10.1027/0044-3409.217.4.180
- Thorpe, K., Hansen, J., Danby, S., Zaki, F. M., Grant, S., Houen, S., & Given, L. M. et al. (2015). Digital access to knowledge in the preschool classroom: Reports from Australia. *Early Childhood Research Quarterly*, 32, 174–182. doi:10.1016/j.ecresq.2015.04.001
- Tokunaga, R. S. (2010). Following you home from school: A critical review and synthesis of research on cyberbullying victimization. *Computers in Human Behavior*, 26(3), 277–287. doi:10.1016/j.chb.2009.11.014
- Ttofi, M. M., Bowes, L., Farrington, D. P., & Lösel, F. (2014). Protective factors interrupting the continuity from school bullying to later internalizing and externalizing problems: A systematic review of prospective longitudinal studies. *Journal of School Violence*, 13(1), 5–38. doi:10.1080/15388220.2013.857345
- Vandebosch, H., & van Cleemput, K. (2009). Cyberbullying among youngsters: Profiles of bullies and victims. *New Media & Society*, 11(8), 1349–1371. doi:10.1177/1461444809341263
- Wang, J., Iannotti, R. J., & Luk, J. W. (2012). Patterns of adolescent bullying behaviours: Physical, verbal, exclusion, rumor, and cyber. *Journal of School Psychology*, 50(4), 521–534. doi:10.1016/j.jsp.2012.03.004 PMID:22710019
- Wang, J., Iannotti, R. J., & Nansel, T. R. (2009). School bullying among adolescents in the United States: Physical, verbal, relational, and cyber. *The Journal of Adolescent Health*, 45(4), 368–375. doi:10.1016/j.jadohealth.2009.03.021 PMID:19766941
- Wolke, D., Copeland, W. E., Angold, A., & Costell, E. J. (2013). Impact of bullying in childhood on adult health, wealth, crime, and social outcomes. *Psychological Science*, 24(10), 1958–1970. doi:10.1177/0956797613481608 PMID:23959952

ADDITIONAL READING

- Ackers, M. J. (2012). Cyberbullying: Through the eyes of children and young people. *Educational Psychology in Practice: Theory, Research and Practice in Educational Psychology*, 28(2), 141–157. doi:10.1080/02667363.2012.665356

- Allison, S., Roeger, L., & Reinfeld-Kirkman, N. (2009). Does school bullying affect adult health? Population survey of health-related quality of life and past victimization. *The Australian and New Zealand Journal of Psychiatry*, 43(12), 1163–1170. doi:10.3109/00048670903270399 PMID:20001416
- Álvarez-García, D., Núñez Pérez, J. C., Dobarro González, A., & Rodríguez Pérez, C. (2015). Risk factors associated with cybervictimization in adolescence. *International Journal of Clinical and Health Psychology*. doi:10.1016/j.ijchp.2015.03.002
- Bannink, R., Broeren, S., Van de Looij-Jansen, P. M., de Waart, F. G., & Raat, H. (2014). Cyber and traditional bullying victimization as a risk factor for mental health problems and suicidal ideation in adolescents. *PLoS ONE*, 9(4), e94026. doi:10.1371/journal.pone.0094026 PMID:24718563
- Bauman, S., Cross, D., & Walker, J. (2013). *Principles of cyberbullying research: Definitions, measures, and methodology*. Hove, UK: Routledge.
- Betts, L. R., & Houston, J. E. (2012). The effects of cyber bullying on children's school adjustment. In J. Jia (Ed.), *Educational stages and interactive learning: From kindergarten to workplace training* (pp. 209–230). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-0137-6.ch013
- Betts, L. R., & Spenser, K. A. (2015). A large can of worms: Teachers perceptions of young peoples experiences using technology. *International Journal of Cyber Behavior, Psychology and Learning*, 5(2), 15–29. doi:10.4018/ijcbpl.2015040102
- Campbell, M. A. (2005). Cyber bullying: An old problem in a new guise? *Australian Journal of Guidance & Counselling*, 15(01), 68–76. doi:10.1375/ajgc.15.1.68
- Casas, J. A., Del Rey, R., & Ortega-Ruiz, R. (2013). Bullying and cyberbullying: Convergent and divergent predictor variables. *Computers in Human Behavior*, 29(3), 580–587. doi:10.1016/j.chb.2012.11.015
- Connell, N. M., Schell-Busey, N. M., Pearce, A. N., & Negro, P. (2014). Badgirlz? Exploring sex differences in cyberbullying behaviours. *Youth Violence and Juvenile Justice*, 12(3), 209–228. doi:10.1177/1541204013503889
- Della Cioppa, V. D., O'Neil, A., & Craig, W. (2015). Learning from traditional bullying interventions: A review of research on cyberbullying and best practice. *Aggression and Violent Behavior*, 23, 61–68. doi:10.1016/j.avb.2015.05.009
- Gámez-Guadix, M., Gini, G., & Calvete, E. (2015). Stability of cyberbullying victimization among adolescents: Prevalence and association with bully-victim status and psychosocial adjustment. *Computers in Human Behavior*, 53, 140–148. doi:10.1016/j.chb.2015.07.007
- Hinduja, S., & Patchin, J. W. (2009). *Bullying beyond the schoolyard: Preventing and responding to cyberbullying*. Thousand Oaks, CA: Sage Publications.
- Hinduja, S., & Patchin, J. W. (2012). *School climate 2.0: Preventing cyberbullying and sexting one classroom at a time*. Thousand Oaks, CA: Sage Publications. doi:10.4135/9781506335438
- Jackson, C. L., & Cohen, R. (2012). Childhood victimization: Modeling the relation between classroom victimization, cyber victimization, and psychosocial functioning. *Psychology of Popular Media Culture*, 1(4), 254–269. doi:10.1037/a0029482
- Kowalski, R. M., & Limber, S. P. (2007). Electronic bullying among middle school students. *The Journal of Adolescent Health*, 41(6), S22–S30. doi:10.1016/j.jadohealth.2007.08.017 PMID:18047942
- Lapidot-Lefler, N., & Dolev-Cohen, M. (2015). Comparing cyberbullying and school bullying among school students: Prevalence, gender, and grade level differences. *Social Psychology of Education*, 18(1), 1–6. doi:10.1007/s11218-014-9280-8

Litwiller, B. J., & Brausch, A. M. (2013). Cyber bullying and physical bullying in adolescent suicide: The role of violent behaviour and substance use. *Journal of Youth and Adolescence*, 42(5), 675–684. doi:10.1007/s10964-013-9925-5 PMID:23381779

Lund, R., Nielsen, K. K., Hansen, D. H., Kriegbaum, M., Molbo, D., Due, P., & Christensen, U. (2009). Exposure to bullying at school and depression in adulthood: A study of Danish men born in 1953. *European Journal of Public Health*, 19(1), 111–116. doi:10.1093/eurpub/ckn101 PMID:19008241

Marr, N., & Field, T. (2001). *Bullycide: Death at play time*. Success Unlimited.

Narayanan, A., & Betts, L. R. (2014). Bullying behaviours and victimization experiences among adolescent students: The role of resilience. *The Journal of Genetic Psychology*, 175(2), 134–146. doi:10.1080/00221325.2013.834290 PMID:24796160

Olweus, D. (1978). *Aggression in the schools: Bullies and whipping boys*. Oxford: Hemisphere.

Olweus, D. (2012). Cyberbullying: An overrated phenomenon? *European Journal of Developmental Psychology*, 9(5), 520–538. doi:10.1080/17405629.2012.682358

Olweus, D. (2012). Comments on cyberbullying: A rejoinder. *European Journal of Developmental Psychology*, 9(5), 559–568. doi:10.1080/17405629.2012.705086

Olweus, D. (2013). School bullying: Development and some important challenges. *Annual Review of Clinical Psychology*, 9(1), 751–780. doi:10.1146/annurev-clinpsy-050212-185516 PMID:23297789

Patchin, J. W., & Hinduja, S. (2012). *Cyberbullying Prevention and Response: Expert Perspectives*. New York: Routledge.

Smith, P. K. (2014). *Understanding school bullying: Its nature & prevention strategies*. London: Sage. doi:10.4135/9781473906853

Stauffer, S., Heath, M. A., Coyne, S. M., & Ferrin, S. (2012). High school teachers perceptions of cyberbullying prevention and intervention strategies. *Psychology in the Schools*, 49(4), 352–367. doi:10.1002/pits.21603

Tangen, D., & Campbell, M. (2010). Cyberbullying prevention: One primary schools approach. *Australian Journal of Guidance & Counselling*, 20(02), 225–234. doi:10.1375/ajgc.20.2.225

Trolley, B. C., & Hanel, C. (2010). *Cyberkids, cyber bullying, cyber balance*. London: Sage Ltd.

Walrave, M., & Heirman, W. (2011). Cyberbullying: Predicting victimisation and perpetration. *Children & Society*, 25(1), 59–72. doi:10.1111/j.1099-0860.2009.00260.x

Wolak, J., Mitchell, K. J., & Finkelhor, D. (2007). Does online harassment constitute bullying? An exploration of online harassment by known peers and online-only contacts. *The Journal of Adolescent Health*, 43(6), 51–58. doi:10.1016/j.jadohealth.2007.08.019 PMID:18047945

Zych, I., Ortega-Ruiz, R., & Del Rey, R. (2015). Scientific research on bullying and cyberbullying: Where we have been and where are we going. *Aggression and Violent Behavior*, 24, 188–198. doi:10.1016/j.avb.2015.05.015

KEY TERMS AND DEFINITIONS

Bash Boards: An online bulletin board or chat room where users can anonymously post messages.

Bullying: A series of aggressive acts directed towards an individual who is often weaker in some way than the person carrying out the act. The aggressive acts can take many forms.

Cyber Bully: An individual who uses technology to carry out acts of bullying.

Cyber Bullying: An act of bullying that can take many forms but involves the use of technology.

Direct Bullying: Involves an act of bullying that the victims is aware of and can be observed.

Indirect Bullying: Involves isolating the victim of bullying from their social world and often occurs without their knowledge.

Psychosocial Adjustment: A reflection of an individual's psychological wellbeing that is influenced by their experiences in the social arena.

Victim: An individual who the acts of bullying are directed towards.

Victimisation: The experience of receiving acts of bullying.

Screen Culture



Ana Melro

University of Aveiro, Portugal

Lídia Oliveira

University of Aveiro, Portugal

INTRODUCTION

Technology and media usage cannot be separated from screens. Examples such as television, computer, smartphone, mobile phone, tablet, e-book reader, multimedia player (music and / or audio), camera and camcorder, watch, digital advertising public panels, virtual reality glasses. In fact, if the goal is to understand what can be considered as a screen, the examples will not cease to increase.

The major goal of the article is to explore the screen definition and the way in which it has emerged in society, in a cross way, almost without being noticed, but ubiquitously, even becoming inseparable from most of the necessary activities. Thus, after almost 20 years of Levinson's (1998 [1997]) reflection and analysis on the need to define a screen taxonomy, it is increasingly relevant to reflect on the screens' nature and on their effects in individuals' daily life. So, how did media and screens step into peoples' lives in such a way a new culture was created and disseminated?

Despite true ontological differences between the artifacts that incorporate screens, there is a progressive process of dilution of the specificities, with the convergence of functionalities and contents. The television becomes interactive and its consumption is increasingly customizable and individualized; the computer and mobile devices allow access to television. One uses the computer to make phone and video calls. The screen as a unifying feature turns out to be the visible side of an ongoing process of convergence that in the

short term will be felt more systematically in the consumption and sociability logics.

This has a lot of implications in all sort of societal levels. In the political sector (e-governance, active citizenship in political decisions); economic sector (new ways of communication and business structures) and cultural sector (e-museums). But also at a micro level, with the need to reorganize familial, labor, scholar and leisure processes around media.

Nowadays, technology is perceived as extension of man (McLuhan (2008 [1964], p. 82). By recognizing the change enhanced by media one can also recognize the effect in the new medium (Federman, 2004, p. 2). The way the above mentioned societal practices changed are intimately related to the way media (and the perceptions of media) also suffered transformations.

A screen culture arises (Chambat & Ehrenberg, 1988), accepted by individuals as a second culture. Media are included in individuals' lives as a second skin, because they are sensitive, ubiquitous and transparent.

Society had suffered major changes and mutations, in order to include media. People comprise media and use them for him/her best purpose. Cyber and screen culture are becoming the focus of social relations of all types (familial, labor and leisure). These were the justifications for the importance of this reflection regarding screen culture.

The article aims to present a state of the art around screen and media uses, their existence and transparency and provide a definition of the concept.

BACKGROUND

Screen Culture: State of the Art

Considering the importance of screens in individuals' daily lives, it is pertinent to draw a chronological path of the main authors that have addressed the topic of screens and were interested in defining the artifacts that usually arise coupled.

Back in the twentieth century, Lev Manovich (1995), when defining what can be considered a screen, noted that screen characteristics can be attributed to a computer monitor, but also to something more unique, such as a painting or a play. Thus, the screen is the frame that separates two different spaces that coexist in some way (Manovich, 1995, p. 1). This will be the screen definition used here, and for the purpose of this article, screen always appears associated with a technological artifact (television, computer or mobile phone). For Manovich (1995), this screen is not neutral. Instead, it has an aggressive status, *"It functions to filter, to screen out, to take over, rendering nonexistent whatever is outside the frame."* (Manovich, 1995, p. 2). Thus and complementing the previous idea with Nelson Zalago's (2010) notion of screen (2010), this is the boundary between the device and the individual receiving the transmitted content (Zagalo, 2010, p. 35).

For an analysis of the differences between the screens studied here, the prospect of Levinson (1998 [1997]) is a starting point. When the author addresses the issue of the twentieth century screens, he warns of the differences in their nature, in particular between the television and the computer, and in this context it shows the need for a taxonomy of screens (Levinson, 1998 [1997] pp. 199-211).

Although radio, photocopying, electronic publishing and fax, the twentieth century can indeed be characterized as the century of the screen. It was so from the beginning. [...] We know our culture, both produced things to the screen, as it was substantially shaped by them. We also know

that the screens are not monolithic in its cultural importance and that different types of things are broadcasted by different types of screens and involve different types of cognitive and emotional processes. Our question, then, is what kinds of things appear in what kind of screens - especially computer screens - and for what purpose. To find an answer, to start building a taxonomy of screens [...] (Levinson, 1998 [1997], p. 199).

Adriana de Souza e Silva (2006) carried her study on the use of social interfaces (designated screens under the definition of the present article) in mediating the relationship between two or more individuals, or in the context of social relations established online. For the author, those interfaces redefine either the communication or the space in which these relationships occur (Silva, 2006, pp. 261-262).

In this perspective, screens are not the only artifacts considered, but, above all, spaces and spaces' convergence, which is something associated to screen usage. This convergence and the emergence of a new space is designed by the author as a hybrid space where the physical and digital converge, but more than that, they are mobile spaces, carried by users in their portable devices, once they are connected to the internet and to other users (Silva, 2006, p. 262).

Screens are, therefore (and increasingly so) a medium, a way of mingle in the real world, with this being referred by Introna and Ilharco (2006) as a "screened world" (Introna & Ilharco, 2006, p. 58). For these authors, the screen is not so much what it represents by itself, but what it transmits to the individual, the message, the content that appears represented on the screen, whether text, images, colors, graphics, etc. (Introna & Ilharco, 2006, p. 62). This way, the user relationship with screens is highly focused on the content and context and not so much on the physical artifact. For that reason, expectations placed on screens are very contextual, and linked to its place in a particular context. For example, at a cinema it is expected that video and image are transmitted, whereas

when opening email it is expected that messages are showed. Screens capture the individual's attention and hold it in the various social contexts (work, leisure, family), but they also make the individual's place in the world clear and show him/her how to be in this world. We look at screens, and we find our way of being in the world (Introna & Ilharco, 2006, pp. 65-66).

Murolo's perspective (2010) follows the direction of screens' unification in media convergence. For the author, the individual is confronted with new screens, new because there is a connection between image and languages which allow for a meaningful construction. New languages, such as photography, animation, audiovisual and multimedia, converge in screens. Quite often these languages' confluence with screens is a bit controversial, in the sense that it tries to embody new meanings in old artifacts, when new artifacts (or new screens) already demand a specific identity (Murolo, 2010, p. 2).

These screens are ubiquitous, including themselves (sometimes meddling) in individuals' social practices "in a usual, confusing and contradictory way." (Murolo 2011, p. 41).

New media and screens attached to it are invisible and define individuals' life through that invisibility, mostly when people stop to think of screens' existence and assume an omnipresence, except when a problem that needs to be solved occurs and gain their visibility state again (Deuze, 2012, p. 111).

It is possible to discuss if this is the spectacle or simulation society, but one cannot not immediately conclude that this is a screens' society, in part because a screen is the frame that divides two different but coexisting spaces (Manovich, 1995, p. 1). Lipovetsky (2010) considers that this society is a screensphere, living in a widespread screen state, with screens anywhere and time, whether at shops, airports, restaurants, bars, subway, cars and planes, of various sizes, textures (smooth, big, miniature, mobile, tactile, graphic, video) and with different capabilities (information, surveillance, entertainment, atmosphere),

imposing a screenocracy where everything is a screen (Lipovetsky, 2010, pp. 10 and 21-22), a new screenculture where transparency is a requirement, involvement and immersion (Zagalo, 2010, p. 51). But it is also demanded a sensuality that promotes an almost intimate relationship between screens and their users, verified by the dissemination of tactile screens all over the places and technological artifacts.

According to Pinto (2005), Lipovetsky (2010), Rivoltella (2010) and Frau-Meigs (2011) we live in a screen society, we are screen individuals.

Divina Frau-Meigs (2011) believes that screens "are the center of all the connection system, as entertainment and services providers. [...]. The screen is present in all spaces, at all times". (Frau-Meigs, 2011, p. 72).

Screens are the intermediaries of individuals' uses of media. Because of that, media assume an epidemic feature, resulting in the creation of a relation of intimacy, seduction and second skin.

Thus, this state of the art will be finished with Cardoso and Quintanilha (2013), regarding the possibility of a sociology of mediation and screen. The authors say "that the focus on the study of the mediation processes and screens role is now divided between three spatial dimensions of social relations frame: the house, the work and intermediate spaces between those. In all three, mediation takes place on the screens we choose to hold or those who accompany us in our pockets, bags or briefcases." (2013, p. 17). Having, therefore, the screen a central role in social, temporal and spatial mediations.

Screen: A Definition Proposal

The screen definition can be quite ambiguous. For example, shadow plays, famous in Eastern archaic societies, were a form of storytelling with moving images in two dimensions, projected continuously on a screen. It was probably the first time that a canvas was used to visualize images and where the utility it could have for theatre plays was apparent. Figure 1 shows this: a canvas is a screen.

Figure 1. Shadow theatre

Source: Moving Image Museum, Portugal



Dating from the eighteenth century, optical boxes served to peek “views”, prints of cities or handmade colorful landscapes. They are designated “Brave New World” because of the fantastic images they gave access to. A box is a screen.

But there are other examples, such as a dark camera where one could see images (Figure 3); a mirror and a light used for images projection on a canvas (Figure 4), in this example the canvas and

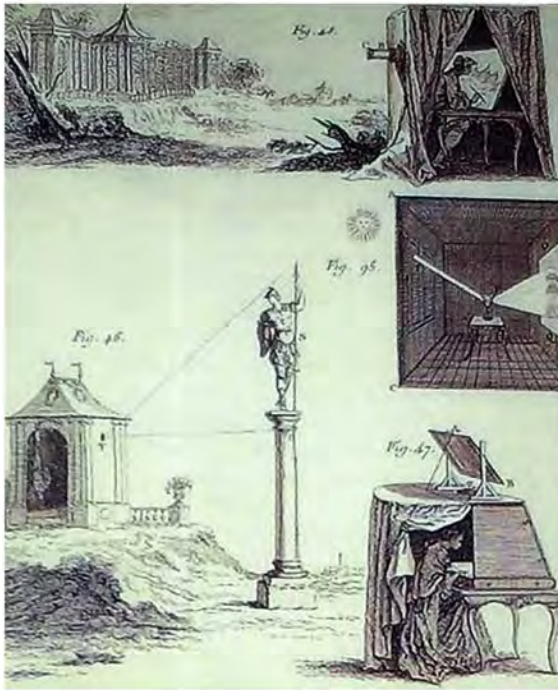
Figure 2. Optical box

Source: Moving Image Museum, Portugal



Figure 3. Dark camera

Source: Moving Image Museum, Portugal



the mirror are screens; or a wall used for image projection is also a screen (Figure 5).

In recent years, it a bigger interaction with the screen and the broadcasted content has been possible, mostly through touch; in more influential ways, such as sending messages (SMS or phone keypad...) to affect what is seen; or even in a social perspective, the messages exchange in chat rooms with others who are watching the same program, for example.

The screen has acquired essential features for its use, such as familiarity, with people integrating new media in their daily lives in such a way that renders screens quite usual and familiar, as opposed to a foreign object. Through familiarity screens also become transparent, being used without their presence noticed. In addition, displays are increasingly touchable and therefore sensitive, i.e., there is no longer such a visible distance between the individual and the screen that conveys him/her to other worlds, they (screens and individuals) are closer to each other; the screen was developed to

Figure 4. Image projection through reflex

Source: Moving Image Museum, Portugal

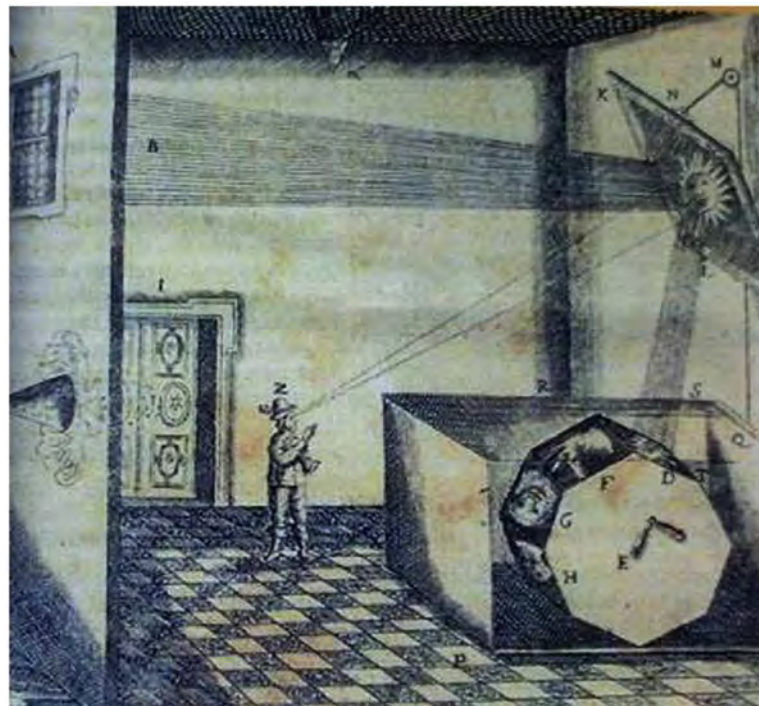
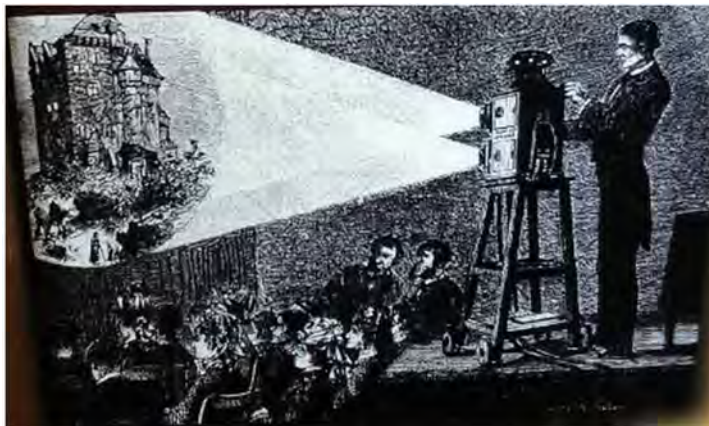


Figure 5. Image projection to a wall

Source: Moving Image Museum, Portugal



allow an approximation of its user physical touch, while increasing the link between devices and the individual. Finally, the screen is the nearest supplier of the information sought and/or found, replacing, in many cases, the mouse and the keyboard, serving as the direct interface between information and individual.

Thus, the article calls upon various definitions to present one of what is a screen within the Screen Culture presented here. The Dictionary of Media and Communications defines this concept as a flat surface on which images are projected (Danesi, 2009, p. 262), and Manovich (1995) complements this definition by stating that it is “a flat, rectangular surface positioned at some distance from your eyes” (Manovich, 1995, p. 1). However, these perspectives are quite reductive of what it is intended to analyze here.

The screen has four essential functions: it is a mediator (between physical and virtual life), a sensations promoter (provides emotions of all kinds when communicating a certain content), a world performer (it plays all the content asked) and a worldviews creator (provides different perspectives of all subjects). You cannot think about media consumption off screens, and this aspect has been thought by the great mass media, to the extent that, increasingly, all have an Internet presence, from radio to the press and are available in all various screens.

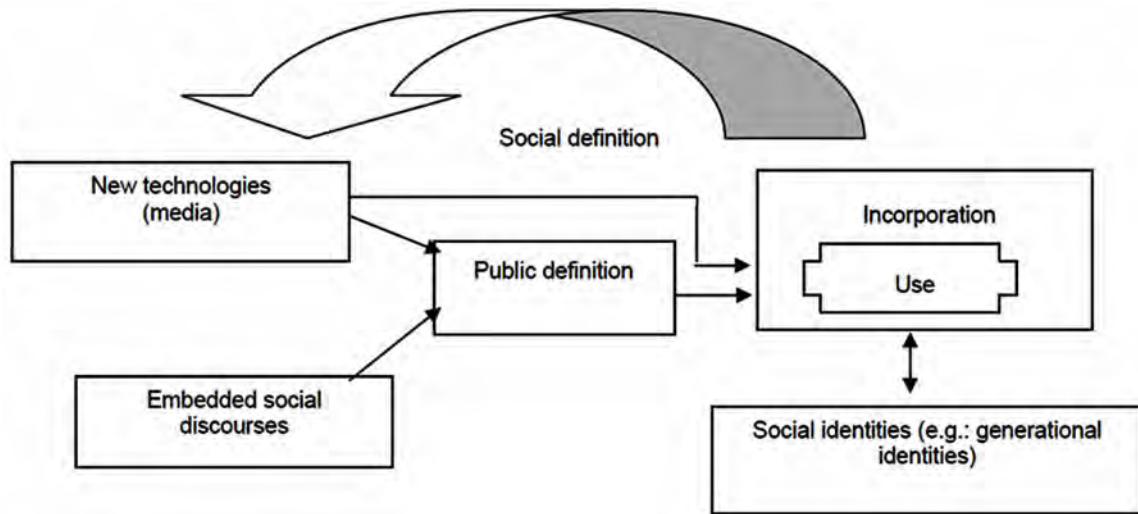
Our screen definition proposal is that this is the ontological somatization of mediation. The screen is what allows the message to be transmitted to the individual, that is, the computer, television and mobile phone monitors. It is something more physical and sensory than virtual. Screens are the connection between the individual and the other side, being the other side what he/she is looking for (or, if not looking for, will eventually going to find), another world and another action.

Media and Screens Appropriation

Relevant to the comprehension of Screen Culture is the understanding of media and screens appropriation. Not only the media is present in the primary socialization, it also became part of people's lives over the years in various contexts (Aroldi & Colombo, 2007, p. 39), i.e. firstly, screens impose their presence to the individual, but later the individuals are the one imposing their presence to the screens (and even contribution to their mutations). The parallel view is the (non-) integration of certain artifacts and tools at a later stage of an individual's life (secondary socialization), in which the familiarity with the media is not as high, but rather a challenge that has to be accepted if you they do not want to be late at

Figure 6. Process of social definition of a new communication technology (or medium)

Source: Aroldi & Colombo, 2007, p. 40



the evolution of the instruments at their disposal (Aroldi & Colombo, 2007, p. 39).

But then, how is a new technology incorporated (e.g., when referring to individuals belonging to the 1950's generation)? Which path is followed to get to the frequent use and unconscious level? Aroldi and Colombo (2007) propose the process of social definition of a new technology. The authors suggest a way in which the approximation of different generations to media is done, considering, on one hand, the technology itself, belonging to a particular generation (being important the era when the technology came about) and, on the other hand, the representation and use of this technology by certain generations (Aroldi & Colombo, 2007, p. 40).

According to the authors, the incorporation occurs in certain social identities, such as generational; but it is also reflected in technologies and social discourses, which, overall, is reflected in a new social setting, since it is an important part of the socialization process, changing even its structures and its symbolic value (Aroldi & Colombo, 2007, p. 40). In fact, following this line of thought, Sáez Vacas (2011) proposes a model of human evolution over time, taking into account the "intelligence-system". According to

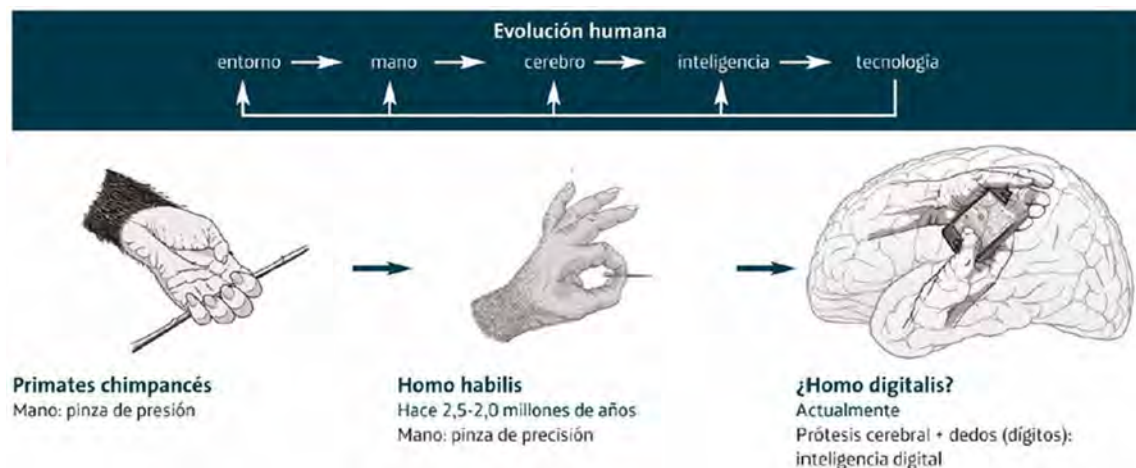
the author, the evolution is carried through certain tools: the environment, the hand, the brain, intelligence and, finally, the technology, which, in turn, uses all of the tools mentioned above. Thus, individuals crossed the primate state to achieve *Homo habilis*, and are now in the state of *Homo digitalis* (Sáez Vacas, 2011, p. 8).

Similar to Sáez Vacas (2011), Ilharco's theory (2007) considers that the individual has been naturally learning how to be in the world, which consequently prepares him/her for how to use technologies that he/she will find, such as mobile phones, computers, televisions, cars or others. Thus, the user will understand media, and when he/she understands the mobile phone, for example, he/she also knows what it means and how to make mobile phone calls. In fact, the author claims that all these artifacts are things taken before being knowable things (Ilharco, 2007, p. 62).

Continuing with Ilharco (2007), and following the thought of Heidegger (1977), in this natural apprehension and appropriation of new media, individuals are shown as part of an efficient orderly process, called *Ge-stell* (or "Enframing") (Heidegger, 1977, p. 19). With the increasing number of technology to be integrated into the daily lives of individuals, the *Ge-stell* arises, "the framed is

Figure 7. Human evolution

Source: Sáez Vacas, 2011, p. 8



re-framed” as a painting within another painting (Ilharco, 2007, p. 66). The broader frame of users’ life appears re-framed in the artefacts, with the internet occupying an increasingly importance.

Thus, it is perceptible that as individuals appropriate media and screens, the more familiar they become with them, assuming a characteristic of continuity between the hands, fingers, brains and technology artifact/screen.

FUTURE RESEARCH DIRECTIONS

Considering the ubiquitous spread of screens, the research perspectives in this field are manifold, from the technological scope of screens as physical artifacts, where the challenge is to create flexible displays that can be wound as a sheet of paper with surfaces easy to read in contexts of high brightness, to the screen as a socio-technical space of social interaction. In this sense, research is challenged by the issue of transmedia narratives of media convergence that finds in the screens the ability to manifest itself. This area encompasses literature, audiovisual, screenwriting and information visualization from big data analysis.

The issue of social temporalities changing with the use of new media was another topic that should be further analyzed. Above all, it should

be understood whether individuals are prepared for those new temporalities and spatialities, or whether individuals’ temporalities are changing and how they have adapted, which could be studied through a longitudinal study.

The safety topic in the use of new media is already of interest to many researchers, especially those that are government-oriented that enable developing ways of a conscious media use. It would be interesting to research this topic through concrete applied measures, especially in rural areas, which are often overlooked by public policies.

It would also be interesting to study the screen as a space of self-projection, namely how new generations build their identities and their self-concept through what is projected of themselves on screen.

Finally, it is also recommended the study of the Screen Culture in the Internet of Things (IoT) context, so as to understand how this new trend gives rise to more and more individuals that are hyperconnected.

CONCLUSION

With the objective of completing some of the ideas discussed here, we return to Sáez Vacas (2011) and verify that there is not a total lack of

technological artifacts in the individuals' daily lives, although the preferred use falls within the computer- internet, mobile phone or television. Currently, screens are more and more touchable, and therefore pervasive screens. So if, according to the author, *Homo digitalis* is characterized by the use of a brain prosthesis, being this the keyboard, the mouse and the screens of all devices, this prosthesis has mutated since the keyboard and mouse were replaced by the screen and the touch,. Then, the emergence of *Homo Digitalis sensitu* can be assumed, for whom the sensations transmitted through contact with the devices and the proximity that unconsciously is created with these devices is increasingly important.

Deuze (2012) suggests that the impact is so profound and overwhelming that one can no longer consider the media as the extension of the man (as did McLuhan (2008 [1964])), because they are no longer external to him/her, so a comparison between media and life is not possible because "we are the media." (Deuze, 2012, pp. 65-66). The screen's growing invisibility and ubiquity has contributed to this purpose, so the focus is now on the task (and eventually the skills required for it performance) and not the tool itself (Weiser, 1994, p. 7). Weiser even compares technology to children, in the sense that both must always be present, are the invisible creations, taken for granted, carried effortlessly throughout life (Weiser, 1994, p. 8). This appropriation is also performed with screens, connective instruments to the content, as stated by Coelho and Neves (2010) "In a lot of what we do and what is done to us, what we perceive and what we think, we can hardly be separated from screens and their perceptual experience" (Coelho & Neves, 2010, p. 102).

New media and screens are the excellence mediators between presence and virtual life, in a logic of social reproduction in both worlds. They are mediating social knowledge and consumption activities, facilitating their appropriation and domestication (Silverstone, Hirsch & Morley, 1992, p. 44). This is something that can be perceived as negative, and this issue should undergo further study, perhaps a longitudinal study over the years

that allows for the understanding of the true effects of technology in individuals' lives.

One of the conclusions is the screen transparency evidence, with the naturalness with which they mingle in the individuals' lives, so that the activities for which it is necessary to use new media are already carried out without individuals being aware of its presence, and this is an essential feature of the development and massification of any new technology, as claimed by Weiser in 1991: "the most profound technologies are those that disappear. They weave themselves into the fabric of everyday life until they are indistinguishable from it." (Weiser, 1991, p. 3). In the opinion of Furtado (2007), we are witnessing the change of the book medium to the screen medium, which brings changes to the relationship between Man, cognition/impairment, culture and body with the world (Furtado, 2007, pp. 106- 107).

We live in a time when three important changes are taking place in the technologies industry that have interference in individuals' lives: one is the entrance in post-PC era, where other artifacts are used in addition to computers, such as laptops, mobile phones or tablets; the other is the cloud era, where much of the information is located in Internet-based servers; and a third, almost paradoxical to the second, is the development of services and applications based on users and content geo-referencing (locative media). Thus, activities and relationships are becoming increasingly mobile, with the internet playing a primary role in the management of these activities and relationships (Maximilien & Campos, 2012, p. 2). In this new logic, Man becomes the gatherer of information, as nomadic were gatherers of food (McLuhan, 2008 [1964], p. 288), even though in this case the man is a nomad connected to its sedentary anchors: an oscillation between local and global, between rooted and rootless (Silva, 2002, p.23). The visions of the future should therefore incorporate this new Man, with all its pendularity tendencies, experimentation and need of constant challenges, especially to what technology is concerned.

REFERENCES

- Almeida, J. (2011). *O essencial sobre o Capital Social* (Vol. 116). Lisbon: Imprensa Nacional Casa da Moeda.
- Aroldi, P., & Colombo, F. (2007). Generational belonging and mediascape in Europe. *Journal of Social Science Education*, 6(1), 34–44.
- Cardoso, G., & Quintanilha, T. L. (2013). Introdução. In *A Sociedade dos Ecrãs. Sociologia dos Ecrãs, Economia da Mediação* (pp. 13-24). Lisbon: Gustavo Cardoso and Edições tinta-da-china, Lda.
- Chambat, P., & Ehrenberg, A. (1988). De la télévision à la culture de écran. Sur quelques transformations de la consommation. *Le Debat*, 5(52), 107–132. doi:10.3917/deba.052.0107
- Coelho, M. Z. S. P., & Neves, J. P. (2010). Conclusão. In M. Z. S. P. Coelho & J. P. Neves (Eds.), *Ecrã, Paisagem e Corpo* (Vol. 20, pp. 102-103). Coimbra: Grácio Editor.
- Collin, S. M. H. (Ed.). (2004 [1987]). *Dictionary of ICT* (4^a ed.). London: Bloomsbury.
- Danesi, M. (Ed.). (2009). *Dictionary of Media and Communications*. New York: M. E. Sharpe.
- Deuze, M. (2012). *Media Life* (1st ed.). Cambridge, UK: Polity Press.
- Federman, M. (2004). *What is the Meaning of the Medium is the Message*. Retrieved from http://individual.utoronto.ca/markfederman/article_mediumisthemessage.htm
- Frau-Meigs, D. (2011). *Penser la société de l'écran. Dispositifs et usages*. Paris: Presses Sorbonne Nouvelle.
- Furtado, J. A. (2007). Fractura digital e literacia: reequacionar as questões de acesso. *Comunicação & Cultura*, (3), 97-111.
- Heidegger, M. (1977). *The question concerning technology, and other essays*. New York: Garland Publishing.
- Ilharco, F. (2007). Where are you? A Heideggerian analysis of the mobile phone. *Comunicação & Cultura*, (3), 59-76.
- Introna, L., & Ilharco, F. (2006). On the Meaning of Screens: Towards a Phenomenological Account of Screenness. *Human Studies*, 29(1), 57–76. doi:10.1007/s10746-005-9009-y
- Levinson, P. (1998). *A Arma Suave - História Natural e Futuro da Revolução da Informação*. Lisbon: Editorial Bizâncio. (Original work published 1997)
- Lipovetsky, G. S., (2010). *O Ecrã Global. Cultura mediática e cinema na era hipermoderna*. (L. F. Sarmiento, Trans.). Lisbon: Edições 70.
- Manovich, L. (1995). An Archeology of a Computer Screen. *Kunstforum International*, 1-15.
- Maximilien, E. M., & Campos, P. (2012). Facts, trends and challenges in modern software development. *International Journal of Agile and Extreme Software Development*, 1(1), 1–5. doi:10.1504/IJAESD.2012.048305
- McLuhan, M. (2008 [1964]). *Compreender os Meios de Comunicação - Extensões do Homem*. Lisbon: Relógio D'Água.
- Murolo, N. (2010). Nuevas pantallas: un desarrollo conceptual. *Razón y Palabra*, (80), 1-13.
- Pinto, M. (2005). A busca da comunicação na sociedade multi-ecrãs: Perspectiva ecológica. *Comunicar*, (25), 259-264.
- Rivoltella, P. C. (2010b). A sociedade multi-ecrãs. Das recomendações educativas à nova mídia-educação. *Educação on-line*, (6), 1-4.

Sáez Vacas, F. (2011). Nativos digitales, inteligencia digital. ¿Homo digitalis? *Revista TELOS (Cuadernos de Comunicación e Innovación)*, (86), 6-8.

Silva, A. S. e. (2006). From Cyber to Hybrid. Mobile Technologies as Interfaces of Hybrid Spaces. *Space and Culture*, 9(3), 261–278. doi:10.1177/1206331206289022

Silva, L. O. d. (2002). *Implicações cognitivas e sociais da globalização das redes e serviços telemáticos. Estudo das implicações da comunicação reticular na dinâmica cognitiva e social da Comunidade Científica Portuguesa* (Doctoral thesis), University of Aveiro. Retrieved May 08, 2013, from <http://www.bocc.ubi.pt/pag/silva-oliveira-lidia-implicacoes-cognitivas-redes.pdf>

Silverstone, R., Hirsch, E., & Morley, D. (1992). Tecnologías de la información y de la comunicación y la economía moral de la familia. In R. Silverstone & E. Hirsch (Eds.), *Los efectos de la nueva comunicación* (pp. 39–57). Barcelona: Editorial Bosch.

Weiser, M. (1991). The Computer for the 21st Century. *Mobile Computing and Communications Review*, 3(3), 3–11. doi:10.1145/329124.329126

Weiser, M. (1994). The World Is Not a Desktop. *Interaction*, 1(1), 7–8. doi:10.1145/174800.174801

Zagalo, N. (2010). Da transparência. In M. Z. S. P. Coelho & J. P. Neves (Eds.), *Ecrã, Paisagem e Corpo* (Vol. 20, pp. 35-54). Coimbra: Grácio Editor.

ADDITIONAL READING

Agger, B. (2011). iTime: Labor and life in a smartphone era. *Time & Society*, 20(1), 119–136. doi:10.1177/0961463X10380730

Agnew, J. A. (2011). Space and Place. In J. A. Agnew & D. N. Livingstone (Eds.), *The SAGE Handbook of Geographical Knowledge* (pp. 316–330). London: SAGE Publications. doi:10.4135/9781446201091.n24

Augé, M. (1992). *Non-lieux. Introduction à une anthropologie de la surmodernité*. Paris: Editions du Seuil.

Borges, J., & Oliveira, L. (2011). Competências infocomunicacionais em ambientes digitais. *Observatorio (OBS*)*, 5(4), 291–326.

Harvey, D. (1990). *The Condition of Postmodernity. An Enquiry into the Origins of Cultural Changes*. Oxford: Blackwell.

Lemos, A. (2007). Ciberespaço e Tecnologias Móveis: processos de Territorialização e Destruturalização na Ciberultura. In A. S. Médola, D. Araújo, & F. Bruno (Eds.), *Imagem, Visibilidade e Cultura Midiática* (pp. 277–293). Porto Alegre: Editora Sulina.

Postman, N. (1993). *Technopoly. The Surrender of Culture to Technology*. New York: Vintage Books.

Rojas, V., Straubhaar, J., Spence, J., Roychowdhury, D., Okur, O., Piñon, J., & Fuentes-Bautista, M. (2012). Communities, Cultural Capital, and Digital Inclusion: Ten Years of Tracking Techno-Dispositions and Techno-Capital. In J. Straubhaar, J. Spence, Z. Tufekci, & R. G. Lentz (Eds.), *Inequity in the Technopolis. Race, Class, Gender, and the Digital Divide in Austin* (pp. 223–264). Austin: University of Texas Press.

Siemens, G. (2005). Connectivism: A Learning Theory for the Digital Age. *International Journal of Instructional Technology & Distance Learning*, 2(1).

Silva, L. O. d., & Abreu, J. F. d. (2003). Cidades digitais: O novo “urbanismo” potencial catalisador da lusofonia. *Anuário Internacional de Comunicação Lusófona*, 1(1), 57–71.

Urry, J. (2000). *Sociology Beyond Societies - mobilities for the twenty-first century*. London: Routledge.

KEY TERMS AND DEFINITIONS

Information and Communication Technologies (ICT): Artefacts needed for production or exchange of information and communication to take place. Thus, it is all the hardware and software that allow individuals to acquire, store, process and distribute information by electronic means, including radio, television, telephone and computers (Collin, 2004 [1987], p. 125).

Media (Uses of): Any symbolic or technological system that enables, structures and amplifies

communication between people, but also provides access to information and content of various types.

Screen: Interface that mediates the relationship between the individual and the content he/she watches (text, image, video, sound and/or multimedia) and allows for interaction, whether by computer, mobile phone/smartphone, TV, tablet, e-book reader, clock, video and photographic camera, public placard, music player/video, among many others.

Social Relationships: Created from the moment that individuals “form families, establish friendships and fellowships, join in groups of cultural, religious, political or other affinity, create organizations and companies.” (Almeida, 2011, p. 39), having these relationships new meanings when the reference is made to online social networks.

Technology Assessment of Information and Communication Technologies



Armin Grunwald

Karlsruhe Institute of Technology, Germany

Carsten Orwat

Karlsruhe Institute of Technology, Germany

INTRODUCTION

Technology Assessment (TA) has developed over the past more than forty years against the background of challenging experiences concerning unintended and often undesirable side effects of science and technology. Development, production, social use, and disposal of technology have often resulted not only in more welfare, employment, health, and other positive achievements, but also in negative or at least ambivalent consequences, including risks to human health, society, and the natural environment. The aim of TA from the very beginning was to contribute to shaping scientific and technological progress and its transformation into innovations according to societal values and goals by investigating and assessing possible impacts and consequences *in advance*, and by transforming this knowledge into advice to decision-makers.

TA has taken up the field of Information Science and Technology (IST) as a subject of study from the 1970s on. Nowadays, this field is of central relevance to TA in a triple respect: (a) as research field *per se*, e.g. with regard to impact dimensions such as privacy, data protection, increasing use of autonomous agents, safety and security, sustainable development, intellectual property rights, regulation, societal vulnerability, *et cetera*. It is (b) also of major and even increasing importance by entering and influencing other fields of technology, e.g. energy supply, military, robotics, logistics, nanotechnology, cognitive sci-

ence, neuroscience, *et cetera*. Finally (c), several new services made available by IST developments are of high utility in TA practice of scientific projects and policy advice, e.g. in the fields of e-participation (Nentwich & König 2012). This article will provide a brief overview of TA with respect to its origin, its development, its objectives, and its current situation in general, followed by a more specific consideration of TA themes and activities in the IST field.

BACKGROUND

Technology Assessment has its roots in specific historical circumstances in the 1960s and 1970s. Activities and concerns in the U.S. political system, in particular in the U.S. Congress, led to the creation of the Office of Technology Assessment (OTA) in 1972 (Bimber, 1996). This origin of TA found a lot of successors in Europe which succeeded in establishing the European Parliamentary Technology Assessment network (EPTA, see www.eptanetwork.org).

Parallel to this development in the political system, far-ranging intellectual changes were taking place. The optimistic belief in scientific and technical progress, which had predominated in the post-Second World War period, came under pressure. Western societies were deeply unsettled by the “Limits of Growth” published by the Club of Rome in 1972, which addressed the limitedness of natural resources. In many fields, problems

with unintended side effects of technology such as pollution and severe accidents became a matter of public debate on further scientific and technological progress. In many countries, social conflicts arose on the occasion of controversial technologies such as nuclear power (from the 1970s on) and genetically modified organisms (from the 1990s on). Ethical questions led to conflicts on the development and use of new technology, in particular in the field of health and human reproduction. Issues of privacy and data protection became a field of controversy, in particular following measures of homeland protection and surveillance strategies after the 9/11 attacks. The challenges led to a complex and multi-dimensional set of objectives and rationales of TA (Grunwald, 2009).

Nowadays, the term “technology assessment” is widely used to designate a broad range of systematic approaches and methods to investigate the conditions for and the consequences of technology and to assess and evaluate them. Its task is to provide knowledge, orientation, and procedures on how to cope with challenges at the interface between technology and society *in both directions*. TA explores and assesses possible impacts and consequences of technology in a prospective manner on the one hand (technology push), and attempts to understand and take up society’s expectations and needs regarding a new technology and direct them to the relevant decision-making processes on the other hand (demand pull). The mission of TA is, thus, to contribute to “a better technology in a better society” (Rip et al., 1995), including reflecting about what the “better” could or should mean in detail and how the respective meaning of the “better” could be determined. There are three partially overlapping branches of TA addressing different targets in the overall technology governance:

1. TA has initially been conceptualized as *policy advice* (Bimber, 1996, Grunwald, 2009), which is still a strong motivation of large parts of TA. The objective is to support policy makers with advice concerning politi-

cal measures which could either influence the further development and use of technology or which themselves could be influenced by new technological developments and achievements. Frequently, policy advice in this sense is about *adequate regulation* (e.g. environmental or safety standards), changes in *research and development funding* (e.g. in the field of new and emerging sciences and technologies such as nanotechnology and synthetic biology), and *political strategies towards sustainable development* involving appropriate technologies. *Parliamentary TA* is an important sub-category of policy-advising TA, showing a high variety of institutional configurations (Cruz-Castro & Sanz-Menendez, 2006).

2. *Participatory TA* has developed approaches to involve citizens, consumers and users, actors of civil society, stakeholders, the media, and the public in different roles at different stages in technology governance (Joss & Belucci 2002, Hennen 2012). According to normative ideas of deliberative democracy, the assessment of technology should be left neither to the scientific experts (expertocracy) nor to the political deciders alone (decisionism) (Habermas 1970). Participative TA procedures are deemed to improve the practical and political legitimacy of decisions on technology. They should make it possible for decisions on technology to be accepted by a larger spectrum of society despite of remaining divergent normative convictions. Several methods have been developed and applied in recent years, such as consensus conferences, citizens’ juries, and focus groups (Joss/Belucci 2002).
3. Building on empirical research on the genesis of technology and on the theoretical framework of social constructivism (Bijker et al., 1987), the idea of *shaping technology* according to social expectations and values emerged. It motivated the development of several approaches, with Constructive TA

(CTA) being the most influential one (Rip et al., 1995). The general idea is not to address policy makers or the public, but rather to approach those groups who are directly involved in the “making of technology”, such as engineers, developers, and planners in companies and publicly-funded research and development centers. Ethical issues and social desires shall, according to this approach, be implemented in technology by enriching and orientating decision-making processes in the design and development phase of new technological systems, products, and services.

An international community evolved around the concept of TA and its various dimensions and diverse objectives, using different concepts and methodologies. Part of this community works in organizations explicitly devoted to TA (e.g., to provide advice to parliaments, cf. the EPTA network mentioned above), part of it is organized in networks (cf. www.netzwerk-ta.net), part is describing its work as systems analysis and life cycle assessment (LCA), and another part is dedicated to research in disciplinary departments of academic or research organizations, such as sociology, philosophy, political science, media studies, economics, or in the STS Community (science, technology & society studies).

TECHNOLOGY ASSESSMENT IN THE IST FIELD

Information and communication technologies (ICTs), related with advanced information science and technology (IST), are among the main driving forces for changes in all modern societies. Development and change have been associated with the penetration of ICT into almost every area of human activities. The computerization of society is characterized to date by parallel and continuing developments: firstly by building a ubiquitous and highly connected ICT infrastruc-

ture interwoven into and with other infrastructures and areas of life, secondly by the further deployment of a growing number of automated and increasingly autonomous technical systems (e.g. production systems, transportation systems, weapons systems), and thirdly by the proliferation of intelligent systems in everyday life. Such technological developments both influencing and influenced by concurrent societal changes give rise to a large manifold of initiatives and projects of technology assessment. In the following they are structured around major thematic directions of socio-technical developments. However, such developments are highly interwoven in reality and often have interrelated implications on privacy, human autonomy, replacement of humans, environmental conditions, etc.:

1. Privacy and security issues are always on the political agenda and subject of research in nearly all TA projects dealing with ICT. However, the attacks of 9/11 and the Snowden revelations have increased awareness of these issues. Technologies of governmental surveillance, biometrical identification, screening of huge data sets on indications of possible terrorist attacks benefit, on the one hand, from the rapid advance in the field of ICT. On the other hand, they provoke societal debates on securing fundamental privacy rights and limitations to or regulation of surveillance practices. Technology assessment aims at contributing to balancing security needs and societal desires to maintain the private sphere (Klüver et al., 2006). For this purpose, assessments of governmental activities of mass surveillance were conducted, from which policy recommendations were derived (Garcia et al., 2015; van den Berg et al., 2015). Related to this are TA projects (e.g., SurPRISE, Prisms, PACTS) dedicated to understanding citizens' opinions and expectations on the balancing between security and privacy and the attitudes about surveillance technologies

(e.g., Čas, 2015). In the last decades, it also became obvious that surveillance by private companies like web tracking, profiling of search engine uses, scanning the use of online services like email, etc. reached mass scale and went beyond the purpose of advertising and marketing. It raised additional privacy concerns and debates on the power of dominant global players in the field of ICT. Thus, TA studies focus on cloud computing and social network sites (Leimbach et al., 2014; Strauß & Nentwich, 2013), shedding light on their implications on privacy and security issues, especially extraterritorial issues through border-crossing data streams. Although TA has taken the role of information intermediaries (US GAO, 2013) and credit scoring (Rothmann et al., 2014) into account, the increased importance of large and combined volumes of data, technologies and practices of data mining, profiling and scoring, and other forms of analytics – often covered by the umbrella term ‘big data’ – is an area of ongoing TA research.

2. IST and the Internet have massively changed value chains and sectors in the economy with ambivalent consequences. Production, transport, logistics, division of labor across countries and continents, international trade and globalization, the workplaces of by far the most employees in industrialized countries – all these fields have been highly impacted or even enabled by the integration of IST, especially Internet-based business and organizational models. The extreme increase of economic efficiency comes at a price not only in this field. The strong increase of competition following economic globalization has, in combination with available IST services, led to adverse effects at the workplace, as a first example. Resulting requirements for permanent availability of employees, increased necessity of constant mobility, competition not only between companies but also between employees in

the same company, the enduring fear of being outranged because one’s own knowledge becomes less relevant, etc. resulted in problematic developments with a strongly increasing number of burn-outs and new debates on work-life balance. The global society’s dependence on a functioning ICT infrastructure is extremely high today and is another example of ambivalent developments. Security issues and vulnerability against hacking and cyber spying are an unwelcomed accompanying effect of the intensive use of ICT in infrastructures. In particular, the use of advanced ICT for the management of larger amounts of fluctuating energy supply from renewables and for enabling dynamic pricing models leads to investigations not only on potential enhancements with the ‘smart grid’ but also on its risks (US GAO, 2011; Ricci et al., 2012). The growing interdependencies between infrastructures, like ICT and electricity infrastructures, have been addressed, for instance, by studying the catastrophic effects of a large-scale and prolonged power outage (Petermann et al., 2011). Another example of this ambivalence can be found in the media industry, where the illegal copying of media products such as music files or videos from Internet sources had threatened the business models of that industry. The reactions were attempts to secure intellectual property rights by technical protection measures, but they got in conflict with consumer interests and rights. Meanwhile, the emergence of ‘prosumers’ actively contributing content through new Internet services has been studied as another important source of change in the media industry (Böhle et al., 2008). Technology assessment is asked in all these ambivalent developments in order to identify chances as well as risks and to provide solutions for dealing with them.

3. The integration of ICTs into society in general and the Internet in particular has also

- been related with the emergence of a “network society” (Castells, 2000). ICTs allow direct, cheap, and fast access to information resources. This aspect gave rise to hopes that citizens would be much better informed about political debates and related arguments and background knowledge, up to expectations for much more transparency in the political system, but the actual developments are ambivalent again. Social networks now have an extremely high number of members unprecedented in human history, but also gain an unparalleled position of monitoring and influencing the social behavior of a vast amount of individuals. Mobilization of pressure groups and networking possibilities supported by the Internet are ambivalent in nature as well. As exemplified during the ‘Arab spring’, this, on the one hand, allows private persons and non-governmental organizations to co-ordinate their activities against powerful institutions. On the other hand, these new opportunities can also be used by terrorist and other radical movements. Technology assessment is therefore not only faced here with a lot of tasks of monitoring and assessing changes in social life, but also with making use of ICT to realize own objectives in political process.
4. More and more visions of ‘ubiquitous computing’, ‘pervasive computing’, and ‘ambient intelligence’ are being realized today, mainly by related concepts and developments that are named with other keywords. The further miniaturization and reduced costs for tags, sensors and connected devices, advancements of ‘intelligent’ systems and components, and the connectivity via diverse forms of Internet communication enable business and organization models called ‘smart home’, ‘smart car’, ‘smart traffic’, ‘smart cities’, etc. as well as ‘Internet of things’ or ‘Internet of everything’. They are realized in almost every area of life including those with great intimacy. Issues of TA research on ubiquitous computing (Bizer et al., 2006; Friedewald & Raabe, 2010) are not getting obsolete but more and more relevant, necessitating focused assessments of the changes in these areas of life. Other issues for TA are emerging, such as the dependence on platforms, issues of accountability, transparency, and trust, or the (ongoing) need for reflecting on privacy conceptions. Also the continuing miniaturization makes elements of ICTs more and more invisible to the human eye. Nanotechnology will certainly support and accelerate these developments (Grunwald, 2012). Such miniaturization, in connection with the limited human eyesight, and the networking of surveillance equipment could substantially obstruct the present control options and regulations for protecting personal data or even make them completely obsolete. The passive surveillance of humans might be supplemented in the more distant future by active surveillance (that the targets themselves might not notice), such as if it proved possible to construct direct technical access to the nervous system or the brain.
 5. Miniaturized technologies including IST and based on nanotechnology offer, according to Roco and Bainbridge (2002), far-reaching perspectives for perceiving even the human body and mind to be formable and for improving them through precisely targeted technical measures. In the idea of “NBIC convergence”, nanotechnology, biotechnology and genetic engineering, information and communication technology, and cognitive science and brain research are assumed to converge and, as a consequence of this convergence, to create radically new opportunities. An example is “neuro-enhancement,” which refers to an enhancement resulting from an implant or medication connected either to the nervous system or directly to the brain. First steps are the technical replacements of natural organs. Examples are cochlear or retinal

- implants to first technically compensate for a *loss* of sensory function (in these cases, of an eye or an ear). As a result of advances in nano-informatics, such as miniaturization or the increased capacity of implants to take up and process data, the spatial dimensions, and performance of the neuro-implants will approach those of natural systems. New IST developments are keys to possibly realize those visions. Technology assessment in this field is more or less to be conceived as a “vision assessment” accompanying the scientific development in its very early stage and helping to better understand chances as well as risks (Grunwald, 2012).
6. Health services are particularly sensitive with regard to the private sphere. The development of small units for preparing diagnoses – “lab on a chip” – can make it possible for comprehensive personalized diagnoses and prognoses to be prepared on the basis of a person’s health data (van Merkerk, 2007). The demands placed on data privacy and thus on the protection of one’s privacy must therefore be high. In particular the strongly increasing fields of tele-medicine and ubiquitous computing in healthcare lead to growing challenges to ensure privacy and to avoid possibilities of misuse (Orwat et al., 2008).
 7. Progress in IST can lead to the loss of human autonomy because it provides technological systems with more and more autonomy for decision-making as is the case in the fields of autonomous robots and autonomous software agents. Self-determination regarding information, a part of an individual’s autonomy, can be limited by technology, just as technological means can be used to gain external control over humans. In particular, the word “technicalization” is frequently associated with a subordination of man to technology, a loss of control, a discontent caused by man’s dependence on technology, and a loss of individuality, emotionality, and spontaneity. Current and future IST developments require thinking anew about the relations between humans and technology with respect to the distribution of autonomy (Decker & Gutmann, 2012; Wiegerling, 2011), which is part of the business of technology assessment involving the ethics of technology (Grunwald, 2012).
 8. IST-enabled innovations have an important place in the debate on e-participation and e-parliament. Petitioning, for example, cannot be simply written off as a relic of a bygone age. Precisely because society is becoming increasingly complex, petitions can provide citizens with an additional – and sometimes also “ultimate” – possibility of drawing attention to grievances and injustices and also submitting suggestions for solving a specific problem. The occasion that prompted the Office of Technology Assessment at the German Bundestag (TAB) to conduct intensive research into the subject of electronic petitioning was the introduction of “public petitions” by the German Bundestag in 2005. These petitions are submitted electronically, published on the Internet, and can be signed and debated on the e-petition platform of the German Bundestag (see www.tab-beim-bundestag.de). Technology assessment was intensively involved in accompanying the process and development of e-petitioning at the German Bundestag by evaluating case studies from other parliaments, by empirical studies on electronic participation and, as is the main task of parliamentary technology assessment, by developing options for parliamentary action (Lindner & Riehm, 2011; Beckert et al., 2011; Riehm et al., 2014).
 9. The huge amount of new technical artifacts related with ICTs such as computers, mobile phones, and servers, which were quickly integrated into added value chains, into the workplace, into education, and into households, influences material flows at the global level as well as the need for energy supply

and thus touches upon ecological concerns. The short life-time of many of these artifacts increases related problems of waste disposal and recycling. While the *Leitbild* of sustainable development calls for more environmentally compatible production and consumption patterns, developments in real society related with ICT seem to go into the opposite direction. Initial hopes on a “de-materialisation” (Berkhout & Hertin 2004) of the economy accompanied by reduced mass flows and energy demand did realize only to a limited extent up to now. An important objective of TA in the field of ICT is, therefore, to help shaping new ICTs to better fit the requirements of sustainable development (Hilty & Ruddy 2002). The method of Life Cycle Assessment and related approaches are of high importance in this respect. Recent research and development efforts are targeting a “Green Information Technology,” in particular aiming at much higher energy efficiency. Beyond regarding individual technologies, there are attempts to better reconcile the development of IST and ICT in the overarching framework of a sustainable information society (Fuchs, 2006). Smart solutions combining technologies, services, and assumed user behavior are developed with the intention, for instance, to reduce greenhouse gas emissions, save energy, enable participation, and reduce poverty. In particular in the context of the transition of energy systems, there are high hopes that ICT solutions contribute to enhancements of energy efficiency (Schippl & Weinberger, 2009).

FUTURE RESEARCH DIRECTIONS

A major development of recent years has been the growing importance of the ideas of “Responsible Development” in scientific and technological

advance and of “Responsible Innovation” in the field of new products, services, and systems (von Schomberg, 2012). They take over the perspective of “shaping technology” mentioned above, extending it to early stages of research and development and bringing together TA, ethics, and STS research.

The notion of Responsible Innovation and related terms will receive increased attention over the next years. In particular, it will be applied to new and emerging sciences and technologies. Because the field of IST again and again creates new and ambitious visions and developments, there will be a strong focus on this field. Some illustrative examples of technological fields with strong IST involvement which need further accompanying reflective analysis are all ‘smart’ developments, the ‘Internet of everything’, the burgeoning importance ascribed to data, increasingly autonomous agents, interdependencies between ICT and other critical infrastructures, or bridging the gap between IST and the human brain, to name but a few.

CONCLUSION

From its very beginning, TA has been confronted with expectations of contributing to research, development, and innovation by adding reflexivity, by including perspectives different from those of scientists, engineers, and managers, by taking into account (even uncertain) knowledge about consequences and impacts of new science and technologies, and by transforming all these elements into advice to policy makers and society. TA has successfully developed several conceptual approaches, methods, and institutional settings to cope with those expectations. Regarding ongoing developments such as

- The further acceleration of scientific and technological advance, in particular in the field of IST,
- The rapidly growing and complexity-increasing penetration of nearly all fields

of technology and areas of life with IST (energy, health, mobility, logistics, homes, bodies, etc.),

- The increasing need for early reflection on economic and social, but also on cultural and ethical impacts of new technologies in industrialized countries,
- The also increasing claim of many citizens and organized groups in modern societies to be involved in decision-making processes on new technologies,
- The worldwide request for a more sustainable development, and
- The new roles of countries such as China, India, Brazil, and South Africa and their research and innovation systems,

it seems obvious that (1) technology assessment will have to further develop in order to be able to meet these challenges, and that (2) IST will play an even more important role in future TA compared to the past because of their cross-cutting, pervasive, and enabling character.

REFERENCES

- Beckert, B., Lindner, R., Goos, K., Hennen, L., Aichholzer, G., & Strauß, S. (2011). *E-public, e-participation and e-voting in Europe – prospects and challenges. Report. Science and Technology Options Assessment (STOA)*. Brussels: European Parliament.
- Berkhout, F., & Hertin, J. (2004). De-materialising and re-materialising: Digital technologies and the environment. *Futures*, 36(8), 903–920. doi:10.1016/j.futures.2004.01.003
- Bijker, W. E., Hughes, T. P., & Pinch, T. J. (Eds.). (1987). *The Social Construction of Technological Systems*. Cambridge, MA: Academic Press.
- Bimber, B. A. (1996). *The politics of expertise in Congress: the rise and fall of the Office of Technology Assessment*. New York: Academic Press.
- Bizer, J., Dingel, K., Fabian, B., Günther, O., Hansen, M., Klafft, M., & Spiekermann, S. et al. (2006). *Technikfolgenabschätzung Ubiquitäres Computing und Informationelle Selbstbestimmung*. Kiel, Berlin: Studie im Auftrag des Bundesministeriums für Bildung und Forschung.
- Böhle, K., Rader, M., Weber, A., & Weber, D. (2008). *Looking Forward in the ICT & Media Industries. Report for the European Parliament, Science and Technology Options Assessment*. Brussels: STOA.
- Čas, J. (2015). *SurPRISE Policy Brief. Report. Project SurPRICE*.
- Castells, M. (2000). *The Rise of the Network Society*. Oxford University Press.
- Cruz-Castro, L., & Sanz-Menendez, L. (2004). Politics and institutions: European parliamentary technology assessment. *Technological Forecasting and Social Change*, 71, 79–96.
- Decker, M., & Gutmann, M. (2012). *Robo- and Informationethics*. Zurich, Berlin: Some Fundamentals.
- Decker, M., & Ladikas, M. (2004). *Bridges between Science, Society and Policy. Technology Assessment – Methods and Impacts*. Berlin: Springer; doi:10.1007/978-3-662-06171-8
- Friedewald, M., & Raabe, O. (2010). Ubiquitous computing: An overview of technology impacts. *Telematics and Informatics*, 28(2), 55–65. doi:10.1016/j.tele.2010.09.001
- Fuchs, C. (2006). Towards a Global Sustainable Information Society (GSIS)? *tripleC*, 4(1), 40–99.

- Garcia, A. G., Velasco, C. C., Zamalloa, E. I., Velasco, E. R., Elejabarrieta, I. E., Lotero, J. H., & Schuste, S. et al. (2015). *Mass Surveillance. Part I – Risks and opportunities raised by the current generation of network services and applications. Report. Science and Technology Options Assessment (STOA)*. Brussels: European Parliamentary Research Service.
- Grunwald, A. (2009). Technology Assessment: Concepts and Methods. In A. Meijers (Ed.), *Philosophy of Technology and Engineering Sciences 9* (pp. 1103–1146). Amsterdam: Elsevier; doi:10.1016/B978-0-444-51667-1.50044-6
- Grunwald, A. (2012). *Responsible Nanobio-technology. Ethics and Philosophy*. Singapore: Panstanford; doi:10.1201/b13147
- Habermas, J. (1970). *Toward a Rational Society*. Boston: MIT press.
- Hennen, L. (2012). Why do we still need participatory technology assessment? *Poiesis und Praxis*, 9(1-2), 27–41. doi:10.1007/s10202-012-0122-5 PMID:23204994
- Hilty, L. M., & Ruddy, T. (2002). Resource Productivity in the Information Age. *Indicators of Sustainable Development. Futura*, 21(2), 76–84.
- Joss, S., & Belucci, S. (2002). *Participatory Technology Assessment – European Perspectives*. London: Academic Press.
- Leimbach, T., Hallinan, D., Bachlechner, D., Weber, A., Jaglo, M., Hennen, L., & Hunt, G. et al. (2014). *Potential and Impacts of Cloud Computing Services and Social Network Websites. Report. Science and Technology Options Assessment (STOA)*. Brussels: European Parliamentary Research Service.
- Lindner, R., & Riehm, U. (2011). Broadening participation through e-petitions? An empirical study of petitions to the German Parliament. *Policy & Internet*, 3(1). doi:10.2202/1944-2866.1083
- Nentwich, M., & König, R. (2012). *Cyberscience 2.0. Research in the age of digital social networks*. Frankfurt a.M.: Academic Press.
- Orwat, C., Graefe, A., & Faulwasser, T. (2008). Towards pervasive computing in health care – a literature review. *BMC Medical Informatics and Decision Making*, 8(26). doi:10.1186/1472-6947-8-26 PMID:18565221
- Orwat, C., & Grunwald, A. (2005). Informations- und Kommunikationstechnologien und Nachhaltige Entwicklung. In B. St. Mappus et al. (Eds.), *Erde 2.0 – Technologische Innovationen als Chance für eine Nachhaltige Entwicklung?* (pp. 242–273)., doi:10.1007/3-540-27113-9_11
- Petermann, T., Bradke, H., Lüllmann, A., Poetzsch, M., & Riehm, U. (2011). *What happens during a blackout – Consequences of a prolonged and wide-ranging power outage. Report*. Berlin: Office of Technology Assessment at the German Bundestag.
- Ricci, A., Faberi, S., Brizard, N., Bougnoux, B., Degel, M., Velte, D., & Garcia, E. (2012). *Smart grids/Energy grids. The techno-scientific developments of smart grids and the related political, societal and economic implications. Report. Science and Technology Options Assessment (STOA)*. Brussels: European Parliament.
- Riehm, U., Böhle, K., & Lindner, R. (2014). *Electronic petitioning and modernization of petitioning systems in Europe. Report*. Berlin: Office of Technology Assessment at the German Bundestag.
- Rip, A., Misa, T., & Schot, J. (1995). *Managing Technology in Society*. London: Academic Press.
- Roco, M. C., & Bainbridge, W. S. (2002). *Converging Technologies for Improving Human Performance*. Arlington, VA: Academic Press.
- Rothmann, R., Sterbik-Lamina, J., & Peissl, W. (2014). *Credit Scoring in Österreich*. Wien: Institute of Technology Assessment of the Austrian Academy of Sciences.

Schippl, J., & Weinberger, N. (2009). *Assessing the potential of ICT to increase energy efficiency and fight climate change – key technologies and prospects. Report. Science and Technology Options Assessment (STOA)*. Brussels: European Parliament.

Strauß, S., & Nentwich, M. (2013). Social network sites, privacy and the blurring boundary between public and private spaces. *Science & Public Policy*, 40(6), 724–732. doi:10.1093/scipol/sct072

USGAO. (2011). *Electricity Grid Modernization: Progress Being Made on Cybersecurity Guidelines, but Key Challenges Remain to be Addressed (GAO-11-117)*. Report. Washington, DC: United States Government Accountability Office.

US GAO. (2013). *Information Resellers – Consumer Privacy Framework Needs to Reflect Changes in Technology and the Marketplace (GAO-13-663)*. Report. Washington, DC: United States Government Accountability Office.

van den Berg, M., de Graaf, P., Kwant, P. O., & Slewe, T. (2015). *Mass Surveillance. Part 2 – Technology foresight, options for longer-term security and privacy improvements. Report. Science and Technology Options Assessment (STOA)*. Brussels: European Parliamentary Research Service.

van Merkerk, R. (2007). *Intervening in Emerging Technologies – A CTA of Lab-on-a-chip technology*. Utrecht: Academic Press.

von Schomberg, R. (2012). Prospects for Technology Assessment in the 21st century: the quest for the “right” impacts of science and technology. In M. Dusseldorp et al. (Eds.), *Technikfolgen abschätzen lernen* (pp. 43–65). Wiesbaden: Academic Press.

WCED. (1987). *Our Common Future. Report by the United Nations World Commission on Environment and Development*. WCED.

Wiegerling, K. (2011). *Philosophie intelligenter Welten*. München: Wilhelm Fink.

ADDITIONAL READING

Hilty, L., Aebischer, B., Andersson, G., & Lohmann, W. 2013. *ICT4S: ICT for Sustainability. Proceedings of the First International Conference on Information and Communication Technologies for Sustainability*. Zurich

Siune, K., Markus, E., Calloni, M., Felt, U., Gorski, A., Grunwald, A.,.... (2009). *Challenging Futures of Science in Society*. Report of the MASIS Expert Group. Brussels.

Som, C., Hilty, L., & Köhler, A. (2009). The Precautionary Principle as a Framework for a Sustainable Information Society. *Journal of Business Ethics*, 85(3), 493–505. doi:10.1007/s10551-009-0214-x

KEY TERMS AND DEFINITIONS

E-Participation: Using ICTs for supporting participation processes in government and governance. The need for broader participation has emerged over the past decades predominantly in Western societies. ICTs provide means to allow a much larger number of people to participate.

Governance: The process of opinion-forming and decision-making. In nation-states, these processes and systems are traditionally administered by the government. Currently we witness a strong movement to better involvement of citizens, stakeholders, and civic society in those processes.

Information Society: The creation, distribution, use, integration, and manipulation of information is of major economic, political, and cultural importance. Its main medium is the widespread use of information technology.

Privacy: The ability of individuals or groups to control the use of personal information about themselves or to keep such information in a defined space.

Responsible Innovation: A term pointing to the need for shaping technology and innovation with respect to criteria of responsibility and human values.

Sustainable Development: The kind of development that meets the needs of the present without compromising the ability of future generations to meet their own needs (WCED, 1987).

Technology Assessment (TA): A scientific, interactive, and communicative process which aims to contribute to the formation of public and political opinion on societal aspects of science and technology (Decker & Ladikas, 2004).

H

Towards an Interdisciplinary Socio-Technical Definition of Virtual Communities

Umar Ruhi

University of Ottawa, Canada

INTRODUCTION

Originally defined as “*social aggregations on the Internet*” (Rheingold, 1993), virtual communities (VCs) refer to interactive online spaces that can potentially enable high levels of information sharing, communication, and social interactions among their members. VCs can also be described as computer supported social networks (Garton, Haythornthwaite, & Wellman, 1997) or “*gathering spots*” on the Internet where individuals and organizations can share common interests and meet differentiated user needs and wants (Baim, 2006).

Since their earliest inception, VCs have been recognized as an important facet of the digital economy, and as a critical success factor for e-Commerce (Figallo, 1998; Hagel & Armstrong, 1997; Preece, Abras, & Maloney-Krichmar, 2004; Ridings & Gefen, 2004). VCs can potentially provide value to their individual members and sponsoring organizations through a variety of internal-facing business applications, such as knowledge sharing and organizational learning, as well as external-facing online activities, such as the provision of commercial and government services (Bughin, 2007; Lee & Suh, 2015; Mačiulienė & Skaržauskienė, 2016; Petouhoff, 2009).

Due to the wide-ranging use-cases and potential benefits of VCs, they have been a subject of study in many academic disciplines, including sociology, psychology, management, communication, computer science, and information systems. The objective of this chapter is to examine various streams of research that have studied VCs, and to facilitate the reader’s understanding of

VCs through an explanation of their underlying concepts and their fundamental properties. Toward this, the chapter specifically reviews the discourse on VCs in research fields adopting a socio-technical lens of analysis, and proposes a socio-technical definition of VCs.

The discussion in this chapter starts with a characterization of VCs from an information systems perspective. This is followed by a brief synopsis of research fields that have form the basis of socio-technical investigations of VCs. Finally, the chapter reviews literature domains that draw upon these research fields and cites seminal definitions from these domains to deliberate and propose an interdisciplinary socio-technical definition of VCs.

Within the various disciplines that study VCs, many researchers often use the terms virtual communities, online communities, web communities, cyber communities, electronic communities and e-communities interchangeably to refer to the same phenomenon (Jones & Rafaeli, 2000; Schoberth, Preece, & Heinzl, 2003). For purposes of this review, our survey of literature includes research that has been conducted surrounding these various notions of VCs. Furthermore, this chapter uses the terms virtual communities and online communities interchangeably.

BACKGROUND

Despite the absence of an agreed upon definition of VCs across research studies, the presence of a technology platform that facilitates interactions

DOI: 10.4018/978-1-5225-2255-3.ch371

Table 1. Virtual community attributes as elements of an IT artifact

Defining Elements of an IT Artifact	Examples of Attributes of Virtual Communities
Information & Communication Technologies	■ Websites and portals hosting the virtual community ■ Groupspace platforms such as social networks, discussion forums, blogs, wikis, mailing lists, newsgroups. ■ Communication tools such as instant messaging and interactive chat applications
Tasks	■ Information exchange in the form of sharing knowledge, asking and answering questions, providing commentary, and collaborating on new content ■ Social interactions in the form of networking with other members, forming relationships, and seeking and providing emotional support
Task Structures	■ Administrative policies and controls ■ Interaction norms and etiquette ■ Technology usability and accessibility ■ Scope of information exchange and social interaction (diverse/narrow) ■ Extent of participation (active/passive)
Task Context	■ Member motivations and interests ■ Membership duration ■ Self-efficacy in technology use and personal knowledge and expertise ■ Member trust levels and sense of community

among members is considered to be a main characteristic of online communities (Donath, 2005; Preece, 2001a; Preece, 2001b). Various information and communication technologies (ICTs) can be used in VCs – including, websites, computer networks, email lists, Usenet newsgroups, discussion forums, Internet chat applications, and networked databases (Coon, 1998; Lapachet, 2001). Additionally, modern technology platforms such as social networking sites, weblogs (popularly known as blogs) for user generated content, and wikis for online collaboration have also been considered in the list of potential technologies that can spawn a VC (Blanchard, 2003; Brailas, Koskinas, Dafermos, & Alexias, 2015; Buss & Strauss, 2009 ; Mačiulienė & Skaržauskienė, 2016). These various technology platforms and their underlying features and functions that enable VCs have been studied by information systems (IS) researchers over a long time. In this section, we offer a characterization of VCs from an IS perspective.

IS research can be considered as the study of the effective use of ICTs and their potential impact on human, organizational, and social world (Gregor, 2006; Hirschheim & Klein, 2003; Khazanchi & Munkvold, 2000). At the theoretical core of IS research is the concept of an *IT artifact*, which

refers to ICTs that act as enabling infrastructure for people and organizations in driving their individual activities and business processes, and ultimately affecting their overall performance and satisfaction levels (Bacon & Fitzgerald, 2001). In viewing VCs as IT artifacts, we adopt Benbasat & Zmud's (2003) normative perspective of the field of IS, and their delineation of the scope of IS research as relating to the IT artifact's immediate nomological network (capabilities, perceptions, uses, practices, behaviour, and impacts linked to the IT artifact). Using this orientation, VCs can be examined in terms of their underlying ICTs and their use by members, while also considering the context and structure of tasks embedded within the VCs and performed by their members. To illustrate the conceptualization of VCs as IT artifacts, Table 1 provides examples of attributes of VCs that align with the defining elements of IT artifacts.

Based on the conceptualization of VCs as IT artifacts, it can be seen that the task structure and task context elements of the IT artifact have a strong sociological basis. To effectively address this dual sociological and technological nature of the IT artifact pertaining to VCs, a socio-technical perspective has been strongly recommended in academic research (Dannecker,

Leimeister, Konana, & Rajagopalan, 2007; Koh, Kim, Butler, & Bock, 2007) and industry practice (Gossieaux & Moran, 2008; Petouhoff, 2009). The socio-technical perspective in IS research is based on the premise that information systems comprise behavioural as well as technological sub-systems that need to be addressed in unison while investigating pertinent IS related phenomena (Becker, Niehaves, & Janiesch, 2007; Lee, 2000). Our ultimate aim in this chapter is to propose a socio-technical definition of VCs that draws upon various streams of research literature.

THEORETICAL UNDERPINNINGS IN SOCIO-TECHNICAL RESEARCH ON VIRTUAL COMMUNITIES

VCS and their underlying facets have been investigated in various research fields, including, sociology, psychology, computer science, information systems, information science, communication studies, economics, management science, and marketing management. In a bibliometric study, Laine (2009) found that the greatest number of publications related to VCs were in the technology related research fields of information science, information systems, and computer science, followed by publications in business and management. Similarly, in a more extensive search across multiple electronic databases, Iriberry and Leroy (2009) found that studies of VCs were more common in information systems and business research areas, but that a strong grounding in the psychology and sociology extant literature also existed.

While a comprehensive coverage of VC research in all these fields is beyond the scope of this chapter, we specifically focus on fields that form the basis of socio-technical research in the extant VC literature. Toward this, based on the commonalities in themes and focus of different VC studies, we classify the extant research on VCs into four categories, namely, i) sociological/psychological; ii) technological; iii) business/management; and iv) economic perspectives.

Table 2 below presents a summary of a sample of studies from these four perspectives by outlining the core issues and common theoretical constructs addressed by each, as well as sample references and top cited publications¹. A brief synopsis of each of the four perspectives follows.

The *sociological/psychological* research perspective of VCs is primarily concerned with the social structure of collectives and the nature of relationships within those collectives. In deliberating social practices and outcomes of participation in VCs, this perspective makes use of concepts from human cognitive and affective processes that lead to the development of psychological trust and sense of community among members. For instance, Etzioni & Etzioni (1999) examine aspects of bonding and culture in a VC and define it through two attributes, namely, a web of affect-laden relationships encompassing group of individuals (bonding), and commitment to a set of shared values, norms, meanings and a shared historical identity (culture). Frameworks based on social capital theory (Coleman, 1988; Jacobs, 1965; Putnam, 1995) and social cognitive theory (Bandura, 1986; Compeau & Higgins, 1995) are frequently used to discuss the development of ties through relational capital including trust, as well as cognitive capital through increased self-efficacy over the course of participation in a virtual community. From a social psychology viewpoint, researchers have also used the concept of psychological sense of community (McMillian & Chavis, 1986) to deliberate the presence of a similar affective and attitudinal construct in the context of VCs (Blanchard, 2007; Blanchard & Markus, 2002; Obst & White, 2004). VC research shows sense of community to be an essential factor in the success and sustainability of VCs (Bagozzi & Dholakia, 2002; Blanchard, 2007; Preece et al., 2004).

Researchers investigating VCs from a *technological* research perspective refer to them in terms of the functionality of the software supporting the interactions in the community (Flake, Lawrence, Giles, & Coetzee, 2002; Lazar, Tsao, & Preece,

Table 2. Research perspectives and principal topics in virtual community studies

Research Perspective	Core Issues	Common Theoretical Constructs	Sample References
Sociological / Psychological	■ Social structure of virtual community environment ■ Factors affecting online social processes and relationships ■ Comparisons with face-to-face communities ■ Cognitive and affective processes of trust building and psychological bonding among community members	■ Social Capital (Structural, Relational, Cognitive) ■ Social Ties (Centrality, Degree, Density etc.) ■ Member Motivations and Social Behaviour ■ Sense of Community	■ Rheingold (1993) ■ Wellman et al. (1996) ■ Romm et al. (1997) ■ Etzioni & Etzioni (1999) ■ Kollock (1999) ■ Wellman & Gulia (1999) ■ Turner et al. (2001) ■ Blanchard & Markus (2003) ■ Eysenbach et al. (2004) ■ Obst (2004) ■ Ling et al. (2005)
Technological	■ Use of appropriate technology platforms for online interactions ■ Interactivity and usability of virtual community interface ■ Supporting sociability through technology tools and features	■ Usability and Sociability ■ Technology Acceptance Factors (Ease of Use, Usefulness, Intention to Use) ■ IS Success Factors (System Quality, Information Quality, System Use, User Satisfaction)	■ Erickson & Kellogg (2000) ■ Ridings et al. (2002) ■ Preece (2001b) ■ Flake et al. (2002) ■ Yoo et al. (2002) ■ Ridings & Gefen (2004) ■ Moore & Serva (2007) ■ Lin & Lee (2006) ■ Wise et al. (2006) ■ Jin et al. (2007) ■ Chen (2007)
Business / Management	■ Fostering relationships with consumers through online communities ■ Leveraging positive consumer behaviour in online customer reference groups ■ Engaging customers and employees to contribute in business functions such as R&D and Sales	■ Consumer Behaviour ■ Loyalty and Commitment ■ Purchase Decisions ■ Knowledge Contribution ■ Co-creation of products and services ■ Competitive Advantage	■ Armstrong & Hagel (1996) ■ Kozinets (1999) ■ McWilliam (2000) ■ Kozinets (2002) ■ Bauer & Grether (2005) ■ Wasko & Faraj (2005; 2000) ■ Lu et al. (2008) ■ Porter & Donthu (2008) ■ Petina et al. (2008)
Economic	■ Sustainability of virtual communities as economic entities within a social context ■ Study of virtual community members as rational utility-maximizers ■ Mechanisms for value creation and resource conversion into benefits for collectives	■ Return on Investment ■ Membership Size ■ Communication Activity Levels ■ Community Growth	■ Hagel & Armstrong (1997) ■ Kollock (1999) ■ Balasubramanian and Mahajan (2001) ■ Butler (2001) ■ Cothrel (2000b) ■ Gu & Jarvenpaa (2003)

1999), as well as issues of usability of the application interface (Chen, 2007; Wise et al., 2006). To bring balance to this approach, through her seminal research work on VCs, Preece (2001c; 2004; 2003) urges system designers to concentrate on sociability aspects of VCs to help understand the social interaction requirements and preferences of community members. Parallel streams of research adopting the technological research perspective exploit the application of more well-established IS models to investigate member participation.

Theoretical constructs from the technology acceptance model (TAM) (Davis, 1989) have been used repeatedly in various contexts to link the constructs of perceived usefulness and perceived ease of use of VC technologies to a member's behavioural intention to continue participation in the VC (Hsu & Lu, 2007; Huang, Hsu, & Yen, 2008; Lin, 2006; Lin, Chiu, & Tsai, 2008). Similarly, constructs of information quality and system quality from the information systems success model (DeLone & McLean, 1992, 2003; Petter, DeLone, & McLean,

2008) have been conceptualized and corroborated to affect the overall satisfaction with and use of the VC (Lin, 2008; Yoo et al., 2002).

Through the *business/management* research perspective of VCs, studies explore applications of online communities in a multitude of business functions including marketing, sales, research and development (R&D), and human resources (Armstrong & Hagel, 1996; McWilliam, 2000; Spaulding, 2009). In an integrated framework to highlight the benefits of VCs in a business context, Spaulding (2009) used the theory of social contract to explain the applications of VCs in each activity across a firm's value chain (Porter, 1980). More specifically, marketing related studies on VCs focus on brand communities and online consumer reference groups. The development of communal bonds among members of a VC to help promote a firm's products and services are a common area of discourse in the extant research literature (Dholakia, Bagozzi, & Pearo, 2004; Kozinets, 1999). In terms of applications for R&D, the value of open innovation communities in improving and advancing products and service offerings has also been discussed (Fuller, Bartl, Ernst, & Muhlbacher, 2004; Ramaswamy, 2009). Finally, a major area of research in the business/management perspective relates to knowledge sharing and contributions in online communities of interest and electronic networks of practice (Wasko & Faraj, 2005; Wasko & Faraj, 2000). More recently, researchers are starting to explore the paradigm of collaborative intelligence in VCs through studies of new technology platforms such as wikis and other social media applications (Brailas et al., 2015; Hoegg, Meckel, Stanoevska-Slabeva, & Martignoni, 2006; Lee & Lan, 2007).

Lastly, under the *economic* research perspective, VCs are viewed as economic entities embedded in a social environment in which value creation and resource utilization aspects can be improved through greater member contributions of informational content (Balasubramanian & Mahajan, 2001). This perspective assumes the actions of VC constituents conform to neoclassi-

cal principles that view society as a collective of egoist entrepreneurs with individual rationality and incentive compatibility (Gu & Jarvenpaa, 2003). As an example of the treatise of VCs from an economic research perspective, consider the research exposition by Balasubramanian and Mahajan (2001) where the authors argue that it is the strength of social interactions and focus on specific market activities in VCs that can help develop economies of scale and lead to a competitive advantage for the sponsoring firm. Another theoretical construct frequently seen in studies pertaining to the economic research perspective is membership size, where resource availability is regarded as directly proportional to membership size and a critical mass of members is required to sustain the interactions in a VC over the long term (Butler, 2001). The concept of return on investment (ROI) has also been discussed using the economic research perspective. Here, the ROI of a VC is viewed as a combination of factors, including: the incremental value of a business (with and without the VC), the size of the community membership, the magnitude of active regular interactions, and the rate of conversion of site visitors to members who engage in commercial transactions (Cothrel, 2000b).

SOCIO-TECHNICAL LITERATURE DOMAINS DISCUSSING VIRTUAL COMMUNITIES

Towards helping us put forward a holistic socio-technical definition of VCs, this section presents a summary of four interdisciplinary literature domains that draw upon some of the theoretical concepts from research disciplines identified in the previous section. These include: i) Computer Mediated Communication (CMC), ii) Community Informatics (CI), iii) Knowledge Management (KM), and iv) Internet Marketing (IM)². An overview of the scope and exposition of research in these four domains is provided below.

Researchers who approach VCs from a *computer-mediated communication (CMC)* standpoint typically conceptualize them as technology tools that facilitate human communication via computers. Primarily based on a technological research perspective, they concentrate their efforts in describing different forms of synchronous, asynchronous or real-time interaction tools used in online communities to exchange text, images, audio and video (Hiltz, 1998; Morris & Ogan, 1996; Olson & Olson, 1998; Sproull & Kiesler, 1991). Additionally, CMC research has also answered the call for the inclusion of sociability (Preece, 2001a) in the overall treatise of virtual community specifications, and there is a trend towards recommending solutions that incorporate sociability as well as usability in the design of VCs (Donath, 2005; Preece, 2001a; Preece, 2001b). Overall, the viewpoints adopted in CMC research can be seen as being geared towards the development of VCs from a human computer interaction (HCI) perspective, as they emphasize the functionality and usability aspects of the interaction media in VCs while taking into account the alignment of these technologies with the sociability objectives of the VC constituents.

The fast emerging interdisciplinary field of *community informatics (CI)* pertains to the design and application of ICTs that can be utilized to enable community processes and achieve community objectives, including a community's social, economic, cultural, or political goals (Gurstein, 2000, 2004). By that token, VCs are often regarded as specific applications in community informatics (Marshall, Taylor, & Yu, 2004; Pitkin, 2001). On a macro level, CI researchers have studied the applicability of VCs and related technologies to facilitate the progress of local, regional and national communities, with frequent discussions of ideas of civic participation, economic development, education and learning, and community service delivery (Gurstein, 2000; Keeble & Loader, 2001; McIver, 2003; Taylor, 2004). Additionally, recent literature in CI has significantly contributed to the body of knowledge outlining success factors

and best practice models for the administration of VC projects (Carroll & Rosson, 2007; Romm & Taylor, 2001). As such, researchers in the field of CI stress the importance of sociological as well as technological ingredients in the overall success of VCs which makes this literature area a worthwhile source of information on socio-technical perspectives of VCs.

Knowledge Management (KM) related research on VCs has primarily explored the effectiveness of different knowledge sharing processes in various online community settings, including virtual communities of practice (Davies, Duke, & Sure, 2004; Droschl, 2004; Ruhi, 2008; Wasko & Faraj, 2005; Wasko & Faraj, 2000), online learning communities (OLCs) (Cothrel, 2000a; Palloff & Pratt, 1999), and online knowledge communities (OKCs) (Bieber et al., 2002; de Vries & Kommers, 2004). Common to the research discourse for these different types of VCs is the discussion on processes and technologies that facilitate information and knowledge sharing. The sharing of tacit knowledge in such communities is often considered to be a socio-psychological process for which effective social activities need to be instigated and sustained as an integral part of the community's institutionalization. For instance, success of knowledge contributions is often framed within the context of social capital constructs such as structural, relational and cognitive capital, and studies discuss how different technology and management interventions can be adopted to improve these social capital dimensions to enable higher level and better quality knowledge contributions (Wasko & Faraj, 2005; Wasko & Faraj, 2000). Hence, a great number of research expositions found in the extant KM literature on VCs also adopt a socio-technical viewpoint of interactions and activities in online communities, and this literature domain constitutes an effective basis for a definition of VCs germane to a socio-technical lens of analysis.

Finally, as an applied field of research, *Internet Marketing (IM)* has been at the forefront of VC research since the earliest conception of

online communities in the commercial realm. Content, Community, Commerce – the 3 Cs of electronic commerce as they are widely known in the internet marketing literature, have together been regarded as the critical success factors that businesses need to manage in order to prosper in online marketing and sales (Chaffey, Mayer, Johnston, & Ellis-Chadwick, 2002). Jagannathan et al. (2002) refer to community as an *“important driver leading to a condition where e-Commerce can occur”*. The focus of discussions in the IM literature is primarily concerned with the generation and maintenance of customer traffic while finding new ways to penetrate current market segments and discover new ones. Recent explorations of consumer reference groups and online customer service communities have focused on motivational factors that facilitate the involvement of customers in various business processes across the sponsoring firm’s value chain (Goel & Mousavudin, 2007; Petouhoff, 2009). Practices related to improving user loyalty, enabling better feedback on products and services, accessing beta testers and early adopters, and promoting word-of-mouth have produced valuable guidelines and principles that relate the role of technology to sociological and psychological consumer behaviour processes (Araujo, 1999). Therefore, these studies from the internet marketing literature domain can also prove useful in proposing a socio-technical characterization of VCs.

AN INTERDISCIPLINARY SOCIO-TECHNICAL DEFINITION OF VIRTUAL COMMUNITIES

From the review of the selected research perspectives and literature domains in the previous sections, it is apparent that although researchers from these subject areas offer discourse-specific descriptions of VCs and their applications, there is a distinct socio-technical undertone to their treatise of VCs. As such, these streams of research offer an adequate basis from which to devise an

interdisciplinary socio-technical definition of VCs. In deliberating such a definition, a review of some of the definitions of VCs from the extant literature in the four areas above was conducted to reconcile and elaborate the primary notions of VCs. As a summary, various definitions proposed by researchers in the four literature domains are contrasted in Table 3.

Among the plethora of definitions that were found, those presented in Table 3 were carefully selected to reflect the viewpoints across the four literature domains. These definitions also vary in the style of their attribution – while some provide ascriptive characterizations of VCs, others are based on their descriptive representations. Researchers employing ascriptive portrayals of VCs focus on their underlying social constructions including the motivations of the members, and the interactions and exchanges among them. This is in contrast to the descriptive accounts of VCs which primarily emphasize their functional goals, often in a technophilic context, highlighting technology features of VCs over their social interaction aspects.

Despite the miscellany of depictions of virtual communities seen in the definitions above, several similarities can be highlighted. These similarities are briefly noted below to conceptualize our proposed definition of VCs.

The first similarity is the aggregation of *“participants”* (the first component of the proposed definition) who share a common *“purpose”* (the second component of the proposed definition). The term participants is used to incarnate both individuals as well as organizations who may partake different roles in a VC. Further, the concept of interaction is logically embedded in the term participation as well. The interactions in a VC revolve around a common purpose in VCs. This common purpose can subsist in a VC in the form of *“shared interests, expertise, problems, ideas, or passions”* (Nilan, Zakari, Guzman, & Zakaria, 2004).

A second similarity is that interactions between members of a virtual community are me-

diated through various technologies. Hence, the presence of a mediating technology “*platform*” constitutes the third component of the proposed definition. Note that most researchers realize the importance of non-computer based technologies in mediating interactions among community members (De Souza & Preece, 2004; Kollock, 1998; Rheingold, 2000). These technologies can include various audio/video conferencing tools, as well as advanced mobile technologies. Hence, the proposed definition of a technology platform is not restricted solely to computer based technologies.

Thirdly, a salient feature that can be noted from Preece’s original (2001a) and subsequent (2001b; Preece et al., 2004) definitions, is the notion of norms and policies. Sociologists assert that the policies and norms in a VC are in fact an ecological extension from traditional real-life communities (Brint, 2001; Valtersson, 1996). Hence, the idea of “*protocols*” constitutes the next essential component of the proposed definition.

Last, an important characteristic of VCs that has recently come under intense discussion is the notion of a sense of belonging or a sense of com-

Table 3. Ten Seminal definitions of virtual communities from various literature domains

Source	Definition	Related Literature Streams			
		CMC	CI	KM	IM
Bagozzi & Dholakia (2002)	Virtual communities [are] mediated social spaces in the digital environment that allow groups to form and be sustained through ongoing communication processes.				●
Bieber et al. (2002)	[A] virtual community include[s] anyone actively interested in, or associated with, a group formed around a particular domain of interest. Dispersed or local, the community requires electronic support to implement a continuous meta improvement strategy in its services.			●	
Cothrel & Williams (1999)	An online community is a group of people who use computer networks as their primary mode of interaction.			●	
Gurstein et al. (2002)	Virtual communities serve a diversity of groups, including people with common interests; groups fostered by particular organizations, industries or marketplaces; those who face similar life circumstances; as well as those who simply wish to socialize, play games or participate in fantasy experiences together on-line.		●		
Hagel & Armstrong (1997)	Virtual communities are computer-mediated spaces where there is a potential for an integration of content and communication with an emphasis on member-generated content.	●			●
Jones & Rafaeli (2000)	Virtual Publics are symbolically delineated computer-mediated spaces, whose existence is relatively transparent and open, that allow groups of individuals to attend and contribute to a similar set of computer-mediated interpersonal interactions.		●		
Lee et al. (2003)	[A virtual community] is a cyberspace supported by computer-based information technology, centered upon communication and interaction of participants to generate member-driven contents, resulting in a relationship being built up.		●		
Plant (2003)	A collective group of entities, individuals or organizations that come together either temporarily or permanently through an electronic medium to interact in a common problem or interest space	●			
Preece (2001a)	An online community consists of people, who interact socially; a shared purpose; policies; and computer systems, to mediate social interaction.	●			
Rheingold (1993, 2000)	[Virtual communities are] social aggregations that emerge from the Net when enough people carry on those public discussions long enough, with sufficient human feeling, to form webs of personal relationships in cyberspace.	●			

munity among VC members (Bagozzi & Dholakia, 2002; Blanchard & Markus, 2003; Koh & Kim, 2003). Also described as a consciousness of kin that acts as an intrinsic connection to other members, researchers attribute this sense of affiliation with the shaping of a member's behaviours in the online community (Kleinman, 2000; Wellman & Gulia, 1999). Feelings of belongingness and self-identity have been regarded as important elements in instilling a psychological ownership among VC members, subsequently leading to improved member outcomes in VCs (Lee & Suh, 2015). Therefore, the "*persona*" that a VC instils among its members to make them feel part of the community and help them express their self-identity constitutes the final component in the proposed definition. Therefore, the "*persona*" that a VC instils among its members to make them feel part of the community constitutes the final component in the proposed definition.

Together, these salient attributes of a virtual community (the 5 Ps discussed above)³ are assembled into components of a nascent definition along the lines of an ascriptive socio-technical

characterization of VCs: *A virtual community is an aggregation of "participants" (individuals or organizations) with a shared "purpose", who have their interactions mediated through a technology "platform" and guided by collective "protocols", and who embrace a cooperative membership "persona" in their online interactions.*

Figure 1 depicts the components of the proposed definition as interlocking gears driving one another, and the interactions constituting member participation in the VC. It is hoped that the proposed definition and its various underlying components can act as a touchstone to guide future research expositions on VC. Some suggestions for the use of this definition are provided in the next section.

FUTURE RESEARCH DIRECTIONS

The definition of VCs proposed in this chapter recognizes that social and technical interaction features in online communities are largely interwoven, and together, these factors play an important role in determining the uses and outcomes of VC initiatives. By drawing attention to various socio-technical facets of VCs, it is hoped that the components in the proposed definition can help VC researchers adopt amongst a multitude of outlooks while investigating VCs.

For IS research in particular, our definition espouses an IT artifact viewpoint of VCs, and it includes various aspects linked to the IT artifact's immediate nomological network. The platform component links to the capabilities of the artifact, while the participants and purpose component characterizes various uses, practices, and behaviour linked to VCs. Similarly, the persona and protocols components of the definition symbolize the perceptions and impacts aspects linked to the use of the IT artifact. Overall, the proposed definition can potentially allow an investigation of VCs in terms of their underlying ICTs and their use by members, while also considering the context and structure of tasks embedded within the VCs and performed by their members.

Figure 1. Components of a socio-technical definition of virtual communities



Our proposed definition can also be utilized as a conceptual framework to provide an analytical foundation for studying member participation in VCs. By combining aspects of social processes as well as attributes of technology artifacts, our definition can complement other analytical frameworks for VCs such as the *online community framework (OCF)* by De Souza & Preece (2004) (also see Akoumianakis, 2008), and the *evidential conceptual model of online community* by McArthur & Bruza (2001). Both these frameworks include broad-based thematic concepts related to the goals and needs of VC members, technology infrastructure requirements for VCs, and human interaction processes in VCs. Our proposed definition also incorporates all of these elements. Therefore, through this definition, we hope to provide an alternative or complementary abstraction of a VC in terms of its basic building blocks.

On the whole, the proposed definition offers a structured and holistic socio-technical lens of analysis for future research pertaining to VCs. Subject areas related to technology development, functionality adoption, implementation and outcome assessment, or practice institutionalization can be facilitated through a framework based on this definition.

CONCLUSION

This chapter offers a summary of various research disciplines and literature domains that investigate VCs from a socio-technical perspective. Subsequently, definitions from seminal research studies are synthesized to propose an interdisciplinary socio-technical definition of VCs.

The proposed definition of VCs can potentially act as a high-level abstraction of online communities – highlighting their five core facets, namely, participants, purpose, platform, protocols, and persona. Together, these facets constitute a structured and holistic socio-technical lens of analysis

that can be used in future research pertaining to VCs. Furthermore, through a focus on one of more underlying elements, the proposed definition can serve as a useful conceptual aid for: i) analyzing the enabling and inhibiting factors in the evolution of VCs; ii) examining different types of member and technology interactions that develop in VCs over time; iii) investigating factors that affect the uptake and outcomes of VC initiatives; and iv) exploring key success factors for effective governance and operation of VCs.

REFERENCES

- Akoumianakis, D. (2008). Designing an 'Electronic Village' of Local Interest on Tourism: The eKoNES Framework. *Interdisciplinary Journal of Information, Knowledge, and Management*, 3, 115–134.
- Araujo, L. d. (1999). *Gaming and virtual communities: Helping a community to build itself*. Paper presented at the Second International Conference on Virtual Communities, Bath, UK.
- Arguello, J., Butler, B. S., Joyce, E., Kraut, R., Ling, K. S., Rosé, C., & Wang, X. (2006). *Talk to me: Foundations for successful individual-group interactions in online communities*. Academic Press.
- Armstrong, A. G., & Hagel, J. (1996). The real value of on-line communities. *Harvard Business Review*, 74(3), 134–141.
- Bacon, C. J., & Fitzgerald, B. (2001). A systemic framework for the field of information systems. *The Data Base for Advances in Information Systems*, 32(2), 47. doi:10.1145/506732.506738
- Bagozzi, R. P., & Dholakia, U. M. (2002). Intentional Social Action in Virtual Communities. *Journal of Interactive Marketing*, 16(2), 2–21. doi:10.1002/dir.10006

- Baim, S. A. (2006). Business Applications of Virtual Communities. In S. Dasgupta (Ed.), *Encyclopedia of Virtual Communities and Technologies* (pp. 13–17). Hershey, PA: Idea Group. doi:10.4018/978-1-59140-563-4.ch004
- Balasubramanian, S., & Mahajan, V. (2001). The economic leverage of the virtual community. *International Journal of Electronic Commerce*, 5(3), 103–138.
- Bandura, A. (1986). *Social Foundations of Thought and Action: A Social Cognitive Theory*. Englewood Cliffs, NJ: Prentice Hall.
- Bauer, H. H., & Grether, M. (2005). Virtual Community. *Journal of Relationship Marketing*, 4(1), 91–109. doi:10.1300/J366v04n01_07
- Becker, J., Niehaves, B., & Janiesch, C. (2007). Socio-Technical Perspectives on Design Science in IS Research. *Advances in Information Systems Development*, 127-138.
- Benbasat, I., & Zmud, R. W. (2003). The identity crisis within the IS discipline: Defining and communicating the discipline's core properties. *Management Information Systems Quarterly*, 27(2), 183–194.
- Bieber, M., Engelbart, D., Furuta, R., Hiltz, S. R., Noll, J., Preece, J., & Walle, B. V. d. et al. (2002). Towards virtual community knowledge evolution. *Journal of Management Information Systems*, 18(4), 11–35.
- Blanchard, A. L. (2003). *Blogs as Virtual Communities: Identifying a Sense of Community in the Julie/Julie Project*. Retrieved from http://blog.lib.umn.edu/blogosphere/blogs_as_virtual.html
- Blanchard, A. L. (2007). Developing a sense of virtual community measure. *Cyberpsychology & Behavior*, 10(6), 827–830. doi:10.1089/cpb.2007.9946 PMID:18085972
- Blanchard, A. L., & Markus, M. L. (2002). *Sense of Virtual Community—Maintaining the Experience of Belonging*. Paper presented at the 35th Annual Hawaii International Conference on System Sciences (HICSS-35'02), Waikoloa, HI.
- Blanchard, A. L., & Markus, M. L. (2003). The experienced “sense” of a virtual community: Characteristics and processes. *The Data Base for Advances in Information Systems*, 35(1).
- Brailas, A., Koskinas, K., Dafermos, M., & Alexias, G. (2015). Wikipedia in Education: Acculturation and learning in virtual communities. *Learning. Culture and Social Interaction*, 7, 59–70. doi:10.1016/j.lcsi.2015.07.002
- Brint, S. (2001). Gemeinschaft revisited: A critique and reconstruction of the community concept. *Sociological Theory*, 19(1), 1–23. doi:10.1111/0735-2751.00125
- Bughin, J. R. (2007). How Companies Can Make The Most of User-Generated Content. *The McKinsey Quarterly*, (August), 2007.
- Buss, A., & Strauss, N. (2009). *Online Community Handbook: Building Your Business and Brand on the Web*. Berkeley, CA: New Riders.
- Butler, B. S. (2001). Membership size, communication activity, and sustainability: A resource-based model of online social structures. *Information Systems Research*, 12(4), 346–362. doi:10.1287/isre.12.4.346.9703
- Carroll, J. M., & Rosson, M. B. (2007). Participatory design in community informatics. *Design Studies*, 28(3), 243–261. doi:10.1016/j.destud.2007.02.007
- Chaffey, D., Mayer, R., Johnston, K., & Ellis-Chadwick, F. (2002). *Internet Marketing: Strategy, Implementation and Practice*. New York, NY: FT Prentice Hall.

- Chen, I. Y. L. (2007). The factors influencing members continuance intentions in professional virtual communities a longitudinal study. *Journal of Information Science*, 33(4), 451–467. doi:10.1177/0165551506075323
- Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology*, 94(S1), 95. doi:10.1086/228943
- Compeau, D. R., & Higgins, C. A. (1995). Application of social cognitive theory to training for computer skills. *Information Systems Research*, 6(2), 118–142. doi:10.1287/isre.6.2.118
- Coon, D. A. (1998). *An Investigation of #Friends Internet Relay Chat as a Community*. Manhattan, KS: Kansas State University.
- Cothrel, J. (2000a). Measuring the success of an online community. *Strategy and Leadership*, 28(3), 17–21. doi:10.1108/10878570010341609
- Cothrel, J., & Williams, R. L. (1999). On-line communities: Helping them form and grow. *Journal of Knowledge Management*, 3(1), 54–60. doi:10.1108/13673279910259394
- Cothrel, J. P. (2000b). Measuring the success of an online community. *Strategy and Leadership*, 28(2), 17–21. doi:10.1108/10878570010341609
- Dannecker, A., Leimeister, J.-M., Konana, P., & Rajagopalan, B. (2007). *Towards a Research Agenda for Virtual Communities*. Paper presented at the Thirteenth Annual Americas Conference on Information Systems (AMCIS 2007).
- Davies, J., Duke, A., & Sure, Y. (2004). OntoShare - An Ontology-based Knowledge Sharing System for Virtual Communities of Practice. *Journal of Universal Computer Science*, 10(3), 262–283.
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *Management Information Systems Quarterly*, 13(3), 319–340. doi:10.2307/249008
- De Souza, C. S., & Preece, J. (2004). A framework for analyzing and understanding online communities. *Interacting with Computers*, 16(1), 579–610. doi:10.1016/j.intcom.2003.12.006
- de Vries, S., & Kommers, P. (2004). Online knowledge communities: Future trends and research issues. *Int. J. Web Based Communities*, 1(1), 115–123. doi:10.1504/IJWBC.2004.004804
- DeLone, W. H., & McLean, E. R. (1992). Information systems success: The quest for the dependent variable. *Information Systems Research*, 3(1), 60–95. doi:10.1287/isre.3.1.60
- Delone, W. H., & McLean, E. R. (2003). The DeLone and McLean model of information systems success: A ten-year update. *Journal of Management Information Systems*, 19(4), 9–30.
- Dholakia, U. M., Bagozzi, R. P., & Pearo, L. K. (2004). A social influence model of consumer participation in network-and small-group-based virtual communities. *International Journal of Research in Marketing*, 21(3), 241–263. doi:10.1016/j.ijresmar.2003.12.004
- Donath, J. (2005). Sociable Media. In C. Ghaoui (Ed.), *Encyclopedia of Human Computer Interaction*. Hershey, PA: Idea Group Publishing.
- Droschl, G. (2004). Communities of Practice: An Integrated Technology Perspective. *Journal of Universal Computer Science*, 10(3), 284–293.
- Erickson, T., & Kellogg, W. A. (2000). Social translucence: An approach to designing systems that support social processes. *ACM Transactions on Computer-Human Interaction*, 7(1), 59–83. doi:10.1145/344949.345004
- Etzioni, A., & Etzioni, O. (1999). Face-to-face and computer-mediated communities, a comparative analysis. *The Information Society*, 15(4), 241–248. doi:10.1080/019722499128402

- Eysenbach, G., Powell, J., Englesakis, M., Rizo, C., & Stern, A. (2004). *Health related virtual communities and electronic support groups: systematic review of the effects of online peer to peer interactions* (Vol. 328). BMJ Publishing Group Ltd.
- Figallo, C. (1998). *Hosting Web communities: Building relationships, increasing customer loyalty, and maintaining a competitive edge*. New York, NY: John Wiley & Sons, Inc.
- Flake, G. W., Lawrence, S., Giles, C. L., & Coetzee, F. M. (2002). Self-organization and identification of web communities. *Computer*, 35(3), 66–70. doi:10.1109/2.989932
- Fuller, J., Bartl, M., Ernst, H., & Muhlbacher, H. (2004). *Community Based Innovation: A Method to Utilize the Innovation Potential of Online Communities*. Paper presented at the 37th Hawaii International Conference on System Sciences (HICSS 2004), Waikoloa, HI. doi:10.1109/HICSS.2004.1265464
- Garton, L., Haythornthwaite, C., & Wellman, B. (1997). Studying Online Social Networks. *Journal of Computer-Mediated Communication*, 3(1). Retrieved from <http://www.ascusc.org/jcmc/vol3/issue1/garton.html>
- Goel, L., & Mousavidin, E. (2007). vCRM: Customer Relationship Management. *The Data Base for Advances in Information Systems*, 38(4), 56–60. doi:10.1145/1314234.1314245
- Gossieaux, F., & Moran, E. (2008). *Tribalization of Business*. Retrieved from <http://www.beelinelabs.com/tribalization/charts/>
- Gregor, S. (2006). The nature of theory in information systems. *Management Information Systems Quarterly*, 30(3), 611.
- Gu, B., & Jarvenpaa, S. (2003). Online discussion boards for technical support: The effect of token recognition on customer contributions. *International Conference on Information Systems (ICIS) 2003 Proceedings*, 10.
- Gurstein, M. (2000). *Community informatics: enabling communities with information and communications technologies*. Hershey, PA: Idea Group Publishing. doi:10.4018/978-1-878289-69-8
- Gurstein, M. (2002). *Weaving Community With Community Fibre: Community Informatics and The Broadband Revolution*. Paper presented at the 35th Annual Hawaii International Conference on System Sciences (HICSS-35'02), Waikoloa, HI.
- Gurstein, M. (2004). Effective Use and the Community Informatics Sector: Some Thoughts on Canada's Approach to Community Technology/Community Access. In M. Moll & L. R. Shade (Eds.), *Communications In the Public Interest Vol. 2: Seeking Convergence in Policy and Practice* (pp. 223-243). Ottawa: Canadian Centre for Policy Alternatives.
- Hagel, J. I., & Armstrong, A. G. (1997). *Net Gain: Expanding Markets through Virtual Communities*. Boston, MA: Harvard Business School Press.
- Hiltz, S. R. (1998). *Collaborative learning in asynchronous learning networks: Building learning communities*. Paper presented at the WEB98, Orlando, FL.
- Hirschheim, R. A., & Klein, H. K. (2003). Crisis in the IS Field? A Critical Reflection on the State of the Discipline. *Journal of the Association for Information Systems*, 4(1), 237–293.
- Hoegg, R., Meckel, M., Stanoevska-Slabeva, K., & Martignoni, R. (2006). *Overview of business models for Web 2.0 communities*. Paper presented at the GeNeMe (Gemeinschaften in Neuen Medien) Workshop 2006, Dresden, Germany.
- Hsu, C. L., & Lu, H. P. (2007). Consumer behavior in online game communities: A motivational factor perspective. *Computers in Human Behavior*, 23(3), 1642–1659. doi:10.1016/j.chb.2005.09.001
- Huang, E., Hsu, M.-H., & Yen, Y.-R. (2008). Understanding participant loyalty intentions in virtual communities. *WSEAS Trans. Info. Sci. and App.*, 5(4), 497–511.

- Iriberry, A., & Leroy, G. (2009). A life-cycle perspective on online community success. *ACM Comput. Surv.*, 41(2), 1-29. doi:<http://doi.acm.org/10.1145/1459352.1459356>
- Jacobs, J. (1965). *The death and life of great American cities*. London, UK: Penguin Books.
- Jagannathan, S., Srinivasan, J., & Kalman, J. (2002). *Internet Commerce Metrics and Models in the New Era of Accountability*. New York, NY: Prentice Hall PTR.
- Jin, X. L., Cheung, C. M. K., Lee, M. K. O., & Chen, H. P. (2007). *Factors Affecting Users' Intention to Continue Using Virtual Community*. Academic Press.
- Jones, Q., & Rafaeli, S. (2000). Time to Split, Virtually: Discourse Architecture and Community Building Create Vibrant Virtual Publics. *Electronic Markets*, 10(4), 214-223. doi:10.1080/101967800750050326
- Keeble, L., & Loader, B. D. (2001). Community informatics: themes and issues. In L. Keeble & B. D. Loader (Eds.), *Community Informatics: Shaping Computer-Mediated Social Relations* (pp. 1-10). New York: Routledge.
- Khazanchi, D., & Munkvold, B. E. (2000). Is information systems a science? An inquiry into the nature of the information systems discipline. *ACM SIGMIS Database*, 31(2), 24-42. doi:10.1145/381137.381141
- Kleinman, S. S. (2000). Social identification in a computer-mediated group for women in science and engineering. *Science Communication*, 21(4), 344-366. doi:10.1177/1075547000021004002
- Koh, J., & Kim, Y. G. (2003). Sense of virtual community: A conceptual framework and empirical validation. *International Journal of Electronic Commerce*, 8(2), 75-94.
- Koh, J., Kim, Y. G., Butler, B., & Bock, G. W. (2007). *Encouraging participation in virtual communities*. Academic Press.
- Kollock, P. (1998). Design Principles for Online Communities. *PC Update*, 15, 58-60.
- Kollock, P. (1999). The economies of online cooperation. In P. Kollock & M. Smith (Eds.), *Communities in cyberspace* (pp. 220-239). London: Routledge.
- Kozinets, R. V. (1999). E-tribalized marketing?: The strategic implications of virtual communities of consumption. *European Management Journal*, 17(3), 252-264. doi:10.1016/S0263-2373(99)00004-3
- Kozinets, R. V. (2002). The field behind the screen: Using netnography for marketing research in online communities. *JMR, Journal of Marketing Research*, 39(1), 61-72. doi:10.1509/jmkr.39.1.61.18935
- Laine, M. O. J. (2009). *Virtual Communities: A Bibliometric Analysis*. Paper presented at the 42nd Hawaii International Conference on System Sciences (HICSS'2009), Waikoloa, HI.
- Lapachet, J. H. (2001). *Virtual communities: The 90's mind altering drug or facilitator of human interaction?* Retrieved from http://is.gseis.ucla.edu/impact/s94/students/jaye/jaye_asis.html
- Lazar, J., Tsao, R., & Preece, J. (1999). One foot in cyberspace and the other on the ground: A case study of analysis and design issues in a hybrid physical and virtual community. *WebNet Journal: Internet Technologies, Applications, and Issues*, 1(3), 49-57.
- Lee, A. S. (2000). *Systems thinking, design science, and paradigms: heeding three lessons from the past to resolve three dilemmas in the present to direct a trajectory for future research in the information systems field*. Keynote Address. Paper presented at the 11th International Conference on Information Management, Kaohsiung, Taiwan.
- Lee, F. S. L., Vogel, D., & Limayem, M. (2003). Virtual Community Informatics: A Review and Research Agenda. *Journal of Information Technology Theory and Application*, 5(1), 47-61.

- Lee, J., & Suh, A. (2015). How do virtual community members develop psychological ownership and what are the effects of psychological ownership in virtual communities? *Computers in Human Behavior*, 45, 382–391. doi:10.1016/j.chb.2014.12.002
- Lee, M., & Lan, Y. (2007). From Web 2.0 to conversational KM: Toward collaborative intelligence. *Journal of Entrepreneurship Research*, 2(2), 47–62.
- Lin, H. F. (2006). Understanding behavioral intention to participate in virtual communities. *Cyberpsychology & Behavior*, 9(5), 540–547. doi:10.1089/cpb.2006.9.540 PMID:17034320
- Lin, H. F. (2008). Determinants of successful virtual communities: Contributions from system characteristics and social factors. *Information & Management*, 45(8), 522–527. doi:10.1016/j.im.2008.08.002
- Lin, W. K., Chiu, C. K., & Tsai, Y. H. (2008). Modeling Relationship Quality and Consumer Loyalty in Virtual Communities. *Cyberpsychology & Behavior*, 11(5), 561–564. doi:10.1089/cpb.2007.0173 PMID:18785818
- Ling, K., Beenen, G., Ludford, P., Wang, X., Chang, K., Li, X., & Rashid, A. M. et al. (2005). Using social psychology to motivate contributions to online communities. *Journal of Computer-Mediated Communication*, 10(4), 00. doi:10.1111/j.1083-6101.2005.tb00273.x
- Lu, Y., Zhao, L., & Wang, B. (2008). *Exploring Factors Affecting Trust and Purchase Behavior in Virtual Communities*. Academic Press.
- Mačiulienė, M., & Skaržauskienė, A. (2016). Emergence of collective intelligence in online communities. *Journal of Business Research*, 69(5), 1718–1724. doi:10.1016/j.jbusres.2015.10.044
- Marshall, S., Taylor, W., & Yu, X. (2004). *Using Community Informatics to Transform Regions*. Hershey, PA: Idea Group Publishing. doi:10.4018/978-1-59140-132-2
- McArthur, R., & Bruza, P. (2001). The ABC's of Online Community. In N. Zhong (Ed.), *Web Intelligence: Research and Development* (pp. 141–147). Berlin, Germany: Springer-Verlag. doi:10.1007/3-540-45490-X_16
- McIver, J. W. (2003). A Community Informatics for the Information Society. In B. Girard & S. Ó. Siochrú (Eds.), *Communicating in the Information Society* (pp. 33–64). Geneva: United Nations Research Institute for Social Development.
- McMillian, D. W., & Chavis, D. M. (1986). Sense of Community: A Definition and Theory. *Journal of Community Psychology*, 14(1), 6–23. doi:10.1002/1520-6629(198601)14:1<6::AID-JCOP2290140103>3.0.CO;2-I
- McWilliam, G. (2000). Building stronger brands through online communities. *Sloan Management Review*, 41(3), 43–54.
- Moore, T. D., & Serva, M. A. (2007). *Understanding member motivation for contributing to different types of virtual communities: a proposed framework*. Paper presented at the 2007 ACM SIGMIS CPR conference on Computer personnel research: The global information technology workforce, St. Louis, MO. doi:10.1145/1235000.1235035
- Morris, M., & Ogan, C. (1996). The internet as mass medium. *Journal of Computer-Mediated Communication*, 1(4).
- Nilan, M. S., Zakari, N., Guzman, I. R., & Zakaria, N. (2004). *Virtual Communities on the Web: Policy and Site Design Based on Facilitating and Hindering Users' Cognitive Movement*. Paper presented at the International Association for Development of the Information Society conference (IADIS), Lisbon, Portugal.
- Obst, P. L., & White, K. M. (2004). Revisiting the Sense of Community Index: A Confirmatory Factor Analysis. *Journal of Community Psychology*, 32(6), 691–705. doi:10.1002/jcop.20027

- Olson, J., & Olson, G. (1998). Research on Computer Supported Cooperative Work. In M. Helander, T. K. Landauer, & P. V. Prabhu (Eds.), *Handbook of Human-Computer Interaction* (pp. 1433–1456). New York, NY: Elsevier Science Inc.
- Palloff, R., & Pratt, K. (1999). *Building learning communities in cyberspace: effective strategies for the online classroom*. San Francisco: Jossey-Bass.
- Pentina, I., Prybutok, V. R., & Zhang, X. (2008). The role of virtual communities as shopping reference groups. *Journal of Electronic Commerce Research*, 9(2).
- Petouhoff, N. L. (2009). *The ROI of Social Media: Customer Service Communities*. Retrieved from <http://www.slideshare.net/drnatalie/dr-natalie-petouhoff-roi-of-social-media-social-media-club-presentation-forrester-research>
- Petter, S., DeLone, W., & McLean, E. (2008). Measuring information systems success: Models, dimensions, measures, and interrelationships. *European Journal of Information Systems*, 17(3), 236–263. doi:10.1057/ejis.2008.15
- Pitkin, B. (2001). *Community Informatics: Hope or Hype?* Paper presented at the 34th Hawaii International Conference on System Sciences (HICSS-34), Maui, HI.
- Plant, R. (2003). Online Communities. *Technology in Society*, 26(1), 51–65. doi:10.1016/j.techsoc.2003.10.005
- Porter, C. E., & Donthu, N. (2008). Cultivating Trust and Harvesting Value in Virtual Communities. *Management Science*, 54(1), 113–128. doi:10.1287/mnsc.1070.0765
- Porter, M. E. (1980). *Competitive advantage: Creating and sustaining superior performance*. Free Press.
- Preece, J. (2001a). *Online Communities: Designing Usability and Supporting Sociability*. New York, NY: John Wiley & Sons Inc.
- Preece, J. (2001b). Sociability and usability in online communities: Determining and measuring success. *Behaviour & Information Technology*, 20(5), 347–356. doi:10.1080/01449290110084683
- Preece, J. (2001c). Sociability and Usability: Twenty Years of Chatting Online. *Behavior and Information Technology Journal*, 20(5), 347–356. doi:10.1080/01449290110084683
- Preece, J., Abras, C., & Maloney-Krichmar, D. (2004). Designing and evaluating online communities: Research speaks to emerging practice. *International Journal of Web Based Communities*, 1(1), 2–18. doi:10.1504/IJWBC.2004.004795
- Preece, J., & Maloney-Krichmar, D. (2003). Online Communities: Focusing on Sociability and Usability. In J. Jacko & A. Sears (Eds.), *Handbook of Human-Computer Interaction* (pp. 596–620). Mahwah, NJ: Lawrence Erlbaum Associates Inc.
- Putnam, R. D. (1995). Tuning in, tuning out: the strange disappearance of social capital in America. *PS: Political Science and Politics*, 664–683.
- Ramaswamy, V. (2009). Leading the transformation to co-creation of value. *Strategy and Leadership*, 37.
- Rheingold, H. (1993). *The Virtual Community: Homesteading on the Electronic Frontier*. Cambridge, MA: Addison-Wesley.
- Rheingold, H. (2000). *The Virtual Community: Homesteading on the Electronic Frontier*. Cambridge, MA: MIT Press.
- Ridings, C. M., & Gefen, D. (2004). Virtual Community Attraction: Why People Hang Out Online. *Journal of Computer-Mediated Communication*, 10(1). Retrieved from http://www.ascusc.org/jcmc/vol10/issue1/ridings_gefen.html
- Ridings, C. M., Gefen, D., & Arinze, B. (2002). Some antecedents and effects of trust in virtual communities. *The Journal of Strategic Information Systems*, 11(3-4), 271–295. doi:10.1016/S0963-8687(02)00021-5

- Romm, C., Pliskin, N., & Clarke, R. (1997). Virtual communities and society: Toward an integrative three phase model. *International Journal of Information Management*, 17(4), 261–270. doi:10.1016/S0268-4012(97)00004-2
- Romm, C. T., & Taylor, W. (2001). *The Role of Local Government in Community Informatics Success Prospects: The Autonomy/Harmony Model*. Paper presented at the 34th Hawaii International Conference on System Sciences (HICSS-34), Maui, HI.
- Ruhi, U. (2008). *A Social Informatics Framework for Sustaining Virtual Communities of Practice* (J. Kisielnicki, Ed.). Hershey, PA: IGI Global. doi:10.4018/978-1-59904-955-7.ch037
- Schoberth, T., Preece, J., & Heinzl, A. (2003). *Online Communities: A Longitudinal Analysis of Communication Activities*. Paper presented at the 36th Hawaii International Conference on System Sciences (HICSS-36), Big Island, HI.
- Spaulding, T. J. (2009). How can virtual communities create value for business? *Electronic Commerce Research and Applications*.
- Sproull, L., & Kiesler, S. (1991). *Connections: New Ways of Working in the Networked Environment*. Cambridge, MA: MIT Press.
- Taylor, W. (2004). Community Informatics in Perspective. In S. Marshall, W. Taylor, & X. Yu (Eds.), *Using Community Informatics to Transform Regions* (pp. 1–7). Hershey, PA: Idea Group Publishing. doi:10.4018/978-1-59140-132-2.ch001
- Turner, J. W., Grube, J. A., & Meyers, J. (2001). Developing an optimal match within online communities: An exploration of CMC support communities and traditional support. *Journal of Communication*, 51(2), 231–251. doi:10.1111/j.1460-2466.2001.tb02879.x
- Valtersson, M. (1996). *Virtual Communities* (Masters Thesis). Umeå University, Umeå, Sweden.
- Wasko, M., & Faraj, S. (2005). Why should I share? Examining social capital and knowledge contribution in electronic networks of practice. *Management Information Systems Quarterly*, 29(1), 4.
- Wasko, M. M., & Faraj, S. (2000). It is What One Does: Why People Participate and Help Others in Electronic Communities of Practice. *The Journal of Strategic Information Systems*, 9(2-3), 155–173. doi:10.1016/S0963-8687(00)00045-7
- Wellman, B., & Gulia, M. (1999). Net surfers don't ride alone. In B. Wellman (Ed.), *Networks in the Global Village*. Boulder, CO: Westview Press.
- Wellman, B., Salaff, J., Dimitrova, D., Garton, L., Gulia, M., & Haythornthwaite, C. (1996). Computer networks as social networks: Collaborative work, telework, and virtual community. *Annual Review of Sociology*, 22(1), 213–238. doi:10.1146/annurev.soc.22.1.213
- Wise, K., Hamman, B., & Thorson, K. (2006). Moderation, response rate, and message interactivity: Features of online communities and their effects on intent to participate. *Journal of Computer-Mediated Communication*, 12(1), 24–41. doi:10.1111/j.1083-6101.2006.00313.x
- Yoo, W.-S., Suh, K.-S., & Lee, M.-B. (2002). Exploring the Factors Enhancing Member Participation in Virtual Communities. *Journal of Global Information Management*, 10(3), 55–71. doi:10.4018/jgim.2002070104

KEY TERMS AND DEFINITIONS

Computer-Mediated Communication: Any form of data exchange (text, images, audio, or video) among two or more end-users utilizing different ICTs for synchronous, asynchronous or real time communication with one another.

HCI (Human-Computer Interaction): An interdisciplinary field of research and practice

concerned with the design, implementation, use, and evaluation of interactive ICTs.

ICTs (Information and Communication Technologies): An umbrella term that includes any communication devices, computing hardware, content tools, as well as software applications.

Interdisciplinary Research: Integration of two or more seemingly exclusive academic disciplines or research fields in order to enhance understanding of a research phenomenon or to create new knowledge.

IT Artifact: Bundles of hardware infrastructure, software applications, informational content, and supporting resources that serve specific goals and needs in personal or organizational contexts.

Socio-Technical Perspective: A research outlook that highlights the interdependence and inextricable linkages between people (sociological systems) and ICTs (technological systems), underscores the co-evolution of these systems, and emphasizes that both systems need to be jointly optimized in order to produce positive practical outcomes.

Virtual Communities: (General definition) an aggregation of individuals or organizations with a shared purpose, who interact through a variety of technology platforms including social networks, online discussion forums, blogs, wikis, email lists or newsgroups.

Virtual Communities: (Proposed definition) an aggregation of participants (individuals or organizations) with a shared purpose, who have their interactions mediated through a technology platform and guided by collective protocols, and who embrace a cooperative membership persona in their online interactions.

H

ENDNOTES

- ¹ The highlighted references in Table 2 are based on top ten citations on virtual communities from the ISI Web of Science Citation Index as reported in Laine (2009).
- ² By no means are these the only socio-technical literature domains that discuss virtual communities. Nevertheless, they provide a modest view of cross-functional perspectives related to sociological/psychological and technological attributes of virtual communities.
- ³ The constituents of the proposed definition have deliberately been named using 5Ps. Acronyms might be passé and this particular one may even be cliché, but it provides a short way of describing the essence of the proposed definition.

The Trajectivity of Virtual Worlds

Christophe Duret

Université de Sherbrooke, Canada

INTRODUCTION¹

According to Stéphane Vial (2016), “there is a profane metaphysics that operates at the heart of the contemporary imagination, which postulates that the human world is bisected by an invisible border that separates the so-called real and the so-called virtual” (p. 135).

However no meaning of “virtual” validates this dichotomy, so it would be more appropriate to speak of “digital monism” (Vial, 2014; Stimler & Vial, 2014) to describe the phenomenon of virtual environments, a formulation which “states that the human reality is a digital-centered hybrid environment made of mixed systems and matters constantly interlinked, that tends to form a single continuous multimaterial artifactual substance.” (Stimler & Vial, 2014, para. 4). This profane metaphysics underpins the opposition between cyberspace and ‘meatspace’, borrowed from cyberpunk literature, or between offline and online. This has been widely disseminated by various web commentators, and also by researchers.

This chapter will propose an ontology of virtual worlds that calls into question the dichotomy between the real and the virtual. This will draw on the concepts of trajectivity and *médiance* (Berque, 2000) in order to describe the way virtual worlds, with their technological and symbolic features, take part in the construction of milieus (or “human environments”) (first part: *Background*). This theoretical proposition will be illustrated with the analysis of *Arcadia*, a virtual world built in *Second Life* (second part: *Second Life’s Arcadia as a Virtual World*). Finally, a mesocriticism will be proposed as a new approach for the study of virtual worlds (third part: *Future research directions*).

BACKGROUND

Researches into virtual worlds in the 1990s and 2000s are characterized by their tendency to consider them separate areas from the real world, betraying a perspective Nathan Jurgenson (2012) calls “digital dualism”. As Doel & Clarke (1999) and Latzko-Toth & Proulx (2006) point out, typical of this dualism are the epistemological postures making virtuality a degraded representation of reality or a solution to correct its flaws, the first reflecting a vision of the virtual as a simulacrum (Baudrillard, 1981), the second being put forward by techno-optimist authors such as Howard Rheingold (2000).

“Virtual” is a term whose meaning varies over history and from one researcher to another. In its long history, it has in turn been synonymous with the potential, the artificial and the simulational (Vial, 2014).

The virtual is known primarily as that which exists potentially, as opposed to what exists actually. Thus, in the eighteenth century, the science of mechanics called “virtual realities of the physics” the matrix of possibilities in which the empirical world represents an actualization (Latzko-Toth & Proulx, 2006). Following on from this, the virtual “is nothing other than an ontological regime, a particular way of being real, that which, in short, *exists without manifesting itself*” (Vial, 2014, p. 179, original in italics). Thus, the virtual is real without being actual (Deleuze, 1996; Granger, 1995; Lévy, 1998).

In the field of optics, the virtual is given a second meaning. For the physicists, the virtual image is an image perceived by the eye that comes from an optical instrument, as opposed to the actual

image, which is present on a screen. According to Vial (2014), the term “real” here as opposed to “virtual” is inappropriate, since in both cases they are discernible realities; the virtual image is real, but artificial.

In the field of IT, finally, the virtual becomes ‘simulational’. That is, as in optics, an artificial form, and more specifically, a “process capable, via programming techniques, of simulating a digital behavior *independently* of the physical medium on which (paradoxically) it *depends*” (Vial, 2014, p. 181, original in italics). While the simulacrum is a matter of lies and illusion, points out Vial (2014), simulations are artificially constructed but ontologically real. The virtual is thus equivalent to a computer model. It translates the concept as an idealized version of the modelled object (Latzko-Toth & Proulx, 2006), since it retains from the latter only a few characteristics and behaviors deemed relevant. This implies, as outlined by Gonzalo Frasca (2003) with his concept of simulation rhetoric, that the model conveys the point of view of the author.

The virtual understood as simulational is not unrelated to the term defined by Charles Sanders Peirce (1974). Indeed, according to him “a virtual X (where X is a common noun) is something, not an X, which has the efficiency (virtus) of an X.” (p. 261) In other words, a virtual object (a simulation) produces the same effect as the simulated object without merging with it. This is the case of an Internet forum, for instance, which, like the agora, is a facilitator of debates and discussions. As such, the virtual cannot be equated to ‘potential’ as it is, according to Peirce, “almost its contrary. For the potential X is the nature of X, but is without actual efficiency” (p. 261)

It is from the late 1980s, mentions Vial (2014), that some digital thinkers merge the meanings of the term “virtual” as both simulational and potential. During this era there appeared a profane metaphysics dichotomizing the real and the virtual, that he describes as a form of phenomenological judgment in response to the shock caused by the arrival of digital interfaces in the perceptual hab-

its of individuals (Vial, 2016). This judgment at first led to the idea of a real world separate from a virtual world. But since then, incorporated into daily practices, the new realities that accompanied the digital revolution, (videogames, social media, etc.) have become commonplace. Thus, “by being integrated into our world experience, they create a new phenomenological viewing point by which virtual beings of the digital system can become world phenomena” (Vial, 2014, p. 186), making it now appropriate to speak of “digital monadism” to describe them (Stimler & Vial, 2014). Neal Stimler and Stéphane Vial subscribe, in this sense, to a third epistemological posture, where virtuality is considered a constitutive phenomenon of reality, as “the principle of ‘hybridization’ of the various levels of reality” (Latzko-Toth & Proulx, 2006, p. 58). In the same vein, Don Slater (2002) questioned the offline / online dichotomy and found that “virtuality is clearly not a feature of the media but one social practice of media use amongst many others.” (p. 544) He thus decompartmentalizes the so-called real and the so-called virtual.

Jurgenson (2012) agrees with Slater when he argues that mobile phones and social media – in the context of the Arab Spring, the Occupy movement, and flash mobs – give rise to “massive gatherings of digitally-connected individuals in physical space [where] digital and physical enmesh to form an ‘augmented reality’.” (p. 83). He therefore echoes the position of Beth Coleman (2011), who describes the emergence of an age of mobile, pervasive, networked connectivity. This network society encompasses a multiplicity of network combinations and might be understood as a continuum that crosses the virtual and the real (the online and the offline). Coleman calls this augmented reality “X-reality” and she argues for a new “understanding of networked media as augmentation of self and world” (p. 83).

Maude Bonenfant (2011) and Vili Lehdonvirta (2010) adopt the same epistemological posture as Stimler and Vial and attribute the false real/virtual dichotomy in MMOGs (massively multiplayer online games) to the concept of magic circle

(Huizinga, 1951), this boundary within which the game is circumscribed, considered an activity isolated from everyday life. This concept results in an exceptionalist perspective of videogames considered objects with sealed borders (see Bartle, 2006; Castronova, 2004; Salen & Zimmerman, 2003) that opposes a departicularising perspective of the games considered as an activity in an osmotic relationship with its environment (society, culture, media landscape, etc.) (see Bonenfant, 2011; Boutet, Carvajal, Ter Minassian, & Triclot, 2013; Lehdonvirta, 2010).

Our own research on *Second Life*'s Gorean role-playing games (Duret, 2014, 2015, 2016a,b) belongs to the departicularizing perspective and supports the principle of hybridization. The research shows that players carry out constant displacements between the different levels of the game (virtual world per se, diegesis, formal structure, etc.), applying a multitude of fictional, ludic, social, technical, and symbolic frameworks in order to make sense of their gaming experience. These displacements and the lability of the frameworks (Goffman, 1974) that assure their organization as a gaming experience are explained by the cognitivist concept of attentional resource allocation. This concept affords the users in virtual worlds an understanding of engagement as a phenomenon located both inside and outside the boundaries of the simulation.

The Trajectivity of Virtual Worlds

In this chapter, the term 'virtual' should be read in the sense of 'simulation' (as both a model and efficiency). To prevent the temptation of digital dualism, we will now ontologically base the reality of virtual worlds by regarding them as technical, symbolic, and technosymbolic extensions of the human milieus. Following the principle of hybridization and of digital monism, as seen above, this signifies the consideration of virtual worlds not as entities belonging to a separate sphere or as digital supports for social activities, but as one component of the milieu, which unites within one web the

digital and non-digital relationships, the symbolic and the technical, the subject and the object, the simulation and the simulated phenomenon. To do this, it will be necessary to call on the concepts of *médiance* and trajectivity.

The geographer Augustin Berque (2000) defines the concept of *médiance* as "the relationship of a society to its environment" (p. 128). It refers to the asymmetric ontological structure of human beings that consists of "the bipartition of our being into two unequal 'halves', one invested in the environment through the technical and symbolic, the other made up of our animal body" (p. 128). These halves, "one internal, one external, one physiologically individualized (... our animal body), the other diffused in the milieu (... our '*médial*' body)" (p. 128) indicates a dynamic identity deployed through a process of externalization of functions, which "gradually rejects all the instruments outside of man" (André Leroi-Gourhan, as cited in Berque, 2000, p. 98). These functions formerly belonged to the animal body of the human being. During the course of evolution, the process took the form of a "triple and mutual engendering" (p. 96), composed of hominization (the passage from animals to humans), anthropization (the objectification of things through technique) and humanization (the subjective transformation of things with the aid of symbols). This triple engendering involves the symbolic, technical and ecological in the same bundle of relationships. It constitutes the milieu, and the concept of *médiance* reflects the meaning given to this milieu.

The milieu is neither subjective nor objective, it is trajective. According to Berque (2000), the concept of trajectivity refers to the fact that the reality of the milieu cannot be understood by modern dualism. This reality is neither subjective (from the perspective of a social actor) nor objective. It is the result of a relation called 'trajectivity'. The trajection is the operation through which the subject and the object grow together in order to produce the concrete reality. They are both embedded in a web of relationships that are technical,

symbolic and ecological. These relationships are deployed over time and on a social rather than an individual scale. The human being comes from such relationships, since it is caught in this frame which constitutes its existential web, its milieu.

According to Berque (2000), the human being externalizes ('cosmizes') his animal body in the environment through technique and internalizes (somatizes) the world in this same body through the symbolic. It is through the process of somatization that one represents the world. However, Berque's externalization/internalization dichotomy cannot take into account the numerous devices whose roles are to reduce the symbolic complexity (viewed in terms of data) of the contemporary milieus in order to organize their representation in an accessible manner. Thus, it is a matter of internalizing the symbolic through technique. Therefore, we will describe these tools as 'technosymbolic'. Technosymbolic tools (e.g. geographical information systems, web search engines, mathematical models, etc.) propose a scaling of the complex contemporary milieus within the limits of human cognitive resources in order to prevent any fragmentary somatization (or 'dys-somatization'). In this sense, they serve as auxiliaries in the cognitive mapping (Jameson, 1991; Lynch, 1960) of the milieus. Virtual worlds as computer models constitute technosymbolic tools. Every model retains only a few characteristics from the modelled object. Therefore, virtual worlds reduce the complexity of the modelled milieu and facilitate their internalization when they symbolically illustrate the data and the behaviors that characterize these milieus.

The argument put forward in this section is that virtual worlds, which are simultaneously symbols (3D simulations), techniques (e.g. communication devices, sets of rules), technosymbolic tools, and navigable spaces, occupy a significant place in this bundle of relationships that make up the *milieu*, in the same way as physical places. At the scale of the *milieu*, virtual worlds are as real as are social interactions, rocks or individuals, which suggest a digital monism. As milieus, virtual

worlds proceed from a technical and symbolic deployment of space as well as exceed their *topos* (their objective and measurable physical space). Therefore, virtual worlds belong to the *médial* body. By doing so, they participate along with the ontological structure of the human being. As simulations, virtual worlds possess, for instance, the efficiency of tangibles places of sociability, yet within the mode of telepresence and in a symbolic space. Finally, virtual worlds constitute technosymbolic tools: in addition to technically extending the human milieus and bringing them within the boundaries of the animal body in the form of symbols (somatization), they serve as auxiliaries in the latter process when they carry out a cognitive scaling.

SECOND LIFE'S ARCADIA AS A VIRTUAL WORLD

Basing virtual worlds ontologically on the concepts of trajectory and *médiance* is a first step. The next step will be to show how this foundation is relevant to the analysis of a specific virtual world. This is proposed through a brief analysis of *Second Life's Arcadia*².

Arcadia is a mountainous region of Greece, in the central and eastern part of the Peloponnese peninsula. This mythical landscape banishes urban life and is populated by shepherds. It proposes a "refusal of the social organization and its constraints and it develops around the themes of the simplicity, the nature, and the pastoral way of life" (Letonturier, 2013, pp. 223-224). Arcadia is found in poetry, opera and painting, depicted by poets and artists as famous as Virgil, Ovid, and Nicolas Poussin.-According to Lyman Tower Sargent (1994), it is reminiscent of utopia and according to Françoise Duvignaud (1994), this utopia is characterized by a sense of nostalgia. It feeds the imagination of exiles haunted by nostalgia for an idealized place whose loss is painful.

Now what about the virtual Arcadia? Arcadia is a "sim" built in *Second Life*, a multi-user virtual

environment inhabited by people via their avatar. “Sims” or “simulators” are three-dimensional virtual worlds. They are leased to users who design and administer them.

Arcadia as a virtual world forms, at the same time, a technical extension of human milieus, a symbolic representation of these milieus, and a technosymbolic tool that facilitates their somatization with the support of cognitive scaling.

Arcadia as an Extension of Human Milieus

The externalization of the technique on the environment takes into account the projection of the animal body in the *médial* body, this social part of beings. *Arcadia* encourages this social dimension as a communication and sociability device.

The housing design and the furnishings foster sociability and conviviality. Indeed, there is a profusion of couches, garden furniture, bar counters, tatamis and “zabutons” (Japanese cushions). In addition, the sim proposes places of sociability like terraces, rotundas, a café, a library, a living room, and a sandy beach. The main architectural feature of *Arcadia* is its rotundity (as can be seen in its rooms, terraces, rotundas, bar counters, etc.), which encourages gatherings, requiring that everyone be seen. Communicational devices (text and voice chat, private messaging, scripted nonverbal language of the avatars) also promote sociability. Finally, access to the sim is open to the public, so everyone can access it freely. As with all physical places that engender sociability (restaurants, bars, cafés, etc.), *Arcadia* provides the same efficiency that foster interactions and communication between users.

Arcadia as a Modelization of Human Milieus

Like the mythical Arcadia, *Arcadia* represents a nostalgic utopia. As such, it models an idealized epoch and place that represents a critique of the social conjuncture in which the modeller lives.

The mythical Arcadia consists of a community governed by simple rules and a pastoral way of life. It belongs to a pre-industrial past where human beings, their habitat, and nature live harmoniously together. The way *Arcadia* was modelled testifies to the modeller’s view on the mythical place and, implicitly, of his own epoch, symmetrically deemed as socially complex, overurbanized, overindustrialized, and alienated from nature. In addition, *Arcadia* participates in the internalization of the world: the lost ideal world, but also the contemporary one, which is deemed flawed in comparison with the former.

Arcadia is a simulation of the mythical eponymous landscape and not a reconstruction of the geographical area. It nevertheless retains from the latter a flat landscape surrounded by steep mountains. *Arcadia* is in fact an archipelago surrounded by a mountain range. The flora and fauna of *Arcadia* are heterogeneous. The first combines specimens from temperate zones (e.g., cypresses, pines) and tropical (e.g., palm trees) with border ponds, waterfalls, cliffs and white sandy beaches. The second combines domestic animals (cows, horses, sheep, llamas, fishes in a garden pond, etc.) as well as wildlife (manatees, dromedaries, flamingos, deers, weasels, etc.): the fertile Arcadia of the shepherds meets the pristine wilderness of the unspoiled islands. This heterogeneity suggests an idealization of nature: it is not an ecosystem that is simulated, but nature itself, thereby returning to an Arcadia viewed as a “fantastical mental landscape” (Duvignaud, 1994, p. 52).

Human dwellings (terraces, winter gardens, café) are not absent from the sim, but they intermingle intimately and harmoniously with nature, effectively forming open architectures endowed with green roofs and within which grows vegetation.

All signs related to urbanization and industrialization are absent from *Arcadia*. On the one hand, it offers landscapes and a habitat on a human scale, as opposed to cityscapes and metropolises which are problematic to map cognitively (Lynch, 1960). On the other hand, there is a total absence

of production. *Arcadia* offers itself as a place of leisure, sociability, and relaxation. The presence of domestic animals coupled with the absence of work suggests the idea of *Arcadia* as a nourishing earth, “the idea of a ‘nature’ that would see humans released from work” (Duvignaud, 1994, p. 52).

So far, the interlacing of relationships of the milieu (its trajectory) is composed of technical (*Second Life* software allowing the design and management of the environment, servers, computers, communication devices, etc.) and symbolic (*Arcadia* as nostalgic utopia) constituents. *Arcadia* therefore combines trajectively both the subjective – symbolic representations of the world and social actors – and the objective – material support consisting of telecommunications networks, servers and computers, as well as the reference objects used in the design of *Arcadia* (beaches, furnishings, mountain ranges, flora and fauna, etc.). The sim also involves a community (the designer and the owner of *Arcadia*, *Second Life* employees, authors who have developed works by taking up and augmenting the *Arcadia* theme, users who explore the sim, socialising and forming a community within it, etc.). In addition, *Arcadia* involves several time frames: the plury-millennial arcadian conventions in art, the multi-centennial traditions of the utopian genre, and the preindustrial and contemporary epochs involved (positively or negatively) by the arcadian utopia. Let’s now see the technosymbolic dimension of *Arcadia*.

Arcadia as a Technosymbolic Device

Along with efficiency, technosymbolicity constitutes another aspect of virtual worlds taken as simulations. *Arcadia* fulfills the role of auxiliary in the somatization of the world when it models it on a small scale. In terms borrowed from Kevin Lynch (1960) and Fredric Jameson (1991), the sim helps its users to cognitively map the world in which they live (the city for the former, the global socioeconomic system of the late-capitalism for the latter). Without mapping ability, this world would appear alienating because it would be too complex to be understood globally.

As a simulation, *Arcadia* takes place on a small scale. In *Second Life*, a sim has a size of 65 536 sqm and only 100 avatars are admitted there simultaneously. *Arcadia* is thus an easy place to cognitively map the technical, symbolic and social levels. Arcadian utopia and the contemporary world are both summoned in this restricted place. The contemporary world is summoned on one hand as non-reference (it exists negatively as a set of flaws to avoid) and on the other through the form of vending machines that inscribe the contemporary intangible economy at the heart of the arcadian utopia. Indeed, vending machines are scattered across the archipelago of *Second Life* in the form of a network that notably includes *Arcadia*. This network can be seen as an atopia (a non-place). Non-places are defined by Marc Augé (1992) as non-identity, non-relational and non-historical places. They are representative of a system that promotes the free movement of men, goods and information flows. Airports, highways, malls and communication networks are representative of non-places. Vending machines allow the sale of artifacts in the sims. They require payment in *Linden* dollars, the currency of *Second Life* convertible into US currency. *Arcadia* offers the sale of miscellaneous objects (folding screens, tatamis, cushions, mobile storage cabinetries, bikinis, DJ mixers, speakers, turntables, etc.) through a set of vending machines tied into the wider networks of *Second Life* and of the intangible economy. Like the communication devices above-mentioned, these virtual vending machines have efficiency (the efficiency of physical stores. Because of these vending machines, the sim is both closed (utopian) and open (as an atopic network connected to the intangible economy) and it transposes a global-scale phenomenon into a local scale.

Additionally, as a technosymbolic tool, *Arcadia* summarizes the problematic state of contemporary *médiance*, translated by the need for nature, a desire felt by human beings to return to an environment of natural living (Bailly & Bourdeau-Lepage, 2011). It is a symptom of what Berque (2000) describes as the decosmization of modernity whereby nature

and things are transformed in systems of objects, and through which the individual is artificially separated from his milieu.

As we have seen previously, *Arcadia* simulates a world in which nature and human habitats intertwine. This utopian relationship reflects a “structural need for unspoiled nature” (Berque 2000, p. 155). The distinction between natural and artificial tends to disappear in this enclave where an attempt to reinscribe the individual in his milieu (a recosmization) is at stake. The vending machines decompartmentalize the arcadian enclave in reconnecting it to the atopic networks of the intangible economy, a presence of a hypertrophied technique at the heart of the *médial* body.

FUTURE RESEARCH DIRECTIONS

Replacing virtual worlds in their milieu overcomes the digital dualism and draws attention to their eco-symbolic-technical dimensions, which are collective and extend through a plurality of time frames. Such an approach could be qualified as a hermeneutics of virtual worlds, a *mesocriticism* (the “meso” prefix here representing “milieu”) capable of uncovering the web of relationships without which a virtual object is nothing more than a meaningless assemblage of pixels, rules, and hardware. The analysis of *Arcadia* is only a draft of this mesocriticism and calls for the development of new methodological tools. Furthermore, when these virtual worlds are games, they come with an extra dimension: a diegesis (the fictional spatiotemporal universe of the game), which the mesocriticism must take into account in future researches.

CONCLUSION

Thinking of the virtual in the context of a bipartition between animal body and *médial* body permits its inclusion in the same ontological definition of the human being, since it becomes one technical,

symbolic, and technosymbolic tool among others whereby the latter invests himself in his milieu and makes sense of it. This conception avoids the false real/virtual dichotomy and hybrids the various levels of the milieu in a trajective reality. In addition, viewing the virtual as a simulation allows us to link together its characteristics that are efficiency and modelling. According to these qualities, virtual worlds constitute communication, selling, and cognitive scaling tools. Finally, the analysis of *Arcadia* as virtual world highlights its paradoxical nature: as a construction both utopian and nostalgic, it provides refuge for anyone who aspires to a return to nature, but without to avoid a technically and symbolically hypertrophied *médial* body: after all, *Arcadia* stays the avatar of an intangible economy. Its inclusion in the economic sphere of *Second Life* gives it the quality of utopias that Jameson (2005) calls “postmodern”, these “islands in the net, a constellation of discontinuous centers, themselves internally decentered” (p. 221). In this sense, *Arcadia* can be compared to contemporary concrete utopias such as Ecofaubourg® (2008), which offers *a new form of city that is more and more conscious of sustainable development and seeks to reconcile urban lifestyles and economic dynamism with environmental conservation* [which is] *structured around various kinds of network... [as its aims are] to be mobile, connected and service-oriented.* (para. 1)

This analysis depicts a technosymbolically constructs diagram of a contemporary milieu in which even the most isolated islands remain connected on a global scale.

REFERENCES

- Augé, M. (1992). *Non-lieux: Introduction à une Anthropologie de la Surmodernité*. Paris: Seuil.
- Bailly, A., & Bourdeau-Lepage, L. (2011). Concilier Désir de Nature et Préservation de l'Environnement: Vers une Urbanisation Durable en France. *Géographie, économie, société*, 13(1), 561-593.

- Bartle, R. (2006). Virtual Worldliness. In J. Balkin & B. S. Noveck (Eds.), *The State of Play: Law, Games and Virtual Worlds* (pp. 31–54). New York: New York University Press.
- Baudrillard, J. (1981). *Simulacres et Simulation*. Paris: Kawade Shobo Shinsha.
- Berque, A. (2000). *Écoumène: Introduction à l'étude des milieux humains*. Belin: Reclus.
- Bonenfant, M. (2011). Les Mondes Numériques ne Sont Pas “Virtuels”: L'Exemple des Jeux Vidéo en Ligne. *Revue des Sciences Sociales*, (45), 60–67.
- Boutet, M., Carvajal, I. C. (de), Ter Minassian, H., & Triclot, M. (2013). Au-delà du Virtuel: Interactions Sociales et Spatiales dans et Autour d'un Univers Vidéoludique. *Médiation et Information*, (37), 103–116.
- Castronova, E. (2004). The Right To Play. *New York Law School Law Review*, (49), 185–210.
- Coleman, B. (2011). *Hello Avatar: Rise of the Networked Generation*. Cambridge, MA: The MIT Press.
- Deleuze, G. (1996). L'actuel et le virtuel. In G. Deleuze, & C. Parnet (Eds.), *Dialogues* (pp. 177–186). Paris: Flammarion.
- Doel, M. A., & Clarke, D. B. (1999). Virtual Worlds: Simulation, Suppletion, S(ed)uction and Simulacra. In M. Crang, Ph. Crang, & J. May (Eds.), *Virtual Geographies: Bodies, Space and Relations*. London: Routledge.
- Duret, C. (2014). Les Cadres de l'Expérience Vidéoludique et la Distribution des Ressources Attentionnelles dans les Jeux de Rôle en Ligne. *Revista Contracampo*, (29), 84–108.
- Duret, C. (2015). Figures Disruptives et Cadres de l'Expérience Vidéoludique dans les JRPEV Goréens. *Sciences du jeu.org*, (3), 1–14.
- Duret, C. (2016a). Jeux de Rôle Néomédiévalistes et Rhétorique de la Démédiation: Les Procédés d'Ouverture et de Fermeture des Hétérotopies Goréennes dans. *Second Life, Communication & Organisation*, (48), 49–60.
- Duret, C. (2016b). Penser les Frontiers des Jeux de Rôle Participatifs en Environnement Virtuel: Le Cas des Jeux de Rôle Goréens de Second Life. In H. Hachour, N. Bouhaï, and I. Saleh (Eds.), *Frontières Numériques & Artéfacts* (pp. 61–106). Paris: L'Harmattan.
- Duvignaud, F. (1994). *Terre Mythique, Terre Fantasmée: L'Arcadie*. Paris: L'Harmattan.
- Ecofaubourg®. (2008). *Reconciling the individual with society and the environment*. Retrieved September 15, 2015, from <http://www.ecofaубourgs.com/en>
- Frasca, G. (2003). *Simulation versus Narrative: Introduction to Ludology*. Retrieved September 20, 2011, from http://www.ludology.org/articles/VGT_final.pdf
- Goffman, E. (1974). *Frame Analysis*. Cambridge, MA: Harvard University Press.
- Granger, G.-G. (1995). *Le Probable, le Possible et le Virtuel*. Paris: Odile Jacob.
- Huizinga, J. (1951). *Homo Ludens: Essai sur la Fonction Sociale du Jeu*. Paris: Éditions Gallimard.
- Jameson, F. (1991). *Postmodernism, or, The Cultural Logic of Late Capitalism*. Durham, NC: Duke University Press.
- Jameson, F. (2005). *Archaeologies of the Future: The Desire Called Utopia and Other Science Fictions*. London: Verso.
- Jurgenson, N. (2012). When Atoms Meet Bits: Social Media, the Mobile Web and Augmented Revolution. *Future Internet*, 4(4), 83–91. doi:10.3390/fi4010083

- Latzko-Toth, G., & Proulx, S. (2006). Le virtuel au pluriel: cartographie d'une notion ambiguë. In S. Proulx, L. Poissant, & M. Sénécal (Eds.), *Communautés Virtuelles: Penser et Agir en Réseau* (pp. 57-76). Québec: Presses de l'Université Laval.
- Lehdonvirta, V. (2010). Virtual Worlds Don't Exist: Questioning the Dichotomous Approach in MMO Studies. *Game Studies*, 10(1).
- Letonturier, É. (2013). Glossaire. In É. Letonturier (Ed.), *Les Utopies* (pp. 223-230). Paris: CNRS éditions.
- Lévy, P. (1998). *Qu'est-ce que le Virtuel?* Paris: La Découverte.
- Lynch, K. (1960). *The Image of the City*. Cambridge, MA: MIT Press.
- Peirce, C. S. (1974). *Collected Papers of Charles Sanders Peirce* (Vol. V-VI). Cambridge, MA: Harvard University Press.
- Rheingold, H. (2000). *The Virtual Community: Homesteading on the Electronic Frontier*. Cambridge, MA: The MIT Press.
- Salen, K., & Zimmerman, E. (2003). *Rules of Play: Game Design Fundamentals*. Cambridge, MA: The MIT Press.
- Sargent, L. T. (1994). The Three Faces of Utopianism Revisited. *Utopian Studies*, 5(1), 1-37.
- Slater, D. (2002). Social relationships and identity online and offline. In L. Lievrouw, & S. Livingstone (Eds.), *Handbook of New Media: Social Shaping and Consequences of ICTS* (pp. 533-546). London: Sage Publications. doi:10.4135/9781446206904.n38
- Stimler, N., & Vial, S. (2014, April). *Digital Monism: Our Mode of Being At The Nexus of Life, Digital Media and Art*. Paper presented at Theorizing the Web 2014, New York, NY.
- Vial, S. (2014). Contre le Virtuel: Une Déconstruction. *Médiation et Information*, (37), 177-188.
- Vial, S. (2016). La Fin des Frontières Entre Réel et Virtuel: Vers le Monisme Numérique. In H. Hachour, N. Bouhaï, & I. Saleh (Eds.), *Frontières Numériques & Artéfacts* (pp. 135-146). Paris: L'Harmattan.

ADDITIONAL READING

- Augé, M. (2010). Retour sur les "Non-Lieux": Les Transformations du Paysage Urbain. *Communications*, (87): 171-177.
- Berque, A. (1986). *Le Sauvage et l'Artifice: Les Japonais Devant la Nature*. Paris: Gallimard.
- Berque, A. (1990). *Médiance, de Milieux en Paysages*. Paris: Belin: Reclus.
- Besnier, J.-M. (1995). Introduction. In G. Cohen-Tannoudji (Ed.), *Virtualité et réalité dans les sciences*. Gif-sur-Yvette: Éditions Frontières. doi:10.1017/CBO9780511519864.002
- Calleja, G. (2011). *In-Game: From Immersion to Incorporation*. Cambridge, MA: The MIT Press.
- Consalvo, M. (2009). There is No Magic Circle. *Games and Culture*, 4(4), 408-417. doi:10.1177/1555412009343575
- Duret, C. (2014). Living the Phantasm of De-mediation: The Priest Kings and the Technology Prohibition in the Gorean Role-Playing Games, *Heidelberg Journal of Religions on the Internet*, (7), 41-60.
- Garfield, R. (2000). Metagames. Retrieved March 5, 2013, from <http://ludusmechanicus.files.wordpress.com/2009/11/garfield.pdf>
- Jones, S. E. (2008). *The Meaning of Video Games: Gaming and Textual Strategies*. New York, NY: Routledge.
- Siang Ang, C., Zaphiris, P., & Wilson, S. (2010). Computer Games and Sociocultural Play: An Activity Theoretical Perspective. *Games and Culture*, 4(5), 354-380. doi:10.1177/1555412009360411

KEY TERMS AND DEFINITIONS

Médiance: The meaning given to the milieu (see Milieu).

Milieu: (Or “human environment”) is the interlacing of the symbolic, technical and ecological in the same bundle of relationships. The milieu is a trajectory reality (see Trajection/trajectivity).

Non-Places: Non-identity, non-relational and non-historical places. Representative of a system that promotes the free movement of men, goods and information flows (e.g. airports, highways, malls and communication networks).

Sims: Three-dimensional virtual worlds built in Second Life.

Trajection/Trajectivity: The trajectory is the dynamics through which the subject and

the object grow together in order to produce the concrete reality. They are both embedded in a web of relationships that are technical, symbolic and ecological. The quality of this dynamics is called trajectivity.

H

ENDNOTES

- ¹ The research on which this article is based has received financial support from the Social Sciences and Humanities Research Council (SSHRC).
- ² Available at <http://maps.secondlife.com/secondlife/Arcadia/72/158/26>. Visited between September 18 and September 23, 2015.

Virtual Hoarding

Jo Ann Oravec

University of Wisconsin – Whitewater, USA

INTRODUCTION

This article outlines hoarding issues involving virtual goods (including databases, videos, images, and digital documents) in workplace, household, and personal contexts. It covers security issues, intellectual property concerns, and matters related to information flow in organizational settings. The article also includes reflections about the moral and personal dimensions of virtual hoarding, with an emphasis on information ethics and the opportunistic appropriation of organizational data (and in some cases, knowledge) for individual gain. Organizations as well as individuals could face substantial losses through the compulsion to “save” virtual goods without developing and carrying out appropriate strategies for managing them.

Virtual hoarding issues may not seem to be critical given the decreased costs of on-site and backup storage as well as relatively-inexpensive storage facilities in the “cloud.” However, data that are not managed in terms of their formats and their storage substrata could certainly present issues for organizations; also, data that are inappropriately removed from the flow of information in organizations may put vital processes at risk. On the level of the individual, personal and professional issues involving the storage of hundreds of thousands of images, videos, and documents with little context or metadata could present comparable problems, albeit on a more contained scale. According to researchers, the reasons that individuals become hoarders range from uncertainty avoidance and OCD (obsessive-compulsive disorder) to opportunistic motives for personal acquisition and perhaps subsequent advancement (Bratotiis, Schmalisch, & Steketee, 2011; Grant, 2014; Oravec, 2015). At the organizational level, the system-level aspects

that lead individuals in particular contexts to hoard data may have close parallels to these individual-level phenomena.

BACKGROUND

The study of virtual or digital hoarding is just emerging, and new case studies, survey results, and other forms of research are being undertaken and disseminated (Oravec, 2015; van Bennekom, Blom, Vulink, & Denys, 2015). Although this article focuses on the hoarding of virtual goods, some discussion of the hoarding of physical objects may provide background and insights. In a variety of social settings, hoarding behavior involving physical items is expanding in its impact; hoarding has been construed as a mental health issue in some organizations (Bratotiis, Schmalisch, & Steketee, 2011) and has been labeled as a disorder in the DSM-5 (*Diagnostic and Statistical Manual of Mental Disorders*, American Psychiatric Association, 2013). Capacities for managing the locations of and access to physical as well as virtual entities are often considered central to competent societal functioning. Lepselter (2011) relates dozens of negative characterizations of hoarding in newspapers, television, and social media. The hoarding of “virtual goods” is also generating concern and has the potential to be even more costly for organizations than its physical correlate. Files that are not properly identified and stored might be seen as appropriately “saved” if placed in the “cloud.” However, if metadata about the files are not available, the files may be essentially worthless, wasting precious organizational resources. Gormley and Gormley (2012) describe the condition of “information clutter” as running

DOI: 10.4018/978-1-5225-2255-3.ch373

parallel with hoarding behavior, a situation that is generally not conducive to conducting efficient workplace operations.

Certain kinds of organizations and professions may be more prone to virtual hoarding issues than others. Peyton (2015) describes the problems some law firms have with digital data management and relates that “Many firms are notorious data hoarders and seem to hold old records without any legitimate justification” (p. 18). Knowledge hoarding has been shown to become a concern in workplace contexts in which uncertainty and loss of trust are issues (Holten, Hancock, Persson, Hansen, & Hogh, 2016; Oravec, 2017). For many professionals, the transition is often difficult from an era in which information was relatively scarce (just a few decades ago) to one in which information resources are overwhelming in size, access, and complexity. Peyton relates that “sub-standard information governance and recordkeeping model” are often found in professional settings, and declares that “Legitimate business justifications for retaining electronic information do not include ‘I may need that information someday—you never know’” (p. 19). “Big data” analytic capabilities may compensate for some non-optimal organizational storage and retrieval practices, but may not provide an adequate overall solution for archival problems (Datskovsky, 2013).

MAIN FOCUS OF THE ARTICLE

Issues, Controversies, Problems

The varieties of virtual goods accumulated by individuals have evolved from simple e-mail address rosters to complex fantasy sports strategies and extraordinarily detailed avatars (Ng and Höpfl, 2011; Oravec, 2015). Good (2013) categorizes an assortment of the “personal media archives” that are often collected and maintained by households, assemblages that include digital images associated with vacations and other family events. Some social media platforms have served to support hoard-

ing behavior in archival activities: for example, Schiele and Hughes (2013) describe hoarding behavior linked with the Pinterest platform, relating how the accumulation of thousands of images becomes part of everyday life in many households. They assert that “consumers lead second lives online claiming ownership to virtual goods and images, using social media to create, control and consume content” (p. 47). In the 1990s, the emergence of the music sharing program Napster brought considerable attention to the hoarding of digital entities, with an emphasis on intellectual property concerns (Newman, 2013). BitTorrent and other kinds of peer-to-peer networks helped to create a “sharing culture” in which hoarding of virtual goods without subsequently allowing others to download them was often portrayed as inappropriate and even “selfish” behavior. The “hoarding ratios” were high for those who accumulated digital materials from the network without making them subsequently available to others (Rando, 2014).

The hoarding of virtual items has not yet been shown to be a direct substitute in psychological terms for the hoarding of physical items. Researchers have provided few clues as to whether the individuals who collect massive quantities of physical goods will have tendencies to collect large amounts of virtual goods as well. So far, little research has been done linking the two and asking whether hoarders in physical realms have tendencies to hoard in virtual platforms as well (Oravec, 2015). Hogan (2015) describes a perspective that may shed light on virtual hoarding phenomena: “the more we encourage the mass hoarding of digital media based on this dumpster model, the more we reinforce the logic of the always-on, always-ready archive” (p. 20). The casualness with which large data repositories can be stored without strategic consideration of how or whether they will be accessed in the future can make individuals feel temporary security; after all, they did not delete the files so they supposedly will be able to access them at a later point, even if the forensics involved (or the excavation, in terms of massive

hoarding) would be expensive. Some of these data resources can subsequently be subpoenaed as part of legal claims concerning various items or services provided by organizations.

Concerns about hoarding are serving to bridge the physical and virtual realms; for instance, St-Pierre-Delorme & O'Connor (2016) developed virtual reality spaces that simulate personal hoarding settings in order to destigmatize the manipulation and storage of physical household objects. From a related perspective, there are some promising prospects for somehow displacing the hoarding of printed photographs and related materials with the accumulation of digital files (switching to the hoarding of digital videos rather than bulky tapes, etc.), in hopes that the virtual items can acquire comparable sentimental linkages to significant personal experiences. Many young people have collected tens of thousands of digital files with music and video content, possibly freeing a great deal of shelf space while maintaining some semblance of personal attachment to the enterprise. Digital music has become a commodity that "opens up greater opportunities for surveillance, advertising, consumption and technological interference" (Morris, 2010, p. 4). Denegri-Knott, Watkins, and Wood (2012) refer to some of these mass accumulation efforts as "stockpiling," which captures some of the lack of selectivity in many virtual compilations. In contrast, Kibby (2009) describes how the acts of accessing, tagging, and storing digital music files can instill meaning into these virtual objects and turn the files into a "collection" rather than just a hoard (p. 428). Sinn and Syn (2014) describe how Facebook became a part of some users' "personal documentation" and digital storage systems, affording a means for them to structure various household archival traces and records. Financial instruments and records in both digital and physical forms can be hoarded, often endangering household finances (Canale & Klontz, 2013).

Workplace settings can raise an assortment of related concerns about the hoarding of virtual entities, including social and ethical issues involving

opportunistic hoarding. Gormley and Gormley (2012) discuss how data hoarding is manifested in employee behavior, stating that "hoarding data can create a false sense of uncertainty avoidance" (p. 90) as individuals accumulate, but do not process and share, organizational information. Some of these hoarding activities can be strategic as individuals retain information that would otherwise be part of regular organization "information flow" for opportunistic reasons. Wiewiora, Trigunarsyah, Murphy, and Coffey (2013) present research on potential organizational culture influences in the levels of sharing versus hoarding in which workplace participants engage. Many organizations do not have clear policies concerning data management; distinctions can be difficult between what is retained because of organizational needs and what is "hoarded" ("Is It Workplace Hoarding," 2014). Managements often view the orderliness of the physical and virtual assets of their organizations to be of high value (Kim, 2014), so hoarding behavior related to virtual goods of various sorts may continue to be of concern. Best practices at the organizational level concerning deletion of unneeded data and other approaches can be of benefit in providing current or potential hoarders with some guidance (van Bussel, Smit, & van de Pas, 2015).

SOLUTIONS AND RECOMMENDATIONS

Some of the strategies that are being examined for the identification and treatment of physical hoarding behaviors may have parallels in virtual realms. *The Hoarding Handbook* (Bratotiis, Schmalisch, & Steketee, 2011) describes an assortment of programs for mitigating hoarding-related problems, some of which are online. Research is also needed on the difficult issue of whether or how to redesign environments so as to accommodate individuals who exhibit hoarding behavior. Labeling hoarders as "disabled" could be problematic in many organizational contexts,

entailing expenditures for counseling and perhaps also the redesign of some spaces (including virtual spaces) and related tasks. In the workplace, the hoarding of pornographic images in organization-owned computers and online networks has also continued; it is especially problematic because of the fact that the individuals involved could readily download and store the materials on their own digital devices (Oravec, 2012). As the demands for disposition of increasing numbers and varieties of virtual items expands, the phenomenon of hoarding is expanding, both in its organizational and personal manifestations. Providing basic support and empowerment for individuals in organizations is part of establishing a just society. Designing workplace environments in which the management of virtual goods can be performed in a less overwhelming, more humane, and less time-consumptive way can aid organizations and communities as well as individuals.

Mayer-Schönberger (2010) outlines some of the “virtues” of data deletion in our current, information-saturated society, including the forgetting of irrelevant and out-of-date information. However, some virtual hoarding behaviors and tendencies may indeed have positive aspects as some individuals in organizations inadvertently retain important communications, images, or documents that are later determined to be of importance. Hoardings can preserve various organizational traces and virtual entities in a relatively-unadulterated “time capsule” (Oravec, 2013). Unearthing these vital virtual entities could be difficult in a hoarding-related system, however, and efforts akin to “data fracking” may be required (Gray & Alles, 2015). Generational differences may also emerge as salient in understanding hoarding behavior and perhaps using these tendencies to best organizational advantage: for example, individuals who hoarded thousands of music files with the Napster platform in the 1990s may have different orientations toward the retaining of virtual goods than many of the individuals who managed personal collections of vinyl albums in the 1960s and 70s.

Marginal accumulation of virtual goods (such as digital documents and musical or video files) has been construed as having fairly little physical impact on the planet (Fried, 2014), except for the problems involved with the production of computing equipment and electricity, which can produce considerable levels of toxicity. Once the basic equipment is manufactured and placed in a secure, temperature-controlled environment, additional virtual goods can be added with little environmental damage. For example, social media users who hoard increasing numbers of friends on Facebook, often without knowing the individuals involved, may be considered as engaging in “hoarding” yet little damage is done to the planet’s ecosystem. In contrast, many varieties of hoarding of physical goods present substantial environmental and public health challenges, such as in the hoarding of pets or perishable food items. In terms of its physical footprint, the collections of images and videos that can be placed even on an inexpensive digital device can dwarf a comparable assortment of canvas, print, or film renditions. However, the hoarding of virtual materials can have other overall negative impacts on society, including potential intellectual property violations.

H

FUTURE RESEARCH DIRECTIONS

For individuals whose hoarding behavior has some linkages to concerns about memory losses (and their related anxieties), the continuous digital recording of everyday events may eventually provide some relief. Some individuals are already engaged in experiments in which their entire lives are being taped (“lifelogging”) so that any slippage in memory could be compensated (Gurrin, Smeaton, & Doherty, 2014). Many professionals, managers, and educators already have a great deal of their effort made “transparent” with extensive documentation, blogging, and surveillance (Oravec, 2004). Perhaps these dramatic ways of supplementing and even enhancing one’s memory will eventually serve to alter some of the dysfunctional

patterns that lead to hoarding, although they may result in different varieties of hoarding syndromes. The growing assortment of forms of “ephemeral communications” (messages that are intended to be received and subsequently disappear) can also present alternative perspectives on the perceived need to hoard digital materials. These genres of communications are supported in Snapchat and comparable messaging platforms (Ganzemuller, 2014). Future research will explore how individuals who need help dealing with virtual hoarding issues can call for help most effectively, in ways comparable to the requests for assistance of physical-good hoarders (Bratiotis, Davidow, Glossner, & Steketee, 2016).

Increasing attention is being directed to the hoarding of dark data and the dangers of indiscriminate storage of files that could possibly harbor dangerous security hazards. The term “dark data” is often used to refer to a database’s content that is “not carefully indexed and stored so it becomes nearly invisible to scientists and other potential users and therefore is more likely to remain underutilized and eventually lost” (Heidorn, 2008, p. 280; Douglass, Allard, Tenopir, Wu, & Frame, 2014). The Internet of Things (IoT) may produce even more opportunities for the production of dark data and their potential hoarding:

Every device will, in the future, be a possible Smart Component on the Internet. Many everyday devices already have a computer chip in them that people do not realize, which if connected to the Internet can start broadcasting or receiving information, either to each other (Machine to Machine or M2M) or to various collection or aggregation points. (Watson, 2015, p. 212)

With the advent of the IoT comes a dramatic increase in the numbers and kinds of devices that have storage capabilities with personal, household, or workplace data. Remote wiping and deletion have already acquired an assortment of mundane applications in deleting such data if these devices are lost or stolen (data that perhaps should have

been deleted long ago in regular information management protocols). For example, when individuals lose smartphones and tablets (or as these devices are stolen) deletion of their files can lessen the severity of the loss. Remote wiping and deletion could indeed be used in systems design to mitigate system problems involving hoarding; it has already been used in copyright-related cases. For instance, in 2009, the book *1984* by George Orwell (1950) was removed from some Kindle devices without the direct consent of the device owners, reportedly for copyright concerns (Van Tassel, 2011; Chan, 2012). Entities other than the immediate owner of an artifact can often update, alter, or delete its localized files; this may lessen the potential for stores of suspicious data in specific IoT artifacts, mobile devices, or other networked entity to endanger the overall systems in which they are embedded.

CONCLUSION

Virtual hoarding behaviors can illuminate a number of issues related to data retention and dissemination in organizational settings and personal life. Relatively few individuals go to extremes, jeopardizing their livelihoods and social lives in the organizational or personal hoarding of data. However, the behaviors of a number of individuals in avoiding the application of standard organizational archiving practices can lead to larger cumulative losses for organizations. From the household and personal perspective, ignoring copyright and choosing to accumulate many thousands of files may serve to erode some of the social supports for the information economy and the funding of creative artists (Newman, 2013). Insights from hoarding of physical goods may help us to understand what is involved in the hoarding of virtual goods. Preston, Kringelbach, Knutson, and Whybrow (2014) propose an “interdisciplinary science of consumption” to develop and conduct research on some of these potential insights and interactions.

Framing cloud services and organizational data archives as kinds of “dumpsters” (Hogan, 2015), ones that can be easily searched or readily discarded, can present damaging prospects for organizations. In order to retain their optimal usefulness, data sources require adequate metadata as well as continuing management and oversight. Many individuals have hoarding tendencies and can emphasize with the “pack rat” syndrome, but such behaviors in workplace contexts can be deleterious to the organization in an assortment of dimensions.

REFERENCES

- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.
- Bratlotis, C., Davidow, J., Glossner, K., & Steketee, G. (2016). Requests for help with hoarding: Who needs what from whom? *Practice Innovations*, 1(1), 82–88. doi:10.1037/pri0000017
- Bratlotis, C., Schmalisch, C. S., & Steketee, G. (2011). *The hoarding handbook: A guide for human service professionals*. New York: Oxford University Press.
- Canale, A., & Klontz, B. (2013). Hoarding disorder: Its more than just an obsession—Implications for financial therapists and planners. *Journal of Financial Therapy*, 4(2), 43–63. doi:10.4148/1944-9771.1053
- Chan, G. K. (2012). Downstream alteration of copyrighted works in a world of licensed, digital distribution. *Columbia Journal of Law & Arts*, 36(2), 261–281.
- Datskovsky, G. (2013). Harnessing big data for competitive advantage. *Information & Management*, 47(2), S6.
- Denegri-Knott, J., Watkins, R., & Wood, J. (2012). Transforming digital virtual goods into meaningful possessions. In M. Molesworth & J. Denegri Knott (Eds.), *Digital virtual consumption* (pp. 76–91). New York: Routledge.
- Douglass, K., Allard, S., Tenopir, C., Wu, L., & Frame, M. (2014). Managing scientific data as public assets: Data sharing practices and policies among fulltime government employees. *Journal of the Association for Information Science and Technology*, 65(2), 251–262. doi:10.1002/asi.22988
- Freedman, A. (2015). Managing personal device use in the workplace. *Suffolk Journal of Trial & Appellate Advocacy*, 20, 284–361.
- Fried, D. (2014). Compulsive hoarding: Psychopathologies of print, phenomenologies of text. *Culture, Theory & Critique*, 55(3), 342–364. doi:10.1080/14735784.2014.892768
- Ganzenmuller, R. G. (2014). Snap and destroy: Preservation issues for ephemeral communications. *Buffalo Law Review*, 62, 1239–1288.
- Good, K. (2013). From scrapbook to Facebook: A history of personal media assemblage and archives. *New Media & Society*, 15(4), 557–573. doi:10.1177/1461444812458432
- Gormley, C., & Gormley, S. (2012). Data hoarding and information clutter: The impact on cost, life span of data, effectiveness, sharing, productivity, and knowledge management culture. *Issues in Information Systems*, 13(2), 90–95.
- Grant, J. E. (2014). Obsessive-compulsive disorder. *The New England Journal of Medicine*, 371(7), 646–653. doi:10.1056/NEJMcp1402176 PMID:25119610
- Gray, G. L., & Alles, M. (2015). Data fracking strategy: Why management accountants need it. *Management Accounting Quarterly*, 16(3).

- Gurrin, C., Smeaton, A. F., & Doherty, A. R. (2014). Lifelogging: Personal big data. *Foundations and Trends in Information Retrieval*, 8(1), 1-125.
- Heidorn, P. B. (2008). Shedding light on the dark data in the long tail of science. *Library Trends*, 57(2), 280–299. doi:10.1353/lib.0.0036
- Hogan, M. (2015). The archive as dumpster. *Pivot: A Journal of Interdisciplinary Studies and Thought*, 4(1), 7-38.
- Holten, A. L., Hancock, G. R., Persson, R., Hansen, Å. M., & Høgh, A. (2016). Knowledge hoarding. *Journal of Knowledge Management*, 20(2), 215–229. doi:10.1108/JKM-06-2015-0222
- Is It Workplace Hoarding... Or Record-Keeping? (2014). *Administrative Professional Today* 40(3), 3.
- Kibby, M. (2009). Collect yourself: Negotiating personal music archives. *Information Communication and Society*, 12(3), 428–443. doi:10.1080/13691180802660644
- Kim, S. E. (2014). Physical workplace as a strategic asset for improving performance in public organizations. *Administration & Society*, 46(5), 496–518. doi:10.1177/0095399713479104
- Lepselter, S. (2011). The disorder of things: Hoarding narratives in popular media. *Anthropological Quarterly*, 84(4), 919–947. doi:10.1353/anq.2011.0053
- Mayer-Schönberger, V. (2010). *Delete: The virtue of forgetting in the digital age*. Princeton, NJ: Princeton University Press.
- Morris, J. W. (2010). *Understanding the digital music commodity* (Doctoral dissertation). McGill University.
- Neal-Barnett, A., & Mendelson, L. (2003). Obsessive Compulsive Disorder in the workplace: An invisible disability. *Women & Therapy*, 26(1/2), 169–178. doi:10.1300/J015v26n01_11
- Newman, J. M. (2013). Copyright freeconomics. *Vanderbilt Law Review*, 66, 1407–1453.
- Ng, Y. R., & Höpfl, H. (2011). Objects in exile: The intimate structures of resistance and consolation. *Journal of Organizational Change Management*, 24(6), 751–766. doi:10.1108/09534811111175733
- Oravec, J. A. (2004). The transparent knowledge worker: Weblogs and reputation mechanisms in KM systems. *International Journal of Technology Management*, 28(7-8), 767–775. doi:10.1504/IJTM.2004.005782
- Oravec, J. A. (2012). The ethics of sexting: Issues involving consent and the production of intimate content. In D. Heider & A. L. Massanari (Eds.), *Digital Ethics* (pp. 129–145). New York: Peter Lang.
- Oravec, J. A. (2013). Not now, perhaps later: Time capsules as communications with the future. *Leonardo Electronic Almanac*, 19(2), 72–83.
- Oravec, J. A. (2015). Depraved, distracted, disabled, or just pack rats? Workplace hoarding persona in physical and virtual realms. *Persona Studies*, 1(2), 75–87. doi:10.21153/ps2015vol-1no2art472
- Oravec, J. A. (2017). The manipulation of scholarly rating and measurement systems: Constructing excellence in an era of academic stardom. *Teaching in Higher Education*, 22(4), 423–436. doi:10.1080/13562517.2017.1301909
- Orwell, G. (1950). 1984. Signet Classics.
- Peyton, A. (2015). Kill the dinosaurs, and other tips for achieving technical competence in your law practice. *Richmond Journal of Law & Technology*, 21, 1–27.
- Preston, S. D., Kringelbach, M. L., Knutson, B., & Whybrow, P. C. (2014). *The Interdisciplinary Science of Consumption*. Cambridge, MA: MIT Press. doi:10.7551/mitpress/9780262027670.001.0001

Rando, D. P. (2014). The music of wish images: Walter Benjamin, filesharing, and utopia. *Journal of Popular Music Studies*, 26(2-3), 321–345. doi:10.1111/jpms.12079

Schiele, K., & Hughes, U. M. (2013). Possession rituals of the digital consumer: A study of Pinterest. In G. Cornelissen, E. Reutskaja, & A. Valenzuela (Eds.), *E-European Advances in Consumer Research* (Vol. 10, pp. 47–50). Duluth, MN: Association for Consumer Research.

Sinn, D., & Syn, S. Y. (2014). Personal documentation on a social network site: Facebook, a collection of moments from your life? *Archival Science*, 14(2), 95–124. doi:10.1007/s10502-013-9208-7

St-Pierre-Delorme, M.-E., & OConnor, K. (2016). Using Virtual Reality in the Inference-Based Treatment of Compulsive Hoarding. *Frontiers in Public Health*, 4, 149. doi:10.3389/fpubh.2016.00149 PMID:27486574

van Bennekom, M. J., Blom, R. M., Vulink, N., & Denys, D. (2015). A case of digital hoarding. *BMJ Case Reports*, bcr2015210814.

van Bussel, G. J., Smit, N., & van de Pas, J. (2015). Digital archiving, green IT and environment: Deleting data to manage critical effects of the data deluge. *Electronic Journal of Information Systems Evaluation*, 18(2), 187–197.

Van Tassel, J. E. (2011). Remote deletion technology, license agreements, and the distribution of copyrighted works. *Virginia Law Review*, 1223–1261.

Watson, D. L. (2015, September). Some security perils of smart living. In *International Conference on Global Security, Safety, and Sustainability* (pp. 211–227). New York: Springer International Publishing.

Wiewiora, A., Trigunaryah, B., Murphy, G., & Coffey, V. (2013). Organizational culture and willingness to share knowledge: A competing values perspective in Australian context. *International Journal of Project Management*, 31(8), 1163–1174. doi:10.1016/j.ijproman.2012.12.014

ADDITIONAL READING

Foster, G. (2014). *Hoarders, Doomsday Preppers, and the Culture of Apocalypse*. New York: Palgrave Macmillan. doi:10.1057/9781137468086

Steketee, G., & Frost, R. (2010). *Stuff: Compulsive Hoarding and the Meaning of Things*. New York: Houghton Mifflin Harcourt.

Vickers, B. D., & Preston, S. D. (2014). The economics of hoarding. *The Oxford Handbook of Hoarding and Acquiring*, 221.

KEY TERMS AND DEFINITIONS

Dark Data: Data that are currently inaccessible or unusable by researchers and users (Heidorn, 2008). Data can be “dark” if they are not stored in appropriate ways or provided with metadata or context.

Ephemeral Communications: Communications that are deleted or otherwise not made available after a certain, relatively-short time period; the Internet messaging application Snapchat provides a platform for such interactions.

Information Flow: The circulation of information between and among various entities and individuals. In organizational contexts, information flow can be blocked if individuals hoard information rather than distributing it when needed.

Intellectual Property: Many kinds of virtual entities have financial and productive value; intellectual property considerations structure how these entities are owned and/or controlled by individuals and organizations for certain periods of time. Intellectual property can include copyrights, trademarks, and patents.

OCD: Obsessive-compulsive disorder: A “common, disabling psychiatric disorder characterized by intrusive and unwanted thoughts, images, or urges that cause distress or anxiety and repetitive thoughts or actions” (Grant, 2014, p. 646). It has been linked to some forms of hoarding behavior.

Remote Deletion (or “Remote Wiping”):

Deletion of digital material from a device or artifact (through a networked connection) without direct physical possession of the device or artifact. This procedure can be used to remove potentially-sensitive material from a stolen phone; it can also be used to delete problematic or suspicious files from an Internet of Things (IoT) device.

Virtual Goods: These are non-physical items that are instantiated in some physical and/or electronic media. Virtual goods are often associated with specific intellectual property categories and values (see “Intellectual Property” above).

Category H

Human Resources Management

Cyberloafing and Constructive Recreation

Jo Ann Oravec

University of Wisconsin – Whitewater, USA

INTRODUCTION

“Cyberloafing” in workplace and educational contexts refers to the uses of computer-related applications and devices in ways or at times that are not directly sanctioned by employers, managers, or teachers. It has often been considered as a kind of “time theft” on the part of employees and students (as described in Block, 2001), possibly decreasing workplace and educational productivity by consuming attention, energies, and resources designated for organizational operations. In contrast, many employees and students have construed cyberloafing as a stress reliever and as a support for personal wellbeing, often with the rationale that they are able to engage effectively in alternating or multitasking between and among their various work and off-work endeavors (Adler & Benbunan-Fich, 2013). “Constructive recreation,” in contrast with cyberloafing, comprises online recreation and gamification initiatives that are designed by employees along with managers; these initiatives are designed to be in synch with productive efforts and support the wellbeing of all organizational participants (Oravec, 2002; 2004a). This article compares and contrasts cyberloafing processes with constructive recreation approaches, the latter involving conscientious consideration of how online leisure activities can enhance workplace and educational activity and improve organizational productivity. The article analyzes some current research trends and public discourse related to cyberloafing; it also describes some constructive recreation approaches that have been explored over the past decades.

Many recent computing technology and gamification advances have helped to blur the conceptual and pragmatic boundaries between

“work” and “play,” distinctions that have great cultural variation (Deterding, Dixon, Khaled, & Nacke, 2011; Oravec, 2015). This article indeed focuses on cyberloafing issues within the US and UK, but international dimensions can also become especially salient in a world of globalized corporate interactions and relations. Cyberloafing practices that are acceptable in one nation may be seen in harsher lights in other places, given cultural and ethical differences that affect how work is structured and evaluated (Cheng, Li, Zhai, & Smyth, 2014; Sheikh, Atashgah, & Adibzadegan, 2015). International and regional variations in organizational approaches to cyberloafing can illuminate other significant aspects of workplace culture (Canaan Messarra, Karkoularian, & McCarthy, 2011), variations that can become salient as many organizations deal with international outsourcers or with units that are rooted in various nations.

BACKGROUND

The term “cyberloafing” emerged in academic and popular discourse in the 1990s as a way of characterizing the growing phenomenon of non-sanctioned online recreation and other activities (Oravec, 2002). Cyberloafing has been defined by Lim (2002) as the “act of employees using their companies’ internet access for personal purposes during work hours” (p. 675). Researchers have generated associated concepts such as Anandaraman, Simmers, and D’Ovidio’s (2011) framing of “personal web usage” or PWR, which is a more specific characterization of cyberloafing activity. As outlined in the “Key Words and Definitions” section below, “cyberbludging” is often used in some nations to refer to individuals’ use of comput-

DOI: 10.4018/978-1-5225-2255-3.ch374

ers in the workplace for recreation while shirking responsibility for work outcomes (Hernandez-Castro, 2016; Liaskos & Sandy, 2004). An alternate term, “cyberslacking,” is often used to label computer usage by students or younger members of the workforce (although it sometimes also emerges in broader discussions of the issue). Other characterizations of these phenomena include “non-work-related computing, cyber deviance, personal use at work, Internet abuse, workplace Internet leisure browsing, and junk computing” (Vitak, Crouse, & LaRose, 2011, p. 1751). Cyberloafing has also been examined as an aspect of the larger construct of “counterproductive workplace behavior” (as formulated in O’Neill, Hambley, and Bercovich, 2014), which also refers to a broader assortment of online workplace dysfunctions such as hacking, manipulation of metrics, and theft of resources (Oravec, 2017). Burnay, Billieux, Blairy, & Larøi (2015) linked some forms of cyberloafing with “obsessive passion” (p. 28). Askew (2012) explores the relationship between cyberloafing and task performance, and Jia, Jia, and Karau (2013) examines the influences of personality and workplace situational factors on cyberloafing. Some researchers portray cyberloafing as a kind of “deviance” (Weatherbee, 2010), despite the wide proliferation of cyberloafing behaviors and their normalization in some settings.

Propositions that employees or students should be forced to focus only on their assigned tasks and not be able to take reasonable discretionary breaks run counter to a great deal of managerial, psychological, and educational research (Lim, 2002; Oravec, 1996). For many decades, social scientists have given support to the assumption that human beings generally require some sort of periodic relief and refreshment as they pursue their workplace and educational endeavors (Metcalf, 1952; Souter, 1940). However, the open access that many employees and students have to computer devices and networking has often set up “contested spaces” in a number of workplaces and schools (Oravec, 2004a); these contestations refer to disputes and confrontations over the right to be

able to define what kinds of diversions and times for recreational activity are deemed as acceptable in the workplace or educational institution.

Cyberloafing issues have changed dramatically in character over the decades. When organizational computer resources were very expensive and tightly controlled (in the 1950s through the early 1970s), informal and unregulated access by employees and students was not considered an issue of widespread importance. Only a relatively few individuals were able to explore the uses of computing technology for recreation in the workplace, although their surreptitious initiatives ultimately resulted in some of the early popular video games (Wolf, 2001). Since the 1970s and the advent of the desktop personal computer (PC), however, opportunities for the non-sanctioned use of computers and networking resources expanded dramatically. Employees could access video games on their desks at work as well as engage in computer network interaction in early electronic bulletin boards or listservs (Oravec, 1996). They could also compose their resumes and cover letters during work breaks and produce them on their employers’ printers for a professional-looking product. The “boss key” or “boss button” became a feature on many applications, giving employees an easy way to appear to be working hard on an organizationally-sanctioned application while engaging in unsanctioned efforts (van Gelder, 1982); by pressing a certain key, usually a function key, the PC’s screen would change to a spreadsheet or other business-oriented image. In that era, managers and administrators at nearly all levels of organizational functioning were faced with the prospect that some of their employees could be utilizing computer technology on company time, possibly consuming expensive organizational resources as well. Cyberloafing provided a wake-up call to managers about the power of computer technology to change everyday organizational functioning, often in complex and dysfunctional ways.

Many of the pioneering research and policy initiatives involving cyberloafing were formulated

well before the advent of ubiquitous social media (such as Facebook and Twitter); some approaches still embody the model of the solitary individual who is involved in cyberloafing by interacting with a particular computer application or online resource rather than using computing to extend and enhance group and community interaction (Oravec, 2002). Social media have expanded the range and amount of cyberloafing activity; it has expanded into more of a communal experience and has linked employees for the sharing of pictures, texts, and videos while at work. For example, through Facebook, Twitter, and YouTube, individuals could engage with each other in workplaces and schools, often creating “virtual breakrooms” or “cyber watercoolers” that stretched across organizational settings. In the past decades in Silicon Valley, employees who devoted many hours online to their intense and demanding jobs developed an assortment of innovative ways to spend their available free time online without straying too far from their workplace settings (Tapia, 2004). Other iterations of cyberloafing have also become more intensely personal and intimate as individuals use their wearable technologies to monitor their medical conditions and enhance their personal experiences while on the job.

MAIN FOCUS OF THE ARTICLE

Issues, Controversies, Problems

Research disciplines that have focused on cyberloafing issues have ranged from information technology and security (as demonstrated by Glassman, Prosch, & Shao, 2015), organizational behavior and leadership studies (Lim, 2002; Jia, Jia, & Karau, 2013), psychology (Wagner *et al.*, 2012), public administration (Hernandez-Castro, 2016; Ozler & Polat, 2012), human resources (Weatherbee, 2010), education (Taneja, Fiore, & Fischer, 2015), and business ethics (Block, 2011). The emerging interdisciplinary field of “cyberpsychology” also developed as a vehicle for studying

cyberloafing and other emerging computer-related phenomena (Anandarajan, Simmers, & D’Ovidio, 2011; Oravec, 2015). With this broad range of research directions and approaches, sometimes confusing amalgams of portraits of cyberloafing phenomena have emerged. From a negative perspective, researchers from various nations have documented billions of lost dollars of productivity over the past decades because of uses of computers in the workplace for non-sanctioned activity (Sheikh, Atashgah, & Adibzadegan, 2015). The opportunity costs involved with cyberloafing are often not included in these estimates; they are difficult to figure as employees and students direct their creative energies toward recreational computer applications other than those relevant for work or school. Blanchard and Henle (2008) distinguish between “minor” cyberloafing (which includes sending and receiving personal email) and “serious” cyberloafing (which includes online gambling as well as interacting with adult-oriented chatrooms and websites). Since the 2000s, “sexting” was identified as a major workplace and school concern, with many cases of managers, employees, and students spending time online during work hours sending or accessing sexually-suggestive messages via computer or phone (Oravec, 2012). Sexting, along with other kinds of serious cyberloafing, engendered legal as well as moral worries for business and educational institutions.

An example of how pervasive and influential cyberloafing can become in organizations involves online shopping. In the early days of the proliferation of the Internet, some kinds of commerce-related cyberloafing were overlooked by managers because they wanted employees to become comfortable with online access and perhaps even obtain useful health or other services online (Belanger & Van Slyke, 2002; Oravec, 2001). Phenomena emerged such as “cyber Monday” (the Monday morning after the popular US post-Thanksgiving shopping day “Black Friday”) in which employees often openly engaged in online shopping with the aid of the computing power

they had on their desks at work, technological access that was often greater than that available at home (Swilley & Goldsmith, 2013). Many online shopping initiatives were indeed conducted with workplace technology rather than home computers because of the increased bandwidth provided in workplace settings. However, as home computing and smartphones proliferated, the need for accommodating such workplace experimentation with computing lessened dramatically. Many employees and managers acquired computer skills at home with the increasingly-inexpensive home and mobile computing technologies, often inspired or driven by their children and friends. Employer tolerance for cyberloafing subsequently diminished in many organizations (Oravec, 2002).

Students using laptops, smartphones, and related devices have ready access to cyberloafing opportunities; restricting the use of these devices has been very difficult in schools and higher education since computers are closely integrated into a great deal of classroom activities. Even if schools install surveillance software on their computer systems, students can use smartphone and smartwatches for non-sanctioned classroom activity. In educational settings, cyberloafing can distract students from maintaining their attention on the teachers' or their peers' presentations during class time, possibly disabling their efforts to achieve their overall educational goals (Taneja, Fiore, & Fischer, 2015). The stress-relieving functions of cyberloafing for students may provide a way for them to remain physically present in a classroom setting while being virtually present elsewhere. Unauthorized use of computing in education is not restricted to students: McBride, Milligan, and Nichols (2013) explored cyberloafing activities among teachers who engaged in non-sanctioned activity while on the job. Education-related issues also involve the engagement of employees in educational activities while at work in circumstances in which their attention is expected to be directed at non-educational tasks. It may seem odd to label educational activities in terms of "loafing," but online educational activities may distract or inter-

fere with critical workplace functions at certain times, possibly leading to workplace mistakes or accidents.

A major issue involving cyberloafing is that of the control and direction of unscheduled and unauthorized computer uses. Many employers, managers, and administrators recognize the importance of stress relief in workplaces and educational institutions. However, many workplace participants are concerned when the organization and not the individual or group determines what computer-enhanced recreation opportunities would be allowed (Belanger & Van Slyke, 2002). The discretionary aspects of cyberloafing have become even more complex as employees and students use their own devices to engage in everyday organizational online interaction, thus often placing themselves outside of the direct surveillance of employers and teachers. The feelings of entitlement to access one's own smartphone or laptop at work can play a role in how cyberloafing behaviors are construed by the parties involved. Individuals have indeed used personal electronics from Sony Walkmans to the latest wearable technologies for decades to provide needed stimulation or relaxation in work environments. However, with the increased cognitive demands involved with many computer applications, new varieties of organizational productivity and safety issues have emerged as individuals attempt to divide their attentions between their engrossing, immersive personal activities and their organization-sanctioned projects. Individual differences are a factor here: some people may indeed be incapable of managing the increased cognitive demands involved with task-break alternation and multitasking while others may be quite comfortable and even improve their performance with the additional stimuli (Adler & Benbunan-Fich, 2013).

Certain aspects of cyberloafing can expose organizations to legal dangers if allowed to continue unchecked. For example, employees and students may access online materials on the job or during class hours that are expressly forbidden in work environments or in society as a whole,

including child pornography or certain forms of illegal gambling. If organizations develop and disseminate clear policies on such problematic access these legal and ethical issues will not be completely solved but at least proactive steps can be taken to address them (Oravec, 2002). Research studies by Ugrin and Pearson (2013) show that “threats, termination, and detection mechanisms are effective deterrents against activities like viewing pornography”(p. 812) but can be less effective toward mitigating some other kinds of cyberloafing (such as personal emailing and social networking).

SOLUTIONS AND RECOMMENDATIONS

Cyberloafing activities can provide useful signals to employers, managers, administrators, and teachers about how tasks are being designed and roles understood. Researchers have linked cyberloafing with such dysfunctions as job burnout (Aghaz & Sheikh, 2016; Oravec, 2002) and procrastination (Lavoie & Pychyl, 2001). Employees or students who have a great deal of time on their hands could turn to cyberloafing simply as a way to mitigate boredom. In some contexts, allowing individuals the leeway to engage in cyberloafing when there is little to do on the job than wait for the next customer can improve employee morale and actually improve overall productivity; this form of personal autonomy may help the employee retain alertness and optimism.

Cyberloafing research and policy development have often served to expand research and practitioner perspectives on the work/play dichotomy. Play (including unfettered exploration in a task-related environment as well as experimentation in personal expression) has been shown to help individuals prepare for the unexpected by presenting varying streams of novel or challenging situations (Oravec, 2015). Play is often construed as an underpinning for children’s intellectual and social development (Oravec, 2000), but how it functions for adults is

less clear. Creating venues for stimulating and invigorating online play in organizations can be a part of constructive recreation. These constructive recreation approaches are not designed to be coercive or compulsive recreation (forced on employees), but opportunities for exploration and expression that are in keeping with the social and cultural settings of the workplace.

Technological approaches to constraining cyberloafing have included surveillance of employee and student activity (Glassman, Prosch, & Shao, 2015). However, monitoring of individuals’ activities in many kinds of jobs (especially knowledge work, creative activity, and technological design) may not provide an accurate measure of their productivity (Oravec, 2004b). Social and policy approaches to cyberloafing have ranged from draconian rules that employees be fired for non-sanctioned computer use to perspectives that are more lenient and situation-sensitive (Block, 2001). Participatory and peer-oriented strategies that permit teams to establish their own affordances and limitations concerning online recreation can help to integrate the knowledge that team members have about their local circumstances as well as harness peer pressure to help deal with specific problems. Simply creating a climate of fear, one in which some small slippage or discretionary break could result in a large penalty) can serve to undermine the benefits of advanced technology innovations and constrain employee openness and creative thinking (Zoghbi-Manrique-de-Lara, 2006).

FUTURE RESEARCH DIRECTIONS

Future directions for research on cyberloafing include its expanding group or collaborative orientations, moving away from the model of cyberloafing primarily being conducted by a solitary and isolated individual. Group-oriented or collaborative cyberloafing provides an attractive venue for individuals to become acquainted or deepen relationships in organizational contexts. However, when individual employees or students

use computers to structure large recreational projects (especially involving sports) dangers can emerge. For example, use of computer networking for office betting pools during “fantasy” sports or major athletic competitions may indeed enhance group cohesion but can also present sizable distractions, often transforming the character of the workplace and diminishing productivity. Research is needed as to how far these large-scale recreational efforts can be supported by organizations before social and task-related damage can occur, including injury to those who are excluded by the group or choose to exclude themselves (for whatever reason).

Cyberloafing is becoming an increasing concern in educational and workplace contexts as the opportunities to spend time on devices and with applications not related to school and work efforts increase. “Eduloafing” can also emerge as a problem as employees use computers to access educational applications in a way that is not in synch with their other organizational activities, potentially resulting in safety and distraction issues. Future directions in research involve understanding why individuals engage in cyberloafing and whether organizations can develop strategies for ensuring that whatever cyberloafing occurs is life-enhancing as well as conducive to productivity. For example, Wagner et al. (2012) explore how the lost sleep time of employees related to daylight savings time shifts can affect the amount of cyberloafing conducted in workplaces. Employee and student self-awareness of individual capabilities and limitations in terms of task alternation and multitasking could help in understanding these phenomena (Adler & Benbunan-Fich, 2013). Compulsive use of online recreation during work or school could signal larger, serious issues involving Internet addiction. Like many other forms of mental health issues, Internet addiction could be debilitating for individuals who cannot balance their online activities with other aspects of their lives; it may need to be handled with advanced medical consideration (Burnay, Billieux, Blairy, & Larøi, 2015).

Yet another emerging research direction involves the “gamification” of workplace and educational efforts (Deterding, Dixon, Khaled, & Nacke, 2011). A growing number of gamification and multigamification initiatives (ways of using gaming approaches and methodologies in everyday activity) have been directed toward infusing game-like components to organizational activity, in part as stress relief but also to increase the engagement of organizational participants (Oravec, 2015). In the near future with gamification initiatives, computer technology applications in workplace and classroom settings may be so attractive and engaging that the basic concept of “cyberloafing” could be problematic; work itself may become increasingly appealing if not addictive (as it apparently already is in some workplaces).

CONCLUSION

Cyberloafing has been shown by some researchers to be a stress reducer, often playing roles in enhancing job performance and even in stimulating creativity. In workplaces in which long hours and intense focus are the norm, such diversion is often perceived as essential in maintaining mental health (Tapia, 2004). However, other researchers and practitioners have stressed the potential hazards cyberloafing presents in terms of decreases in productivity as well as increases in safety concerns as individuals are distracted or disoriented (Lim & Chen, 2012). Situational approaches that incorporate “constructive recreation” strategies can help organizational participants deliberate safety and social problems and map out potential solutions (Oravec, 2002; 2004a). For example, in many knowledge work settings in which computing technology plays substantial roles cyberloafing may be an acceptable way of taking a break in order to regain perspective or generate new ideas; in other settings (such as controlling heavy machinery) a similar approach toward taking breaks could result in tragedy. Such technologies as smartphones and smartwatches make it increasingly feasible

for employees and students to be surreptitious in their cyberloafing. Important and complex new issues are also emerging as individuals “eduloaf” by engaging in immersive educational experiences while on the job, possibly setting themselves up for cognitive overload.

As outlined in this article, cyberloafing phenomena have changed in character over the past several decades with advances in technology making computing more collaborative (with social media) and more personal (with mobile and wearable technologies). Cyberloafing will probably take on even more new forms as computer technology is integrated ever more closely into our everyday lives. The individual who focuses on his or her smartwatch to ascertain stress levels or monitor heartbeats may be conceived as a form of “cyberloafing” (for taking time out of assigned work hours for these tasks). However, without monitoring the wearable technology, the employee may eventually suffer in terms of health. If this individual works in an air controller facility and has jets to monitor, the situation may be viewed as quite different from that of the office worker who is handling routine paperwork and takes a quick break between cases. Equipping employees and students to be aware of their personal capabilities in terms of alternating between work and play, as well as to be conscientious and sensitive to others’ needs in their use of computer technology in recreation, will be part of the overall movement in organizations to mitigate the dysfunctional aspects of cyberloafing.

REFERENCES

- Adler, R. F., & Benbunan-Fich, R. (2013). Self-interruptions in discretionary multitasking. *Computers in Human Behavior*, 29(4), 1441–1449. doi:10.1016/j.chb.2013.01.040
- Aghast, A., & Sheikh, A. (2016). Cyberloafing and job burnout: An investigation in the knowledge-intensive sector. *Computers in Human Behavior*, 62, 51–60. doi:10.1016/j.chb.2016.03.069
- Anandarajan, M., Simmers, C. A., & D’Ovidio, R. (2011). Exploring the underlying structure of personal web usage in the workplace. *Cyberpsychology, Behavior, and Social Networking*, 14(10), 577–583. doi:10.1089/cyber.2010.0136 PMID:21728817
- Askew, K. L. (2012). *The relationship between cyberloafing and task performance and an examination of the theory of planned behavior as a model of cyberloafing*. USF Graduate Theses and Dissertations. Retrieved from <http://scholarcommons.usf.edu/etd/3957>
- Belanger, F., & Van Slyke, C. (2002). Abuse or learning? *Communications of the ACM*, 45(1), 64–65. doi:10.1145/502269.502299
- Blanchard, A. L., & Henle, C. A. (2008). Correlates of different forms of cyberloafing: The role of norms and external locus of control. *Computers in Human Behavior*, 24(3), 1067–1084. doi:10.1016/j.chb.2007.03.008
- Block, W. (2001). Cyberslacking, business ethics and managerial economics. *Journal of Business Ethics*, 33(3), 225–231. doi:10.1023/A:1012002902693
- Burnay, J., Billieux, J., Blairy, S., & Larøi, F. (2015). Which psychological factors influence Internet addiction? Evidence through an integrative model. *Computers in Human Behavior*, 43, 28–34. doi:10.1016/j.chb.2014.10.039
- Canaan Messarra, L., Karkoulis, S., & McCarthy, R. (2011). To restrict or not to restrict personal internet usage on the job. *Education, Business and Society: Contemporary Middle Eastern Issues*, 4(4), 253–266. doi:10.1108/17537981111190042
- Cheng, L., Li, W., Zhai, Q., & Smyth, R. (2014). Understanding personal use of the Internet at work: An integrated model of neutralization techniques and general deterrence theory. *Computers in Human Behavior*, 38, 220–228. doi:10.1016/j.chb.2014.05.043

- Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011, September). From game design elements to gamefulness: defining gamification. In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments* (pp. 9-15). ACM. doi:10.1145/2181037.2181040
- Garrett, R. K., & Danziger, J. N. (2008). On cyberslacking: Workplace status and personal Internet use at work. *Cyberpsychology & Behavior*, 11(3), 287–292. doi:10.1089/cpb.2007.0146 PMID:18537498
- Glassman, J., Prosch, M., & Shao, B. B. (2015). To monitor or not to monitor: Effectiveness of a cyberloafing countermeasure. *Information & Management*, 52(2), 170–182. doi:10.1016/j.im.2014.08.001
- Hernandez-Castro, W. (2016). *An empirical assessment of employee cyberslacking in the public sector* (Order No. 10096094). Available from ProQuest Dissertations & Theses A&I. (1780296267). Retrieved from <http://search.proquest.com/docview/1780296267?accountid=14745>
- Holguin, E. S. (2016). *Strategies Functional Managers Use to Control Cyberloafing Behaviors*. ProQuest Dissertations & Theses A&I. (10141379)
- Jia, H., Jia, R., & Karau, S. (2013). Cyberloafing and personality: The impact of the Big Five traits and workplace situational factors. *Journal of Leadership & Organizational Studies*, 20(8), 358–365. doi:10.1177/1548051813488208
- Kim, S. J., & Byrne, S. (2011). Conceptualizing personal web usage in work contexts: A preliminary framework. *Computers in Human Behavior*, 27(6), 2271–2283. doi:10.1016/j.chb.2011.07.006
- Lavoie, J. A., & Pychyl, T. A. (2001). Cyberslacking and the procrastination superhighway: A web-based survey of online procrastination, attitudes, and emotion. *Social Science Computer Review*, 19(4), 431–444. doi:10.1177/089443930101900403
- Liaskos, J., & Sandy, G. A. (2004). An evaluation of Internet use policies of Victorian local government. *Australian Journal of Public Administration*, 63(4), 90–100. doi:10.1111/j.1467-8500.2004.00405.x
- Lim, V. K. (2002). The IT way of loafing on the job: Cyberloafing, neutralizing and organizational justice. *Journal of Organizational Behavior*, 23(5), 675–694. doi:10.1002/job.161
- Lim, V. K., & Chen, D. J. (2012). Cyberloafing at the workplace: Gain or drain on work? *Behaviour & Information Technology*, 31(4), 343–353. doi:10.1080/01449290903353054
- McBride, J., Milligan, J., & Nichols, J. (2013). “Cyberslacking” in the classroom: The reactions of classroom teachers. *College Student Journal*, 47(1), 212–218.
- Metcalf, H. G. (1952). Recreation education. *Journal of the American Association for Health, Physical Education, and Recreation*, 23(2), 19–20.
- ONeill, T. A., Hambley, L. A., & Bercovich, A. (2014). Prediction of cyberslacking when employees are working away from the office. *Computers in Human Behavior*, 34, 291–298. doi:10.1016/j.chb.2014.02.015
- Oravec, J. A. (1996). *Virtual individuals, virtual groups: Human dimensions of groupware and computer networking*. New York: Cambridge University Press. doi:10.1017/CBO9780511574986
- Oravec, J. A. (2000). Interactive toys and children’s education: Strategies for educators and parents. *Childhood Education*, 77(2), 81–85. doi:10.1080/00094056.2001.10521635
- Oravec, J. A. (2001). On the proper use of the internet: Self-help medical information and on-line health care. *Journal of Health & Social Policy*, 14(1), 37–60. doi:10.1300/J045v14n01_03 PMID:11374297

- Oravec, J. A. (2002). Constructive approaches to Internet recreation in the workplace. *Communications of the ACM*, 45(1), 60–63. doi:10.1145/502269.502298
- Oravec, J. A. (2004a). When work morphs into play: using constructive recreation to support the flexible workplace. In M. Anandarajan (Ed.), *Personal web usage in the workplace: a guide to effective human resource management* (pp. 46–60). Hershey, PA: Idea Group Publishing. doi:10.4018/978-1-59140-148-3.ch003
- Oravec, J. A. (2004b). The transparent knowledge worker: Weblogs and reputation mechanisms in KM systems. *International Journal of Technology Management*, 28(7-8), 767–775. doi:10.1504/IJTM.2004.005782
- Oravec, J. A. (2012). The ethics of sexting: Issues involving consent and the production of intimate content. In D. Heider & A. L. Massanari (Eds.), *Digital Ethics* (pp. 129–145). New York: Peter Lang.
- Oravec, J. A. (2015). Gamification and multigamification in the workplace: Expanding the ludic dimensions of work and challenging the work/play dichotomy. *Cyberpsychology (Brno)*, 9(3), 59–71. doi:10.5817/CP2015-3-6
- Oravec, J. A. (2017). The manipulation of scholarly rating and measurement systems: Constructing excellence in an era of academic stardom. *Teaching in Higher Education*, 22(4), 423–436. doi:10.1080/13562517.2017.1301909
- Ozler, D. E., & Polat, G. (2012). Cyberloafing phenomenon in organizations: Determinants and impacts. *International Journal of eBusiness and eGovernment Studies*, 4(2), 1-15.
- Sheikh, A., Atashgah, M. S., & Adibzadegan, M. (2015). The antecedents of cyberloafing: A case study in an Iranian copper industry. *Computers in Human Behavior*, 51, 172–179. doi:10.1016/j.chb.2015.04.042
- Souter, S. H. Jr. (1940). Why recreation for prisoners. *Jail Association Journal*, 2, 19.
- Swilley, E., & Goldsmith, R. E. (2013). Black Friday and Cyber Monday: Understanding consumer intentions on two major shopping days. *Journal of Retailing and Consumer Services*, 20(1), 43–50. doi:10.1016/j.jretconser.2012.10.003
- Taneja, A., Fiore, V., & Fischer, B. (2015). Cyber-slacking in the classroom: Potential for digital distraction in the new age. *Computers & Education*, 82, 141–151. doi:10.1016/j.compedu.2014.11.009
- Tapia, A. H. (2004). The power of myth in the IT workplace: Creating a 24-hour workday during the dot-com bubble. *Information Technology & People*, 17(3), 303–326. doi:10.1108/09593840410554201
- Ugrin, J. C., & Pearson, J. M. (2013). The effects of sanctions and stigmas on cyberloafing. *Computers in Human Behavior*, 29(3), 812–820. doi:10.1016/j.chb.2012.11.005
- van Gelder, L. (1982, December). Superplayers run the gauntlet. *PC Magazine*, 131.
- Vitak, J., Crouse, J., & LaRose, R. (2011). Personal Internet use at work: Understanding cyberslacking. *Computers in Human Behavior*, 27(5), 1751–1759. doi:10.1016/j.chb.2011.03.002
- Wagner, D. T., Barnes, C. M., Lim, V. K., & Ferris, D. L. (2012). Lost sleep and cyberloafing: Evidence from the laboratory and a daylight saving time quasi-experiment. *The Journal of Applied Psychology*, 97(5), 1068–1076. doi:10.1037/a0027557 PMID:22369272
- Weatherbee, T. G. (2010). Counterproductive use of technology at work: Information & communications technologies and cyberdeviancy. *Human Resource Management Review*, 20(1), 35–44. doi:10.1016/j.hrmr.2009.03.012
- Wolf, M. J. (2001). *The medium of the video game*. Austin, TX: University of Texas Press.

Zoghbi-Manrique-de-Lara, P. (2006). Fear in organizations: Does intimidation by formal punishment mediate the relationship between interactional justice and workplace internet deviance? *Journal of Managerial Psychology*, 21(6), 580–592. doi:10.1108/02683940610684418

KEY TERMS AND DEFINITIONS

Constructive Recreation: Recreation is “constructive” when it is in synch with pending work responsibilities, allowing individuals to use time not consumed by workplace demands in ways that equip them to face future tasks with greater energy and expanded perspectives (Oravec, 2002). Constructive recreation is also in keeping with technological constraints, as exemplified by the organizations that allow online recreation but place limits during certain hours to avoid system overload.

Cyberbludging (or “Cyber Bludging”): Workplace and community cyberloafing in Australia and some other nations is often labeled as “cyberbludging” (Hernandez-Castro, 2016; Liaskos & Sandy, 2004), with the word “bludging” having the connotation of a “shirking of responsibility.”

Cyberslacking: A use of computer equipment or computer networking interaction in a way that is not in keeping with the overall goals of the organization. The term is often associated with the activities of individuals who are younger in relation to the average age of employees and managers in the workplace.

Gamification: Generally characterizes approaches through which video games and other technology-enhanced gaming systems (such as video games and virtual reality) are being used to transform workplace activities. Strategies for gamification can vary from simple contests to

complex segments derived from familiar or newly-invented games.

Internet Addiction: Syndrome in which individuals engage in compulsive behavior in relation to their Internet activity. Researchers and practitioners have sometimes construed Internet addiction as merely one form of compulsivity and others differentiate it because of its computer-related focus and the particular qualities of Internet interaction.

Mindfulness: The state of engaging in conscientious and focused attention to the current task at hand and one’s own associated mental state, or upon a limited and constrained set of stimuli.

Multitasking: Attempts to parcel or divide attention between or among two or more tasks. In the context of cyberloafing, this would entail attempting to engage in recreational activity at the same time one is immersed in a workplace project.

Serious Games: Games that are directly developed for the purpose of enhancing some aspect of educational, political, social, or workplace interaction. Serious games can serve to deepen the players’ understandings of some dimensions of the human experience, often providing the capability to experiment with and explore in depth different aspects of social and interpersonal interaction.

Stress Reduction: Approaches to assist individuals in coping with stressors in various contexts, such as workplaces and schools. Stressors include noxious or toxic aspects of the social or physical environment as well as overbearing demands for personal performance.

Time Theft: A use of time on the job that involves matters not productively related to employer and organizational concerns. Individuals who engage in time theft can be seen as committing offenses against their employers as well as against their co-workers who must make up for their lost efforts.

Influencing People and Technology Using Human Resource Development (HRD) Philosophy

Claretha Hughes

University of Arkansas, USA

Matthew W. Gosney

University of Colorado – Health, USA

Cynthia M. Sims

Clemson University, USA

INTRODUCTION

Historically, and notably since Zakaria's (2010) statement that "technology and globalization are shattering the middle class" in America (p. 31), American workers are continuing to be displaced by the productivity gains of technology and the competition for cheap laborers in a globalized economy. Human Resource Development (HRD) professionals are seeking ways to address rapid, continuous changes in technology and some, unfounded, fears and concerns regarding the influence of globalization on middle class workers. Globalization efforts do not need to equal unemployment for American middle class workers. Globalization can add opportunities for American middle class workers to increase their job prospects by learning new skills including how to leverage technological innovations for their advantage within the global job market. HRD professionals have an opportunity to assist workers with their skill development and improvement.

HRD professionals are engaged with workers in training and development, career development, and organization development initiatives (Mankin, 2001; Swanson & Holton, 2001). These initiatives are vital to employee development within organizations and many employees seek training and development opportunities so that they can

build successful careers and the ability to adapt to organizational and marketplace changes.

Workers are beginning to recover from the Great Recession of 2007 and HRD professionals are tasked with making their employees' recovery more viable. Aguinis and Kraiger (2009) suggest that there was "[a]n important challenge for the practice of training... to integrate the training function with employee selection, performance, management, rewards, and other human resource practices (Aguinis, 2009; Aguinis & Pierce, 2008; Cascio & Aguinis, 2005)" (p. 467). Their assessment is currently still needed and will be needed for the foreseeable future as artificial intelligence using robots, computer technology, simulations, avatars, and other technological innovations are being used to realign and displace workers. The current post-fordism era is associated with significant changes to American culture, organizations, and individuals (Heffernan, 2000). However, the influence of technology on the development of people has rarely been discussed within the HRD research literature (Githens, Dirani, Gitonga, & Teng, 2008; Hughes, 2010; 2012).

HRD, a relatively new field of study, has scholars and professionals who are continuously examining and testing its theories and philosophy. Organizations are mutable and strive to succeed through the people and technology in which it

invests and employ. The management and development of both people and technology is essential to the competitive advantage of organizations. The management of these critical domains is led by individuals whom often have limited interaction and different vantage points through which they examine and determine success. The dynamics of HRD theory and philosophy on the relationships between people and technology within organizations is an area requiring more exploration within the HRD field.

This chapter explores the extent to which HRD philosophy influence the relationship between people and technology. We have examined whether or not HRD professionals and researchers deny that there is a viable relationship between people and technology, and whether or not HRD professionals and researchers are limiting the field because of their ethical beliefs which juxtapose people to technology.

BACKGROUND

HRD has evolved from sociology, to business, to education and is still looking for a place of its own within academia (McLean, Lynham, Azevedo, Lawrence, & Nafukho, 2008). HRD has a rich history that is not well known or explained in the HRD research literature (Gosney, 2014; Gosney & Hughes 2015). Without an explanation and chronicling of its history, HRD's search for a clear philosophy is continuous and debatable amongst HRD researchers and professionals.

HRD Theory Building

Swanson (2001) identifies three foundational theories of HRD: psychological theory, economic theory, and systems theory that make up the legs of his three-legged stool model; however, they are not the only theories that can support the field. Some of the theories that have contributed to defining HRD are as follows:

1. Commonly held theories of HRD (Weinberger, 1998);
2. Operational definitions of expertise and competence (Herling, 2000);
3. Organization development: An analysis of the definitions and dependent variables (Egan, 2002);
4. An investigation into core beliefs underlying the profession of HRD (Ruona, 1999); and
5. Philosophical foundations of HRD practice (Ruona & Roth, 2000).

There are also examples of HRD theory building efforts which include:

1. Systems theory applied to human resource development (Jacobs, 1989);
2. Foundations of performance improvement and implications for practice (Swanson, 1999);
3. A theory of intellectual capital (Harris, 2000);
4. A theory of knowledge management (Torraco, 2000); and
5. The development and validation of a model of responsible leadership and performance (Lynham, 2000).

These theory building efforts have all contributed to the emerging field of HRD; however, McLean et al. (2008) suggested that HRD has much to learn from "EVERY field and discipline that touches on human behavior" (p. 249). Gosney (2014) and Gosney and Hughes (2015) propose in Gosney's Model of Modern Era Theory & Practice Generation in HRD that HRD theory and practice is influenced from a variety of sources, including philosophy and psychology. Openness to new concepts and formulations provide opportunity for continuous development within HRD.

HRD Philosophy

An oftentimes unexplored element of theory-building is an examination of the philosophies upon which the theory rests. Philosophy, literally

defined as the love of wisdom (Jaspers, 1951), is the continuous search for understanding of how the universe works. Or, as aptly described by Paulsen (1907), “[i]n the most general sense of the term, philosophy is simply the continually-repeated attempt to arrive at a comprehensive and systematic knowledge of the form and connection, the meaning and import of all things” (p. 2-3). Inherent in that comprehensive and systematic knowledge are assumptions – assumptions of time, space, and purpose. As Whitehead (1926) noted, “Every philosophy is tinged with the colouring of some secret imaginative background, which never emerges explicitly into its trains of reasoning” (p. 9). Jaspers (1951) suggests that “there is no escape from philosophy. The question is only whether a philosophy is conscious or not, whether it is good or bad, muddled or clear. Anyone who rejects philosophy is himself unconsciously practicing a philosophy” (p. 12).

If we are to take Jaspers (1951) at his word, HRD practices a philosophy through the theories it espouses and methodologies it embraces. Those theories, then, carry with them assumptions – assumptions regarding free will, purpose, even of the nature of time (Gosney & Hughes, 2015). Gosney and Hughes (2015) explore the history of HRD with the specific intent of unearthing the underlying philosophies of HRD and the assumptions upon which those theories rest. They identify three major philosophical influences on HRD theory and practice: Empiricism, Humanism, and Systems Theory. Each philosophical pillar undergirds much of how HRD is realized both in theory and in practice today.

Empiricism as an informing philosophy became distinct in HRD at the rise of the Industrial Revolution (Gosney & Hughes, 2015) and evidences can be seen in both current theory and practice. As a discipline that borrows heavily from the social sciences (Reio & Batista, 2014), HRD has embraced an empirical methodology in testing theory (Callanan & Connor, 2015; Nimon & Astakhova, 2015). It has also leaned upon empirical and positivist philosophy in the development

of tools such as personality assessments (Lohman, 2004; Messmann & Mulder, 2012), selection and job-fit assessment (Berr, Church, & Waclawski, 2000), and instructional design and learning (Mazur, 1994; Wilson, Jonnassen, & Cole, 1993).

Humanism is a philosophy that has noted influence in HRD theory, as evidenced by Vince (2003) who suggests that “HRD policies and practices have traditionally been informed by a tired humanism . . .” (p. 559). This humanism, with its genesis in the Human Relations movement post-Industrial Revolution (Barnard, 1938; Follett, 1919; Mayo, 1945) directly impacts HRDs conceptualization of employee engagement and its attendant influence on organizational productivity (Cardus, 2013; Elliott & Turnbull, 2003; Shuck & Rocco, 2014). In its influence of HRDs conceptualization of employee engagement, humanism influences HRD theory. Humanism also influences HRD practice through its methods of executive coaching (Athanasopoulou & Dopson, 2015; Mulvie, 2015) as well as learning and instructional design methodologies (Hiemstra & Brockett, 1994; Knowles, Holton, & Swanson, 2015).

Finally, the philosophy of systems theory is pervasive in both HRD theory and practice. With roots in Aristotelian teleology (Bertalanffy, 1969; Falcon, 2015), systems theory rose to prominence as an influencing philosophy of HRD in conjunction with the emergence of Organization Development (Gosney & Hughes, 2015; Swanson, 2001b). Systems theory as a philosophy informs HRD theory most definitively in the arena of change management (Hayes, 2014) but also in theory-building around how knowledge is disseminated throughout an organization (Hughes & Gosney, 2016a; Senge, 1990). The influence of systems theory philosophy in the practical aspects of HRD are also apparent, such as the utilization of survey search (Likert, 1958) communities of practice (Chang & Jacobs, 2012) and action learning methodologies (Yeo & Marquardt, 2015). Kuchinke’s (2000) call for institutional theory (a systems theory coming out of sociology) to be utilized as a frame-work for HRD research

is another example of the influence of systems theory philosophy in HRD.

While Gosney and Hughes (2015) have explored the underlying philosophies of HRD, there is as of yet a lack of general awareness on the part of both theoreticians and practitioners of the influence of philosophy on the discipline. While HRD is not alone in such lack of awareness (Slife, 2000), and while others within the discipline have called for a more robust and thorough approach to philosophy within HRD (Ruona & Lynham, 2004), HRD has yet to fully embrace and explore the influence of philosophy on its attendant theory and practice. The reasons for this are likely numerous, but among them may be the influence of an additional philosophy that is predominant in the environment in which HRD is currently practiced: pragmatism. Gosney (2016) explores the influence of pragmatism on HRD today, particularly as a potential cause for the challenges in bridging the gap between scholars and practitioners in HRD.

As HRD matures as a discipline, and as new theories and methodologies emerge to meet the challenges of today's organizations, the influence of philosophies and their underlying assumptions will continue unabated. There is historical and current evidence to suggest that a number of philosophical influences, such as empiricism, humanism, systems theory, and pragmatism (Gosney, 2016; Gosney & Hughes, 2015), wield notable influence on both the theory and practice of HRD. As new theories and methodologies are proposed it benefits the discipline to acknowledge and consider these philosophical assumptions – particularly as they relate to assumptions of the individuals to whom the theories and methodologies will be applied.

People, Technology, and HRD

Much of the research regarding technology and HRD centers upon computer technology and how it relates to educating employees in the workplace. Performance management systems, performance support systems, Learning Management Systems

(LMS's), enterprise social networks (e.g., Yammer), social networks (e.g., Twitter, Facebook), professional networks (e.g. LinkedIn) and mobile devices that manage employee training and development, career development, and organization development are important tools for HRD professionals. Current trends suggest the need for organization's to be device agnostic, so employees can bring their devices to work and use the device of their choice in the workplace. HRD professionals and researchers must go beyond managing the systems to helping employees operate technological equipment. These types of technology systems that are displacing middle class workers (Zakaria, 2010) because they are not equipped or educated to operate high technology and cannot be trained quickly enough to meet the global needs of organizations. Instead, companies are bringing in foreign workers to meet immediate workplace needs in the US.

Due to the workplace realities of this millennium, HRD must recognize the need to support employees who are likely to have several short tenure positions within multiple organizations across their careers. HRD can embrace this trend and ensure that their employee develop transportable and marketable information technology (IT) skills. Examples of such skills include the ability to create a web page, troubleshoot software, and debug hardware. To aid the development of technological competencies, HRD will need to facilitate employees' adoption of a technology identity if HRD is to drive meaningful employee and organizational change. Possessing a technology identity will fuel individuals' desire to be lifelong technology learners and developers of technology. This will benefit the employees, their organizations and the broader society.

Githens, Dirani, Gitonga, and Teng (2008) conducted a meta-analysis of technology-related research in HRD publications from 2000 to 2006. Their findings revealed that only 10% of 1675 articles in five journals were related to technology; implying a lack of interest in technology by HRD researchers and professionals. This is discouraging

because Swanson (2008) noted the following assumptions regarding the role of human expertise and skill they bring to organizations:

1. Organizations are human-made entities that rely on human expertise in order to establish and achieve their goals.
2. Human expertise is developed and maintained through HRD processes for the mutual long-term and short-term benefits of sponsoring organizations and individuals involved.
3. HRD professionals are advocates of individual, team, work-processes, and organizational integrity. (p.331)

If HRD is to advocate the success of work processes, how can they do so without an understanding of technology's role in workplace success? How can HRD retool its workforce without first looking at their own technology skillset? It is difficult to model or advocate behavior that you do not personally use. People and technology are essential to the functioning of an organization. Hughes' (2010) People as Technology Conceptual Model is an integrative part of an organization's value creation strategy.

The PT conceptual model described the potential value that technology development and people development can bring to an organization. Operationalization of this model (Hughes, 2012) has the potential to be a promising management practice (Espedal, 2005) that will provide a solution when organizations struggle to implement their best practices without the best people in the right jobs at the right time and the best technology (Brache, 2002; Espedal, 2005; Martelli, 1998; Pfeffer, 1994; Stewart, 1997). People and technology must blend to the extent that the process is a win-win for employees and organization (London & Diamante, 2002). This concept is useful within all business strategies including JIT, Lean manufacturing, ISO, TQM, and Six Sigma because it enhances the foundation of the strategy, the people, and the process (Snell & Dean, 1992).

Much of the HRD technology related research "focused on educational technology in higher education settings, while non-profit organizations and government/military settings were under-represented. Overall, non-training topics were [also] underrepresented" (Githens, et al., 2008, p. 1). The opportunity for research in technology related research within HRD is extensive and needed. Re-training of the American workforce in high tech skills is an area where HRD researchers and professionals' skills are warranted. Training and development and career development areas of HRD in particular are being influenced by the technological advances that are being made in the global economy (Zakaria, 2010). How organizations adjust to these developments impacts the future of its employees and its existence.

FUTURE TRENDS

HRD is continuously evolving as organizations are continuously changing processes and procedures, developing employees, and trying to remain competitive in a global economy.

HRD professionals and researchers must also be willing to continuously change and evolve to help organizations achieve their goals. Some of the key strategies are:

1. HRD professionals and researchers must end the debate about which is more important between people and technology (Hughes, 2010, 2012; Kuchinke, 2000; Murphy & Garavan, 2009) and begin to learn from every field that touches human behavior (McLean et al., 2008). A focus on the history or organization theory and its evolution from sociology to business to HRD would enhance the development and understanding of HRD philosophy (Gosney & Hughes, 2015).
2. HRD professionals and researchers must find a way to bridge the gap between research and practice (Hughes & Gosney, 2016b). They must come to terms with the fact that

organizations are struggling to reach their goals without effective and efficient strategies from the HRD field.

3. HRD professionals and researchers must understand the relationship between people and technology (Hughes, 2014) and not focus on one or the other in isolation. They must develop methods of valuing both people and technology within the workplace (Hughes, 2012).
4. Technological innovations in the workplace continues an explosive upward trend. HRD researchers and professionals must understand these innovations to ensure that as people are developed their roles are not diminished or further eliminated but enhanced by technology.
5. HRD professionals and researchers must work to meet the business needs of workers in all organizations and provide technological based solutions along with their people based solutions. HRD professionals and researchers cannot plan people strategies without understanding technological implications (Betz, 1993; Hughes 2014).
6. Ultimately, HRD must consider public policy as a potential future area of research. Understanding the broader labor relations and economic implications of their decisions within the public policy arena can only strengthen HRD practice.

CONCLUSION

HRD professionals and researchers do not appear to deny that there is a relationship between people and technology; yet, there is very limited research that supports their effort to understand the relationship of people and technology within organizations. HRD professionals' and researchers' ability to expand their research and practice are limited by their beliefs concerning the juxtaposition of people to technology and/or their focus upon issues that are not adding to the

knowledge base in these specific topic areas. A mind shift needs to occur before progress can be made in this area of inquiry: people, technology and HRD. Understanding the history, philosophy and theories of HRD provides an opportunity for HRD researchers and professionals to build a solid foundation upon which they can enhance human behavior within organizations. They can begin to blend research and practice in a way that allows individuals and organizations to thrive and achieve competitive goals. It will also allow them to value the relationship between people and technology so that organizations can maximize the efforts of both people and technology and not one or the other (Carrig & Wright, 2006). Without these changes HRD professionals will never win a seat at the 'table' in corporate America or any other organization even though it is central to training and development, career development, and organization development. The future is optimistic for HRD researchers and professionals, and there are numerous ways that HRD can be developed to transform institutions to meet organizations' needs for global competitiveness.

REFERENCES

- Aguinis, H. (2009). *Performance management* (2nd ed.). Upper Saddle River, NJ: Pearson Prentice Hall.
- Aguinis, H., & Kraiger, K. (2009). Benefits of training and development for individuals and teams, organizations, and society. *Annual Review of Psychology*, 60(1), 451–474. doi:10.1146/annurev.psych.60.110707.163505 PMID:18976113
- Aguinis, H., & Pierce, C. A. (2008). Enhancing the relevance of organizational behavior by embracing performance management research. *Journal of Organizational Behavior*, 29(1), 139–145. doi:10.1002/job.493
- Athanasopoulou, A., & Dopson, S. (2015). *Developing leaders by executive coaching: Practice and evidence*. Oxford, UK: Oxford University Press.

- Barnard, C. I. (1938). *The functions of the executive*. Cambridge, MA: Harvard University Press; doi:10.1093/acprof:oso/9780199681952.001.0001
- Berger, N. O., Kehrhahn, M. T., & Summer-ville, M. (2004). Research to practice: Throwing a rope across the divide. *Human Resource Development International*, 7(3), 403–409. doi:10.1080/1367886042000246003
- Berr, S. A., Church, A. H., & Wacławski, J. (2000). The right relationship is everything: Linking personality preferences to managerial behaviors. *Human Resource Development Quarterly*, 11(2), 133–157. doi:10.1002/1532-1096(200022)11:2<133::AID-HRDQ4>3.0.CO;2-T
- Betz, F. (1993). *Strategic technology management*. New York, NY: McGraw-Hill.
- Brache, A. P. (2002). *How organizations work: Taking a holistic approach to enterprise health*. New York, NY: John Wiley & Sons.
- Callanan, J. L., & Connor, G. (2015). The competing interests of paradigm and praxis in critical HRD research: incorporating quantitative methods to enact critical practice. In M. N. K. Saunders & P. Tosey (Eds.), *Handbook of Research Methods in Human Resource Development*. Cheltenham, UK: Edward Elgar Publishing. doi:10.4337/9781781009246.00031
- Cardus, M. (2013). The five levers of employee engagement. *Journal for Quality and Participation*, 36(2), 28–31.
- Carrig, K., & Wright, P. M. (2006). *Building profit through building people: Making your workforce the strongest link in the value-profit chain*. Alexandria, VA: Society for Human Resource Management.
- Cascio, W. F., & Aguinis, H. (2005). *Applied psychology in human resource management* (6th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.
- Chang, J., & Jacobs, R. L. (2012). Determinants and outcomes of employee participation in a strategic community of practice: A mixed-method approach. *Human Resource Development Quarterly*, 23(3), 341–362. doi:10.1002/hrdq.21141
- Egan, T. M. (2002). Organization development: An examination of definitions and dependent variables. *Organization Development Journal*, 20(2), 59–70.
- Elliott, C., & Turnbull, S. (2003). Reconciling autonomy and community: The paradoxical role of HRD. *Human Resource Development International*, 6(4), 457–474. doi:10.1080/13678860210155395
- Espedal, B. (2005). Management development: Using internal or external resources in developing core competence. *Human Resource Development Review*, 4(2), 136–158. doi:10.1177/1534484305276217
- Falcon, A. (2015, Spring). Aristotle on causality. In E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*. Retrieved from: <http://plato.stanford.edu/archives/spr2015/entries/aristotle-causality/>
- Follett, M. P. (1919). Community is a process. *The Philosophical Review*, 28(6), 576–588. doi:10.2307/2178307
- Githens, R. P., Dirani, K., Gitonga, J. W., & Teng, Y. T. (2008). Technology-related research in HRD publications: An analysis of content and metaperspectives from 2000 to 2006. *Human Resource Development Quarterly*, 19(3), 191–215. doi:10.1002/hrdq.1236

- Gosney, M. W. (2014). *Theory and practice: A historical examination of the assumptions and philosophy of human resource development* (Unpublished doctoral dissertation). University of Arkansas.
- Gosney, M. W. (2016). The interplay between theory and practice in HRD: A philosophical examination. In C. Hughes & M. W. Gosney (Eds.), *Bridging the Scholar-Practitioner Gap in Human Resource Development*. Hershey, PA: IGI-Global. doi:10.4018/978-1-4666-9998-4.ch003
- Gosney, M. W., & Hughes, C. (2015). *The history of human resource development: Understanding the unexplored philosophies, theories, and methodologies*. New York, NY: Palgrave MacMillan.
- Harris, L. (2000). A theory of intellectual capital. In R. W. Herling & J. M. Provo (Eds.), *Strategic perspectives on knowledge, competence, and expertise* (pp. 22–37). San Francisco, CA: Berrett-Koehler.
- Hays, J. (2014). *The theory and practice of change management* (4th ed.). New York, NY: Palgrave MacMillan.
- Heffernan, N. (2000). *Capital, class, and technology in contemporary American culture: Projecting post-Fordism*. Sterling, VA: Pluto Press.
- Herling, R. W. (2000). Operational definitions of expertise and competence. *Advances in Developing Human Resources*, 6(4), 8–21. doi:10.1177/152342230000200103
- Hiemstra, R., & Brockett, R. G. (1994). From behaviorism to humanism: Incorporating self-direction in learning concepts into the instructional design process. In H. B. Long (Ed.), *New Ideas about Self-Directed Learning*. Norman, OK: University of Oklahoma Press.
- Hughes, C. (2010). People as technology conceptual model: Towards a new value creation paradigm for strategic human resource development. *Human Resource Development Review*, 9(1), 48–71. doi:10.1177/1534484309353561
- Hughes, C. (2012). *Valuing people and technology in the workplace: A competitive advantage framework*. Hershey, PA: IGI Global. doi:10.4018/978-1-4666-0240-3
- Hughes, C. (2014, Spring). Conceptualizing the five values of people and technology development: Implications for human resource management and development. *Workforce Education Forum*, 37(1), 23–44.
- Hughes, C., & Gosney, M. (Eds.). (2016a). *Bridging the scholar-practitioner gap in Human Resource Development*. Hershey, PA: IGI Global. doi:10.4018/978-1-4666-9998-4
- Hughes, C., & Gosney, M. W. (2016b). Human resource development as a knowledge management system: The importance of bridging the scholar-practitioner gap. In C. Hughes & M. W. Gosney (Eds.), *Bridging the Scholar-Practitioner Gap in Human Resource Development*. Hershey, PA: IGI-Global. doi:10.4018/978-1-4666-9998-4.ch001
- Jacobs, R. (1989). Systems theory applied to human resource development. In D. B. Gradous (Ed.), *Systems theory applied to human resource development* (pp. 27–60). Alexandria, VA: American Society for Training and Development.
- Jacobs, R. L. (1999). Partnership research: Ensuring more useful HRD collaborations. In K. P. Kuchinke (Ed.), *AHRD Conference Proceedings* (pp. 874–879). Bowling Green, OH: Academy of Human Resource Development.
- Jaspers, K. (1951). *Way to wisdom*. New Haven, CT: Yale University Press.
- Knowles, M., Holton, E. F. III, & Swanson, R. A. (2015). *The adult learner: The definitive classic in adult education and human resource development* (8th ed.). New York, NY: Routledge.
- Kuchinke, K. P. (2000). Debates over the nature of HRD: An institutional theory perspective. *Human Resource Development International*, 3(3), 279–283. doi:10.1080/13678860050128474

- Likert, R. (1958, March-April). Measuring organizational performance. *Harvard Business Review*, 41–50.
- Lohman, M. C. (2004). The development of a multirater instrument for assessing employee problem-solving skill. *Human Resource Development Quarterly*, 15(3), 303–321. doi:10.1002/hrdq.1105
- London, M., & Diamante, T. (2002). Technology focused expansive professionals: Developing continuous learning in the high-technology sector. *Human Resource Development Review*, 1(4), 500–524. doi:10.1177/1534484302238438
- Lynham, S. A. (2000). *The development and validation of a model of responsible leadership for performance*. St. Paul, MN: University of Minnesota Human Resource Development Research Center.
- Mankin, D. A. (2001). A model for human resource development. *Human Resource Development International*, 4(1), 65–85. doi:10.1080/13678860121714
- Martelli, J. (1998). Training for new technology: Midwest steel company. In W. Rothwell (Ed.), *Linking HRD programs with organizational strategy* (pp. 85–96). Alexandria, VA: ASTD.
- Mayo, E. (1945). *The social problems of an industrial civilization*. Boston, MA: Harvard University Press.
- Mazur, J. E. (1994). *Learning and behavior* (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.
- McLean, G. N., Lynham, S. A., Azevedo, R. E., Lawrence, J. E. S., & Nafukho, F. M. (2008). A response to Wang and Swansons article on National HRD and theory development. *Human Resource Development Review*, 7(2), 241–258. doi:10.1177/1534484308316486
- Messmann, G., & Mulder, R. H. (2012). Development of a measurement instrument for innovative work behavior as a dynamic and context-bound construct. *Human Resource Development International*, 15(1), 43–59. doi:10.1080/13678868.2011.646894
- Mulvie, A. (2015). *The value of executive coaching*. New York, NY: Routledge.
- Murphy, A., & Garavan, T. (2009). The adoption and diffusion of an NHRD standard: A conceptual framework. *Human Resource Development Review*, 8(1), 3–21. doi:10.1177/1534484308330019
- Nimon, K. F., & Astakhova, M. (2015). Improving the rigor of quantitative HRD research: Four recommendations in support of the general hierarchy of evidence. *Human Resource Development Quarterly*, 26(3), 231–247. doi:10.1002/hrdq.21219
- Paulsen, F. (1907). *Introduction to philosophy* (2nd ed.; F. Thilly, Trans.). New York, NY: Henry Holt & Co.
- Pfeffer, J. (1994). *Competitive advantage through people: Unleashing the power of the workforce*. Boston, MA: Harvard Business School Press.
- Reio, T. G., & Batista, L. C. (2014). Psychological foundations of HRD. In N. E. Chalofsky, T. S. Rocco, & M. L. Morris (Eds.), *Handbook of Human Resource Development*. Hoboken, NJ: John Wiley & Sons, Inc. doi:10.1002/9781118839881.ch1
- Ruona, W. E. A. (1999). *An investigation into core beliefs underlying the profession of human resource development* (Unpublished doctoral dissertation). University of Minnesota, St. Paul, MN.
- Ruona, W. E. A., & Lynham, S. A. (2004). A philosophical framework for thought and practice in human resource development. *Human Resource Development International*, 7(2), 151–164. doi:10.1080/13678860310001630665

- Ruona, W. E. A., & Roth, D. (Eds.). (2000). Philosophical foundations of human resource development practice. *Advances in Developing Human Resources*, 2(3), 7.
- Senge, P. M. (1990). *The fifth discipline*. New York, NY: Doubleday.
- Short, D. C., Sherlock, J. J., & Sugrue, B. (2004). Time to recognize those who link research and practice. *Human Resource Development Quarterly*, 15(3), 259–262. doi:10.1002/hrdq.1102
- Shuck, B., & Rocco, T. S. (2014). Human resource development and employee engagement. In C. Truss, K. Alfes, R. Delbridge, A. Shantz, & E. Soane (Eds.), *Employee Engagement in Theory and Practice*. New York, NY: Routledge.
- Slife, B. D. (2000). The practice of theoretical psychology. *Journal of Theoretical and Philosophical Psychology*, 20(2), 97–115. doi:10.1037/h0091300
- Snell, S. A., & Dean, J. W. (1992). Integrated manufacturing and human resource management: A human capital perspective. *Academy of Management Journal*, 35(3), 467–504. doi:10.2307/256484
- Stewart, T. A. (1999). *Intellectual capital: The new wealth of organizations*. New York, NY: Doubleday.
- Swanson, R. A. (1982). Industrial training. In W. H. Mitzel (Ed.), 5th encyclopedia of educational research (pp. 864–870). New York, NY: Macmillan.
- Swanson, R. A. (1999). The foundations of performance improvement and implications for practice. In R. Torraco (Ed.), *Performance improvement theory and practice: Advances in developing human resources* (pp. 1–25). Thousand Oaks, CA: Sage. doi:10.1177/152342239900100102
- Swanson, R. A. (2001a). The discipline of human resource development. In R. A. Swanson & E. F. Holton (Eds.), *Foundations of human resource development* (pp. 88–100). San Francisco, CA: Berrett-Koehler Publishers.
- Swanson, R. A. (2001b). Human resource development and its underlying theory. *Human Resource Development International*, 4(3), 299–312. doi:10.1080/13678860110059311
- Swanson, R. A., & Holton, E. F. (2001). *Foundation of human resource development*. San Francisco, CA: Berrett-Koehler Publishers.
- Swanson, R. A., & Torraco, R. J. (1994). Technical training's challenges and goals. *Technical & Skills Training*, 5(7), 18–22.
- Torraco, R. J. (2000). A theory of knowledge management. In R. W. Herling & J. M. Provo (Eds.), *Strategic perspectives on knowledge, competence, and expertise* (pp. 38–62). San Francisco, CA: Berrett-Koehler.
- Vince, R. (2003). The future practice of HRD. *Human Resource Development International*, 6(4), 559–563. doi:10.1080/13678860310001630656
- von Bertalanffy, L. (1969). *General systems theory* (revised ed.). New York, NY: George Braziller.
- Weinberger, L. A. (1998). Commonly held theories of human resource development. *Human Resource Development International*, 1(1), 75–93. doi:10.1080/13678869800000009
- Welch, J. (2005). *Winning*. New York, NY: Collins.
- Whitehead, A. N. (1926). *Science and the modern world*. Cambridge, UK: Cambridge University Press.
- Wilson, B., Jonassen, D., & Cole, P. (1993). Cognitive approaches to instructional design. In G. M. Piskurich (Ed.), *The ASTD handbook of instructional technology* (pp. 21.1–21.22). New York: McGraw-Hill.
- Yeo, R. K., & Marquardt, M. J. (2015). (Re) interpreting action, learning, and experience: Integrating action learning and experiential learning for HRD. *Human Resource Development Quarterly*, 26(1), 81–107. doi:10.1002/hrdq.21199

Yorks, L. (2005). Nothing so practical as a good theory. *Human Resource Development Review*, 4(2), 111–113. doi:10.1177/1534484305276176

Zakaria, F. (2010, November). Restoring the American dream. *Time*, 176(18), 30–35.

KEY TERMS AND DEFINITIONS

Axiology: The study of the nature, types, and governing criteria of values and value judgments.

Epistemology: The branch of philosophy that studies the nature of knowledge, in particular its foundations, scope, and validity.

Exogenous: Originating outside an organism or system.

Globalization: To become international or start operating at the international level, or cause something, especially a business or company, to become international.

Mutable: Tending or likely to change.

Ontology: The most general branch of metaphysics, concerned with the nature of being; the study of existence.

Operationalize: To put something to use or into operation.

Philosophy: A set of basic principles or concepts underlying a particular sphere of knowledge.

Performance Appraisal



Chandra Sekhar Patro

Gayatri Vidya Parishad College of Engineering (Autonomous), India

INTRODUCTION

In every organization, employees play a crucial part in determining its growth and survival. Employees are the eventual assets of any organization. The relation between the appraisal structure and the organization's approach is used as a strategic tool to accomplish the organization's vision. Performance appraisal is a controversial management tool searching for answers to ubiquitous tribulations in system design and administration (Thayer, 1987). This will evidently be a key factor in communicating values, promoting flexibility and maximizing individual potentials and contributions. Performance appraisal is a formal management system which evaluates the quality of an individual's performance in an organization. Performance appraisal is defined as a systematic description of individual job-relevant strengths and weaknesses, for the purposes of decision making regarding an individual's performance. In another term, performance appraisal is a process of evaluating the behaviour of the employees in the workplace, or it can also be referred as a process of giving feedback on employees' performance. The practice of performance appraisal tool has become the heart of the human resource management system in the organizations. It is a tool not only for accessing the performance of the employees but also the whole organizational performance.

The concept of performance appraisal was first time used during the First World War. At the instance of Water Drill Scott the U.S Army adopted the Man to Man rating system for evaluating military personnel. This concept came for industrial workers during 1920-30 under which efficient workers used to be identified and paid wages incentives and that scheme was popu-

larly known as "merit rating programs". In the early fifties, performance appraisal techniques began to be used for technical, professional and managerial personnel. Performance appraisal is a developmental tool which is used for all round development of an individual. It is the assessment of performance on an individual in a systematic way (Armstrong, 2006). It helps identify ways to improve one's job performance thus benefiting organization and the society as a whole. Rao (2005) defines that "performance appraisal is a method of evaluating the behaviour of employees in the work spot, normally including both the quantitative and qualitative aspects of job performance."

The term performance appraisal is sometimes referred as performance review, employee appraisal, performance evaluation, employee evaluation, employee rating, merit evaluation, or personnel rating. Performance appraisal is a system that involves a process of measuring, evaluating, and influencing employees' attributes, behaviour and performance in relation to a pre-set standard or objective. Usually, the appraisal of employee's performance is prepared by his immediate supervisor or manager. This process normally requires the supervisor to fill out a standardized assessment form that evaluates the individual on different dimensions and then discuss the results of the evaluation with the employee. In most of the organizations the performance appraisal systems stay one of the immense paradoxes of effectual human resource management. The appraisal system provide valuable performance information to a number of critical human resource activities, such as the allocation of rewards, e.g., merit pay, promotions; feedback on the development and assessment of training needs; other human resource systems evaluation, e.g., selection predictors; and

DOI: 10.4018/978-1-5225-2255-3.ch376

performance documentation for legal purposes (Cleveland, Murphy, & Williams, 1989). Efficient implementation of the performance appraisal process can help the organization in diverse ways. This process offers the benefit of identifying the employees' skills and serving the employees' development needs and career ambitions.

BACKGROUND

Performance appraisal is a complex tool of the human resource management process in any organization. Performance appraisal is not just about rating employees, it is a basis for administrative decisions such as promotion, allocation of financial rewards, employee development and identification of training needs (Meenakshi, 2012). Appraisal is preceded by establishing common objectives or a description for the job, identifying precise job expectations, providing feedback and, when necessary provides coaching (Hillman, Schwandt, & Bartz, 1990). The assessment of employees' performance is one of the most common practices in almost every organization, and so performance appraisal is an essential procedure for the better performance of employees and the organization itself (Karimi, Malik, & Hussain, 2011). Fletcher (2004) believes that the general aims of performance appraisal also include motivating staff, succession planning and identifying potential, promoting manager-subordinate dialogue and formal assessment of unsatisfactory performance. Many businesses regularly use performance appraisal scores to determine the distribution of pay, promotions, and other rewards; however, few organizations attempt to evaluate how employee perceptions of performance appraisal fairness impact employee attitudes and performance (Swiercz, Bryan, Eagle, Bizzotto, & Renn, 2012). The organizations should implement performance appraisal practice in the best possible way; there is the need to develop a good feedback system, appropriate and adequately filing, discussing appraisal results, design ways to

communicate appraisal results, review appraisal on due attention, participatory appraisal rating system and have to design procedure to make aware of every employee about the appeal process (Bekele, Shigutu, & Tensay, 2014).

According to Cawley, Keeping, & Levy (1998) subordinates reactions to performance appraisal can be a way of measuring their outlook towards the system. The main reactions that can be assessed are their satisfaction from the appraisal, the utility, whether they felt they were fairly appraised, how motivated they were from the appraisal and the accuracy of the system. Poon (2004) identified the usefulness of performance appraisal as a managerial decision tool depends partly on whether or not the performance appraisal system is able to provide accurate data on employee performance. Rees and Porter (2003) stated 360-degree appraisal as a process that involves the key people in a person's network of working relationships making assessments of an employee's performance and their subordinate being appraised is then given structured feedback; this may involve feedback from subordinates and any key outside parties, if it is practicable. DeCenzo and Robbins (2007) feel that appraisers should only rate in those areas in where they have substantial job knowledge. They should be as close as possible to the organizational level of the employee being evaluated. If the appraiser is not in a position where they can observe the persons work behaviour then there is a greater chance of inaccuracies.

According to Muo (2007) performance appraisal entails the systematic, organised and formalised process of evaluating individual employee's job related strengths and weaknesses with a view to providing feedback on which performance adjustment can be made. Thus, performance appraisal has both evaluative and developmental objectives. Nickols (2007) argues that the performance appraisal takes lot of time and energy, and can create frustration that can undermine the teamwork and climate of trust. Lack of communication, very variable appraisal standards and personal biases and values that replace organization standards,

etc., are all factors that influence the appraisal process. Gibson, Harvey, & Harris (2007) viewed that performance appraisal form a foundation for many human resources functions, effectively setting the standards to drive recruiting efforts, and it is customary to use these criteria in hiring, promoting, evaluating and equitably compensating employees, and forming the basis for many employee training programmes. Boachie-Mensah & Seidu (2012) advises that employees are likely to embrace and contribute meaningfully to the performance appraisal scheme, if they recognize it as an opportunity for personal development, a chance to be visible and demonstrate skills and abilities and an opportunity to network with others, but if the employees perceive this as an unreasonable effort by management to try to closer supervise and gain control over tasks they carry out, they won't welcome the scheme as easily. Mwema & Gachunga (2014) argued that organizations should appraise their employees often through utilized targets, accomplishments, organization goals, time management and efficiency for performance measure purposes as it would lead to increased in employee's productivity.

OBJECTIVES OF THE STUDY

1. To examine the usefulness of various performance appraisal tools adopted in an organization.
2. To find out the employees' satisfaction level on performance appraisal process.
3. To explore the gaps and the recent trends in implementation of performance appraisal system in an organization.

CLASSIFICATION OF PERFORMANCE APPRAISAL TOOLS

Performance appraisals generally review the past behaviour and so provide an opportunity to reflect on past performance. But to be successful

they should also be used as a basis for making development and improvement plans and reaching agreement about what should be done in the future. Employee performance appraisal has two forms i.e., formal (systematic) and informal (non-systematic) appraisal. Informal appraisal means continuous evaluation of an employee by her/his superior during the work process (Dědina & Cejthamr, 2005). Formal employee appraisal is a formal organizational process conducted on a systematic basis in order to enable a comparison between the expected individual (group) and real performance (Giangreco *et al.*, 2012). Appraisal takes place annually between the manager and the employee. However there are number of trends that are changing the style and relationship of the appraisal.

According to the authors (Deb, 2006; Khurana *et al.*, 2010; Randhawa, 2007), traditional appraisal methods focus on performance i.e., paired or group comparisons, ranking, rating scales and reports, essay method, checklists, a critical/key incident method, forced choice and distribution methods, field reviews. On the other hand, modern methods are partly similar to methods focusing on the future that estimate an employee's growth potential. These methods are Management by Objectives, Assessment Centre, Behaviorally Anchored Rating Scales (BARS), Psychological Appraisals, 360 Degree feedback, and human resources accounting.

Some authors have proposed methods of performance appraisal that include the combination of both traditional and modern methods of appraising the performance of an employee. Mathis and Jackson (2011) classified checklists, graphic rating scales and Behaviorally Anchored Rating Scales (BARS) as scaling methods of performance appraisal and stated that the Narrative methods are used in cases requiring written or oral appraisal which includes free essay, reports and critical incident methods. In BARS, the ratter will act as observer not the judge. It helps the ratter to focus on specific desirable and undesirable incidents of work behaviour which can serve as examples

in discussing a rating. BARS use behavioural statements or concrete examples to illustrate multiple levels of performance for each element of performance.

Griffin (2012) stated that the appraisal methods such as Behaviorally Anchored Rating Scales (BARS), Rating Scales, Ranking and Comparison methods are targeted to evaluate and assess the current outputs of the employees under standardized conditions, and these are classified as Judgmental methods (Pride, Hughes, & Kapoor, 2012). Bogardus (2007) classified that appraisal techniques as comparative, rating, narrative, and behavioral methods. Comparative methods compare work performance of individuals - the most common type is the paired comparison, ranking, forced distribution method and forced choice method. The most common appraisal methods are checklists and rating scales. Finally, the group of behavioral methods covers the Behaviorally Anchored Rating Scales (BARS) method.

METHODOLOGY

The present study is an insight on the impact of practicing the performance appraisal system and the satisfaction experienced by the employees working in these organizations. For the survey, the employees (teaching and non-teaching staff) working in different private engineering colleges and located in Visakhapatnam city of Andhra Pradesh, India are considered. The staff members having a minimum five years of experience were considered for the study. The employees list was obtained from the official websites of engineering colleges and a simple random sampling method was employed for the study.

The data presented in the study are collected from both the primary and secondary sources. For the literature survey, secondary source of data is used from various published sources such as, books, journals, websites, and studies done previously by the researchers. The primary data

was collected from the respondents by distributing questionnaires personally through emails. The purpose of the study was explained to the employees and was asked to rate the scales. They were also assured of confidentiality of their responses. The questionnaire was sent or personally distributed to 200 employees and 120 valid and fair responses were measured for the study.

This study therefore seeks to investigate the employees' satisfaction towards the existing performance appraisal tools and fill this knowledge gap by determining the impact of performance appraisal on employee productivity in the organizations.

DATA ANALYSIS AND DISCUSSIONS

In the study, a descriptive analysis is performed to know the perceptions of the respondents with regard to the performance appraisal process. The responses for each specific statement are compared using the mean and standard deviation score. The degree of agreement or disagreement of the respondent for each statement is analyzed by using a five point scale ranging from Strongly Agree (5) to Strongly Disagree (1). To measure the level of employee satisfaction on the implementation of performance appraisal in their organizations, the mean and standard deviation of the indicators were analyzed with the help of descriptive statistics of SPSS software. The higher the mean score, the more that respondent agreed with the indicators and vice-versa. The standard deviation scores also indicate the degree to which responses varied from each other; the higher the SD score, the more variation in the responses. The results and findings with regard to the practice of performance appraisal process in the organizations are discussed in the paper.

The views of the respondents on the eleven indicators of performance appraisal process with mean and standard deviation scores are shown in Table.1.

Table 1. Employees' satisfaction level on performance appraisal process

Sl. No.	Indicators	Mean	Standard Deviation
1	Satisfied with existing appraisal process used to evaluate the performance	3.32	1.15
2	Performance appraisal process help me to know about my current performance level	3.51	1.20
3	The organization conducts performance appraisal in the best possible way	3.25	1.12
4	Encourage continuous feedback between the appraisee and the supervisor	3.05	1.19
5	Improve quality of work through better planning and fair participation in appraisal	3.28	1.15
6	Links the individual performance with organizational performance and goals	3.30	1.18
7	Current Performance appraisal process is fair and unbiased	3.28	1.15
8	If not satisfied with performance appraisal score, there is appeal process	3.10	1.10
9	The organization recognizes the good performers rather than criticizing the poor performers	2.95	1.22
10	Provide information on HR issues such as promotions, delegation, training, deployment, and rewards	3.20	1.19
11	Performance appraisal is valuable to me as well as to my organization	4.10	0.96
Total		3.29	1.14

The analysis on the level of respondents satisfaction on performance appraisal process shows that one indicator is high i.e., Performance Appraisal is valuable to me as well as to my organization with a mean score of 4.10 and SD of 0.96. The satisfaction level of respondents is moderate with mean score of 3.51 and SD of 1.20 for the indicator, performance appraisal process helps me to know about my current performance level. However, the respondent's level of satisfaction is low for the other eight indicators as the mean score ranged from a maximum of 3.32 to a minimum of 2.95.

As a result from this finding it can be generalized that, the respondents are not satisfied concerning with the existing appraisal process used to evaluate the performance (3.32, 1.15); organization conducts performance appraisal in the best possible way (3.25, 1.12); encourage continuous feedback between the appraisee and the supervisor (3.05, 1.19); improve quality of work through better planning and fair participation in appraisal (3.28, 1.15); links the individual performance with organizational performance

and goals (3.30, 1.18); current Performance appraisal process is fair and unbiased (3.28, 1.15); if not satisfied with performance appraisal score there is appeal process (3.10, 1.10); the organization recognizes the good performers rather than criticizing the poor performers (2.95, 1.22); and provide information on HR issues such as promotions, delegation, training, deployment, and rewards (3.20; 1.19).

The overall response for the eleven indicators specifies the mean score as 3.29 and standard deviation (SD) score as 1.14. Therefore, the mean score indicates that there is low level of satisfaction towards the current performance appraisal process, which implies that the respondents are dissatisfied. Thus, from the statistical analysis shown in table.1 it is observed that the intention of the employees towards the performance appraisal process practiced in their organizations is moderate and dis-satisfied. Therefore, such remarked areas of dis-satisfaction are the indicators and the root causes for employees' negative satisfaction regarding to the performance appraisal process of their organization.

GAPS IN IMPLEMENTATION OF PERFORMANCE APPRAISAL

The appraisal process should directly reflect an employee's performance and not on any biases of a supervisor. Generally, this is unworkable to do perfectly as most raters either intentionally or unintentionally commit errors. Raters need to be aware of these biases, so that their effect on the appraisals can be limited or eliminated. The important issues and discrepancies concerning performance appraisal in the present circumstances are:

1. Most of the indicators used for measuring the performance of the employees are not quantifiable in nature, making it difficult to measure the performance.
2. Appropriate information is not provided to the employees on the performance appraisal process followed in the organization and the appraisals are not conducted on regular basis.
3. The performance appraisal processes are faced with a lot of obstacles, the most prominent being lack of quantifiable indicators of the performance.
4. In some of the organizations the objectives are unchallenging, unrealistic and not timely reviewed and updated.
5. Due to the lack of accountability and job security, most of the employees have a *laissez faire* attitude towards their work.
6. There is often a lot of bias and subjectivity involved in the ratings given by the superiors due to unprofessional and unstructured approach towards the appraisal process.
7. Performance measuring, rating and review systems have become more detailed, structured and individual specific than before.
8. The focus of the performance appraisals is turning towards career development relying on the dialogues and discussions with the superiors.

RECENT TRENDS IN PERFORMANCE APPRAISAL

The problems in the implementation of the performance appraisal processes are being anticipated and efforts are being made to overcome them. Most of the organizations are moving towards 360-degree feedback system due to the various challenges faced by the management. The emergence of following concepts and the following trends related to performance appraisal can be seen in the global scenario.

1. **360 Degree Appraisal:** The 360 degree appraisal process is also known as 'multi-rater feedback' and is the most comprehensive appraisal where the feedback about the employees' performance comes from all the sources that come in contact with the employee on his job. Organizations are increasingly using feedback from various sources such as peer input, customer feedback, and input from superiors. Different forms with different formats are being used to obtain the information regarding the employee performance.
2. **Team Performance Appraisal:** Most of the performance appraisal techniques are formulated with individuals in mind i.e. to measure and rate the performance of the individual employee. Therefore, with the number of teams increasing in the organizations, it becomes difficult to measure and appraise the performance of the team. The answer to this problem that is being adopted by many companies is to measure both the individual and the team performance. Sometimes, team based objectives are also included in the individual performance plans.
3. **Rank and Yank Strategy:** This strategy of appraisal is can also be referred as 'Up or out policy'. It refers to the performance appraisal model in which best-to-worst ranking

methods are used to identify and separate the poor performers from the good performers. Then the action plans and the improvement opportunities of the poor performers are discussed and they are given to improve their performance in a given time period, after which the appropriate HR decisions are taken. Some of the organizations following this strategy are Ford, Microsoft and Sun Microsystems.

4. **Information and Communication Technology:** The adoption and utilization of Information and Communication Technology (ICT) in performance appraisal is gradually becoming a major requirement for improvement in the efficiency and effectiveness of employee performance and productivity. The emergence of technology has brought great changes in evaluating the performance of the employees. The innovations in information processing, telecommunications and related technologies, known collectively as information and communication technology (ICT) is having profound impact public as well as organizations by easing enquiry, saving time, and improving service delivery (Yasuharu, 2003). The rapid advances in technology drastically changed the traditional ways in which information was processed, communications conducted, and services made available (Sarfo, 2007). The technology based appraisal will help the organizations in perceived usefulness, saving time and other resources. This method is practiced in many large organizations where the usage of technology has become more essential.

Instead of measuring employees' performance and pointing out where they fall short, the organizations will achieve more results by finding ways to fine-tune and improve their systems. These days the managers just need to document that they have in fact had regular conversations with their employees and if there are any problems the

managers are expected to make note of it. This creates the paper trail that will support any eventual disciplinary action or termination.

FUTURE RESEARCH DIRECTIONS

In the present conditions, organizations need to adopt new performance appraisal tools that allow greater involvement of both the raters and the ratees. Just like other management practices the employee appraisal process should also be dynamic. There is continuous need for reviewing and updating the appraisal process to match along with the organizational changes and current management practices. Performance appraisal as a controversial tool of human resource tool also needs the whole system to be transparent and have control in each step.

This study highlights employees' satisfaction on performance appraisal and the gaps in implementation of appraisal system. But the research on how to establish control, appeal and transparent system is limited in this study. Therefore, for further studies, it is important to explore how the organizations can establish the formal appeal and control system, and keep transparency in performance appraisal. This study can be further enhanced to know the execution of technology based appraisal process and improve the system specifically, and make it work professionally.

CONCLUSION

A quality performance appraisal process must be put in practice by an organization to exert the employee's maximum efforts towards realization of organizational objectives and goals. In situations where employees are not aware of what they are expected to perform and the consequences that their performance would bring to them, it is difficult to anticipate better work performance and commitment towards the organization. The objective of this study was to assess the impact

of performance appraisal on employees' performance, as performance appraisal has been a controversial tool with its long lasting impacts on the employees' efficiency and overall organizational productivity. Thus, the findings of the study show serious managerial implications in order to create a good perception in the minds of the employees. In doing so, the organization may have to implement performance appraisal practice in the best possible way so as to change the perception of employees by realizing their needs. The outcomes of the study indicate that the negative perception of employees towards performance appraisal process has its own impact on the employees' work performance and overall organizational productivity.

The organizations should have to give due attention to continually improve and maintain employee affective organizational commitment and should make conscious efforts by creating different mechanisms like creating trust on the mind of employee about performance appraisal, establishing organizational plans for the career and development of employee, and providing opportunity to employee to suggest improvement in the ways things are done. Therefore, the management should introduce better incentive, opportunity for internal growth and development, smooth and transparent work communication system.

REFERENCES

- Armstrong, M. (2006). *Performance Management: Key Strategies and Guidelines* (3rd ed.). London, UK: Kogan Page Ltd.
- Bekele, A. Z. M., Shigutu, A. D., & Tensay, A. T. (2014). The Effect of Employees' Perception of Performance Appraisal on Their Work Outcomes. *International Journal of Management and Commerce Innovations*, 2(1), 136–173.
- Boachie-Mensah, F., & Seidu, P. A. (2012). Employees perception of performance appraisal system: A case study. *International Journal of Business and Management*, 7(2), 73–88. doi:10.5539/ijbm.v7n2p73
- Bogardus, M. A. (2007). *PHR/SPHR: Professional in Human Resources Certification*. Indianapolis, IN: Wiley Publishing.
- Cawley, B. D., Keeping, L. M., & Levy, P. E. (1998). Participation in the performance appraisal process and employee reactions: A meta-analytic review of field investigations. *The Journal of Applied Psychology*, 83(4), 615–633. doi:10.1037/0021-9010.83.4.615
- Cleveland, J. N., Murphy, K. R., & Williams, R. E. (1989). Multiple uses of performance appraisal: Prevalence and correlates. *The Journal of Applied Psychology*, 74(1), 130–135. doi:10.1037/0021-9010.74.1.130
- Deb, T. (2006). *Strategic Approach to Human Resource Management: Concept, Tools & Application*. New Delhi: Atlantic Publishers and Distributors.
- Decenzo, D. A., & Robbins, S. P. (2002). *Human Resource Management* (7th ed.). New York: John Wiley & Sons Inc.
- Dědina, J., & Cejthamr, V. (2005). *Management a organizační chování*. Praha: Grada Publishing.
- Fletcher, C. (2004). *Appraisal and Feedback, Making Performance Review Work*. London, UK: Chartered Institute of Personnel and Development.
- Giangreco, A., Carugati, A., Sebastiano, A., & Al-Tamimi, H. (2012). War outside, ceasefire inside: An analysis of the performance appraisal system of a public hospital in a zone of conflict. *Evaluation and Program Planning*, 35(1), 161–170. doi:10.1016/j.evalprogplan.2010.11.004 PMID:21168215

- Griffin, W. R. (2012). *Fundamentals of Management*. South Western Educational Publishing.
- Hillman, L., Schwandt, D., & Bartz, D. (1990). Enhancing Staff Members Performance through Feedback and Coaching. *Journal of Management Development*, 9(3), 20–27. doi:10.1108/02621719010135110
- Karimi, R., Malik, M. I., & Hussain, S. (2011). Examining the relationship of performance appraisal system and employee satisfaction. *International Journal of Business and Social Science*, 2(22), 243–247.
- Khurana, A., Khurana, P., & Sharma, H. L. (2010). *Human Resource Management*. Delhi, India: Star Offset.
- Mathis, L. R., & Jackson, H. J. (2012). *Human Resource Management: Essential Perspectives*. South-Western Cengage Learning.
- Meenakshi, G. (2012). Multi source feedback based performance appraisal system using Fuzzy logic decision support system. *International Journal on Soft Computing*, 3(1), 91–106. doi:10.5121/ijsc.2012.3108
- Murphy, K. R., & Cleveland, J. N. (1995). *Understanding performance appraisal: Social, organizational, and goal-based perspectives*. Thousand Oaks, CA: Sage.
- Mwema, N. W., & Gachunga, H. G. (2014). The influence of performance appraisal on employee productivity in organizations: A case study of selected WHO offices in East Africa. *International Journal of Social Sciences and Entrepreneurship*, 1(11), 324–337.
- Nickols, F. (2007). Performance appraisal: Weighed and found wanting in the balance. *Journal for Quality and Participation*, 30(1), 13–16.
- Patro, C. S. (2013). The Role of Human Resource Management in Implementation of TQM. *International Journal of Computer Science and Management Research*, 2(6), 2689–2695.
- Pride, M. W., Hughes, R. J., & Kapoor, J. R. (2012). *Business*. Mason, OH: South-Western Publishing.
- Randhawa, G. (2007). *Human Resource Management*. New Delhi, India: Atlantic Publishers and Distributors.
- Rao, V. S. P. (2005). *Human Resource Management: Text and Cases* (2nd ed.). New Delhi, India: Excel Books.
- Rees, W. D., & Porter, C. (2003). Appraisal pitfalls and the training implications-part 1. *Industrial and Commercial Training*, 35(7), 280–284. doi:10.1108/00197850310501677
- Sarfo, F. K. (2007). *Educational Technology* (2nd ed.). Kumasi: UEW Press.
- Swiercz, P. M., Bryan, N. B., Eagle, B. W., Bizzotto, V., & Renn, R. W. (2012). Predicting employee attitudes and performance from perceptions of performance appraisal fairness. *Business Renaissance Quarterly*, 7(1), 25–46.
- Thayer, F. (1987). Performance Appraisal and Merit pay Systems: The Disasters Multiply. *Review of Public Personnel Administration*, 7(2), 36–53. doi:10.1177/0734371X8700700204
- Yasuharu, U. (2003). *Effects of Information Systems Investment in Banking Industry*. Boston: MacGraw-Hill Irwin.

ADDITIONAL READING

- Armstrong, M. (2009). *Armstrong's Handbook of Performance Management: An Evidence-based Guide to Delivering High Performance*. UK: Kogan Page.
- BasuMihir, K. (1978). *Performance Appraisal in India*. New Delhi, India: Vision Books.
- George, A. L., & Bennett, A. (2005). *Case Studies and Theory Development*. Cambridge, MA: MIT Press.

Goel, D. (2010). *Performance Appraisal and Compensation Management: A Modern Approach*. New Delhi, India: PHI Learning Private Ltd.

Grubb, T. (2007). Performance Appraisal Reappraised: It's Not All Positive. *Journal of Human Resources Education*, 1(1), 1–22.

Haldar, U. K., & Sarkar, J. (2012). *Human Resource Management*. New Delhi, India: Oxford University Press.

Manoharan, T. R., Muralidharan, C., & Deshmukh, S. G. (2009). Employee Performance Appraisal Using Data Envelopment Analysis: A Case Study. *Research and Practice in Human Resource Management*, 17(1), 92–111.

Roberts, G. E. (1998). Perspectives on enduring and emerging issues in performance appraisal. *Public Personnel Management*, 27(3), 301–320. doi:10.1177/009102609802700302

Sanjeevni, G. (2012). Employee Survey on Performance Appraisal System. *International Journal of Social Sciences and Interdisciplinary Research*, 1(6), 124–141.

Tziner, A., & Murphy, K. R. (1999). Additional evidence of attitudinal influences in performance appraisal. *Journal of Business and Psychology*, 13(3), 407–419. doi:10.1023/A:1022982501606

KEY TERMS AND DEFINITIONS

360 Degree Feedback: The information regarding the performance of an employee is derived from a number of stakeholders like immediate supervisors, team members, customers, peers and self.

Appraise: An individual being evaluated on work performance.

Appraiser: An individual responsible for evaluating an individual's work performance.

Behaviorally Anchored Rating Scales (BARS): Behaviorally Anchored Rating Scales is a formatted performance appraisal method based on making rates on behaviors or sets of indicators to determine the effectiveness or ineffectiveness of employees work performance.

Management by Objectives (MBO): In this method, the managers or employers set a list of objectives and make assessments on their performance on a regular basis, and finally make rewards based on the results achieved.

Performance Appraisal (PA): Performance Appraisal is a process of evaluating the employee's performance in terms of job requirements and setting job standards in an organization.

Psychological Appraisal: This method assesses the employee's potential for future performance rather than the past one. It focuses on the employee's emotional, intellectual, and motivational and other personal characteristics affecting his/her performance.

Work Performance: It is an accomplishment of the assigned tasks for achieving organization's goal.

Promoting Strategic Human Resource Management, Organizational Learning, and Knowledge Management in Modern Organizations



Kijpokin Kasemsap

Suan Sunandha Rajabhat University, Thailand

INTRODUCTION

The added value of human resource management (HRM) is strongest when HRM decisions are linked to organizational strategy (Schalk, Timmerman, & van den Heuvel, 2013). Strategic human resource management (SHRM) practices are developed, implemented, and executed based on a deliberate linkage to organizational strategy (Huselid, Jackson, & Schuler, 1997). SHRM predicts electronic human resource management (e-HRM) outcomes and the relationship appears context-dependent in modern organizations (Marler & Fisher, 2013).

Learning has an important position in the development of employees and their expertise (Melo & Beck, 2015). Organizational learning theory indicates that knowledge must be integrated throughout organization to facilitate strategic renewal (Lionzo & Rossignoli, 2013). Organizational learning incorporates the organizational system, dynamic mechanism, and effective HR activities toward improving the competency and capability of organizational employees in the modern workforce (Sabir & Kalyar, 2013). Organizational learning can significantly improve the family firms' ability to counter by stimulating entrepreneurship (Zahra, 2012). Knowledge management (KM), strategic orientation, and organizational innovation are correlated with organizational performance (Kasemsap, 2014a).

This article aims to bridge the gap in the literature on the thorough literature consolidation

of SHRM, organizational learning, and KM. The extensive literature of SHRM, organizational learning, and KM provides a contribution to practitioners and researchers by describing the theory and applications of SHRM, organizational learning, and KM in order to maximize the value of SHRM, organizational learning, and KM in modern organizations.

BACKGROUND

Strategic human resource management (SHRM) literature emerged about 30 years ago (Lengnick-Hall, Lengnick-Hall, Andrade, & Drake, 2009). The strategic evolution perspective explains how the expected outcomes of SHRM may vary depending on what decisions are made and what paths are taken in implementation (Lengnick-Hall et al., 2009). Most organizational strategy research offers some rationale to account for performance differences across organizations or to account for strategic differences that presumably have an impact on organizational performance (Barnett & Burgelman, 1996). SHRM is defined as the pattern of planned HR deployments and activities intended to enable an organization to achieve its goals (Wright & McMahan, 1992).

Organizational learning refers to the process of developing new knowledge and insights derived from the common experiences of people within the organization (Slater & Narver, 1995). Organizational learning is described as an important process that expands the organization's ability

DOI: 10.4018/978-1-5225-2255-3.ch377

to accomplish effective actions by improving its performance and results (Chiva & Alegre, 2009). Organizational learning includes the acquisition, dissemination, and use of knowledge (Argote, McEvily, & Reagans, 2003). Organizational learning is an extremely useful process for generating new ideas (Fernandez-Mesa & Alegre, 2015). Knowledge-sharing behavior positively mediates the relationships between organizational learning and organizational innovation and between KM and organizational innovation in the learning organizations (Kasemsap, 2014b).

Organizational policy perspectives can promote both SHRM and e-HRM toward increasing organizational learning, KM, and competitive advantage by encouraging the effective coordination and partnerships among organizations and combining employees' learning and development opportunities available to employees through modern learning technologies. Executives and HR managers should organize particular emphasis on employees' learning and development by implementing SHRM and e-HRM. The establishment of SHRM-related learning and development is the necessary component of organizational learning and KM strategies in modern organizations.

FACETS OF STRATEGIC HUMAN RESOURCE MANAGEMENT, ORGANIZATIONAL LEARNING, AND KNOWLEDGE MANAGEMENT

This section emphasizes the concept of SHRM; the concept of e-HRM; the importance of SHRM in modern organizations; and the current trends of organizational learning and KM in modern organizations.

Concept of Strategic Human Resource Management

With respect to the integration of HR issues in strategic decision making, the different levels of integration can be considered (Golden & Ra-

manujam, 1985). HRM is present at the highest management level and is crucial to organizational strategy (Bennett, Ketchen, & Schultz, 1998). Connecting HRM and strategic decision making improves organizational performance (Huselid, 1995). HRM has an added value in terms of influencing organizational performance (Liu, Combs, Ketchen, & Ireland, 2007). Stakeholder interests and contextual factors exert an influence on HRM policy choices (Guest, 1987).

In the SHRM literature, scholars focus on strategic outcomes, such as organizational performance (Becker & Huselid, 1998), strategic alignment (Schuler & Jackson, 1987), and competitive advantage (Wright, Dunford, & Snell, 2001). SHRM is defined as the significant activities affecting the behavior of individuals in their efforts to meet the strategic needs of the business (Schuler, 1992). SHRM implies that HR objectives have to be chosen in accordance with the general strategy of organization (Schalk et al., 2013). SHRM not only requires vertical alignment with an organization's business strategy, but each business activity also needs to be aligned with other HR activities (Lepak, Liao, Chung, & Harden, 2006).

This section is dealing with the concept of SHRM and the next section is dealing with the concept of e-HRM. SHRM is the strategic practice of attracting, developing, rewarding, and retaining employees with the important goal of increasing various benefits to both employees as individuals and organization as a whole.

Concept of Electronic Human Resource Management

The studies of electronic human resource management (e-HRM) began around 1995 (Strohmeier, 2007). Ruël et al. (2004) defined e-HRM as a way of implementing HRM strategies, policies, and practices in organizations through the conscious and directed support of and with the full utilization of the Internet. Strohmeier (2007) defined e-HRM as the application of information technology (IT) for both networking and supporting the interaction

of at least two individual or collective actors in their shared performing of HR activities. HR staff with substantial e-HRM implementations need the increased competency in the knowledge of business, functional HR delivery, and technology expertise (Bell, Lee, & Yeung, 2006).

Developing value from e-HRM should be possible if the organization utilizes e-HRM to acquire and develop superior human capital and to produce organizational knowledge, social networks that promote information sharing, and organizational processes that produce innovative responses in competitive markets (Marler, 2009). Large companies increasingly are turning to data mining to help identify and retain their qualified talents regarding talent management (Bondarouk, 2014). Data mining is the computational procedure of pioneering schemes in the large data sets regarding methods at the integration of artificial intelligence, machine learning, statistics, and database systems (Kasemsap, 2015a). Talent management introduces new strategic goals to streamline hiring and leadership succession processes (Kasemsap, 2016a).

This section is dealing with the concept of e-HRM and the next section is dealing with the importance of SHRM in modern organizations. With the advent of the Web 2.0, e-HRM can meet the challenge of becoming strategic, flexible, customer-oriented, and cost-effective by applying IT. Regarding strategic management, e-HRM can lower administrative costs, increase productivity, expedite response times, improve decision making, and enhance customer service toward improving all aspects of HRM in modern organizations.

Importance of Strategic Human Resource Management in Modern Organizations

Lengnick-Hall et al. (2009) indicated that SHRM involves three major stages (i.e., industrial organizational (I/O) economic perspective, resource-based view (RBV) perspective, and strategic evolution perspective). The I/O eco-

nomics perspective focuses on fit between HR practices and business strategy. Industry characteristics define the performance potential for companies based on the industry in which they operate and then within an industry, a firm's strategic positioning determines the extent to which it successfully competes to achieve competitive advantage.

The RBV perspective focuses on what and why HR resources and bundles lead to positive organizational outcomes (Lengnick-Hall et al., 2009). HRM can be strategic either as a specific capability or as instrumental in developing organizational resources (e.g., human capital and intellectual capital) that are rare and imperfectly imitable (Wright et al., 2001). Regarding strategic evolution perspective, the divergence between intended and implemented SHRM practices suggest the possibility that the outcomes of SHRM practice may vary depending on what decisions are made and what paths are taken during implementation (Marler & Fisher, 2013). Evolutionary strategic theorists view strategic responses to specific problems or opportunities that occur in the environment to be rationally bounded, which departs from the highly rationalized behavior (Gavetti & Rivkin, 2007).

This section is dealing with the importance of SHRM in modern organizations and the next section is dealing with the current trends of organizational learning and KM in modern organizations. SHRM is an essential business partner in organizational success, as opposed to a necessity for legal compliance or compensation. SHRM strategically utilizes organizational resources and talent within HR functions to make organizations more effective in the modern workforce. Organizational learning and KM allow for organizational employees to share knowledge and learn exactly what is relevant to their specific tasks toward encouraging human capital and knowledge creation. Organizational learning and KM also increase information sharing, communication, understanding, and the quality of decisions strategically made in modern organizations.

Current Trends of Organizational Learning and Knowledge Management in Modern Organizations

Organizational learning is conducted through either the exploratory or exploitative approach (Kim & Atuahene-Gima, 2010). Exploratory learning refers to the acquisition and learning of information and knowledge from outside current customer and competitor boundaries, and often involves experimentation and risk-prone projects (Kim & Atuahene-Gima, 2010). Ugurluoglu et al. (2013) stated that organizational learning is the procedure of increasing effective organizational movements through knowledge and understanding. Organizational learning is necessary in rapidly changing environments, through seeking new knowledge and the practical use of existing knowledge (Dayaram & Fung, 2014). Leaders of global businesses should organize information and communication technology (ICT) skills for organizational employees to enlarge their knowledge in global business (Kasemsap, 2014c).

Organizational learning positively affects export intensity (Bengtsson, 2004). A number of previous studies have viewed exporting as a process of knowledge accumulation during which company identifies and exploits opportunities abroad (Brouthers, Nakos, Hadjimarcou, & Brouthers, 2009). Knowledge renewal and exploitative learning regarding foreign markets increase exports (Balabanis, Theodosiou, & Katsikea, 2004) because firms that learn from their experience are able to export faster and with fewer mistakes. Firms effectively change by learning through the commitment decisions that they make to strengthen their position in the foreign market (Johanson & Vahlne, 2009). Organizational learning partially mediates the relationship between entrepreneurial orientation and performance and fully mediates the link between learning orientation and performance (Real, Roldán, & Leal, 2014).

The global economies thrive on cutting-edge knowledge, which drives research and organizational innovation in competitive environments (Kasemsap, 2016b). Organizational learning, HRM, and KM capability are related to improved organizational performance in modern organizations (Kasemsap, 2015b). Economic development is enhanced through the knowledge-based organizations to create new market and new industry platforms (Kasemsap, 2016c). KM is a functional set of behaviors, processes, and technologies that are designed for managing information more efficiently to improve learning, decision making, innovation, and other keys to business success (Kasemsap, 2016d).

Lifelong learning and KM have become a fundamental goal of education policies, both at a national and international level (Kasemsap, 2016e). Effective KM is the key driver of new knowledge and new ideas to the innovation process, innovative products, services, and solutions (Kasemsap, 2016f). IT and KM applications effectively improve the strategic tools for providing the direct link between customers and tourism organizations, thus encouraging the communication channels in global tourism (Kasemsap, 2016g). Organizational culture, organizational climate, and KM are positively related to job performance (Kasemsap, 2016h).

FUTURE RESEARCH DIRECTIONS

The classification of the extensive literature in the domains of SHRM, organizational learning, and KM will provide the potential opportunities for future research. In the knowledge economy, characterized by information, speed, and innovation, effective HRM is key to successful strategy execution. Human capital management (HCM) is an approach to employee staffing that perceives people as assets (human capital) whose current value can be measured and whose future value can be enhanced through organizational invest-

ment in the modern workforce. Communities of practice (CoPs) help promote a growing cycle of knowledge sharing activities that allow for the members to regularly meet, reflect, and evolve in the KM environment (Kasemsap, 2016i).

Organizations that embed the practices of human capital and competency across a range of HRM activities effectively create and develop a boundary spanning culture connecting with various organizational disciplines in the global knowledge economy (Kasemsap, 2016j). Cultural influences are changing dramatically, as organizational cultures are no longer dependent on local resources to formulate their characteristic tastes, preferences, and organizational behavior (Kasemsap, 2015c). The relationships among SHRM, organizational learning, KM, HCM, CoPs, and cross-cultural organizational behavior in the modern workplace will be the beneficial topics for future research directions.

CONCLUSION

This article highlighted the concept of SHRM; the concept of e-HRM; the importance of SHRM in modern organizations; and the current of organizational learning and KM in modern organizations. SHRM, organizational learning, and KM are the essential drivers of productivity growth and competitive advantage in the digital age. SHRM interacts with other departments within an organization in order to understand their goals and to create a strategy that aligns with organizational goals. Effective SHRM requires favorable organizational resources, such as the modern technologies (e.g., e-HRM, the Internet, and social media platforms) and the highly qualified employees through organizational learning, KM, professional development, and talent management.

Organizational learning and KM occur due to the interactions of organizational structure, strategy, environment, technology, and culture.

Organizational learning and KM become popular among organizations which are interested in facilitating competitive advantage, innovation, and effectiveness in the knowledge economy. SHRM, organizational learning, and KM are vital for modern organizations to sustain competitive advantage and promote organizational innovation. Regarding SHRM and e-HRM perspectives, organizational policies and agreements between employers and employees are the crucial perspectives of facilitating workplace learning in the effective KM environment. Promoting SHRM, organizational learning, and KM has the potential to enhance organizational performance and achieve strategic goals in modern organizations.

REFERENCES

- Argote, L., McEvily, B., & Reagans, R. (2003). Managing knowledge in organizations: An integrative framework and review of emerging themes. *Management Science*, 49(4), 571–582. doi:10.1287/mnsc.49.4.571.14424
- Balabanis, G., Theodosiou, M., & Katsikea, E. (2004). Export marketing: Developments and a research agenda. *International Marketing Review*, 21(4/5), 353–377. doi:10.1108/02651330410547081
- Barnett, W. P., & Burgelman, R. A. (1996). Evolutionary perspectives on strategy. *Strategic Management Journal*, 17(S1), 5–19. doi:10.1002/smj.4250171003
- Becker, B., & Huselid, M. A. (1998). High performance work systems and firm performance. *Research in Personnel and Human Resources Management*, 16(1), 53–101.
- Bell, B. S., Lee, S. W., & Yeung, S. K. (2006). The impact of E-HR on professional competence in HRM: Implications for the development of HR professionals. *Human Resource Management*, 45(3), 295–308. doi:10.1002/hrm.20113

- Bengtsson, L. (2004). Explaining born globals: An organisational learning perspective on the internationalisation process. *International Journal of Globalisation and Small Business*, 1(1), 28–41. doi:10.1504/IJGSB.2004.005616
- Bennett, N., Ketchen, D. J. Jr, & Schultz, E. B. (1998). An examination of factors associated with the integration of human resource management and strategic decision making. *Human Resource Management*, 37(1), 3–16. doi:10.1002/(SICI)1099-050X(199821)37:1<3::AID-HRM2>3.0.CO;2-0
- Bondarouk, T. (2014). *Orchestrating the e-HRM symphony*. Retrieved from <https://www.utwente.nl/academischeplechtigheden/oraties/archief/2007-2014/oratieboekje-bondarouk.pdf>
- Brouthers, L. E., Nakos, G., Hadjimarcou, J., & Brouthers, K. D. (2009). Key success factors for successful export performance for small firms. *Journal of International Marketing*, 17(3), 21–38. doi:10.1509/jimk.17.3.21
- Chiva, R., & Alegre, J. (2009). Organizational learning capability and job satisfaction: An empirical assessment in the ceramic tile industry. *British Journal of Management*, 20(3), 323–340. doi:10.1111/j.1467-8551.2008.00586.x
- Dayaram, K., & Fung, L. (2014). Organizational learning in the Philippines: How do team and individual learning contribute? *Asia Pacific Journal of Human Resources*, 52(4), 420–442. doi:10.1111/1744-7941.12039
- Fernandez-Mesa, A., & Alegre, J. (2015). Entrepreneurial orientation and export intensity: Examining the interplay of organizational learning and innovation. *International Business Review*, 24(1), 148–156. doi:10.1016/j.ibusrev.2014.07.004
- Gavetti, G., & Rivkin, J. W. (2007). On the origin of strategy: Action and cognition over time. *Organization Science*, 18(2), 420–439. doi:10.1287/orsc.1070.0282
- Golden, K., & Ramanujam, V. (1985). Between a dream and a nightmare: On the integration of the human resource management and strategic business planning processes. *Human Resource Management*, 24(4), 429–452. doi:10.1002/hrm.3930240405
- Guest, D. A. (1987). Human resource management and industrial relations. *Journal of Management Studies*, 24(5), 503–521. doi:10.1111/j.1467-6486.1987.tb00460.x
- Huselid, M. A. (1995). The impact of human resource management on turnover, productivity, and configurational performance predictions. *Academy of Management Journal*, 38(3), 635–672. doi:10.2307/256741
- Huselid, M. A., Jackson, S. E., & Schuler, R. S. (1997). Technical and strategic human resource management effectiveness as determinants of firm performance. *Academy of Management Journal*, 40(1), 171–188. doi:10.2307/257025
- Johanson, J., & Vahlne, J. E. (2009). The Uppsala internationalization process model revisited: From liability of foreignness to liability of outsidership. *Journal of International Business Studies*, 40(9), 1411–1431. doi:10.1057/jibs.2009.24
- Kasemsap, K. (2014a). Strategic innovation management: An integrative framework and causal model of knowledge management, strategic orientation, organizational innovation, and organizational performance. In P. Ordóñez de Pablos & R. Tennyson (Eds.), *Strategic approaches for human capital management and development in a turbulent economy* (pp. 102–116). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-4530-1.ch007
- Kasemsap, K. (2014b). The role of knowledge sharing on organisational innovation: An integrated framework. In L. Al-Hakim & C. Jin (Eds.), *Quality innovation: Knowledge, theory, and practices* (pp. 247–271). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-4769-5.ch012

Kasemsap, K. (2014c). The role of social networking in global business environments. In P. Smith & T. Cockburn (Eds.), *Impact of emerging digital technologies on leadership in global business* (pp. 183–201). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-6134-9.ch010

Kasemsap, K. (2015a). The role of data mining for business intelligence in knowledge management. In A. Azevedo & M. Santos (Eds.), *Integration of data mining in business intelligence systems* (pp. 12–33). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-6477-7.ch002

Kasemsap, K. (2015b). Developing a framework of human resource management, organizational learning, knowledge management capability, and organizational performance. In P. Ordoñez de Pablos, L. Turró, R. Tennyson, & J. Zhao (Eds.), *Knowledge management for competitive advantage during economic crisis* (pp. 164–193). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-6457-9.ch010

Kasemsap, K. (2015c). The role of cultural dynamics in the digital age. In B. Christiansen & J. Koeman (Eds.), *Nationalism, cultural indoctrination, and economic prosperity in the digital age* (pp. 295–312). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-7492-9.ch014

Kasemsap, K. (2016a). Promoting leadership development and talent management in modern organizations. In U. Aung & P. Ordoñez de Pablos (Eds.), *Managerial strategies and practice in the Asian business sector* (pp. 238–266). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-9758-4.ch013

Kasemsap, K. (2016b). The roles of knowledge management and organizational innovation in global business. In G. Jamil, J. Poças-Rascão, F. Ribeiro, & A. Malheiro da Silva (Eds.), *Handbook of research on information architecture and management in modern organizations* (pp. 130–153). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-8637-3.ch006

Kasemsap, K. (2016c). Advocating entrepreneurship education and knowledge management in global business. In N. Baporikar (Ed.), *Handbook of research on entrepreneurship in the contemporary knowledge-based global economy* (pp. 313–339). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-8798-1.ch014

Kasemsap, K. (2016d). The roles of e-learning, organizational learning, and knowledge management in the learning organizations. In E. Railean, G. Walker, A. Elçi, & L. Jackson (Eds.), *Handbook of research on applied learning theory and design in modern education* (pp. 786–816). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-9634-1.ch039

Kasemsap, K. (2016e). The roles of lifelong learning and knowledge management in global higher education. In P. Ordoñez de Pablos & R. Tennyson (Eds.), *Impact of economic crisis on education and the next-generation workforce* (pp. 71–100). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-9455-2.ch004

Kasemsap, K. (2016f). Creating product innovation strategies through knowledge management in global business. In A. Goel & P. Singhal (Eds.), *Product innovation through knowledge management and social media strategies* (pp. 330–357). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-9607-5.ch015

Kasemsap, K. (2016g). The roles of information technology and knowledge management in global tourism. In A. Nedelea, M. Korstanje, & B. George (Eds.), *Strategic tools and methods for promoting hospitality and tourism services* (pp. 109–138). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-9761-4.ch006

Kasemsap, K. (2016h). A unified framework of organizational perspectives and knowledge management and their impact on job performance. In A. Normore, L. Long, & M. Javidi (Eds.), *Handbook of research on effective communication, leadership, and conflict resolution* (pp. 267–297). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-9970-0.ch015

- Kasemsap, K. (2016i). Utilizing communities of practice to facilitate knowledge sharing in the digital age. In S. Buckley, G. Majewski, & A. Giannakopoulos (Eds.), *Organizational knowledge facilitation through communities of practice in emerging markets* (pp. 198–224). Hershey, PA: IGI Global. doi:10.4018/978-1-5225-0013-1.ch011
- Kasemsap, K. (2016j). Analyzing the roles of human capital and competency in global business. In S. Sen, A. Bhattacharya, & R. Sen (Eds.), *International perspectives on socio-economic development in the era of globalization* (pp. 1–29). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-9908-3.ch001
- Kim, N., & Atuahene-Gima, K. (2010). Using exploratory and exploitative market learning for new product development. *Product Development & Management Association*, 27(4), 519–536.
- Lengnick-Hall, M. L., Lengnick-Hall, C. A., Andrade, L. S., & Drake, B. (2009). Strategic human resource management: The evolution of the field. *Human Resource Management Review*, 19(2), 64–85. doi:10.1016/j.hrmr.2009.01.002
- Lepak, D. P., Liao, H., Chung, Y., & Harden, E. E. (2006). A conceptual review of human resource management systems in strategic human resource management research. *Research in Personnel and Human Resources Management*, 25, 217–271.
- Lionzo, A., & Rossignoli, F. (2013). Knowledge integration in family SMEs: An extension of the 4I model. *Journal of Management & Governance*, 17(3), 583–608. doi:10.1007/s10997-011-9197-y
- Liu, Y., Combs, J. G., Ketchen, D. J. Jr, & Ireland, R. D. (2007). The value of human resource management for organizational performance. *Business Horizons*, 50(6), 503–511. doi:10.1016/j.bushor.2007.07.002
- Marler, J. H. (2009). Making human resources strategic by going to the Net: Reality or myth? *International Journal of Human Resource Management*, 20(3), 515–526. doi:10.1080/09585190802707276
- Marler, J. H., & Fisher, S. L. (2013). An evidence-based review of e-HRM and strategic human resource management. *Human Resource Management Review*, 23(1), 18–36. doi:10.1016/j.hrmr.2012.06.002
- Melo, S., & Beck, M. (2015). Intra and interorganizational learning networks and the implementation of quality improvement initiatives: The case of a Portuguese teaching hospital. *Human Resource Development Quarterly*, 26(2), 155–183. doi:10.1002/hrdq.21207
- Real, J. C., Roldán, J. L., & Leal, A. (2014). From entrepreneurial orientation and learning orientation to business performance: Analysing the mediating role of organizational learning and the moderating effects of organizational size. *British Journal of Management*, 25(2), 186–208. doi:10.1111/j.1467-8551.2012.00848.x
- Ruël, H. J. M., Bondarouk, T. V., & Looise, J. K. (2004). E-HRM: Innovation or irritation. Explorative empirical study in five large companies on web-based HRM. *Management Review*, 15(3), 364–380.
- Sabir, H. M., & Kalyar, M. N. (2013). Firm's innovativeness and employee job satisfaction: The role of organizational learning culture. *Interdisciplinary Journal of Contemporary Research in Business*, 4(9), 670–686.
- Schalk, R., Timmerman, V., & van den Heuvel, S. (2013). How strategic considerations influence decision making on e-HRM applications. *Human Resource Management Review*, 23(1), 84–92. doi:10.1016/j.hrmr.2012.06.008
- Schuler, R. S. (1992). Strategic human resource management: Linking people with the needs of the business. *Organizational Dynamics*, 21(1), 18–32. doi:10.1016/0090-2616(92)90083-Y
- Schuler, R. S., & Jackson, S. E. (1987). Organizational strategy and organization level as determinants of human resource management practices. *Human Resource Planning*, 10(3), 125–141.

Slater, S. F., & Narver, J. C. (1995). Market orientation and the learning organization. *Journal of Marketing*, 59(3), 63–74. doi:10.2307/1252120

Strohmeier, S. (2007). Research in e-HRM: Review and implications. *Human Resource Management Review*, 17(1), 19–37. doi:10.1016/j.hrmr.2006.11.002

Ugurluoglu, O., Aldogan, E. U., & Dilmac, E. (2013). The impact of managers perceptions of learning organizations on innovation in healthcare: Sample of Turkey. *The International Journal of Health Planning and Management*, 28(2), e158–e168. doi:10.1002/hpm.2143 PMID:23047766

Wright, P. M., Dunford, B. B., & Snell, S. A. (2001). Human resources and the resource-based view of the firm. *Journal of Management*, 27(6), 701–721. doi:10.1177/014920630102700607

Wright, P. M., & McMahan, G. C. (1992). Theoretical perspectives for strategic human resource management. *Journal of Management*, 18(2), 295–320. doi:10.1177/014920639201800205

Zahra, S. A. (2012). Organizational learning and entrepreneurship in family firms: Exploring the moderating effect of ownership and cohesion. *Small Business Economics*, 38(1), 51–65. doi:10.1007/s11187-010-9266-7

ADDITIONAL READING

Alegre, J., & Chiva, R. (2013). Linking entrepreneurial orientation and firm performance: The role of organizational learning capability and innovation performance. *Journal of Small Business Management*, 51(4), 491–507. doi:10.1111/jsbm.12005

Argote, L. (2015). An opportunity for mutual learning between organizational learning and global strategy researchers: Transactive memory systems. *Global Strategy Journal*, 5(2), 198–203. doi:10.1002/gsj.1096

Arthur-Mensah, N., & Shuck, B. (2014). Learning in developing countries: Implications for workforce training and development in Africa. *New Horizons in Adult Education and Human Resource Development*, 26(4), 41–46. doi:10.1002/nha3.20084

Bingham, C. B., Heimeriks, K. H., Schijven, M., & Gates, S. (2015). Concurrent learning: How firms develop multiple dynamic capabilities in parallel. *Strategic Management Journal*, 36(12), 1802–1825. doi:10.1002/smj.2347

Borrás, S., & Højlund, S. (2015). Evaluation and policy learning: The learners' perspective. *European Journal of Political Research*, 54(1), 99–120. doi:10.1111/1475-6765.12076

Cho, S. H., Song, J. H., Yun, S. C., & Lee, C. K. (2013). How the organizational learning process mediates the impact of strategic human resource management practices on performance in Korean organizations. *Performance Improvement Quarterly*, 25(4), 23–42. doi:10.1002/piq.21127

Chronéer, D., & Backlund, F. (2015). A holistic view on learning in project-based organizations. *Project Management Journal*, 46(3), 61–74. doi:10.1002/pmj.21503

Eldor, L., & Harpaz, I. (2016). A process model of employee engagement: The learning climate and its relationship with extra-role performance behaviors. *Journal of Organizational Behavior*, 37(2), 213–235. doi:10.1002/job.2037

Erkelens, R., van den Hooff, B., Huysman, M., & Vlaar, P. (2015). Learning from locally embedded knowledge: Facilitating organizational learning in geographically dispersed settings. *Global Strategy Journal*, 5(2), 177–197. doi:10.1002/gsj.1092

Evans, S. D., & Kivell, N. (2015). The transformation team: An enabling structure for organizational learning in action. *Journal of Community Psychology*, 43(6), 760–777. doi:10.1002/jcop.21756

- Froehlich, D., Segers, M., & van den Bossche, P. (2014). Informal workplace learning in Austrian banks: The influence of learning approach, leadership style, and organizational learning culture on managers' learning outcomes. *Human Resource Development Quarterly*, 25(1), 29–57. doi:10.1002/hrdq.21173
- Hotho, J. J., Lyles, M. A., & Easterby-Smith, M. (2015). The mutual impact of global strategy and organizational learning: Current themes and future directions. *Global Strategy Journal*, 5(2), 85–112. doi:10.1002/gsj.1097
- Huang, L., & Luthans, F. (2015). Toward better understanding of the learning goal orientation–creativity relationship: The role of positive psychological capital. *Applied Psychology*, 64(2), 444–472. doi:10.1111/apps.12028
- Hung, Y. H., Lin, C. F., & Chang, R. I. (2015). Developing a dynamic inference expert system to support individual learning at work. *British Journal of Educational Technology*, 46(6), 1378–1391. doi:10.1111/bjet.12214
- Karkoulis, S., Messarra, L. C., & McCarthy, R. (2013). The intriguing art of knowledge management and its relation to learning organizations. *Journal of Knowledge Management*, 17(4), 511–526. doi:10.1108/JKM-03-2013-0102
- Kim, T. (2015). Impacts of learning interventions on organizational human capital and performance. *Performance Improvement Quarterly*, 27(4), 7–28. doi:10.1002/piq.21180
- Lai, J., Lui, S. S., & Tsang, E. W. K. (2016). Intrafirm knowledge transfer and employee innovative behavior: The role of total and balanced knowledge flows. *Journal of Product Innovation Management*, 33(1), 90–103. doi:10.1111/jpim.12262
- Lee, J., Zo, H., & Lee, H. (2014). Smart learning adoption in employees and HRD managers. *British Journal of Educational Technology*, 45(6), 1082–1096. doi:10.1111/bjet.12210
- Li, Y., Chen, H., Liu, Y., & Peng, M. W. (2014). Managerial ties, organizational learning, and opportunity capture: A social capital perspective. *Asia Pacific Journal of Management*, 31(1), 271–291. doi:10.1007/s10490-012-9330-8
- Lim, D. H., Yoon, S. W., & Park, S. (2013). Integrating learning outcome typologies for HRD: Review and current status. *New Horizons in Adult Education and Human Resource Development*, 25(2), 33–48. doi:10.1002/nha.20015
- Manuti, A., Pastore, S., Scardigno, A. F., Giancaspro, M. L., & Morciano, D. (2015). Formal and informal learning in the workplace: A research review. *International Journal of Training and Development*, 19(1), 1–17. doi:10.1111/ijtd.12044
- Marshall, J., & Rossett, A. (2014). Perceptions of barriers to the evaluation of workplace learning programs. *Performance Improvement Quarterly*, 27(3), 7–26. doi:10.1002/piq.21173
- Naveh, E., Katz-Navon, T., & Stern, Z. (2015). Active learning climate and employee errors: The moderating effects of personality traits. *Journal of Organizational Behavior*, 36(3), 441–459. doi:10.1002/job.2000
- Peltokorpi, V., & Hasu, M. (2015). Moderating effects of transformational leadership between external team learning and research team performance outcomes. *R & D Management*, 45(3), 304–316. doi:10.1111/radm.12091
- Sung, S. Y., & Choi, J. N. (2014). Do organizations spend wisely on employees? Effects of training and development investments on learning and innovation in organizations. *Journal of Organizational Behavior*, 35(3), 393–412. doi:10.1002/job.1897 PMID:25598576
- Thomas, N., & Vohra, N. (2015). Development of network measures for knowledge processes: A relational framework. *Knowledge and Process Management*, 22(2), 126–139. doi:10.1002/kpm.1465

Waight, C. L. (2015). Learning during the integration phase of mergers and acquisitions: Perspectives from learning and development professionals. *Performance Improvement Quarterly*, 28(1), 7–26. doi:10.1002/piq.21184

Yeo, R. K., & Li, J. (2013). In pursuit of learning: Sensemaking the quality of work life. *European Journal of Training and Development*, 37(2), 136–160. doi:10.1108/03090591311301662

Za, S., Spagnoletti, P., & North-Samardzic, A. (2014). Organisational learning as an emerging process: The generative role of digital tools in informal learning practices. *British Journal of Educational Technology*, 45(6), 1023–1035. doi:10.1111/bjet.12211

KEY TERMS AND DEFINITIONS

Human Resource: The resource that resides in the knowledge, skills, and motivation of people.

Human Resource Management: The process of hiring and developing employees so that they become more valuable to the organization.

Knowledge Creation: The formation of new ideas through the interactions between explicit and tacit knowledge in individual human minds.

Knowledge Management: The strategies and processes designed to identify, capture, structure, value, organize, and share an organization's intellectual assets to enhance its performance and competitiveness.

Learning: The measurable and relatively permanent change in behavior through experience, instruction, and study.

Learning Organization: The organization that obtains knowledge in order to survive in a rapidly changing environment.

Organizational Learning: The organization-wide continuous process that enhances its collective ability to accept, make sense of, and respond to the internal and external change.

Strategic Human Resource Management: The proactive management of the employees of an organization.

Technology, Learning Styles, Values, and Work Ethics of Millennials

Harish C. Chandan
Argosy University, USA

INTRODUCTION

Millennials represent people who were born between 1981 and 1997. In 2017, they are 20 to 36 years old. The Millennials are also called “Y” generation, digital natives, “net generation” or “me generation”. At present, the Millennials are university students and are entering the work force. The evolving technology has shaped their learning styles, values and work ethics. This chapter summarizes the influence of technology on learning styles, values and work ethics of Millennials. Implications of the different learning styles of the Millennials for the instructors’ teaching style, leadership style of university administrators and student services are discussed. The work ethics of Millennials are explored and implications for managerial and mentoring practices are discussed.

BACKGROUND

The entire adult population can be viewed as four generational cohorts – Silent (1928-1945), Baby Boomer (1946-1964), X (1965-1980), Y (1981-1997). The numbers in the parenthesis represent the range of birth years. There are slight variations in the birth years among different researchers (Pew Research Center, 2015).

Each generation has different values and viewpoints, which were shaped by their social, economic and political environment when they were growing up and their current level of responsibility. Different generational cohorts have different perceptions of each other and have different expectations. These perceptions are sometimes

based on positive and negative stereotypes reported in the popular press. These views are often not subjected to objective empirical scrutiny and there are variations within a generation (Macky et al., 2008). This generation gap can be a source of conflict and misunderstanding at the work place (Meriac et al., 2010).

TECHNOLOGY, NARRATIVE ABOUT MILLENNIALS, AND THEIR VALUES

Some of the major influences in a person’s development include peers, parents, popular culture, major political and social events. The differences in the social context and different shared life experiences of different generations lead to different beliefs, values and attitude towards work. The conventional view of the Millennials as narcissistic, self-absorbed, distrustful, anxious, cynical and lonely in the current competitive job-market may not be correct. There is no empirical evidence that the Millennials have a bad attitude. The Boomers’ perception of Millennials may reflect the old guard’s bias to new generations (Kowske, Rasch and Wiley, 2010). It may be that Millennials are adapting to the changing world that other generations are trying to resist (Ellin, 2014).

A lot of narratives about Millennials come from other generational cohorts and may reflect the biases of that generation (Twenge, 2009). The parenting styles, political events, social and cultural trends, technology and economic events during the time a generation is growing up contribute to the evolution of the psychosocial characteristics of a generation (Strauss and Howe, 1997; Howe

and Strauss, 2000). The differences in the social context and different shared life experiences of different generations lead to different beliefs, values, expectations, and attitude towards education and work. The popular press has both positive and negative stereotypes about Millennials, Table 1.

Millennials use texting more than e-mail as a regular mode of communication. They are comfortable with a wide variety of media including blogs, reviews and social networks to openly express their interests and feelings (Hershatter and Epstein, 2010). The Millennials watch less television and are not unduly influenced by the mainstream media. They are much more resistant to advertisements than previous generations (Ciminillo, 2005). Millennials are constantly adapting to the new forms of social media. The e-mail and Facebook may have become the grandpa's social media as the Millennials switch to Instagram. Millennials multitask using different technologies. Their multi-tasking behavior includes communicating with many people while playing Xbox (Putre, 2013; PwC, 2013). They are more connected digitally but may be socially isolated.

In today's hypermodern times, assembling sociality has become more challenging (Warde, 1999). The Millennials use social media to coordinate, stack or shift their social interactions to coordinate with personal schedules. Millennials migrate across a range of devices and platforms at on-line and off-line sites. Media are used for

entertainment and socializing (Botterill et al., 2015; Strauss and Howe, 2000; Hoover, 2009; PwC, 2008, Johnson Controls, 2010).

TECHNOLOGY AND LEARNING STYLES OF MILLENNIALS

Incorporation of technology in the classroom does enhance learning and is moderated by the student characteristics (Krentler and Willis-Flurry, 2005). As an example, the model of social learning environment as inquiry-based on cloud technology is being applied for enhancing the critical thinking skills and collaborative learning (Meepian and Wannapiroon, 2013). Multimedia and hypermedia Technologies affects positively on both learning in a content area and learning to use technology itself (Kinzer and Leu 1997). Technology can enhance literacy development, impact language acquisition, provide greater access to information, support learning, motivate students, and enhance their self-esteem (Boster et al., 2004; Tracey & Young, 2006).

Millennial students are used to instantaneous answers without much deliberate thought. This hinders the development of critical thinking. They do not realize that their lack of attention span or ability prevents them from focusing and contemplating the material deeply. They think that that they are not able to think critically due to lack of time. Millennials are comfortable learning through webinars, social media and virtual meetings. They are proficient users of productivity software for increased productivity at the workplace. Millennials find the right information and eliminate the unnecessary information (Ellin, 2014).

Raised by the overly involved 'helicopter parents', Millennials crave for immediate feedback, praise, attention and guidance. Many Millennials students lack critical thinking skills required for inductive and deductive reasoning (Monaco and Martin, 2007). They lack the original thought due to excessive reliance on the Internet. There is lack of concern for accuracy and validity of their

Table 1. Positive and negative stereotypes of millennials (Source: Twenge and Campbell, 2008)

Positive Stereotypes	Technology and Social media savvy, Multi-tasking, Digitally connected, Confident, Self-expressive, Liberal, Upbeat and Open to change, Civic minded with a sense of both local and global community
Negative Stereotypes	Narcissistic, Sense of entitlement, Coddled, Self-promotional, Opinionated, Whiny and Needy, Seek constant feedback and immediate gratification, Lack of diligence, Poor task performance and shirking behaviors

research sources. They are inclined to trust peer opinion and public consensus without verification (Hershatter & Epstein, 2010). Millennials need to practice critical thinking through meta-cognition, i.e. they need to be aware of what they are thinking and how they are thinking. In other words, they need to be able to verbalize the steps involved in solving a problem. Millennials need to be engaged with co-operative learning exercises and need help with analyzing their own learning strategies (McGlynn, 2005).

One has to recognize the students' generational characteristics and learning style and match it with the teaching style of instructors (Benfer and Shanahan, 2013). The Millennials thrive on working collaboratively in a flexible and open work or classroom environment. At Syracuse University, a collaborative classroom where new infrastructure is designed to promote collaborative, active learning is replacing the traditional classroom. The collaborative classroom features 10 interconnected LED monitors, reconfigurable tables, mobile whiteboards, dual overhead projectors, and a high tech teaching station. The classroom is designed for students to explore their course material in ways that are not possible in traditional classrooms. Instructors design their classroom activities to foster team-based learning (Syracuse Engineer, 2015).

Being technology savvy, the Millennials' educational expectations, socialization characteristics and learning styles are different as compared with previous generations. This requires changes in the educational programs, student services, and instructional strategies (Lowery, 2004). Universities and Colleges are redesigning student services and educational facilities including textbooks, library services and teaching methods to best serve the Millennials. Millennials prefer flexibility and mobility to learn in their own time and on their own terms. Millennials do multi-tasking, appreciate experiential learning, active engagement and prefer teamwork (Oblinger and Raines, 2002).

Freshmen retention is a critical component to the academic and financial success of an educa-

tional institution (Tinto, 2009). The persistence and retention of the Millennial freshmen college students is a challenge (Rickes, 2009). About 75% of the freshman students who do not return for sophomore year, more than 60% will not return to the same institution (Burshong, 2009). To improve the freshman retention for Millennials, universities need to focus on four new strategies including freshman focused activities, helping with effective study skills, more supportive instructor-student relationship and offering academic advisement support. The collaborative and interactive learning environment, ongoing academic guidance, study and time management skills development help a social and academic transition for the freshman students (Turner and Thompson, 2013). Universal design in learning and instruction helps with furthering educational goals for students. This leads to better learning outcomes and educational success (Much et al., 2014; Palmer, 2015).

TECHNOLOGY AND WORK ETHICS OF MILLENNIALS

For the first time in history, four distinct generations are coexisting in the work force—Millennials, Generation X, Baby Boomers and Traditionalists, Table 2 (PRC, 2015).

The *multi-dimensional work ethic profile* (MWEP) is a 7-dimensional construct and is defined as set of beliefs and attitudes reflecting the behavior at work including centrality of work, self-reliance, hard work, leisure, morality/ethics, delay of gratification and wasted time (Miller et al., 2002). Pogson et al. (2003) used MWEP to examine differences in work ethic as a function of one's career stage and found that differences in various dimensions of work ethic exist as a function of career stage. The limitations of this study were that the generational differences were confounded with age/career stage and the measurement equivalence of MWEP across the three career stages was not tested. Meriac et al., 2010 measured the generational differences in work

Table 2. Characteristics of four generations at work

Generation	% of Work Force	Characteristics at Work
Millennials (1981-1996)	27.7	Technology savvy, Respect for diversity, Socially conscious, Confident, Optimistic, Achievement oriented
Gen X (1965-1980)	19.8	Entrepreneurs, Self-reliant, Realist, Results-oriented, informal, fun,
Baby Boomers (1946-1964)	26.4	Workaholics, More conventional, Competitive, Optimistic, Personal gratification
Traditionalists (1928-1945)	13	Respect for rules, Conformity, Loyal, Hard work, Frugal

ethic using MWEP and examined the measurement equivalence across the three generational cohorts – Baby Boomers, Generation X and Millennials. It was found that the Millennials differ from other generations on several dimensions of the MWEP. However, the lack of measurement equivalence in several comparisons indicated that the different generations interpret the item content in the survey differently (Meriac et al., 2010).

From managerial standpoint, it may be more important to focus on learning about what the Millennials' values and attitudes regarding work are and not focus on the relevance of the generation differences. Comparing work ethics for different generations at a given time has several complexities including the effect of age, career-stage and applicability of the instrument used to measure the different generations.

Seven distinct characteristics have been identified for the Millennials in the workplace including *technological savvy*, *collaborative creativity*, *multi-tasking*, *work-life balance*, *social responsibility*, *education*, *diversity*, *open to change*, (Bannon et al., 2011; Innovation book). Each of seven characteristics has managerial implications in terms of mentoring and managing Millennials.

Being *technological savvy* includes workplace mobility, wireless devices, texting, and connected on social networks. At work, Millennials expect to have access to mobile and portable technologies. Millennials can positively contribute to the transformation of information and communication technologies to virtualization including cloud computing. Virtualization eliminates the need for

back-up copies of data on real storage devices. Cloud computing enables the storing of data for clients and employees. Companies can use the technology savvy characteristic of Millennials for learning new skills and training them through on-line simulations and case studies including new employee orientation, customer relations and ethical behavior. Their social habits include social networking, connecting, talking, sharing, tagging and creating and distributing content. Firms need to exploit the business advantages of social networking to better communicate and understand their Millennial employees, customers and other stakeholders.

Millennials are comfortable building relationships online. According to Mark Zuckerberg, "People have really gotten comfortable not only sharing more information and different kinds, but more openly and with more people. That social norm is just something that has evolved over time". The Face book founder Mark Zuckerberg was born in 1984 and is a Millennial. (Mossberg and Swisher, 2010).

Social media is used within and outside a company. Companies can incorporate social networking into the internal communication channels. Outside a company, the use of social media can help with company-customer, recruitment and obtaining background information on job applicants. Within a company, employees can share their organizational experience to share a story. In some cases, employees can intentionally expose the organizational issues and practices to resist organizational control. IBM encourages

employees to exchange ideas through blogs, and online communication. It promotes a dialogue between their partners, clients and other stakeholders. IBM employees use social media to informally communicate with other employees, managers and leaders on internal social media. BBC limits employee's social media use and has a clear social media policy where BBC-related photographs, offensive comments and videos are considered a disciplinary offense. Millennials perceive higher person-organization fit (POF) for a company with organizational policies supporting employees' social media use (Cho et al., 2013). Sharing information on-line can involve some work-related transgressions like blogging or tweeting negatively about their own company. Millennials need to be coached regarding social media policy of the company and information confidentiality.

The *collaborative creativity* characteristics of Millennials' is reflected in their preference for working collaboratively in a flexible and open work environment. Millennials are very comfortable with *multi-tasking*. They are able to process several incoming stimuli simultaneously and are excellent at multitasking. They are able to take in large doses of information from multiple sources simultaneously.

Millennials have benefited from playing video games in terms of decision-making, problem solving and thinking on their feet. Video games force an individual to analyze the options available and react quickly. Video games increase manual dexterity, eye-hand coordination and speed up neural pathways (Shotton, 1989). Millennials benefit from video games in terms of motivation creativity and cognitive skills as games aspire players to conquer a level to reach more difficult level. Millennials have developed hypertext minds that allow them to gather information rapidly from multiple sources and make connections between various sources of data.

The fourth of the seven characteristics of the Millennials is *work-life balance*, which is a big concern for them. They believe that the quality

of their work is more important than the number of hours spent in the traditional office. Flex Time is the preferred working mode for Millennials. This has a potential to transform the traditional nine-to-five office schedule for corporations to a more elastic timeframe. In addition, it has already transformed the parental leave policy. Facebook joined tech companies Amazon, Netflix and Microsoft Corp. in extending parental leave. In doing so, they are addressing the perception that today's workplace culture is incompatible with family and personal life. Facebook is extending its parental leave policy to four months paid time off for all employees around the world (Lori Matloff Goler, Head of human resources at Facebook, WSJ, 2016). A high task variety, high task identity, and a moderate level of both received interdependence and task significance are recognized as common job characteristics of knowledge workers across generations (Hernaus and Poloski, 2014).

Many Boomers expect that the Millennials will 'pay their dues' and 'earn respect'. Millennials want to be respected and valued for what they contribute, not for how long they have been with the organization (Deloitte, 2009). When Boomers see that the Millennials' do not have the same workaholic approach, they doubt the Millennials' ability to fit into the existing workgroup and culture (Expectancy theory) and demean them (Myers and Sadaghiani, 2010).

Millennials get bored easily and dislike menial tasks. They expect to move ahead quickly based on skills rather than experience (Eiser, 2009). Millennials will likely to want to work fewer hours (Deal, Altman and Rogelberg, 2010). They want their employers to be socially and environmentally conscious. They want meaningful relationships with clients and peers. They prefer flatter organizational structures (Shaikh, 2010). Millennials prefer brief communications with interactivities. They prefer webinars that entertain and educate (Dorsey, 2010).

A majority of the Millennials says that the older generation's work ethic and moral values are better than their own generation (Pew Research

Center, 2010). A lot of popular literature exists on perceptions about Millennials' work related attitudes and values, but there is limited empirical scientific data on comparing work ethics of different generational cohorts (Mackay et al., 2008).

The organizations have to develop strategies for managing the interactions between generations including mentoring, team working, and the design of the physical and virtual workplace to maximize productivity. The changing pattern of work and life dictate that organizations have to adapt their organizational culture to meet the demands and expectations of new generations in the workplace (Bannett, 2012).

The differences in work-related attitudes and values among different generations can be a source of misunderstandings and conflict in the workplace (Twenge and Campbell, 2008).

FUTURE RESEARCH DIRECTIONS

There is very little research available about the most effective way of teaching Millennials. More research is needed to improve retention of Millennial students in colleges by updating the educational curriculum in a way that promotes a collaborative and interactive student-centered environment (Toohey et al., 2016). The role of multi-tasking on the critical thinking skills for Millennials needs to be understood better. Instructors have to be retrained for student centered learning.

The personality traits do not consistently predict the task performance and shirking behaviors like cyber loafing. Based on cognitive psychology, the performance on the Cognitive Reflection Test (CRT) is a good indicator of task performance and shirking behaviors (Corgnet et al., 2015). Research is needed to understand the role of CRT in understanding the cyber loafing phenomenon among the Millennials.

Managers need to question their assumptions about the Millennials, encourage open communication, provide positive feedback frequently and provide negative feedback with discretion

(Keene and Handrich, 2010). Research is needed to learn about effective ways of providing feedback and mentoring the Millennials. More research is needed about how organizations would benefit by bringing social media platforms (e.g. Yammer) into their internal communication strategy and increase employee involvement and commitment. Inter-generational mentoring may be a good way to enhance communication and understanding between the generations. More research is necessary about the effectiveness of inter-generational mentoring. We need to understand the Millennials as they are and not through the lens of the other generations' perceptions.

CONCLUSION

The Millennials are redefining what it means to be an individual in today's complex, connected, global and competitive society. Currently, Millennials are the largest generation in United States. Half of them are already in the workforce. Through the increasing use of ICT technologies, Millennials are reshaping the businesses, organizations, university student services and teaching practices including student-centered and collaborative learning. Firms have to figure out how to effectively engage the Millennials to achieve productivity gains, higher job satisfaction, and lower turnover rate. To improve retention, organizations have to create an open workplace culture, frequent communication, flexible work styles, skill-based compensation and career structure, career development and work-life balance opportunities that resonate with the aspirations of the Millennials. Teachers, leaders, managers and human resource professionals need to understand the Millennials' use of technology to enable them to update the teaching, student services and organizational practices to resonate more with the Millennials who represent the future of the society where one needs skills to go beyond the information age to move into the conceptual age where the big picture and innovation are essential.

REFERENCES

- Abraham, R., & Harrington, C. (2015). Consumption Patterns of the Millennial Generational Cohort. *Modern Economy*, 6(01), 51–64. doi:10.4236/me.2015.61005
- Bannon, S. (n.d.). Ford, Kelly and Meltzer, L., (20110, “Understanding Millennials in the Workplace. *The CPA Journal*, (Nov), 61–65.
- Benfer, E. A., & Shanahan, C. F. (2013). Educating the Invincibles: Strategies for Teaching the Millennial Generation in Law School. *Clinical Law Review*, 20(1), 1–37.
- Bennett, J., Pitt, M., & Price, S. (2012). Understanding the Impact of Generational Issues in the Workplace. *Facilities*, 30(7/8), 278–288. doi:10.1108/02632771211220086
- Botterill, J., Bredin, M., & Dun, T. (2015). Millennials’ Media Use: It is a Matter of Time. *Canadian Journal of Communication*, 40, 537–551.
- Burshong, S. (2009). Freshman retention Drops, Except at 2-Year Colleges. *The Chronicle of Higher Education*, 55(21), A17.
- Census. (2015). Retrieved from <http://www.census.gov/popclock>
- Chandler, N. (2013). Millennials, Entrepreneurs and the Hungarian Workplace of the Future- Preliminary Findings of a Pilot Study. *Vezetéstudomány*, 46(11), 15-24
- Cho, J., Park, D. J., & Ordonez, Z. (2013). Communication-Oriented Person –Organization Fit as a Key Factor of Job-Seeking Behaviors: Millennials Social Media Use and Attitudes Towards Organizational Social Media Policies. *Cyberpsychology, Behavior, and Social Networking*, 16(11), 794–799. doi:10.1089/cyber.2012.0528 PMID:23848961
- Ciminillo, J. A. (2005). Elusive Generation Y Demands Edgier Marketing. *Automotive News*, 79(6144), 28B.
- Corngnet, B., Hernan, R., & Mateo, R. (2015). Cognitive Reflection and the Diligent Worker: An Experimental Study of Millennials. *PLOS One*. DOI:10.1371/journal.pone.0141243
- Deal, J., Altman, D., & Rogelberg, S. (2010). Millennials at Work: What We Know and What We Need to Do (if anything). *Journal of Business and Psychology*, 25(2), 109–199. doi:10.1007/s10869-010-9177-2
- Deloitte. (2009). *Generation Y: Powerhouse of the Global Economy*. Retrieved from <http://www2.deloitte.com/content/dam/Deloitte/us/Documents/human-capital/us-consulting-hc-generationy-snapshot-041509.pdf>
- Dorsey, R. (2010). Inside the Mind of a Gen Y Employee. *SCM*, 14(2), 7.
- Eiser, B. (2009). Managing the Millennials. *Pennsylvania CPA Journal*. Retrieved from <http://www.picpa.oeg/Content/39996.aspx>
- Ellin, A. (2014, March). The Beat (Up) Generation. *Psychology Today*, 56-62.
- Ethics Resource Center (ERC), (2009). *Millennials, Gen X and Baby Boomers: Who’s Working at Your Company and What do They Think About Ethics*. National Business Ethics Survey.
- Fry, R. (2016). *Millennials overtake Baby Boomers as America’s largest generation*. Retrieved from <http://www.pewresearch.org/fact-tank/2016/04/25/millennials-overtake-baby-boomers/>
- Hernaus, T., & Pološki Vokic, N. (2014). Work Design for Different Generational Cohorts: Determining Common and Idiosyncratic Job Characteristics. *Journal of Organizational Change Management*, 27(4), 615–641. doi:10.1108/JOCM-05-2014-0104
- Hershat, A., & Epstein, M. (2010). Millennial and the World of Work—An Organization and Management Perspective. *Journal of Business and Psychology*, 25(2), 211–223. doi:10.1007/s10869-010-9160-y

- Hewlett, S., Sherbin, L., & Sumberg, K. (2009). How Gen Y and Boomers Will Reshape Your Agenda. *Harvard Business Review*, 87(7/8), 71–76. PMID:19630257
- Hoover, E. (2009). The Millennial Muddle. *The Chronicle of Higher Education*, 56(8), A1–A34.
- Howe, N., & Strauss, W. (2000). *Millennials Rising- The Next Great Generation*. New York: Knopf Doubleday Publishing Group.
- Johnson Controls. (2010). *Generation Y and Workplace*. Annual Report. Retrieved from <http://www.johnsoncontrols.com/content/us/en/search.html?q=Millennials>
- Keene, D. L., & Handrich, R. R. (2010, September). Between Coddling and Contempt: Managing and Mentoring Millennials. *The Jury Expert*, 1-14.
- Kowske, B., Rasch, R., & Wiley, J. (2010). Millennials (Lack of) Attitude Problem: An Empirical Examination of Generational Effects on Work Attitudes. *Journal of Business and Psychology*, 25(2), 265–279. doi:10.1007/s10869-010-9171-8
- Krentler, K. A., & Willis-Flurry, L. A. (2005). Does Technology Enhance Actual Student Learning? The Case of Online Discussion Boards. *Journal of Education for Business*, 80(6), 316–321. doi:10.3200/JOEB.80.6.316-321
- Lin, Y., Li, Y., & Hou, X. (2015). Utilitarian Orientation and Performance: Evidence from Chinese Millennial Generation Employees. *Social Behavior and Personality*, 43(9), 1463–1476. doi:10.2224/sbp.2015.43.9.1463
- Lowery, J. W. (2004). Student Affairs for a New Generation. *New Directions for Student Services*, 106(106), 87–98. doi:10.1002/ss.127
- Macky, K., Gardner, D., & Forsyth, S. (2008). Generational Differences at Work: Introduction and Overview. *Journal of Managerial Psychology*, 23(8), 857–861. doi:10.1108/02683940810904358
- McGlynn, A. G. P. (2005). *Teaching Millennials, Our Newest Cultural Cohort*. Retrieved from www.eddigest.com
- Meepian, A., & Wannapiroon, P. (2013). Design of Social Learning Environment as Inquiry-Based on Cloud Technology for Enhancing the Critical Thinking Skill and Collaborative Learning. *International Journal of e-Education, e-Business, e- Management Learning*, 3(3), 253–257.
- Meriac, J. P., Woehr, D. J., & Banister, C. (2010). Generational Differences in Work Ethic: An Examination of Measurement Equivalence Across Three Cohorts. *Journal of Business and Psychology*, 25(2), 315–324. doi:10.1007/s10869-010-9164-7
- Miller, M. J., Woehr, D. J., & Hudspeth, N. (2002). The Meaning and Measurement of Work Ethic: Construction and Initial Validation of a Multidimensional Inventory. *Journal of Vocational Behavior*, 60(3), 451–489. doi:10.1006/jvbe.2001.1838
- Monaco, M., & Martin, M. (2007). The Millennial Student: A New Generation of Learners. *Athletic Training Education Journal*, 2, 42–46.
- Mosberg, W., & Swisher, K. (2010, June 7). Getting Personal, Bur Too Personal?. *Wall Street Journal*.
- Much, K., Wagener, A. M., Breitreutz, H. L., & Hellenbrand, M. (2014). Working With the Millennial Generation: Challenges Facing 21st Century Students From the Perspective of University Staff. *Journal of College Counseling*, 17(1), 37–46. doi:10.1002/j.2161-1882.2014.00046.x
- Myers, K., & Sadaghiani, K. (2010). Millennials in the Workplace: A Communication Perspective on Millennials Organizational Relationships and Performance. *Journal of Business and Psychology*, 25(2), 225–238. doi:10.1007/s10869-010-9172-7 PMID:20502509

- Palmer, J. S. (2015). "The Millennials Are Coming": Improving Self-Efficacy in Law Students Through Universal Designs in Learning. Retrieved from [http://engagedscholarship.csuohio.edu/clevstlrev/63\(3\)](http://engagedscholarship.csuohio.edu/clevstlrev/63(3))
- Pew Research Center. (2010). *Millennials: Confident, Connected, and Open to Change*. Retrieved from <http://www.pewsocialtrends.org/2010/02/24/millennials-confident-connected-open-to-change/>
- Pew Research Center. (2015). *Comparing Millennials to Other Generations*. Retrieved from <http://www.pewresearch.org/fact-tank/2015/03/19/how-Millennials-compare-with-their-grandparents/>
- Pogson, C., Cober, A., Doverspike, D., & Rogers, J. (2003). Differences in Self Reported Work Ethic Across Three Career Stages. *Journal of Vocational Behavior*, 62(1), 189–201. doi:10.1016/S0001-8791(02)00044-1
- Putre, L. (2013, September). The March of the Millennials. *H&HN*, 38-42. Retrieved from www.hhnmag.com
- PwC's NextGen. A Global Generational Study. (2013). Retrieved from <http://www.pwc.com/us/en/people-management/publications/nextgen-global-generational-study.html>
- Rapoza, J. (2008, March). Those Tech-Savvy Millennials. *eWeek*, 49.
- Rickes, P. C. (2009). Make Way for Millennials! How Today's Students are Shaping Higher Education Space. *Planning for Higher Education*, 37(2), 7–17.
- Rudon, T. (2015). *Ten Benefits of Video Games*. Retrieved from http://www.selfgrowth.com/articles/10_Benefits_Of_Video_Games.html
- Shaikh, P. (2010). *Misunderstood Gen Y in the Workplace: Challenges Faced by the Millennial Generation*. Retrieved from <http://trainingpd.suite101.com/article.cfm/misunderstood-gen-y-in-the-workplace>
- Smith, K. T. (2011). Digital Marketing Strategies that Millennials Find Appealing Motivating or Just Annoying. *Journal of Strategic Marketing*, 19(6), 489–499. doi:10.1080/0965254X.2011.581383
- Society for Human Resource Management (SHRM). (2001). *Generation Y's Impact on the Workplace*. Retrieved from <http://www.shrm.org/Research/FutureWorkplaceTrends/Pages/0301b.aspx>
- Steinmetz, K. (2015). Help! My Parents are Millennials. *Time*, 37–43.
- Strauss, W., & Howe, N. (1997). *The Fourth Turning: An American Prophecy*. New York: Broadway Books.
- Syracuse Engineer. (2015). New Classroom Encourages Teamwork in Engineering Education. *Syracuse Engineer*, 41.
- Toohy, S. L., Wray, A., Wiechmann, W., Lin, M., & Boysen-Osborn, M. (2016). Ten Tips for Engaging the Millennial Learner and Moving an Emergency Medicine Residency Curriculum into the 21st Century. *Western Journal of Emergency Medicine: Integrating Emergency Care with Population Health*, 17(3), 337–343. doi:10.5811/westjem.2016.3.29863 PMID:27330668
- Turner, P., & Thompson, E. (2013). College Retention Initiatives Meeting the Needs of Millennial Freshman Students. *College Student Journal*, 94–104.
- Twenge, J. M. (2014). *Generation Me: Why Today's Young Americans are More Confident, Assertive, Entitled and More Miserable than Ever Before*. New York: Atria Paperback.
- Twenge, J. M., & Campbell, S. M. (2008). Generational Differences in Psychological Traits and Their Impact on Workplace. *Journal of Managerial Psychology*, 23(8), 862–877. doi:10.1108/02683940810904367
- Twenge, J. M., & Campbell, S. M. (2009). *The Narcissism Epidemic – Living in the Age of Entitlement*. Free Press, Simon and Schuster.

U.S. Census Clock. (2015). Retrieved from <http://www.census.gov/popclock/>

Wall Street Journal. (2016). Retrieved from <http://blogs.wsj.com/digits/2015/11/27/facebook-expands-parental-leave-to-match-zuckerberg-plans>

Ward, A. (1999). Convenience Food: Space and Timing. *British Food Journal*, 101(7), 518–527. doi:10.1108/00070709910279018

Zacher, H. (2015). Using Lifespan Developmental Theory and Methods as a Viable Alternative to the Study of Generational Differences at Work. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 8(3), 342–346. doi:10.1017/iop.2015.47

ADDITIONAL READING

Braskamp, L. A., Trautvetter, L. C., & Ward, K. (2016). *Putting Students First: How Colleges Develop Students Purposefully*. John Wiley & Sons.

Coomes, M. D., & DeBard, R. (2011). *Serving the Millennial Generation: New Directions for Student Services*. John Wiley & Sons.

Espinoza, C., & Ukleja, M. (2016). *Managing the Millennials: Discover the Core Competencies for Managing Today's Workforce* (2nd ed.). John Wiley & Sons.

Harvey, E., & Clark, S. (2016). *Millennials vs. Boomers- Listen, Learn and Succeed Together*. Simple Truths, an Imprint of Sourcebooks Inc.

Knowlton, D. S., & Hagopian, K. J. (2013). *From Entitlement to Engagement: Affirming Millennial Students' Egos in the Higher Education Classroom: New Directions for Teaching and Learning*. John Wiley & Sons.

Pletka, B. (2007). *Educating the Net Generation: How to Engage Students in the 21st Century*. Santa Monica Press.

Rainer, T. S., & Rainer, J. (2011). *The Millennials*. B & H Publishing Group.

Tulgan, B. (2015). *“Not Everyone Gets a Trophy- How to Manage Millennials”*, (2015) (2nd ed.). John Wiley & Sons.

KEY TERMS AND DEFINITIONS

Critical Thinking: Critical thinking is a combination of skills including skepticism (questioning everything) evidence-based reasoning and not emotional response, self-awareness about assumptions and biases, and open-mindedness about alternative explanations.

Cyber -Loafing: A type of shirking behavior at work that involves using the Internet at work for personal use while pretending to do legitimate work.

Generational Cohorts: The entire adult population can be viewed as four generational cohorts – Silent (1928-1945), Baby Boomer (1946-1964), X (1965-1980), Y (1981-1997). The numbers in the parenthesis represent the range of birth years.

Millennials: Millennials represent people who were born between 1981 and 1997. In 2017, they are 20 to 36 years old. The Millennials are also called “Y” generation, digital natives, “net generation”, Trophy generation or “me generation”.

Work-Life Balance: A concept that promotes a balance between work and leisure, career development and lifestyle activities including family, health, spiritual development and pleasure.

Category I

Industrial Engineering and Informatics

Cuckoo Search Algorithm for Solving Real Industrial Multi-Objective Scheduling Problems



Mariappan Kadarkarainadar Marichelvam
Mepco Schlenk Engineering College, India

Mariappan Geetha
Kamaraj College of Engineering and Technology, India

INTRODUCTION

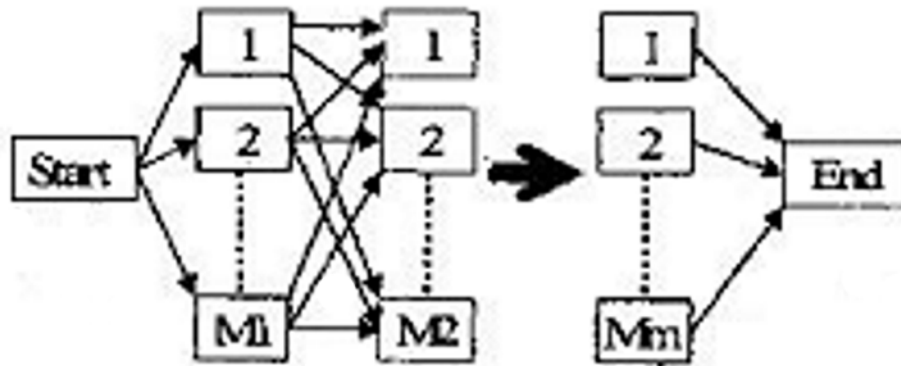
Scheduling is an essential task in our day-to-day life that helps us to shape up our daily activities. The arrival and the departure of airplanes have to be scheduled in an airport. The class hours and the examinations are scheduled in the schools and Universities. Schedules are prepared in industries too. Today, industries need meticulous planning and scheduling to meet the customer demands. Due to globalization and liberalization the attributes of the customers have changed. Hence, industries ought to satisfy them by improving the quality, reducing the price and despatching the goods on time. Scheduling is one of the most important decision making processes. Scheduling is defined as a process of allocating resources over time to perform the assigned tasks effectively (Baker & Trietsch 2009). The machines, equipment, facilities, computers and operators are the important resources in all organizations. Effective scheduling leads to improve the productivity, reduce the inventory, improve the production efficiency, minimize the production time and cost and hence increase the efficiency of the production system. Different types of scheduling environments were addressed by Pinedo (1995). Among them the hybrid flow shop (HFS) environment plays a vital role as many industries resemble it. The HFS has also been called as a flexible flowshop, multiprocessor flow shop, flexible flow line, flow shop with multiple processors or a flow shop with

parallel machines by Ribas et al. (2010). Different operations are performed in different machines in a simple flow shop. The HFS consists of a set of production stages in which each stage has multiple parallel machines. An HFS consists of both the flow shop and parallel machine environments. The parallel machine scheduling system involves the scheduling of a set of immediately available jobs, each on one of the parallel machines. The simple flow shop scheduling system is described as the sequencing of a set of immediately available jobs through each of the ordered work centers. There are two or more work centers in this system but only one machine at each work center. In the HFS, some stages may have only one machine. But, at least one stage should have two or more parallel machines. The machines may be identical, uniform or non-uniform. Some of the jobs may skip some of the stages in the HFS environment. The jobs flow in unidirectional in the HFS environment. Though many researchers have addressed the HFS scheduling problems for more than 40 years, only a few researchers have addressed the real-industrial scheduling problems with multiple objectives which are conflicting naturally with each other. Hence, in this paper a bi-objective HFS scheduling problem is considered. The objective is to minimize the makespan and mean flow time. The layout of an M-stage hybrid flow shop environment is given in figure 1.

Hoogeveen et al. (1996) proved that a two-stage hybrid flow shop scheduling problem is NP-hard

DOI: 10.4018/978-1-5225-2255-3.ch379

Figure 1. Layout of the hybrid flow shop environment



in the strong sense even if there is only one machine on the first stage and two machines on the second stage. Hence, we cannot find the optimal solutions for these problems in a reasonable time. As the hybrid flow shop scheduling problems are NP-hard problems, the problems cannot be solved by exact algorithms. Researchers proposed many heuristics and meta-heuristics to solve the hybrid flow shop scheduling problems. Cuckoo search algorithm is a recently developed meta-heuristic algorithm. In this chapter, the cuckoo search algorithm is proposed to solve the multi-objective hybrid flow shop scheduling problems.

BACKGROUND

Hybrid flow shop scheduling problem was first proposed by Arthanari and Ramamurthy (1971). Researchers have developed many heuristics and meta-heuristics to solve such problems and to obtain optimal or near optimal solutions with considerably less computational time. Lee and Vairaktarakis (1994) proposed heuristics to minimize makespan for hybrid flow shop scheduling problems. Riane et al. (1998) proposed efficient heuristics to minimize makespan for a three-stage flow shop problem. Brah and Loo (1999) proposed a heuristic for flow shop scheduling problems with multiple processors. Oğuz et al. (2003) also proposed some heuristics to solve multiprocessor

task scheduling in a two-stage flow shop scheduling problems.

Recently, many meta-heuristics have been widely applied for solving the hybrid flow shop scheduling problems. Engin and Döyen (2004) proposed an artificial immune system algorithm for solving the hybrid flow shop scheduling problems to minimize the makespan. Yang et al. (2004) applied the tabu search simulation optimization algorithm to solve the multiprocessor flow shop scheduling problems. They applied the algorithm to solve the scheduling problem of a ceramic capacitor manufacturing company. Oğuz and Ercan (2005) have presented a genetic algorithm for the hybrid flow shop scheduling problems with multiprocessor tasks. They used a local search algorithm as a decode method to obtain the objective function value. They also proposed a new crossover operator in their work. Ruiz and Maroto (2006) have also addressed a genetic algorithm that consists of various new crossover operators to minimize makespan for the hybrid flow shop scheduling problems. The sequence dependent setup time and machine eligibility were considered by them. Ying and Lin (2006) developed an ant colony system (ACS) approach to solve the hybrid flow shop scheduling problems. Alaykíran et al. (2007) also proposed an improved ant colony optimization algorithm to minimize makespan for multistage hybrid flow shop scheduling problems. Jungwattanakit et al.

(2008) formulated a 0–1 mixed integer program for hybrid flow shop scheduling problems with unrelated parallel machines. They have developed a polynomial heuristic and also a genetic algorithm to minimize makespan and total number of tardy jobs. Tseng and Liao (2008) proposed a particle swarm optimization algorithm to solve the hybrid flow shop scheduling problems to minimize makespan. Jungwattanakit et al. (2009) compared the performance of three different meta-heuristic algorithms (genetic algorithm, tabu search and simulated annealing) to minimize the convex sum of makespan and the number of tardy jobs for flexible flow shop problems with unrelated parallel machines. Jouglet et al. (2009) proposed a memetic algorithm to minimize makespan for the hybrid flow shop scheduling problems. They used a constraint programming based branch and bound algorithm as the local search in their work. Kahraman et al. (2010) solved the multiprocessor task scheduling in multistage hybrid flowshop using a parallel greedy algorithm approach. Khalouli et al. (2010) suggested an ant colony optimization algorithm for the hybrid flow shop scheduling problems to minimize the sum of the total earliness and tardiness. Rashidi et al. (2010) presented an improved multi objective parallel genetic algorithm for hybrid flow shop scheduling with unrelated parallel machines to simultaneously minimize the makespan and the maximum tardiness. Ruiz and Vázquez-Rodríguez (2010) presented a detailed literature review on exact, heuristics and meta-heuristics methods that have been proposed for the hybrid flow shop scheduling problems. They reported the different industries that resemble the hybrid flow shop scheduling environment and pointed out that only a very small percentage of works concentrated on real industrial scheduling problems. An efficient genetic algorithm was developed by Engin et al. (2011) to solve the hybrid flow shop scheduling with multiprocessor task problems. Behnamian and Ghomi (2011) solved the hybrid flow shop scheduling problems to minimize makespan and total resource allocation costs using genetic al-

gorithm hybridized with variable neighbourhood search. The machine and resource dependent processing time was considered by them. They have developed some random problems and compare their results. Wang et al. (2011) proposed a simulated annealing algorithm for the hybrid flow shop scheduling problems to minimize the makespan. Marichelvam and Prabaharan (2012) applied the bat algorithm for solving the realistic hybrid flowshop scheduling problems to minimize makespan and mean flow time. Mousavi et al. (2013) also proposed a simulated annealing algorithm hybridized with some local search methods to minimize the makespan and total tardiness on a hybrid flowshop environment. Marichelvam et al. (2013) solved the hybrid flow shop scheduling problems using the bat algorithm. Marichelvam et al. (2014) also developed a discrete firefly algorithm for the multi-objective hybrid flowshop scheduling problems. Makespan and flow time are the objective functions considered by them. Marichelvam and Prabaharan (2014) developed an improved hybrid genetic scatter search (IHGSS) algorithm for multistage hybrid flow shop scheduling problems with missing operations. They hybridized the genetic algorithm and scatter search algorithm to minimize the makespan. Marichelvam and Geetha (2014) addressed the tri-objective multistage hybrid flow shop scheduling problems. They solved the problems using the discrete firefly algorithm.

Cuckoo search algorithm is one of the recently developed population-based meta-heuristic algorithms, developed by Yang and Deb (2009). Layeb (2011) applied the quantum inspired cuckoo search algorithm for solving the knapsack problems. Valian et al. (2011) developed an improved cuckoo search algorithm for solving the global optimization problems. Gherboudj et al. (2012) proposed a discrete binary version of cuckoo search algorithm for solving the 0-1 knapsack problems. Layeb and Boussalia (2012) applied a novel quantum inspired cuckoo search algorithm for bin packing problem. Marichelvam (2012) proposed an improved hybrid Cuckoo Search metaheuristics algorithm for permutation flow shop scheduling problems

to minimize the makespan. Basu and Chowdhury (2013) proposed the cuckoo search algorithm for solving the economic dispatch problems. Burnwal and Deb (2013) solved the scheduling optimization problems of flexible manufacturing system using the cuckoo search-based approach. Civicioglu and Besdok (2013) presented a conceptual comparison of the cuckoo search algorithm, particle swarm optimization, differential evolution and artificial bee colony algorithms. Gandomi et al. (2013) solved the structural optimization problems using the cuckoo search algorithm. Kaveh and Bakhshpoori (2013) steel frame design optimization problems using the cuckoo search algorithm. Ouyang et al. (2013) developed a novel discrete cuckoo search algorithm for spherical traveling salesman problem. Vo et al. (2013) applied the cuckoo search algorithm for non-convex economic dispatch problems. Yang and Deb (2013) developed a multi-objective cuckoo search for design optimization problems. Marichelvam et al. (2014) proposed an improved cuckoo search algorithm for solving the hybrid flow shop scheduling problems to minimize makespan. A Discrete cuckoo search algorithm was proposed by Ouaarab et al. (2014) for solving the travelling salesman problem. Yang and Deb (2014) presented the recent advances and applications of the cuckoo search algorithm.

From the above literature review, one can easily conclude that the applications of the cuckoo search algorithm to solve the multi-objective scheduling problems require further extensive studies. Hence, in this paper the cuckoo search algorithm is proposed to minimize makespan and flow time for multistage hybrid flow shop scheduling problems.

MAIN FOCUS

This chapter considers the hybrid flow shop scheduling problems. The objective is to minimize the makespan and mean flow time. Makespan is defined as the completion time of the last job in the production system. Makespan is a performance measure used to improve the efficiency of the production system. Mean flow time is defined as

the average time spent by the jobs in the production system. Reduction in mean flow time will result the reduction in work-in-process inventory. It also reduces the average response time. Hence, minimization of makespan and mean flow time are considered in this chapter. Marichelvam et al. (2014) proposed the mathematical model for the hybrid flow shop scheduling problems to minimize the weighted sum of makespan and mean flow time. The same mathematical model for the objective function of minimization of weighted sum of makespan and mean flow time is considered in this chapter. They considered several assumptions. The same assumptions are considered in this chapter also. The assumptions are:

- All n jobs are available at the beginning of scheduling.
- Each stage has infinite storage capacity.
- One machine can process only one job at a time.
- One job can be processed by only one machine at any time.
- For all the jobs, the processing times at each stage are known in advance and deterministic.
- Job set-up times are sequence-independent and are included in the job processing time of the jobs at the corresponding stage.
- Travel time between consecutive stages is negligible.
- Preemption is not allowed.

CUCKOO SEARCH ALGORITHM

Cuckoo search (CS) algorithm is a new nature-inspired metaheuristic algorithm developed by Yang and Deb (2009). Cuckoo search algorithm was inspired by the obligate brood parasitic behavior of some cuckoo species in combination with the Lévy flight behavior of some birds and fruit flies in nature. The breeding behaviour and the Lévy flights will be discussed in the following sections.

CUCKOO BREEDING BEHAVIOUR

Some of the cuckoo species lay their eggs in the nests of other host birds. The cuckoos often select the recently spawned nests instinctly. They may remove others eggs to increase the hatching probability of their own eggs. Some host birds can engage direct conflict with the intruding cuckoos. If a host bird discovers the eggs are not their owns, they will either throw these alien eggs away or simply abandon its nest and build a new nest elsewhere. Some cuckoos have evolved in such a way that female parasitic cuckoos are often very specialized in the mimicry in colour and pattern of the eggs of a few chosen host birds. This will reduce the probability of their eggs being abandoned. This also increases their reproductivity. Furthermore, the timing of egg-laying of some cuckoos is also amazing. The cuckoos often choose a nest where the host bird just laid its own eggs. In general, the cuckoo eggs hatch slightly earlier than their host eggs. Once the first cuckoo chick is hatched, the first instinct action it will take is to evict the host eggs by blindly propelling the eggs out of the nest. This will increase the cuckoo chick's share of food provided by its host bird. Moreover, a cuckoo chick can also mimic the call of host chicks to gain access to more feeding opportunity.

RULES OF CUCKOO SEARCH ALGORITHM

Based on the brooding behaviour of cuckoos, the basic steps of cuckoo search are described using the following three idealized rules/approximations:

1. Each cuckoo lays one egg (solution) at a time, and dumps its egg in a randomly chosen nest. That is, an egg represents a solution. As there is one egg in one nest, then it can be assumed that an egg is equivalent to a nest and a solution.
2. The best nests with high quality eggs/solutions will carry out to the next generation.

Here is the best means the solution with the best (minimum) objective values (for minimization problems).

3. The egg laid by a cuckoo can be discovered by the host bird with a probability P_a and a nest will then be built. That is to say, a fraction P_a of the n nests being replaced by new nests (with new random solutions at new locations).

For minimization problems the quality or fitness function value may be the reciprocal of the objective function. Each egg in a nest represents a solution and the cuckoo egg represents a new solution. Therefore, there is no difference between an egg, a nest and solution. The pseudo code of the cuckoo search algorithm is presented below.

```

Start
Objective function  $f(x)$ ,  $x = (x_1, x_2, \dots, x_d)$ 
Generate initial population of  $n$  host nests  $x_i$  ( $i = 1, 2, \dots, n$ )
While ( $t < \text{MaxGeneration}$ )
  Get a cuckoo randomly (say  $a$ ) by Lévy flights
  evaluate its quality/fitness  $F_a$  [proportional to  $f(x)$ ]
  Choose a nest among  $n$  (say  $b$ ) randomly
  if ( $F_a$  is better than  $F_b$ ),
    replace solution  $b$  by the new solution;
  end
  A fraction ( $P_a$ ) of worse nests is abandoned and new ones are built;
  Keep the best solutions (or nests with quality solutions);
  Rank the solutions and find the current best
end while
Post process results and visualization
End
  
```


When generating new solutions $x_a^{(t+1)}$ for a^{th} cuckoo a Lévy flight is performed using the equation (1).

$$x_a^{t+1} = x_a^t + \alpha s \otimes H(p_a - \varepsilon) \otimes (x_j^t - x_k^t) \quad (1)$$

where x_j^t and x_k^t are two different solutions selected randomly by random permutation, $H(u)$ is a Heaviside function, ε is a random number drawn from a uniform distribution, and S is the step size. On the other hand, the global random walk is carried out by using Lévy flights

$$x_i^{t+1} = x_i^t + \alpha L(s, \lambda), L(s, \lambda) = \frac{\lambda \Gamma(\lambda) \sin(\pi \lambda / 2)}{\pi} \frac{1}{s^{1+\lambda}}, (s \gg s_0 > 0). \quad (2)$$

Lévy flights essentially provide a random walk while their random steps are drawn from a Lévy distribution for large steps ($1 < \lambda \leq 3$). Here the step size vector L is drawn from the power-law distribution as given in equation (2).

LÉVY FLIGHTS

In nature, animals search for food in a random manner. In general, the foraging path of any animal is effectively a random walk because the next move is based on the current location and the transition probability to the next location. The direction of movement depends implicitly on a probability which can be modeled mathematically. The Levy flights model involve a lot of small steps, interspersed with occasional very large excursions. Yang and Deb (2013) have pointed out that the Lévy flights are a random walk whose step length is drawn from Lévy distribution. The foraging path is very important as the stopping points of a Levy flight are fractal (scale invariant is the main point here), and in complex ecosystems the distribution of food is also fractal. (i.e. there are some large

areas without food). To avoid spending too much time in such unproductive areas, animals need to develop search strategies that generate a fractal distribution of stopping points. Levy flights have this property. The turning points and the trajectory of the Lévy flights are presented in Figure 2. The cuckoo search algorithm consists of three parameters P_a , α , and λ . Among them the parameters P_a and α are very important to obtain better solutions. In the literature, these parameters of the cuckoo search algorithm are kept constant. But, in this thesis work, the parameters P_a , α , and λ are not kept constant. The parameters of P_a and α are determined by the design of experiments. The values used in the simulations are based on extensive parametric studies and the parameters would be $n=15$ to 40 , $\lambda=1.5$, and $\alpha=0.01 \cdot \text{abs}(\text{Ub}-\text{Lb})$, where Ub and Lb are the upper bound and lower bound, respectively, of the problem of interested.

SOLUTIONS AND RECOMMENDATIONS

In order to validate the performance of the proposed algorithm a real industrial scheduling problem is considered in this chapter. The scheduling problem of a leading furniture manufacturing is considered in this chapter. The furniture manufacturing company is one of the leading industries in Tamilnadu, India. The layout of the collaborative company is depicted in Figure 3. The number of stages are five and the number of machines in each stage are presented in Table 1 and the processing time of jobs is presented in Table 2.

To validate the performance of the proposed algorithm the results are compared with many other algorithms addressed in the literature. Mean relative deviation index proposed by Marichelvam et al. (2014) is used in this paper as a performance measure in this chapter. MRDI is one of the most important performance measures used in the scheduling literature. Lower MRDI will be the indication of better performance of the proposed algorithm. The MRDI is calculated using the

Figure 2. Turning points and trajectory of the Lévy flights

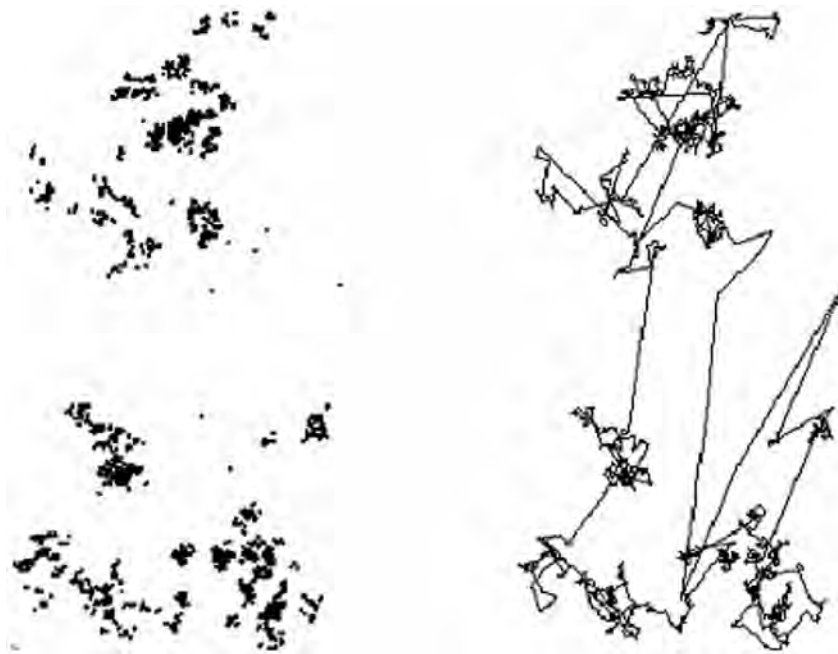


Figure 3. Layout of the case study company

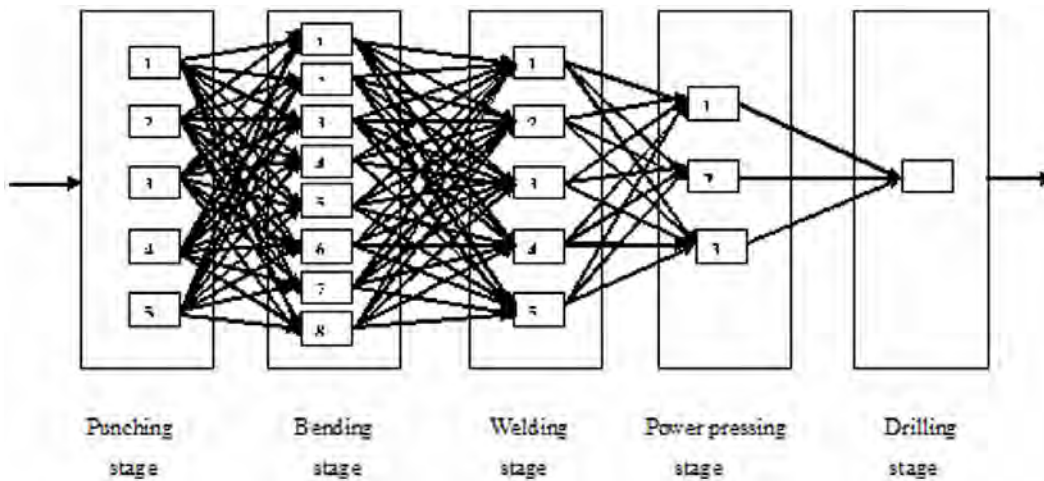


Table 1. Number of machines in each stage

Sl. No.	Stage	Number of Machines
1	Punching	5
2	Bending	8
3	Welding	5
4	Power pressing	3
5	Drilling	1

Table 2. Processing time of jobs (in seconds)

Stages → Jobs ↓	Punching	Bending	Welding	Power Pressing	Drilling
1.	39	28	0	0	0
2.	61	66	110	0	0
3.	10	17	52	0	0
4.	22	0	0	0	0
5.	35	60	0	0	0
6.	34	66	70	0	0
7.	42	0	0	0	0
8.	42	0	0	50	0
9.	34	250	30	0	0
10.	12	18	25	0	0
11.	0	0	0	10	0
12.	0	0	0	20	0
13.	52	90	0	0	0
14.	0	0	0	20	0
15.	0	0	0	42	0
16.	12	22	0	0	32
17.	0	0	0	8	0
18.	0	0	0	6	0
19.	0	0	0	10	0
20.	0	0	0	6	0

equation (3). The MRDI is comparison of different algorithms is presented in Table 3.

$$MRDI = \sum_{l=1}^R \frac{(Z^* - Z_{meta})}{Z^*} \times 100 / R \quad (3)$$

Where,

Z^* = best objective function value

Z_{meta} = objective function value obtained by the different algorithms

R = number of runs (20)

From the result table, it is concluded easily that the performance of the proposed cuckoo search algorithm is better than many other algorithms

addressed in the literature for the multi-objective real industrial scheduling problems. This indicates the effectiveness of the proposed algorithm.

Table 3. MRDI Comparison of different algorithms for the test problems

Sl. No.	Algorithms	MRDI
1	Ant colony optimization algorithm	3.91
2	Cuckoo search algorithm	0
3	Genetic algorithm	2.23
4	Particle swarm optimization algorithm	1.56
5	Simulated annealing algorithm	2.77



FUTURE RESEARCH DIRECTIONS

In this chapter, the multi-objective hybrid flow shop scheduling problems are considered. However, many assumptions are made in this chapter. For instance, the setup time is not addressed in this chapter. Consideration of setup time and transportation time is a future research scope of this work. Consideration of due date related criteria such as earliness and tardiness would be another interesting scope of this research. The cuckoo search algorithm may also be applied for other optimization problems. The proposed algorithm might be hybridized with other algorithms.

CONCLUSION

The multi-objective hybrid flow shop scheduling problem is addressed in this chapter. A recently developed cuckoo search algorithm is proposed to minimize makespan and mean flow time for the hybrid flow shop scheduling problems. The algorithm is validated with a real industrial data. The performance of the proposed cuckoo search algorithm is compared with other algorithms such as ant colony optimization, genetic algorithm, particle swarm optimization and simulated annealing and proved to be better.

REFERENCES

- Alaykran, K., Engin, O., & Dyen, A. (2007). Using ant colony optimization to solve hybrid flow shop scheduling problems. *International Journal of Advanced Manufacturing Technology*, 35(5-6), 541–550. doi:10.1007/s00170-007-1048-2
- Arthanari, T. S., & Ramamurthy, K. G. (1971). An extension of two machines sequencing problem. *Opsearch*, 8(1), 10–22.
- Baker, K. R., & Trietsch, D. (2009). *Principles of Sequencing and Scheduling*. John Wiley & Sons, Inc. doi:10.1002/9780470451793
- Basu, M., & Chowdhury, A. (2013). Cuckoo search algorithm for economic dispatch. *Energy*, 60, 99–108. doi:10.1016/j.energy.2013.07.011
- Behnamian, J., & Ghomi, S. F. (2011). Hybrid flowshop scheduling with machine and resource-dependent processing times. *Applied Mathematical Modelling*, 35(3), 1107–1123. doi:10.1016/j.apm.2010.07.057
- Brah, S. A., & Loo, L. L. (1999). Heuristics for scheduling in a flow shop with multiple processors. *European Journal of Operational Research*, 113(1), 113–122. doi:10.1016/S0377-2217(97)00423-2
- Burnwal, S., & Deb, S. (2013). Scheduling optimization of flexible manufacturing system using cuckoo search-based approach. *International Journal of Advanced Manufacturing Technology*, 64(5-8), 951–959. doi:10.1007/s00170-012-4061-z
- Civicioglu, P., & Besdok, E. (2013). A conceptual comparison of the Cuckoo-search, particle swarm optimization, differential evolution and artificial bee colony algorithms. *Artificial Intelligence Review*, 39(4), 315–346. doi:10.1007/s10462-011-9276-0
- Engin, O., Ceran, G., & Yilmaz, M. K. (2011). An efficient genetic algorithm for hybrid flow shop scheduling with multiprocessor task problems. *Applied Soft Computing*, 11(3), 3056–3065. doi:10.1016/j.asoc.2010.12.006
- Engin, O., & Dyen, A. (2004). A new approach to solve hybrid flow shop scheduling problems by artificial immune system. *Future Generation Computer Systems*, 20(6), 1083–1095. doi:10.1016/j.future.2004.03.014
- Gandomi, A. H., Yang, X. S., & Alavi, A. H. (2013). Cuckoo search algorithm: A metaheuristic approach to solve structural optimization problems. *Engineering with Computers*, 29(1), 17–35. doi:10.1007/s00366-011-0241-y

- Gherboudj, A., Layeb, A., & Chikhi, S. (2012). Solving 01 knapsack problems by a discrete binary version of cuckoo search algorithm. *International Journal of Bio-inspired Computation*, 4(4), 229–236. doi:10.1504/IJBIC.2012.048063
- Hoogeveen, J. A., Lenstra, J. K., & Veltman, B. (1996). Minimizing the makespan in a multiprocessor flow shop is strongly NP-hard. *European Journal of Operational Research*, 89(1), 172–175. doi:10.1016/S0377-2217(96)90070-3
- Jouglet, A., Oğuz, C., & Sevaux, M. (2009). Hybrid flow-shop: A memetic algorithm using constraint-based scheduling for efficient search. *Journal of Mathematical Modelling and Algorithms*, 8(3), 271–292. doi:10.1007/s10852-008-9101-1
- Jungwattanakit, J., Reodecha, M., Chaovalitwongse, P., & Werner, F. (2008). Algorithms for flexible flow shop problems with unrelated parallel machines, setup times, and dual criteria. *International Journal of Advanced Manufacturing Technology*, 37(3-4), 354–370. doi:10.1007/s00170-007-0977-0
- Jungwattanakit, J., Reodecha, M., Chaovalitwongse, P., & Werner, F. (2009). A comparison of scheduling algorithms for flexible flow shop problems with unrelated parallel machines, setup times, and dual criteria. *Computers & Operations Research*, 36(2), 358–378. doi:10.1016/j.cor.2007.10.004
- Kahraman, C., Engin, O., Kaya, İ., & Öztürk, R. E. (2010). Multiprocessor task scheduling in multistage hybrid flow-shops: A parallel greedy algorithm approach. *Applied Soft Computing*, 10(4), 1293–1300. doi:10.1016/j.asoc.2010.03.008
- Kaveh, A., & Bakhshpoori, T. (2013). Optimum design of steel frames using cuckoo search algorithm with lévy flights. *Structural Design of Tall and Special Buildings*, 22(13), 1023–1036. doi:10.1002/tal.754
- Khalouli, S., Ghedjati, F., & Hamzaoui, A. (2010). A meta-heuristic approach to solve a JIT scheduling problem in hybrid flow shop. *Engineering Applications of Artificial Intelligence*, 23(5), 765–771. doi:10.1016/j.engappai.2010.01.008
- Layeb, A. (2011). A novel quantum inspired cuckoo search for knapsack problems. *International Journal of Bio-inspired Computation*, 3(5), 297–305. doi:10.1504/IJBIC.2011.042260
- Layeb, A., & Boussalia, S. R. (2012). A novel quantum inspired cuckoo search algorithm for bin packing problem. *International Journal of Information Technology and Computer Science*, 4(5), 58–67. doi:10.5815/ijitcs.2012.05.08
- Lee, C. Y., & Vairaktarakis, G. L. (1994). Minimizing makespan in hybrid flowshops. *Operations Research Letters*, 16(3), 149–158. doi:10.1016/0167-6377(94)90026-4
- Marichelvam, M. K. (2012). An improved hybrid Cuckoo Search (IHCS) metaheuristics algorithm for permutation flow shop scheduling problems. *International Journal of Bio-inspired Computation*, 4(4), 200–205. doi:10.1504/IJBIC.2012.048061
- Marichelvam, M. K., & Geetha, M. (2014). Solving tri-objective multistage hybrid flow shop scheduling problems using a discrete firefly algorithm. *International Journal of Intelligent Engineering Informatics*, 2(4), 284–303. doi:10.1504/IJIEI.2014.067190
- Marichelvam, M. K., & Prabakaran, T. (2012). A bat algorithm for realistic hybrid flowshop scheduling problems to minimize makespan and mean flow time. *ICTACT Journal on Soft Computing*, 3(1), 428–433. doi:10.21917/ijsc.2012.0066
- Marichelvam, M. K., & Prabakaran, T. (2014). Performance evaluation of an improved hybrid genetic scatter search (IHGSS) algorithm for multistage hybrid flow shop scheduling problems with missing operations. *International Journal of Industrial and Systems Engineering*, 16(1), 120–141. doi:10.1504/IJISE.2014.057946



- Marichelvam, M. K., Prabakaran, T., & Yang, X. S. (2014). Improved cuckoo search algorithm for hybrid flow shop scheduling problems to minimize makespan. *Applied Soft Computing*, 19, 93–101. doi:10.1016/j.asoc.2014.02.005
- Marichelvam, M. K., Prabakaran, T., & Yang, X. S. (2014). A discrete firefly algorithm for the multi-objective hybrid flowshop scheduling problems. *IEEE Transactions on Evolutionary Computation*, 18(2), 301–305. doi:10.1109/TEVC.2013.2240304
- Marichelvam, M. K., Prabakaran, T., Yang, X. S., & Geetha, M. (2013). Solving hybrid flow shop scheduling problems using bat algorithm. *International Journal of Logistics Economics and Globalization*, 5(1), 15–29. doi:10.1504/IJLEG.2013.054428
- Mousavi, S. M., Zandieh, M., & Yazdani, M. (2013). A simulated annealing/local search to minimize the makespan and total tardiness on a hybrid flowshop. *International Journal of Advanced Manufacturing Technology*, 64(1-4), 369–388. doi:10.1007/s00170-012-4031-5
- Oğuz, C., & Ercan, M. F. (2005). A genetic algorithm for hybrid flow-shop scheduling with multiprocessor tasks. *Journal of Scheduling*, 8(4), 323–351. doi:10.1007/s10951-005-1640-y
- Oğuz, C., Ercan, M. F., Cheng, T. E., & Fung, Y. F. (2003). Heuristic algorithms for multiprocessor task scheduling in a two-stage hybrid flow-shop. *European Journal of Operational Research*, 149(2), 390–403. doi:10.1016/S0377-2217(02)00766-X
- Ouaarab, A., Ahiod, B., & Yang, X. S. (2014). Discrete cuckoo search algorithm for the traveling salesman problem. *Neural Computing & Applications*, 24(7-8), 1659–1669. doi:10.1007/s00521-013-1402-2
- Ouyang, X., Zhou, Y., Luo, Q., & Chen, H. (2013). A novel discrete cuckoo search algorithm for spherical traveling salesman problem. *Applications of Mathematics*, 7(2), 777–784.
- Pinedo, M. (1995). *Scheduling: theory, algorithms, and systems*. Englewood Cliffs, NJ: Prentice-Hall.
- Rashidi, E., Jahandar, M., & Zandieh, M. (2010). An improved hybrid multi-objective parallel genetic algorithm for hybrid flow shop scheduling with unrelated parallel machines. *International Journal of Advanced Manufacturing Technology*, 49(9-12), 1129–1139. doi:10.1007/s00170-009-2475-z
- Riane, F., Artiba, A., & Elmaghraby, S. E. (1998). A hybrid three-stage flowshop problem: Efficient heuristics to minimize makespan. *European Journal of Operational Research*, 109(2), 321–329. doi:10.1016/S0377-2217(98)00060-5
- Ribas, I., Leisten, R., & Framiñan, J. M. (2010). Review and classification of hybrid flow shop scheduling problems from a production system and a solutions procedure perspective. *Computers & Operations Research*, 37(8), 1439–1454. doi:10.1016/j.cor.2009.11.001
- Ruiz, R., & Maroto, C. (2006). A genetic algorithm for hybrid flowshops with sequence dependent setup times and machine eligibility. *European Journal of Operational Research*, 169(3), 781–800. doi:10.1016/j.ejor.2004.06.038
- Ruiz, R., & Vázquez-Rodríguez, J. A. (2010). The hybrid flow shop scheduling problem. *European Journal of Operational Research*, 205(1), 1–18. doi:10.1016/j.ejor.2009.09.024
- Tseng, C. T., & Liao, C. J. (2008). A particle swarm optimization algorithm for hybrid flow-shop scheduling with multiprocessor tasks. *International Journal of Production Research*, 46(17), 4655–4670. doi:10.1080/00207540701294627
- Valian, E., Mohanna, S., & Tavakoli, S. (2011). Improved cuckoo search algorithm for global optimization. *International Journal of Communications and Information Technology*, 1(1), 31–44.

Vo, D. N., Schegner, P., & Ongsakul, W. (2013). Cuckoo search algorithm for non-convex economic dispatch. Generation. *Transmission & Distribution, IET*, 7(6), 645–654. doi:10.1049/iet-gtd.2012.0142

Wang, H. M., Chou, F. D., & Wu, F. C. (2011). A simulated annealing for hybrid flow shop scheduling with multiprocessor tasks to minimize makespan. *International Journal of Advanced Manufacturing Technology*, 53(5-8), 761–776. doi:10.1007/s00170-010-2868-z

Yang, T., Kuo, Y., & Chang, I. (2004). Tabu-search simulation optimization approach for flow - shop scheduling with multiple processors-a case study. *International Journal of Production Research*, 42(19), 4015–4030. doi:10.1080/00207540410001699381

Yang, X. S., & Deb, S. (2009, December). Cuckoo search via Lévy flights. In *Nature & Biologically Inspired Computing, 2009. NaBIC 2009. World Congress on* (pp. 210-214). IEEE.

Yang, X. S., & Deb, S. (2013). Multiobjective cuckoo search for design optimization. *Computers & Operations Research*, 40(6), 1616–1624. doi:10.1016/j.cor.2011.09.026

Yang, X. S., & Deb, S. (2014). Cuckoo search: Recent advances and applications. *Neural Computing & Applications*, 24(1), 169–174. doi:10.1007/s00521-013-1367-1

Ying, K. C., & Lin, S. W. (2006). Multiprocessor task scheduling in multistage hybrid flow-shops: An ant colony system approach. *International Journal of Production Research*, 44(16), 3161–3177. doi:10.1080/00207540500536939

ADDITIONAL READING

Marichelvam, M. K., & Geetha, M. (2016). A hybrid cuckoo search metaheuristic algorithm for solving single machine total weighted tardiness scheduling problems with sequence dependent setup times. *International Journal of Computational Complexity and Intelligent Algorithms*, 1(1), 23–34. doi:10.1504/IJCCIA.2016.077463

Marichelvam, M. K., & Geetha, M. (2016). Application of novel harmony search algorithm for solving hybrid flow shop scheduling problems to minimise makespan. *International Journal of Industrial and Systems Engineering*, 23(4), 467–481. doi:10.1504/IJISE.2016.077698

Marichelvam, M. K., & Prabakaran, T. (2015). Solving realistic industrial scheduling problems using a multi-objective improved hybrid particle swarm optimisation algorithm. *International Journal of Operational Research*, 23(1), 94–129. doi:10.1504/IJOR.2015.068752

Yang, X. S., & Deb, S. (2010). Engineering optimisation by cuckoo search. *International Journal of Mathematical Modelling and Numerical Optimisation*, 1(4), 330–343. doi:10.1504/IJMMNO.2010.035430

Yildiz, A. R. (2013). Cuckoo search algorithm for the selection of optimal machining parameters in milling operations. *International Journal of Advanced Manufacturing Technology*, 64(1-4), 55–61. doi:10.1007/s00170-012-4013-7

KEY TERMS AND DEFINITIONS

Cuckoo Search Algorithm: A recently developed meta-heuristic algorithm to solve the optimization problems.

Makespan: Makespan is defined as the completion time of the last job to leave the system.

Mean Flow Time: Mean flow time is defined as the average time spent by the jobs in the production system.

NP-Hard Problems: Non – deterministic polynomial time hard problems.

Scheduling: Scheduling is defined as a process of allocating resources over time to perform a collection of tasks.



The Trends and Challenges of 3D Printing

Edna Ho Chu Fang

University of Malaya, Malaysia

Sameer Kumar

University of Malaya, Malaysia

INTRODUCTION

3D printing is the new wave of technology advancement in the world of architecture, design and manufacturing. Also known as rapid prototyping, 3D printing is a type of additive manufacturing technology where a 3D object is created by laying down subsequent layers of material at the mm scale. 3D printers print objects by reading a CAD design file or by scanning an object (Sachs et al., 1992). Today, 3D printing is applied in various industries such as footwear, jewelry, architecture, engineering and construction, aerospace, dental and medical industries, education, consumer products, automotive and industrial design. Some claim that 3D printing will put an end to traditional manufacturing primarily since 3D printing imposes a tool-less process. Product parts can be specifically designed to avoid assembly lines, as well as ensuring maximum utilization of raw materials. In this article, the authors discuss the state-of-the-art of 3D printing its future direction.

BACKGROUND

In May 1980, Dr Kodama from Japan filed the very first patent application for Rapid Prototyping technology. Unfortunately, he did not file the subsequent full patent specification before the one year deadline after the application. Hence, in 1986, Charles (Chuck) Hull filed the first patent for stereolithography apparatus (SLA). He was the first to invent the SLA machine in 1983. After obtaining the patent, he went on to co-found 3D

Systems Corporation, which is one of the largest organizations operating in the 3D printing world today. During the mid-nineties, the 3D printing sector started to diverge into two specific areas. First, there was the high end of 3D printing, which saw the production of complex parts. These applications include the medical, aerospace, jewelry and automotive sectors. Then there was the lower end of the market, which saw a price war among many 3D printer manufacturers, highlighting improvements in speed, accuracy and materials.

In 2007, 3D Systems came up with the first 3D printer which was priced under \$10,000. The first commercial 3D printer was offered for sale in January, 2009. It was based on the RepRap concept, and came in a kit form. Makerbot Industries also developed commercial printers in April of the same year. 2013 saw Stratasys acquiring Makerbot. It was a year of significant growth and consolidation for 3D printing.

Materials for 3D printing were very limited during the early days of the technology. Today, there is an array of different types of materials available for choice.

The first step in 3D printing is to design the 3D digital model using a CAD program or scan the object with a 3D scanner. The model will then be 'sliced' into layers and converted into a printer-readable file. The printing material will be added one layer at a time.

Different materials are suited for different 3D printing technologies. Some 3D printers process powdered materials which utilize a light source to fuse layers of the powder together to make the desired shape. Others process polymer resin

materials and utilize a laser to solidify the resin in ultra thin layers. Another method is the jetting of fine droplets using materials and a binder to fix the layers. One of the most commonly used 3D printing technology is the stereolithography (SLA) technology. This technology utilizes photocuring resins as raw material. New resins that are being developed will combine transparency, heat resistance and toughness.

The second commonly used technology is called 'Fused Deposition Modeling', as invented by Scott Crump, a co-founder of Stratasys Inc. The FDM technology uses thermoplastic resins as raw materials. It is the simplest 3D additive manufacturing technology, in which the thermoplastic resin softens when heat is applied. The third type of technology is called 'Selective Laser Sintering (SLS)'. This builds objects by using a laser to fuse together layers of a mixture of different powdered raw materials. The fourth type of technology is called 'Multi-jet modeling (MJM)'. Objects are built up from the layering of powder through an inkjet-like print head that also sprays a binder solution to glue the required granules together. The raw materials associated with this type of technology are sand mold or nylon resins.

THE TECHNICAL ISSUES OF 3D PRINTING

3D printing has revolutionized our society from providing medical advances; to scalable production of everything from product parts to buildings. There are, however, many issues that accompany this technology. Two main issues of 3D printing are the technical problems and the controversies. This section will be divided into two parts. Part one presents the technical problems of 3D printing, while part two presents the controversies of 3D printing.

As 3D printing is getting more widespread, the issue of its quality is always being questioned. In 3D printing, the quality of the printed object is linked to the printing speed; and the printing

speed is linked to the raw materials' thermoplastic properties. Different raw materials are supposed to be printed at different speeds. To ensure a beautifully printed object, the speed of printing has to be just right – not too slow and not too fast either.

Some of the common printing problems are warping, stringing, gaps in the top layers, under-extrusion, over-extrusion, pillowing, layer-misalignment, elephant foot, etc. Furthermore, the raw materials that can be used for 3D printers are still rather limited. Presently, most of the commercial 3D printed products consist of one single material. The idea of printing electronic goods such as smart phones is already being researched but not yet foreseeable. The tough challenge here is to add different types of materials that fulfill their functionality to make up various parts of an electronic component.

Last but not least, 3D printing requires post-processing. An additive manufacturing (AM) machine cannot add finishing touches to the printed object. It requires a manually intensive process. The printed objects will still have to undergo post-printing processes such as friction-weld, paint, sand, rivet and so on.

The Controversies of 3D Printing

The controversies surrounding 3D printing include infringement of intellectual property laws, fabrication of weapons and drugs for crime purposes, compliance with FDA safety standards and ethical considerations.

Intellectual Property Laws

Online platforms such as GrabCad and Thingiverse provide users with the CAD design files needed for 3D printing. Registered users can download, upload, design and modify a 3D model on these online platforms. Infringement of copyright laws happens through the way users obtain the CAD design files. They either: (1) create an object design file; (2) modify an existing design; (3) scan the object (Mendis & Secchi, 2015). The main ques-



tion here is whether a CAD file can be protected under the copyright law. Firstly, for a design to be eligible for protection, it has to qualify as a literary work. Many arguments are in favor of the notion that a CAD file is not a type of computer program (literary work), hence it is not eligible for protection under the copyright law.

Rideout (2012) asserts that CAD files cannot be categorized as copyrightable software in the USA because they are just “triangular representations of a 3D object”. Unlike computer programs, the CAD files do not control the way 3D printers operate. He justifies that a 3D design file isn’t a literary work and is more likely to be considered under “graphic and sculptural works”. Then again, another issue will be raised if 3D designs were to be considered as sculptural works. Art works have their own set of copyrights. Apparently, a 3D model does not infringe copyright laws if the CAD file is used to create an object for use and not as an artwork. This is showcased in the example of Andrew Ainsworth, a props maker who in the film *Star Wars*, printed the Stormtroopers’ white helmets and justified that it was for use and not as an artwork. The court ruled in favor of Ainsworth.

Whilst the USA stands by the notion that a CAD file cannot be protected under the copyright law, the UK believes that a CAD file could be protected. As for now, the copyright status of CAD design files remains unclear.

Fabrication of Weapons or Drugs for Crime Purposes

3D printing provides users with countless possibilities. As much as it can be used for good purposes, it can also be misused when it falls into the hands of people with ill intentions. Anyone with sufficient modal can own a 3D printer and start printing guns or illegal drugs at the comfort of their homes. They are only limited by the costs and technological capabilities of the printers, as well as the availability of the raw materials.

Ever since Cody Wilson, a law student who became the first man to ever print a gun called

‘the Liberator’ from a 3D printer, firearms enthusiasts are constantly modifying gun designs to come up with more durable guns made from a combination of 3D printed materials and parts taken from traditionally manufactured guns. 3D printed guns pose threats to the society because metal detectors cannot detect them.

Generally, the United States allows its citizens to own guns though each state has its own regulatory laws regarding gun ownership. For example, in California, people with a negative track record or people with mental illness are prohibited from owning firearms. 3D printing enables almost anyone to possess a firearm through self-manufacturing.

The group of people who called for firearms to be banned in the States fear that 3D printed guns might be used for crimes or mass shootings. Even if the government does ban 3D printed guns, the technological advances in firearm printing will not stop there as it is part of the European history and culture to own guns.

While owning a firearm is legal, how about owning a gun design file? The province of New South Wales of Australia recently enacted a law that bans gun design files. The amendment made to the Firearms Act 1996 states that possessing 3D digital blueprints for manufacturing firearms is a crime punishable with a penalty of up to 14 years in jail. Individuals who possess digital gun blueprints for research purposes are legally exempted from the law.

The legal stand of manufacturing own firearms differ amongst countries all over the world. Some countries such as Australia and Japan have banned the possession of gun design files, while the US does not. Currently, making own firearms is legal in the US as long as it is intended for personal use only and not for commercial purposes.

Manufacturing of Illegal or Counterfeit Drugs

As with the printing of weapons, criminals might misuse 3D printing to manufacture illegal or

counterfeit drugs. Professor Lee Cronin and a team of 45 researchers at Glasgow University are researching on additive manufacturing technology that would allow drugs to be produced through 3D printing. It was discovered that bathroom sealant which acts as the “ink” for creating reaction chambers was a suitable raw material for the 3D printer. The printer can inject “chemical inks” that would create more complex molecules through a series of reactions (“3-D Printing: The Potential Implications and Challenges for Law Enforcement”, 2015).

Cronin stated that almost all drugs are made of the similar components of oxygen, carbon, and hydrogen with vegetable oils and paraffin. He foresees that in the future, people will be able to download drug recipes from online platforms and use 3D printers to produce medicines at their homes. When this happens, 3D printed drugs might not just be a possible crime threat, but a social threat as well.

Besides the issue of printing illegal drugs, individuals who want to earn easy money can also manufacture counterfeit drugs that produce similar effects, but have different chemical components from today’s controlled substances. Due to the drugs’ distinct chemical compositions, these counterfeit drugs may not be covered by current laws and would hence be technically legal. They might also be more dangerous than the controlled substances because less research is done to understand their chemical composition and discover the side effects of these drugs.

Compliance With FDA Safety Standards

All food, toys, cosmetic and medical products made with 3D printing technologies must comply with the regulations outlined by the U.S. Food & Drug Administration (FDA). How can we ensure that 3D printed foods fulfill all the necessary safety standards? Despite its promise, 3D printing is still relatively immature when it comes to safety and quality control standards compared to conventional food manufacturing process.

3D printed food is still at its early testing phase as researchers are trying to incorporate nutrition into the printed food. 3D printed food creates a new dilemma for the FDA because the authorities are unsure whether the current GMP and food reviewing procedures are sufficient and applicable to the 3D printing industry, or should new regulations be made, said Claudia Lewis, a partner at Venable LLP.

Currently, 3D printers are capable of printing pastries, desserts and burgers or even personalized food items and can cater to people with special dietary requirements. However, a lot of stakeholders in the 3D food printing industry are not fully aware that food elements have to be regulated. The food products will not get to market unless they get some sort of approval by the FDA. How the FDA decides to regulate the food depends on which category the product belongs to. For example, FDA regulations for beverages depend on whether the beverage is classified as a conventional drink or a dietary supplement. It is required at this point for the FDA to seek to understand more about the food industry and for the industry to ask the agency for guidance.

Ethical Considerations

3D printing has made its presence known in the field of medical science and technology. Through tissue engineering, 3D printing is utilized to build biological substitutes that mimic human bodily tissues and can be used to heal wounds or restore failing organs.

Traditionally, tissue engineering starts by taking a small biopsy from a patient. The cells from the biopsy are isolated, expanded and seeded on a natural or synthetic scaffold. After the cells differentiate and proliferate in the scaffold, the resulting tissue construct is transferred to a bioreactor for it to mature before it can be used for transplantation. Through this approach, researchers have successfully printed the trachea, ear cartilage, urinary bladder and so on. However, this approach is still limited in terms of accuracy and reproductibility that can be achieved as more complex multicellular



structures with vascular network integration has to be built (Lucas & Spiegel, 2015).

The innovation in bioprinting is relatively new and there aren't many similar procedures with which to compare them. There aren't many prior case studies or statistics to be used for weighing the possible contradictions and risks. Patients have to fully trust the work of the biomedical engineers and doctors. Besides, the cost of the procedure will be very high, hence raising the question of medical procedure coverage by health insurance companies. Is this procedure worth the cost and risks when it is still at its early stages of research and implementation?

Still at its early stages of research, it is only ethical for the researchers to consider the patients' safety in using bioprinted tissues or organs. The safety of the patient throughout the healing process is the utmost concern with bioprinting (Lucas & Spiegel, 2015).

Though there are biomedical engineers who have successfully printed various body parts, it will take a considerable amount of years before commercialized printing of compatible organs could take place. Currently, only microtissues can be printed (Varkey & Atala, 2015). The Food and Drug Administration will have to screen through all bioprinted tissues and organs for effectiveness and safety, and also weigh the pros and risks involved.

A number of regulatory laws will need to be cleared before the first lab-printed organ becomes commercially available. Technological advancements in terms of software and advances in medical science and technology are needed before the application of bioprinting can occur in a widespread manner (Varkey & Atala, 2015).

SOLUTIONS AND RECOMMENDATIONS

Technical problems can be easily resolved by creating a 3D printer that is of higher sophistication, is durable and cost-efficient. More raw materials can be sought after and test-run for printing a wider range of objects in the future.

The U.S. government should implement copyright laws for CAD design files specifically on those of firearms. Though it is legal to own a firearm, the government can pass a law that prohibits the manufacturing of firearms for selling purposes. The law should also implement a strict screening of eligible firearm owners before gunsmiths start selling to them. It is reasonable that only sound-minded and good-intentioned folks are capable of handling firearms for logical reasons. The authority of printing drugs should be strictly given to registered medical institutions, practitioners and researchers to ensure quality control. Law enforcement will have to alter its investigation methods for drug crimes.

Furthermore, in terms of 3D printed food, the FDA must ensure that all food printed should adhere to the safety standards. Food printed has to be edible and contains nutrition readily absorbed by the human body.

Lastly, 3D bioprinting should be ethical by taking care of the patient's safety and concern (first and foremost). More tests has to be run in the field of organ printing to ensure maximum compatibility of cells and functioning after an organ transplant.

FUTURE RESEARCH DIRECTIONS

3D printing of hybrid objects is still an under-researched topic as most of the research done is on single-raw material printed products. The future of 3D printing would be promising if objects that require different raw materials for different functioning parts can be printed completely without the need of product assembling.

Another interesting field for research is 3D printed food. Currently, there isn't sufficient research that ensures the safety of 3D printed food in compliance with the regulations outlined by the FDA. Once 3D printed food passes the FDA regulations, they can start to be produced commercially and should garner great interest in the market.

CONCLUSION

3D technology will continue to evolve. Many stakeholders are adopting it due to the benefits it promises. A technology will prove to be useful if it falls into the right hands. Though 3D printing technology is used in weapon manufacturing, it is also being used to improve the lives of mankind. In the future, 3D printing will most probably be used to print human organs. An issue of concern is raised regarding the possibility of printing 3D drugs, through the assembly of chemical compounds at a molecular level. In the future, patients can print medicine on their own. It is only hoped that with regulations, the potential of 3D printing technology can be harvested for good and not for worse.

REFERENCES

- Bose, S., Vahabzadeh, S., & Bandyopadhyay, A. (2013). Bone tissue engineering using 3D printing. *Materials Today*, 16(12), 496–504. doi:10.1016/j.matmod.2013.11.017
- Campbell, T., Williams, C., Ivanova, O., & Garrett, B. (2011). *Could 3D printing change the world. In Technologies, Potential, and Implications of Additive Manufacturing*. Washington, DC: Atlantic Council.
- Connell, J. L., Ritschdorff, E. T., Whiteley, M., & Shear, J. B. (2013). 3D printing of microscopic bacterial communities. *Proceedings of the National Academy of Sciences of the United States of America*, 110(46), 18380–18385. doi:10.1073/pnas.1309729110 PMID:24101503
- Estes, A. C. (2015, January 06). *3D-Printed Guns Are Only Getting Better, and Scarier*. Retrieved April 26, 2016, from <http://gizmodo.com/3d-printed-guns-are-only-getting-better-and-scarier-1677747439>
- Grunewald, S. J. (2015, August 19). *The Shutty Hybrid 3D Printed 9mm Pistol Raises Questions About 3D Printed Gun Control*. Retrieved April 26, 2016, from <https://3dprint.com/89919/shutty-hybrid-3d-printed-pistol/>
- Holliman, K. (2015, April 20). *Innovations in 3D Printed Food Outpacing FDA Regulations - Food Quality & Safety*. Retrieved April 26, 2016, from <http://www.foodqualityandsafety.com/article/innovations-in-3d-printed-food-outpacing-fda-regulations/>
- Journal, M. (2015). *The potential of 3D printing*. Retrieved from http://www.meti.go.jp/english/publications/pdf/journal2013_08a.pdf
- Lucas, C., & Spiegel, M. (2015). *3D printing skin grafts for facial reconstruction*. Academic Press.
- Manyika, J., Chui, M., Bughin, J., Dobbs, R., Bissson, P., & Marrs, A. (2013). *Disruptive technologies: Advances that will transform life, business, and the global economy* (Vol. 12). McKinsey Global Institute New York.
- Mendis, D., & Secchi, D. (2015). *A Legal and Empirical Study of 3D Printing Online Platforms and an Analysis of User Behaviour*. Academic Press.
- Peter, V., & Kinscö, I. (2013). *Developing an evaluation and progress methodology to underpin the intervention logic of the Action Plan to Boost Demand for European Innovations*. EU.
- Petrack, I. J., & Simpson, T. W. (2013). 3D printing disrupts manufacturing: How economies of one create new rules of competition. *Research-Technology Management*, 56(6), 12–16. doi:10.5437/08956308X5606193
- Ratto, M., & Ree, R. (2012). Materializing information: 3D printing and social change. *First Monday*, 17(7). doi:10.5210/fm.v17i7.3968
- Rideout, B. (2012). Printing the impossible triangle: The copyright implications of three-dimensional printing. *The Journal of Business Entrepreneurship & the Law*, 5(1), 6.

Sachs, E., Cima, M., Williams, P., Brancazio, D., & Cornie, J. (1992). Three dimensional printing: Rapid tooling and prototypes directly from a CAD model. *Journal of Engineering for Industry*, 114(4), 481–488. doi:10.1115/1.2900701

Schwartz, C. (2015, March). 3-D Printing: The Potential Implications and Challenges for Law Enforcement. *The Police Chief*, 82, 64–69.

Sissons, A., & Thompson, S. (2012). *Three Dimensional Policy: Why Britain needs a policy framework for 3D printing*. London: The Big Innovation Centre.

Stahl, H. (2015). *3D Printing – Risks and Opportunities*. Retrieved from <http://www.oeko.de/oekodoc/1888/2013-532-en.pdf>

Stratasys. (2015). *Digital Dentistry 3D printing makes digital dentistry happen*. Retrieved from <http://www.stratasys.com/industries/dental>

Varkey, M., & Atala, A. (2015). Organ Bioprinting: A Closer Look at Ethics and Policies. *Wake Forest JL & Pol'y*, 5, 275.

ADDITIONAL READING

Childers, E. P., Wang, M. O., Becker, M. L., Fisher, J. P., & Dean, D. (2015). 3D printing of resorbable poly (propylene fumarate) tissue engineering scaffolds. *MRS Bulletin*, 40(02), 119–126. doi:10.1557/mrs.2015.2

Lee, N. (2016). Revisiting the Principle of Technological Neutrality in Patent Protection in the Age of 3D Printing Technology and Cloud Computing. In *TRIPS plus 20* (pp. 361–388). Springer Berlin Heidelberg.

Lieben, L. (2016). Regenerative medicine: The future of 3D printing of human tissues is taking shape. *Nature Reviews. Rheumatology*, 12(4), 191. doi:10.1038/nrrheum.2016.29 PMID:26935277

Loy, J., Tatham, P., Healey, R., & Tapper, C. L. (2016). 3D Printing Meets Humanitarian Design Research: Creative Technologies in Remote Regions. *Creative Technologies for Multidisciplinary Applications*, 54.

ONeill, B., Wang, D. D., Pantelic, M., Song, T., Guerrero, M., Greenbaum, A., & O'Neill, W. W. (2015). Reply: the role of 3D printing in structural heart disease: all that glitters is not gold. *JACC: Cardiovascular Imaging*, 8(8), 988–989. doi:10.1016/j.jcmg.2015.04.011 PMID:26271095

Rideout, B. (2012). Printing the impossible triangle: The copyright implications of three-dimensional printing. *The Journal of Business. Entrepreneurship & the Law*, 5(1), 6.

Varkey, M., & Atala, A. (2015). Organ Bioprinting: A Closer Look at Ethics and Policies. *Wake Forest JL & Pol'y*, 5, 275.

Xing, J. F., Zheng, M. L., & Duan, X. M. (2015). Two-photon polymerization microfabrication of hydrogels: An advanced 3D printing technology for tissue engineering and drug delivery. *Chemical Society Reviews*, 44(15), 5031–5039. doi:10.1039/C5CS00278H PMID:25992492

KEY TERMS AND DEFINITIONS

Additive Manufacturing (AM): The process of making 3D objects by applying layer upon layer.

CAD Design File: Stands for computer-aided design file; used by engineers or architects to create technical illustrations.

Elephant Foot: The base of the model is slightly bulging outwards.

Friction-Weld: A welding process that generates heat through mechanical friction between workpieces to fuse materials.

Gunsmithing: The act of making, selling and repairing firearms.

Pillowing: Bumps or holes are present on the top surface of the printed model.

Post-Processing: Furbishing or decorating stage of the printed model.

Rapid Prototyping: A set of techniques used to quickly fabricate a scale model of a part using 3D CAD data.

Scalable Production: A production that can cope and perform under an expanding workload.

Stringing: Unsightly strings of plastic between parts of the printed model.

Warping: The print bends upwards at the base of the model and is no longer parallel with the print platform.



Category I

Information Resources Management

Advanced Model of Complex Information System



Miroslav Svitek

Czech Technical University in Prague, Czech Republic

INTRODUCTION

The proposed methodology represents the new ideas that come from defined mathematical assumptions like information-electrical or information-mechanical analogies. From these simple assumptions a lot of information physical principles can be derived like for example information flow, information content, information power, etc.

Analogies among electrical, mechanical and information circuits seem to be efficient attempts for problems solving within *systems engineering* by Vlček (1999). Concepts of *information power* and significant proximity of the measure of information and knowledge could enable upgrading these analogies for solving even wider class of tasks.

BACKGROUND

Data mean a change of state, for example from 0 to 1 or from 1 to 0, where the state vector is not necessarily only digital or one-dimensional. Every such change can be described with the use of a quantity of *information* in bits.

Information theory was founded by Shannon (1948) and his colleagues in the 1940s and was associated with coding and data transmission, especially in the newly emerging field of radar systems, which became a component of defensive systems during the Second World War.

Syntactic (Shannon) information has been defined as the degree of probability of a given event and has replied to the question: how often

a message appears? For example, by telling you that the solar system would cease to exist tomorrow, I would be giving you the maximum information possible, because the probability of this phenomenon occurring is nearly equal to zero. The probability model of information so defined has been used for the designing of self-repairing codes, digital modulations and other technical applications. Telecommunications specialists and radio engineers were concentrating on a probabilistic description of encoded data and on the minimizing of probability errors during data transmission.

The model-theoretical work of *semantic information* was done by Carnap and Bar-Hiller (1953). On the other hand, semantic information asks: how often a message is true? Zadeh (1965) introduced the theory of *fuzzy sets* as functions that map a value, which might be a member of a set, to a number between zero and one, indicating its actual degree of membership.

Currently, a number of interesting results have been discovered in the field of *quantum information science*, taking as their basis the foundations of quantum physics and using for modeling of complex systems those principles that do not arise in classical physics, such as *entanglement* and *quantization*. In the technical literature, we read that the behavior of entangled states is very odd. Firstly, it spreads rapidly among various phenomena, where for this spreading it makes use of a property known as *entanglement swapping*. The quantum information quantity in bits can be measured e.g. by *von Neumann entropy* in Vedral (2006) which measures the amount of uncertainty contained within the density operator taking into

account also wave probabilistic features like entanglement, quantization or bosonic / fermionic quantum behavior by Svítek (2012).

On the basis of the information theories, a number of methods and algorithms have emerged that attempt to eliminate or minimize indefiniteness and to do a better job of extracting the real, useful information from data. An excellent example is the *Bayes method* by Peterka (1981), which interprets the density of probability not as a description of a random quantity, but rather as a description of the indefiniteness of the system, i.e. how much information we have available about the monitored system. The system itself might be completely deterministic (describable without probability theory), but we may have very little available information about the system. When performing continuous measurement, we obtain more and more data, and therefore more information as well about our system, and our system begins to appear to us to be more definite. The elimination of indefiniteness therefore increases the quantity of information we have about the monitored system.

When eliminating indefiniteness, we also have to bear in mind the possibility of a change to the context of the event or phenomenon. There is plenty of testimony available to us from live witnesses, but there is none from dead ones, and this gives us an asymmetrical set of observations by Taleb (2010). It brings to mind the well-known saying that history is not written by the losers.

Once indefiniteness has been eliminated, one may proceed to the *interpretation of information*, or in other words, to the determination of how to reconstruct the described system, or how to build a more or less perfect model of it using the information. This task already belongs to the theory of systems, where it is necessary to identify the state vector, individual processes of the system etc. There emerges from this a knowledge system, which is able to describe the given object appropriately.

Information systems in Kelly and Cegielski (2009) aim to support management, operation and decision making, e.g. expert system, geographical

information system, enterprise system, etc. Components of computer-based information systems are hardware, software, databases, networks and procedures.

INFORMATION PHYSICS

Information behavior in Fidher, Erdelez and McKechnie (2005) means human behavior in relation to sources and channels of information - both active and passive information-seeking and use in different contexts. Information can be treated as physical matter or physical entity, and as such can be studied using the same methods as those applied by physics to study the properties of the physical matter. That is why new field *information physics* is established.

If we wish to apply information theory to the natural sciences, we should begin by describing information systems using a mathematical tool similar to one by which physics is described. We would therefore use analogies between various physical systems. In every such system, there exists potential and kinetic energy, to which there are corresponding quantities of potential and flow. In electronics, for example, voltage is defined as the quantity of potential, and electric current is the quantity of flow.

Let us then introduce a unit for the current of information and call it *information flow*, which is measured in bits per second and describes the input or output of information per unit of time. We can analogously define a quantity of potential, which we would call *information content*, which determines the quantity of work per bit (in Joules per bit). Information content for information systems (IT/ICT) can be defined as the number of 'success events' in the system per bit of information, and one may expect that if received information is significant, in information system a sequence of 'success events' is activated that orders our system. This also means that in order for us to obtain any concrete information content, we would already have to have done work, such as studying, searching for documentation, preparation etc. On the other

hand, it could be the other way around: the given information content might enable us to obtain a certain quantity of energy or (nowadays) funding.

From knowledge of information flow and information content, one can define other information physics quantities. One of the important quantities can be *information power*, defined by Svítek, Votruba and Moos (2010) as the product of information flow times information content. Analysis easily reveals that the unit of information power is work per second realized thanks to the received bit of information. For information systems (IT/ICT), information power is defined as the number of “success events” per second caused by the receipt of one bit of information. By introducing the quantity of information power, one can demonstrate that the impact of information is maximized if the received information flow is appropriately processed by the recipient and transformed into the best possible information content. If there is a flow of valuable information that the recipient is incapable of processing, the information power level is low. On the other hand, if the recipient is able to make good use of the information flow, but the flow does not carry needed information, the result is likewise a low level of information power.

We can take these ideas even further and introduce a phase shift between the information flow and content, thereby arriving at the definition of active and reactive information power by Svítek (2011). We can imagine *active information power* as power caused by information, which is transformed directly into concrete physical events. *Reactive information power* represents our emotions, which of course do not perform any work, but which support our decision making. Worth mentioning in this context is a well-known Bible story: King Solomon proposes to have a child split in half when two women are fighting over it, but because of her emotions, the real mother would rather give up her child than let it be killed.

An interesting area of information physics is the perception of time, which we can imagine as the number of biological events taking place in our bodies with a given frequency. If the information

power intake is slower than our biological clocks, we have the feeling that time is moving slowly, but if intake is faster, we have the feeling that time is being well used.

INFORMATION: PHYSICAL ANALOGIES

In Svítek, Votruba, Moos (2010) the extended Frege’s information concept was presented based on result of Vlček (1999). In Figure 1 basic information quantities are given:

$O_i(t)$ - a set of rated quantities on an object,

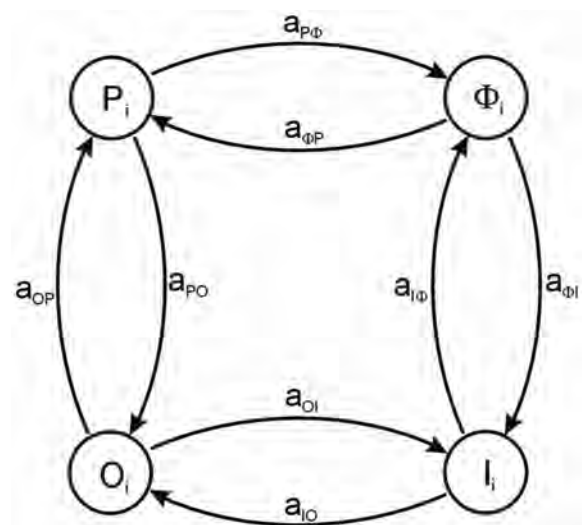
$P_i(t)$ - a set of states,

$\Phi_i(t)$ - a set of syntactic strings (data flow),

$I_i(t)$ - a set of information images of state quantities.

In physics, the state is a complete description of a system in terms of parameters at a particular moment in time. Thermodynamic state is a set of

Figure 1. Frege’s functional concept of information image origin and action



physical quantities (e.g. temperature, pressure, and composition) describing variable properties.

Other parameters representing links between information quantities are as follows:

a_{op} – identification,

a_{po} – invasivity,

$a_{p\phi}$ – projection in a set of symbols and syntactical strings,

$a_{\phi p}$ – uncertainty correction and identification,

$a_{\phi i}$ – interpretation, information origin,

$a_{i\phi}$ – language constructs reflection,

a_{io} – relation of functions and structural regularity,

a_{oi} – integrity verification.

The circuit theoretician Chua (1971) introduced the basic concept of electrical components together with the relations between them as it is shown in Figure 2. There are six different math-

ematical relations connecting pairs of the four fundamental circuit variables:

$q(t)$ - charge,

$\phi(t)$ - magnetic flux,

$i(t)$ - electric current,

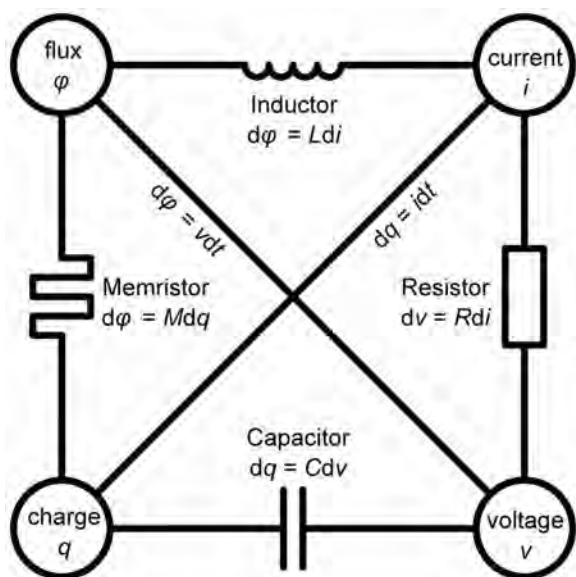
$v(t)$ - voltage.

From electrical variables definition we know that the charge is the time integral of the current. Faraday's law tells us that the flux is the time integral of the electromotive force, or voltage. There should be four basic circuit elements described by relations between variables: resistor, inductor, capacitor and memristor. Chua's concept is famous due to an envisioned new electrical component named "memristor" that provides a functional relation between charge $q(t)$ and flux $\phi(t)$.

Let us use the analogies between Chua's electrical and Frege's information concepts. Electrical charge $q(t)$ is measured in coulombs [c] and can be understood as basic electrical unit in analogy with the information unit measured in bits [bit]. Quantum physics really defines the quantum information unit called q-bit. We speak of (input / output) data changes $O_i(t)$ through which the system's model can be estimated.

Magnetic flux $\phi(t)$ is naturally related to the system states $P_i(t)$ which should be viewed as the piece of extracted knowledge (e.g. parameters estimated based on the observed data, understanding why system behaves in such a way, etc.). Time evolution of system states characterizes the system behavior. Magnetic flux is measured in webers [wb] corresponding to Joule multiplied by second and divided by coulomb [J.s/c]. In agreement with our analogy the states $P_i(t)$ are measured in [J.s/bit] which notes how much energy can be ex-

Figure 2. Chua's concept of electrical quantities



tracted / delivered based on one bit of information within one second.

Electrical current $i(t)$ measured in coulomb per second [c/s] is related to the information (syntax) flow $\Phi_i(t)$ that describes dynamical property of data changes in bits per second [bit/s]. The information flow $\Phi_i(t)$ typically represents the system input / output data flows per time.

Electrical voltage $v(t)$ leads to the analogical definition of the information (semantic) content $I_i(t)$ which characterizes the knowledge content measured in Joule per bit [J/bit]. For information systems (IT/ICT), the information content in [J/bit] can be alternatively defined as the number of “success events” caused by the receipt of one bit of information.

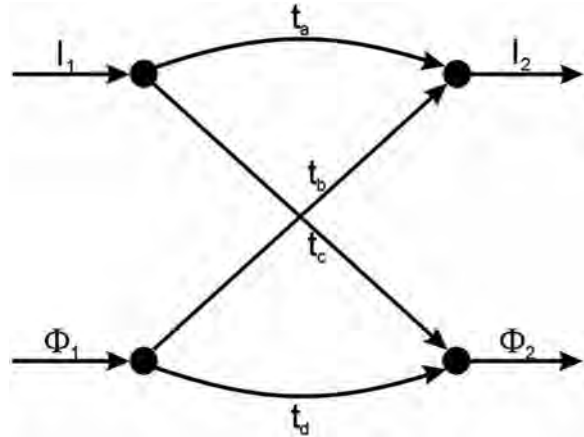
COMPLEX INFORMATION SYSTEMS

On the basis of the previous definitions, we can continue with the quantities we have introduced, information flow and information content, as well as with the quantities we have derived from them, such as information power. For the sake of simplicity, let us imagine an information subsystem as an *input-output information gate* given on Figure 3.

Between the input ports, input information content is available, and input information flow enters the system. Between the output ports, it is possible to obtain output information content, and output information flow leaves the system.

We can furthermore assume that this subsystem is open and is capable of using its surroundings as a source for drawing energy. Kauffman (2004) introduced the term *autonomous living agents*, which are characterized by the ability to direct and release energy. Kauffman is also the originator of the idea that the orderliness that is characteristic of living systems is defined by a series of actions leading to the dissemination of macroscopic work.

Figure 3. Information gate (Φ - information flow of data measured in bits per second, I - information content measured in Joule per bits)



Let us now examine the input-output information gate we have created. Input quantities can describe purely intellectual operations. Input information content includes our existing knowledge, and input information flow describes the change to the environment in which our gate operates and the tasks we want carried out (target behavior). All of the valuable, long-term information gained in this way can be used for the targeted release of energy, where at the output of the input-output gate, there may be information content on the order of millions of Joules per bit (or profits in millions of dollars). The output information flow serves as a model of the providing of such services or knowledge.

The basis of information systems is the ability to interconnect individual information subsystems, or in our case, input-output information gates. It is very easy to imagine the serial or parallel ordering of these subsystems into higher units. A very interesting model is feedback of information subsystems, because it leads to non-linear characteristics, to information systems defined at the limit of stability and other interesting properties. In this manner one may define *information filters*,

which are able to select, remove or strengthen a particular component of information.

FUTURE RESEARCH DIRECTIONS: POSSIBLE EXAMPLE

In the context of information physics, it is also necessary to deal with the problem of teaching, because the information subsystem called a *teacher* may be regarded as a source of information content. The teacher has prepared this information content for years with respect to both the content as such (optimizing the information content) and its didactic presentation (optimizing the information flow), so that the knowledge can be passed on to a subsystem known as a *student*. If we assume that the teacher subsystem has greater information content than the student subsystem, after their interconnection, the information flow will lead from the teacher to the student, so that the information content of the two subsystems will gradually balance out.

The students receive the information flow and increase their information content. If the students are not in a good mood, or if the information flow from the teacher is confused, the students are unable to understand the information received and to process it, so as to increase their information content. With the help of the reactive information power mentioned above, which concerns the emotional aspects of the recipient and the source, i.e. the student and the teacher, it is possible to create a situation where the students' sensitivity to the received information flow is maximized, so that they are able to process it appropriately and transform it into information content. Analogously, the teacher being in a good mood can lead to the creation of better information flow.

The individual components and subsystems of information systems can behave in different ways, and their behavior can be compared to everyday situations in our lives. A characteristic of politicians is their ability to use even a small input of information content to create a large output infor-

mation flow. They have the ability to take a small amount of superficially understood content and to interpret and explain it to the broadest masses of people. On the other hand, a typical professor might spend years receiving input information flow and input information content, and within her/his field, he/she may serve as a medium for transmitting a large quantity of output information content. The professor, however, might not spread the content very far, sharing it perhaps only with a handful of enthusiastic students.

It is hard to find an appropriate system to combine the characteristics of the different information subsystems described above, but it is possible to create a group of subsystems - system alliances introduced by Votruba and Novák (2010), where these characteristics can be combined appropriately. In this way, one can model a company or a society of people who together create information output that is very effective and varied, leading to improved chances for the survival and subsequent evolution of the given group. Such approach yields into *society knowledge* firstly introduced by Hayek (1945).

Through an appropriate combination of its internal properties, our information alliance can react and adapt to the changing conditions of its surroundings. Survival in alliances thus defined seems more logical and natural than trying for a combination of all necessary processes within the framework of one universal information subsystem.

CONCLUSION

For many years, I have been observing how much difficulty natural sciences have with explaining the basis of life, the behavior of living beings or of society, and how such important terms as *information* or *knowledge* have gradually crept into thinking in these fields.

It is as if we were building a house. We would need material (or mass), we would need plenty of workers (or energy), but without knowledge of



the plans for when and how to build, we could not erect the house. Information and knowledge are therefore the things that could enrich the natural sciences, enabling them to describe more faithfully the world around us.

In this text I tried to present basic approach to the information physics and to the complex information systems. I personally believe that the capturing of processes in the world around us with the help of information and knowledge subsystems organized into various interconnections, especially with feedback, can lead to the controlled dissemination of macroscopic work as described by Kauffman (2004), and after the overcoming of certain difficulties, even to the description of the behavior of living organisms.

I am convinced that the future will see a convergence of the physical sciences, life sciences and engineering. I would even allow myself to go a bit further, to consider even convergence with the humanities, because I am convinced that the laws of behavior of human society described, for example, in sociology or political science will be able to be better understood using the tools of information physics.

From the examples given above, I see the possibility for linking the physical world with the world of information, because every information flow must have its transmission medium, which is typically a physical object (e.g. physical particles) or a certain property of such an object. The case is again similar with information content, which also must be encoded through a real, physical system. The operations defined above the information systems can then likewise be depicted in a concrete physical environment.

Presented ideas may be controversial though in the information science field, as it continues the debate as to whether information can be treated as a thing or physical entity or whether it is socially constructed and does not exist outside of the human mind as was defined by Penrose (1989). This chapter should at least acknowledge that debate.

REFERENCES

- Bar-Hiller, Y., & Carnap, R. (1953). Semantic Information. *The British Journal for the Philosophy of Science*, 4(14), 147–157. doi:10.1093/bjps/IV.14.147
- Chua, L. O. (1971). Memristor - The Missing Circuit Element. *IEEE Transactions on Circuit Theory*, CT-18(5), 507–519. doi:10.1109/TCT.1971.1083337
- Fidher, K. E., Erdelez, S., & McKechnie, L. (2005). *Theories of Information Behavior*. ASIST Monograph Series.
- Hayek, F. A. (1945). The Use of Knowledge in Society. *The American Economic Review*, (4): 519–530.
- Kauffman, S. (2004). *Autonomous Agents. In Science and Ultimate Reality: Quantum Theory, Cosmology, and Complexity*. Cambridge University Press. doi:10.1017/CBO9780511814990.032
- Kelly, R. R., & Cegielski, C. G. (2009) *Introduction to Information Systems: Enabling and Transforming Business* (3rd ed.). Academic Press.
- Penrose, R. (1989). *The Emperor's Mind: Concerning Computers, Minds and The Law of Physics*. Oxford University Press.
- Peterka, V. (1981). Bayesian Approach to System Identification. In P. Eykhoff (Ed.), *Trends and Progress in System Identification* (pp. 239–304). Elmsdorf, NY: Pergamon. doi:10.1016/B978-0-08-025683-2.50013-2
- Shannon C.E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.*, 27(379), 623.
- Stonier, T. (1990). *Information and Internal Structure of the Universe*. Springer-Verlag London. doi:10.1007/978-1-4471-3265-3
- Svítek, M. (2011). Wave probabilistic information power. *Neural Network World*, 21(3), 269–276.

Svítek, M. (2012). Wave probabilistic functions for quantum cybernetics. *IEEE Transactions on Systems Man and Cybernetics, Part C-Applications and Reviews*, 42(2), 233–240. doi:10.1109/TSMCC.2011.2115237

Svítek, M. (2015). Towards Complex System Theory. *Neural Network World*, 25(1), 241–247.

Svítek, M., Votruba, Z., & Moos, P. (2010). Towards Information Circuits. *Neural Network World*, 20(2), 241–247.

Taleb, N. N. (2010). *The Black Swan: The Impact of the Highly Improbable*. New York: Random House and.

Vedral, V. (2006). *Introduction to Quantum Information Science*. Oxford University Press. doi:10.1093/acprof:oso/9780199215706.001.0001

Vlček, J. (1999). Systémové inženýrství [Systems Engineering]. ČVUT.

Votruba, Z., & Novák, M. (2010). Alliance Approach to the Modelling of Interfaces in Complex Heterogenous Objects. *Neural Network World*, 20(5), 609–619.

Zadeh, L. A. (1965). Fuzzy Sets. *Information and Control*, 8(3), 338–353. doi:10.1016/S0019-9958(65)90241-X

ADDITIONAL READING

Kauffman, S. (2000). *Investigation*. Oxford University Press.

Moos, P., & Malinovský, V. (2006). *Information systems and technologies, Monography NNW No.5*. Czech Academy of Science.

Stigler, G. (1961). The Economics of Information. *Journal of Political Economy*, 69(3), 213–225. doi:10.1086/258464

Svítek, M. (2006). *Dynamical Systems with Reduced Dimensionality, Edice monografií NNW č.6*. Praha: Neural Network World.

Svítek, M. (2009). *Quantum System Theory – Principles and Applications*. VDMVSG Saarbrücken.

KEY TERMS AND DEFINITIONS

Fuzzy Information: Determines the actual degree of set membership (fuzzy set) to a number between zero and one.

Information Content: Determines the quantity of work per bit and it is measured in Joules per bit.

Information Flow: Describes the input or output of information per unit of time and it is measured in bits per second.

Information Power: Determines the product of information flow and information content and it is measured in Joules per second.

Quantum Information: Determines how to integrate information theory with quantum mechanics, by studying how information can be stored with and retrieved from a quantum mechanical system. The primary piece of information in quantum information theory is the qubit, an analog to the bit (0 or 1) in classical information theory.

Semantic Information: Describes the content level of a given event and replies to the question: how often a message is true?

Shannon Entropy: Is a measure of the uncertainty in a random variable and quantifies the expected value of the information (in bits) contained in a message.

Syntactic Information: Describes the degree of probability of a given event and replies to the question: how often a message appears?

Von Neumann Entropy: Is the extension of Shannon entropy to the field of quantum mechanical systems described by a density matrix.

Computer Information Library Clusters



Fu Yuhua

CNOOC Research Institute, China

INTRODUCTION

As well-known, the concept of “set” has been constantly expanded and developed. For example, it can be expanded and developed into the concepts of “generalized set” and the like. At present, many theories and practices demonstrate the viability of big data analysis as a global business activity. In order to meet the needs of big data analysis and the like, we consider that the concepts of “generalized set” and the like should be expanded and developed into the concept of “computer information library clusters” (CILC).

DEFINITIONS AND NOTATIONS

In this section we shall present some basic definitions and notations.

Definition 1: The expanded and developed result of set is defined as generalized and hybrid set.

Definition 2: The expanded and developed result of generalized and hybrid set is defined as library.

Definition 3: The expanded and developed result of libraries related to computer information is defined as “computer information library clusters” (CILC).

Definition 4: In this paper, the variation principle is defined as the following standard form

$$\Pi = \min_0$$

where: $\min_0$ is the minimum and its value should be equal to zero.

More definitions and notations can be found in (Fu, 2013), (Fu, 2016a), and (Fu, 2016b) respectively.

BACKGROUND

In (Fu, 2016a), generalized and hybrid set can be created with neutrosophy and quad-stage method. In which, generalized and hybrid set are discussed firstly; based on this, the concepts of “problem set”, “solution set”, “principle set”, “law set”, “theory set”, “formula set”, and the like are presented; Secondly the combination or synthetic body of generalized and hybrid sets is named as “library” (various generalized and hybrid sets can be put into the related “library”); such as “mathematics library”, “physics library”, “natural science library”, “social science library”, and the like. As for the constitution of “library”, referring to quad-stage method and Chinese ancient “Complete Library of Four Branches of Books”, the concept and methodology of a special “Four-library” (including four sub-libraries: information library, question library, correlation library, and achievement library) are proposed. Neutrosophy and quad-stage method can also be used to solve many practical problems within the framework of “set” and “library”; for example, based on the analyses of one “Four-library”, jointly solving problem of advance of planet’s perihelion with partial results of law of gravity and general relativity (these two theories belong to “gravitational theory set”); and jointly expanding “uncertainty principle” to “certainty-uncertainty principle set” (including three principles in different conditions:

“certainty principle”, “uncertainty principle”, and neutral (fuzzy) “indeterminacy principle”) with Heisenberg inequality and Ozawa inequality. Finally, with the help of the concepts of “generalized and hybrid set” and “library”, we introduce the concepts of “variation principle of set” and “variation principle of library”, and establish a kind of “partial and temporary unified theory of mathematics so far”.

Based on the concept of “library” presented in (Fu, 2016a), this paper presents the concept of “computer information library clusters” (CILC).

CONSTITUTION OF “COMPUTER INFORMATION LIBRARY CLUSTERS” (CILC)

There are various ways and means to form CILC. For instance, CILC can be considered as the “total-library”, and consists of several “sub-libraries”. The example is the “Four-library” presented in (Fu, 2016a) (including four sub-libraries: information library, question library, correlation library, and achievement library). As another example, in CILC, a “total-library” can be set up, and a number of “sub-libraries” are side by side with the “total-library”. In which, the relationships between “total-library” and “sub-libraries” are also various. In (Fu, 2016a), the “total-library” may include all or part of the contents (such as directory, abstracts, etc) of sub-libraries. Of course, generally the “total-library” includes part of the contents of sub-libraries only.

EXAMPLES OF “COMPUTER INFORMATION LIBRARY CLUSTERS” (CILC)

For the sake of convenience, only discuss the situation that the “total-library” consists of a number of “sub-libraries”. In which, “sub-library” can be

divided into “first order sub-library”, “second order sub-library”, and the like.

Firstly, we discuss “natural science computer information library clusters”. If it is seen as the “total-library”, then the “sub-library” can be constituted in the following three ways: (1) constituted in accordance with discipline, (2) constituted in accordance with name, (3) constituted in accordance with A to Z sequence.

If constituted in accordance with discipline, the “first order sub-libraries” include: “mathematics library”, “physics library”, “chemistry library”, “biology library”, “medicine library”, and so on. In “mathematics library”, “second order sub-libraries” include: “algebra library”, “geometry library”, “trigonometry library”, “calculus library”, and so on.

If constituted in accordance with name, the “first order sub-libraries” include: “Newton library”, “Einstein library”, “Archimedes library”, “Euclid library”, “Qian Xuesen library”, and so on. In “Newton library”, “second order sub-libraries” include: “Newton mathematics library”, “Newton physics library”, “Newton mechanics library”, and so on.

If constituted in accordance with A to Z sequence, the “first order sub-libraries” include: “information library that begin with the letter A” to “information library that begin with the letter Z”. In “information library that begin with the letter A”, “second order sub-libraries” include: “information library that begin with the letters AA” to “information library that begin with the letters AZ”.

Secondly, the “computer information library clusters” (CILC) can also be constituted according to the different requirements and different points of interest. Specially, for CILC, the operation functions can be added.

For example, in (Fu, 2013), for unified dealing with the problems of natural science, applying least square method, “partial and temporary unified theory of natural science so far” can be

expressed in the following form of “partial and temporary unified variation principle of natural science so far”

$$\Pi_{\text{NATURE}} = \sum_1^n W_i \int_{\Omega_i} F_i^2 d\Omega_i + \sum_1^m W_j' S_j^2 = \min_0 \quad (1)$$

where: the subscript NATURE denotes that the suitable scope is all of the problems of natural science, all of the equations $F_i = 0$ denote so far discovered (derived) all of the equations related to natural science (their suitable scopes are Ω_i), all of the equations $S_j = 0$ denote so far discovered (derived) all of the solitary equations related to natural science (they are suitable on solitary points or in some special cases, the meanings can be found in (Fu, 2013)), W_i and W_j' are suitable positive weighted constants, and $\min_0$ was introduced in (Fu, 1989), indicating the minimum and its value should be equal to zero.

In this way, the theory of everything to express all of natural laws, described by Hawking that a single equation could be written on a T-shirt, is partially and temporarily realized in the form of “partial and temporary unified variation principle of natural science so far”.

However, the disadvantages of such a “partial and temporary unified theory of natural science so far” are also very obvious, namely it is disorganized and lack of layers.

In order to avoid these disadvantages, we introduce some concepts according to the thought of “computer information library clusters” (CILC).

The first one is the concept of “variation principle of library”.

The meaning of “variation principle of library” is the variation principle formed by using least square method (LSM) to process all equations and equalities included in the library.

By using least square method (LSM) to process the equations and equalities related to “library of quadratic equation of one unknown”, we have

the following “variation principle of library of quadratic equation of one unknown”

$$\Pi_{\text{QEOOlibrary}} = \Pi_1 + \Pi_2 + \Pi_3 + \dots = \min_0$$

where: the subscript QEOOlibrary denotes “library of quadratic equation of one unknown”, $\Pi_1 = F_1^2, \Pi_2 = F_2^2, \Pi_3 = F_3^2, \dots$

Similarly, for “library of cubic equation of one unknown”, we have “variation principle of library of cubic equation of one unknown” ($\Pi_{\text{CEOOlibrary}} = \min_0$); for “library of quartic equation of one unknown”, we have “variation principle of library of quartic equation of one unknown” ($\Pi_{\text{Q'EOOlibrary}} = \min_0$); and the like.

In addition, we expand the concepts of equation and solution, and discuss point equation, line equation, plane equation, solid equation, sub-domain equation, whole-domain equation, and the like; as well as point solution, line solution, plane solution, solid solution, sub-domain solution, whole-domain solution, and the like in (Fu, 2016b). Accordingly, the concepts of “point equation library”, “line equation library”, “plane equation library”, “solid equation library”, “sub-domain equation library”, “whole-domain equation library”, and the like; as well as “point solution library”, “line solution library”, “plane solution library”, “solid solution library”, “sub-domain solution library”, “whole-domain solution library”, and the like, can be presented. Furthermore, the related “variation principle of library” can be presented also.

For “natural science computer information library clusters” (natural science CILC), applying “variation principle of library”, “partial and temporary unified theory of natural science so far” with different degrees can be established as follows

$$\Pi_{\text{NATURElibrary}} = \Pi_{\text{MATHsublibrary}} + \Pi_{\text{PHYSsublibrary}} + \Pi_{\text{CHEMsublibrary}} + \dots = \min_0 \quad (2)$$

where: $\Pi_{\text{MATHsublibrary}}$ is “variation principle of mathematics library”, $\Pi_{\text{PHYSsublibrary}}$ is “variation principle of physics library”, $\Pi_{\text{CHEMsublibrary}}$ is “variation principle of chemistry library”.

For more applications of “variation principle of library”, referring to the examples in (Fu, 2016a) that solving problem of advance of planet’s perihelion with partial results of law of gravity and general relativity (now, these two theories belong to “gravitational theory library”); and jointly expanding “uncertainty principle” to “certainty-uncertainty principle library” (including three principles in different conditions: “certainty principle”, “uncertainty principle”, and neutral (fuzzy) “indeterminacy principle”) with Heisenberg inequality and Ozawa inequality (now, these two inequalities belong to “inequality library”).

At present, as running various big data analysis, we can put all data into CILC in advance, then, for a special analysis, choosing related equations and formulas and applying “variation principle of library”, thus we can find the optimum or suitable result.

FUTURE RESEARCH DIRECTIONS

Referring to the concept of “natural science CILC”, the concepts of “social science CILC”, “natural science and social science CILC”, and the like, can be presented.

Similarly, referring to the concept of “computer information library clusters”, the concepts of “computer and non-computer information library clusters”, “earth information library clusters”, “solar system information library clusters”, “Milky Way galaxy information library clusters”, “universe information library clusters”, and the like, can be presented.

CONCLUSION

This paper discusses the concept of “computer information library clusters”(CILC), as well as other concepts of “information library clusters”. The examples in this paper show that, with the operation function, “computer information library clusters”, as well as other “information library clusters”, will have a wide range of applications. At present, the theories and methods proposed in this paper can pave the way for many practical applications, such as big data analysis and the like.

REFERENCES

- Fu, Y. H. (1989). New solution for problem of advance of Mercury’s perihelion. *Acta Agronomica Sinica*, 30(4), 350–354.
- Fu, Y. H. (2013). Neutrosophic Examples in Physics. *Neutrosophic Sets and Systems*, 1, 26–33.
- Fu, Y. H. (2016a). *Creating Generalized and Hybrid Set and Library with Neutrosophy and Quad-stage Method. In Handbook of Research on Generalized and Hybrid Set Structures and Applications for Soft Computing*. Hershey, PA: Sunil Jacob John.
- Fu, Y. H. (2016b). Point Equation, Line Equation, Plane Equation etc and Point Solution, Line Solution, Plane Solution etc —Expanding Concepts of Equation and Solution with Neutrosophy and Quad-stage Method. *Neutrosophic Sets and Systems*, 11, 67–73.

KEY TERMS AND DEFINITIONS

Computer Information Library Clusters (CILC): The expanded and developed result of libraries related to computer information is defined as “computer information library clusters” (CILC). For example, CILC can be considered as the “total-library”, and consists of several “sub-libraries”. As another example, in CILC, a “total-library” can be set up, and a number of “sub-libraries” are side by side with the “total-library”.

Partial and Temporary Unified Variation Principle of Natural Science So Far and Partial and Temporary Unified Theory of Natural Science So Far: For unified dealing with the problems of natural science, applying least square method, “partial and temporary unified theory

of natural science so far” can be expressed in the form of “partial and temporary unified variation principle of natural science so far”. In this way, the theory of everything to express all of natural laws, described by Hawking that a single equation could be written on a T-shirt, is partially and temporarily realized in the form of “partial and temporary unified variation principle of natural science so far”.

Variation Principle of Library: The meaning of “variation principle of library” is the variation principle formed by using least square method (LSM) to process all equations and formulas included in the library. It can pave the way for many practical applications, such as big data analysis and the like.



The Impact of the Impact of Meta-Data Mining From the SoReCom “A.S. de Rosa” @-Library

Annamaria Silvana de Rosa

Sapienza University of Rome, Italy

Laura Dryjanska

Sapienza University of Rome, Italy

Elena Bocci

Sapienza University of Rome, Italy

INTRODUCTION

The Social Representations Theory was founded by Serge Moscovici in 1961, and after more than 50 years, the study of social representations, originally specifically European, is currently a multilingual, worldwide discipline with a substantial body of literature, involving leading scholars from social psychology and other social sciences (de Rosa, 2011, 2013a, 2013b, 2016b; Sammut, Andreouli, Gaskell, & Valsiner, 2015; Lo Monaco, Delouvée, & Rateaux, 2016). The field of Social Representations represents a unifying meta-theoretical perspective on the social construction of knowledge and its relation to socially situated practices. It has important applications for the public and private sectors, acting as a bridge among disciplines including psychology, social psychology, sociology, cultural studies with pragmatic approach to language, semiotics, socio-history, anthropology, and communication studies (also including multidisciplinary approaches from computer sciences and new technologies) with important implications for institutional and organizational contexts, culture and health practices, inter-group relations, ideology and politics, economics, the environment, etc. Interested in how scientific knowledge is transformed by lay

people and the media into common sense, it is also a supra-disciplinary field because it has activated a conversation among social, human and natural scientists in a wide range of internationally recognised research programmes. These concern public understanding of the sciences and discoveries in various fields, such as medicine, environmental studies, biology, informatics, economy, political science, etc., and the social representations of complex new multidisciplinary topics like biogenetic foods, medical innovation, globalisation and climate change, forms of interaction through new media, the risk society, immigration, minority groups, racism and multiculturalism, human rights, European integration and enlargement, etc. Consequently, this field is open to trans-disciplinary and multi-methodological research approaches (experimental and field work). The “objects” studied have a strong societal impact and important practical applications “within” and “for” society in the political, economic and social spheres.

The diffusion of the theory of social representations have been an object of analysis in the bibliometric domain in the light of the critical debate, which still animates the community of scientists, stimulating meta-reflexive discussion and view exchanges among the members of our

scientific community on the preferable publishing options and collaborative strategies in the current editorial and academic scenario (de Rosa, 2015a).

The objective of this chapter is to discuss the “*impact of the impact*” of the social representations theory in the bibliometric culture era, *even beyond the journal’s impact*. At this purpose we will present both:

- A selection of results on the dissemination of the theory across the continents based on data and meta-data concerning authors’ countries and institutional affiliations, comparing articles published in “*not indexed*” and “*indexed*” journals as derived from the two largest bibliometric databases (Isi-web of Science Thomson and Reuters and Scopus-Elsevier);
- And the *geo-mapping* of the wider scientific production in Social Representations.

BACKGROUND

Background question concerns what is the value of the scientific networking, training and documentation activities in the new academic scenario dominated by the bibliometric assessment culture and by the impact of the technology to the science production and sharing (data-driven science, big data, open data, open access, etc.).

The spirit that animates our interest in considering the “*impact of the impact*” is coherent with the main assumptions of the “*impact beyond the impact factor*” (Zupanc, 2014), and with the conclusions/recommendations of final assessment report of the Higher Education Founding Council for England, Wales and Northern Ireland (HE-FCE) (<http://www.hefce.ac.uk/pubs/rereports/Year/2015/metrictide/Title,104463,en.html>)

- There is considerable skepticism among researchers, universities, representative bodies and learned societies about the broader use of metrics in research assessment and management.

- Peer review, despite its flaws, continues to command widespread support as the primary basis for evaluating research outputs, proposals and individuals. However, a significant minority is enthusiastic about greater use of metrics, provided appropriate care is taken.
- Carefully selected indicators can complement decision-making, but a ‘variable geometry’ of expert judgement, quantitative indicators and qualitative measures that respect research diversity will be required.
- There is legitimate concern that some indicators can be misused or ‘gamed’: journal impact factors, university rankings and citation counts being three prominent examples.
- The data infrastructure that underpins the use of metrics and information about research remains fragmented, with insufficient interoperability between systems.
- Analysis concluded that no metric can currently provide a like-for-like replacement for REF peer review.
- In assessing research outputs in the REF, it is not currently feasible to assess research outputs or impacts in the REF using quantitative indicators alone.

and of the following 2015 report on “*Future approaches to quality assessment in England, Wales and Northern Ireland*”: <http://www.hefce.ac.uk/pubs/year/2015/201511/>.

More closely to our disciplinary field this spirit is coherent with the *opinion piece for the Bulletin of the European Association of Social Psychology*, written together with other internationally recognized social psychologists convened in a small meeting in Lausanne (June 12-14, 2013) to reflect on the new conformism dominating research practices in social psychology and to launch debate within the European Association of Social Psychology (EASP).

Another motivational driving force for involvement in the empirical investigation in the field of bibliometric culture is closely related to improve-



ment of the So.Re.Com. “A.S. de Rosa” @-library, and its exploitation as a scientific tool to conduct empirical research on the development and dissemination of the social representations literature around the world (by continents, countries, cities and even single institutions of author affiliation) and across the generations of scientists belonging to different academic cultures and contexts and working in different thematic areas, by different paradigmatic and methodological approaches. Adding the bibliometric indexes to the rich set of data and meta-data detected for each bibliographic item and filed in the bibliographic and meta-theoretical analysis repositories is intended to evaluate also the “*impact of the impact*” of the literature inspired by the social representations theory. Reconstructing the kind and evolution of inter-individual and inter-institutional co-operations is another goal of the wider meta-theoretical analysis research program. Further publications will be dedicated to the dynamics of the knowledge epidemiology via the inter-institutional collaborations between authors belonging to institutions in different countries and continents: *who* works with *whom*, (on what) and *where*?

GEO-MAPPING THE GLOBAL DISSEMINATION OF THE SOCIAL REPRESENTATIONS THEORY AND ITS IMPACT IN THE BIBLIOMETRIC CULTURE ERA

Coherently with the title of this chapter, our goal is to discuss some aspects of “*the IMPACT of the IMPACT*”, presenting a selection of results visualized according to a technique designed ad hoc (de Rosa, 2014b) for geo-mapping the development and the dissemination of the theory across the continents, over several generations of scientists, and according to thematic, paradigmatic and methodological approaches. Given the limited space available, here we will discuss only some of the comparative analyses based on “*big data*” and “*meta-data*” filed in our SoReCom “A.S. de Rosa” @-library repositories, concerning authors’

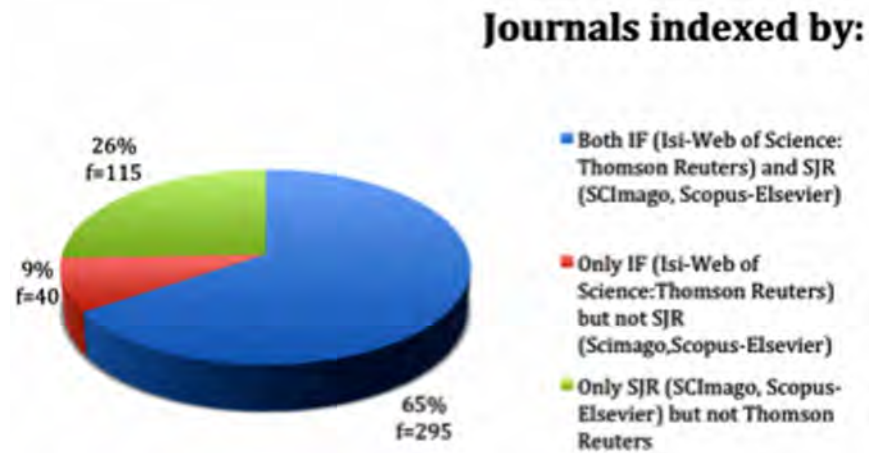
countries and institutional affiliations, years of publication by decades, thematic areas, language of publication, etc. and in particular:

1. Those related to the various bibliometric indexes (Impact Factor and SJR) as derived from the two largest bibliometric databases: Isi-web of Science Thomson and Reuters and Scopus-Elsevier, presenting the geo-mapping by Continents according to the country of the author’s institutional affiliation when considering comparatively the inclusion of the scientific production in both (a1.) or exclusively in just one of the two bibliometric databases of indexed journals (a.2);
2. The *geo-mapping* of the wider scientific production in Social Representations and comparative results with the restricted sources included in the indexed bibliometric databases;
3. The relevance of the crosscutting thematic choice for the articles in the top six “indexed journals”.

Sources for the Empirical Analyses and Main Results

We may define our data source as “big data” or almost “big data”, because up to now our SoReCom “A.S. de Rosa” @-library is the most comprehensive library specialised in Social Representation, compared the universe of the total literature. Although we are aware of classical definition of big data (see among others: Boyd & Crawford, 2012; Mayer-Schönberger & Cukier, 2013), we assume that “There is no one definition of Big Data. Thought about in simple terms, Big Data involves datasets that are far larger than those traditionally examined in journals like this one. Yet there has always been considerable variation in the size of datasets, ranging from small experimental studies to large samples involving census or polling data. Size alone is therefore an insufficient descriptor” (Parks, 2014: 255).

Figure 1. The distribution of 450 “indexed journals according to the inclusion in the bibliometric databases



The bibliographic sources of the empirical data used for our analyses were extracted from a larger number of 10325 bibliographic references (as of July 2015), of which 8740 items specifically related to social representations and communication (including also books, book chapters, conference presentations, web documents, manuscripts, university reports, Master and PhD theses, etc.) filed in the repositories of the *SoReCom “A.S. de Rosa” @-Library* (de Rosa, 2014a).

The results show that out of a total of 3239 articles:

- 1747 (54%) articles have been published in 624 “not indexed” journals
- 1492 (46%) in 450 “indexed” journals included in at least one of the bibliometric databases (Isi-web of Science Thomson and Reuters WoS and Scopus-Elsevier). The 450 journals are distributed as illustrated in the Figure 1 according to the criterion of inclusion in both or at least one of the two bibliometric databases: although the majority (65%, $f=295$) are indexed by both IF and SJR, those indexed only by IF (included in Isi-Web of Science) represent the 9% ($f=40$) compared to the 26% ($f=115$) indexed by SJR (included in Scopus-Elsevier).

This wide range of 450 “indexed” journals is in itself an indicator of the impact of the Social Representations theory and research field, which has gained its visibility beyond the borders of social psychology and, thanks to its supra-disciplinary epistemological power of attractiveness, has reached audience in journals from many disciplinary and thematic fields.

Based on a selection of 3239 articles related to Social Representations and Communication published in journals, the analyses of the results (Figure 2 and Figure 3) show that 1088 (34%) are included in both the databases from which the bibliometric indexes were extracted:

- *Impact Factor* (Thomson and Reuters) = 1174 articles, of which only 81 (2%) were included exclusively in Isi-web of Science;
- *SJR* (Scopus-Elsevier) = 1.414 articles, of which only 323 (10%) were included exclusively in Scopus.

Inspection of the frequencies distribution of indexed journals included in one or two bibliometric databases by journal’s “country ranking” clearly shows the dominance of English-speaking countries: United Kingdom (especially for the scientific production published in indexed journals by Scopus-Elsevier: SJR SCImago) for Europe

Figure 2. Sources: the distribution of 3239 articles published in 1074 “not indexed” and “indexed journals”

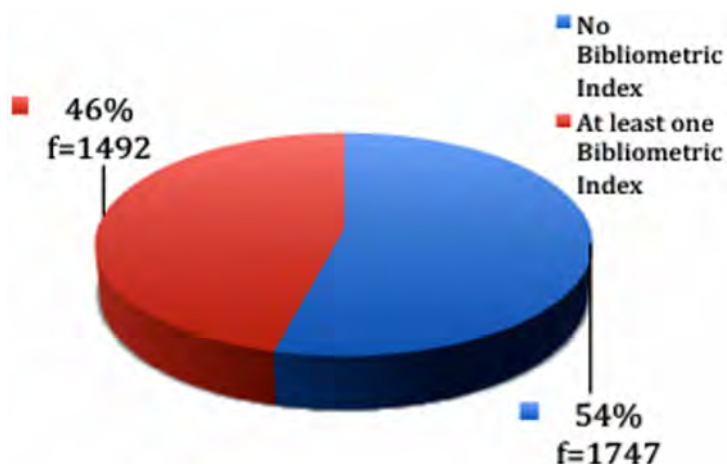
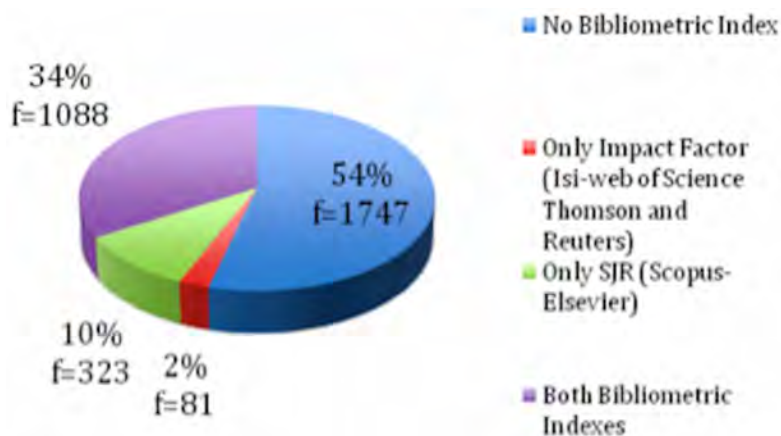


Figure 3. The distribution of 3239 articles published in “not indexed” journals or included in one or both of the “bibliometric databases”

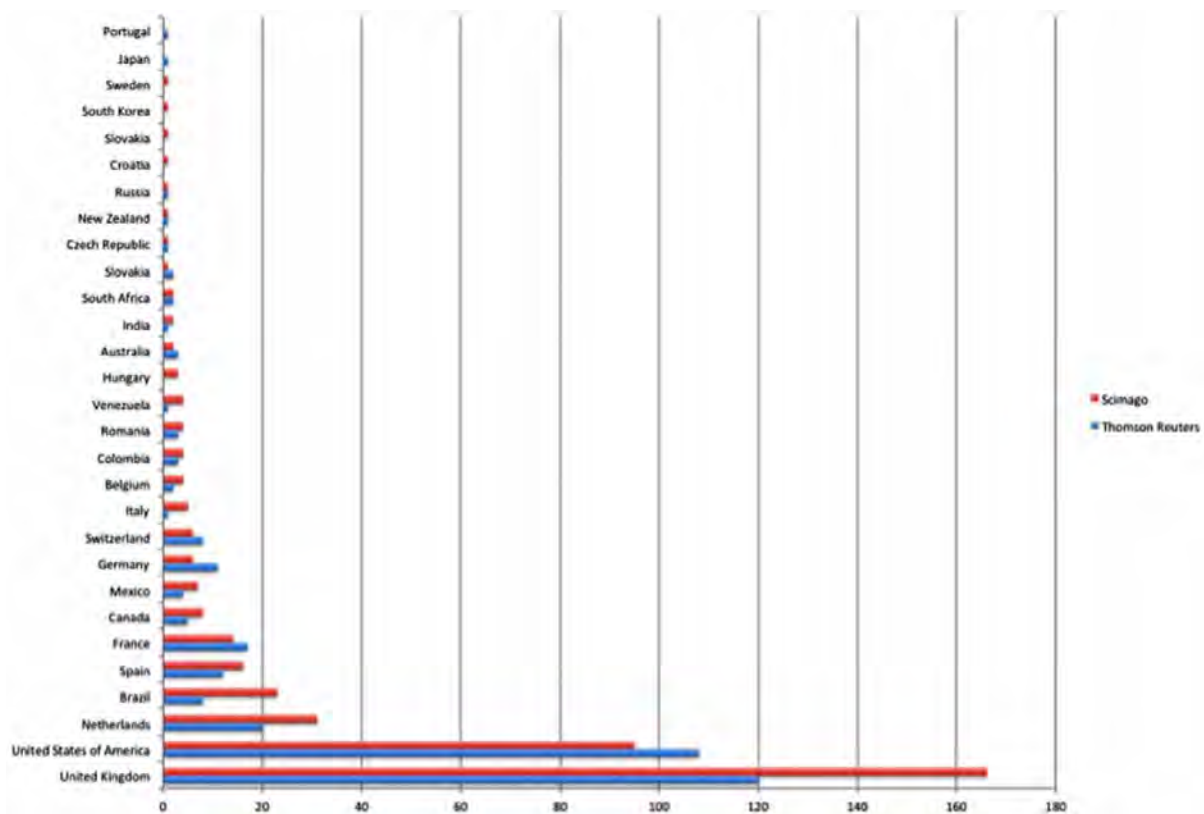


as a continent and the United States (the only North American country appearing for the scientific production published in indexed journals by Isi-web of Science: IF), for the continent of America, followed by the Netherlands (a country where scientists publish mainly in English and are strongly linked with the British and North American scientific world and where the headquarters of the Elsevier publishing multinational company is located) (see Fig. 4).

Geo-Mapping the Wider Scientific Production in Social Representations and Comparative Results with the Restricted Sources Included in the Indexed Bibliometric Databases

Comparison of the results based on the indexed journals with the geo-mapping by the authors' institution continents based on all 3234¹ articles on

Figure 4. The frequencies distribution by “country ranking” based on the 450 “indexed journals” included in at least one or two bibliometric databases *Isi-Web of Science (IF)* and *Scopus (SJR SCImago)*



social representations published in journals filed in the *SoReCom “A.S. de Rosa” @-library* bibliographic repository (and not only in the 450 indexed journals by “country ranking” as represented in the Figure 8) evidences the prominence of Europe, and even more so of Latin America (compared to North America), for articles published in non indexed journals; whereas North America, Asia and Oceania are the prominent continents for articles published in indexed journals, compared to Europe (where the theory was initially generated and spread) and to Latin America (where the theory has been largely disseminated in the last two decades).

The next graphs – realised by the software *Tableau Desktop* (<http://www.tableausoftware.com>) – visualize the massive data based on the wider scientific production in Social Representations (on 3234 articles on social representations published in journals filed in the *SoReCom “A.S. de Rosa”*

@-library bibliographic repository) respectively on worldwide scale (Figure 5) with respective zooms on Continents and geo-cultural American Regions (Europe: Figure 6; Latin America: Figure 7; North America: Figure 8; Oceania: Figure 9; Asia: Figure 10; Africa: Figure 11). These graphs visualize on world wide scale the absolute frequencies for each country the distinction between the articles published in “non indexed journals”, those published in journal indexed in both or one of the bibliometric data bases (Thomson and Reuters for Impact Factor or Scopus-Elsevier for SJR). Therefore they clearly show – both on worldwide scale with a zoom on country by country in each continent - the comparative results based on the wider scientific production in the format of 3234 articles with the restricted sources included in one or both the indexed bibliometric databases, as already analytically described in the previous sections.

Figure 5. The absolute frequencies’ distribution of 3234 in articles (filed in the repositories of the SoReCom “A.S. de Rosa” @-Library) published in “non indexed” and “indexed journals” on worldwide scale according to the country of the author’s institutional affiliation

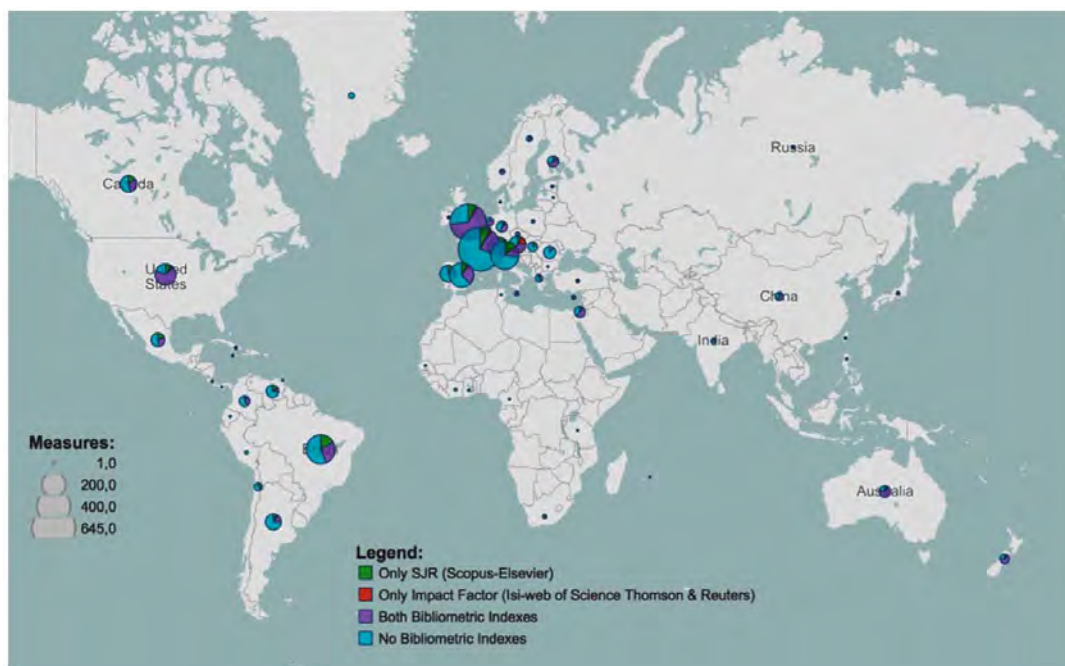


Figure 6. The absolute frequencies’ distribution of 2203 articles (filed in the repositories of the SoReCom “A.S. de Rosa” @-Library) published in “non indexed” and “indexed journals” by the continent Europe according to the country of the author’s institutional affiliation



Figure 7. The absolute frequencies' distribution of 587 articles (filed in the repositories of the SoReCom "A.S. de Rosa" @-Library) published in "non indexed" and "indexed journals" by the continent Latin America according to the country of the author's institutional affiliation



Figure 8. The absolute frequencies' distribution of 259 articles (filed in the repositories of the SoReCom "A.S. de Rosa" @-Library) published in "non indexed" and "indexed journals" by the continent North America according to the country of the author's institutional affiliation



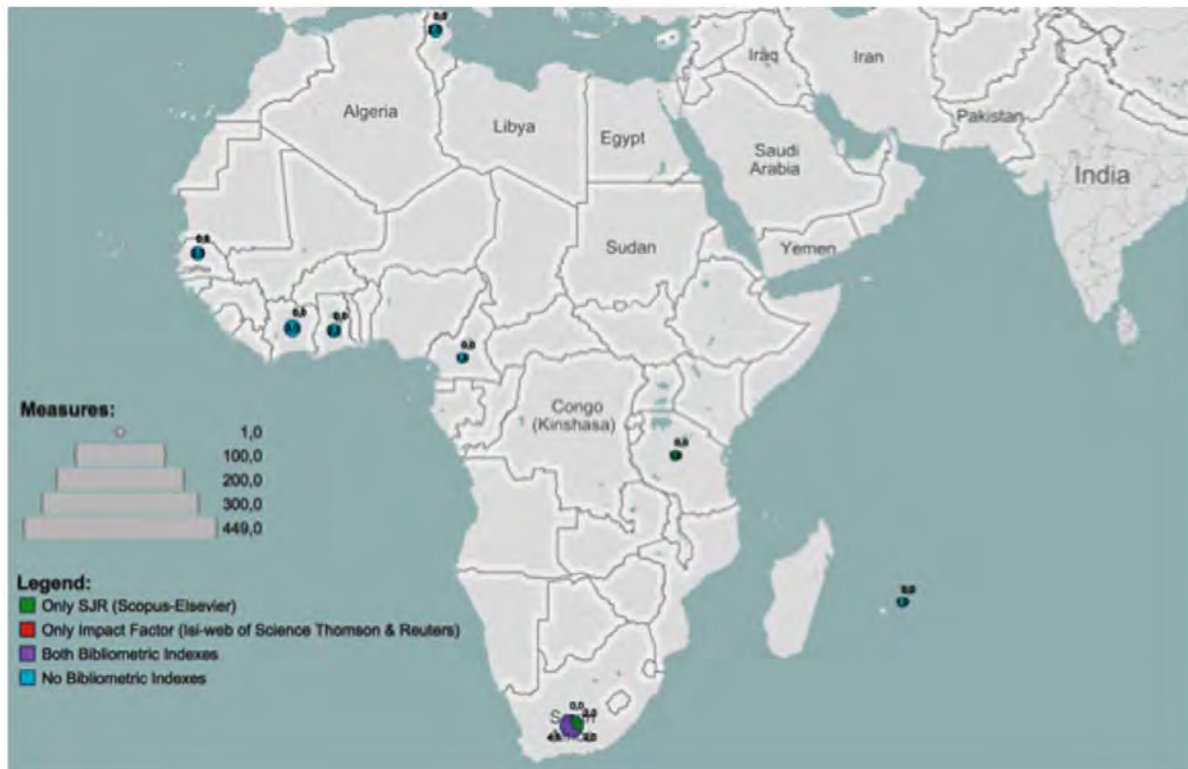
Figure 9. The absolute frequencies’ distribution of 77 articles (filed in the repositories of the SoReCom “A.S. de Rosa” @-Library) published in “non indexed” and “indexed journals” by the continent Oceania according to the country of the author’s institutional affiliation



Figure 10. The absolute frequencies’ distribution of 89 articles (filed in the repositories of the SoReCom “A.S. de Rosa” @-Library) published in “non indexed” and “indexed journals” by the continent Asia according to the country of the author’s institutional affiliation



Figure 11. The absolute frequencies' distribution of 19 articles (filed in the repositories of the SoReCom "A.S. de Rosa" @-Library) published in "non indexed" and "indexed journals" by the continent Africa according to the country of the author's institutional affiliation



An even more impressive paradoxical effect emerges when the prominence of the three countries (United Kingdom, United States and Netherlands) in the distribution of 450 indexed journals including publications on social representations (Figure 4) by "country ranking" is compared with the geo-mapping of the entire scientific production (8348 items specifically related to social representations and communication, including also books, book chapters, conference presentations, web documents, manuscripts, university reports, master and PhD theses, etc.) filed in the SoReCom "A.S. de Rosa" @-library by continents and countries according to the affiliation countries of authors' institutions² (Figure 12).

A clear prominence of the European countries (especially France, UK, Italy, Spain and Switzerland) with 5488 publications, followed by the Latin American countries (especially Brazil, Argentina,

Mexico and Venezuela) with 2112 publications – among the list of 72 countries represented worldwide - emerges in the production-diffusion of the literature related to this scientific field.

Not only in terms of the prominence the geo-mapping of the literature produced by authors belonging to different Continents/Countries differ when comparing the restricted sources extracted by the 450 "indexed journals" with the entire production filed in the SoReCom "A.S. de Rosa's" @-library, but also in terms of the number of diversified countries and therefore of different geo-cultural contexts which contribute to the development of theory and research in the supra-disciplinary field of Social Representations and Communication. In fact the number of the worldwide countries represented in the entire corpus of 8348 items are 72 (among them 44 countries corresponding to author's publications

Figure 12. The worldwide % frequencies distribution of 8348 items specifically related to social representations and communication (including also books, book chapters, conference presentations, web documents, manuscripts, university reports, master and PhD theses, etc.) filed in the repositories of the SoReCom “A.S. de Rosa” @-Library by Continents according to the country of the author’s institutional affiliation



in journals indexed in both bibliometric data bases, 55 countries corresponding to author’s publications in not indexed journals), whilst the list of the “country ranking” included in at least one of the bibliometric databases is restricted to 29 (24 countries for articles indexed only in Isi-Web of Science for IF and 27 only in Scopus-Elsevier for SJR) (see Figure 4).

Issues, Controversies, Problems

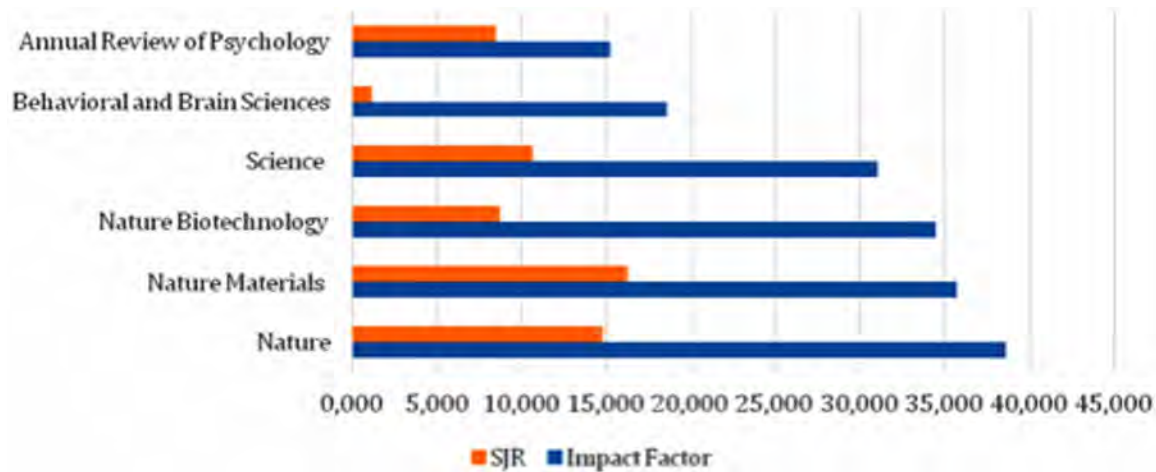
It is evident from the statistical results visualized on the geo-maps that a serious epidemiological study on the diffusion of a scientific field cannot take into account partial sources of information (only articles and not also books; only articles in indexed journals and not the entire scientific production)³.

The analysis of the relevance of the scientific production also requires further refinement by cross-referencing many other interesting meta-data, like for example the inter-institutional collaboration “within” and “between” Continents and

Countries and even single Institution, paradigmatic approaches, preferred research designs and methodological options, thematic choices, etc.: by way of example, by analysing the significant link between the source of publication and the *thematic choice*, as evident on looking at the intersection between the top six “indexed” journals (selected by Impact Factor more than 10) and the topic chosen by the authors⁴, including multidisciplinary subjects (engineering or biochemistry, genetics and molecular biology, or agricultural and biological science, neuroscience, psychology) and just one alone in psychology, dealing with cross-cutting edge topics like biogenetics, nanotechnology, biotechnology (Figure 13).

This comment about the *relevance of the cross-cutting thematic choice, multidisciplinary domains and applied societal value of the topics* receives an external validation in the explanations provided by the authors of the 24 top cited papers – selected from an impressive initial search results including more than 13 million records (as of June 11, 2014) later refined, to include only

Figure 13. The top six indexed journals (selected by impact factor more than 10) including articles strictly related to social representations



full research articles while excluding reviews, editorials or book chapters - indexed in Scopus across various disciplines between 2001 and 2011 interviewed by Halevi and Moed (2014) “to seek their insight about why they think their papers are as highly cited as they are”. In fact one of the three observations advanced by them on the basis of the author’s comments is:

“The analysis of citing disciplines shows that research, regardless of its disciplinary origin, crosses subject-specific domains and has impact on a wide range of areas, some of which are quite surprising. It is plausible that the growing ability of researchers to be exposed to and read a wider range of literature encourages the transfer of knowledge from one discipline to another.”

The other two observations concern: a) the applied value for the knowledge development of the description of computer software and b) the inter-institutional and international collaboration among researchers interested on issues of global concern:

“It is noticeable that 4 out of the 10 articles featured here describe the development of computer software. The practice of citing computer software when used in a study is a part of this phenomenon. Regardless of the subject field, the

computational tools developed and written about are highly cited.

Out of the 10 selected articles, 6 are the result of a scientific collaboration between two or more researchers. Collaboration is seen across institutions and countries which could be a result of a common global concern to damaging phenomena related to the environment.” (Halevi & Moed, 2014)

Due to page limits it is impossible to include here the full list of the 450 “indexed” journals in which the 1492 (46% of our corpus) articles has been published and that we have categorised in different classes according to their ranking in the bibliometric databases Isi-Web of Science and Scopus-Elsevier and also of the list of 624 “not indexed” journals, in which the 1747 (54% of our corpus) articles have been published. These results deserve a specific publication, providing both a contribution to the systematic geo-mapping of the editorial channels for the dissemination of the social representation theory and research literature around the world, for illustrating the dynamic of knowledge in context produced according to the generational level of the scientists, their language and cultural constrains, the inter-institutional collaboration within and between countries and

continents etc. and at the same time a strategic tool for orienting the new generation of early stage researchers, faced by the bibliometric academic culture era for their career perspectives.

Solutions and Recommendations

The social representations of the new bibliometric culture (as an articulated set of opinions, beliefs, attitudes, values systems, social meanings, shared practices, etc.) should be set in relation not only with the authors' bibliometric profiles, but also with other variables that characterize the individual's scientific style and his/her options among academic practices such as:

1. Scientists' individual preferences to publish articles, book chapters or books, also taking into account the relation with their academic roles and generational levels;
2. Their preferred publishing strategies (as single authors, in teams of collaborating co-authors, etc.);
3. Preferred language in which to publish (national, English, other);
4. Internationalization of the scientists' careers (international collaboration, international project collaboration or leadership, study period abroad, etc.);
5. Their preferred approach: experimental/quantitative vs qualitative (and to what extent exclusively);
6. Several indicators of their scientific productivity and institutional recognition (research grants, awards, leadership roles, appointment to international committees, editorial responsibilities and so on).

FUTURE RESEARCH DIRECTIONS

In the near future we intend to extend this investigation in several directions, from the analysis of “The Impact of the Impact” based on textual literacy sources to the analysis of “the social rep-

resentations of the bibliometric culture” among the authors themselves, by using multiple survey methods of a projective and structured nature (free associations, interviews, focus groups, questionnaires, conversational data extracted from academic social networks). We plan to conduct the study on our worldwide academic community, the So.Re.Com. THEmatic NETwork (de Rosa, 2006), including its training structure (the European/International Joint Ph.D. in Social Representations and Communication: de Rosa, 2015b, 2016a), as a case-study in the social sciences.

This investigation will be further develop a multi-year research program led by de Rosa (2013a, 2013b, 2016b; de Rosa, Dryjanska, & Bocci, 2017) guided by the main goal of evaluating *the impact of the scientific production driven by the Social Representations theory in the social arena faced with social demand*. The aim of this project launched in 1994 is to take stock of the scientific field developed in more than 50 years by conducting an empirical meta-theoretical analysis of the literature on Social Representations, mapping the development of different paradigms, the related research methods, the thematic areas and their impact on the various applied fields within the multi-generational community of scientists and across different geo-cultural contexts.

CONCLUSION

Given the importance of the on-going controversial debate on academic quality evaluation systems based on bibliometrics (*see additional readings*) and knowledge production using big data, we are confident that this investigation will attract the interest of the scientists belonging to our scientific community and motivate them to collaborate.

Taking into account the debate which still animates and divides the community of scientists and institutional leaders, differently affecting not only their personal careers but also their intergroup relations, and their personal, social and scientific identities, according to the discipline, the genera-

tional level, and anchoring in different geo-cultural contexts, our study – supported by empirical data – may stimulate meta-reflexive discussion and exchanges of views among the members of our scientific community on the preferred publishing options and collaborative strategies in the current editorial and academic scenario.

It is out of the doubts, that also for the very well-reputed informetricians, that “more research is needed into communication, publication and citation and evaluation practices in Social Sciences and Humanities (a “science of SSH”)” (Moed, 2014). This is even truer in the era of the Science 2.0, characterised by the progressive computerisation of the research process, which is going to deeply modify the research practices (Nielsen, 2012).

Our scientific community will not be exempted from being affected by these fast modified research practices and the pioneer study presented in this chapter will need to be periodically updated to show a clear dynamic picture of the impact of the Social Representation literature over the world and across the disciplines.

REFERENCES

- Boyd, D., & Crawford, K. (2012). Critical questions for big data. *Information Communication and Society*, 15(5), 662–679. doi:10.1080/1369118X.2012.678878
- de Rosa, A. S. (1994). From theory to meta-theory in S.R.: The lines of argument of a theoretical-methodological debate. *Social Sciences Information. Information Sur les Sciences Sociales*, 33(2), 273–304. doi:10.1177/053901894033002008
- de Rosa, A. S. (2006). Social Representation and Communication Thematic Network: A case study for monitoring the development of a scientific Community. In I. Labhrainn, C. McDonald Legg, D. Schneckenberg & J. Wildt (Eds.), *The Challenge of eCompetence in Academic Staff Development* (pp. 49-57). Galway: CELT, NUI.
- de Rosa, A. S. (2011). 50 anos depois: a ‘Psychanalyse, son image et son public’ na era do Facebook. In A. Almeida, Z. Trindade, & F. Santos (Eds.), *Teoria das Representações sociais - 50 anos*, (pp. 491-561). Technopolitik e Centro Moscovi, Brasília: MEMORES/UERJ, REPSO/UFES, LABINT/UFPE.
- de Rosa, A. S. (2013a). Taking stock: a theory with more than half a century of history. In A.S. de Rosa (Ed.), *Social Representations in the “social arena”* (pp. 1-63). Routledge.
- de Rosa, A. S. (2013b). Research fields in social representations: snapshot views from a meta-theoretical analysis. In A. S. de Rosa (Ed.), *Social Representations in the “social arena”* (pp. 89–124). New York: Routledge.
- de Rosa, A. S. (2014a). The *So.Re.Com.* “A.S. de Rosa” @-library: a digital tool for integrating scientific documentation, networking and training purposes in the supra-disciplinary field of Social Representations and Communication. In M. Khosrow-Pour (Ed.), *Encyclopaedia of Information Science and Technology* (pp. 260–271). Hershey, PA: IGI Global.
- de Rosa, A. S. (2014b). “Geo-mapping” Tutorial for visualizing statistical data on geographical maps. In A.S. de Rosa (Ed.), *Guidelines for the development of the So.Re.Com. “A.S. de Rosa” @- library*.
- de Rosa, A. S. (2015a). The Use of Big-Data and Meta-Data from the So.Re.Com A.S. de Rosa @-Library for Geo-Mapping the Social Representation Theory’s Diffusion over the World and its Bibliometric Impact. *9th International Technology, Education and Development Conference, INTED2015*, Madrid: INTED Publications.
- de Rosa, A. S. (2015b). The European/International Joint PhD In Social Representations and Communication: a Pioneer Case of Triple “I” (International, Interdisciplinary, Intersectoral) Networked Joint Doctorate. In *Inted 2015 Proceedings*. Madrid: INTED Publications.

de Rosa, A. S. (2016a). The European/International Joint PhD in Social Representations and Communication: a triple “I” networked joint doctorate, *Book of Proceedings of the 2nd International Conference on Development in Doctoral Education and Training* (Oxford, UK, 30th -31st March 2015) (D. Halliday & G. Clarke Eds.), pp. 47-60, London: Epigeum part of Oxford University Press, <http://www.ukcge.ac.uk/ICDDDET>. ISBN 978-0-9928799-3-8

de Rosa, A. S. (2016b). Mise en réseau scientifique et cartographie de la dissémination de la théorie des représentations sociales et son impact à l’ère de la culture bibliométrique. In G. Lo Monaco, S. Delouée, & P. Rateaux (Eds.), *Les représentations sociales* (pp. 51–68). Brussels, Belgium: Editions de Boeck.

de Rosa, A. S., & d’Ambrosio, M. L. (2008). International conferences as interactive scientific media channels: the history of the Social Representations theory through the eight editions of ICSR from Ravello (1992) to Rome (2006). *Rassegna di Psicologia*, 2, 153-207.

de Rosa, A. S., Dryjanska, L., & Bocci, E. (2017). Profiling authors based on their participation in Academic Social Networks INTED2017, (Valencia, SPAIN, 6-8th of March, 2017). In *Inted 2017 Proceedings*, (pp. 1061-1072). Madrid: INTED Publications. <http://library.iated.org/publications/INTED2017>

Halevi, G., & Moed, H. F. (2014). 10 years of research impact: top cited papers in Scopus 2001-2011. *Research Trends*, 38. Available from: <http://www.researchtrends.com/issue-38-september-2014/10-years-of-research-impact/>

Lo Monaco, G., Delouée, G., & Rateaux, P. (Eds.). (2016). *Les représentations sociales*. Brussels, Belgium: Editions de Boeck.

Mayer-Schönberger, V., & Cukier, K. (2013). *Big Data: A revolution that will transform how we live, work, and think*. Boston, MA: Houghton Mifflin Harcourt.

Moed, H. F. (2014). *New development in the evaluation of HSS research*. Workshop Internazionale “La valutazione della ricerca nelle Humanities and Social Science”, ANVUR (National Agency for the Evaluation of Universities and Research Institute), Rome, Italy. Retrieved from http://www.anvur.org/index.php?option=com_content&view=article&id=693:workshop-internazionale-la-valutazione-della-ricerca-nelle-humanities-and-social-sciences-it&catid=48&Itemid=363&lang=it

Moscovici, S. (1961/1976). *La psychanalyse son image et son public. Etude sur la représentation sociale de la psychanalyse*. Paris, France: Presses Universitaires de France.

Nielsen, M. (2012). *Reinventing discovery: The new era of networked science*. Princeton, NJ: Princeton University Press.

Parks, M. R. (2014). Big Data in Communication Research: Its Contents and Discontents. *Journal of Communication*, 64(2), 355–360. doi:10.1111/jcom.12090

Sammur, G., Andreouli, E., Gaskell, G., & Valsiner, J. (2015). *The Cambridge handbook of social representations*. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9781107323650

Zupanc, G. K. (2014). Impact beyond the impact factor. *Journal of Comparative Physiology. A, Neuroethology, Sensory, Neural, and Behavioral Physiology*, 200(2), 113–116. doi:10.1007/s00359-013-0863-1 PMID:24264238

ADDITIONAL READING

Adam, D. (2002). Citation analysis: The counting house. *Nature*, 415(6893), 726–729. doi:10.1038/415726a PMID:11845174

Allen, L., Jones, C., Dolby, K., Lynn, D., & Walport, M. (2009). Looking for landmarks: The role of expert review and bibliometric analysis in evaluating scientific publication outputs. *PLoS ONE*, 4(6), e5910. doi:10.1371/journal.pone.0005910 PMID:19536339

- Bar-Illan, J. (2008). Which h-index? A comparison of WoS, Scopus and Google Scholar. *Scientometrics*, 74(2), 257–271. doi:10.1007/s11192-008-0216-y
- Batista, P. D., Campiteli, M. G., & Kinouchi, O. (2006). Is it possible to compare researchers with different scientific interests? *Scientometrics*, 68(1), 179–189. doi:10.1007/s11192-006-0090-4
- Bornmann, L. (2010). Towards an ideal method of measuring research performance. Some comments to the Opthof and Leydesdorff (2010) paper. *Journal of Informetrics*, 4(3), 441–443. doi:10.1016/j.joi.2010.04.004
- Cronin, B., & Sugimoto, C. (2014). *Beyond bibliometrics. Harnessing multidimensional indicators of scholarly impact*. Cambridge, MA: MIT Press.
- Das, A. K., Arora, P., & Bhattacharya, S. (2012). Webliography of STI Indicator Databases and Related Publications. *Journal of Scientometric Research*, 1(1), 86–93. doi:10.5530/jscires.2012.1.14
- De Bellis, N. (2009). *Bibliometrics and Citation Analysis: From the Science Citation Index to Cybermetrics*. Lanham, MD: Scarecrow Press.
- DORA. (2012) San Francisco Declaration on Research Assessment. Putting Science into the Assessment of Research. <http://am.asch.org/dora/>
- Ernst, R. E. (2010). The follies of citation indices and academic ranking lists. A brief commentary to Bibliometrics as weapons of mass citation. *Chimia*, 64(1/2), 90. doi:10.2533/chimia.2010.90 PMID:21137692
- Fanelli, D. (2010). Do pressures to publish increase scientists' bias? An empirical support from US states data. *PLoS ONE*, 5(4), published online 21 April.
- Garfield, E. (1955). Citation indexes for science. A new dimension in documentation through association of ideas. *Science*, 122(3159), 108–111. doi:10.1126/science.122.3159.108 PMID:14385826
- Garfield, E. (2006). The history and meaning of the journal impact factor. *Journal of the American Medical Association*, 295(1), 90–93. doi:10.1001/jama.295.1.90 PMID:16391221
- Kurtz, M., & Bollen, J. (2010). Usage bibliometrics. *Annual Review of Information Science & Technology*, 44(1), 1–64. doi:10.1002/aris.2010.1440440108
- Moed, H. F. (2002). The impact-factors debate: The ISIs uses and limits, correspondence. *Nature*, 415(14), 731–732. doi:10.1038/415731a PMID:11845176
- Moed, H. F. (2005). *Citation analysis in research evaluation*. Dordrecht: Springer.
- Moed, H. F. (2010). Measuring contextual citation impact of scientific journals. *Journal of Informetrics*, 4(3), 265–277. doi:10.1016/j.joi.2010.01.002
- Molinié, A., & Bodenhausen, G. (2010). Bibliometrics as Weapons of Mass Citation. *CHIMIA International Journal for Chemistry*, 64(1), 78–89. doi:10.2533/chimia.2010.78 PMID:21137691
- Moskovkin, V. M., Bocharova, E. A., & Balashova, O. V. (2014). Journal benchmarking for strategic publication management and for improving journal positioning in the world ranking systems. *Campus-Wide Information Systems*, 31(2/3), 82–99. doi:10.1108/CWIS-11-2013-0066
- Seglen, P. O. (1997). Why the impact factor of journals should not be used for evaluating research. *British Medical Journal*, 314(7079), 498. PMID:9056804
- Torres-Salinas, D., Cabezas-Clavijo, A., & Jiménez-Contreras, E. (2013). *Altmetrics* New indicators for scientific communication in Web 2.0. *Comunicar*, 21(41), 53–60. doi:10.3916/C41-2013-05
- Van Raan, A. F. J. (2005). Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods. *Scientometrics*, 62(1), 133–143. doi:10.1007/s11192-005-0008-6

KEY TERMS AND DEFINITIONS

Bibliometric Indexes: Indexes calculated on the basis of the number of citations and other factors, applied to scientific journals, and gradually also to other types of scientific production, such as conference abstracts and books. Among the main bibliometric indexes: Impact Factor (Isi-web of Science by Thomson and Reuters) and SJR (Scopus-Elsevier).

Big Data: According to some sources (Boyd & Crawford, 2012; Mayer-Schönberger & Cukier, 2013) data volumes in the range of petabytes and beyond, which exceed the capacity of most current online storage and processing systems; critical issues related to big data include their volume, velocity, variety, value and veracity. However according to Parks (2014: 255) data volume does not qualify in itself the “big data”, because the criteria should be relative to the “representativeness” of the specific field analysed.

Data Mining: Application of specific algorithms for extracting patterns from data; the term primarily used by statisticians, data analysts and the management information systems (MIS) communities who find useful patterns in data.

Geo-Cultural Context: Specific geo-location that takes into account the cultural characteristics of inhabitants, such as common language, history, customs or art; for example, insular, Mediterranean, North European, Latin American geo-cultural contexts.

Modelling Approach: A paradigmatic approach to social representations developed by de Rosa (2013a, 2013b) that integrates different methods and techniques based on multiple communicative channels (oral, textual, iconic, behavioral, etc.) coherently with the articulation of different theoretical constructs (like social representations, place-identity, social memory, attitudes, prior knowledge and channels of information, etc.) and dimensions (informational, evaluative, emotional, imaginal/experiential, etc.).

So.Re.Com. “A.S. de Rosa” @-Library: A multi-purpose web-platform for integrating

scientific documentation, networking and training in the field of Social Representations and Communication (So.Re.Com.) (de Rosa, 2014a).

Social Representations: A theory developed by Serge Moscovici (1961/1976), which – according to de Rosa (1994) can be operationalized on three different levels: a) social representations as *phenomenon* – “ways of knowing” characteristic of social reality that emerge in daily life during interpersonal communication and are directed toward comprehension and control of the physical-social environment; b) *theory* of social representations – the collection of conceptual definitions, methodological operations and formulation of constructs that have social representations as their object; c) *meta-theory* of social representations – the collection of critical comments, ripostes and comparisons with other theoretical models which emerges from the critical debate on the theory of social representations. For a stock of the wide scientific field developed in more than 50 years since 1961 see de Rosa (2011, 2013a), de Rosa & d’Ambrosio (2008).

ENDNOTES

¹ The source of data presented in this section is equal to 3234 rather than 3239 articles, because it was impossible to identify the institutional affiliation for 5 authors and therefore to include them in the geo-mapping.

² At the time of writing this article (March 2016) it has been possible to identify the country of the author’s Institution affiliation for 8348 items of the 8740 references specifically related to social representations and communication (including also books, book chapters, conference presentations, web documents, manuscripts, university reports, Master and PhD theses, etc.) filed in the repositories of the *SoReCom “A.S. de Rosa” @-Library* and therefore to proceed with statistical analysis the ranking by

continents and countries in the production/ dissemination of the literature.

- ³ It should be noted that currently Scopus-Elsevier and WoS start to cover - at different extent - also books and book chapters upon review of quality standard of publishing houses, but - despite the big investment es-

pecially by Scopus in the field of humanities and social sciences – the sources are still limited.

- ⁴ Gaskell, G. (1997). *Europe ambivalent on biotechnology*. *Nature*, 387, 845–847. <http://www.nature.com/nature/journal/v387/n6636/full/387845a0.html>

Information and Its Conceptual Perspectives

José Poças Rascão

Institute Polytechnic of Setúbal, Portugal

INTRODUCTION

Information has attributes/positivist concepts that some authors use to better understand it. Others are critical about information subjectivism. This chapter discusses conceptual perspectives of information. The aim is to study the concepts of information from some areas of knowledge of the social sciences, cognitive and business looking to propose a concept of information in the field of information science. The methodology of the study is formed by bibliographical research. It is concluded that the concept of information is still in full development.

The concept of information depends on the perception of information (Kirk, 1999). Although this is not something that shocks, raises some interesting questions and research opportunities. For example, as we get the information? What are the claims, limits and consequences of these perceptions? Can these perceptions be described and why?

Fundamental Concepts

The perception of information¹ not only influences our view of information, but also our perception of information system (Klein & Hirschleim, 1987), our perception of communication (Mokros, 1993, Schement, 1993) and the conduct of research (Newman, 2001, Schement 1993). This means that the perception of information, which we prefer to call information concepts, have a profound influence in the field of information science.

The information concept fascinates many scientists from different fields such as biology,

psychology, computer science, sociology, economics, management, political science, statistics, philosophy, communication and information studies (Mokros, 1993, Newman, 2001, Ruben 1993, Schement, 1993). In all these fields the information is an important concept, but at the same time none of them can claim the information as being relevant only for them.

The information should be viewed as an interdisciplinary concept. This means that the concepts of information must be studied in different disciplines. It also means that the concepts of information are not only relevant in the field of information science.

On the concept of inter-disciplinarity of information no deal has emerged and no unifying theory is presented as imminent (Schement, 1993). When information is defined “the abundance and diversity confuse us” (Braman, 1989, p. 233). A tempting conclusion that we reached is that the meaning of information depends on the context. While many argue that we need a theoretical perspective of information (Devlin, 1999, Aefiner, 1999, Newman, 2001). We do not intend to define a theoretical perspective, but only present the different concepts in different disciplines, as well as a critical analysis of the different concepts.

Newman (2001) describes a variety of concepts in different sciences that can be grouped as follows:

- Probabilistic concept;
- Concept of information processing;
- Ecological concept of info;
- Social and organizational concept of information.

The probabilistic concept of information is that low-probability events represent high information content. An important application of this concept is the information theory Shannon and Weaver (1949, in: Newman, 2001). In this theory the mathematical representation of the transmission of a message is presented as if the information was a measure of predictability.

Logic, cybernetics and philosophy also correlate the information with the probability (Fisher, 1934, Carnap & Bar-Hillel, 1952, Popper, 1965, Mackay, 1969 in: Newman, 2001). But these concepts differ in important ways, as for example, in the interpretation of probability and on the semantics of the information. With respect to the semantics of the information, many concepts see the information as reduction of uncertainty.

The concept of information processing (or cognitive concept) focuses on the thought of cognitive psychology. However this concept, thinking and information processing are seen as analogous. It is clear that the information is the product of thought and that increases knowledge about anything. The model of the cognitive process and the internal representation are the first concern of this approach.

The concept of ecological information is not created, but is present in the world, from the environment, in a given situation. Organizations collect this information actively from the outside world. An important extension of the ecological approach is the situation theory. This reconstructs itself in a mathematical basis, and makes a clear distinction between information (content or information) and its representation.

The social and organizational concept of information falls within the sphere of labour: work associated with the concept of information economy. In this category, the information relates to the processing of the same and the information pyramid model is often used. In this model a data must be processed to produce information and the information should be processed to produce knowledge.

An important ingredient of the information economy is the quantification of “work information” and “information product”, used among other things to show the importance of knowledge in the modern economies (Wallerstein 2000, Murteira, 2001, Brandt, 1995, Nicholas, 2000, Handy, 1990, Hauknes, 1999). In the well-known effort of Porat (1997, in: Newman, 2001) it is clear that the information is associated with the reduction of uncertainty.

The research of information science focuses on the process of information in the Organization and the need for information of managers in support of decision-making. The satisfaction of this requirement can result in a reduction of uncertainty, which contributes to a better decision-making (Schement, 1993).

Philosophical Concept of Information

Belkin (1978) contributed with many studies for an important problem of information science: the question of the definition of an appropriate concept of information for information science. Although Belkin discusses the concepts of information used only in information science, many of these concepts were originated from other fields and/or are used in a wide variety of these (Belkin, 1978, p. 82):

- **Information as Fundamental Category:** The information is seen as something that is essential to the existence of the universe, as the base, but a different category of matter;
- **Information as Property of Matter and Consciousness:** The information is not regarded as a special category, but as property of matter (i.e., objective information) and or property of conscience or reflection of an individual (that is, subjective information);
- The information as a social-scientific information is based on the classification

of Mikhailov, Chernyi and Giliarevskii (1975, in: Belkin, 1978). This classification divides the intuitive idea of social and non-social information, semantics social information and not semantic and scientific and unscientific semantic information. According to Mikhailov, Chernyi and Giliarevskii, the information is limited by the social science;

- **Information as Event:** Information is seen as the expression of the mental image that occurs when we receive a message;
- **The Information Framework:** The information is seen not as an event, but as the resulting structure for this event. Example, information is the resulting structure in the mind of a sensory data or some experience;
- The information as a probability of occurrence of an event comes from Information theory of Shannon and Weaver (1949);
- **Information as Message:** Vague concept in which the information is confused with the content of a communication.

Concept of Information in the Context of Decision-Making

Braman (1989) suggests a hierarchy of information definitions that are used in the context of decision-making. The hierarchy is based on three dimensions: the level of opportunity, the level of complexity and associated with the power (which is guaranteed for the info, streams and use). These dimensions group the definitions of information into four groups:

- Information as a resource;
- Information as something useful;
- The information as a standard perception;
- The information as an essential feature of society.

Information as a resource is associated with the lowest level of opportunity, complexity and power, while the information as essential feature

of society is associated with the highest level in these three dimensions.

- Information as a resource treats the information as an isolated and distinct entity without power in itself. The information is divided into the parts that make up the body of knowledge or information flows in which can be organized (Braman, 1989, p. 236).
- The information as useful thing focuses on the process of exchange of information between people. This concept requires the chain production, through which the information gains economic value (Porter, 1980). The chain includes steps such as the creation, processing, and distribution. This implies a greater complexity of social structure “including suppliers, customers and the Organization in order to keep the market” (Braman, 1989, p. 238).
- Information as a standard perception requires the information and the context. The information “has a past and a future, is affected by the stimulus of casual factors and the environment” (Braman, 1989, p. 238). Compared to information as useful thing, the scope of the phenomenon covered by this concept is extended. The information can be used to articulate social structures. This definition sees the information as an element of reducing uncertainty.
- Information as essential feature of information-oriented society as “an active function built in context” (Braman, 1989, p. 239). The information becomes an actress who affects the environment and creates a social structure.

This definition treats the information as an essential feature of society. Is applied to all phenomena and processes in which information is involved and can be applied to social structure with some degree of articulation and complexity (Braman, 1989, p. 241).

Concept of Information Process, as Knowledge and as Thing

Buckland (1991) identifies three “primary uses” of the term information:

- The information as a process;
- Information as knowledge;
- Information as thing.

The information process refers to the act of informing/be informed. When someone is reported, what he/she knows is changing. The “information as knowledge” refers to what is perceived as process information. Is the knowledge that is communicated. Buckland (1991) see information as reduction of uncertainty, as a special case of “information as knowledge”.

Some information increases the uncertainty. Information as thing refers to things that are looked at as being informative, things become informed. Buckland also examines different things (data, text, subject material, events) and concludes that the whole thing is or should be informative. Argues that the virtue of being the information as thing is situational and depends on subjective judgments.

Buckland (1991) summarizes the main tree concepts of information, in terms of two distinctions:

- Between entities and processes;
- Between intangible and tangible assts.

Buckland (1991) distinguishes four aspects of information, but only three regarding the use of information. The fourth aspect of information is information processing. This refers to the execution, manipulation and deduction of new forms or versions of information, as a thing.

Concept of Inter-Disciplinarily of Information

Ruben (1992, 1993) has different propositions to “provide an inter-disciplinarity in the relationship information-communication” (Ruben, 1992, p.

22). Ruben does not justify this classification and does not refer any example of these concepts in the literature.

However sees the information as an inter-disciplinary concept focused on the relationship between information and communication. Here explicitly outlines a wide variety of fields, such as biology, economics, and Cybernetics, mathematics, sociology and communication studies:

- The information has potential significance for a living system, but this potential is not yet updated;
- The information is that which has been transformed and configured for use by an individual;
- The information includes the sharing of information/knowledge base of society and other social systems.

Concept of Information in Communication

Schement (1993) reviews 22 definitions of information from different fields, such as economic, physical, information and communication and information science. Although his focus is the study of information and communication, he parses their inter-disciplinarity, because the definitions of the different fields are compared. On the basis of these definitions distinguishes “fundamental terms whose sketch of current of thought is the nature of the information” (Schement, 1993, p. 7).

Information as thing treats the information as a thought, being a thing (non-material). According to Schement this concept is the most widely used of the three concepts. Two examples of this concept are:

- “information is an entity; but a thing which has no mass or energy” (Diener, 1989 in: Schement, 1993);
- The information is a consistent collection of data or messages organized that have meaning or may be used by the human system (Ruben, 1988 in: Schement, 1993).



The information can also be a process. This concept of information process sees the information as the phenomenon to inform or to change a given situation. An important sub-theme of this concept is the vision of the information as the reduction of uncertainty, a common view among economists, managers and computer scientists.

The last concept, the information as the product of manipulation is seen as thought is something that must be manipulated to exist. Example:

Information is produced, as a result of a process on the data. (Hayes, 1969, in: Schement, 1993)

According to Schement these perceptions of information are related to different perceptions of communication. He argues that these two concepts are inextricably bound by each other.

Information Concept in the Real World

According to Gelepithis (1999) information is the central concept for the community of information sciences. A considerable number of disciplines related to information has been engaged in the development of other closed concepts, related to information (ex: sign, symbol and meaning (Shannon and Weaver, 1949, in: Newman, 2001).

Gelepithis (1999) is concerned with the clarification of these concepts and their consequences in the fields of information science. However its proposal is not present in the table of contents of the various concepts of information in different disciplines. Gelepithis (1999) presents seven concepts of information:

- The information in terms of the probability of a signal;
- The information as a state;
- The information in terms of knowledge and meaning the mental level and as a mental entity and not material;
- The information in terms of concept of the sign as a primitive;
- The information conceived in terms of the world tree;
- The information in terms of true condition;
- The information as a basic property of the universe.

The problem with these concepts is that they are very brief. The information in terms of sign as a primitive is referred to by Stamper (1985) in that he proposes the semiotics (sign theory), as an appropriation of the information theory (Shannon and Weaver, 1949). He argues that the idea of a sign is “the very primitive in which is based the information science” (Gelepithis, 1999, p. 195).

The signals can be described as physical things-objects, events, or properties of objects and events — and are available to represent a function in human behaviour. The information is in fact a measure of some property of a signal. The measures differ from each other (e.g. entropy measure and subjective measure) and in addition information has different meanings.

The information conceived in terms of Popperian world tree design is the basis of Popper & Eccles (1977) they argue that we just accept things as real if they can interact with material things. He distinguishes three realities or three worlds:

- 1 World:** The world of physical objects and states;
- 2 World:** The world of states of consciousness (e.g., subjective knowledge, experience of creative imagination);
- 3 World:** The world of knowledge towards objective (e.g. products of the human mind, theoretical systems, scientific problems).

According to Popper these worlds interact with each other. However still unclear what Gelepithis (1999) is mean by information conceived in terms of these three worlds.

CRITICAL ANALYSIS OF THE CONCEPTS OF INFORMATION

Philosophical Questions

It is important to make the distinction between information and its representation, being one of the most common misconceptions confuses information with its representation (Devlin 1999, p. 27). Buckland (1991) states that the representation is often seen as information and Ruben (1993) stresses that seeing the representation as information is “appropriate and useful for some purposes.” So the information as something material (representation) can be seen as an important concept.

Atmanspacher (1999) and Fuchs-Kittowski (1999) describe the information as something material or not, and the philosophical question is related to the distinction between material and non-material². Information as material thing is related to the human mind and is seen for example by:

- Buckland (1991) – Information as thing;
- Ruben (1993) – Information about the surrounding environment;
- Belkin (1978) – Information as property of matter.

The concept of information as useful and as a resource is basically variations of information as something material. According to Haefner (1999) information has a material level, but in reality does not exist, though Buckland (1991) claims that we often use the word information as if it were something material.

The concept of information as mental thing is often equalled to knowledge, (Buckland, 1991, Ruben, 1993, 1978, Belkin Gelepithis, 1999, Anderson, 1980, in: Newman 2001). According to Miller (2002) this is also a habit in the literature of knowledge management. He alerts us to a vision because the information itself has no intrinsic meaning. Information can provoke or evoke the thoughts of others³.

There are also some researchers who relate explicitly to information as something different besides of matter and mind. For example for Devlin (1999) “information is not physics nor is in the mind of the other: our thoughts are located inside of our heads, but the information is out of it.

The information exists somewhere between the physical and the mental world. It is reasonable and wonderful to see the information as something that exists in the public domain. Gelepithis (1999) and Belkin (1978) also refer to something other than matter and mind.

Gelepithis (1999) presents a concept in which the information is a basic property of the universe and Belkin (1978) introduces the concept in that the information is a key resource, such as the base. Another example is Fuchs-Kittowski (1999) and Zikmund (2000) who argue that the information is neither in the article nor in the mind, but makes the connection between the physical environment and the ideal.

One can criticize or comment these visions, but we just summarize the possible views as regards to the position of the information found in the literature:

- Information as subject material;
- The subject of mental information;
- Information as an abstract entity (not material or mental)

We hold these positions as information changes as thing; the information is treated as an object, an entity, a thing. The three variations can be found in the literature: the Based on the similar description between information concepts groups we concluded that the information as thing information and as process are the basic concepts of information, is informational, a mental object and an abstract entity referred to the context.

The Information vis-a-vis Social

The constructed nature of information is also present on the concepts of Braman (1989) “perception

of a pattern” and “constitutive element of society”. Also the concept of Ruben (1992) refers to the social nature of the information. The information includes the sharing of information/knowledge, bases of society and other social systems. Belkin (1978) refers to “relational system of scientific facts, theories, hypotheses, and laws, publications that are the result of human observation, of nature and of itself”.

When we removed the restriction to scientific information we obtain greater resemblance to the concept of Ruben. Also the world of Popper 3 (1977) refers to this concept of information (Tully 1985). On the basis of this research we suggest the third basic concept of information, information as social construction. This concept refers to the sharing and building the information base of the social system.

Information as Probability

Belkin (1978) and Gelepithis (1999) distinguish the concept of information related to the likelihood and both refer to Shannon (1949). In the formulation of Shannon information as feature selects the desired message out of possible messages. The amount of information associated with the selected message is related to the probability of this message be sent. The expected message does not result in information⁴.

The concept of Shannon (1949) has been used by many other fields, more than was originally proposed: telecommunications. At the same time the use of the concept of information of Shannon’s is much criticized (Belkin, 1978, Borgman, 1999, Bryant, 2001, Callaos, Callaos, B., (2002) because it has a very limited vision of the concept of information⁵.

Sveiby (1998) points out that Shannon was unhappy with the word information on his theory and he was advised to use the word “entropy” rather than information. To Shannon information equals “entropy”, i.e. chaos/confusion. This contradiction with Wiener (1948, in: Sveiby, 1998) takes into account that the information is the same as

the structure (“negative entropy”)⁶, that is, “to play the world’s structured” (Sveiby, 1998, p. 4).

Although the concept of Shannon information is criticized has also been used by some researchers as a point of departure. However we consider the concept of Shannon information, as an information base concept. The concept of information as structure, which seems to contradict the concept of Shannon, has been found in Belkin (1978) and Gelepithis (1999).

Concepts of Information Related to Information as Thing

We argue that the concepts presented earlier might be related with the information like this: the information as the product of manipulation (Schement, 1993)), the information as a resource, the information as convenience (Braman, 1989), the information as data and the information as message herself (Belkin, 1978).

Information as the Product of Manipulation

We see the concept of Schement information (1993), as the product of manipulation, as an example of how information thing. Remember the description of this concept: “one thing that should be handled with the purpose to exist”. With this thing a process can be associated; in this case a process of manipulation.

Information as a Resource and as a Convenience

Form the description of Braman (1989) it can be concluded that the information as a resource and as a convenience can be seen as an example of information as thing (Lauer 2001, p. 43). Outlines the information as a resource, as an entity, as a commodity and as a resource with economic power⁷.

According to Braman (1989) the focus of the attention of the information as a convenience is the process of changing the information. In other

words, when we see information as something we can still focus on the procedural aspects associated with this thing.

The Message Itself

Belkin (1978) proposes the concept of information as “the message itself”. This concept is not described clearly, but seems to refer to “message content”. In terms of basic concepts this means the information as no material thing (e.g., mental object or abstract).

Data, Information, and Knowledge

Although the concepts of data and information with interchangeable frequencies are used, these are not the same thing. Information is not the same thing as data, although the two concepts are often confused by the subtle distinction between the two concepts is essential.

The data does not carry any sense or meaning of the facts, images or sounds, since they lack relational elements essential to the establishment of a complete meaning, requiring a relational structure internal to a cognitive purpose. The data become information when its creator joins them meaning (Davenport and Prusak, 1998).

Data are partial representations of facts, images or sounds, have no meaning by themselves, since they do not lead to the understanding of the facts or situations and converted into information, when introduced into an information structure (global) acquired (Steven Alter, 1992, Davenport and Prusak, 1998, Keith Devlin, 1999).

The concept of information as is commonly seen in the field of scientific research. In this vision of the information as given with meaning attributed (Checkland and Howell, 1998, p. 95) or simply: $\text{information} = \text{data} + \text{meaning}$ ^{8,9} (Devlin, 1999, p. 33).

While it seems logical to think that this meaning is given by humans, Checkland and Howell (1998) point out that some information in the scientific

field definitions assume that the machine assigns meaning. This suggests that there is only a single meaning related to the information.

The concept of information can be understood under very different perspectives. The information is an object created by man, having the purpose to represent an identifiable event for him in the real world, integrating and relating a set of records or data, (Le Moigne, 1979, Steven Alter, 1992). Is the difference that makes the difference (Bateson, 1972).

In terms of equation:

$\text{Information} = \text{data} + \text{meaning}$

When a person internalizes the information to the point that one can use, we call it *knowledge* (Zikmund, 2000). This is a fluid mixture of experience, values, contextual information and expert insight, structured to provide a framework for evaluating and incorporating new experiences and information.

Knowledge is information as valid and accepted, integrating data, actions, information and sometimes hypotheses. The knowledge needs someone triages combine and interpret the information. The information can be considered as a “*substance*” which can be acquired, stored and possessed by a person or group and transmitted from person to person or from group to group.

$\text{Knowledge} = \text{Information internalized} + \text{ability to use}$

Knowledge is thus a mixture of information, experience and understanding which provide a structure that can be applied in the evaluation of new information or new situations (William Zikmund, 2000).

The knowledge is fundamentally and inherently inside people. These are much more complex and unpredictable individual level than an entire society, so it is not surprising that knowledge work much harder to get that information. The



knowledge exists primarily within the people, is an integral part of the human unpredictability and complexity (Davenport and Prusak, 1998).

Reduction of Uncertainty

The reduction of uncertainty is often related to the concept of probability. The probabilistic approach more the vision of reducing uncertainty with the function semantics; the effect of information is the reduction of uncertainty. This reduction depends on the likelihood of the message being sent.

The concept of uncertainty reduction of Belkin (1978) sees the reduction of uncertainty as a particular effect; It changes the State of knowledge. Braman (1989), Buckland (1991) and Schement (1993) see the reduction of uncertainty as an instant or an example of the concepts proposed by them.

All these aspects seem to be necessarily related with the sense of uncertainty. Only based on linguistic Association we see the reduction of uncertainty as a mental process and so the concept of information as uncertainty reduction as a special case of case information as a process.

Information as Structure

Belkin (1978) distinguishes information as structure and Gelepithis (1993) distinguishes information as requested. Gelepithis does not make the distinction of this concept, is unclear about what is this concept. It seems to us that order and structure refer to the same thing.

Belkin (1978) refers to an effect of information, that is, the change in structure. This change in structure can occur in multiple “systems” (Belkin, 1978, p. 79), for example (or structured container image). In the case of the structure of the mind, this concept refers to something mental; the mental thing or the mental process. In the description of Belkin (1978) remains unclear what he meant with the structure of other “systems”.

It is important to note that this concept has no relationship with the Wiener (1948), the information structure (Belkin, 1978, p. 79), which

describes the information as “a structured part of the world”. This description of Wiener (1948) seems to see the information as a thong. Sveiby (1998) points-in the aftermath of the concept of Wiener, which is often suggested by “adding value” or selection, interpretation and information update, the highest of the information (e.g. knowledge) that can be retrieved¹⁰.

True Condition

Gelepithis (1999) proposes the concept of “information in terms of real condition”. It is not clear what he means by this. Other researchers do not describe this concept, by which we consider not to be very important and is not a basic concept.

Signal as Primitive

It explain briefly what the meaning of signal as primitive is and its relationship to information. Information has different meanings, depending on the area of semiotics. On the basis of Stamper (1985) is the description of these meanings. We argue that the information as probability is an important empirical concept. The information as given (which relate to the information like thing) is an important concept for the syntax.

For the semantics we suggest the information as mental process that is pragmatic and even information as social construction is an important concept. So the concept of information as a sign of a primitive is not in fact a simple concept; It is a basic principle of semiotics and organization work of various ideas with regard to information.

Popper

Gelepithis (1999) also proposes the concept of “information conceived in terms of design Properian of the three words”. Previously we explain briefly these three words and we would like to remind you that it is not clear what Gelepithis (1999) mean by this concept. Although this is true, it is appropriate to recall that the basic concept of the

three trees can be related to these three words of Popper (World 1, World 2, world 3).

Information as something becomes a material variation of the world 1, information as something becomes a variation of information process for the world 2 and information how thing is an abstract variation of information as social construction into the world 3. Information only as probability we do not know where to place it.

Some concepts are used in various areas (e.g. the reduction of uncertainty) and may be related to more than one basic concept. Because most concepts are related to the basic concepts and therefore argue that these fundamentals reflect (or feature or summarize) previously concepts of information.

We argue that the basic concepts reflect the most important uses of information in the literature. The main argument is that the concepts presented in previously (in which the basic concepts are based) are in our opinion a pure reflection and understanding of the areas in which the information is perceived in the literature.

FUTURE RESEARCH DIRECTIONS

The discussion presented aimed to propose the understanding of the concept of information, in the context of information science under the systemic perspective, whereas epistemological, scientific and practical.

Although you can delimit the relevant study and distinguishable object, the information, to the information science is essential to establish an epistemological justification of the concept of information, as a basis for the understanding of this object, in all areas of knowledge.

The feature of information is its inter-disciplinarity character, typical of areas of expertise that arise in the context of Post modernity. Therefore, just as in information science, methods, models and theories in the area are derived from other areas of knowledge.

The discussions raised on the concept of information are of the utmost importance to the information science and society. Who knows, therefore, that the work does not end here. This is only their embryonic stage.

FINAL CONSIDERATIONS

The concept of information has been consolidated as a positivist paradigm in information science with the incorporation of theories, concepts and methods of various areas of expertise and manifests itself in various subareas that comprise the information science.

The paradigm shares with positivism all its features, as well as its limitations, being the principal the inability to capture what the method can not handle to grasp: the subjective information, with various interpretations and inserted into the historical-cultural experience.

Several studies show the complexity of the phenomenon-information-to theories proposing to study it, including information science, on the one hand, and its binding to the humanities and social sciences, on the other. The positivist nature studies reaffirm the concept of information in objectivist perspective, without consideration the subject and the socio-cultural contexts, which takes the information as given and not as a construction.

REFERENCES

- Alter, S. (1992). *Information System - A Management Perspective*. Addison-Wesley.
- Atmanspacher, H. (1999). Cartesian cut, Heisenberg cut and the concept of complexity. In *The quest for a unified theory of information: Proceedings of the second international conference of the foundation science*. Gordon and Breach Publishers.
- Bateson, G. (1972). *Steps to an Ecology of Mind*. Nova Iorque, Balentine Books.

- Bausch, K. (2000). The *evolution of embodied information: a mosaic*. OLD Saybrook 2 Conference at State University of West Georgia.
- Belkin, N. J. (1978). Information concepts for information science. *The Journal of Documentation*, 34(1), 55–85. doi:10.1108/eb026653
- Braman, S. (1989). Defining information: An approach for policymakers. *Telecommunications Policy*, 13(3), 233–242. doi:10.1016/0308-5961(89)90006-2
- Buckland, M. (1991). Information as thing. *Journal of the American Society for Information Science*, 42(5), 351–360. doi:10.1002/(SICI)1097-4571(199106)42:5<351::AID-ASI5>3.0.CO;2-3
- Callaos & Callaos. (2002). Toward a systemic notion of information: Practical consequences. *Information science*, 5(1).
- Checkland, P., & Holwell, S. (1998). *Information, systems and information systems: making sense of the field*. John Wiley & Sons.
- Davenport, T., & Prusak, L. (1998). *Working Knowledge*. Cambridge, MA: Harvard University Press.
- Devlin, K. (1999). *Info Sense: turning information into knowledge*. New York: W. H. Freeman and Company.
- Fisher, A. G. B. (1934). *The Clash of Progress and Security*. London: Kelley.
- Fuchs-Kittowski, K. (1999). Information – neither matter or mind: on the essence and on the evolutionary stage conception of information. In *The quest for a unified theory of information: proceedings of the second international conference of the foundation science*. Gordon and Breach Publishers.
- Gelepithis, P. A. M. (1999). A rudimentary theory of information: Consequences for information science and information systems. In *The quest for a unified theory of information: proceedings of the second international conference of the foundation science*. Gordon and Breach Publishers.
- Haefner, K. (1999). Foreword. In *The quest for a unified theory of information: proceedings of the second international conference of the foundation science*. Gordon and Breach Publishers.
- Handy, C. (1990). *The Age of Unreason*. Boston: Harvard Business Review.
- Kirk, J. (1999). Information in Organizations: Directions for Information Management. *Information Research*, 4(3). Retrieved from <http://Informationr.net/ir/4-3/paper57.html>
- Klein, H. K., & Hirschleim, R. A. (1987). Social change and the future of information systems development. In R. J. Boland & R. A. Hirschleim (Eds.), *Critical issues in information systems research*. John Wiley & Sons.
- Lauer, T. W. (2001). Questions and Information: Contrasting metaphors. *Information Systems Frontiers*, 3(1), 41–48. doi:10.1023/A:1011497503953
- Le Moigne, J. L. (1979). Informer la décision ou décider de l’information. *E.S./ Sciences de Gestion*, (1), 863-888.
- Mackay, D. M. (1969). *Information, Mechanism and Meaning*. Cambridge, MA: MIT Press.
- Miller, F. J. (2002). I = 0 Information has no intrinsic meaning. *Information Research*, 8(1). Retrieved from <http://Informationr.net/ir/8-1/paper140.html>
- Mokros, H. B. (1993). The impact of a native theory of information on two privileged accounts of personhood. In J. R. Schement & B. D. Ruben (Eds.), *Between communication and information*. Transaction Publishers.
- Murteira, M., Nicolau, I., Mendes, V. E., & Martins, A. (2001). *Serviços Informacionais e transição para a economia do conhecimento em Portugal*. IDEG/ISCTE, estudo realizado para o GEPE do Ministério da Economia.
- Newman, J. (2001). Some observations on the semantics of information. *Information Systems Frontiers*, 3(2), 155–167. doi:10.1023/A:1011439305567

Popper, K., & Eccles, J. C. (1965). La ciencia normal y sus peligros. In *La crítica y el desarrollo del conocimiento*. Grijalbo.

Popper, K., & Eccles, J. C. (1977). *The self and its brain: an argument for interactionism*. London: Taylor & Francis Group. doi:10.1007/978-3-642-61891-8

Porat, M. U. (1997). *The Information Economy*. Publication Spatial OT.

Porter, & Millar. (1985, July-August). How Information Gives You Competitive Advantage. *Harvard Business Review*.

Prusak, J., & McGee. (1995). *Gerenciamento Estratégico da Informação*. Editora Campus, Rio de Janeiro.

Ruben, B. D. (1992). The communication-information relationship in system-theoretic perspective. *Journal of the American Society for Information Science*, 43(1), 15–27. doi:10.1002/(SICI)1097-4571(199201)43:1<15::AID-ASI2>3.0.CO;2-K

Ruben, B. D. (1993). Integrating concepts for the Information Age: communication, information, mediation and institutions. In J. R. Schement & B. D. Ruben (Eds.), *Between communication and information*. Transaction Publishers.

Schement, J. R. (1993). Communication and information. In J. R. Schement & B. D. Ruben (Eds.), *Between communication and information*. Transaction Publishers.

Shannon, C. E., & Weaver, W. (1948). The Mathematical Theory of Communication. *Bell System Technical Journal*, 27, 379–423, 623–656.

Shannon, C. E., & Weaver, W. (1949). *The Mathematical Theory of Communication*. Chicago: University of Illinois Press.

Stamper, R. K. (1973). *Information in Business and Administrative Systems*. London: Batsford.

Stamper, R. K. (1985). Towards a theory of information – Information: Mystical fluid or a subject for scientific enquiry? *The Computer Journal*, 28(3), 195–199. doi:10.1093/comjnl/28.3.195

Sveiby, K. (1998). *What is Information?* Retrieved from <http://InformationR.net/ir/8-1/paper144.html>

Tully, C. J. (1985). Information, human activity and the nature of relevant theory. *The Computer Journal*, 28(3), 206–210. doi:10.1093/comjnl/28.3.206

Wallerstein, I. (2000). Globalization or the Age of Transition? – A Long Term view of the Trajectory of the World System. *Economia Global e Gestão*, 5.

Wersig, G. (1993). Information science: The study of postmodern knowledge usage [mar]. *Information Processing & Management*, 29(2), 229–239. doi:10.1016/0306-4573(93)90006-Y

Zikmund, W. G., (2000), *Business Research Methods* (6th ed.). Dryden Press.

Zorinho, C., Serrano, A., & Lacerda, P. (2003). *Gerir em Complexidade*. Edições Sílabo.

ADDITIONAL READING

Bjorn-Anderson, N. (1980). *The Human Side of Information Processing*. Amsterdam: North-Holland.

Brown, J. S., & Duguid, P. (2000). *The Social Life of Information*. Harvard Business School Press.

Dion, E. (1997). *Invitation à la Théorie de l'Information, Points*. Paris: Seuil.

Mckean, J. (1999). *Information Masters*. John Wiley & Sons.

Ross, W. T. Jr, & Dreyer, E. H. (1992). Making Inferences about Missing Information: The effects of existing Information. *The Journal of Consumer Research*, 19(Junho).

Sabaté, F. Tarragó, (1989), *Fundamentos de Economía de la Empresa*, 2ª edição, Librería Hispano Americana, Barcelona

Stiglitz, J. E. (1985). Information and Economic Analysis: A Perspective. *The Economic Journal*, 95, 21–41. doi:10.2307/2232867

Svenonius, E. (2000). *The Intellectual Foundation of Information Organization*. MIT Press.

Truijens, O. (2002), *Towards a Resource-based Theory of Information Strategy*, British Academy of Management Annual Conference 2002, September 9-11, Hammersmith, Vigand, R., A. Picot, et al., (1997), *Information, Organization and Management – expanding markets and corporate boundaries*, John Wiley and Sons

Vreeken, A. (2002), *Notions of Information: a review of literature*, PrimaVera Working Paper 2002-13, University van Amsterdam

Wisse, P. (2001), *Metapattern: Information Modeling as Enneadic Dynamics*, PrimaVera Working Paper 2001-04, University van Amsterdam

KEY TERMS AND DEFINITIONS

Data: The data does not carry any sense or meaning of the facts, images or sounds, since they lack relational elements essential to the establishment of a complete meaning, requiring a relational structure internal to a cognitive purpose.

Information: The information as “data with value for decision making”. The description of this concept refers to the effect that information, such as reduction of uncertainty, we consider, as an example of how to process information.

Knowledge: The information can be considered as a “substance” which can be acquired, stored and possessed by a person or group and transmitted from person to person or from group to group. The information has certain stability and perhaps is better faced as existing at the society level.

Objective Information and Subjective Information: To distinguish two main concepts of information: objective information and subjective information. In its systemic concept, the information will be considered: subjective information quadruple objective information, and both processes that relate.

Profitability of Information: “Information is a complex concept, ubiquitous since the emergence of life and inherent in any evolutionary process and thus served and continues to serve as a framework for reflection so much in biology, physics, psychology, in management, in Linguistics, as in many other branches of science”.

ENDNOTES

- ¹ The distinction between definition and perception, is that the definition is what characterizes the defined phenomenon, whereas the perception (concept) is the process of looking at the phenomenon. Accepting the idea of perception becomes easy to look into the usefulness of perception with more reason than the universally true definition (Belkin, 1978, p. 58).
- ² Descartes is known for his clear separation between mind (not material) and body (material). This is sometimes called the Cartesian form.
- ³ Sveiby (1998) points out that many scientists in experimental psychology and the science of information see information as meaning herself, information has a meaning independent of the user.
- ⁴ In general terms may mean that information is just information is something new. Another step is to say that the information should make a difference. “information is the difference that makes a difference” (Bateson in: Bausch and in 2000: Sveiby 1998).
- ⁵ Shannon wasn’t very interested with the meaning and the semantic aspects of communication. He should focus on the problem of engineering (see e.g. Sveiby 1998).

- ⁶ According to Bryant (2001) this contradiction disappears once resulted in the focus of Shannon (along with Weaver), in the performance potential of technology by passing the signal and Wiener focused him on human participation.
- ⁷ Treat information as a resource or convenience has given rise to many discussions in the economic literature (Babe, 1994, Belkin, 1978, p. 64/5, Braman, 1989 p. 238).
- ⁸ I've always been curious that a unit (Physics) of the two operands is always a formula.
- ⁹ In this view the information has an intrinsic meaning.
- ¹⁰ The relationship between knowledge and information is a question debated in the literature, especially in the literature of knowledge management (AlHawamdeh 2002, Wilson 2002). One problem is that different perceptions of knowledge exist.



Open Data Repositories in Knowledge Society

Nadim Akhtar Khan

University of Kashmir, India

Sara Sohrabzadeh

Tehran University of Medical Science, Iran

Garvita Jhamb

University of Delhi, India

INTRODUCTION

“Open means anyone can freely access, use, modify, and share for any purpose subject, at most, to requirements that preserve provenance and openness. (Open Definition, 2015).” Open data is data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and share alike (Open Data Handbook, n.d). The term “Open Data” thus refers to data and information beyond just governmental institutions and includes those from other relevant stakeholder groups such as business/industry, citizens, Non Profit Organizations (NPOs) and Non Government Organizations (NGOs), science or education including the World Bank, the United Nations, The Guardian and the Open Knowledge Foundation (Bauer & Kaltenbock, 2012). Open data is regarded as an essential component of modern day research as it facilitates carrying advance research activities in different subject domains at much broader levels. Open Data is an important resource for supporting decision making activities in Government agencies; research Institutions, Higher Educational Institutions etc. The research activities as well as policy making decisions are highly benefited by the availability of open data. Many countries have come forward to embrace the concept of Openness especially facilitating Open access to government data as such data is deemed to be public data by law. The open data

facilitates transparency and democratic control, participation, self-empowerment, improved or new products and services, innovation, improved efficiency and effectiveness of government services, impact measurement of policies, new knowledge from combined data sources and patterns in large data volumes. Open data as such is of much economic importance and is being reused world over in different setups. It improves government efficiency in addressing different issues concerning common people. It is paving ways for modern social fabrication and adding economic values to the existing data sets for efficient utilization resulting in new knowledge creation and more specializations. There should be no restrictions to its reuse like legal, financial or technological etc (Open Data Handbook, n.d).

The key requirements for Open Data include free availability of data and access and also format for redistribution and reuse of that data (SPARC, 2013). Good open data can be linked to so that it can be easily shared and talked about; is available in a standard, structured format, so that it can be easily processed; has guaranteed availability and consistency over time, so that others can rely on it; is traceable, through any processing, right back to where it originates, so others can work out whether to trust it (Open Data Institute, 2015). Harnad (2005) discusses that open access has two complementary routes: the gold road and the green road. Publishing in open-access journals

is the golden road. It calls on publishers to adopt open access policies whereas green road calls on researchers to self-archive their works in institutional repositories and publishes it on the author's institutional Web site. The more data is made openly available in a useful manner, the greater the level of transparency and reproducibility and hence the more efficient the scientific process becomes, to the benefit of society. This viewpoint is becoming mainstream among many funders, publishers, scientists, and other stakeholders in research (Molloy, 2011).

Heery and Anderson (2005) define digital repositories as digital collections in which the content is deposited and organized in a database, either by the content creator or some other means. Such a repository is based on an architecture that manages content as well as metadata and offers a minimum set of basic services while the purpose of a digital repository is to store and manage learning objects (McGreal, 2007). These repositories provide services to faculty, researchers, and administrators who want to collect research, historic, and creative materials (Drake, 2004). One characteristic common to all open repositories is the fact that they have been created by using software packages with open source licensing and developed by working groups linked to the open archives initiative community. This is based on the implementation of a shared protocol: Open Archives Initiative-Protocol for Metadata Harvesting (OAI-PMH). The essential feature of OAI-PMH is that it provides great ease of implementation because components of the OAI model basically involve a metadata harvester, a search and retrieval interface and a shared repository made up of a collection of individual repositories. These are combined with a set of requests and responses made through the hypertext transfer protocol (Bravo & Diez, 2007).

Data Repository is a logical partitioning of data where multiple databases which apply to specific applications or sets of applications reside (Online Learning, 2014). It is a somewhat general term used to refer to a physical or logical destination

or place designated for data storage for multiple databases, records or files. The term data repository can be used as either a local location that is directly accessible to the user without having to travel across a network or for a network location (Research Data Alliance, 2014). Generally, a data repository serves as the centerpiece of an open data effort. It serves as a central location to find data, a venue for standardizing practices, and a showpiece of the use of that data. Some repository software allows syndication, permitting other organizations to automatically incorporate their own data (How to Open Data, 2015). Open data repository as such is the warehouse of open data, where data warehouse refers to central repository for all or significant parts of the data collected. It includes data that can be freely used, reused and redistributed by anyone (National Informatics Centre, 2013).

Open data repositories as such provide ample opportunities for the scientific communities not only to showcase their research but making their research assets visible and distinct at global level. These repositories thus incorporate much richer features including:

Web Syndication

Web syndication refers to the creation of web feeds available from a particular website in order to provide people with update of recently added content. The two main formats of web syndication include RSS and Atom. This technology is incorporated in open data repositories to make their users updated by retrieving latest information from sites of interest and informing them about the changes in the content.

Support for Varied Data Types

These repositories also have been maintained to host different types of Open Data sets with the help of different software technologies. These include images, graphs, audio-visual data, network-based data, etc. These types of data have been used by



repositories to make their content more authentic, meaningful, attractive and easy to understand to increase the use of resources in these repositories.

Persistent Identifiers

Persistent identifier is an identifier which provides an appropriate representation of the resource and can be updated when resource change location so that resource will be available for longer period of time. Open data repositories make use of different persistent identifier systems like Digital Object Identifier (DOI), Uniform Resource Name (URN), Persistent Uniform Resource Locator (PURL), hdl, etc to make their available data persistent, globally accessible and citable over longer period of time and across platforms.

OAI-PMH Complaint

The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) provides an application-independent interoperability framework based on metadata harvesting. OAI-PMH complaint has great ease in implementation and flexibility to adapt to any context to provide information about any sort of resource, whether physical or digital (Bravo & Diez, 2007). OAI-PMH compliance enhances the visibility of the repositories and hence increases its usage that encourages use of it by these repositories.

BACKGROUND

The development of open data repositories started with the open access movement which took concrete shape and was described in declarations, the Budapest open access initiative, 2001 and The Bethesda statement on open access publishing, 2003 (Suber, 2003). Berlin Declaration 2003 brought together international experts with the aim of developing a new web-based research environment using the Open Access Paradigm as a mechanism for having scientific knowledge

and cultural heritage accessible worldwide (Max Planck Society, 2015). The exchange of scientific information is vital for the establishment of open access archives, which can prove as a faster, cheaper, and simpler route to reach the ultimate goal of universal open access in the future (Kirsop, Chan & Arunachalam, 2007). Trencheva and Todorova (2014) in their study on Open access to scientific information concluded that Librarians and information professionals are called to help their users to take advantage of the extraordinary potential of OA and participate in the creation of repositories and to encourage publication in the available repositories of scientific content. All these factors lead to development of open repositories as worthy information channels for scholarly communication. Many studies focused on open institutional repositories (Chang, 2003; Yeates, 2003) revealed that institutional repositories provide a way to display the research development of an institution through orderly collection, organization, preservation and providing access to its research output.

Repositories can be further categorized into Disciplinary and Institutional Repositories. A disciplinary repository is subject-oriented, without organizational or geographic limit. These types of repositories collect and provide access to the literature and scholarly outputs for a single discipline or a set of related disciplines. These repositories accumulate their data through authors' self-uploading of articles. Disciplinary repositories like PubMed or arXiv use standard set of content types and dissemination mechanisms and have few interaction tools (Adamick & Zellen, 2010). These aim to provide a rapid and convenient way for scientists to exchange their research results with colleagues (Luce, 2001).

On the other hand Johnson (2002) defines Institutional repositories as digital archives of the intellectual products created by the faculty, researchers, and students of an institution and accessible to end users both within and outside the institution. Lynch (2003) states that "An institutional repository is the collective intellectual

output of an institution recorded in a form that can be preserved and exploited. It collects, manages, disseminates and preserves scholarly works created in a digital form by faculty and students in universities and colleges. It can offer a set of services including digital content, submission, organization, access, distribution and preservation". Crow (2002) states that their content may comprise of preprints and other works-in-progress, peer-reviewed articles, monographs, data sets and other ancillary research material, conference papers, electronic theses and dissertations, and gray literature. Institutional repositories provide a centralized framework in which faculty members, researchers, scholars, and others can build their digital collections. These repositories provide digital collections with an infrastructure and stability that can sustain changes (Gibbons, 2004).

Trends in Open Data Management and Availability

Open public data are not only considered as an economic asset that can contribute to new products, services and efficiency gains, they are also seen as a key driver in the promotion of transparency (European Commission, 2011 as cited in Hellberg and Hedstrom (2015)). There are many versions of "open data" such as open data, open government data, linked open data and public sector information. "Government data" is explained as "any data and information produced or commissioned by public bodies". "Open Government Data" is government data that is openly available. It has been defined by eight principles, which outline that to be assessed as open; the data should be complete, primary, timely, accessible, machine-processable, non-discriminatory, non-proprietary and license-free (Ubaldi, 2013). According to Lluís, (2014) Open Data projects could reinforce the strategies of records preservation in databases and could mean an opportunity for historical archives if they take charge of the maintenance of the Open Data Archive. This could be relevant in the future for the transformation of data from information

systems into documentary heritage. Open Data products are easy to use and manage, and they do not present any problems regarding the data privacy restrictions. Sicilia, Garcia-Barriocanal and Sanchez-Alonso, (2016) highlight that the preservation and availability of research data is a major concern as it aspects core principles of scientific practice, including repeating and contrasting experiment and sharing findings. This concern has resulted in a number of initiatives and services that offer long-term preservation of research data in a broad sense, many of them exposing data and its description in open access form. Trends in the growth, development and use of Open repositories and availability of open harvesting tools have enabled increased visibility of such repositories. These repositories are becoming significantly important for scholarly communication. They give better visibility to the research papers that they hold because of free access and downloading facilities (Jayakanth, Minj, Silva & Jagirdar, 2008). These repositories proved as blessing for developing countries as they enhanced the research capacity of such countries by providing free interchange of essential scientific knowledge combined with capacity building. Open repositories have tremendous impact as they provide timely and open access to research and scholarship, increase the potential research impact of archived publications, supporting alternative forms of journal publishing and novel forms of digital scholarship by preserving and making accessible academic digital objects, datasets, and analytic tools that exist outside of the traditional scholarly publishing system (Chan, 2004). Open data repositories as such are created and maintained to provide universal and free access to such information products in electronic format as a means of facilitating research and scholarship (Reitz, 2002). Childs, McLeod, Lomas and Cook, (2014) examined the issues and opportunities of open research data and advocated that technological development and the obligation from research-funding bodies to maintain research data and make it open for sharing and reuse are offering new challenges to the research community



which generates pressure on them to react, rather than to lead a discussion about which data is suitable for sharing and reuse, and of the processes that are required to enhance proper research that is responsive to current and future scientific and societal developments.

Kim and Lee, (2014) while analyzing the Global data repository status based on Korea, China and Japan revealed that the repository operation events operated by the government is inadequate in Asia and there are no repository operation case in learning, open data, research data and web observatory, types in Korea, China and Japan. Reichman, Jones and Schildhauer (2011) in their study on “Challenges and Opportunities of Open Data in Ecology” revealed that standardization of methods and development of robust metadata can increase data access but reproducibility of analyses is also important. The establishment of well-curated, federated data repositories will provide a means to preserve data while promoting attribution and acknowledgement of its use. This is further supported by the view that the research papers available in open access repositories get more citations than subscription based papers (Norris, Oppenheim and Rowland, 2008). Xia and Nakanishi (2012) further examined the open access availability of journal articles and their citation conditions and revealed that articles in high-ranked journals do not have a higher open access rate, and articles in lower-ranked journals have a greater increase rate of citations if they are freely accessible.

Open publication of data as such is nowadays a clear upward trend, for which it is easier to assign human and financial resources. That is why these projects should be considered as an opportunity to include the perspective of digital preservation. With the open data initiative rolling out, researchers will gain experience with making their research data open and in reusing open data in their own research (Childs, McLeod, Lomas & Cook, 2014). One reason for emerging excitement about educational data mining is the increasing availability of student learning data. These data

are coming from many sources including data from schools, like standardized tests and student and teacher background variables and videos of classroom interactions. The data can also drive improved models of student cognition, affect, and learning that can be used to improve On-line assessment and on-line learning (Koedinger, Cunningham, Skogsholm & Leber, n.d). For each discipline or scientific domain there is a particular interpretation of datasets and research data, their nature, and collection procedures and the variations in the way that data are described with metadata and the problems associated with sharing (Gomez, Mendez, & Hernandez-Perez, 2016). Open Data Repositories as such incorporate varied aspects ranging from creation, processing, modeling, linking and availability of Open data sets in different setups ranging from research institutions, educational establishments to Government agencies for supporting education, research and decision-making activities. In modern times as such Open Data repositories result in ease of sharing data with greater potential for reuse resulting in citation impact and reputation building (Johnston, 2016).

Scenario of Open Data Repositories in Different Setups

Number of Open Data Repositories has been established in different domains to support innovation, creativity, teaching/Learning activities, and research activities in specific subject areas. The availability of Open Data especially Open data for research, Open Government data has paved new means to establish credibility and accountability of funding and government agencies. Many countries are embracing the concept and have come up with number of repositories hosting open data. Few such repositories have been discussed below:

European Union Open Data Portal¹

The European Union Open Data Portal is the single point of access to a growing range of data from

the institutions and other bodies of the European Union (EU). Data are free to use and reuse for commercial or non-commercial purposes. The portal aims to promote their innovative use and unleash their economic potential. It also aims to help foster the transparency and the accountability of the institutions and other bodies of the EU. The EU Open Data Portal is managed by the Publications Office of the European Union. Implementation of the EU's open data policy is the responsibility of the Directorate-General for Communications Networks, Content and Technology of the European Commission.

The OD500 Global Network²

The OD500 Global Network is an international network of organizations that seek to study the use and impact of open data. Coordinated by the Governance Lab (GovLab) the OD500 Global Network enables participating organizations to analyze open data in their country in a manner that is both globally comparative and domestically specific. The OD500 Global Network starts from the assumption that only by mapping the use of open data within and across countries, can new approaches for understanding the economic and social impact of open government data be generated.

Open Data Commons³

Open Data Commons exists to provide legal solutions for open data and is an Open Knowledge Foundation project run by its Advisory Council and is a not-for-profit effort working for the benefit of the general open knowledge community. In March 2008 it launched the first ever open data license: the Public Domain Dedication and License (PDDL).

ArcGIS Open Data⁴

Open Data allows organizations to use the ArcGIS platform to provide the public with open access

to their authoritative data. Organizations can create a website with their own look and feel and specify Open Data groups to share specific items. The general public can use Open Data sites to search by topic or location, download data in multiple formats, and view data on an interactive map and in a table. ArcGIS Open Data currently works with data from hosted feature layers, ArcGIS Server feature services, ArcGIS Server map services, ArcGIS Server image services, CSV files, hosted tables.

Open Government Data (OGD) Platform India⁵

Open Government Data (OGD) is a platform for supporting Open Data initiative of Government of India. The portal is intended to be used by Government of India Ministries/ Departments their organizations to publish datasets, documents, services, tools and applications collected by them for public use. It intends to increase transparency in the functioning of Government and also open avenues for many more innovative uses of Government Data to give different perspective. Open Government Data Platform India has 4 (four) major modules implemented on a single Drupal instance: Data Management System (DMS), Content Management System (CMS), Visitor Relationship Management (VRM), and Communities.

Open Data for Africa⁶

The African Development Bank Group (AfDB) committed to support statistical development in Africa has a large team of researchers who focus on the production of statistical data on economic and social situations. The data produced by the institution's statistics department constitutes the background information in the Bank's flagship development publications. Besides its own publication, the AfDB also finances studies in collaboration with its partners. Through its statistics department the AfDB is contributing to the effective development of RMC statistical capacity and



systems for the provision of timely and reliable data for policy formulation, implementation and evaluation and monitoring of progress towards achieving Millennium Development Goals (MDGs) and, Poverty Reduction Strategies (PRSs). It is also responsible for developing and managing databases on economic, social and development cooperation statistics in Africa in collaboration with RMCs and regional and international agencies (ECA, IMF, OECD, UN, World Bank, etc.).

CERN Open Data Portal⁷

The CERN Open Data portal is the access point to a growing range of data produced through the research performed at CERN. It disseminates the preserved output from various research activities, including accompanying software and documentation which is needed to understand and analyze the data being shared. The portal adheres to established global standards in data preservation and Open Science: the products are shared under open licenses; they are issued with a digital object identifier (DOI) to make them citable objects in the scientific discourse. The Open Data portal focuses on the release of data that comprises data directly related to publications, simplified data formats for analysis in outreach and training exercises, reconstructed data and simulations as well as the analysis level software to allow a full scientific analysis and also covers basic raw level data and their associated software and allows access to the full potential of the experimental data.

ALICE⁸

A Large Ion Collider Experiment (ALICE) is the experiment at the Large Hadron Collider devoted to the physics of matter at infinitely small scale exploiting heavy-ion collisions. Its scientific program aims at answering a series of fundamental questions. The available primary ALICE datasets contain a limited sample of specially selected interaction events recorded from p-p and Pb-Pb collisions collected in 2010. The data format is

the ESD (event summary data) format generated by the ALICE raw data reconstruction software. These datasets are not statistically representative and therefore cannot be used to perform full fledged scientific studies. They rather serve to demonstrate the use of the ALICE analysis tools. A set of simplified datasets derived from the primary ones for use in the provided master class analyses. These data contain exclusively the information needed by the master classes and are not useful for a different purpose.

Marine Geoscience Data System⁹

The Marine Geoscience Data System (MGDS) provides a suite of tools and services for free public access to marine geoscience research data acquired throughout the global oceans and adjoining continental margins. The system evolved from projects initiated from 2003-2004 with funding from the National Science Foundation to provide data management services for the U.S. Antarctic Program (USAP), the Ridge 2000 and MARGINS programs, and for active source seismic data. The Data Systems and Services include Marine-Geo Digital Library, Community Data Portals, Antarctic Data Resources, Global Bathymetry, Visualization and Analysis Tools.

National Centers for Environmental Information (NCEI)¹⁰

NCEI is responsible for hosting and providing access to one of the most significant archives on Earth, with comprehensive oceanic, atmospheric, and geophysical data. From the depths of the ocean to the surface of the sun and from million-year-old sediment records to near real-time satellite images, NCEI is a leading authority for environmental information. NCEI continues the tradition of excellence, unmatched expertise, and trusted, authoritative data that the previous three Data Centers (the National Climatic Data Center, the National Geophysical Data Center, and the National Oceanographic Data Center) established.

The top priority during the near future is to provide full spectrum of atmospheric, oceanographic, coastal, and geophysical products and services.

eCrystals¹¹

eCrystals - Southampton is the archive for Crystal Structures generated by the Southampton Chemical Crystallography Group and the EPSRC UK National Crystallography Service. The information contained within each entry of this archive is all the fundamental and derived data resulting from a single crystal X-ray structure determination, but excluding the raw images. The results have not been externally refereed, but the information supplied should enable any reader to check the reliability and validity directly, since all the files provided are freely available for download. This eCrystals repository is made available under the Open Data Commons Attribution License.

PubChem¹²

PubChem, released in 2004, provides information on the biological activities of small molecules. PubChem is organized as three linked databases within the NCBI's Entrez information retrieval system. These are PubChem Substance, PubChem Compound, and PubChem BioAssay. PubChem also provides a fast chemical structure similarity search tool. Links from PubChem's chemical structure records to other Entrez databases provide information on biological properties. These include links to PubMed scientific literature and NCBI's protein 3D structure resource. Links to PubChem's bioassay database present the results of biological screening. Links to depositor web sites provide further information. A PubChem FTP site, Download Facility, Power User Gateway (PUG), Standardization Service, Score Matrix Service, Structure Clustering, and Deposition Gateway are also available.

Geological Society of America Data Repository¹³



The Geological Society of America established in 1888, provides access to elements that are essential to the professional growth of earth scientists at all levels of expertise and from all sectors: academic, government, business, and industry. The Geological Society of America (GSA) is a global professional society with a growing membership of more than 26,000 individuals in 115 countries. The GSA Data Repository was established in 1974. It is an open file in which authors of articles in journals can place information that supplements and expands on their article. These supplements will not appear in print but may be obtained from GSA.

ISSUES AND CHALLENGES

The major challenges facing Open Data includes issues related to availability of data created with public money as Open Data with high priority, creating an environment for openness by starting embedding Open Data in everyday's tasks resulting in shifting from policy-making process to a more evidence-based, data-powered one. Besides improving public confidence is very important by showcasing the benefits that data can bring to society in every sphere. The infrastructure improvements concerning Open Data needs to be addressed by developing Information Infrastructure at different levels (Sollazzo, 2015). The area of Open data including availability of Open data sets in different establishments through open repositories is so vast and provides conducive environment for researchers to chalk out strategies for addressing different challenges concerning open data in future.

FUTURE RESEARCH DIRECTIONS

Future of Open data will be highly influenced by the development and adaptation of open data standards, integrated portals, open data feeds, live feeds, data consistency, publishing tools, Government policies, which have power to enforce open data initiatives and goals, data literacy and demystification of open data, increased awareness, transparency, and Open source tools(Cordivano,2015).

Further Research needs to be carried on analyzing the changing dimensions in the procedures associated with the availability of open data in future. These research activities must focus on innovation, creativity and long term persuasion. Steps need to be taken in order to convince different stake holders including Government agencies, funding bodies, research institutions etc to make their data sets open for people, so that it will help establishing a more transparent and efficient system with credibility and efficiency. Open Data projects needs to be prioritized in different setups so as to explore and identify different issues concerning Open data repositories in terms of designing and management of different datasets and formats. Such activities will facilitate a more convenient and organised open research environment for addressing different issues concerning common man in future.

CONCLUSION

The benefits of Open data can be entirely realized only after incorporating technical skills among different stake holders especially in order to ascertain the quality of data made available and the associated procedures for organizing and making multiple data sets retrievable through different channels. According to IFLA trend report by 2020 governments will have realized that it takes more than rhetoric to deliver effective and transparent open government services. Currently Open Data and services especially Open Government data is mainly supply driven, when in reality it needs to

be driven by citizen demand (IFLA, 2015). There is a need for greater collaboration and cooperation among all stakeholders to solidify the gains and keep the momentum.

REFERENCES

- Adamick, J., & Reznik-Zellen, R. (2011). Representation and recognition of subject repositories. *D-lib Magazine*, 17, 9-10. doi:10.1045/september2010-adamick
- Bauer, F., & Kaltenbock, M. (2012). *Linked Open Data: The Essentials. A Quick Start Guide for Decision Makers*. Retrieved from <http://www.reeep.org/LOD-the-Essentials.pdf>
- Bravo, B. R. & Diez, L. A. (2007). E-science and open access repositories in Spain. *OCLC Systems & Services: International Digital Library Perspectives*, 23(4), 363 – 371. 10.1108/10650750710831501
- Budapest Open Access Initiative. (2001). *Budapest Open Access Initiative*. Retrieved from www.Soros.org/openaccess
- Chan, L. (2004). Supporting and enhancing scholarship in the digital age: the role of open-access institutional repositories. *Canadian Journal of Communication*. Retrieved from http://eprints.rclis.org/5648/1/Chan_CJC_IR.pdf
- Chang, S.H. (2003). Institutional repositories: the library's new role. *OCLC Systems & Services*, 19(3), 77-79. 10.1108/10650750310733199
- Childs, S., McLeod, J., Lomas, E., & Cook, G. (2014). Opening research data: Issues and opportunities. *Records Management Journal*, 24(2), 142-162. 10.1108/RMJ-01-2014-0005
- Cordivano, S. (2015). *Building the Future of Open Data: Part 1*. Retrieved from <http://www.azavea.com/blogs/atlas/2015/04/building-the-future-of-open-data-part-1/>

Crow, R. (2002). *The case for institutional repositories: A SPARC position paper*. Washington, DC: SPARC. Retrieved from http://www.sparc.arl.org/sites/default/files/media_files/instrepo.pdf

Drake, M. (2004). Institutional Repositories Hidden Treasures. *Searcher (Medford, N.J.)*, 12, 41–45. Retrieved from <http://www.infotoday.com/searcher/may04/drake.shtml>

Gomez, N-D., Mendez, E., & Hernandez-Perez, T. (2016). Social sciences and humanities research data and metadata: A perspective from thematic data repositories. *El Profesional de la Información*, 25(4), 545-555. 10.3145/epi.2016.jul.04

Harnad, S. (2005). *Fast-Forward on the Green Road to Open Access: The Case Against Mixing Up Green and Gold*. Retrieved from <http://eprints.ecs.soton.ac.uk/10675/1/index.html>

Heery, R., & Anderson, S. (2005). *Digital repositories review*. Joint Information Systems Committee. Retrieved from <http://www.ahds.ac.uk/preservation/digital-repositories-review-2005.pdf>

Hellberg, A.-S., & Hedstrom, K. (2015). The story of the sixth myth of open data and open government. *Transforming Government: People, Process and Policy*, 9(1), 35-51. 10.1108/TG-04-2014-0013

How to Open Data. (2015). *Data Repositories*. Retrieved from <http://how-to.readthedocs.org/en/latest/The-Basics-of-Open-Data/Data-Repositories/>

IFLA. (2015). *The future impact of open government and open data initiatives*. Retrieved from <http://trends.ifla.org/expert-meeting-summary/the-future-impact-of-open-government-and-open-data-initiatives>

Jayakanth, F., Minj, F., Silva, U., & Jagirdar, S. (2008). ePrints@IISc: Indias first and fastest growing institutional repository. *OCLC Systems & Services*, 24(1), 59–70. doi:10.1108/10650750810847260

Johnson, R. K. (2002). Institutional Repositories: Partnering with Faculty to Enhance Scholarly Communication. *D-Lib Magazine*, 8, 11. Retrieved from <http://www.dlib.org/dlib/november02/johnson/11johnson.html>

Johnston, L. (2016). *Data Repositories: The Answer that Actually Came with a Question*. University of Massachusetts and New England Area Librarian e-Science Symposium. Retrieved from http://escholarship.umassmed.edu/escience_symposium/2016/program/5

Kim, S., & Lee, W. (2014). Global data repository status and analysis: based on Korea, China and Japan. *Library Hi Tech*, 32(4), 706-722. 10.1108/LHT-06-2014-0064

Kirsop, B., Arunachalam, S., & Chan, L. (2007). Access to scientific knowledge for sustainable development: Options for developing countries. *Ariadne*, 52. Retrieved from www.ariadne.ac.uk/issue52/kirsop

Koedinger, K. R., Cunningham, K., Skogsholm, A., & Leber, B. (n.d.). *An open repository and analysis tools for fine-grained, longitudinal learner data*. Retrieved from http://pslclatashop.org/about/pubs/16_Koedinger_45.pdf

Lluís Esteve Casellas Serra. (2014). The mapping, selecting and opening of data. *Records Management Journal*, 24(2), 87 – 98. doi:10.1108/RMJ-01-2014-0008

Luce, R. E. (2001). E-prints Intersect the Digital Library: Inside the Los Alamos arXiv. *Issues in Science and Technology Librarianship*. Retrieved from <https://indico.cern.ch/event/410468/session/12/attachments/831393/1149886/12Luce.pdf>

Lynch, C. A. (2003). Institutional repositories: essential infrastructure for scholarship in the digital age. *ARL: A Bimonthly Report*, (226). Retrieved from www.arl.org/resources/pubs/br/br226



- Max Planck Society. (2015). *Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities*. Retrieved from <http://openaccess.mpg.de/Berlin-Declaration>
- McGreal, R. (2007). *A typology of learning object repositories*. Retrieved from <http://auspace.athabascau.ca/handle/2149/1078>
- Molloy, J. C. (2011). The Open Knowledge Foundation: Open Data Means Better Science. *PLoS Biology*, 9(12), e1001195. <http://doi.org/10.1371/journal.pbio.1001195>
- National Informatics Centre. (2013). *Implementation Guidelines for National Data Sharing and Accessibility Policy (NDSAP)*. Retrieved from https://data.gov.in/sites/default/files/NDSAP_Implementation_Guidelines-2.1.pdf
- Norris, M., Oppenheim, C., & Rowland, F. (2008). *The citation advantage of open-access articles*. Wiley. ASIS&T. doi:10.1002/asi.20898
- Online Learning. (2014). *What is Data Repository*. Retrieved from <http://www.learn.geekinterview.com/data-warehouse/dw-basics/what-is-data-repository.html>
- Open Data Handbook. (n.d.). *Open data. Open knowledge*. Retrieved from <http://opendatahandbook.org/guide/en/what-is-open-data/>
- Open Data Institute. (2015). *What makes data open?* Retrieved from <https://theodi.org/guides/what-open-data>
- Reichman, O. J., Jones, M. B., & Schildhauer, M. P. (2011). Challenges and opportunities of open data in ecology. *Science*, 331(6018), 703-5. DOI: 10.1126/science.1197962
- Reitz, J. M. (2002). *ODLIS – Online Dictionary for Library and Information Science*. Retrieved from <http://vlado.fmf.uni-lj.si/pub/networks/data/dic/odlis/odlis.pdf>
- Research Data Alliance. (2014). *Data Repository*. Retrieved from http://smw-rda.esc.rzg.mpg.de/index.php/Data_Repository
- Sicilia, M. A., Garcia-Barriocanal, E., & Sanchez-Alonso, S. (2016). *Community curation in open dataset repositories: Insights from Zenodo*. Academic Press.
- Sollazzo, G. (2015). *Four challenges for the future of Open Data*. Retrieved from <https://data.gov.uk/blog/four-challenges-future-open-data>
- SPARC. (2013). *Open Data*. Retrieved from <http://www.sparc.arl.org/issues/open-data>
- Suber, P. (2003). *Bethesda Statement on Open Access Publishing*. Retrieved from <http://nrs.harvard.edu/urn-3:HUL.InstRepos:4725199>
- The Open Definition. (n.d.). *The Open Definition*. Retrieved from <http://opendefinition.org/>
- Trencheva, T. S. & Todorova, T. Y. (2014). Open access to scientific information: comparative study in DOAJ. *Library Management*, 35. 10.1108/LM-08-2013-0070
- Ubaldi, B. (2013). *Open government data: towards empirical analysis of open government data initiatives*. OECD Working Papers on Public Governance, No. 22. OECD Publishing. 10.1787/5k46bj4f03s7-en
- Xia, J., & Nakanishi, K. (2012). Self-selection and the citation advantage of open access articles. *Online Information Review*, 36(1), 40-51. 10.1108/14684521211206953
- Yeates, R. (2003). Institutional repositories. *Vine*, 33(2), 96-101. doi:10.1108/03055720310509064

ADDITIONAL READING

- Gronbaek, K., & Trigg, R. H. (1999). *From Web to workplace: Designing open hypermedia systems*. Cambridge, Mass: MIT Press.
- Halsall, F., & Halsall, F. (1992). *Data communications, computer networks, and open systems*. Wokingham, England: Addison-Wesley Pub. Co.

Janert, P. K. (2011). *Data analysis with open source tools*. Sebastopol, CA: O'Reilly.

Miller, D. R., & Clarke, K. S. (2004). *Putting XML to work in the library: Tools for improving access and management*. Chicago: American Library Association.

Moore, S. A. (2014). *Issues in Open Research Data*. s.l. Ubiquity Press. doi:10.5334/ban

Pascoe, J. A. (2014). *A PhD student's experience with open research data*. TU Delft Library.

R.S. Means Company. (1985). *Open shop building construction cost data*. Kingston, MA: R.S Means Co.

Thomas, A., & Rothery, A. (2005). Online repositories for learning materials: the user Perspective. *Ariadne* (45). Retrieved from www.ariadne.ac.uk/issue45/thomas-rothery

Ubaldi, B. (2013). *Open government data: Towards empirical analysis of open government data initiatives*. Paris: OECD. doi:10.1787/19934351

Willinsky, J. (2006). *The access principle: The case for open access to research and scholarship*. Cambridge, Mass: MIT Press.

KEY TERMS AND DEFINITIONS

Digital Repositories: A digital repository is where digital content and assets are stored and can be searched and retrieved for later use.

Institutional Repository: An institutional repository is an online archive for collecting, preserving, and disseminating digital copies of the intellectual output of an institution, particularly a research institution.

Open Access: An unrestricted online access to peer reviewed scholarly research.

Open Data: Open data is data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and share alike.

Open Government Data: Data produced or commissioned by government or government controlled entities. Data which is open as defined in the Open Definition – that is, it can be freely used, reused and redistributed by anyone.

ENDNOTES

- ¹ <https://open-data.europa.eu/en/about>
- ² <http://www.opendata500.com/>
- ³ <http://opendatacommons.org/about>
- ⁴ <http://doc.arcgis.com/en/open-data/>
- ⁵ <https://data.gov.in/about-us>
- ⁶ <http://opendataforafrica.org/about/us>
- ⁷ <http://opendata.cern.ch/about>
- ⁸ <http://opendata.cern.ch/about/ALICE>
- ⁹ <http://www.marine-geo.org/about/overview.php>
- ¹⁰ <https://www.ncei.noaa.gov/>
- ¹¹ <http://ecrystals.chem.soton.ac.uk/>
- ¹² <http://pubchem.ncbi.nlm.nih.gov/about.html>
- ¹³ <http://www.geosociety.org>

Quantum Information Science Vis-à-Vis Information Schools

P. K. Paul

Raiganj University, India

D. Chatterjee

Seacom Skills University, India

A. Bhuijali

Raiganj University, India

INTRODUCTION

Quantum Information Science (QIS) is a combination of Quantum Science (which combines Radio Physics, Condensed Physics, and Electronics) and Information Science (which combines Computer Science, Information Technology, Mathematics, Information Studies, and Documentation Studies and so on. Quantum Information Science (QIS) is actually an extension of Quantum Computing. Quantum Information Science (QIS) is mistakenly taken as Quantum Information Theory but it has several differences with this. Quantum Information Science (QIS) is mainly responsible for improved and faster acquisition, transmission and processing of information. The 20th century is marked by three monumental achievements, namely, computer science, quantum physics and information theory, which have not only stunned the civilized world but also ushered into a new world – a new paradigm of science and technology. The information technology revolution of this century was directed towards the miniaturization of electrical circuitry on silicon chips on one side, enabling performance to double every 18 months (Moore's Law) and the deepening understanding of classical information in coding, cryptography and computational complexity on the other side. Thus a new era dawned with the promise of a new design strong enough to handle the ensuing new problems. Considerable planning has already

been devoted to the challenges of designing and fabricating devices at the atomic scale using nanotechnology and quantum theory. Indeed quantum effects are being exploited to perform important, otherwise impossible information processing tasks, in addition to the creation of unbreakable codes and possibly a quantum computer to perform easily computations that would have taken millions of hours for its performance. Quantum Information Science (QIS) is actually an extension of Quantum Computing. Quantum Information Science (QIS) is mistakenly taken as Quantum Information Theory but it has several differences with this. Quantum Information Science (QIS) is mainly responsible for improved and faster acquisition, transmission and processing of information. Quantum Information Science (QIS) is actually nothing but an application and integration of Quantum Science and Information Science principles (See Figure 1). Thus Quantum Science is mainly rooted by the Quantum Physics principles and then Quantum Computing and lastly Quantum Information Science. QIS is an existing field which need an intellectuality and mentality towards achievements. Quantum Information System is the ultimate result of healthy Quantum Information Science (QIS) practice. This paper talks about Quantum Information Science (QIS); its background and origin. Possibilities of Quantum Information Science (QIS) as an academic domain and possible courses emphasize-

DOI: 10.4018/978-1-5225-2255-3.ch386

ing challenges and opportunities to introduce it in I-School or Information Science departments with specialization are explored.

BACKGROUND AND EXISTING LITERATURE

Practically, Quantum Information Science [QIS] is combination of some more subfield. The main field from Quantum Side is Quantum Physics, Quantum Mechanics, Quantum Computing and other hand the domain of Information side; which is combination of Management Science, Computer Science, Information Technology, Cognitive Science, and Operation Research and so on (Bouwmeester et al., 1997; Ahn, 2000; Paul, 2013a).

In other words, Quantum Information Science (QIS) is a combination of Quantum Science (which combines Radio Physics, Condensed Physics, Electronics) and Information Science (which combines Computer Science, Information Technology, Mathematics, Information Studies, Documentation Studies and so on, which has possibilities of healthy Information Infrastructure and speedy information access between information channels, information networks, information centre to information networks and so on (Devoret, 2013; Paul, 2013b). The wider benefit and possibilities of Quantum Information Science (QIS) attract thousands of scientists and technological experts in the field of physics and mathematics and its integration into information science. The Information Technology revolution of the past several decades has been driven by steady advances in the miniaturization of electronic circuits on silicon chips, allowing performance to double roughly every 18 months (Moore's law). Quantum Computing and Information Science mainly differs in the perspective of different uses and scope. Here physical form of information has a qualitative rather than merely quantitative bearing on how efficiently information can be processed (DiVincenzo, 2001; Paul, 2015). The economic impact of this quantum information

science (QIS) is going to be no less profound. The complexity of the economic world with its colossal information will be resolved to generate a wholly new economic paradigm. So it is important that the necessary foundation of QIS be built to provide tools to solve those problems and enable progress move forward towards more specific technical and economic growth (See Table:1 for more clarification).

OBJECTIVE AND HYPOTHESIS

The main aim and objective of the study include:

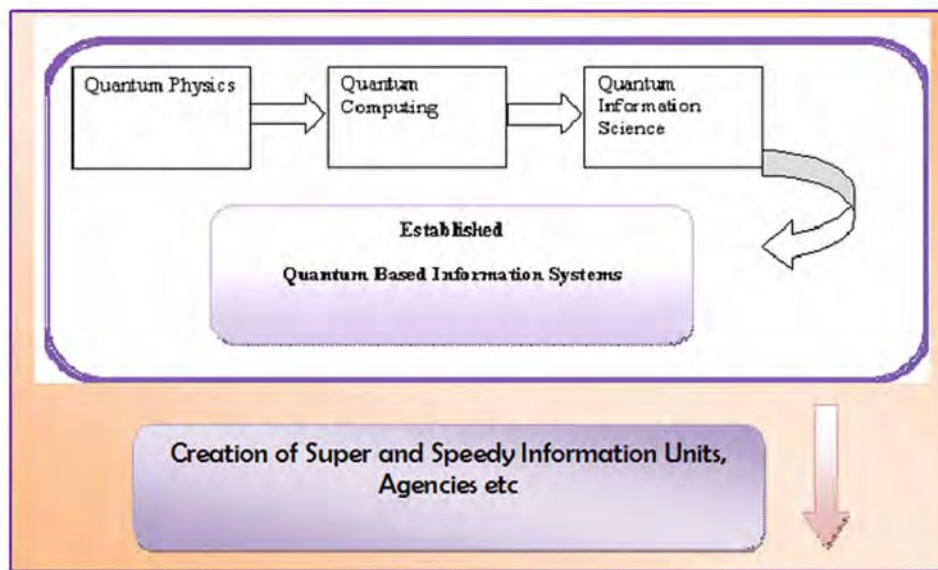
- To know about the Quantum Information Science (QIS) and its nature.
- To learn about Quantum Information Science (QIS) and its basic nature.
- To find out main advantages and engineering applications of Quantum Information Sciences.
- To find out possibilities of I-School based Quantum Information Science (QIS) programmes.
- To know about attributes and characteristics of modern Quantum Information Science (QIS).

QUANTUM INFORMATION SCIENCE [QIS] AS A DOMAIN AND GROWING INTERDISCIPLINARY NATURE

The interdisciplinary Quantum Information Science [QIS] is providing a new way into fundamentals problems related to Computing as well as Physics (NSF, 1999). Though it is important to note that, Quantum Information Science based on basic Mathematical principles and which are not new but integration is new that produce several new information mechanism powered by Quantum BITS (Leuenberger & Loss, 2001).

Quantum Information Science (QIS) is a kind of science supported by several technologies

Figure 1. Road map of establishment of quantum information sciences to learn about I-School and its emerging nature



mainly responsible for information collection, selection, organization, processing, management and dissemination with super speed and dynamic movement from one place to another within a moment (Dutt et al., 2007). Quantum Information Science (QIS) is mainly backed by optics and quantum physics; though it is now treated as a subject with far reaching implication for basic mathematical as well as physical science. Data and information are rapidly growing. Today it is very difficult to manage information thus emerging world wide both in professional practice and in academic researches. Fundamentally it is dedicated to both computer and physical science. In general Quantum Information Science (QIS) may

seem to be a combination of Quantum Science and Information Science; but practically these two include some more subjects (NSF, 1999). So ultimately Quantum Information Science (QIS) is a knowledge cluster of Computer Science, Physics, Information Studies, Management Studies, Mathematical Science, Nano-Technology. NSF, USA mentions that Quantum Information Science (QIS) is providing profound new insight into fundamental problems relating to both computation and physical science. Practically the backbone of QIS is Quantum Theory which originated nearly 80 years back and Information Science discovered nearly 70 years back (Furusawa et.al., 1998).

Table 1. Some disparities between general and Quantum Information Science

General Information Science	Quantum Information Science [QIS]
• Classical Information can be read and copied without disturbed. • In General and classical Information, it is tough or impossible to exploited perform task. • Today's general Computer will take billion of years to find the prime factor. • It is small-scale Computing depended. • To find out a particular item in general computer will take N to search.	• Information carried by a Quantum System flouts such Common sense principles. • Quantum Information Science [QIS] can be exploited to perform task. • Whereas Quantum Information Science [QIS] based computer will take just second for that. • It is large-scale Computing depended. • Where as in Quantum Information Science [QIS], Computer will take square root of N.

QIS AND TECHNOLOGICAL FEATURES

The bright future of QIS warns us about several problems in its development. We have already seen that two obvious properties of classical information are

- It can be read and copied without being disturbed.
- The state of a composite system can be fully specified by specifying the state of each of its parts.

But it is noteworthy that a quantum information system flouts such common sense principles. This implies that a quantum information system can be used to perform tasks where a classical system will fail to do (Knill et.al. 1998). Recent researches have revealed that noisy quantum devices can reliably store and process suitably encoded quantum states. Normally quantum states that arise in the intermediate stages of a quantum computation are extraordinarily fragile; but if a logical qubit is encoded not as a single physical qubit but in the form of entanglement, it becomes far more robust (NSF, 1999).

There is no doubt that the new quantum error correcting codes and fault tolerant methods will be essential for any future effort to create, maintain and manipulate intricate many-qubit quantum states (Home et.al., 2009). Hopefully the quantum information processing of moderate complexity will be soon possible in a variety of implementations, showing the prospect of a full-scale quantum computer. It is worth mentioning here that the advancement in the technology of quantum cryptography is quite satisfactory and closer to commercial realization. It can be emphatically remarked that researches in QIS will have an impact not only in nanoscale engineering and precision metrology but also in basic sciences and defence research (Imamog et. al, 1999). These imply that the future advances in researches will require the efforts of expertise

from the disciplines of mathematics, computer science, physics, chemistry, material science and engineering and information theory. Thus as one of the greatest scientific ideas of the twentieth century, QIS will set a new order of science in the twenty first century.

QIS: EMERGING APPLIED INFORMATION SCIENCE

The novelty of the quantum effects lies not merely in the quantitative but in the qualitative improvement also, otherwise impossible. This intellectually exciting field has incredible technological implications on mathematics and physics also and has started providing a wholly new language for describing the mysteries of nature and a new way of thinking about a wide variety of scientific and technological questions (Virginia, US, 2006; Maurer et.al., 2012). The new scientific tool, i.e. quantum computer is expected to play the most vital role in this transformation (See Figure 2).

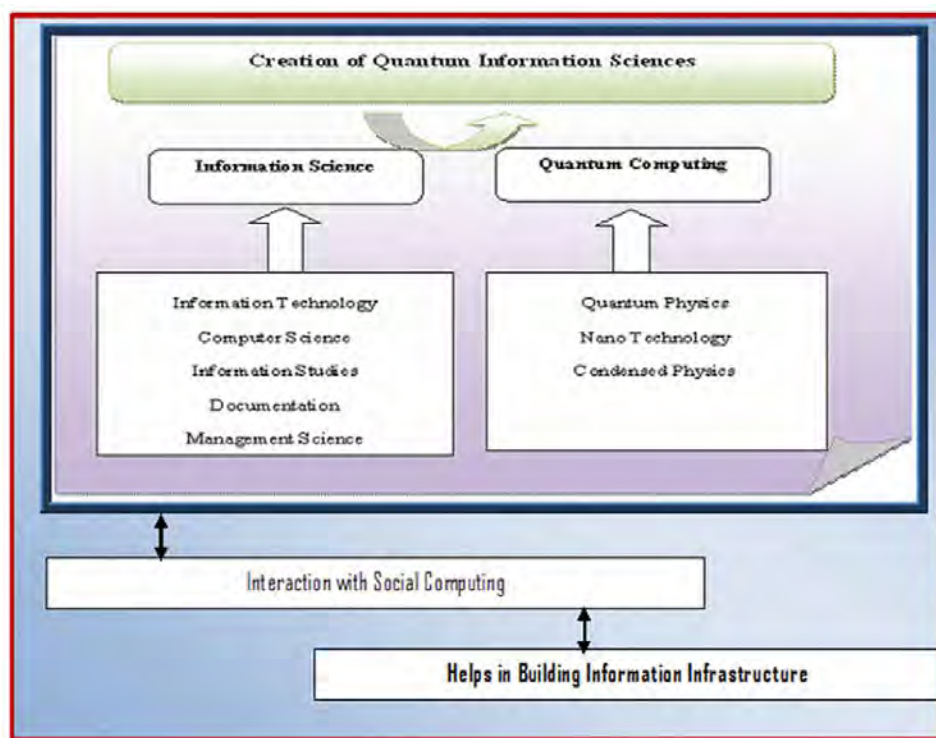
INFORMATION SCHOOL VIS-A-VIS QIS AND QUANTUM COMPUTING

Information Schools are responsible for providing manpower in the field of Information and Technologies and Computing with the focus of Information. Information Schools are abbreviated as I-School and many schools and departments are make a consortium and designated as Information School (Paul et.al., 2015). There are many opportunities come with such schooling in terms of courses, research, job potentialities and interdisciplinary skills (Karski et. al., 2009).

Though it deals with some problems and challenges; this paper is talks about such preliminary aspects, emerging trends and possibilities in Indian context with international agenda. In other words, Information School or I-School is a kind of schooling concept; where several information programmes are tied up. Information Schools are



Figure 2. The interdisciplinary knowledge cluster of quantum information science and its side by side promotional affairs



next generation schools which believe not only in healthy information channel and information infrastructure building but also interaction between Information-Technology people. In many perspectives it is motivated towards information and computing application to the various parts of the society and industries. Several courses are listed in I-School; but the flagship programme is Information Science; though a new nomenclature 'Information Science and Technology' (IST) started recently for better Information and Technological solution.

Some new Information School programmes specialized in nature like Geographical Information Science, Chemical Information Science, Bio-Informatics, Music Informatics have come up by now. Thus Information Science has a long tradition of introducing interdisciplinary programmes which match with other domains like Computer Science, Chemical Science, Geography, Biology. Quantum Information Science

(QIS) may also be initiated as a full-fledged QIS Degree programme like BSc/MSc-Information Science (QIS). Quantum Information Science (QIS) initiated in I-School need joint efforts of Physical Science professionals who have expertise in Quantum Computing and Information Science professionals who believe in Technocratic Information Solutions.

QIS: EMPOWERING RAPID AND QUICK INFORMATION SYSTEM

Quantum Information Science (QIS) is mainly responsible for faster and transparent information transfer within a moment from any where and by dedicated structure powered by Quantum Computing, Quantum Physics, and Information Theory (Shor 1994). The understanding of classical type of information and data, coding, cryptography and computational complexity acquired in the preced-

ing decades has laid foundations for extension to the quantum realm (Vlastakis et.al., 2013). The development of healthy and dynamic laboratory techniques has provided the basic infrastructure for manipulation and monitoring the characteristics of single automatic, electronic and nuclear systems (Zoller et.al., 2005). These two are the main factors behind Quantum Information Science (QIS).

Quantum Information Science (QIS) is a wonderful practicing field and also a tool for building healthy and sophisticated information systems and network building (Kiraz et.al., 2004). It allows faster communication; for examples Quantum Computers can perform the search in a time proportional to the square root of N . The communication between process and within CPU and large complex computing systems may interact and transfer in a much more speedy way. Fundamentally main pillar of Digital Information Services are general computers which process classical information encoded in bits, where as Quantum Information System based computers process information encoded in Quantum bits or QBITs (Zhang et.al., 2007). Thus, Quantum Information Science affects both internal and external information systems.

QUANTUM INFORMATION SCIENCE (QIS): EMERGING FEATURES

Quantum Information Science (QIS) may be attributed as under (and also depicted in Figure 3):

- It may transfer, deliver and collect large amount of data within a moment which is tough in general Information Science Practices.
- Finding the prime factor of a number may take long time where as Quantum Information Systems based on computer may perform that task within a moment.
- Information explosion may be done in a sophisticated way with the help of Quantum Information Science (QIS).

- After Information exchange, cryptographic protocol may be devised in which privacy is ensured by the principles of fundamentals physics.
- The large scale quantum computer may boost super Quantum Technology.
- According to NSF, the main achievement is generalisation or extension of classical theory that apply when information is reprinted as a quantum static rather than in terms of classical bits.
- Quantum states are actually highly unstable and cannot be observed without being distributed.
- Virtually, QIS ensures that quantum systems of modest size are endowed with truly vast computational systems.
- The main goal of Quantum Informatics is that it is the characteristics of quantum entanglement.
- Practically, the compressibility of quantum message can be quantified.

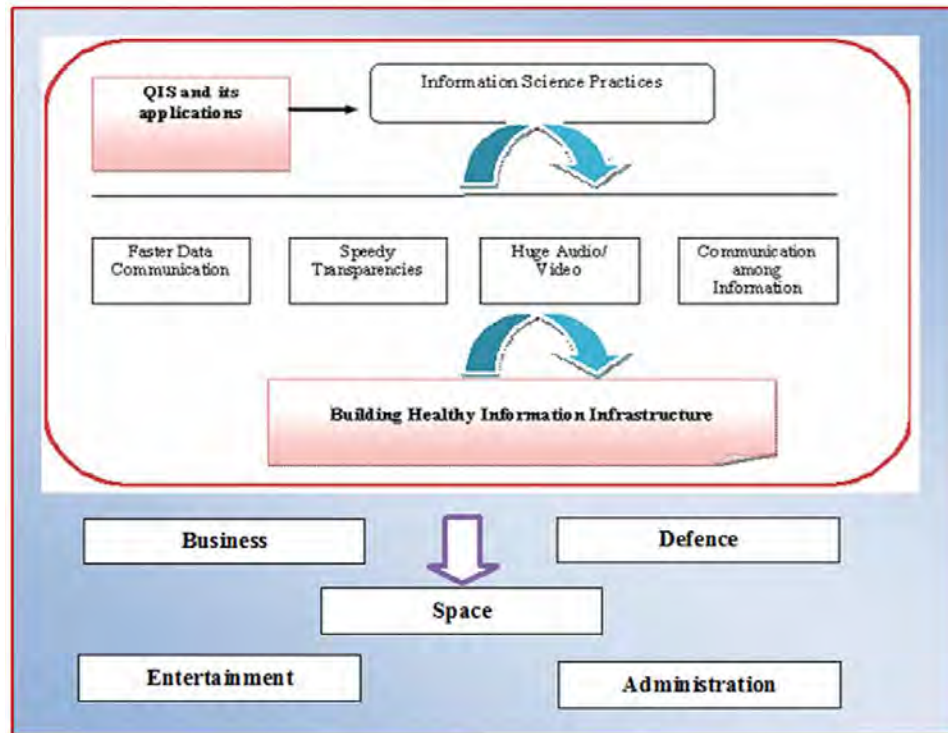
SUBFIELDS OF QUANTUM INFORMATION SCIENCE

Quantum Information Science is mainly responsible for the speedy and transparent information transparency. Building Quantum powered Information Science practice is mainly possible with Quantum Information Science, when as it is depends on Quantum Computing. Thus, the Quantum Information Science is the longer periphery than Quantum Information Systems and lastly Quantum Computing.

Virtually, Quantum Information Science is combine with so many tools, techniques and theories which includes- Quantum Statistical Mechanism, Quantum Gravity, Quantum Computational Complexity, Quantum Communication, Quantum Cryptography, Quantum Information Processing, Conditional Quantum entropy and so on (Paul et al., 2015). However, some Quantum expert and Information expert claim and express that, apart



Figure 3. Depicted the wonderful role of quantum information sciences



from the area and field mention above some are also important fact or subject of expert Quantum Information Science, these are as follows:

- Quantum Error Connection,
- Quantum Communication Complexity,
- Quantum entanglement,
- Quantum Dense Coding,
- Shor Factoring Algorithm,
- Grover Search Algorithm,
- Classical Information Theory,
- Quantum Systems,
- Quantum Mechanism,
- Quantum Teleportation and so on.

FINDINGS

- Quantum Information Science (QIS) is a popular name in Quantum world; which includes quantum computing, quantum physics and quantum science.

- Information Science and Quantum Science makes it a professional practice and also an academic domain; earlier it was mainly treated as a practice field and as Quantum Information Systems.
- It delivers QBITS in its process in all segments rather than classical BITS.
- Practically the quantum channel and its capacity can be enhanced if the communicating parts share a super dense coding (See Table 2).

FUTURE RESEARCH POTENTIALITY AND DIRECTION

Quantum based Computer will be much more faster and speedy communication between computer networks and Information Networks. Here one Information Centre may communicate with another and one Information Centre with Information Networks and Information Systems. Quantum

Table 2. Some possible courses related to information science vis-à-vis quantum sciences in the I-schools

QIS in Information Science Context	QIS in Computer Science Context	QIS in Engineering Context
• BSc/MSc- Quantum Information Science • BS/MS- Quantum Information Science • BS/BSc/MS/MSc-Information Science (QIS) • BSc/MSc./BS/MS-Quantum Informatics	• BSc/MSc- Quantum Information Science (Computing Context) • BS/MS- Quantum Information Science (Computing Context) • BS/BSc/MS/MSc-Computer Science (QIS) • BSc/MSc./BS/MS-Computing (Quantum Informatics)	• BTech/MTech-Quantum Information Science • BE/ME-QIS • BTech/MTech/BE/ME-Computer Science, (Quantum Information Science)

Information Science (QIS) is still practiced in small processors; but it needs to be initiated in the information systems and networks. Noisy quantum channel should be removed by healthy attributes; Use of channel twice in the same direction may also send information; Research in Quantum Information Science (QIS) is needed in IT, CS, Information Science, Information Systems schools for further development of QIS. It is essential to check that Quantum key distribution provide secrecy of generating such key information. Academic Institutions need to do several things for advancement of Quantum Information Service for internet and future data Management. More over the following aspects need to take care—

- It is essential to start Quantum based Networks and Quantum based internet.
- Quantum Information Science [QIS] need to start as full-fledged programme in Indian University; however programme may be start as pass course in BSc Programmes.
- Proper funding is essential to builds Quantum based Information Infrastructure.
- Better cooperation in Computer Science, Information Science, Management Science and Physical Science may bring more interdisciplinary research.

CONCLUSION

Quantum Information Science (QIS) is fully responsible for faster as well as transparent in-

formation transfer mechanism within a moment from anywhere and by dedicated structure powered by Quantum Computing, Quantum Physics, and Information Theory. For healthy QIS practice several things need to be taken care of, out of which Quantum Searching, Quantum Communication Complexity, Quantum Error correcting codes are important ones. The aspects of Quantum are long time practiced in Atomic, Molecular and Optical Physics (AMO). Information Science and similar academic programmes need to be initiated as a full fledged academic programmes on this emerging interdisciplinary domain for building sophisticated speciality, small and large Information channels and of course Big Data or Information Infrastructure Management. Research and innovation in Quantum Information Science (QIS) is required in Informatics, Computing, Information Science, Information Systems schools and other academic units for further development of QIS as a scientific and interdisciplinary domain.

REFERENCES

- Ahn, J., Weinacht, T. C., & Bucksbaum, P. H. (2000). Information storage and retrieval through quantum phase. *Science*, 287(5452), 463–465. doi:10.1126/science.287.5452.463 PMID:10642542
- Bouwmeester, D., Pan, J.-W., Mattle, K., Eibl, M., Weinfurter, H., & Zeilinger, A. (1997). Experimental Quantum Teleportation. *Nature*, 390(6660), 575–579. doi:10.1038/37539

- Devoret, M. H., & Schoelkopf, R. J. (2013). Superconducting circuits for quantum information: An outlook. *Science*, 339(6124), 1169–1174. doi:10.1126/science.1231930 PMID:23471399
- DiVincenzo, D. P. (2001). Dogma and heresy in quantum computing. *Quantum Information & Computation*, 1(4), 1–6.
- Dutt, M. G., Childress, L., Jiang, L., Togan, E., Maze, J., Jelezko, F., & Lukin, M. D. et al. (2007). Quantum register based on individual electronic and nuclear spin qubits in diamond. *Science*, 316(5829), 1312–1316. doi:10.1126/science.1139831 PMID:17540898
- Furusawa, A. (1998). Unconditional Quantum Teleportation. *Science*, 282(5389), 706–709. doi:10.1126/science.282.5389.706 PMID:9784123
- Home, J. P., Hanneke, D., Jost, J. D., Amini, J. M., Leibfried, D., & Wineland, D. J. (2009). Complete methods set for scalable ion trap quantum information processing. *Science*, 325(5945), 1227–1230. doi:10.1126/science.1177077 PMID:19661380
- Imamog, A. et al.. (1999). Quantum Information Processing Using Quantum dot spins and cavity QED. *Physical Review Letters*, 83(20), 4204–4207. doi:10.1103/PhysRevLett.83.4204
- Karski, M., Forster, L., Choi, J.-M., Steffen, A., Alt, W., Meschede, D., & Widera, A. (2009). Quantum walk in position space with single optically trapped atoms. *Science*, 325(5937), 174–177. doi:10.1126/science.1174436 PMID:19589996
- Kiraz, A., Atatüre, M., & Imamoglu, A. (2004). Quantum-dot single-photon sources: Prospects for applications in linear optics quantum-information processing. *Physical Review A*, 69(3), 032305. doi:10.1103/PhysRevA.69.032305
- Knill, E., & Laflamme, R. (1998). Power of One bit of Quantum Information. *Physical Review Letters*, 81(25), 5672–5675. doi:10.1103/PhysRevLett.81.5672
- Leuenberger, M. N., & Loss, D. (2001). Quantum computing in molecular magnets. *Nature*, 410(6830), 789–793. doi:10.1038/35071024 PMID:11298441
- Maurer, P. C., Kucsko, G., Latta, C., Jiang, L., Yao, N. Y., Bennett, S. D., & Lukin, M. D. et al. (2012). Room-temperature quantum bit memory exceeding one second. *Science*, 336(6086), 1283–1286. doi:10.1126/science.1220513 PMID:22679092
- NSF. (1999). QIS: An Emerging Field of Interdisciplinary Research and Education in Science and Engineering. In *Proceedings of Workshop QIS*. NSF.
- Ourjoumtsev, A. (2006). Generating optical Schrödinger kittens for quantum information processing. *Science*, 312(5770), 83–86. doi:10.1126/science.1122858 PMID:16527930
- Paul, P.K. (2013a). Quantum Information Science: Emerging Basic Science Focused Information Science domain. *Abhinav National Journal of Science and Technology*, 2(9), 7-14.
- Paul, P. K. (2013b). Quantum Information Science-Domain of Future Information Management in Organization and Enterprises: Emerging Interdisciplinary Scenario. *Journal of Management*, 3(2), 543–551.
- Paul, P. K. (2015). MSc Information Science (Quantum Information Science) Specialization: Its Requirement and Proposed Model for Future I-School Dedicated to Next Generation Information-Technology-Community Interaction. *SRELS Journal of Information Management*, 52(2), 9–15.
- Shor, P. W. (1994, November). Algorithms for quantum computation: Discrete logarithms and factoring. In *Foundations of Computer Science, 1994 Proceedings., 35th Annual Symposium on* (pp. 124-134). IEEE.

Vlastakis, B., Kirchmair, G., Leghtas, Z., Nigg, S. E., Frunzio, L., Girvin, S. M., & Schoelkopf, R. J. et al. (2013). Deterministically encoding quantum information using 100-photon Schrödinger cat states. *Science*, 342(6158), 607–610. doi:10.1126/science.1243289 PMID:24072821

Zhang, Z. J., & Cheung, C. Y. (2007). Minimal classical communication and measurement complexity for quantum information splitting. *Journal of Physics. B, Atomic, Molecular, and Optical Physics*, 41(1), 015503. doi:10.1088/0953-4075/41/1/015503

Zoller, P. (2005). Quantum information processing and communication. *The European Physical Journal D-Atomic, Molecular, Optical and Plasma Physics*, 36(2), 203–228.

ADDITIONAL READING

Biolatti, E., Iotti, R. C., Zanardi, P., & Rossi, F. (2000). Quantum Information Processing with Semiconductor Macroatoms. *Physical Review Letters*, 85(26), 5647–5650. doi:10.1103/PhysRevLett.85.5647 PMID:11136068

De Belle, J. S., & Heisenberg, M. (1994). Associative odor learning in *Drosophila* abolished by chemical ablation of mushroom bodies. *Science-AAAS-Weekly Paper Edition-including Guide to Scientific Information*, 263(5147), 692–694. PMID:8303280

Doi, J. (2012, November). Peta-scale lattice quantum chromodynamics on a Blue Gene/Q supercomputer. In *High Performance Computing, Networking, Storage and Analysis (SC), 2012 International Conference*. pp. 1-10

Geremia, J. M., Stockton, J. K., & Mabuchi, H. (2004). Real-time quantum feedback control of atomic spin-squeezing. *Science*, 304(5668), 270–273. doi:10.1126/science.1095374 PMID:15073372

Hadfield, R. H. (2009). Single-photon detectors for optical quantum information applications. *Nature Photonics*, 3(12), 696–705. doi:10.1038/nphoton.2009.230

Lukin, M. D., Fleischhauer, M., Cote, R., Duan, L. M., Jaksch, D., Cirac, J. I., & Zoller, P. (2001). Dipole blockade and quantum information processing in mesoscopic atomic ensembles. *Physical Review Letters*, 87(3), 037901. doi:10.1103/PhysRevLett.87.037901 PMID:11461592

Steger, M., Saeedi, K., Thewalt, M. L. W., Morton, J. J. L., Riemann, H., Abrosimov, N. V., & Pohl, H.-J. et al. (2012). Quantum Information Storage for over 180 s using donor spins in a ²⁸Si semiconductor vacuum. *Science*, 336(6086), 1280–1283. doi:10.1126/science.1217635 PMID:22679091

Tamulis, A.,.... (2004). Quantum mechanical design of molecular computers elements suitable for self-assembling to quantum computing living systems. In *Solid State Phenomena* (Vol. 97, pp. 173-180). doi:10.4028/www.scientific.net/SSP.97-98.173

Vidal, G., Latorre, J. I., Rico, E., & Kitaev, A. (2003). Entanglement in quantum critical phenomena. *Physical Review Letters*, 90(22), 227902. doi:10.1103/PhysRevLett.90.227902 PMID:12857342

Vion, D. et al.. (2002). Manipulating the quantum state of an electrical circuit. *Science*, 296(5569), 886–889. doi:10.1126/science.1069372 PMID:11988568

Zheng, S. B., & Guo, G. C. (2000). Efficient scheme for two-atom entanglement and quantum information processing in cavity QED. *Physical Review Letters*, 85(11), 2392–2395. doi:10.1103/PhysRevLett.85.2392 PMID:10978018



KEY TERMS AND DEFINITIONS

Computing: Computing includes designing and building hardware and software systems for a wide range of purposes; processing, structuring, and managing various kinds of information; doing scientific studies using computers; making computer systems behave intelligently; creating and using communications and entertainment media; finding and gathering information relevant to any particular purpose, and so on.

Information: Information as a concept has a diversity of meanings, from everyday usage to technical settings. Generally speaking, the concept of information is closely related to notions of constraint, communication, control, data, form, instruction, knowledge, meaning, mental stimulus, pattern, perception, and representation.

Information School: I-Schools are academic units which work for professional and academic solution of Information and Technologies. Information Schools are responsible for providing manpower in the field of Information and Technologies and Computing. I-Schools gained popularity during 1990's when internet reached the common people and other commercial ventures. It is in this time the explosion of information has lead the development of healthy information management approach. Earlier evolution of I-Schools provided such knowledge as Library Science, Communication Science, Mass-Communication, Computer Science and Information Technology.

Information Science: Information Science is actually a discipline of disciplines. It has come from the subject which deals with information, information services, information activities and digital information. So mainly from library science, archive science, documentation science, computer science it has originated; however this is not truly computer science nor library science. This is an interdisciplinary science; which is a combination of Library Science, Computer Science, Information Technology, Management Science, Operation Research and the subjects that deal with information.

Interdisciplinary: Connection between some subjects irrespective of domain and direct and indirect connection with main facet or sub field.

Quantum Computing: Computers which are run according to the power and system of Quantum Bit and only employed for faster computer operation.

Quantum Information Science: Quantum Information Science [QIS] is a new field which is responsible for faster and advance Communication powered by QBIT. Actually Quantum is the main pillar of Quantum Science. The integration of Quantum Science and Information Science may treat as Quantum Information Science [QIS]. Quantum Computer plays an important role to build healthy Quantum Information Science [QIS] practice. Today's super computer and in future computer QBIT may solve hard problem and may able in speed up communication in Computation.

Towards a General Theory of Information



Laura L. Pană

Polytechnic University of Bucharest, Romania

INTRODUCTION

Today's society and our entire epoch are often characterized as the information society, respectively the information epoch, which is the current stage of the technical era. The technology developed on information bases – Information Technology – is now exhaustively studied and further rapidly developed. Consequently, technical information is nowadays the most accessed. This article attempts to bring nearer to completeness the theory of information by analyzing social and human information, existing in various, human-like and human-made structures, processes and activities. The complex status of information is based on its double quality, namely that of an essential component of existence at all its structural levels, and that of a founding concept for many scientific, technical and philosophical disciplines. An interdisciplinary study of information is thus accomplished, by using findings from several scientific and philosophical disciplines, from Information Epistemology or Information Aesthetics to Neuroinformatics and Neurorobotics. Various information types, specific for different existential levels, such as structural information, functional information or free information will be characterized and defined, as well integrated in a general overview of the world of information. New research themes such as information values, information efficiency and information responsibility are proposed at the end of article.

BACKGROUND

Intellectual life is dominated today by information-related activities such as information generation,

information processing, evaluating, managing and using. Under these conditions, even society, as well as the whole existence could appear as a hierarchical succession of informational structures, processes and activities. This information-centered vision is supported by a series of scientific representations and philosophical outlooks. Some outstanding findings of these cognition area show, respectively argue that information is present, active and has a decisive importance at each layer and in every field of existence: in nature, society and technology. In the field of physics, various structural layers of existence are often associated with certain informational features. Sometimes the hypothesis of generation of distinct existence forms by various information contents is formulated. At other times, even a distinct kind of matter is postulated – the so named *informatter* –, characterized by some deep constitutive information sorts, such as electric charge, lepton or baryon number, charm and strangeness (Drăgănescu, pp. 52, 228, 219).

The synthetic concept of *informational energy* was coined and elaborated by means of mathematical statistics (Onicescu, pp. 4-5). The formula associated with this concept makes possible an overall characterization of the organization state of a system: when the system is completely disorganized, its informational energy is minimal, and when the states of the system tend to be reduced to a single state, its organization is total and the informational energy to maximum.

Energy, organization and information are correlated and within different technical approaches. Norbert Wiener was the first who explicitly stated that information is that what counteracts the degradation of energy as the increasing of disorder in systems, and who defined information as negative entropy or *negentropy*. In other, kindred, but more

DOI: 10.4018/978-1-5225-2255-3.ch387

nuanced approaches, as that of J. Guillaumaud, who accept only an analogy between negentropy and information, this latter is understood as *potential anti-entropy*, with the same role: that of a measure of the degree of organization of a system. Recently, datastore technologies were enriched with *active anti-entropy* applications, by which can be developed faster-growing Web-based, mobile and social networking activities, as well as cloud services, used to solve conflicts between information object replicas stored on different nodes or between concurrent client updates, and devoted to avoid physical data loss and corruption, and other events that distributed systems are built to handle. A *passive anti-entropy* system heals information object conflicts only when a read request reaches the web service from a client: any conflicts between objects that are not read by clients will go undetected. In an active anti-entropy technology, conflict resolution runs as a continuous background process, by which databases are updated in real time as new writes come in, which reduces the time that it takes to detect and repair missing or divergent replicas.

Some authors studied *information entropy* and more, its role in *social entropy*. They demonstrated that the quantity of information in a social system depends only on its complexity and dynamics (Şileţchi & Lascu, p. 19), but also that important in social entropy reducing are not only order and disorder, but also orientation or disorientation of social processes (p. 37-53); a well oriented social development can stabilize the variation of social entropy and thus can decrease the probability of crises in the society - nature interaction as in society (p. 130). Aiming to clarify some structural and operational aspects of human psychology by applying an informational research perspective, some scholars discover that information itself can be alienated in psychiatric diseases or can be alienating, when induces incertitude in individual or public mind, and indecision at social level. They use the notion of *informational conscience* (Pamfil & Ogorescu, p. 186) and propose a genetic definition of information (pp. 190-201).

By the concept of *quasi-information* R. Ruyer aims to explain and to give a content to the idea that information is inserted in matter even from the layer of elementary particles, but he also believes that the term designates everything that can be subsumed to notions such as potentiality, sense, finality, instinct or competency. He specifies that consciousness is the most important, because it alone can induce order in the world.

Other authors, such as G. Simondon, refuse to reduce the extension of information as a concept to its technical meanings, and argue that it perfectly responds simultaneously to the exigencies of quantum physics, biology and psychology or social studies, because it is linked only with energy and structure, and because of its purely operative feature (Simondon, pp. 219-220). He goes even further and – as shoved by one of his disciples – reforms the concept of information both in terms of system's meta-stability and of processes of transductivity (Bardin, p. 27).

An even larger extension of the term is proposed by J. Zeman, who studies the applications of the concept of information channel both to cosmic information flows and epistemic processes, and shows the role of information as the main negentropic factor in knowledge and learning (Zeman, 247). The concept of *informational time* and its measurement in these kinds of processes are initiated by him (p. 253).

Several scientific disciplines highlight the status of information in existence by imposing the concept of *structural information*. This notion, which also circulates in the form of “molecular information”, is used in biology and in macromolecular biochemistry, which is also called informational biochemistry. In these scientific areas have been developed methods of calculating the amount of structural information necessary to obtain complex structures (copolymers) or to measure the entropy of cell in multiplication (functional or abnormal).

Renowned representatives of Information Science and Technology are recognized for their practical approaches which are, however, in deep

accordance with the largest philosophical outlooks, specifies that "a huge number of physical things, events and actions can be described as being information" (Dertouzos, p.59). Other objects, like computing systems are able not only to use and to produce information, but also to assist humans in inaugurating new ways of knowledge (p. 173). An aspect of this forecast is illustrated by other authors, who consider that computers are powerful tools to process, to understand and to build a large variety of non-sentential representations of information, and to increase the efficiency of human reasoning (Barwise & Etchemendi, pp. 112). Such kind of technical approaches and endeavors lead to a more general concept of information that transcends the technical notion of information.

At the social level of existence the notion of information circulates with multiple significations, such as the meaning of a string of conventional signs; the content of a message; the denotation of a concept; the sense of a statement, the content of knowledge or the totality of the available elements of an expert system. From the perspective of this article, the significance of an action, the products of various activity fields, the sense of evolution of an event, and even the generative structure of a class of entities have to be added, because all these entities, structures or processes contain or even constitute information organized in various forms.

MAIN FOCUS OF THE ARTICLE

Two main objectives will be followed in this article, namely:

- To show the hierarchy of natural, social and human information structures, processes and activities;
- To integrate the various information levels, structures and features in order to ensure a valid and suitable scientific and philosophical foundation of a General Theory of Information.

The World of Information and Its Hierarchical Structure



At each level of existence can be identified forms of structuring and development, substantial contents and energetic processes that generate, incorporate or suppose information. But the existential status of information is different at distinct levels of existence and not independent of their energetic and structural properties, as Simondon stated. The physical level of existence contains information as a result of substantial processes and of energetic interactions. For the biotic level of existence is characteristic that it uses information from the environment as also its own information, with a well-determined finality. Human mind produces and processes information which is super-structured to the previous, physical and biotic levels, here information being characterized by structural and functional autonomy over previous levels. In social systems information is freed from some material constraints (physical and biotic), as well as from its subjective (psychic) determinations, at the high level intellectual activities, and is instead captured in spiritual structures in cultural forms. This is possible because:

- The social cybernetic complex of connections is open
- Social cybernetics admits feedforward and feedback type connections, not only feedback type ones
- A part of information is self-generating; people are initiators, assessors and users of this kind of information

Synthetically speaking, the physical universe contains *structural information*, the living world is organized and evolves through *functional information*, and the social system (which is an extension of natural life and the result of development in mental life) produces *free information*.

Natural living world closes information in the cybernetic circle of relationships between organism, heredity and environment. In the adaptation

process, it is the organism that changes its internal structure and functioning rather than the medium itself. In these circumstances, the information is constantly consumed, recycled and regenerated in the same forms, without this process to substantially affect the energetic or informational balance of nature. At the psychological level, the information flow is open, namely in two main directions: that determined by the purposes and results of its processing in the person's internal environment, and that opened by external (natural and cultural) possibilities. The psyche itself becomes autonomous from its own physical and biotic support, sometimes even of that social. The next level of existence, that social, is characterized by an omni-directional circulation of information. In some of its forms – mainly those highly abstract –, information may be released from all its previous bonds and, very important, a process of new information generation and even self-generation becomes possible. At this level, the object, the means and the products of activity are now becoming more and more informational.

Social Information and Its Internal Structure

Any social information has a complex internal structure, composed of at least three concentric layers.

Structural information is that which originates in the succession of levels at which information is manifested in the whole existence and which contains, at the same time, the general fund of information on the social environment. It can also be called *contextual information*, as well as *environmental* or *background information*. This type of information is included in any social information, no matter how specialized, but in concise or diffuse, implicit or tacit ways. Therefore, it cannot yet be explored and exploited by artificial information systems. Each type of social information (whether scientific, technical, political, moral or artistic) integrates, in addition to its specific content, several contextual elements and

the structural level of social information includes, in turn, and knowledge in the hard, scientific meaning, such as universal laws.

Systemic information is the specific informational content of a social unit. Such a unit may be a whole social system, a cultural area or, for instance, one of the sciences which belong to the scientific culture. The last unit of science is theory, because an acquaintance has no meaning and cannot be true than under or in terms of a theory. It can be also part of an informational unit by excellence, such as a sentence, a theorem, or a specialized database. Systemic information can be called *specific information* too, because contains information about the object, means and products of activity in well defined social fields;

- It is the one involved in the production of novelty in the social system.
- This category of information is systematized in disciplines and developed through coherent research programs: scientific, technical, philosophical etc.
- If, at the first social information level, the main function of information is that of prerequisite for knowledge, at the second level it is that of foundation for social action, be it theoretical or practical.

Active information is that which “puts to work” specific information, and at this level, structural and specific scientific information are translated into practical rules, working methods and appropriate techniques for different activity fields. This type of information can also be called *generative information*, for the following major reasons:

1. Contains in an unmediated manner the generative structure on which relies the production of new social structures;
2. Represents and permanently reproduces the deep intellectual structure of humankind itself;
3. Constitutes a condition of manifestation of his creative faculty;

4. It is the information that can be considered the core of individual creativity, and can therefore be named *nuclear information*;
5. A well informed and cultivated person, more exactly, the personality - the only creative social entity - becomes the source of the new and significant social information.

The active, generative information can, in turn, to be organized in various ways.

Information-Shift Phenomena of the Information Age

Several changes which shape the features of the information structure of present society can be here only concisely outlined and described.

- **The Wrap-Up Phenomenon of Social Information:** Each unit of social information is composed of concentric information strata; in our times a few meaningful restructuring phenomena occur within the information clew:
 - The mental functions, operations and processes, as well as the conceptual structures which constitute the generative structure of the entire cognitive and creative activity, are now modeled and simulated by the computer, that changes its own status of intelligence amplifier and takes over or even change a few cognitive structures, functions and activities;
 - Second information stratum constitutes the most consistent part of cognitive information, that becomes, almost entirely, scientific and technical information; it is composed of
 - The results of various specific creativity fields, together with
 - The appropriate cognitive instruments and skills, languages and cognitive styles, which are also changed and adapted to

new knowledge work types and environments;

- The exterior stratum now becomes, by contrary, more and more inconsistent, even if it has, in this context, a comparable role with that of atmosphere among other geo-spheres. It provides two types of information, namely:
 - Data, representations and beliefs about nature, civilization and culture;
 - Information about the socio-technical system, mainly about information structuring, processing and transmitting systems, techniques and skills.
- **Present Changes at Various Levels of Social Information:**
 - Transition phenomena from systemic information to general information that has as a result a phenomenon of informational “subduction” and finally, the rising of the general level of social information;
 - Processes of transformation from general information to active information, often manifested as “melting” and dissipation of the pyramid of knowledge;
 - A constant deepening of general information and the corresponding hypertrophy of specific information, a process which results in the building of a double pyramid, with the inversed pyramid of specific information in an impossible equilibrium on the pick of the pyramid of general information.

This analytic but concise presentation of the variety of information forms, of their contemporary hierarchy and of their present changes allows the building of an integrative perspective and of a general overview on information, able to show the totality of structural levels at which is or can



be organized the amount and the variety of information which is generated, stored, processed and used worldwide today.

SOLUTIONS AND RECOMMENDATIONS

Taking the envisaged spectrum of scientific results as acquired bases, and the above presented and synthesized original findings as well-founded knowledge, a number of explanatory definitions of information can be formulated here, from more factual to more general ones.

Information is the generative structure by which the world is self-organized, by using the existing content and the specific properties of its levels, and by taking shapes selected from the set of inherent possibilities of these levels. (1)

Information appears, from this perspective, as an intern-structural component of the whole existence. More specifically said, the world is structured through various kinds of components, properties, relationships, functions and products, some of which being informational. This statement may result in a new design of the concept.

Information is an irreducible manifestation of the entire hierarchy of existence. (2)

Given the multiple and active relation of information with the nature, structure, functions, relations and interactions of systems, as well as with the sense of evolution of systems which compose the whole existence, a new conceptual determination of information follows.

Information belongs to the very essence of existence. (3)

A more detailed analysis leads to another ontological definition.

Information is the set of internal, specific and stable properties of entities and processes, both material and energetic. (4)

Thus, information is not only belonging to the essence of existence, as shown in (3).

Information is the essence of existence. (5)

But the most promising hypothesis on the nature of information may be derived from N. Wiener's conclusion that information is neither matter none energy.

This thesis allows the possibility to see reality as including three basic forms of existence: matter, energy and information.

Therefore, another even more powerful definition of information can be considered here.

Information is a fundamental form of existence, just like matter and energy. (6)

Information can be interpreted quantitatively as a physical entity, such as mass and energy, but it is qualitatively different from them, because it refers always to internal, specific and stable features of a particular system or of an organization layer of systems. As shown, it is just the opposite of physical entropy, is negative or anti-entropy.

Information differs from both matter and energy, but exists beside and together with them.

Very important for the aim to build a General Theory of Information is that an information status may be acquired both by matter and energy, under certain conditions and – at the social level especially - in the presence of some appropriate devices.

At conceptual level, matter, energy and information may be mutually defined. Matter is named “slowed energy” by a few physicists, while energy may be seen as wavy matter. Information is matter reduced to its deepest internal structure, while reported to energy information can be seen as a both structure and evolution generating power.

A distinction made between two kindred concepts - informativeness and informativity – is also useful to explain the very status of information within existence.

Informativeness regards a human or artificial information source or product (a sentence or a diagram, a device or a system that processes signals) and shows the information potentiality of the source. Barwise and Etchmendy specified (1998) that the concept expresses the amount of information in a signal, that can be measured by the number of possibilities eliminated by that signal. L. Floridi studied the *degree of informativeness* (Floridi, 2003, p. 54), as the possibility to calculate it (p. 55) and established that the average informativeness of a string of symbols depends on the probability of occurrence of each symbol. Thus, more possibilities of occurrence eliminated, the more information can be extracted.

Informativity is a concept which can be applied to a larger category of entities or processes and - from the perspective of this work, as on the basis of the already gathered arguments - even to the totality of existing entities. Informativity is a fundamental property of existence, just like dynamism or connectivity, structurality and infinity. Informativity consists of the capacity of universe, as well as of each its component, of integrating, of structuring and developing, of transmitting and using information. Thus, each entity or process and product can be both informed and informer, informing and receiving information.

The path towards a General Theory of Information is also paved by contributions of the philosopher M. Bense, who also studied mathematics and physics, and formulated the *co-reality* hypothesis, within his aesthetic theory, which passed from metaphysical aesthetics to mathematical aesthetics and then to informational aesthetics. Bense also coined the concept of *aesthetic information* but, more important, he made a generalization of the concept of “aesthetic state”, applied both to natural, cultural and technical objects, any object of reality being both a support for sign (vehicle of

information) and a sign. In his view, information is not an aspect of reality, none one of its properties or functions, but is another reality.

A number of other ways to study information – mathematical and technical - also arrive at meaningful generalizations. Any event can be treated with informational tools and, in conclusion, reality itself, in its entirety, consists of various information states and their changes, i.e. events, and has an informational nature or simply, is information.

Brain itself, as well as the whole string of mental processes and products it supports can be studied and represented as different-order information states (Pană, 2016). A series of scientific and philosophical disciplines explore these processes and products from an informational perspective. Neuroinformatics studies brain both as a biotic system structured from many kind of specific cells with their own genetic information, and under its wave aspect, as in the Brit Brain Wave Project (2011-2012), and with not only theoretical, but also practical purposes: NASA invests in Neurorobotics to create Robonauts or even more, spatial robots, which will be send as light beams that will contain all the necessary information to be materialized where they will work.

Information Epistemology, where the terms *information permeability* and *information competency* were coined, is now continued with means of Information Science and Technology, e. g. by studying how “knowledge objects” are update by various, upward, horizontal or downward informational permeability processes, which have a specific role in information propagation within society (Borgo & Pozza, pp. 99-100).

Experimentally validated sets of information competence are elaborated and used now in universities and described in collaborative works such as (Pinto et al.). These tests were also adapted to self-assess and report perceptions of information competence along with intergroup experiments, in which variables such as text comprehension, interest, quality and credibility were evaluated (pp. 440, 446). Information competency has more



than individual and special, educational meanings and roles, as showed in (Pană, 2000, pp. 167-168). The term covers:

- The capability of a person or of a system to search, acquire and store information,
- The ability of efficiently use the available information, and mainly,
- The human power to generate new information.

Other, social aspects of information competence are to:

- Gather, select, order and to update information with social content and importance;
- Structure social information in order to meet the specific needs of various decision authorities;
- Keep social information ready to be used;
- Distribute it at all institutional levels with economic, social or educational competences, in order to ensure an effective use of the totality of available social information.

An information continuum, which roams through every field of existence and which connect all its structural layers may be distinguished. The social space-time continuum makes evident the same information continuum, but with a few particularities, such as:

1. A differentiated request and use of information flows and reserves, which may be excessive, sufficient or even deficient, in distinct moment and various situations;
2. Not only the quantity of information that circulates in specific fields and channels, but also the presence of social opportunities for using or generating new information is also important;
3. The degree in which information is involved in social processes and is, thus, productive, has large variations;

4. Socialization of information exceeds its technical, economic and administrative using: both humans and societies are information-oriented;
5. Technological responsibility is here extended and now even deepened by a new kind of responsibility - *informational responsibility*;
6. A special type of social efficiency is now emerging – *informational efficiency* – in the more and more large and dense field of information activities.

The present society is characterized by information thinking, information action, information humankind as well as by information machine and, mainly in this, *stricto sensu*, it can be taken as an information society.

FUTURE RESEARCH DIRECTIONS

- To deepen Information Science and Information Technology by studying technical information and its specificity. Several ways to do this are:
 - To discuss with new theoretical instruments the present meanings and relationships between concepts such as data, information, cognition, action and responsibility, in order to re-evaluate the results of earlier debates, animated by scholars such as J. P. Dupuy, Sybille Krämer, J. Mingers or G. Rapohl;
 - To envisage technical information as part of social information, as well as its formatting social role, by studying new social realities such as informational thinking, informational action, informational machine, informational man and informational society;
- To continue exploring human structures and processes such as brain, mind, cognition, consciousness and creation, from an informational perspective;

- To extend the study devoted to the world of information in order to develop an authentic information culture;
- To study the constitutive entities of the information culture: *information values*.

CONCLUSION

A General Theory of Information is now both necessary and possible. Its building is even more feasible than a general theory of physical particles, of physical interactions or of the main physical forces, which are constantly followed but still unrealized goals of scientific community. These theoretical constructs are hard to be accomplished because human scholars aim to produce universal truths although their findings are acquired in a “corner” of the universe. By contrary, humans are in the center of the information world and more, human brain and therefore humankind, are the most complex and advanced information systems. However, to build a General Theory of Information, a series of prerequisites are still needed, ones that are consisting not necessarily of a newer or greater amount of information, but of new approaches, interdisciplinary and integrative inquiries, and maybe mainly of a more advanced reflection capability.

REFERENCES

- Bardin, A. (2015). Reforming the Concept of Information. In A. Bardin (Ed.), *Epistemology and Political Philosophy in G. Simondon, Individuation, Techniques, Social Systems*, (Vol. 19, pp. 27–31). Dordrecht: Springer.
- Barwise, J., & Etchemendi, J. (1998). Computers, Visualization, and the Nature of Reasoning. In T. W. Bynum & J. H. Moor (Eds.), *The Digital Phoenix: How Computers are Changing Philosophy*. Oxford, UK: Blackwell Publishers.
- Borgo, S., & Pozza, G. (2009). Disentangling Knowledge Objects. In R. Ferreiro & A. Oltramari (Eds.), *Formal Ontologies Meet Industry* (pp. 92–101). Berlin, Germany: IOS Press.
- Dertouzos, M. (2000). *What Will Be: How the New World of Information Will Change Our Lives, translated from English*. Bucharest: Technical Publishing.
- Drăgănescu, M. (1989). Depths of the Material World. In M. Drăgănescu (Ed.), *The Ring of Material World*. Bucharest: Scientific Publishing.
- Dupuy, J. P. (1980). Myths of the Informational Society. In K. Woodward (Ed.), *The Myths of Information: Technology and Postindustrial Culture*. Madison, WI: Coda Press.
- Floridi, L. (2003). Information. In L. Floridi (Ed.), *The Blackwell Guide to the Philosophy of Computing and Information* (pp. 40–61). Oxford, UK: Blackwell. doi:10.1111/b.9780631229193.2003.00006.x
- Șilețchi, M., & Lascu, A. (1978). *Information and Entropy in Social Processes*. Bucharest: Academic Publishing.
- Krämer-Friedrich, S. (1986). Information Measurement and Information Technology: A Myth of the Twentieth Century. In C. Mitcham & A. Hünig (Eds.), *Information Technology and Computers in Theory and Practice* (pp. 17–28). Dordrecht: D. Reidel Publishing Company. doi:10.1007/978-94-009-4512-8_2
- Mingers, J. (1997). The Nature of Information and its Relationship to Meaning. In R. L. Winder et al. (Eds.), *Philosophical Aspects of Information Systems* (pp. 73–84). London: Taylor and Francis.
- Onicescu, O. (1966). Informational Energy: A Component of a Statistical Barometer of Systems. *Revista de Statistica*, 11(1), 4–5.
- Pamfil, E., & Ogorescu, D. (1973). *Psychology and Information*. Bucharest: Scientific Publishing.

- Pană. (2006). Co-evolution of Human and Artificial Cognitive Agents. *Proceedings of Computers and Philosophy: An International Conference*, (pp. 366–378). Laval, France: Institut Teknologi de Laval, Le Mans Université.
- Pană. (2007). From Information Flows and Nets to Knowledge Groups and Works, *European Computing and Philosophy Conference*. Twente University.
- Pană. (2016). *Structure and Content in Brain Work: How Ideas are Generated? Efficiency types and degrees in Human and Artificial Intelligence*. Paper presented at the National Conference of the Romanian Brain Research Group, Bucharest, Romania.
- Pană, L. (2000). Technical Values. In L. Pană (Ed.), *Philosophy of Technical Culture* (pp. 162–170). Bucharest: Technical Publishing.
- Pană, L. (2004). *Philosophy of Information and Information Technology*. Bucharest: Politehnica Press.
- Pană, L. (2008). An Interdisciplinary and Integrative Model of Intellectual Activities. *Noema*, 7(1), 120–137.
- Pană, L. (2008). Sociomatic Systems: A Challenge for Social and Intellectual Invention. In R. Trapl (Ed.), *Cybernetics and Systems* (Vol. 2, pp. 325–330). University of Vienna.
- Pană, L. (2010). Crucial Intellectual Events in the History of Information Science and Technology. *Noesis*, 34(1), 171–196.
- Pană, L. (2015, May). *New Ways of Thinking Associated with the Virtual Work Environment*. Paper presented at the Annual Symposium of the Logic, Methodology and Philosophy of Science Division of the Romanian Academy, Bucharest, Romania.
- Pinto, M. (2013). Designing and Implementing Web-Based Tools to Assess Information Competences of Social Science Students at Spanish Universities. In S. Kurbanoglu et al. (Eds.), *Worldwide Commonalities and Challenges in Information Literacy Research and Practice* (pp. 443–449). New York: Springer. doi:10.1007/978-3-319-03919-0_59
- Rapohl, G. (1986). Information Does Not Make Sense or: The Relevance Gap in Information Technology and its Social Dangers. In C. Mitcham & A. Huning (Eds.), *Information Technology and Computers in Theory and Practice* (pp. 63–74). Dordrecht: D. Reidel Publishing Company. doi:10.1007/978-94-009-4512-8_5
- Simondon, G. (2005). *L'individuation à la lumière des notions de forme et d'information, Première édition 1964, Paris, P.U.F., Éditions Jérôme Milon*. Grenoble: Collection Krisis.
- Zeman, J. (1975). Information, Knowledge and Time. In L. Kubátka & J. Zeman (Eds.), *Entropy and Information in Science and Philosophy*. Praha, Amsterdam: Elsevier.

ADDITIONAL READING

- Burgin, M. S. (1997). *Fundamental Structures of Knowledge and Information*. Kiev: Academy of Information Sciences.
- Deegan, M., & Tanner, S. (2002). *Digital Futures: Strategies for the Information Age*. London: Library Association Publishing.
- Dobrescu, R., & Al Iordache, D. (2010). Informing Science as Conceptual Framework for Developing Information Systems. In R. Dobrescu and D. Al Iordache, *Complexity and Information* (pp. 61–71), Bucharest, Editura Academiei Romane.

Drăgănescu, M. (2001). The Information and Knowledge Society: Vectors of the Knowledge Society. In Fl. Gh. Filip (Ed.), *Information Society – Knowledge Society: Concepts, Solutions and Strategies for Romania*, (pp. 43-112), Bucharest, Editura Expert.

Pană, L. (2000). Informational Thinking, Informational Action, Informational Man and Informational Society. In L. Pană (Ed.), *Philosophy of Technical Culture*, (pp. 439-452), Bucharest, Editura Tehnica.

Pană, L. (2002). Characteristics of Knowledge in the Technical Intellectual Environment of Informational Type. *Noema*, 1(1), 116–128.

Pană, L. (2004). Information Management. In L. Pană (Ed.), *Philosophy of Information and Information Technology*, (pp. 103-149).

Pană, L. (2004). The Technical Intellectual Environment or the Artificial Infosphere. In L. Pană (Ed.), *Philosophy of Information and Information Technology*, (pp. 49 – 55).

Pană, L. (2005) Values Computing and Motives Understanding in Social Evolution Modeling, ECAP'05, *European Computing and Philosophy Conference*, NTNU, June 2-4 2005, Trondheim, Norway.

Pană, L. (2006). Information Ethics and Moral Intelligence. *Noema*, 5(1), 116–145.

Pană, L. (2006). Intellectual Techniques and Information Technologies in the Transition to the Knowledge Society. *Noesis*, 30-31(1), 69–85.

Pană, L. (2006). The Information Machine and the Ethics of Intelligent Machine. In L. Pană (Ed.), *Cultural Models of Knowledge Society from the Perspective of Technical Culture* (pp. 42–69). Bucharest: Politehnica Press.

Pană, L. (2008). Consciousness as a Third-Order Information State, ECAP'08, *European Computing and Philosophy Conference*, 17-18th June 2008, Montpellier, France.

Pană, L. (2009). From the Machine Man to the Informational and Artificial Man. In L. Pană (Ed.), *Infinity, Possibility, Prediction* (pp. 106–119). Bucharest: Politehnica Press.

KEY TERMS AND DEFINITIONS

Free Information: An information variety which is produced in society in high level intellectual activities, and which is able even of self-generation, under specific human or technical conditions.

Information: Is a basic existence form, the third one after matter and energy.

Informativity: A fundamental property of the whole existence, beside dynamism and connectivity, structurality and infinity; it consists of the capacity of universe, as well as of each its component of integrating, of structuring and developing, of transmitting and using information.

Social Information: Is a multileveled, multifunctional and multi-directed information flow and net.

Structural Information: A kind of information present at all existential levels – nature, society, technology and human mind.

Structurality: An objective, universal and essential feature of existence which consists of its capacity of self-organizing in more and more complex systems and to constitute an infinite succession of relatively homogenous and relatively stable organizing levels.

Systemic Information: Is an information sort characteristic to various social units, from types of civilization, cultural area and social systems, to educational systems and scientific disciplines.

Category I

Information Retrieval

Analysis and Assessment of Cross–Language Question Answering Systems



Juncal Gutiérrez-Artacho
University of Granada, Spain

María-Dolores Olvera-Lobo
University of Granada, Spain

INTRODUCTION

Within the sphere of the Web, the overload of information is more notable than in other contexts. Thus, too often, on planning a search with tools from the Web (search engines, directories, or meta-search engines), the number of web pages found proves excessive and not all of them are relevant or useful for the objectives of the user.

Cross-language information retrieval (CLIR) is an active sub-domain of Information Retrieval (IR). Like IR, CLIR is centered upon the search for documents, reconciling queries and documents which are written in different languages. Cross-lingual information access covers a wide range of tasks that enable users to access information in languages other than their own, including IR, Question Answering, information extraction and summarization. CLIR enables users to find information in languages they do not know, but CLIR search results are not immediately useful because a separate translation techniques must be applied before the user can read the results. This type of systems has opened a new research field that examines the most effective methods for IR and investigates which resources are required for a correct translation.

Question answering systems (QAS) are presented as an alternative to the traditional Information Retrieval (IR) systems, seeking to offer precise and understandable answers to factual questions instead of showing the user a list of documents related to a given search (Jackson

and Schilder, 2005). The functioning of QAS is based on short-answer models (Blair-Goldensohn et al., 2004), and the main advantage that it offers is that the user does not have to consult complete documents to collect the information needed, as the system provides the correct answer in the form of a number, a noun, a short phrase, or a brief fragment of text.

In relation to the types of QAS, the present work focuses on the analysis and assessment of multilingual and cross-language QAS. These systems need to include some type of linguistic translation resource, tool, or technique for the proper retrieval of the result, since the QAS can retrieve the answer from a collection of documents written in languages differing from the one in which the question is formulated. Given that the QAS is presented as a substantial advance in the improvement of IR (Kolomiyets and Moens, 2011), it becomes necessary to determine its effectiveness for the final user. With this aim, 7 studies were undertaken to evaluate: a) in the first two, the linguistic resources and tools used in these systems for multilingual retrieval (Research 1; Research 2); and b) the performance and quality of the answers of the main monolingual and multilingual QA of general domain and specialized domain in the Web (QuALiM, SEMOTE, START, TrueKnowledge, and HONqa) in response to different types of questions (of definition, of facts, and of lists) and subjects (e.g. art, literature, biology, medicine, names, history, economy, or sports), so that different evaluation means can

be applied (Research 3, Research 4, Research 5, Research 6, Research 7).

BACKGROUND

In the field of CLIR tools are being created that can greatly assist specialists in their work; as well as helping other users find a wide variety of information. These tools are evolving but several years of study and research are still needed to improve implementations. One of the main difficulties facing these tools is the task of translating queries made by users and the documentary sources found in response (Diekema, 2003). Given the current expansion in research, development, and the creation of CLIR systems, it was considered worthwhile analysing and evaluating the resources used by one type of these systems: multi-lingual QAS.

Frequently, a keyword query entered into a web search tool (search engine or meta-search engine) to satisfy a user's information need, provides too many result pages – many of which are useless or irrelevant to the user. In effect, modern IR systems allow us to locate documents that might have the associated information, but the majority of them leave it to the user to extract the useful information from an ordered list (Dwivedi & Singh, 2013). In contrast to the IR scenario, a QAS processes questions formulated into Natural Language instead of keyword based queries, and retrieves answers instead of documents (Peñas et al., 2012). Therefore, the usefulness of these types of systems for quickly and effectively finding spe-

cialized information has been widely recognized (Diekema et al., 2004).

While the development of QAS represents progress, the systems nevertheless suffer restrictions. Many were only developed as prototypes, or demonstration versions, and few were marketed. Some researchers have designed and created systems that were presented and discussed at various forums and conferences. However, because the usefulness of the systems was limited to very specific contexts, or because of problems of implementation, only a few of these systems were later developed for end users.

These circumstances have fuelled academic interest in CLIR, and the techniques of natural language processing. This interest led to many conferences dealing exclusively, or partially, with CLIR – such as TREC, the Cross-Language Evaluation Forum (CLEF), the NII Text Collection for IR Systems (NTSIR), the Language Resources and Evaluation Conference (LREC), among others. However, research on CLIR, which mostly began in 1996, has not led to commercial success and so dissemination was limited.

METHODOLOGY

For the purposes of in-depth analysis of QAS, a series of studies has been proposed which are gathered together in seven research articles published in international magazines with peer review and a chapter in a book. The following table provides a summary of the aspects addressed in the analysis performed.

Table 1. Summary table of the contents of the publications

A1: Analysis Oriented at Resources and Linguistic Tools in Multilingual QAS		
Research 1: Analysis of the linguistic resources used in CLEF		Research 2: General analysis of linguistic resources used in QAS
A2: Evaluation Oriented at the System		
Research 3: Identification and analysis of the evaluation measures used for the testing of QAS		
Research 4: Evaluation of QAS as efficient sources of terminological information.	Research 5: Evaluation of general domain QAS against those of specialist domain in the biomedical field.	Research 6: Evaluation of the efficacy of the functioning of general domain QAS.
Research 7: Final evaluation. The objective and subjective evaluation of the multilingual search for answers		



The first analysis (A1) consists of two studies focusing on the analysis and evaluation of the resources and linguistic tools used by multilingual QAS to solve the problem of translation. The first study centres exclusively on the publications presented at *Conference and Labs of Evaluation Forum*, as it is the conference that has most committed itself to this type of system, every year holding a *track* dedicated to this area. The second study extends to a number of contributions made in the field, analysing all the conferences, congresses and forums in the area of IR.

The second analysis (A2), called 'Evaluation oriented at the system', consists of three related objectives. Firstly, all the measures used in the *QA tracks* of the three main QAS evaluation campaigns: *Text Retrieval Conference (TREC)*, *Conference and Labs of Evaluation Forum (CLEF)* and *NTCIR (research 3) Conference*. The different specific *tasks* or *labs* created in each *QA track* have been identified, as have the type of evaluation questions and the evaluation measures used in the different competitions analysed. This first part of the analysis has been essential in order to detect the most recurrent evaluation measures.

The second part of this evaluation consists of three tests (*research 4*, *research 5* and *research 6*) where there is analysis, evaluation and comparison of the monolingual and multilingual QAS, general and specialist, available on the Web; and also of a fourth analysis, where a final evaluation has been made (*research 7*), which brings together all the preceding objectives. This consists of an objective evaluation, as it analyses and evaluates the answers offered by the multilingual QAS, and a subjective evaluation, as it incorporates the results of the earlier studies and analyses the sources of information from which the answers have been taken.

This series of tests, with a perspective as much quantitative as qualitative, has allowed for the evaluation of the functioning and quality of the monolingual and multilingual QAS available on the Web, as well as those which, although developed, have not yet been made available to end users.

RESULTS

First Analysis

The basic premises that have guided this analysis were: a) the use of QAS enables research in multiple languages, provides faster responses, and increases the likelihood of the user obtaining the right answer; b) cross-language QAS try to overcome the language barrier, which is one of the most maxim in the IR field; c) the linguistic tools that most affect efficiency in the field of cross-language QAS must be identified.

This analysis is primarily intended as a general purpose analysis and aims to encompass translation in the study of cross-language QAS. The second general aim is to identify and analyse the linguistic resources and tools found in these systems. Specific objectives include identifying the main types of language resources and tools useful in the multi-lingual IR processes associated with cross-language QAS, and establishing how much use is made of these tools by cross-language QAS.

A semi-empirical methodology was adopted for the two researches and the collection of data about the tools and linguistic resources employed by these systems; as well as their use and implementation. The first stage of the studies focused on identifying the major conferences, meetings, and forums that address cross-language QAS. The aim was to find, analyze, and compare the different types of linguistic resources. Although a growing number of IR conference are held each year, not all include a section devoted exclusively to QAS, and even fewer tackle multi-lingual aspects. In the first research some 165 papers published between 2000 and 2008 at the above, and other, conferences were reviewed, and 315 papers research published between 2000 and 2011 at the above in the second. Over 75% of the published articles were presented at CLEF.

Five main types of linguistic resources used in multi-lingual QA systems were identified following an analysis of the literature. The main resource types were databases, corpora, dictionaries, ontologies, and thesauri. These resources and tools,

along with their various types and subtypes, do not run in the same way and use differing methods of processing information. Sometimes, a single resource was insufficient and several resources were used together to achieve better results.

We found that automatic translators remain the most popular option, despite the fact that the authors of the papers recognize the resulting problems of ambiguity. It is the low computational cost and ease of storage that accounts for the popularity of automatic translation among developers. In our opinion, this tool can be adequate for IR when combined with other resources. However, there have been some changes in the use and incorporation of these resources and tools. The three most popular traditional resources (automatic translators, dictionaries, and corpora) are gradually leaving a widening gap for others — such as ontologies and the free encyclopaedia Wikipedia. In addition, other approaches such as computational grammars are slowly attracting more researchers who are experienced in handling the results they produce. This data suggests that we may see unexpected changes in the future and this area deserves to be studied and evaluated in future research.

Second Analysis

The evaluation of QAS is a major research area that needs attention, especially with the emergence of domain-oriented question answering systems based on natural language understanding and reasoning (Sing et al., 2005). We can find different analyses referring to the evaluation of QA systems, such as those focusing on the evaluation of QA systems on the Web (Athenikos & Han, 2010; Cruchet et al., 2009; Fahmi, 2009; Jacquemart & Zweigenbaum, 2003; Lee et al., 2006; Olvera-Lobo & Gutiérrez-Artacho, 2010, 2011a, 2011b, 2011c, 2012, 2013a, 2013b, 2014, 2015; Terol et al., 2007; Yu et al., 2007; Zweigenbaum, 2005, 2009) or in some of the international evaluation forums (Peñas et al., 2012; Forner et al., 2010).

The third research is a brief review of the TREC, CLEF and NTCIR Conferences from the QA perspective. We present a historical overview of 15 years of QA evaluation tracks using the method of systematic review. We have examined identified the different tasks or specific labs created in each QA track, the types of evaluation question used, as well as the evaluation measures used in the different competitions analyzed. Of the conferences, it is CLEF that has applied the greater variety of types of test question (factoid, definition, list, causal, yes/no, amongst others). NTCIR, held on 13 occasions, is the conference which has made use of a greater number of different evaluation measures. Accuracy, precision and recall have been the three most used evaluation measures in the three campaigns.

After having identified the different evaluation measures proposed in the tasks to be used in the evaluation of the QA systems, it can be seen that around 50 evaluation measures of very different types have been used. It is the traditional measures of IR (precision, recall, accuracy) which continue to be the most used. However, we find variances in the different tracks since the measures are applied so that they are suited best to the particularities of each task and the evaluation needs of each competition. Another of the commonly used measures is MRR (Mean Reciprocal Rank) proposed by TREC, which is very useful for the evaluation of QA systems as it makes it possible to take into account all the answers retrieved by the system and to assign it a reciprocal value in accordance with the ranking of the system. The traditional F-measure, which combines with recall and precision, is also prominent.

In the fourth, fifth and sixth researches, we took a sample of 150-200 definitional questions about different medical from the webpage WebMD, and a sample of 500 questions of different types (definitional, factoid, yes/no and list) about different issues from CLEF and TREC forum. Results obtained for the six systems analysed (START, MedQA, TrueKnowledge, SEMOTE, QuALiM and HONqa) allowed us to evaluate their effective-

ness and their use of different information sources. To evaluate the effectiveness of the systems, we applied several measures of evaluation: MRR, FHS (First Hit Success), TRR (Total Reciprocal Rank), MAP (Mean Average Precision), AveP (Average Precision), Precision and Recall.

Despite certain limitations on the part of both systems (a lack of accessibility for the general public, and insufficient development in some specific areas), we were able to confirm that these QAS are very useful in the retrieval of valid information, with responses from both proving coherent and precise to an acceptable degree. They also help in understanding the information collected and are set to become one of the key tools available to index and organise specialised information. In addition, these approaches together with others are slowly attracting more researchers who are experienced in handling the results they produce. These data suggest that we may see unexpected changes in the future and this area deserves to be studied and evaluated in future research.

Finally, in the seventh research, an analysis has been made of the results offered by the QA multilingual biomedical system HONqa, available in the Web. The study has used a set of 120 biomedical definitional questions (What is...?), taken from the medical website WebMD, which were formulated in English, French, and Italian. The answers have been analysed using a series of specific measures (MRR, TRR, FHS, precision, recall, MAP), as well as sources used to get the responses. The study

confirms that for all the languages analysed the functioning effectiveness needs to be improved, although in the multilingual context analysed the questions in the English language achieve better results for retrieving definitional information than in French and Italian. This may be because the information sources from which the answers are elicited are more numerous in English and have a more appropriate structure for this purpose than in the other languages studied.

The open-domain systems retrieved fewer answers than the restricted-domain systems, with START in last place followed closely by QuALim. In the restricted-domain QA systems, the results increased substantially, especially in the case of HONqa, while MedQA was slightly better than the open-domain systems. The correct answers were present to the greatest degree in START, open-domain system. In relation to the inexact answers, the QA system with the lowest average was HONqa (7.99%). The remaining QA systems offered similar averages. Finally, the incorrect answers were also lower in the open-domain systems, like START.

The value obtained by applying the MRR and FHS evaluation measures indicates that MedQA best ranks answers, as the first correct answer appears at the top of the list of results. This is significant in that the system does not use any sort of rank algorithm to carry out this process but rather it always recurs to the same order of answers based on the knowledge source. FHS

Table 2. Summary of each research in the evaluation oriented at the system

Evaluation Oriented at the System		
Research 3: Identification and analysis of the evaluation measures used for the testing of QAS		
Research 4: Evaluation of QAS as efficient sources of terminological information. QAS: START and MedQA Evaluation measures: MRR, TRR, FHS and precision	Research 5: Evaluation of general domain QAS against those of specialist domain in the biomedical field. QAS: START, QuALim, MedQA and HONqa Evaluation measures: MRR, TRR, FHS and precision	Research 6: Evaluation of the efficacy of the functioning of general domain QAS. QAS: START, QuALim, SEMOTE and TrueKnowledge Evaluation measures: MRR, TRR, FHS, precision, AveP and MAP
Research 7: Final evaluation. The objective and subjective evaluation of the multilingual search for answers QAS: START, QuALim, MedQA and HONqa Evaluation measures: MRR, TRR, FHS, precision, recall, MAP and AveP		

turns out to be a very relevant metric, as users often tend to focus on the first answer obtained, disregarding the rest. The other restricted-domain QA system, HONqa, is in the opposite situation, as it offers the lowest FHS value but has the second highest average in MRR. The precision of the two open-domain QA systems is quite similar, and there is not a great deal of difference between the two metrics, which is explained by the fact that the answers retrieved for each question usually fluctuate from one to three.

However, the TRR metric is greater in the restricted-domain systems as it considers the value of each correct answer based on its position in the list of results. Naturally, this value increases when more results are offered.

However, the recall shows us that the two open-domain QA systems are less exhaustive than the restricted-domain ones. So START presents a recall value of 21.87% and an average of retrieved answers of 1.6, and the values of QuALiM are 22.13%, referring to the recall measure, and an average of three retrieved answers.

Finally, the correlation between the metrics used in this study shows a significant correlation between TRR and MRR because they focus on similar characteristics of the answers retrieved by QA systems.

FUTURE RESEARCH DIRECTIONS

Despite the numerous advantages that these new IR systems offer, the underlying restrictions of the QAS limited their immediate popularity, as well as their general implementation on the web. We believe that this line of investigation may also be continued by extending the analysis and evaluation of the new multilingual systems that have appeared in recent years. As regards Cross-Lingual Information Retrieval in the Translation discipline, it is of considerable interest for the development of cross-lingual resources and tools. Carrying out interdisciplinary and transdisciplinary studies in IR and translation must be continued in order to enrich this area of research.

CONCLUSION

The conclusions obtained from the two evaluations carried out were as follows:

1. In the first analysis, although the first study shows a trend towards the most traditionally popular tools, the second study reflects a little change towards new and more versatile resources. In general, the majority of the resources grow in popularity, which presents a promising future. Although the combination of various resources needs to be studied better, the analysis suggests that, in order to achieve full coverage, the use of two or more resources are necessary for translation, especially for certain language pairs.
2. In the second analysis, although the different forums cannot be compared with each other since each has its own history, tradition and objectives, the third study shows a new scenario which is an excellent testing ground for the relevance of using traditional methods of evaluation, and for proposing and adapting new measures appropriate for this environment.

The system evaluation studies show that the six QAS can be considered good resources for general and specialist information retrieval. As regards the multilingual aspect, the answer retrieval still has a long way to go to reach the levels of efficacy of the general retrieval systems. These systems must improve in terms of information retrieval and presentation of answers.

REFERENCES

- Athenikos, S. J., & Han, H. (2010). Biomedical question answering: A survey. *Computer Methods and Programs in Biomedicine*, 99(1), 1–24. doi:10.1016/j.cmpb.2009.10.003 PMID:19913938

- Blair-Goldensohn, S., McKeown, K., & Schlaikjer, A. H. (2004). Answering Definitional Questions: A Hybrid Approach. *New Directions in Question Answering*, 4, 47–58.
- Cruchet, S., Gaudinat, A., Rindflesch, T., & Boyer, C. (2009). What about trust in the Question Answering world? *AMIA 2009 Annual Symposium*.
- Diekema, A. R. (2003). *Translation Events in Cross-Language Information Retrieval: Lexical ambiguity, lexical holes, vocabulary mismatch, and correct translations* (Dissertation). University of Syracuse.
- Diekema, A.R., Yilmazel, O. & Liddy, E.D. (2004). *Evaluation of restricted domain Question-Answering systems*. Center for Natural Language Processing.
- Dwivedi, S. K., & Singh, V. (2013). Research and Reviews in Question Answering System. *Procedia Technology*, 10, 417–424. doi:10.1016/j.protcy.2013.12.378
- Fahmi, I. (2009). *Automatic term and relation extraction for medical question answering system* (Dissertation). Groningen Dissertations of Linguistics.
- Forner, P., Giampiccolo, D., Magnini, B., Peñas, A., Rodrigo, Á., & Sutcliffe, R. (2010). Evaluating multilingual question answering systems at CLEF. *Proceedings of 7th International Conference on Language Resources and Evaluation*, 2774–2781.
- Jackson, P., & Schilder, F. (2005). Natural Language Processing: Overview. In Brown (ed.), *Encyclopedia of Language & Linguistics* (vol. 2, pp. 503-518). Amsterdam: Elsevier Press.
- Jacquemart, P., & Zweigenbaum, P. (2003). Towards a Medical Question-Answering System: a Feasibility Study. *Proceedings of Medical Informatics Europe*, 463–468.
- Kolomityets, O., & Moens, M. F. (2011). A Survey on Question Answering Technology from an Information Retrieval Perspective. *Information Sciences*, 181(24), 5412–5434. doi:10.1016/j.ins.2011.07.047
- Lee, M., Cimino, J., Zhu, H. R., Sable, C., Shanker, V., Ely, J., & Yu, H. (2006). *Beyond Information Retrieval –Medical Question Answering*. Washington, DC: Estados Unidos.
- Olvera-Lobo, M. D., & Gutiérrez-Artacho, J. (2010). Question-answering systems as efficient sources of terminological information: Evaluation. *Health Information and Libraries Journal*, 27(4), 268–274. doi:10.1111/j.1471-1842.2010.00896.x PMID:21050369
- Olvera-Lobo, M. D., & Gutiérrez-Artacho, J. (2011a). Language resources used in multilingual Question Answering Systems. *Online Information Preview*, 35(4), 543–557. doi:10.1108/14684521111161927
- Olvera-Lobo, M. D., & Gutiérrez-Artacho, J. (2011b). Evaluation of Open- vs. Restricted-Domain Question Answering Systems in the Biomedical Field. *Journal of Information Science*, 37(2), 152–162. doi:10.1177/0165551511398575
- Olvera-Lobo, M. D., & Gutiérrez-Artacho, J. (2011c). Multilingual Question-Answering System in Biomedical Domain on the Web: An evaluation. *Lecture Notes in Computer Science*, 6941, 83-88.
- Olvera-Lobo, M.D. & Gutiérrez-Artacho, J. (2012). Language Resources for Translation in Multi-Lingual Question Answering Systems. *Translation Journal*, 16(2).
- Olvera-Lobo, M. D., & Gutiérrez-Artacho, J. (2013a). Evaluación del rendimiento de los sistemas de búsqueda de respuestas de dominio general. *Revista Española de Documentación Científica*, 36(2), e009. doi:10.3989/redc.2013.2.921

Olvera-Lobo, M. D., & Gutiérrez-Artacho, J. (2013b). *Searching Health Information in Question-Answering Systems*. In *Handbook of Research on ICTs for Healthcare and Social Services: Developments and Applications* (pp. 474–490). IGI Global.

Olvera-Lobo, M. D., & Gutiérrez-Artacho, J. (2014). Overview of Translation Techniques in Cross-Language Question Answering during the Last Decade. *Encyclopedia of Information Science and Technology*, 3, 4747–4755.

Olvera-Lobo, M. D., & Gutiérrez-Artacho, J. (2015). Question Answering Track Evaluation in TREC, CLEF and NTCIR. *New Contributions in Information Systems and Technologies. Advances in Intelligent Systems and Computing*, 353, 13–22. doi:10.1007/978-3-319-16486-1_2

Peñas, A., Magnini, B., Forner, P., Sutcliffe, R., Rodrigo, A., & Giampiccolo, D. (2012, June). Question answering at the cross-language evaluation forum 2003---2010. *Language Resources and Evaluation*, 46(2), 177–217. doi:10.1007/s10579-012-9177-0

Sing, Ardil, Wong, & Sahib. (2005). Response Quality Evaluation in Heterogeneous Question Answering System: A Black-box Approach. *Proceedings of World Academy of Science, Engineering and Technology*, 9.

Terol, R. M., Martínez-Barco, P., & Palomar, M. (2007). A knowledge based method for the medical question answering problem. *Computers in Biology and Medicine*, 37(10), 1511–1521. doi:10.1016/j.combiomed.2007.01.013 PMID:17374369

Yu, H., Lee, M., Kaufman, D., Ely, J., Osheroff, J. A., Hripcsak, G., & Cimino, J. (2007). Development, implementation, and a cognitive evaluation of a definitional question answering system for physicians. *Journal of Biomedical Informatics*, 4(3), 236–251. doi:10.1016/j.jbi.2007.03.002 PMID:17462961

Zweigenbaum, P. (2005). Question answering in biomedicine. *Proceedings Workshop on Natural Language Processing for Question Answering*, 1–4.

Zweigenbaum, P. (2009). Knowledge and reasoning for medical question-answering. *Proceedings of the 2009 Workshop on Knowledge and Reasoning for Answering Questions*, 1–2 doi:10.3115/1697288.1697289

ADDITIONAL READING

Abusalah, M., Tait, J., & Oakes, M. (2005). Literature review of cross language information retrieval. In *Proceedings of the World Academy of Science, Engineering and Technology* (Vol 4) (pp. 175–177)

Fluhr, C. (1996) Multilingual Information Retrieval. *Survey of the State of the Art in Human Language Technology*, at: cslu.cse.ogi.edu/HLT-survey/ch8node7.html

Gey, F. C., Kando, N., & Peters, C. (2005). Cross language information retrieval: The way ahead. *Information Processing & Management*, 41(3), 415–431. doi:10.1016/j.ipm.2004.06.006

Kelly, D., & Sugimoto, C. R. (2013). A systematic review of interactive information retrieval evaluation studies, 1967–2006. *Journal of the American Society for Information Science and Technology*, 64(4), 745–770. doi:10.1002/asi.22799

Moghadas, S. I., Ravana, S. D., & Raman, S. N. (2013). Low-cost evaluation techniques for information retrieval systems: A review. *Journal of Informetrics*, 7(2), 301–312. doi:10.1016/j.joi.2012.12.001

Oard, D. W., & Diekema, A. (1998). Cross language information retrieval. *Annual Review of Information Science & Technology*, 33, 223–256.

Olvera-Lobo, M. D. (2009). Cross-language Information Retrieval on the Web. In M. M. Cruz-cunha, E. F. Oliveira, A. J. V. Tavares, & L. G. Ferreira (Eds.), *Handbook Of Research On Social Dimensions Of Semantic Technologies And Web Services* (pp. 704–719). doi:10.4018/978-1-60566-650-1.ch034

Olvera-Lobo, M. D., & García-Santiago, L. (2010). Analysis of errors in the automatic translation of questions for translingual QA systems. *The Journal of Documentation*, 66(3), 434–455. doi:10.1108/00220411011038485

Olvera-Lobo, M.D. & Robinson-García, N. (2009). Tratamiento lingüístico de las preguntas en los sistemas de búsqueda de respuestas. *El profesional de la documentación*, 18 (2): 180–187.

Radev, D. R., Qi, H., Wu, H., & Fan, W. (2001). *Evaluating Web-based Question Answering Systems. Technical Report*. University of Michigan.

Salton, G. & Mc Gill, M.J. (1983). *Introduction to Modern Information Retrieval*. Nueva York: Mc Graw-Hill Computer Series

KEY TERMS AND DEFINITIONS

CLIR: CLIR (*cross-lingual information retrieval*) involves at least two languages in this process.

Cross-Language QAS: The Cross-Lingual Question Answering systems are a set of coordinated monolingual systems in which each extracts responses from a collection of separate monolingual documents.

Evaluation Measures: Many different measures for evaluating the performance of information retrieval systems have been proposed. The evaluation of QA systems has been carried out by the traditional evaluation measures based on the relevance (precision and MAP), and also, by other specific measures like MRR, TRR and FHS.

FHS: First Hit Success (FHS) assigns a value of 1 if the first answer offered is correct, and a value of 0 if it is not (thus it considers only the answer that appears in the first place on the list of results).

Information Retrieval: Fully automatic process that responds to a user query by examining a collection of documents and returning a sorted document list that should be relevant to the user requirements as expressed in the query.

MAP: Mean Average Precision: For systems that return a ranked sequence of documents, it is desirable also to consider the order in which the returned documents are presented. MAP measures the Average Precision of a set of queries for which the answers are arranged by relevance.

MRR: Mean Reciprocal Rank (MRR) assigns the inverse value of the position in which the correct answer is found (1 if the first, $\frac{1}{2}$ if the second, $\frac{1}{4}$ if the fourth, and so on), or zero if there is no correct response. This measure considers only the first correct response shown on the list of results offered by the system, and the final value is the average of the values found for each question. MRR assigns a high value to the responses that were in the highest classification positions.

Question Answering Systems: As an alternative to traditional IR systems they give correct and understandable answers to factual questions – rather than just offering a list of documents related to the search.

Translation: The process of translating words or text from one language into another.

TRR: Total Reciprocal Rank (TRR) is useful for evaluating the existence of several correct responses offered by a system to the same query. In these cases, it is not sufficient to consider only the first correct response in the evaluations, and thus TRR takes all of them into account and assigns a weight to each answer according to its position in the list of results retrieved. Thus, if the second and fourth answers on the list of results are correct for a question, the TRR value would be $\frac{1}{2} + \frac{1}{4} = \frac{3}{4}$.



Challenges in Collecting Qualitative Data for Information Systems Studies

Tiko Iyamu

Cape Peninsula University of Technology, South Africa

Irja Naambo Shaanika

Namibia University of Science and Technology, Namibia

INTRODUCTION

In research, problem is not necessarily something that is broken, but phenomenon which require further or an in-depth investigation for a fresh perspective. Thus, every research necessitates a problem statement, goal and objectives, which determines the data collection methods. The data type can be either quantitative or qualitative. According to Seidman (2012), depending on the objectives of the study, either the qualitative or quantitative research methods are selected for data collection. However, both methods can be selected, which is referred to as a mixed method (Barbour 2013; Silverman 2013).

The choice of research methods is critical in that they influences the way in which data is collected and analysed. According to Myers and Avison (2002,p70), qualitative research methods were developed in the social sciences to enable researchers to study social and cultural phenomena. The primary purpose of qualitative research is to understand a phenomenon as it is seen by respondents within a period and space. This is achieved by studying the respondents in their natural environments. Yin (2010) takes the argument further and states that the events and ideas emerging from qualitative research can represent the meanings given to real life events by the people who live them, not the values, perceptions or meanings held by researchers. However, the meaning which individuals and groups give or associate to events of Information systems

and technologies (IS/IT) has never been easy for researchers to understand.

BACKGROUND

IS/IT are used to support and enable organisations' operations including strategic intents. IS/IT does not operate in vacuum, but in socially constructed environments. According to Iyamu, Sekgweleo and Mkhmazi (2014), IS is not only made up of technology by its self, it also includes human and non-human actors, making it more complex than often seen from afar. The multifaceted nature of information systems does not make studies in the field easier. Also, there is a great diversity in the research methods and approaches that are employed in IS studies (Myers & Avison, 2002). However, it is believed that qualitative research methods are being used increasingly in evaluation of IS/IT studies (Kaplan & Maxwell, 2005). Qualitative methods are often employed to study the socio-technical aspects of IS, and to help researchers including postgraduate students to draw conclusions on why things happen in the way that they do (Iyamu, 2010).

Theoretically, many postgraduate students in the field of IS are knowledgeable about data collection methods, techniques and approaches. However, in practice, there are numerous challenges in how their knowledge is applied (Hen-nink, Hutter & Ajay, 2011). This has led to many students not able to complete their studies, or take

longer to do so. This is the primary motivation of this study. This chapter discusses hands-on experience, reveals pitfalls and challenges in collecting qualitative data, using semi-structured technique, towards achieving research objectives. The remainder of this article is divided into six main parts. The first and second covers literature review. The third discusses the processes that are involved in data collection. The fourth presents the major challenges that are encountered when the semi-structured method is employed in data collection. Future research is stated in the sixth part. Finally, a conclusion is drawn.

RESEARCH QUALITATIVE DATA

The qualitative research methods are considered to be most appropriate for studying the social world. Hence the type of data is critical, from the perspectives of actors' intentional and unintentional acts. Erickson (2012) emphasises that the essential purposes of qualitative research methods are to document in detail, the conduct of everyday events, and to identify the meanings according to those who participate and witness them. According to Kaplan and Maxwell (2005), qualitative methods are primarily inductive. In the beginning the focus is on a specific phenomena and based on that, the findings can be generalised. Thus, qualitative methods does not allow definition of variables before the research process begins (Carla, 2013), but rather researchers studies phenomena without predictions and assumptions of outcome, making data collection critical.

The qualitative data enables researchers to examine and understand the relationship and interactions between human and non-human factors, which do not necessarily depend on quantity. Stahl (2013) states that understanding of socio-technical phenomena aims to lead to better or better usage of systems. In the same context, Erickson (2012) explains that the qualitative data helps to find and examine: (1) detailed information about implementation, actors and processes; (2) identify the nuances of subjective understanding that motivate

various participants in a setting; and (3) identify and understand change over time.

Qualitative data is connected to research methods, such as case study, ethnography and action research. Myers and Avison (2002) discusses the case study, ethnography, action research and grounded theory as qualitative research methods that can be used in IS. The options broaden, and instill depth in IS studies in the use of qualitative research methods. According to Yin (2009), a case study is an empirical inquiry that investigates a contemporary phenomenon within its real-life context. Madnick, Wang, Lee and Zhu (2009) supports the use of case study method, stating that it enable an in-depth inquiry of a single instance or event that can lead to deeper understanding of why and how that event happened. The ethnography research method is similar to the case study approach. It allows the researcher to participate overtly or covertly in people's daily lives for an extended time. Ethnography, according to Ritchie, Lewis, Nicholls and Ormston (2013), involves understanding of the social world of particular objects or subjects typically via immersion in their communities. In IS studies, the ethnographer join and become part of the IS environments and observe actors' behaviours and their interactions with the systems, which they are part of. A different perspective from the case study and ethnographic, is the action research. The action research requires experiments in which participants collaborate with the researcher to identify and solve problem of social practice, in real life situations. Järvinen (2007) upholds that action research includes the researcher as an active participant rather than a passive observer.

QUALITATIVE RESEARCH METHODS: DATA SOURCE

The data collection method involves the gathering of data relating to the subject (research) from sources, such as interview (Yin, 2009). According to Babbie and Mouton (2001), data source includes data collected by the researcher, such as transcripts



of the interviews and existing written documents. The data used in this study was gathered from a postgraduate student's field work. The candidate was registered with the department of Informatics, at the Namibia University of Science and Technology. The candidate selected the qualitative, case study approach, using the semi-structured interview technique for the data collection. Nineteen participants were interviewed in the process. This included senior and junior employees, from both technology and business units of the organisations that were used as cases in the study. The interviews were recorded, using a tape recorder, and later transcribed to a total of 39,000 words. The analysis of the data was guided by the Activity Theory. From the study, an enterprise architecture framework for the Namibian government-wide was developed. The data collection process was critical to the study.

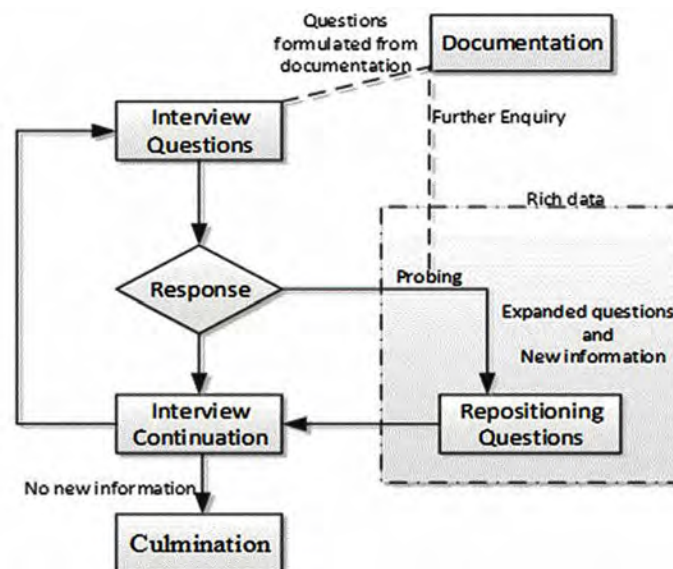
QUALITATIVE RESEARCH METHODS: DATA COLLECTION PROCESS

Data collection is an important aspect of research, as it helps to determine the evaluation, success or

failure of IS studies. The data serve as the empirical evidence that is needed to resolve the problem at hand as expressed in a problem statement of the study. In qualitative, the main emphasis is on data collection which is in words rather than in numerical values. This priority of emphasis does not mean that information about frequency of events is irrelevant to qualitative enquiry (Erickson, 2012). Qualitative researchers also calculate totals and record frequencies of events when and where required, depending on the objectives of the study.

Qualitative data collection techniques that are used in IS studies include observations, interviews, documentation and open ended questionnaires. The techniques can be used complimentary with each other. The suitability of the techniques depends on the objectives of the study. The process becomes even more complex, but might be more holistic when more than one technique is selected and used. Figure1 below illustrates data collection process, which involves two techniques, semi-structured interviews and documentation, from qualitative perspective. The processes as explained below should be read with the diagram to gain better understanding of the process and its complexity.

Figure 1. Data collection process



Interview Questions

In interpretive studies, there are different types of interviews, which include structured, semi-structured and unstructured interviews. As shown in Figure 1, the focus is on the semi-structured interview technique. In achieving the research objectives, interview questions are extracted from the research questions. The interview questions form the written guidelines that are used to elicit information from the participants. The guidelines help the researcher to maintain consistencies, uniformity and orders during data collection, which contribute to improving the quality of the data.

The interview questions can also be generated from the documentations relating to the study. This occurs when the researcher finds items in documentations that can be linked to the research objectives. Hence it is necessary to request access to documentations that are related to the study before interview is embarked upon. However, data obtained from documentations could also sometimes require further clarification. In such an instance, the researcher incorporates them in the interview questions for clarity purposes. This process enhances data quality and richness.

Documentation

Documentation is a technique used to obtain or access documents that are related to phenomena being studied. The documents can be in hard or soft (electronic) copies. Documentation as a technique is used in data collection, to provide valuable data that could otherwise had been difficult or impossible to obtain if only the interviews technique was employed. The process was considered to be good, and did improve the quality of data. The process is often used for comparing and evaluating complementarity and alignment between participants' responses and what has been documented in the context of the study.

As revealed from experience-based, documents obtained for the purpose of a study should be reviewed as soon as it is practical, and not after the interview process is complete. This is primarily

because the documents can help the researcher to identify critical areas that might require further exploration and probing. Thus the researcher incorporates into the questions for the forthcoming interviews, therefore has the opportunity to probe and sought clarity.

Response

Response is the subjective opinion and perspective which interviewees associate and provide to the subject in question. In the process of the interview, the researcher should take written notes in order to capture as much as possible from the interviewees' responses. Thus, the researcher is strongly advised to listen attentively even if the interview is recorded. By so doing, the researcher stays focus and is able to grasp as much details, therefore, able to identify areas that need further probing and clarification.

Also, to be able to probe the responses further, the researcher must be knowledgeable about the area of study. Lack of knowledge of the subject obviously limits the researcher's ability to explore and probe the interviewees' responses. Hence it is critical to have done much work in the areas of literature review and studying related documentations before data collection is undertaken.

Probing is very crucial in interpretive studies, using the semi-structured interview technique in that it allows the researcher to expand on interview questions. This leads to collection of new information and enhances the data richness. There are likely gaps in the data as a result of lack of probing, which often emanate during analysis. There is no limit to the number of questions during probing of response. As shown in Figure 1, response can be iterative as probing continues. To end the probing process, the researcher repositions his or her questions.

Repositioning Questions

The purpose of repositioning is to avoid diverting away from the original topics during probing. It allows the interviewer to reconnect back with the



interview questions (guidelines). Reconnection with the interview questions enables the researcher to maintain the flow of the questions, towards achieving his or her objectives. Reposition become even more necessary when both interviewer and interviewee are passionate and very knowledgeable about the discourse, otherwise, the researcher lost track of the direction of the responses.

The implication of not connecting back to the interview questions as quickly as possible could be costly, in that the researcher might end up collecting quantitative data for qualitative purposes. This happens when “true or false” are offered as responses during probing. This is as a result of inconsistencies, which manifest from lack of flow in the questions being asked. Also, this could lead to respondents getting lost or confused in the process, as disjoint arise in the manner in which the interview questions are being posed. This makes it difficult to continue with the interview either from the interviewee’s or interviewers, or both perspectives. In addition, it becomes difficult or impossible for the researcher to link the sets of data during analysis.

Interview Continuation

Continuation is the step-by-step process of data collection, using the interview technique. It prevent disjoint, and strengthens focus during data collection, when using the interview technique within the qualitative paradigm. It happens, whether probing occurred or not in the process. However, if interviewee’s response requires further probing, continuation is done after the repositioning process is complete. If no probing is required, the researcher proceeds to the next question as outlined in the interview guideline.

Using the semi-structured technique, the interview process can be iterative. However, the process has to stop at some point. The process ceases when point of saturation is achieved. Saturation occurs when there are no more new and unique information is forthcoming, but rather repetition of what has been said or opined by other respondents.

To notice point of saturation, the researcher must be familiar with the data he or she has collected in the process. This can be achieved through transcribing of the recorded interviews immediately after each interview is conducted. As revealed from our experience, accumulating the interviews materials is a bad practise, and it limits the researcher’s chances of gathering rich data. This is primarily because the researcher will not be familiarised with the data that he or she has already collected, in order to identify where (if any) gaps exist.

Culmination

Culmination is referred to as, when data collection processes come to an end at some point during the study. This happens whether point of saturation is achieved or not. The end point could be initiated either by the researcher or the interviewee – when the researcher has exhausted the questions on the guideline, or the interviewee’s response leads to closure.

Culmination is considered to be the end of discussion between the researcher and the participants on matters relating to the phenomena being studied. Thereafter, the researcher performs data cleansing in preparation for analysis. This is done by correcting proof reading and editing, as well as appropriately labelling of the texts from the interviews. Without reaching the point of culmination, it is difficult to avoid disjoint and inconsistency in the data, which of course determines the findings.

MAJOR CHALLENGES IN AN INTERPRETIVE DATA COLLECTION PROCESS

The case study and semi-structured from the qualitative methods are considered to be more efficient for collection of rich data in IS studies. However, the data collection processes are not always as easy as we are meant to believe. Researchers often encounter challenges in collecting qualita-

tive data. This is caused by two main factors: (i) how well the researcher is knowledgeable about the technique he or she has employed for the data collection; and (ii) how familiar is the researcher with the environment within which data is being collected. However, the challenges vary from one researcher to the other. This section presents and discusses practical experience during a researcher's data collection in an interpretive study, using the semi-structured interview technique.

The most critical challenges that are often repeatedly encountered by researchers, particularly aspiring researchers and postgraduate candidates are discussed in this section. The challenges which include unconscious biasness, data size, knowledge of the subject and transcribing process determines and shapes the outcome of data that is collected in an interpretive study.

Expectation and Unconscious Biasness

In the qualitative paradigm, the interpretive approach attempt to understand a phenomenon that is being studied by exploring the meanings people assign to them, and within context (Lee, 1994). It does not predefine the phenomenon. However, researchers continue to misconstrue the guidelines of the interpretive approach as empirically gathered in the study that we did use as a case. The researcher explained:

When I went out to gather data I had expectations of the types of answers that I wanted to gather. I wanted to hear answers that sound good to my hearing, such that are well articulated and explained. These I had considered sensible according to my subjective reasoning at the time.

The interpretive research approach does not predefine dependent and independent variables, but focuses on human sense making within context (Myers, 1997). Even though it is difficult to completely separate the researcher from the phe-

nomenon, there shouldn't be specific expectation from the participants. However, some researchers do, as explained in this case:

I went to gather data with a mind-set of certain expectation from the respondents. So, as respondents shared and explained their views and perspectives which were completely different from what I had expected, I immediately concluded that they were not serious. Most of the times, I walked out of respondents' offices being angry, and felt that my time was wasted.

Following Klein and Myers (1999), it is against the principle of interpretive study to collect data with a pre-set mind. It poses a danger of reaching a conclusion before the evidences are gathered, and therefore defeating the purpose of the study. This triggers fundamental questions, such as, on what basis were judgements carried out on the respondents' views and opinions. According to the researcher in this case, most of the data that was thought to be useless became the corner stones, focal points during the analysis.

Data Size

The usefulness and relevance of qualitative data is not determined by the size or volume, but by its richness. According to the researcher, some interviews were too short, in that they lasted less than forty minutes. This was considered to be less useful at the time, even before the analysis was conducted. However, researcher does contribute to the size of the data that is collected as explained below:

In some interviews, when I did not get what I had expected, I immediately drew a conclusion that the respondent was not able to answer the question. As a result, he or she would not be able to answer the next question as well. I therefore lost interest and not continue. This contributed to making some of the interviews period to be short.

If the goal of the research problem is not to test the validity of a model where all the variables which influence a phenomenon or process is already known, then there is no need for expectation of answers from respondents. Also, the interpretive approach does not set out to test hypotheses; it is aimed at producing an understanding within context, whereby the phenomenon being studied influences and is influenced by the social context. Such culminative decision helps determine the data size, and could be attributed to the researcher's knowledge of the subject.

Knowledge of the Subject

Consultation on wide range of literature helps the researcher to gain deeper knowledge of the field of study. However, it has little or no influence on the data collection. According to Rowlands (2005), limitations in literature and the nature of the problem influence authors. Thus enabling the researcher to build on his or her knowledge, this is necessary in order to be able to engage in a robust conversation with the respondents during data collection. Otherwise, the researcher will be short of knowing what, when and how to probe the responses. This was evident in the study that was used in this case, as explained by the researcher:

Some of the interviewees were too short, due to the lack of probing of respondents' responses. The respondents would share their views, but in some cases I failed to follow up on them. I didn't think fast enough to see that an area needs further probing.

Some researchers learn as they progress in their studies. Others had vast knowledge about the phenomenon before they embarked on the data collection stage. On one hand, limited knowledge does not allow the researcher to take advantage of the flexibility which the interpretive approach presents, using semi-structured technique. On the other hand, vast knowledge could make the researcher lead the participants to specific direc-

tion or influence their views. The balance is a challenge, although it has been trivialised for long among researchers and postgraduate students.

Data Transcribing

The text and voice data are recorded with various types of infrastructures, such as cellular phones and tapes. Transcribing the recorded data can be rigorous, but critical to the richness. As such, it should be carried out with a high level of diligent and appropriateness. This includes the use of reliable infrastructures, for accurate audio. The researcher narrated her experience and challenges as follows:

The tape recorder was at some point not working properly, I only discovered that after the interview was completed. As a result, some of the valuable information was lost. It was difficult to reschedule an appointment due to the respondent's busy schedule. Also, the respondent had been disappointed if asked to repeat himself.

To organise and clean the data for analysis wholly depend on how well the researcher is familiar with the data. Also, data can get distorted during transcribing and documentation. Hence the role of the research in the management of the data is critical. One of the factors in the management of data is briefly explained as follows:

I did not complete transcribing some interviews while they were still fresh in my memory, and they piled up. When I had to transcribe them, I did not hear some of the words and sentences, and it became difficult to remember what respondent did say.

The outcome of one interview is a guide to conducting the next, in that it is used for corrective measure and improvement. It is therefore necessary for the researcher who transcribes the data to listen attentively to the audio. Also, to try to make sense of the data as the process continues.

This is to avoid discarding of valuable data. Thus the researcher explained, based on her experience:

When transcribing, listen attentively because the outcome of every interview is a guide to the next interview. As you transcribe the data, you identify gaps, due to lack of probing. So your next interview should be seen as an opportunity to fill the gaps.

Two critical factors, analysis and differentiation of data should be avoided during transcribing of the interviews' data. It is advisable to avoid analysis of the data until the transcribing is complete. This is mainly because sense making could be made out of "half of the data" only. Also, the researcher is not in a position to eliminate or consider data to be useful or useless until the transcribing is complete, and the data has been cleaned. Otherwise, the probability that the researcher would be biased is high. This implies that the authenticity of the outcome of the study could be affected.

FUTURE RESEARCH DIRECTIONS

For future research, the challenges revealed in this study can be explored to understand how they manifest themselves during data analysis and interpretation.

CONCLUSION

The qualitative methods enable the studies of new areas and theories on why and how the change takes place. Qualitative data helps researchers to interrogate and understand human behaviours as they interact with one another, as well as with objects in the development, implementation and use of IS/IT. Thus, it is critical to be unbiased in the collection of data. The quality and credibility of research data gives practitioners more confidence in considering the outcome in their practices and purposes.

Also, this chapter is intended to help researchers and postgraduate, to avoid pitfalls in their studies. It unveils some of the challenges which are commonly encountered by researchers and postgraduate in IS studies. Although some of the findings from the study might be common, they cannot be taken for granted. The article presents hands-on experience and uncovers critical pragmatic issues from the problem under study. The empirical and hands-on nature of the study is its main contribution, which makes it useful, especially for aspiring IS researchers.

REFERENCES

- Babbie, E., & Mouton, J. (2001). *The practice of social research*. Cape Town: Oxford University Press.
- Barbour, R. (2013). *Introducing qualitative research: A student's guide*. Sage (Atlanta, Ga.).
- Carl, W. (2013). *Introducing qualitative research in psychology*. McGraw-Hill International.
- Erickson, F. (2012). Qualitative Research Methods for Science Education. In B. Fraser, K. Tobin, & C. McRobbie (Eds.), *Second international handbook of science education* (pp. 451–1469). Springer, Netherlands. doi:10.1007/978-1-4020-9041-7_93
- Hennink, M., Hutter, I., & Ajay, B. (2011). *Qualitative Research Methods*. London: Sage Publications Ltd.
- Iyamu, T. (2013). Underpinning Theories: Order-of-Use in Information Systems Research. *Journal of Systems and Information Technology*, 15(3), 1–13. doi:10.1108/JSIT-11-2012-0064
- Iyamu, T., Sekgweleo, T., & Mkhomazi, S. (2013). Actor Network Theory in Interpretative Research Approach. In Y. Dwivedi, H. Henriksen, D. Wastell, & R. De (Eds.), *Grand Successes and Failures in IT* (pp. 605–610). Springer Press. doi:10.1007/978-3-642-38862-0_41

- Järvinen, P. (2007). Action research is similar to design science. *Quality & Quantity*, 41(1), 37–54. doi:10.1007/s11135-005-5427-1
- Kaplan, B., & Maxwell, J. A. (2005). Qualitative research methods for evaluating computer information systems. In *Evaluating the organizational impact of healthcare information systems* (pp. 30–55). New York: Springer. doi:10.1007/0-387-30329-4_2
- Klein, H., & Myers, M. (1999). A set of principles for conducting and evaluating interpretive field studies in information systems. *Management Information Systems Quarterly*, 23(1), 67–94. doi:10.2307/249410
- Madnick, S., Wang, R., Lee, Y., & Zhu, H. (2009). Overview and framework for data and information quality research. *ACM Journal of Data and Information Quality*, 1(1), 2–22.
- Myers, M. (1997). Interpretive research in information systems. In J. Mingers & F. Stowell (Eds.), *Information Systems: An Emerging Discipline* (pp. 239–26). London: McGraw-Hill.
- Myers, M., & Avison, D. (2002). An introduction to qualitative research in information systems: A reader. *Sage (Atlanta, Ga.)*.
- Ritchie, J., Lewis, J., Nicholls, C., & Ormston, R. (Eds.). (2013). *Qualitative research practice: A guide for social science students and researchers*. Sage.
- Rowlands, B. (2005). Grounded in practice: Using interpretive research to build theory. *The Electronic Journal of Business Research Methodology*, 3(1), 81–92.
- Seidman, I. (2012). *Interviewing as qualitative research: A guide for researchers in education and the social sciences* (4th ed.). New York: Teachers College Press.
- Silverman, D. (2013). *Doing qualitative research: A practical handbook* (4th ed.). London: Sage.
- Stahl, B. (2013). Interpretive accounts and fairy tales: A critical polemic against the empiricist bias in interpretive IS research. *European Journal of Information Systems*, 23(1), 1–11. doi:10.1057/ejis.2012.58
- Yin, R. (2009). *Case inquiry research, design and methods* (4th ed.). London: Sage Publications.
- Yin, R. (2010). *Qualitative research from start to finish*. Guilford Press.
- Yunos, Z., & Ahmad, R. (2014). The application of qualitative method in developing a cyberterrorism framework. *Proceedings of 2014 International Conference on Economics, Management and Development*, 133–137.

ADDITIONAL READING

- Barbour, R. (2013). *Introducing qualitative research: A student's guide*. Sage (Atlanta, Ga.).
- Bazeley, P. (2013). *Qualitative data analysis: Practical strategies*. Sage (Atlanta, Ga.).
- Doody, O., & Noonan, M. (2013). Preparing and conducting interviews to collect data. *Nurse Researcher*, 20(5), 28–32. doi:10.7748/nr2013.05.20.5.28.e327 PMID:23687846
- Guest, G., Namey, E. E., & Mitchell, M. L. (2012). *Collecting qualitative data: A field manual for applied research*. Sage (Atlanta, Ga.).
- Morgan, D. L., Ataie, J., Carder, P., & Hoffman, K. (2013). Introducing dyadic interviews as a method for collecting qualitative data. *Qualitative Health Research*, 23(9), 1276–1284. doi:10.1177/1049732313501889 PMID:23925406
- Smith, J., & Firth, J. (2011). Qualitative data analysis: The framework approach. *Nurse Researcher*, 18(2), 52–62. doi:10.7748/nr2011.01.18.2.52.c8284 PMID:21319484

KEY TERMS AND DEFINITIONS

Continuation: Sequential order in an interview process.

Culmination: A sudden dead-end.

Probing: Examination and further question based on response to previous question.

Repositioning: Readjustment to original structure of a guideline.

Saturation: A point where nothing new is forthcoming.



Cognitive and Psychological Factors in Cross-Language Information Retrieval

Rowena Li

Bayside High School Library, USA

INTRODUCTION

With the rapid growth of computer and communication technology, a global interconnected information infrastructure is quickly constructing through the Internet. As the information is able to travel beyond geographical and spatial borders via the Internet, and more and more people around the world have gained access to global networks, the language boundaries have to be crossed in order to make the global communication possible. Statistics shows that 74% of Internet users are non-English speakers (MiniWats Marketing Group, 2015). However, English is still the leading language in global communication environment. English language dominates 54.4% of the websites (W3Techs.com, 2015). As the Internet has become one of the major communication mechanisms for information storage, retrieval, and dissemination, users need the ability to locate and retrieve information wherever, whenever and in whatever the language it has been stored. However, most of the search engines currently available can only provide monolingual information retrieval, which means that the retrieval can only be conducted in the same language as the query language. Cross-language information retrieval (CLIR) has become increasingly important to facilitate the effectiveness of information exchange among different languages. As a result, the study and development of tools and technology of cross-language information retrieval have gained greater attention over the past decade. While a lot of research has focused on the effectiveness of system functionality, few studies have examined information needs and so-

cial aspects related to cross-language information retrieval. This article aims to speculate the human and social aspects of cross-language information retrieval. It explores CLIR users' unique social and cultural contexts, their psychological and cognitive structures, and their distinctive relevance judgment. It examines in depth the barriers embedded in cultural, linguistic, and cognitive dimensions, which might hinder further advancement in cross-language information retrieval.

BACKGROUND

Information Seeking

Information has been traditionally viewed as a message transmitted from sender to receiver through a channel which may reduce or increase its ambiguity (Shannon & Weaver, 1949). In this view, information is external and objective, as well as structured and measurable. Dervin (1976, 1980, & 1983) rejected this traditional approach and proposed her sense-making theory, which concerns the behavioral, cognitive and social aspects of information seeking. She introduced information seeking as a concept that an individual has to constantly "make sense" of his situation to move physically and cognitively across a gap in front of him through time and space. This model sees information seeking as systematic, subjective, situational, individual, cognitive, and holistic (Dervin & Nilan, 1986). In her own way, Dervin recognizes the similarity in information seeking process, as well as understands the uniqueness

of individuals in information seeking situation. Saracevic, Kantor, Chamis, and Trivison (1988) associated information seeking process with users' *perception of the problem, intent for use of the information, internal knowledge state, and public knowledge expectations*. Wilson's model (1981) states that an individual's situation and social role, his/her psychological and cognitive states influence his/her context of information need. Other cognitive approaches to information seeking were also presented in the past two decades. Bates' "berry-picking" model (1989) states that information seeking is not linear, but is constantly modified by the feedback occurred in information seeking process. Belkin's ASK (Anomalous State of Knowledge) model states that information seeking arises when an individual cannot identify the gap in his knowledge state (Belkin, Oddy, & Brooks, 1982). Taylor's (1991) situational theory believes that information seeking is not only "based on subject matter, but on other elements of the context within which a user lives and works" (p. 218). Kuhlthau's (1991) Information Search Process approach believes that individuals seek meaning rather than answers. Affective states are influenced by uniqueness of information situation and mood states. And personal involvement influences the information seeking process. Ingwersen's (1992, 1996) cognitive model states that cognition occurs in all stages of information seeking, and individuals experience a variety of cognitive modeling. Furthermore, Ellis's information seeking theory (1989), Schamber's relevance (Schamber, Eisenberg, & Nilan, 1990), and Wilson's situational relevance (1973) have also influenced information seeking research.

In general, for the past two decades, the study of information has experienced a paradigm shift. It concentrates more on subjectivity (instead of objectivity) of information, on holistic (instead of atomistic) view of experience, and on internal cognition (instead of external behavior). And at the same time, users have become a part of information process and a significant factor in the study of information seeking and retrieval.

Information Retrieval and Cross-Language Information Retrieval



Information retrieval (IR) is "a field concerned with the structure, analysis, organization, storage, searching, and retrieval of information" (Salton, 1968, p. v). It is the process of query formulation, matching, selection, evaluation and representation. The result is presented in the form of a ranked list in a descending order according to their relevance to the query. Information retrieval process is evaluated by the effectiveness and relevance of this returned list to the search query.

Cross-language information retrieval (CLIR) is a special case of information retrieval. It focuses on retrieving documents in languages other than query language. This process involves not only information retrieval, but also query translation to locate relevant documents. CLIR is defined as "Given a query in any medium and any language, select relevant items from a multilingual multimedia collection which can be in any medium and any language, and present them in the style or order most likely useful to the user, with identical or near-identical objects in different media or languages appropriately identified" (Fluhr et al., 1999).

CLIR Models and Approaches

The process of CLIR involves query formation, document pre-processing, index matching, document selection, document examination, document delivery, and relevance feedback (Oard & Diekema, 1998).

Of all the stages of information retrieval process, matching is the most important and problematic one. To better match a query to the indexed terms of pre-processed documents, four matching strategies have been identified. They are *cognate matching*, *query translation*, *document translation*, and *interlingual techniques* (Oard & Diekema, 1998).

Oard and Diekema (1998) indicated that *cognate matching* is the process of matching words

from one language to another according to the words' spelling or pronunciation. This approach is used to find matches when the query words are not in the dictionary, but only under a condition - the two languages are in the same language group. *Query translation* is the most used strategy. The query is translated into target languages. Afterwards, the translation is used to match the documents for relevant retrieval. The advantage of query translation is its efficiency. However, since the query contains only a few words, the translation might lead to lexical ambiguity due to lack of context. *Document translation* is to automatically translate documents into query language. It involves online real-time machine translation system for users to navigate the Internet. However, it requires large system functionality. Since it is costly and time-consuming, it is only limited to smaller collections. *Interlingual techniques* convert both query and index terms from the documents into a unified language-independent representation. Multilingual thesauri are used in this concept. Since the controlled-vocabulary representation generates only one term for each concept, query and document representations can map their words respectively onto their unique terms in the thesauri. This approach reduces translation ambiguity. However, it requires additional storage space for the translated documents.

CLIR Users and Their Information Needs

Users have various information needs. As users possess different language skills and information seeking tasks, their information needs and information seeking behavior are different. CLIR users basically fall in two categories:

1. Monolinguals search documents/images in target language(s) that they do not know by entering a query in a language they know;
2. Multilinguals search documents/images in target language(s) they know by entering a query in one language.

Large organizations may also be users of CLIR. Commercial online services such as Dialog and LexisNexis have developed retrospective market searches and alert services (Oard, 1997). CLIR may help expand the market for these services. Individual government and international institutions also have the demand to search a large amount of multilingual documents. Journalists may look for news and stories in foreign language newspapers, and business analysts might gather foreign business information and provide service to different countries. Translators may use CLIR services to serve clients in other countries, and information specialists may search as intermediaries for patrons (Petrelli, Beaulieu, Sanderson, & Hansen, 2002). Furthermore, scholars may search documents for current development in their research fields.

CLIR Information Seeking Behavior

There has been a paradigm shift from a system-centered approach toward a user-centered approach in the study of human information seeking behavior since the 1980s. Onscreen reading time and printing behavior have been used as a useful source of implicit feedback to indicate the relevance of the retrieval output (Kim, Oard, & Romanik, 2000). Personality and situation (Petrelli et al., 2002), cognition state (Foster, 2004), the language used (Al-Wreikat, Rafferty, & Foster, 2015) also influence users' information seeking behavior. In addition, knowledge of target language affects users' general information seeking behavior. Users make faster relevant judgments based on the retrieval output if they know target language well (Józsa, Köles, Komlódi, Hercegfli, & Chu, 2012).

HUMAN AND SOCIAL ASPECTS OF INFORMATION SEEKING

Social and Cultural Contexts in CLIR

In the process of CLIR, machine translation is used to translate either the query or the full text

documents. However, translation errors impact CLIR performance and these errors occur because of two issues: the lack of equivalent lexical terms in target language, and conceptual domains of the query are different from those of the retrieved documents in target language.

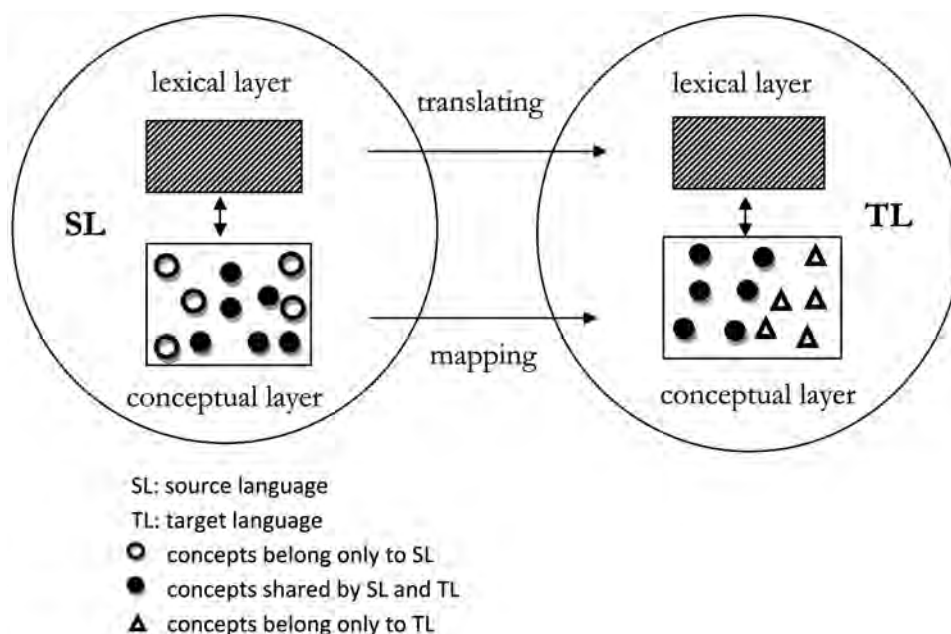
The process of translation involves not only the technical and structural transferring information, but also cultural transferring from one language to another (Gransow, 2001). Some shared perceptions and concepts are universal between languages. Some concepts exist only in a specific context because of cultural differences. As a result, the receiver might not be able to perceive the world presented by the sender through mere linguistic translation. Especially with non-translatable indigenous concepts, the words-to-words translation is simply not enough (Riggs, 1987).

Language reflects culture. Through translation, the culture of source language should be reflected in target language via a linguistic filter. Therefore, translation is not merely a word-for-word process; rather, it filters through linguistic, stylistic and cultural phenomenon transparently. Translation is also tied to other aspects of hu-

man knowledge, such as psychology, sociology, anthropology, history, linguistics and geography (Castro-Paniagua, 2000). An exclusion of all these factors in translation will result in transferring only a partial picture of the given messages across languages. When the cultures between the two languages are totally different, the culture embedded in source language should be transferred and explained into target language and be adopted into target language. Otherwise, the translation only delivers the linguistic meaning of the words, not the culture embedded.

Essentially, a word has two representative levels, a lexical level and a conceptual level. When a word is emic, which exists within a culture and determined by local custom and belief, it is hard to map this word at both its lexical and conceptual levels. This creates a lexical hole in machine translation. Even when a word is etic, which is not culturally specific and can be applied to any cultural practices, the conceptual level of a word might not be easily mapped to that of a corresponding word in another language, because some cultural elements may not be the same. Figure 1 demonstrates a translation process.

Figure 1. Mapping model for translating a word from one language to another



This model shows even though a source word can be mapped to its target language on a lexical level, some of the cultural elements might be lost in translation process, because those cultural elements at the conceptual level are local and specific to that culture. As a result, users might be confused with the mismatch between the language and the context. For example, 江南一叶 is a well-known Chinese phrase, which cannot be found in any dictionaries. Figure 2 indicates two translation processes taken from lexical and conceptual levels with and without the knowledge of Chinese culture. The images from the two translations are totally different due to implementing or missing cultural elements on the conceptual level.

Unfortunately, in the study of CLIR, great efforts have been made to solve the problems of words mismatch and ambiguity. Little research has focus on the consideration of mapping cultural-situational concepts from source language to target language in translation process.

Psychological and Cognitive Structure in CLIR

It is clear that translation errors are not just the result of linguistic or lexical mismatches, they also relate to translator's cognitive state, which links the linguistic choices to culturally determined conceptual preferences. Generally, four cognitive variables affect users and context of information seeking: *problem*, *internal knowledge state*, *intent*, and *public knowledge estimate* (Saracevic et al., 1988). The initial information seeking process starts with the emergence of a cognitive dissonance when the user realizes there is a gap between his internal knowledge state and reality (Saracevic et al., 1988; Wilson, 1997). With the intention of filling this gap, he makes his public knowledge estimate prior to information seeking process. During his information seeking, users bring two types of "knowledge structures" - episodic memory (personal information) and semantic memory

Figure 2. Translation process models with/without the knowledge of culture

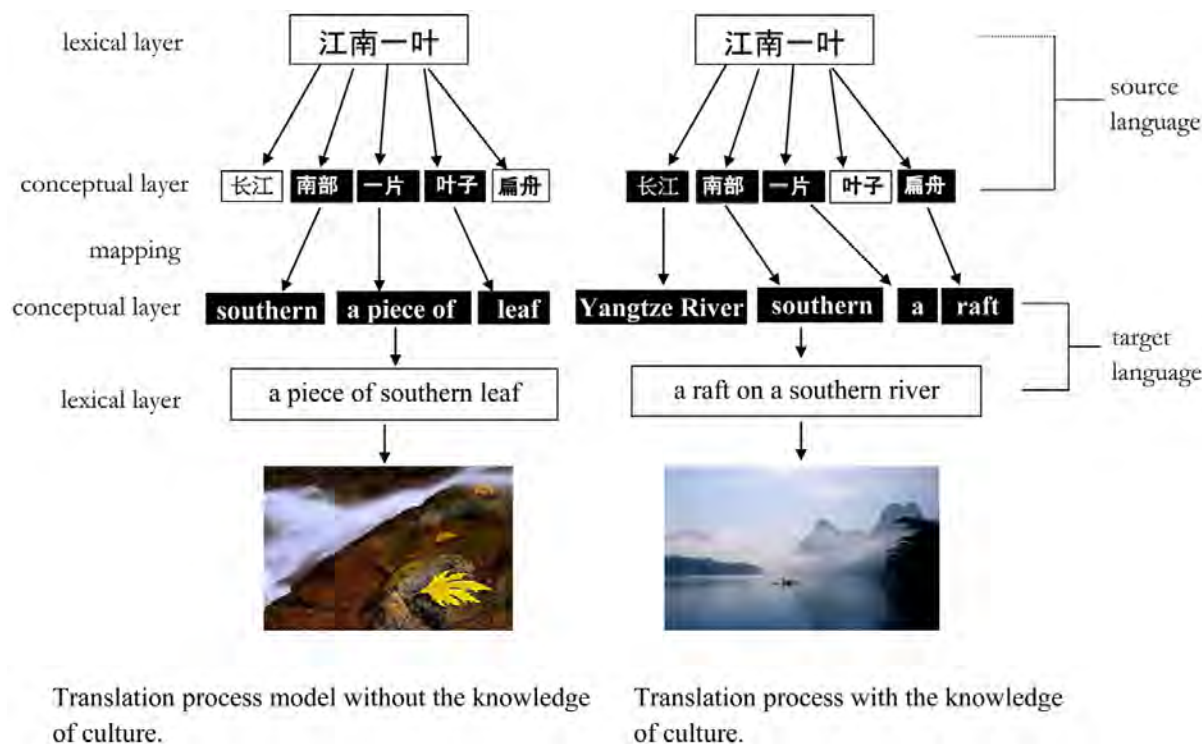
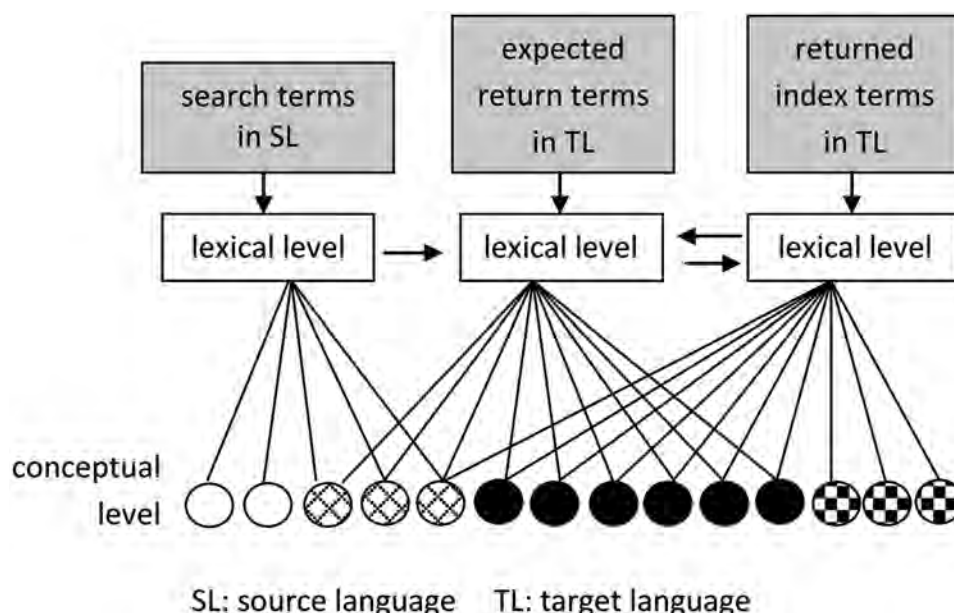


Figure 4. Cognitive state of a multilingual in a CLIR process



language in their mind and compare them to the original search terms.

Figure 4 shows that each language is stored separately at the lexical level. However, at the conceptual level, words in each language access a shared semantic representation. The features in conceptual memory activated by the word in source language in lexical memory may not necessarily be activated by its target language translation (de Groot, 2002; Luna & Peracchio, 2002). As a result, although the word is translated lexically, the meanings of the two words might be different semantically, because the contexts in which the words are being used are different between cultures and languages. This indicates that culture and cognitive states of the information seeker influence the process of information seeking.

Subjective Relevance in CLIR

Relevance has been a major criterion for evaluating the effectiveness of information retrieval. It is a measure in degrees of how close the relationship is between a query and the contents of the retrieved documents. Traditionally, relevance has

been judged through laboratory models, which consist of three parts: a document database, a set of topical requests, and a machine-run binary assessment algorithm. This model has been challenged recently by an emphasis on user-centered information design and human interaction in information retrieval. As a result, researchers start to focus their studies on subjective relevance in addition to objective relevance (Ingwersen & Järvelin, 2005). Subjective relevance refers to the usefulness of documents in relation to the goals in a situation perceived by the user himself subjectively (Borlund & Ingwersen, 1998). This has changed the course of relevance study - relevance is measured not only by the algorithm produced from binary computer programs, users' personal information needs, language skills, knowledge, perceptions, interaction, and intentions influence relevance measure as well. In a sense, relevance expanded to include its subjective (instead of objective), dynamic (instead of static) and situational (instead of algorithmic) dimensions (Schamber, 1994). The once fixed measure of system-based relevance becomes dynamic when applied to diverse users and various situations.

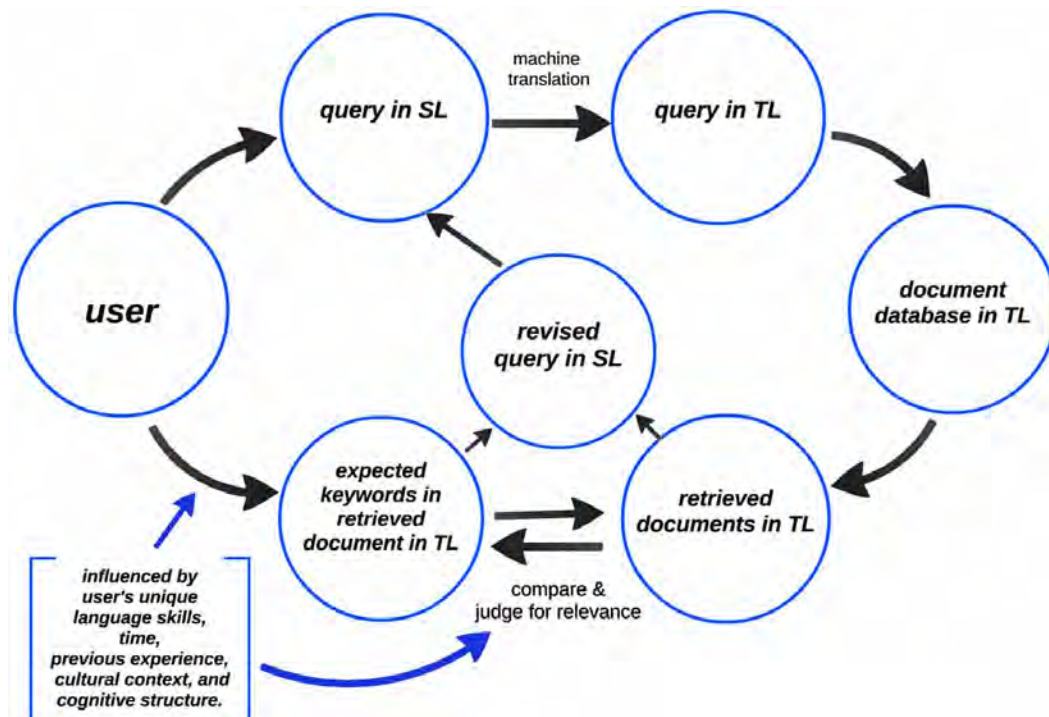
Since CLIR is a special case of information retrieval, user's language skills and cognitive structures should be considered in its relevance measure. In the process of CLIR, a multilingual types a query in source language and expects to retrieve relevant documents in target language. As he sends in the query, he has some index words envisioned in his mind in target language. This envisioned index words are conditioned by his language skills and his cultural knowledge in the conceptual level. When the documents are retrieved in target language, he compares and evaluates the relevance of the retrieved documents based on his expected index words pre-formed in his mind, and this process also depends on his language skills and his cultural backgrounds. Figure 5 demonstrates subjective relevance process in CLIR.

Users are the final judges of relevance measure, since information retrieval systems are built for users instead of lab-run system assessment. Users'

language skills and their knowledge of culture have a great impact on their information seeking experience and information seeking behavior. It also influences their relevance judgments.

1. The more the user knows the subject matter, the easier the judgment is made on the relevance;
2. The more the user knows target language and its cultural contexts related to the subject matter, the fewer documents of a retrieval set he tends to judge relevant;
3. The more the user knows target language and its cultural context related to the subject matter, the less time he spends on the retrieval process;
4. The more the user knows target language and its cultural context related to the subject matter, the more frequent he revises the query and starts new rounds of information retrieval.

Figure 5. Cross-language information retrieval model with the influence of the knowledge of culture on subjective relevance



FUTURE RESEARCH DIRECTIONS

After more than a decade of research on cross-language information retrieval, information retrieval systems have been greatly improved to accurately translate queries and to effectively retrieve relevant information in target language. However, recall and precision still remain as the top two challenges. When we look into the problems of ambiguity and mismatch, the dimensions of cultural context, psychological and cognitive structure, and situational relevance are not to be overlooked. Further study of these three dimensions will open another window in designing a user-centered system and in reinforcing the concept of user-system interactions.

The four models presented in this article are not generated from surveys, observances, or laboratory research. They have not been substantiated through data collection and analysis. Further studies need to concentrate on verifying these models statistically. Further investigations may also focus on defining the associations of multilinguals' fluency level in target language and their relevance judgment. Observations and lab samplings might be used to collect data and to establish a foundation for further study.

CONCLUSION

Cross-language information retrieval system users are unique in their information seeking process. Their psychological and cognitive states are different from those of monolingual information system users. Their language skills, social and cultural backgrounds, situational and personal perceptions become important variables in measuring relevance in information retrieval. To establish a user-centered interactive cross-language information retrieval system, an early feedback system should be considered. As indicated, monolinguals compare their query to the returned index terms in source language and make relevance judgment (Figure 3), and multilinguals compare their expected index terms in target language to the

returned index terms (Figure 4), an interactive system should be set up to present multiple query translations right after the query is entered. As a result, before machine matches query to index terms in database documents, users are able to choose the best lexical translation equivalents that are culturally appropriate in particular situational contexts. In this way, the number of retrieved irrelevant documents shall be reduced and its recall and precision will be improved.

In general, it is imperative to focus on the human and social aspects in the examination of cross-language information retrieval process. The influence of culture and users' target language skills on the process of CLIR should be further investigated. Although much attention has been concentrated on solving the problems of words mismatch and ambiguity in the course of translation, users' cognitive and psychological development and progress involved in CLIR process should never be overlooked.

REFERENCES

- Al-Wreikat, A., Rafferty, P., & Foster, A. (2015). Cross-language information seeking behaviour English Vs Arabic. *Library Review*, 64(6/7), 446–467. doi:10.1108/LR-04-2015-0044
- Bates, M. (1989). The design of browsing and berrypicking techniques for the online search interface. *Online Information Review*, 13(5), 407–424. doi:10.1108/eb024320
- Belkin, N. J., Oddy, R. N., & Brooks, H. M. (1982). ASK for information retrieval: Part I. Background and theory. *The Journal of Documentation*, 38(2), 61–71. doi:10.1108/eb026722
- Borlund, P., & Ingwersen, P. (1998). Measures of relative relevance and ranked half-life: performance indicators for interactive IR. In *Proceedings of the 21st annual international ACM SIGIR conference on research and development in information retrieval* (pp. 324–331). New York, NY: ACM. doi:10.1145/290941.291019

Castro-Paniagua, F. (2000). *English-Spanish translation, through a cross-cultural interpretation approach*. New York: University Press of America.

de Groot, A. M. (1992). Determinants of word translation. *Journal of Experimental Psychology*, 18(5), 1001–1018.

de Groot, A. M. (2002). Lexical representation and lexical processing in the L2 user. In V. Cook (Ed.), *Portraits of the L2 User* (pp. 32–63). Buffalo, NY: Multilingual Matters LTD.

Dervin, B. (1976). Strategies for dealing with human information needs. *Journal of Broadcasting*, 20(3), 323–333. doi:10.1080/08838157609386402

Dervin, B. (1980). Communication gaps and inequities: Moving toward a reconceptualization. In B. Dervin & M. Voight (Eds.), *Progress in Communication Sciences* (Vol. 2, pp. 73–112). Norwood, NJ: Ablex.

Dervin, B. (1983). An overview of sense-making research: Concepts, methods and results. Paper presented at the annual meeting of the. Dallas, TX: International Communication Association. Retrieved from <http://communication.sbs.ohio-state.edu/sense-making/art/artdervin83.html>

Dervin, B., & Nilan, M. (1986). Information needs and uses. In M. Williams (Ed.), *Annual Review of Information Science and Technology* (Vol. 21, pp. 3–33). White Plains, NY: Knowledge Industry Publications.

Dufour, R., & Kroll, J. E. (1995). Marching words to concepts in two languages: A test of the concept mediation model of bilingual representation. *Memory & Cognition*, 23(2), 166–180. doi:10.3758/BF03197219

Ellis, D. (1989). A behavioral approach to information retrieval system design. *The Journal of Documentation*, 45(3), 171–212. doi:10.1108/eb026843

Fluhr, C., Frederking, R. E., Oard, D., Okumura, A., Ishikawa, K., & Satoh, K. (1999). Multilingual (or Cross-lingual) Information Retrieval. *Proceedings of the Multilingual Information Management: Current Levels and Future Abilities*.

Foster, A. (2004). A nonlinear model of information-seeking behavior. *Journal of the American Society for Information Science and Technology*, 55(3), 228–237. doi:10.1002/asi.10359

Gransow, B. (2001). Nontranslatable: Indigenous concepts in social science research on China. *Asian Journal of Social Science*, 29(2), 262–284. doi:10.1163/156853101X00073

Ingwersen, P. (1982). Search procedures in the library—analyzed from the cognitive point of view. *The Journal of Documentation*, 38(3), 165–191. doi:10.1108/eb026727

Ingwersen, P. (1992). *Information Retrieval Interaction*. Los Angeles, CA: Taylor Graham.

Ingwersen, P. (1996). Cognitive perspectives of information retrieval interaction: Elements of a cognitive IR theory. *The Journal of Documentation*, 52(1), 3–50. doi:10.1108/eb026960

Ingwersen, P., & Järvelin, K. (2005). *The turn: Integration of information seeking and retrieval in context*. Springer.

Józsa, E., Köles, M., Komlódi, A., Hercegf, K., & Chu, P. (2012). Evaluation of search quality differences and the impact of personality styles in native and foreign language searching tasks. In *Proceedings of the 4th Information Interaction in Context Symposium* (pp. 310–313). ACM. doi:10.1145/2362724.2362782

Kim, J., Oard, D., & Kathleen, R. (2000, May). *Using implicit feedback for user modeling in internet and intranet searching*. Retrieved November 19, 2015, from <https://terpconnect.umd.edu/~oard/pdf/clistr0001.pdf>



- Kuhlthau, C. C. (1991). Inside the search process: Information seeking from the users perspective. *Journal of the American Society for Information Science American Society for Information Science*, 42(5), 361–371. doi:10.1002/(SICI)1097-4571(199106)42:5<361::AID-ASI6>3.0.CO;2-#
- Luna, D., & Peracchio, L. A. (2002). Uncovering the cognitive duality of bilinguals through word association. *Psychology and Marketing*, 19(6), 457–475. doi:10.1002/mar.10020
- MiniWats Marketing Group. (2015). *Internet world stats: Usage and population statistics*. Retrieved October 1, 2015, from <http://www.internetworldstats.com/stats7.htm>
- Oard, D. W. (1997). Serving Users in Many Languages: Cross-Language Information Retrieval for Digital Libraries. *D-Lib Magazine*. Retrieved from <http://www.dlib.org/dlib/december97/oard/12oard.html>
- Oard, D. W., & Diekema, A. R. (1998). Cross-language information retrieval. *Annual Review of Information Science*, 33. Retrieved from http://digitalcommons.usu.edu/itls_facpub/18/
- Petrelli, D., Beaulieu, M., Sanderson, M., & Hansen, P. (2002). User requirement elicitation for cross-language information retrieval. *The New Review of Information Behavior Research-Studies of information seeking in context*, 3, 17–35.
- Riggs, F. W. (1987). Indigenous concepts: A problem for social and information science. *International Social Science Journal*, 39(4), 607–617.
- Salton, G. (1968). *Automatic information organization and retrieval*. New York: McGraw-Hill.
- Saracevic, T., Kantor, P. B., Chamis, A. Y., & Trivison, D. (1988). A study of information seeking and retrieving: Part 1, background and methodology. *Journal of the American Society for Information Science American Society for Information Science*, 39(3), 161–176. doi:10.1002/(SICI)1097-4571(198805)39:3<161::AID-ASI2>3.0.CO;2-0
- Schamber, L. (1994). Relevance and information behavior. *Annual Review of Information Science & Technology*, 29, 3–48.
- Schamber, L., Eisenberg, M. B., & Nilan, M. S. (1990). A re-examination of relevance: Toward a dynamic, situational definition. *Information Processing & Management*, 26(6), 755–775. doi:10.1016/0306-4573(90)90050-C
- Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana, IL: University of Illinois Press.
- Taylor, R. S. (1991). Information use environments. In B. Dervin (Ed.), *Progress in Communication Sciences* (Vol. 10, pp. 217–225). Norwood, NJ: Ablex.
- W3Techs.com. (2015). *Usage of content languages for websites* [Fact sheet]. Retrieved from <http://w3techs.com/>
- Wilson, P. (1973). Situational relevance. *Information Storage and Retrieval*, 9(8), 457–471. doi:10.1016/0020-0271(73)90096-X
- Wilson, T. (1997). Information behavior: An inter-disciplinary perspective. In P. Vakkari, R. Savolainen, & B. Dervin (Eds.), *In Proceedings of the International Conference on Research in Information Needs, Seeking and Use in Different Context* (pp. 39–50). Los Angeles, CA: Taylor Graham.
- Wilson, T. D. (1981). On user studies and information needs. *The Journal of Documentation*, 37(1), 3–15. doi:10.1108/eb026702

ADDITIONAL READING

- Belkin, N. J. (1980). Anomalous states of knowledge as a basis for information retrieval. *Canadian Journal of Information Science-Revue Canadienne Des Sciences De L'Information*, 5(1), 133–143.

Belkin, N. J., Oddy, R. N., & Brooks, H. M. (1982). ASK for information retrieval: Part II. Results of a design study. *The Journal of Documentation*, 38(3), 145–164. doi:10.1108/eb026726

Grefenstette, G. (1998). The problem of cross-language information retrieval. In G. Grefenstette (Ed.), *Cross-language Information Retrieval* (pp. 1–9). Boston: Kluwer Academic; doi:10.1007/978-1-4615-5661-9_1

Harter, S. P., & Hert, C. A. (1997). Evaluation of information retrieval systems: Approaches, issues, and methods. In M. E. Williams (Ed.), *Annual Review of Information Science and Technology* (Vol. 32, pp. 3–94). Medford, NJ: Information Today.

Saracevic, T. (2007). Relevance: A review of the literature and a framework for thinking on the notion in information science. Part II: nature and manifestations of relevance. *Journal of the American Society for Information Science and Technology*, 58(13), 1915–1933. doi:10.1002/asi.20682

KEY TERMS AND DEFINITIONS



Cross-Language Information Retrieval:

The process of locating relevant documents from a large collection of data in a language which is different from the query language.

Information Retrieval: The process of locating relevant documents from a large collection of data in physical and electronic forms.

Objective Relevance: The degree of precision and recall of the retrieved information to the query measured objectively by machine-run binary assessment algorithms.

Source Language: The language used in a query.

Subjective Relevance: The degree of precision and recall of the retrieved information to the query perceived subjectively by the user.

Target Language: The language used for the retrieved information, usually different from source language.

A Fast and Space-Economical Algorithm for the Tree Inclusion Problem

Yangjun Chen

University of Winnipeg, Canada

Yibin Chen

University of Winnipeg, Canada

INTRODUCTION

Let T be a rooted tree. We say that T is *ordered* and *labeled* if each node is assigned a symbol from an alphabet Σ and a left-to-right order among siblings in T is specified. Let v be a node different of the root in T with parent node u . Denote by $delete(T, v)$ the tree obtained by removing the node v from T , by which the children of v become part of the children of u as illustrated in Figure 1.

Given two ordered labeled trees P and T , called the pattern and the target, respectively. We may ask: Can we obtain pattern P by deleting some nodes from target T ? That is, is there a sequence $v_1, \dots, v_k$ of nodes such that for

$$T_0 = T \text{ and}$$

$$T_{i+1} = delete(T_i, v_{i+1}) \text{ for } i = 0, \dots, k-1,$$

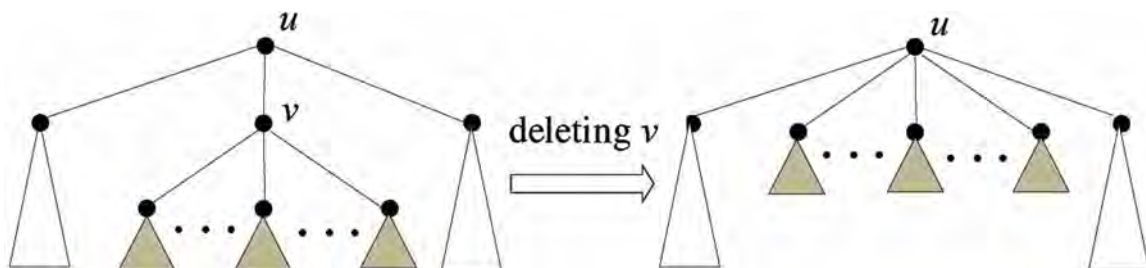
we have $T_k = P$? If this is the case, we say, P is included in T (Kilpeläinen and Mannila, 1995).

Such a problem is called the *tree inclusion problem*. It has many applications in the computer engineering as described below.

BACKGROUND

The first interesting application of the tree inclusion is used as an important query primitive for XML data (Mannila and R  iha, 1990), where a structured document database is considered as a collection of parse trees that represent the structure of the stored texts and the tree inclusion is used as a means of retrieving information from them. As an example, consider the tree shown in Figure 2, representing an XML document for the book *Arts of Programming* authored by (Knuth, 1969). One might want to find this book in an XML database by forming a pattern tree as shown in Figure 3 as a query, which can be obtained by deleting some nodes from the tree shown in Figure 2. Thus, a tree inclusion checking needs to be conducted to evaluate this query.

Figure 1. Illustration of node deletion



DOI: 10.4018/978-1-5225-2255-3.ch391

Figure 2. A XML document (target) tree

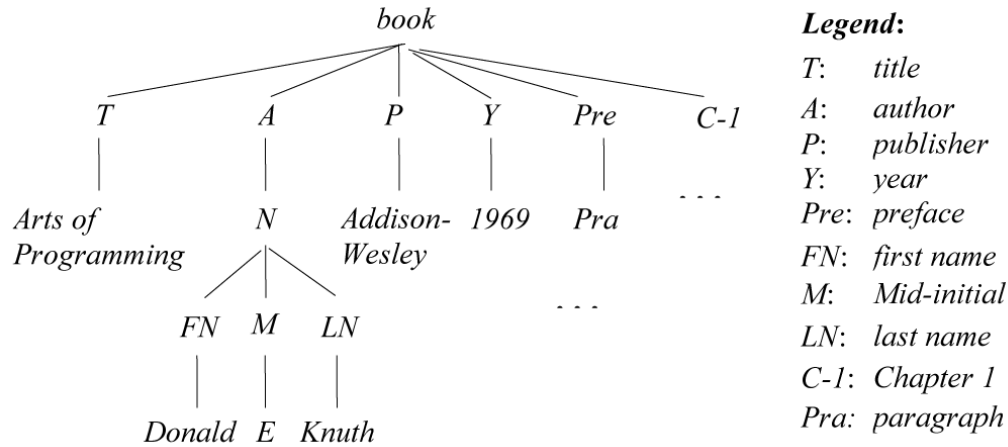
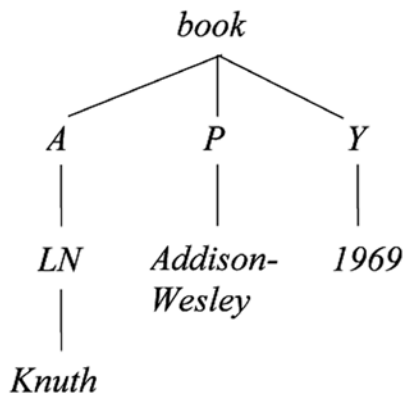


Figure 3. A pattern tree



Another application of this problem is to query the grammatical structures of *English* sentences, which can also be represented as an ordered tree since a sentence can always be divided into several ordered components such as noun phrases, verb phrases, and adverbs; and a noun phrase itself normally contains an article and a noun while a verb phrase may contain a verb, a noun phrase, an adverb, and so on. To check whether a concrete sentence is grammatically correct, we will represent it as a pattern tree and make a tree inclusion checking against some target grammatical tree structures.

A third application of the ordered tree inclusion is the video content-based retrieval. According to (Rui and Huang, 1999), a video can be successfully

decomposed into a hierarchical tree structure, in which each node represents a scene, a group, a shot, a frame, a feature, and so on. Especially, such a tree is an ordered one since the temporal order is very important for video. Some other areas, in which the ordered tree inclusion finds its applications, are the scene analysis, the computational biology, such as *RNA* structure matching (Lyngs, Zuker and Pedersen, 1999), and the data mining, such as tree mining discussed in (Zaki, 2002), just to name a few.

Up to now, the best algorithm for this problem requires $O(|T| + |P|)$ space and $\Theta(|T| \cdot |\text{leaves}(P)|)$ time (Alonso and Schott, 1993; Chen, 1998; Chen and Chen, 2006; Bille and Gørtz, 2011), where $\text{leaves}(P)$ stands for the set of the leaves of P .

In this chapter, we propose an efficient algorithm for this problem. Its time and space complexities are bounded by $O(|T| \cdot \min\{h_p, |\text{leaves}(P)|\})$, and $O(|T| + |P|)$, respectively, where h_p is the height of P , defined to be the number of edges on the longest downward path from the root to a leaf node.

BASIC DEFINITIONS

In this section, we mainly define the notations that will be used throughout the paper. Let T be a

labeled tree that is ordered, i.e., the order between siblings is significant. We denote the set of nodes and edges by $V(T)$ and $E(T)$, respectively. The size of T is denoted by $|T|$.

Technically, it is convenient to consider a slight generalization of trees, namely forests, which are defined to be a set of disjoint trees. A tree T consisting of a specially designated node $root(T) = t$ (called the root of the tree) and a forest $\langle T_1, \dots, T_k \rangle$ (where $k \geq 0$) is denoted as $\langle t; T_1, \dots, T_k \rangle$. We also call T_j ($1 \leq j \leq k$) a direct subtree of t .

Let u, v be two nodes in T . If there is path from node u to node v , we say, u is an ancestor of v and v is a descendant of u . In this paper, by *ancestor* (*descendant*), we mean a proper ancestor (descendant), i.e., $u \neq v$. We will use $u \Rightarrow v$ to represent that u is a proper ancestor of v .

The ancestorship in a tree can be checked very efficiently by using a kind of tree encoding (Chen, 2006), which labels each node v in a tree with an interval $I_v = [a_v, b_v]$, where b_v denotes the rank of v in a *post-order* traversal of the tree. Here the ranks are assumed to begin with 1, and all the children of a node are assumed to be ordered and fixed during the traversal. Furthermore, a_v denotes

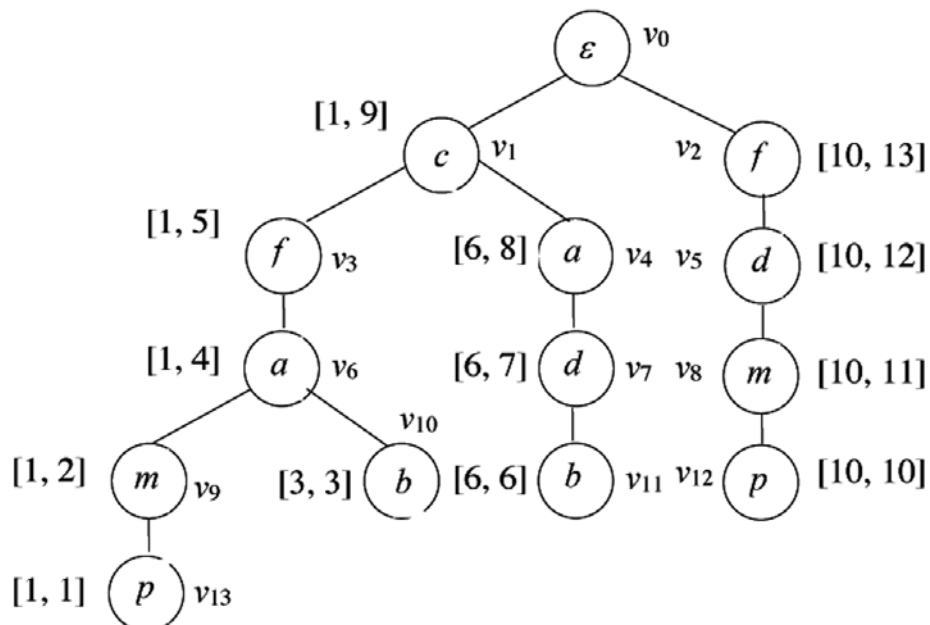
the lowest rank for any node u in $T[v]$ (the subtree rooted at v , including v). Thus, for any node u in $T[v]$, we have $I_u \subseteq I_v$ since the post-order traversal visits a node after all its children have been visited. In Figure 4, we illustrate such a tree encoding, assuming that the children are ordered from left to right. It is easy to see that by interval containment we can check whether two nodes are on a same path. For example, $v_3 \Rightarrow v_{10}$, since $I_{v_3} = [1, 5]$, $I_{v_{10}} = [3, 3]$, and $[3, 3] \subset [1, 5]$; but v_9 is not reachable from v_2 since $I_{v_2} = [10, 13]$, $I_{v_9} = [1, 2]$, and $[1, 2] \not\subset [10, 13]$.

Let $I = [a, b]$ be an interval. We will refer to a and b as $I[1]$ and $I[2]$, respectively.

Lemma 1: For any two intervals I and I' generated for two nodes in a tree T , one of four relations holds: $I \subset I'$, $I' \subset I$, $I[2] < I'[1]$, or $I'[2] < I[1]$.

Based on Lemma 1, the left-to-right ordering of nodes can also formally be defined. A node u is said to be to the left of v if they are not related by the ancestor-descendant relationship and v fol-

Figure 4. Illustration for tree encoding



lows u when we traverse T in preorder. Then, u is to the left of v if and only if $I_u[2] < I_v[1]$.

In the following, we use $<$ to represent the left-to-right ordering. Also, $v \leq v'$ iff $v < v'$ or $v = v'$. The following definition is due to [9].

Definition 1: Let F and G be labeled ordered forests. We define an ordered embedding (φ, G, F) as an injective function $\varphi: V(G) \rightarrow V(F)$ such that for all nodes $v, u \in V(G)$,

1. $\text{label}(v) = \text{label}(\varphi(v))$; (label preservation condition)
2. $v \Rightarrow u$ iff $\varphi(v) \Rightarrow \varphi(u)$, i.e., $I_u \subset I_v$ iff $I_{\varphi(u)} \subset I_{\varphi(v)}$; (ancestor condition)
3. $v < u$ iff $\varphi(v) < \varphi(u)$, i.e., $I_v[2] < I_u[1]$ iff $I_{\varphi(v)}[2] < I_{\varphi(u)}[1]$. (sibling condition) $\square$

If there exists such an injective function from $V(G)$ to $V(F)$, we say, F includes G , F contains G , F covers G , or say, G can be embedded in F . Figure 5 shows an example of an ordered tree inclusion.

Let P and T be two labeled ordered trees. An embedding φ of P in T is said to be *root-preserving* if $\varphi(\text{root}(P)) = \text{root}(T)$. If there is a root-preserving embedding of P in T , we say that the root of T is an occurrence of P . Figure 5b also shows an example of a root preserving embedding. According to (Kilpeläinen and Mannila,

1995), restricting to root-preserving embedding does not lose generality. In fact, the method to be discussed works top-down and always tries to find root-preserving subtree embeddings. Throughout the rest of the paper, the outdegree of v in a tree is denoted by $d(v)$ while the height of v is denoted by $h(v)$. The height of a leaf node is set to be 0. In addition, we refer to the labeled ordered trees simply as trees.

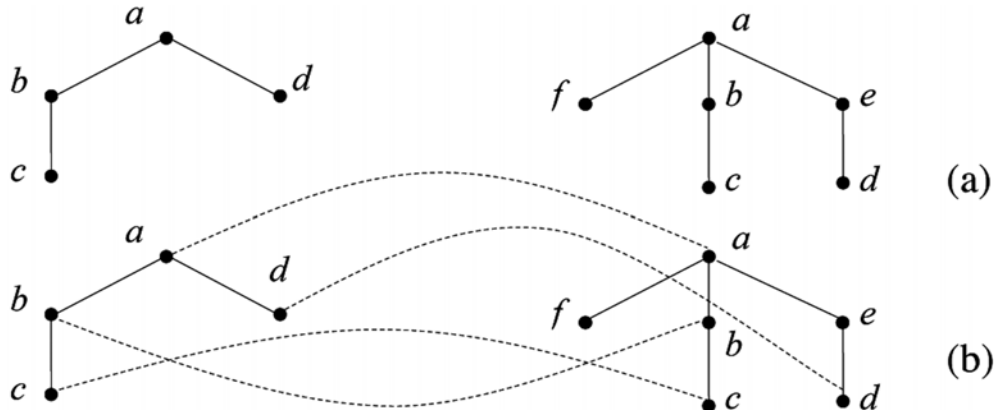
Throughout the rest of the paper, the outdegree of v in a tree is denoted by $d(v)$ while the height of v is denoted by $h(v)$. The height of a leaf node is set to be 0. In addition, we refer to the labeled ordered trees simply as trees.

MAIN IDEA: CUTS

The main idea of our algorithm consists in a mechanism called *cut checking* introduced into a top-down tree search to get rid of useless computation.

Let $T = \langle t; T_1, \dots, T_k \rangle$ ($k \geq 0$) be a tree and $G = \langle P_1, \dots, P_q \rangle$ ($q \geq 0$) be a forest. We handle G as a tree $P = \langle \nu_G; P_1, \dots, P_q \rangle$, where ν_G represents a virtual node, matching any node in T . Note that even though G contains only one single tree it is considered to be a forest. So a virtual root is added. Therefore, each node in G , except the virtual node, has a parent.

Figure 5. (a) The tree on the left can be included in the tree on the right by deleting the nodes labeled: f and e ; (b) the embedding corresponding to (a)



Consider a node v in $G = \langle P_1, \dots, P_q \rangle$ with children $v_1, \dots, v_k$. We use a pair $\langle [i, j], v \rangle$, called an *interval* rooted at v , to represent an ordered forest $\langle G[v_i], \dots, G[v_j] \rangle$ made up of a series of subtrees rooted at $v_i, \dots, v_j$, respectively. Especially, $\langle [1, i], v \rangle$ (or simply denoted as $\langle i, v \rangle$) represents an ordered forest containing the first i subtrees of v : $\langle G[v_1], \dots, G[v_i] \rangle$. If v is ν_G , or a node on the left-most path in P_1 , $\langle i, v \rangle$ is called a *left-corner* of G [5]. Obviously, $\langle i, \nu_G \rangle$ is a left-corner, representing the first i subtrees in G : $P_1, \dots, P_i$. So, $\langle q, \nu_G \rangle$ stands for the whole G .

In addition, we will use $\langle i, v \rangle$ to represent the forest $\langle G[v_{i+1}], \dots, G[v_k] \rangle$, referred to as the complement of $\langle i, v \rangle$. When it is clear from a context, we may use $\langle G[v_i], \dots, G[v_j] \rangle$ and $\langle [i, j], v \rangle$ interchangeably without causing any confusion.

Let u be a node on the left-most path in P_1 . Let $\langle i, v \rangle$ be a left-corner of $G = \langle P_1, \dots, P_q \rangle$. If $v = u$, we say that $\langle i, v \rangle$ and u are *level-equal*, denoted as $\langle i, v \rangle \cong u$. If v is an ancestor of u , we say, $\langle i, v \rangle$ is higher than u , denoted as $\langle i, v \rangle \succ u$. Then, $\langle i, v \rangle \preceq u$ represents that $\langle i, v \rangle$ is higher than or level-equal to u .

In particular, we will use $A(T, G) = \langle i, v \rangle$ to represent a checking of G against T , returning a left-corner $\langle i, v \rangle$ in G with the following properties:

- If $i > 0$ and v is not the left-most leaf node, it shows that
 - The first i subtrees of v can be embedded in T ;

- For any $i' > i$, $\langle i', v \rangle$ cannot be embedded in T ; and
- For any v 's ancestor u on the left-most path in G , there exists no $j > 0$ such that $\langle j, u \rangle$ is able to be embedded in T .
- If $i = 0$ or v is the left-most leaf node of G (denoted as $\rho(G)$), it indicates that no left-corner of G can be embedded in T .

In this sense, we say, $\langle i, v \rangle$ is the *highest* and *widest* left-corner which can be embedded in T .

Now we consider a tree T and a forest G shown in Figure 6, in which each node in T is identified with t_i , such as t_1, t_2, t_{11} , and so on; and each node in G is identified with p_j . Besides, each subtree rooted at t_i (p_j) is represented by T_i (resp. P_j).

In order to check whether T includes $G = \langle P_1, P_2 \rangle$, we can first check whether T_1 alone includes G . That is, we will perform a recursive call as follows:

$$A(T, \langle P_1, P_2 \rangle) \rightarrow A(T_1, \langle P_1, P_2 \rangle).$$

Assume that $A(T_1, \langle P_1, P_2 \rangle)$ returns $\langle i, v \rangle$. We may have one of three cases:

Case 1: $\langle i, v \rangle = \langle 2, \nu_G \rangle$.

Case2: $\langle i, v \rangle = \langle 1, \nu_G \rangle$.

Case 3: $v \neq \nu_G$, but a node on the left-most path in P_1 . That is, T_1 contains only a left-corner not higher than p_1 .

Figure 6. A tree and a forest

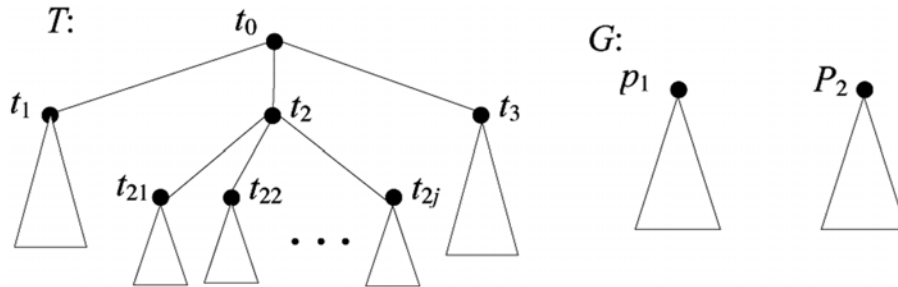
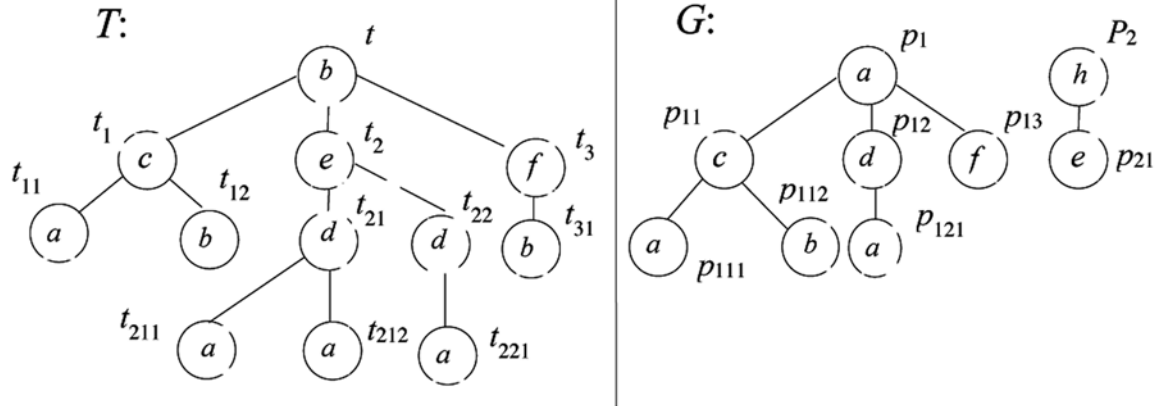


Figure 7. A target and a pattern tree



In Case 1, T_1 contains G .

In Case 2, T_1 contains only P_1 , and we will call $A(T_2, \langle P_2 \rangle)$ in a next step.

In Case 3, we will continually check whether T_2 alone is able to include G (by calling $A(T_2, \langle P_1, P_2 \rangle)$). This time, however, we will use v (in $\langle i, v \rangle$) to control the working process to cut off part of the computation once we find that a left-corner higher than v cannot be produced. It is because such a computation will not make any contribution to the final result due to the following observation.

Assume that $A(T_2, \langle P_1, P_2 \rangle)$ returns $\langle i', v' \rangle$ with $v = v'$ or $v \Rightarrow v'$. Then, in a next step, we will check T_3 against $\langle P_1, P_2 \rangle$ by calling $A(T_3, \langle P_1, P_2 \rangle)$. If its return left-corner is higher than v , then we will use this left-corner as the return value of $A(T, \langle P_1, P_2 \rangle)$. Then, $\langle i', v' \rangle$ is not used. If its return left-corner is not higher than v , we will make a *supplement checking* of $\langle T_2, T_3 \rangle$ against $\langle i, v \rangle$ to see whether $\langle T_2, T_3 \rangle$ is able to embed some subtrees in $\langle i, v \rangle$. Assume that $\langle T_2, T_3 \rangle$ embeds the first j subtrees in $\langle i, v \rangle$. Then, the return value of $A(T, \langle P_1, P_2 \rangle)$ should be $\langle i + j, v \rangle$. In this case, $\langle i', v' \rangle$ is not used, either, in terms of the following analysis:

If $v \Rightarrow v'$, or $v = v'$ but $i' \leq i$, $\langle i', v' \rangle$ is obviously useless for the final result. However, even

if $v = v'$ with $i' > i$, it is still useless since in this case, there is definitely an integer $j \geq i' - i$ such that $\langle T_2, T_3 \rangle$ embeds the first j subtrees in $\langle i, v \rangle$, and the supplement computation will find this embedding.

The above discussion shows that if $A(T_2, \langle P_1, P_2 \rangle)$ cannot return a left-corner higher than v , the corresponding work is futile and should be avoided. However, avoiding the whole work seems not possible. Yet we can really effectively block a significant part of the useless computation by using the partial results obtained in the previous steps, as illustrated in Example 1.

Example 1: Consider the tree T and the forests G shown in Figure 7.

In order to check whether T includes G , we will first check T_1 against $G = \langle P_1, P_2 \rangle$. Obviously, T_1 is not able to embed G . However, it can embed P_{11} and therefore the return value of this checking should be $\langle 1, p_1 \rangle$. In a next step, we will check T_2 against G , and try to see if T_2 alone is able to embed G . But this time, p_1 will be utilized to control the process. More specifically, it will effectively block the checking of T_2 against P_{11} since this checking can only possibly return a left-corner not higher than p_1 . $\square$

We refer to a node like p_1 in Example 1 as a *cut*.

ALGORITHM

In this section, we present our basic algorithm $A(T, G)$ to check a tree $T (= \langle t; T_1, \dots, T_k \rangle)$ against a forest $G (= \langle P_1, \dots, P_q \rangle)$. The algorithm works in a multiple recursive way in the sense that different kinds of recursive calls will be carried out in terms of different characteristics of inputs. As a whole, two general cases need to be recognized:

In Case 1, we have $G = \langle P_1 \rangle$; or $G = \langle P_1, \dots, P_q \rangle$ with $q > 1$, but $|T| \leq |P_1| + |P_2|$. In this case, what we can do is to check T against P_1 since it is not possible for T to embed more than one subtree in G .

In Case 2, we have $G = \langle P_1, \dots, P_q \rangle$ with $q > 1$, and $|T| > |P_1| + |P_2|$. In this case, we will check $\langle T_1, \dots, T_k \rangle$ against the whole G since in this case we may have a sequence of subtrees $T_1, \dots, T_{l_m}$ with each being able to embed some subtrees in G . For this reason, we define two subfunctions: α -function and β -function, used to handle Case 1 and Case 2, respectively:

$$\alpha(T, P_1) = \langle i, v \rangle, \quad (1)$$

where $\langle i, v \rangle$ is a highest and widest left-corner in P_1 , which can be embedded in T .

$$\beta(\langle T_1, \dots, T_k \rangle, G) = \langle j, u \rangle, \quad (2)$$

where $\langle j, u \rangle$ is a highest and widest left-corner in G , which can be embedded in $\langle T_1, \dots, T_k \rangle$.

Here, our intention is quite straightforward:

In Case 1 we will call $\alpha(T, P_1)$ and in Case 2 we will call $\beta(\langle T_1, \dots, T_k \rangle, G)$. However, during the working process, they may call each other recursively.

Additionally, in Case 2, the return value $\langle j, u \rangle$ of $\beta(\langle T_1, \dots, T_k \rangle, G)$ needs to be further checked as follows:

- If $u \neq p_1$'s parent, check whether $\text{label}(t) = \text{label}(u)$ and $j = d(u)$. If it is not the case, the return value of $A(T, G)$ is the same as $\langle j, u \rangle$. Otherwise, the return value of $A(T, G)$ should be set to $\langle 1, u$'s parent $\rangle$, showing that T includes $G[u]$. (For this reason, $d(v_G)$ is set to be ∞ , larger than the outdegree of any node in both T and G .)
- If $u = p_1$'s parent, the return value of $A(T, G)$ is the same as $\langle j, u \rangle$, showing that T embeds $\langle P_1, \dots, P_j \rangle$.

By using the α -function and the β -function, the algorithm can be described as shown in Function 4, in which node v (as the third input) works as a cut.

*Function 4. $A(T, G, v)$ (*Initially, cut v is set to be $\rho(G)$.*)*

Input: $T = \langle t; T_1, \dots, T_k \rangle, G = \langle P_1, \dots, P_q \rangle$.

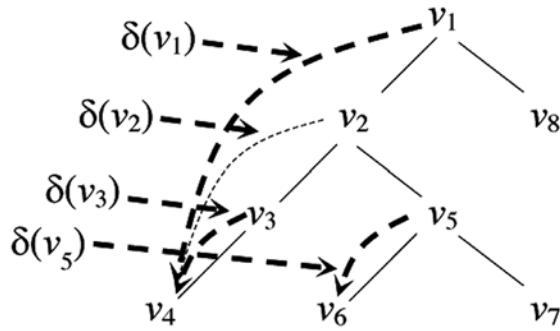
Output: a left corner.

Begin

```

1.   if  $p_1$ 's parent is not an ancestor of  $v$  then return  $\langle 0, \rho(G) \rangle$ ;
2.   if  $(q = 1 \text{ or } |T[t]| \leq |G[p_1]| + |G[p_2]|)$  (*Case 1*)
3.   then return  $\alpha(T, P_1, v)$ 
4.   else if  $\text{label}(t) = \text{label}(v)$  (*Case 2*)
5.       then  $\langle j, u \rangle := \beta(\langle T_1, \dots, T_k \rangle, G, v$ 's first child);
6.       else  $\langle j, u \rangle := \beta(\langle T_1, \dots, T_k \rangle, G, v)$ ;
7.       if  $v \neq p_1$ 's parent
8.       then if  $(\text{label}(t) = \text{label}(u) \wedge j = d(u))$ 
9.           then return  $\langle 1, u$ 's parent  $\rangle$ ;
10.      return  $\langle j, u \rangle$ ;
end
```

Figure 8. A pattern tree



In the following, both the α -function and β -function will be discussed in great detail.

• α -function

In order to implement the α -function, we need to associate each node v in G with a link to the left-most leaf node in $G[v]$, denoted as $\delta(v)$, as illustrated in Figure 8.

Let v' be a leaf node in G . $\delta(v')$ is defined to be a link to v' itself. So in Figure 8, we have $\delta(v_1) = v_1$, $\delta(v_2) = v_2$, $\delta(v_3) = v_3$, $\delta(v_4) = v_4$, $\delta(v_5) = v_5$, $\delta(v_6) = v_6$, $\delta(v_7) = v_7$, and $\delta(v_8) = v_8$. Denote by $\delta^{-1}(v')$ a set of nodes x such that for each $v \in x$ $\delta(v) = v'$. Thus, in Figure 8, we have $\delta^{-1}(v_4) = \{v_1, v_2, v_3, v_4\}$, $\delta^{-1}(v_6) = \{v_5, v_6\}$, $\delta^{-1}(v_7) = \{v_7\}$, and $\delta^{-1}(v_8) = \{v_8\}$. Let p_1 be the root of P_1 . We also have $\rho(G) = \delta(p_1)$.

Let $T = \langle t; T_1, \dots, T_k \rangle$, $G = \langle P_1, \dots, P_q \rangle$. In $\alpha(T, P_1)$, altogether five different cases as listed in Figure 9 should be checked.

Obviously, in Case (1-1), where t is a leaf node, we will check whether $\text{label}(t) = \text{label}(\delta(p_1))$ since $\delta(p_1)$ is the only left-corner which can possibly be covered by t . If it is the case, return $\langle 1, \text{parent of } \delta(p_1) \rangle$. Otherwise, return $\langle 0, \delta(p_1) \rangle$.

In Case (1-2), where $|T| > 1$, but $|T| < |P_1|$ and/or $h(t) < h(p_1)$, we will make a recursive call $A(T, \langle P_{11}, \dots, P_{1j} \rangle)$, where $\langle P_{11}, \dots, P_{1j} \rangle$ is a forest containing all the direct subtrees of p_1 . The return value of $A(T, \langle P_{11}, \dots, P_{1j} \rangle)$ is used as the return value of $\alpha(T, P_1)$. It is because in this case, T is not able to embed the whole P_1 . So we will try to find whether T is able to embed a left-corner within $\langle P_{11}, \dots, P_{1j} \rangle$.

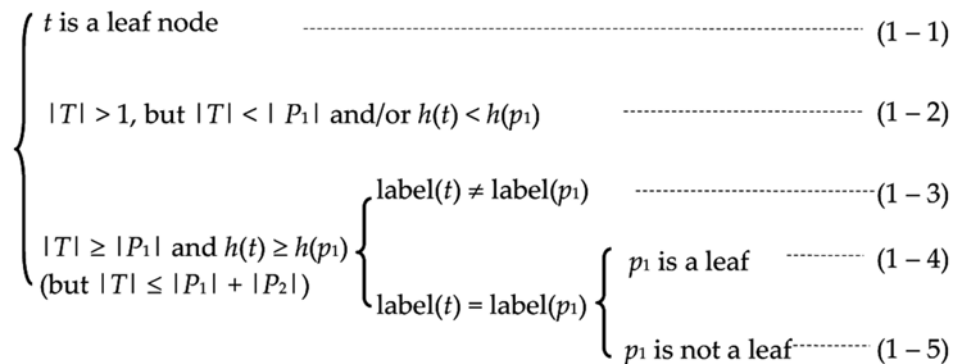
If $|T| \geq |P_1|$ and $h(t) \geq h(p_1)$ (but $|T| \leq |P_1| + |P_2|$), we further distinguish among three cases: Case (1-3), (1-4) and (1-5).

In Case (1-3), we have $\text{label}(t) \neq \text{label}(p_1)$, and we will call $\beta(\langle T_1, \dots, T_k \rangle, \langle P_1 \rangle)$ to see whether $\beta(\langle T_1, \dots, T_k \rangle)$ is able to embed P_1 .

In Case (1-4), we have $\text{label}(t) = \text{label}(p_1)$ and p_1 is a leaf node. In this case, we return $\langle 1, p_1$'s parent $\rangle$.

In Case (1-5), we have $\text{label}(t) = \text{label}(p_1)$, but p_1 is not a leaf node. In this case, we need to call $\beta(\langle T_1, \dots, T_k \rangle, \langle P_{11}, \dots, P_{1j} \rangle)$. Assume that the return value of $\beta()$ is $\langle i, v \rangle$. We need to do an extra checking:

- If $\text{label}(t) = \text{label}(v)$ and $i = d(v)$, the return value of $A(T, G)$ is set to be $\langle 1, v$'s parent $\rangle$.

Figure 9. Different cases to be checked in α -function

- Otherwise, the return value of $\alpha(T, G)$ is the same as $\langle i, v \rangle$.

According to the above discussion, we give the following formal algorithm for the α -function (Function 5).

In Function 5, due to the use of cuts, Case (1-5) is further divided into two subcases:

1. $p_1 = v$. In this case, we will call $\beta(\langle T_1, \dots, T_k \rangle, \langle P_{11}, \dots, P_{1j} \rangle, p_{11})$ with the cut being set to be p_{11} . It is because in this case the main checking of the β -function execution may reveal that $\langle T_1, \dots, T_k \rangle$ is able to embed the whole $\langle P_{11}, \dots, P_{1j} \rangle$. In the case, the return value of $\alpha(T, G, v)$ will be set to $\langle 1, p_1 \text{'s parent} \rangle$, higher than v . So it is a useful computation; and downgrading the cut from $v = p_1$ to p_{11} will let it go through. On the other hand, p_{11} will effectively prohibit any possible further supplement checking in this β -function execution since such a checking can only bring out a left corner lower than p_{11} and will not be used.
2. $p_1 \approx v$. In this case, we will call $\beta(\langle T_1, \dots, T_k \rangle, \langle P_{11}, \dots, P_{1j} \rangle, v)$, by which v is directly transferred since we must have $p_{11} \approx v$ and

no useful computation can be eliminated by cut v .

- β -function

In comparison with the α -function, the β -function is more interesting. It is designed to handle the general Case 2. Let $F = \langle T_1, \dots, T_k \rangle$ and $G = \langle P_1, \dots, P_q \rangle$. Denote by t_l the root of T_l ($l = 1, \dots, k$). Denote by p_j the root of P_j ($j = 1, \dots, q$). In $\beta(F, G, v)$, we will make a series of calls $A(T_l, \langle P_{j_1}, \dots, P_{j_x} \rangle, v_l)$, where $l = 1, \dots, x \leq k, j_1 = 1, j_1 \leq j_2 \leq \dots \leq j_x \leq q$, controlled as follows.

1. Two index variables l, j are used to scan $T_1, \dots, T_k$ and $P_1, \dots, P_q$, respectively. (Initially, l is set to 1, and j is set to 0.) They also indicate that $\langle P_1, \dots, P_j \rangle$ has been successfully embedded in $\langle T_1, \dots, T_l \rangle$.
2. Let $\langle i_l, u_l \rangle$ be the return value of $A(T_l, \langle P_{j+1}, \dots, P_q \rangle, v_l)$. If $u_l = p_{j+1}$'s parent, set j to $j + i_l$ and v_{l+1} to p_j . Otherwise, j is not changed. Set l to $l + 1$. Go to (2) (i.e., repeat this step.)
3. The loop terminates when all T_l 's or all P_j 's are examined. (Figure 10 helps for illustration of this iteration process.)

Function 5. $\alpha(T, P1, v)$

Input: $T = \langle t; T_1, \dots, T_k \rangle, P_1 = \langle p_1; P_{11}, \dots, P_{1j} \rangle$.

Output: a left corner.

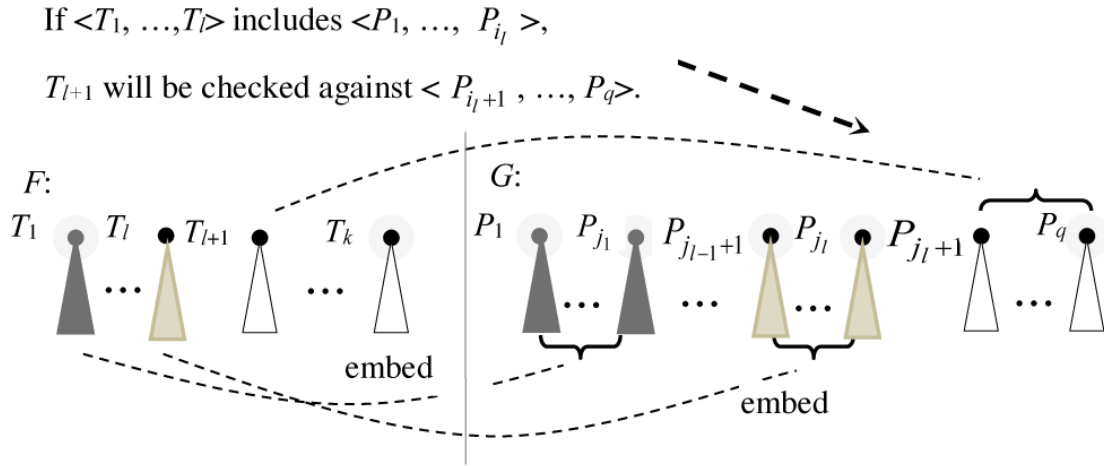
begin

```

1. if (1-1) then if label( $t$ ) = label( $\delta(p_1)$ )
2.           then return  $\langle 1, \delta(p_1) \text{'s parent} \rangle$ 
3.           else return  $\langle 0, \delta(p_1) \rangle$ ;
4. if (1-2) then return  $A(T, \langle P_{11}, \dots, P_{1j} \rangle, v)$ ;
5. if (1-3) then return  $\beta(\langle T_1, \dots, T_k \rangle, \langle P_1 \rangle, v)$ ;
6. if (1-4) then return  $\langle 1, p_1 \text{'s parent} \rangle$ 
7. if (1-5-i) then  $\langle j, u \rangle := \beta(\langle T_1, \dots, T_k \rangle, \langle P_{11}, \dots, P_{1j} \rangle, p_{11})$ 
8. if (1-5-ii) then  $\langle j, u \rangle := \beta(\langle T_1, \dots, T_k \rangle, \langle P_{11}, \dots, P_{1j} \rangle, v)$ ;
9. if  $j = d_{out}(u)$  and label( $t$ ) = label( $u$ ) then return  $\langle 1, u \text{'s parent} \rangle$  else re-
turn  $\langle j, u \rangle$ ;

```

end

Figure 10. Illustration for an execution of β -function

4. If $j > 0$ when the loop terminates, $\beta(F, G)$ returns $\langle j, p_1 \text{'s parent} \rangle$, indicating that F contains $P_1, \dots, P_j$. Otherwise, $j = 0$, indicating that even P_1 alone cannot be embedded in any T_λ ($\lambda \in \{1, \dots, \kappa\}$). However, in this case, we need to continue looking for a highest and widest left-corner $\langle i, u \rangle$ in P_1 , which can be embedded in F . This can be done as follows.

- a. Let $\langle i_1, v_1 \rangle, \dots, \langle i_k, v_k \rangle$ be the return values of $A(T_1, \langle P_1, \dots, P_q \rangle, v_1), \dots, A(T_k, \langle P_1, \dots, P_q \rangle, v_k)$, respectively. Since $j = 0$, each $v_l \in \delta^{-1}(\rho(G))(G)$ ($l = 1, \dots, k$).

If each $i_l = 0$, the return value of $\beta(F, G)$ should be $\langle 0, (G) \rangle$. Otherwise, there must be some v_l 's (higher than v) with $i_l > 0$. We call such a node a *non-zero point*. Find the first non-zero point v_f with children $w_1, \dots, w_s$ such that v_f is not a descendant of any other non-zero point. Then, we will check $\langle T_{f+1}, \dots, T_k \rangle$ against $\langle P[w_{i_f+1}], \dots, P[w_s] \rangle$. This can be done by a recursive call $\beta(\langle T_{f+1}, \dots, T_k \rangle, \langle P[w_{i_f+1}], \dots, P[w_s] \rangle)$. Let y be a number such that $\langle P[w_{i_f+1}], \dots, P[w_{i_f+1}] \rangle$ can be embedded in $\langle T_{f+1}, \dots, T_k \rangle$. The return value of $\beta(F, G)$ should be set to $\langle i_f + y, v_f \rangle$. $\square$

In the above process, (1), (2) and (3) together are referred to as a *main computation* while (4) alone as a *supplement computation*. In addition, special attention should be paid to the condition under which a supplement computation is conducted:

- $j = 0$, and
- There exists at least a non-zero point.

We refer to this condition as the *supplement checking condition* (SCC-condition for short). In terms of the above discussion, we give the following formal algorithm for the β -function (Function 6).

In Function 6, we have two *while*-loops: one from line 2 to 7 and the other from line 12 to 15. In the first *while*-loop, we do the main computation to find a largest j such that $\langle T_1, \dots, T_k \rangle$ embeds $\langle P_1, \dots, P_j \rangle$. In the second *while*-loop, the *supplement computation* will be conducted. However, this is done when the SCC-condition is satisfied.

In order to record the first non-zero point which is not a descendant of any other non-zero point, variable f is used. Initially, f is set 0. Therefore, if no non-zero point is found, we must have $f = 0$ after the main computation is completed. So only when $j = 0$ and $f > 0$, the SCC-condition is satisfied and the supplement computation will be performed (see lines 8 and 9), in which we check

Function 6. $\beta(F, G, v)$

Input: $F = \langle T_1, \dots, T_k \rangle$, $G = \langle P_1, \dots, P_q \rangle$.

Output: a left corner.

begin

```

1.       $l := 1$ ;  $j := 0$ ;  $u := v$ ;  $f := 0$ ;
2.      while ( $j < q$  and  $l \leq k$ ) do                (*main checking*)
3.           $\langle i_l, u_l \rangle := A(T_l, \langle P_{j+1}, \dots, P_q \rangle, u)$ 
4.          if ( $u_l = p_1$ 's parent and  $i_l > 0$ ) then  $\{j := j + i_l; u := p_j;\}$ 
5.          else if ( $u_l$  is an ancestor of  $u$  and  $i_l > 0$ )
6.              then  $\{u := u_l; f := l;\}$ 
7.               $l := l + 1$ ;
8.          if  $j > 0$  then return  $\langle j, p_1$ 's parent  $\rangle$ ;
9.          if  $f = 0$  then return  $\langle 0, \delta(p_1) \rangle$ ;
10.         let  $w_1, \dots, w_s$  be the children of  $u_f$ ;                (*supplement checking*)
11.          $l := f + 1$ ;  $j := i_f$ ;
12.         while ( $j < s$  and  $l \leq k$ ) do
13.              $\langle i_l, v_l \rangle := A(T_l, \langle G[w_{j+1}], \dots, G[w_s] \rangle, w_{j+1})$ ;
14.             if ( $v_l = v_f$  and  $i_l > 0$ ) then  $j := j + i_l$ ;
15.              $l := l + 1$ ;
16.         return  $\langle j, u_f \rangle$ ;

```

end

$\langle T_{f+1}, \dots, T_k \rangle$ against $\langle P[w_{i_f+1}], \dots, P[w_s] \rangle$, where $w_{i_f+1}, \dots, w_s$ are all those children of the first non-zero point v_f such that the subtrees rooted at them are not covered by $\langle T_1, \dots, T_f \rangle$. (Notice that $\langle i_f, v_f \rangle$ is the return value of $A(T_f, \langle P_1, \dots, P_q \rangle$ with $i_f > 0$.) The following example helps for illustration. In this example, we trace the execution of our basic algorithm when applied to the tree T and the forest G shown in Figure 7.

CONCLUSION

In this paper, a new algorithm is proposed to solve the ordered tree inclusion problem. Up to now, the best algorithm for this problem needs quadratic time. However, ours requires only $O(|T| \cdot \min\{h_p, |\text{leaves}(P)|\})$ time and $O(|T| + |P|)$ space, where T and P are a target and a pattern tree (forest), respectively; h_p is the height of P and $\text{leaves}(P)$ is the set containing all the leaf nodes of P . The critical

concepts of our algorithm are the *left-corner* and *cuts*, which enables us to develop a deep insight into the tree inclusion problem and extend it to a more general one to return a left corner as a result. In practice, the general problem seems to be more useful than the original one since if P cannot be embedded in T , we may want to know whether any part of P can be embedded in T . In addition, our algorithm is more efficient than any existing method for the problem by using cuts to skip over useless computations.

FUTURE WORK

In the near future, we will work on the unordered tree inclusion, by which the order of sibling is not important. According to (Kilpeläinen and Mannila, 1995), this problem is NP -complete. Therefore, our research will focus on the approximate algorithms. For example, we may relax the matching criteria to allow more than one siblings in a pattern

to be mapped to a single node in a target as long as they share the same label as done by *Twig Joins* (Chen, 2006). We can also allow *k-mismatches* as done by the string matching problem, which can be more useful for the *RNA* structure matching, and the data mining.

REFERENCES

- Alonso, L., & Schott, R. (1993). On the tree inclusion problem. *Proceedings of Mathematical Foundations of Computer Science*, 211-221. doi:10.1007/3-540-57182-5_13
- Bille, P., & Gørtz, I. L. (2011). The Tree Inclusion Problem: In Linear Space and Faster. *ACM Transaction on Algorithms*, 7(3), 38:1-38:47.
- Chen, W. (1998). More efficient algorithm for ordered tree inclusion. *Journal of Algorithms*, 26(2), 370–385. doi:10.1006/jagm.1997.0899
- Chen, Y. (2006, October). On the Stack Encoding and Twig Joins. *WSEAS Transactions on Information Science & Applications*, 3(10), 1865–1872.
- Chen, Y., & Chen, Y. B. (2006). A New Tree Inclusion Algorithm. *Information Processing Letters*, 98(6), 253–262. doi:10.1016/j.ipl.2005.11.022
- Chen, Y., & Chen, Y. B. (2012). A Linear-Space Top-down Algorithm for Tree Inclusion Problem. *Proc. 2nd Int. Conf. on Computer Science and Service System (CSSS2012)*, 2127-2131. doi:10.1109/CSSS.2012.30
- Chen, Y., & Chen, Y. B. (2014). A Time and Space Efficient Algorithm for Tree Inclusion Problem. *Proc. International Conference on Future Communication, Information and Computer Science (FCICS 2014)*.
- Cheng, H. L., & Wang, B. F. (2007). On Chen and Chens new tree inclusion algorithm. *Information Processing Letters*, 103(1), 14–18. doi:10.1016/j.ipl.2007.02.006
- Kilpeläinen, P., & Mannila, H. (1995). Ordered and unordered tree inclusion. *SIAM Journal on Computing*, 24(2), 340–356. doi:10.1137/S0097539791218202
- Knuth, D. E. (1969). *The Art of Computer Programming* (Vol. 1). Reading, MA: Addison-Wesley.
- Lyngs, R. B. Zuker, M., & Pedersen, C.N.S. (1999). Internal loops in RNA secondary structure prediction. *Proceedings of the 3rd annual international conference on computational molecular biology (RECOMB)*, 260-267.
- Mannila, H., & Räihä, K.-J. (1990). On Query Languages for the p-string data model. In H. Kangassalo, S. Ohsuga, & H. Jaakola (Eds.), *Information Modelling and Knowledge Bases* (pp. 469–482). Amsterdam: IOS Press.
- Richter, T. (1996). A new algorithm for the ordered tree inclusion problem. *Lecture Notes of Computer Science: Vol. 1264. Proceedings of the 8th Annual Symposium on Combinatorial Pattern Matching (CPM)* (pp. 150-166). Springer.
- Rui, Y., Huang, T. S., & Mehrotra, S. (1999, September). Y. and Huang, T. S. and Mehrotra, S. (1999) Constructing table-of-content for videos. *ACM Multimedia Systems Journal*, 7(5), 359–368. doi:10.1007/s005300050138
- Zaki, M. (2002) Efficiently mining frequent trees in a forest. *Proc. of KDD*. doi:10.1145/775047.775058

ADDITIONAL READING

- Bruno, N., Koudas, N., & Srivastava, D. (2002) ‘Holistic twig joins: optimal XML pattern matching’, *Proc. SIGMOD Int. Conf. on Management of Data*, June, pp.310–321, Madison, Wisconsin. doi:10.1145/564691.564727
- Chen, S., Li, H.-G., Tatemura, J., Hsiung, W.-P., Agrawa, D., & Canda, K. S. (2006) Twig2Stack: bottom-up processing of generalized-tree-pattern queries over XML documents, *Proc. VLDB*, September, pp.283–294, Seoul, Korea.

Chen, Y., & Che, D. (2006). Efficient Processing of XML Tree Pattern Queries. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, 10(5), 738–743. doi:10.20965/jaciii.2006.p0738

Chen, Y., & Zou, L. (2011). Unordered tree matching: The evaluation of tree pattern queries, Int. J. Information Technology. *Communications and Convergence*, 1(3), 254–279.

Gou, G., & Chirkova, R. (2007) Efficient algorithms for evaluating XPath over streams, *Proc. SIGMOD*, 12–14 June. doi:10.1145/1247480.1247512

KEY TERMS AND DEFINITIONS

Bottom-Up Tree Search: A systematic way to explore a tree, by which a node will be visited after its children are visited.

Interval: A pair of integers $[a, b]$ where b denotes the rank of v in a post-order traversal of a trie. Here the ranks are assumed to begin with 1, and all the children of a node are assumed to

be ordered and fixed during the traversal. In addition, a denotes the lowest rank for any node u in the subtree rooted at v . Its purpose is to check the reachability. Let $[a, b]$ and $[c, d]$ be the intervals associated with v and u , respectively. If $[c, d] \subset [a, b]$, then u is reachable from v through a path in the tree.

Ordered Tree Inclusion: Mapping a pattern tree into a target tree, by which the labels and ancestor/descendant relationship, as well as left-to-right ordering of nodes should be preserved.

Top-Down Tree Search: A systematic way to explore a tree, by which a node will be visited before its children are visited.

Tree Matching: Mapping a pattern tree into target tree, by which the labels and parent/child relationship of nodes should be preserved.

Twig Matching: A kind of relaxed unordered tree inclusion, by which more than one node in a pattern can be mapped to a single node in a target as long as labels and ancestor/descendant relationship of nodes are preserved.

Unordered Tree Inclusion: A tree inclusion, by which the ordering of siblings is not important.

Information Seeking Models in the Digital Age



Mudasir Khazer Rather
University of Kashmir, India

Shabir Ahmad Ganaie
University of Kashmir, India

INTRODUCTION

Information Seeking Behavior of students in higher academic institutions is an exhaustive and complex process. The theoretical representation of such process is very difficult to comprehend, thus the graphical or pictorial representation will make things easier to understand and the same principle forms the base of Information Seeking Models. A model may be defined as a structure for thinking about a perceived problem and may evolve into a statement of the relationships among theoretical propositions. Information seeking models diagrammatically represent the complex tasks of information seeking process. Most Information Seeking Behavior models are generally the statements, often in the form of diagrams that attempt to explicate an information-seeking activity, the causes and consequences of that activity, or the relations among stages in information Seeking Behavior (Wilson, 1999). Information seeking models aim to describe the process that a user follows to satisfy his information need and while fulfilling that need, he approaches towards formal and informal information sources or available services which finally results in success or failure to retrieve desired information. A number of models have been designed by various authors and researchers from time to time globally relevant to information needs and seeking behavior of users in various academic institutions. Some models also highlight major as well as minor factors that

may directly or indirectly influence the Information Seeking Behavior of users.

BACKGROUND

Most models of information behavior are generally the statements, often in the form of diagrams that attempt to explicate an information-seeking activity, the causes and consequences of that activity, or the relations among stages in information-seeking behavior. Behavior may be defined as the more general field of investigation with information-seeking Behavior being seen as a sub-set of the field, particularly concerned with the variety of methods people employ to discover, and gain access to information resources, and information searching Behavior being defined as a sub-set of information-seeking, particularly concerned with the interactions between information user (with or without an intermediary) and computer-based information systems.. The first model for study of Information Seeking Behavior was proposed by James Krikelas in 1983. This model suggests that the steps of information seeking process are as follows:

1. Perceiving a need,
2. The search,
3. Finding the information, and
4. Using the information which results in either satisfaction or dissatisfaction.

Over the period of four decades a number of information seeking behavior models have been propounded by many researchers globally (Sawant, 2015). Robson and Robinson (2015) reveal that *Model* presents practical vision into the information seeking behavior of users and the factors that influence them. A variety of models like that of Kuhlthau, Dervin, Wilson, Ellis etc. describe the information seeking process of researchers and students thereby highlight important activities, services, actions and issues related to their information search (Infomatters, 2006). Bates (2005) reveals that *Models* are most useful at the description and prediction phases of understanding a process. Proper explanation of a phenomena results in a 'Theory'. It is believed that most of the theories in LIS are still at the modeling stage. Models are of great importance in the improvement of theory. They are a kind of proto-theory, a tentative anticipation set of associations, which can then be validated by means of various tests. McKenzie (2003) found that many information seeking models are limited in their ability to explain everyday life information seeking. These are generally related to the studies of scholars or professionals and some have been designed using a *cognitive approach* to model building. However, Robson and Robinson (2013) state that the existing models have some elements in common and most of these in the field of Library and Information Science focus on information seeking behavior of users. Nkomo (2009) divulges that scholars within Library and Information Science as well as outside the field have designed several information seeking models to sketch the information seeking behavior of researchers and students. It can therefore be supposed that the models somehow map out the development of information seeking and number of such models have been designed like *Ellis 1993 model*, *Kuhlthau's 1992 model* etc. These models have been applied by researchers to examine the information seeking behavior of various users in different academic institutions of the world (Kakai, Odongo & Bukenya, 2004).

Information Seeking Models

A number of models have been designed from time to time globally by various authors and researchers relevant to information needs and seeking behavior of users in various academic institutions. Some of the famous models of information seeking behavior are discussed as:

Models Proposed by T.D. Wilson

T.D. Wilson has put forth a series of models of Information Seeking Behavior including Wilson Model (1981), Wilson Model (1996) Wilson Model (1999) etc.

Wilson's Model (1981)

Wilson propounded an Information Seeking Behavior model in 1981 (Figure 1) in which he suggested two things:

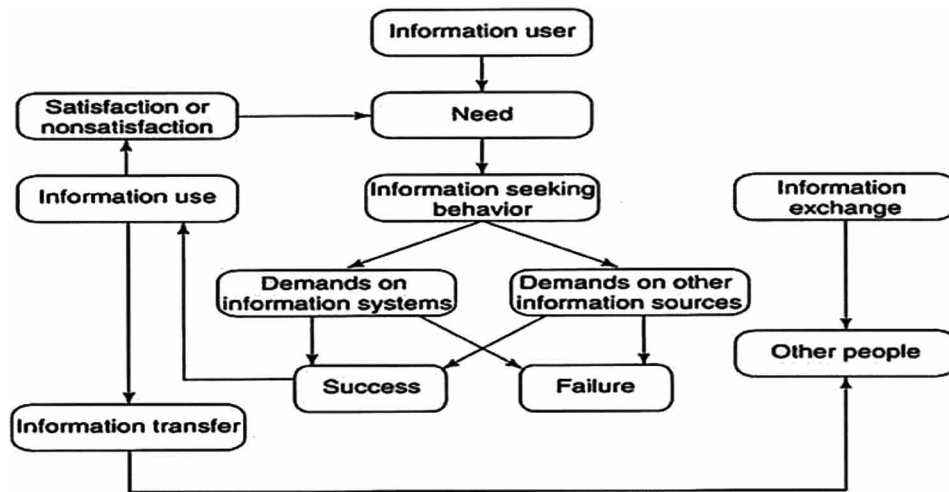
- Information need is not a primary need rather it is a secondary need that arises out of needs of a more basic kind.
- While satisfying his information needs the user come across various problems or barriers.

Drawing upon definitions in psychology, Wilson states that the basic needs can be defined as physiological, cognitive or affective. He further reveals that the needs arise as a result of personal life issues or the environments (political, economic, technological, etc.) within which that life or work takes place. He then suggests that the barriers that hamper the information search arise as a result of satisfying those needs (Wilson, 1981).

Wilson's Model (1996)

Wilson proposed another model in 1996 as a modification to his earlier model of 1981 which presents a cycle of information activities from the rise of information need to the information use of

Figure 1. Wilson's 1981 model of information-seeking behavior (Wilson, 1981)



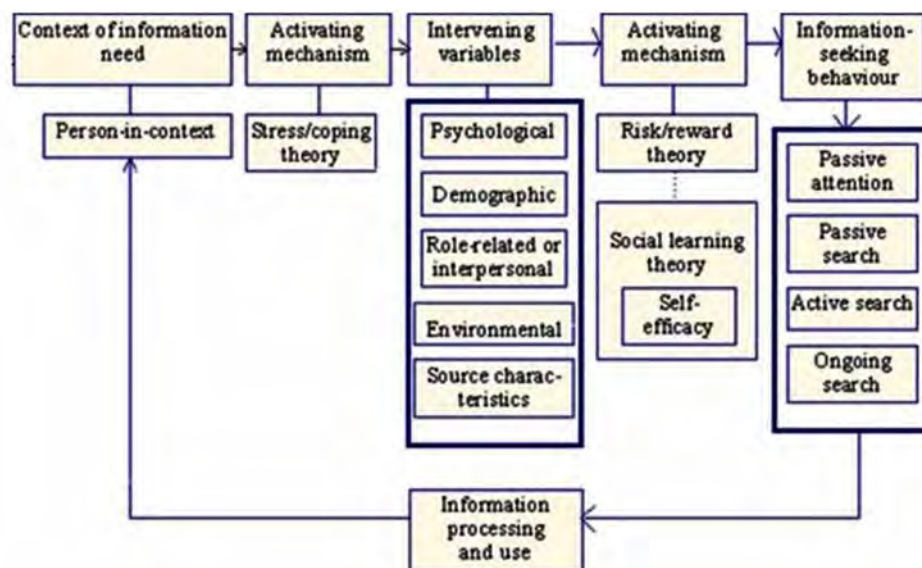
an information seeker. It also shows the intervening variables like (Demographical, Psychological etc.) which play a pivotal role in influencing the Information Seeking Behavior of the user. According to Wilson, it is not necessary that every need of the user lead to Information Seeking Behavior. The model identifies not only impending personal variables and modes of seeking information but also highlight relevant theories (Social learning

theory, Risk/reward theory) of motivations behind the user's search behavior.

Wilson's Model (1999)

Another, information seeking model (Figure 3) proposed by Wilson was again on 1999 which stresses on the intricacies of context of information seeking process like that of Dervin's sense making metaphor. This model of Wilson is com-

Figure 2. Wilson's (1996) model of information seeking behavior



monly known as the Macro-model. This model comprises of three important aspects viz:

- Why information seeking is more likely to occur in response to some needs more than others.
- Why some information sources are more used by the users than others.
- Why user's opinions of their own competence influences their success in meeting an information goal.

According to Wilson, as feedback of user is necessary aspect of overall information seeking process so in his model, he stresses on information process and invokes a feedback cycle in which he highlights *information seeking* as *iterative* at many steps rather than successive (Wilson, 1999).

Model Proposed by Brenda Dervin

Brenda Dervin is one of the famous Psychologists who proposed a well-known model on the cognitive dimension of Information Seeking Behavior of users. Her model (Figure 4) comprises of three important steps i.e. Situation, Gap and Use. In the model, she describes that human's progress through space and time (Situation) whereby they come across a cognitive gap which results in the creation of information need. According to her, such gaps are bridged by using the new information prior to proceeding further (Dervin, 1992).

She further extended it in the form of an important theory known as *Dervin's Sense-making Theory*.

Dervin's Sense Making Theory of 1996

Although, Brenda Dervin developed the Sense-making theory after years of efforts together and proposed it in 1996 in the form of a diagrammatic sketch as given in Figure 5 but she called it as theory rather than a model as she states it as "*a set of assumptions, a theoretic perspective, a methodological approach, a set of research methods, and a practice*" designed to handle with information perceived". She focuses on four significant aspects of user's Information Seeking Behavior viz:

- A Situation (Experience, background etc.) in space & time which determines the circumstances in which information problems arise.
- A Gap (Doubt, ambiguity etc.) which examines the difference between the contextual situation and the desired situation (e.g. ambiguity).
- An Outcome (Consequences, impact etc.) that is the result of the Sense-Making process.
- A Bridge (Ideas, Thoughts etc.) which indicates the means and the techniques used for closing the gap between situation and outcome (as cited in Wilson, 1999).

Figure 3. Wilson's model (1999) of Information Seeking Behavior (Problem Solving Model)

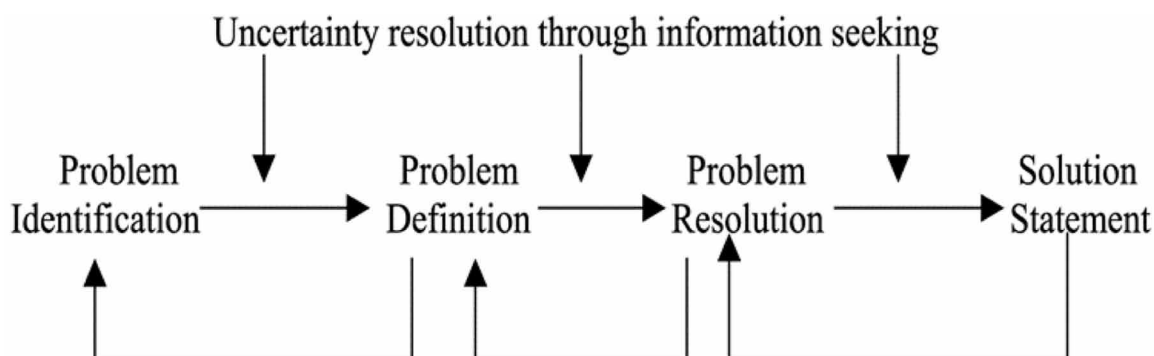
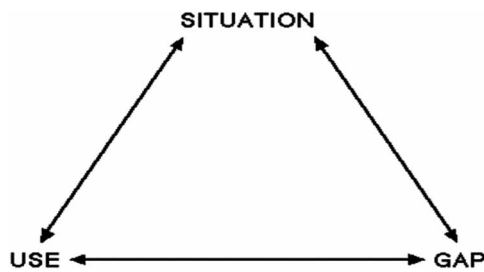


Figure 4. Dervin's Model



Model proposed by Carol Kuhlthau

Carol Kuhlthau's Model (1991)

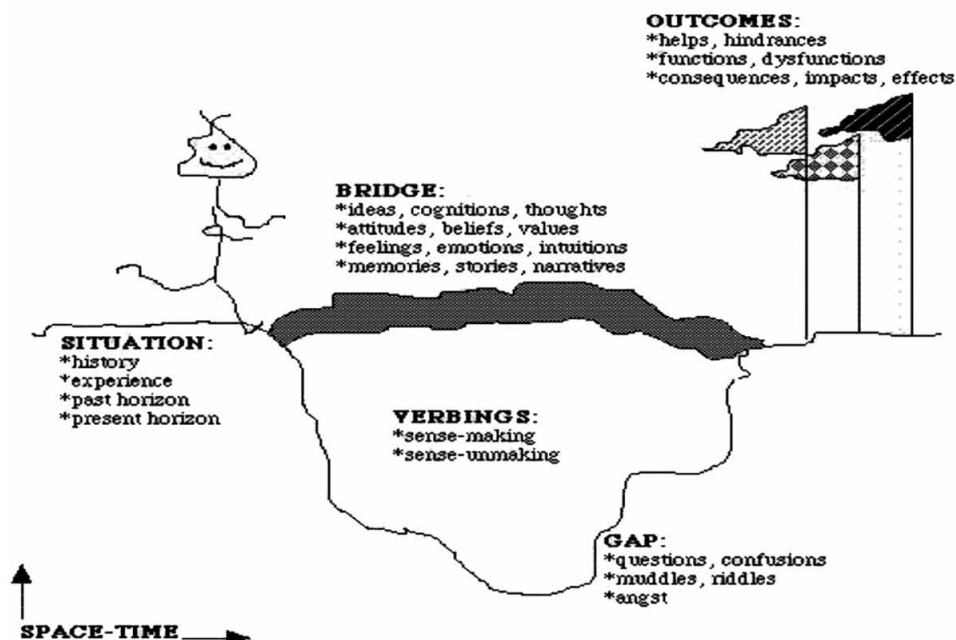
Carol Kuhlthau propounded an Information Search Model (ISM) in 1991 (Figure 6) and is based on Kelly's Theory who believes that learning is a process of testing constructs. According to Kuhlthau, reduction of ambiguity and uncertainty is the basic motivator of research. She divided information seeking process into different stages. She relates each of these stages with the emotional state of researcher. For example, she relates *Anxiety* with the recognition of uncertainty at the first stage i.e. *Initiation*. The other five stages and related emo-

tional states (listed in parenthesis) are: Selection (*optimism*), Exploration (*confusion/frustration/doubt*), Formulation (*clarity*), Collection (*confidence*) and Presentation (*relief/satisfaction or disappointment*) (Kuhlthau, 1991).

Kuhlthau 'Six stages' are discussed below:

1. **Initiation:** This is the first stage of the process; expressed by feelings of uncertainty and more general thoughts with a need to be aware of or connect new to existing knowledge.
2. **Selection:** In this stage, selection of a general topic with some common feelings of hopefulness takes place in order to identify most significant areas of the topic.
3. **Exploration:** It is the third step of the process in which scrutinizing to extend personal perception and decrease the feelings of doubt and confusion about the topic takes place.
4. **Formulation:** Here the researcher focuses on the process by means of encountered information accompanied by feelings of increase in confidence.

Figure 5. Dervin's sense-making Theory (Wilson, 1999)



5. **Collection:** In the fifth stage, the researcher interacts smoothly with the information system with confidence as the topic is analyzed by selecting and evaluating information.
6. **Presentation/Closure:** It is the final stage of the process which ends the process with a sense of confidence or failure depending on how useful are the findings (As cited in Baro, Onyenania & Osaheni, 2010).

Carol Kuhlthau's Model (1992)

Kuhlthau (1992) developed another model of the ISP as a modification to her earlier model of 1991. Although, her earlier model included the feelings of the researcher with the different stages of search process but this model also encompasses the development of thoughts about a research topic, and the actions of seeking and using sources. The model goes beyond the mere mechanics of information seeking; it incorporates three realms: *the affective (feelings), the cognitive (thoughts), and the physical (actions and strategies)*. These realms are common to each stage of the search process, as described below (Figure 7).

Kuhlthau's model suggests that user is an active participant in information search process. According to her the student's knowledge grows as he interacts with the information. He uses his

cognitive behavior and cognitive strategies such as brain storming, thinking, predicting, consulting, choosing, identifying, selecting, defining and confirming. Although, the model is silent about how to convert information and data into knowledge but it highlights the affective feelings of users such as nervousness, hesitation, uncertainty, confusion, anxiety, anticipation, doubt, optimism, and confidence in the search process. This model focuses on the search process and the activities associated with finding information rather than how to use, synthesize, and evaluate the information that is found.

Model Proposed by Ellis

Ellis propounded a model (Figure 8) on the basis of studies related to the Information Seeking Behavior of social scientists, research scholars and research scientists. His model also comprises of six categories of information seeking activities viz:

1. Starting
2. Chaining
3. Browsing
4. Differentiating
5. Monitoring
6. Extracting.

Figure 6. Kuhlthau's model of the information searching process (Kuhlthau, 1991)

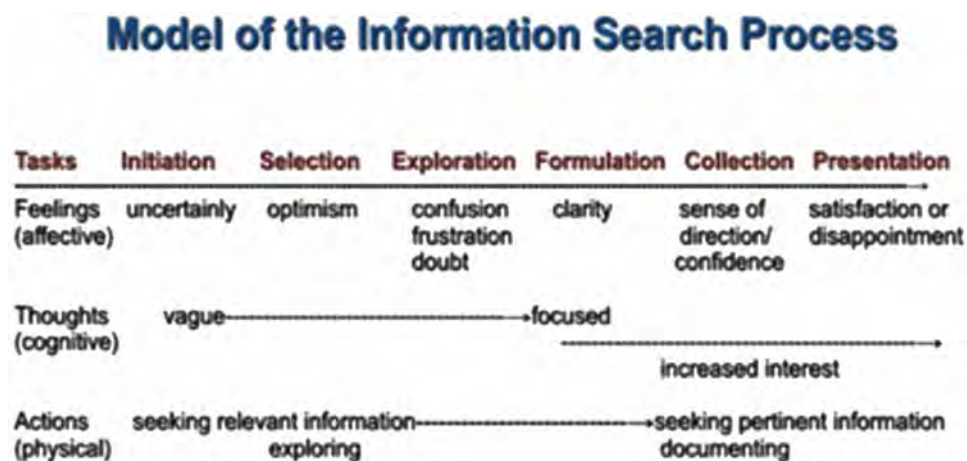


Figure 7. Kuhlthau's Model

Image source: <http://people.ualgary.ca/~ahayden/kuhlth.gif>

TASK	THOUGHTS	FEELINGS	ACTIONS	STRATEGIES
STAGE 1: TASK INITIATION	Contemplating assignment Considering possible topics	Apprehension Uncertainty	Talking with others Browsing library collection	Brainstorming Discussing
STAGE 2: TOPIC SELECTION	Consider personal interest, project requirements, info available, time	Confusion Anxiety Anticipation	Consulting with informal mediators Preliminary search of the library	Discussing possible topics Using general sources for overview of possible topics
STAGE 3: PREFOCUS EXPLORATION	Become informed about a general topic Seeking focus Identifying possible focuses Inability to express precise info needed	Confusion Doubt Threat Uncertainty	Locating relevant info Reading Taking notes Making bibliographic citations	Reading Intentionally seeking focus Listing descriptors
STAGE 4: FOCUS FORMULATION	Predicting outcome Using criteria of personal interest, project requirements, info available, time Identify ideas	Optimism Confidence in ability to complete task	Reading notes for themes	Making a survey of notes Listing focus Choosing focus Combining focus
STAGE 5: INFORMATION COLLECTION	Seeking info to support focus Defining and extending focus Gathering pertinent info Organizing info	Realization of amount of work to be done Confidence in ability to complete task Increased interest	Using the library Requesting specific sources Taking detailed notes	Using descriptors to find pertinent info Comprehensive searches Using indexes Requesting assistance
STAGE 6: SEARCH CLOSURE	Identifying need for additional info Consider time restrictions Diminishing relevance Exhausting resources	Sense of relief Sometimes satisfaction Sometimes disappointment	Rechecking sources Confirming info	Returning to library to make summary search Rechecking info

These activities are discussed as:

1. **Starting:** It is the first step and it consists of those activities that user carries out at the beginning of searching needed information. Here, the user selects the sources of information that are relevant to his information needs. He may be familiar with the identified sources that have been used before or unfamiliar about these depending upon his experience and knowledge. These sources also help him in finding more relevant sources by means of pointing to, suggest, or recommended additional sources or references.
2. **Chaining:** It is the second step of information search which begins after starting stage. Chaining can be either backward or forward. Backward chaining is generally carried out by researchers and scientists. Here, references from an initial source are followed. While as, forward chaining identifies and follows up on other sources that refer to an initial source or document. Although it can be an effective way of broadening a search, forward chaining is much less commonly used (Al-Muomen, Morris & Maynard, 2011).
3. **Browsing:** In this step, the user starts browsing his desired information by looking tables of contents, lists of titles, subject headings, names of organizations or persons, abstracts, conclusions and summaries etc. Here, the user browses information relevant to his needs. He uses formal as well as informal sources.
4. **Differentiating:** In the fourth step, user filters and scans the collected information by noticing the differences between the nature and quality of information offered by various sources. Experience and previous knowledge plays a significant role in this process.
5. **Monitoring:** This step deals with the activity of keeping oneself abreast with the latest developments in his specific field of study by regularly following particular sources of information. Here, the user keeps track on the Core sources which may furnish latest and relevant information to his needs. Core sources may include professional or

researcher groups, experts, authentic publications etc.

6. **Extracting:** It is the final step of Ellis Model in which the user systematically extracts the desired information from the relevant sources. The user extracts information either directly from the sources or indirectly by searching through online indexes, databases, bibliographies etc (Choo, Detlor & Turnbull, 1999).

According to Ellis, each step in his model is very important and each of these varies in according to the environment in which the user satisfies his information need. But no matter what the circumstances are, the process must start with the Starting feature, and end with the Ending. Ellis model is similar to Wilson's model to a good extent.

Comparison Between Ellis and Kuhlthau's Model

Both the Ellis as well as Kuhlthau's models discuss the overall process of Information Seeking Behavior of an individual while searching his desired information. Wilson (1999) combined the two models (Figure 9) together into one model

to easily mark the differences and similarities between the views of the two authors.

Wilson (1999) is of the opinion that though both the models have many similarities but there occurs many differences as well. According to him, Ellis model has the flexibility in its application to various situations as he states that the information behavior characteristics of an individual vary in accordance to the circumstances in which he seeks his desired information which is not seen in case of Kuhlthau's models as she has put different stages in the model on the basis of her analysis of information Behavior. So, the two models oppose. Ellis's model is more authentic compared to Kuhlthau's as it is based on experimental research and has been tested in a number of studies. One more important difference between the two models is that Kuhlthau's model includes feelings and thoughts of an individual while seeking necessary information and the way in which these thoughts and feelings gradually change as the process proceeds while as Ellis's model focuses on the logical steps that the individual takes to obtain useful information.

Thus, it can be concluded that the two models oppose in a way and match each other in another with each model examining the information seeking process from a different angle.

Figure 8. Ellis Model

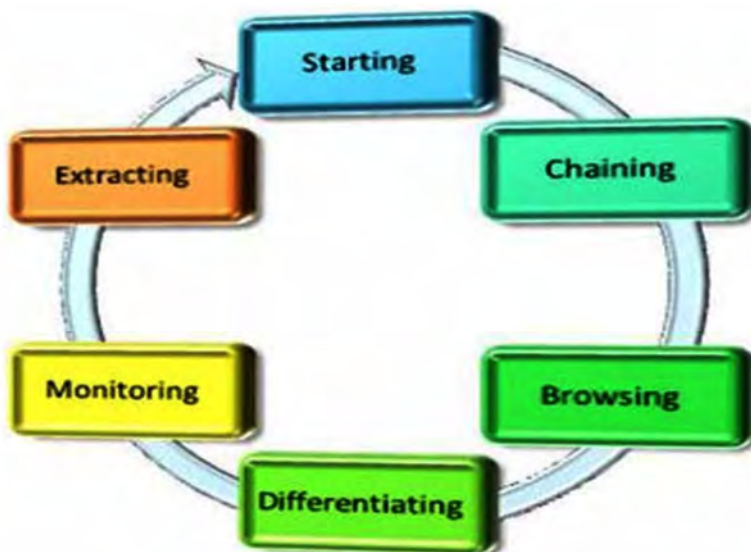
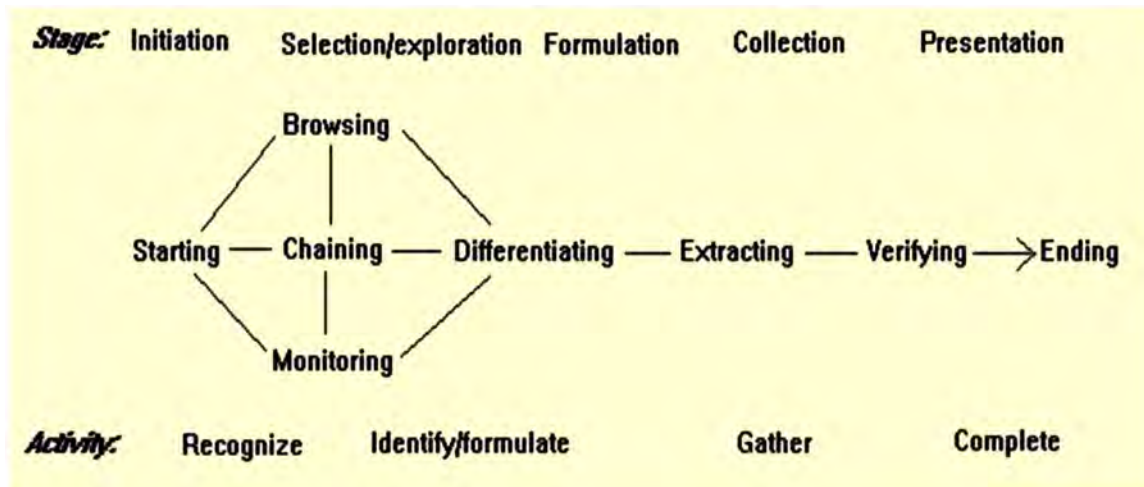


Figure 9. Comparison of Ellis' and Kuhlthau's model (Wilson, 1999)



Model Proposed by Eisenberg and Berkowitz (1992)

Another model of Information Seeking Behavior was proposed by Eisenberg and Berkowitz in 1992 popularly known as Big Six Model (Figure 10) which closely resembles to that of Kuhlthau's model. This model again comprises of six logical steps and represents general approach to information problem-solving. The sequence of steps changes with different search ventures but each step is necessary for the successful resolution of an information problem.

The steps of *Big Six Skills* model are discussed as:

1. **Task Definition:** This is the first step of Big Six model. Here, an individual defines the problem or the topic about which he wants to seek information. He defines what he needs to do, what kind of information he needs to collect etc before rushing on to the strategies to retrieve information. According to Eisenberg and Berkowitz, most researchers spend very less time in understanding and defining their topic. They prefer to proceed to the next steps of information search process rather than reflecting on the type of information that they need to find. They further add
2. **Information Seeking Strategies:** This step helps in decision making i.e. the decision of selecting most appropriate and relevant sources of information among a plethora of others to get needed information. This step focuses on the methodology or the strategy to be followed to acquire the objectives of topic in hand.
3. **Location and Access:** This is one of the most important steps of Big Six Model which involves implementation of the information Seeking Strategy. It focuses on inculcating skills for using information access tools, organization of information sources in libraries, making aware about various parts of a book, use of search tools and search strategies in databases etc. Although, traditional libraries used to teach specific skills needed to use various tools to its users but the problem with teaching those skills is that users lack the understanding of how these skills can be implemented effectively in other new situations. As the Big Six Skills Model focuses on problem-solving approach so it

provides the solution by means of focusing on overall information problem-solving process rather than teaching them the skills that are restrained to a particular situation. By teaching a general problem-solving approach, the students are better equipped to utilize new and unique sources of information

4. **Use of Information:** In this step, students use information in more sophisticated and systematic way. They use skills like interacting, dialoguing, scanning, questioning, and reflecting on the information. They are able to decide what kind of information is more valuable and extract that to satisfy their information need. In Kuhlthau's Model, stage 3 and 5 are similar to as Big Six Model's step 2, 3 and 4. These all focus on finding and locating information and then using that information to satisfy desired information.
5. **Synthesis:** In this step, the extracted information is analyzed, restructured, repacked and applied in new and different form. It is sometimes applied directly like writing formulas but sometime it is also used in documenting the research articles which is a complex process and in that case it involves combining related information, reorganizing

and manipulating the gathered information. Synthesis is simply converting the collected information into knowledge.

6. **Evaluation:** It is the last step of the model in which analysis, assessment and examination of the whole information seeking process is carried out at the end. It determines the usefulness, efficiency and effectiveness of the process. This step examines whether the information found at the end related to the problem is providing real solution to the problem, satisfies the specified objectives of the topic or meets the desired information needs of researchers or the information seeker. Self-evaluation leads to lifelong information literacy of students to know their strengths and weaknesses to complete information seeking processes and can help them to take right decisions in solving such problem.

Model Proposed by Urquhart and Rowley (2007)

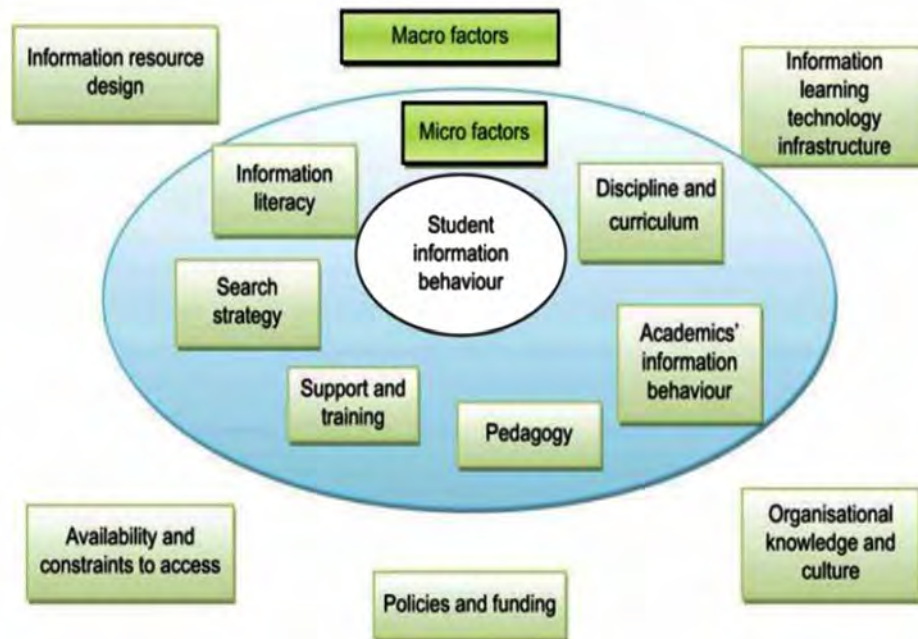
Urquhart and Rowley (2007) proposed a model (Figure 11) for representing the Information Seeking Behavior of students especially while accessing

Figure 10. Big Six Skills Model

Image source: http://www.summit.k12.co.us/cms/lib04/CO01001195/Centricity/Domain/606/big_six.png



Figure 11. Urquhart and Rowley Model



electronic resources of information. In their model they have highlighted *macro and micro factors* that influence the Information Seeking Behavior of students. According to Urquhart and Rowley, the *micro factors* are the factors that are related to the personality of student which includes search strategy, information literacy, pedagogy, discipline etc. The *macro factors* include information resource design, information learning technology infrastructure, organizational knowledge and culture, policies and funding etc (Urquhart and Rowley, 2007).

FUTURE RESEARCH DIRECTIONS

Models are diverse in nature including Information Seeking Behavior models and Information Search Models. The significance of these models lies in the fact that these help in understanding and examining the Information Seeking Behavior of researchers and prove beneficial to know appropriate methodologies and techniques to attain the objectives of a research problem. These help in better understanding of tasks, activities,

available services, infrastructure of universities and problems faced by users while searching for desired information thereby help administrators and higher authorities to take right decisions, offer best services, implement tech savvy infrastructure and satisfy the information needs of users especially scholars in more systematic and sophisticated way. Further, the models may also assist in examining the impact of technological advancements and information availability channels on the information seeking behavior of researchers. This study has a broader scope for future research. It can be extended to examine and highlight the information seeking behavior of Professors, Lecturers, Teachers, Post Graduate & under Graduate students associated with teaching and learning processes in different Colleges and Universities of the world. Further, research can be extended towards exploring information retrieval and information searching models.

One more area for future research is to examine the factors influencing the information seeking behavior of users in university libraries and design model for the same.

CONCLUSION AND SUGGESTION

Information seeking models diagrammatically represent the complex tasks of information seeking process. A number of information seeking models have been proposed by various authors like Wilson, Kuhlthau, Ellis etc. in different years. Most of the models comprise of many steps or stages. Models are diverse in nature including Information Seeking Behavior models and Information Search Models. The significance of these models lies in the fact that these help in understanding and examining the Information Seeking Behavior of researchers and prove beneficial to select best methodologies and techniques to attain the objectives of a research problem. These can help in better understanding of tasks, activities, available services, infrastructure of universities and problems faced by users while searching for desired information thereby help administrators and higher authorities to take right decisions, offer best services, implement tech savvy infrastructure and satisfy the information needs of users especially students and scholars in more systematic and sophisticated way. Further, it is also concluded that the models assist in comprehending all the aspects associated with the information seeking process that may be emotional, cognitive or practical. These form a foundation and provide a framework for novel researchers to design and implement new models of Information Seeking Behavior of users in their respective organizations or institutions to improve the services; facilitate authentic and relevant information resources and to eliminate all the problems and issues that users face while searching for desired information.

REFERENCES

- Al-Muomen, N., Morris, A., & Maynard, S. (2012). Modelling information-seeking behavior of graduate students at Kuwait University. *The Journal of Documentation*, 68(4), 430–459. doi:10.1108/00220411211239057
- Baro, E., Onyenania, G. O., & Osaheni, O. (2010). Information Seeking Behavior of undergraduate students in the humanities in three universities in Nigeria. *Program: Electronic Library and Information Systems*, 76(2), 109–117. doi:10.1108/00242531011047055
- Bates, M. J. (2005). An introduction to meta-theories, theories, and models. *Theories of Information Behavior Information Today*, 1(4). Retrieved from: <https://pages.gseis.ucla.edu/faculty/bates/articles/information-behavior.html>
- Choo, C. W., Detlor, B., & Turnbull, D. (1999). *Information Seeking on the Web--An Integrated Model of Browsing and Searching*. DOI: 10.1108/00242531011047055
- Dervin, B. (1992). From the mind's eye of the user The sense-making qualitative-quantitative methodology. In *Qualitative Research in Information Management* (pp. 68-70). Englewood, CO: Libraries Unlimited.
- Infomatters. (2006). *No more information seeking models please*. Retrieved from <http://sentra.ischool.utexas.edu/~adillon/blog/archives/34>
- Kakai, M., Odongo, R. I., & Bukenya, I. M. N. K. (2004). A study of the information seeking behavior of undergraduate students of Makerere University. *World libraries*, 14(1), 1–22.
- Kuhlthau, C. (1991). Inside the search process information seeking from the users perspective. *Journal of the American Society for Information Science*, 42(5), 361–371. doi:10.1002/(SICI)1097-4571(199106)42:5<361::AID-ASI6>3.0.CO;2-#
- McKenzie, P. J. (2003). A model of information practices in accounts of everyday-life information seeking. *The Journal of Documentation*, 59(1), 19–40. doi:10.1108/00220410310457993
- Nkomo, N. (2009). *A comparative analysis of the web information seeking behavior of students and staff at the University of Zululand and the Durban University of Technology* (Doctoral dissertation). Retrieved from: <http://hdl.handle.net/10530/389>

Robson, A., & Robinson, L. (2013). Building on models of information behavior Linking information seeking and communication. *The Journal of Documentation*, 69(2), 169–193. doi:10.1108/00220411311300039

Sawant, S. (2015). Krikelas' Model of Information Seeking Behavior (1983). In M. Al-Suqri, & A. Al-Aufi (Eds.), *Information Seeking Behavior and Technology Adoption: Theories and Trends* (pp. 82–93). Hershey, PA: Information Science Reference. Doi:10.4018/978-1-4666-8156-9.ch005

Urquhart, C., & Rowley, J. (2007). Understanding student information behavior in relation to electronic information services: Lessons from longitudinal monitoring and evaluation, part 2. *Journal of the American Society for Information Science and Technology*, 58(8), 1188–1197. doi:10.1002/asi.20562

Wilson, T. (1999). Models in information behavior research. *The Journal of Documentation*, 55(3), 249–270. doi:10.1108/EUM0000000007145

Wilson, T. D. (1981). On user studies and information needs. *The Journal of Documentation*, 37(1), 3–15. doi:10.1108/eb026702

Wilson, T. D. (1996). *Information behavior an interdisciplinary perspective*. Retrieved from: www.informationr.net/tdw/publ/infbehav/cont.html

ADDITIONAL READING

Adio, G., & Arinola, A. (2012). Information Needs and Information-Seeking Behavior of Agricultural Students AAT LAUTECH, Ogbomoso. *The official publication of the Pacific Northwest Library Association*, 76(3).

Al-Ansari, H. (2006). Internet use by the faculty members of Kuwait University. *The Electronic Library*, 24(6), 791–80. doi:10.1108/02640470610714224

Alazemi, T. R. (2015). *Users' Information Seeking Behaviors, their Interactions and Experience with the Academic Library Web Interface*. University of Salford. Retrieved from: <http://usir.salford.ac.uk/36705/1/Final%20Thesis.pdf>

Campbell, I. (2000). *The ostensive model of developing information-needs*. Glasgow University, Glasgo. Retrieved from: www.dcs.gla.ac.uk/~iain/papers/phd.pdf

Caroline, F. T., & Cees, A. W. (2010). Developing scales for information-seeking behavior. *The Journal of Documentation*, 66(1), 46–69. doi:10.1108/00220411011016362

Ganaie, S. A., & Khazer, M. (2015). Diversity of information sources in the digital age an overview. *Journal of Advancements in Library Sciences*, 2(2), 53–61.

Tanackovic, F. S., Horvatic, I. F., & Badurina, B. (2015). European Union information in an acceding country. *Library Hi Tech*, 33(1), 143–158. doi:10.1108/LHT-10-2014-0103

KEY TERMS AND DEFINITIONS

Information Behavior: Information behavior refers to how people approach and handle information.

Information Literacy: The ability to know when there is a need for information, to be able to identify, locate, evaluate, and effectively use that information for the issue or problem at hand.

Information Need: The perception of a lack of information that provokes one to develop a need for it.

Information Seeking Behavior: A special case of problem solving which, includes recognizing and interpreting the information problem, establishing a plan of search, conducting the search, evaluating the results, and if necessary, iterating through the process again.

Model: A model may be defined as a structure for thinking about a perceived problem and may evolve into a statement of the relationships among theoretical propositions.

An Insight Into Deep Learning Architectures

Nishu Garg

VIT University, India

Nikhitha P

VIT University, India

B. K. Tripathy

VIT University, India

INTRODUCTION

For molding the world to incredible amplitude, appropriate learning of computers has inevitable participation. This ability of the computer of deploying the world can be referred as intelligence. For making a machine intelligent, it has to go through a series of training processes where the information is stored in an organized way making it easier to relate the so gathered information with the real life scenarios. Storing all the information in the machine manually is a troublesome task, especially when the information is abundant like in the case of sophisticated artificial intelligence tasks. This is why the learning algorithms have gained attention among researches to store huge information at a stretch. Many learning algorithms are established and capable of understanding the view, but failed to express efficiently in natural language. Semantic understanding of these algorithms was restricted by some degree which is the necessity and should be expanded. So it was figured out that the prior algorithms were incapable of maintaining the interaction towards the humans through various semantic and visual mediums. To resolve this problem of AI task, Deep architectures were introduced, so that the machines can be trained and can be made efficient in the areas they are lacking.

Information retrieval have widely classified into two subdivisions: image retrieval and the text retrieval (A.G. Abby, 2000). Both approaches were emerged in the mid 60-70's. In order to carry out the several tasks like indexing, searching and retrieval data has to be well organized. In this work, the focus is on the image retrieval perspective. Image retrieval is divided into query or text based and content based approach (P.A. Vikhar, 2010). Our investigation is centralized around the first methodology. Text based image retrieval is basic and fundamental one which makes use of searching through simple query word (T. Westerveld., J.C Gemert., R. Cornacchia., D Hiemstra., A. P. Vries 2005). Earlier work to this text based approach was database management systems which binds the images with text. Multi-dimensional indexing, data modeling and query evaluation are the exemplary research lines. But the fore mentioned procedures turned out ineffectual for the complex data link structures, language pairs, sophisticated image data characteristics, question answering, annotation and context retrieval. These made a way to design a new approach that addresses the various issues.

The basic question found out while natural language processing is "Is there exists any analog program loaded in computer which converts a portion of textual data represented in English into a

computer understandable data model that clearly depict the actual text meaning?”. While taking care of this problem, computer specialists have to get down for fewer objectives that represent lesser aspects of text data. Natural language processing is a flexible architecture which focuses on learning representations useful for the specific tasks. The NLP had evolved in a way for seeking a model which converts human understandable language (English in the beginning) into a machine friendly data representation that conveys the meaning of text without any ambiguity. The earlier times the evolutionary idea behind the NLP seemed to be very broad and vast for the researchers so they settled down for lesser goals of representing limited textual information by extracting simpler representations. Text processing in natural language processing encompasses by bag-of-words, removal of frequent words from the input text query (stop words), identifying the root words (stemming), spell check, text and controlled vocabulary mapping and named entity recognition.

BACKGROUND

The concept of neural networks derived from the human nervous system by the way of answering to a question which seems to be simple yet more complex while approaching “can a computer mimic the human nervous system?”. Human brain is highly complex nonlinear parallel computer. Structural constituents of human brain are known as neurons which are interconnected in complex way (Lee, T.S., Mumford, D.2003). The task of recognition through a neural network achieved by teaching the computer what to do in each and every step with a set of pre-defined procedures that is it is more like if a computer has to identify a photograph first it has to be taught by feeding various photographs then it go for a matching with the original photograph with the ones which it stores in the large database repository. It gives out the picture which carries a best resemblance with original one. While answering to the ques-

tion another constraint researchers faced is human brain processing speed which is of 10^{-3} s where as a silicon IC response time is 10^{-9} s.

The evolution of the deep machine learning starts from convolutional neural networks, deep neural networks and deep belief networks eras. Convolutional neural networks are a kind of special multi- layer neural networks which are inspired from biology. It is similar to other neural networks and they are trained with a version of back propagation algorithm but only differ in the architecture. CNN are mainly designed to recognize visual patterns directly from pixel images with minimal preprocessing. The hurdles faced by feed forward neural networks towards image processing could be resolved by the entry of CNN to some extent. The major constraint that normal FFNN facing was the rapidly growing weights with the dimensionality of the input, since every object in one unit is connected to every object in next unit. That makes it slow in learning for high dimensional domain of vision. Learning to recognize an object wouldn't even propagate the same object presented in different visual field, it's because the every pair of neurons between different units has separate weights which would be involved in the calculation. A CNN comprises of various convolutional and sub sampling layers alternatively took after by completely joined layers. The input to a convolutional layer is an $m \times m \times r$ picture where m is the tallness and width of the picture and r is the quantity of channels, e.g. a RGB picture has $r=3$. The convolutional layer will have k layers (or portions) of size $n \times n \times q$ where n is littler than the measurement of the picture and q can either be the same as the quantity of layers or littler and may shift for every part. The measure of the channels offers ascend to the generally joined structure which are each convolved with the picture to create k highlight maps of size $m-n+1$. Every guide is then sub sampled ordinarily with mean or max pooling over $p \times p$ adjoining districts where p ranges between 2 for little pictures (e.g. MNIST) and is normally not more than 5 for bigger inputs. Either before or



after the sub sampling layer an added substance predisposition and sigmoidal nonlinearity is connected to every component map.

As a solution to this architecture, conventional neural networks came, which takes the input from two dimensional spaces and should reduce the number of parameters involved in the training. Convolutional neural networks can be best explained with an experiment made in cat's visual cortex. Vann leclun and toshua bengio are two scientists who tried to understand the organization of visual cortex of cat which has complex cell arrangement. These cells are called receptive field which is very sensitive to small regions of input space and it is tiled in a fashion that would cover the entire visual field. These cells are like filters which are local to input space and it helps in the strong local correlation of the input image. In addition to that cells has been divided into two, simple cells which is used to respond to edge like stimulus patterns and complex cells which have larger receptive fields, invariant to the exact position of stimulus. Till now visual cortex is considered as the most powerful vision system.

Generally images are composed of large number of variables called pixels. A fully connected network with hundreds of hidden layers will have several tens of thousands of weights. This requires larger training set and high memory requirements which rule out certain hardware implementation. The basic problems with these unstructured nets are they don't have a built in invariance to the distortions and translations for images on speech inputs. Moreover, the input to the fixed size neural network should be size normalized and centered in the input field. The input can be character images, other 2-D, 1-D signals but there is no such preprocessing is perfect. As an example, hand written characters which can be normalized at word level causing variations with respect to size, slant and position of individual characters. Apart from that it also differs according to the writing style. These combined variations cause variations in the position of distinctive features in input objects.

In general, a fully connected network with fixed size have the capability to produce output by learning the inputs those are invariant with respect to such variations where as convolutional neural network are invariant to shift operation. Another drawback of a fully connected neural network is that it doesn't give importance to the local structure of the input. For e.g., images which have strong 2-D topology where nearby pixels are highly correlated. In CNN, local extraction of features is done by making the receptive field of hidden layers to be local.

MAIN FOCUS OF THE ARTICLE

Issues, Controversies, Problems

Indeed, the shortcomings of CNN paved a way for newer deep neural network architecture. For learning the complex functions equipped with high level abstractions which is employed in various AI task, we requisite deep architectures (B.Yoshua 2009). Deep architecture can depict the multiple levels of nonlinear data. After that the era of deep neural networks came. As the image retrieval being one of the tedious tasks the learning from an image makes its complexity a one level higher. This is where deep architecture methodology comes into existence. Prior to image retrieval an image has to be learned since learning phase comprises of various feature learning activities. Deep neural network is one such technique which involves learning from low level features of the image to more detailed high level features.

These features are deeply influenced by some factors which every image possesses such as orientation, lightning, position. These can be referred as the factors of variation. Explicit knowledge about this factors and making a relationship with the interaction patterns of these factor make the deep neural network to recreate the scene. In structural point of view deep neural network consists of multiple numbers of deep layers having nonlinear

operations. Each deep layer itself comprises of more abstracted operations. This series of hidden layers placed between the input layer and output layer enables sharing and reusing of information by applying training over these layers. Here training and learning comes along the two processes are very different. The training function is the overall algorithm that is used to train the neural network to recognize a certain input and map it to an output. A common example is back propagation and its many variations and weight/bias training. A learning function deals with individual weights and thresholds and decides how those would be manipulated. These usually (but not always) employ some form of gradient descent. The more the number of such intermediate layers the more features can be learned. In our proposed work we mainly focused our attention towards deep neural network for learning from the features at various layers. The ability of deep neural network to employ learning at each and every layer makes our model much intelligent.

Feature learning is the one of the classical problems in machine learning. In our model we utilized the idea of feature fusion from the different latent features. In contrast to the learning from features the designed method supports latent feature learning. Deep neural network strengthen the task of learning by applying the feature fusion from low level features to higher level features. DNN and DBN are two methodologies that well address the deep learning problem. The approach we are designed can be extended to any domain where feature learning and natural language processing becomes a vital part. A good citation to this can be social media where the data possess diverse, heterogeneity, noise and interconnectedness. Also DNNs relates to over fitting as the result of addition of layers of abstraction, which allow them to project rare dependencies in the training data. This led the way to DBN's evolutions.

Deep Belief Network

The other type of deep neural network is deep belief network (G.E Hinton, 2006). It includes unsupervised layer wise pre training similarly like in a deep feed forward network. DBN is unsupervised generative model which mix the directed and undirected interactions between the variables that consist of input layer or hidden layer. The top two layers always form a Restricted Boltzmann machine with undirected interactions and other layers form the directed interactions with Bayesian network. The DBN is not a feed forward neural network.

$$P(x, h^1, \dots, h^l) = P(h^l) \left(\prod_{k=1}^{l-1} P(h^k | h^{k+1}) \right) P(x | h^1) \quad (1)$$

In 1 it depicts the architecture for the deep belief network. The above equation has 1 layer with joint distribution between joint hidden layers h^k and vector x . $P(h^{l-1}, h^l)$ is the joint distribution in the top layer of restricted Boltzmann machine, here x corresponds to lowest level of the hidden layer that is h_0 and $P(h^{k-1} | h^k)$ is the visible joint distribution associated with Restricted Boltzmann machine at the level k of the DBN $P(h^k | h^{k+1})$ forms the Restricted Boltzmann machine.

We can show in 1 a generalize model with consist of three hidden layer. The full distribution of the deep belief network with three hidden layers is given below.

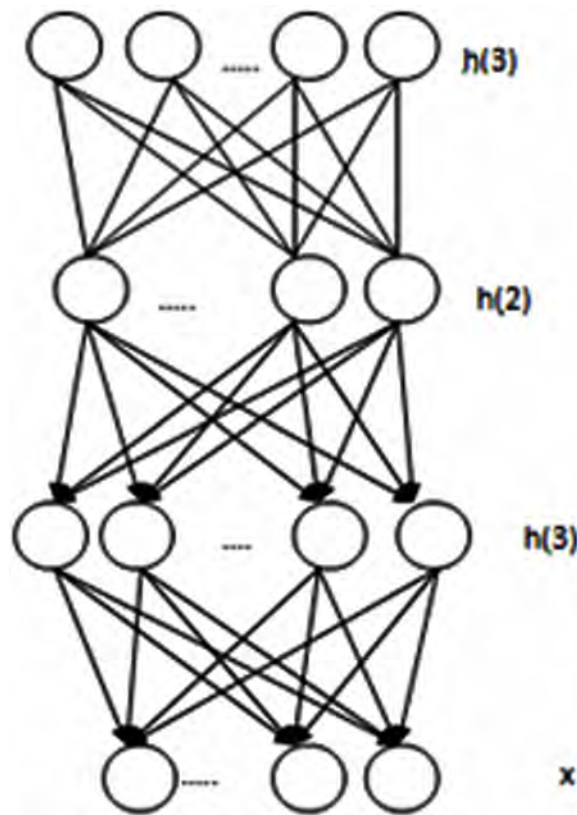
$$P(x, h^{(1)}, h^{(2)}, h^{(3)}) = P(h^{(2)}, h^{(3)}) P(h^{(1)} | h^{(2)}) P(x | h^{(1)}) \quad (2)$$

$$P(h^{(2)}, h^{(3)}) = \exp \left(h^{(2)\top} W^{(3)} h^{(3)} + b^{(2)\top} h^{(2)} + b^{(3)\top} h^{(3)} \right) / Z \quad (2.1)$$

$$P(h^{(1)}, h^{(2)}) = \pi_j P(h_j^{(1)}, h_j^{(2)}) \quad (2.2)$$

$$P(x, h^{(1)}) = \pi_i P(x_i, h_i^{(1)}) \quad (2.3)$$

Figure 1. Deep Belief Network



In 2, $P(x, h^{(1)}, h^{(2)}, h^{(3)})$ is the joint distribution over the input layer (x) and the hidden layers are (h^1, h^2, h^3). The distribution over $P(h^{(2)}, h^{(3)})$ is given by 2.1. here $P(h^{(1)} | h^{(2)})$ is the distribution of $h^{(1)}$ given by $h^{(2)}$, it doesn't depend on $h^{(3)}$, same with the other hidden and input layers. $W^{(3)}$ and $b^{(3)}$ is the weight and bias of the nodes in the third layer. Now in this kind of the network the training procedure starts with the pertaining of the top layer that is the restricted Boltzmann machine and then training the layer preceding it, this idea of initialization has formed the stacking mechanism in the DBN. A good initialization can play a crucial role in training the DBN. Otherwise training of deep belief network is the hardest.

Restricted Boltzmann Machine

The new type of neural network for unsupervised learning is Restricted Boltzmann machine (Le Roux, N., Bengio, Y, 2008). A restricted/layered Boltzmann machine (RBM) has either bit or scalar node values, an array for each layer, and between those are scalar values potentially for every pair of nodes one from each layer and an adjacent layer. The unsupervised learning only uses the inputs $x^{(i)}$ for learning and automatically extracts useful information from our data. RBM is the non-directed graphical model that finds the distribution over input vector ' x ' using the hidden layer binary inputs. ' W ' represents the connections between the visible layer and the hidden layers.

Here 'x' and 'h' are the random variables. In (3) the energy function can be calculated which on making exponential gives the distribution over the networks shown in 3.1. Here Z is a partial function which is intractable.

Energy function:

$$E(x, h) = b' - x - c'h - h'Wx \quad (3)$$

Distribution:

$$P(x, h) = \exp(-E(x, h)) / Z \quad (3.1)$$

SOLUTIONS AND RECOMMENDATION

As keeping in mind various constraints for the learning the latent features especially from an image the authors have proposed a novel technique. The proposed algorithm is capable of retrieving images with 'n' keywords, minimum of 'n' image data classes and 'n' context classes. Any NLP query can be the input to this algorithm. Then the query undergoes some pre-processing methodologies and the refined (Query, Keyword) pair is fed as an input to deep belief network. The pre-processing on query involves removing the frequently used words i.e., stop words, suffix stripping (M.F Porter, 1980) for yielding the root words, and spell check (Victoria J. Hodge, Austin Jim, 2001) to identify grammatical errors. The (Query, Keyword) pair extraction is followed by the pre-processing step. This (Query, Keyword) pair (q', K) serves as the input to first layer of deep belief network.

Numerous relations exist between the pairs and among them semantically correct pairs are extracted. Thus the relation score and semantic score can be treated as input to the successive layers of DBN. Ultimately the deep belief network points to the well-defined context classes. Fuzzy decision making approaches can aid in the mapping between the context classes and the data set for image retrieval.

FUTURE RESEARCH DIRECTIONS

Deep architectures are a promising research area, which yield more researchers to add on assorted techniques with an improved performance in the training, learning tasks. Future enhancement to this work can incorporate multi-modality to latent feature learning alongside selecting the adaptive kernel function automatically for learning through deep belief network.

CONCLUSION

The deep learning architectures are the paradigm shift in the study of artificial neural networks that evolved with an objective of optimizing the training task. The deep complex non linear structure makes this model to understand diverse latent features through various learning methodologies. This article lightens the way towards the deep architecture in the several domains. The innovative methodology depicted in this work for image retrieval aids the characteristics of natural language processing built under utilizing the deep belief network concepts. The NLP query is ensemble into various latent query features and fed into deep belief network as inputs. The proposed learning strategy makes use of restricted Boltzmann machine. After consecutive learning extract the image classes thereby retrieve the images of the analogous query.

REFERENCES

- Hodge & Austin. (2001). A comparison of a novel neural spell checker and standard spell checking algorithms. *Pattern Recognition*, 35(11), 2571–2580.
- Abby, A. G. (2000). Image Information Retrieval: An Overview of Current Research. *Informing Science*, 3(2), 63–66.

Chai, J. Y., Zhang, C., & Jin, R. (2007). An Empirical Investigation of User Term Feedback in Text based Targeted Image Search. *ACM Transactions on Information Systems*, 25(1), 1–25. doi:10.1145/1198296.1198299

Hinton, G. E. (2006). *To recognize shapes, first learn to generate images. Technical Report UTML TR 2006-003*. University of Toronto.

Le Roux, N., & Bengio, Y. (2008). Representational power of restricted Boltzmann machines and deep belief networks. *Neural Computation*, 20(6), 1631–1649. doi:10.1162/neco.2008.04-07-510 PMID:18254699

Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America*, 20(7), 1434–1448. doi:10.1364/JOSAA.20.001434 PMID:12868647

Porter, M. F. (1980). An algorithm for suffix stripping. *Program*, 14(3), 130–137. doi:10.1108/eb046814

Westrvel, T., Gemert, J. C., Cornacchia, R., Hiemstra, D., & Vries, A. P. (2005). An Integrated Approach to Text and Image Retrieval. *Proceedings of the TRECVID Workshop*

Yoshua, B. (2009). Learning Deep Architectures for AI. *Foundations and Trends in Machine Learning*, 2(1), 1–127. doi:10.1561/22000000006

KEY TERMS AND DEFINITIONS

Convolutional Neural Networks: A multi layer neural network similar to artificial neural networks only differs in its architecture and mainly built to recognize visual patterns from image pixels.

Deep Architectures: The deep learning architectures model higher level abstractions of data by learning through the complex abstract features embedded in the data with the help of its deep complex architecture.

Deep Belief Network: It is the type of deep neural network which is an unsupervised generative model.

Deep Neural Networks: An artificial neural network with multiple hidden processing neurons possessing higher performance in learning tasks utilizing its non linear complex architecture with the aid of back propagation algorithm. It is a model that recreates a picture by making relationships among feature of data.

Information Retrieval: The process of bringing down to the relevant information from various information resources and it can be of text retrieval, image retrieval etc.

Latent Query Features: The hidden semantic meaning of the input query such as antonyms, synonyms, homonym and polysemy.

Online Information Retrieval Systems Trending From Evolutionary to Revolutionary Approach



Zahid Ashraf Wani

University of Kashmir, India

Huma Shafiq

University of Kashmir, India

INTRODUCTION

Information has been available since the evolution of the world and is being applied every day. McCreddie and Rice (1999) summarize the concept of information that has been proposed over the years in various studies and explains information in four different ways. Firstly, as a representation of knowledge, studies are of the opinion that information is a representation of knowledge. Earlier when there was no concept of internet and digital media, the printed documents were assumed to be the primary representation of information. But with the advent of ICT, this view point has changed from print to the electronic media (Lievrouw, 1988). Secondly, as a data available in the environment, some of the researchers view information in broader paradigms which may include objects, sounds, smells, events, activities, artefacts, even any phenomenon of nature, etc. The data may not convey any message as such but can be informative if interpreted appropriately (Buckland, 1990; Taylor, 1991). Thirdly, as a part of communication process. Various researchers tend to believe that information is a part of communication process among individuals of the society. Social elements play vital role in processing and inferring the information which are not treated as physical or cognitive activities apart from the work of individuals but regular intrinsic activities which constitute the very nature of people while interacting in organizations (Atwood & Dervin,

1982; Solomon, 1997a; Solomon, 1997b; Solomon, 1997c). Lastly, as a resource or commodity. Some researchers stress upon information being a resource or a thing which can be generated, acquired, replicated, disseminated, employed, organised, traded, sold and exchanged. It is transmitted from a sender to receiver in the form of a message and is later interpreted at the receiving end (Arrow, 1979; Bates, 1988; Buckland, 1991; Hirschleifer & Riley, 1992). But all this makes sense to end user if this information is retrieved and received at right time with minimum of fuss and intricacies. That is where concept of information retrieval system comes to play its role, more so in cyberage when zillions of terabytes of information is generated every year.

Man by its intrinsic curiosity to gain new insights of himself and surroundings has always been on the path to search and sift whatever satisfies his intellectual quest. Every time he searches for some specific information, he draws strategy in mind to retrieve the desired information. Nowadays in order to retrieve the desired information one must have good knowhow of search mechanism. It all started at the fade end of the 20th century, with the invention of the World Wide Web that revolutionized the whole world and with this followed the exponential growth of information. Information seeking was already the basic nature of a man, and now two decades later, it has become a part and parcel of our day to day lives. The trend of information search via the traditional library

DOI: 10.4018/978-1-5225-2255-3.ch394

environments has been entirely evolved over the years to a virtual space called Web, where all this information can exist. Web contains all sorts of digitized as well as born digital formats of information which are just a click away from the users. With each passing day, World Wide Web gained popularity, thus searching through the web, also referred to as the information retrieval on the web, has become an important research area.

The Web has the features of being an information space having free access with no particular focal point to control and scrutinize the dissemination of information resulting in many challenges including the maintenance of authenticity and quality of information. Thus, various computational tools are being used in order to perform the search processes over the web among which the most widely adopted approaches are web directory and search engine services. In a web directory, the data in the web is organized in a classified manner having some hierarchical structure making it equivalent to the traditional library system. Search engines, on the other hand, are simply the web portals used to find information on the web. These have the capability of indexing large segments of the web and then storing the information in their databases. In contrast to a web directory, the processes of a search engine are carried out automatically (Zhu, 2011). Apart from web directories and search engines, there are a number of information retrieval tools that have evolved over the years revolutionizing the concept of information retrieval. In this perspective, the present study delves deep and highlight various trends in online information retrieval from primitive to modern ones. Furthermore, the study has made an attempt to visualize the future requirements and expectation keeping in view the ever-increasing dependence on diverse species of information retrieval tools.

BACKGROUND

The evolution of man has begun millions of years ago. With this, information has also evolved right from the beginning. This information has passed on to generation after generation with so many advancements through its way. Primitive man may not have farsighted the importance of storing this information, but with time it started getting clear for him that unaided memory, actions and speech cannot be sufficient to recapitulate all the information and transmit the same to future generations. Consequently, he started storing this information with whatever source he could get hold on, e.g. stones, leaves, parchments, clay tablets, wax tablets, etc. Historians and anthropologists make this advancement of primitive man evident from various live examples available even to this date like cave paintings, pictographic and hieroglyphic forms of writing, primitive sculptures, etc. Most of these storage media have flaws of deterioration in one way or the other making it obvious for people to search for other ways of storing information that could last long. Subsequently, Egyptians came up with the idea of stalks of papyrus plant pounded flat and pressed into a relatively acceptable writing surface, resulting in the existence of long strips of stalks that could be rolled. These scrolls thus became one such early forms where information could be stored. But it was difficult to locate a particular piece of information from such long scrolls. A great technical development of information storage on scrolls came when these long scrolls were cut and bound together in the form of what we today call as a 'book'. The early set up of holding the stored information was not much of a problem to organize as it was only confined to a limited stock and keepers of this information were thoroughly familiar with its entire content. This stood true only upto the invention of printing press in 16th century which revolutionized the concept of information production. This great invention enabled man to create new records with far better pace than the manual approach and duplicate the same in ample quantity. This

eventually led to the concept of organizing and managing the information that has been gathered over the centuries (Gull, 1955).

With information storage came the concept of information retrieval. Few American scientists thought of this concept after World War II for better ways to handle scientific knowledge as much of this important knowledge was being lost due to poor handling of overwhelming information in scientific literature. Though the concept of information retrieval was originally intended for scientific literature, it is applied to all the available information throughout the globe. The credit for development of information retrieval systems goes to the prophetic article "As we may think" by Vannevar Bush in 1945 where he proposed the invention of a special machine "MEMEX" to accommodate the records in the form of a microfilm. According to Bush, the machine was easy to build but designing a workable indexing system was the real problem to face (Wilds, 1961).

The notion of information retrieval has been a problem ever since man started saving his thoughts and surrounding information he became aware of. However, the concept of 'Information Retrieval' was already prevailing in the scholarly community but was unknown until Calvin Mooers coined the term in early 1950s. While reviewing the previous literature, it has been observed that there is a variation in the exact year when the term was actually coined. Studies claim that the term was coined in 1949 (Fairthorne, 1968), some say in 1950 (Mishra, 2014), some in 1951 (Hjørland, 2006), while some are of the opinion that the term was coined in 1952 (Onwuchekwa & Jegede, 2011; Sparck & Willet, 1997).

Information retrieval refers to the area of study that deals with searching for documents, information available in the documents, structured databases like digital libraries, or web and retrieving the relevant information to the users for their use. It is an interdisciplinary field based on information science, computer science and cognitive psychology. The basic goal of information retrieval was to create such systems (for instance,

bibliographic databases) which could provide access to a number of information resources like books, journals, etc. Now-a-days, the most visible information retrieval applications are web search engines (Dineta, Chevalierb & Tricot, 2012).

TRENDS IN INFORMATION RETRIEVAL SYSTEMS

1. Indexing Approach

Mooers (1960), though being the one to coin the term himself, after a decade has the notion that there are yet no complete solutions for retrieval of large collections of information. At that time, information retrieval was not regarded as the sheer discovery and providing information.

Arnold (1958) on the other hand gives an overview of previous ten years about various aspects in the field of information retrieval. The study shows that over the past decade, researchers have been concerned with different possibilities of improving how information can be stored in catalogues and how to prepare an exhaustive list of entry points for an index so that effectiveness could be achieved in the process of information retrieval. The researchers working on index formation have been inspired by Mortimer Taube's Uniterm indexing system. Besides Taube, with the help of Batten's punched card system, introduced his method of referring a user to the relevant information based on previous research works.

Studies have revealed that evolution of information retrieval systems started with various retrieval tools like card catalogues, indexes, etc. (Shaw, 1951; Voigt, 1956).

There have been many types of card catalogs that are observed to be used for early information retrieval purposes. Edge-notched cards (also known as edge-punched cards) being one of earlier tools used where all indexing characteristics of one particular record are punched at one place or card. Another type of card is aspect card which include Uniterm card, Peek-a-boo card, Batten card, etc.



Here, each characteristic of an index is recorded separately. One more form of card catalogs include punched cards (also called Hollerith cards). These punched cards are similar to edge-notched cards except that machines are involved while searching and thus a large number of index entries can be expected (Wilds, 1961).

Gull (1955) speaks about punched card accounting machines which needed human supervision so that information could be organized, stored and later retrieved. Thus, analog and digital computers have been given preference which can operate on large amounts of information utilizing minimum possible time. These computers could also be used with lesser human supervision for many hours when programmed accurately.

Moreover, Dietrich (1960) while studying about machine retrieval of Pharmacological data has been of the opinion that punched cards along with machine sorting can be helpful in organizing and retrieving the Pharmacological data.

Kendrick (1964) studies about information storage and retrieval revealing that punched cards are being used which helped for machine sorting, print outs, etc. and eventually magnetic tapes, discs or drum storage would help simplifying the practical applicability of various integrated systems of electronic information storage and retrieval once those are introduced.

Taube (1956) in his study discusses about the machine retrieval and makes it clear that machine retrieval at that time was not as satisfactory as retrieval using tools like card catalogs or to be precise, printed cards. He further says that even though card catalogs show better results than machine retrieval but at the same time these card systems do not fit for large scale storage and retrieval systems.

2. Computer-Based Approach

The emergence of computerized information retrieval services began in early 1970's. Since then, a number of information retrieval services have appeared while a decade later, indexing and

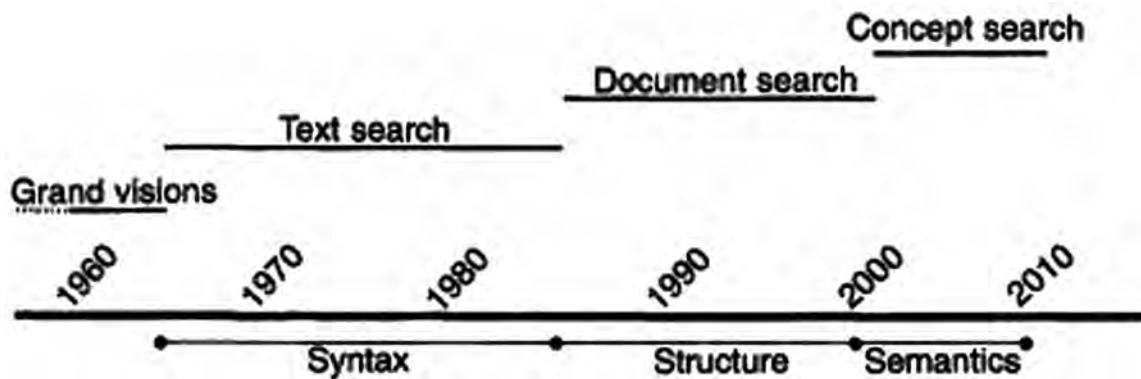
abstracting services started gaining importance in the field of information retrieval. Some of these services include Educational Resources Information Center (ERIC), National Information Center for Educational Media (NICEM), Exceptional Child Educational Resources (ECER), etc. Such information retrieval services help in an efficient and rapid retrieval of information. Searching via ERIC is boosted by using its controlled vocabulary "*Thesaurus of ERIC Descriptors*". Apart from this, users can use keywords from title or abstracts which may act as additional access points for searching thus increasing the efficiency of searching (which cannot be seen in manual searching). Search strategy modifications like broadening, limiting, or refining the search yield more relevant search results, again increasing the efficiency of retrieval systems (Gerke, 1981).

Computer-based information retrieval systems have grown to the full bloom in late 1970s and early 1980s. Bawden (1984) thus talks about the use of various indexing and abstracting services for Statistical Literature. The study shows a trend of using various indexing and abstracting services (like Index Medicus, Science Citation Index, Mathematical Reviews, Statistical Theory and Method Abstracts, etc.) during the present decade for an efficient retrieval of relevant information.

Indexing, abstracting, thesauri formation, file formatting, etc. have always been the topics of interest for researchers in the field of library and information science. With the advent of high speed computers, and a simultaneous interest in the field of information retrieval, the latest trend is directed towards assessing the effectiveness of various computerized file structures and search strategies besides, comparing the utility of index languages as well (Levine, 1985).

According to Salton (1991), the records stored in information retrieval systems during this era are typically identified by using Boolean combinations of index terms. The retrieval systems are designed in a way that provide only those results which match the exact combinations used in search queries.

Figure 1. Rough timeline of the generations of information retrieval in digital libraries. Source (Schatz, 1997)



3. Bibliographic Databases Approach

Early 1970s witnesses the development of on-line systems allowing interactive searching. These online systems thus paved way to the development of various bibliographic databases for information retrieval which can hold information about millions of records. These online information retrieval systems used inverted file organization and also incorporated various search techniques like Boolean operators, truncation, etc. helping users formulate their queries in a better possible way and retrieve better results (Salton, 1980).

Doszkocs, Rapp and Schoolman (1980) in their study talk about the development of simple retrieval systems based on the use of natural language which can be achieved by dealing with the ambiguities of syntax and grammar. It has been seen from the recent studies about National Library of Medicine that various prototypes using English language interface have been developed, e.g. MEDLINE, Hepatitis Knowledge Base, TOX-LINE, etc. The basis for the development of such prototypes was a change from complex retrieval systems to simple ones using natural language of users and understanding that language in order to provide relevant results to the query.

Leider (1964) discusses about microfilm retrieval system whose concept was envisioned by Vannevar Bush in his visionary article regarding Memex. Records can be saved in miniaturized

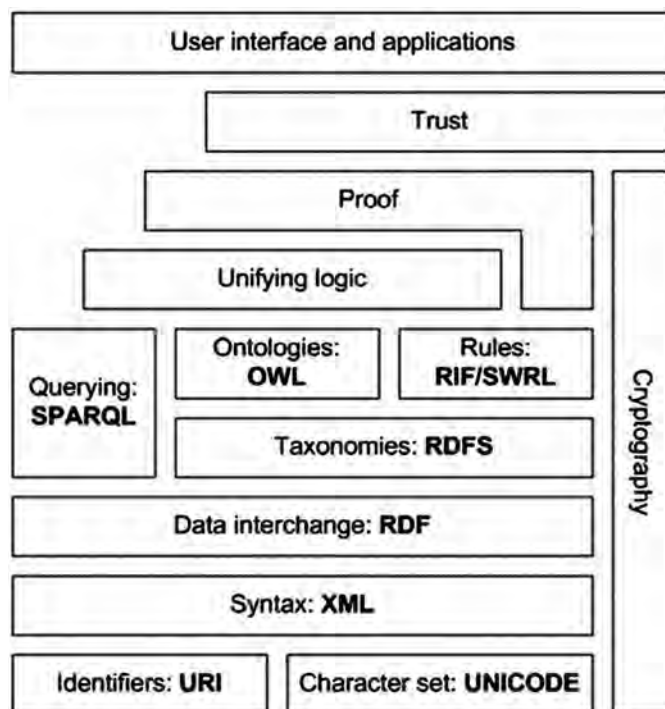
forms in microfilms. In this context, Leider says that microfilm retrieval systems not only list the document numbers but provide full text records about a particular information which can be reproduced in the form of hard copies later by users.

4. Web Based Approach

Right from the development of online information retrieval era, the fundamental functions involved in information storage and retrieval processes had not changed for over three decades but Schatz (1997) has envisioned fundamental technology for large collection retrieval changing for future generations from syntactic to semantic approach. Figure 1 shows the Grand vision proposed by Schatz. Schatz further says that the recent advances in World Wide Web (WWW) has enabled users of online information retrieval systems to search for information sources across the global network with the help of internet.

Xie and Cool (1998), Ahmed et al. (2005), Islam and Panda (2007) provide a scenario of emerging interactive information retrieval systems offering sophisticated system features such as use of natural language queries, rank documents as per relevancy, automatic relevance feedback, provide full text of documents, etc. Furthermore, Xie and Cool (1998) give emphasis on another emerging concept of web based searching as is it intrinsically interactive providing various significant

Figure 2. Semantic web architecture. Source: (Thangaraj & Sujatha, 2014)



features like navigation to different databases, use of multiple interactive search strategies, mapping to thesaurus terms, etc.

Chowdhury (1999) also emphasizes on the significance of World Wide Web which led to the use of internet for information exchange in all spheres of day to day life. Libraries started offering various services via internet like Online Public Access Catalog (OPAC), online databases vendors like DIALOG, and CDROM database vendors like SilverPlatter, etc.

Ide and Salton (1971) while studying about interactive search strategies in information retrieval predict that retrieval effectiveness can be achieved more significantly by adopting adaptive interaction techniques which would help in extracting information from users thus improving the organization of data space. Hence leading to better and effective search and retrieval processes.

5. Semantic Approach

The concept of semantic search in the field of information retrieval though has been available since early 1980s (Croft, 1986; Guarino et al., 1999), but could not gain much importance until Berners-Lee proposed the concept of semantic web technology in 2000 who defines it as an extension of web where information is given a well-defined meaning assisting users and information retrieval systems to work in cooperation with each other (Berners-Lee, Hendler & Lassila, 2001; Lou & Qiu, 2014).

Figure 2 shows the architecture of semantic web as proposed by Berners-Lee. The main objectives of semantic web include semantic query processing and finding semantic information from unstructured information available on the web (Thangaraj & Sujatha, 2014).

Since the architecture suggested by Tim Berners-Lee in 2000, various studies have been carried out proposing structural layouts of semantic web. Ilyas et al. (2004) has suggested a conceptual architecture for semantic search engines laying main emphasis on inference engine. This model

has been created using a complete knowledge base using a relational database and claiming that the model shows high performance as data can be directly retrieved from the knowledge base reducing the noise to the maximum level. Later in 2007, Zhi-Qiang et al. (2007) proposes another framework of semantic web search engine based on domain ontology. This architecture is based on a crawling mechanism as seen in present day search engines like Google, Bing, etc. and uses an already constructed ontology based information extraction algorithm to retrieve information from crawled web pages.

Fernández et al. (2011), Kara et al. (2012) talk about the advances in information retrieval from contemporary keyword approach to conceptualization approach (i.e. semantic approach) as semantic approach has been in its evolution phase at the time of these studies. Renl (2009) also discusses more advanced areas of information retrieval focusing on emerging three areas viz. cross-lingual, multimedia and semantic-based information retrieval. Firstly, cross-lingual approach involves the retrieval of information in more than one language. The resulting information can be later translated into the language of user query so that users may understand the gist of retrieved document. This approach became necessary with increasingly globalized economy. Secondly, apart from finding different types of media on internet, various information storage systems like computers started being used to store this media other than text such as audio, video, images, etc. and thus methods for rapid and accurate retrieval of such media also became a necessity resulting in multimedia information retrieval. Finally semantic information retrieval tries to go beyond the traditional approach of retrieval and is based on concept based approach i.e. it tries to define concepts in documents and queries rather than just focusing on query terms, thus aids in improved and efficient retrieval of information.

Bodoff (2006) while discussing about the conventional information retrieval interfaces (i.e. the last decade of 20th century and the beginning of

21st century) says that these interfaces have been designed for analytical approach of searching rather than just a browsing approach, the web based interfaces otherwise involve browsing approach.

Traditional information retrieval systems lead to low precision and high recall. However, in case of semantic based information retrieval systems, an improvement in the relevancy of retrieved results is observed as these systems understand the search intentions and conceptual meaning of queries entered by users (Remi & Varghese, 2015).

To crown it all it would be in the fitness of things to shed some light on Google - the most popular search engine, and the first preference for most of the online users when it comes to searching information on World Wide Web, a detailed account of Google and its searching mechanism is given as follows:

Google Search, owned by Google Inc., is commonly referred to as Google Web Search or simply Google. It is the most widely used search engine managing more than three billion searches per day. It is also ranked as the most used search engine in the United States as of September 2016. Google is commercial web search engine and is available in 123 different languages, thus having multilingual interface.

Google Search Mechanism

When users submit any search term in Google search, a very long list of mostly relevant search results (often in millions) from all over the web is presented to them instantly. There are three main processes that Google search undergoes while performing the search and delivering the results:

Crawling

Crawling when defined is simply a process in which Google index is generated by adding new and updated pages with the help of Googlebot. Googlebot is a computer program that is used to fetch hundreds and thousands of web pages in order to create an index called Google index. Googlebot



is many a times synonymously used for spider, robot, or simply bot. The algorithmic process used by Googlebot is that the program first determines which sites need to be crawled, how often these sites need to be crawled, and finally how many pages need to be crawled from each site.

The crawling process in Google starts with previously created list of web page URLs and eventually the list is increased with recently uploaded web pages by various data providers. Googlebot keeps on detecting links on every site and adding the same to the index. Whenever any new site, any change in previous site, or any dead link is detected, it eventually updates the Google index in order to keep it up to date.

Indexing

After crawling any page, Googlebot processes that page so that an index can be compiled from all the words in that page along with the location of each word on the page. It further includes all the information in key content tags and attributes in the index. It is important to note that Googlebot has the ability to process most of the content types except for few like some rich media files.

Serving Results

This process is concerned with the communication between the system and the user. This process starts when a user submits any query and ends with displaying the results to the users. After the submission of a search query, the system searches its index for related information and tries to return back the most relevant results to the user. This relevancy is determined by so many factors, and PageRank is one such factor. Page rank (named after one of the founders of Google, Larry Page) is an algorithm which ranks websites by determining their importance for a search query. This is done by counting how many links are associated with a particular website or page and how qualitative these links are, thus determining the importance of that particular page i.e. more the

links associated to a page, more is the importance of that page. All the factors that negatively affect the search results like spam links are always kept in check by Google so that user experience can be constantly improved. Google is trying hard to make its interface more user-friendly and some features like “Did you mean” and Google Autocomplete (Stemming) are best examples to explain the same. Since the time of the user is very important (Ranganathan’s Fourth Law of Library Science), Google has designed these features to help users save their time by displaying popular queries, related terms, or even common spelling errors (Google, 2015).

However, the trend is changing from having robust search mechanism to personalized information delivery service based on previous search behavior of the user what we also call in our parlance proactive or anticipatory service. Such trend is fast picking up and is adapted and developed by most of the top notch search/software platforms like Google, Microsoft and Apple, viz. Google Assistant, Cortana and Siri respectively. Such applications are revolutionizing the information retrieval world by providing users a personalized experience. Whenever users search any new information with these search applications, a personalized search profile is automatically created in these applications which gets more intelligent and search efficient as we go. Anything new appears or gets updated on web that is related to the personalized search profile an alert pops up for users via these applications. These applications, thus, are intelligent personal assistants and knowledge guides that always keep users updated about their interests and help them get desired information retrieved without much efforts.

FUTURE RESEARCH DIRECTIONS

Future of the Web is relying on how intelligent and efficient web retrieval tools will become. The smart web is relying on further development of these wonder tools. Therefore, developing proper

structure linkages and relations between the data is imperative for any search tool. The semantics of data is paramount in the success of any online retrieval tool. So it would be interesting to study the stages of developing this semantics by prominent search tools, and how far they have attained their goals. Besides investigating how far major search tools will allow metadata cross walks for better retrieval efficacy can be an interesting area of future research.

CONCLUSION

With the advent of Gutenberg moving printing press, information began to be produced, reproduced and communicated at large scale and people began to think to devise systems to facilitate easy and efficient discovery of requisite information. The much prominent amongst the earliest systems were indexing cards which provided very few directions to reach to the information sources. However, there has always been unprecedented production of information which encourages, rather compels to devise new and to improve upon existing systems. Information retrieval systems also witnessed machine-controlled indexing systems which were more or less machine equivalent of card system but with brighter scope of easy information retrieval. In the current era dominated by presence of electronic gadgets in every sphere of human activity, information retrieval systems have taken a leap ahead making use of artificial intelligence and semantic web. It would be more exciting to observe what is next to come in the domain of human-machine and machine-machine resource discovery and access. With each new approach, the retrieval relevancy of information retrieval systems is moving towards more balance between its precision and recall. The semantic web and context based personalization are evolving as unique modes to offer precision based information to the user. These two new features if properly synchronized can yield better efficiency and higher level of satisfaction with the information

retrieval system and can enormously help users to avoid retrieving noise to great extent. Besides researchers are also getting very close to evolve a comprehensive multimedia content retrieval mechanism which can retrieve precise part of multimedia file and play the specific portion out of whole. We expect to use these revolution information retrieval systems in near future which can further alleviate the difficulties faced by the end users.

REFERENCES

- Ahmed, S. M. Z., McKnight, C., & Oppenheim, C. (2005). A study of learning and retention with a web-based IR interface. *Journal of Librarianship and Information Science*, 37(1), 7–16. doi:10.1177/0961000605052155
- Arnold, D. V. (1958). Information Retrieval: A pragmatic approach. *The Journal of Documentation*, 14(4), 183–189. doi:10.1108/eb026259
- Arrow, K. J. (1979). The economics of information. In M. L. Dertouzos & J. Moses (Eds.), *The computer age: A twenty year view* (pp. 306–317). Cambridge, MA: MIT Press.
- Atwood, R., & Dervin, B. (1982). Challenges to sociocultural predictors of information seeking: A test of race versus situation movement state. In M. Burgoon (Ed.), *Communication yearbook* (Vol. 5, pp. 549–570). New Brunswick, NJ: Transaction Books.
- Bates, B. J. (1988). Information as an economic good: Sources of individual and social value. In V. Mosco & J. Wasco (Eds.), *The political economy of information* (pp. 76–94). Madison, WI: University of Wisconsin Press.
- Bawden, D. (1984). Computer-Based Online Information Retrieval for Statistical Literature. *Journal of the Royal Statistical Society. Series A (General)*, 147(1), 78–86. doi:10.2307/2981738

- Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. *Scientific American*, 284(5), 28–37. doi:10.1038/scientificamerican0501-34 PMID:11341160
- Bodoff, D. (2006). Relevance for browsing, relevance for searching. *Journal of the American Society for Information Science and Technology*, 57(1), 69–86. doi:10.1002/asi.20254
- Buckland, M. K. (1990). Information retrieval and the knowledgeable society. *Proceedings of the 53rd Annual Meeting of the American Society for Information Science* (pp. 239-244). Medford, NJ: Learned Information.
- Buckland, M. K. (1991). Information as thing. *Journal of the Association for Information Science and Technology*, 42(5), 351–360. doi:10.1002/(SICI)1097-4571(199106)42:5<351:AID-ASI5>3.0.CO;2-3
- Chowdhury, G. G. (1999). The internet and information retrieval research: A brief review. *The Journal of Documentation*, 55(2), 209–225. doi:10.1108/EUM0000000007144
- Croft, W. B. (1986). User-specified domain knowledge for document retrieval. *Proceedings of the 9th annual international ACM conference on research and development in information retrieval (SIGIR 1986)*, 201–206.
- Dietrich, E. V. (1960). Machine Retrieval of Pharmacological Data. *Science, New Series*, 132(3439), 1556-1557. Retrieved from: <http://www.jstor.org/stable/pdf/1707062.pdf>
- Dineta, J., Chevalierb, A., & Tricot, A. (2012). Information search activity: An overview. *European Review of Applied Psychology*, 62(2), 49–62. doi:10.1016/j.erap.2012.03.004
- Doszkoecs, T. E., Rapp, B. A., & Schoolman, H. M. (1980). Automated Information Retrieval in Science and Technology. *Science, New Series*, 208(4439), 25-30. Retrieved from: <http://www.jstor.org/stable/1683789>
- Fairthorne, R. A. (1968). Repartee: The Limits of Information Retrieval. *The Journal of Library History (1966-1972)*, 3(4), 363-369. Retrieved from: <http://www.jstor.org/stable/25540130>
- Fernández, M., Cantador, I., López, V., Vallet, D., Castells, P., & Motta, E. (2011). Semantically enhanced Information Retrieval: An ontology-based approach. *Web Semantics: Science, Services, and Agents on the World Wide Web*, 9(4), 434–452. doi:10.1016/j.websem.2010.11.003
- Gerke, R. (1981). Instructional Strategies through Computerized Information Retrieval. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 55(1), 9–11. doi:10.1080/00098655.1981.9958195
- Google. (2015). *How Google search works*. Retrieved from: <https://support.google.com/webmasters/answer/70897?hl=en>
- Guarino, N., Masolo, C., & Vetere, G. (1999). OntoSeek: Content-based access to the web. *IEEE Intelligent Systems*, 14(3), 70–80. doi:10.1109/5254.769887
- Gull, C. D. (1955). Implications for the Storage and Retrieval of Knowledge. *The Library Quarterly: Information, Community, Policy*, 25(4), 333-343. Retrieved from: <http://www.jstor.org/stable/pdf/4304468.pdf>
- Hirschleifer, J., & Riley, J. G. (1992). *The analytics of uncertainty and information*. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9781139167635
- Ide, E., & Salton, G. (1971). *Interactive Search Strategies and Dynamic File Organization in Information Retrieval*. Retrieved from: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.232.6072&rep=rep1&type=pdf>
- Ilyas, Q. M., Kai, Y. Z., & Talib, M. A. (2004). A Conceptual Architecture for Semantic Search Engine, *IEEE*, 605-610.

- Hjørland, B. (2006). Information retrieval (IR). *Core Concepts in Library and Information Science (LIS)*. Retrieved from: http://www.iva.dk/bh/Core%20Concepts%20in%20LIS/articles%20a-z/information_retrieval.htm
- Islam, A., & Panda, K. C. (2007). Web-based information retrieval trends of researchers: A case study of Sambalpur University (India). *The Electronic Library*, 25(6), 757–765. doi:10.1108/02640470710837173
- Kara, S., Alan, O., Sabuncu, O., Akpınar, S., Cicekli, N. K., & Alpaslan, F. N. (2012). An ontology-based retrieval system using semantic indexing. *Information Systems*, 37(4), 294–305. doi:10.1016/j.is.2011.09.004
- Kendrick, W. B. (1964). Towards better information storage and retrieval. *Mycology*, 56(5), 781–782. Retrieved from: <http://www.jstor.org/stable/3756635>
- Leider, A. (1964). Modern Systems of Information Storage and Retrieval. *American Water Works Association*, 56(5), 569–574. Retrieved from: <http://www.jstor.org/stable/41263895>
- Levine, H. G. (1985). Principles of Data Storage and Retrieval for Use in Qualitative Evaluations. *Educational Evaluation and Policy Analysis*, 7(2), 169–186. doi:10.3102/01623737007002169
- Lievrouw, L. A. (1988). Four programs of research in scientific communication. *Knowledge in Society*, 1(2), 6–22. doi:10.1007/BF02687210
- Lou, W., & Qiu, J. (2014). Semantic information retrieval research based on co-occurrence analysis. *Online Information Review*, 38(1), 4–23. doi:10.1108/OIR-11-2012-0203
- McCreadie, M., & Rice, R. E. (1999). Trends in analyzing access to information. Part I: Cross-disciplinary conceptualizations of access. *Information Processing & Management*, 35(1), 45–76. doi:10.1016/S0306-4573(98)00037-5
- Mishra, R. (2014). Trends in Information Retrieval. In P. K. Singh, A. Kumar., V. Trivedi, & V. Trivedi (Eds.), *Opportunities and Challenges of the Institutional Library in Rural Areas. Pre seminar proceedings of Shri Shakti Degree College, Sankhahari, Ghatampur, Kanpur Nagar*. Bharati Publishers & Distributors. Retrieved from: http://ssdckanpur.org/images/final_book_with_cover_page.pdf#page=136
- Mooers, C. N. (1960). The next twenty years in information retrieval. *American Documentation*, 11(3). Retrieved from: <http://search.proquest.com/openview/e5357d2518ba652d3dca072a22c95a40/1?pq-origsite=gscholar&cbl=1818555>
- Onwuchekwa, E. O., & Jegede, O. R. (2011). Information Retrieval Methods in Libraries and Information Centers. *An International Multidisciplinary Journal*, 5(6), 108–120. doi:10.4314/afrrrev.v5i6.10
- Remi, S., & Varghese, S. C. (2015). Domain Ontology Driven Fuzzy Semantic Information Retrieval. *Procedia Computer Science*, 46, 676–681. doi:10.1016/j.procs.2015.02.122
- Ren, F. (2009). Advanced Information Retrieval. *Electronic Notes in Theoretical Computer Science*, 225, 303–317. DOI:10.1016/j.entcs.2008.12.082
- Salton, G. (1980). Automatic information retrieval. *IEEE Computer*, 41–56. Retrieved from: <https://www.computer.org/csdl/mags/co/1980/09/01653784.pdf>
- Salton, G. (1991). Developments in Automatic Text Retrieval. *Science, New Series*, 253(5023), 974–980. Retrieved from: <http://www.jstor.org/stable/2878789>
- Schatz, B. R. (1997). Information Retrieval in Digital Libraries: Bringing Search to the Net. *Science, New Series*, 275(5298), 327–334. Retrieved from: <http://www.jstor.org/stable/2890992>

- Shaw, R. R. (1951). Management, Machines, and the Bibliographic Problems of the Twentieth Century. In J. H. Shera & M. E. Egan (Eds.), *Bibliographic Organization* (pp. 200–225). Chicago: University of Chicago Press.
- Solomon, P. (1997a). Discovering information behavior in sense-making. I: Time and timing. *Journal of the Association for Information Science and Technology*, 48(12), 1097–1108. doi:10.1002/(SICI)1097-4571(199712)48:12<1097:AID-ASI4>3.0.CO;2-P
- Solomon, P. (1997b). Discovering information behavior in sense-making. II: The social. *Journal of the Association for Information Science and Technology*, 48(12), 1109–1126. doi:10.1002/(SICI)1097-4571(199712)48:12<1109:AID-ASI5>3.0.CO;2-Y
- Solomon, P. (1997c). Discovering information behavior in sense-making. III. The person. *Journal of the Association for Information Science and Technology*, 48(12), 1127–1138. doi:10.1002/(SICI)1097-4571(199712)48:12<1127:AID-ASI6>3.0.CO;2-W
- Sparck, K., & Willett, P. (1997). Overall Information. In J. K. Sparck & P. Willett (Eds.), *Readings in information retrieval*. San Francisco: Morgan Kaufmann Pub.Inc.
- Taube, M. (1956). Machine Retrieval of Information. *Library Trends*, 5(2), 301–308. Retrieved from https://www.ideals.illinois.edu/bitstream/handle/2142/5694/librarytrendsv5i2_opt.pdf?sequence=3#page=116
- Taylor, R. S. (1991). Information use environments. In B. Dervin (Ed.), *Progress in Communication Sciences* (Vol. 10, pp. 217–225). Norwood, NJ: Ablex.
- Thangaraj, M., & Sujatha, G. (2014). An architectural design for effective information retrieval in semantic web. *Expert Systems with Applications*, 41(18), 8225–8233. doi:10.1016/j.eswa.2014.07.017
- Voigt, M. J. (1956). The Trend toward Mechanization in Libraries. *Library Trends*, 5(2), 193–204. Retrieved from https://www.ideals.illinois.edu/bitstream/handle/2142/5694/librarytrendsv5i2_opt.pdf?sequence=3#page=116
- Wilds, T. (1961). Information Retrieval. *The American Archivist*, 24(3), 269–282. Retrieved from <http://www.jstor.org/stable/40290892> doi:10.17723/aarc.24.3.y36258ju1j434001
- Xie, H. I., & Cool, C. (1998). Online Searching in Transition: The Importance of Teaching Interaction in Library and Information Science Education. *Journal of Education for Library and Information Science*, 39(4), 323–331. doi:10.2307/40324307
- Zhi-Qiang, D. U., Jing, H. U., Hong-Xia, Y. I., & Jin-Zhu, H. U. (2007). The Research of the Semantic Search Engine based on the Ontology. *WiCom IEEE*, 360-363.
- Zhu, Z. (2011). *Improving Search Engines via Classification* (Doctoral dissertation). University of London, London, UK. Retrieved from: <http://www.dcs.bbk.ac.uk/research/recentphds/zheng.pdf>

ADDITIONAL READING

- Beaulieu, M. (2000). Interaction of information searching and retrieval. *The Journal of Documentation*, 56(4), 431–439. doi:10.1108/EUM0000000007122
- Brand-Gruwel, S., Wopereis, I., & Walraven, A. (2009). A descriptive model of information problem-solving while using Internet. *Computers & Education*, 53(4), 1207–1217. doi:10.1016/j.compedu.2009.06.004
- Khan, L., McLeod, D., & Hovy, E. (2004). Retrieval effectiveness of an ontology-based model for information selection. *The International Journal on Very Large Data Bases*, 13(1), 71–85. doi:10.1007/s00778-003-0105-1

Lopatovska, I., & Arapakis, I. (2011). Theories, methods and current research on emotions in library and information science, information retrieval and human computer interaction. *Information Processing & Management*, 47(4), 575–592. doi:10.1016/j.ipm.2010.09.001

Markey, K. (2007). 25 years of end-user searching, part 2: Future research directions. *Journal of the American Society for Information Science and Technology*, 58(8), 1123–1130. doi:10.1002/asi.20601

Onwuegbuzie, A. J., & Jiao, Q. G. (2004). Information search performance and research achievement: An empirical test of the anxiety-expectation mediation model of library anxiety. *Journal of the American Society for Information Science and Technology*, 55(1), 41–54. doi:10.1002/asi.10342

Pak, R., & Price, M. M. (2008). Designing an information search interface for younger and older adults. *Human Factors*, 50(4), 614–628. doi:10.1518/001872008X312314 PMID:18767521

Sutcliffe, A., & Ennis, M. (1998). Towards a cognitive theory of information retrieval. *Interacting with Computers*, 10(3), 321–351. doi:10.1016/S0953-5438(98)00013-7

Wang, Y. D., & Forgionne, G. (2006). A decision-theoretic approach to the evaluation of information retrieval systems. *Information Processing & Management*, 42(4), 863–874. doi:10.1016/j.ipm.2005.06.005

White, R. W., Jose, J. M., & Ruthven, I. (2006). An implicit feedback approach for interactive information retrieval. *Information Processing & Management*, 42(1), 166–190. doi:10.1016/j.ipm.2004.08.010

KEY TERMS AND DEFINITIONS

Bibliographic Database: An organized collection of references to the available literature. Some primary examples include MeSH, MEDLINE, Web of Science, Scopus, etc.

Information Retrieval: An activity of tracing information from a collection of information resources and recovering back the relevant results as per needs of information seekers.

Indexing: A mechanism facilitating the retrieval of relevant information with the help of an alphabetical list of keyword terms, subject headings, etc. called as index.

Knowledge Base: A database of information that is collected, manipulated, revalidated and accumulated with the passage of time so as to aid in better informed decision making in future endeavours. Here concepts like artificial intelligence and semantic web are made use to facilitate services like reasoning, explanation, etc.

Ontology: A formal representation of knowledge as a set of concepts within a particular domain and the relations between these concepts.

Semantic Web: An extension of World Wide Web providing standardized way where information systems understand the meaning of hyperlinked information and show relationships between web pages.

World Wide Web: An information space invented by Tim Berners-Lee in 1989 and available through internet where information resources are available and identified by Universal Resource Locators (URL) and linked with each other by hypertext links. The term ‘World Wide Web’ now-a-days is synonymously used simply with ‘Web’.



Category I

IT Research and Theory

Adaptive Networks for On-Chip Communication



Mário Pereira Vestias

Instituto Politécnico de Lisboa, Portugal

INTRODUCTION

Networks-on-Chip (NoC) are a scalable inter-connection network for on-chip communication capable to integrate a high number of processing elements. Scalability, energy efficiency and reliability are among the most important advantages of this new communication paradigm. Hundreds or even thousands of cores can be integrated in a single device using a NoC structure without facing the non-scalability problems associated with bus-based structures or point-to-point connections which are usually irregular and harder to route. Global on-chip communication with long wires thus not scales down with increasing clock frequency. The new communication paradigm decouples the cores from the network, reducing the need for global synchronization, reduces the number of global wires and the energy consumption of cores can be individually controlled. NoC are also reliable since fault-tolerant techniques can be implemented from hardware redundancy to adaptive routing protocols that look for alternative paths for a communication.

The first generation of NoC solutions considers regular topologies, typically 2D meshes under the assumption that the wires' layout is well structured in such topologies. Routers and network interfaces between IP cores and routers are mainly homogeneous so that they can be easily scaled up and facilitate modular design. All advantages of a NoC infrastructure were proven with this first generation of NoC solutions.

However, soon, the designers started to be worried about the two main disadvantages associated with NoCs, namely, area and speed overhead.

Routers of a NoC need space for buffers, routing tables, switching circuit and controllers. On the other side, direct bus connection is always faster than pipelined connections through one or more routers since these introduce latency due to packaging, routing, switching and buffering.

In a first attempt to consider area and latency in the design process, designers considered that regular NoC structures may probably be adequate for general-purpose computing where processing and data communication are relatively equally distributed among all processing units and traffic characteristics cannot be predicted at design time. But, many systems developed for a specific class of applications exhibit an intrinsic heterogeneous traffic behavior. Since routers introduce a relative area overhead and increase the average communication latency, considering a homogeneous structure for a specific traffic scenario is definitely a waste of resources, a communication performance degradation and an excessive power consumption.

Application specific systems can benefit from heterogeneous communication infrastructures providing high bandwidth in a localized fashion where it is needed to eliminate bottlenecks (Benini & De Micheli, 2002), with sized communication resources to reduce area utilization, and low latency wherever this is a concern.

Homogeneous and heterogeneous solutions of first generation NoCs follow different design methodologies but have one thing in common: their architectures are found at design time and are kept fixed at runtime, i.e., the topology and the architecture of the routers are fixed at design time. Apparently, this is not a problem, but since

several applications may be running with the same NoC, the same topology and router will generally not be equally efficient in terms of area, performance and power consumption for all different applications. The efficiency of both homogeneous and heterogeneous solutions can be improved if runtime changes are considered. A system running a set of applications can benefit from the runtime reconfiguration of the topology and of the routers to improve performance, area and power consumption considering a particular data communication pattern. Customization of the number of ports, the size of buffers, the switching techniques, the routing algorithms, the switch matrix configuration, etc. should be considered in a reconfigurable NoC. Both general purpose and application-specific System-on-Chips (SoCs) will benefit from using dynamically reconfigurable NoCs since the performance and power consumption of data communication can be optimized for each application.

The second generation of NoCs are dynamic or adaptive providing a new set of benefits in terms of area overhead, performance, power consumption, fault tolerance and quality of service compared to the previous generation where the architecture is decided at design time. To improve resource efficiency and performance, the NoC must consider adaptive processes at several architectural levels, including the routing protocols, the router, the network interface and the network topology.

This article focuses on adaptive Networks-on-Chip. Adaptive topologies and adaptive routers are described in the following sections.

BACKGROUND

More than a hundred of proposals of NoC architectures can be found in the literature (Salminen, Kulmala & Hämäläinen, 2008). These NoC proposals differ in the used topology, the routing and the switching schemes, the design metrics and the target application. Routers have been also extensively studied, designed and implemented

with different flit widths, buffer sizes, switching and routing mechanisms considering latency, area, power consumption, fault-tolerance, quality-of-service, among other metrics. However, only few works have considered adaptive techniques where the NoC can be adapted to the communication requirements statically or at runtime.

Deterministic routing always uses the same route for a particular destination without considering any information about the state of the network. Adaptive routing considers the state of the network, such as the status of a link or buffer, to route data, to route data across the network. Compared to adaptive, deterministic routing requires fewer resources while guaranteeing an ordered packet arrival. On the other hand, adaptive routing provides better throughput and lower latency by allowing alternate paths. Deterministic routing is more appropriate if the traffic generated by the application is predictable, while adaptive deals better with irregular networks and/or stochastic traffic. Deterministic routing usually has poor capacity to equally distribute the traffic along all links of the network since the routes are statically assigned independently of the traffic requirements. On the other hand, highly adaptive algorithms have the potential to reach a uniform utilization of resources and provide fault tolerance. These algorithms distribute the traffic through all links to reduce congestion. However, the efficiency of highly adaptive algorithms is compromised by the necessity to guarantee deadlock free scenarios. Generally, having these algorithms adaptable requires a number of virtual channels (Bjerregaard & Mahadevan, 2006) increasing the cost of the solution compared to that using deterministic routing. Several adaptive routing methods were proposed, as described in the next section.

The proposed adaptive routing algorithms run under the same topology and router configurations, and thus the optimization that can be gained with the adaptive routing is limited by the structure of the NoC and their network components. Besides, highly adaptive routing algorithms are expensive in terms of virtual channels and, consequently,

increase the area overhead. Reducing the number of virtual channels usually means that the adaptiveness of the algorithms and hence the efficiency of the approach are reduced. Adaptable communication structures can be implemented with reconfigurable technology. Given a particular application or traffic behavior, it is possible to dynamically change the topology according to the interconnection requirements. The basic idea would consist on keeping the number and size of resources and then dynamically change the way the routers are interconnected to form the network topology.

The topology of the network and the routing mechanism to be followed are two important design optimizations to be considered in the development of a NoC. However, router optimizations are also fundamental to reduce area overhead and latency. Several aspects may be considered in the optimization process of the router: number and size of buffers, switch matrix structure and number and size of links. The optimization of the switch matrix is also very important since for specific technologies (e.g., FPGA – Field Programmable Gate Array) this is one of the most area consuming resource. Buffers also take a significant portion of the area and power consumption of a NoC (De Michelli & Benini, 2006) and so their size must be carefully minimized. Reducing the size of buffers has a negative impact over the performance of the NoC, especially when the network becomes congested. However, recent works (Martini et al., 2007; Medardoni et al., 2007) show that as the switch buffer size is decreased, the clock frequency can be increased.

Runtime switching is another technique that tries to improve the NoC by changing the switching protocol. For example, circuit switching is more appropriate to guarantee quality of service, while packet switching is more efficient for best effort traffic. Therefore, it is natural that according to the traffic type, the router dynamically changes the switching type.

According to above state-of-the-art, second generation NoCs are still in its infancy. Basic run-

time topology configuration and dynamic switching together with adaptive routing algorithms were the first attempts towards a dynamically reconfigurable NoC. Runtime router configuration with an adaptive structure is still an open issue that must be mitigated so that an integrated runtime NoC will be designed.

ADAPTIVE NETWORK-ON-CHIPS

A network-on-chip is enough to support the communication of a wide range of different applications in a single hardware platform. The NoC supports different bandwidth and quality-of-service requirements. Generically, the flexibility is achieved employing an over-engineering bandwidth capacity. The problems associated are that the NoC interconnection network takes a significant number of resources and it is in general under-utilized by a given application. Adaptive networks-on-chip try to statically or dynamically redistribute the resources and available bandwidth following the traffic requirements of each specific application. This will permit to reduce the required resources and consequently the area and power consumption of a NoC.

In the following, several adaptive techniques are analyzed including adaptive topologies and adaptive routers.

ADAPTIVE TOPOLOGIES

A few state-of-the-art proposals of dynamically reconfigurable NoCs have focused their attention into the reconfigurable topology. Stensgaard & Sparsø (2008) present a NoC architecture with a reconfigurable topology. The logical topology is configured at design time based on the communication requirements of the application. Topology switches are introduced between the network and the routers allowing links to be connected to a port of a router or to another link, whether using packet switching or circuit switching, respectively.



In the CuNoC approach (Jovanvic et al., 2007) the reconfigurable device approach is filled with small units that can establish a communication between two cores. This approach suffers from huge power consumptions and high latency due to the significant number of units in a communication path. CoNoChi (Pionteck, Koch & Albrecht, 2006) is an adaptable NoC. The reconfigurable device is divided in a matrix where each cell can hold a computational module, and a switch or a point-to-point link. This approach has major drawbacks, including high area overhead and latency. Ahmad, Erdogan & Khawam (2006) present a dynamically reconfigurable NoC, where routing, switching and data packet size can be dynamically reconfigured, but resources are fixed at design time. Kumar et al. (Kumar et al., 2007) follow a similar approach where runtime configuration reduces to a simple change of the contents of a memory. Topology, buffers and port connections are determined at design time. In (Rana et al., 2008), a runtime reconfigurable NoC is presented using an FPGA. The NoC can add or remove express lines, and perform run-time NoC topology and routing table reconfiguration to deal with traffic congestion.

Topology customization at runtime is achieved at the cost of extra area overhead and latency increase. In fact, the topology is somehow optimized for a specific runtime application permitting to use the available bandwidth more efficiently but if we consider the occupied resources the efficiency of the solution typically decreases. Reconfigurable properties of FPGA technology are used to implement adaptable topologies but still suffer from high delays to reconfigure the interconnection topology.

ADAPTIVE ROUTERS

The structure and the behavior of an adaptive router can be statically or dynamically reorganized given the quests for specific traffic bandwidth requirements. For example, a router may decide at runtime to use a second parallel link or

increase the buffer size to a specific direction. Using an adaptable structure for the router will provide better performance since its architecture adapts to the traffic requirements to achieve the best possible performance. An adaptive router tries to achieve the same performance of a static router using fewer resources and better switching and routing decisions. Reducing the number of resources (e.g., buffers, links, etc.) for packet routing is compensated by the flexibility to redirect the resources and efforts of the router to where it is most needed. Several customizations can be considered for an adaptive router (e.g., routing, switching, buffers). Runtime adaptive routers were considered in (Véstias, M. & Neto, H., 2010) where several adaptive techniques are proposed to dynamically customize a router to network traffic.

Adaptive Routing

Adaptive routing algorithms can be used to dynamically change the routes given the traffic requirements and the links' congestion. The approach is very attractive from the point of view of the compiler since the routing process is managed by the network controller. The disadvantage compared to static routing is that it needs a well-designed protocol to maintain information about the utilization of the links for the router to take good routing decisions.

An alternative is to determine the routes of each application at compile time and then change the routing tables before starting the execution of each particular application. This option is attractive for the network since it just has to support the reload of new routes, while it is harder for the compiler that must determine the best routes for each application. Compared to the approach using adaptive routing algorithms this is easier to achieve and possibly will offer better performances, but only applies to applications whose traffic pattern is known at compile time. On the other side, adaptive routing algorithms have a much broader usage since they can be used in all types of applications, whether specific or generic.

Soon, researchers have started to improve the performance of routing algorithms using adaptive techniques (Duato, 1993; Chiu, 2000). Solutions exist that base their routing decisions on the load state of the buffers of neighbor routers (Ascia, et al., 2006), (Dong et al., 2006). Another one dynamically determines its routes using the distance to destination and current link bandwidth utilization (Faruque, Ebi, Henkel, 2007). Others Hu & Marculescu (2004) have proposed a smart routing for networks-on-chip which is a combination of static and dynamic routing. The router switches between Odd-Even and XY routing using a simple congestion monitoring. XY routing was also improved to load balance the traffic (Dehyadgari et al., 2005). Pseudo adaptive XY-routing uses the deterministic XY-routing algorithm under low congestion and adaptive routing under heavy traffic conditions. A fully adaptive algorithm was proposed in (Bartic et al., 2005) using an operating system to program the routing tables. Li et al. (Li, Zeng, & Jone, 2006) proposed a dynamic XY (DyXY) fully adaptive routing scheme. Virtual channels are used to guarantee deadlock avoidance. The EDXY adaptive algorithm was proposed in (Lotfi-Kamran, P., et al., 2010) as an enhancement of the DyXY algorithm that includes a congestion-aware mechanism. In (Chang, E.-J., et al., 2015) an adaptive routing scheme is proposed that simplifies the implementation of an adaptive routing algorithm based on ant colony optimization (Hsin, H.-K. et al., 2015). In (Kumar, M. et al., 2014) an adaptive and deadlock free routing algorithm for a 2D mesh topology is presented. The algorithm uses all available minimal/non-minimal paths between source and destination nodes. A packet is routed along the non-minimal path when minimal paths get congested at the neighboring nodes.

Deterministic routing algorithms is usually the best choice for uniform or regular traffic patterns, while adaptive routing is preferable to forward non-uniform or bursty traffic or in the presence of unreliable links. Another disadvantage of adaptive routing is that packets may arrive to the

destination from different paths and with different latencies and packets may arrive out of order, requiring dedicated resources, including memory, for packet ordering.

Adaptive Switching

Switching can be dynamically changed between wormhole or virtual cut through and circuit switching to guarantee the requirements of some delay sensitive traffic. Runtime switching was considered by Wang et al. (Wang, Gu & Wang, 2007) that proposed a hybrid switch mechanism to guarantee quality of service for real-time applications. In this approach, circuit switching is used for real-time traffic and virtual cut through switching is used for best effort traffic. At runtime, data is identified according to the type of switching to be used for their commutation within each router.

The decision about the type of traffic is taken at the source of the data, i.e., the processing unit. Then, the network interface adds one bit to the header of the packet to indicate the type of traffic. When the router receives the packet it checks the type of traffic. In the case of best effort traffic, the packet is sent to the appropriate output using wormhole or virtual cut through switching. Otherwise, the router initiates a process to reserve an output port for circuit switching, namely: (1) requests a dedicated link, (2) the associated output arbiter grants the link as soon as the link is free and (3) the input releases the link at the end of the message. All the routers belonging to the path from the source to the destination processing units (circuit) do the same port reservation. The circuit is kept reserved until the end of the transmission of the message or set of messages associated with a specific flow.

Adaptive switching is usually used to support concurrent communication of best-effort traffic and real-time traffic. Circuit switching is used for real-time traffic, while wormhole or virtual cut through switching is used for other types of traffic.

Adaptive Buffering

In general, buffers are not equally utilized. Therefore, some are under-utilized while others are overutilized. Adaptive routing can alleviate somehow this unbalanced utilization but in some cases we could consider adaptive buffer resize.

Adaptive buffer resize have also been proposed for NoC adaptability. In (Al Faruque, Ebi, Henkel, 2008), the architecture uses a runtime agent to observe the presence of faults and remap tasks, reroute packets or reallocate buffers. In (Concatto, C., et al., 2009) a router can resize a buffer at run-time by moving slots buffers from one port to another. In (Matos, D., et al., 2009) buffer adaptation is based on a flow sensor that watches the traffic at run time, then a control equation adapts itself to change the buffer size of each port to achieve better performance. In (Soteriu, V., et al., 2009), the authors use a shared buffer scheme to improve the input-buffering technique. The solution statically groups pairs of inputs to each of the two buffers. In (Wang, L., 2009), a centralized buffer structure is proposed, which dynamically allocates buffer resources using linked lists. In (Véstias, M., 2011), a dynamic buffer resize technique was proposed for FPGA.

Buffer resizing is an expensive process in terms of resources. So, in general its applicability is very limited. Static resize of buffers achieves very good and efficient results but less efficient results are obtained with adaptive buffer resize due to the complexity associated with the hardware circuitry needed to dynamically resize of buffers.

FUTURE RESEARCH DIRECTIONS

There are a number of dynamic customizations that can be utilized in the design of NoC, including adaptive topologies, adaptive routing, adaptive switching, adaptive buffer resize and also adaptive network interfaces (Diguet, J.-P. 2014). Each technique achieves different tradeoffs between performance, area and power consumption.

Also, the methods are correlated, that is, using one method reduces the effectiveness of another and so the appropriate combination of techniques must be subjected to a careful analysis in order to optimize the final network-on-chip solution. The adaptability of the NoC increases the size of the design space. So, the development of efficient design methodologies and tools to help the designer explore such huge space is therefore a very important topic of research. One possible approach is to iteratively apply the customization techniques starting with those considered more efficient to improve the performance or the area.

A major open problem in the design of adaptive Networks-on-Chip is which set of adaptive techniques should be used for a specific architecture to obtain the most efficient solution in terms of area, performance, power consumption and/or other requirements. In a lower level of the design process, more efficient adaptive solutions are required either taking advantage of the technology (e.g. reconfigurability of FPGA technology) or considering new architectural structures with efficient adaptability. Considering the number of works produced within this subject, it seems that adaptive routing is the easiest the apply to a NoC. It does not mean that it is the most efficient solution. Topology optimization is in general a very efficient approach but unless it is applied statically, dynamically changing the topology is a complex process that requires some reconfigurability or adaptability of the hardware.

All adaptive customization methods rely on timely information about the congestion of the network. Typically, adaptive methods rely on neighbor information about the state of buffers and the utilization of links. While this is a simple solution it is important to research other data gathering methods relying on information beyond local information; what can be achieved in terms of performance, area and power consumption considering the costs associated with the data gathering system.

Static versus dynamic adaptation must be considered in all adaptation techniques. Static

adaptation requires knowing traffic patterns at compile time but reduces the complexity of the NoC as well as the process of gathering and analyzing traffic data at runtime. On the other side, dynamically adaptable NoCs requires dynamic traffic analysis. Static adaptation is probably most appropriate for data oriented applications, while the dynamic method is for control oriented. Which one is most appropriate for a determined application should also be explored. An hybrid static-dynamic solution should also be investigated, where some adaptable parts of the NoC are static while others are dynamic.

Many applications have real-time tasks with real-time traffic associated. Several works have proposed techniques to deal with such real-time requirements, namely: reservation of virtual channels, prioritization of traffic and circuit switching for real-time traffic. Additional aspects must be considered in the design of adaptive NoC for real-time applications: the influence of the discussed adaptation methods and how they can be used in the design of routers with support for real-time traffic.

Power consumption is an important aspect of many on-chip systems. Several works have analyzed the power consumption of NoC and how to reduce the power consumption. Similar analysis must be considered for adaptive NoC to see how the adaptation techniques affect power consumption.

Recent many-core architectures with hundreds or more cores are a challenge for adaptable NoC architectures. In these cases, hierarchical interconnection structures with different communication topologies, including buses, crossbars and point-to-point links, are a promise of better communications than just using a single NoC. These cases, require adaptable NoCs to be designed for hierarchical heterogeneous interconnection structures where adaptability may be applied at all levels or NoC may have to be reconfigurable to model different communication structures.

CONCLUSION

The article describes the backgrounds of adaptive Networks-on-Chip. Different ways and aspects of adaptability are explained and state-of-the-art implementations are referenced.

A considerable number of publications, patents and implementations have appeared during the last 50 years about adaptive interconnection networks. However, there is still work to be done on the design of adaptive NoC, to find how different techniques achieve different results, how do they influence each other and how far can we go with these adaptive techniques. Adaptive topology at low cost is an open research problem less researched when compared to adaptive routing techniques. However, adapting the topology seems to be very efficient to improve communication delay and latency. So, the research about adaptive NoC is an actual topic of research.

REFERENCES

- Ahmad, B., Erdogan, A., & Khawam, S. (2006). Architecture of a dynamically reconfigurable NoC for adaptive reconfigurable MPSoC. *The 1st NASA/ESA Conference on AHS*, 405–411. doi:10.1109/AHS.2006.25
- Al Faruque, M. A., Ebi, T., & Henkel, J. (2008). ROAdNoC: Runtime Observability for an Adaptive Network on Chip Architecture. *IEEE/ACM International Conference on Computer-Aided Design*, 543-548.
- Ascia, G., Catania, V., Palesi, M., & Patti, D. (2006). Neighbors-on-Path: A new selection strategy for On-Chip Networks. *4th IEEE/ACM/IFIP Workshop on on Embedded Systems for Real Time Multimedia*, 79-84. doi:10.1109/ES-TMED.2006.321278



- Bartic, T., Mignolet, J.-Y., Nollet, V., Marescaux, T., Verkest, D., Vernalde, S., & Lauwereins, R. (2005). Topology Adaptive Network-on-Chip Design and Implementation. *IEE Proceedings. Computers and Digital Techniques*, 152(4), 467–472. doi:10.1049/ip-cdt:20045016
- Benini, L., & De Micheli, G. (2002). Networks on Chips: A New SoC Paradigm. *Computer*, 35(1), 70–78. doi:10.1109/2.976921
- Bjerregaard, T., & Mahadevan, S. (2006). A Survey of Research and Practices of Network-on-Chip. *ACM Computing Surveys*, 38(1), 1–51, es. doi:10.1145/1132952.1132953
- Chang, E.-J., Hsin, H.-K., Chao, C.-H., Lin, S.-Y., & Wu, A.-Y. (2015). Regional ACO-Based Cascaded Adaptive Routing for Traffic Balancing in Mesh-Based Network-on-Chip Systems. *IEEE Transactions on Computers*, 64(3), 868–875. doi:10.1109/TC.2013.2296032
- Chiu, G. (2000). The odd-even turn model for adaptive routing. *IEEE Transactions on Parallel and Distributed Systems*, 11(7), 729–728. doi:10.1109/71.877831
- Concatto, C., Matos, D., Carro, L., Kastensmidt, F., Susin, A., & Kreutz, M. (2009). NoC Power Optimization Using a Reconfigurable Router. *IEEE Symposium on VLSI*, 235–240.
- De Micheli, G., & Benini, L. (2006). *Network on chips*. Morgan Kaufmann.
- Dehyadgari, M., Nickray, M., Afzali-kusha, A., & Navabi, Z. (2005). Evaluation of pseudo adaptive XY routing using an object oriented model for NOC. *Proceedings of the International Conference on Microelectronics*, 204–208. doi:10.1109/ICM.2005.1590068
- Diguet, J.-P. (2014). Self-Adaptive Network On Chips. *27th Symposium on Integrated Circuits and Systems Design*, 1–6.
- Dong, W., Al-Hashimi, B., & Schmitz, M. (2006). Improving Routing Efficiency for Network-on-Chip through Contention-Aware Input Selection. *Proceedings of the Asia and South Pacific Design Automation Conference*, 36–41.
- Duato, J. (1993). A new theory of deadlock-free adaptive routing in wormhole networks. *IEEE Transactions on Parallel and Distributed Systems*, 4(12), 1320–1331. doi:10.1109/71.250114
- Faruque, M., Ebi, T., & Henkel, J. (2007). Run-time Adaptive on-chip Communication Scheme. *Proceedings of the International Conference on Computer Aided Design*, 26–31.
- Hu, J., & Marculescu, R. (2004). DyAD – smart routing for networks-on-chip. *Proceedings of the 41st Annual Conf. Design and Automation*, 260–263.
- Jovanovic, S., Tanougast, C., Weber, S., & Bobda, C. (2007). CUNoC: A scalable dynamic NoC for dynamically reconfigurable FPGAs. *International Conference on Field Programmable Logic and Applications*, 753–756. doi:10.1109/FPL.2007.4380761
- Kumar, A., Hansson, A., Huisken, J., & Corporaal, H. (2007). An FPGA design flow for reconfigurable network-based multi-processor systems on chip. *Proceedings of the Design, Automation and Test in Europe Conference and Exhibition*, 1–6. doi:10.1109/DATE.2007.364577
- Kumar, M., Laxmi, V., Gaur, M., Ko, S.-B., & Zwolinski, M. (2014). CARM: Congestion Adaptive Routing Method for On Chip Networks. *27th International Conference on VLSI Design*, 240–245.
- Li, M., Zeng, Q.-A., & Jone, W.-B. (2006). DyXY - a proximity congestion-aware deadlock-free dynamic routing method for network on chip. *Proceedings of Design Automation Conference*, 849–852.

Lotfi-Kamran, P., Rahmani, A. M., Daneshtalab, M., Afzali-Kusha, A., & Navabi, Z. (2010). EDXY - A low cost congestion-aware routing algorithm for network-on-chips. *Journal of Systems Architecture*, 56(7), 256–264. doi:10.1016/j.sysarc.2010.05.002

Martini, F., Bertozzi, D., & Benini, L. (2007). Assessing the Impact of Flow Control and Switching Techniques on Switch Performance for Low Latency NoC Design. *First Workshop on Interconnection Network Architectures On-Chip*.

Matos, D., Concatto, C., Kolgeski, A., Carro, L., Kastensmidt, F., Susin, A., & Kreutz, M. (2009). Adaptive router architecture based on traffic behavior observability. *Proceedings of the 2nd international Workshop on Network on Chip Architectures*, 17-22. doi:10.1145/1645213.1645219

Pionteck, T., Albrecht, C., & Koch, R. (2006). A Dynamically Reconfigurable Packet-Switched Network-on-Chip. *Proceedings of the Conference on Design, Automation and Test in Europe*, 136-137.

Pionteck, T., Koch, R., & Albrecht, C. (2006). Applying partial reconfiguration to networks-on-chips. *Proceedings of International Conference on Field Programmable Logic and Applications*, 1-6.

Rana, V., Atienza, D., Santambrogio, M., Sciuto, D., & De Micheli, G. (2008). A Reconfigurable Network-on-Chip Architecture for Optimal Multi-Processor SoC Communication. *16th IFIP/IEEE International Conference on Very Large Scale Integration*, 321-326.

Salminen, E., Kulmala, A., & Hämäläinen, T. (2008). *Survey of Network-on-Chip Proposals. WHITE PAPER, OCP-IP*. Tampere University of Technology.

Soteriu, V., Ramanujam, R., Lin, B., & Peh, L.-S. (2009). A High- Throughput Distributed Shared-Buffer NoC Router. *IEEE Computer Architecture Letters*, 8(1).

Stensgaard, M., & Sparsø, J. (2008). ReNoC: A Network-on-Chip Architecture with Reconfigurable Topology. *2nd ACM/IEEE International Symposium on Networks-on-Chip*, 55-64. doi:10.1109/NOCS.2008.4492725

Véstias, M. (2011). A Dynamic Buffer Resize Technique for Networks-on-Chip on FPGA. *VII Southern Programmable Logic Conference*, 227-232. doi:10.1109/SPL.2011.5782653

Véstias, M., & Neto, H. (2010). *Dynamically Reconfigurable Networks-on-Chip using Runtime Adaptive Routers*. In *Dynamic Reconfigurable Network-on-Chip Design: Innovations for Computational Processing and Communication*. IGI Global.

Wang, K., Gu, H., & Wang, C. (2007). Study on Hybrid Switching Mechanism in Network on Chip. *Proceedings of the 7th International Conference on ASIC*, 914-917.

Wang, L., Zhang, J., Yang, X., & Wen, D. (2009). Router with Centralized Buffer for Network-on-Chip. *Proceedings of the 19th ACM Great Lakes symposium on VLSI*, 469-474. doi:10.1145/1531542.1531649

ADDITIONAL READING

Abba, S., & Lee, J.-A. (2013) “Examining the Performance Impact of NoC Parameters for Scalable and Adaptive FPGA-Based Network-on-Chips,” *Fifth International Conference on Computational Intelligence, Modelling and Simulation*, pp.364-372. doi:10.1109/CIMSim.2013.65

Al Faruque, M. A., Ebi, T., & Henkel, J. (2012). “AdNoC: Runtime Adaptive Network-on-Chip Architecture,” *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 20(2), 257–269.

- Chang, E.-J., Hsin, H.-K., Lin, S.-Y., & Wu, A.-Y. (2014). Path-Congestion-Aware Adaptive Routing With a Contention Prediction Scheme for Network-on-Chip Systems. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 33(1), 113–126. doi:10.1109/TCAD.2013.2282262
- Dahir, N., Mak, T., Al-Dujaily, R., & Yakovlev, A. (2013). Highly adaptive and deadlock-free routing for three-dimensional networks-on-chip. *IET Computers & Digital Techniques*, 7(6), 255–263. doi:10.1049/iet-cdt.2013.0029
- Ebrahimi, M. (2013). Fully adaptive routing algorithms and region-based approaches for two-dimensional and three-dimensional networks-on-chip. *IET Computers & Digital Techniques*, 7(6), 264–273. doi:10.1049/iet-cdt.2013.0034
- Hao Shu. Jiang-Yi Shi; Yue Hao; Pei-Jun Ma; Zhao Xu, (2012), “DANoC: A dynamic adaptive Network on Chip architecture,” IEEE 11th International Conference on Solid-State and Integrated Circuit Technology (ICSICT), pp.1-3.
- Hsin, H.-K., Chang, E.-J., Su, K.-Y., & Wu, A.-Y. (2015). Ant Colony Optimization-Based Adaptive Network-on-Chip Routing Framework Using Network Information Region. *IEEE Transactions on Computers*, 64(8), 2119–2131. doi:10.1109/TC.2014.2366768
- Lin, S.-C., Chang, E.-J., Chen, Y.-Y., Hsin, H.-K., & Wu, A.-Y. (2014). High performance adaptive routing for Network-on-Chip systems with express highway mechanism. *IEEE Asia Pacific Conference on Circuits and Systems*, pp.1-4. doi:10.1109/APCCAS.2014.7032704
- Mak, T., Cheung, P. Y. K., Lam, K.-P., & Luk, W. (2011). Adaptive Routing in Network-on-Chips Using a Dynamic-Programming Network. *IEEE Transactions on Industrial Electronics*, 58(8), 3701–3716. doi:10.1109/TIE.2010.2081953
- Manevich, R., Cidon, I., Kolodny, A., Walter, I., & Wimer, S. (2011) “A Cost Effective Centralized Adaptive Routing for Networks-on-Chip,” *14th Euromicro Conference on Digital System Design*, pp.39-46. doi:10.1109/DSD.2011.10
- Pratomo, I., & Sebastien, P. (2012) “Impact of design parameters on performance of adaptive Network-on-Chips,” *International Conference on High Performance Computing and Simulation (HPCS)*, pp.724-725. doi:10.1109/HPC-Sim.2012.6267003
- Rettkowski, J., & Gohringer, D. (2014). RAR-NoC: A reconfigurable and adaptive routable Network-on-Chip for FPGA-based multiprocessor systems. *International Conference on ReConfigurable Computing and FPGAs*, pp.1-6.
- Samman, F. A., Hollstein, T., & Glesner, M. (2013). Runtime Contention and Bandwidth-Aware Adaptive Routing Selection Strategies for Networks-on-Chip. *IEEE Transactions on Parallel and Distributed Systems*, 24(7), 1411–1421. doi:10.1109/TPDS.2012.200

KEY TERMS AND DEFINITIONS

Adaptive Buffering: Refers to a technique that can change static or dynamically the structure or the behavior of buffers inside a router of a network-on-chip given the traffic requirements and the links’ congestion.

Adaptive Network-On-Chip: A network-on-chip with some form of adaptability to the traffic patterns. The adaptation of the network-on-chip can be static if it is done at design time or dynamic if it is done at runtime.

Adaptive Routers: Refers to network-on-chip routers that static or dynamically adapt themselves to the traffic requirements.

Adaptive Routing: Refers to routing algorithms that can change the routes static or dynamically given the traffic requirements and the links’ congestion.

Adaptive Switching: Refers to switching algorithms that can change static or dynamically the switching method inside a router given the traffic requirements and the links' congestion.

Adaptive Topology: A network-on-chip topology whose structure and connections can be

changed static or dynamically according to the traffic requirements.

Network-On-Chip: A new communication paradigm consisting of a scalable interconnection network for on-chip communication capable to integrate a high number of processing elements.



Business Model Innovation– Oriented Technology Management for Emergent Technologies

Sven Seidenstricker

Fraunhofer Institute for Industrial Engineering, Germany

Ardilio Antonino

Fraunhofer Institute for Industrial Engineering, Germany

THE CHALLENGE IN TECHNOLOGY MANAGEMENT

During past decades the requirements of research institutions and technology-oriented companies has changed radically. Competition for customers and markets, as well as within research and development, is carried out at global level. Furthermore, the high speed dynamics of technology markets lead to constantly shorter product and technology life cycles.

Traditionally, technology transfer into the market was conducted by cost-cutting and/or innovation. However, in many cases research institutions as well as companies concentrated mainly on the improvement of their current value proposition (technology and/or product). Potential in providing new services and organisational innovations as well as innovations within the business model often remained unused. However, sticking rigidly to traditional business methods is often no longer possible and can lead to the viability and survivability of institutions or companies being endangered. In particular, research institutions in the area of applied research face this problem, as their mission is to transfer technologies into the market.

Therefore, it will be increasingly significant for research institutions and technology-oriented companies to pursue developments and trends

in industries such as markets, customer groups, technology fields, etc. and to identify growth areas and applications outside their current market and field of application. Thus, R&D effort can be reached sooner. Different applications often require several concepts of technology commercialisation. The concept can differ in terms of the number of essential partners, the value chain, the kind of value proposition, the kind of revenue achievement, etc. Technology exploitation claims a holistic consideration of all relevant issues in terms of developing new business models for the emergent technology (Nair & Paulose, 2014; Boons et al., 2013).

Unfortunately, the majority of research institutions, as well as technology-oriented companies, do not deal enough with these challenges (Bezerra, Barquet et al., 2013). Often they are not sensitised to, or are rather scared of, the laborious process behind the identification of new markets and professional technology commercialisation. The purpose of this paper is to describe some crucial steps to improve the outcome of technology management, in particular professional technology commercialisation to support institutions of applied research and technology-oriented companies and to detect white-spots in the technology and research landscape from a market perspective to increase efficiency and effectiveness in the development of emergent technologies.

OBJECTIVES OF BUSINESS MODEL INNOVATION-ORIENTED TECHNOLOGY MANAGEMENT

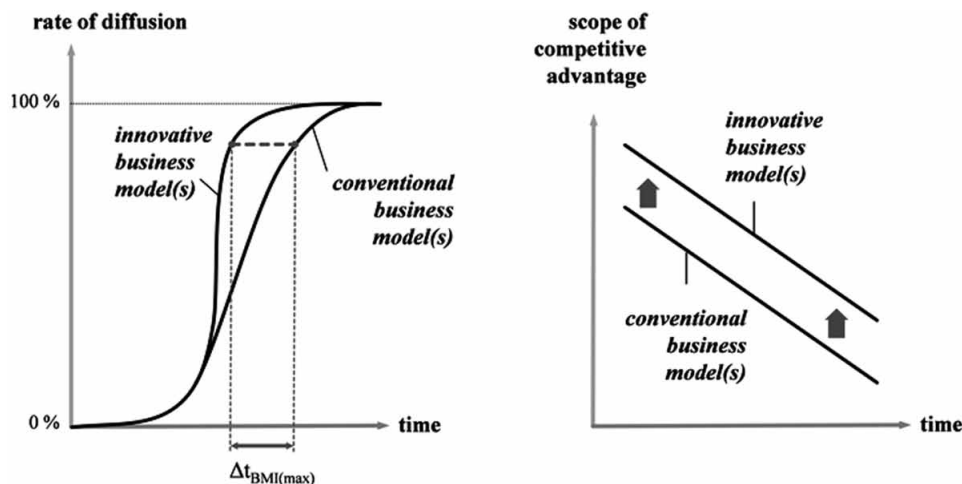
The diffusion of new technologies into the market, even if they have a large potential, is difficult. Many technologies require a period of years before they are adopted by the social system. In some cases, technologies never make it into the market (Baden-Fuller & Haefliger, 2013; Gómez & Vargas, 2012).

The diffusion of emergent technologies into the market needs to aim at increasing the diffusion rate (quantity) and decreasing the adoption time. The diffusion rate is the relative speed at which an innovation is adopted by members of a social system (Rotolo, Hicks & Martin, 2015; Planing, 2014; Rogers, 2003). A good technology diffusion rate implies a bigger market demand and, therefore, greater potential for the company or institution in terms of marketing their technology. Decreasing the adoption time brings advantages in terms of competition as competitors need to develop their technologies and/or adopt their business models in order to address the same market. However, if the gradient is too high, companies and/or institutions can face problems with the provision of the technology to the customer. In this case, the demand is much higher than existing resources

can handle. The big challenge for research institutions and companies is to achieve these objectives successfully. Smart Home Technologies or Green Technologies are examples of how commercialisation has had a huge impact on the success of technology as a whole (Karakaya, Nuur, & Hidalgo, 2016; Bohnsack, Pinkse & Kolk, 2014).

Technology commercialisation and detecting markets at an early stage of technology development has been considered in the research on technology management and for the last 10 years. At the same time, another research topic, called business model innovations, was becoming more and more important. Some leading researchers into innovations announced the idea that business model innovations were more significant than technology innovations. However, some of the business model innovations wouldn't exist without the possibilities of new technology. Nevertheless, business model innovations can boost emergent technologies, so this approach should be used to increase the diffusion rate and enhance the competitive strength of a new technology from the beginning (Figure 1). Business models should offer advantages along the value chain (not only for the end user, but also for all other stakeholders) and dissolve industry structures and market barriers (which could result in competitive advantages). Through new business models, the rate of

Figure 1. Boosting emergent technologies through business model innovations

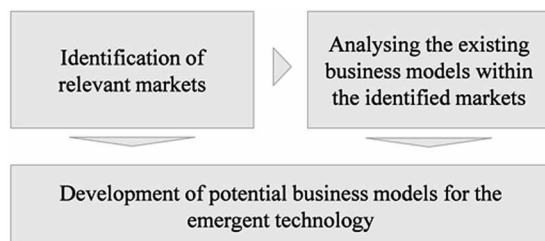


technology-diffusion of research institutions and technology-oriented companies can be increased by addressing new markets and/or improving within established markets.

HOW TO ACHIEVE THE AFOREMENTIONED OBJECTIVES?

Developing innovative business models does not exactly follow the guidelines described nor can it be predicted, but some crucial steps can be recognised in order to enhance the probability of creating a successful business model. The first step is a systematic identification of potential markets. Within these identified markets the analysis of existing business models should be the second step. Based on these two steps new attractive and competitive business models should be developed. The development of different ideas and approaches which challenge the established market leader and existing business models represents the third step. The idea development and the assessment of different ideas in the early stage of technology development allows for an influence on the research and development activities of the new technology and can be used to derive certain crucial technical requirements and adapt the technology development roadmap positively. All three steps support the increase of the diffusion rate, the decrease of the adoption time and the optimisation of the competitive advantage; these steps are shown in Figure 2.

Figure 2. Steps to develop new business models for emergent technologies



Identification of Relevant Markets

In order to be able to discover relevant markets for technologies, the current and potential applications for such technologies have to be identified. Current applications in this context means applications which are already distributed in the market for the technology under consideration, while potential applications are either addressed by a competitive technology or do not yet exist in the market.

The identification of applications for a technology is either orientated towards methods for idea generation or can be supported by intelligent patent research. In both cases, the functional profile of the technology plays a major role. The functional profile for the technology covers the whole set of functions and its performance parameters. This profile will be mapped to different branches through patent searches or a creativity session based on branch lists (like NACE-Code etc.). If the accepted function of a technology appears in patents, the associated IPC-Code gives a hint of the application field.

From the discovered applications, those which could already be addressed by the technology, and for which, therefore, no further development of the technology was needed, were dismissed from further consideration. Applications whose technological implementation was highly unlikely within the time-period under consideration were also excluded from further consideration. Outputs from this phase are made up of attractive potential applications for the technology. However, some of these markets cannot be addressed immediately by the technology due to performance or cost.

Analysing the Existing Business Models Within the Identified Markets

The determination of existing business models requires a concept that can describe and analyse 'real' business activities. There are many different business model concepts available in business model research; however, despite their differences,

which are illustrated by many papers, there is widespread recognition that business models have to be understood as the connection of many different business elements and that business models can be described and designed by using those elements.

However, there is disagreement regarding what the fundamental business elements are and how they characterise a business model. Shafer, Smith and Linder (2005) and Onetti et al. (2012) sought to compare different approaches in terms of business models and their elements to find an answer to this question. In consideration of these studies and additional concepts, which were published within the last three years, there are six essential business model elements, namely (Seidenstricker & Linder, 2014):

- Value proposition: this business model element includes all components which create value for the stakeholders of a business model. The value that a business model creates has to be formulated as precisely as possible.
 - Competencies and key resources: this includes the competencies and key resources of a company. These are all key-assets of a business model to capture value for customers and establish unique selling points on one's own behalf.
 - Channels and customer relations: the transmission of the value proposition into an offer and the distribution of this offer to the customer are the realisation of 'channels and customer relations'. The purpose of this is to support the development of customer relations.
 - Value chain and processes: on the one hand, this business model element considers the value chain of the business model referring to the industries and, on the other, the essential processes needed to accomplish the value proposition to the customer.

- Network and partners: this describes and defines the relationships to suppliers, R&D partners and other service providers needed for the business model. The customers could also be partners, particularly when they represent Lead-Users.
- Revenues: the configuration of the financial returns for services around the offer is included in 'revenues'. This element forms the economic basis for the business model.

These six business model elements will be used to establish a business model framework structured by a system theoretical approach. This implies that the business model elements can also be split into different sub elements or options. This technique will be used in the next step which provides different design options for several business model elements (Figure 3).

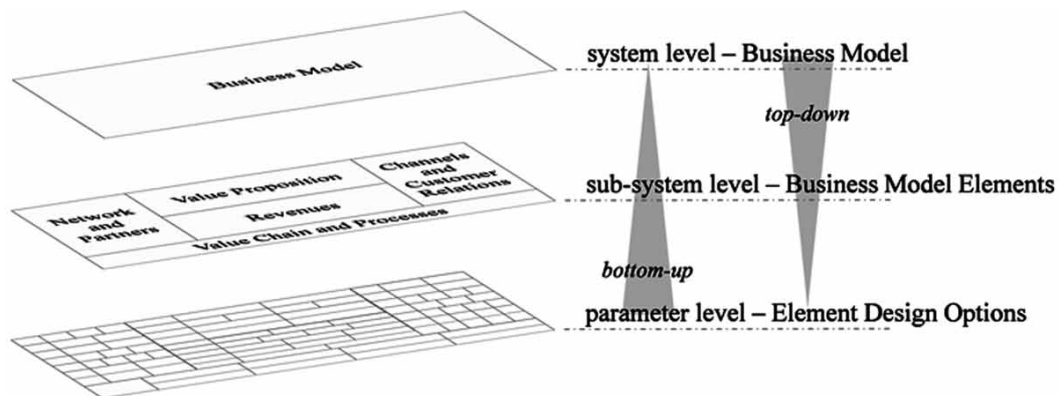
Identifying Gaps and Leverages to Create Business Model Innovations

The identification of potential markets and the analysis of existing business models should be provided for the business model idea development process, so that they can be used to detect some gaps in the existing business models and derive approaches for business model innovations. The goal in each idea creation process is to develop a huge number of different ideas without any boundaries and limitations. The complexity in the development of new business models needs to be reduced so that idea development has to be carried out for each business model element separately. Each business model element can also be split into different design options and parameter. To merge all the parts into a coherent business model, the combinations in the business models have to be verified and adjusted finely, so that bottom-up and top-down processing are applied (Figure 3).

Analog to this proceeding and to supporting the idea development as much as possible on the



Figure 3. Business Model Framework



level of sub-elements and parameters, different generic design options and parameters should be provided for each business model element. For instance, in the business model element 'revenues' five design parameters can be named each of which can contain different design options. The design options should be described generically and reflect the whole range and typical design possibilities. In the case of the business model element 'revenues', the 'kind of revenue achievement' and 'calculation base' referenced directly to the revenue achievement, and the others, covers the area of the pricing system. In all business model elements such a matrix should be established with parameter and design options. This procedure was derived from morphological

analysis, which structures the problem and splits it into its influencing variables.

After the development of different ideas in each business model element the 'sub-ideas' should be combined and merged to form consistent business model ideas. The development of business model ideas can be realised by 'mentally' combining the individual sub-ideas, similarly to the morphological analysis. The starting point for developing a business model concept logically can be any business model element. It is even advisable to vary the starting point for solution development. Furthermore, when developing solution paths, there is an additional need for other sub-ideas to become visible or more sub-ideas to be generated. If necessary, the matrices with the design

Figure 4. Design options in business model element 'revenues'

kind of revenue achievement	direct revenue achievement	indirect revenue achievement	direct and indirect revenue achievement	
calculation base	result	level	period	usage
approach of pricing	demand-oriented	cost-oriented	competitive-oriented	
price mechanism	fixed one-sided	fixed double-sided	not fixed	
price bundling	non		mixed	

areas and generic design parameters need to be considered again at this point. Further sub-ideas detected have to be added to the table. The aim of this iterative approach is to close possible gaps and to identify new sub-ideas and business model ideas, respectively.

To prioritise the developed business model ideas, all ideas can be evaluated by their potential to change the market, their competitive strength and the effort required for their implementation. A roadmap can enable the timeline to be visualized and show interdependences between business model ideas, R&D outcomes and technology requirements. By using the roadmap, R&D gaps or white-spots in the technology development can be identified, and especially for research institutes this information can be useful in adapting their technology development strategy, focusing their resources and gaining a competitive advantage in comparison to other research institutes.

FUTURE RESEARCH DIRECTIONS

Future work should continue the application of the developed approach to extend the depth and range of business model-oriented technology management. Especially in analysis of existing business models, more research is needed into which technologies are used and which would be of great assistance for research institutions in the commercialisation of their technologies.

In the idea development process, in particular during the creative phase, it might be useful to make use of some more tools, so that more radical ideas can be created. In particular, this might allow for the discovery of potential from digitalisation and how connecting different information sources can support development to create disruptive business model innovations.

The road mapping approach and the tracking of developed ideas and of other conducted studies will help feedback on the effectiveness of such proceedings to be gathered. A dynamic consideration of technology development proj-

ects to generate new business model innovations should also be the subject of further research. As a result, a set of methods within the framework of Business Model Engineering could be created to support companies and research institutions to successfully advance their technologies (at different development stages) into the market.

CONCLUSION

The concept of business model innovation is becoming significantly more important in research and practice and could also advance in technology management for emergent technologies. In terms of the commercialisation of technologies in particular, the business model concept enables a performing, holistic and efficient transfer of technologies into the market. This enables companies and research institutions to identify future potential fields of application for the technology and to secure the feedback of market information relevant to its development.

Hence, especially for emerging technologies where markets are not yet clear, essential 'vital levers' can be identified and directive recommendations concerning future developments can be given. This improves the diffusion of the technology into the market and technology competitiveness.

The business model framework allows the complexity of business models to be broken down in practice and should support the identification of leverage points for business model innovations. After such a breakdown, it is as important that all ideas and leverages are merged into consistent business models. Supporting the creative process and reducing barriers to creative thinking should also be considered during this process. Morphological boxes with design parameter and options are a useful tool to increase the quantity and quality of possible business model ideas. The three aforementioned steps cannot be completed in an exactly predefined sequence; often an iterative process is required. Nevertheless, an in-depth analysis at



the beginning is the first milestone to success in developing new disruptive business models.

Prioritising might also be an important step when focusing on important technology developments; in particular, for technology-driven small and medium-sized companies it is essential that limited resources are used with care to increase outcomes. Tracking and adapting ideas and concepts as well as the progress of implementation and feedback from the market are quite important. Roadmapping and Project Management Tools could be helpful to stay on track and push towards a realisation of the developed business model concepts.

This paper is focused mainly on the generation and evaluation of new business model ideas for emergent technologies. However, it is just the first step – but nevertheless an important component – in advancing the diffusion of a technology and maximising the scope of its competitive strength from the beginning. It should be helpful for scientists and engineers involved in technology development to keep these three steps in mind and to aim to benefit the whole organisation or company.

REFERENCES

- Baden-Fuller, C., & Haefliger, S. (2013). Business Models and Technological Innovation. *Long Range Planning*, 46(6), 419–426. doi:10.1016/j.lrp.2013.08.023
- Bezerra Barquet, A. P., de Oliveira, M. G., Román Amigo, C., Pinheiro Cunha, V., & Rozenfeld, H. (2013). Employing the business model concept to support the adoption of product–service systems. *Industrial Marketing Management*, 42(5), 649–655.
- Bohnsack, R., Pinkse, J., & Kolk, A. (2014). Business models for sustainable technologies: Exploring business model evolution in the case of electric vehicles. *Research Policy*, 43(2), 284–300. doi:10.1016/j.respol.2013.10.014
- Boons, F., Montalvo, C., Quist, J., & Wagner, M. (2013). Sustainable innovation, business models and economic performance: An overview. *Journal of Cleaner Production*, 45(4), 1–8. doi:10.1016/j.jclepro.2012.08.013
- Gómez, J., & Vargas, P. (2012). Intangible resources and technology adoption in manufacturing firms. *Research Policy*, 41(9), 1607–1619. doi:10.1016/j.respol.2012.04.016
- Karakaya, E., Nuur, C., & Hidalgo, A. (2016). Business model challenge: Lessons from a local solar company. *Renewable Energy*, 85(1), 1026–1035. doi:10.1016/j.renene.2015.07.069
- Nair, S., & Paulose, H. (2014). Emergence of green business models: The case of algae bio-fuel for aviation. *Energy Policy*, 65(2), 175–184. doi:10.1016/j.enpol.2013.10.034
- Onetti, A., Zucchella, A., Jones, M. V., & McDougall-Covin, P. P. (2012). Internationalization, innovation and entrepreneurship: Business models for new technology-based firms. *Journal of Management & Governance*, 16(3), 337–368. doi:10.1007/s10997-010-9154-1
- Planing, P. (2014). *Innovation Acceptance: The Case of Advanced Driver-Assistance Systems*. Berlin: Springer Gabler. doi:10.1007/978-3-658-05005-4
- Rahim, N. A., Mohamed, Z. B., & Amrin, A. (2015). Commercialization of Emerging Technology: The Role of Academic Entrepreneur. *Procedia: Social and Behavioral Sciences*, 169(1), 53–60. doi:10.1016/j.sbspro.2015.01.285
- Rogers, E. M. (2003). *Diffusion of Innovations*. New York: Free Press.
- Rotolo, D., Hicks, D., & Martin, B. R. (2015). What is an emerging technology? *Research Policy*, 44(10), 1827–1843. doi:10.1016/j.respol.2015.06.006

Seidenstricker, S., & Linder, C. (2014). A morphological analysis-based creativity approach to identify and develop ideas for BMI: A case study of a high-tech manufacturing company. *International Journal of Entrepreneurship and Innovation Management*, 18(5-6), 409–424. doi:10.1504/IJEIM.2014.064716

Shafer, S. M., Smith, H. J., & Linder, J. C. (2005). The power of business models. *Business Horizons*, 48(3), 199–207. doi:10.1016/j.bushor.2004.10.014

ADDITIONAL READING

Albar, F. (2011). *Technology assessment: Forecasting future adoption of emerging technologies*. Berlin: Erich Schmidt.

Baden-Fuller, C., & Morgan, M. S. (2010). Business models as models. *Long Range Planning*, 43(2/3), 156–171. doi:10.1016/j.lrp.2010.02.005

Balachandra, P., Nathan, H. S. K., & Sudhakar Reddy, B. (2010). Commercialization of sustainable energy technologies. *Renewable Energy*, 35(8), 1842–1851. doi:10.1016/j.renene.2009.12.020

Bell, F., Fletcher, G., Greenhill, A., Griffiths, M., & McLean, R. (2014). Making MadLab: A creative space for innovation and creating prototypes. *Technological Forecasting and Social Change*, 84(5), 43–53. doi:10.1016/j.techfore.2013.09.004

Bode, A., Castro-Arce, J., Heida, B., Henschel, C., Wechsung, A., & Wojcicka, J. (2015). Process Technology Licensing: An Interface of Engineering and Business. *Computer Aided Chemical Engineering*, 37, 93–100. doi:10.1016/B978-0-444-63578-5.50012-8

Budde, L., Nagler, O., & Friedli, T. (2015). A Method to Set up a Complexity Index to Improve Decision-making Performance. *Procedia CIRP*, 36, 53–58. doi:10.1016/j.procir.2015.01.052

Calvo-Mora, A., Navarro-García, A., & Periañez-Cristobal, R. (2015). Project to improve knowledge management and key business results through the EFQM excellence model. *International Journal of Project Management*, 33(8), 1638–1651. doi:10.1016/j.ijproman.2015.01.010

Chesbrough, H. W. (2010). Business Model Innovation. Opportunities and Barriers. *Long Range Planning*, 43(2/3), 354–363. doi:10.1016/j.lrp.2009.07.010

Copani, G., & Urgo, M. (2015). Innovative flexibility-oriented business models and system configuration approaches: An industrial application. *CIRP Journal of Manufacturing Science and Technology*, 11(11), 10–17. doi:10.1016/j.cirpj.2015.05.008

Dedehayir, O., Nokelainen, T., & Mäkinen, S. J. (2014). Disruptive innovations in complex product systems industries: A case study. *Journal of Engineering and Technology Management*, 33(3), 174–192. doi:10.1016/j.jengtecman.2014.06.002

Desyllasa, P., & Sako, M. (2013). Profiting from business model innovation: Evidence from Pay-As-You-Drive auto insurance. *Research Policy*, 42(1), 101–116. doi:10.1016/j.respol.2012.05.008

Donmeza, B., Matsonb, Z., Savanb, B., Farahanib, E., Photiadisb, D., & Dafoeb, J. (2014). Interruption management and office norms: Technology adoption lessons from a product commercialization study. *International Journal of Information Management*, 34(6), 741–750. doi:10.1016/j.ijinfomgt.2014.08.003

Eyring, M. J., Johnson, M. W., & Nair, H. (2011). New Business Models in Emerging Markets. Targeting the middle market can be lucrative - but companies won't be able to deliver unless they start from scratch. *Harvard Business Review*, 89(1/2), 89–95.



- Funkhouser, E., Blackburn, G., Magee, C., & Rai, V. (2015). Business model innovations for deploying distributed generation: The emerging landscape of community solar in the U.S. *Energy Research & Social Science*, 10(11), 90–101.
- Geroski, P. A. (2000). Models of technology diffusion. *Research Policy*, 29(4/5), 603–625. doi:10.1016/S0048-7333(99)00092-X
- Guo, Y., Huang, L., & Porter, A. L. (2010). The research profiling method applied to nano enhanced, thin film solar cells. *R & D Management*, 40(2), 195–208. doi:10.1111/j.1467-9310.2010.00600.x
- Johnson, M. W., Christensen, C. M., & Kagermann, H. (2008). Reinventing Your Business Model. *Harvard Business Review*, 86(12), 50–59. PMID:18681297
- Klenner, P., Hüsigg, S., & Dowling, M. (2013). Ex-ante evaluation of disruptive susceptibility in established value networks - When are markets ready for disruptive innovations? *Research Policy*, 42(4), 914–927. doi:10.1016/j.respol.2012.12.006
- Landry, R., Amara, N., Cloutier, J.-S., & Halilem, N. (2013). Technology transfer organizations: Services and business models. *Technovation*, 33(12), 431–449. doi:10.1016/j.technovation.2013.09.008
- Lichtenthaler, U., Lichtenthaler, E., & Frishammer, J. (2009). Technology commercialization intelligence: Organizational antecedents and performance consequences. *Technological Forecasting and Social Change*, 76(3), 301–315. doi:10.1016/j.techfore.2008.07.002
- Lubik, S., Lim, S., Platts, K., & Minshall, T. (2013). Market-pull and technology-push in manufacturing start-ups in emerging industries. *Journal of Manufacturing Technology Management*, 24(1), 10–27. doi:10.1108/17410381311287463
- Oo, C., Tsai, J.-C., & Kao, H. D. (2015). There is no better time than the present: Nanotechnology as a disruptive innovation for drug development. *Drug Discovery Today*, 20(6), 645–647. doi:10.1016/j.drudis.2015.03.013 PMID:25835226
- Osterwalder, A., & Pigneur, Y. (2010). *Business Model Generations. A Handbook for Visionaries, Game Changers and Challengers*. New Jersey: John Wiley & Sons.
- Reinhardt, R., & Gurtner, S. (2015). Differences between early adopters of disruptive and sustaining innovations. *Journal of Business Research*, 68(1), 137–145. doi:10.1016/j.jbusres.2014.04.007
- Souto, J. E. (2015). Business model innovation and business concept innovation as the context of incremental innovation and radical innovation. *Tourism Management*, 51(12), 142–155. doi:10.1016/j.tourman.2015.05.017
- Teece, D. J. (2010). Business Models, Business Strategy and Innovation. *Long Range Planning*, 43(2/3), 172–194. doi:10.1016/j.lrp.2009.07.003
- Watson, D. (2005). *Business models. Investing in companies and sectors with strong competitive advantage*. Hampshire: Harriman House Press.
- Zhang, J., Gang, H., & Jianwen, Y. (2010). *Business model innovation. From Customer Value Perspective*. Paper presented at 3rd International Conference on Information Management, Innovation Management and Industrial Engineering. Kunming. doi:10.1109/ICIII.2010.577
- Zott, C., & Amit, R. (2010). Business Model Design: An Activity System Perspective. *Long Range Planning*, 43(2/3), 216–226. doi:10.1016/j.lrp.2009.07.004
- Zott, C., Amit, R., & Massa, L. (2011). The Business Model: Recent Developments and Future Research. *Journal of Management*, 37(4), 1019–1042. doi:10.1177/0149206311406265

KEY TERMS AND DEFINITIONS

Business Model: Business model describes the logic of the value generation for customers, stakeholders and the company or research institution.

Business Model Innovation: Business model innovation is a new way to generate value for customers and stakeholders and enables the establishment of competitive advantage for the company or research institution.

Business Model-Oriented Technology Management: Business model-oriented technology management integrates the concept of business model innovation, which means identifying, developing and analysing business concepts, in the management of technologies and technology knowledge.

Diffusion: Diffusion describes the accumulated level of a technology in the market.

Diffusion Rate: Diffusion rate refers to the current quantitative scope of a technology's diffusion.

Emergent Technology: Emergent technology is new and coming to the fore.

Technology: Technology uses physical principles and scientific laws to realise technical functions.

Technology Management: Technology management covers all aspects of organising technology knowledge, technology forecasting, technology development, technology commercialisation and technology usage through the technology life-cycle.



Cognitive Mapping in Support of Intelligent Information Systems

Akbar Esfahanipour

Amirkabir University of Technology, Iran

Ali Reza Montazemi

McMaster University, Canada

INTRODUCTION

Cognitive mapping techniques consist of a set of procedures to capture perceived relationships of attributes related to ill-structured decision problems that decision makers have to face. This paper provides an overview of the application of cognitive maps (CMs) in the design and development of intelligent information systems. Here, CM is used as a set of techniques to identify subjective beliefs and to portray those beliefs and their relationships externally as follows:

- Causal mapping is used to investigate the cognition of decision-makers. A causal map represents a set of causal relationships (i.e., cause and effect relationships) among constructs within a system. For example, Figure 1 shows that better sanitation facilities, causing an initial improvement in health, led to an increase in the city's population. This growth led to more garbage, more bacterial, and therefore more disease. Causal map aids: 1) in identification of irrelevant data, 2) to evaluate the factors that affect a given class of decisions, and 3) enhances the overall understanding of a decision maker's environment, particularly when it is ill-structured.
- Semantic mapping, also known as *idea mapping*, is used to explore an idea without the constraints of a superimposed structure. A semantic map visually organizes related

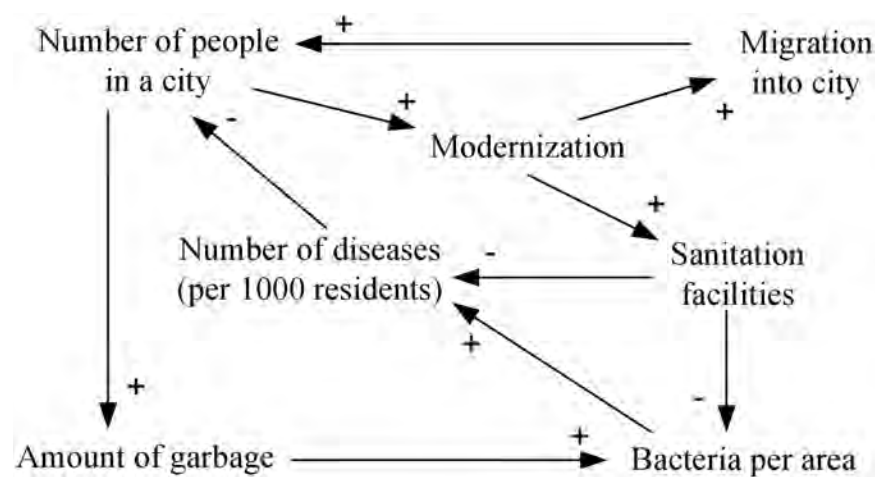
concepts around a main concept with tree-like branches. Figure 2 depicts different types of transportation, organized in three categories: land, water, and air. This technique facilitates communication between end-users and system analysts in support of information requirements analysis.

- Concept mapping is a useful tool for organizing and representing concepts (events or objects) and their interrelationships in a particular domain. Each concept is designated with a label. The relationship between two concepts in a concept map is referred to as a proposition; propositions connect concepts to form a meaningful statement. Relationships between concepts are associative. For example, in Figure 3, two concepts of "plants" and "flowers" are associated via "may have" that form the proposition of "plants may have flowers." Describing complex structures with simple propositions improve quality of conceptual modeling in the development of information systems.

BACKGROUND

Cognitive Map (CM) has been employed to capture, store and retrieve expert knowledge in support of the design and development of intelligent information systems. CM is a representation of the relationships that are perceived to exist among the elements of a given environment. Taking any

Figure 1. Causal map for public health issues



two of these elements, the concern is whether the state or movement of the one is perceived to have an influence on the state or movement of the other (both static and dynamic relationships can be considered) (Montazemi & Conrath, 1986). CMs have been used to describe experts' tacit knowledge about a certain problem, particularly in ill-structured decision problems (Axelrod, 1976; Montazemi & Chan, 1990; Amer et al. 2015). Tacit knowledge is personal knowledge, shared

and exchanged through direct and face-to-face contact among actors (Eden, 1988).

There are different perspectives of knowledge within organizations (Nonaka, 1994). Thus, it seems appropriate to use knowledge management categories to identify different applications of cognitive map in the design and development of intelligent information systems. Alavi and Leidner (2001) provide a framework that is grounded in the sociology of knowledge and is based on the

Figure 2. Semantic map for different types of transportation

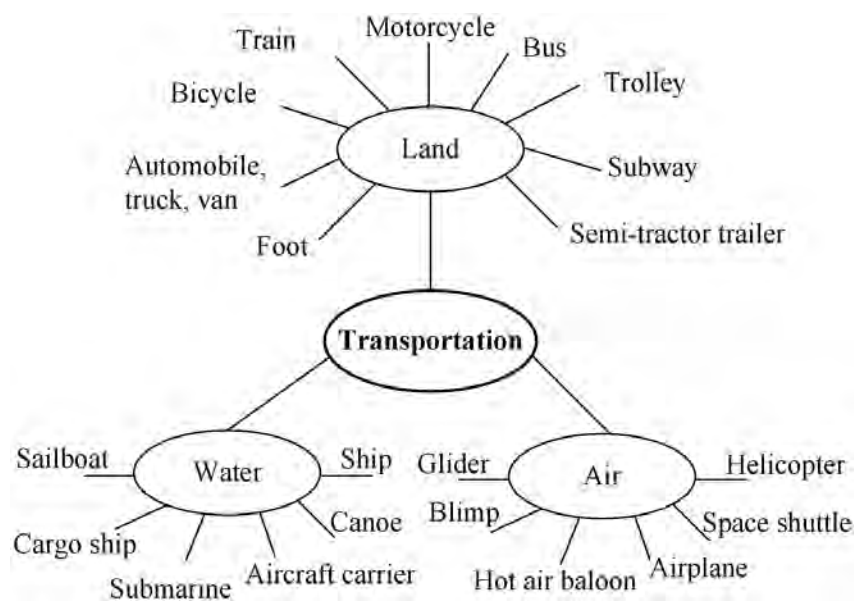
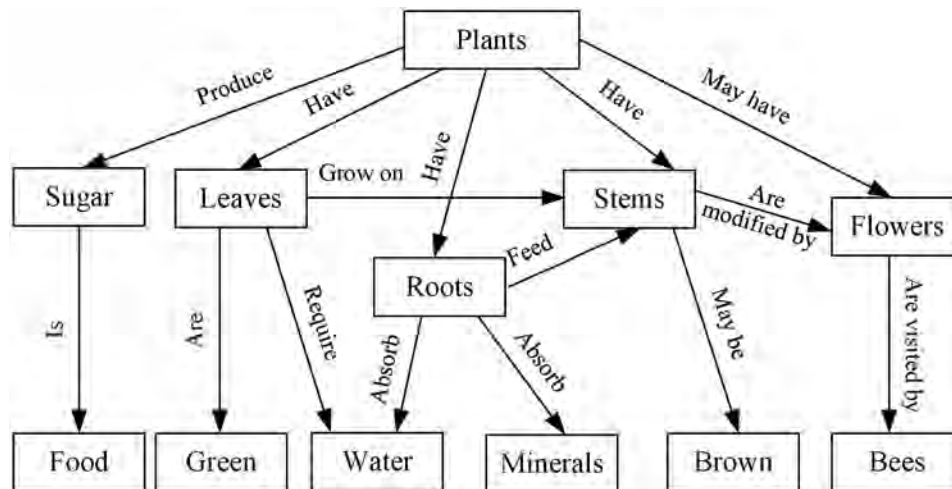


Figure 3. Concept map for plants



view of organizations as social collectives and “knowledge systems.” They contend that organizations as knowledge systems consist of four sets of socially enacted “knowledge processes” as follows:

- **Knowledge Application:** Those activities concerned with deploying knowledge in order to produce goods and services. Information technology can enhance knowledge application by facilitating capture, updating and accessibility of organizational directives.
- **Knowledge Storage/Retrieval:** This is also referred to as *organizational memory*, which includes knowledge residing in various component forms, including written documentation, structured information stored in electronic databases, codified human knowledge stored in expert systems, and tacit knowledge acquired by individuals.
- **Knowledge Transfer:** Transfer of knowledge to locations where it is needed and can be used. This can occur at various levels: transfer of knowledge between individuals, from individuals to explicit sources, from individuals to groups, between

groups, across groups, and from groups to the organization.

- **Knowledge Creation:** Developing new content or replacing existing content within the organization’s tacit and explicit knowledge. Through social and collaborative processes as well as through the cognitive processes of the individual, knowledge is created, shared, amplified, enlarged and justified in organizational settings.

A literature search shows that application of CMs in intelligent information systems can be found in support of the above four categories of knowledge processes, as depicted in the Appendix, Table 1. A brief description of each of the above four categories within the context of CMs is presented next.

KNOWLEDGE PROCESS THROUGH CMs

Knowledge Application

Cognitive mapping techniques (e.g., causal mapping, semantic mapping and concept mapping) have been used to improve the processes of the

design and development of information systems. They include improvement of conceptual modeling (Siau & Tan, 2005a), user–database interaction (Siau & Tan, 2006) and applicability of the resulting information systems (Siau & Tan, 2005b). Siau and Tan (2006) proposed a framework of user–database interaction in support of knowledge application. According to this framework, user–database interaction has three dimensions -- content, structure and style. The content (i.e. semantics) dimension is determined by the user's data needs. The structure dimension is determined by both the user's perceived data model (constituent structure) and by the query language (syntactic structure). The style dimension is dictated by user–database interface.

CM techniques have been applied in diverse areas for knowledge processing. For example, Papageorgiou et al. (2012) formalized medical knowledge using Fuzzy Cognitive Map (FCM) and semantic web approach in support of therapy decision making in medicine treatment. Here, FCMs are applied for the case of modeling and knowledge integration of clinical practice guidelines in the form of if-then fuzzy rules, and used the semantic web tools to implement the FCMs. Their semantic web approach, to formalize the FCMs, is based on two notions: 1) to develop a common model of integration and use of data from disparate sources; and 2) to link data with real world objects, for machines and humans to access data on different bases related by semantic relations. FCM techniques have also been used to develop diagnosis decision support systems based on clinical guidelines in the form of case-based FCMs (Douali et al., 2015). In another application of FCMs, Amer et al. (2015) develop a national-level wind energy roadmap through scenario planning in which multiple future scenarios are developed using FCMs. Also, Kim et al. (2016) suggest a FCM-based approach for scenario building from expert opinions.

CM techniques can also help decision makers in design of risk management strategies. For example, risk managers are interested in

incorporating stakeholder beliefs into their risk assessment planning process. Techniques for diagramming stakeholder mental models enable risk managers to better understand stakeholder beliefs and perceptions concerning risk in support of their risk management strategies. These risk management (RM) techniques can be categorized as: decision-analysis-based mental modeling, concept mapping, and semantic web analysis (Wood et al., 2012). For example, Van Winsen et al. (2013) propose a CM technique, combining the three types of RM techniques, to identify and represent risk perception faced by farmers. Lopez & Salmeron (2014) develop an FCM, i.e., the first RM technique, in support of ERP maintenance risk management. They modeled ERP maintenance project outcomes and risks perceptions. Aju Kumar et al. (2015) propose an FCM technique to assess human reliability in plant maintenance environment.

Knowledge Storage/Retrieval

Cognitive mapping techniques have been also applied to acquire a decision maker's tacit knowledge when faced with ill-structured decision problems such as the extraction of expert knowledge to forecast stock prices (Montazemi & Chan, 1990), the building of expert systems to control electronic data interchange (Lee & Lee, 2007), for negotiation in a B2B system (Lee & Kwon, 2006) and for analysis of firms with regard to their creditworthiness (Noh et al., 2000). CMs have been used to develop a specific method to analyze complex problems. For example, Satur and Liu (1999) developed an inference method in support of geographic information systems and Kwon et al. (2002) proposed a structural analysis method for a DSS in support of urban planning. Noh et al. (2000) adopted CMs to formalize tacit knowledge and proposed case-based reasoning as a tool for storage/retrieval of CM-driven tacit knowledge in the form of frame-typed cases. Lee et al. (2013) applied fuzzy cognitive maps (FCM) to industrial marketing planning. They proposed



an agent based inference method to overcome dynamic relationships, time lags, and reusability issues of FCM evaluation.

Lu et al. (2014) proposed a modeling and prediction approach of time series based on synergy of high-order fuzzy cognitive map and fuzzy c-means clustering. They contend that this approach ameliorate the prediction problem of large-scale time series. This is extraction of explicit knowledge from historical time series data that can be stored/retrieved in the form of CMs.

Knowledge Transfer

Cognitive maps have been applied in group decision making for the exchange of ideas and opinions between group members in regard to a specific problem. For instance, Caliskan (2006) used the CM technique to facilitate the evaluation of transport investment alternatives using a group of experts. Sengupta and Te'eni (1993) studied the impact of cognitive feedback on group decision making at three levels: individual, interpersonal and collective. They contend that for the development of a group decision support system (GDSS), cognitive feedback should be considered as an integral part at every level, and that a human-computer interaction should be designed as an effective transition across the components of feedback at all levels. Khan and Quaddus (2004) present a methodology for the development and analysis of FCM as a GDSS tool. This methodology consists of two phases: development and application. In the development phase, an FCM is created for each group member and these are then merged to produce a group FCM. The application phase consists of static analysis of concepts and causal relationships, and dynamic analysis of the simulated system over time. Jetter and Sperry (2013) propose an FCM methodology in support of stakeholder engagement in product planning. Their method helps product planners to systematically capture, understand, and assess stakeholder needs and their interdependencies. Their approach consist

of six basic steps for organizing FCM modeling projects: Clarification of project objectives and information needs (Step 1); Plans for knowledge elicitation (Step 2); Knowledge capture (Step 3); Conceptual (Step 4) and detailed (Step 5) design of FCM; Test, Interpretation, and Validation of model results (Step 6). Jetter & Kok (2014) contend that such a six-step FCM modeling framework would be applicable in future studies to better integrate expert, stakeholder, and indigenous knowledge by creating scenarios that bridge the gap between quantitative analysis and qualitative story lines.

FCMs have been also used to develop a fuzzy cognitive agent to provide personalized recommendations to on-line customers (Miao et al., 2007). The agent learns users' preferences from the most recent cases and helps customers to make inferences and decisions. Xirogiannis and Glykas (2007) propose the application of the fuzzy causal characteristics of FCMs to generate a hierarchical and dynamic network of interconnected maturity indicators in e-business strategy formulation exercises. To improve semantic inference, Lee & Kwon (2014) propose an ontological semantic inference method that combines the CM and semantic influence. This approach reuses a pre-defined cognitive map and provides an ontological semantic inference mechanism in decision making environments by reducing the degree of cognitive complexities in a large cognitive map. Here, the predefined knowledge, which is represented as a CM, is processed through semantic inference method to enhance decision making.

Knowledge Creation

CM techniques have been applied for knowledge creation. Nonaka (1994) define the four "modes" of knowledge creation as 1) socialization: conversion of tacit knowledge to new tacit knowledge through interaction of individuals, sharing ideas and experiences and learning from each other; 2) externalization: conversion of tacit knowledge to new explicit knowledge such as articulation of best practices or lessons learnt; 3) combination:

conversion of explicit knowledge to new explicit knowledge by merging, categorizing, reclassifying and synthesizing existing explicit knowledge; 4) internalization: conversion of explicit knowledge to new tacit knowledge that could be achieved through the learning and understanding that result from reading about a specific topic.

Given that tacit knowledge is difficult to formalize and communicate (Polanyi, 1966), socialization, which occurs through conversion of such tacit knowledge to a new form of tacit knowledge, is a complex task to model. As a result, CMs have not been applied directly to model socialization mode of knowledge creation. However, the four aforementioned knowledge creation modes are interdependent and intertwined so that each mode relies on, contributes to, and benefits from the other modes (Alavi & Leidner, 2001). Hence, it is possible to use one or more intermediate conversion steps to model socialization using CMs. That is, conversion of tacit knowledge to a new explicit knowledge (i.e. externalization) and then converting the obtained explicit knowledge to a new tacit knowledge (i.e. internalization). We may need to use a combination mode before internalization, however.

CM allows modeling of mental models of decision makers, and it leads to clarification and structuring of the experts' thought processes when faced with ill-structured problems (Rodhain, 1999). Thus, CMs can be used as a means of externalization. For example, Rodhain (1999) uses CMs to extract business strategy mental models of managers for determining a portfolio of projects. Other pertinent applications include application of CMs to extract experts' tacit knowledge in order to forecast stock prices (Montazemi & Chan, 1990) and to analyze information requirements (Montazemi & Conrath, 1986).

FCM has been used in group decision making to construct a causal knowledge base in stock investment analysis (Lee & Kim, 1997), in electronic data interchange (EDI) controls (Lee & Lee, 2007), in technology road mapping (Amer

et al. 2015), and in scenario building (Kim et al., 2016) as a means of externalization, so that the resulting CMs could be transferred, managed and stored as explicit knowledge.

Individuals' CMs could be different in a specific domain because perceived attributes and/or perceived causal relationships of CMs may be different from one person to another. These differences between CMs, which are called *content difference* (Langfield-Smith & Wirth, 1992) exist because of different levels of individual experience and expertise. Langfield-Smith and Wirth (1992) propose quantitative measures for analyzing content difference between two or more CMs. These measures would assist in providing a more objective basis for qualitative analysis. Eden (2004) discusses the use of CMs as a qualitative method for structuring ill-structured decision problems. Wang (1996) uses neural network as a quantitative method to measure differences between CMs of individuals over time. The comparisons of CMs provide useful information for a decision maker when considering individual points of view. Providing such information can be considered as combination, and analyzing of this information by decision makers to get an insight into a given problem can be considered as the internalization mode of knowledge creation. León et al (2013) propose a model for analyses learning, clustering and aggregation of FCM of decision makers' cognitive change at different levels of abstraction.

Learning methods have been used to improve knowledge represented by means of FCMs. These learning methods include the non-linear Hebbian learning rule (Papageorgiou et al., 2003), genetic learning method (Stach et al., 2005), unsupervised learning technique (Papageorgiou et al., 2006), and Imperialist Competitive Learning Algorithm (ICLA) (Ahmadi et al., 2015). Improving the embedded knowledge of CMs through machine learning methods can be considered as combination mode of knowledge creation that can improve the user's tacit knowledge (i.e., internalization).



For example, Lee et al. (2002) propose a three-phased web-mining inference amplification (WEMIA) based on inference logic of FCM. The first phase is used to extract association rules, and based on the extracted rules in the second phase, corresponding FCMs are developed. The advantages of causal knowledge over association rules are that it can provide a more refined inference mechanism than can the association rules (Lee et al., 2002). The final phase is used to apply inference amplification procedures to the causal knowledge: to eliminate rule redundancy, to search for the directly and indirectly chained rules, and to amplify inferences about the logical relationships between rules. In this case, knowledge is created thanks to the conversion of some explicit knowledge (i.e. association rules) to another form of explicit knowledge (i.e. FCM), which is referred to as *knowledge combination*. Here, knowledge is improved by the use of inference methods to derive rich semantics which are suitable for and understandable by the decision maker. This is referred to as the *internalization mode of knowledge creation*.

Thus, CM techniques can be applied directly toward externalization and combination modes of knowledge creation and can contribute indirectly toward internalization and eventually to the socialization mode of knowledge creation.

FUTURE RESEARCH DIRECTIONS

This literature review shows diverse use of CM in the varied domain of decision making processes under uncertainty. We expect increased sophistication in capturing CM and its application in support of intelligent information systems. In particular, CM can play a major role in realizing the usefulness of the Semantic Web, which is the outgrowth of many diverse desires and influences, all aimed at making better use of the Web as it stands. The Semantic Web is portrayed as: 1) a universal library, to be readily accessed and

used by humans in a variety of information use contexts; 2) the backdrop for the work of computational agents completing sophisticated activities on behalf of their human counterparts; and 3) a method for federating particular knowledge bases and databases to perform anticipated tasks for humans and their agents. Intelligent agents can use CM of users to filter information from the Semantic Web, making it possible to manage organizational knowledge using Intranet and the vast resources available from the Internet.

CONCLUSION

Cognitive mapping techniques have been applied in design and development of intelligent information systems as a means of acquiring knowledge from domain experts, manipulating knowledge to create new contents, storing/retrieving knowledge for later use, and sharing knowledge between group members to deal with ill-structured decision problems. Successful applications of cognitive mapping methodologies reveal the efficiency and flexibility of these techniques, which can be potentially utilized by developers of intelligent information systems as follows:

- Providing a rich picture of ill-structured problems for decision makers,
- Facilitating group decision making through clear exchange of ideas and information,
- Dynamic systems in uncertain environments using fuzzy cognitive maps,
- Providing cognitive feedbacks to improve decision makers' decision processes when faced with ill-structured decision problems,
- Improving quality of conceptual modeling of systems, user-system interactions and usability of the resulting systems, and
- Better identifying end-users' information requirements and improve user-developer communications.

REFERENCES

- Aguilar, J. (2002). Adaptive random fuzzy cognitive maps. In F. J. Garijio, J. C. Riquelme, & M. Toro (Eds.), *IBERAMIA 2002, Lecture Notes in Artificial Intelligence 2527* (pp. 402–410). Berlin: Springer-Verlag.
- Ahmadi, S., Forouzideh, N., Alizadeh, S., & Papageorgiou, E. I. (2015). Learning Fuzzy Cognitive Maps using Imperialist Competitive Algorithm. *Neural Computing & Applications*, 26(6), 1333–1354. doi:10.1007/s00521-014-1797-4
- Aju Kumar, V. N., Gandhi, M. S., & Gandhi, O. P. (2015). Identification and Assessment of Factors Influencing Human Reliability in Maintenance Using Fuzzy Cognitive Maps. *Quality and Reliability Engineering International*, 31(2), 169–181. doi:10.1002/qre.1569
- Alavi, M., & Leidner, D. E. (2001). Review: knowledge management and knowledge management systems: conceptual foundations and research issues. *Management Information Systems Quarterly*, 25(1), 107–136. doi:10.2307/3250961
- Amer, M., Daim, T. U., & Jetter, A. (2015). Technology roadmap through fuzzy cognitive map-based scenarios: The case of wind energy sector of a developing country. *Technology Analysis and Strategic Management*. doi:10.1080/09537325.2015.1073250
- Axelrod, R. (1976). *Structure of decision*. Princeton, NJ: Princeton University Press.
- Caliskan, N. (2006). A decision support approach for the evaluation of transport investment alternatives. *European Journal of Operational Research*, 175(3), 1696–1704. doi:10.1016/j.ejor.2005.02.035
- Douali, D., Csaba, H., De Roo, J., Papageorgiou, E. I., & Jaulent, M. C. (2015). Diagnosis Support System based on clinical guidelines: Comparison between Case-Based Fuzzy Cognitive Maps and Bayesian Networks. *Computer Methods and Programs in Biomedicine*, 113(1), 133–143. doi:10.1016/j.cmpb.2013.09.012 PMID:24599907
- Eden, C. (1988). Cognitive mapping: A review. *European Journal of Operational Research*, 36(1), 1–13. doi:10.1016/0377-2217(88)90002-1
- Eden, C. (2004). Analyzing cognitive maps to help structure issues or problems. *European Journal of Operational Research*, 159(3), 673–686. doi:10.1016/S0377-2217(03)00431-4
- Jetter, A. J., & Kok, K. (2014). Fuzzy Cognitive Maps for futures studies-A methodological assessment of concepts and methods. *Futures*, 61, 45–57. doi:10.1016/j.futures.2014.05.002
- Jetter, A. J., & Sperry, R. C. (2013). Fuzzy Cognitive Maps for Product Planning: Using Stakeholder Knowledge to Achieve Corporate Responsibility. In *Proceedings of the 2013 46th Hawaii International Conference on System Sciences* (pp. 925–934). doi:10.1109/HICSS.2013.238
- Khan, M., & Quaddus, M. (2004). Group decision support using fuzzy cognitive maps for causal reasoning. *Group Decision and Negotiation Journal*, 13(5), 463–480. doi:10.1023/B:GRUP.0000045748.89201.f3
- Kim, J., Han, M., Lee, Y., & Park, Y. (2016). Futuristic data-driven scenario building: Incorporating text mining and fuzzy association rule mining into fuzzy cognitive map. *Expert Systems with Applications*, 57, 311–323. doi:10.1016/j.eswa.2016.03.043
- Kwon, H., Im, I., & Van de Walle, B. (2002). Are you thinking what I am thinking? –A comparison of decision makers' cognitive maps by means of a new similarity measure. *Proceedings of the 35th Hawaii International Conference on System Sciences*. doi:10.1109/HICSS.2002.993998

- Langfield-Smith, K., & Wirth, A. (1992). Measuring differences between cognitive maps. *The Journal of the Operational Research Society*, 43(12), 1135–1150. doi:10.1057/jors.1992.180
- Lee, H., & Kwon, S. J. (2014). Ontological semantic inference based on cognitive map. *Expert Systems with Applications*, 41(6), 2981–2988. doi:10.1016/j.eswa.2013.10.029
- Lee, K. C., & Kim, H. S. (1997). A fuzzy cognitive map-based bi-directional inference mechanism: An application to stock investment analysis. *Intelligent Systems in Accounting, Finance & Management*, 6(1), 41–57. doi:10.1002/(SICI)1099-1174(199703)6:1<41::AID-ISAF119>3.0.CO;2-J
- Lee, K. C., Kim, J. S., Chung, N., & Kwon, S. J. (2002). Fuzzy cognitive map approach to web-mining inference amplification. *Expert Systems with Applications*, 22(3), 197–211. doi:10.1016/S0957-4174(01)00054-9
- Lee, K. C., & Kwon, S. J. (2006). The use of cognitive maps and case-based reasoning for B2B negotiation. *Journal of Management Information Systems*, 22(4), 337–376. doi:10.2753/MIS0742-1222220412
- Lee, K. C., Lee, H., Lee, N., & Lim, J. (2013). An agent-based fuzzy cognitive map approach to the strategic marketing planning for industrial firms. *Industrial Marketing Management*, 42(4), 552–563. doi:10.1016/j.indmarman.2013.03.007
- Lee, K. C., & Lee, S. (2003). A cognitive map simulation approach to adjusting the design factors of the electronic commerce web sites. *Expert Systems with Applications*, 24(1), 1–11. doi:10.1016/S0957-4174(02)00078-7
- Lee, K. C., & Lee, S. (2007). Causal knowledge-based design of EDI controls: An explorative study. *Computers in Human Behavior*, 23(1), 628–663. doi:10.1016/j.chb.2004.11.003
- León, M., Nápoles, G., Bello, R., Mkrtchyan, L., Depaire, B., & Vanhoof, K. (2013). Tackling Travel Behaviour : An approach based on Fuzzy Cognitive Maps. *International Journal of Computational Intelligence Systems*, 6(6), 1012–1039. doi:10.1080/18756891.2013.816025
- Lopez, C., & Salmeron, J. L. (2014). Dynamic risks modelling in ERP maintenance projects with FCM. *Information Sciences*, 256, 25–45. doi:10.1016/j.ins.2012.05.026
- Lu, W., Yang, J., Liu, X., & Pedrycz, W. (2014). The modeling and prediction of time series based on synergy of high-order fuzzy cognitive map and fuzzy c-means clustering. *Knowledge-Based Systems*, 70, 242–255. doi:10.1016/j.knosys.2014.07.004
- Miao, C., Yang, Q., Fang, H., & Goh, A. (2007). A cognitive approach for agent-based personalized recommendation. *Knowledge-Based Systems*, 20(4), 397–405. doi:10.1016/j.knosys.2006.06.006
- Montazemi, A. R., & Chan, L. (1990). An analysis of the structure of expert knowledge. *European Journal of Operational Research*, 45(2-3), 275–292. doi:10.1016/0377-2217(90)90193-F
- Montazemi, A. R., & Conrath, D. W. (1986). The use of cognitive mapping for information requirements analysis. *Management Information Systems Quarterly*, 10(1), 45–55. doi:10.2307/248879
- Noh, J. B., Lee, K. C., Kim, J. K., Lee, J. K., & Kim, S. H. (2000). A case-based reasoning approach to cognitive map-driven tacit knowledge management. *Expert Systems with Applications*, 19(4), 249–259. doi:10.1016/S0957-4174(00)00037-3
- Nonaka, I. (1994). A dynamic theory of organizational knowledge creation. *Organization Science*, 5(1), 14–37. doi:10.1287/orsc.5.1.14

- Papageorgiou, E. I., Roo, J. D., Huszka, C., & Colaert, D. (2012). Formalization of treatment guidelines using Fuzzy Cognitive Maps and semantic web tools. *Journal of Biomedical Informatics*, 45(1), 45–60. doi:10.1016/j.jbi.2011.08.018 PMID:21911082
- Papageorgiou, E. I., Stylios, C., & Groumpos, P. P. (2006). Unsupervised learning techniques for fine-tuning fuzzy cognitive map causal links. *International Journal of Human-Computer Studies*, 64(8), 727–743. doi:10.1016/j.ijhcs.2006.02.009
- Papageorgiou, E. I., Stylios, C. D., & Groumpos, P. P. (2003). Fuzzy cognitive map learning based on nonlinear Hebbian rule. In T. D. Gedeon & L. C. C. Fung (Eds.), *AI 2003, Lecture Notes in Artificial Intelligence 2903* (pp. 256–268). Berlin: Springer-Verlag; doi:10.1007/978-3-540-24581-0_22
- Polanyi, M. (1966). *The tacit dimension*. London: Routledge & Kegan Paul.
- Rodhain, F. (1999). Tacit to explicit: transforming knowledge through cognitive mapping - an experiment. Proceedings of Special Interest Group on Computer Personnel Research (SIGCPR '99).
- Rodriguez-Repiso, L., Setchi, R., & Salmeron, J. L. (2007). Modelling IT projects success with fuzzy cognitive maps. *Expert Systems with Applications*, 32(2), 543–559. doi:10.1016/j.eswa.2006.01.032
- Satur, R., & Liu, Z. Q. (1999). A contextual fuzzy cognitive map framework for geographic information systems. *IEEE Transactions on Fuzzy Systems*, 7(5), 481–494. doi:10.1109/91.797974
- Sengupta, K., & Teeni, D. (1993). Cognitive feedback in GDSS: Improving control and convergence. *Management Information Systems Quarterly*, 17(1), 87–113. doi:10.2307/249511
- Siau, K., & Tan, X. (2005a). Improving the quality of conceptual modeling using cognitive mapping techniques. *Data & Knowledge Engineering*, 55(3), 343–365. doi:10.1016/j.datak.2004.12.006
- Siau, K., & Tan, X. (2005b). Technical communication in information systems development: The use of cognitive mapping. *IEEE Transactions on Professional Communication*, 48(3), 269–284. doi:10.1109/TPC.2005.853933
- Siau, K., & Tan, X. (2006). Cognitive mapping techniques for user–database interaction. *IEEE Transactions on Professional Communication*, 49(2), 96–108. doi:10.1109/TPC.2006.875074
- Stach, W., Kurgan, L., Pedrycz, W., & Reformat, M. (2005). Genetic learning of fuzzy cognitive maps. *Fuzzy Sets and Systems*, 153(3), 371–401. doi:10.1016/j.fss.2005.01.009
- VanWinsen, F., deMey, Y., Lauwers, L., Van Passel, S., Vancauteran, M., & Wauters, E. (2013). Cognitive mapping: A method to elucidate and present farmers risk perception. *Agricultural Systems*, 122, 42–52. doi:10.1016/j.agsy.2013.08.003
- Wang, S. (1996). A dynamic perspective of differences between cognitive maps. *The Journal of the Operational Research Society*, 47(4), 538–549. doi:10.1057/jors.1996.59
- Wood, M. D., Bostrom, A., Bridges, T., & Linkov, I. (2012). Cognitive Mapping Tools: Review and Risk Management Needs. *Risk Analysis*, 32(8), 1333–1348. doi:10.1111/j.1539-6924.2011.01767.x PMID:22340369
- Xirogiannis, G., & Glykas, M. (2007). Intelligent modeling of e-business maturity. *Expert Systems with Applications*, 32(2), 687–702. doi:10.1016/j.eswa.2006.01.042

KEY TERMS AND DEFINITIONS

Cognitive Map: A representation of the relationships which are perceived to exist among the elements of a given environment.

Explicit Knowledge: Codified knowledge which refers to knowledge that is transmittable in formal and systematic language.

Fuzzy Cognitive Map: An extended and fuzzified version of the cognitive map that enables causal relationships to have fuzzy weights.

Knowledge Creation: The creation of new content based on the organizational tacit and explicit knowledge.

Knowledge Transfer: The transfer of organizational knowledge from one entity to another entity within/between organizations.

Tacit Knowledge: Personal knowledge which could be shared and exchanged through face-to-face contact among actors.

APPENDIX



Table 1. Articles related to application of CMs in intelligent information systems

Author (Year)	Method	Application	Knowledge Perspective*			
			1	2	3	4
Wang, 1996	used neural networks in a three- dimensional framework defined by initial input to the system, time horizon and thresholds.	measuring differences between CMs	✓			
Montazemi, & Conrath, 1986	applied CM for the evaluation of the performance of insurance claim representatives	Information requirements analysis	✓			✓
Montazemi & Chan, 1990	used CMs to extract the structure of expert knowledge	stock price forecasting	✓			✓
Satur & Liu, 1999	used contextual FCMs to draw inferences from quantitative and qualitative descriptions	Geographic Information Systems	✓			✓
Wood et al, 2012	Review of techniques for assessing and diagramming stakeholder mental models	Risk management	✓	✓		✓
León et al, 2013	Used FCM to model travel behavior	Transportation policy		✓	✓	✓
Kim et al, 2016	Used FCM from expert opinion	Secnarion Building	✓			✓
VanWinsen et al, 2013	Used CM to present farmers' risk perception	Risk analysis of agriculture systems	✓			✓
Lee & Kim, 1997	used FCMs to construct a causal knowledge base and perform a bi-directional inference	stock investment analysis	✓	✓		✓
Eden, 2004	used CMs to structure problems or issues		✓			
Lee & Lee, 2007	used FCM to develop a causal knowledge-based expert system	EDI controls	✓	✓		✓
Papageorgiou et al, 2012	Used FCM and semantic web to formalize medical knowledge	therapy decision making in medicine treatment	✓	✓	✓	✓
Douali et al 2015	Used case-based FCMs for medical diagnosis	Clinical diagnosis Support System	✓	✓	✓	✓
Lee & Lee, 2003	used simulation of CM to adjust the design factors of the EC web sites	designing EC website	✓			
Lee et al., 2002	used FCMs to develop causal knowledge base from mined association rules	web-mining inference amplification	✓			
Kwon et al., 2002	proposed a method to assess the structural similarities among the FCMs	DSS for urban planning decision problem	✓			✓
Lopez & Salmeron, 2014	Built FCMs for ERP maintenance risks	Risk analysis of ERP maintenance project				✓
Aguilar, 2002	developed an adaptive FCM based on the random neural network model	modeling of dynamic system	✓			✓
Papageorgiou et al., 2003	used unsupervised Hebbian algorithm to nonlinear units for training FCMs	learning methods	✓			✓
Papageorgiou et al., 2006	used unsupervised learning techniques for fine-tuning weights of concepts in FCMs	industrial process control	✓			✓
Stach et al., 2005	used genetic algorithm to generate FCMs from historical data	learning methods	✓			
Ahmadi et. Al 2015	Used Imperialist competitive algorithm to generate FCMs from historical data	learning methods	✓	✓		

continued on following page

Table 1. Continued

Author (Year)	Method	Application	Knowledge Perspective*			
			1	2	3	4
Lee et al., 2013	Used FCM for strategic marketing planning for industrial firms	Marketing planning	✓			✓
Lee & Kwon, 2006	used CM to formalize tacit knowledge and applied CBR techniques for storing and retrieving CMs.	reasoning in B2B negotiations		✓		✓
Noh et al., 2000	used CM to formalize tacit knowledge, and applied CBR techniques to store and reuse tacit knowledge	credit analysis problem		✓		✓
Sengupta, & Te'eni, 1993	applied cognitive feedback in GDSS to improve decision making.				✓	
Amer et al, 2015	Used FCM-based scenarios to develop technology roadmap	Wind energy sector of a developing country		✓		✓
Miao et al., 2007	used a fuzzy cognitive agent to provide personalized recommendations to on-line customers	e-commerce application			✓	✓
Caliskan, 2006	used CM and AHP technique as a DSS	evaluation of transport investment alternatives			✓	✓
Siau & Tan, 2006	used CM techniques (causal mapping, semantic mapping, and concept mapping) to improve user–database interaction.	user–database interface analysis and design				✓
Lu et al, 2014	used high-order FCM and fuzzy c-means clustering for time series prediction	modelling and prediction of time series	✓	✓	✓	
Siau & Tan, 2005a	used CM techniques to improve the quality of conceptual modeling.	conceptual modeling				✓
Siau & Tan, 2005b	used CMs to improve the usability of information systems	Technical Communication in IS Development				✓
Aju Kumar et al, 2015	Used FCMs for assessment of factors influencing human reliability	Human reliability of plant maintenance	✓	✓		✓
Khan & Quaddus, 2004	used FCM to improve GDSS usability	group decision support environment			✓	
Xirogiannis & Glykas, 2007	used FCM to model hierarchical and distributed nature of e-business maturity	e-business strategy formulation			✓	✓
Jetter & Sperry, 2013	Used FCM to capture, integrate and analyze stakeholders' mental models about product impacts and needs	planning tool for product development teams			✓	✓
Rodriguez-Repiso et al., 2007	used FCM for mapping success, modeling Critical Success Factors (CSFs) perceptions and their relationships between them	IT project success				✓

1. Knowledge creation 2. Knowledge storage/retrieval 3. Knowledge transfer 4. Knowledge application

Computer-Assisted Parallel Program Generation



Shigeo Kawata

Utsunomiya University, Japan

INTRODUCTION

Parallel computation is widely employed in scientific researches, engineering activities and product development. Parallel program writing itself is not always a simple task depending on problems solved. Large-scale scientific computing, huge data analyses and precise visualizations, for example, would require parallel computations, and the parallel computing needs the parallelization techniques. In this Chapter a parallel program generation support is discussed, and a computer-assisted parallel program generation system P-NCAS is introduced.

Computer assisted problem solving is one of key methods to promote innovations in science and engineering, and contributes to enrich our society and our life toward a programming-free environment in computing science. Problem solving environments (PSE) research activities had started to enhance the programming power in 1970's. The P-NCAS is one of the PSEs; The PSE concept provides an integrated human-friendly computational software and hardware system to solve a target class of problems (Kawata, 2014). For example, a PSE generates a computer program automatically to solve differential equations (Boonmee et al., 1998a; Boonmee et al., 1998b; Fujio et al., 1998; Fujita et al., 2000; Gallopoulos, et al., 1991; Gallopoulos et al., 1994; Hirayama et al., 1988; Houstis et al., 1992; Kawata et al., 2000; Okochi et al., 1994; Rice et al., 1984; Umetani, 1985). In the PSE concept, human concentrates on target problems, and a part of problem solving, which can be solved mechanically, is performed by computers or machines or software.

The concept of the computer-assisted program generation has been opening the new style of the computer programing to reduce the programing hard task. Huge computer software may include errors and bugs. The errors or malfunction of the software infrastructure may induce uncertainty and accordingly serious accidents in our society (Einarsson, 2005; Kawata et al, 2012). The programing process tends to include mechanical parts, that means, mechanically programmable parts. When the application area of the software is limited in a reasonable size, a part of the software would be mechanically generated. For example, scientific research-oriented programs would have a similar program structure depending on the numerical scheme.

P-NCAS supports scientists and engineers to generate parallel programs for problems described by partial differential equations (PDEs). P-NCAS presents a remarkable capability of visualization and steering of all the processes required for the generation of parallel programs. In P-NCAS users input problem description information including PDEs, initial and boundary conditions, discretization scheme, algorithm and also comments on the problem itself as well as the parallelization information. P-NCAS supports a domain decomposition method for the parallelization, and the SPMD (single program multi data) model is employed. In P-NCAS users can see and edit all the information through the visualization and editing windows. The program flow is also visualized by a Problem Analysis Diagram (PAD). Even through the flow visualization window users can modify the program flow. Finally P-NCAS outputs the corresponding parallel program in the C language,

DOI: 10.4018/978-1-5225-2255-3.ch398

and at the same time a document for the program including the problem itself, the program flow, the PDEs, the initial and boundary conditions, the discretization method, the numerical algorithm employed and the variable definitions.

In this chapter first the PSE concept is briefly introduced, and then the details of the P-NCAS system is described for the parallel program generation support. Finally an example application result is presented including a load balancing result.

BACKGROUND OF COMPUTER-ASSISTED PROBLEM SOLVING ENVIRONMENT (PSE) IN SCIENTIFIC COMPUTING

PSE is defined as follows: “A system that provides all the computational facilities necessary to solve a target class of problems. It uses the language of the target class and users need not have specialized knowledge of the underlying hardware or software” (Gallopoulos, Houstis & Rice, 1994). In computing sciences, we need computer power, excellent algorithms and programming power in order to solve scientific and engineering problems leading to scientific discoveries and development of innovative new products. The computer power and the computing algorithms have been developed extraordinarily, and have provided tremendous contributions to sciences, engineering and productions. On the other hand, the programming power has not been developed well, compared with the computer power and the algorithm power. The concept of PSE was initially proposed to support the programming power in science and engineering, and has been explored for decades.

So far computer simulation has contributed to researches, productions and developments as well as experimental and theoretical methods. New researches tend to require new computer programs to simulate phenomena concerned. In developing new products engineers would also need new computer programs to develop new

products effectively. They may have to develop the new programs or learn how to use the programs for the product development. Human power still contributes greatly to develop and write the new computer software. They like to devote themselves to solve their target problems, but not to develop or learn the computer programs. The PSEs would also help them develop the computer software including parallel programs or learn how to use the software system.

In these days PSE covers a various wide area, for example, computer-assisted program generation, education support, CAE (Computer Aided Engineering) software learning support, grid/cloud computing support, factory management support including plant factory management, program execution support, uncertainty management in scientific computing, etc. There are many PSE examples studied so far. In the references of (Ford & Chatelin, 1987; Fujita et al., 2006; Gaffney, & Houstis, 1992; Houstis et al, 1997; Houstis, Rice, Gallopoulos & Bramley, 2000; Kawata, Tago, Umetani & Minami, 2005; Kawata et al., 2012; Kawata, 2014; Gaffney & Pool, 2007; Teramoto et al., 2007) one can find the example PSEs.

COMPUTER-ASSISTED PARALLEL PROGRAM GENERATION PSE - P-NCAS -

P-NCAS supports computer-assisted parallel program generation for PDEs-based problems in FDM (Finite Difference Method) based on a domain decomposition and MPI (Message-Passing Interface) functions (MPI, 2012). In addition, parallel programs generated in P-NCAS have load-balancing functions; one is a function for a static load balance and the other is that for the dynamic load balance. Figure 1 shows the P-NCAS work flow for the parallel program generation.

Through the interface of the PSE system of P-NCAS (Boonmee et al., 1998a; Boonmee et al., 1998b; Kawata et al., 2000; Fujita et al., 2000; Kawata, 2014) users input problem information

itself including PDEs, initial and boundary conditions, discretization and computation schemes, information for the domain decomposition and meanings for each equation and symbol. P-NCAS visualizes all the information to the users, and the users also can steer the program generation process through the P-NCAS interface. Then P-NCAS outputs a program flow, a C-language source code for the problem and also a document for the program and for the problem. In P-NCAS a user inputs his/her problem, and P-NCAS guides the users to solve the problem. The P-NCAS contains all the information of the problem, PDEs, the discretization scheme, the parallelization method, the mesh information, the equation manipulation method and results, the program structure designed, the variables employed, the constant definitions and the program itself. Therefore, P-NCAS also generates the corresponding document for the program generated in P-NCAS (Inaba et al., 2004).

P-NCAS helps users generate MPI-based parallel simulation programs based on PDEs (Fujita et al., 2000). P-NCAS supports a domain decomposition in a design of a parallel numerical simulation program, and the domain decomposition is designed or steered by users through a visualization window in P-NCAS. After designing the domain decomposition, the parallel program itself is designed and generated in P-NCAS, and the designed parallel program is visualized and steered by a Problem Analysis Diagram (PAD). In P-NCAS, MPI functions (MPI, 2012) are employed for the message passing, and a single program multiple data (SPMD) model is supported. The visualization and steering capabilities provide users a flexible design possibility of parallel programs.

Figure 2 shows an example image of the domain decomposition in P-NCAS. At least the boundary data for each domain decomposed are required to complete the computation in the adjacent processor (see Figure 3). In P-NCAS the MPI functions are also automatically inserted to complete the parallel data communication programming. After specifying the domain decomposition information in P-NCAS, the parallel program is generated and provided to the users.

The visualization and steering capabilities are enabled by the tree-type data structure shown in Figure 4 (Boonmee et al., 1998a) in P-NCAS. The tree-type data structure is employed to express all the data and information in P-NCAS including PDEs, each term of the PDEs, each symbol, boundary and initial conditions, the parallelization information for the domain decomposition and the program flow. Each PDE is discretized by a discretization method, which the user specifies. The equation algebraic manipulation is also performed smoothly based on the tree structure. In the tree-type structure each component has a link to its meaning. Based on the meanings and definitions of the components, P-NCAS helps users to prevent users' input or steering errors. When the users modify or steer the equation or one term of PDEs or some symbols, the steering results propagate and are reflected to all the relating components immediately through the links. By using the data structure, P-NCAS also visualizes easily all the information and provides the steering capability to users (see Figures 5-7).

Figure 5 presents an example description of an input problem information, and Figure 6 shows an example domain decomposition information. Through the P-NCAS visualization windows, for examples, shown in Figures 5 and 6, one can check all the information and can also edit the information. In P-NCAS, after setting all the information for the problem description, the discretization information and the parallelization information through the P-NCAS windows, all the information is visualized to the users and the users can edit all the information through the windows. The discretization of each PDE is also performed automatically; depending on the discretization information which users input through the P-NCAS windows, the PDEs are discretized and manipulated appropriately according to the PDEs solving scheme. Then P-NCAS designs the parallel program for the problem, and outputs the parallel program and the corresponding document. Figure 6 shows an example MPI program automatically generated in P-NCAS.

In addition, P-NCAS supplies users an option for a selection of a load balancing method, that is, an equal load balancing method or a dynamic load balancing method. When all the CPUs or machines are equivalent and the machine power is static, the equal load balancing method is appropriate. When the machines are heterogeneous or the machine power changes dynamically, the dynamic load balancing method is appropriate. Users can specify one of them, when they input the parallelization information through P-NCAS. In the dynamic load balance P-NCAS generates additional functions to measure all the machine power dynamically together with the corresponding computing program generated. The additional load balance measure functions measure the machine power regularly during the computation, and dynamically the domain sizes decomposed are modified to minimize the computation time.

Figure 8 shows an example parallelization performance result for a shock wave propagation in a 2-dimensional compressible-fluid simulation: a parallel program generated in P-NCAS is used to measure the parallelization performance. In Figure 8 the static load balance method is used on uniform machines. In order to check the dynamic load balance function automatically generated by P-NCAS, during the computation an additional load was applied as shown in Figure 9 (the left graph): by the additional load the computation time increases much in this specific case, if the static load balancing is used. When the dynamic load balancing method is selected in this example case, P-NCAS generates the functions, which measure the load balance of each machine dynamically, and according to the measured result each domain size is changed and adjusted dynamically to minimize the computation time. The right graph in Figure 9 demonstrates the viability of the dynamic load balancing functions generated in P-NCAS, and the computation time reduction is significant in this case.

FUTURE RESEARCH DIRECTIONS

At present P-NCAS supports the parallel program generation for PDEs-based problems by FDM based on the domain decomposition parallelization method and MPI. P-NCAS demonstrates that for a specific scientific problem area described by PDEs the parallel program generation is well supported mechanically by computers.

One of the near future works should be a parallel program generation support for FEM. FEM is also widely used in industries and scientific areas. One of future directions is a program generation support for specific computer architectures: a program generation support oriented to, for example, GPU computers or vector computers. GPU is now widely available, but it needs a special care for its programming. In addition to the directly relating areas to P-NCAS, a programming-free environment should be studied further to reduce the heavy task of the programming. So far one way has been to build CAE (Computer-Aided Engineering) software to reduce the programming work. P-NCAS and its future capabilities would contribute to build up CAEs. CAE is one of good ways to reduce the programming task. However, to learn how to use huge CAE softwares is another hard task, and this hard work should be also supported. On this point, PSE would also help users learn CAEs, and CAE-navigation PSEs help users learn or work on huge CAE systems.

Another important future direction must be to support scientific computing without uncertainty (Einarsson, 2005; Kawata et al., 2012; Kawata, 2014). In the process of scientific computing, the origin of uncertainty comes from physical model uncertainty, mathematical model errors, unknown input data, unknown boundary condition errors, numerical model errors, insufficient numerical precision, round-off error, floating point precision, programming errors, data processing errors or uncertainty, visualization errors, human errors, etc. So far the uncertainty or the errors in scientific

computing have caused serious accidents and disasters (Einarsson, 2005). Future PSEs should support to reduce the uncertainties in scientific computing.

CONCLUSION

We have presented a concept of computer-assisted parallel program generation, and the P-NCAS system is introduced. The PSE concept and a brief history of PSE research activities are also presented in computing science. The PSE researches have explored capabilities of enhancement of program-

ming power. PSEs have also started to provide a reliable tool to help engineers and scientists generate programs including parallel programs. There would not be an almighty PSE to solve all the problems, but each problem area would be solved by a specific PSE, for example, a PSE for a parallel program generation for PDEs-based problems by FDM and the domain decomposition; Another parallel program generation PSE would be needed for problems by FEM (Finite Element Method) or BEM (Boundary Element method) or particle methods, for example. A lot of PSE studies have demonstrated that each PSE in each area works well to support users to

Figure 1. P-NCAS work flow for the computer assisted scientific parallel program generation support. P-NCAS inputs the domain-decomposition parallelization information, partial differential equations (PDEs), initial and boundary conditions, discretization method and algorithm, and outputs a C language program. The PDEs are automatically discretized and the program is generated mechanically. P-NCAS is a white box system, in which users can see and steer all the processes of program generation. P-NCAS system contains all the information for program generation, including basic equations, discretization schemes, discretized equations, boundary and initial conditions, parallelization scheme, mesh structure, program structure, and definitions of variables and constants. Therefore, a document for the corresponding program is also generated together with the program itself.

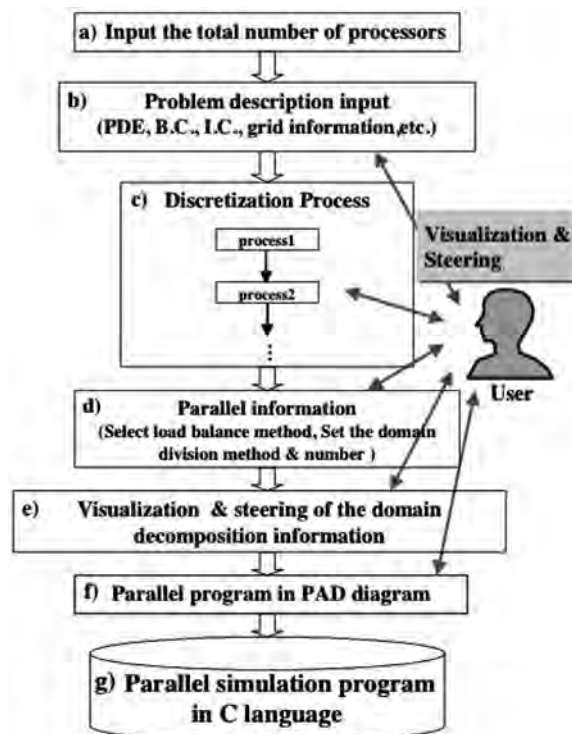


Figure 2. An example image of the domain decomposition in P-NCAS

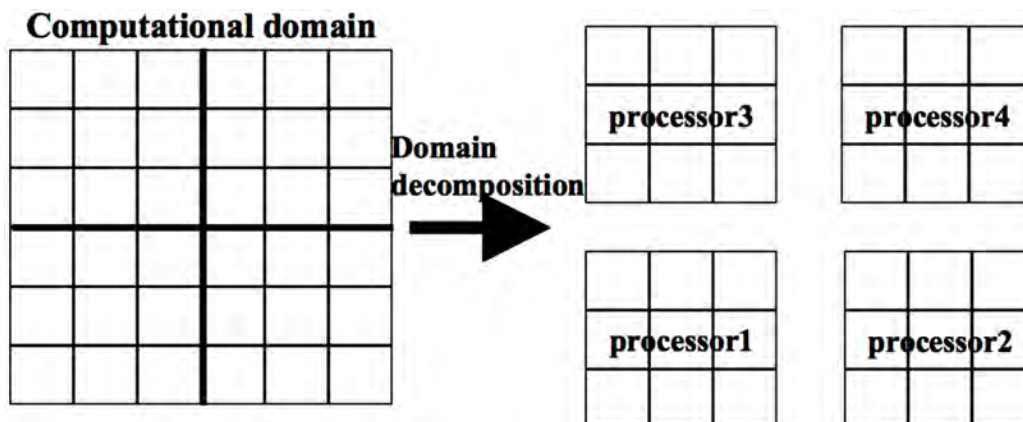


Figure 3. Boundary data are communicated by the MPI functions in P-NCAS, in which FDM is supported

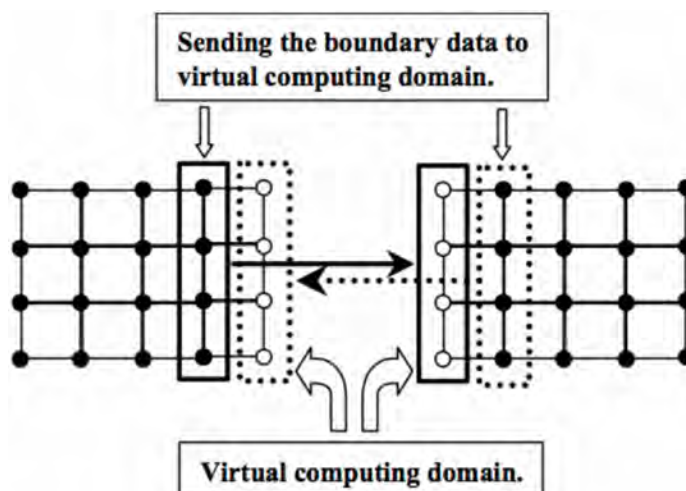


Figure 4. An example tree-type data structure used in P-NCAS. In this example, a term of an equation is expressed by the tree, and each symbol has a link to the symbol definition.

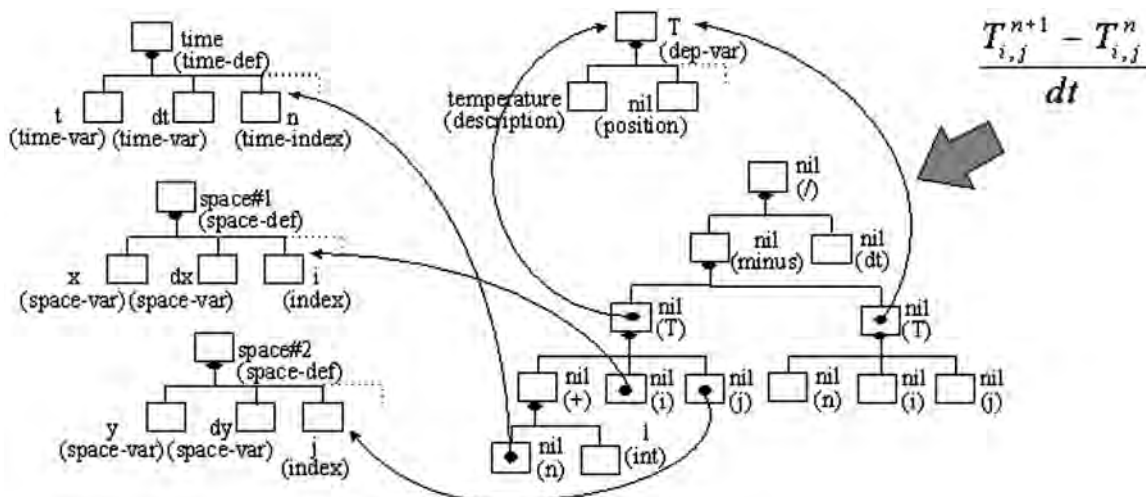


Figure 5. An example PDE-problem description in P-NCAS. Through each window users would edit the input description.

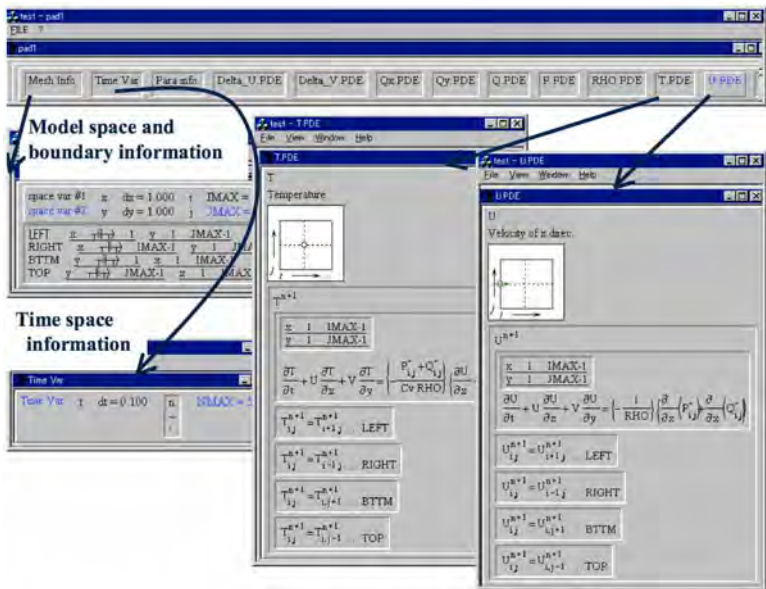


Figure 6. Input and visualization of domain decomposition information.

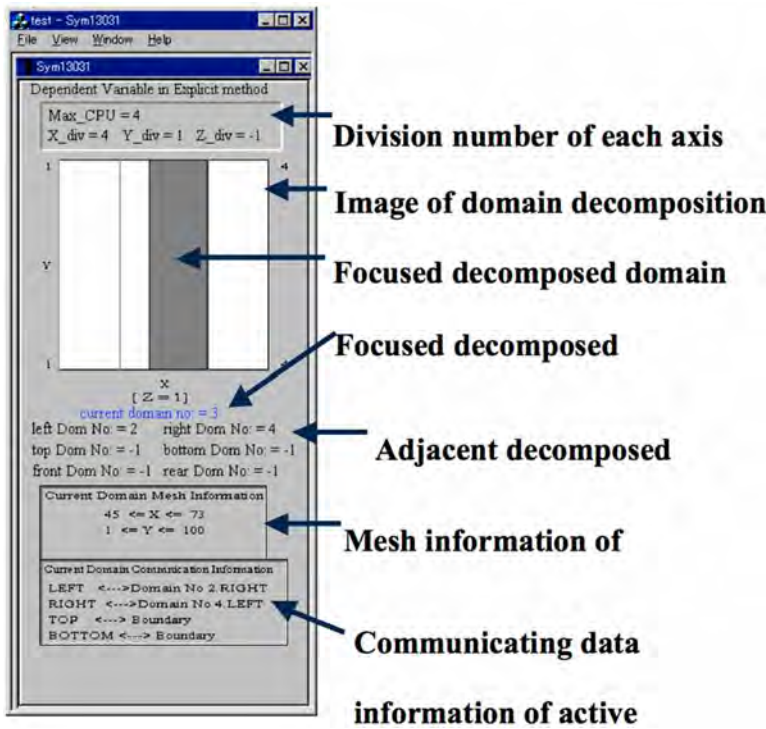


Figure 7. Visualization of MPI functions designed for a domain decomposition information in P-NCAS

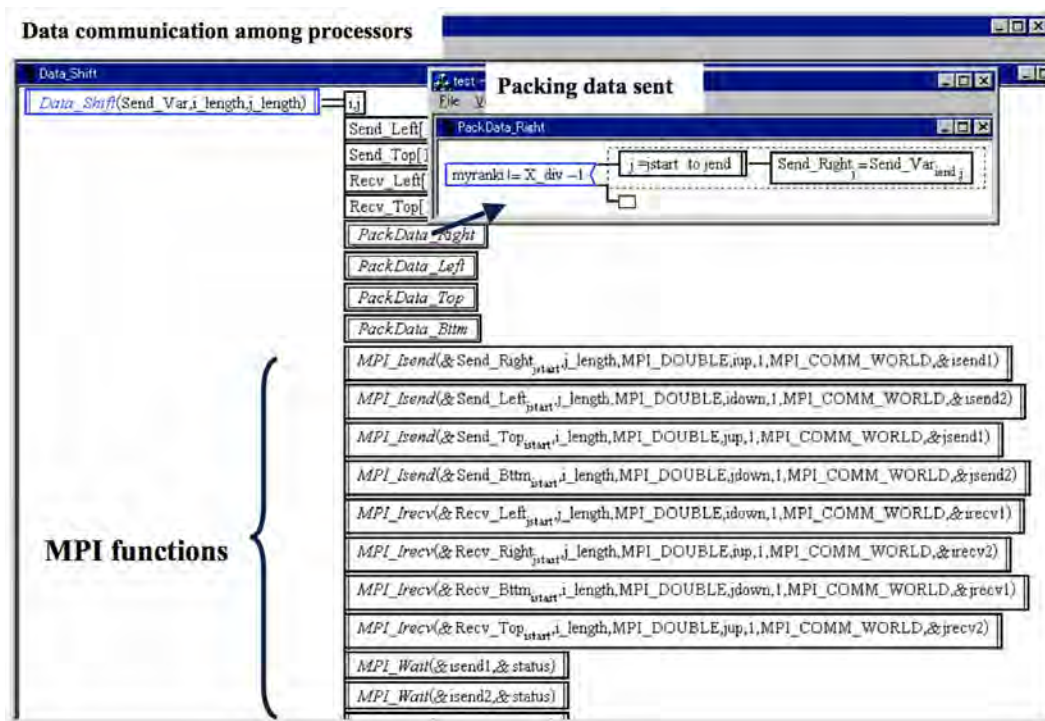


Figure 8. Speed up performance for a 2-dimensional fluid simulation (shock wave propagation) by a program generated by P-NCAS

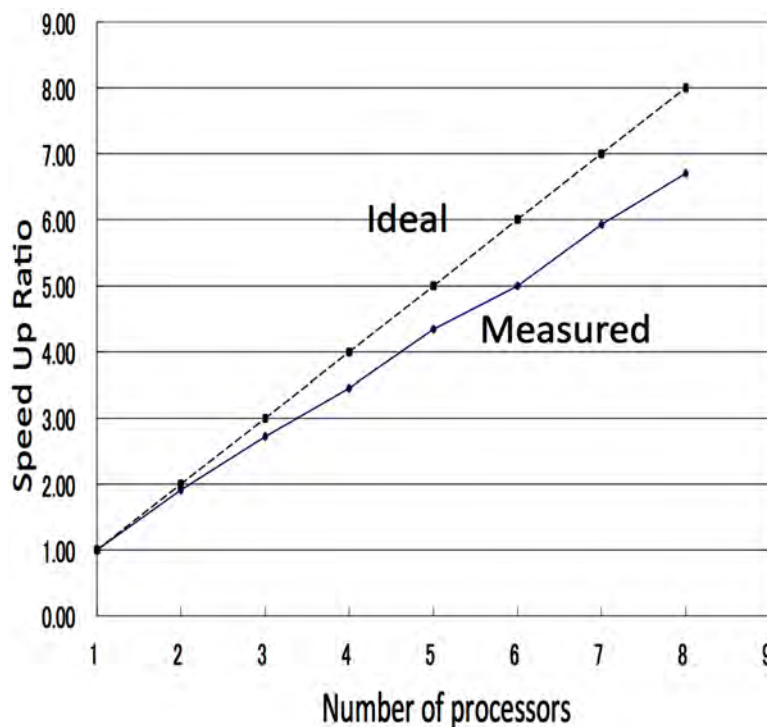
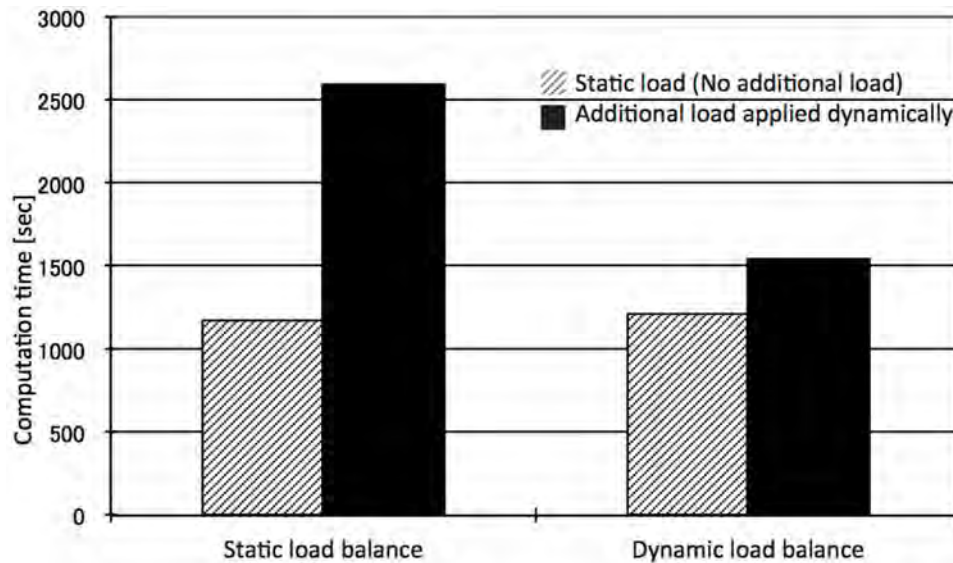


Figure 9. A performance test result for a dynamic load balance. An additional load was applied during the computation, so that the computation time increases as shown in the left graph. When the dynamic load balancing function generated automatically by P-NCAS is used to this specific case, the domain size changes automatically depending on the machine load, and the result of the dynamic load balancing in the right graph shows the viability of the dynamic load balance functions generated by P-NCAS.



solve their target problems effectively. Even for uncertainty management in computing science PSE supports realistic tolls to avoid numerical errors or uncertainties (for example, Einarsson, 2005; Kawata et al., 2012).

In this Chapter we presented P-NCAS for the parallel program generation support, and the results shown demonstrate that the P-NCAS, a parallel program generation support system, works very effectively to solve the problems in parallel. The results in P-NCAS show that the PSE concept and P-NCAS are viable for the problem solving in computing science. PSEs would be appropriate to enhance the programming power and to reduce the costs and heavy tasks in scientific computing as a reliable society infrastructure.

ACKNOWLEDGMENT

The work is partly supported by JSPS (Japan Society for the Promotion of Science), MEXT (Ministry of Education, Culture, Sports, Science & Technology in Japan), CORE (Center for Optical Research and Education, Utsunomiya University, Japan), Collaboration Center for Research and Development of Utsunomiya University, IFIP WG2.5, JSCES (Japan Society for Computational Engineering and Science) and PSE (Problem Solving Environment) Research group Japan. The authors also would like to express their appreciations to C. Boonmee, A. Fujita, T. Teramoto, H. Usami, Y. Manabe, Y. Hayase, M. Matsumoto, J. Rice, E. Houstis and friends in PSE Research group in Japan and in IFIP WG2.5 for their helps and fruitful discussions.

REFERENCES

- Boonmee, C., & Kawata, S. (1998a). Computer-Assisted Simulation Environment for Partial-Differential-Equation Problem, 1. Data Structure and Steering of Problem Solving Process. *Transaction of the Japan Society for Computational Engineering and Science*, Paper No. 19980001.
- Boonmee, C., & Kawata, S. (1998b). Computer-Assisted Simulation Environment for Partial-Differential-Equation Problem: 2. Visualization and Steering of Problem Solving Process. *Transaction of the Japan Society for Computational Engineering and Science*, Paper No. 19980002.
- Einarsson, B. (Ed.). (2005). *Accuracy and Reliability in Scientific Computing*. Society for Industrial and Applied Mathematics. doi:10.1137/1.9780898718157
- Ford, B., & Chatelin, F. (Eds.). (1987). *Problem Solving Environments for Scientific Computing*. North-Holland.
- Fujio, H., & Doi, S. (1998). Finite Element Description System as a Mid-Layer of PSE. *Proceedings of Conference on Computation Engineering and Science*, 3(2), 441-444.
- Fujita, A., Teramoto, T., Nakamura, T., Boonmee, C., & Kawata, S. (2000). Computer-Assisted Parallel Program Generation System P-NCAS from Mathematical Model-Visualization and Steering of Parallel Program Generation Process. *Transaction of the Japan Society for Computational Engineering and Science*, Paper No. 20000037.
- Fuju, H., Kawata, S., Sugiura, H., Saitoh, Y., Usami, H., Yamada, M.,... Hayase, Y. (2006). Scientific Simulation Execution Support on a Closed Distributed Computer Environment. *The 2nd IEEE International Conference on e-Science and Grid Computing*.
- Gaffney, P. W., & Houstis, E. N. (Eds.). (1992). *Programming Environments for High-Level Scientific Problem Solving*. North-Holland.
- Gaffney, P. W., & Pool, J. C. T. (Eds.). (2007). *Grid-Based Problem Solving Environments*. Springer. doi:10.1007/978-0-387-73659-4
- Gallopoulos, E., Houstis, E., & Rice, J. R. (1991). *Future Research Directions in Problem Solving Environments for Computational Science*. Technical Report CSRD Report No.1259. Report of a workshop on Research Direction in Integrating Numerical Analysis, Symbolic Computer, Computational Geometry, and Artificial Intelligence for Computational Science, Washington DC.
- Gallopoulos, E., Houstis, E., & Rice, J. R. (1994). Computer as Thinker/Doer: Problem-Solving Environments for Computational Science. *IEEE Computational Science & Engineering*, 1(2), 1-23. doi:10.1109/99.326669
- Hirayama, Y., Ishida, J., Ota, T., Igai, M., Kubo, S., & Yamaga, S. (1988). Physical Simulation using Numerical Simulation Tool PSILAB. *The 1st Problem Solving Environment Workshop*, 1-7.
- Houstis, E. N., Gallopoulos, E., Bramley, E., & Rice, J. R. (1997). Problem-Solving Environment in Computational Science. *IEEE Computational Science & Engineering*, 4(3), 18-21. doi:10.1109/MCSE.1997.615427
- Houstis, E. N., & Rice, J. R. (1992). Parallel ELLPACK, a development environment and problem solving environment for high performance computing machines. In P. Gaffney & E. N. Houstis (Eds.), *Programming Environments for High-Level Scientific Problem Solving* (pp. 229-241). North-Holland.
- Houstis, E. N., Rice, J. R., Gallopoulos, E., & Bramley, R. (Eds.). (2000). *Enabling Technologies for Computational Science, Framework, Middleware and Environments*. Kluwer Academic Publishers. doi:10.1007/978-1-4615-4541-5
- Inaba, M., Fuju, H., Kitamuki, R., Kawata, S., & Kikuchi, T. (2004). Computer-Assisted Documentation in a Problem Solving Environment (PSE) for Partial Differential Equation Based Problems. *Transactions of the Japan Society for Computational Engineering and Science*, 20040025.

Kawata, S. (2014). Computer Assisted Problem Solving Environment (PSE). In Encyclopedia of Information Science and Technology, Third Edition. IGI Global.

Kawata, S., Boonmee, C., Fujita, A., Nakamura, T., Teramoto, T., Hayase, Y., & Matsumoto, M. et al. (2000). Visual Steering of the Simulation Process in a Scientific Numerical Simulation Environment—NCAS-. In E. N. Houstis & J. Rice (Eds.), *Enabling Technologies for Computational Science* (pp. 291–300). Kluwer. doi:10.1007/978-1-4615-4541-5_23

Kawata, S., Inaba, M., Fujiu, H., Sugiura, H., Saitoh, Y., & Kikuchi, T. (2005). Computer-Assisted Liaison among Modules in a Distributed Problem Solving Environment (PSE) for Partial Differential Equation Based Problems. *Transactions of the Japan Society for Computational Engineering and Science*, 20050029.

Kawata, S., Kobashi, H., Ishihara, T., Manabe, Y., Matsumoto, M., Barada, D., & Usami, H. et al. (2012). Scientific Simulation Support Meta-System: PSE Park - with Uncertainty Feature Information -. *International Journal of Intelligent Information Processing*, 3(2), 66–76. doi:10.4156/ijiip.vol3.issue2.9

Kawata, S., Tago, Y., Umetani, Y., & Minami, K. (Eds.). (2005). PSE Book: Computer assisted Problem Solving Environment in computing science. Tokyo: Baifukan.

MPI (Message passing Interface). (2012). *MPI-CH2*. Retrieved October 2, 2012, from <http://www.mcs.anl.gov/research/projects/mpich2/>

Okochi, T., Konno, C., & Igai, M. (1994). High Level Numerical Simulation Language DEQSOL for Parallel Computers. *Transaction of Information Processing Society of Japan*, 35(6), 977–985.

Rice, J. R., & Boisvert, R. F. (1984). *Solving Elliptic Problems Using ELLPACK*. Springer-Verlag.

Teramoto, T., Okada, T., & Kawata, S. (2007). A Distributed Education-Support PSE System. *The 3rd IEEE International Conference on e-Science and Grid Computing*, 516–520.

Umetani, Y. (1985). DEQSOL A numerical Simulation Language for Vector/Parallel Processors. *Proceedings of IFIP TC2/WG22 (International Federation for Information Processing, Working Group 2.2)*, 5, 147–164.

KEY TERMS AND DEFINITIONS

Parallel Computing: Computations performed in parallel on distributed computers or processors. Parallel computing program needs a special treatment to divide computations into many sub-computations on the distributed computers or processors. Special parallel computers would need special programming technique to divide the task into many sub-tasks or to achieve the communication among processors or computers if needed.

PSE (Problem Solving Environment): Defined as “A system that provides all the computational facilities necessary to solve a target class of problems. It uses the language of the target class and users need not have specialized knowledge of the underlying hardware or software”. An example typical PSE generates a computer program automatically to solve differential equations. PSE provides integrated human-friendly computational services and facilities. At present PSE covers a rather wide area, for example, program generation support PSE, education support PSE, CAE software learning support PSE, grid/cloud computing support PSE, job execution support PSE, e-Learning support PSE, uncertainty management in scientific computing, and PSE for PSE generation support. A parallel computing support PSE, P-NCAS is introduced in this Chapter.

Scientific Simulation: Computation to describe a real world in computer.

Constrained Nonlinear Optimization in Information Science

William P. Fox

Naval Postgraduate School, USA

INTRODUCTION

A company manufactures new smart-phones that are supposed to capture the market by storm. The two main inputs components of the new smart-phone are the circuit board and the relay switches that make the phone faster and smarter and give it more memory.

The number of smart-phones to be produced is estimated to equal $E = 200x_1^{\frac{1}{3}}x_2^{\frac{1}{2}}$, where E is the number of smart-phones produced and x_1 & x_2 are the number of circuit board hours and the number of relay hours worked, respectively. Such a function is known to economists as a *Cobb–Douglas function*. Laborers are paid by the type of work they do: the circuit boards and the relays for \$5 and \$10 an hour, respectively. We want to maximize the number of smart-phones to be made if we have \$150,000 to spend on these components in the short run.

Lagrange multipliers can be used to solve *nonlinear optimization problems* (NLPs) in which all the constraints are equality constrained. We consider the following type of NLPs as shown by Equation (1):

$$\text{Maximize (Minimize) } z = f(x_1, x_2, \dots, x_n) \quad (1)$$

Subject to

$$g_1(x_1, x_2, \dots, x_n) = b_1$$

$$g_2(x_1, x_2, \dots, x_n) = b_2$$

...

$$g_m(x_1, x_2, \dots, x_n) = b_m$$

DOI: 10.4018/978-1-5225-2255-3.ch399

In our smart-phones example, we find we can build an equality constrained model. We want to maximize

$$E = 200x_1^{\frac{1}{3}}x_2^{\frac{1}{2}}$$

subject to the equality constraint

$$5x_1 + 10x_2 = 150,000$$

Problems in information science and technology such as this can be modeled using constrained optimization. We begin our discussion with equality constrained optimization, then discuss the inequality constrained optimization, and finally discuss some numerical methods to approximate the solutions.

BACKGROUND

The general constrained nonlinear programming (NLP) problem is to find x^* as to optimize $f(X)$, $X = (x_1, x_2, \dots, x_n)$ subject to the constraints of the problem shown in equation (2).

$$\text{Maximize or Minimize } f(x_1, x_2, \dots, x_n)$$

subject to

$$g_i(x_1, x_2, \dots, x_n) \begin{cases} \geq \\ = \\ \leq \end{cases} b_i \quad (2)$$

for $i = 1, 2, \dots, m$.

Classical constrained optimization appeared with equality constraints and Lagrange multiplier named for Joseph Lagrange in the late 1700's. It was almost two hundred years later when Kuhn-Tucker (1951) presented their famous Kuhn-Tucker (KT) conditions. Scholars later found that Karsuh (1939) had done considerable work on his thesis in the area of constrained optimization and thus, was added his name was added to create the Karsh-Kuhn-Tucker (KKT) conditions. Bellman (1952; 1957) created dynamic programming in the 1950's to handle sequential constraints in optimization.

Methods to ease computation and complexity of certain forms of NLP were created to ease the solving of such problems. Quadratic programming (Bazaraa et al., 1993) and Wolfe's method (1959) showed how to convert some NLPs into linear programs. Separable programming was another heuristic procedure described by Winston (2002) and Bazaraa et al. (1993). Complexities in problems have led to other heuristic models. The evolutionary strategy optimization technique was created in the early 1960s and developed further in the 1970s and later by Ingo Rechenberg, Hans-Paul Schwefel and his co-workers. Further development was done in evolutionary strategies by Michalewicz (1999). Strides in computer science led to the development of genetic algorithms, (Homaifar et al. 1994; Stender, 1994; Burke et al., 2005) and then to particle swarm optimization (Kennedy, 1995; 1997; Parsopolous et al., 2002). The complexity of the problems dictates the methods to be used. Cuckoo search is the most recent optimization method (Yeng et al, 2009)

MAIN FOCUS

Introduction and Basic Theory

To solve NLPs in the form of Equation (2), we associate a Lagrangian multiplier, λ_i , with the i th constraint and form the Lagrangian equation to get Equation (3):

$$L(X, \lambda) = f(X) + \sum_{i=1}^m \lambda_i (b_i - g_i(X)) \quad (3)$$

The computational procedure for Lagrange multipliers requires that all the partials of this Lagrangian function, Equation (4), must equal zero. These partials are the *necessary conditions* of the NLP problem. These are the conditions required for $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$ to be a solution to Equation (3).

$$\text{The Necessary Conditions} \quad (4)$$

$$\partial L / \partial X_j = 0 \quad (j = 1, 2, \dots, n \text{ variables}) \quad (4a)$$

$$\partial L / \partial \lambda_i = 0 \quad (i = 1, 2, \dots, m \text{ constraints}) \quad (4b)$$

Definition: $\mathbf{x}$ is a regular point if and only if $\nabla g_i(x)$, $i = 1, 2, \dots, m$ are linearly independent.

Theorem 1

- Let (2) be a maximization problem. If f is a concave function and each $g_i(x)$ is a linear function, then any point satisfying Equation (4) will yield an optimal solution.
- Let (2) be a minimization problem. If f is a convex function and each $g_i(x)$ is a linear function, then any point satisfying Equation (4) will yield an optimal solution.

The Hessian matrix is used to determine if a function was convex, concave, or neither. We also note that the above theorem limits our constraints to linear functions. What if we have nonlinear equality constraints?

We can use the bordered Hessian in the sufficient conditions. Given the bivariate Lagrangian function as in

$$L(x_1, x_2, \lambda_i) = f(x_1, x_2) + \sum_{i=1}^m \lambda_i (b_i - g_i(x_1, x_2))$$

The bordered Hessian is

$$BdH = \begin{bmatrix} 0 & g_1 & g_2 \\ g_1 & f_{11} - \lambda g_{11} & f_{12} - \lambda g_{12} \\ g_2 & f_{21} - \lambda g_{21} & f_{22} - \lambda g_{22} \end{bmatrix}$$

We find the determinant of this bordered Hessian as Equation (5):

$$|BdH| = \det \begin{bmatrix} 0 & g_1 & g_2 \\ g_1 & f_{11} - \lambda g_{11} & f_{12} - \lambda g_{12} \\ g_2 & f_{21} - \lambda g_{21} & f_{22} - \lambda g_{22} \end{bmatrix} = g_1 g_2 (f_{21} - \lambda g_{21}) + g_2 g_1 (f_{12} - \lambda g_{12}) - g_2^2 (f_{11} - \lambda g_{11}) - g_1^2 (f_{22} - \lambda g_{22}) \quad (5)$$

The sufficient condition for a *maximum*, in the bivariate case with one constraint, is that the determinant of its bordered Hessian is positive when evaluated at the critical point. The sufficient condition for a *minimum*, in the bivariate case with one constraint, is that the determinant of its bordered Hessian is negative when evaluated at the critical point.

If x is a regular point and $g_i(x) = 0$ (constraints are satisfied) then $M = \{y | \nabla g_i(x) \bullet y = 0\}$ defines a plane tangent to the feasible region at x .

Lemma: If x is regular, $g_i(x) = 0$, and $\nabla g_i(x) \bullet y = 0$, then $\nabla f(x) \bullet y = 0$.

The Lagrange multiplier conditions are exactly the same for a minimization problem as a maximization problem. This is why these conditions alone are not sufficient conditions. Thus, a given solution can either be a maximum or a minimum. To determine whether the point found is a maximum, minimum, or saddle point, we use the Hessian.

The Lagrange multiplier, λ , has an important modeling interpretation. It is the “shadow price” for scarce resources. Thus, λ_i is the shadow price of the i th constraint. Thus, if the right-hand side (RHS) is increased by a small amount Δ in a maximize or a minimize problem, then the optimal solution will change by $\lambda_i \Delta$. We illustrate the shadow price both graphically and computationally.

Graphical Interpretation of Lagrange Multipliers

The method of a Lagrange multiplier has a geometric interpretation. This geometric interpretation involves the gradients of both the function and the constraints. Initially, let’s consider only one constraint,

$$g(x_1, x_2, \dots, x_n) = b$$

so that the Lagrangian equation simplifies to

$$\nabla f = \lambda \nabla g$$

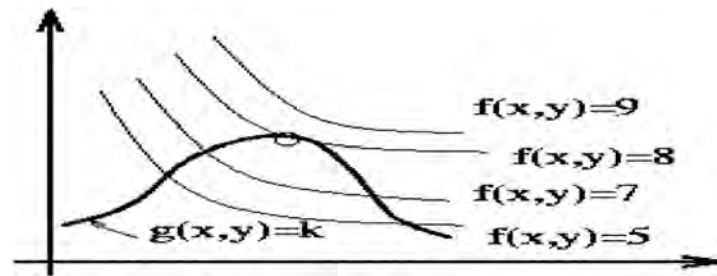
The solution is the point in x where the gradient vector, $\nabla g(x)$, is perpendicular to the surface. The gradient vector, ∇f , always points in the direction in which f increases fastest. At both maximums and minimums, this direction must also be perpendicular to S . Thus, because both ∇f and ∇g both point along the same perpendicular line, then $\nabla f = \lambda \nabla g$. This is where the contours of the objective function are tangent to the constraints, see Figure 1.

Computational Method of Lagrange Multipliers with Applications

Consider the set of equations in Equation (3). This gives $m + n$ equations in the $m + n$ unknowns (x_j, λ_i) . Generally speaking this is a difficult problem to solve without a computer except for simple problems. Also, because the Lagrange multipliers are necessary conditions only (not sufficient), we may find solutions (x_j, λ_i) that are not optimal for our NLP. We need to be able to determine the classification of points found in the solution to the necessary conditions. Commonly used methods as justification include:

- the Hessian matrix or
- the bordered Hessian, $\text{Det}[BH] = 0$.

Example 1: Cobb-Douglass function

Figure 1. Equality constrained Lagrange multipliers for levels curves $f=5, 7, 8$, and 9

Recall the problem suggested in the introduction of the chapter. A company manufactures new smart-phones that are supposed to capture the market by storm. The two main inputs components to the e-phone are the circuit board and the relay switches. The number of smart-phones produced is estimated to equal $E = 200x_1^{\frac{1}{3}}x_2^{\frac{1}{2}}$ where E is the number of smart-phones produced and x_1 & x_2 are the number of circuit board hours and the number of relay hours worked, respectively. Such a function is known to economists as a *Cobb–Douglas function*. Our laborers are paid by the type work they do: the circuit boards and the relays for \$5 and \$10 an hour. We want to maximize the number of smart-phones to make if we have \$150,000 to spend on these components in the short run.

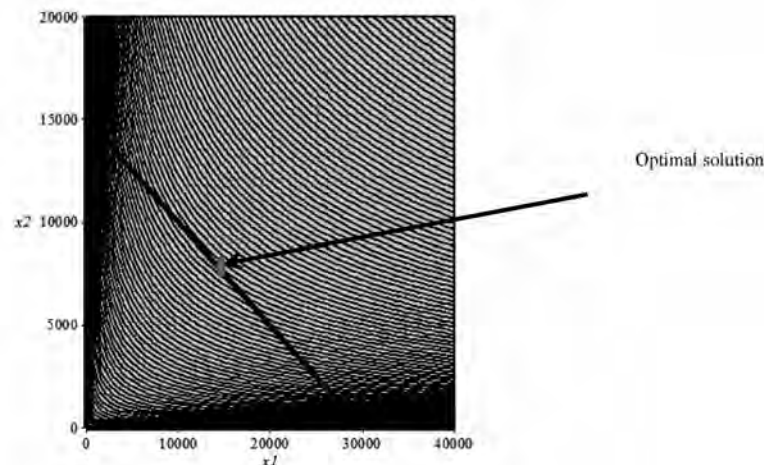
$$\text{Maximize } E = 200x_1^{\frac{1}{3}}x_2^{\frac{1}{2}}$$

subject to the equality constraint

$$5x_1 + 10x_2 = 150,000.$$

Using the method of Lagrange multipliers as described, we find we can make 434,388.51 in selling smart-phones using 12,000 relays and 9,000 circuit boards hours of labor with $\lambda = 2.4132$, see Figure 2. The Hessian is negative definite so we have found the maximum. If the constraint is increase by \$1, we can increase our profits by making 2.41 more smart-phones.

Figure 2. Equality optimization of smart-phone problem



INEQUALITY CONSTRAINTS: KARUSH-KUHN-TUCKER NECESSARY AND SUFFICIENT CONDITIONS

Introduction to KTC

In most realistic problems, many of the constraints are inequalities. These constraints form the boundaries for the solution. The general form of the NLP we present is shown by Equation (6):

$$\begin{aligned} &\text{Maximize (minimize) } z = f(x_1, x_2, x_3, \dots, x_n) \\ &\text{Subject to} \\ &g_1(x_1, x_2, \dots, x_n) \geq b_1 \\ &g_2(x_1, x_2, \dots, x_n) \geq b_2 \\ &\cdot \\ &\cdot \\ &\cdot \\ &g_m(x_1, x_2, \dots, x_n) \geq b_m \end{aligned} \quad (6)$$

One classic method for solving NLPs of this type, Equation (6), is the Karush-Kuhn-Tucker condition (KKTC).

Basic Theory of Constrained Optimization

During this discussion of KTC, we are concerned with problems of the form (6a): maximize or minimize $f(X) = f(x_1, x_2, \dots, x_n)$ subject to

$$g_i(x_1, x_2, \dots, x_n) \begin{cases} \geq \\ = \\ \leq \end{cases} b_i \quad (6a)$$

for $i = 1, 2, \dots, m$.

We allow the constraints to be equal and either less than or equal to or greater than or equal to. We have previously discussed Lagrange multipli-

ers to solve problems with equality constraints. Recall that with Lagrange multiplier the optimal solution actually fell on one constraint or at an intersection of several constraints. With the inequality constraints, the solution no longer must lie on a constraint or at an intersection point of constraints. This concept poses new problems. We need a method to account for the position of the optimal solution relative to each constraint. This KKTC procedure involves setting up a Lagrangian function of the decision variables $\mathbf{X}$, the Lagrange multipliers λ , and the slack or surplus variables $\mathbf{U}$. The $\mathbf{X}_j$ are the decision variables $(x_1, x_2, \dots, x_n)$, the $-\lambda_i$ are the shadow prices for the i th constraint, and the $\mathbf{U}_i$ are either added (slack variables from $\leq$ constraints) or subtracted (surplus variables from $\geq$ constraints). Thus, with the sign of $\mathbf{U}_i$ we are able to accommodate both $\leq$ and $\geq$ constraints.

We set up this generic Lagrangian function, equation (7):

$$L(X, \lambda, U^2) = f(X) + \sum_{i=1}^m \lambda_i (g(X_i) + (\pm U^2) - b_i) \quad (7)$$

U_i^2 depends on the type of inequality constraint, we either need to add or subtract a quantity U_i^2 to make the inequality into an equality. The computational procedure for the KTC requires that all the partials of this Lagrangian function equal zero. These partials are the *necessary conditions* of the NLP problem. These are the conditions required for $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$ to be a solution to Equation (7).

$$\text{The Necessary Conditions} \quad (8)$$

$$\partial L / \partial X_j = 0 \quad (j = 1, 2, \dots, n) \quad (8a)$$

$$\partial L / \partial \lambda_i = 0 \quad (i = 1, 2, \dots, m) \quad (8b)$$

$$\partial L / \partial U_i = 0 \text{ or } 2U_i \lambda_i = 0 \quad (i = 1, 2, \dots, m) \quad (8c)$$

The following two theorems give the sufficient conditions for $x^* = \{x_1, x_2, \dots, x_n\}$ to be an optimal solution to the NLP given in Equation (8).

The Sufficient Conditions (9)

- (9a) **Minimum:** If $f(x)$ is a convex function and each of the $g_i(x)$ are convex functions, then any point that satisfies the necessary conditions is an optimal solution. An optimal point is a point that minimizes the function subject to the constraints. λ_i is greater than or equal to zero for all i .
- (9b) **Maximum:** If $f(x)$ is a concave function and each of the $g_i(x)$ are convex functions, then any point that satisfies the necessary conditions is an optimal solution. An optimal point is a point that maximizes the function subject to the constraints. λ_i is less than or equal to zero for all i .

If these above conditions are not completely satisfied, then we may use another method to check the nature of a potential stationary or regular point, such as the bordered Hessian.

Bordered Hessian

The bordered Hessian is a symmetric matrix of second partials of the Lagrangian:

$$H_B = \partial^2 L / \partial (X_j^2, \lambda_k^2) \{j = 1, 2, \dots, n, k = 1, 2, \dots, m\}$$

We can determine, if possible, the nature of the stationary point by classifying the bordered Hessian. This method is only valid leading to max or min points. If the bordered Hessian is indefinite, then another method should be used.

Complementary Slackness

The KTC computational solution process uses these necessary conditions for solving 2^m possible cases, where m equals the number of constraints. The value 2 comes from the possible conditions

placed on λ_i : It either equals zero or does not equal zero. There is actually more to this process because it really involves the complementary slackness condition embedded in the necessary condition, $2 U_i \lambda_i = 0$. Thus, either U_i equals zero and λ_i does not equal zero or U_i is greater than or equal to zero and λ_i is equal to zero. This ensures that the complementary slackness conditions are satisfied while the other necessary conditions from Equations 7a and 7b are solved.

It is these complimentary slackness necessary conditions, Equation 6c, that leads to the solution process on which we focus our computational and geometric interpretation. We have defined U_i^2 as a slack or surplus variable. Therefore, if U_i^2 equals zero, then our point is on the i th constraint; and if U_i^2 is greater than zero, then the point does not lie on the i th constraint. Furthermore, if the value of U_i^2 is undefined because it equals a negative number, then the point of concern is infeasible.

Computational KKTC With Information Science Application

Example 2: Revisit Cobb-Douglass, Smart-phones, with modifications

Recall the problem suggested in the introduction of the chapter. A company manufactures new smart-phones that are supposed to capture the market by storm. The two main inputs components to the e-phone are the circuit board and the relay switches. The number of smart-phones produced

is estimated to equal $E = 200x_1^{\frac{1}{3}}x_2^{\frac{1}{2}}$ where E is the number of smart-phones produced and x_1 & x_2 are the number of circuit board hours and the number of relay hours worked, respectively. Our laborers are paid by the type work they do: the circuit boards and the relays for \$5 and \$10 an hour. We want to maximize the number of smart-phones to make if we have at most \$150,000 to spend on these components in the short run. We also have capital restrictions on labor. These are:

\$2 for type x_1 and \$20 for type x_2 with a total available capital of \$125,000.

$$\text{Maximize } E = 200x_1^{\frac{1}{3}}x_2^{\frac{1}{2}}$$

Subject to:

$$5x_1 + 10x_2 \leq 150,000$$

$$2x_1 + 20x_2 \leq 125,000$$

$$x_1, x_2 \geq 0$$

Using the method of KKTC, we find we can make 356,513 smart phones by using 21875 and 4062.5 relays respectively. Further our two shadow prices are -0.2611 and -2.06 respectively, which imply that a 1 unit increase in total relays increases production by about 0.261 units while a \$1 increase in capital increases production by 2.06 units. Thus, we would prefer a \$1 investment in capital. Figure 3 shows the graphical analysis. The boarded Hessian is negative definite so we found the maximum.

Numerical and Heuristic Methods

Not all optimization problems are solvable in closed form. For that reason more and more procedures are created to solve these problems. Fox

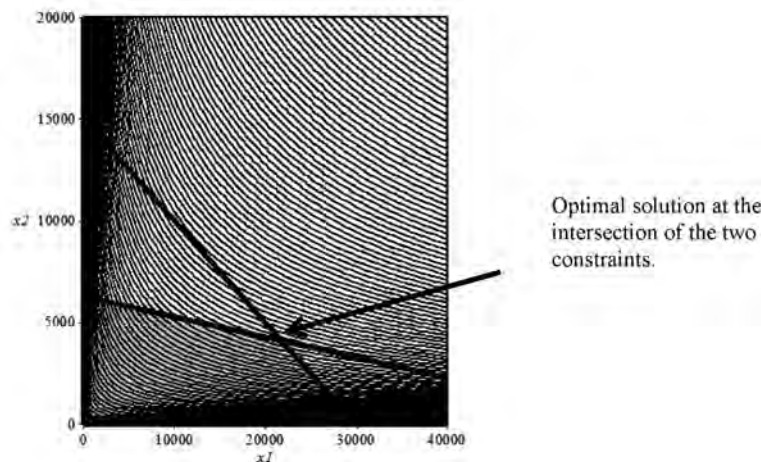
(1990) presented a heuristic algorithm to solve a non-linear objective function with both linear and non-linear constraints as applied to component selection and user specification problems. Xanthopoulos et al. (2011) used a hybrid genetic algorithm to solve constrained optimization in Just-in-Time manufacturing systems in supply chain optimization. Homaifar et al. (1994) and Michalewicz (1991) developed algorithms for genetic algorithm to solve constrained optimization problems with good results. Alvarez-Vasquez et al. (2000) addressed the optimization of citric acid production rate.

Some classical methods include quadratic programming and separable programming methods that convert certain NLP into linear programs (see Bazaraa et al., 1993). Additionally newer numerical optimization method have been created such as simulated annealing, differential evolution, particle swarm optimization, genetic algorithms, and cuckoo search.

- **Quadratic:** Quadratic programming problems (QPP) is a special case of nonlinear optimization where the highest degree of x is squared. Several algorithms exist to solve QPPs.

Shim (1983) discussed some applications of QPP in business.

Figure 3. Graphical analysis of example 2.



- **Separable:** Separable programming is a another special case of nonlinear optimization where we might use approximating functions with piecewise linear functions. Stefanov (2001) presents a good discussion separable programming.
- **Feasible Directions:** This method was invented in 1960 by Zoutendijk and many revisions have been made to improve it over the years. Fox et al. (1993) showed a method of starting points to improve convergence for some of these methods such as feasible directions. Canale (2010) has a good summary of these methods.
- **Particle Swarm Optimization (PSO):** This is a computational method that solves the problem by iteratively trying to improve the candidate solution that was discovered by Kennedy and Eberhart in 1995, and modified by Shi in 1998.. It used a population of candidate solutions, called particles, and moves them around the search space. This is expected to move the swarm toward the best solutions. It was founded for simulating social behavior thus lends itself nicely to information science. Good example are found in Bonyadi et al (2016).
- **The Nelder–Mead Method** is a commonly applied numerical method used to find the minimum or maximum of an objective function in a multidimensional space. It is applied to nonlinear heuristic optimization method in problems for which derivatives may not be known. The Nelder–Mead technique was devised by Nelder and Mead (1965).
- **Simulated Annealing (SA)** is a probabilistic technique for approximating the global optimum of a given constrained optimization problem that was described by Scott Kirkpatrick, C. Daniel Gelatt and Mario P. Vecchi in 1983. Specifically, it is a meta-heuristic to find good approximate global optimization solution in large search spaces.. It is often used when the search space is discrete (e.g., all tours that visit a given set of cities). For problems where finding the precise global optimum is less important than finding an acceptable local optimum in a fixed amount of time, simulated annealing may be preferable to other alternatives. Simulated annealing interprets slow cooling as a slow decrease in the probability of accepting worse solutions as it explores the solution space. Accepting worse solutions is a fundamental property of metaheuristics because it allows for a more extensive search for the optimal solution.

Table 1. Free and Open Source software

Name	License	Brief information
ADMB	BSD	nonlinear optimization framework, using automatic differentiation
APMonitor	BSD	MATLAB Toolbox and Python APIs to Mixed Integer Nonlinear Programming Solvers
ASCEND	GPL	mathematical modeling system
IPOPT	CPL	large scale nonlinear optimization for continuous system
JOptimizer	ASL	open source java library for convex optimization
OpenOpt	BSD	free numerical optimization framework for solving NLP, LP, MIP, QP, with automatic differentiation features.
OpenSolver	GPL	open source Excel addin for solving LP and MIP problems
OPTI Toolbox	BSD	MATLAB Toolbox for solving linear, nonlinear, continuous and discrete optimization problems.
MOEA Framework	LGPL	open source Java library for multi-objective optimization

For Proprietary software, see http://en.wikipedia.org/wiki/List_of_optimization_software.

- **Genetic Algorithms:** In the field of artificial intelligence it is a search heuristic that mimics the process of natural selection. This heuristic is routinely used to generate useful solutions. Genetic algorithms belong to the larger class of evolutionary algorithms (EA), which generate solutions to optimization problems using techniques inspired by natural evolution.
- **Cuckoo Search (CS)** was developed by Yang and Deb (2009). It was inspired by the cuckoo and how they lay their eggs in nests hosted by other birds. Some host birds can engage direct conflict with the intruding cuckoos. For example, if a host bird discovers the eggs are not their own, it will either throw these alien eggs away or simply abandon its nest and build a new nest elsewhere. Cuckoo search idealized such breeding behavior, and thus can be applied for various optimization problems.

Other techniques worthy of consideration are differential search and random search methods which can be found in optimization books.

Technology

There exist many software packages to assist with solving constrained, nonlinear optimization problems. These include and are not limited to LINGO, MAPLE, MATLAB, MATEMATICA, and EXCEL's (SOLVER). We provide a very abbreviated list below of other free software: (from http://en.wikipedia.org/wiki/List_of_optimization_software.)

FUTURE TRENDS

As problems get more complex, the need for more sophisticated and easy to use software will be sought that solve these difficult and complex problems quickly and allow for good sensitivity analysis.

CONCLUSION

We provided a survey of techniques to solve constrained optimization problems in information science and technology.

REFERENCES

- Alvarez-Vasquez, F., Gonzales-Alcon, C., & Torres, N. (2000). *Metabolism of citric acid production by Aspergillus niger: Model definition, steady-state analysis and constrained optimization of citric acid production rate*. New York, NY: John Wiley and Sons.
- Bazaraa, M. S., Sherali, H., & Shetty, C. (1993). *Nonlinear Programming Theory and Algorithms*. New York, NY: John Wiley & Sons.
- Bellman, R. (1952). On the Theory of Dynamic Programming. *Proceedings of the National Academy of Sciences*.
- Bellman, R. (1957). *Dynamic Programming*. Princeton, NJ: Dover.
- Bonyadi, M. R., & Michalewicz, Z. (2016). Particle swarm optimization for single objective continuous space problems: A review. *Evolutionary Computation*, 1–54. doi:10.1162/EVCO_r_00180 PMID:26953883
- Canale, R., & Charpa, S. (2010). *Numerical Methods for Engineers*. London, UK: McGraw Hills Publishers.
- Ecker, J., & Kupperschmid, M. (1988). *Introduction to Operations Research*. New York, NY: John Wiley and Sons.
- Fox, W. (1990). *Application of Pseudo Boolean Models to Weapon System Design* (Dissertation). Clemson University.
- Fox, W. (2012). *Mathematical Modeling with Maple*. Boston, MA: Cengage publishers.

- Fox, W. P., & Combs, T. (1995). Finding a Better Starting Point for the Method of Feasible Directions. *Computers in Education Journal*, V(2), 27–32.
- Giordano, F., Fox, W., & Horton, S. (2012). *A first course in mathematical modeling* (5th ed.). Boston, MA: Cengage Publishers.
- Hillier, F., & Lieberman, G. (1995). *Introduction To Operations Research*. New York, NY: McGraw-Hill Co.
- Homaifar, A., Lai, S., & Qi, X. (1994). Constrained optimization via genetic algorithms. *Simulation*, 62(4), 242–254. doi:10.1177/003754979406200405
- Kennedy, J. (1997). The particle swarm: social adaptation of knowledge. *Proceedings of IEEE International Conference on Evolutionary Computation*, 303–308. doi:10.1109/ICEC.1997.592326
- Kennedy, J., & Eberhart, R. (1995). Particle Swarm Optimization. *Proceedings of IEEE International Conference on Neural Networks*, 1942–1948. Retrieved from <http://www.engr.iupui.edu/~shi/Conference/psopap4.html>
- Kennedy, J., & Eberhart, R. (1995). Particle Swarm Optimization. *Proceedings of IEEE International Conference on Neural Networks*, 1942–1948. doi:10.1109/ICNN.1995.488968
- Kennedy, J., & Eberhart, R. C. (2001). *Swarm Intelligence*. Morgan Kaufmann.
- Michaelwicz, Z., & Koziel, S. (1999). Evolutionary algorithms, homomorphous mappings, and constrained parameter optimization. *Evolutionary Computation*, 7(1), 19–44. doi:10.1162/evco.1999.7.1.19 PMID:10199994
- Michalewicz, Z. (1995). Genetic algorithms, numerical optimization, and constraints. In L.J. Eshelman (Ed.), *Proceeding of the 6th annual international conference on genetic algorithms* (pp. 151–158). Morgan Kaufman Publisher.
- Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *The Computer Journal*, 7(4), 308–313. doi:10.1093/comjnl/7.4.308
- Parsopoulos, K. E., & Vrahatis, M. N. (n.d.). Particle Swarm Optimization Method for Constrained Optimization Problems. In P. Sincak, J. Vascak, V. Kvasnicka, & J. Pospichal (Eds.), *Intelligent Technologies - Theory and Applications: New Trends in Intelligent Technologies*, (pp. 214–220). Academic Press.
- Rechenberg, I. (1971). *Evolutionsstrategie – Optimierung technischer Systeme nach Prinzipien der biologischen Evolution* (PhD thesis). Reprinted by Fromman-Holzboog (1973).
- Shi, Y., & Eberhart, R. C. (1998). A modified particle swarm optimizer. *Proceedings of IEEE International Conference on Evolutionary Computation*, 69–73.
- Shi, Y., & Eberhart, R. C. (1998). A modified particle swarm optimizer. *Proceedings of IEEE International Conference on Evolutionary Computation*, 69–73.
- Shim, J. (1983). A survey of quadratic programming applications to business and economics. *International Journal of Systems Science*, 14(1), 105–115. doi:10.1080/00207728308926445
- Stefanov, S. (2001). *Separable programming: Theory and methods*. Kluwer Academic Publishers. doi:10.1007/978-1-4757-3417-1
- Winston, W.L. (2002). *Introduction to Mathematical Programming Applications and Algorithms* (4th ed.). Belmont, CA: Duxbury Press.
- Wolfe, P. (1959). The Simplex Method for Quadratic Programming. *Econometrica*, 27(3), 382–398. doi:10.2307/1909468
- Xanthopoulos, A. & Koulouriotis, D. (2011). Constrained optimization of JIT manufacturing systems with hybrid genetic algorithms. In *Supply Chain Optimization, Design, and Management*. IGI-Global.

ADDITIONAL READING

- Arsham, H. (1992). Postoptimality analysis for the transportation problem. *The Journal of the Operational Research Society*, 43(2), 121–139. doi:10.1057/jors.1992.18
- Arsham, H. (1993). Managing project activity-duration uncertainties. *Omega: International Journal of Management Science*, 21(2), 111–122. doi:10.1016/0305-0483(93)90043-K
- Arsham, H. (1998). Managing cost uncertainties in transportation and assignment problems. *Journal of Applied Mathematics and Decision Sciences*, 2(1), 65–102. doi:10.1155/S1173912698000042
- Arsham, H. (2004). Global optima for linearly constrained business decision models. *Scientific Journal of Administrative Development*, 2(1), 27–53.
- Arsham, H. (2004). Global optima for linearly constrained business decision models. *Scientific Journal of Administrative Development*, 2(1), 27–53.
- Arsham, H. (2005). Tight bounding of continuous functions over polyhedrons: A new introduction to exact global optimization. *International Journal of Pure and Applied Mathematics*, 19(3), 393–413.
- Barrientos, O., & Correa, R. (2000). An algorithm for global minimization of linearly constrained quadratic functions. *Journal of Global Optimization*, 16(1), 77–93. doi:10.1023/A:1008306625093
- Chen, X., & Fukushima, M. (1999). Proximal quasi-Newton methods for nondifferentiable convex optimization. *Mathematical Programming*, 85A(2), 313–334. doi:10.1007/s101070050059
- Cheng, R., & Jin, Y. (2015). A social learning particle swarm optimization algorithm for scalable optimization. *Information Science*, 291, 43–60. doi:10.1016/j.ins.2014.08.039
- Clerc, M. (2012). Particle swarm optimization, http://clerc.maurice.free.fr/pso/SPSO_descriptions.pdf
- Craven, B., & Islam, S. (2005). *Optimization in Economics and Finance*. New York, NY: Springer.
- Dreyfus, S. (2002). Richard Bellman on the birth of dynamic programming. *Operations Research*, 50(1), 48–51. doi:10.1287/opre.50.1.48.17791
- Eles, P. (2010). Simulated Annealing Department of Computer and Information Science (IDA) Linköpings universitet <http://www.ida.liu.se/~petel/>
- Elton, E., Gruber, M., Brown, S., & Goetzman, W. (2003). *Modern Portfolio Theory and Investment Analysis* (pp. 182–208). New York, NY: John Wiley & Sons.
- Gandomi, A., Yang, X., & Alavi, A. (2013). Cuckoo search algorithm: A metaheuristic approach to solve structural optimization problems. *Engineering with Computers*, 29(1), 17–35. doi:10.1007/s00366-011-0241-y
- Gavin, H. (2016). *The Nelder-Mead Algorithm in Two Dimensions CEE 201L*. Uncertainty, Design, and Optimization Department of Civil and Environmental Engineering Duke University.
- Gupta, P., & Bhatia, D. (2000). Multiparametric analysis of the maximum tolerance in quadratic programming problems. *Operations Research*, 37(1), 36–46.
- Hassan, R., Cohan, B. & de Weck, O. () A comparison of particle swarm optimization and the genetic algorithm. Massachusetts Institute of Technology, Cambridge, MA, 02139 Gerhard Venter§ Vanderplaats Research and Development, Inc., Colorado Springs, CO, 80906
- Haugen, R. (1997). *Modern Investment Theory* (pp. 92–130). Upper Saddle River, NJ: Prentice Hall.
- Ibidapo-Obe, O., Asaolu, O., & Badiru, A. (2002). A new method for the numerical solution of simultaneous nonlinear equations. *Applied Mathematics and Computation*, 125(1), 133–140. doi:10.1016/S0096-3003(00)00125-9

- Jacobson, L. (2013). Simulated annealing for beginners. <http://www.theprojectspot.com/tutorial-post/simulated-annealing-algorithm-for-beginners/6>
- Kallio, M., & Salo, S. (2000). An interior point method for solving systems of linear equations and inequalities. *Operations Research Letters*, 27(3), 101–107. doi:10.1016/S0167-6377(00)00036-5
- Kaul, R., Bhatia, D., & Gupta, P. (1999). Tolerance approach to sensitivity analysis in quadratic programs. *Opsearch*, 36(1), 1–9.
- Kirkpatrick, S. (1984, March). Optimization by simulated annealing: Quantitative studies. *Journal of Statistical Physics*, 34(5), 975–986. doi:10.1007/BF01009452
- Kirkpatrick, S., Gelatt, C., & Vecchi, M. (1983). Optimization by Simulated Annealing S. New Series: Vol. 220. No. 4598. (May 13, 1983) (pp. 671–680). Stable; URL <http://links.jstor.org/sici?sici=0036-8075%2819830513%293%3A220%3A4598%3C671%3A0BSA%3E2.0.CO%3B2-8>
- Klein, K., & Neira, J. (2013). *Nelder-Mead Simplex Optimization Routine for Large-Scale Problems: A Distributed Memory Implementation*. Comput Econ. Springer Science Business Media New York; doi:10.1007/s10614-013-9377-8
- Konak, A., Coit, D., & Smith, A. (2006). Multi-objective optimization using genetic algorithms: A tutorial. *Reliability Engineering & System Safety*, 91(9), 992–1007. doi:10.1016/j.ress.2005.11.018
- Kucherenko, S., & Sytsko, Y. (2005). Application of deterministic low-discrepancy sequences in global optimization. *Computational Optimization and Applications*, 30(3), 297–318. doi:10.1007/s10589-005-4615-1
- Lynch, B. (2006) Optimizing with Genetic Algorithms <https://www.msi.umn.edu/sites/default/files/OptimizingWithGA.pdf>
- Mathews, J., & Fink, C. (2004). Numerical Methods Using Matlab, 4th Edition, 2004 John H. Mathews and Kurtis K. Fink ISBN: 0-13-065248-2 Prentice-Hall Inc. Upper Saddle River, New Jersey, USA <http://vig.prenhall.com/>, pp 430-434
- McCall, J. (2005). Genetic algorithms for modeling and optimization. *Journal of Computational and Applied Mathematics*, 184(1), 205–222. doi:10.1016/j.cam.2004.07.034
- Mijangos, E., & Nabona, N. (2001). On the first-order estimation of multipliers from Kuhn-Tucker systems. *Computers & Operations Research*, 28(3), 243–270. doi:10.1016/S0305-0548(99)00102-1
- Nocedal, J., & Wright, S. (2006). *Numerical optimization*. New York, NY: Springer.
- Poli, R., Kennedy, J., & Blackwell, T. (2007). Particle swarm optimization An overview, *Springer Science and Business Media*, LLC 2007.
- Rajabioun, R. (2011). Cuckoo Optimization Algorithm. [Cuckoo Optimization Algorithm Ramin Rajabioun* Control and Intelligent Processing Centre of Excellence] [CIPCE] [, School of Electrical and Computer Engineering, University of Tehran, Tehran, Iran]. *Applied Soft Computing*, 11(8), 5508–5518. www.elsevier.com/locate/asoc doi:10.1016/j.asoc.2011.05.008
- Rini, D. Particle Swarm Optimization: Technique, System and Challenges, *International Journal of Computer Applications* (0975 – 8887) 14(1), January 2011.
- Schrijver, A. (2003). *Combinatorial optimization: Polyhedra and efficiency*. New York, NY: Springer Verlag.
- Sherali, H., Choi, G., & Tuncbilek, C. (2000). A variable target value method for nondifferentiable optimization. *Operations Research Letters*, 26(1), 1–8. doi:10.1016/S0167-6377(99)00063-2

Steinbach, M. (2001). Markowitz revisited: Mean-variance models in financial portfolio analysis. *SIAM Review*, 43(1), 31–85. doi:10.1137/S0036144500376650

Stonedahl, F., & Wilensky, U. (2008). NetLogo Particle Swarm Optimization model, <http://cc1.northwestern.edu/netlogo/models/ParticleSwarmOptimization>. Center for Connected Learning and Computer-Based Modeling. Northwestern University, Evanston, IL.

van Laarhoven, P., & Aarts, E. (1987). *Simulated Annealing: Theory and Application*. Springer Publishers. Springer Netherlands. doi:10.1007/978-94-015-7744-1

van Laarhoven, P., & Aarts, E. H. L. D. Reidel, Dordrecht. (1988). Simulated annealing: Theory and applications, *Acta Applicandae Mathematica*, 12(1). Pp. 108-111.

Yamashita, H., & Yabe, H. (2005). Quadratic convergence of a primal-dual interior point method for degenerate nonlinear optimization problems. *Computational Optimization and Applications*, 31(2), 123–144. doi:10.1007/s10589-005-2177-x

Yang, X., & Deb, S. (2009). Cuckoo Search via Lévy Flights “Cuckoo search via Lévy flights”, in: *Proc. of World Congress on Nature & Biologically Inspired Computing (NaBIC 2009)*, December 2009, India. IEEE Publications, USA, pp. 210-214 (2009). doi:10.1109/NABIC.2009.5393690

KEY TERMS AND DEFINITIONS

Binding Constraint: A constraint that is satisfied at equality.

Constraints: A resource constraint is one that either limits the amount of a resource or sets a minimum to amount of a resource requirement.

Decision Variables: The variables that change that affect the solution.

Feasible Region: The convex region bounded by the constraints.

Initial Values: Starting values for numerical methods. Default values are usually all 0's or all 1's but care must be taken to insure convergence.

Kuhn-Tucker: A nonlinear optimization methodology that involved inequality constraints.

Lagrange: A nonlinear optimization methodology that involved equality constraints.

Lagrange Multiplier: The Lagrange multiplier, λ , is used in sensitivity analysis in optimization problems. In general the objective function changes by the value of λ for each unit change in the resource.

Numerical Methods: When a function cannot be solved in closed form an appropriate numerical technique may be used to approximate the solution.

Optimization: The process of maximizing or minimizing an objective function.

Decimal Hardware Multiplier



Mário Pereira Vestias
INESC-ID/ISEL/IPL, Portugal

INTRODUCTION

IEEE-754 2008 has extended the standard with decimal floating point arithmetic. Human-centric applications, like financial and commercial, depend on decimal arithmetic since the results must match exactly those obtained by human calculations without being subject to errors caused by decimal to binary conversions. Decimal Multiplication is a fundamental operation utilized in many algorithms and it is referred in the standard IEEE-754 2008. Decimal multiplication has an inherent difficulty associated with the representation of decimal numbers using a binary number system. Both bit and digit carries, as well as invalid results, must be considered in decimal multiplication in order to produce the correct result.

This article focuses on algorithms for hardware implementation of decimal multiplication. Both decimal fixed-point and floating-point multiplication are described, including iterative and parallel solutions.

BACKGROUND

Usually, humans perform arithmetic operations using decimal arithmetic. However, computers do it with binary arithmetic. It means that performing decimal operations in a computer without support for decimal arithmetic is subject to errors from representing decimal numbers, converting them and rounding. In fact, it is easy to find decimal numbers that cannot be represented exactly in binary format (e.g., 0.1). Several examples exist where errors due to binary calculation of decimal numbers are obtained. A clarifying example came from the Vancouver Stock Exchange (Quinn, K.,

1983), where due to rounding errors an initial index value of 1000.000 dropped to 574.081 instead of the correct result of 1098.892.

In fact, the business and commercial markets were one of the triggers for the importance of decimal computer arithmetic since many commercial databases have more than 50% of the numerical data represented in decimal (Tsang, A. & Olschanowsky, M., 1991). In these cases, to avoid errors with undesirable consequences it is important to have a complete system to support decimal arithmetic.

At the era of electronic computers, both binary and decimal arithmetic functions were considered. We had computer systems, like the ENIAC (Goldstine, H. & Goldstine, A., 1996) and IBM 650 (Knuth, D., 1986) implementing arithmetic functions in decimal, and others like EDSAC (Wilkes, M., 1997) and EDVAC (Williams, M., 1993) that adopted binary based arithmetic implementations. Both arithmetic systems were still considered after the advent of transistorized computers with decimal numbers represented with four bits following different representations, like in Binary-Coded Decimal (BCD) format. However, soon binary arithmetic was adopted by most computer systems since at that time scientific computing, whose operations could be more efficiently implemented in binary, were more in demand than financial computing that requires decimal arithmetic to avoid costs from representation errors. So, binary became very popular, while decimal was supported only by some computers in the 1960s and 1970s.

Precise decimal arithmetic operations with binary based computing systems are done in software. In some cases, these binary-based computing systems include some specific hardware

DOI: 10.4018/978-1-5225-2255-3.ch400

instructions that are hardware supported and so software algorithms can take advantage of them to speed-up execution. Several languages include primate decimal datatypes, including Ada, COBOL, and SQL. Several other languages support the GDAS (General Decimal Arithmetic Specification) (Cowlshaw, M., 2008), including the IBMCDecNumberLibrary (Cowlshaw), the Java BigDecimal (Sun Microsystems), Eiffel Decimal Arithmetic (Crismer), Python Decimal (Batista), among others. Decimal floating point extensions conforming to the IEEE 754-2008 standard were proposed for C (JTC 1, 2007) and C++ (JTC 1, 2008) languages. These extensions were supported by GNU C compiler 4.2 release. Intel has also developed a decimal floating-point math library (Intel) that implements decimal floating-point arithmetic specified in IEEE 754-2008.

Hardware support for decimal arithmetic is needed if the percentage of time spent executing decimal functions from these software libraries is relevant. Two different perspectives have emerged in the end of the last decade. Wang (Wang, L.-K., et al., 2007), examined several financial benchmarks and concluded that the time spent on executing decimal operations ranged from 33.9% to 93.1%. On the contrary, a research from Intel (Cornea, M. & Crawford, J., 2007) concluded that most commercial applications spend less than 5% executing decimal operations. Therefore, hardware for decimal arithmetic is not a priority in the design of Intel's processors. In fact, Intel x86 processors offer only a set of eight fixed-point decimal arithmetic instructions, and Motorola 68K reduces this set to just five instructions. On the other side, several IBM's processors include a considerable support for decimal arithmetic. The S/390 processor (ESA/390, 2001) includes a dedicated decimal adder to execute decimal fixed point addition, subtraction, multiplication and division. The last two, are executed iteratively using additions and subtractions. The IBM System z9 (Duale, A. et al., 2007) and System z10 (Schwarz, E., Kapernick, J., & Cowlshaw, M., 2009) already include a decimal floating-point arithmetic unit

in conformance with IEEE 754-2008 standard. The GNU C Compiler (GCC) 4.3 Release and several compilers from IBM (e.g., IBM XL C/C++ (IBM)) were extended and developed to utilize the dedicated instructions and hardware units present in these IBM's processors.

DECIMAL MULTIPLICATION

Hardware implementations for decimal multiplication can be classified according to the type of operands to be multiplied as fixed or floating-point, whether the operands are fixed or floating-point, respectively.

Fixed-point multiplication follows generically the typical hand process that starts by generating partial products, followed by reduction of the partial products using decimal addition. The process can be either based on iterative or parallel algorithms. In the iterative approach partial products are generated and accumulated step-by-step in an iterative process, while in the parallel case partial products are generated in parallel and merged with an adder tree.

Decimal floating-point multiplication typically uses a fixed-point decimal multiplier to multiply the trailing significant fields, together with exponent addition, rounding and sign calculation, similar to a binary multiplier.

Designs for both fixed- and floating-point multipliers were proposed in different target technologies: Application Specific Integrated Circuit (ASIC) and Field Programmable Gate Arrays (FPGAs). FPGA-based solutions are flexible and can be configured with different operand sizes. Many existing solutions are optimized for specific technologies and therefore may not be the most appropriate solution when migrated from one technology to the other.

In the following, the basic algorithms and architectures of each type of decimal multiplier are introduced together with state-of-the-art proposals based on each of these types and technologies.

DECIMAL FIXED-POINT MULTIPLICATION

Whether in the iterative or in the parallel method, partial product generation in a fixed-point multiplication can be either the result of digit by digit, digit by word or word by word multiplication. In a digit by digit multiplication process each digit of the multiplier is multiplied by a digit of the multiplicand. A faster approach is to multiply each digit of the multiplier by the whole multiplicand, that is, a multiple of the multiplicand is accumulated for each digit of the multiplier. Word by word multiplication is a method that seeks to reduce the number of partial products at the cost of more complex partial product generation.

Digit by Digit Multiplication Design

Different designs are obtained for digit by digit iterative multiplication depending on the how the digits are traversed. One approach consists on multiplying a digit of the multiplier by all digits of the multiplicand from the least significant digit to the most significant digit. Thus, the multiplier must run all digits for each digit of the multiplier (see figure 1a).

At each step, a new digit by digit product must be added to the accumulated partial product using an n digit adder. Also, the partial product must be aligned right after multiplying a digit of the multiplicand and realigned left for each new digit of the multiplier

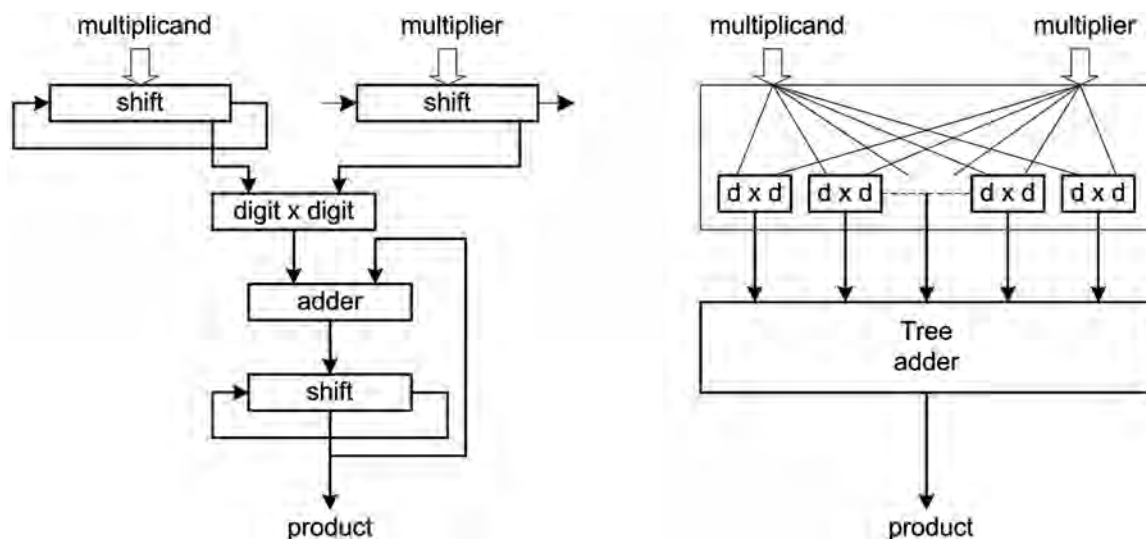
One of the disadvantages of this approach is that we need an n digit adder. A different approach consists on choosing the digits of the multiplicand and of the multiplier to be multiplied at each step so that the digits of the product are obtained in order from the least significant digit. For example, consider the multiplication of $A = A_1A_0$ and $B = B_1B_0$, the method calculates the product in the following order:

$$A \times B = A_0 \times B_0 + (A_0 \times B_1 + A_1 \times B_0)10 + (A_1 \times B_1)100$$

In the example, after the first $A_0 \times B_0$ multiplication, the partial product is shifted. Then, accumulate all digit by digit products with weight 10 and shift. Finally, accumulate the final product with weight 100.

The algorithm requires a more elaborate control over the shifts since the partial product only

Figure 1. Digit by digit multiplier with each digit of the multiplier multiplied by all digits of the multiplicand. a) Iterative and b) Parallel



shifts after accumulating all products of the same weight. On the other side, the width of the adder just has to be enough to accumulate all products of the same weight, that is, $\log_{10}(n \times 9^2)$ digits, where n is the number of digits of the multiplier.

In a digit by digit parallel design all digit products are generated in parallel and then accumulated using some form of adder tree (see Figure 1b). The solution requires more hardware, but it is faster. Considering the multiplication of $n \times n$ digits, the digit by digit method runs in $n \times n$ steps, while with the parallel approach all partial products are generated in a single and the same cycle followed by some more cycles to add all partial products.

Either iterative or parallel an important aspect in a digit by digit approach is how to multiply a digit by a digit. The product is either obtained from a lookup table or from an implementation with random logic. A closer look at the possible products reveals that from the total of 100 possible input combinations there are only 37 possible products. This redundancy can be reduced by recoding the digits of the operands. One such example illustrating the advantages of recoding was proposed in (Carlough, S., & Schwarz, E., 2006) and (Erle et al., 2005) with the digit range $\{-5, -4, -3, -2, 0, -1, 1, 2, 3, 4, 5\}$ which produces only 15 different unsigned products.

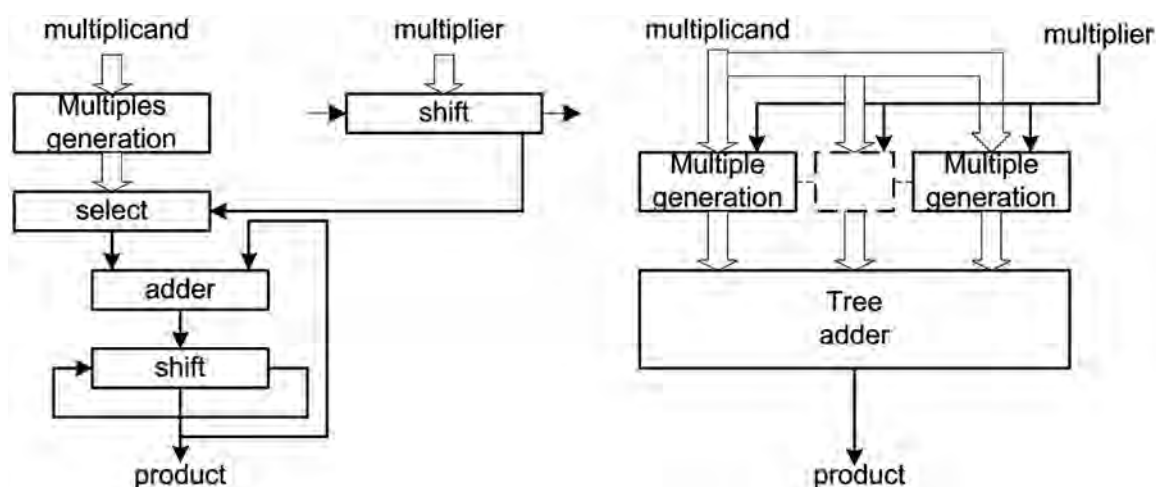
Only a few works have followed a digit by digit multiplication approach to design decimal multipliers. In (Larson, R., 1973) and (Ueda, T., 1995), digit by digit product is implemented with a lookup table. While in (Jaberipur, G. & Kaivani, A., 2007) the product is generated with combinational logic. The problems associated with this approaches are the complexity associated with generation of a digit by digit multiplier, which in a parallel solution requires a considerable amount of logic, and the number of cycles required in an iterative solution.

Digit by Word Multiplication Design

A faster approach for decimal multiplication is to multiply each digit of the multiplier by the whole multiplicand, that is, a multiple of the multiplicand is accumulated for each digit of the multiplier (see word by digit multiplier in Figure 2).

The multiples of the multiplicand may be generated on the fly or a priori and stored. Usually, only a subset of the ten possible multiples are generated and then added to generate the complete set of multiples. In (Busaba, F., 2004) even multiples are precomputed $\{2X, 4X, 6X, 8X\}$ and stored. The odd multiples are generated from the even multiples adding X . In (Erle, M. and Schulte, M., 2003) the multiples $\{X, 2X, 4X, 5X\}$ are precomputed. These multiples have the

Figure 2. Digit by word multiplier. a) Iterative and b) Parallel



advantage of being calculated without carry propagation. In (Vazquez, A. et al., 2007) the multiplier is based on signed-digit radix-10 recoding of the multiplier and precomputation of multiples $\{-5X, -4X, \dots, -X, 0, X, \dots, 4X, 5X\}$. The partial products are added using a BCD 4221 carry-save adder. This work has been followed in (Vázquez, A., Antelo, E. & Bruguera, J., 2014) other redundant codes, namely BCD excess-3 and an overloaded BCD representation. In (Lang, T. & Nannarelli, A., 2006) the multiplier is recoded to a signed digit representation as $Y_i = Y_H \cdot 5 + Y_L$, with $Y_H \in \{0, 1\}$ and $Y_L \in \{-2, -1, 0, 1, 2\}$. In this case, only multiples $2X$ and $5X$ have to be generated and a $10'$ complement to generate negative multiples. The method generates two partial products for each multiplier digit.

A few iterative digit by word multipliers have been proposed. In (Fukuta, M., 1987) the multiplicand is iteratively multiplied by one digit of the multiplier to generate a partial product. Partial products are then added to produce the final decimal product. Other decimal multipliers have been proposed based on sequential units, such as (Kenney, R., Schulte, M., & Erle, M., 2004) and (Erle, M. & Schulte, M., 2003), where a set of multiplicand multiples is generated in a preprocessing step and then selectively added according to the value of the multiplier digits. (Kaivani, A., Chen, L., & Ko, S., 2012) proposes a high frequency sequential multiplier using a fast carry-save adder. Sutter et al. (Sutter, G., et al., 2009) also propose an iterative decimal multiplier using an $N \times 1$ BCD multiplier and a BCD carry chain adder that accumulates the partial results in FPGA. Véstias in (Véstias, M., & Neto, H., 2011) proposed an iterative approach based on binary arithmetic also in FPGA.

Many more approaches have considered a parallel approach (Dadda, L. & Nannarelli, A., 2008), (Jaberipur, G. & Kaivani, A., 2009), (Vázquez, A., Antelo, E. & Montuschi, P., 2010), (Guardia, M., 2012), (Véstias, M. & Neto, N., 2012), (Han, L. & Ko, S.-B., 2013), (Zhu, M., Baker, A. & Jiang, Y.,

2013) using some different internal representations and different tree adders in both technologies.

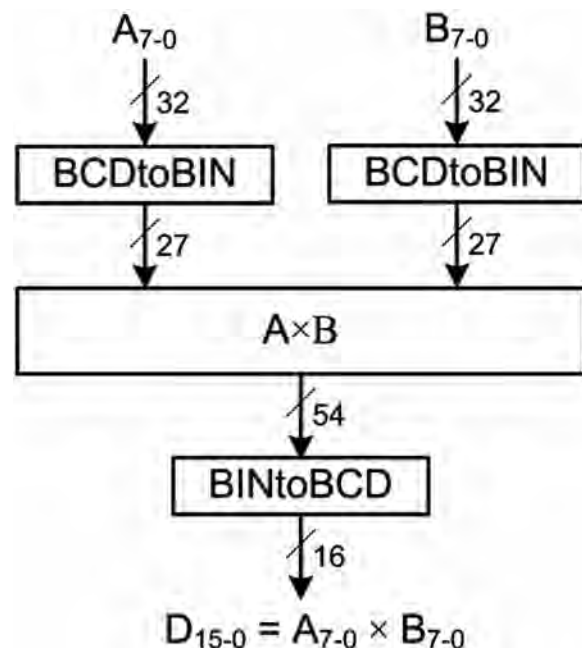
Word by Word Multiplication Design

Considering word by word multiplication as the case in which a partial product results from the multiplication of more than one digit from the multiplier and from the multiplicand, the typical approaches that consider multiples generation of the multiplicand does not apply.

Considering word by word multiplication an approach was considered in (Neto, H. & Véstias, M., 2008) and extended in (Véstias, M. & Neto, H., 2010) using binary arithmetic implemented in FPGA. The decimal operands are first converted to binary, multiplied in binary and then the result is converted back to decimal (see example with 8-digit operands in Figure 3).

In the example, 8-digit operands are converted to binary. The result is a 27-bit number. A 27×27 multiplier produces a 54-bit product to be converted to a 16-digit number. Digits of the

Figure 3. Decimal multiplier using a binary multiplier



multiplier and the multiplicand can be grouped (e.g., 8-digit operands can be grouped in two 4-digit numbers and then partially multiplied as 4×4) in which case several partial products are generated. This method can be implemented as iterative or parallel. This approach is particularly attractive when binary multipliers are available, like in a microprocessor or in an FPGA where embedded binary multipliers are available. A major problem of this solution is that binary to decimal conversion is a slow process using the shift and add-3 algorithm. For large numbers, a fast parallel implementation of this algorithm consumes a considerable amount of hardware resources. Recently, Neto and Véstias (Neto, H. & Véstias, M., 2008) have proposed a new binary to decimal converter that is faster and uses about half of the resources when compared to the traditional shift and add-3 algorithm. With this new converter, the decimal multiplication using binary multipliers becomes quite competitive with the more common approaches based on direct decimal manipulation. A similar contribution was done in (Juang T.-B.; Chiu Y.-M. (2013) with a new high-speed binary to BCD converter for decimal multiplication.

DECIMAL FLOATING-POINT MULTIPLICATION

Decimal floating-point (DFP) multiplication has been proposed and implemented many years ago. For example, the IBM Type 1620 computer already considered variable precision decimal floating-point features. But it was only recently that decimal floating-point formats were standardized with the published IEEE 754-2008 standard. This brought a set of proposals for designing decimal floating-point units including multiplication.

IEEE 754-2008 Standard

IEEE 754-2008 published in August, 2008, defines three binary and two decimal floating-point

number formats. The inclusion of DFP format was probably the most relevant aspect considered in the standard.

While in the binary floating-point format (BFP) the operands and the result must be normalized, in the DFP format these do not have to be normalized. Hence, considering a number with precision p , a product with $p - m$ digits will have m different equivalent representations, as long as the exponent has sufficient range to allow the shift of the leading or trailing non-zero digits. For example, using 4 digits of precision, the number 45×10^8 can be represented as 450×10^7 or 4500×10^6 . This flexibility has led to the introduction of the concept *preferred exponent*. In case of the multiplication, the *preferred exponent* of $A \times B$ is $(E^A + E^B)$.

IEEE 754-2008 defines three formats for DFP: *decimal32*, *decimal64* and *decimal128*. Data in each format is encoded in three fields: (1) one bit sign S ; (2) a combination field G and (3) a trailing significand field. The trailing significant field represents all decimal digits of the number, except the most significant. Decimal digits in the field are either encoded in *Densely Packed Decimal* (DPD) format (M. F. Cowlishaw, 2002) or in *Binary Integer Decimal* (BID) format. The DPD encoding is a specific format to compress three decimal digits into 10 bits, instead of the 12 bits required using simple Binary Coded Decimal.

The operations specified in the standard include addition, subtraction, multiplication, division, square-root and fused multiply-addition; two new decimal operations: samequantum and quantize, used for exact computations; comparisons; and different types of conversions.

Five rounding modes are considered in the standard and also five exceptions (see IEEE 754-2008 standard for further details).

Floating-Point Multiplication Design

Decimal floating-point multiplication typically uses a fixed-point decimal multiplier to multiply

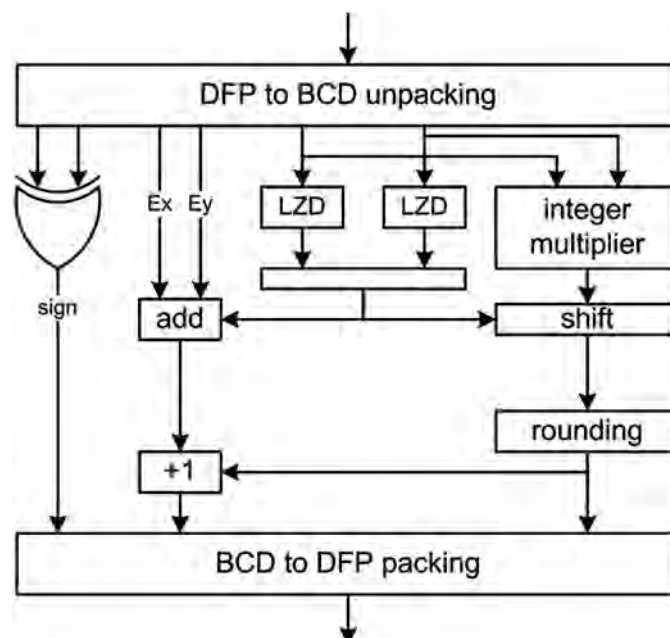
the trailing significant fields, together with exponent addition, rounding and sign calculation. The preferred exponent rule, for exact results, is Q_x (exponent of x) + Q_y (exponent of y). As in the binary case, exactness may be computed by normalizing the product and check the round and sticky bits. Another strategy to check the exactness is to predict if the sum of leading zeros (LZ) of both multiplier and multiplicand is larger than the number of digits of each operand. This is generally the common case and thus the product is exact, the calculated exponent is correct and needs no normalization. Otherwise, if the result is inexact or too large to be representable with the preferred exponent then the result of the multiplication needs to be normalized and then rounded (see Figure 4).

In a typical implementation, the operands are first unpacked to extract the sign, the exponent and the significand field. The signs of both operands are XORed to determine the sign of the product. The exponents are added and the significand fields are multiplied. The product is normalized with a left shift of $\min(LZ_x + LZ_y, \text{operands size})$ to remove the

leading zeros up to the operands size. The output of the integer multiplier is rounded, the exponents are adjusted according to the rounding requirements and the product is packed again from BCD to decimal floating-point format.

There are implementations of decimal floating-point multipliers in commercial microprocessors from IBM, namely Power6 (Eisen, L., et al., 2007) and z10 (Webb, C., 2008). Both architectures use the same word by digit integer multiplier. The implementation improves the execution of decimal multiplication compared to software solutions, but is relatively slow compared to binary multiplication since the multiplication is iterative. A slightly faster iterative floating-point multiplier compliant with IEEE 754-2008 was proposed in (Erle, M., et al., 2007) and a faster solution was proposed in (Hickman, B., et al, 2007) based on a fixed-point parallel multiplier. In (Navarro, G.; et al., 2013), a multiplier that operates on binary integer decimal (BID) encoded decimal floating-point (DFP) numbers is proposed. The multiplier is compliant with the IEEE 754-2008 Standard.

Figure 4. Generic architecture of a decimal floating-point multiplier



FUTURE RESEARCH DIRECTIONS

Decimal multiplication is an important operation for applications where errors are generated from representing the decimal numbers in binary are unacceptable. Decimal multiplication and, in general, decimal arithmetic has not received much attention until 1990s. Now, this field of research is getting much attention and became a hot topic of research.

In particular, decimal multiplication is already part of some commercial processors with dedicated decimal arithmetic instructions. However, the area still needs a set of real benchmarks to help us determine how far should we go about dedicated decimal arithmetic hardware, and in particular, decimal multiplier. What gains can be achieved with dedicated decimal multipliers and this influences the performance of binary arithmetic.

In fact, the integration of decimal arithmetic units with binary arithmetic units deserves further research. How to best integrate decimal hardware into an existing binary arithmetic unit is also an open question. A few solutions already combine decimal floating-point and binary floating-point in a single arithmetic unit (Tsen, S., 2009). Combining decimal and binary arithmetic is advantageous in terms of area but degrades the performance of binary multiplication. Therefore, it is expected to find combined decimal/binary multipliers to have latencies and throughputs close to those of isolated binary multipliers. The emphasis to achieve such similarities should be on the carry save adders used to add partial products so that the binary carry save adder can be used for both radices.

Decimal multiplication based on binary arithmetic also needs further research since it would be possible to integrate binary/decimal converters, a decimal multiplier and a binary multiplier in a single unit. Efficient binary/decimal converters are a decisive aspect for the success of this approach. Binary-based decimal multiplication is a type of combined binary/decimal operation and should be further researched. Processors or devices where binary multipliers are available can

be easily extended with decimal multiplication by using dedicated binary to decimal and decimal to binary converters in hardware or in software. Up to date, hardware for conversions between binary and decimal introduce a considerable delay in the execution time of the operation and so, decimal multiplication based on binary arithmetic is still slower than decimal based multipliers. However, further research over binary-decimal converters may change the scenario.

Word-by-word multiplication should be further investigated since it potentially reduces the number of partial products and consequently the complexity of the adder tree, including its layout routing. The problem has to do with the generation of partial products from multiples of the multiplicand with more than a single digit.

Iterative solutions are attractive since they reduce considerably the area of the design. However, even digit-by-word iterative multiplication may be too costly in terms of latency. Hence, iterative word-by-word solutions where words represent a group of digits should be further considered as a way to reduce latency while keeping an acceptable design area.

Fused operators must also be further investigated. For example, the fused multiply-add operation is very common and specified in the IEEE 754-2008 standard. But, only a few works have considered this problem. Therefore, this operations needs to be explored both in the context of a processor and also as a dedicated hardware fused unit.

Most general-purpose processors still do not include explicit hardware support for the main decimal operations since most of the generic applications they are supposed to execute are binary-based. A different approach could consider an application specific processor for decimal-based applications or an heterogeneous solution with a general-purpose processor and an FPGA that can be reconfigured to run any operation or function with decimal arithmetic. This solution can be easily achieved and have shown very good speed-ups when running decimal arithmetic.



Extended decimal precision arithmetic is still very immature and needs further research at least to consider the decimal128 format of the standard. Some previous decimal multipliers are scalable and can therefore be easily extended to support this format. Once again FPGAs are very attractive to implement high-precision multiplication not only with 128 digit operands but also any other number of digits.

CONCLUSION

The article describes the backgrounds of fixed and floating-point decimal multiplication. The typical architectures of both iterative and parallel multipliers are explained and state-of-the-art implementations for each solution are provided, including some recoding techniques to get faster or smaller implementations.

A considerable number of publications, patents and implementations have appeared during the last 50 years about binary arithmetic. On the other side, decimal arithmetic hardware is still in its infancy but already achieved very interesting results compared to those on binary arithmetic. Hardware designers for decimal arithmetic want to narrow the performance gap between decimal and binary arithmetic so that applications start using more decimal operations. So, the research about decimal multiplication, in particular, is an actual topic of research.

Recent publications proposed new operand recoding and new methods for binary to decimal conversions to improve the delay of decimal hardware multipliers. Therefore, both approaches are still considered, that is, direct decimal multiplication and binary-based decimal multiplication.

Whether decimal operands should be integrated with processors or in co-processing devices, like FPGAs, is also an open research question. If not used in processors, all decimal multiplier proposals will only be used as hardware blocks in some dedicated hardware implementation.

REFERENCES

- Batista, F. (n.d.). *Decimal Data Type*. Retrieved from <http://www.python.org/dev/peps/pep-0327>
- Busaba, F., Slegel, T., Carlough, S., Krygowski, C., & Rell, J. (2004). Design of the Fixed Point Unit for the z990 Microprocessor. *Proceedings of the 14th ACM Great Lakes Symposium on VLSI*, 364-367. doi:10.1145/988952.989040
- Carlough, S., & Schwarz, E. (2006). *Decimal Multiplication Using Digit Re-coding*. U.S. Patent, #7,136,893.
- Cornea, M., & Crawford, J. (2007). IEEE 754R Decimal Floating-Point Arithmetic: Reliable and Efficient Implementation on Intel® Architecture Platforms. *Intel Technology Journal*. Retrieved from <http://www.intel.com/technology/itj/2007/v11i1/s2-decimal/1-sidebar.htm>
- Cowlishaw, M. (2002). Densely Packed Decimal Encoding. *IEE Proceedings. Computers and Digital Techniques*, 149(3), 102-104. doi:10.1049/ip-cdt:20020407
- Cowlishaw, M. (2008). *General Decimal Arithmetic Specification*. Retrieved from <http://speleotrove.com/decimal/decarith.html>
- Cowlishaw, M. (n.d.). *IBM decNumber Library*. Retrieved from <http://www.alphaworks.ibm.com/tech/decnumber>
- Crismer, P. (n.d.). *Eiffel Decimal Arithmetic Library*. Retrieved from <http://eiffelzone.com/esd/eda/index.html>
- Dadda, L., & Nannarelli, A. (2008). A Variant of a Radix-10 Combinational Multiplier. *Proceedings IEEE International Symposium on Circuits and Systems*, 3370-3373. doi:10.1109/ISCAS.2008.4542181
- Duale, A., Decker, M., Zipperer, H.-G., Aharoni, M., & Bohizic, T. (2007). Decimal Floating-Point in Z9: An Implementation and Testing Perspective. *IBM Journal of Research and Development*, 51(1.2), 217-228. doi:10.1147/rd.511.0217

- Eisen, L., Ward, J. W., Tast, H.-W., Mading, N., Leenstra, J., Mueller, S. M., & Carlough, S. R. et al. (2007). IBM Power6 Accelerators: VMX and DFU. *IBM Journal of Research and Development*, 51(6), 663–684. doi:10.1147/rd.516.0663
- Erle, M., & Schulte, M. (2003). Decimal Multiplication Via Carry-Save Addition. *IEEE International Conference on Application Specific Systems, Architectures and Processors*, 348–358. doi:10.1109/ASAP.2003.1212858
- Erle, M., Schulte, M., & Hickman, B. (2007). Decimal Floating-Point Multiplication via Carry-Save Addition. *18th Symposium on Computer Arithmetic*, 46–55. doi:10.1109/ARITH.2007.14
- Erle, M., Schwarz, E., & Schulte, M. (2007). Decimal Multiplication with Efficient Partial Product Generation. *17th IEEE Symposium on Computer Arithmetic*, 21–28.
- ESA. (2001). *390 Principles of Operation, Decimal Arithmetic Instructions*. IBM.
- Fukuta, M., Ishikawa, S., Ohtsuki, T., Oshima, Y., & Yabe, H. (1987). *Apparatus for decimal multiplication*. U.S. Patent 4 677 583.
- Goldstine, H., & Goldstine, A. (1996). The Electronic Numerical Integrator and Computer (ENIAC). *IEEE Annals of the History of Computing*, 18(1), 10–16. doi:10.1109/85.476557
- Gonzalez-Navarro, S., Tsen, C., & Schulte, M. J. (2013). Binary Integer Decimal-Based Floating-Point Multiplication. *IEEE Transactions on Computers*, 62(7), 1460–1466. doi:10.1109/TC.2012.79
- Guardia, M. (2012). Implementation of a fully pipelined BCD multiplier in FPGA. *VIII Southern Conference on Programmable Logic*, 1–6.
- Han, L., & Ko, S.-B. (2013). High-Speed Parallel Decimal Multiplication with Redundant Internal Encodings. *IEEE Transactions on Computers*, 62(5), 956–968. doi:10.1109/TC.2012.35
- Hickman, B., Krioukov, A., Erle, M., & Schulte, M. (2007). A Parallel IEEE P754 Decimal Floating-Point Multiplier. *25th Conference on Computer Design*, 296–303. doi:10.1109/ICCD.2007.4601916
- IBM. (n.d.). *IBM C and C++ Compilers*. Retrieved from <http://www.ibm.com/-software/awdtools/xlcpp>
- IEEE 754-2008. (2008). *IEEE Standard for Floating-Point Arithmetic*. IEEE Computer Society.
- IEEE Working Group of the Microprocessor Standards Subcommittee. (2008). *IEEE Standard for Floating-Point Arithmetic*. New York: The Institute of Electrical and Electronics Engineers.
- Intel. (n.d.). *Intel Decimal Floating-Point Math Library*. Retrieved from <http://softwarecommunity.intel.com/articles/eng/3687.htm>
- Jaberipur, G., & Kaivani, A. (2007). Binary-Coded Decimal Digit Multipliers. *IET Computers & Digital Techniques*, 1(4), 377–381. doi:10.1049/iet-cdt:20060160
- Jaberipur, G., & Kaivani, A. (2009). Improving the Speed of Parallel Decimal Multiplication. *IEEE Transactions on Computers*, 58(11), 1539–1552. doi:10.1109/TC.2009.110
- JTC 1/SC 22/WG 14. (2007). *TR 24732: Extensions for the Programming Language C to Support Decimal Floating Point Arithmetic*. Retrieved from <http://www.open-std.org/jtc1/sc22/wg14/www/docs/n1312.pdf>
- JTC 1/SC 22/WG 14. (2008). *TR 24733: C++ Decimal Floating Point Arithmetic Extensions*. Retrieved from <http://www.open-std.org/jtc1/sc22/wg21/docs/papers/2008/n2732.pdf>
- Juang, T.-B., & Chiu, Y.-M. (2013). High-speed binary to binary-coded-decimal converters for decimal multiplications. *International SoC Design Conference (ISOCC)*, 370–371. doi:10.1109/ISOCC.2013.6864053



- Kaivani, A., Chen, L., & Ko, S. (2012). High-frequency sequential decimal multipliers. *IEEE International Symposium on Circuits and Systems*, 3045-3048.
- Kenney, R., Schulte, M., & Erle, M. (2004). High frequency decimal multiplier. *IEEE International Conference on Computer Design: VLSI in Computers and Processors*, 26–29. doi:10.1109/ICCD.2004.1347893
- Knuth, D. (1986). The IBM 650: An Appreciation from the Field. *IEEE Annals of the History of Computing*, 8(1), 50–55. doi:10.1109/MAHC.1986.10010
- Larson, R. (1973, December). High Speed Multiply Using Four Input Carry Save Adder. *IBM Technical Disclosure Bulletin*, 2053-2054.
- Neto, H., & Véstias, M. (2008). Decimal multiplier on fpga using embedded binary multipliers. *Proceedings IEEE Field Programmable Logic and Applications*, 197–202. doi:10.1109/FPL.2008.4629931
- Neto, H., & Véstias, M. (2010). Parallel Decimal Multipliers using Binary Multipliers. *VI Southern Programmable Logic Conference*, 73-78.
- Quinn, K. (1983). Ever Had Problems Rounding Off Figures? This Stock Exchange Has. *Wall Street Journal*.
- Schwarz, E., Kapernick, J., & Cowlishaw, M. (2009). Decimal Floating-Point Support on the IBM System z10 Processor. *IBM Journal of Research and Development*, 53(1/2).
- Sun Microsystems. (n.d.). *BigDecimal Java Class*. Retrieved from <http://java.sun.com/j2se/1.5.0/docs/api/java/math/BigDecimal.html>
- Sutter, G., Todorovich, E., Bioul, G., Vazquez, M., & Deschamps, J.-P. (2009). FPGA implementations of BCD multipliers. *Proceedings IEEE International Conference on Reconfigurable Computing and FPGAs*, 36–41.
- Tsang, A., & Olschanowsky, M. (1991). *A Study of Database 2 Customer Queries*. IBM Technical Report 03.413. IBM.
- Tsen, S., Gonzalez-Navarro, S., Schulte, M., Hickmann, B., & Compton, K. (2009). A Combined Decimal and Binary Floating-Point Multiplier. *20th IEEE International Conference on Application-specific Systems, Architectures and Processors*, 8-15. doi:10.1109/ASAP.2009.28
- Ueda, T. (1995). *Decimal Multiplying Assembly and Multiply Module*. U.S. Patent. #5,379,245.
- Vazquez, A., Antelo, E., & Bruguera, J. D. (2014). Fast Radix-10 Multiplication Using Redundant BCD Codes. *Computers, IEEE Transactions on*, 63(8), 1902-1914. doi:10.1109/TC.2014.2315626
- Vázquez, A., Antelo, E., & Montuschi, P. (2010). Improved Design of High-Performance Parallel Decimal Multipliers. *IEEE Transactions on Computers*, 59(5), 679–693. doi:10.1109/TC.2009.167
- Véstias, M., & Neto, H. (2011). Iterative decimal multiplication using binary arithmetic. *VII Southern Conference on Programmable Logic*, 257-262.
- Véstias, M., & Neto, H. (2012). Parallel Decimal Multipliers and Squarers Using Karatsuba-Ofman's Algorithm. *15th Euromicro Conference on Digital System Design*, 782-788. doi:10.1109/DSD.2012.101
- Wang, L.-K., Tsen, C., Schulte, M., & Jhalani, D. (2007). Benchmarks and Performance Analysis for Decimal Floating-Point Applications. *25th International Conference on Computer Design*, 164-170. doi:10.1109/ICCD.2007.4601896
- Webb, C. (2008). IBM z10—The Next-Generation Mainframe Microprocessor. *IEEE Micro*, 28(2), 19–29. doi:10.1109/MM.2008.26
- Wilkes, M. V. (1997). Arithmetic on the EDSAC. *IEEE Annals of the History of Computing*, 19(1), 13–15. doi:10.1109/85.560726

Williams, M. (1993). The Origins, Uses, and Fate of the EDVAC. *IEEE Annals of the History of Computing*, 15(1), 22–38. doi:10.1109/85.194089

Zhu, M., Baker, A., & Jiang, Y. (2013). On a parallel decimal multiplier based on hybrid 8421–5421 BCD recoding. *IEEE 56th International Midwest Symposium on Circuits and Systems*, 1391–1394.

ADDITIONAL READING

Akkas, A., & Schulte, M. (2011) “A decimal floating-point fused multiply-add unit with a novel decimal leading-zero anticipator,” *IEEE International Conference on Application-Specific Systems, Architectures and Processors*, pp.43–50. doi:10.1109/ASAP.2011.6043235

Al-Khaleel, O., Al-Qudah, Z., Al-Khaleel, M., Papachristou, A., & Wolff, F. (2011) “Fast and compact binary-to-BCD conversion circuits for decimal multiplication,” *29th IEEE International Conference on Computer Design*, pp.226–231. doi:10.1109/ICCD.2011.6081401

Bhattacharya, J., Gupta, A., & Singh, A. (2010) “A high performance binary to BCD converter for decimal multiplication,” *International Symposium on VLSI Design Automation and Test*, pp.315–318. doi:10.1109/VDAT.2010.5496752

Cowlshaw, M. F. (2003) “Decimal floating-point: algorithm for computers,” *Proceedings 16th IEEE Symposium on Computer Arithmetic*, pp.104–111. doi:10.1109/ARITH.2003.1207666

Dadda, L., Pisoni, M., & Santambrogio, M. D. (2012) “A Parallel-Serial Decimal Multiplier Architecture,” *IEEE 15th International Conference on Computational Science and Engineering*, pp.310–317. doi:10.1109/ICCSE.2012.50

General Decimal Arithmetic. <http://speleotrove.com/decimal/>

Gonzalez-Navarro, S., Tsen, C., & Schulte, M. J. (2013). Binary Integer Decimal-Based Floating-Point Multiplication. *IEEE Transactions on Computers*, 62(7), 1460–1466. doi:10.1109/TC.2012.79

IEEE Standard for Floating-Point Arithmetic. (2008) “*IEEE Std 754-2008*”, pp.1–70.

Vázquez, A., & Dinechin, F. (2010) “Efficient implementation of parallel BCD multiplication in LUT-6 FPGAs,” *International Conference on Field-Programmable Technology*, pp.126–133. doi:10.1109/FPT.2010.5681767

KEY TERMS AND DEFINITIONS

Binary-Coded Decimal: A binary code of decimal numbers where each digit is represented with four bits.

Decimal Fixed-Point Multiplication: Refers to the multiplication operation where the inputs are represented in radix ten fixed-point numbers.

Decimal Floating-Point Multiplication: Refers to the multiplication operation where the inputs are represented in radix ten real numbers.

Decimal Multiplication: Refers to the multiplication operation where the operands are represented with some decimal digit encoding.

Densely Packed Decimal: An encoding scheme in which three BCD digits are represented with ten bits.

Iterative Decimal Multiplication: A decimal multiplication method where partial products are generated and accumulated step-by-step in an iterative process.

Parallel Decimal Multiplication: A decimal multiplication method where partial products are generated and added according to a parallel process.

Digital Divide



Patrick Flanagan

St. John's University, USA

INTRODUCTION

Since 1991 when the World Wide Web (WWW) was first made available to the public, the WWW has revolutionized the ways the global community engages each other economically, politically, and socially. Its impact has been historically unprecedented. While the availability of and access to the WWW appears to be ubiquitous, it is not. The expansion of this marvelous Information Communication Technology (ICT) has not penetrated certain areas of the world resulting in a “digital divide.” This chapter discusses this digital divide. It first defines the term and how scholars have understood the digital divide. It then moves to discuss the origins of the term in popular literature and official government documents. From there, the chapter moves to present concrete evidence of how the digital divide has negatively impacted the global community. Finally, it names and evaluates the efforts of different organizations and agencies to resolve the digital divide. It concludes with a prospectus on the future challenges of information communication technology vis-à-vis the digital divide.

BACKGROUND

The digital divide is a term that describes the gap between those who have access to information communication technology (ICT) and those who have limited or no access. This distinction, however, between the “haves” and the “have nots” can be too basic a delineation (Compaine, 2001; Hawkins, 2006; Selwyn, 2004; Warschauer, 2002). What is “had” and “not had” is much more compre-

hensive involving available physical equipment, utility resources (for instance, electricity), and technological skills. While the “have nots” can be those who do not have effective access to information communication technology, the “haves” can include those who have a computer, but with no or limited connection to the Internet, with a rather dated dialup and not a broadband connection, or those who connect through a mobile phone. ICT has transformed significantly political, social, and economic engagement in connected parts of the global village. Without effective widespread access to ICT, the digital divide further alienates citizens within and among countries of the world and amplifies divides already established ethnic, gender, income, and geographic inequalities. Both government agencies and scholars have studied carefully the digital divide and have suggested creative ways to ensure access to equipment, education, and viable signal connections in order to maximize fuller participation in this dynamic global ICT phenomenon.

A review of literature early on in the rollout of the WWW reveals attentiveness to more than just lack of access to the rich technological resources some enjoyed. In their assessment of the digital divide, scholars highlight that the chasm is much more complex than its original sense involving widespread inequalities on various political, economic, educational, demographic, ability, and gender levels (Alampay, 2006; Barzilai-Nahon, 2006; Colle and Roman, 2001; Dagron, 2001; DiMaggio, Hargittai, and C & S, 2004; Fink and Kenny, 2003; Norris, 2001; Parkinson, 2005; Potter, 2006; Simpson et al, 2004; and Warschauer, 2003). While admitting, for example, the excitement of the Internet’s impact for optimizing

networking in the global village, Norris (2001) raised some critical questions as to whether or not the Internet would evolve into a democratic participatory medium offering equal advantages for engagement or would it only reinforce dominance and inequality. Beyond a binary construction of the digital divide rendering it more complex, Norris describes three divides that called for a response: the global divide that focused on access; the social divide that alienated people; and, the democratic divide that illustrated the use or lack of use of the Internet for political purposes. Van Dijk and Hacker (2003) identify psychological, material, skill, and usage factors that influence this access. Hilbert (2004) focuses on the gender divide while Preiger and Hu (2006) study the racial divide, both further specifications of the digital divide. Kularski and Moller (2012) further specify the digital divide focusing on technological skill gap. The challenge, Kularski and Moller note, involves more than supplying ICT equipment and ensuring access points to digitally excluded people. Users need to be trained how to use technology optimally for their needs.

Castells (2001) highlights the strong relationships between the different geographies of the Internet: technological geography, geography of users and economic geography. Castells affirms that Internet is the technological tool able to distribute the informational power, knowledge and capability to connect people into different networks. To be disconnected means to be marginalized in the global system. For this reason the sentence about the need for the underdeveloped countries to start from the real needs of the third world (health, culture, water and electricity), before thinking to Internet, reveals a deep misunderstanding. Without Internet no Country has the possibility to generate the resources able to satisfy the needs linked to the development. Following Castells (1996) Internet has the control and accessibility to informative flows, although not homogeneously dislocated in the world, configuring new geo-political balances on the basis of new geo-informational maps of the

“Internet Galaxy”. It is a fact, that the relationship between science and both society and media, is profoundly changed. In the last two decades, we have been observing a structural transformation of traditional communicational channels, where tele-communication was used to connect people physically separated from each other, toward a new pattern of “connected presence” (Castells, Fernandez-Ardevol, Qiu & Sey, 2007). In this new model, small gestures or signs of attention may be at least as important as the message content itself.

MAPPING THE DIGITAL DIVIDE

46% of the world population is connected to the Internet, a far cry from 1% in 1995 (Internet Live Stats, 2016). Statistics report considerable growth in digital usage for the past sixteen years (Internet World Stats, 2016). However, evidence also amplifies that the African, Asian, and Middle Eastern regions still lag in Internet connections significantly impacting the larger global digital divide. 2.1 billion of the 3.3 billion Internet users lived in 10 countries (Internet Live Stats). The rest comprised the bottom quarter and among 191 other countries of the world. Heading the list of countries were China, the United States, India, Japan, and Brazil. The most connected countries are Iceland, Faeroe Islands, Norway, Bermuda, Andorra, Denmark, and Liechtenstein all averaging approximately 96% Internet penetration per population rates. The least connected countries of the world are, in ascending order, Eritrea, Timor-Leste, Burundi, Somalia, Guinea, and Niger, all with no more than 2% Internet penetration rates.

A number of different efforts have been made by scholars and journalists to illustrate visually the digital divide. Chris Harrison from the Human Interactions Institute at Carnegie Mellon University tracked Internet access over a four year period. In a 2011 map, Harrison showed the increased user connections throughout the world and, at the same time, demonstrated Internet ac-

cess remained unattainable for areas of the world where access could be transformative for their human condition.

In 2011, Gregor Aisch, a graphics editor at the New York Times, used IP addresses taken from GeoLiteCity database by MaxMind to visualize the global digital divide. He did qualify his findings by noting that his employment of IP addresses did not recognize the widespread use of mobile devices in Africa. Harrison and Aisch capture well in their dark spots of the world the locations and extent of the digital divide.

The digital divide also has garnered attention from leadership in the public and private sector given that the global community has become quickly enriched with technology. As a world increasingly dependent on technology for its rich resources, consistent efforts have been made to analyze and make provisions to resolve the digital divide. Any technological deficit only furthers digital inequalities making it difficult for the least connected countries to compete, participate, resource, and communicate the information superhighway. The World Summit on the Information Society declared that the global challenge for the new millennium is to build a society “where everyone can create, access, utilize and share information and knowledge, enabling individuals, communities and peoples to achieve their full potential in promoting their sustainable development and improving their quality of life” (World Summit).

Since 1995, the United States Department of Commerce’s National Telecommunications and Information Administration’s (NTIA) has studied the issue of ICT diffusion publishing regularly its findings and strategies to resolve the unequal distribution (1998, 1999, 2000, 2002, 2004, 2007, 2009, 2010, 2011, 2013, and 2014). Its 1995 research report “Falling Through the Net: A Survey of the ‘Have Nots’ in Urban and Rural America” is often credited with coining this term “digital divide,” the NTIA report acknowledged the revolution that had begun impacting many Americans positively. Although, the report’s

findings offered a broad perspective on the digital divide, the term “digital divide” itself can be found nowhere in the government document. Earlier in 1994, Steve Case, the then-AOL Chairman and CEO, identified the digital divide in his remarks: “There’s no single solution to bridging the digital divide... We must take steps now so that in the Internet Century, no children are left behind” (Norris, 2000). A more descriptive appreciation of the phrase “digital divide,” however, is known to have more popular origins in print media. In January 1996, New York Times journalist Gary Andrew Poole reported on the digital divide in education between two schools in the center of the technologically rich Silicon Valley (Poole, 1996). One school had access to the latest computers while another proximate one had outdated ones offsetting the array of opportunities the former enjoyed. Later that same year, Los Angeles Times writer Amy Harmon used the term “digital divide” to describe the social distance she felt as a wife felt towards her husband preoccupied with surfing the Web (Harmon, 1996). Despite its novel origin as a social construction in describing a marital exchange, Harmon offered initial insight into how an individual, particularly a woman in this case, and in the overall historical schema of the global digital divide, could be estranged from the obvious engaging resources of the Web. Interest in the digital divide was raised to national prominence in October 1996. In an address to the people of Knoxville, Tennessee, then-President Bill Clinton and then-Vice President Al Gore used the term “digital divide” twice to amplify their concern for those who would be left out of the American dream (Wilson, 2009).

In July 1998, the NTIA published “Falling Through the Net: A Survey of the ‘Haves’ and ‘Have Nots’ in Rural and Urban America” with “Falling Through the Net II: New Data on the Digital Divide.” In July 1999, the NTIA released its third publication “Falling Through the Net: Defining the Digital Divide.” The NTIA published “Falling Through the Net: Toward Digital Inclusion” in 2000 and “A Nation Online: Internet Use



in America” in 2002. Broadband was the focus of the NTIA’s 2004 report “A Nation Online: Entering the Broadband Age.” In 2007, the NTIA published “Networked Nation: Broadband in America” and continued its discussion of broadband service with its 2009 publication of “Digital Nation: 21st Century America’s Progress Towards Universal Broadband Internet Access.” November 2010 saw the release of its document “Exploring the Digital Nation: Home Broadband Internet Adoption in the United States” and in 2011 “Digital Nation - Expanding Internet Usage.” The NTIA’s 2013 publication “Exploring the Digital Nation: America’s Emerging Online Experience” tackled the digital divide by identifying first what Americans do online and its 2014 report “Exploring the Digital Nation: Embracing the Mobile Internet” illustrated the move from personal computers to mobile devices.

In addition to the regular thorough analysis that the United States NTIA has done to highlight the disconnectedness of so many of its citizens since the diffusion of technology to the public, the European Union (EU) also has sought to resolve the challenge of the digital divide. In 2010, the EU published “Europe 2020 Strategy” which has sought “smart, sustainable and inclusive growth for European Economy” so as “to exit the crisis and prepare the EU economic for the challenges of the next decade” (European Commission, 2015). Member nations acknowledged that “in a changing world, we want the EU to become a smart, sustainable and inclusive economy. These three mutually reinforcing priorities should help the EU and the Member States deliver high levels of employment, productivity and social cohesion.”

A “Digital Agenda for Europe” is one Europe 2020 Strategy’s priorities, a “flagship initiative.” The EU recognized that digital engagement through ICTs is critical to growth and sustainability. It could not be a mere bystander as technology transforms the global community. The European Commission earmarked 1 billion Euros to extend digital inclusion to rural areas in the hopes of generating significant economic growth. All 27 member states of the European

Union were charged with evaluating the digital divide specific to their own country and working in concert with other nations to resolve it. Unlike the United States’ NTIA’s regular updated published reports EU created a website <http://ec.europa.eu/digital-agenda/> tailored to offer citizens and officials access to analytics and monitor progress in each EU country.

The digital divide also has been a central concern on the international level. In 2000, the G8 formed the Digital Opportunity Task Force. This task force was a collaborative effort among governments, private sector enterprises, and not for profit organizations to brainstorm strategies as to how to resolve the digital divide and deliver technology to marginalized parts of the world (Hart, 2001). In 2001, the United Nations Information and Communication Technologies Task Force took up the same task and continues its mission to today. This task force disseminates information about the digital divide, builds alliances with corporations and nonprofits, and works closely with people from less connected countries in both connecting them to the Internet and sustaining them in their technological development.

In 2012, UN Secretary General Ban-Ki Moon called for renewed efforts in bridging the digital divide, noting that “All people must be able to make the best use of information and communications technology to help create the future we want” (United Nations News Centre, 2012). Moon’s words build upon the United Nations’ Human Rights Council unanimous resolution to identify internet access and online freedom of expression as a basic human right (Estes, 2011). Yet, this goal remains elusive given the fact that, according to the United Nations’ International Telecommunication Union’s own data, while around one third of the global population was online, almost two thirds of the population was not.

BRIDGING THE DIGITAL DIVIDE

Evident progress has been made to get people online. A number of creative contemporary solu-

tions that have sought to close the digital divide in the global village. This section will highlight only three of the many creative endeavors initiated by strong partnerships between less connected countries, businesses, and human interest organizations.

The One Laptop per Child (OLPC) initiative was founded by Nicholas Negroponte in 2005 in a joint partnership with leading technology corporations. OLPC distributes those familiar white computers and more recently tablets with green antennae. The attraction is its sticker price which at one point in its history was 100 USD. While the project is not without criticism from those suggesting other human matters should be prioritized, evidence indicates that OLPC has reduced digital inequality and promoted education, health, and welfare (Cotten, Hale, Moroney, O'Neal, and Borch, 2011; Selwyn, 2013; and Warschauer, 2012).

The British Council and Microsoft have partnered to upgrade technology in Kenya by creating a “Badiliko,” in kiSwahili, a “digital hub,” for students to gain full and free access to the Internet in over 100 schools. This collaboration also involves the training of teachers. It is a small effort to provide students with online learning resources and connect them with peers and teachers throughout the world considering the magnificent number of students enrolled in education (BBC News, 2013; Gbolagade Adekanmbi and Bopelo Boitshwarelo, 2012).

Another program that seeks to bridge the digital divide is the Close the Gap organization, based out of Brussels, Belgium. In its mission statement, Close the Gap acknowledges that it seeks to “bridge the digital divide by offering high-quality, pre-owned computers donated by European companies to educational, medical and social projects in developing and emerging countries. Unfortunately, laudatory efforts such as these are not always successful in bridging the digital divide for a variety of reasons, including “outdated and underfunded programs, lack of comprehensive solutions, uninspired models of technology in education, and non-holistic ap-

proaches” (Evans, 2010). A history of debt in the global South, too, has precluded unconnected nations from having the infrastructure and capital to sustain hopeful endeavors only pushing the further into a downward spiral of further fiscal distress, disease, and illiteracy. Admittedly, wealthier nations with a surplus of equipment, generally dated and sometimes inoperable, donate technology equipment to poor nations. However though, historically there tends to be little to no follow-up educational tutoring for the digital illiterate; no sustainable strategies to access the Internet given infrastructure and operational barriers; and, no ongoing support for computer repair.

IMPETUS FOR RESOLVING THE DIGITAL DIVIDE

Technology empowers people. The last decade has demonstrated this. In the Occupy Movement in Northern Africa, technology was the key medium to offsetting a history of injustices and tyrannies. The national leadership in Egypt was so threatened by what could be accomplished by technological communication that it shut down access to the Internet for five days. China has done similarly. Access to the Internet “created a world of carnival, community, and contention in and through cyberspace and how in this process they have transformed personhood, society, and politics” as “a response to the grievances, injustices, and anxieties caused by the structural transformation of Chinese society” (Yang, 2009).

Being connected to the Internet has the capacity to revolutionize peoples through access to informational resources and communication platforms, but technology must be accessible in order to harness its power. Unfortunately, for much of the world, particularly those in the southern regions of the globe, the prospect of having a decent computer equipment and a sustainable infrastructure to support broadband access seems as realizable as relief from their persistent impoverishment, hunger, and illiteracy.



Technology is not the panacea to the panorama of problems plaguing the world. Closing the digital divide and affording people Internet access could lead to a lessening of the world's deleterious problems through the Internet's capabilities. Poverty, oppression, hunger, human trafficking, clean water, violence and war, to name some of the more pressing global concerns, may not be able to be fully resolved through digital connectivity, but could be ameliorated by using the Internet as a *catalyst*. Although, not all tech leaders agree with my claim that bridging the digital divide would alleviate many of the problems plaguing developing nations. Bill Gates reminds the global community that technology alone will not solve basic problems inherent to developing nations (Stone, 2013). Yet, access to technology is increasingly becoming an important component to alleviating poverty and increasing opportunities for marginalized populations across the world. However, narrowing the digital divide has the capacity to inform and connect people to previously unavailable resources. Expanding online access levels can level the playing field in the global economy and offer a constellation of advantages for connected individuals.

Technology extends the range of possible outcomes and without an Internet connection, unequal opportunities for economic, political, and social growth persist. The digital divide eclipses the opportunities to access the wealth of resources available on the Web for individuals, communities, and nations for education, collaboration, and significant development.

FUTURE RESEARCH DIRECTIONS

Resolving the digital divide requires a multi-dimensioned response and will continue to entail a great deal of strategic analysis and priority planning. Further research to study carefully many of the critical concerns raised in this chapter will be important to sustain momentum to close

the digital gap. While the penetration of mobile devices throughout the global village has been hailed as a significant advancement in bridging the digital divide, mobile telephony is only a step. A person, using the mobile phone for Internet access in its present form, cannot engage fully the wider resources on the Web; a larger Internet connected computer or tablet is ideal. In addition to securing physical provisions for technologically handicapped nations, access to reliable and economically reasonable Internet signals also are as important.

Another area of future research must involve ongoing training and development. Internet connected computers cannot merely be given to digitally challenged nations with the expectation that they will be able to maximize the use of technology. While the technological learning curve can be thin in younger populations who are more willing to take risks and learn through experimenting with technology, strategic efforts to ensure training and subsequent updating of the wide range of computer users will be critical. As advancements in computer technology become available, it also would be important that work be done regularly to supply the more marginalized and vulnerable communities with contemporary machinery, stronger Internet signals, and updated software. In the latter case, the Open Source Movement offers marginalized peoples opportunities to use common software programs and, at the same time, participate in tweaking the software (Bergquist, Ljungberg, & Rolandsson, 2011).

A final area that researchers need to study more deeply focuses on the motivations on leaders and nations to restrict access to the ICT. It is evident that the plethora of information and connections offered through ICT can maximize global engagement and industrious innovation. At the same time, there are some leaders of nation states that are threatened by the resultant thoughts and actions after interacting with the massive capabilities the Internet offers. Research into how cultures have advanced through their engagement with ICT and

how Luddite countries might move beyond its technological paralysis would be helpful to both individual citizens and the global community.

CONCLUSION

The revolutionary reach of ICT has extended to many parts of the world offering different global communities significant opportunities for growth in economic, educational, political, and social areas. Unfortunately, a goodly number of global citizens do not enjoy having access to technological equipment and being connected to the powerful resources available through this celebrated historical advancement. This reality commonly known as the digital divide further heightens socio-economic disparities by handicapping the potential to connect for information and communication. In a rather unfortunate unfolding of events, the Internet which was intended to connect people and build livelihoods, in some instances, only has isolated people further and deepened already existing inequalities. This chapter examined the digital divide using insights and data from academic and government resources. Both scholars and officials acknowledge that digital divide is closing through specific efforts of organizations and governments. At the same time, both of these constituencies admit much more work has to be done to bridge the digital gap, particularly with only 40% of the world's population connected. An integrated response must involve access, use, skill, language, background and environment. Accessibility to connected technology leads to greater inclusion, resolving digital inequalities often reflective of deeper economic, political, social, educational, and demographic inequities.

REFERENCES

- Adekanmbi, G., & Boitshwarelo, B. (2012). Collaboration in Distance Education in Sub-Saharan Africa: Trends, Trials and. *Cases on Cultural Implications and Considerations in Online Learning*, 375.
- Aisch, G. (2011). *Global digital divide*. Retrieved April 30, 2016, from <http://old.driven-by-data.net/about/global-digital-divide/#/0>
- Alampay, E. (2006). Beyond access to ICTs: Measuring capabilities in the information society. *International Journal of Education and Development Using ICT*, 2(3).
- Barzilai-Nahon, K. (2006). Gaps and bits: Conceptualizing measurements for digital divide/s. *The Information Society*, 22(5), 269–278. doi:10.1080/01972240600903953
- BBC News. (2013). *Kenya IT Hubs Launched for Primary Schools*. Retrieved April 30, 2016, from <http://www.bbc.co.uk/news/world-africa-24147333>
- Bergquist, M., Ljungberg, J., & Rolandsson, B. (2011). A Historical Account of the Value of Free and Open Source Software: From Software Commune to Commercial Commons. In *Open Source Systems: Grounding Research* (pp. 196–207). Springer Berlin Heidelberg.
- Castells, M. (1996). The Rise of the Network Society: The Information Age: Economy. *Society and Culture*, I.
- Castells, M. (2001). *The Internet Galaxy: Reflections on the Internet, Business, and Society*. Oxford University Press. doi:10.1007/978-3-322-89613-1
- Castells, M., Fernandez-Ardevol, M., Qiu, J., & Sey, A. (n.d.). *Mobile Communication and Society: A Global Perspective (Information Revolution and Global Politics)*. MIT Press.



- CBC News. (2011). Internet 'kill switch' easy target in Egypt. *CBC News*. Retrieved from <http://www.cbc.ca/news/technology/internet-kill-switch-easy-target-in-egypt-1.1110730>
- Colle, R. D., & Roman, R. (2001). Editorial: The Telecenter Environment in 2002. *The Journal of Development Communication*, 12(2), 15.
- Compaine, B. M. (2001). *The digital divide: Facing a crisis or creating a myth?* MIT Press.
- Cotten, S. R., Hale, T. M., Moroney, M. H., O'Neal, L., & Borch, C. (2011). Using affordable technology to decrease digital inequality: Results from Birmingham's one laptop per child XO laptop project. *Information Communication and Society*, 14(4), 424–444. doi:10.1080/1369118X.2011.559266
- Dagron, A. G. (2001). Prometheus riding a Cadillac? Telecentres as the promised flame of knowledge. *The Journal of Development Communication*, 12(2), 803–824.
- DiMaggio, P., Hargittai, E., Celeste, C., & Shafer, S. (2004). Digital inequality. *Social Inequality: From Unequal Access to Differentiated Use*, 355–400.
- Estes, A. (2011). The U.N. declares Internet access a human right. *National Journal*. Retrieved April 30, 2016, from <http://www.nationaljournal.com/dailyfray/the-u-n-declares-internet-access-a-human-right-20110606>
- European Commission. (2015). *Digital agenda for Europe: A Europe 2020 initiative*. Retrieved April 30, 2016, from <http://ec.europa.eu/digital-agenda/>
- Evans, C. (2012). Five reasons we have yet to close the digital divide. *Microsoft in Education Blog*. Retrieved April 30, 2016, from, http://blogs.technet.com/b/microsoft_in_education/archive/2012/02/02/five-reasons-we-have-yet-to-close-the-digital-divide.aspx
- Fink, C., & Kenny, C. J. (2003). Whether the digital divide?. *Info*, 5(6), 15–24.
- Harmon, A. (1996). Daily Life's Digital Divide. *Los Angeles Times*, p. A1. Retrieved April 30, 2016, from http://articles.latimes.com/1996-07-03/news/mn-20785_1_digital-technology
- Harrison, C. (n.d.). Global scale: City-to-City Connections with the Internet. *The geographic advantage: an AAG companion website for understanding the changing planet*. Retrieved April 30, 2016, from <http://geographicadvantage.aag.org/index.cfm?action=main.module§ionID=3&moduleID=8&pageID=29>
- Hart, J. (2002). *The digital opportunities task force: The G8's effort to bridge the global digital divide*. Retrieved April 30, 2016, from <http://www.g7.utoronto.ca/conferences/2004/indiana/papers2004/hart.pdf>
- Hawkins, B. L., & Oblinger, D. G. (2006). The Myth about the Digital Divide. *EDUCAUSE Review*, 41(4), 12–13.
- Hilbert, M. (2011, December). Digital gender divide or technologically empowered women in developing countries? A typical case of lies, damned lies, and statistics. *Womens Studies International Forum*, 34(6), 479–489. doi:10.1016/j.wsif.2011.07.001
- Internet Live Stats. (2016). Retrieved April 30, 2016, from <http://www.internetlivestats.com/internet-users/>
- Kularski, C., & Moller, S. (2012). The digital divide as a continuation of traditional systems of inequality. *Sociology*, 51(1), 1–23.
- Maxmind. (2015). *GeoLite legacy downloadable databases*. Retrieved April 30, 2016, from <http://dev.maxmind.com/geoip/legacy/geolite/>
- National Telecommunications and Information Administration. (1995). *Falling through the net: a survey of the 'have-nots' in rural and urban America*. Retrieved April 30, 2016, from <http://www.ntia.doc.gov/ntiahome/fallingthru.html>

National Telecommunications and Information Administration. (1998). *Falling through the net II: new data on the digital divide*. Retrieved April 30, 2016, from <http://www.ntia.doc.gov/ntiahome/net2/falling.html>

National Telecommunications and Information Administration. (1999). *Falling through the net: defining the digital divide*. Retrieved April 30, 2016, from www.ntia.doc.gov/ntiahome/fttn99

National Telecommunications and Information Administration. (2000). *Falling through the net: toward digital inclusion*. Retrieved April 30, 2016, from <http://www.ntia.doc.gov/ntiahome/fttn00/contents00.html>

National Telecommunications and Information Administration. (2002). *A nation online: Internet use in America*. Retrieved April 30, 2016, from <http://www.ntia.doc.gov/report/2002/nation-online-internet-use-america>

National Telecommunications and Information Administration. (2004). *A nation online: entering the broadband age*. Retrieved April 30, 2016, from <http://www.ntia.doc.gov/report/2004/nation-online-entering-broadband-age>

National Telecommunications and Information Administration. (2007). *Networked Nation: Broadband in America*. Retrieved April 30, 2016, from <http://www.ntia.doc.gov/legacy/reports/2008/NetworkedNation.html>

National Telecommunications and Information Administration. (2010). *Exploring the digital nation: home broadband Internet adoption in the United States*. Retrieved April 30, 2016, from <http://www.ntia.doc.gov/report/2010/exploring-digital-nation-home-broadband-internet-adoption-united-states>

National Telecommunications and Information Administration. (2011). *Digital Nation - Expanding Internet Usage*. Retrieved April 30, 2016, from <http://www.ntia.doc.gov/report/2011/digital-nation-expanding-internet-usage-ntia-research-preview>

National Telecommunications and Information Administration. (2013). *Exploring the digital nation: America's emerging online experience*. Retrieved April 30, 2016, from <http://www.ntia.doc.gov/report/2013/exploring-digital-nation-americas-emerging-online-experience>

National Telecommunications and Information Administration. (2014). *Exploring the digital nation: embracing the mobile Internet*. Retrieved April 30, 2016, from <http://www.ntia.doc.gov/report/2014/exploring-digital-nation-embracing-mobile-internet>

Norris, P. (2000). The worldwide digital divide. In *Paper for the Annual Meeting of the Political Studies Association of the UK*. London School of Economics and Political Science.

Norris, P. (2001). *Digital divide: Civic engagement, information poverty, and the Internet worldwide*. Cambridge University Press. doi:10.1017/CBO9781139164887

Oxford English Dictionary Online. (2004). New York: Oxford University Press. Retrieved April 30, 2016, from <http://dictionary.oed.com/>

Parkinson, S. (2005). *Telecentres, access and development: experience and lessons from Uganda and South Africa*. IDRC. doi:10.3362/9781780441429

Poole, G. A. (1996, January 29). A new gulf in American education, the digital divide. *New York Times*, p. 2.

Potter, A. B. (2006). Zones of silence: A framework beyond the digital divide. *First Monday*, 11(5). doi:10.5210/fm.v11i5.1327

Prieger, J., & Hu, W. (2006). *An empirical analysis of indirect network effects in the home video game market*. Retrieved April 30, 2016, from: <http://ssrn.com/abstract=941223>

Selwyn, N. (2004). Reconsidering political and popular understandings of the digital divide. *New Media & Society*, 6(3), 341–362. doi:10.1177/1461444804042519



Selwyn, N. (2013). Empowering the World's Poorest Children? A Critical Examination of One Laptop per Child. *The Politics of Education and Technology: Conflicts, Controversies, and Connections*, 101.

Simpson, L., Daws, L., & Pini, B. (2004). Public internet access revisited. *Telecommunications Policy*, 28(3), 323–337. doi:10.1016/j.telpol.2003.10.001

Stats, I. W. Usage and Population Statistics. (2016). Retrieved April 30, 2016, from <http://www.internetworldstats.com/stats.htm>

Stone, B. (2013). Bill Gates on his foundation's health and education campaigns. *Bloomberg Businessweek*. Retrieved April 30, 2016, from <http://www.businessweek.com/articles/2013-08-08/bill-gates-on-his-foundations-health-and-education-campaigns#p2>

United Nations News Centre. (2012). *Ban stresses power of information technology to facilitate sustainable development*. Retrieved April 30, 2016, from http://www.un.org/apps/news/story.asp?NewsID=40869#.VXiLXc_BzRY

United States Department of Commerce's National Telecommunications and Information Administration. (n.d.). *Internet and Computer Use Studies and Data Files*. Retrieved April 30, 2016 from <http://www.ntia.doc.gov/data>

Van Dijk, J., & Hacker, K. (2003). The digital divide as a complex and dynamic phenomenon. *The Information Society*, 19(4), 315–326. doi:10.1080/01972240309487

Warschauer, M. (2002). Reconceptualizing the digital divide. *First Monday*, 7(7). doi:10.5210/fm.v7i7.967

Warschauer, M. (2004). *Technology and social inclusion: Rethinking the digital divide*. MIT Press.

Warschauer, M. (2012). The digital divide and social inclusion. *American Quarterly*, 6(2), 131.

Wilson, J., & Wilson, H. (2009). Digital Divide: Impediment to ICT and Peace Building in Developing Countries. *American Communication Journal*, 11(2), 1–9.

World Summit on the Information Society. (2003). *Declaration of principles*. Retrieved April 30, 2016, from http://portal.unesco.org/ci/en/files/25347/11895190701DPGeneva_en.pdf/DPGeneva_en.pdf

Yang, G. (2013). *The power of the Internet in China: Citizen activism online*. Columbia University Press.

KEY TERMS AND DEFINITIONS

Access: The physical proximity to computer technology, electricity, and an Internet signal.

Affordability: The financial means to access or purchase computer equipment, electricity and an Internet connection.

Broadband: Wide bandwidth data transmission using a fixed cable or fiber optics or combination, or through wireless satellite services, to conduct multiple signals.

Connectivity: The vibrancy and speed of an Internet connection often dependent on the Internet service provider (ISP) used.

Digital Divide(s): A multi-dimensional reality where only 40% of the world's population has access to the Internet.

Internet: A computer network that can be accessed through a dial-up modem, digital or cable network like, satellite, or broadband connection.

Marginalization: The relegation or neglect of a specific group of people to the fringes of society.

World Wide Web: The user-friendly platform on which users interact with others and resource information on the Internet.

Eight Tips for the Theme, “Data and Forecasts”



Alessio Drivet

GeoGebra Institute of Torino, Italy

INTRODUCTION

Paul was a common octopus living (January 2008, October 2010) in a public aquarium at the marine life center in Oberhausen, Germany. Paul experienced some international notoriety during the 2010 World Cup when it was used to “predict” the results of football matches in which the German national football team was involved and the final (that was not played by Germany). Paul’s predictions were all correct.

The octopus belongs to Octopodidae family and has eight tentacles, therefore we imagine that each of these gives us an indication as to which are the best strategies to analyze resources and obstacles of teaching probability and statistics.

Mathematical artifacts shown in the figures are the basis of a project that could be called educational gaming. The collection of materials found, purchased, and constructed is the result of many years of research work (Drivet, 2013). They are taken from <https://sites.google.com/site/oggettimatematici/home> site. Currently there are 210 objects and, of these, 40 relate to the mentioned subject (Figure 1).

BACKGROUND

Numbers, geometry, relations and functions, data and forecasts are the four themes that characterize the classic division that has gradually made its way (albeit with some linguistic difference) within the Programme for International Student Assessment (PISA), the TIMSS (Trends in International Mathematics and Science Study) and,

with regard to Italy, the UMI (Italian Mathematical Union), the INVALSI (National institute for the evaluation of the education and training system) and the National Guidelines for the curriculum.

This text will tackle issues related to data and forecast, only partially similar to the traditional concepts of probability and statistics.

A first starting point is provided by this quote: “Probability is the theory which organizes the world of chance phenomena. One has to start with phenomena which beg to be organized: intuition about unpredictable events, games of chance, occurrences which seem to happen without regularity etc. One then has to teach the learner to constitute the mental object for himself/herself. This is the stage preceding concept attainment but too often ignored” (Kapadia & Borovcnik, 2012).

As a matter of fact, there is a difference between the theory of probability and the theme of the data and forecasts. The task of data and forecasts means analyzing and interpreting data, developing conclusions and reasoning also with the aid of graphic representations, consciously using the calculation tools and the potential offered by IT applications.

EIGHT TIPS

First Idea: The Different Registers

The following quote makes it perfectly clear the key concept.

“For example, if you throw an ideal dice cube and one wonders what is the probability of getting either 1 or 6, it can be answered in different ways,

DOI: 10.4018/978-1-5225-2255-3.ch402



sets (such as lists of patients' pulse rates or baseball players' batting averages)"(Pfenning, 1998).

Figure 2. Dice



Table 1. Register and Representations

Registers	Representations
Linguistic	2 of 6
Fractional	1/3
Decimal	0.33
Percentage	33.3%

Second Idea: The Role of Combinatorics

One question: the introduction of combinatory calculus in school teaching has a mathematical autonomous value or is it a simply tool in order to teach probability?

The discussion would take us away; here we examine only the utilitarian aspect underlining the fundamental principle of the calculation, namely the principle of multiplication.

A typical example is the slot machine and the definition of the number of possible winning combinations. The ideal would be to have a concrete object (Figure 3).

The goal is to be able to achieve a winning combination between the figures that appear on the reels. To play, simply pull the arm that controls the reels of slot machines, wait until they stop, and see if the output is a winning combination or not.

The odds of winning vary according to the type of slot.

Taking a cue from the slot machine in the picture we can see that there are three reels and five symbols: BAR, cherries, apple, 7, \$.

Lowering the arm, the wheels move independently to form an arrangement among the 5^3 possible.

You win if you get a triple BAR or a triple 7, or two BAR and a 7 or two 7 and a BAR.

In total you have eight winning combinations.

The probability of winning is therefore equal to $8/125 = 0.064$.

In other words: if you play 1 \$, in case of victory, you should get $0,064 \cdot S = 1$, i.e. 15.63 \$.

In practice, the pay-out will be much lower, and “You will play at a loss.” You have to know

Figure 3. Slot machine



that in Italy for every euro introduced in slot machines a portion of that money is transferred to the Treasury and to the managers. Tobacco shops or coffee shops machines distribute only 75 percent of the collected amount!

The basic idea (in teaching) is that the items should be introduced, first, by an intuitive point of view, using a variety of examples and, only later, by mathematical formalization.

Third Idea: Coins and Dice in Various Probabilistic Approaches

All the textbooks use examples referring to coins, dice, cards, marbles, etc. The reason is known: you can build simple examples that will (should?) leave room for more complex random situations.

This approach can be improved by expanding the range of “coins” and “dice” used to make a transition logically motivated by the classical definition to the frequentist probability. As examples



(Drivet, 2013), we cite thumbtacks (Figure 4) or astragals (Figure 5).

The thumbtacks have two possible ways to fall and therefore could be treated like a coin with two faces. A thumbtack can drop with the tip downwards or upwards, but in this case you cannot apply the probability in the classical sense, because the two events do not have the same probability just because of the geometrical structure of the object.

In order to provide an assessment of probability (for example the exit of the tip up which we will call Head) the thumbtack should be launched a sufficiently high number of times in order to get a valuation given by the ratio between favorable outcomes and attempts.

Since there are different types of thumbtacks you can direct students to get them to justify an argumentation on the probability of the event “head”.

Figure 4. Thumbtacks



Figure 5. Astragal



Astragal (Talus or Anklebone) is one of the group of foot bones known as the tarsus. It has particularly regular proportions and, since ancient times, taken from goat and mutton, it was like a sort of with 4-sided dice. Of course it is not a fair dice.

Also in this case the objective is to compare the results of a regular dice with those of the talus.

A small digression: Astragal was used in games of chance, but also to predict the future. In Iceland it was (and it is) used to have an answer to a question. First it is read the following magic sentences that, in English, more or less sounds like this: “Tell me, my Vala, the answer to my question. I shall gladden you with gold, and satisfy you with silver if you tell me the truth. But I shall drown you in urine and burn you in fire if you lie to me”.

Then astragal is launched and if the bone falls with the convex face upwards the answer will be positive, if it shows the concave surface the answer is no, otherwise there is no certain answer.

Fourth Idea: Information Technology

Computer software offers the opportunity for students to learn about modelling. It enables students to build their own models in order to describe data and to generate simulations that can be explored.

“Computer-based assessment gives students the opportunity to work with larger data sets and provides the computational power and data handling capacities they need to work with longer

available sets. Students are given the opportunity to choose appropriate tools to manipulate, analyze, and represent data, and to sample from data populations. “(PISA, 2012).

In class, we work too often with invented data. Often the presented data are not real and not even realistic. Sometimes they fit so close to the model then they seem completely artificial. (Gattuso, 2011).

It will be up to the teacher, based on the age of the students of available resources and skills, to decide whether to use a spreadsheet as Excel (Salkind, 2016) or Calc, Geogebra (Prodromou, 2014). R (Ferreira, Kataoka, & Karrer, 2014) or online available tools such as Core Math Tools (Hirsch, Keller, Fonger & Edson, 2013).

An area that presents an interesting development possibility is the statistical packages that allow you to manipulate, summarize, and analyze data using statistical techniques and methods. Particularly interesting is the evolution of the Apps for Mobile Device (Edwards, 2015).

Fifth Idea: Small and Big Data

A question arises; in the school, it makes sense to do statistics with a few data? A clear answer does not exist, but it is clear that confine itself to processing only a few components (“guys, calculate the arithmetic mean of these five numbers”) is limited and often misleading.

There not here the place to deepen the important problem of the so-called big data. Nevertheless, is certainly useful to work with a data set sufficiently wide to provide materials for not trivial considerations and inferences.

Let us clarify this aspect. In the school it is unthinkable to work on large data sets (often not available) that would require not only appropriate software, but also a different approach. Martin Lindstrom (2016) says: “Big Data is all about finding correlations, but Small Data is all about finding the causation”.

The balance should be a dataset sufficiently significant but easy to handle.

For many problems and questions, small data are enough. Everything processed in Excel is small data.

Anthropometric data of sufficient amplitude are therefore strongly recommended. For example, in Italy, you can use the data collected from CIRDIS (<http://cirdis.stat.unipg.it/index.php?canale=160&lang=ita>).

Another suggestion (Frictionless Data) can be found at <http://data.okfn.org/> where a small set of lightweight data package provide a base structure on which tooling and integration can build.

Sixth Idea: The Probability is Not Just a Game

It is true that the birth of the first probabilistic concepts is closely linked to games of chance, but it would be wrong to introduce students only to this aspect. Without pretending to go into detail it is necessary to present a number of areas in which you can use stochastic models, for example: environmental (weather forecast), economic (predictive models), medical (diagnostic), social (insurance).

It is desirable the use of artefacts, spreadsheet and Web resources such as mortality tables.

Weather Forecast

A rain gauge (Figure 6) is an instrument used by meteorologists to gather and measure the amount of liquid precipitation over a set period of time.

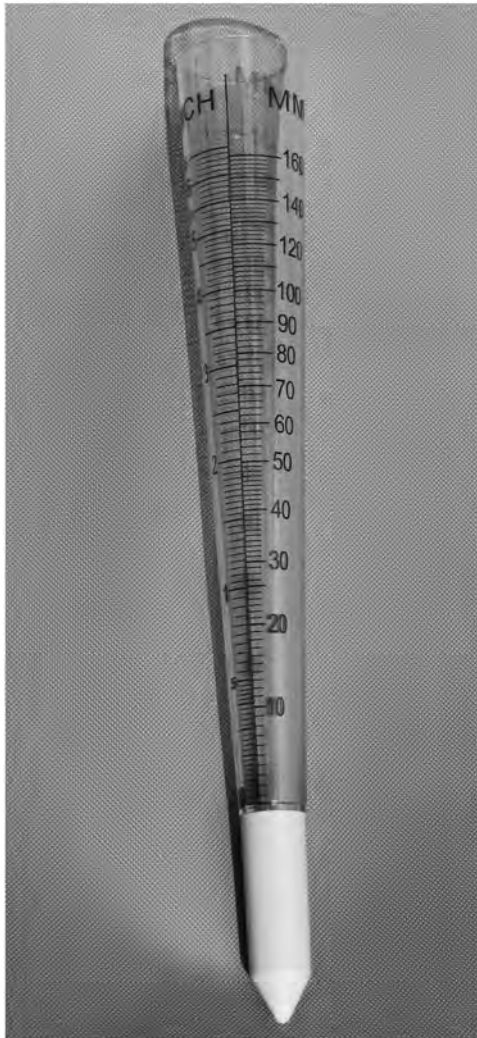
Analysis of weather conditions is an ideal tool to introduce statistical and probabilistic concepts. The theme is very timely, given the general level of interest shown by the population.

Predictive Models

The Big Mac (Figure 7) index was invented by The Economist in 1986 as a light-hearted guide to whether currencies are at their “correct” level. It is based on the theory of purchasing-power parity (PPP), the notion that in the long run exchange



Figure 6. Rain gauge



rates should move towards the rate that would equalize the prices of an identical basket of goods and services (in this case, a burger) in any of the two countries whose currencies are compared.

The Big Mac PPP exchange rate between two countries is obtained by dividing the price of a Big Mac in one country (in its currency) by the price of a Big Mac in another country (in its currency). This value is then compared with the actual exchange rate.

Figure 7. Big Mac



Diagnostic

An ultrasound examination can provide a starting point for a statistical survey (Gross, Callen, Filly, 1982).

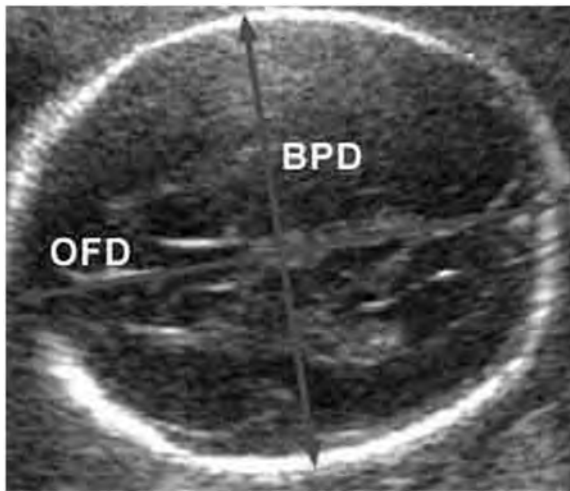
By indicating with BPD the bi-parietal diameter and with OFD the occipital-frontal diameter (Figure 8) the problem can be formulated as follows: “From these ultrasound data, which is the formula that can more accurately estimate the head circumference (CC) admitted that the machine being used is not sufficient to measure it?” (Drivet, 2006)

In addition to reasoning of geometric type, the solution of the problem also involves references of statistical type.

In the literature we can find numerous formulas to be compared with the empirical data.

1. $(BPD + OFD) / 2 * 3.14$
2. $\sqrt{((BPD^2 + OFD^2) / 2)} * 3.14$
3. $\sqrt{(BPD * OFD)} * 3.14$
4. $(3 * (BPD / OFD + 2/2) - \sqrt{((BPD / OFD * 2 + 3/2) * (3 * BPD / OFD + 2/2))}) * 3.14$
5. $BPD * OFD + 3 + 1.57 *$
6. $(BPD + OFD) * 1.62$

Figure 8. Echography



The comparison between the different formulas and the estimate of the circumference obtained by the method of least squares, from $z = ax + by$ model, allows students to compete not with the more or less realistic usual problem, but with a sufficiently wide and challenging case study.

Insurance

A life table is a concise way of showing the probabilities of living or dying for a member of a particular population at a particular age.

The table is used by actuaries, demographers etc. to present the mortality experience of a population. This method is applicable to the analysis of many measurable processes and mainly in the insurance field. For detailed analysis, see Namboodiri & Suchindran (2013).

From the tables it is possible to obtain information to answer questions like: what is the probability that a woman aged 20 years will survive fifty years? What is the average number of years of life remaining for men who have just reached their 40th birthday?

The idea that the probability is not calculated as the ratio between the number of heads and coin flips, but as the ratio of survivors of a cohort, is always amazing (Figure 9).

Seventh Idea: The Dynamic Statistic

“Many teachers firmly believe that most students hate statistics, perceive it to be a necessary but painful class to teach, and imagine that it will naturally result in poor course evaluations. Unfortunately, these myths can become self-fulfilling prophecies”. (Hulsizer-Woolf, 2009).

This assertion is perhaps deliberately extreme. It is not actually true that students have a difficult relationship with the statistic. They simply are faced to a not credible simulacrum.

Construction of surveys, research, and creation graphics are essential, but a real quantum leap you can get giving dynamism and historical meaning to the statistics. A good practice is to make use of resources such as the Interactive maps, for example: IMFDATAMAPPER, MAPPINWORLDS or GAPMINDER (Figure 10).

Gapminder Foundation is a non-profit venture registered in Stockholm, Sweden, that promotes the use and understanding of statistics and other information about social, economic and environmental development at local, national and global levels.

Eighth Idea: Modelling

“Modelling is an essential process and the introduction of basic probabilistic notions raises specific issues [...]. Students meet a new difficulty when they have to link probabilistic notions to reality” (Batanero, Burrill, Reading, 2011).

The approach to patterns associated with random events is perhaps the culmination of an educational path. It allows you to switch from probability to statistics in a concrete and meaningful mode. This is one of the cornerstones of the argument: under very general conditions the simultaneous action of a large number of random factors leads to a substantially predictable effect.

“Only with a large number of single events randomness is lost and you fall under the control of statistical laws”. (Eigen & Winkler, 1986).

An excellent example of modelling is due to the above mentioned authors.

Eight Tips for the Theme, "Data and Forecasts"

Figure 9. Mortality table

d-mo istat.it [Vista per SINGOLA AREA](#) [DOWNLOAD](#) [NOTE](#)

Anno: 2014 [Salva](#) [Stampa](#)

Tavola completa: ☐
Tavola ridotta: ☐

Maschi: ☐
Femmine: ☐

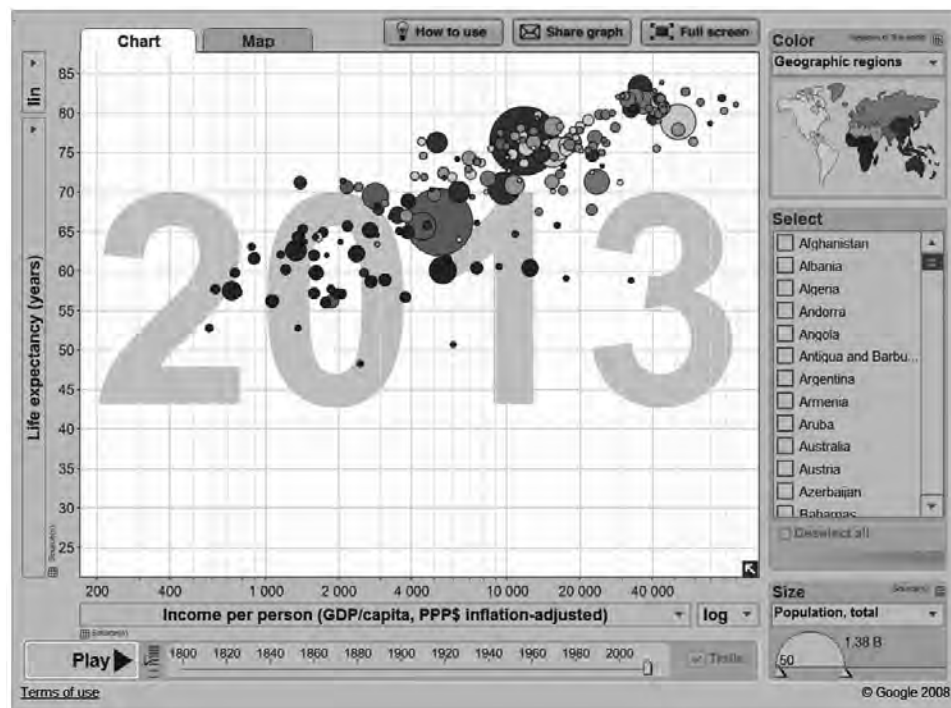
Ripartizione:
Regione:
Provincia:

[Tavola](#)

Tavole di mortalità della popolazione italiana
Ripartizione: Italia - Maschi - Anno: 2014

Età x	Sopravvivenzi l_x	Decessi d_x	Probabilità di morte (per mille) q_x	Anni vissuti L_x	Probabilità prospettive di sopravvivenza P_x	Speranza di vita e_x
0-4	100000	392	3.92380	498229	0.9994025	80.281
5-9	99608	42	0.41695	497931	0.9995809	75.594
10-14	99566	51	0.50888	497722	0.9990652	70.625
15-19	99515	148	1.48937	497257	0.9980574	65.659
20-24	99367	224	2.25383	496291	0.9976129	60.753
25-29	99143	246	2.47959	495106	0.9973627	55.884
30-34	98897	287	2.90295	493801	0.9965732	51.017
35-39	98610	395	4.00164	492108	0.9952044	46.158
40-44	98216	578	5.88295	489749	0.9923846	41.333
45-49	97638	945	9.67642	486019	0.9874979	36.561
50-54	96693	1525	15.77269	479943	0.9797275	31.892
55-59	95168	2438	25.61635	470213	0.9667838	27.360
60-64	92730	3922	42.29547	454594	0.9463260	23.009
65-69	88808	5884	66.25780	430194	0.9149797	18.906
70-74	82924	8930	107.68789	393619	0.8642634	15.060
75-79	73994	12796	172.93115	340191	0.7708847	11.558
80-84	61198	18322	299.39006	262248	0.6199615	8.415
85-89	42876	20628	481.10389	162584	0.4285403	5.895
90-94	22248	14905	669.95396	69674	0.2637073	4.053
95-99	7343	6177	841.27218	18373	0.1124940	2.793
100-104	1166	1111	953.48493	2067	0.0319165	1.830
105-109	54	54	992.78403	56	0.0051749	1.223
110-114	0	0	999.41664	0	0.0004562	0.873
115-119	0	0	999.97226	0	0.0000239	0.682

Figure 10. Gapminder



To simulate the trend of a population you imagine a game in which black and white pawns are the elements of a population. A chessboard 6x6 is the space where you play the various processes. Two players have randomly put their pieces on the board so that each occupies half of the available squares (Figure 11). Then we introduce the rules of the game that will lead to very different asymptotic probability distributions.

- “Drifting” You flip a coin. If the result is heads, white removes a black pawn and replaces it with a white from its reserves. Otherwise, a white pawn is replaced with a black one.
- “Balance” You roll two dice: the checker from the box corresponding to the coordinates resulting from the shooting is removed and replaced with a spare pawn of the opponent.
- “All or nothing” You roll two dice: the checker from the box corresponding to the coordinates resulting from the roll is doubled at the expense of color opponent. Therefore, if you have obtained a square with a white pawn removes any black pawn and replace it with another white from the reserve.

In the presented cases the asymptotic probability will lead, respectively, to uniform distribution, Gaussian curve, two possible singularities (all or nothing).

Issues

“Probability does not exist” (De Finetti, 1974).

The paradoxical statement, means that the probability is not an immanent property, but a feature of our reasoning about the world.

In this sense, the previous statement opens a window on the subjective probability that is too often neglected in school while would be useful as a first step towards the exercise of critical thinking and a start to the practice of argumentation.

Just one example: how trustworthy are the predictions of bookmakers regarding football?

We have to collect the data on the concordance between prediction (considering winning the team with the lowest quotation) and result to draw conclusions. Merely by way of curiosity, in Italy the percentage, calculated on the last two championships, is equal to 50%!

Figure 11. The chessboard as model



Controversies

“Probability is increasingly taking part in the school mathematics curriculum; yet most teachers have little experience with probability and share with their students a variety of probabilistic misconceptions. Probability is difficult to teach for various reasons, including disparity between intuition and conceptual development even as regards apparently elementary concepts. Since an education that only focuses on technical skills is unlikely to help teachers overcome their erroneous beliefs, it is important to find new ways to teach probability to them, while at the same time bridging their content knowledge and their pedagogical content knowledge” (Batanero et al., 2005).

The raised issues, which will be taken up to the conclusions, are particularly interesting. To these we could add another level of discussion, the contrast between teaching for problems and teaching for patterns. According Carla Rossi (1994) the most appropriate path for teaching probability and statistics should be inductive and follow the scheme:

significant problem → resolving model → results (forecast) → comparison with data → possible modification of the previous model

Problems

Brousseau (1983) describes very clearly some pedagogical obstacles.

Knowledge can be a source of distortions in the sense that the student uses this knowledge to produce the correct answer in a given and known context. When this knowledge is used out of context, it generates errors. The student ignores the contradictions produced by the obstacle and resists to establishing a deeper knowledge.

“Moreover, errors and difficulties do not arise in a random, unpredictable way. Frequently it is possible to discover regularities in them, to find some association with other variables of the

proposed tasks, of the subjects, or of the present and past circumstances” (Batanero et al., 1994).

Working with objects, doing research, making deductions, involves commitment and effort. We must never forget that students tend to apply the criteria of applied mathematics: profit maximization and cost minimization.

Trying to identify objects and situations that speak of a “different Mathematics”, actually embodied in everyday life, could provide more opportunity to treat the innumeracy that seems to afflict much of our society.

FUTURE RESEARCH DIRECTIONS

For the understanding of probabilistic thinking we must to emphasize two aspects. The first concerns the concept of intuitive knowledge as adaptation to the surrounding environment shaped by experience. The second aspect shows that probabilistic intuition differs from the intuition that one has in other mathematical and scientific areas.

Especially in the latter case it is important to find the links between probability and the world around us, with all its objects but also moods or abstract concepts. In this sense, it is crucial to start from significant problems and artifacts to build a solid framework.

SOLUTIONS AND RECOMMENDATIONS

“However, it is also crucial to connect the introduction of probability with the statistical aspects that are involved in everyday life. Students are quite often unable to appropriately take statistical information into account to evaluate probabilities. On the other hand, it is quite remarkable that teachers lack specific knowledge in this field” (Garuti, Orlandoni & Ricci, 2008).

More than solutions it is appropriate to give ideas to those teachers who want to use, at least

in part, this approach. The recommendation is to sparingly use the objects, over a school year, choosing from those which can be situated in a curriculum. For example, you may decide to develop a theme such as the probability by using a dozen objects with a path that, starting from the simplest objects (dice, cards, coins) can get to touch more complex issues such as the binomial distribution (Quincunx or Galton box).

CONCLUSION

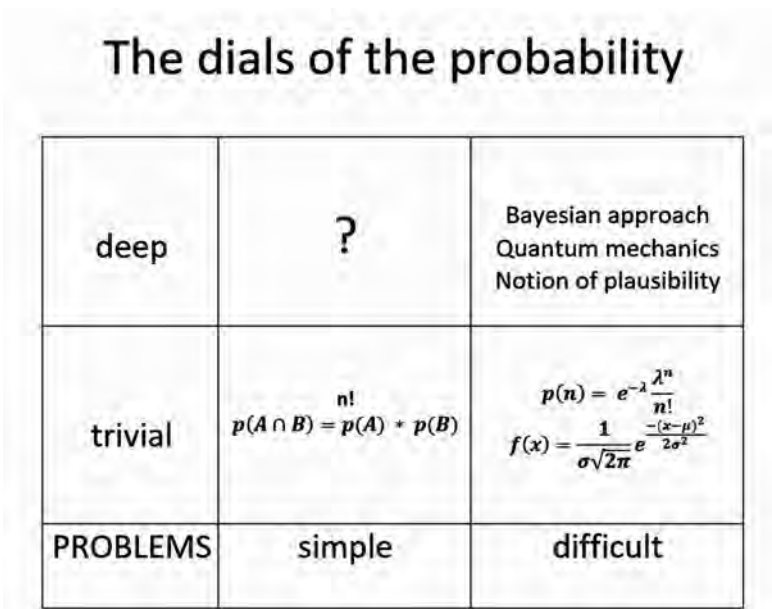
We can make some assumptions about why, in school, probability and statistic are, at all levels, as Cinderella. We can single out some aspects of the problem:

- 1. Lack of specific training of teachers in the context Data and Forecasts. Without wishing to generalize it is normal to meet many teachers (or aspiring) who claim to have a very poor university education in this area;
- 2. More limited experimentation and educational research than other nuclei, although

there are praiseworthy local or national initiatives;

- 3. Reluctance to broaden the discourse over repetitive examples of textbooks when the tools of probability and statistics can refer to a very wide range of possible fields of application. For example:
 - a. Environmental: weather forecast, earthquake risk map
 - b. Biological: genetic
 - c. Economic: insurance, forecasting models
 - d. Physical: statistical mechanics
 - e. Legal: evidence assessment
 - f. IT: big data processing
 - g. Language: comparison to award author styles, encryption
 - h. Fun: games of chance
 - i. Medical: diagnostic, screening, spreading epidemics
 - j. Social: data mining, queue management, transport organization
 - k. Technology: quality control, stochastic communication networks

Figure 12. The dials of the probability



4. Limited use of IT tools, not just those complexes although free as R or KNIME, but especially instruments with a favorable learning curve as Excel. It takes only several lines of instruction to turn a spreadsheet into a powerful simulation tool;
5. Excessive use of examples and abstract formulations with exclusive reference to the algebraic language. Here the examples can be multiplied: from combinatorial formulas to Bayes Theorem is a continuous run to remove the student from the discipline pleasure. A reflection should refer to the analysis of simple and deep problems (Figure 12).

Anyway, we must not forget that: “This branch of mathematics is the only one, I believe, in which good writers frequently get results which are entirely erroneous” (Pierce, 1982).

REFERENCES

- Arrigo, G. (2010). Le misconcezioni degli allievi di scuola primaria relative al concetto di probabilità matematica. *Bollettino dei docenti di Matematica*, Bellinzona.
- Batanero, C., Biehler, R., Maxara, C., Engel, J., & Vogel, M. (2005). Using simulation to bridge teachers content and pedagogical knowledge in probability. *15th ICMI Study Conference: The Professional education and development of teachers of mathematics*.
- Batanero, C., Burrill, G., & Reading, C. (2011). *Teaching Statistics in School Mathematics-Challenges for Teaching and Teacher Education*. Springer Science & Business Media.
- Batanero, C., Godino, J. D., & Roa, R. (2004). Training teachers to teach probability. *Journal of statistics. Education*, 12(1), 1–19.
- Batanero, C., Godino, J. D., Vallecillos, A., Green, D. R., & Holmes, P. (1994). Errors and difficulties in understanding elementary statistical concepts. *International Journal of Mathematical Education in Science and Technology*, 25(4), 527–547. doi:10.1080/0020739940250406
- Brousseau, G. (1983). Les obstacles épistémologiques et les problèmes en mathématiques. *Recherches en Didactique des Mathématiques*, 4(2), 164–198.
- De Finetti, B. (1974). *Theory of probability: Vol. 1. No. 2*. New York: Wiley.
- Drivet, A. (2006). Modelli per analisi ecografiche. *Induzioni*, 32(1), 1000–1011.
- Drivet A. (2013). *La cassetta degli attrezzi*. Pubblicato dall’Autore.
- Drivet, A. (2013). *The toolbox: objects and tools for doing mathematics. Didactic Strategies and Technologies for Education*. IGI Global.
- Drivet, A., Della Penna, A., & Signore, D. A. (2012). Artefatti per l’insegnamento della probabilità: Il laboratorio di ottobre scienza= Tools for Teaching Probability. *Induzioni*, 45(2), 79–94.
- Edwards, H. P. (2015). Probability and Statistics Apps for Mobile Devices: A Review. *Integrating Touch-Enabled and Mobile Devices into Contemporary Mathematics Education*, 242.
- Eigen, M., & Winkler, R. (1986). *Il gioco*. Adelphi, Milano.
- Ferreira, R. D. S., Kataoka, V. Y., & Karrer, M. (2014). *Teaching Probability With The Support Of The R Statistical Software 13*. Academic Press.
- Garuti, R., Orlandoni, A., & Ricci, R. (2008). *Which probability do we have to meet? A case study about statistical and classical approach to probability in students’ behaviour*. Reading: A. Rossman.

Gattuso, L. (2011). L'enseignement de la statistique: Où, quand, comment, pourquoi pas? *Statistique et Enseignement*, 2(1), 5–30.

Gross, B. H., Callen, P. W., & Filly, R. A. (1982). The relationship of fetal transverse body diameter and biparietal diameter in the diagnosis of intra-uterine growth retardation. *Journal of Ultrasound in Medicine*, 1, 361–365. PMID:6152958

Hirsch, C., Keller, B., Fonger, N., & Edson, A. (2013). Core Math Tools. *Common Core Mathematics Standards and Implementing Digital Technologies*, 1.

Hulsizer, M. R., & Woolf, L. M. (2009). *A guide to teaching statistics: innovations and best practices* (Vol. 10). John Wiley & Sons.

Kapadia, R., & Borovcnik, M. (Eds.). (2012). *Chance encounters: Probability in education* (Vol. 12). Springer Science & Business Media.

Lindstrom, M. (2016). *Small Data: The Tiny Clues that Uncover Huge Trends*. St. Martin's Press.

Namboodiri, K., & Suchindran, C. M. (2013). *Life table techniques and their applications*. Academic Press.

OECD. (2012). *Pisa Mathematics Framework*. OECD Publishing.

Peirce, C. S. (2009). *Writings of Charles S. Peirce: A Chronological Edition* (vols. 1-6). Indiana University Press.

Pfenning, N. (1998). *Chances Are: Making Probability and Statistics Fun to Learn and Easy to Teach*. Prufrock press inc.

Prodromou, T. (2014). GeoGebra in Teaching and Learning Introductory Statistics. *Electronic Journal of Mathematics & Technology*, 8(5).

Rohatgi, V. K., & Saleh, A. M. E. (2011). *An introduction to probability and statistics* (Vol. 910). John Wiley & Sons.

Rossi, C. (1994). *Ragionamento induttivo-ragionamento deduttivo: problemi e implicazioni nell'insegnamento del calcolo delle probabilità e della statistica*. Atti del XVI convegno nazionale sull'insegnamento della matematica: Probabilità e Statistica nella scuola secondaria, Notiziario della Unione Matematica Italiana, suppl. al n.7, 27-42.

Salkind, N. J. (2016). *Statistics for People Who (Think They) Hate Statistics: Using Microsoft Excel 2016*. Sage Publications.

Steinbring, H. (1991). The theoretical nature of probability in the classroom. In *Chance encounters: Probability in education* (pp. 135–167). Springer Netherlands. doi:10.1007/978-94-011-3532-0_5

ADDITIONAL READING

Alavi, O., Mohammadi, K., & Mostafaeipour, A. (2016). Evaluating the suitability of wind speed probability distribution models: A case of study of east and southeast parts of Iran. *Energy Conversion and Management*, 119, 101–108. doi:10.1016/j.enconman.2016.04.039

Brijlall, D. (2014). Exploring the pedagogical content knowledge for teaching probability in middle school: A South African case study. *International Journal of Educational Sciences*, 7(3), 719–726.

Bui, D. T., Pradhan, B., Lofman, O., Revhaug, I., & Dick, Ø. B. (2013). Regional prediction of landslide hazard using probability analysis of intense rainfall in the Hoa Binh province, Vietnam. *Natural Hazards*, 66(2), 707–730. doi:10.1007/s11069-012-0510-0

Carniel, N. (2016). Measuring risk attitude in travel insurance: a case study.



Di Forti, M., Marconi, A., Carra, E., Fraietta, S., Trotta, A., Bonomo, M., & Stilo, S. A. et al. (2015). Proportion of patients in south London with first-episode psychosis attributable to use of high potency cannabis: A case-control study. *The Lancet Psychiatry*, 2(3), 233–238. doi:10.1016/S2215-0366(14)00117-5 PMID:26359901

Ghiselli, S., & Pesenti, L. (2015). Determining factors in the job search strategies: A multivariate analysis. *Sociologia del Lavoro*, (137): 95–114. doi:10.3280/SL2015-137006

Goerlandt, F., & Montewka, J. (2015). A framework for risk analysis of maritime transportation systems: A case study for oil spill from tankers in a ship–ship collision. *Safety Science*, 76, 42–66. doi:10.1016/j.ssci.2015.02.009

Vandecasteele, I., Rivero, I. M., Sala, S., Baranzelli, C., Barranco, R., Batelaan, O., & Lavalle, C. (2015). Impact of shale gas development on water resources: A case study in Northern Poland. *Environmental Management*, 55(6), 1285–1299. doi:10.1007/s00267-015-0454-8 PMID:25877457

KEY TERMS AND DEFINITIONS

Artifacts: Sometimes indicates an object in a digital environment.

Big Data: A term for data sets that are so large or complex that traditional data processing applications are inadequate.

Combinatorics: A branch of math concerning the study of finite or countable discrete structures.

Dynamic Statistic: The use of dynamic graphics for the representation of statistical data.

Information Technology: The use of a device to store, retrieve, transmit and manipulate data.

Predictive Modelling: Use of statistics to predict outcomes.

Probability: A number between 0 and 1 that measure the likelihood of an event.

Statistic: Discipline that has as its goal the study of quantity and quality of a phenomenon.

Stochastic Modelling: A model that depends on random variables.

Exploring New Handwriting Parameters for Writer Identification



Verónica Inés Aubin

Universidad Nacional de La Matanza, Argentina

Jorge Horacio Doorn

Universidad Nacional del Oeste, Argentina & Universidad Nacional de La Matanza, Argentina

INTRODUCTION

In the XIX century, Giovanni Morelli -Verona, February 25th 1816 – Milán, February 28th 1891- introduced a paradigmatic change in the detection of forgery of famous paintings (Ginzburg & Davin, 1980). Before him, famous picture authentication was based on global indicators such as general illumination, perspective, body position, smiles, and gaze direction, among others. However, such well known characteristics are easy to imitate.

On the contrary, Giovanni Morelli believed that forgeries should be detected watching minor details, such as those less influenced by the pictorial school to which the artist belonged. Every artist automatically produces, almost unconsciously, some details such as nails, fingers, toes, earlobes. These details were considered by Morelli as notoriously revealing, since they appear when the artist's control relaxes and individual impulse take control of his or her actions. Following his conceptual model, Morelli created a large catalog of the minor details of many artists such as Boticelli, Leonardo, Rafael and others. Obviously, trained observers were needed to apply the comparisons advocated by Morelli.

Morelli's method has had a cultural influence over many fields where the search for identifying details began. This change of paradigm led to the use of smaller details in many fields such as people identification. Later, the necessity for better people identification techniques increased. In 1879, Alphonse Bertillon (1893) created an

anthropometric method based on many physical measurements. A few years later, Francis Galton proposed a simpler identification method based on fingerprints. This miniaturization process reached a possible end when started the use of DNA for people identification.

In the specific field of handwriting recognition domain, the evolution followed the same pattern, since old practices, based on global characteristics, such as geometric observable information was replaced by new practices based in parameters which are the result of measurements of smaller characteristics. This chapter proposes some new smaller characteristics, barely observable even by zooming of captured images of the manuscript text.

The automatic processing of handwriting samples is part of the computational biometric. It applies qualitative and quantitative techniques by means of capturing, visualizing and analyzing handwriting. The main applications are writer identification and text understanding. Two significantly different situations appear: on line and off line data capturing. In the former the samples are obtained in a dedicated framework, where the writing instrument and the surface have several sensors. In the latter, the unique information available comes from the residues left on the paper. This chapter deals with the second situation. Width, gray value, direction and other parameters of the residual manuscript text were influenced by the psychomotor characteristics of the writer. Some of these personal parameters may be estimated from the observable properties of the written text.

Any automatic or semi-automatic handwriting recognition process is always composed by at least two activities: i) recognition parameters calculation and ii) comparison of sample parameters with a data bank of registered parameters. Parameter calculation and parameter comparison are quasi independent activities, since almost every parameter set may be used by almost every comparison technique. Very infrequently a better efficiency and efficacy may be obtained by combining certain parameters sets with a given comparison technique. This chapter proposes some new parameters and enhances the understanding of some already in use parameters, leaving the choice of the comparison technique to later studies. The authors believe that the major contribution of the research whose results are presented in this chapter is the better comprehension of the mechanism that creates the residues left on the paper, since some unproven and largely used hypothesis are criticized in some degree.

BACKGROUND

The information recovered off-line from the static residues left on the paper is called pseudo-dynamic. It is clearly different from the actual dynamic information captured on-line during the writing activity. There are many contributions proposing different pseudo-dynamic characteristics, most of them oriented to signature authentication, which is a sub-problem of identification of the author of free text. Some of them are summarized in the next paragraphs.

The work of Ammar et al. (1986) the histogram of the gray level of the signature is analyzed and a pressure threshold is proposed. The signature is described in terms of the percentage of pixels whose gray value that exceeds such threshold, along with other parameters such as the minimum and maximum gray value and the dynamic range of the signature image. Those parameters are compared with a set of signature parameters using classifiers based in Euclidian distance.

A combination of geometric and pseudo-dynamic parameters is used in Huang and Yan (1997). It is based on strokes geometric parameters such as contour, skeleton and the high pressure regions; these parameters feed a classifier based on neural networks.

The proposal introduced in (Fang, 1999), is oriented to geometry; it places the focus on strokes whose size exceeds a length threshold. A softness index is calculated for every one of these strokes. The quotient between the number of soft strokes and the total number of strokes becomes an identification parameter.

A strategy based on pseudo-dynamic characteristics is described in Mitra et al. (2005). It is oriented to the detection of elaborated forgeries. The pressure applied by the writing instrument, estimated by using the gray level of the image, is considered the most relevant parameter. The strategy chooses low pressure pixels as identification parameters. No information about the paper used, the writing instrument or the base where the paper was resting is given.

Another proposal oriented to geometric parameters can be found in Oliveira et al. (2005). It is biased to recognition or authentication of signatures, since it uses parameters not available on free text. Several new geometric parameters are introduced such as the quotient between high and wide, a softness index and inclination in relation with a baseline, among others. Using a similar set of geometric parameters, Lv et al. (2005) propose a Chinese signature verification method. In addition to geometric parameters they include two pseudo-dynamic characteristics: the pixel gray level distribution and the wide of the line. Both parameters are supposed to indicate the pressure applied during the writing process. No indication about the paper, the writing instrument or the base used is given.

In two consecutive articles Vargas et al. (2008, 2011) propose two methods for signature authentication or recognition. Both are biased to signatures. In the first article, the geometric center of the signature is calculated and a polar geometric

system is established on it. An image depicting the distribution of high pressure pixels, contained in polar sectors, is created. The polar sectors are defined by angle and radius limits. This image is binarized using a threshold which is calculated using the gray level histogram of this image. In the second article, the pixel gray level distribution is analyzed using the wavelet transform and the patterns of this distribution is determined by means of the co-occurrence matrix.

In an article oriented to paleographic authors recognition the relation of wide of the stroke and the direction of the trace is analyzed (Brink, Smit, Bulacu, & Schomaker, 2012). The importance of this relationship for author identification is highlighted. In paleographic text this relationship is described in terms of the writing instrument used which was a capillarity-based liquid ink pen. These conclusions are extended to modern handwritings texts. In this case the causes of such influence are attributed to the increasing of the force applied by the stroke authors when the line is going in certain directions. Individual variations are supposed to be partially deliberated and partially unconscious.

NEW PARAMETERS

When dealing with pseudo-dynamic parameter, practically all authors use pixel gray level, its distribution along the stroke, its histogram or other view. Implicitly or explicitly this attribute is referred as the *pressure* applied by the writing instrument on the paper (Ammar, Yoshida, & Fukumura, 1986; Fang et al., 1999; Lv, Wang, Wang, & Zhuo, 2005; Mitra, Banerjee & Ardil, 2005; Oliveira, Justino, Freitas, & Sabourin, 2005; Plamondon, & Srihari, 2000; Vargas, Ferrer, Travieso, & Alonso, 2008, 2011). Even more than that, many times the gray level attribute is directly mentioned as *high pressure* pixels or *low pressure* pixels. In a few articles the parameter *wide of the line* is also attributed to the *pressure* applied (Lv et al., 2005). In the consulted publications, those cited and many others, no indication about the

paper used, the writing instrument or the base used is given. Moreover, neither it was found an article proving the relation between the *pressure* and the gray level nor a citation to such sort of experience. This lack of data motivated the authors of this chapter to do some experiments on that matter. The link between pressure and gray value seemed to be so evident that there was no intention to refute or falsify such relationship (Popper, 1963). The purpose was just to know more about it: to see if it is or not a linear relationship or if it has any peculiarity that may help to understand more clearly the residues left on the paper or surface over which the writing process takes place.

Strictly the use of the word *pressure* in this context is ambiguous since it is not clear if either it is used with the common meaning of the *exertion of force upon a surface by an object* or the more precise semantic taken from physics science as the *force per unit area* (Merriam-Webster Inc., 2004). In what follow, the word *pressure* is replaced by the word *force*.

Vertical Strokes

A very basic experimental array was set up using the followings elements:

- *A writing instrument.*
- *A set of small spheres* of controlled mass.
- *A tube* able to hold the writing instrument and the spheres.

The friction between the tube and the spheres or the writing instrument was ensured to be irrelevant. Inserting the writing instrument in the tube and the some of the spheres on top of it allowed having the possibility of creating strokes, applying controlled forces. At this time the tube and the writing instrument were applied perpendicularly over the paper.

This experimental array may vary upon the use of different writing instruments, different quality of papers and upon the number of sheets. In an exploratory experience like this one, there was no



intention of covering all possible combinations of such factors. For that reason, only one few combinations were used. Changes in the ink color were done, since this factor which was perceived at that time, as being the most important one.

Cross Section and Gray Level

Every stroke was digitalized in such a way that ensures a perfect horizontal image. Figure 1 shows the gray level of pixels belonging to a vertical line crossing the stroke. The gray level of the pixels located before and after the peak represents the gray value g_p , of the unwritten portion of the sheet. For each pixel of the skeleton, the gray level of the darkest pixel on such vertical segment g_d is chosen. The medium gray level of all pixels chosen is taken as the gray level of the stroke; and it is shown in the Figure 2a. To calculate a wide of the stroke, for each pixel of the skeleton a reference gray level g_r is calculated as:

$$g_r = (g_d + g_p) / 2$$

The cross section is defined here as the number of pixel of the vertical segment whose gray value coincides or is darker than g_r . Strictly, the cross section is the wide of the stroke having at least a given gray level. The medium of such cross section wide along the whole stroke is presented in Figure 2b, as a function of the force applied.

Using different spheres, straight strokes under forces from 10 grams -force to 150 grams-force were obtained. Figure 2a presents the influence of the force applied on the cross section wide of the stroke. Figure 2b depicts the relationship of the gray level of the stroke with the force applied. The influence of the force applied on the gray level is almost linear when the applied force varies between 0 and 50 grams-force. If stronger forces are applied the gray level changes very little. For the cross section, the influence of the force keeps its linearity up to 150 grams-force.

Inclined Strokes

Concurrently with the controlled strokes with specific forces, direction and position of the writing instrument, samples of totally uncontrolled free writings were taken, using the same writing instrument, same paper and same number of sheets. The observation of the gestural actions of the writers and the residues left on the sheet of paper, lead to propose a new hypothesis linking the cross section and gray level with the position of the writing instrument.

The experimental array was slightly modified by cutting the tube in a predefined angle. This forces the writing instrument to have a non vertical contact with the sheet of paper. The position of the writing instrument may be in any position in relation with the direction of the stroke. To

Figure 1. Gray value along a line vertical to the trace

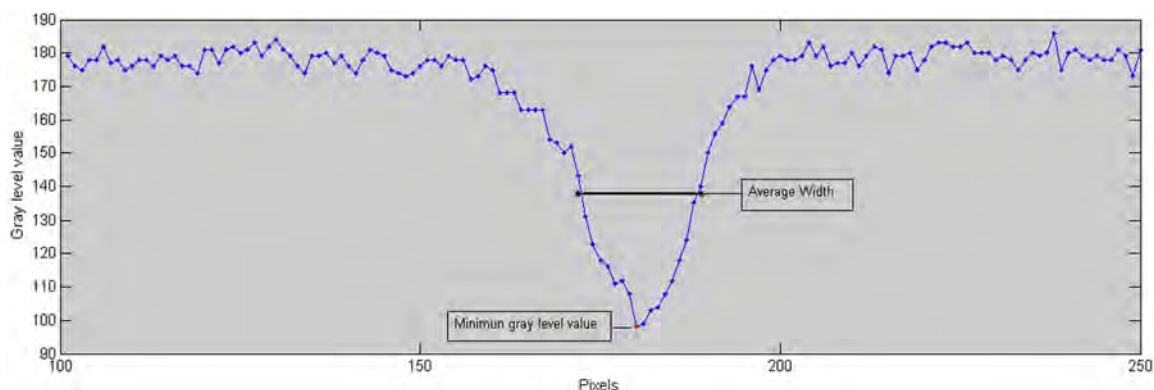
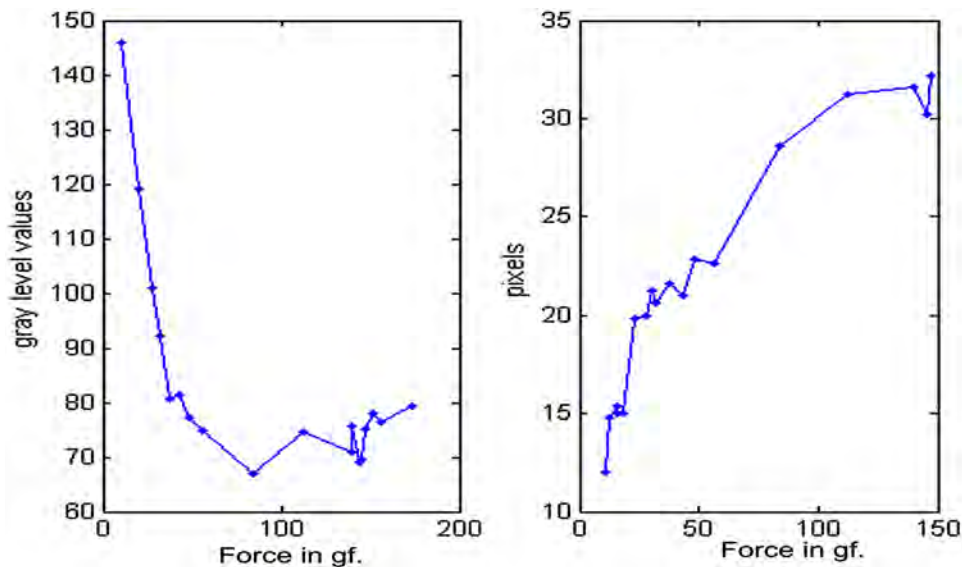


Figure 2. (a) Gray level of darker pixels for different forces. (b) Cross section of the stroke for different forces



reduce the number of samples, three archetypical positions were used:

- **Inclined Against the Stroke:** The projection of the position of the writing instrument coincides with the fragment of the stroke already created.
- **Inclined Forward the Stroke:** The projection of the position of the writing instrument coincides with the fragment of the stroke not created yet.
- **Inclined Perpendicularly to the Stroke:** The projection of the position of the writing instrument falls at a right angle with the stroke.

Using different tubes, inclined strokes samples were obtained for 30 and 45 degrees angles between the writing instruments and the vertical. For angles forward the stroke, the whole stroke looks lighter than the vertical one with similar force. On the other hand when the angle is against the stroke, its gray value is very similar to the vertical one. These results may be observed with just a few samples. Perhaps, controlled experiences involving more samples will distinguish vertical and

inclined against the stroke gray values. It should be noted, that the experimental array was based on the mass of a reduced set of small spheres. Table 1 depicts the average gray value of inclined and vertical strokes under two different forces.

Additionally, samples obtained for strokes when the inclination is perpendicular to the stroke show darkness very similar to those obtained with vertical strokes with similar forces. However, these samples show an important new characteristic: the gray distribution on the wide of the trace is notoriously asymmetric. The darkest pixel on a line perpendicular to the stroke appears away from the skeleton. This is rather different from what is shown in Figure 1 where the darkest point and the skeleton coincide or they are close to each other. When the inclination increases, both the asymmetry and the distance between the skeleton and the darkest pixel also increase. Table 2 contains some data related to the distance between the skeleton and the darkest point.

The asymmetry of the gray level across the stroke may also be visualized comparing the average gray value on the pixels on both sides of the skeleton. Table 3 contains come data on that matter.

Figure 3. Experimental array

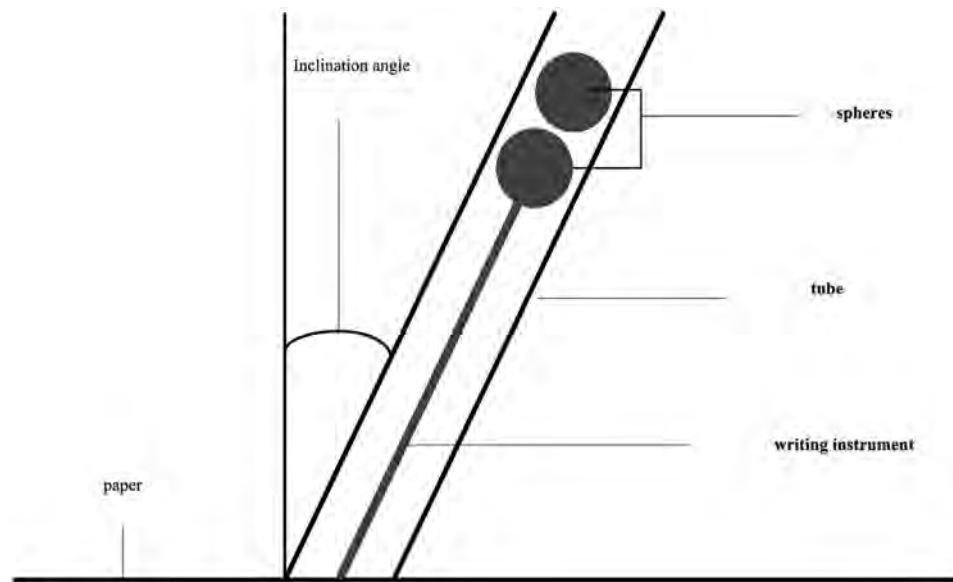


Table 1. Gray level of vertical and inclined stroke under different forces

Force in gf.	Vertical	45° Inclined Against the Stroke
75	169,33	187,53
90	141,87	157,52

Table 2. Distance, in number of pixels from the darkest pixel and the skeleton

Sample	30° Angle	45° Angle
1	20,99	28,29
2	24,45	29,68
3	22,29	28,48
4	26,99	30,04

SOLUTIONS AND RECOMMENDATIONS

Just with the data collected on this preliminary experiment, it is possible refute the widely accepted assumption that the gray level of the stroke is an indication of the force applied by the writer. Actually the gray level of the pixels depends, at least, upon two factors: the force applied and the inclination of the writing instrument.

On handwritings free of restrictions most of the strokes are curves. On this curves the distance between the darkest pixel and the skeleton varies along stroke. The perpendicular to the tangent to the skeleton of the curve at the position where the asymmetry reaches a maximum provides a mean to estimate of the direction at which the writing instrument is pointing. The distance between the darkest pixel and the skeleton is an indication of

Table 3. Average and median of the gray value of pixels on both sides of the skeleton

		1	2	3	4	5
Right Half	Average	180,57	166,75	165,63	169,40	165,30
	Median	180,57	173,78	168,63	175,00	166
Left Half	Average	202,34	189,55	190,25	187,70	176,47
	Median	202,35	190,41	190,26	189,41	176

the angle at which the writing instrument was hold. Other significant pixels of the stroke are those where the line created by the darkest pixels coincides with the skeleton. In these points the tangent to the skeleton is another indication of the inclination of the writing instrument.

FUTURE RESEARCH DIRECTIONS

All results presented in this chapter are qualitative since they only established the importance of the studied parameters. No quantitative relationships among observed residues left on the paper and dynamic gestures of the writer, were obtained.

Two main line of research may clearly be pointed out at this time:

- **Determine the Influence of:**
 - The type of paper.
 - The surface where the paper is resting.
 - The type of writing instrument.
- **Include These New Parameters in Classifiers:** To determine their contribution to improve the identification of the writers.

CONCLUSION

This chapter deals with approaches for extracting characteristics from handwriting text. These approaches are non invasive and of low cost. They are of low cost because they only need an image of the text. They are non invasive, since the original text is not modified in any way. Additionally, they allow multiple analyses. They are focused on the small deformation produced on the paper and the distribution of the gray level in every section of the strokes.

Results obtained coincide with already published results Ammar et al. (1986), Mitra et al. (2005), and Vargas et al. (2008) in the sense that the gray value and the cross section [Lv2005] are highly repetitive for strokes of the same grapheme created by the same author. These results were

extended from the point of view of the mechanisms that influence both the grey value and the cross section of the stroke. The implicit hypothesis that only the force applied is the driving cause was refuted since both values are also influenced by the relationship between the inclination of the writing instrument and direction of the movement.

It is also determined that it is more revealing to use the evolution of the gray level along the stroke than the use of histograms (Ammar et al., 1986; Mitra et al., 2005; Vargas et al., 2008). Histograms hide very valuable information such as the distance between the darkest points of the stroke and the skeleton. When this distance is:

- **Near to Zero:** The tangent to the stroke is an indication of the position of the writing instrument.
- **Maximum:** The perpendicular to the tangent is an indication of the position of the writing instrument.

Finally, the distance between the line of the darkest pixels and the skeleton suggest the inclination of the writing instrument.

REFERENCES

- Ammar, M., Yoshida, Y., & Fukumura, T. (1986) A new effective approach for off-line verification of signatures by using pressure features. In *Proceedings of International Conference on Pattern Recognition* (pp. 566-569).
- Bertillon, A. (1893). *Identification anthropométrique: instructions signalétiques*. Impr. administrative.
- Brink, A. A., Smit, J., Bulacu, M. L., & Schoemaker, L. R. B. (2012). Writer identification using directional ink-trace width measurements. *Pattern Recognition*, 45(1), 162–171. doi:10.1016/j.pat-cog.2011.07.005

Fang, B., Wang, Y. Y., Leung, C. H., Tang, Y. Y., Kwok, P. C., Tse, K. W., & Wong, Y. K. (1999). A smoothness index based approach for off-line signature verification. In *Proceedings of the Fifth International Conference on Document Analysis and Recognition, 1999*. (pp. 785-787). IEEE. doi:10.1109/ICDAR.1999.791905

Ginzburg, C., & Davin, A. (1980, April). Morelli, Freud and Sherlock Holmes: clues and scientific method. In *History Workshop* (pp. 5-36). Editorial Collective, History Workshop, Ruskin College. doi:10.1093/hwj/9.1.5

Huang, K., & Yan, H. (1997). Off-line signature verification based on geometric feature extraction and neural network classification. *Pattern Recognition, 30*(1), 9-17. doi:10.1016/S0031-3203(96)00063-5

Lv, H., Wang, W., Wang, C., & Zhuo, Q. (2005). Off-line Chinese signature verification based on support vector machines. *Pattern Recognition Letters, 26*(15), 2390-2399. doi:10.1016/j.patrec.2005.04.013

Merriam-Webster Inc. (2004). *Merriam-Webster's collegiate dictionary*. Merriam-Webster.

Mitra, A., Banerjee, P., & Ardil, C. (2005). Automatic Authentication of handwritten documents via low density pixel measurements. *International Journal of Computational Intelligence, 2*(4), 219-223.

Oliveira, L. S., Justino, E., Freitas, C., & Sabourin, R. (2005). The graphology applied to signature verification. In *12th Conference of the International Graphonomics Society* (pp. 286-290).

Plamondon, R., & Srihari, S. N. (2000). Online and off-line handwriting recognition: A comprehensive survey. *Pattern Analysis and Machine Intelligence. IEEE Transactions on, 22*(1), 63-84.

Popper, K. (1963). *Conjectures and refutations* (Vol. 7). London: Routledge and Kegan Paul.

Vargas, J., Ferrer, M. A., Travieso, C. M., & Alonso, J. (2008). Off-line signature verification based on high pressure polar distribution. In *Proceedings International Conference on Frontiers in Handwriting Recognition*.

Vargas, J. F., Ferrer, M. A., Travieso, C. M., & Alonso, J. B. (2011). Off-line signature verification based on gray level information using texture features. *Pattern Recognition, 44*(2), 375-385. doi:10.1016/j.patcog.2010.07.028

ADDITIONAL READING

Baxes, G. A. (1994). *Digital image processing: principles and applications*. John Wiley & Sons, Inc.

Nelson, W., & Kishon, E. (1991, October). Use of dynamic features for signature verification. In *Systems, Man, and Cybernetics, 1991. 'Decision Aiding for Complex Systems, Conference Proceedings. 1991 IEEE International Conference on* (pp. 201-205). IEEE. doi:10.1109/ICSMC.1991.169685

Sakamoto, D., Morita, H., Ohishi, T., Komiya, Y., & Matsumoto, T. (2001, May). On-line signature verification algorithm incorporating pen position, pen pressure and pen inclination trajectories. In *International Conference on Acoustics, Speech and Signal Processing* (pp. 993-996). IEEE. doi:10.1109/ICASSP.2001.941084

Wijesoma, W. S., Yue, K. W., Chien, K. L., & Chow, T. K. (2001). Online handwritten signature verification for electronic commerce over the internet. In *Web Intelligence: Research and Development* (pp. 227-236). Springer Berlin Heidelberg. doi:10.1007/3-540-45490-X_27

Zhang, T. Y., & Suen, C. Y. (1984). A fast parallel algorithm for thinning digital patterns. *Communications of the ACM, 27*(3), 236-239. doi:10.1145/357994.358023

KEY TERMS AND DEFINITIONS

Authentication: The acceptance or the rejection of the identity given by a person.

Biometric Systems: Assemble of equipments, hardware and software whose purpose is to authenticate or identify persons automatically. They are based on the capture of several biometric parameters and the comparison of them with the same set of parameters already known of identified persons.

Biometry: The science that procures to obtain numeric value from biological phenomena and the use of such data in the automatic identification or authentication of human beings.

Co-Occurrence Matrix: A statistical matrix use for measuring texture in images. Could be considered an image by itself.

Graphology: It is a pseudoscience that claims its ability to identify the writer, indicate his or her psychological state at the time of writing, or evaluate his or her personality by means of the analysis of the physical characteristics and patterns

of handwriting. The term is sometimes incorrectly used to refer to forensic document examination.

Identification: The determination of the identity of a person, by a biometric system.

Off-Line: Acquisition of biometric parameter from the residues left by the human being actions. In handwriting analysis these residues are: the gray level and the wide of the strokes including the form of the stroke.

On-Line: Acquisition of biometric parameters in real time. In handwriting analysis, the writing instrument, the surface or both have sensors capturing position, trajectory, force and inclination.

Pseudo-Dynamics Characteristics: Parameters acquired off-line used as estimators for on-line data.

Questioned Document Examination: Forensic phrase used to refer documents that are potentially disputed in a court of law. The primary purpose is to provide evidence about a suspicious or questionable document. Evidence might include alterations, chain of possession, damage, forgery, origin and authenticity, among aspects.



Haptics–Based Systems Characteristics, Classification, and Applications

Abeer Bayousuf

King Saud University, Saudi Arabia

Hend S. Al-Khalifa

King Saud University, Saudi Arabia

Abdulmalik Al-Salman

King Saud University, Saudi Arabia

INTRODUCTION

With advances in technology, individuals today can interact with computer technology, not only through their visual and auditory senses, but also through touch. The sense of touch is unlike other senses such as vision and hearing; its uniqueness lies in the fact that it enables a bi-directional exchange of information between users and the virtual or real environment, whereas the other senses, such as sight and hearing, are uni-directional forms of interaction, taking place from the computer application to the user.

A haptics-based system utilizes the tactile sense of touch, the kinesthetic sense, or both types to enable human–computer interaction. The source of the touch under investigation could be animate (e.g. humans), inanimate (e.g. machines), or a combination of both. The items or environments being touched could be real, virtual, or a combination of both (M. Eid, Orozco, & Saddik, 2007), (Saddik, 2007). In addition, a haptics-based system may complement other sensory information such as vision, audition, or both, to make an environment appear more realistic.

Nowadays, haptics-based systems have had a significant impact in a variety of fields, such as surgical simulation, rehabilitation, technology designed to help the blind and visually impaired, education and training, entertainment and leisure activities such as video games, art and design,

online consumer activity, security, and mobile phones. A survey of haptic-based applications can be found in (M. Eid et al., 2007), (Saddik, 2007), (Abdulmotaleb El Saddik, Mauricio Orozco, Mohamad Eid, & Jongeun Cha, 2011) and (Escobar-Castillejos, Noguez, Neri, Magana, & Benes, 2016). Despite this, haptics-based applications are still in their infancy (Saddik, 2007), and their development is ongoing.

In this chapter we present the concept of haptics, its characteristics, classification and applications. Thus, the outline of the chapter is as follows: Section 2 defines the concept of haptics. Section 3 presents the components of haptics-based system architecture in a virtual environment. Section 4 provides the characteristics and classification of haptic devices. Section 5 presents examples of haptics-based systems in different areas. Finally, section 6 concludes the chapter with future research directions.

BACKGROUND: THE CONCEPT OF HAPTICS

The concept of haptics is derived from the Greek verb *haptesthai* which means ‘to touch’. It refers to the science of sensation and manipulation by touch for the purpose of perception or modification of the environment (Abdulmotaleb El Saddik et al., 2011), (M. Eid et al., 2007), (Mihelj, Novak, &

Begus, 2014). Touch is divided into two forms: the tactile (also called cutaneous) sense, and the kinesthetic (also called proprioceptive or force) sense (Abdulmotaleb El Saddik et al., 2011), (Benali-khoudja, Hafez, Alex, & Kheddar, 2004). The tactile sense refers to the registering of physical contact with a real object through skin receptors. It provides the individual with information about an object; this includes texture, pressure, temperature, wetness, softness, friction, vibration, and shape. Conversely, the kinesthetic sense of touch means feeling motion through awareness of position and movement of body parts as well as muscular efforts that are conveyed to the individual by sensory receptors in joints, tendons, and muscles when touching and manipulating items. The kinesthetic sense tells the individual about the movement of joints, velocity, muscle control, and weight. The human hand features both types of touch (Abdulmotaleb El Saddik et al., 2011), (Mihelj et al., 2014).

The term haptics was introduced in the field of psychology in the early 20th century to provide an explanation of the human perception of objects by touch. In the 1970s and 1980s, robotics researchers redefined the term to refer to the remote control of robots and their environment by touch (Abdulmotaleb El Saddik et al., 2011), (M. Eid et al., 2007). The early 1990s were the 'golden age' of haptics, due to the widespread appearance on the consumer market of different haptic devices such as the tactile mouse and the joystick. The field of computer haptics, introduced to integrate haptic

devices with computers (Abdulmotaleb El Saddik et al., 2011), (M. Eid et al., 2007), is associated with the design and development of algorithms and software to generate and display forces and tactile sensations to the user (Abdulmotaleb El Saddik et al., 2011).

HAPTICS-BASED SYSTEM ARCHITECTURE

The components of haptics-based system architecture in a virtual environment consist of the human operator, the haptic device, haptic rendering, the simulation engine, and audio-visual rendering, as shown in Figure 1.

The interaction between a human and a virtual environment in a haptics-based system is easy to understand. It is best described as follows:

Human Operator

The human operator manipulates the haptic device via the touch sense in either a tactile or kinesthetic manner, or both. Most haptics-based systems in a virtual environment interact with individual users through their hands, including fingertips and palms (see Figure 2), or through both hands (see Figure 13). However, the touch sense is distributed all over the body. As a result, Boian et al. have developed haptics-based systems that interact with an individual's foot (Boian, Deutsch, Lee, Burdea, & Lewis, 2003).

Figure 1. Haptics-based system architecture in virtual environments (Salisbury, Conti, & Barbagli, 2004)

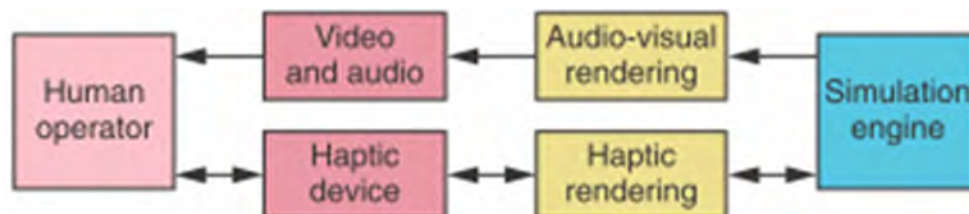


Figure 2.



Haptic Device

Haptic devices are mechanical devices that enable communication between humans and machines by using the sense of touch (Sofronia, Savii, & Davidescu, 2010). A haptic device consists of one or more sensors and actuators. The sensors measure the interactions between a proxy and the environment, whereas the actuator applies the response mechanical action command (Abdulmotaleb El Saddik et al., 2011). Usually, haptic devices include a software toolkit to control the device by programming. Both the hardware and software of a haptic device are called the haptic interface (Robles-De-La-Torre, 2008). Haptic devices are classified according to the type of feedback into two categories: tactile feedback devices and force feedback devices. More information about haptic devices can be found in section 4.

Haptic Rendering

At the heart of a haptic-based application is haptic rendering. This enables a human to make use of a haptic interface in order to touch, feel, and manipulate virtual objects (Cagatay Basdogan & Mandayam A. Srinivasan, 2002). Haptic rendering refers to the group of algorithms that are used to calculate forces/torques and generate tactile or

kinesthetic sensations to signalize interaction between a proxy point inside the virtual environment and the virtual objects within the environment (Abdulmotaleb El Saddik et al., 2011), (Saddik, 2007), (Salisbury et al., 2004).

Haptic rendering is divided into three categories: collision detection algorithms, collision response algorithms, and control algorithms. The collision detection algorithm detects when and where a proxy meets a virtual object (Abdulmotaleb El Saddik et al., 2011), (Salisbury et al., 2004). The collision response algorithm calculates the ideal interaction force/torque for a force device and/or interprets tactile parameters, such as texture, as actuation commands for a tactile device (Abdulmotaleb El Saddik et al., 2011). The ideal reaction force cannot be applied to the user for various reasons, including limitations of the haptic device such as friction, inertia, and finite sensor resolution. In addition, there are limitations to the possible exerted force and direction by the haptic interface. Moreover, the interaction between the human operator and the haptic device is continued in time, whereas the interaction between haptic device and haptic rendering occurs in discrete-time signals. To counteract this, the control algorithm has to minimize the error between ideal and applicable reaction, with the provision of kinesthetic constraints on the

virtual environment and the decrease of inertia in free haptic movement (Abdulmotaleb El Saddik et al., 2011), (Salisbury et al., 2004).

The control algorithm for interaction between haptic device and virtual environment is divided into impedance control and admittance control. In impedance control, the user moves the haptic device, and it sends the haptic data to virtual environment application. In contrast, the users in the admittance control exert force on the haptic device for movement, which reacts as displacement for both input and output haptic data. This type of control provides free haptic device movement and produces movement with more force and stiffness (Escobar-Castillejos et al., 2016), (Mihelj et al., 2014). In the haptics-based system, haptic rendering draws comparisons between a proxy and all virtual objects. If a collision is detected, haptic rendering will calculate the reaction force/torque for kinesthetic sensation and/or actuator commands for tactile sensation. If no collision is detected, no reaction feedback will be calculated. This means that the user has free movement of the haptic device and does not affect any other virtual object. This haptic rendering loop is called the Servo Loop Rate or haptic update rate, and is repeated at specific time intervals such as 1 kHz (Abdulmotaleb El Saddik et al., 2011), (Saddik, 2007), (Salisbury et al., 2004).

There are different types of haptic rendering algorithms; they are classified according to feedback type to: virtual environment model and interaction methods. Haptic rendering produce different kinds of tactile stimuli such as temperature, texture, stiffness and friction. Each stimuli is rendered in a specific manner and presented through a specific display (Mihelj et al., 2014). In contrast, the kinesthetic rendering algorithms are usually different according to the number of Degree Of Freedom (DOF) for haptic device (Salisbury et al., 2004). Moreover, the haptic rendering algorithms are different according to virtual environment model such as rendering for object shape, deformable object, dynamic object, haptic playback, and haptic interaction through the network (Cagatay

Basdogan & Mandayam A. Srinivasan, 2002). Furthermore, the interaction method between haptic device and virtual environment is classified into: point-based, ray-based and 3D object made by a grouping of points, lines and polygon (Cagatay Basdogan & Mandayam A. Sinivasan, 2002), (Escobar-Castillejos et al., 2016).

Simulation Engine

This component computes the behavior of the virtual environment over a period of time. For example, visual rendering detects collisions among all objects in a virtual environment, while haptic rendering detects collisions in the immediate area of the proxy (Abdulmotaleb El Saddik et al., 2011). Moreover, the rendering update rates are different between haptic, visual, and audio rendering. The visual rendering update rate of 60 Hz is enough to display continuous motion on the screen that is visible to the human eye. Also, the audio rendering sample rate for a quality audio signal is 44.1 kHz for each audio frame which contains 10 ms (Taylor, Chandak, Antani, & Manocha, 2009). In contrast, the minimum haptic update rate in force feedback haptic rendering is 1 kHz. If the update rates fall below that value, the haptic device starts fluctuating and vibrating when we make virtual contact with solid surfaces (Saddik, 2007). A noticeable interaction delay is seen between the human operator and the force feedback from a haptic device (Robles-De-La-Torre, 2008). The best response for tactile vibration is higher than 300 Hz (Jeffrey J. Berkley, 2003). One way to solve the problem of the differences between the various rendering update rates is to use multiple threads (Cagatay Basdogan & Mandayam A. Srinivasan, 2002); however, the process of synchronizing different threads can be a challenge.

Audio-Visual Rendering

Audio-visual rendering calculates audio and graphic movement in a virtual environment (Abdulmotaleb El Saddik et al., 2011), (Saddik, 2007),



(Salisbury et al., 2004). Visual rendering (graphic rendering) refers to the generation of an image from a sense file through the use of computer programs with the ability to display tens of times per second. The sense file contains information about objects, texture, lighting, and shading as descriptions of virtual senses. Audio rendering provides realistic sound for interactive applications. It consists of two processes: propagation and rendering. The propagation process computes the path from the sound source to the listener and takes into account reflection and diffraction. The rendering phase generates specialized audio signals (Taylor et al., 2009).

HAPTIC DEVICES CHARACTERISTICS AND CLASSIFICATION

Haptic devices are input and output devices that trace the movements of the user (input) and that produce a touch sensation, either tactile or force (output) (Sofronia et al., 2010). This section provides characteristics and classification of haptic devices along with examples.

Characteristics of Haptic Devices

Each haptic device has different attributes, and these vary widely depending on the specific application of the device. The attributes of haptic devices are described in detail in (Abdulmotaleb El Saddik et al., 2011) and (Sofronia et al., 2010). Table 1 summarizes the characteristics of haptic devices and offers a short description of each characteristic.

Classification of Haptic Devices

Haptic devices are categorized into two groups based on the kind of feedback they provide; the two groups are tactile feedback devices and force feedback devices. These are described as follows:

Tactile Feedback Devices

Tactile feedback devices are designed to reproduce tactile parameters as accurately as possible. The parameters include texture, temperature, shape, roughness, pressure, rigidity, and vibration, among others; these are conveyed by skin receptors. However, no haptic device combines all tactile parameters (Benali-khoudja et al., 2004). Commercially available tactile devices are classified into vibration devices and pin array devices. Vibration tactile devices (also called vibrotactile devices) provide vibration upon touching virtual objects. An example is the CyberTouch by the company CyberGlove Systems, whereby vibrotactile stimulators on a glove produce vibration for the five fingers and the palm of the hand; see Figure 2. The tactile mouse is another example of a vibrotactile device. Its shape is similar to that of a standard mouse, but it produces vibration when the cursor is moved over a Graphical User Interface (GUI); one example is the iFeel mouse from Logitech (Abdulmotaleb El Saddik et al., 2011).

Pin array devices have generally advanced due to a significant amount of research performed in that field. They usually consist of a matrix of pins that differs in actuator technology, numbers of pins, space between pins, frequency, workspace, maximum force, resolution, and motion mode (raised/lower pins). In recent years, a number of pin array devices for blind individuals have been developed, such as VideoTIM3 and Braille Sense U2. VideoTIM3 from Abtim is a tactile device that converts printed text and simple graphics into a raised/lower matrix of pins (see Figure 3). The device consists of a camera held in the fingers and a refreshable matrix of pins to represent text in vibrotactile form (“VideoTIM3: READING DEVICE for the BLIND,” n.d.).

A further use of tactile feedback devices is to generate sensations of temperature and pressure. For example, CM Research, Inc. has developed the Displaced Temperature Sensing System (DTSS) that produces temperature feedback via a thimble (see Figure 4). The device can work as an input or

Table 1. Characteristics of haptic devices

Characteristic	Description
Workspace	The real-world space that can be reached by the end-effector of a haptic device. The workspace is classified into traditional and rotational. The point is represented in a traditional workspace as a Cartesian coordinate; it is represented in a rotational workspace as an angle range in pitch, yaw, and roll.
Degree Of Freedom (DOF)	Describes the amount of possible haptic device motion (right-left, up-down, forward-backward, and rotation) through the number of traditional and rotational variables in the device. For example, 3DOF means the device has three variables and can move right-left, up-down and forward-backward. Each haptic device has two sets of DOF: one for input (position sensing) and one for output (force feedback).
Position resolution	The smallest amount of end-effector movement that the device's sensors can detect. Haptic devices have two kinds of resolution: one for input (position sensing) and one for output (force feedback).
Precision	Accuracy of the haptic device. Haptic devices have two types of precision: one for input (position sensing) and one for output (force feedback).
Repeatability	Accuracy of identifying the physical position as being identical with the virtual position.
Grounding location	The base of the haptic device, either ground-based or body-based.
Maximum exertable force	The maximum output force that can be produced by a haptic device over a very small time interval such as milliseconds.
Minimum exertable force	The minimum output force that can be produced by a haptic device over a very small time interval.
Continuous exertable force	The output force exerted over an extended period.
Dynamic exertable force	Ratio of the maximum and minimum force.
Stiffness	The ability to mimic a rigid object.
Friction/damping	Force of resistance motion either with velocity (called damping or viscous friction) or without velocity (called Coulomb friction).
Backdrivability	Friction in the motors and cable of the haptic device.
Inertia	Mass at end-effector of haptic device.
Weight	The weight of the haptic device.
Interface	Item that connects the device with a computer, such as USB interface.
Bandwidth	Range of frequencies which end-effector provides force feedback.
Haptic refresh rate/ servo loop rate	Speed of haptic rendering from position sensors to force feedback response.
Maximum acceleration	Maximum changes in velocity.
Device latency	Total time from sending a command to the haptic device to receiving a response from the device.
System latency	The total time delay within the haptic virtual environment. IR includes sensing a position, computing the force feedback response, sending feedback to the device, and reading the next position.

Figure 3.



output channel and is operated through a controller unit or through a computer (Abdulmotaleb El Saddik et al., 2011). In addition, the researchers of the Toshiba Nuclear Engineering Laboratory have developed actuators in ring forms on a VPL DataGlove that interacts through pressure from the fingertips (Benali-khoudja et al., 2004).

Force Feedback Devices

Force feedback devices provide force that is conveyed to the human sensory receptors in tendons, joints, and muscles. Force feedback devices have two main functions: first, to determine the position of sensors and/or forces that are exerted by the human body; and second, to apply appropriate force and/or position of end-effectors to the user depending on the limitations of the haptic device.

Force feedback devices are classified according to design into ground-based and body-based devices. With ground-based force feedback devices, the base of the device is placed on the ground or on a desktop, and the user touches a part of the device; examples include joysticks, steering wheels, mice, linkage-based devices, and tension-based devices (also called string-based devices). On the other hand, body-based force feedback devices are worn or attached to any part of the body; examples include gloves, suits, and exoskeletal devices. Force feedback joysticks, mice, and steering wheel are cheap, portable, and are usually used for gaming, but they are not very accurate and allow only for a limited range of movement. For example, the G27 Racing Wheel from Logitech is a force feedback steering wheel haptic device; it uses force feedback to feel roads, bumps, and walls in racing games ("G27 Racing Wheel," n.d.). Logitech WingMan Force 3D joysticks are one example of force feedback joysticks. The device measures the angle or direction of movements and exchanges virtual objects on screen or video games. Also, the WingMan Force 3D joystick produces forces depending on user action, such as shooting, explosions, and acceleration.

Linkage-based devices consist of a robotic arm connected to an end-effector. The arm tracks the location of the end-effector and is capable of exerting a force to the end-effector (Jeffrey J. Berkley, 2003). There are three types of end-effectors in linkage-based devices: a knob, which the user grips by hand; the stylus/pen, which the user holds like a pen; and a thimble, into which the user inserts a finger. Linkage-based devices have various professional applications because they include more controls and haptic features; however, they permit only a single point of contact with the virtual environment. Commercially available linkage-based devices that are available and not limited are the Novint Falcon, Omega devices, Delta devices, Phantom haptic devices, and the Freedom 6S. The Novint Falcon from Novint Technologies has three arms, which extend from the device and which are joined to a knob for user control (see Figure 4).

The Omega 3 and the Delta 6 from Force Dimension are other examples of knob end-effector haptic devices. The design of these devices bears similarities to the Novint Falcon, but they provide high force feedback for a sustained period of time. The Omega 3 provides 3-DOF translation for both position sensing and force feedback. The Omega 3 and the Delta 6 are shown in Figures 5 and 6 respectively.

Examples of commercially available linkage-based force feedback devices with a stylus/pen end-effector are the Phantom Omni and the Omega 6 as shown in Figures 7 and 8 respectively. The Phantom from SensAble Devices, Inc. (now part of the Geomagic company) is a family of pen-based force feedback devices. It has three lines of commercially available haptic devices: the Phantom Omni (recently renamed the Geomagic Touch), the Phantom Desktop (recently renamed the Geomagic Touch X) and the Phantom Premium.

An example of a commercially available thimble-based force feedback device is the Phantom Premium 1.0 from SensAble Technologies, Inc. (now part of Geomagic). It includes a stylus

Figure 4. The Novint Falcon (El Saddik et al., 2011)



Figure 5. The Omega 3 (“Force Dimension - Products,” n.d.)



and thimble to produce forces to one fingertip (see Figure 9).

Tension-based devices consist of cables connected to a point of contact. The device senses the point position through determining the length of each cable. Moreover, tension-based devices produce force feedback on the point through creating tension in the cables. Tension-based devices have large workspaces, less inertia, and more accurate force feedback; they also provide forces to multiple points of contact (Jeffrey J. Berkley, 2003). A few commercially available tension-based devices

exist. For example, SPIDAR-8 was developed to manipulate virtual objects and produce forces via eight fingertips (thumb, index finger, middle finger, and ring finger on both hands) (Sato, 2002) (see Figure 10).

Examples of commercially available body-based force devices are the CyberGrasp and the Interactor vest. The CyberGrasp is an exoskeletal glove from CyberGlove Systems, which produces force feedback independently for each finger and the palm. Its full range motion of the hands includes shoulder, elbow, wrist, and finger (see Figure

Figure 6. The Delta 6 (“Force Dimension - Products,” n.d.)



Figure 7. The Phantom Omni (“Haptic Devices from Geomagic,” n.d.)



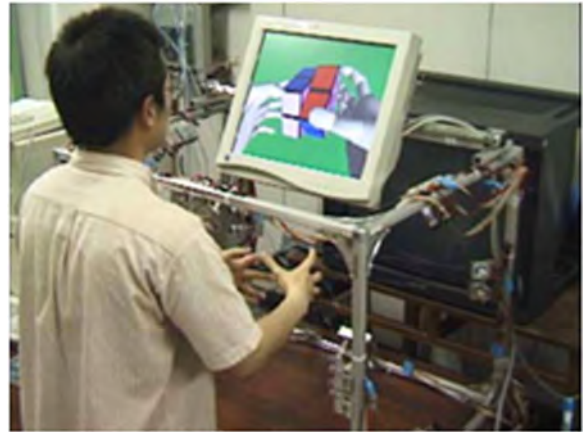
Figure 8. The Omega 6 (“Force Dimension - Products,” n.d.)



Figure 9. The Phantom Premium 1.0 (“Haptic Devices from Geomagic,” n.d.)



Figure 10. SPIDAR-8 (Sato, 2002)



11). The Interactor vest from Aura systems, Inc. is a rigid plastic backpack called a vest, which is worn over the upper torso. The device converts sound signals into forces and includes speakers with five levels of volume. The force feedback is increased according to the volume level. The interactor vest is plugged into the computer, TV, CD stereo, or video game (Doug Horton, 2004) (see Figure 12).

Figure 11. CyberGrasp (“cyberglovesystems products,” n.d.)



APPLICATIONS

There are different haptics-based systems developed for different areas. In this section, we present examples of haptics-based systems (see Table 2) that were proposed for medical simulation, rehabilitation, education and training, security, shopping, web application, collaborative environment, handwriting and emotion fields.

Figure 12. Interactor vest (Horton, 2004)



Table 2. Summary of haptics-based systems

Application Area	Description	Reference
Medical simulation	This approach enables a haptic palpation interaction for virtual reality-based medical simulation. The Palpation is a physical examination technique where objects are touched with fingers to determine their size, shape, consistency and location.	(Ullrich & Kuhlen, 2012)
rehabilitation	The system combines haptic, virtual reality and game technologies in order to assist patients in the training of the upper limb movement. The system provides visual, audio and force feedback.	(Covarrubias, Mansutti, Bordegoni, & Cugini, 2014)
education and training	Visual Haptic Jawi (HV-JAWI) is a handwriting training system designed for teaching Jawi cursive handwriting to primary school students. The application supports three different modes of handwriting training: fully guided, partially guided, and copy mode. In addition, the system supports assessment for student handwriting through online Jawi handwriting recognition and visualization of the student trajectory.	(Amin, Zaman, & Ahmad, 2013)
security	The system proposed incorporation of four haptic features in a haptic enabled authentication procedure, including stiffness, static friction, magnetic, and pop-through.	(Potocny, McNulty, Maiga, & Zadeh, 2015)
web application	The researchers proposed a framework to include haptic interaction in HTML based Web applications. Each HTML element provides specific type of force feedback. In addition, it supports independent haptic device	(Comai, Mazza, & Guarinoni, 2012)
collaborative environment	The haptic researchers have developed Collaborative Haptic Audio-Visual Environments (C-HAVE) that allow multi-users, who each has a haptic interface, to manipulate shared objects in the environment, either collaboratively or remotely.	(Abdulmoteleb El Saddik et al., 2011)
emotion	Proposed a system that produces a haptic mirror for visually impaired. The system captures facial expressions directly from a webcam and presented through haptic interface.	(Yasmin, McDaniel, & Panchanathan, 2014)

FUTURE RESEARCH DIRECTIONS

Recently, contactless haptic devices are a new trend in the development of haptic devices. They produce tactile feedback without any physical contact with the user. The contactless haptic devices depend on different technology such as air-jet, acoustic radiation pressure and laser (Arafsha, Zhang, Dong, & Saddik, 2015).

Furthermore, to develop more complex haptic applications, haptic multimedia was introduced; it is defined as “the acquisition of spatial, temporal, and physical knowledge of the environment through the human touch sensory system and the integration and/or coordination of this knowledge with other sensory displays such as audio, video, and text in a multimedia system” (M. Eid et al., 2007). Multimedia haptic applications are also referred to as Haptic Audio-Visual Environment (HAVE) (Saddik, 2007). Moreover, the term of

tele-haptics was introduced for transmission haptic sense over network (Shen & Shirmohammadi, 2008). Recently, the field of affective haptics was introduced for capture and display emotions through the sense of touch; it is defines as “the acquisition of human emotions through the human touch sensory system, the processing of emotion related haptic data to detect affect, and the display of emotional reactions via haptic interfaces. Emotions may be solely communicated through the sense of touch or coordinated/integrated with other sensory displays (such as audition or vision) in a multimedia system” (M. A. Eid & Osman, 2016).

Finally, the research trend in the development of haptics-based systems is the focus on haptic interaction for mobile device such as smartphones and tablets. The haptic interactions for mobile devices are not limited to simple tactile vibration, but also the proposed kinesthetic feedback can potentially be incorporated into mobile devices

(Kwon, Yang, & Cho, 2010). For example, researchers have proposed a haptic stylus to access visual images from smartphones. The haptic stylus generates force and tactile feedback (Tian, Song, & Chen, 2016).

REFERENCES

- G27 Racing Wheel. (n.d.). Retrieved July 13, 2016, from http://support.logitech.com/en_us/product/g27-racing-wheel
- Amin, M. M., Zaman, H. B., & Ahmad, A. (2013). Visual haptic approach complements learning process of Jawi handwriting skills. In *2013 5th International Conference on Information and Communication Technology for the Muslim World (ICT4M)* (pp. 1–6). <http://doi.org/doi:10.1109/ICT4M.2013.6518888>
- Arafsha, F., Zhang, L., Dong, H., & Saddik, A. E. (2015). Contactless haptic feedback: state of the art. In *2015 IEEE International Symposium on Haptic, Audio and Visual Environments and Games (HAVE)* (pp. 1–6). <http://doi.org/doi:10.1109/HAVE.2015.7359447>
- Basdogan, C., De, S., Kim, J., Muniyandi, M., Kim, H., & Srinivasan, M. A. (2004). Haptics in minimally invasive surgical simulation and training. *IEEE Computer Graphics and Applications*, 24(2), 56–64. doi:10.1109/MCG.2004.1274062 PMID:15387229
- Basdogan, C., & Srinivasan, M. A. (2002). Haptic Rendering in Virtual Environments. In *Handbook of virtual environments: design, implementation, and applications*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Benali-khoudja, M., Hafez, M., Alex, J., & Kheddar, A. (2004). *Tactile Interfaces: a State-of-the-Art Survey*. Academic Press.
- Berkley. (2003, May 5). *Haptic Devices*. Mimic Technologies Inc. Retrieved from <http://www.learningace.com/doc/2617500/e4571e9f-d42a964d14262d2c0edf99a4/white-paper-haptic-devices>
- Boian, R. F., Deutsch, J. E., Lee, C. S., Burdea, G. C., & Lewis, J. (2003). *Haptic effects for virtual reality-based post-stroke rehabilitation*. Presented at the 11th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems. <http://doi.org/doi:10.1109/HAPTIC.2003.1191289>
- Burdea, G. C., & Coiffet, P. (2003). output devices: graphics, three-dimensional sound, and haptic displays. In *Virtual Reality Technology*. John Wiley & Sons.
- Covarrubias, M., Mansutti, A., Bordegoni, M., & Cugini, U. (2014). Interacting Game and Haptic System Based on Point-Based Approach for Assisting Patients after Stroke. In K. Miesenberger, D. Fels, D. Archambault, P. Peñáz, & W. Zagler (Eds.), *Computers Helping People with Special Needs* (pp. 289–296). Springer International Publishing. Retrieved from http://link.springer.com/chapter/10.1007/978-3-319-08596-8_46
- Cyberglovesystems Products. (n.d.). Retrieved July 20, 2016, from <http://cyberglovesystems.com/all-products>
- Eid, M., Orozco, M., & Saddik, A. E. (2007). A guided tour in haptic audio visual environments and applications. *Int. J. Adv. Media Commun.*, 1(3), 265–297. doi:10.1504/IJAMC.2007.013918
- Eid, M. A., & Osman, H. A. (2016). Affective Haptics: Current Research and Future Directions. *IEEE Access*, 4, 26–40. doi:10.1109/ACCESS.2015.2497316
- El Saddik, A., Orozco, M., Eid, M., & Cha, J. (2011). *Haptics Technologies: Bringing Touch to Multimedia*. Springer. doi:10.1007/978-3-642-22658-8

- Escobar-Castillejos, D., Noguez, J., Neri, L., Magana, A., & Benes, B. (2016). A Review of Simulators with Haptic Devices for Medical Training. *Journal of Medical Systems*, 40(4), 1–22. doi:10.1007/s10916-016-0459-8 PMID:26888655
- Force Dimension - Products. (n.d.). Retrieved July 17, 2016, from <http://www.forcedimension.com/products>
- Haptic Devices from Geomagic. (n.d.). Retrieved July 12, 2016, from <http://www.geomagic.com/en/products-landing-pages/haptic>
- Horton, D. (2004). *AVSIM Commercial Hardware*. Retrieved April 18, 2013, from http://www.avsim.com/pages/0604/aura/aura_interactor.htm
- Kwon, D. S., Yang, T. H., & Cho, J. (2010). Trend & prospects of haptic technology in mobile devices. In *2010 IEEE International Symposium on Industrial Electronics* (pp. 3778–3783). <http://doi.org/doi:10.1109/ISIE.2010.5637683>
- Mihelj, M., Novak, D., & Begus, S. (2014). Haptic Modality in Virtual Reality. In *Virtual Reality Technology and Applications* (pp. 161–194). Springer Netherlands. doi:10.1007/978-94-007-6910-6_7
- Potocny, J., McNulty, S., Maiga, K., & Zadeh, M. H. (2015). On the Incorporation of Haptic Effects in Security Authentication. In *2015 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (pp. 469–473). <http://doi.org/doi:10.1109/SMC.2015.93>
- Robles-De-La-Torre, G. (2008). Principles of haptic perception in virtual environments. In M. Grunwald (Ed.), *Human Haptic Perception: Basics and Applications* (pp. 363–379). Birkhäuser Basel. doi:10.1007/978-3-7643-7612-3_30
- Saddik, E. (2007). The Potential of Haptics Technologies. *IEEE Instrumentation & Measurement Magazine*, 10(1), 10–17. doi:10.1109/MIM.2007.339540
- Salisbury, K., Conti, F., & Barbagli, F. (2004). Haptic rendering: Introductory concepts. *IEEE Computer Graphics and Applications*, 24(2), 24–32. doi:10.1109/MCG.2004.1274058 PMID:15387225
- Sato, M. (2002). *Development of String-based Force Display: SPIDAR*. VSM2002 (The Eighth International Conference on Virtual Systems and Multi Media), Gyeongju, Korea.
- Shen, X., & Shirmohammadi, S. (2008). Tele-Haptics. In B. Furht (Ed.), *Encyclopedia of Multimedia* (pp. 842–848). Springer, US. doi:10.1007/978-0-387-78414-4_227
- Sofronia, R. E., Savii, G., & Davidescu, A. (2010). *Haptic devices in engineering and medicine*. Presented at the 2010 International Joint Conference on Computational Cybernetics and Technical Informatics (ICCC-CONTI). <http://doi.org/doi:10.1109/ICCCYB.2010.5491247>
- Taylor, M. T., Chandak, A., Antani, L., & Manocha, D. (2009). RESound: interactive sound rendering for dynamic virtual environments. In *Proceedings of the 17th ACM international conference on Multimedia* (pp. 271–280). New York, NY: ACM. <http://doi.org/doi:10.1145/1631272.1631311>
- Tian, L., Song, A., & Chen, D. (2016). A Novel Haptic Stylus for Mobile Terminal. In F. Bello, H. Kajimoto, & Y. Visell (Eds.), *Haptics: Perception, Devices, Control, and Applications* (pp. 338–349). Springer International Publishing. doi:10.1007/978-3-319-42321-0_31
- Ullrich, S., & Kuhlen, T. (2012). Haptic Palpation for Medical Simulation in Virtual Environments. *IEEE Transactions on Visualization and Computer Graphics*, 18(4), 617–625. doi:10.1109/TVCG.2012.46 PMID:22402689
- VideoTIM3. READING DEVICE for the BLIND. (n.d.). Retrieved July 19, 2016, from http://www.abtim.de/home__e_/home__e_.html

What is Haptics? (2010). Retrieved July 15, 2016, from http://neilforrest.zxq.net/what_is_haptics.php

Yasmin, S., McDaniel, T., & Panchanathan, S. (2014). A Haptic-Based Application for Active Exploration of Facial Expressions by the Visually Impaired. In G. Bebis, R. Boyle, B. Parvin, D. Koracin, R. McMahan, J. Jerald, ... M. Carlson (Eds.), *Advances in Visual Computing* (pp. 357–366). Springer International Publishing. Retrieved from http://link.springer.com/chapter/10.1007/978-3-319-14364-4_34

Yu, W., & Brewster, S. (2002). Comparing Two Haptic Interfaces for Multimodal Graph Rendering. In *Proceedings of the 10th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems* (p. 3). Washington, DC: IEEE Computer Society. Retrieved from <http://dl.acm.org/citation.cfm?id=795682.797511>

KEY TERMS AND DEFINITIONS



Human–Computer Interaction (HCI): The study of how user interacts with computers.

Haptics-Based System: A software that uses touch sense for interacting with the user.

Haptic Device: A machine that uses touch sense as input, output or both.

Haptic Rendering: A group of algorithms that are used to calculate and generate tactile or kinesthetic sensations.

Kinesthetic Sense: Information about touch sense through motion of joints, tendons, and muscles during touch object motion.

Tactile Sense: Information about touch sense through a physical contact between object and surface of the body. This information includes texture, pressure, temperature, wetness, softness, friction, vibration, and shape.

Virtual Environment: A software that simulates the real world or imagination.

The Holon/Parton Structure of the Meme, or The Unit of Culture

J. T. Velikovsky

University of Newcastle, Australia

INTRODUCTION

This chapter identifies the formal structure of *the unit of culture*,¹ a unit also known as the *meme* (Dawkins, 1976) as: the *holon/parton* - as a resolution of a longstanding open problem in both Information Science and Information Technology² and across all domains of culture including the media, the arts, entertainment and science. The holon/parton is a fractal hierarchical structure, the formation of which is governed by the laws of physics as proposed in (Wilson 1998). Previously there was no universal definition of the unit of culture, resulting in more than three hundred varying definitions of culture (Baldwin et al 2006) and no consensus (van Peer et al 2007).

This chapter thus presents *a universal structure for the unit of culture* (the meme) in order to facilitate commensurate empirical analyses of culture (and its component parts, or units) across disciplines - and to enable structural, functional and behavioral comparisons between biocultural artifacts (memes). This allows for the accurate tracking and understanding of the transmission of culture including within the domains of Information Science and Information Technology. The structure of the unit of culture - viewed as the *holon-parton* - is a conceptual, theoretical, practical and scientific tool for identifying and analyzing units of culture (as parts and also as wholes) within all symbol systems in culture.³

BACKGROUND TO THE PROBLEM

Knowledge (i.e., retained information, or culture) emerges from creative⁴ biocultural human systems at an exponential rate. As two examples: the number of published scientific papers doubles every nine years (Bornmann & Mutz, 2014) and the number of new books published each year (approximately 700,000 per year in the U.S.) doubled in the five years from 2008 to 2012 (Bowker, 2013). The question emerges: What structural rules (laws) govern the formation of these units of culture (i.e., research papers, chapters, books, phrases, words) - or ideas, processes and products?

Knowledge is retained information, and is considered 'culture'. However 'culture' has been a problematic term as there is currently no consensus across all of domains and disciplines on the term. In 1952, 164 definitions of culture were extant (Kroeber & Kluckhohn, 1952) and in 2006 the list had been extended to over 300 definitions (see Baldwin, Faulkner, Hecht, & Lindsley, 2006, pp. 139-226). The article 'Evolution of Culture, Memetics' in the *Encyclopedia of Complexity and Systems Science* states 'The lack of a universally accepted meme definition and the vagueness of meme boundaries... indeed make empirical studies less evident' (Heylighen & Chielens, 2009, p. 3217).

The following identification of the structure of the unit of culture (the meme) - namely the *holon/*

parton - provides a solution to the problem of defining and thus analyzing culture and therefore has universal (multi-, cross- and trans-disciplinary) applications in allowing for comparative empirical studies of culture (or knowledge, or memes).

THEORETICAL PERSPECTIVE

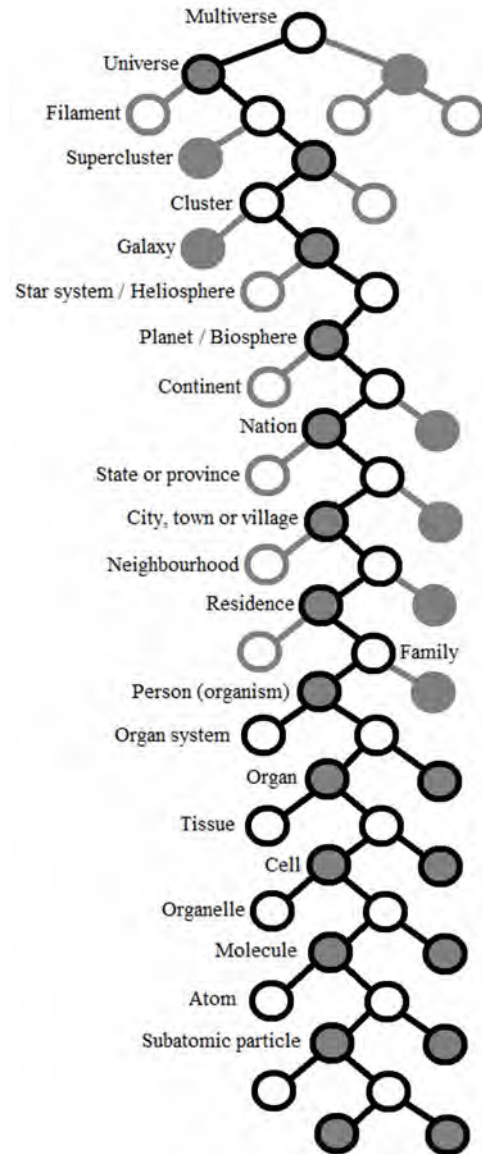
The theoretical perspective adopted in this chapter is Evolutionary Systems Theory.⁵ *Introduction to Systems Philosophy: Toward a New Paradigm of Contemporary Thought*, Laszlo (1972) incorporates Living Systems Theory (Miller, 1968)⁶ and Bunge's hierarchical structures (Bunge, 1969) to provide a framework for understanding universal structures spanning from subatomic physics, through biology, chemistry, organisms, and social systems to the cosmos (Laszlo, 1972, pp. 29, 177-180). An illustration of this view is shown in Figure 1.

In this view the largest-to-smallest levels of structural hierarchical organization are all governed by three laws of physics, namely three laws of holarchies (see below). Laszlo (1972) employs Koestler's holon theory in systems (Laszlo, 1972, pp. 55-118, 252, 255, 272-114), as Koestler had previously extended General Systems Theory (Koestler, [1978] 1979, pp. 31-32; von Bertalanffy, 1950) from biological systems to social holarchies (i.e., hierarchies of social holon/partons) and to linguistic hierarchies (Koestler, [1964] 1989, pp. 287-290; [1967] 1989, pp. 103, 198-109; [1978] 1979, pp. 27-62). As noted in the *Key Terms and Definitions* section (below) in simplest terms the *three laws of holarchies* (three key behaviors of holon/partons) are as follows.

As *units*, holon/partons:

- Compete, and/or co-operate (and/or engage in *co-opetition*) 'sideways' – namely with other holon/partons on the same level;
- Integrate 'upwards' – namely into the larger holon/parton 'above'; and
- Control and command the holon/partons on the level 'below' – namely their own component holon/partons

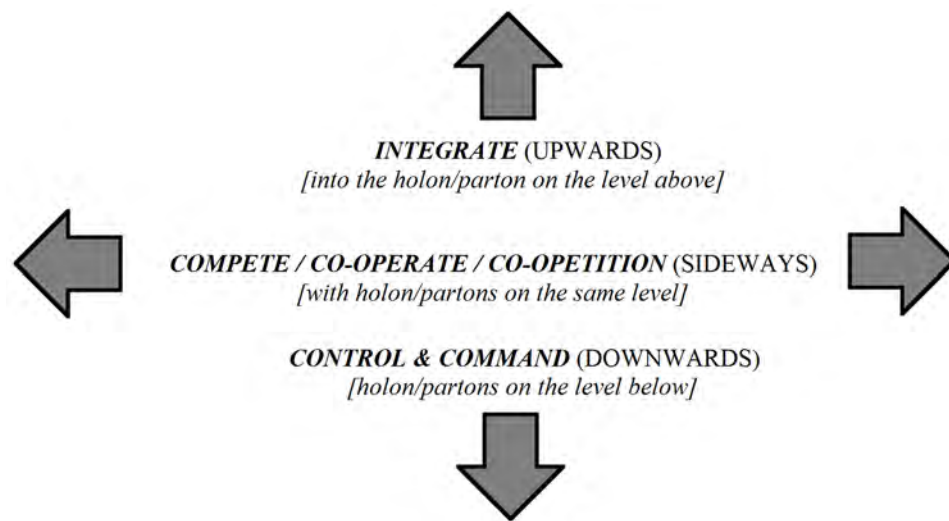
Figure 1. The Evolutionary Systems View, from the multiverse through to subatomic particles (derived from Laszlo 1972); image copyright IGI Global (Velikovsky 2016, p. 211)



This basic concept of 'the three laws of holarchies' is illustrated in Figure 2.

These three laws are laws of physics (and of systems) applying both to biology and also as seen here in bioculture (aka 'culture') a possibility suggested in *Consilience: The Unity of Knowledge* (Wilson, [1998] 1999, pp. 60, 291, 293).

Figure 2. The three laws of holon/partons, or, holarchies (derived from Koestler 1964-1978); image copyright IGI Global (Velikovsky 2016, p. 212)



THE DEFINITION OF THE UNIT OF CULTURE AS: THE HOLON/PARTON

The definition of the unit of culture (the meme) is thus: the holon/parton. The holon/parton *structure* of the unit of culture (meme; i.e. idea, and/or process, and/or product) can be viewed in Figure 3.

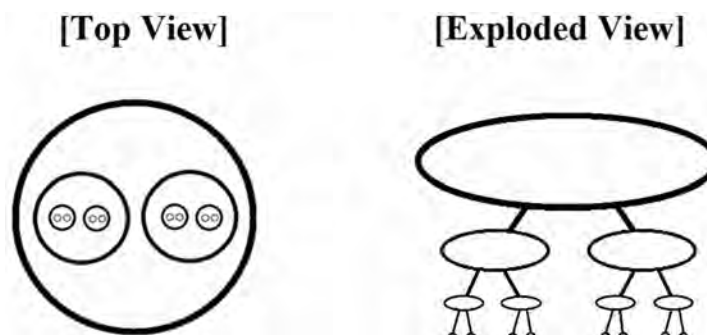
This structure is fractal (self-similar on smaller scales). This is not to imply that *all* units of culture can (nor should be) be subdivided into exactly *two* parts infinitely; the number of components on each lower (smaller) levels of the holarchy may increase exponentially.⁷ As one example: in a single novel there may be around 100,000 words;

in a 90-minute movie (at 24 frames per second) around 130,000 still-image frames.⁸

THE HOLON/PARTON AS THE UNIT OF CULTURE IN VARIOUS COMMUNICATION MEDIA

With the above holon/parton schema of all (bio) culture in mind, below are hierarchical templates for various communication media as memes (units of culture) and their component units (on the levels below) each considered as holon/partons.

Figure 3. The holon/parton structure of the unit of culture - or the meme (idea, process, product); image copyright IGI Global (Velikovsky 2016, p. 217)





The three laws of holarchies apply to the units on each level; namely each of these whole units: (1) integrate *upwards*; (2) compete and co-operate *sideways*, and (3) control and command, *downwards*.

The Novel

- All extant novels
- All libraries, and all retail outlets for novels
- Library, or retail outlet
- All novel genres
- Specific Genre
- Specific Novel
- Chapter
- Paragraph
- Sentence
- Word
- Letter (and punctuation mark, including spaces)
- Idea

As an analysis of the laws governing these holon/partons: individual *novels* compete with other individual novels for reader attention, thus for canonical status, and for both library and retail shelf-space in the field (Boyd, 2009; Van Peer, 1996, 1997; Moretti 2000); as do transmedia stories, literary genres, and sub-genres; the specific words, sentences, paragraphs and chapters (and also any images) that have survived the creation and editing process are those retained in the published work; in the completed work they also co-operate (or, *operate together*) to attract the attention of; to convey information and meaning to; and, possibly to evoke emotional or intellectual engagement (the 'flow' state, or Narrative Transportation) in the reader, i.e. the audience, or field (Boyd, Carroll, & Gottschall, 2010; Velikovsky 2014a).

Movie

- All extant movies
- All movie genres

- Specific Genre
- Specific Movie
- Act
- Sequence
- Scene
- Shot
- Action, and/or Dialogue (e.g. line of dialogue)
- (Dramatic) beat
- Single (still image) frame
- Idea

As with the 'Novels' example above each unit (or holon/parton) on the same holarchic level of a movie *integrates* 'upwards'; *competes/co-operates* 'sideways'; and also *controls* (i.e., provides organizing principles namely sets of 'rules' for) the units on the level 'below' it.

Videogame

- All extant videogames
- All videogame genres
- Specific Genre
- Specific Videogame
- Game Level
- Game Environment
- Building
- Room
- Puzzle/Hazard
- Game event
- Idea

Theatre (Play)

- All plays
- All play genres
- Specific Genre
- Specific Play
- Act
- Scene
- (Dramatic) Beat
- Moment
- Idea

Popular Song

- All popular songs
- All popular song genres
- Specific Genre
- Specific Song
- Intro / Verses / Chorus / Bridge / Solo / Coda (or commensurate, or alternate parts)
- Word / Chord
- Letter / Musical note/ Beat
- Idea

Poem

- All poems
- All poetic genres
- Specific Genre
- Specific Poem
- Stanza
- Word
- Letter
- Idea

Joke

- All jokes
- All joke genres
- Specific Genre
- Specific Joke
- Setup / Feed / Punchline (or commensurate or alternate parts; e.g., one-liner)
- Word
- Idea

Spoken Language

- Phrase / Sentence
- Word
- Morpheme
- Phoneme

Science

- All science
- Domain of science

- Scientific discipline
- Scientific paradigm
- Metatheory
- Theory (including: Laws, Assumptions, Initial Conditions)
- Objective Phenomena
- Fact
- Idea

Religion

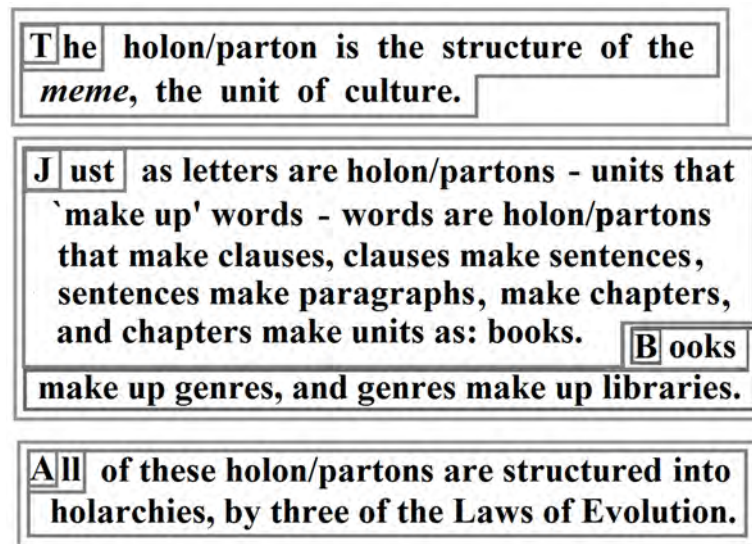
- All religions
- Specific Religion
- Specific Faith
- Metaphysics
- Scripture
- Idea

Bioculture is composed of extrasomatic symbol systems (e.g., written and spoken language, mathematics, musical notation, paintings, drawings, and so on). This holon/parton (and *holarchy*) structural tendency of symbolic biocultural artifacts as outlined above has been hinted at by prior theorists including Herbert Simon (1996):

Symbolic Systems... systems of human symbolic production. A book is a hierarchy in the sense in which I am using that term. It is generally divided into chapters, the chapters into sections, the sections into paragraphs, the paragraphs into sentences, the sentences into clauses and phrases, the clauses and phrases into words. We may take the words as our elementary units, or further subdivide them, as the linguist often does, into smaller units. If the book is narrative in character, it may divide into "episodes" instead of sections, but divisions there will be. The hierarchic structure of music, based on such units as movements, parts, themes, phrases, is well known. (Simon, 1996, pp. 187-188)

This holon/parton structure Simon (1996) indicates (without labelling it as such) can be viewed as in Figure 4.

Figure 4. A diagrammatic analysis of the holon/parton structure of three paragraphs of text; image copyright IGI Global (Velikovsky 2016, p. 220)



Evolutionary psychologist Steven Pinker states: (with 'chunks' here considered as units)

Several clauses are joined or embedded in a sentence; several sentences make up a paragraph; several paragraphs make up a section, several sections a chapter, several chapters a book. A text with this hierarchical structure is easy for a reader to assimilate because at any level of granularity, from clauses to chapters, the passage can be represented in the reader's mind as a single chunk... (Pinker 2014, p. 142)

THE HOLON/PARTON AS THE UNIT OF STORY (THE NARREME)

The unit of story (unit of narrative) has been discussed by Barthes (1966), (Dorfman, 1969) and many other researchers however it is clear that the narreme was not satisfactorily defined by Barthes in 1966 nor since (Baikadi & Cardona-Rivera, 2012). Stories (narratives) are part of bioculture. Since stories are composed of ideas, words, phrases, (and/or cinema shots, images, sounds) and so on, the unit of culture (the meme

- as holon/parton) can thus be applied to the *unit of story* (the narreme).

Boyd (2009) notes narrative plots are structured focused on agents (characters) with an: intention, action, obstacle/s, and outcome. A story (or narrative), and its component parts, namely *scenes* (both considered as holon/partons) display a fractal structure - namely both the story (the *whole*) and the scene (the *part*) include the characteristics of: intention, action, obstacle, and outcome.

ISSUES, CONTROVERSIES, AND PROBLEMS

The Complexity perspective and the use of Systems Theory is no longer seen as controversial in the Social Sciences given 'The Complexity Turn' of the mid-1990s⁹ however the integration of the Systems Model of Creativity and the Evolutionary Systems Theory perspective has yet to become widespread in Mass Communication Studies (see also: Fulton & McIntyre, 2013).

Memetics has been in a (Kuhnian) pre-paradigm state since Dawkins (1976) as the universal structure of the unit of culture (the meme) has

not previously been identified; however *the holon/parton* presents a solution to this problem.¹⁰ While not every single retained cultural artifact in existence adheres to the holon/parton structure when analyzed, canonical cultural artifacts (as opposed to ‘archive’ *sensu* Moretti 2000) do so.

FUTURE RESEARCH DIRECTIONS

Future research directions and applications of the structure of the unit of culture (the meme) in Information Science and Information Technology include empirical investigations on the number and the type (quantitative and qualitative analyses) of units of culture (as holon/partons) that compose those canonical memes (i.e., the ideas, processes and products that are retained and reproduced in culture) compared to those in the non-canonical (archive) units which are de-selected by the field (audience) for that domain of culture.

The 2011 article ‘Can We Measure Memes?’ (McNamara, 2011) asks if fMRI technology can measure memes in the brain, proposing the distinction of *internally* and *externally* represented memes (‘i-memes’ and ‘e-memes’). The article concludes that if memes (units of culture) *can* be accurately defined and if cognitive science can thus measure memes this would be a significant advance with far-reaching implications in cultural evolution for many domains and disciplines well beyond mere questions around how language works (McNamara 2011, online).

CONCLUSION

This chapter identifies the unit of culture aka the meme (an idea, a process or a product) as the holon/parton (and thus the holarchy). This Evolutionary Systems Theory view of bioculture adopts an Applied Evolutionary Epistemology approach (Gontier 2012) and aims to provide a solution of the problem of the definition of the

term ‘culture’ and thus a theory to enable accurate empirical transdisciplinary investigations of bioculture across all communication media including the arts, entertainment and sciences - including Information Science and Information Technology.

This holon/parton view of units of culture supports E. O. Wilson’s hypothesis in *Consilience* (Wilson, 1998) that laws of physics (laws of Evolution and Systems Theory - namely the three laws of holarchies) apply to the formation and growth of bioculture, illuminating the unit of culture (the meme) where ideas, processes and products correlate respectively with: Worlds Three, Two, and One in Sir Karl Popper’s ‘Three Worlds’ theory (Popper, 1978b; Boyd 2016).

REFERENCES

- Altmann, G., & Koch, W. A. (1998). *Systems: New Paradigms for the Human Sciences*. Berlin: W. de Gruyter. doi:10.1515/9783110801194
- Baikadi, A., & Cardona-Rivera, R. E. (2012). Towards finding the fundamental unit of narrative: A Proposal for the Narreme. *Proceedings of the 2012 Workshop on Computational Models of Narrative (CMN2012)*, 44-46.
- Baldwin, J. R., Faulkner, S. L., Hecht, M. L., & Lindsley, S. L. (Eds.). (2006). *Redefining Culture: Perspectives Across the Disciplines*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Barthes, R., & Duisit, L. (1966). An Introduction to the Structural Analysis of Narrative. *New Literary History*, 6(2), 237-272. doi:10.2307/468419
- Blaikie, N. W. H. (2007). *Approaches to Social Enquiry* (2nd ed.). Cambridge, UK: Polity Press.
- Bornmann, L., & Mutz, R. (2014). Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *Journal of the Association for Information Science and Technology*. doi: arXiv:1402.4578v3 [cs.DL]

- Bowker. (2013). *Self-Publishing In the United States, 2008-2013*. New Providence, NJ: Bowker (A ProQuest Affiliate).
- Boyd, B. (2009). *On The Origin Of Stories: Evolution, Cognition, and Fiction*. Cambridge, MA: Belknap Press of Harvard University Press.
- Boyd, B. (2010). Art and Evolution: The Avant-Garde as Test Case—Spiegelman in The Narrative Corpse. In B. Boyd, J. Carroll, & J. Gottschall (Eds.), *Evolution, Literature and Film: A Reader*. New York: Columbia University Press.
- Boyd, B. (2016). Poppers World 3: Origins, Progress, and Import. *Philosophy of the Social Sciences*, 46(3), 1–21. doi:10.1177/0048393116640282
- Boyd, B., Carroll, J., & Gottschall, J. (2010). *Evolution, Literature and Film: A Reader*. New York: Columbia University Press.
- Bradie, M., & Harms, W. (2012). Evolutionary Epistemology. *The Stanford Encyclopedia of Philosophy*. Retrieved from <http://plato.stanford.edu/archives/win2012/entries/epistemology-evolutionary>
- Bunge, M. (1969). The Metaphysics, Epistemology and Methodology of Levels. In L. Whyte, A. Wilson, & D. Wilson (Eds.), *Hierarchical Structures*. New York: Elsevier.
- Buss, D. M. (2012). *Evolutionary Psychology: The New Science of the Mind* (4th ed.). Boston: Pearson Allyn & Bacon.
- Campbell, D. T. (1960). Blind Variation and Selective Retention in Creative Thought as in Other Knowledge Processes. *Psychological Review*, 67(6), 380–400. doi:10.1037/h0040373 PMID:13690223
- Campbell, D. T. (1974). Evolutionary Epistemology. In P. A. Schlipp (Ed.), *The Philosophy of Karl Popper* (Vol. 1, pp. 413–459). La Salle.
- Capra, F., & Luisi, P. L. (2014). *The Systems View of Life: A Unifying Vision*. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511895555
- Chaisson, E. (2001). *Cosmic Evolution: The Rise of Complexity in Nature*. Cambridge, MA: Harvard University Press.
- Csikszentmihalyi, M. (1988). Society, Culture, and Person: A Systems View of Creativity. In R. J. Sternberg (Ed.), *The Nature of Creativity* (pp. 325–339). New York: Cambridge University Press.
- Csikszentmihalyi, M. (1996). *Creativity: Flow and the Psychology of Discovery and Invention* (1st ed.). New York: HarperCollins.
- Csikszentmihalyi, M. (2014). The Systems Model of Creativity and Its Applications. In D. K. Simonton (Ed.), *The Wiley Handbook of Genius*. Chichester, UK: John Wiley & Sons Ltd. doi:10.1002/9781118367377.ch25
- Csikszentmihalyi, M., & Nakamura, J. (2006). Creativity Through the Life Span from An Evolutionary Systems Perspective. In C. Hoare (Ed.), *Handbook of Adult Development and Learning* (pp. 243–254). New York: Oxford University Press.
- Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection, or The Preservation of Favoured Races in the Struggle For Life*. London: J. Murray.
- Dawkins, R. (1976). *The Selfish Gene*. Oxford, UK: Oxford University Press.
- Dena, C. (2009). *Transmedia Practice: Theorising The Practice Of Expressing A Fictional World Across Distinct Media And Environments* (PhD Thesis). Department of Media and Communications, University of Sydney. Retrieved from <http://www.christydena.com/phd/>
- Dennett, D. C. (1995). *Darwin's Dangerous Idea: Evolution and the Meanings of Life*. New York: Simon & Schuster.

- Dorfman, E. (1969). *The Narreme in the Medieval Romance Epic: An Introduction to Narrative Structure*. Toronto: University of Toronto Press.
- Feynman, R. P. (2005). *Don't You Have Time To Think?* London: Allen Lane (Penguin).
- Fulton, J., & McIntyre, P. (2013). Futures of Communication: Communication Studies ~ Creativity. *Review of Communication*, 13(4), 269–289. doi: 10.1080/15358593.2013.872805
- Gershenson, C., & Heylighen, F. (2005). How Can We Think Complex? In K. A. Richardson (Ed.), *Managing Organizational Complexity: Philosophy, Theory and Application* (pp. 47-62). Greenwich: IAP - Information Age Pub. Inc.
- Gontier, N. (2012). Applied Evolutionary Epistemology: A new methodology to enhance interdisciplinary research between the life and human sciences. *Journal of Philosophy and Science*, 4, 7-49.
- Gontier, N. (2014). Evolutionary Epistemology. *The Internet Encyclopedia of Philosophy*. Retrieved 7th April, 2015, from <http://www.iep.utm.edu/evo-epis/>
- Habash, G. (2012). *Average Book Length: Guess How Many Words Are In A Novel*. Retrieved from http://www.huffingtonpost.com/2012/03/09/book-length_n_1334636.html
- Helmenstine, A. M. (2014). *How Many Atoms Are There in a Human Cell?* Retrieved from <http://chemistry.about.com/od/biochemistry/f/How-Many-Atoms-Are-There-In-A-Human-Cell.htm>
- Heylighen, F., & Chielens, K. (2009). Evolution of Culture, Memetics. In R. A. Meyers (Ed.), *Encyclopedia of Complexity and Systems Science* (pp. 3205–3220). Larkspur, CA: Springer. doi:10.1007/978-0-387-30440-3_189
- Kerrigan, S., & Velikovsky, J. T. (2015). Examining Documentary Transmedia Narratives Through 'The Living History of Fort Scratchley' Project. *Convergence*. doi: 10.1177/1354856514567053
- Koestler, A. (1979). *Janus: A Summing Up*. London: Pan Books.
- Koestler, A. (1989). *The Act of Creation*. London: Arkana. (Original work published 1964)
- Koestler, A. (1989). *The Ghost in the Machine*. London: Arkana. (Original work published 1967)
- Kroeber, A. L., & Kluckhohn, C. (1952). *Culture: A Critical Review of Concepts and Definitions*. Cambridge, MA: The Museum.
- Laszlo, E. (1972). *Introduction to Systems Philosophy: Toward a New Paradigm of Contemporary Thought*. New York: Gordon and Breach.
- Lin, Y., Duan, X., Zhao, C., & Xu, L. (2013). *Systems Science: Methodological Approaches*. Boca Raton, FL: CRC Press.
- Macdonald, I. W. (2013). *Screenwriting Poetics and the Screen Idea*. New York: Palgrave Macmillan. doi:10.1057/9780230392298
- McIntyre, P. (2012). *Creativity and Cultural Production: Issues for Media Practice*. Basingstoke, UK: Palgrave Macmillan. doi:10.1057/9780230358614
- McNamara, A. (2011). Can we measure memes? *Frontiers in Evolutionary Neuroscience*, 3(1). doi:10.3389/fnevo.2011.00001 PMID:21720531
- Miller, J. G. (1968). Living Systems: Basic Concepts. In W. Gray, F. J. Duhl, & N. D. Rizzo (Eds.), *General Systems Theory and Psychiatry*. Boston: Little, Brown & Co.
- Miller, J. G. (1978). *Living Systems*. New York: McGraw-Hill.
- Mobus, G. E., & Kalton, M. C. (2014). *Principles of Systems Science*. New York: Springer.
- Moretti, F. (2000). The Slaughterhouse of Literature. *MLQ: Modern Language Quarterly*, 61(1), 207–227. doi:10.1215/00267929-61-1-207

- Peer, W. V., Hakemulder, J., & Zyngier, S. (2012). *Scientific Methods for the Humanities*. Amsterdam: John Benjamins Pub. Co. doi:10.1075/lal.13
- PGA. (2010). *Producers Guild of America Code of Credits*. Retrieved from http://www.producers-guild.org/?page=coc_nm#transmedia
- Pinker, S. (2014). *The Sense of Style: The Thinking Person's Guide to Writing in the 21st Century*. New York: Penguin.
- Polson, D., Cook, A.-M., Brackin, A., & Velikovsky, J. T. (Eds.). (2014). *Transmedia Practice: A Collective Approach*. London: ID-Press.
- Popper, K. R. (1963). *Conjectures and Refutations: The Growth of Scientific Knowledge. (Essays and Lectures.)*. London: Routledge & Kegan Paul.
- Popper, K. R. (1978a). Natural Selection and the Emergence of Mind. *Dialectica*, 32(3-4), 339–355. doi:10.1111/j.1746-8361.1978.tb01321.x
- Popper, K. R. (1978b). *Three Worlds*. Salt Lake City, UT: University of Utah.
- Popper, K. R. (1979). *Objective Knowledge: An Evolutionary Approach* (Rev. ed.). Oxford, UK: Oxford University Press.
- Popper, K. R. (1999). *All Life is Problem Solving*. London: Routledge.
- Richards, C. E. (2012). Complementarities in Physics and Psychology. In L. J. Miller (Ed.), *The Oxford Handbook of Psychology and Spirituality*. Oxford, UK: Oxford University Press. doi:10.1093/oxfordhb/9780199729920.013.0005
- Runco, M. A., & Jaeger, G. J. (2012). The Standard Definition of Creativity. *Creativity Research Journal*, 24(1), 92–96. doi:10.1080/10400419.2012.650092
- Sadowski, P. (1999). *Systems Theory as an Approach to the Study of Literature: Origins and Functions of Literature*. Lewiston, NY: E. Melan Press.
- Sawyer, R. K. (2012). *Explaining Creativity: The Science of Human Innovation* (2nd ed.). New York: Oxford University Press.
- Simon, H. A. (1991). *Models of My Life*. New York: Basic Books.
- Simon, H. A. (1996). *The Sciences of the Artificial* (3rd ed.). Cambridge, MA: MIT Press.
- Simonton, D. K. (2012a). Creative Thought as Blind Variation and Selective Retention: Why Creativity is Inversely Related to Sightedness. *Journal of Theoretical and Philosophical Psychology*, 33(4), 253–266. doi:10.1037/a0030705
- Simonton, D. K. (2012b). Fields, Domains, and Individuals (Chapter). In M. D. Mumford (Ed.), *Handbook of Organizational Creativity* (pp. 67–86). Oxford, UK: Elsevier Science. doi:10.1016/B978-0-12-374714-3.00004-5
- Van Peer, W. (1996). Canon Formation: Ideology or Aesthetic Quality? *The British Journal of Aesthetics*, 36(2), 97–108. doi:10.1093/bjaesthetics/36.2.97
- Van Peer, W. (1997). Two Laws of Literary Canon: Growth and Predictability in Canon Formation. *Mosaic: A Journal for the Interdisciplinary Study of Literature*, 30(2), 113–132.
- Van Peer, W., Hakemulder, J., & Zyngier, S. (2007). *Muses and Measures: Empirical Research Methods for the Humanities*. Newcastle, UK: Cambridge Scholars.
- Velikovsky, J. T. (2013). *The Holon/Parton Structure of the Meme, the Unit of Culture*. Retrieved from <http://storyality.wordpress.com/2013/12/12/storyality-100-the-holonic-structure-of-the-meme-the-unit-of-culture/>
- Velikovsky, J. T. (2014a). *Flow Theory, Evolution & Creativity: or, 'Fun & Games'*. Paper presented at the Interactive Entertainment 2014 (IE2014), Newcastle, Australia. doi:10.1145/2677758.2677770

Velikovsky, J. T. (2016). The Holon/Parton Theory of the Unit of Culture (or the Meme, and Narreme) in Science, Media, Entertainment and the Arts. In A. Connor & S. Marks (Eds.), *Creative Technologies for Multidisciplinary Applications*. New York: IGI Global. doi:10.4018/978-1-5225-0016-2.ch009

von Bertalanffy, L. (1950). An Outline of General System Theory. *The British Journal for the Philosophy of Science*, 1(2), 134–165. doi:10.1093/bjps/1.2.134

Warfield, J. N. (2006). *An Introduction to Systems Science*. Hackensack, NJ: World Scientific. doi:10.1142/6058

West, G. B., & Brown, J. H. (2005). The origin of allometric scaling laws in biology from genomes to ecosystems: Towards a quantitative unifying theory of biological structure and organization. *The Journal of Experimental Biology*, 208(9), 1575–1592. doi:10.1242/jeb.01589 PMID:15855389

Wilson, E. O. (1999). *Consilience: The Unity of Knowledge*. New York: Knopf: Random House.

Zimmer, C. (2011). 100 Trillion Connections: New Efforts Probe and Map the Brain's Detailed Architecture. *Scientific American*. Retrieved from <http://www.scientificamerican.com/article/100-trillion-connections/>

KEY TERMS AND DEFINITIONS

Creativity: The standard definition of creativity is ‘original and useful’ namely a biocultural artifact (i.e., idea, process, product) judged ‘original and useful’ (‘novel and appropriate’) by a consensus of the audience (i.e. the field) for a specific domain in culture (Runco & Jaeger, 2012). Examples of domains in culture include movies, literature, popular music, videogames, astronomy, physics, chemistry, biology, psychol-

ogy, sociology, anthropology and all their various subdomains (including genres and subgenres).

Culture: ‘Culture’ is knowledge encoded in symbol systems in the form of ideas, processes and products (Csikszentmihalyi, 1996) including in science, the arts, media, religion, and languages. Examples of ideas include scientific theories (e.g., gravity, general and special relativity, evolution) and also literary and media characters and narratives (Sherlock Holmes, Anna Karenina, Harry Potter); examples of processes include cognitive processes: writing an email, having a conversation, reading a book and in science, pasteurization; examples of cultural products include the hand-axe, the spear, the wheel, words, books, movies, songs, the printing press, the telescope, the microscope, the car, the computer and iPod. As there is no culture without biology a synonym for ‘culture’ is ‘bioculture’.

Evolution (Biological, Cultural, Cosmic): Evolution can most simply be defined as ‘descent with modification’ (Chaisson, 2001; Darwin, 1859). Here the concept of evolution includes the (post-1940s) Modern Evolutionary Synthesis (aka ‘Neo-Darwinism’) including Multilevel Selection Theory (Gontier, 2012). ‘The evolutionary process requires variation, differential survival and reproductive success, and inheritance’ (Capra & Luisi, 2014, p. 200; see also Dennett, 1995, p. 343). Laszlo (1972) explains atoms, biology and social systems as classes of natural (and sometimes designed) systems ‘sharing invariant fundamental properties’ (Laszlo, 1972, p. 56) and demonstrates that evolution occurs across natural systems (Laszlo, 1972, pp. 57-117).¹¹ Systems and artifacts (inputs and outputs) are selected (or deselected and thus ‘falsified’) by their environment (Popper 1999). The iterative and recursive Evolutionary Algorithm (namely: variation, selection, transmission) satisfies rather than optimizes.¹²

Holarchy: Koestler defines a holarchy as a hierarchy of self-regulating holons (Koestler, 1967, p. 103). Holarchies function according to the three

laws of holarchies - which are also three laws of evolution in Systems Theory (see Laszlo, 1972, pp. 55-118, 176-180) - namely: (1) competition and/or cooperation (and/or co-opetition), 'sideways' with other holon/partons on the same level; (2) integration upwards, into the larger holon/parton on the level 'above'; and (3) control and command of holon/partons on the level 'below' (Koestler, 1964, 1967, 1978).

Holon: (sensu Koestler 1967) 'A holon ... is something that is simultaneously a whole and a part. and a part' (Koestler 1967, p. 48).

Holon/Parton: A portmanteau synthesis of the terms holon and parton. A consilient (science and the arts) synthesis of these two terms used herein is 'holon/parton' as this term emphasizes the dual or 'Janus-faced' (Koestler, 1979, p. 27) nature of these whole/part entities; they are a whole and also a part at the same time in both biological and in socio-cultural systems and also in biological and biocultural units.

Meme: The unit of culture (Dawkins 1976). Identified (thus defined) here as: the holon/parton (see: Velikovsky 2016).

Parton: (sensu Feynman 1975, 2005) The term in subatomic physics that Richard P. Feynman (Feynman, 2005, p. 278) created for the equivalent structure of: the holon.

Systems Theory: The study of systems and their environments - including system inputs, outputs, relations and interactions including the concept of emergence where the whole system (its structure, function and behavior) is more than the sum of its parts. For reviews of Systems Theory, Complexity and Emergence in biology and culture see (Sepänen in Altmann & Koch, 1998, pp. 180-302), (Sadowski, 1999, pp. 7-10), (Gershenson & Heylighen. (2005). (Warfield, 2006, pp. 38-46), (Lin, Duan, Zhao, & Xu, 2013, pp. 1-22). Mobus & Kalton, 2014, 32-40.

Transmedia: A narrative extended across more than one media (e.g., across a movie, novel, videogame, nonfiction documentary and so on). Adaptations of a narrative from one media to another (e.g. novel-to-film) and prequels and sequels

in the same media are not transmedia (see: Dena, 2009; Jenkins, 2011; Polson, Cook, Brackin, & Velikovsky, 2014; Kerrigan & Velikovsky, 2015).

ENDNOTES

- ¹ See (Velikovsky, 2013; Velikovsky, 2016).
- ² As Khosrow-Pour notes in the *Dictionary of Information Science and Technology*: 'the primary focus of IST as a discipline is broadening the science of information processing, management, and dissemination' (Khosrow-Pour 2007, p. v).
- ³ This structure of the unit of culture synthesizes concepts from (Koestler 1967) and Feynman (2005).
- ⁴ See the standard definition of creativity (Runco & Jaeger, 2012).
- ⁵ See the (evolutionary) DPFi systems model of creativity (Csikszentmihalyi 1988, 1996, 2014) and Csikszentmihalyi, M., & Nakamura, J. (2006). See also, Simonton (2012a, 2012b), Sawyer ([2006] 2012, p. 123, on the DPFi 'sociocultural' model of creativity), and see McIntyre (2012) and also Macdonald (2013 p. 123) on the same Domain, Person, Field interaction (DPFi, or DIFi) systems model of creativity.
- ⁶ See also the subsequent updated edition, *Living Systems* (Miller, 1978).
- ⁷ In human biology for example there can be around 10^{14} units when different levels (and units) emerge in the holarchy: there are around 10^{14} atoms in cell (Helmenstine, 2014), around 10^{14} cells in a human body (Richards, 2012, p. 70) and around 10^{14} synapses (connections between neurons) in the human brain (Zimmer, 2011). See also (West & Brown, 2005).
- ⁸ *Amazon.com* notes that the 'average' book on Amazon has 64,500 words (Habash, 2012).
- ⁹ See (Blaikie, 2007, pp. 206-214).
- ¹⁰ One reason for the abandoning of the *Journal of Memetics* (JoM, 1997-2005) and dispersal

of the nascent field of Memetics is that: the structure of the meme (the unit of culture) was not previously identified. Dennett noted in 1995 that ‘The prospects for elaborating a rigorous science of memetics are doubtful’ (Dennett, 1995, p. 369). Various scholars in Memetics (e.g., Blackmore, Farncombe, Dennett et al.) are persisting in solving this hard problem (see Hull, 1982 in Dennett,

1995, p. 352); see also for example (Velikovsky, 2014a, 2014b).

¹¹ Herbert Simon (1996) distinguishes between natural and artificial systems (Simon, 1996).

¹² Simon (1991) rightly notes: ‘natural selection only predicts that survivors will be fit enough, that is, fitter than their losing competitors; it postulates satisficing, not optimizing’ (Simon, 1991, p. 166).

ICT Standardization



Kai Jakobs

RWTH Aachen University, Germany

INTRODUCTION

According to the definition adopted by the International Organization for Standardization (ISO) a standard is a document, “established by consensus and approved by a recognized body, that provides, for common and repeated use, rules, guidelines or characteristics for activities or their results, aimed at the achievement of the optimum degree of order in a given context” [ISO, 2015]. The fact that standards are established ‘by consensus’¹ distinguishes them from legislation. Typically, the use of standards is voluntary. However, through legislation they may become mandatory (e.g. many health and safety standards) or ‘quasi-mandatory’ (e.g. Harmonized European Standards).

Standards – in a very general sense – have been with humankind for quite some time. About 4,000 years ago the first alphabets emerged, enabling new forms of communication and information storage. Around the 7th century BC the Lydians invented the first coin-based currency; it established the basis for easier inter-regional trading. The advent of the railroad in the 19th century resulted in a need for technical standards, e.g. those that enabled compatibility between individual parts of technical artifacts, defining e.g. the width of railway gauges, the diameter of screws, etc. This was once more reinforced when mass production generated a demand for interchangeable parts. In parallel, the invention of the electric telegraph in 1837 triggered the development of standards in the field of electrical communication technology. In 1865, the International Telegraph Union – to become the International Telecommunication Union² (ITU) in 1932 – was founded by twenty nation states. The other major international standards setting bodies, the International Electrotechnical

Commission (IEC) and ISO, were founded in 1906 and 1947, respectively.

In the field of Information and Communication Technologies (ICT) international standards are the major mechanism to ensure interoperability between systems. Frequently, ICT standards also describe a commonly agreed platform upon which innovations can be based. Moreover, standards in general are a valuable means of technology transfer. They have also been used as policy tools – for example, they are a major pillar of the European Single Market. Standards’ potential economic implications must also not be under-estimated. A new standard may be used to extend a market, or even help open up a new one. Intellectual Property Rights (IPR), mostly patents, also play a major role here. That is, standards must also be considered as strategic tools. The above suggests that standards are not just technical documents but that they may have ramifications well beyond technology. Accordingly, they should no longer be considered as pure ‘public goods’, i.e. as something that is non-rival and non-excludable [Deneulin & Townsend, 2007]. Rather, these days standards are typically seen as impure public goods, or club goods; they are non-rival and excludable, just like e.g. satellite TV, cinemas or private parks (see also e.g. [Hawkins, 2009]). This is largely due to the frequent incorporation of IPR into standards.

Against this background, this chapter will first briefly look at the links between standards and standardization, innovation and economics. It will then offer a brief description of a typical standards setting process. The complex ‘web’ of Standards Setting Organizations in the ICT sector will be discussed next. Subsequently, a flexible tool to describe the characteristics of an SSO will be discussed. It can be deployed by firms to

DOI: 10.4018/978-1-5225-2255-3.ch406

identify the SSO that is best suited for a planned standardization activity. This selection is only part of the fairly complex task of standardization management, which will be described next. Finally, the chapter will briefly discuss national standardization strategies.

BACKGROUND

Not so long ago standardization and innovation were considered as almost mutually exclusive (see e.g. [Hemenway, 1975], reported in [Farrell & Saloner, 1985]). This has changed by now. In fact, close links between standardization and innovation may often be identified. Today, standardization is no longer considered an impediment to innovation. However, the unqualified claim that ‘standards foster innovation’ does not fully reflect reality either. Swann & Lambert [2010] observe that standards do both – enable and constrain innovation – but that the enabling aspect is much more important. Specifically, they note that “... *standardization does constrain activities but in doing so creates an infrastructure to help trade and subsequent innovation. Standardization is not just about limiting variety by defining norms for given technologies in given markets. Standardization helps to achieve credibility, focus and critical mass in markets for new technologies*” (p. 370).

That is, especially in the field of ICT many standards describe a commonly agreed platform upon which innovations can be based and marketed. Accordingly, standards may be, and indeed are, used as strategic tools. For example, a new standard can extend a market, or even help open up a whole new one (just think what GSM did for mobile communication). On the other hand, backing and subsequently being locked into a ‘wrong’ technology (i.e. one that does not get standardized) may well ruin at least smaller companies.

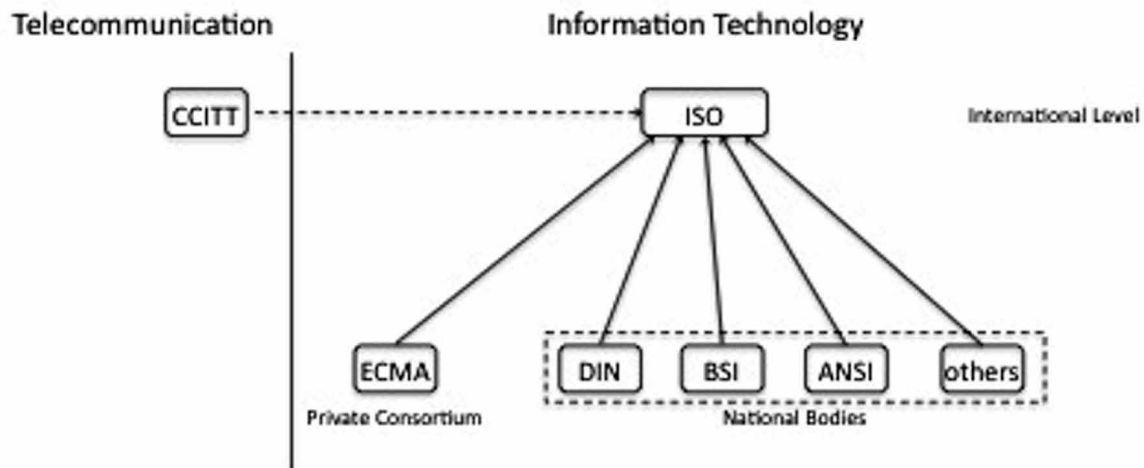
What’s more, “*Standards are not only technical questions. They determine the technology that will implement the Information Society, and consequently the way in which industry, users, consumers and administrations will benefit from it*” [CEC, 1996; p. 1]. That is, those that develop ICT standards today at the same time shape much of the ICT environment we all will use in the future. And if they do their job properly, i.e. if they develop standards that meet the needs of all stakeholders (including e.g. individuals, communities, businesses, and governments) society at large stands to benefit.

From a macro-economic perspective, DIN [2004] finds that “*standardization contributes to GDP growth at the rate of about one percentage point per annum*”. This is in line with a number of similar studies from other countries. Blind [2013] reports that the contribution of standards to the growth rate in different countries was equivalent to 0.9% in Germany, 0.8% in France and Australia, 0.3% in the UK and 0.2% in Canada. Adopting a micro-economic perspective, ISO [2011, 2012] did a number of studies in companies from different business and countries. Based on Porter’s value chain [Porter & Kramer, 2011] these studies show that the implementation of standards can provide economic benefits to firms between 0.5% and 4% of their annual revenues.

By now, participation in standards setting has become a major strategic tool for many firms. Frequently, a firm’s aim will be to influence the process for its own benefit. This holds particularly – though by no means exclusively – for large companies. Other potential motivations include e.g. networking, intelligence gathering and increasing credibility (see e.g. [Jakobs, 2013]).

In addition, standards are a valuable means of technology transfer, which makes them also relevant also for academia and research institutions. Standardization also provides a legal platform for co-operation with competitors and (potential) customers.

Figure 1. The ICT Standardization Universe in the 1970s (adapted from Jakobs & Kritzner, 2009)



SETTING ICT STANDARDS

The ICT ‘Standardization Universe’

Back in the 1970s, the standardization universe was still rather simple and well-structured. Telecommunications standards (called ‘Recommendations’) were developed by the Consultative Committee for International Telephony and Telegraphy (CCITT; the predecessor to ITU-T), a then quasi-monopoly UN organization. This sector was almost completely separate from IT standardization, which was lead by ISO and its national member organizations. That is, standards were almost exclusively set by ‘formal’ Standards Developing Organizations (SDOs; i.e. e.g. ISO, CCITT, ITU-T). The one private consortium in existence was the European Computer Manufacturers Association (ECMA). Figure 1 depicts this situation.

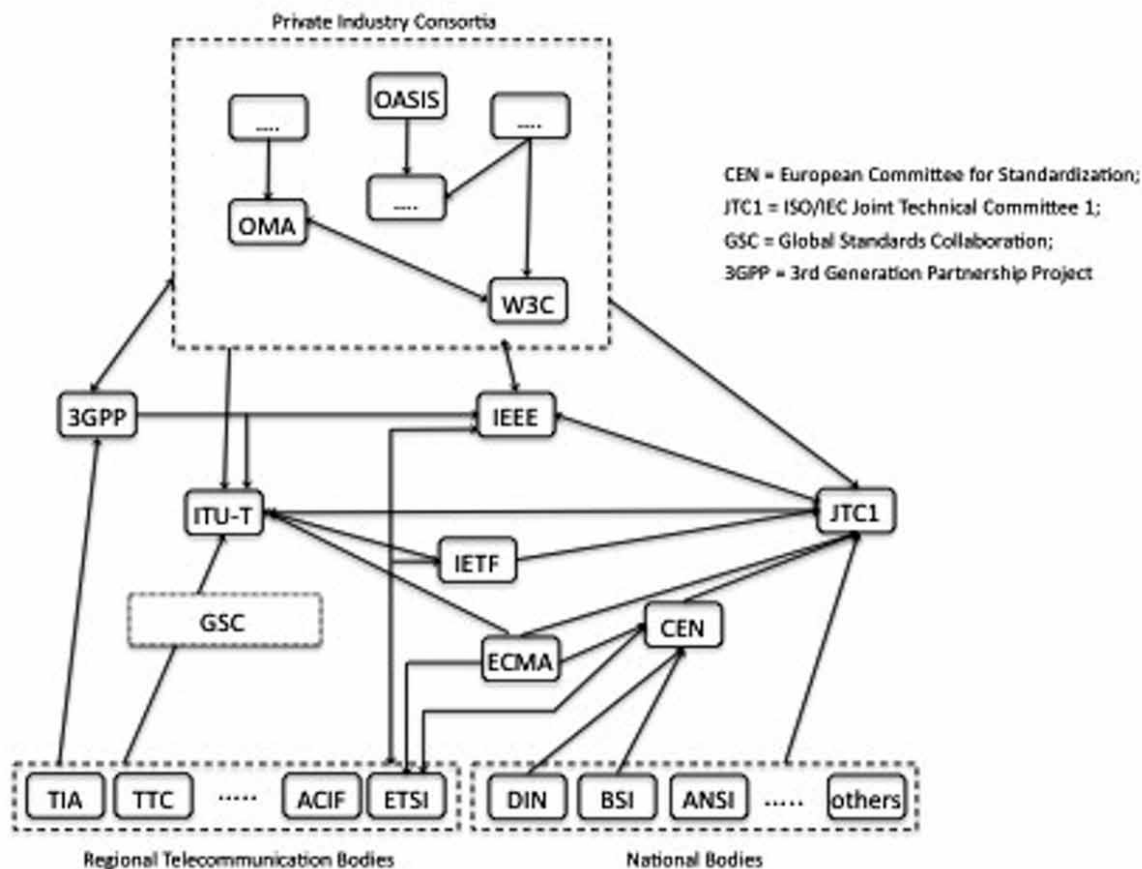
These days, an extensive web of Standards Setting Organizations³ (SSOs) are active at the international (e.g. ISO and ITU), the regional (e.g. the European Telecommunications Standards Institute, ETSI) and the national level (e.g. the American National Standards Institute, ANSI; see Figure 2). Their output is frequently referred to as ‘de-jure’ standards, although these standards do

not have any regulatory power per-se. In parallel, a plethora of private industry fora and consortia (a couple of hundred), such as the World Wide Web Consortium (W3C) and the Open Mobile Alliance (OMA), produce what is often termed ‘de-facto’ standards. The processes through which both types of standards are developed are typically very similar (see below). The required level of consensus may differ, though. It should be noted that both ‘de-facto’ and ‘de-jure’ standards may gain some regulatory power, e.g. through references in legal documents.

Figure 2 also shows that the former clear-cut distinction between Information Technology on the one hand and the Telecommunication sector on the other has largely disappeared. This reflects the situation on the market, where the two sectors are hardly distinguishable any more (virtually every computer can communicate and every smart phone is also a small computer).

This highly complex structure has led to a considerable degree of fragmentation. Frequently, several SSOs are active in similar or overlapping domains. As a result, there may well be competition either generally between SSOs that cover similar ground or temporarily between different SSOs’ Working Groups working on similar projects. Consequently, there is an urgent need

Figure 2. The ICT Standardization Universe today (excerpt; adapted from Jakobs & Kritzner, 2009)



for co-ordination. Given the diverse nature of the individual SSOs (see also below) this may be extremely difficult to achieve.

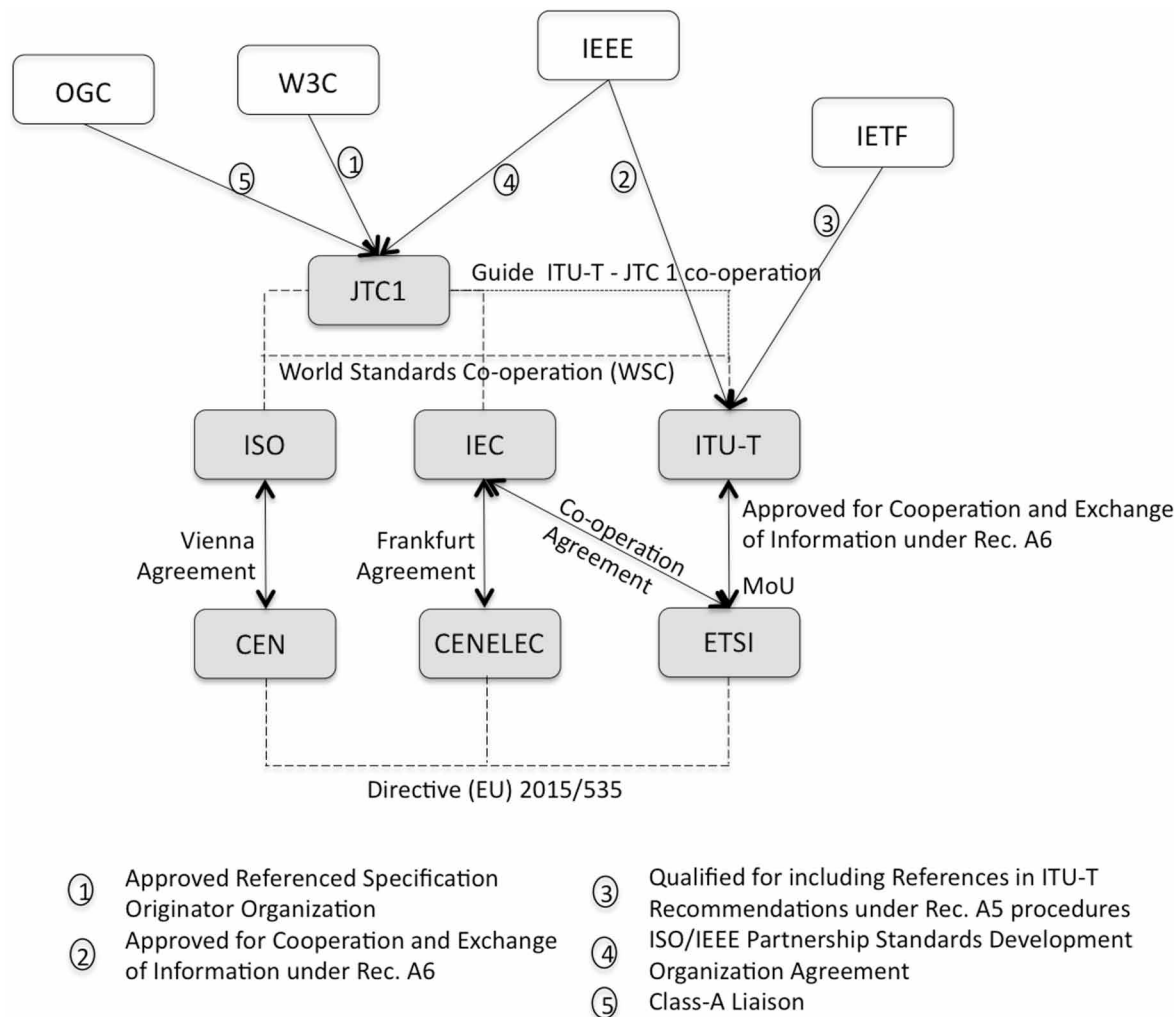
Nevertheless, a comparably high level of co-ordination exists between ISO and the regional and national SDOs, to avoid inconsistencies and contradictions between international, regional and national standards. This coordination is based on the WTO's agreement on 'Technical Barriers to Trade' [WTO, 2015]. In a nutshell, it states that national bodies shall use international standards and remove contradicting national ones whenever possible. Figure 3 shows different types of bilateral agreements.

Following a different approach, the Global Standards Collaboration (GSC; see Figure 2) provides for the regular exchange of work programs and other information between its members, the

regional Telecommunication standards bodies and the ITU. ETSI Partnership Projects represent a related approach to co-ordination. These projects coordinate groups of regional SDOs working towards a common goal. 3GPP⁴, which is active in the field of wireless communication, is a prominent example.

Coordination of the activities of consortia working in overlapping fields is not always desired; some consortia were primarily established to compete with others. Where this is not the case, cooperation is mostly based on different types of bilateral liaison agreements. These agreements are, in most cases, either rather generic (e.g. a general Memorandum of Understanding) or very specific (e.g. relating to one specific standard). Overall, cooperation between consortia is not nearly as pronounced as between SDOs.

Figure 3. Different types of bi-lateral agreements between SSOs (adapted from Jakobs, 2013)



The Standards Setting Organizations (SSOs)

The characteristics of individual SSOs may differ considerably, in various respects. This holds for both SDOs and consortia. Therefore, the traditional distinction between these groups is not overly helpful. A more flexible and meaningful approach was introduced by Updegrove [2006]. It is based on a set of attributes that can be applied to describe an SSO. These attributes fall into four categories:

- General
- Membership

- Standards setting process
- Output

An attribute is a tuple <type, value>. Table 1 summarizes the defined types and provides a brief description for each; the values need to be determined individually for each SSO.

An SSO's IPR policy is probably the single most important criterion. In principle, each party that contributes to the work of an SSO on a specific standard is required not only to disclose any IPR that is relevant for this standard, but also to license it. This holds particularly (though not exclusively) for any 'Standards Essential Patents' (SEPs). If a

Table 1. SSOs' attributes and descriptions (taken from Jakobs & Kritzner, 2009)

General	Description
Internal	
overall goals business models	These provide some general background information.
governance	'Governance' is important for those who wish to influence the goals and activities.
finances staffing	'Finances' and 'staffing' are important to evaluate the chances of an SSB's long-term survival.
IPR policy	IPR policy should not deter any stakeholders.
External	
reputation	Joining an SSB with a poor reputation, or with a poor record, will not make too much sense in many cases.
competition	Do competing SSBs exist, or are they likely to emerge in the near future?
liaisons	Is the SSB prepared to co-operate with other entities?
Membership	Description
Quantitative: • overall # of members • membership levels • membership fees	Provides some information about the relevance of the SSB. In many cases, some are more equal than others. Who decides, who does the work, and are these the same? May also help to determine a prospective new member's chances of active contribution and decision.
Qualitative: • key players involved? • 'active' members • individuals' capacity	Are those organizations important in the specific sector on board? Which are missing? Which members are prepared to invest resources into activities like e.g. editing documents, leading WGs, etc.? Is there a 'corporate whip', or do reps decide independently? What are the SSB's rules?
Standards Setting Processes	Description
Timing: • time for TC/WG establishment, • average time until finalization	Indicates the effort/time required to finalize a new standards project. Helps to determine whether or not a (perceived) window of opportunity can be met.
Process characteristics: • openness • transparency • required level of consensus • observation of due process • decision mechanism	Basically to determine how fair the process is? Do I have a say, or do the big guys determine? Can I complain, and to whom? Etc.
Beyond standardisation: • implementations required? • proof of interoperability required • conformance testing	Indicate a more 'praxis-oriented' approach; crucial if products/services are to hit the market pretty soon.
Output	Description
Types of 'products'	Do an SSB's deliverables suit my needs? I do not always need a full blown standard, and may be prepared to swap consensus for speed.
Quantitative aspects: • standards output • # of implementations	Do they actually produce standards that are widely implemented? Indicates importance of output; if nobody implements the standards participation is hardly worth the effort.
Dissemination: • PAS submitter? • specifications for free?	Related to 'liaisons' above. Is an SSB prepared to bring its specifications into the public domain? And are they prepared to do it for free?
Follow-up activities: • standards maintenance • impact considered? • consistency checks in place?	How does an SSBs react to changes in technology/ requirements? And to errors/inaccuracies? Do they reflect on what they are doing? Are standards checked for consistency with others?

Table 2. Project stages and associated documents in the ISO/IEC process (taken from ISO, 2015).

Project Stage	Associated Document	
	Name	Abbreviation
Preliminary stage	Preliminary Work Item	PWI
Proposal stage	New Work Item Proposal ^a	NP
Preparatory stage	Working Draft(s) ^a	WD
Committee stage	Committee Draft(s) ^a	CD
Enquiry stage	Enquiry Draft ^b	ISO/DISIEC/CDV
Approval stage	Final Draft International Standard ^c	FDIS
Publication stage	International Standard	ISO, IEC or ISO/IEC
a	These stages may be omitted.	
b	Draft International Standard in ISO, committee draft for vote in IEC.	
c	May be omitted.	

standard cannot be implemented without infringing a patent this patent is called an SEP. Under the ‘Royalty Free’ (RF) regime licenses for such patents have to be granted for free⁵. Under a ‘(Fair), Reasonable and Non-Discriminatory’ (F)RAND regime the IPR holder may ask for licensing fees⁶. The SSO does not normally interfere with the negotiations between the IPR holder and the potential licensee(s). No generally accepted definition or algorithm for the calculation of FRAND exists and the courts are frequently called upon to settle disputes (see e.g. [Updegrove, 2013]).

The Standardization Process

In general, the standardization process should make sure that its outcome (i.e. the standards) meets the needs of as many groups of stakeholders as possible. To come at least close to this goal, the World Trade Organization (WTO) has defined a set of “... *principles and procedures that should be observed, when international standards ... are elaborated ...*” [WTO, 2015]. Specifically, these include transparency, openness, impartiality and consensus, effectiveness and relevance, coherence and to address the concerns of developing countries. Most major SSOs more or less observe these principles although formally they are only relevant for national and international SDOs.

Terminology, procedural and structural details may differ⁷, but all major SSOs have adopted a standardization process that is subdivided into several stages. Table 2 exemplary summarizes the stages of the ISO/IEC process. Once a new Work Item has been agreed upon, a Working Group (WG; either an existing one or a newly created one) is charged with the development of the associated standard. When the WG has agreed upon a draft specification the members of the parent Technical Committee or Subcommittee (TC/SC) need to reach consensus. Depending on the number and nature of comments received the specification may be returned to the originating WG for further elaboration. Eventually, a vote will be cast at TC/SC level. If the approval criteria (in terms of majority and number of negative votes) are not met a revised version will be produced. The latter two steps may differ between SSOs and may have several iterations.

It should once more be highlighted that virtually all SSOs follow a variation of this process⁸.

Not unlike the processes, the organizational structures of major SSOs are similar. In most cases both their technical work and their internal organization are hierarchically structured. Figures 4 and 5 exemplary show ISO’s governance and work structure, respectively.

Figure 4. ISO's governance structure

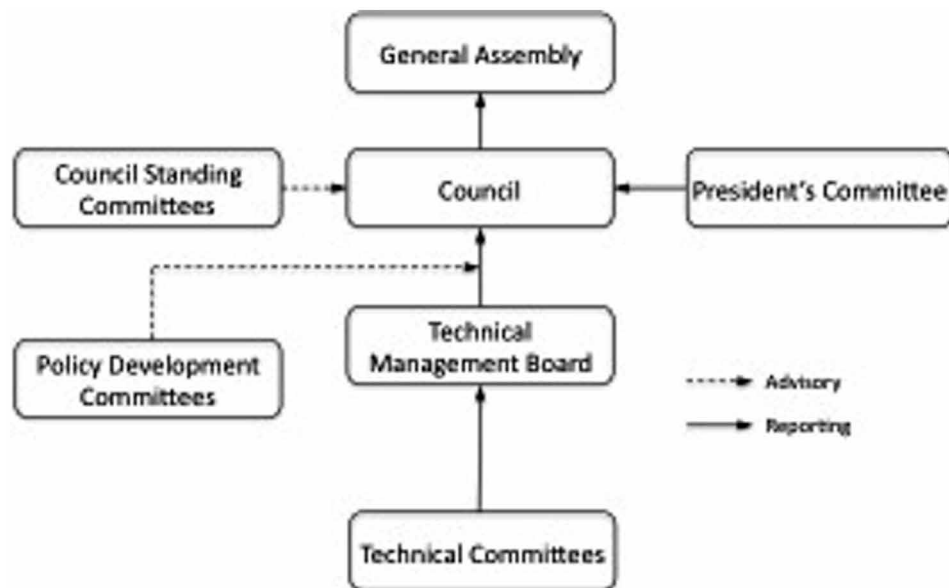
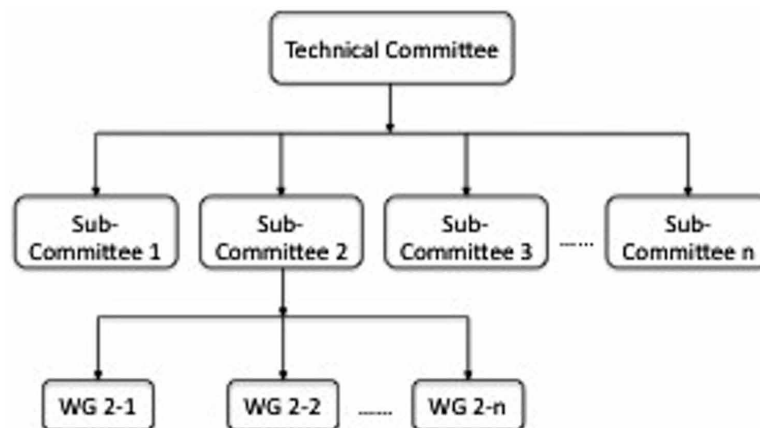


Figure 5. ISO's work structure

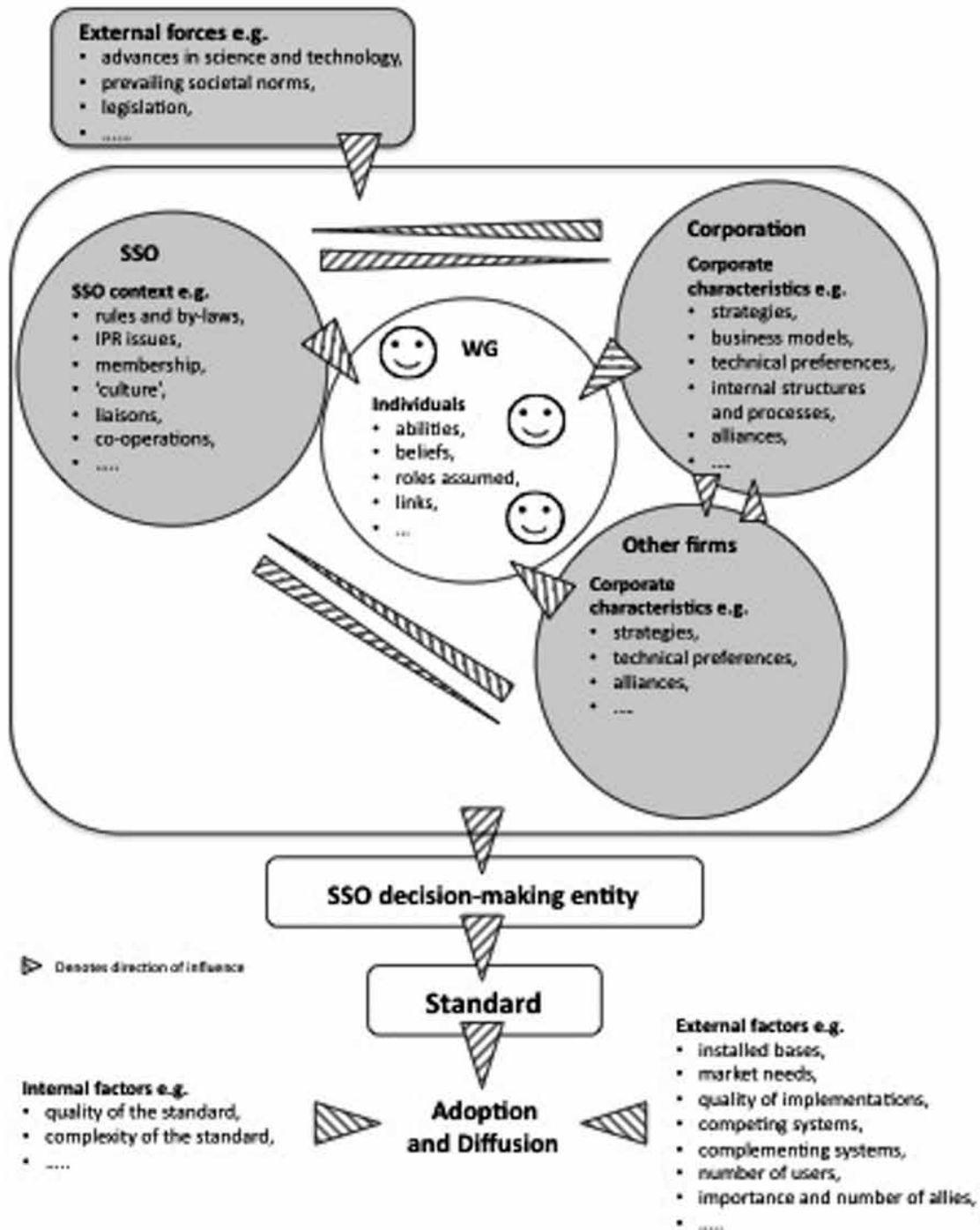


Typically, a TC covers a rather broad field. For example, ISO/IEC's Joint Technical Committee 1's (JTC 1) scope is 'Information Technology'. It is subdivided into 20 Subcommittees (SCs). Each Subcommittee, in turn, may comprise a number of Working Groups (WGs; more than 100 in total in JTC1). This is where the bulk of the actual technical work is done⁹. The more strategic decisions are normally taken at SC/TC level. This is also where the decisive voting is done.

MANAGING CORPORATE STANDARDIZATION

Standardization management refers to the set of activities a firm needs to perform in order to achieve its goal(s) in standards setting. That is, it needs to be clear about these goals in the first place. In many cases, standardization will be used as a tool to support a firm's business strategy in a given sector. To this end, standards setting activities and the business strategy need to be aligned.

Figure 6. Some factors that influence standards development (based on Jakobs, 2013).



Potential activities in standardization may range from simple intelligence gathering via pro-actively starting standardization initiatives to the foundation of a dedicated standards consortium to develop and promote standards for a particular field.

Once these corporate goals are clear, a number of factors need to be taken into account, including.

- The (potentially) relevant legal, moral, ethical, societal boundary conditions.

- The characteristics of the individual SSOs and the links that exist between them (see above). Here as well IPR issues will play a major role in many cases.
- The obvious allies and opponents that will also participate and their technical and strategic interests.
- Potential alliances that could be formed based on similar/complementing interests..
- The nature of the contributions to be made (e.g. technical, requirements, editorial).
- The content of the contributions.
- The size of the team, the roles and the necessary capabilities of its members.

The latter point is frequently ignored. However, at the end of the day standards are developed by a group of people who may cooperate or compete in a WG. These people's capabilities, backgrounds, views and agendas will to a considerable degree shape the final standard (see e.g. [Isaak, 2007], [Jakobs, 2011]). It should be noted that in addition to their technical expertise these people also need to be good communicators with a reasonable level of diplomatic and negotiation skills. It is, however, not always clear up-front whether or not they will actually act in the best interest of their respective employer [Jakobs, 2011]. Not least because of this uncertainty the selection of representatives with suitable expertise and experience and their briefing is a managerial issue whose importance should not be under-estimated.

Figure 6 summarizes and illustrates the different factors that need to be taken into account in corporate standardization management (without claiming completeness).

National Standardization Strategies

Not only firms have realized the importance of standards. Almost all industrialized nation states now have developed their own national standardization strategy¹⁰. Typically, standards are seen as important for the economy, policy making

and societal issues. A typical statement reads “*Standards are a fundamental building block for the economy and society*” [NSSF, 2003, p. 4]).

Almost all national strategies highlight the increasing globalization of markets and the resulting equally increasing need for globally valid, unique and accepted standards. At the same time, they also aim to provide a competitive advantage for the respective national firms and to strengthen the respective national position in the global market.

A number of other aspects may also be found in most national strategies. These include, for instance, the links between research, innovation and standardization. Along similar lines, they highlight the role of standards as a vehicle for the transport of innovations to the market as well as for the convergence of technologies (e.g. broadcasting, telecommunications and IT). The support of adequate education about standardization is another goal. This holds for both formal curricula and training on the job.

To achieve these goals, the need for an adequate national standardization infrastructure is stressed. This also includes the adaptation of current processes to shorter product life cycles and increasing consumer requirements. Specifically, the need for adequate (public) funding of the standardization infrastructure is also highlighted, not least in the light of the economic importance of standards.

Despite these similarities a number of differences between national strategies may also be identified. For one, the special needs and requirements of Small and Medium-sized Enterprises (SMEs) are only addressed in the strategies of the European countries. And while all strategies aim to improve the national performance in globalized markets the US strategy quite openly aims to ‘export’ its sector-based national standardization system, not least to counter the export of European technology and practices to other nations through the European standards processes. The Chinese strategy is the only one that explicitly highlights ICT. ‘Next Generation Information Technology Industry’ is the second of the 13 major areas iden-

tified where science and technology shall support the development of new technical standards. The support of R&D (Research and Development) for standardization is another focus of the strategy, an aspect the other strategies only touch upon.

FUTURE DIRECTIONS

In the past couple of years standardization has seen a development that may eventually change the whole standardization landscape – the mergers of formerly separate sectors. The – almost finished – integration of (tele)communication and information technology led to ICT. A next step – currently being taken – is the integration of ICT and broadcasting. IPTV and Internet Radio are well-known cases in point.

Such mergers of technologies will continue, and they are likely to do so on a broad scale. These emerging new technological fields will also be based on technical standards. Examples here would include, among others, Intelligent Transport Systems (ITS; ICT, transport telematics, traffic engineering), electric vehicles (ICT, automotive, power supply), e-health (ICT, medicine) and, perhaps ultimately, the Internet of Things (IoT) and Cyber-Physical Systems (CPSs; ICT and pretty much everything else).

These merging technologies represent a considerable problem for standards setting. They will require co-operation between standardization entities with very different cultures, from equally different backgrounds and with very different technology life cycles. On top of that, ICT in general, and specifically technologies like e-health and CPS, need to address issues of information security and safety and of trust. These aspects as well will need to be considered in the standardization process, adding yet another dimension of complexity. It remains to be seen if and how these problems will be addressed in practice.

CONCLUSION

Standards are crucially important, both economically and technically. This holds especially for the ICT sector, which simply would not function without globally accepted and implemented interoperability standards. Ongoing developments like ITS, e-health and the IoT, all of which require considerable interdisciplinary co-operation, will further increase the importance of standards and the underlying standardization processes.

REFERENCES

- Blind, K. (2013). *The impact of standardization and standards on innovation. Compendium of Evidence on the Effectiveness of Innovation Policy Intervention*. Retrieved 27 April 2015, from https://www.nesta.org.uk/sites/default/files/the_impact_of_standardization_and_standards_on_innovation.pdf
- CEC (Ed.). (1996). *Communication from the Commission to the Council and the Parliament on Standardization and the Global Information Society: The European Approach*. COM(96) 359 final. Retrieved 2 March 2015, from <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:1996:0359:FIN:EN:PDF>
- Coallier, F., & Motoei, A. (1997). Introduction to Software Engineering Standards. In K. E. Emam, W. Melo, & J. N. Drouin (Eds.), *SPICE: The theory and practice of software process improvement and capability determination* (pp. 1–18). Chichester, UK: Wiley.
- Deneulin, S., & Townsend, N. (2007). Public goods, global public goods and the common good. *International Journal of Social Economics*, 34(1/2), 19–36. doi:10.1108/03068290710723345
- DIN (Ed.). (2004). *Economic Benefits of Standardization: Summary of Results*. Berlin: Beuth Verlag.



- Farrell, J., & Saloner, G. (1985). Standardization, compatibility, and innovation. *The Rand Journal of Economics*, 16(1), 70–83. doi:10.2307/2555589
- Hawkins, R. (2009). Business Models and the Dynamics of Supply and Demand for Standards. In K. Jakobs (Ed.), *Information Communication Technology Standardization for E-Business Sectors: Integrating Supply and Demand Factors* (pp. 31–51). Hershey, PA: IGI. doi:10.4018/978-1-60566-320-3.ch004
- Hemenway, D. (1975). *Industry-wide Voluntary Product Standards*. Cambridge: Ballinger.
- Isaak, J. (2007). The role of individuals and social capital in POSIX standardization. *International Journal of IT Standards and Standardization Research*, 4(1), 1–23. doi:10.4018/jitsr.2006010101
- ISO. (Ed.). (2011). *Economic benefits of standards – International case studies – Volume 1*. Retrieved 27 April 2015, from http://www.iso.org/iso/home/store/publication_item.htm?pid=PUB100288
- ISO. (Ed.). (2012). *Economic benefits of standards – International case studies – Volume 2*. Retrieved 27 April 2015, from http://www.iso.org/iso/home/store/publication_item.htm?pid=PUB100311
- ISO. (Ed.). (2015). *ISO/IEC Directives, Part 1 – Consolidated ISO Supplement – Procedures specific to ISO*. Retrieved 27 April 2015, from <http://isotc.iso.org/livelink/livelink?func=ll&objId=17159827&objAction=Open&nexturl=%2F%2Flivelink%2F%2Flivelink%3Ffunc%3D%26objId%3D4230452%26objAction%3Dbrowse%26sort%3Dsubtype>
- Jakobs, K. (2011). How People and Stakeholders Shape Standards – The Case of IEEE 802.11. In J. Filipe & J. Cordeiro (Eds.), *Web Information Systems and Technologies 2010, LNBIP* (pp. 1–13). Berlin: Springer. doi:10.1007/978-3-642-22810-0_1
- Jakobs, K. (2013). Strategic positioning in ICT Standardisation. In K. Jakobs et al. (Eds.), *Standards: Boosting European Competitiveness* (pp. 201–216). Aachen: Mainz Publishers.
- Jakobs, K., & Kritznier, J. (2009). How to Select the Best Platform for ICT Standards Development? In K. Jakobs (Ed.), *Information and Communication Technology Standardization for E-Business: Integrating Supply and Demand Factors* (pp. 81–96). Hershey, PA: IGI Global. doi:10.4018/978-1-60566-320-3.ch007
- Morales-Trujillo, M. E., Oktaba, H., & Piattini, M. (2015). The making of an OMG standard. *Computer Standards & Interfaces*, 42, 84–94. doi:10.1016/j.csi.2015.05.001
- NSSF. (Ed.). (2003). *National Standardization Strategic Framework*. Retrieved 27 April 2015, from <http://www.strategicstandards.com/files/UK.pdf>
- Porter, M. E., & Kramer, M. R. (2011). Creating shared value. *Harvard Business Review*, 89(1/2), 62–77.
- Swann, G. M. P., & Lambert, R. (2010). Why do Standards Enable and Constrain Innovation? In *Proceedings of the 15th EURAS Conference* (pp. 356–376). Aachen: Mainz Academic Publishers.
- Updegrove, A. (2006). *What Standards Organization Should I Join?* Retrieved 27 April 2015, from <http://www.consortiuminfo.org/essentialguide/participating1.php>
- Updegrove, A. (2013). *Judge Robart’s Opinion in Motorola vs. Microsoft and the Future of FRAND*. Retrieved 27 April 2015, from <http://www.consortiuminfo.org/bulletins/#testimony>

WTO (Ed.). (2015). *World Trade Organization (ed.). Decisions and recommendations adopted by the WTO committee on Technical Barriers to Trade, Rev. 12*. Retrieved 27 April 2015, from https://docs.wto.org/dol2fe/Pages/FE_Search/FE_S_S006.aspx?MetaCollection=WTO&SymbolList=&Serial=&IssuingDateFrom=&IssuingDateTo=&CATTITLE=DECISIONS+and+recommendations+ADOPTED+BY+THE++WTO+COMMITTEE+on+Technical+Barriers+to+trade++SINCE+1+JANUARY+1995&ConcernedCountryList=&OtherCountryList=&SubjectList=&TypeList=&AutoSummary=&FullText=&FullTextForm=&ProductList=&BodyList=&OrganizationList=&ArticleList=&Contents=&CollectionList=&RestrictionTypeName=&PostingDateFrom=&PostingDateTo=&DerstrictionDateFrom=&DerstrictionDateTo=&ReferenceList=&Language=ENGLISH&SearchPage=FE_S_S001&ActiveTabIndex=0&&languageUIChanged=true#

KEY TERMS AND DEFINITIONS

Information and Communication Technology (ICT): Encompasses computer hardware and software as well as the communication networks necessary to process and transmit data.

Innovation: The implementation of a new or significantly improved product (good or service), process or new organizational method.

Intellectual Property Rights (IPR): IPR is the legal right granted with the aim to protect a creation of the intellect. These include e.g. patents, trademarks and copyright.

Standard: Document, established by consensus and approved by a recognized body, that provides, for common and repeated use, rules, guidelines or characteristics for activities or their

results, aimed at the achievement of the optimum degree of order in a given context.

Standardization: The process through which a standard is developed.

Standards Setting Organization: An organization that provides a platform for the development of standards.

World Trade Organization (WTO): The WTO deals with the global rules of trade between nations. Its main function is to ensure that trade flows as smoothly, predictably and freely as possible.

ENDNOTES

- ¹ In short, according to [ISO, 2015] consensus means a general agreement, characterized by the absence of sustained opposition to substantial issues. Consensus need not imply unanimity.
- ² ITU is a United Nations agency for information and communication technology issues.
- ³ This term is used to denote both formal SDOs and private standards consortia.
- ⁴ <http://www.3gpp.org/>.
- ⁵ The World Wide Web Consortium applies the RF regime.
- ⁶ All SDOs and most consortia apply the (F) RAND regime.
- ⁷ In the following, ISO terminology will exemplary be used.
- ⁸ See [Morales-Trujillo et al., 2015] for a detailed description of a process leading to a de-facto standard.
- ⁹ Depending on the SSO's bylaws WG members either act in a personal capacity or represent their employer or country, respectively.
- ¹⁰ A compilation of such national strategies may be found through <http://www.strategicstandards.com/Perspectives.html>.

Immersing People in Scientific Knowledge and Technological Innovation Through Disney's Use of Installation Art

Jonathan Lillie

Loyola University Maryland, USA

Michelle Jones-Lillie

Lillie Pad Studios, USA

INTRODUCTION

This chapter uses Disney's extremely successful exhibits at the 1964-65 New York World's Fair as iconic examples of art installations designed to provide strong experiences of scientific progress and technological innovation. The chapter therefore explores how Disney and installation artists have used this form to create immersive, memorable experiences for audiences. The goal is to identify methods for using installation art to convey scientific knowledge to general audiences to foster greater understanding and inspire future generations toward scientific and technological discovery.

When Walt Disney began to sketch out his plans for Disneyland, which opened in 1955, he was explicit in this desire to bring the same narrative craft to rides and exhibits that he demanded of his feature films. He wanted visitors to (actually) step inside, immerse themselves in the story and experience the wonder and emotional connection directly. While the Disneyland and World's Fair exhibits are certainly not the first examples of the construction of large-scale, immersive narrative experiences, they have been extremely influential in the areas of art, education and culture. At the 1964-65 fair in New York, Walt Disney wanted to demonstrate the vast improvement in installation/dark ride technology over the 1939 World's Fair (also in New York) and its influential future-gazing exhibits such as GM's "Futurama." In later

years the Disney Company, groups of artists, and institutions such as museums have drawn on the methods used by these iconic exhibits to develop installations to convey concepts and trends in science and technology.

Disney designed and built four installation exhibits for the 1964-65 fair: Ford's "Magic Skyway"; "The Carousel of Progress" for General Electric's Progressland Pavilion; "Great Moments with Mr. Lincoln" for the State of Illinois; and "It's a Small World" co-sponsored by Pepsi and UNICEF. The narrative craft of exhibits such as Carousel of Progress (Figure 1) and the Magic Skyway (Figures 2 and 3) are compared below to works within the genre of installation art, which has developed greatly since the 1960s. Similar to Disney, many artists have deployed immersive installation art exhibits to envelop audiences in a detailed aesthetic and conceptual narrative. Some artists, as well as educational institutions, have used experiential installations for addressing or presenting scientific concepts, and much potential exist for further work in this direction.

HISTORICAL BACKGROUND AND LITERATURE REVIEW

The 1950s and 1960s were vital for the development of installation art. Installation was one of several anti-commodity forms such as conceptual art, performance and body art, earth works, and

DOI: 10.4018/978-1-5225-2255-3.ch407

Figure 1. “Father,” dog, and electric oven from the Walt Disney World revised version of the Carousel show (2006, public domain, photo by SteamFan available on Wikipedia)



new forms of sculpture, which emerged in these decades (Morse, 1998, p. 4). However the birth of the genre is often traced back to the surrealist exhibits held in Paris in 1938, 1957, 1959 and in New York in 1942, which inspired the contemporary “Art Installation movement” insofar as the artists desire to display their artworks according to their own aesthetics rather than allowing a museum curator to take on that task (thus, defining the exhibition of the collection of artworks as an integral element of the artwork as a whole). The first of these was the 1938 *Exposition Internationale du Surréalisme*, a surrealist exhibition held in Paris that included Salvador Dali’s installation “Rain Taxi” (Tomkins, 1996, p. 364). In the 1960s and 70s installation art was certainly explored, however the occurrences were sporadic and did not truly reach the height of its label as an artistic movement until the 1990s (Bishop, 2005, p. 8). In the 1980s installation art began making its way to the forefront in major international exhibitions such as *The Venice Biennale*, *Sao Paulo Biennial*, and *the Skulptur Projekte, Munster*, as a way to fill large architectural or industrial spaces with a sensorial experience aiming to leave guests with a more tangible memory of the exhibition. Increasingly,

more and more post-1960’s art venues sought to exhibit installation art as an outcome. Currently, Installation Art is the backbone of many international biennials and triennials for its ability to completely immerse the guests in the art and alter the consciousness of the guests through sensory perception (Bishop, 2005, p. 37).

The Disney exhibits at the fair are examples of installation art in several aspects. First, as opposed to more long-standing forms of two-dimensional art, installations must, of course, be installed by the artist(s), typically in a prepared space. Often in more immersive works the boundaries of the space, such as interior walls, have been modified to the specific needs of the piece, transforming it into a new “environment.” Following the surrealist expositions of the 1930s and 40s, the next important development of the genre emerged with the experiential “environments” and “happenings” created in the 1960s by artists such as Allan Kaprow, who were reacting against the commercial boom of Abstract Expressionist paintings in European and American galleries (Bishop, 2005, p. 23). As opposed to Disney’s commercial engineering approach,¹ the spaces for these Kaprow’s environments tended to either be disused

Figure 2. The Magic Skyway (© 1965. With permission from the Collections of The Henry Ford)



spaces such as abandoned warehouses and caves, or galleries extensively altered by the artists.

The fair exhibits are also installations in that they accomplished “the actual construction of a passage for bodies or figures in space and time” (Morse, 1998, p. 154). As art historian Margaret Morse describes, the character of installation and the materials inside the installation are “meaningful within the whole pattern of orientations and constraints on the passage of either the body of the visitor or of conceptual figures through various modes of manifestation—pictorial, sculptural, kinesthetic, aural, and linguistic” (p. 154). Indeed, the Disney installations at the fair took advantage of all of these ways to engage visitors. Disney developed several modes of manifestation to achieve this vision for the Magic Skyway, which took passengers in modified Mustang convertibles on a journey of ‘human progress’ from the invention of speech by early humans to techno-cities of the future. “What we want to provide guests at the Ford Pavilion is an entirely original experience, something no one has ever seen or done before,” Walt promised in his sales pitch to Ford. “It could never happen in real life, but we can achieve the

illusion by creating an adventure so realistic that visitors will feel they have lived through a wonderful, once-in-a-lifetime experience” (Ford Times, December 1963). Of particular note are the sculptural and kinesthetic techniques used in the Skyway and Carousel. The Skyway (Figure 3) first allowed visitors a “stunning” view of the fairgrounds before traveling back to “the dawn of life.” It was here that audio-animatronic dinosaur sculptures battled in a foggy prehistoric landscape. The next scenes presented “cavemen” in comic scenes inventing speech, writing, and the wheel (Figure 3). These sculptures moved by a pneumatic technology developed a few years earlier for the jungle cruise in Disneyland. GE’s funding of the Carousel (as well as the money for the Lincoln robot for the Illinois pavilion) allowed Disney’s engineers to take audio animatronics to the next level to create the “lifelike” stereotypical white, mid-western, American family (Figure 1). The heavy reliance on audio animatronics, aside from practical considerations, contributed to a sense of technological wonderment and authenticity. These mechanical bodies in movement shared the same space with the slowly moving visitors. And

the kinesthetic aspect of being moved through the exhibit was just as important to the visitors' experience as the robot characters were. Both the Skyway and Carousel² were thematically based on a linear movement through time, witnessing a segment of human history.

Compared to most installation artwork the Disney installations were over-proscribed and constraining to visitors, more like the classic cinematic experience of being shunted to a dark seat to observe with an omniscient narrator to guide you through the story (the exhibits used narrator voices, music, and sound effects). Morse explains how the theater and film can be explored through Plato's "Simile of the Cave" where the spectator is divided from the realm to be contemplated. In the cave, "the machinery that creates the vision of another world is largely hidden, allowing the immobilized spectator to sink into an impression of its reality with horror or delight but without danger from the world on view," Morse notes.³

It is here that we can find perhaps the greatest difference between the Disney exhibits at the fair and installation art. What Disney hoped to learn

how to perfect at the fair and new rides at Disneyland was the immersification of the modern film. Creating a space where visitors are able to actually go inside the narrative with their bodies, not just their minds, as the cinema experience seems to privileged. Most installation art, on the other hand, try to force visitors to be mindful and critical of where they are, what they are doing, and what is being done to them. Many art historians (Hawkins, 2010; Morris, 1998) see immersive installation art as particularly useful in drawing attention to the fact that cognition is a product of embodied experience. For example, artist Natalie McKeever's installation "Internal Worlds" (2013, Figure 4) immerses the body in a technological apparatus to not only make the visitor more aware of her or his own body – visitors see and here their own biorhythm data in real time – but also of their bodily existence in a technology saturated society. This installation is "an ontological study of humans' body-mediated perception in a world" (Lillie, 2014) where perception is increasingly mediated through ubiquitous information and communication technology.

Figure 3. Skyway "Cavemen" (© 1965. With permission from the Collections of The Henry Ford)



Figure 4. McKeever's "Internal Worlds" (© 2013 N. McKeever. Used with permission)



IMMERSIVE NARRATIVE

V. IMMERSIVE PLAY

"The underlying premise of installation appears to be that the audiovisual experience supplemented kinesthetically can be a kind of learning not with the mind alone, but with the body itself," writes Morse (1998, p. 158). Certainly this description can be applied to Disney's installations. The fair exhibits were undoubtedly so powerful due to how the visitors' bodies were moved through a *narrative*. Narratology scholar Gerald Prince notes that the word narrative has emerged from the Latin *gnarus*, which is translated to "knowing" or "acquainted with." *Gnārus* comes from the Indo-European root *gnā*, which means "to know." Narrative therefore "is a particular mode of knowledge," Prince concludes. "It does not merely reflect what happens; it discovers and invents what can happen. It does not simply record events; it constitutes and interprets them as meaningful parts of meaningful wholes, whether the latter are situations, practices, persons, or societies" (Prince, 2000, p. 129). With Disney's use of installation, rides and exhibits were, and still are, most often

used to convey narratives about technological progress (e.g., many of the original and current features at the Epcot Center theme park in Florida, USA). The careful crafting⁴ of narrative invites cognition, decoding and often dominant readings. Knowledge presentation of any kind faces the same challenge as any message created for consumption, finding a successful way to convey a concept. Disney solution to this conundrum is found in its devotion to the complete storytelling process: every aspect of the narrative and how to tell it was, and still is, absolute. Whereas Disney's exhibits tried very hard to encourage audiences to connect pro-social emotions to specific concepts, most installation artists design their works to be open to a wide range of interpretations. Both Disney and most installation artists, however, strive also to evoke an emotional reaction.

The installation artist leaves the location once set up is complete, leaving the visitor to perform the work on her own, or at the same time as other visitors. Rarely are installations as scripted and narrative driven as the Disney exhibits,⁵ but media scholars recognize that with forms such as television and film, the mind of the viewer is

usually highly active, making meaning from what they see and hear. Disney's exhibitions displayed masterful use of textual references, often nostalgia, for visitors' minds to decode, often with an emotional response. Whether the classic T-rex vs. Stegosaurus fight and caveman antics of the Skyway, the long-suffering livingroom patriarch in his armchair of the Carousel, or the austere, inspiring audio-animatronic President Lincoln of the Illinois pavilion, visitors are able to access the many textual references to American culture and history so that the experience seems bigger and more significant than just a tour through a large aluminum building. And audiences were also always primed in the pavilion pre-shows for the amazement that they were about to experience. In contrast, installation artists such as Ilya Kabakov often arrange related objects in a room, or series of room, in an attempt to tell a story that is very much open to visitors' interpretations (*see* Bishop, 2005, p. 15-16, for a description of Kabakov's narrative installations).

Unlike the scripted, controlled experience that Disney preferred, most installation art is often predicated on letting, or even forcing, the audience to *play*, in other words the journey through the narrative is often a choose-your-own adventure of sorts. While the artist guides the audience there is much more room for a wide variety of experiences and interpretations. As in other areas of the arts, installation artists' uptake of critical/postmodern theory (more Foucault [1977], Baudrillard [1981], and Barthes [1957] in this discipline than Jameson [1991]) understands that no reading of a text can be determined, and in fact that may not be desirable. Although almost all installation art is conceptually grounded, immersive installations are often designed to be experienced through the full range of human senses, but also upping the cognitive ante by encouraging audiences to actively think about and critically analyze what they are experiencing and feeling. Therefore, installation art audiences are often left feeling that they made their own conclusions based on their ability to think critically as oppose to being swept through an experiential

journey merely to be entertained. The relationship between the artist and the audience seems agreed to be about reaching an intellectual enlightenment of a sort, which of course is an admirable goal for many undertakings of scientific knowledge presentation, especially in educational settings. Educational destinations such as zoos, aquariums, and science centers have for several decades continued to increase the number and diversity of interactive exhibits (Allen, 2004), acknowledging that there are many ways to convey knowledge and you never know which method will end up connecting with which visitor.

The importance for interactivity, some call it "play," for installation artists is similar to the importance of visitor exploration and discovery for immersive installation artists and science or cultural museums alike. Sue Allen, the director of visitor research and evaluation at the Exploratorium in San Francisco, acknowledges that the self-guided exploration that makes science museums so fun and novel also makes them difficult to design for learning. "On the exhibit floor there is no accountability, no curriculum, no teachers to enforce concentration, no experienced guide to interpret and give significance to the vast amounts of stimulus and information presented," she writes. "Without restrictions, visitors have complete freedom to follow their interests and impulses as they move through a public space packed with exhibits all vying for attention" (2004, p. 18). Immersive installations, however, generally do not suffer from this problem, since they are typically a single exhibit focused on one particular experience and concept.

INSTALLATION FOR KNOWLEDGE PRESENTATION

In a reputable volume on representation in scientific practice Lorraine Daston reminds us that the "specter of the perfect copy still haunts the history of science, despite the best efforts of scientists and artists to explain that the most faithful rendering



is often not one that could be mistaken for the original” (Daston, 2014, p. 319). Presentations or demonstrations of scientific concepts will often fall short when they do not make sense to lay audiences and cannot address the “bigger picture” or contexts. Of course all such endeavors must be simulations of a sort, and there is a danger in the possibility that audiences will believe that simulations are indeed exact replications. Many children left the Magic Skyway feeling they had glimpsed a close approximation of what dinosaurs and “cavemen” were actually like; some visitors watching the Abraham Lincoln robot give its speech thought it was a human actor. And the Carousel purported to tell a truism story of the history of the American home. On the other hand, narrative storytelling is one of the most fundamental forms of knowledge acquisition for our species, the other being direct experience. But both of these are considered flawed by the paradigm of objectivity. This is perhaps one of the reasons science loves “facts” in the form of numbers, stats, and equations born out by experimentation and/or detailed observation. They seem more like truth, but we need language to convey them. Also, the things such facts are telling about are typically so far abstracted from the reader of such knowledge that it is difficult to place it in context, to “make sense” of it in the context of our everyday lives. Which is where storytelling, narrative, comes in. Can you imagine if Disney’s follow up to *Frozen* were a strong message on the need for every person and community to act to combat global warming? But of course any heavy-handed message would detract from the narrative and would therefore the concept would be severely lightened and simplified to make it more digestible for multiple publics and corporate sponsors. Installation art on the other hand is well suited to presenting scientific knowledge, acknowledging and desiring for the narrative to be highly subjective but also emotionally powerful.

Part of the power in the Disney rides was that they were literally and figuratively a journey through the narrative. In the Carousel of Progress and Magic Skyway the forced movement of the

audience was central to experiencing the concept of the movement through time and history. Yet the experiential possibilities were constrained and forced by the design of the immersive narrative apparatus. Interactive scientific exhibits are similarly constrained by the apparatus design. A web-based science exhibit is experienced and constrained by the technology: keyboard, screen, the possible modes of user interaction programmed into the webpage and again limited by what current technologies can enable. Regional science museums and discovery centers on the other hand are place-based with open-architectures, built to engage children’s sense of play and exploration. Typically many individual exhibits inside these centers use methods of knowledge presentation, but the overall purpose is not merely rote learning of scientific facts and concepts, but perhaps more vital, cultivating a love of learning about the world and a desire to know more. In some respects many immersive art installations are more like these science centers for kids than the Disney exhibits and rides. Morse notes that we “lack the vocabulary for kinesthetic ‘insights,’ for learning at the level of the body ego and its orientation in space. Perhaps such learning principles might be considered ‘Deweyian,’ a ‘figuring within’ as opposed to the ‘reading’ of literature or the ‘imagining’ of pictorial art” (Morse, 1998, p. 153). However educational institutions in sciences, whether engineering, medicine, chemistry, etc., have long understood that knowledge as it is desired to be applied must be learned in part through experience. For an example of installation used in science education, consider “the Giant Heart” (Figure 5) exhibit at the Franklin Institute in Philadelphia, where visitors walk through the gallery-sized heart installation and learn about the mechanics of the human body.

CONCLUSION

Although the Disney model of dark-ride installations are by many measurements highly success-

Figure 5. “The Giant Heart” at the Franklin Institute (© 2010 Cary Whitt. Used with permission)



ful, there are also many reasons why scientists or educational institutions might not go down that path to presenting knowledge with a strict narrative script for large audiences. First of all: It takes a lot of money! Walt Disney was mainly interested in participating in the 1964-65 fair because the company did not have the funds to support all the research and development needed to expand Disneyland with high-tech dark rides. He needed the fair's rich corporate sponsors for that. Similarly, like large science studies, large installations typically require funding. Certainly funded installations that present scientific knowledge are to be applauded not shunned, but the trick is to allow artists to use their conceptual knowledge and skills and not force perceived modes of entertainment into the installation, which in fact might close down rather than open up visitors' abilities to critically engage. Second, when a highly narrative immersive installation is intended to be at least in part entertaining, the tendency will usually be to eschew complex facts for amusement and narrative simplicity/coherence. While science and history museums are important institutions for experiential education, where installation is often

one form of knowledge presentation, such venues are often burdened with many other tasks – such as boosting attendance, growing endowments, developing age-appropriate curricula for school fieldtrips and the like (*see* Silberman's, 2005, critique of museums).

Installation can be a powerful tool for moving people. They may move people physically and emotionally, as the Disney exhibits at the fair, or they may move people to critical thinking, or merely move people out of their comfort zone. Certainly, for many artists, one of the reasons for making art is to compel people to both have an emotional reaction to a work but also to think critically. Yet this is also a dual purpose for the making and conveying of scientific knowledge. The title of the popular web/social media curator, “I Fucking Love Science” is very illustrative of this point. Of course, unlike Disney, where the desired emotional responses are positive (wonder, amusement, etc.) some of the most powerful installations hope to invoke unexpected, disjointed feelings from visitors, provoking them to reflect on how they are feeling and why. A variety of studies in the biological sciences have shown that

emotional experiences are typically better stored in memory (see Bower, 1992; Canli et al., 2000; Christianson, 1992). For example, a recent study of visitors to a wildlife exhibit found that four months later visitors recalled emotional responses much more than factual information and that they connected these emotions to positive changes in their pro-conservation attitudes and behaviors (Ballantyne, Packer, & Sutherland, 2011). This could be the reason so many visitors to the 1964-65 New York fair report fond and detailed memories of the Disney installations even decades later (Anderson, 1995). While Disney wanted to create powerful, pro-social entertainment, for many installation artists, however, the purpose of their artwork is to provide visitors with experiences that are powerful because they are out of the ordinary – sometimes unsettling, other times mesmerizing – but typically using an immersive, multiple media approach to impart such feelings.

FUTURE RESEARCH DIRECTIONS

Conveying scientific knowledge and technological innovation faces the same challenge as any message created for public consumption: finding ways to convey a concept. Disney as well as many installation artists realized that conveying the message can be a much more powerful experience when audiences are compelled to feel. Cognition is never merely conveying of information. Human cognition might be computer like in some respects, but as phenomenologists have realized for decades information must be routed through the whole body-as-apparatus to go through the process. Disney's method for this was cinematic, and thus the company's devotion to the complete storytelling process, every aspect of the narrative and how to tell it, was absolute. Artists on the other hand take many different approaches to developing a 'structure of feeling' through installation. The majority of colleges and universities that offer degree programs in art offer courses in

Installation Art. And it so happens that colleges and universities are at the forefront of science and technology discovery and education. Therefore collaboration between Art departments/faculty/students and their counterparts in the Sciences would be the logical vehicle for developing installations that convey new science and/or technology concepts. Such cross disciplinary ventures are likely to be supported by educational institutions that are increasingly trying to get students to make connections among their different disciplines of study. Furthermore these collaborations can be flashpoints for developing partnerships with local tech industries, science centers, and museums.

REFERENCES

- Allen, S. (2004). *Designs for learning: Studying science museum exhibits that so more than entertain*. Wiley InterScience. Retrieved from www.interscience.wiley.com
- Anderson, P. (1995). *Persistence of Vision: An Unofficial Historical Journal celebrating the Creative Legacy of Walt Disney*. Salt Lake City, UT: Anderson Publishing.
- Ballantyne, R., Packer, J., & Sutherland, L. A. (2011). Visitors memories of wildlife tourism: Implications for the design of powerful interpretive experiences. *Tourism Management*, 32(4), 770–779. doi:10.1016/j.tourman.2010.06.012
- Barthes, R. (1957). *Mythologies*. Paris: Seuil.
- Baudrillard, J. (1981). *Simulacra and Simulation*. Ann Arbor, MI: University of Michigan Press.
- Bishop, C. (2005). *Installation Art: A critical history*. London: Routledge.
- Bower, G. H. (1992). How might emotions affect learning. In *The handbook of emotion and memory: Research and theory*, (pp. 3-31). Academic Press.

Canli, T., Zhao, Z., Brewer, J., Gabrieli, J. D., & Cahill, L. (2000). Event-related activation in the human amygdala associates with later memory for individual emotional experience. *The Journal of Neuroscience*, 20(19), RC99–RC1. PMID:11000199

Cardiff, J., & Bures-Miller, G. (2001). Retrieved September 16, 2014, from: http://www.cardiffmiller.com/artworks/inst/paradise_institute.html

Christianson, S.-A. (1992). *The handbook of emotion and memory: research and theory*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Daston, L. (2014). Beyond Representation. In J. V. Catelijne, M. Lynch, & S. Wolgar (Eds.), *Representation in Scientific Practice Revisited* (pp. 319–322). London: MIT Press. doi:10.7551/mitpress/9780262525381.003.0016

De Oliveira, N., Oxley, N., & Petry, M. (2003). *Installation Art in the New Millennium*. London: Thames & Hudson.

Exposition Internationale du Surréalisme. (2013). Retrieved September 16, 2014, from: http://en.wikipedia.org/wiki/Exposition_Internationale_du_Surréalisme

Foucault, M. (1977). Discipline and Punish: The Birth of the Prison (A. Sheridan, Trans.). New York: Vintage. Ford Times, 57(5).

Hawkins, H. (2010). The argument of the eye? The cultural geographies of installation art. *Cultural Geographies*, 17(3), 321–340. doi:10.1177/1474474010368605

Jameson, F. (1991). *Postmodernism: The Cultural Logic of Late Capitalism*. Durham, NC: Duke University Press.

Lillie, J. (2014). Twelve years of NMEDIAC: An introduction to the Winter 2013-14 issue. *NMEDIAC, The Journal of New Media & Culture*, 9(1).

Morse, M. (1998). Video installation art: The body, the image, and the space-in-between. In M. Morse (Ed.), *Virtualities: television, media art, and cyberculture*. Bloomington, IN: Indiana University Press.

Prince, G. (2000). On Narratology (Past, Present, Future). In M. Mcquillan (Ed.), *The Narrative Reader*. London: Routledge.

Silberman, N. A. (2005). Beyond Theme Parks and Digitized Data: What Can Cultural Heritage Technologies Contribute to the Public Understanding of the Past?. *European Commission*, 2005, 9–12.

Tomkins, C. (1996). *Marcel Duchamp: A Biography*. London: Holt Paperbacks.

Young, W. P. (2001). *Let's Ride! The Magic Skyway*. Retrieved April 20, 2012, from Nywf64.com/ford08.shtml

ADDITIONAL READING

Bierman, J. H. (1976). The Walt Disney Robot Dramas. *The Yale Review*, LXVI, 223–236.

Davies, H. M., & Onorato, R. J. (1997). Blurring the boundaries: installation art, 1969-1996. Museum of Contemporary Art, San Diego.

Newhouse, V. (2005). *Art and the Power of Placement*. New York: Monacelli Press.

Reiss, J. H. (2001). *From margin to center: the spaces of installation art*. MIT Press.

Rosenthal, M. L. (2003). *Understanding installation art: from Duchamp to Holzer*. Prestel Pub.

Smith, D., & Steven, C. (1999). *Disney: The First Hundred Years*. New York: Hyperion.

Smith, M. L. (1993). Making Time: Representations of Technology at the 1964 World's Fair. In *The Power of Culture: Critical Essays in American History*, eds. Richard W. Fox and T. J. Jackson Lears. Chicago: U of Chicago, Suderburg, E. (Ed.). (2000). *Space, site, intervention: situating installation art*. U of Minnesota Press.

Weiner, L. Y. (1997). There's a Great Big Beautiful Tomorrow: Historic Memory and Gender in Walt Disney's Carousel of Progress.. *Journal of American Culture*, 2(1), 111–117. doi:10.1111/j.1542-734X.1997.00111.x

Wilson, S. (2002). *Information arts: intersections of art, science, and technology*. MIT press.

KEY TERMS AND DEFINITIONS

1964-1965 New York World's Fair: Between April 22, 1964 and October 17, 1965 fifty-one million people visited the New York World's Fair held at Flushing Meadow in Queens, New York, on the site of the 1939 World's Fair. Due to a stipulation that no nation can host a World's Fair more than once a decade and that an official World's Fair was held in Seattle in 1962, the New York Fair did not enjoy the official sanction of the Bureau of International Expositions.

Dark Ride: A “dark ride” is the term often used to describe large indoor installations produced for the purpose of entertainment at theme parks such as Disneyland. They tend to be heavily immersive and themed environments. They are called by other names as well, such as “ghost train” in the UK and Australia.

Disney, Walt: Walt E. Disney (born December 5, 1901 – died December 15, 1966) is a famous American media pioneer. The Walt Disney Company, founded by Walt in 1923, is best known for its animated fantasy feature films, theme parks, and television channels ABC, ESPN and the family of Disney channels catering to children and teens.

Installation Art: An art movement or genre based on three-dimensional work. It was not considered a true art movement until the 1980s, but has roots in surrealism exhibitions of the 1930 and 40s. Most installation artwork comprises multiple objects, multiple media, and takes up a whole room or many rooms in a gallery or venue.

Kaprow, Allan: Kaprow (1927–2006) was an American artist known for organizing art exhibitions in the 1950s and 60s that he termed “environments” due to the artists’ altering of the art venues combined with thematic performances.

Narrative: A narrative is simply a story or version of events communicated through language. Scholars in many disciplines including English, media studies, linguistics, cultural theory, and anthropology have studied how humans communicate and form communities using different forms of narrative. Since installations use many pieces of art, they can be used together to create a narrative of sorts.

ENDNOTES

- ¹ For the fair Disney's architects and “imagineers” drew up specs for the buildings that had to be constructed to house its feature exhibits. Despite this large-format construction, however, all four exhibits along with almost all of the other structures associated with the fair were torn down. The vast majority of installation artwork is only temporarily installed in a location. More successful and well-funded installations may bounce among museums and galleries as traveling exhibits. The Carousel, Small World and Lincoln robot were all later installed as slightly different versions in Disneyland and later duplicated (and again slightly revised) more than a decade later for Disney World in Florida.
- ² For the Carousel Disney engineers proposed and developed the first revolving audience

theater: a “doughnut on railroad wheels” containing six compartmentalized sets of seats revolved around the center that contained six stages allowing a 3,900 people per hour capacity (Anderson, 1995, p. 53). The stages showed the animatronic family enjoying state of the art home electronics from the 1890s, 1920s, 1940s, and late 1960s. The central concept of technological progress was literally felt as the whole group of seats moved as one scene ended to the next scene that represented ideal middle-class domestic life decades later.

³ Morse continues: “Cinema spectators immobilized in darkness were like the prisoners in Plato’s Cave, but they are not held in place by chains but by machines of desire, enjoying the impression of mastery over an imaginary world” (Morse, 1998, p. 156-157; See Morse for a discussion of film theory’s critique of Rene Descartes’s [hence “Cartesian”] arguments about human experience). Critical theories have drawn particular attention to how to the modern era of film has been dominated by Cartesian narratives/design, where bodies are placed, immobile, in a black box to observe the neat, manufactured realities on the other side of the screen.

⁴ The Disney exhibits were successful due to their rigid devotion to narrative craft. The use of technology made the exhibits seem to come to life, enhanced visitors’ sense of wonder and evoked a positive emotional response. Thus the techno-enabled narratives were made it more likely that these experiences would imprint into memory. And the rides were designed to enable as many people as possible to have a similar pro-social narrative experience. The company was still so geared toward cinematic narrative construction in the 1950s and 60s that the lead project managers for the fair exhibits

and Disneyland were called “art directors.”

Perhaps the fair exhibit with the most long-term success and impact was “It’s a Small World,” which – 50 years later – is currently in operation in all six major Disney theme parks around the globe. Although “Walt’s little boat ride” was a tour through 26 world cultures, he wanted a strong, consistent design style to drive the narrative. Mary Blair, who provided concept art for recent Disney films including *Cinderella*, *Alice in Wonderland* (1951) and *Peter Pan* (1953) was brought in to provide styling themes for the ride including the child sculptures, faux buildings and landscaping, and facades inside and outside the building that housed the ride. Similar to Disney’s feature films, catchy songs were also central to the narratives of the Carousel (“It’s a Great Big Beautiful Tomorrow”) and Small World.

⁵ For the Disney exhibits most non-scripted physical interaction between visitors and the exhibits were not desired due to the need for high per-hour capacity. But such interactivity was also viewed as a detractor from the carefully crafted narrative. Ford Motor managers wanted Ford salesmen to act the role of chauffeurs in the Skyway mustangs, which they thought would certainly result in increased sales of their newest model (the mustang’s official debut date was the opening day of the fair). But Walt personally talked them out of this plan, saying it would detract from the wondrous experience that the Skyway would impart (Anderson, 1995, p. 55). Disney’s devotion to cinematic narrative was so complete that the instances when the narrative was interrupted by temporary mechanical glitches, the extent to which the experience is manufactured and constraining becomes quickly evident.

Indicators of Information and Communication Technology

Gulnara Abdrakhmanova

National Research University Higher School of Economics, Russia

Leonid Gokhberg

National Research University Higher School of Economics, Russia

Alexander Sokolov

National Research University Higher School of Economics, Russia

INTRODUCTION

Over the past twenty years, technologies using microelectronics for collection, storage, processing, retrieval, transmission, and presentation of data, texts, images, and sound, collectively known as ICT have completely changed all people's activities. The rapid proliferation of ICT and their impact on all spheres of modern life (see Ahmad et al. [2004]) – production processes, the interaction of individuals and organisations with public authorities, the development of social infrastructure, and privacy issues – has stimulated the interest to statistical analysis of the ICT sector's prospects at the national and regional levels. Influence of ICT led to emergence of a new socioeconomic configuration commonly referred to as "Information Society".

BACKGROUND

An Information Society is usually understood as a society that makes extensive use of information networks and technologies, produces large quantities of ICT goods and services, and has a diversified content industry. Both theoretical and practical issues related to measuring different aspect of Information Society has been increasingly addressed by many authors during the last 20 years (see for example Blank, Groselj [2014];

Dolničar et al. [2014]; Billon et al. [2016]). The key three thematic "pillars" related to the Information Society are as follows (Figure 1):

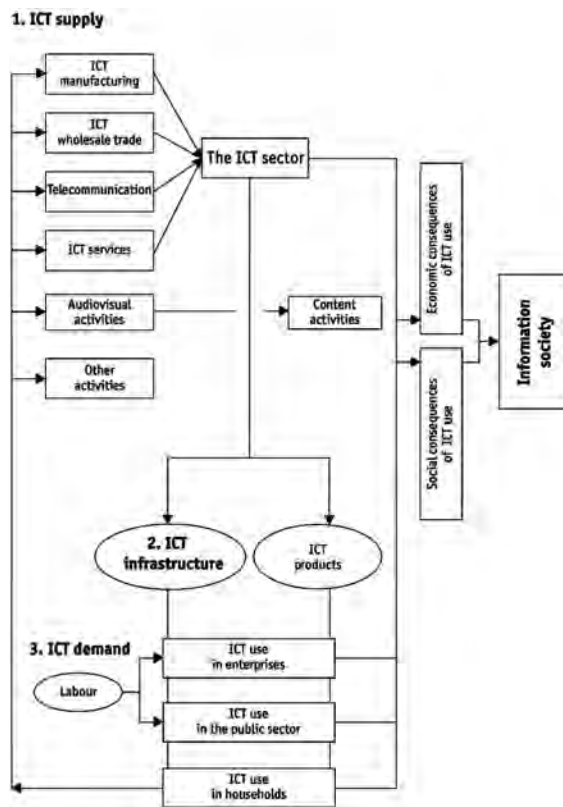
- **ICT Sector and the Supply of ICT:** Which industries it includes, how important they are for the national economy, how many enterprises are involved and how many persons are employed, which types of products and services are produced, and what is the total turnover?
- **Technical Infrastructure:** Including the penetration rates of landline and mobile telephone networks, the number of computers per inhabitant and the number of Internet connections (whether or not a country is ready to become an information-based society).
- **ICT Demand:** Which enterprises and individuals are using ICT products and services? Which technologies are being used, and why? What barriers hamper a country's integration in the global Information Society?

Despite the fact that the Information Society Statistics is a relatively new area, international organisations have already adopted a set of harmonized measurement standards and allied indicators, and put together the necessary methodological foundations. The Organisation for Economic Co-

DOI: 10.4018/978-1-5225-2255-3.ch408

Figure 1. The Conceptual Model of the Information Society

Source: [Gokhberg and Buegh-Nielsen, 2007].



operation and Development (OECD), the Statistical Office of the European Union (Eurostat), the International Telecommunication Union (ITU), and the United Nations Conference on Trade and Development (UNCTAD), the Partnership on Measuring ICT for Development (Partnership) play the leading role in this effort (Table 1).

Statistical standards are developed by international organisations' working groups in cooperation with national statistical services, in particular: the OECD Working Party on Measurement and Analysis of the Digital Economy, the Eurostat Working Group on Information Society Statistics, the ITU Expert Group on Telecom/ICT Indicators, the ITU Expert Group on ICT Household Indicators. Use of international statistical standards is usually voluntary; however, to ensure full-scale participation in international data exchanges, and to be able to assess the country's prospects in the

Table 1. Existing International standards for ICT statistics

Area covered by standards	References
ICT supply	
The ICT sector and content and media sector definition by economic activity in accordance with international classifications	OECD, 2011 UNCTAD, 2009
The ICT product definition by goods and services in accordance with international classifications (CPC, HS)	OECD, 2011 UNCTAD, 2009
Core list of ICT sector and trade in ICT goods indicators	Partnership, 2016
ICT infrastructure	
ITU Telecommunications/ICT Indicators	ITU, 2014 OECD, UNCTAD
Core list of ICT infrastructure and access indicators	Partnership, 2016
ICT demand	
ICT usage by businesses (enterprises) survey	Eurostat, 2016a OECD, 2015 UNCTAD, 2009
Core list of ICT access and use by enterprises indicators	Partnership, 2016
ICT access and usage by households and individuals survey	Eurostat, 2016a OECD, 2015 ITU, 2014
Core list of ICT access and use by households and individuals indicators	Partnership, 2016

context of global trends, most nations follow the international recommendations on measuring ICT.

INTERNATIONAL STANDARDS FOR THE INFORMATION SOCIETY STATISTICS

Information Society Statistics is one of the newest branches of social and economic statistics aimed at studying all aspects of activities related to production of ICT goods and services, proliferation and usage of ICT in the economy, social and public sectors, and private life (HSE, 2014).

Information Society Statistics has the following objectives:

- Regular collecting and processing relevant data;
- Studying ICT development trends, assessing policies aimed at supporting further development of the Information Society;
- Conducting international comparisons, calculating international rankings;
- Providing reliable, and comprehensive data to users;
- Publishing working papers, manuals, outlooks, data books and databases.

The main international body developing relevant methodologies, from the very beginning, has been the OECD. Several issues of the OECD Guide to Measuring the Information Society were published, summarising the experience accumulated over many years of studies (2005, 2009, the latest edition – 2011). The 2011 Guide presents basic concepts, definitions, and classifications, and includes recommendations on collecting internationally compatible statistical data in the following areas:

- **The Information Society:** The conceptual model in statistical terms;
- **ICT Products:** Definitions and classifications relating to ICT goods and services, measurement of international trade in ICT goods, prices and quality of ICT products;
- **ICT Infrastructure:** Access, quality, investments, tariffs;
- **ICT Supply:** The ICT sector and its impact, other suppliers of ICT, ICT patenting activity;
- **ICT Demand by Businesses:** The model survey of ICT usage by businesses includes proposed statistical standards for e-business and e-commerce, ICT investment and expenditure by business, and the economic impacts of ICT investment and use;
- **ICT Demand by Households and Individuals:** The model survey of ICT access and usage by households and individuals; includes a discussion of e-commerce

and the social and economic impacts of ICT use by households and individuals;

- **Content and Media Sector:** Statistical issues related to information and electronic content, and more recent work on defining the content and media sector and its products;
- **The International Scene and the Road Ahead:** Overview of the international activities on measuring ICT, and future challenges (OECD, 2011).

Currently most of the countries and many international organisations apply, though to a different extent, the OECD-proposed methodological approaches to statistical surveys and analyses of the ICT sphere.

Basic OECD statistical standards for measuring ICT usage are harmonised with Eurostat statistical methodology and data collection practices adopted by the EU countries (*Eurostat, 2016a*). In its turn, the European experience prompted further development of the OECD's and other international organisations' activities in this area.

The ITU provides definitions, which serve a basis for developing ICT infrastructure indicators. Its guide (ITU, 2011) is a key reference document for collection of internationally compatible indicators on telecommunications/ICT based on administrative sources (i.e. supply-side data mainly provided by operators). Due to the rapid development of the telecommunications sector, indicators have to be regularly adjusted. The impressive growth and transformation of the mobile communications and Internet sectors made it necessary to review the existing definitions. The current version of the Handbook includes definitions and methodological clarifications for 81 internationally agreed indicators and corresponding sub-indicators, discussed by the ITU. It fully reflects the main aspects of monitoring development of infrastructure: landline and mobile telephone communications, Internet, tariffs for communication services, investments in telecommunications, and other relevant indicators.

UNCTAD's activities include collecting and analysing national statistics, as well as providing consulting and training services to countries willing to produce internationally compatible ICT indicators. The Manual for the Production of Statistics on the Information Economy (UNCTAD, 2009) contains recommendations for national statistical offices on methodological and practical aspects of statistical surveys of ICT development and application in the economy. It includes definitions, data sources, and methodologies for calculating important indicators, questionnaires, and analysis of methodological issues commonly arising in the course of relevant statistical surveys, together with suggestions on dissemination of statistical information.

The international institutes promotes international comparability of indicators through them outreach activities. The Partnership on Measuring ICT for Development is a good example (Partnership, 2010).

The main activities of Partnership are developing and maintaining a Core List of ICT Indicators. Other activities include compilation and dissemination of ICT data, and provision of technical assistance to help statistical agencies collect data for the Core List of ICT Indicators. The latest version of the Core List is based on a supply/demand conceptual model of the Information Society and covers the following areas: ICT infrastructure and access; ICT access and use by households and individuals; ICT access and use by enterprises, ICT sector and trade in ICT goods; ICT in education; ICT in government (UN, 2016).

Now we observe explosive growth of online economic activities including Internet access provision and Internet-based services. So a new direction of the measurement activities is developed – the study of the contribution of Internet economy to GDP. The international studies can be classified to the following criteria: the direct impact of the Internet on the economy (BCG, 2014; McKinsey, 2011; Bakshi, 2016), indirect economic impact of the Internet (Stiglitz., Sen,

Fitoussi, 2009; BCG, 2010), and its impact on the social sphere (OECD, 2013; Shah, Kwak, Holbert, 2011; Morton, 2006; Greenstein, McDevitt, 2011).

INDICATORS OF ICT AND CONTENT-RELATED ACTIVITIES

All ICT and content-related activities (information economy) are divided into two sectors: the ICT sector and the Content and Media sector.

The central aspect of a statistical study of the ICT sector is identifying the ICT-related activities (industries), and defining of goods and services produced in this area. The OECD definition of the ICT sector is based on the UN International Standard Industrial Classification (ISIC).

The definition of the ICT sector has been updated a few times. The latest was in line with adoption of the modified ISIC Rev. 4 and introduced as an official alternative analytical aggregation. Now the activities (industries) in the ICT sector are grouped into ICT manufacturing industries, ICT trade industries and ICT services industries. Their list which complies with the above definition contains the following positions:

ICT Manufacturing Industries

- Manufacture of electronic components and boards;
- Manufacture of computers and peripheral equipment;
- Manufacture of communication equipment;
- Manufacture of consumer electronics;
- Manufacture of magnetic and optical media.

ICT Trade Industries

- Wholesale of computers, computer peripheral equipment and software;
- Wholesale of electronic and telecommunications equipment and parts.



ICT Service Industries

- Software publishing;
- Telecommunications, included Wired telecommunications activities, Wireless telecommunications activities, Satellite telecommunications activities, and Other telecommunications activities;
- Computer programming, consultancy and related activities, included Computer programming activities, Computer consultancy and computer facilities management activities, and Other information technology and computer service activities;
- Data processing, hosting and related activities; web portals, included Data processing, hosting and related activities, and Web portals;
- Repair of computers and communication equipment, included Repair of computers and peripheral equipment, and Repair of communication equipment (UN, 2008).

Design of the Classification of ICT products used the same principles. ICT product classification enables statistical agencies to produce two types of data: about production, based on the Central Product Classification (CPC), and about imports (exports) of relevant products, based on the Harmonised System (HS).

In addition to the ICT product classification, there is an obvious need for a harmonised classification of ICT services. The services classification became possible, based on the new CPC Ver. 2 (OECD, 2011), which provided a detailed structure of the CPC Ver. 2 subclasses linking them with the ISIC Rev. 4 and HS 2007.

An internationally agreed definition of the Content and Media sector (comprising organisations whose economic activities involve production, publication, and/or dissemination of content) was developed in 2007–2008 on the initiative of the OECD, based on the ISIC Rev. 4. The sector covered the following industries:

- Publishing of books, periodicals and other publishing activities, included Book publishing, Publishing of directories and mailing lists, Publishing of newspapers, journals and periodicals, and Other publishing activities;
- Motion picture, video and television programme activities, included Motion picture, video and television programme production activities; Motion picture, video and television programme post-production activities, Motion picture, video and television programme distribution activities, and Motion picture projection activities;
- Sound recording and music publishing activities;
- Programming and broadcasting activities, included Radio broadcasting, and Television programming and broadcasting activities;
- Other information service activities, included News agency activities, and Other information service activities n.e.c. (UN, 2008).

Today, core indicators on ICT sector and Content and Media sector are as follows: value added, share in total value added, employment, share in total employment.

INDICATORS OF ICT INFRASTRUCTURE

Statistical surveys of ICT infrastructure describe results of telecommunication activity (as a segment of the ICT sector), and of broadcasting organisations as a segment of the Content and Media sector. These include communication services output, revenues, and number of personnel.

ICT infrastructure indicators include: access to the Internet, telephone, telegraph, satellite communications; radio communications; volume of Internet and mobile traffic. These include numbers

of Internet users with mobile, fixed, and broadband access, mobile subscribers, etc.

Indicators measuring accessibility of communication services, such as tariffs for local landline telephone calls, mobile communications, and landline broadband Internet access, also play an important role.

Due to convergence of networks, services, and devices, indicators on ICT infrastructure now include television indicators. Development of the television segment is measured by availability of mandatory TV channels and the population's coverage by digital on-air and cable TV broadcasting.

These indicators are important not only to measure the level of ICT availability, their use for the estimation of productivity (Byrne, Fernald, Reinsdorf, 2016).

INDICATORS OF ICT USAGE BY BUSINESSES, HOUSEHOLDS (INDIVIDUALS), AND THE SOCIAL SPHERE (HEALTH AND EDUCATION)

The main areas of statistical surveying of ICT usage in the business sector are availability of ICT, objectives of the Internet usage, and results of such usage.

Statistical surveys of ICT usage by households and individuals measure availability of and access to ICT; use of mobile phones, computers, and the Internet (places of access, purposes and frequency of usage); ICT-related skills, etc.

Internationally applied indicators of ICT usage by individuals reflect differences between countries, and between various population groups in the same country, by the following parameters:

- Application of cutting-edge technologies to provide higher-quality Internet access;
- Use of opportunities provided by the Internet;
- Replacement of conventional forms of socialising, work, leisure, information

search, shopping, etc. with new Internet-based formats;

- Factors promoting or hindering ICT proliferation among the population.

Particular attention is paid to e-commerce and e-government indicators. Related aspects are analysed for enterprises and individuals.

E-commerce is the sale or purchase of goods or services conducted over computer networks by methods specifically designed for the purpose of receiving or placing of orders. The payment and the delivery of the goods or services do not have to be conducted online. E-commerce transactions exclude orders made by manually typed e-mail messages (Eurostat, 2015).

E-government is electronic contacts via the Internet with public authorities and some public services. Contacts through manually typed e-mails should be excluded. Contact and interaction with public authorities and public services include websites concerning citizen obligations (e.g. tax declaration, notification of moving), rights (e.g. social benefits), official documents (e.g. ID card, birth certificate), public educational services (e.g. public libraries, information on the enrolment in schools or universities), public health services (e.g. services of public hospitals) (Eurostat, 2015).

Surveying ICT usage in the social sphere, health, and education is also developing. Such surveys have a structure similar to those monitoring ICT usage in the business sector. They also cover certain specific issues, such as, in health—development of telemedicine, application of ICT to create and maintain medical registries and databases, provide electronic services to patients, etc.; in education, availability of various kinds of ICT in all types of educational institutions is measured, along with conditions for efficient application of ICT in teaching, ICT usage by teachers and students, introduction of new ICT-based teaching formats and techniques, etc.

All these standards use for estimating to indicators for monitoring and benchmarking on



sustainable development, international ranking, ICT development by region and country.

The Eurostat collected and analyzed data of the Information Society in the European Union (EU), focusing on the availability of ICT usage by individuals and within households (Eurostat, 2016b) and in enterprises (Eurostat, 2016c). The development of the Information Society is regarded as critical to meet the demands of society and the EU economy. The countries in other regions also show more active in implementing standards and shaping the data.

SOLUTIONS AND RECOMMENDATIONS

Despite this extensive methodological groundwork ICT development challenges of measurement.

First of all it relate of broadband Internet definitions. Broadband is defined as a service providing download speeds of at least 256 Kbit/s (ITU, 2011). In order to adequately reflect the development of high-speed Internet for broadband should consider starting with a speed of at least 2 Mbit/s.

Specifically, conducting surveys of ICT usage by households and individuals, international organisations typically consider all people living on the same premises, both related and unrelated to each other, as a household, if at least one of them belongs to the target population group. Recommended age range for the surveyed population is 16–74. Countries may collect data about individuals outside this age group (they might be particularly interested in ICT usage by people aged 75+). Many OECD member countries have actually extended this group and also survey people under 15. In about a half of the OECD member states the target group's age range was extended to 10–72. Some countries, such as Korea, Mexico, and the USA, even survey children aged 3–6. Many countries have also eliminated the upper age limit of the target group. Inclusion, the world over, of increasingly large population groups into the target group is due to the wide proliferation

of information technologies, and their penetration of people's everyday life at any age.

Some problems exist with collecting data for ICT infrastructure indicators. Currently, most of the ITU member countries provide data about revenues from all telecommunication services; from mobile communications; and data on investments in telecommunications. Meanwhile methodology for producing such data, and its availability remain a problem in a number of developing countries. In some cases, different methodologies may result in a significant distortion – up to 20–30%, which complicates comparing countries on equivalent basis and producing reliable global statistics.

FUTURE RESEARCH DIRECTIONS

Statistical approaches and indicators described above should be seen as basic ones. They are quite suitable for application in practically all countries. At the same time, ICT-involving processes are inherently so dynamic, and change so rapidly that statistical measuring techniques must be updated and improved all the time. There is a need to supplement conventional monitoring of Information Society with experimental indicators reflecting the Internet's impact on the economic growth (quantitative assessment of the Internet-induced openness's consequences in such areas as trade, innovation, entrepreneurship, and labour productivity); social wellbeing; countries' independence; their inclusion in the Internet environment; efficiency of digital resources' usage; managing digital security and privacy risk.

The Digital Economy now permeates countless aspects of the world's economy, impacting sectors as varied as banking, retail, energy, transportation, education, publishing, media or health. Information and Communication Technologies are transforming the ways social interactions and personal relationships are conducted, with land-line, mobile and broadcast networks converging, and devices and objects increasingly connected to forming the Internet of things.

To reflect the latest trends in the emerging Digital Economy, the Internet dissemination in all activities (horizontal penetration), digitalisation of each sector of the economy, and of every public service provision should be measured, together with development of broadband technologies, the Internet of things, personal data protection, children's network security and overall cyber-security, ICT impact on the environment, etc. (OECD, 2016).

The growth of ICTs has resulted in a rapid increase of new (including big) data sources, in particular from the ICT industry. The most international institutes with national experts are looking into innovative ways to utilize big data as a new data source, and to overcome important data gaps (UN, 2016). So it is very perspective way for development statistics.

CONCLUSION

Rapid development of ICT contributed to emergence of profound social and economic changes, new forms of socialising and interaction, and transformation of the whole society. An interest and a need to study this phenomenon arose, on national and international levels. At the same time demand for statistical data describing the new processes and trends has emerged, to support relevant political decision-making.

Official statistical data sufficient to monitor development of Information Society was always seen as an important tool for shaping international and national policies.

Three major sources of data describing ICT include the following: production statistics, telecommunications statistics, and ICT usage statistics. The first type of data, which characterises the ICT sector, Content and Media sector, and ICT products, is based on data of business statistics and foreign trade statistics. ICT infrastructure indicators are calculated on the basis of data provided by communication networks' operators. On the other hand, data on ICT usage are usually specially

collected and processed through targeted surveys designed to measure ICT usage by enterprises, households and individuals.

For compatibility and benchmarking purposes, it is important to adopt uniform methodological approaches to statistical measurement. This aspect is given particular attention when international standards are developed. Significant progress in measuring ICT was made in recent years, including monitoring of new technologies' and communication services' proliferation, horizontal penetration, and structural changes they engender. New study areas have emerged, allowing to measure the major transformations brought about by development of ICT.

Basic statistical standards actually applied by international organisations and various countries include the following:

- The ICT sector and Content and Media sector definition by economic activity in accordance with ISIC Rev. 4;
- The definitions and classifications relating to ICT goods and services based on CPC Ver. 2, HS2007;
- Apply other branch statistics (business statistics, foreign trade data, the system of national accounts, etc.) to calculate ICT supply and infrastructure indicators;
- A flexible data collection system; providing of standard questionnaire for specialised businesses and households (individuals) surveys;
- Adoption of agreed definitions (including for e-commerce, e-government, etc.);
- Development of internationally harmonised statistical indicators systems for measuring ICT, and principles of their calculation allowing to produce internationally compatible statistics, conduct international benchmarking, and analyse global information development trends.

However, given a rapid progress of ICT, its deeper and deeper penetration in all areas, and



comprehensive impact on economic and social development and the quality of life, there is an obvious need to further develop Information Society Statistics at both national and international levels to study new technologies and develop new data collection techniques, Big Data methods.

REFERENCES

- Ahmad, Schreyer, & Wölfl. (2004). ICT Investment in OECD Countries and its Economic Impacts. *The Economic Impact of ICT – Measurement, Evidence and Implications*. Paris: OECD.
- Bakshi, H. (2016). How can we measure the modern digital economy?. *Significance*, 13(3), 6–7.
- BCG. (2010). *The Connected Kingdom. How the internet is transforming the U.K. Economy*. Boston: The Boston Consulting Group.
- BCG. (2014). *Greasing the Wheels of the Internet Economy*. Retrieved from https://www.bcgperspectives.com/content/articles/digital_economy_telecommunications_greasing_wheels_internet_economy/
- Billon, M., Lera-Lopez, F., & Marco, R. (2016). ICT use by households and firms in the EU: Links and determinants from a multivariate perspective. *Review of World Economics*, 152(4), 629–654. doi:10.1007/s10290-016-0259-8
- Blank, G., & Groselj, D. (2014). Dimensions of Internet use: Amount, variety, and types. *Information Communication and Society*, 17(4), 417–435. doi:10.1080/1369118X.2014.889189
- Byrne, Fernald, & Reinsdorf. (2016). *Does the United States have a Productivity Slowdown or a Measurement Problem?* BPEA Conference Draft. Retrieved from https://www.brookings.edu/wp-content/uploads/2016/03/ByrneEtAl_ProductivityMeasurement_ConferenceDraft.pdf
- Dolničar, V., Prevodnik, K., & Vehovar, V. (2014). Measuring the dynamics of information societies: Empowering stakeholders amid the digital divide. *The Information Society*, 30(3), 212–228. doi:10.1080/01972243.2014.896695
- Eurostat. (2015). *Methodological Manual for Statistics on the Information Society*. Eurostat.
- Eurostat. (2016a). *Methodological Manual for Statistics on the Information Society*. Eurostat.
- Eurostat. (2016b). *Digital economy and society statistics – households and individuals*. Retrieved from http://ec.europa.eu/eurostat/statistics-explained/index.php/Digital_economy_and_society_statistics_-_households_and_individuals
- Eurostat. (2016c). *Digital economy and society statistics – enterprises*. Retrieved from http://ec.europa.eu/eurostat/statistics-explained/index.php/Information_society_statistics_-_enterprises
- Gokhberg, L., & Bøegh-Nielsen, P. (Eds.). (2007). *Information Society Statistics in the Russian Federation: Harmonization with International Standards*. Moscow: State University – Higher School of Economics.
- Greenstein, S., & McDevitt, R. C. (2011). The broadband bonus: Estimating broadband Internet's economic value. *Information*, 35(7), 617–632.
- HSE. (2014). *Information Society: 2014. Data Book*. Moscow: National Research University – Higher School of Economics.
- ITU. (2011). *Handbook for the Collection of Administrative Data on Telecommunications/ICT*. Geneva, Switzerland: ITU.
- ITU. (2014). *Manual for Measuring ICT Access and Use by Households and Individuals*. Geneva, Switzerland: ITU.
- ITU. (2015). *Measuring the Information Society*. Geneva, Switzerland: ITU.

McKinsey. (2011). *Sizing the Internet economy. Internet matters: The Net's sweeping impact on growth, jobs and prosperity*. McKinsey Global Institute.

Morton, F.S. (2006). Consumer Benefit from Use of the Internet. In *Innovation Policy and the Economy* (Vol. 6). Cambridge, MA: National Bureau of Economic Research.

OECD. (2011). *OECD Guide to Measuring the Information Society*. Paris: OECD Publishing.

OECD. (2013). *Measuring the Internet Economy*. Paris: OECD Publishing.

OECD. (2015). *Digital Economy Outlook*. Paris: OECD Publishing.

OECD. (2016). *OECD Ministerial Meeting on the Digital Economy: Innovation, Growth and Social Prosperity*. Paris: OECD Publishing.

Partnership. (2010). *Partnership on Measuring ICT for Development Core List of ICT Indicators*. Geneva, Switzerland: ITU.

Partnership. (2016). *Partnership on Measuring ICT for Development Core List of ICT Indicators*. March 2016 version. Retrieved from http://www.itu.int/en/ITU-D/Statistics/Documents/coreindicators/Core-List-of-Indicators_March2016.pdf

Shah, D. V., Kwak, N., & Holbert, R. L. (2011). 'Connecting' and 'Disconnecting' with Civic Life: Patterns of Internet Use and the Production of Social Capital. *Political Communication*, 18(2), 141–162.

Stiglitz, J.E., Sen, A., & Fitoussi, J.-P. (2009). *Report by the Commission on the Measurement of Economic Performance and Social Progress*. Commission on the Measurement of Economic Performance and Social Progress.

UN. (2008). International Standard Industrial Classification of All Economic Activities (ISIC). The fourth revised edition. Statistical papers. Series M No. 4/Rev.4. New York: United Nations.

UN. (2012). *Forty-third session of Statistical Commission. United Nations. Item 2 of the provisional agenda. Economic and Social Council, 28 February – 2 March 2012*. New York: UN.

UN. (2014). *Forty-fifth session of Statistical Commission. United Nations. Item 3 (g) of the provisional agenda. Economic and Social Council, 4 – 7 March 2014*. New York: UN.

UN. (2016). *Report of the Global Working Group on Big Data for Official Statistics. Forty-seventh session of Statistical Commission. United Nations. Item 3(c) of the provisional agenda. Economic and Social Council, 8 – 11 March 2016*. New York: UN.

UNCTAD. (2009). *UNCTAD Manual for the Production of Statistics on the Information Economy*. New York: UNCTAD.

UNESCO. (2009). *UNESCO Guide to Measuring Information and Communication Technologies (ICT) in Education*. Institute for Statistics. Montreal, Canada: UNESCO-UIS Publishing.

ADDITIONAL READING

European Commission. Digital Agenda Key Indicators. Retrieved July 20, 2016, from http://digital-agenda-data.eu/datasets/digital_agenda_scoreboard_key_indicators/indicators

ITU. (2014) New telecommunication/ICT indicators from administrative data sources, 2011–2013. http://www.itu.int/pub/D-IND-ITC_IND_HBK-2011 (Accessed 21.10.2016).

OECD. (2014). *Measuring the Digital Economy: A New Perspective*. Paris: OECD Publishing.

Prakash, S., Baron, Y., Liu, R., Proske, M., & Schlossler, A. (2014). Study on the Practical Application of the New Framework Methodology for Measuring the Environmental Impact of ICT—Cost/Benefit Analysis. European Commission, Brussels, Studie.

Schlichter, B. R., & Danylchenko, L. (2014). Measuring ICT Usage Quality for Information Society Building. *Government Information Quarterly*, 31(1), 170–184. doi:10.1016/j.giq.2013.09.003

UN (2014). UN E-Government Survey: E-government for the future we want. New York: UN.

UN. (2016). Forty-seventh session of Statistical Commission. United Nations. Item 3 (i) of the provisional agenda. Economic and Social Council, 8 – 11 March 2016. New York and Geneva: UN.

UNCTAD (2015). Information Economy Report – Unlocking the Potential of E-commerce for Developing Countries. Geneva, Switzerland: UN.

Waller, P., Irani, Z., Lee, H., & Weerakkody, V. (2014). Lessons on Measuring e-Government Satisfaction: An Experience from Surveying Government Agencies in the UK. [IJEGR]. *International Journal of Electronic Government Research*, 10(3), 37–46. doi:10.4018/ijegr.2014070103

KEY TERMS AND DEFINITIONS

ICT Sphere: Supply, demand and usage of ICT.

ICT Usage Statistics: A sub-area of Information Society statistics studying all aspects of

activities involving application of ICT in the business sector, social sphere, public administration, households, and by individuals.

Information and Communication Technologies (ICT): Technologies using microelectronics for collection, storage, processing, retrieval, transmission, and presentation of data, texts, images, and sound (HSE, 2014).

Information Industry Statistics: A sub-area of Information Society statistics studying all aspects of the ICT and content-related activities.

Information Society Statistics: The newest branch of social and economic statistics aimed at studying all aspects of activities related to the production of ICT goods and services, distribution and usage of ICT in the economy, social and public sectors, and private life (HSE, 2014).

International Statistical Standards for Measuring ICT: A set of internationally accepted recommendations on statistical surveying of Information Society, specifying requirements to statistical data and describing methodology for building basic indicators, based on harmonisation of national statistical standards.

International Statistics: Harmonised statistical data collected and published by international organisations based on standardised programmes, to conduct regional and international comparisons.

Information Technologies and Social Change



Muhammet Ali Köroğlu

Uşak University, Turkey

Cemile Zehra Köroğlu

Uşak University, Turkey

INTRODUCTION

Change is an inevitable fate for all the communities of the earth. The thing making social change today more important is the speed of this change. In previous historical periods of humanity, changes lasting for centuries today occur in only a few years. On this scale, change requires a dissolution and restructuring from the interaction between generations to all social institutions. To adapt to such radical changes is not easy for any society. In fact, components of the social structure are protectionist and change slowly. Structures such as the culture, family, understanding of religion, economy, politics, and education do not change easily. But today, it is seen that all of these institutions change rapidly. Certainly social change depends on many factors such as cultural, environmental, demographic and so on. This is a natural result of the interaction between man and nature. It is the process of adapting human to life and nature. Thus, a slow and controlled change occurs. However, through the development of communication technologies in the 21st century, in terms of speed and direction, social change takes an unpredictable state.

The industrial revolution, affected mostly humanity and forced it to change via communication technologies. With the Industrial revolution, especially in Western Europe, the modern type of society emerged. Modern society organizes in hegemony of reason and scientific knowledge.

The idea of progress is dominant depending on the Enlightenment philosophy. Accordingly, societies progress continuously from primitive towards the modern. Modern society is the last stop. Industry creates an urban life depends on factory rather than a rural, earthbound life. Economics becomes an activity that occurs outside the households. Family and education change in structure and function. Politics ceases to be an event in king's palace. Relations between the governing and the governed, re-establish in the framework of citizenship.

Enlightened every citizen's political participation within the framework of wisdom and knowledge are accepted. Another important development emerged with the Industrial Revolution and modern society in mass media. Developing since the mid-19th century, high-circulation newspapers and magazines became the most important agents of change in modern society with the mass media such as the radio, cinema, and television. Especially since the 1920s, the role of the media and mass media have been discussed in influencing human attitudes and behavior. Thus, they have affected individuals and communities from the health to leisure, from the consumption to religion and family. They have accelerated social change in an unpredictable manner. Therefore, converter effects of communication technologies on the social structures and social relations will remain as the investigation area of social sciences for a longer time.

BACKGROUND

Classical sociology seeks social change in the relationship between social structures. Accordingly, an element of social structure affects the another structure by changing. Institutions such as religion, economy, family, and education create a social stability. As the first sociologists of modernity, Simon, Comte, Durkheim, Marx, Spencer and, Weber, in order to discover the laws of social change, try to produce information from social phenomena. Thus, an extensive literature occurs on the issue. In this context, theoretical approaches such as structuralism, conflict theories, evolutionism, functionalism and so on. can be mentioned (Comte, 2001; Weber, 2000; Durkheim, 2006).

In addition to these grand theories, with the development of communication technologies, theories about the culture created by these tools and its impact on human behavior and society remain. Studies on the impact of communication technologies on society and human relations gained momentum especially between the two world wars. Theoretical approaches such as Syringe Model, Limited Effects Approach, Usages and Gratifications Approach and so on. Approaches such as Criticism of Culture Industry of the Frankfurt School is important for literature. However, the approaches expressed in this literature deal with the impact of information and culture transmitted via information technologies on human. It is criticized that information and culture transmitted by the media and the mass media deactivate the society and the society is forced to consume a culture which is not produced by itself.

However, the internet and social networks opened to the public after 1985, giving the opportunity to participate. And their impact on society and human behavior is much deeper. They changed the traditional acceptance about crime, family, education, and leisure completely. Today, no social institution, is as effective on people and society as information technologies. Therefore, the interaction of communication technologies and society, should be re-evaluated based on existing

literature. Thus, it can be understood that how every social structure influenced and transformed by this process. This is the aim of this study.

THEORETICAL APPROACHES

When analyzing social sciences, the type of modern society presented after Enlightenment and Industrial Revolution have not sufficiently recognized the importance of communication technologies. Until the use of the Internet as public, as a means of mass media, newspapers, magazines, radio, and television have mediated a one-sided communication. Today, Internet and social media emerging attached to it, have come to the fore. Internet and social networks enclose social life in every aspect and the importance of it is increasing with each passing day. Thus, the platform created by the communication technologies become the center of life. There are significant effects of communication technologies on the images of human about life and world. Realities are created largely through the media in social and political life (Ergül, 2013, p.165).

Media concept includes both newspaper, and television as yesterday's communication means and internet and social networks. Studies deal with relations between means of mass communication and society have gone through several points. Until 1940, there was a strong belief that the media is a very powerful and has a convincing role is dominant. The dominant approach in this period is the Syringe model. Accordingly, media inject messages into the human brain like injecting drugs into the body (Gökalp, 2013, p. 27). Messages transferred by mass media to the recipients affect their behavior. Thus, to impress, to mobilize, to manipulate the masses and to change their attitudes, and to create perceptions are possible. According to this approach, one who control the mass media has the power to divert the masses as wished.

The first studies on the mass media were concentrated during the two world wars. To be under-

stood of importance of the propaganda during the war had an effect on this situation. Propaganda used by the parties in the war, had been effective on soldiers and civilians. This led to a separate propaganda war (Kağıtçıbaşı, 2012, p.192). This theory is a Social Psychological adaptation of Behavioral Approach in Psychology. Accordingly, the individual receives the transmitted message, learns the message, and reveals attitudes and behaviors accordingly. Hence, especially in the marketing field, there has been a number of studies on the power of advertising on consumer behavior.

Another important era in the field of communication technologies is the period between the 1940s and the 1960s. During this period, the political atmosphere in the world softened and the totalitarian threat passed. With the rise of the welfare state and economic relaxation, the service sector became widespread. Official monopoly began to disperse in communication technologies. Under these circumstances, instead of “Approach of Unlimited Power of Communications Technology”, “Understanding of the Limited Effects” developed. According to the “Limited Effects Model”, communication technologies reinforces the already existing attitudes and behavior in the receiver. The audience developed mechanisms such as selective attention, selective perception and selective recall against the message transmitted via communication technologies. So the audience protected himself against a message conflicting with his idea, attitudes, and behaviors (Türkoğlu, 2010, pp.113,114).

Another approach that emphasizes the limitations of the impact of communication technologies is the “Uses and Gratifications Approach”. In this model developed from a functionalist perspective, communication technologies and given messages are less important. In contrast, it is important that people use communication technologies to meet their needs. The audience is not in a passive consumers position. On the contrary, according to their needs, act selectively (Mutlu, 1994, p.140). According to this, different people can use the technology for different purposes in different ways.

This and similar approaches about communication technologies have been developed in conjunction with social theories such as Functionalism and Pluralism. Both approaches are not critical to society and politics. Functionalism has installed the role of the transfer of values to preserve a stable society to communication technologies. However, Pluralism has assumed that different groups in society can make their voice heard through communication technologies (Gökalp, 2013, p.31). Although Functionalism is a classical Sociological theory, Pluralism is specific to modern society. According to this approach, through communication technologies, each social formation from political parties to pressure groups or minority groups can express themselves. Communication technologies can not be the monopoly of a single group, therefore, groups competing with each other can express themselves.

Thus, communication technologies, along with other structures of society can produce and protect values which create collective consciousness and make it alive. Therefore, in general, media and privatization of the media, even to be dependent of media to some economic powers through advertisements have not been considered as problems. In terms of analysis of culture created by communication technologies, The Frankfurt School has a very important place. The Frankfurt School represents a critical tradition. This criticism is not limited by media or communication technologies. In fact, the Frankfurt School makes a comprehensive criticism of reason, science, and social world based on Enlightenment. In this context, means of mass communication are considered as instruments of culture industry. Culture, art and information transmitted by the communication technologies are produced for mass and consumption. Receivers consume this culture passively. Before everything this is an economic sector. It works according to the principles of profitability. This relationship elicits a quality problem.

According to this approach, the industrial production of cultural heritage commodify culture. Modern civilization ascribes to everything the



similarity. Culture industry to meet the numerous requests produces standard goods (Bal, 2013, p.108). The goal here is to force individuals into appropriate consumption patterns which belong to their status groups and to justify the existing economic structure. Additionally, technical, economical, and organizational requirements also determine the production of culture (Atiker, 1998, p.52). Furthermore, mass culture has created artificial and incorrect needs in individuals through the culture industry. As a result, it turned into an anonymizer ideology passivating masses (Stauth & Turner, 2000, p.407). All mass culture under monopoly is an identical culture and at the same time, It is a bad culture as a result of the combination of culture and entertainment.

Between advertising and culture industry has formed an identity in terms of being orientation methods. With the victory of the advertisement on cultural industries, consumers have begun to feel compelled to use the products of this industry (Bottomore, 2013, p.63). Thus, the culture industry, instead of directing consumers to think, has accustomed them to accept the ready. As a result of this, individuals have become passive tools of mass culture. Different communities living in different regions of the world have already begun to consume the products of the culture industry. This has meant the erosion of cultural diversity and one-dimensional human.

Culture industry criticism of the Frankfurt School is still worth discussing in many ways. However, there has been new advances in communication technologies after the 1970s. The keyword of this advancement known as Information Revolution has been “combination”. With advances in the information storage and its operating, large quantities of information could be shared instantly with systems such as advanced satellite and cable (Kejanlioğlu, 2004, p. 81). Thus, computers, in a very short period of time, have become indispensable by wrapping human life. To meet the needs of people, many versions of the programs and softwares have been produced. Development is still ongoing and there is no limit.

Through computer technologies, the Internet and social networks have become involved in human life, so everyday life has almost been moved into the virtual world. Internet is an alternative environment to the real-life world where many relationships and efforts from shopping to information access, education to artistic efforts, interactive relationships to images and information stocks have been experienced (Esmer, 2011, p. 249). Internet and social networks are different from the communication means such as newspaper and television. Such mass media is one-way and has a centralized structure. The internet is not centralized. Communication is bilateral. The user is not only receiver and consumer but also producer and distributor. Internet, due to its interactive nature, is a constantly transforming and reshaping communication environment (Gökalp, 2013, p. 47).

Therefore, the reality produced in such a communication world is temporary, rootedless and recurring. In fact, according to Baudrillard (Baudrillard, 2014, pp. 110,112,160), it consists of simulation because now the reality is not produced as rational, depending on an origin; it is reproduced through an operational model. In today's world where the principles of simulation is decisive, the reality only proves the existence of the model. Additionally, more information and news are produced every day, but less meanings are produced. Communication technologies destroying the contents produced by themselves are also destroying the social area. Now, virtual and reality are mixed and virtual has become the preferred to the reality. In short, simulations are more real. Thus, the world of everyday life is lost and society is getting individualized.

SOCIAL INSTITUTIONS AND TECHNOLOGY

How does such a communication world affect the basic structures of the society? Today, in this context, it is necessary to concentrate and do more

studies. Indeed, institutions such as religion, family, economy, education, politics and leisure are subjected to a conversion. Direction and end of this conversion is non-specific.

When evaluated specifically, family is a universal institution enabling the continuation of generation and socializing people. Family unity, for both sides, refers to a meaning and value above all. Partners come together to combat all kinds of challenges that life confronts with them. In today's world of communication, family has been reduced to contractual, spatial, and sexual partnerships. The rate of divorce is almost more than the rate of marriage. Of course, in this case, the share of culture industry is huge. All over the world, in works produced by the culture industry, eternal love, eternal happiness, unlimited romanticism are offered as the values of marriage. With marriage, couples understand that eternal love is a literary fantasy and then they begin to complain about lack of understanding. Also many researches show that intensive use of communication technologies has negative effects on the level of satisfaction in marriage. Indeed, the purpose of use of new communication technologies such as social media by couples represents different meanings in terms of gender norms. (Dew & Tulane, 2015, p. 626; Sekam, 2011, p. 274). Therefore, the purpose of use of communication technologies is an issue that should be questioned.

Today, divorce is one of the most serious family problems. There has been a significant increase in divorce rates throughout the world. Divorce leads to serious personal and social problems for individuals (TUIK, 2006). Feeling of failure, loneliness, concerns of being unloved, and stigma are some of these consequences. If they have children, this leads to other problems and challenges (Zastrow, 2013, p. 261). Therefore, the social functions of the family are fading. In contrast, cohabitation, virtual wedding, virtual infidelity, virtual love will be more valid. Still, current functions of family such as continuation of generation, socialization of the new individuals, providing emotional satisfaction will be more

minimized. In fact, with today's technology it is possible to live love someone in Asia with who living in America. The same person can find a surrogate mother who lives in Europe, send the sperm to have a child and grow his child through social services. Considering that children socialize via internet more than family, this will be even more rational choice.

Another social institution affected by the communication technologies is politics. Politics is a universal and fundamental institution providing the public order and performing general administrative. As such, it is a parallel phenomenon to the existence of humanity (Aydin, 1997, p. 143). In the very long history of humanity, politics was an event closed to the participation and limited by the palace surroundings. However, with the development of written communication technologies, political issues have become topics that more people are interested and have information about it.

Governance has been forced to be more transparent than in the past. Especially after the development of radio and television broadcasting, governance has been able to hide less things. With the development of Internet and social networks, state secrets have decreased more. However, this does not mean that everything is transparent. Regarding the issue, Habermas considers that communication technologies are tools for both formation and transformation of public sphere which is a political participation area. Initially, newspapers and publishing houses are tools providing public expression of ideas. However, later, news has become itself an economic value and these tools are isolated from the public contents (Habermas, 2002, p. 318, 319). However, today, there are numerous comments that Internet contributes to the spread of the pluralism.

Accordingly, the Internet increases social participation. In this regard, communication technologies are great equalizers (Bozkurt & Baştürk, 2013, pp. 411, 412). In contrast, those who disagree with techno-policy express that the arguments in favor of electronic democracy do not understand the nature of politics. Democracy



includes political negotiations, judgment, and face to face discussions. This negotiation occurs under the guidance of rational reasoning and principles of political discourse and takes shape in a mutual discussion aiming to reach a consensus. However, all communications and discussions arising from the media are the death of policies based on rational negotiations. Additionally the internet is a distorted form of political debate drawn into the manipulative, extreme, and sensational (Kellner, 2004, p. 716).

Theoretically, people from every class, status, ethnic, or minority groups can share and defend all kinds of ideas on the Internet independently from the state repression and legal limits. This is a pluralism. However, this does not always create a rational negotiation or a social consensus. Nevertheless, the virtual environment can mediate to the opposition and restrict government's actions. In this sense, the control of the virtual environment is virtually impossible. For example, during the process of the Arab Spring, the opposition was organized via social media and able to conduct mass actions. In this respect, social networks allow the creation of virtual congregations and communities exceeding color, status and national boundaries. Therefore, there will be many social movements on a global scale in the future.

Similarly, economic relations will continue to be made over the Internet and face-to-face relations will be further reduced. The culture created by ICTs will continue to determine people's need for goods and services and their consumption preferences. Capitalism and its values will be more widespread on a global scale. How this process will affect the global economic inequality is unclear. Likewise, educational institutions will be forced to move away from their traditional format. Options such as education at home and distance learning will occur more smoothly. The institution affected by the development in communication technologies as a minimum will be religion.

Although religious insights and congregation patterns are influenced by this process, religion is still the source of truth and meaning for many

individuals and society. Communication technologies can increase pluralism in religious thought, contribute to the development of religious thought and create new religious congregation patterns. Discussions about the truth value of religion as a total system, still will continue to be carried out within the framework of the relationship between religion and science.

FUTURE RESEARCH DIRECTIONS

When evaluated generally, during the development of communication technologies, many studies have been conducted in terms of impact of message on the recipients. During the two world wars, controlling communication technologies was understood as almost to control the whole society. In determining human behavior, has been focused on the impact of communication technologies. But after the 1940s, especially with the development of Social Psychology, there has been more concentration on the multidimensionality of human behavior. The idea that the mass which is collocutor of the communication technologies is not completely passive has gained importance. Together with the emergence of the internet and widespread use of it, a new period has begun. The most important feature of this period is becoming communication interactive.

Studies on communication technologies are carried out in the framework of themes like advertising, marketing, management, attitude and behavior change and so on. However, more studies are needed to be done about the impacts of communication technologies on the basic institutions and structures of societies. In particular, a new era has begun with internet; technology and network has affected more all societies structurally. Therefore, there is a need for more research regarding impacts of communication technologies specifically on the family, religion, education, culture, politics and public policies. Similarly, transformation of everyday life should be evaluated in this context. Dimensions of communication technologies which

contribute to human freedom and creativity should be taken forward.

CONCLUSION

Communication technologies have shown a very rapid development since the mid-19th century. Specifically, through radio and television broadcasting, they have entered into people's lives more and gained an important place. Thus, the space and distances have lost meaning. About the impact of mass media on human life and society, the Frankfurt School has represented critical tradition. Social processes taking place on the concepts of mass society, consumption and culture industry has been criticized. This criticism has been in line with modern society's acceptance and values. Many of these criticisms still remain actual.

Real revolution in communication technologies has been realized through the publicity of internet that enables interactive interaction. Today, people spend a portion of their lives in front of TV screens, as they spend a large portion of them on Internet. Social relations are affected more than the new process. On one hand, individuals are lonely, but on the other hand they socialize on the internet. In this way, sense of loneliness is further increased. Society is more individuated, life world and the field of everyday affairs are narrowing. The most basic and universal institutions of humanity transform functionally. The future will be somewhere in the intersection of virtual and reality. Pessimism and hope for humanity have the same value. It is very difficult to predict the future or changes will be experienced. However, it is sure that all aspects of individual and social life will be affected and transformed.

REFERENCES

- Atiker, E. (1998). *Modernizm ve Kitle Toplumu*. Ankara, Türkiye: Vadi Yayınları.
- Aydın, M. (1997). *Kurumlar Sosyolojisi*. Ankara, Türkiye: Vadi Yayınları.
- Bal, H. (2013). *İletişim Sosyolojisi*. Bursa, Türkiye: Sentez Yayıncılık.
- Baudrillard, J. (2014). *Simulakrlar ve Simülasyon*. Ankara, Türkiye: Doğu-Batı Yayınları.
- Bottomore, T. (2013). *Frankfurt Okulu ve Eleştirisi*. İstanbul, Türkiye: Say Yayınları.
- Bozkurt, V., & Baştürk, Ş. (2013). Bilgi Toplumu ve Türkiye. In M. Zenzirkıran (Ed.), *Dünden Bugüne Türkiye'nin Toplumsal Yapısı* (pp. 407-420). Bursa, Türkiye: Dora Yayınları.
- Comte, A. (2001). *Pozitif Felsefe Kursları*. İstanbul, Türkiye: Sosyal Yayınlar.
- Dew, J., & Tulane, S. (2015). The Association Between Time Spent Using Entertainment Media and Marital Quality in a Contemporary Dyadic National Sample. *Journal of Family and Economic Issues*, 36(4), 621–632. doi:10.1007/s10834-014-9427-y
- Durkheim, E. (2006). *Toplumsal İşbölümü*. İstanbul, Türkiye: Cem Yayınevi.
- Ergül, H. (2013). Medya. In E. Gökalp (Ed.), *Sosyal Bilimlerde Temel Kavramlar* (pp. 164-194). Eskişehir, Türkiye: Anadolu Üniversitesi Yayınları.
- Esmer, H. (2011). Küreselleşme ve Dijital Çağ. In B. Gülmez (Ed.), *Kültür Tarihi* (pp. 138-262). Eskişehir, Türkiye: Anadolu Üniversitesi Yayınları.
- Gökalp, E. (2013). Medya, Kitle İletişimi ve Toplum. In N. Suğur (Ed.), *İnsan ve Toplum* (pp. 22-57). Eskişehir, Türkiye: Anadolu Üniversitesi Yayınları.
- Habermas, J. (2002). *Kamusal Yaşamın Yapısal Dönüşümü*. İstanbul, Türkiye: İletişim Yayınları.
- Kağıtçıbaşı, Ç. (2012). *Günümüzde İnsan ve İnsanlar*. İstanbul, Türkiye: Evrim Yayınları.

Kejanlıoğlu, D.B. (2004). *Türkiye’de Medya’nın Dönüşümü*. Ankara, Türkiye: İmge Yayınları.

Kellner, D. (2004). Tabandan Küreselleşme: Radikal Demokratik Bir Teknopolitikaya Doğru. In M. Özbek (Ed.), *Kamusal Alan* (pp. 715-735). İstanbul, Türkiye: Hil Yayınları.

Mutlu, E. (1994). *İletişim Sözlüğü*. Ankara, Türkiye: Ark Yayınları.

SEKAM. (2011). *Türkiye’de Aile Ailenin Yapısal Özellikleri ve Değişimi*. İstanbul, Türkiye: Sekam Yayınları.

Stauth, G., & Turner, B. S. (2000). Nostalji, Postmodernizm ve Kitle Kültürü Eleştirisi. In M. Küçük (Ed.), *Modernite versus Postmodernite* (pp. 405-424). Ankara, Türkiye: Vadi Yayınları.

TUIK. (2006). *Türkiye İstatistik Kurumu Aile Yapısı Araştırması*. Retrieved from www.tuik.gov.tr

Türkoğlu, N. (2010). *Toplumsal İletişim*. İstanbul, Türkiye: Urban Yayınları.

Weber, M. (2000). *Sosyoloji Yazıları*. İstanbul, Türkiye: İletişim Yayınları.

Zastrow, C. (2013). *Sosyal Hizmete Giriş*. İstanbul, Türkiye: Nika Yayınları.

ADDITIONAL READING

Adaklı, G. (2006). *Türkiye’de Medya Endüstrisi. Neoliberalizm Çağında Mülkiyet ve Kontrol İlişkileri*. Ankara, Türkiye: Ütopya Yayınevi.

Adorno, T.W. (2014). *Kültür Endüstrisi Kültür Yönetimi*. İstanbul, Türkiye: İletişim Yayınları.

Anderson, B. (1991). *Imagined Communities: Reflections of the Spread of Nationalism*. Londra. Verso.

Castells, M. (2009). *Communication Power*. UK: Oxford University Press.

Castells, M. (2010). *The Rise of the Network Society, The Information Age: Economy, Society and Culture-1*. UK. Oxford: Blackwell.

Dertrouzos, M. L. (1998). What Will Be: How the New World of Information Will Change Our Lives. New York: Harper Collins Publishers. Dur- sun, Ç. (2001). TV Haberlerinde İdeoloji. Ankara, Türkiye: İmge Yayınevi.

Habermas, J. (2001). *İletişimsel Eylem Kuramı*. İstanbul, Türkiye: Kabalcı Yayınları.

Kling, R. (2000). Learning About Information Technologies and Social Change: The Contribution of Social Informatics. *The Information Society*, 16(3), 217–232. doi:10.1080/01972240050133661

Meder, M. (2001). Bilgi Toplumu ve Toplumsal Değişim. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi*, 9, 72–81.

KEY TERMS AND DEFINITIONS

Cultural Structure: Spiritual aspect of social structure emphasizing human relationships.

Culture Industry: Industrialization of culture and commodification of cultural products.

Information Technologies: The tools that are generating, accumulating, and, protecting information.

Mass Culture: A form of culture produced by the culture industry.

Modernity: The name of values system emerged in the 17th century in Europe.

Social Change: The change of social structure, culture and, social behaviors in a society.

Social Structure: A set of organized relationships in a society.

*i*Schools Promoting “Information Science and Technology” (IST) Domain Towards Community, Business, and Society With Contemporary Worldwide Trend and Emerging Potentialities in India



P. K. Paul

Raiganj University, India

D. Chatterjee

Seacom Skills University, India

INTRODUCTION

There are so many opportunities that come with iSchooling in terms of courses, research, job potentialities and interdisciplinary skills. Though it deals with some problems and challenges; this paper is talks about such preliminary aspects, emerging trends and possibilities in Indian context with international agenda. The chapter is illustrated some possible programme and nomenclature in this context of People-Information-Technology interaction. This is treated as a healthy and sophisticated academic innovation in the field of information, knowledge, computing and technologies where some schools have been merged into a big academic unit or school. USA is a pioneer in this regard where several information schools such as Information studies, Computer science, Management Science, Knowledge Management have been merged. And in emerging context, altogether such combined subjects called as ‘Information Science & Technology’.

BACKGROUND AND ORIGIN OF I-SCHOOLS

I-Schools gained popularity during 1990’s when internet reached the common people and other commercial ventures. It is in this time the explo-

sion of information has lead the development of healthy information management approach. Earlier evolution of I-Schools provided such knowledge as Library Science, Communication Science, Mass-Communication, Computer Science and Information Technology. But the interdisciplinary research lead to a closer interaction among these departments and subjects and ultimately a new academic unit called I-Schools or Information Science/ Information Science & Technology (IST) Schools or Information Schools evolved. I-Schools gained popularity during 1900 -2000 (*Abeysekera, I. and Guthrie, J. (2004), and Agarwarl, Suren, 1989*). In this period many new I-Schools emerged not only in USA but also in Europe, and African countries, though the trend was much less in Australia and Asian countries including India. I-Schools in some universities evolved as new academic and research units. However in some of the universities it was established by merging the related disciplines with computing and/or information. I-Schools today has gained a separate recognition in the academic world due to its significant role and incredible benefits.

METHODOLOGY

It is a fact that, still research on nature & domain of Information Science is tough to find out in

DOI: 10.4018/978-1-5225-2255-3.ch410

academic community; hence it is urgent to find out the role of Information Science or Information Science & Technology (IST) in several community; and complete Social development by the review of literature and other mode of Scientific Research. Practically, it is still tough to dig out the latest area and field of Information Science and this research work will be helpful to find out such domain. To conduct this study various tools and techniques have been used, out of which review of literature played an important role. Web Review plays an important role to find out running courses in the field of Information Science & Technology (IST) and other information and related departments. To assess educational situation and make SWOT analysis of this field, some websites were searched; these are official websites of UGC; AICTE; Government of India. The MSc- Information Science curriculum, Courses and programme structure of more than 600+ educational institutes have been assessed during this study and the main link of UGC, AICTE and MHRD have been used to find out the latest of Education Information in India.

FEATURES OF I-SCHOOLS

The establishment on *iSchool* originated during 1988, when eminent educationist and Information Scientist, Toni Carbo, School of Information Science, University of Pittsburgh; Donald Marchan, School of Information Studies, Syracuse University and Richard Lytle, College of Information Science/Information Science & Technology (IST) and Technology, Drexel University joint hands for establishment of Information association called *iSchool* Caucus. After few years in 2001, the three members' team expands to 'Gang of Five' and then in 2003 some more dean of several Information related schools join the forum and as so August, 2015 total 65 institutes fall under the same. I-Schools or Information Schools are the academic units which are dedicated to the information solution through academic and research outputs. It is primarily responsible for the collection, selection, organization, processing, management

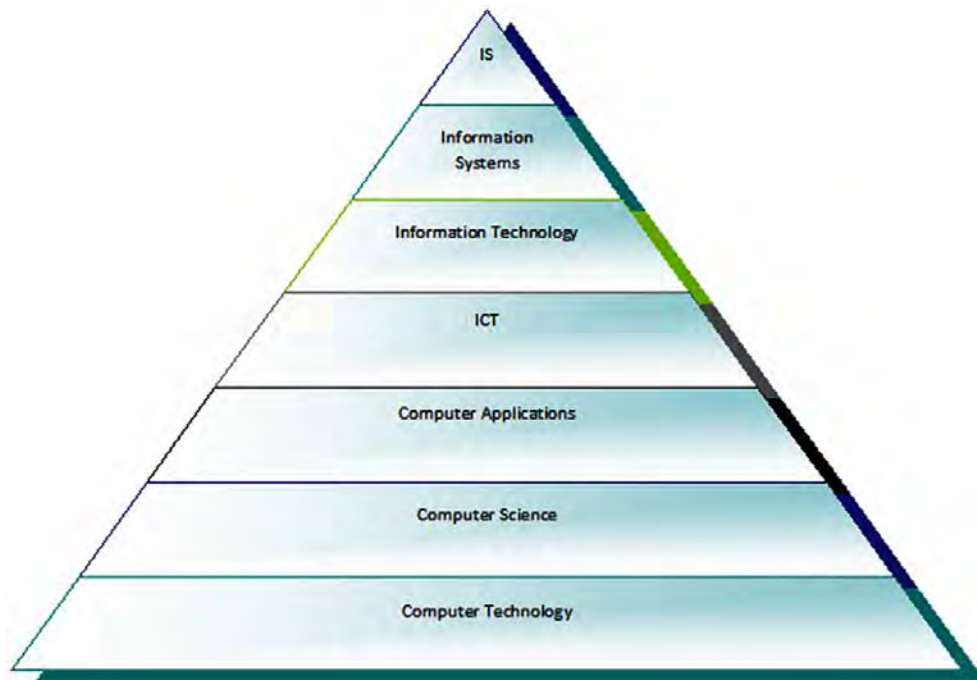
and dissemination of information with the help of Technologies. The main difference between IT Schools and I-Schools is that, IT Schools provide Technological Manpower and solution such as database, Networking, Software, Hardware and Core Computing, Multimedia and image processing, whereas I-Schools are responsible for the Information solution or knowledge Management solution with the help of Technologies such as Database Management, Networking Technology, Multimedia Technology, Communication Technology and Web Technology. Thus I-Schools are dedicated to information solution powered by technologies (*Balwan Singh, 2004 & Paul, b, Kumar, 2012,*). I-Schools are also responsible for Information and technological application in society, commerce, health and governance. Thus I-Schools deal much more with societies and believe in interaction between Information, Technology and Society Where as IT Schools deal just about application rather than societal approach. LIS schools deal with information solution in library and similar foundations, I-Schools deal and works with several domains and institutions (*Vickery, B.C., & Vickery, A. 1987 & Paul, c 2012*).

I-SCHOOLS AND I-PROGRAMS: EXISTING AND POSSIBILITIES

I-Schools offer so many programs but majority of them are in Applied Science. The flagship program in most of the academic units is Information Science and the degree as MS/MSc (Information Science/Information Science & Technology). However apart from these, I-Schools offer some specialized information programs also such as—

- **MSc/MS/BSc/BS:** Information Science/Information Science & Technology (IST) (Human Computer Interaction)
- **MSc/MS/BSc/BS:** Usability Engineering
- **MSc/MS/BSc/BS:** Digital Repository
- **MSc/MS/BSc/BS:** Communication Systems
- **MSc/MS/BSc/BS:** Database System

Figure 1. From Smaller and Broader periphery of Information and Technology related field



However some Universities also offer domain-based Information Programs such as Telecommunication Technology, Communication Systems, Multimedia Systems, with full fledged BS/MS degree (White, H. D., & McCain, K. W., 1997). Information Schools also offer domain based Information Programs as full fledged academic degree with specialization in Information related programs (Though Information Schools not at all offer conventional IT program such as Computer Science, IT, Computing, Software Engineering and so on.) (Buckland, Michael K and Liu 1995 and Bansal, 2005, C. Wang, 2010). The programs are as follows:-

- **BSc/BS/MSc/MS:** Information Science/Information Science & Technology (IST)
- **MSc/MS/BSc/BS:** Geoinformatics
- **MSc/MS/BSc/BS:** Medical Informatics
- **MSc/MS/BSc/BS:** Bio- Informatics
- **MSc/MS/BSc/BS:** Music-Informatics
- **MSc/MS/BSc/BS:** Library Management
- **MSc/MS/BSc/BS:** Chemo Informatics
- **MSc/MS/BSc/BS:** Quantum Informatics

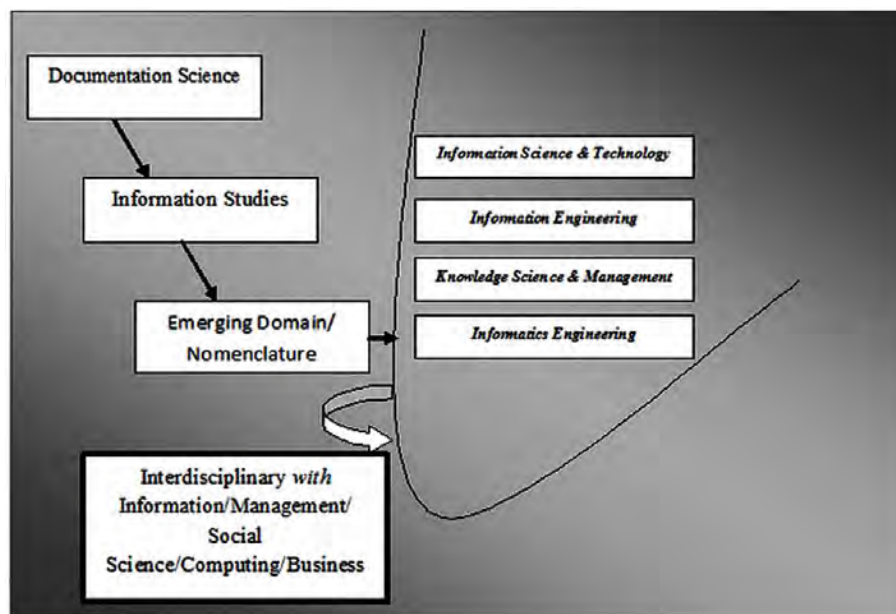
- **MSc/MS/BSc/BS:** Social Computing
- **MSc/MS/BSc/BS:** Social Informatics
- **MSc/MS/BSc/BS:** Archive Studies and so on

Thus Information Science/Information Science & Technology (IST) may possibly be offered in IT degree programs such as –

BCA/MCA (with domain based specialization such as Geo-Informatics/Medical Informatics / Bio-Informatics/Library Management/Bio Informatics/Quantum Information Science/Social Informatics/Social Computing).

Information Science/Information Science & Technology (IST) can also be offered as Fundamental program in BCA/MCA (Information Science/Information Science & Technology), which will be better for healthy computer Information Infrastructure Management (Wersig, G., & Neveling, U. 1975). It is possible to start new programs in I-Schools such as Engineering, BTech/MTech (IS) or BTech/MTech in the ECE/IT or related fields with IS specialization or it's any domain based specialization.

Figure 2. Advances of Information Programs with Interdisciplinary approaches



I-SCHOOLS IN INDIA: CURRENT SCENARIO AND FUTURE POTENTIALS

It is a fact that in India any strategy has not been at all adopted by the universities and higher Educational Institutes. India is one of the largest education hub in the world with more than 600 universities, including Central, State, Private and Deemed universities. Apart from universities, Institutes of National importance such as NIIT's, IIT's, IIM's, over 38000 Colleges serving for Indian students. Today nearly 5000 Engineering colleges and Polytechnics also provide oxygen to technical manpower and technological solution (Paul, f 2012). In India, most available information related subjects are Information Science (very few), Library Science (very common), Communication Studies (mainly in universities), journalism and mass communication. However, some universities also offers corporate communication, Information Systems, Electronic Information Systems which also fall under the I-School periphery.

The most common program of Information Science/Information Science & Technology (IST) in India is Library and Information Science which actually deals with library, Information Centers and similar foundations, i.e., the course is based on Information Science and Library Science interaction. Around 300 universities offer LIS program under the nomenclature BLIS, MLIS and in some cases BLIB/MLIB with MPhil and PhD level. The LIS programs are available in Regular and Distance modes and also online and off campus modes in many universities. This program is suitable for library management and deals with so many library aspects including some Information and Technological Foundations. The manpower such as Librarian, Cataloguer, Classification-ers, Indexer, Abstractor, Reference In-charge, Documentation Officer, Information Officer are provided by Library and Information Science.

The other Information related programs like Communication studies are available in very few universities and provide jobs such as, Communication Officer, Corporate Manager, and Information Manager. The Journalism and Mass Communica-

tion course provide manpower to the field of news paper and communication industries (Cohen, E. B., & Malgorzata, N.2006; Martin, 1998).

Thus, in India most of the Information programs are concerned with Library or communication industry. The general biased-free information program 'Information Science/Information Science & Technology (IST)' is available only in some of the institutes. Among them, BIT, Mesra is the first one who introduced MSc Information Science. The same program is offered by MCIS, Manipal University, Karnataka, IEM-Salt Lake, KITM-Buniadpur, DATM-Alipurduar, Techno India-Salt Lake, Techno India-Hooghly, DSMS- Durgapur. Periyar University is another which offers MSc (Information Science) and BSc(Information Science). This is the only university which offers BSc (IS). However depending upon application and special condition some universities permit admission in MPhil/PhD- Information Science. Most Universities offer individual courses and course gradients. Information Science (IST) in MCIS and Periyar University are biased with IT and Computing rather than main facet of I-Schools Information or knowledge. However BIT, Mesra offers MSc- Information Science which is based on both computing and Information Fundamentals (Such as indexing, abstracting, cataloguing and classification), Thus students becomes eligible in both library related jobs opportunities as well as in the Computing positions which are based on information. Information Scientist, CTO, CIO, IN, Network Administrator, System Administrator, Web Administrator, and Usability Engineer are the main job positions which may be availed of by the Information Science students. But the supply of students is very insignificant in India.

Information Schools with I-School flagship/ nomenclature are run by International School of Information Management (ISIM), Mysore University which offers a special information focused M.Tech. Programme i.e. M.Tech.-Information System and Management first time in Asia.

Specialized domain based Information Programs such as health Informatics, Medical Infor-

matics, Quantum Information Science, Chemical Informatics, Geo-Informatics are offered in very few educational institutes. However such programs are not offered in I-Schools or Information Science/Information Science & Technology (IST) Programs/ Departments. The programs are mainly offered in the related fields such as Geo Informatics, in Geo Science/ CSE/IT departments, Bio Informatics, Health Informatics, in Medical Programme. Medical Informatics, Chemo-Informatics, Quantum Informatics, Music Informatics are not offered in any of the universities.

WAY TO BUILD I-SCHOOLS IN INDIA

Information Science/Information Science & Technology (IST)/IST is very much diverse in nature and focused with so many domains and challenges its style depending upon need, place, country, school & international nature. Though, this research based chapter work is included to find out latest on IST/IS worldwide listed school (listed by Asia Pacific iSchools consortium); which is also indirectly helps to design programme on IST/IS in Indian concentration. But focuses is on designing with MSc/BSc programme. Though still potentiality is exists on nomenclature with BE/ME/BTech/MTech and so on.

Information Science/Information Science & Technology (IST) (and similar) as a domain become more popular and emerging with both Computational and traditional point of view. Depending upon demand and country the nature of Information Science/Information Science & Technology (IST)/IST are different with several nomenclatures. In this chapter review is carried out in the context of Information Science/Information Science & Technology (IST) as a programme; which is mainly concentrated on Asia and India. For critical analysis of Information Programs and domain so many nomenclatures are handle. As Information Science/Information Science & Technology (IST) is a big domain (combine with Computer Science, Computer Application,



Computer Engineering, Information Systems, Information Technology, Information Studies, Management Information Systems, and other domain); thus, in this chapter computing related programs are also included in Indian context. Chapter mentions about some IST Courses or Programs which are available with 'Information Science nomenclature throughout India Information Schools in India can be built institutionally or supported by the Government or Universities for the tremendous importance of information and knowledge.

Though Information Schools can be built by merging related departments of the universities, such as IT, CSE, CA, LIS, Mass Communication, departments of management and social studies may also be involved in I-Schools for better professional and research results. A full-fledged I-School/department may be established with some fulltime faculty and some adjunct professors from the IT, CSE, CA, Management departments depending upon the need. However, another approach may be starting of full-fledged Information Schools as University. In some countries universities with full-fledged Information Science programs have already been initiated. The Wuhan University of China has established a school focusing on IST. Instead of schools/department, in many countries special full-fledged universities have been established. For example, Tokyo University of Information Science & Technology (IST) & Technology, Japan; Beijing University of Information Science & Technology (IST) & Technology, China; University of Information Science & Technology (IST) & Technology, Ohrid-Macedonia. These Universities are the prime examples of IST development, application and movement. It is possible to establish a university with IT department, CSE, Management, Information Systems department with a strong focus on Information and Knowledge and Human Interaction with Technology and common people. Here the department will not only deal with conventional structure but also will be able to work like a big department. Thus it will promote interdisciplinary researches.

Engineering Colleges in India who are big players can start I-Schools or its related programs such as ME, ECE, IT, CSE which shall integrate each other. They can offer Information Science/Information Science & Technology (IST) as B.Tech./M.Tech.-ECE- Information Science & Technology (IST), B.Tech./M.Tech.-IT- Information Science & Technology (IST), B.Tech./M.Tech.-CSE- Information Science & Technology (IST).

I-SCHOOLS AND ENGINEERING INFRASTRUCTURE IN INDIA

India has sufficiently many Engineering institutes and academic units in Indian universities and degree colleges. It has nearly 5000+ technical institutes which include engineering colleges (*run under the affiliation of a university and approval of AICTE*), Polytechnics, Management and Administrative Colleges. Engineering Degree programs and general Degree courses are offered including 'BE or Bachelor of Engineering' / 'B.Tech. or Bachelor of Technology'.

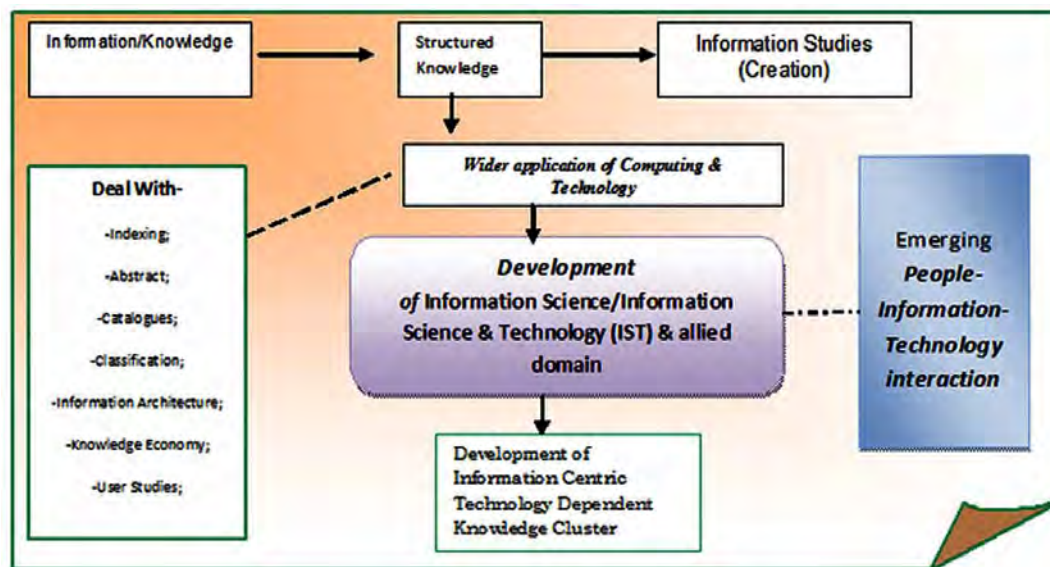
This degree is offered mainly in core Engineering subjects or Soft Engineering subjects. The most common subjects are

- Electronics and Communication Engineering (ECE).
- Electrical and Electronics Engineering (EEE).
- Electrical Engineering (EE).
- Civil Engineering (CE).
- Mechanical Engineering (ME).
- Aeronautical Engineering (AE) and so on.

However, apart from core subjects mentioned above some modern and interdisciplinary subjects are also offered in Indian Universities/ Engineering Colleges in their BE/BTech programs. These programme subjects are:

- Bio Technology.
- Structural Engineering.

Figure 3. The nature of Information Studies emphasizing origin of Information Science/Information Science & Technology (IST) (IS)



- Water and Resource Engineering.
- Environmental Engineering.
- Printing Technology.
- Geo Engineering and so on.

However, the mentioned subjects are offered in the Master of Engineering (ME)/ Master of Technology (MTech) programs with some specialization apart from the main subjects such as MTech-Electronic and Communication Engineering (*VLIS Design*). Here VLIS Design may be treated as specialization in Electronics and Communication Engineering (ECE). Thus this way some popular and flagship programs of Information Schools/ I-Schools may be introduced with Engineering Degree (*Like-BTech/MTech*). The possible degree programs may be:

- BTech/MTech- Mechanical Engineering (Information Science/Information Science & Technology (IST)).
- BTech/MTech- Electronics and Communication Engineering (Information Science/Information Science & Technology (IST)).

- BTech/MTech- Electrical and Electronics Engineering (Information Science/Information Science & Technology (IST)).
- BTech/MTech- Computer Science and Engineering (Information Science/Information Science & Technology (IST)).
- BTech/MTech- Information Technology (Information Science/Information Science & Technology (IST)).
- BTech/MTech- Software Engineering (Information Science/Information Science & Technology (IST)).
- BTech/MTech- Communication Engineering (Information Science/Information Science & Technology (IST)).

Some other engineering domains which are related to Information Schools such as Information Science/Information Science & Technology (IST)/ Information Systems/Information and Knowledge Management/Digital Media Communication can also play a pivotal role in this regard. Thus, it is possible to offer Information Programs only with the Engineering disciplines having Computer or Electronics or Information domains.

It is a fact that BTech/MTech Degree is still not offered with Information Science/Information Science & Technology (IST) as the main subject; but there is a huge possibility with 'Information Science/Information Science & Technology (IST)' or related information programs with Engineering Degree. If it is not a full-fledged department, then it can be offered with some other department such as Electronics and Communication Engineering/ Computer Science and Engineering/ Information Technology/ Software Engineering/ Mechanical Engineering and so on. Adjunct Professors may be a good alternative in this situation.

Virtually, MTech-Information Science/Information Science & Technology (IST) is also possible with so many specializations (*like BTech/MTech- Computer Science and Engineering offered with VLIS Design, Information Security, Embedded Systems and so on*).

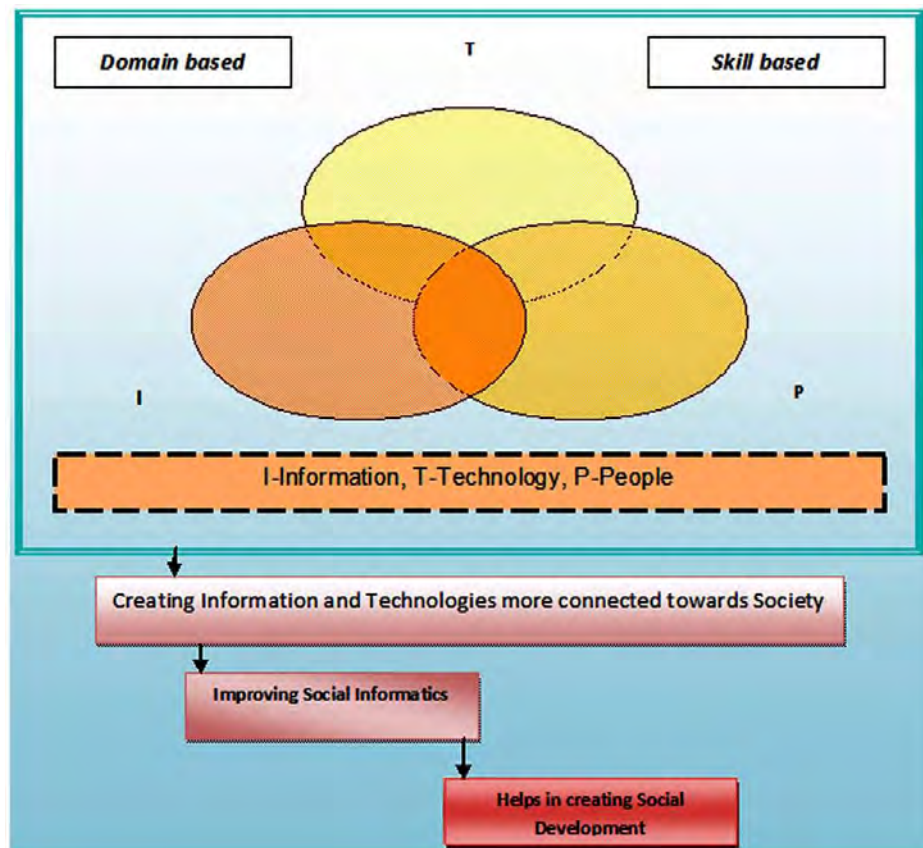
However, the concerned specializations need to be offered only with information and knowledge domain. Some possible degrees with Information Science/Information Science & Technology (IST) specializations may be as follows:

- BTech/MTech/BE/ME- Information Science/Information Science & Technology (IST) (Cloud Computing).
- BTech/MTech/BE/ME- Information Science/Information Science & Technology (IST) (Usability Engineering).
- BTech/MTech/BE/ME- Information Science/Information Science & Technology (IST) (Human Computer Interaction).
- BTech/MTech/BE/ME- Information Science/Information Science & Technology (IST) (Information Architecture).
- BTech/MTech/BE/ME- Information Science/Information Science & Technology (IST) (Interaction Design).
- BTech/MTech/BE/ME- Information Science/Information Science & Technology (IST) (Green Informatics).
- BTech/MTech/BE/ME- Information Science/Information Science & Technology (IST) (Geo-Informatics).
- BTech/MTech/BE/ME- Information Science/Information Science & Technology (IST) (Database Management).
- BTech/MTech/BE/ME- Information Science/Information Science & Technology (IST) (Multimedia Information Systems).
- BTech/MTech/BE/ME- Information Science/Information Science & Technology (IST) (Web Technology).
- BTech/MTech/BE/ME- Information Science/Information Science & Technology (IST) (Big Data Management).
- BTech/MTech/BE/ME- Information Science/Information Science & Technology (IST) (Knowledge Engineering).
- BTech/MTech/BE/ME- Information Science/Information Science & Technology (IST) (Networking Technology).
- BTech/MTech/BE/ME- Information Science/Information Science & Technology (IST) (Library Management).
- BTech/MTech/BE/ME- Information Science/Information Science & Technology (IST) (Decision Support Systems).
- BTech/MTech/BE/ME- Information Science/Information Science & Technology (IST) (ERP).
- BTech/MTech/BE/ME- Information Science/Information Science & Technology (IST) (Library Automation); and so on.

Apart from specialization in skill based practice subjects, BTech/MTech-Information Science/Information Science & Technology (IST) may be offered in some theoretical subjects as well such as

- BTech/MTech/BE/ME- Information Science/Information Science & Technology (IST) (Digital Dived).

Figure 4. The main components of Information Science/Information Science & Technology (IST) (IS) and its ultimate results



- BTech/MTech/BE/ME- Information Science/Information Science & Technology (IST) (Information Literacy).
- BTech/MTech/BE/ME- Information Science/Information Science & Technology (IST) (Information Systems).
- BTech/MTech/BE/ME- Information Science/Information Science & Technology (IST) (Information Resource Management).
- BTech/MTech/BE/ME- Information Science/Information Science & Technology (IST) (Knowledge Organization); and so on.
- BTech/MTech/BE/ME- Information Science/Information Science & Technology (IST) (Health Informatics).
- BTech/MTech/BE/ME- Information Science/Information Science & Technology (IST) (Bio Informatics).
- BTech/MTech/BE/ME- Information Science/Information Science & Technology (IST) (Geo Informatics).
- BTech/MTech/BE/ME- Information Science/Information Science & Technology (IST) (Chemo Informatics).
- BTech/MTech/BE/ME- Information Science/Information Science & Technology (IST) (Quantum Informatics).
- BTech/MTech/BE/ME- Information Science/Information Science & Technology (IST) (Medical Informatics).

As Information programs have connection with other domains, Information Science/Information Science & Technology (IST) in Engineering may also be offered in subjects such as:-

- BTech/MTech/BE/ME- Information Science/Information Science & Technology (IST) (Nuero Informatics).
- BTech/MTech/BE/ME- Information Science/Information Science & Technology (IST) (Business and Retail Information Systems).
- BTech/MTech/BE/ME- Information Science/Information Science & Technology (IST) (Social Computing/Informatics).

Thus Information Science/Information Science & Technology (IST) and other Information Programs (*Like- Information Systems/Information and Knowledge Management/Digital Media Communication*) may be introduced in Engineering colleges and universities with main Degree (like *MTech-Information Science/Information Science & Technology* or specialization of Information Science/Information Science & Technology (IST) in other Engineering Degrees like BTech/MTech- Electrical and Electronics Engineering (Information Science & Technology); or MTech-Information Science & Technology (IST) with specialization in core and allied Information Areas with skill based nomenclature (BTech/MTech/BE/ME- Information Science & Technology (IST) (Usability Engineering) /(Cloud Computing) or domain based nomenclature like BTech/MTech- (Health Informatics)/ (Quantum Informatics) and so on. Joint Venture, Collaboration, Adjunct Professorship may be good solutions at the initial stage to introduce these Information Programs in India; especially in Engineering domain. Thus I-Schools may get new oxygen this way.

CHALLENGES

This Research & Development is specially designed to find out latest on 'Information Science or Information Science & Technology including contemporary programme, nature, duration, specialization worldwide. But it is very much tough to find out latest scenario of Information Sciences

in universities worldwide (located more than 200 countries).

Hence here I-School organizations mainly chosen (which combines IST offering schools/programs worldwide); as they are not only consortium of Directory of 'IST/IS' offering schools but also included only such institutes who's agenda is interaction of 'Information-Technology-People'. Thus, this research work is based on this agenda; and only analyzes and, interpreted and includes programme/nature of Information Science/Information Science & Technology (IST) concentrated on interaction on 'Information-Technology-People'.

As far as Indian educational aspects are concerned, it is one of the largest country in the world with more than 35000 higher educational institutions/ colleges, 500+ Engineering/Technical Institutes, 700+ Universities and nearly two hundreds of centrally funded Institutes including IITs, NITs, IEST, IESR, a large number of institution under CSIR (12). Yet efforts to build I-Schools or information related academic unit/s programs are very less. Governmental supports, planning from educationalists, University level efforts is minimum. However many students of IT/ CSE are absorbed by the MNCs and other employers and then they make them ready for their job. But a simple way is to get ready-professionals is by introducing I-Schools or its related academic programs. Want of Planning, funding, efforts and less academic and professional associations in this field is the main reason for undeveloped or under developed situation of I-Schools in India.

FUTURE RESEARCH DIRECTIONS

Information Science/Information Science & Technology (IST) is very much diverse in nature and focused with so many domains and challenges its style depending upon need, place, country, school & international nature. Though, this research work is included to find out latest on IST/IS worldwide listed school. Information Science/Information Science & Technology (IST) is an important are



of Applied Science and Technology. This paper illustrates so many aspects of this domain and current situation and future possibilities in Indian and worldwide scenario. Here for future I-Schools, so many courses are proposed with related department and possible offering. Thus in future it is possible to design and develop each programme depending upon educational norms and requirements of various countries. 'Time Limit' is another limitation of the research work. The scope of the research is very broad; hence works may be carried out in so many sub divided areas depending upon further review, need and so on.

CONCLUSION

'Technologies are main pillar in this age' and if it is Information Technology then there is no doubt regarding the slogan. Internet was the last emerged tool of Computing and recently some more tools and name are shining such as Cloud Computing & Virtualization, Big Data Management, Intelligent Analytics, Green Computing & Technology (N. Chandrakant, 2011 & 2013). And all such fields are need to incorporate in the Information Science/Information Science & Technology (IST) and or allied domain. Information Science in general may help concerned field in many ways. It is able to develop a relationship between information-technology and people or stakeholders of the concerned field. With thousands of educational institutes it is possible to build independent I-Schools separately but if it is impossible initially it may be built by joint ventures with other departments. Employers, associations need to develop awareness to give pressure for starting I-Schools to prepare professionals from the I-Schools who are not only able in information but also in technologies and their applications in the society / community or respective fields.

REFERENCES

- Abeysekera, I., & Guthrie, J. (2004). How is intellectual capital being reported in a developing nation?. *Accounting in Emerging Economies*, (S2), 149-169.
- Agarwal, S. (1989). *Development of Documentation in India: Social Science Information*. New Delhi, India: Concept Pub. Co.
- Bansal. (2005). Securing the future of Information: Digitization and Preservation of Documents in e-Format. *DESIDOC Bulletin of Information Technology*, 25(1), 19-26.
- Buckland, M., & Liu, Z. (1995). History of information science. *Annual Review of Information Science & Technology*, 30, 385-416.
- Buckland & Liu. (1995). History of Information Science. *Annual Review of Information Science and Technology*, 30, 385-416.
- Chandrakant, N. (2011). Maximizing Lifetime of Wireless Sensor Network by using energy efficient middleware service. *International Journal of Innovative Technology and Creative Engineering*, 1, 20-24.
- Chandrakant, N. (2013). Green Computing and Mobile Cloud Computing inspired Middleware for Next Generation. *International Journal of Advanced Research in Computer Science and Electronics Engineering*, 2(7), 542-545.
- Cohen, E. B., & Malgorzata, N. (2006). Learning Objects and E-Learning: An Informing Science Perspective. *Interdisciplinary Journal of Knowledge and Learning Objects*, 2(2), 18-22.
- Marco, G. (1996). Two false dogmas of Information Science. *New Library World*, 97(11), 11-14. doi:10.1108/03074809610148766

Martin, S. B. (1998). Information technology, employment, and the information sector: Trends in information employment 1970–1995. *Journal of the American Society for Information Science*, 49(12), 1053–1069. doi:10.1002/(SICI)1097-4571(1998)49:12<1053::AID-ASI2>3.0.CO;2-U

Paul, P. K. (2008). *Information Science: Trends and Analysis 2003-07* (Dissertation). University of North Bengal, India.

Paul, P. K., Chatterjee, D., Sridevi, K. V., & Dangwal, K. L. (2012). Computer and Information Science (CIS) Education in the Universities of India: Emphasizing Central Universities – A Brief Study. *International Journal of Marketing and Trade Policy*, 01(04), 277–284.

Paul, P. K., & Kumar, K. (2012). Green Computing Vis-a-Vis Information Science - Indian Perspective'. *International Journal of Computer Science and Engineering Systems*, 06(04), 167–171.

Paul, P. K., & Sridevi, K. V. (2012). Information Science (IS) Education: Challenges, Issues and Opportunities in Indian context. *International Journals of Mathematics and Engineering with Computers*, 03(02), 87–93.

Paul, P. K., & Dangwal, K. L. (2012). Green Computing: Opportunities and Problems in the perspective of Developing Countries. *Asian Journal of Applied Science and Technology*, 1(4), 15-19.

Singh & Kapila. (2004). Search Engines: Tools For Library. *Annals of Library And Information Studies*, 3(S1), 93-98.

Wang, C., Wang, Q., Ren, K., & Lou, W. (2010). Privacy-Preserving Public Auditing for Data Storage Security in Cloud Computing. *Proceedings - IEEE INFOCOM*, 1–9.

Wersig, G., & Neveling, U. (1975). The phenomena of interest to information science. *Information Scientist*, 9, 127–140.

White, H. D., & McCain, K. W. (1997). Visualization of literatures. *Annual Review of Information Science & Technology*, 32, 99–168.

ADDITIONAL READING

Boyd Rayward, W. (1997). The Origin of Information Science and the International Institute of Bibliography/ International Federation for Information and Documentation. *Journal of the American Society for Information Science*, 48(4), 289–300. doi:10.1002/(SICI)1097-4571(199704)48:4<289::AID-ASI2>3.0.CO;2-S

KEY TERMS AND DEFINITIONS

AICTE: All India Council for Technical Education responsible for the Engineering, Technical and Management Education in India.

Computing: Includes designing and building hardware and software systems for a wide range of purposes; processing, structuring, and managing various kinds of information; doing scientific studies using computers; making computer systems behave intelligently; creating and using communications and entertainment media; finding and gathering information relevant to any particular purpose, and so on.

Information: Information as a concept has a diversity of meanings, from everyday usage to technical settings. Generally speaking, the concept of information is closely related to notions of constraint, communication, control, data, form, instruction, knowledge, meaning, mental stimulus, pattern, perception, and representation.

Information School: Academic units which work for professional and academic solution of Information and Technologies. Information Schools are responsible for providing manpower

in the field of Information and Technologies and Computing. I-Schools gained popularity during 1990's when internet reached the common people and other commercial ventures. It is in this time the explosion of information has led the development of healthy information management approach. Earlier evolution of I-Schools provided such knowledge as Library Science, Communication Science, Mass-Communication, Computer Science and Information Technology.

Information Science: A discipline of disciplines that came from the subject which deals with information, information services, information activities and digital information. So mainly from

library science, archive science, documentation science, computer science it has originated; however this is not truly computer science nor library science. This is an interdisciplinary science; which is a combination of Library Science, Computer Science, Information Technology, Management Science, Operation Research and the subjects that deal with information.

Interdisciplinary: Connection between some subjects irrespective of domain and direct and indirect connection with main facet or sub field.

UGC: University Grants Commission is the main regulator of Universities of India, Including Central, State, Private and Deemed.



Logic Programming for Intelligent Systems

James D. Jones

Liberty University, USA

INTRODUCTION

In what seem to be never-ending quests for automation, integration, seamlessness, new genres of applications, and “smart systems”, all of which are fueled in part by technological changes, intellectual maturity (or so one thinks), and out-of-the-box thinking that says “surely, there must be a better way”, one dreams of a future. That future consists of many different implications and dimensions, but materially, part of that future consists of devices that compute. These devices are not limited to small gadgets, but include all machinery that makes any kind of decision, whether that decision is autonomous, or imposed by an external act or agent. We dream of a future of endless computation, convenience, and sophistication, orders of magnitude above where we are today.

Limiting the present discussion to non-trivial complex computations, this paper suggests a different paradigm for “intelligent systems” than what is commonly perceived as being sufficient. That paradigm is *Logic Programming*, which is the use of logic to represent and reason about knowledge. Further, this paradigm has been blessed with reasoning formalisms specified in the language of Logic Programming which endow us with the ability to do complicated, human-like reasoning.

These reasoning formalisms that provide incredible reasoning power, include: a well developed theory of multiple forms of negation, an understanding of open domains and the closed world assumption, default reasoning with exceptions, reasoning with respect to time (i.e., a solution to the frame problem, and introspection with regard to previous beliefs), reasoning about actions, introspection about current beliefs (as well as past beliefs), and maintaining multiple

views of the world simultaneously (which enables reasoning with uncertainty.)

This paper suggests that logic programs employing recent advances in semantics and in knowledge representation formalisms provide a more robust framework in which to develop very intelligent systems in any domain of knowledge or application. The author has performed work applying this paradigm and these reasoning formalisms in the areas of financial applications, security applications, and enterprise information systems.

BACKGROUND

This is a primer for computing professionals, and not an authoritative work on the state-of-the-art of Logic Programming. Logic programming presents us with an excellent tool to develop a variety of powerful intelligent systems. What will be discussed is a particular version of logic programming called “ASP”, which means “Answer Set Programming”. “Version” in this context is defined by the semantics of the language, and not by performance issues. ASP is currently the most popular and the most well researched Logic Programming language. (To demonstrate the importance that ASP is receiving in the field, an entire issue of “AI Magazine” has been devoted to ASP. This journal is the flagship journal for the Association for the Advancement of Artificial Intelligence, the premier scholarly and practitioner association in the field of Artificial Intelligence. The issue mentioned is the most current issue at the time of this writing, Volume 37 Number 3, Fall 2016.)

While there are still issues that need to be addressed in Logic Programming, and while

DOI: 10.4018/978-1-5225-2255-3.ch411

there may be additional non-logical techniques to complement logic-based systems, it is reasonable to believe that logic will form the cornerstone of any serious machine intelligence in the future. Consider that the goal of the field of artificial intelligence (in fact, one of the early goals of computing itself) is to build “HAL”, the all-knowing, self-sufficient computer of science-fiction movies (Clarke, 1968). To this end, it behooves us to understand, use, and further refine this paradigm.

In this paper, the syntax and of semantics of ASP shall be presented. The early forms of this language were initially called the Stable Model Semantics, and later, Disjunctive Logic Programming (Gelfond & Lifschitz, 1988; Gelfond & Lifschitz 1991). This language is a purely declarative language (as opposed to procedural languages which dominate industry, and as opposed to functional languages which are the main competitor in the field of AI.) Declarative languages have their roots in logic programming (Kowalski, 1974; Kowalski, 1979), they are defined in the syntax and semantics of standard Prolog (Colmerauer, et. al., 1973; Clark 1978), and they employ the work on nonmonotonic logic (Reiter, 1980; Moore, 1985).

The semantics of ASP are arguably the most well-known and most well-developed semantics in logic programming. That there are other competing semantics is not of concern. Other semantics will differ from the ASP primarily at the extremes. Also, the semantics of ASP are conservative: a system built upon these semantics believes only what it is forced to believe. As already stated, the semantics of ASP are the most popular semantics in the logic programming research community.

It is the hope here that ASP and the accompanying knowledge representation techniques presented will be appreciated by the practitioner community to the extent that ultimately, reasoning systems will be based upon these ideas. The material presented here is self-contained.

ANSWER SET PROGRAMS (ASP)

The language of ASP is mathematically precise, and well understood among researchers and

those “in the know”. A correct discussion of this would necessarily consists of several unambiguous, mathematical definitions. However, for the purposes of this work, definitions will be kept to a minimum, and will favor an intuitive bent, rather than a rigorous mathematical bent. Nonetheless, some formalities are necessary. An intuitive discussion may be viewed as “sloppy” in the sense of mathematical rigor, but it is necessary and easier to understand for those not acquainted with the field. It is this audience to which this paper is being written.

The presentation here will be “backwards” in the sense of starting with the final result, and explaining the component pieces. A mathematical presentation would do the opposite: build a foundation, such that when the final result is presented, all the pieces have already been discussed.

Syntax

The syntax of any language is the grammar that defines how to construct legal sentences in that language. The syntax of ASP is as follows. A rule (the only building block of an ASP program) is of the form:

$$l_0 \text{ or } \dots \text{ or } l_i \leftarrow l_{i+1}, \dots, l_m, \text{ not } l_{m+1}, \dots, \text{ not } l_n \quad (\text{Gelfond \& Kahl, 2014})$$

where the l s are “literals”, $i \geq 0$, $m \geq 0$, and $n \geq 0$.

The symbol l is just an abstraction for a “predicate” with the “terms” appropriate for that predicate. (All these predicates and terms are defined merely by their use in an ASP program.) These would be some examples of literals:

```
male(john)
married(john, sue)
employee(john, accounting, 1/25/2012)
itIsHotToday()
```

The terms of the last two literals need further technical explanation, but suffice it to say that even these literals *could* be correct.

The left side of the ‘ $\leftarrow$ ’ symbol is called the “head” of the rule, and the right side of that symbol is called the “body” of the rule. This is a rule of inference: if the right side of the rule is true, then the left side of the rule must be true. (That is, the head can be inferred to be true.) There can be zero or more literals on the left side of the ‘ $\leftarrow$ ’ symbol; there can be zero or more literals in the part of the body prior to the sequence of “not” symbols; and there can be zero or more literals each individually preceded by the symbol “not”.

The above examples of literals only tell part of the story. Those literals are “positive”. In a proper context, they could be making the claim that something is true. (For instance, that it is true that “john” is a “male”.) There are also “negative” literals. For instance

$\neg$ married(john, sue)

in the proper context, could be stating that it is true that *john* and *sue* are not married. (Perhaps this should be stated more strongly: it is absolutely false that *john* and *sue* are married.) The symbol $\neg$ is called “strong negation”. A literal can be positive or negative.

The symbol “not” means that something is “unknown”. Mathematically, it means that something is “unprovable”. According to the rules of the specific ASP program, it can not be proven to be true. This symbol is called “weak negation”.

Semantics

The syntax of a language describes how to construct legal sentences in that language. The semantics of that language describe what those sentences mean.

As a prelude, the notion of *satisfiability* needs to be discussed. The ultimate goal is to arrive at an *answer set* for a program. All interactions with a program (queries about whether or not something is true) will be made with respect to this answer set. This answer set is a minimal set of literals that *satisfies* the rules of the ASP program. The following is the definition of satisfiability.

“A set S of ground literals **satisfies**:

1. l **if** $l \in S$;
2. $\text{not } l$ **if** $l \notin S$;
3. $l_0 \text{ or } \dots \text{ or } l_n$ **if** for some $1 \leq i \leq n$, $l_i \in S$
4. a set of ground extended literals **if** S satisfies every element of this set;
5. rule r **if**, whenever S satisfies r ’s body, it satisfies r ’s head.” (Gelfond & Kahl, 2014, p. 15)

A “ground literal” is a literal in which there are no variables. For instance,

male(john)

is a ground literal, whereas

male(X)

is not.

A “ground extended literal” is a ground literal that optionally may or may not be preceded by the symbol “not”, such as

not visiting(john, sue).

If set S from above satisfies all the rules of the ASP program, and if set S is minimal (with respect to set inclusion, meaning that there is not proper subset of S which also satisfies all the rules of the ASP program), then S is an answer set for the program. It may be possible to have more than one answer set for a program. (Multiple answer sets can be induced either by: 1) weak negation (the “not” symbol), or 2) a head in which disjunction (the “or” symbol) appears. There are ways to ensure that there exists only one answer set, but that discussion is beyond this paper.

In its simplest, abstract sense, the goal of any program is to answer queries. To “run the payroll”, what is being asked is “What is the net pay of each employee for this pay period”. Different questions might include things such as “what is John’s wife’s name”, “should the A/C be turned on

to maintain a temperature of 78 degrees”, “should I shoot down the incoming missiles”.

Let us consider a query *Q* which is a literal. In determining how an ASP program responds to that query, the answer is:

1. *Yes*, if *Q* is an element of ALL the answer sets of the program
2. *No* if the complement of *Q* is an element of ALL the answer sets of the program
3. *Unknown* otherwise

This discussion of semantics has omitted some important aspects, namely:

1. The semantics of a program that includes disjunctive heads (the notion of an answer set being minimal necessitates that when required, and whenever possible, only one disjunct of that head is an element of the answer set),
2. The meaning of query *Q*, when *Q* is a conjunctive query (each element of the conjunction must be true with respect to All the answer sets for the query to be true, or at least one element is false with respect to ALL the answer sets for the query to be false, unknown otherwise), and
3. The meaning of query *Q*, when *Q* is a disjunctive query (only one element of the query needs to be true in ALL the answer sets for the query to be true, or each element of the disjunction must be false in ALL the answer sets for the query to be false, unknown otherwise.)

Some Reasoning Techniques

Defaults with Exceptions

Default reasoning is a key area of human mental activity. Simply put, default reasoning is “jumping to a conclusion”. Throughout any typical

day, a human will jump to many conclusions. It is this area of reasoning that differentiates between human reasoning and mathematical logic. Mathematical logic is “sound and complete”. “Soundness” means that a conclusion is correct under all circumstances. New information will not change the truth of such a conclusion. “Hunches”, on the other hand (which humans are so good at) are not sound conclusions. The conclusions may ultimately be right, or they may turn out to be wrong. Yet, without “hunches” or “jumping to conclusions”, people could not operate. Humans do not live in this world with complete knowledge, therefore, much of the time we are guessing.

This is an area of reasoning that has proven difficult for computers. Work in this area originated with (Reiter, 1980). In the context of logic programming, it is more clearly understood with the work of (Gelfond, 1994).

As an example, assume *sam* is a bird. Let us decide if *sam* flies. This is formalized as

```
flies(X) ←
bird(X), (3)
not ab(flies, bird, X), (4)
not ¬flies(X). (5)
```

Line (5) is necessary to prevent our program from concluding an inconsistency. It is not desired for the program to conclude that a specific bird flies if it can elsewhere conclude that that same bird does not fly.

Line (4) is a crucial part of the formalism. *Ab* is called the *abnormality predicate*. It is a 3-ary predicate. Its use is to state that something is abnormal. The first term is the predicate which we are trying to conclude, in this case, *flies*. The second term is the predicate which states that “normally, *X*’s are *Y*’s”. In this case, that predicate is *bird*. The statement is that normally, birds fly.

Line (3) lists all the “normal” criteria for concluding the head of the rule. The entire rule reads: “if *X* is a bird, and if it is not known that



X is abnormal with respect to flying, and if it cannot be concluded that X does not fly, then it is concluded that X flies.

In this example, given just the rule above, and the fact

bird(sam)

then it can be concluded that *sam* flies.

Let us expand the example by adding the following information:

penguin(sam)
ab(flies, penguins, X)

These rules state that *sam* is a penguin, and that penguins are abnormal with respect to flying. Now, our program will not conclude that *sam* flies.

In order to allow our program to jump to safe conclusions, we create such defaults rules. We then must also list the conditions under which the defaults are abnormal. This is a very modular approach.

(Zhang 2003) presents the notion of having priorities among defaults. Not only does this allow us to express a hierarchy of defaults, but can also ensure a unique answer set, which greatly simplifies computation.

Selectively Implementing CWA

CWA (the “closed world assumption”) is appropriate in applications for which it is appropriate to assume that the reasoner has complete knowledge. Such an application would be an airline schedule. If we ask the reasoner if there is a 3:00 flight from Dallas to Chicago on Friday, the answer is easily determined. If there exists a record in the flight table for such a flight, then the answer is yes (equivalently, true.) Otherwise, it is false that such a flight exists.

CWA assumes that you have complete knowledge of the domain you are reasoning about. For some applications (such as an airline schedule), this is appropriate. For other applications, this is

disastrous. What we need, and what we illustrate here, is the ability to selectively incorporate CWA. (That is, to use it when it is needed, and to not use it when it is not needed.) Consider a human resources database that defines the *employee* relation with rules of the form:

employee(john, accounting, 40000)

which states that *john* is an employee who works in the *accounting* department, earning a salary of \$40,000. Suppose we ask the query

employee(mary, X, Y)?

which asks whether or not *mary* is an employee. Suppose that this person does not work for our company, and therefore, such a fact is not in the database. Since neither the query nor its negation can be proven, the result would be *unknown*. However, this does not match normal intuition. In this scenario, it is clearly desired that the system respond to such a query with “no”.

The desired result can be achieved by applying CWA to the *employee* relation. This can be done with the following rule:

$\neg employee(X, Y, Z) \leftarrow$
not employee(X, Y, Z)

This rule means that if it is not believed that X is an employee, then it is definitively false (by way of default reasoning) that X is an employee. For every predicate (or relation) for which CWA is desired, such a rule very similar to the above needs to be added.

Reasoning About Actions and Time

A very significant area of knowledge representation (especially as it relates to logic programming) is reasoning about actions and time. (Baral, et. al., 1997a; Baral & Gelfond, 1997b; Gelfond & Lifschitz, 1993) formalize these ideas in extended logic programs.

At first glance, it may appear that reasoning about actions is an irrelevant exercise. However, all computer processing can be formalized in this framework. Any entity can be abstracted as an agent. Agents perform actions. Any event can be abstracted as an action. Actions cause information to change. (Things may become true or false due to an action.) Within a theory of actions, information is represented as *fluents*. Any piece of information can be abstracted as a fluent. A theory of actions can be used to represent any area of human knowledge, or any area of human endeavor.

A fluent is defined as in the following.

fluent(Type,Item).

This formula has two parameters (terms). The first parameter identifies the type of fluent. There are two values for the type of fluent: *i* means the fluent is inertial, and *n* means the fluent is not inertial. An inertial fluent retains its value from one time period to the next, unless an action causes its value to change. A non-inertial fluent retains its value only for the present time period (or the present duration of time, whichever is appropriate.) The second parameter *m* identifies the fluent itself. Consider the following example. *fluent(i,m)*

In this example, suppose that *m* is a message which means “the books are out of balance”. It is an inertial fluent, meaning that if the books are out of balance, they will remain out of balance until something happens to change that fact. On the contrary, if the books are not out of balance (meaning they are balanced), then they will remain balanced until something happens to change that fact. This fluent (or proposition) can be true, false, or unknown.

Assume that the truth value of this fluent is “true”. Since the fluent is inertial, it will remain true that the books are out of balance until something causes it to change its truth value. An example of something that could cause it to change its truth value would be the action of making a journal entry to balance the books. Another example of

something which could change the truth value of this fluent is that the error causing the imbalance could be discovered and corrected.

What causes this fluent to have any meaning at all is an assertion as to whether the fluent is true or not. It is the collection of these assertions which form our knowledge base, or our set of beliefs. The general form for such an assertion is *holds*(*Fluent*, *Time*, *Truth_vale*).

Hence, *holds(m,1005,0)* means that fluent *m* is *false* at time 1005. (To relate this back to our example, “it is explicitly false that the books are out of balance at time 1005”.) Similarly, *holds(m, 1005, 1)* means that fluent *m* is *true* at time 1005. A fluent does not ever have to be asserted as either true or false; it could be “unknown”. There is not a separate designation for “unknown”. Rather, both formulae *holds(m,1005,0)* and *holds(m, 1005, 1)* would be missing from our set of beliefs.

A powerful feature of fluents is that unlike other variables in typical programming languages, a fluent can have very many values. The value of interest is determined by the context of time. Assume that we have the formula *holds(m, 1005, 1)*. Since this fluent is inertial, it will remain true the next time period, unless something happens to change that fact. Therefore, we will also have *holds(m, 1006, 1)*. We will continue to have similar formulae for each time period until the value of the fluent changes. Suppose that quite some time later at time period 2233, an action causes the books to become “balanced”. We will have the formula *holds(m, 2233, 0)*. Suppose that a very long time later at time period 102054321, the books became out of balance again. We will then have the formula *holds(m, 102054321, 1)*. A year later, we have all this knowledge available to us, and we can determine that at time period 1005 the books became out of balance, they remained out of balance until time period 2233, and they remained balanced from time period 2233 until time period 102054321.

To give a more complete picture of a theory of actions, there are underlying rules of significance. Some of those rules define the “law of inertia”.

What this means is that there are rules that infer that if an inertial fluent held in one time period, it will continue to hold the next time period unless something changes that fact. There are also rules that define the actions of the domain being reasoned about. In particular, those rules will define the impact of actions. For instance, if I turn on a light with paint on my finger, it will then be true that the light is on, it will be true that I can see what I am doing, it will be true that there is paint on the light switch, it will be true that the nocturnal insects are scurrying for cover, etc.

Other Knowledge Representation Issues

There are many other knowledge representation issues which have been identified, carefully analyzed, and formalized. (Baral, 2003; Brachman et.al., 2004; Davis, 1990), and to a lesser extent (Genesereth, et. al., 1987) address many of these. Some of the issues they identify are generic and apply to any domain of reasoning. Others are domain specific. The latter three books (i.e., all except (Baral, 2003)) represent the ideas in *first order logic* (FOL). Formulae in FOL would have to be translated into logic programs.

FUTURE RESEARCH DIRECTIONS

Further Development of Recent Research

There are several ideas found in recent research worthy of deeper consideration. These may assist in tailoring systems which are either more correct, or better suited to serve a particular reasoning constraint. (Delgrande, et. al., 2003; Eiter, et. al., 2003) discuss how to give preference among rules. This would be objectionable to a “purist” view, since a logician would take the posture that we are concerned about all that a program entails.

Nevertheless, this is very much akin to what happens in conflict resolution of rule-based systems. Further, as will be pointed out in a moment, this approach lends itself to creating a unique (i.e., single) answer set.

An inconsistent program is one that infers P and $\neg P$, for some atom P . We mentioned earlier how to cast default rules so that the default rules themselves will not introduce an inconsistency. There may be other opportunities for inconsistencies to be introduced. (Nieuwenborgh and Vermeir, 2006) find the “best” answer set for such programs. (Balduccini & Gelfond, 2003) take another approach. They have developed *CR-Prolog* which restores consistency in answer set programs (without sacrificing multiple answer sets, as the “best” answer set approach does.)

Answer set programming allows us to represent uncertainty with its ability to represent multiple views of the world, and with its semantics for negation as failure. To further our ability to represent and reason about uncertainty, (Baral, et. al., 2004) introduces the concept of combining answer set programming with probabilistic reasoning. This paradigm is called *P-log*. (Gelfond, et. al., 2006) combines the consistency restoring abilities of *CR-Prolog* with the probabilistic features of *P-log*.

Implementation Issues

An implementation issue is the software platform (akin to a “compiler”) to use. In short, we are faced with two choices: use a commercial Prolog product, or use a research tool. The advantages of using a commercial product are many. Some of these include: a wide user-base, an efficient product, an easy-to-use interface, product stability, etc. However, the most compelling disadvantage is that a Prolog product limits the representation and reasoning abilities. On the other hand, a research tool (such as *SModels*) implements the full features of extended logic programs. (In fact, it implements the full features of disjunctive

logic programs, which is higher in the hierarchy.) Each of the advantages of a commercial product is a disadvantage for a research tool.

The criteria for selecting a commercial product vs. a research tool is simple: is the ability to represent multiple ways of viewing the world more important than all the benefits of a commercial product? We live in such a monolithic world (in terms of information processing) that in many cases the answer will be “no”. Hence, the choice in these situations would be a commercial Prolog product. To make this choice does not mean to entirely negate the power of logic programming. The companion paper mentioned before discusses several knowledge representation techniques. If the product implements negation as failure, and if designers/users can accept a monolithic view of the world (which to this point, they have), then expert systems can still be advanced by implementing the sophisticated representation methodologies of the companion paper. On the other hand, for the truly adventurous, the use of the full power of the language of extended logic programs in an environment that truly supports the language (such as SMOBELS) will provide reasoning abilities far beyond current practice (albeit, at the price of a clumsy interface.)

CONCLUSION

We see logic programming to be a viable, more powerful, and more robust alternative to both traditional expert systems technology, and typical Prolog systems. Acceptance of this paradigm, and its associated reasoning paradigms, will advance any field of computing which desires to achieve some form of robust intelligence. Systems based upon these ideas will be capable of far more complex reasoning.

The ASP is a powerful, popular semantics in the field of logic programming. A challenge for the practitioner would be to decide how to implement this language. The practitioner is faced with

two basic choices: either implement a subset of the language in a commercially available Prolog product or implement the full language in one of the newer tools coming to market (which likely would be the immediate successor of a research tool that is not very user-friendly.)

REFERENCES

- Balduccini, M., & Gelfond, M. (2003). Logic Programs with Consistency-Restoring Rules. *AAAI Spring Symposium 2003*, 9-18.
- Baral, C., Gelfond, M., & Rushton, N. (2004). Probabilistic Reasoning with Answer Sets. In *Proceedings of Logic Programming and Non-Monotonic Reasoning*. LPNMR.
- Clark, K. (1978). Negation as Failure. In H. Gallaire & J. Minker (Eds.), *Logic and Databases* (pp. 293-322). New York: Plenum Press.
- Clarke, A. C. (1968). *2001: A Space Odyssey*. New York, NY: New American Library.
- Colmerauer, A., Kanoui, H., Paser, R., & Russet, P. (1973). *Un Systeme de Communication Homme-machine en Francais. Technical report*. Groupe de Intelligence Artificielle Universitat de Aix-Marseille.
- Delgrande, J. P., Schaub, T., & Tompits, H. (2003). A Framework for Compiling Preferences in Logic Programs. *Theory and Practice of Logic Programming*, 3(2), 129-187. doi:10.1017/S1471068402001539
- Eiter, T., Faber, W., Leone, N., & Pfeifer, G. (2003). Computing Preferred Answer Sets by Meta-Interpretation in Answer Set Programming. *Theory and Practice of Logic Programming*, 3(4), 463-498. doi:10.1017/S1471068403001753
- Eiter, T., Gottlob, G., & Mannila, H. (1997, September). Disjunctive Datalog. *ACM Transactions on Database Systems*, 22(3), 364-418. doi:10.1145/261124.261126

Gelfond, M. (1994). Logic Programming and Reasoning with Incomplete Information. *Annals of Mathematics and Artificial Intelligence*, 12(1-2), 89–116. doi:10.1007/BF01530762

Gelfond, M. (2002). Representing Knowledge in A-Prolog, Computational Logic: Logic Programming and Beyond. *Essays in Honour of Robert A. Kowalski*, 2408(Part II), 413–451.

Gelfond, M., & Kahl, L. (2014). Knowledge Representation, Reasoning, and the Design of Intelligent Agents. New York, NY: Cambridge. doi:10.1017/CBO9781139342124

Gelfond, M., & Lifschitz, V. (1988). The Stable Model Semantics for Logic Programming. *5th Intl Conference on Logic Programming 1988*.

Gelfond, M., & Lifschitz, V. (1990). Logic Programs with Classical Negation. *Logic Programming: Proceedings of the 7th Int'l Conf.*

Gelfond, M. & Lifschitz, V. (1991). Classical Negation in Logic Programs and Disjunctive Databases. *New Generation Computing*, (9).

Gelfond, M., Rushton, N., & Zhu, W. (2006). Combining Logical and Probabilistic Reasoning. *AAAI Spring Symposium 2006*, 50-55.

Jones, J. D. (2006, November). A Different Paradigm for Expert Systems: An Introduction to Logic Programming and Related Knowledge Representation Issues. *Expert Systems: International Journal of Knowledge Engineering and Neural Networks*, 23(5), 342–355. doi:10.1111/j.1468-0394.2006.00415.x

Kowalski, R. (1974). Predicate logic as a programming language. *Information Processing*, 74, 569–574.

Kowalski, R. (1979). *Logic for problem solving*. North-Holland.

Lifschitz, V. (1989). *Logical Foundations of Deductive Databases*. *Information Processing 89*. North-Holland.

Moore, R. (1985). Semantical Considerations on Nonmonotonic Logic. *Artificial Intelligence*, 25(1), 75–94. doi:10.1016/0004-3702(85)90042-6

Nieuwenborgh, D. V., & Vermeir, D. (2006). Preferred Answer Sets for Ordered Logic Programs. *Theory and Practice of Logic Programming*, 6(1-2), 107–167. doi:10.1017/S1471068404002315

Reiter, R. (1980). A logic for default reasoning. *Artificial Intelligence*, 13(1-2), 81-132.

Zhang, Y. (2003). Two Results for Prioritized Logic Programming. *Theory and Practice of Logic Programming*, 3(2), 223–242. doi:10.1017/S1471068402001552

ADDITIONAL READING

Brewka, G., & Eiter, T. (1998). Preferred Answer Sets for Extended Logic Programs. *Artificial Intelligence*, 109(1-2), 297–356. doi:10.1016/S0004-3702(99)00015-6

Erdem, E., Gelfond, M., & Leone, N. (2016). Applications of ASP. *AI Magazine*, 37(3), 53–68.

Janhunen, T., & Niemela, I. (2016). The Answer Set Paradigm. *AI Magazine*, 37(3), 13–24.

Lifschitz, V. (2016). Answer Sets and the Language of Answer Set Programming. *AI Magazine*, 37(3), 7–11.

KEY TERMS AND DEFINITIONS

Artificial Intelligence: The field of research aimed at getting a computer to perform tasks which one would normally associate with intelligence. Mathematician Alan Turing proposed what is now called “the Turing test” to define artificial intelligence. Loosely speaking place a computer in one secluded room, and a human in another secluded room. If an observer queried these two entities, and could not identify which entity was

the computer, then the computer had achieved intelligence on a par with a human.

Atom: Consists of an n-ary predicate symbol and n-terms. An example would be married(john,sue).

Common Sense Reasoning: A field of research in the artificial intelligence community that seeks to endow a computer with “common sense”. Many applications of artificial intelligence are highly successful. As an example, consider the fact that IBM’s “Big Blue” has defeated the world chess champion on multiple occasions. Yet, all computer applications perform miserably when required to perform outside of the realm for which they were designed. Put differently, the performance of humans degrades gracefully, while the performance of computers plummets rapidly when performing unanticipated tasks.

Extended Literal: A literal possibly preceded by the logical connective “not”. All of the following are extended literals - male(john), $\neg$ male(john), not male(john), not $\neg$ male(john).

Ground: There are no variables for a term; all terms are object constants. For instance, male(X) is not ground, but male(john) is ground. ‘X’ is a variable. ‘john’ is an object constant.

Knowledge Representation: A field of research within artificial intelligence that seeks to methodically identify ways in which to represent and reason about knowledge of a particular domain.

Literal: An atom or its negation. Examples would be married(john,sue), or $\neg$ married(john,sue).

Logic Programming: A field of research within artificial intelligence that seeks to use logic to represent and reason about knowledge. Initially, pure mathematical logic was used (aka “first order logic”). However, as the need to mimic human thought processes became prevalent (such as “rules of thumb”, or “jumping to conclusions”), the field broke away from first order logic. The field still has a very strong relationship to first order logic. However, the field employs nonmonotonic techniques, and higher orders of logic.

Strong Negation: Use of the ‘ $\neg$ ’ connective. This connective means “it is absolutely false that...”.

Weak Negation: Use of the “not” connective. This connective means “it is not provable that...”.



Methods for Improving Alias Rejections in Comb Filters

Gordana Jovanovic Dolecek

Institute INAOE Puebla, Mexico

INTRODUCTION

Decimation is the process of decreasing the sampling rate by an integer M in the digital domain. This process is used in sub band coding, filter banks, communication systems, oversampled A/D (analog/digital) converters, among others. If the signal is not appropriately filtered, the unwanted replicas of the main spectrum of the decimated signal, called aliasing, will be present. Therefore, to prevent aliasing in the decimated signal, the signal must be first filtered by a low pass filter, called anti-aliasing, or decimation filter. As a result, the process of decimation consists of two principal stages: filtering and down sampling (decreasing the sampling rate by integer M). The integer value M is also called the decimation factor.

In order to avoid high order decimation filters, the decimation is usually performed in two or more stages. The most simple decimation filter is comb filter, (Hogenauer, 1981) which usually works at high input rate. The transfer function of the comb filter is given as:

$$H(z) = \left[\frac{1}{M} \frac{1 - z^{-M}}{1 - z^{-1}} \right]^K, \quad (1)$$

where K is the order of the comb, and z is a complex variable.

The magnitude characteristic of the comb filter, expressed in digital frequency ω , is given as:

$$|H(e^{j\omega})| = \left| \frac{1}{M} \frac{\sin(\omega M / 2)}{\sin(\omega / 2)} \right|^K. \quad (2)$$

The magnitude characteristic must be flat in the pass band of interest. Additionally, in order to eliminate aliasing, the comb filter must have a high attenuation in the so called folding bands, which are the bands around the zeros of the comb filter.

BACKGROUND

A simple method to increase the attenuation in the comb folding bands consists of increasing the order of the comb filter, as shown in Figure 1. Here, the decimation factor M is 10 and the values of order K are 1, 3, and 5. The zooms in the pass bands in Figure 1 show that the pass band droop increases with the increase of the order of the comb filter.

The objective of this article is to present the methods proposed so far to increase the bandwidths around the comb zeros, and thus increase the comb alias rejection.

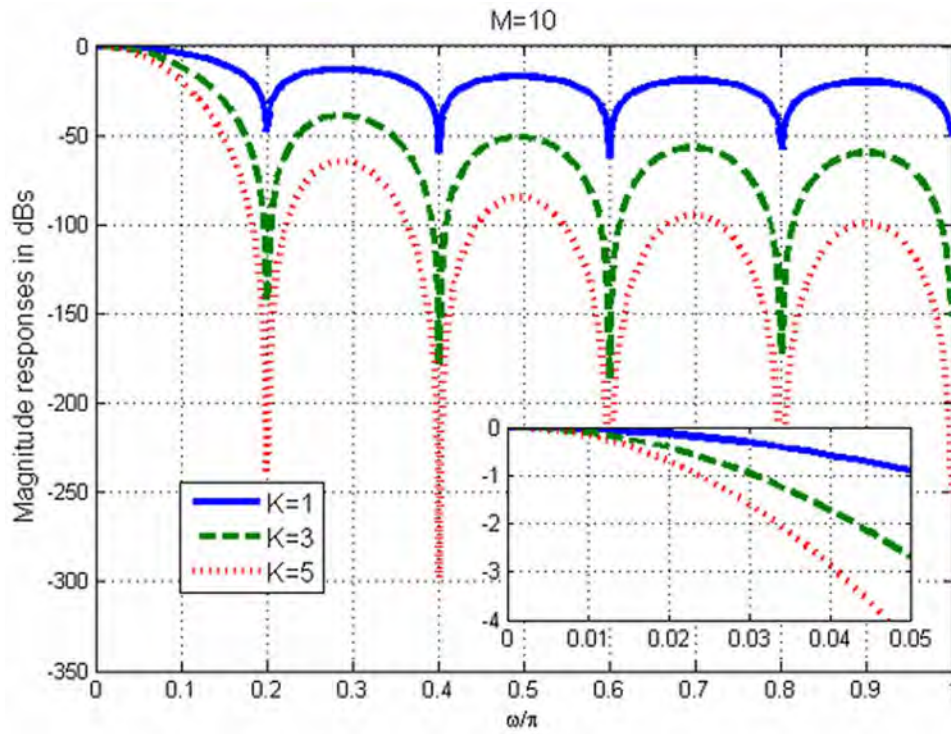
Methods for increasing comb alias rejections are primarily based on comb zero rotation. Additional methods based on cosine filters, Chebyshev polynomials, cyclotomic polynomials, and cascade of combs with different decimation factors, among others, are also used in practice.

REVIEW OF THE METHODS BASED ON COMB ZERO ROTATION

Zero Rotation

The comb zero rotation was introduced by (Presti, 2000). The method consists of applying a clock-

Figure 1. The overall comb magnitude characteristics and the pass band zooms for $M=10$ and different filter orders



wise rotation of β radians to any zero of comb filter, thus obtaining the following transfer function:

$$H_u(z) = \frac{1}{M} \frac{1 - z^{-M} e^{j\beta M}}{1 - z^{-1} e^{j\beta}}. \quad (3)$$

By applying the rotation in the reverse direction, the following transfer function is obtained:

$$H_v(z) = \frac{1}{M} \frac{1 - z^{-M} e^{-j\beta M}}{1 - z^{-1} e^{-j\beta}}. \quad (4)$$

The filters (3) and (4) both have complex coefficients. However, their cascade has real coefficients:

$$\begin{aligned} H_r(z) &= H_u(z) H_v(z) \\ &= \frac{1}{M} \frac{1 - z^{-M} e^{j\beta M}}{1 - z^{-1} e^{j\beta}} \frac{1}{M} \frac{1 - z^{-M} e^{-j\beta M}}{1 - z^{-1} e^{-j\beta}} \\ &= \frac{1}{M^2} \frac{1 - 2 \cos(\beta M) z^{-M} + z^{-2M}}{1 - 2 \cos(\beta) z^{-1} + z^{-2}} \end{aligned} \quad (5)$$

Figure 2(a) shows the zeros of a comb filter with $M=5$ and $K=1$. Similarly, Figure 2(b) shows the rotated zeros from (5), taking $\beta=0.03$. Note that the zeros of (5) are rotated around the original comb zero positions.

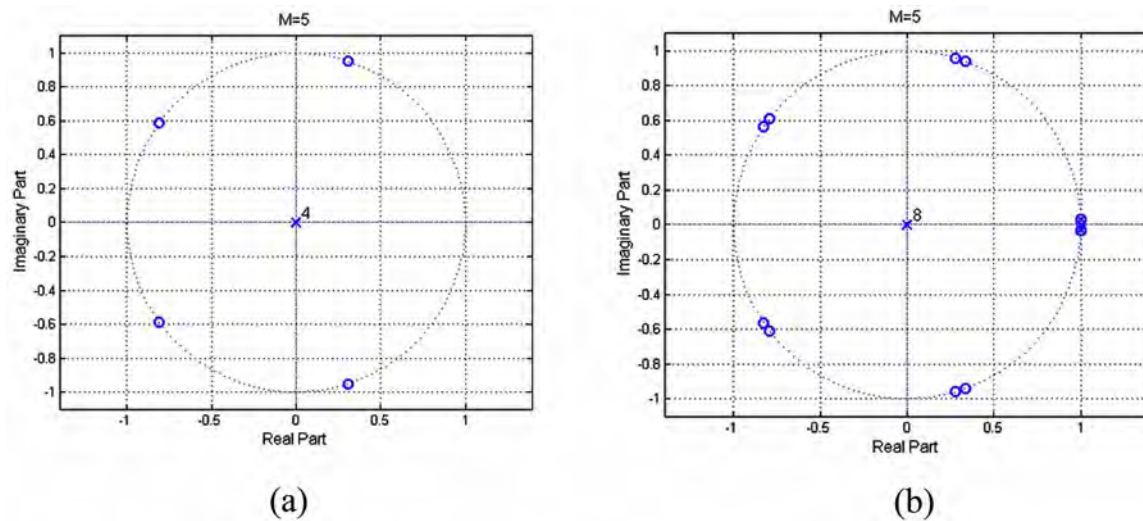
The cascade of the filter (5) and the comb filter is called rotated sinc (RS) filter:

$$\begin{aligned} H_R(z) &= H(z) H_r(z) \\ &= \frac{1}{M} \frac{1 - z^{-M}}{1 - z^{-1}} \frac{1}{M^2} \frac{1 - 2 \cos(\beta M) z^{-M} + z^{-2M}}{1 - 2 \cos(\beta) z^{-1} + z^{-2}} \end{aligned} \quad (6)$$

Figure 3 compares the magnitude responses of a comb filter with parameters $M=5$, $K=3$ and the corresponding RS filter. The first folding band zooms are also shown.

It can be observed that due to the rotated zeros, the RS filter provides wider folding bands and consequently has better aliasing rejection than the corresponding comb filter and without additional deterioration in the pass band region. The method

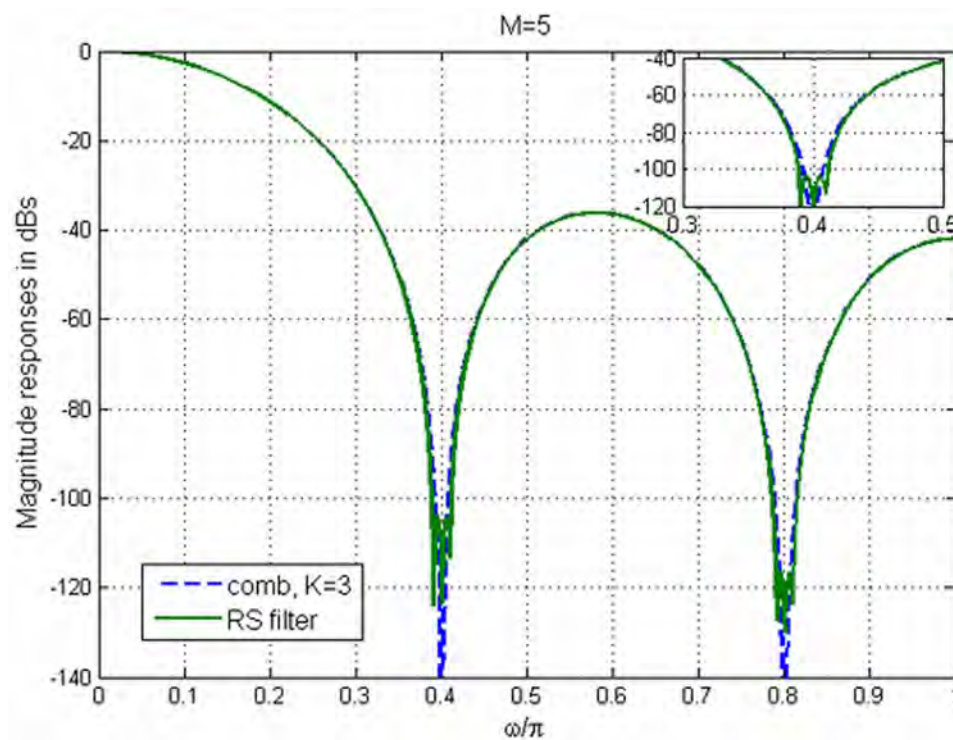
Figure 2. Pole-zero plots (a) Comb filter, $M=5$, $K=1$. (b) Filter (5), $M=5$, $\beta=0.03$



is generalized in (Laddomada, 2007) wherein a generalized comb filter (GCF) is introduced in which the parameter β is optimized to achieve

the optimal alias rejection for a given M and K . However, the comb zeros rotation introduces some problems, discussed in next section.

Figure 3. Magnitude responses of comb filter $M=5$, $K=3$ and the corresponding RS filter



Problems

The comb zero rotation improves the comb alias rejection but in turns introduces the following problems:

- Two multipliers are introduced, one working at high input rate.
- There is possible instability due to loss of pole-zero cancellation on the unit circle, in the case of the finite precision of coefficients.
- The most important first comb folding band has less attenuation than other folding bands and also has the worst case aliasing occurs.

As it mentioned before, the popularity of comb decimation filter is due to its simplicity, (comb filter is a multiplier less filter). Consequently, introducing multipliers for improvement of its characteristic degrades its principal features. As a result, it is desirable that the filter used to improve comb filter characteristic be also a multiplierless filter.

The solution to avoid possible instability is to eliminate the poles on the unit circle using a non recursive form for the rotation term.

Different methods have been proposed to solve above mentioned problems. In the following are presented some of the methods based on comb-zero rotation

1. Method in (Jovanovic Dolecek & Mitra, 2004)

In (Jovanovic Dolecek & Mitra, 2004), a modification of the original comb zero rotation method was proposed for the case in which the decimation factor M is a product of two integers M_1 and M_2 : $M = M_1 M_2$. To this end, the modified comb decimation filter has two stages with the different filter orders K_1 and K_2 :

$$H_m(z) = \left[\frac{1}{M_1} \frac{1 - z^{-M_1}}{1 - z^{-1}} \right]^{K_1} \left[\frac{1}{M_2} \frac{1 - z^{-M_2}}{1 - z^{-M_1}} \right]^{K_2} = H_1(z) H_2(z^{M_1}) \quad (7)$$

The filter $H_2(z^{M_1})$ can be moved to a lower rate which is M_1 times less than the high input rate, becoming:

$$H_2(z) = \left[\frac{1}{M_2} \frac{1 - z^{-M_2}}{1 - z^{-1}} \right]^{K_2} \quad (8)$$

Next, the zero rotation is applied only to the comb $H_2(z)$, resulting in the following modified rotated term:

$$H_{rm}(z) = \frac{1}{M^2} \frac{1 - 2 \cos(\beta M) z^{-M} + z^{-2M}}{1 - 2 \cos(\beta M_1) z^{-M_1} + z^{-2M_1}} \quad (9)$$

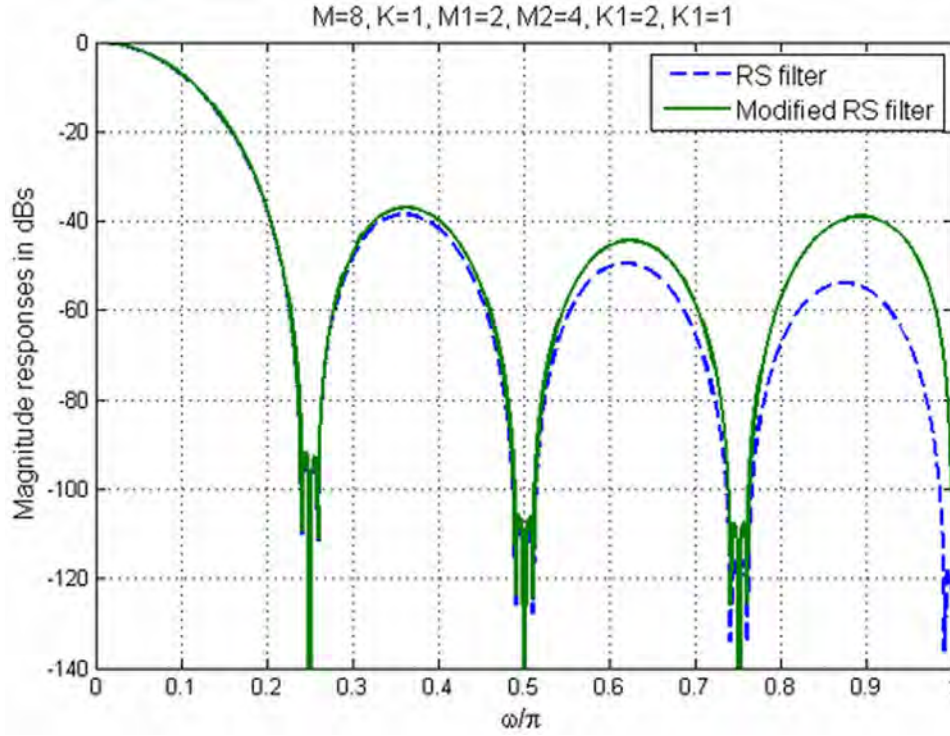
The transfer function of the modified RS filter is given as:

$$H_{Rm}(z) = H_m(z) H_{rm}(z) = H_1(z) H_2(z^{M_1}) \frac{1}{M^2} \frac{1 - 2 \cos(\beta M) z^{-M} + z^{-2M}}{1 - 2 \cos(\beta M_1) z^{-M_1} + z^{-2M_1}} \quad (10)$$

In contrast to the original rotated term (5), the rotated term in (9) works at rate which is M_1 times less than the high input rate. This price is that not all folding bands are improved, as shown in Figure 4, where the modified RS filter is compared with the original RS filter using $M=8$, $K=2$, $M_1=2$, $M_2=4$, $K_1=2$, $K_2=1$, and $\beta=0.03$.

The method is generalized in (Jovanovic Dolecek & Mitra, 2005) for the multistage comb decimator.

Figure 4. Comparison of magnitude responses of the RS and the modified RS filters



2. Method in (Jovanovic Dolecek, 2010)

In order to improve the alias rejection in the first folding band and to eliminate possible instability, the method proposed in (Jovanovic Dolecek, 2010) applies the zero rotation only in the first folding band, yielding:

$$\begin{aligned} H_{ZR}(z) &= k(1 - z^{-1}e^{-j\beta})(1 - z^{-1}e^{j\beta}) \\ &= k(1 - 2\cos(\beta)z^{-1} + z^{-2}), \end{aligned} \quad (11)$$

where k is the normalizing constant introduced to ensure that the magnitude characteristic is equal to 1 at $\omega = 0$.

The transfer function of the modified comb is given as

$$H_m = \left[\frac{1}{M} \frac{1 - z^{-M}}{1 - z^{-1}} \right]^K k \left(1 - 2z^{-1} \cos \left(\frac{2\pi}{M} - \frac{\pi}{(\beta_o + 2)M} \right) + z^{-2} \right), \quad (12)$$

where β_o is chosen to ensure that the rotated zeros are in the first folding band, $\beta_o = 0.99$.

To improve the alias rejection in subsequent folding bands, the expanded cosine filters are cascaded with the modified comb (12):

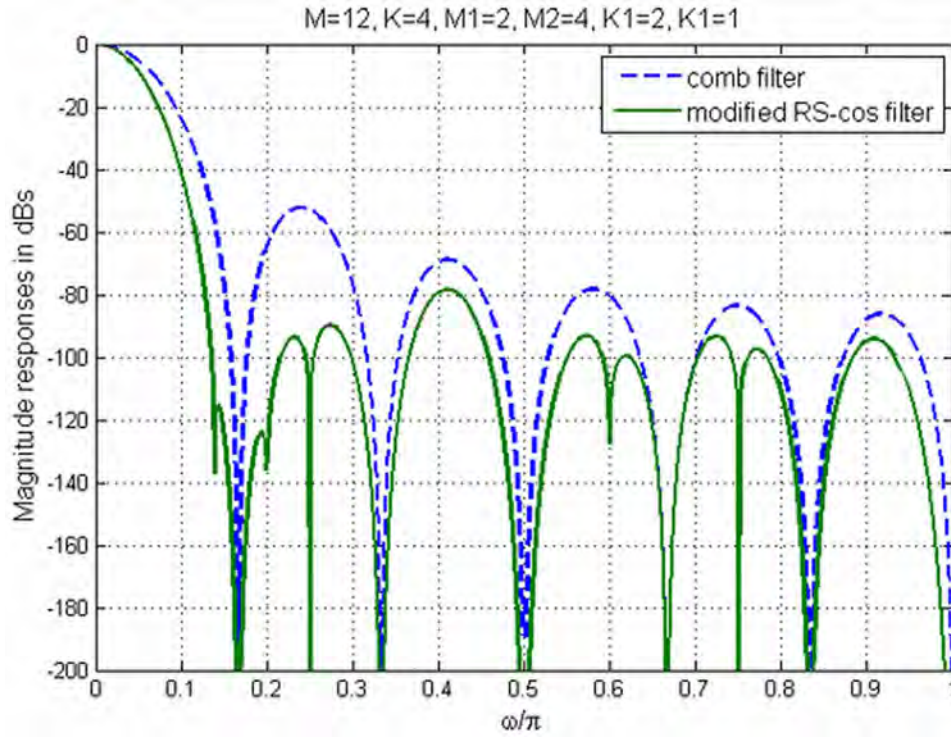
$$H_{\cos}(z) = \prod_{i=1}^{M/2} \left[\frac{1}{2} (1 + z^{-N_i}) \right]^{K_1}. \quad (13)$$

As a result, the following modified comb transfer function is obtained:

$$H_{m\cos}(z) = H_m(z)H_{\cos}(z). \quad (14)$$

Figure 5 compares the magnitude responses of the filter (14) and the comb filter for $M=12$, $K=4$, $N_i=1$, taking $i=1, \dots, 6$, and $K_1=1$.

Figure 5. Comparison of the magnitude responses of the RS and the modified cosine filter



3. Method in (Jovanovic Dolecek & Fernandez-Vazquez, 2010)

In (Jovanovic Dolecek & Fernandez-Vazquez, 2010) a method is presented wherein, the rotation term is expressed in a non recursive form, thus avoiding the possible instability due to the loss of pole-zero cancellation on the unit circle:

$$H_{r,nr}(z) = \frac{1}{M^2} \left[\sum_{i=0}^{M-2} c_i (z^{-i} + z^{-(2M-2-i)}) + c_{M-1} z^{-(M-1)} \right], \quad (15)$$

where

$$c_i = 2 \cos(\beta) c_{i-1} - c_{i-2}, \quad i = 0, \dots, M-2, \\ c_0 = 1; c_1 = 2 \cos(\beta). \quad (16)$$

Note that the filter (15) is symmetric and consequently has linear phase. The number of coefficients in (15) is equal to $M-1$.

As one example, using $M=5$, and $\beta=0.03$ in (16) the coefficients c_i are obtained as follows:

$$c_1=1.9991; c_2=2.9964; c_3=3.9910; c_4=4.9820. \quad (17)$$

The corresponding rotated term (15) is given as:

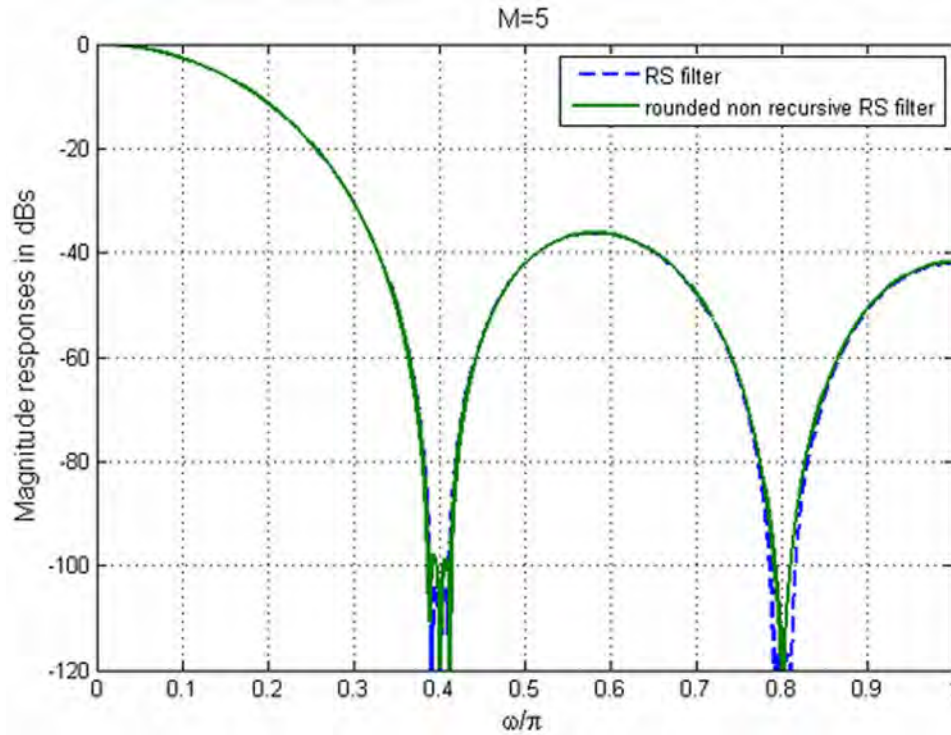
$$H_{r,nr}(z) = \frac{1}{25} \left[\begin{aligned} &(1 + z^{-8}) + 1.9991(z^{-1} + z^{-7}) \\ &+ 2.9964(z^{-2} + z^{-6}) \\ &+ 3.9910(z^{-3} + z^{-5}) + 4.9820z^{-4} \end{aligned} \right]. \quad (18)$$

It can be observed that this filter is equivalent to the recursive filter (5) for $M=5$ and $\beta=0.03$, and being in the non recursive form does not suffer from possible instability.

Using the rotation term (15), the transfer function of the corresponding RS filter is obtained as:

$$H_{RS} = H(z)H_{r,nr}(z). \quad (19)$$

Figure 6. Comparison of magnitude responses of the RS and the rounded non recursive RS filters



Note that this filter requires multipliers.

To eliminate the presence of multipliers, the method proposed in (Jovanovic Dolecek & Fernandez-Vazquez, 2010) rounds the coefficients c_i using the rounded constant $r=2^k$:

$$C_i = r \text{ round}(c_i/r), \quad (20)$$

where $\text{round}(x)$ means rounding of x to the nearest integer.

As a result of the rounding of the coefficients in (15), the following transfer function is obtained:

$$H_{r,nrr}(z) = A \left[\sum_{i=0}^{M-2} C_i (z^{-i} + z^{-(2M-2-i)}) + C_{M-1} z^{-(M-1)} \right], \quad (21)$$

where C_i can be implemented as adds and shifts, thus resulting in a multiplierless filter, and A is the normalization constant.

The rounded RS filter is given as:

$$H_{RSr} = H(z)H_{r,nrr}(z). \quad (22)$$

The rounding introduces a small distortion of the magnitude response. The level of the distortion is determined by the rounding constant.

As an example, by considering $r=2^6$ the following integer coefficients C_i in (21) are obtained from (17):

$$C_1=r128; C_2=r192, C_3=r255, C_4=r319. \quad (23)$$

Figure 6 compares the magnitude responses of RS and rounded non recursive RS filters, for $M=5$ and $\beta=0.03$, using $r=2^6$.

The principal problem in this method is a high number of coefficients which increase with the increase of the value of M . Similarly, in the rounded case, the number of adders is high and increases with the increase of the decimation factor M . As a result, the method is convenient only for small values of M .

4. Method in (Jovanovic Dolecek & Laddomada, 2015)

The problem of the increased number of the coefficients is solved by proposing a two-stage comb structure where the rounded rotation term is applied only at the second stage. Considering $M=M_1M_2$, the rotation term is given as:

$$H_{r, nr2}(z) = \frac{1}{M_2^2} \left[\sum_{i=0}^{M_2-2} c_i (z^{-i} + z^{-(2M_2-2-i)}) + c_{M_2-1} z^{-(M_2-1)} \right], \quad (24)$$

where the coefficients c_i are computed using (16) taking M_2 instead of M .

Note that as in (16), the coefficients c_i do not depend on the M but only on the rotation angle β .

The filter (24) has (M_2-1) coefficients which is less than in method (Jovanovic Dolecek & Fernandez-Vazquez, 2012), which requires $M-1$ coefficients.

The modified RS filter has the following transfer function:

$$H_{mRS}(z) = [H_1(z)]^{K_1} [H_2(z^{M_1})]^{K_2} H_{r, nr2}(z^{M_1}). \quad (25)$$

As an example, by considering $M=8$ and $M_1=2$, $M_2=4$, and $\beta=0.25$ we have

$$\begin{aligned} H_{r, nr2}(z) &= \frac{1}{16} \left[\sum_{i=0}^2 c_i (z^{-i} + z^{-(8-2-i)}) + c_3 z^{-3} \right] \\ &= \frac{1}{16} \left[(1 + z^{-6}) + c_1 (z^{-1} + z^{-5}) + c_2 (z^{-2} + z^{-4}) + c_3 z^{-3} \right], \end{aligned} \quad (26)$$

where

$$c_1 = 2 * \cos(0.25) = 1.9378.$$

$$c_2 = 2 * c_1 * \cos(0.25) - 1 = 2.7552. \quad (27)$$

$$c_3 = 2 * c_2 * \cos(0.25) - c_1 = 3.4012.$$

Note that here one needs only 3 coefficients, while in the method in (Jovanovic Dolecek, Fernandez-Vazquez, 2012) one needs 7 coefficients. Figure 7 compares magnitude responses of the filter (25) and the comb filter. (Parameters of the filter (25) and comb are: $M_1=2$, $M_2=4$, $K_1=2$, $K_2=1$, and $M=8$, and $K=2$, respectively).

Note that all folding bands are improved except for the last folding band. Similar to the method in (Jovanovic Dolecek & Fernandez-Vazquez, 2012), the coefficients c_i can be rounded, thus obtaining the multiplierless filter.

REVIEW OF METHODS BASED ON DIFFERENT CONCEPTS

Cosine Filters

The special case of comb filter for $M=2$, is a so-called cosine filter:

$$H(z) = \frac{1}{2} \frac{1 - z^2}{1 - z^{-1}} = \frac{1}{2} (1 + z^{-1}). \quad (28)$$

The name cosine is due to the cosine form of the magnitude response of the filter (28):

$$|H(e^{j\omega})| = \cos(\omega / 2). \quad (29)$$

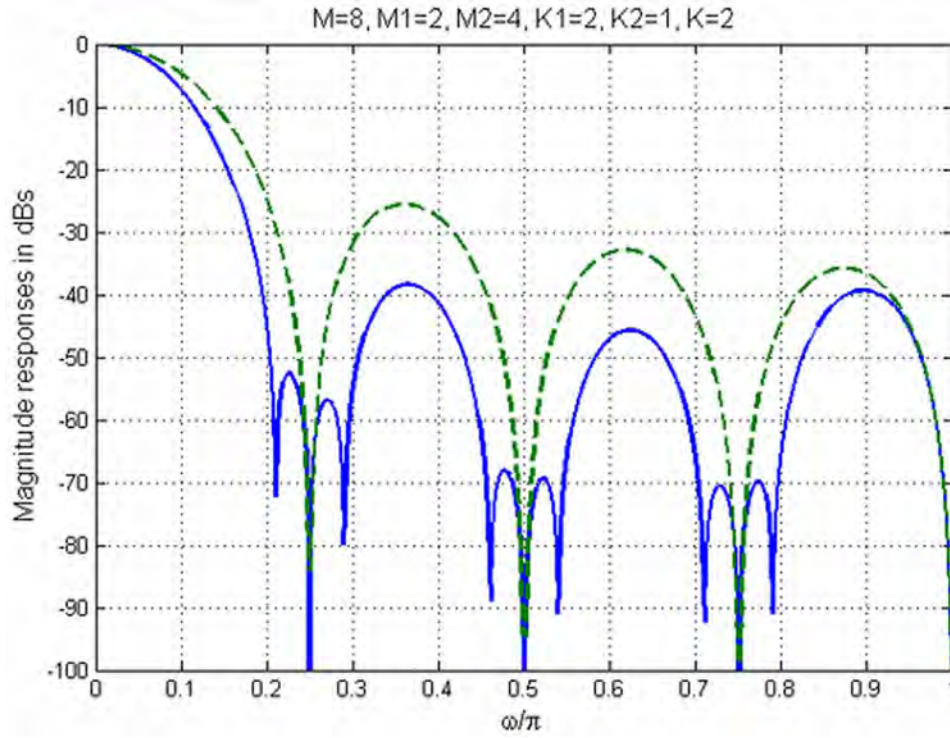
The expanded (by integer N) filter is obtained by replacing each delay z^{-1} by z^{-N} . From (28), the transfer function of the expanded cosine filter is given as:

$$H(z^N) = \frac{1}{2} (1 + z^{-N}). \quad (30)$$

From (29), the magnitude response of the expanded cosine filter is given as:



Figure 7. Comparison of magnitude responses of the comb and the two-stage non recursive RS filters



$$|H(e^{jN\omega})| = \cos(N\omega / 2). \quad (31)$$

By properly choosing the integer N , the zeros of the expanded cosine filter may fold down into the comb folding bands, and thus increase comb alias rejection.

1. Method in (Jovanovic Dolecek & Laddomada, 2013)

The cascade of two expanded cosine filters is proposed. The transfer function is given as:

$$H_{\cos}(z) = \frac{1+z^{-N_1}}{2} \frac{1+z^{-N_2}}{2}, \quad (32)$$

where N_1 and N_2 are integers.

The magnitude response of (32) is given as:

$$|H_{\cos}(e^{j\omega})| = |\cos(\omega N_1 / 2) \cos(\omega N_2 / 2)|. \quad (33)$$

The parameters N_1 and N_2 depend on the value of the decimation factor M . The parameters N_1 and N_2 are chosen to get zeros on the left and right sides of the first comb zero, respectively. There are two cases depending if M is even, or odd.

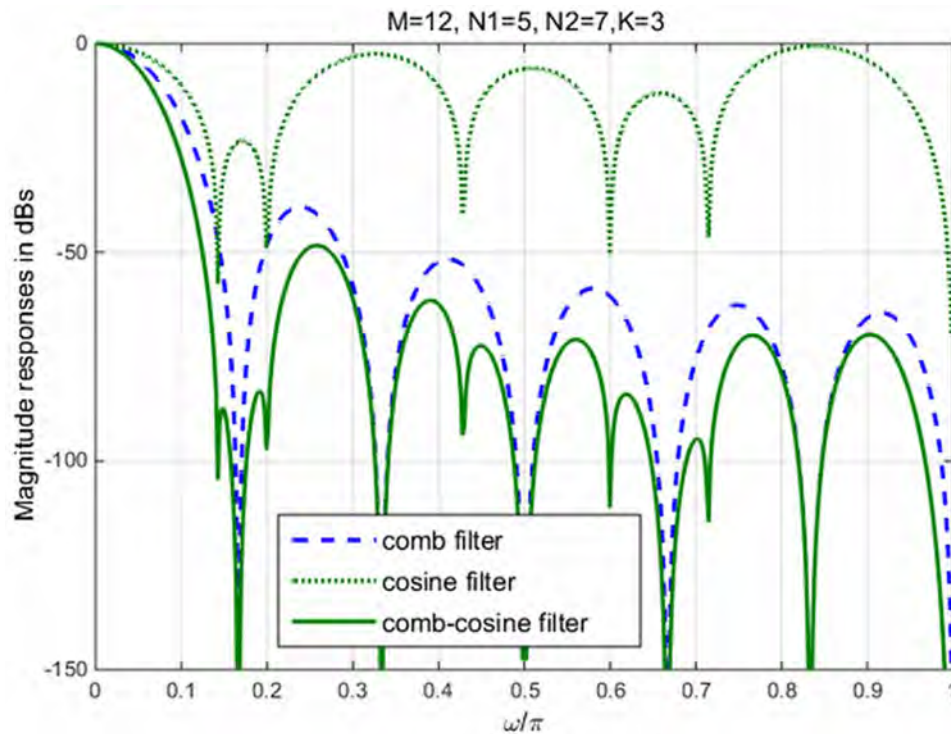
$$M \text{ even} : N_1 = \frac{M}{2} - 1; N_2 = \frac{M}{2} + 1. \quad (34)$$

$$M \text{ odd} : N_1 = \left\lfloor \frac{M}{2} \right\rfloor; N_2 = N_1 + 1. \quad (35)$$

The cosine-based comb filter is obtained by cascading the comb and cosine filters (32):

$$H_{\text{comb-cos}}(z) = A \left[\frac{1-z^{-M}}{1-z^{-1}} \right]^K (1+z^{-N_1})(1+z^{-N_2}). \quad (36)$$

Figure 8. Magnitude responses of the comb filter, cosine filter and comb-cosine filter



As an example consider comb filter with $M=12$ and $K=3$. According to (34), the corresponding values for N_1 and N_2 are 5 and 7, respectively. Figure 8 presents the magnitude responses of the comb filter, cosine filter (32) and comb-cosine filter (36). Note that, due to the cosine filter, the comb first folding band is wider and has an increased attenuation. However, not all folding bands are equally improved. Best improvements are obtained in the first and fourth folding bands.

2. Method in (Garcia Robles & Jovanovic Dolecek, 2016)

This paper describes novel, low complexity non-recursive comb-cosine decimation structures for decimation factors which can be presented as a power of two. The alias rejection is improved by inserting cosine filters into a certain stage of a comb non-recursive structure. The position of

cosine filters in the structure results in a trade-off between the improvement of alias rejection and area and power consumption.

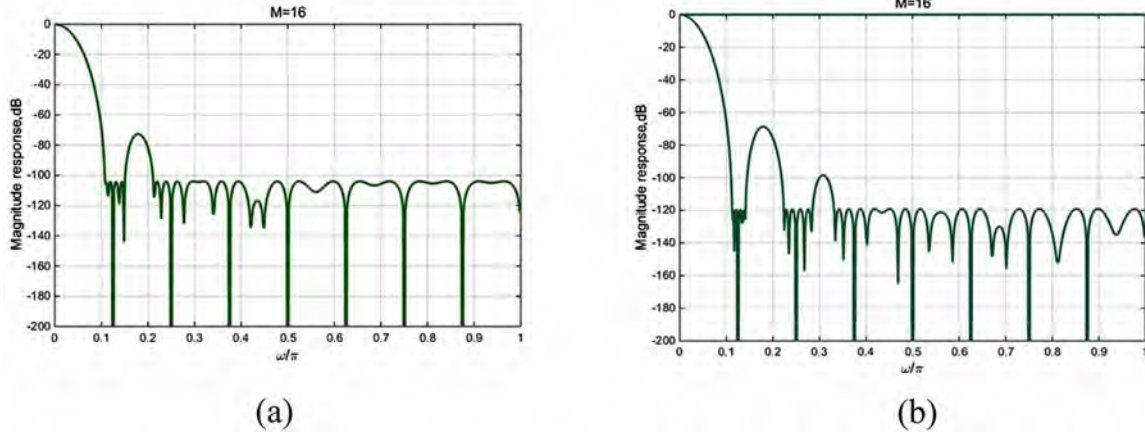
Chebyshev Polynomials

The stop bands of a comb decimation filter can be improved by sharpening with a Chebyshev polynomial, using a previously reported comb variant. As a result, comb multiple zeros are separated into an equiripple stop band.

1. Method in (Coleman 2012)

The method uses the sharpening comb with a Chebyshev polynomial. The advantage of the Coleman's method is that the same controlled attenuations and widths across all folding bands are obtained.

Figure 9. Magnitude responses of the Coleman's comb filters. (a) Minimum attenuations: 100dB. (b) Minimum attenuations: 120dB



As an example, consider $M=16$ and the 5th degree Chebyshev polynomial $T_5(x)=5x-20x^2+16x^5$. Figures 9(a) and 9(b) show the magnitude responses of the Coleman's comb filters for two values of the minimum attenuations: 100dB, and 120dB.

Cascade of Combs with Different Decimation Factors

The cascade of comb filters of different orders and with different decimation factors may result in higher attenuation than the equivalent comb filter.

1. Method in (Stosic, Milic & Pavlovic, 2015)

Considering the equivalent comb filter of order $7L+3$, where L is a given integer, and the decimation factor M , the following cascade is proposed:

$$H_{\text{cascad_comb}}(z) = [H_1(z)]^{L+1} [H_2(z)]^{L+1} [H_3(z)]^{L+1} [H_4(z)]^L [H_5(z)]^L [H_6(z)]^L [H_7(z)]^L, \quad (37)$$

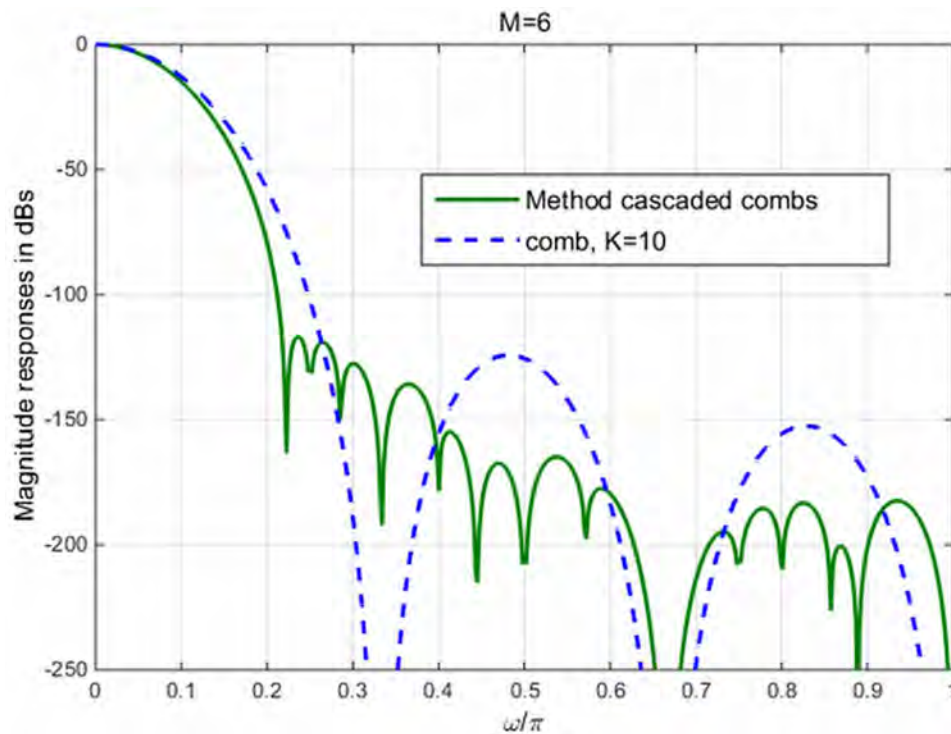
where

$$\begin{aligned} H_1(z) &= \frac{1}{M-3} \frac{1-z^{-(M-3)}}{1-z^{-1}}; \\ H_2(z) &= \frac{1}{M} \frac{1-z^{-M}}{1-z^{-1}}; \\ H_3(z) &= \frac{1}{M+3} \frac{1-z^{-(M+3)}}{1-z^{-1}}; \\ H_4(z) &= \frac{1}{M-2} \frac{1-z^{-(M-2)}}{1-z^{-1}}; \\ H_5(z) &= \frac{1}{M-1} \frac{1-z^{-(M-1)}}{1-z^{-1}}; \\ H_6(z) &= \frac{1}{M+1} \frac{1-z^{-(M+1)}}{1-z^{-1}}; \\ H_7(z) &= \frac{1}{M+2} \frac{1-z^{-(M+2)}}{1-z^{-1}}. \end{aligned} \quad (38)$$

As an example, for $M=6$ and $L=1$, the equivalent comb filter has an order of $K=10$. Figure 10 compares the magnitude response of the filter (37) and the equivalent comb filter. Note that this method is not dedicated to improving alias rejection in the comb folding bands but to improving the alias rejection in overall comb stop band, providing in this example a minimum attenuation of about 120 dB.

Similar proposals but with different comb filter decimation factors and orders are proposed

Figure 10. Magnitude responses of equivalent comb filter and method in (Stosic, Milic & Pavlovic, 2015)



in (Stosic & Pavlovic, 2014a), (Stosic & Pavlovic, 2014b), (Milic & Pavlovic, 2014), (Stosic & Pavlovic, 2016).

SOLUTIONS AND RECOMMENDATIONS

The review of the proposed methods for the improvement of comb alias rejection shows that there is a trade-off between the quality of improvement and the filter complexity. In comparison with the original comb zero rotation method, the advantages of the methods in which the rotation term is in a non-recursive form, thus avoiding the possible loss of pole-zero cancellations on the unit circle which leads to instability. Similarly, the methods based on cosine filters and Chebyshev polynomials have advantage of non recursive form. Additionally, the advantage of Chebyshev polynomial method is the

controlled and equal attenuations in all folding bands. The methods based on a cascade of different comb filters provide high attenuation in all stop bands at the cost of the increased complexity.

FUTURE RESEARCH DIRECTIONS

The future applications of decimation systems, especially in communications and in A/D converters impose high demands on the design of decimation filters with high alias rejections and low complexity. Therefore the future trends will be in developing new methods for improving comb alias rejections with the aim to increase the comb alias rejection as much as possible and to decrease the corresponding filter complexity as much as possible. The methods could be based either on comb zero rotation or on new mathematical developments and new structures.

CONCLUSION

In this paper we treated the problem of increasing the comb alias rejection, since the comb filter itself does not provide enough rejection of aliasing. The original comb-zero rotation method rotates the zeros of the comb filter around their original positions, thus introducing two new zeros in each folding band. The inserted zeros increase the widths of the folding bands and increase the attenuation across the folding bands. However, the disadvantage of the method is the introduction of two multipliers, one working at high input rate. Additionally, the rotated filter has poles on the unit circle which are perfectly canceled with the zeros, as far as the infinite precision of the filter coefficients is concerned. However, in the case of finite precision, this cancellation can be lost leading to instability. Different methods have been developed to address one or more issues associated with the original zero-rotation method with the rotation term either in recursive or non recursive forms. The methods with the rotation term in a non recursive form do not have the instability problem. The methods based on cosine filters, Chebyshev polynomials, and the cascaded comb filters have the advantage of non recursive form. Future work can be dedicated to a multiplierless design of filters for the improvement of comb alias rejection with a decreased number of adders, and improved alias rejection in all folding bands.

REFERENCES

- Coleman, J. O. (2012). Chebyshev stop bands for CIC decimation filters and CIC- implemented array tapers in 1D and 2D. *IEEE Transactions on Circuits and Systems. I, Regular Papers*, 59(12), 2956–2968. doi:10.1109/TCSI.2012.2206435
- Garcia Robles, A., & Jovanovic Dolecek, G. (2016). On non-recursive comb-cosine decimation structures. In *Proceedings of the IEEE Latin American Symposium on Circuits and Systems LASCAS 2016*, (pp. 1-4), Florianapolis, Brazil: IEEEExplore.
- Hogenauer, E. (1981). An economical class of digital filters for decimation and interpolation. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 29(2), 155–162. doi:10.1109/TASSP.1981.1163535
- Jovanovic Dolecek, G. (2010). Simplified rotated sinc (RS) filter for Sigma-Delta A/D conversion, In *Proceedings of International Conference on Green Circuits and Systems ICGCS 2010*, (pp. 283-288). Shanghai, China: IEEEExplore. doi:10.1109/ICGCS.2010.5543051
- Jovanovic Dolecek, G., & Fernandez-Vazquez, A. (2012). On non recursive rotated comb filter, In *Proceedings of Midwest Symposium MWCAS 2012*, (pp. 474-477). Boise, ID: IEEEExplore.
- Jovanovic Dolecek, G., & Laddomada, M. (2013). An improved class of multiplier less decimation filters: Analysis and design. *Digital Signal Processing*, 23(5), 1773–1782. doi:10.1016/j.dsp.2013.05.011
- Jovanovic Dolecek, G., & Laddomada, M. (2015). Design of two-stage non recursive rotated comb decimation filters with droop compensation and multiplier less architecture. *Journal of the Franklin Institute*, 352(3), 913–929. doi:10.1016/j.jfranklin.2014.11.016
- Jovanovic Dolecek, G., & Mitra, S. K. (2004). Efficient multistage comb-modified rotated sinc (RS) decimator. In *Proceedings of European Signal Processing Conference EUSIPCO-2004*, (pp. 1425-1428). Vienna, Austria: IEEEExplore.



Jovanovic Dolecek, G., & Mitra, S. K. (2005). A new multistage comb-modified rotated sinc (RS) decimator with sharpened magnitude response. *IEICE Transactions on Information and Systems: Special Issue on Recent Advances in Circuits and Systems*, E88-D(7), 1331–1339. doi:10.1093/ietisy/e88-d.7.1331

Laddomada, M. (2007). Generalized comb decimation filters for Sigma-Delta A/D converters: Analysis and design. *IEEE Transactions on Circuits and Systems. I, Regular Papers*, 54(5), 994–1005. doi:10.1109/TCSI.2007.895528

Milic, D. N., & Pavlovic, V. D. (2014). A new class of low complexity lowpass multiplier-less linear-phase special CIC FIR filters. *IEEE Signal Processing Letters*, 21(12), 1511–1515. doi:10.1109/LSP.2014.2343212

Presti, L. L. (2000). Efficient modified-sinc filters for sigma-delta A/D converters. *IEEE Transactions on Circuits and Systems-II: Analog and Digital Signal Processing*, 47(11), 1204–1213. doi:10.1109/82.885128

Stošić, B. P., Milić, D. N., & Pavlović, V. D. (2015). New CIC filter architecture: Design, parametric analysis and some comparisons. *Journal of the Institution of Electronics and Telecommunication Engineers*, 61(3), 244–250. doi:10.1080/03772063.2015.1021386

Stošić, B. P., & Pavlović, V. D. (2014a). Design of selective CIC filter functions. *AEÜ. International Journal of Electronics and Communications*, 68(12), 1231–1233. doi:10.1016/j.aeue.2014.07.002

Stošić, B. P., & Pavlović, V. D. (2014b). On design of a novel class of selective CIC FIR filter functions with improved response. *International Journal of Electronics and Communications*, 68(8), 720–729. doi:10.1016/j.aeue.2014.02.013

Stošić, B. P., & Pavlović, V. D. (2016). Design of new selective CIC filter functions with passband-droop compensation. *Electronics Letters*, 52(2), 115–117. doi:10.1049/el.2015.3333

ADDITIONAL READING

Harris, F. J. (2004). *Multirate signal processing for communications systems*. New Jersey, USA: Prentice Hall.

Jovanovic Dolecek, G. (Ed.). (2001). *Multirate systems: Design and applications*. Hershey, PA, USA: IGI Global.

Jovanovic Dolecek, G., & Diaz Carmona, J. (2005). A new cascaded modified CIC-cosine decimation filter. In *Proceedings IEEE Symposium on Circuits and Systems, ISCAS 2005*. (pp. 3733–3736). Kobe, Japan: IEEEExplore. doi:10.1109/ISCAS.2005.1465441

Lyons, R. G. (2010). *Understanding digital signal processing* (3rd ed.). New Jersey, USA: Prentice Hall.

Milic, L. (2009). *Multirate filtering for digital signal processing: MATLAB Applications*. Hershey, PA, USA: IGI Global. doi:10.4018/978-1-60566-178-0

Salgado, G. M., Jovanovic Dolecek, G., & de la Rosa, J. M. (in press). Low power two-stage comb decimation structures for high decimation factors. *Analog Integrated Circuits and Signal Processing*.

KEY TERMS AND DEFINITIONS

Aliasing: Unwanted replicas of the main spectrum of the decimated signal.

Comb Filter: Filter with all coefficients equal to unity.

Cosine Filter: Special case of comb filter for $M=2$. It has magnitude response in the cosine form.

Digital Filter: A filter that performs mathematical operations on a discrete-time signal.

Expanded Filter By N: The filter obtained by replacing each delay in the original filter by N delays.

Filter: The system which has function to reduce or enhance certain aspects of input signal,

as for example to remove unwanted parts of the signal, such as random noise, or to extract useful parts of the signal, such as the components with a certain frequencies.

Non-Recursive Form: The transfer function has only the nominator.

Recursive Form: The transfer function has the nominator and the denominator.

Rotated Comb: Comb filter in which all zeros are rotated around the original comb zero positions.

RS Filter: The cascade of rotated comb and comb filter.

A Paradoxical World and the Role of Technology in Thana–Capitalism



Maximiliano Emanuel Korstanje
University of Palermo, Argentina

INTRODUCTION

We live in digital times, where events are covered and disseminated by the media in seconds to a much wider audience (Warschauer & Matuchniak, 2010). Technology not only altered our current means of production, but also the ways life is lived as well as our habits, behaviours and social interaction with others (Gold 2012). The seminal texts of Jacques Ellul about a “technological society” shed light on the negative effects of autocracy imposed by the expansion of technological breakthroughs. The legitimacy of elite is enhanced at the time workforce accepts the cultural value of efficacy as the best of possible worlds. One day, humankind will be strictly controlled by Machines which will be characterized by rationality, artificiality and automatism. Ellul acknowledged that technology was conducive to the culture of capitalism posing instrumentality as mediator between citizens and institutions (Ellul, 1964). Scholars and thinkers are divided respecting to the role played by digital technology in the liberal world of consumers. Not too far from the legacy of Max Weber, Ellul was pioneer in adamantly alerting to what extent technology promotes an atmosphere of further alienation for human beings. However, others voices as Guy Sorman (2008) claims that the forces of progress activate conservative counter-reactions that are oriented to prevent a more egalitarian society. Detractors of technology and its progress only are limited to tell part of the truth, which means the aftermaths of new techniques in the fields of economy but ignoring those achievements promoted by technology as the expansion of life expectative

or the improvements in healthcare overt recent years. As Korstanje and Skoll put it, neither good nor bad technology depends on the use people did. Concerned by the paradoxes of Chernobyl, modernity showed that technology enrooted in a world of complexities and uncertainty would be as “a run-away train” very hard to control. The paradox was that the same instrument will make of our life a safer place to dwell become in a global threat that very well jeopardizes our existence in this planet (Korstanje & Skoll 2014). Here some questions arise: what is the role of Technology in our modern World, is technology a mechanism of control or censorship in democratic societies?, in what way?

Although technology introduced a plenty of liberties and rights for humankind, which are protected by democracy, it resulted in a much deeper disciplinary mechanism that leads to censorship. To put this in other terms, in Medieval Times, writers, thinkers whose text defied the authority of King or Catholic Church were jailed, tortured and condemned to the stake. The power of coercion emanating from Prince exerted violence as an efficient instrument of dissuasion, circumscribing the Leviathan’s whims. The dissemination of books was limited to those authors who were conducive to status quo. In this respect, the power was endorsed by the capacity of prince to create terror in others. Rather, in postmodern times, censorship is preferably achieved by over-production without limits and no matter whom or under what theme the writer focuses on. For example, once we key in Google the name Karl Marx or Max Weber we will get thousands of records of different studies containing or citing both scholars. Since

DOI: 10.4018/978-1-5225-2255-3.ch413

our limited mind can only be read part of these records (not all), we only are restricted to have a partial viewpoint of the problem. In the world of consumption, where liberty plays a crucial role in order for consumers to channel their desire in many directions, knowledge is over-produced to cause misunderstanding in readers. The larger the bibliography consulted, less the derived understanding. For those readers who are not specialized in sociology it is almost impossible to understand modernity only accessing to ten or twenty works bought in bookstore. This happens simply because the censorship in postmodern times is based on the liberty administered by technology to produce without order in many directions. Conducive to mass-consumption, freedom and democracy delineate the contours of societal order making the produced commodities affordable to consumers, but in so doing opens the doors for an atmosphere of conflict and discontent as never before.

Though this point will be explained in detail in the following sections, let's explain the term Thana Capitalism is used to denote a new stage of capitalist system. In this emergent facet not only technology played a vital role subordinating social practice to gazing but also paved the ways for the needs of captivating the suffering of others. In times of Thana-capitalism, risk sets the pace to death as a mediator between citizens and their institutions. We mean to Thana, as a derived term from Thanatos (Greek) from Death. In the days of Thana Capitalism, global audiences gaze spectacles which are based on news, content of disasters, mass-death or trauma. Citizens who are prone to gaze others' death enthrall their own status as a part of privileged-class.

BACKGROUND

Over decades, common questions asked on the evolution of science in Occident. Three different schools explored the interconnection of technology in the maturation of science. Econometrics, which initiate the first family of studies, signalled

to mathematical algorithms to produce top-down knowledge. Everything which cannot be visually tested was systematically rejected as a source of genuine knowledge. Human habits not number were the key factors that distinguished these emerging disciplines from economy. If classic science was prone with the premise of objectivity between *observer and observed*, sociology and anthropology gradually instilled a hermeneutics dialectics of knowledge which based on individual experience gave innovative outcomes (Cuevas 2005). Though this premise was already-evinced by young Marx, since the means of production mould the ideology for the society to understand environment, it is tempting to confirm that technologies are subsumed under an historical dialectic imposed externally by elite. This suggests that things are used to accomplish broader goals, but the underlying meaning for what we use these different objects is externally and culturally imposed by a privilege class (Marx, 1975; 1967; Marx & Engels, 1983). This raises a more than interesting question, is technology a platform to alienation?

For H. Marcuse, the alienatory nature of technology rested on its capacity to alter the cultural background of society, which legitimized the asymmetries of classes (Marcuse 1991). Capitalism has imposed an "economy of desire" which absorbs the wages of work-force (Korstanje 2015). In this respect, David Harvey (1989) argues that modern means of production blurred the different of space and time, accelerating the gift-exchange process. The oil embargo after 1972 produced a serious crisis to West which decentralized the economic system to the extent, that the concept of reality as it was envisaged by Enlightenment set the pace to multi-layered view of reality (Harvey 1989). In this context, French philosophy pivoted a radical criticism against the role played by technology and science in this every-increasing process. From Weber and Sombart's days, there was no other critical thinker as Jean Baudrillard that attacked the epistemological core of capitalism. For Baudrillard, capitalism has successfully expanded in view of the alternations of the

epistemological viewpoint in what reality means. The visual technology in these times has not only altered the perceptions of citizens accelerating the intersection of time and space, but the proliferation of textuality. In sharp contrast to other ages, modernity is based on the rise of computers where the meaning of publicity associates to the electronic World. To understand this better, he cites the example of Precogs, in *Minority Report* film, anticipate to crime before it is committed. This reminds that the visual appearance of events does not imply they are happening in present; as these crimes were forecasted by Precogs in a future that modified the present of potential criminal, now pseudo-events are virtually shaping our sensual realm (Baudrillard, 1986; 1995; 2001, 2002). As Virilio puts it, technology played a leading role not only blurring the borders between past and time, but effacing the difference between reality and pseudo-reality. In consonance with Baudrillard, Virilio understands that modern science remains subordinated to the interests of market, instead of searching in the true, to solve the problems of citizens. More interested in supporting the financial mega corporations, science embraces the paradigms of simulation, which reflects complex algorithms to understand the future. One of the chief aims of science, adjoined to technology, is to prevent the risks that may place capitalist system in jeopardy (Virilio 2010).

Virilio's account is oriented to exploring the limitations of modern science which today are not concerned in understanding issues, but forecasting hazardous events in order for the current system of consumption not to be affected. Doubtless, it opens the doors for a more creative discussion to understand to what extent science and the society of risk are inextricably interlinked.

KNOWLEDGE IN THE SOCIETY OF RISK

The idea of a society of risk was coined by German Sociologist Ulrich Beck in observance of the role

of nuclear technology in Chernobyl, Ukraine. This sudden disaster not only shocked public opinion worldwide, but also evinced the vices and limitations of experts in the manner technology are manipulated. Those instruments which originally would be featured for the benefits of humankind now result in global risks that put society upside down. In this vein, Beck adheres to the thesis that the presence of risk makes societies more egalitarian since all classes are united by the same problems. Paradoxically, this accident created a radical shift in how knowledge is produced and disseminated to lay-people. With the advent of the society of risks, consumption replaced ethics in the line with the fact that hope for full employment sets the pace to insecurity and fear. While lay-people advanced in their knowledge of the world, risks appealed to a process of "reflexibility" where top-down authority of experts was placed under the scrutiny (Beck, 2002; 2009; 2015).

Following this ground-breaking contribution, Anthony Giddens goes on the open points unexplored in Beck's account. For Giddens, capitalism and industrialism are different projects which originally worked dissociated each other. Taking his cue from the "attachment theory", Giddens understands that social ties are represented in the ontological security developed by individuals from their childhoods. At the time, the emerging force of tradition is eroded by post-modernity; the ontological security of citizens will be seriously damaged. Trust serves as a "protective cocoon" in order for people to face the daily threats produced by the system. Without the blessing of trust, lay-citizens are more frightening of future and will look shelter in technology. The process of reflexivity engendered more horizontal circuits of authority, and of course, the system of experts is confronted to the knowledge accessed by lay-people. To wit, the legitimacy of experts in the fields of medicine, to set an example, will depend on the credibility the professional instill in patients. In this stage, trust plays a crucial role configuring the social basis of a society that enthralled global risks as the mediators between citizens and their institutions

(Giddens & Pierson, 1998; Giddens 1991, 1999). In consonance with Virilio, Giddens advises that in a near future, knowledge will be regularized by means of reflexivity in a way that will lead science to be prone to discover elements that help forecasting risks instead of studying facts. At a first look, what would be more than interesting to discuss is to what extent, in post-capitalist societies, the use of digital-technology produces a more efficient censorship.

In regards to the dichotomy of Surveillance and censorship, a must-read book is *Liquid Surveillance*, authored by D. Lyon & Z Bauman (2013). Both agree though 9/11 has not created the current state of hyper-surveillance in modern capitalist societies, no less true is that this traumatic event accelerated its condition of reproduction. The already existent technologies are posed not only to make for public spaces, or households a safer place, but also to remind the importance of those persons who can pay for private security or closed-circuits of satellite control. This book suggests two significant ideas. First and foremost, the state is unable to protect their citizens because the power was conferred to trade. Modern nation state is obliged to give solutions for problems created elsewhere. This engendered a sense of anomie, by which the citizens feel vulnerable. Secondly, the introduction of surveillance technology makes an unsafe world. The quest of order that characterizes the human existence is determined by the needs of change. The paradox lies in everything what we do, is to create a sense of stability we never will reach before death. If the society of risk imagined by Beck considered the risk as a result of action, the liquid surveillance goes in another direction.

As this argument given, Rebecca Fiske reviews the intersection of the state of exception with law-making process, a strand ignited by Schmitt and recently continued by Agamben. Fiske starts from the premise that any law derives from the authority of Gods, and remain affordable for humans in case of emergency. The paradox with this doctrine is that while the notion of exemption supposes breaking the law, sovereign serves as a

symbolic gatekeeper. While this applies very well for tribal societies, in industrial organizations where divinity exerts a moderate influence, the sense of exemption prevails. Fiske acknowledges that after 9/11 US committed some human right violations, in the name of security and well-being. While emergencies allow the delegation of further powers to executive branch, the judiciary often does not express any worry. This pattern would be very dangerous for democracy because it leads to dictatorship. After 9/11 the society of risk passed to a new stage where the spectacle of disasters would be the mainstream cultural value. Risk society mutated to Thana Capitalism, in the same way, the sense of protection set the pace to the needs of gazing others' death (Korstanje 2016). This will be fleshed out in next lines.

THE RISE OF THANA CAPITALISM

In one point, not only Giddens but also Beck took the wrong side of the road. There is nothing as the end of class-society in postmodernity. Neither classes have disappeared, nor are citizens living in a more egalitarian society. In the society of risk, some groups make the decisions that generate those risks other more vulnerable groups will face. Given this condition, those classes which monopolize the financial resources to mitigate risks are situated at the top of society, while lower classes are condemned to experience the catastrophes triggered by decisions that are made in external circles. Lower-classes are far from controlling risks in a Thana-Capitalism, they are subject to witness their effects alone.

Throughout Medieval Times, the access to knowledge was restricted to few hands. Without the print invented by Johannes Gutenberg's, the gathered information was monopolized by Catholic Church and scholars. Those agents, who defied status quo, were coactively repressed and condemned to death. This was an oft-used tactic of censorship that instilled panic in population. Violence and death were everywhere and a peas-

ant has an expectative of life of almost 30 years. However, the invention of print changed everything, enhancing the accessibility to books. The sense of control exerted by the Prince changed to subtler forms which were based on alienation and standardization. As Marshal McLuhan puts it, printing press accelerated a real revolution in European thinking to the extent to produce a mega system of global communication (Global Village) where the efficacy of the system was given by its ability to gather information. However, in this universe, technologies are not inventions only, which very well are affordable to users, but means to produce “identities” that form the character of peoples. High-tech enrooted in the culture to produce habits in users. In one direction, the creation of movable types was a turning point to reproduce thousands of books, which led towards a society of information, whereas unfortunately it was created a “tyranny of visual culture” cemented a *standardized-culture* (McLuhan, 2011).

In our current days, the society of risks examined by Giddens and Beck has gone forever, or at least it has mutated into a new form, *Thana-Capitalism*. Risk-Capitalism essentially was associated to bio politics and the needs of life, Thana Capitalism centres on the opposite, death. In sharp contrast to Medieval Age, in *Thana-Capitalism* people have a high expectancy of life, but live frightened by the possibility of death. Consumers are fascinated to witness spectacles where others are affected, killed or tortured. Death has become in the main cultural value in this new society, where 24 hours per day, TV and Journalism only transmit news linked to death. News that oscillates between disasters, terrorist attacks or local crimes daily captivate an audience more interested to feel more special than consternated for a run-a-way World. In many newscasts, it appears a message that say “images may hurt your sensibility”, or Warning, videos and images not for children”. At a closer look, even if audience in the developed societies are bombarded to catastrophic news with the end of keeping the society in an on-going

state of emergency, no less true is that death was become in a form of entertainment.

Following this argument, Phillipe Aries observed how in modern times though the expectances of life have notably extended; the symbolism of death has been more savage and disrupting for social order. We are now accustomed to live more years than our ancestors and it seems as though we have advanced in our efforts to discipline death, but now we are more afraid about the security of our loved relatives (Aries, 1975). To wit, five indicators are of paramount importance to observe the end of the Risk-Capitalism and the rise of *Thana-capitalism*.

We mean OMG as object Management Group. Further, Table 1 shows differences between Risk and Thana-Capitalism. The society drawn by Giddens and Beck corresponds with Risk-Capitalism. The organization of this system was centred on a decentralized mean of production where globalization attracted heavy investment in any part of the world. The maximization of profits and the reduction of costs was the main goal of risk-society. Even under-developed countries were more attractive for investors because the lower work-force costs respecting to central economies. As Taylor Gooby (2004) noted, the same technology that expanded the expectancy of life and enhanced the quality of life of elderly class produced a serious imbal-

Table 1. Differences of Risk and Thana-Capitalism

	Risk-Capitalism	Thana-Capitalism
Politics	Decentralized	Chaotic
Economy	Segments	Total Competence
Production	Consumption	Survival of the fittest
Main Value	Risk	Death
Event	Oil Embargo (1972)	11 September of 2001
Reaction	Risk Management	Victimization
Pressure Group	OMG	Victims
Censorship.	Fear	Over-production of information.

ance in the social security costs, because active workforce should absorb an ever growing demand of retirees that threatened the balance of the system. Secondly, the adoption of new technology to efficiently maximize profits of capital-owners reduced the supply of jobs for a precaritized blue-collar class. All these events conjoined to the oil Embargo caused by the Arab Israeli war, produced a serious glitch in the well fare state that initiated a new version of capitalism, more based on fear as a dissuasive element of control. In governing through Crime, Simon evinced how Elite appealed to fear in order for workforce to accept economic policies otherwise would be widely rejected (Simon 2007). Neoliberalism alluded to the sensationalism and fear disasters wake up in society in order to ensure a logic of exploitation over workforce (Klein 2007). Since the main cultural value in the society is risk, the reaction of citizens corresponds with the management of risk, which is associated to the needs of protection. As this backdrop, citizens go to private market in the absence of nation-state, where by the paying of a quota of capital they get a temporal sense of security. By means the buying insurances for protection the society of risk still produces new commodities to be exchanged in a global market. The tactic of discipline or censorship over citizens depends on the degree of fear elite can produce in society. Frightened by the presence of global fears that can end with the life as we know, citizens are sensitive to the market to pay for the security the state does not grant. However, one thing is using technology to predict disasters, and other go to the high-tech to produce a “theatralization of disasters”. In the second case, death is enthralled as an attractive factor of western culture in the contemporary World.

During the risk society, people lived afraid the disaster will be next to take place. The specialized literature suggests that risks are not disasters. Risk is a potential danger which can be occur or not. Therefore, the society of risk was oriented to a future. Understanding the dangers hidden in the future will be an efficient attempt to protect the

society. Thana-capitalism, rather, promotes disasters in order to entertain a global audience dubbed as “death-seekers”. Seeing how others die serves as a symbolic platform for watcher to feel how outstanding he or she is. This is the reason why, terrorism or news related to terrorist attacks are covered instantly by the media. We are fascinated by the state of disaster produced by risk society. Technology not only has disciplined the real external or internal threats but it has commoditized lower classes’ suffering as a new type of cultural entertainment (Korstanje 2016).

Doubtless, 9/11 was a founding event for the rise of Thana Capitalism. From this moment on, not only Western audience acknowledged that the world was an unsafe place to be, but they witnessed how their high-tech and transport applications can be used as weapons to the cultural core of the world. In this new era, coaction and violence were futile and censorship was politically manipulated to cause misunderstanding in lay-people. The term Thana-Capitalism stems from the word *Thanaptosis* which was originally coined by poet William C Bryant to symbolize the cycle of life and death. We live in order for others die, and we will die so that others can live. Furthermore, *Thanatos* (Θάνατος) comes the name given to a demon who represented the personification of “death”. In this new stage capitalism signalled to a much complex system of production which exploited the symbolism of death. Since the conditions of production passed from a decentralized model to a chaotic one, where all agents enter in competence with others for surviving, the doctrine of Thana-capitalism follows doctrines of social Darwinism where protection set the pace to *the Survival of the strongest*. Sometimes, holiday-making and vacations give some hints on how societies are changing. Our parents went to beaches or relaxing sites to spend their days while now tourists are exciting to visit spaces of disaster as Chernobyl, Katrina in New Orleans or even the ground zero in New York. Sociologically, if holidays-makers have developed an apollonian sense of beautiness, now they are in quest of spaces of mass-death and suf-

fering to revitalize the psychological frustrations they daily experience in their respective societies. Death has gained traction in thana-capitalism, captivating the admiration and becoming in a real tourist-attraction for many international segments (Bloom 2000; Stone & Sharpley 2008; White & Frew 2013). Knowing further on what purposes people travel exhibits a valid way of understanding on what values society is organized (Korstanje 2015). Is dark tourism a sign of egoism or a real attempt of understanding Death?

Several decades back, George H Mead, one of the fathers of symbolic interactionism, questioned why paradoxically many people are prone to read or listen of bad news presented by journalism, at the time they show preference by these types of news. What is our fascination for other's suffering? He assertively concludes that the self is configured by its interaction with others. This social dialectic alludes to anticipation and interpretation as two pillars of communication-process. The self feels happiness by other's suffering, because it represents a rite necessary to avoid or think in own pain. Starting from the premise the self is morally obliged to assist the other to reinforce its sentiment of superiority, Mead adds, this is the ethical nature of social relationship (Mead, 2009). The same remarks may apply for *thana-capitalism*.

Undoubtedly, the myth of Noah is one of the stories of the Bible that modern cinema has commoditized. This myth not only explains but also legitimize the material asymmetries produced by Thana-Capitalism. Let's remind readers that this God annoyed by the corruption of human beings, mandated to Noah to construct an ark. His divine mission consisted in gathering a pair by specie to achieve the preservation of natural life. The world destroyed by a great flood will be reconstructed by the specie protected by Noah. At a first glance, as the myth was ethically formulated, a formal message is based on the importance of nature and the problem of sin, corruption. But unconsciously, it poses the dilemma of competition where Thana-capitalism is centered. At any tournament or game, there can be only one winner.

Not only the creation but also Noah is witness of other's death, other's mass-death. The curiosity and fascination for death comes from this founding myth. The fact is that it can be found in Reality shows as big brother, a game widely studied by sociologists and detractors of visual technology rests on this principle. The main message of this game is only one participant is the chosen one to win. The doctrine of a selective-oriented salvation, enrooted in Protestantism claims for (though from different angles) understanding death as a token of vulnerability, lack of purity. The question whether people are happier when disaster takes place in other neighborhoods, it is still unclear the intersection of disasters and social bondage (Korstanje 2015; 2016).

It is important not to lose the sight that shrines reminding spaces of disasters are dispositiffs politically enrooted in the "spectacle of disaster", a term coined by Jean Baudrillard. Thematizing on disasters helps to the process of recovery of society in post disaster context, not only because society understands the lesson given by nature, but also poses as valid effort to discipline death. As disasters, death comes at any moment of life. This engenders much anxiety in survivors. In post-disasters environs, survivors develop a much deeper process of mourning to orchestrate the resiliency. They elaborate special rites to overcome the traumatic event that inflicted an extreme sentiment of pain. Any victim, before the climate of destruction, realizes that Gods were benevolent after all. Survivors, that way, embrace a climate of superiority by their subsistence was given by outstanding characteristics such as bravery, moral virtue and strength. This type of reaction helps community to recover to adversity but may generate sentiments of nationalism, superiority or ethnocentrism if it is not limited. The superiority of survivors, in this vein, depends on the other's misfortune. Unless this sentiment of narcissism to be regulated, it produces a clear rupture that affects social solidarity. Survivors feel they are super-heroes or special people touched by Gods, their mandates not only are sacred but also just. In

this moment, this sentiment of superiority feeds back a much deeper stage of victimization where survivors feel that pain is the only mean to reach happiness. Quite aside from this, their losses make them special for others who have not faced the same situation. This is exactly the instance where trust is undermined in order for the loyalty of citizens to be subject to market-place. Whatever the case may be, Thana-capitalism in sync with existent visual technology, not only exploits these types of climates but also obscures the causality of events. French ethnologist, Marc Augé acknowledged that the mass-media portrays tragic events blurring the connection between causes and consequences. News or stories focus on the effects instead on a clear diagnosis of reasons behind. As a result of this disasters' are continuously repeated once and once again (Augé, 2002). Not surprisingly, mass-media disseminates on a daily basis, news containing stories of victims to reinforce the above described sentiment of Narcissism.

CONCLUSION

In Thana-capitalism death is the main commodity which is produced and disseminated by the media worldwide. As in dark tourism practices, people are interested to gaze the Other's suffering but in fact far from being closer to others, it reinforces their happiness not to be touched by death. Since thana-capitalism is based on a social Darwinism where all are struggling against all to survive, others' death endorses an exclusive sign of supremacy for those who are still alive. Metaphorically speaking, in thana-capitalism life is understood as a thrilling trace where only few selected participants reach the glory. As in reality-shows as Big Brother, or Films of Hunger Games, Thana Capitalism stimulates the competence of all against all. Ideologically, narcissism allows the participation of innocent people in this cynic game. This suggests that participants of these games not only are unfamiliar with their real probabilities to fail, in view the fact that they are moved by their

own egocentrism. That way, Status quo which promotes the logic of divide and rule undermines any efforts of other groups to take the power. In these scenarios, never participants are sufficient stronger to break the arbitrariness as well as the law of elite. Of course, because this law cannot be broken, is the reason why Thana Capitalism reproduces serious material asymmetries in society where few citizens gain almost all, once the rest is pressed to live in the ruins. Without social Darwinism, Thana Capitalism, which should be framed into a new stage of capitalism, would never exist. In this system censorship is not exerted by coercion, which is symbolically rechannelled by the figure of death, but by over-production of scattered information. Articulated in a more efficient way, now modern users have access to a lot of information but at the same time their understanding is in decline.

As discussed throughout this chapter, at time of researching for some issues the records visible in the media prevent a coherent understanding of social issues, this multiplication of information paved the ways for a paradoxical situation where only few hands monopolize what can be or not written. If in medieval times, the censorship was oriented to violence now it is related to liberty and the scattered information that leads to misinformation. This is only possible in a new era, dubbed as "Thana-Capitalism". Nowadays, technology offers a fertile ground to visualize the suffering of "Others" connecting our self to death, but at the same time stimulating the over-production of information (which oscillates between books, magazines, papers to visual material) that keeps lay-people aside from the understanding of events. With the benefits of hindsight, we confirm that death, censorship and technology are inextricably intertwined. In medieval times where death was a very common thing in day-to-day life, produced-information was reduced to limited circles. The lower technical condition of production was prone to use violence to exercise censorship to unauthorized voices. People frightened death in the same extent than god. However, in secularized

Thana-capitalism people reject the possibility of a “heaven”, or eternal shelter. There is nothing beyond death which sounds interesting for consumers of Thana-capitalism. The main value of this society is not life, but death. Once death has been liberated from her cage, society neglected God adopting technology as a valid vehicle to expand life. The belief in heaven as an exemplary centre where only chosen souls enters have been replaced by an atmosphere of eternal competence here in hearth, where participants struggle to gain recognition. This technology is not restricted; it has been democratized to be used by wider audiences worldwide. However, this resulted in an unexpected consequence. The unlimited overproduction of digital information produced a subtle censorship by the decline of understanding. By abundance, the produced knowledge is scattered in thousand pieces almost impossible to gather into a puzzle. Last but not least, thana-capitalism foments an eternal atmosphere of competence only if real violence is reduced to zero or to the minimum desirable degree. If this thesis is correct, what is the position of violence in this model?.

LINES FOR FUTURE RESEARCH

The argument held here may find much criticism in liberal thinkers, who argues that capitalism helped not only to crystalize democracy but making of this world a less violent realm (Pinker, 2011). However, the problem of capitalism rests on the belief that few deserve much, while the rest are on the ruin. Almost 2% of global population concentrates 90% of produced wealth. In a recent book, Maximiliano Korstanje (2015) alarmed that this postmodern world can be compared to the film *Hunger Games*, or the reality show *Big Brother*. In both settings, participants are dominant of their conciseness because they remain unfamiliar with the real probabilities to fail. These competences, like liberal market, are based on the premise of social Darwinism that claims for “the survival of strongest”, which means that the glory of only one

equals to the failure of the whole rest. Participants not only over-value their own skills, but are confident of their strongholds. The stimulation of competition in the labour market, emulated by entertainment industry, resulted in two interesting dynamics. The vulnerability of rank-and-file workers associated to the decline of well-fare state facilitated the capital-owners to increase their profits and wealth, while risk was adopted as a new value for modern workers. After all, the spectacle of disasters confers to workers the idea they are co-manager of their own fate. As things come from worse to worst, workers not only are responsible by decisions others make, but in Thana-Capitalism, elite eludes its responsibilities to fail in paving the conditions for a fairer wealth distribution.

On another hand, capital flourishes in a climate of liberty to move across the world. The doctrine of liberty (democracy) stimulates in workers the needs of consuming, in terms of Donohue, a type of freedom from “want”. The change from a society of producers to consumers took hit after 40s decade. Ideologically, Americans have felt “superior” to other nations because they are enthralled as the main democratic and prosperous society; although more egalitarian at the surface, American citizens are subject to more work and consumption but less leisure. This happens because, in a pro consumer society, workers are bombarded with emulation and advertising creating the needs to buy. This not only jeopardized their real liberty to choose, but affects seriously to democracy. Detractors of capitalism, who pushed their focus on the arbitrariness of producers, were involuntarily responsible or conducive to the formation of a global society of consumers. Those denunciations on an economy that protect the interests of producers as well as the needs to adopt consumption to break the material asymmetries among classes, were two guiding concepts to the formation of a globalized version of capitalism, which paradoxically was prone to mass-consumption but producing serious vulnerabilities (Donohue 2003). Despite the hot-debate ignited in this chapter, we feel much deeper



investigation is necessary in order to provide an all-encompassing model to understand these times where death seems to be exchangeable in every spheres of society. Precisely, in Thana-Capitalism digital information acts as mechanism of censorship successfully orchestrated by over-abundance.

REFERENCES

- Aries, P. (1975). *Western attitudes toward death: From the middle ages to the present* (Vol. 3). Baltimore, MD: John Hopkins University Press.
- Augé, M. (2002). *Diario de guerra: el mundo después del 11 de septiembre*. Barcelona: Gedisa.
- Baudrillard, J. (1986). *America*. New York: Verso.
- Baudrillard, J. (1995). *The Gulf War Did Not Take Place*. Sydney: Power Publications.
- Baudrillard, J. (2001). *Impossible Exchange*. New York: Verso.
- Baudrillard, J. (2002). *Screened Out*. New York: Verso.
- Beck, U. (1992). *Risk Society: Towards a New Modernity* (Vol. 17). London: Sage.
- Beck, U. (2009). *World at Risk*. Cambridge, UK: Polity Press.
- Beck, U. (2015). *Ecological politics in an age of risk*. New York: John Wiley & Sons.
- Blom, T. (2000). Morbid-Tourism – a postmodern market niche with an example from Althrop. *Norwegian Journal of Geography*, 54(1), 29–36.
- Cuevas, A. (2005). The many Faces of Science and Technology relationships. *Essays in Philosophy*, 6(1), article 3
- Donohue, K. (2003). *Freedom From Want*. Baltimore, MD: John Hopkins University Press.
- Ellul, J. (1964). *The technological Society*. New York: Knopf.
- Fiske, R. (2016). The borders of Corruption: living in the state of exemption. In R. Cropf & T. Bagwell (Eds.), *Ethical Issues and Citizens Rights in the era of Digital Government Surveillance* (pp. 1–15). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-9905-2.ch001
- Giddens, A. (1991). *Modernity and self-identity: Self and society in the late modern age*. Stanford, CA: Stanford University Press.
- Giddens, A. (1999). Risk and responsibility. *The Modern Law Review*, 62(1), 1–10. doi:10.1111/1468-2230.00188
- Giddens, A., & Pierson, C. (1998). *Conversations with Anthony Giddens: Making sense of modernity*. Stanford, CA: Stanford University Press.
- Gold, M. K. (2012). *Debates in the digital humanities*. Minneapolis, MN: University of Minnesota Press. doi:10.5749/minnesota/9780816677948.001.0001
- Harvey, D. (1989). *The conditions of postmodernity: An enquiry into the origins of cultural change*. Blackwell.
- Klein, N. (2007). *The shock doctrine: The rise of disaster capitalism*. New York: Macmillan.
- Korstanje, M., & Skoll, G. (2014). Evaluating the Risks of Technological evolutions. In *Encyclopedia of Business Analytics and Optimization*, (pp. 873-887). Academic Press.
- Korstanje, M. E. (2015). *A Difficult World, examining the roots of Capitalism*. New York: Nova Science Publishers.
- Korstanje, M. E. (2016). *The Rise of Thana Capitalism and Tourism*. Abingdon, UK: Routledge.
- Lyon, D., & Bauman, Z. (2013). *Liquid surveillance: A conversation*. Cambridge, UK: Polity.
- Marcuse, H. (1991). *One Dimensional Man. Studies in ideology of advanced industry society*. London: Routledge.

Marx, K. (1967). Capital Volume 1: The Process of Capitalist Production. New York, NY: International Publishers.

Marx, K. (1975). Economic and Philosophical Manuscripts. In Early Writings. Harmondsworth, UK: Penguin Books.

Marx, K., & Frederick, E. (1983). Manifesto of the Communist Party. In E. Kamenka (Ed.), *The Portable Karl Marx* (S. Moore, Trans.; pp. 203–241). New York, NY: Penguin Books.

McLuhan, M. (2011). *The Gutenberg galaxy: The making of typographic man*. Toronto: University of Toronto Press.

Mead, G. H. (2009). *Mind, self, and society: From the standpoint of a social behaviourist* (Vol. 1). Chicago: University of Chicago press.

Pinker, S. (2011). *The better angels of our nature: Why violence has declined* (Vol. 75). New York: Viking.

Sennett, R. (2011). *The corrosion of character: The personal consequences of work in the new capitalism*. New York: WW Norton & Company.

Simon, J. (2007). *Governing through crime: How the war on crime transformed American democracy and created a culture of fear*. Oxford, UK: Oxford University Press.

Stone, P., & Sharpley, R. (2008). Consuming Dark-Tourism a Thanatological Perspective. *Annals of Tourism Research*, 35(2), 574–595. doi:10.1016/j.annals.2008.02.003

Taylor-Gooby, P. (2004). *New risks, new welfare: the transformation of the European welfare state*. Oxford, UK: Oxford University Press. doi:10.1093/019926726X.001.0001

Virilio, P. (2010). *The University of Disaster*. Cambridge, UK: Polity Press.

Warschauer, M., & Matuchniak, T. (2010). New technology and digital worlds: Analyzing evidence of equity in access, use, and outcomes. *Review of Research in Education*, 34(1), 179–225. doi:10.3102/0091732X09349791

Weber, M. (1958). Capitalism and rural society in Germany. In *Max Weber: Essays in sociology*, (pp. 363–85). Academic Press.

White, L., & Frew, E. (2013). *Dark Tourism: place and identity: managing and interpreting dark places*. London: Routledge.

ADDITIONAL READING

Alderman, E., & Kennedy, C. (1997). *The right to privacy*. New York: Vintage.

Altheide, D. L. (2004). Consuming terrorism. *Symbolic Interaction*, 27(3), 289–308. doi:10.1525/si.2004.27.3.289

Altheide, D. L. (2006). Terrorism and the Politics of Fear. *Cultural studies critical methodologies*, 6(4), 415–439.

Altheide, D. L. (2009). *Terror post 9/11 and the media* (Vol. 4). New York: Peter Lang.

Bazan, E. B. (2002). *The Foreign Intelligence Surveillance Act*. Nova Publishers.

García-Peñalvo, F. J., González-González, J. C., & Murray, M. (2012). Myelvin: A web-based informal learning platform for languages practice. [IJKSR]. *International Journal of Knowledge Society Research*, 3(1), 26–39. doi:10.4018/ijksr.2012010103

Howie, L. (2007). The terrorism threat and managing workplaces. *Disaster Prevention and Management: An International Journal*, 16(1), 70–78. doi:10.1108/09653560710729820



- Korstanje, M. E. (2013). Preemption and terrorism. When the future governs. *Cultura (Asociación de Docentes de la Universidad de San Martín de Porres)*, 10(1), 167–184.
- Korstanje, M. E. (2015). The spirit of Terrorism: Tourism, Unionization and Terrorism. *Pasos: Revista de turismo y patrimonio cultural*, 13(1), 239–250.
- Korstanje, M. E., & Clayton, A. (2012). Tourism and terrorism: Conflicts and commonalities. *Worldwide Hospitality and Tourism Themes*, 4(1), 8–25. doi:10.1108/17554211211198552
- Korstanje, M. E., Tzanelli, R., & Clayton, A. (2014). Brazilian World cup 2014: Terrorism, tourism, and social conflict. *Event Management*, 18(4), 487–491. doi:10.3727/152599514X14143427352391
- Kraft, E., & Wang, J. (2010). An exploratory study of the cyberbullying and cyberstalking experiences and factors related to victimization of students at a public liberal arts college. [IJT]. *International Journal of Technoethics*, 1(4), 74–91. doi:10.4018/jte.2010100106
- Lyon, D. (2002). Surveillance Studies: Understanding visibility, mobility and the phenetic fix. *Surveillance & Society*, 1(1), 1–7.
- Lyon, D. (2003). *Surveillance after september 11* (Vol. 11). Cambridge: Polity.
- Lyon, D. (2004). Globalizing Surveillance Comparative and Sociological Perspectives. *International Sociology*, 19(2), 135–149. doi:10.1177/0268580904042897
- Lyon, D. (2007). *Surveillance studies: An overview*. Cambridge: Polity.
- Lyon, D., & Zureik, E. (Eds.). (1996). *Computers, surveillance, and privacy*. Minneapolis: U of Minnesota Press.
- Mason, R. O. (1986). Four ethical issues of the information age. *Management Information Systems Quarterly*, 10(1), 5–12. doi:10.2307/248873
- OBrien, M., Tzanelli, R., Yar, M., & Penna, S. (2005). The Spectacle of Fearsome Acts: Crime in the Melting Pot in Gangs of New York. *Critical Criminology*, 13(1), 17–35. doi:10.1007/s10612-004-6111-9
- Palanisamy, R., & Mukerji, B. (2012). Security and Privacy issues in e-Government. IGI Global, 236–248.
- Reitz, J. C. (2006). E-government. *The American Journal of Comparative Law*, 54, 733–754.
- Shaw, L. E. (2004). USA Patriot Act of 2001, the Intelligence Reform and Terrorism Prevention Act of 2004, and the False Dichotomy between Protecting National Security and Preserving Grand Jury Secrecy, The. *Seton Hall Law Review*, 35, 495.
- Tzanelli, R. (2004). Giving gifts (and then taking them back): Identity, reciprocity and symbolic power in the context of Athens 2004. *Journal for Cultural Research*, 8(4), 425–446. doi:10.1080/1479758042000310117
- Tzanelli, R. (2006). Capitalizing on value: Towards a sociological understanding of kidnapping. *Sociology*, 40(5), 929–947. doi:10.1177/0038038506067516
- Wade, L. (2003). Terrorism and the internet: Resistance in the information age. *Knowledge, Technology & Policy*, 16(1), 104–127. doi:10.1007/s12130-003-1018-4

KEY TERMS AND DEFINITIONS

Digital Surveillance: The use of digital technology to control others, citizens and the life of a nation.

Censorship: It means the suppression of free speech rights or any right to express own ideas.

Risk Society: The ways a society reacts against the rise of risks. Risk society alludes to the needs of forecasting future to prevent potential threats that may affects societal order.

Thana-Capitalism: This represents a new term just coined in this manuscript where death situates as the main commodity of good exchange process and current economic systems. Thana capitalism replaced risk society after 9/11.

Globalization: Refers to a cultural project of integration of economies and networks which leads to multiculturalism and interchange of worldviews.



Performance Measurement of Technology Ventures by Science and Technology Institutions

Artie W. Ng

The Hong Kong Polytechnic University, Hong Kong

Benny C. F. Cheung

The Hong Kong Polytechnic University, Hong Kong

Peggy M. L. Ng

The Hong Kong Polytechnic University, Hong Kong

INTRODUCTION

Developing and transitional economies are keen to advance their science and technology in augmenting social and economic progress so as to catch up with developed economies. While the concept of national innovation system gains significant interest among emerging economies, institutions for science and technological innovation have been established to nurture the growth and expansion of technology ventures (Freeman, 1995). Since the early 1980s, institutions in the USA, UK, Germany, France, and Japan have been established to stimulate their countries' economic development through technological innovation and advancements of their emerging industries. Local universities have been involved as well. For instance, Massachusetts Institute of Technology (MIT) in the U.S. developed its Industrial Liaison Program in 1948 (Liu & Jiang, 2001). In China, the University-Industry Cooperation Committee of Tsinghua University (UICCTU) was established in 1995, providing various collaborative services for the member companies including IBM, Motorola, Hitachi, and NEC (Liu & Jiang, 2001).

In recent decades, an increasing number of science and technology institutions (STIs) have been founded to pursue the development of high technology companies in partnership with enter-

prises with the aim of undertaking joint projects for the assessment of new technology ventures (De Coster & Butler, 2003). The growing importance of technological innovation in national economic development has been increasingly recognized. China, as a transitional economy, has stepped up its Research and Development (R&D) expenditures and advocates institutional collaboration with the private sector for innovation performance (Boeing, Mueller, & Sandner, 2016; Kafourous, Wang, Pipropoulos, & Zhang, 2016). Hong Kong, China's special administrative region, established its Innovation and Technology Bureau in early 2016 as an effort of the government to foster innovation and technology industry in collaboration with local universities and its Science Park.

However, challenges remain significant in performance measurement and management for STIs' effective development, given the intention to reduce the rate of failures among early-stage ventures. Prior studies looked into issues with managing performance of technology ventures (Ganotakis & Love, 2012; Li & Atuahene-Gima, 2001). These studies that focus on specific dynamics of growth and development from the standpoint of resources allocation process did not fully examine pertinent externalities. Externalities, such as market competitions, applications, and commercialization of emerging disruptive technolo-

gies, would further complicate the environment and potential success of new technology ventures.

This chapter explores the dynamics between resources management and external landscapes as critical consideration in performance measurement and management of technology ventures. Based on an interdisciplinary literature review, a pertinent framework is developed with a set of monitoring areas and success indicators. The authors argue that STIs would have to utilize their expertise in specific clusters as intangible resources to assess product development and innovation capabilities as competitive strengths among the ventures. However, there could be limitation to the efficacy of a top-down, mechanical planning and control approach adopted by STIs to facilitate the innovative development of an emerging technology sector, given the constant dynamics of external markets.

BACKGROUND

Technology Monitoring and Assessment by STIs in Emerging Economies

STIs require a systematic corporate approach to monitor and assess the ongoing development of science and technology in order to capture the next big waves. At a country level, the concept of national innovation system is supported by the use of technology foresight techniques to strengthen its effectiveness among developed countries (Martin & Johnston, 1999). The use of a national innovation system is also advocated for the promotion and formation of partnership for effective technology development within a developing country (Hall, Bockett, Taylor, Sivamohan, & Clark, 2001). Bradford, Kinzey, and Gunn (1991) pointed out the importance of having a structured approach to monitor programs in both public and private sectors on a global basis. Their study also proposed a schematic to outline both internal and external aspects, as well as to keep abreast of technological changes in order to ensure a healthy development

of an organization. Nevertheless, expertise is needed to provide in-depth and timely review of the associated technology sectors.

Lichtenthaler (2004) used a case study to outline a framework that demonstrates the need to assess technological change through structural, hybrid, and informal forms for coordinating technology intelligence processes. This framework is made of coordinative forms of technology intelligence at both company and technology levels. The dynamic environments of technology ventures, as investigated by Zahra and Bogner (2000), are highly crucial to the determination of these concerns' internal development and exploitation of technological resources, and therefore to their performance and even survival. Under such a competitive environment, technology ventures are pressured to develop new products, upgrade them, and strengthen R&D in a timely manner, especially under the highly competitive markets of developed economies.

However, STIs could be hindered by their internal bureaucracy and formal constraints while the external environment changes rapidly. As North (1990) noted in his examination of institutional performance,

The major role of institutions in a society is to reduce uncertainty by establishing a stable (but not necessarily efficient) structure to human interaction. But the stability of institutions in no way gainsays the fact that they are changing.... Institutions typically change incrementally rather than in discontinuous fashion. (p. 6)

In light of emergence of disruptive technologies, an institution could have neglected such changing externalities and missed an opportunity to counter swiftly.

Challenges in Nurturing Early-Stage Technology Ventures

The emergence of technology ventures creates positive effects in modern economic development fostering innovations of industries; however,



early-stage technology ventures typically face a number of challenges. One of the main concerns is related to these enterprises' optimal utilization of limited resources during their early development. Chan and Lau (2005) examined the criticalness of resources provided by the incubators to technology start-ups in a science park. In financing the expansion of a venture, evidence shows that agency costs could become significant when there are less tangible assets, more growth options, and lower asset specificity (Gompers, 1995). As a result, venture capitalists will incur increased monitoring costs, especially when R&D intensities are higher (Gompers, 1995).

New technology ventures suffer from a "liability of newness", which can be seen as the root problems and struggle to survive in their early stages (Aspelund, Berg-Utby, & Skjevdal, 2005). Thus, managing internal capabilities in the commercialization process and ensuring resource development effectiveness are crucial in assuring young technology ventures' survival and growth (Burgelman, 2012). Another challenge for new ventures is the conspicuous lack of organizational experience leading to "exacerbating" the cost of internationalization; therefore, a large resource endowment is considered an advantage so as to enhance survival and growth (Sapienza, Autio, George, & Zahra, 2006).

Range of Intellectual Capital as Organizational Resources and Performance Drivers

The resource-based view (RBV) of technology ventures focuses on a bundle of intangible resources which are valuable, scarce, and difficult to imitate, but combined to significantly affect a venture's performance (Lee, Lee, & Pennings, 2001). The theory of RBV considers intangible resources as critical for technology ventures. In order to grow technology ventures, the acquisition of intangible resources, such as human resources and managerial resources, are substantial in es-

tablishing ventures' durable competitive advantages and performance (Brinckmann & Hoegl, 2011). Eisenhardt and Schoonhoven (1990), as well as Brüderl and Preisendörfer (2000), further emphasized the attributes of ventures' top executives, including executives' education and gender, management-specific skills, and level of work experience, would help boost up the success of technological start-ups. In an earlier study, Cooper, Gimeno-Gascon, and Woo (1994) noted that initial human capital and financial resources are key factors to predict a new venture's future performance. As young ventures lack experienced entrepreneurs, financial resources are becoming significant for new ventures. Although financial resources are not based on RBV as they are tradable, non-scare and invaluable, they can be considered as a sustainable competitive advantage for new technology ventures (Lee et al., 2001).

In order to facilitate subsequent growth to another stage, functional specializations as well as heterogeneous resources, composed of tangibles and intangibles, should be available to support the growth of technology ventures (Kazanjian & Drazin, 1990). Lee et al. (2001) revealed that entrepreneurial orientation includes organizational processes, methods, and styles used to implement ventures, whereas technological capabilities are related to patents protected by law, technological knowledge, and production skills.

In reviewing "spin-off" high-technology ventures originated from research performed by a university, De Coster and Bulter (2003) found that these ventures are able to maintain relatively more competitive advantages than the others due to their protected intellectual property. The authors also noted the criticalness of the product innovation level and market potentials. An optimal combination of intangible and tangible resources, considered as intellectual capital, during various stages of development would provide a potential source of competitive advantage for technology ventures in the new economy (Hayton, 2005; Ng, 2006, Ng 2008).

Product Innovation Capabilities

Prior studies indicated that product innovation is a critical driver in measuring technology performance (Li & Atuahene-Gima, 2001). As managerial and financial resources are normally limited for new ventures, product innovation typically involved a certain level of risk (Eisenhardt & Schoonhoven, 1990). In order to succeed, technology ventures must swiftly devise their new products and subsequently innovate them, especially within a market that could be influenced by disruptive technologies (Nambisan, 2002; Schilling & Hill, 1998). Product development itself could be an inherently risky process, in particular when new technology is adopted (Littler, Leverick, & Bruce, 2003). Its production effectiveness and the time to market would further make a new product fragile to competition. The overall product innovation capabilities need to be assessed carefully as part of a venture's performance evaluation.

Dynamic Performance Measurement System for Technology Ventures

Under highly competitive pressures within a technology sector, new ventures with support of limited financial resources struggle to sustain their performances as closely monitored by their equity stakeholders. In particular, venture capitalists are found to be an important group of investors who would use both formal and informal means to monitor the performance of their invested portfolio of ventures while constantly performing strategic analysis of the external competitions (Gorman & Sahlman, 1989). Despite this rather unstructured role, their monitoring style reflects a dynamic characteristic, as a venture's performance needs to be examined in light of the externalities.

Laitinent (2002) further explained the concept of a dynamic performance measurement system which would enable the stakeholders, including the management, to utilize an integrated approach to actively review competitiveness and the internal resource allocation processes. Bititci, Garengo,

Dörfler, and Nudurupati (2012) pointed out the real challenge lies in the development of a dynamic performance measurement system as a social system that enables learning in "autopoietic networks". Bititci et al.'s (2012) study suggests that taking a holistic approach in the development of a dynamic performance measurement system would recognize the concurrent nature of challenges that the stakeholders face in technological contexts.

Alignment of Interests Among the Stakeholders

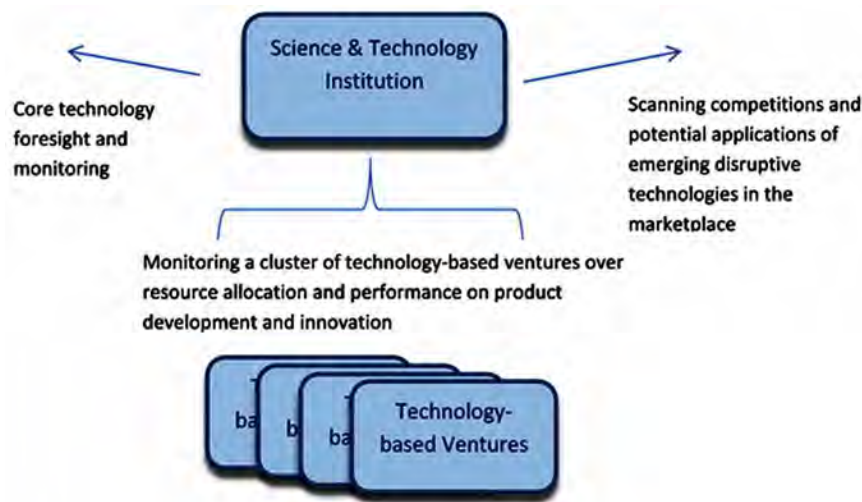
As a whole, there should be consideration for the alignment of interest among the stakeholders within a performance measurement system in order to facilitate a new venture's goal congruence (Chenhall, 2005). Objectives differ across stakeholders and may have different time horizon; thus, at institutional level, stakeholders need to share a long-term interest and reward system that motivates them to enhance performance. At organizational level, the founding management team's performance needs to be incentivized. In order to avoid myopic behavior, the design of the alignment of interests is particularly crucial for the institutions which are not necessarily motivated by economic interest on an individual basis, but for their institutional objectives. Effective communication should be utilized to accelerate stakeholders' alignment of interests. Indeed, continuous and direct interaction among stakeholders is the basis to establish common institutional objectives (Al-Kwafi & Ahmed, 2013).

Proposed Framework

As Neely, Gregory, and Platts (1995) pointed out, as no single measure can represent all the salient characteristics of an activity, a set of measures is necessarily required for a performance measurement system. Figure 1 shows the framework of a dynamic performance measurement system for technology ventures. It consists of three dimensions, suggesting the dynamic relationship be-



Figure 1. Framework of dynamic performance measurement system for technology ventures



tween a STI and a cluster of technology ventures under its supervision and monitoring. First, STIs need to maintain a timely technology foresight with constant assessments and identification of relevant core technologies. Second, STIs should assess the technological position of ventures in relation to their competitors with respect to new product innovation and commercialization by exploiting new technologies (Geum, Lee, Kang, & Park). Third, STIs have a critical role in the assessment and monitoring of emerging technologies, as well as of external competitors from the market place, in order to make appropriate judgment about their ventures' performance.

In addition to reviewing the composition of intellectual capital within these ventures, the supervising institutions need to proactively review the progress of their product development and innovation in light of the evolving externalities. Institution-related resources, such as the technological infrastructure provided by the institution, quality of its staff, and industrial networks could make a difference in the venture development (González-Pernía, Kuechle, & Peña-Legazkue, 2013). These STIs in turn are required to develop their internal capabilities, including funding, government support, and high quality research staff (Liu & Jiang, 2001). Such capabilities enable researching and analyzing the landscape

of technological innovation within their core technology clusters while anticipating potential disruptive technologies.

Scanning Competitions and Potential Applications of Emerging Disruptive Technologies

It is important for STIs to evaluate potential applications of emerging disruptive technologies so as to assess the relative competitive position regarding potential disruptive technologies of ventures. This can be done by comparing the key disruptive technology's performance of the ventures and of the major competitors on the focal mainstream (Yu & Hang, 2010). An STI may also collect pertinent information and data from the customers, manufacturers, dealers, consultants, and experts. Some may adopt an importance-performance matrix as a tool to evaluate the performance of internal and external services using two measurement dimensions: importance to customer and performance of services (Slack, 1994). Leong and Tan (1992) also developed an importance-performance matrix approach to assess a firm's national competitive superiority, which was measured by two dimensions: importance to investors and country's competitiveness.

In order to evaluate potential applications of emerging disruptive technologies, a technology strength-competitiveness matrix is proposed in Figure 2. Potentials of disruptive technologies can be measured by two dimensions: strength of technological innovation within and competitive technology position. Under strength-competitiveness matrix, the latter's competitive position in the marketplace is evaluated with respect to specific technological innovation in product development, commercialization and resulting market share in comparison with external competitors. The ultimate goal of each technology venture is to achieve the goal of the technology leader and highly competitive technology position (high-lead) of the matrix. Internal strength of technological innovation refers to a venture's ability to develop potentially disruptive technologies on a continuous basis, largely through dedicated R&D activities.

Areas of Monitoring and Success Indicators

A set of structured tools need to be developed in order to monitor the performance of the technology ventures that are engaged in formal

collaboration with STIs and to keep track of these ventures' dynamic performance within a technology cluster. Relevant success indicators in the areas of resources capabilities, emerging disruptive technology, scanning competitions. and performance of new technology ventures are proposed to allow STIs to monitor key aspects of performance (see Table 1).

FUTURE RESEARCH DIRECTIONS

In future studies, the framework proposed in this article will be integrated with specific performance indicators for an evaluation from early-stage to expansion-stage ventures. Alignment of interests between an institution and the ventures being monitored needs to be examined in order to ensure the overall validity of this dynamic performance measurement system. Specific case studies on a science and technology cluster can provide evidence about the relevance of this proposed framework while examining the needs to develop particular performance indicators for the cluster.

Other studies may also explore the influence of a national culture over performance management for technological innovation. In particular, a top-down innovation system under a command economy largely driven by central planning, such as China, should be explored for its differences in institutional performance versus a highly market-based system in the West, which is mostly driven by free flow of information, rationing, and motivation under a competitive market economy. Empirical studies or in-depth case studies under national cultural logics can provide evidence about the relevance of various components of this proposed framework. Comparative studies with Western economies, that have prevailed in intangibles and software developments, would further reveal how emerging technology enterprises from China have grown and developed from its sizable domestic market, with emphases on technological infrastructure and hardware, (e.g., China has identified its high-speed railway technology as

Figure 2. Strength-competitiveness matrix on disruptive technologies

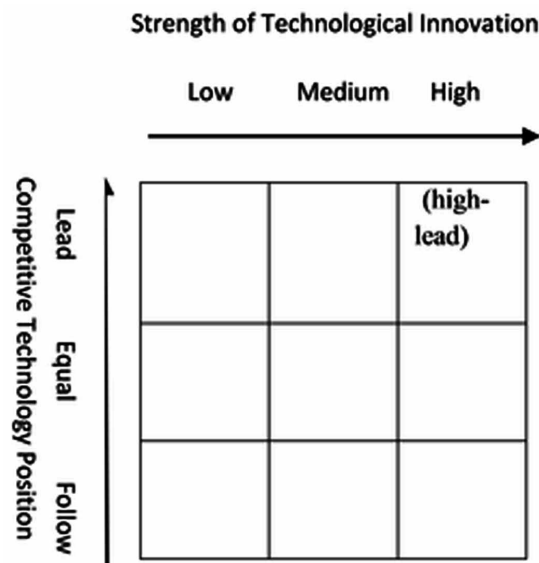


Table 1. Areas of monitoring and success indicators

Areas of Monitoring	Characteristics	Indicators
Resources capabilities	Entrepreneurial orientation, technological capabilities / knowledge assets, and human capital	• Number of high quality R&D employees (Lee et al., 2001). • Range of technologies that were internally developed (Ahuja & Katila, 2001). • Effectiveness of R&D investment (Motohashi, 2005). • Number of patents obtained or submitted (Motohashi, 2005; Perkmann, Neely, & Walsh, 2011). • Overall technological capabilities pertinent to new product design and development (Lee et al., 2001).
Performance of a cluster of technology ventures	Ability to sustain continuous growth and development as a new business venture.	• Exponential growth in employment (Brüderl & Preisendörfer, 2000). • Time to market or product commercialization process (Snow Fjeldstad, Lettl, & Miles, 2011). • Sales volumes (Lee et al., 2001). Sales growth (Lau & Bruton, 2011).
Emerging disruptive technologies	Potential applications of emerging disruptive technologies.	• Count of emerging disruptive technologies (Yu & Hang, 2010). • Commercialization opportunities (Frishammar, Lichtenthaler, & Rundquist, 2012). • Strength of technological innovation (see Fig. 1)
Scanning competitions	External competitions with the global marketplace.	• The level of relative technology competitive position (see Fig. 1)
Venture-institution collaboration/ integrative efforts	Success in collaboration between technology ventures and institution.	• Joint objective setting (Perkmann et al., 2011) • Interaction intensity (Lee et al., 2001) • Effective formal and informal communication (Al-Kwafi & Ahmed, 2013)

part of its national strategy for its infrastructure development with timely commercialization through domestic market penetrations).

CONCLUSION

In this article, a framework that integrates critical knowledge about technology monitoring and assessment of potential disruptive technologies from the standpoint of an STI is proposed on the basis of an interdisciplinary literature review. An effective integration of emerging disruptive technologies into venture strategy is conceived as a critical aspect for the development and growth of technology ventures. An STI is contemplated as a technology assessor that monitors a cluster of ventures continuously pursuing technological innovation to enhance their products. The proposed framework and its dimensions are introduced to help evaluate potential applications of emerging disruptive technologies and assess the relative competitive position of a new venture in harnessing pertinent disruptive technologies within a timeline of commercialization. The areas of monitoring and indicators with time lag potentials are devised to reflect the key dimensions embedded in a dynamic performance measurement system over time.

Further, we suggest there is a dynamic relationship between an STI and a cluster of technology ventures under its supervision and monitoring. An STI has a critical role in monitoring emerging technologies as well as external competitions from the global marketplace in order to make appropriate judgment about the performance of these ventures. Instead of maintaining a top-down, mechanical approach in its strategy formulation and implementation seemingly useful for infrastructural projects, the challenge for an STI in an emerging economy is to assess internal capabilities and more importantly externalities of emerging disruptive technologies over time, being critical to the development and growth of technology ventures striving for product innovation against competitions. Moreover, performance indicators

need to be adapted over time for instituting a culture that supports the dimensions of the proposed performance measurement system.

Nevertheless, the ability of an STI to interact with externalities and to anticipate disruptive technologies remains a challenge. As noted by North (1990), “*Institutions typically change incrementally rather than in discontinuous fashion.*” It is even more so for STIs to embrace changes with timely technology foresights as they have to deal with clusters that are undergoing rapid technological developments and emerging competitive forces on a global basis.

REFERENCES

- Ahuja, G., & Katila, R. (2001). Technological acquisitions and the innovation performance of acquiring firms: A longitudinal study. *Strategic Management Journal*, 22(3), 197–220. doi:10.1002/smj.157
- Al-Kwafi, O., & Ahmed, Z. U. (2013). Accessing external knowledge by Chinese firms: A conceptual framework. *Journal of Technology Management in China*, 8(1), 5–17. doi:10.1108/JTMC-04-2013-0021
- Aspelund, A., Berg-Utby, T., & Skjevdal, R. (2005). Initial resources influence on new venture survival: A longitudinal study of new technology-based firms. *Technovation*, 25(11), 1337–1347. doi:10.1016/j.technovation.2004.06.004
- Bititci, U., Garengo, P., Dörfler, V., & Nudurupati, S. (2012). Performance measurement: Challenges for tomorrow. *International Journal of Management Reviews*, 14(3), 305–327. doi:10.1111/j.1468-2370.2011.00318.x
- Boeing, P., Mueller, E., & Sandner, P. (2016). China's R&D explosion—Analyzing productivity effects across ownership types and over time. *Research Policy*, 45, 159–176.
- Bradford, A. W., Kinzey, B. R., & Gunn, M. E. (1991). A structured approach for monitoring science and technology developments. *International Journal of Technology Management*, 6(1-2), 91–111.
- Brinckmann, J., & Hoegl, M. (2011). Effects of initial teamwork capability and initial relational capability on the development of new technology-based firms. *Strategic Entrepreneurship Journal*, 5(1), 37–57. doi:10.1002/sej.106
- Brüderl, J., & Preisendörfer, P. (2000). Fast-growing businesses: Empirical evidence from a German study. *International Journal of Sociology*, 30(3), 45–70. doi:10.1080/15579336.2000.11770218
- Burgelman, R. A. (2012). Managing the internal corporate venturing process. *Sloan Management Review*, 25(2), 33–48.
- Chan, K. F., & Lau, T. (2004). Assessing technology incubator programs in the science park: The good, the bad and the ugly. *Technovation*, 24(10), 1215–1228.
- Chenhall, R. H. (2005). Integrative strategic performance measurement systems, strategic alignment of manufacturing, learning and strategic outcomes: An exploratory study. *Accounting, Organizations and Society*, 30(5), 395–422. doi:10.1016/j.aos.2004.08.001
- Cooper, A. C., Gimeno-Gascon, F. J., & Woo, C. Y. (2002). Initial human and financial capital as predictors of new venture performance. *Journal of Business Venturing*, 9(5), 371–395. doi:10.1016/0883-9026(94)90013-2
- De Coster, R., & Bulter, C. (2003). Assessment of proposals for new technology ventures in the UK: Characteristics of university spin-off companies. *Technovation*, 25(5), 535–543. doi:10.1016/j.technovation.2003.10.002

- Eisenhardt, K. M., & Schoonhoven, C. B. (1990). Organizational growth: Linking founding team, strategy, environment, and growth among US semiconductor ventures, 1978-1988. *Administrative Science Quarterly*, 35(3), 504–529. doi:10.2307/2393315
- Freeman, C. (1995). The ‘National System of Innovation’ in historical perspective. *Cambridge Journal of Economics*, 19(1), 5–24.
- Frishammar, J., Lichtenthaler, U., & Rundquist, J. (2012). Identifying technology commercialization opportunities: The importance of integrating product development knowledge. *Journal of Product Innovation Management*, 29(4), 573–589. doi:10.1111/j.1540-5885.2012.00926.x
- Ganotakis, P., & Love, J. H. (2012). The innovation value chain in new technology-based firms: Evidence from the UK. *Journal of Product Innovation Management*, 29(5), 839–860. doi:10.1111/j.1540-5885.2012.00938.x
- Geum, Y., Lee, S., Kang, D., & Park, Y. (2011). Technology roadmapping for technology-based product–service integration: A case study. *Journal of Engineering and Technology Management*, 28(3), 128–146. doi:10.1016/j.jengtecman.2011.03.002
- Gompers, P. (1995). Optimal investment, monitoring, and the staging of venture capital. *The Journal of Finance*, 50(5), 1461–1489. doi:10.1111/j.1540-6261.1995.tb05185.x
- González-Pernía, J. L., Kuechle, G., & Peña-Legazkue, I. (2013). An assessment of the determinants of university technology transfer. *Economic Development Quarterly*, 27(1), 6–17. doi:10.1177/0891242412471847
- Gorman, M., & Sahlman, W. A. (1989). What do venture capitalists do? *Journal of Business Venturing*, 4(4), 231–248. doi:10.1016/0883-9026(89)90014-1
- Hall, A., Bockett, G., Taylor, S., Sivamohan, M. V. K., & Clark, N. (2001). Why research partnerships really matter: Innovation theory, institutional arrangements and implications for developing new technology for the poor. *World Development*, 29(5), 783–797. doi:10.1016/S0305-750X(01)00004-3
- Hayton, J. C. (2005). Competing in the new economy: The effect of intellectual capital of corporate entrepreneurship in high-technology new ventures. *R & D Management*, 35(2), 137–155. doi:10.1111/j.1467-9310.2005.00379.x
- Kafouros, M., Wang, C., Piperopoulos, A., & Zhang, M. (2016). Academic collaborations and firm innovation performance in China: The role of region-specific institutions. *Research Policy*, 44(3), 803–817. doi:10.1016/j.respol.2014.11.002
- Kazanjian, R. K., & Drazin, R. (1990). A stage-contingent model of design and growth for technology based new ventures. *Journal of Business Venturing*, 5(3), 137–150. doi:10.1016/0883-9026(90)90028-R
- Laitinen, E. K. (2002). A dynamic performance measurement system: Evidence from small Finnish technology companies. *Scandinavian Journal of Management*, 18(1), 65–99. doi:10.1016/S0956-5221(00)00021-X
- Lau, C. M., & Bruton, G. D. (2011). Strategic orientations and strategies of high technology ventures in two transition economies. *Journal of World Business*, 46(3), 371–380. doi:10.1016/j.jwb.2010.07.011
- Lee, C., Lee, K., & Pennings, J. M. (2001). Internal capabilities, external networks, and performance: A study on technology-based ventures. *Strategic Management Journal*, 22(6-7), 615–640. doi:10.1002/smj.181
- Leong, S. M., & Tan, C. T. (1992). Assessing national competitive superiority: An importance-performance matrix approach. *Marketing Intelligence & Planning*, 10(1), 42–48. doi:10.1108/02634509210007902

- Li, H., & Atuahene-Gima, K. (2001). Product innovation strategy and the performance of new technology ventures in China. *Academy of Management Journal*, 44(6), 1123–1134. doi:10.2307/3069392
- Lichtenthaler, E. (2004). Technological change and technology intelligence process: A case study. *Journal of Engineering and Technology Management*, 21(4), 331–348. doi:10.1016/j.jengtecman.2004.09.003
- Littler, D., Leverick, F., & Bruce, M. (2003). Factors affecting the process of collaborative product development: A study of UK manufacturers of information and communications technology products. *Journal of Product Innovation Management*, 12(1), 16–32. doi:10.1016/0737-6782(94)00025-B
- Liu, H., & Jiang, Y. (2001). Technology transfer from higher education institutions to industry in China: Nature and implications. *Technovation*, 21(3), 175–188. doi:10.1016/S0166-4972(00)00045-6
- Martin, B. R., & Johnston, R. (1999). Technology foresight for wiring up the national innovation system: Experiences in Britain, Australia and New Zealand. *Technological Forecasting and Social Change*, 60(1), 37–54. doi:10.1016/S0040-1625(98)00022-5
- Motohashi, K. (2005). University–industry collaborations in Japan: The role of new technology-based firms in transforming the National Innovation System. *Research Policy*, 34(5), 583–594. doi:10.1016/j.respol.2005.03.001
- Nambisan, S. (2002). Complementary product integration by high-technology new ventures: The role of initial technology strategy. *Management Science*, 48(3), 382–398. doi:10.1287/mnsc.48.3.382.7724
- Neely, A., Gregory, M., & Platts, K. (1995). Performance measurement system design: A literature review and research agenda. *International Journal of Operations & Production Management*, 15(4), 80–116. doi:10.1108/01443579510083622
- Ng, A. W. (2006). Reporting intellectual capital flow in technology-based companies: Case studies of Canadian wireless technology companies. *Journal of Intellectual Capital*, 7(4), 492–510. doi:10.1108/14691930610709130
- Ng, A. W. (2008). The value relevance of intangible assets, human capital and corporate governance disclosed in prospectuses of growth enterprises: Evidence from Hong Kong. *International Journal of Learning and Intellectual Capital*, 5(3/4), 431–454. doi:10.1504/IJLIC.2008.021021
- North, D. C. (1990). *Institutions, Institutional Change and Economic Performance*. New York: Cambridge University Press. doi:10.1017/CBO9780511808678
- Perkmann, M., Neely, A., & Walsh, K. (2011). How should firms evaluate success in university–industry alliances? A performance measurement system. *R & D Management*, 41(2), 202–216. doi:10.1111/j.1467-9310.2011.00637.x
- Sapienza, H. J., Autio, E., George, G., & Zahra, S. A. (2006). A capabilities perspective on the effects of early internationalization on firm survival and growth. *Academy of Management Review*, 31(4), 914–933. doi:10.5465/AMR.2006.22527465
- Schilling, M. A., & Hill, C. W. L. (1998). Managing the new product development process: Strategic imperatives. *The Academy of Management Executive*, 12(3), 67–81.
- Slack, N. (1994). The importance-performance matrix as a determinant of improvement priority. *International Journal of Operations & Production Management*, 14(5), 59–75. doi:10.1108/01443579410056803

Snow, C. C., Fjeldstad, Ø. D., Lettl, C., & Miles, R. E. (2011). Organizing continuous product development and commercialization: The collaborative community of firms model. *Journal of Product Innovation Management*, 28(1), 3–16. doi:10.1111/j.1540-5885.2010.00777.x

Yu, D., & Hang, C. C. (2010). A reflective review of disruptive innovation theory. *International Journal of Management Reviews*, 12(4), 435–452. doi:10.1111/j.1468-2370.2009.00272.x

Zahra, S. A., & Bogner, W. C. (1999). Technology strategy and software new ventures performance: Exploring the moderating effect of the competitive environment. *Journal of Business Venturing*, 15(2), 135–173. doi:10.1016/S0883-9026(98)00009-3

ADDITIONAL READING

Koberg, C. S., Uhlenbruck, N., & Sarason, Y. (1996). Facilitators of organizational innovation: The role of life-cycle stage. *Journal of Business Venturing*, 11(2), 133–149. doi:10.1016/0883-9026(95)00107-7

Liu, M. C., & Chen, S. H. (2012). MNCs offshore R&D networks in host countrys regional innovation system: The case of Taiwan-based firms in China. *Research Policy*, 41(6), 1107–1120. doi:10.1016/j.respol.2012.03.016

Mindruta, D. (2013). Value creation in university-firm research collaborations: A matching approach. *Strategic Management Journal*, 34(6), 644–665. doi:10.1002/smj.2036

Segers, J. P. (2015). The interplay between new technology based firms, strategic alliances and open innovation, within a regional systems of innovation context. The case of the biotechnology cluster in Belgium. *Journal of Global Entrepreneurship Research*, 5(1), 1–17. doi:10.1186/s40497-015-0034-7

Yan, X., & Yu, C. C. (2013). Strengths and weaknesses of Hong Kongs technology and innovation industry with reference to the extended open innovation model. *Journal of Science and Technology Policy in China*, 4(3), 180–194. doi:10.1108/JSTPC-10-2013-0010

KEY TERMS AND DEFINITIONS

Disruptive Technologies: Emerging technologies that challenge the established players in an existing industry with potential displacement of its range of products and services through technological innovation.

Performance Measurement: The approach in collecting, analyzing and reporting a range of financial and non-financial indicators on past performance of an organization with implications for its future.

Technological Innovation: The process of new product or service development through integration of complementary technologies for adoption by the marketplace.

Technology Ventures: Emerging business entities during their early development and growth with exploitation of technologies and transforming such technologies into new products or services for rapid business growth and development.

Science and Technology Institution: Large established organizations that facilitate the development of science and technology through various research and development (R&D) as well as technological innovation activities.

The Skills of European ICT Specialists



Francesca Sgobbi

University of Brescia, Italy

INTRODUCTION

Specialists in Information and Communication Technologies (ICTs) are professionals “who have the ability to develop, operate and maintain ICT systems” and for whom “ICTs constitute the main part of their job” (OECD, 2004, p.219). ICT specialists play a crucial role in supporting the growth of the information society. However, some signals exist that in mature industrialized economies such as the European Union (EU) the supply of ICT specialists has been falling behind the demand and the evolution of their professional profile may not accompany the growing request for non-technical complementary skills (Eurostat, 2015a; Hüsing *et al.*, 2015).

After a brief survey of the international literature on skill-related issues that may affect the further development of ICT-based applications this article depicts the state-of-the-art of the professional skills supplied by ICT specialists in 11 EU countries based on data from the OECD Survey of Adult Skills.

The first part of the empirical analysis focuses on the skills profile of ICT personnel from EU countries and examines to what extent the higher skills displayed by ICT specialists depend on a different distribution of demographic characteristics and job characteristics compared to the rest of the workforce. The second part of the empirical analysis focuses on the relationship between skills and wages and tests whether employers recognize an occupation-specific wage premium to ICT specialists.

BACKGROUND

Pervasiveness characterizes ICT solutions as general purpose technologies, i.e., technologies whose innovative applications impact everyday household life besides spanning across different business sectors (Jovanovic and Rousseau, 2005). This pervasiveness rose hopes that ICT-based innovations would stimulate cross-industry virtuous circles between investments, innovation, productivity, consumption, and employment able to solve the “jobless growth dilemma” apparent in the USA and the EU in the 1990s (Selhofer, 2000). However, the complementarities between existing and new technologies (Davis and Wright, 1999) and the lack of digital skills among ICT providers and end users soon pointed out risks that skill shortage may obstacle the achievement of expected benefits (Selhofer, 2000).

After twenty years of debate on the occupational effects of ICTs the picture is still unclear. Spiezia *et al.* (2016) suggest that ICTs cause a drop in labor content per unit, hence reducing the demand for labor, but at the same time they raise labor productivity, hence increasing the convenience of labor compared to other productivity factors. Overall effects on labor demand are expected to disappear in the long run, due to a reallocation of labor from traditional sectors to innovative, ICT-intensive ones. Nevertheless, the current framework is still undergoing adjustment processes. Investments in ICTs raised the demand for labor in OECD countries between 1990 and 2007, but reduced it afterwards (Spiezia *et al.*,

2016). In addition, after 2007 the decline in labor demand has been accompanied by polarization between high-skilled and poorly-skilled jobs at the expense of middle-educated workers (Michaels *et al.*, 2014). Job polarization happens because, due to their programmable nature, ICTs tend to substitute labor in case of routine tasks, which prevail among middle-skilled workers. In contrast, knowledge-intensive non-routine tasks concentrate among high-skilled employees whereas labor-intensive non-routine tasks, often concerning the provision of personal services, prevail among low-skilled workers (Autor *et al.*, 2003).

If the diffusion of ICTs has so far deluded the expectations of jobs growth in the whole economy, focus on ICT employment provides a more positive picture. According to Hüsing *et al.* (2015), the ICT professional workforce in Europe in 2014 comprised 7.5 million workers, or 3.5% of the European workforce, with United Kingdom, Germany, France, Italy, Spain, Poland, and the Netherlands accounting for three quarters of total EU ICT professionals. OECD countries present a similar share of ICT specialists (3.6% of workforce in 2014, OECD, 2016).

ICT specialists were not affected by the recent economic crisis. Eurostat (2016) reports a 4% annual growth rate between 2006 and 2014. OECD (2015) reports stable employment levels within the ICT sector in the last decade (2004-2013), which results in growth in the number of ICT specialists, who comprise at least 3% of total employment in most OECD countries, at the expenses of less professionalized roles.

If non-ICT workers are increasingly required to master ICT generic skills besides the core skills of their professional domain, ICT-specialists face the challenge of adding non-technical ICT-complementary skills to their traditional expertise. The emphasis on “soft skills” signals a shift in the nature of ICT specialists’ work and impacts the design of educational curricula (Chillas *et al.*, 2015). Besides facing the challenge of adjusting their curricula European faculties of Computer Science, one of the main sources of ICT special-

ists, have seen a decline in the number of their students and graduates (Hüsing *et al.*, 2015).

The risk of lower inflows from tertiary education, together with some signals of hard-to-fill vacancies, has raised concerns about a possible shortage of ICT specialist skills. However, not all studies agree on this point. OECD (2016) reports stable or even decreasing vacancy rates since 2007 in OECD countries. In addition, 38% of EU enterprises which recruited or tried to recruit ICT specialists in 2014 reported difficulties in filling ICT vacancies (Eurostat, 2015a). However, ICT-specialist recruiters represent only 20% of EU firms. This means that only about 3% of EU enterprises report ICT-related hard-to-fill vacancies and this figure has not changed from 2012 to 2014 (OECD, 2016).

ICT SPECIALISTS IDENTIFICATION AND DATA

A quantitative analysis of the skills provided by European ICT specialists requires a clear identification of these professionals, usually based on either the industry they work in or their occupation. Given the pervasiveness of ICT jobs across industries and enterprise functions (Eurostat, 2015a; Hüsing *et al.*, 2015) an occupation-based approach seems more appropriate to identify ICT specialists and characterize their skills. Eurostat (2015b) details a list of occupations used to operationalize a statistical definition of ICT specialists. Relevant occupations are identified based on the 4-digit level of the 2008 release of the International Standard Classification of Occupations (ISCO-08).

This article resorts to the Eurostat criterion to identify EU ICT specialists based on the OECD Survey of Adult Skills (Figure 1). This survey was developed within the wider OECD Programme for the International Assessment of Adult Competencies (PIAAC Survey). The PIAAC Survey, run between August 2011 and March 2012, inquires

a representative sample of individuals aged 16-65 in 22 OECD countries (OECD, 2013).

The PIAAC Survey provides suitable data to analyze the skill profile of European ICT specialists, since information on education and learning is complemented by a large set of questions concerning experience in the labor market, working conditions, and personal and household background. The dataset also includes the outcomes of field tests for the assessment of individual proficiency in literacy, numeracy, and, for some countries, problem solving in a technology-rich environment. Unfortunately the Eurostat approach to identify ICT specialists requires information on 4-digit ISCO-08 occupations, which not all EU countries collect. For this reason the empirical analysis is restricted to 11 EU countries (Belgium, Czech Republic, Denmark, France, Germany, Italy, Netherlands, Poland, Slovakia, Spain, and United Kingdom). Nevertheless the selected sample accounts for 80.1% of EU population and for 85.5% of EU gross domestic product in 2014.

The distribution of ICT specialists in the examined EU countries as a percentage of total employment resulting from PIAAC data is quite similar to that reported by Eurostat (2016). In almost all countries the share of ICT specialists is above 3% of total employment, which corresponds to an estimate of over 5.3 million ICT specialists.

DO EUROPEAN ICT SPECIALISTS PROVIDE BETTER SKILLS?

Table 1 compares individual and demographic characteristics of ICT specialists and other workers in the examined countries, whereas Table 2 reports information on job characteristics. Both tables outline that ICT specialists comprise a very peculiar subgroup of the total workforce.

In line with the data reported by Eurostat (2016), males largely prevail among ICT specialists, who are on average younger than the rest of the workforce. Average educational qualifications of ICT specialists are always higher than those

achieved by other employees. ICT specialists differ from the remaining working population also in terms of personal background. On average, the parents of ICT specialists achieved higher qualification levels in all the examined countries. In addition, with the only exception of France the share of native speaking population is always higher among ICT specialists.

Three columns in Table 1 report the outcomes of the PIAAC Survey field tests of interviewees' literacy, numeracy, and (for a subset of countries) problem-solving along a 0-500 scale. Despite large cross-country variance, the scores achieved by ICT specialists in cognitive skills tests are always significantly higher compared to the rest of the workforce, especially in the case of numeracy skills. The last column in Table 1 reports a measure of non-technical skills surveyed by PIAAC (Soft skills). The normal variable Soft skills is calculated by means of a factor analysis on 12 variables that assess how often, in a 1-5 scale, individuals are asked to provide leadership, communication, or relational skills. As in the case of cognitive skills, ICT specialists declare more frequent provision of soft skills than other workers.

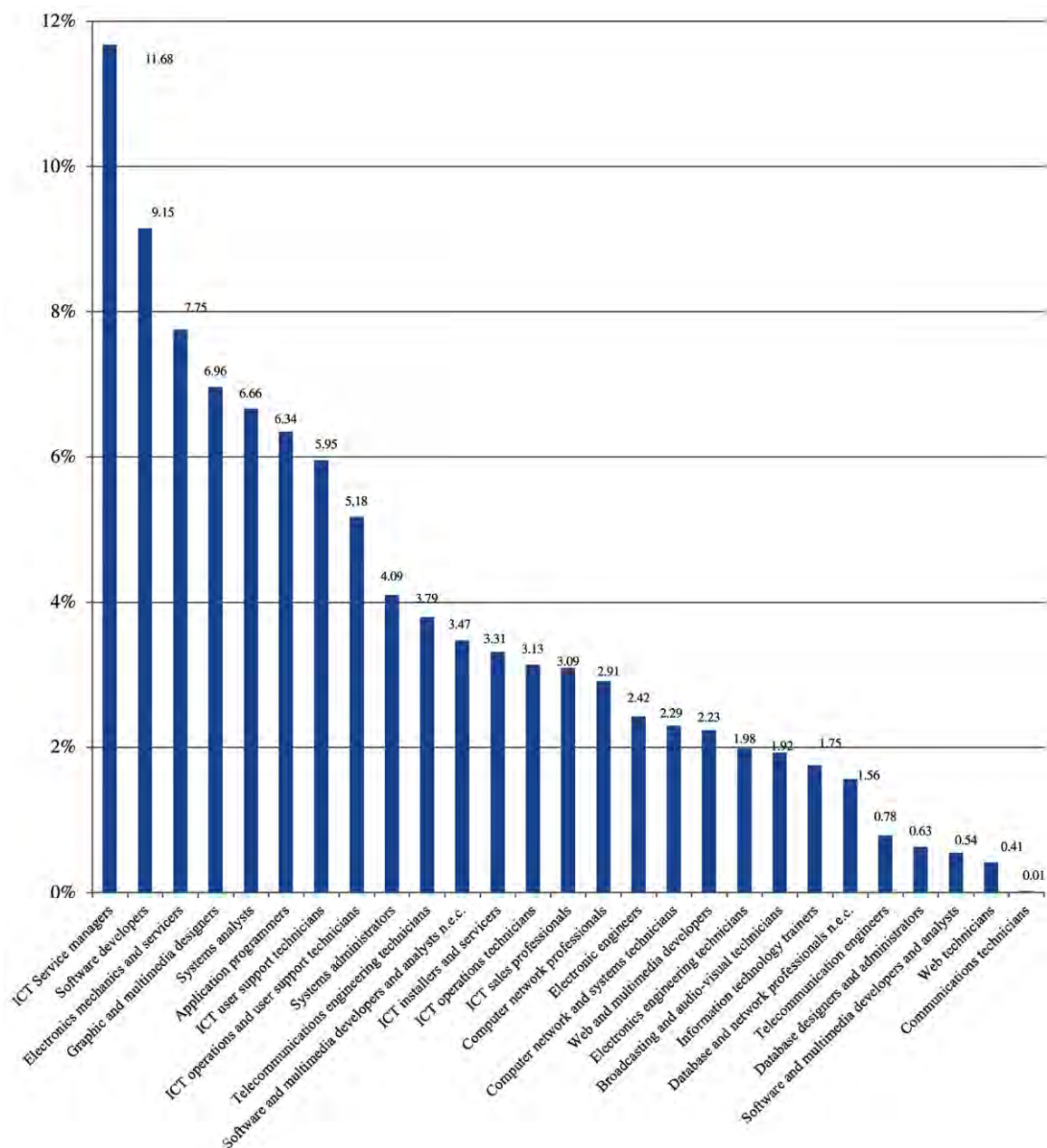
Table 2 shows that the demographic and individual favorable conditions enjoyed by ICT specialists correspond to better employment conditions. ICT professionals concentrate among skilled occupations. The more dynamic nature of the labor market for ICT specialists also in a period of economic crisis is confirmed by the higher number of jobs held in the five years before the PIAAC Survey. The share of ICT specialists and non-ICT specialists with a temporary contract is similar only for Belgium and Italy. In all other countries ICT specialists enjoy a significantly higher employment security, especially in the Czech Republic and the United Kingdom. The last column in Table 2 shows that also the frequency of part-timing is much lower among ICT specialists compared to other workers.

Data in Tables 1 and 2 confirm that ICT specialists significantly differ from the rest of the working population of the examined EU countries



Figure 1. EU ICT specialists by occupation

Source: Elaboration on PIAAC data



in terms of both individual and job characteristics. The higher proficiency in cognitive skills (literacy, numeracy, and problem solving) and the more intense provision of soft skills assessed by the PIAAC Survey in the case of ICT specialists may consequently follow from the concentration of favorable characteristics among ICT specialists—a

higher educational qualification, a more favorable family background, an occupation that provides more learning opportunities—rather than depending on these professionals’ core capabilities.

To assess whether ICT specialists provide higher cognitive skills after taking into account the impact of other employee-specific, job-spe-

Table 1. Individual characteristics: ICT specialists vs. other workers

Country	ICT Specialist	Women [%]	Age [Years]	Schooling [Years]	Parent with Tertiary Education [%]	Native Speaker [%]	Literacy [0-500]	Numeracy [0-500]	Pb. Solving [0-500]	Soft Skills
Belgium	Yes	20.2	39.2	14.4	38.6	90.0	310.4	320.1	312.9	0.325
	No	47.2	41.7	13.0	25.4	91.7	280.8	285.9	281.5	0.025
Czech Republic	Yes	5.6	38.3	15.5	33.2	93.4	307.2	316.8	311.1	0.318
	No	46.2	40.8	13.4	14.1	96.1	276.2	278.2	281.0	0.036
Denmark	Yes	21.4	39.3	14.0	41.3	88.3	302.2	315.8	309.0	0.596
	No	49.1	41.1	12.8	32.1	89.7	275.4	283.6	284.5	0.256
France	Yes	18.1	38.8	14.6	35.7	88.9	295.8	303.6	n.a.	0.279
	No	48.8	41.1	11.8	18.2	88.7	265.4	260.0	n.a.	-0.089
Germany	Yes	13.2	39.9	15.0	58.8	90.8	309.5	316.7	313.7	0.212
	No	47.2	41.6	13.8	35.1	87.4	273.4	277.7	282.9	-0.130
Italy	Yes	18.0	39.9	13.2	14.2	97.4	284.1	290.0	n.a.	0.388
	No	40.2	41.3	11.3	5.5	88.2	254.1	254.1	n.a.	-0.102
Netherlands	Yes	10.5	38.8	14.6	32.5	90.3	320.3	317.2	318.4	0.398
	No	48.0	40.1	13.5	25.6	87.2	288.3	285.4	289.4	0.028
Poland	Yes	17.1	35.2	14.8	36.5	99.4	300.9	294.5	316.5	0.205
	No	45.2	39.5	13.3	14.5	98.0	271.0	265.1	272.3	-0.017
Slovakia	Yes	13.4	35.2	15.5	35.1	97.5	299.5	308.7	300.2	0.118
	No	45.8	40.7	13.7	12.7	95.9	278.7	284.2	281.4	-0.176
Spain	Yes	25.2	37.3	14.6	31.3	97.1	294.7	291.8	n.a.	0.205
	No	46.1	41.3	12.1	13.1	94.2	259.0	255.2	n.a.	-0.051
United Kingdom	Yes	12.4	39.5	13.9	34.1	86.1	308.8	308.6	314.7	0.570
	No	47.9	40.2	13.2	26.6	87.7	277.9	268.6	282.5	0.339
All 11 Countries	Yes	15.5	38.9	14.4	37.7	91.6	303.1	306.5	314.1	0.337
	No	46.3	40.9	12.8	21.2	90.0	269.6	267.7	282.1	-0.013

cific, or country-specific features Table 3 reports the results of an Ordinary Least Square (OLS) regression where the dependent variable is the numeracy score achieved in the PIAAC Survey. Focus on numeracy skills is justified by the relative importance of quantitative techniques for ICT specialists and by the high correlation (never lower than 0.710) existing among the three measures of cognitive skills provided by PIAAC. The last two lines of column “Numeracy proficiency” in Table 1 show that, on average, ICT specialists scored 306.5 points in the PIAAC numeracy test,

whereas non-ICT specialists scored 267.7 points out of 500, for a differential of 38.8 points. The estimate in Table 3 shows that 24.9 points of this differential depend on a disproportioned concentration of characteristics that favor proficiency in numeracy among ICT specialists. However, the remaining 13.9 points (significant and positive coefficient of variable ICT Specialist in Table 3) represent the better performance that ICT specialist on average provided compared to individuals endowed with similar characteristics. Compared to the rest of the workforce EU ICT specialists

Table 2. Job characteristics – ICT specialists vs. other workers

Country	ICT Specialist	Skilled Occupation [%]	Semiskilled White-Collar Occupation [%]	Semiskilled Blue-Collar Occupation [%]	Elementary Occupation [%]	No. Jobs in the Last 5 Years	Temporary Contract [%]	Part-Time Contract [%]
Belgium	Yes	84.6	0.9	10.9	1.7	1.7	8.0	5.4
	No	45.7	25.7	17.8	8.1	1.6	8.7	17.4
Czech Republic	Yes	90.3	0.2	5.4	1.7	2.1	9.5	5.5
	No	33.6	22.8	33.8	5.7	1.8	19.3	6.0
Denmark	Yes	87.9	0.0	4.0	2.0	1.9	7.7	9.2
	No	41.3	20.9	16.6	6.6	2.1	17.1	19.0
France	Yes	91.1	0.0	4.0	2.0	1.7	14.1	4.8
	No	39.2	25.5	21.6	10.5	1.8	17.1	14.9
Germany	Yes	74.5	0.3	15.3	0.0	1.7	17.1	9.1
	No	33.9	26.5	21.4	5.8	1.7	22.8	23.6
Italy	Yes	88.8	0.0	6.4	0.8	1.5	23.3	8.4
	No	29.9	27.6	28.0	9.8	1.5	22.3	17.0
Netherlands	Yes	91.9	0.0	2.6	0.0	2.0	17.2	10.0
	No	48.0	24.4	11.0	5.9	2.0	26.9	36.7
Poland	Yes	74.6	0.0	13.4	3.7	2.4	30.9	11.0
	No	36.1	20.0	30.1	5.8	1.8	34.5	11.0
Slovakia	Yes	82.0	0.0	13.6	2.8	1.6	15.4	3.5
	No	40.9	21.1	27.3	6.3	1.6	20.5	7.3
Spain	Yes	82.2	0.0	6.8	0.0	2.2	19.4	20.9
	No	31.9	32.2	18.6	11.3	1.7	27.7	14.1
United Kingdom	Yes	79.3	0.0	15.3	0.0	2.0	10.8	8.4
	No	36.2	33.6	16.0	9.2	1.8	20.6	26.1
All 11 Countries	Yes	82.3	0.1	10.3	0.8	1.9	16.4	9.3
	No	35.6	27.1	21.9	8.2	1.7	22.9	19.3

provide a statistically significantly higher level of cognitive skills as proxied by numeracy skills.

Similarly to Table 3, Table 4 reports the outcomes of an OLS regression that tests the determinants of the frequency of soft skills provision. Also in this case the coefficient of variable ICT Specialist is statistically significant. However, its negative sign shows that the higher soft skills declared by ICT specialists (last column in Table 1) are totally due to the heterogeneous distribution of soft skills drivers across the workforce. When compared with individuals with the same

age, gender, education, family background, occupation, employment conditions, and country of residence ICT specialists reveal lower frequency of soft skills provision.

DO EUROPEAN ICT SPECIALISTS EARN HIGHER REWARDS?

The higher educational and professional human capital of ICT specialists is associated with higher hourly and monthly average gross earnings (Table

Table 3. The determinants of numeracy proficiency

	Coefficient	Robust Std. Err.	
Constant	156.117	5.209	***
ICT Specialist	13.886	1.692	***
Age	1.631	0.229	***
Squared Age	-0.024	0.003	***
Female	-13.425	0.762	***
Native speaker	26.329	1.421	***
Years of education	5.358	0.147	***
Parent with secondary education ^(a)	7.429	0.900	***
Parent with tertiary education ^(a)	15.293	1.069	***
Skilled occupation ^(b)	22.335	1.552	***
Semiskilled white-collar occupation ^(b)	12.924	1.507	***
Semiskilled blue-collar occupation ^(b)	5.425	1.614	***
Temporary contract	-5.353	0.935	***
Part-time contract	-2.456	1.035	**
Czech Republic ^(c)	-10.839	1.405	***
Denmark ^(c)	-2.300	1.024	***
France ^(c)	-15.656	1.057	***
Germany ^(c)	-13.174	1.128	***
Italy ^(c)	-17.638	1.405	***
Netherlands ^(c)	-2.650	1.095	**
Poland ^(c)	-24.625	1.288	***
Slovakia ^(c)	-7.312	1.163	***
Spain ^(c)	-22.083	1.175	***
United Kingdom ^(c)	-11.926	1.272	***
R-Squared	0.357		

Dependent variable: Numeracy proficiency; OLS regressions; 32,902 weighted observations

* Statistically significant at the.10 level; ** at the.05 level; *** at the.01 level

^(a)Reference category: Parents with primary education or below

^(b)Reference category: Elementary occupation

^(c)Reference category: Belgium

Table 4. The determinants of soft skills provision

	Coefficient	Robust Std. Err.	
Constant	-2.207	0.093	***
ICT Specialist	-0.214	0.031	***
Age	0.034	0.004	***
Squared Age	0.000	0.000	***
Female	-0.178	0.014	***
Native speaker	0.285	0.024	***
Years of education	0.058	0.003	***
Parent with secondary education ^(a)	0.073	0.017	***
Parent with tertiary education ^(a)	0.138	0.020	***
Skilled occupation ^(b)	1.040	0.028	***
Semiskilled white-collar occupation ^(b)	0.694	0.026	***
Semiskilled blue-collar occupation ^(b)	0.176	0.028	***
Temporary contract	-0.202	0.017	***
Part-time contract	-0.285	0.019	***
Czech Republic ^(c)	0.043	0.029	
Denmark ^(c)	0.307	0.020	***
France ^(c)	0.028	0.022	***
Germany ^(c)	-0.106	0.022	***
Italy ^(c)	0.118	0.026	***
Netherlands ^(c)	0.071	0.021	***
Poland ^(c)	-0.029	0.025	
Slovakia ^(c)	-0.228	0.024	***
Spain ^(c)	0.045	0.024	*
United Kingdom ^(c)	0.439	0.024	***
R-Squared	0.370		

Dependent variable: Soft Skills; OLS regressions; 32,606 weighted observations

* Statistically significant at the.10 level; ** at the.05 level; *** at the.01 level

^(a)Reference category: Parents with primary education or below

^(b)Reference category: Elementary occupation

^(c)Reference category: Belgium

5). To ensure comparability of data across countries earnings are expressed in US dollars and corrected to account for purchasing power parity (PPP). In addition, they include bonuses for wage and salary earners and self-employed.

As in the case of ICT specialists' skills discussed above the higher earnings observed compared to other employees may depend on two different phenomena. Higher earnings could follow from an over-representation of characteristics

Table 5. Earnings: ICT specialists vs. other workers

Country	ICT Specialist	Hourly Earnings [PPP Corrected \$US]	Monthly Earnings [PPP Corrected \$US]
Belgium	Yes	26.7	4,700
	No	23.4	3,639
Czech Republic	Yes	12.6	2,227
	No	9.2	1,620
Denmark	Yes	29.6	5,015
	No	24.0	3,780
France	Yes	20.7	3,465
	No	15.8	2,493
Germany	Yes	26.6	4,551
	No	20.2	3,152
Italy	Yes	20.1	3,659
	No	16.5	2,563
Netherlands	Yes	26.2	4,241
	No	22.9	3,224
Poland	Yes	12.6	1,914
	No	10.3	1,688
Slovakia	Yes	14.3	2,549
	No	10.4	1,627
Spain	Yes	21.1	3,061
	No	15.4	2,368
United Kingdom	Yes	28.6	4,819
	No	18.8	3,071
All 11 Countries	Yes	23.7	3,972
	No	17.3	2,707

usually associated with high wages among ICT specialists, such as higher educational qualifications or superior hierarchical positions. However, the higher earnings in Table 5 may also reflect an occupation-specific premium recognized to the unique and scarce capabilities provided by ICT specialists.

Information provided by the PIAAC Survey allows to disentangle the source of the wage advantage of ICT specialists by estimating OLS regressions of a wage equation (Mincer, 1974),

which explains the natural logarithm of wage with the drivers of individual human capital. When statistically significant, the coefficients estimated for independent variables appraise the percent increase of individual wage caused by a unit increase of the corresponding independent variables (McGuinness, 2006). The inclusion of a binary variable that takes value 1 for ICT specialists and 0 for other workers allows assessing the existence of a significant wage premium for ICT professionals.

Table 5 reports the coefficients estimated for two wage equations by means of OLS regressions. In both models the dependent variable is the natural logarithm of hourly gross earnings. Model 1 provides a baseline specification that controls for individual education and educational mismatch, individual experience, and individual-specific, job-specific, and country-specific effects. The latter regressors, always jointly significant, are not displayed in Table 6. Model 2 adds to the independent variables of Model 1 three interactive factors to explore whether years of education, gender, and proficiency in numeracy skills are rewarded differently for ICT specialists compared to the rest of the workforce.

In general terms, Model 1 in Table 6 confirms the findings of previous studies of wage determinants (McGuinness, 2006). However, for the purposes of this contribution the key variable is ICT Specialist. The positive and significant coefficient of this variable shows that, *ceteris paribus*, ICT specialists enjoy 3.4% higher hourly earnings compared to other workers. The estimated average premium is lower than the figures reported in Table 4, according to which the hourly wage premium of ICT specialists averages 37% for all countries (23,7 US\$ per hour for ICT specialists compared to 17,3US\$ per hour for other workers). This outcome suggests that, on average, the overrepresentation of wage-enhancing characteristics among ICT specialists explains 33.6 points of the observed 37% differential. Yet, the remaining 3.4% is the reward recognized to unique and scarce capabilities provided by ICT specialists.

Table 6. The determinants of monthly earnings

	Model 1			Model 2		
	Coefficient	Robust Std. Err.		Coefficient	Robust Std. Err.	
Constant	1.723	0.039	***	1.731	0.040	***
ICT Specialist	0.034	0.017	*	-0.320	0.135	**
Years of education	0.037	0.002	***	0.037	0.002	***
ICT specialist*Years of education				0.010	0.008	
Overeducated	-0.084	0.010	***	-0.084	0.010	***
Undereducated	0.101	0.011	***	0.102	0.011	***
Tenure	0.037	0.002	***	0.037	0.002	***
Squared tenure	-0.001	0.000	***	-0.001	0.000	***
Female	-0.160	0.009	***	-0.163	0.009	***
ICT specialist*Female				0.096	0.044	**
Native speaker	0.004	0.016		0.005	0.016	
Parent with secondary education ^(a)	0.007	0.010		0.007	0.010	
Parent with tertiary education ^(a)	0.027	0.012	**	0.027	0.012	**
Numeracy	0.001	0.000	***	0.001	0.000	***
ICT specialist*Numeracy				0.001	0.000	*
Soft skills	0.062	0.005	***	0.062	0.005	***
Skilled occupation ^(b)	0.210	0.021	***	0.211	0.021	***
Semiskilled white-collar occupation ^(b)	-0.013	0.021		-0.013	0.021	
Semiskilled blue-collar occupation ^(b)	0.071	0.021	***	0.071	0.021	***
Temporary contract	-0.164	0.012	***	-0.163	0.012	***
Part-time contract	0.110	0.014	***	0.111	0.014	***
Public sector	0.054	0.008	***	0.055	0.008	***
R-Squared	0.479			0.480		

Dependent variable: logarithm of gross hourly earnings; OLS regressions; 30,363 weighted observations; the regressions include 10 country-specific binary variables

* Statistically significant at the .10 level; ** at the .05 level; *** at the .01 level

^(a)Reference category: Parents with primary education or below

^(b)Reference category: Elementary occupation

The exam of the coefficients estimated for Model 2 in Table 6 confirms that employers attach a higher value to some individual characteristics of ICT specialists compared to other workers. This is not the case of years of education (non-significant coefficient of interacted variable ICT specialist*Years of education), whereas the wage penalty of female employees more than halves for female ICT specialists as a result of the sum of the negative and significant coefficient of variable

Female and the positive and significant coefficient of the interacted term ICT specialist*Female. If a 10-point increase in the score of the PIAAC numeracy test is associated with a 1% increase of gross hourly wage for all workers, this advantage doubles in the case of ICT specialists, thanks to the composition of the coefficients of variables Numeracy and ICT specialist*Numeracy. In contrast with cognitive skills, soft skills do not impact the earnings of ICT specialists differently

from other workers. When adding an interactive term ICT Specialist*Soft skills to Model 2 the estimated coefficient is not statistically significant.

FUTURE RESEARCH DIRECTIONS

Up-to-date information on the evolution of ICT specialists' skill profiles is needed to design and support educational, vocational, and training programs able to comply with changes in demand requirements. Three main research lines are crucial to improve the quality of available information on ICT specialists and their skills. First, agreement is needed to operationalize a statistical definition of ICT specialists, today still identified based on the only partially overlapped criteria of membership in specific industries or specific occupations. Second, longitudinal datasets focused on ICT specialists are required to assess skills evolution and to avoid estimate biases due to unobserved individual heterogeneity. Third, given the potential of ICTs in supporting economical catch up the collection of statistical information on ICT specialists should extend to developing countries.

CONCLUSION

The European Union frames ICT skills as a crucial resource to prevent the loss of key ICT jobs to other regions of the world (Eurostat, 2015a). Unfortunately, the performance of the EU ICT sector has been lagging behind its main competitors in most recent years (OECD, 2015). To catch up with competitors EU countries need to value and develop the human capital of their ICT specialists. The proposed empirical analyses point out at least three possible paths to pursue this goal.

1. In all examined countries ICT specialists display higher human capital and enjoy better employment conditions. In addition,

the superior cognitive skills of ICT specialists command better earnings. Employment policies should leverage on this evidence to attract high-potential labor inflows.

2. Also after accounting for the heterogeneous distribution of employee-specific and job-specific characteristics ICT specialists confirm to provide higher levels of cognitive skills. In contrast, they provide lower soft skills compared to otherwise similar individuals. Educational curricula and training programs need adjustments to avoid the risk of skill shortage in ICT-complementary skills.
3. Female workers and non-native workers may improve their labor market outcomes by pursuing a career as ICT specialist. Employment policies may focus on these segments of the labor market to enlarge the number of candidates to ICT jobs.

REFERENCES

- Autor, D. H., Richard, F. L., & Murnane, J. R. (2003). The skill content of recent technological change: An empirical exploration. *The Quarterly Journal of Economics*, 118(November), 1279–1333. doi:10.1162/003355303322552801
- Chillas, S., Marks, A., & Galloway, L. (2015). Learning to labour: An evaluation of internships and employability in the ICT sector. *New Technology, Work and Employment*, 30(1), 1–15. doi:10.1111/ntwe.12041
- David, P. A., & Wright, G. (2003). General purpose technologies and surges in productivity: Historical reflections on the future of the ICT revolution. In P. A. David & M. Thomas (Eds.), *Economic Future in Historical Perspective* (pp. 135–166). Oxford, UK: Oxford University Press for the British Academy.



Eurostat. (2015a). *ICT specialists - statistics on hard-to-fill vacancies in enterprises*. Retrieved Jun 30, 2016, from http://ec.europa.eu/eurostat/statistics-explained/index.php/ICT_specialists_-_statistics_on_hard-to-fill_vacancies_in_enterprises

Eurostat. (2015b). *ICT specialists in the labour force*. Retrieved November 2, 2016, from http://ec.europa.eu/eurostat/cache/metadata/en/isoc_skslf_esms.htm

Eurostat. (2016). *ICT specialists in employment*. Retrieved November 2, 2016, from http://ec.europa.eu/eurostat/statistics-explained/index.php/ICT_specialists_in_employment

Hüsing T., Korte W. B., & Dashja E. (2015). *e-Skills in Europe. Trends and Forecasts for the European ICT Professional and Digital Leadership Labour Markets (2015-2020)*. Bonn: empirica.

Jovanovic, B., & Rousseau, P. L. (2005). General Purpose Technologies. In Ph. Aghion & S. Durlauf (Eds.), *Handbook of Economic Growth* (pp. 1181–1224). Amsterdam: Elsevier.

Mas, M., & de Guevara Radoselovics, J. F. (2015). The 2015 Predict Report: An analysis of ICT R&D in the EU and beyond. In I. K. Rohman, G. De Prato, & P. Desruelle (Eds.), Report EUR 27510 EN. Luxembourg: Publications Office of the European Union.

McGuinness, S. (2006). Overeducation in the labour market. *Journal of Economic Surveys*, 20(3), 387–418. doi:10.1111/j.0950-0804.2006.00284.x

Michaels, G., Natraj, A., & Van Reenen, J. (2014). Has ICT polarized skill demand? Evidence from eleven countries over twenty-five years. *The Review of Economics and Statistics*, 96(1), 60–77. doi:10.1162/REST_a_00366

Mincer, J. (1974). *Schooling, experience and earnings*. New York: Columbia University Press.

OECD. (2004). *OECD Information Technology Outlook*. Paris: OECD.

OECD. (2013). *Technical Report of the Survey of Adult Skills (PIAAC)*. Paris: OECD.

OECD. (2015). *Digital Economy Outlook 2015*. Paris: OECD.

OECD. (2016). *Skills for a digital world. DSTI/ICCP/IIS(2015)10/FINAL*. Paris: OECD.

Selhofer, H. (2000). Skills shortage vs. job creation: A review of empirical evidence on the Issue of ICTs and employment. *Vierteljahrshefte zur Wirtschaftsforschung*, 69(4), 510–526. doi:10.3790/vjh.69.4.510

Spiezia, V., Polder, M., & Presidente, G. (2016). *ICTs and jobs: Complements or substitutes? The effects of ICT investment on labour demand by skills and by industry in selected OECD countries. DSTI/ICCP/IIS(2016)1/FINAL*. Paris: OECD.

ADDITIONAL READING

European Commission. (2016). *New Skills Agenda for Europe*. Retrieved Jun 30, 2016, from <http://ec.europa.eu/social/main.jsp?catId=1223&langId=en>

Llorens-Garcia, A., Llinas-Audet, X., & Sabate, F. (2009). Professional and interpersonal skills for ICT specialists. *IT Professional*, 11(6), 23–30. doi:10.1109/MITP.2009.132

OECD. (2013). *OECD Skills Outlook 2013. First Results from the Survey of Adult Skills*. Paris: OECD.

OECD. (2015). *Proposal for a EUROSTAT-OECD definition of ICT specialists. DSTI/ICCP/IIS(2015)7/REV1*. Paris: OECD.

Sabadash, A. (2013). *ICT Employment Statistics in Europe: Measuring Methodology. EUR 26126 EN*. Luxembourg: Publications Office of the European Union.

KEY TERMS AND DEFINITIONS

Cognitive Skills: Intrinsic individual characteristics that can be measured by achievement tests and taught in formal educational or training programs.

ICTs: Information and Communication Technologies. ICTs comprise methods and systems to transfer, archive, retrieve, and elaborate information, digital information included.

ICT Specialist: A professional who has the ability to develop, operate, and maintain ICT systems and for whom ICTs constitute the main part of the job.

Job Polarization: Concentration of labor demand towards jobs in the high and the low tails of skills requirements distribution.

PIAAC Survey: International assessment of the competences of adult population run in 22 OECD countries in 2012 within the wider Programme for the International Assessment of Adult Competencies (PIAAC).

Skill Shortage: Difficulty to fill open vacancies due to lack of candidates endowed with relevant skills.

Skills for the Digital Economy: Skills to develop, use, and manage ICT-based products and processes in business and household life.

Soft Skills: Intrinsic individual characteristics hardly measurable by achievement tests. Soft skills are broadly applicable across different job titles, occupations, and industries.

A Trust Case-Based Model Applied to Agents Collaboration



Felipe Boff

Lutheran University of Brazil (ULBRA), Brazil

Fabiana Lorenzi

Lutheran University of Brazil (ULBRA), Brazil

INTRODUCTION

The interaction among people as well as the interaction among systems can, at times, be seen from similar perspectives. People and systems construct their relationship network supported by information. People have the ability to retain information about things and, based on that, pursue their requirements. Computational systems can also function in a similar manner; however, this is highly dependent upon how they were created. As time passes, a person may lose contact with someone or develop stronger ties if they see this relationship as favorable. Systems and agents can also cease to relate to other systems or start new relationships, but, for this, actions external to the agent or system may be necessary.

One of the feelings that can bring comfort to interpersonal relationships is trust. Based on this feeling, people allow other people they trust to participate more frequently or intensely in their lives. The relationship between two systems is treated here as collaboration, where each agent, according to its computational purpose, executes operations requested by other agents. Computational systems, however, can have difficulties in evaluating their collaboration with other systems. Quantifying collaboration as appropriate or detrimental can require complex algorithms and several exceptions to a given rule. The proposal presented here suggests the creation of a method which evaluates the relationship between two agents to generate a rating, called *general trust rating*.

All collaborations between agents are unique and generate equally unique results. These results can be seen as a history of the collaboration between agents. This history is rich in information which can be used to underpin the generation of a trust rating. For this, the information has to be stored. The more information that is stored, the more complex the process of extraction has to be. Due to this, a Case-Based Reasoning approach is used to store this information. The collaboration between agents generates a new case in the case database, which comprises the history of this collaboration. Whether this is beneficial or not for the agent can be concluded by reviewing the whole collaboration history. By using the method developed, the correlation of this history can be quantified, generating a trust rating.

This paper presents a method which was developed to quantify the relationship between agents through the generation of a rating. As in human relations, agents may have a high degree of trust in other agents when the relationship is analyzed from a wider perspective. Nevertheless, the agent is likely to have a different trust rating for specific activities due to the environment it is exposed to or to its purpose. In order to allow this differentiation, the proposed method uses case-based reasoning for each collaboration request to find similar previous collaboration. It selects from this base of previous collaboration cases those that were similar to the current demand, generating a situational trust rating for it. The weighting of the situational with the general trust rating allows

evaluating the collaboration considering a particular situation as well as the history of collaboration between the agents.

In the following sections first we present some related work and how the important points are set in the proposed method. Then we illustrate how the method was implemented, where functionalities are separated into modules so as to provide segregation and isolation of these functionalities: we present the tests that were run and the obtained results that corroborate the efficiency of the method. Finally, we present our conclusions and some future work we have planned.

BACKGROUND

Related Work

The conceptual definition of an autonomous agent is understood as being something that perceives its environment through sensors (physical or virtual) and acts upon this environment through effectors (Russel & Norvig, 2003). These effectors can resolve problems or determine actions, acting on the environment to carry out sets of goals or tasks for which they were designed.

A multi-agent system (MAS) consists of a group of agents hosted in an environment where autonomous agents collaborate with other agents for solving problems (Challenger et al., 2016). Agents can be competitive - when an agent attempts to maximize its performance measures over those of other agents - or cooperative, when the agents function in such a manner that they complete their goals and tasks through the knowledge of other agents.

MAS are exploited in several contexts nowadays to provide e-services, as e-Commerce, e-Learning and so on. Those systems are distributed across the internet and are hosted by different companies or structures. Thus, they have a decentralized control and are subject to constant changes during their operational life. The relationship between these systems requires the existence of components that can interact flexibly in dynamic and uncertain

environments, with the ability to meet their goals or complete their tasks. (Buccafurri et al., 2014).

In both forms of relationship between agents - cooperative and competitive - it is necessary for the agent to have knowledge of the behavior of those agents with which it has a relationship. This necessity becomes of greater importance when the relationship between the agents occurs in open MAS, given that the differences in structures or controls can impact directly on the task results.

Collaboration Between Agents

The collaboration between agents can be defined as the division of a global task, decomposed into a number of agents' subproblems that can be solved efficiently based on the collaboration of agents (Hsieh & Lin, 2014). This definition makes it possible to understand that collaboration between agents has direct impact on the final result of a task.

Even when two agents have exactly the same purpose and receive exactly the same demand, the outcomes can be different. In addition, since they may have different knowledge, they can be hosted in different structures or have different relationships. It is important that, due to this, collaboration is made between agents that possess the best knowledge, the best performance and also the lowest error rate.

The possibility of interaction between ever wider and more disperse systems has resulted in the creation of numerous models of interaction, such as collaboration (Pynadath & Tambe, 2002) and negotiation (Kraus, 2001). However, the application of these models to open MAS has presented a major challenge; each agent can interact with agents of the same or of other categories. In the literature, various interaction paradigms have been defined. Agents can be competitive if they have conflicting objectives and cooperative otherwise (Barbati, Bruno & Genovese, 2012).

All of the interaction models created presuppose that the relationship between systems will result in a common objective and in the same level of information. Nevertheless, in practice, it is almost impossible to obtain an adequate level

of information about the environment or about the interactive properties of the partners, the possible strategies and priorities (Russel & Norvig, 2003).

Trust as a Determinant for Collaboration

Given the context that has been presented, the agents are exposed to elevated levels of uncertainty when taking decisions. Thus, they need to have trust in other agents as a method of reducing the uncertainty associated with interactions in open MASs.

Among the different definitions of trust, the one that follows reflects the reason why it is so necessary for agents to have knowledge of the environment to which they are exposed: "Trust is a belief an agent has that the other party will do what it says it will (being honest and reliable) or reciprocate (being reciprocate for the common good of both), given an opportunity to defect to get higher payoffs" (Ramchurn, Huynh, & Jennings, 2004). The utilization of trust in MAS environments is very diversified in approach (Rosaci, 2012) (Khosravifar et al., 2009) and in the form of definition (Sabater & Sierra, 2005). Even in approaches that support their definition by the same principle, trust has to be quantified as a translation of descriptive variables. By this the theory stated by Marsh (1994) a solid combination of well-defined patterns and suitable calculation approach.

According to Marsh (1994), trust can be separated into three different aspects: *basic*, *general*, and *situational*. Basic trust is a reflection of all the past experience of an agent in any situation - that is, without differentiating between these experiences among interlocutors. General trust represents the trust of an agent in a specific interlocutor, considering all the interactions between these two. Situational trust reflects the trust of an agent in an interlocutor in a specific situation - in this evaluation, only those interactions that are similar to the situation for which the trust rating is desired are considered. These definitions will be detailed later in this paper.

In addition to the separation in three aspects, Marsh chose to treat trust as a numeric interval, in the range $[-1, +1]$. This range represents the possible trust that an agent has in another considering the definitions given in the preceding paragraph. Each of the three different aspects of trust uses the same range.

Two points have to be emphasized with respect to the interval defined by Marsh: 1) distrust and no trust are different states and should be treated differently, in which the value of 0 (zero) in the trust interval does not mean distrust, but that there has been no trust established in the first place. This rating could be the initial value in a relationship; 2) while a rating of -1 reflects total distrust in an interlocutor, the same situation is not possible for a rating of +1, as a value of +1 represents absolute trust or "blind trust", which means this agent is trusted regardless of the situation - something that does not refer to the concept of trust, given that this requires the evaluation of the situation as a presupposition for collaboration. To avoid these problems of interpretation of the trust ratings Marsh created a stratification of trust - somewhat similar to fuzzy logic for trust -, which is presented in table 1.

Table 1. Stratification of trust values (Marsh, 1994)

Value Range	Label
+1	Blind Trust
> 0.9	Very High Trust
0.75 to 0.9	High Trust
0.5 to 0.75	High Medium Trust
0.25 to 0.5	Low Medium Trust
0 to 0.25	Low Trust
-0.25 to 0	Low Distrust
-0.5 to -0.25	Low Medium Distrust
-0.75 to -0.5	High Medium Distrust
-0.9 to -0.7	High Distrust
< -0.9	Very High Distrust
-1	Complete Distrust

These concepts require the agent to be capable of weighting indicators - be they quantitative or qualitative - in a consolidated manner, so as to calculate a trust rating for each agent for each of the basic, general and situational aspects. The ability to consider these indicators is to be found through trust models.

On the basis of calculated trust rating for other agents, an agent can choose to interact with those that have a higher trust rating. On the other hand, a lower trust rating allows an agent to avoid interactions with agents that have these lower ratings. The increase in interaction due to higher trust ratings or the reduction of interaction where the trust rating is low tends to generate a result that better meets the demand.

The Calculation of Trust Rating

The trust model proposed by Marsh (1994) considers only direct interactions between agents. It means that there is no increase or reduction in the trust rating without some collaboration. As previously discussed, the model is divided into three different types of trust: basic, general and situational, with each type having a different formula for evaluation and calculation.

Case-Based Reasoning for the Calculation of the Trust Rating

The use of CBR is directly related to the solution of the problem. CBR uses past experiences for trust have showed good results (Floyd, Drinkwater & Aha, 2014) (Abedinzadeh & Sadaoui, 2014). Problems are solved through the reuse and, at times, adaptation of known solutions to similar problems. The review phase is of great importance to the final results of systems that use this methodology. This phase is identified as essential to the process of human learning: the confirmation of a plausible solution to a problem (Stefan & Weiß, 1991).

This paradigm of human review during the CBR cycle is very important for the system to

learn. However, it requires external interference in the system for the validation of its results.

When the use of CBR systems is evaluated for the measurement of the trust ratings of agents, the first paradigm to be overcome is identified: the need for human interaction in the evaluation of results. In open MAS there are many agents and proportionally many more interactions between them. The need for human evaluation of the trust indicator generates a bottleneck, reducing considerably the effectiveness of a system that had been presumed to be automatic. The existence of previous cases of interaction between two agents can serve as a means of comparison to guide future requirements. It is also possible that past cases can be re-evaluated, allowing their outcome to have a different effect on the result. For example, in the event that a given agent increases its response time on some occasions, this situation will certainly impact on the final result of the required task for a calling agent which should result in a decrease in the trust rating of these agents. However, conventional methods do not consider the complete history of the relationship and do not permit a complete view of the collaboration of these agents. Through the review process of CBR it is possible to see isolated incidents as connected incidents because of their similarities, allowing the trust rating to reflect this situation.

Reuse of Cases Not Only for the Measurement of the Trust Rating

According to Bedi & Vashisth (2014) computing trust is a problem of reasoning under uncertainty, requiring the prediction and anticipation by an agent (the evaluator) of the future behavior of another agent (the target). With the use of CBR it is possible that all interactions be treated as cases and, thus, be stored in the case base. Through the reuse of cases it is possible to calculate the situational trust rating considering only those cases that are similar to the current collaboration, thus foreseeing to predict the agent's behavior.

The approach suggested by Marsh (1994) combined with the CBR approach, allows that the general trust rating of an agent to be high. However, given a particular interaction request, the situational trust rating can be calculated according to the results of the latest interactions in this context and be divergent in relation to the general rating. As an example, a sales agent has a high trust rating for a payment agent, which, in turn, has methods for consulting the scheduled payments. However, interactions directed to this end generate worse results for other agents. Given this situation, despite the global trust rating being high, demands for scheduled payments could be sent to other agents that have better ratings for the case in question.

Method for Calculation of the Trust Rating

The relationships between agents present major challenges that demand a large degree of adaptability of the agents and understanding of the environment to which they are exposed. The method for the calculation of the trust rating proposed in this paper addresses one form of evaluation of the relationship between agents in a multi-agent system. The goal of the proposed method is to provide the agents - in a fast and efficient manner - with a list of trust ratings for the agents available to carry out a requirement. The differential of this method is the evaluation of the general trust rating of an agent in accordance with the cases that represent the relationship with another agent, here called the situational trust rating. Therefore, through a correlation between general and situational trust ratings is achieved a new trust rating based on the overall relationship and similar interactions among agents. This method also provides a differential with respect to the evaluation of relational scenarios between the agents. The collaboration will entitle the evaluation of trust from one agent to others: the *caller* agent (the one who is asking for the collaboration), and the agents that can execute the demand from the

caller. Given that three trust ratings are created: the basic, considering the trust rating of an agent (the caller) to the system; a general, considering the trust of the caller agent in another; and another situational which reflects the trust that *caller* agent has for another in specific situations similar to the desired collaboration.

The general trust rating - which covers all of the interactions between agents - and the situational rating will be considered in the evaluation of the trust relationship of one agent with another. Given this, the method developed allows not only the general relationship between the agents to be considered at all times, but also the calculation to weight the trust rating obtained from similar situations against the required collaboration, inhibiting those cases in which a high general trust rating would prevail over the relationship between the agents. Through the proposed calculation, the system searches for the best collaboration for the situation presented rather than the best collaboration with the best agent.

The collaboration between agents is treated as a case possessing information that permits it to be classified and differentiated with respect to other collaboration demands between the two agents. Figure 1 shows that after each collaboration request, the method will use the information about this collaboration to recalculate the global trust rating between the agents and this update will be done based on the information from the current request weighted against previous cases of collaboration. It means that past cases are relevant to the calculation. This method allows the history in the trust rating to be considered, ensuring that collaboration momentarily rated positive or negative does not distort the trust rating.

The great differential of the proposed method addresses the use of similar cases to the current collaboration request for the calculation of the situational trust rating for the request, considered along with the general trust rating of the agent. The basic process for this calculation is shown in Figure 2, in which the method provides the agents with a means of evaluation that is more directed

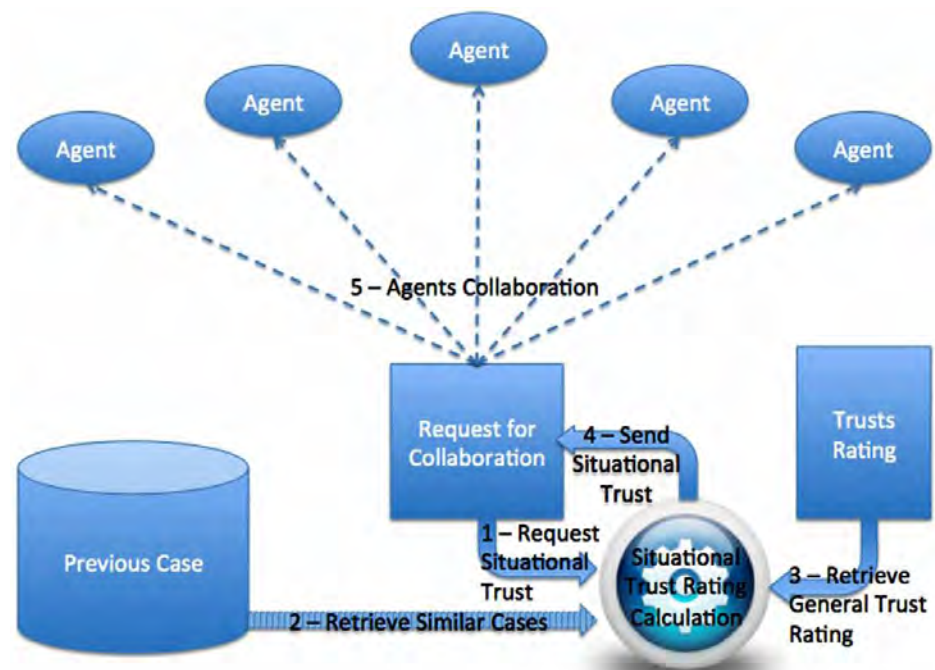
Figure 1. Basic process for updating trust ratings. Source: the authors



to the collaboration demands, allowing the trust rating to reflect the effective trust for that agent more precisely given that demand.

The cases are obtained from the collaboration between agents themselves. The second step in the sequence for collaboration shown in figure

Figure 2. Basic process for the calculation of the demand trust rating. Source: the authors



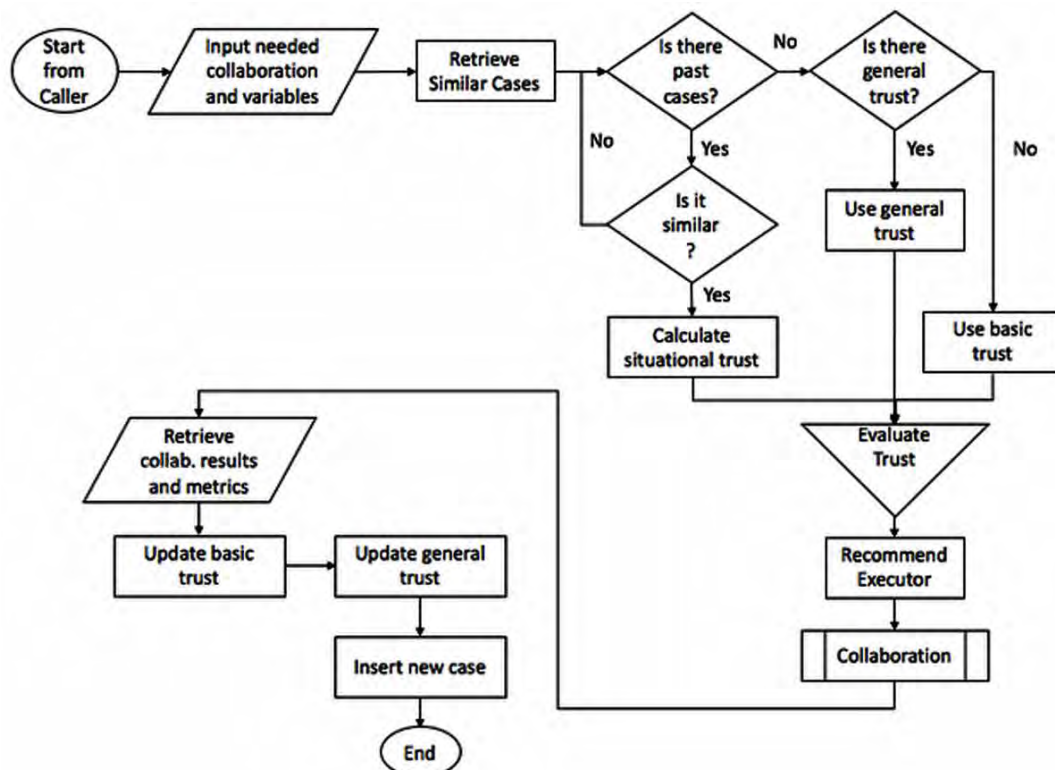
1 indicates the storage of the collaboration in a database. Trust rating needs cases for its validation, but these did not exist yet for a first interaction. Thus, we have a situation of double dependency, in which the trust rating depends on cases and the generation of cases depends on the trust rating. This situation is most likely at the start of a relationship when there are no pre-existent cases. In order to deal with this issue, the initial trust rating may be arbitrated, i.e. originated through the use of the basic trust rating of another agent, permitting the system to consider the agent for the next collaboration request. The adoption of an initial trust rating of zero, that does not indicate distrust has also been set up, but that there has been no trust established yet.

Hierarchical algorithms were considered for the selection of cases that should be used in the calculation of the situational trust rating. They treat the relationship between variables and allow for the creation of new variables based on

information originated from the case itself or even from the systems.

The trust ratings are not calculated in a pairwise manner because all the measures are made by one side (by the *caller* of the collaboration). All calculation considers the results achieved by the *caller* perspective. That means all metric used to measure the trust between *caller* and executor are set and evaluate by the *caller* agent. The system is focused in figuring out the most trustful agents to collaborate from the caller perspective. When collaboration is needed the *caller* agent has a list of agents that can provide that interaction and their respective trust in different perspectives: trust in the agent itself, that is indicated by the general trust which covers all of the interactions between agents; trust in the agent during collaboration similar to the desired one, indicated by situational trust which is measured considering all cases similar to the intended collaboration; and the basic trust, which is the trust of the caller on the network it is

Figure 3. Algorithm flowchart. Source: the authors



exposed. Considering all prior interactions without filtering the executor, the Basic trust indicates if the available network is trustworthy to the *caller* agent. Figure 3 shows the meta algorithm flowchart flowed by the proposed method.

Given that all relations generate new cases, there is a tendency toward the increase of the case base for agents that have high volumes of collaboration. As a means of avoiding a huge case base, a mechanism was created to exclude less expressive or older cases. The mechanism searches for cases that have the least impact on the trust rating, trying to ensure that the withdrawal of cases from the base does not alter the global trust rating. The mechanism also considers the quantity of cases similar to the case being withdrawn, ensuring that unique cases remain in the case base.

As a means of presenting trust as a number that can be subject to calculation, the method was created with two possibilities for demonstrating this value, the first being from an external source - stored together with the case information and referenced in future calculations. The second, who was implemented, was through the indication of a variable that reflects the trust in the agent. For this paper, the variable of response time was selected to indicate trust - how the calculation works will be explained later. It is important to stress that the method constructed could consider more than one variable as determinants of trust, for example response time and number of items returned.

In order to obtain a trust value for a task, agent or method through execution time, the formula constructed considered that all agents have trust 0 at the start of the system. If there was no trust information in the case database then the first collaboration would have a trust of 0.99 if it finished successfully or -1 in the case that it concluded with an error. From the second case, the system will calculate the average execution time for a task, agent and method. On the basis of this calculation there will be a deviation in respect to the collaboration in question to indicate trust. Collaborations with performance slower than average will have reduced trust and collaborations with faster-than-

average performance will have increased trust. Collaborations that result in errors always reduce the trust in proportion to their deviation.

It is important to mention that trust in our approach can be measured in any context. Our approach does not create limitations to any level of evaluation. However, since the system is based on the relationship between *caller* - the agent who is demanding the collaboration - to executor on the service perspective it is more likely to be used on inter-organizational or multi-organizational. In these two approaches the caller agent will always be the organization that evaluates the trust among other company or service. In inter-organizational the system is trying to identify which of its service is the most trustful. In multi-organizational, the executor will be another organization and the caller agent is expanding its evaluation to the market looking for the most trustful organization among several companies that provide the desired collaboration.

Tests and Validation

Several tests scenarios were created to validate the presented method, based on a determined and fictional environment. We created an environment isolated as much as possible from external variables to run the tests. As to simplify test scenarios, the environment was created with two agents (*fCalculadora* and *iCalculator*) with similar purposes - adding two numbers - and the agent caller. The intention of caller's collaboration was to add two numbers. Focusing on situational trust calculation, these scenarios considered fixed values for utility and importance, so the tests showed only the variation in the trust itself.

In each scenario we interfered in a different variable. There is a scenario that we change the day of execution to impact the situational trust. A collaboration performed on Monday may be different from collaboration on Friday; perhaps these are two situational trusts. In another scenario we forced the collaboration to result in an error, this situation has to affect negatively all the trust, in

a higher degree the situational trust, in a median level the general trust and in a lower degree the basic trust. This goal was to prove that trust is a reflection of all interactions, however the impact on the rating will happen accordingly how representative is the case to the trust index (basic, general and situational).

All test scenarios executed had their basic trust rating equal to 0.1546, general trust rating -0.0335973 for the agent *fCalculadora* with method “soma” and for the agent *iCalculator* method “add” the rating was 0.342863. According to the table of stratification trust values, these ratings meant the system had low trust in all its environment, the agent *iCalculator* had a trust rating classified as low medium trust and the agent *fCalculadora* had a low distrust.

The first test scenario aimed to verify the situational trust considering these hypothetical values:

- First sum value: 10.00125;
- Second sum value: 10.002;
- Execution date and time: 11/16/2014 at 08:29:00 PM
- Execution weekday: Sunday.

Those values were considered in the process of retrieving similar cases to this one and are also included in the similarity calculation. Figure 4 shows the result for this test scenario. Through the retrieved cases, the situational trust rating for the “add” method was 0.33, creating a weighted trust of 0.3364315 (the weighted trust is the average between situational trust and general trust). According to the weighted trust, the recommendation was directed to the method “add”. More importantly, this test scenario did not update the case base and, therefore, there were no trust ratings update.

The second test scenario considered the same execution date, but the added values were greatly reduced. Thus, the retrieved algorithm selected different cases from the first test scenario. The following values were considered in this test:

Figure 4. Results obtained from the first test scenario. Source: the authors

```

-----Task Information-----
Task Runtime: 5
Task Trust: 0.192279
-----
Method: add
Method Runtime: 3
Method General Trust: 0.342863
Method Situational Trust: 0.33
Weighted Trust: 0.3364315
-----
Method: soma
Method Runtime: 2
Method General Trust: -0.0335973
Method Situational Trust: 0.23935127
Weighted Trust: 0.10287698
-----Suggested Method-----
|Id: 1 - add
-----
Done!!!

```

- First sum value: 10125;
- Second sum value: 102;
- Execution date and time: 11/16/2014 at 08:29:00 PM
- Execution weekday: Sunday.

The basic and general trust ratings were the same as the first test scenario. Figure 5 shows the results obtained. The main difference from the previous run is the suggested method – “soma”. This happened because of the increase of 247% in the situational trust rating on the suggested method when both runs were compared. In addition, there was a reduction from 0.33 to -0.17134847 in the situational trust rating for the “add” method. This reduction may have happened because either the elapsed time or the error rating from the retrieved case was higher.

For the third test scenario, the same values of the first scenario were maintained, but the execution date and time were modified to the middle of the week. In the first scenario, the weekday of execution was during the weekend and the shift had been moved from the evening to the morning.

Figure 5. Results obtained from the second test scenario. Source: the authors.

```

-----Task Information-----
Task Runtime: 5
Task Trust: 0.192279
-----
Method: add
Method Runtime: 3
Method General Trust: 0.342863
Method Situational Trust: -0.17134847
Weighted Trust: 0.08575726
-----
Method: soma
Method Runtime: 2
Method General Trust: -0.0335973
Method Situational Trust: 0.5703666
Weighted Trust: 0.26838467
-----Suggested Method-----
|Id: 2 - soma
-----
Done!!!

```

The focus on this scenario was to prove the efficiency of the retrieval algorithm and effectiveness of calculations considering changes that happen in the context of collaboration.

The values considered in this run were:

- First sum value: 10.00125;
- Second sum value: 10.002;
- Execution date and time: 11/19/2014 at 08:29:00 PM;
- Execution weekday: Wednesday.

The results justify the recommendation of the “add” method, but as shown in figure 6 there was a reduction in the situation trust level and weighted trust level of this method. This behavior can be explained by displacement of the execution date. Although the other values were kept, the change in the execution date forces the similarity calculation to adapt its results, so new cases are considered then. This scenario shows how the algorithms have the ability to learn with previous cases.

In the last test we applied the same information used in the previous test. However, we updated the collaboration result as an error. This modification allowed the reduction of the basic and general ratings, as shown in figure 7.

Figure 6. Results obtained from the forth test scenario. Source: the authors

```

-----Task Information-----
Task Runtime: 5
Task Trust: 0.192279
-----
Method: add
Method Runtime: 3
Method General Trust: 0.342863
Method Situational Trust: 0.083101414
Weighted Trust: 0.21298221
-----
Method: soma
Method Runtime: 2
Method General Trust: -0.0335973
Method Situational Trust: 0.23935127
Weighted Trust: 0.10287698
-----Suggested Method-----
|Id: 1 - add
-----
Done!!!

```

Figure 7. Results obtained from the fifth test scenario. Source: the authors

```

-----Task Information-----
Task Runtime: 5
Task Trust: 0.192279
-----
Method: add
Method Runtime: 3
Method General Trust: 0.342863
Method Situational Trust: -0.17134847
Weighted Trust: 0.08575726
-----
Method: soma
Method Runtime: 2
Method General Trust: -0.0335973
Method Situational Trust: 0.5703666
Weighted Trust: 0.26838467
-----Suggested Method-----
|Id: 2 - soma
-----
-----Updated Trust-----
Basic Trust: -0.0064342767
Method General Trust: -0.355732
-----
Case inserted in the database
Done!!!

```

An execution with performance lower than the average of the past cases reduces the trust ratings, but this reduction is proportional to the percentage of reduction in the performance. When the collaboration occurs with any error - such as timeout or returned error - the reduction is higher due to all the collaboration reversion in the trust rating. Thus, the greater historical cases based the lower reduction curve.

CONCLUSION

This paper presented a trust case-based model applied in multi-agent environments to deal with the collaboration between agents. This method quantifies agent collaboration and generates trust ratings. Therefore, more efficient collaboration between agents is obtained, improving the overall efficiency of multi-agent systems.

Although Marsh's (1994) definition has more than 20 years, it still presents itself as a powerful approach to deal with the concept of trust in the computational environment. This paper was developed joining solid concepts and an innovative manner to tackle the problems related to the evaluation of agent's collaboration.

Several aspects of the collaboration were represented in the proposed method through the use of different measures for trust. Such as: The basic trust index that considers all collaborations from the caller to the environment. General trust gives the vision of the caller agent to another agent considering all previous collaborations. Situational trust goes another step deeper, considering the trust based on past collaborations similar to the desired one. Those three levels create a great flexibility to represent trust, without creating fixed barriers on qualitative variables.

Due to the heavy need of past information the situational trust was the most challenging measure to be implemented. The approach suggested here by the using of CBR proved very handful since

it created a practical way to evaluate and select the relevant information. Different tests were run, showing that the situational confidence rating is extremely sensitive to the context of the collaboration.

Therefore, the evaluation and selection of past cases are critical components in this method. The correct maintenance in the cases database as well the right approach regarding its indexing process are also crucial for achieving the best results. As noted in the tests, the proposed method could change the way cases were retrieved, adding greater assertiveness and flexibility to the context modification.

FUTURE RESEARCH DIRECTIONS

As a future research, this method can provide information to reputation models and also could be extended to measure trust in a pairwise manner. As developed, the method does not share information regarding their collaboration to other agents. Another improvement can be to allow the executor to evaluate its trust in the caller. Currently, just the caller can evaluate the executor. This extension is somewhat relevant in an environment where the executor does not know the caller. Furthermore, this evaluation of trust would allow the executor to avoid being exploited by the caller.

REFERENCES

- Aamodt, A., & Plaza, E. (1994). Case-based reasoning: Foundational issues, methodological variations, and system approaches. *AI Communications*, 7(1), 39–59.
- Abedinzadeh, S., & Sadaou, S. (2014). A trust-based service suggestion system using human plausible reasoning. *Applied Intelligence*, 41(1), 55–75. doi:10.1007/s10489-013-0495-8

- Barbati, M., Bruno, G., & Genovese, A. (2012). Applications of agent-based models for optimization problems: A literature review. *Expert Systems with Applications*, 39(5), 6020–6028. doi:10.1016/j.eswa.2011.12.015
- Bedi, P., & Vashisth, P. (2014). Empowering recommender systems using trust and argumentation. *Information Sciences*, 279(1), 569–586. doi:10.1016/j.ins.2014.04.012
- Buccafurri, F., Comi, A., Lax, G., & Rosaci, D. (2014). A Trust-Based Approach to Clustering Agents on the Basis of Their Expertise. *Agent and Multi-Agent Systems: Technologies and Applications*, 296(1), 47–56.
- Challenger, M., Mernikb, M., Kardasa, G., & Kosarb, T. (2016). Declarative specifications for the development of multi-agent systems. *Computer Standards & Interfaces*, 43(1), 91–115. doi:10.1016/j.csi.2015.08.012
- Floyd, M. W., Drinkwater, M., & Aha, D. W. (2014). How Much Do You Trust Me? Learning a Case-Based Model of Inverse Trust. *Case-Based Reasoning Research and Development. 22nd International Conference, ICCBR 2014, Cork, Ireland, September 29, 2014 - October 1, 2014. Proceedings*, 8765(1), 125–139.
- Hsieh, F., & Lin, J. (2014). A dynamic scheme for scheduling complex tasks in manufacturing systems based on collaboration of agents. *Applied Intelligence*, 41(2), 366–382. doi:10.1007/s10489-014-0521-5
- Khosravifar, B., Gomrokchi, M., Bentahar, J., & Thiran, P. (2009) Maintenance-based Trust for Multi-Agent System. *Proceedings of the 8th International Conference on Autonomous Agents and Multiagent Systems, Int. Foundation for Autonomous Agents and Multiagent Systems*, 8(1), 1017-1024.
- Kraus, S. (2001). *Strategic Negotiation in Multi-Agent Environments*. Cambridge, MA: MIT Press.
- Marsh, S. P. (1994). *Formalising Trust as a Computational Concept*. Stirling, UK: University of Stirling.
- Pynadath, D., & Tambe, M. (2002). Multiagent teamwork: Analyzing the Optimality and Complexity of Key Theories and Models. In *Proceedings AAMAS '02 The First International Joint Conference on Autonomous Agents and Multi-Agent Systems (formerly known as Autonomous Agents)* (vol. 1, pp. 873-880). Bologna, Italy: ACM. doi:10.1145/544862.544946
- Ramchurn, S. D., Huynh, D., & Jennings, N. R. (2004). Trust in multi-agent systems. *The Knowledge Engineering Review*, 19(1), 1–25. doi:10.1017/S0269888904000116
- Rosaci, D. (2012). Trust measures for competitive agents. *Knowledge-Based Systems*, 28(1), 38–46. doi:10.1016/j.knsys.2011.11.011
- Russel, S., & Norvig, P. (2003). *Artificial Intelligence*. Prentice-Hall.
- Sabater, J., & Sierra, C. (2005). Review on Computational Trust and Reputation Models. *Artificial Intelligence Review*, 24(1), 33–60. doi:10.1007/s10462-004-0041-5
- Stefan, K. A., & Weß, S. (1991). Case-Based Knowledge Acquisition, Learning and Problem Solving for Diagnostic Real World Tasks. *Proceedings EKAW-91, European Knowledge Acquisition Workshop*, 1, 48-67.

ADDITIONAL READING

- Aydoğan, R., Sharpanskykh, A., & Lo, J. (2015). A Trust-Based Situation Awareness Model. *European Conference on Multi-Agent Systems EUMAS 2014: Multi-Agent Systems*, 19–34.

Biswas, S., Chakraborty, M., Singh, H. R., Devi, D., Purkayastha, B., & Das, A. (2016). Hybrid case-based reasoning system by cost-sensitive neural network for classification. *Soft Computing*, 1–18.

Craw, S. (2016). Case-Based Reasoning. *Encyclopedia of Machine Learning and Data Mining*, 1–9.

Fan, Z., Li, Y., Wang, X., & Liu, Y. (2014). Hybrid similarity measure for case retrieval in CBR and its application to emergency response towards gas explosion. *Expert Systems with Applications*, 41(5), 2526–2534. doi:10.1016/j.eswa.2013.09.051

Leake, D. (2015). Problem Solving and Reasoning: Case-Based. *International Encyclopedia of the Social & Behavioral Sciences* (Second Edition), 2(1), 56–60.

Liu, X., Datta, A., & Lim, E. (2015). *Computational Trust Models and Machine Learning*. CRC Press.

Pinyol, I., & Sabater-Mir, J. (2013). Computational trust and reputation models for open multi-agent systems: A review. *Artificial Intelligence Review*, 40(1), 1–25. doi:10.1007/s10462-011-9277-z

Tavakolifard, M., Herrmann, P., & Öztürk, P. (2009). Analogical Trust Reasoning. *IFIP International Conference on Trust Management, IFIPTM 2009: Trust Management III*, 149–16325.



KEY TERMS AND DEFINITIONS

Case-Based Reasoning: An approach that solves new problems using solutions applied in previous cases.

Collaboration: When two or more agents collaborate to perform a task.

General Trust: Represents the trust of an agent in a specific interlocutor, considering all the interactions between these two.

Multi-Agent Systems: A system composed by agents that are able to interact among themselves and within an environment.

Situational Trust: Reflects the trust of an agent in an interlocutor in a specific situation.

Trust Rating: The level of the trust among the agents.

Trust: A measurable level of the probability which an agent knows that another agent will perform a particular action.

Understanding and Assessing Quality of Models and Modeling Languages

John Krogstie

Norwegian University of Science and Technology, Norway

INTRODUCTION

An important challenge for organizational activity is to effectively represent and transfer *knowledge*. One reason why humans have excelled as species is our ability to create common stories and represent, reuse and transfer this as knowledge across time and space. Whereas in most areas of human conduct, one-dimensional natural language texts are the main way of expressing and sharing knowledge, we see the need for and use of two and many-dimensional forms of knowledge representational to be on the rise. A form of representation which plays an increasingly important role in information systems and enterprise development is *conceptual models* (Krogstie, Opdahl, & Brinkkemper, 2007), which are diagrams expressed in some (semi-) formal visual language (e.g. nodes interconnected with edges), describing some area of interest. In (Krogstie, 2007), the following vision was stated: “Over time the use of modelling will become an established way of expressing knowledge in all fields of human conduct. Everyone (not only expert ‘modelers’) is involved in the process of developing and activating models made in adapted modelling language.” Examples of models could be organization charts, strategy and goal breakdown structures, business process models, or models of the information to be contained in a database. The *quality* of a conceptual model will strongly affect decisions based on the model, and can therefore be of vital importance to the stakeholders.

According to general model theory (Stachowiak, 1973) there are three common characteristics of models: *Representation*, *Simplification* and *Pragmatic orientation*.

- **Representation:** Models represents something else than the model itself.
- **Simplification:** Models possess a reductive trait in that they represent only a subset of attributes of the phenomenon being modelled.
- **Pragmatic Orientation:** Models have a substitutive function in that they substitute a certain phenomenon as being conceptualized by a certain subject in a given temporal space with a certain *intention* or operation in mind.

Thus a model is not just a representation of something else; it is a conscious construction to achieve a certain goal beyond the making of the model itself. Whereas modeling techniques traditionally have been used to create intermediate artifacts in systems analysis and design, modern modeling methodologies support a more active role for the models. For instance in Business Process Management (BPM) (Havey, 2005), Model Driven Architecture (MDA) and Model-driven Software Engineering (MDSE) (Brambilla, Cabot & Wimmer, 2012), Domain specific modeling (DSM) (Kelly & Tolvanen, 2008), Enterprise Architecture (EA) (Lankhorst, 2005), Enterprise modeling (EM) (Sandkuhl, Stirna, Persson & Wißotzki, 2014), Interactive Models (Krogstie & Jørgensen, 2004) and Active Knowledge Modelling (AKM) (Lillehagen & Krogstie, 2002; Lillehagen & Krogstie, 2008), the models are used directly as part of the information system of the organization. At the same time, similar modeling techniques are also used for sense-making and communication, model simulation, quality assur-

ance, and requirements specification in connection to more traditional forms of information systems and enterprise development (Krogstie, Dalberg & Jensen, 2008).

Since modeling techniques are used in such a large variety of tasks with different goals, it is hard to assess whether a model is sufficiently *good* to achieve the goals. To provide guidance in this process, the latest version of *SEQUAL*, a framework for understanding quality of models and modeling languages, will be presented in this chapter.

BACKGROUND

Since the early 90ties, many researchers have worked on quality of models. Work on *SEQUAL* can be traced back to at least 1993 (Lindland, 1993). Sindre and Lindland in particular collaborated on the next step (Lindland, Sindre & Sølvsberg, 1994). Although an elegant framework which was easily applicable for understanding important aspects of quality of models, several other works pointed to the need for extending the framework. Important inspirations in this regard was the work on 3 dimensions of requirements engineering (Pohl, 1993), work related to the semiotic ladder presented in early versions of the IFIP 8.1 FRISCO framework (Lindgren, 1990) and work on social construction of ‘reality’ and models thereof of the domain, which is typically not as ideal and objectively given in practice that as the original framework took as an outset (Berger & Luckmann, 1966). Specifically the framework of Pohl also pointed to the need for achieving *agreement* among the stakeholders of the model.

These extensions, in addition to a specific focus on requirements specification models resulted in the framework presented in (Krogstie, Lindland & Sindre, 1995). At the same time Shanks and Moody (Moody & Shanks, 1994) started their work on quality of data models. Becker, Rosemann and Schütte (1995) focused on the quality of process

models. Later a number of other researchers, e.g. (Nelson, Poels, Genero & Piattini, 2011) have worked within this area.

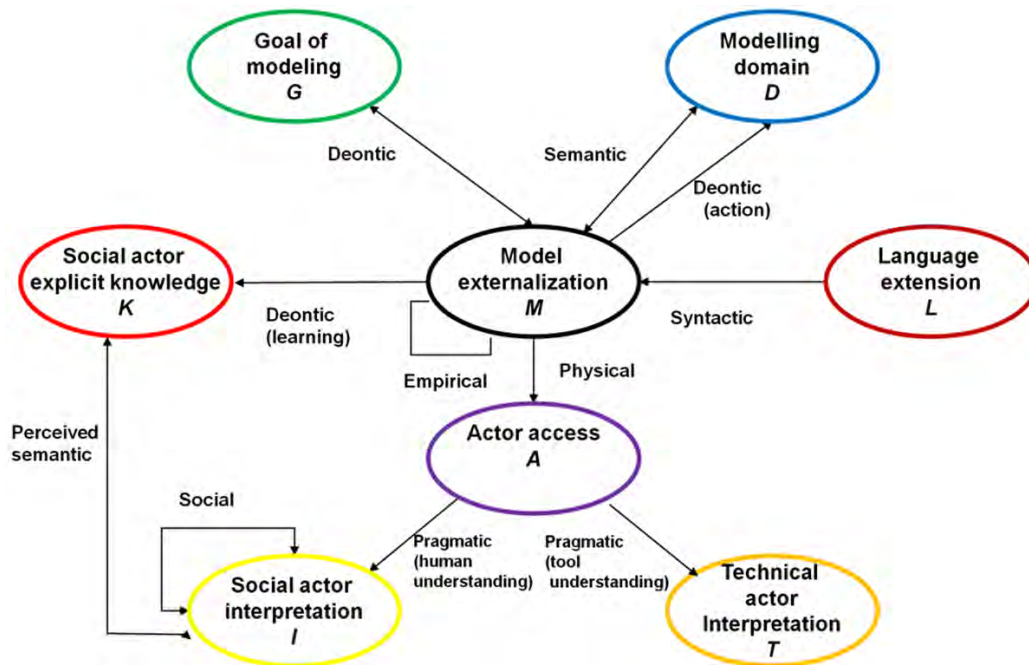
In hindsight the work done on *SEQUAL* can be framed as design science research (Hevner et al, 2004), with the quality framework as the core artifact. Whereas the early validation was primarily analytical, later work e.g. together with Moody (Moody, Sindre, Brasethvik & Sølvsberg, 2002) also included empirical evaluations, and practical applications of the framework have been reported in (Heggset, Krogstie & Wesenberg, 2014; Heggset, Krogstie & Wesenberg, 2015b). The framework has been developed through a number of iterations, and have also in some cases been established as part of the knowledge base e.g. in the development of a framework for quality of maps (Nossum & Krogstie, 2009). The current version of the framework is described in (Krogstie, 2012a) where also newer work on quality of modeling languages (including the work presented by Moody (2009)) is incorporated.

To summarize, *SEQUAL* has three unique properties compared to the early work on quality of models:

- It distinguishes between quality characteristics (goals) and means to potentially achieve these goals by separating what you are trying to achieve from how to achieve it.
- It is based on a constructivistic worldview, recognizing that significant models are usually created as part of a dialogue between the many stakeholders involved in modelling, whose knowledge of the modelling domain changes as modelling takes place.
- It is closely linked to linguistic and semiotic concepts. In particular, the core of the framework including the discussion on syntax, semantics, and pragmatics is parallel to the use of these terms in the se-



Figure 1. SEQUAL framework for discussing quality of models



miotic theory of Morris. Also the work in FRISCO on the semiotic ladder took the work of Morris as an outset, but has extended this with physical, empirical, and social aspects which we have adapted.

FRAMEWORK FOR QUALITY OF MODELS

The current framework is illustrated in Figure 1. Quality has been defined referring to the correspondence between statements belonging to the following sets (the sets depicted as ellipses):

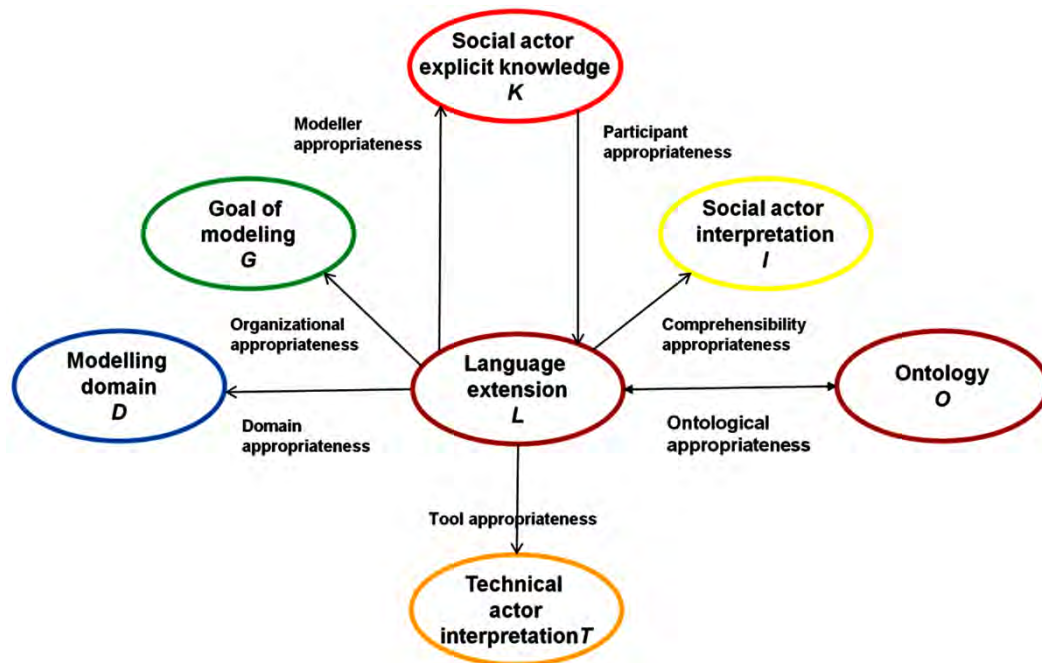
- **G**, the set of goals of the modeling task.
- **L**, the language extension, i.e., the set of all statements that are possible to make according to the rules of the modeling languages used.
- **D**, the domain, i.e., the set of all statements that can be stated about the situation.
- **M**, the externalized model itself.

- **A**, the part of the model that can be accessed by one or more actor, actors being either persons and tools.
- **K**, the explicit knowledge relevant to the domain of the audience.
- **I**, the social actor interpretation, i.e., the set of all statements that the audience interprets that an externalized model consists of.
- **T**, the technical actor interpretation, i.e., the statements in the model as ‘interpreted’ by modeling tools.

The main quality types as illustrated as relationships in Figure 1 are:

1. Physical quality: The basic quality goal is that the externalized model **M** is available to the relevant actors **A**.
2. Empirical quality deals with comprehensibility when a visual model **M** is read by different social actors. Before evaluating empirical quality, physical quality should be addressed.

Figure 2. Seven areas of language quality



3. Syntactic quality is the correspondence between the model M and the language extension L . Before evaluating syntactic quality, physical quality should be addressed.
4. Semantic quality is the correspondence between the model M and the domain D . This includes both validity and completeness. Before evaluating semantic quality, syntactic quality should be addressed. Domains can be divided into two parts, exemplified with a software requirements model:
 - i. Everything the computerized information system (CIS) is supposed to do (for the moment ignoring the different views the stakeholders have on the CIS to be produced).
 - ii. Constraints on the model because of earlier baselined models such as system level requirements specifications, enterprise architecture models, statements of work, and earlier versions of the requirement specification model to which the new requirement specification model must be compatible.
5. Perceived semantic quality is the similar correspondence between the social actor interpretation I of a model M and his or hers current knowledge K of domain D . Before evaluating perceived semantic quality, pragmatic quality (see below) should be addressed.
6. Pragmatic quality is the correspondence between the available part of the model M (i.e. A) and the actor interpretation (I and T) of it. One differentiates between social pragmatic quality (to what extent people understand the model) and technical pragmatic quality (to what extent tools can be made that can interpret the model). Before evaluating pragmatic quality, empirical quality should be addressed.
7. The goal defined for social quality is agreement among social actor's interpretations (I). Before evaluating social quality, perceived semantic quality should be addressed.
8. The deontic quality of the model relates to that all statements in the model M contribute to fulfilling the goals of modeling G , and that

all the goals of modeling *G* are addressed through the model *M*. In particular, one often includes under deontic quality learning and domain change.

Language quality goals are looked upon as means in the framework. 7 areas of language quality are differentiated as illustrated in Figure 2:

1. Domain appropriateness. This relates the language and the domain. Ideally, the conceptual basis must be powerful enough to express anything in the domain, not having what (Wand & Weber 1993) terms construct deficit. On the other hand, you should not be able to express things that are not in the domain, i.e. what is termed construct excess (Wand & Weber 1993). Domain appropriateness is primarily a mean to achieve semantic quality.
2. Ontological appropriateness. Whereas domain appropriateness discusses the languages appropriateness relative to a particular domain, when discussing expressiveness in general, one might want to do that relative to some general framework or ontology developed for the field. Any modeling approach or perspective taken includes some level of ontological commitment, and general discussions on construct deficit and excess is typically done relative to some existing ontology.
3. Comprehensibility appropriateness relates the language to the social actor interpretation. The goal is that the participants in the modeling effort using the language understand all the possible statements of the language. Comprehensibility appropriateness is primarily a mean to achieve empirical and pragmatic quality.
4. Participant appropriateness relates the social actors' knowledge of the modeling language (i.e. do the participants know the language being used). Participant appropriateness is primarily a mean to achieve semantic and pragmatic quality.
5. Modeler appropriateness: This area relates the language extension to the knowledge of the modeler. The goal is that there are no statements in the explicit knowledge of the modeler that cannot be expressed in the language. Modeler appropriateness is primarily a mean to achieve semantic quality.
6. Tool appropriateness relates the language to the technical audience interpretations. For tool interpretation, it is especially important that the language lend itself to automatic reasoning. This requires formality (i.e. both formal syntax and semantics being operational and/or logical), but formality is not necessarily sufficient, since the reasoning must also be efficient to be of practical use. This is covered by what we term analyzability (to be able to exploit any mathematical semantics efficiently) and executability (to be able to exploit any operational semantics efficiently). Different aspects of tool appropriateness are means to achieve syntactic, semantic and pragmatic quality (through formal syntax, mathematical semantics, and operational semantics respectively).
7. Organizational appropriateness relates the language to standards and other organizational needs and constraints within the organizational context of modeling. These are expected to support deontic quality in particular to support efficient modeling within the organization.

Issues and Recommendations for Dealing With These Issues

Whereas SEQUAL over the year has evolved to a very comprehensive framework, it has been found hard to understand and use in practice both by practitioners and researchers in the field (Reijers, Mendling and Recker 2010). One reason that the framework is complex to use in practice is that it

is very generic. Inspired by (Moody, 2009), suggesting the need for an inheritance hierarchy of quality frameworks, we have started to develop a number of specialized version of SEQUAL applicable for specific types of modeling, see e.g. the specialization of SEQUAL for business process models in (Krogstie, 2012b, Krogstie, 2016).

As illustrated in fig. 3 below there are a number of specializations of SEQUAL. As with GoM (Becker et al. 1995), we have a first level meant to be relevant for all node-edge-oriented models (SEQUAL-GEN). A second level looks on a particular type of models. Finally one can have specific guidelines on an even more detailed level, e.g. for business process modelling using BPMN. In real-world modelling activities we find that all quality levels are important, but the *weight* on different levels is different based on different goals of modelling.

Basis for the different specializations can be found in (Krogstie, 2012a):

- **SEQUAL- SRS:** Software requirements specifications (Krogstie, 1999; Krogstie, 2001), including quality of goal models (Krogstie, 2008)
- **SEQUAL - DM:** Data models (Krogstie, 2013a)
- **SEQUAL - DQ:** Data quality (Krogstie, 2013b) and data integration quality (Krogstie, 2015))
- **SEQUAL - IM:** Interactive models (Krogstie & Jørgensen, 2002)
- **SEQUAL - EM:** Enterprise models (Krogstie & Arnesen, 2004)
- **BPM:** Business Process Models (Krogstie, 2012b; Krogstie, 2016), in particular BPM using BPMN (Recker, Rosemann & Krogstie, 2007, Aagesen & Krogstie 2010, Aagesen & Krogstie, 2015)
- **SEQUAL - ONT:** Ontologies (Hella & Krogstie, 2010)

Figure 3. Specialization of SEQUAL based on modelling perspective

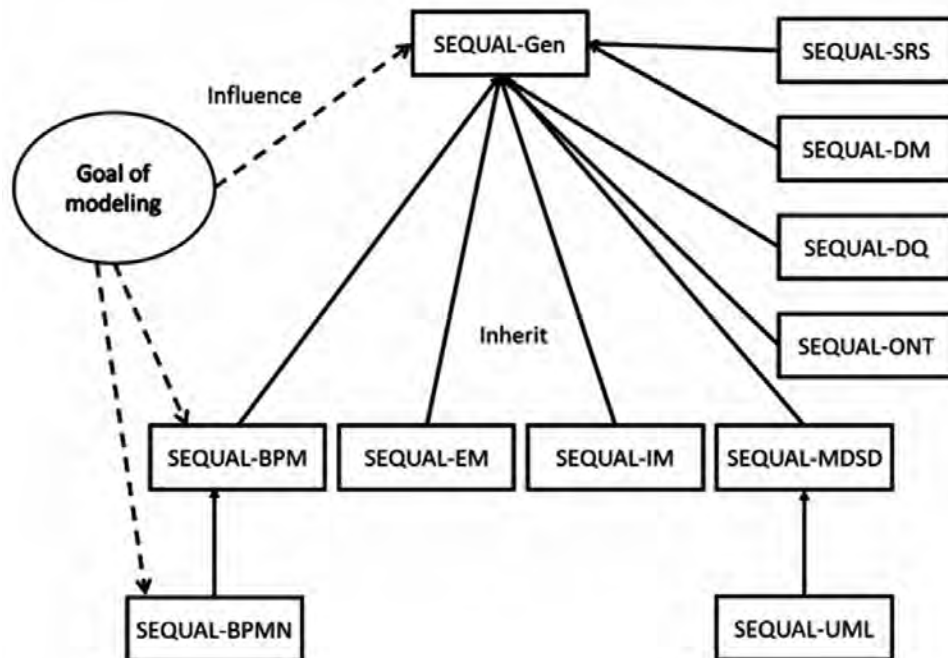


Table 1. Types of models according to temporal aspects and purpose

Type of Model	Past	Present	Future
Ideal model	Ideal model of the past	Reference model	Ought-to-be model
Simulated model (what-if)	Possible model of the past	Possible model	Possible model of the future
Model espoused	As-was model	As-is model	To-be model
Model in use	Actual as-was model	Actual as-is model	Workaround model
Motivational model	Past burning-platform model	Burning platform model	Burning platform model

- **SEQUAL - MDSD:** Models used in Model-driven Software Development, with a focus on quality of UML-models (Krogstie, 2003)
- **MAPS - MAPQUAL:** (not in the figure) (Nossum & Krogstie, 2009)

Another dimension of specialization relates to the type of model. Traditionally, one have mainly differentiated between as-is (descriptive) and to-be (prescriptive models). Recent work (Fossland & Krogstie, 2015) has illustrated the need for a larger set of model types. In Table 1, we list relevant situations, along the temporal and ontological axes.

First of all models can be of past situations, the present, or a potentially future situation. One look here primarily at the present and future, noting that in several areas, also to keep track of the processes used in the past is important, e.g. in areas where compliance is important and where one can get into situations where the authorities questions the adherence to compliance rules in the past. Models can at all temporal stages be looked upon as being

- **Ideal:** A model of an ideal situation in the area, ignoring contextual restrictions such as legacy systems and installed base.
- **Simulated:** A model that differs in some way to the actual state-of-things, e.g. to be able to play what-if analysis.
- **Model Espoused:** The official model in a restricted area.

- **Model in Use:** How the situation actually is (or was). This should ideally not be so different from the model espoused, but in practice we often experience this.
- **Motivational Model:** A simulation which depicts a defensive approach i.e. what happen if nothing is done. This is often termed the burning-platform model.

FUTURE RESEARCH DIRECTIONS

By making the framework more operational and instrumental, with precise guidelines that can be implemented in modeling tools to directly support the modeling process, and with stronger validation in real, industrial case studies one can take the SEQUAL-framework a great leap forward

Secondary objectives to achieve this are to validate and extend the existing SEQUAL framework, in particular to also support the process of modeling to end up with and evolve models of high quality. In particular to extend the framework to support the process of

- Enterprise modeling, including business process modeling and evolution of interactive models.
- Requirements modeling, including the integrated representation of functional and non-functional requirements. Business process models can also play a role here.
- Improve modeling approaches in enterprise modeling and requirements modeling to take these extension into account.

An appropriate methodology for this type of research is design science research (Hevner et al., 2004) which aims to improve or extend human and organizational capabilities by building new artifacts. Traditionally, four types of artifacts are described; constructs, models, methods and instantiations. A framework for design science is presented by Peffers et al. (2006). In future research on quality of models, we propose in particular to use the following research methods:

- Experiments with individual modelers, investigating how they develop models on a limited domain. We will follow an experimental approach for controlled experiments (Gopalakrishnan, Sindre & Krogstie, 2012) using tools such as Cheetah (Pinggera et al., 2012) that can track detailed modeling activities. From literature (Krogstie, 2012a; Pinggera et al., 2012) starting hypothesis for this setting exist, thus we will first have an objective-centered initiation. When tools and methods are developed to support this modeling-setting, studies having a design and development centered initiation can be performed.
- Group experiments, investigating how groups of modelers with different expertise jointly develop models on open problem areas. Participatory modeling (Gjersvik et al., 2004, Sandkuhl et al, 2014) can be done with a main modeler as in socio-technical walkthrough (Prilla & Jahnke, 2012) using a traditional tool (like Cheetah), or with parallel input from all the participants, in which collaborative tools like CubeBPM is appropriate (Nolte et al 2015). We are currently experimenting with these types of tools, but will need to start out with a problem-centered initiation of the research.
- Longitudinal case-studies of modeling in practice, studying the modeling process of collaborators already applying the core SEQUAL framework (e.g. Statoil (Wesenberg, 2011, Heggset, Krogstie &

Wesenberg, 2015a)). The first phase of this will have a problem-centered initiation, whereas we in the end hope to have research being initiated by external clients, evaluating the results.

CONCLUSION

More active approaches to the use of models in IT and enterprise development are increasing. In enterprise modeling and process modelling in particular one aim at representing knowledge in a format available for a wide variety of stakeholders. These techniques are being taken into use to an increasing degree in areas such as health care, public administration, the oil and gas industry, the building industry, defense, and higher education to make it possible to be able to manage the increased usage and dependency on ICT. One striking aspect is that the number and variety of stakeholders that will need to relate to models increase. Given the increased educational level in most countries, with an increasing group of people with master degrees, it is not unlikely that also more people will be able to relate to these types of abstractions, given that one of the things that you are exposed to in a master study of most types, is how to deal with abstractions. To reach the vision of modeling of the people, by the people, for the people (Krogstie, 2007), visual modeling must become as easy for e.g. designers and engineers as scribbling in order for them to express their knowledge while performing work, reflecting on this for learning and excelling in their roles. This will also enable users to capture contextual dependencies between roles, tasks, information elements and the views required for performing work without having to go through traditional systems developers to have enhanced support for their work. The importance on supporting judgment on quality of these models will thus increase, and also the usefulness of frameworks for quality of models and modeling languages such as SEQUAL.

REFERENCES

- Aagesen, G., & Krogstie, J. (2010). Analysis and design of business processes using BPMN. In J. vom Brocke & M. Rosemann (Eds.), *Handbook on Business Process Management*. Springer. doi:10.1007/978-3-642-00416-2_10
- Aagesen, G., & Krogstie, J. (2015). BPMN 2.0 for modeling business processes. In J. vom Brocke & M. Rosemann (Eds.), *Handbook on Business Process Management 2.0 edition*. Springer. doi:10.1007/978-3-642-45100-3_10
- Becker, J., Rosemann, M., & Schütte, R. (1995). Guidelines of Modelling (GoM). *Wirtschaftsinformatik*, 37(5), 435–445. (in German)
- Berger, P., & Luckmann, T. (1966). *The Social Construction of Reality: A Treatise in the Sociology of Knowledge*. Penguin.
- Brambilla, M., Cabot, J., & Wimmer, M. (2012). *Model-driven Software Engineering in practice*. Moregan Claypool publishers.
- Fossland, S., & Krogstie, J. (2015). Modeling As-is, Ought-to-be and To-be: Experiences from a Case Study in the Health Sector. *Proceedings PoEM 2015*.
- Gjersvik, R., Krogstie, J., & Følstad, A. (2004). Participatory Development of Enterprise Process Models. In J. Krogstie, K. Siau, & T. Halpin (Eds.), *Information Modeling Methods and Methodologies*. Idea Group Publishers.
- Gopalakrishnan, S., Krogstie, J., & Sindre, G. (2012). Capturing Location in Process Models: Comparing Small Adaptations of Mainstream Notation. *International Journal of Information System Modeling and Design*, 3(3), 24–45. doi:10.4018/jismd.2012070102
- Havey, M. (2005). *Essential Business Process Modelling*. O'Reilly.
- Heggset, M., Krogstie, J., & Wesenberg, H. (2014). Ensuring quality of large scale industrial process collections: Experiences from a case study. *Proceedings of Practice of Enterprise Modeling*, 11-25.
- Heggset, M., Krogstie, J., & Wesenberg, H. (2015a). Understanding Model Quality Concerns when Using Process Models in an Industrial Company. *Proceedings from EMMSAD 2015*.
- Heggset, M., Krogstie, J., & Wesenberg, H. (2015b). The Influence of Syntactic Quality of Enterprise Process Models on Model Comprehension. *Proceedings from CAiSE Forum*.
- Hella, L., & Krogstie. (2010). A Structured Evaluation to Assess the Reusability of Models of User Profiles. *Proceeding of EMMSAD - Conference on Evaluating Modeling Methods in Systems Analysis and Design*. doi:10.1007/978-3-642-13051-9_19
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science in Information Systems Research. *Management Information Systems Quarterly*, 28, 75–105.
- Kelly, S., & Tolvanen, J.-P. (2008). *Domain-Specific Modelling: Enabling Full Code Generation*. John Wiley & Sons. doi:10.1002/9780470249260
- Krogstie, J. (1999). *Using Quality Function Deployment in Software Requirements Specification*. Paper presented at the Fifth International Workshop on Requirements Engineering: Foundations for Software Quality (REFSQ'99).
- Krogstie, J. (2001). A Semiotic Approach to Quality in Requirements Specification *Proceedings of IFIP 8.1. Working Conference on Organizational Semiotics*.
- Krogstie, J. (2003). Evaluating UML Using a Generic Quality Framework. In L. Favre (Ed.), *UML and the Unified Process* (pp. 1–22). IRM Press. doi:10.4018/978-1-93177-744-5.ch001

- Krogstie, J. (2007). Modelling of the People, by the People, for the People. In J. Krogstie, A. Opdahl, & S. Brinkkemper (Eds.), *Conceptual Modelling in Information Systems Engineering*. Springer Verlag. doi:10.1007/978-3-540-72677-7_19
- Krogstie, J. (2008). Integrated Goal, Data and Process modeling: From TEMPORA to Model-Generated Work-Places. In P. Johannesson & E. Sørderstrøm (Eds.), *Information Systems Engineering From Data Analysis to Process Networks* (pp. 43–65). IGI. doi:10.4018/978-1-59904-567-2.ch003
- Krogstie, J. (2012a). *Model-based Development and Evolution of Information Systems: A Quality Approach*. Springer. doi:10.1007/978-1-4471-2936-3
- Krogstie, J. (2012b). Quality of Business Process Models. *Lecture Notes in Business Information Processing*, 134, 76–90.
- Krogstie, J. (2013a). Quality of Conceptual Data Models. *Proceedings 14th ICISO*.
- Krogstie, J. (2013b). A Semiotic Framework for Data Quality. *Proceedings EMMSAD 2013*.
- Krogstie, J. (2015, February). Capturing Enterprise Data Integration Challenges Using a Semiotic Data Quality Framework *Business & Information Systems Engineering*, 57(1), 27–36.
- Krogstie, J. (2016). *Quality in Business Process Modeling*. Springer. doi:10.1007/978-3-319-42512-2
- Krogstie, J., & Arnesen, S. (2004). Assessing Enterprise Modeling Languages using a Generic Quality Framework. In J. Krogstie, K. Siau, & T. Halpin (Eds.), *Information Modeling Methods and Methodologies*. Idea Group Publishing.
- Krogstie, J., Dalberg, V., & Jensen, S. (2005). Using a Model Quality Framework for Requirements Specification of an Enterprise Modeling Languages. In K. Siau (Ed.), *Advanced Topics in Database Research* (Vol. 4). Idea Group Publishing. doi:10.4018/978-1-59140-471-2.ch006
- Krogstie, J., Dalberg, V., & Jensen, S. M. (2008). Process modeling value framework. In Y. Manolopoulos, J. Filipe, P. Constantopoulos & J. Cordeiro (Eds.), *Selected papers from 8th International Conference, ICEIS 2006* (Vol. LNBIP 3, pp. 309–321). Paphos, Cyprus: Springer. doi:10.1007/978-3-540-77581-2_21
- Krogstie, J., & Jørgensen, H. D. (2002). Quality of Interactive Models. *Proceedings of First International Workshop on Conceptual Modelling Quality*.
- Krogstie, J., & Jørgensen, H. (2004). Interactive Models for Supporting Networked Organisations. *Proceedings of 16th Conference on advanced Information Systems Engineering (CAiSE 2004)*.
- Krogstie, J., Lindland, O. I., & Sindre, G. (1995). Defining Quality Aspects for Conceptual Models. *Proceedings of IFIP8.1 working conference on Information Systems Concepts (ISCO3): Towards a Consolidation of Views*. doi:10.1007/978-0-387-34870-4_22
- Krogstie, J., Opdahl, A., & Brinkkemper, S. (Eds.). (2007). *Conceptual Modelling in Information Systems Engineering*. Springer Verlag. doi:10.1007/978-3-540-72677-7
- Lankhorst, M. (2005). *Enterprise Architecture at Work - Modelling, Communication and Analysis*. Berlin: Springer-Verlag.
- Lillehagen, F., & Krogstie, J. (2002). Active Knowledge Models and Enterprise Knowledge Management. *Proceedings of International Conference on Enterprise Integration and Modelling Technology (ICEIMT'02)*.

- Lillehagen, F., & Krogstie, J. (2008). *Active Knowledge Modelling of Enterprises*. Springer. doi:10.1007/978-3-540-79416-5
- Lindgren, P. (Ed.). (1990). A framework for information systems concepts. Interim report FRISCO.
- Lindland, O. I. (1993). *A Prototyping Approach to Validation of Conceptual Models in Information Systems Engineering* (Unpublished doctoral dissertation). IDT, NTH.
- Lindland, O.I., Sindre, G., & Sølvsberg, A. (1994, March). Understanding Quality in Conceptual Modeling. *IEEE Software*.
- Moody, D. L. (2009). The “Physics” of Notations: Toward a Scientific Basis for Constructing Visual Notations in Software Engineering. *IEEE Transactions on Software Engineering*, 35(6), 756–779. doi:10.1109/TSE.2009.67
- Moody, D. L., & Shanks, G. G. (1994). What Makes a Good Data Model? Evaluating the Quality of Entity Relationship Models.: *Proceedings of the 13th International Conference on the Entity-Relationship Approach (ER'94)*, 94–111. doi:10.1007/3-540-58786-1_75
- Moody, D. L., Sindre, G., Brasethvik, T., & Sølvsberg, A. (2002). Evaluating the Quality of Process Models: Empirical Testing of a Quality Framework. *LNCS*, 2503, 214–231. doi:10.1007/3-540-45816-6_36
- Nelson, H.J., Poels, G., Genero, M., & Piattini, M. (2011). A conceptual modeling quality framework. *Software Quality Journal*.
- Nguyen, L., & Swatman, P. A. (2001) Managing the Requirements Engineering Process. *Proceedings REFSQ 2001*.
- Nolte, A., Brown, R. A., Poppe, E., & Craig, A. (2015). Exploring collaborative modeling of business processes on large interactive touch display walls. In *Collaboration Meets Interactive Surfaces (CMIS)*, Madeira, Portugal.
- Nossum, A., & Krogstie, J. (2009). Integrated Quality of Models and Quality of Maps. *Proceedings EMMSAD 2009*. doi:10.1007/978-3-642-01862-6_22
- Peffer, K., Tuunanen, T., Gengler, C. E., Rossi, M., Hui, W., Virtanen, V., & Bragge, J. (2006) The Design Science research process. *Proc. of DESRIST*, 83–106.
- Pinggera, J., Zugal, S., Weidlich, M., Fahland, D., Weber, B., Mendling, J., & Reijers, H. (2012). Tracing the Process of Process Modeling with Modeling Phase Diagrams. *Proceedings ER-BPM'11*, 370–382. doi:10.1007/978-3-642-28108-2_36
- Pohl, K. (1993). The Three Dimensions of Requirements Engineering. *LNCS*, 685.
- Prilla, M., & Jahnke, I. (2012). The Socio-Technical Walkthrough for Participatory Process Design. In *Implementing International Services*. Springer.
- Recker, J., Rosemann, M., & Krogstie, J. (2007). Ontology- versus pattern-based evaluation of process modeling language: A comparison. *Communications of the AIS*, 20, 774–799.
- Reijers, H., Mendling, J., & Recker, J. (2010) Business Process Quality Management. In J. vom Brocke & M. Rosemann (Eds.), *Handbook on Business Process Management I*. Springer.
- Sandkuhl, K., Stirna, J., Persson, A., & Wißotzki, M. (2014). *Enterprise Modeling: Tackling Business Challenges with the 4EM Method*. Springer. doi:10.1007/978-3-662-43725-4
- Sindre, G., & Krogstie, J. (1995). Process heuristics to achieve requirements specification of feasible quality. *Proceedings Second International Workshop on Requirements Engineering: Foundations for Software Quality (REFSQ'95)*.
- Stachowiak, H. (1973). *Allgemeine Modelltheorie*. Wien: Springer. doi:10.1007/978-3-7091-8327-4

Wand, Y., & Weber, R. (1993). On the Ontological Expressiveness of Information Systems Analysis and Design Grammars. *Journal of Information Systems*, 3(4), 217–237. doi:10.1111/j.1365-2575.1993.tb00127.x

Wesenberg, H. (2011). Enterprise Modeling in an Agile World. *Proceedings of the 4th conference on Practice of Enterprise Modeling*. doi:10.1007/978-3-642-24849-8_10

ADDITIONAL READING

Jiang, L., Barone, D., Borgida, A., & Mylopoulos, J. (2009). Measuring and Comparing Effectiveness of Data Quality Techniques. *Proceedings CAiSE, 2009*, 171–185.

Mendling, J., Reijers, H. A., & Recker, J. (2010). Activity labeling in process modeling: Empirical insights and recommendations. *Information Systems*, 35(4), 467–482. doi:10.1016/j.is.2009.03.009

Moody, D. (2005). Theoretical and Practical Issues in Evaluating the Quality of Conceptual Models. *Data & Knowledge Engineering*, 55(3), 243–276. doi:10.1016/j.datak.2004.12.005

de Oca, I. M. M., Snoeck, M., Reijers, H. A., & Rodríguez-Morffi, A. (2015). A systematic literature review of studies on business process modeling quality. *Information and Software Technology*, 58, 187–205. doi:10.1016/j.infsof.2014.07.011

Price, R., & Shanks, G. (2005). A semiotic information quality framework: Development and comparative analysis. *Journal of Information Technology*, 20(2), 88–102. doi:10.1057/palgrave.jit.2000038

KEY TERMS AND DEFINITIONS

Analysis Model: A model developed to learn all aspects of a problem domain to determine the best way to solve a specific set of user needs.

Conceptual Model: Models in some (semi-) formal visual modeling language describing some area of interest.

Design Model: A model developed to represent the optimal technical solution of a specified user need (as represented in a requirements model).

Enterprise Model: A model with similar abstraction level as an analysis model meant to capture the totality of relevant aspects for enterprise development.

Model: An abstraction represented in a modeling language.

Modeling Language: A language (i.e. a set of symbols, and rules for how to combine these symbols) to represent knowledge.

Requirements Model: A model to represent the external requirement to a system without taking into account how the system looks inside.

Visual Modeling Language: A diagrammatic modeling language, i.e. where the models made in the language are 2-dimensional diagrams.



Utilizing Information Science and Technology in Franchise Organizations

Ye-Sho Chen

Louisiana State University, USA

INTRODUCTION

Franchising involves with granting and receiving business rights. The one granting the business rights is called the *franchisor* and the one receiving the right to operate in accordance with the rules is called the *franchisee* (Justis & Judd, 2002). Information technology (IT) has been widely used in today's businesses. In his best seller, *Business @ the Speed of Thought*, Bill Gates (1999) wrote: "Information Technology and business are becoming inextricably interwoven. I don't think anybody can talk meaningfully about one without talking about the other." Thus, to see how IT is used in franchising (Repack & Repack, 2010), one needs to know how franchising really works. The objective of this paper is to propose an attention-based IT infrastructure that is grounded in the information science of cultivating the relationship building between the franchisors and their franchisees which will ultimately lead to the success of the franchise organizations.

BACKGROUND

In addition to the popular growth strategy for many businesses, franchising has emerged over the years as a pathway to wealth creation for entrepreneurs (Justis & Vincent, 2001). This paper first discusses the information science of franchising, including the day-to-day operations at both the franchisor headquarters and the franchisee outlets; the franchisor/franchisee relationship and the essential indicators needed to pertain and flourish the good relationship; and the inevita-

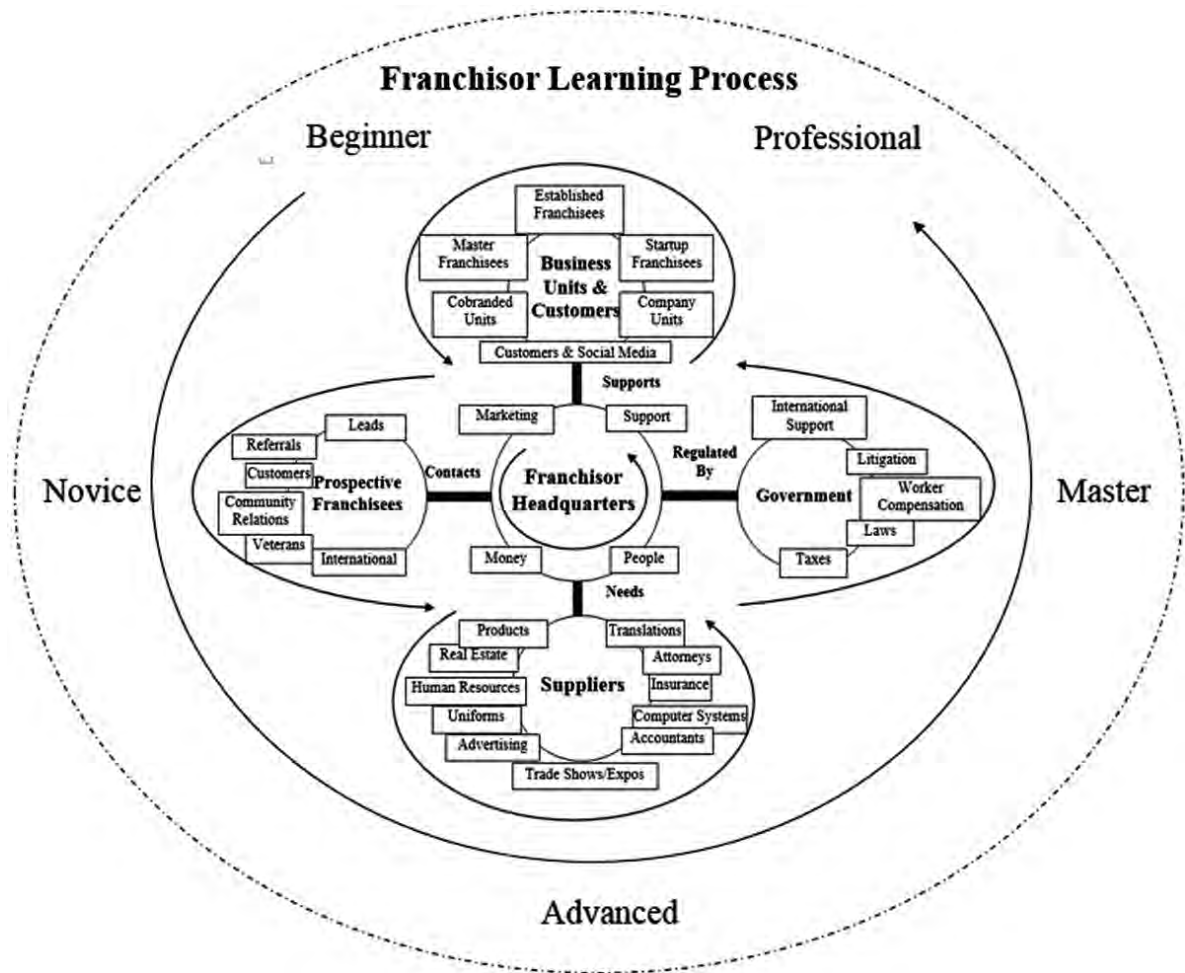
bility of collaborative learning and innovation, which leads us to the discussion of the working knowledge development among the franchisor and the fellow franchisees (Dickey, 2003). Second, we discuss that the proposed attention-based IT infrastructure will enable the knowledge sharing and dissemination between the franchisor and the franchisee (Dixon & Quinn, 2004); and suggest outsourcing the initial architectural stages of the IT infrastructure to trusted applications service providers.

UNDERSTANDING THE FRANCHISOR

In this section we examine the day-to-day operational activities at the franchisor headquarters. Figure 1 illustrates the interactions of the franchisor with all four of its entities: business units & customers, prospective franchisees, suppliers, and government; as well as performing relevant activities (represented by rectangles): marketing its products and services, assisting in creating distinguished brand names indispensable in attracting new customers, selling to the franchisees, and handling the diversified financing quandaries.

The franchisor headquarters is required to provide both initial and ongoing support/service to all business units (Lindblom & Tikkanen, 2010). Business units & customers here include company units, all of the start-up, established and mastered franchisees, the co-branded units, and customers & social media. Among the six different types of business units & customers, the franchisor needs to have intense concentration on supporting the

Figure 1. Understanding the Franchisor Learning Process



start-up franchisees, since a good start is as efficient as the half way completion of any task. On the other hand, established and mastered franchisees are the ones in need of appealing incentives (e.g., having cobranded units) in order to encourage growth and expansion. Company units are typically used as role models for the franchisees. To expand the business, the franchisor ought to select and contact the prospective partners (franchisees). The partner selection process is crucial to the success of franchising and requires exceptional attention. Prospective franchisees can be contacted through: leads, referrals, consumers, community and media, public services, and international contacts. Franchise suppliers can be anywhere from products and

goods distributors up to business service providers, such as real estate agents, human resources providers, uniform vendors, marketing and advertising agents, trade show and exposition organizers, accountants, information systems vendors, insurance providers, attorneys, translators, and many others. Franchisors also need to comply with regulations that govern the sales of the franchises and business transactions in the places where the business located. The overall legal landscape of franchising is complex which includes: federal, state, and international taxes; local, regional, and global laws; insurances, such as workers compensation; possibilities of litigations from government, customers, and franchisees; and supports for international expansion.

As is seen in Figure 1, the franchisor interacts with various business entities and conducts various deeds to fulfil his/her obligations. Moreover, the franchisor goes through a learning process, composed of five stages (Justis & Judd, 2002): (1) Beginner – learning how to do it; (2) Novice – practicing doing it; (3) Advanced – doing it; (4) Master – teaching others to do it; and (5) Professional – becoming the best that you can be. Once attaining the advanced stages of development, most preceding struggles have been overcome. However, further convoluted and challenging enquiries will arise as the franchise continues the expansion. This is especially factual once the system reaches the “Professional” stage, where various unpredicted and intricate problems could arise. Bud Hadfield (1995), the founder of Kwik Kopy franchise and the International Center of Entrepreneurial Development, explained it the finest: “The more the company grows, the more it will be tested.” To capture the franchisor learning process in Figure 1, a counter-clockwise round arrow is used in describing the five stages. It depicts the increasing intensity of learning in every stage as the franchise struggles to survive and thrive. In addition, a dashed circle around the franchisor learning process denotes that the learning process is without limits. For example, as the system expands, the real estate sub-activity will become much more complex, since the quandary of territory encroachment becomes more significant and harder to manage.

UNDERSTANDING THE FRANCHISEE

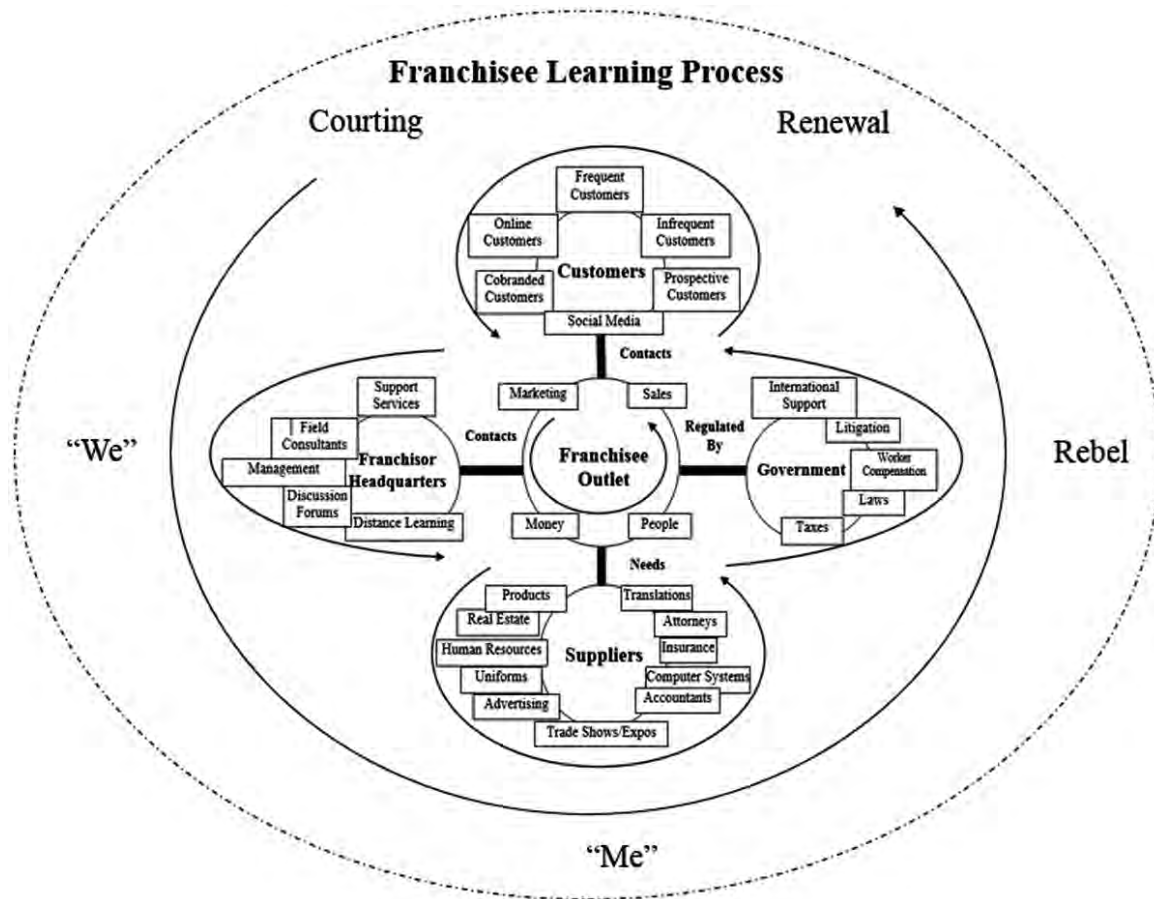
As illustrated in Figure 2, the franchisee sells to customers, perform marketing and advertising activities, handle financial/accounting issues, and manage sales people (Paswan & Wittmann, 2009). Akin to the franchisor, the franchisee outlets (shortened form, franchisee) work together with four entities: customers, franchisor headquarters, suppliers and government. The franchisee custom-

ers are divided into six categories and consist of potential, infrequent, frequent, online, co-branded, and social media. Supports from the franchisor headquarter may include demonstrations from field representatives, training and continued education from the management groups, in addition to discussion forums and distance learning. Suppliers of the franchisee are similar to those of the franchisor. They include both products and goods distributors, business service providers such as real estate agents, human resources providers, uniform vendors, marketing and advertising agents, trade shows and exposition organizers, accountants, information systems vendors, insurance providers, attorneys, translators, and others. The franchisee is regulated by the franchising laws at local, state, and federal level. The regulatory framework contains of: federal, state, and international taxes; local, regional, and global laws; insurances such as workers compensation; possibilities of litigations from government, customers, and franchisees; and support for international expansions.

Developing a high-quality “family” relationship between the franchisor and the franchisee is believed to be the most significant factor for the success of a franchise (Justis & Judd, 2002). To understand how the relationship is developed, one needs to know the franchisee learning process as depicted in Figure 2 (Schreuder, Krige, & Parker, 2000):

- **The Courting Phase:** Both the franchisee and the franchisor are eager with the relationship. This typically corresponds to the Beginner stage of the franchisee.
- **The “We” Phase:** The relationship starts to deteriorate, but the franchisee still values the relationship. This typically corresponds to the Novice stage of the franchisee.
- **The “Me” Phase:** The franchisee starts to question the reasons for payments to their franchisors. They may start to think that the success so far is purely of his/her own work. This typically corresponds to the advanced stage of the franchisee.

Figure 2. Understanding the Franchisee Learning Process

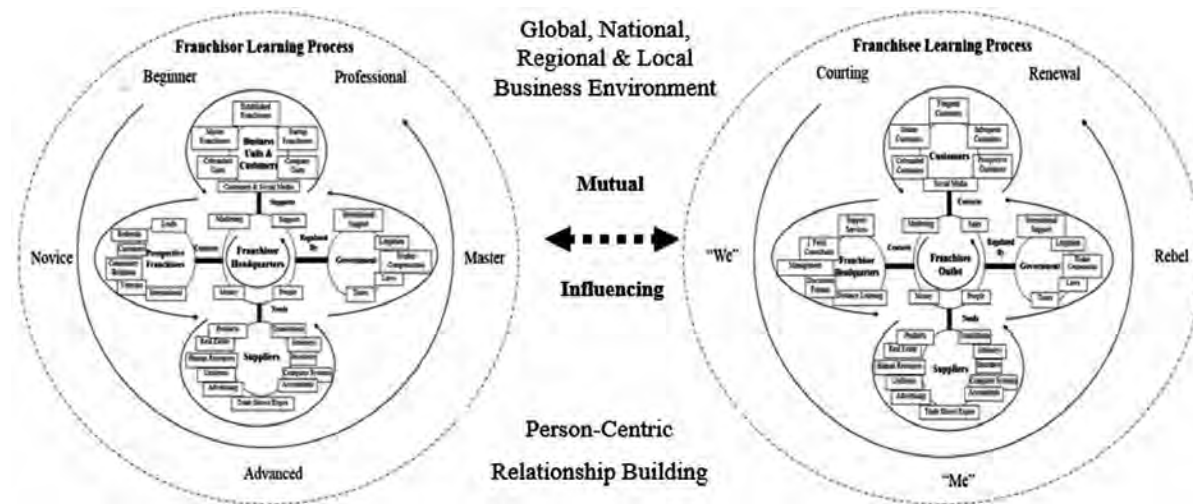


- **The Rebel Phase:** The franchisee starts to challenge the restrictions being placed upon. This typically corresponds to the Master stage of the franchisee. The rebel ones tend to be those who know the system very well and are capable of influencing others to follow them.
- **The Renewal Phase:** The franchisee realizes that the “win-win” solution is to continue teaming up with the franchisor to develop the system. This typically corresponds to the Professional stage of the franchisee.

Thus, the major challenge for the franchisor is to turn a “Rebel” franchisee into the “Renewal” one. A viable solution for the franchisor is to provide a learning and innovative environment

where the franchisee can continue contributing to the growth of the system. Successful collaborative learning and innovations will provide a strong incentive for the Professional franchisees to continue their Renewal relationship with the firm, which in turn will have positive impact on maintaining the good relationship with other franchisees and recruiting new ones. On the other hand, constant failures in this stage will intensify the Rebel franchisees to desert the firm, which in some cases lead to the demise of the franchise. To capture the franchisee learning process in Figure 2, a counter-clockwise round arrow is used in describing the five stages. It depicts the increasing intensity of learning in every stage as the franchisee struggles to survive and thrive. In addition, a dashed circle around

Figure 3. Understanding the Franchisor/Franchisee “Family” Relationship



the franchisee learning process denotes that the learning process is without limits.

DEVELOPING THE FRANCHISOR/FRANCHISEE FAMILY RELATIONSHIP

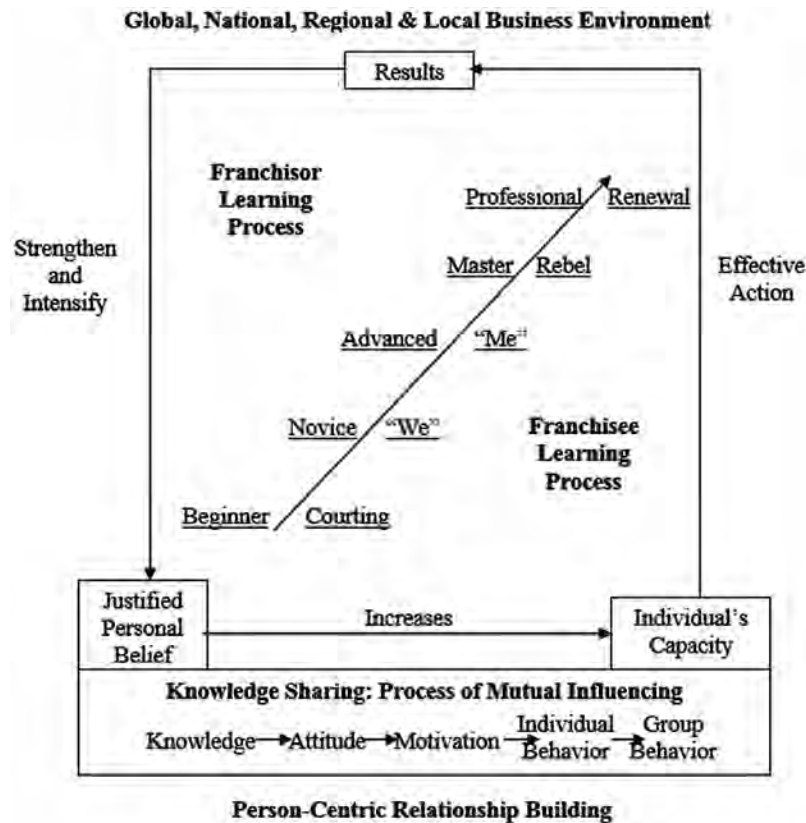
As the franchisor/franchisee progresses through the learning process, the good relationship gradually develops a mutual influencing process (Justis & Vincent, 2001), depicted in Figure 3 with a bi-directional arrow. In the figure, it shows that the franchise system is operated in the dynamic business environment of global, national, regional, and local communities. The “family” relationship is developed through a mutual learning process of person-centric relationship building. By going through the processes of learning and influencing, both the franchisor and the franchisee gain the progressive working knowledge of relationship management with the consumers and suppliers. The franchisor and the franchisee in Figure 3 are surrounded with dashed lines, indicating that there is no limit to the learning process.

Figure 3 shows the actual usage of IT in the franchise business. For instances, ERP-systems are used in the value chain in franchise business,

including Point-of-Sales systems (Tang & Lee, 2008) used at the store-units to manage the tasks of consumer relationships; and Electronic-Data-Exchange systems used in the supply chain to manage the tasks of supplier relationships (Jun & Xi, 2009). Figure 3 also provides a good impression over IT used in the franchise business and ultimately a good starting point for those interesting in starting a franchise business. Figure 3 also shows how the actual usage of IT in the franchise business differs from the ones used in other industries such as banks or car producers. That is, IT in franchising focuses more on managing and perfecting the franchisor-franchisee relationship through mutual learning process.

Lying behind the successful collaboration is the working knowledge of the franchise firm. Knowledge is defended as “a justified personal belief that increases an individual’s capacity to take effective action.” (Alavi & Leidner, 1999). Knowledge becomes “working” when the action produces results. When knowledge produces results, the personal confidence strengthens, intensifies, and justifies. The more superior the individual’s capacity becomes, the better results are attained. This spiral-up cycle of working knowledge development is very important in the context of franchising (Paswan, Wittmann, &

Figure 4. Developing the Franchisor/Franchisee “Family” Relationship



Young, 2004). Figure 4 shows that the development process is incrementally developed through the five stages of the spiral-up learning process for the franchisor and the franchisee respectively.

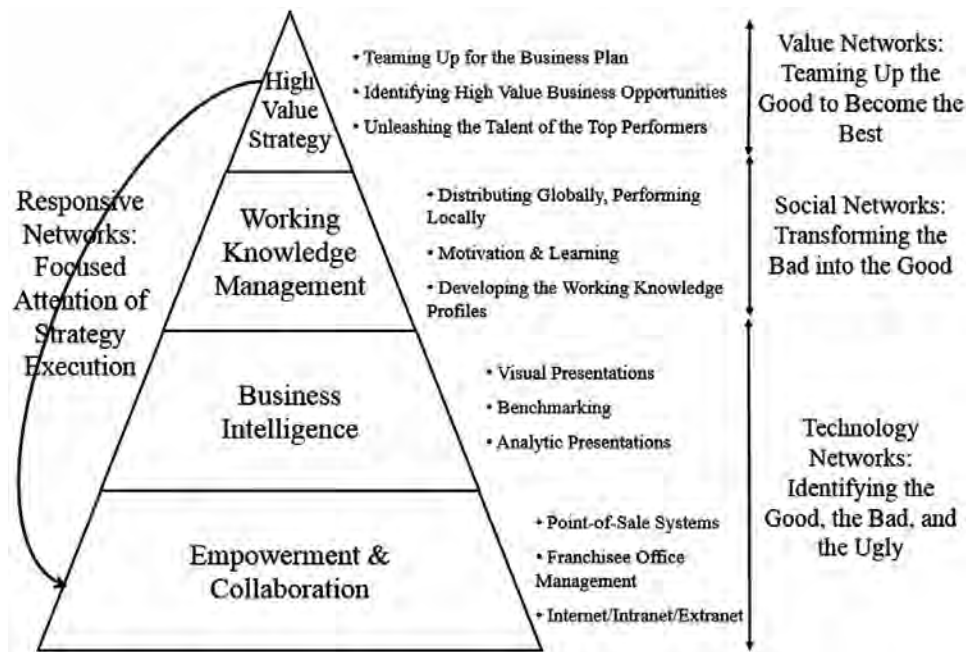
The foundation of the learning process is the capability of sharing and coaching the working knowledge throughout the franchise system (Windsperger & Gorovaia, 2011). The process of mutual influencing for knowledge dissemination consists of five steps (Justis & Vincent, 2001): (1) Knowledge, the proven abilities to solve problems in the franchise environment; (2) Attitude, constructive ways of presenting and sharing the working knowledge; (3) Motivation, incentives for learning or teaching the working knowledge; (4) Individual Behavior, the strengths of the participants to learn and enhance the working knowledge; and (5) Group Behavior, collaborative ways to create, disseminate, and manage the hard-earned working knowledge. The franchisor/franchisee

“family” relationship building in Figure 4 is also surrounded with dashed line, denoting that the relationship is enlarged and expanded without limits as the franchisee incrementally learns the working knowledge through the influencing of the franchisor and the fellow franchisees. By going through the processes of learning and influencing, both the franchisor and the franchisee progressively gain the working knowledge and manage the franchisor/franchisee family relationship efficiently and effectively.

AN ATTENTION-BASED IT INFRASTRUCTURE IN FRANCHISE ORGANIZATIONS

In an information-rich world, Herbert Simon (Nobel laureate in Economics in 1978) wrote: “a wealth of information creates a poverty of atten-

Figure 5. An Attention-based IT Infrastructure in Franchise Organizations



tion.” (Simon, 1971) As such, the “proper aim of a management information system is not to bring the manager all the information he needs, but to recognize the manager’s environment of information so as to reduce the amount of time he must devote to receiving it.” (Simon, 1971) From the discussions above regarding the information science of franchising, it is obvious that an attention-based IT infrastructure in franchising shall be one devoted to enabling the building of a good “family” relationship between the franchisor and the franchisee. In Figure 5, we propose such IT architecture (Chen, Chen, & Wu, 2005, 2007) for franchise organizations to manage the immense information produced by the firms.

The architecture, adapted from Gates’ Digital Nervous Systems (1999), consists of the following four layers:

- **Empowerment and Collaboration:** Point-of-sale and office management systems are used to empower the franchisees (Chen, Justis, & Chong, 2002). Internets, Extranets, and Intranets networking the franchisor, the franchisees, customers and

suppliers are deployed to improve collaboration (Chen, Chong, & Justis, 2002).

- **Business Intelligence:** Data warehousing and data mining techniques are used to convert volumes of data into reports and benchmarks for management to glean business intelligence (Chen, Justis, & Watson, 2000; Chen, Justis, & Chong, 2008; Chen, Zhang, & Justis, 2005). Both Empowerment and Collaboration and Business Intelligence are enabled by technology networks with the objective of identifying the good, the bad, and the ugly in the system to gain the attention of the decision makers (Chen, Liu, Zeng, & Azevedo, 2012).
- **Working Knowledge Management:** Intranet-based systems consisting of the skills of coaching/influencing others and working knowledge profiles (Chen, Hammerstein, & Justis, 2002) are implemented for knowledge sharing and learning within the franchise business. A Distance-Learning curriculum (Chen, Chong, & Justis, 2000) of working knowledge modules can also be deployed. The working knowledge management is

empowered by social networks with the objective of transforming the bad to the good (Chen, Seidman, & Justis, 2005).

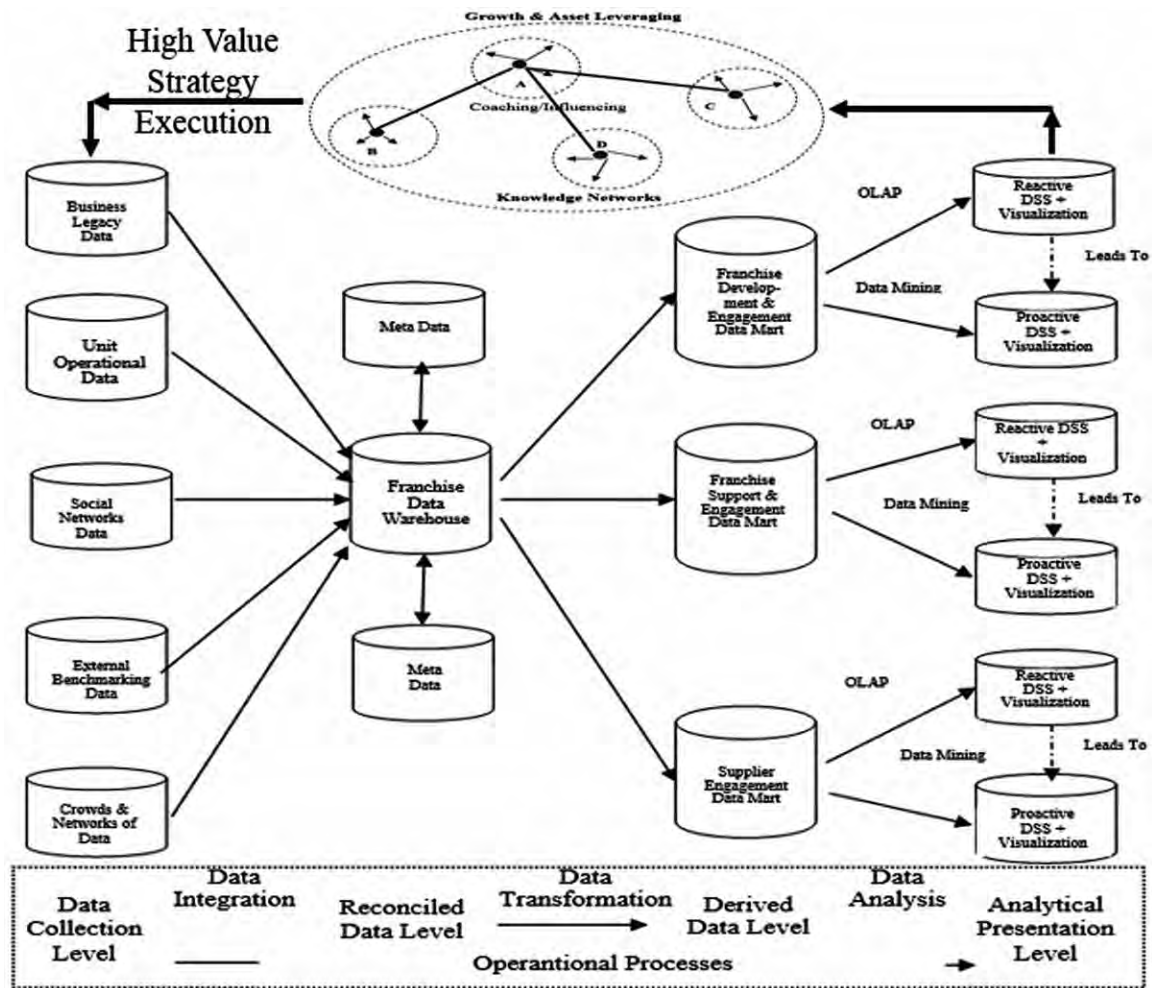
- **High Value Strategy:** Value networks (Chen, Justis, & Wu, 2006) are developed with the goal of improving working knowledge profiles and applying effectual reasoning principles (Chen, Yuan, & Dai, 2004) to create high value strategy for business opportunities (Chen, Justis, & Yang, 2004).
- **Strategy Execution:** The new strategy is executed by empowering decision makers in their collaborations (Chen, Justis, & Wu, 2008) via the responsive networks.

To demonstrate how the attention-based IT infrastructure is related to the activities of the franchisor, the franchisee (shown in Figures 1 and 2), and the “family” relationship building (shown in Figures 3 and 4), we show in Figure 6 the character of the business process, the information flow in the franchise business, and clarify the rationale behind the necessity of the attention-based IT infrastructure. The foundation of the architecture, adapted from Inmon (1996), consists of four levels: (1) data collected from the empowerment and collaboration activities of the franchisor (Figure 1) and the franchisee (Figure 2), including business legacy data, units operational data, social networks data, external benchmarking data, crowds & networks of data; e-business strategy shall be one empowering the franchisor and the franchisees to do their activities (Chen, Chong, & Justis, 2002); (2) reconciled data in the franchise data warehouse (Chen, Justis, & Watson, 2000), including franchise data warehouse and meta data; (3) derived data residing in data marts based on various franchisee-centred segmentations such as franchise development and support; and (4) business intelligence presentations and reports generated from analytical data analysis; for example, management reporting and benchmarking could be produced by online analytical processing (OLAP) periodically which periodically may lead

to more proactive analysis of top/low performer attributes using the data mining techniques (Chen, Justis, & Chong, 2008).

The business intelligence reports will help the franchise system to identify key top performers and their success attributes. With their involvement, working knowledge profiles can be developed and dissimilated throughout the knowledge networks of the business using coaching/influencing techniques (Chen, Hammerstein, & Justis, 2002). The top of Figure 5 depicts such a knowledge sharing and dissemination idea described in Figure 3 in detail. There are four franchisees (A - D) in the figure. Each franchisee (a dot) has his/her personal knowledge network (arrows pointing out of the dot). The knowledge network may include the customers' likes and dislikes, the kind of employees to hire, the competitors' and suppliers' pricing strategies, the social needs in the local community. Each franchisee is surrounded with a circle with dashed lines, meaning there is no limit to the personal knowledge network. Supposing franchisee A is the top performer and is charged and rewarded with coaching/influencing other franchisees to survive and thrive. Thus, clusters (connected dots) of knowledge network are formed and surrounded with a circle with dashed lines, meaning there is no limit for improving and leveraging the assets of the business (Chen, Justis, & Yang, 2004). Using the attention-based IT architecture as a guide, we recommend franchise organizations to outsource IT in the first two layers to application service providers (ASPs) (Chen, Ford, Justis, & Chong, 2001). The concept of subscribing IT services through ASPs has special appeal in the franchising industry because an ASP can duplicate success for former similar franchises quickly and economically.

Figure 6. Implementing the Attention-based IT Infrastructure in Franchising



FUTURE RESEARCH DIRECTIONS

With the popularity of using social media to develop social networking for growth (Perrigot & Pénard, 2013), one may find amount of data available on social media sites such as Facebook and Twitter. The need for analysing the “big data” to identify meaningful patterns for useful decision making (He et.al, 2013) is becoming a necessity, especially for those franchises interested in going abroad, e.g., to emerging markets such as BRICS (Brazil, Russia, India, China, and South Africa). Developing soft landings curricu-

lum (Chen, Watson, Liu, Cornachione, and Wu, 2010; Chen, Watson, and Azevedo, 2011) using the resource-based strategy (Chen, Watson, and Azevedo, 2013; Chen, Watson, Cornacchione, and Azevedo, 2013) is a smart strategy for franchising in emerging markets.

CONCLUSION

Franchising has been popular as a growth strategy for small businesses; it is even more so in today’s global and e-commerce world (Chen & Wu, 2007).

Although IT is quite important in franchising, IT researchers have largely ignored this arena. One major reason is that few IT researchers are vested in the knowledge of how franchising functions, and without this intimate knowledge it is difficult to implement effective IT systems. Based on years of academic and consulting experiences in franchising and IT, an effective framework of IT in franchising is proposed. At the heart of the framework is to manage working knowledge and use IT to build up the “family” relationship between the franchisor and the franchisee.

REFERENCES

- Alavi, M., & Leidner, D. E. (1999). Knowledge management systems: Issues, challenges, and benefits. *Communications of the Association for Information Systems*, 1(7), 1.
- Chen, Y., Chen, G., & Wu, S. (2005). Issues and opportunities in e-business research: A Simonian perspective. *International Journal of E-Business Research*, 1(1), 37–53. doi:10.4018/jebr.2005010103
- Chen, Y., Chen, G., & Wu, S. (2007). A Simonian approach to e-business research: A study in netchising. In I. Lee (Ed.), *E-business innovation and process management* (pp. 133–164). Hershey, PA: Idea Group. doi:10.4018/978-1-59904-277-0.ch007
- Chen, Y., Chong, P., & Justis, R. T. (2000). Franchising knowledge repository: A structure for learning organizations. *Proceedings of the 14th Annual International Society of Franchising Conference*.
- Chen, Y., Chong, P. P., & Justis, R. T. (2002). E-business strategy in franchising: A customer-service-life-cycle approach. *Proceedings of the 16th Annual International Society of Franchising Conference*.
- Chen, Y., Ford, C., Justis, R. T., & Chong, P. (2001). Application service providers (ASP) in franchising: Opportunities and issues. *Proceedings of the 15th Annual International Society of Franchising Conference*.
- Chen, Y., Hammerstein, S., & Justis, R. T. (2002). Knowledge, learning, and capabilities in franchise organizations. *Proceedings of the 3rd European Conference on Organizational Knowledge, Learning, and Capabilities*.
- Chen, Y., Justis, R., & Watson, E. (2000). Web-enabled data warehousing. In M. Shaw, R. Blanning, T. Strader, & A. Whinston (Eds.), *Handbook of electronic commerce* (pp. 501–520). Springer-Verlag. doi:10.1007/978-3-642-58327-8_24
- Chen, Y., Justis, R., & Wu, S. (2006). Value networks in franchise organizations: A study in the senior care industry. *Proceedings of the 20th Annual International Society of Franchising Conference*.
- Chen, Y., Justis, R., & Wu, S. (2008). Skew distributions and the sizes of franchise firms: A study in the senior care industry. *Proceedings of the 22nd Annual International Society of Franchising Conference*.
- Chen, Y., Justis, R. T., & Chong, P. P. (2002). Franchising and information technology: A framework. In S. Burgess (Ed.), *Managing information technology in small business: Challenges and solutions* (pp. 118–139). Hershey, PA: Idea Group Publishing.
- Chen, Y., Justis, R. T., & Chong, P. P. (2008). Data mining in franchise organizations. In J. Wang (Ed.), *Data warehousing and mining: Concepts, methodologies, tools, and applications* (pp. 2722–2733). Hershey, PA: Idea Group Publishing. doi:10.4018/978-1-59904-951-9.ch169

- Chen, Y., Justis, R. T., & Yang, H. L. (2004). Global e-business, international franchising, and theory of netchising: A research alliance of east and west. *Proceedings of the 18th Annual International Society of Franchising Conference*.
- Chen, Y., Liu, C., Zeng, Q., & Azevedo, R. F. L. (2012). E-business strategy in franchise relationship management. *International Journal of Strategic Technology and Applications*, 3(1), 55–67.
- Chen, Y., Seidman, W., & Justis, R. (2005). Strategy and docility in franchise organizations. *Proceedings of the 19th Annual International Society of Franchising Conference*.
- Chen, Y., Watson, E., & Azevedo, R. F. L. (2011). Soft Landings Curriculum of U.S.-China Entrepreneurship. *China Currents*, 10(2).
- Chen, Y., Watson, E., Cornacchione, E., & Azevedo, R. F. L. (2013). “Flying High, Landing Soft”: An Innovative Entrepreneurial Curriculum for Chinese SMEs Going Abroad. *Journal of Chinese Entrepreneurship*.
- Chen, Y., Watson, E., Liu, C., Cornacchione, E., & Wu, S. (2010). *Soft Landing Curriculum of Entrepreneurship in Emerging Markets*. Academy of International Business 2010 Annual Meeting, Rio de Janeiro, Brazil.
- Chen, Y. & Wu, S. (2007). E-business research in franchising (Invited editorial preface). *Information Management*, 20(1/2), 9.
- Chen, Y., Yuan, W., & Dai, W. (2004). Strategy and nearly decomposable systems: A study in franchise organizations. *International Symposium on “IT/IS Issues in Asia-Pacific Region, Co-sponsored by ICIS-2004*.
- Chen, Y., Zhang, B., & Justis, R. T. (2005). Data mining in franchise organizations. *Encyclopaedia of Information Science and Technology*, (pp. 714–722). Academic Press.
- Chen, Y. S., Watson, E., & Azevedo, R. F. (2013). E-Strategy and Soft Landings for Franchising in Emerging Markets. *Trends in E-Business, E-Services, and E-Commerce: Impact of Technology on Goods, Services, and Business Transactions: Impact of Technology on Goods, Services, and Business Transactions*, 148.
- Dickey, M. H. (2003). The effect of electronic communication among franchisees on franchisee compliance. *Journal of Marketing Channels*, 10(3–4), 111–132. doi:10.1300/J049v10n03_06
- Dixon, H., & Quinn, B. (2004). Franchising and the Internet: An exploratory study of franchisor web sites. *Internet Research*, 14(4), 311–322. doi:10.1108/10662240410555324
- Gates, W. (1999). *Business @ the speed of thought*. Warner Books.
- Hadfield, B. (1995). *Wealth within reach*. Cypress Publishing.
- He, W., Zha, S., & Li, L. (2013). Social media competitive analysis and text mining: A case study in the pizza industry. *International Journal of Information Management*, 33(3), 464–472. doi:10.1016/j.ijinfomgt.2013.01.001
- Inmon, W. H. (1996). *Building the data warehouse*. John Wiley & Sons.
- Jun, W., & Xi, Z. (2009). Research on logistic distribution system of retail chain. *Proceedings of the 16th International Conference on Industrial Engineering and Engineering Management* (pp. 618–621), IEEE Explore. doi:10.1109/ICIEEM.2009.5344376
- Justis, R. T., & Judd, R. J. (2003). *Franchising* (3rd ed.). DAME Publishing.
- Justis, R. T., & Vincent, W. S. (2001). *Achieving wealth through franchising*. Adams Media Corporation.

Lindblom, A., & Tikkanen, H. (2010). Knowledge creation and business format franchising. *Management Decision*, 48(2), 179–188. doi:10.1108/00251741011022563

Paswan, A. K., & Wittmann, C. M. (2009). Knowledge management and franchise systems. *Industrial Marketing Management*, 38(2), 173–180. doi:10.1016/j.indmarman.2008.12.005

Paswan, A. K., Wittmann, C. M., & Young, J. A. (2004). Intra, extra, and internets in franchise network organizations. *Journal of Business-To-Business Marketing*, 11(1-2), 103–129. doi:10.1300/J033v11n01_07

Perrigot, R., & Pénard, T. (2013). Determinants of e-commerce strategy in franchising: A resource-based view. *International Journal of Electronic Commerce*, 17(3), 109–130. doi:10.2753/JEC1086-4415170305

Repack, D. L., & Repack, W. F. (2010). A case study of the adoption and use of information systems technology in a business-format franchise. *Information Resources Management Journal*, 23(1), 17–34. doi:10.4018/irmj.2010102602

Schreuder, A. N., Krige, L., & Parker, E. (2000). The franchisee lifecycle concept—A new paradigm in managing the franchisee-franchisor relationship. *Proceedings of the 14th annual International Society of Franchising Conference*.

Simon, H. A. (1971). Designing organizations for an information rich world. In M. Greeberger (Ed.), *Computers, communications, and the public interest* (pp. 38-52). Baltimore, MD: The Johns Hopkins Press.

Tang, L. L., & Lee, C. M. (2008). The Business performance of convenience stores installing POS systems: An empirical study of 7-Eleven. *International Journal of Integrated Supply*, 4(1), 60–74. doi:10.1504/IJISM.2008.016068

Windsperger, J., & Gorovaia, N. (2011). Knowledge attributes and the choice of knowledge transfer mechanism in networks: The case of franchising. *Journal of Management and Governance*, 15(4), 617–640. doi:10.1007/s10997-009-9126-5

ADDITIONAL READING

Burgess, S. (Ed.). (2002). *Managing information technology in small business: Challenges and solutions*. Hershey, PA: Idea Group Publishing. doi:10.4018/978-1-93070-835-8

Chen, C. Y., Lee, W. I., Kuo, H. M., Chen, C. W., & Chen, K. H. (2010). The study of a forecasting sales model for fresh food. *Expert Systems with Applications*, 37(12), 7696–7702. doi:10.1016/j.eswa.2010.04.072

Chen, S. Y., & Liu, X. (2004). The contribution of data mining to information science. *Journal of Information Science*, 30(6), 550–558. doi:10.1177/0165551504047928

Cho, W., & Olsen, M. D. (1998). A case study approach to understanding the impact of information technology on competitive advantage in the lodging industry. *Journal of Hospitality & Tourism Research* (Washington, D.C.), 22(4), 376–394. doi:10.1177/109634809802200404

Dant, R. P., & Nasr, N. I. (1998). Control techniques and upward flow of information in franchising in distant markets: Conceptualization and preliminary evidence. *Journal of Business Venturing*, 13(1), 3–28. doi:10.1016/S0883-9026(97)00040-2

Duan, Y., Mullins, R., & Hamblin, D. (2001). Training for e-commerce success in SMEs. *Information Technologies in SMEs: Challenges and Solutions*.



- Ernst, D. (2005). The new mobility of knowledge: digital information systems and global flagship networks. *Digital formations: it and new architectures in the global realm*, 89-114.
- Gharavi, H., Love, P., & Cheng, E. (2004). Information and communication technology in the stockbroking industry: An evolutionary approach to the diffusion of innovation. *Industrial Management & Data Systems*, 104(9), 756-765. doi:10.1108/02635570410567748
- Gómez, R. S., González, I. S., & Vázquez, L. (2010). Multi-unit versus single-unit franchising: Assessing why franchisors use different ownership strategies. *Service Industries Journal*, 30(3), 463-476. doi:10.1080/02642060802252027
- Kaushik, P. D., & Singh, N. (2004). Information technology and broad-based development: Preliminary lessons from North India. *World Development*, 32(4), 591-607. doi:10.1016/j.worlddev.2003.11.002
- Kumar, J. A., & Ganesh, L. S. (2009). Research on knowledge transfer in organizations: A morphology. *Journal of Knowledge Management*, 13(4), 161-174. doi:10.1108/13673270910971905
- Li, L. (2012). Effects of enterprise technology on supply chain collaboration: Analysis of China-linked supply chain. *Enterprise Information Systems*, 6(1), 55-77. doi:10.1080/17517575.2011.639904
- Manohar, B. M. (2005). Information and communication technology applications in development: India as a role model for other developing countries. *Information development*, 21(1), 47-52.
- Michael, S. C. (2000). The effect of organizational form on quality: The case of franchising. *Journal of Economic Behavior & Organization*, 43(3), 295-318. doi:10.1016/S0167-2681(00)00125-6
- Michael, S. C. (2003). First mover advantage through franchising. *Journal of Business Venturing*, 18(1), 61-80. doi:10.1016/S0883-9026(01)00085-4
- Mitsuhashi, H., Shane, S., & Sine, W. D. (2008). Organization governance form in franchising: Efficient contracting or organizational momentum? *Strategic Management Journal*, 29(10), 1127-1136. doi:10.1002/smj.702
- Perry, C., Cavaye, A., & Coote, L. (2002). Technical and social bonds within business-to-business relationships. *Journal of Business and Industrial Marketing*, 17(1), 75-88. doi:10.1108/08858620210415217
- Pirkul, H., Narasimhan, S., & De, P. (1987). Firm expansion through franchising: A model and solution procedures. *Decision Sciences*, 18(4), 631-645. doi:10.1111/j.1540-5915.1987.tb01551.x
- Quanjun, X. (2003). Exploring a New Theory: Self-organization Theory of the Firm. *Nankai Business Review*, 3, 012.
- Sen, K. C. (2001). Information asymmetry and the franchise decision. *Journal of Marketing Channels*, 8(1-2), 91-109. doi:10.1300/J049v08n01_05
- Sorenson, O., & Sørensen, J. B. (2001). Finding the right mix: Franchising, organizational learning, and chain performance. *Strategic Management Journal*, 22(6-7), 713-724. doi:10.1002/smj.185
- Spohrer, J., Maglio, P. P., Bailey, J., & Gruhl, D. (2007). Steps toward a science of service systems. *Computer*, 40(1), 71-77. doi:10.1109/MC.2007.33
- Stanworth, J., Purdy, D., & Price, S. (2003). Franchising as a source of technology transfer to developing countries. In D. Welsh & I. Alon (Eds.), *International franchising in industrialized markets: Western and Northern Europe* (pp. 87-103). Chicago, USA: CCH Incorporated.



Tatnall, A., & Gilding, A. (2005). Actor-network theory in information systems research. *Encyclopedia of Information Science and Technology*. IGI Global, 42-46.

Wang, J. Y., & Fan, Z. P. (2005). Game Analysis of Resource Sharing Decisions in Collaborative Knowledge Creation. *Journal-Northeastern University Natural Science*, 26(10), 1017.

KEY TERMS AND DEFINITIONS

Attention-Based IT Infrastructure: An IT infrastructure which is able to sort through volumes of data and produce right information at the right time for the right persons to consume.

Franchisee Life Cycle: The stages a franchisee goes through in the franchise system: Courting, “We”, “Me”, Rebel, Renewal.

Franchisee: The individual or business who receives the business rights and pay the royalties for using the rights.

Franchising: A business opportunity based on granting the business rights and collecting royalties in return.

Franchisor/Franchisee Learning Process: The stages of learning, including Beginner, Novice, Advanced, Master, and Professional.

Franchisor/Franchisee Relationship Management: The vital factor for the success of a franchise, including: Knowledge, Attitude, Motivation, Individual Behavior, and Group Behavior.

Franchisor: The individual or business who grants the business rights.

Category I

IT Security and Ethics

Computer Fraud Challenges and Its Legal Implications



Amber A. Smith-Ditizio
Texas Woman's University, USA

Alan D. Smith
Robert Morris University, USA

INTRODUCTION

Computer Fraud

Computer fraud and hacking attempts have been publicized for more than a century. Although customers only think of computers and smart-phones being hacked, there are examples in the early 19th century where phone lines were hacked. Cybercrime is a fast growing area and has drastically increased over the years. Its business model is evolving and the market is profitable for criminals. New activities have emerged as technology advances. Traditionally, consumers and businesses were lax with security as hackers could easily encrypt and infect any technological device. Hence, cybercriminal activities grew rampant in the global economy. Security protection, government involvement, and leading software companies have become strategic partners in combating cybercriminal activities. However, despite all these efforts, cybercrime is still growing. There are many strategic solutions to this growing epidemic, such as investing in anti-virus software and commonsense approaches to password protection. In order to reduce the amount of cybercriminal activity occurring globally, action needs to be taken immediately.

The targets are computers or anything device connected to the Internet, such as tablets or smart-phones (Sundarambal, Dhivya, & Anbalagan, 2010). Hackers affect the cybersecurity of large companies, government agencies and regular

customers, especially if competitive or personal information is stolen for ransom or extortion purposes. In the majority of incidents, it is relatively simple to trace back to the hacker, as many are nonprofessionals with little experience. However, it has become increasingly difficult to catch more sophisticated hackers. Although, if and when they are caught, there are significant penalties that come with hacking and computer fraud, many have argued that these penalties are not severe enough to deter such activities (Beldona & Tsatsoulis, 2010; Mohanty, et al., 2010; Smith, 2007). Some have suggested that such crimes are inevitable as IT systems become increasingly complex and globally interconnected (Dharni, 2014; Latha & Suganthi, 2015; Chand, et al., 2015; Han, et al., 2015; Soon, et al., 2015).

Exploring Types of Computer Fraud

To illustrate these trends, Stewart and Shear of SecureWorks™ have examined many hacker markets and found that cybercriminals are increasing their activity of stealing information (Clarke, 2013). Stewart is Dell's SecureWorks™ Director of Malware Research for the Counter Threat Unit (CTU) and independent researcher Shear have done much research into the dark marketplace that is frequented by cybercriminals. There are online tutorials for novice hackers to learn the trade for under US\$1. For example, one can access Social Security card numbers, name and address of customers for US\$250. Cy-

bercriminals can gain control of computers for US\$20 to 50. Customers can also hire someone to hack a website for US\$100 to 200. However, not everything is cheaper. The price of botnets, spam and malicious software, has increased from US\$90 to US\$600-1,000. Multiple sellers advertise “satisfaction guaranteed” on the data, which is designed to capture the attention of a potential or practicing hacker. It seems the traditional business model of value-added activities works well in the more hidden and illegal markets as well. However, organizations’ drives to understand and anticipate their customers’ needs ultimately forces management to connect valuable and vulnerable corporate systems to the general public and, thus, cyberthieves (Daim, Basoglu, & Tanoglu, 2010; Daramola, Oladipupo, & Musa, 2010; Dominic, Goh, Wong, & Chen, 2010; Kapur, Gupta, Jha, & Goyal, 2010; Keramati & Behmanesh, 2010).

Dell and its software SecureWorks™ are well-known and highly respected IT-security providers. Software and management at Dell has been investigating this illegal market for some time. Dell, as a provider of security systems, has in place this particular security software in over 61 countries with 4100 clients and has been providing top of the line service for the past 16 years. Dell SecureWorks™ provides a relatively quick warning to its clients when a cyber-attack is happening. It also provides a prediction of where the cyberattack is coming from and what it is trying to get from the computer. After the security system finds a cyberattack the system then works to get rid of it and tries to prevent the cyberattack from happening again. Counter Threat Platform (CTP) powers Dell SecureWorks™. CTP analyzes a total of 150 billion networks to find any possible threats, generate information, and find any information that could lead to a cyberattack (Alderete & Gutiérrez, 2014). Overall, CTP allows Dell’s SecureWorks™ to prevent, detect, respond, and predict.

Cybercrime an industry that has shown double digit growth year after year within the current struggling global economy. Cybercriminal activities have been on the rise and are becoming more

profit driven as its business model is evolving and new cybercriminal activities are emerging. Out of dozens of underground markets surveyed by SecureWorks™ researchers, a subsidiary that specializes in cybersecurity and data protection, it was found that business is booming. Not only have prices gone down on many of the services offered, but offerings have expanded as well. From basic hacking offerings to infecting networks of computers with the use of botnets, “underground hackers are monetizing every piece of data they can steal or buy and are continually adding services so other scammers can successfully carry out online and in-person fraud” (Lawrence, 2014). Cybercriminal activity has become a major global problem.

China, for example, tops the list for online crime hotspots with 83% of respondents residing there having been victimized (Bertolucci, 2016). According to the Norton Cybercrime Report (“Cybercrime Report,” 2011), 41% do not have an updated security software suite to protect their personal and online data in the U.S. Within this study, more than 65% have claimed that they have fallen victim to viruses and malware attacks, online scams, phishing, social network hacking, credit card fraud, and sexual predation. Many people are uneasy about the safety of online commerce, as they feel that the majority of online criminals operate within countries where prosecution is unlikely. Illegal downloads, digital piracy, and digital harassment are some of the other more common cybercriminal activities committed on a daily basis (Shinder, 2011).

Cybercrime is up 10.4% over the previous year as stated by Kassner (2016). In general, cybercrime has risen over US\$1.13 million domestically in the 2014 alone. Web-based attacks, denial of services, malicious insiders, viruses, worms, Trojans, malicious code, phishing and social engineering, malware, stolen devices, and botnets are among the top cybercriminal activities have cost the U.S. and other countries the most money in terms of lost productivity, cybertheft, and counterintelligence efforts. The actual price of cybercriminal

activities paid by hacker has decreased over the years. For example, the cost for a Remote Access Trojan (RAT) has dropped considerably in recent years, on part to the increased availability of free ones now available online (Lawrence, 2014). However, the price of botnets has increased from US\$600 to 1,000 for 5,000 of botnets within the U.S. With activities expanding and services more readily available, cybercrime is an ever-expanding industry and is on the rise.

Security protection, government involvement, and leading software companies have become competitive forces for cybercriminal activities. More security protection is being added towards platforms, PCs, mobile devices, and tablets. Security product vendors are also on the rise when it comes to cloud computing. As mentioned by Kassner (2016), cloud computing affords cybercriminals perfect opportunities. However, the use of cloud-based fraud detection can be used against cybercriminals. In general, applications of cloud services can pick up on patterns of criminal activity that would not otherwise be obvious with the collection and sharing of information from the millions of devices across the global economy.

Governments have also become a competitive force towards cybercrime. The U.S., as well as the U.K., has been investing heavily in their security infrastructures. The Department of Homeland Security within the U.S. is combating cybercrime by investigating a wide range of cybercrimes and conducting high-impact investigations to disrupt and defeat cybercriminals. The U.S. Secret Service maintains many electronic crimes task forces that are known to concentrate on identifying/locating global crime connected to cybercrime, bank fraud, data breaches, to name a few ("Combating cybercrime," 2016). The U.S. Immigration and Customs Enforcement, Homeland Security Investigations, and Cyber Crimes Center, are involved in cybercrime investigations by providing computer based technical services to support both domestic and international investigations in cross-border crime. The U.K. has recently spent over 63 million pounds in building

up its resources for fighting cybercrime. Large companies, such as Microsoft, have invested their resources in efforts aimed at tracking down and prosecuting cybercriminals. Microsoft, as well as other leading companies, has teamed up with NASA and the U.S. Department of Defense to develop international standards of making IT equipment far more secure than it is today. Despite the challenges presented towards cybercriminals as many cybercriminals having been prosecuted and their illicit organizations shut down, cybercrime is still on the rise and has soared 107% percent in 2015 (Watson, 2016).

Even though Dell is a leading IT-security service provider, there are several other companies that are close behind in terms of competitive offers ("SecureWorks competitors - CB insights," 2016). FireEye™ and Intellitactics™ are the two main competitors. FireEye™ provides its users with firewall, IPS, anti-virus, and gateway protections. These different types of security protocols are designed to protect the user's computer from many threats that may appear using day-to-day systems, such as e-mail and customer-service operations and normal e-commerce offerings. FireEye™ is a public service company that originated from and initial public offering (IPO). As an IPO, initial sales to the public after being released from a private company to a public company, specifically, FireEye™ is a provider that specializes in preventing cyberattacks, such as zero-day and advanced persistent threat (APT) attacks that bypass traditional defenses and compromise over many networks. Many zero-day attacks are purposefully designed to expose computer applications' vulnerabilities; hence there has been no time to develop a proper defense against such attacks. Unfortunately, many zero-day exploits are totally unexpected and frequently used by cybercriminals for malicious purposes. FireEye™ attempts to provide solutions that supplements signature-based firewalls, IPS, anti-virus, and gateways, and provides cross-enterprise, signature-less protection against cyberattacks and malware problems on clients' files.

To add to this misery, APTs utilizes many different phases and methods to compromise a computer network in order to avoid detection by anti-cyberthreat software in order to gain access to important information through thefts of information leaked out have relatively long time periods. Both FireEye™ and Intellitactics™ have specific safeguards against such attacks, but technology advancements always work to undermine such solutions.

Intellitactics™, such provides information security software for comprehensive enterprise security management, is a competitor in a different way than FireEye™. Intellitactics™ is part of the Trustwave® corporate initiative, which is a larger security business. Intellitactics™ is more of a pinpointed security solution. Intellitactics™ is more of a security information and event management system solution. This essentially means that the security system is more used by a business that stores important information on a computer or a business that stores planned events on a computer.

BACKGROUND

Businesses Affected by Hackers

Virtually every modern and web-connected business uses computer systems to keep track of its records and keep in contact with customers in order to make transactions smooth and relatively seamless. Unfortunately, these same systems also put a company at risk; not just the company that is affected it is also their clients (Aeron, Kumar, & Janakiraman, 2010; Barra, Savage, & Tsay, 2010). Firewalls and encryptions can be put into place and properly maintained, but if a determined hacker knows what he/she is targeting; undoubtedly, it is only a matter of time before these systems are eventually compromised unless significant steps and constant monitoring processes are implemented. Some of the major companies have been hacked recently (e.g., Sony, Home Depot, and Target) know this scenario too well. For example,

Sony Pictures Entertainment was made aware of the hack on November 24, 2014, but there is a possibility that the hacker had been in the system for almost a year (Fitz-Gerald, 2014). Names, addresses and social security numbers were exposed of over 47,000 current and former employees. Of course, as conspiracy theories develop, there was a theory that it was a current or former employee that was responsible for the hack and not terrorists' organizations as it had been originally speculated. Hence, as suggested by Barrett and Yadron (2016), companies and governmental agencies must be trusting partners in the war on cyberattacks and fighting computer hackers if such security breaches that was experienced by Sony in 2014 will become commonplace as warned by U.S. President Barack Obama stated in February 2016.

The industry position for individuals or companies that are illegally selling data on the Internet, unfortunately, still practices a number of traditionally practiced business models. These underground hacking markets are constantly changing with the new technology that is allowing for cheaper options, which makes it easier for hackers to access customers' computers. For an example, one year the black market price for a RAT could be up to US\$250, and then the next year it could be down to US\$20 ("Underground Hacking Markets Report," 2014). Such inconsistent markets puts pressure on the hackers not knowing if this data will be high or low on black market, which forces hackers to innovate and more dangerous ways to infect computers on a global scale. Unfortunately, what may be at risk can be both intellectual property and personally identifiable information (PII) that ties information directly to an individual company and/or customer (Shukai, Chaudhari, & Dash, 2010).

In terms of state-sponsored cyberterrorism, North Korea is typically blamed for many widespread cyberattacks, regardless of proof of their true degree of involvement. This is part to its suspected involvement in preventing a movie, *The Interview*, from being formally released. The film, about assassination of the President of North Korea, the hackers, identifying as "The

Guardians of Peace,” tried to prevent its release through public discourse. The hackers did not leak everything at once but slowly and information is still being released to this day. The hackers promised they would stop the leaks if Sony cancelled *The Interview*’s theatrical release. Personal e-mails between company employees and actors were released, as well as confidential salaries of celebrities and top employees. Despite this pledge, significant and personally damaging information was getting released daily. Management at Sony went ahead with the movie’s release, but instead in theaters, customers could buy the movie at home because movie theaters were getting threatened if they aired the movie. It was in December, 2014, that the FBI stated that North Korea appear to be involved although the government of North Korea officially denies it and threatened the White House and pentagon if the U.S. government levies any punishments. Following the Sony hack, other studio companies took precautions and cancelled their own North Korea-themed movies. Currently, Russia is accused of interfering with the 2016 U.S. presidential elections.

Another well-publicized major incident of a company that was targeted by hackers was Target. A few days before Thanksgiving, 2014 a malware was installed on Target’s security and payment systems at POS, so that every time a customer’s card was swiped it would take the card number and store it in a separate system for the hacker to retrieve at a later date, possibly suggesting insider involvement (Riley, Elgin, Lawrence, & Matlack, 2014). On “Black Friday,” when many customers do their Christmas shopping, over 40 million customer accounts were breached. In the aftermath, over 90 lawsuits have been filed against Target from customers and bank for negligence. In March 2015, a federal judge approved an offer of 10 million from Target to settle a class action lawsuit if the customers affected can actually prove they were affected since banks usually reimburse 100% for card fraud (Riley & Pagliery, 2015).

These documented breaches on global companies that traditionally make significant profits

have spent considerable financial and human resources cybersecurity on their data and networks that support them. Obviously, such investments do not guarantee protection (Kim & Tadisina, 2010; Marthandan & Tang, 2010). But how do the small businesses successfully protect their customers’ information? Based on prior studies, over 90% of small businesses in the U.S. do not use data protection for their business or their customer’s information and, hence, this is costing business and their customers’ considerable financial and other resources (Warren, 2015). A small business can start off by making sure their employees are not visiting the wrong kind of websites and getting cybersecurity insurance. Retail and finance business are at higher risk than restaurants but they are still considered a target.

From a strategic perspective on the nature of cybercrime, the competitive forces for a company illegally selling data on the Internet are unique (Smith, 2007). The availability of automatic identification and data capture technologies (AIDC) and its relatively low cost have helped to develop the underground cybercrime industry. The cybercrime industry varies widely on the enforcement of government regulations and legality of the information exchanged (Jarrett, 2016). The barriers-to-entry are low, as anyone with enough knowledge to gain the valued information can become a seller; therefore, the threat of new entrants is high (Elysee, 2015). Buyers have a high degree of bargaining power, as there are nearly endless sources of this information available on the Web and in person. Since customer service is important in the industry, as there are many sellers and there is a need to differentiate their services. Sellers have a relatively high degree of bargaining power as well, since there is little enforcement of laws around the sale. For example, if a seller sells a “bad” ID that does not work, the buyer has little means of getting its money back and will likely not use the law to enforce a botched sale or contract.

A number of authors (Aeron, et al., 2010; Barra, et al., 2010; Shukai, et al., 2010) have discussed the available and effective security defenses that



have been developed and continue to grow in effectiveness for all of the security threats mentioned. No security defense can be completely effective, but many of these can be used to help avoid major breaches, known threats, ordinary hackers, and security ignorance. Perhaps the most obvious defense is the personal firewall, which provides antivirus and anti-spyware. Firewalls help devices by blocking or limiting the use of Bluetooth and Wi-Fi unless asked to connect. This is difficult to ensure that every device follow because many devices encourage the user to allow connection to Bluetooth for all types of traffic. Antivirus and antispyware scan network traffic for signatures of known malware. These types of mobile security are not widely used on mobile devices but will become more available due to the target malware has on the devices as technology progresses. As suggested by Friedman and Hoffman (2008), there are limitations of firewalls, antivirus and anti-spyware, that are based on their inability to block malware that goes through the port that is in use. Also, virus writers and cyber criminals have found ways to evade signature based antivirus and antispyware by using short-span attacks and designer malware. Because of this additional defenses are available as a sort of back up to the backup protection. Intrusion prevention systems examine network traffic into and out of the computers and identify sketchy behavior. This form of protection establishes a baseline of normal online behavior and is alerted when there is a significant change.

For the defense against phishing and social engineering, education may be the best defense. This includes being informed on what accurate looking e-mail appears from their banks or institutions, what phishing messages, what unknown websites may appear based on URL observing and just the knowledge that giving out personal information to someone online. Content and spam filtering are known best for maintaining a large list of known websites that are known for spamming or housing malware and it in many ways blacklists them and blocks them from incoming. Other

ways of preventing communication interception and spoofing can be accomplished by disabling unneeded communication, password encryption, virtual private networks, and to restrict access to hot spots. A user can disable these unneeded communication methods and ports by changing configuration settings on the system being used and avoid hackers from unknowingly obtaining information without the device users consent. Good password encryption will help the device carrier to use algorithms in order to protect the device and applications the user may have important information on as well.

Defense against loss and theft of devices are better known as many users already use locks on their devices to keep only them from accessing their device. Some devices have “time-bomb” capability, so if the device is lost or stolen and someone tries to obtain information from it, it will wipe it and destroy data on the device. Many devices also have back up or a recovery mode, which replaces lost or stolen files. Device control can be used to prevent files from being written to USB and requires that files be encrypted. By encrypting files, regardless of a loss or theft, files can remain confidential and are less likely to be leaked to outside sources.

CONCLUSION

Legal Solutions Are Inadequate

In conclusion, fraud is a crime that will always be relevant and present, even as the technology to protect against it grows and strengthens. Laws and regulations put in place to prohibit such actions can only do so much in combating against such threats. Many cases arise from fraudulent activities and will vary in degree of damage, but with each case is the degree of the crime, unique to the case. For every business that uses computer systems, managers must become beware of threats with the hope of a better functioning system and a rise in profit. Strategic recommendations for

security providers vary depending on the type of security you provide. For companies, like Intelitactics™, it is to provide reassurance that their customer's information will be kept safe from the unauthorized public. To do this properly, they market protective measures to ensure that hackers cannot get into stored information. For companies, like SecureWorks™ and FireEye™, where customers are working more with individuals than other businesses, it is equally important to gain their trust by providing successful documentation and cases. It is important to keep customers in constant contact and insure that if a problem is detected that they will be alerted immediately. Keeping customers updated on the innovative hacker technologies would gain customer trust by keeping them in the information loop and alert.

It is important to raise the awareness of the potential punishment for those who may think of illegally purchasing a product or service from hackers and look for a strategic alliance where they can work with another company or companies to raise the defenses against hackers. Specifically, in the late 2015, Lifetime, Netflix, Pandora, and Hulu accounts were being hacked and sold online for as low as US\$1. For the average consumer with little technical background information, it seems like a great purchase bargain, but since these accounts are stolen, there are consequences to the consumer if they are found with such account. With the proper investment into the awareness by Internet users, hackers would have no market or interest to sell stolen accounts. If such companies as SecureWorks™ were to partner with each other to raise account security standards, these accounts would be much more challenging to hackers, as they typically go after soft targets. Ultimately, as evident from the researchers cited in this chapter, a major threat to the global economy is the general lack of Internet security. With advocates for stronger defense systems like those, there should be a stronger movement towards more secure networks and acts at tracking down sites used to sell hacked goods. In essence, there was a variety of information presented in this chapter regarding

how the reader could arm himself/herself in the probability of an attack by a hacker through better education of the nature of cyberattacks.

Strategic Implications

In order to attempt to catch up to this growing epidemic, fast and decisive actions are required. Even though mobile phones and smart devices have remained relatively safe, users are vulnerable targets, as presented in the 2011 Norton Cybercrime Report. Consumers have to become more protected by installing security programs not only on their PCs, but on their mobile devices and tablets as well. Every year it is getting easier to hack computers and counterfeit items such as passports. It is also becoming more “mainstream” in the e-marketplace. The cost to hack websites, for example, has fallen 33% in the last year alone. It has become easier to be a hacker or hire a hacker than it would be to invest in advanced security protection software. In a matter of years hacking will become as easy as going to a drug-store if it is not stopped soon.

The continuous demand and growth of cybercrime has led to the development of a new industry to offset its negative impact; cybersecurity. Corporations and individuals can protect their electronic gadgets by installing anti-virus programs and file restrictions that blocks access as well as detecting and eliminating malicious ware before it expands. Anti-virus software is the most common solution to combat lower degree of cybercrime activity such as Trojans and worms. Another strategic measure that can be taken is to pay close attention to the source of e-mails and their content; if the information seems unfamiliar or the sender is unknown, the receiver should eliminate the e-mail without opening any attachment or providing any information.

Besides computer infiltration, spam and phishing, another common form of cybercrime is identity theft. This criminal activity is possible due to the amount of information that is stored in computers, social media, employment forms and



e-mails. The easiest way for criminals to obtain personal information is through social media, therefore it is better if accounts on websites such as Facebook or MySpace are accessible by friends only, and all pictures that contain addresses or telephone numbers are blurred or removed. While it might seem innocent to fill out personality tests and surveys, it is better to investigate the source of these; Facebook, for example, allows independent developers to create customized applications without verifying their identity and intentions. In many cases, identity theft can be extremely difficult to detect in the beginning since criminals tend to keep a low profile. An obvious way to fight this can be by monitoring accounts and financial information on a constant basis and to report all suspicious activity to the law enforcement activities. Companies such as Discover notify customers of transactions that takes place a specified distance from a set location, and Chase Bank, for example, asks for verification codes every time a new smartphone or computer is used for a balance inquiry.

Commonsense approaches include never use the same password for more than one account as well as the use of common information for security questions. Much of cybercrime can be stopped in the effective use of passwords and 2-step verification methods, using access codes that are texted or sent via voice mail. Social media and acquaintances may provide enough information to third parties unintentionally and might allow access to bank accounts, retirement funds, e-mails and other accounts. It is highly suggested to elaborate passwords that contain numbers, different characters and unrelated dates and names. These measures can help fight cybercrime since there is no better solution than restricting the amount of information to which someone has access.

In conclusion, to fight cybercrime it is necessary to constantly monitor accounts that involve trading or savings for any suspicious activity, observe the sources of suspicious e-mails and avoid opening files or filling formats that seem illegitimate. Restrict privacy settings in social

media networks and destroy physical evidence of important information, such as old utility bills or any addresses from envelopes, birthday dates, etc. If completing purchases online, using trusted sources like PayPal®; if that option is not available, use prepaid gift cards to keep bank or credit card account information safe. Installing and maintaining an anti-virus program can help prevent possible malware from infecting the user's computer. The use of different passwords and security questions for each account can also aide in the protection of being hacked. Companies and countries that have gone beyond the current standard security protocol have achieved great results in the prevention of cybercriminal activities.

FUTURE RESEARCH DIRECTIONS

Much of the information presented in this chapter typically applies to many daily business situations. They are particularly relevant in the era of big data analytics and the growth of identity theft, evident to the widespread nature of malicious attacks. The increasing likelihood of being a victim of cyberhacking is well documented by academics and practitioners alike. Fortunately, essentially every cyberthreat has a potential solution that is attainable by consumers seeking prevention, but the legal professional and governmental action leaves much to be desired in dealing with these threats.

Ultimately, perhaps the best advice when customers and organizations are confronted with these threats is the need to promote education and commonsense prevention. Cybercriminals will always seek soft targets, generally caused by lack of easy prevention techniques that could easily be prevented if proper education is made more universal. Further research needs to deals with this problem of making such educational opportunities more available and easy-to-use. The particular needs of each individual in regards to defense in dealing with the various ways they could be attacked must be addressed. If organizations can use big data analytics to e-personalize their products and services; why not apply the same

technologies to customize defense and safeguards against cyberattacks? For example, when it comes to phishing, of which there are countless victims, with the proper education they can learn to detect what is fraudulent and withhold their information from getting out to these cybercriminals. Hence, essentially every attack is preventable regardless of whether it is a phishing scheme, malware attack, data communication interception, loss or theft of device internal breach, and/or a virus attack. Future research that is aimed at educating potential victims would provide much relief in these ever increasing cybercriminal problems. The authors of this chapter suggest that further research should continue on all levels of cyberattacks. Society, in general, is becoming more vulnerable to such attacks and educating customers and managers about different threats and opportunities made available to counter hackers that are becoming more intelligent on ways to undermine the various precautions put in place, makes the needs to make potential victims more intelligent about cyberthreats even more pressing.

REFERENCES

- Aeron, H., Kumar, A., & Janakiraman, M. (2010). Application of data mining techniques for customer lifetime value parameters: A review. *International Journal of Business Information Systems*, 6(4), 530–546. doi:10.1504/IJBIS.2010.035744
- Alderete, M. V., & Gutiérrez, L. H. (2014). Drivers of information and communication technologies adoption in Colombian services firms. *International Journal of Business Information Systems*, 17(4), 373–397. doi:10.1504/IJBIS.2014.065553
- Barra, R. A., Savage, A., & Tsay, J. J. (2010). Equational zero vector databases, non-equational databases, and inherent internal control. *International Journal of Business Information Systems*, 6(3), 354–377. doi:10.1504/IJBIS.2010.035050
- Barrett, D., & Yadron, D. (2016). Sony, U.S. Agencies fumbled after cyberattack: Lack of information and consultation led to flip-flops, confusion. *The Wall Street Journal*. Retrieved February 26, 2016 from <http://www.wsj.com/articles/sony-u-s-agencies-fumbled-after-cyber-attack-1424641424>
- Beldona, S., & Tsatsoulis, C. (2010). Identifying buyers with similar seller rating models and using their opinions to choose sellers in electronic markets. *International Journal of Information and Decision Sciences*, 2(1), 1–16. doi:10.1504/IJIDS.2010.029901
- Bertolucci, J. (2016). Cybercrime is rampant around the world, says study. *PCWorld*. Retrieved March 7, 2016 from http://www.pcworld.com/article/205051/Norton_Study_Says_Cyber-crime_is_Rampant.html
- Chand, M., Raj, T., & Shankar, R. (2015). A comparative study of multi criteria decision making approaches for risks assessment in supply chain. *International Journal of Business Information Systems*, 18(1), 67–84. doi:10.1504/IJBIS.2015.066128
- Clarke, E. (2013). *The underground hacking economy is alive and well*. Dell SecureWorks™. Retrieved February 26, 2016 from <https://www.secureworks.com/blog/the-underground-hacking-economy-is-alive-and-well>
- Combating cybercrime. (2016). *Cybersecurity*. Homeland Security. Retrieved March 7, 2016 from <https://www.dhs.gov/topic/combating-cyber-crime>
- Daim, T., Basoglu, N., & Tanoglu, I. (2010). A critical assessment of information technology adoption: Technical, organisational and personal perspectives. *International Journal of Business Information Systems*, 6(3), 315–335. doi:10.1504/IJBIS.2010.035048

- Daramola, J. O., Oladipupo, O. O., & Musa, A. G. (2010). A fuzzy expert system (FES) tool for online personnel recruitments. *International Journal of Business Information Systems*, 6(4), 444–462. doi:10.1504/IJBIS.2010.035741
- Dharni, K. (2014). Exploring information system evaluation in Indian manufacturing sector. *International Journal of Business Information Systems*, 17(4), 453–468. doi:10.1504/IJBIS.2014.065564
- Dominic, P. D. D., Goh, K. N., Wong, D., & Chen, Y. Y. (2010). The importance of service quality for competitive advantage – with special reference to industrial product. *International Journal of Business Information Systems*, 6(3), 378–397. doi:10.1504/IJBIS.2010.035051
- Elysee, G. (2015). An empirical examination of a mediated model of strategic information systems planning success. *International Journal of Business Information Systems*, 18(1), 44–66. doi:10.1504/IJBIS.2015.066126
- Fitz-Gerald, S. (2014). Everything that's happened in the Sony leak scandal. *Vulture*. Retrieved February 26, 2016 from <http://www.vulture.com/2014/12/everything-sony-leaks-scandal.html>
- Friedman, J., & Hoffman, D. (2008). Protecting data on mobile devices: A taxonomy of security threats to mobile computing and review of applicable defenses. *Information, Knowledge, Systems Management*, 7(1-2), 159–180.
- Jarrett, M. H. (2016). *Prosecuting computer crimes*. Office of Legal Education Executive Office for United States Attorneys. Retrieved February 26, 2016 from <http://www.justice.gov/criminal/cybercrime/docs/ccmanual.pdf>
- Kapur, P. K., Gupta, A., Jha, P. C., & Goyal, S. K. (2010). Software quality assurance using software reliability growth modelling: State of the art. *International Journal of Business Information Systems*, 6(4), 463–496. doi:10.1504/IJBIS.2010.035742
- Kassner, M. (2014). The New Phenomenon report shows cybercrime is on the rise. *TechRepublic*. Retrieved March 7, 2016 from <http://www.techrepublic.com/article/new-ponemon-report-shows-cybercrime-is-on-the-rise/>
- Keramati, A., & Behmanesh, I. (2010). Assessing the impact of information technology on firm performance using canonical correlation analysis. *International Journal of Business Information Systems*, 6(4), 497–513. doi:10.1504/IJBIS.2010.035743
- Kerr, O. S. (2010). Vagueness challenges to the Computer Fraud and Abuse Act. *Minnesota Law Review*, 1561. Available at SSRN: <http://ssrn.com/abstract=1527187>
- Kim, E., & Tadisina, S. (2010). A model of customers initial trust in unknown online retailers: An empirical study. *International Journal of Business Information Systems*, 6(4), 419–443. doi:10.1504/IJBIS.2010.035740
- Latha, T. J., & Suganthi, L. (2015). An empirical study on creating software product value in India - an analytic hierarchy process approach. *International Journal of Business Information Systems*, 18(1), 26–43. doi:10.1504/IJBIS.2015.066125
- Lawrence, D. (2014). Christmas Shopping for Cybercriminals. *BloombergBusiness*. Retrieved March 7, 2016 from <http://www.bloomberg.com/bw/articles/2014-12-15/current-hacker-underground-markets-are-booming>
- Marthandan, G., & Tang, C. M. (2010). Information systems evaluation: An ongoing measure. *International Journal of Business Information Systems*, 6(3), 336–353. doi:10.1504/IJBIS.2010.035049
- Mohanty, R., Ravi, V., & Patra, M. R. (2010). The application of intelligent and soft-computing techniques to software engineering problems: A review. *International Journal of Information and Decision Sciences*, 2(3), 233–272. doi:10.1504/IJIDS.2010.033450

Report, C. (2011). *Norton*. Retrieved March 7, 2016 from http://www.symantec.com/content/en/us/home_homeoffice/html/cybercrimereport/

Riley, C., & Pagliery, J. (2015). Target will pay hack victims \$10 million. *CNN Money*. Retrieved March 2, 2016 from <http://money.cnn.com/2015/03/19/technology/security/target-data-hack-settlement/>

Riley, M., Elgin, B., Lawrence, D., & Matlack, C. (2014). Missed alarms and 40 million stolen credit card numbers: How Target blew it. *Bloomberg Business*. Retrieved February 26, 2016 from <http://www.bloomberg.com/bw/articles/2014-03-13/target-missed-alarms-in-epic-hack-of-credit-card-data>

SecureWorks competitors - CB insights. (2016). Retrieved February 23, 2016, from <http://www.cbinsights.com/company/secureworks-competitors>

Shinder, D. (2011). Cybercrime: Why it's the new growth industry. *IT Security*. Retrieved March 7, 2016 from <http://www.techrepublic.com/blog/it-security/cybercrime-why-its-the-new-growth-industry/>

Shukai, L., Chaudhari, N. S., & Dash, M. (2010). Selecting useful features for personal credit risk analysis. *International Journal of Business Information Systems*, 6(4), 419–443. doi:10.1504/IJBIS.2010.035745

Smith, A. D. (2007). Identity theft and e-fraud driving CRM information exchanges. In A. Gunasekaran (Ed.), *Modeling and Analysis of Enterprise Information Systems* (pp. 110–133). Hershey, PA: IGI Publishing. doi:10.4018/978-1-59904-477-4.ch005

Soon, J. N. P., Mahmood, A. K., Yin, C.-P., Wan, W.-S., Yuen, P.-K., & Heng, L.-E. (2015). Barebone cloud IaaS: Revitalisation disruptive technology. *International Journal of Business Information Systems*, 18(1), 26–43. doi:10.1504/IJBIS.2015.066130

Sundarambal, M., Dhivya, M., & Anbalagan, P. (2010). Performance evaluation of bandwidth allocation in ATM networks. *International Journal of Business Information Systems*, 6(3), 398–417.

Underground, H. M. R. (2014). *SecureWorks*. Retrieved February 27, 2016 from <https://www.secureworks.com/assets/pdf-store/white-papers/wp-underground-hacking-report.pdf>

Warren, J. (2015). Why US small businesses should worry about cybersecurity and how to act. *IT Governance*. Retrieved March 2, 2016 from <http://www.itgovernanceusa.com/blog/why-us-small-businesses-should-worry-about-cybersecurity-and-how-to-act/>

Watson, L. (2015). Crime soars 107% as cyber offences included for the first time – as it happened. *Crime*. Retrieved March 7, 2016 from <http://www.telegraph.co.uk/news/uknews/crime/11932670/Cyber-crime-fuels-70-jump-in-crime-levels.html>

ADDITIONAL READING

Code, U.S. 12. (2016). Legal Information Institute, Cornell University. [Online]. Retrieved February 26, 2016 from http://www.law.cornell.edu/uscode/html/uscode18/usc_sec_18_00001030----000-.html

Computer and information technology occupations. (2016). *Occupational Handbook*. U.S. Bureau of Labor Statistics. [Online]. Retrieved March 4, 2016 from <http://www.bls.gov/ooh/computer-and-information-technology/home.htm>

Han, W., Ada, S., Sharman, R., Gray, R. H., & Simha, A. (2015). Factors impacting the adoption of social network sites for emergency notification purposes in universities. *International Journal of Business Information Systems*, 18(1), 85–106. doi:10.1504/IJBIS.2015.066129

Kapitanyan, M. (2009). Beyond wargames: How the Computer Fraud and Abuse Act should be interpreted in the employment context. [Online]. Retrieved February 26, 2016 from https://www.researchgate.net/publication/265939376_Beyond_WarGames_How_the_Computer_Fraud_and_Abuse_Act_Should_Be_Interpreted_in_the_Employment_Context

More than a decade of data tracking the evolution of cyber threats. (2015). *Verizon Data Breach Investigations Reports*. [Online]. Retrieved February 23, 2016, from <http://www.verizonenterprise.com/DBIR/>

Ponemon study shows the cost of a data breach continues to increase. (2016). [Online]. Retrieved March 4, 2016 from <https://www.ponemon.org/news-2/23>

Skibell, R. (2003). Cybercrimes & misdemeanors: A reevaluation of the Computer Fraud and Abuse Act. *Berkeley Technology Law Journal*, 18(3), 909–944.

Smith, A. D. (2004). Cybercriminal and cyberterrorism activities and its affects on online consumer confidence. *Online Information Review*, 28(3), 224–234. doi:10.1108/14684520410543670

Smith, A. D., & Lias, A. R. (2005). Identity theft and e-fraud as critical CRM concerns. *International Journal of Enterprise Information Systems*, 1(2), 17–36. doi:10.4018/jeis.2005040102

Theohairs, M. (2016). Computer and Internet crime laws. [Online]. Retrieved February 26, 2016 from <http://www.criminaldefenselawyer.com/crime-penalties/federal/computer-crimes.htm>

KEY TERMS AND DEFINITIONS

Advanced Persistent Threats (APT): Advanced Persistent Threats typically employs a number of phases and techniques to break into a network, avoid detection, and gather customers' or companies' information over an extended time period. Unfortunately, such well-planned and extensive strategy is difficult to detect, as it many techniques and phases may be extremely complex and expensive to develop a counteroffensive that is successful.

Direct Cyberattack: Direct cyberattack causes data communications interception at the hands of a specific hacker who may have located a specific person or Wi-Fi area in order to perform such cybercrime. A type of direct attach is direct access, where the criminal gains physical access to either the computer or its network in order to compromise security by loading worms, Trojans, compromising data, etc.

Personally Identifiable Information (PII): Personally identifiable information (PII) refers to all information, including personal financial and healthcare information that can be traced to an individual customer or patient.

Remote Access Trojan (RAT): In basic terms, Remote Access Trojans (RATs) are essential tools available to cyberthieves with unrestricted limited access to computers. By using stolen access privileges through properly submitted passwords, all important business and personal data, the entire customer's or company's information can be readily accessed and retrieved.

Zero-Day Attack: This term has been used in various forms and meaning. It terms of this chapter, this term refers to exploitable software or hardware vulnerabilities that have been attacked when there has been no previous information of a system flaw by the general information security community. In such a situation, unfortunately, there is no vendor fix or software security patch that have been developed or made available to the public to correct or protect against such an attack.

Cost Estimation and Security Investment of Security Projects



Yosra Miaoui

University of Carthage, Tunisia

Bouthaina Fessi

University of Carthage, Tunisia

Nouredine Boudriga

University of Carthage, Tunisia

INTRODUCTION

Project management is an important task that should be performed when dealing with security project, since it allows avoiding different project failures. This task is an effective methodical approach of planning, organizing, leading, and controlling resources to achieve organization's goals. It involves, thus, identifying requirements, determining clear objectives, and balancing the triple constraints scope, time, and cost (Institute, 2013).

It is noted that the management of software and security projects are not performed in the same way, due to several reasons, related mainly to the software intangibility, complexity, conformity, and flexibility. It is also shown that the parameters involved in the security cost estimation differ from those of software cost estimation. Therefore, the developed methods should be adapted to consider security specificities.

In this context, organization's project managers should estimate the cost associated to a security project during its design. This estimation should include the computation of the optimal security level and residual risk accepted by the organization. Moreover, it should consider managerial aspects regarding, for example, the effort required for security monitoring of the new assets to be acquired or updated, the security training of the technical staff, the update of the managerial deci-

sional system, and the development of policy and procedures related to the use of information processing facilities, instead of only considering the industrial source coding of the security packages.

Another significant aspect, which should be carefully examined when dealing with security, is related to security investment. The financial budget allocated to security should be well established and managed to avoid under or over expenses. Different security investment models are developed in the literature using various techniques and examining several features. Most of them have focused on determining the optimal security investment allocation based on budgetary aspect, economic, and financial constraints. Recent works are interested to examine more specific security features when assessing the required investment, such as the system vulnerabilities, attacks type, risk factors, data privacy, and insurance.

This chapter aims at examining two aspects related to security project: cost estimation and investment assessment. First, the characteristics of security projects are stressed on and the importance of adopting management is determined. Then, the chapter presents the different cost estimation models dedicated to security project and discusses the technical and managerial factors affecting the cost estimation and the management of project. In addition, a sample review of research works directed toward security investment models is determined. These models are organized according

to the type of issues and aspects handled to compute the optimal amount of security investment. Finally, the chapter discusses future directions that could be investigated to make available useful models for cost estimation and investment on security projects.

SECURITY PROJECTS MANAGEMENT FRAMEWORKS

In this section, we examine the objective and features of security project and show the importance of the management task when dealing with these projects.

Definition and Characteristics

Security project is a specific type of project that implements a set of tasks to protect and secure a considered information system from attacks and potential threats. It lies usually outside the core functions of the business and aims to protect critical involved resources.

Security project is different from a software project for at least five features which characterize it:

- First, the security project requires a better knowledge of the security threats and vulnerabilities surrounding the activity of the enterprise, their future severity, and the evolving techniques they will implement. In addition, managing efficiently a security project assumes that the remaining risk related to damaging attacks, unobserved during project design, will be confined in the future. Moreover, the necessity to provide a response to significant attacks supposes that an efficient monitoring of activity and risk assessment are guaranteed.
- Second, the output of security projects is complicated and may include: (1) a security policy customized to the enterprise, its activity, and its environment describing the rules to be enforced, the security Procedures to be observed, the detection to perform, and the security invariants to comply with. While a security policy acts as a specification of the security project, it differs from a specification in the way it involves the activity of an incident response team and the procedures it triggers on the occurrence of attacks; (2) a set of preventive, detective, and responsive systems to be deployed in a way that allows flexible configuration and real time connectivity to other information and decision systems in the enterprise; and (3) a set of countermeasures to introduce on preventive and monitoring tools to thwart future attacks along with security procedures and guidelines to be followed by people when immediate reactivity on the managerial procedures is needed.
- Third, estimating security project complexity is influenced by different factors, such as the size of the information system to secure in terms of number and size of interconnected sub-networks; the number of users exploiting the information system resources including customer users connecting remotely and administrators; and the complexity of the security policy defining the measures to implement and rules to enforce.
- Fourth, a great part of the work in a security project is added on the customer network in order to: (i) interview the users, security administrators, and managers, especially during the risk analysis phase to identify the managerial issues that can affect the project; (ii) integrate and configure the developed solution and update the existing networking solutions to guarantee an adequate inter-operation; and (iii) assess the robustness of the resulting security system through auditing and penetration testing.
- Fifth, a security project has to keep under control the variation of the security robust-

ness of the project over time. This need is difficult to achieve since vulnerability models cannot handle a good perception of the vulnerability evolution.

Management Role in Security Project

Security project, as all types of projects, needs a project management activity in order to better pursue its objective. Indeed, this latter is considered as a process that helps producing successful projects by ensuring that the key elements of the projects are included in the right order at the right time.

A successful security project requires a good management of ten key elements (Stahl, 2008): (1) project scope, (2) project time and schedule, (3) project cost and budget, (4) project resources (internal and external), (5) contract management, (6) procurement management, (7) project communications, (8) quality management, (9) project risk, and (10) cross-organizational coordination.

The project management activity consists in organizing, planning, and scheduling tasks. These later represent the most time-consuming project management activities. Therefore, elaborating and following a security project plan is a useful way before implementing security solutions. The developed security plan will be customized to fit the type and the needs of the organization to protect. The major steps include identifying threats, analyzing and prioritizing threats, developing plans and strategies to reduce the occurrence likelihood of the identified threats, and elaborating contingency plans.

It is noted that various benefits are determined when adopting project management to security project. In fact, the core advantage of performing the management is to avoid some projects failures, related for example to over estimation of the budget allocated to security project, unrealistic timelines for the different security tasks achievement, and large definition of the project scope. Moreover, adopting project management provides a strong framework for roles and responsibilities. This is crucial since the role of security is often misun-

derstood by many within an organization. A highly defined role for security can help alleviate friction between the security program and the rest of the organization (Gentil, 2015). In addition, the management allows defining for each process group of the project (i.e. initiation, planning, execution, controlling, and closing) the set of issues to be solved as it is specific for each group.

COST ESTIMATION MODELS FOR SECURITY PROJECTS

This section provides a brief survey of the developed cost estimation models for security projects. It also highlights the importance of integrating technical and managerial factors in these cost estimation models.

Review of Existing Cost Estimation Models

The developed cost estimation models for security projects are not as popular as those dedicated to software projects (Boehm, 2000). The ones described in the literature are cited in the following.

Labor Time Estimation Model (LTEM)

The model is described in (Sato, 2009) to estimate the labor time of information security audit projects through the estimation of the labor time of its four phases, i.e., planning, implementation, reporting, and improvement. The estimation is based on factors related mainly to the type of the audit form, the operation mode, the enterprise size, and the penetration degree of information security management. These parameters are measured through performing quantitative and regression analysis of data collected from past audit projects.

Despite the suitability of the LTEM model to assess security audit projects, it is noted that it is not appropriate to estimate the cost of security projects due to a number of limitations. In fact, the model considers few parameters which affect the



effort and the cost of conducting security projects. Factors related to the information system to be audited or reporting on the auditing team skills could be considered. In addition, the applied size estimation technique is only based on the number of employees involved in the audit process, and does not consider the information system components that could affect it. Another important limitation of this model lies in its lack of consideration of the security policy, which could be used as a baseline for auditing, analysis, compliance checking of the information system, and the correctness of the configuration of the security solutions to the security rules. Therefore, the security policy could introduce additional labor time on the audit activity, especially, when inconsistency is noticed. Furthermore, a security policy can be used in size estimation as it provides more constraints than a specification.

A Security Cost Model (SECOMO)

This model aims at estimating the effort required to conduct a security project (Krichene, 2003). It provides a basis for determining how much time, cost, and personnel are required for security projects. It is founded similarly to the Constructive Cost Model (COCOMO), which is a hybrid cost estimation model based on regression and expertise based techniques.

SECOMO defines three security cost models, classified into basic, intermediate, and advanced. These models differ according to their complexity, size, and the strength of the security policy they implement. In addition, they are based on the size of the network to be secured and use a set of cost drivers, decomposed in effort multipliers and scale factors, that give a measure of the security task complexity and where their effects are different. Effort multipliers have linear effect on the estimation, whereas, scale factors have an exponential impact.

Despite the contribution of this model to the management of security projects, some limitations could be identified. First, SECOMO is tailored to

enterprises developing security software. It cannot be used, as it is, by any enterprise assessing the total cost of a security project it wants to acquire for itself. Different used parameters could not be estimated accurately, such as developing team capability and tool experience since they are intrinsic to a software development company. The second limit is related to the inability of SECOMO to consider the managerial effort related to the integration of the solutions to systems dealing with strategic intelligence and the organizational security documents. In fact, a large set of managerial and organizational tasks within the enterprise should be accounted during the cost estimation as they are included in the enterprise's security policy. In particular, the awareness activity following the deployment of a security solution is not considered in SECOMO.

A Managerial Issues-Aware Cost Estimation of Enterprise Security Projects (MCE)

The MCE cost estimation model is an extension of SECOMO and COCOMO models. It is developed to conduct a security project and integrates the impact of managerial and technical aspects related to five types of tasks, namely monitoring, awareness, decision making, re-engineering, and integration and compliance (Fessi, 2013). The developed model allows managers predicting the cost associated to a security solution according to the enterprise's needs, and thus they (i.e. managers) could define and manage efficiently their project budget.

The developed cost model depends on the size of the information system, the size of the existing security policy (SP), the enterprise size, in addition to a set of cost drivers. It allows estimating numerically the managerial impact on the cost associated to a security project, compared with the other existing models. Moreover, it gives an important role to the security policy supporting the security project, as it represents a key element for estimating the required effort and the expected cost.

It is noted that behind the provided benefits, the developed model is closely tied to the content of security policy (SP), which is considered as a static document during the project estimation time. In reality, SP must be adapted to the environment where the enterprises operate. This latter is characterized by its complexity and uncertainty where threats and vulnerabilities increase continuously, attackers change their behavior over time, and resources may show dynamic features. Moreover, MCE lacks of tools for the estimation of the financial cost associated to the residual risk which is not addressed by security project. In fact, by considering the risk value, the enterprise could take the appropriate decisions to invest or not in the security project. We notice that this limitation is shared with the other developed cost estimation models.

Technical and Managerial Factors Support

The presented security project cost estimation models (Krichene, 2003; Fessi, 2013; Satoh, 2009) show the involvement of two types of factors: technical and managerial. The first set of factors are those related to the enterprise assets and to the security solution, whereas the second set is associated to the enterprise management features, such as organizational, human, and risk management. This integration of managerial factors besides technical ones into the estimation of the cost of the security project allows more accuracy to determine security budget allocation.

As it is mentioned beforehand, several factors are identified in the presented models. For the LTEM model, factors that influence the labor times include the type of business, the audit form, the operation mode, the penetration degree of information security management, and the enterprise size. For the SECOMO model, 23 parameters are defined and referred to as effort multipliers factors (i.e., product, personal, project, and information system) and scale factors

(i.e., precedentedness, team cohesion, project maturity, and security strategy). And for the MCE model, 19 parameters are identified and are allocated according to their relation to information system size, security policy, enterprise size and the defined five tasks (monitoring, awareness, decision making, re-engineering, and integration and compliance). Despite this variation of parameters between the presented models, it is noticed that some parameters are considered in two or all models, but have different designations such as audit form parameter in LTEM and audit frequency parameter in SECOMO; the enterprise size parameter presented in LTEM model corresponds to multi-site information system parameter in SECOMO model and it is also used as a key measure in the MCE model.

The different cost drivers could be assessed either quantitatively or qualitatively. Quantitative measures are determined on the basis of historical data, defined functions, statistical measures, or even experts opinion. Qualitative measures, in the other hand, require the intervention of experts to define the scale level of each parameter, and then approximate measure could be affected to it. For example, in SECOMO, the whole majority of parameters related to personal, are qualitative parameters, while in MCE model, strategic intelligence and awareness factors are qualitative and should be determined by experts.

SECURITY INVESTMENT MODELS FOR SECURITY PROJECTS

Several research works are developed to compute the optimal amount of security investment from the financial and economic perspectives. Recent works are directed toward integrating new aspects related mainly to cyber-insurance, security vulnerabilities patching, and game theory in cybersecurity investment. These models are described in the following paragraphs.



Economics of Security Investment Models

Determining the optimal amount of security investment which minimizes the risk of security attack is still an open issue for decision-makers, due to several factors such as the limited financial budget allocated to security information systems, the near-impossibility of eliminating security residual risk, and the dynamic and evolutionary aspect of different types of vulnerabilities and threats facing the firm's information systems.

Several works have been developed to handle security investment problem using a multitude of economic and financial approaches and techniques, such as risk-return analysis (Hausken, 2014; Gordon, 2002), utility theory (Huang, 2013; Miaoui, 2015a), game theory (Cavusoglu, 2008; Hua, 2011) and real options theory (Li, 2007; Tatum, 2009; Ullrich, 2013). The security optimal investment model based on the maximization of expected net benefits was first proposed in (Gordon, 2002). In this model, the decision-makers, which are assumed to be risk-neutral, maximize their expected incomes from security investment by comparing the marginal financial benefits and the marginal financial costs of information security. The model analyzes the impact of several parameters, such as system's vulnerabilities and potential financial loss on the amount of optimal security investment considering two classes of security breaches, namely targeted attacks and distributed attacks.

In (Huang, 2008), utility based model is used to determine the optimal security investment with the assumption of risk averse decision-makers. Considering each class of security breach probability, as assumed in (Gordon, 2002), different scenarios are conducted to analyze the relationships between optimal security investment levels and the potential loss, the extent of the decision maker's risk aversion, and the investment effectiveness. In an extension work, the authors developed an analytic model that optimizes the allocation of security investment to defend against concurrent

heterogeneous attacks on a firm's information systems (Huang, 2013). The model considers the budget constraint that all firms face and is based on the concept of scale-free networks to obtain breach probabilities for different classes of attacks.

Another context of security investment is examined in (Miaoui, 2015b), where the firms outsource or offshore their business operations. In this case, particular aspects related to bilateral security risk and data privacy are considered when computing their security investments. The developed model considers the involvement and interdependency of several stakeholders (e.g., attackers, the privacy protection firms, data sellers, and outsourcing companies) in conducting and protecting against privacy attacks in the context of an outsourced business activity. Moreover, the authors designed different models related to security threats on data privacy and consider the particular aspects of privacy attacks when determining the expected monetary loss and benefits. They also conducted a numerical analysis to assess the impact of the quality of detection and reaction to privacy breaches on the amount of optimal investment and residual risk.

Cyber-Insurance Investment

Several investments can be adopted by companies to protect their information systems against damaging security incidents, such as self-protection, self-insurance and cyber-insurance. The self-protection investment includes the development of security policy, the design of architectures and the deployment of security solutions (including, but not limited to anti-virus, firewalls or intrusion detection mechanisms) to mitigate the occurrence of security breaches, while the self-insurance investment is an internal investment made by an enterprise to cover himself from a financial potential loss. To reduce the large amount of potential losses induced by cyber-attackers, an investment in cyber-insurance can be undertaken by an enterprise (insured) to transfer the financial risk of security breaches to a third party partner (insurer).

An insurance contract is, thus, established between the two parties (insurer and insured) including in particular, the insurance premium to be paid by the insured firm and the reimbursement rate at which the insurance company will cover the amount of potential loss.

While the transfer of risk through an insurance company was suggested, by practitioners and academics, to cover the expenses and losses induced by security breaches (cost of notification and response to security incidents, loss or corruption of data), complete the existing set of security countermeasures and manage the security residual risk (after investments in cyber-security have already been made), the growing evolution of cyber risk and the interdependency of information security risks (due to interconnections of computers across different firms) have made the cyber insurance investment particularly complex, unclear, and more challenging (Zhao, 2013).

Indeed, as a classic insurance, the cyber-insurance raises several issues related to: a) pricing which is the amount of premium to charge to the insured firm in turn of potential loss coverage. To determine the appropriate price, insurance company should accurately estimate the cyber risk incurred by the insured firm. However, some estimation bias can be introduced when the applicant firms have high risk interdependencies due to their high information systems openness to external environment and/or their degree of information sharing through outsourcing operations; b) adverse selection which is a hidden action problem arises when the insured firm does not communicate its real security risk in order to benefit from lower risk premium. To deal with this problem, the insurance company can estimate the security risk of the applicant's resources, the systems vulnerabilities and the threats through security audit. As the security vulnerabilities and threats severity, frequency and evolution with respect to time are increasing, security incidents occurrence probability as well as the potential loss

to be reimbursed by insurer are also increasing; and c) moral hazard which is an asymmetric information problem arising from the non-ability of the insurance company to observe the behavior of insured firm after purchasing cyber-insurance. The latter may behave recklessly in order to increase the likelihood of insurance claims.

In fact, although an insurance contract does not prevent security breaches from occurring, some firms may find it cost sensitive to invest in cyber insurance than in self-protection. In the worst case, untrusted insured firms could create fictive self-attacks and ripe benefits from the insurance company. Therefore, a high premium or a low coverage (e.g., use of deductibles) should be given to firms that do not take the appropriate measures to reduce the probability of loss, or that are unable to prove that the incident is not self-generated.

In this context, several models were developed to cope with these issues (Gordon, 2003; Herath, 2011; Böhme, 2010). In particular, cyber insurance has been proposed as a promising solution to manage security risks and to optimize security investment. In (Ogut, 2005) and (Bolat, 2008), the impact of cyber-insurance investment is studied, when firms face correlated risk on the cost and efficiency of security investments, and on the increase of incentives to invest in self-protection. In (Grossklags, 2008), a model of security investments interactions using game theory was proposed. The authors show how the players decide strategically to shift between investment in self-protection and investment in cyber insurance. Recent work proposed by (Miaoui, 2015a) is focused on the importance of security investment in forensic investigation to collect digital evidence and generate provable insurance claims to ensure a better reimbursement of loss in case of security breaches occurrence. The developed investment model is based on three interdependent investment types: investment in self-defense, investment in insurance, and investment in forensic readiness.

Security Vulnerabilities Patching Investment

The detection and the patching of vulnerabilities represent important tasks that keep an information system secure. Security administrators should typically proceed as soon as possible with the application of patches or the modification of security policy rules (e.g., deactivation of vulnerable services, ports filtering) to prevent potential exploitation of the identified vulnerabilities. In practice, different reasons may delay or prevent a firm from the immediate application of security patches, such as: (1) the number of patches to be deployed is important and their implications on security vary accordingly so that it is necessary to sort and prioritize vulnerabilities, which is a labour-intensive task; (2) there may be many available fixes for an identified vulnerability (i.e., official ones released by the vendor, and other ones released by the security community experts), so a test of compliance with existing software should be performed; and (3) a maintenance period, where the production system needs to be rebooted or temporary disconnected, could introduce an economic loss if it is required.

This delayed application of patches could lead to serious economic losses and could have severe security consequences. Therefore, a firm should design a patch management strategy to determine the time and the appropriate patch to deploy for the detected vulnerability. This strategy will keep the safety of the system, reduce the operational cost, and reduce the security risk of loss to an acceptable level. In this context, some research works are directed toward this issue and propose economic models based on game theoretic approaches to derive the optimal strategy and frequency of patches update. This strategy depends on the operation cost model, the required downtime period, the patch release policy of the software provider, and the stochastic vulnerability release. The developed model in (Cavusoglu, 2006) defines an objective function to minimize the damage cost of exploiting an unpatched vulnerability and the update patch

to shield vulnerability at the client company. This latter is assumed to incur attacks either because it is waiting for the release of a patch or because the patch is released but it is waiting the end of the update cycle. For the customer firm, the software vendor firm is assumed to release vulnerability fixes every predefined period of time, incurring consequently two different losses: patch development and release losses, and reputation cost.

Game Theory in Security Investments

Regarding the previous approaches, where the security investment is considered as an optimization problem that determines the best compromise between the security budget to spend and the residual risk to accept, several recent works are interested to use the game theory as a security tool to model the cyber environment under which defenders and offenders interact together. Such a theory will capture the behaviour of the attacker which varies depending on the strategy of the enterprise in defending and responding to security attacks. In most real scenarios of cyber protection, several actors interact together, including insurance companies, privacy protection firms, defender, third party companies, network and security providers, and even security authorities. All of these actors try to improve their expected benefits and therefore they adapt their behaviour to the visible or hidden properties of the cyber environment (available resources, loss of incur, identification ability, etc.).

In the context of security economics, several applications of game theory were proposed during the last years. In (Nagurney, 2015), information asymmetry between sellers and buyers is introduced to determine the optimal security investment at the seller side. In fact, contrarily to sellers which have a good knowledge of their security investment and the residual security risk they incur, and want to maximize their expected benefit from selling their products on the Internet market space (and therefore execute

secure transactions); buyers set their preference regarding the product prices based on the average provided security level of the seller (this security level is imprecise due to information asymmetry and is not related to individual sellers, but simply reflects the average security level in the industry). The used game theory determines the Nash equilibrium by which sellers can assess the cyber security investment, the expected damages in case of attacks, and the attacks probabilities.

In (Panaousis, 2014), the game theory is used as a tool to model the cyber security environment of a firm, as well as a non cooperative control game between a defender and an attacker. The former abstracts decision makers in security to protect the firm's assets and the latter abstracts all untrusted parties that aim to compromise the firm's assets. In this work, a security control is defined by the firm with respect to the direct and indirect damages that it can avoid on all targets. The game is modelled based on a two player minimax game, to allow defining one security control to implement, by which the more the attacker gains, the higher will be the firm's loss. Moreover, the proposed model implements a multi-objective optimization problem to derive the optimal allocation of the security budget, since different security controls are required by the firm to defend against the different security vulnerabilities.

Another model is developed based on the game theory approach (Fielder, 2014) to support making decisions regarding the allocation of security resources, including the time consumed by administrators in security their systems. The proposed model assumes that an attacker could harm many targets, and each target is threatened using a single vulnerability. Therefore, to defend against attacks, two security levels are considered, namely baseline and best practice. The former describes the probability of an attack, while the latter denotes the percentages of attacks being defended.

FUTURE RESEARCH DIRECTIONS



This section presents a number of future research directions related to unsolved challenging issues by the current cost estimation models of security projects. It also examines new research trends expected to determine an optimal security investment model.

Cost Estimation Models

To enhance the cost estimation of security projects, a set of consideration should be involved in the computation task. The first concern is related to the management of uncertainty and dynamicity related to the environment where the solution will be deployed. In fact, the cost estimation model should be extended to include additional factors that manage this uncertainty and cope with the variation of the protected environment by the security solution. It is, therefore, important to develop proactive cost models that tolerate modifications on the security policy during the achievement of the security projects to consider new threats and vulnerabilities.

The risk factors are other important features that should be considered when estimating the cost of security projects. They are related to the risk associated to future security breaches in the operational period of the developed security solution and to the remaining residual risk after deploying the security solution. For the first type, it is necessary to develop a model that anticipates the evolution of risk, associated to the unpredictable evolution of vulnerabilities and threats over time, and integrates their variable severities in order to enhance the cost-effectiveness of the developed security projects. The second type of risk allows the improvement of the efficiency of security solution in the future, since providing a complete security solution is impossible due to the constraints of the allocated budget to the security managers.

Security Investment Models

Different aspects might be considered to improve the developed security investment models. First, it is noted that when designing an investment model, the main aim is set on the efficiency of the security solution to protect the information system regardless its usefulness and its added value for the company's shareholders. It is, thus, more useful to develop a model that allows the achievement of its primary objective and to yield profit through including new parameters related to creation of value for the shareholder (such as dividends, profitability of funds, governance and financial transparency), in addition to parameters associated to the evolution and dynamicity of vulnerabilities and threats.

The second point to deal with concerns the security investment when private data are protected by outsourcing firms. In this case, the type of the subcontracting firms has an impact on the investment computation. This type is related to the degree and the nature of control (i.e. partial, global) over the outsourced resources (e.g., data, applications, services). In addition, related parameters to private business data should be taken into account as they could have significant impacts on determining the investment model, such as the value of the information to be protected, the degrees of sensitivity of this information to security attacks, the degrees of attractiveness of business data to attackers, etc.

CONCLUSION

The project management is an important task to be performed when dealing with security project to avoid project failures. Different researches assume that security and software projects are managed in the same way using the same tools and techniques. Nevertheless, it is noted that there are differences between the two projects and additional issues should be considered when managing security project.

A number of study were conducted around different aspects of security projects but remaining issues related specifically to the dynamicity and evolution of the environment, attacks, and system resources are needed to be resolved in order to develop an efficient cost estimation model and to better assess the budget allocated to security task. Therefore, additional efforts still needed.

This chapter focuses on the security project management and the existing cost estimation models. It also presents a review of the security investment models and describes the latest developed works. In addition, the chapter underlines the relationships between cost estimation and security investment when designing security projects and examines the future directions for the development of efficient models.

REFERENCES

- Boehm, B., Abts, C., & Chulani, S. (2000, November). Software development cost estimation approaches - a survey. *Annals of Software Engineering*, 10(1), 177–205. doi:10.1023/A:1018991717352
- Böhme, R., & Schwartz, S. (2010, June 7-8). Modeling Cyber-Insurance: Towards a Unifying Framework. *9th Annual Workshop on the Economics of Information Security (WEIS 2010)*. Harvard University.
- Bolot, J. C., & Lelarge, M. (2008, April 13-18). A New Perspective on Internet Security using Insurance. *Proceedings of the 27th Conference on Computer Communications (INFOCOM 2008)*. IEEE. doi:10.1109/INFOCOM.2008.259
- Cavusoglu, H., Raghunathan, S., & Yue, W. (2008). Decision-theoretical and game-theoretical approaches to it security investment. *Journal of Management Information Systems*, 25(2), 281–304. doi:10.2753/MIS0742-1222250211

- Cavusoglu, H., & Zhang, J. (2006). Economics of security patch management. *The fifth Workshop on the Economics of Information Security (WEIS 2006)*.
- Fessi, B. A., Miaoui, Y., & Boudriga, N. (2013, June). A managerial issues-aware cost estimation of enterprise security projects. *Journal of Information Security Research*, 4(2), 81–89.
- Fielder, A., Panaousis, E., Malacaria, P., Hankin, C., & Smeraldi, F. (2014, June 2-4). Game Theory Meets Information Security Management. *Proceedings of the 29th IFIP TC 11 International Conference (SEC 2014)*, 15-29.
- Gentile, M., Collette, R., & CISOHandbook.com Team. (2015). Project Management As A Security Touch-Point. In *Project Management and Security*. CISOHandbook.com.
- Gordon, L. A., & Loeb, M. P. (2002, November). The economics of information security investment. *ACM Transactions on Information and System Security*, 5(4), 438–457. doi:10.1145/581271.581274
- Gordon, L. A., Loeb, M. P., & Lucyshyn W. (2003, May 31). Information security expenditures and real options: A wait-and-see approach. *Computer Security Journal*, 19(2).
- Grossklags, J., Christin, N., & Chuang, J. (2008, April 21-25). Secure or Insure? A Game-Theoretic Analysis of Information Security Games. *Proceedings of the 17th International Conference on World Wide Web*, 209-218.
- Hausken, K. (2014, April). Returns to information security investment: Endogenizing the expected loss. *Information Systems Frontiers*, 16(2), 329–336. doi:10.1007/s10796-012-9390-9
- Herath, H., & Herath, T. (2011, February 27). Copula Based Actuarial Model for Pricing Cyber-Insurance Policies. *Insurance Markets and Companies: Analyses and Actuarial Computations*, 2(1), 7–20.
- Hua, J., & Bapna, S. (2011). Optimal IS Security Investment: Cyber Terrorism vs. Common Hacking. In D. F. Galletta & T.-P. Liang (Eds.), *ICIS. Association for Information Systems*.
- Huang, C. D., & Behara, R. S. (2013, January). Economics of Information Security Investment in the Case of Concurrent Heterogeneous Attacks with Budget Constraints. *International Journal of Production Economics*, 141(1), 255–268. doi:10.1016/j.ijpe.2012.06.022
- Huang, C. D., Hu, Q., & Behara, R. S. (2008, August). An economic analysis of the optimal information security investment in the case of a risk-averse firm. *International Journal of Production Economics*, 114(2), 793–804. doi:10.1016/j.ijpe.2008.04.002
- Krichene, J., Boudriga, N., & Guemara, S. (2003). SECOMO: An estimation cost model for risk management. *Proceedings of international conference on Telecom (ConTel'03)*. doi:10.1109/CONTEL.2003.176966
- Li, J., & Su, X. (2007). Making Cost Effective Security Decision with Real Option Thinking. *International Conference on Software Engineering Advances (ICSEA)* (p. 14). IEEE Computer Society. doi:10.1109/ICSEA.2007.50
- Miaoui, Y., Boudriga, N., & Abaoub, E. (2015a). Insurance Versus Investigation Driven Approach for the Computation of Optimal Security Investment. *Proceedings of the 19th Pacific Asia Conference on Information Systems (PACIS 2015) on IT and Open Innovation*.
- Miaoui, Y., Boudriga, N., & Abaoub, E. (2015b). Economics of Privacy: A Model for Protecting Against Cyber Data Disclosure Attacks. *Proceedings of the 3rd Information Systems International Conference (ISICO 2015)* (pp. 1-15). Surabaya, Indonesia: Procedia Computer Science, Elsevier (Scopus Indexed). doi:10.1016/j.procs.2015.12.165

Nagurney, A., & Nagurney, L. S. (2015, August). A game theory model of cybersecurity investments with information asymmetry. *NETNOMICS: Economic Research and Electronic Networking*, 16(1-2), 127–148. doi:10.1007/s11066-015-9094-7

Ogut, H., Menon, N., & Raghunathan, S. (2005). Cyber Insurance and IT Security Investment: Impact of Interdependence Risk. *Proceedings of Fourth Workshop on the Economics of Information Security (WEIS 2005)*.

Panaousis, E. A., Fielder, A., Malacaria, P., Hankin, C., & Smeraldi, F. (2014, November 6-7). Cybersecurity Games and Investments: A Decision Support Approach. Lecture Notes in Computer Science: Vol. 8840. *Proceedings of the 5th International Conference (GameSec 2014)* (pp. 266-286). Springer. doi:10.1007/978-3-319-12601-2_15

Project Management Institute. (2013). *A guide to the project management body of knowledge (PMBOK® Guide)* (5th ed.). Author.

Satoh, N., & Kumamoto, H. (2009). Estimation model of labor time at the information security audit and standardization of audit work by probabilistic risk assessment. *International Journal of Computers*, 3(3), 311-320.

Stahl, S., & Pease, K. A. (2008). *A Success Strategy for Information Security Planning and Implementation*. Los Angeles, CA: Citadel Information Group, Inc.

Tatsumi, K. I., & Goto, M. (2009). Optimal Timing of Information Security Investment: Real Options Approach. In *The 8th Annual Workshop on the Economics of Information Security (WEIS 2009)*. University College London.

Ullrich, C. (2013). Valuation of IT Investments Using Real Options Theory. *Business & Information Systems Engineering*, 5(5), 331–341. doi:10.1007/s12599-013-0286-0

Zhao, X., Xue, L., & Whinston, A. B. (2013, March 11). Managing Interdependent Information Security Risks: Cyberinsurance, Managed Security Services, and Risk Pooling Arrangements. *Journal of Management Information Systems*, 30(1), 123–152. doi:10.2753/MIS0742-1222300104

KEY TERMS AND DEFINITIONS

Awareness: Is the extent to which an individual who has access to the information system assets is aware of. It is related to the importance of security and dangerousness of attacks, the enterprise's security requirements, and its responsibilities regarding the enforcement of security inside the information system.

Cost Estimation: Is an approximation of the probable cost of a project computed on the basis of the cost of all resources that will be charged to complete project activities.

Economic of Information Security: Is the discipline that applies different economic theories to resolve information security problems.

Information Security: Is the protection of the information system's resources from unauthorized access, use, disclosure, disruption, modification, perusal, inspection, recording or destruction.

Information System: Is a set of interconnected components (technology, process and people) that collect, process, store, and distribute information to sustain decision making and control in an enterprise.

Optimal Security Investment: Is the amount of investment which maximizes the Esperance of the gained loss and minimizes the risks of security attacks.

Project Management: Is a common framework providing managers with principles, techniques, and tools needed to manage project team effort efficiently and to meet successfully the enterprise's project objectives.

Residual Risk: Is a quantification of the risk or the degree of exposure that the protected information system will incur, after deciding to counter or eliminate known risk.

Security Attack: Is any form of malicious actions taken to harm the security of information system components. An action is classified as malicious with respect to the enterprise security policy.

Security Policy: Is a document that lays the framework for information system security of the enterprise. Through this framework, a security project team can draw intelligible objectives, plans, rules and formal procedures required to manage and protect the sensitive enterprise information system from different attacks.

Security Project: Is a set of activities that aim to protect and secure an information system from attacks and potential threats.

Security Risk: Is the likelihood that enterprise assets (i.e. information, systems and network infrastructures, data, programs and applications) be targeted by a successful attack.

Threat: Is an indication about a potential event that can harm the security of the protected resource. A threat can turn to a security attack once a vulnerability that can be exploited is found.

Vulnerability: Is a defect or weakness in information system's assets or mechanisms, which could lead to a security breach when exploited by malicious entities.



Development of Personal Information Privacy Concerns Evaluation

Anna Rohunen

University of Oulu, Finland

Jouni Markkula

University of Oulu, Finland

INTRODUCTION

Personal data collection and utilization are increasingly taking place today as a part of the application of personal data intensive systems and services. Both individuals and data collecting organizations benefit from this. Personal data are collected and processed by private companies and public organizations for various purposes, for example, for delivery of more personalized services and for marketing. Despite its usefulness, extensive personal data collection raises privacy concerns among data subjects, and these concerns are also discussed in public very often. For example, vehicle GPS tracking-based kilometer taxation has been recently under debate. This debate has shown that vehicle tracking data can also be used for purposes other than taxation, and that it can be combined with other data, for example, for producing traffic information services by private or public organizations. Information privacy concerns derive from data subjects' desire to not be monitored, and their worries about the consequences of the use of their data. Privacy concerns often decrease data subjects' willingness to disclose their personal data or to use services that require personal data disclosure. For this reason, privacy concerns may lead to non-adoption of new services and technologies, dropping out of them, or a decline in data disclosure (i.e., omitting data or providing false information). To address these issues, we need to understand how to evaluate

privacy concerns in current and future evolving service development contexts. We need insights into how privacy concerns have been evaluated in earlier contexts in order to adapt evaluations to new contexts.

In many countries, legislation sets the foundation for protecting personal data privacy and provides a framework for implementing privacy protection methods and technologies. However, it does not really suppress privacy concerns of the data intensive services' users. People's privacy concerns need to be understood to apply legislation in present-day data collection contexts, characterized by rapid technological change, expanded data collection, diverse uses for collected data, and possibilities to monitor individuals' behavior and combine data from different sources. This understanding can be gained with well-designed *privacy concerns evaluation instruments*. When privacy concerns are evaluated and analyzed, their negative effects on personal data disclosure can be mitigated. In this way, more efficient promotion of personal data intensive services and realization of their benefits for both service users and providers can be reached. In practice, privacy concerns can be addressed by various means based on their evaluation. First, the means of privacy protection and the real risks of data disclosure can be communicated to the data subjects and the general public. Second, data subjects can be given control over their information, and benefits can be offered to them for disclosing information.

Third, privacy-preserving systems and service design can be facilitated by taking service users' privacy concerns into account.

Several researchers have contributed to development of information privacy concerns evaluation instruments since the beginning of the 1990s. Due to evolving technologies and new data collection contexts, the existing instruments do not necessarily match data subjects' privacy concerns anymore. Therefore, the validity of these instruments should be examined for their subsequent development and use. We have addressed this challenge by carrying out an analysis of the most widely used privacy concerns evaluation instruments. Through this analysis, we have gathered information specifically on different aspects of individuals' privacy concerns (referred to as privacy concerns' *dimensions* in the instruments) and how they should be taken into account in the instruments' development. We have identified both privacy concerns' core dimensions that have remained unchanged in the evaluations with time and the types of context dependent dimensions to be incorporated into evaluation instruments. When summarizing the results of our analysis, we pay attention to the fact that in addition to being valid and up-to-date, evaluation instruments should also be made easy-to-use enough. In this way, they can be applied to the practical development of personal data intensive services.

In this article, an overview of the existing privacy concerns evaluation instruments will be given and complemented with an outline of their future development. At first, the historical development of privacy concerns evaluation instruments is described, and the most widely used key evaluation instruments from different decades are introduced. Next, an analysis of these key evaluation instruments is presented, focusing on the privacy concerns' dimensions and their changes with technological development and evolving data collection contexts. After this, recommendations on how to utilize the existing evaluation instruments are given, as well as suggestions for future

research dealing with validation and standardization of the instruments.



BACKGROUND

Opportunities for automatic processing of personal data for business purposes have evolved from the first electronic records in the 1950s to present-day comprehensive data collection and processing systems with different data sources and diverse uses. With technological development enabling this change, companies are showing increasing interest in personal data use for developing their products and services and making their operations more effective.

The rapid progress of information technology in the 1960s enabled big enterprises to, for the first time, establish extensive databanks for their customers' personal data. Later on, data warehousing type systems made it possible to easily combine, process, and analyze the collected data for corporate decision making. Along with these changes, discussion on information privacy was evoked, bringing out the need for personal data protection (cf. Westin, 1967). With the launch of e-commerce and other Internet-based services in the mid-1990s, again, new and expanded opportunities for collecting and utilizing personal data of these systems' users appeared. For example, compared to traditional customer records, customers could now be profiled and their preferences could be identified in more detail based on their clickstream. A few years later—specifically, after the introduction of smart phones—development of location and mobility data-based services gained momentum, creating possibilities for gathering even more detailed and extensive data on individuals and their behavior. As a whole, nowadays, the current technology makes possible large-scale personal data collection, integration of data sources of different types, and combination of separate data pools for diverse uses of the collected data. This enables the production of big data that is highly

valuable to society, as it has the potential to boost both economic growth and utility for individuals. Increased efficiency, quality, and productivity can be reached, and customers' needs can be better met through big data use. On the other hand, it is clear that there are substantive privacy issues associated with the present-day personal data collection. These issues are being regulated by legislation which needs to be adapted to individuals' privacy behavior and personal data collection technologies and contexts. The need for regulation of automatic data processing with ongoing technological development is reflected, for example, by the General Data Protection Regulation reform in the EU (European Commission, 2012).

Evaluation of individuals' information privacy concerns is not necessarily a straightforward task, as people may be concerned about different aspects of privacy (which are usually referred to as privacy concerns *dimensions*). There has been a tendency to learn to understand and take into account privacy concerns for decades, and these concerns have been measured in opinion polls and as a part of research studies as well. However, privacy concerns evaluation instruments and their dimensionality have been developed in a methodical way only since the beginning of the 1990s. Prior to this, information privacy concerns evaluations were relatively fragmented in nature, in that they considered privacy concerns to be either a unidimensional construct or varying in their dimensions. Culnan (1993) presented some burning questions related to secondary use and publishing of personal data for direct marketing purposes in the off-line context. She stated that companies collecting consumers' personal data may find it difficult to pursue the opportunities provided by data collection technologies if they do not comply with information practices responding to consumers' privacy concerns. This challenge has been extensively studied in the literature on information privacy concerns and privacy behavior, indicating that individuals' privacy concerns often lead them to decline to disclose personal data if not being informed of appropriate privacy

practices. For this reason, it is crucial to have well-designed, validated evaluation instruments for information privacy concerns evaluation. The development of these instruments initially focused on collection of off-line personal data such as customers' demographic and purchase data (e.g., Culnan, 1993; Smith et al., 1996; Stewart & Segars, 2002). With the development of e-commerce and Internet-based services, evaluation instruments specific to the Internet context started to become popular (e.g., Malhotra et al., 2004; Dinev & Hart, 2004; Castaneda et al., 2007). These instruments' items were adapted from the existing instruments, taking into account online threats and unforeseen uses of information. As for individuals' location and mobility data collection through smart mobile devices with GPS tracking, questionnaire items on continuous, large-scale mobility data collection were further developed in some studies on privacy behavior in this context (e.g., Junglas et al., 2008; Xu & Gupta, 2009; Raschke et al., 2014). People are exposed all the time to the possibility of increasingly severe privacy losses due to present-day production of big data together with mobile devices' and body sensors' collection of personal data. As data collection of this kind enables continuous monitoring of individuals and their health and physical condition, data that are often considered highly sensitive information, the need for checking the validity of the existing privacy concerns evaluation instruments is increasing.

Information privacy concerns are changing along with evolving personal data collection contexts and the trend towards increasingly comprehensive data collection and modeling of data subjects' behavior. People's attitudes towards data use may change as it becomes a part of our everyday life and we accommodate to the present-day inclusive data collection culture (cf. Nosko et al., 2010). On the other hand, people are continuously exposed to privacy related news and public discussions and hence are increasingly aware of possible privacy issues regarding personal data use. All these aspects should be taken into account when

developing privacy concerns evaluation instruments for new personal data intensive services; in other words, evaluation instruments should be adapted to changes in evaluation contexts and the data disclosure culture.

PRIVACY CONCERNS EVALUATION IN EVOLVING PERSONAL DATA INTENSIVE SERVICES

Present-day and future personal data intensive services are challenging privacy concerns evaluation due to evolving technologies and new data collection contexts. We need to understand existing evaluation instruments' historical evolution to further develop and validate them. Therefore, we present a historical description and analysis of key evaluation instruments and their dimensions in the literature. Our description and analysis covers the most widely used evaluation instruments that date back to different decades and represent different data collection contexts.

Evolution of Privacy Concerns Evaluation Instruments

In the times of electronic records and data warehousing systems, the data collected from customers typically consisted of their contact information and demographic and purchase data. These data were not only used for customer service and companies' operational and strategic aims, they were also often sold for secondary purposes, specifically to direct marketing companies. At the beginning of the 1990s, Culnan (1993) conducted a survey on data subjects' attitudes towards secondary use of their personal data for direct marketing. She identified two dimensions of privacy concerns: *loss of control* over information and *unauthorized secondary use* of information. Culnan's work was followed by an instrument developed by Smith et al. (1996) for evaluating data subjects' concerns about data collecting companies' organizational information practices for personal data use. Smith

et al. identified four dimensions of information privacy concerns, one of them similar to Culnan's dimension of *unauthorized secondary use*. The other three dimensions were *data collection* (i.e., whether excessive data are collected on a data subject), *improper access* to personal information (i.e., within an organization, whether a person without the "need to know" is able to access personal information stored in the files), and *errors* in personal data (e.g., accidental errors or obsolete data). Stewart and Segars (2002) further developed the instrument by Smith et al. Their results suggested that data subjects are concerned about all of the dimensions of organizational information practices simultaneously, rather than any one dimension in particular. Stewart and Segars also found that separate dimensions are interrelated and that the *control* over the information dimension possibly accounts for these interrelationships.

At the beginning of the 2000s, new information privacy concerns evaluation instruments, intended for the Internet context, were introduced. An instrument by Dinev and Hart (2004) for Internet users' privacy concerns evaluation was based on the instruments by Smith et al. and Culnan and Armstrong (1999), and these existing instruments were modified to reflect the nature of the online context. The results by Dinev and Hart suggested an instrument with only two privacy concerns' dimensions: information *finding* on the Internet (i.e., possibilities to track data subjects' activities and their personal information) and information *abuse*. Dinev and Hart also measured *perceived ability to control* information disclosure, recognizing it a separate construct from privacy concerns. However, they pointed out that different results could have been obtained if *need for control* was measured instead of *perceived ability to control*, and that need to control information disclosure may already have been captured by privacy concerns. Almost at the same time as Dinev and Hart, Malhotra et al. (2004) presented their instrument for evaluating Internet users' information privacy concerns related specifically to data disclosure in e-commerce. They proposed three privacy con-



Table 1. Comparison of the key information privacy concerns evaluation instruments

Dimensions Incorporated Into the Instruments	Unauthorized Secondary Use	Control	Collection	Errors	Improper Access	Awareness	Finding	Abuse
Culnan (1993)	x	x						
Smith et al. (1996)	x		x	x	x			
Stewart & Segars (2002)	x	x	x	x	x			
Dinev & Hart (2004)		(x)					x	x
Malhotra et al. (2004)		x	x			x		
Castaneda et al. (2007)	x		x					

cerns' dimensions by drawing on social contract theory: *collection* of personal information, *control* over the collected information, and *awareness* of information use. The items incorporated into the collection dimension were based on Smith et al. and adapted to the Internet context by slightly changing their wording. Castaneda et al. (2007) further developed privacy concerns' measurements in the Internet context. They aimed to distinguish between general Internet privacy concerns and e-commerce web site-specific privacy concerns. Two dimensions were identified for these concerns, namely concern for personal information *collection* and concern for personal information *use*, the latter one corresponding to *unauthorized secondary use*. Castaneda et al.'s instrument items in the *collection* dimension include the aspects of *loss of control* and *improper access*, reflecting the intrusive side of the Internet context.

The recent development of information privacy concerns evaluation comprises themes such as identification of different conceptualizations of privacy concerns (Hong and Thong, 2013), revisiting of existing privacy concerns evaluation (Sipior et al., 2014), privacy concerns' dependency on the types of digital applications for which data are disclosed (Bergström, 2015), and comparison of different privacy concerns evaluation instruments as a part of a privacy behavior model (Fodor and Brem, 2015). All this research reflects the nature of present-day personal data collection, characterized by high volume and continuity of data collection, digital traces, and possibilities of data mining.

It is worth noting that aspects corresponding to privacy concerns' dimensions are defined in data privacy legislation, but from the data controller's (the body collecting data and legally responsible for its processing) viewpoint instead of the viewpoints of individuals whose data are collected. The aforementioned EU privacy regulation aims to minimize personal data collection by requiring specified, predetermined, and legitimate purposes of use (cf. *collection*, *finding*, and *abuse*). It also sets conditions on data disclosure to third parties (cf. *unauthorized secondary use*), imposes data subjects' right of access to their data and right to demand its rectification or removal (cf. *control*), and sets right to obtain rectification of incorrect data (cf. *errors*). Further, the regulation requires implementation of appropriate technical and organizational data protection measures (cf. *improper access*), informed consent for data processing (cf. *awareness*), and provision of data subjects with specified information on processing of their data (cf. *awareness*). It should be borne in mind, however, that not all individuals' information privacy concerns are covered by legislation. This applies specifically to concerns brought up by the new data collection contexts.

A comparison of the described key information privacy concerns evaluation instruments, as regards their dimensions, is presented in Table 1.

Information privacy concerns evaluation instruments by Smith et al. (1996) and Malhotra et al. (2004) have served as standard evaluation instruments in both offline and Internet contexts.

Subsequent studies have often used them as a starting point or reference for their instrument development. The instruments by Smith et al. and Malhotra et al. have been adapted into the evolving data collection contexts by modifying the wording of their items and adding new items relevant to the context. When moving to the Internet context, items originally dealing with companies' data collection and privacy practices in the offline context have been reworded to reflect online data collection. New items on Internet threats incorporated into the instruments deal with stealing or misuse of submitted information, uncertainty about its subsequent use, and continuous tracking of individuals' actions on the Internet. Furthermore, a tendency to adapt instruments better to increasing personal data collection and monitoring of individuals' behavior can be identified in some recent studies (cf. Mao & Zhang 2013, Raschke et al. 2014).

Analysis of Privacy Concerns' Dimensions

Three privacy concerns' dimensions are incorporated into most evaluation instruments regardless of their application contexts: extent of *data collection*, *unauthorized secondary use* of information, and *control* over personal information. Specifically, *collection* is relevant to different kinds of data collection contexts, and it possibly captures the tendency toward increasingly intrusive data collection. It is worth noting that when adapting the instruments to the new contexts, the *collection* dimension has specifically been modified. In the instrument by Smith et al., developed for off-line contexts, the *collection* dimension's items reflect data subjects' views on and reactions to personal data collection by companies, that is, whether data collection and its extent bothers them, and whether they hesitate to disclose their data. Malhotra et al. adapted these items to their online context study. Castaneda et al., instead, based their items on the possibilities of non-transparent and unauthorized data collection in online contexts.

Unauthorized secondary use, which mainly refers to direct marketing purposes in early evaluation instruments, is a relevant dimension in present-day data collection contexts as well due to the diverse uses and possibilities of combining the data. This dimension has remained relatively unchanged despite the transition in data collection contexts. It deals with data use and sharing without the data subject's permission and using of it for other purposes that it was collected for. *Control* over personal information is a key concept related to the definition of information privacy and has hence been a part of privacy concerns evaluations since their inception. It can be said that its role in privacy protection is becoming even more important with the evolution of data collection contexts and individuals' will to decide the types and uses of their data they should disclose. The *control* dimension reflects data subjects' need for control, their perceptions of loss of privacy, and their opinions on the importance of being able to make decisions about their data collection and use.

Unlike the three dimensions described above, some privacy concerns' dimensions seem to be context dependent. *Awareness* about the use and processing of collected data is an aspect closely related to *control* because the ability to control information requires knowledge and understanding about its uses. This dimension likely becomes more important with increasing, more diverse data collection. It is noteworthy that the instruments intended for online contexts do not incorporate dimensions of *improper access* or *errors* that reflect data subjects' demand for procedures to protect personal data and to ensure its accuracy. These dimensions have been highly relevant to earlier offline contexts' data collection with electronic records accessible to companies' employees and prone to errors with regard to data input. In present-day contexts, *improper access* to data can be considered included in concerns about tracking. This idea may support the exclusion of the *improper access* dimension from the instruments when more parsimonious and simple instruments are needed.

In conclusion, it seems that there are some core dimensions to be incorporated even in present and future privacy concerns evaluation instruments. On the other hand, the relevant dimensions may be dependent on the characteristics of the data collection context. Interrelationships between dimensions and their overlapping should also be taken into account when developing evaluation instruments.

SOLUTIONS AND RECOMMENDATIONS

Information privacy concerns evaluation instruments can be used to gain an understanding of users' privacy concerns regarding personal data intensive services. The existing instruments provide a solid, validated base for the development of privacy concerns evaluation. For this reason, they should be used as a starting point for evaluations and then adapted to the context in question. Our analysis showed that information privacy concerns' core dimensions are incorporated into privacy concerns evaluation instruments independently of the data collection context. These core dimensions are extent of *data collection*, *unauthorized secondary use* of information, and *control* over personal information. Adaptation of evaluation instruments to the data collection context can be done by varying the rest of the instruments' dimensions. Privacy concerns' dimensions specific to different data collection contexts typically reflect the extent of data collection, the level of data collection's continuity (i.e., its potential for tracking users' behavior either online or in a physical environment), and diversity of subsequent uses of the collected data. It is worth noting that in continuously evolving data collection contexts, new privacy concerns' dimensions may emerge regarding the context's nature and, for example, perceived sensitivity of data. In our research, we have identified the studies presented in this article as the key studies

involving evaluation instruments that should be taken into account in privacy concerns evaluation.

FUTURE RESEARCH DIRECTIONS

This article has presented an analysis of key privacy concerns evaluation instruments and their dimensions. Due to technological changes and new data collection contexts, privacy concerns' dimensions should be investigated even in future research, specifically to identify the context-specific dimensions to be applied in the evaluations. As the literature contains a large number of studies on information privacy concerns evaluation, the present analysis should be advanced by analyzing these studies systematically. In addition to the scientific literature, it would be reasonable to extend the analysis to the privacy aspects present in legislation. Similarly, people's changing attitudes towards personal data collection as a part of cultural change should be taken into account in future research. In this way, a validated, standardized, and easy-to-use evaluation framework could be constructed for facilitating application of knowledge to different and new data collection contexts. Relations between separate privacy concerns' dimensions could be another relevant topic of research in order to determine whether different dimensions of privacy concerns always need to be incorporated into the evaluation instruments or whether there are situations in which simplified instruments could be used.

CONCLUSION

This article explored how existing information privacy concerns evaluation instruments could be adapted to match data subjects' privacy concerns in evolving data collection contexts. An overview was presented of the historical development of the information privacy concerns evaluation instruments, spanning a period from the beginning of

the 1990s, with its data warehousing systems, to the 2000s, with its continuous monitoring of individuals' actions on the Internet. This overview was followed by an analysis of the key evaluation instruments, which provided insight into their validity for present-day personal data collection contexts. The analysis focused on the dimensions incorporated into privacy concerns evaluation instruments. It showed that these dimensions have changed with evolving data collection contexts and the tendency toward extensive, continuous personal data collection. However, there are core dimensions that seem to be valid even for current data collection contexts. These dimensions, namely extent of *data collection*, *unauthorized secondary use* of information, and *control* over personal information, can be used as a starting point for privacy concerns evaluation. Evaluation should then be adapted to the data collection context in question by varying the instrument's dimensions. Regarding future research, it is suggested to start construction of a validated, standardized, and easy-to-use evaluation framework for facilitating privacy concerns evaluations in different contexts.

REFERENCES

- Bergström, A. (2015). Online privacy concerns: A broad approach to understanding the concerns of different groups for different uses. *Computers in Human Behavior*, 53, 419–426. doi:10.1016/j.chb.2015.07.025
- Castaneda, J. A., Montoso, F. J., & Luque, T. (2007). The dimensionality of customer privacy concern on the Internet. *Online Information Review*, 31(4), 420–439. doi:10.1108/14684520710780395
- Culnan, M. J. (1993). How did they get my name? An exploratory investigation of consumer attitudes toward secondary information use. *Management Information Systems Quarterly*, 17(3), 341–363. doi:10.2307/249775
- Culnan, M. J., & Armstrong, P. K. (1999). Information privacy concerns, procedural fairness, and impersonal trust: An empirical investigation. *Organization Science*, 10(1), 104–115. doi:10.1287/orsc.10.1.104
- Dinev, T., & Hart, P. (2004). Internet privacy concerns and their antecedents—Measurement validity and a regression model. *Behaviour & Information Technology*, 23(6), 413–422. doi:10.1080/01449290410001715723
- European Commission. (2012). *Proposal for a regulation of the European Parliament and of the Council on the Protection of Individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation), COM (2012) 11 final*. Brussels: European Commission.
- Hong, W., & Thong, J. Y. L. (2013). Internet privacy concerns: An integrated conceptualization and four empirical studies. *Management Information Systems Quarterly*, 37(1), 275–298.
- Malhotra, N. K., Kim, S. S., & Agarwal, J. (2004). Internet users information privacy concerns (IUIPC): The construct, the scale, and a causal model. *Information Systems Research*, 15(4), 336–355. doi:10.1287/isre.1040.0032
- Mao, E., & Zhang, J. (2013). The role of privacy in the adoption of location-based services. *Journal of Information Privacy and Security*, 9(2), 40–59. doi:10.1080/15536548.2013.10845678
- Nosko, S., Wood, E., & Molema, S. (2010). All about me: Disclosure in online social networking profiles: The case of FACEBOOK. *Computers in Human Behavior*, 26(3), 406–418. doi:10.1016/j.chb.2009.11.012
- Raschke, R. L., Krishen, A. S., & Kachroo, P. (2014). Understanding the components of information privacy threats for location-based services. *Journal of Information Systems*, 28(1), 227–242. doi:10.2308/isyss-50696

Sipior, J. C., Ward, B. T., & Connolly, R. (2013). Empirically assessing the continued applicability of the IUIPC construct. *Journal of Enterprise Information Management*, 26(6), 661–178. doi:10.1108/JEIM-07-2013-0043

Smith, H. J., Milberg, J. S., & Burke, J. S. (1996). Information privacy: Measuring individuals concerns about organizational practices. *Management Information Systems Quarterly*, 20(2), 167–196. doi:10.2307/249477

Stewart, K. A., & Segars, A. H. (2002). An empirical examination of the concern for information privacy instrument. *Information Systems Research*, 13(1), 36–49. doi:10.1287/isre.13.1.36.97

Westin, A. F. (1967). *Privacy and freedom*. New York: Atheneum.

ADDITIONAL READING

Acquisti, A. (2009). Nudging privacy: The behavioral economics of personal information. *IEEE Security and Privacy*, 7(6), 82–85. doi:10.1109/MSP.2009.163

Chellappa, R. K., & Sin, R. G. (2005). Personalization versus privacy: An empirical examination of the online consumers dilemma. *Information Technology Management*, 6(2-3), 181–202. doi:10.1007/s10799-005-5879-y

Chen, K., & Rea, A. I. (2004). Protecting personal information online: A survey of user privacy concerns and control techniques. *Journal of Computer Information Systems*, 44(4), 85–92.

Eastlick, M. A., Lotz, S. L., & Warrington, P. (2006). Understanding online B-to-C relationships: An integrated model of privacy concerns, trust, and commitment. *Journal of Business Research*, 59(8), 877–886. doi:10.1016/j.jbusres.2006.02.006

European Commission (1995). Directive 95/46/EC. Official Journal of the European Communities (No L 281/31).

Li, Y. (2011). Empirical studies on online information privacy concerns: Literature review and integrative framework. *Communications of the Association for Information Systems*, 28, 453–496.

Li, Y. (2012). Theories in online information-privacy research: A critical review and an integrated framework. *Decision Support Systems*, 54(1), 471–481. doi:10.1016/j.dss.2012.06.010

Liu, C., Marchewka, J. T., Lu, J., & Yu, C.-S. (2005). Beyond concern—A privacy-trust-behavioral intention model of electronic commerce. *Information & Management*, 42(2), 289–304. doi:10.1016/j.im.2004.01.003

Liu, Z., Bonazzi, R., Fritscher, B., & Pigneur, Y. (2011). Privacy-friendly business models for location-based mobile services. *Journal of Theoretical and Applied Electronic Commerce Research*, 6(2), 90–107. doi:10.4067/S0718-18762011000200009

Lwin, M., Wirtz, J., & Williams, J. D. (2007). Consumer online privacy concerns and responses: A power-responsibility equilibrium perspective. *Journal of the Academy of Marketing Science*, 35(4), 572–585. doi:10.1007/s11747-006-0003-3

Son, J. Y., & Kim, S. S. (2008). Internet users' information privacy-protective responses: A taxonomy and a nomological model. *Management Information Systems Quarterly*, 32(3), 503–529.

The Privacy Act of 1974, 5 U.S.C. § 552a (1974).

U.S. Department of Health, Education & Welfare (1973). Records, computers and the rights of citizens. Report of the Secretary's Advisory Committee on automated personal data systems. Chapter IV: Recommended safeguards for administrative personal data systems. Washington, DC: U.S. Government Printing Office.

Van Slyke, C., Shim, J. T., Johnson, R., & Jiang, J. J. (2006). Concern for information privacy and online consumer purchasing. *Journal of the Association for Information Systems*, 7(6), 415–444.

Zhou, T. (2011). The impact of privacy concerns on user adoption of location-based services. *Industrial Management & Data Systems*, 111(2), 212–226. doi:10.1108/02635571111115146

KEY TERMS AND DEFINITIONS

Information Privacy: An individual's capability to control disclosure of information about him/her to others by determining the type, extent, uses, and users of the data to be disclosed.

Information Privacy Concerns: An individual's concerns about collection, processing,

and use of information about him/her, and about the related consequences.

Personal Data: Any information relating to an identified or identifiable natural person (“data subject”); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural, or social identity (EC Directive 95/46/EC).

Personal Data Intensive Services: Services substantially based on collection, processing, and utilization of users' personal data for service provision. These services can be produced by private companies, public sector organizations, or non-governmental parties.

Privacy Concerns Evaluation Instrument: A way or method to measure individuals' privacy concerns, for example, a survey.



Digital Video Watermarking Using Diverse Watermarking Schemes

Yash Gupta

Maulana Abul Kalam Azad University of Technology, India

Shaila Agrawal

Maulana Abul Kalam Azad University of Technology, India

Susmit Sengupta

Maulana Abul Kalam Azad University of Technology, India

Aruna Chakraborty

Maulana Abul Kalam Azad University of Technology, India

INTRODUCTION

With the advent of internet in 1967, it has revolutionized the fields of communication and computer unlike anything before (Maity & Kundu, 2002). Since then it has grown exponentially and has become a vast information reserve. Nowadays people prefer to search the internet than looking into any book, to gain knowledge on the subject that intrigues them.

Internet has now become the easiest way of sharing information and with the growth of social networking sites, doing the latter with the masses has become even easier and it can be done quite briskly. People now buy storage spaces over the internet so that they can have access to their works from anywhere in the world. With the increase of information and digital content over the internet the need and necessity of multimedia security and copyright protection arises (Agrawal, Gupta, & Chakraborty, 2015; Natarajan & Makhdumi, 2009). So in order to stop theft and lose of fidelity of digital content we need to develop techniques to safeguard the digital contents (Lai, & Tsai, 2010). Digital Watermarking is one such technique that is used for copyright protection of digital media.

BACKGROUND

What Is Digital Watermarking?

Copyright Protection incorporates a logo or some ownership information into the digital media without affecting its perceptibility (Agrawal, Gupta, & Chakraborty, 2015; Yeo, & Yeung, 1997). Hence, in case of a conflict, the logo can effectively be extracted from the digital media in order to claim the ownership rights. Watermarking is a process of embedding some data called the watermark or the digital signature into the digital media (Sinha, Bardhan, Pramanick, Jagatramka, Kole, & Chakraborty, 2011). Here the researchers will primarily focus on Digital Video Watermarking.

What Is Digital Video Watermarking?

Digital Video Watermarking is a method of copyright protection of videos in which a watermark is added to the original video without affecting its perceivable quality (Yeo, & Yeung, 1997; Doerr, & Dugelay, 2003). For a watermarking scheme to be used for copyright protection, it should fulfill two criterions i.e. it must be robust against attacks like signal processing and lossy compression and

it should not lead to loss of fidelity (Al-Khatib, Al-Haj & Lama-Rajab, 2008). Some watermarking techniques require the original video for the detection of the watermark. This is called a non-blind watermark detection method while there are some watermarking techniques which do not require the original video for watermark detection (Maity & Kundu, 2002). This is called blind watermark detection method and is usually preferred.

Types of Watermarks

As per human perception the watermark can be of three types, visible, invisible and dual. (Potkar, & Ansari, 2014). As the name suggests, a visible watermark is perceivable but an invisible watermark is not. For e.g. a watermark is to be added to an image or a video then in case of a visible watermark, the watermark can be seen on the image or on the video frames but in case of invisible watermarks, a layman would not be able to differentiate whether a watermark has been added to the digital media or not. A Dual watermark incorporates both visible and invisible watermarks to the video. Here the invisible watermark is kept as a backup for the visible watermark. The researchers' can choose anyone of them as per our requirements and necessities.

Factors to be Considered While Embedding a Watermark on a Video

While embedding a digital image watermark on the video the following things should be kept in mind:

1. **Imperceptibility:** The watermark should not be visible in the copyrighted video (Maity, & Kundu, 2011).
2. **Robustness:** It should be impossible for the attacker to detect or extract or manipulate the watermark (Maity, & Kundu, 2011).
3. **Fidelity:** The perceivable quality of video should not degrade.
4. **Security:** It should be impossible to generate a duplicate of the authentic watermark

to claim false ownership. The watermark should also be non-invertible.

5. **Verification:** The watermark can be extracted from the watermarked video to prove ownership.
6. **Constant Bit Rate:** As the transmission channel bandwidth has to be obeyed so the watermarked video should also have the same bit rate.

Classification of Video Watermarking

Many Video watermarking schemes have been proposed till date, while some are deployed on uncompressed videos, the others are deployed on the compressed versions (Hartung, & Girod, 1998; Meng, & Chang, 1998). Video Watermarking is classified into two categories based on the methods of embedding the watermark bits in the host video. The two categories are:

1. **Spatial Domain Technique:** In this method the watermark can be embedded and detected by directly manipulating the pixel intensity values of the video frame (Sinha, Bardhan, Pramanick, Jagatramka, Kole, & Chakraborty, 2011).
2. **Transform Domain Technique:** Here the spatial pixel values of the host video are altered according to some predefined algorithm. The watermark is dispersed throughout the host video and this makes the technique robust against malicious geometric attacks (Sinha, Bardhan, Pramanick, Jagatramka, Kole, & Chakraborty, 2011).

Thus, transform domain watermarking schemes ensures more imperceptibility, randomness in the distribution of the watermark and have also proven to enhance robustness against geometric attacks. Hence, it is better to use the transform domain technique rather than the spatial domain technique (Sinha, Bardhan, Pramanick, Jagatramka, Kole, & Chakraborty, 2011).



Types of Attacks on Video Watermarks

A successful attack is one in which the watermark has been modified in a manner to prevent its successful extraction without affecting the quality of the video significantly. A successful attack does not mean that the digital media cannot be restored to its original state but it means that watermark detection and extraction techniques are no longer useful for claiming ownership rights as the watermark has been modified (Potkar & Ansari, 2014). The attacks are of two types (Singh & Chadha, 2013):

1. **Intentional Attacks:** These attacks are done deliberately to prevent successful extraction of the watermark. Types of Intentional attacks are:
 - a. **Geometric Attacks:** Here minor geometric distortions are made to the video frames to de-synchronize the extraction procedure. The changes induced in the video frame can be negligibly perceived but it prevents successful extraction of the watermark (Potkar & Ansari, 2014). It instills attacks like rotation, scaling, etc. to the video frames. This type of attacks usually affects each frame of the video (Voloshynovskiy, Pereira & Pun, 1999; Husain, 2012).
 - b. **Statistical Attacks:** Videos are susceptible to attacks which are not applicable in the case of images. These attacks take advantage of this inherent redundancy between the frames to remove the watermark. Attacks like frame averaging, frame swapping and statistical analysis comes under this head (Potkar & Ansari, 2014).
 - c. **Protocol Attacks:** This type of attacks try to destroy the digital content itself rather than trying to remove or destroy the watermark. Examples are Copy attack and watermark inversion.

In copy attack the watermark is copied from one frame to another without any knowledge of the watermarking technique used. A watermarking scheme is not resistant to copy attack then the watermark used for verification cannot be trusted as it might have been copied from somewhere else (Potkar & Ansari, 2014).

- d. **Removal Attacks:** This type of attacks usually employ filtering techniques to remove the embedded watermark. Its examples are spatial averaging filter, Wiener de-noising filter (Potkar & Ansari, 2014).
2. **Unintentional Attacks:** These attacks are introduced into the system by normal signal processing and compression but are not done intentionally. Types of Unintentional attacks are (Potkar & Ansari, 2014):
 - a. Analog to Digital and Digital to Analog Conversion.
 - b. Scaling and Cropping Operations.
 - c. Frame Rate Conversion.
 - d. Aspect Ratio Conversion.
 - e. Quantization,
 - f. Compression, and
 - g. Noise Addition.

Applications of Video Watermarking

According to Van Huyssteen's thesis the applications of video watermarking are numerous but they can be brought down under six heads, they are:

1. **Transaction Tracking:** It is used to track the manner in which the content was distributed in the system or through multiple points. In situations where illegal distribution of data occurs then it should be ideally possible for us to track the source from where the distribution has occurred in order to identify the misappropriating party. This is done by embedding a unique identifier into the media at the time of playback which can be

extracted at a later time (Potkar & Ansari, 2014; Singh, Jain & Jain, 2013).

2. **Broadcast Monitoring:** It is used by the broadcasters or content owners to check if their content had been completely broadcasted or just a part of the material had been used. The watermark is automatically extracted to verify the latter. The watermark is usually embedded by the content owner itself while it is extracted either by monitoring sites in the broadcast chain or by a third party at the receiving end (Potkar & Ansari, 2014; Singh & Chadha, 2013).
3. **Content Authentication:** The content is watermarked with a semi-fragile watermark, which is designed to be affected by signal transformations. So on tampering with the content the watermark gets either destroyed or modified and can be used to prove that the content is not authentic. It is used for detecting tampering of the content (Potkar & Ansari, 2014; Singh, Jain & Jain, 2013).
4. **Ownership Identification:** The watermark is extracted from the digital content and verified to prove or claim ownership rights in case of any conflicts. The pirates may also embed their own watermark to the digital content and in such a situation it becomes quite difficult to tell that which one is the original watermark. This phenomenon is also termed as ownership deadlock problem (Potkar & Ansari, 2014; Singh, Jain & Jain, 2013).

Video Watermarking Schemes Currently in Use

The transform domain techniques which are commonly being used today are Discrete Wavelet Transform (DWT), Discrete Cosine Transform (DCT), Discrete Fourier Transform (DFT) (Jiansheng, Sukang & Xiaomei, 2009; Chaturvedi & Basha, 2012; Rahman, 2013). There are other techniques as well like Discrete Walsh Transform, Discrete Hartley Transform (DHT), Discrete

Kekre Transform (DKT), Singular Value Decomposition, Principal Component Analysis, etc.

These techniques are often combined to form new video watermarking schemes so as to produce better results. For example DCT along with DWT (Jiansheng, Sukang & Xiaomei, 2009), or DWT along with SVD and PCA (Sinha, Bardhan, Pramanick, Jagatramka, Kole, & Chakraborty, 2011) forms different watermarking schemes and the list continues. The basis of comparison of these techniques is by calculating the Peak Signal to Noise Ratio (PSNR) values and Normalization Coefficient (NC) value. PSNR gives us a measure about the imperceptibility of the watermark in an attacked video frame. The higher the PSNR values higher is the imperceptibility (Sinha, Bardhan, Pramanick, Jagatramka, Kole, & Chakraborty, 2011). NC is used to find the similarity between the extracted and original watermark. Its peak value is 1. It is used to make judgments on extraction fidelity (Sinha, Bardhan, Pramanick, Jagatramka, Kole, & Chakraborty, 2011). So the higher the PSNR and NC values, better is the watermarking scheme.

Discrete Wavelet Transform

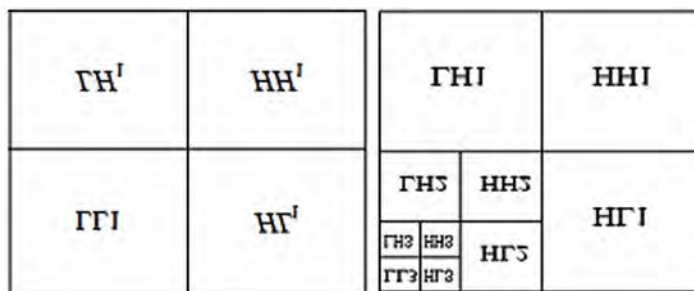
Discrete Wavelet Transform or (DWT) is used widely in the field of signal processing. A two-dimensional DWT is a combination of two single 1-D DWT's applied to both the horizontal and vertical directions. 2-D DWT is used to decompose the image into lower resolution approximation sub-band (LL) as well as horizontal (HL), vertical (LH) and diagonal (HH) detail components, (Figure 1). The embedding process is carried out in the LL sub-band to make the watermarked image withstand lossy compression (Shaikh, Khan & Kelkar, 2012).

Discrete Fourier Transform

Discrete Fourier Transform or DFT is one of the oldest techniques of image compression. A Fourier Transform converts a signal into sinusoids. It takes



Figure 1. DWT sub-bands



as input a time-continuous signal and its output is a signal in frequency-domain. However, DFT takes as input samples of the signal in time-domain and converts them into samples in frequency-domain. DFT is easily calculated in computers using Fast Fourier Transform or FFT algorithms. The time complexity in this calculation is $O(N \log N)$.

Discrete Cosine Transform

Discrete Cosine Transform (DCT) is another commonly used transform technique in signal processing. It aims at converting an image from its spatial domain to frequency domain in order to make it robust against different attacks like contrast adjustment, low pass filtering, etc. (Chaturvedi & Basha, 2012). Discrete Cosine transform is defined by the equations, (1) and (2).

$$f(mn) = a(j)a(k) \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} f(mn) \cos\left[\frac{(2m+1)j\pi}{2N}\right] \cos\left[\frac{(2n+1)k\pi}{2N}\right] \quad (1)$$

and the Inverse Discrete Cosine transform is given as:

$$f(mn) = \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} a(j)a(k) f(mn) \cos\left[\frac{(2m+1)j\pi}{2N}\right] \cos\left[\frac{(2n+1)k\pi}{2N}\right] \quad (2)$$

Singular Value Decomposition

Singular Value Decomposition is a technique for expressing a matrix as a product of a diagonal matrix and two orthogonal matrices (Seema, 2012). The SVD of a given image A in the form of a matrix is defined by equation (3).

$$A = USV^T \quad (3)$$

where U and $V \in \mathbb{R}$ and are $N \times N$ unitary matrices and $S \in \mathbb{R}$ with dimensions $N \times N$ is a diagonal matrix.

COMPARISON BETWEEN DIFFERENT SCHEMES

1. **DFT and DWT:** DFT is helpful in determining the frequency components present in a

Figure 2. Representation of singular value decomposition

$$SVD(A) = \begin{bmatrix} U_{1,1} & \dots & U_{1,n} \\ U_{2,1} & \dots & U_{2,n} \\ \vdots & \vdots & \vdots \\ U_{n,1} & \dots & U_{n,n} \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \sigma_m \end{bmatrix} \begin{bmatrix} V_{1,1} & \dots & V_{1,n} \\ V_{2,1} & \dots & V_{2,n} \\ \vdots & \vdots & \vdots \\ V_{n,1} & \dots & V_{n,n} \end{bmatrix}^T$$



signal. It can even be used on a signal of variable frequency components. However, it does not show the temporal localization of these frequency components. This can be achieved by using DWT which illustrates which frequency component of the signal is present at a given time. DWT can, thus, also be applied on non-stationary signals. Hence, DWT has replaced DFT in most of the applications because of its extended usability. The PSNR value from a watermarking scheme using DWT is better than that of a scheme using DFT.

2. **DWT and SVD:** DWT is the outcome of the electrical treatment being given to an image. SVD is a procedure applied to a matrix to approximate some values in the final SVD matrix with a view of minimizing the useful data part of the original matrix. Here, an image is treated on the basis of its storage in a system, i.e., in the form of a matrix. It is to be noted that the original data is subject to manual alterations and the SVD procedure may not always yield the desired results. The same is not true for DWT where the user cannot meddle with the electrical signal directly.

Conventional Watermarking Methodology

The original watermarking methodology involved the application of a single decomposition technique with the algorithm to embed the watermark. Eventually, a combination of two or more techniques began being used to yield better results in terms of the PSNR. The combinations included DFT-DCT, DWT-DCT, DWT-SVD, etc. (Podilchuk, & Delp, 2001). Here, one decomposition algorithm followed the other. However, in video watermarking applications, the same watermarking scheme(s) were used for all video frames, with the fidelity of the scheme being same for all the video frames. Thus, the researchers have developed a new idea to watermark digital videos with great diversity in the schemes.

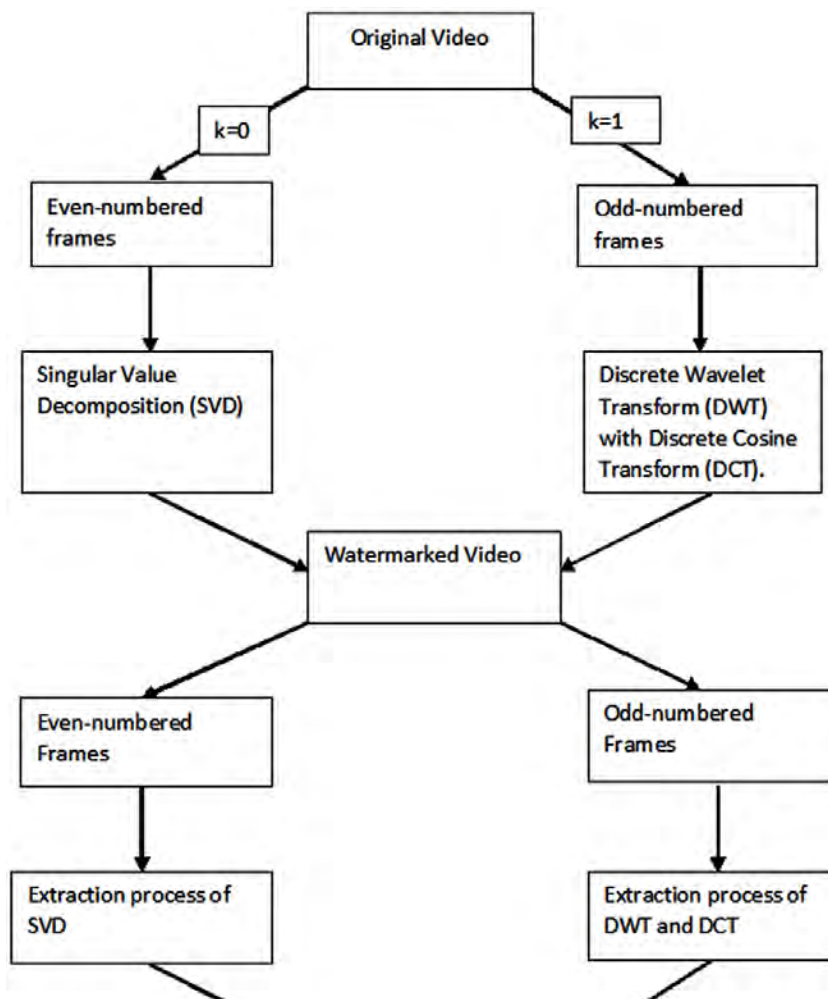
Method Used

The researchers realized that the inherent property of videos being composed of frames could be used to improve the efficiency of watermarking algorithm in terms of its PSNR and NC values (Agrawal, Gupta, & Chakraborty, 2015). Thus, a Composite Watermarking Technique had been proposed where different watermarking scheme(s) could be applied to different sets of video frames. Then, the watermark could be extracted from each set of frames for all of them to be superimposed to give the final extracted watermark. The number of different sets of frames in which the video is divided and the manner in which the sets are formed depends on whatever seems suitable.

Suppose a video contains ' N ' frames and it is to be divided into ' n ' number of sets. There is a variable ' i ' whose value ranges from 1 to N and a variable ' k ' whose value is given by $(i \bmod n)$ and its value ranges from 0 to $n-1$. Thus, on performing modular function on every i^{th} frame, we obtain a value of k by the equation, $k = i \bmod n$ and allot i to the k^{th} set. This gives us n distinct sets $(0 - (n-1))$. Then for each different set, a different watermarking scheme(s) would be applied to all frames in it.

The researchers have implemented this watermarking technique by dividing the video into two sets of frames, to obtain two values of k as 0 and 1. The set corresponding to $k = 0$ is the set of even numbered frames and that corresponding to $k = 1$ is the set of odd numbered frames. Here, DWT has been applied to the odd numbered frames with a DCT transformed watermark while another scheme of SVD is applied to the even numbered frames. The extraction process again follows the same approach for dividing the frames and each set has its own extraction procedure. The final extracted watermark is obtained by superimposing the individual watermarks obtained from the different sets (Agrawal, Gupta, & Chakraborty, 2015). The algorithms for the aforesaid scheme are given in the next section.

Figure 3. The flow chart for the proposed method with $n=2$



Watermarking Algorithms

The algorithms proposed for embedding and extracting the watermark are given below (Agrawal, Gupta, & Chakraborty, 2015).

Step 1: Watermark Embedding

Step 1: Input the original video and extract the individual frames.

Step 2: For odd,-numbered frames follow algorithm 1, cited by Chaturvedi, N., & Basha, D. S. (2012). Comparison of Digital Image watermarking Methods DWT & DWT-DCT on the Basis of PSNR. *International Journal of Innovative Research in Science, Engineering and Technology*, 1(2).

Step 3: For even-numbered frames follow algorithm 2, based on SVD (Rahman, 2013; Seema, 2012).

Step 4: Combine the outputs obtained from step 5 of algorithm 1 and from step 6 of algorithm 2 to obtain the watermarked video.

Algorithm 1

Step 1: Input the watermark, change it to YUV form (Y component– stands for luminance, i.e. brightness of the color, U and V components determines the color) (gray scale) from RGB (Red - Green – Blue color model) form and extract the luminance or the Y component of it.

Step 2: Apply 2-D DCT to the extracted Y component.

Step 3: For every odd numbered frame, change it from RGB form to YUV form and apply 2-D DWT to the Y component of it.

Step 4: Resize the 2-D DCT transformed watermark according to the LL part of the image and embed it into the LL part with a strength alpha by equation (4).

$$LL1 = LL + \alpha * WM \quad (4)$$

Here, LL is the approximation image obtained from 2-D DWT, WM is the resized watermark matrix and LL1 is the watermarked approximation image.

Step 5: Reconstruct the original frame from LL1 by applying inverse 2-D DWT to the Y component and changing it back to RGB form.

Algorithm 2

Step 1: Input the watermark, change it from RGB to YUV form and apply SVD on the Y component of it.

Step 2: To the even numbered frames extracted from step 1 of algorithm 1, change the color format from RGB to YUV and extract the Y component of it.

Step 3: Apply SVD on the Y component of the frames.

Step 4: Embed the watermark into the frames with strength alpha (same as that in algorithm 1), using equation (5).

$$Sf1 = Sf + \alpha * Sw \quad (5)$$

Here, Sf is the singular matrix obtained from single value decomposition of the video frame, Sw is the singular matrix obtained through single value decomposition of the watermark and Sf1 is the final singular matrix.

Step 5: Reconstruct the Y component of the watermarked frame by applying equation (6).

$$Y1 = Uf * Sf1 * Vf' \quad (6)$$

where, Uf and Vf are the matrices obtained by SVD of the video frame.

Step 6: Change the watermarked frame obtained from YUV to RGB format.

Step 2: Watermark Extraction

Step 1: Input the watermarked video, together with the matrices, Uw and Vw obtained by performing SVD on the watermark image.

Step 2: Extract the individual frames. For odd-numbered frames, follow algorithm 1 for extraction, cited by Chaturvedi & Basha 2012.

Step 3: For the even-numbered frames, follow algorithm 2, cited in (Shaikh, Khan & Kelkar, 2012; Lai & Tsai, 2010).

Step 4: Resize the image matrices obtained from step 5 of algorithm 1 and from Step 5 of algorithm 2 to match their dimensions.

Step 5: Add the two images to obtain the final watermark image.

Algorithm 1

Step 1: Convert the frames from RGB TO YUV color format and separate the Y component.

Step 2: Perform 2-D DWT on the Y component.

Step 3: From the LL1 component obtained from step 3, obtain the watermark by the following equation.

$$WM = (LL1 - LL) / \alpha \quad (7)$$

Here, WM is the embedded watermark component and LL is the Y component of the original video frame.

Step 4: Perform inverse 2-D DCT on the output of step 4 to obtain the Y component of the watermark.

Step 5: Convert the YUV watermark image to RGB format.

Algorithm 2

Step 1: From the extracted even-numbered frames, convert the color format of the image from RGB to YUV.

Step 2: Obtain the Y component and perform SVD on it to get $Uf1$, $Sf1$ and $Vf1$.

Step 3: From $Sf1$, obtain Sw using the equation,

$$Sw = (Sf1 - Sf) / \alpha \quad (8)$$

Step 4: From Sw obtained from the previous step, reconstruct the Y component using the following equation,

$$Y = Uw * Sw * Vw' \quad (9)$$

where, Uw and the Vw are the matrices obtained from the SVD of the original watermark in the embedding process.

Step 5: Convert the image from YUV to RGB color format.

ALGORITHM RESULTS

The performance of the method mentioned above is measured in terms of the imperceptibility and robustness against possible attacks like filtering, noise addition, geometric attacks. The results are given in Table 1.

Peak Signal to Noise Ratio

The Peak Signal to Noise Ratio (PSNR) is a measure of the imperceptibility of the watermark. A high value of PSNR indicates more imperceptibility. The PSNR is calculated by (10).

$$PSNR = 10 \log_{10} (MAX_i^2 / MSE) \quad (10)$$

where MAX_i is the maximum possible value of a pixel in an image, MSE is the mean squared error which is calculated by (11).

$$MSE = \left(\frac{1}{mn} \right) \sum_{i=0}^m \sum_{j=0}^n [I(i, j) - I'(i, j)]^2 \quad (11)$$

where I and I' are pixel values at location (i, j) of the original and the extracted frames respectively.

Normalized Correlation Coefficient

The normalized coefficient (NC) is used to determine the robustness of the watermarking and has a peak value of 1 (Sinha, Bardhan, Pramanick, Jagatramka, Kole, & Chakraborty, 2011). The formula for calculating NC is given in (12).

$$NC = \frac{\sum_i \sum_j W(i, j) W'(i, j)}{\sqrt{\left(\sum_i \sum_j W(i, j) \right) \sqrt{\left(\sum_i \sum_j W'(i, j) \right)}}} \quad (12)$$

Where W and W' represents the original and extracted watermarks respectively.

Algorithm 1 and 2 are applied on a video, with the watermark and the extracted watermark is obtained.

The value of PSNR and NC for the proposed method is found to be 65.073dB and 0.547 respectively. These are the values without noise or any other attack. The values of both PSNR and NC under some of the attacks are given in Table 1.

FUTURE RESEARCH DIRECTIONS

The method used by the researchers holds immense potential in furthering researches related to the field of digital video watermarking. The use of composite methodology is highly promising. Many different schemes can be combined based

Table 1. Values of PSNR and NC under different attacks

Attacks	PSNR (dB)	NC
Gaussian Noise	53.397	0.509
Median Filtering	61.356	0.530
Poisson Noise	65.097	0.547
Salt & Pepper Noise	47.749	0.349
Sharpening Filter	45.532	0.968

on their productivity and also their compatibility with each other. The value of n can be varied to accommodate as many such combinations as required. When $n=1$, the same scheme can be applied to all the frames. Thereafter, combinations can be used.

CONCLUSION

The method used is potent of diversifying the use of the same set of schemes by applying them in the same or different sets of frames. The values of PSNR and NC are different for such diversification. For example, DWT and SVD can be applied to one set of frames in the given algorithm followed by DCT in another set of algorithms. The result is not the same as that for the demonstrative algorithm. The whole idea behind the use of such technology is in the improvement of the quality of the extracted watermark. Most of the attacks are more destructive for a set of schemes and cause negligible changes in the other. The choice of schemes should be such that the effects of different attacks cause different degrees of modification, so that the extracted watermarks from different sets could together yield a more resilient watermark.

REFERENCES

- Agrawal, S., Gupta, Y., & Chakraborty, A. (2015, October). A Composite Approach to Digital Video Watermarking. In *International Congress on Information and Communication Technology*. ICICT.
- Al-Khatib, T., Al-Haj, A., & Lama Rajab, H. M. (2008). *A Robust Video Watermarking Algorithm 1*. Academic Press.
- Chaturvedi, N., & Basha, D. S. (2012). Comparison of Digital Image watermarking Methods DWT & DWT-DCT on the Basis of PSNR. *International Journal of Innovative Research in Science, Engineering and Technology*, 1(2). Retrieved from ece.mits.ac.in
- Doerr, G., & Dugelay, J. L. (2003). A guide tour of video watermarking. *Signal Processing Image Communication*, 18(4), 263–282. doi:10.1016/S0923-5965(02)00144-3
- Doërr, G., & Dugelay, J. L. (2004). Security pitfalls of frame-by-frame approaches to video watermarking. *Signal Processing. IEEE Transactions on*, 52(10), 2955–2964. doi:10.1109/TSP.2004.833867
- Hartung, F., & Girod, B. (1998). Watermarking of uncompressed and compressed video. *Signal Processing*, 66(3), 283–301. doi:10.1016/S0165-1684(98)00011-5
- Husain, F. (2012). A survey of digital watermarking techniques for multimedia data. *Int'l Journal of Electronics and Communication Engineering*, 2(1), 37–43.
- Jiansheng, M., Sukang, L., & Xiaomei, T. (2009, May). A digital watermarking algorithm based on DCT and DWT. In *International Symposium on Web Information Systems and Applications (WISA'09)* (pp. 104-107). Academic Press.



- Lai, C. C., & Tsai, C. C. (2010). Digital image watermarking using discrete wavelet transform and singular value decomposition. *Instrumentation and Measurement. IEEE Transactions on*, 59(11), 3060–3063.
- Langelaar, G. C., Setyawan, I., & Lagendijk, R. L. (2000). Watermarking digital image and video data. A state-of-the-art overview. *IEEE Signal Processing Magazine*, 17(5), 20–46. doi:10.1109/79.879337
- Lin, Y. R., Huang, H. Y., & Hsu, W. H. (2006, August). An embedded watermark technique in video for copyright protection. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on* (Vol. 4, pp. 795-798). IEEE.
- Maity, S. P., & Kundu, M. K. (2002, December). Robust and Blind Spatial Watermarking. In *Digital Image. ICVGIP*.
- Maity, S. P., & Kundu, M. K. (2011). Performance improvement in spread spectrum image watermarking using wavelets. *International Journal of Wavelets, Multiresolution, and Information Processing*, 9(01), 1–33. doi:10.1142/S0219691311003931
- Meng, J., & Chang, S. F. (1998, October). Embedding visible video watermarks in the compressed domain. In *Image Processing, 1998. ICIP 98. Proceedings. 1998 International Conference on* (Vol. 1, pp. 474-477). IEEE.
- Natarajan, M., & Makhdumi, G. (2009). Safeguarding the digital contents: Digital watermarking. *DESIDOC Journal of Library & Information Technology*, 29(3), 29–35. doi:10.14429/djlit.29.249
- Podilchuk, C., & Delp, E. J. (2001). Digital watermarking: Algorithms and applications. *Signal Processing Magazine, IEEE*, 18(4), 33–46. doi:10.1109/79.939835
- Potkar, A. N., & Ansari, S. M. (2014). Review on Digital Video Watermarking Techniques. *International Journal of Computers and Applications*, 106(11).
- Rahman, M. (2013). *A dwt, dct and svd based watermarking technique to protect the image piracy*. arXiv preprint arXiv:1307.3294
- Seema, S. S. (2012). DWT-SVD based efficient image watermarking Algorithm to achieve high robustness and perceptual quality. *International Journal of Advanced Research in Computer Science and Software Engineering*, 2(4), 75–78.
- Shaikh, H., Khan, M. I., & Kelkar, Y. (2012). A Robust DWT Digital Image Watermarking Technique Basis On Scaling Factor. *International Journal of Computer Science. Engineering and Applications*, 2(4), 63.
- Singh, A., Jain, S., & Jain, A. (2013). A Survey: Digital Video Watermarking. *International Journal of Scientific & Engineering Research*, 4(7), 1261–1265.
- Singh, P., & Chadha, R. S. (2013). A survey of digital watermarking techniques, applications and attacks. *International Journal of Engineering and Innovative Technology*, 2(9), 165–175.
- Sinha, S., Bardhan, P., Pramanick, S., Jagatramka, A., Kole, D. K., & Chakraborty, A. (2011). Digital video watermarking using discrete wavelet transform and principal component analysis. *International Journal of Wisdom Based Computing*, 1(2), 7–12.
- Thakur, M. K., Saxena, V., & Gupta, J. P. (2010, July). A Performance analysis of objective video quality metrics for digital video watermarking. In *Computer Science and Information Technology (ICCSIT), 2010 3rd IEEE International Conference on* (Vol. 4, pp. 12-17). IEEE. doi:10.1109/ICCSIT.2010.5564962

Voloshynovskiy, S., Pereira, S., & Pun, T. (1999). Watermark attacks. *Erlangen Watermarking Workshop 99*.

Yeo, B. L., & Yeung, M. M. (1997, October). Analysis and synthesis for new digital video applications. In *Image Processing, 1997. Proceedings, International Conference on* (Vol. 1, pp. 1-4). IEEE.

KEY TERMS AND DEFINITIONS

Fidelity: The quality of the media (image, video and other) should not degrade on addition of the watermark to it.

Imperceptibility: The watermark should not be visible in the copyrighted video.

Robustness: The watermark should be robust against a wide range of attacks, so that the attacker cannot extract or manipulate the watermark.

Security: It should be impossible to create a duplicate watermark as this will lead to ambiguity during ownership verification.

Verification: The watermark can be extracted from the media and compared with the original watermark to prove ownership.

Watermark: It is a logo or digital signature which is used in watermarking.

Watermarking: It is the process of embedding a logo or digital signature in a media (image, video and other) for Copyright protection.



Ethical Computing Continues From Problem to Solution

Wanbil William Lee

The Computer Ethics Society, Hong Kong & Wanbil & Associates, Hong Kong

INTRODUCTION

Ethical Computing (Lee, 2015a) is instrumental in identifying and reaching a near-ideal solution to the problems arising from an environment that is *technology-driven information-intensive*. These problems raise techno-ethical issues, particularly information security concerns. *Post-implementation* and *post-contract* problems are cases in point. Many of these problems could have been avoided, occur; because we are either insensitive to or ignorant of their ethical implications. As a result, we could reach only a partial, compromised solution at best.

Ideally, the solution is not only technically efficient, financially viable and legal admissible, but also ethically acceptable, socially desirable, and in many situations environmental-friendly (the so-called *hexa-dimension criteria*) (Lee, 2015b & 2015d). In addition, the solutions sought must be capable of balancing the potential inter-conflicts among these demands or satisfying the five or six criteria. Given the conditions, a deep understanding of the basic ethical principles and the requisite technical know-how (the *requisite competence*) are necessary, and shifting our view on risk culminating in a new type of risk called techno-ethical risk or simply *ethical risk* and adopting a new tool of analysis to cater for the new risk (the *additive*) are also required (Lee, 2015c).

BACKGROUND

The Problem

Symptoms

Contemporary business organizations rely increasingly on information technology to accumulate and process the data needed to advertise their services and products, aiming at, for example, capturing market share and attracting customers. As a result, the marketplace becomes more transparent, and the consumers are better-informed thus more demanding. Consequently, more information is required, technology is increasingly relied upon to handle the increased amount of information generated and demanded, and new technological facilities such as call center and weblining (for marketing), and Big Data, Internet of Things, Cloud Computing (for communication and data management) are developed to aid business, to provide the processing power and contain the information explosion. This forms a vicious circle: *increase in demand for information leads to increase in reliance on technology, and increase in use of technology consumes and generates more information, culminating in the so-called technology-driven information-intensive phenomenon* (Lee, 2015d), raises ethical issues, and creates techno-ethical problems. Of these problems, many are commonly found in post-implementation and post-contract situations.

The one described below is typical of techno-ethical problems in post-implementation situations:

An online monitoring system was implemented successfully to replace an existing offline help-desk platform a high-tech facilities distributor. The new system enables help-desk staff to see exactly what is on the users' screens and to respond to users' requests for assistance quickly. Impressed by the fast response time and the increased user satisfaction, the executive vice president (EVP) asks the chief information officer (CIO) to have a copy of the system installed in her office for she wants to use this surveillance function to track down drug dealing allegedly occurring on company premises. The EVP's request raises indeed ethical dilemma for the Chief Information Officer (CIO).

Described in the following is another example of techno-ethical problems in post-contract situations:

The system specification (spec) was approved by the provider and client, and a contract duly signed. A senior project consultant assigned to the project discovered a fault in the inventory control function in the spec. The client is a fashion boutique so the fault is critical. The fault was confirmed after the consultant's site visit. Keep quiet or tell the boss or the client or both is a dilemma the consultant faces.

Causes

What causes this kind of problem is not the technology because the technology per se is neutral. What creates techno-ethical risks is the use of the technology. Two major reasons are cited here to make the point. First, we are insensitive to or ignorant of the ethical aspect of the problems and its ramifications because mainly of our lack of deep appreciation of the ethical principles. Second, we have hitherto treated risk as a technical, corporate/

personal matter when it is in fact a managerial, social concern under the influence of the so-called *flawed education across science and technology* and *misinterpretation of risk*. Three cases are described in the following to exemplify situations where techno-ethical risks arise:

First: Planning to replace a corporate legacy system by a web-based facility but concentrating on potential economic efficiency such as improved speed, elimination of redundancy or even reduced head-counts only, will miss such adverse consequences as deterioration in morale and end-user dissatisfaction (due to the disturbance to inertia).

Second: Evaluating information governance of a computer-based system but failing to include an audit of or a check for ethical issues will run the risk of a deficient information security management review.

Third: Assessing softlifting (the software equivalent of shoplifting) by focusing on the economic and legal impact such as infringement of copyright law and leaving out the social impact such as personal use of sensitive proprietary information will result in a risk of an incomplete assessment.

An Ideal Solution

As alluded earlier, an ideal solution is required to satisfy the hexa-dimension criteria, and is structured around, requisite competence and additive to supplement the requisite competence which alone is not sufficient.

REQUISITE COMPETENCE

Technical Know-How

Professional practitioners and layman users of IT presumably possess the basic rudiments. So this primary competence can be safely assumed.



The Basic Ethical Principles

IT professionals of various ranks, including CIOs, tend to offer support when asked for an opinion on the importance of computer ethics, but when pressed for elaboration as to what computer ethics is or why it is important, many may respond in silence. In order to proceed, we need a real appreciation of the basic ethical principles. To know what these theories are and how they can be used to explain the dilemmas, let us illustrate using the Edward Snowden episode. Was he a whistle-blower or simply a rumor-monger? Some respect him, calling him a hero; others disapprove of his actions, calling him a traitor. Is he defensible on ethical grounds? One might have heard of these arguments: “Snowden is not the only one. There are plenty other whistle-blowers,” or “If Tom, Dick and Harry can do it, why not Edward Snowden?” These arguments are based on the concept of relativism.

Relativism:

Right or wrong is not absolute, that is, what is right for you may not be right for me or I can decide to do what I think is right for me, but you can decide what is right for you.

Hence, if one person thinks it is right to say Snowden is a hero, but another individual does not think so, the argument is pointless, as it allows two people to decide right and wrong for themselves. In the end, no moral distinction between the opinions of the two individuals can be made. Certainly, the debate does not tell us whether Snowden’s actions are morally right or wrong.

Snowden is no ordinary worker; he is a professional, one who engages in a job that handles a highly sophisticated commodity—confidential information. He was an employee of the US National Security Agency (NSA). In this capacity, Snowden appears to be wrong, disloyal to his employer in stealing and disclosing confidential information without authority. However, while in the role of a professional, Snowden is expected

to respect professionalism and to observe his professional code of conduct, he has, as a person, a duty to himself and to his moral convictions. This duty-based argument is based on the theory of deontology.

Deontology:

An act is morally right if it is done out of a sense of duty.

So, as an employee, Snowden failed as he was disloyal and leaked confidential information. But, as a professional, he was right in exposing the stealth act because he was acting in accordance with professional conduct. While helpful in defending duty-bound actions, this principle is inherently troublesome because the actor owes responsibility to a multitude of stakeholders, and each of the stakeholders has its own aims that may be conflicting with one another.

Next, let us think of the impact or the consequences of Snowden’s actions. The consequences may be beneficial or harmful. Snowden might have done something “good” for the victims in particular and the world at large, and “bad” for NSA and the US government. This result-based argument, which is influenced by the concept known as consequentialism.

Consequentialism:

An action is morally right if its consequences are beneficial, and an action is morally wrong if its consequences are harmful.

The consequentialist argument is not sufficient and raises such questions as for whom or for how many is the result good, how good or how bad. This argument needs to be supplemented with a utilitarian view.

Utilitarianism:

An action is morally wrong if its harmful results are greater than its beneficial results, and all persons should pursue the greatest good for the greatest number of people.



A utilitarian argument may be useful to suggest the issue: for whom or what purpose the good result is beneficial or the bad result is harmful, but it raises further questions that include, among others, quantifying and comparing the results.

Besides these, others which are noteworthy are the Categorical Imperative and Social Contract Theory, and the so-called Golden/Silver Rule.

Kantian Categorical Imperative:

Never treat human beings merely as means to an end; always treat them as ends in themselves.

Social Contract Theory:

An action is morally wrong if someone's rights are being violated.

Golden Rule:

"Do unto others as you would have them do unto you" or "we should do to others what we would want others to do to us"; this is derived from "Do not do unto others what you would not have them do unto you" (sometimes referred as Silver Rule).

ADDITIVE

As can be seen, none of these principles alone can help to arrive at the ideal solution. To meet this end, a mix of some or all of these principles is necessary. Obvious, having the requisite knowledge is not enough, and something in addition is needed. That something, it is argued, comprises a new concept of risk, Ethical Risk, and a new tool for analysis of the dilemmas, Ethical Matrix (Lee, 2015a). This new view of Computer Ethics forms the core of the Ethical Movement in Cyberspace (Ethical Movement) (Lee, 2015b) advocated by the Computer Ethics Society (iEthics) (The Computer Ethics Society, 2012).

A New Concept of Risk

Connecting Security/Risk and Ethics

The human users of technology who create the risks, take them as *techno-economic risks* (due to *misinterpretation of risk* and the *flawed education*) (Lee, 2015a). It has been argued that abusing or not abusing the technology depends on the users' sense of morality which is influenced by many factors, amongst which is their understanding of the ethical principles. Using the technology in contradiction to ethical principles constitutes a different type of risk called *ethical risk vis-à-vis* risk of a technical, legal or financial nature, because it is a technical and managerial concern and should be measured in financial, legal and moral terms with equal priority.

For example, taking home from work a USB containing corporate data for personal convenience runs the risk of breaching the security for privacy or confidentiality. This is a situation where threat (t) exploits vulnerability (v), thus causing loss, damage or destruction to an asset (a), or $R = f(t, v, a)$. In this case, the person who takes home an un-encrypted USB in which confidential data are stored is a Threat, taking home an USB is a Vulnerability, and this action is a violation of data privacy, thus a risk.

This new risk is related partly to technology (the USB) and partly to people (both the perpetrator and the victims); it is therefore a risk of technological cum social nature in cyberspace, a new risk that can be called a *techno-ethical risk*, simply *ethical risk*. Figure 1 below depicts this phenomenon.

In cyberspace, as in the physical world, security relies on trust; trust depends on privacy; privacy is breached if law is not observed and if ethical principles are not followed. So, you would trust or disclose anything confidential to only those who respect privacy; you would trust your security in

Figure 1. Techno-ethical risk (connecting philosophy/technology)

Source: Lee, 2015d.



the protecting agent who would act according to ethical principles. Hence, trust will be lost or damaged, and undesirable consequences will follow if there is no privacy. Figure 2 is a sketch of the interrelationship of these entities.

Figure 3 combines the figures 1 and 2, and also shows the influence of law on privacy, and the link of techno-ethic risk to cybersecurity.

Shifting Our View of Risk

Security problems, be they of a technical or non-technical nature, is rooted in human mistakes to which none of us is immune. Wherever and whenever is there vulnerability, there is threat ready to exploit. Risk will result whenever a threat is actually carried out. To mitigate *risk* (that is,

Figure 2. Connecting security/risk and ethics

Source: Lee, 2015d.

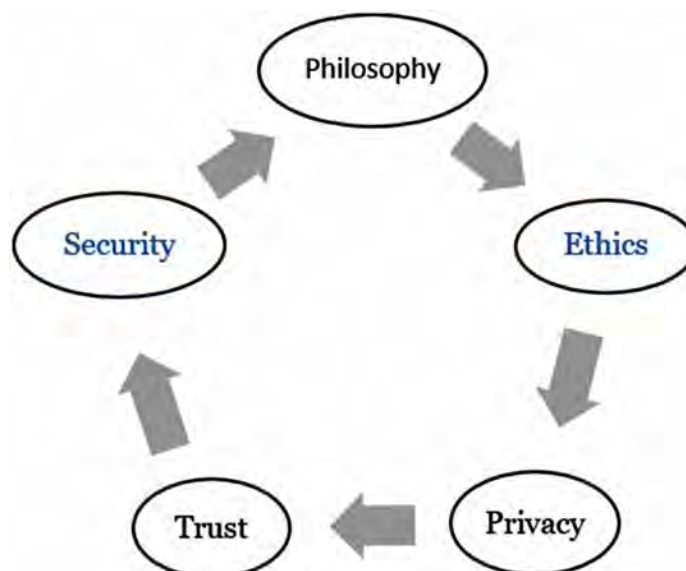
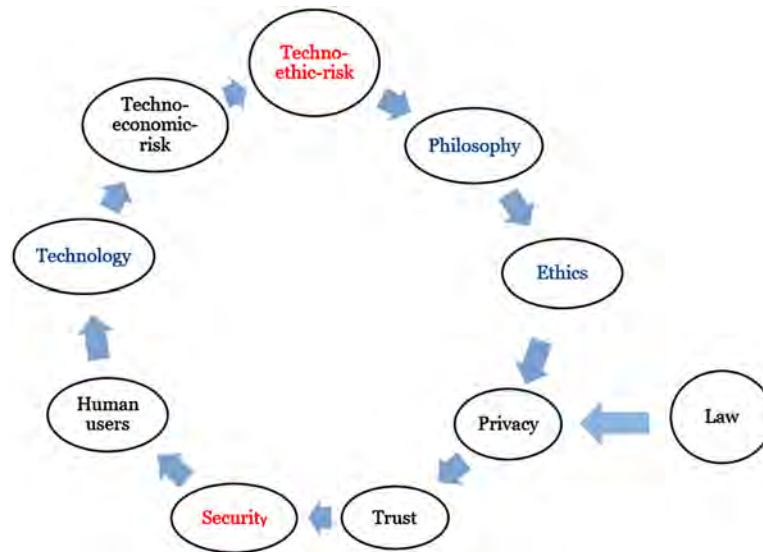


Figure 3. Connecting philosophy/technology and ethics/(security/risk)
Source: Lee, 2015d.



the damage, loss or destruction of we what we want to protect), we must identify *threat* and deal with *vulnerability*. Obviously, risk is a function of vulnerability and threat, or $r = f(v, t)$, and exposure of risk is a function of probability (the likelihood that risk occurs) and damage (of technical, financial and ethical nature) or $R = f(p, d)$.

Computer Ethics, as advocated by the Computer Ethics Society, can be viewed as a double duality representing a dual mission: computer ethics as a practice and a discipline, and as a dual function: computer ethics playing the role of a different type of risk and an alternative category of (Lee, 2015a). The impact of knowledge of Computer Ethics on students' behaviour, as hypothesized by in a classroom study, empirically supports the anti-risk mechanism role of Computer Ethics (Lee and Chan, 2008). The double duality (Lee, 2014-15) is reproduced below, shown as Figure 4.

The support of the Lee-Chan hypothesis, which was deduced from the empirical data of investigating the effect of Computer Ethics in a Computer Science syllabus (Lee & Chan 2008), showing i) 62% of the students claimed that they

never heard nor aware of Computer Ethics and were not sure if they used the computer ethically before taking the course (Question ii response) and 72% understood Computer Ethics and would use the computer according to ethical principles after finishing the course (Question iv response), ii) a small percentage (4.5% - Question viii responses) of the students claimed that they were indifferent to or did not care about Computer Ethics (Question viii response), and iii) 2.5% (the union of Question vi and Question vii response) claimed that they would not change their view (Lee, 2015a). It should be of interest to note that the number of participants in the annual survey averages about 100 over the six years (2006 to 2012). It is also worthy of further investigation to ascertain of the 2.5% of students who claimed not to change their view, 1% is clearly anti-ethics, what view the 1.5% of the students (Question vi response) would adopt. On account of a typo discovered in the response rate (Question viii), which should be 4.5% and not 5.5%, the corrected Table 1 (Lee, 2015a) is reproduced.

Figure 4. Conceptual graph: double duality

Source: Lee, 2014-2015.

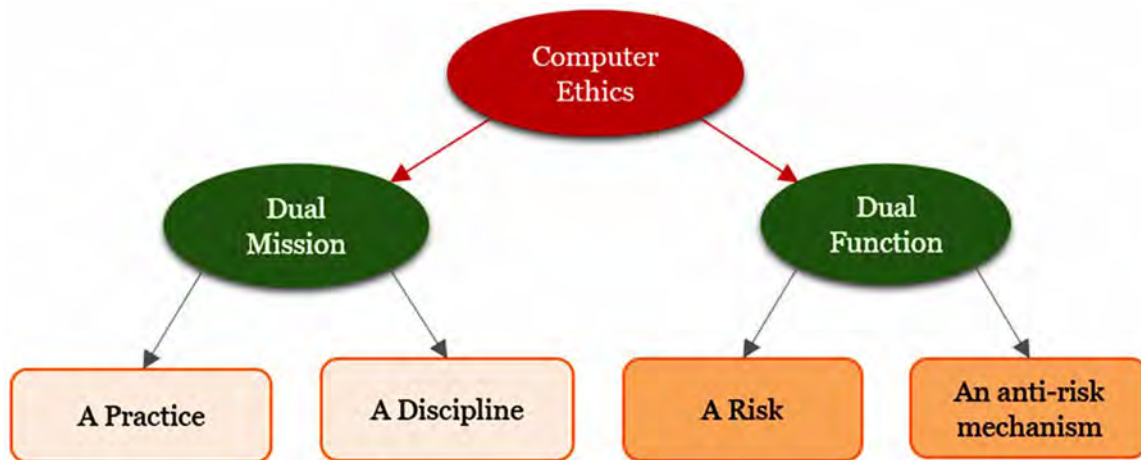


Table 1. Average response rates

Questions before Attending the Course	Rate
I heard or was aware of computer ethics, but I did not know if I used my computer ethically.	8%
I never heard of, nor was I aware of, computer ethics and I was not sure if I used my computer ethically.	62%
I never heard of, nor was I aware of, computer ethics and I thought I used my computer ethically.	30%
Questions after Attending the Course	Rate
I understand computer ethics and I will use my computer according to ethical principles.	72%
I understand computer ethics and I will not use my computer according to ethical principles.	21%
I will not change my view because I do not fully understand the subject of computer ethics.	1.5%
I will not change my view because I do not subscribe to the subject of computer ethics.	1%
I am indifferent because I do not care much about computer ethics.	4.5%

Source: Lee & Chan (2008).

A New Tool for Ethical Analysis

Ethical Matrix

Ethical Matrix, newly adopted by Ethical Computing as an additional approach to its tools and methods for performing ethical analyses (Lee, 2015a), poises to supplement the requisite competence. Checking for potential ethical and social impact (in addition to technical and economic efficiency) adds a step or steps to the other established anti-risk countermeasures. For example, going through the process of applying the Ethical Matrix method will force decision makers to

consider adverse consequences. This may reveal such risks as user morale and satisfaction or potentially undesirable consequences of a social or moral nature that would otherwise be missed in the typical anti-risk checks and audits, in addition to technical and economic efficiency issues such as improved speed, elimination of redundancy or reduced head counts. This makes computer ethics a different kind of anti-risk mechanism *vis-à-vis* the extant risk countermeasures including the following.

- Technical access control comprises physical and logical access controls such as

passwords, firewalls, intrusion prevention systems (IPS), secure web, email gateways, cryptographic algorithms and Anti-Virus Systems or Software (AVS). However, these mechanisms can handle only the physical, tangible errors or risks, but not risks of a sociotechnical nature.

- Computer law is useful as a reference framework for remedy and can be a forceful deterrent to prevent abusive acts by virtue of its power to punish. However, it is not as effective as expected: It is too slow to combat rapidly developed cybercrimes, and the process of creating new laws when needed is too complex and time consuming. Furthermore, people are, in general, reluctant to proceed with legal action or they opt for legal action only as a last resort.
- Risk analysis relies on the probability of occurrence and the associated damage or loss that the analyst estimates based on the threats identified and the related vulnerabilities in a risk assessment exercise. It is helpful in detecting risks but not very useful in preventing or powering them.
- Computer audit is useful for detecting deviation from set policies of performance and checking and verifying that compliance is properly enforced and results are consistent with the set standards, but in physical, cost-benefit terms only. The audit-based mechanism is, therefore, limited as a deterrent.

These extant countermeasures are being rendered impotent by such emerging applications and technologies as the Internet of Things (IoT), big data and cloud computing, all of which are complex and sophisticated, and by the ever-lurking perpetrators who are always ready to crack any new countermeasures soon after they are developed and released. Anti-risk is becoming more difficult. New anti-risk mechanisms are, therefore, called

for in order to strengthen the weakened existing mechanisms.



Illustration of Ethical Matrix

The Story

The story about an online monitoring system implemented by Libnaw, a high-tech facilities distributor, and the dilemmas that CIO faced thereof, is that described earlier in this article to exemplify a post-implementation problem. The dilemmas created and needed to be addressed include 'deviation from the firm's policy for acquisition approval (company policy), privacy invasion, staff being treated as a means, and staff morale (staff), corporate image damage (HRM), professionalism (CIO), duty (EVP). (Lee, 2014)

The Dilemma

The EVP's request worries the CIO.

First, managerial-technical functional conflict: The online surveillance capability in the hands of the EVP will become a *managerial control function*. This deviates from the originally approved purpose: to provide the help-desk staff with an online diagnosis and problem solving aid (*a technical support function*).

Second, management-staff interest conflict: Watching (by the EVP or someone from her office) over the users' screens without the users knowing it has the potential of breaching personal privacy in the work place. This certainly alarms both the staff (for fear of privacy invasion) and the firm (for concern over corporate image and personnel welfare).

Third, professional interest-contractual obligation conflict: Entertaining the EVP's request implies using the surveillance function for spying on the staff. This affects not only the CIO (with respect to professionalism and deontological obligations) but also the EVP (on her duty to protect the corporate image from possible damage), the

staff (about being exploited by one or more rotten apples) and the firm (about staff morale and company policy).

Analysis

The stakeholders identified are the firm, staff, EVP, and CIO, and the categories of ethical values considered relevant to the identified stakeholders' concern are well-being, autonomy and justice/fairness. The results of a first-cut analysis is shown in Table 2.

The next steps including modifying, deleting or adding, stakeholders and/or category of ethical values and the final steps including quantifying and comparing the concerns, and concluding are left for the interested readers to contemplate.

REMEDIAL ACTION: FROM CONCEPT TO PRACTICE

Scenario 1: Chuck's Dilemma: Contract Spec Fulfilled but Problem Unsolved

The Story

Alluded to earlier in the description of a post-contract problem, the story goes like this: Chuck is a senior project consultant working for A1 ASP, an application software provider. A1 signed a contract with Alwaays21, a fashion boutique in town, for the provision of a total system, and

Chuck was assigned to take full charge of the project. Reading through the system specification, Chuck was hit by a couple of loopholes which could cause intermittent logic errors in the update routine of the inventory control function. These errors, though rarely occur, could, if triggered, be vital to the time-critical fashion business. Chuck has to decide: Keep quiet or tell (the boss and/or the client).

Keep Quiet

Chuck is very tempted keep quiet about his finding since it is probably the most popular option. Also, relativism allows him to go the way he thinks is right. Keeping quiet can save him the trouble of confronting the boss and/or the client and avoid suffering a highly likely unpleasant reaction. But taking this option may be against his sense of duty, and worries him when thinking of the potential, unpleasant consequences.

Tell the Boss or the Client or Both

As an employee, Chuck knows that he is obliged to protect company property, to uphold company welfare, and to be loyal to the boss. As a professional, he should look after the client's welfare as well as that of the employer and other stakeholders. The Golden Rule tells Chuck to treat Alwaays21 as he would treat A1; Categorical Imperative advises Chuck to be fair to both parties. Chuck begins to incline to telling both the boss and the client. However, he hesitates as the issue of extra

Table 2. The first-cut results

Respect for Stakeholders	Well Being	Autonomy	Justice/Fairness
Firm	Personnel welfare, corporate image	Personnel protection	Staff morale
Staff	Personal privacy	Freedom of personal movement	Exploitation by minority
EVP	Job security	Firm's welfare	Entitlement of support resources
CIO	System utility	Professionalism	Distribution of computer resources

Source: Lee (2014).

cost pops up. The extra cost might incur as a result of increase in labour cost and resource for implementing the amended version. Who is to bear that extra cost? Has he got a case? It occurs to Chuck that a consequentialist argument would help.

Tell the Boss

When told, as Chuck expects, the boss is upset about the oversight (the flaw in the inventory control function design), and furious when the issue of extra cost is raised. The boss appears unwilling to bear the full cost.

Chuck makes these points on utilitarian grounds that the value associated with the potential damage to A1's reputation plus the cost of potential law suit and the ruin to the relationship with Alwaays21 or loss of its business will far exceed the extra cost, and this may also lead to serious morale issues, loss of customer trust including that of Alwaays21, damage to brand name, among others. In deontological terms, implementing a project knowingly with inherent weakness is an act of dishonesty, a violation of the company's code of conduct and a deviation from fair trade practice.

The boss's attitude seemed softened.

Tell the Client

Dolly appeared shocked upon learning the problem. She demanded to have the vulnerabilities dealt with thoroughly before implementation, or to withdraw the contract.

In response, Chuck argued that Alwaays21 should bear some or even equal responsibility: the contract was signed by both parties as a result of a mutually agreed feasibility study and a fully validated spec (on deontological grounds).

A1 doesn't have to reveal the problem voluntarily in the first place and just went on to implement according to the contract (in relativistic terms).

Further, the extra cost is negligible compared to the potential loss of business and the interrup-

tion to operation, etc. if the hidden errors pop up and the cost for amending and repairing in future if the system is implemented now according to the original design (a similar argument with the boss in consequentialist terms).

Now A1 has run the altruistic extra mile. This honest and ethical practice should be encouraged and praised, not rebuked and punished.

Closure

In the end, both A1 and Alwaays21 agree to amend the design and the contract, and to contribute equally to the extra cost.

Scenario 2: Taking Home a USB from Work

Taking home something like a biro from work is so common that it happens all the time, and probably nobody really cares about it. But taking home a USB is a different kettle of fish, and should not be taken lightly. People who do that unconsciously or for a purpose should realize that it is company property; it may contain proprietary data which may or may not have been encrypted; leaving where it belongs may be vulnerable to leaking (the content), physical damage to or loss of the device; and the potential liability to breach of privacy policy, thus incurring cost for law suit, and resulting in bad publicity, diminished trust, and so on.

The Story

One day in the early afternoon, Alex took with him a USB when he left office his (in Causeway Bay) in a hurry because he was trying to get to an evening seminar in Central in time and intended to continue working on his assigned project at home after the seminar. Alice accidentally spotted the act when passing Alex's office on her way to the staff pantry.



Complication

- Alex and Alice had been dating for almost two years and are getting steady.
- Betty (the boss) recommended to promote Alex only last week.
- The USB contains classified information, forbidden by company rules to leave company premises.
- The USB is not encrypted.

Dilemmas

- For Alice: report the case or not.
- For Alex: pretend to be ignorant of the rules or confess and defend on ground that though deviating from the rule, the intent of taking the USB away from office was to work at home, a contribution to increase in productivity.
- For Betty, the boss, when Alice reported Alex's act: to reprimand, to dismiss the report, or what?

Concerns

- Loyalty (Alex and Alice have for each other due to their intimate relationship, and for the company)
- Professionalism (Both Alex and Alice regard themselves as IT professionals)
- Alex's pending promotion (a concern shared by Betty, Alex, Alice and the team)

- Upholding company welfare, protecting company property
- Observing data privacy policy

Analysis

Obviously, there are four stakeholders: Alice, Alex, Betty and the team. The ethical values considered relevant to the identified stakeholders' concern are deontology, utilitarianism and consequentialism. The results of a first-cut analysis is shown in Table 3.

FUTURE RESEARCH DIRECTIONS

Ethical Matrix is a viable tool for ethical analysis of cyber-security issues. Application of this tool should be further developed and the first to develop is a quantification scheme for prioritizing remedial actions on the concerns, for only after a metric is developed and well validated can the concerns be compared and remedial actions taken. Interest in Computer Ethics is growing, though not as fast as expected; it must grow as cybercrimes grow; it needs time to accelerate as we take time to shift our view of risk (Lee, 2015c). Ethical Computing is to grow in importance in combating the constantly increase of ethical issues due to the alleged increase in hacking instances, insider perpetrators, security and privacy concerns associated with the emerging technology such as Big Data, Cloud Computing, and Internet of Things.

Table 3. The first-cut results: the USB case

Respect for Stakeholders	Deontology	Utilitarianism	Consequentialism
Alice	loyalty to Alex, company, professionalism	Alex's promotion	violation of data privacy policy
Alex	productivity	promotion	data protection
Betty	firm's welfare & property	her recommendation of Alex's promotion	data leaking
The team	team spirit & morale	promotion opportunity, fairness of treatment	data privacy

Source: Lee (2015d).

CONCLUSION

Many of the problems in the technology-driven information-intensive environment that could have been avoided or mitigated, occur because the ethical aspect of these problems is ignored or not recognized. It has been argued that solutions for these cyber-problems need to meet the six criteria: technical efficiency, financial viability, legal admissibility, social desirability, environmental sustainability, and ethical acceptability. It has been further argued that a deep understanding of the basic ethical principles, a shift in our view of ethics to include Ethical Risk, and a new tool for analysis of Ethical Risk (Ethical Matrix) are required. A completely ideal solution may not be readily and easily achieved. Some remedial action has been illustrated by means a few cases.

Computer Ethics that connects Technology (security and risk) and Philosophy (ethics and privacy). Ethical Computing as a practice that aims at establishing the ethical standards for guiding specialists and laymen alike through the ethical dilemmas are getting more important as a weapon against cyber-abuses that cause damage in excess of billions of US dollars. Further development is necessary and worthwhile.

REFERENCES

- Lee, W. W. (2014). Why Computer Ethics Matters to Computer Auditing. *ISACA Journal*, 2, 48–52.
- Lee, W.W. (2014-15). *Ethical, Legal & Social Issues*. Postgraduate Diploma in eHealth Informatics, The University of Hong Kong.
- Lee, W. W. (2015a). Ethical Computing. In M. Khosrow-Pour (Ed.), *Encyclopedia of Information Science and Technology* (3rd ed.; pp. 2991–2999). doi:10.4018/978-1-4666-5888-2.ch292
- Lee, W. W. (2015b). Ethical Movement: An alternative anti-crime mechanism in cyberspace. *Trusted Computing – A Security Oxymoron or The Future? 16th Info-Security Conference*.
- Lee, W.W. (2015c). *e-Crime & understanding Risk & Ethics in cyberspace*. Inaugural e-Crime Congress, Hong Kong.
- Lee, W.W. (2015d, September 11). *Ignorant of ethical pitfalls can hurt information security – Implications for IT users & Information Security Professionals*. The Hong Kong Jockey Club.
- Lee, W. W., & Chan, K. C. C. (2008). Computer Ethics: A Potent Weapon for Information Security Management. *Information Systems Audit & Control Journal*. Retrieved from www.isaca.org/jonline
- ADDITIONAL READING**
- Johnson, D. G. (2009). *Computer Ethics* (4th Ed.). Upper Saddle River, NJ: Prentice Hall Josephson Institute of Ethics. (1999). Five Steps of Principled Reasoning. Retrieved from http://www.ethicsscoreboard.com/rb_5step.html
- Josephson Institute of Ethics. (2002). Making Ethical Decisions: A 7-Step Path. Retrieved from <http://blink.ucsd.edu/finance/accountability/ethics/path.html>
- Kallman, E. A., & Grillo, J. P. (1996). *Ethical Decision Making and Information Technology*. Watersonville, CA: McGraw-Hill.
- Lee, W. W. (2010). *Information Security Management: Semi-intelligent Risk-analytic Audit*. VDM Verlag.
- Mephram, B., Kaiser, M., Thorstensen, E., Tomkins, S., & Millar, K. (2006). *Ethical Matrix Manual*. The Hague: LEI [http://www.ethicaltools.info/content/ET2ManualEM\(Binnenwerk45p\).pdf](http://www.ethicaltools.info/content/ET2ManualEM(Binnenwerk45p).pdf)

Moor, J. H. (1985). What is Computer Ethics? *Metaphilosophy*, 16(4), 266–275. doi:10.1111/j.1467-9973.1985.tb00173.x

KEY TERMS AND DEFINITIONS

Additive: This is a supplement to the requisite competence which alone is not sufficient, and a requirement of a change in viewing risk as having not only technical, physical, financial and legal but also ethical and social implications, and a tool or method to cater for this type of risk.

Computer Ethics: Applied Ethics referring ethical issues arising out of using the computer and its peripherals.

Ethical Computing: The practice of Computer Ethics and can be regarded as a branch of Computing, somewhat akin to Green Computing, Mobile Computing, and Cloud Computing. Whereas the others deal with tangible aspects of Computing, Ethical Computing handles techno-ethical aspects.

Ethical Risk (Short for Techno-Ethical Risk): Abusing technology or technology-based systems constitutes a risk called techno-ethical risk or ethical risk. Using the computer or developing computer-based systems or software inconsistent with ethical principles results in computer-ethical risk or simply ethical risk.

Hexa-Dimension Criteria: This refers to an ideal state expected of a solution that is technically efficient, financially viable, legally admissible, socially desirable, and ethically acceptable. Another dimension “environmentally sustainable” was added making a hexa-dimension model was developed and presented for consideration as a metric for the design of a code of conduct corporate-wide or functional unit based (Lee, 2015d).

Ideal Solution: To solve the problem which is required to satisfy the penta-dimensional criteria, the proposed solution is structured around the two major components, *requisite competence* and

additive (to supplement the requisite competence which alone is not sufficient).

Requisite Competence: This comprises the technical-know (the rudimentary skills and knowledge to use the computer) and a deep understanding of the basic ethical principles.

Technology-Driven Information-Intensive: This is a phenomenon that characterized the contemporary society and attributes to the vicious circle: increase in demand for information leads to increase in reliance on technology, and increase in use of technology consumes and generates more information. On the one hand, to capture market share and attract customers, contemporary business organizations rely increasingly on information technology to accumulate and process data needed to advertise services and products. Hence, the marketplace becomes more transparent, and the consumers are better-informed thus more demanding. More information is required and technology is more heavily relied upon. On the other hand, to aid business, to increase the processing power, and to handle the increase in data volume, new technological facilities such as call center and weblining (for marketing), and Big Data, Internet of Things, Cloud Computing (for communication and data management) are developed and used by business but these facilities consume and generate more data. In turn, reliance on technology increases to handle the information explosion.

The Basic Ethical Principles: For the purpose of this exposition, included are relativism, deontology, consequentialism and utilitarianism, categorical imperative, Social Contract Theory, and Golden Rule.

The Computer Ethics Society: Was established in 2012, aiming to promote professional and ethical standards, and to foster an open platform for sharing and interchanging experience and knowledge of moral issues associated with the use and development of ICT-based systems (www.iEthicsSoc.org).

The Problem: Referred to in this article: Its symptom is the phenomenon called technology-driven information-intensive. Its cause is techno-ethical risk or *ethical risk*. This type of risk created because the user not only is insensitive to or ignorant of the ethical aspect of the problems and its ramifications, attributing to a lack of deep ap-

preciation of the ethical principles, but also treats risk as a technical, corporate/personal matter when it is in fact a managerial, social concern under the influence of the so-called *flawed education across science and technology* and *misinterpretation of risk*. Post-contract problems and post-implementation problems are typical examples.



Group Signature System Using Multivariate Asymmetric Cryptography

Sattar J. Aboud

University of Bedfordshire, UK

INTRODUCTION

The group signature lets the group of people to sign document anonymously on behalf of other group. In the case of the dispute, the designated director can open the signature to disclose the identity of its generator. To the degree that we know the majority of the group signatures are relied on the known schemes, such as RSA and ElGamal. However, these schemes could be broken when quantum computers appear. The problem typed multivariate asymmetric key cryptography is the notable option to common asymmetric schemes for its possible to withstand future attacks of quantum computers. The initial group signature scheme relied on the multivariate asymmetric cryptography that is introduced in this chapter. The proposed scheme have two extraordinary attributes. In the first one, the group signatures are divided to dissimilar time intervals. The signatures are linkable in the same time interval, but un-linkable among dissimilar time intervals. In the second one, the duties of the group director is restricted. The group director does not allow him to open the signature without the assist from the verifier. These attributes are vital in selected uses such as e-voting schemes. The concept of the proposed scheme is straightforward and its security bases on both an arbitrary hash function and an isomorphism of polynomial problem.

In 1991, Chaum-Heyst presented the first idea of group signature. The group signature scheme give permission to the group of people to sign the documents on behalf of the group. The verifier can only inform that the signature is signed by the person from the group, but cannot determine

the identity of the signer. In addition, the verifier cannot differentiate if the two signatures are published by the same person of the group. But, in special case such as official dispute, the designated group director can open the signature to disclose the identity of its generator. At the same time, no one even the group director can forge the signature of other group people.

The characteristics of group signature construct it smart for many specific applications, like e-voting, e-cash and e-games. For instance, in e-voting systems, the electorate are not allowable to vote many times. Thus the count authority should be capable to differentiate the reduplicate votes without opening an election. Furthermore, there is a rule exist supervision authority to constraint the duties of the count authority and promise the fairness of the voting in the voting system. Thus, the group signature schemes cannot be employed the e-voting systems straight. Most of the group signatures are using known cryptography schemes, such as RSA and ElGamal. However, the algorithm proposed by Shor illustrates that solving the factoring integers and the discrete logarithms can be achieved in polynomial time on the quantum computer. If the quantum computers become a reality, the common asymmetric key cryptography under these problem, such as RSA and Elliptic curve will be broken. multivariate asymmetric key cryptography is studied to be one of the best option. The security basis of multivariate asymmetric key cryptography is the information that solving the set of multivariate polynomial formulas over the finite field is the NP-hard problem. Quantum computers do not seem to have any benefit if managing this NP-hard problems, and

DOI: 10.4018/978-1-5225-2255-3.ch424

it appears that we cannot recover the solution to the set of polynomial formulas efficiently even in the future. Furthermore, multivariate asymmetric key cryptography schemes are more efficient than common asymmetric key cryptography. It makes them appropriate for restricted computing tools, for example smart cards. Different multivariate asymmetric key cryptography schemes have been presented.

QUANTUM COMPUTING THREATENS

Quantum computing threatens definite techniques and does not threaten others. Public key encryption, is being used considerably for securing the internet payments, banking transactions, and also emails and webs. The majority of today cryptography schemes are using public-key cryptography, that is in fact secure anti-attacks from contemporary computers.

Suppose that quantum cryptography can easily break many schemes by inverse the computing private-keys and quicker than the classical computer. While quantum cryptography are still in their early stages and non-equipped, with publicly known new quantum computers, small to attack traditional cryptography algorithms, many public authorities have begun to know the risk included if this technology becomes the practical applications. Since quantum computers is to process huge amounts of information in the quite short of time.

Traditional cryptography schemes provide computational security but is not ensure perfect or resistant security. The power of the existing cryptography algorithms based on composite mathematical problems, for example integer factoring, elliptic curve and discrete logarithm problem. Such difficulties can deciphered by applying large-scale quantum cryptography and thus can simply break traditional schemes. Consequently, experts have started planning new encryption schemes that are considered quantum-resistant that cannot be broken as fast as traditional algorithms.

The National Security Agency recognized the quantum processing threat by publicly announcing their strategies for changing to quantum unbreakable methods. However, the quantum processing threat has increased over public key infrastructure that is applied greatly in protecting the webs.

Quantum cryptography can be attack both symmetric schemes such as block ciphers, and asymmetric schemes such as RSA and DSA. Such cryptography can break each single public key algorithm in the small amounts of time. Quantum methods, for example Shor scheme, can be applied to retrieve the RSA key in polynomial time, but quantum cryptography with sufficient power at present is not existed.

Post-quantum encryption is being utilized for designing cryptography methods that are believed to be secure anti-attack by quantum processors. It is expected that 2048-bit RSA keys can be defeated on the quantum computer containing 4000 qubits and 100 million gates. Even if there are some public-key methods that are believed resistant, they are not utilized currently, since Quantum algorithms is based on difficult and complicated mathematical problems to give security that is stronger than conventional cryptography. If quantum cryptography becomes the truth, it will product in re-engineering and improvements in existing encryption schemes.

BACKGROUND

Due to its significant attributes, group signature draw many authors attentions. However, in addition to the first scheme proposed by Chaum-Heyst, there are many other schemes appear. For example, Chen-Pedersen in 1994 presented the new group signature scheme which conceal an identity of the signer categorically and permitted new persons to associate the group. Also, in 1997 Camenish-Stadler introduced an efficient group signature scheme for large groups, with the size of public keys and the size of signatures are free of the group persons. In



2000, Ateniese *et al.* presented the group signature scheme relied on other number theory topics with good efficiency and robust security. In 2003 Chen-Yang proposed another digital signature scheme called TTS/4. It is from the family of multivariate cryptography schemes. They claim that TTS/4 is competitive and better compare with other schemes. In 2004, Ding introduced the new variant scheme from the concept of perturbation. This scheme employs the set of small number linearly independent functions. The difference between the proposed thought and the similar concept of the concealed field equation and Oil-Vinegar scheme is that perturbation is internal, while other schemes are external perturbation. In 2010, Petzoldt, *et al.* they extend the thought of the Rainbow signature scheme. They claim that the construction of their scheme able to decrease the length of the public-key up to 62%. In 2011, Jintai Ding published a patent contains three schemes to construct new multivariate public-key schemes that are better in security and efficiency. The three schemes are entitled the internal perturbation plus, the enhanced internal perturbation and the multi-layer Oil-Vinegar construction. In 2012, Noaki Ogura, proposed scheme concerning the multivariate public-key scheme. In 2014, Ning Huang analyzed and solved the errors of the medium field multivariate public-key encryption scheme.

In the multivariate asymmetric-key cryptography schemes, the asymmetric-key is the set of polynomials $p = (p_1, \dots, p_m)$ in integers $x = (x_1, \dots, x_n)$, with all parameters and coefficients are in $K = GF(k)$. This is typically achieved by:

$$P = S \circ Q \circ T : K^n \rightarrow K^m$$

or

$$\begin{aligned} P : w = (w_1, \dots, w_n) \in K^n &\xrightarrow{T} x = M_T w + c_T \\ &\xrightarrow{Q} y = M_S y + c_S = (z_1, \dots, z_m) \in K^m \end{aligned}$$

The central map Q be owned by the certain class of quadratic maps whose inverse does calculated straightforwardly. The maps S and T are affine and bijective. The private-key contain the central map Q and affine maps S, T .

MAIN FOCUS OF THE CHAPTER

In this section, we are going to describe the contribution, then the traditional schemes, the solution scheme, then the discussions and finally the conclusions and remarks.

The Contribution

This Chapter presents a new group signature scheme using multivariate asymmetric cryptography. Compared with the existed signature schemes, The proposed scheme is applicable to e-voting schemes and can convince the requirements of e-voting schemes because it has two important characteristics, Traceability and Unlinkability. Traceability denotes that a group director cannot open the signature alone. He has to collaborate with a verifier to disclose an identity of the signer. Unlinkability denotes that the group signature can be split accordance to time durations. Then signatures are linkable in the same time range, but un-linkable between dissimilar time periods. Therefore, the count authority can notice the double-votes prior to open them. Thus, there are two features in the proposed signature, for count and supervision authority. Also, the size of signatures and the calculation overhead are private from the group members in the proposed scheme. So, it is efficient for large groups.

The Proposed Group Signature Scheme

The customary group signature schemes permit the user to sign documents anonymously on the behalf of his group. The verifier can check that the signature was signed by the user of the group,

nevertheless cannot determine which user of the group signed it. Also, the signatures are unlinkable, which means that provided two group signatures, the verifier cannot distinguish if they are published by the same group user or not. However, the unlinkability seems not modified to some uses for example the e-voting schemes. The verifier must be capable to differentiate the over-votes, in the meantime the anonymity must be certain. In addition, in customary group signatures the group director can open any group signature randomly.

Although in the e-voting schemes, there is typically the supervision organization to assure the fairness of the voting. Thus we want to constraint the duties of the group director. Thus, the customary group signatures cannot be useful with e-voting schemes straight. In order to solve these difficulties, we introduce the new group signature scheme fitted with e-voting schemes. The proposed scheme is described as follows.

1. **Unlinkability:** The group signature is divided into time duration. The signatures in the same time range are linkable. The verifier can check if two signatures are published by a same group member. But the signatures between dissimilar time period are unlinkable. They are entirely anonymous. This property is useful in the e-voting schemes. For instance, in the voting there are numerous positions require to be chosen. Every duration for one position. As the votes in the same period are linkable, the verifier is capable to differentiate the over-votes without opening them. All at once, the ballots between dissimilar ranges are un-linkable, thus ballots for diverse position are anonymous.
2. **Traceability:** The group director and the verifier cannot disclose an identity of the signature individually, but they can do it jointly.
3. **Correctness:** The group signature published by a valid group entity and should be accepted by a verifier.

4. **Unforgeability:** Only the legitimate group entities are capable to sign documents on behalf of the group.
5. **Undeniable:** Both the group entity and the group director is capable to sign documents on behalf of other group entities. This also causes the group signature undeniable.
6. **Anonymity:** Provided the legitimate group signature, no entity can disclose the identity of a real signer, unless the group director opens the signature with an assist of the verifier.

In the proposed scheme, the group director cannot open the signature randomly. His duties is restricted. This characteristic is significant in some application such as the e-voting schemes. The management usually contains the count authority and the supervision authority. Neither of them can disclose an identity of the vote alone. They have to act jointly to open the vote to disclose a real identity. This characteristic lets the voting more open.

Signature of Knowledge

In this section, we present the signature of knowledge using multivariate asymmetric key cryptography, which is the construction block of the proposed scheme.

Isomorphism of Polynomials Problem

The isomorphism of polynomials problem was first presented by Patarin in 1996. It is the basic problem of multivariate encryption because it is connected with the difficulty of the key recovery of the schemes. The idea of Isomorphism of Polynomials is illustrated as follows.

Suppose that $K = GF(k)$ is the finite field. All parameters are over the field K . Assume that A are the set of u quadratic formulas with n parameters $x_1, \dots, x_n$ that provide y values from x values:



$$y_k = \sum_i \sum_j Y_{ijk} x_i x_j + \sum_i u_{ik} x_i + \delta_k \quad \text{where } k = 1, \dots, u$$

Given B is the set of u quadratic formulas with n parameters $x'_1, \dots, x'_n$ that provide y' values from x' values:

$$y'_k = \sum_i \sum_j Y'_{ijk} x'_i x'_j + \sum_i u'_{ik} x'_i + \delta'_k \quad \text{where } k = 1, \dots, u$$

Suppose that T is the bijective affine transformation of the parameters x'_i where $1 \leq i \leq n$ with S is the bijective affine transformation of the parameters y_k , such that $1 \leq k \leq u$. Then

$$T(x'_1, \dots, x'_n) = (x_1, \dots, x_n) \\ S(y_1, \dots, y_u) = (y'_1, \dots, y'_u)$$

When there is the transformation pair (S, T) to convince $B = S \circ A \circ T$, we said to be A and B are isomorphic, and the bijective affine transformation pair (S, T) is the isomorphism from $(A$ to $B)$. When A and B are two public sets of u quadratic formulas, and A and B are isomorphic, get isomorphism (S, T) from $(A$ to $B)$.

Signature of Knowledge Protocol

Zero-knowledge proofs permit the prover to show the knowledge of the secret, without disclose any important information. Patarin introduced the non-interactive zero-knowledge proofs of knowledge scheme using an isomorphism of polynomial problem [13]. In this Chapter, we will use the simplicity but practical scheme we entitled it signature of knowledge which is as follows.

Assume K is the finite field. Given n, u, q are three variables, and H is the secure hash function $H : (0, 1)^* \rightarrow (0, 1)^q$ that maps the binary bits of random size to q -bit hash value. Assume A and B are two sets of u quadratic formulas with n parameters as follows:

$$A : y_k = \sum_i \sum_j Y_{ijk} x_i x_j + \sum_i u_{ik} x_i + \delta_k$$

where $k = 1, \dots, u$

$$B : y'_k = \sum_i \sum_j Y'_{ijk} x'_i x'_j + \sum_i u'_{ik} x'_i + \delta'_k$$

where $k = 1, \dots, u$

Suppose m is a document to be signed. Request $q + 1$ parameters $V = h, (S_1, T_1), \dots, (S_q, T_q)$ the signature of knowledge involved (A, B) and document m , when the following conditions are convinced: $C_i = S_i \circ A \circ T_i$, if $h(i) = 0$ or $C_i = S_i \circ B \circ T_i$, if $h(i) = 1$ with $1 \leq i \leq q$, and $h = H(m \parallel C_1 \parallel C_2 \parallel \dots \parallel C_q)$.

When a prover identifies the isomorphism (S, T) from $(A$ to $B)$ means $B = S \circ A \circ T$, then a signature of knowledge can be created as follows.

The prover chooses q arbitrary bijective affine transformation pairs $(S'_1, T'_1), \dots, (S'_q, T'_q)$ which appear like:

$$T'_1(x_1^{(1)}, \dots, x_n^{(1)}) = (x_1, \dots, x_n), S'_1(y_1, \dots, y_u) = (y_1^{(1)}, \dots, y_u^{(1)}) \\ T'_q(x_1^{(q)}, \dots, x_n^{(q)}) = (x_1, \dots, x_n), S'_q(y_1, \dots, y_u) = (y_1^{(q)}, \dots, y_u^{(q)})$$

The prover finds:

$$C_1 = S'_1 \circ A \circ T'_1 \\ C_2 = S'_2 \circ A \circ T'_2 \\ : \\ C_q = S'_q \circ A \circ T'_q$$

Then, gets q sets of u quadratic formulas with n parameters as follows:

$$y_k^{(1)} = \sum_i \sum_j Y_{ijk}^{(1)} x_i^{(1)} x_j^{(1)} + \sum_i u_{ik}^{(1)} x_i^{(1)} + \delta_k^{(1)}$$

where $k = 1, \dots, u$

$$y_k^{(q)} = \sum_i \sum_j Y_{ijk}^{(q)} x_i^{(q)} x_j^{(q)} + \sum_i u_{ik}^{(q)} x_i^{(q)} + \delta_k^{(q)}$$

where $k = 1, \dots, u$

The prover finds a hash outcome $h = H(m \parallel C_1 \parallel C_2 \parallel \dots \parallel C_q)$. Assume that the binary format of result h is introduced as $h(q) \dots h(2)h(1) \in (0,1)^q$. The prover finds $(S_i, T_i) = (S'_i, T'_i)$, if $h(i) = 0$, or $(S'_i \circ S^{-1}, T'^{-1} \circ T'_i)$ if $h(i) = 1$ where $1 \leq i \leq q$. Therefore, the signature of knowledge V is built.

If the verifier receives the public key (A, B) , the document m , and the signature of knowledge V , the steps of the signature verification is as follows.

The verifier finds q sets of u quadratic formula by

$$C_i = S_i \circ A \circ T_i, \text{ if } h(i) = 0 \text{ or } C_i = S_i \circ B \circ T_i, \text{ if } h(i) = 1 \text{ where } 1 \leq i \leq q$$

The verifier finds the hash value $h = H(m \parallel C_1 \parallel C_2 \parallel \dots \parallel C_q)$, and verifies if the new hash value h' is equal to h .

The Description of the Proposed Scheme

In this section, we introduce the new group signature scheme using multivariate asymmetric key cryptography, deal with the e-voting schemes.

Scheme Initialization

The group director setup the scheme parameters as follows:

1. Suppose that n, u, q are three integers. Also, let $K = GF(k)$ is the finite field.
2. H is the secure hash function $H : (0,1)^* \rightarrow (0,1)^q$ that maps the binary bits of random size to q -bit hash value.

3. Q is the central map chosen by a group director.
4. Every entity i has a secret key $(L_{1,i}, L_{2,i})$, and a public key $A_i = L_{1,i} \circ Q \circ L_{2,i}$.

Joining Group

Each user wants to join the group to get the ability of signing on behalf of the group. Assume that the user i needs to join the group.

1. User i chooses arbitrary bijective affine transformation pair (S_i, T_i) and maps A_i to B_i by $B_i = S_i \circ A_i \circ T_i$. The, A_i and B_i are isomorphic and affine transformation (S_i, T_i) is an isomorphism from $(A_i$ to $B_i)$. This pair (A_i, B_i) is the public key of user i in the group, while (S_i, T_i) is the secret key of user i in the group.
2. User i makes the signature of knowledge of (A_i, B_i) , represented by V_i , and passes the knowledge V_i and (A_i, B_i) to the group director. In this case, the interaction between a user i and a group director must be secure.
3. The group director checks the signature of knowledge. If it is true, he stores the knowledge V_i and (A_i, B_i) . if not, he erases the knowledge V_i and (A_i, B_i) , and reject the user i request.

Signing Documents

The group director can split the group signature to a number of time periods. The amendment of the period is ruled by the group director. So as to alter the period, the group director should do the following two phases.

First Phase: Change Period

The steps of this phase are as follows:



1. The group director chooses the public keys of various group users. These users form the sign group.
2. For each public key (A_i, B_i) in a sign group, the director makes the arbitrary bijective affine transformation pair $(S_{1,i}, T_{1,i})$, and finds

$$\begin{aligned} A_i^{(1)} &= S_{1,i} \circ A_i \circ T_{1,i} \\ B_i^{(1)} &= S_{1,i} \circ B_i \circ T_{1,i} \end{aligned}$$

The group director issued $(A_i^{(1)}, B_i^{(1)})$ to the bill-board. Then he posts each affine transformation pair $(S_{1,i}, T_{1,i})$ to the related user i in the sign group by a secure communication.

Second Phase: Signing

Inside every period, group users in the sign group can sign any documents on the behalf of the group, where the signatures are linkable. Assume that group user i needs to sign the document m , entity should perform the second phase as follows. The steps of this phase are as follows:

1. The user i passes $(A_i^{(1)}, B_i^{(1)})$ to verifier v .
2. The verifier v verifies that if $(A_i^{(1)}, B_i^{(1)})$ is in the bill-board. If yes, he chooses the arbitrary bijective affine transformation pair $(S_{2,i}, T_{2,i})$ and posts it to user i . Then he finds $A_i^{(2)} = (S_{2,i} \circ A_i^{(1)} \circ T_{2,i}, B_i^{(2)} = S_{2,i} \circ B_i^{(1)} \circ T_{2,i})$ and stores $(A_i^{(2)}, B_i^{(2)})$ and $(S_{2,i}, T_{2,i})$ in the database.
3. The user i finds $(A_i^{(2)}, B_i^{(2)})$. Then, he calculates the isomorphism (A'_i, B'_i) from $A_i^{(2)}$ to $B_i^{(2)}$ by:

$$\begin{aligned} S'_i &= (S_{2,i} \circ S_{1,i}) \circ S_i \circ (S_{2,i} \circ S_{1,i})^{-1} \\ T'_i &= (T_{1,i} \circ T_{2,i})^{-1} \circ T_i \circ (T_{1,i} \end{aligned}$$

With (S'_i, T'_i) user i can produce the signature of knowledge of document m and $(A_i^{(2)}, B_i^{(2)})$ indicated by $V_i^{(2)}$. Therefore $[V_i^{(2)}, (A_i^{(2)}, B_i^{(2)})]$ is the group signature published by user i .

Verifying Signatures

Upon receiving the group signature $[V_i^{(2)}, (A_i^{(2)}, B_i^{(2)})]$, a verifier v verifies the validity of the signature by two steps.

1. The verifier v decrypts the $(A_i^{(1)}, B_i^{(1)})$ by

$$\begin{aligned} A_i^{(1)} &= S_{2,i}^{-1} \circ A_i^{(2)} \circ T_{2,i}^{-1} \\ B_i^{(1)} &= S_{2,i}^{-1} \circ B_i^{(2)} \circ T_{2,i}^{-1} \end{aligned}$$

Then he verifies $(A_i^{(1)}, B_i^{(1)})$ if it is in the bill-board. If yes, he moves to the next step. If not the signature is invalid.

2. The verifier v verifies if $V_i^{(2)}$ is the valid signature of knowledge of document m and $(A_i^{(2)}, B_i^{(2)})$. If yes, he accepts the signature. If not the signature is invalid.

Linking Signatures

Suppose that the two group signatures $[V_i^{(2)}, (A_i^{(2)}, B_i^{(2)})]$ and $[V_j^{(2)}, (A_j^{(2)}, B_j^{(2)})]$. So, to check the double-signature is easy.

1. The verifier verifies if $(A_i^{(2)}, B_i^{(2)}) = (A_j^{(2)}, B_j^{(2)})$ then the two signatures are published by the same user. If not move to the next step.
2. The verifier decrypts the $(A_i^{(1)}, B_i^{(1)})$ and the $(A_j^{(1)}, B_j^{(1)})$ as follows:

$$(A_i^{(1)}, B_i^{(1)}) = (S_{2,i}^{-1} \circ A_i^{(2)} \circ T_{2,i}^{-1}, S_{2,i}^{-1} \circ B_i^{(2)} \circ T_{2,i}^{-1})$$

$$(A_j^{(1)}, B_j^{(1)}) = (S_{2,j}^{-1} \circ A_j^{(2)} \circ T_{2,j}^{-1}, S_{2,j}^{-1} \circ B_j^{(2)} \circ T_{2,j}^{-1})$$

If $(A_i^{(1)}, B_i^{(1)})$ and $(A_j^{(2)}, B_j^{(2)})$ are still dissimilar then the two signatures are published by unlike group users. If not, they are published by the same user.

- **Remark:** When the two signatures are from dissimilar periods, they are un-linkable. It means that the outcome of the Link operation is published by dissimilar group users.

Opening Signatures

To open the group signature to disclose the real identity of the signer, the group director has to assist with a verifier. For instance, $[V_i^{(2)}, (A_i^{(2)}, B_i^{(2)})]$ the group director receives the affine transformation pair $(S_{2,i}, T_{2,i})$ from the verifier. Then the group director finds (A_i, B_i) using $(S_{2,i}, T_{2,i})$.

$$A_i = (S_{2,i} \circ S_{1,i})^{-1} \circ A_i^{(2)} \circ (T_{1,i} \circ T_{2,i})^{-1}$$

$$B_i = (S_{2,i} \circ S_{1,i})^{-1} \circ B_i^{(2)} \circ (T_{1,i} \circ T_{2,i})^{-1}$$

The isomorphism from $(A_i^{(2)}, B_i^{(2)})$ to (A_i, B_i) is $((S_{2,i} \circ S_{1,i})^{-1}, (T_{1,i} \circ T_{2,i})^{-1})$. Therefore the group director obtains an identity of the real signer.

Security of the Proposed Scheme

The proposed scheme properties are as follows.

- **Unlinkability:** The signatures are linkable in the same time period, but un-linkable in dissimilar time periods.

Proof: In one period, group user j can obtain only one license (S_1, T_1) from the group director.

Therefore he can only generate one $(A_j^{(1)}, B_j^{(1)})$. He can inquire a verifier twice for (S_2, T_2) and (S'_2, T'_2) then generates $(A_j^{(2)}, B_j^{(2)})$ and $(A_j'^{(2)}, B_j'^{(2)})$ but a verifier will decrypt the same $(A_j^{(1)}, B_j^{(1)})$ using (S_2, T_2) and (S'_2, T'_2) respectively. Thus the signatures in the same time period are linkable. Whereas in dissimilar periods, the user j will obtain dissimilar licenses (S_1, T_1) and (S'_1, T'_1) from group director, and he generates $(A_j^{(1)}, B_j^{(1)})$. The verifier can decrypt them using (S_2, T_2) and (S'_2, T'_2) but he cannot decrypt the origin (A_j, B_j) . Thus, the signatures between dissimilar time periods are unlikable.

- **Traceability:** Both the group director and the verifier cannot disclose an identity of the signature alone, but they can disclose it jointly.

Proof: In the proposed scheme, after the computing of $(A_i^{(1)}, B_i^{(1)})$ the signer i maps the $(A_i^{(1)}, B_i^{(1)})$ to $(A_i^{(2)}, B_i^{(2)})$ by using an affine transformation pair $(S_{2,i}, T_{2,i})$ provided by a verifier. So, the group director cannot sets up the relationship between $(A_i^{(2)}, B_i^{(2)})$ and $(A_i^{(1)}, B_i^{(1)})$ without $(S_{2,i}, T_{2,i})$. Thus, the group director cannot disclose an identity of the signature. Also, a verifier cannot sets up the relationship between $(A_i^{(1)}, B_i^{(1)})$ and original (A_1, B_1) . Thus, the verifier cannot open the signature alone, unless otherwise there is a cooperation between the group director and the verifier to decrypt the identity of the signature.

- **Correctness:** The group signature published by a valid group user should be accepted by a verifier. where u is the polynomial numbers in Q , n is the parameter numbers.
- **Unforgeability:** The group user is capable to sign documents on behalf of the group.

Table 1. The best algorithms for isomorphism of polynomials

Problem	Sub-Case	Complexity
Isomorphism of Polynomials	degree = 2, $u = n$, inhomogeneous	$O(n^3) / O(n^6)$
	degree = 2, $u = n$, homogeneous	$O(n^{3.5} \cdot q^{n/2})$

Proof: The verifier can verify the public keys in the bill-board. Thus, to obtain the transformation pair $(S_{2,i}, T_{2,i})$ from verifier and produce the signature, the user i has to obtain the qualification from the group director.

- **Undeniable:** Both the group user and the group director are not capable to sign the documents on behalf of other group users.

Proof: Assume that the group user j needs to sign documents on the behalf of a group user i . The user j can inquire the group director for the affine transformation pair (S_1, T_1) and inquire a verifier for an affine transformation pair (S_2, T_2) . Then he finds $(A_i^{(2)}, B_i^{(2)})$:

$$\begin{aligned} A_i^{(2)} &= S_2 \circ S_1 \circ A_i \circ T_1 \circ T_2 \\ B_i^{(2)} &= S_2 \circ S_1 \circ B_i \circ T_1 \circ T_2 \end{aligned}$$

To produce the signature of knowledge $(A_i^{(2)}, B_i^{(2)})$ he must discover the isomorphism from $(A_i^{(2)}$ to $B_i^{(2)})$. It means that he must resolve the isomorphism of polynomial problem. It is arithmetically difficult without the secret key (S_i, T_i) of user i .

- **Anonymity:** Provided the valid group signature, no one can disclose an identity of the real signer, unless the group director opens the signature with an assist of the verifier.

Proof: Provided the group signature $[V_i^{(2)}, (A_i^{(2)}, B_i^{(2)})]$, to disclose an identity of the

signer, a hacker has to discover the affine transformation pair (S'_i, T'_i) that maps (A_i, B_i) to $(A_i^{(2)}, B_i^{(2)})$. This means that he has to resolve an isomorphism of polynomial problem. It is arithmetically difficult without $(S_{1,i}, T_{1,i})$, and $(S_{2,i}, T_{2,i})$ which hold by the group director and the verifier respectively.

COMPARISONS

We compare the proposed scheme with some other group signature schemes. The proposed scheme is the group signature scheme using multivariate asymmetric key cryptography, and is possible the first group signature scheme that withstands the future attacks of quantum computers. Thus, we can only compare the proposed scheme with the schemes using conventional cryptography. The early proposed group signature schemes are impractical for large groups, as the size of signatures. In the proposed scheme, the calculation of the operations are based on the number of group users. The size of signatures and the size of user keys are determined. But the size of keys of group director are linear in the number of group users, thus decreasing the storage space of the scheme can be part of the future work. But, this scheme is efficient for large groups. Also, compared with other schemes, the proposed scheme can be used in e-voting schemes directly because of two unique characteristics, the linkability and the traceability. Table 2 shows the comparisons between dissimilar group signatures.

Table 2. The comparisons between dissimilar group signatures

Characteristic	Chaum [1]	Petersen [25]	Ateniese [5]	Proposed Scheme
Cryptography Assumption	Discrete Log	Discrete Log	RSA and Diffie-Hellman	Isomorphism of Polynomial
Quantum Attack	Yes	Yes	Yes	No
Sign	$O(n)$	$O(n)$	Constant	Constant
Verify	$O(n)$	$O(n)$	Constant	Constant
Open	$O(n)$	$O(n)$	Constant	Constant
Size of Signature	$O(n)$	$O(n)$	Constant	Constant
Size of Key	$O(n)$	Constant	Constant	Director: $O(n)$ User: Constant

FUTURE RESEARCH DIRECTIONS

In the future work, we will consider conditions that undertaking an attack of Faugere and Spaenlehaner versus Akiyama *et al.* to make it works better. Also, we can produce keys that are more arbitrary than the keys of Faugere and Spaenlehaner applied then we can proved the effectiveness of their attack by our tests. In their attack we can infer data of the divisor polynomial f by finding the resulting of selected polynomials according to the encrypted document. Then, we can examine the scheme by using the theory of Gröbner basis computation.

However, we can introduced another multivariate public-key system by applying wild automorphisms, which are deeply examines in affine algebraic geometry. However, we create the private key as selected polynomial functions created by ρ – derivations, and announce the composition of such functions. As, we employ polynomial functions which ranges are high, the length of keys and the time complexity of scheme tend to be considerable. Thus, we have to decrease the key length and the time complexity.

CONCLUSION

This Chapter presents the group signature scheme under multivariate asymmetric key cryptography. Its security based on the isomorphism of polynomial problem and the secure hash function. Compared with other group signature schemes, the proposed scheme is design to e-voting schemes. Because of two characteristics, the unlinkability and the traceability, the proposed scheme can convince the additional requirements of the e-voting schemes. Also, the size of signatures and a calculation overhead are regardless of the number of group users. It is efficient for large groups.

REFERENCES

- Akiyama, K., Goto, Y., & Miyake, H. (2009). An Algebraic Surface Cryptosystem, public-key cryptography-PKC09. Lecture Notes in Computer Science, 5443, 425-442.
- Ateniese G., Camenisch J., Joye M., and Tsudik G., (2000). A practical and provably secure coalition-resistant group signature scheme. In *Advances in Cryptology-CRYPTO 2000*. Springer.

- Camenisch, J., & Stadler, M. (1997). Efficient group signatures schemes for large groups. In *Advances in Cryptology-CRYPTO 1997*. Springer.
- Chaum, D., & Heijst van. (1991). Group signatures. In *Advances in Cryptology-EUROCRYPT 1991*. Springer.
- Chen, L., & Pedersen, T. (1994). New group signature schemes. In *Advances in Cryptology-EUROCRYPT 1994*. Springer.
- Ding, J. (2004). A new variant of the Matsumoto-Imai cryptosystem through perturbation. In *Public-key cryptography-PKC 2004*. Springer.
- Ding, J. (2011). *Method to produce new multivariate public key cryptosystems*. Patents US 7961876 B2.
- Fangere, J., & Spaenlehaner, P. (2010). Algebraic Cryptanalysis of the PKC2010 Algebraic Surface Cryptosystem. Public-Key Cryptography PK-2010. Springer.
- Huang, N. (2014). Analysis of a multivariate public-key cryptosystem and Its application in software copy protection. *Journal of Software*, 9(8), 8. doi:10.4304/jsw.9.8.2010-2017
- Li. (2011). *Gröbner Bases in Ring Theory*. World Scientific Publishing.
- Ogura, N. (2012). *On multivariate public-key cryptosystem* (PhD Dissertation). Tokyo Metropolitan University.
- Patarin, J. (1996). Hidden fields equations and isomorphism of polynomials: Two new families of asymmetric algorithms. In *Advances in Cryptology-EUROCRYPT 1996*. Springer.
- Petersen, H. (1998). How to convert any digital signature scheme into a group signature scheme. In *Security Protocols*. Springer.
- Petzoldt, A., Bulygin, S., & Buchmann, J. (2010). A multivariate signature scheme with a partially cyclic public-key. *Proceedings of the 2nd international conference on symbolic computation and cryptography (SCC 2010)*, 229-235. doi:10.1007/978-3-642-17401-8_4
- Shor, P. (1994). Algorithms for quantum computation: Discrete logarithms and factoring. *Proceeding 35nd Annual Symposium on Foundations of Computer Science*, 124-134. doi:10.1109/SFCS.1994.365700
- Yang, B., & Chen, J. (2003). A more secure and efficacious TTS signature scheme. In *Information Security and Cryptology-ICISC 2003*. Springer.

KEY TERMS AND DEFINITIONS

Cryptography: Is the science and study of secret writing.

Hacker: Is a term used by some to mean a clever programmer and by other, especially those in popular media to mean someone who tries to break into computer systems.

Group Signature Scheme: Is a method for allowing a member of a group to anonymously sign a message on behalf of the group.

Multivariate Cryptography: Is the generic term for asymmetric cryptography based on multivariate polynomials over finite fields. Solving systems of multivariate polynomial equations is proven to be NP-hard. Today multivariate quadratics can be used only to build signatures.

Quantum Computers: Is the direct use of quantum-mechanical phenomena, to perform operations on data. Quantum computers are different from digital computers based on transistors. Whereas digital computers require data to be encoded into binary digits (bits), quantum computation uses quantum bits which can be in super positions of states.

Isomorphism of Polynomials Problem: Is a well-known problem studied in multivariate cryptography. It is related to the hardness of the key recovery of some cryptosystems.

NP-Hard Problems: Is a class of problems that are, at least as hard as the hardest problems in NP. More precisely, a problem H is NP-hard when every problem L in NP can be reduced in polynomial time.

Hexa-Dimension Code of Practice for Data Privacy Protection



Wanbil William Lee

Wanbil & Associates, Hong Kong

INTRODUCTION

Contemporary cyberspace inhabitants live and work in a *technology-driven information-intensive* era, a *phenomenon* born out of a *vicious circle*. The consequence is a mixture of blessing and nightmare. Data protection emerges as a critical concern and data privacy protection an urgent and vital problem for information security management. Given the situation, the front-line information security personnel is among the first to bear the brunt and in dire need of a pragmatic guidance.

A recently developed set of International Data Privacy Principles was proposed as a reference considered to be useful for tackling the *first need* (Zankl, 2016). Hong Kong's Personal Data (Personal) Privacy Ordinance (PDPO), a data protection principle-based law like many others legislated by western jurisdictions, which has been in force for a number of years (Chang, 2016), can make a contribution the *second need*. To address the *third need*, a framework that comprises a 6-d code based on the 6-d metric and an operationalization scheme is recommended.

A first-cut version of the framework was recently presented to an audience of Information Security Management specialists (Lee, 2015a). The rationale of the metric, the definition of the code and its worthiness for recommendation, and an indicative guideline to operationalize the code, are described in this article.

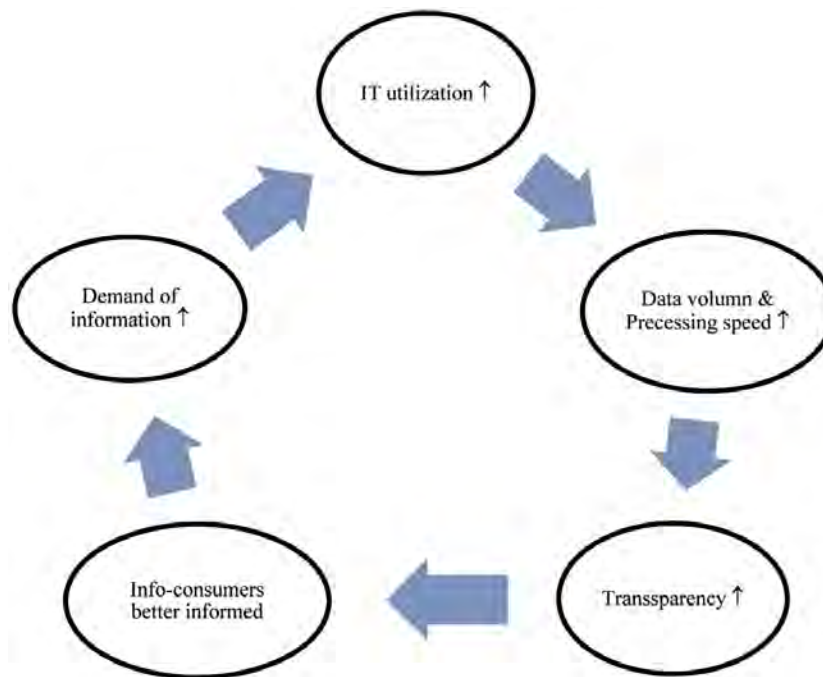
BACKGROUND

Technology-Driven Information-Intensive Phenomenon and the Vicious Circle

Netizens are provided with such technologies as *Customer Relationship Management*, *Web-lining* and *Call Centre*, and so on; they can by means of these facilities conduct their daily activities more efficiently and effectively, and optimize the outcome of these activities, because they are better-informed and able to innovate marketing, to accelerate business promotion, to enlarge data storage capacity and communication coverage, to increase retrieval facilities, and to improve transaction speed in a more transparent and open environment. But then they will need to rely increasingly heavily on the technologies. While transparency and communication keep on improving, more and more data are consumed and correspondingly generated. Briefly, "the happier the consumers of information, the higher the demand for more information and the more the technology is used, the more the data and the speedier the processing, the more transparent the cyberspace, and the happier the consumers" – a *vicious circle* is formed; the *technology-driven information-intensive phenomenon* created. (See Figure 1).

The consequence is good and bad. The good is the accelerated arrival of such technologies as Big Data, Cloud Computing, Internet of Things

Figure 1. A conceptual graph of the circle and the phenomenon



and a plethora of social engineering tools. These technologies enable integration of massive, scattered datasets, efficient interpretation of the integrated data, and speedier communication of the information. An obvious benefit is that with a huge amount of information being made available, the cyber-world becomes more transparent and netizens are better informed. And the bad is that there emerges a plethora of additional security threats bred in the loopholes in the new technologies, in the use of them or in the facilities enabled by the massive volume of data they generate, which the cyber-miscreants are ever lurking around to exploit when detected. However, that numerous clandestine activities are brought to light, for example, the Snowden episode (*South China Morning Post*, 2013) and the Panama Papers leak (Wilson, 2016), can be beneficial to some people/organizations and adversary to others.

Data Privacy Protection Problem

The problem is rooted in the way the data are collected about and are used adversely against

data subjects, and the *right* of the data subjects to that data, and has to deal the *techno-ethical-risk* which is originated in the way the technology is applied and the human users of the technology and the data, though the technology per se is neutral, yet the use of the technology is not (Parker, 1986; Neumann, 1995; and Williams, 1997). The protection of data privacy means to ensure that data privacy is not breached, that is, the data content is securely stored and accessed only with proper authority, and that the right of the data subjects not abused. Solving the problem means developing, reviewing, and implementing information protection and data privacy policies, standards, guidelines and processes (the policies), and mitigation measures; and to identify new technology risks; and to ensure that the policies are appropriately enhanced, communicated and complied with, and that mitigation measures are properly implemented.

The issue of data privacy invasion and the harmful impacts thereof have extended beyond the individuals and corporations to the regional/national governments, instanced by the UK House

of Commons' advocating a data ethics council and more urgency on data privacy, and sometime later the UK Government accepting the data ethics council proposal (*Computer Weekly*, 2016). These are strong reasons to suggest that data privacy security is no longer a personal and corporate concern but also a concern of governments, and international organizations, much more so than ever before.

Data Privacy and Concept of Data Privacy Category

Data privacy, which is basically about collecting and using the data about individuals and organizations, and the right of the data subjects to that data, is hitherto taken as a personal (individual) issue; it can be impersonal (corporate) as well. It is *personal data privacy* if the data subjects refer to individuals; it is *corporate data privacy* if the data subjects refer to organizations including government agencies. For example, the secret formula that a corporation adopts in determining the special fringe benefits in addition the total annual remuneration (which is normally published in the annual report) as a means to retain the service of their top-ranked staff is not trade secret but corporate privacy, and accessing or revealing without authority is an act of breaching corporate data privacy.

This the categorization is necessary. There are two supporting reasons for this view. The first reason is that an individual making a decision differently depends on his/her role: "as a member of an organization, the individual makes decisions not in relationship to personal needs and results, but in an impersonal sense as part of the organizational intent" (Barnard, 1938). The second reason is that the consequences of data privacy breach might mean differently to the individual vis-à-vis a group (organization) due to four considerations. The first consideration is the data owner's right to control over the access to and use of the information. The second consideration is the owner's integrity and provision of confidential data that include such

personal data as biometric measurement (in the case of individuals) or corporate data as the formula adopted by a company to determine the hidden portion of benefits in each senior executive's total annual remuneration (hidden because such information is usually withheld from a company's annual report) (in the case of business concerns) (bodily privacy). The third consideration is the data being observed by surveillance and the like (behavioural privacy). The last consideration is freedom of communication that includes speech or expression and dialogue with third parties (communicational privacy).

The dichotomy is recognized in the International Data Privacy Principles (Zankl, 2016), but generally missed in the extant data privacy laws which are inherently inadequate or unable to judge corporate data privacy infringement. It pays, it is argued, all stakeholders, particularly those involved in law making and enforcement, and the information security scholars and practitioners to be clear about the hitherto muddled concept.

Data Privacy Protection and Changes in Demand and Status

Data protection has long been regarded as a technical and economic matter and requires technical actions to ensure accuracy, completeness, integrity and availability of the data content, and to admit only authorized access. However, the impact of scope and level of data privacy breach has changed, extending from personal/corporate to governmental and the UK government pronouncement alluded earlier proves a case in point). Also, the nature data privacy breach has changed because the problem and the consequences thereof are found to be of techno-economic as well as legal, ethical, social and ecological.

Limitations of the Extant Policies and Code of Conduct/Practice

The policies and standards in current use are invariably *dated* due to the rapid advances of the



technologies, *limited* to cope with techno-socio threats, a product of the technology-driven information phenomenon, and *inadequate* to deal with the well-equipped and cunning cyber-miscreants. This is an obvious shortcoming.

Further, the policy statements tend to be vague and less than user-friendly or simply difficult to absorb and follow, easily forgotten and probably ignored in the end. For example, a 70-page Code of Professional Conduct for the Registered Medical Practitioners (Hong Kong Medical Council, 2009) is certainly unpalatable for the busy doctors. Another case in point is that when asked, the Deputy Commissioner of a police force could not tell what the code was, though he admitted being aware of one in existence and posted on the force's web-site. Alluded to earlier, the text of the policy and the code tend to be difficult to comprehend, wordy and lengthy. This is another limitation.

On the one hand, the extant policies usually focus on the quantitative/tangible/technical efficiency (aiming at achieving targets against such indicators as *return on investment* (ROI), *net present value* (NPV) and *payback period* based on cost-benefit (Portney, 2008) and risk (Wharton, 1992; Moore, et al, 1976) considerations), and satisfying the minimum requirements of the law), and in some cases, ethical acceptability embedded in professionalism but hardly touch or miss the qualitative/intangible social desirability (despite the rhetoric by some organizations in response to *corporate social responsibility* CSR) and ecological sustainability (despite Green Computing which advocates environment-friendly and paper-free design and application). On the other hand, professionals (in the field of Science and Technology, like others) are nurtured under the influence of the *misinterpretation of risk* and the *flawed curricula* across science and technology (Lee, 2015b), and their decision-making mentality developed along the Simon doctrine which is essentially to “select the alternative, among those available, which will lead to the most complete achievement of your goals” (Simon, 1976b).

The goals are not necessarily personal despite the personal possessive adjective, according to Barnard (1938) who asserted that “the decisions that an individual makes are quite distinct from his personal choice” so as a member of an organization, the individual makes decisions not in relationship to personal needs and results, but in an impersonal sense as part of the organizational intent. This is why the policy, a product of our decision, and also the code of practice/conduct is inherently technical and financial, and evaluated in terms of financial performance/economic value, and found short when dealing with many a problem commonly found in *post-implementation* and *post-contract* situations, for example, that IT professionals face daily are of an ethical and/or a social nature. This is an inherent weakness, the hidden reason.

Given a code of conduct, an organization lacks effective measure to encourage staff to familiarize with and to act according to the code. It is not uncommon to find that the code exists in name and not in substance. This is arguably a limitation, definitely an operational issue.

RECOMMENDED SOLUTION

Information security policy is propelled to centre-stage by the protection problem, and the front-line functional group is among the first to bear the brunt. To alleviate some burdens on the front-line information security staff who are always working against a tight schedule, a handy reference or a quick guide in the form of a code of conduct or practice, into which the gist of the adopted policies is incorporated, as an aid, was conceived. This document will contribute to cultivate a corporate culture of ethics conducive to the compliance with the data privacy protection policy organization-wide. A framework that comprises a 6-d code based on the 6-d metric and an operationalization scheme to make the code work, is proposed.

The Hexa-Dimension Metric and its Rationale

The hexa-dimension LESTEF (pronounced “*lest ye forget*”), which was initially conceived as a checklist for decision making in technology-driven and information-intensive organizations, is proposed for developing a code of practice or corporate charter as an aid for ensuring and enhancing information security policies including data privacy preservation. By implication, hefty amount of money is invested in the expensive technologies, thus it is a matter of course to expect a justifiable return on that the investment (*financial viability*) and an optimized utilization of the technologies (*technical effectiveness*). In the same vein, huge amount of information that is generated can be abused, thus *legal validity*, *ethical acceptance*, and *social acceptability*. Running the technologies consumes a lot of natural resources (electricity, water, paper, etc.), and generates air, noise pollution, thus *ecological sustainability* – the six attributes representing efficiency (requirements) of the hexa-dimension model.

The Hexa-Dimension Code of Practice/Conduct

Code of Conduct

In general, a code of conduct or practice comprises a set of rules and regulations that encapsulate the gist of the policies and standards adopted by the organization. It has arguably the capability to provide a guide, a handy reference, and a reminder to the stakeholders, especially the constantly busy front-line operational staff.

If properly and appropriately formulated and articulated, the code would be useful in permeating the policies and standards throughout the organization and around, thus cultivating a corporate-wide ethical, professional conduct. While such codes may serve to deter potential offensive actions, they are limited in driving home those rules or standards as they can rely only and purely on the

moral obligation of the stakeholders concerned primarily because violation of the code by itself does not in general attract any criminal charges. Auxiliary measures must put in force to arrive at desirable results, such as executive actions that empower to grant rewards and to impose punishment (for example, annual performance appraisal).

The Hexa-Dimension Code

Organizations of various varieties might have some kind of code of practice in place. However, the extant codes invariably tend to focus on technical, financial and probably legal issues; they are insufficient when considering also social, ethical and probably ecological issues. The hexa-dimension code based on the hexa-dimension model poises to make up for the deficiency.

Operationalization Schedule

Prerequisite

To operationalize requires the support of the Board of Directors with respect to corporate policy aspects and the supporting infrastructures that include the organization’s human resources management (HRM), legal, finance, and ICT functional units with respect to technical support and reference, and mandatorily, and an appraisal of ethical consistency in conduct be included in an annual performance review (by the HRM Department). In addition, and ideally, GCF Computing (Patrignani & Whitehouse, 2014) be adopted. The characteristic features of GCF Computing are provided in Table 1.

Procedure

There are three major preparatory steps (summarized in Table 2):

- Identify the relevant critical factors depending on the target end-users (corporate-wide or a functional unit, nature of opera-



Table 1. GCF computing

GCF	Target	Enabler	Contribute to
Good Computing	Workable & technically efficient deliverables	Internet technologies, Big Data, Cloud Computing, Ubiquitous computing (Evasive Computing), etc.	Legal, Technical & Financial efficiencies
Clean Computing	Energy & materials saving	Green Computing	Environmental efficiency
Fair Computing	Corporate social responsibility	Ethical Computing (Lee, 2015b)	Social & Ethical efficiencies

Table 2. Schedule of operationalization

Generic				User specific	
Attributes/ Factors	Remarks	Control	Targets	Drivers (*)	Supported by
Legal validity	Abiding the law, rule & regulations currently in force	External	Local law and convention	Corporate counsel /legal dept.	Good Computing
Social desirability	Beneficial to all stakeholders including the general public	External	Local culture and habit; standards for perceivable benefits to primary stakeholders and the public		Fair Computing
Ecological sustainability	Friendly to natural and industrial/business environment	External	Policy/guidelines for energy/ materials saving; ecology related law (if any)		Clean Computing
Ethical acceptability	Consistent with ethical principles	Internal	Corporate social responsibility; corporate culture	PR, HRM	Fair Computing
Technical effectiveness	Use up-to-date technology to meet targets effectively and efficiently	Internal	Sufficient & necessary human & technological resources to support infrastructural support	ICT, HRM	Good Computing
Financial viability	financially sound and viable	Internal	Corporate policy (performance metrics: operational goal, risk absorption standards & limits, etc.)	Finance, Operations	Good Computing

tion – for example, environmental impact is critical for a mining company or a factory, but may probably be skipped for an Information Security Unit)

- Secure the support of the Board of Directors and the supporting infrastructures including in an annual performance review to assess awareness and compliance of the code in carrying out assignments/duties (by the HRM Department).
- Determine a schedule for quantifying the elements of each factor for measuring, prioritizing and balancing the factors. Depending on the attributes/factors deter-

mined, the steps to be taken to measure the effectiveness.

Briefly,

- Legal validity is to ensure that the laws current in force and the organization's rules and conventions are abided at the advice of the corporate counsel or the Legal Department, and is measured by the extent and degree the law is obeyed.
- Social desirability is to make sure that the benefit of all stakeholders including that the public has been considered and deliberated

consistent with local culture and habit and can be measured by survey results of the extent of public benefited based on the ethical principles of utility and consequence.

- Ecological sustainability means the actions taken must be friendly (not harmful) to the environment and is measured by indication of energy/materials saving, noise and air pollution, radiation emission rate and so on using EIA (environmental impact assessment) (Hong Kong Government).
- Ethical acceptability demands that the actions taken must be consistent with ethical principles with the help of the corporation's Public Relations Unit and Human Resource Management Department, and can be measured using the Ethical Matrix method (Lee, 2015b) to assess the consistency with the relevant ethical principles of duty and reciprocity.
- Technical effectiveness means using up-to-date technology to meet targets effectively and efficiently.
- Financial viability is to ensure that the actions taken are financially sound and viable consistent with the corporation's performance metric on operational goal, risk absorption standards and limits, etc. as advised by the Finance and Operations experts, and is measured by ROI and NPV.

FUTURE RESEARCH DIRECTIONS

The concept that a decision should meet the six criteria is new, but each criterion in the hexa-dimension metric model is not new. The metric has the potential to serve others purposes, for example, Management Science, in enhancing identification and evaluation of the relevant factors of the traditional and popularly adopted decision model like the rationality and logic based Simon model (1976) which is no longer sufficient because decision makers are not always rational and sometimes non-rational, or judgmental decision making based procedural or bounded rationality

(Gigerenzer, 2001) and escalation of commitment decision making that “under certain conditions, decision makers who make an initial decision become overly committed to the original choice and then subsequently make decisions biased by psychological commitment” (McCarthy, Schoorman and Cooper (1993, p. 9). Also, that the decision variables are formulated and measured mainly in economic terms must be updated as they are also legal, ethical, social, and more recently, ecological in nature. This raises many questions that seek an answer, to name a few:

- The hexa-dimension code can be a prototype for designing a code of practice for an IT organization, as proposed in this article. At this exploratory stage, is efficiency needs to be demonstrated, field-tested and validated against actual results. An on-site study of its effectiveness as a handy reference at the different levels of the corporate is also a worthwhile exercise.
- The code, which aims to be a handy reference for the stakeholders, particularly employees, would not by itself brings about the results desired; supporting infrastructures including human resources management action such annual performance appraisal and so on, must be in place and work in concert. This raises two research questions: How the mechanisms concerned are coordinated and cooperated, and can corporations or functional units expect an effective enforcement of the code of practice and an ethical atmosphere be cultivated corporate-wide or within a functional unit after a compulsory annual appraisal scheme that includes an assessment of ethical/social elements in the assigned duty is in force.
- An empirical study of applicability and acceptability of the hexa-dimension metric to enhance the chance and scope of finding factors likely to bring about ethical, social and ecological repercussion on decision making warrants serious consideration.

CONCLUSION

The information security professionals are in urgent need of effective and pragmatic guidance as the information security function in a technology-driven information-intensive environment becomes more complicated and harder due to the new socio-techno risks and data privacy protection becomes a primary concern as privacy infringement occurs more frequently and leads to potentially devastating consequences to the individuals, corporations and governments. The newly developed framework, though at this embryonic stage of growth still awaiting empirical evidence and popular acceptance in the IT arena, poises to play a useful role, in addition to the Personal Data (Privacy) Ordinance (Hong Kong's privacy law) (Chang, 2016), the International Data Privacy Principles (Zankl, 2016), to release the burden on as the extant policies and standards. It noteworthy that

- An international standard is useful and necessary when regular (hard) law often fails to handle issues such as those arising from inter-border data transfer or cross-border trading and when formulating corporate policies that meet not only legal but also ethical demands.
- Privacy laws of many jurisdictions are based on the principle-based OECD Privacy Guidelines. Many of these principles have been implemented by jurisdictions across the globe including Australia, Canada, Hong Kong and the UK, and the ethical dimensions of these principles have been realized and expressed in legal terms. Organizations have found themselves insufficient to follow just the letters of the law, but also the spirit, as the public understand and expect their privacy rights.
- The hexa-dimension Code of Conduct/ Practice aims at alerting stakeholders to

exercise their duties such they are not breaking the law in force; not harmful to the individuals and society at large; not wasteful of the resources available (the computer facilities, the workforce, the budget, etc.); and are environmentally-friendly. Since different organizations have their own unique policies, thus code of conduct, there can be no universal receipt but a general guideline. Since the code by itself is limited in enforcing compliance with the policies and standards, a framework was developed and a 3-step procedure are recommended.

- Decision makers may face unavoidable competing circumstances when considering the issues of ethical and social nature, and balancing the competing issues is necessary in resolving them. Apart from the linear menus, the Ethical Matrix (Lee, 2015b) is worthy of consideration.
- The hexa-dimension code of practice will be of interest not only to the information security personnel but also to the practitioners and scholars in Information Science and Technology and perhaps other fields such as Information Systems Audit, Management, and Public Policy.

REFERENCES

- Association for Computing Machinery. (1992). *Code of Ethics and Professional Conduct*. Retrieved February 13, 2016, from <http://www.acm.org/about-acm/acm-code-of-ethics-and-professional-conduct>
- Barnard, C. I. (1938). *The Functions of the Executive*. Cambridge, MA: Harvard University Press. Retrieved April 13, 2016, from https://en.wikipedia.org/wiki/Herbert_A_Simon#CITEREFBarnard1938

House of Commons Science and Technology Committee calls for Data Ethics Council. (n.d.). Retrieved February 13, 2016, from <http://www.computerweekly.com/news/4500272963/House-of-Commons-Science-and-Technology-Committee-calls-for-Data-Ethics-Council>

Government accepts data ethics council proposal. (n.d.). Retrieved on 28 April 2016, from <http://www.computerweekly.com/news/450294474/Government-accepts-data-ethics-council-proposal>

Donaldson, J. (1992). The Ethics of Risk Management. In J. Ansell & F. Wharton (Eds.), *Risk Analysis, Assessment and Management*. Chichester, UK: John Wiley & Sons.

Gigerenzer, G. (2001). *Decision Making: Nonrational Theories*. In *International Encyclopedia of the Social and Behavioral Sciences* (Vol. 5, pp. 3304–3309). Oxford, UK: Elsevier. doi:10.1016/B0-08-043076-7/01612-0

Hong Kong Government. (n.d.). *A guide to the EIA (Environmental Impact Assessment) Ordinance*. Retrieved on April 25, 2016, from <http://www.epd.gov.hk/eia/english/guid/index1.html>

Lee, W. W. (2015a). Ignorant of ethical pitfalls can hurt information security: Implications for IT users & Information Security Professionals. Special Seminar on Information Security, the Hong Kong Jockey Club.

Lee, W. W. (2015b). Ethical Computing. In M. Khosrow-Pour (Ed.), *Encyclopedia of Information Science and Technology* (3rd ed.; pp. 2991–2999). doi:10.4018/978-1-4666-5888-2.ch292

McCarthy, A. M., Schoorman, F. D., & Cooper, A. C. (1993). Reinvestment decision by entrepreneurs: Rational Decision-making or escalation of commitment? *Journal of Business Venturing*, 8(1), 9–24. doi:10.1016/0883-9026(93)90008-S

Moore, P. G., Thomas, H., Bunn, D. W., & Hampton, J. M. (1976). *Case Studies in Decision Analysis*. Penguin Books.

Neumann, P. G. (1995). *Computer Related Risk*. New York, NY: ACM Press /Addison-Wesley.

Parker, D. B. (1986). Computer Crime. *Financial Executive*, 2(12), 31–33.

Patrignani, N., & Whitehouse, D. (2014). Slow Tech: A quest for good, clean and fair ICT. *Journal of Information, Communication and Ethics in Society*, 78-9.

Portney, P. R. (2008). Benefit-Cost Analysis. In D. R. Henderson (Ed.), *Concise Encyclopedia of Economics* (2nd ed.). Indianapolis, IN: Library of Economics and Liberty.

Simon, H. A. (1976a). From Substantive to Procedural Rationality. In S. J. Latsis (Ed.), *Method and Appraisal in Economics*. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511572203.006

Simon, H. A. (1976b). *Administrative Behaviour* (3rd ed.). New York: The Free Press.

South China Morning Post. (2013). Retrieved April 17, 2016, from <http://www.scmp.com/news/hong-kong/article/1259422/edward-snowden-let-hong-kong-people-decide-my-fate>

Wharton, F. (1992). Risk Management: Basic Concepts and General Principles. In J. Ansell & F. Wharton (Eds.), *Risk Analysis, Assessment and Management*. Chichester, UK: John Wiley.

Williams, K. (1997). Safeguarding Companies from Computer/Software Fraud. *Management Accounting*, 78(8), 18.

Wilson, S. (2016). The Panama Papers scandal. *MoneyWeek*. Retrieved April 17, 2016, from <http://moneyweek.com/the-panama-papers-scandal>



ADDITIONAL READING

Berthold, M. (1994). Hong Kong's data privacy proposals, *Privacy Law and Policy Reporter*, 124. Retrieved February 13, 2016, from <http://www.austlii.edu.au/au/journals/PrivLaw-PRpr/1994/124.html>

International Journal of Green Computing (IJGC). Retrieved February 13, 2016, from www.igi-global.com/journal/international-journal-green-computing.../1175

Lee, W. W. (2014). Why Computer Ethics Matters to Computer Auditing. *ISACA Journal*, 2, 48–52.

Lee, W. W., & Chan, K. C. C. (2008). Computer Ethics: a Potent Weapon for Information Security Management. *Information Systems Audit & Control Journal*, December (www.isaca.org/jonline)

Moore, P. G., & Thomas, H. (1976). *The Anatomy of Decisions*. Marmonsworth: Penguin Books. *OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data*. Retrieved February 13, 2016, from <http://www.oecd.org/sti/ieconomy/oecdguidelinesonthe protection of privacy and transborder flows of personal data.htm>

Pazowski, P. (2015). *Green computing: latest practices and technologies for ICT sustainability*. Retrieved February 13, 2016, from www.toknowpress.net/ISBN/978-961-6914-13-0/papers/ML15-377.pdf

PDPO. (Hong Kong's privacy law) *The Personal Data (Privacy) Ordinance, Chapter 486*. Retrieved February 13, 2016, from <https://www.pcpd.org.hk/english/files/pdpo.pdf>

KEY TERMS AND DEFINITIONS

Data Privacy: Can be personal data privacy if the data subjects refer to individuals or corporate data privacy if the data subjects refer to organizations including government agencies.

GCF Computing: Slow Tech is not a technology that is slow but a movement that parallels with the same concept of the Slow Food movement. Slow Food International claimed in 1989 that the Slow Food Movement was set up to “counter the rises of fast food and fast life”: to strive for good food (food that tastes good, is a pleasure to eat and selected according to its quality), clean food (food that is produced by such a process that respects the environment and should promote biodiversity and sustainability), and fair food (food that is cultivated and produced by ways that must respect the farmers). GCF Computing (Good Computing, Clean Computing, Fair Computing) was proposed and defined following the Slow Tech concept (Patrignani & Whitehouse, 2014).

International Data Privacy Principles: Comprise 13 principles (Zankl, 2016).

Operationalization Scheme: Acquire Board's endorsement and infrastructural support from other relevant departments, and determine the relevant criteria/factors according to the nature of the problem and the target users and a quantification system for measuring against and balancing the criteria.

PDPO Data Protection Principles (DPP): (1) Data Collection and Purpose; (2) Accuracy and Retention; (3) Data Use; (4) Data Security; (5) Openness; and (6) Data Access and Correction (https://www.pcpd.org.hk/english/data_privacy-law/ordinance_at_a_Glance/ordinance.html).

PDPO: Personal Data (Privacy) Ordinance (Hong Kong's privacy law), enacted in 1995 (<https://www.pcpd.org.hk/english/files/pdpo.pdf>).

Post-Contract Problem (an example): The specification was approved by provider and client, and contract duly signed. A senior project consultant who was assigned to the project discovered a fault in the specification, a fault in the inventory control function. Inventory control is a critical function for the business of the client, a fashion boutique, because supply and demand of the goods for sale is time-critical and zero error-tolerance is

expected according to the specification. The fault was confirmed after the consultant's site visit. The consultant faces the dilemma: Keep quiet or tell the boss or the client or both.

Post-Implementation Problem (An Example): An online monitoring system was implemented to replace an existing offline help-desk platform. The Executive Vice President (EVP) is impressed by the performance, particularly the online monitoring capability provided by the system, and asks the Chief Information Officer (CIO) to have a copy of the system installed in her office to track drug dealing allegedly taken place on company premises. Dilemmas that need to be addressed include deviation of approval of acquisition, privacy invasion, corporate image damage, professionalism (CIO), duty (EVP); treated as a means (staff); staff morale & corporate policy (Firm).

Technology-Driven Information-Intensive Phenomenon: A state that characterizes the modus operandi in cyberspace and attributes to the vicious circle: More and more information is needed to sustain more transparent business operations, thus satisfying the more demanding consumers of information; higher utilization of more technologies is needed to generate more information to handle the enlarged data volume, and to further increase the processing speed.

The Data Privacy Protection Problem: To develop and implement data privacy policies, standards, guidelines and processes (the policies), to ensure that the policies are appropriately enhanced, communicated and complied with, and to devise a set of effective mitigation measures.

The Solution: A *hexa-dimension code of practice*, a list of rules and regulations that encapsulates the gist of the policies and standards adopted by the organization, and embraces the six criteria.

Information and Communication Technology Ethics and Social Responsibility

Tomas Cahlik

Charles University Prague, Czech Republic & University of Economics Prague, Czech Republic

INTRODUCTION

Information and Communication Technologies (ICTs) have penetrated during the last 20 years all human activities everywhere on the Earth. Humanity has entered into the information age, virtual reality and even virtual worlds have been created.

The basic ethical questions stay as they have always been: How are we to live? What are we to be? Basic answers are, of course, that we ought to live good lives and be good persons.

The aim of this article is:

- To specify what “living a good life” and “being a good person” could be in the information age;
- To identify some challenges and opportunities ICTs offer in this context.

Having absolutely stabilized basic questions and basic answers makes the methodology of ethics quite different from the methodology in sciences. In sciences, one starts with a thorough review of previous research, specifies some new and interesting research question, makes hypotheses about possible answers and bases argumentation on data. In ethics, one reflects problems of the current age in a mirror that was created centuries ago and has been polished by many ethical reflections ever since. Forms of ethical texts are rich: dialogs, even poems, but the most used form is an essay.

BACKGROUND

Literature review in research articles is used for showing that the research described in the article fits into research themes that are interesting for contemporary research community. Literature review in ethical reflections is used differently, just for illustration of ideas that have been published in the area of interest and for “opening the scene”.

Looking into the Web of Science database in September 2015 and using keywords “information technology”, “ethics” and “social responsibility” 60 entries are obtained (from that 55 articles or conference proceedings), 31 entries being published since 2010. This reveals not high but steady and increasing activity on the interdisciplinary border between ICTs and that part of ethics that is linked to social responsibility.

Looking closer into the content of those articles, following themes can be identified in the last decade:

- Ethical questions linked with the creation and use of “big data”, including creation of agreed standards of good practice - e.g. (Rizk&Choueiri, 2006), (Light,& McGrath, 2010), (Celen, & Seferoglu, 2013);
- Development of sustainable information society - e.g. (Tsai&Chen, 2013), (Busch, 2011), (Niemela&Ikonen&Leikas&Kantola&Kulju&Tammela&Ylikauppi, 2014)



in the sense of an inclusive and environmentally friendly society; application of precautionary principle in the development of ICTs (Som&Hilty& Kohler, 2009);

- Corporate social responsibility of both ICT suppliers and users – e.g. (Tsai&Chen, 2013), (Busch, 2011), (Vaccaro&Madsen, 2009), including suggestions for standards of good practice (Patrignani&Whitehouse, 2014) and how to enable consumers to push companies to behave ethically with the use of ICTs (Watts& Wyner, 2011);
- University social responsibility (Arntzen, 2010); new teaching and learning culture based on ICTs (Stepien, 2010).

This indicates research activity that is driven by applications and can be contrasted with the research activity from the years before, that was pushed by theoretical considerations. (Lianos, 2000) e.g. starts with sociological concepts and identifies the threat that ICTs can atomize society through making development of personal trust obsolete. Lianos uses credit card as an example: one does not need to be trusted by the provider of money, the only thing that is relevant is the validity of the card. Technical norms replace moral and social norms.

Research fields may have different dynamics. It is quite usual that after the theoretical development some themes or even the whole field dissolve in applications. Nevertheless, after some time, both practitioners and applied researchers may find it useful to return to more generalizing theoretical reflection. In this article, the most general level of ethical reflection is being considered.

Ethics as practical philosophy offers a lot of valuable ideas that have been generalized from the real life problems of the whole human history. In this article, especially ideas developed by utilitarian philosophers, existential philosophers and the proponents of the “virtue ethics” can be identified. Two more areas of ethics linked with this article are “ethics of norms” and “casuistry”. Readers are advised to find more detailed summaries of those three schools and two areas on web.

“LIVING GOOD LIFE” AND “BEING A GOOD PERSON” IN THE INFORMATION AGE

Issues, Controversies, Problems

Let us identify three challenges and three opportunities ICTs directly offer in the context of ethics. This gives a framework to this article from which the discussion of economic costs and benefits stays mostly out, even if economic costs and benefits surely have ethical impacts.

Specification of the first challenge builds upon the above described conclusion in (Lianos, 2000) that technical norms replace moral and social norms. Problem is that the network of moral and social norms has always been considered as something absolutely necessary for the identity of specific society and that the dissolution of this network means atomization and threatens the whole society.

The second challenge is linked with the current research frontier of ICTs, with virtual reality. Economic and social thinking is strongly based on utilitarianism and its basic persuasion that for having a good life pleasures must be maximized and pains minimized. ICTs allow creating a virtual reality in which the choice of pleasures is almost unlimited and in which we are able to avoid pain. The basic question here is the authenticity of such a life. It can be easy and pleasant but it loses any sense.

Above the basic utilitarian life level, another level ought to be built in which we ask if our activities are right. What is the right activity has been discussed in Western thinking for centuries since Plato and two basic outcomes are as follows:

- In a right activity, humans are used not as tools only. E.g. authors of new software ought to take into consideration not only how much they earn but if the software allows better self-realization of users, too;
- Real circumstances must be taken into consideration in following the basic target of

“being better humans”. E.g. software firms must follow the profit motive as well as the moral one to stay in the business.

ICTs have not changed these basic outcomes but have surely changed the real circumstances and this opens the stage for discussion of what the right activities are in these changed circumstances. All new technologies have always done this, but what makes ICTs specific is the scale and speed of change.

One lives a good life if one realizes his or her potential in right activities. A person can develop its potential and become a virtuous person through cultivating its virtues and fighting its vices. Good people are virtuous people and there exists a positive correlation between being good people and doing right activities. So there exists a strong link between being a good person and living a good life.

The second challenge of ICTs lies in the possibility that people could be sucked too much into the virtual reality, enjoy pleasures there and try to avoid reality. Their lives would lose authenticity and they would not leave the basic utilitarian life level. Impacts on both the individual and society would be analogic to impacts of drugs addiction.

There exists a positive feedback between the first and second challenge: instead of being identified with a real society existing at some geographical place, people are identified with the usually anonymous social group sharing the same virtual reality. In the process of atomization of real society, the motivation for its members to find substitutes of social ties in virtual reality increases. This closes the loop.

In this context, we could ask about the geo-political impact of having socially atomized countries in which a lot of citizens avoid reality together with socially homogenous countries where people live more in reality. Let us imagine a possibly violent conflict and ask who would probably win in such a conflict? Or let us imagine mass immigration of people from the second type of countries to a country of the first type. What would happen? It is surely not socially responsible to ignore these scenarios.

The first opportunity of the virtual reality ICTs have created lies in quick detection of suspect economic, political and other social activities in a real society and in facilitating the discourse about such activities. By definition, discourse is discussion in which participants are free in the sense that they are not influenced by exogenous incentives or constraints in the expression of their opinions. Virtual reality, if it is free - without censorship - and anonymous offers an optimal environment for discourses on any topic. Discourses on suspect social activities help to create socially accepted moral norms and push firms and politicians to social responsibility.

The second opportunity lies in the enormous educational potential of the virtual world. For Aristotle, being educated was considered to be a virtue by itself, even if not being a moral virtue. For Plato, if people just knew what were right they would behave in that way. It may be too optimistic but we know that there exists a negative correlation between educational level in a country and e.g. crime.

Instead of discussing this intuitively clear opportunity on general level let us illustrate how ICTs can be used for benchmarking and looking for best practices in ethical issues. In (Codes of Ethics Collection), there are thousands of different ethical codes. Each of them presents norms that have been accepted by some organization (firm, association, university etc.). Ethical Codes are classified according to professional categories, one of them being “Computer and Information Science”. In September 2015, 56 organizations are found in this category, some of them with ethical codes for specific activities, some of them with different versions of the same code showing how it has been developed in time.

Any organization interested in development of its ethical norms can use this enormous source to compare the norms it already has with norms of its peers in the Codes of Ethics Collection. In an internal discourse, it can then formulate and accept the best for itself.

We can see another positive feedback, now between both opportunities: educational material easily accessible through ICTs can be used in discourses and additional knowledge created in discourses becomes easily accessible through ICTs.

The third opportunity lies in using ICTs for supporting the creation of virtues and the third challenge in supporting the creation of vices. The lists of moral virtues and vices have been changing during the centuries since Aristotle and there is not full consent about what ought to be on the lists. The most of virtues and vices get stronger with the repetition of either good or bad activities.

Let us take the (List of virtues) and (List of vices) from the webpage <http://www.virtuescience.com/>, choose some of them and analyze how they can be strengthened by ICTs. (We could surely take any other list, but the advantage of these two lists is that the list of virtues has 120 items and the list of vices 17 only. This supports the optimistic attitude towards human nature.)

“Compassion” and “Charity”: ICTs allow watching repeatedly and with high frequency poor people around the World what develops compassion among the most of watchers; psychologically especially among young people who are in the same time the most knowledgeable users of ICTs. This positive correlation can make this effect quite strong. ICTs offer a very easy way (e.g. donation through SMS) to do charitable activities what makes compassionate people to do such activities repeatedly. This strengthens the virtue charity.

“Curiosity”, “Commitment” and “Creativity”: Anderson in (Anderson, 2010) describes a nice example how web video powers global innovation. A slum boy watches video with street dance. Watching wakes up his curiosity and he makes commitment to master street dance himself. He surely has to be highly creative in transferring the knowledge from video into reality. There are millions of different activities and millions of web users, many of them behaving similarly as the slum boy, so ICTs have a very strong effect on these three virtues.

“Enthusiasm” and “Cooperation”: For economists, it is really difficult to explain the very existence of the open source software or Wikipedia because of non-existence of material incentives. Possible explanation is following: Knowledgeable users of ICTs are usually quite enthusiastic about the possibilities ICTs offer. They form a global community and identify strongly with this community. This has created the situation in which they like to cooperate not for material benefit but for achieving a shared goal. This works if individual costs of cooperation are negligible, what they with the use of ICTs can really be. Social benefits are enormous.

“Chastity” as a virtue or “Lack of Chastity” as a vice: In this context, pornographic content of web sites ought to be discussed. Watching pornography repeatedly creates habit - sometimes addiction – and weakens chastity. This discussion is not relevant for all World societies, e.g. in Japan people do not consider chastity as having sexual context. Watching pornography is clearly linked with the second challenge of ICTs discussed above; real sexual activity can shift to virtual reality what could have demographic impacts.

“Sloth”: Let us understand laziness as aversion to be active in the real world. Being active in that part of virtual reality that overlaps with reality perhaps weakens sloth; ICTs can increase the efficiency of our real activity what increases our motivation for doing it. Problem is that ICTs have created an enormous potential for procrastination – for doing less urgent but pleasurable tasks in preference to more urgent but less pleasurable ones. The seductive power can be strong and yielding to temptation is surely positively correlated with sloth.

Solutions and Recommendations

The positive impact of above discussed opportunities can neutralize or outweigh the negative impact of above discussed challenges - threats. We cannot say for sure what happens because we know that all systems with positive feedbacks can considerably and dis-continually change very



quickly their behavior and that it is impossible to predict when the change comes and what the behavior of the system would be after the change. Similar situation has occurred with the spread of each new technology and some recommendations can be done based on historical experiences.

It would not have any sense to try to stop the development of ICTs, it is simply impossible. What is possible is to try to increase the costs of ICTs activities that are negative from the ethical point of view and to increase benefits of activities that are positive from the ethical point of view. If it ought to be done, how to do it and who ought to do it opens a lot of questions. Let us take as an example the possibility of censorship.

Censorship can decrease negative impact through forbidding access to pages that could create addiction. Problem is that it constrains freedom in the virtual reality that is the basic condition for using up the opportunities of ICTs. In each society, some equilibrium level of censorship has been created that reflexes different opinions on this issue. Both opinions and the equilibrium shift in time, we can imagine that in Western societies more www pages than child pornography could be blocked and in China some of the currently blocked www pages could be unblocked.

Previous example shows that discourse of ethical issues linked with ICTs can never stop and that all members of society have responsibility to take part in it.

FUTURE RESEARCH DIRECTIONS

Following research questions would deserve more attention:

- What is the impact of ICTs on social identity? Could possible changes in social identity have a geo-political impact?
- What other virtues and vices are influenced by ICTs and how?

- Is it possible to increase the costs of ICTs activities that are negative from the ethical point of view, ought it to be done and by whom?
- Is it possible to increase the benefits of ICTs activities that are positive from the ethical point of view, ought it to be done and by whom?

CONCLUSION

ICTs offer a lot of possibilities and present some challenges – threats - for living a good life in the ethical sense and being a good person. It is impossible to predict if the positive impacts outweigh the negative ones. Anyway, it is impossible to stop the development of ICTs. The question is if the society ought to try to increase the costs of ICTs activities that are negative from the ethical point of view and to increase benefits of activities that are positive from the ethical point of view, who ought to do it and how. All members of society have responsibility to participate in discourse of this question.

REFERENCES

- Anderson, C. (2010). *How web video powers global innovation*. Retrieved September 30, 2015 from <https://www.ted.com/>
- Arntzen, A. A. B. (2010). Knowledge, Technology and University Social Responsibility (USR): A Conceptual Framework. *Proceedings of the 11th European Conference on Knowledge Management*, 1-2, 74-81.
- Busch, T. (2011). Capabilities in, capabilities out: Overcoming digital divides by promoting corporate citizenship and fair ICT. *Ethics and Information Technology*, 13(4), 339–353. doi:10.1007/s10676-010-9261-3

Celen, F. K., & Seferoglu, S. S. (2013). Investigation of Elementary School Students' Opinions Related to Unethical Behavior in the use of Information and Communication Technologies. *2nd World Conference on Educational Technology Research*, 83, 417-421. doi:10.1016/j.sbspro.2013.06.082

Codes of Ethics Collection. (n.d.). Retrieved September 30, 2015 from <http://ethics.iit.edu/ecodes/introduction>

Lianos, M. (2000). Dangerization and the end of deviance - The institutional environment. *The British Journal of Criminology*, 40(2), 261-278. doi:10.1093/bjc/40.2.261

Light, B., & McGrath, K. (2010). Ethics and social networking sites: A disclosive analysis of Facebook. *Information Technology & People*, 23(4), 290-311. doi:10.1108/09593841011087770

List of Virtues and List of Vices. (n.d.). Retrieved September 30, 2015 from <http://www.virtue-science.com/vicelist.html>

Niemela, M., Ikonen, V., Leikas, J., Kantola, K., Kulju, M., Tammela, A., & Ylikauppila, M. (2014). Human-Driven Design: A Human-Driven Approach to the Design of Technology. *ICT and Society*, 431, 78-91.

Patrignani, N., & Whitehouse, D. (2014). Slow Tech: The Bridge between Computer Ethics and Business Ethics. In K. Kimppa, D. Whitehouse, T. Kuusela, & J. Phahlamohlaka (Eds.), *ICT and Society* (Vol. 431, pp. 92-106). doi:10.1007/978-3-662-44208-1_9

Rizk, N. J., & Choueiri, E. M., & Insticc. (2006). The concept of ethics in electronic qualitative research. *ICEIS 2006: Proceedings of the Eighth International Conference on Enterprise Information Systems: Information Systems Analysis and Specification*, 126-134.

Som, C., Hilty, L. M., & Kohler, A. R. (2009). The Precautionary Principle as a Framework for a Sustainable Information Society. *Journal of Business Ethics*, 85(S3), 493-505. doi:10.1007/s10551-009-0214-x

Stepien, T. (2010). Transcultural and transdisciplinary competences. Technological transformations of society and education. *3rd International Conference of Education, Research and Innovation (Iceri2010)*, 4815-4824.

Tsai, H. Y., & Chen, W. C. (2013). Exploring Intelligent Information Communication Services: A Case of Corporate Social Responsibility. In X. H. Liu, K. F. Zhang, & M. Z. Li (Eds.), *Manufacturing Process and Equipment*, Pts 1-4 (Vol. 694-697, pp. 3636-3641). doi:10.4028/www.scientific.net/AMR.694-697.3636

Vaccaro, A., & Madsen, P. (2009). Corporate dynamic transparency: The new ICT-driven ethics? *Ethics and Information Technology*, 11(2), 113-122. doi:10.1007/s10676-009-9190-1

Watts, S., & Wyner, G. (2011). Designing and theorizing the adoption of mobile technology-mediated ethical consumption tools. *Information Technology & People*, 24(3), 257-280. doi:10.1108/09593841111158374

KEY TERMS AND DEFINITIONS

Casuistry: Branch of ethics that uses case studies to make clear moral principles and norms in typical situations. All ethical codes specify some norms of conduct. These norms are usually explained with the use of different case studies.

Existential Philosophy: Branch of Philosophy that starts with individual existence and the problem of its being in the World. Important predecessor was Søren Kierkegaard (19th century), among

main representatives we can find Martin Heidegger and Jean-Paul Sartre (both 20th century). They looked for authentic existence and have found different solutions depending especially on the religiosity of the specific philosopher.

Ethics: Practical philosophy, scientific analysis of moral contents.

Ethics of Norms: Branch of ethics that analyses the structure of norms and what they are based on.

Norms: Norms are the shared and sanctioned rules. Sanctions can be both formal (e.g. law) and informal (e.g. pressure of social group).

Utilitarianism: Branch of ethics that bases moral reasoning on consequences of actions. Actions are good if they maximize pleasure and

minimize pain. Among grounders are Jeremy Bentham and John Stuart Mill (18th -19th century). It has penetrated the mainstream economics. It can be shown that total utility of consumers can be increased by redistribution.

Virtue Ethics: Branch of Ethics that considers humans as being bearers of given or developed virtues. Humans have to develop virtues and fight with vices for becoming good. Basic concepts go back as far as to Plato and Aristotle (5th, 4th century BCE).

Virtual Reality: Created and accessed by ICTs. It usually overlaps with “reality” but the overlapping is changing and fuzzy. Extremes of Virtual Reality are Virtual Worlds, in which “avatars” of humans can live parallel lives.

Intrusion Tolerance Techniques



Wenbing Zhao

Cleveland State University, USA

INTRODUCTION

Intrusion tolerance refers to the capability of maintaining the system availability and integrity despite malicious attacks. Intrusion tolerance has been a hot research area for more than a decade and various techniques have been introduced to achieve various degrees of intrusion tolerance (Castro & Liskov, 2002; Chai & Zhao, 2014 June; Chai & Zhao, 2014 August; Deswarte et al., 1991; Verissimo et al., 2003; Yin et al., 2003; Zhao, 2013; Zhao, 2014). Such techniques can tolerate intrusion attacks in two respects: (1) a system continues providing correct services (may be with reduced performance) and (2) no confidential information is revealed to an adversary. The former can be achieved by using the replication techniques, as long as the adversary can only compromise a small number of replicas. The later is often built on top of secret sharing and threshold cryptography techniques. Plain replication is often perceived to reduce the confidentiality of a system, because there are more identical copies available for penetration. However, if replication is integrated properly with secret sharing and threshold cryptography, both availability and confidentiality can be enhanced.

BACKGROUND

In this section, we introduce some basic security and dependability concepts and techniques related to intrusion tolerance. A secure information system is one that exhibits the following properties (Pfleeger & Pfleeger, 2002):

- **Confidentiality:** Only authorized users have access to the information.
- **Integrity:** The information can be modified only by authenticated users in authorized ways. Any unauthorized modification can be detected.
- **Availability:** The information is available whenever a legitimate user wants to access it.

Confidentiality is often achieved by using encryption, authentication, and access control. Encryption is a reversible process that scrambles a piece of plaintext into something uninterpretable. Encryption is often parameterized with a security key. To decrypt, the same or a different security key is needed. Authentication is the procedure to verify the identity of a user that wants to access confidential data. Access control is used to restrict what an authenticated user can access.

Integrity can be protected by using secure hash functions, message authentication code (MAC) and digital signatures. For data stored locally, including the application binary files, a checksum is often used as a way to verify data integrity. The checksum can be generated by applying an oneway secure hash transformation on the data. Before the data is accessed, one can verify its integrity by recomputing the checksum and comparing it with the original one. The integrity of a message transmitted over the network can be guarded by a MAC. A MAC is generated by hashing on both the original message and a shared secret key (and often with a sequence number as well). If it is tampered with, the message can be detected in a way similar to that for checksum. For stronger

protection, a message can be signed by the sender. A digital signature is produced by first hashing the message using a secure hash function, and then encrypting the hash using the sender's private key.

High availability is achieved by using replication, checkpointing and recovery techniques. Replication is a technique that relying on running redundant copies of an application so that if one copy fails, the services can be provided by the remaining copies. Checkpointing means to take a snapshot of the state of a replica. The saved state can be used to bring a new or a restarted replica up to date. Checkpointing is also useful to avoid log buildup (when a checkpoint is taken, all previously logs can be garbage collected). Recovery techniques concern the tasks of removing faulty replicas, repairing them, and reintegrating them back to the system.

INTRUSION TOLERANCE TECHNIQUES

Intrusion tolerance is built on top of two fundamental techniques: replication and secret sharing/threshold cryptography (Deswarte et al., 1991). In the context of intrusion tolerance, a very general fault model must be used because a compromised replica might exhibit arbitrary faulty behaviors. Such a fault model is often termed as Byzantine fault (Lamport et al., 1982).

Byzantine Fault Tolerance

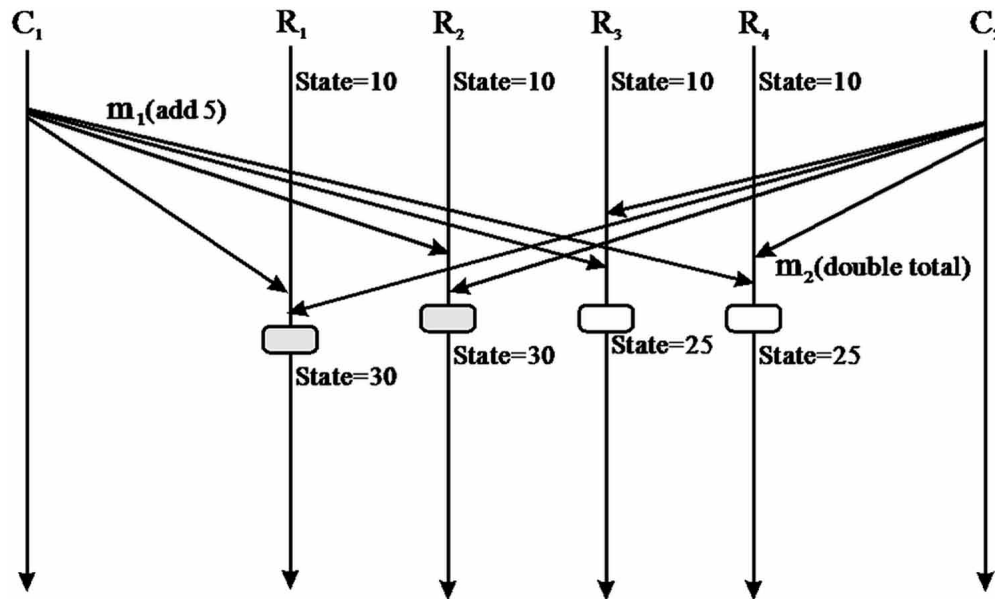
An intrusion attack might bring a service down, or compromise the integrity of a service. An effective defense is to introduce redundancy into the system, i.e., to replicate critical components in the system. Assuming that an intrusion attack can only penetrate a small fraction of the replicas, the service availability and integrity can be preserved by the remaining correct replicas. However, achieving this goal is not trivial – we must ensure consistent execution of all correct replicas despite the attacks launched by faulty replicas.

A Byzantine faulty replica may use all kinds of strategies to prevent the normal operations of a replica. In particular, it might propagate conflicting information to other replicas or components that it interacts with. To tolerate f Byzantine faulty replicas in an asynchronous environment, we need to have at least $3f+1$ number of replicas (Castro & Liskov, 2002). An asynchronous environment is one that has no bound on processing times, communication delays, and clock skews. Internet applications are often modeled as asynchronous systems. Usually, one replica is designated as the primary and the rest are backups.

There are two different approaches to Byzantine fault tolerance. In a Byzantine quorum system (Malkhi & Reiter, 1997), read and write operations issued by some clients are applied on a set of data items (which consists of the state of a service). It is assumed that the read and write operations are synchronized. A read operation retrieves information from a quorum of correct replicas and a write operation applies the update to a quorum of correct replicas. In a system with $3f+1$ replicas, a quorum can be formed by $2f+1$ replicas so that any two quorums overlap by at least $f+1$ replica, among which at least one is not faulty. This guarantees the correct operations of the quorum-based system.

A more general method is the state-machine based approach (Schneider, 1990). In the state-machine based approach, a replica is modeled as a state machine. The state change is triggered by remote invocations on the methods offered by the replica. This approach is applicable to a much wider range of applications. Consider a client server application where the server is replicated using the state-machine based approach (Castro & Liskov, 2002). The client first sends its request to the primary replica. The primary then broadcasts the request message to the backups and also determines the execution order of the message. To prevent a faulty primary from intentionally delaying a message, the client starts a timer after it sends out a request. It waits for $f+1$ identical replies from different replicas. Because at most f

Figure 1. An example scenario showing the importance of totally ordering incoming requests



replicas are faulty, at least one reply must come from a correct replica. If the timer expires before it receives a correct reply, the client broadcasts the request to all server replicas. This enables the correct replicas to detect the primary failure so that a new primary can be elected (this is often called a view change).

All correct replicas must agree on the same set of input messages with the same execution order. In other words, the request messages must be delivered to the replicas reliably in the same total order. To understand this better, consider the scenario illustrated in Figure 1. There are four server replicas R_1 , R_2 , R_3 , and R_4 , and two clients C_1 and C_2 . Client C_1 multicasts a request m_1 to all the replicas, and concurrently, the other client C_2 multicasts another request m_2 to all the replicas. Without controlling the ordering of the messages, it might happen that R_1 and R_2 deliver m_1 ahead of m_2 , and R_3 and R_4 deliver m_1 after m_2 . Assume that the initial state at each replica can be represented as an integer with a value of 10, the request m_1 asks for the addition of 5 to the state, and the request m_2 asks for the doubling of the state value. With the different ordering, state at R_1 and R_2 after processing m_1 and m_2 would be 30 while the state

at R_3 and R_4 after processing m_2 and m_1 would be 25. Hence, the states of the replicas diverge due to inconsistent request delivery ordering.

Furthermore, replicas may not process the same request deterministically, i.e., given the same request (delivered in the same order to all replicas), they may not produce identical reply. There are many factors that can cause non-deterministic execution of a request, e.g., differences in local clocks, process identifiers, and many other local resources that might be referenced by the replicas, and multithreading. These factors must be controlled properly so that the replicas appear to execute deterministically. This is a rather difficult task in the face of Byzantine fault. Until recently, the issue was addressed in a very limited manner, for example, by assuming that the replicas are single-threaded and all non-deterministic operations are known a priori. The primary replica determines the values to be used for all replicas and disseminates them to the backups. The backups subsequently verify the proposed values. If the primary is detected to be faulty, a view change is initiated. Recently, we introduced a framework that systematically handles various types of replica nondeterminism (Zhang et al., 2011).

Byzantine fault tolerance mechanisms tend to suffer from scalability problems. Amir et al. (2006) have tackled the size scalability problem by using a hierarchical replication architecture so that Byzantine fault tolerance can be achieved in large scale application running over wide area networks. Aiyer et al. (2005) considered the challenges faced by ensuring Byzantine fault tolerance over multi-administrative domains (administration scalability problems) and invented mechanisms to cope with selfish members, in addition to arbitrary faulty members.

Secret Sharing and Threshold Cryptography

Another aspect of intrusion tolerance is to protect confidential information (e.g., security keys) even if some replicas have been penetrated (Deswarte et al., 1991). If all the secrets are maintained by a single process, an adversary may obtain these secrets by compromising the process. To defend against such an attack, each secret is divided into multiple shares and each share is stored in a separate process. To obtain the secret, an adversary must now break into a significant number of processes. This is the basic objective of secret sharing.

A popular secret sharing scheme is the (k, n) threshold scheme, where n is the total number of shares and k is the minimum number of shares needed to reconstruct the secret. No useful information can be obtained as long as the number of shares collected is less than k , the threshold. This scheme is first proposed by Shamir (1979) and implemented using polynomial interpolation. In this scheme, each share is of similar size to the original secret and shares can be dynamically added or deleted.

The (k, n) threshold scheme is quite expensive computationally and space-wise. Therefore, they are used only to protect the most crucial secret such as security keys. It might not be practical

to apply them to file systems or databases. Consequently, a more cost effective scheme, called Fragmentation-Replication-Scattering (FRS), is proposed (Deswarte et al., 1991). It was initially designed to provide intrusion tolerance for file systems and was later extended to object-based systems (Fabre & Randell, 1992). The FRS scheme involves three steps. First, a file is partitioned into many smaller pieces. (To enhance the confidentiality, the file can be encrypted prior to the fragmentation step.) Second, each piece is replicated. Finally, the pieces are distributed pseudo-randomly according to some algorithm to the storage sites. The fragmentation and scattering protect data confidentiality against intrusion attacks because to obtain the file, an adversary must first find out the locations of the fragments belonging to the file and then penetrate all the sites that store the fragments. The replication protects the availability of the file so that even if some pieces are destroyed by an adversary, there are enough copies left to reconstruct the original file.

Despite the elegance of the secret sharing schemes, the secret must be reconstructed before it can be used. This poses a security threat because if the process that performs this task is compromised, the secret may be exposed. This prompted the development of threshold cryptography (Desmedt & Frankel, 1989). In threshold cryptography, security operations such as encryption, decryption, signature generation and verification can be performed by a group of processes without reconstructing the shared secret. Threshold cryptography has been applied primarily to public key based security services by sharing the private key among a group of processes (Zhou et al., 2002). The shares can be proactively updated to further enhance the security. By sharing the private key, each process can produce a partial signature. If a client obtains enough number of partial signatures, it can compute the complete signature. During this process, the private key is never reconstructed.

Protecting Both Availability and Confidentiality

Byzantine fault tolerance ensures service high availability, but it does not protect data confidentiality against intrusions. Threshold cryptography guarantees both confidentiality and high availability of some security operations, but not general services one might use. All secret sharing schemes require the reconstruction of the secret at a trusted site, which may be vulnerable to intrusion attacks.

COCA (Zhou et al., 2002) is the first attempt to integrate a Byzantine quorum system with threshold cryptography in the context of a certificate authority (CA) service. In COCA, the most critical state, i.e., the service private key (used to generate signatures for the certificates issued to clients) is shared among the replicas to prevent an adversary from stealing it. There are other states in a CA service, such as the certificate issued to the clients. A client can query and update the certificate information through the CA services. The high availability of the CA services is provided by the group of CA replicas. To enhance failure resiliency, the replicas are periodically restarted proactively with a correct binary image and the latest state collected from other correct replicas. The server keys (used to communicate securely among the replicas) are also periodically refreshed.

FUTURE TRENDS

Even though there is moderate success in applying intrusion tolerance techniques to a few specific applications, it remains to be seen how to introduce intrusion tolerance to other types of information systems, such as credit card processing systems, online banking systems, and e-commerce systems. The lack of effective solutions has left many mission critical systems vulnerable. This is evidenced by the increasingly frequent reports of high impact security breaches that resulted in massive disclosure of confidential information (Baker et al., 2011), such as the break-in of CardSystems

Solutions, which exposed the confidential data of more than 40 million credit card holders (Dash & Zelle, 2005), the Veteran Affairs incident, which has caused the potential loss of personal data of more than 26.5 million veterans to the hands of adversaries (Dunham, 2006), and recently the Sony PlayStation online network break, which led to exposure of personal data of over 77 million user accounts (Baker & Finkle, 2011).

There are many open issues to be resolved before we see widespread adoption of intrusion tolerance techniques. The most interesting research problem seems to be the design of a systematic methodology that allows direct operation on the shared secret that is dispersed among a number of processes. This will eliminate the vulnerability introduced by relying on a trusted process to reconstruct the secret. Another urgent issue is how to address replica non-determinism problem, especially that caused by multithreading. Practical information systems are very complicated and contain extensive non-deterministic operations. Unfortunately, all Byzantine fault tolerance strategies require deterministic execution of replicas. Even though the non-determinism resulted from a replica's accessing of some local resources (such as clock values and file descriptors) can be easily controlled if the access pattern is known in advance (Yin et al., 2003), there is no straightforward solution to render a multithreaded application deterministic without significantly impacting the application's performance and correctness (Zhao, 2005). Recently, we proposed a drastically different architecture to avoid replicating execution, which would eliminate the need for controlling the replica nondeterminism (Zhao, 2008).

Furthermore, intrusion tolerance favors design diversity to avoid common vulnerabilities among the replicas. N-version programming was proposed as a potential solution by Chen and Avizienis (1978). However, it might not be an economically viable solution due to its heavy software development cost. Alternative, less costly solutions are needed to diversify the



replica implementations. Code randomization appears to be a good approach to achieve replica diversity (Forrest, 1997).

Intrusion tolerance design does not conflict with the effort of intrusion prevention and intrusion detection (Verissimo et al., 2003). In fact, they are all indispensable techniques for building secure and dependable systems. Without applying good intrusion prevention techniques, the application would contain too many vulnerabilities for the replication strategy to take effect. Similarly, intrusion detection plays an essential role in intrusion tolerant systems. To remove a faulty replica and to recover it subsequently, the fault must be detected as quickly as possible. Intrusion detection also helps deter future intrusion attacks if there is enough information logged that can be used to prosecute the intruders (Dunlap et al., 2002).

CONCLUSION

We believe that the research and development of intrusion tolerant systems will gain more momentum as more and more services are offered online. The expectation of such services is high, considering their essential roles in everyday operations of businesses and individuals as well. The impact of service unavailability and security breaches will only be more serious. In this chapter, we have surveyed the state of the art techniques for building intrusion tolerant systems. We also illustrated a few most urgent open issues for future research. Finally, we pointed out that to build secure and dependable systems, we need a concerted effort in intrusion prevention, intrusion detection, and intrusion tolerance.

REFERENCES

- Aiyer, A., Alvisi, L., Clement, A., Dahlin, M., Martin, J., & Parth, C. (2005). BAR fault tolerance for cooperative services. *ACM Symposium on Operating Systems Principles*. Brighton, UK: ACM Press.
- Amir, A., Danilov, C., Dolev, D., Kirsch, J., Lane, J., Nita-Rotaru, C., & Zage, D. et al. (2006). Scaling Byzantine fault-tolerant replication to wide area networks. *Proceedings of International Conference on Dependable Systems and Networks*. Philadelphia, PA: IEEE Computer Society Press. doi:10.1109/DSN.2006.63
- Baker, L. B., & Finkle, J. (2011). *Sony PlayStation suffers massive data breach*. Reuters.
- Baker, W., Hutton, A., Hylender, C. D., Pamula, J., Porter, C., & Spitler, M. (2011). *2011 data breach investigations report*. Verizon RISK Team. Available: www.verizonbusiness.com/resources/reports/rp_databreach-investigations-report-2011_en_xg.pdf
- Castro, M., & Liskov, B. (2002). Practical Byzantine fault tolerance and proactive recovery. *ACM Transactions on Computer Systems*, 20(4), 398–461. doi:10.1145/571637.571640
- Chai, H., & Zhao, W. (2014, June). Towards trustworthy complex event processing. In *Proceedings of the 5th IEEE International Conference on Software Engineering and Service Science* (pp. 758-761). IEEE.
- Chai, H., & Zhao, W. (2014, August). Byzantine Fault Tolerant Event Stream Processing for Autonomic Computing. In *Proceedings of the 12th IEEE International Conference on Dependable, Autonomic and Secure Computing* (pp. 109-114). IEEE. doi:10.1109/DASC.2014.28

- Chen, L., & Avizienis, A. (1978). N-version programming: A fault-tolerance approach to reliability of software operation. *Proceedings of International Symposium on Fault Tolerant Computing*. Toulouse, France: IEEE Computer Society Press.
- Dash, E., & Zeller, T. (2005, June 18). MasterCard says 40 million files are put at risk. *New York Times*.
- Desmedt, Y., & Frankel, Y. (1990). Threshold cryptosystems. *Lecture Notes in Computer Science*, 435, 307–315. doi:10.1007/0-387-34805-0_28
- Deswarte, Y., Blain, L., & Fabre, J. (1991). Intrusion tolerance in distributed computing systems. *Proceedings of the IEEE Symposium on Research in Security and Privacy*. Oakland, CA: IEEE Computer Society Press.
- Dunham, W. (2006, May 22). Personal data on millions of U.S. veterans stolen. *Computerworld*.
- Dunlap, W., King, S., Cinar, S., Barsai, M., & Chen, P. (2002). ReVirt: Enabling intrusion analysis through virtual-machine logging and replay. *Proceedings of the Symposium on Operating Systems Design and Implementation*, 211–224. doi:10.1145/1060289.1060309
- Fabre, J., & Randell, B. (1992). An object-oriented view of fragmented data processing for fault and intrusion tolerance in distributed systems. *Lecture Notes in Computer Science*, 648, 193–208. doi:10.1007/BFb0013899
- Forrest, S., Somayaji, A., & Ackley, D. (1997). Building diverse computer systems. *Proceedings of the Workshop on Hot Topics in Operating Systems*. Cape Cod, MA: IEEE Computer Society Press.
- Lamport, L., Shostak, R., & Pease, M. (1982). The Byzantine generals problem. *ACM Transactions on Programming Languages and Systems*, 4(3), 382–401. doi:10.1145/357172.357176
- Malkhi, D., & Reiter, M. (1997). Byzantine quorum systems. *Proceedings of the ACM Symposium on Theory of Computing*. El Paso, TX: ACM Press.
- Pfleeger, C., & Pfleeger, S. (2002). *Security in Computing* (3rd ed.). Prentice Hall.
- Schneider, F. (1990). Implementing fault-tolerant services using the state machine approach: A tutorial. *ACM Computing Surveys*, 22(4), 299–319. doi:10.1145/98163.98167
- Shamir, A. (1979). How to share a secret. *Communications of the ACM*, 22(11), 612–613. doi:10.1145/359168.359176
- Verissimo, P., Neves, N., & Correia, M. (2003). Intrusion-tolerant architectures: Concepts and design. *Lecture Notes in Computer Science*, 2677, 90–109. doi:10.1007/3-540-45177-3_1
- Yin, J., Martin, J., Venkataramani, A., Alvisi, L., & Dahlin, M. (2003). Separating agreement from execution for Byzantine fault tolerant services. *Proceedings of the ACM Symposium on Operating Systems Principles*. Bolton Landing, NY: ACM Press. doi:10.1145/945445.945470
- Zhang, H., Zhao, W., Moser, L. E., & Melliar-Smith, P. M. (2011). Design and implementation of a Byzantine fault tolerance framework for non-deterministic applications. *Software, IET*, 5(3), 342–356. doi:10.1049/iet-sen.2010.0013
- Zhao, W. (2008, May). Towards practical intrusion tolerant systems: a blueprint. In *Proceedings of the 4th Annual Workshop on Cyber security and information intelligence research: developing strategies to meet the cyber security and information intelligence challenges ahead* (p. 19). ACM. doi:10.1145/1413140.1413162
- Zhao, W. (2013, April). Towards practical intrusion tolerant systems. In *Proceedings of the IET International Conference on Information and Communications Technologies* (pp. 280–287). IET.

Zhao, W. (2014, August). Application-Aware Byzantine Fault Tolerance. In *Proceedings of the 12th IEEE International Conference on Dependable, Autonomic and Secure Computing* (pp. 45-50). IEEE. doi:10.1109/DASC.2014.17

Zhao, W., Moser, L., & Melliar-Smith, P. (2005). Deterministic scheduling for multithreaded replicas. *Proceedings of the IEEE International Workshop on Object-oriented Real-time Dependable Systems*. Sedona, AZ: IEEE Computer Society Press. doi:10.1109/WORDS.2005.26

Zhou, L., Schneider, F., & van Renesse, R. (2002). COCA: A secure distributed online certification authority. *ACM Transactions on Computer Systems*, 20(4), 329–368. doi:10.1145/571637.571638

ADDITIONAL READING

Atighetchi, M., Pal, P., Webber, F., Schantz, R., Jones, C., & Loyall, J. (2004). Adaptive cyberdefense for survival and intrusion tolerance. *IEEE Internet Computing*, 8(6), 25–33. doi:10.1109/MIC.2004.54

Bessani, A. N., Sousa, P., Correia, M., Neves, N. F., & Verissimo, P. (2007). Intrusion-tolerant protection for critical infrastructures.

Bessani, A. N., Sousa, P., Correia, M., Neves, N. F., & Verissimo, P. (2008). The CRUTIAL way of critical infrastructure protection. *Security & Privacy, IEEE*, 6(6), 44–51. doi:10.1109/MSP.2008.158

Chai, H., & Zhao, W. (2012). Byzantine Fault Tolerance as a Service. In *Computer Applications for Web, Human Computer Interaction, Signal and Image Processing, and Pattern Recognition* (pp. 173-179). Springer Berlin Heidelberg. doi:10.1007/978-3-642-35270-6_24

Deswarte, Y., & Powell, D. (2006). Internet security: An intrusion-tolerance approach. *Proceedings of the IEEE*, 94(2), 432–441. doi:10.1109/JPROC.2005.862320

Feng, D., & Xiang, J. (2005, July). Experiences on intrusion tolerance distributed systems. In *Proceedings of the 29th Annual International Computer Software and Applications Conference* (Vol. 1, pp. 270-271). IEEE.

Garcia, M., Bessani, A., Gashi, I., Neves, N., & Obelheiro, R. (2011, June). OS diversity for intrusion tolerance: Myth or reality? In *Proceedings of the IEEE/IFIP 41st International Conference on Dependable Systems & Networks* (pp. 383-394). IEEE.

Kim, J., Bentley, P. J., Aickelin, U., Greensmith, J., Tedesco, G., & Twycross, J. (2007). Immune system approaches to intrusion detection—a review. *Natural Computing*, 6(4), 413–466. doi:10.1007/s11047-006-9026-4

Liu, P. (2002). Architectures for intrusion tolerant database systems. In *Proceedings of the 18th Annual Computer Security Applications Conference* (pp. 311-320). IEEE.

Min, B. J., & Choi, J. S. (2004). An approach to intrusion tolerance for mission-critical services using adaptability and diverse replication. *Future Generation Computer Systems*, 20(2), 303–313. doi:10.1016/S0167-739X(03)00146-8

Pal, P., Webber, F., Schantz, R. E., & Loyall, J. P. (2000, October). Intrusion tolerant systems. In *Proceedings of the IEEE Information Survivability Workshop* (pp. 24-26).

Ramasamy, H. V., Pandey, P., Cukier, M., & Sanders, W. H. (2008). Experiences with building an intrusion-tolerant group communication system. *Software, Practice & Experience*, 38(6), 639–666. doi:10.1002/spe.848

Sousa, P., Bessani, A. N., Correia, M., Neves, N. F., & Verissimo, P. (2007, December). Resilient intrusion tolerance through proactive and reactive recovery. In *Dependable Computing, 2007. PRDC 2007. 13th Pacific Rim International Symposium on* (pp. 373-380). IEEE. doi:10.1109/PRDC.2007.52

Sousa, P., Bessani, A. N., Dantas, W. S., Souto, F., Correia, M., & Neves, N. F. (2009, June). Intrusion-tolerant self-healing devices for critical infrastructure protection. In *Dependable Systems & Networks, 2009. DSN'09. IEEE/IFIP International Conference on* (pp. 217-222). IEEE. doi:10.1109/DSN.2009.5270333

Verissimo, P., Correia, M., Neves, N. F., & Sousa, P. (2009). Intrusion-resilient middleware design and validation. *Information Assurance. Security and Privacy Services*, 4, 615–678.

Wang, F., Uppalli, R., & Killian, C. (2003, October). Analysis of techniques for building intrusion tolerant server systems. In *Proceedings of the IEEE Military Communications Conference* (Vol. 2, pp. 729-734). IEEE.

Ye, N. (2001). Robust intrusion tolerance in information systems. *Information Management & Computer Security*, 9(1), 38–43. doi:10.1108/09685220110366786

Young, M., Kate, A., Goldberg, I., & Karsten, M. (2010, June). Practical robust communication in DHTs tolerating a byzantine adversary. In *Proceedings of the 2010 IEEE 30th International Conference on Distributed Computing Systems* (pp. 263-272). IEEE. doi:10.1109/ICDCS.2010.31

Young, M., Kate, A., Goldberg, I., & Karsten, M. (2013). Towards Practical Communication in Byzantine-Resistant DHTs. *IEEE/ACM Transactions on Networking*, 21(1), 190–203. doi:10.1109/TNET.2012.2195729

Zhao, W. (2007, October). BFT-WS: A Byzantine fault tolerance framework for web services. In *Proceedings of the Eleventh International IEEE EDOC Conference Workshop* (pp. 89-96). IEEE. doi:10.1109/EDOCW.2007.6

Zhao, W. (2007) A Byzantine fault tolerant coordination for Web services atomic transactions. In *Proceedings of the 3rd IEEE International Symposium on Dependable, Autonomic and Secure Computing* (pp. 37-44). Columbia, MD. doi:10.1109/DASC.2007.10

Zhao, W. (2009). Design and implementation of a Byzantine fault tolerance framework for Web services. *Journal of Systems and Software*, 82(6), 1004–1015. doi:10.1016/j.jss.2008.12.037

Zhao, W. (2014). *Building dependable distributed systems*. John Wiley & Sons. doi:10.1002/9781118912744

Zhao, W. (2015). Optimistic Byzantine fault tolerance. *International Journal of Parallel, Emergent and Distributed Systems*, 1-14 (preprint).

KEY TERMS AND DEFINITIONS

Byzantine Fault Tolerance: A replication-based technique used to ensure high availability of an application subject to Byzantine fault.

Byzantine Fault: It is used to model arbitrary fault. A Byzantine faulty process might send conflicting information to other processes to prevent them from reaching an agreement.

Byzantine Quorum System: The system offers read and write services to its clients on a set of replicated data items. A read operation retrieves data from a quorum of correct replicas and a write operation applies the update to a quorum of correct replicas. Any two quorums must overlap by at least one correct replica.

Fragmentation Redundancy Scattering: A secret sharing scheme that involves the following three steps: fragmenting a file, replicating each fragment, and distributing the replicated fragments to different storage sites.

Intrusion Tolerance: It refers to the capability of maintaining the system availability and integrity despite malicious attacks.

(k, n) Threshold Scheme: A secret is divided into n shares. To reconstruct the secret, at least k shares are needed. No useful information can be obtained from k-1 shares.

Replica Consistency: The states of the replicas of an application should remain to be identical at the end of the processing of each request. Replica consistency is necessary to mask a fault in some replicas.

Threshold Cryptography: Security operations such as encryption, decryption, signature generation and verification can be performed by a group of processes without reconstructing the shared secret. Threshold cryptography utilizes (k, n) threshold schemes internally.

New Perspectives of Pattern Recognition for Automatic Credit Card Fraud Detection



Addisson Salazar

Universitat Politècnica de València, Spain

Gonzalo Safont

Universitat Politècnica de València, Spain

Alberto Rodriguez

Universidad Miguel Hernández de Elche, Spain

Luis Vergara

Universitat Politècnica de València, Spain

INTRODUCTION

The automatic detection of frauds in financial operations using credit cards is a challenge issue that has been increasingly studied. The rapid expansion of information and communication technologies has expanded the potential to emulate legitimate operations by fraudsters. The solution to that problem has to be able to be adaptive since the behavior of frauds is changing constantly in time; to handle the detection in data with a very small ratio of fraud amount to legitimate operations, e.g., $5e-5$; and accomplish operation requirements of very low false alarm ratios in real-time processing. Thus, several approaches have been proposed from pattern recognition and machine learning areas.

Main issues related with the problem of automatic credit card fraud detection (ACCFD) and proposed solutions are discussed from theoretical and practical standpoints. The perspective of detection analyses from receiving operating characteristic (ROC) curves and business key performance indicators (KPI) are jointly analyzed (Girgenti & Hedley, 2011) (Wells, 2011) (Montague, 2010). Therefore, a new conceptual framework for ACCFD considering modern techniques such as decision fusion and surrogate data is outlined. There are only a few references from the research field

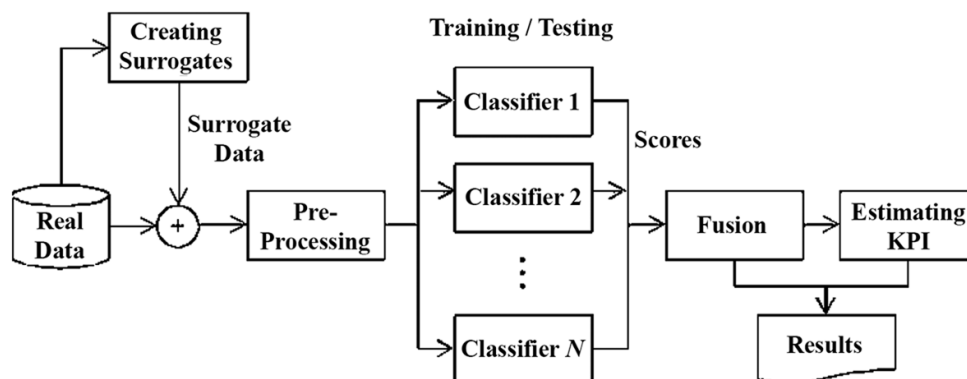
of signal processing for ACCFD, see for instance (Salazar, Safont, Soriano, & Vergara, 2012).

A case of study that combines different proportions of real and surrogate data is included. Several scenarios considering different single and combined methods are considered. ROC and KPI curves are analyzed bearing in mind numeric and operational requirements. The sensitivity of the methods to different proportions of fraud/legitimate ratios is tested. Thus, limitations and advantages of the studied methods are demonstrated.

BACKGROUND

Cyber-security and privacy have become very important subjects of research in recent years. This research spans many different fields, such as: security in the physical layer of wireless communications (Poor, 2012)); database security (Sankar, Rajagopalan, & Poor, 2013); distributed systems (Pawar, El Rouayheb, & Ramchandran, 2011); and biometrics (Lifeng, Ho, & Poor, 2011). One activity where the security and privacy mechanisms are critical is the e-commerce by using credit cards. This application features a massive volume of on-line transactions that are continuously exposed to frauds. Fraud detection

Figure 1. Outline of the signal processing procedure



in credit card transactions is a critical problem affecting large financial companies and involving annually loss of billions of dollars (Bhattacharyya, Jha, Tharakunnel, & Westland, 2011).

Basically two strategies can be raised. The first consists of defining the problem as one-class classification, and thus, characterizing the largest data population (the legitimate transactions) and considering all the data with different characteristics as outliers (Hodge & Austin, 2004) (Tax & Duin, 2001). The second strategy is to define the problem as a two-class classification characterizing legitimate and fraudulent transaction data. We have concentrated in this later detection approach which takes full advantage of the available labeled data.

There is extensive literature that reviews and provides taxonomies and comparisons about the large number of ACCFD methods that have been developed during the last two decades (e.g., (Danas, 2015)). However, only few of these references are from the research field of signal processing. The particular characteristics of ACCFD make this a challenging problem for signal processing algorithms (Salazar, Safont, Soriano, & Vergara, 2012). Optimum design of the algorithms depends on the detection models employed to estimate the multidimensional joint distribution of the random variables underlying the data.

Figure 1 shows an outline of the proposed signal processing procedure. The multivariate surrogate data is obtained following the methods explained in (Salazar, Safont, & Vergara, Surrogate techniques for testing fraud detection algorithms in credit card operations, 2014). The pre-processing step consists of applying principal component analysis (PCA) to reduce dimensionality of the data preserving 95% of data variance.

SURROGATE DATA

Bank enterprises collect large amount of historical records corresponding to millions of credit cards operations, but, unfortunately, only a small portion, if any, is open access. This is because, e.g., the records include confidential customer data and banks are afraid of public quantitative evidence of existing fraud operations (Bhattacharyya, Jha, Tharakunnel, & Westland, 2011). A solution is to generate synthetic records which replicate as much as possible the behavior of the real data. Surrogate techniques give an approach to this problem. Surrogates algorithms have been extensively used to detect the possible presence of non-linearities in a given time series realization. Basically, surrogate replicas of the original data are generated trying to preserve the correlation (second-order statistic) and amplitude distribution (first-order

statistic). These replicas are generated in such a form that linearity applies. Then some statistics for detecting possible presence of non linearities are computed from both the surrogate replicas and the original data. If significant differences are given, then existence of non-linearity is decided (Prichard & Theiler, 1994).

The main idea of surrogate techniques is to synthesize new stationary data by randomization in the Fourier domain. For multivariate series, cross-correlation among the variables should be conserved. The randomization in the Fourier domain is chosen so that the differences of phase between components stay the same. Let $X(n)$ be a multivariate series with M components ($x_j(n)$ with $n=0, \dots, N-1$ being its j -th component). For initialization, the Fourier transform of each component can be computed:

$$(Fx_j)(f) = \sum_{n=0}^{N-1} x_j(n) e^{-i2\pi \frac{nf}{N}} = Ax_j(f) e^{i\psi_{x_j}(f)} \quad (1)$$

A classical multivariate surrogate can be computed using the following expression:

$$S(n) = [s_1(n), \dots, s_M(n)]^T, \quad (2)$$

$$s_j(n) = \frac{1}{N} \sum_{f=0}^{N-1} Ax_j(f) e^{i(\psi_{x_j}(f) + \Theta(f))} e^{-i2\pi \frac{nf}{N}},$$

where $\Theta(f)$ is a random phase i.i.d. uniform in $[0, 2\pi]$.

However the algorithm of (1) does not impose any constraint to the surrogates that should be restricted to have the same marginal distributions and covariance as the original series (variables). The iterative IAAFT (Iteratively Amplitude Adjusted Fourier Transform) algorithm that projects on the two constraints (the covariance function expressed in the Fourier domain and the prescribed marginal distributions was proposed in (Schreiber & Schmitz, 2000). We applied a new extension of IAAFT algorithm considered in (Borgnat, Abry, & Flandrin, 2012) using a reduced set of real credit

card data to generate surrogate multivariate series of credit card data transactions.

SCORES AND LIKELIHOOD RATIO: A SIGNAL PROCESSING PERSPECTIVE FOR ACCFD

Detection of credit card frauds may be approached from a signal processing perspective. It is comparable to the classical problem of detecting signals in a noise background. The “signal presence” hypothesis (H_1) would be equivalent to “fraud presence”, meanwhile “absence of signal” hypothesis (H_0) would correspond to “absence of fraud”. In both cases hypothesis H_1 has much lower a priori probability than hypothesis H_0 , hence Neyman-Pearson criterion is the most appropriate. Implementation of a Neyman-Pearson

detector is made by the test $\Lambda(\mathbf{x}) = \frac{f(\mathbf{x} / H_1)}{f(\mathbf{x} / H_0)} \underset{H_0}{\overset{H_1}{>}} \lambda$,

where $f(\mathbf{x} / H_i)$ is the probability density of the observation vector $\mathbf{x}$ conditioned to hypothesis H_i , and λ is a threshold selected to fit an acceptable probability of false alarm (PFA). $\Lambda(\mathbf{x})$ is the likelihood ratio (LR) and its computation is the essential signal processing required to implement the optimum detector. Determination of LR mainly depends on the assumed models about the signal and noise. Different optimum detectors may be designed assuming multivariate Gaussian noise and different amount of knowledge about the signal, e.g., matched filter, subspace matched filter, and energy detector.

Similarly, the usual test in fraud detection is of the form $s(\mathbf{x}) \underset{H_0}{\overset{H_1}{>}} t$, where $s(\mathbf{x})$ is a score or

qualification between 0 and 1, assigned to the transaction under analysis. The higher the value $s(\mathbf{x})$, the higher the probability of a fraudulent transaction. The threshold t filters the transactions that would be passed to the human operator. Actually, the daily number of false positive frauds



which the human operators can manage imposes a limit in the PFA in much the same manner that it happens in other areas like radar signal detection. For those filtered transaction, the score $s(\mathbf{x})$ is part of the information taken into account by the operator to arrive to a final decision about authorization.

A natural candidate for $s(\mathbf{x})$ is $P(H_1/\mathbf{x})$, the “a posteriori” probability of having a fraud conditioned to the observation. It is straightforward to show the relation between $s(\mathbf{x})$ and $\Lambda(\mathbf{x})$ by using the Bayes theorem:

$$\begin{aligned} s(\mathbf{x}) &= P(H_1 / \mathbf{x}) = \frac{f(\mathbf{x} / H_1)P(H_1)}{f(\mathbf{x})} = \\ &= \frac{f(\mathbf{x} / H_1)P(H_1)}{f(\mathbf{x} / H_1)P(H_1) + f(\mathbf{x} / H_0)P(H_0)} = \\ &= \frac{\Lambda(\mathbf{x})}{\Lambda(\mathbf{x}) + \frac{P(H_0)}{P(H_1)}} \Leftrightarrow \Lambda(\mathbf{x}) = \frac{s(\mathbf{x}) \frac{P(H_0)}{P(H_1)}}{1 - s(\mathbf{x})} \end{aligned} \quad (3)$$

So we see the LR is implicitly considered given a scoring function, and, vice versa, a score value may be computed from the LR. In consequence, a test based in comparing the score (computed as the a posteriori probability of H_1) with a threshold is an LR test:

$$s(\mathbf{x}) \underset{H_0}{\overset{H_1}{>}} t \quad \Leftrightarrow \quad \Lambda(\mathbf{x}) \underset{H_0}{\overset{H_1}{>}} \frac{t}{1-t} \frac{P(H_0)}{P(H_1)} \quad (4)$$

Notice that the observation vector $\mathbf{x}$ will be formed by a set of heterogeneous features such as money amount, commerce code, country code, operation date and so on, that are extracted directly from the records. Besides, taking into account that the behavior of fraudsters normally emulates regular customer behavior in directly measured features; additional indirect features can be estimated to highlight differences (e.g., transaction velocity, transaction acceleration,

and country commuting). Thus, it is not easy to find parametric models for the statistical distributions of the observations under fraud and no fraud conditions. Hence, deriving optimum sufficient statistics or implementing generalized likelihood ratio tests (GLRT) are not practical options. However, it is usual that sets of training observations for both hypotheses are available, so that supervised learning of the LR or of a score function is possible. This is an advantage in comparison with signal detection, where sets of the expected signals are difficult to measure or to generate: it is difficult to learn $f(\mathbf{x} / H_1)$ from observation samples.

Other differences between these conceptually equivalent problems are in the type of performance measures. In signal detection the Receiving Operating Characteristic (ROC) curves (probability of detection, PD, as function of probability of false alarm, PFA) are the most usual. In addition, the dependence of PD on the signal-to-noise-ratio (SNR) for a given PFA, is very familiar. SNR is not a practical concept in fraud detection, as the presence of a fraud modifies the distribution of $\mathbf{x}$ in an arbitrary and unpredictable form rather than a simple change in energy. Moreover, PD is a relative measure of good performance as the final goal is to save money, i.e., it is better to detect one fraud involving a lot of money than many frauds of small amounts of money. Thus the Value Detection Rate (VDR) defined as the percentage of saved money is usually preferred. In addition, the Account False to Positive Rate (AFPR) defined in (5) is considered as a sample estimate of the average number of useless interventions in non-fraudulent operations produced for every detected fraud.

$$\begin{aligned} AFPR &= \frac{\text{True positives} + \text{False positives}}{\text{True positives}} = \\ &= 1 + \frac{\text{False positives}}{\text{True positives}} \end{aligned} \quad (5)$$

In conclusion, there are clear similarities between signal and fraud detection, but some differences exist, positioning the fraud problem somewhere in between signal processing and pattern recognition areas.

FUSION OF SCORES

Fraud detection is a difficult problem due to the arbitrary and variant nature of the statistical models under both hypotheses. This imposes the need of adaptive supervised learning of reasonable (probably non-optimum) detectors from datasets of millions of labeled transactions. The extremely large amount of available legitimate operations in comparison with the moderate number of confirmed frauds introduces an additional difficulty for a balanced learning of the models under both hypotheses. All this suggests that approaches based on deriving one “sophisticated” detector, having many degrees of freedom and/or parameters to be dynamically and continuously learned, will not be an appropriate approach. Instead, the idea of combining or fusing complementary robust detectors is tempting (some formal ideas about this intuitive concept emanates from the so-called evidence theory (Shafer, 1976)). Fusion of detectors has been considered since long time ago in different areas and under different terminologies: multimodal fusion (Atrey, Hossain, El Saddik, & Kankanhalli, 2010), mixture of experts (Yuksel, Wilson, & Gader, 2012) and classifier combiners (Kittler, Hatef, Duin, & Matas, 1998), to mention a few, but there is convenient to focus into the fraud detection context.

Let us assume that N different detectors are designed to detect possible frauds. Different designs may be the result of using different training sets on the same detector structure, of using different detector structures or both. In any case, every detector i produces a score s_i and we must find a fusion function to compute a final score s_f . Then, finding the optimum fusion function is

a design objective. This can be approached in a similar manner to the original problem of determining one score from the observation vector $\mathbf{x}$. We may consider that the vector of individual scores $\mathbf{s} = [s_1 \dots s_N]^T$ is a new observation vector, and then s_f may be computed from the LR of scores

$$s_f(\mathbf{s}) = \frac{\Lambda(\mathbf{s})}{\Lambda(\mathbf{s}) + \frac{P(H_0)}{P(H_1)}} \quad \Lambda(\mathbf{s}) = \frac{f(\mathbf{s} / H_1)}{f(\mathbf{s} / H_0)} \quad (6)$$

So, ultimately, the optimum fusion function depends on the joint probability densities functions of the scores under both hypotheses, $f(\mathbf{s} / H_1)$ and $f(\mathbf{s} / H_0)$. If we assume that the scores are independent and identically distributed random variables then (6) can be expressed as

$$s_f(\mathbf{s}) = \frac{\frac{f(s_1 / H_1) \dots f(s_N / H_1)}{f(s_1 / H_0) \dots f(s_N / H_0)}}{\frac{f(s_1 / H_1) \dots f(s_N / H_1)}{f(s_1 / H_0) \dots f(s_N / H_0)} + \frac{P(H_0)}{P(H_1)}} \quad (7)$$

In that case non-parametric estimates of the marginals can be obtained from training set of scores and introduced in the computation of the final score (plug-in methods, (Jain A. K., 2000)). But the fusion problem becomes more complicated if statistical independence among the individual scores cannot be assumed (Vergara, Soriano, Safont, & Salazar, 2016). Actually, presence of dependence can be expected if all the detectors share the same input $\mathbf{x}$. Unless Gaussianity could be assumed, capturing the dependence of multivariate random variables is not an easy problem. Gaussian and Non-Gaussian Mixtures Models are possible options (Tzikas & Likas, 2008) (Salazar, Vergara, Serrano, & Igual, 2010), they require the estimation of many parameters and the determination of the best number of mixture components. Another alterna-

tive is based on the use of copulas. Copula methodology has been used during years in econometrics, although it is only recently considered in the signal processing community (Iyengar, Varshney, & Damarla, 2011). The multivariate probability density function is factorized in the marginals and a copula function which captures dependence. The copula function is the multidimensional probability density function of uniform variables derived from the original dependent variables. A family of parametric functions exists having the required copula properties, so that a variety of dependence models are possible. Another option is α -integration (Soriano, Vergara, Ahmed, & Salazar, 2015), which is expressed as

$$s_f(\mathbf{s}) = \begin{cases} \left(\sum_{i=1}^N w_i \cdot s_i^{\frac{1-\alpha}{2}} \right)^{\frac{2}{1-\alpha}}, & \alpha \neq 1 \\ \exp \left(\sum_{i=1}^N w_i \cdot \log(s_i) \right), & \alpha = 1 \end{cases} \quad (8)$$

First, every score is modified by a nonlinear function controlled by the α parameter. Then, a weighted mean of the modified scores is computed, and the nonlinear inverse function is finally applied to the weighted mean. The α parameter and the weights w_i $i=1, \dots, N$ can be estimated so that a predefined cost function is minimized. One interesting aspect of α -integration is that some simple cases are included for particular values of α and w_i . Thus, assuming that $w_i = 1/N$, we see that $\alpha = -1, 1, 3$ respectively renders the arithmetic mean, the geometric mean and the harmonic mean. Similarly, $\alpha = \infty / -\infty$ is equivalent to compute the minimum/maximum.

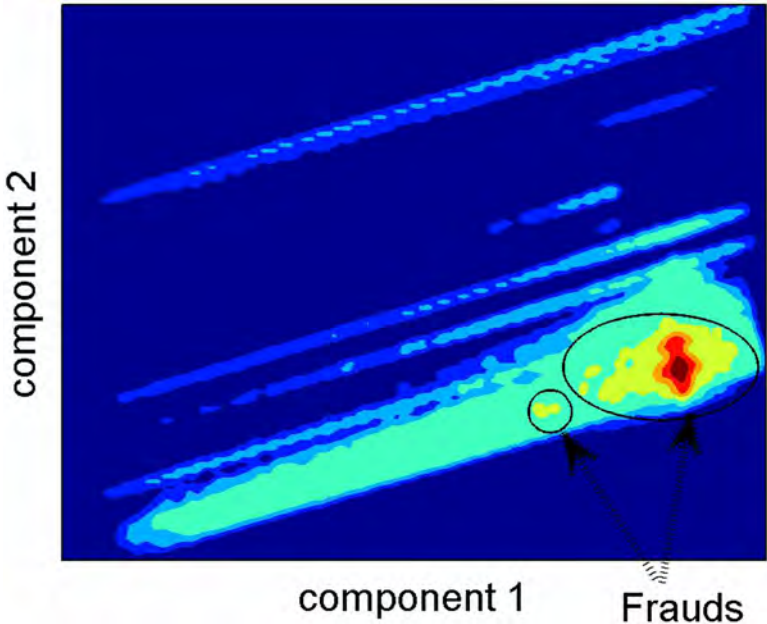
The proposed signal processing procedure was applied using a set of real data provided by a large financial company. These data were combined with surrogate data following the methods described in (Salazar, Safont, & Vergara, Surrogate techniques for testing fraud detection algorithms in credit card operations, 2014). Thus, a total dataset of 37.000.000 transactions were available (90%

surrogate). These dataset correspond to about 3 million credit cards with a total of 2005 fraudulent transactions from 682 credit cards. The proportion between fraudulent and legitimate transactions is different from the real proportion, but it is small enough for evaluating the methods. This dataset was divided 5% for training stage and 95% for testing stage. Figure 2 shows the distribution of the data in the first two components space estimated by PCA. It can be seen that the fraud data are crowded together the genuine data making difficult the separation of the two kinds of data. Although the genuine data have been sub-sampled 100 times, it can be seen the high disproportion between the two kinds of data.

Seven different representative detectors were selected. Three of them are individual detectors selected from generally applied methods: Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA) and NonGaussian Mixture Models (NGM) (Salazar, On Statistical Pattern Recognition in Independent Component Analysis Mixture Modelling, 2013). Two more were considered because they have been previously used in the ACCFD problem: Random Forest (RF) and Support Vector Machines (SVM). Finally two fusion methods were applied. The first is α -integration (α -INT). The second fusion method was based on the use of a Gaussian copula (COP). Several experiments show that SVM and RF did not contribute to improve fused results, so the final methods included in the fusion step were LDA, QDA and NGM.

Figure 3 shows the ROC curves corresponding to transactions summarized by cards for the seven methods. Actually, the detectors are applied to every transaction, but once a transaction is detected as fraudulent the associated card is considered fraudulent in the observed period of analysis. Thus ROC curves of transactions may differ from ROC curves of cards, being these latter more significant in the ACCFD context. In Figure 3 we only represent the acceptable margins for the PFA [0, 0.04]. We can see that PDs under 0.5 are obtained; this can be unacceptable in other

Figure 2. Scatter plot of the data in two component space. Localization of fraud data is highlighted.



applications but are very significant in ACCFD as this can potentially lead to saving big amount of money. Finally, we deduce from Figure 3 the following main conclusions. First, LDA is clearly below the rest of methods, thus suggesting that Gaussian models are not appropriate in this case. Second, fusion methods improve the individual detectors, especially for the lowest values of PFA.

KPI CURVES

KPI are estimated after processing in order to evaluate the results from the business operational perspective. They provide different information than the ROC curves because they take into account not only the transactions themselves, but also variables such as grouping of transactions using card number; the amount of each transaction; and the period of time until the fraud is detected (VDR and AFPR were defined above):

- **ADR:** Account Detection Rate, the percentage of detected fraudulent cards.

- **ADT:** Average Detection Time, the average amount of fraudulent transactions on the same card required for detecting its fraudulent use.

Figure 4 shows the results obtained with KPI to be compared with Figure 3. It represents VDR as function of the normalized number of daily alerts (number of daily alerts divided by the total number of processed cards in a day). In general, results are similar but a significant difference is present, in this case, α -INT performs better than COP when VDR instead of PD is computed. There are several discussions that can be made from KPI results, but they are outside the scope of this work.

SURROGATE AND FRAUD/LEGITIMATE RATIOS

Figures 2 and 3 show results of the surrogate algorithm for legitimate and fraud operations, respectively. Figure 1a shows 500 samples of the surrogate data for the first eight components. Figure 1b shows the histogram of the eight com-

Figure 3. ROC estimated from the results of detection methods

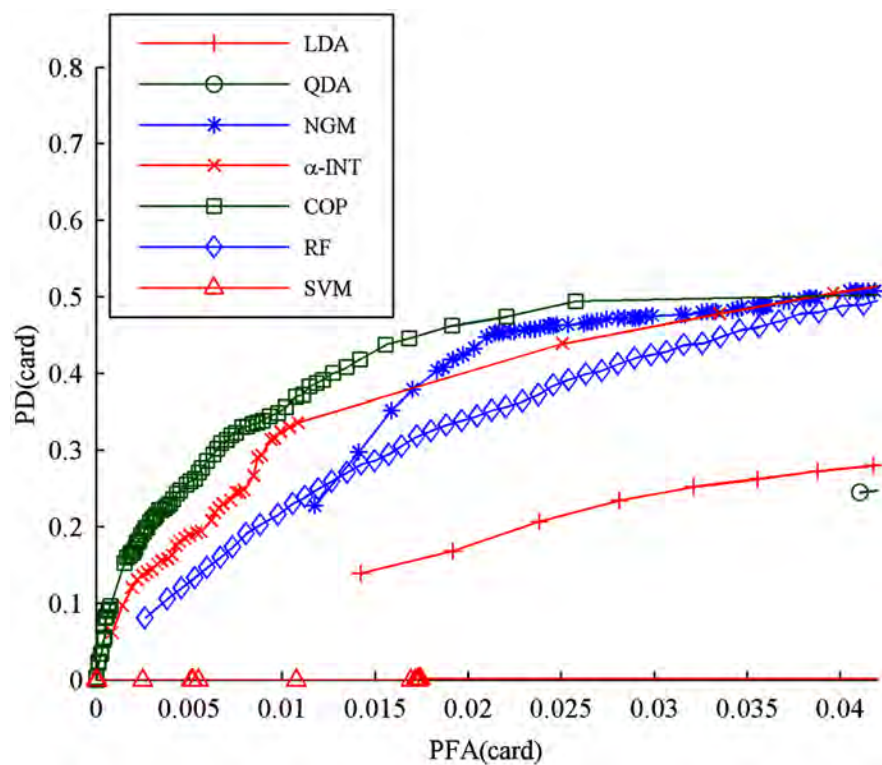


Figure 4. VDR with respect to normalized alerts per day estimated from detection methods

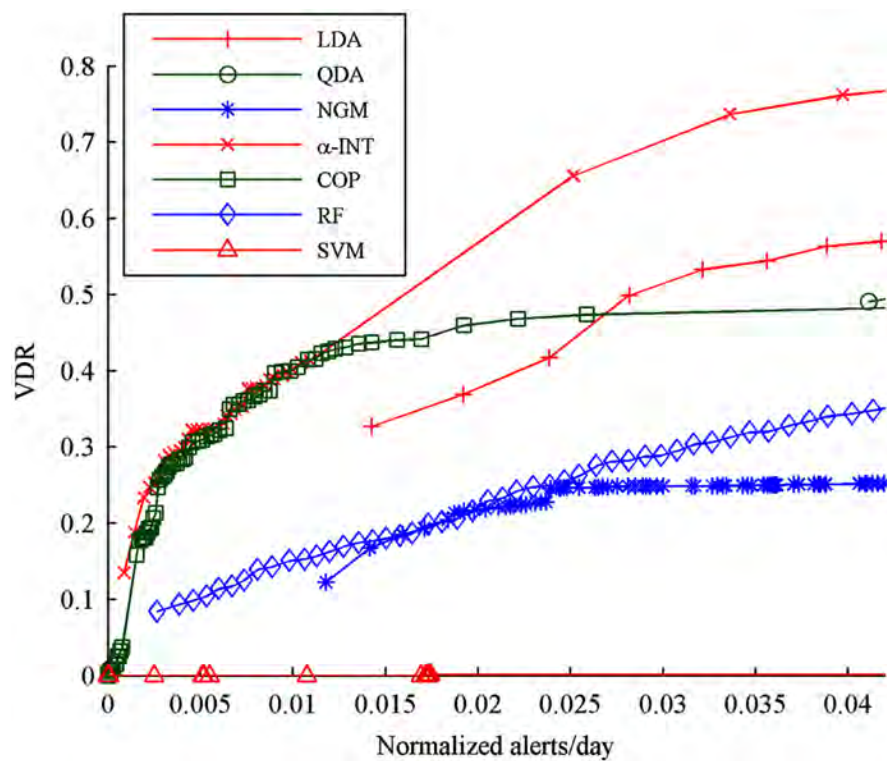
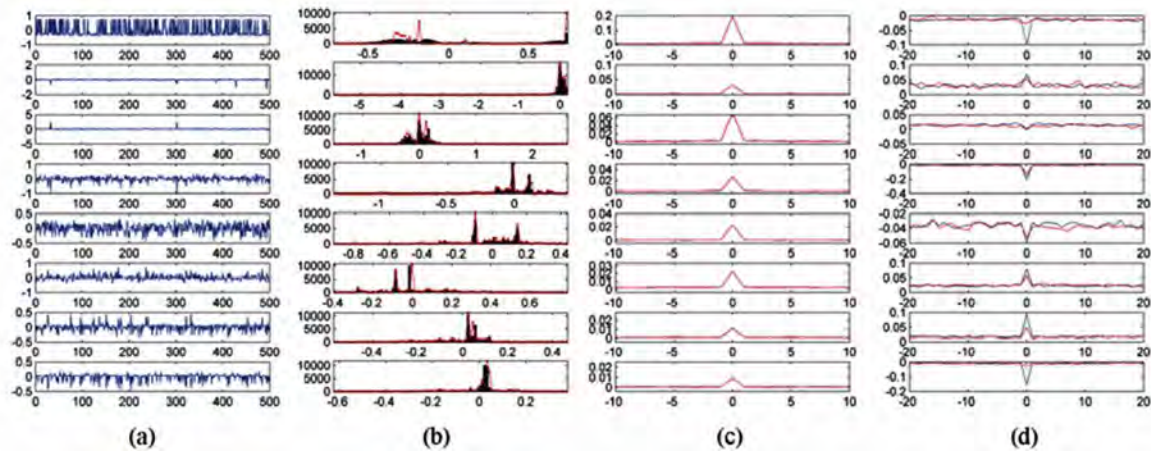


Figure 5. Results of surrogate data from legitimate operations



ponents in surrogate data, in black bars, with the histogram of the legitimate operations superimposed in red. Note, that those distributions are non-Gaussians. Figure 1c compares the autocorrelations of each surrogate component (in blue) with the corresponding autocorrelation of the legitimate data (in red). Finally, Figure 1d shows the cross-correlations of the first component with the other components (first seven rows) and the cross-correlation of the second component with the third component (eighth row). The cross-correlations of the surrogate data are shown in blue, and the cross-correlations of the legitimate data are shown in red. It can be seen that surrogate

legitimate operations were very close to the real legitimate operations from standpoints of marginal distributions of the components and autocorrelation and cross-correlations between variables.

Figure 7a shows the joint probability density functions (pdf) of real (blue) and surrogate (red) legitimate operations. Similarly, joint pdfs for fraud operations are shown in Figure 7b. Although the joint pdfs are calculated for all 8 components at once, they are shown in slices of two components to make the result more readable. The results in Figure 7 show that the surrogate data had very similar joint pdf to that of real legitimate data. The peaks of the pdf for the surrogates were located at

Figure 6. Results of surrogate data from fraud operations

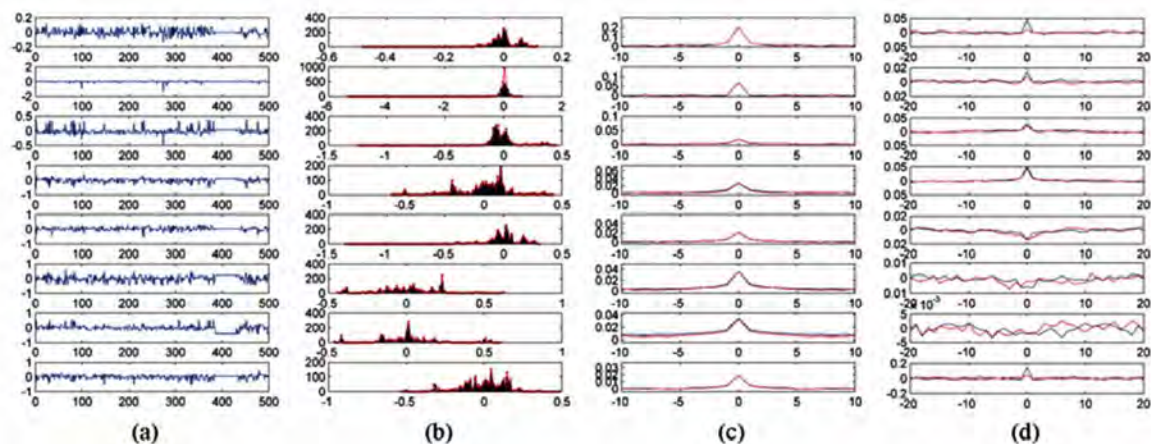
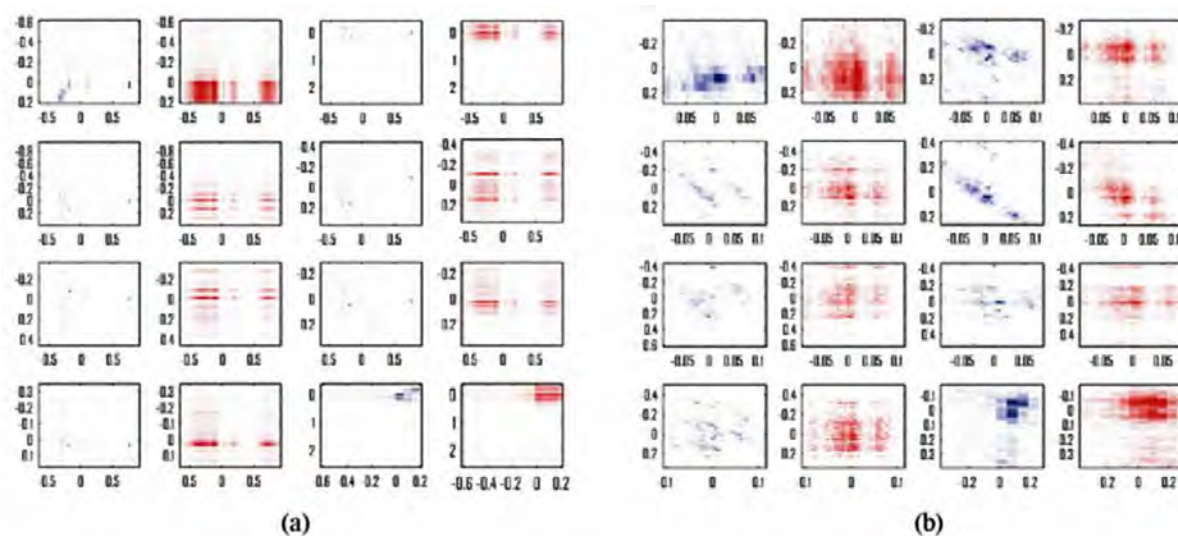


Figure 7. Joint distribution of a) legitimate and b) fraud operations



the same positions as the peaks of the pdf for the true data, and the shapes of both pdfs were very close in all cases. The pdf of the surrogates were more diffuse (i.e. less discrete), which might be caused by the amount of non-stationarity in the legitimate operations.

Similarity of the surrogates and the real operations was also tested by performing ACCFD using the method α - INT explained above. A certain amount of randomly-selected operations were replaced with surrogate operations and the

detection was repeated. The amount of replaced operations was changed from 0% (only true operations) to 50%, 75%, and 100% (only surrogate operations). The obtained ROC curves are shown in Figure 8 showing that the curves are similar in all cases even for low false alarm values (Figure 8b). There are some differences in value for very low amounts of FPR. In all cases, the lower the amount of surrogate operations, the closer the ROC became to that of the real operations (all the curves were very close for FPR above 0.06).

Figure 8. ROC curves for different percentages of surrogate data: a) complete curves; b) zoom in a zoom of interest

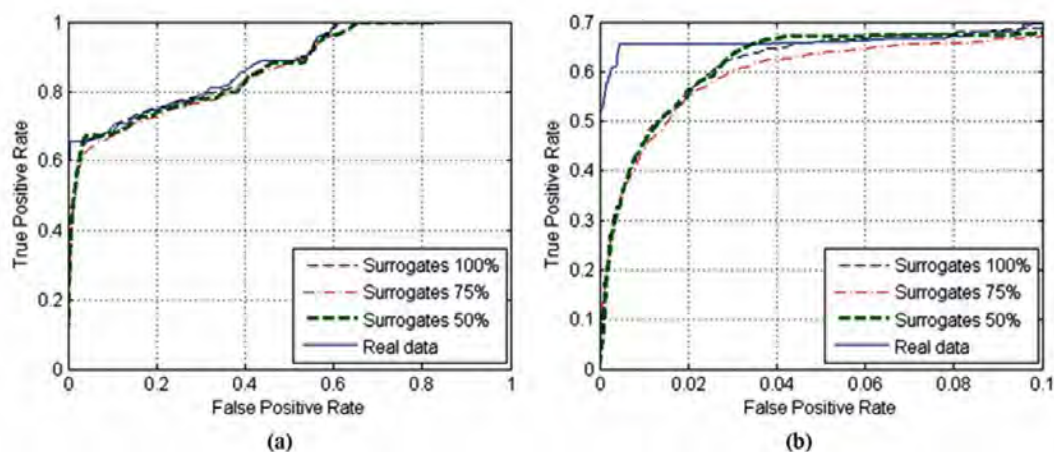


Figure 9. ROC curves for different proportions of fraud/legitimate operations obtained by different methods

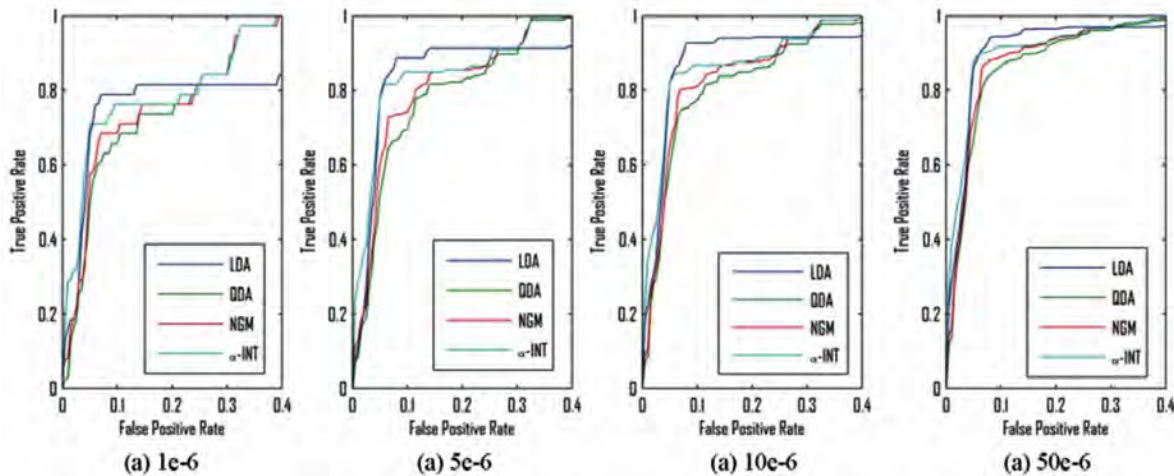


Figure 9 shows the results of applying different algorithms for ACCFD changing the fraud/legitimate ratio. It can be seen that the performance decreases the ratio decreases with α -INT method obtaining best results for all ratios. Furthermore, α -INT performance shows robustness at very-low false alarm values. Thus, exploiting complementarities of basic detectors seems to be immune for the fusion method.

FUTURE RESEARCH DIRECTIONS

Improvement of fusion techniques to take full advantage of complementarities among detectors is still required. A challenging issue is to consider in the fusion method that the behavior of the detectors for transaction scoring is not regular for the whole dataset. Also changing fraud dynamics might continuously distorts the complementarities among the detectors making tighter the design of an optimum fusion method. Methods that incorporate prediction and temporal dependence should be devised for ACCFD.

Besides, the combination of different kinds of general detectors presented here, there are other alternatives that can be explored. Several versions of a kind of detector that are trained in different

ways can be combined, for instance, using different versions of the data, searching specialized decision trees for each of the features, etc. Some examples of that kind of methods are boosting, bagging, and mixtures of experts. One of the critical issues for these ensemble approaches is to adequately handle with the so-called bias-variance tradeoff. Small bias indicates that training data has been learned well (small error of prediction) and small variance indicates capability of generalization, i.e., unobserved data might be predicted with small error. The problem consists of a low variance causes a low bias and vice-versa.

Technological constraints and prospective arising from big data context including data storage, privacy and security, and time processing issues are also subjects for future research.

CONCLUSION

We have presented a new application having unquestionable interest for the signal processing community. Some similarities and differences with classical detection problems in the signal processing area have been emphasized. Then it has been considered the use of fusion as a powerful technique to derive algorithms having

reasonable performance to solve this very complex problem. Experimental results using large real and surrogate datasets have been showed, where the improvements of using fusion of scores have been illustrated. Different well-known methods were compared with the new approaches. The detection results obtained for different mixtures of real and surrogate data were comparable under real requirements of very low false alarm.

An important aspect to have in mind is the context. Thus, arriving to probabilities of detection of 40% (to give a number) may be useless in other areas like radar detection, but can be extremely significant in credit card fraud detection, as it could imply a big amount of money saving. Also the usual figures of merit should be replaced by the KPI. In this sense, much work remains to arrive to methods suited to “modified cost functions”, which better reflect the actual optimization to be reached, or to practical methods that can select the best detectors for a simultaneous fitting of the KPI.

ACKNOWLEDGMENT

This work was supported by Spanish Administration and European Union under grant TEC2014-58438-R and Generalitat Valenciana under grant PROMETEO II/2014/032.

REFERENCES

- Atrey, P., Hossain, M., El Saddik, A., & Kankanhalli, M. (2010). Multimodal fusion for multimedia analysis: A survey. *Multimedia Systems*, 16(6), 345–379. doi:10.1007/s00530-010-0182-0
- Bhattacharyya, S., Jha, S., Tharakunnel, K., & Westland, J. C. (2011). Data mining for credit card fraud: A comparative study. *Decision Support Systems*, 50(3), 602–613. doi:10.1016/j.dss.2010.08.008
- Borgnat, P., Abry, P., & Flandrin, P. (2012). Using surrogates and optimal transport for synthesis of stationary multivariate series with prescribed covariance function and non-Gaussian joint distribution. *Proceedings of ICASSP* (págs. 3729-3732). Kyoto, Japan: IEEE. doi:10.1109/ICASSP.2012.6288727
- Danenas, P. (2015). Intelligent financial fraud detection and analysis: A survey of recent patents. *Recent Patents on Computer Science*, 8(1), 13–23. doi:10.2174/2213275907666141101001436
- Girgenti, R. H., & Hedley, T. P. (2011). *Managing the risk of fraud and misconduct: meeting the challenges of a global, regulated and digital environment*. McGraw-Hill.
- Hodge, V., & Austin, J. (2004). A survey of outlier detection methodologies. *J. of Art. Intell.*, 22(2), 85–126.
- Iyengar, S. G., Varshney, P. K., & Damarla, T. (2011). A Parametric Copula-Based Framework for Hypothesis Testing Using Heterogeneous Data. *IEEE Transactions on Signal Processing*, 59(5), 2308–2319. doi:10.1109/TSP.2011.2105483
- Jain, A. K., Duin, P. W., & Jianchang Mao, (2000). Statistical pattern recognition: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1), 4–37. doi:10.1109/34.824819
- Kittler, J., Hatef, M., Duin, R. P., & Matas, J. (1998). On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3), 226–23. doi:10.1109/34.667881
- Lifeng, L., Ho, S. W., & Poor, H. V. (2011). Privacy-Security Trade-Offs in Biometric Security Systems - Part I: Single Use Case. *IEEE Trans. Inf. Forensics Security*, 6(1), 122–139. doi:10.1109/TIFS.2010.2098872
- Montague, D. A. (2010). *Essentials of online payment security and fraud prevention*. Wiley.

- Pawar, S., El Rouayheb, S., & Ramchandran, K. (2011). Securing Dynamic Distributed Storage Systems Against Eavesdropping and Adversarial Attacks. *IEEE Transactions on Information Theory*, 57(10), 6734–6753. doi:10.1109/TIT.2011.2162191
- Poor, H. V. (2012). Information and inference in the wireless physical layer. *Wireless Communications*, 19(1), 40–47. doi:10.1109/MWC.2012.6155875
- Prichard, D., & Theiler, J. (1994). Generating surrogate data for time series with several simultaneously measured variables. *Physical Review Letters*, 73(7), 951–954. doi:10.1103/PhysRevLett.73.951 PMID:10057582
- Salazar, A. (2013). *On Statistical Pattern Recognition in Independent Component Analysis Mixture Modelling*. Berlin, Germany: Springer. doi:10.1007/978-3-642-30752-2
- Salazar, A., Safont, G., Soriano, A., & Vergara, L. (2012). Automatic credit card fraud detection based on non-linear signal processing. *46th Annual IEEE International Carnahan Conference (ICCST)* (pp. 207–212). Boston, MA: IEEE. doi:10.1109/CCST.2012.6393560
- Salazar, A., Safont, G., & Vergara, L. (2014). Surrogate techniques for testing fraud detection algorithms in credit card operations. *48th International Carnahan Conference on Security Technology (ICCST2014)* (pp. 1–7). Rome, Italy: IEEE. doi:10.1109/CCST.2014.6986987
- Salazar, A., Vergara, L., Serrano, A., & Igual, J. (2010). A General Procedure for Learning Mixtures of Independent Component Analyzers. *Pattern Recognition*, 43(1), 69–85. doi:10.1016/j.patcog.2009.05.013
- Sankar, L., Rajagopalan, S. R., & Poor, H. V. (2013). Utility-Privacy Tradeoff in Databases: An Information-theoretic Approach. *IEEE Trans. Inf. Forens. Sec.*, 8(6), 838–852. doi:10.1109/TIFS.2013.2253320
- Schreiber, T., & Schmitz, A. (2000). Surrogate time series. *Physica D. Nonlinear Phenomena*, 142(3–4), 346–382. doi:10.1016/S0167-2789(00)00043-9
- Shafer, G. (1976). *A Mathematical Theory of Evidence*. Princeton University Press.
- Soriano, A., Vergara, L., Ahmed, B., & Salazar, A. (2015). Fusion of Scores in a Detection Context Based on Alpha Integration. *Neural Computation*, 27(9), 1983–2010. doi:10.1162/NECO_a_00766 PMID:26161815
- Tax, D., & Duin, R. (2001). Uniform object generation for optimizing one-class classifiers. *Journal of Machine Learning Research*, 2, 155–173.
- Tzikas, D. G., Likas, A. C., & Galatsanos, N. (2008). The variational approximation for Bayesian inference. *IEEE Signal Processing Magazine*, 25(6), 131–146. doi:10.1109/MSP.2008.929620
- Vergara, L., Soriano, A., Safont, G., & Salazar, A. (2016). On the fusion of non-independent detectors. *Digital Signal Processing*, 50, 24–33. doi:10.1016/j.dsp.2015.11.009
- Wells, J. T. (2011). *Corporate fraud handbook: prevention and detection* (3rd ed.). Wiley.
- Yuksel, S. E., Wilson, J. N., & Gader, P. D. (2012). Twenty Years of Mixture of Experts. *IEEE Trans. On Neural Networks and Learning Systems*, 23(4), 1177–1193. doi:10.1109/TNNLS.2012.2200299 PMID:24807516

ADDITIONAL READING

- Safont, G., Salazar, A., Rodriguez, A., & Vergara, L. (2014). On Recovering Missing Ground Penetrating Radar Traces by Statistical Interpolation Methods. *Remote Sensing*, 6(8), 7546–7565. doi:10.3390/rs6087546

Safont, G., Salazar, A., Soriano, A., & Vergara, L. (2012). Combination of multiple detectors for EEG based biometric identification/authentication. *Proceedings of 46th Annual 2012 International Carnahan Conference on Security Technology (ICCST)* (págs. 230-236). Boston: IEEE. doi:10.1109/CCST.2012.6393564

Salazar, A., Vergara, L., & Miralles, R. (2010). On including sequential dependence in ICA mixture models. *Signal Processing*, 90(7), 2314–2318. doi:10.1016/j.sigpro.2010.02.010

Slavakis, K., Giannakis, G. B., & Mateos, G. (2014). Modeling optimization for big data analytics. *IEEE Signal Processing Magazine*, 31(5), 18–31. doi:10.1109/MSP.2014.2327238

Zhai, Y., Ong, Y. S., & Tsang, I. W. (2014). The emerging big dimensionality. *IEEE Computational Intelligence Magazine*, 9(3), 14–26. doi:10.1109/MCI.2014.2326099

KEY TERMS AND DEFINITIONS

Classifier: A computational method that can be trained using known labeled data for predicting the label of unlabeled data. If there's only two

labels (also called classes), the method is called “detector”.

Fusion of Scores: A process for efficient combination of a set of scores or probabilistic grades granted by a set of detectors/classifiers. Evaluation of combination performance depends on the application objectives.

KPI: Key performance indicator is an index that defines the performance of a process. In the case of ACCFD, roughly, the set of KPIs serves to translate ROC results into business terms.

Pattern Recognition: Automatic process of extracting, representing, and splitting conspicuous characteristics from a dataset to produce several subsets that can be associated with concepts normally accepted by humans.

Receiving Operating Characteristic (ROC) Curve: A numeric tool for the evaluation of a detection process that implements comparison in a coordinate plane of probability of detection and probability of false alarm at different operating points from 0 to 1.

Surrogate Data: Data that are computationally generated that behave with statistical features similar to that of real data.

Privacy, Algorithmic Discrimination, and the Internet of Things



Jenifer Sunrise Winter

University of Hawaii at Manoa, USA

INTRODUCTION

The Internet of Things (IoT) is a paradigm encompassing a wide range of developments that enable everyday objects to be tagged and uniquely identified over the Internet. Although there is no single definition for the Internet of Things, competing visions agree that it relates to the integration of the physical world with the virtual world, with any object having the potential to be connected to the Internet via short-range wireless technologies, such as radio frequency identification (RFID), near field communication (NFC), or Wireless Sensor Networks (WSNs). This merging of the physical and virtual worlds will “enable the Internet to reach out into the real world of physical objects” (Internet of Things Conference Organizing Committee, 2010). Further, it will allow increased instrumentation, tracking and measurement of the natural world, enabling analytic tools to enhance business management processes and offer citizens increased convenience and safety (Uckelmann, Harrison, & Michahelles, 2010).

The IoT is imagined as a “backbone for ubiquitous computing, enabling smart environments to recognize and identify objects, and retrieve information from the Internet to facilitate their adaptive functionality” (Weber & Weber, 2010, p. 1). In this regard, the IoT is an emerging global architecture that will enable enhanced machine intelligence to automate the exchange of goods and services. In addition to improving supply chain management, this integration of tags and sensor networks will also be employed in diverse application scenarios, including smart appliances

and smart homes, disaster warning, structural engineering, farming, and in-vivo health applications (Atzori, Iera, & Morabito, 2010). This chapter will introduce the Internet of Things, address its definition and related concepts, outline anticipated application areas, highlight challenges, and discuss privacy and surveillance concerns.

BACKGROUND

Related Areas

Current research agendas focus on the IoT ecosystem—networks of physical objects embedded with the ability to sense, and sometimes act upon, their environment, as well as related communication, applications, and data analytics (Gartner, 2014). The IoT is often mentioned in relation to other, overlapping research paradigms, particularly Ubiquitous Computing, Pervasive Computing, and Ambient Intelligence, research agendas that address the integration of myriad, heterogeneous objects into the everyday environment. Weiser’s (1991) vision of Ubiquitous Computing emerged in the late 1980s and emphasized the potential of multiple computers per person, in a variety of forms, to activate the physical environment and make computational intelligence an extension of human activity. Ubiquitous Computing research is distinguished by its human-centered focus and has increasingly addressed interaction contexts (Abowd, Ebling, Hunt, Lei, & Gellersen, 2002). The related concept of Pervasive Computing (Hoffnagle, 1999) emerged as a corporate vision

at IBM during the late 1990s. This agenda has focused on the technical systems required to embed numerous, networked devices throughout the environment. Over time, the two research communities have overlapped, and the two leading conferences, ACM's Pervasive and UbiComp, merged in 2013. Ambient Intelligence research has been guided by the European Union's Fifth Framework Programme (Information Society Technologies, 1998-2002) and focuses on embedded devices, particularly those in smart homes, which are context-sensitive and tailored towards personal needs. While the IoT overlaps technical developments in these related areas, it is distinguished by several concepts. These include 1) goals for an architecture that provides billions, or trillions, of heterogeneous objects with unique identifiers that allow them to interact over a global network; and 2) an emphasis on machine-to-machine (M2M) communication. Although all of these paradigms tend to focus on near-term visions of potential future environments (Dourish & Bell, 2011), the IoT is already manifest in various ways today.

Origin and Evolution of the Concept

Kevin Ashton is credited with the first use of the phrase "Internet of Things" in 1999. He focuses on the potential of M2M intelligence to capture real-time data about the physical world and use it without direct human oversight (Ashton, 2009), stating that the goals of IoT research and development focus on endowing computers

with their own means of gathering information, so they can see, hear and smell the world for themselves, in all its random glory. RFID and sensor technology enable computers to observe, identify and understand the world—without the limitations of human-entered data. (Ashton, 2009, para. 5)

Although the concept arose in the context of supply chain management, he emphasizes that the IoT now applies to nearly any aspect of the physical world.

Since Ashton's first use of the phrase, the IoT has been associated with a wide variety of academic, government, and corporate research projects. An early and substantial influence was the Massachusetts Institute of Technology's Auto-ID Center, which was formed in 1999 to create an RFID-based IoT architecture for global supply-chain management. In 2003, the Auto-ID Center closed, and EPCglobal took over its commercial work. This effort spun off into the Auto-ID Labs, seven academic research labs that focus on building a global infrastructure using RFID or other short-range, wireless technologies. The Auto-ID Labs developed the Electronic Product Code (EPC), an RFID tag that uniquely identifies a product, and the EPCglobal stack is the de facto standard for retail and consumer goods industries.

Other large-scale efforts include IBM's Smarter Planet strategy initiated in 2008 (IBM, 2008), which has focused on the potential of instrumenting the physical world with tens of billions (perhaps trillions) of interconnected sensors. Here, the IoT is envisioned as a way to fuel economic growth, business and government efficiency, improve physical security, and enhance scientific knowledge. Similarly, HP's Central Nervous System for the Earth (CeNSE) initiative relies on billions of nanoscale sensors to detect vibration and motion (HP, n.d.)

The IoT has also figured prominently in national and international technology policy strategies. Following the International Telecommunication Union's special report on the IoT (International Telecommunication Union, 2005a), the European Union's Directorate, General Information Society and Media shifted concern to the IoT (European Commission, Community Research and Development Information Service, 2012) with its i2010 policy framework for the information society and media (2005-2010). Under its Seventh Framework Programme (FP7), the EU funded the CASAGRAS (Coordination and support action for global RFID-related activities and standardization) program, which has provided a framework for research to aid the European Commission in

navigating international issues related to RFID, in particular the IoT (CASAGRAS, 2009). This project defines the IoT as

A global network infrastructure, linking physical and virtual objects through the exploitation of data capture and communication capabilities. This infrastructure includes existing and evolving Internet and network developments. It will offer specific object-identification, sensor and connection capability as the basis for the development of independent cooperative services and applications. These will be characterised by a high degree of autonomous data capture, event transfer, network connectivity and interoperability. (CASAGRAS, 2009, p.10)

Simultaneously, the industry-driven policy initiative EPoSS (European Technology Platform on Smart Systems Integration) has sought to direct research and innovation needs related the IoT (European Technology Platform on Smart Systems Integration, n.d.).

China is also taking a lead role in IoT research and development. In 2009, Chinese Premier Wen Jiabao announced that IoT will be central to China's coordinated, national strategic development (GSMA, 2015). The IoT appears in China's 12th Five Year Plan, which guides policy from 2011–2015 (Hvistendahl, 2012). This has led to a number of IoT research parks throughout China, including an Internet of Things Center established in Shanghai and the City of Wuxi redesigned as a research and development center.

Components of the Internet of Things

Atzori et al. (2010) argue that the IoT arises from three different, overlapping visions: 1) things-oriented visions focusing on real-world objects embedded with RFID or sensor technologies; 2) Internet-oriented visions; and 3) semantic-oriented visions. The components of each of these are addressed next.

‘Things’ are objects in the natural world that we wish to instrument. Each thing must be connected to a communication device, such as an RFID chip, a sensor, or an actuator. These can be embedded in a variety of objects, including mobile phones, automobiles, home appliances, clothing, or even the human body. In addition to RFID-tagged objects or NFC-embedded smartphones, things may include nanoelectronics, sensors, or other embedded systems that detect, measure, compute, and communicate (Vermesan et al., 2011). Data about these objects can include location-based information, data from the environment that the object is sensing, or other, stored data. According to Vermesan et al. (2011), things are

expected to become active participants in business, information, and social processes where they are enabled to interact and communicate among themselves and with the environment by exchanging data and information “sensed” about the environment, while reacting autonomously to the “real/physical” world events and influencing it by running processes that trigger actions and create services with or without direct human intervention. (p. 10)

These developments will be supported by the emergence of linked data, standards and practices for connecting structured data via the World Wide Web, creating a massive, global data space that can be navigated and processed by machine intelligence (Heath & Bizer, 2011). This will enable the Semantic Web (Berners-Lee, 2000), a web of data that can be processed by machines without human intervention.

Networks are a second area of research concern. The maturity of RFID, NFC, and wireless sensor networks (based on IEEE 802.15.4) is poised to allow the interconnection of various smaller networks to the Internet. However, whereas the Internet relies on the TCP/IP protocol stack for end-to-end transmission, there is no single standard to connect these intranets of objects to the Internet. A variety of communication standards



operating at various layers are in development. To reach a global scale and maximize the interoperability of heterogeneous systems, there is a need for development of flexible, open standards.

Semantic specifications, the third area of interest, are concerned with how to represent, store, organize, and query objects and data. Here, research focuses on the developments of autonomous software agents and semantic standards for machine communication. Vermesan et al. (2011) highlight the need for “ontology-based semantic standards [that] will enable mapping and cross-referencing between them, in order to enable information exchange” (p. 42).

Application Areas

Applications for the IoT are numerous and diverse, addressing tasks such as measuring location, temperature, acceleration, pollution, or chemical components. Data collected from these sensors is expected to enhance resource management in virtually every sector. The following is a list of potential uses:

- **Logistics/Supply Chain Management:** e.g., item tracking (Ashton, 2009); re-stocking; payment systems (Uckelmann & Harrison, 2010; Atzori et al., 2010).
- **Scientific Research:** e.g., remote or small scale experiment monitoring.
- **Health Care/Biomedical:** e.g., ambient sensors for independent living; implantable or edible medical devices (CERP-IoT, 2010).
- **Environment:** e.g., natural disaster prediction, such as flood, fire, earthquake, tsunami warning systems (CERP-IoT, 2010); chemical, and gas leak identification; pollution and temperature monitoring (Hvistendahl, 2012); water potability testing.
- **Security:** e.g., motion-sensitive camera activation; access control; radiation moni-

toring (Ishigaki, Matsumoto, Ichimiya, & Tanaka, 2013); intrusion detection (Khan, Khan, Zaheer, & Khan, 2012).

- **Structural Engineering:** e.g., monitoring and identifying faults in buildings, roads, or bridges (Agrawal & Lal Das, 2011).
- **Food Safety:** e.g., testing (Hvistendahl, 2012) and recall of tainted food (CERP-IoT, 2010).
- **Agriculture:** e.g., monitoring hydration, chemical composition, or soil quality; live-stock tracking (CERP-IoT, 2010).
- **Smart Cities:** Infrastructure monitoring; management of smart grids to govern cost- and resource-efficient use of energy (Khan et al., 2012; Atzori et al., 2010); “Green ICT” to lower environmental impact (Vermesan et al., 2011).
- **Smart Homes:** e.g., automatic lighting and power allocation (CERP-IoT, 2010).
- **Transportation:** Aerospace part authentication (CERP-IoT, 2010); sensor-enabled roads; assisted driving (Atzori et al., 2010).
- **Augmented Reality Maps:** e.g., maps equipped with tags enabling NFC-equipped phones to automatically interact with Web services (Atzori et al., 2010).
- **Social Networking:** e.g., automated updates of information on social networks (Atzori et al., 2010).
- **Transparency of Government Data:** Anti-corruption (Bertot, Jaeger, & Grimes, 2010).

PRIVACY AND THE INTERNET OF THINGS

An estimated Zettabyte (1,000,000,000,000,000,000 bytes) of data will flow over the Internet in 2016 (Cisco, 2014), as much data as in the entire history of the Internet since its creation in 1969. However, threats to privacy introduced by the IoT are not merely related

to the increased magnitude of data collection. The following are unique challenges to privacy in the context of the IoT (Winter, 2013):

1. **Lack of Transparency:** Small components are not necessarily visible. One may not be aware when and where data is being collected (Langheinrich, 2001).
2. **Sheer Number of Objects Monitoring the Environment:** Even under a scheme where one could “opt out” of data sharing, the interface to accomplish this would be unmanageably complex. Similarly, if an individual wishes to deactivate, or temporarily halt, an object’s communication potential, the number of objects might make this difficult.
3. **Novel Data Types:** Billions, or trillions, of everyday objects, even the human body itself, may be equipped with networked sensors. Many novel data types can be collected. For example, a variety of biometric data in aggregate form.
4. **Increased Aggregation of User Data:** As data analytics is increasingly embedded into the core of business and operational functions (Davenport, Barth, & Bean, 2012), there is a strong economic incentive to model user data. Thus, data may be collected indiscriminately for future use.
5. **De-Anonymization of Personal Data:** Data that has been anonymized in order to meet legal requirements can be “re-personalized” via data mining techniques (Schwartz & Solove, 2011; Winter, 2016).
6. **M2M Processing and Communication:** Machine intelligence will be used both to collect and to analyze these data, increasing both the scope and analytic potential of big data analytics.

Unjust Algorithmic Discrimination

Unjust discrimination based on big data analytics is a growing concern (e.g., Custers, 2013, boyd,

Levy, & Marwick, 2014). Various data about our daily lives are being mined to reveal associations or behaviors, leading to potential political or economic discrimination by governments and corporations in the form of offering different services, products, or prices to individuals based on their data profile (Turow, 2012; Winter, 2015). Examples include:

- Insurers are beginning to allocate risk differently (Upturn, 2014). As new forms of price differentiation emerge, risk calculations by insurers will no longer be spread across a large pool of people, leading to higher burdens (e.g., denial of coverage or higher costs) for some.
- In the criminal justice system, data-based risk assessment is now being considered for use at sentencing. The State of Pennsylvania may allow “some offenders considered low risk to get shorter prison sentences than they would otherwise or avoid incarceration entirely. Those deemed high-risk could spend more time behind bars” (Barry-Jester, Casselman, & Goldstein, 2015, para. 4).
- Data gathered via the smart grid might be used to collect personal information. The National Institute of Standards and Technology observed that smart grid data might reveal specific lifestyle information that could be used by insurers or commercial service providers (National Institute of Standards and Technology, 2010).
- Researchers were able to uniquely identify males who had shared personal DNA sequences on Internet genealogy forums based on publicly available data (Gymrek, McGuire, Golan, Halperin, & Erlich, 2013). While in the United States the *Genetic Information Nondiscrimination Act* is intended to prevent abuse of genetic data, the financial incentives for mining these data are immense, and big data ana-



lytics may enable non-protected “proxy” fields (i.e., other patterns or data-based evidence that is not explicitly protected) to be used instead (Barocas & Selbst, 2016).

In 2013, news coverage about the existence of Internet surveillance programs by the United States’ National Security Agency and its partners in other countries was credited for growing public awareness and concern about government surveillance (Pew Research Center for the People and the Press, 2013). This growing concern may cause additional resistance to IoT developments. Further, self-censorship due to privacy concerns may limit citizens’ freedom of access to information or discussion of issues relevant to democratic decision-making.

Challenges for the Internet of Things

The IoT research community is addressing a wide variety of technical and social issues. Overall, key challenges for development of the IoT include:

- Development of a universal architecture based on open, layered standards (Vermesan et al., 2011).
- Network scalability (Uckelmann et al., 2011).
- Integration of non-IP-based things into the network (Uckelmann et al., 2011).
- Design of energy-efficient objects (Atzori et al., 2010).
- Addressing schemes (Vermesan et al., 2011).
- Autonomous agent design (Vermesan et al., 2011).
- Data storage, representation, and querying standards (Vermesan et al., 2011).
- Mobility management (Atzori et al., 2010).
- Creation of large-scale testbeds for real-world experimentation (Gluhak, Krco, Nati, Pfisterer, Mitton, & Razafindralambo, 2011).

- Spectrum efficiency (Khan, Khan, Zaheer, & Khan, 2012).
- Object and network security (Weber, 2010).

Each of these areas presents a set of robust challenges, and *privacy* should be considered throughout the design process and product lifecycle, rather than as an afterthought. Below, solutions and recommendations for this complex problem are outlined.

SOLUTIONS AND RECOMMENDATIONS

There is no single solution to address privacy concerns related to the Internet of Things. A complex and ongoing approach addressing three domains – sociological, technical, and regulatory – is required (International Telecommunication Union, 2005b).

The sociological domain includes public education and ongoing negotiation about acceptable uses in various contexts. Fostering discussion with a broad range of stakeholders at an early stage of IoT development is necessary to ensure that these developments align with community values and goals (Winter, 2008).

The technological domain includes an array of privacy enhancing technologies (PETs). PETs can include software that allows users to set privacy preferences, “hide” from others on the network, or erase traces of activity. Other means include encryption tools, peer-to-peer systems (P2P), or the ability to switch off tags or sensors (Weber & Weber, 2010). All of these approaches have limitations and, due to the scale and heterogeneity of the IoT, do not provide a satisfactory solution to privacy concerns. Cavoukian and Kursawe (2012) argue that privacy must be introduced during the earliest stages of system development. Privacy by design is a design framework with principles that protect personal privacy while

enabling high-quality data collection and analysis. Using the principles in the framework, a variety of systems with diverse goals and contexts can be designed with privacy at the core, rather than as an afterthought.

Because the IoT is a global system, the regulatory domain is particularly complex. Weber and Weber (2010) argue that the time to create legal protections is before major problems arise. Weber (2010) argues that a new privacy approach is required. “The nature of the IoT asks for a heterogeneous and differentiated legal framework that adequately takes into account the globality, verticality, ubiquity and technicity of the IoT” (p. 30). He acknowledges that neither industry self-regulation nor national laws will be sufficient. As Hildner (2006) observes, “Experience demonstrates that legally unenforceable self-regulation will not be a sufficient limitation on RFID’s threat to privacy” (p. 159). On the other hand, omnibus laws (such as those in the European Union) may be difficult to enforce outside of their jurisdiction. Weber (2010) argues for the development of key principles to be set by a transparent, international organization complemented with more detailed regulation originating in the private sector. However, the feasibility of this approach is uncertain. In December, 2013, the United Nations General Assembly adopted a consensus resolution strongly supporting a right to privacy (United Nations News Centre, 2013). The IoT is a complex sociotechnical system comprised of a plurality of actors, networks, institutions, and contexts, and efforts to regulate privacy at an international level may not be sufficient. Instead, we may see global privacy regulations continue at a variety of levels and contexts rather than as a global, coordinated approach.

FUTURE RESEARCH DIRECTIONS

As outlined above, there are a number of critical technical challenges for the Internet of Things,

including issues of network scalability, integration of non-IP-based objects into the network (Uckelmann et al., 2011), standards for data storage, representation, and querying (Vermesan et al., 2011), and spectrum efficiency (Khan, Khan, Zaheer, & Khan, 2012). Further, there is a need for increased focus on security and privacy (Weber, 2010). Focusing on privacy, each of the three domains outlined by the International Telecommunication Union (2005b) – sociological, technical, and regulatory – present fertile areas for research related to privacy and the emerging Internet of Things. Within the sociological domain, research is needed to better understand users’ behaviors and attitudes related to IoT use, particularly in specific contexts of use. Within the technological domain, many aspects of IoT network and object design will benefit from robust security- and privacy-enhancing technologies, especially those that are built into overall system design. Within the regulatory domain, there is a need for development of an internationally-recognized set of key principles related to privacy rights. Research is also required to map and assess the distributions of social benefits and harms related to the IoT in order to determine whether the resulting distributions serve the public good. Further, there is a need for studies that promote a clearer understanding and assessment of privacy regulation at national, regional, and international levels.

CONCLUSION

This chapter introduced the IoT and provided an overview of its main components and various research initiatives that guide its development. The IoT is likely to have many beneficial applications that impact nearly every sector, and a number of these noted in the literature were addressed. Further, key challenges were outlined, including the issues of privacy and unjust algorithmic discrimination. Foresight activities involving a broad range of stakeholders are necessary both



for the formulation of legislation and to inform system design. IoT systems should be designed with privacy in mind from the start.

REFERENCES

- Abowd, G. D., Ebling, M., Hunt, G., Lei, H., & Gellersen, H.-W. (2002). Context-aware computing. *IEEE Pervasive Computing/IEEE Computer Society [and] IEEE Communications Society*, 1(3), 22–23. doi:10.1109/MPRV.2002.1037718
- Agrawal, S., & Lal Das, M. (2011, December). *Internet of Things – a paradigm shift of future Internet applications*. Paper presented at the Second International Conference on Current Trends in Technology (NUICONE 2011), Ahmedabad. doi:10.1109/NUIConE.2011.6153246
- Ashton, K. (2009). That ‘Internet of Things’ thing. *RFID Journal*. Retrieved May 11, 2010, from <http://www.rfidjournal.com/article/view/4986>
- Atzori, L., Iera, A., & Morabito, G. (2010). The Internet of Things: A survey. *Computer Networks*, 54(15), 2787–2805. doi:10.1016/j.comnet.2010.05.010
- Barocas, S., & Selbst, A. D. (2016). Big data’s disparate impact. *California Law Review*, 104.
- Barry-Jester, A. M., Casselman, B., & Goldstein, C. (2015, August 4). *Should prison sentences be based on crimes that haven’t been committed yet?* Retrieved from <http://fivethirtyeight.com/features/prison-reform-risk-assessment/>
- Berners-Lee, T. (2000). *Weaving the Web: The past, present and future of the World Wide Web by its inventor*. London: Texere.
- Bertot, J. C., Jaeger, P. T., & Grimes, J. M. (2010). Using ICTs to create a culture of transparency: E-government and social media as openness and anti-corruption tools for societies. *Government Information Quarterly*, 27(3), 264–271. doi:10.1016/j.giq.2010.03.001
- boyd, d., Levy, K., & Marwick, A. (2014). The networked nature of algorithmic discrimination. In S. Gangadharan (Ed.), *Data and discrimination: Collected essays* (pp. 53–57). Washington, DC: Open Technology Institute – New America Foundation.
- CASAGRAS. (2009). *RFID and the inclusive model for the Internet of Things: Final report*. EU Framework 7. Retrieved from [http://www.grifs-project.eu/data/File/CASAGRAS%20FinalReport%20\(2\).pdf](http://www.grifs-project.eu/data/File/CASAGRAS%20FinalReport%20(2).pdf)
- Cavoukian, A., & Kursawe, K. (2012). Implementing privacy by design: The smart meter case. *Proceedings of the 2012 IEEE International Conference on Smart Grid Engineering* (pp. 1–8). Piscataway, NJ: IEEE. doi:10.1109/SGE.2012.6463977
- CERP-IoT. European Union, Cluster of European Research Projects on the Internet of Things. (2010). *Vision and challenges for realising the Internet of Things*. Brussels: European Commission – Information Society and Media.
- Cisco. (2014, June 10). *The Zettabyte era: Trends and analysis*. Cisco White Paper. San Jose, CA: Cisco Systems. Retrieved from http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/VNI_Hyperconnectivity_WP.pdf
- Custers, B. (2013). Data dilemmas in the information society: Introduction and overview. In B. Custers, T. Calders, B. Schermer, & T. Zarsky (Eds.), *Discrimination and privacy in the information society: Data mining and profiling in large databases* (pp. 3–26). New York: Springer. doi:10.1007/978-3-642-30487-3_1
- Davenport, T. H., Barth, P., & Bean, R. (2012). *How ‘big data’ is different*. Retrieved from <http://sloanreview.mit.edu/article/how-big-data-is-different/>

Dourish, P., & Bell, G. (2011). *Divining a digital future: Mess and mythology in ubiquitous computing*. Cambridge, MA: The MIT Press. doi:10.7551/mitpress/9780262015554.001.0001

European Commission, Information Society and Media. (2008). *Internet of Things in 2020: Roadmap for the future. European Technology Platform on Smart Systems Integration*. Version 1.1. Author.

European Technology Platform on Smart Systems Integration (EPoSS). (n.d.). *Objectives and mission*. Retrieved from <http://www.smart-systems-integration.org/public/about/objectives-mission>

Gartner. (2014, November 11). *Gartner says 4.9 billion collected 'Things' will be in use in 2015*. Retrieved from <https://www.gartner.com/newsroom/id/2905717>

Gluhak, A., Krco, S., Nati, M., Pfisterer, D., Mitton, N., & Razafindralambo, T. (2011). A survey of facilities for experimental Internet of Things research. *IEEE Communications Magazine*, 49(11), 58–67. doi:10.1109/MCOM.2011.6069710

GSMA. (2015). *How China is scaling the Internet of Things*. Retrieved from <http://www.gsma.com/newsroom/wp-content/uploads/16531-China-IoT-Report-LR.pdf>

Gymrek, M., McGuire, A. L., Golan, D., Halperin, E., & Erlich, Y. (2013). Identifying personal genomes by surname inference. *Science*, 339(6117), 321–324. doi:10.1126/science.1229566 PMID:23329047

Haggerty, K. D., & Ericson, R. V. (2006). The new politics of surveillance and visibility. In K. D. Haggerty & R. V. Ericson (Eds.), *The new politics of surveillance and visibility* (pp. 3–25). Toronto: University of Toronto Press.

Heath, T., & Bizer, C. (2011). Linked data: Evolving the Web into a global data space. *Synthesis Lectures on the Semantic Web: Theory and Technology*, 1(1), 1–136.

Hildner, L. (2006). Defusing the threat of RFID: Protecting consumer privacy through technology-specific legislation at the state level. *Harvard Civil Rights-Civil Liberties Law Review*, 41, 133–176.

Hoffnagle, G. F. (1999). Preface. *IBM Systems Journal*, 38(4), 502–503. doi:10.1147/sj.384.0502

HP. (n.d.). *CeNSE*. Retrieved from <http://www8.hp.com/us/en/hp-information/environment/cense.html#Uo1MKMQQZhE>

Hvistendahl, M. (2012). China pushes the Internet of Things. *Science*, 336(6086), 1223–1223. doi:10.1126/science.336.6086.1223 PMID:22679075

IBM. (2008, November). *A mandate for change is a mandate for smart*. Conversations for a smarter planet series, 1. Retrieved from http://www.ibm.com/smarterplanet/global/files/us_en_us_general_smarterplanet_overview.pdf

International Telecommunication Union. (2005a). *The Internet of Things*. Geneva: International Telecommunication Union.

International Telecommunication Union. (2005b). Privacy and Ubiquitous Network Societies: Background paper. *ITU Workshop on Ubiquitous Network Societies*. Geneva: International Telecommunication Union.

Internet of Things Conference Organizing Committee. (2010). *Internet of Things*. Retrieved from <http://www.iot2010.org/outline/>

Ishigaki, Y., Matsumoto, Y., Ichimiya, R., & Tanaka, K. (2013). Development of mobile radiation monitoring system utilizing smartphone and its field tests in Fukushima. *IEEE Sensors Journal*, 13(10), 3520–3526. doi:10.1109/JSEN.2013.2272734

Khan, R., Khan, S. U., Zaheer, R., & Khan, S. (2010). Future Internet: The Internet of Things architecture, possible applications and key challenges. *10th International Conference on Frontiers of Information Technology*, 257–260. Doi:10.1109/FIT.2012.53

- Langheinrich, M. (2001). Privacy by design: Principles of privacy-aware ubiquitous systems. In *Proceedings of the 3rd International Conference on Ubiquitous Computing* (pp. 273-291). London: Springer-Verlag. doi:10.1007/3-540-45427-6_23
- National Institute of Standards and Technology. (2010). *Introduction to NISTIR 7628: Guidelines for smart grid cyber security*. Gaithersburg, MD: NIST.
- Neumann, P. G., & Weinstein, L. (2006). Risks of RFID. *Communications of the ACM*, 49(5), 136. doi:10.1145/1125944.1125971
- Pew Research Center for the People and the Press. (2013). *Few see adequate limits on NSA surveillance program*. Retrieved from <http://www.people-press.org/files/legacy-pdf/7-26-2013%20NSA%20release.pdf>
- Schwartz, P. M., & Solove, D. (2011). The PII problem: Privacy and a new concept of personally identifiable information. *New York University Law Review*, 86, 1814–1894.
- Turow, J. (2012). *The daily you: How the new advertising industry is defining your identity and worth*. New Haven, CT: Yale University Press.
- Uckelmann, D., & Harrison, M. (2010). Integrated billing mechanisms in the Internet of Things to support information sharing and enable new business opportunities. *International Journal of RF Technologies: Research and Applications*, 2(2), 73–90.
- Uckelmann, D., Harrison, M., & Michahelles, F. (2010). An architectural approach towards the future Internet of Things. In D. Uckelmann et al. (Eds.), *Architecting the Internet of Things*. Berlin: Springer-Verlag Berlin Heidelberg.
- United Nations News Centre. (2013). *General Assembly backs right to privacy in digital age*. Retrieved from http://www.un.org/apps/news/story.asp?NewsID=46780&Cr=privacy&Cr1=#.UuW_5hB6dD8
- Upturn. (2014). *Civil rights, big data, and our algorithmic future*. Retrieved from <https://bigdata.fairness.io/>
- Vermesan, O., Friess, P., Guillemin, P., Gusmeroli, S., Sundmaeker, H., Bassi, A., & Doody, P. et al. (2011). Internet of Things strategic research roadmap. In O. Vermesan & P. Freiss (Eds.), *Global technological and societal trends from smart environments and spaces to green ICT* (pp. 9–52). Aalborg: River Publishers.
- Weber, R. H. (2010). Internet of Things: New security and privacy challenges. *Computer Law & Security Report*, 26(1), 23–30. doi:10.1016/j.clsr.2009.11.008
- Weber, R. H., & Weber, R. (2010). *Internet of Things: Legal perspectives*. Berlin: Springer-Verlag Berlin Heidelberg. doi:10.1007/978-3-642-11710-7
- Weiser, M. (1991). The computer for the twenty-first century. *Scientific American*, 265(3), 94–101. doi:10.1038/scientificamerican0991-94
- Winter, J. S. (2008). Emerging policy problems related to ubiquitous computing: Negotiating stakeholders visions of the future. *Knowledge, Technology & Policy*, 21(4), 191–203. doi:10.1007/s12130-008-9058-4
- Winter, J. S. (2013). Surveillance in ubiquitous network societies: Normative conflicts related to the consumer in-store supermarket experience in the context of the Internet of Things. *Ethics and Information Technology*. doi:10.1007/s10676-013-9332-3
- Winter, J. S. (2015). Algorithmic discrimination: Big data analytics and the future of the Internet. In J. S. Winter & R. Ono (Eds.), *The future Internet: Alternative visions*. New York: Springer. doi:10.1007/978-3-319-22994-2_8

KEY TERMS AND DEFINITIONS

Ambient Intelligence (AmI): Describes an environment in which many embedded, networked devices exist throughout the environment. Typically, these are tailored towards personal needs and aware of context. This research program is often associated with smart homes. Ambient Intelligence research has been guided by the European Union's Fifth Framework Programme (Information Society Technologies, 1998-2002).

Electronic Product Code (EPC): Is a standard that seeks to provide unique identification for RFID tags. It was originally created by MIT's Auto-ID Center and is currently directed by EPCglobal, an organization dedicated to the global standardization of EPC. The EPCglobal stack is the de facto standard for retail and consumer goods industries.

NFC (Near Field Communication): Is a collection of standards that enable the wireless exchange of data between mobile communication devices in close proximity. NFC is currently being used to exchange personal data and as in electronic payment systems.

RFID (Radio-Frequency Identification): Is the use of an object (typically referred to as an RFID tag) incorporated into a product for the purpose of identification and tracking using short-range wireless communication. RFID tags store information and can be read at short range with an RFID reader.

Ubiquitous Computing: Is a research paradigm that emerged from anthropological studies at Xerox Parc in the late 1980s. Marc Weiser proposed a human-centric vision of a many-to-one relationship between computers and humans.



The Protection Policy for Youth Online in Japan

Nagayuki Saito

Ochanomizu University, Japan

Madoka Aragaki

Business Breakthrough University, Japan

INTRODUCTION

In recent years, Internet use among young people has been associated to various social problems in many different countries. Examples include miscommunication by text, billing fraud, access to illegal content, and contact with ill-intentioned people. Especially in Japan, the use of smartphones has spread very rapidly among teenagers and young adults since 2012, generating tremendous changes in their online environment. These changes have triggered the abovementioned problems.

To tackle these problems, “the Act on the Development of an Environment that Provides Safe and Secure Internet Use for Young People” (Act No. 79 of 2008) was enforced in April 2009 in Japan. Because Article 3 defines the skills needed to use the Internet efficiently, it is important to empower teenagers and young people to develop risk management skills by using Internet effectively.

For this reason, it is crucial to optimize educational policy to meet the needs of young people. Doing so will require criteria on which to review the current policy. It is also important to evaluate the Internet literacy of teenagers and young people, and to reform educational policy and its implementation to reflect the results of this assessment.

This study aims to develop ILAS, the Internet Literacy Assessment Indicator for Students, making it a more effective and visible tool for developing young people’s coping skills, reducing their online risks, and enabling them to use the Internet

more safely. This indicator will be evidence based and designed to optimize educational policy; it can play an important role as a decision-making system for designing effective educational policy.

BACKGROUND

Review of Evidence Based Policy Making

OECD (2012a) advised all stakeholders to reduce online risks and provide a safer Internet environment. This recommendation obliges every stakeholder to provide a safer online environment for teenagers and young people. To provide effective protection, it is important to implement a youth protection policy at every level of government, as well as in the private sector and educational organizations. Without clear role definitions, it will be difficult to implement a concrete protection policy.

The most effective way to solve these problems is to think about each problem separately, clarifying the political tasks each sector should deal with. One tactic that can help to achieve this is to adopt an Evidence Based Policy (EBP).

An EBP is an approach derived from Evidence Based Medicine; it was proposed by Gordon Guyatt at Manchester University in Canada (Tsutani 2000). EBP is used in areas such as social policy, educational policy, and welfare policy (Sowaki 2010). The OECD (2007) has argued

that EBP-based policy making enables people and organizations to choose clear and simple evidence from among many options. EBP has been widely adopted in various policy areas for evidence based policy making.

Nishimura (2005) pointed out that evidence should be based on “objective and politically neutral statistical indicators.” Such evidence would gain public understanding and help to establish trust between government and society (OECD 2004). In addition, the OECD (2012) has emphasized the need to set indicators as metrics of the evidence, allowing people to visualize the actual condition of each political area.

From these discussions, it seems clear that EBM can be effective in supporting rational decision making for effective educational policy implementation. One key measure to promote the policy will involve establishing an indicator to evaluate the evidence.

Review of the Indicators Adopted in Each Educational Policy Area

In reviewing previous studies related to EBP, this section will focus on studies carried out at the level of government. Examples include the “Flash Eurobarometer” implemented for EU member countries and the “Fact-finding Survey on Young People’s Online Usage Environment” carried out by the Japanese Cabinet Office. In addition, this study will clarify the differences between these earlier studies and ILAS, touching on the social and academic impact of this study.

1. Review of the International Situation

The EU has been a pioneer in conducting research on this problem, launching the “Safer Internet Program,” which conducted actual condition surveys on Internet use in member countries. The EU also carried out quantitative research on children in 2003 and 2004. In 2005, the European Commission conducted a face-to-face questionnaire survey of children from member countries (European

Commission 2005); in 2007, the group interview method was used (European Commission 2007). In the following year, the EC conducted an actual condition survey of children’s Internet use and guardian control policies (European Commission 2008).

These studies focused on children’s actual usage of the Internet, geographic features, psychological conditions, parental controls, and educational policies at home. Their results offered insight into the Safer Internet Program and provided basic data to both InSafe, an international organization that advocates information literacy and morals, and INHOPE, an international hotline against harmful content (European Commission 2009).

In the UK, former Prime Minister Tony Blair appointed Dr. Tanya Byron (2008) to study children’s Internet use and address the emerging related social problems. Her results produced detailed policy proposals aimed at clarifying the different roles of the government and private sector. Following Byron’s review, the Office of Communications released “Ofcom’s Response to the Byron Review” in March 2008, highlighting the need to enhance the media literacy of children and their guardians as well as to implement self-regulation by industry groups. This report also identified the regulation of content delivery and user access as crucial methods for the self-regulation of industry groups (Office of Communications 2008).

In the United States, additional surveys were conducted by private and non-political organizations and think tanks. For example, the Family Online Safety Institute (2013), which advocates online safety for children and their families, conducted qualitative and quantitative studies on 558 children between the ages of 13–17, living in the US. Their results showed that the number of children connecting online through mobile phones increased from 43% in 2012 to 64% in 2013. The Pew Research Center (2012) researched the experience of parental control, surveying 802 guardians of children aged 12–17. Their results revealed that women (51% vs. 49% of men) under 40 years old



(54% vs. 48% over 40), who were white (59% vs. 31% African American) and had higher household incomes were the main users of parental control methods. In addition, the Cyberbullying Research Center (2015) reported the results of a survey on the impact of cyberbullying on the lives of 15,000 American children since 2002. The Center's results suggested that the number of children exposed to cyberbullying had increased from 18.8% in 2007 to 36.6% in 2014.

The Organization for Economic Cooperation and Development (OECD) has emphasized the importance of evidence based policymaking. The OECD Council adopted a "Recommendation on the Protection of Children Online" in February 2012, aiming to establish a foundation to promote co-regulation by each stakeholder for efficient policy implementation to address the identified problems (OECD 2012). This recommendation focused on the need for an evidence based policymaking approach for managing complex policy through enhanced policy coordination. For example, it suggested that enhanced Internet literacy among children and their guardians would be a viable solution, to be determined by evaluating actual usage situations.

Saito (2015) reported the results of an Internet Literacy Assessment Indicator for Students (ILAS), developed through a collaborative effort by the OECD Science, Technology and Industry Directorate and the Japanese Ministry of Internal Affairs and Communications; this project targeted adolescents, measuring their capacity to use the Internet safely and securely. This report investigated the differences in the ILAS test score of students who had bad experiences online and those who had not. It found that the percentage of questions answered correctly by students who had bad experiences online was higher than that of those who had no bad experiences (71% vs. 65%). Thus, adolescents can develop online literacy by coping with bad experiences through Internet usage. These reports recommended that the policy to be implemented not be too strict (e.g., an online ban). It recommended that policies worldwide

should support adolescents in developing online literacy through online experiences, while promoting filtering measures, to be implemented through a step-by-step approach.

2. Review of the Japanese Situation

In Japan, the Cabinet Office has conducted an annual "fact-finding survey on young people's online usage environment" to assess the online usage of teenagers and young adults, and their guardians' approach to protection since 2009. The main data gathered reveal hours of online use, geographical locations, and aspects of youth behavior. Guardians were asked about filtering and rules at home.

Article 3 of the "Act on Establishment of Enhanced Environment for Youth's Safe and Secure Internet Use" Appendix defines the role of the government in terms of conducting reviews: three years after implementation, appropriate measures will be summarized and undertaken. Following Article 3, a survey has been conducted to evaluate the improvement of the young people's online environment. Its main purpose was to gather basic data to evaluate the online use and behavior of teenagers and young adults, as well as filtering recognition, filtering usage, and the drawbacks of filtering.

The Importance of the Internet Literacy Assessment Indicator

As mentioned above, every governmental organization is constantly evaluating children's online usage. These results reveal the actual situation, in regard to children's online usage, education, parental education, and views of parental control. By reflecting these results for educational policy making, it will be possible to establish a protection policy for children and their parents. Political costs will be kept to a minimum by appropriately implementing effective policies.

To enhance the effectiveness of EBP, "children's ability to avoid online risks" should be

evaluated, even though this factor was not included in the EU's, "Flash Eurobarometer" or the Japanese Cabinet Office research. In this section, it will be argued that problems can be tackled by evaluating "children's ability to avoid online risks."

1. Adjusting the Strength of Child Protection Measures

By evaluating children's ability to avoid risks, it is possible to adjust the strength of technical measures for child protection, such as filtering. When filtering became obligatory, the Ministry of Internal Affairs and Communications working team focusing on illegal and harmful online content argued that there were three major problems.

The first is the uniformity problem: the question of whether uniform filtering is appropriate, given the wide range of risk competency among children of elementary school to high school age. The second problem involved filtering technology. Under the white list algorithm, many harmless websites are excluded; only official websites can be accessed. On the other hand, a black list algorithm blocks all websites in certain categories, regardless of whether these websites are harmful or not. The last problem involves convenience: whether controlling access to community websites that are essential to children's social life is appropriate or not (Ministry of Internal Affairs and Communications 2008).

By evaluating children's ability to avoid risks, it will be possible to adjust the strength of technical measures for child protection to reflect a child's developmental status and computer literacy level. This adjustment will not only optimize the level of online protection, but will also ensure that children have as much freedom in their online activities as possible.

2. Optimization of the Awareness Educational Policy

By evaluating the children's ability to avoid risk, it will be possible to optimize the content of awareness education. Awareness education has

depended on the discretion of educators and a limited discussion about whether to include or make it a priority to teach children about online risks.

The indicator will make visible every risk avoidance ability that children should acquire; as a result, educational content will be optimized. The optimization of educational content can and should also be customized to reflect children's skills.

Previous research results were not able to offer enough adjustment for child protection. For example, earlier research evaluated the number and type of educational experiences, but failed to assess how much online risk avoidance ability was acquired. It was difficult to assess child protection policies, because it was impossible to evaluate which types of ability were inadequate.

If online risk avoidance ability is evaluated, it will be possible for education to focus on building the abilities that children lack. Moreover, by knowing which ability is insufficient, it will be possible to change the specifications of online services for children, and to review several policies based on current laws, such as the "Act for the Establishment of an Enhanced Environment for Young People's Safe and Secure Internet Use."

3. Policy Making that Reflects Personal and Regional Backgrounds will be Possible Through this Analysis

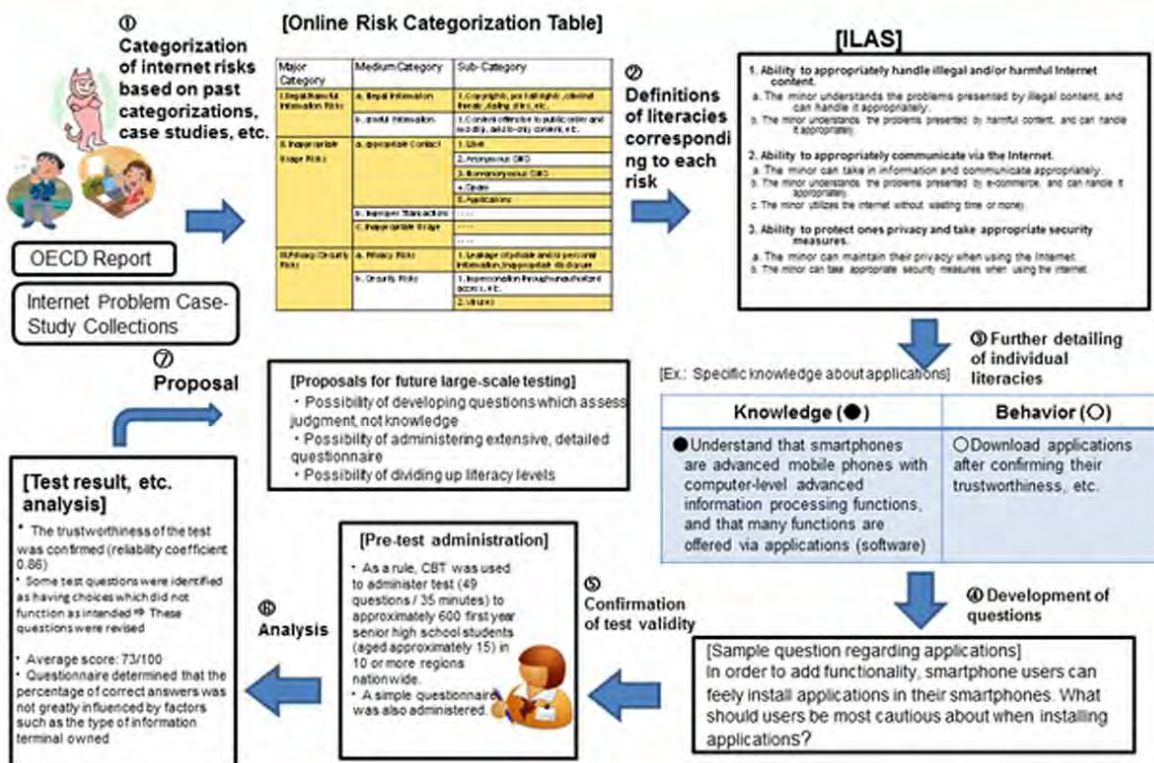
Although this study has aimed to establish indicators that can evaluate the online risk avoidance ability of children, it will also help to design protection policy that reflects personal and regional backgrounds; it can achieve this by analyzing the relationship between children's test scores and background information. For example, by analyzing the relationship between risk avoidance ability test scores and regional data, it will be possible to show children's average risk avoidance ability in each region. This will make it possible to establish child protection policies that reflect regional differences. Implementing customized awareness education in each region will enhance children's risk avoidance ability and lead to both a regional and national enhancement of skills.



Figure 1. Overview of the Internet Literacy Assessment Indicator for Students

Source: National Institute for Information and Communications Policy (2012a)

Overview of This Initiative



SOLUTIONS AND RECOMMENDATIONS

The Development and Implementation of ILAS by the Japanese Government

These discussions show that evaluating young people's online literacy and reforming and implementing educational policy that reflects these results will optimize the educational policies of all stakeholders for each childhood developmental phase. It is therefore important to develop Internet literacy indicators based on previous EBP to evaluate children's risk avoidance ability. By using risk avoidance ability as evidence, it will enable to offer appropriate protection to children, while ensuring their online freedom.

In Japan, the Ministry of Internal Affairs and Communications has developed the Internet Literacy Assessment Indicator for Students (ILAS) in 2012 in order to implement the OECD recommendations at governmental level; ILAS aims to measure children's online literacy skills and to evaluate their ability to cope with typical online risks.

ILAS systematically defined "the ability all children should acquire against online risks" (see Table 1). It is the indicator used to evaluate these abilities by means of a test (the National Institute for Information and Communications Policy 2012a). It is similar to the OECD's PISA test, which evaluates the academic skills of 15-year-old students; however, ILAS evaluates online risk coping skills. Like the PISA test, ILAS evaluates children's ability to cope with online risks and



Table 1. Classification table of online risks

Main Category	Significant Category	Sub-Category
I Illegal & Harmful Information Risks	A Illegal Information	1 Copyright, portrait rights, criminal threats, dating sites, etc.
	B Harmful Information	1 Content offensive to public order and morality, adults-only content, etc.
II Inappropriate Usage Risks	A Inappropriate Contact	1 Libel
		2 Anonymous SNS
		3 Non-anonymous SNS
		4 Spam
		5 Applications
	B Improper Transactions	1 Fraud, sale of improper products, etc.
	C Inappropriate Usage	1 Excess Internet consumption
		2 Dependence
III Privacy & Security Risks	A Privacy Risks	1 Leakage of private and/or personal information, inappropriate disclosure
	B Security Risks	1 Impersonation through unauthorized access, etc.
		2 Viruses

Source: National Institute for Information and Communications Policy (2012a)

visually represents its results by area. It also provides “criteria for judgement” for governmental and regional decision-making, online companies’ social responsibility, the activities of child protection organizations, and school decision-making.

In 2012, a pre-test was given to 2,464 10th grade students from 23 high schools to evaluate online literacy in Japan. ILAS was used to evaluate online risk avoidance ability and another questionnaires used to assess awareness of cell phone risks. By analyzing these indicators, researchers tried to evaluate the relationship between risk avoidance ability and cellphone use (see Table 2).

Visualization of Children’s Online Literacy by ILAS

In this section, the data collected from 2,464 students through the ILAS pre-test will be discussed.

1. Literacy in Each Risk Category

When analyzing online literacy as an ILAS risk category, the percentages of questions answered correctly in the sections on “irregular trading risks

Table 2. Overview of the pre-test

Time Required	50 Minutes
Number of questions	49 items (multiple-choice)
Contents	Description of the CBT: 10 minutes Pre-questionnaire: 5 minutes Implementation of the Pre-Test: 35 minutes Post questionnaire: 5 minutes
Investigation period	May 30, 2012 to July 30, 2012
Subjects	2,464 first year high school students
Number of schools	23 Schools

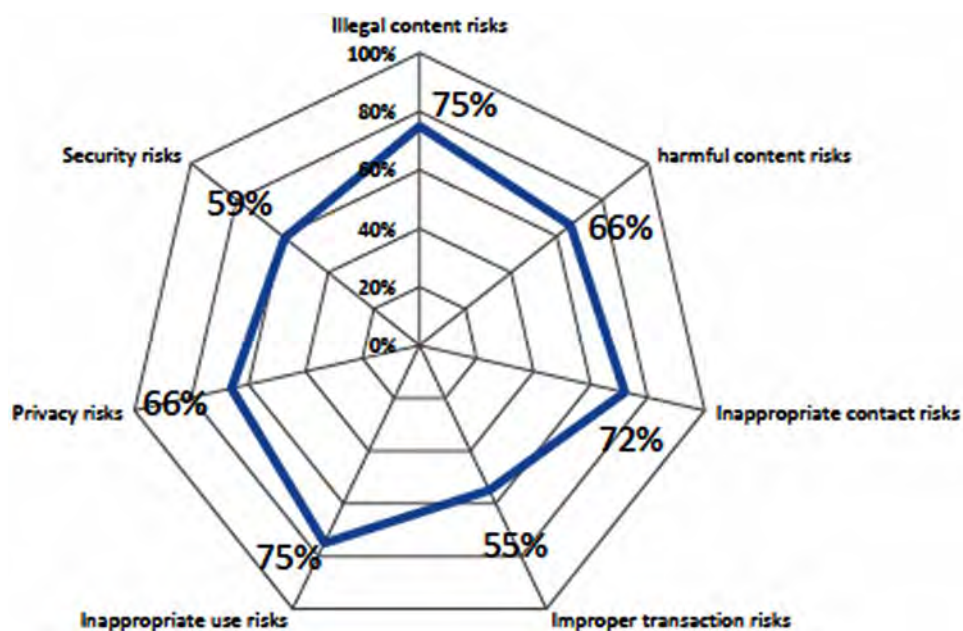
Source: National Institute for Information and Communications Policy (2012b)

(55.09%)” and “security risks (59.48%)” were lower than other categories of risk (see Figure 2).

When asked about “irregular trading risks,” it is likely that 15-year-old students will have had limited experience of online trading. Although it is possible to enhance their literacy through online trading, it is important to teach them the risks of online trading before they engage in it, to enhance their risk competency and avoid actual risk.

Next, “security risks” may be a challenge for students who are not familiar with technical

Figure 2. Average percentages of questions answered correctly in each risk category (n=2,464)
Source: Saito (2015)



subjects because they involve technology. It is important to ensure that students have a chance to learn about security, and teaching them in small steps to avoid learning failure. These things are important for developing the awareness education plan.

When thinking about the policy-making strategy of each stakeholder, governmental and regional organizations with an obligation to provide social education and online companies with social responsibility obligations must review

their educational programs to make sure that they include opportunities to learn about “irregular trading risks” and “security risks.”

Educators and parent-teacher associations should establish systematic teaching bodies and enhance their teaching ability. Parents should learn about these risk areas and should enhance their ability to protect and control their children (see Table 3).

2. Literacy and the Population of Cities

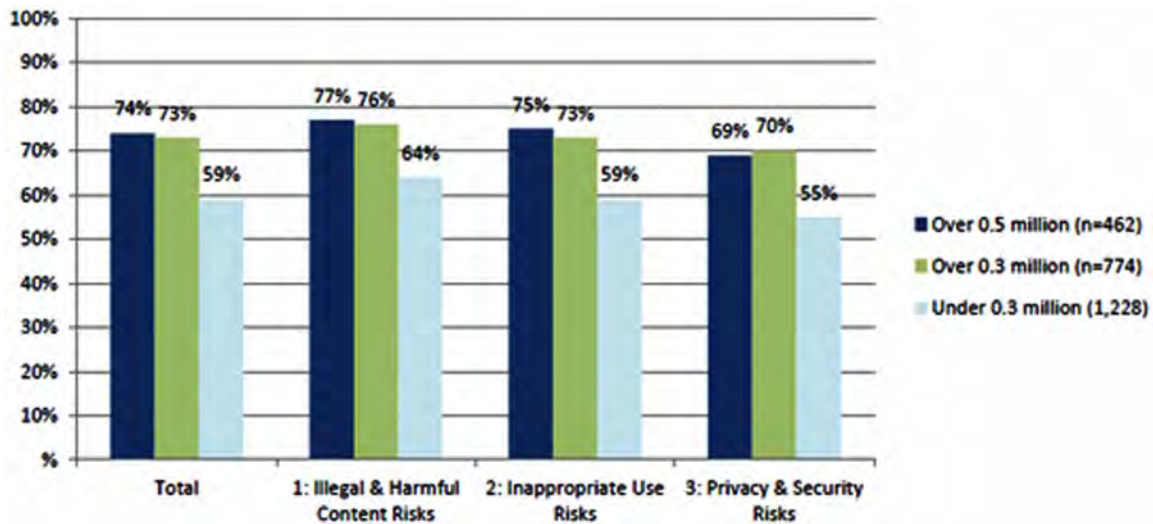
Next are the results about the relationship between children’s online literacy and the population of their cities. These showed that the percentage of questions answered correctly by students from cities with populations of less than 300,000 were low in 3 risk category areas. In addition, the total percentage of questions answered correctly was 74% in cities of than 500,000; 73% in cities of more than 300,000 people; and 59% in cities of less than 300,000 people ($F(2,461) = 192.06$, $p < 0.001$). Thus, as population declined, the percentage of questions answered correctly also declined (see Figure 3).

Table 3. Political challenges of awareness education at each layer

Stakeholder	Political Challenge
Governmental and regional organizations	Enhancement of awareness education relating to “irregular trading risks” and “security risks”
Educators and parent-teacher associations	Establish systematic teaching bodies and enhance their ability to teach students about “irregular trading risks” and “security risks.”
Parents	Learn about “irregular trading risks” and “security risks” and enhance their ability to protect and control their children.

Figure 3. The relationship between the percentage of questions answered correctly by students and the population size of their cities (n=2,464)

Source: Saito (2015)



These results may reflect the fact that larger cities offer more access to communication services, promoting frequent communication between students in daily life enhancing their online literacy.

Results suggest that online literacy differs, depending on the city population, and that smaller cities need to be aware of the potential for lower levels of literacy. The lower literacy rates in small cities may reflect several factors, including the real world communication divide, and a lack of social capital when acquiring online information.

When thinking about the implementation of awareness education, small cities tend to have more limited social educational networks than larger cities. For example, larger cities have organized bodies, which host social education events, such as workshops and symposiums—regularly providing awareness education. Smaller cities have fewer such organizations and face difficulties in providing awareness education. Although online networks reach every city, regardless of offline networks, the same set of online risks exists for all children in all cities, regardless of size. It is therefore important to provide awareness education throughout the country.

To tackle this problem, government should collaborate with local authorities and regional branches to provide nation-wide awareness education. The private sector should communicate with its regional branches and provide awareness education activities as an aspect of corporate social responsibility.

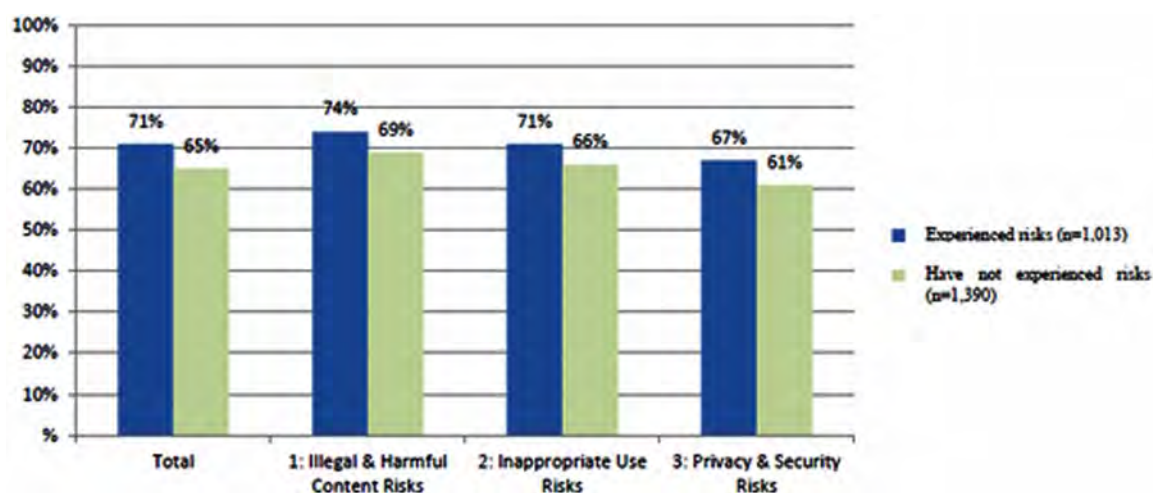
In addition, educational committees and parent-teacher associations need to hold workshops to enhance their teaching ability in a systematic

Table 4. Challenges related to regional differences in online literacy

Stakeholder	Political Challenge
Governmental and regional organizations	- Government should collaborate with local authorities to provide nation-wide awareness education - Private sector should communicate with its regional branches and provide awareness education locally
Educators and parent-teacher associations	Hold local workshops to enhance their teaching ability in a systematic way
Parents	Attend local workshops actively and enhance their protection and control ability

Figure 4. Cross analysis of the percentage of questions answered correctly by students with bad online experiences and students without such experiences (n=2,403)

Source: Saito (2015)



way. Parents should actively attend local workshops and enhance their protection and control ability (see Table 4).

3. Bad Online Experiences and Online Literacy

Next, the percentage of questions answered correctly by students with bad online experiences, and students without such experiences were compared. Results showed that the percentage of questions answered correctly by students with bad online experiences was higher than that of students without such experiences (71% vs 65%, $t(2,235.53) = 8.07, p < 0.001$) (see Figure 4). There was a particularly large difference in three areas of privacy security risks (67% vs 61%). The results suggested that students gained online literacy skills by dealing with bad online experiences.

Results showed that students who had had bad online experiences had higher levels of online literacy than students who had not had such experiences. This suggests that too much regulation of Internet use could have a negative effect on students' ability to learn online literacy. Thus, children's online freedom and protection should be balanced in any child protection initiative (OECD 2012b).

To deal with such problems, government should support appropriate technical regulation and private activities that aim to establish social systems. It should also provide many opportunities for awareness education so that children can gain risk avoidance ability. To promote effective policy

Table 5. Challenges resulting from too much online regulation

Stakeholder	Political Challenge
Governmental and regional organizations	- Government should support appropriate technical regulation and private activities aiming to establish social systems. Also, government should provide wide opportunities for awareness education to children, so that they can acquire risk avoidance ability. In addition, the government should conduct continual research on children's online literacy. - The private sector should work on appropriate self-regulation and promote awareness education so that children themselves can gain risk avoidance ability.
Educators and parent-teacher associations	Communicate with homes and advocate continuously for awareness raising.
Parents	Enhance their own protection and control ability at home, and help their children gain online literacy.

implementation, the government should also continual research on children's online literacy.

The private sector should work on appropriate self-regulation through cooperation with government, and should promote awareness education so that children themselves can gain risk avoidance ability.

In addition, schools and parent-teacher associations should avoid banning online activities for children. Instead, they should communicate with homes and advocate continuously for awareness raising. Parents need to enhance their protection and control ability at home, and help their children can gain online literacy (see Table 5).

FUTURE RESEARCH DIRECTIONS

This study showed the future direction of policymaking for each layer of stakeholders working for youth protection based on ILAS analysis. The government, local authorities, and private companies form the macro-layer; schools and parental organizations constitute the meso-layer; and parents form the micro-layer.

Further research is needed on the problem, including study of "irregular trading risk" and "security risk," which were pointed out by this study. In addition, the problem of reaching out to small cities to provide awareness education has room for further inquiry, such as (1) evaluation of the implementation of child protection policy by each stakeholder and (2) analysis of the cooperation and synergy between each stakeholder.

CONCLUSION

This study developed an appropriate test to evaluate children's ability to use the Internet safely and securely. It discussed the problems involved in promoting awareness education in weaker areas, such as "irregular trading risk" and "security risk." It also discussed regional differences in online literacy and the importance of helping children to

acquire risk avoidance ability through their own efforts. To tackle these problems, this study has proposed political strategies that each stakeholder (including government, local authorities, companies, schools, and parents) should use to promote child safety at each childhood developmental stage. Our proposals can support the decision-making of every stakeholder enhance the effectiveness of this policy.

The next step, following on from this study, is to implement the recommended political strategy for other stakeholders. This study provides basic evidence to support policy making, thus contributing to a safe and secure online environment for children and young people.

ACKNOWLEDGMENT

This study could not have been completed without the cooperation of the members of the Committee on the Internet Literacy Assessment Indicator for Students, Japan and the Ministry of Internal Affairs and Communications, Japan.

This research was partially supported by the Ministry of Education, Science, Sports and Culture, through a Grant-in-Aid for Scientific Research (C), 2014-2016 (26330389, Nagayuki Saito).

REFERENCES

- Byron, T. (2008). *Safer children in a digital world: The report of the Byron Review*. DCSF Publications. Retrieved September 10, 2016, from <http://webarchive.nationalarchives.gov.uk/20130401151715/http://www.education.gov.uk/publications/eOrderingDownload/DCSF-00334-2008.pdf>
- Cabinet Office. (2014). *Survey report on youth Internet environment*. Retrieved September 10, 2016, from the Cabinet Office homepage: <http://www8.cao.go.jp/youth/youth-harm/chousa/h24/net-rating/pdf-index.html>



- Cyberbullying Research Center. (2015). *Summary of Our Cyberbullying Research (2004-2015)*. Retrieved September 10, 2016, from <http://cyberbullying.org/summary-of-our-cyberbullying-research/>
- European Commission. (2005). *Special Eurobarometer: Safer Internet*. Retrieved September 10, 2016 from <http://polis.osce.org/library/f/3652/2821/EU-EU-RPT-3652-EN-2821>
- European Commission. (2007). *Qualitative Study: Safer Internet for Children*. Retrieved September 10, 2016, from http://ec.europa.eu/public_opinion/archives/quali/ql_safer_Internet_summary.pdf
- European Commission. (2008). *Flash Eurobarometer: Towards a Safer Use of the Internet for Children in the EU—A Parents' Perspective Analytical Report*. Retrieved September 10, 2016 from http://ec.europa.eu/public_opinion/flash/fl_248_en.pdf
- European Commission. (2009). *Safer Internet Programme 2005–2008. (Safer Internet Plus)*. Retrieved September 10, 2016, from http://europa.eu/legislation_summaries/information_society/Internet/l24190b_en.htm
- Family Online Safety Institute. (2013). *Teen Identity Theft, Fraud, Security, and Steps Teens are Taking to Protect Themselves Online*. Retrieved September 10, 2016 from file:///C:/Users/nasaito/Downloads/Teen_Identity_Theft_Full_Report_Nov_13.pdf
- National Institute for Information and Communications Policy. (2012a). 「青少年のインターネット・リテラシー指標(指標開発編)」[Internet Literacy Indicator for Youth—Development of Indicators]. NICT. Retrieved September 10, 2016 from www.soumu.go.jp/iicp/chousakenkyu/data/research/survey/telecom/2012/ilas2012-report-build.pdf
- National Institute for Information and Communications Policy. (2012b). 「青少年のインターネット・リテラシー指標(実態調査編)」[Internet Literacy Indicator for Youth—Investigation of actual conditions]. NICT. Retrieved September 10, 2016, from www.soumu.go.jp/iicp/chousakenkyu/data/research/survey/telecom/2012/ilas2012-report-survey.pdf
- Nishimura, K. (2005). The Need for Evidence-Based Policy. *Economic Review (Federal Reserve Bank of Atlanta)*, 1, 4–7.
- OECD. (2004). *Statics, Knowledge and Policy: OECD World Forum on Key Indicators*. Retrieved September 10, 2016 from <http://www.oecd.org/site/worldforum06/36422528.pdf>
- OECD. (2007). *Knowledge Management: Evidence in Education—Linking Research and Policy*. Paris: OECD Publishing.
- OECD. (2012). *The Protection of Children Online: Risks Faced by Children Online and Policies to Protect Them*. Paris: OECD Publishing.
- Pew Research Center. (2012). Parents, Teens, and Online Privacy. The Berkman Center for Internet & Society at Harvard University.
- Saito, N. (2015). *Internet Literacy in Japan*. Paris: OECD Publishing. doi:10.1787/18151965
- Saito, N., & Aragaki, M. (2012). *Analysis on political direction of non-formal parental education to develop a safer Internet environment for adolescents* (Vol. 23). Tokyo: Japan Society of Research and Information on Public and Co-operative Economy.
- Sowaki, H. (2010). The Development of an Evidence-Based Education Policy in England. *National Institute for Educational Policy Research*, 139, 153–168.

Tsutani, K. (2000). The Cochrane Collaboration and Systematic Review—Its Role in the EBM Movement. *Journal of the National Institute of Public Health*, 49(4), 313–319.

ADDITIONAL READING

Byron, T. (2010). *Do We Have Safer Children in a Digital World?—A Review of Progress Since the 2008 Byron Review*. Available at <http://media.education.gov.uk/assets/files/pdf/d/do%20we%20have%20safer%20children%20in%20a%20digital%20world%202010%20byron%20review.pdf>

Japan Internet Safety Promotion Association. (2015). *ILAS2015 Final Report*. Tokyo: Japan Internet Safety Promotion Association.

Livingstone, S., Haddon, L. G. A., & Ólafsson, K. (2011). Risks and safety on the Internet: the perspective of European children: full findings and policy implications from the EU Kids Online survey of 9–16 year olds and their parents in 25 countries, EU Kids Online, Deliverable D4, EU Kids Online Network, London, UK.

Ministry of Internal Affairs and Communications. (2009). *Final report of Study Group on Measures to Illegal and Harmful Information on the Internet: Internet Safety Promotion Program*. Tokyo: Ministry of Internal Affairs and Communications.

Ofcom (2008). *Ofcom's Response to the Byron Review*, Office of Communications, http://stakeholders.ofcom.org.uk/binaries/research/telecoms-research/Byron_exec_summary.pdf

Saito, N., & Aragaki, M. (2011). An experiment in collaborative education to enhance adolescents' respect for social norms in Internet use. *Journal of the Japan Information-Culture Society*, 18(2), 60–67.

Saito, N., Tanaka, E., & Yatuzuka, E. (2014). Evolving Challenges to the Development and Assessment of Information Literacy Education for Online Safety in Japan. *Journal of Cases on Information Technology*, 16, 21–44.

Saito, N., Yoshida, T., & Akahori, K. (2012). Development and Evaluation of the Internet Literacy Assessment Indicator for Students, *Research Report of JSET Conferences*, 2012 (3), pp. 4–50, Tokyo.

Tanaka, E., Saito, N., & Yatuzuka, E. (2011). *A Usage Analysis of the Mobile Literacy e-Learning Program "Mobami" Regarding Access to Content and Rulemaking Tendencies*, World Conference on e-Learning in Corporate, Government, Healthcare, and Higher Education 2011, Honolulu, Hawaii: AACE, pp. 436–441, USA.

Tapscott, D. (2008). *Grown up digital: How the net generation is changing your world*. New York, NY, USA: McGraw-Hill.

KEY TERMS AND DEFINITIONS

Awareness Education: Education designed to enhance knowledge/attitudes that can help to avoid various online risks ensure appropriate Internet use.

Evidence Based Policy: A policy making method based on actual proof. It was first advocated as Evidence Based Medicine by Gordon Guyatt, University of Manchester, Canada in 1991. It has subsequently been used widely in social policy, educational policy, and social welfare policy.

Internet Literacy Assessment Indicator for Students (ILAS): An indicator developed to evaluate online literacy, especially the effectiveness of coping skills/morals in reducing online risks/threats. It was released by the Ministry of Internal Affairs and Communications in September, 2012.

Online Risks: Inclusive term covering various online risks, including cyberbullying, online addiction, cyber grooming, security problems, and the leakage of personal information.

Risk Categories: Classification of online risks defined by ILAS. ILAS classifies 3 major

online risks as Illegal & Harmful Information Risks, Inappropriate Usage Risks, and Privacy & Security Risks. It also defines 7 significant categories and 13 sub-categories. In addition, it defines 186 coping skills, which can be used to protect against online risks.

Security of Identity-Based Encryption Algorithms



Kannan Balasubramanian

Mepco Schlenk Engineering College, India

M. Rajakani

Mepco Schlenk Engineering College, India

INTRODUCTION

The concept of Identity Based Cryptography was proposed in (Shamir, A., 1984) which introduced the idea of using arbitrary strings such as e-mail addresses and IP Addresses to form public keys with the corresponding private keys being created by the Trusted Authority (TA) who is in possession of a system-wide master secret (Srinivasan, S., 2010). Then a party, Alice who wants to send encrypted communication to Bob need only Bob's identifier and the system-wide public parameters. Thus the receiver is able to choose and manipulate the public key of the intended recipient which has a number of advantages. While Identity Based Cryptography (IBC) removes the problem of trust in the public key, it introduces trust in the TA. As the TA uses the system-wide master secret to compute private keys for users in the system, it can effectively recompute a private key for any arbitrary string without having to archive private keys. This greatly simplifies key management as the TA simply needs to protect its master secret.

Some of the earlier Identity Based Cryptosystems proposed such as the one by Cocks (Cocks, C., 2010) and Boneh (Boneh, et.al., 2007) were not based on mathematics of pairings. The Identity based cryptosystem (the term Identity Based Cryptography refers to this set of algorithms whereas the term Identity Based Cryptosystem refers to a specific algorithm) was introduced by Boneh and Franklin (Boneh, et.al., 2001). An Identity Based Encryption or IBE (the term IBE is used to denote

a specific Identity Based Cryptosystem) scheme has the following four algorithms: Setup, KeyDer, Enc and Dec. This chapter discusses the algorithms of the IBE schemes and compares them based on the implementation efficiency. An extension to the basic IBE scheme is the Hierarchical IBE proposed by Horwitz and Lynn (Horwitz, et.al., 2001).

In contrast to the basic standard model of IBE, a Random Oracle Model (Bellare, et.al., 1993) may be used where proofs of security are obtained by replacing hash functions with "Random Oracles" that output truly random values for every distinct output. This chapter discusses IBE schemes based on the Random Oracle Model IBEs and compares them with the standard model IBE.

An extension of the above schemes with multiple Trusted Authorities (TAs) instead of a single TA is also possible. An architecture for the implementation of the IBE is discussed along with the security of the various schemes.

BACKGROUND

The public key encryption is a cryptographic system that uses two keys -- a *public key* known to everyone and a *private* or *secret key* known only to the recipient of the message. When user Alice wants to send a secure message to user Bob, she uses Bob's public key to encrypt the message, Bob then uses his private key to decrypt it. An important element to the public key system is that the public and private keys are related in

such a way that only the public key can be used to encrypt messages and only the corresponding private key can be used to decrypt them. Moreover, it is virtually impossible to deduce the private key if you know the public key. Users will exchange public keys; this transaction does not need to be done in a secure manner because the release of public keys does not threaten the security of any private information. After this swap, someone who wishes to send private information to another user will encrypt the data with the intended recipient's public key and then pass along the encrypted message. The recipient, who will keep his or her private key secure under any circumstance, can use the private key to decrypt the encoded message.

Keys in public-key cryptography, due to their unique nature, are more computationally costly than their counterparts in secret-key cryptography. Asymmetric keys must be many times longer than keys in secret-cryptography in order to boast equivalent security. Keys in asymmetric cryptography are also more vulnerable to brute force attacks than in secret-key cryptography (Halevi, S. et.al., 1987). Public-key cryptography also has vulnerabilities to attacks such as the man in the middle attack. In this situation, a malicious third party intercepts a public key on its way to one of the parties involved. The third party can then instead pass along his or her own public key with a message claiming to be from the original sender. An attacker can use this process at every step of an exchange in order to successfully impersonate each member of the conversation without any other parties having knowledge of this deception. In order to tackle the issues surrounding the generation, distribution and safekeeping of the private and public keys and also simplify the process of obtaining the public keys, the identity based Encryption was invented.

Many different Identity Based Cryptosystems have been proposed. Some of them use the concept of a Random Oracle. A popular methodology for designing cryptographic protocols consists

of the following two steps. One first designs an ideal system in which all parties (including the adversary) have oracle access to a truly random function, and proves the security of this ideal system. Next, one replaces the random oracle by a "good cryptographic hashing function" (such as MD5 or SHA), providing all parties (including the adversary) with a succinct description of this function. Thus, one obtains an implementation of the ideal system in a "real-world" where random oracles do not exist. This methodology was formulated by Bellare and Rogaway (Bellare, M., et.al, 1993) and has been used in many works.

AN IBE SCHEME

An IBE scheme is defined in terms of four algorithms *Setup*, *KeyDer*, *Enc* and *Dec*:

- **Setup:** On input 1^k outputs a master public key mpk which includes system parameters $params$, and a master secret key msk . We assume that $params$ contains descriptions of the message and ciphertext spaces, $MsgSp$ and $CtSp$. This algorithm is randomized.
- **KeyDer:** A Key derivation algorithm that on input mpk and msk and identifier id , returns a private key usk_{id} . This algorithm may or may not be randomized.
- **Enc:** An encryption algorithm that on input mpk , identifier id and a message $m \in MsgSp$, returns a ciphertext $c \in CtSp$ and is written as $c = Enc(mpk, id, m)$. This algorithm is usually randomized and if randomness is emphasised it is written as $c = Enc(mpk, id, m; r)$.
- **Dec:** A decryption algorithm that on input mpk , a private key usk_{id} and a ciphertext $c \in CtSp$ returns either a message $m \in MsgSp$ or a failure symbol $\perp$.

PAIRINGS IN ELLIPTIC CURVE CRYPTOGRAPHY

Pairing based Cryptography which builds on the foundation of Elliptic Curve Cryptography has become an active area of research since the work of Boneh and Franklin (Boneh, D., et al.,). Pairing based cryptography is introduced here to show how Identity based Cryptosystems can be constructed.

Let q be a large prime and m an integer with $m \geq 1$. Let F_{q^m} be a finite field with q^m elements. Here q denotes the characteristic of the field and m the extension degree. $F_{q^m}^*$ denotes the multiplicative group of F_{q^m} . The Elliptic curve E over F_{q^m} is denoted by E / F_{q^m} and is defined to be the set of elements $(x, y) \in F_{q^m} \times F_{q^m}$ satisfying an equation of the form

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_5,$$

where $a_i \in F_{q^m}$ for $i=1,2,3,4,5$.

A point $P = (x, y) \in F_{q^m} \times F_{q^m}$ is said to be on the curve if it satisfies the above equation. $E(F_{q^m})$ represents the set of points on the curve and together with the point at infinity forms an additive Abelian Group.

Suppose that $E(F_{q^m})$ has cyclic subgroups G_1 and G_2 of prime order p . Let G_T be the cyclic subgroup of $F_{q^m}^*$ of prime order p . Then an admissible bilinear pairing is a function e which maps a pair of elliptic curve points in G to an element G_T ,

$$e : G_1 \times G_2 \rightarrow G_T,$$

having the following properties:

- **Bilinearity:** Let $P \in G_1, Q \in G_2$ and $a, b \in Z_p^*$, $e(aP, bQ) = e(P, Q)^{ab}$

- **Non-Degenerate:** $e(P, Q) \neq 1$
- **Efficiently Computable:** There must be an efficient algorithm that computes the map e for any pair of inputs.

A survey of applications of pairings in cryptography can be found in (Okamoto, T., 2006).

The mathematics behind pairing is quite involved and is usually treated as a black-box which can be used in the implementation of the security protocols.

THE BONEH-FRANKLIN SCHEME

The steps of the Boneh-Franklin scheme are as follows:

Setup

On input 1^k the algorithm works as follows:

- Runs a Pairing Generation algorithm to obtain (G, G_T, e, p, g) where $e: G \times G \rightarrow G_T$ is a bilinear map
- Sets $params = (G, G_T, e, p, g, H_1, H_2, l)$.
- Sets $h = g^s$ where s is selected from Z_p^*
- Sets $mpk = (params, h)$ and $msk = s$.

Here, $H_1: \{0,1\}^* \rightarrow G$, $H_2: G_T \rightarrow \{0,1\}^l$ for some $l=l(k)$, $MsgSp = \{0,1\}^l$, $CtSp = G \times \{0,1\}^l$.

KeyDer

On input mpk , msk and $id \in \{0,1\}^*$, sets the private key corresponding to $usk_{id} = H_1(id)^{msk}$

Enc

On input mpk , to encrypt a message $m \in MsgSp$, under the identifier $id \in \{0,1\}^*$,

- Parses mpk as $(params, h)$.



- Sets $T = e(H_1(id), h)^r$ where r is selected from Z_p^* .
- Outputs $c = (u, v) = (g^r, m \oplus H_2(T))$.

Dec

On input mpk , ciphertext c and the private key usk_{id} corresponding to identifier $id \in \{0,1\}^*$,

- Parses c as $(u, v) \in G \times \{0,1\}^1$,
- Sets $T = e(usk_{id}, u)$.
- Outputs $m = v \oplus H_2(T)$.

COCKS SCHEME

The Cocks Scheme (Cocks, C., 2001) of Identity Based Encryption is as follows:

Setup

The Trusted Authority gets the following inputs to generate a private key:

- An RSA modulus $n=pq$, where p, q are two private prime numbers which satisfy $p \equiv q \equiv 3 \pmod{4}$.
- A message space $MsgSp = \{-1, 1\}$ and a ciphertext space $CtSp = Z_n$.
- A secure common hash function $f: \{0,1\}^* \rightarrow Z_n$.

KeyDer

Input: Parameters generated by Setup and an arbitrary ID

Output: The private key r .

- Generate a which satisfies $\left(\frac{a}{p}\right) = 1$ with a deterministic procedure ID.
- Let $r = a^{\frac{n-5-p-q}{8}}$ which satisfies $r^2 = \pm a \pmod{n}$.

Encrypt

Input: Parameters generated by Setup, ID of the sender and message M .

Output: Corresponding ciphertext C .

- Select a random t which satisfies $m = \left(\frac{t}{p}\right)$, where m is an arbitrary bit of M .
- Let $c_1 = t + at^1 \pmod{n}$ and $c_2 = t - at^{-1}$
- Send $s = (c_1, c_2)$ to the recipient.

Decrypt

Input: Ciphertext C and the private key and parameters generated by the Trusted Authority.

Output: Original message M .

- Let $\alpha = c_1 + 2r$ if $r^2 = a$, otherwise $\alpha = c_2 + 2r$.
- Return $m = \left(\frac{\alpha}{n}\right)$.

AUTHENTICATED IBE

Here we discuss the authenticated IBE encryption.

Setup

1. Get a security parameter k and then generate a prime number q and Groups G_1 and G_2 of order q , and an admissible bilinear map $e: G_1 \times G_2 \rightarrow G_2$. Select a random generator $P \in G_1$.
2. Generate a random number $s \in Z_q^*$. Let $P_{pub} = sP$.
3. The Trusted Authority selects a random generator $g \in G_1$ and hash functions $H_1: F_q \times G_2 \rightarrow \{0,1\}^n$, $H_2: \{0,1\}^* \rightarrow G_1$, $H_3: \{0,1\}^* \times \{0,1\}^* \rightarrow F_q$, $H_4: \{0,1\}^n \rightarrow \{0,1\}^n$.
4. Return params $= \langle q, G_1, G_2, g, g^s, e, n, P, P_{pub}, H_1, H_2, H_3, H_4 \rangle$ and master key $= s$.

KeyDer

1. The Trusted Authority extracts the private key of user ID_A : $d_A = H_2(ID_A)^s$

Authenticated Encrypt

User A (ID_A) uses private key d_A and another user's id to encrypt a message $M \in \{0,1\}^*$

1. Select a random number $r \in \{0,1\}^n$
2. Let $c_1 = H_3(r, M)$ and $c_2 = e(d_A, H_2(ID_B))$.
3. Return the ciphertext $C = \langle r \otimes H_1(c_1, c_2), E_{H_4(r)}(M) \rangle$.

Authenticated Decrypt

User B uses A's ID ID_A , his private key d_B and params to decipher the ciphertext $\langle U, V, W \rangle$

1. Let $c_2 = e(H_2(ID_A), d_B)$.
2. Let $r = V \oplus H_1(U, c_2)$
3. Let $M = D_{H_4(r)}(W)$.
4. Compare U and $H_3(r, M)$.
5. If $U \neq H_3(r, M)$, discard the ciphertext, otherwise return the message M .

THE HIERARCHICAL IBE SCHEME

In this scheme, every user has an n -tuple ID in the hierarchy tree. The n -tuple ID is composed by the IDs of the user and his ancestors. All users in the i -th level are denoted by $Level_i$. The root of the hierarchy tree $Level_0$ is the trusted Authority.

The following five algorithms form the Hierarchical IBE scheme (Cheng, P., et al., 2012):

Root Setup

1. Based on the security parameter k , generate a big prime q .

2. Use q to generate two fields G_1 and G_2 which satisfy the bilinear map $e: G_1 \times G_1 \rightarrow G_2$
3. Pick an arbitrary element P_0 in G_1 and a random number S_0 in Z_q (q is non zero) as the master-key. Calculate the system parameter $Q_0 = S_0 P_0$.
4. Generate two hash functions $H_1: \{0,1\} \rightarrow G_1, H_2: G_2 \rightarrow \{0,1\}^n$.

Lower-Level Setup

For each user $E_i \in Level_i$, Specify a random number $s_i \in Z_q$

Extract

1. For each user E_i with $ID = (ID_1, ID_2, \dots, ID_i)$, its parent calculates $P_i = H_1(ID_1, ID_2, \dots, ID_i) \in G_1$, where s_0 is the identity of G_1 .
2. Return the private key $S_i = S_{i-1} P_i = \sum_{j=1}^i s_{i-j} P_j$ of E_i and parameter $Q_i = s_i P_0$

Encrypt

1. For a message M and $ID = \langle ID_1, ID_2, \dots, ID_i \rangle$, calculate
 - a. $P_i = H_1(ID_1, ID_2, \dots, ID_i) \in G_1$
2. For any $r \in Z_n$, return the ciphertext: $C = \langle rP_0, rP_2, \dots, rP_i, M \oplus H_2(g_r) \rangle$, $g = e(Q_0, P_1) \in G_2$

Decrypt

For ciphertext $C = \langle U_0, U_2, \dots, U_i, V \rangle$ and $ID = \langle ID_1, ID_2, \dots, ID_i \rangle$, return the message:

$$M = V \oplus H_2 \left(\frac{e(U_0, S_i)}{\prod_{i=2}^t e(Q_{i-1}, U_i)} \right)$$

THE ARCHITECTURE OF AN IBE SYSTEM

The RFC5408 document (<https://tools.ietf.org/html/rfc5408>) describes an architecture for the Identity Based Encryption system. A brief description is provided below:

Identity-based encryption (IBE) is a public-key encryption technology that allows a public key to be calculated from an identity and a set of public mathematical parameters and that allows for the corresponding private key to be calculated from an identity, a set of public mathematical parameters, and a domain-wide secret value. An IBE public key can be calculated by anyone who has the necessary public parameters; a cryptographic secret is needed to calculate an IBE private key, and the calculation can only be performed by a trusted server that has this secret.

The calculation of both the public and private keys in an IBE system can occur as needed, resulting in just-in-time creation of both public and private keys. The ability to calculate a recipient's public key, in particular, eliminates the need for the sender and receiver to interact with each other, either directly or through a proxy such as a directory server, before sending secure messages. A characteristic of IBE systems that differentiates them from other server-based cryptographic systems is that once a set of public parameters is fetched, encryption is possible with no further communication with a server during the validity period of the public parameters.

The server components required for an IBE system are the following:

- **A Public Parameter Server (PPS):** IBE public parameters include publicly-sharable cryptographic material, known as IBE public parameters, and policy information for an associated PKG. A PPS provides a well-known location for secure distribution of IBE public parameters and policy information that describe the operation of a PKG.
- **A Private-Key Generator (PKG):** The PKG stores and uses cryptographic material, known as a master secret, which is used for generating a user's IBE private key. A PKG accepts an IBE user's private key request, and after successfully authenticating them in some way, returns their IBE private key.

Sending a message that is IBE encrypted: In order to send an encrypted message, an IBE user must perform the following steps:

1. Obtain the recipient's public parameters

The public parameters of the recipient's system are needed to perform IBE operations. Once a user obtains these public parameters, he can perform IBE encryption operations. These public parameters may be available at a PPS that is operated by the user's organization, one that is operated by the sender's organization, or by a different organization entirely.

2. Construct and send an IBE-encrypted message

In addition to the IBE public parameters, all that is needed to construct an IBE-encrypted message is the recipient's identity, the form of which is defined by the public parameters. When this identity is the same as the identity that a message would be addressed to, then no more information is needed from a user to send them an encrypted message than is needed to send them an unencrypted message. This is one of the major benefits of an IBE-based secure messaging system. Examples of identities are individual, group, or role identifiers.

Receiving and Viewing an IBE-Encrypted Message

In order to read an IBE-encrypted message, a recipient of such a message parses it to find the URI (Uniform Resource Identifier) or IRI (In-

ternationalized Resource Identifier) he needs in order to obtain the IBE public parameters that are required to perform IBE calculations as well as to obtain a component of the identity that was used to encrypt the message. Next, the recipient carries out the following steps:

1. Obtain the IBE public parameters

An IBE system's public parameters allow it to uniquely create public and private keys. The recipient of an IBE-encrypted message can decrypt an IBE-encrypted message if he has both the IBE public parameters and the necessary IBE private key. The public parameters also provide the URI or IRI of the PKG where the recipient of an IBE-encrypted message can obtain the IBE private keys.

2. Obtain the IBE private key from the PKG

To decrypt an IBE-encrypted message, in addition to the IBE public parameters, the recipient needs to obtain the private key that corresponds to the public key that the sender used. The IBE private key is obtained after successfully authenticating to a private key generator (PKG), a trusted third party that calculates private keys for users. The recipient then receives the IBE private key over a secure connection.

3. Decrypt the IBE-encrypted Message

The IBE private key decrypts the CEK (Content-Encryption Key). The CEK is then used to decrypt the encrypted message.

It may be useful for a PKG to allow users other than the intended recipient to receive some IBE private keys. Giving a mail-filtering appliance permission to obtain IBE private keys on behalf of users, for example, can allow the appliance to decrypt and scan encrypted messages for viruses or other malicious features.

When requesting a private key, a client has to transmit three parameters:

1. The IBE algorithm for which the key is being requested
2. The identity for which it is requesting a key
3. Authentication credentials for the individual requesting the key

The identity for which a client requests a key may not necessarily be the same as the identity that the authentication credentials validate. This may happen, for example, when a single user has access to multiple aliases.

When following the above protocol, attacks may be possible on the system. The following attacks have been identified:

- Passively monitor information transmitted between users of an IBE system and the PPS and PKG
- Masquerade as a PPS or PKG
- Perform a denial-of-service (DoS) attack on a PPS or PKG
- Easily guess an IBE users authentication credential

Additional Information on the structure of the identity and the format of the request protocols can be obtained from RFC5408 (<https://tools.ietf.org/html/rfc5408>)

SECURITY NOTIONS AMONG IBE SCHEMES

Let $A = (A_1, A_2)$ be an adversary and A_1 and A_2 are both Probabilistic Polynomial time algorithms. At the first stage, given the system parameters, the adversary computes and outputs a challenge template τ . A_1 can output some state information s which can be transferred to A_2 . At the second stage the adversary is issued a challenge ciphertext



y^* generated from τ by a probabilistic function in a manner depending on the goal. We say the adversary successfully breaks the scheme if she achieves her goal (<https://eprint.iacr.org/2005/253.pdf>).

Four security goals are considered: One-wayness, Indistinguishability, Semantic Security and non-Malleability.

- **One-Wayness:** Here one-wayness is defined using a two-stage experiment: A_1 is run on the system parameters $param$ as input. At the end of A_1 's execution the adversary outputs $(s; id)$, such that s is state information (possibly including $param$) which she wants to preserve, and id is the public key which she wants to attack. One plaintext x is randomly selected from the message space M beyond adversary's view. A challenge y is computed by encrypting x with the public key id . A_2 tries to compute what x was.
- **Indistinguishability:** In this scenario A_1 is run on $param$, and outputs $(x_0; x_1; s; id)$, such that x_0 and x_1 are plaintexts with the same length. One of x_0 and x_1 is randomly selected, say x_b , beyond adversary's view. A challenge y^* is computed by encrypting x_b with id . A_2 tries to distinguish whether y^* was the encryption of x_0 or x_1 .
- **Semantic Security:** In this scenario, A_1 is given $param$ and outputs (M', hf, s, id) . Here the distribution of M' is designated by A_1 and (M', hf) is the challenge template τ . A_2 receives an encryption y^* of a random message x^* drawn from M' . The adversary then outputs a value v . The adversary hopes that $v = f(x^*)$. The adversary is successful if this can be done with a probability significantly more than any simulator does. The simulator tries to do as well as the adversary without knowing the challenge ciphertext y^* nor accessing any oracle.

- **Non-Malleability:** In this scenario, A_1 is given $param$, and outputs a triple (M', s, id) . A_2 receives an encryption of y^* of a random message x_1 . The adversary then outputs a description of a relation R and a vector $\vec{y}$ of ciphertexts. It is assumed that $y \notin \vec{y}$. The adversary hopes that $R(x_1, \vec{x})$ holds. We say the adversary is successful if this can be done with a probability significantly more than that which $R(x_0, \vec{x})$ holds. Here x_0 is also a plaintext chosen uniformly from M' independently of x_1 .

The attack models for Public Key Encryption (PKE) are Chosen Plaintext attack (CPA) (Goldwasser, S., et.al., 1984), non-adaptive chosen ciphertext attack (CCA1) (Rackoff, C., et.al., 1991), and adaptive chosen ciphertext attack (CCA2) (Dolev, D., et.al., 1991). Semantic Security is widely accepted as the natural goal of encryption scheme it formalizes an adversary's inability to obtain any information about the plaintext from a given ciphertext. IND-CCA2 is considered to be the "right" standard security notion for PKE.

For a public key encryption, the adversary has access, as anybody to the encryption key. It can encrypt any plaintext of its choice. Hence, the basic attack is called "Chosen Plaintext Attack" or CPA in short. But the adversary may have access to more information and namely some decryptions. This is modeled by an access to the decryption oracle.

An adversary is called non-adaptive chosen-ciphertext adversary (or a lunchtime adversary denoted by CCA1 adversary) if it can access the oracle before the challenge ciphertext is known only. An adversary is called an adaptive chosen-ciphertext adversary (or CCA2 adversary if it can access the oracle whenever it wants, that is before and after the challenge ciphertext is known with the sole restriction not to use it on the challenge itself. Due to the particular mechanism, the adversaries are given more power in IBE than in PKE. Essentially the adversaries have access



to the *key extraction oracle*, which answers the private key of any queried public key (identity). The IND-CCA2 security for PKE can be extended to IND-ID-CCA2 security for IBE.

Among the four IBE schemes presented, the Boneh-Franklin scheme is semantically secure under the random oracle model under the Bilinear Diffie Hellman Assumption.

FUTURE RESEARCH DIRECTIONS

In almost all the existing schemes on IBE with a small number of exceptions, there is a single global TA issuing keys to all users in the system and all ciphertexts are created using the public parameters of the that single global TA. In practice there will be more than TA each issuing private keys to a set of users. In addition, some users may have keys issued by more than one TA. In such scenarios, it is not unreasonable to assume that the different TAs may share some common system parameters. This will necessitate that the system parameters may have to be reused. This and other issues in the multi TA setting have not been explored fully (Srinivasan, S., 2010a).

CONCLUSION

The Identity based encryption scheme has been put forward by the cryptographic Community to solve some the issues in obtaining the private and public keys in the Public key encryption. By using an id such as an e-mail address, the user can easily generate a public key. The receiver can obtain the corresponding private key from a Trusted Authority. We can also assume more than one Trusted Authority for providing private keys for the users. This chapter presented a number of Identity Based Encryption schemes such as the Cocks's Scheme, Boneh-Franklin Scheme, Authenticated IBE and Hierarchical IBE. It also presented an architecture of the IBE scheme describing how the users will interact with the Trusted Authority and perform encryption and decryption.

REFERENCES

- Bellare, M., & Rogaway, P. (1993). Random Oracles are practical: a paradigm for designing efficient protocols. *ACM Conference on Computer Communications Security*, 62-73. doi:10.1145/168588.168596
- Boneh, D., Craig, G., & Michael, H. (2007). *Space-Efficient Identity based Cryptosystem without pairings*. IEEE Computer Society.
- Boneh, D., & Franklin, M. K. (2001). Lecture Notes in Computer Science: Vol. 2139. *Identity-Based Encryption from the Weil Pairing*. Springer.
- Chatterjee, S., & Sarkar, P. (2011). *Identity Based Encryption*. Springer. doi:10.1007/978-1-4419-9383-0
- Cheng, P., Gu, Y., Lv, Z., Wang, J., Zhu, W., Chen, Z., & Huang, J. (2012). A Performance Analysis of Identity Based Encryption Schemes. *Lecture Notes in Computer Science*, 7222, 289–303. doi:10.1007/978-3-642-32298-3_19
- Cocks, C. (2001). Lecture Notes in Computer Science: Vol. 2260. *An Identity Based Cryptosystem based on Quadratic Residues*. Springer.
- Dolev, D., Dwork, C., & Naor, M. (1991). *Non-malleable Cryptography*. *STOC*, 91, 542–552.
- Goldwasser, S., & Micali, S. (1984). Probabilistic Encryption. *Journal of Computer and System Sciences*, 28(2), 270–299. doi:10.1016/0022-0000(84)90070-9
- Halevi, S., & Krawczyk, H. (1999). Public-key cryptography and password protocols. *ACM Transactions on Information and System Security*, 2(3), 230–268. doi:10.1145/322510.322514
- Horwitz, J., & Lynn, B. (2002). Lecture Notes in Computer Science: Vol. 2332. *Toward hierarchical identity-based encryption*. Springer.
- Joye, M., & Neven, G. (Eds.). (2009). *Identity-Based Cryptography*. IOS Press.

Martin, L. (2008). *Introduction to Identity-Based Encryption*. Artech House.

Okamoto, T. (2006). On Pairing Based Cryptosystems. *Lecture Notes in Computer Science*, 4341, 50–66. doi:10.1007/11958239_4

Phan, D. H., & Pointcheval, D. (2004). On the Security Notions of Public Key Encryption Schemes. *LNCS*, 3352, 33–47.

Rackoff, C., & Simon, D. R. (1991). Non-interactive zero-knowledge proof of knowledge and chosen ciphertext attack. *LNCS*, 576, 433–444.

Shamir, A. (1984). *Identity-Based Cryptosystems and Signature Schemes*. CRYPTO.

Srinivasan, S. (2010). *Identity-Based Encryption: Progress and Challenges*. *Information Security Technical Report*. Elsevier.

Srinivasan, S. (2010a). *New Security Notions for Identity Based Encryption* (Ph.D thesis). Royal Holloway, University of London.

ADDITIONAL READING

Hansen, D. M. (2009). Pairing Based Cryptography A Short Signature Scheme using Weil Pairing, M.Sc.Thesis, <http://www.sagemath.org/files/thesis/hansen-thesis-2009.pdf>

Küchlin, W. (1987). Public key encryption. *ACM SIGSAM Bulletin*, 21(3), 69–73. doi:10.1145/29309.29320

Menezes, A. (1991). An Introduction to Pairing Based Cryptography <http://cacr.uwaterloo.ca/~ajmenez/publications/pairings.pdf>

KEY TERMS AND DEFINITIONS

Adversary: An attacker who may try to guess the secret key or the plaintext from a given ciphertext.

Authentication: refers to verifying the identity of a particular user.

Ciphertext: The term ciphertext refers to the output message from an encryption algorithm.

Hash Algorithm: This produces a fixed size output given a message as input. This output can be used to verify message contents.

Identity Based Cryptography: The algorithms that generate the public key using an id of the user. The private key should be obtained from a Trusted Authority.

Identity: Refers to a string that is used to identify an individual. For example, an e-mail address.

Pairing Based Cryptography: The cryptographic algorithms that use the Mathematics of pairings. The most commonly used pairings include Tate Pairing, Weil Pairing etc.

Plaintext: The term plaintext refers to the input message to an encryption algorithm.

Private Key: The secret key in the public key encryption algorithms. This key is kept secret.

Public Key: The Public key is one of the keys used in public key encryption algorithms. This key is made available publicly.

Random Oracle Model: This model assumes the presence of oracle which returns answers to queries similar to Hash functions. It is suggested that the Oracle be implemented through Hash functions.

Signatures: Signing of a message by one's private key. The hash of a message is encrypted with the private key.

Trusted Authority: The Trusted Authority is a third party trusted to distribute keys. In the Identity Based Encryption Schemes, the Trusted Authority provide secret keys.

Steganography Using Biometrics



Manashee Kalita
NERIST, India

Swanirbhar Majumder
NERIST, India

INTRODUCTION

Steganography is one of the techniques which is used to provide security to the information. There are many other techniques available to do so. Those are cryptography, steganography and watermarking. Cryptography scrambles the message using some encryption algorithm with some secret key. When the receiver receives the scrambled message (cipher), he/she decrypt the message using the proper key (same or different). Last two methods, steganography and watermarking are very much similar to both the methods come from the set of data hiding techniques, but with a different objective. Watermarking is a technique where the cover image is digitally marked using some data hiding technique. The method has some way to logically extract the mark without destroying or harming the cover image. On the other hand, for steganography, the matter of concern is the hidden message only, not the cover image. Steganography pay attention to the degree of imperceptibility where watermarking concentrates on the number robustness of the method. Application of watermarking are copy control, authentication, device control, proof of ownership, etc. Steganography mainly aims to provide the security to the information.

The word steganography is derived from Greek. The Greek word “stego” means cover and “grafia” means writing. The goal of steganography is to conceal the very existence of any secret information in the cover media file. The cover media is any media which usually doesn’t come under suspicion. Selection of cover media has

being changed with the change of technology. In the ancient time, the cover was different, such as messenger’s body part, some natural picture, usual greeting letter, etc.

BACKGROUND

Steganography is a prehistorical practice. From the ancient times, steganography has been using to provide security to the confidential information. Italian mathematician Jerome Cardan reinvented Chinese ancient secret writing method. In that method two parties share a paper mask with holes and after that fill up the blank spaces. The final message appears as an innocuous text. Many secret writing techniques were invented during World War II, such as null cipher, microdot, invisible ink, etc. In the 5th century, BC Hiatus wanted to send some message to his friend secretly. He shaved one of the trusted slave’s head and tattooed a message on it. The slave was sent after his hair grew back. During World War II, Morse codes were encoded in pictures, like long blades of grass indicate dashes and dots were indicated by short blade.

The word biometrics is also originated from Greek word “Bio” which means life and “metric” means measure. Biometric define the measurement of statistical analysis of people’s physical and behavioral characteristics. This is mainly used for authentication, access control, identification. Nowadays, authentication tool/machine developers start to prefer biometrics characteristics as identification or authentication measure rather than

passwords, smart card, etc. Because biometrics is a property which can defines or identify “who are you.” Various biometric characteristics are being used by different authentication machine such as palm geometry, fingerprints, iris, face, skin, etc.

Physical characteristics are related to the feature of the body, such as palm veins, retina, face recognition, DNA, fingerprint, hand geometry, etc. On the other hand, the behavioral characteristic is related to the behavior of a person. It includes signature, voice, gait, typing speed, handwriting, etc. Biometric gets the preference to be a reliable authentication measure than a password, smart card, etc. because biometric characteristics are virtually impossible to steal. Therefore, biometric starts dominating the field of authentication. We can observe the large application of biometric in regular life, e.g. Bank employees use Thumbprint to login into their system, in many universities, offices use biometric punching machine where the biometric feature of employees is used to keep the attendance.

Now, if we focus on the steganography using biometrics, it can be done in two ways, one hides your biometric information in some cover file, and another is the reserve one, i.e., biometric information will carry some secret information. Here a brief discussion related to these two mentioned types are presented.

LITERATURE SURVEY

Anil K. Jain et al. (2003) proposes another method to hide biometric information using steganography. They discuss two scenarios. In the first one, the authors embed the fingerprint information (minutiae) into another fingerprint image, so that attackers do not suspect that the visible fingerprint image is not the actual one. The stego fingerprint image is again encrypted using a secret key to increase the security level. In the second scenario, the minutiae and facial information (Eigen—face coefficient) are hidden in the fingerprint image. Again this stego image embedded in the smart card for authentication.

Hussain Ud-Din et al. (2006) proposes a system for providing better security to the online shopping customers. In this system, there are three main phases. The first phase extract the feature of the biometric information (fingerprint). In the 2nd phase the sensitive information of the electronic shopping card is encrypted and in the final phase extraction biometric feature and encrypted information are embedded in cover image (the image of the shopping card).

The author Yinghuua Lu et al. (2008) proposes a method to improve the security of biometric authentication using lossless and content-based hidden watermarking algorithm. Chaos is employed by the method to encrypt the watermark and the initial condition for chaos are generated by the biometric image of the user. Watermark includes like ID, Palmprint no. etc. which are embedded in the palm print image. Again, the stego palm print image is embedded in the cover image to provide the security to the biometric information as well as authenticate by watermarking the palm print image.

A paper by Abbas Chaddad and Joan Condell (2008), discusses the efficiency of embedding information in the skin tone color space. Anjali A. Shejul and Prof. U.L Kulkarni (2010) proposed a method where secret information is embedded by finding the location to embed using the biometric concept. A DWT based steganography algorithm is used for embedding the information in the skin tone of the human body.

A combination of cryptography, biometric and steganography approach is made by authors Hisham Al-Assam et al. (2013). In the method, the biometric feature is used for remote authentication. On the other, to protect biometric feature of individual, steganography is used. To embed the biometric feature vector, Random LSB scheme is used. Another similar approach is made by Indradip Banarjee et al. (2014) where they integrate the face extraction geometry into the cover image using DWT. Another method combines this three concept again which is suggested by Ashsa Ali, Liyamol (2010). This method uses RC5 encryption

algorithm to encrypt the user's data and embeds the encrypted information into the fingerprint image using some LSB based steganography algorithm.

The author Shivendra Katiyar along with the co-authors (2011) suggests an online voting system where steganography and cryptography have been used to provide added security. The biometric feature, thumb impression is used for authentication. The proposed system also uses the system generated a secret key and SSN no. to log into the voter's account. The system concentrates on the authentication of the voter's identity, but not the confidentiality of the voter's biometric features.

For secure transmission of medical information, a method is proposed by S. Barkathunisha et al. (2013) In this method, they provide a technique to enforce security to the medical information where patient need to send their report to the expert in remote station.

The method proposed by Rasher D. Rashid (2013) hides the feature of a face biometric into a cover image. The authors calculated the PSNR value for the invisibility of the system and recognition rate is calculated using Euclidian Distance. To extract the feature of the face, the method divides the face into multiple frequency bands and each band is divided into non-overlapping blocks. From the blocks, local binary pattern histogram is extracted.

Włodzimerz Kasprzak et al. (2015) suggests a method which uses a printed steganography to authenticate the photo in identity card. This method proposes a steganography technique for authentication of a printed face image.

A. Kapezynski (2011) proposed a hybrid method using biometric and steganography. This approach considers physical and behavioral biometric characteristics. The physical characteristics, the fingerprint is used to hide the behavioral biometric keystroke. Embedding is done by LSB substitution.

D. Goyal proposed Steganography authentication concept using face and voice recognition for secure banking system through mobile. The author uses multiple biometric features for identification.

In the proposed system, the face is used to login into the banking server. After successful logged in, the voice of the user is used to provide better security to the transmission phase. The voice is transmitted using steganography algorithm.

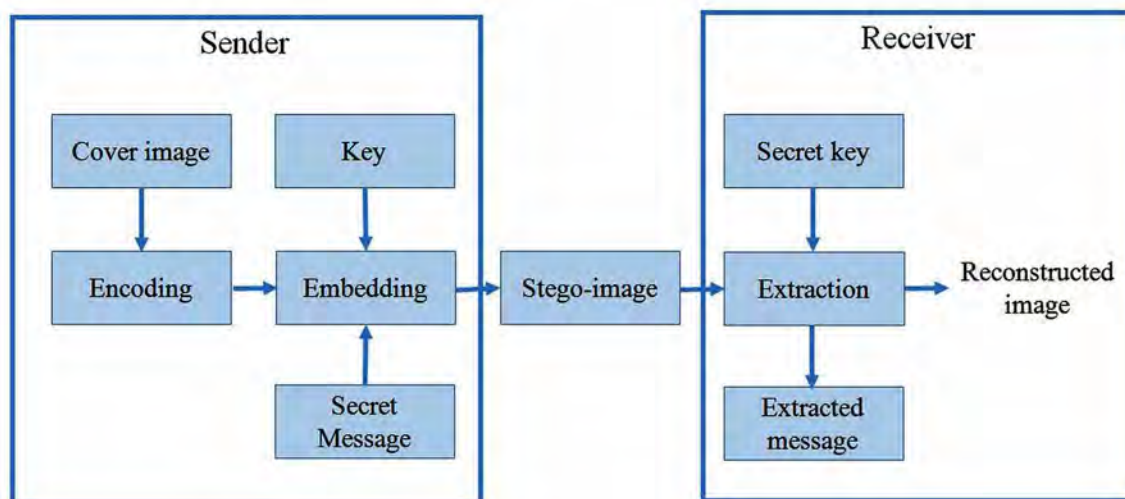
P. Kamble proposed a method for ATM security using multimodal biometrics system with steganography technique. The system embeds the code in the image of the palm print of the user and image is sent to the authentication server with an iris image of the same user.

Many researchers work on steganography using DWT and biometric feature. Sarthi k, in 2013, proposed a biometric based steganography method using circular folding in DWT domain. The method hides the secret message in the skin tone region. To detect skin and non-skin, HSV color space is used. The cover image is cropped after detection of skin tone. DWT is performed on the cropped image. The third plane of the cropped cover images is rotated by 180° using circular folding technique. Next DWT is performed on the folded image and the high-frequency sub-band is used to hide the secret information. Anjali A Shejwal, S. Barve, N. Lavanya (2012) also proposed similar methods where skin tone and DWT have been used.

It has been observed that researchers are constantly searching for a solution for secure online voting system. But, very less work have done in this field. An online voting concept has proposed by N. Gandhi (2014), where the concept of steganography, cryptography using biometrics are used. Similarly, G.M. Kamau (2013) proposed a hybrid steganography method using biometrics (fingerprints) for casting a vote remotely. Using enhanced least significant bit steganography method, the fingerprint template with the voter details are embedded in the voter face image with a Stego key. S. Lokhande (2012) also presented a method using biometrics, cryptography, and steganography with the conjunction of GSM module. The method uses cryptography algorithm (MD5) to generate a unique key using users ID card, users key and time factor. The unique key is embedded in the thumbprint image and send to the authen-



Figure 1. Basic block diagram of steganography



tication server. Few more similar approaches are proposed by N. Malwade (2013), V. J Lakshmi (2014), S Bhattacharyya (2014), Shanthini (2012). S.A.Tambe (2014) approached a method where the fingerprint is used as a cover image to embed the ID of the voter. Embedding process is done in the transform domain using DWT.

STEGANOGRAPHY

A basic steganography system is mainly composed of embedding process and extracting process. Figure.1 illustrates basic block diagram of a steganography system.

The sender selects an unsuspected cover image to embed the secret message. Sometimes, some “key” is used in the embedding procedure, which is optional. The embedding process produces the stego-image and then it is transmitted over the network. The receiver has to run the extracting algorithm with the proper key (if any) to get the message back accurately.

Steganography is a very broad area of application. With the digital revolution, not only mankind gets benefited, but also terrorists and criminals take the advantage to fulfill their suspicious act.

Here, we are focusing only on the positive side of the development. Steganography has many useful applications such as copyright control, smart ID where details of the individual are embedded in the photos, synchronization of video audio secret information circulation etc. While designing steganography method to use in the different field, we always have to consider the main three properties of it. These are-

1. **Imperceptibility:** It is the first property one needs to keep in mind. The basic motive of steganography is to conceal the existing of any message. Higher perceptibility indicates better steganography method.
2. **Capacity:** With good imperceptibility of the stego file, the designer must look after the capacity of the method also. Application of steganography is not limited. So, as per the requirement steganography method should be capable of providing good embedding capacity.
3. **Robustness:** This is also an important property. It is the capacity of the method to withstand the message after some attack or some manipulation operation, such as cropping, rotating, etc.

Some other properties, such as security of the key used in the method, various cost such as computational cost, time cost, are also required to consider.

With the development of steganography, the counter measures of it have also developed in parallel, i.e., steganalysis. Steganalysis is the art of detecting hidden message in the cover file without having any knowledge of the steganography method. There are various steganalysis methods, such as histogram analysis, statistical analysis, etc.

Steganography algorithm can be implemented in various domains. Mainly there are two domains, but further, these are subdivided into many domains. Different researchers visualize the type of steganography methods in various categories. Some of them have divided into six categories depending on the hiding technologies: Statistical method, Distortion techniques, Substitution, Cover generation techniques, Spread spectrum techniques and Transform domain techniques. On the other hand, some researchers have divided depending on the domain, where most of bits are hidden. Those are mainly spatial domain steganography, transform domain steganography, Spread spectrum steganography and model based steganography. To discuss the type steganography methods briefly, we consider the second type division case.

Spatial Domain Steganography

In Spatial domain steganography method, the embedding process is done by directly manipulating the pixel values of the cover image. These type of methods are less complex and less robust. Various approaches in spatial domain are discussed below:

1. **LSB Substitution:** This is the simplest spatial domain technique where the message is embedded in the least significant bit (rightmost). This method can provide good embedding capacity but vulnerable to even minor manipulation, such as rotation, scal-

ing, cropping and addition of noise, lossy compression, etc. Also, it is not robust against stego attacks.

2. **Gray Level Modification:** This technique doesn't embed or hide the data. It uses a mathematical function to map data by modifying their gray level. To map data within the cover image, this technique uses some odd even mapping. For example, odd values map 1 and even maps 0. This method can provide high embedding capacity with low complexity.
3. **Pixel Value Differencing:** This method uses the difference between adjacent pixels. A larger difference indicates the presence of an edge, and according to human visual properties, the human eyes can tolerate more changes in edge areas than that of flat areas. This property leads to high embedding capacity without noticeable perceptibility. Some methods have been developed using this concept and gain good results. These methods are also vulnerable to steganalysis method and stego attacks.
4. **Quantization Index Modulation (QIM):** Brain Chen (2001), proposed QIM algorithm which has higher embedding capacity. It also allows to control robustness and distortion. To embed the information in the cover file using QIM, it first modulates an index or sequence of indices with the embedded information and then quantized the host signal with associated quantizer or sequence of the quantizer. There are some other techniques come under this category, such as vector quantization, singular value decomposition, side-match vector quantization, search order code (SOC).
5. **Multiple Base Notational System (MBNS):** In this method, secret information is converted into symbols (e.g. integer numbers). The pixels of a cover image are modified in such a way that the remainders are equal to symbols after dividing the pixel values by the bases. The degree of the local variation



of pixel values in a cover image determines the specific bases. This method achieves high payload capacity and performs better in terms of PSNR quality factor and Watson's metric.

6. **Prediction Based Steganography:** Direct manipulation or alteration of pixel value leads to the significant distortion. Predictive-based steganography is suggested to overcome this issue where pixel values are predicted using predictor. Here, prediction error values are modified to embed data instead of modifying the pixel values. The gradient adjusted prediction and median edge are mainly used in prediction based image coding.

Transform Domain

All digital images are composed of low and high-frequency component. Low frequency is represented by flat and plain areas whereas sharp transition and edges contribute in high-frequency component. Changes in the low frequency are easily visible to the human visual system. Therefore, embedding an equal number of information in both frequencies is not feasible. It embeds the message bits in optimally chosen coefficients. Coefficients are calculated from the cover image using some transformation function. There are a number of transformation functions available, such as DCT, DWT, Integer transform, Haar transform, Hadamard transform, contourlet transform, ridgelet transform, ripplelet transform, DD DT DWT etc. These methods are computationally complex than spatial domain and robust against attacks and image manipulation operations. Some popular transform functions are:

1. **Discrete Cosine Transformation:** In 1974, researcher community working on an image compression has suggested DCT, which turns like a great achievement in this field. The DCT can be considered as a discrete time version of the Fourier cosine series. The com-

putational complexity of DCT is $O(n \log n)$, like FFT, DFT. DCT is a real-valued with fewer coefficient and better approximation of a signal. The 2D DCT compression $c(u,v)$ of a $N \times N$ image signal $f(x,y)$ can be defined as:

$$c(u, v) = \alpha(u) \alpha(v) \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x, y) \cos \left[\frac{(2x+1)u\pi}{2N} \right] \times \cos \left[\frac{(2y+1)v\pi}{2N} \right]$$

where,

$$\alpha(u) = \begin{cases} \frac{\sqrt{1}}{N}, & \text{for } u = 0; \\ \frac{\sqrt{2}}{N}, & \text{for } u = 1, 2 \dots N-1 \end{cases}$$

In 2016, Fengyong Li uses DCT to steganalyze over large scale social network with high order joint features and clustering ensembles.

2. **Discrete Wavelet Transformation:** The word wavelet has been used for decades in digital signal processing and exploration geophysics. First wavelet transformation is Haar wavelet proposed by Alfred haarin 1909. This is a multiresolution transform method. The wavelet transform for a square integration function $f(t)$ is defined by-

$$w_f(a, b) = \int_{-\infty}^{\infty} f(t) \varphi_{a,b}^*(t) dt$$

where $*$ denotes complex conjugate, $\varphi_{a,b}^*(t) = \frac{1}{a} \varphi \left(\frac{t-b}{a} \right)$, a basis (mother) wavelet $\varphi(t)$ by dilations and transition. Here, a and b are scale factor and transition factor respectively. $\frac{1}{a}$ is the normalization factor for energy normal-

ization. For different application different basis function are available. A basis can be modified according to the requirement. Some popular basis are – Haar Basis, Gaussian wavelets, Gabor wavelets, Daurbechies basis.

3. **Walsh Transformation:** In 1923, Walsh defined transform, but in 1823 Hadmard had also achieved a similar result by the application of certain orthogonal matrices which called Hadamard matrices. Hadmard matrices contain only ± 1 . It is equivalent to multidimensional DFT of size $2 \times 2 \times 2 \dots 2 \times 2$ and can be considered as being built out of size two discrete transformation.
4. **Curvelet Transform:** This technique was proposed by Candes and Donoho in 1999. It can be viewed as an extension of wavelet transformation and it became more popular in the same field. This transformation is multi-scale directional transform technique which is an almost optimal non adaptive sparse representation of object with edge.
5. **Contourlet Transformation:** It is also a multiresolution directional tight frame which design to efficiently approximate image. This transformation was proposed by Do and Vetterli in 2002. They construct a discrete domain multiresolution and multidirectional expansion. Non separable filter banks are used to do so in the same manner as wavelet filter bank. This technique represents at 2D image with smooth corners. Contourlet has the properties of multiresolution, localization, directionality, critical Sampling and anisotropy. Method proposed a filter Bank structure to obtain sparse expansion of 2D image having a smooth contour. To implement this double filter bank, the Laplacian pyramid is used for capturing point of discontinuities and to link point of discontinuities into linear structure, a direction filter bank is used.

Spread Spectrum

A well-known concept in digital communication, spread spectrum was proposed by Harvel et al. (1999). The method involves in spreading the bandwidth of a narrowband signal across a wide band of frequencies like white noise. At the decoder side, during extraction process image restoration and error control techniques can be used. It is difficult to detect because the energy of the narrow band signal is low in any frequency band after spreading. The resulting signal is embedded in the cover image to get the stego image. Due to the difference in power of signals (embedded and cover image), the SNR of stego image is very low which result in low detectability. This method provides good embedding capacity and also maintain the robustness against statistical attacks. Statistical properties are preserved because the secret message is spread throughout the cover image.

Model Based Steganography

In 2003, P Shallee proposed a spatial domain technique called Model-based steganography which based on the statistical properties of a cover medium. In this method, the cover image is divided into two parts, and one part is selected for embedding message without altering the statistical properties. A popular adaptive method is proposed by Hoki which it is known as “A block Complexity based Data Embedding (ABCDE). Here noisy block is replaced by the block obtained by embedding the message.

BIOMETRICS

We all are aware of unauthorized access to information. There are a number of methods to prevent unauthorized access. The most common way is to use username and password combination. However, mainly there are three ways to authenticate a user-



1. Something you know. For example, password.
2. Something you have. For example, token.
3. Something you are. For example, measurable physical features.

These three pillars of authentication can be used separately or combinedly for stronger authentication. In some situations, only one way is not sufficient to provide stronger authentication. For example, it would be better if some biometric information is required while withdrawing money using ATM card because ATM card can be stolen, pin code can be guessed.

Biometric assures individual's unique physical and behavioral characteristics to authenticate or recognize their identity. Physical biometric features include fingerprints, facial, Iris, hand palm geometry, retina, etc. Behavioral characteristics are handwriting, signature, voice, keystroke pattern, gait, etc.

The biometric system works by scanning the fingerprint or iris to get the pattern of it, capturing a digital color/grayscale image for face recognition, etc. The feature of the image is transformed into the template using mathematical functions. These biometric templates are added to the database for further identification. A biometric system works for two different purposes; one for adding the template to the database for further reference and another is to find a match with the existing template.

Biometric has occupied the field of authentication because of its special characteristics. Among all characteristics, biometric feature has chosen for authentication based on the following characteristic.

1. **Uniqueness:** It is the most desirable and primary characteristic so that chances of occurring same characteristic of two people is minimum.
2. **Stability:** Chosen feature should not change over time.

3. **Ease of Capturing:** Capturing of the biometric template should be convenient to the user with preventing misrepresentation of the feature.

Because of the special characteristic of biometric, it occupies the place of most secure and convenient authentication tool. This section discusses some of the popular biometric features.

Physical and Behavioral Characteristics

Physical biometric comprises some of the shape or composition of the body like a fingerprint, hand geometry, face, ear, retina, etc. Most of the physical biometrics are time invariant and also satisfy the character of the uniqueness with stability.

1. **Fingerprint:** It is the unique physical characteristic which features can be recorded as arches, loop, whorl pattern with the ridges, minutiae, and furrows. Matching Fingerprint can be done in three ways:
 - a. Minutiae based,
 - b. Correlation based,
 - c. Ridge features.
2. **Hand Geometry:** An optical camera is used by hand geometry based system to capture two orthogonal two-dimensional image of palm and sides of hand. A number of dimensional measures are collected including height, length, width, distance between joints, etc. The system does not consider the fingerprint or Palm print.
3. **Palm Print:** Palm print system takes the image of palm. It is different from hand geometry.
4. **Face:** A quality digital camera is required to capture the facial images of the user. Characteristic of face are analyzed. Many software applications have developed which have facial image authentication, due to the ease of capturing the facial image.

5. **Facial Thermography:** It is similar to the face recognition, except an infrared camera, which is used to capture the image. It detects heat patterns emitted from the skin and created by the branching of blood vessels. Thermograms patterns are highly distinctive.
6. **Vein Recognition:** The vein pattern structure of a user is image processed and stored. It is believed by many researchers that vein biometric recognition can produce high accuracy rate than fingerprint recognition.
7. **Blood Pulse:** It is not as strong authentication measure as it has a high false matching rate. It can be measured on a finger with an infrared sensor.
8. **Retina:** This biometric analyses the layer of blood vessel situated in the back of the eye. The network of a blood vessel in the retina is not genetically determined, and therefore even identical twin do not share the same pattern. Some retina based biometric system used a low-intensity light source through an optical coupler to scan the unique pattern of retina. This kind of system has low false, positive, high reliability, high cost. It is not user convenient because it requires close contact with the reading device.
9. **Iris:** This type of biometric system analyses features of the colored ring of tissue that surrounds the pupil. Iris recognition systems apply mathematical pattern recognition technique to the images of irises of an individual's eye. There are many advantages of iris matching, besides its speed of matching, they are extreme resistant to false matches, stability, and extremely visible organ of the eyes. It also works with contact lenses, eye glasses and non-mirror sunglasses.

Apart from these, there are many more features, such as tongue print, dental scan, ear, nail bed, etc. Here, only the popular ones have been discussed. Biometrics, such as, DNA matching, ear, odor are also used but because of many issues, these characteristics are not so popular than

others. Now, let us give a brief overview of some behavioral characteristic.



1. **Signature:** Signature has been widely used for authentication, and it is also a very old practice. Signature authentication may be either static or dynamic. Static authentication uses only the geometric feature whereas dynamic includes some more features such as velocity, acceleration, pressure, the trajectory of the signature, etc.
2. **Voice:** It is a combination of physical in behavioral characteristics. Physical characteristics, vocal tracts, mouth, nasal cavities, and lips are included and under behavioral, emotional and physical state are considered. Traditionally voice-based authentication divided into text dependent and text independent categories. Text independent authentication is more complex than text dependent because in text dependent, the speaker speaks a predetermined phrase and in text independent, no constraint exists. Regardless of their classification, voice based authentication faces a lot of challenges because of the variability like emotional state of the speaker, misspoken phrases, environmental noise, etc.
3. **Keystroke:** This kind of authentication systems aim to capture the latency period between keystroke and hold time of the user's keyboard interaction to provide a unique representation of each user. The main advantage of this technique is that it allows continuous authenticate since the user can be analyzed over a large period.
4. **Gait:** This type of system authenticate people by the way they walk. To create gait signature, some models are built, based on the temporal and spatial matrices of the human motion. However, gait is not supposed to be very distinctive across individuals. Therefore, it is not suitable for high security scenario. It also involves video sequence analysis, which may be comparatively expensive.

Behavioral characteristic is less stable than physiological. As time passes, changes come in handwriting, voice and gait. These characteristics are also influenced by one's emotional state, physical state, environment, etc. Therefore, the most common biometrics used for authentication is Fingerprint and Iris. These biometric features satisfy the main required properties for authentication. The properties make them popular are:

1. **Stability:** Both the characteristics remain constant in an entire lifetime.
2. **Uniqueness:** Finding two identical fingerprint or Iris is nearly impossible.
3. **Flexibility:** Fingerprint or iris recognition system can be integrated easily into the existing system. It can also be operated as standalone.

Other biometrics are also used, but all biometric features are not equally efficient, secure and user-friendly. For example, facial recognition system fails in case of identical twins and for many other factors like the beard, facial hairs, glasses, scars, etc. Voice, signature, gait also have many issues, like voice, signature is prone to forgers and are affected by environment, time variant, etc. On the other hand, some biometric analysis is costly to use for public authentication, such as DNA matching.

Fingerprint is the most common user-friendly and distinctive biometric. Same is applicable for iris biometric. In spite of all points, people prefer to use fingerprint by fingerprint scanner than placing an eye in IR scanner camera.

STEGANOGRAPHY USING BIOMETRICS

When this biometric information is used to provide security, authentication, then simultaneously they also need security, since biometrics represent a person physiologically and behaviorally. This information is itself very confidential to a person. For providing security, steganography can employ

in this situation. While transferring biometric information over network we have to take care of its confidentiality. Here we have two type of use of biometrics in steganography. In one type, we can use steganography to provide security to the biometric information. In a different kind, we can use biometric information to provide security to secret information.

Security to Biometrics

To prevent the misuse or leaking of any biometric information of a person, one can take advantage of steganography method. We have already discussed that how steganography works and how does it efficient in providing security. Figure 2 shows that the biometric information, the fingerprint is embedded in the cover image 'baboon'. For embedding, any steganography method can be used as the requirement. There are various techniques with different methods in different domains which are already discussed in the former section.

Security by Biometrics

Biometrics image could be a very good/unsuspicious cover image. When a sender sends a biometric image over the network, then generally it is not suspected by the intruder, as because the confidential data biometrics is sent as it is. In that case, we can use the biometric image to hide some secret information. There are some other cases also, where biometric require to hold some information secretly. For example, in some medical report, where the patient's personal information (e.g. name, age, sex etc.) should be kept hidden to others except the receiver, the information is embedded in the medical image secretly. Figure 3 shows the use of biometric information as cover image in steganography. Figure 3(a) shows hiding of secret information in iris biometrics and Figure 3(b), the information of the patient is concealed in the MRI report.

Biometrics characteristics of an individual are very confidential since it defines the person physi-



Figure 2. Security to biometric data

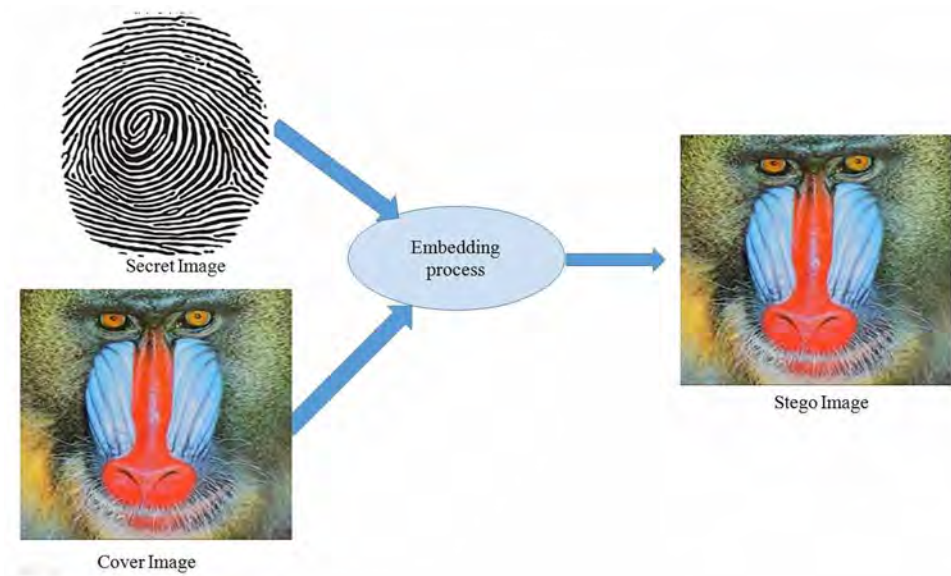
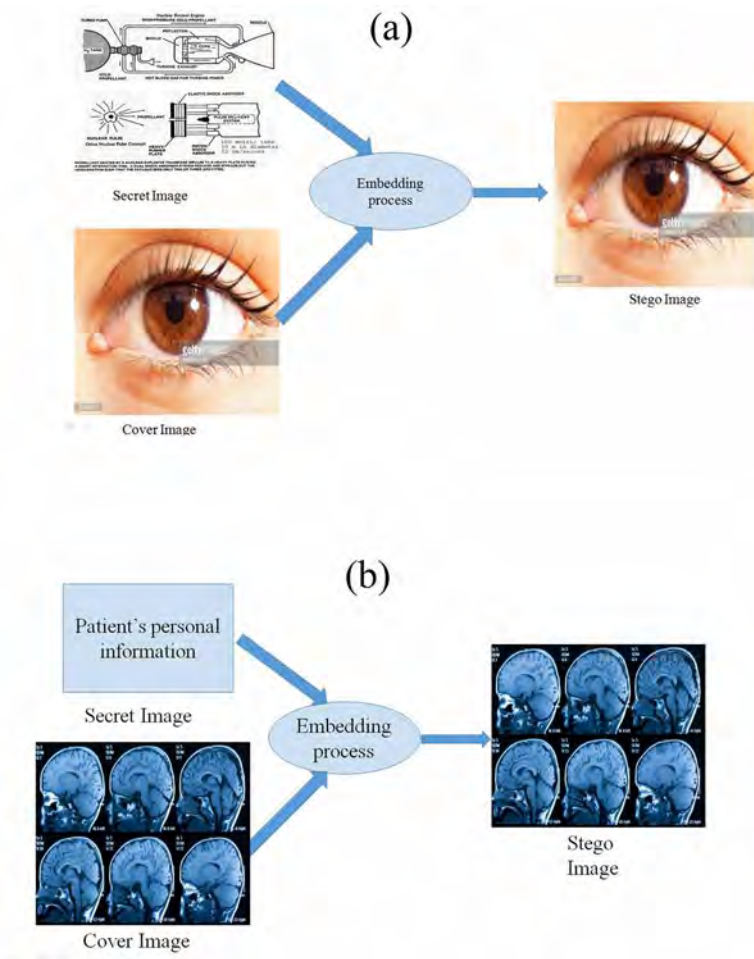


Figure 3. Biometrics image as cover image



cally and behaviorally. Security of these always should be a matter of concern for researchers. We know that many encryption algorithms being used to assist the security enhancement of biometric system, but there are still some issues to resolve. In 2001, Ratha addressed eight primary sources of attacks in biometrics system.

1. Fake biometric at the sensor;
2. Resubmission of old digitally stored biometrics signal (typical replay attack in voice recognition);
3. Override feature extraction;
4. Tampering with the feature representation in network environment;
5. Override matcher;
6. Tampering with stored templates in database;
7. Channel attack between stored templates and the matcher during biometric data or feature template transmission;
8. Decision level override.

We know that the cryptography algorithm scrambles the message, which attracts the intruder to decode the message. To avoid this situation, steganography is used. For example, a person needs to send his fingerprint for some normal online authentication. If that person's activities are being tracked by some intruder without his knowledge, then his fingerprint will also be leaked. That fingerprint may be misused (e.g. online fund transfer, personal account access, etc.) by the intruder further.

SOLUTIONS AND RECOMMENDATIONS

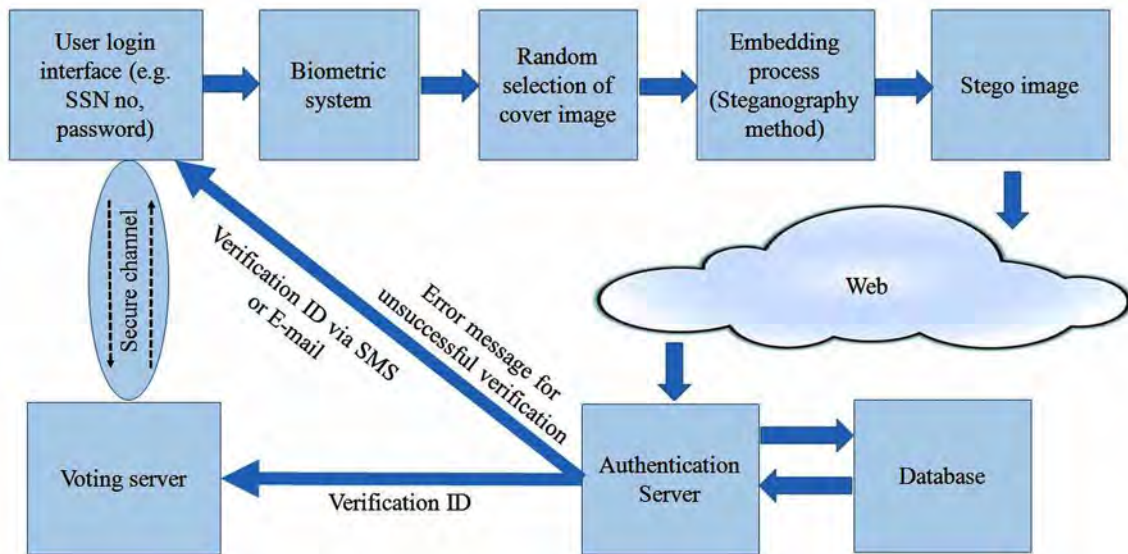
From the literature survey, it is found that biometrics has extensively been used by the developer to provide security and authenticity. Many researchers and developers use a biometric feature such as skin, iris, etc. to hide the secret information. But, the security of biometric information is not much been focused. D. Goyal proposed a steganography authentication system. In that method, the face and voice of the user are embedded in the cover

image and is sent to the bank server. Bank server contacts authentication server to verify and then the transaction takes place.

Here, will present a hypothetical secure voting system for casting a vote remotely. As a responsible citizen, we all wanted to establish the voting right. But due to our busy scheduled life, we may not able to cast our vote each time. It would be beneficial if we could cast our vote through a secure trusty online system. The basic working of an example system is prototype below:

1. Login to the user account through user name and password. The username could be any unique ID, such as SSN no, Adhar no, Voter ID, etc.
2. The system will ask to provide the thump print through some fingerprint scanner which will be integrated part of the system.
3. Without user's knowledge, the system will select a cover image randomly from its database and embeds this information onto the selected cover image using some robust steganography algorithm.
4. The stego image is sent to the authentication server.
 - a. If biometric verification fails, then an error message will be sent to the user.
 - b. If user has already cast a vote, then also an error message will be sent.
 - c. First, the server will check whether the user has already cast a vote or not. If not, then it will send the extracted data to the authentication server database. After verification, a verification number will be generated. The confirmation number will be sent to the user as well as to the voting server as shown in Figure 4 (via sms or email).
5. If the verification number match between the user and voting server, then a secure channel will be established between the two parties otherwise communication link will be disconnected with a message. The connection will be for a limited period only.

Figure 4. Block diagram of secure e-voting system



6. Now the user can cast their vote before the session expires.

Here multimodal biometric authentication can also be applied. With fingerprint, we can use iris also for stronger authentication. For scanning fingerprint, any standard fingerprint scanner can be used. Many technologies have been used including optical, capacitive, RF, thermal, piezoresistive, ultrasonic, piezoelectric.

Biometric Steganography

Fingerprint as the biometric feature is a very good choice for authentication. It holds the important properties of good biometrics: stability, uniqueness, user-friendly, etc. In steganography, the primary goal is to achieve high imperceptibility and it happens when the size the image or message to be embedded is as small as possible. In case of a fingerprint, we can represent it as a binary image without losing any important information. Converting a grayscale image to binary image saves eight times bit. Therefore using this concept we can achieve a very high-quality stego image.

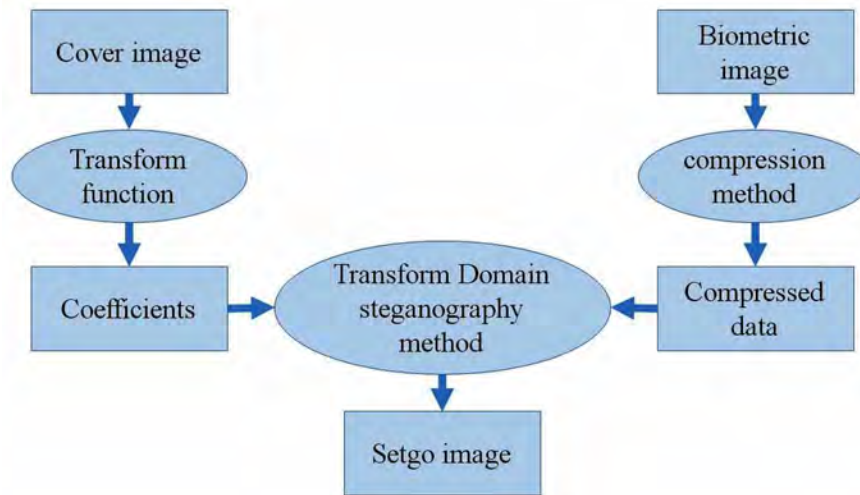
Now, for steganography method, considering the robustness property of transform domain, any

one of the feasible methods can be adopted among transform domain methods. Again for the secret biometric image, we can compress the secret image to reduce the size, so that we can achieve better quality stego image. Figure 5 shows the block diagram with a compression method. Compression can be lossless or lossy. Both the methods are useable here. With lossy compression (e.g. DCT, DWT compression) fingerprint can successfully authenticate a user.

FUTURE RESEARCH DIRECTIONS

Information security has been increased here by the amalgamation of steganography and biometrics to a next level. But being a new area very less joint literature and research is available in this mixed area. The necessity of protection of intellectual property rights is increasing day by day due to most of the information and data being stored in cloud space and being readily available on the internet. Thus the inclusion of both physical and behavioral biometrics in all possible permutations and combinations with transform domain variants of steganography both on software and realization on hardware is to be addressed in future. Then based

Figure 5. Transform domain steganography using compression method



on computational complexity as well as response time and robustness best possible techniques need to be analyzed and optimized to obtain the best possible ways of enhancing security. Thereby, the future research in this area must target these areas.

CONCLUSION

In this chapter presents an introduction to the concept of steganography with a scope. It also discussed the basics of biometrics and different type of it. Steganography with the idea of biometrics such as biometric image as a cover image, biometric image as secret image have been presented. Application of these entire concepts is shown in an online voting system which is now on demand.

REFERENCES

Ahmed, N., Natarajan, T., & Rao, K. R. (1974, January). Discrete Cosine Transform. *IEEE Transactions on Computers*, C-23(1), 90–93. doi:10.1109/T-C.1974.223784

Al-Assam, H., Rashid, R., & Jassim, S. (2013). Combining steganography and biometric cryptosystems for secure mutual authentication and key exchange. *8th International Conference for Internet Technology and Secured Transactions (ICITST)*, 369-374. doi:10.1109/ICITST.2013.6750224

Ali, A., Aliyar, L., & Nisha, V. K. (2010). RC5 encryption using key derived from fingerprint image, Computational Intelligence and Computing Research (ICCIC). *2010 IEEE International Conference on, Coimbatore*, 1-4. doi:10.1109/ICCIC.2010.5705778

Banerjee, I., Bhattacharyya, S., Mukherjee, S., & Sanyal, G. (2014). Biometric steganography using face geometry. *TENCON 2014 - 2014 IEEE Region 10 Conference*, 1-6. doi:10.1109/TENCON.2014.7022450

Barkathunisha, S., & Meenakumari, R. (2013). Secure transmission of medical information using IRIS recognition and steganography. *International Conference on Computation of Power, Energy, Information and Communication (ICCPEIC)*, 89-92. doi:10.1109/ICCPEIC.2013.6778504

- Barve, S., Nagaraj, U., & Gulabani, R. (2011). Efficient and Secure Biometric Image Steganography using Discrete Wavelet Transform. *International Journal of Computer Science & Communication Networks*, 1(1).
- Bhattacharyya, S., Banerjee, I., Chakraborty, A., & Sanyal, G. (2014). Biometric Steganography Using Variable Length Embedding. *International Journal of Computer, Electrical, Automation, Control and Information Engineering*, 8(4).
- Cheddad, A., Condell, J., Curran, K., & Kevitt, P. (2010). Digital image steganography: Survey and analysis of current methods. *Signal Processing*, 90(3), 727–752. doi:10.1016/j.sigpro.2009.08.010
- Cheddad, A., Condell, J., Curran, K., & Kevitt, P. M. (2008). Biometric Inspired Digital Image Steganography. Engineering of Computer Based Systems, 2008. ECBS 2008. *15th Annual IEEE International Conference and Workshop on the*, 159-168. doi:10.1109/ECBS.2008.11
- Chen, B., & Wornell, G. W. (2001). Quantization index modulation: A class of provably good methods for digital watermarking and information embedding. *IEEE Transactions on Information Theory*, 47(4), 1423–1443. doi:10.1109/18.923725
- Chen, B., & Wornell, G.W. (2011). Quantization index modulation: A class of provably good methods for digital watermarking and information embedding. *IEEE Trans. Inform. Theory*, 47, 1423–1443.
- Gandhi, N. (2014). Study on Security of Online Voting System Using Biometrics and Steganography. *IJCSC*, 5(1), 29-32.
- Gonzalez & Woods. (n.d.). *Digital Image Processing*. Pearson.
- Goyal, D., & Wang, S. (n.d.). *Steganographic Authentications in conjunction with Face and Voice Recognition for Mobile Systems*. Academic Press.
- Hioki, H. (2010). A data embedding method using BPCS principle with new complexity measures. Proc. of Pacific Rim Workshop on Digital Steganography.
- Hussam, U.-D., Ihmaidi, A., Al-Jaber, A., & Hudaib, A. (2006). Securing Online Shopping Using Biometric Personal Authentication and Steganography. *Proceedings of 2nd International Conference on Information & Communication Technologies*, 233-238. doi:10.1109/ICTTA.2006.1684376
- Inthavisas, K., & Lopresti, D. (2012, March). Secure speech biometric templates for user authentication. *IET Biometrics*, 1(1), 46–54. doi:10.1049/iet-bmt.2011.0008
- Itkis, G., Chandar, V., Fuller, B. W., Campbell, J. P., & Cunningham, R. K. (2015, September). Iris Biometric Security Challenges and Possible Solutions: For your eyes only? Using the iris as a key. *IEEE Signal Processing Magazine*, 32(5), 42–53. doi:10.1109/MSP.2015.2439717
- Jain, A. K., & Uludag, U. (2003). Hiding Biometric Data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(11).
- Kahn, D. (2005). The History of steganography. Chapter Information Hiding. *Lecture Notes in Computer Science*, 1174, 1–5.
- Kalita, M., & Tuithung, T. (2016). A comparative study of steganography algorithms of spatial and transform domain. In *National Conference on Recent Trends in Information Technology NCIT2015* (pp. 9-14). IJCA Press.
- Kamau, G.M., Kimani, S., & Mwangi, W. (2013). A general Purpose Image-Based Electors Smart Card Using an Enhanced Least Significant Bit Steganographic Method for Information Hiding: A case study of the Kenyan Electoral Process. *International Journal of Computer Science Issues*, 10(2).

- Kapezynski, A., & Banasik, A. (2011). Biometric Logical Access Control Enhanced by Use of Steganography Over Secured Transmission Channel. *The 6th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications*. doi:10.1109/IDAACS.2011.6072859
- Kasprzak, W., Stefańczyk, M., & Wilkowski, A. (2015). Printed steganography applied for the authentication of identity photos in face verification. *2nd International Conference on Cybernetics (CYBCONF)*, 512-517. doi:10.1109/CYBConf.2015.7175987
- Katiyar, S., Meka, K. R., Barbhuiya, F. A., & Nandi, S. (2011) Online Voting System Powered By Biometric Security Using Steganography. *2011 Second International Conference on Emerging Applications of Information Technology*, 288-291. doi:10.1109/EAIT.2011.70
- Lai, L., Ho, S. W., & Poor, H. V. (2011, March). Privacy-Security Trade-Offs in Biometric Security Systems—Part I: Single Use Case. *IEEE Transactions on Information Forensics and Security*, 6(1), 122–139. doi:10.1109/TIFS.2010.2098872
- Lakshmi, V. J., Vineka, P., & Anbarasu, V. (2014). Biometrics and Steganography based Secure Online Voting System. *International Journal of Research in Engineering & Advanced Technology*, 2(2).
- Lavanya, N., Manjula, V., & Rao, N. V. K. (2012). Robust and Secure Data Hiding in Image Using Biometric Technique. *International Journal of Computer Science and Information Technologies*, 3(5), 5133–5136.
- Li, F., Wu, K., Lei, J., Wen, M., Bi, Z., & Gu, C. (2016, February). Steganalysis Over Large-Scale Social Networks With High-Order Joint Features and Clustering Ensembles. *IEEE Transactions on Information Forensics and Security*, 11(2), 344–357. doi:10.1109/TIFS.2015.2496910
- Lokhande, S., Sawant, D., Sayyad, N., & Yengul, M. (2012). YE-Voting through Biometrics and Cryptography- Steganography Technique with conjunction of GSM Modem. *International Journal of Computer Applications*.
- Lu, Y., Li, X., Qi, M., Li, J., Fu, Y., & Kong, J. (2008). Lossless and Content-Based Hidden Transmission for Biometric Verification. *Intelligent Information Technology Application, 2008. IITA '08. Second International Symposium on*, 462-466. doi:10.1109/IITA.2008.351
- Majumder, S., & Das, T. S. (2013). *Watermarking of Data Using Biometrics, Handbook of Research on Computational Intelligence for Engineering, Science, and Business*. IGI Global.
- Malwade, N., Patil, C. Chavan, S., & Raut, S. Y. (2013). Secure Online Voting System Proposed By Biometrics And Steganography. *International Journal of Emerging Technology and Advanced Engineering*, 3(5).
- Odelu, V., Das, A. K., & Goswami, A. (2015, September). A Secure Biometrics-Based Multi-Server Authentication Protocol Using Smart Cards. *IEEE Transactions on Information Forensics and Security*, 10(9), 1953–1966. doi:10.1109/TIFS.2015.2439964
- Potdar, V. M., & Chang, E. (2004). Gray level modification steganography for secret communication. *Proc. of 2nd IEEE International Conference on Industrial Informatics*, 223–228.
- Prabakaran, G., Bhavani, R., & Kanimozhi, K. (2013). Dual transform based steganography using wavelet families and statistical methods. *Proc. of the 2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering*, 287–293. doi:10.1109/ICPRIME.2013.6496488
- Rane, S., Wang, Y., Draper, S. C., & Ishwar, P. (2013, September). Secure Biometrics: Concepts, Authentication Architectures, and Challenges. *IEEE Signal Processing Magazine*, 30(5), 51–64. doi:10.1109/MSP.2013.2261691

Rashid, R. D., Jassim, S. A., & Sellahewa, H. (2013). Covert exchange of face biometric data using steganography. *5th Computer Science and Electronic Engineering Conference (CEEC)*, 134–139. doi:10.1109/CEEC.2013.6659460

Sajedi, H., & Jamzad, M. (2009). Contsteg: Contourlet-based steganography method. *Wireless Sensor Network*, 1(03), 163–170. doi:10.4236/wsn.2009.13022

Sallee, P. (2003). Model-based steganography. Proc. of the 2nd International Workshop on Digital Watermarking, 2939, 154–167.

Santhi, K., & Kumar, A. M. N. (2013). Biometrics based Steganography using Circular Folding in DWT Domain. *International Journal of Computer Applications*, 61(10).

Shanthini, B., & Swamynathan, S. (2012). Multimodal Biometric-based Secured Authentication System using Steganography. *Journal of Computer Science*, 8(7), 1012–1021. doi:10.3844/jcssp.2012.1012.1021

Shejul, A. A., & Kulkarni, U. L. (2010). A DWT Based Approach for Steganography Using Biometrics, Data Storage and Data Engineering (DSDE). *2010 International Conference on*, 39–43. doi:10.1109/DSDE.2010.10

Shejul, A. A., & Kulkarni, U. L. (2011). A Secure Skin Tone based Steganography Using Wavelet Transform. *International Journal of Computer Theory and Engineering*, 3(1).

Stallings. (n.d.). *Cryptography and network security*. Pearson.

Subhedara & Mankarb. (2014). Current status and key issues in image steganography: A survey. *Computer Science Review*, 13–14, 95–113.

Tambe, S.A., Joshi, N.P., & Topannavar, P.S. (2014). Steganography & Biometric Security Based Online Voting System. *International Journal of Engineering Research and General Science*, 2(3).

Woodward, J. D. (1997, September). Biometrics: Privacys foe or privacys friend? *Proceedings of the IEEE*, 85(9), 1480–1492. doi:10.1109/5.628723

Wu, D., & Tsai, W. H. (2003). A steganographic method for images by pixel value differencing. *Pattern Recognition Letters*, 24(9–10), 1613–1626. doi:10.1016/S0167-8655(02)00402-6

Zhang, X., & Wang, S. (2004). Vulnerability of pixel-value differencing steganography to histogram analysis and modification for enhanced security. *Pattern Recognition Letters*, 25(3), 331–339. doi:10.1016/j.patrec.2003.10.014

ADDITIONAL READING

Bertillon, A. (1985). La couleur de l'iris. *The Review of Scientific Instruments*, 36(3), 65–73.

Bodade, R. M., Talbar, S. N., & Ojha, S. K. (2008). Iris recognition using rotational complex wavelet filters: A novel approach. *IEEE International Conference on Innovations in Information Technology, IIT2008*, pp. 658–662. AIAin. doi:10.1109/INNOVATIONS.2008.4781770

Boles, W., & Boashash, B. (1998). A human identification technique using images of the iris and wavelet transform. *IEEE Transactions on Signal Processing*, 46(4), 1185–1188. doi:10.1109/78.668573

Chicago, Illinois. Xie, L., & Arce, G. (1998). Joint wavelet compression and authentication watermarking. *Proceedings of the IEEE International Conference on Image Processing*, Vol. 2, (pp. 427–431).

Chicago, Illinois. Yao, P., Li, J., Ye, X., Zhang, Z., & Li, B. (2006). Iris recognition algorithm using-modified log Gabor filters. *IEEE 18th International Conference on Pattern Recognition, ICPR 2006*, (pp. 461–464). China.

- Daugman, J. (1992). High confidence personal identification by rapid video analysis of iristexture. *Proceedings of the IEEE International Conference on Security Technology, Crime Countermeasures*, (pp. 50-60).
- Daugman, J. (1993). High confidence visual recognition of persons by a test of statistical independence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(11), 1148–1161. doi:10.1109/34.244676
- Daugman, J. (2004). How iris recognition works. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(1), 21–30. doi:10.1109/TCSVT.2003.818350
- Jain, A., Bolle, R., & Pankanti, S. (2002). *Biometrics: Personal identification in networked society*. Kluwer Academic Publishers.
- Kundur, D., & Hatzinakos, D. (1999). Digital watermarking for telltale tamper proofing and authentication. *Proceedings of the IEEE*, 87(7), 1167–1180. doi:10.1109/5.771070
- Lin, E. T., & Delp, E. J. (1999). A review of fragile image watermarks. *Proceedings of the Multimedia and Security Workshop*, (pp. 25-29).
- Liu, R., & Tan, T. (2002). A SVD-based watermarking scheme for protecting rightful ownership. *IEEE Transactions on Multimedia*, 4(1), 121–128. doi:10.1109/6046.985560
- Ma, L., Tan, T., Wang, Y., & Zhang, D. (2004). Efficient iris recognition by characterizing key local variations. *IEEE Transactions on Image Processing*, 13(6), 739–750. doi:10.1109/TIP.2004.827237 PMID:15648865
- Memon, N., Shende, S., & Wong, P. (1999). On the security of the Yueng-Mintzer authentication watermark. *Final Program and Proceedings of the IS&T PICS 99*, (pp. 301- 306).
- Rao, R. M., & Bopardikar, A. S. (2001). *Wavelet transforms: Introduction to Theory and applications*. Pearson Education Asia; doi:10.1117/1.482718
- Ross, A. A., Nandakumar, K., & Jain, A. K. (2006). *Handbook of multibiometrics*. Springer.
- Seitz, J. (2005). *Digital watermarking of digital media*. Information Science Publishing; doi:10.4018/978-1-59140-18-4
- San Jose, California. Wu, M., & Liu, B. (1998). Watermarking for image authentication. *Proceedings of the IEEE International Conference on Image Processing*, Vol. 2, (pp. 437-441).
- Savanna, G. Nixon, M. S., & Aguado, A. S. (2002). *Feature extraction and image processing*. Oxford, UK: Newness Publishers.
- Shnayderman, A., Gusev, A., & Eskicioglu, A. M. (2004). *A multidimensional image quality measure using singular value decomposition*. Bellingham, WA: Society of Photo-Optical Instrumentation Engineers; doi:10.1117/12.530554
- Stinson, D. (1995). *Cryptography theory and practice* (3rd ed.). Boca Raton, FL: CRC Press.
- Sun, Z., Wang, Y., Tan, T., & Cui, J. (2004). Robust direction estimation of gradient vector field for iris recognition. *International Conference on Pattern Recognition*, (pp. 783-786).
- Swanson, M., Kobayashi, M., & Tewfik, A. (1998). Multimedia data embedding and watermarking technologies. *Proceedings of the IEEE*, 86(6), 1064–1087. doi:10.1109/5.687830
- Wayman, J., Jain, A., Maltoni, D., & Maio, D. (2005). *Biometric systems: Technology, design and performance evaluation*. Springer. doi:10.1007/b138151
- Wildes, R. P., Asmuth, J. C., Green, G. L., & Hsu, S. C. (1994). A system for automated iris recognition. *Proceedings of 2nd IEEE Workshop on Application of Computer Vision*, (pp. 121-128). doi:10.1109/ACV.1994.341298
- Wolfgang, R., & Delp, E. (1996). A watermark for digital images. *Proceedings of the IEEE International Conference on Image Processing*, 3, 219–222. doi:10.1109/ICIP.1996.560423

Wolfgang, R., & Delp, E. (1999). Fragile watermarking using the VW2D watermark. Proceedings of the IS&T/SPIE Conference on Security and Watermarking of Multimedia Contents, (pp. 204-213).

Zhang, D., Jing, X., & Yang, J. (2006). *Biometric image discrimination technologies*. Hershey, PA: Idea Group Publishing; doi:10.4018/978-1-59140-830-7

Zhu, X., Zhao, J., & Xu, H. (2006). A digital watermarking algorithm and implementation based on improved SVD. *Proceedings of the 18th IEEE Computer Society International Conference on Pattern Recognition (ICPR'06)*.

KEY TERMS AND DEFINITIONS

Biometrics: It refers to the technology to measure physical and behavioral characteristics of an individual, such as fingerprint, palm print, iris, DNA, retina, signature, keystroke etc. Biometric characteristics are extensively used in authentication and verification process as most of the characteristics are unique.

Cover Media: In steganography, a carrier media is selected to hide the message or to carry the message in it, that carrier media is called cover media. A cover media can be any unsuspecting file, such as natural picture, family photo, video, audio clip, etc.

Cryptography: Cryptography is a technique to secure a communication. It converts the message into some meaningful text (cipher) using encryption algorithm, so that no one can read that message without knowing the decryption algorithm. Depending on the secret key used in the algorithm it is divided into two categories: symmetric - if one key is used for encryption as well as decryption and asymmetric - if two different keys are used for encryption and decryption.

Steganography: It is a technique to provide security to the secret information using data hiding technique. In steganography, a cover media is selected to hide the information. Depending on the place of data hiding it is mainly divided into two categories: Spatial domain - if information is embedded directly in the pixel values and Transform domain - if information is embedded in the coefficient values which are calculated from the pixel values of the cover image.

Stego Image: Stego image is the output of the embedding process. Stego image contains the hidden message either in pixel values or in optimally selected coefficients.

Watermarking: It is a data hiding technique whose primary motive is to authenticate. Watermarking can be visible or invisible. Here the cover image is also important along with the embedded message. Robustness is the main criteria for selection of embedding algorithm.



Usable Security

Andrea Atzeni

Politecnico di Torino, Italy

Shamal Faily

Bournemouth University, UK

Ruggero Galloni

Square Reply S.r.l., Italy

INTRODUCTION

Recent decades have been characterized by the growth of information technologies in the private and public sectors. The positive impact that ICT has on job performance, as well as the expansion and creation of business opportunities for companies, count as the main drivers for this growth. This growth led to the proliferation of distributed applications and physical devices, and the diffusion of technologies that facilitate social participation and social interaction. All these applications, devices and interactions may contain important information, or give access to sensitive data, putting them at risk.

The rapid diffusion of technology has led to the reduction of active security monitoring, as well as the lack of technically competent people in control of applications and devices. Moreover, the increment in social interaction increases the damage other people can directly or indirectly cause.

Traditionally, security is only considered as strong as its weakest link, and people were considered as the weak links (Schneier, 2003). This thinking triggers a vicious circle. (Adam & Sasse, 1999) stated that users are informed as little as possible on security mechanisms took by IT departments, precisely because they are seen as inherently untrustworthy. Their work has shown that users were not sufficiently aware of security issues and tend to build their own (often inaccurate) models of possible security threats.

Users have a low perception of threats because they lack the necessary information to understand their importance. According to (Sasse & al., 2001) blaming users for a security breach is like blaming human error rather than bad design. Security has, therefore, a human dimension that must be neither ignored nor neglected. The increase in the number of breaches may be attributed to designers who fail to sufficiently consider the human factor in their design techniques. Thus, to undo the Gordian knot of security, we must provide a human dimension to security.

BACKGROUND

Human-Computer Interaction (HCI) is a field concerned with the interaction between people and technology, and how this supports humans in completing tasks to achieve one of more specific goals. Traditionally, it has been involved in analyzing and improving usability.

HCI has been an active area of research since the 1980s. It has focused on improving the design of user interfaces, and helping users transforming their goals into productive actions for the computers. Improving user interfaces and usability is important because poorly designed interfaces increase the potential for human error. In particular, human behavior is largely goal-driven, therefore the execution of activities which help the users to achieve their goals is the main key to create a

usable system. So, when a user “engages with a complex system of rules that change as the problem changes” (e.g. an interface does not present information clearly and coherently with a user mental model), it leads to “Cognitive Friction” (Cooper, 2004).

The “Cognitive Friction” is a by-product of the information age, and it is more evident in all the computing devices lacking a natural cause-effect relation between user input and device output, e.g. when similar inputs result in different outputs.

When a person is dealing with the cognitive friction, ancestral mechanisms of the human being come into play. As result, in this case, users cannot be modeled as purely rational beings. Thus, to understand users’ behavior, and to appreciate how systems can be made usable, we need to consider the following factors:

- Users are driven by goals. People are naturally prone to pursuing goals. In achieving this, according to Krug “every question mark adds to our cognitive workload, distracting our attention from the task at hand” (Krug, 2005). This, according to Norman (Norman, 2002), creates usability issues, because it introduces the cognitive friction into play and leads users to make mistakes, which sometimes can also result into security flaws;
- Users do not read the instructions. Users proceed by trial and are not interested in reading manuals, instructions or documentation. For most of the users, it is not important to know how to do something, until the moment in which it is not necessary to use it (Krug, 2005);
- Users follow the path of least resistance. Several studies in the field of HCI have shown how users, in their task to accomplish a goal, tend to seek the path requiring them less effort (e.g. (Norman, 2002)). Once they find the first reasonable option allowing them to perform the desired action, it becomes irrelevant to them if it

is not the most efficient and safe option. Furthermore, users have no incentive to improve. When users “find something that works - no matter how badly – they tend not to look for a better way” (Krug, 2005). Some operations can be inconvenient from the point of view of performance, others, in the long run, can cause damage to the system: users may be unaware of it until problems show up for the first time.

While many research studies in HCI has been focused in defining what usability is and, consequently, intervene in improving user interfaces, several studies have shown that the “ease of use” cannot be limited to those aspects alone (Whitten & Tygar, 1999) (Balfanz & al., 2004).

To increase the acceptance of the security mechanisms, conventional wisdom suggests it is sufficient to make them easier through a more usable user interface. In practice, however, it is not enough to provide a proper user interface, even in the case it is supported by specific configuration guidelines. This is what Whitten and Tygar argue, in their study “Why Johnny cannot encrypt” (Whitten & Tygar, 1999), which is a seminar paper in the usable security literature. This study focuses on analyzing data and email encryption of the security software Pretty Good Privacy 5.0 (PGP). They showed that user errors have not decreased, despite years of improvements to the graphical interface. This has led to additional studies looking beyond the interfaces.

This field of study, which deals with analyzing the usability issues related to security, is called HCI-Sec and was founded in 2000 by Whitten as a mailing list on Yahoo! Groups. It has been said that HCI-Sec “only rarely received significant attention as a primary subject for study” (Balfanz & al., 2004), this despite the fact that “usability remains one of the most pressing and challenging problems for computer security” (Whitten & Tygar, 1999).

Although HCI-Sec has only recently gained momentum, initial studies have their roots in 1975,



when (Saltzer & Schroeder, 1975) argued that the usability was an essential component of a secure system. In their seminar “The protection of Information in Computer System” they presented eight basic principles that serve as guidelines for the design of systems aimed at protecting information. The principle of the “Psychological Acceptability” is one of them, and states: “it is essential that the human interface is designed for ease of use so that users routinely and automatically apply the protection mechanisms correctly”.

Since then, little work has been focused on HCI-Sec and, as a result, the security systems are sometimes poorly designed, leading to cases where users seek alternative interactions with the system or completely avoid the security mechanisms. Given the difficulty in making IT systems usable, it is unsurprising that the problem of “aligning usability and security” has been almost neglected until the beginning of the early 90s of the last century.

According to Fléchaïs, it is wrong to justify such a dearth of research as a tension between usability and security. Until that decade, the research community was more focused on technical trade-offs, such as for example the realization of robust encryption on low energy consumption microprocessors (Flechaïs & Sasse, 2005). This is reasonable, because, before the growth of the Internet, security was mainly a physical concern, and physical thinking was based on a military mindset.

The problem of usability in security, however, was not limited to this, and already existed during the 80s and 90s. For example, it was already possible to improve usability and possibly weaken security by automating common tasks. Brad Reid (Reid, 1987), argued that programmer convenience is the antithesis of security because it becomes intruder convenience if the programmer’s account is compromised. Reid mentioned a “programmer” because at that time the main users of the computer systems were mostly researchers or computer science specialists who possessed some programming aptitude. These individuals

possessed technical skills, received specific training and were, therefore, prone to ignore usability, and failing to identify the security implications this might have.

In the 1990s, with the diffusion of personal computers and the mass adoption of the Internet, the problem of usable security has remained virtually unexplored and did not leverage pre-existing HCI research. The initial solution to the problem addressed the symptoms rather than the root cause, by updating the anti-virus software, or the patching software in known problems. Therefore, research has been focused more on short-term practical gains, rather than long-term design changes that attend to both usability and security. A further problem is that few developers are trained in usability, or have significant software security experience.

The advent of HCI-Sec introduced the idea of security as an important consideration for usability, while usability is an important aspect of security (Cranor & Garfinkel, 2005). Therefore, if the purpose of HCI was to ensure that users would reach their goals by the use of better interfaces, HCI-Sec aims to ensure that users are able to achieve these goals also in the most secure way.

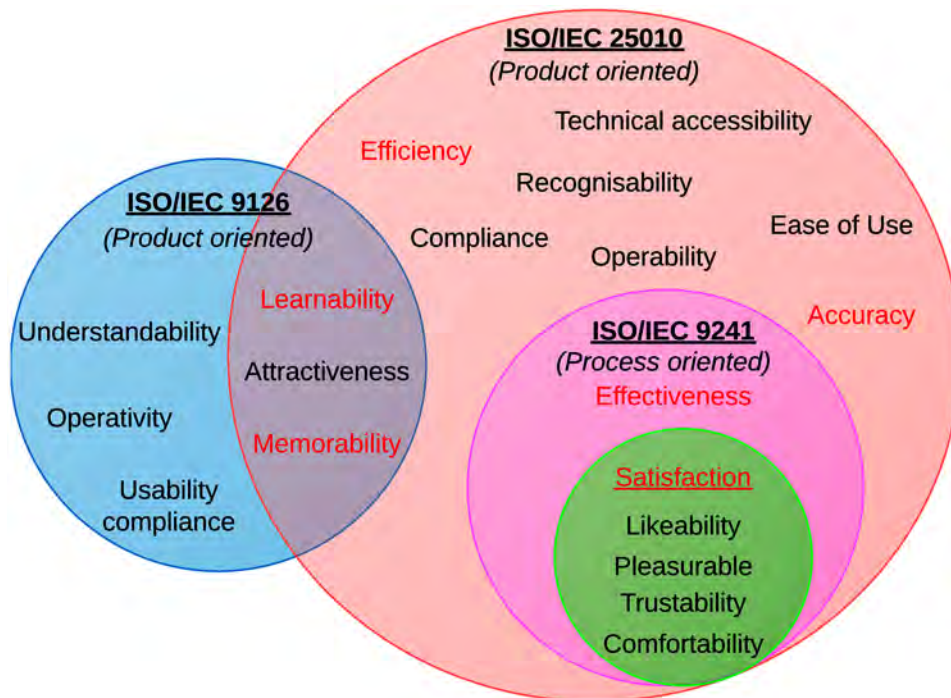
MAIN FOCUS OF THE CHAPTER

The aim of this chapter is to analyze the goals and state of the art of usability and security to determine where and how they can be effectively “aligned”.

ISSUES, CONTROVERSIES, PROBLEMS

Usability has become a key factor in the quality of the software and has a determining role in productivity and acceptance (Cranor & Garfinkel, 2005). This term has more than one meaning, though; it refers multiple concepts that may or may not be taken together. Some are based on *execution time*, *performance*, *user satisfaction* and *ease of learn-*

Figure 1. Usability definitions according to ISO/IEC 9241-11, ISO/IEC 9126 and ISO/IEC 25010



ing (also known as *learnability*). Thus, it remains something that is not unequivocally defined, and is subject to interpretation based on the stake one has in usability (Hertzum, 2010).

Over the years, the International Organization for Standardization (ISO) itself produced various, and sometimes conflicting, definitions of usability (Figure 1). These definitions can be classified into two main categories: *product-oriented* and *process-oriented*. The former category provides definitions of the qualities that belong to the final product. This appears to be a reasonable approach, because software usability is essential for end-users, as crucial for achieving particular tasks quickly and effectively. The latter category focuses on the methodological aspects of obtaining usability: for a software developer, usability describes the internal attributes of a system, including concerns such as quality of design, documentation, and maintenance.

These various points of view and the different requirements have resulted in contrasting perspectives on usability, carried out by several groups

of experts in a non-uniform and inconsistent way. For example, some terms have different meanings and labels. In document ISO/IEC 9241-11 (ISO, 1998) *learnability*, as a quality of a software, is designed as the “time of learning”, while in ISO/IEC 9126 (ISO, 2012) is defined as “comprehensible input and output, instruction readiness, messages readiness”.

Discrepancies among the standard can be even more significant: in the standard ISO/IEC 9126, usability is defined in a product-oriented way as a set of attributes that bear on the effort needed for use and on the individual assessment of such use, by a stated or implied set of users. The qualitative properties to be achieved are Understandability, Learnability, Operability, Attractiveness and Usability compliance (Abran & al, 2003). The standard that replaced it, ISO/IEC 25010, focuses on the user’s goals and on how fast they are achieved, in addition to user satisfaction with the system. In this document, usability replaces learnability property in favour of *operability* (Lew & Olsina, 2010), which is described as the degree

to which the product has attributes that enable it to be understood, to be learned, to be used, and to be attractive to the user, when used under specific conditions.

With this definition of operability, the properties to be achieved are *appropriateness, recognisability, ease of use, learnability, attractiveness, technical accessibility* and *compliance*. The same standard describes however also a different model, *Quality in Use*, in which the usability appears described as the extent to which a product can be used by specific users to achieve specified goals with effectiveness, efficiency and satisfaction without adverse consequences in a specified context of use.

The process-oriented point of view was defined in the document ISO/IEC 9241, which is a suite of international standards on “Ergonomics of Human System Interaction”. In Part 11, the definitions of usability from different perspectives are grouped together. The key components are effectiveness and satisfaction. The former describes the interactions from the point of view of process efficiency and puts the focus on the results and valuable assets. The latter requires carefulness on the user’s needs. The standard attempts to explain how to identify the information that has to be taken into account when evaluating usability in terms of measures of user performance and user satisfaction.

The criterion of satisfaction is very difficult to be measured and, for this reason, additional usability factors have been proposed in Part 2 of ISO/IEC 25010. They are *likeability, pleasurable, comfort and trust*.

Given the subjectivity and the different contexts in which the term “usability” can be used, Kainda and Flechais (Kainda et al., 2010) proposed to consolidate it in six key factors which are defined as:

- **Effectiveness:** A system is only usable if its users can achieve intended goals, and effectiveness is measured by whether users are able to complete a particular task or not;

- **Satisfaction:** A system must be accepted by users, otherwise it is bound to fail, even if is usable;
- **Accuracy:** A system demands may have an impact on the user’s tasks. For example, a system may require 100% accuracy in an providing information, such as a pin code or a password. However, this accuracy, is not always achievable by the user, making the system unusable;
- **Efficiency:** To guarantee usability, a system must ensure that each user’s goals are achievable within an acceptable amount of time and effort;
- **Memorability:** A system may require users to memorize secrets, namely passwords. This may be problematic since the users are cognitively burdened with credentials, and other secrets;
- **Knowledge:** This corresponds to the Learnability property. However, the user may not attempt to learn or understand the system, as users tend to care only about the parts of the system of interest to them. Therefore, knowledge of the security mechanisms or policies is required by the user;

These characteristics can be measured in different ways. Effectiveness, satisfaction, efficiency and memorability can be measured directly, while accuracy and knowledge are measured indirectly, i.e. the first set of characteristics can be measured directly by quite simple empirical indicators, while the latter are typically derived by combining more indicators.

Another HCI-sec problem is that adequate usability is essential in specific security mechanism (e.g. authentication process), but the requirements for achieving it and a high level of security may collide (Braz & Robert, 2006). For example, in the case of password-based authentication, many usability principles (e.g. use shortcuts in case of frequent use, provide informative feedback)

contrast with best practices (e.g. password must not be showed during typing, and only success or failure must be reported, to mitigate social engineering and guessing).

SOLUTIONS AND RECOMMENDATIONS

From the perspective of HCI, there are several principles for building a system that is “quick to use and relatively error-free” (Johnson, 2007). One of the most important of these is ensuring the system “does what the user wants” without “complicating the user’s task”. Another important aspect is to evaluate the usability of a system. In this regard, the System Usability Scale (SUS) (Brooke, 1996) is a widely accepted base. The SUS is designed to give a quick impression of the overall usability of a product. It consists of ten questions (e.g. “I found this system unnecessarily complex”, “I found the various functions in this system were well integrated”) rated on a Likert scale, resulting in an overall 0-100 value, where 100 represents excellent levels of usability. SUS has been adopted in many contexts, also in HCI-sec (De Witt, J. Kuljis, 2006) since quick to complete, thus avoiding user frustration and ensure answer accuracy.

From the perspective of methods and procedures used in HCI, many of them have been adapted to the HCI-Sec. The main difference is the focus on a balanced trade-off between usability and security.

The methodologies between HCI and HCI-Sec differ for at least five key aspects that are analyzed in the following paragraphs, detailing related recommendation as well.

THE SECONDARY GOAL

People do not generally sit at their computers wanting to manage their security; rather they want to send mail, browse web pages or download software.

Traditionally, security definitions have been defined around attackers. Unfortunately, doing so ignores the legitimate and non-malicious use, and also may adversely affect the system (Kainda & al, 2010). Users may not have the perception of damaging the system or, through making certain actions or inactions, bypassing security systems, putting their assets risk. Users must be constantly made aware of the operations involving security and the system must ensure that it is hard to make catastrophic errors. Furthermore, if such events occur, user actions should be reversible. To illustrate how this can be achieved, consider the implementation of dialog boxes requiring confirmation of a particular action. The implementation of the “Empty Trash” feature in desktop operating systems typically allows accidentally deleted files to be recovered, or the “Undo” button, now present in many desktop applications, allows for an action to be reverted. In a business setting, backup and redundancy of servers are amongst the systems used to avoid potential damages even from non-malicious users. Unfortunately, as Garfinkel has demonstrated, even “Empty Trash” functionality can behave in a manner inconsistent with a corresponding interface design (Garfinkel, 2005).

ABSTRACTION

Security policies are usually phrased as abstract rules that are easily understood by programmers but “alien” and unintuitive to many members of the wider user population.

(Johnson, 2007) proposes a focus on learnability and memorability, properties which, as we have discussed, belong to Usability. Facilitating the learning process is possible, by creating a consistent lexicon transmitted through the user interface. This is convenient since it was discovered that a particular trend also applies to IT users: they prefer not to invest time in training or reading manuals, but in learning the functionality of the system through the exploration of the user interface (Krug, 2005).



THE HIDDEN FAILURE

It is difficult to provide good feedback for security management and configuration because configurations are complex and not easy to summarize.

Making a secure system does not guarantee its security because the system must also be installed and used in a secure way. (Bishop, 2005) noted that the configuration is a key component of security because it is during the configuration of a system that it is defined who will interact with the system and how. Practitioners and security staff often make mistakes in applying default software configurations, ignoring the fact that different configurations lead to different security contexts. For example, a computer configured to be secure in a university research environment could be considered insecure in a military installation. In the former, information might be made accessible to the whole class or research group while, in the latter, they might be accessible only on a need to know basis.

One method used to counteract and minimize the adverse effects of an incorrect configuration is the "fail-safe default" principle (Saltzer & Schroeder, 1975). This states that the safest solution is a default configuration without any permission granted. During the configuration phase, a security responsible task selects the correct permissions for each system function and group of users. When configuring a firewall, this principle corresponds to the whitelist configuration: everything that is not explicitly allowed is forbidden by default. This policy contrasts with the blacklist, which grants any permission by default and chooses specifically the ones to forbid. The former one, despite being more difficult to handle, forces discussion on any permission to be enabled, thereby making the system more secure.

THE BARN DOOR

Once a secret has been left accidentally unprotected, even for a short time, there's no way to be sure it has not already been read by an attacker.

Once sensitive data or vital assets for the company are compromised and made public by mistake, it is possible that attackers will use it for their own advantage. There are several ways to approach this problem. You might try to avoid social engineering attacks, where even expert users fall victim to if channels of communication they trust and use regularly are compromised. This risk can be prevented using anti-fraud mechanisms, aimed at preventing phishing through e-mail or other channels. Should an attacker successfully obtain sensitive information such as passwords, private keys or credit card numbers, it should also be possible to erase and getting new information. In the case of commercially sensitive intellectual property, DRM can also be implemented, which can control access to resources, and revoke permissions in the event of a successful attack. DRM technology is, however, complex to maintain and not without its own usability issues (Favale & al., 2016)

THE WEAKEST LINK

The security of a networked computer is like a chain: it is only as strong as its weakest component.

It is generally recognized that the user is often the weak point of a computer system from a security perspective. However, as discussed, this creates a vicious circle in which users are kept unaware of what the security mechanisms are. Therefore, users are driven to the creation of their own security views, which fail to align with reality. To avoid this issue, security mechanisms should be complemented by specific guidelines that take into account the specific constraints of security mechanisms, minimizing discrepancies introduced by users with different backgrounds and skills.

FUTURE RESEARCH DIRECTIONS

Future research direction will address the problem of measuring usable security in a more systematic and practical way, or, as a first milestone toward the goal, understand if there are real advantages and tangible benefits resulting from formative and summative usability assessment processes. Such an assessment is not easy because of the previously discussed complex variables into play. Moreover, there are several aspects that are influenced by contextual conditions, such as economic resources, time, and other economic or innovation drivers.

To address these evaluation issues, we need to understand what it means to precisely evaluate usable security. Only then will it be possible to identify what methods can be effectively used in the analysis processes.

Also, all these facets activate different levels of sub-choices that depend on several variables and the context of use. One of these could, for example, be advancing the project in time: do you want to evaluate a system in its initial stage of development, or at a different iteration of the same application? It has already been noted that, while considering security at an early stage of a software product's design is virtuous, design techniques may be needed that specifically consider security at a later stage (Faily, 2015).

For any improvements, development should be scientifically measurable. As Lord Kevin said over 200 years ago "if you cannot measure it, you cannot improve it", meaning that without a scientifically sound evaluation methodology, would be difficult to draw any objective conclusions and take any proper improving actions (Atzeni & Liroy, 2005).

Measurement should not be an end in itself, but lead to something analogous to a benchmark, which is a result or a group of results that can become a point of reference and standard; this enables comparison and judgment on how good or bad things are.

CONCLUSION

Human behavior is goal-driven, therefore each aspect of a system interacting with users, security included, should be organized to help users to achieve their goals. In particular, security must be embedded paying attention to usability aspects, to avoid "cognitive friction".

Since usable security principles can be applied both to final product and to the production process, from one hand it is necessary to adopt methodologies to understand and measure the usability of the final artifact, from the other, all production line components should be considered in light of usability effectiveness, starting from the earliest steps in building software. This ensures that the quality of usable security is 'built into' the final product, and diagnose the feedback that allows the project to be changed before its final release.

The ability to diagnose and correct an error in the usability of a software before entering the market is in itself a significant benefit, as are methods that can also be used to determine quality variations between two iterations of a given software. It is, therefore, necessary to describe and scientifically evaluate the properties of usability and security as two correlated factors, even if they are both difficult to quantify and even define. To facilitate quantification and definition, past literature split up usability and usable security in more atomic pieces (e.g. effectiveness, satisfaction, accuracy, efficiency, memorability) to make them more identifiable and comparable.

Finally, a scientifically sound usability assessment is a target of great interest. Further research is welcome because it is a complex problem (even when decomposed in sub-parts like satisfaction or memorability) and because the context of a product under evaluation can introduce influencing variables, enlarging the problem complexity.



REFERENCES

- Abran, A., Khelifi, A., Suryn, W., & Seffah, A. (2003). Usability Meanings and Interpretations in ISO Standards. *Software Quality Journal*, 11(4). DOI:10.1023/A:1025869312943
- Adams, A. Sasse, M.A. (1999). Users are not the enemy. *Communications of the ACM*, 42(12). DOI:10.1145/322796.322806
- Atzeni, A., & Lioy, A. (2005). Why to adopt a security metrics? A little survey, *Proc. of QoP'05 - Quality of Protection*, 1(12).
- Balfanz, D., Durfee, G., Grinter, R. E., & Smetters, D. K. (2004). In Search of Usable Security: Five Lessons from the Field. *IEEE Security & Privacy*, 2(5). DOI doi:10.1109/MSP.2004.71
- Bishop, M. (2005). Psychological Acceptability Revisited. *Security and Usability: Designing Secure Systems that People Can Use*, 1(12).
- Braz, C., & Robert, J. M. (2006). Security and usability: the case of the user authentication methods. *18th Conference on l'Interaction Homme-Machine*, 199-203. doi:10.1145/1132736.1132768
- Brooke, J. (1996). SUS-A quick and dirty usability scale. *Usability Evaluation in Industry*, 189(194), 4-7.
- Cooper, A. (2004). *The Inmates Are Running the Asylum: Why High Tech Products Drive Us Crazy and How to Restore the Sanity* (2nd ed.). Sams Publishing.
- Cranor, L. F., & Garfinkel, S. (2005). *Security and Usability: Designing Secure Systems that People Can Use*. O'Reilly.
- De Witt, A. J., & Kuljis, J. (2006). Aligning Usability and Security-A Usability Study Of Polaris. *Proc. of the Symp. On Usable Privacy and Security*. doi:10.1145/1143120.1143122
- Faily, S. (2015). Engaging Stakeholders during Late Stage Security Design with Assumption Personas. *Information and Computer Security*, 435(446).
- Favale, M., McDonald, N., Faily, S., & Gatzidis, C. (2016). Human Aspects in Digital Rights Management: The Perspective of Content Developers. *Fourth International Workshop on Artificial Intelligence and Law*.
- Flechais, I., & Sasse, M.A. (2005). Developing Secure and usable software. *OT2003*. DOI 10.1234/12345678
- Garfinkel, S. (2005). Sanitization and Usability. *Security and Usability: Designing Secure Systems that People Can Use*, 293(317).
- Hertzum, M. (2010). Images of Usability. *International Journal of Human-Computer Interaction*, 567(600).
- ISO/IEC. (1998). *ISO/IEC 9241-11 - Guidance on Usability management*. ISO.
- ISO/IEC. (2012). *ISO/IEC 27002:2007 - Product quality*. ISO.
- Johnson, J. (2007). *Common User Interface Design Don'ts and Dos*. Interactive Technologies.
- Kainda, R., Flechais, I., & Roscoe, A. W. (2010). Security and Usability: Analysis and Evaluation. *ARES '10 International Conference on Availability, Reliability, and Security*, 275(282). Doi:10.1109/ARES.2010.77
- Krug, S. (2005). *Don't Make Me Think: A Common Sense Approach To The Web Usability*. New Riders Pub.
- Lew, P., Li, Z., & Olsina, L. (2010). Usability and user experience as key drivers for evaluating GIS application quality. *International Conference on Geoinformatics*, 1(6). doi:10.1109/GEOINFORMATICS.2010.5567803

Norman, D. A. (2002). *The Design of Everyday Things*. Basic Books.

Reid, B. (1987). Reflections on some recent widespread computer break-ins. *Communications of the ACM*, 103(105). doi:0.1145/12527.315716

Saltzer, J. H., & Schroeder, M. D. (1975). The protection of information in computer systems. *Proceedings of the IEEE*, 1278(1308). doi:10.1109/PROC.1975.9939

Sasse, M. A., Brostoff, S., & Weirich, D. (2001). Transforming the 'Weakest Link' - a Human/Computer Interaction Approach to Usable and Effective Security. *BT Technology Journal*, 122(131). doi:10.1023/A:1011902718709

Schneier, B. (2003). *Beyond fear: Thinking sensibly about security in an uncertain world*. Copernicus Books. doi:10.1057/palgrave.sj.8340200

Whitten, A., & Tygar, J. D. (1999). Why Johnny can't encrypt: a usability evaluation of PGP 5.0. *USENIX Security Symposium*. Retrieved from <http://dl.acm.org/citation.cfm?id=1251421.1251435>

ADDITIONAL READING

Nielsen, J. (1993). *Usability Engineering*. Academic Press.

KEY TERMS AND DEFINITIONS

Cognitive Friction: The affinity friction between the user and the software that originates in the user mind when a product does not behave the way the user expects (e.g. a button on the screen that does not trigger any action when the user press it). (<https://www.linkedin.com/pulse/20140801230851-205508682-what-the-heck-is-cognitive-friction>).

Comfort (ISO/IEC 25010): The extent to which the user is satisfied with physical comfort.

Effectiveness: The properties which measures to what extent interactions achieve objective process efficiency indicators (i.e. concrete results of user actions while using the addressed product).

Likeability (ISO/IEC 25010): The extent to which the user perceives achievement of pragmatic goals, including successful subjective results of use and consequences of use.

Memorability: A factor which measures how much a product require users to memorize secrets (e.g. passwords or passphrases).

Operability: The degree to which the product has attributes that enable it to be understood, be learned, be used and be attractive to the user, when used under specific conditions.

Pleasurable (ISO/IEC 25010): The extent to which the user is satisfied with his perceived achievement of hedonistic goals of stimulation, identification and evocation and associated emotion responses.

Psychological Acceptability: A founding principle of usable security stating that “*it is essential that the human interface is designed for ease of use so that users routinely and automatically apply the protection mechanisms correctly*”.

Process-Oriented Usability: The categorization of usability aiming to achieve it addressing the characteristics of the process to obtain the final product (e.g. documentation and design effort).

Product-Oriented Usability: The categorization of usability aiming to achieve it addressing the final products characteristics (e.g. learning curve to use the product).

Satisfaction: The property which measures to what extent the user's needs are subjectively satisfied by the product.

Trust (ISO/IEC 25010): The extent to which the user is persuaded that the product will behave as intended.



Category K

Knowledge Management

Boosting the Social Development of the Majority Through the Creation of a Wireless Knowledge Society

K

Danilo Piaggese
Framericas, USA

INTRODUCTION

The rapid advances and pervasive diffusion of information and communication technology (ICT), combined with the growth of the wireless Internet, has led to deep transformations in economic, social and institutional structures. ICT applications affect the performance of businesses and the efficiency of markets, foster the empowerment of citizens and communities as well as their access to knowledge, and contribute to strengthening and redefining governance processes at all institutional levels. Nevertheless, as all major and wide-ranging technological advances, the deployment of ICT is at the same time creating enormous opportunities and posing daunting challenges to the Majority in the emerging economies (EE).

According to C.K. Prahalad (2005) in his book, *The Fortune at the Bottom the Pyramid*, “The distribution of wealth and the capacity to generate incomes in the world can be captured in the form of an economic pyramid. At the top of the pyramid are the wealthy, with numerous opportunities for generating high levels of income. More than 4 billion live at the Base of the Pyramid (BOP), on less than \$2 per day. Those are the Majority.”

This paper presents some of the successful sustainable ICT practices aiming at boosting the social development of the Majority contributing to the creation of a wireless and inclusive Knowledge Society. It also offers a road map for the international financial institutions, particularly the Multilateral Development Banks (MDBs), aiming at supporting ICT for development programs benefitting EE.

DOI: 10.4018/978-1-5225-2255-3.ch434

BACKGROUND

The United Nations Millennium Declaration (United Nations, n.d.) noted that efforts to make internet access available to all and to harness the power of ICT could contribute toward the achievement of the Millennium Development Goals (MDGs), thereby creating “digital opportunities” in development. The ongoing debate on the new set of UN Sustainable Development Goals (SDGs) reignites the interest for the ICT in a contest of emerging knowledge economies and societies.

The ICT can facilitate the participation of lower income populations, the majority at the base of the pyramid, according to the definition given in (Prahalad, 2005) in the development process by directly tackling relevant aspects, which precisely hinder their integration into social and economic development. Such aspects concern:

- Limited knowledge and literacy which impairs access to skills and jobs (education);
- Poor health and sanitary conditions limiting employability and risk-taking attitudes (health);
- Scarcity of economic opportunities (economy);
- Limited involvement in civic life and in the democratic processes, as well as uneasy access to public services (government).

The following successful sustainable practices, show how ICT can help reducing the risks of exclusion related to the aspects cited above, thus contributing to the integration of lower income

populations into social and economic development. Based on the lesson learned from the practices, we draw some conclusions and offer some recommendations. These recommendations are the basis for forward-looking scenarios that can be realized through the deployment of ICT towards the attainment of an inclusive economic growth process for all, meeting social development and poverty reduction objectives, as expressed in the United Nations Millennium Declaration.

SUSTAINABLE PRACTICES IN ICT FOR SOCIAL DEVELOPMENT AND POVERTY REDUCTION

ICT in Human Capital Development

Inequalities in access to education—especially high-quality education that prepares young people for employment opportunities in an inclusive knowledge society and to become active citizens in complex, market-driven, democratic societies—are a critical barrier to reducing poverty and increasing economic growth. Near-universal access to the Internet via low-cost networks enables teacher training, enhances student access to traditional teaching materials via Internet distribution, and allows the introduction and use of new and advanced multi-media resources and learning tools. The young generation takes readily to computers and such resources, and there is

evidence that classroom access to ICT tools can improve learning and help motivate students to stay in school.

Best practices for ICT-enhanced classroom education have been slow to emerge, in part because of the high cost of providing computers, appropriate curricula and adequate teacher training. Nevertheless, there are a number of concrete examples that show the effectiveness of widespread, small-scale experimentation and pilot projects which, coupled with careful evaluation, provide best practice ICT applications for formal and informal education (see Table 1)

ICT in Health and Social Services

The improvement in the delivery of health care services in geographically remote and rural areas is one of the most promising and clearly demonstrated applications of ICT in social development. Evidence suggests that improved health outcomes have been achieved through various applications of ICT solutions. In particular, ICT is being used in many developing countries and communities to facilitate: (a) remote consultation, diagnosis and treatment through the use of digital cameras to download images onto a computer and transfer them to doctors in nearby towns; (b) collaboration and information exchange among physicians; (c) ICT-based medical research through the use a network of satellites and ground stations to submit data for clinical trials; (d) medical training through

Table 1. Committee to democratize information and communication technology in Brazil

The <i>Comitê para Democratização da Informática</i> (CDI), a nongovernmental, nonprofit organization fosters the social inclusion of less-privileged social groups with ICT as a tool to encourage education and active citizenship. It works to create opportunities for young people to free themselves from poverty and social exclusion through the implementation of community Computer Science and Citizenship Schools (EICs). The CDI views computer literacy as a vehicle for creating employment opportunities and promoting civic participation, formal education, literacy, concern for the ecology, health, human rights, and nonviolence. The CDI also uses ICT to benefit low-income communities and institutions by assisting individuals with special needs (including the physically and mentally disabled, the visually impaired, homeless children, prisoners, and indigenous populations). CDI invests in the community's capacity to organize its own educational programs. Since its inception in Rio de Janeiro in 1995, CDI has provided support to 130 communities in the establishment of autonomous and financially self-sustaining EICs. In order to promote digital inclusion, CDI entered into partnerships with national and international philanthropic organizations, companies, government agencies and individual donors. There are currently 965 EICs using the methodology and model developed by CDI. In a recent evaluation, carried out by an external consulting group, 86 percent of the students stated that CDI schools had a positive impact on their lives (such as going back to regular schools, making new friends and staying away from drugs). The experience has been replicated throughout the world and, today, CDIs operate in Argentina, Chile, Colombia, Guatemala, Honduras, Mexico and Uruguay as well as in Angola, Japan and South Africa.

ICT-enabled delivery mechanisms; and (e) access to centralized data repositories connected to ICT networks that enable remote healthcare professionals to keep abreast of medical knowledge.

Moreover, ICT provides considerable benefits and capabilities when applied to disease prevention and response efforts during epidemics (see Table 2).

As far as the social services are concerned, it is important to mention the CETID project, a successful good practice of the State of Sao Paulo, Brazil, for the social inclusion of people with disabilities through ICT. CETID stands for “Center of Excellence in Technology and Innovation for the social inclusion of people with Disabilities”. This successful practice is the result of the Knowledge economy approach to the activities of the “Sao Paulo State Secretariat for the rights of people with disabilities”. By the simultaneous and seamless investments in Innovation, Education, ICT, Habilitating policies and Institutional regimes, the Secretariat created a virtuous cycle that led to the creation of the Center that benefits people with disabilities.

ICT in Financial Services

ICT tools can drive down transactions costs for financial services such as microfinance and a widening range of banking, insurance, and other services for low-income groups, particularly as their delivery expands beyond nonprofit groups

and becomes more widespread (see Table 3). For example, the expanded use of ICT and the Internet can reduce the transaction costs of remittances in a way that brings higher social benefits for all parties involved in these transactions. Nonetheless, there is still much to do. Transaction systems adapted to serve low-income communities via text messaging over mobile phones have been successful in Asia and Africa, and may prove useful in Latin America and the Caribbean as well. ICT offers several approaches to expanding access to electronic transactions and banking services via remote transaction devices for microfinance that work over mobile phone networks; smart cards that can store account balances, transaction histories, and positive IDs such as a fingerprint. The next generation of mobile phones and bank services are capable of conducting transactions automatically via very short-range radio, potentially turning phones into electronic wallets.

Wireless Technologies and Community-Based Communications Services

Largely for reasons of cost, most rural communities and many low-income urban communities lack effective and affordable local phone systems. These communities usually rely on a few pay phones or shared mobile phones (which are used sparingly) and most calls are made to numbers outside the community. Yet historically, where affordable lo-

Table 2. Alerta DISAMAR: an innovative disease surveillance system in Peru

Real-time reporting of disease outbreaks as well as ordering medicines and supplies via software such as the VOXIVA toolset pioneered in Peru (which is now used worldwide) has dramatically increased the efficiency of health services delivery. Such solutions combined with handheld devices assist frontline health workers in efficiently collecting and reporting patient data, and contain digital diagnostic and treatment guides for a wide variety of conditions. Alerta DISAMAR is a pilot disease surveillance program implemented in 2002 by the Peruvian Navy, which relies on novel technology from VOXIVA Inc. Alerta DISAMAR allows users to collect, store, disseminate and analyze data on diseases under surveillance, reported through any type of telephone or the Internet. Designated users receive automatic notification of selected reports via e-mail, voice mail, or SMS message. Health officials can communicate with remote health professionals using voice mail as if they were e-mails to individuals or to predetermined groups of users. The system sends timely automatic reports on the number of cases of eight of the most reported diseases compared to the previous three weeks and a monitoring of cases that do not have yet a final diagnosis. Alerta DISAMAR shows a sustained improvement in coverage, data quality, baseline incidence rates, and outbreak detection capability. Alerta is a flexible and stable program with minimal recurring costs. It is a prime example of sustainable technology transfer for improving public health in a country with limited resources. More than 18,042 health events have been reported so far using this program. In addition, 18 outbreaks have been detected that have led to recommendations for improving vector control and food procurement processes.

Table 3. PRODEM FFP's multilingual smart ATMs for microfinance in Bolivia

Bolivian PRODEM Private Financial Fund (PRODEM FFP) offers low-income communities and micro, small and medium enterprises a wide range of savings, credit and money transfer services. Its 65-branch network is the largest in the country and spans both urban and rural areas. To expand its market, improve its services in even the most remote areas and help overcome barriers such as illiteracy, the company has developed and deployed a new technology-based solution that employs smart cards, digital fingerprint recognition technology, and Smart ATMs, as well as stand-alone, voice-driven Smart ATMs in local languages with color-coded touch screens.

cal phone systems exist, typically 60 percent of all phone traffic is within the community. Thereby, a low-cost local phone system can make universal access a reality in many communities of EE (see Table 4).

The technological potential of the project has emerged, via local Wireless Fidelity Networks (WiFi) and Voice-Over-Internet (VOIP) telephony using peer-to-peer systems. For example, technologies such as Skype are having a global impact on long distance calling. In addition to making a wide range of voice-driven e-government and commercial services accessible and affordable, low-cost wireless networks, VOIP and high-capacity broadband networks, enable access to a great variety of Internet services and information via a computer or other converged device.

One of the benefits of voice-driven or voice-accessible services—especially if also made available in indigenous languages—is overcoming literacy and computer skill barriers. Other benefits also include ending rural isolation, enhanced family solidarity, increased access to information and services, improved ability to find employment and, at a community level, higher economic capacity and productivity and wider citizen participation in

democratic processes. The experience shows that this can be achieved with affordable computers and Internet access, especially if these services are delivered through local entrepreneurs or community access facilities that can assist in computer and Internet usage.

Nevertheless, not all the countries in Latin America and the Caribbean (LAC) have made VOIP legal or removed restrictions to its use. Likewise, few countries have made frequencies available for unrestricted WiFi use, permitted open competition for telecom and Internet services, or allowed community-based systems exemption from legal/natural monopolies. Beyond regulatory restrictions, the business environment in many countries still poses barriers to entrepreneurs and the creation of small businesses. There have been few pilot projects that promote best practices with rapidly emerging ICT technologies such as those cited. As a result, the LAC Region still lags behind many Asian, and even some African countries in realizing the potential development benefits from widespread ICT access and emerging wireless and no conventional communication technologies (see Table 5).

Table 4. Communications and social services in the Amazon

The Amazon Association, a Brazilian NGO, and the Solar Electric Light Fund, a nonprofit organization based in Washington, D.C., teamed up to provide basic services and economic opportunities to the Caboclo Indians by means of broadband wireless Internet. Many members of this indigenous community lack basic health care, education and economic opportunities. In the absence of phones, electricity or Internet infrastructure, solar panels are being used to power a permanent satellite uplink to a local telecenter, which was built in four days and is expected to be self-sustainable in four years. In addition, it is expected that connectivity will be extended to other communities along the Jauaperí River in a cost effective manner by using WiFi technologies and sharing costs. According to the Wireless Internet Institute, “communicating on a regular basis with the outside world has provided a tremendous psychological lift to the community.” The Amazon Association is able to stay in contact with community members via e-mail, reports on problems at the reserve, participates in decision-making, and requests supplies and medicines. Wireless technologies are an important means for outreach, empowering a community and promoting civic involvement. The project has become a source of income for local craftswomen, who can now sell their wares through the Amazon Association.
--

Table 5. Vodacom's community cell phones in South Africa

Vodacom Community Services started operating under a 1994 government mandate to provide telecommunications services in disadvantaged communities in South Africa. Vodacom developed an innovative way to meet this mandate, via entrepreneur-owned and operated phone shops. These have provided affordable communication services to millions of South Africans, and empowered thousands of previously disadvantaged individuals with income-generating opportunities and lasting business skills. The Community Services program now provides over 23,000 cellular lines at over 4,400 locations throughout South Africa. By investing so extensively in disadvantaged communities, Vodacom is also investing in its own future by creating a distribution channel for its services, in addition to a well-recognized brand name. Though full returns will materialize in the future, the company is confident that its investments will pay off directly through increased sales and indirectly through a stronger, better-connected South African economy.

A WIRELESS KNOWLEDGE SOCIETY¹

Table 5 illustrates what Professor C. K. Prahalad of the Michigan Business School poses in his book *The Fortune at the Bottom of the Pyramid: Eradicating Poverty Through Profits*, which highlights the collective purchasing power potential of the world's 4 to 5 billion poorest people (the Majority). He urges efforts to promote an enabling business climate that goes beyond the promotion of social corporate responsibility as means to foster local well-being, while generating a strong revenue and profit potential base in the long term. Moreover, a recent study by the London Business School found that, in a typical developing country, an increase of ten mobile phones per 100 people boosts GDP growth by 0.6 percentage points. The study concludes that wireless solutions are concrete examples of "technologies that help people help themselves" (Avgerou, Ciborra, Cordella, Kallinikos, & Smith, 2005) by: promoting a multi-stakeholder partnership framework delineating the effective participation of the public sector and civil society, while creating the incentives for socially responsible private investment; strengthening the provision of ICT-based public social services and promoting social inclusion, while maintaining the role of the private sector as the main source of innovation; stimulating macroeconomic growth by facilitating access to knowledge and information through increased connectivity and appropriate ICT solutions for marginalized and lower-income populations, thereby tapping a strong market potential; designing and adopting long-term ICT investment frameworks in human development.

By removing barriers to the entry of new and lower-cost technologies (such as fixed wireless) and of new communication and ICT-based services (such as Voice-Over-Internet phone service), and ensuring open competition, it will be possible to increase entrepreneurial activity and expand private sector investment. Nevertheless, the direct involvement of the public sector continues to be fundamental in sharing and expanding service into remote areas and to serve as a catalyst for the effective delivery of social services. This also reduces risk for private enterprises and helps comply with commitments to serve certain regions, communities and marginalized or disadvantaged groups, while also allowing market forces to trigger actions and operate effectively.

The World Summit on the Information Society (WSIS) asserted the importance of the use of information and communication technologies (ICTs) for the achievement of the Millennium Development Goals (MDGs).²

In particular, wireless Internet technologies deserve particular attention not only because of their importance to the development process, but also because of their ability to reduce the costs of providing ICT access and ICT-enabled services to underserved areas.

The cost of providing access to voice and data networks using wireless technologies, even in rural areas, approached in 2005 US\$300 per subscriber, compared to US\$1,000 for fiber optic or copper networks in urban areas. In addition, wireless technology does not imply significant sacrifices in quality or throughput. Furthermore, while copper or fiber optics usually cost between US\$20,000 and US\$40,000 per kilometer of

connectivity, hundreds of kilometers of wireless connectivity can be provided for US\$50,000 (Best, 2003). This cost advantage (which is widening with increasing global use of wireless technologies and resulting economies of scale) can be used to connect rural areas in a cost-effective and even self-sustainable fashion.

Given its unprecedented cost advantages wireless technologies promise more than a cost-effective way of providing access to underserved populations. For governments, wireless technologies offer a way to deliver valuable services (such as health and education), as well as government services, to citizens living in remote areas.³ In some cases, rural populations can pay for the capital costs of any necessary equipment as well as the recurring costs of Internet connection. For example, poor rural residents of Tamil Nadu, India, have funded telecenters for as little as US\$3 in revenue per day by using a wireless technology similar to WiMax called corDECT. An analysis of Latin America and the Caribbean shows that the same could be accomplished in the region for approximately US\$3 to US\$14, covering between 100 and 500 households. Recurring monthly costs for wireless Internet connectivity for education purposes in Sucre, Colombia, amount to US\$5 per student per year. The latent demand for communication services from poor rural populations could bring financing and sustainability even without government aid. The following examples from selected EE countries provide a broad picture of possible successful wireless applications:

Providing Education in Sucre, Colombia

The state of Sucre is marked by high poverty and illiteracy rates and, until some years ago, uncontrolled violence. After security was re-established, population increased by 30 percent without a commensurate increase in the state budget. The *Corporación Politécnica Nacional de Colombia*, a partnership between the government

and the private sector, was created in an attempt to control high school dropout rates and improve the cost-effectiveness of education.⁴

Through a grant (that is considered a social investment) from the Ministry of Education for approximately US\$220,000, the *Politécnica* connected 11 schools across the state using wireless technologies, provided new computers, and trained local staff at each school in hardware and LAN maintenance. Recurring costs of the Internet connection, software upgrades, ISP uplink costs, and network infrastructure amount to US\$5 per student per year, which, so far, has been funded by the state government.

The initial response was not as good as expected. The availability of Internet did not by itself promote its use, and neither did the services offered when the project was launched. However, after adding customized services, including course management software as well as human and capital resources, software usage increased. Students and professors have reacted very positively to the Internet connection and the software, and they now routinely create websites and participate in classes via videoconferences from state universities. This experience shows that simply providing Internet connections is not enough to promote its usage or address development problems. In addition to providing connectivity, investments should be made to develop customized services.

Voice and Services in the Dominican Republic

A village-area network (VAN) was installed in Bohechio, Dominican Republic, in March 2001 as part of the LINCOS project headed by the Costa Rican Foundation for Sustainable Development. The rural community of Bohechio, one of the least developed communities in the country, is situated in a mountain range and has a population of approximately 7,000 people (Best, 2003).

The VAN covers an area of approximately one square kilometer and took three days to install.

The cost was less than US\$20,000 (prices fell significantly in recent years) for six computers, two telephones, a multifunction fax/scanner/ printer, a cash machine, an environmental testing lab, FM radio station broadcasting ability, a big-screen TV, and a telemedicine unit. A VSAT provides uplink connectivity via satellite link and, using an 802.11b network, gives access to wireless devices, including PDAs and voice- over-IP telephones at a radius of up to 1 kilometer and at speeds of up to 11 Mbit/second (Best, 2003).

Applications currently provide assistance to schools, a medical clinic and farmers. The VoIP solution and associated IP-based telephones installed around the town compete with the telecom provider, offering rates of US\$0.18 per minute on calls to the city, compared to US\$0.30 on fixed lines. INDOTEL, the national telecommunications regulatory authority, has relaxed its regulations so that spectrum used in the project is valid and has offered significant support and partnership (Best, 2003).

E-Commerce in Ecuador

The “world’s first Wi-Fi linked e-payments network” was created in the Mall of San Marino in Guayaquil, Ecuador, through a partnership between the US e-payments equipment and services provider, Verifone, and the wireless equipment provider MediaNet, a subsidiary of D-Link. The wireless network eliminated multiple dial-up phone lines and Internet access points, using instead a central uplink and IP- based Wi-Fi to connect Verifone’s “multimodal network access point-of-sale terminals” in the stores. This resulted in lower charges for long- distance calls for merchants and 24-hour/7-day availability. The speed at which payments were processed and data transferred improved by 350 percent, with an average of four seconds per transaction. Verifone’s marketing director for Latin America and the Caribbean stated that the company expects “wireless-enabled POS terminals to become more mainstream over the next several years, particu-

larly when the solution is delivered with the right partners” (Burger, 2004).

PDAs and E-Health in Uganda

Malaria, tuberculosis and AIDS have ravaged many African countries and have been largely responsible for the current health crisis in the continent. SATELLIFE, a Boston non-profit organization, has been using WideRay servers, rugged plastic boxes equipped with a GSM cell phone and Strong ARM processor in an effort to help ameliorate the critical situation.

Equipped with wireless-capable PDAs in rural areas in Uganda, health workers, many of whom had never used a computer before, are able to relay information to the Ugandan Ministry of Health on drugs needed by the population and local epidemics. The International Development Research Centre (2003) notes that, prior to this innovation, reports were hand written and requests for medicines took months to reach those able to fulfill them. The introduction of the PDAs has significantly reduced the time lag to receive needed medicines in rural areas, and has also made it possible to relay information back to healthcare workers in the field regarding the most cost-effective approaches or latest treatments for endemic diseases. In addition, reporting errors have been drastically reduced, improving the quality of the information available to policymakers. As a result of improved information on the health status and burden of disease of the population, important improvements can be expected in the management of national health budgets.

The application of these new low-cost technologies in Africa shows the type of improvements in social well-being that can be brought about by the use of wireless technologies. Like the example from the project in the Dominican Republic, this project demonstrates the success of using simpler technologies (i.e., PDAs with wireless) in cases where those involved may not know how to use computers.

E-Government in India⁵

Through e-government services a government can stay connected to remote populations and, with fee-based servicing, can ensure a steady stream of revenue so that local communities are able to run self-sustainable access points.

The Gyandoot project in the Dhar district of Madhya Pradesh established a set of telekiosks (small, self-service telecenters) that bring e-government services to the local population. The government has promised a one-week response time on local grievances for a fee of approximately US\$0.20. This has become the third most popular service, behind market prices and job availability (Best & Maclay, 2002). Additional services provided include assistance in obtaining legal certificates for land deeds, e-mail, village auctions, on-line matrimonial listings, and educational guides. Although most of the current network is based on wired infrastructure, wireless technology is being used to reach isolated and remote areas.

LESSONS LEARNED AND RECOMMENDATIONS

Based on the learning derived from the preceding examples and others, the following actions are suggested for the promotion and use of ICT for development:

- The creation or strengthening of the institutional capacity with the participation of the public, private and civil society sectors, in order to promote and foster the dissemination and use of ICT through programs and initiatives that build a participatory information and knowledge society.
- Raising awareness and creating the conditions for the design and implementation of pilot initiatives and the replication of best practices in priority areas including ICT for social development and poverty reduc-

tion, ICT for governance, and ICT for economic growth, among others.

- Undertaking the actions necessary to promote a digital inclusion for all.
- Creating the organizational conditions for the identification of concrete and comprehensive national ICT investment plans which effectively delineate the participation of the various sectors and stakeholders.
- Implementing and/or strengthening national and regional regulatory frameworks to, among others, promote competition and loosen restrictions on the telecommunications market, allow for the expansion of media convergence, and promote and facilitate research and technological innovation.

The Multi-Lateral Development Banks (MDBs) can play a relevant role in supporting ICT for development programs. Indeed, the MDBs are aware that they must continuously update their financial and technical assistance instruments to effectively adapt to the demand of their borrowing member countries; they have to support activities to provide and facilitate access to efficient ICT tools for the exchange of information, ideas and knowledge; and they have to support the deployment of ICT to promote a more inclusive socioeconomic development that extends benefits to the poor.

In particular, the MDBs could establish collaboration agreements with the technology-based private sector by taking into consideration:

- The extensive need for new and modern technologies in pilot and demonstrative projects, and whose results and applications can be replicated in operations of a wider scale.
- The goal to make the private sector from both, donor and borrowing member countries of the Banks, as active stakeholders in the development process through ICT.

- The aim to provide opportunities for the private sector to participate in the efforts of the countries towards social development and poverty reduction, while creating value added opportunities and the expansion of ICT-based services.
- The example of the Inter-American Development Bank (IADB) that created the “Facility in Information and Communication Technology for Development”, which set the institutional, financial and legal means for the Bank to effectively partner with leading private sector firms and civil society institutions with the goal of contributing to the attainment of the MDGs by the countries of the Region and, specifically, through the value added ICT solutions described above.

The MDBs have the opportunity to take a leadership role in shaping the enormous potential of wireless ICT in the EE. One overarching principle is that the MDBs should not in any way replace or hinder the significant efforts of the private sector. There are, therefore, four primary initiatives related to wireless technologies that the MDBs can lead: convening policy discussions across sectors, streamlining existing country operations through pilot projects, generating and sharing best practices, and fostering a regional policy consensus.

FUTURE RESEARCH DIRECTIONS

The hypothesis that ICT can have significant positive impact upon socio-economic development indicators is based generally on three propositions: (1) speed, because ICT are based on increasingly sophisticated digital instruments, and because rapidity of processing is one of their principal attributes, the use of ICT to deliver development services increases the speed at which those services become available to development beneficiaries; (2) reach, because ICT can make use of massive communication means, the number of potential

beneficiaries is greatly augmented; and (3) cost, as a function of a larger customer base and faster deployment rate, lower costs in the delivery of development services can be expected of the use of ICT as compared to other (traditional) instruments.

However, there is a significant gap between current and available knowledge regarding optimal conditions to make the most of the introduction of ICT to the development tool-box. In particular, the following research priorities have been identified:

- Impact and end-result (increase in income; number of direct and indirect beneficiaries; reduction in costs of service to the service provider, etc.) that depict more accurately the differential impact of the inclusion of ICT in a development context, capable of being adapted to different cultural and socio-economic settings;
- Development of cooperative arrangements to achieve critical mass in infrastructure (communications, R&D) investment available to the poorest countries, where conditions are not present to maximize the usefulness of ICT investment;
- Reliable, low cost, patent systems, particularly for communal, historic, or otherwise not individual innovation products;
- Effective communication mechanisms between supply and demand of ICT to cover the needs and acquisition ability of special interest groups and different income levels;
- ICT public policy needed to compensate market failures, and differential needs for low and medium income developing economies.

CONCLUSION

Examples of past programs, projects and activities where ICT has been instrumental to improve the conditions of beneficiary populations allow the following conclusions regarding the impact of ICT in a development context:

- ICT substantially lowers transaction costs in most economic activities;
- ICT increases the organizational capacity of private firms and their ability to manage data and information to base their decisions;
- ICT improves market effectiveness by broadening buyers and sellers' access to information –for example information on market price; volume, location and characteristics of demand; supply options;
- Investment in R&D explains up to 75% of the differences in total factor productivity growth rates, once externalities have been factored-in (Griliches, 1979, as cited in IADB, 2014), with productivity growth identified as a result, not a cause, of R&D investment (Rouvinen, 2001, as cited in IADB, 2014).
- In poorer countries, investment in R&D can have negative results due to the absence of trained human capital, apt scientific infrastructure, capable private sector, and effective public sector incentives.

In fact, in recent years, the EE have experienced a significant progress in the penetration of ICT, exceeding the growth achieved in the sector by other regions in the world. Nevertheless, the current levels continue to fall short of those in industrialized nations with respect to overall connectivity and widespread access, in addition to the adoption of ICT products and services by the private sector (especially by SMEs), as a means to improve their productivity, market access and overall competitiveness.

Priority actions have been identified in several fields: public policy; private sector investment; external financing; research; dissemination; among others. Success stories and case studies must attain wider audiences in order to make potential benefits of ICT for development a reality for lower income communities and countries. There is ample room for collaboration between higher and lower income economies in today's globalized world,

where innovation and opportunity are not limited to one side of the economic spectrum, but are available to all as long as adequate instruments and farsighted leadership are at hand.

REFERENCES

Avgerou, C., Ciborra, C., Cordella, A., Kallinikos, J., & Smith, M. (2005). *The role of information and communication technology in building trust in governance: toward effectiveness and results*. Washington, DC: Inter-American Development Bank.

Best, M. (2003). The Wireless Revolution and Universal Access. In *Trends in Telecommunications Reform*. International Telecommunication Union.

IADB. (2014). *Innovation*. Washington, DC: Science and Technology Sector Framework Document, Inter-American Development Bank.

Prahalad, C. K. (2005). *The fortune at the bottom of the pyramid*. Wharton School Publishing.

United Nations. (n.d.). *United Nations Millennium Declaration*. Retrieved from <http://www.un.org/millennium/declaration/ares552e.htm>

ADDITIONAL READING

Avgerou, C., Ciborra, C., Cordella, A., Kallinikos, J., & Smith, M. (2005). *The role of information and communication technology in building trust in governance: toward effectiveness and results*. Washington, D.C.: Inter-American Development Bank.

Awotwi, J. E. (2011). The Ghana Community Network (GCNet): Implementation Challenges. In D. Piaggese, K. J. Sund, & W. Castelnovo (Eds.), *Global Strategy and Practice of E-Governance: Examples from Around the World* (pp. 209–230). New York: IGI Global. doi:10.4018/978-1-60960-489-9.ch012

- Aziz, M. (2011). Implementing ICT for Governance in a Post-Conflict Nation: a Case study of Afghanistan. In D. Piaggese, K. J. Sund, & W. Castelnovo (Eds.), *Global Strategy and Practice of E-Governance: Examples from Around the World* (pp. 185–208). New York: IGI Global. doi:10.4018/978-1-60960-489-9.ch011
- Best, M. (2003). The Wireless Revolution and Universal Access. In *Trends in Telecommunications Reform*. International Telecommunication Union.
- Hsieh, J. J., Rai, A., & Keil, M. (2011). Addressing Digital Inequality for the Socioeconomically Disadvantaged Through Government Initiatives: Forms of Capital That Affect ICT Utilization. *Information Systems Research*, 22(2), 233–253. doi:10.1287/isre.1090.0256
- IADB. (2006). *Sustaining Development for all: Expanding Access to Economic Activity and Social Services*. Washington, D.C.: Inter-American Development Bank.
- Kaplan, W. A. (2006). Can the Ubiquitous Power of Mobile Phones be Used to Improve Health Outcomes in Developing Countries? *Globalization and Health*, 2(2), 9. doi:10.1186/1744-8603-2-9 PMID:16719925
- Kozma, R. (2005). National policies that connect ICT-based education reform to economic and social development. *Human Technology*, 2(1), 117–156. doi:10.17011/ht/urn.2005355
- Piaggese, D., Sund, K. J., & Castelnovo, W. (Eds.). (2011). *Global Strategy and Practice of E-Governance: Examples from Around the World*. New York: IGI Global. doi:10.4018/978-1-60960-489-9
- Prahalad, C. K. (2005). *The fortune at the bottom of the pyramid*. Wharton School Publishing.
- Puri, S. (2007). Integrating Scientific with Indigenous Knowledge: Constructing Knowledge Alliances for Land Management in India. *Management Information Systems Quarterly*, 12(31), 355–379.
- Raghunathan, A. (2005). *The economic advantage of wireless infrastructure for development*. Washington, D.C.: Inter-American Development Bank.
- Rashid, A. T., & Elder, L. (2009). Mobile Phones and Development: An analysis of IDRC-Supported Projects. *EJISDC*, 2(36), 1–16.
- Rouviren, P. (2002). Characteristics of Product and Process Innovators. Some Evidence from the Finnish Innovation Survey. *Applied Economics Letters*, 9(9), 575–580. doi:10.1080/13504850110108102
- Seinmueller, E. (2010). Economics of Technology Policy. In B. H. Hall & N. Rosenberg (Eds.), *The Economics of Innovation*. Amsterdam: North Holland.
- United Nations Millennium Development Goals. (n.d.). Retrieved from <http://www.un.org/millenniumgoals>
- Vital Wave Consulting. (2009) mHealth for Development: The Opportunity of Mobile Technology for Healthcare in the Developing World, UN Foundation-Vodafone Foundation Partnership, Washington, D.C. and Berkshire, UK.
- Vitro, R., & Piaggese, D. (Eds.). (2005). *The knowledge economy in development: perspectives for effective partnerships*. Washington, D.C.: Inter-American Development Bank.
- WSIS (2003) *Declaration of Principles, Building the Information Society: a global challenge in the new Millennium*, World Summit on the Information Society, WSIS-03/GENEVA/DOC/4-E, A7

KEY TERMS AND DEFINITIONS

Information and Communication Technology for Development (ICT4DEV): The universe of internet-related applications that in the different sectors of a Knowledge Society (KS), support boosting the living conditions of people, particularly in EE.

Knowledge Economy (KE): An economy based on knowledge as its principal asset. In a KE, knowledge is the new currency and ICT is a tool to enabling its attainment. Particularly important for emerging economies (EE).

Knowledge Society (KS): A society where tacit and explicit knowledge represent its main backbone. In a KS, local and indigenous knowledge are valued; imported knowledge is adopted whenever it could be adapted to local conditions; and new knowledge can be created to serve societal purposes.

Sustainable Development (SD): The set of economic, social, environmental, and institutional conditions that guarantee a constant or increased flux of benefits for the members of a given society. For societal development to be sustainable, such conditions would have to tend towards equity, fair distribution of benefits, and durability in the use of natural and financial resources.

Sustainable Development Goals (SDG): The new set of SD United Nations goals to be attained by 2030 in order to improve the living conditions worldwide. In the SDG, ICT4DEV is an enabling common denominator supporting tool.

ENDNOTES

¹ This chapter largely draws from the work of Mr. Adithya Raghunathan: *The Economic Advantage of Wireless Infrastructure for Development*. Mr. Adithya Raghunathan was a consultant with the Information and Communication Technologies for Develop-

ment Division (SDS/ICT), IADB, when this paper was written. The author is grateful to Edward Malloy and Jeff Cochrane of USAID, Beatriz Acosta of Corporación Politécnica Nacional (Colombia), and Colin Maclay and Michael Best of the Harvard Berkman Center for Internet and Society who graciously agreed to be interviewed for this report. He also acknowledges the comments of Daniel Aghion of the Wireless Internet Institute. In addition, the author recognizes the dedicated and detailed analysis provided by this paper's reviewers: María Paula Duque Samper, Ministry of Communications, Colombia; Harry Silverstone, EION Wireless; and Guillermo Castillo and Maximilian Spiess, SDS/ICT.

² A declaration was issued during the first phase of meetings, which were held in Geneva, Switzerland on December 10 to 12, 2003.

³ This includes such remote areas as the Andes or the Amazon. See case study on the Amazon Association Project.

⁴ Sixty-three percent of eligible students enrolled in primary school and only 8 percent did so in secondary school. Approximately 1 percent of students finished secondary school and went on to post-secondary education.

⁵ Based on information from the article "Gyandoot: A Community-Owned, Self-Sustainable and Low-Cost Rural Intranet Project" at <http://gyandoot.nic.in/gyandoot/intranet.html>.

Communities of Practice as a Source of Open Innovation

K

Diane-Gabrielle Tremblay
University of Quebec, Canada

INTRODUCTION

In this entry, we first define this new form of learning and knowledge management that is communities of practice. We present the concept as described by the creators of the concept but also comment on the role of these communities in organizational learning or informal learning. We follow with some of the results, centering on the conditions of success and challenges that emerge, as well as limits in the learning and sharing process, which are often underestimated. We highlight some results from a research on communities of practice in Canada, in particular the main conditions and challenges of such new modes of knowledge creation and management, which don't always work automatically. We compare these results to other recent research. Research clearly confirms that participants' commitment and motivation in the project, dynamism and continuity of leadership, organizational support and recognition of employees' involvement are the key elements in a community of practice, and they can contribute to open innovation.

COMMUNITIES OF PRACTICE

The term 'communities of practice' was first used by Brown and Duguid (1991), by Lave and Wenger (1991), and finally by Wenger (1998; Wenger et al., 2002, 2000). It refers to the idea of sharing information and knowledge within a small group, as well as to the value of informal learning for a group and an organization. As is usually the case today, we consider people use technologies (com-

puter, cell phone, ipad, etc.) to exchange with each other, but also to keep track of some information and knowledge the group wants to stock. Wenger et al. (2002) describe a community of practice as a group of participants who:

Don't necessarily work together every day, but they meet because they find value in their interactions. As they spend time together, they typically share information, insight, and advice. They help each other solve problems. They discuss their situations, their aspirations, and their needs. They ponder common issues, explore ideas, and act as sounding boards. They may create tools, standards, generic designs, manuals, and other documents – or they simply develop a tacit understanding that they share. However they accumulate knowledge, they become informally bound by the value that they find in learning together. This value is not merely instrumental for their work. It also accrues in the personal satisfaction of knowing colleagues who understand each other's perspectives and of belonging to an interesting group of people. Over time, they develop a unique perspective on their topic as well as a body of common knowledge, practices, and approaches. They also develop personal relationships and established ways of interacting. They may even develop a common sense of identity. They become a community of practice. (pp. 4-5)

In the 90s, observers mainly studied informal communities that were created spontaneously in a workplace. However, over the years, there has been increasing interest in creating and cultivating such communities in workplaces (Swan et al.,

DOI: 10.4018/978-1-5225-2255-3.ch435

2002; Wenger et al., 2002). Also, more recently, there has been more and more interest in seeing companies and organizations in general as a group of communities of practice and more and more interest in the leadership and empowerment dimension (Cordery et al., 2015), as well as on the impact on innovation (Müller & Ibert, 2015). Many of these communities are teleworking communities or distributed communities (Friberger & Falkman, 2013), often active in an agile and lean environment (Paasivaara & Lassenius, 2014) that use information and communication technologies, and this was the case in the communities we studied.

The following definitions help us to better understand what this concept actually means (Mitchell, 2002):

- Communities of practice are people who share a concern, a set of problems or a passion about a topic, and deepen their knowl-

edge and expertise in this area by interacting on an ongoing basis

- A group whose members regularly engage in sharing and learning, based on their common interests

Wenger et al. (2002) as well as Mitchell (2002), among others, indicate that communities of practice take on various forms, and Table 1 highlights the differences that exist between types of communities. In the cases we studied, communities were of the structured distributed type, most of them being formally supported by one organization, a few being inter-organizational, but all having to do with their work activity and not personal interests, as is more often the case in the informal type of community. Over the years, inter-organizational or inter-cluster interactions (Cusien & Loubaresse, 2015) have become more important, and there is as much interest in these types of communities as in those organized within a single firm, while

Table 1. various forms of communities of practice

Aspects	Informal	Supported	Structured
Objective	Provide a discussion forum for people with affinity of interest or needs within their practice	Build knowledge and capability for a given business or competency area	Provide a cross-functional platform for members who have common objectives and goals
Affiliation	Self-joining or peer invited	Self-joining, member invited or manager suggestion	Selection criteria outlined Invited by sponsors or members
Sponsorship	No organizational sponsor	One or more managers as sponsors	Business unit or senior management sponsorship
Mandate	Jointly defined by members	Jointly defined by members and sponsor(s)	Defined by sponsor(s) with endorsement of members
Organizational support	General endorsement of communities of practice Provision of standard collaborative tools	Discretionary managerial support in terms of resources and participation Supplemented array of tools and facilitation support	Fully-fledged organizational support on the same basis as organizational segments Budget allocation as part of business plans
Infrastructure	Most likely meets face-to-face; primary contact Has a means of communication for secondary contact	Uses collaborative tools Meets face-to-face on a regular basis	Uses sophisticated technological infrastructure to support collaboration and store knowledge objects generated in the community Highly enabled by technology
Visibility	So natural that it may go unnoticed	Visible to colleagues affected by the community's contribution to practice	Highly visible to the organization through targeted communication efforts that are stewarded by sponsors.

Source: Davel and Tremblay (2011)

the boundaries of the firm becomes an important issue (Contu, 2014).

Much existing literature centers on face to face communities (Gherardi & Nicolini, 2000), but many communities actually function in a context of distance or telework (Hildreth & Wright., 2000). This brings us to the issue of virtual communities of practice, which are more and more common in a global environment. In our view, this virtual dimension is an important aspect of communities of practice theory, especially in the global context.

At first, authors mainly studied informal communities that were created spontaneously in a workplace. However, over the years, there has been increasing interest in the creation of such communities in workplaces (Wenger & Snyder, 2000), and even in the creation of teleworking communities that use information and communication technologies. Task interdependence appears crucial (Cordery et al., 2015).

The advantages of communities are said to be the following: informal diffusion of relevant knowledge, exchange of knowledge between peers and, as a result, improvement of innovation and productivity.

Much of the literature centers on face to face communities, while many function in a context of distance or telework, which brings us to talk of virtual communities of practice, which is what we studied in our project, and is more and more common in a global environment. In our view, this virtual, or rather, distance dimension is an important dimension, especially in the global context, and it requires more detailed analysis, which is why this research was done.

Also, much of the literature does not take into account the temporal dimension, but it does appears essential in analyzing organizations, since mutual engagement appears to be crucial (Gau, 2014) and this engagement tends to evolve over time. The most detailed model of the evolution of communities of practice was presented by Wenger et al. (2002), but in our view it again presents a very normative portrait of communities of practice, from which reality often departs.

Wenger et al. (2002) and Bourhis and Tremblay (2004) define five stages. At the beginning, the community is an informal network, a potential community. It then unites itself and acquires maturity, and then momentum, and becomes productive (Mitchell, 2002) until at some point, an event makes it essential for the community to change or renew itself. Again, this seems a little normative in comparison to the real life of communities, and we wanted to better understand the variables which influence the life of communities.

Let us now turn to conditions of success of these communities. First, group work always requires a number of conditions and the communities of practitioners (CoPs) are not an exception to this rule. On the contrary, these conditions are certainly even more important in the CoP context, since participants must in principle share tacit knowledge, collectively build up knowledge, and solve production or service problems. In this context, in our view, the social relations between actors and demographic characteristics cannot be neglected, although they tend to be in the literature on communities of practice. We therefore turned to the literature on collaborative learning to dig into the reality more deeply and put forward new questions to the participants of the communities.

One of the main conditions mentioned in the literature on collaboration and collaborative learning concerns the engagement and commitment of participants to the task or the community, as well as the interest and motivation of individuals to work together as a group (Gau, 2014). In the Community of practice (CoP) literature, some authors refer to a “joint enterprise” to describe the mission or common objective that participants give to a CoP. However, few authors have determined how to foster this commitment, which appears to be taken for granted regardless of the context and the social relations of work, whereas in reality this is not the case. Second, many authors emphasize the importance of having a shared set of resources or what could be referred to as “common baggage,” or common language, in order to facilitate exchanges and avoid misunderstandings

and conflicts. We now turn to some results from our study of CoPs in order to support some of the elements put forward here.

RESULTS

The results are derived from an action research on a dozen communities of practitioners (CoPs) conducted under the aegis of the *Centre francophone d'informatisation des organisations* (CEFRIO). In fact, seven CoPs have actively participated in the research. One hundred and eighty (180) participants answered questionnaires on starting up a CoP and slightly less than 100 participants answered evaluation questionnaires six months later. In addition, focus groups and recordings of critical incidents in each of the communities were also conducted so as to better understand the dynamics of each of the CoPs. We will focus on the aspects related to learning and training, paying particular attention to the conditions and challenges that emerge from our results.

Let us first present a few demographic characteristics of our respondents. The majority of respondents, that is 60% (105 out of 173 respondents to this question), were aged from 35 to 49. However, there is a good differentiation in ages, which permits some statistical analyses, as we will see later. There were 61% women against 39% men (105 women, 68 men out of 173 respondents).

Attainment of Objectives

Although the objectives of the communities of practitioners studied differed (Tremblay, 2008), they were mainly aimed at learning through exchange and collaboration. From this perspective, it is interesting to note how the objectives have evolved over time. When the communities were starting up, the objectives identified by the participants were usually related to exchange and sharing of information and knowledge, better utilization of delocalized resources, as well as the creation of a collective memory - objectives which actually

pertain to learning through interaction, something common to most CoPs (Lee et al., 2015).

It must first be stressed that in our research, the majority of respondents to this question had mixed feelings about the success and usefulness of the community, even though they think that it has had a positive impact on the work climate. Thus, although the participants do not appear to be enthusiastic, collaboration within the CoPs seems to be rather positive (Bourhis & Tremblay, 2004), and networking and co-learning are indeed an important objective of CoPs (Jimenez-Zarco et al., 2015).

However, after a few months of work in a virtual CoP, the achievement of objectives seemed to be uneven. In fact, although certain CoPs felt that they had achieved their objectives (Tremblay, 2008, 2004a), this was not so true of other CoPs. Perhaps it was still too soon to assess the achievement of objectives since, unlike project teams or groups, CoPs are not supposed to have a specific schedule and they have to learn new operating modes in a short time.

Concerning the partial achievement of the objectives of CoPs, there are various possible reasons for this, including the frequent change of CoP leader, the loss of interest on the part of management or participants, or the lack of time for participation. However, it must be stressed that developing learning and experimenting with a new problem-solving approach, which were not always among the objectives considered to be the most important at first, seemed to have been relatively well achieved by a number of CoPs and these forms of learning are greatly appreciated by the participants. There appear to be criteria and conditions for CoPs to function and clearly, this type of arrangement cannot be transferred anywhere or globalized without taking into account these considerations.

It must be stressed that all of the CoPs operated with a knowledge-sharing telesoftware. The participants were either not very familiar with the software or had to more or less master it in a few months, depending on how easy or difficult it was

for them to use this software and the time—which is generally limited—that they had. The use of software such as Knowledge Forum or Lotus Notes, which was different in each case, allowed CoP participants to exchange messages. These were then grouped together on a space and could be reviewed and re-organized according to the themes discussed in the exchanges. In principle, this is how virtual (i.e., teleworking) communities must jointly develop knowledge.

We analyzed the data on success or attainment of objectives according to various demographic variables, but only two (gender and age) came out significantly in some of the analyses. For various reasons, often lack of variance in the respondents, the other variables tested did not show up as significant: level of schooling, professional category, language but could be the object of other research in the future. There were 178 respondents in the first phase of the research (1st questionnaire) and 106 in the second phase, six months later, which essentially addressed issues of impact and results. The success of the CoP was evaluated in different ways, amongst which the attainment of the strategic and operational objectives of the CoP.

Sources of Satisfaction

Satisfaction or dissatisfaction with the CoP are an important issue, as they impact on engagement of participants (Gau, 2014). In general, participants appreciated the pertinence of the topics addressed in the exchanges in relation to their work, the collaboration between members, the solving of work problems, the establishment of consensus, group work, and the development of new skills. They were slightly more critical of the quality of the exchanges, which was viewed differently by the various CoPs. It must, however, be noted that younger participants seemed to appreciate all these aspects more than participants aged 50 or over. More in-depth analysis is needed to determine whether age alone explains this finding or whether other variables might be more important in the explanation, such as the specific local context,

the topic, the leadership (Paasivaara & Lassenius, 2014) or the level of engagement (Gau, 2014).

Participants were also asked to assess different aspects of their experience. It was clear that the most interesting aspect for participants was learning from other people as well as exchanging and sharing information and knowledge in a supportive environment and this is found regularly over the years as a main source of interest and engagement (Paasivaara & Lassenius, 2014; Gau, 2014). Nevertheless, it is interesting to note that the majority of our participants thought that they had learned more from others than had contributed to the exchanges themselves. It thus seems that there was a deficit in active participation by CoP members, since many of them remained somewhat on the periphery of the community's central core, in what is referred to as "peripheral participation." (more detail in Tremblay, 2004, 2005a, 2008).

It must be noted that women's involvement in the project was often slightly higher than that of men, at least according to their own evaluation. More research needs to be done however on this issue of involvement and participation, according to gender, since other elements of context (organizational culture, passionate leadership, decision making process, the interest in the topic, etc.; cf. Paasivaara & Lassenius, 2014) might also explain the stronger involvement of some participants.

Sources of Dissatisfaction

The main sources of dissatisfaction identified by our participants relate to the lack of recognition of participation by the employer, sometimes also the lack of peer recognition, and in particular the too often limited time (given the objectives), spent on the community's activities. In fact, the majority of participants were not released from other tasks to participate in the CoP and this activity therefore ate up their working time (Tremblay, 2004). However, the most satisfied CoP in this regard is made up of a group of some 20 female health professionals, whose CoP was not supported by their employer but by a professional association,

and thus the participants used their personal time to participate (Tremblay, 2004). Once again, motivation and engagement to the project emerged as the key variables in the success of this CoP, as other authors have also shown (Gau, 2014). Participants were willing to put personal time in a project because the knowledge acquired and the achievements seemed to be worth their while.

In contrast, in other cases, the achievements were apparently too minor or not sufficiently visible or satisfactory. This negative view was confirmed by the fact that the majority did not think that the CoP activity would be recognized in their performance evaluation, career progression, and skills assessment. However, it seemed that participants were generally more optimistic about the recognition of their learning by colleagues, although this did not yield concrete results in career terms.

It must be noted that most of the participants in the CoPs we studied did not know each other well beforehand, but were designated to participate in these CoPs. Therefore not all of them were volunteers and as trust-building is important, this may have been an issue here. Moreover, one CoP in which most participants did not know each other at all – composed of the female health professionals – was the most successful case in our view, which means that other factors (mutual engagement in this case, *cf.* Gau, 2014) can compensate for prior acquaintance. Nevertheless, the latter is deemed to be important by many authors, as it is considered to be a source of trust and greater collaboration between participants. Indeed, it was found that although prior acquaintance can make it easier to collaborate in certain CoPs, it is not a sufficient condition for them to achieve their objectives. Thus, although being in the habit of collaborating can result in trust, which is generally considered to be essential to collaboration and learning, it is evident that participants need additional motivation to move the CoP forward and achieve its objectives. Moreover, it should be noted that women spent twice as long as men on

CoP activities, on average, one hour versus half an hour for men.

These observations contribute to the existing literature on communities of practice, since it nuances the importance of prior acquaintance and organizational support, often considered as determinant characteristics of the success of communities of practice. Our research tends to highlight the importance of commitment, personal and mutual engagement (as in Gau, 2014) and interest co-learning (Jimenez-Zarco et al., 2015). It may also be that the most successful groups had less difficulty working from a distance or teleworking (Paasivaara & Lassenius, 2014; Tremblay, 2003, 2002).

Our research had some limits of course, but recent research seems to center on the same questions, which should be explored further: Which groups are more motivated by this form of co-learning (Jimenez-Zarco et al., 2015)? How can one develop trust and passion for the topic (Paasivaara & Lassenius, 2014) in order for people to be willing to share knowledge? To what extent is a passionate leadership essential for success (Paasivaara & Lassenius, 2014)?

CONCLUSION

To conclude, a number of factors related to the conditions and challenges associated with CoPs are summarized in order to identify those which would help promote the wider use of these co-learning practices (Jimenez-Zarco et al., 2015).

The participants in our research mentioned that participants' engagement was considered to be a crucial factor in the success of CoPs, as was also observed by Gau (2012). In fact, the most successful CoP was one in which the participants' engagement was indeed important (Tremblay, 2004). However, other factors can play a role in explaining the more mixed success of other cases: for example, the frequent change of leaders and the lack of dynamic leadership (Paasivaara and

Lassenius, 2014), the degree of decentralization of the firm (Dumitru and Enachescu, 2015). These various factors must be taken into account when developing co-learning through communities of practitioners. One of the communities which functioned the best was headed by a female health professional who had been hired specifically for this task, and therefore invested the necessary time in the project (Tremblay, 2005a, 2005b) and provided dynamic leadership (Paasivaara & Lassenius, 2014).

It was also shown that the support offered to participants by the organization can be a factor of success. However, our results indicated that most of the participants would not necessarily have wanted more resources or training (in conflict management, communication or problem solving) even though few had received the training. Therefore, our findings suggest that training and support resources are not as important in the success of CoPs as is indicated in the literature, while leadership and engagement appear crucial (Paasivaara & Lassenius, 2014; Gau, 2014). The engagement of participants appears to sometimes compensate for the lack of organizational support. Participants indicated that their organization's interest in the CoP had not increased over time and this also seems to be one of the challenges associated with the medium- and long-term viability of CoPs. Again, one of the most successful cases was one composed entirely of female health professionals which was outside of the employer organizations and therefore this indicates that organizational support is not necessarily the most important source of success, personal engagement appearing to be as determinant (Gau, 2014; Tremblay, 2008).

In any case, although relatively new, this CoP formula offers interesting prospects for learning, but we can see that it cannot be generalized without considering various dimensions including leadership and engagement, and various characteristics of the community need to be taken into account,

since they may have an impact on success and attainment of objectives.

FUTURE RESEARCH DIRECTIONS

Research clearly needs to be pursued on issues such as leadership, engagement, goal-setting, decision-making processes, as well as the conditions and factors of success. Also the differences between formal and informal CoPs should be analyzed in more detail, as there has been relatively little comparative work on this (Tebogo et al., 2015). There has been work on informal CoPs in but it has been strongly concentrated in the field of education; there has been much less work in management on the issues and challenges that CoPs pose in terms of leadership and engagement, much of the literature being of a normative nature. This is why we wanted to contribute by doing some empirical research to test to what extent learning and sharing was actually occurring and to try to identify the factors that make it happen. Over the years, various authors have contributed to the literature (Paasivaara & Lassenius, 2014; Gau, 2014), but research needs to be pursued on these two crucial factors that are leadership and engagement. In our view, there also needs to be more research on inter-cluster or inter-organizational interactions, to go beyond the boundaries of the firm.

ACKNOWLEDGMENT

This research was initially conducted under the aegis of the Cefrio (www.cefrio.qc.ca), but was updated under the aegis of the Canada Research Chair on Socio-Organizational Challenges of the Knowledge Economy (www.teluq.ca/chaireeco-savoir) and the CURA Center on Work-Life Articulation over the Life Course (www.teluq.ca/aruc-gats).

REFERENCES

- Bourhis, A., & Tremblay, D.-G. (2004). *Les facteurs organisationnels de succès des communautés de pratique virtuelles*. Québec: Cefrio. http://www.cefrio.qc.ca/media/uploader/Facteurs_organisationnels.pdf
- Brown, J. S., & Duguid, P. (1991). Organizational learning and communities of practice: Toward a unified view of working, learning, and innovation. *Organization Science*, 2(1), 40–57. doi:10.1287/orsc.2.1.40
- Contu, A. (2014). On boundaries and difference: Communities of practice and power relations in creative work. *Management Learning*, 45(3), 289–316.
- Cordery, J. L., Cripps, E., Gibson, C. B., Soo, C., Kirkman, B. L., & Mathieu, J. E. (2015). Managing a New Collaborative Entity in Business Organizations: Understanding Organisational Communities of Practice Effectiveness. *The Journal of Applied Psychology*, 96(6), 1234–1245.
- Cusin Julien & Élodie Loubaresse. (2015). L'interclustering. De la communauté de pratique aux réseaux d'innovation. *Revue française de gestion*, 1(246), 13–39.
- Davel & Tremblay. (2011). *Formation et apprentissage organisationnel. La vitalité de la pratique*. Québec: Télé-université/Presses de l'Université du Québec.
- Dumitru, D., & Enăchescu, V. (2015). Communities of Practice as a Mean for Decentralization. *Procedia - Social and Behavioral Sciences*, 187, 752–756.
- Frativelli, F., Negri, F., & Cori, E. (2015). Social media communities of practice: Reputation at risk. *International Journal of Management Cases*, 17(4), 102–119.
- Friberger, M. G., & Falkman, G. (2013). Collaboration processes, outcomes, challenges and enablers of distributed clinical communities of practice. *Behaviour & Information Technology*, 32-6, 519–531.
- Gau, W.-B. (2014, February 21). A Study on Mutual Engagement in Communities of Practice. *Procedia: Social and Behavioral Sciences*, 116, 448–452.
- Gherardi, S., & Nicolini, D. (2000). The organizational learning of safety in communities of practice. *Journal of Management Inquiry*, 9(1), 7–18. doi:10.1177/105649260091002
- Hildreth, P. M., Kimble, C., & Wright, P. (2000). Communities of practice in the distributed international environment. *Journal of Knowledge Management*, 4(1), 27–37. doi:10.1108/13673270010315920
- Jiménez-Zarco, A. I., González-González, I., Saigí-Rubió, F., & Torrent-Sellens, J. (2015). The co-learning process in healthcare professionals: Assessing user satisfaction in virtual communities of practice. *Computers in Human Behavior*, 51, 1303–1313.
- Kislov, R. (2014). Boundary discontinuity in a constellation of interconnected practices. *Public Administration*, 92(2), 307–323.
- Lave, J., & Wenger, E. (1991). *Situated Learning. Legitimate Peripheral Participation*. Cambridge, MA: Cambridge University Press. doi:10.1017/CBO9780511815355
- Lee, L., Reinicke, B., Sarkar, R., & Anderson, R. (2015). Learning Through Interactions: Improving Project Management Through Communities of Practice. *Project Management Journal*, 46(1), 40–52.
- Mitchell, J. (2002). *The potential for communities of practice*. John Mitchell and Associates.

Müller, F. C., & Ibert, O. (2015). (Re-)sources of innovation: Understanding and comparing time-spatial innovation dynamics through the lens of communities of practice, Review Article. *Geoforum*, 65(October), 338–350.

Swan, J., Scarbrough, H., & Robertson, M. (2002). The construction of communities of practice in the management of innovation. *Management Learning*, 33(4), 477–496. doi:10.1177/1350507602334005

Tebogo Keikotlhaile, R., Ekambaram, A., Halvorsen, S. B., & Klakegg, O. J. (2015, July 2nd). Formalising the Informal? – Finding a Balance between Formal Teams and Communities of Practice in a Project-based Organisation. *Procedia: Social and Behavioral Sciences*, 194, 105–114.

Tremblay, D.-G. (2002). Work organization and satisfaction in the context of telework. *Revista de administrao de empresas*, 42(3), 54–65.

Tremblay, D.-G. (2003). Telework: A New Mode of Gendered Segmentation. Results from a study in Canada. *Canadian Journal of Communication*, 28(4), 461–478.

Tremblay, D.-G. (2004). Communities of practice: A gendered analysis of their functioning and results based on a Canadian Research. *Proceedings of the Communication au colloque annuel de la Society for the Advancement of Socio-Economics*. Retrieved from www.sase.org

Tremblay, D.-G. (2005a). Virtual Communities of Practice: Explaining different effects in two organizational contexts. *Canadian Journal of Communication*, 30(3), 367–382.

Tremblay, D.-G. (2005b). Virtual communities of practice: an organizational form that can support and foster knowledge? In *The Passion for Learning and Knowing. Proceedings of the 6th International Conference on Organizational Learning and Knowledge* (vols. 1-2). Trento: University of Trento e-books. Retrieved from http://eprints.biblio.unitn.it/archive/00000828/01/Volume_I_02.pdf

Wenger, E. (1998). Communities of Practice – Learning as a social system. *Systems Thinker*, 9(5).

Wenger, E., McDermott, R., & Snyder, W. (2002). *Cultivating Communities of Practice. A guide to managing knowledge*. Boston: Harvard Business School Press.

Wenger, E., & Snyder, W. (2000). Communities of Practice: The Organisational Frontier. *Harvard Business Review*, 78(1), 139–145.

KEY TERMS AND DEFINITIONS

Communities of Practice: The term was first used by Brown and Duguid (1991) and by Lave and Wenger (1991) and then by Wenger et al. (2002). It refers to the idea of sharing information and knowledge within a small group, as well as to the value of informal learning for a group and an organization.

Telework: Work from a distance of the main office, either at home or in telework centers, generally using technologies.

Indigenous Knowledge Systems

Osarumwense Iguisi

University of Benin, Nigeria

Osaro Rawlings Igbinomwanhia

University of Benin, Nigeria

INTRODUCTION

The importance of cultural values to management practices in Africa has become increasingly obvious in recent years as many expectations of African organizations and institutions created and managed along lines of Western textbooks and models have not achieved the expected results of sustainable economic development and growth. In Africa, we have very limited knowledge about its cultural values and the consequences it poses for African intellectuals and management practice in African organizations.

While the African intellectuals are very knowledgeable about accepted theories and models of the Western world, knowledge about cultural values of their society is limited. The African intellectuals are not equipped enough to understand the obligations imposed on them by Western cultural values in which they have been socialized and the traditional environment in which they were born and raised, thus, making their ability to contribute something original to the development of their society limited.

Because of failure of the Westernized African managers to identify and take advantage of the 'growth-positive' cultural values of their society for effective management practice that the relevancy of western management theories and models utilized in training managers in Universities and business schools to managing organizations in Sub-Saharan Africa comes into question.

Africans have ways of exercising power and authority at the workplace, ways of motivating

and rewarding people to make them work harder. Neither the institutions nor the political structures put in place by the colonizers and the so-called African intellectuals to Westernized Africa have respected these indigenous knowledge structures, but much of them have survived in the traditions and cultural values of the African people. However, unlike in Europe and most part of Asia, the attempted modernization or Westernization after independence has completely neglected the native cultural traditions and tried to transfer or transplant ready-made western management theories and models to traditional African soil. The results of these transformations, in most cases, have been disappointing.

The objective of this paper is to draw attention to the relevance of cultural values in management philosophy with the purpose of contributing to a culturally viable and appropriate theories and practice of management in Sub-Saharan Africa. Management practice in Africa requires identification of "growth-positive" and "growth-negative" culture-based factors.

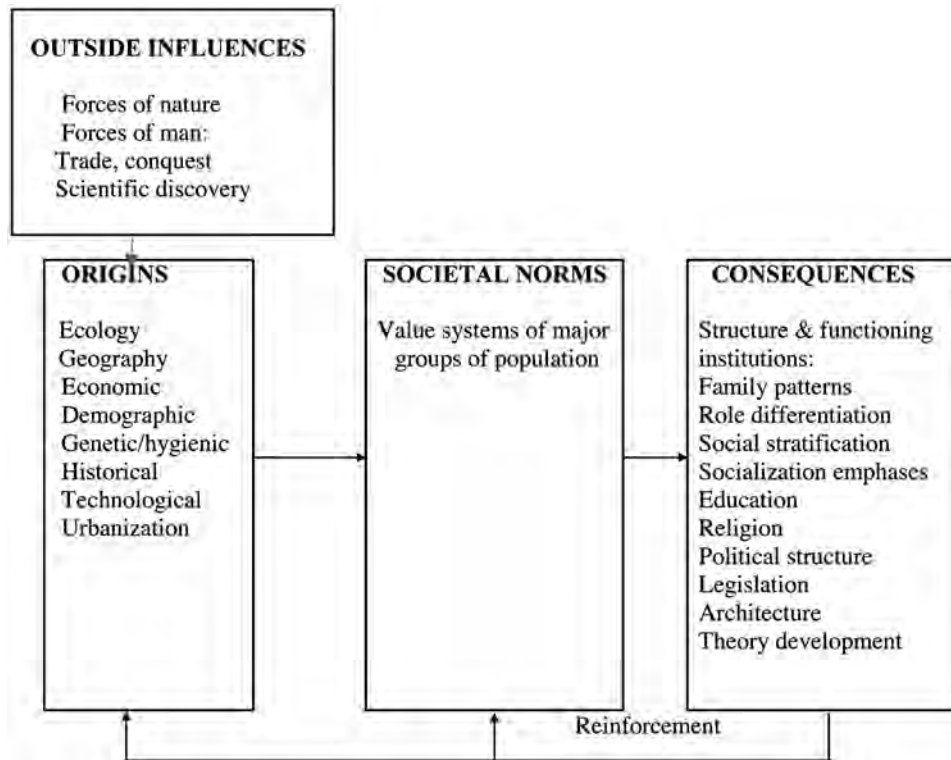
BACKGROUND

Culture Stabilization Patterns

The model of Figure 1, taken from Iguisi and Hofstede (1993), indicates how we assume culture patterns in a country to stabilize themselves through feedback loops, but also to change under the influence of outside forces.

Figure 1. The stabilization of culture patterns

Source: Igusi and Hofstede (1993, p. 3)



The system of this model in Figure 1 implies that one cannot understand one element—such as, management practice philosophy within the local environment—without its societal and cultural value context.

In the center is a system of societal norms, consisting of the value systems shared by major groups of the population. Their origins are in a variety of ecological factors (in the sense of factors affecting the physical environment). The societal norms have led to the development and pattern maintenance of institutions in society with a particular philosophy, structure and way of functioning. These include the family, education systems, politics, and legislation. These institutions, once they have become facts, reinforce the societal norms and the ecological conditions that led to them. According to Hofstede, in a relatively closed society, such a system will hardly change

at all. Institutions may change, but this does not necessarily affect the societal norms; and when these remain unchanged, the persistence influence of a majority value system patiently smooths the new institutions until their structure and functioning is again adapted to the societal norms. Change comes mainly from the outside, through forces of nature (change of climate, silting up of harbors) or forces of man (trade, colonization, scientific discovery) (Hofstede, 1980). The arrow of outside influences is deliberately directed at the origins, not at the societal norms themselves. It is believed that norms change rarely by direct adoption of outside values, but rather through a shift in ecological conditions: technological, economical, and hygienic. In general, the norm shift will be gradual unless the outside influences are particularly violent (Hofstede, 1980a).

The Place of Cultural Values in African Philosophy Discourse

The value of a thing, be it an object or a belief, is normally defined as its worth. Just as an object is seen to be of a high value that is treasured, our beliefs about what is right or wrong that are worth being held are equally treasured. According to Idang (2015) a value can be seen as some point of view or conviction which we can live with, live by and can even die for. This is why it seems that values actually permeate every aspect of human life. For instance, we can rightly speak of religious, political, social, aesthetic, moral, cultural and even personal values. We have observed elsewhere that there are many types and classifications of values. As people differ in their conception of reality, then the values of one individual may be different from those of another. Life seems to force people to make choices, or to rate things as better or worse as well as formulate some scale or standard of values. Depending on the way we perceive things we can praise and blame, declare actions right or wrong or even declare the scene or objects before us as either beautiful or ugly. Each person, as we could see, has some sense of values and there is no society without some value system (Idang, 2009, p. 4).

Whether we are aware of it or not, the society we live in has ways of daily forcing its values on us about what is good, right and acceptable. We go on in our daily lives trying to conform to acceptable ways of behaviour and conduct. Persons who do not conform to their immediate society's values are somehow called to order by the members of that society. If a man, for instance, did not think it wise to make honesty a personal value, and it is widely held by his immediate society that truth telling is a non-negotiable virtue, it would not be long before such an individual gets into trouble with other members of his society. This shows that values occupy a central place in a people's culture. It forms the major bulwark that sustains a people's culture, making it more down-to-earth and real. Elsewhere, we have seen African culture

as "all the material and spiritual values of the African people in the course of history and characterising the historical stage attained by Africa in her developments" (Idang, 2009, p. 142). This simply means that there is a peculiar way of life, approach to issues, values and world views that are typically African.

Based on cultural considerations, some forms of behaviour, actions and conduct are approved while others are widely disapproved of. To show the extent of disapproval that followed the violation of values that should otherwise be held sacred, the penalty was sometimes very shameful, sometimes extreme. Antia (2005) says that "what a people hold to be true, right or proper with regard to those things explains much of the cultural traits by which they become identified". What Antia calls "traits" here can as well be called values; and Etuk (2002) writes that "no group of people can survive without a set of values which holds them together and guarantees their continued existence".

Values, whether "growth-positive" or "growth-negative", occupies a very wide area in the discipline of philosophy. To show the fundamental importance of values, it is regarded as a core area in philosophy, together with knowledge and reality. When we are dealing with actions that a people see as good or bad, right or wrong, praiseworthy or blame-worthy, we are dealing with the aspect of value theory that rightly falls under ethics or moral philosophy. Having seen the centrality of values to African culture and any culture for that matter, it can be stated that the "growth-positive" and "growth-negative" values of culture are what give it uniqueness and identity. Let us now look at African philosophy discourse.

African Philosophy Discourse

Anthropologist and sociologists have come across multi-farious patterns of family, kinship, and political, religious, economic and technological organisations. The variety extends to morals, beliefs, art and languages. It is now generally accepted that societies have divergent ways of expressing their

thoughts about themselves and their world and of organizing their activities. The non-industrialized societies with the most primitive technology have in their explanations of natural processes and the supernatural, organised their lives on these assumptions, which may seem outlandish to the industrialized or industrializing men who have their own way of conceptualising reality by separating the natural from the supernatural.

Several African writers such as Mbiti (1969), Abimbola (1976) Oruka (1990, 1998), and Gbadegesin (1991, 1998) have all attempted to elaborate on some aspects of African philosophy though they have not clearly stated this attempt. Onwuejeogwu (1978, 1997), extensive empirical fieldwork in the study of African divination and social action, demonstrates that African traditional philosophy is based on the relationship linking the past, present and future. Accordingly, this conception is expressed in an African genealogical structure that models the collective existence and actions of contemporaries (the living), in terms of predecessors (ancestors), and successors (the unborn). In this system of beliefs, grandfathers under certain metaphysical conditions are expected to re-incarnate in their grandchildren: thus the past is linked with the future through the present (Onwuejeogwu, 1978).

In the African traditions, it is the community that defines the person as a person. "Ubuntu" as a translation of the Xhosa expression "Umuntu ngumuntu ngabantu" means the person is a person through other persons, and this expresses a typical African conception of a person. Mbigi (1995) defined "Ubuntu" as the sense of solidarity or brotherhood, which arises among people within marginalized or disadvantaged groups. They acknowledged that it is not culturally specific to African people or communities but may be manifest elsewhere and they restate the collective shared experience of African peoples. Jackson (2005) and Mangaliso (2001) provide literal translation of Ubuntu as "I am who I am through others". Karsten and Illa (2005) see Ubuntu as humaneness and pervasive spirit of caring and

community, harmony and hospitality, respect and responsiveness that individuals and groups display for one another. Ubuntu provides a strong philosophical base for the community concept of management (Khoza, 1994). Mbigi (1997) has listed the following relevant principles of Ubuntu as the spirit of unconditional African collective contribution, solidarity, acceptance, dignity, stewardship, compassion and care, hospitality and legitimacy. Ubuntu is an African worldview that is rooted and anchored in people's daily lives.

On the basis of an extensive review of literature in African studies, Ahiauzu (1986) has shown that there exists a great deal of similarity in the general ways of life of black Africans. For example, he provided empirical evidence that there clearly exists an African thought-system, quite different from that of other societies. According to Ahiauzu, African thought system draws largely from proverbial social thought, while theoretical thinking in traditional African society is mystical based mainly on the actions of gods, ancestral spirits and other kinds of spirits (Ahiauzu, 1986). Numerous studies of African cosmologies have highlighted the influences of gods and different kinds of spirits on the life of the African (Horton, 1962, 1954; Uchendu, 1965; Barber, 1981; Lienhardt, 1961; Middleton, 1960; Fortes, 1959; Biko, 1984, 1998; Wiredu, 1995; Prinsloo, 1998). If it is accepted, therefore, that the natives of Africa can be identified by certain common characteristics in their ways of life, as suggested by a host of scholars of African studies (Blyden, 1908; Donohugh, 1935; Busia, 1967; Sithole, 1959; Wilson, 1936), particularly an identifiable unique system of social thought (Otite, 1978; Adesanya, 1958; Horton, 1967; Maquet, 1972; Ahiauzu, 1986; Kopyoff, 1997; Fanon, 1997), then there might exist an indigenous African Management philosophy.

African Management Philosophy

In the last few years a body of literature has arisen in response to Africa's relegation to the margins of global considerations of management practice.

This field of study has become known as African management philosophy defined as:

The practical way of thinking about how to effectively run organisations be they in the public or private sectors on the basis of African ideas and in terms of how social and economic life is actually experienced in the region. Such thinking must be necessarily interwoven with the daily existence and experience in Africa and its contextual reality. (Edoho, 2001, p. 74)

Proponents of African management philosophy argue that Africa's effort to engineer authentic development will continue to be unsuccessful until indigenous management systems are established and institutionalised (Anyansi-Archibong, 2001; nken & Jones, 1997; Edoho 2001; Kamoche, 2000; Mangaliso, 2001; Mbigi, 1997, 2005; Ngambi, 2004). The call for indigenous approaches to management falls within the broader call for an African Renaissance that seeks to reclaim the identity of Africans (Makgoba, 1999; Mbeki, 1998; Mudimbe, 1988; Nzelibe, 1986). It is also consistent with post-colonial theory, which calls for the colonised to reclaim a culture of their own, a history of their own, aesthetics of their own, all based on an essence of their own, free from and independent of the images of the 'Other' (Said, 2002; Prasad, 1997; Ghandi, 1998; Spivak, 1990).

If Africa was better managed in the past, what went wrong and how can it be reclaimed? Colonialism is identified as the culprit for the often corrupt and ineffective management of organisations in many African countries. The underlying belief is that if indigenous African management can be reclaimed and re-institutionalised in Africa, there would be a positive effect on resolving significant problems of poverty, economic stagnation and development.

Proponents of African management philosophy look to the history of Africa and the presence of indigenous knowledge systems that resulted in effective management during the pre-colonial era (Iguisi 2012, Edoho, 2001; Kiggundu, 1993;

Ngambi, 2004; Nzelibe, 1986). Describing management systems during the pre-colonial period remains problematic because of the scarcity of written documentation of such systems. However, this has not prevented scholars from offering descriptions of African management systems. In offering such descriptions, scholars draw heavily from the literature in African studies and writings of African historians (e.g.; Mazrui, 1998; Iguisi & Deva 2014).

How are the basic dimensions of African management and leadership described in these writings? While writers (Ahiauzu, 1997; Mutabazi, 2002; Mbigi, 2005) often point to the vast diversity in Africa, the focus is on offering generalised descriptions of African management. Whereas Western management thought is said to advocate Eurocentricism and individualism, African management thought is said to emphasize traditionalism, communalism, cooperative teamwork, and mythology. Traditionalism has to do with the adherence to accepted customs, beliefs and practices that determine accepted behaviour, morality and characteristics of individuals in African society.

African management philosophy is also said to be characterised by a strong belief in the individual's relation to nature and supernatural beings, and connections between the individual and ancestors (Mbigi, 1997, 2005). Nzelibe (1986, p.12) argues that the continuity from the material to the spiritual is the universal basis of African management philosophy.

FUTURE RESEARCH DIRECTIONS

Consequences for Organisations in Africa

One remarkable scientific comment made by Ahiauzu (1985b) in his empirical study of African workers in Nigeria and a host of other African countries, was that when ordinary Africans are observed in traditional workplaces, such as vil-

lage farms, local communal work organisations, and self-employment where African indigenous managerial and work organisational practices are largely adopted, the Africans are normally seen to be relatively relaxed and hardworking. But when they come to town and become employed in government public service organisations, industrial or other modern work organisations, their attitudes to work and general work behaviour falls far below expectation.

Adopting the foregoing view as a frame of reference, and pondering on the behaviour and general reactions of the Africans in modern organisations, one important question comes to mind. The question is: is it logical for management scholars and social scientists, to attempt to obtain a genuine understanding and explanation of the work behaviour of Africans in modern organisations, by adopting constructs and theoretical frameworks developed by social scientists in the Western world? Mitchell (1967) has made a statement that assists us in addressing the question. The author argues that social scientists:

Are prone to think their work is the outcome of a play of free intelligence over logically formulated problems. They may acknowledge that their ideas have been influenced by their reading and the teaching which they were wise enough to choose, but they seldom realize that their free intelligence has been moulded by the circumstances in which they have grown up; that their minds are social products; that they cannot in any serious sense transcend their environment. (pp. 36-37)

Mitchell's argument suggests that Western management and theories are the products of the minds of persons, which have unconsciously been moulded by the circumstances in which those theorists have grown up.

There is the belief in African philosophy that all the provinces of meaning of an African in his everyday life are interrelated. Because of the strong relationship between the work and non-work realities of Africans, and because of

inherent inability to achieve a complete leap of consciousness when shifting from non-work to work realities, the major structural features that characterise the self-perception and relational processes of Africans in their non-work world, cannot be ignored while developing management and organisational behaviour theories to be adopted in understanding and explaining their behaviours in modern organisations.

The challenge in African management discourse is to reflect on the different cultural and philosophical discourses, management theories and research methodologies to present an African methodological perspective upon which to develop appropriate theoretical constructs and frameworks to be used in understanding and explaining the values and behaviours of African managers in modern organisations, bearing in mind the nature of the elements that structure their work and non-work worlds.

CONCLUSION

It is truly disappointing to find that, despite available evidences (see Frank, Hofstede, & Bond, 1991; Hofstede, 2005, 2011; Boon, 2005; Iguisi, 1994, 1998, 2014), towards the role of cultural values in management for African scholars, leaders, managers, and international development agencies that have been prescribing Western formula for African development to have not recognized that culture does have greater part to play in their consideration for management and economic development issues in Africa.

This paper has discussed the issues of the role of culture in management and a number of issues in Western management theories and assumptions that are culturally conditioned. Effectiveness and appropriateness within a given culture, and judged according to the values of the culture, call for management theories and assumptions reflective of the local culture. The point we are making here comes out more clearly if one draws from both oral tradition and written historical records.

Before the advent of colonial administration, the old African villages and towns had effective public administrative mechanism, which the village and town heads, chiefs and kings administered. The use of anthropological concepts in this context will help in the development of appropriate organizational behavior and effective management practice congruous with the local cultures and traditions.

African scholars and practicing managers aware of the possibilities of their cultural values for modern management practice can hold their head erect and put their best foot forward at any feat they may attempt. They need not assume the role of black-Europeans or Americans to be good scholars or managers; there is enough strength in African cultures and traditions to enable them perform wonders in any modern organization. African scholars, managers, and practitioners' conscious of their cultural values, more importantly, the 'growth-positive' cultural value factors in their societies can perform best in modern organizations.

REFERENCES

- Abimbola, W. (1976). *Ifa: An Exposition of Ifa Literary Corpus*. Oxford University Press.
- Adesanya, A. (1958). Yoruba Metaphysical Thinking. *Journal of West African Studies*, 5, 25–37.
- Ahiauza, A. I. (1985a). "The Influences Shaping Work Behaviour in African Organisations" Organisation. *Forum (Chicago, Ill.)*, 1(2).
- Ahiauza, A. I. (1985b). Towards a Diagnostic Approach to Motivating the Nigerian Worker. In *Managing Nigeria's Economic System*. Heinemann.
- Ahiauza, A. I. (1986). The African thought-system and the work behaviour of the African industrial man. *International Studies of Management & Organization*, 16(2), 37–58.
- Ahiauza, A. I. (1997). Type 'A' Method for Management Research in Africa: Enunciating a culture-specific paradigm. In O. Igusi (Ed.), *Culture Dimensions to Appropriate Management in Africa*. Paris: UNESCO and E-AMARC Publications.
- Antia, O. R. U. (2005). *Akwa Ibom Cultural Heritage: Its Incursion by Western Culture and its Renaissance*. Abnny Publishers.
- Anyansi-Archibong, C. B. (2001). Towards an African-oriented management theory. In F. M. Edoho (Ed.), *Management challenges for Africa in the twenty-first century: theoretical and applied perspectives* (pp. 63–72). Westport, CT: Praeger.
- Barber, K. (1981). How Man Makes God in West Africa: Yoruba Attitudes Towards the Orisa. *Africa*, 51(03), 724–745. doi:10.2307/1159606
- Behrend, H. (1988). *Die Zeit Geht Krumme Wege*. Frankfurt: Campus.
- Biko, S. (1984). The definition of black consciousness. In *The African Philosophy Reader*. Routledge London.
- Blunt, P., & Jones, M. (1992). *Managing organizations in Africa*. Berlin, New York: Walter de Gruyter. doi:10.1515/9783110850031
- Blyden, E. W. (1908). *African Life and Customs*. London: C M Phillips.
- Boon, E. M. (2005). Knowledge Systems and Traditional Versus Modern Social Security Systems in Africa: A Case Study of Ghana. In E. K. Boon & L. Hens (Eds.), *Indigenous Knowledge Systems and Sustainable Development: Relevance for Africa*. Delhi: Kamla-Raj Enterprise.
- Boon, M. (1996). *The African Way: The Power of Interactive Leadership*. Johannesburg: Zebra Press.
- Busia, K. A. (1967). *Africa in Search of Democracy*. London: Routledge and Kegan Paul.

- Cameron, J. (1975). Traditional and Western education in Mainland Tanzania: an attempt at synthesis. In G. N. Brown & M. Hiskett (Eds.), *Conflict and Harmony in Education in Tropical Africa* (pp. 350–362). London: Allen & Unwin.
- Donohugh, A. C. L. (1935). Essentials of African Culture. *Africa.*, 8(03), 329–339. doi:10.2307/3180417
- Edoho, F. M. (2001). Management in Africa: The quest for a philosophical framework. Westport, CT: Academic Press.
- Etuk, U. A. (2002). *Religion and Cultural Identity*. Ibadan: Hope Publication.
- Fanon, F. (1997). On National Culture. In R. R. Grinker & C. B. Steiner (Eds.), *Perspectives on Africa: A reader in culture, history & representation*. Blackwell Publishers.
- Forde, C. D., & Jones, G. I. (Eds.). (1954). *African Worlds*. Oxford.
- Franke, R. H., Hofstede, G., & Bond, M. H. (1991). Cultural Roots of Economic Performance: A Research Note. *Strategic Management Journal*, 12(S1), 165–173. doi:10.1002/smj.4250120912
- Gabin, K. (1997). *Urban Life and Traditional Models: A Study of the Social Networks in a Secondary Town*. Côte d'Ivoire-the Example of Abogville.
- Gandi, L. (1998). *Postcolonial history: A Critical Introduction*. Edinburgh, UK: Edinburgh University press.
- Hofstede, G. (2001). *Culture Consequences* (2nd ed.). Thousand Oaks, CA: Sage.
- Hofstede, G. (2005). *Culture Consequences: Comparing Values, Behaviour, Institutions and Nations*. Sage.
- Hofstede, G. (2011). *Dimensionalizing Cultures: The Hofstede Model in Context*. Sage.
- Hofstede, G., & Bond, M. H. (1988). The Confucius connection: 'From cultural roots to economic growth. *Organizational Dynamics*, 16(4), 4–21. doi:10.1016/0090-2616(88)90009-5
- Horton, R. (1962). African Traditional Thought and Western Science. *Africa*, 37(01), 50–71. doi:10.2307/1157195
- Idang, G. E. (2009). *The Mind-Body problem in African Culture. From Footmarks to Landmarks on African Philosophy*. Lagos: Obaroh and Ogbinaka Publishers.
- Idang, G.E. (2015). African culture and values. *Phronimon*, 16(2).
- Iguisi, O. (1997). *The Role of Culture in Appropriate Management and Indigenous Development in Africa*. Paper delivered at the African seminar on Culture Dimensions to Appropriate management and Sustainable Development in Africa.
- Iguisi, O. (2007). Management and Indigenous Knowledge Systems: An Analysis of Motivational Values Across Cultures. In E. K. Boon & L. Hens (Eds.), *Indigenous Knowledge Systems and Sustainable Development: Relevance for Africa*. Delhi: Kamla-Raj Enterprice.
- Iguisi, O. (2009). Motivation-Related Values across Cultures. *International Journal of African Business and Management*.
- Iguisi, O. (2012). *Cultural Dynamics in African Management Practice: Leadership, Motivation, Recruitment and Promotion*. LAP LAMBERT Academic Publishing.
- Iguisi, O. (2014). Cultural Vualues and Motivation in Nigeria Work Settings. In U. Hacioglu & H. Dincer (Eds.), *Globalization and Governance in the International Political Economy*. IGI Publishing. doi:10.4018/978-1-4666-4639-1.ch018
- Iguisi, O., & Hofstede, G. (1993). *Industrial Management in an African Culture*. Maastricht: IRIC University of Limburg Press.

- Ijere, N. O. (1988b). *Management Potential of Self-Help Organizations*. Centre for Rural Development and Cooperatives. Nsukka, Nigeria: University of Nigeria.
- Jackson, T. (2005). *Hybridisation of people Management Systems: Western MNCs in Africa*. London: Routledge.
- Kamoche, K. (2002). Introducing human resource management in Africa. *International Journal of Human Resource Management*, 13(7), 993–997. doi:10.1080/09585190210131258
- Karsten, L., & Illa, H. (2005). Ubuntu as a key African management concept: Contextual background and practical insights for knowledge application. *Journal of Managerial Psychology*, 20(7), 607–620. doi:10.1108/02683940510623416
- Kaseke, E. (1997). Social Security in Eastern and Southern Africa Organizations, Issues and Concept in Modern and Traditional Arrangements. *Journal of Social Development in Africa*, 121(2), 39–47.
- Khoza, R. (1994). The need for an Afrocentric approach to management. African Management, Knowledge Resources (PTY) Ltd.
- Khoza R. (2001), *African leadership challenge*. Acceptance speech at the UNISA Leadership in Practice Award.
- Kiggundu, M. (1993). The Challenge of Management Development in Sub-Saharan Africa. In *Managing Organizations in Africa*. Readings, Cases, and Exercises. Berlin: de Gruyter.
- Kopytoff, I. (1997). Ancestors as Elders in Africa. In *Perspectives on Africa: A reader in culture, history, & representation*. Blackwell Publishers.
- Lamb, D. (1985). *The Africans* (Revised edition). New York: Random House.
- Lienhardt, G. (1961). *Divinity and Experience: The Religion of the Dinka*. London: Oxford University Press.
- Makgoba, W. (1999). *African renaissance*. Cape Town: Mafube Publishing.
- Mangaliso, M. P. (2001). Building competitive advantage from ubuntu: Management lessons from South Africa. *The Academy of Management Executive*, 15(3), 23–32. doi:10.5465/AME.2001.5229453
- Maquet, J. (1972). *Africanity: The Cultural Unity of Black Africa*. London: Oxford University Press.
- Mazrui, A. (1998). Africa between Nationalism and Nationhood: A Political Survey. *Journal of Black Studies*, 23–44.
- Mbeki, T. (1998). *Africa: The time has come*. Cape Town: Tafelberg.
- Mbigi, L. (1997) Ubuntu: The African Dream in Management. In *Africa: Knowledge Resources*. Academic Press.
- Middleton, J. (1954). Some Social Aspects of Lugbara Myth. *Africa*, 24(03), 189–199. doi:10.2307/1156423
- Middleton, J. (1960). *Lugbara Religion: Ritual and Authority among an East African People*. London: Oxford University Press.
- Mitchell, W.C. (1967). *Types of Economic Theory*. New York: Kelley.
- Mudimbe, V. Y. (1988). *The invention of Africa: Gnosis, Philosophy, and the order of knowledge*. Indiana University Press.
- Mutabazi, E. (2002). Preparing African Leaders. In C. B. Derr, S. Roussillon, & J. Boumais (Eds.), *Cross-cultural approaches to leadership development*. Westport, CT: Quorum Books.
- Ngambi, H. (2004). African Leadership: lesson from the chiefs. In T. A. Meyer & I. Boninehhi (Eds.), *Conversations in leadership: South African Perspectives* (pp. 107–132). Johannesburg: Knowledge Resources.

Nzelibe, C. O. (1986). The Evolution of African Management Thought. *International Studies of Management & Organization*, 16(26), 6–16. doi:10.1080/00208825.1986.11656427

Onwuejeogwu, M. A. (1978). *A Symbolism Among the Nri: A System of Social Action*. Unpublished PhD thesis.

Oruka, H. O. (1998). Indigenous thinkers and modern debate on African philosophy. Leiden: Brill. In *The African Philosophy Reader*. Routledge.

Otite, O. (1978). The Study of Social Thought. In O. Otite (Ed.), *Themes in African Social and Political Thought*. Enugu: Fourth Dimension Publishers.

Prasad, A. (1997). *The colonizing consciousness and representations of the order: a postcolonial critique of the discourse of oil*. Thousand Oaks, CA: Sage.

Prinsloo, E. D. (1998). Ubuntu culture and participatory management. In P. H. Croetzee & A. Roux (Eds.), *The African Philosophy Reader*. London: Routledge.

Said, E. (2002). Imaginative geography and its representations: Orientalising the oriental. In P. Essed & D. T. Goldberg (Eds.), *Race critical theories* (pp. 15–37). Malden, MA: Blackwell Publishers.

Schultz, M. (1992). Postmodern pictures of culture. *International Studies of Management & Organization*, 22(2), 15–35. doi:10.1080/00208825.1992.11656578

Sithole, N. (1959). *African Nationalism*. Cape Town: Oxford University Press.

Spivak, G. C. (1990). *The post-colonial critic*. New York: Viking.

Takyi-Asiedu, S. (1993). Some socio-cultural factors retarding entrepreneurial activity in Sub-Saharan Africa. *Journal of Business Venturing*, 8(2), 91–98. doi:10.1016/0883-9026(93)90013-U

Tshiyembe, M. (1998). La science politique africainiste et le statut théorique de l'État Africain: Un bilan négatif. *Politique Africaine* (Paris, France), 71.

Uchendu, V. C. (1965). *The Igbo of Southeast Nigeria* Holt. New York: Rinehart and Winston.

Wilson, G. (1936). An African Morality. *Africa*, 9(01), 75–99. doi:10.2307/1155242

Wiredu, K. (1995). The moral foundations of an African culture, Person and community. In *The African Philosophy Reader*. Routledge.

KEY TERMS AND DEFINITIONS

Culture: The interactive aggregate of common characteristics that influence a human group's response to its environment, determines the identity of a human group in the same way as personality determines the identity of an individual.

Indigenous: Traditions; social, cultural, economic and political characteristics that are distinct to a particular group of people or society. Indigenous beliefs, systems and knowledge of practices for the sustainable management of resources.

Values: Important and lasting beliefs or ideals shared by the members of a culture about what is good or bad and desirable or undesirable. They are what the civilized and rational minds cherish, esteem, prize preciously and have inclined attitude for. They are principles or standards of a person or society for the good or benefit of the person or the society concerned.

Integrating Knowledge Management and Business Processes

John Steven Edwards

Aston University, UK

INTRODUCTION

Knowledge has been a subject of interest and enquiry for thousands of years, since at least the time of the ancient Greeks, and no doubt even before that. “What is knowledge” continues to be an important topic of discussion in philosophy.

More recently, interest in *managing* knowledge has grown in step with the perception that increasingly we live in a knowledge-based economy. Drucker (1969) is usually credited with being the first to popularize the knowledge-based economy concept by linking the importance of knowledge with rapid technological change. Karl Wiig coined the term knowledge management (hereafter KM) for a NATO seminar in 1986, and its popularity took off following the publication of Nonaka and Takeuchi’s book *“The Knowledge Creating Company”* (Nonaka & Takeuchi, 1995). Knowledge creation is in fact just one of many activities involved in KM. Others include identifying, acquiring, sharing, retaining, refining, and using knowledge. Heisig (2009) compared 160 different KM frameworks: no fewer than 117 of them included a list of activities. Global interest in KM, both academic and practical, has continued to increase throughout the last two decades, but as these numbers indicate, consensus on the theory underpinning KM remains some way off.

In this article, first the different types of knowledge are outlined, then comes a discussion of various routes by which knowledge management has been implemented. A business process-based route, which enables people, processes and tech-

nology to fit together, is growing in popularity as the best way to deliver effective KM that is integrated into what the organization does. Some examples of the business processes route in use are then given. Finally there is a look towards the future.

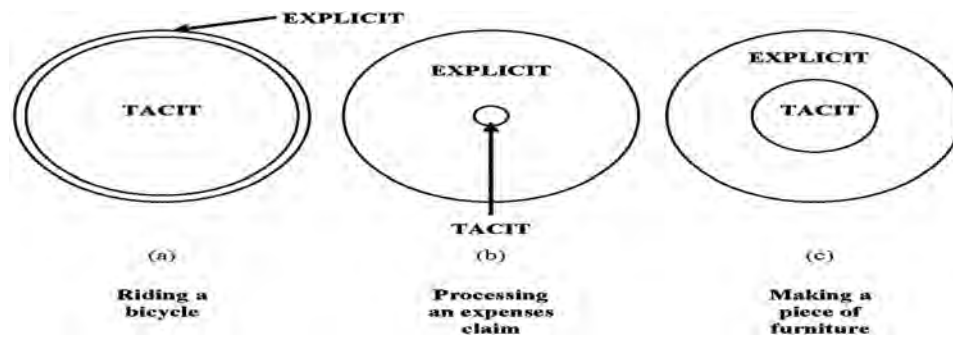
BACKGROUND

Types of Knowledge: Tacit and Explicit

Nonaka and Takeuchi’s book (1995) popularized the concepts of tacit and explicit knowledge, as well as KM more generally. They based their thinking on that of Michael Polanyi (1966), expressed most memorably in his phrase “we know more than we can tell”.

It is important to realize that tacit and explicit knowledge are not mutually exclusive concepts. Rather, any piece of knowledge has both tacit and explicit elements, as shown in Figure 1. The size of the inner circle represents the proportion of tacit knowledge: the tacit core at the heart of the knowledge that we “cannot tell”. Figure 1(a) shows a case where the knowledge is almost entirely tacit, as in riding a bicycle. Figure 1(b) shows mainly explicit knowledge, where the tacit core is very small, for example how to process a claim for travel expenses in an organization. Figure 1(c) shows an intermediate case, such as making a piece of furniture, where substantial amounts of both tacit and explicit knowledge are involved.

Figure 1. The relationship between tacit and explicit knowledge



KM Strategies

Hansen, Nohria and Tierney (1999) identified that there are two fundamental KM strategies, codification and personalization. Codification concentrates more on explicit knowledge (typically relying very heavily on information technology), personalization more on tacit knowledge (stressing interactions between people). They advocate that an emphasis on one fundamental KM strategy but also including an element of the other, in an 80-20 proportion, is likely to be the most successful.

ROUTES TO IMPLEMENTING KM

Managers have to translate the goals of any strategic initiative into practical, implementable reality. Even with a clear KM strategy, many organizations find it difficult to implement knowledge management systems successfully, especially to integrate KM into the organization properly. Identifying *who* should be involved in knowledge management, *what* knowledge is being managed, and *why* is it being managed can each be problematic. The routes organizations have attempted to follow can be put into five generic categories.

Knowledge World Route

The practical focus in Nonaka and Takeuchi (1995) was very much on knowledge creation. As a result, organizations attempting to follow

their principles for other aspects of KM, such as sharing or retaining knowledge, found it difficult to integrate abstract ideas about knowledge into what the organization actually does, or could do, or should do. Often only the “why” was considered, not the “who” or even the “what”.

Functional Route

This organizes the implementation around the existing organizational structure. The most commonly found structural elements intended to facilitate learning and knowledge sharing in organizations are departmental groupings based on functions. These have clear advantages for integration in terms of what might be called professional development and allegiance. Davenport and Prusak (1998) report examples of successful knowledge transfer between groups of surgeons, and groups of tunneling engineers, amongst others. However, this functional route also has the disadvantage that it encourages the compartmentalization of knowledge, so that integration may be only within that function, not the whole organization. Indeed, professional divisions can actively prevent sharing of knowledge. It has, for example, taken decades for hospital doctors in the UK National Health Service to allow other professionals such as pharmacists and physiotherapists to participate in decision-making about treatment of individual patients on an equal footing. More broadly, modern Western medical science has come to separate “diet” and “medication”, at least until the very

recent past, in a way that Chinese medicine, for example, never has done. The “functional silos” mentality that tends to result from running an organization in this manner was recognized by authors such as Hammer (1990) when KM was in its infancy. However, they are still commonly encountered, as pointed out by Edwards (2011).

This route is only effective when KM can be bounded within a single function. Where the need for KM to be integrated within the organization is broader, for example transferring knowledge between functions, this route will be counter-productive.

IT-Driven Route

This route assumes that the KM strategy is entirely one of codification, usually into some form of repository. In all but the smallest of organizations, such a codification task evidently requires considerable IT support, and the thrust of this route is that once the “correct” form of IT support for managing knowledge has been chosen, it is simply a matter of a great deal of hard work.

Since it ignores the personalization element of the KM strategy, this route only works in a very limited range of situations where the “what” questions are paramount, and as in Figure 1(b) little tacit knowledge is involved, for example where the main KM task is managing the knowledge held by a company in the form of patents. In other circumstances it may not achieve any improvement in KM at all. One example of this from the author’s experience is of a heavy manufacturing firm. Knowledge management in this organization was seen solely as an information systems issue; the KM group was part of the information systems department. The “solution” was seen in terms of the implementation of a knowledge sharing system based on Lotus Notes™. However, there was no real consideration as to who would share what knowledge with whom or for what specific purpose (“why”). Consequently, the eventual use of the installed IT was poor; the only really successful use was by the knowledge management

project team itself, where the “who” and “why” questions had been properly addressed, not just the “what” questions.

People-Centric Route

By contrast, this route assumes that the KM strategy is entirely one of personalization. By definition, this will answer all the “who” questions that might be involved in KM. Thus in organizations where there is general consensus on “what” knowledge is important and “why” it needs to be managed, such a route may prove effective.

However, this is only likely to be the case in small firms. Larger organizations have become increasingly diverse in their activities, and in the range of specialized knowledge that they need to access. This means that consensus even on what knowledge the organization has, never mind what is important, may be difficult to achieve. On the one hand, it may not be easy for a particular specialist to fully appreciate “what the organization does”. Equally, even the most conscientious senior manager will find it literally impossible to understand all the expertise and knowledge possessed by the specialists in his or her organization. To repeat the much-used quotation from Hewlett Packard CEO Lew Platt (Davenport & Prusak, 1998, p. xii) “If HP knew what HP knows, we would be three times as profitable.”

Business Processes Route

The basic weakness of both the IT-driven and people-centric routes is that they do not explicitly connect with “what the organization does”, thus leaving a potential implementation gap between KM and the business itself. Various management thinkers have presented models of “what the organization does”, for example:

- Porter’s value chain (Porter, 1985),
- Earl’s view of core processes (Earl, 1994), the ones that are done directly for external customers,

- Beer's "System Ones" (Beer, 1985), the systems that make the organization what it is,
- Core competences/competencies as espoused by Hamel and Prahalad (1994).

One of the eight principles of the ISO9000 family of standards for Quality Management Systems is the use of a process approach (ISO, 2012). The ISO9000 term product (or service) realization process is equivalent to Earl's core process or Porter's primary activity.

Although there are some significant differences between these models, their common theme is that the effectiveness – indeed the competitive advantage – of organizations depends not on how they are structured, or on what IT or people resources they have, but on what they do. In the terminology of this article, this means their underlying business processes. Note that business processes exist equally in not-for-profit organizations and the author has always found them happy to use the term.

Business processes possess five characteristics that justify their use as a foundation for implementing knowledge management in organizations.

1. Business processes have identifiable customers, whether internal or external. Knowledge is of little relevance to the organization unless put to use for a customer of some kind.
2. Business processes cut across organizational boundaries. Knowledge flows do not need to, and should not, obey the artificial boundaries within an organization.
3. Business processes consist of a structured set of activities. Choosing the appropriate way to structure activities is an important part of the relevant knowledge.
4. Business processes need to be measured. Without some form of measurement as a comparison, knowledge cannot be validated.
5. While the parts of a business process are important, the overriding requirement is that the overall process works. Valid knowledge in an organizational context must take a holistic view.

An additional argument (Braganza, 2001), is that viewing knowledge management in terms of an organization's processes gives a much-needed *demand-side* view of knowledge. This is complementary to the supply-side view of knowledge that stems, for example, from considerations of 'data leading to information leading to knowledge'. Beer and Earl particularly concentrate on this demand-side perspective. Beer indeed goes even further, to include the informal processes and activities of the organization as well as the more formalized ones. A demand-driven view clearly offers better potential for integration of KM with the rest of "what the organization does".

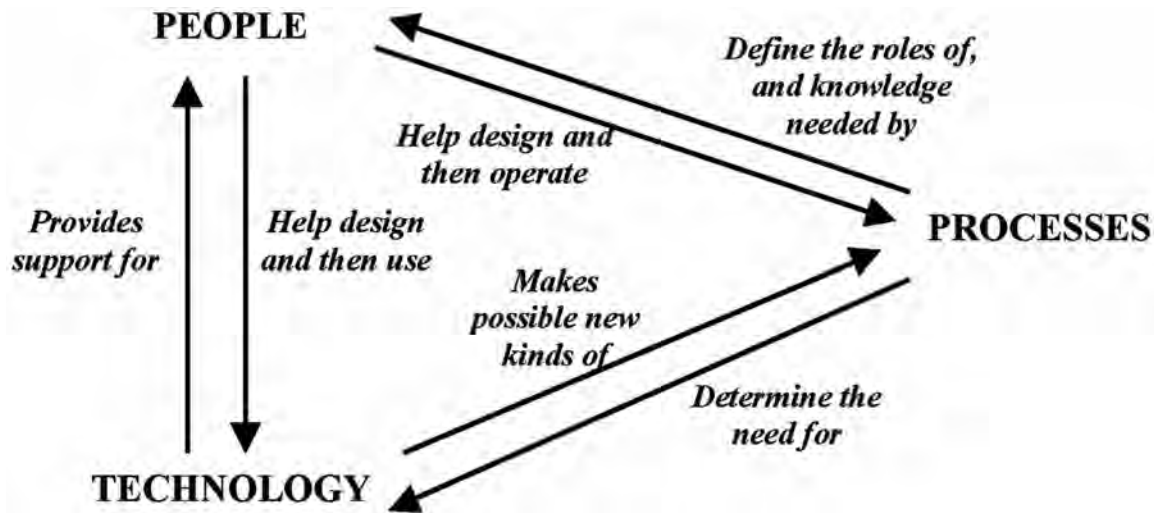
Thus we see that the knowledge that an organization requires must be related not just to what that organization does, but also to how it does it. Therefore people in organizations should benefit from thinking about this knowledge, and how to manage it, by reference to that organization's business processes.

PEOPLE, PROCESSES, AND TECHNOLOGY

As we have seen, effective KM requires the consideration of both tacit and explicit knowledge. The need is to coordinate people, processes and technology successfully using some kind of KM system. It is important to realize that there is more to a KM system than just technology, and that any deliberate, conscious attempt to manage knowledge in an organization amounts to a KM system. The interaction of the three elements people, processes and technology is shown in Figure 2.

Not only does a knowledge management system consist of more than technology, it is important to realize that the technology used to support KM does not have to be "KM software". Several studies continue to find that generic software such as e-mail, blogs or an Intranet is at least as important as specific software (Edwards, Shaw, & Collier, 2005; Zhou & Fink, 2003).

Figure 2. People, processes, and technology in a KM system



THE DEVELOPMENT OF THE BUSINESS PROCESSES ROUTE TO KM

The business processes route to KM is becoming more widely adopted. In the earlier days of KM it was seen as a “niche” approach, leading to the development of what was termed Business Process Oriented Knowledge Management (BPO-KM) (Heisig, 2001). However, BPO-KM has now become “mainstream”, as shown by examples such as Apostolou, Abecker, and Mentzas (2007), who take a process view of the activities of a management consulting firm, to develop KM support for the delivery of its services; and Mas-Machuca and Costa (2012) who study critical success factors for KM, also in the consulting sector.

Examples of the business processes route may now also be found in many other sectors and countries. The Singapore Ministry of Manpower (Fung, 2006) has used business strategies and processes to drive its blended KM approach rather than “plunging straight into analyzing the KM elements” (Fung, 2006, p. 31). Fung concluded that this dramatically increased the business relevance of the KM initiatives. Several organizations in the construction industry have used an approach based on incorporating KM

within a process improvement model (Jeong, Kagioglou, Haigh, Amaratunga, & Siriwardena, 2006). Geisel (2005) gives examples of how the insurance industry in the USA and in Europe has been building knowledge into its business processes. Dayan and Evans (2006) consider KM as a way to support the process-focussed Capability Maturity Model, and explain how this approach has been used in KM at Israel Aircraft Industries. Barcelo-Valenzuela, Sanchez-Schmitz, Perez-Soltero, Rubio, and Palma (2008) look at KM for one of the administrative processes of a Spanish university.

Corallo, Lazoi, and Secundo (2012) use a process view of New Product Development to consider knowledge integration in collaborative projects in the Italian aerospace industry, presenting comprehensive recommendations for managers working in similar supply chain contexts, be they prime contractors or lower-tier suppliers. Figueiredo (2013) examines learning and innovation in Motorola’s subsidiary in Brazil. A process approach leads to recommendations about the acquisition, assimilation and internalization of knowledge from elsewhere, whether that is within Motorola or outside it. Hustad (2010) describes corporate memory in a multinational company in the marine insurance industry, and especially how

knowledge is shared across boundaries such as geographical locations, professional specialisms and project teams.

It is becoming more and more common for researchers to take the benefits of a process approach to KM as an assumption. Han and Park (2009, p.7441) observe “[I]f knowledge is separated from the business process context, it does not lead to the ability to take the right action for target performance”. Mas-Machuca and Costa (2012, p.1310) included as an assumption the need “to integrate KM in the business process”. Soares, Santoro, and Baiao (2013, p.1451) state that “An organization that aligns KM with its business process is able to identify gaps, correct mistakes and update more quickly”. Bitkowska (2015, p.54) comments “Knowledge, which is a source of competitive advantage, cannot be considered in isolation from the actions implemented by the company.”

Cao, Thompson, and Triche (2013) recently made a connection not previously observed, by regarding the fit between business processes (task) and KM systems (technology) as an example of the task-technology fit theory, as originally put forward by Goodhue and Thompson (1995), that has found wide application in the information systems field. They find support for better performance from a business processes route in three case studies: a business communication provider, a networking infrastructure solutions provider, and a GPS technology provider.

Although this and many other examples involve substantial use of information technology, that does not have to be the case. The author’s group worked with a component manufacturer in the aerospace industry, whose KM initiative also has an explicit process focus. Typically their manufacturing processes use a machine operated by one person. The operators’ choice of the best way to retain and share knowledge does not use IT at all (except for word processing software and a printer). The agreed best operating procedure,

with illustrations, is put on a laminated sheet of paper mounted near the machine, which includes the names of the people who had contributed to designing the procedure. A suitable pen is provided to annotate the laminated sheet. At regular intervals office staff come round to produce a revised version of any of these “Standard Operating Sheets” that have been annotated.

Most significantly of all in terms of demonstrating the movement of this approach into the mainstream, a survey by the American Productivity and Quality Center (APQC) in 2011 found that more than half of the organizations surveyed were integrating KM into their critical processes and workflows at the enterprise level (Trees, 2011). It is much easier to do this if a business processes perspective is adopted, as shown by two recent studies in Taiwan. Hung, Tsai, Lee, and Chau (2015) demonstrate the mediating effect of business process outcomes between KM enablers (KM infrastructure and KM capability) and KM effectiveness, using data from hospitals and financial services. The assumption that KM needs to be aligned with the business processes is implicit in the way they measure the outcomes (efficiency, effectiveness, flexibility). Wu and Chen (2014) use a similar model on data from the manufacturing sector, with business process capability as the mediating factor between knowledge assets/knowledge process capability and organizational performance.

FUTURE RESEARCH DIRECTIONS

During the 21st century, business processes have become part of the mainstream of management thinking. Strategic management textbooks, e.g. Johnson, Whittington, Scholes, Angwin, and Regnér (2013), now routinely cover the business process view of organizations. As explained in the previous section, over the past 5-10 years it has begun to be accepted that the most effective

way to develop knowledge management in an organization is to integrate it with the organization's business processes.

It seems clear, therefore, that the integration of knowledge management and business processes will become even more common in future, and that this will encourage the development of ever more appropriate information technology for supporting it. Equally, new information technologies will enable new types of process. Social media and smart phones, for example, offer possibilities for the business and for supporting KM which are still only in the early stages of exploitation. They bring challenges, too: for example, unstructured and unrestricted "tagging" of online items may be as much confusing as helpful (Kim, Breslin, Kim, & Choi, 2010). Intelligent agents, wearable devices, and the potential for end-user access to the results of analytics based on big data (especially using new developments in visualization) all present both opportunities and challenges for KM. Perhaps the biggest challenge for integrating KM and business processes remains the need to be able to measure key activities: measurement is at the core of any process approach, but measuring KM is far from easy.

CONCLUSION

This article has considered the implementation of knowledge management in organizations by looking at five different generic routes towards achieving it: knowledge world, functional, IT-driven, people-centric and business processes. While each of these routes has some merit, acceptance is now growing that the business processes route, delivering integration between knowledge management and "what the organization does", is the most effective approach. It is thus likely to be increasingly chosen in the future.

REFERENCES

- Apostolou, D., Abecker, A., & Mentzas, G. (2007). Harmonising codification and socialisation in knowledge management. *Knowledge Management Research & Practice*, 5(4), 271–285. doi:10.1057/palgrave.kmrp.8500156
- Barcelo-Valenzuela, M., Sanchez-Schmitz, G., Perez-Soltero, A., Rubio, F. M., & Palma, J. (2008). Defining the problem: Key element for the success of knowledge management. *Knowledge Management Research & Practice*, 6(4), 322–333. doi:10.1057/kmrp.2008.22
- Beer, S. (1985). *Diagnosing the system for organisations*. Chichester, UK: Wiley.
- Bitkowska, A. (2015). The orientation of business process management toward the creation of knowledge in enterprises. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 25(1), 43–57. doi:10.1002/hfm.20533
- Braganza, A. (2001). Knowledge (mis)management...and how to avoid it. *Information & Management*.
- Cao, Q., Thompson, M. A., & Triche, J. (2013). Investigating the role of business processes and knowledge management systems on performance: A multi-case study approach. *International Journal of Production Research*, 51(18), 5565–5575. doi:10.1080/00207543.2013.789145
- Corallo, A., Lazoi, M., & Secundo, G. (2012). Inter-organizational knowledge integration in collaborative NPD projects: Evidence from the aerospace industry. *Knowledge Management Research & Practice*, 10(4), 354–367. doi:10.1057/kmrp.2012.25
- Davenport, T. H. (1993). *Process innovation: Reengineering work through information technology*. Boston, MA: Harvard Business School Press.
- Davenport, T. H., & Prusak, L. (1998). *Working knowledge: How organizations manage what they know*. Boston: Harvard Business School Press.

- Dayan, R., & Evans, S. (2006). KM your way to CMMI. *Journal of Knowledge Management*, 10(1), 69–80. doi:10.1108/13673270610650111
- Drucker, P. F. (1969). *The age of discontinuity*. London: Heinemann.
- Earl, M. J. (1994). The new and the old of business process redesign. *The Journal of Strategic Information Systems*, 3(1), 5–22. doi:10.1016/0963-8687(94)90003-5
- Edwards, J. S. (2011). A process view of knowledge management: It ain't what you do, it's the way that you do it. *Electronic Journal of Knowledge Management*, 9(4), 297–306.
- Edwards, J. S., Shaw, D., & Collier, P. M. (2005). Knowledge management systems: Finding a way with technology. *Journal of Knowledge Management*, 9(1), 113–125. doi:10.1108/13673270510583009
- Figueiredo, P. N. (2013). Embedding with multiple knowledge sources to improve innovation performance: The learning experience of Motorola in Brazil. *Knowledge Management Research & Practice*, 11(4), 361–373. doi:10.1057/kmrp.2012.19
- Fung, M. (2006). Breaking silos at Singapore Ministry of Manpower. *Knowledge Management Review*, 9(2), 30–33.
- Geisel, R. W. (2005). The marriage of knowledge to business processes. *Business Insurance*, (3): 18–20.
- Goodhue, D. L., & Thompson, R. L. (1995). Task-technology fit and individual performance. *Management Information Systems Quarterly*, 19(2), 213–236. doi:10.2307/249689
- Hamel, G., & Prahalad, C. K. (1994). *Competing for the future*. Boston, MA: Harvard Business School Press.
- Hammer, M. (1990). Re-engineering work: Don't automate, obliterate. *Harvard Business Review*, 68(4), 104–112.
- Hammer, M., & Champy, J. (1993). *Reengineering the corporation.: A manifesto for business revolution*. London: Nicholas Brealey.
- Han, K. H., & Park, J. W. (2009). Process-centered knowledge model and enterprise ontology for the development of knowledge management system. *Expert Systems with Applications*, 36(4), 7441–7447. doi:10.1016/j.eswa.2008.09.031
- Hansen, M. T., Nohria, N., & Tierney, T. (1999). What's your strategy for managing knowledge? *Harvard Business Review*, 77(2), 106–116. PMID:10387767
- Heisig, P. (2001). Business process oriented knowledge management. In K. Mertins, P. Heisig, & J. Vorbeck (Eds.), *Knowledge management best practices in Europe* (pp. 15–43). Berlin: Springer.
- Heisig, P. (2009). Harmonisation of knowledge management. *Journal of Knowledge Management*, 13(4), 4–31. doi:10.1108/13673270910971798
- Hung, S.-Y., Tsai, J. C.-A., Lee, W.-T., & Chau, P. Y. K. (2015). Knowledge management implementation, business process, and market relationship outcomes: An empirical study. *Information Technology & People*, 28(3), 500–528. doi:10.1108/ITP-12-2013-0209
- Hustad, E. (2010). Exploring knowledge management initiatives in a multinational company: Creating a corporate memory through networking. In E. Tomé (Ed.), *Proceedings of the 11th European Conference on Knowledge Management* (pp. 476–485). Reading, UK: Academic Publishing.
- ISO. (2012). *Quality management principles*. Retrieved 22 October, 2015, from http://www.iso.org/iso/qmp_2012.pdf
- Jeong, K. S., Kagioglou, M., Haigh, R., Amaratunga, D., & Siriwardena, M. L. (2006). Embedding good practice sharing within process improvement. *Engineering, Construction, and Architectural Management*, 13(1), 62–81. doi:10.1108/09699980610646502

Johnson, G., Whittington, R., Scholes, K., Angwin, D., & Regnér, P. (2013). *Exploring strategy: Text & cases* (10th ed.). Cambridge, UK: Pearson.

Kim, H.-L., Breslin, J., Kim, H.-G., & Choi, J.-H. (2010). Social semantic cloud of tags: Semantic model for folksonomies. *Knowledge Management Research & Practice*, 8(3), 193–202. doi:10.1057/kmrp.2010.10

Mas-Machuca, M., & Costa, C. M. (2012). Exploring critical success factors of knowledge management projects in the consulting sector. *Total Quality Management & Business Excellence*, 23(11), 1297–1313. doi:10.1080/14783363.2011.637778

Nonaka, I., & Takeuchi, H. (1995). *The knowledge-creating company, how japanese companies create the dynamics of innovation*. New York: Oxford University Press.

Polanyi, M. (1966). *The tacit dimension*. Garden City, NY: Doubleday.

Porter, M. E. (1985). *Competitive advantage: Creating and sustaining superior performance*. New York: Collier Macmillan.

Soares, D. C., Santoro, F. M., & Baiao, F. A. (2013). Discovering collaborative knowledge-intensive processes through e-mail mining. *Journal of Network and Computer Applications*, 36(6), 1451–1465. doi:10.1016/j.jnca.2013.02.007

Trees, L. (2011, December 12). *Knowledge sharing in the flow: Survey results*. Retrieved October 27, 2015, from <https://www.apqc.org/knowledge-base/documents/knowledge-sharing-flow-survey-results>

Wu, I.-L., & Chen, J.-L. (2014). Knowledge management driven firm performance: The roles of business process capabilities and organizational learning. *Journal of Knowledge Management*, 18(6), 1141–1164. doi:10.1108/JKM-05-2014-0192

Zhou, A. Z., & Fink, D. (2003). Knowledge management and intellectual capital: An empirical examination of current practice in Australia. *Knowledge Management Research & Practice*, 1(2), 86–94. doi:10.1057/palgrave.kmrp.8500009

ADDITIONAL READING

Becerra-Fernandez, I., & Sabherwal, R. (2010). *Knowledge management: Systems and processes*. Armonk, NY: M.E.Sharpe.

Nonaka, I., Toyama, R., & Hirata, T. (2008). *Managing flow: A process theory of the knowledge-based firm*. Houndmills: Palgrave Macmillan; doi:10.1057/9780230583702

KEY TERMS AND DEFINITIONS

Business Process: A structured, measured set of activities designed to produce a specified output for a particular customer or market (Davenport, 1993, p. 5).

Business Process Reengineering: The fundamental rethinking and radical redesign of business processes to achieve dramatic improvements in critical, contemporary measures of performance, such as cost, quality, service and speed (Hammer & Champy, 1993, p. 32).

Demand-Driven View of Knowledge: A view of knowledge stemming from the requirements of the organization; for example, what knowledge is needed to carry out a particular activity and how can it be applied?

Explicit Knowledge: Knowledge that has been (or can be) codified and shared with others.

Knowledge Management: Supporting and achieving the creation, sharing, retention, refinement, and use of knowledge (generally in an organizational context).

Knowledge Management Software: Software specifically intended for knowledge management, such as data mining and “people finder” software.

Knowledge Management System: A combination of people, processes and technology whose purpose is to perform knowledge management in an organization.

Supply-Driven View of Knowledge: A view of knowledge stemming from the knowledge itself rather than its uses. Often related to a continuum data-information-knowledge.

Tacit Knowledge: Knowledge that is difficult or impossible to express, except by demonstrating its application.

K

Intellectual Capital Measurement

Lukasz Bryl

Poznan University of Economics and Business, Poland

INTRODUCTION

In the contemporary knowledge-based economy the importance of assets related to human, his knowledge and abilities is on the rise. Nowadays Intellectual capital (IC) and intangibles as wealth production factors take precedence compared to physical assets. Especially in the ICT industry, in which Intellectual capital has its origin, IC is perceived as the driver of innovation, growth and competitive advantage of the companies. Moreover in the ICT industry it is the Intellectual capital mostly responsible for the value creation. ICT is being created by Intellectual capital, while at the same time ICT enable efficient Intellectual capital management.

The aim of the article is to present the current state of knowledge concerning Intellectual capital and its measurement methods. Although IC may be identified and calculated on the micro, macro and industry level, this paper deals with the IC measurement on the enterprise level solely. First part of the article is concentrated on the introduction to the IC notion and its main forms, second presents the most common IC measurement methods, while the third part is the analysis of controversies and usability of chosen method to provide a general vision of IC measurement concept.

BACKGROUND

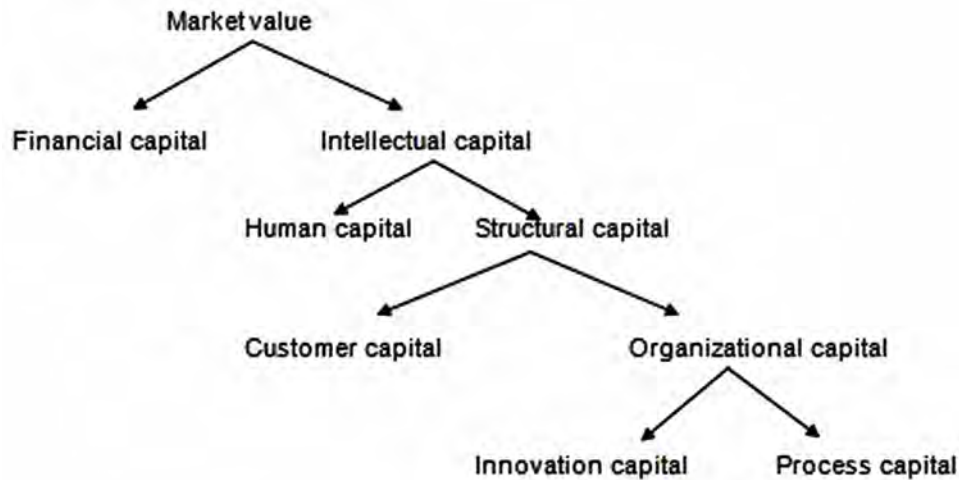
First use of the term *Intellectual capital* took place in 1958 when two financial analysts recognized as the most important element in the information technology companies their *Intellectual capital*. As a result of changes in the structure of econo-

mies in many countries, the term has become a subject of interest and study of a wider group of researchers.

Abeysekera (2006, pp. 61) recognizes Intellectual capital as a form of knowledge that is not posted in the traditional financial reporting. Brooking (1998, pp. 12) believes that Intellectual capital are combined intangible resources that allow organizations to function. Stowe (2001, pp. 86), in turn, argues that Intellectual capital is the ability to use the knowledge possessed by a person or a company to make better use of human and natural resources. In the discussion on the importance of Intellectual capital it is often emphasized that it may have a significant impact on achieving and maintaining a competitive advantage. Stewart (2003, pp. 32) considers the Intellectual capital as a sum of all knowledge the company staff has and what can provide a competitive advantage, manifested in market value which exceeds book value. Edvinsson and Malone (2001, pp. 17) argue that Intellectual capital may be recognized as the difference between the market value and the book value. Intellectual capital may include: patents, processes, people skills and experience, technologies, information about customers and suppliers (Stewart, 1997, pp. 71).

To sum up IC shall be perceived as intangible assets created by human and his knowledge that have not been entirely disclosed in the balance sheet but play a crucial role in the contemporary business environment in terms of enterprises competitive advantage. Moreover, there shall be stated the difference between the commonly used terms of: Intellectual capital and intangibles. IC is a broader notion than Intangibles - Intangibles shall be associated with these knowledge-based

Figure 1. Forms of Intellectual capital (Edvinsson & Malone, 2001, p. 45)



assets that can be reported and valued in the financial statements of the enterprises. In this sense Intangibles are part of IC.

Defining Intellectual capital is a complex activity, as it has several dimensions and may be presented in many ways. Figure 1 provides one of the most common and widely accepted classification of IC.

Intellectual capital has two main forms, which are human capital and structural capital. Human capital includes employee's collective knowledge, competence, abilities and the power of brain. Structural capital, however, is divided into customer capital and organizational capital. Customer capital encompasses relationships with customers, suppliers, industry associations and market channels. Organizational capital, that may be further divided into innovation and process capital, are enterprise policy and procedures, software applications, research and development programs, patents and training courses (Seetharaman, Sooria, & Saravanan, 2002, pp. 130).

Researchers argue that knowledge and consciousness about the IC existence is solely not the ultimate goal of analyzing this complex phenomena. Far more useful is the identification and measurement of IC and its components, as it is crucial for the proper valuation of the enterprises

(especially stock exchange companies). In this sense it helps investors to take strategic decisions concerning portfolio structure. Moreover, proper measurement of IC can provide a better view on the entity assets what plays a significant role in the mergers and acquisition transactions.

The most well-known and at the same time basic classification of measurement methods of Intellectual capital is the division proposed by Sveiby (2001), consisting of four groups:

1. Direct Intellectual capital Methods that focus on the study of certain (chosen) intangible assets.
2. Market Capitalization Methods which are basing on the difference between the market value and the book value, representing the value of Intellectual capital.
3. Return on Assets Methods, which examine the profitability of individual assets (including intangibles) involved in the company during the analyzed period.
4. Scorecard Methods, that, similarly to the direct measurement methods focus on the determination of individual components of Intellectual capital with such a difference that they rarely allow for estimating their monetary value.

Methods described above have been distinguished with the use of two criteria: possibility of IC valuation and measurement level. In that sense Market Capitalization, Return on Assets Methods and Direct Intellectual capital Measurement Methods provide the value of IC, however the first two apply to the organizational level, while the last one allows to identify certain components of IC within an organization. In addition, Scorecard Methods do not provide exact IC value, however they enable the identification of certain IC elements in a given entity.

Although there are more than 40 widely recognized IC measurement methods, due to the limited space, Table 1 provides information only on selected, most renowned methods.

DISCUSSION ON THE INTELLECTUAL CAPITAL MEASUREMENT METHODS

Although, there is a plethora of Intellectual capital measuring methods on the theoretical background, proper IC calculations still cause serious problems and controversies. According to the above mentioned groups, Direct Methods provide data only on chosen IC components and as they do not need to measure IC elements in financial terms, their outcomes are not very useful for financial and comparing purposes.

More holistic approach, and thus better view on IC as an entire resource (capital) attributed to an enterprise provide Scorecard Methods, that

Table 1. Summary of the main Intellectual capital measurement methods

Method	Author, Year	Description
Direct Intellectual Capital Methods		
The Value Explorer	(Andriessen & Tiessen, 2000)	Method distinguishes IC as five types of intangibles: (1) Assets and endowments, (2) Skills and tacit knowledge, (3) Collective values and norms, (4) Technology and explicit knowledge, (5) Primary and management processes.
Intangible Assets Valuation	(Sullivan, 2000)	Methodology concentrating only on the valuation of intellectual property.
Accounting for the Future (AFTF)	(Nash, 1998)	Model is based on future, projected, discounted cash-flows. IC is perceived as Value added that is being estimated as a difference between AFTF at the end and at the beginning of the period.
Inclusive Valuation Methodology (IVM)	(McPherson, 1998)	Final IC value is the combination of Monetary Value Added and Intangible Value Added. Model uses hierarchies of weighted indicators.
Total Value Creation (TVC)	(Anderson & McLean, 2000)	Discounted future cash-flow are used to re-examine the effect of events on the planned activities.
Technology Broker	(Brooking, 1996)	IC is divided into four components: (1) Human-centred Assets, (2) Intellectual Property Assets, (3) Market Assets and (4) Infrastructure Assets. IC value is calculated on the basis of in-depth analysis of answers on 20 questions covering mentioned IC components.
Market Capitalization Methods		
Tobin's q	(Tobin & Brainard, 1968)	The ratio between the stock market value and the replacement cost of its assets. Commonly, formula is presented as the ratio between market and book value. Figures above 1 indicate there is a hidden value (IC) which is not reported in the traditional company accounting.
Market Value Added (MVA)	-	The easiest and most common IC valuation. IC is the difference between current stock market value and its book value.

continued on following page

Table 1. Continued

Method	Author, Year	Description
Investor Assigned Market Value (IAMV)	(Standfield, 1998)	IAMV uses the market value of the company as the true value and divides it into: tangible capital, realized IC, IC erosion and Sustainable Competitive Advantage (SCA).
Return on Assets Method		
Calculated Intangible Value (CIV)	(Stewart, 1997)	CIV is based on the assumption that an enterprise generates IC when its ROA is higher than mean industry ROA; thus showing the capability to generate higher profits with the use of the same tangible assets what is explained by the IC presence.
Human Resources Costing	(Johansson, 1997)	Model concentrates on the impact of human resources on the company performance. IC is calculated as the difference between contribution made by human assets and salary expenditures.
Knowledge Capital Earnings (KCE)	(Lev, 1999)	Financial outcome of the company (net profit/loss) is the result of involving physical, financial and intangible capital. By taking the mean rate of return each of the capital it is possible to estimate the outcome attributed to the intangible capital.
Economic Value Added (EVA)	(Stewart, 1997)	EVA shows if the company generates profits for the shareholders by comparing its pre-tax operating profit with its cost of capital. Useful method for fostering strategic decisions in an enterprise.
Value Added Intellectual Coefficient (VAIC)	(Pulic, 1997)	Divides capital involved in the company into 3 components: physical capital, structural capital and human capital. The efficiency of creating value by each of the components is calculated.
Scorecard Methods		
Balanced Scorecard	(Kaplan & Norton, 1996)	It is a set of indicators referring to four major firm's perspectives: (1) financial, (2) customer, (3) internal process and (4) learning, which are based on strategic objectives of an enterprise.
Intangible Assets Monitor	(Sveiby, 1997)	It is based on estimating the capability of creating value from intangible assets in terms of four indicators: (1) growth, (2) renewal, (3) utilization/efficiency and (4) risk reduction/stability according to the company strategy.
IC-Index	(Roos, Roos, Dragonetti & Edvinsson, 1997)	Single IC index that puts together all indicators referring to the intellectual properties and its components. IC Index values are then compared to the market value changes.
Skandia Navigator	(Edvinsson & Malone, 1997)	One of the first methods; has been developed for the purpose of estimating IC in the Swedish enterprise, Skandia. IC is calculated with the help of 164 indices. 73 are traditional, whereas 91 intellectually based. Together they cover five components referring to the past (financial components), present (customer, process, human) and future (renewal and development). With the help of Skandia Navigator it is possible to distinguish between different forms of IC.
Holistic Approach Value	(Rambøll, 1999)	Model weights nine key areas of corporate performance to give true and fair value of the company. Eight key areas: (1) Values and management, (2) Strategic processes, (3) Human resources, (4) Structural resource, (5) Consultancy, (6) Customer results, (7) Employee results, (8) Results are intangible and measured by the indices that provide information about the true above the market value.

Source: Own work based on <http://www.sveiby.com/articles/IntangibleMethods.htm>

measure IC from the bottom up, can be applied at every level of an organization and provide useful data on firm's IC. However, as Scoreboard methods rely on the contextual and subjective indicators, their meanings may be ambiguous and impossible for inter-industry comparisons. Furthermore, their monetary valuation is not possible, hence they give only a broad image of IC, its components and potential influence on company performance, without indicating exact value. Analyzing one of the Direct Intellectual capital Methods, Intangible Assets Monitor there shall be stated that indices taken into account are subjective, and thus may vary and therefore provide different results. The example of Intangible Assets Monitor is presented in the Table 2.

Another (different) group of IC measurement methods form Market Capitalization Methods and Return on Assets Methods. These methods provide data on the exact IC value. Their main advantages are: relatively simple calculation, no major problems with the acquisition of data and (sometimes) the ability to assess which of the assets (tangible or intangible) bring the most benefits to the company. Moreover, by gathering IC values in a long-term, there is an opportunity

to observe changes in time, what may be useful for the diagnosis of main barriers of the company strategic development.

However, they also have disadvantages. The most common methods are: Tobin's q and Market Value Added (MVA) which is based on the previously mentioned indicator. First one states, that by comparing Market Value with the Book Value (Total equity) in the form of a ratio, there can be determined the IC existence (or lack of existence). When Tobins' q is higher than 1, the company possesses IC. This assumption is highly controversial, since it is based on the efficient market hypothesis. However it should be noted that whether markets are efficient enough to capitalize all disclosed information about IC and how many times a market needs to capitalize it stays unclear (Zavertiaeva, 2016). Secondly it should be stressed that Market Capitalization Methods are only relevant to the public companies that shares are traded on the stock exchange. This causes another problem, which is the high volatility of market value, and thus IC value, since the stock price is the result of daily supply and demand. Another problem is that values received are strongly dependent upon the macroeconomic environment. In the boom

Table 2. The example of intangible assets monitor

Competencies	Internal Structure	External Structure
Growth/Renewal Indices		
• Average professional experience of the employee (years) • Share of the customers increasing the knowledge resources of the company in total sales • total competencies – sum of all professional employee years	• Investments in IT as a percentage of sales • Share of customers strengthening the internal structure in total sales • R&D expenditures as sales percentage	• Company revenue growth • Share of customers strengthening the firm image in total sales
Efficiency Indices		
• Value added per one expert • Value added per one employee	• Changes in the proportion of auxiliary staff to all staff	• Sales per one client
Stability Indices		
• Experts rotation • Average number of working years of the professional staff • Average age of experts	• Employee rotation • Average number of working years of auxiliary staff • Share of employee with the work experience shorter than two years	• Share of regular customers • Share of sales per five largest enterprise customers

Source: Own work based on Sveiby (1997)

period they tend to be overvalued, whereas in the recession – undervalued. Hence to eliminate the influence of stock exchange volatility Tobin's q is being modified to the following formula:where:

MV - market value of shares,
 B_LL - book value of long-term liabilities,
 B_TI - book value of inventories,
 B_CL - book value of current liabilities,
 B_CA - book value of current assets,
 B_TA - book value of total assets.

Numerous studies on Tobin's q ratio have been conducted so far. One of them is the study by Kong, Xiao, and Liu (2010, pp. 6-33). Authors have made research on 313 listed companies on the Shenzhen Stock Exchange and found out that the mean Tobin's q ratio was 1,43 what, according to the theory, constitutes IC presence.

The last recommended method of measuring Intellectual capital among methods based on market capitalization is the IAMV model that formula is following:

CV = value of physical + realized capital + IC erosion +SCA

where:

CV - value of the company
 IC - Intellectual capital
 SCA - Sustainable competitive advantage

Although the formula itself seems to be easy understandable, the most ambiguous term is the notion of *Sustainable competitive advantage* that is explained as the long-term competitive advantage of the firm that is not easily duplicable or surpassable by the competitors and thus allows to perform at a higher level than others in the same industry. This theoretical approach may be, however, calculated in many ways.

Among ROA methods there may be also doubts observed that mainly refer to the subjective taken criteria having substantial impact on the final result of IC values. This notion is especially adequate to the Calculated Intangible Value (CIV) and Knowledge Capital Earnings (KCE) methods. CIV formula, designed by Stewart (1997) for IC calculations may be presented as the ratio below:

Main controversies related to the CIV formula emerge in terms of calculating the discount rate and mean industry ROA. First one seems to be a major problem, since under the term *industry* there should be understood the most similar group of peer enterprises so that the comparison would be legitimated. This occurs to be a problem due to four reasons. First, peer group should derive from possible same industry and country, second, peer group should be quite numerous, third, only companies that provide reliable financial accounting can be taken into account, and fourth, there may be compared entities that are in the analyzed period on the same or similar level of development,

Figure 2. Modified formula for Tobin's Q

$$\text{Tobin's } q = \frac{MA + B_{LL} + B_{TI} - B_{CL} - B_{CA}}{B_{TA}}$$

Figure 3. CIV formula by Stewart (1997)

$$IC_{CIV} = \frac{(\text{Income before Tax} - ROA_{\text{Industry}}) \times (\text{Total assets} - \text{Intangibles}) \times (1 - \text{Tax rate}_{\text{effect}})}{\text{discount rate}}$$

thus peer group may not include large TNC's and SME at the same time.

Furthermore, according to the controversial discount rate assumptions, the most common way of the CIV calculations are based on the weighted average cost of capital (WACC) that consists of: cost of equity and cost of debt. First one is estimated, primarily, by using the Capital Asset Pricing Model (CAPM), that, in its basic form, includes three variables: risk-free rate, risk premium and beta. Not going into detail, on the theoretical and empirical level there is no one clear method of calculating these values, hence the final result of the cost of equity and WACC may be different.

In the research on CIV method Aho, Ståhle, and Ståhle (2011, pp. 27-35) have tested the possible connections of CIV to profitability, market value and different capital assets in firms listed on the Finnish Stock Exchange during 2004-2008. Correlation analysis showed that CIV value is positively associated with the firm's market value and different capital assets, but not with profitability and financial assets.

Second, controversial method among ROA methods is the Knowledge Capital Earnings (KCE), developed by Lev (1999). It assumes that the financial result of the company is due to the use of physical, financial and intangible assets. Taking into account the average rate of return on physical and financial assets, it is possible to determine the earnings attributed to knowledge, and after discounting it - the value of this capital. The formula for KCE is following:

$$\text{Financial result} = \alpha \times \text{physical assets} + \beta \times \text{financial assets} + \delta \times \text{intangible resources}$$

where:

α - return on physical assets

β - the profitability of financial assets

δ - the profitability of intangible assets

KCE relies on the average rates of return on investment in physical, financial and intangible assets. Lev proposes 7%, as a return rate for physical capital investment, what is attributed to the mean rate of return for physical capital in the entire economy. Rate of return for the financial assets is 4,5% which is the average rate of return of 10-year US bonds during 1980-1990. However, the rate of return for intangible assets has been set for 10,5%, which is the average rate of return on stock investments in biotechnology and software enterprises during 1980-1990. What part of the profit is created by the individual types of resources is highly controversial. Moreover, proposed by Levy values for the rates of return seem to be obsolete nowadays, thus nowadays researchers tend to calculate the rates of return on their own using Lev's methodology.

There is also one unique method of IC measurement among ROA methods that is the Value Added Intellectual Coefficient (VAIC) that assesses what is the return on the capital invested in the company. Its high value means that the company uses resources effectively, both tangible and intangible. This also tells how Intellectual capital is able to create value-added (Pulic, 2000, pp. 702-714). In this method capital that is used by the company is divided into physical (employed) capital, human capital and structural capital. The problem with this method is that VAIC does not really show the value of IC, in fact VAIC provides data only on IC effectiveness. The basic estimation in the VAIC method is the calculation of Value Added which is the sum of operating profit (EBIT), amortization/depreciation and human capital. Moreover, it assumes certain values from the traditional accounting as IC or its components. For example human capital (as part of IC) is the total value of human resources expenditures (basic remuneration, performance premium and trainings), whereas structural capital is the difference between value added and human capital. Basing on the above assumptions and provided methodology several studies have been conducted to measure IC effectiveness with the use of VAIC

method. One of them was the study conducted by Ståhle, Ståhle, and Aho (2011, pp. 531 - 551). Authors analyzed 125 listed Finnish companies. The mean VAIC value in 2006-2008 was 2,8%, however mean Value Added amounted to 485 million EUR. Moreover, authors found out that VAIC values were significantly positively correlated with ROA, however there was no relation found between VAIC and Market Value.

SOLUTIONS AND RECOMMENDATIONS

As stated in the beginning, although the notions of IC and intangibles are not the same, they are partly overlapping terms. Thus some part of IC (in form of intangibles) may be valued with the use of traditional accounting tools. Although unambiguous measurement of IC is still a problem, there has been some recognized attempts undertaken for IC components (Intangibles) identification. At the same time many entities gradually adopt international recommendations in terms of intangible assets quantification. According to IFRS 3 there are following intangible assets distinguished (International Financial Reporting Standard Foundation, 2008, pp. 113-118):

- **Marketing-Related**
 - (Trademarks, trade names, service marks, collective marks and certification marks, Trade dress, Newspaper mastheads, Internet domain names, Non-competition agreements);
- **Technology-Based**
 - (Patented technology, Computer software and mask works, Unpatented technology, Databases, including title plants, Trade secrets, such as secret formulas, processes and recipes);
- **Customer-Related**
 - (Customer lists, Order or production backlog, Customer contracts and related customer relationships, Non-contractual customer relationships);

- **Artistic-Related**
 - (Plays, operas and ballets, Books, magazines, newspapers and other literary works, Musical works such as compositions, song lyrics and advertising jingles, Pictures and photographs, Video and audiovisual material, including motion pictures or films, music videos and television programs);
- **Contract-Based**
 - (Licensing, royalty and standstill agreements, Advertising, construction, management, service or supply contracts, Lease agreements, Construction permits, Franchise agreements, Operating and broadcast rights, Servicing contracts, such as mortgage servicing contracts, Employment contracts, Use rights, such as drilling, water, air, timber cutting and route authorities).

These intangible assets might be identified in the company balance sheet, valued and prepared for the potential sale to the third parties. It should be noted that, according to the international standards of accounting, only these components of intangible assets can be reported in the financial statements (Nimtrakoon, 2015, pp. 4). The remaining assets, that individual distinction is impossible, may be subject to sales transactions only within the sale of the whole company. One of such assets is called goodwill, which is defined as a difference between the purchase price of a particular entity or its parts and the fair value of the acquired net assets. As the notion is a subject of discussion Johnson and Petrone (1998) identified following categories of goodwill:

- Excess of the fair values over the book values of the acquiree's reported net assets;
- Fair values of not identifiable intangible assets;
- Fair value of synergies from combining the acquirer's and acquiree's businesses;
- Overpayment by the acquirer.

Studies on goodwill clearly state that it can be demonstrated in terms of accounting standards in the company's balance sheets when the entity is sold for a higher value than the total value of its assets including liabilities. Then the surplus is treated as an additional value and appears on the company balance sheet

FUTURE RESEARCH DIRECTIONS

Current state of knowledge on the IC estimations and tools shows that finding an ultimate and best measurement methods of IC is not a widely accepted and finished process. Although there has been much done in terms of IC identification, there is still a lot of effort needed to establish widely recognized method(s) for calculating its value. Proper procedure for IC valuation shall include following basic features: it shall be comparable in time, among countries and across industry. Moreover, it should be free of any subjective criteria (or at least reduced to the minimum essential), so that IC value could not be questioned because of imperfect methodological assumptions.

CONCLUSION

On the basis of undertaken studies there may be several conclusion formed concerning the problem of Intellectual capital measurement. As a result there can be stated:

- Lack of one widely accepted and universal definition of Intellectual capital,
- Coexistence of the concept of Intellectual capital and value creation term implies the conclusion that Intellectual capital makes sense only if it generates the value growth in the organization,
- Interchangeability of the terms: Intellectual capital, value of intangible, intellectual assets, intangible assets, assets hidden, intellectual resources and intangible resources,
- Intellectual capital provides useful knowledge, with the help of which substantial financial results may be achieved,
- The most common classification of the Intellectual capital is the division on: human, customer and organizational capital. There are tools for valuing each of them,
- Great controversies with the Intellectual capital measurement, as a result of methods diversity and taken assumptions,
- Lack of one widely accepted measurement method of Intellectual capital – urgent need for new tools and procedures.

REFERENCES

- Abeysekera, I. (2006). The project of Intellectual capital disclosure: Researching the research. *Journal of Intellectual Capital*, 7(1), 61–77. doi:10.1108/14691930610639778
- Aho, S., Ståhle, S., & Ståhle, P. (2011). A critical assessment of Stewarts CIV method. *Measuring Business Excellence*, 15(4), 27–35. doi:10.1108/13683041111184080
- Edvinsson, L., & Malone, M. (1997). *Intellectual capital: Realizing Your Company's True Value by Finding Its Hidden Brainpower*. New York, NY: HarperCollins Publishers.
- International Financial Reporting Standard Foundation. (2008). *International Financial Reporting Standards 3*. London: Business Combinations.
- Johnson, L. T., & Petrone, K. R. (1998). Is goodwill an asset? *Accounting Horizons*, 12(3), 293–303.
- Kaplan, R. S., & Norton, D. P. (1996). *The Balanced Scorecard: Translating Strategy into Action*. Boston, MA: Harvard Business School Press.
- Kong, D., Xiao, T., & Liu, S. (2010). Asymmetric information, firm investment and stock prices. *China Finance Review International*, 1(1), 6–33. doi:10.1108/20441391111092246

Lev, B. (2001). *Intangibles. Management, Measurement, and Risk*. Washington, DC: Brookings Institution Press.

Nimtrakoon, S. (2015). The relationship between Intellectual capital, firms market value and financial performance. *Journal of Intellectual Capital*, 16(3), 587–618. doi:10.1108/JIC-09-2014-0104

Pulic, A. (2000). VAIC - An Accounting Tool for IC Management. *International Journal of Technology Management*, 20(5-8), 702–714. doi:10.1504/IJTM.2000.002891

Seetharaman, A., Sooria, A. H. B. Z., & Saravanan, A. S. (2002). Intellectual capital accounting and reporting in the knowledge economy. *Journal of Intellectual Capital*, 3(2), 128–148. doi:10.1108/14691930210424734

Ståhle, P., Ståhle, S., & Aho, S. (2011). Value added intellectual coefficient (VAIC): A critical analysis. *Journal of Intellectual Capital*, 12(4), 531–551. doi:10.1108/14691931111181715

Stewart, T. A. (1997). *Intellectual capital. The New Wealth of Organizations*. New York, NY: Doubleday.

Stewart, T. A. (2003). *The Wealth of Knowledge. Intellectual capital and the Twenty-First Century Organization*. New York, NY: Doubleday.

Stowe, C., & Kwiatkowski, S. (2001). *Knowledge cafe for Intellectual capital*. Warsaw: Leon Kozminski Publishing.

Sveiby, K. E. (1997). *The New Organizational Wealth: Managing and Measuring Knowledge Based Assets*. San Francisco, CA: Berrett-Koehler Publishers.

Sveiby, K. E. (2001). *Methods for Measuring Intangible Assets*. Retrieved from <http://www.sveiby.com/articles/IntangibleMethods.htm>

Zavertiaeva, M. (2016). Portfolio forming decisions: The role of Intellectual capital. *Journal of Intellectual Capital*, 17(3), 439–456. doi:10.1108/JIC-10-2015-0087

ADDITIONAL READING

Drucker, P. (1999). Knowledge – Worker Productivity. The Biggest Challenge. *California Management Review*, 41(2), 79–94. doi:10.2307/41165987

Guthrie, J., Petty, R., & Ricceri, F. (2006). The voluntary reporting of Intellectual capital: Comparing evidence from Hong Kong and Australia. *Journal of Intellectual Capital*, 7(2), 254–271. doi:10.1108/14691930610661890

Guthrie, J., Ricceri, F., & Dumay, J. (2012). Reflections and projections: A decade of Intellectual capital accounting research. *The British Accounting Review*, 44(2), 68–82. doi:10.1016/j.bar.2012.03.004

Harrison, S., & Sullivan, P. H. Sr. (2000). Profiting from Intellectual capital Learning from Leading Companies. *Journal of Intellectual Capital*, 1(1), 33–46. doi:10.1108/14691930010324124

Kaplan, R. S., & Norton, D. P. (2004). Measuring the Strategic Readiness of Intangible Assets. *Harvard Business Review*, (February): 2–10. PMID:14971269

Roos, G., Roos, J., Dragonetti, N., & Edvinsson, L. (1997). *Intellectual capital: Navigating in the New Business Landscape*. Oxford: Macmillan. doi:10.1007/978-1-349-14494-5

Sullivan, P. (2000). *Value-Driven Intellectual capital*. New York, NY: John Wiley & Sons.

Williams, S. M. (2001). Is Intellectual capital performance and disclosure practices related? *Journal of Intellectual Capital*, 2(3), 192–203. doi:10.1108/14691930110399932

Wyatt, A. (2005). Accounting Recognition of Intangible Assets: Theory and Evidence on Economic Determinants. *Accounting Review*, 80(3), 967–1003. doi:10.2308/accr.2005.80.3.967

Zhang, D. (2003). Intangible assets and stock trading strategies. *Managerial Finance*, 29(10), 38–56. doi:10.1108/03074350310768508

KEY TERMS AND DEFINITIONS

Book Value: Value of Total Assets less Liabilities.

Customer Capital: Relationships with customers, suppliers, industry associations and market channels.

Efficient Market Hypothesis: Assumption stating that existing share prices always incorporate and reflect all relevant information and thus are traded at a fair value.

Financial Statement: Corporate document presenting company performance during studied

period, typically consisting of: balance sheet, cash-flows and income statement.

Human Capital: Employee's collective knowledge, competencies, abilities and the power of brain.

Intellectual Capital: Hidden intangible assets based on the human knowledge that are not entirely reported in the financial statement of the enterprise and are playing significant role in the knowledge-based economy.

Knowledge-Based Economy: Economy which is directly based on the production, distribution and use of knowledge and information.

Market Value: Value of an enterprise calculated as a number of shares multiplied by unit share price.

Organizational Capital: Enterprise policy and procedures, software applications, research and development programs, patents and training courses.

ROA: Return on assets; the formula is calculated as the ratio of net profit/loss to total assets.

Knowledge Acquisition on Dante Alighieri's Works



Elvira Immacolata Locuratolo

ISTI-CNR, Italy

Valentina Bartalesi Lenzi

ISTI-CNR, Italy

INTRODUCTION

Towards a Digital Dante Encyclopedia is a three-year Italian National Project that aims to build a digital library endowed with services to support scholars in creating, evolving and consulting a digital encyclopedia of Dante Alighieri and of his works (Alighieri, 2011, 2014). Studies concerning the semantic representation of the notes to Dante Alighieri's works can be found in Bartalesi, Locuratolo, Meghini and Versienti (2013). A semantic network of Dante's works and their contextual knowledge is described in Andriani, Bartalesi, Locuratolo, Versienti, Meghini, Tavoni and Versienti (2013).

This article is a research paper related with one of the project objectives, i.e. the identification and the implementation of visualization diagrams to support the of knowledge acquisition on Dante's works. The collaboration between humanists and computer scientists was essential for the purposes of this activity. Moreover, knowledge from humanists was advantageously exploited to individuate the right diagrams; knowledge from computer scientists was advantageously exploited to solve methodological problems of research, implementation and graphical representation.

The approach followed for the purposes of this activity was based on three phases, respectively called identification, implementation and graphical representation. With regard to the identification

phase, two high level structures were introduced: the former, called Dante work's sources, is concerned with the cited primary sources; the latter, called Dante work's structure, is concerned with the structure of Dante's works. For each of them, significant types of histograms were identified. The structure Dante work's sources is applicable to both all the Dante's works/a selected Dante's work. In the former case, the histograms that can be derived from the structure are sufficient to fulfill the project purposes; in the latter case, the histograms derived from both the structures, Dante work's sources and Dante work's structure, need to be related.

The implementation phase is concerned with a table summarizing the structures and the knowledge from humanists. The significant types of citation histograms are represented in a model that does not reduce the completeness of the graphical information. The model is formed by a table containing four columns identified by the 4-tupla (Domain, Range, Graphical Type, Refinement).

With respect to the graphical representation, the histogram is a function depending from the parameters Domain and Range. The domain is represented on the horizontal axis of a cartesian diagram, whereas the citation distributions are represented on the vertical axis. The possibility to refine the acquired knowledge, as well as the possibility to increase it is provided. An examples of citation histograms is given.

BACKGROUND

The study concerning the semantic representation of the knowledge contained in commentaries to Dante Alighieri's works, especially focused on primary sources which are the works of other authors cited by Dante in his texts, is summarized. The approach is composed of several steps: the first step of them started from the analysis of an Excel style sheet, where some pieces of knowledge included in a commentary of *Convivio* (Alighieri, 2014), a philosophical essay composed by Dante Alighieri in the years between 1304 and 1307, were organized and reported by an Italian scholar. In the Excel sheet, every note is given in a row, and is composed of the following pieces of knowledge:

- Number of chapter of the annotated text, represented as a pair of the form book chapter (e.g., 1.01 indicates the first chapter of the first book of *Convivio*);
- Number of the paragraph;
- The text fragment to which the note applies (e.g., “Sì come dice lo filosofo nel principio della Prima Filosofia”);
- The text of the note;
- The last three columns give information about a primary source reported into the note, structured as:
 - Author (e.g. Aristotele);
 - Title (e.g. Metafisica);
 - Thematic area (e.g. Aristotelismo).

In the second step, in order to create an ontology in RDF (Manola & Miller, 2004) for the semantic representation of the previous pieces of knowledge, the terms belonging to vocabularies used in the Digital Libraries domain were identified. For this purpose, several existing ontologies, e.g. CIDOC-CRM (Doerr, 2003), FRBR (Tillet, 2005), FaBiO (Peroni & Shotton, 2012) SKOS (Miles, Matthews, Wilson & Brickley, 2005) were investigated, and the terms that we considered useful to represent

our knowledge were chosen. In order to describe all the knowledge embedded in the commentary, we added classes and properties (Bartalesi, Locuratolo, Meghini & Versienti, 2013).

In the third phase, we built a RDF graph that represents the semantics of the commentary structured according to the ontology. After this preliminary study on commentary to *Convivio*, we used the developed ontology to represent other works of Dante as well as the knowledge carried by commentaries to them. Since the structure of knowledge reported in the Excel file is valid also for other Dante's works (i.e. *Monarchia* (Alighieri, 2014), *De Vulgari Eloquentia* (Alighieri, 2011), and *Vita Nova* (Alighieri, 2011)), the ontology was populated with other knowledge. Three experts extracted the previous reported pieces of knowledge from commentaries to these Dante's works, using a semi-automatic tool we developed. Finally, the RDF graph was stored into a Virtuoso triple store (Erling & Mikhailov, 2009).

The technical advantages of our approach are essentially three: (1) researchers can add classes and relationships to our ontology, thereby refining it, (2) our ontology can be linked to other ontology to extend the represented domain, (3) any user can download and use our model freely, using the paradigm of Linked Data.

Web services were developed to visualize the knowledge stored in the graph in form of charts and tables and CSV format. Using SPARQL query language to extract knowledge from Virtuoso, the services make it possible to address several tasks carried out by the scholars building an encyclopedia of Dante's works, starting with the visualization of the distribution of primary sources both in time and in the works of Dante. The overall goal is to shed light on the cultural context in which Dante wrote his works and on the development of Dante's reference library over time. Our diachronic analysis, in fact, aims at representing the evolution of Dante's knowledge about primary sources.

MAIN FOCUS OF THE ARTICLE

The focus of this article is concerned with one of the main research activity supporting the *Towards a Digital Dante Encyclopedia* project: the approach followed for the identification and the implementation of visualization histograms useful to the digital humanities scholars, during the acquisition of knowledge on Dante's works. Firstly, high level structures are introduced. Secondly, an implementation model is given. Finally, a discussion is provided. The high level structures make it possible to identify the types of significant histograms of all the Dante's works/a Dante work. The implementation model offers the minimal number of histogram types need to preserve completeness. The discussion focuses on educational aspects of mathematics and statistics underlying the histogram representation; on the approach to acquire knowledge on Dante's works starting from the visualization diagrams, and on research advancements. The achievements of the *Towards a Digital Dante Encyclopedia* projects, in particular, the description of the web application development (Le fonti dantesche, n.d.); of the search forms for queries and of histogram examples are provided in (Tavoni, Andriani, Bartalesi, Locuratolo, Meghini & Versienti, 2014).

SOLUTIONS AND RECOMMENDATIONS

In this section, the solution to the identification and implementation of visualization diagrams able to support the knowledge acquisition on Dante's works is proposed. The approach is based on three phases, respectively called identification, implementation and graphical representation.

With regard to the identification phase, two high level structures were introduced: the former, called Dante work's sources, is concerned with the primary sources cited in Dante's works; the latter, called Dante work's structure, is concerned with the structure of Dante's work. For each of

them, the significant types of citation histograms were first identified. The structure Dante work's sources is applicable to both the Dante's works/a selected Dante's work. In the former case, this structure is sufficient to fulfill the project purposes; in the latter case, the histogram types of both the structures Dante work's sources and Dante work's structure need to be related. The two structures are described in the next sections.

Dante Work's Sources

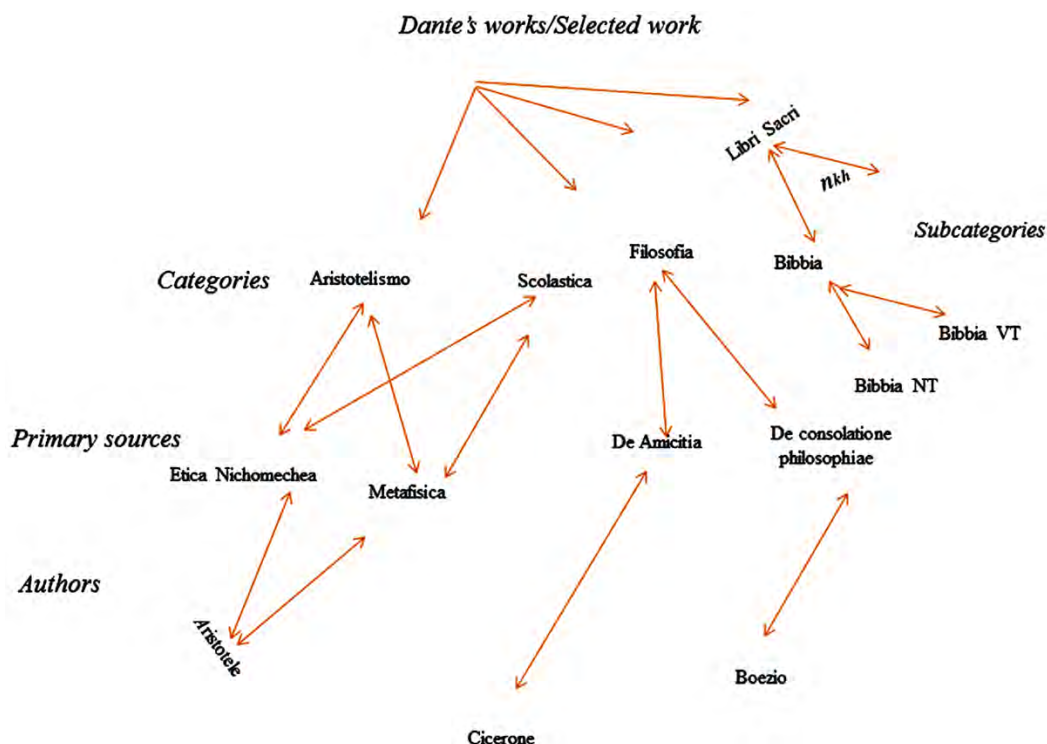
Dante work's sources is a high level structure exploited to identify the histogram types useful to the scholars. The structure is identified by the triple (categories, primary sources, authors). An example of this structure is provided in Figure 1. The label at the highest level is representative of the Dante's works/a selected Dante work.

The labels at the first level represent the categories under which the citations of the Dante's works/a selected Dante work fall. The categories level is organized in sub-categories defining a graph: moreover, the citations that fall under a category can be further broken. Conversely, the citations that fall under sub-categories fall also under a higher category. Thus, the branches of the category tree are oriented from the top to bottom level.

The labels at the level of primary source represent the sources cited in the Dante's texts. The citations that fall under a category/sub-category can be related to more primary sources. In Figure 1, the Aristotelism category is related to both the primary sources *Etica Nichomechea* and *Metafisica*. Conversely, each primary source can fall under many categories/subcategories. As an example, the citations to the primary source *Metafisica* fall under both the categories *Aristotelismo* and *Scolastica*.

Each node at the author level represents the author of one or more primary sources, as an example, *Aristotele* is the author of both the primary sources *Etica Nichomechea* and *Metafisica*, conversely, each primary source has been

Figure 1. Dante work's sources



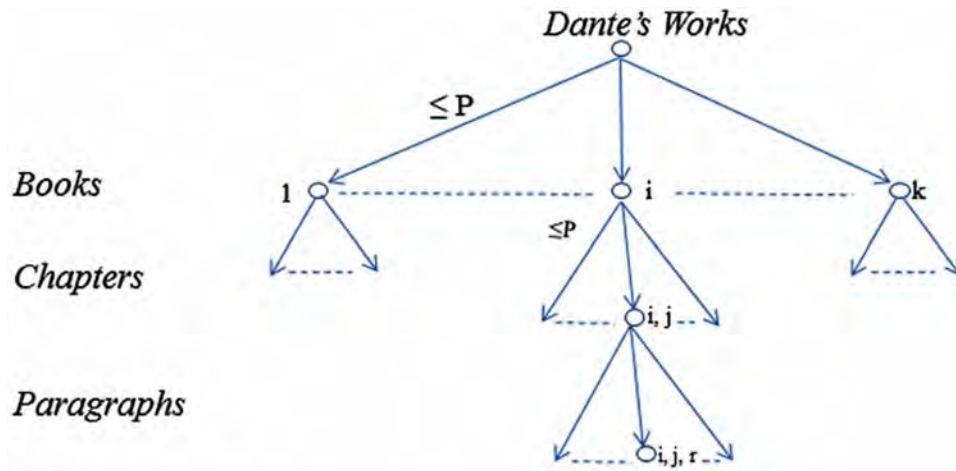
written by one and only one author; thus, only a link starts from a primary source to its author. Conversely, more links can start from an author to the primary sources.

In order to identify the significant types of histograms, the levels of the structure Dante work's sources are examined moving along the oriented directions of the triple (Categories, Primary sources, Authors). The oriented links are exploited to identify the refinements of the identified histogram types. As an example, if the citation histograms of a Dante's work that fall under categories have been identified, these histograms can be refined into subcategories. The types of citation histograms of Dante's works/a selected Dante's work determined from the structure Dante work's sources are the following:

- Citation histograms that fall under *categories*;
- Citation histograms that fall under *categories* refined in *subcategories*;

- Citation histograms that fall under *categories* refined in *cited primary sources*;
- Citation histograms that fall under *categories* refined in *authors of cited sources*;
- Citation histograms that fall under *sub-subcategories* of a *category*;
- Citation histograms that fall under *sub-subcategories* of a *category* refined in *cited primary sources*;
- Citation histograms that fall under *sub-subcategories* of a *category* refined in *authors of cited primary sources*;
- Histograms to *cited primary sources*;
- Histograms to *cited primary sources* refined in *categories/sub-categories*;
- Histograms to *authors of cited primary sources*;
- Histograms to *authors of cited primary sources* refined in *categories/sub-categories*.

Figure 2. Dante's work structure



Dante Work's Structure

The Dante work's structure is a tree identified by the triple of levels (books, chapters, paragraphs) and by the part of relation, a reflexive, asymmetric and transitive relation, symbolically represented by ($\leq P$). This will mean that the books are part of a work, the chapters are part of a book and the paragraphs are parts of a chapter. The root of the structure represents a Dante's work, as an example the *Convivio* work. Each node of the structure is characterized by an index if the node belongs to the book level, as an example the i^{th} book, two indices, if the node belongs to the chapter level, as an example the j^{th} chapter of the i^{th} book, and three indices if the node belongs to the paragraphs level, as an example the k^{th} paragraph of the j^{th} chapter of the i^{th} book. The tree branches are uni-directionally oriented from top downwards. Figure 2 represents the Dante Work's structure.

The tree levels (books, chapters, paragraphs) are examined moving from the top to the bottom level of the structure; analogous examination is performed for the two level subtrees and for the single level subtrees.

The following are the identified types of citation histograms determined by the Dante work's structure:

- Histograms of citations in the *books* of a Dante work.
- Histograms of citations in the *books* of a Dante works refined in *chapters*.
- Histograms of citations in the i^{th} book of a Dante's work.
- Histograms of citations in the i^{th} book of a Dante's work refined in *chapters*.
- Histograms of citations in to j^{th} chapter of the i^{th} book of a Dante's work refined in *paragraphs*.

In order to complete the list of the possible citation histograms types for a *Dante's work*, the above two sets of histograms have been related as follows:

- Histograms of citations in the *books* of a Dante's work with refinements in *categories/primary sources/authors of primary sources*.
- Histograms of citations in the i^{th} book of a Dante's work with refinements in *categories/primary sources/authors of primary sources*.
- Citation histograms of a Dante's work that fall under *categories/subcategories* with refinements in *books/chapters of the i^{th}*

- *book/ paragraphs of the j^{th} chapter of the i^{th} book.*
- Histograms of citations of a Dante's work to *cited primary sources* with refinements in *books/chapters of the i^{th} book, paragraphs of the j^{th} chapter of the i^{th} book.*
- Histograms of citations of a Dante's work to *authors of cited primary sources* with the refinements in *books/ chapters of the i^{th} book/ paragraphs of the j^{th} chapter of the i^{th} book.*

The Implementation Phase

The implementation phase is concerned with a table containing four columns identified by the 4-pla (Domain, Range, Graphical Type, Refinement). This table summarizes the high level structures and the knowledge from humanists. The significant types of citation histograms are represented in a minimal model that does not reduce the completeness of the graphical representations. The minimal model is not-unique.

In the first rectangle of the *Domain* column, the following options, related to the root of Dante work's source and to Dante work's structure can be chosen:

- Dante's works involving the set of all the Dante's works;

- Dante work involving a selected work of Dante;
- Part of Dante work involving a part (book, chapter, paragraph) of a Dante's work.

The same options can also be chosen in the second, in the third and in the fourth rectangles of the Range column. In the second rectangle of the Domain column, the following options, related to the former level of the Dante work's source, can be chosen:

- Categories involving the set of all the categories under which the citations can fall;
- Category involving a selected category;
- Sub-category involving a part of a selected category.

In the third and in the fourth rectangles of the Domain column, the following options, respectively related to the second and third level of the Dante work's source can be chosen:

- Primary sources involving the works cited by Dante in his works;
- Authors involving the authors of the cited works.

The options of Categories, Primary Sources and Authors proposed in the second, in the third

Table 1. The minimal model

Domain	Range	Graphic Type	Refinements
<i>Dante's works</i> <i>Dante work</i> <i>Part of Dante work</i>	<i>Categories</i> <i>Category</i> <i>Sub-categories</i>	→ 2 → 3 → 4	→2 in sub-categories → 3 in parts → 4 in works
<i>Categories</i> <i>Category</i> <i>Sub-categories</i>	<i>Dante's works</i> <i>Dante work</i> <i>Part of Dante work</i>	→ 5	→5 in parts
<i>Primary Sources</i>	<i>Dante's works</i> <i>Dante work</i> <i>Part of Dante work</i>	→ 6	→6 in parts
<i>Authors</i>	<i>Dante's works</i> <i>Dante work</i> <i>Part of Dante work</i>	→ 7	→7 in parts

and in the fourth rectangles of the Domain column are the same of the first rectangle of the Range column. The graphical types of histograms and their refinements are summarized in the third and in the fourth column of the table. If the Dante's works option is selected, no refinement is possible; otherwise, refinement by parts is always possible. Further, if the option categories is selected, a possible refinement is in sub-categories. Analogously, if the option author is selected, the histogram can be refined in works.

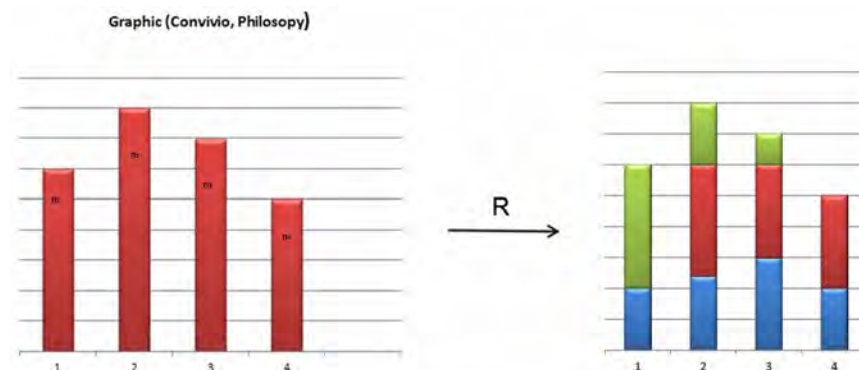
The graphical representation of the citation histograms is a function depending from the parameters Domain and Range, i.e., Graphic (Domain, Range). An option is selected for both the component of the couple (Domain, Range), as an example, Graphic (Convivio, Philosophy). In this case, the result of the function is a citation histogram enclosing the Convivio books on the horizontal axis; the distribution of the citations that fall under the Philosophy category on the vertical axis. The histogram can be refined as an example in chapters. Another possible refinement is for example in sub-categories. In Figure 3, the citation histogram of the proposed example is provided.

Further advancements of this work were concerned with the possibility to increase the knowledge acquired by the scholars with respect to other parameters, as an example the time parameter; thus, knowledge can also be extracted according to chronological order.

DISCUSSION

Statistical analysis is generally based on data proposed in tables or in visualization diagrams. In this chapter, data visualized through histograms are the starting point to support the scholars in the acquisition of knowledge on Dante's works. In order to explain how knowledge can be acquired, let us start from the function: Graphic (Domain, Range), in mathematics $f(x,y)$. For each couple of values, one for the Domain component and one for the Range component, one and only one histogram is returned. In Figure 3, the chosen values are: (Convivio, Philosophy) and the result of the Graphic (Convivio, Philosophy) function is a histogram. The histogram is visualized on a Cartesian diagram; on the horizontal axis, the four Convivio books are represented. The indices $n_1, \dots, n_4$ are the histogram bar heights. For the purposes of this discussion, it is not necessary to have the real citation distribution. The example evidences that all the Convivio chapters cite primary sources that fall under the Philosophy category, but that the citation distribution changes with the chapters. The possibility to refine the histogram exploiting one of the possible refinements proposed in Table 1 is also provided; for example, the citations falling under the Philosophy category can be refined in sub-categories represented by different colours. Let us now suppose that the former component of the couple (Domain, Range) is fixed to the Convivio value, and let us change only the latter

Figure 3. Citation histogram and refinement



component. In this way, the histograms of the *Convivio* books that cite primary sources falling under different categories, Philosophy, Bible, Aristotelism, are returned to the scholars. Comparisons among the obtained histograms give rise to data from which knowledge can be derived.

For advancement: Let us suppose that the former component of the couple (Domain, Range) is represented by intervals of time, in statistics temporal classes; and that the latter component is fixed to the Philosophy category. In this example, the intervals of time are represented on the horizontal axis of the Cartesian diagram, and the numbers of citations that fall under the chosen category are represented by the histogram bar heights. The histogram evidences the variability in time of the Dante's works citations that fall under the Philosophy category. Let us now suppose that other categories are chosen and that new histograms are returned to the scholars. Different types of comparison can be performed among histograms. Knowledge concerning the evolution in time of Dante's cultural background is an example of knowledge that can be derived following this approach.

FUTURE RESEARCH DIRECTIONS

The methodology and the tools developed within the *Towards a Digital Dante Encyclopedia* project can be easily reusable to represent the primary sources of other authors of the Italian or International Literature, different from Dante Alighieri. Future work is concerned with:

- The evaluations of the representational adequacy of the ontology by scholars, expert of Dante Alighieri's works;
- The usability of the implemented Web services through user tests.

The collaboration between humanists and computer scientists is worthy for possible future projects.

CONCLUSION

A research article related with one of the main activities of the *Towards a Digital Dante Encyclopedia* project is presented. The focus is on the identification and the implementation of visualization diagrams to support the knowledge acquisition on Dante's works. The approach proposed to achieve this objective is based on three phases. Two high level structure; an implementation model and a discussion are proposed within these phases. An example of citation histogram and refinement is provided. As achievements of this project, a web application to support scholars in writing a complete encyclopedia on Dante Alighieri was developed. The knowledge included in the semantic network concerning Dante's works can be extracted, and according to different parameters (chronological order, type of source, works, etc.) data about primary sources cited by Dante can be visualized. Conflicting aspects of Dante's biography and evolution in time of its culture can also be explored.

ACKNOWLEDGMENT

The authors thank Prof. Mirko Tavoni and Dr. Paola Andriani – Pisa University; Dr. Carlo Meghini and Loredana Versienti – CNR, for the useful collaboration during the project. A thank to Dr. Paola F. Canonico and to Dr. Andrea Vannozzi for the technical realization of the figures in this article.

REFERENCES

- Alighieri, D. (2011a). Opere.: Vol. I. *Rime, Vita Nova, De Vulgari Eloquentia*, a cura di Guglielmo Gorni. Milano, Italia: Mondadori.
- Alighieri, D. (2011b). Opere.: Vol. I. *Rime, Vita Nova, De Vulgari Eloquentia*. a cura di Mirko Tavoni. Milano, Italia: Mondadori.

Alighieri, D. (2014a). *Opere.: Vol. II. Convivio, Monarchia, Epistole, Egloghe, a cura di Giancarlo Fioravanti*. Milano, Italia: Mondadori.

Alighieri, D. (2014b). *Opere.: Vol. II. Convivio, Monarchia, Epistole, Egloghe, Milano, Mondadori, 2014, a cura di Diego Quaglioni*. Milano, Italia: Mondadori.

Bartalesi, V., Locuratolo, E., Versienti, L., & Meghini, C. (2013). A preliminary study on the semantic representation of the notes to Dante Alighieri's *Convivio*. In *Proceedings of the 1st International Workshop on Collaborative Annotations in Shared Environment: metadata, vocabularies and techniques in the Digital Humanities*. New York: ACM Press. doi:10.1145/2517978.2517983

Doerr, M. (2003). The CIDOC CRM - an ontological approach to semantic interoperability of metadata. *AI Magazine, Special Issue*, 24(3), 75–92.

Erling, O., & Mikhailov, I. (2009). RDF Support in the Virtuoso DBMS. In *Networked Knowledge Networked Media* (pp. 7–24). Berlin: Springer. doi:10.1007/978-3-642-02184-8_2

Le fonti dantesche. (n.d.). Retrieved from <http://dante1.isti.cnr.it:8080/perunaenciclopediadantescadigitale/>

Manola, F., & Miller, E. (2004). *RDF Primer. W3C Recommendation 10 February 2004*. Available at: <http://www.w3.org/TR/rdf-primer>

Miles, A., Matthews, B., Wilson, M., & Brickley, D. (2005) SKOS core: simple knowledge organisation for the web. *International Conference on Dublin Core and Metadata Applications*, 3–10.

Peroni, S., & Shotton, D. (2012). FaBiO and CiTO: Ontologies for describing bibliographic resources and citations. *Web Semantics: Science, Services, and Agents on the World Wide Web*, 17, 33–43. doi:10.1016/j.websem.2012.08.001

Tavoni, M., Andriani, P., Bartalesi, V., Locuratolo, E. I., Meghini, C., & Versienti, L. (2014). Towards a Semantic Network of Dante's Works and Their Contextual Knowledge. In *DH 2014 – Digital Humanities 2014*. Lausanne, Switzerland: UNIL-University of Lausanne.

Tillett, B. (2005). What is FRBR? A conceptual model for the bibliographic universe. *The Australian Library Journal*, 54(1), 24–30. doi:10.1080/00049670.2005.10721710

ADDITIONAL READING

Altman, M. (2006). *An overview of digital libraries: issues and trends* (pp. 1–10). Cambridge: Academic Research Microsoft.

Association for Computers and the Humanities. (1990). Guidelines for the encoding and interchange of machine-readable texts. C. M. Sperberg-McQueen, & L. Burnard (Eds.). ACH, ACL, ALLC.

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. *Scientific American*, 284(5), 28–37. doi:10.1038/scientificamerican0501-34 PMID:11341160

Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data-the story so far. *Semantic Services, Interoperability and Web Applications: Emerging Concepts*, 205–227.

Davidson, D. (2001). *Essays on actions and events: Philosophical essays* (Vol. 1). Oxford University Press. doi:10.1093/0199246270.001.0001

Doerr, M., & LeBoeuf, P. (2007). Modelling intellectual processes: the FRBR-CRM harmonization. In *Digital libraries: Research and development* (pp. 114–123). Springer Berlin Heidelberg. doi:10.1007/978-3-540-77088-6_11

Jordanous, A., Lawrence, K. F., Hedges, M., & Tupman, C. (2012). Exploring manuscripts: sharing ancient wisdoms across the semantic web. In *Proceedings of the 2nd International Conference on Web Intelligence, Mining and Semantics* (p. 44). ACM. doi:10.1145/2254129.2254184

McCarthy, J. (1987). Epistemological problems of artificial intelligence. *Readings in artificial intelligence*, 459.

Russell, S., & Norvig, P. (1995). Artificial intelligence: a modern approach.

Vanhoutte, E. (2004). An Introduction to the TEI and the TEI Consortium. *Literary and Linguistic Computing*, 19(1), 9–16. doi:10.1093/lc/19.1.9

KEY TERMS AND DEFINITIONS

Dante Work's Source: High level structure representative of the *Dante's works/a selected Dante's work*. The structure is identified by the labels of the triple (*categories, primary sources,*

authors) and by bi-oriented links relating the labels from one level to another. One and only one link starts from a label at the *author* level to a label at the *primary source* level.

Dante Work's Structure: Tree identified by a *root*, representative of a Dante work, by the triple of levels (*books, chapters, paragraphs*) and by the part of relation, symbolically represented by ($\leq P$).

Digital Humanities: Digital humanities is an area of research and teaching at the intersection of computing and the disciplines of the humanities.

Linked Data: Linked Data describes a method of publishing structured data so that it can be interlinked and become more useful through semantic queries.

Ontology: Is a model for describing the world that consists of a set of types, properties, and relationship types.

Semantic Web: The Semantic Web is an extension of the Web through standards by the World Wide Web Consortium (W3C). The standards promote common data formats and exchange protocols on the Web, most fundamentally the Resource Description Framework (RDF).

Knowledge Management for Development (KM4D)



Alexander G. Flor

University of the Philippines, Philippines

INTRODUCTION

Knowledge management, as a discipline, was borne out of the need to systematically leverage the intellectual assets of an organization to achieve corporate goals. We can trace its origins to Nonaka and Takeuchi's treatise on the knowledge creating company (1995), which was informed by the tacit-explicit knowledge dichotomy of Polanyi (1967). However, it was Davenport and Prusak (1997) who introduced the term knowledge management to the wider management science community.

At the very onset, knowledge management was intended as a tool for the private sector, a means to increase an organization's IQ, as Bill Gates (1999) puts it. Soon after its introduction to the corporate world, however, it was embraced by both the government sector and the international development assistance community owing to the fact that governments and international agencies are, by nature, knowledge organizations. These include United Nations agencies; international financial institutions such as the World Bank, the International Monetary Fund (IMF) and the International Fund for Agricultural Development (IFAD); regional financial bodies such as the Latin American Development Bank, the Asian Development Bank and the African Development Bank; and bilateral aid agencies as well, such as USAID, AusAID, CIDA, JICA, AFD, DFID and others.

Considering the urgent nature of the societal problems addressed by this area, the emphasis of the KM4D community for the past twenty-five years has been on practice rather than on research or theory building. There have been attempts to define the field, but the desirability of a grounded

approach to the discourse necessitated sound referents found in its practice. Thus, even academic journals devoted to this area (e.g., KM4D Journal) focused on field practice documentation and evaluation. Nevertheless, an attempt at defining KM4D by differentiating it from conventional KM can serve as a starting point of theory.

It is the intention of this chapter to differentiate corporate KM from KM4D. It enumerates KM4D sectors and themes used by the international development assistance community in the past two and a half decades. Furthermore, the chapter attempts to present a proto-typology of corporate KM and KM4D for purposes of future theory building and of thickening the knowledge management discourse.

BACKGROUND

Among the first to apply knowledge management to the development agenda and the public sphere was Stephen Denning, who headed the Knowledge Management Program of the World Bank. Denning (2000) employed a technique which he calls organizational storytelling to champion KM among his colleagues. He presents the beginnings of the KM4D narrative in the following account:

As a manager in the World Bank in 1996, I had been trying to communicate the idea of knowledge management and to get people to understand and to implement it. At that time in that organization, knowledge management was a strange and generally incomprehensible idea. I used the traditional methods of communicating with no success... In

my desperation, I was willing to try anything and eventually I stumbled on the power of a story, such as the following: "In June 1995, a health worker in a tiny town in Zambia logged on to the website for the Center for Disease Control in Atlanta Georgia and got the answer to a question on how to treat malaria....

This was June 1995, not June 2001. This was not the capital of Zambia but a tiny place six hundred kilometers away. This was not a rich country: this was Zambia, one of the poorest countries in the world. But the most important part of this picture for us in the World Bank is this: the World Bank isn't in the picture. The World Bank doesn't have its know-how accessible to all the millions of people who made decisions about poverty. But just imagine if it had... This story had helped galvanize staff and managers to imagine a different kind of future for the organization and to set about implementing it. Once knowledge management became an official corporate strategy later that year, I continued to use similar stories to reinforce and continue the change. The efforts were successful: by 2000, the World Bank was benchmarked as a world leader in knowledge management. (Denning, 2000)

There is reason to believe that the roots of knowledge management in the public sphere extended before Nonaka and Takeuchi. It may be traced to a group of pre-World War II Austrian academics who represented a school of thought, known as knowledge economics. Its luminaries were Hayek (1937) and Schumpeter (1942). The movement situated knowledge as a major economic variable but it was overshadowed by the Keynesian school, which dominated post-Bretton Woods economic theory and practice. In the sixties, it was resurrected in the United States by Machlup (1962) and later on by Porat (1978) albeit under a new brand, information economics. It was Machlup and Porat who introduced the Agricultural-Industrial-Information Age(s)

trichotomy as well as the concept of information economy or information society.¹

The KM4D discourse is likewise premised on the pitfalls of mishandling knowledge when it is not distinguished from information or wisdom, in part inspired by the T.S. Elliot poem, *The Rock*, a stanza of which reads:

Knowledge of speech, but not of silence

Knowledge of words, and ignorance of the Word

All our knowledge brings us nearer to our ignorance

All our ignorance brings us nearer to death

But nearness to death, no nearer to God

Where is the life we have lost in living?

Where is the wisdom we have lost in knowledge?

Where is the knowledge we have lost in information?

MAIN FOCUS OF THE CHAPTER

Differentiations

Paraphrasing Leibmann (1999), knowledge management is a nascent or newly emerging discipline that considers an organization's intellectual capital as a manageable and potentially profitable asset. Each and every organization possesses intellectual capital that may purposively be employed to achieve organizational goals, objectives, or targets. Currently, this intellectual capital can be managed mainly through the use of information and communication technology. Gates (1999) uses the digital nervous system metaphor, comparing an organization to an organism with a nervous system technologically enabled by computer

hardware, software and networks. This network of workstations and servers (known as a KM system) programmed to facilitate knowledge sharing and reuse is, to an organization, what a nervous system is to an organism.

Knowledge management for development or KM4D also considers intellectual capital as a manageable asset but it is leveraged not to increase profit but to further the development agenda. KM4D is likewise founded on knowledge science and knowledge economics. It is similarly anchored upon Davenport and Prusak's KM theory and thus adopts a digital environment. However, it is applied within the development context and targeted at the achievement of national, regional or global development goals.

From the point of view of governments and development assistance agencies, knowledge management need not be exclusively applied to organizations. It can very well be employed to non-going concerns such as projects and to larger systems, communities, or development sectors. For instance, knowledge within academia, scientific circles, or communities of practice, can be managed, particularly so with social media and content management systems.

A Prototypology

How exactly does KM4D differ from conventional KM?

Based on twenty-five years of KM4D practice among international development agencies, one can differentiate conventional KM from KM4D in terms of the following: purpose; considerations; thrust; levels of utilization; system functionalities and applications; and system content.

The purpose of conventional KM is economic in nature, specifically the creation of wealth using intellectual assets. It may be said that the purpose of KM is to enhance a company's ability to conduct its core business. Assuming that corporations are primarily profit-oriented, it may be assumed that the end goal of KM is to increase the company's

bottom line. On the other hand, KM4D owing to its origins in the international development assistance sector has always been linked to development goals. From 2000 to 2015, these were called the Millennium Development Goals or MDGs. The MDGs were restructured into what is now known as the Sustainable Development Goals or SDGs for the 2015 to 2030 timeframe. Both the MDGs and SDGs will be discussed in a separate section in this chapter.

In terms of primary considerations, conventional KM guards intellectual assets closely and is geared at managing these assets efficiently. Thus, intellectual property, security of and accountability for these assets figure prominently in its list of considerations. However, KM4D believes in the synergistic or non-zero-sum nature of knowledge. Hence, knowledge is non-proprietary and is considered as a free commodity to be shared openly.

This leads to the concept of knowledge sharing and reuse, the primary goal of knowledge management as proposed by Leibmann (1999). For conventional KM, knowledge sharing and reuse is encouraged freely within the corporate organization but rarely outside it. In KM4D, the parameters of knowledge sharing and reuse extend beyond the organization to all sectoral stakeholders. The thrust of conventional KM is internal knowledge sharing. The thrust of KM4D is external knowledge sharing.

This difference is seen in the design and utilization of KM and KM4D systems. Content of conventional KM systems tends to be demand driven while the content of KM4D systems is supply driven. The former is determined by organizational requirements while in the latter, affirmative action is taken to supply knowledge and information needs of stakeholders. Conventional KM systems are primarily for corporate use of a going concern. On the other hand, KM4D systems may be designed for use at the project level, the agency level, even the sectoral or thematic level. Traditionally, users of KM systems are communities of practice or CoPs. Users of

KM4D systems involve communities of interest (CoIs) and communities of champions (CoCs). Lastly, KM4D systems go beyond conventional messaging and collaboration, file sharing, content management and search functionalities but extend to online learning.

Thus, we arrive at the following proposed typology of conventional KM and KM4D highlighting their differentiations:

Information Technology and Knowledge Management

Intellectual capital comes in two forms: explicit knowledge and tacit knowledge. Explicit knowledge is documented and easily made available in operations manuals, research studies, reports and journals. On the other hand, tacit knowledge is more difficult to manage, because it resides inside a person's head. This, in fact, is the traditional view of knowledge science: that knowledge resides within individuals and thus cannot be effectively managed. As Sarkar (1980) aptly puts it,

knowledge begins with perception and ends with realization. In other words, it requires subjectivization, the conceptual union of the subject (or the knower) and the object (or the known).

Nonaka and Takeuchi (1995) believe that in knowledge organizations, tacit knowledge should be converted to explicit knowledge for it to become a manageable asset. Gates (1999) asserts that information technology has facilitated the digital capture of tacit knowledge and thus enabled its sharing and reuse. Prior to the development of information technology, intellectual capital was considered difficult to manage because tacit knowledge, in particular, primarily resides in the individual. A research organization can lay claim to a paid individual's time or even skills. But it cannot ordinarily share ownership of an individual's knowledge permanently unless it is documented and copyrighted by the organization, which is relatively rare. With information technology, such knowledge can be captured, stored and shared electronically, in short, managed. In its electronic form, knowledge becomes part of content.

Table 1. Prototypology of conventional KM and KM4D

	Conventional KM	KM4D
Purpose	To increase company's bottom line To enhance company's ability to conduct its core business	To further development goals
Considerations	Security and accountability	Synergy
Sharing Parameters	Within the organization	All development stakeholders
Thrust	Internal knowledge sharing (within the company)	External knowledge sharing (for project/agency beneficiaries and other stakeholders)
Content	Demand driven	Supply driven
Levels of Utilization	Organizational level Among communities of practice (CoPs)	Project level Agency level Sectoral level Thematic level Among communities of interests (CoIs) Among communities of practice (CoPs) Among communities of champions (CoCs)
System Functionalities/ Applications	Messaging and Collaboration File sharing Documents/content management Search	Messaging and Collaboration File sharing Documents/content management Search Online learning

FUTURE RESEARCH DIRECTIONS

Knowledge management research may be categorized as thematic or sectoral. By 2005, the international development assistance community had seen it fit to classify development programs under two major categories: sectors and themes. Sectors include agriculture, health, education, environment, natural resources and others. Themes involve crosscutting concerns such as governance, gender, poverty, sustainability, and climate change. KM4D likewise adopts these sectors and themes.

KM4D interventions and, subsequently, research were designed and implemented to contribute to long-term targets of any one of the eight Millennium Development Goals or MDGs:

- **MDG 1:** Eradicate extreme poverty
- **MDG 2:** Achieve universal primary education
- **MDG 3:** Promote gender equality and empower women
- **MDG 4:** Reduce child mortality
- **MDG 5:** Improve maternal health
- **MDG 6:** Combat HIV/AIDS, malaria and other diseases
- **MDG 7:** Ensure environmental sustainability
- **MDG 8:** Global partnership for development

Identified during the Millennium Summit of 2000, these goals were expected to be achieved by 2015. By the beginning of 2015, it has become clear that the MDGs will have to be replaced by a new set of goals, now called Sustainable Development Goals of SDGs. The list has been expanded from eight (8) MDGs to seventeen (17) SDGs:

- **SDG 1:** End poverty in all its forms everywhere
- **SDG 2:** End hunger, achieve food security and improved nutrition and promote sustainable agriculture

- **SDG 3:** Ensure healthy lives and promote well-being for all at all ages
- **SDG 4:** Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all
- **SDG 5:** Achieve gender equality and empower all women and girls
- **SDG 6:** Ensure availability and sustainable management of water and sanitation for all
- **SDG 7:** Ensure access to affordable, reliable, sustainable and modern energy for all
- **SDG 8:** Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all
- **SDG 9:** Build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation
- **SDG 10:** Reduce inequality within and among countries
- **SDG 11:** Make cities and human settlements inclusive, safe, resilient and sustainable
- **SDG 12:** Ensure sustainable consumption and production patterns
- **SDG 13:** Take urgent action to combat climate change and its impacts*
- **SDG 14:** Conserve and sustainably use the oceans, seas and marine resources for sustainable development
- **SDG 15:** Protect, restore and promote sustainable use of terrestrial ecosystems, sustainably manage forests, combat desertification, and halt and reverse land degradation and halt biodiversity loss
- **SDG 16:** Promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable and inclusive institutions at all levels
- **SDG 17:** Strengthen the means of implementation and revitalize the global partnership for sustainable development

The intention now is to leverage the collective knowledge of the international development assistance community for achieving these goals by 2030. How this can be effectively and efficiently done should be the subject of extensive research within the academic community shaping the KM4D discourse.

CONCLUSION

Knowledge management began in the private sector but has since been adopted by governments and international development agencies alike. In the mid-nineties, Stephen Denning established the Knowledge Management Program of the World Bank, which has served as the model for applying KM to international development assistance or KM4D. In defining KM4D, this chapter reflected the need for a grounded approach by using field practice experience to differentiate corporate KM from KM4D.

The chapter argues that conventional KM and KM4D differ in terms of purpose, considerations, thrust, levels of utilization, system functionalities and applications, and system content. It concludes that KM4D research can best be classified under sectors and themes used by the international development assistance community in the past two and a half decades.

REFERENCES

Davenport, T., & Prusak, L. (1997). *Information Ecology: Mastering the Information and Knowledge Environment*. New York: Oxford University Press.

Davenport, T., & Prusak, L. (1997). *Working Knowledge, How Organizations Manage What They Know*. Boston: Harvard Business School Publishing.

Denning, S. (2000). *How Storytelling Ignites Action in Knowledge-Era Organizations*. New York: Butterworth Heinemann.

Gates, W. (1999). *Business at the Speed of Thought: Using a Digital Nervous System*. New York: Time Warner.

Hayek, F. (1937). *Economics and Knowledge. Economica*. London: Wiley.

Jussawalla, M., Lamberton, D., & Karunaratne, N. (Eds.). (1989). *The Cost of Thinking*. Ablex Publishing.

Lave, J., & Wenger, E. (1991). *Situated learning. Legitimate peripheral participation*. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511815355

Leibmann, M. (1999). *A way to KM solutions: Things to consider when building knowledge management solutions with Microsoft technologies*. World Wide Technical Services, Microsoft Corporation.

Machlup, F. (1962). *The Production and Distribution of Knowledge in the United States*. Princeton University Press.

Negroponte, N. (1995). *Being Digital*. New York: Alfred P. Knopf.

Nonaka, I., & Hirotaka, T. (1995). *The Knowledge-Creating Company*. New York: Oxford University Press.

Open Working Group. (2015). *Proposal for Sustainable Development Goals*. Full report of the Open Working Group of the General Assembly on Sustainable Development Goals. UN Document A/68/970. Retrieved from <http://undocs.org/A/68/970>

Polanyi, M. (1967). *The Tacit Dimension*. London: Routledge and Kegan Paul.

Porat, M. (1977). *The Information Economy*. Washington, DC: US Department of Commerce.

Sarkar, P. R. (1980). *The Faculty of Knowledge (Parts 1 to 5). A Few Problems Solved*. Ananda Nagar: Gurukula University.

Schumpeter, J. (1942). *Capitalism, Socialism and Democracy*. New York: Harper.

ADDITIONAL READING

APQC (American Productivity and Quality Center). (1996). *Knowledge management—consortium benchmarking study, Final Report*. Houston: APQC.

Balaji, V., Bagga, M., Kumar, R., Lavanya, K. N., Sabitha, P., Prabhakar, T. V., & Yadav, V. et al. (2010). *Agrotags – a tagging scheme for agricultural digital objects*. Berlin, Heidelberg: Springer. doi:10.1007/978-3-642-16552-8_4

Batchelor, S. (2002). *Using ICTs to generate development content. IICD Research Report 10*. The Hague: International Institute for Communication and Development.

Boom, D. 2005. The Asian Development Bank's knowledge management framework. *KM4D Journal* 1(2): 69-75 www.km4dev.org/journal

Cecchini, S., & Scott, C. (2003). Can information and communications technology applications contribute to poverty reduction? Lessons from Rural India. *Information Technology for Development*, 10(2), 73–84. doi:10.1002/itdj.1590100203

Chapman, R., & Slaymaker, T. (2002). ICTs and rural development: review of the literature, current interventions, and opportunities for action. *ODI Working Paper 192*. London: Overseas Development Institute.

Colle, R. D., & Roman, R. (2002). *Content for ICT initiatives. International Congress of Information Conference on information, knowledge and society, challenges of a new era*. Havana, Cuba. April 22-26, 2002.

Disterer, G. (2001). Individual and social barriers to knowledge transfer. *Proceedings of the 34th Hawaii International Conference on System Sciences*, Hawaii. doi:10.1109/HICSS.2001.927138

Flor, A. G. (2001). eDevelopment and Knowledge Management: ICT Applications for Sustainable Development. Los Baños, Laguna: SEAMEO Regional Center for Graduate Study and Research in Agriculture.

Flor, A. G. (2002). Knowledge Management: An Emerging Model for Agriculture and Environment Education Development. Searching for New Models of Agriculture Education in a Disturbed Environment (Samuel T. Mancebo, Marilyn B. Perlas, and Frances Muriel L. Tuquero, Editors.). Bangkok: Food and Agriculture Organization of the United Nations Regional Office for Asia and the Pacific.

Flor, A. G. (2003). *Ethnovideography: Video Based Indigenous Knowledge Systems*. Los Baños, Laguna: SEAMEO Regional Center for Graduate Study and Research in Agriculture.

Flor, A.G. (2011). *Knowledge Management for Climate Change: Strategic and M&E Plans (2012-2016)*. Quezon City: Department of Environment and Natural Resources and *Gesellschaft für Internationale Zusammenarbeit*

Flor, A. G. (2013). Exploring the downside of open knowledge resources: The case of indigenous knowledge systems. *Open Praxis. Journal of the International Council for Open and Distance Education*, 5(1), 75–80.

Flor, A. G., Amoloza, E. M., & Flor, A. G. Jr. (2011). *Design, Development and Testing of an Indigenous Knowledge Management System Using Mobile Device Video Capture and Web 2.0 Protocols*. Los Baños: UP Open University.

Jamali, D., & Zurayk, R. (2004) A community of practice approach to knowledge management Presented at the 5th European Conference on Knowledge Management, 30 September-1 October 2004, Unpublished paper.

Kollock, P. (1999). The economies of online co-operation: gifts and public goods in cyberspace. In M. Smith & P. Kollock (Eds.), *Communities in cyberspace*. London: Routledge.

Llewellyn, R. S. (2007). Information quality and effectiveness for more rapid adoption decisions by farmers. *Field Crops Research*, 104(1-3), 148–156. doi:10.1016/j.fcr.2007.03.022

Marwick, A. D. (2001). Knowledge Management Technology. *IBM Systems Journal*, 40(4), 814–830. doi:10.1147/sj.404.0814

McGrath, S., & King, K. (2004). Knowledge based aid: A four agency comparative study. *International Journal of Educational Development*, 24(2), 167–181. doi:10.1016/j.ijedudev.2003.10.008

Sulaiman, R. V., Hall, A., Kalaivani, N. J., Dorai, K., & Reddy, T. S. (2012). Necessary but not sufficient: Critiquing the role of Information and Communication Technology in putting knowledge into use. *Journal of Agricultural Education and Extension*, 18(4), 331–346. doi:10.1080/1389224X.2012.691782

Sulaiman, R. V., Thummuru, L., Hall, A., & Dijkman, J. (2011). Tacit knowledge and innovation capacity: Evidence from the Indian livestock sector. *Knowledge Management for Development Journal*, 7(1), 32–44. doi:10.1080/19474199.2011.593855

Wenger, E. (1998). *Communities of practice: learning, meaning, and identity*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511803932

KEY TERMS AND DEFINITIONS

Communities of Champions (CoC): Online as well as offline social networks who advocate a specific development policy, strategy or intervention. These advocates are often the *de facto* spokespersons of stakeholders in the development process.

Communities of Interest (CoI): Networks of persons interested in a common problem who bond together online as well as offline.

Communities of Practice (CoP): Online as well as offline social networks who share a passion for a practice. Their practice improves with regular online and offline interactions as theorized by Lave and Wenger (1991).

Development Stakeholder: Roughly categorized as funding agencies, government agencies, nongovernment organizations (NGOs), civil society and beneficiaries of development aid.

Knowledge Management (KM): An emerging discipline that treats intellectual capital as an asset that may be managed through information and communication technologies.

Knowledge Management for Development (KM4D): Undertakings, systems and products that leverage intellectual capital in the pursuit of national, regional or global development goals.

Prototypology: A prototype typology that is used as an initial device in the theory building process.

ENDNOTE

- ¹ Defined as a society whose economy is information based, i.e. the majority of its workforce is made up of information workers or the greater part of its GNP may be attributed to information products, services and labor (Porat, 1978).

Knowledge Management From the Metaphorical Perspective

K

Magdalena Bielenia-Grajewska
University of Gdansk, Poland

INTRODUCTION

The purpose of this chapter is to present the notions connected with how knowledge management is communicated. The focus of investigation is on the place of metaphors in creating, acquiring and sharing data. The metaphorical side of knowledge management is discussed from both theoretical and practical perspectives. Thus, selected theories are supported with practical examples coming from diversified sources on knowledge management. The discussion aims at showing the positive and negative role of metaphors in various stages of knowledge management.

BACKGROUND

The way knowledge is coded and decoded has changed throughout centuries. Nowadays we can observe the growing role of technology in creating and sharing knowledge in educational (e.g. Turula, Mikołajewska & Stanulewicz, 2015) and professional settings (Bielenia-Grajewska, 2016). Thus, traditional ways of accessing and using knowledge are supported, and in some cases substituted, with virtual sources of data. Apart from the plethora of knowledge resources, the phenomenon of knowledge poverty resulting from limited or no access to information still exists. There are different types of knowledge poverty that can be categorized by taking into account such notions as purpose, target audience,

and form of dissemination. Thus, the following typology of knowledge poverty can be applied: scientific knowledge poverty, everyday knowledge poverty, professional knowledge poverty, social knowledge poverty, cultural knowledge poverty and linguistic knowledge poverty. The last type of knowledge poverty can be understood in at least two ways. First, linguistic knowledge poverty may be represented by no or limited familiarity with the official language spoken in a country one has to study or work in. The second type of linguistic knowledge poverty is connected with the minor positions of some languages or dialects in a given community (Bielenia-Grajewska, 2015b). Another phenomenon connected with restricted understanding and using knowledge is attention crash, represented in limited concentration resulting from information overload. It can be observed especially among people exposed to many impulses (mainly of a technological nature) at the same time. Nowadays this phenomenon is associated with constant access to emails, sms, chatting and talking on the phone (Bielenia-Grajewska, 2014b). One method of dealing with attention crash and knowledge poverty is to handle proper and effective communication. In addition, the linguistic turn in knowledge management is connected with the dynamic character of knowledge. For example, Wittgenstein (1992) highlights that it cannot be determined how long knowledge, skills and understanding last. Consequently, the character of knowledge management (KM) requires proper linguistic forms that will facilitate the process of

perception, comprehension and dissemination of new concepts (Bielenia-Grajewska, Carayannis & Campbell, 2013b).

LANGUAGE AND KNOWLEDGE

Recent publications written by the representatives of different studies draw the attention of readers to the discursive dimension of knowledge management. An example of such domains is discursive psychology. As Crane (2016) discusses in her contribution, interaction is crucial in knowledge creation and knowledge sharing. Thus, talk and text play an important role in KM. The narrative side of knowledge management is also stressed by Schreyögg and Koch (2005) who elaborate on the issue of narration and storytelling as well as their place in KM. In addition, the link between language and culture is discussed by researchers including Holden and Glisby (2014).

The functionality of language in knowledge management can be discussed through the prism of micro, meso and macro perspectives. Starting with the micro approach, language is applied to denote a new concept or a novel idea. The meso level is represented by the usage of linguistic tools in shaping texts. As far as the macro level is concerned, language is also used as a tool facilitating knowledge flows between different domains (Bielenia-Grajewska, 2012). It should also be highlighted that language and knowledge management do not exist in a vacuum; they shape and simultaneously they are shaped by other domains. An example is education - modern education differs from that which could be observed some years ago as teaching and learning of the twenty-first century involve not only various learners in terms of their age, gender and background, but also diversified methods of encoding and decoding knowledge. For example, with the appearance of MOOCs and other educational offers available online, modern education is not as restricted in terms of geography, prior levels of knowledge or types of accessible educational tools as it was

before. Consequently, knowledge is also gained in activities that were previously associated exclusively with the ludic character of our life. An example is the application of serious games in learning (e.g. Bielenia-Grajewska, 2016).

The way language facilitates knowledge management can also be discussed by taking into account literal and non-literal language. Focusing on the complexity of figurative language and its multidimensionality, symbolic communication in KM is given a detailed study in this contribution.

METAPHORS

The purpose of metaphors is to make people perceive one sphere of experience in terms of another, whereas cross-domain interconnections facilitate the observation of similarities between two domains within a given metaphor (Lakoff & Johnson, 2003). Metaphors are used in different domains of life, including, among others, investment banking (Bielenia-Grajewska, 2009), leadership (Lancaster, 2015), neuromanagement (Bielenia-Grajewska, 2013a), translation (Guldin, 2016) and biology (Boldt, 2016). The functions of metaphors are as follows. First, they are efficient in denoting novel objects or phenomena. An example is the area of investment banking, characteristic of new terms describing such economic notions as new strategies of mergers and acquisitions or derivatives (Bielenia-Grajewska, 2009). In other words, symbolic language facilitates the process of introducing new products or services, providing efficient linguistic tools to accompany economic or technological development (Bielenia-Grajewska, 2012). Their “freshness of expression”, novelty in form, makes them popular in coding and decoding different forms of knowledge. Wittgenstein (1992) underlines that a new word is like a fresh seed thrown in the soil of discussion. In the case of KM, new terms coined in professional settings may enrich specialized knowledge and they should facilitate communication between laymen and professionals. Moreover, metaphors are useful in

handling risk communication; they can be used to discuss e.g. food-borne diseases (Bielenia-Grajewska, 2015c). It should also be mentioned that metaphors are not passive communication tools; linguistic symbols construct the reality, imposing a given perspective (Kövecses, 2015). Moreover, Wittgenstein (1989) compares words to acorns from which an oak tree may grow. This comparison may be used to discuss the place of metaphors in KM discourse. Symbolic language, if used effectively, can facilitate an understanding of novel domains as well as coding and decoding new data. Metaphors can be used in business education, to transfer new knowledge from professors to students or from experts to laymen. An example is the usage of metaphorical names relying on the domain of gold to denote the names of participants, instruments and strategies of the modern economic discourse (Bielenia-Grajewska, 2015d). The metaphorical names include golden coffin, golden hello, golden umbrella or golden parachute. The next important feature of symbolic language is its attractiveness; metaphorical names draw one's attention more easily than literal expressions (Charteris-Black, 2007). For example, such M & A (mergers and acquisitions) strategies as black knight or porcupine defense sound more attractive for laymen than their non-figurative descriptions. Another key feature of metaphors is their contextuality (Kövecses, 2015). Metaphors do not exist in a vacuum; they reflect the reality they are part of. The next key issue is the economical character of symbolic communication. In comparison with standard elaborate definitions, metaphorical names occupy (in most cases) up to two words. This feature is especially important in headlines or sneak peaks. Applying metaphors in the discussion on knowledge can also facilitate an understanding of knowledge as a complex concept. Bratianu (2015) in his contribution discusses such examples as *knowledge as energy*, and related subcategorizations, such as cognitive knowledge as a mechanical energy, emotional knowledge as a thermal energy and knowledge dynamics as energy thermodynamics. The functionality of

metaphors in knowledge management can also be discussed through the lens of deliberateness. The purposes and conscious usage of metaphors in KM can be studied from the perspective of Deliberate Metaphor Theory, stressing the non-automatic and intentional use of symbolic language in modern communication, for example, in journalism (Steen, 2015).

Modern contributions on knowledge also stress the role of identity in different stages and methods of KM, especially in technological settings. The growing role of virtual education is also described in new contributions devoted to the role of symbolic language in learning outside the regular classroom. For example, Ligorio, Iodice and Manca (2016) discuss how metaphors not only stimulate knowledge enhancement in virtual settings, but also build a common identity of online learners.

Technology has also influenced the way we perceive the surrounding reality and ourselves. For example, Carr (2016) draws our attention to the way the metaphorical perception of memory has changed. In the past, bees and nectars were used to depict the relation between memory and intellect (e.g. Seneca), whereas nowadays computers and database are used to mirror the above-mentioned relation.

Metaphors and Organizations

The dialogic character of modern organizations is discussed in different contributions (e.g. Bushe & Marshak, 2015), including the knowledge perspective. Organizational knowledge is defined by Bratianu (2015) as the concept encompassing cognitive, emotional and spiritual knowledge (the triple helix metaphor). The linguistic dimension of companies and institutions with regard to KM is created and exercised largely by symbolic language. Metaphors play different functions in organizational communication. First, they are used to stress organizational culture. An example is the study conducted by Bielenia-Grajewska on companies operating in the food industry. This

research has shown that such organizational metaphors as organization as a teacher, organization as a network, organization as a protector, organization as a traditionalist, organization as a travel guide, and organization as a family are used by companies to depict their activities and policies (Bielenia-Grajewska, 2014a). Other contributions (e.g. Zamparini & Lurati, 2016) stress the role of multimodal metaphors in shaping company identity. Since organizations do not only produce or offer services but also knowledge that accompanies their merchandise, metaphors can be used to make their products or services more visible on the competitive market. Organizational innovations are often accompanied by metaphorical slogans or pictorial metaphors to draw the attention of potential stakeholders.

Functions of Metaphors in Knowledge Management-Stage Perspective

The functions of metaphors in different knowledge processes can be discussed by taking into account the key characteristics of different stages connected with managing knowledge. The most representative features are presented by using the Four-dimensional model of metaphorical KM communication (Figure 1), focusing on knowledge management and the place of metaphors in it. The discussion is an enhanced version of the one presented in Bielenia-Grajewska (2015a).

Metaphors and Knowledge Creation

Metaphors turn out to be efficient in the first stage of knowledge management, mainly, in creating novel data and information. Their efficiency and simplicity often result in a better understanding of concepts among specialists working on new approaches or products. Taking into account heterogenic working groups (constituting people representing different professions and educational

Figure 1. Five-dimensional model of metaphorical KM communication (created by the current author)



backgrounds), symbolic language facilitates effective communication, especially when highly specialized notions and technical data are involved.

Metaphors and Knowledge Encoding

Knowledge creators should find relevant tools and channels for encoding novel data. Metaphors play an important role in putting new ideas into linguistic representation known and understood by a large number of users who differ in terms of their professional knowledge, age or educational background.

Metaphors and Knowledge Decoding

Bailin and Grafstein (2016) discuss the role of metaphors in text readability, analyzing such notions as the time needed to decode symbolic messages, in comparison with literal communication. It should be mentioned, however, that some researchers underline the potential negative side of using metaphors in knowledge decoding. Since metaphors rely on one's individual understanding, it may lead to the misinterpretation of information.

For example, the substitution of real numbers with metaphorical names to depict contagious diseases may lead to the wrong perception of risk.

Metaphors and Knowledge Application

Metaphors are used in applying knowledge since they make complex ideas easier to implement when the explanation is supported with concepts coming from well-known domains. It is especially important when knowledge is applied for users who do not possess things like professional or technical expertise or when individuals differ as far as their educational background is concerned.

Metaphors and Knowledge Sharing

This function of metaphors is visible in offering quick and efficient knowledge communication. Metaphors stimulate knowledge exchange, especially among users with diversified level of knowledge on a given topic. In the case of laymen-professionals communication, metaphors facilitate effective dialogue by offering commonly understood connotations. Symbolic language is also efficient for discussions conducted by specialists representing different professions and knowledge branches.

SOLUTIONS AND RECOMMENDATIONS

The discussion of the role of metaphors in knowledge management shows how symbolic language can be used in various spheres of life. As has been presented here, metaphors can influence knowledge management in both positive and negative ways. Thus, they have to be chosen carefully in order to facilitate knowledge flows. In addition, taking the dynamic character of KM into account, knowledge creators should select the linguistic

tools that will meet the needs and expectations of participants in knowledge communication.

FUTURE RESEARCH DIRECTIONS

It can be predicted that metaphors will be an important tool of knowledge communication in the future. There are several areas that are likely to influence the metaphorical dimension of knowledge management. The first one is neuroscience, with its tools and methods used in social studies and humanities as well as management. For example, the usage of fMRI (functional magnetic resonance imaging) can facilitate understanding about how people perceive metaphors and how they react to them. Another important branch is technology, offering new ways of coding, decoding, storing and disseminating knowledge. For example, the growing role of video games has led to the appearance and popularity of audiovisual metaphors used in designing player characters (Fahlenbrach & Schröter, 2016). Taking into account the growing role of technology in the twenty-first century, together with the increasing place of non-human actors in interactions (Tatnall & Davey, 2015) and the broad application of computer-assisted tools in linguistic analysis, especially in computational linguistics, another topic of interest is the role of computers in interpreting metaphors (Weber Russell, 2016). The next future direction can be the shift from metaphor as a research object to metaphor as a research tool (Cameron, 2014). Thus, novel research methodologies may offer symbolic communication as a way of studying knowledge management.

CONCLUSION

As has been presented here, metaphors are important in different stages of knowledge management, such as creation, encoding, decoding and

dissemination. This discussion mainly centered on the positive sides of symbolic language in knowledge management. However, it should be remembered that metaphors may also have a negative impact on KM. As with other linguistic choices, the selection of metaphors should mirror the topic to be covered as well as the audience to be addressed.

REFERENCES

- Bailin, A., & Grafstein, A. (2016). *Readability: Text and Context*. Basingstoke, UK: Palgrave Macmillan. doi:10.1057/9781137388773
- Bielenia-Grajewska, M. (2009). The role of metaphors in the language of investment banking. *Ibérica*, 17, 139–156.
- Bielenia-Grajewska, M. (2012). Linguistics. In W. S. Bainbridge (Ed.), *Leadership in Science and Technology. A Reference Handbook* (pp. 41–48). Thousand Oaks, CA: SAGE Publications, Inc. doi:10.4135/9781412994231.n5
- Bielenia-Grajewska, M. (2013a). International Neuromanagement. In D. Tsang, H. H. Kazeroony, & G. Ellis (Eds.), *The Routledge Companion to International Management Education* (pp. 358–373). Abingdon, UK: Routledge.
- Bielenia-Grajewska, M. (2014a). CSR Online Communication. The Metaphorical Dimension of CSR Discourse in the Food Industry. In R. Tench, B. Jones, & W. Sun (Eds.), *Communicating Corporate Social Responsibility: lessons from theory and practice* (pp. 311–333). Bingley, UK: Emerald. doi:10.1108/S2043-9059(2014)0000006030
- Bielenia-Grajewska, M. (2014b). Attention Crash. In K. Harvey & G. J. Geoffrey (Eds.), *Encyclopedia of Social Media and Politics* (pp. 88–89). SAGE.
- Bielenia-Grajewska, M. (2015a). The role of figurative language in knowledge management. In M. Khosrow-Pour (Ed.), *Encyclopedia of Encyclopedia of Information Science and Technology* (pp. 4728–4736). Hershey, PA: IGI. doi:10.4018/978-1-4666-5888-2.ch464
- Bielenia-Grajewska, M. (2015b). Knowledge poverty. In M. Odekon (Ed.), *The SAGE Encyclopedia of World Poverty* (pp. 882–884). Thousand Oaks, CA: SAGE.
- Bielenia-Grajewska, M. (2015c). Metaphors and risk cognition in the discourse on foodborne diseases. In J. M. Mercantini & C. Faucher (Eds.), *Risk Cognition* (pp. 89–105). Berlin: Springer.
- Bielenia-Grajewska, M. (2015d). Rola języka symbolicznego w edukacji ekonomicznej na przykładzie nazw metaforycznych wykorzystujących domenę złota. *Edukacja Ekonomistów i Menadżerów*, 3(37), 57–70.
- Bielenia-Grajewska, M. (2016). Serious games and their application in creating corporate identity. In R. Nakatsu, P. Ciancarini, & M. Rauterberg (Eds.), *Handbook of Digital Games and Entertainment Technologies* (pp. 1–18). Singapore: Springer. doi:10.1007/978-981-4560-52-8_53-1
- Bielenia-Grajewska, M., Carayannis, E. G., & Campbell, D. F. J. (2013b). Linguistic Dimension of Creativity, Invention, Innovation, and Entrepreneurship. In E. G. Carayannis (Ed.), *Encyclopedia of Creativity, Invention, Innovation and Entrepreneurship* (pp. 1206–1215). London: Springer. doi:10.1007/978-1-4614-3858-8_478
- Boldt, J. (2016). *Synthetic Biology: Metaphors, Worldviews, Ethics, and Law*. Wiesbaden: Springer. doi:10.1007/978-3-658-10988-2
- Bratianu, C. (2015). *Organizational Knowledge Dynamics: Managing Knowledge Creation, Acquisition, Sharing, and Transformation*. Hershey, PA: IGI.

- Bushe, G. R., & Marshak, R. J. (2015). *Dialogic Organization Development: The Theory and Practice of Transformational Change*. Oakland, CA: Berrett-Koehler Publishers.
- Cameron, L. (2014). Metaphor in Spoken Discourse. In J. P. Gee & M. Handford (Eds.), *The Routledge Handbook of Discourse Analysis* (pp. 342–355). Abingdon: Routledge.
- Carr, N. (2016). Misled by Metaphor. In S. Groes (Ed.), *Memory in the Twenty-First Century. New Critical Perspectives from the Arts, Humanities, and Sciences* (pp. 67–69). Basingstoke, UK: Palgrave Macmillan.
- Charteris-Black, J. (2007). *The communication of leadership: the design of leadership style*. Abingdon, UK: Routledge.
- Crane, L. (2016). *Knowledge and Discourse Matters. Relocating Knowledge Management's Sphere of Interest onto Language*. Hoboken, NJ: Wiley.
- Fahlenbrach, K., & Schröter, F. (2016). Embodied Avatars in Video Games. Audiovisual Metaphors in the Interactive Design of Player Characters. In K. Fahlenbrach (Ed.), *Embodied Metaphors in Film, Television, and Video Games: Cognitive Approaches* (pp. 251–268). New York: Routledge.
- Guldin, R. (2016). *Translation as Metaphor*. Abingdon, UK: Routledge.
- Holden, N., & Glisby, M. (2014). The Nonaka-Takeuchi model of knowledge conversion: A discussion of many contexts of Japanese history and culture. In A. Örtenblad (Ed.), *Handbook of research on knowledge management* (pp. 366–387). Cheltenham, UK: Edward Elgar. doi:10.4337/9781783470426.00031
- Kövecses, Z. (2015). *Where Metaphors Come From: Reconsidering Context in Metaphor*. Oxford, UK: Oxford University Press. doi:10.1093/acprof:oso/9780190224868.001.0001
- Lakoff, G., & Johnson, M. (2003). *Metaphors we live by*. Chicago: The University of Chicago Press. doi:10.7208/chicago/9780226470993.001.0001
- Lancaster, S. (2015). *Winning Minds: Secrets From the Language of Leadership*. Basingstoke, UK: Palgrave Macmillan. doi:10.1057/9781137465948
- Ligorio, M. B., Iodice, M., & Manca, S. (2016). In E. Gola & F. Ervas (Eds.), *Metaphor and Communication* (pp. 235–248). Amsterdam: John Benjamins Publishing Company. doi:10.1075/milcc.5.13lig
- Schreyögg, G., & Koch, J. (2005). *Knowledge Management and Narratives: Organizational Effectiveness Through Storytelling*. Berlin: Erich Schmidt Verlag GmbH & Co KG.
- Steen, G. (2015). Developing, testing and interpreting deliberate metaphor theory. *Journal of Pragmatics*, 90, 67–72. doi:10.1016/j.pragma.2015.03.013
- Turula, A., Mikołajewska, B., & Stanulewicz, D. (2015). *Insights into Technology Enhanced Language Pedagogy*. Frankfurt am Main: Peter Lang. doi:10.3726/978-3-653-04995-4
- Weber Russell, S. (2016). *Computer Interpretation of Metaphoric Phrases*. Boston: Walter de Gruyter Inc.

ADDITIONAL READING

- Auer, P. (1998). *Code-Switching in Conversation: Language, Interaction and Identity*. London: Routledge.
- Bielenia-Grajewska, M. (2013). Genderlect and Linguistic Styles. In E. H. Kessler (Ed.), *Encyclopedia of Management Theory*. Thousand Oaks, CA: SAGE. doi:10.4135/9781452276090.n93
- Bielenia-Grajewska, M. (2013). The Heteroglossic Linguistic Identity of Modern Companies. *Management and Business Administration. Central Europe*, 4(123), 120–131.

Bielenia-Grajewska, M. (2014). Corporate Online Social Networks and Company Identity. In K. Michael & R. Abbas (Eds.), *Privacy and Security Issues in Social Network. Encyclopedia of Social Network Analysis and Mining (ESNAM)* (pp.270-275). Springer. doi:10.1007/978-1-4614-6170-8_337

Bielenia-Grajewska, M. (2015). Linguistic Convergence. In: W. Bainbridge (Ed.), *Handbook of Science and Technology Convergence* (pp. 591-605). Springer International Publishing. doi:10.1007/978-3-319-04033-2_84-1

Bielenia-Grajewska, M. (2015). The Role of Gossip in the Creation and Miscreation of Company Identity from the Perspective of Actor-Network Theory. [IJANTTI]. *International Journal of Actor-Network Theory and Technological Innovation*, 7(4), 40–55. doi:10.4018/IJANTTI.2015100104

Cornelissen, J. P. (2008). Metaphor. In R. Thorpe & R. Holt (Eds.), *The SAGE Dictionary of Qualitative Management Research* (pp. 128–129). London: SAGE Publications, Ltd.

Cortés de los Rios, M. E. (2010). Cognitive Devices to Communicate the Economic Crisis: an Analysis through Covers in The Economist. *Ibérica*, 20, 81-106.

Cortés de los Rios, M. E. (2014). La Construcción de Significado Conceptual en la Publicidad Comercial. *Comunicación y Cultura*, 2(1), 23–37.

Forceville, Ch. (1996). *Pictorial Metaphor in Advertising*. London: Routledge. doi:10.4324/9780203272305

Jäkel, O. (1997). *Metaphern in abstrakten Diskurs-Domänen. Eine kognitiv-linguistische Untersuchung anhand der Bereiche Geistestätigkeit, Wirtschaft und Wissenschaft*. Frankfurt: Peter Lang GmbH.

Knowles, M., & Moon, R. (2006). *Introducing Metaphor*. Abingdon: Routledge.

McCloskey, D. (1995). Metaphors economists live by. *Social Research*, 62, 215–237.

Mladenov, I. (2006). *Conceptualizing Metaphors. On Charles Peirce's marginalia*. Abingdon: Routledge.

Tilley, Ch. Y. (1999). *Metaphor and material culture*. Malden, MA: Blackwell Publishers Ltd.

Wittgenstein, L. (1989). *Bemerkungen über die Philosophie der Psychologie*. Frankfurt am Main: Suhrkamp Verlag.

Wittgenstein, L. (1992). *Bemerkungen über die Farben. Über Gewissheit. Zettel. Vermischte Bemerkungen*. Frankfurt am Main: Suhrkamp Verlag.

Zamparini, A., & Lurati, F. (2016). Being different and being the same: Multimodal image projection strategies for a legitimate distinctive identity. *Strategic Organization*, 1-34.

KEY TERMS AND DEFINITIONS

Figurative Language: Language relying on linguistic tools having non-literal meaning, such as metaphors, metonymies, similes, paradoxes, idioms, puns, etc.

Knowledge Management: Strategies and processes connected with managing knowledge at different stages, such as creation, encoding, decoding, application and sharing.

Metaphor: A non-literal linguistic device relying on known domains to denote unknown or less known domains.

Multimodal Metaphor: A non-literal linguistic device using known domains to denote known or less known domains by applying verbal and pictorial notions.

Pictorial Metaphor: A non-literal linguistic device using known domains to denote known or less known domains by applying pictures, drawings and other forms of pictorial representation; also called a visual metaphor.

Verbal Metaphor: A non-literal linguistic device using known domains to denote known or less known domains by applying words or phrases.

Theory and Practice of Online Knowledge Sharing



Will W. K. Ma

Hong Kong Shue Yan University, Hong Kong

INTRODUCTION

Online knowledge sharing refers to the online communication of knowledge so that knowledge is learned and applied by an individual. Several key aspects of the term can be identified. First, it concerns interactions among individuals who communicate with one another. Second, the use of the term “online” signifies a focus on social interaction through online connections and/or online environments. Third, it involves the exchange of knowledge. Individuals come to understand the knowledge they acquire through the process of online communication and social interaction. More importantly, individuals are able to apply such knowledge in future in a similar or different context. Online communication among individuals thus enables knowledge sharing to take place in an online learning environment such that the learner understands the knowledge acquired and is able to apply it in similar or other contexts. Such a definition distinguishes the *online* context from traditional means of communication and differentiates knowledge sharing from purely emotional support.

BACKGROUND

Knowledge sharing represents a critical step in the knowledge-creation process from a variety of knowledge perspectives. Individuals create knowledge and become knowledgeable or de-

velop expertise. The key to learning is to locate knowledgeable others and to learn from them. However, some knowledge cannot be explicitly described and transferred without a certain kind of interaction between experts and learners. In addition, knowledge may not exist in knowledgeable others or experts, but may be sticky knowledge embedded in communities of practice. Knowledge can be shared only through social interaction among individuals who truly understand the practice within a specific context. Furthermore, the social dimension of knowledge has evolved as an integral part of the process of knowledge creation and knowledge acquisition. Various knowledge perspectives emphasize that knowledge is imparted through continuous transactions between individuals, with social interaction at the forefront of such transactions. For example, Schraw (2006, p. 246) argued that learners actively construct meaning rather than simply assimilating it in a passive manner. The social dimension of learning therefore plays a central role in the construction of knowledge. Schraw suggested that most contemporary educators support such a constructivist view in one way or another. Prawat (1996) suggested that individual-social interaction underlies the knowledge-construction process and rejects the notion that the locus of knowledge is in the individual (p.217). Rather, knowledge is the product of a perfect inferential system between the individual and the social environment. The post-positivist perspective on knowledge states that knowledge sharing is a process of consensual

understanding situated in everyday experience, and holds that knowledge is negotiated among those who encounter and use it.

THE CURRENT STATUS OF THE KNOWLEDGE LITERATURE

Nature of Individual Knowledge Sharing and Learning

A popular socio-cultural theory of learning is that of Lev Semyonovich Vygotsky, who explained the mechanism by which knowledge is acquired and represented through knowledge sharing and social interaction (Vygotsky, 1978). This mechanism comprises two planes: the social/individual plane and the public/private plane. Learning starts on the social plane, with learners acquiring new concepts and strategies through interactions with more knowledgeable others. Individual learners then use and extend the concepts and strategies thus acquired to other contexts, and meanings and interpretations are initiated through social interactions (social to individual). Learning then emerges in the public domain, with the knowledge being used by more knowledgeable others and made available to learners. Through interactions within the public domain, individual learners understand, adjust, and implement the knowledge they have learned in the private domain (public to private). Harre (1984) and Wertsch and Bivens (1992) concluded that the success of learning is based on the assumptions that knowledgeable members of a culture will assist others to learn and that learners will actively engage in learning activities to facilitate higher mental functions. Social interactions are initiated among individual learners and naturally result in knowledge sharing.

Organizational Learning and Knowledge Sharing

Ikujiro Nonaka put forward a dynamic theory of organizational knowledge creation that posits social interaction among individuals as the only

means by which tacit-to-tacit and tacit-to-explicit knowledge sharing can take place. At the fundamental level, knowledge is created by individuals (Nonaka, 1994, p. 17). From an organizational point of view, organizations cannot create knowledge without individuals. Organizations provide a context designed to encourage individuals to create knowledge. Through social interaction in informal communities, organizations amplify the knowledge created by individuals, transform such knowledge, and legitimize it through formal notions of a hierarchical structure. Specifically, Nonaka suggested that knowledge appears in two forms: tacit and explicit. Based on these two forms of knowledge, knowledge is created through four modes of knowledge conversion: from tacit to tacit (socialization), explicit to explicit (combination), tacit to explicit (externalization), and explicit to tacit (internalization). Social interaction appears to be the key conversion process by which tacit and explicit knowledge is created. For example, an individual can acquire tacit knowledge through their interactions with individuals. By way of observation, imitation and practice, individuals gain shared experience and share one another's thinking processes. The key to gaining tacit knowledge is shared experience through a process known as socialization. Another means of extracting tacit knowledge from explicit knowledge involves the use of social processes that combine explicit knowledge held by individuals (p.19). Individuals exchange and combine their existing explicit knowledge through meetings and telephone conversations to re-configure, re-categorize, or re-contextualize such knowledge, thus leading to the formation of new knowledge. The theory of knowledge creation thus underlines the importance of knowledge sharing and the key role it plays in creating new knowledge.

Community of Practice and Knowledge Sharing

Etienne Wenger (1998) developed the theory of communities of practices to explain knowledge embedded in social practices. According to this

communities of practice theory, engagement in social practice is the fundamental process by which humans learn. This concept suggests that learning is a process of social participation, a notion further supported by the fact that learner participation is associated with positive effects on learning, satisfaction and retention. Brown and Duguid (2001) argued that too much attention is often paid to the idea of community and too little to the implications of practice (p.198). They suggested that work practice is critical to understanding the acquisition of knowledge at work. Although it is not greatly connected to the community or to the structure of a formal or informal community, knowledge should be viewed through the lens of practice and the perspective of participation. Knowledge developed through participation will be extended to the social sphere if work practices are social. Knowledge is thus viewed less as something mandated by structure or dictated by culture and more as something that participation helps to create (p.202). In the study of the largest online professional community of practice of teachers in Taiwan, Tseng and Kuo (2014) found that closer connections among online community members can lead to greater recognition of and altruism towards others. The development of social relationships among online members helps them obtain potential resources and reliable support through their social network. Online membership fosters a prosocial attitude that increases their willingness to share useful resources and solve other members' problems both emotionally and instrumentally. Knowledge sharing behaviors, in terms of knowledge giving and knowledge receiving, predicted prosocial commitment and performance expectation. Jahnke (2010) explored the dynamics of social roles in a computer-supported social group structure, in particular, called Socio-Technical Community. The results suggested that a group evolves from an informal trust-based community with few formal roles to a formal Socio-Technical Community structure where the social mechanisms,

not the software architecture, support knowledge management process.



Extrinsic and Intrinsic Motivation for Knowledge Sharing

The willingness to share knowledge is regarded as a prosocial form of behavior, both among individuals in interpersonal relationships and among employees engaging in organizational behavior. While this type of prosocial behavior is practiced on a voluntary basis rather than being forced, the key challenge in knowledge sharing is how to motivate people to share knowledge: why should I share my knowledge with you? Prior studies utilize a number of theories to explain the willingness to share knowledge online, for example, community of practice (e.g., Tseng & Kuo, 2014); uses and gratifications model (e.g., Lee & Jang, 2010); theory of reasoned action (e.g., Arpaci & Baloglu, 2016); theory of belonging (e.g., Ma & Yuen, 2010); intrinsic motivation of altruism (e.g., Diep et al., 2016; Ma & Chan, 2014); self-determination theory (e.g., Pe-Than et al., 2014); sense of community (e.g., Yilmaz, 2016), etc. They examine how both extrinsic and intrinsic motivations affect the intention to share online knowledge and/or actual knowledge-sharing behavior. To cite some of the findings, for example, Diep, Cocquyt, Zhu, and Vanwing (2016) examined altruism and perceived learning benefits as a measure of performance expectancy where social capital with regard to adult learners' online participation. Sense of belonging and norms of reciprocity are found to significantly related to online participation. Similarly, Ma and Chan (2014) draws on the theory of belonging and the intrinsic motivation of altruism to explore the factors contributing to social media knowledge sharing behavior. They found that perceived online attachment motivation and perceived online relationship commitment, and altruism are significantly related to social media knowledge sharing. Furthermore, Yoo and Gretzel (2011) examined the effect of personality

traits on content creation in comparison to only information consumption. Results showed that personality has been found to be a particularly influential trait significantly related to perceived barriers to content creation, motivations to engage in content creation and specific creation behaviors. Another study by Yilmaz (2016) explored and found that academic self-efficacy and sense of community are related to knowledge sharing behavior. In particular, connectedness to the community, learning perception in the community, the self-efficacy of the students on the cognitive applications in the courses and their social status in the community are related to knowledge sharing behavior. In the study of perceived enjoyment for information sharing, Pe-Than, Goh, and Lee (2014) found that perceived needs for autonomy, competence, and relatedness are related to perceived enjoyment. In another study among information technology majoring undergraduates in Turkey, Arpaci and Baloglu (2016) found that cultural collectivism has a positive and significant impact on attitudes toward and subjective norms with regard to knowledge sharing.

An Interpersonal Perspective for Knowledge Sharing

A review of prior studies identifies an important motivation that triggers frequent social interactions, which form the basis of knowledge sharing. Baumeister and Leary (1995) suggested that the need to belong is a fundamental human motivation for frequent and regular social interaction. They defined the need to belong as “a need to form and maintain at least a minimum quantity of interpersonal relationships, [which] is innately prepared (and hence nearly universal) among human beings” (p.499). According to this theory, people are naturally driven toward establishing and sustaining belongingness. Thus, the theory of the need to belong may explain the motivation for social interaction in knowledge sharing through the mechanisms of affiliation motivation (to form

social bonds) and relationship commitment (to maintain those bonds). To satisfy this desire, individuals will try their best to form and maintain relationships through frequent social contact and interaction with others. This provides a basis for explaining the motivation for knowledge sharing. The need to belong stimulates goal-directed activities designed to satisfy it. Motivated by the social goal of satisfying the need to belong, people tend to seek out interpersonal contacts and cultivate possible relationships, and continue doing so until they have reached a minimum level of social contact and relatedness. A recent empirical study found that online knowledge sharing has significant direct relationships with online perceived attachment motivation (the motivation to form a relationship) and online perceived relationship commitment (the motivation to maintain a relationship). Following this line of thought, Ma and Yuen (2010) explored and found significant relationship between interpersonal relationship motivations and online knowledge sharing.

RESEARCH ON ONLINE KNOWLEDGE SHARING PRACTICES

It is commonly argued that the use of information technology improves knowledge sharing through learner participation and interaction in both traditional and online learning. A review of recent empirical studies shows there are several streams of research into online learning that deal with knowledge sharing. The first stream considers online learning provided through a shared platform on which peer learners interact, often in the form of discussion forums, and in which knowledge sharing occurs through the continuous interaction of peer learners engaging in asynchronous written communication (e.g., Lee & Jang, 2010). The second stream examines online learning in shared workplaces that allow peer learners to interact to complete a common task, in which knowledge sharing occurs through continuous interaction

among peer learners who learn by doing (e.g., Cevik, 2015). Another stream holds that online learning provides a transparent demonstration of individual outcomes, and that knowledge sharing occurs through continuous exposure to best practices and learning by observation among peer learners (e.g., Feng & Ye, 2016). A final strand of the literature highlights that online learning provides a centralized meeting place for community building and that knowledge sharing occurs naturally in the presence of human resources and expertise (e.g., Jahnke, 2010).

Many empirical studies of online knowledge sharing measure such sharing in terms of social interaction, participation and engagement, and its relationship to learning. For example, online discussions have been examined by assessing messages posted by learners (e.g., Mazzolini & Madison, 2007); collaborative learning environments have been investigated by assessing interactional activity, participation patterns, and their effects on performance (e.g., Lin & Lai, 2013); and collaboration environments supported by different levels of technology have been assessed by investigating the extent to which learners share knowledge and the resulting effect on individual outcomes (e.g., Eid & Al-Jabri, 2016). Knowledge-building communities have been studied by investigating the sharing of learners' knowledge with other members of the community (Pi, Chou, & Liao, 2013). All of the empirical evidence in these studies indicates that fully utilizing an online learning environment can improve online knowledge sharing.

The examples given below illustrate the best practices using technology to support knowledge sharing in a learning environment.

Asynchronous Virtual Classrooms

Hiltz and Wellman (1997) established one of the first asynchronous virtual classrooms (the NJIT Virtual Classroom) to empirically test knowledge sharing through virtual community building and asynchronous communication. With the help of

the virtual classroom setting, such students can interact online with their instructors and other peer learners at any time and in any place (p.49). This virtual classroom project generated a number of subsequent studies. One such study (Wellman et al., 1996) found that the majority (71%) of students stated they had better access to their professors, and 73% found the virtual classroom "more convenient" overall. Collaborative learning took place online and learning communities were found to exist.

Knowledge Forums

Marlene Scardamalia and Carl Bereiter developed the CSILE (Computer Supported International Learning Environments) system at the Ontario Institute for Studies in Education (Singh & Means, August, 1995). Established in 1986, it is considered the first networked system designed for collaborative learning and has been used as a research program within Toronto schools for a number of years. The CSILE system functions as collaborative learning environment that helps form a learning community using the same database capable of displaying both text and graphics. The dialogue that ensues eventually leads to the accumulation of knowledge. In 1995, Knowledge Forum was developed under the World Wide Web environment to replace CSILE and facilitate its ongoing use to help and support knowledge-building communities. In a recent study, Lee and Jang (2010) explored the motivations to knowledge sharing in online knowledge forums. They found that stronger affiliative tendency, higher self-esteem, or stronger public individuation are more likely to contribute to the open information repository. On the other hand, they also found that these psychological traits significantly enhance individual's intention to share knowledge on a public web site only when other users' presence is rendered salient and individual contributions are visibly acknowledged.

Social Networking Sites

The recent development of social networking sites brought new opportunities for knowledge sharing and learning. Social networking is a basic human need to associate, communicate, and collaborate with one another, and to share ideas, feelings, observations, and thoughts with others. With the help of social networking platforms developed in recent years, social networks form online and allow us to interact, share, educate, and connect to extend our physical interpersonal relationships to build trust and to form relationships in the virtual world (Safko, 2013, pp. 46-47). Recent studies examined the process of knowledge sharing in social networking and explained the process with social networking theory and analysis. For example, Eid and Al-Jabri (2016) examined the various categories of social networking sites use including chatting and online discussion, creating knowledge and information content, file sharing, and enjoyment and entertainment by undergraduate students. They found that there are significant positive relationships between both chatting and online discussion and file sharing and knowledge sharing, and entertainment and enjoyment with student learning. Lin and Lai (2013) utilized social network analysis to find that social network awareness increases participant rate and opportunities of knowledge sharing and helps promote learning achievement. In particular, students with higher centrality (regardless of degree and closeness) are more likely to take advantage of the social network position to ask for help; easily become target helpers that peers seek to; utilize the social network more frequently; and have better learning achievement. Another study (Pi et al., 2013) found that reputation affects knowledge sharing attitude; sense of self-worth affects attitude through subjective norm; social networking sharing culture (fairness, identification, and openness) is related to knowledge sharing intention through attitude and subjective norm.

MEASUREMENT OF ONLINE KNOWLEDGE SHARING

Ko et al. (2005) argued that although many scholars have conceptualized knowledge sharing or knowledge transfer, relatively few have attempted to measure it directly (p.68). Based on the study of Argote and Ingram (2000), Ko et al. (2005) define knowledge sharing as “the communication of knowledge from a source so that it is learned and applied by a recipient” (p.60). They developed and empirically tested a six-item instrument in which three items address the learning component and the other three items assess the application component.

Based on the investigations of Zander and Kogurt (1995) and Ko et al. (2005), Brown et al. (2006) examined the perceived value of the knowledge-sharing process to determine whether and how much knowledge sharing occurred. They argued that perceptions are more useful in measuring across-domain knowledge sharing, whereas specific tests of knowledge content are narrow in their application because of their context-specific nature (p.2). Based on person-to-person knowledge sharing, this study put forward and empirically tested a ten-item scale to measure the value of knowledge sharing composed of four components: efficiency, quality, learning, and understanding. The study generated four key findings. First, a knowledge-sharing process helps employees save time in performing their jobs; second, knowledge sharing improves the quality of work performed; third, for knowledge sharing to be truly valuable, it must enhance learning; fourth, for knowledge sharing to be truly beneficial, it must facilitate sharing of the deep structures necessary to apply that knowledge (i.e., understanding). More recently, Ma and Yuen (2011a, 2011b) developed and validated an online knowledge-sharing scale, using two samples from two different online learning environments. This is a five-item instrument in which two items measure the learning component and three items assess the application component.

CONCLUSION

Although the issue of what factors motivate knowledge sharing remains an open question, Facebook usage statistics clearly show that sharing is a fundamentally natural process. For instance, Facebook statistics shows that the site has 1.13 billion daily active users on average; photo uploads total 300 million per day; 4.5 billion likes generated daily; 4.75 billion pieces of content shared daily (Zephoria.com, 2016). The popularity of this social media site demonstrates that people frequently engage in social interaction and sharing. What we are less certain of and which therefore requires further exploration, is what inhibits us from sharing, especially in an online knowledge-sharing context. For example, although it is known that learning conceptions and learners' self-concepts affect learners' approaches to learning and learning outcomes (e.g., Burnett, Pillay, & Dart, 2003), do different learning conceptions affect interaction and knowledge sharing among peer learners? Studies of interpersonal relationships have shown that differences between individuals affect how people approach others. Nevertheless, although technology enhances our ability to communicate, does it enhance online knowledge sharing or inhibit it instead? For example, Waycott, Sheard, Thompson, and Clerehan (2013) identified a number of concerns while sharing students' work including students' fear of copying, poor online conduct, and the risk that students may feel exposed when publishing their work online. Thus, sharing process creates tension between the collaborative and participatory nature of the social networking, and the competitive and individual nature of the undergraduate assessment in formal education.

FUTURE RESEARCH DIRECTIONS

Do we understand the above concerns well enough? This would also be an important topic for further exploration.

REFERENCES

- Argote, L., & Ingram, P. (2000). Knowledge transfer: A basis for competitive advantage in firms. *Organizational Behavior and Human Decision Processes*, 82(1), 150–169. doi:10.1006/obhd.2000.2893
- Arpaci, I., & Baloglu, M. (2016). The impact of cultural collectivism on knowledge sharing among information technology majoring undergraduates. *Computers in Human Behavior*, 56, 65–71. doi:10.1016/j.chb.2015.11.031
- Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, 117(3), 497–529. doi:10.1037/0033-2909.117.3.497 PMID:7777651
- Brown, J. S., & Duguid, P. (2001). Knowledge and Organization: A Social-Practice Perspective. *Organization Science*, 12(2), 198–213. doi:10.1287/orsc.12.2.198.10116
- Brown, S. A., Dennis, A. R., & Gant, D. B. (2006). Understanding the factors influencing the value of person-to-person knowledge sharing. *Proceedings of the 39th Hawaii International Conference on System Sciences*. IEEE Computer Society. doi:10.1109/HICSS.2006.516
- Burnett, P. C., Pillay, H., & Dart, B. C. (2003). The Influences of Conceptions of Learning and Learner Self-concept on High School Students Approaches to Learning. *School Psychology International*, 24(1), 54–66. doi:10.1177/0143034303024001621
- Cevik, Y. D. (2015). Assessor or assessee? Investigating the differential effects of online peer assessment roles in the development of students problem-solving skills. *Computers in Human Behavior*, 52, 250–258. doi:10.1016/j.chb.2015.05.056

- Diep, N. A., Cocquyt, C., Zhu, C., & Vanwing, T. (2016). Predicting adult learners online participation: Effects of altruism, performance expectancy, and social capital. *Computers & Education*, 101, 84–101. doi:10.1016/j.compedu.2016.06.002
- Eid, M. I. M., & Ai-Jabri, I. M. (2016). Social networking, knowledge sharing, and student learning: The case of university students. *Computers & Education*, 99, 14–27. doi:10.1016/j.compedu.2016.04.007
- Feng, Y., & Ye, H. (2016). Why do you return the favor in online knowledge communities? A study of the motivations of reciprocity. *Computers in Human Behavior*, 63, 342–349. doi:10.1016/j.chb.2016.05.007
- Harre, R. (1984). *Personal being: A theory for individual psychology*. Cambridge, MA: Harvard University Press.
- Hiltz, S. R., & Wellman, B. (1997). Asynchronous Learning Networks as a Virtual Classroom. *Communications of the ACM*, 40(9), 44–49. doi:10.1145/260750.260764
- Jahnke, I. (2010). Dynamics of social roles in a knowledge management community. *Computers in Human Behavior*, 26(4), 533–546. doi:10.1016/j.chb.2009.08.010
- Ko, D.-G., Kirsch, L. J., & King, W. R. (2005). Antecedents of knowledge transfer from consultants to clients in enterprise system implementations. *Management Information Systems Quarterly*, 29(1), 59–85.
- Lee, E.-J., & Jang, J.-W. (2010). Profiling good Samaritans in online knowledge forums: Effects of affiliative tendency, self-esteem, and public individuation on knowledge sharing. *Computers in Human Behavior*, 26(6), 1336–1344. doi:10.1016/j.chb.2010.04.007
- Lin, J.-W., & Lai, Y.-C. (2013). Online formative assessments with social network awareness. *Computers & Education*, 66, 40–53. doi:10.1016/j.compedu.2013.02.008
- Ma, W. W. K., & Chan, A. (2014). Knowledge sharing and social media: Altruism, perceived online attachment motivation, and perceived online relationship commitment. *Computers in Human Behavior*, 39, 51–58. doi:10.1016/j.chb.2014.06.015
- Ma, W. W. K., & Yuen, A. H. K. (2010). Understanding online knowledge sharing: An exploratory theoretical framework. In P. Tsang, S. K. S. Cheung, V. S. K. Lee & R. Huang (Eds.), *Hybrid Learning: Third International Conference, ICHL2010, LNCS 6248* (pp. 239–248). Berlin: Springer-Verlag.
- Ma, W. W. K., & Yuen, A. H. K. (2011a). Gender Differences of Knowledge Sharing in Online Learning Environment. In R. Kwan et al. (Eds.), *ICHL2011, LNCS 6837* (pp. 116–128). Springer-Verlag.
- Ma, W. W. K., & Yuen, A. H. K. (2011b). Understanding Online Knowledge Sharing: An Interpersonal Relationship Perspective. *Computers & Education*, 56(1), 210–219. doi:10.1007/978-3-642-22763-9_11
- Mazzolini, M., & Maddison, S. (2007). When to jump in: The role of the instructor in online discussion forums. *Computers & Education*, 49(2), 193–213. doi:10.1016/j.compedu.2005.06.011
- Nonaka, I. (1994). A Dynamic Theory of Organizational Knowledge Creation. *Organization Science*, 5(1), 14–37. doi:10.1287/orsc.5.1.14
- Pe-Than, E. P. P., Goh, D. H. L., & Lee, C. S. (2014). Making work fun: Investigating antecedents of perceived enjoyment in human computation games for information sharing. *Computers in Human Behavior*, 39, 88–99. doi:10.1016/j.chb.2014.06.023
- Pi, S.-M., Chou, C.-H., & Liao, H.-L. (2013). A study of Facebook groups members knowledge sharing. *Computers in Human Behavior*, 29(5), 1971–1979. doi:10.1016/j.chb.2013.04.019

- Prawat, R. S. (1996). Constructivisms, modern and postmodern. *Educational Psychologist*, 31(3/4), 215–225. doi:10.1080/00461520.1996.9653268
- Safko, L. (2013). *The fusion marketing bible*. New York: The McGraw Hill.
- Schraw, G. (2006). Knowledge: Structures and processes. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of educational psychology* (2nd ed.; pp. 245–263). Mahwah, N.J.: Lawrence Erlbaum.
- Singh, R., & Means, B. (1995). *Computer-Supported Intentional Learning Environments (CSILE). Education and Technology Reform*. Retrieved from <http://www2.ed.gov/pubs/EdReformStudies/EdTech/csile.html>
- Tseng, F.-C., & Kuo, F.-Y. (2014). A study of social participation and knowledge sharing in the teachers online professional community of practice. *Computers & Education*, 72, 37–47. doi:10.1016/j.compedu.2013.10.005
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Waycott, J., Sheard, J., Thompson, C., & Clerehan, R. (2013). Making students work visible on the social web: A blessing or a curse? *Computers & Education*, 68, 86–95. doi:10.1016/j.compedu.2013.04.026
- Wellman, B., Salaf, J., Dimitrova, D., Garton, L., Gulia, M., & Haythornthwaite, C. (1996). Computer networks as social networks: Collaborative work, telework, and virtual community. *Annual Review of Sociology*, 22(1), 213–238. doi:10.1146/annurev.soc.22.1.213
- Wenger, E. (1998). *Communities of Practice: Learning, Meaning, and Identity*. New York: Cambridge University Press. doi:10.1017/CBO9780511803932
- Wertsch, J. V., & Bivens, J. (1992). The social origins of individual mental functioning: Alternatives and perspectives. *Quarterly Newsletter of the Laboratory of Comparative Human Cognition*, 14(2), 35–44.
- Yilmaz, R. (2016, October). 92016). Knowledge sharing behaviors in e-learning community: Exploring the role of academic self-efficacy and sense of community. *Computers in Human Behavior*, 63, 373–382. doi:10.1016/j.chb.2016.05.055
- Yoo, K.-H., & Gretzel, U. (2011). Influence of personality on travel-related consumer-generated media creation. *Computers in Human Behavior*, 27(2), 609–621. doi:10.1016/j.chb.2010.05.002
- Zander, U., & Kogut, B. (1995). Knowledge and the speed of the transfer and imitation of organizational capabilities: An empirical test. *Organization Science*, 6(1), 76–92. doi:10.1287/orsc.6.1.76
- Zephoria.com. (2016). *The top 20 valuable Facebook statistics*. Retrieved from <https://zephoria.com/top-15-valuable-facebook-statistics/>
- Zhang, J., Scardamalia, M., Lamon, M., Messina, R., & Reeve, R. (2007). Socio-cognitive dynamics of knowledge building in the work of 9- and 10-years-olds. *Educational Technology Research and Development*, 55(2), 117–145. doi:10.1007/s11423-006-9019-0

ADDITIONAL READING

- Bereiter, C. (2002). *Education and Mind in the Knowledge Age*. Hillsdale, NJ: Lawrence Erlbaum.
- Ghadirian, H., Ayub, A. F. M., Silong, A. D., Bakar, K. B. A., & Zadeh, A. M. H. (2014). Knowledge sharing behavior among students in learning environments: A review of literature. *Asian Social Science*, 10(4), 38. doi:10.5539/ass.v10n4p38

Grover, V., & Davenport, T. H. (2001). General Perspectives on Knowledge Management: Fostering a Research Agenda. *Journal of Management Information Systems*, 18(1), 5–21.

Hiltz, S. R. (1994). *The Virtual Classroom: Learning without Limits via Computer Networks*. Norwood, NJ: Ablex Publication.

Nonaka, I., & Takeuchi, H. (1995). *The Knowledge-Creating Company: How Japanese Companies Create the Dynamic of Innovation*. New York: Oxford University Press.

Polanyi, M. (1966). *The Tacit Dimension*. London: Routledge & Kegan Paul.

KEY TERMS AND DEFINITIONS

Asynchronous Communication: A mediated form of communication in which the sender and the receiver are not concurrently engaged in communication.

Computer Supported Collaborative Learning: A kind of learning through sharing and construction of knowledge that takes place via social interaction using a computer or through the Internet.

Explicit Knowledge: The kind of knowledge which has been or can be articulated, codified, and stored in certain media that can be readily transmitted to others.

Extrinsic Motivation: The kind of motivation which comes from outside of the individual.

Intrinsic Motivation: The kind of motivation which is driven by an interest or enjoyment in the task itself, and exists within the individual rather than relying on any external pressure.

Online Knowledge Sharing: online communication of knowledge so that knowledge is learned and applied by an individual (Ma & Yuen, 2011b).

Social Networking Site: A social networking service (also social networking site, SNS or social media) is an online platform that is used by people to build social networks or social relations with other people who share similar personal or career interests, activities, backgrounds or real-life connections.

Tacit Knowledge: The kind of knowledge which is difficult to transfer to another person by means of writing it down or verbalizing it.

Virtual Communities: A social network of individuals who interact through specific media, potentially across geographical and political boundaries in order to pursue mutual interests or goal.

Visualization as a Knowledge Transfer



Anna Ursyn

University of Northern Colorado, USA

INTRODUCTION

There is growing significance of both visual literacy and knowledge of visualization, where visualization means communication of data, information, and knowledge with graphical representations. Knowledge visualization has become a cross-disciplinary, interactive culture and the element of the utmost importance in science education because of a need to convey the information to students about advances in technologies. Possibly, visualization is the best way of learning, teaching, or sharing the data, information, and knowledge because it amplifies cognition, outperforms text-based sources, and increases our ability to think and communicate. For all these reasons visualization ability should be introduced and trained since kindergarten.

This text presents selected concepts, methods, and tools related to visualization of data, information, and knowledge. It presents some approaches to the concept of visualization and the ways it mediates between the user and the physical world. It overviews visualization tools and applications, and discusses the importance of visualization methods for the current educational strategies.

BACKGROUND

Information is usually presented in numerical, graphic, or diagrammatic form; it may be shown as a sketch, drawing, diagram, plan, outline, image, geometric relationship, map, music and dance notation, object, interactive installation, or a story. Diagrams visualize information in a pictorial yet abstract (rather than illustrative) way: as plots, line-graphs and charts, or the engineers or archi-

itects' blueprints. Complicated presentations of data organization and interpretation, for example governmental statistics are easier to comprehend in a graphic than in a numerical form, when they serve as explanatory tools for the data sets. Thus, visualizations change numerical data into graphs, clouds (Chen, 2010), tree visualizations (Shneiderman, 2014; Lima, 2014), network data, time-based, interactive, metaphorical visualization designs, and other formats.

Visualization means the communication of information with graphical representations. At the present time, visualization means using the computer, which transforms data into information, and then visualization converts information into picture forms. Graphic images and symbols convey and express the meaning of abstract data, which lets us comprehend data and make discoveries, decisions, or explanations about patterns or individual items (Shneiderman, 1996). Thus communication through visualization is at the same time pictorial and linguistic. It is socially and culturally conditioned, based on familiar linguistic patterns, as in a 'pie chart' metaphor for market shares, or a 'starry night' metaphor showing data in 3D (Bertschi & Bubenhofer, 2005).

Cognitive way of learning and teaching may involve cooperation with specialists in several disciplines. Computer scientists and artists apply visual way of presentation while working, for example with mathematicians, anthropologists, designers, and architects to conduct computer analysis of facades and architectural details. Professionals performing scientific presentations and researchers in fields of natural sciences, medicine, pharmacology, biology, geology, or chemistry examine and visualize symmetry and patterns in natural and human-made structures. Many artists

DOI: 10.4018/978-1-5225-2255-3.ch443

have created masterpieces this way. Visualization serves as an efficient tool that assists practitioners creating communication media-art, installations, animated video or film, architectural projects, designing newspapers and magazines, or working on website design. Users apply visualizations to understand how data analyses and queries relate to each other. From simple charts and data graphics to 3D multi-user virtual reality environments happening in real time with human interaction possible, visualizations let us fly around the organized data, comprehend, and make decisions (Chen, 2010, 2011). Structural modeling of the relationships may involve the use of graphs, trees, or cones; detecting proximity and connectivity; clustering and classification using word search; multi-dimensional-scaling; network analysis; glyphs (single graphical units portraying many variables by adapting their properties) on charts and graphs; virtual structures; applying complex network theory, and network representations (Chen, 2010).

Visualization enhances communication through information display with the use of letters, numerals, art, graphic design, visual storytelling, signs, symbols, and application software. Drawing basic shapes like squares, triangles, and circles connected by lines and arrows, and then inserting simple drawings inside of these shapes creates visualization of our concepts. Graphs, diagrams, or animations can visualize messages as well. Examples of the non-visual creations are multi-modal interactive data presentations, sonifications, and haptic/touch interfaces, for example pressure sensitive interfaces. Visualization has also been considered a semiotic process because of the use of signs to present ideas.

There is a wide range of visualization techniques, still growing along with the developments in computing and information technology. Visualizations help explore and understand complex data, communicate, and navigate on the web. Visualization industry uses software and programming solutions, along with the great amount of the cultural, historical, and architectural research.

Examples may include a study on the Napoleon's campaign (Tufte, 1983,1992) or a Periodic Table of Visualization Methods (Lengler, & Eppler, 2008). Time-based, 3D, and augmented reality applications serve for the military, intelligence, aviation and air traffic control, medical education, and other purposes.

Metaphorical Language of Visualization

Visual metaphors make a basic structure in visualization because they describe relations among data, organize information in a meaningful way, and combine creative imagery with the analytic rationality of conceptual diagrams. A metaphor indicates one thing as representing another, difficult one, thus making mental models and comparisons. Instead of developing a nomenclature specific for computing, we apply metaphors – names of familiar items and actions for organizing computing-related items and activities: we open a new window or a file with a mouse, put them in a folder, we cut, copy, and paste, place icons on a desktop, use tools and search engines, canvas, mailbox, documents, in-and-out boxes, and a web portal. The desktop metaphor is now fading because cell phones and tablets are replacing PCs as the main gateway to the Internet. We may use graphics or show virtual environments, often shaped by artist's fantasy; the success and quality of any visualization depend on imagination, the retrieval of necessary data, the choice of a metaphor, and the delivery method: whether to apply animation, interconnection, or interaction. Visual metaphors can either be natural objects or phenomena (e. g., mountains, tornados) or artificial, man-made objects (e. g., a bridge, a temple), activities (climbing), or concepts (war, family). Metaphors organize, structure information, and convey an insight through characteristics or associations. Metaphors may convey topics such as theological events or encyclopedias' tables of contents, and serve as classification systems (Lima, 2011).

Communication involves language, and pictures need a caption. Because language is metaphorical, our thoughts are mostly metaphorical. To communicate knowledge in a dynamic, interactive way, often in real-time, visualizations provide two-dimensional and tree-dimensional metaphors, familiar and understandable in social and cultural terms. Metaphors traditionally used in visualizations show programs or data as natural objects, such as a solar system, or man-made objects, for example a city, architecture, house, parking lot, metro, library, street, but also they map the data to facial expressions, video games, or nested boxes. A city metaphor may represent a software package; navigating through software may enhance program understanding or lessen costs and effort of software production and maintenance.

VISUALIZATION DOMAINS

The most important domains in visualization are: data visualization, information visualization (when somebody presents what has been done with the data), and knowledge visualization. New approaches to visual way of processing change with the growth of communication media, social networking, cloud computing, and GPS.

Data Visualization

Data visualization presents the sets of data in a visual form, and thus enables us to go from the abstract numbers in a computer program (ones and zeros) to visual interpretation of data. Sabol (in Bertschi, Bresciani, Crawford, Goebel, Kienreich, Lindner, Sabol, & Moere, 2011, p. 333) describes data as “sequences of numbers or characters representing qualitative or quantitative attributes of specific variables. Many types of physical or numerical data may include 1D linear, 2D (e.g., a map), 3D (e.g., molecule or architectural models), multidimensional (scattergrams, clusters), and other types. Data visualization applies tools that have been listed as techniques for spatial

and geospatial data, imaging multivariate data, visualization of trees, graphs, and networks, text representations, and interaction techniques (Ward, Grinstein, & Keim, 2010).

Information Visualization

Information Visualization means the use of computer-supported, “interactive visual representations of abstract data to amplify cognition” (Bederson & Shneiderman, 2003; Card, Mackinlay, & Shneiderman, 1999). Ben Shneiderman included theoretical approaches and discussion about visualization as a scientific discipline. He proposed the Visual Information Seeking Mantra and also the task-by-data-type taxonomy. Mantra serves as a methodological guidance to practitioners who design novel systems. According to the classic Shneiderman’s information visualization seeking mantra (1996), visualization should support seven high level tasks: overview, zoom, filter, detail-on-demand, relate, history, and extract. Craft and Cairns (2005, 2008) elaborated this notion in prescriptive way. It has been also described as representation plus interaction, which often derives new insights. Creating visualizations includes transforming raw data (that may be abstract, semantic, or verbal information in hypertext, www, or other text documents) into structured data as data tables, converting data for calculations of their attributes, and visual mapping of important structures into the abstract visual structures. Users transform these structures on a screen changing their shape, color, size, or location. Designing information visualization has been described by Shneiderman as importing and cleaning data, combining visual representations with textual labels, finding related information, viewing large volumes of data, integrating data mining, integrating with analytical reasoning techniques, collaborating with others, achieving universal usability, and evaluation (Fowler, 2012).

Information visualization may have a great variety of applications. It may serve scientific theories and fields including applications of graph

theory, geometric modeling and imaging, visual analytics, virtual environments, geo-analytics, biomedical informatics and visualization, web visualization, cultural heritage knowledge visualization, aesthetics, visualization in software engineering, architecture, visualization in built and rural environments, and many other fields. It serves for conveying information in online journalism, business management, technical writing, social networks, and education. Software architecture can be understood easier when a visual query helps us to find patterns of highly connected components in a node-link diagram. Visualization techniques are used in conjunction with medical treatment including cancer treatment.

Network and web-search result visualization became the main carrier of information. Visual search engines use information visualization, data mining, and semantic web. Navigation systems make a website opening fast, visually appealing, containing all information needed, and easy to find. Web architectures use search engines, manage large databases, and create web interfaces based on the concept of the semantic web. Surfing the web involves visualizing and manipulating data in multiple dimensions, using Java, 2D and 3D interaction metaphors, and data mining.

Knowledge Visualization

Knowledge Visualization uses visual representation to transfer insights to create, integrate, and apply knowledge, rather than data, between individuals; it focuses on the recipients, other types of knowledge, and on the process of communicating different visual formats (Burkhard, Meier, Smis, Allemang, & Honish, 2005). It has been widely accepted that knowledge visualization provides new knowledge in the process of communication between at least two people (Card, Mackinlay, & Shneiderman, 1999). Techniques are aimed at explanation and presentation of knowledge, so they enhance cognitive processes of users and reduce cognitive load for working memory.

Knowledge visualization can be created without the use of a computer. A cave drawing, a map, or a mind map drawn with a pen on paper may serve as examples of knowledge visualization generated without a computer. Knowledge visualization has been present in different disciplines and in various modes since early days of civilization. Tables of the past acted as visualization modalities (Marchese, 2011). Many kinds of tables preserved since ancient times; the Near East Akkadian clay tablets, Sumerian accounting tables, Aztec calendars, or the Egyptian stele Rosetta Stone, as well as the medieval chronicles, canon tables, and calendars are representations of early genres in information visualization. Analysis of these tables demonstrates the constant need to visualize abstract data, information, and knowledge.

Knowledge visualization specialists use computer-based (and also non-computer-based) graphic representations, such as information graphics (infographics), objects, sketches, conceptual diagrams, images, concept maps, animations, interactive visualizations, storyboards, and visual metaphors to produce solutions concerning readability, simplification, and effectiveness of presentation for a wide spectrum of users. Designers co-work with communication science specialists for social network users (such as cell phone users, e-mail archives, criminal networks, or underage audience sensitive messages). Knowledge visualization could contribute design- and user-specific representations, e.g., a map, metro, aquarium, solar system, or flower metaphor for users with limited visual literacy (Kienreich, in Bertschi et al., 2011).

Scientific Visualization

Scientific Visualization was established in 1985 at the National Science Foundation panel. McCormick, DeFanti, and Brown (1987) defined scientific visualization as a field in computer science that encompasses user interface, data representation, processing algorithms, visual

representations, and other sensory presentation such as sound or touch (McCormick et al., 1987). Abstract or model-based scientific visualizations present real objects in a digital way. Scientific visualization (mapping from the computed to perceptual representations) focuses on maximizing human understanding, usability, effectiveness, and efficiency. Physically based data are defined, selected, transformed, and represented directly from the data according to space coordinates, such as geographic data or computer tomography data of a body for medical use. Visualization as storytelling comprises narratives, interactive graphics, explanatory and animated graphics, and multimedia. Analysts, decision makers, engineers, or emergency-response teams depend on the ability to analyze information contained in the data; they also search how users navigate the database.

Concept Mapping

Concept Mapping represents the structure of information visually and builds knowledge models useful for strategic planning, product development, market analysis, decision-making, and measurement development. Concept maps support the learning, construct knowledge, reduce cognitive load, and improve recall of information. Knowledge maps visualize knowledge by showing changes and interrelationships; they help us design strategies and build assessment. Network and web-search result visualization became the important carrier of information. Visual search engines use information visualization, data mining, and semantic web.

Visual Analytics

Visual Analytics provides automated analysis of large amounts of dynamic information and the closed loop approach with visualization and interaction utilizing multi-touch surfaces; combines computational and visual methods using visualization, data mining, and statistics; uses abstract

visual metaphors, mathematical deduction, and human intuitive interaction. Visual analytics serves for studying the entire genome of an organism at many abstraction levels: cells, organisms, and ecosystems, in formats and scales such as molecules, gene networks, and signaling networks. It also serves medicine, environmental research, and national security. Tag cloud visualization is a text-based visual representation of a set of tags.

Information Aesthetics

Information Aesthetics forms a cross-disciplinary link between information visualization and visualization art. Information visualization is often evaluated on representing data in pleasurable and intelligible way. Visualization not only makes the unseen visible, it is building a meaningful net of associations and connotations. The developments in computing technology may cause that aesthetics of knowledge visualization may be measured as the balance between art literacy and technological literacy: visual competence in the art, design, and technological solutions in visualization. Aesthetic values are considered important in mathematics, science, and computing, including the aesthetic computing and aesthetic issues related to digital environment.

In the field of product design, product visualization supports market research and advertising; a grasp of visualization technologies is a requisite for designing effective visual marketing. Critics and audience expect the inclusion of functionality, effectiveness, and efficiency into evaluation of the product. This needs consideration whether an object's design fulfills the functional, ergonomic, aesthetic, material- and space-related demands, along with the product's comfort, simplicity, elaboration, message, or easiness of use. Visual literacy supports the quality design and product aesthetics. For Edward Tufte (1983/2001, 1992/2005), good design has two key elements found in simplicity of design and complexity of data. Frøkjær, Hertzum, and Hornbæk (2000) connected usability with ef-

fectiveness – the accuracy and completeness with which users achieve certain goals; efficiency – the relation between effectiveness and the resources expanded in achieving them; and satisfaction – the user’s comfort with and positive attitudes towards the use of the system.

Product visualization can extract its high-quality features and present them as quite distinct images in a way that everybody can apprehend the product’s values and buy the marketed article. The objectives of aesthetic studies also move toward the effectiveness, efficiency, and easiness to understand (a low cognitive cost) of visual presentation, not exclusively to the beauty of an image. Informative art works are not only aesthetical objects but also contain information. Researchers usually associate aesthetics with readability, and readability with understanding.

Tools for visualization may serve as in terms of cognitive instruments such as utilization of a metaphor, simulations, or the use of layers, details, and complexity in a website. Tools and instruments for visualization are derived from many domains; they may include scanners, microscopes, and cameras, along with many kinds of applications for recording and measuring in real time, micro agents, or bots. In terms of computing one may apply computer graphics tools such as Photoshop, and also the computer networking related concepts such as cloud computing. Clustering technique (clusters are subsets of observations) is used in data mining for statistical analysis, pattern recognition, and bioinformatics.

Infographics refers to tools and techniques involved in graphical representation of data, information, and knowledge, mostly in journalism, art, and storytelling. Models of information management address the data presentation, mapping, and issues related to the temporal dimension. A model may mean a scaled copy of some object (often using interactive modeling systems), or a mathematical model describing physical laws and the behavior of physical objects.

CURRENT APPROACHES AND APPLICATIONS

Visualization techniques may support the present-day trends in science education strategies. They allow presenting the mathematical theories and their proofs, along with the computing science related theories as two-dimensional and three-dimensional constructs. For educational the purpose, placing a great emphasis on developing visual literacy in students may be beneficial in fulfilling this task. For example, students may create mathematically programmed sculptures, or the fractal based art works, sceneries, and backdrops. Some versed in programming would be coding music- and dance-related shapes and form, thus creating music visualization.

Current means of delivering knowledge include for example, the use of 3D printing (Mercuri & Meredith, 2014) or augmented reality (Bredl, Groß, Hünninger, & Fleischer, 2012). 3D printing technology and open source printers just arrived to the school environment (Irwin, Opplinger, Pearce, & Anzalone, 2015); Schelly, Anzalone, Wijnen & Pearce, 2015). In the 3D printing technology based on the rapid prototyping process, additive processes are used, with successive layers of thermoplastic material laid down according to a computer program (Grujović, Radović, Kanjevac, Borota, Grujović, & Divac, 2011; Ravikumar, Khan, Mohanty, Sageer, & Aigali, 2015). Thus for example, the Voronoi tower can be created. Two- and three-dimensional Voronoi diagrams (Voronoi tessellations) are now commonly applied in architectural concepts, and other science and technology fields. Using the 3D printing technology and dinosaur fossils from the American Museum of Natural History paleontology collections, a group of high school students learned to think like paleontologists while producing models of dinosaurs as part of the innovative program “Capturing Dinosaurs: Reconstructing Extinct Species Through Digital Fabrication” (American Museum, 2013).

In educational terms, scientific visualization may present the art-science cooperative learning projects in order to make knowledge comprehensible both to the students and to a wide audience. Artists used to transform patterns and repetitions to apply the unity or symmetry in their compositions (for example, by examining a Fibonacci sequence, prime numbers and magic squares, a golden section, or tessellation techniques). Mathematicians, computing scientists, and artists used to apply visual metaphors as a cognitive tool to visualize the world's structure and our knowledge. For example, hierarchical structures are predominantly analyzed with the use of a tree metaphor. Manuel Lima called the tree figure the most ubiquitous and long-lasting visual metaphor, "through which we can observe the evolution of human consciousness, ideology, culture, and society" (Lima, 2014, p. 42). Figure 1 presents a work of a student Hunter Shioshita of the University of Northern Colorado entitled "The History of Rock and Roll" designed in a tree format.

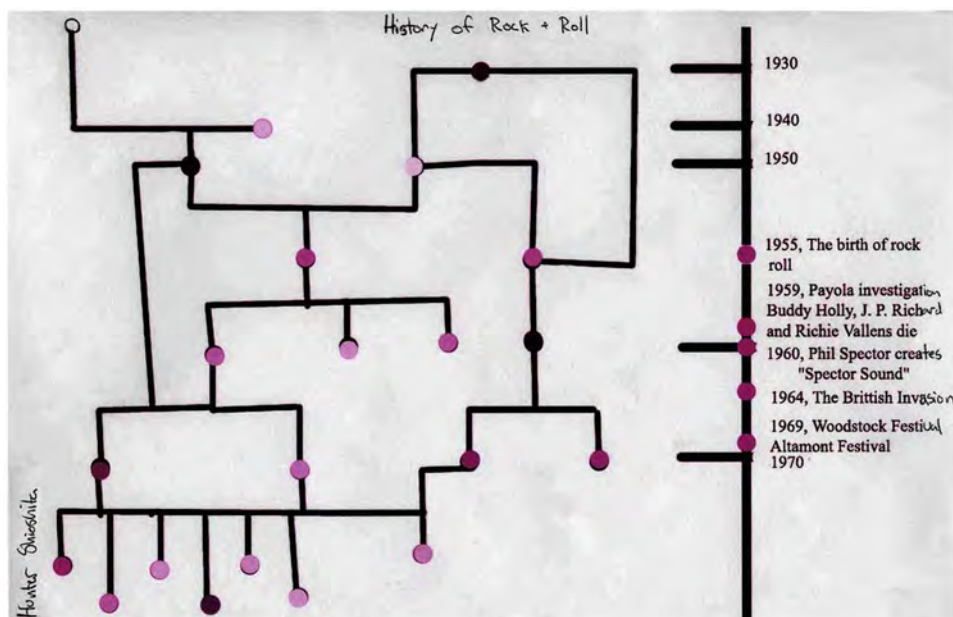
According to Lima, visualizations fall into all three categories of science, design, and art; they are used as a tool for understanding data – i.e.

discovering patterns, connections, and structure. As a scientific tool, information visualization serves for discovery of new knowledge; as a design tool it facilitates the perception of patterns and evokes emotions in the viewers; as art it is a technique to produce something non-utilitarian and aesthetically interesting (Lima, 2011, p. 12). In the Manuel Lima's book (2011), Lev Manovich indicates the important features that make information visualization unique: projects are visually dense, with more data; they show relations between data; in aesthetical terms, they show complexity (chaos theory, emergent complexity theory) rather than reduction (breaking down into the simplest elements).

SOLUTIONS AND RECOMMENDATIONS

Combining visualization techniques with the recent advances in the selected fields of science would comply with the actual trends in education described as the STEM (science, technology, engineering, and mathematics) and the STEAM

Figure 1. Hunter Shioshita, *The History of Rock and Roll* (© 2005, H. Shioshita. Used with permission)



(science, technology, engineering, arts, and mathematics) programs in education. Integration of science, technology, and art offers a response to the actual trends in education described as the STEAM (science, technology, engineering, art, and mathematics). Many educators have been working toward transforming STEM into STEAM as a framework for teaching across the disciplines (STEM to STEAM, 2014; Gonzalez & Kuenzi, 2012; Maeda, 2012). This trend results from the growing technological literacy in small children who can now use software to learn science, mathematics, programming, and storytelling.

The existing and coming tendencies and solutions in educational and instructional environment should include solving problems through knowledge visualization. Learning programming languages should enhance programs for Studio Art. Students should learn, create, and play together globally through educational games, VR, teaching oriented apps, social networking, and meaningful communication, K through PhD (DeVry University, 2014). Technologies, and practices aimed at enhancing visual literacy and thus supporting learning about science may include toys, games, puzzles, apps, models (often involving 3D printing), animations, simulations (automatic pilot, control tower), and augmented reality environments, among other solutions. Perhaps there is a need for focus on games, apps, and ways to interact through social networking that would bring more learning opportunities, actively converting exchange of ideas into somehow curiosity-based learning realm.

In business corporations and commercial companies there is a trend to include the arts in business, and thus support teambuilding, communication, and leadership. It is also a response to the economical and social demands for the arts-based development training, which became used in corporations. Not only we should integrate scientific, technical, and artistic topics, but also construct, envision, and explain through visuals and interactivity (Honey, Pearson, & Schwein-gruber, 2014).

FUTURE RESEARCH DIRECTIONS

Research on neuronal networks in the brain and its organizational principles may shift the focus of visualization trends. For example, John O'Keefe, along with May-Britt Moser and Edvard I. Moser performed a study on grid cells in brain that was awarded the 2014 Nobel Prize in Physiology or Medicine. According to John O'Keefe, grid cells perform as "inner GPS" that helps the brain navigate through the world" (Devlin, 2014). The cells work together to perform cognitive functions; they provide a context-independent metric representation of local environment (Giocomo, Moser, & Moser, 2011). The firing of the grid cells forms a positioning system in the brain, which tells about the animal's speed and running direction at each location in the environment (Moser, Moser, & Roudi, 2014). Computational models of grid cells present how the neurons fire within a tessellating pattern, and how firing properties depend not primary on sensory signals but rather on internal self-organizing principles.

Current perspective on visualization techniques may evolve accordingly to the developments in technologies that provide an extension or augmentation of the sensory apparatus of the student or child (Eisenberg & Eisenberg, 2015). Novel types of scientific instrumentation: for example, telescopes and microscopes are extending the range of vision, and Geiger counters extend sensitivity to phenomena such as radioactive decay. In a way, eyeglasses, hearing aids, or walking sticks and prosthetic devices work that way by improving or repairing our senses. Another venue in sensory augmentation is in the burgeoning area of wearable technology, pervasive/ubiquitous computing, and prosthetics (Eisenberg & Eisenberg, 2015). There seems to be a great potential for visualization of the nanosize objects and events because nano objects are invisible and inaccessible otherwise.

Developments in visualization may go in parallel with those in other scientific areas. For example, the quality of imaging with optical microscopy depends on the right selection of a microscope

according to the imaging depth needed (de Grand & Bonfig, 2015). Super resolved fluorescence microscopy brought optical microscopy into the nano dimension and won the 2014 Nobel Prize (Chang, 2014). The U.S. National Nanotechnology Initiative defines nanotechnology as “the understanding and control of matter at dimensions between approximately 1 and 100 nanometers, where unique phenomena enable novel applications. Encompassing nanoscale science, engineering, and technology, nanotechnology involves imaging, measuring, modeling, and manipulating matter at this length scale” (Nanowerk, 2015).

Materials seen in the nano scale used to display the size-dependent properties resulting from quantum effects. These properties are distinct from those of materials controlled by laws of classical physics; hence their physical, chemical, and biological characteristics may be different at the macroscopic scale (Nanowerk, 2015). For example, visualizing butterfly wing scales look not the same when seen at different magnifications: a wing that is dark blue when looked at without magnification becomes yellow at 220x magnification, purple at 5000x magnification, and green at 20000x magnification (PennState Modules, 2011).

Future applications of visualization techniques may surely include bioimaging and numerous related bio inspired technologies in science, engineering, medicine, pharmacology, and many other disciplines. The discipline of bioimaging explores biological structures and functions to create information visualizations in two, three, or four dimensions. Bioimaging researchers gather and analyze information from sources such as light waves, nuclear magnetic resonance, x-rays, or ultrasounds. Current methods of bioimaging may comprise optical light microscopy, several medical imaging techniques such as magnetic resonance imaging, nuclear magnetic resonance imaging, magnetic resonance tomography, functional magnetic resonance imaging, and many other advanced approaches involving microspec-

troscopy, to investigate structures at the scale of micrometers.

Other fields that involve visualization comprise Raman spectroscopy, photoacoustic imaging techniques, bioimaging with fluorescence, magnetic supernanoparticles, bioimaging with laser near-infrared (NIR) fluorescence, imaging with multifunctional nanostructures, among a number of other technologies. For example, researchers are developing the infrared camouflage tape for military use inspired by the squid skin properties (No author, Squid skin, 2015). Recently, imaging up to 8 mm into a fixed tissue is possible in multiphoton microscopes. These systems do not require slicing and thus damaging tissues, and thus can be used to provide neurobiological and other information about deep, intact structures. Lasers with dual-wavelength ultrafast (femtosecond) pulses provide simultaneous activation of multiple neurons (Arrigoni & Gallaher, 2015).

CONCLUSION

Significant characteristics of current culture of information systems include visualization techniques, interactive visualizations, and the visual content analysis applied to data and knowledge. The use of visualization techniques in various kinds of presentations plays pivotal role in communication media and educational programs. Nature derived metaphors serve as the enrichment of bio-inspired, interdisciplinary models and support data visualization, information and knowledge visualization, data mining, and semantic web. This caused a need for combining metaphorical visualization with artistic principles and working on metaphorical way of learning and teaching using art and graphic metaphors for the information and communication purposes. Visualization that help making a progress in science provide a set of reasons and a logical basis why visualization ability should be taught and trained since early childhood.

REFERENCES

- American Museum of Natural History. (2013). *Students Use 3D Printing to Reconstruct Dinosaurs*. Retrieved November 10, 2015, from https://www.youtube.com/watch?v=_KBxG1_WO8k
- Arrigoni, M., & Gallaher, N. (2015, February-March). Diverse Applications Drive Lasers for Multiphoton Microscopy. *BioPhotonics*, 32-36.
- Bederson, B., & Shneiderman, B. (2003). *The Craft of Information Visualization: Readings and Reflections*. San Francisco, CA: Morgan Kaufmann Publishers.
- Bertschi, S., Bresciani, S., Crawford, T., Goebel, R., Kienreich, W., Lindner, M., & Moere, A. V. et al. (2011). What is Knowledge Visualization? Perspectives on an Emerging Discipline. *Proceedings of the Information Visualisation 15th International Conference*, 329-336. doi:10.1109/IV.2011.58
- Bertschi, S., & Bubenhofer, N. (2005). Linguistic learning: a new conceptual focus in knowledge visualization. *Proceedings of the Information Visualisation 9th International Conference*, 383-389. doi:10.1109/IV.2005.71
- Bredl, K., Groß, A., Hünninger, J., & Fleischer, J. (2012). The Avatar as a Knowledge Worker? How Immersive 3D Virtual Environments may Foster Knowledge Acquisition. *Electronic Journal of Knowledge Management*, 10(1), 15-25.
- Burkhard, R., Meier, M., Smis, J. M., Allemang, J., & Honish, J. (2005). Beyond Excel and PowerPoint: Knowledge Maps for the Transfer and Creation of Knowledge in Organizations. In *iV, International Conference on Information Visualisation*, 403-408.
- Card, S., Mackinlay, J., & Shneiderman, B. (1999). *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann Publishers.
- Chang. (2014, October 14). Nobel Laureates Pushed Limits of Microscopes. *The New York Times*. Retrieved November 11, 2015, from http://www.nytimes.com/2014/10/09/science/nobel-prize-chemistry.html?_r=0
- Chen, C. (2010). *Information Visualization: Beyond the Horizon* (2nd ed.). Springer.
- Chen, C. H. (2011). *Emerging topics in computer vision and its applications*. World Scientific Publishing Company. doi:10.1142/8103
- Craft, B., & Cairns, P. (2005). Beyond Guidelines: What Can We Learn from the Visual Information Seeking Mantra? *Proceedings of the 9th International Conference on Information Visualization*, 110-118. doi:10.1109/IV.2005.28
- Craft, B., & Cairns, P. (2008). Directions for Methodological Research in Information Visualization. *Proceedings of the 12th International Conference on Information Visualisation*. London, UK: IEEE Computer Society Press. doi:10.1109/IV.2008.88
- De Grand, A., & Bonfig, S. (2015, January). Selecting a Microscope Based on Imaging Depth. *BioPhotonics*, 26-29.
- Devlin, H. (2014, November 11). John O'Keefe awarded Nobel Prize for discovery of brain's 'GPS' system. *The Times*. Retrieved November 10, 2015, from <http://www.thetimes.co.uk/tto/science/article4228469.ece>
- DeVry University. (2014). *Web Game Programming Degree Specialization*. Retrieved November 11, 2015 from <http://www.devry.edu/degree-programs/college-media-arts-technology/web-game-programming-about.html>
- Eisenberg, M., & Eisenberg, A. (2015). Sensory Extension as a Tool for Cognitive Learning. In A. Ursyn (Ed.), *Handbook of Research on Maximizing Cognitive Learning through Knowledge Visualization* (pp. 72-83). IGI Global. doi:10.4018/978-1-4666-8142-2.ch002

- Frøkjær, E., Hertzum, M., & Hornbæk, K. (2000). Measuring Usability: Are Effectiveness, Efficiency, and Satisfaction Really Correlated? *CHI 2000 Conference Proceedings*, 345-352.
- Giocomo, L. M., Moser, M.-B., & Moser, E. I. (2011). Computational models of grid cells. *Neuron*, 71(4), 589–603. doi:10.1016/j.neuron.2011.07.023 PMID:21867877
- Gonzalez, H. B., & Kuenzi, J. J. (2012). *Science, Technology, Engineering, and Mathematics (STEM) Education: A Primer*. Congressional Research Service. Retrieved November 11, 2015, from <http://fas.org/sgp/crs/misc/R42642.pdf>
- Grujović, N., Radović, M., Kanjevac, V., Borota, J., Grujović, G., & Divac, D. (2011). 3D printing technology in education environment. *34th International Conference on Production Engineering*, 29-30.
- Honey, M., Pearson, G., & Schweingruber, H. (Eds.). (2014). *National Academy of Engineering and National Research Council. STEM Integration in K-12 Education: Status, Prospects, and an Agenda for Research*. Washington, DC: The National Academies Press.
- Irwin, J. L., Opplinger, D. E., Pearce, J. M., & Anzalone, G. (2015). Evaluation of RepRap 3D Printer Workshops in K-12 STEM (Program/Curriculum Evaluation). *122nd ASEE Annual Conference & Exposition*. doi:10.18260/p.24033
- Lengler, R., & Eppler, M. J. (2008). *A Periodic Table of Visualization Methods for Management*. Retrieved November 10, 2015, from http://www.visual-literacy.org/periodic_table/periodic_table.html
- Lima, M. (2011). *Visual Complexity: Mapping Patterns of Information*. New York: Princeton Architectural Press.
- Lima, M. (2014). *The Book of Trees: Visualizing the Branches of Knowledge*. New York: Princeton Architectural Press.
- Maeda, J. (2012). *STEM to STEAM: Art in K-12 is Key to Building a Strong Economy*. Edutopia. Retrieved November 11, 2015, from <http://www.edutopia.org/blog/stem-to-steam-strengthens-economy-john-maeda>
- Marchese, F. T. (2011). Exploring the Origins of Tables for Information Visualization. *Proceedings of the Information Visualisation 15th International Conference*, 395-402. doi:10.1109/IV.2011.36
- Mercuri, R., & Meredith, K. (2014). An educational venture into 3D Printing. *Integrated STEM Education Conference (ISEC)*, 1-6.
- Moser, E. I., Moser, M.-B., & Roudi, Y. (2014). Network mechanisms of grid cells. *Phil. Trans. R. Soc. B*, 369(1635), 20120511. doi:10.1098/rstb.2012.0511 PMID:24366126
- Nanowerk. (2015). *Introduction to nanotechnology*. Retrieved August 10, 2015, from http://www.nanowerk.com/nanotechnology/introduction/introduction_to_nanotechnology_1.php#ixzz3iXtNEYyL
- Penn State Modules. (2011). *Nano4Me.org. NACK educational resources*. The Pennsylvania State University. Retrieved November 11, 2015, from <http://nano4me.live.subhub.com/categories/modules>
- Ravikumar, R., & Mohsin Khan, I. (2015). Design & Development of a 3D Printer. *Proceedings of 12th IRF International Conference*.
- Schelly, C., Anzalone, G., Wijnen, B., & Pearce, J. M. (2015). Open-Source 3-D Printing Technologies for Education: Bringing Additive Manufacturing to the Classroom. *Journal of Visual Languages and Computing*, 28, 226–237. doi:10.1016/j.jvlc.2015.01.004
- Shneiderman, B. (1996). The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In *Proceedings of the IEEE Symposium on Visual Languages*, (pp. 336-343). Washington, DC: IEEE Computer Society Press. doi:10.1109/VL.1996.545307

Shneiderman, B. (2014). *Treemaps for space constrained visualization of hierarchies*. Retrieved August 10, 2015, from: <http://www.cs.umd.edu/hcil/treemap-history/>

STEM to STEAM. (2014). Retrieved September 9, 2015, from <http://stemtosteam.org/>

Tufte, E. R. (1983/2001). *The Visual Display of Quantitative Information* (2nd ed.). Cheshire, CT: Graphics Press.

Tufte, E. R. (1992/2005). *Envisioning Information*. Cheshire, CT: Graphics Press. Third printing with revision.

Ward, M., Grinstein, G. G., & Keim, D. (2010). *Interactive data visualization: foundations, techniques, and applications*. Natick, MA: A K Peters Ltd.

ADDITIONAL READING

Klanten, R., Bourquin, N., Ehmann, S., van Heerden, F., & Tissot, T. (Eds.). (2008). *Data Flow: Visualising Information in Graphic Design*. Die Gestalten Verlag.

Lima, M. (2011). *Visual complexity: Mapping patterns of information*. New York: Princeton Architectural Press.

Lima, M. (2014). *The Book of Trees: Visualizing the Branches of Knowledge*. New York: Princeton Architectural Press.

Yau, N. (2011). *Visualize this: The flowing data guide to design visualization and statistics*. Wiley.

Yau, N. (2013). *Data Points: Visualization That Means Something*. Wiley.

KEY TERMS AND DEFINITIONS

Algorithm: A mathematical sequence of instructions telling how to carry on computation to implement it as a program. Algorithm serves for

solving a complex problem by writing a sequence of simpler, unambiguous steps.

Concept Map: A graphical two-dimensional display of knowledge. Concepts, usually presented within boxes or circles, are connected by directed arcs that encode, as linking phrases, the relationships between the pairs of concepts.

Data Visualization: Information abstracted in a schematic form to provide visual insights into sets of data. Data visualization enables us to go from the abstract numbers in a computer program (ones and zeros) to visual interpretation of data. Text visualization means converting textual information into graphic representation, so we can see information without having to read the data, as tables, histograms, pie or bar charts, or Cartesian coordinates.

Infographics: Tools and techniques involved in graphical representation of data, mostly in journalism, art, and storytelling.

Information Visualization: Often characterized as representation plus interaction, means the use of computer-supported, interactive visual representations of abstract data to amplify cognition and derive new insights.

Knowledge Visualization: Uses visual representations to transfer insights and create new knowledge in the process of communicating different visual formats.

Pattern: The regular order existing in nature or in a manmade design. In nature patterns can be seen as symmetries (e.g., snowflakes) and/or structures having fractal dimension such as spirals, meanders, or surface waves. In computer science, design patterns serve in creating computer programs. In the arts, pattern is an artistic or decorative design made of recurring lines or any repeated elements. We can see patterns everywhere in nature, mathematics, art, architecture, and design. Patterns make a basis of ornaments, which are specific for different cultures.

Scientific Visualization: Presents real, abstract, or model-based objects in a digital way directly from the data. Visualization as storytelling comprises narratives, interactive graphics, explanatory and animated graphics, and multimedia.

Category L

Language Studies

Mobile Testing System for Developing Language Skills

Svetlana Titova

Far Eastern Federal University, Russia

INTRODUCTION

Recent research demonstrated that language skills can be enhanced through mobile technologies that transformed drastically foreign language learning/teaching experience offering immediate diagnosis of learning problems and design of new assessment models (Cooney & Keogh, 2007); creating mobile networking collaboration (Lan, Sung & Chang, 2007; Pemberton, Winter & Fallahkhair, 2010), enhancing autonomy (Murphy, Bollen & Langdon, 2012); providing instant feedback and a personalized learning experience (Voelkel & Bennett, 2013; Oberg & Daniels, 2013); enabling teachers to create new formats of problem-solving tasks based on augmented reality (Cook, 2010; Driver, 2012).

But in spite of the plethora of research in the area of mobile learning, it challenges instructors to examine how the pedagogical potential provided by mobile technologies relates to their teaching aims, methods, and subject matter because there is not yet consistent MALL methodology. There is a need for a new educational framework for mobile testing apps implementation aimed at developing learner skills rather than just assessing learner knowledge. The hypothesis to guide the framework of this research includes enquiry-based learning pedagogy and educational opportunities provided by ubiquitous devices such as interactivity and immediate feedback. This study supported by both current m-learning theory and enquiry-based approach focuses on working out a methodological framework for mobile testing system implementation into the language classroom.

BACKGROUND

Today teachers who would like to meet the expectations of a new generation of mobile natives need to follow a transformational approach (Puente, 2011) to the development of language skills based on creative use of mobile technologies within the learner-oriented environment. The main prerequisite of this environment functioning is collaborative peer-learning approach. The social framework, learners' expertise, and cultural practices are gaining importance, the role of the devices is becoming less important. Mobility is no longer defined through the devices, but through "the learners' abilities to act flexibly in ever changing and self-constructed learning contexts" (Seipold, 2011, p.32). Seipold argues that only if teachers provide spaces to learners to act according to their interests, agency, and cultural practices, innovative use of the devices can be discovered by learners (2011).

The research framework is also based on Mishra and Koehler's model for implementing new technologies into teaching - Technological Pedagogical Content Knowledge (TPCK) (2006). This approach suggests teachers should be aiming to reach a point where their traditional content and pedagogical knowledge is enhanced by technological knowledge. According to TPCK framework, a new tool complements teachers' knowledge and skills. This theoretical perspective suggests that learning is affected and modified by the tools employed for it and that reciprocally these tools are adjusted in the way they are used for learning. As Stockwell and Hubbard argue: "Let the language

learning task fit the technology and environment, and let the technology and environment fit the task” (2013, p.9). The Substitution Augmentation Modification Redefinition model developed by Puentedura (2011) can be used as a complement to TPACK. According to this model the use of new tech tools in education may lead either to the enhancement of education (augmentation and substitution phases) or to the real transformation (redefinition and modification phases). Redefinition is the highest transformation phase which allows for a completely new format of tasks and activities that were previously impossible. This approach also offers a perspective in which the pedagogical considerations shape the design of mobile learning.

Other important theories that have been influential to work out the framework of this research are enquiry-based learning and behaviorist approach. Many researchers today highlight social aspects of mobile technologies proposing complex structures of m-learning pedagogy built on Vygotsky’s hypothesis about the importance of discussions in an educational context (Sharples, Taylor, & Vavoula, 2007). Enquiry-based learning is a shift away from passive methods to the problem-based methods through which students are expected to construct their own knowledge and understandings by taking part in supported processes of enquiry (Kahn & O’Rourke, 2005).

Danaher, Gururajan, and Hafeez-Baig (2009) propose the m-learning structure based on three principles: engagement, presence, and flexibility. Presence is characterized as an interaction which is sub-divided into three types: cognitive, social and teaching. Kearney, Schuck, Burden and Aubusson (2012) argue that the main constituents of m-learning pedagogy are personalization, authenticity, and collaboration. Mobile technologies enable instructors to create collaboration environment that motivates students to learn for themselves, bringing a research-based approach to the subject. This interactive, “dialogic models of learning are similar to the processes of participation in research” (Sambell, 2010, 56).

Ubiquitous access to information via mobile devices potentially enables a paradigmatic shift in education, it changes the way classes are managed and the instructor’s role (Beatty, 2004). Kahn and O’Rourke (2005) argue that enquiry-based learning encourages students to seek out new evidence for themselves and support peer learning approach. This approach implies a principal change in the paradigm of teaching due to the fact that mobile devices effectively “act as accelerators of the social discourse” (DeGani, Martin, Stead, & Wade, 2010, p.181).

THE PEDAGOGICAL POTENTIAL OF THE MOBILE TESTING SYSTEM PELE TO ENHANCE LANGUAGE LEARNING

The mobile testing system PeLe for handheld devices, developed at Sør-Trøndelag University, enables instructors to deliver a test through any mobile device, assess it and provide timely feedback to both individual students and a group of students immediately after a test. Using Pele students can respond to tests electronically and the teacher can see on the screen what is happening during the test. The technological characteristics and the pedagogical potential of PeLe are summed up in table 1.

PeLe suits perfectly to evaluate both group dynamics and individual student results, it was primarily used in our research for *formative assessment* or low stake assessment which serves to give learners feedback on their performance (Sambell, 2010).

METHODOLOGY

Research Objectives

Mobile Language Learning (MOBILL) was an international project involving two institutions Sør-Trøndelag University College (HiST, Norway),



Table 1. Technological characteristics and the pedagogical potential of PeLe

Technological Characteristics of PeLe	Pedagogical Potential
Multiple choice tests can be delivered either via PeLe or in the written form	• Teaching in technologically limited environments • No need for profound tech preparation
Instant visualization of the test results	• Group dynamics evaluation • Formative assessment approach • Enhance learner motivation
Immediate test assessment and feedback	• Timely diagnosis of teaching/learning problems • Instant feedback on learning problems • Revision of teaching strategies
Student Response System is installed	• Can be used to conduct in-class surveys • Encourage peer discussions and post-test activities
Database stores both group and individual results	• E-portfolio approach • Formative assessment approach • Evaluation of group dynamics
It returns individual feedback to students' mobile devices	• Individual approach to each student • E-Portfolio approach
Multiple-choice questions are a mixture of text and images	• Visualization of learning materials • Enhancing learner motivation
The teacher can see on his screen what is happening during the test	• Evaluation of student learning progress • Any aspect of student output is under control
Equipment necessary: one internet-enabled teacher computer and internet-enabled student mobile devices	• Teaching in technologically limited environments • No need for bulky costly equipment
Use of student own devices	• No need for tech instructions - familiar devices

and Lomonosov Moscow State University (LMSU, Russia). The project was conducted during two periods from September 2013 to January 2014, and from February to May 2014. The key objective of this research was to work out sound pedagogical strategies on how to implement PeLe into the traditional language classroom (pedagogical perspective), thus introducing some improvements to the piloting tool (technological perspective). At the first period teachers from LMSU and HiST piloted PeLe and try to develop Mobile Assisted Language Learning Model (MALLM) based on enquiry-based approach and formative assessment. The following research questions were proposed:

1. What is the PeLe pedagogical potential to develop student language skills?
2. Does PeLe intervention impact assessment patterns of the traditional classroom and foster the development of learner language skills?

3. How can enquiry-based methods be effectively implemented into MALLM?
4. Does the proposed MALLM impact student motivation and to what extend?

THE PROJECT IMPLEMENTATION AND RESEARCH DESIGN: METHODOLOGY OF PELE INTERVENTION

The methodological framework of the MALL model based on PeLe implementation includes both enquiry-based methods such as collaborative and peer learning post-test activities, brainstorming, problem-solving activities, group discussions and mobile learning opportunities such as immediate feedback, formative assessment, and interactivity. We offered the following procedure of PeLe intervention employed in this study:

1. Setting up the Assessment Template

The teacher sets up the PeLe assessment template choosing the number of questions, alternatives, and the correct answer. In this case the teacher has three options: he can use the printed version of the test, he can show the test on the IWB or he can save the test on PeLe.

2. The Test is on.

Students take the multiple-choice test and the teacher gets the answers monitoring this process on his computer screen. When the students are picking up the correct answers using their devices the teacher can follow the group dynamics and each student test results. At this stage the teacher can see what kind of problems students have as a group and individually.

3. Test Submission and Instant Feedback

The teacher takes some time (usually about 3-4 minutes) to analyze test results, then students are shown their group scores on each test item in the form of a diagram. The teacher chooses the items that were the most difficult ones for students. After that, the pedagogical agent selects scaffolding from the range of post-test activities: teacher explanation, a group discussion, a pair discussion or SRS-supported activity.

4. Post-Test Activities

The teacher has to pick up a post-test activity aimed at improving test results. The type of activity depends upon the group results on a test question.

We offered the following activities: teacher explanation, group discussion, pair discussion. To figure out what types of post-test activities were the most effective ones we asked the teachers to fill in the grids on post-test activities used after each test.

5. Second Voting on the Tough Questions

After post-test activities, students take the second vote on the discussed questions using either SRS installed or PeLe. At this stage, the teacher can also provide immediate feedback comparing and demonstrating group results of the first and second trial.

RESEARCH PARTICIPANTS

The project target groups consisted of Norwegian and Russian learners of English, all at approximately the same language level (B1) according to the Common European Framework for Languages (see table 2). In the first period of experimentation, students enrolled in a preparatory English course at LMSU were randomly assigned to 3 experimental groups (EXG) and 1 control group (CG). Students of the experimental groups took a series of PeLe supported formative tests as volunteers using handheld devices. The control group was tested by pen and paper method.

Table 2. The project target groups

	Group Size	Male	Female	Number of Formative Tests	Entrance test MS	Entrance Test Score SD
EXG 1	12	1	10	7	81	5,97
EXG 2	8	1	7	7	64	8,75
EXG 3	9	3	6	7	66	15.09
CG	7	0	7	0	60	7,09

DATA COLLECTION

Data collection was done in 3 cycles:

1. The intervention of PeLe tests as formative assessment tools from September to December 2013 in three experimental groups. Quantitative data of the first and the second voting of PeLe tests were analyzed by mean and standard deviation, students' t-test results. The grid on post-test activities used after each test were completed by the teachers of the experimental groups.
2. Quantitative data of the final tests were gathered in control and experimental groups to compare overall performance at the end of the semester, the data were analyzed by mean and standard deviation, t-test results.
3. The post-intervention questionnaire asks students to reflect on their attitude to PeLe integration. Qualitative data were gathered to help explain quantitative findings.

RESEARCH RESULTS AND DATA DISCUSSION

Cycle 1

In the first cycle, the students of the three experimental groups took PeLe multiple-choice grammar and vocabulary tests according to PeLe methodology intervention. In the experimental groups, formative Pele tests were provided in the form of in-class grammar and vocabulary tests. Students responded with their smart phones or tablets. They had access to PeLe tests by using Wi-Fi in class.

The students of the control group were taught in the traditional way, they took the placement and final tests for summative assessment, they were not supposed to take formative PeLe supported tests with immediate feedback on their results.

According to the methodology of PeLe intervention we collected the quantitative data on the results of the first voting (FV) and the results of the second voting (SV) of the experimental groups (EXG) and analyzed the results. To assess the magnitude of any significant changes following the intervention, effect sizes were calculated according to the methods of mathematical statistics - the standard deviation (SD) and mean score (MS) for each group (see table 3). Based on the calculations, we can conclude that the effectiveness of the formative tests increased after post-tests activities. These results indicated that there was a substantial improvement in group 2 and 3 where the entrance test mean scores were 64 and 66 correspondingly although there was not the so substantial improvement in group 1 where the entrance test mean result was much higher - 88. In other words, it turned out that PeLe supported approach was more beneficial in our case for the groups with lower language level: FV MS1= 56< SV MS1= 75, FV MS2 =57< SV MS2=77. Statistical differences between the first and second voting of the PeLe supported test in the experimental groups were assessed using student's t-test for independent samples, as appropriate. T-test results of the first and second voting in group 2 and group 3 statistically are quite significant: EXG2 T-test= 3,0512 EXG3 T-test=3,2342. For group 1 T-test is 0,5023, by conventional criteria, this difference is considered to be not statistically significant.

Table 3. The mean score, standard deviation, t-test of the first and second voting of the PeLe supported test in the experimental groups

	FV MS	SV MS	FV SD	SV SD	FV and SV T-Test
EXG1	85	87	7,9	8,05	0,5023
EXG2	56	75	13,11	10,17	3,0512
EXG3	57	77	14,44	8,40	3,2342

Our study demonstrated that the second voting results in the experimental groups were better than test results of the first trial. The main reasons for that could be immediate feedback on group test results and post-test activities offered by the teachers in the experimental groups. Hattie and Timperley (2007) emphasize that effective feedback needs to provide information that specifically relates to the task so that students can develop error detection strategies and use the feedback to tackle more challenging tasks.

To answer the third question of our research we tried to figure out, first, what kind of post-test activity led to any improvement in student performance so was the most efficient one, second, whether it was the correlation between the group item score and the type of activity offered by the instructor. Our grid data analysis demonstrated that on average in each group 5.125 activities were offered after each test. The most frequently used by the instructors activity was the class discussion - 48% of the activities offered, then comes teacher explanation - 32,5% of the activities offered, the least frequently used activity was group discussion activities - 19,5% of all the activities (see table 4).

Class discussions are likely to be used more frequently because for the instructors it was easier to be a class facilitator and to initiate discussions asking open-ended questions that provoke further discussion, stimulate deeper exploration and challenge student thinking, encouraging them to seek new ways to work with problems and situations. But in this case only carefully formulated questions

can stimulate the generation of ideas and interest in what students have to say, providing clues as to whether students are 'on track' (Kahn & O'Rourke, 2005). Whereas peer tuition approach seems to be very time and effort consuming on the part of the teacher. Although experimental language groups were not numerous so it was quite comfortable for the learners and the instructor to have a discussion. Peer-tuition approach in language classes might also be complicated due to the fact that the language of instruction was foreign for the learners.

Our data analysis demonstrated that the most efficient post-test activity was the class discussion because the increase in the second trial test results was significant. These increases were statistically more significant for group 2 $T\text{-test}=3,053$, $FV\ MS2=56 < SV\ MS2=75$, and for group 3 $T\text{-test}=3,2342$, $FV\ MS3=57 < SV\ MS3=77$. In each group 21 class discussion activities were used (see table 5). These results substantiate the value of group discussions for student understanding, confirming the significance of constructivist approach in mobile learning (Arnesen, Korpås, Hennissen, & Stav, 2013).

The next question we tried to figure out is the correlation between the group item score and the type of activity offered. Our grid data analysis demonstrated that if from 50% of the group members gave incorrect item answer the instructors preferred teacher explanation technique. If from 50 to 75% of the group members gave the correct item answer the instructors offered either pair or class discussion activities. In this case,

Table 4. Number and type of post-test activities

	Teacher Explanation		Class Discussion		Group Discussion		Total Number of Activities		The Average Number of Activities Per Group
	%	N	%	N	%	N	%	N	
EXG 1	34,4%	11	53,1%	17	12,5%	4	26%	32	5,125
EXG2	37,5%	18	43,8%	21	19,9%	9	39%	48	
EXG3	25,6%	11	48,8%	21	25,6%	11	35%	43	
All Groups	32,5%	40	48%	59	19,5%	24		123	

Table 5. Correlation between second voting test results and the post-activities offered

	SV MS	FV and SV T-Test	Number of post- test activities		
			Teacher Explanation	Class Discussion	Group Discussion
EXG 1	87	0,5023	11	17	4
EXG 2	75	3,0512	18	21	9
EXG 3	77	3,2342	11	21	11

the teacher choice depends on the two important considerations whether it is necessary to give some additional input to initiate the discussion and whether the chosen activity helped save the time assigned for the test and its discussion.

Cycle 2

The learners of both control and experimental groups were given the same placement and final tests. These tests were used for summative assessment. The overall mean scores were included to compare overall performance of the control and experimental groups after the implementation of the intervention. The data collected on the mean scores of the entrance and final tests in control and experimental groups suggest that introduction of PeLe tests helped improve academic performance in the experimental groups in mean results of final test (see table 6) whereas the control group demonstrated just a slight increase in mean scores (60>62). Statistical differences between the two tests were also assessed using Student's t-test for independent samples: t-test results in group 1 are 1.807, in group 2 - 2.6201, in group 3 - 1.2405. In the control group t-test is the lowest - 0,7025. These data suggest that the introduction of PeLe

supported approach helped improve student performance in all three experimental groups.

These results support our hypothesis that collaborative enquiry-based learning and educational opportunities provided by handheld devices formative assessment led to a significantly better exam performance of the students who took PeLe quizzes, compared to those who did not. The summative part of PeLe tests ensured a high completion rate, whereas the formative part provided students with prompt feedback and gave them information on what they needed to do to improve their performance (Voelkel, 2013). The increase in the overall exam results was encouraging, but not conclusive to show that only PeLe tests were beneficial. It could also be due to timely feedback on test results, collaborative post-test activities and administration of formative tests.

Cycle 3

At this cycle, an online survey was administered to elicit students' own perceptions of PeLe intervention. The quantitative data were supplemented by student feedback gained from a post-study questionnaire. The post-study questionnaire contains 11 questions aiming to get student views on the

Table 6. Final and entrance test results in control and experimental groups

GROUPS	Entrance Test MS	Entrance Test SD	Final Test MS	Final Test SD	T-TEST
EXG1	81	5.97	87	8.23	1.8027
EXG2	64	8.75	78	14.58	2.6201
EXG3	66	15.09	73	10.30	1.2405
CG	60	7.09	62	4.88	0,7025



Table 7. Results of the post-study questionnaire

	Strongly Agree	Agree	Disagree	Strongly Disagree	Median Score
1. Mobile devices are the best tools to be used for language practice	17	4	3	0	3,58
2. PeLe tests helped me understand the topic in focus	16	6	2	0	3,58
3. PeLe tests helped me get ready for midterm and final a lot	16	4	2	2	3,41
4. Instant feedback after PeLe tests was very supportive and encouraging for my learning	22	2	0	0	3,91
5. Post-test activities made me better understand grammar rules	14	6	4	0	3,41
6. Post-test activities were very helpful and timely	17	5	2	0	3,62
7. Re-voting on SRS after post-test activities helped me correct my mistakes	11	6	3	4	3,00
8. Activity switching kept me engaged in class	10	8	4	2	3,08
9. The use of mobile devices and tasks based on PeLe was fun and changed my attitude to learning	10	8	4	2	3,08
10. PeLe tests were frustrating, they complicated my learning a lot	0	0	4	20	1,16
11. PeLe based tests were motivating	16	6	2	0	3,58

strengths and weaknesses of PeLe intervention. The questionnaire was completed by 24 students (22 female, 2 male) of the experimental groups. Responses are provided in table 7.

In response to statement 1 and statement 2 students expressed high levels of satisfaction (3,58) with PeLe supported tests undertaken on mobile devices. The majority of students commented favorably on the fact that mobile tests helped them understand the topic in focus and get ready for midterms and final with responses to the second statement averaging.

The largely positive reaction to statements 3, 4, 5 and 6, where the mean scores were 3,41, 3,91, 3,41 and 3,62 respectively, emphasizes that immediate feedback on test results was very supportive and encouraging for student learning. Students appreciated the prompt feedback they got on their own understanding of material. In response to statement 7, the average was 3,00. The students were very positive about the fact that re-voting after post-test activities helped them correct their mistakes and promote reflection on

grammar practice. The positive reaction to statement 8, where the mean score was 3,08, indicates that activity switch approach kept the students involved. Statements 9,10 and 11 were designed to get students attitude to PeLe intervention. The majority of students claimed that the use of mobile devices was fun and changed their attitude to learning. It supports the idea that the availability of mobile devices that learners possess makes it an attractive supplement to other forms of teaching and learning a second language (Stockwell & Hubbard, 2013).

In reaction to statement 10, the majority of students disagree on the fact that PeLe tests were frustrating and complicated their learning a lot. The average for statement 11 was 3,58 proving the idea that formative assessment tests promote feedback that seeks to empower students to become motivated and committed to exercising more control over their own learning (Sambel, 2010).

Our findings suggest that students place heavy emphasis on the value of instant, timely feedback on their tests as well as on post-test activities

that stood them in good stead in improving their grammar skills. Despite the overall positive responses, there was also notable ambivalence: several respondents combined positive comments with criticisms. For example, answering statement 7 four students strongly disagreed with the idea that re-voting on tough test items after post-test activities helped them correct their mistakes and promote reflection on their grammar practice. Nonetheless, our analysis demonstrates clearly that the vast majority of students found PeLe implementation to be very appealing.

FUTURE RESEARCH DIRECTIONS

We hope that our research will provide some constructs for pedagogical thinking about enhancing MALL with new mobile-assisted assessment methodology. Formative assessment practice (Black & Wiliam, 2009) lay at the heart of the project's innovative pedagogic approach, so the approach offers a practical way of embodying assessment for language learning environments. Although mobile testing system Pele holds promise but more research is needed to determine its effects upon developing not only grammar skills but also some other skills such as speaking, writing, listening.

CONCLUSION

This research indicated that mobile apps integration into language learning could be efficient, especially if combined with enquiry-based and peer learning approaches and the pedagogical potential provided by mobile testing systems. The experimental results suggest that the MALLM approach combining m-learning and enquiry-based learning theory and formative assessment is most advisable. The research results supported our hypothesis that collaborative enquiry-based learning and educational opportunities provided by handheld devices formative assessment led

to a significantly better exam performance of the students who participated in formative PeLe quizzes, compared to those who did not.

REFERENCES

- Arnesen, K., Korpås, G. S., Hennissen, J., & Stav, J. B. (2013). Experiences with use of Various Pedagogical Methods Utilizing a Student Response System – Motivation and Learning Outcome. *The Electronic Journal of E-Learning*, 11(3), 169–181.
- Beatty, I. (2004). Transforming Student Learning with Classroom Communication Systems. *EDUCAUSE Research Bulletin*, 3. Retrieved April 13, 2016, from <http://net.educause.edu/ir/library/pdf/ERB0403.pdf>
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5–31. doi:10.1007/s11092-008-9068-5
- Cook, J. (2010). Mobile phones as mediating tools within augmented contexts for development. In E. Brown (Ed.), *Education in the wild: contextual and location-based mobile learning in action* (pp. 23-26). University of Nottingham. doi:10.4018/jmbl.2010070101
- Cooney, G., & Keogh, K. (2007). Use of mobile phones for language learning and assessment for learning, a pilot project. *Proceedings of the 6th Annual International Conference on Mobile Learning*. Retrieved April 13, 2016, from <http://www.learnosity.com>
- Danaher, P., Gururajan, R., & Hafeez-Baig, A. (2009). Transforming the practice of mobile learning: Promoting pedagogical innovation through educational principles and strategies that work. In H. Ryu & D. Parsons (Eds.), *Innovative mobile learning: Techniques and technologies* (pp. 21–46). Hershey, PA: Information Science Reference/IGI Global. doi:10.4018/978-1-60566-062-2.ch002

- DeGani, A., Martin, G., Stead, G., & Wade, F. (2010). E-learning Standards for an M-learning world – informing the development of e-learning standards for the mobile web. *Research in Learning Technology*, 25(3), 181–185.
- Driver, P. (2012). Pervasive Games and Mobile Technologies for Embodied Language Learning. *International Journal of Computer-Assisted Language Learning and Teaching*, 2(4), 23–37. doi:10.4018/ijcallt.2012100104
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. doi:10.3102/003465430298487
- Kahn, P., & O'Rourke, K. (2004). *Guide to curriculum design: enquiry-based learning*. York, UK: Higher Education Academy, Imaginative Curriculum Network.
- Kearney, M., Schuck, S., Burden, K., & Aubusson, P. (2012). Viewing mobile learning from a pedagogical perspective. *Research in Learning Technology Journal*, 20(1), 21–34.
- Lan, Y.-J., Sung, Y.-T., & Chang, K.-E. (2007). A mobile-device-supported peer-assisted learning system for collaborative early EFL reading. *Language Learning & Technology*, 11(3), 130–151. Retrieved from <http://ilt.msu.edu/vol11num3/pdf/lansungchang.pdf>
- Mishra, P., & Koehler, M. (2006). Technological Pedagogical Content Knowledge: A Framework for Teacher Knowledge. *Teachers College Record*, 108(6), 1017–1054. doi:10.1111/j.1467-9620.2006.00684.x
- Murphy, P., Bollen, D., & Langdon, C. (2012). Mobile technology, collaborative reading, and elaborative feedback. In J. Díaz-Vera (Ed.), *Left to my own devices: Learner autonomy and mobile-assisted language learning innovation and leadership in English language teaching*. Bingley, UK: Emerald Group Publishing Limited.
- Oberg, A., & Daniels, P. (2013). Analysis of the effect a student-centred mobile learning instructional method has on language acquisition. *Computer Assisted Language Learning*, 26(2), 177–196. doi:10.1080/09588221.2011.649484
- Pemberton, L., Winter, M., & Fallahkhair, S. (2010). Collaborative mobile knowledge sharing for language learners. *Journal of the Research Center for Educational Technology*, 6(1), 144–148.
- Puentedura, R. R. (2011). *A brief introduction to TPACK and SAMR*. Retrieved April 13, 2016, from <http://hippasus.com/rrweblog/archives/2011/12/08/BriefIntroTPCKSAMR.pdf>
- Sambel, K. (2010). Enquiry-based learning and formative assessment environments: Student perspectives 2010. *Practitioner Research in Higher Education*, 4(1), 52–61.
- Seipold, J. (2011). A critical perspective on mobile learning: Results of a heuristic analysis of the scientific process and a hermeneutic analysis of mobile learning practice. *Mobile learning: Crossing boundaries in convergent environments*, 31–34.
- Sharples, M., Taylor, J., & Vavoula, G. (2007). A theory of learning for the mobile age. In R. Andrews & C. Haythornthwaite (Eds.), *The SAGE handbook of e-learning research* (pp. 122–138). London: Sage. doi:10.4135/9781848607859.n10
- Stockwell, G., & Hubbard, P. (2013). *Some emerging principles for mobile-assisted language learning*. Monterey, CA: The International Research Foundation for English Language Education.
- Voelkel, S., & Bennett, D. (2013). Combining the formative with the summative: The development of a two-stage online test to encourage engagement and provide personal feedback in large classes. *Research in Learning Technology*, 21(1), 75–92.

KEY TERMS AND DEFINITIONS

Enquiry-Based Approach: a learner-centered approach that emphasizes higher order thinking skills.

Formative Assessment: monitors student learning to provide ongoing feedback that can be used by instructors to improve their teaching and by students to improve their learning.

Interactive Learning: is a pedagogical approach that incorporates social networking into course design and delivery.

MALL: mobile assisted language learning.

Mobile Learning: learning methods and materials that involve the use of mobile phones or handheld devices.

Mobile Testing System: mobile apps design specifically for testing skills and knowledge.

Summative Assessment: evaluates student learning at the end of an instructional unit by comparing it against some standard or benchmark.

Nominalizations in Requirements Engineering Natural Language Models



Claudia S. Litvak

Universidad Nacional de La Matanza, Argentina & Universidad Nacional del Oeste, Argentina

Graciela Dora Susana Hadad

Universidad Nacional del Oeste, Argentina

Jorge Horacio Doorn

Universidad Nacional del Oeste, Argentina & Universidad Nacional de La Matanza, Argentina

INTRODUCTION

The first activity of the development of a software system is to define its requirements. Requirements engineer must interact mainly with clients and users among other stakeholders (Macaulay, 1993). He or she has to understand the context in which the future system will act and he or she must carefully consider the reasons that conducted to the decision of developing such system. Requirements engineer's responsibilities include establishing a fluid communication with all stakeholders to produce reliable documents that will be used later in the software development process (Leite et al., 2004). Usually the culture, knowledge and skills of clients and users are rather different from those of the software development experts. As part of the communication with clients and users, the requirements engineer must clearly show them the characteristics of the software system that he or she is conceiving to attend clients and users' issues. Requirements Engineering process has two main activities: to understand the application domain, and to correctly define the services that the future software system will provide (Leite et al., 2004). As a general practice, both activities involve the development of models that describe such domain. It is also a common practice to develop such models in natural language to enhance the communication among stakeholders (Rolland & Ben Achour, 1998; Leite et al., 2004; Seiff et al.,

2009); however, this introduces some obstacles, such as ambiguity, incompleteness and poor information structuring (Zowghi & Gervasi, 2002; Berry & Kamsties, 2004; Leite et al., 2005; Doorn & Rida, 2009; Hadad et al., 2015). All of these inconveniences come from the natural language itself. As a consequence, Requirements Engineering has become more and more involved with linguistic considerations. Furthermore, it should be kept in mind that language conveys culture and knowledge (Nettle & Romaine, 2000; Fishman, 1999). Thereby, the terminology of clients and users holds application domain knowledge. Therefore, to document and to slightly formalize the relevant words or phrases heard from clients and users or read from documents is a valuable practice. In other words, creating a glossary of such terminology helps the Requirements Engineering process in two relevant ways: it eases the understanding of the application domain and it reduces the ambiguity of the oral communication with clients and users and the ambiguity of every produced document (Hadad, Doorn, & Kaplan, 2009).

However, the glossary construction itself introduces some drawbacks. The most important of these drawbacks, not yet treated, is the presence of nominalizations, either in the clients and users' terminology or in the produced glossary. The former is a possible source of ambiguity while the latter is a hint of an adequate or inadequate

DOI: 10.4018/978-1-5225-2255-3.ch445

creation of glossary symbols by the requirements engineer. Both may cause defects in the glossary. Nominalization refers to the construction of nouns from verbs or adjectives. Linguistic authors have largely studied nominalization in many languages such as English, French, German, Russian, Spanish, etc. (Alexiadou, 2001; Bisetto & Melloni, 2005; Grimshaw, 1990; Rothmayr, 2009; Rozwadowska, 1997). The simplest way to describe verb nominalization is by means of the phrase *action of and effect of*. In some cases, nominalization occurs only by *action of*; while in others only occurs by *effect of*.

In this chapter, the influence of nominalization on the quality of Requirements Engineering documents is analyzed. The requirements engineer should be aware of the substantial differences in meaning, that sometimes arise when using the nominal mode of a verb or its verbal mode, since the action and the effect of a verb nominalization may produce synonyms or even homonyms.

BACKGROUND

The work presented in this chapter is based on the lessons learned in several research projects where many study cases were created using Scenarios and Language Extended Lexicon (LEL) models (Leite et al., 1997; Leite et al., 2000; Leite et al., 2004). LEL is a glossary proposed by Leite and Franco (1993). Scenarios are structured natural language descriptions of situations that occur in the application domain (Leite et al., 2000).

The LEL is itself a glossary with roles and structure different from the usual ones. It is composed by a set of symbols, which are words or phrases peculiar and frequently used in the application domain. Each symbol is identified by a name or names. An acronym or an abbreviation may be also a name of a term, only if present in the application domain. In case of synonyms the more relevant name is used as the main key entry. Every symbol has two types of descriptions; this particular structure makes the difference with

other glossaries. The first type, called Notion, is the usual one and describes the denotation of the word or phrase, that is, it defines what the symbol is. The second, called Behavioral Response, describes the connotation of the word or phrase, that is, it describes how the symbol acts in the application domain; this description is not usually present in other glossaries and enriches the knowledge about the symbol and the context at hand. LEL symbols contain hypertext links pointing to directly related entries.

When the denotation of the term acquires several meanings, it indicates the existence of homonyms, which forces the creation of more than one entry in the lexicon. The absence of any behavioral response indicates that the symbol does not belong to the LEL.

LEL entries are classified in four types according to its general use in the application domain. The types are: Subject, Object, Verb and State. Table 1 shows the LEL model.

Symbols of type object, verb and state may be affected by a linguistic transformation, called nominalization, which may hide the type of the symbol and may produce confusions between verbs and objects, or between verbs and states. Nominalization is a source of ambiguities, especially in behavioral responses. It has not received the necessary attention in the Requirements Engineering literature (Berry & Kamsties, 2004; Kovitz,

Table 1. Language EXTENDED LEXICON model

LEL: representation of the symbols in the application domain language. Syntax: {Symbol} _i ^N
Symbol: entry of the lexicon that has a special meaning in the application domain. Syntax: {Name} _i ^N + {Notion} _i ^N + {Behavioral Response} _i ^N
Name: identification of the symbol. More than one represents synonyms. Syntax: Word Phrase
Notion: denotation of the symbol. Syntax: Sentence
Behavioral Response: connotation of the symbol. Syntax: Sentence

Source: (Leite et al., 2000)

1998). This chapter is devoted to analyze and control the influence of nominalization on the quality of Requirements Engineering documents. Proposals to reduce the possible negative influence of nominalization are described.

The scenario model is a structure composed of the entities: title, goal, context, resources, actors, episodes, exceptions and the attribute constraint (see Table 2). Actors and resources are an enumeration. Title, goal, context and exceptions are declarative sentences while episodes are a set of sentences expressed in a simple language that give an operational description of behavior. A scenario must satisfy a goal that is reached by performing its episodes. Episodes represent the main course of action but they include also variations or possible alternatives. While performing episodes an exception may arise, signaling an obstacle to goal achievement. The treatment of the exception may or may not satisfy the original goal. All descriptions in the Scenario must maximize the use of LEL symbols, as a way to reduce ambiguity. This is another reason that strengthens the necessity of a high quality LEL creation.

VERB NOMINALIZATIONS

Verb nominalization may be interpreted as *eventive* with the semantic of *action of* and *stative* when it takes the semantic of *effect of*. Eventive nominalizations denote an activity and they are frequently combined with indications of location, duration or date. Eventive nominalizations inherit the arguments of the verbal base in such a way that they are capable of expressing the same semantic roles that the verb itself. Stative nominalizations are related with the result of the action of the verb. In most cases this result is an entity or a state. The following sentences contain the nominalization of the verb *publish*. The first sentence contains an eventive nominalization and the second a stative one.

- The software system must register the number of copies of every *publication*.
- The *publication* of books, and even of journals, requires resources that have to be available on time in order to minimize costs.

Ambiguity, completeness, and structuring of natural language documents are related among them. A slightly formalized glossary contributes to improve completeness since it contains a set of application domain notions materialized through subjects, objects, activities and states and their interrelations. The perception by the requirements engineer of any of those glossary elements helps to discover more information about the related glossary symbols. This clearly contributes to reduce incompleteness.

The experience acquired along many study cases led to observe the weaknesses that occur by nominalization of verb symbols. The semantic flaws produced due to nominalizations are of such nature that sometimes is quite different to register the verb itself or its nominalization in the LEL. The actual alternatives of verb nominalization may be:

1. Verb and its nominalization have exactly the same meaning.
2. Verb and its nominalization have different meaning.
 - a. Just verb semantic is relevant in the application domain.
 - b. Just nominalization semantic is relevant in the application domain.
 - c. Both verb and its nominalization semantic are relevant in the application domain.
3. There is more than one nominalization for the same verb.

The purpose of the LEL is to represent the vocabulary used in the application domain, therefore, nominalizations could appear or not as part of



Table 2. Scenario model

Scenario: description of a situation in the application domain. Syntax: Title + Goal + Context + {Resources}₁^N + {Actors}₁^N + {Episodes}₂^N + {Exceptions}
Title: identification of the scenario. In the case of a sub-scenario, the title is the same as the episode sentence (see below in the Episode definition), without the constraints. Syntax: Phrase ([Actor Resource] + Verb + Predicate)
Goal: aim to be reached in the application domain. The scenario describes the achievement of the goal. Syntax: [Actor Resource] + Verb + Predicate
Context: composed by at least one of the following sub-components: Geographical Location: physical set of the scenario. Temporal Location: time specification for the scenario development. Precondition: initial state of the scenario. Syntax: {Geographical Location} + {Temporal Location} + {Precondition} where Geographical Location is: Phrase + {Constraint} where Temporal Location is: Phrase + {Constraint} where Precondition is: [Subject Actor Resource] + Verb + Predicate + {Constraint}
Resources: relevant physical elements or information that must be available in the scenario. Syntax: Name + {Constraint}
Actors: persons, devices or organisation structures that have a role in the scenario. Syntax: Name
Episodes: set of actions that details the scenario and provides its behaviour. An episode can also be described as a scenario. Syntax (using partial BNF): <episodes>::= <group series> <episode series> <group series>::= <group> <group> <non-sequential group> <group series> <group> <group>::= <sequential group> <non-sequential group> <sequential group>::= <basic sentence> <sequential group> <basic sentence> <non-sequential group>::= # <episode series> # <episode series>::= <basic sentence> <basic sentence> <episode series> <basic sentence> <basic sentence>::= <simple sentence> <conditional sentence> <optional sentence> <simple sentence>::= <episode sentence> CR <conditional sentence>::= IF <condition> THEN <episode sentence> CR <optional sentence>::= [<episode sentence>] CR where <episode sentence> is described: ((([Actor Resource] + Verb + Predicate) ([Actor Resource] + [Verb] + Title)) + {Constraint})
Exceptions: usually reflect the lack or malfunction of a necessary resource. An exception hinders the achievement of the scenario goal. The treatment of the exception may be expressed through other scenario. Syntax: Cause [(Solution)] where Cause is: Phrase ([Subject Actor Resource] + Verb + Predicate) where Solution is: Title
Constraint: a scope or quality requirement referring to a given entity. It is an attribute of Resources, basic Episodes or sub-components of Context. Syntax: ([Subject Actor Resource] + Must [Not] + Verb + Predicate) Phrase
+ means composition, {x} means zero or more occurrences of x, () is used for grouping, stands for or and [x] denotes that x is optional

that vocabulary. Nominalizations are not a problem provoked by the Requirements Engineering process or by the requirements engineer himself or herself. It is an obstacle that he or she must manage by identifying the variants of the term

and by establishing the meaning of those variants in the application domain. When capturing the vocabulary used in the application domain, it could happen and it usually happens that the requirements engineer does not perceive one or

more of the possible meanings of a nominalized verb. Once this happens, it is very likely that he or she describes only the captured meaning keeping all the omissions undiscovered.

SOLUTIONS AND RECOMMENDATIONS

Each one of the cases enumerated in the previous section should be treated in a particular way, as shown in Table 3. When the requirements engineer runs across with a verb symbol, he or she does not know in which case the verb falls. To avoid omissions or ambiguities he or she has to find out how the symbol is used in the application domain, and then to which case it belongs in order to take out the correct action.

Table 3 strictly considers the possible consequences of the use of nominalizations without paying attention to the particular semantic of the verb itself or its nominalizations. Actually a verb or a nominalization may have more than one interpretation in the application domain, producing eventually homonyms. Therefore, any of the cells of third column of Table 2 may produce more symbols than it is indicated. Watching how the nominalizations are used is not enough to detect such additional homonyms cases. This happens with any single symbol of the LEL, regardless if it is the nominalization of a verb or not. The particular semantic of each symbol should be clearly understood in order to discover hidden or partially hidden homonyms, and possible synonyms.

Table 4 shows an example about the Study Case Production Management of Cardboard Boxes. In

Table 3. Guidelines to manage verbs and nominalizations

Syntactic Form	Meaning	Action
Used as a verb	Action of	Register as a verb symbol
	Effect of	N/A
	Both	N/A
Used as a nominalization	Action of	Register as a verb symbol
	Effect of	Register as an object or state symbol
	Both	Register both as a verb symbol and as an object or state symbol. Label both as homonyms
Used as a nominalization with more than one nominalization	Action of	Register each of them as a verb symbols. Consider that eventually any nominalization may be a synonym of another
	Effect of	Register each of them as object or state symbols. Consider that eventually any nominalization may be a synonym of another
	Both	Register as a verb symbol and as an object or state symbols. Consider that eventually any nominalization may be a synonym of another
Used both forms	Action of	Register both forms as a verb symbol, choose the most used form as key entry name and the least used as synonym (see Table 3)
	Effect of	N/A
	Both	Register both as a verb symbol and as an object or state symbol (see Table 4). Consider that the nominalization may be also a synonym of the verb form, then label both entries as homonyms
Used both forms with more than one nominalization	Action of	Register all forms as verb symbols. Consider that any nominalization may be a synonym of the verb form; choose the most used form as the key entry name and the other as synonym. Label homonyms if applicable
	Effect of	N/A
	Both	Register as a verb symbol and as object or state symbols. Consider that one of the nominalizations may be a synonym of the verb form; choose the most used form as the key entry name and the other as synonym. Consider that eventually any nominalization may be a synonym of another. Label homonyms if applicable (see Table 5)

Table 4. Verb and its nominalization used as action of, being both synonyms; Study Case: Production management of cardboard boxes

Symbol Name/s	Plan Production / Production Planning	Type: Verb
Notion	• Process to establish the <i>production schedule</i> for all the shifts of the next week. • It is performed every Tuesday.	
Behavioral Response	• <i>Manufacturing programs</i> for each of the 21 shifts per week are generated. • Deployments boxes are organized to <i>reduce cardboard scrap</i>. • <i>Production orders</i> received from a <i>collaborative supplier</i> are considered <i>urgent orders</i>.	

this example, the requirements engineer detects two different synonyms, used both in that context, for the same action: Plan Production and Production Planning.

In Table 5 there are two different symbols, one is a Verb symbol, Pay, and the other is an Object symbol, Payment, belonging to the same study case, Management of Treatments given by Medical Providers, where the Object refers to the verb nominalization used as effect of.

In Table 6 the symbol Authorize Treatment is a Verb, and there are three different nominalizations. The first one, Treatment Authorization, is a synonym of Authorize Treatment, being both verbs. Treatment Authorization as a Verb is also the homonym of Treatment Authorization as an Object. The third nominalization has generated a State symbol Authorized Treatment.

FUTURE RESEARCH DIRECTIONS

This chapter presents the results of an exploratory research which has allowed discovering, as a regular pattern, the presence of many nominalizations of verbs that are relevant in the application domain. Frequently, such nominalizations have created different types of defects in the glossaries. The knowledge of the existence of those defects and the understanding of their cause has allowed proposing guidelines in order to reduce quantitative and qualitative consequences of such defects. These guidelines have been proven to be effective in few and small study cases. It is planned to perform a set of more structured experiments to quantify the efficiency of the proposed guidelines and to obtain reliable statistics and metrics of the quality enhancement of the glossaries and the importance of this on the whole Requirements Engineering process.

Table 5. Verb and its nominalization used as effect of; Study Case: Management of treatments given by medical providers

Symbol Name/s	Pay	Type: Verb
Notion	• Action performed by the <i>treasury staff</i> producing an economic compensation to a <i>medical provider</i>. • It takes place after the <i>payment</i> has been approved by the <i>medical director</i>.	
Behavioral Response	• The <i>treasury staff</i> issues the <i>payment</i> thru the <i>system</i>. • The <i>treasury staff</i> notifies the <i>medical provider</i> upon the issued <i>payment</i>.	
Symbol Name/s	Payment	Type: Object
Notion	• Document issued by the <i>treasury staff</i>, establishing the cancellation of a debt acquired due to a <i>treatment</i> given by a <i>medical provider</i>.	
Behavioral Response	• It is registered in the system. • It requires the <i>director</i> approval before given to the <i>medical provider</i>. • A copy signed by the <i>medical provider</i> is stored.	

Table 6. Verb and two nominalizations used as action of and effect of, producing synonyms and homonyms; Study Case: Management of treatments given by medical providers

Symbol Name/s	Authorize Treatment / Treatment Authorization	Type: Verb
Notion	• Action that allows the <i>patient</i> to receive <i>treatment</i>. • It is performed at the <i>patient service office</i>.	
Behavioral Response	• The staff checks the <i>medical credit</i> of the <i>patient</i> thru the <i>system</i>. • The staff checks the <i>clinical order</i> given by the <i>patient</i>. • The staff issues an <i>authorization</i>. • The staff delivers the <i>authorization</i> to the <i>patient</i>. • The <i>patient</i> annexes the <i>clinical order</i> to the <i>authorization form</i> in order to get its <i>treatment</i>.	
Symbol Name/s	Treatment Authorization / Authorization	Type: Object
Notion	• Document issued by the <i>patient service staff</i>, allowing a <i>patient</i> to get a <i>treatment</i>.	
Behavioral Response	• It is generated thru the <i>system</i> based on the <i>clinical order</i> exhibited by the <i>patient</i>. • It is signed by the <i>service director</i>. • It is assigned a sequential number and an expiration date.	
Symbol Name/s	Authorized Treatment	Type: State
Notion	• Condition that a <i>treatment</i> takes after <i>treatment authorization</i>.	
Behavioral Response	• The <i>patient</i> can get a <i>treatment</i> thru a <i>medical provider</i> before its expiration date. • After the <i>patient</i> gets the <i>treatment</i>, the <i>medical provider</i> issues an <i>invoice</i> to receive the <i>payment</i>.	

CONCLUSION

The research whose results are reported in this chapter may be summarized as:

- The relevant words and phrases of any application domain include several verbs.
- The full understanding of the verbs is important for improving the knowledge about the application domain.
- Nominalization of verbs obscures the perception of their eventive or stative semantic.
- Frequently one of the possible verb nominalizations hides the perception of the others.
- Losing an eventive or a stative meaning prevents discovering important information.

REFERENCES

- Alexiadou, A. (2001). *Functional Structure in Nominals. Nominalization and ergativity*. Amsterdam, The Netherlands: John Benjamins Publishers; doi:10.1075/la.42
- Berry, D. M., & Kamsties, E. (2004). Ambiguity in requirements specification. In *Perspectives on software requirements* (pp. 7-44). Springer US. doi:10.1007/978-1-4615-0465-8_2
- Bisetto, A., & Melloni, C. (2005, September). Result Nominals: A Lexical-Semantic Investigation. In *On-line Proceedings of the Fifth Mediterranean Morphology Meeting (MMM5)* (pp. 393-412). Bologna: University of Bologna.

- Doorn, J. H., & Ridao, M. (2009). Completeness Concerns in Requirements Engineering. In M. Khosrow-Pour (Ed.), *Encyclopedia of Information Science and Technology* (2nd ed., pp. 619–624). Hershey, PA: IGI Global, Information Science Reference; doi:10.4018/978-1-60566-026-4.ch101
- Fishman, J. (1999). *Handbook of Language and Ethnic Identity*. Oxford, UK: Oxford University press.
- Grimshaw, J. (1990). *Argument structure*. Cambridge, MA: MIT Press.
- Hadad, G. D., Doorn, J. H., & Kaplan, G. N. (2009). Creating Software System Context Glossaries. In M. Khosrow-Pour (Ed.), *Encyclopedia of Information Science and Technology* (2nd ed., Vol. 2, pp. 789–794). Hershey, PA: IGI Global, Information Science Reference; doi:10.4018/978-1-60566-026-4.ch128
- Hadad, G. D. S., Litvak, C. S., Doorn, J. H., & Ridao, M. N. (2015). Dealing with Completeness in Requirements Engineering. In M. Khosrow-Pour (Ed.), *Encyclopedia of Information Science and Technology* (3rd ed., pp. 2854–2863). Hershey, PA: IGI Global, Information Science Reference; doi:10.4018/978-1-4666-5888-2.ch279
- Kovitz, B. L. (1998). *Practical software requirements: A manual of content and style*. Greenwich: Manning.
- Leite, J. C. S. P., Doorn, J. H., Hadad, G. D., & Kaplan, G. N. (2005). Scenario inspections. *Requirements Engineering*, 10(1), 1–21. doi:10.1007/s00766-003-0186-9
- Leite, J. C. S. P., Doorn, J. H., Kaplan, G. N., Hadad, G. D., & Ridao, M. N. (2004). Defining System Context using Scenarios. In J. S. C. P. Leite & J. H. Doorn (Eds.), *Perspectives on Software Requirements* (pp. 169–199). Norwell, MA: Kluwer Academic Press; doi:10.1007/978-1-4615-0465-8_8
- Leite, J. C. S. P., & Franco, A. P. M. (1993, January). A strategy for conceptual model acquisition. In *Proceedings of IEEE International Symposium on Requirements Engineering* (pp. 243–246). San Diego, CA: IEEE Computer Society Press.
- Leite, J. C. S. P., Hadad, G. D., Doorn, J. H., & Kaplan, G. N. (2000). A Scenario Construction Process. *Requirements Engineering*, 5(1), 38–61. doi:10.1007/PL00010342
- Leite, J. C. S. P., Rossi, G., Balaguer, F., Maiorana, V., Kaplan, G. N., Hadad, G. D. S., & Oliveros, A. (1997). Enhancing a Requirements Baseline with Scenarios. *Requirements Engineering Journal*, 2(4), 184–198. doi:10.1007/BF02745371
- Macaulay, L. (1993). Requirements capture as a cooperative activity. In *Proceedings of IEEE International Symposium on Requirements Engineering* (pp. 174–181). San Diego, CA: IEEE Computer Society Press.
- Nettle, D., & Romaine, S. (2000). *Vanishing Voices*. New York: Oxford University Press.
- Rolland, C., & Ben Achour, C. (1998). Guiding the construction of textual use case specifications. *Data & Knowledge Engineering*, 25(1-2), 125–160. doi:10.1016/S0169-023X(97)86223-4
- Rothmayr, A. (2009). *The structure of stative verbs* (Vol. 143). John Benjamins Publishing; doi:10.1075/la.143
- Rozwadowska, B. (1997). *Towards a unified theory of nominalizations: External and internal eventualities* (No. 2031). Wdawn. Uniwersytetu Wrocławskiego.
- Seyff, N., Maiden, N., Karlsen, K., Lockerbie, J., Grünbacher, P., Graf, F., & Ncube, C. (2009). Exploring how to use scenarios to discover requirements. *Requirements Engineering*, 14(2), 91–111. doi:10.1007/s00766-009-0077-9

Zowghi, D., & Gervasi, V. (2002). The Three Cs of Requirements: Consistency, Completeness, and Correctness. In *Proceedings of 8th International Workshop on Requirements Engineering: Foundation for Software Quality* (pp. 155-164). Essen, Germany: Essener Informatik Beitiage.

ADDITIONAL READING

Alexiadou, A. (2011). *The aspectual properties of nominalization structures*. Amsterdam: John Benjamins; doi:10.1075/la.178.11ale

Alexiadou, A., & Rathert, M. (Eds.). (2010). *The syntax of nominalizations across languages and frameworks* (Vol. 23). Berlin: Walter de Gruyter; doi:10.1515/9783110245875

Browne, G. J., Pitts, M. G., & Wetherbe, J. C. (2007). Cognitive stopping rules for terminating information search in online tasks. *Management Information Systems Quarterly*, 31(1), 89–104.

Firesmith, D. (2005). Are Your Requirements Complete? *Journal of Object Technology*, 4(1), 27–43. doi:10.5381/jot.2005.4.1.c3

Giannakidou, A., & Rathert, M. (Eds.). (2009). *Quantification, definiteness, and nominalization* (Vol. 24). Oxford: Oxford University press.

Koptjevskaja-Tamm, M. (2002). *Nominalizations*. New York: Routledge.

Pustejovsky, J. (Ed.). (2012). *Semantics and the Lexicon* (Vol. 49). Dordrecht: Springer Science & Business Media.

Rathert, M., & Alexiadou, A. (Eds.). (2010). *The semantics of nominalizations across languages and frameworks* (Vol. 22). Berlin: Walter de Gruyter; doi:10.1515/9783110226546

Sager, N. (1981). *Natural language information processing*. Massachusetts: Addison-Wesley Publishing Company, Advanced Book Program.

Zucchi, A. (2013). *The language of propositions and events: Issues in the syntax and the semantics of nominalization* (Vol. 51). Dordrecht: Springer Science & Business Media.

L

KEY TERMS AND DEFINITIONS

Ambiguity: Phrase or sentence that does not have a single clear meaning. It provokes two or more possible interpretations.

Application Domain: The context in which the software system will be used.

Completeness: The state or condition of having all the necessary or appropriate parts of a model or document.

Language Extended Lexicon: A glossary composed by a set of symbols, which are words or phrases peculiar and frequently used in a given application domain. The lexicon symbols have an additional description, not present in other glossaries. It contains hypertext links that interconnect symbols.

Natural Language Model: A model in narrative text produced during Requirements Engineering process, with a structure and content easily read by all stakeholders.

Nominalization: The result of forming a noun from a verb, or the act of forming a noun from a verb. Nominalization itself is the nominalization of the verb *nominalize*. In this chapter it is used as a verb and as the effect of the verb (a state).

Omission: Information that has not been included in a model or document. It is the nominalization of the verb *omit*. In this chapter it is used as the effect of the verb (a state).

Scenario: The description of a situation that currently happens or will happen in the application domain. The former are called Current Scenarios and the latter are referred as Future Scenarios.

Word Formation Study in Developing Naming Guidelines in the Translation of English Medical Terms Into Persian

Ali Akbar Zeinali

Universiti Sains Malaysia, Malaysia

INTRODUCTION

The medical translation is a critical tool for communication between patients and health care professionals. A correct translation finds itself significant when a medical translator, as a skilled mediator, acts accurately and precisely by transferring the messages between a medical professional and a patient. This study is of utmost importance as it aims to find guidelines specific to translation procedures of English medical terms into Persian through word formation processes for linguists and translators. Such guidelines can be employed by translators to find the required equivalents. Persian linguists from the Persian Language and Literature Academy (PLLA) and Medical Science Academy in Iran may also find it beneficial. It may be effective and applicable to information technology and machine translation systems linguistic databases. This study presents the findings obtained by a comparative analysis of the Persian translation equivalents found in “اهیرامی بی لملل نی ب یراذگدک یامنهار” (Guide to ICD-9-CM in Persian) using the universal naming guidelines and local naming principles in Persian, with regard to their pairs in English selected from medical terms in “ICD-9-CM.” The study went through the research questions to investigate the naming characteristics of the English-Persian medical terms with respect to Sager’s criteria (1990, p. 88) who provides 12 criteria for the perfect idealized requirement in a serious controlled condition for naming, the PLLA principles, the effectiveness of the adopted TPs in naming, the contribution of the morphosemantic

DOI: 10.4018/978-1-5225-2255-3.ch446

factors to the naming process and to propose the specific naming parameters for the translation of English-Persian medical terms. The objectives of the study were also in line with the research questions. The analysis focused on Sager’s naming criteria, the PLLA naming principles, and the applied translation procedures on the selected English medical terms into Persian equivalents during the translation process.

BACKGROUND

According to Ashuri (1995, p.29), one of the problems for Persian today is the fact that it is a combination of ancient writings and translation works. Scientific language of new works mostly deals with translation together with a lot of English and French words and syntactical structures. He suggests that the word formation is a solution for enriching a language with new concepts.

Beheshti (1999, pp.25-31) explains that one of the most significant contemporary linguistic issues is scientific-technical word formation, which is based on the language grammar and linguistic principles. She discovered that only 50 percent of the total terms studied had the equivalences in Persian; thus, indicating that the translators have not yet shown interest in employing Persian equivalents in their works. Further investigation is crucial to uncover the underlying reason. According to her, equivalent findings or naming of imported medical terms should be based on the features specific to medical terms. This means that medical terms, either in the source language

or the target language, should be studied to find their systematic characteristics and some patterns in order to help the translators or linguists in the word formatting process or naming imported terms in the future. She suggests a study on the patterns based on the term characteristics of morphology, etymology, word formation and translation procedures.

Naseri et al. (2011, pp.41- 47) believe that the present problems of Persian in the areas of science and technology are due to the application of foreign language structures, lack of consistencies in scientific terms, and no consensus among authors and translators. Sadeqi (1993) believes that the solution to these problems is stabilizing the scientific and technical terminology at the basic levels and mass media. He states that the scientific terminologies are not consistent and in some cases, there are various Persian equivalents for a single foreign word in different dictionaries, or there are various foreign equivalents for a single Persian word. As a result, it is claimed that the scientific expressions are inconsistent (Kafi, 1992). According to Beheshti (1999, p.31), a language that borrows a large number of words suffers from negative consequences in the general language and can denigrate the capacity for people to speak their native language (i.e. such as German language in interaction with the English language). The Persian language in interaction with other languages will lose its capabilities and will be converted to a totally different language if it is not protected against the foreign words imported from many languages. Therefore, activating the Persian language capabilities for naming the scientific and non-scientific foreign words is considered as a necessity which should be based on definite principles or patterns.

MAIN FOCUS

It is of the interest of this study to investigate what types of universal naming criteria and principles are followed by the selected data. While naming

process should follow the international naming guidelines, the state or local naming principles may not be skipped, as creation of an equivalent in the target language should be lexically and semantically based on the cultural, terminological, linguistic and language factors of the target language.

The naming criteria suggested by Sager (1990, p.88) provides guidance on the creation of terms to be referred to when the translator deals with the difficulties in coining translated terms because of different structures and term formation techniques found in various languages. It remains difficult to generalize on a broad level. Sager provides 12 criteria for the perfect idealized requirement in a serious controlled condition for naming. "Criterion" is abbreviated as "C" in this study and it is followed by a numerical digit showing the number of criterion suggested by Sager. For example, "C1" represents the first criterion Sager has proposed for term formation. These criteria are outlined as below:

- C1. The term must relate directly to the concept.
It must express the concept clearly. A logical construction is advisable.
- C2. The term must be lexically systematic. It must follow an existing lexical pattern and if the words are of foreign origin, a uniform transcription must be preserved.
- C3. The term must conform to the general rules of word formation of the language which will also dictate the word order in compounds and phrases.
- C4. Term should be capable of providing derivatives.
- C5. Terms should not be pleonastic (i.e. no redundant repetition, e.g. combining a foreign word with a native word having the same meaning).
- C6. Without sacrificing precision, terms should be concise and not contain unnecessary information.
- C7. There should be no synonyms whether absolute, relative or apparent.



- C8. Terms should not have morphological variants.
- C9. Terms should not have homonyms.
- C10. Terms should be monosemic.
- C11. The content of the terms should be precise and not overlap in meaning with other terms.
- C12. The meaning of the terms should be independent of context.

In this study, Persian word formation structures will be processed according to the Naming Guidelines approved by the Persian Language and Literature Academy (2009). These principles, obligatory for the word formation teams, have been approved by PLLA, as word designation and equivalent findings for the imported terms are the concerns of the Academy according to Article 2 and 3 of the second provision of the Islamic Republic of Iran constitution. "Principle" is abbreviated by "P" which is followed by a numerical digit showing the Principle number approved by PLLA. For example, "P2" is the second principle which is proposed by PLLA for naming. PLLA provides nine principles for naming which are outlined as below:

- P1. Persian words should be based on "basic Persian language today" i.e. the common language used by the learnt societies.
- P2. Word designation should adhere to the correct grammar structure of the Persian language.
- P3. Word designation should avoid dissonance and cacophony and should be based on Persian phonetics. As a result, the designated word, as much as possible, should be shorter than its pair in the source language.
- P4. Equivalents with derivational and inflectional capabilities and also words with noun, adjective and verb possibilities are required to determine word designation priorities.
- P5. Word designation should consider the following hierarchy as criteria:
 - P5.1 Older common and accustomed words in the Persian language

- P5.2 Neologisms according to Persian word formation processes using Persian words
- P5.3 Arabic words, common and accustomed in the Persian language
- P5.4 Neologisms according to Persian word formation processes using Arabic words common in the Persian language
- P5.5 Words borrowed from several Persian language dialects in Iran today
- P5.6 Words borrowed from medieval and ancient Persian Languages
- P6. In word designation, the priority is given to the words with clear meaning compared to the ones with ambiguous meaning.
- P7. In word designation, especially in scientific areas, one word should be designed for any word with special meaning regardless of any probable diversity.
 - Nota bene: Designation of multiple equivalents is applicable for any foreign word with several scientific definitions, according to scientific tradition, history and convention.
- P8. No equivalent is needed for foreign words that are universally known and accepted.
- P9. The Language Academy Council will make the final decision in case of failure in word designation process based on common Persian language patterns.

Based on a comparison between the Persian medical equivalents with Sager's naming criteria and the PLLA naming principles, the goal of the study is to determine the similarities and differences of the compatible and incompatible terms (Persian equivalents) with respect to the applied translation procedures and the applied word formation processes and to identify those with the highest number of frequencies.

The compatible and incompatible equivalents with Sager's criteria and the PLLA Principles provide the data for this study and enable statistical descriptive and qualitative analysis to be carried

out. The collected data included a population of 339 Persian medical terms from the target text with regard to their pairs from the source text, which is sufficient with respect to the statistical descriptive method of analysis. It should, however, be noted that the selected terms reflected only the results limited to the selected system of the human body (the musculoskeletal system and connective tissue). Furthermore, the results obtained in this study do not necessarily reflect all the guidelines available for the translation of all the medical terms of other human body systems mentioned in ICD-9-CM into Persian.

This study establishes a conclusion about the integration of the findings of the study. It begins with a discussion of inferences drawn from the study and is followed by a comparative description of the findings and recommendations.

Data Analysis and Methods

The study needs preliminary information for analyzing the data including grammar, morphology, etymology and syntax. The Persian term will be broken down into components or morphemes based on the Katamba morphology (1994) and Tabatabaee's structures (2009). The analysis will focus on the characteristics of the Persian terms in accordance to the naming guidelines provided by ISO and criteria suggested by Sager (1990, p.88). The terms will subsequently be processed based on the principles recommended by the PLLA.

This is designed to uncover the fact that those English medical terms with certain characteristics can be translated through certain translation procedures and word formation processes and are compatible with certain Persian word formation patterns. It should be noted that in this research, a positive or compatible equivalent indicates the positive compatibility of the equivalent with all the PLLA principles and Sager's criteria; and a negative or incompatible equivalent indicates negative

or no compatibility of the equivalent with at least one of the PLLA principles or Sager's criteria. The conclusion will be uncovering the effective naming parameters or patterns for translating the medical terms from English into Persian.

The analysis of each term from the source language and its pair in the target language is performed according to the arranged elements of a data table (Table 1). For better explanation, suppose that the term *ازولانی* [azolāni ostoxāni] has been selected among the related terms of the selected system of the human body in the target text.

Among the selected terms, a few have equivalents that are composed of two separate words, one with borrowing and derivation processes and the other with only the derivation process. As in the example, the single Greco-Latin term of "musculoskeletal" (Dorland's Medical Dictionary, 2005) is composed of two free root words of "musculus" and "skeleton" which have been embedded by the derivational Latin suffix "al". Like other examples of this group of terms, the equivalent consists of two separate words, *ازولانی* [azolāni] /azola:nI/ (muscular), which is Arabic, and *استوخانی* [ostoxāni] /ostoxa:nI/ (skeletal), which is Persian. The former matches with the borrowing process, while the latter is a native word in the target language. Both of them indicate derivation in their word formation processes as well. The suffix "i" as an adjectival suffix is observed in both Persian words. In Persian, it is not possible to combine the two words together by adding the suffix only to one of them, so the translator is required to produce two independent adjective words which do not make a phrase. In this example the translation procedure is through translation since "musculoskeletal" is a compound word.

Table 1 shows the analysis of the equivalent produced and it could be concluded that this compound equivalent is not applicable due to its unconformity to Sager's 7 and 8 criteria and



Table 1. Data table

English Term		Musculoskeletal	
Equivalent in Persian		عضلانی و استخوانی (azolāni + ostoxāni)	
Gloss		muscular + skeletal	
Back Translation		muscular-skeletal	
Parts of Speech		adjective	
Morphological Analysis		azolāni [muscular]	ostoxān+i [skeletal]
		root	root + suffix
Parts of Speech		adjective	adjective
Tabatabaee's Persian Structure		balanced structure Adj+ - + Adj. = Adj	
Morphosyntactic Structure		Adj + Adj	
Word Formation		(borrowing, derivation) + derivation	
Translation Procedure		through translation + substitution	
Sager's Criteria		PLLA Principles	
Result		Result	

mismatches with principle 7 in naming Principles of PLLA. However, it indicates that there is a compatible structure with the Balanced Structure of Tabatabaee's structures i.e. adj + - + adj.

The problem with the above example is multiplicity. With respect to Sager's criteria 7 and 8, it is found that another synonym like [azole i] "عضلانی" (muscular) can be used that follows the Persian structure, i.e. [azole] (root) + i (adjectival suffix), while [azole] "عضله" (muscle) is borrowed from Arabic as an independent word. On the other hand, [azolāni] "عضلانی" is also imported from Arabic as an independent adjective. Since [azolāni] is an Arabic adjective, it cannot be matched with any Persian structure, because there is no [āni] suffix in Persian structures that can combine the borrowed noun [azole]. Thus, no derivability can be found for [azolāni] as it is an adjective imported directly from another language. It violates principle 7 in the naming PLLA principles.

Findings

The findings of the equivalents under this study can be classified into two groups: compatible

and incompatible equivalents. The compatible terms are the ones which follow all of Sager's criteria together with all of the PLLA principles. These terms or words are compatible with all the requirements for naming. In other words, the compatible terms automatically present those features of translation procedures which are effective for naming the Persian medical terms. The incompatible terms are the ones which do not follow one or more naming criteria suggested by Sager or the naming principles recommended by the PLLA. Such terms or words are incompatible with all or some of the naming requirements for naming. In other words, the incompatible terms automatically present those features of translation procedures which are ineffective for naming the Persian medical terms. Table 2 indicates the frequency of distribution for these two groups of equivalents under this study.

The study identified that around 67% of the terms are incompatible with the basal naming guidelines employed in the study (Figure 1). Generally, the findings of this study are not limited to the incompatible equivalents, but to the compatible ones also. The compatible equivalents can indicate

Table 2. Frequency and percentage of the compatible and incompatible equivalents

Equivalents	Frequency	Percent
Compatibles	112	33.04
Incompatibles	227	66.96
Total	339	100

the effective translation procedures and naming requirements for the secondary word formation of English medical terms into Persian. According to Sager (1990, pp.80-83), the terms which belong to scientific and technological innovations basically differ from the terms accompanying the transfer of scientific and technological knowledge from one linguistic society to another: “while the former is spontaneous, the latter can be designed and engineered”. Primary term formation is a process starting with concept formation in a scientific area. Such a process is out of external control, and is therefore monolingual and affected by “existing patterns of terms already created”.

The incompatible ones illustrate the incorrect methods which should not be followed in the translation processes; however, they can be changed to compatible ones by employing the guidelines provided through this study. The incompatible

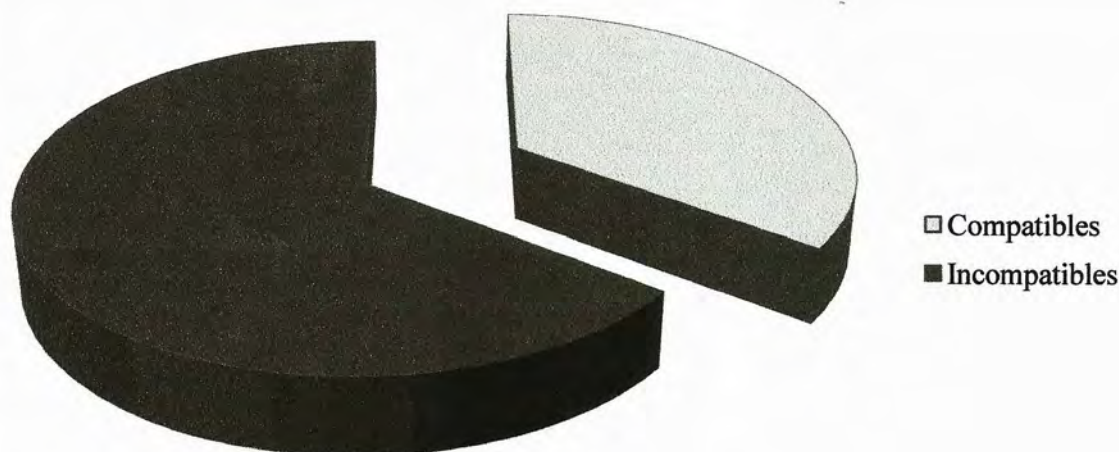
frequencies, as well as the applied translation procedures, are important findings of this research and will be discussed in the next sections.

Compatible Terms or Equivalents

The finding shows that Criteria 10, 11 and 12 (C10, C11 and C12) are the criteria which are fulfilled by the terms in this study. It means that firstly, all the equivalents under this study follow some semantic criteria. Secondly, the respective terms belong to one meaning each and the meaning does not overlap with other terms; and thirdly their meanings can be learnt out of context.

Given the PLLA principles, the compatible principles that resulted in the equivalents in this study are principles 1, 5, 8 and 9 (P1, P5, P8 and P9). Based on the analysis, all equivalents in this study have been formed in the Persian language area today, and they follow the regularities of word designation discussed in P5. Principles 8 and 9 are items suggested by PLLA to help the translators and linguists in case of inquiry. Principle 5 has been classified in lexicology category. As all equivalents in this study are compatible with P5, some of the terms are conflicting with those principles suggested by PLLA in semantic area.

Figure 1. Distribution of the compatible and incompatible equivalents



On the other hand, some equivalents are totally compatible with all Sager's criteria and the PLLA principles. Those terms are formed by through translation procedure of the ones classified by Newmark (1988, p.81) including transference, naturalization, through translation, shift, modulation, recognized translation, literal translation etc. For some terms, substitution and naturalization have been identified as the translation procedures, and they are utilized in the translation process, which lead the equivalents to be compatible with all basal guidelines. It indicates that through translation and naturalization are two effective procedures in translating English medical terms into Persian if they are applied independently and are not combined with any other procedure.

Incompatible Terms or Equivalents

The most incompatible equivalents have been found the ones which do not fulfill the Sager's criteria 4 and 7 together with principle 3 of the PLLA. Such terms, with a frequency of 54 occurrences, consists of 16% of the total terms and 26% of the incompatible terms in this study. The applied translation procedure in these equivalents include (substitution + shift + through translation) with 50 terms and (shift + through translation + naturalization) with 4 terms. It indicates that the majority of incompatibility in the equivalents belongs to (substitution + shift + through translation). Substitution is a translation procedure suggested by C. Heah (1989, p.191) in which borrowing is employed as a word formation process.

The study also found that the next high frequency of incompatibilities is related to the equivalents which do not fulfill criterion 4 and principle 3 with a frequency of 48 terms consisting of 14% of the total terms and 23% of the incompatible terms in this study. The applied translation procedures for these equivalents are (substitution + shift) with 40 terms and shift with 8 terms. It is clear that the second important area which is targeted by incompatibilities in this study is (substitution + shift).

Results

The finding of the current study is as below:

1. The equivalents in this study are more incompatible with the guidelines in the lexicology area compared to the semantics area.
2. With respect to Sager's criteria, around 52% of the equivalents have been constructed without any derivational capability (-C4).
3. With respect to the PLLA Principles, around 56% of the equivalents have translations that are longer than their pairs in the first language (-P7).
4. With respect to the applied translation procedures, most of the equivalents are incompatible with Sager's criteria and those equivalents are constructed by (substitution + shift + through translation) feature of translation procedures.
5. 22% of the terms mentioned in the previous item do not have any derivational capabilities (-C4) and 15% have at least one more or other synonym (-C7).
6. With respect to the applied translation procedures, around 33% of the equivalents constructed by (substitution + shift + through translation) feature of translation procedures are incompatible with the PLLA principles.
7. 22% of the terms mentioned in the previous item have longer translations than their pairs in the first language (-P3). This frequency has been repeated for terms without any derivational capability (-P4) and for those that have not been constructed with special meaning (-P7).
8. The medical Persian equivalents incompatible with C4 can be categorized into three categories:
 - a. Equivalents that have been constructed in a phrase pattern which do not leave any opportunity for derivational discussion. Such equivalents should be made into single nouns in derivational area and not into phrases.

- b. Equivalents that do not have any derivation in the source language term pair; therefore, there is no need to explain them in derivational area.
 - c. Equivalents that have been translated incorrectly by the translators which do not follow the derivational capability.
9. The main problem of incompatibility of Persian medical terms with Sager's criteria or the PLLA principles is due to borrowing and substitution procedures. Therefore, naming through word formation should be done carefully with respect to the compatibility with basal guidelines.

THE RECOMMENDED GUIDELINES SPECIFIC TO THE TRANSLATION OF ENGLISH MEDICAL TERMS INTO PERSIAN

The findings and discussions conducted the study to many guidelines specific to medical translation, providing facilities for English-Persian translation of medical terms. Although the current medical equivalents in Persian, either compatible or incompatible, are largely used in the Persian language society, the following guidelines recommended by this study, will be applicable as the rectification of the existing translation procedures and naming Persian medical equivalents in the future. The guidelines are as follows:

1. Through translation and naturalization are the translation procedures applicable in the translation of English medical terms into Persian if no other translation procedure is involved simultaneously and if the equivalent is a single word and not in a phrase structure.
2. The source language medical term which appears in the form of an adjective or a derived word can be translated into Persian by through translation or naturalization if no other translation procedure is involved simultaneously. In this case, the equivalent

- will be a single word with derivational capability in the form of a noun or adjective.
3. Substitution and through translation can be employed simultaneously to translate an English medical term into Persian if the equivalent is a single word with derivational capability.
 4. Naturalization and through translation procedures can be employed simultaneously in the translation process of an English medical term into Persian if the equivalent is a single word with derivational capability.
 5. In Persian which is a language without any combining form morphemes, the third principle of PLLA cannot be applied in the naming of words or in the word formation process of the medical terms, unless the combinational capability of the Persian language is modified.
 6. Obviously, derivational capability will not be discussed in the naming process of the Persian medical terms if the medical equivalent construction is conducted toward phrasal word formation. Therefore, C4 is not applicable in order to produce such terms or equivalents.
 7. If the (substitution + shift + through translation) feature is employed for the translation of English medical terms into Persian, for the component of the phrase which has been formed by substitution (or borrowing), the translation procedure should adhere to either Sager's criteria or the PLLA principle. Such situations can be observed for single-word English medical terms, while their equivalents are in phrase forms.
 8. If the constructed equivalent is a noun phrase, the translator should consider a single-word synonym as the 'illustrate' equivalent for derivational applications.



IMPLICATIONS

The ever-increasing number of imported words and specific terms into the native language is due to the rapid development of technology and science. It is challenging to find appropriate equivalents for terms or technical words because of the lack of effective translation guidelines applicable in the translation process. Research needs to be conducted to solve the problem. The goal of this study is to propose guidelines specific to translation procedures of English medical terms into the Persian language through term formation processes. If there is a specific guideline for naming the terms, translators will not seek equivalents based on individual preferences and there should be consensus within the field and this will reduce the discrepancy in establishing terminology equivalents. It will also be beneficial for Persian linguists from the PLLA and Medical Science Academy in Iran, as they can finally come across standardized and acceptable equivalents in Persian.

Borrowing several equivalents with different word origins for just one imported word illustrates that none of the available equivalents has been effectively adopted. The problems in scientific and technical language are due to imitating or borrowing language structures and grammar from several languages, which is the result of a lack of fixed or standardized scientific terminology or the disagreement among authors and translators (Sadeqi, 1993, pp.125-6).

Regarding the equivalents incompatible with either Sager's criteria or the PLLA principles, it should be emphasized that the problematic area for the translation process of the English medical terms into Persian has been identified by this study. However, this study highlights that caution needs to be exercised by translators or linguists not to adopt some features of translation procedures.

RECOMMENDATION FOR FURTHER STUDIES

This research focused on the study of one of the human body systems classified by ICD-9-CM. Future students, researchers, linguists and translators should conduct further investigations in other systems of the human body to find more comprehensive or universal results. Indeed, this study is considered a step towards recognizing any unknown problem in the medical translation process. Therefore, the following investigations and studies are recommended:

1. To find basic and general guidelines specific to the translation of English medical terms into Persian, further research is needed to include other human body systems. Comparing the findings and conclusions with the findings and conclusion of this study will allow us to have more reliable and fundamental translation approaches or naming guidelines.
2. Other languages, given their basic language characteristics, may follow this research methodology to identify their problems with translation of the English medical terms into their language, which will allow them to have effective guidelines.
3. The guidelines in this study may be applicable to those languages with characteristics similar to those of the Persian language, which need further investigation by related linguists, terminologists and translators.
4. This research style can be employed for the terms in other technological and scientific fields to find their potential problems with the translation process of the English terms into their equivalents in Persian, which suggests effective guidelines that are unique to the related field.
5. The guidelines in this study may be applicable to other technological and scientific fields, which depend on the source language and the target language (Persian).



6. The applied research methodology, the employed research style and the resulted guidelines in this study may be effective and applicable to information technology and machine translation systems linguistic databases.

CONCLUSION

In conclusion, this research attempts to look at this issue from the naming secondary term formation perspective based on the two basal guidelines and the translation procedures employed in translating medical terms from English into Persian and highlights the compatibility of the equivalents with respect to translation procedures and word formation processes, finding problematic areas and presenting guidelines unique to translation of English medical terms into Persian.

REFERENCES

- Ashuri, D. (1995). *Scientific Inspiration and Scientific Language, Revision of the Persian Language*. Tehran: Markaz Press.
- Beheshti, M. (1999). Etymology and Word Designation in German Language. In Sh. M. Azizi, & M. Qolami (Eds.), *Proceeding of the First Conference on Word Designation and Terminology* (p.27, 36-48). Tehran: Academic Press Center
- Bowker, L. (2006). *Lexicography, Terminology and Translation. Text-Based Studies in Honour of Ingrid Meyer*. Ottawa: The University of Ottawa Press.
- Heah, C. L. H. (1989). The Influence of English on The Lexical Expansion of Bahasa Malaysia. Kuala Lumpur: Dewan Bahasa dan Pustaka. Ministry of Education.
- Katamba, F. (1994). *Morphology*. Houndmills, UK: McMillan.
- Naseri, M. H., Hesami, T., Ghanbari, A., & Dalilan, S. (2011). Trends in Persian Medicinal Terminology, A Progressing Field of Interdisciplinary Research. *Journal of Paramedical Sciences.*, 2(2), 41–47.
- Newmark, P. (1988). *A Textbook of Translation*. New York: Prentice Hall International.
- PLLA (The Persian Language and Literature Academy) (2009). *Naming Principles and Regularities Together with Descriptions*. Tehran: Naming Department.
- Qazi Jahani, B., & Tabrizi, A. (2005). *Dorland's Medical Dictionary – English into Persian*. Tehran: Aryan Teb.
- Sadeqi, R. (1993). The Persian Language Experiments in Science. In A. Kafi (Ed.), *Proceeding of the Persian Language and Science Language Seminar* (pp. 2-72, 125-6). Tehran University.
- Safdari, R., Mehdiqolikhani, R., Azizi, A., Qazi Saeedi, M., Minavand, B., & Hemmati, M. (1973). *Guide to ICD-9-CM*. Tehran: Research Sector of Iranian Ministry of Healthcare and Medical Education.
- Sager, J. C. (1990). *A Practical Course in Terminology Processing*. Amsterdam: John Benjamins Publishing Company. doi:10.1075/z.44
- Tabatabaee, A. (2009). *Word Structure and Grammar*. Tehran: Culture, Art and Communication Research Center.
- United States of America. American Medical Association. (2007). *ICD-9-CM, 9th Revision*, (vols. 1-3). Washington, DC: Government Printing Office.
- Yule, G. (1988). *The Study of Language*. Cambridge, UK: Cambridge University Press.

ADDITIONAL READING

Cabre, M. T. (1999). *Terminology: Theory, Methods and Applications*. Amsterdam, Philadelphia: John Benjamins Publication. doi:10.1075/tlrp.1

Finch, C. A. (1969). *An Approach to Technical Translation*. New York: Pergamon Press Ltd.

Hilton, A., & Skrutkowski, M. (2002). Translating Instruments into Other Languages: Development and Testing Processes. *Cancer Nursing*, 25(1), 1–7. doi:10.1097/00002820-200202000-00001 PMID:11838715

Kin, W. F. (2009). *Applied Medical Terminology Research*. Rockville Pike. U.S. National Library of Medicine.

Lee-Janke, H. (1998). Training in Medical Translation with Emphasis on German. In H. Fischbach (Ed.), *Translation and Medicine* (p. 89). Amsterdam: John Benjamins Pub. doi:10.1075/ata.x.10lee

Newmark, P. (1979). *A Layman's View of Medical Translation*. Retrieved July 13, 2013, from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1597100>

Sager, J. C., Dungworth, D., & McDonald, P. F. (1980). *English Special Languages: Principles and Practice in Science and Technology*. Weisbaden: Brandstetter.

Zeinali, A. A. (2009). *The Translation of English Medical Terms into Persian*. (Unpublished Master's dissertation). USM, Penang.

KEY TERMS AND DEFINITIONS

Cacophony: According to Encyclopedia Britannica, “cacophony”, the opposite of “euphony”, refers to combination of words that produce harsh discordance of sound. Linguistically, euphony and cacophony belong to the study of inherent pleasantness or unpleasantness of the sound of certain utterances. This knowledge is called “phonaesthet-

ics”. Morphologically, this word is derived from two Greek word parts that mean “voice-sound” and “aesthetics.” Despite cacophony, euphony is achieved through the use of vowel sounds in words. Vowel sounds, which are more easily pronounced than consonants, are more euphonic; the longer vowels are the most melodious. See dissonance.

Compatibility: In this study, the characteristics of the equivalents or terms will be investigated and discussed according to the compatibility status with Sager's criteria and the PLLA principles. In other words, it will be studied if the terms follow the naming guidelines (as compatible equivalents) or appear out of naming guidelines (as incompatible equivalents).

Effectiveness: “Effective” or “ineffective” are the terms which are employed in this study for the effectiveness of the applied translation procedures in the translation process of English medical terms into Persian. It is evaluated based on four terminological factors suggested by Meyer and Bowker (2006: p.117) which belong to morphosemantics in linguistics. Conciseness, absence of competing terms, derivative form capability, and compliance with the rules of the language are the four factors that can all contribute to the effectiveness of the applied translation procedures in translating the English medical terms into Persian.

Morphosemantics: Morphosemantics is generally a knowledge in linguistics, pertaining to morphological analysis combined with a semantic interpretation of words.

Naming Guidelines: It refers to the criteria or principles which should be followed in naming or word designation process. Naming guidelines can be categorized into the ones which are common for all the languages around the world, namely universal or international guidelines, and the ones which are specific to a certain language society namely local guidelines.

Naming / Word Designation Process: Naming occurs once a new concept, object, phenomenon and the like appears. In this process, a name has undergone multiple attempts and word formation processes (Sager, 1990: p.63).

Translation Procedures: It refers to particular translation courses which are applicable to sentences and the smaller units of language. (1988). (p. 81). Newmark.

Word Formation Processes: It speaks of the continuous development of new terms and new

applications of old terms leading to language productivity in the way a language is evolved based on its user's requirements (Yule, 1988: p.52). There are different ways for word formation to enter lexicon.

L

Category L

Learning Assessment and Measurement

Implicit Cognitive Vulnerability



Caroline M. Crawford

University of Houston – Clear Lake, US

INTRODUCTION

Implicit Cognitive Vulnerability is a theoretical construct, a concept that engages in an understanding revolving around how a person understands information, shifts information from short-term memory into long-term memory access, and the vulnerability within the learner's cognitive processes that support or abandon the learner's ability to conceptualize the new information within prior learned information. Vygotsky (1933/1966, 1935, 1981) referred to this as the conceptual framework of understanding, fitting new information within the previously learned information. The intriguing shift in thought, referred to as Implicit Cognitive Vulnerability, suggests that the discomfort and tension surrounding learning new information and engaging with this information in new and different ways introduces a level of cognitive dissonance and conceptualized resistance to the person's re-envisioning information in new and different ways. This vulnerability, specifically cognitive vulnerability, is the inherent undergirding towards a rationalization revolving around understanding new information in new and different, creative ways as the information moves from short-term memory and into implicit, long-term conceptual frameworks of understanding. The objectives of the article is to frame an understanding of Implicit Cognitive Vulnerability as a developing learning theory that supports one's understanding of the instructional process. This includes the supportive development of a "comfortableness" within the instructional environment that supports the learner's cognitive efforts and developing understanding of prior information with new information, as well as the creativity involved in "thinking outside the box" within a safe and supportive environment.

DOI: 10.4018/978-1-5225-2255-3.ch447

BACKGROUND

Cognitive vulnerability within the realm of cognitive psychology has a focus upon states of emotional depression (Haeffel & Hames, 2014; Hanklin & Abramson, 2001; Matthews & MacLeod, 2005; McGinn, Nooner, Cohen & Leaberry, 2015). However, within the learning process, one might suggest that cognitive vulnerability is not so much a state of depression, as much as a state of cognitive dissonance wherein the learner is attempting to understand information in new and different ways, by not only framing the information within the previously developed knowledge base, or conceptual framework of understanding (Vygotsky, 1933/1966, 1935, 1981) as well as the learner's need for support within Maslow's Hierarchy of Needs (1943, 1954, 1962, 1968, 1964/1970a, 1970b) that frames a five stage model that frames the deficiency needs (physiological, safety, social, esteem) and growth need (self-actualization) as vitally important areas of impact upon a learner's motivational needs. One may suggest that the vulnerability inherent within the learning process engages Vygotsky's conceptual framework of understanding the specific subject matter that also includes the connectedness of the new information with previously understood information, with Maslow's Hierarchy of Needs that engages in the safety and esteem aspects of the community of learning (Klamma, Rohde & Stahl, 2005; Swan, 2002). This also supports Wittgenstein's (1961) work related to a learner's understanding of information within the realm of their own socially communicated and socially corrective understandings that frame learned information within a social context that may be perceived as the community of learning environment

as well as the larger community of practice that also includes the larger social realm of influence.

The importance of Implicit Cognitive Vulnerability within a community of learning is clearly articulated through an understanding of the deficiency needs as defined by Maslow (physiological, safety, social, esteem) that clearly embraces the impact of the social realm upon the learner that includes not only the instructor but also the collegial learners. This includes the vitally integral support of the community of learners as a safe, supportive area that embraces self-efficacy (Bandura, 1977, 1986, 1997) as a motivational effort, including the expectancy construct (Parsons & Goff, 1978), supporting Vroom's expectancy theory (1964) that, according to Holdford and Lovelace-Elmore (2001), Vroom supports an understanding that "intensity of work effort depends on the perception that an individual's effort will result in a desired outcome" (p. 8). This supports an understanding towards the self-actualization and creativity that are inherent not only towards Bloom's Cognitive Taxonomy (Bloom, 1956, 1984; Bloom, Englehart, Furst, Hill & Krathwohl, 1956; Bloom & Krathwohl, 1956; Krathwohl, Bloom & Masia, 1964). This also includes Anderson and Krathwohl's revised Taxonomy for the Digital Age (2001) focus upon higher order thinking skills achievements but also towards supporting the creativity inherent within higher order thinking skills that may be described as "out of the box" creative thinking in new and different ways that are inherently vulnerable arenas into which one may journey boldly.

MAIN FOCUS OF THE ARTICLE

The primary concerns revolving around the conception of Implicit Cognitive Vulnerability have to do with the impact of the instructional facilitator, the collegial learners, and the subject matter prior experience on the part of the learner. Each aspect within the instructional environment is integrally

important to the impact upon the learner's cognitive understanding and implicit memory retention and retrieval. Simplistically stated the triad of components, or constituents, within this instructional process are imperative towards supporting and engaging the learner's successes.

Instructional Facilitator

The instructional facilitator is the learning community mentor, wherein the tone of the learning community is set and the level of collegial support by all members within the instructional community is gauged. The instructional facilitator's style of instructional presentation of materials to the learners, the ways that the students engage with the information, and the levels of direct engagement with higher order thinking skills efforts is directly designated and supported by the instructional facilitator. As well, the level of collegiality and professionalism of the learners within the instructional environment is set up and supported by the instructional facilitator, towards being a supportive environment in which learners are indeed colleagues who support the engagement with the information while also engaging in either supportive or non-supportive "out of the box" nontraditional and creative thought processes. This understanding and engagement may more fully address using the information in new and different ways, that not only integrate the new knowledge within each learner's conceptual framework of understanding, as checked and corrected by the instructional facilitator as well as learner colleagues, but also more fully supports the integration of new information into each learner's implicit memory attainment. The "comfortableness" modeled by the instructional facilitator guides the learners towards a collegial engagement within the bound learning environment, wherein a sense of "comfortableness" and creative engagement with the subject matter and learner colleagues may develop into a sense of community engagement that supports the learner's

motivational and self-efficacy efforts as undergirded by belonging and engaging as an important member of the learning community.

Collegial Learners

The collegial learner is the second component of the triad. The leadership, guidance and mentorship of the instructional facilitator directly impacts the collegial learners, not only as singular learners but also within a collegial community of learners. The concept of Vroom's Expectancy Theory (1964) lives within a performance-reward correlation that supports the instructional facilitator's mentorship and expectations revolving around the learning community that is perceived and rewarded. Towards understanding the learning community expectations as set by the instructional facilitator, the learner colleagues develop an understanding as regards boundaries, communication and relational expectations, as well as levels of performance expectation that is set with not only the information learned but also the relational performance between learner colleagues. Expectations as regards supporting all members of a community are modeled and embraced by the instructional facilitator, with the expectation that the learner colleagues follow suite and engage in similar levels of quality engagement and expectation. The expectations revolving around perceived effort, perceived performance and collegial respect and support are integrally important towards supporting the instructional process that occurs within the learning community. The ability of the collegial learners to delve into cognitively vulnerable efforts, directly understood as being potentially uncomfortable and vulnerable new ways of thinking about the information being learned, is an inherent undergirding of the learning process that results in the learners engaging with the information towards the higher order thinking skills arena as designated by Bloom (1956, 1984; Bloom, Englehart, Furst, Hill & Krathwohl, 1956; Bloom & Krathwohl, 1956; Krathwohl, Bloom &

Masia, 1964) and Anderson and Krathwohl (2001); specifically, the learners are engaging with the information at the levels of analysis, synthesis, evaluation and creation of information in new and different ways. The level of cognitive vulnerability of the learners, to engage with co-learners and with the instructional facilitator within unclear, creative and vulnerable levels of understanding the information being learned, brings forward not only a cognitive dissonance but also a cognitive vulnerability within the learner's own sense of learner self and self-efficacy. The learner's desire to support a level of self-efficacy and learner self while engaging in cognitively vulnerable efforts will naturally develop a level of cognitive resistance, as engaging in each level of higher order thinking skills is uncomfortable and rationally difficult to achieve. The instructional facilitator supports the collegial learner's efforts within the community, supporting, correcting, coaching and rewarding each learner's efforts towards developing a stronger conceptual framework of understanding as regards the new information's integration into long-term memory, as well as the ability of each learner towards developing a comfortableness with the concept of sharing their own creativity and creative ideas with others in a cognitively vulnerable manner. The more fully the instructional facilitator supports the collegial learner's efforts towards working with the information in cognitively vulnerable manners, the more fully the learners will embrace the new information within their own schemas of understanding and implicit memory engagement.

Subject Matter

The subject matter is the final component of the triad. Different groups of learners bring forward different conceptual understandings as revolves around their own past experiences, their own prior knowledge base, and their own conceptual frameworks of understanding as regards the subject matter addressed within an instructional



environment. This understanding of the learner's background and levels of prior knowledge attainment as well as levels of engagement and expertise with the subject matter will guide the instructional facilitator's instructional design for the instructional experience. Meaning, if the learners within the community of learning are novices with the specific subject matter, then the learners will require a significantly different approach towards working with the knowledge and remembering it, to understanding the knowledge, and then the inherent application of the new knowledge; each of these aspects are considered lower order thinking skills within Bloom's (1956, 1984; Bloom, Englehart, Furst, Hill & Krathwohl, 1956; Bloom & Krathwohl, 1956; Krathwohl, Bloom & Masia, 1964) and Anderson and Krathwohl's (2001) cognitive taxonomies. The instructional facilitator must work with the learners towards ensuring that the subject matter is understood and have fully engaged with the knowledge at these lower levels of thinking skills in carefully orchestrated progression, so as to then more fully engage with the subject matter at higher orders of thinking and subject matter engagement. As the collegial learners progress through each level of subject matter understanding and engagement (as designated by Bloom's and Anderson and Krathwohl's work), the cognitive vulnerability of the learners is recognized as cognitive dissonance and cognitive vulnerability, as the learners strive towards overcoming a sense of cognitive resistance and achieves a higher level of subject matter understanding, engagement and the slow shift of information from short-term memory towards long term memory engagement and understanding. The subject matter is an integral aspect of the instructional process, specifically as regards the conception of Implicit Cognitive Vulnerability, as different information is appropriately learned in different manners by different groups of learners. To more simply state this, different information is learned by different learners in different ways.

SOLUTIONS AND RECOMMENDATIONS

An interesting discussion revolves around Implicit Cognitive Vulnerability and ways to deal with issues, controversies or problems associated with the concept of Implicit Cognitive Vulnerability. One instructional style does not meet the needs of every single learner, even if the subject matter is similar in presentation and prior knowledge attainment and understanding. Much of this may be the ways that different instructional facilitators present the subject matter to the learners; as well, much of this may be due to the community of learning that is developed and the trust that bonds the collegial learners within the instructional environment.

Instructional Facilitator

Again this discussion returns to the instructional facilitator, as the instructional design and presentation of the subject matter within an instructional environment is framed by the instructional facilitator's philosophical beliefs systems as regards the beliefs revolving around how people learn, why people learn, framing the information for the learners, as well as the level of community support and engagement necessary towards modeling a conceptual framework of understanding. This occurs through social discourse and communicative choices that engage with the subject matter being learned. The quality of instructional experience is framed by the instructional facilitator, although the subject matter and the community of learning's collegial learners also are integrally engaged in the instructional successes that support the concept of Implicit Cognitive Vulnerability. The instructional environment in which the learners engage with the information is initially framed and supported by the instructional facilitator, with decisions made as to the "comfortableness" and cognitive safety of the learners throughout their instructional experience. As well, the support

towards “thinking outside the box” in new and different ways can be reinforced and sustained by the instructional facilitator to different levels of being and different motivational and self-efficacy senses of belonging.

Reflective Practitioner

Solutions that embrace the concept of a reflective practitioner may be framed through Cruikshank’s (1987) Reflective Teaching Model that fully engages the instructional facilitator in analyzing their own instructional practices and peccadillos. This analysis more fully supports the realization of an instructional facilitator’s unique ways of “being” and of “teacher presence” that also suggests one’s ability to engage in conflict resolution within a professional and mutually inclusive and respectful manner. The reflective understandings that are derived from an instructional facilitator’s reflective practices and ability to critically analyze one’s own strengths, weaknesses and areas of necessary improvement suggest that not only is the instructional facilitator open and cognitively vulnerable in one’s own manner of being, but also offers a more self-serving recognition that the instructional facilitator is more attuned to the needs of the collegial learners and a stronger mentorship understanding may engage a stronger sense of engagement. This stronger sense of engagement and respectful understanding of the needs of the collegial learners within the instructional community more fully supports the concept of Implicit Cognitive Vulnerability. Recognizing, supporting and engaging the developing self-efficacy of the learners associated with the subject matter, as well as within the learner’s own sense of ability, “comfortableness” and internalized sense of motivation within the instructional environment can be more fully realized through the reflective practitioner’s own sense of engagement and reflective understandings of the learner’s efforts associated with the subject matter, with the learning community, with the cognitive efforts towards not only “comfortableness” but also stretching one’s own

analytical abilities towards creativity that may be described as “thinking outside the box” in new and different ways that are recognized as safely occurring within a bound learning community.



FUTURE RESEARCH DIRECTIONS

Directions for future research are intriguing, as the concept of Implicit Cognitive Vulnerability as an understanding revolving around the learning processes and needs of learners is in its mere infancy. Recognition of the learner’s need for cognitive support throughout the instructionally relevant learning process is inherent, while recognition of the learner’s cognitive dissonance and cognitive vulnerability as an in-between aspect of concern as the learner progresses from the lower order thinking skill domains towards each successively higher cognitive taxonomy thinking skill domain (Anderson & Krathwohl, 2001; Bloom, 1956, 1984; Bloom, Englehart, Furst, Hill & Krathwohl, 1956; Bloom & Krathwohl, 1956; Krathwohl, Bloom & Masia, 1964) must be more fully understood and supported. More fully supporting the learner at these sensitive points of informational engagement and vulnerability will aid in successively lowering the potential levels of resistance in the learner’s cognitive efforts with the conceptual framework of understanding and implicit memory attainment that shifts a learner from novice subject matter understandings towards more experienced understandings revolving around the subject matter. The recognition that the community environment is integrally important is also vital towards the success of the learner, as the social aspects of the learning process directly impacts the learner’s successful attainment of the information.

A stronger understanding of the learner’s cognitive dissonance and cognitive vulnerability, specifically as it impacts the learner’s ability to understand and work with the subject matter informational attainment must be more fully addressed. The social engagement and vulnerable

cognitive points throughout the learning process must also be more fully understood. These are integrally important aspects to consider, within the Digital Age impact of not only traditional instructional environments and the influence of the instructional facilitator within this environment, but just as vital are the instructional environments of blended learning, online learning and mobile learning. The dichotomy that currently exists within the instructional environments must be addressed through future research opportunities.

CONCLUSION

The impact of the instructional facilitator, the subject matter and the collegial learner environmental community engage in impacting the learner's Implicit Cognitive Vulnerability. The viability of this theory cannot be more strongly emphasized, as the impact of the instructional environment's framing and support structure embraces a learning community that is also viable towards directly impacting the learner's successful attainment of the information desired.

REFERENCES

- Anderson, L.W., & Krathwohl. (Eds.). (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84(2), 191–215. doi:10.1037/0033-295X.84.2.191 PMID:847061
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Bandura, A. (1997). *Self-Efficacy: The exercise of control*. New York, NY: W. H. Freeman.
- Bloom, B. S. (Ed.). (1956). *Taxonomy of educational objectives: The classification of educational goals: Handbook I, cognitive domain*. New York: Longmans, Green.
- Bloom, B. S. (1984). *Taxonomy of educational objectives*. Boston, MA: Allyn and Bacon.
- Bloom, B. S., Enghart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (Eds.). (1956). *Taxonomy of educational objectives, the classification of educational goals, Handbook I: cognitive domain*. New York: Longmans.
- Bloom, B. S., & Krathwohl, D. R. (1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals, by a committee of college and university examiners. Handbook I: Cognitive Domain*. New York: Longman, Green.
- Cruikshank, D. R. (1987). *Reflective Teaching: The preparation of students of teaching*. Reston, VA: Association of Teacher Educators.
- Haefel, G. J., & Hames, J. L. (2014, January). Cognitive vulnerability to depression can be contagious. *Clinical Psychological Science*, 2(1), 75–85. doi:10.1177/2167702613485075
- Hanklin, B. L., & Abramson, L. Y. (2001). Development of gender differences in depression: An elaborated cognitive vulnerability-transactional stress theory. *Psychological Bulletin*, 127(6), 773–796. doi:10.1037/0033-2909.127.6.773 PMID:11726071
- Holdford, A. D. A., & Lovelace-Elmore, B. (2001). Applying the principles of human motivation to pharmaceutical education. *Journal of Pharmacy Teaching*, 8(4), 1–18. doi:10.1300/J060v08n04_01
- Klamma, R., Rohde, M., & Stahl, G. (Eds.). (2005) Special Issue on: Community-Based Learning: Explorations into Theoretical Groundings, Empirical Findings and Computer Support. *SigGroup Bulletin*, 24(4), 1-100.



Krathwohl, D. R., Bloom, B. S., & Masia, B. B. (1964). *Taxonomy of educational objectives: The classification of educational goals. Handbook 11: The affective domain*. New York: David McKay.

Maslow, A. H. (1943). A Theory of Human Motivation. *Psychological Review*, 50(4), 370–396. doi:10.1037/h0054346

Maslow, A. H. (1954). *Motivation and personality*. New York: Harper and Row.

Maslow, A. H. (1962). *Towards a psychology of being*. Princeton, NJ: D. Van Nostrand Company. doi:10.1037/10793-000

Maslow, A. H. (1964/1970a). *Religions, values, and peak experiences*. New York: Penguin.

Maslow, A. H. (1968). *Toward a Psychology of Being*. New York: D. Van Nostrand Company.

Maslow, A. H. (1970b). *Motivation and personality*. New York: Harper & Row.

Mathews, A., & MacLeod, C. (2005, February). Cognitive vulnerability to emotional disorders. *Annual Review of Clinical Psychology*, 1(1), 167–195. doi:10.1146/annurev.clinpsy.1.102803.143916 PMID:17716086

McGinn, L. K., Nooner, K. B., Cohen, J., & Leaberry, K. D. (2015, August). The role of early experience and cognitive vulnerability: Presenting a unified model of the etiology of panic. *Cognitive Therapy and Research*, 39(4), 508–519. doi:10.1007/s10608-015-9673-9

Parsons, J. E., & Goff, S. B. (1978). Achievement & motivation: Dual modalities. *Journal of Educational Psychology*, 13(1), 93–96. doi:10.1080/00461527809529199

Swan, K. (2002). Building communities in online courses: The importance of interaction. *Education Communication and Information*, 2(1), 23–49. doi:10.1080/1463631022000005016

Vroom, V. H. (1964). *Work and motivation*. New York: Wiley.

Vygotsky, L. S. (1933/1966). Play and its role in the mental development of the child. *Social Psychology*, 12(6), 62–76.

Vygotsky, L. S. (1935). *Mental development of children during education*. Moscow: Uchpedzig.

Vygotsky, L. S. (1981). The genesis of higher mental functions. In J. V. Wertsch (Ed.), *The concept of activity in Soviet psychology*. Armonk, NY: Sharpe.

Wittgenstein, L. (1961). *Tractus logico-philosophicus* (D. F. Pars & B. F. McGuinness, Trans.). New York: Humanities Press.

ADDITIONAL READING

Crawford, C. M. (February 17, 2015). Vulnerability in learning. In J. M. Spector, T. Johnson, D. Ifenthaler, W. Savenye & M. Wang (Eds.) *The Sage Encyclopedia of Educational Technology*, pp. 831–833. Los Angeles, CA: SAGE Publications, Inc. Doi:10.4135/9781483346397

Crawford, C. M. (2016). *Instructor immediacy and authenticity: Engaging in cognitive vulnerability within the online instructional environment. Creating Teacher Immediacy in Online Learning Environments* (pp. 15–36). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-9995-3.ch002

Crawford, C. M., & Semeniuk, M. S. (2016). *Metaphoric Representations of Cognitive Understanding via a Stairway Approach: Implicit Cognitive Vulnerability Theory through a Progressive Cognitive Taxonomical Approach. Research Highlights in Technology and Teacher Education, 2016. A publication of the Society for Information Technology and Teacher Education*. Waynesville, North Carolina: Association for the Advancement of Computing in Education.

Crawford, C. M., & Smith, M. S. (2014). Re-thinking Bloom's Taxonomy: Implicit Cognitive Vulnerability as an Impetus towards Higher Order Thinking Skills. In J. Zing (Ed.), *Exploring Implicit Cognition: Learning, Memory, and Social-Cognitive Processes* (pp. 86–103). Hershey, PA: Information Science Reference (an imprint of IGI Global); doi:10.4018/978-1-4666-6599-6.ch004

Groenewold, N. A., Roest, A. M., Renken, R. J., Opmeer, E. M., Veltman, D. J., van der Wee, N. J. A., de Jonge, P., Aleman, A., & Harmer, C. J. (2015, March). Cognitive vulnerability and implicit emotional processing: Imbalance in frontolimbic brain areas? *Cognitive, Affective, & Behavioral Neuroscience* 15(1), pp. 69-79. doi: 10.3758/s13415-014-0316-5

Kosnes, L., Whelan, R., ODonovan, A., & McHugh, L. A. (2013, September). Implicit measurement of positive and negative future thinking as a predictor of depressive symptoms and hopelessness. *Consciousness and Cognition*, 22(3), 898–912. doi:10.1016/j.concog.2013.06.001 PMID:23810864

Kruijt, A.-W., Antypa, N., Booij, L., de Jong, P. J., Glashouwer, K., Penninx, B. W. J. H., & Van der Does, W. (2013). Cognitive Reactivity, Implicit Associations, and the Incidence of Depression: A Two-Year Prospective Study. *PLoS ONE*, 8(7), e70245. doi:10.1371/journal.pone.0070245 PMID:23922962

van Tuijl, L. A., de Jong, P. J., Sportel, B. E., de Hullu, E., & Nauta, M. H. (2014, March). Implicit and explicit self-esteem and their reciprocal relationship with symptoms of depression and societal anxiety; A longitudinal study in adolescents. *Journal of Behavior Therapy and Experimental Psychiatry*, 45(1), 113–131. doi:10.1016/j.jbtep.2013.09.007 PMID:24135033

KEY TERMS AND DEFINITIONS

Cognition: Cognition is a term that focuses upon the thought processes of a person. Within this discussion, the term revolves around the mental processes of persons involved in the instructional process.

Cognitive Vulnerability: The term is primarily implemented within psychology as suggesting a person's depressive state. Within this discussion, the term does not suggest a level of mental depression. Instead, this term focuses upon a sense of dissonance, meaning discord or conflict, within a person's mental processes while learning new information but more directly this term focuses upon the learner's efforts towards using information in new and different ways of understanding.

Collegial Learner: This term is offered as a label for all learners within an instructional environment. This term is implemented with the desire to reflect an instructional community environment wherein the students come together and support each other's instructional understanding as well as supporting each other's knowledge attainment while using the knowledge in new and different ways of understanding.

Community of Learners: This term is an overarching understanding of the group of students, also including the instructional facilitator, who come together with the intention to learn information while also supporting the larger group's instructional understandings and efforts. This term reflects a philosophical understanding, that learning is not a singular activity but, instead, is a socially supported effort.

Conceptual Framework of Understanding: This term was coined by Vygotsky (1933/1966, 1935, 1981) and means that all information is remembered by a person in a mixed understanding with other information. No learned information stands alone; instead, the information is linked within each person's understanding with other prior knowledge, whether learned or experienced information. This occurs in a social context wherein the person learning the information, or

re-learning prior information, is making connections of understanding between all prior knowledge points that connect with the current information under study. The checking of understanding occurs in a social context, with re-learning and re-connecting prior knowledge information occurring as is appropriate.

Instructional Facilitator: This term reflects the role of the teacher within any instructional environment. The reason for the term is the understanding that the teacher is not only a “sage on the stage” but is also a mentor, a coach, and a Socratic questioner as is necessary and appropriate towards supporting the learning efforts within the instructional processes. Facilitating each learner’s efforts is the role of the teacher.

Reflective Practitioner: This term focuses upon the professionalism of the teacher, suggesting that the continuous analysis of the teacher’s strengths, weaknesses and areas of further development are necessary and appropriate towards

recognizing the ever evolving understanding of the profession.

Subject Matter: This term reflects the information that is meant to be taught by the teacher, and learned by the students.

Vulnerability: This term describes the cognitive understanding of the instructional process, wherein the student engages in differentiated understanding and efforts associated with the subject matter being learned. The student’s comfortableness with the information, the student’s comfortableness and engagement within the instructional environment (including the level of trust that the student feels towards the teacher and the other students) can further enhance each student’s ability to openly articulate creative thoughts and ideas, feeling a sense of safety that the student will not be derided for potentially incorrect understandings or “thinking outside the box” creativity that may be unusual understandings and ideas implemented with the subject matter under study.



Learning Analytics

Constanța-Nicoleta Bodea

Bucharest University of Economic Studies, Romania

Maria-Iuliana Dascalu

University Politehnica of Bucharest, Romania

Radu Ioan Mogos

Bucharest University of Economic Studies, Romania

Stelian Stancu

Bucharest University of Economic Studies, Romania

INTRODUCTION

Reinforcement of the technology-enhanced education transformed education and training into a data-intensive domain. The software systems implemented in the education institutions (e.g. student information systems, library systems, university timetabling systems, attendance monitoring systems, lecture capture/media streaming systems, ePortfolios, assessment systems, curriculum maps, personal tutor system, student engagement system, social learning platforms etc.) made available large amounts of data related to the learning/teaching processes, etc. Educational data reflects not only the students' background, but also the students' performance. Each action carried out by students in different educational activities is stored and can be analyzed. The amount of data stored in different systems is impressive, for this reason the big data related technologies (Hadoop, Cloud etc.) are increasingly present in the IT infrastructure of education institutions.

As in many other data-intensive domains, the interest for data analysis through various business intelligence solutions is growing in the education domain. Improvement of educational activities by continuous analysis of collected data is a common practice in many educational institutions. More and more software tools are used for: modeling students' learning behavior (in order to understand

how students are engaged into online activities and how their performance can be increased), identification and evaluation of students' performance factors (in order to predict performance/lack of performance), prediction of dropout and retention at very early stages, monitoring students' progress and comparing them to their peers, in order to increase professors' awareness during learning activities, personalizing the educational processes mainly by choosing specific learning content, or improving feedback and assessment services, in order to increase students' awareness about the progress and their contribution in learning activities. By applying different business analytics methods, all these applications were associated to a new emerging field of technology-enhanced education, named learning analytics.

Learning analytics (LA) contributes to the quality assurance and improvement in education institutions (Sclater, Peasgood, & Mullan, 2016). The students and teachers could improve their learning and teaching processes. Also, the administrators could use LA as diagnostic tool for identifying issues and for improving the modules and degree programmes design. LA can lead to the development of new quality metrics and standards in the education domain and to demonstrate the compliance of education institutions to them. Some of the main positive effects of LA are the boosting of retention rates (LA allows to identify

students at risk and efficiently act with advices and support at very earlier stages) and the adoption of the adaptive learning, through the personalization of educational experiences. Students have the opportunity to take control over their own learning, having the possibility to monitor their current performance in real-time and help them to make informed choices about what to study. LA survey results (Newland, Martin, & Ringan, 2015; Sclater et al., 2016) show a clear adherence of stakeholders to LA adoption initiatives.

Although there is a growing interest for LA, the progress in the implementation of LA applications is still inconclusive, mainly due to the reduced maturity of technology-enhanced education and the lack of specific LA expertise. According to the HeLF survey (Newland et al., 2015), nearly half of UK higher education institutions have not yet implemented LA and about a third are working for LA implementation. The most common applications are the students' retention and the learning process enhancement. Responsibility for LA implementation is undertaken by different departments (Quality, Planning, Registry, IT Services, eLearning, Academic Development, Learning Services, Academic etc.). Three quarters of senior management have limited levels of understanding of the possible LA benefits and outcomes.

Another study undertook in the Danube Region of Europe (Bodea et al., 2016; Dascalu, Bodea, Stancu & Purnus, 2016) reveals that only 55.9% of the professors use a learning management system (LMS) in class, especially for content sharing and communication via forums and chats and social networks for keeping the connection with their students. Most of them (81.3%) consider the development of a social learning platform with LA functionalities as very relevant initiative for Danube Region. Regarding the students, most of those who participated to the survey use LMS in classes (75.2%), in order to share content (82.5%), to be assessed (48.5%) and collaborate (40.2%). Most students consider forums, webinars and

chats to foster online collaboration. The majority of students (88%) consider the development of a social learning platform to be relevant or very relevant for the Danube Region and more than half (58.6%) see advanced learning analytics as a mandatory feature of such a platform.

The article starts by defining LA, with relevant views on the literature. A discussion about the relationships between LA, educational data mining and academic analytics is included in the background section of the article. In the main section of the article, the learning analytics, as an emerging trend in the educational systems is described, by discussing the main issues, controversies, problems on this topic. The final part of the article presents the future research directions and the conclusion.

BACKGROUND

There are various definitions of Learning Analytics (van Harmelen & Workman, 2012; Siemens, 2010; Cooper, 2012; Ferguson et al., 2015). According to the definition proposed by the Society for Learning Analytics Research (SoLAR), LA involves “measurement, collection, analysis and reporting of data about students and learning contexts in order to understand and optimize learning and the environments in which it takes place” (van Harmelen et al., 2012). LA means “intelligent use of data produced by students and analysis of existing patterns in the data, finding information and social connections to predict and advice on learning paths” (Siemens, 2010). At the same time, the cultural values of contemporary society have changed: according to Harvard Business Review, the online learning environments are seen as the great educational revolution (Tamny, 2013). LA aims to explain the impact of cultural changes on the learning process and to optimize it accordingly. LA is, therefore, more than metrics; it is a process (Bischel, 2012). LA represents the usage of analytic techniques on educational data,



being often associated with business intelligence, data mining, web analytics, academic analytics, action analytics and predictive analytics.

According to Romero and Ventura (2013), educational data mining (EDM) means “developing, researching, and applying computerized methods to detect patterns in large collections of educational data that would otherwise be hard or impossible to analyze due to the enormous volume of data within which they exist”. Papamitsiou and Economides (2014) examined the literature on experimental case studies on EDM and LA which were conducted in 2008-2013. They conclude that EDM and LA represent two areas oriented towards the inclusion and exploration of big data capabilities in education, by using analytic techniques to evaluate rich data sources, identify patterns within the data and exploit these patterns in decision making. Associated to these two domains there is an ecosystem of methods and techniques for data gathering, processing and reporting, in order to inform and empower learners, teachers and educational organizations about performance and goal achievement, and facilitate decision making accordingly. Chatti, Dyckhoff, Schroeder and Thüs (2012) also consider EDM and LA as being two separate communities which share similar goals and focus but having different origins, techniques and areas of applications. While the LA adopts a holistic framework, seeking to understand the educational system as a whole, EDM applies a reductionist approach by seeking for individual patterns of specific components of the educational systems. Analyzing the differences, similarities and evolution of EDM and LA, Linan and Perez (2015) consider the application process as being the main difference between EDM and LA, while Siemens and Baker (2012) identify the following distinctions between EDM and LA: aims (EDM is interested in the data-driven discovery, while LA is focused on human learning process), the approach (EDM reduces systems to components, while LA analyzes the whole system), the roots (EDM is rooted on student modeling, while LA is related to semantic technologies) and applied

methods (EDM employs methods of classification, clustering and associations mining, while LA focuses on concept analysis, social network analysis, sentiment analysis and discourse analysis). Bienkowski, Feng and Means (2012) consider that LA is a multidisciplinary domain, being connected not only with education, computer science and statistics, as it is the case of EDM, but also with sociology, psychology, linguistics and learning sciences. In this regards, LA is different from EDM being more learning science-driven than data-driven. This make that LA uses different types of data compared to EDM when the students and learning contexts are monitored and analyzed, for example: preferences, affect observations, forum and discussion posts. These disciplinary influences make EDM to be more IT oriented, unlike LA that is more learning oriented.

Academic Analytics (AA) represents the usage of analytical techniques to support decision making in education institutions. Instead of learning and teaching processes, AA is focused on the business of education institutions, with marketing and administrative orientation (Ferguson, 2012; Siemens & Long, 2011). The main target group for AA are the managers/executives and not the students and teachers.

LEARNING ANALYTICS DRIVERS AND BARRIERS

LA has a rapid evolution, due to several opportunities which exist. But, there are also many barriers limiting the LA expansion in education and training.

An important driver of LA evolution is the financial pressure facing many educational institutions in our days. The budget constraints create strong demands for expenses rationalization and enhancing the efficiency and effectiveness of all education activities. Better retention rates, an increased satisfaction of the students and other stakeholders and, finally, higher enrolment figures lead to the improvement of the economic situation

of educational institutions. LA allows teachers and administrators to early predict students' dropout and to identify factors for increasing the retention. Having this information, teachers can increase the students' engagement in learning activities and consequently increase their satisfaction and retention.

The demands of students and other stakeholders are also increased. The students, as consumers which pay fees for education services expect more personalized and motivating learning experiences and better academic performance. Personalized interventions such as better recommendations/advice or formative assessments increase the motivation of students for learning (Bodea, Dascalu & Lipai, 2012). LA contributes in producing meaningful feedback and a better pedagogical support for the students. LA allows identification and evaluation of various factors as indicators of performance, for prediction purposes but also for deciding actions to increase students' performance. Some analytics are used by teachers to monitor the performance of students, being able to adapt their teaching strategy during the course delivery and make changes in content and processes.

The integration of different systems implemented in education institutions into virtual learning systems/platforms (VLEs) or learning management systems (LMSs) represents an important LA driver. The amount of data available on these platforms is impressive, considering the fact that not only the students' personal and performance data are stored, but also educational resources (e-books, videos and audios when delivered lectures are captured but also multi-media educational resources, serious games, scenarios, simulators). Considering this huge amount of data, the big data related technologies (Hadoop, Cloud) are increasingly present in the educational infrastructure. The existence of VLEs/LMSs, making available large amount of data represents a good premise for adopting LA. The integration of different data sources into one single "learning warehouse" allows analytics functions to be added to VLEs/LMSs.

Massive, open, online courses (known as MOOCs) paradigm represents a flexible online courses delivery model, allowing the participants to choose when and where to study and the level of involvement). MOOCs were considered as a 'learning revolution', mainly because provide a global access (no entry requirements or fees) and participation at good education and training (Katsarova, 2015). MOOCs are mainly developed by private companies in partnership with universities or individuals, on open-source web platforms. *Coursera* (developed by Stanford University) and *edX* (MIT and Harvard University) are very well-known examples of MOOCs. MOOCs are usually characterized by a large number of enrolled students with different backgrounds, learning experience, education levels, language skills, objectives and needs, a high dropout rates and different patterns of participation (Clow, 2013). In this context, there is an important potential for LA. Available data from thousands of students can be used for designing new courses and new training processes for new student generations. The concept of adaptive MOOCs has also emerged, based on the educational personalization achieved through analytics in MOOCs environments. Sonwalkar (2013) presents an aMOOC platform, developed by using the Amazon Web Services' cloud architecture. It is expected that the research interest in aMOOCs to increase in the next years.

Social networks became important facilitators in information exchange, engagement and overall connection between peers. The social learning is based on the usage of social networks in e-learning activities. The strengths of social learning are: it is fully accessible and affordable, it boosts collaboration and active learning and it enhances learner engagement. The weaknesses of social learning are the following: it is distracting and not secure as a learning management system, it limits the control one may have over the learning content and it cannot provide feedback on performance (Pappas, 2015). A promising direction of LA research is the *social learning analytics*. The efforts in this direction are more exploratory than explanatory. Shum and Ferguson (2012) identify several sub-



domains of social LA: social network analysis to see the connection between students, educational resources and ideas; discourse analysis, to see how to build knowledge; content analysis of actions taking place in the virtual environment; analysis of the student's mood, meaning what motivates him to learn and context analysis, supported in particular by mobile technologies, which provide information about the learning habits, for example when and where learning takes place.

The availability of different LA tools and the experience already gained in other business domains in applying analytics methods represents an important LA driver in order to explore better the student learning processes and to find efficient ways to improve them. The HeLF survey (Newland et al., 2015) finds that institutions applied or consider a variety of different LA solutions, including their own in-house developments. Table 1 presents some examples of LA tools.

Table 1. LA tools examples

Learning Analytics Tools	Examples
LA extensions for LMS	LA extension for the Khan Academy platform (Ruipérez-Valiente, Muñoz-Merino, Leony & Kloos, 2015), for a better understanding of educational processes LA extension for virtual learning worlds developed in Opensim (Fernández-Gallego, Lama, Vidal & Mucientes, 2013)
"Opportunistic" learning tools	Active Campus – an application that can connect students and resources, through alerts about friends, colleagues or potential educational events in one's proximity (Griswold et al., 2004)
Context analysis tools	MOBilearn (Syvanen, Beale, Sharples & Ahonen, 2005)
LA tools for virtual learning environments	GRAPPLE Visualisation Infrastructure Service (GVIS) platform, which aggregates data about users working in different virtual learning environments and create their profiles (Mazzola & Mazza, 2010)
LA tools for mobile applications	LearnTracker, a mobile application that tracks the learning process of students (Tabuenca, Kalz, Drachsler & Specht, 2015).

There are not only opportunities but also barriers which affect the LA dynamics, such as:

- Lack of expertise, for a significant number of teachers and managers with regard to LA. In order to mitigate this problem, it is important to increase acceptance and develop a data-driven culture in educational environments (Romero et al., 2013).
- Barriers related to the ethics and personal privacy (Greller & Drachsler, 2012). Ethics must be taken into account in all stages, from data gathering to the interpretation of outputs and decision making, for instance, by avoiding statements that could lead to discriminatory treatments when working with gender, social status, race, home country, religious beliefs, ideology or disability. Similarly, issues related to the ownership of students' data, which differ from country to country, need to be considered.

LEARNING ANALYTICS METHODS AND AREAS OF APPLICATIONS

The main groups of learning analytics methods are the following:

- Reporting Analytics

These methods allow reporting on the past situations. Various types of reports can be generated, such as: reports for selected activities, students, items and periods of time; real-time reports, for ongoing activities; regular activity reports, regarding the students for selected students, a given period and courses/activities etc. Blackboards and Google Analytics also offers several types of reports. Data visualization methods, such as: heat maps, learning curves and Isax1s can be also included in this group, assisting the distillation of data for human judgment.

- Automated analytics, such as: clustering, factor analysis, social network analysis and domain structure discovery.

They provide models for end-users to understand and deal with the present challenges (for example: cognitive models of students, which represent their skills and declarative knowledge, students' learning behavior models or models of domain structure). Even if these results/models are not produced by predictive analytics, they can be used to make predictions.

- Predictive analytics, such as: classifiers, regressors, association rule mining, correlation mining, sequential pattern mining and latent knowledge estimation methods. All these methods create predictive models to deal with future challenges.

Table 2 presents the main actual areas of application for the learning analytics with relevant methods and examples. Baker and Yacef (2009) study the proportion of works reporting the usage of each method group, during the period 1995 – 2009. From 1995 until 2005, the most applied methods were the relationship mining methods (43%) and predictive ones (28%). Human judgment or exploratory data analysis (17%) and clustering

(15%) were also popular. In contrast, in 2008-2009, the importance of relationship mining decreased (9%), while predictive methods reached 1st place (42%, of papers from 2008 reported the usage of predictive methods). The usage of human judgment and clustering methods remained relatively constant (12% and 15%, respectively). The application of the item response theory, Bayesian nets and Markov decision processes increased (28%).

SOLUTIONS AND RECOMMENDATIONS

The future of LA relies on the technological progress but also how some organizational strategic issues are managed by educational institutions.

Technological developments in the field of LA platforms address the openness and the architectural modularity. Openness has a critical role in the LA future, meaning usage of open source LA software (such as Weka, Kettle, Pentaho, R or Python), alignment to open standards and APIs for LA (such as: Experience API (xAPI) or IMS Caliper/Sensor API), adherence to open models and open content/access. Openness with regards to ethics/privacy issues means transparency and building the trust of end-users. Open Academic Learning Analytics Initiative (OAAI) is one example of the

Table 2. Areas of application and examples

Main Areas of Application	Learning Analytics Methods	Examples
Increase (self-) reflection and (self-) awareness	Reporting and automated analytics	Baker, Corbett & Wagner (2006) Santos et al. (2012)
Creating real time alerts for stakeholders	Reporting and automated analytics	Baron, J. (2015).
Improve feedback and assessment services	Reporting, automated and predictive analytics	Barla et al. (2010)
Student modeling	Automated and predictive analytics	Jeong & Biswas (2008)
Recommendation of resources	Automated and predictive analytics	Antonenko, Toy & Niederhauser (2012) Romero and Ventura (2013)
Scientific inquiry in education	Automated and predictive analytics	Baker, Corbett & Wagner (2006) Ming & Ming (2012)
Prediction of performance	Predictive analytics	Yadav & Pal (2012)
Prediction of dropout and retention	Predictive analytics	Ueno (2004)

strategic relevance of openness in the evolution of LA platforms. The architecture modularity means that the critical functions (data collection, storage, analysis and action) have to be assured by distinct components/modules of the platform. Jisc's learning analytics architecture (Sclater et al., 2016) implements this modularity concept. At the heart of Jisc architecture is the learning analytics processor where predictive analytics are carried out, and lead to action coordinated by the alert and intervention system. Data visualizations are available in a series of dashboards. The student applications allow learners to compare their performance with others.

The organizational strategic issues related to LA are the following: the organizational Leadership, culture and competences, assuring access to the educational data, ethics and privacy.

FUTURE RESEARCH DIRECTIONS

The future research in LA should address the actual limitations of the LA solutions (Ferguson, 2012; Agudo-Peregrina, Iglesias-Pradas, Conde-González, & Hernández-García, 2014) which negatively affect the extension of LA. The main relevant limitations are the following:

- Lack of sufficient application of science education in the implementation of LA (which would have the following benefits: optimized planning of the learning process, effective implementation of the principles of teaching and increasing students' self-awareness).
- Poor operation of different datasets (both in terms of the origin and type- mobile data, biometric, emotional, and so on) and their combination; we should understand that the current trend is to use multiple virtual learning environments or learning manage-

ment systems such as Moodle or informal social networks, thus a meaningful analysis of learner's data should be done on the data collected from all the environments in which he/she is active.

- Focusing less on student data analysis (taking into account their profile, career goals, motivational factors, etc.) and more on grades and institutional needs, which is not a solid approach.
- Lack of a clear set of ethical rules, taking into account the ownership and management of data.
- Lack of consensus regarding data relevance (what data is really important to be measured, understood and used in subsequent modelling) as being other relevant limitation of the current LA solutions.

LA research already starts to shift from technological aspects towards an educational focus. Also, it is expected that more LA research will be done in social learning analytics.

CONCLUSION

LA is a promising emerging domain, with associated strengths, weaknesses, opportunities and threats. LA strengths include the huge volumes of available educational data, availability of powerful methods for data analysis and a growing learning theory. LA weaknesses are the potential misinterpretation of data, lack of the education systems integration and lack of expertise in implementing LA platforms. LA opportunities include the existence of open standards for increasing the compatibility across systems and achievements in business analytics in other domains. LA threats are mainly related to the ethical and data privacy issues.

REFERENCES

- Agudo-Peregrina, Á. F., Iglesias-Pradas, S., Conde-González, M. Á., & Hernández-García, Á. (2014). Can we predict success from log data in VLEs? Classification of interactions for learning analytics and their relation with performance in VLE-supported F2F and online learning. *Computers in Human Behavior*, 31, 542–550. doi:10.1016/j.chb.2013.05.031
- Antonenko, P. D., Toy, S., & Niederhauser, D. S. (2012). Using cluster analysis for data mining in educational technology research. *Educational Technology Research and Development*, 60(3), 383–398. doi:10.1007/s11423-012-9235-8
- Baker, R. S. J. D., Corbett, A. T., & Wagner, A. Z. (2006). Human classification of low-fidelity replays of student actions, in: Ikeda, M., Ashlay, K. & Chan, T. (eds.), *Proceedings of the 8th International Conference on Intelligent Tutoring Systems*. Springer.
- Baker, R. S. J. D., & Yacef, K. (2009). The State of Educational Data Mining in 2009: A Review and Future Visions. *Journal of Educational Data Mining*, 1(1), 3–17.
- Bienkowski, M., Feng, M., & Means, B. (2012). *Enhancing teaching and learning through educational data mining and learning analytics: An issue brief*. US Department of Education, Office of Educational Technology.
- Bischel, J. (2012). *Analytics in higher education: benefits, barriers, progress and recommendations*. EDUCAUSE Centre for Applied Research. Retrieved July 30, 2016, from <https://library.educause.edu/resources/2012/6/~media/6f422b4bed3a439ba3925cd992144811.ashx>
- Bodea, C. N., Dascalu, M. I., & Lipai, A. (2012). Clustering of the Web Search Results in Educational Recommender Systems. In O. C. Santos & J. G. Boticario (Eds.), *Educational Recommender Systems and Technologies: Practices and Challenges* (pp. 154–181). Hershey, PA: Information Science Reference. doi:10.4018/978-1-61350-489-5.ch007
- Bodea, C. N., Mogos, R. I., Dascalu, M. I., Moldoveanu, A., Dragoi, G., Huemann, M., & Schilder, M. (2016). *Emerging Technologies: Adding Dimensions to Lifelong Learning*. IEEE Zooming Innovation in Consumer Electronics International Conference (ZINC), RT-RK Institute for Computer Based Systems. doi:10.1109/ZINC.2016.7513657
- Chatti, M. A., Dyckhoff, A. L., Schroeder, U., & Thüs, H. (2012). A reference model for learning analytics. *International Journal of Technology Enhanced Learning*, 4(5), 318–331. doi:10.1504/IJTEL.2012.051815
- Clow, D. (2013). MOOCs and the funnel of participation. *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge*. doi:10.1145/2460296.2460332
- Cooper, A. (2012). *What is Analytics? Definition and Essential Characteristics*. Retrieved July 30, 2016, from <http://publications.cetis.org.uk/2012/521>
- Dascalu, M.-I., Bodea, C. N., Stancu, S., & Purnus, A. (2016). *Social Learning Analytics: a Mandatory Step for E-Education Success in Danube Region*. 15th International Conference on Informatics in Economy, Cluj-Napoca, Romania.
- Ferguson, R. (2012). Learning analytics: Drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, 4(5/6), 304–317. doi:10.1504/IJTEL.2012.051816

- Ferguson, R., Macfadyen, L. P., Clow, D., Tynan, B., Alexander, S., & Dawson, S. (2015). Setting learning analytics in context: Overcoming the barriers to large-scale adoption. *Journal of Learning Analytics*, 1(3), 120–144. doi:10.18608/jla.2014.13.7
- Fernández-Gallego, B., Lama, M., Vidal, J. C., & Mucientes, M. (2013). Learning Analytics Framework for Educational Virtual Worlds, *Procedia. Computer Science*, 25, 443–447.
- Greller, W., & Drachsler, H. (2012). Translating Learning into Numbers: A Generic Framework for Learning Analytics. *Journal of Educational Technology & Society*, 15(3), 42–57.
- Griswold, W. G., Shanahan, P., Brown, S. W., Boyer, R., Ratto, M., Shapiro, R. B., & Truong, T. M. (2004). ActiveCampus: Experiments in Community-Oriented Ubiquitous Computing. *Computer*, 37(10), 73–81. doi:10.1109/MC.2004.149
- Jeong, H., & Biswas, G. (2008). Mining student behavior models in Learning-by-teaching environments. *Proceedings of the 1st International Conference on Educational Data Mining*, 127–136.
- Katsarova, I. (2015). *Higher education in the EU Approaches, issues and trends; In-depth analysis*. European Parliamentary Research Service, European Union.
- Linan, L. C., & Perez, A. A. J. (2015). Educational Data Mining and Learning Analytics: Differences, similarities, and time evolution. *RUSC Universities and Knowledge Society Journal*, 12(3), 98–112. doi:10.7238/rusc.v12i3.2515
- Mazzola, L., & Mazza, L. (2010). GVIS: A Facility for Adaptively Mashing Up and Representing Open Learner Models. *Proceedings of EC-TEL*, 554–559. doi:10.1007/978-3-642-16020-2_53
- Newland, B., Martin, L., & Ringan, N. (2015). *HeLF UK HE Report on Learning Analytics*. Retrieved July 30, 2016, from <http://helfuk.blogspot.ro/2015/10/uk-he-learning-analytics-survey-2015.html>
- Papamitsiou, Z., & Economides, A. (2014). Learning Analytics and Educational Data Mining in Practice: A Systematic Literature Review of Empirical Evidence. *Journal of Educational Technology & Society*, 17(4), 49–64.
- Pappas, C. (2015). *How To Use Facebook For eLearning: What eLearning Professionals Should Know*. Retrieved July 30, 2016, from <http://e-learningindustry.com/using-facebook-for-elearning-ultimate-guide-for-elearning-professionals>
- Romero, C., & Ventura, S. (2013). Data mining in education. *WIRES Data Mining and Knowledge Discovery*, 3(1), 12–27. doi:10.1002/widm.1075
- Ruipérez-Valiente, J. A., Muñoz-Merino, P. J., Leony, D., & Kloos, C. D. (2015). ALAS-KA: A learning analytics extension for better understanding the learning process in the Khan Academy platform. *Computers in Human Behavior*, 47, 139–148. doi:10.1016/j.chb.2014.07.002
- Sclater, N., Peasgood, A., & Mullan, J. (2016). *Learning Analytics in Higher Education A review of UK and international practice Full report*. Retrieved July 30, 2016, from: <https://www.jisc.ac.uk/reports/learning-analytics-in-higher-education>
- Shum, S. B., & Ferguson, R. (2012). Social Learning Analytics. *Journal of Educational Technology & Society*, 15(3), 3–26.
- Siemens, G. (2010). *What Are Learning Analytics?* Retrieved July 30, 2016, from <http://www.elearnspace.org/blog/2010/08/25/what-are-learning-analytics/>

Siemens, G., & Baker, R. S. J. D. (2012). Learning analytics and educational data mining: Towards communication and collaboration. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 252–254. doi:10.1145/2330601.2330661

Siemens, G., & Long, P. (2011). Penetrating the Fog: Analytics in Learning and Education. *EDUCAUSE Review*, 46(5), 30.

Sonwalkar, N. (2013). The First Adaptive MOOC: A Case Study on Pedagogy Framework and Scalable Cloud Architecture – Part I. *MOOCs Forum*, 1(P), 22–29. doi:10.1089/mooc.2013.0007

Syvanen, A., Beale, R., Sharples, M., & Aho-nen, M. (2005). Supporting Pervasive Learning Environments: Adaptability and Context Awareness in Mobile Learning. *Proceedings of IEEE International Workshop on WMTE*, 251 – 253. doi:10.1109/WMTE.2005.62

Tabuenca, B., Kalz, M., Drachsler, H., & Specht, M. (2015). Time will tell: The role of mobile learning analytics in self-regulated learning. *Computers & Education*, 89, 53–74. doi:10.1016/j.compedu.2015.08.004

Tamny, J. (2013). Online Education Will Be the Next ‘Bubble’ To Pop, Not Traditional University Learning. *Forbes*. Retrieved July 30, 2016, from <http://www.forbes.com/sites/johntamny/2013/06/09/online-education-will-be-the-next-bubble-to-pop-not-traditional-university-learning/#4b2b888d694d>

Ueno, M. (2004). Online Outlier Detection System for Learning Time Data in E-Learning and its Evaluation. *Proceedings of the International Conference on Computers and Advanced Technology in Education*, 248–253.

van Harmelen, M., & Workman, D. (2012). *Analytics for Learning and Teaching*. Retrieved July 30, 2016, from: <http://publications.cetis.org.uk/2012/516>

Yadav, S. K., & Pal, S. (2012). Data Mining: A Prediction for Performance Improvement of Engineering Students using Classification. *World of Computer Science and Information Technology Journal*, 2(2), 51–56.



ADDITIONAL READING

Barla, M., Bielíková, M., Ezzeddinne, A. B., Kramár, T., Šimko, M., & Vozár, O. (2010). On the impact of adaptive test question selection for learning efficiency. *Computers & Education*, 55(2), 846–857. doi:10.1016/j.compedu.2010.03.016

Baron, J. (2015). *Learning Analytics in the United States: Surveying the Landscape*. Retrieved July 30, 2016, from: <http://projet-cela.fr/wp-content/uploads/2015/12/Learning-Analytics-in-the-United-States-Surveying-the-Landscape.pdf>

Guillot, C., Guillot, R., Kumar, V., & Kinshuk. (2015). MUSIX: Learning Analytics in Music Teaching. In Y. Li, M. Chang, M. Kravcik, E. Popescu, R. Huang, Kinshuk, & N. Chen, State-of-the-Art and Future Directions of Smart Learning (pp. 267–271). Singapore: Springer S.

Lewkow, N., Feild, F., Zimmerman, N., Riedesel, M., Essa, A., Boulanger, D.,... Kode, S. (2016). A Scalable Learning Analytics Platform for Automated Writing Feedback. *Proceedings of the Third ACM Conference on Learning@Scale* (pp. 109–112). Cambridge, MA: ACM. doi:10.1145/2876034.2893380

Ming, N. C., & Ming, V. L. (2012). Predicting Student Outcomes from Unstructured Data, *Proceedings of the 2nd International Workshop on Personalization Approaches in Learning Environments*, 11–16.

Santos, J. L., Govaerts, S., Verbert, K., & Duval, E. (2012). Goal-oriented visualizations of activity tracking: A case study with engineering students, in: S. Buckingham Shum, D. Gasevic, & R. Ferguson (eds.), *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, New York, 143–152. doi:10.1145/2330601.2330639

Siemens, G. (2013). Massive Open Online Courses: Innovation in Education? In R. McGreal, W. Kinuthia, & S. Marshall (Eds.), *Open Educational Resources: Innovation* (pp. 5–16). Vancouver, Canada: Research and Practice.

KEY TERMS AND DEFINITIONS

Academic Analytics: Advanced data analysis for improving the performance of education organizations.

Educational Data Mining: Data analysis for detecting patterns in large collections of educational data.

Learning Analytics: Advanced data analysis for understanding and optimizing learning and teaching. It is an emerging field of technology-enhanced education.

Social Learning Analytics: Learning analytics applied in social learning.

Social Learning: The usage of social networks in e-learning activities.

Predicting Students Grades Using Artificial Neural Networks and Support Vector Machine



Sajid Umair

National University of Sciences and Technology (NUST), Pakistan

Muhammad Majid Sharif

National University of Sciences and Technology (NUST), Pakistan

INTRODUCTION

Education is not limited to just studying books, cramming notes and passing exams. There are a lot of other activities and habits that can affect your education. If we know what type of habits are having a bad impact on our children's studies we can try to eradicate these habits and we can replace the one's with better impact. In fact changing daily lifestyle can improve children's learning skills. In recent times, learning analytics is being used by faculty members in their courses to help students learn and improve. Everything considered, these analytics play important part for the prediction of students' performance and its improvement. The definition of Learning analytics is that, the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs (Long & Siemens, 2011). Learning analytics is used for improving student's performance, to predict it and for the purpose of retaining because with learning analytics faculty, institutions and students are able to make data-driven decisions about student's success and retention (Uhler & Hurn, 2013). A classroom has a variety of students with performances on a broad range. Some of these students show great care and perseverance and are self-motivated but there are also those that have difficulty studying and understanding. Teachers are there to guide students and to motivate them. But observing

each and everyone's learning attitudes all over the periods in the semester is a hard task. Different regular assessment methods have been used in previous studies such as e-learning logs, test marks and questionnaires to observe and assess students' learning behavior. Although, teacher's observation plays important role to increase students educational situation, they get only a few cases requiring to their needs, mostly due to their experience in the class. In this chapter our focus is on finding those skills that are having bad effect on learning environment and suggest them to the parents. We used various data mining methods to find the outcomes and also verified by our class labels that how well our algorithm works. Data mining methods and tools we used in this paper are introduced in this part by a sequence in which they were used.

In this chapter first we will discuss the introduction part in detail. After this we will explain in detail, the literature review of the chapter and also discuss some important topics which play key role in this chapter. After this we discuss the implementation and results part. And at the end we discuss the conclusion and future work section.

BACKGROUND

In educational environments, it is very important to predict student's performance. To amplify a students' performance is a long-term goal in all aca-

demographic institutions in their learning environment. Now a day, data mining technique 'Educational Data Mining' (EDM) is used on a large-scale to automatically analyze the student's performance and his behavioral data with learning environments. The use of text mining is a new trend in EDM that extends data mining on text data. A lot of experiments have been done in past couple of years in areas to predict students' academic performance. A couple of methods have also been applied in Machine learning area to obtain useful/important data and for the prediction of future data trends.

Now we want to discuss in details about the Educational Data Mining (EDM). According to the websites being used for the educational data mining, the educational data mining is an emerging field. In this field we use and developed different type of methods which are used to investigate the different data types, which we collect from the field of education, and in future that data is used for the prediction and understanding of students. Moreover different experts (Baker & Yasef, 2009; Romero et al., 2010) categorize the educational data mining work in different categories like visualization, regression, statistics, classification and clustering. Educational data mining is used for different purposes in different fields like learning performance, judge students, increase learning process, pilot students learning, give evaluation and learning recommendation of student learning behavior, judge learning elements, judge problems and abnormal learning situations, and gain broad understanding of education (Baker, 2009; Gorissen, Bruggen & Jochems, 2012). Some Literature related to the educational data mining is given below. Gorissen and their group used the educational data mining technique and analyzed the student interaction with the record learning (Gorissen, Bruggen & Jochems, 2012). The lecture capture system is used for data and they combine and used that data with the survey data. They found variations and similarities between the actual data and students reports. The proposed data for the students have a bigger chance of qualifying the

exam. At the end they get a result and say that the difference between the actual data and the verbal report, the analysis will no longer depend on verbal report. Jovanovica and his team applied and tested the classification models which were used for the predicting the performance of students. And also applied the models related to clustering for the categorization of students on the basis of their cognitive styles in e-learning environment (Jovanovica et al., 2012). They prove and suggest that the model related to classification is very helpful and useful for students, teachers and business peoples.

Parack also used different algorithms related to data mining for profiling students on the basis of academic results and records like attendance, exam marks, practical exam and team work score etc and cluster them with the help of K-mean method (Parack, Zahid, & Merchant, 2012). They also found that the data mining is very helpful and important to explore and detect the important data relate to the profiling students.

Wu He also explained a technique which he used to exploit huge students data like chat messages and online questions and that data collected automatically with the help of system, that system is called live video streaming system (He, 2013). The results of text mining and data, acknowledge interesting patterns in the interaction of students and teachers. At the result the students related to health and education asked more questions as compared to the science and engineering students. Two courses related to education also have positive correlation between the final grades of students and number of questions.

Now we discuss the text mining and its role and importance. Text mining is the extracting and finding the important or useful data, directions, models, patterns, trends from the unordered text like HTML file, emails, text document etc. Different applications of text mining are summarization of text, analysis of links, extraction of information and data, clustering (Ananiadou, 2008). The text mining is also used for the extraction, integration, identification, and managing the data related to

the education and research. Text mining is an automated technique (Abdousa, 2010). Now a days, there are some knowledge related to text mining that emphasizes how we use this method for the purpose of test learning data. Tane and their group used the text mining methods and clustered data related to the learning according to the similarities and titles (Tane, Schmitz, & Stumme, 2004). Hung also used the clustering method to test the visualized patterns and e-learning literature (2012). Goda also presented a model related to the prediction of grades on the basis of comments of students. This method used the PCN technique. The result of this experiment shows that the students get lower PCN scores, then the prediction performance of the student grades becomes lower (Goda et al., 2013). Sorour also proposed a method which is used for the prediction of student's grades and this model is based on c-comments information (Sorour, Mine, Goda, & Hirokawa, 2014). They enforced K-mean method and LSA method on two class information/data and then get the outcomes of the applied methods by finding accuracy of grades.

C-comments are also used for the prediction of student's grades (Goda et al., 2013). C-comments shows the performance and understanding of the subject duration of the class and time of the class. Some examples of C-comments are given below

I fully understood the subject and feel courage to make different functions. I did not complete the entire exercise because I could not learn the methods and time was out.

The data collected related to comments in different classes and number of students is 123. They read Goda's course and they studied 15 lessons. In this experiment we collect comments from students from lesson 7 to lesson 15. The subject of lessons are programming related to C language. Some students are not present there for they did not provide the feedback or comments.

Five grade categories are used for the prediction of a student's grades from their comments to classify their marks. The process was done by having average marks of a student's report assigned 3 times and their attendance rate. The correspondence between grades of a student and the marks of a student are shown in Table 1 that is an output of their tests. For example, there are 40 in grade A with marks between 89 and 80.

Artificial Neural Networks

Artificial Neural Networks are basically inspired by biological neural networks (Jain et al., 1996), this model of neural networks was first proposed in 1943 by Warren McCulloch and Walter Pitts, and they presented their first idea based on math-

Table 1. Students grades and the range of marks

	S	A	B	C	D
	90-100	80-89	70-79	60-69	0-59
0					
5					
10					
15					
20					
25					
30					
35					
40					
45					

emational algorithm specially using threshold logic (McCulloch & Pitts, 1943). Artificial Neural Networks were not very popular at first because of too much computational power was required to process them and they used to take long time to develop a model. But as the time passed availability of computational power became easy but in 1969 Marvin Minsky published two major issues in artificial neural networks one of them was the use of single layer (Minsky & Seymour, 1969). But the concept of Multiple Layers in deep learning that came in late 2000s renewed the scientist's interest in artificial neural networks (Hinton, 2007). The model of Artificial Neural Network is shown in Figure 1.

Artificial Neural Networks usually consist of a large number of inputs and some hidden layers usually one to two, though more hidden layers are possible but that makes the model too complex to implement. Connections between inputs, hidden layers and outputs are usually weighted and these weights are the real part of Artificial Neural Networks we are interested in. Whenever we want to devise a new model we give new inputs at the start, known as training set, then we set a learning rate and start the algorithm. Back Propagation (explained later) method starts adjusting the weights, once these weights become constant up

to the fraction we chose, this model is ready. Now if we want to predict a new value we just give new inputs and use our devised model's weights and the outputs at the end are our new predicted values. Every new model depends only on the weights between these connections. Our goal is to modify the weights between the connections so that they represent the model we want to prepare for our purpose. ANN layers model is shown in Figure 2.

Back Propagation is the method that is usually used to modify the weights or we can say to learn the weights that represents our model. Back Propagation is very slow indeed it can take thousands of iterations to learn the weights. But once the weights are learned and the model is built it makes all the work very easily and the predictions or classification is very fast after that (Werbos, 1974). So we don't care how much time it takes to learn the weights as long as its outputs are correct, which means error rate is very low.

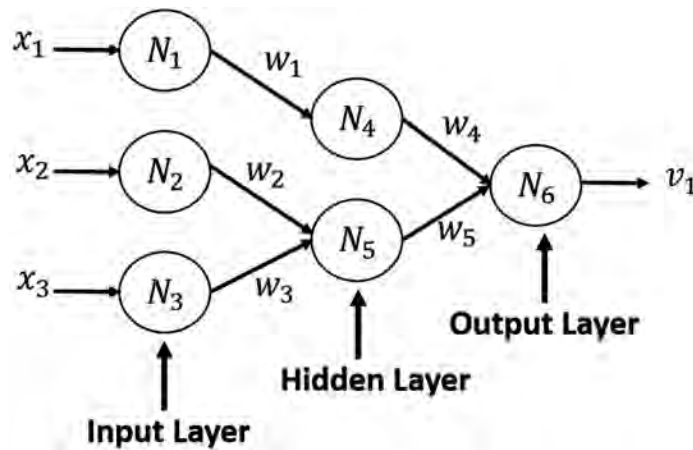
Latent Semantic Analysis

This is also one of the famous neural language processing techniques that is widely being used to find relationship between the words. It works on the fact that if words are similar in meaning it means they can be correlated (Landauer et al.,

Figure 1. Artificial neural network model



Figure 2. ANN layers model



1998). Latent Semantic Analysis usually works in the form of matrix in which element (i, j) describes the occurrence of term i in document j (this can be, for example, the frequency). X Will look like this (Dougherty, Natow & Vega, 2012).

d_j
↓

$$t_i^T \rightarrow \begin{bmatrix} x_{1,1} & \dots & x_{1,n} \\ \vdots & \ddots & \vdots \\ x_{m,1} & \dots & x_{m,n} \end{bmatrix}$$

In our case instead of using different documents we have to use different aspects coming from same data source.

Support Vector Machines

Support Vector Machines are based on supervised learning model, which means some of our data which we intend to use as our training set (Hearst et al., 1998). Support vector machines are used for categorization of hypertext and text and also categorize their applications. Support vector machines are also used for transductive settings and also for standard inductive settings. Support vector

machines is also used for labeling the training data. We can also use the Support vector machines for image classification. Support vector machines are best from other search schemes because the accuracy of the Support vector machines are higher than the other methods and they provide the best and fast results compare to the others machines or methods Support vector machine is also useful in medical field. In support vector machine we basically find support vectors, that are tuples from the original data, to define a hyper plane that classifies our data in two portions (In our case, is this habit good or bad for child). Support Vector Machines model is shown in Figure 3.

We want to maximize the margin between two dotted lines and the middle one solid line is the hyper line, in case of multidimensional data it's a hyper plane, is the one that classifies our data. Finding this hyper plane can be very tricky sometimes and also requires a lot of computations. To reduce the computations we use different kernel tricks (Tang, Guo, & Jinghuai, 2009), like one of them is shown in the Figure 4 below.

IMPLEMENTATION

In this section we discuss the methodology to predict the performance of student. We use Latent Semantic Analysis (Landauer, Foltz, &

Figure 3. Support vector machine model

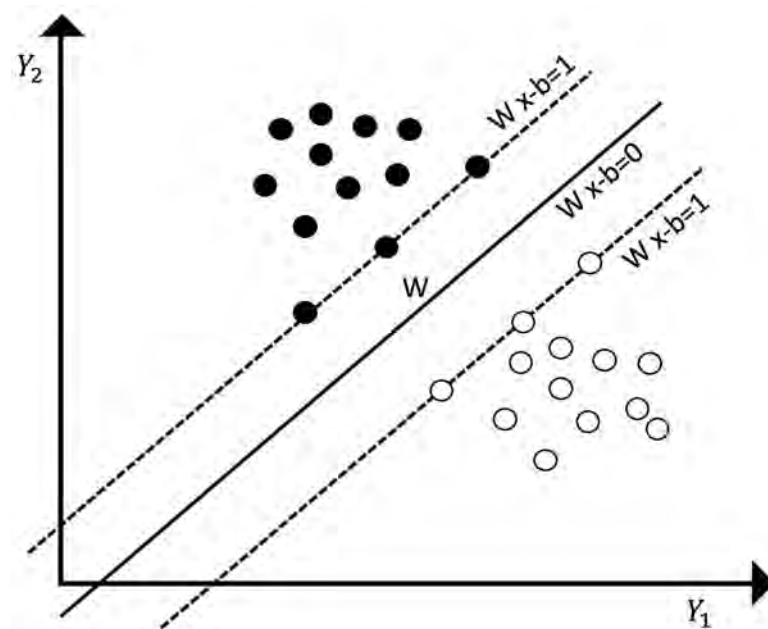
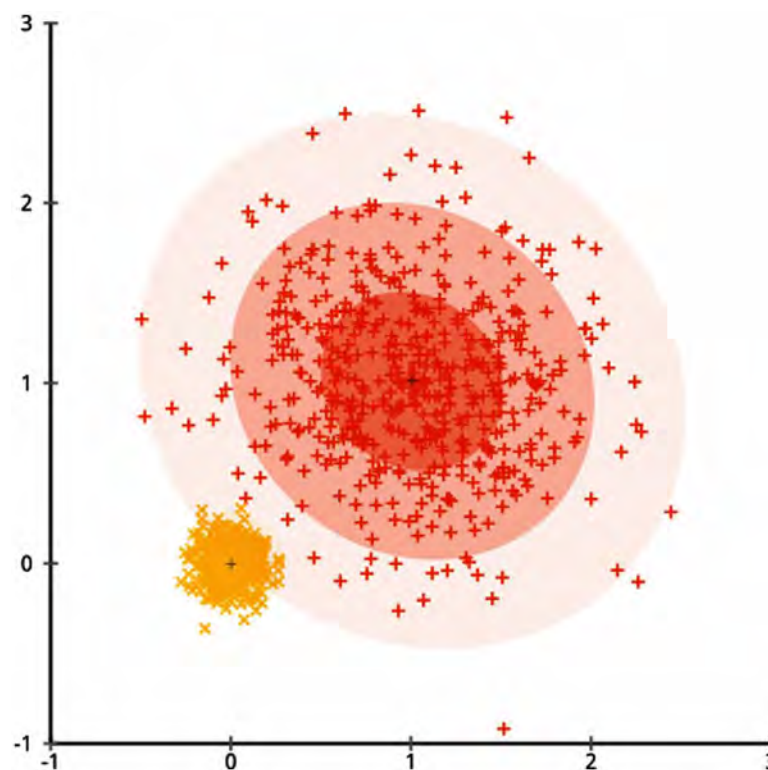


Figure 4. Gaussian Kernel model



Laham, 1998) technique for prediction of student performance. In our work first we analyze our data and then we use) Latent Semantic Analysis method to predict the results of student and in the end we use Artificial Neural Networks model for classification. Our implementation needs the following steps.

1. First we collect the data related to the students like his remarks after the lecture, which includes: Level of interest, level of understandability, level of difficulty, practical knowledge. Teacher's remarks about student's participation, which includes: Behavior, social skills, communication, reading writing.
2. In the second phase we build a data set from the student remarks and give a fine shape to this data set. So in this part we separate all the parts of speech and find the occurrence of the work in the remarks or comments. After this we still need data cleaning as there can be a lot of empty occurrences or typing errors that prevents this data to fall into any category and now apply different methods like entropy to equalize or balance the frequencies occurrences. This data cleaning step is also known as preprocessing.
3. Now we apply the Latent Semantic Analysis model to find the relationship between the different words and analyze the words.

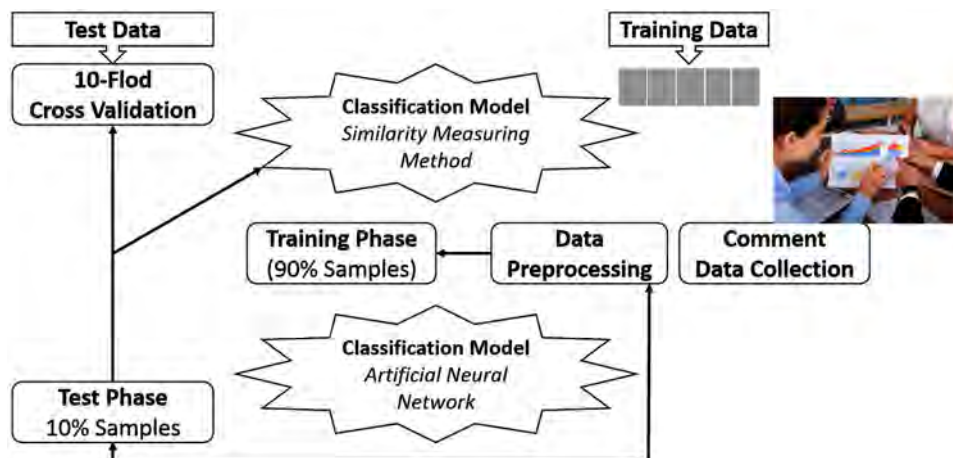
Further details of this step are discussed in later part of the paper.

4. When we get the results after applying the Latent Semantic Analysis model then we use the Artificial Neural Networks model for the classification of words or comments. This is also called the training phase.
5. In this phase we find the accuracy of the model / classifier. This phase is called testing phase. Flow of work model is shown in Figure 5.

For testing and finding accuracy of the model we use the cross validation method and we select K-folds for cross validation (Kohavi, 1995). For example we select 10 folds the then we select 9 for training data and the remaining 1 fold for test data, selection of data in each fold can also effect the output of our model. So while selecting data we consider two type of selections.

1. Random selection of data for every fold, which is naïve and enhances the chances of error, so it is not appreciated.
2. Supervised selection of data, in this method we distribute the data on basis of its type so that every fold of data has all types of data available in it. It is called well distributed data and can be handful to get better results, we use this type of distribution.

Figure 5. Work flow of model



Now let's see how our method works. According to the method first of all we find the similarities between the data. For this we analyze the data and we find this data contains different verbs, objects, nouns etc. Now in next step we create a vector model of these words. Vector model is actually the matrix which contains the words and their frequencies or weights. When we build the matrix then we normalize or balance the weights in the matrix. For this we use the global frequency or weight. The formula used to find the global weigh is

$$g_i = 1 + \sum_{j=1}^n (p_{ij} \log(p_{ij})) / \log(n)$$

g_i In the formula is basically an overall weight found from the whole set of remarks for this model. Whereas p_{ij} is the number of occurrences of word 'i' in a remark 'j' and 'n' is the total number of remarks used to devise the model. Flow diagram for finding similarities between the data is shown in Figure 6.

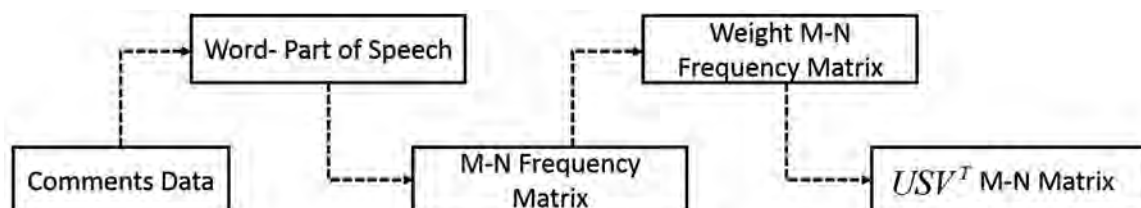
Latent Semantic Analysis is a powerful technique which is used to represent the language. This technique is basically representation of our English language and grammar data in the form of mathematics (Dearwester et al., 1990). There are three levels or steps of Latent Semantic Analysis,* the first is to simulate the natural language and find the frequency of every single word in the data, then decompose the data and build the matrix and show the words in the matrix along with their weights, and the last step is that the data is shown in the matrix in the form of a vec-

tor Latent Dirichlet Allocation work with the help of decomposition of a single value (Horn & Axel, 2003). For Latent Dirichlet Allocation we use the given formula $A = USV^T$. Latent Semantic Analysis; is used for the new data or comment and the formula for Latent Semantic Analysis is $q = q^T U_K S_{k-1}$.

Now when the data is divided and the method is applied on this data then we build different categories of the data and we give names to those clusters like cluster 1, cluster 2 and so on. Then we store the data in different clusters with the help of K-Mean clustering method. When the Latent Semantic Analysis is applied on the data then we perform this clustering method. We build the clusters according to the comments like we say that 62 % of one type of data lies in cluster 1 and 43 % * of data lies in cluster 2 and so on. According to our experiment we say that we have 5 grades like A + , A , B + , B and C + . Now 62 % of data in cluster 1 is of one type let's say A+ and 43 % of data in cluster 2 is A+, this is just a supposition as in our case every time different data comes in and percentage can change according to our data.

Every cluster has its own center which is called centroid (Jain, Murty & Flynn, 1999). Now we calculate the distance of each comment with the all cluster centers also known as centroid and the comment is placed in the cluster who's centroid has the least distance from the notation of this comment and this process is continued until all the comments are finished (Jain, Murty & Flynn, 1999). When the construction of clustering* phase is done then we perform the k-fold cross validation process.

Figure 6. Model of work flow



Now we have to calculate the prediction for the new item or comment. When a new comment occurs we have to decide in which cluster we should store this comment. For this purpose we use the measurement of similarity* method (Resnik, 1995). The formula used for finding similarity is

$$\text{Similarity} = \frac{S_{new} * S_k}{\|S_{new}\| * \|S_k\|} = \frac{S_{new} * S_k}{\sqrt{\sum_{i=1}^k S_{new}^2} * \sqrt{\sum_{i=1}^k S_k^2}}$$

In this formula S_{new} is the new item and S_k is the number of items in the cluster. Now we select the * Support Vector Machines to classify our data. The given diagram shows the details for the Support Vector Machines that how Support Vector Machines work in our method. When we collect the data and all the results are found then we give these results as an input to the Support Vector Machines which contains the pervious academic performance, personality information and biographical data of the student, these three categories are assigned to the Support vector machine as an input. The output of this Support Vector Machine that we designed also contains three sections behaviors, abilities of language and future academic performance. The detail Figure 7 is given below.

Date of birth and *gender are the example of bibliographical data. Past academic records means

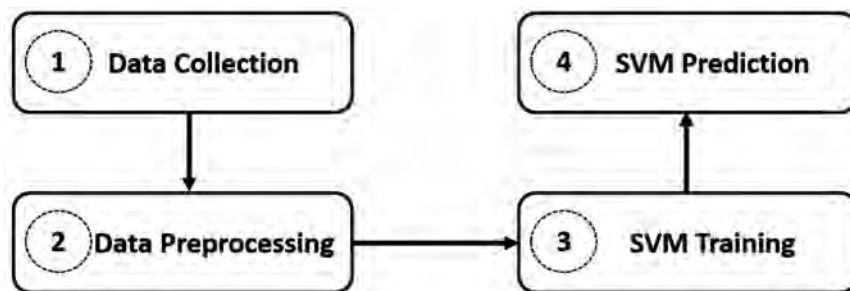
that the students*grades or result of previous exam. Output results contains the future academics performance which means that the grades or marks of student in future subject like math, data mining etc. behavior tells us the abilities of the student like the student have leadership quality etc. for further explanation gender means that female or male, order of birth means that among the brothers and sisters the student order of birth. Support Vector Machine method was used to solve this problem which gets the input features or data and finds the correlation between different features and provide the final result related to the target results. The flow of the operations is show in the given diagram. First we collect the data then in second step we perform preprocessing on that data it means that we clean our data and remove all the noise from the data and we set the data for future processing. Now we select some data for training and select some data for testing. The flow of process is show in the Figure 8.

At the end we use the Artificial Neural Networks for the prediction and classification or selection of data it means that when the Latent Semantic Analysis operations are finished then we use three level perceptron for the calculation for student grads and performance. We create a perceptron for every input. The first layer shows the input layer, each input to the Artificial Neural Networks is the result of Latent Semantic Analysis. The middle layer is called hidden layer which contain a lot of neurons. The third layer

Figure 7. SVM Presiction model



Figure 8. Experiment procedure



is called output layer which gives us the results. In the diagram we get the 5 output results which go to its own neuron and each neuron is called the grade of student. Like grade S to grade C. Artificial Neural Networks with different Layers and outputs are shown in the Figure 9.

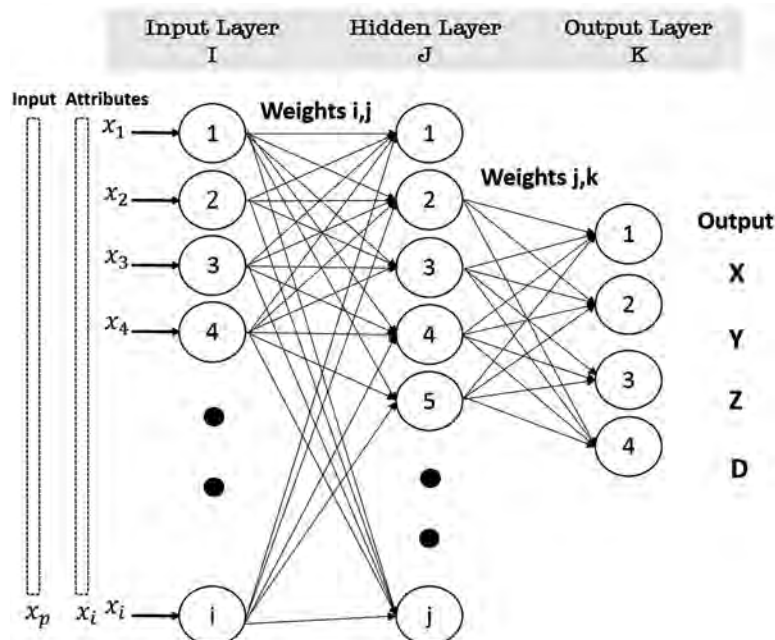
The back propagation plays a very important role in the Artificial Neural Networks because it updates the weights with the help of gradient descent. In Artificial Neural Networks we use the learning rate which is used to adjust or decrease the error and also the weights of the paths. In our case learning rate was kept low to get better results

so it is comparatively a slower method but gives us more accurate results. We perform the operations again and again until the error become minimized.

FUTURE RESEARCH DIRECTIONS

In future a greater set of remarks can be used to find a better model. We are also planning to use Dynamic Programming, another very popular algorithm, to predict the grades and analyze the impact of different habits of student on his final grades.

Figure 9. ANN with different layers and outputs





CONCLUSION

Latent Semantic Analysis when combined with Support Vector Machine found to be very useful. Overall we found that after matching our predicted values with the labeled data our prediction model gives very accurate results with quite an improvement over the previously used model. We deduced that, because of Support Vector Machines we are getting better results but we used Artificial Neural Network system with slow learning rate, its performance was better but overall speed of the whole learning process was slow so still there is a room for improvement especially in the process of learning speed.

ACKNOWLEDGMENT

The authors would like to thank D. Khalid Latif, Assistant Professor, School of Electrical Engineering and Computer Sciences (SEECs), National University of Sciences and Technology (NUST) Islamabad Pakistan.

REFERENCES

Abdousa, M. (2010). A predictive study of learner satisfaction and outcomes in face-to-face, satellite broadcast, and live video-streaming learning environments. *The Internet and Higher Education*, 13(4), 248–257. doi:10.1016/j.ihe-duc.2010.04.005

Ananiadou, S. (2008). *National centre for text mining: Introduction to tools for researchers*. Retrieved from <http://www.jisc.ac.uk/publications/publications/>

Baker, R. S. J. D. (2009). *Data mining for education*. In *International Encyclopedia of Education* (3rd ed.). Oxford, UK: UK Elsevier.

Baker, R. S. J. D., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining*, 1(1).

Blei, D. M., Andrew, Y. N., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 31(3), 993–1022.

Bridgeman, B., Pollack, J., & Burton, N. (2008). *Predicting grades in different types of college courses*. College Board Research Report.

Burton, N. W., & Ramist, L. (2001). Predicting success in college: SAT studies of classes graduating since 1980. *College Board Research Report*, 2.

Jain, A.K., & Mohiuddin, K.M. (1996). Artificial Neural Networks. A Tutorial. *IEEE Computers & Society*, 29(3).

Cooper, C. I. (2010). Predicting Persistence of College Freshmen Using Neural Networks. *Technological Developments in Education and Automation*, 145–148.

Dearwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by Latent Semantic analysis. *Journal of the American Society for Information Science*, 41(6), 391–407. doi:10.1002/(SICI)1097-4571(199009)41:6<391::AID-ASII>3.0.CO;2-9

Dougherty, K. J., Natow, R. S., & Vega, B. E. (2012). Popular but Unstable: Explaining Why State Performance Funding Systems in the United States Often Do Not Persist. *Teachers College Record*, 114(3), n3.

Educational Attainment in the United States. (2012). Retrieved from <http://www.census.gov/hhes/socdemo/education/data/cps/2012/tables.html>

Goda, K., Hirokawa, S., & Mine, T. (2013). *Correlation of grade prediction performance and validity of self-evaluation comments*. Academic Press.

- Gorissen, P., Bruggen, J. V., & Jochems, V. (2012). Usage reporting on recorded lectures. *International Journal of Learning Technology*, 7(1), 23–40. doi:10.1504/IJLT.2012.046864
- He, W. (2013). Examining students online interaction in a live video streaming environment using data mining and text mining. *Computers in Human Behavior*, 29(1), 90–102. doi:10.1016/j.chb.2012.07.020
- Hearst, M. A., Dumais, S. T., Osman, E., Platt, J., & Scholkopf, B. (1998). Support Vector Machines. *IEEE Intelligent Systems & their Applications*, 13(4), 18–28. doi:10.1109/5254.708428
- Hinton, G. H. (2007). Learning Multiple Layers of Representation. *Trends in Cognitive Sciences*, 11(10), 428–434. doi:10.1016/j.tics.2007.09.004 PMID:17921042
- Horn, D., & Axel, I. (2003). Novel clustering algorithm for microarray expression data in a truncated SVD space. *Bioinformatics (Oxford, England)*, 19(9), 1110–1115. doi:10.1093/bioinformatics/btg053 PMID:12801871
- Hung, J. (2012). Trends of e-learning research from 2000 to 2008: Use of text mining and bibliometrics. *British Journal of Educational Technology*, 43(1), 5–15. doi:10.1111/j.1467-8535.2010.01144.x
- Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: A review. *ACM Computing Surveys*, 31(3), 264–323. doi:10.1145/331499.331504
- Jovanovica, M., Vukicevica, M., Milovanovica, M., & Minovica, M. (2012). Using data mining on student behavior and cognitive style data for improving e-learning systems: A case study. *International Journal of Computational Intelligence Systems*, 5(3), 597–610. doi:10.1080/18756891.2012.696923
- Kohavi, R. (1995). *A study of cross-validation and bootstrap for accuracy estimation and model selection*. Paper presented at the International Joint Conference on Artificial intelligence.
- Landauer, K. T., Foltz, P. W., & Laham, D. (1998). An Introduction to Latent Semantic Analysis. *Discourse Processes*, 25(2-3), 259–284. doi:10.1080/01638539809545028
- Long, P., & Siemens, G. (2011). Penetrating the analytics in learning and education. *EDUCAUSE Review*, 32–40.
- Marvin, M., & Papert, S.A. (1969). *A seminal work that developed the modern theory of computational geometry and established fundamental limitations of loop-free connectionist learning machines*. Academic Press.
- McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 5(4), 115–133. doi:10.1007/BF02478259
- Minsiky, M., & Seymour, P. (1969). Retrieved from <http://hdl.handle.net/1721.1/6701>
- National Conference of State Legislatures. (2012). Retrieved from <http://www.ncsl.org/research/education/performance-funding.aspx>
- Parack, S., Zahid, F., & Merchant, Z. (2012). *Application of data mining in educational databases for predicting academic trends and patterns*. Paper presented at the IEEE International Conference on Technology Enhanced Education. Retrieved from <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6208617>
- Resnik, P. (1995). *Using Information Content to Evaluate Semantic Similarity in a Taxonomy*. Paper presented at the 14th International Joint Conference on Artificial Intelligence.
- Romero, C., Espejo, P. G., Zfra, A., Romero, J. R., & Ventura, S. (2010). Web usage mining for predicting final marks of students that use model courses. *Computer Applications in Engineering Education*.

Sorour, S. E., Mine, T., Goda, K., & Hirokawa, S. (2014). *Efficiency of lsa and k-means in predicting students' academic performance based on their comments data*. Paper presented at the International Conference on Computer Supported Education.

Tane, J., Schmitz, C., & Stumme, G. (2004). *Semantic resource management for the web: An e-learning application*. Paper presented at the WWW conference, New York, NY. doi:10.1145/1013367.1013369

Tang, Y., Guo, W., & Jinghuai, G. (2009). Efficient Model Selection for Support Vector Machine with Gaussian Kernel Function. *Computational Intelligence and Data Mining, IEEE Symposium*, 40-45.

Uhler, B. D., & Hurn, J. E. (2013). Using learning analytics to predict (and improve) student success: A faculty perspective. *Journal of Interactive Online Learning*, 1(1), 17-26.

Umair, S., Muneer, U., Zahoor, M. N., & Malik, A. W. (2015). *Mobile computing: issues and challenges*. Paper presented at the 12th International Conference on High-capacity Optical Networks and Enabling/Emerging Technologies (HONET), Islamabad, Pakistan. doi:10.1109/HONET.2015.7395438

Vapnik, V. (1995). *The Nature of Statistical Learning Theory*. Springer. doi:10.1007/978-1-4757-2440-0

Venezia, A., Callan, P. M., Finney, J. E., Kirst, M. W., & Usdan, M. D. (2005). *The governance divide: A report on a four-state study on improving college readiness and success*. Academic Press.

Werbos, P. J. (1974). *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences*. Harvard University.

Wigdahl, J., Heileman, G. L., Slim, A., & Abdallah, C. T. (2014). *Curricular efficiency: What role does it play in student success?* Paper presented at the 121st ASEE Annual Conference and Exposition IEEE, Indianapolis, IN.

Zhang, L. (2009). Does state funding affect graduation rates at public four year colleges and universities? *Educational Policy*, 23(5), 714-731. doi:10.1177/0895904808321270



ADDITIONAL READING

Blei, D. M., Andrew, Y. N., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 31(3), 993-1022.

Bridgeman, B., Pollack, J., & Burton, N. (2008). *Predicting grades in different types of college courses*. Retrieved from College Board Research Report.

Burton, N. W., & Ramist, L. (2001). Predicting success in college: SAT studies of classes graduating since 1980. *College Board Research Report*, 2.

Cooper, C. I. (2010). Predicting Persistence of College Freshmen Using Neural Networks. *Technological Developments in Education and Automation*, 145-148.

Educational attainment in the United States. (2012). Retrieved from <http://www.census.gov/hhes/socdemo/education/data/cps/2012/tables.html>

Ful, J. H., Chang, J. H., Huang, Y. M., & Chao, H. C. (2012). *A Support Vector Regression-based Prediction of Students' School Performance*.

Marvin, M., & Papert, S. A. (1969). Perceptron's. *A seminal work that developed the modern theory of computational geometry and established fundamental limitations of loop-free connectionist learning machines*.

National Conference of State Legislatures. (2012). Retrieved from <http://www.ncsl.org/research/education/performance-funding.aspx>

Umair, S., Muneer, U., Zahoor, M. N., & Malik, A. W. (2015). *Mobile computing: issues and challenges*. Paper presented at the 12th International Conference on High-capacity Optical Networks and Enabling/Emerging Technologies (HONET), Islamabad Pakistan. doi:10.1109/HONET.2015.7395438

Vapnik, V. (1995). *The Nature of Statistical Learning Theory*. Springer. doi:10.1007/978-1-4757-2440-0

Venezia, A., Callan, P. M., Finney, J. E., Kirst, M. W., & Usdan, M. D. (2005). *The governance divide: A report on a four-state study on improving college readiness and success*. CA: Retrieved from San Jose.

Wigdahl, J., Heileman, G. L., Slim, A., & Abdallah, C. T. (2014). *Curricular efficiency: What role does it play in student success?* Paper presented at the 121st ASEE Annual Conference and Exposition IEEE. Indianapolis.

Zhang, L. (2009). Does state funding affect graduation rates at public four year colleges and universities? *Educational Policy*, 23(5), 714–731. doi:10.1177/0895904808321270

KEY TERMS AND DEFINITIONS

ANN: ANN stands for Artificial Neural Network. Artificial Neural Networks usually consist of a large number of inputs and some hidden layers usually one to two. Connections between inputs, hidden layers and outputs are usually weighted and these weights are the real part of Artificial Neural Networks.

EDM: EDM stands for Educational Data Mining. The educational data mining is an emerging field. In this field we use and developed different type of methods which are used to investigate the different data types.

LSA: LSA stands for Latent Semantic Analysis which is famous neural language processing techniques that is widely being used to find relationship between the words. It works on the fact that if words are similar in meaning it means they can be correlated.

SVM: SVM stands for Support Vector Machines which is supervised learning model, which means some of our data which we intend to use as our training set. Support vector machines are used for categorization of hypertext and text and also categorize their applications. Support vector machines are also used for transductive settings and also for standard inductive settings.

The Relationship Between Online Formative Assessment and State Test Scores Using Multilevel Modeling



Aryn C. Karpinski
Kent State University, USA

Jerome V. D'Agostino
The Ohio State University, USA

Anne-Evan K. Williams
Billings Middle School, USA

Sue Ann Highland
Grand Canyon University, USA

Jennifer A. Mellott
Kent State University, USA

INTRODUCTION

The main goal of the current study was to examine the relationship between online formative assessments (FAs) and summative, yearly state proficiency test scores. Specifically, the relationship between one online formative assessment (FA) program in reading, known as the Diagnostic Online Reading Assessment (DORA), and state test scores in reading (i.e., the Colorado Student Assessment Program [CSAP]) was examined in four cohorts across elementary, middle, and high school in beginning in the 2004/2005 academic year and ending in 2009/2010. This investigation used Hierarchical Linear Growth Modeling (HLGM; i.e., Multilevel Modeling) to address the following research question: (1) What is the relationship between online formative assessment score growth and state test score growth?

Formal and informal FAs are one of many teaching methods that have been used to increase student performance on end-of-course, academic year, and other high-stakes achievement tests for decades and has a large research base to support

these practices (e.g., Black & Wiliam, 1998a). Additionally, summative assessment data (e.g., yearly state proficiency tests) are continually used as indicators of school and district performance for policymakers and the public. However, these summative data are of little use in the day-to-day activities of teachers in diagnosing student learning progress and modifying teaching strategies, as is done in the FA process (e.g., Black, 2015). Because this collection of abstract theories and research methods have transitioned into actual teaching practices, it is important to build the literature surrounding technology-based methods as teachers continue to use FA in the classroom.

The purposes in conducting this study include the following: (1) To support the burgeoning literature outlining the role of technology in general in teaching and learning, and (2) To bolster support for federal initiatives and administrative demands for more efficient ways to meet state standards. As technology-based assessment is gradually used to support and/or replace traditional forms of evaluation, the need to examine the extent to which these methods are educationally sound is

in high demand. Overall, information presented in this study can provide practical implications to district-wide implementation of supplemental reading instruction in an online environment.

BACKGROUND

E-learning (i.e., learning that is facilitated by electronic technologies) is referred to as part of the equipment of 21st Century scholarship (Buzzetto-More & Guy, 2006). However, e-learning is only half of the equation as government mandates have required schools to use data to inform decision making. The use of data has necessitated the development of improved information technology and access to computers and high-speed Internet in schools (Petrides, 2006). Thus, the other half of the equation is the use of data rendered from e-learning, or e-assessment, which entails using electronic technologies to drive student learning and assessment as with FA (Ridgway, McCusker, & Pead, 2004).

FA can be briefly defined as the use of diagnostic formal and informal assessments to provide feedback to teachers and students over the course of instruction for the purpose of improving performance and achievement (e.g., Black, 2015; Boston, 2002). Previous research in this area has primarily focused on traditional FA practices (e.g., paper-and-pencil quizzes), with the current literature beginning to examine the effectiveness of Internet-based, automated FA programs (e.g., Chua & Don, 2013; Kingston & Nash, 2011). The overall consensus from the traditional body of literature is that FA is an essential component of classroom procedure, and that its proper use can raise standards and achievement (e.g., Black & Wiliam, 1998a; Carlson, Borman, & Robinson, 2011; Gulikers, Biemans, Wesselink, & van der Wel, 2013; Merino & Beckman, 2010), with the latest studies of technology-based FA beginning to echo these findings. Many theories have attempted to describe FA in terms of multilevel relationships (i.e., students, teachers, schools, school districts,

etc.), with few studies focusing on statistically accounting for these nested associations, and hardly any examining technology-based FA practices (Black & Wiliam, 2009).

Regarding previous studies of online FA, the overwhelming majority of these studies have examined college-age populations in the university setting, usually within one course (e.g., Buchanan, 2000; Jenkins, 2004). In addition, past and current FA research has thoroughly examined the relationship between measures of FA and performance on a summative, usually end-of-course or final exam, but not state proficiency test scores. This area of research is just beginning to use more sophisticated statistical analyses, which is in contrast to the many qualitative studies summarizing student perceptions of a technology-based platform for quizzes/exams (Hunt, Hughes, & Rowe, 2002; Peat & Franklin, 2002). Additionally, due to the novelty of the mode of online or computerized administration, understandably research is lacking in longitudinal data analysis, with few studies examining multiple years of data across several cohorts.

METHODS

Existing DORA data were provided from one school district in Colorado from an online testing company, and existing CSAP data were provided from the same school district by the Colorado Department of Education. The data were selected via collaboration with the testing company and one school district in Colorado. The selected school district was one that gave permission to use their student demographic information and state test scores. It was necessary to have permission from both parties as the only way to examine correlated growth is to link the data via a shared student ID number. Additionally, this particular school district was selected because they had fewer missing data, with all students having graduated or left the school district at the present time. Data were linked anonymously producing four cohorts

of students across grades 3 through 11. It was hypothesized that student FA score growth will be significantly and positively related to student state test score growth.

The analytical method used was a *Two-Level Time-Varying Covariate* HLGM (Singer & Willett, 2003). According to Raudenbush and Bryk (2002), investigating a relationship with a growth trajectory of another variable of interest is common practice in multilevel modeling. Time-varying covariates are defined as person-level characteristics that are measured and may change over time, and are related to the outcome (O'Connell & McCoach, 2004). Thus, the measurements across time and other Level 1 predictors form a nested structure when combined with other student level variables (i.e., Level 2).

Time (i.e., in months with one unit being every 3 to 6 months) and DORA scores (i.e., the time-varying covariate) were used in Level 1. Demographic covariates such as gender (i.e., SEX), ethnicity (i.e., ETHNIC), SES (i.e., FREERED,

which is Free/Reduced lunch status), and ESL/ELL status were incorporated into Level 2 of the model. CSAP scores were used as the outcome variable (see Table 1). Four models were run using the DORA subtests (i.e., Word Recognition [WR], Oral Vocabulary [OV], Spelling [SP], and Reading Comprehension [RC]) as the time-varying covariate in separate models. These subtests were chosen because they did not have ceiling effects compared to the other subtests.

DATA

Existing DORA and CSAP data were used ($N = 208$) from 2004/2005 to 2009/2010 for students in grades 3 through 11. For the CSAP, students' test scores from the spring of 2005 were the first data point. Two state reading test scores were obtained before DORA implementation and three afterwards. For DORA, students' test scores from fall of 2006 represented the first data point.

Table 1. Hierarchical linear growth model level 1 and level 2 formulae and explanation of symbols

Formula Component	Explanation
Level 1 Model	$Y_{ti} = \pi_{0i} + \pi_{1i}(\text{DORA})_{ti} + \pi_{2i}(\text{Time})_{ti} + e_{ti}$
Y_{ti}	Student's CSAP score for time t for student i
$(\text{Time})_{ti}$	Elapsed years/months since DORA implementation
$(\text{DORA})_{ti}$	Time-varying predictor for a student at a given time point
π_{0i}	Intercept/Student's initial CSAP score
π_{1i}	Linear growth coefficient (for DORA)
π_{2i}	Growth rate over all years/months
e_{ti}	Individual student error
Level 2 Model	$\pi_{0i} = \beta_{00} + \beta_{01}(\text{SEX})_i + \beta_{02}(\text{ETHNIC})_i + \beta_{03}(\text{ESLELL})_i + \beta_{04}(\text{FREERED})_i + r_{0i}$ $\pi_{1i} = \beta_{10} + \beta_{11}(\text{SEX})_i + \beta_{12}(\text{ETHNIC})_i + \beta_{13}(\text{ESLELL})_i + \beta_{14}(\text{FREERED})_i + r_{1i}$ $\pi_{2i} = \beta_{20} + \beta_{21}(\text{SEX})_i + \beta_{22}(\text{ETHNIC})_i + \beta_{23}(\text{ESLELL})_i + \beta_{24}(\text{FREERED})_i + r_{2i}$
π_{0i}	Individual-specific CSAP score parameter (initial status)
π_{1i}	Individual-specific CSAP score parameter (DORA growth)
π_{2i}	Individual-specific CSAP score parameter (growth rate)
β_{00}	Baseline expectation (initial CSAP status) for the demographics coded 0
β_{10}	Expected change of the CSAP controlling for the DORA time-varying covariate
β_{20}	Expected change of the CSAP for the demographic predictors coded as 0
r_{0i}, r_{1i}, r_{2i}	Residuals

Students were tested approximately three times during the school year, with a possible total of 11 DORA assessments for the years investigated in the current study. See Table 2 for the sample demographic information.

MEASURES

DORA tests were across seven subtests: (1) High-Frequency Words, Phonics, Phonemic Awareness, Word Recognition (WR), Oral Vocabulary (OV), Spelling (SP), and Reading Comprehension (RC). With the exception of the SP subtest which is a generative test, all test items are multiple-choice. DORA results are returned as grade-level equivalencies. The CSAP is administered each spring, yielding a single, scaled score (i.e., reading score in the current study) for each student every year. The state scores were based on a scale that ranged from 0 to 1000 depending on the grade level assessed, with a score of approximately 550 as the cut-point for proficiency each year. At each grade level, students are assessed using 40 to 70 multiple choice items depending on the grade level, developed to assess student knowledge of grade-level indicators, identified as the Colorado Model Content Standards for that particular grade level. The tests across grades were vertically equated.

RESULTS

The main assumptions in Hierarchical Linear Growth Modeling pertain to the functional form of the model examining linearity, and the stochastic part of the model involving normality and homoscedasticity. Compound symmetry (i.e., a sufficient condition for sphericity; Maxwell & Delaney, 1990) can be relaxed in a multilevel framework (Raudenbush & Bryk, 2002). Linearity, normality, and homogeneity of variance (i.e., homoscedasticity), however, are assumptions that are typical of any General Linear Model (GLM) approach. Thus, these assumptions were checked

Table 2. Demographic information of the sample (N = 208) for grades 3 through 11 across the 2004/2005 to 2009/2010 academic years

Demographic Information	n (%)
Cohort	
1	47 (22.6)
2	52 (25.0)
3	48 (23.1)
4	61 (29.3)
Gender	
Male	104 (50.0)
Female	104 (50.0)
Ethnicity	
White (Non-Hispanic)	142 (68.3)
Hispanic	60 (28.8)
Black (Non-Hispanic)	3 (1.4)
Asian/Pacific Islander	2 (1.0)
American Indian/Alaskan Native	1 (.5)
Free/Reduced Lunch	
Eligible	93 (44.7)
Not Eligible	115 (55.3)
English as a Second Language/English Language Learner	
Yes	26 (12.5)
No	182 (87.5)

and upheld to ensure unbiased estimates of population effects.

The full model-building strategy was implemented for each DORA subtest as the time-varying covariate (i.e., One-way Random Effects Analysis of Variance [ANOVA], Unconditional Model, Conditional Growth Model, Full Model; Raudenbush & Bryk, 2002; Snijders & Bosker, 2012). Only the Full Models will be discussed below (see Table 3). Each column in the table provides the results from one HLGGM analysis (i.e., one per DORA subtest). The first row in the table provides the estimated average student DORA scores at Time 0, the standard error for these estimates (in parentheses), and whether the estimates were significantly greater than zero.

Table 3. Hierarchical linear growth models for Colorado Student Assessment Program (CSAP) scores and time predicting online formative assessment (Diagnostic Online Reading Assessment [DORA]) subtests ($N = 208$)

Fixed Effects	Word Recognition	Oral Vocabulary	Spelling	Reading Comprehension
Initial CSAP Status (π_{0i})				
Intercept (β_{00})	634.38 (4.94)***	632.19 (5.08)***	637.47 (4.66)***	634.33 (4.70)***
Sex (β_{01})	5.94 (6.27)	9.66 (6.17)	8.40 (5.85)	5.33 (6.03)
Ethnicity (β_{02})	-2.85 (7.80)	4.75 (7.10)	-4.49 (7.17)	3.15 (7.32)
ESL/ELL (β_{03})	-36.56 (10.51)**	-36.58 (11.23)*	-41.19 (10.30)***	-20.37 (11.35)
Free/Reduced Lunch (β_{04})	-16.13 (6.65)*	-16.73 (6.44)**	-16.24 (6.13)**	-16.88 (6.79)*
DORA Growth Rate (π_{1i})				
Intercept (β_{10})	1.22 (.84)	.31 (.84)	4.16 (1.09)***	2.93 (.66)***
Sex (β_{11})	.78 (1.03)	2.77 (1.18)*	1.81 (1.33)	-.14 (.94)
Ethnicity (β_{12})	-1.05 (1.58)	2.99 (1.60)	-1.47 (1.57)	1.05 (1.13)
ESL/ELL (β_{13})	2.01 (1.97)	-1.66 (2.23)	-2.30 (2.31)	1.91 (1.65)
Free/Reduced Lunch (β_{14})	.44 (1.02)	.36 (1.36)	-.90 (1.45)	-.64 (1.02)
CSAP Growth Rate (π_{2i})				
Intercept (β_{20})	11.89 (1.45)***	12.43 (1.52)***	9.66 (1.52)***	9.15 (1.39)***
Sex (β_{21})	.78 (1.88)	-1.20 (2.03)	-.97 (1.99)	1.68 (1.97)
Ethnicity (β_{22})	.13 (2.25)	-2.97 (2.18)	-.43 (2.24)	-2.44 (2.25)
ESL/ELL (β_{23})	-1.18 (3.45)	1.91 (3.58)	2.69 (3.69)	-3.67 (3.98)
Free/Reduced Lunch (β_{24})	.61 (1.89)	1.09 (1.96)	2.48 (1.88)	2.42 (2.12)

Note. * $p < .05$; ** $p < .01$; *** $p < .001$. The columns are the four DORA reading subtests. ESL/ELL = English as a Second Language/English Language Learner.

All initial DORA average scores across the four subtests were significantly greater than zero. The growth rates for all subtests were significantly greater than zero. Finally, the DORA and CSAP covariation results indicated that student gain in DORA over time covaried positively and significantly with their CSAP gain on the SP and RC subtests. In comparing the growth rates for SP and RC, both were significantly positively related to the state test in reading. For every one-point increase in SP score, there was a 4.16-point increase in the state test score, and for every one-point increase in RC score, there was a 2.93-point increase in the state test score. Therefore, student performance on the SP subtest resulted in faster growth on the CSAP compared to RC.

DISCUSSION

The goal in this study was to examine if student CSAP growth is related to student DORA growth. The hypothesis was that student CSAP growth will be significantly and positively related to student DORA growth. Thus, the relationship of interest in addressing this research question is with the time-varying covariate and the state reading test. The hypothesis was partially supported in the Full Models controlling for the demographics in that the DORA scores for Spelling and RC were positively and significantly related to state reading test scores, indicating that these subtests are demonstrating a correlated growth in students reading to the state testing regardless of demographic status.

As noted in the results section, the time-varying growth rates for SP and RC from the Full Models were both significantly and positively related to the state test in reading; however, student performance on the SP subtest had faster growth on the state test compared to RC. This is an interesting finding as typically the RC subtest is viewed as the most similar in structure and content to state reading tests (Let's Go Learn, Inc. ©, 2013). The RC subtest attempts to access the semantic domain of a learner's reading abilities. Children silently read passages of increasing difficulty and answer questions about each passage immediately after they read it. The questions for each passage are broken up into three factual questions, two inferential questions, and one contextual vocabulary question. This is typically how many state reading tests structure their exams.

As indicated above, the SP subtest surprisingly was related to the fastest state test score growth rates in students. SP is a generative process as opposed to a decoding or meaning-making process as seen in most assessments of reading comprehension, which does not support the finding as noted above. Additionally, it is natural for young readers' spelling abilities to lag a few months behind their reading comprehension abilities (Bear, Invernizzi, Templeton, & Johnston, 2000). Overall, the significant findings from the Full Models indicate that as modes of online FA, SP and RC online FA subtests are related to faster state test score growth rates, with SP producing the highest growth rate.

With regards to the non-significant findings for WR and OV, it is not alarming to see that the WR subtest growth was not correlated with CSAP growth, as the testing of word identification skills out of context is typically not a skill that is the focus of standardized reading assessments (Let's Go Learn, Inc. ©, 2013). As for OV, a significant correlation between this subtest and state test score growth was expected, as a learner's knowledge of words and what they mean is an important part of the reading process (Butler, Marsh, Sheppard, & Sheppard, 1985). The knowledge of word meanings affects the extent to which the learner

comprehends what is read, such as in more traditional standardized reading tests. This subtest asks students to select the picture that correctly corresponds to a word they hear. This audio-visual format may explain the non-significant relationship between the more standardized paper-and-pencil format for most reading state tests.

SOLUTIONS AND RECOMMENDATIONS

Although causal inference is limited, the demonstrated relationship can provide teachers and administrators with evidence to warrant the continued use of technology-based FA practices. Specific to the results, the SP and RC subtests are further supported as a learning tool to gauge, or perhaps predict, student performance on the reading state test. The fact that DORA use was significantly and positively related to the state test even when student demographics were controlled, suggests that DORA SP and RC have something additional to offer.

For teachers/educators, focusing on a student's SP and RC growth can potentially add to growth on the state reading test. For example, if a teacher can raise students' DORA SP subtest score by one point, he or she can expect to see a 4.16-point increase on the reading state test on average. And if a teacher can raise students' DORA RC subtest score by one point, he or she can expect to have a 2.93-point increase on the state test (i.e., every three to six months). Thus, if a state reading test is given once a year (i.e., every 12 months) and this mode of online FA renders a one-point increase every three to six months, students' scores on the state test are predicted to grow between 8.32 and 16.64 points that year.

Teachers (and administrators) will also benefit from the results by garnering support for the use of online FA from a practical perspective. One major benefit is the ease of disseminating feedback to students after an assessment, and using the automated, specialized feedback to

diagnose problems and quickly remedy these issues in time for the state exam. Buchanan (2000) noted that the individualized feedback makes this mode of online FA ideal for large, multi-section, introductory-level college courses. In the case of the current study, this mode can also be deemed ideal for large classrooms of elementary, middle, and high school students, in which their teachers may not have the time or resources to give specialized feedback to everyone.

Another implication is the practical advantage of ease of assessment for the large number of students being tested in the educational system. For FA to be most effective, quality feedback should be provided at frequent intervals, and testing a large number of students frequently with specialized feedback can advocate the use of a technology-based mode of FA. Implications also extend to the cost surrounding mass testing. Since a positive relationship was indicated between online FA scores and state test scores, this may allow administrators to have the necessary support to purchase site licenses and invest money in such programs, which are generally cheaper to administer frequently in bulk.

For administrators, the demand for school systems, individual schools, and teachers to be accountable for student performance has increased considerably over the past two decades. This demand for accountability relates to a direct measurement of attainment of educational standards and objectives. The results from this research question support the use of online FA tools as a way to measure and attain various educational standards such as having students pass and excel on the end-of-year, summative state exam. Overall, these results provide some support for administrative demands to find more efficient ways to meet state standards. The significant and positive relationship between the scores may help get support for schools to obtain the funds needed for programs to alleviate some constraints of mass assessment.

Society, in general, benefits from these findings by demonstrating how online FA practices may have the potential to give educators a more

efficient and consistent way to make accurate, data-driven, informed decisions when evaluating student progress. This technology-enhanced efficiency has the potential to support independent student learning and facilitate future lifelong learners in our society. Rarely do teacher practices and methods have such depth of empirical support and is considered a universal best practice. Because of this widespread acceptance, evidencing how technology-based methods of FA influence achievement becomes paramount in demonstrating student learning gains and success in school and beyond.

FUTURE RESEARCH DIRECTIONS AND LIMITATIONS

Although the usefulness of correlational studies and related research questions have a place in the research process, causal conclusions cannot be stated. Future research should consider implementing a similar design, but also obtain an adequate control group. With regards to general threats to internal validity, some limitations are apparent (Cook & Campbell, 1979). Maturation is a limitation, since the existing data sampled took place over months in academic years (Kazdin, 2003). Another methodological limitation includes the use of one school district. Future studies should include multiple school districts from a range of rural and urban areas and involve public and private schools as well.

With regards to the DORA and CSAP data, the DORA data are from 2006/2007 to current and the CSAP data range from 2004/2005 to present as well. Moreover, there are more DORA time points than CSAP time points. Thus, data were not collected at the same time points (i.e., it was approximated). Generally, HLGMs can accommodate time-unstructured data such as the above; however, the accuracy and validity of results may depend on how closely the data are measured (i.e., same time/day compared to several days/weeks apart) in a time-varying covariate model (Biesanz,

Deeb-Sossa, Papadakis, Bollen, & Curran, 2004). Although this could be considered problematic, analyzing only the data points that “matched” compared to all the data did not change the substantive results.

Analyzing only students in grades 3 through 11 may be considered another weakness. Students in grades Preschool through 2 and grade 12 were not included because this study focused on the state test and regularly administered FAs, which only occur between grades 3 through 10. State testing in Colorado begins in grade 3, and grades 11 and 12 are given college preparatory exams and high school exit exams. Additionally, DORA is administered more frequently in younger grade levels, and at least three time points are necessary to analyze the data. Future studies should consider analyzing all grade levels with complete data from multiple districts.

CONCLUSION

The body of FA literature has unanimously heralded the benefits of the diagnostic use of assessment to inform curriculum and instruction, and consequently, improve student performance and achievement. Previous research in this area has primarily focused on traditional FA practices. More recently with the technology movement in schools, the literature is beginning to examine the effectiveness of Internet-based FA, with the latest studies of this modern mode FA beginning to replicate these findings. The current study attempted to add to this literature base by examining one online FA program and its relationship to a summative state proficiency test.

It was hypothesized that student online FA growth would be related to state test score growth. This hypothesis was partially supported in that these online FA scores demonstrated a correlated growth with the state test scores regardless of demographic status. The findings can provide some support to the burgeoning literature outlining the role of online FA in teaching and learning. Internet-mediated teaching and assessment

is becoming commonplace in the classroom, and is more frequently being used to replace traditional modes of student assessment. The need to examine the extent to which these methods are educationally sound is in high demand. Results from this study can not only add to the literature base theoretically and methodologically, but also practically, by bolstering support for federal initiatives and administrative demands for more efficient, technology-based ways to encourage teachers to invest their time in this mode of FA, and in turn, meet state standards and increase student achievement.

REFERENCES

- Bear, D. R., Invernizzi, M., Templeton, S., & Johnston, F. (2000). *Words their way: Word study for phonics, vocabulary, and spelling instruction* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.
- Biesanz, J. C., Deeb-Sossa, N., Papadakis, A. A., Bollen, K. A., & Curran, P. J. (2004). The role of coding time in estimating and interpreting growth curve models. *Psychological Methods*, 9(1), 30–52. doi:10.1037/1082-989X.9.1.30 PMID:15053718
- Black, P. (2015). Formative assessment – An optimistic but incomplete vision. *Assessment in Education: Principles, Policy & Practice*, 22(1), 161–177. doi:10.1080/0969594X.2014.999643
- Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–75. doi:10.1080/0969595980050102
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5–31. doi:10.1007/s11092-008-9068-5
- Boston, C. (2002). The concept of formative assessment. *Practical Assessment, Research & Evaluation*, 8(9). Retrieved from <http://PARE-online.net/getvn.asp?v=8&n=9>



- Buchanan, T. (2000). The efficacy of a world-wide web mediated formative assessment. *Journal of Computer Assisted Learning*, 16(3), 193–200. doi:10.1046/j.1365-2729.2000.00132.x
- Butler, S., Marsh, H. W., Sheppard, M. J., & Sheppard, J. L. (1985). Seven-year longitudinal study of the early prediction of reading achievement. *Journal of Educational Psychology*, 77(3), 349–361. doi:10.1037/0022-0663.77.3.349
- Buzzetto-More, N., & Guy, R. (2006). Incorporating the hybrid learning model into minority education at a historically black university. *Journal of Information Technology Education*, 5, 153–164.
- Carlson, D., Borman, G. D., & Robinson, M. (2011). A multistate district-level cluster randomized trial of the impact of data-driven reform on reading and mathematics achievement. *Educational Evaluation and Policy Analysis*, 33(3), 378–398. doi:10.3102/0162373711412765
- Chua, Y. P., & Don, Z. M. (2013). Effects of computer-based educational achievement test on test performance and test takers motivation. *Computers in Human Behavior*, 29(5), 1889–1895. doi:10.1016/j.chb.2013.03.008
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago, IL: Rand McNally.
- Gulikers, J. T. M., Biemans, H. J. A., Wesselink, R., & van der Wel, M. (2013). Aligning formative and summative assessments: A collaborative action research challenging teacher conceptions. *Studies in Educational Evaluation*, 39(2), 116–124. doi:10.1016/j.stueduc.2013.03.001
- Hunt, N., Hughes, J., & Rowe, G. (2002). Formative automated computer testing (FACT). *British Journal of Educational Technology*, 33(5), 525–535. doi:10.1111/1467-8535.00289
- Jenkins, M. (2004). Unfulfilled promise: Formative assessment using computer-aided assessment. *Learning and Teaching in Higher Education*, 1, 67–80.
- Kazdin, A. E. (2003). *Research design in clinical psychology*. Boston, MA: Allyn & Bacon.
- Kingston, N., & Nash, B. (2011). Formative assessment: A meta-analysis and a call for research. *Educational Measurement: Issues and Practice*, 30(4), 28–37. doi:10.1111/j.1745-3992.2011.00220.x
- Let's Go Learn, Inc. (2013). *Let's Go Learn Research*. Retrieved November 21, 2013, from <http://www.letsgolearn.com/lgl/site/research>
- Maxwell, S. E., & Delaney, H. D. (1990). *Designing experiments and analyzing data: A model comparison perspective*. Belmont, CA: Wadsworth.
- Merino, K., & Beckman, T. (2010). Using reading curriculum-based measurements as predictors for the measure academic progress map standardized test in Nebraska. *International Journal of Psychology: A Biopsychosocial Approach*, 6, 85 – 98.
- OConnell, A. A., & McCoach, D. B. (2004). Applications of hierarchical linear models for evaluations of health interventions: Demystifying the methods and interpretation of multilevel models. *Evaluation & the Health Professions*, 27(2), 119–151. doi:10.1177/0163278704264049 PMID:15140291
- Peat, M., & Franklin, S. (2002). Supporting student learning: The use of computer-based formative assessment modules. *British Journal of Educational Technology*, 33(5), 515–523. doi:10.1111/1467-8535.00288
- Petrides, L. (2006). Data use and school reform. *T.H.E. Journal*, 33(8), 38–41.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Newbury Park, CA: Sage Publications.
- Ridgway, J., McCusker, S., & Pead, D. (2004). *Literature review of e-assessment*. Bristol, UK: Nesta Future Lab.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York, NY: Oxford University Press. doi:10.1093/acprof:oso/9780195152968.001.0001

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). London, UK: Sage Publications.

ADDITIONAL READING

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2002). *Working inside the black box*. London, England: Nelson Publishing Company.

Black, P., & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139–149.

Brookhart, S. M. (2007). Expanding views about formative classroom assessment: A review of the literature. In J. H. McMillan (Ed.), *Formative classroom assessment: Research, theory and practice*. New York, NY: Teachers College Press.

Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, 58(4), 438–481. doi:10.3102/00346543058004438

Fuchs, L. S., & Fuchs, D. (1986). Effects of systematic formative evaluation: A meta-analysis. *Exceptional Children*, 53, 199–208. PMID:3792417

Hox, J. J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Erlbaum.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2), 254–284. doi:10.1037/0033-2909.119.2.254

Shepard, L. A. (2009). Commentary: Evaluating the validity of formative and interim assessment. *Educational Measurement: Issues and Practice*, 28(3), 32–37. doi:10.1111/j.1745-3992.2009.00152.x

KEY TERMS AND DEFINITIONS

Covariation: Variation or variance that is correlated between two or more variables.

E-Learning: Learning that uses electronic technology or media (e.g., the Internet) to access education outside of the traditional brick and mortar classroom.

Formative Assessment: Formal and informal assessment methods conducted by educators concurrent with student learning used to adapt teaching and learning activities to improve student achievement.

Multilevel Modeling: Statistical models (e.g., generalizations of linear models such as linear regression) of parameters that vary at more than one level (e.g., nested data).

Proficiency Test: An exam that evidences how competent or skilled a student or learner is in a particular activity or field of study.

Summative Assessment: Assessment methods that are used to evaluate student learning and/or achievement at the end of an instructional cycle.

Time-Varying Covariate: This statistical term (also called a time-dependent covariate) is used in multilevel growth modeling or survival analysis, and indicates that a covariate in the growth model is not constant throughout.

Category L

**Library Science and
Administration**

Change Leadership Styles and Behaviors in Academic Libraries

John Kennedy Lewis

Salve Regina University, USA

INTRODUCTION

Academic libraries in the 21st century are in the midst of constant change caused by advances in technology, shifting demands of faculty and students, declining budgets, transformations in publishing and in other methods of acquiring scholarly research. Librarians working in this environment must cope with continuous change while still providing research support to their primary clientele - faculty and students. Determining the best methods of change leadership to cope with these forces is vital. As technology has increasingly impacted the workplace, best practices for leading employees through change initiatives have become increasingly common. As a new millennium is entered and the pace of change accelerates the importance of determining the best leadership styles for assisting followers in these transitions is even more imperative.

Traditionally change leadership research has fallen into two camps – those that emphasize the importance of the leader's role in facilitating change, and those that look at change leadership as a situational contingency that is moderated by traditional leadership behaviors and styles (Herold, Fedor, Caldwell, & Yi, 2008). The former approach focuses on what leaders should do when confronted with a change initiative and how to encourage followers to embrace change. This approach is typified by the work of Kotter. This article concentrates on the latter approach which believes that certain leadership styles are more useful in situations of change and is typified by the work of Bass and Riggio.

BACKGROUND

Academic libraries have been involved in a process of continuous change over the past thirty years due to the impact of technology on all aspects of librarianship. From the automated card catalog, web based database searching, remote off-campus access, streaming video, downloadable e-books, and patron driven acquisition, advances in technology have continually affected and changed the way librarians perform their jobs. These technological advances have created numerous pressures on librarians ranging from shifting job responsibilities to frequent library reorganizations.

The Disruptive Innovation Era

Libraries remained hierarchical organizations well into the 1960's. The major shift in academic library organization was the direct result of technological advances. The introduction of the computer to the library workplace was the type of "disruptive innovation" discussed by Christensen (2003) in his seminal work "The Innovator's Solution." Academic libraries adopted two new technologies which changed the workplace for librarians: first the online public access catalog (OPAC), followed by online databases. The OPAC changed the way patrons located print materials in the library. It also changed the way librarians performed their duties. By the late 1990's the majority of academic libraries had switched over to online access of databases (Kenan, 2012). An academic librarian without technological skill cannot function in the 21st century academic library.

Traditional Leadership Methods

Reinforcement/Transactional Leadership

This style is typified by the type of manager who rewards good employees and corrects poor performance with negative feedback or withholding wanted gain. This can involve various types of positive and negative reinforcement. Bass and Bass (2008) defined contingent reward as a constructive transaction. "The leader assigns a task or obtains agreement from the follower on what needs to be done and arranges for psychological or material rewards of followers in exchange for satisfactorily carrying out the assignment" (p. 623). Reinforcement leadership makes an employee's job simple – the employee is told what to do and instructed how to do it. Some employees enjoy this type of work environment – they know exactly where they stand and little independent thought or initiative is required. However, for motivated and innovative employees, such an environment can become tedious and stifling. Suwannarat (1994) found that contingent reward was the most commonly used transactional leadership behavior by library directors while active/passive management by exception was used slightly less often (p. 77).

Consultative Leadership

Consultative leadership is typified by the type of manager who asks for and encourages employee feedback but reserves the final decision making power. Consultative leadership is still quite common in many academic research libraries. Most library directors will meet with librarians when a change initiative is contemplated to receive feedback from those who will be affected. However, this type of leader tends to only accept feedback which agrees with their point of view. Negative feedback is often either ignored or punished. Library leaders who practice this style promote apathy in their followers. Librarians soon realize their opinions and expertise are not valued

and tune out from such discussions. They accept change passively or resist quietly. The deluded leaders who employ such a leadership style take the silence for agreement, when it is in fact apathy, disgruntlement and quiet resistance to change.

Contingency Theories of Leadership

There are various theories of contingency leadership which propose that the best leadership style varies with the actual situation encountered. Situational leadership dictates that different problems often require different leadership styles. Another contingency theory of leadership is the path-goal theory of leadership. Path-goal theory leadership cuts across all the leadership styles as at various times a leader may be exhibiting transactional, transformative, consultative, directive or even autocratic behaviors. It is incumbent upon the leader to select the correct leadership behavior to fit each circumstance.

Podell (2012) applied path-goal theory to academic library leadership. Ernest Shackleton was used to exemplify how path-goal theory can be utilized by library leaders. Despite having his ship crushed by ice, Shackleton was able to lead his crew to safety without the loss of a single man. Wolcott (1984) investigated path-goal theory in academic libraries and found the theory to be inapplicable. He attributed the negative results to librarians having a low need-for-independence. However, it seems more likely that criticisms of path-goal theory as being overly reliant on expectancy theory which does not account for emotional reactions to decision dilemmas or human motivations such as self-concepts were responsible for the lack of confirmation (Yukl, 2013).

Situational leadership has also been a leadership style addressed in the library literature. Dragon (1979) proposed that a new theory of leadership was needed and suggested that situational leadership could be the foundation of this new theory. Library supervisors were found to be higher on the initiating structure dimension than other groups including corporation presidents



and U.S. Senators. The library supervisors were lower on the consideration dimension than all other groups except aircraft engineers and highway patrol officers. “Subordinates in this study perceived their supervisors as behaving in ways more likely to initiate structure and bring order to their work groups than in ways which would increase the sense of group well-being” (p. 62).

Bull (2000) used the Managerial Grid to investigate leadership styles of library directors in small, Christian academic libraries. She found that there were two primary leadership styles employed in this type of academic library, and they were on exact opposite corners of the grid: impoverished and country club. Most of the directors were either task masters unconcerned with the welfare of their followers or humanitarians who put their employees first.

Mitchell (1987) found that the Contingency Model was not supported in an academic library environment (p. 94). Mitchell concluded as had other researchers that the LPC was not an effective measure of leadership style. Mitchell suggested that the reason the model was not supported was because librarians were a co-acting group rather than an interacting group (p. 101). This logic seems faulty for while academic librarians perform solo work they also perform group tasks on a daily basis. It seems more likely that the Mitchell’s findings did not support Fiedler’s Contingency Model of Leader Effectiveness because of its previously mentioned flaws.

Participatory Leadership

There are many leadership styles and behaviors that can be classified as participatory/ democratic leadership. Goleman (2002) defined a democratic leader as one who utilizes an employee’s knowledge, and creates a group commitment to the resulting goals. Despite general agreement that participatory leadership is desirable and leads to more job satisfaction and higher morale actual research findings have been spotty. Yukl (2013) believes the lack of strong, consistent research

findings is primarily the result of poor methodology or measurement limitations.

Participatory leadership has been touted as a panacea in academic libraries for many years. The team model involved more librarians in the decision making process but did not change the end result – the library director still had final say on any major decision. Many academic library directors saw team formation as a way to dismantle the unwieldy committee structure that was prevalent in large academic libraries (Nussbaumer & Merkley, 2010). The imposition of team structures lead to less participatory leadership, not more. The problems with participatory leadership were well covered by Lichtenstein (2000) “participatory management requires skill; most people need to be trained in the process; improper use of participatory management may cause serious resentment among staff; and may serve as a mask for poor leadership.” (p. 34).

Casey (2011) found that participatory management practices were either in place or being implemented in her study of strategic change in three academic libraries. She found considerable evidence of participatory management at all three schools either through a team based decision making structure or through committee activities. Strecker (2010) found that academic librarians prefer a team based structure over a hierarchical organization. Librarians felt that working in teams best assisted them in adopting and learning new technologies. While there was a preference for teamwork the vast majority of the libraries still employed a hierarchical structure.

LEADERSHIP STYLES THAT SUPPORT CHANGE

Transformational Leadership

James MacGregor Burns first proposed the theory of transformational leadership in relation to political leaders such as President John F. Kennedy in his seminal work, *Leadership* (Burns, 1978). Bass

and Riggio (2006) believed that transformational leaders “help followers grow and develop into leaders by responding to individual followers’ needs by empowering them and by aligning the objectives and goals of the individual followers, the leader, the group, and the larger organization” (p. 3). Bass proposed that there were four transformational behaviors that leaders could practice: inspirational motivation, idealized influence, intellectual stimulation, and individualized consideration (Bass & Bass, 2008).

According to Alire (2007) transformational leadership should be the most effective leadership style for academic libraries due to the need for constant change and adaptation. “Libraries are experiencing rapid change as a result of the technological transformation of society, higher education, and service delivery” (p. 171). Information seeking behavior of library patrons has also drastically changed due to competition from commercial providers (Google) and the technological competencies of Generation X and Millennials (Alire, 2007).

Hernon, Powell and Young in a two part study looked at the qualities needed to become an ARL library director. While never addressing transformational leadership directly, the list of attributes created included many transformational leadership qualities. Top valued traits included: shapes change; builds a shared vision, builds relationships, thinks outside the box, is innovative; and articulates direction - all qualities of a transformational leader.

Suwannarat (1994) measured the transformational leadership behaviors exhibited by academic library directors in ARL libraries. He found that directors primarily exhibited transformational leadership behaviors according to self-reports and their subordinate’s reports. Interestingly, “library directors were found to rate themselves significantly higher than did subordinates on idealized influence, inspirational leadership, intelligent stimulation, individualized consideration, contingent reward, extra effort and effectiveness”

(p. 102). Essentially the leaders were much more impressed with their job performance and leadership behaviors than their followers.

Baird (2010) conducted a Delphi study to look at how library deans lead their organizations during periods of frequent change in comparison with academic deans of the same institutions. There was little difference between the two types of leaders; both took actions “designed to effectively protect and lead their units” (p. 152). She pointed out a problem with current library science curricular standards, “an education that purports to prepare professionals might need to build depth and specializations, perhaps through certificate programs and post-graduate internships, to give a solid grounding in leadership theories supporting the core knowledge of library science” (p. 161).

Albritton (1993) directly compared transformational and transactional leadership styles in university libraries and their perceived effect on organizational effectiveness. Her findings were noteworthy confirming that transformational leadership behaviors were both preferred and perceived as more effective when compared to transactional leadership behaviors. One hypothesis that was supported in the research was that transformational leadership factors such as “charisma, inspiration, intellectual inspiration, and individualized consideration” when compared to transactional leadership factors such as contingent reward and management by exception would lead to “three outcome measures: extra effort of subordinates; satisfaction with leader; and leadership effectiveness” (p. 145). She also found that transformational leadership was “associated with perceptions of higher levels of planning, organizing, leading, evaluating, managing, and resource development” by respondents (p. 255). These findings correspond with Wang (2011) who found that “when a leader displays higher levels of transformational leadership, he or she will be perceived as exerting more proactive feedback seeking behaviors by subordinates” (p. 67).

Shared Leadership

Historically leadership research has centered around one individual – the leader. However, in actuality leadership is seldom about one person all great leaders have great lieutenants. Pearce and Conger (2003) point to the shifts in how work is performed, the importance of cross-functional teams to explain the rise of shared leadership. Other factors include the need for speedier leadership decisions, the fact that each employee possesses unique knowledge, expertise and skills, and that a single CEO cannot possess “all the leadership skills and knowledge necessary to guide complex organizations in a dynamic and global marketplace” (p. 2). Shared leadership should be welcomed by academic librarians. As Pearce (2004) points out “shared empowering leadership emphasizes building self-influence skills that orchestrate performance while preserving autonomy, as such, it may be particularly suited to knowledge workers, who often desire autonomy on the job” (p. 54).

Shared leadership can also mean co-leadership or tri-leadership instead of one leader. O’Toole, Galbraith and Lawler (2002) defined shared leadership in this manner. This definition of shared leadership is rather limited in that it connotes not sharing leadership with followers but merely dividing up responsibilities among leaders. The main advantage of this type of shared leadership is that it leads to smooth succession planning.

Newman (1998) called for academic librarians from within the ranks to seize the leadership opportunities available to them. His concept of shared leadership was quite limited as he pointed to collection management, library instruction, and technology as areas where academic librarians could make decisions. The idea of shared leadership in academic libraries hasn’t advanced much beyond Newman’s limited view in the past fifteen years. In a recent book Hernon (2013) has a very limited concept of shared leadership in academic libraries. Library directors should form teams, “ensure that the tasks assigned to it are clear, delineate the

team’s authority, provide stable membership over time, and monitor discreetly that team members work well together” (p. 152). This is not shared leadership it is merely replacing department heads with teams and reserving ultimate decision making power for the library director.

Franklin (1999) called for shared leadership in academic libraries. “Generally speaking, today’s college and research staff are highly educated, self-motivated, and possess specialized skills, these are all characteristics that support a shared leadership model” (p. 22). Cawthorne (2010) conducted the only study of shared leadership in academic libraries to date. He found the highest agreement with the idea that middle managers are accountable for decisions within the scope of their responsibilities. This is hardly surprising as it merely measures hierarchical decision-making authority. He found little evidence of true shared responsibility. Middle managers agreed they shared information with top managers, but many felt top managers did not reciprocate. The majority of respondents also did not agree with the statement that “ideas presented at all levels of staff in the library are equally considered” (p. 154).

A few libraries, however, have achieved some type of shared leadership. One model that has worked at Dickinson College is collegial management and rotating leadership. McKinzie (2000) states, “In 1975 the librarians at Dickinson did something that few small academic libraries have ever done, they scrapped the traditional hierarchical model of leadership, introduced a collegial pattern of management with a rotating chair, and implemented a holistic vision of librarianship” (p. 2). The librarians modeled their structure on a typical campus academic department that is run by a rotating chair. Dickinson College was the first library to adopt such a leadership model, since then several other small college libraries have followed suit. He points out there are numerous advantages to such a model - there is no possibility of being stuck long term with a bad leader since each chair serves only a three year term, all librarians

have an opportunity for leadership, every voice is given equal weight, and the image of librarians is enhanced on the campus.

Carriveau, Viers and Ziebell (2008) also touted the benefits of collaborative leadership. Ripon College decided to save money by not replacing a library director who left but instead created a tripartite governing body of department heads. Such collaborative leadership is rare in the mainly hierarchical organization of academic libraries. The benefits are many including avoiding the “departmentalization and territoriality that exists in many libraries” (p. 307).

Self-Leadership

The ultimate end of the continuum of leadership behaviors is self-directed leadership. This means that each employee is a leader, and that all workers are completely autonomous and responsible for all decision-making in their milieu. Charles Manz (1986) is the only management theorist who has advocated for what he called self-leadership. “Building on Bandura’s social learning theory, Manz argued for fostering self-planning, self-direction, self-monitoring, and self-control, which could replace otherwise needed supervision” (Bass and Bass, 2008, p. 871). However, even Manz believed that self-leadership required a superleader who would oversee all autonomous individuals. Wassenaar and Pearce (2012) believe self-leadership could be a “possible substitute for a more traditionally appointed, vertical leader” (p. 367).

Deiss (1999) in a best practices article claimed that through improving self-awareness, seeking feedback to improve performance, and creating a clear personal vision, librarians could benefit from self-leadership. However, her article did not address self-leadership as a replacement for supervision by others. Topper (2009) in a short article described self-leadership as a method of self-improvement which could lead one to be a better person and leader.

FURTHER RESEARCH DIRECTIONS



There are many areas rife for future research in the area of change leadership. One is the effect of gender on change leadership. The research seems to suggest that gender does have an impact on leadership. Men and women have different innate strengths and weaknesses in change leadership but this has yet to be empirically proven. Age is another factor that seems to affect change leadership. Some leadership behaviors seem to increase or improve with age while others decline. The biggest controversy in the research on change leadership is the nature/nurture debate. Are leadership behaviors innate or can they be developed through training? Further research needs to focus on identifying the most effective change leadership behaviors for the future. In today’s rapidly changing, globalized world behaviors that are important now may be unimportant in the near future, and little used behaviors may become vital. The impact of situation and context on change leadership styles also needs further research. Specifically on prioritizing which change leadership behaviors seem to be the most potent in the most situations.

CONCLUSION

Change leadership in academic libraries has been an understudied topic in library literature. Research methods such as ethnography, in-depth interviews, and observation allow a researcher to get a closer look at change leadership styles. Standard tools such as questionnaires mainly result in self-perceived efficacy which while interesting is not as valuable. Another difficulty not just with change leadership but with the whole area of organizational change in academic libraries is lack of measurement and assessment. Many articles are opinion based or best practices but not much literature addresses the effects of different leadership styles; nor is there much proof that organizational changes have improved service, job satisfaction, or staff morale.

REFERENCES

- Albritton, R. A. (1993). *Transformational vs. transactional leadership in university libraries: A test of the model and its relationship to perceived library organizational effectiveness* (Doctoral Dissertation). Retrieved from Proquest Dissertations and Theses. (9328961)
- Alire, C. A. (2007). Diversity and Leadership. In P. Hernon & N. Rossiter (Eds.), *Making a Difference: Leadership and academic libraries* (pp. 169–179). Westport, CT: Libraries Unlimited.
- Baird, L. (2010). *Colliding scopes: Seeing academic library leadership through a lens of complexity* (Doctoral Dissertation). Retrieved from Proquest Dissertations and Theses. (3455099)
- Bass, B. M., & Bass, R. (2008). *The Bass handbook of leadership: Theory, research and managerial applications* (4th ed.). New York, NY: Free Press.
- Bass, B. M., & Riggio, R. E. (2006). *Transformational leadership* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Bolden, R. (2011). Distributed leadership in organizations: A review of theory and research. *International Journal of Management Reviews*, 13(3), 251–269. doi:10.1111/j.1468-2370.2011.00306.x
- Bull, S. I. (2000). *Managerial styles of library directors in institutions of the Council for Christian Colleges & Universities* (Doctoral Dissertation). Retrieved from Proquest Dissertations and Theses. (3023328)
- Burns, J. M. (1978). *Leadership*. New York, NY: Harper & Row.
- Carriveau, A., Viers, V., & Ziebell, C. (2008). Collaborative leadership at Ripon College. *College & Undergraduate Libraries*, 15(3), 301–313. doi:10.1080/10691310802258224
- Casey, A. M. (2011). *Strategic priorities and change in academic libraries* (Doctoral Dissertation). Retrieved from Proquest Dissertations and Theses. (3467461)
- Cawthorne, J. E. (2010). Leading from the middle of the organization: An examination of shared leadership in academic libraries. *Journal of Academic Librarianship*, 36(2), 151–157. doi:10.1016/j.acalib.2010.01.006
- Deiss, K. J. (1999). Personal best. *College & Research Libraries News*, 60(3), 193. Retrieved from <http://crln.acrl.org/>
- Dickinson, D. (1978). Some reflections on participative management in libraries. *College & Research Libraries*, 39(4), 253–262. doi:10.5860/crl_39_04_253
- Dragon, A. (1979). Leader behavior in changing libraries. *Library Research*, 1(1), 53–66. Retrieved from <http://www.sciencedirect.com/science/journal/07408188>
- Franklin, B. R. (1999). Beyond shared leadership: The importance of learning in a shared leadership model. *College & Research Libraries News*, 60(1), 22–23. Retrieved from <http://crln.acrl.org/>
- Goleman, D., Boyatzis, R., & McKee, A. (2002). *Primal leadership: Realizing the power of emotional intelligence*. Boston, MA: Harvard Business School Press.
- Hernon, P., & Pors, N. O. (Eds.). (2013). *Library leadership in the United States and Europe: A comparative study of academic and public libraries*. Santa Barbara, CA: Libraries Unlimited.
- Hernon, P., Powell, R. R., & Young, A. P. (2001). University library directors in the Association of Research Libraries: The next generation, part one. *College & Research Libraries*, 62(2), 116–145. Retrieved from <http://crl.acrl.org/> doi:10.5860/crl.62.2.116
- Hernon, P., Powell, R. R., & Young, A. P. (2002). University library directors in the Association of Research Libraries: The next generation, part two. *College & Research Libraries*, 63(1), 73–90. doi:10.5860/crl.63.1.73

Herold, D. M., Fedor, D. B., Caldwell, S., & Yi, L. (2008). The effects of transformational and change leadership on employees commitment to a change: A multilevel study. *The Journal of Applied Psychology*, 93(2), 346–357. doi:10.1037/0021-9010.93.2.346 PMID:18361637

Kenan, S. K. (2012). *Perceptions of personnel at selected Texas community colleges regarding the impact of technology on their libraries* (Doctoral Dissertation). Retrieved from Proquest Dissertations and Theses. (3504939)

Lichtenstein, A. A. (2000). Participatory Management: A Critical Look. *Journal of Library Administration*, 31(1), 29–41. doi:10.1300/J111v31n01_02

Manz, C. C. (1986). Self-Leadership: Toward an Expanded Theory of Self-Influence Processes in Organizations. *Academy of Management Review*, 11(3), 585–600. doi:10.5465/AMR.1986.4306232

McKinzie, S. (2000). Twenty-five years of collegial management: The Dickinson College Model of Revolving Leadership and Holistic Librarianship. *Library. Philosophy and Practice*, 2(2), 1-8. Retrieved from <http://www.webpages.uidaho.edu/~mbolin/>

Mitchell, E. S. (1987). *Leadership style in academic libraries: A test of Fiedler's contingency model of leadership effectiveness* (Doctoral Dissertation). Retrieved from Dissertations & Theses: Full Text. (AAT 8723277).

Newman, G. C. (1998). Leading from within: Leadership within the ranks of academic librarians. In T. F. Mech & G. B. McCabe (Eds.), *Leadership and academic librarians* (pp. 108–119). Westport, CT: Greenwood Press.

Nussbaumer, A., & Merkley, W. (2010). The path of transformational change. *Library Management*, 31(8), 678–689. doi:10.1108/01435121011093441

OToole, J., Galbraith, J., & Lawler, E. E. III. (2002). When two (or more) heads are better than one: The promise and pitfalls of shared leadership. *California Management Review*, 44(4), 65–83. doi:10.2307/41166143

Peak, M. (2008). *Organizational technological change and the research university: The exploration of transformation in an academic research library* (Doctoral Dissertation). Retrieved from Proquest Dissertations and Theses. (3312181).

Pearce, C. L. (2004). The future of leadership: Combining vertical and shared leadership to transform knowledge work. *The Academy of Management Executive*, 18(1), 47–57. doi:10.5465/AME.2004.12690298

Pearce, C. L., & Conger, J. A. (2003). *Shared leadership: Reframing the hows and whys of leadership*. Thousand Oaks, CA: Sage Publications.

Podell, R. (2012). Using Shackleton's concepts to improve library management: Applying the lessons of history's greatest leader. *Journal of the Library Administration & Management Section*, 8(2), 4–21. Retrieved from <http://www.nyla.org/>

Strecker, B. L. (2010). *Academic librarians' perceptions of teamwork and organizational structure in a time of rapid technological change* (Doctoral Dissertation). Retrieved from Proquest Dissertations and Theses. (3434167).

Suwannarat, P. (1994). *Library leadership in research university libraries* (Doctoral Dissertation). Retrieved from Proquest Dissertations and Theses. (9429735).

Topper, E. F. (2009). Self-leadership: Road to personal excellence. *New Library World*, 110(11/12), 561–563. doi:10.1108/03074800911007578



Wang, L. (2011). *An alternative explanation for transformational leadership effectiveness: The role of leader proactive feedback-seeking behavior in the relationship between transformational leadership and subordinate performance* (Doctoral Dissertation). Retrieved from Dissertations & Theses: Full Text. (AAT 3460812)

Wassenaar, C. L., & Pearce, C. L. (2012). The nature of shared leadership. In D. V. Day & J. Antonakis (Eds.), *The nature of leadership* (2nd ed.; pp. 363–389). Thousand Oaks, CA: Sage Publications.

Wolcott, C. (1984). *The relationship between the leadership behavior of library supervisors and the performance of their professional subordinates (Path-goal theory, management)* (Doctoral Dissertation). Retrieved from Dissertations & Theses: Full Text. (AAT 8417654)

Yukl, G. (2013). *Leadership in organizations* (8th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

ADDITIONAL READING

Bolden, R. (2011). Distributed leadership in organizations: A review of theory and research. *International Journal of Management Reviews*, 13(3), 251–269. doi:10.1111/j.1468-2370.2011.00306.x

Dent, E. B., & Goldberg, S. (1999). Challenging resistance to change.. *The Journal of Applied Behavioral Science*, 35(1), 25–41. doi:10.1177/0021886399351003

Gang, W., In-Sue, O., Courtright, S. H., & Colbert, A. E. (2011). Transformational leadership and performance across criteria and levels: A meta-analytic review of 25 years of research. *Group & Organization Management*, 36(2), 223–270. doi:10.1177/1059601111401017

Hicks, D. (2007). *Perpetual beta: A study of the attitudes and opinions of Nova Scotian academic librarians towards change*. (Doctoral Dissertation). Retrieved from Proquest Dissertations and Theses. (MR31590)

Hobman, E. V., Jackson, C. J., Jimmieson, N. L., & Martin, R. (2011). The effects of transformational leadership behaviours on follower outcomes: An identity-based analysis. *European Journal of Work and Organizational Psychology*, 20(4), 553–580. doi:10.1080/1359432X.2010.490046

Hoffman, B. J., Bynum, B. H., Piccolo, R. F., & Sutton, A. W. (2011). Person-organization value congruence: How transformational leaders influence work group effectiveness. *Academy of Management Journal*, 54(4), 779–796. doi:10.5465/AMJ.2011.64870139

Ismail, A., Zainuddin, N., & Ibrahim, Z. (2010). Linking participative and consultative leadership styles to organizational commitment as an antecedent of job satisfaction. *Unitar E-Journal*, 6(1): 11-26. Retrieved from <http://www.unirazak.edu.my/research/ejournal/ejournal.htm?latest.htm>

Kelly, T. (2010). A positive approach to change: The role of appreciative inquiry in library and information organisations. *Australian Academic & Research Libraries*, 41(3), 163-177. Retrieved from <http://www.alia.org.au/publishing/aarl/>

Oreg, S., & Berson, Y. (2011). Leadership and employees reaction to change: The role of leaders personal attributes and transformational leadership style. *Personnel Psychology*, 64(3), 627–659. doi:10.1111/j.1744-6570.2011.01221.x

Reed, D. L. (2011). *Leading the comprehensive community college library: Defining, aligning, and supporting innovation and change*. (Doctoral Dissertation). Retrieved from Proquest Dissertations and Theses. (3493207)

KEY TERMS AND DEFINITIONS

Academic Librarians: Academic librarians are those library employees who have obtained a Masters in Library Science or a Masters in Information Science and quite often a second Master's degree in a subject discipline. Occasionally some academic librarians also possess a PhD in a subject discipline as well.

Academic Library: An academic library is a library associated with a college or university which supports the mission of the institution and the research needs of its faculty, staff and students. Although it is possible that a purely virtual library could exist, no brick and mortar institution has yet taken that step. Instead, most academic libraries consist of a physical space (usually quite large but on occasion as small as one room) and a staff of librarians and other employees that manage the building and its collections.

Academic Library Director: An academic library director is the person designated by an institution of higher education to manage a library building and its staff. Depending on the size of the institution and library this person may be designated a director, dean, provost or vice-president.

Consultative Leadership: Also known as democratic leadership, in this style leaders talk to their followers and get their input before making a decision.

Contingency Theories of Leadership: There are various contingency theories of leadership but they include Path-Goal theory, Hersey and

Blanchard's Situational Leadership theory, Blake and Mouton's Managerial Grid, and Fiedler's Contingency Model of Leader Effectiveness.

Participatory Leadership: Participatory leaders allow their followers some input into the decision-making process and often delegate some decision making to their followers.

Shared Leadership: Also called distributed leadership, it may be practiced in situations where the management of an organization is too complicated for a single individual.

Situational Leadership: Situational leadership, originally called the Life Cycle Theory of Leadership, by its progenitors Paul Hersey and Ken Blanchard, proposes that there is no best method of leadership. Instead leadership depends on the situation.

Transactional Leadership: Transactional leadership occurs when the leader rewards or disciplines the follower depending on their level of performance. Transactional leadership relies on contingent rewards or on active/passive forms of management by exception.

Transformational Leadership: Transformational leadership occurs when leaders and followers interact in a manner that raises both to higher levels of motivation and morality. Typically transformational leadership consists of four behaviors displayed by leaders: idealized influence (charisma), intellectual stimulation, individualized consideration and inspirational motivation.

L

Changing Expectations of Academic Libraries

Jennifer Ashley Wright Joe
Western Kentucky University, USA

INTRODUCTION

The digital age has fundamentally changed how academic libraries operate. With the advent of electronic resources the job descriptions and duties of librarians have expanded to include many of the same roles that IT professionals traditionally play. These roles include dealing with computers and software on a scale not seen before, as libraries attempt to stay current and relevant by adding computers to their building. How each library has risen to meet these new challenges may be different, but there are a few things that remain consistent among them. Academic libraries in the modern age are expecting a different type of education and experience from their librarians. The administration expects them to deal with virtual materials and virtual resources, which requires different skills and expertise. These libraries have also adapted their mission statements and functions to reflect the change in their resources and personnel. Some of these changes have come slowly over the last two decades, while others have come fast, but they will not be the only changes that libraries will have to endure. Technology is continuing to change at a pace not seen before in history, and patrons are expecting the library to embrace more and more of it.

BACKGROUND

Most people would say that the digital age was born with the internet. However digital technologies are much older than that, dating back to computers built in the 1960s. Libraries at this time were

using MARC records, which were digitized and are still in use today (Arms, 2012). Though digital databases and e-books were still far in the future, libraries began using this technology early on to improve the quality of the services that they were offering their patrons. Technology was used to streamline libraries on both the back end and the front end, from the records that catalogers used to the way that patrons found materials in the library. Card catalogs became obsolete as computing power increased and those records were able to be digitized. The libraries still had tons of physical materials for people to use in their research, but the digital catalog helped make it easier for people to find those materials.

In addition, telephones, not traditionally considered digital technology, helped the library answer questions without patrons having to be in the library. This was a marked change from previous interaction with patrons, because much of the traditional reference interview relies on the body language of the patron and picking up on clues that they do not know they are giving off. Without the visual cues, librarians had to sharpen their ability to hone in on problems. Just because the patron was calling with their question did not mean that they were certain about the information they needed. This remote-access librarianship paved the way for other methods of contacting librarians, including e-mail, instant messenger, and chat services. Learning from what they had done with regards to telephone interactions, librarians were able to adapt their reference interactions to serve their patron populations through these methods. While this greatly increased the contact librarians had with patrons, it did decrease the

number of people coming into the library, and that was only the start.

Technological advances led to even more new horizons for the library. When the world wide web and its contemporaries were established, they brought with them a new type of resource, the web page, which librarians had to learn how to vet for accuracy and timeliness (Arms, 2012). Though there were many contenders for what would become the internet, the World Wide Web persevered as the accepted software, albeit with many changes over time. These changes included the addition of colors and images, as well as behind-the-scenes information that could be registered within the properties of the web page, but there were still questions about the authenticity of the information being presented on many webpages. Because anyone could build a webpage, this led to misinformation and confusion among library patrons, a problem that still continues in various forms to this day. Academic libraries in their research capacity especially struggled with this problem as students began incorporating webpages into their research. Confusing the matter were many valid websites which held information that could not easily be accessed in other ways. Organizations and government departments began releasing content on the internet. Access became easy, but tracing the source became hard. Furthermore, studies at this time showed that, while students knew that the library had a webpage, and had access to the internet, those surveyed felt that the library and the internet were two separate things (D'Esposito, 1999, p. 458), leading to more concern that students were not getting the best advice when it came to research materials. At the academic library, researchers, in this case students, also had to contend with professors who would not let them use online resources for fear of misinformation. This practice of barring that type of resource led to students not fully understanding how to vet sources on the World Wide Web for their own everyday use.

Then, databases began forming to house materials that were both available in print and digitally. This led to database aggregators such as Ebscohost, JStor, and CSA. Libraries found themselves having to allocate more and more of their budgets to these electronic resources. In some ways, these database aggregators made things easier for the academic library, because digital resources were available in packages that did not initially require much thought from the librarians. However, as libraries continued to sign agreements with these aggregators, they found that there was considerable overlap in the various collections, and that dropping parts of these agreements would increase the price of the packages, essentially giving libraries less information for more money (Zimmerman, 2010). However, libraries could not afford to just drop out of the digital age entirely; students, now familiar with web searching, began expecting resources to be available to them online, without them ever having to step foot in a library. Digital technologies allowed the universities these libraries were associated with to offer classes online for the first time, swelling their enrollment numbers. However, academic libraries saw a decline in their patron numbers even as enrollment continued to grow. This juxtaposition of events required a response from the academic library.

ACADEMIC LIBRARIES RESPOND TO THE DIGITAL AGE

Early in the transition, academic libraries realized that they would need to change to continue to stay relevant. Some of these changes were relatively simple, and included adding the most popular databases as they first began, or offering technology classes for their students who had missed out on those opportunities in high school. However, as public libraries and school libraries increased their technological advancements, academic libraries found themselves having to go ever farther to keep up with the demand of the digital age.



EDUCATION AND EXPERIENCE

When the digital age first began, the Master's in Library Science (MLS) was the accepted degree for librarianship, and had been since 1924 (Kennedy, et al, 2007). Library programs taught cataloging, reference instruction, and how to prepare research guides, among other things. The focus was primarily on interaction with the patron, regardless of the level with which one was dealing. Even as technology progressed and these interactions were completed over the telephone or internet, the emphasis was still on patron interaction.

The introduction of the Master's in Information Science (MIS) and the Master's in Information and Library Science (MLIS) broadened what constituted library education. Classes offered in these programs were more technology driven, and included metadata, database search systems, and social networking, the idea being that if librarians could understand the inner workings of the technology they were dealing with, they could better assist their patron population with that technology. In many programs, these classes were offered alongside the traditionally offered programs, and would often be listed as interchangeable in the course catalog. For example, at Indiana University, in the School of Informatics (formerly the School of Library and Information Science), Representation and Organization of Information, a class that focuses on theories of classification and metadata, is offered as an alternative to the traditional cataloging class. Students enrolled in the MLS program at IU also have the opportunity to take MIS classes as electives, while many others choose to earn a dual degree in both programs.

In the present age, now that libraries are more comfortable with technology, any of the three degrees discussed can help a job seeker secure a position in an academic library. Many librarians still hold the traditional Master's in Library Science, including younger librarians, but the content of those degrees has shifted dramatically since the first library science degree was granted. About 78% of accredited programs have

some technological requirement for all students, regardless of whether these students pursue library science or information science, with 11 programs (approximately 25%) requiring technology skills before admission (Scripps-Hoekstra, Carroll & Fotis, 2014). There is an overwhelming idea that even if a future librarian is not interested in technology, there are some basic skills they will need anyway to perform certain job functions. This has proven to be very helpful in preparing them for the changes in the materials and resources that the academic library uses.

Job-listings also changed with the increase in technological demands. Libraries began recruiting librarians that had experience with electronic resources, database management, and other technological experience. The number one change in job-listings is the advent of the new position, electronic resources manager, or electronic resources librarian. Whatever the permutation of the title, these jobs require intimate knowledge of virtual resources and virtual materials on a scale that had not been seen before. Equally important to the familiarity with the resources was a familiarity with vendor language and licensing agreements that differed from the traditional print publishers. Many of the licensing agreements contain carryover language from print resources (Masango, 2005), but these agreements can be complex and daunting. Library schools accredited by the ALA have increased their technology-related course offerings in recent years (Maceli, 2015), perhaps as a response to these demands and earlier criticisms of the breadth of technological course offerings.

VIRTUAL RESOURCES

Virtual materials or electronic resources (e-resources) are now the backbone of many academic library collections. These e-resources include databases teeming with journal articles, conference papers, and book reviews, as well as e-books through various systems, and mp3s and streaming video for media collections. While the process of

collecting these materials has grown increasingly complicated over the last two decades, it was not until recently that the importance of electronic resource managers (ERMs) and management software was known. While many smaller universities struggle to catch up by hiring ERMs now, other universities were pioneers in this field. Many of them designed their own software to catalog and maintain electronic collections at a time when few, if any alternatives existed. These systems became forerunners of the electronic resource management systems that many libraries use today, but libraries are still finding that it is occasionally beneficial to create systems for themselves, because electronic resource management does not have a one-size-fits-all solution. Complicating matters is the fact that these people and systems were not deemed necessary until very late in the process. With virtual materials now taking up large amounts of the collection, both in terms of number of items and amount of money in the budget, special attention needs to be paid to the collection process.

Some materials, however, will not be digitized. There may be concerns with the process of digitizing them, or with the quality that will result from digitizing. Some digitization projects will be rejected due to copyright concerns. In these cases, there will have to be extra effort from the library to make sure their patrons are aware of what is available in the library, so that these materials are used. Still other academic libraries will have special collections, such as historical items or rare books that would not fit into the digitization model. These libraries are still a vital part of the library system at a university, and can be a great draw for researchers not just from the area but from around the globe. Marketing these physical materials alongside digital materials is vital to their continued existence. Otherwise, the materials will be deaccessioned from non-use.

The other option libraries have in the face of unused materials is to place them in archives. This is not always feasible for libraries due to budget

constraints, but it is an option to consider when faced with materials that there may not be many existing copies of. In that case, a dark archive would be the best solution. In addition to keeping these materials in the library's possession, it ensures that there will be an existing print copy of the material available to the world in the case of the online vendor going out of business. This is not only important for the institution holding the last remaining physical copy of an item, but for all libraries that would like to have continued access to items. It is, however, important to exhaust all efforts to find an extant digital copy of materials before embarking on a digitization project at the university level, because it can be costly, both in money and reputation, if mistakes are made.

There are special copyright rules associated with making copies and digitizing materials with the intent to archive them, and this can sometimes circumvent traditional copyright problems with regards to digitalization. There has been quite a bit of discussion of this in law journals since the *Authors Guild, Inc. v. Google, Inc.* case in 2013, in which Authors Guild, Inc. challenged Google, Inc.'s legal right to digitize copyrighted materials, on the grounds that Google, Inc. would be profiting from the digitization. The judge in the case ruled in Google, Inc.'s favor, citing that the in depth indexing that was the direct result of digitization was substantively transformative, because it would help researchers be able to better find books (*Authors Guild, Inc. v. Google, Inc.*, 2013). This optimistically opens the door for more digitization projects in the future, though it does not negate the time and money required for such an endeavor.

Also of particular interest in libraries when it comes to electronic resources is the cost of materials. The cost of online databases has risen steadily in the last decade and will continue to do so unless suitable alternatives are found (Zimerman, 2010). While many libraries are ready to give in to the cost and adjust their subscriptions and purchases based on it, there are alternatives already available.

One such alternative is the open access journal. The Directory of Open Access Journals lists over nine thousand journals in eighteen categories, each housing scores of smaller disciplines. Not only are these journals as good as their for-pay counterparts, but often they are actually preferred by researchers in their field, for a variety of reasons. Among these reasons are higher impact rankings and lower costs to both researchers and institutions (Joint, 2009). In addition, materials in open access journals can be supplemented by online repositories sponsored by universities. These repositories hold pre-prints of articles, as well as other materials, depending on the legal agreements the universities can make with the publisher. Many of them also house work that undergraduates and graduate students have done during their time at the university, making for a more well-rounded research body. Not only can students see what is being pursued by researchers in their field, but they can also see what work their fellow students have been doing, encouraging collaborative work within disciplines and without.

MISSION STATEMENTS AND LIBRARY FUNCTIONS

The rise of digital material required a response from the librarians out in the field doing work with patrons and acquiring and cataloging materials in technical services, but also from the administration governing these librarians. It has long been accepted that universities would have mission statements at both the university level and at the department level, which would be organized in tandem so that the university was working toward one united goal. As their roles on the university campus changed, academic libraries needed coherent mission statements that encompassed all the work that they were doing, not just their role as repositories for the physical book. Mission statements began including language about the technology, itself. For example, in Western Ken-

tucky University's mission statement, the fourth tenant is, "Maximize digital technologies and develop networked resources that enhance learning (WKU Library Advisory Council, 2013)." These new mission statements do not just deal with library technology as a burden that has to be dealt with, but as a tool for improving library resources and service. Activities that go along with these mission statements include building information commons, where students can gather and partake of information through a variety of means, including print and digital. They often also have televisions tuned to news headlines and group study areas for students to work collaboratively.

This reflects the changing library functions. What was once a quiet area for individual study is now a collaborative area filled with students working in groups to finish projects. In a way, this is just the library catching up to what has been done in the classroom for decades. Group projects had often been assigned to students in years past, and these projects were often carried out in the computer lab, or shut away somewhere in a group study room. However, libraries can, with technology, now embrace these group projects and make it easier for students to work. Now the library is a place where that work can be done easily, with access to all of the materials the group would need. Though there are librarians on duty to help students with questions, the stereotype of the shushing librarian is being shattered. This was a change necessitated by falling patron counts; as students began accessing materials on-line, libraries had to find other reasons for students to come into the library. Libraries began marketing their resources, rather than just relying on word of mouth for students to come in to the library. One of the library's best resources is the space that they provide for students to work, and this was a resource that was being underused because the space had not adapted to fit changing work demands.

Changes in how librarians view resources, and what constitutes a resource, have also facilitated

new growth in the library. Many libraries, both academic and public, have makers' spaces in which patrons can experience new technology that they might not be able to access or afford outside of the library setting. The tools that constitute makers' spaces can be either digital or physical in nature. Examples include video production software, laser cutters, and builder kits. Depending on the patron population, any or all of these might be available at the library, and are usually tailored to meet the demands of certain majors. Though maintenance and materials can be costly, these types of tools have become more resources for the library to offer, and best of all, from a statistical standpoint, they still require a physical presence in the library.

CHALLENGES TO TRADITIONAL ROLES

Some people have raised questions about the role of catalogers and reference librarians in the digital age. It is true that many of the materials acquired by the academic library are now digital and therefore come with their own records that need little to no processing, this does not signal the death of cataloging. Instead, it frees up the experienced cataloger's time, allowing them the opportunity to catalog the many and varied items that libraries might acquire that are not digital and not typical formats, such as materials used in makerspaces, in which educational resource libraries in particular are interested. This has led to a shift in how catalogers are trained, perhaps, but not a change in their necessity; in an interview, Allyson Carlyle, an associate professor at the University of Washington, emphasizes a focus on the teaching catalogers how to think in the best ways for cataloguing, rather than teaching them the rote behaviors of cataloging (Dull and Parks, 2011).

Reference librarians, too, have seen some changes to their job roles. More of an emphasis has been placed, for example, on curating collections, as the balance of print resources and electronic resources is more important than ever

(Terrell, 2015). It is up to the reference librarian to sort through not only information but also information sources, and deliver those that meet their approval to their patrons. It is also imperative that reference librarians become instructors, instructing their patrons on the appropriate ways to find information and how to vet it; patrons may be able to use search engines for answers, but they need help in selecting the best answer.

FUTURE RESEARCH DIRECTIONS

The challenges that libraries face today could potentially sound the death knell for smaller academic libraries. Students no longer believe that libraries have the monopoly on research materials, and they are right. Google scholar has become a serious challenge to libraries despite a pushback against it from librarians, because it is attached to a name that nearly everyone knows—Google. Most students, especially undergraduates, can find the materials they need for their class projects from the comfort of their own home computer. Part of the way to combat this problem is to embrace the supposed enemy. Google scholar can be a very effective tool when used properly, and if a student has the system set up correctly, they can even use Google Scholars' search features to search their own libraries. This search feature has been immensely popular with students, but it is not the only thing they need when they research. Other databases are still important to the research process and may have search interfaces more suited to the search topic at hand. Libraries have to become creative when trying to draw in students to their libraries and to use their resources.

Drawing students into the library can be accomplished through a strong connection to the faculty. Students see faculty every day and are more likely to take their advice when it comes to research and their grades. Set up library instruction sessions for classes, and be sure the students know what's in it for them when they attend. Explain to the student what the difference between searching in Google



and searching in a database is, and how the latter can actually be easier and quicker to do once they know how to do it. Running through search questions in the various databases can demonstrate to them the different types of results that they can get. Furthermore, it will show them that difficulty sometimes presented in retrieving materials when one is not logged in to a university database.

Other suggestions for making a library more inviting and adaptable to the changing digital world include creating collaborative learning spaces that integrate technology with the learning experience, adding outlets and power stations to facilitate the use of laptops, tablets, and other portable electronic devices. Information commons like the one discussed here are an excellent way to facilitate this change and are very popular subjects for funding and grants at this time. They provide a seamless entrance for the library into the world of technology and provide a valuable resource for the students. Furthermore, if a campus has a technology center that is not part of the university library system, an information common can artfully redirect traffic back to the library that the library might be losing to other sources that seem to be more convenient. This is in turn better for the students, who will have academic librarians on hand to answer questions when or if they arise, making the research process easier and encouraging them to continue their studies without getting discouraged.

All of the change that has occurred in libraries, and all of the change that is still yet to come, requires a change in how success is measured. A library whose success is measured by the number of people walking through their physical gates and using their brick-and-mortar building is going to be disappointed by the numbers that they see, even if those numbers are bolstered by things like information commons. Universities are being more flexible with their class models, offering more classes online, and therefore many people may never step foot on the university campus, much less in the university library. Rather than letting these students languish, struggling on their own to

accomplish research while their counterparts on campus have all the help they could want, university libraries should be working with advisors and classes to make sure that they are also reaching out to these students. There are a variety of ways that this can be done, including web tutorials, research guides, and videotaped presentations on library resources that can be embedded in the software that the university uses for their online classes. The tools to reach these students are available, and helping them should count in statistics, even if they are not coming to the university. All ways that the university resources are accessed should be given equal weight, because the university is still providing a valuable service, whether librarians are conversing with patrons through email, on the telephone, or in face-to-face interactions. Even keeping track of patrons accessing virtual resources through OpenURL can be an eye-opener to the way that the academic library is helping their students (Wright, 2014).

CONCLUSION

Today's library exists in a very different world from the days when they first became popular. Much of the library's content is digital, and can be accessed online. Some materials, however, cannot be digitized at this time because of technical or copyright concerns. Identifying these materials and integrating them into the curriculum is important for their continued value and to increase patron counts in libraries. Equally important to libraries is the constant innovation of new resources and outreach methods to serve the patron population. Changing how they measure success will also help libraries justify and evaluate programs, because it will highlight the way that their services have changed over the years. For the most part, libraries are already keeping track of statistics in many and varied ways, and acknowledging that these changes exist. However, they are not doing enough to put more emphasis on those reference interactions and that resource usage, when they



should be. By embracing the digital age, libraries will survive and thrive far into the future.

REFERENCES

Arms, W. Y. (2012). The 1990s: The formative years of digital libraries. *Library Hi Tech*, 30(4), 579–591. doi:10.1108/07378831211285068

Authors Guild, Inc. v. Google Inc., 954 F.Supp.2d 282 (S.D.N.Y. 2013)

DEsposito, J. E., & Gardner, R. M. (1999). University students perceptions of the Internet: An exploratory study. *Journal of Academic Librarianship*, 25(6), 456–461. doi:10.1016/S0099-1333(99)00078-6

Dull, M. E., & Parks, B. (2011). Cataloging Education and Cataloging Futures: An Interview with Allyson Carlyle. *Serials Review*, 37(2), 116–119. doi:10.1080/00987913.2011.10765361

Joint, N. (2009). The Antaeus column: Does the open access advantage exist? A librarians perspective. *Library Review*, 58(7), 477–481. doi:10.1108/00242530910978172

Kennedy, K., Russell Gonzalez, S., & Cenzer, P. (2007). Student Perspectives on Library School Degrees and the Hiring Process. *Journal of Education for Library and Information Science*, 48(4), 284–293.

Maceli, M. (2015). Creating Tomorrows Technologists: Contrasting Information Technology Curriculum in North American Library and Information Science Graduate Programs against Code4lib Job Listings. *Journal of Education for Library and Information Science*, 56(3), 198–212. doi:10.12783/issn.2328-2967/56/3/3

Masango, C. A. (2005). Digital licence agreements and their effects on acquisitions and academic library users. *South African Journal of Library and Information Science*, 71(2), 127–135.

Scripps-Hoekstra, L., Carroll, M., & Fotis, T. (2014). Technology Competency Requirements of ALA-Accredited Library Science Programs: An Updated Analysis. *Journal of Education for Library and Information Science*, 55(1), 40–54.

Terrell, H. (2015). Reference is Dead, Long Live Reference: Electronic Collections in the Digital Age. *Information Technology & Libraries*, 34(4), 55–62. doi:10.6017/ital.v34i4.9098

WKU Library Leadership Council. (2013). *Dean's Office – Library Mission*. Retrieved from <http://www.wku.edu/library/deansoffice/mission.php>

Wright, J. (2014). Considerations of a Digital Age: The Hows and Whys of Electronic Resource Management from a Collection Development Perspective. In N. Patra, B. Kumar, & A. Pani (Eds.), *Progressive Trends in Electronic Resource Management in Libraries* (pp. 17–30). doi:10.4018/978-1-4666-4761-9.ch002

Zimerman, M. (2010). Periodicals: Print or electronic? *New Library World*, 111(9/10), 426–433. doi:10.1108/03074801011089341

ADDITIONAL READING

Adam, S. (2010). Preserving authenticity in the digital age. *Library Hi Tech*, 28(4), 595–604. doi:10.1108/07378831011096259

Albu, C., & Predescu, C. (2011). Information resources development - challenge and impact among librarians and users in digital age. *Proceedings Of The International Conference On Library & Information Science / Conferinta Internationala De Biblioteconomie Si Stiinta Informari*, 143–148.

Becker, B. W. (2015). Advances in Technology and Library Space. *Behavioral & Social Sciences Librarian*, 34(1), 41–44. doi:10.1080/01639269.2015.996496

Chamberlain, E. (2012). Investigating Faster Techniques for Digitization and Print-on-Demand. *New Review Of Academic Librarianship*, 18(1), 57–71. doi:10.1080/13614533.2012.660769

Given, L. M., & McTavish, L. (2010). Whats old is new again: The reconvergence of libraries, archives, and museums in the digital age. *The Library Quarterly*, 80(1), 7–32. doi:10.1086/648461

Hazen, D. (2011). Lost in the Cloud: Research Library Collections and Community in the Digital Age. *Library Resources & Technical Services*, 55(4), 195–204. doi:10.5860/lrts.55n4.195

LaGuardia, C. (2012). Library Instruction in the Digital Age. *Journal of Library Administration*, 52(6/7), 601–608. doi:10.1080/01930826.2012.707956

Lippincott, J. (2015). Libraries and the Digital University. *College & Research Libraries*, 76(3), 283–295. doi:10.5860/crl.76.3.283

Roland, L., & Bawden, D. (2012). The Future of History: Investigating the Preservation of Information in the Digital Age. *Library & Information History*, 28(3), 220–236. doi:10.1179/1758348912Z.00000000017

Zabel, D., Shank, J. D., & Bell, S. (2011). Blended Librarianship: [Re]Envisioning the Role of Librarian as Educator in the Digital Information Age. *Reference and User Services Quarterly*, 51(2), 105–110. doi:10.5860/rusq.51n2.105

KEY TERMS AND DEFINITIONS

Academic Libraries: Libraries affiliated with universities or colleges.

Digital Age: From about 1980 on, coinciding with the advent of the World Wide Web.

Digitization: Scanning or otherwise manipulating print materials into an electronic format.

Electronic Resources: Library materials available in an electronic format.

Mission Statement: Unified theory, plan, or goal of an organization or subsection of the organization.

Digital Archives for Preserving and Communicating Architectural Drawings



Roberta Spallone

Politecnico di Torino, Italy

Francesca Paluan

Politecnico di Torino, Italy

INTRODUCTION

Archives created by architects are unique and significant sources for scientific research in understanding architecture in all its features connected with multiple disciplines (history of architecture and representation, conservation, city planning and landscape). Contemporary architectural archives contain relations which cannot easily be read and contextualized because of the documents organization and order, given by the archival necessities and the producers themselves. According to the archival science, the documents reorganization is forbidden even if it could be helpful for their analyses. As researchers in the architectural representation field, the authors believe that the mere digitization of archival drawings is not sufficient to transmit the potential of knowledge and meaning that it is caught by the direct study of the sources. The use of digital databases at microscopic and macroscopic scale allows the contextualization of the archival data, in order to help the public to elaborate information and acquire new relations. The authors investigate this phenomenon underlining the benefits and the risks connected to the digitization of the contemporary architectural archives, focusing on their communicative potentialities thanks to the tools of the digital representation, in relation to the necessities of fruition and preservation of the original archival materials.

BACKGROUND

Contemporary architectural archives constitute more than a documentary heritage, recently recognized and validated (Domenichini & Tonicello, 2004). Since the end of the Seventies, effectively, the principal North-American and North-European institutions engaged in the knowledge and diffusion of Architecture have been working hard and more carefully into conservation and enhancement of contemporary architectural drawings, which need specific standards for their description, preservation and fruition guaranteed by international associations, like the ICAM (International Confederation of Architectural Museums) and the ICA (International Council on Archives) since 1979.

The international archives take part to the ICA, whereas the archival cooperation in the European Union sphere is developed by a working group called European Archives Group (EAG), born in 2006.

Such a recent interest about the architectural documentary heritage, in fact, is carried out by different fields of research that starts from the analysis of the drawings and is elaborated according to numerous purposes, the study of the representation, the history of architecture and city planning. Its value is moreover justified by the same meaning taken: in addition to being a proof of a phase in the design process and a historical

memory, the drawing may be considered a useful document to the reconstruction of the events happened to an artifact, or it could be evaluated as a work of art. For this motive, in the same years various institutions arose, engaged in the promotion and fruition of architecture in general, and particularly focused on the safeguard of the documentary heritage, made up of heterogeneous material. This trend, still of north European and North American origin, led to the foundation of museums or hybrid structures containing collections of architectural documents. Because of the documents complexity and heterogeneity, the need to establish new descriptive standard of the objects was immediately felt within the cataloging range. On the international level, the current state provides for standard (taken from the Library System and Document) derived initially from the archival cataloging specifications and gradually structured according to the new descriptive requirements for the architectural documents. Currently the standards most commonly used are those dictated by the International Council on Archives, such as: the ISAD (G), that is the General International Standard Archival Description; the ISAAR (CPF), that is the International Standard Archival Authority Record for Corporate Bodies, Persons and Families; the ISDIAH that is the International Standard for Describing Institutions with Archival Holdings (Ministero per i beni e le attività culturali, Direzione generale per gli archivi & Servizio documentazione e pubblicazioni archivistiche, 2003).

Since the nineties, the archival administration has arranged multiple informative systems, which matter as a kind of computerized registry of the archives.

The difficulties of managing information concerning the architectural documents arise from the different materials constituting this cultural heritage: not only drawings and correspondence, but pictures, scale models, audio and video recordings. The collections have been therefore expanded and layered as new materials have been introduced as support to the architects' project activities.

Since the twentieth century, the traditional paper supports have been juxtaposed by other types of support, including acetates and derivatives from plastics. Another new type of worksheets has been introduced by the technologies and software for the architectural representation since the early nineties. These innovations have led to significant impact not only in the design phases, but also in the products in the matter of architectural drawings in digital formats. This phenomenon has involved the formulation of other cataloging approaches, as well as preservative and communicative strategies.

For this reason, several international projects have been activated to study the phenomenon of cataloging and preservation of the architectural design files born in digital format.

Among these plans, the International Research on Permanent Authentic Records in Electronic Systems (InterPARES) aims at "developing the knowledge essential to the long-term preservation of authentic records created and/or maintained in digital form and providing the basis for standards, policies, strategies and plans of action capable of ensuring the longevity of such material and the ability of its users to trust its authenticity" (www.interpares.org). This program, trained by the School of Library, the Archival and Information Studies at the University Of British Columbia, in Vancouver, has already developed four steps of research since 1998.

The archives of contemporary architecture are organized according to a structure in which the material is interfaced with the public, in accordance with the shared methodologies for the data search. The first research means that the user has to consult is the catalog of the archival groups that is an ordered and systematic list of the archival units (the individual documents) with sufficient details to identify them. This search tool is structured according to the database, a data archive of the archive itself relating to the same topic or more correlated to each other, whose management - which involves the entry, modification and research operations - is fulfilled by means of dedicated software. These databases are managed

by the DBMS (Database Management System), systems for the administration of large amounts of data that could be shared by several applications and a plurality of users, which could be increased without restrictions, supporting natively the multi-tasking and networking (Bartolomei, 2001, p. 97).

The catalog of a database system stores the metadata of each object inserted. “Metadata is a structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource” (Guenther & Radebaugh, 2004).

In an architectural archive, as in other fields, metadata could be descriptive, improving the research and the identification of the object they refer to; structural, listing the object compounds; administrative, containing technical information about the rights management and preservation of the resource. For this reason, even architectural archives have developed a strict organogram made up of different groups facing with the different aspects of cataloging. Such structure is significant because it expresses the main actual challenges to be faced: the archive management complexity related also to the digital assessment, both in static (web pages, with the names and locations of the resources “hard-coded” in the HTML) and dynamic form (thanks to specific tools supporting web applications able to extract and manipulate the information). However, metadata stored in a digital database “can help organize electronic resources, facilitate inter-operability and legacy resource integration, provide digital identification and support archiving and preservation” (Guenther & Radebaugh 2004). On the other hand, a database is more susceptible to the technological obsolescence and intentional, or unintentional, alteration, the same as the digital “objects” collected in the archive. The linked open data (LOD) about the theme of the web of data (semantic web) are currently being examined as part of the technical-computer and librarian academic field, in the North America and North Europe, as a tool for the linking of Cultural Heritage information for libraries, archives and museums, through ap-

plications that can query data, drawing inferences using vocabularies (Mazzini & Ricci, 2011).



MAIN FOCUS OF THE ARTICLE

Issues, Controversies, Problems

In the course of time, documents of architecture present different support typologies and data characters, providing insights for numerous possible interpretations.

The need for standardization in the cataloging process, required not only to the interoperability between archives and institutions, but also to facilitate the finding information clashes with a wide variety of materials in the archives of contemporary architecture, sometimes hardly recognizable in their material and graphic composition. The diversification of supports, graphic and photocopying techniques, in fact, is made more difficult to understand since the technical and technological innovations have been introduced in the design studios. The archives of contemporary architecture are defined on the basis of the project activities and are differentiated in the time and the needs of the works. Consequently, each architectural archive represents a reality per se, both for the composition and the extension of the documentary corpus, both for the arrangement and the organizing ways adopted for it. Therefore, the study of an architectural archive involves primarily the recognition of the material preserved, as well as to understand and place each document into the project refers to, as a product of architecture, engineering, city planning and design: so the skills of the archivist and the architect are essential. Therefore, also with regard to the filing, the documents heterogeneity requires the adoption of specific methodologies of cataloging, differentiating materials into: graphic ones (sketches, preparatory and/or demonstrative drawings, and survey and design tables, cartographic material and preliminary documentation); textual ones (loose papers or collected in envelopes and files,

manuscripts and typescripts bound or not, administrative documents, correspondence, advertising or documentary material, extracts from magazines, clippings, books); photographic ones (positive and negative on paper, film, glass); digital ones (audio and video recordings, computer files on various media); objects (scale models and sculptures, tools used in professional activity, paintings, samples of materials). These types could be grouped as aid materials for work or as products under preparation phase (design, definition, execution, construction site); they could be made also for different purposes (competitions, contracts, promotions) and customers (private or public). Since the twentieth century, the material traditionally used in the preparation of the architectural drawings, has been juxtaposed by new supports and techniques. In addition, the same paper differs significantly in its composition depending on the era it was produced in. The same argument is valid to the tracing papers, which differ mainly with the introduction of new materials and related techniques of graphic reproduction. So, in a contemporary architectural archive there are miscellaneous materials, including: traditional paper, cardboard, tracing paper of different matrix (traditional and/or tissue paper, vegetal parchment, etc.), sheets of acetate. Even the graphic techniques of the architectural drawing are extremely varied (pencil, charcoal, pastels, inks of various types, etc.), sometimes of artistic origin. Another peculiar aspect of the contemporary architectural archives is the use of the reproduction techniques of the drawings, necessary for the project activities: the need to represent quickly different executive variations has involved the drawings duplication using techniques that have appeared since the twentieth century, resulting from the evolution of photography and the research on light-sensitive substances. Therefore, in an architectural archive, multiple copies of the same drawing could be often grouped, made by means of different techniques such as the blueprint, the ectography (or aniline process), the heliography (or diazotypes) and the xerography, that lend themselves to the use of other supports than the traditional ones. The radex corresponds

to this case, because it is a heliography print on a polyester sheet. The architectural documents, also, show a series of infographic data that should be interpreted. In relation to the design practice and representation techniques these data refer to, drawings transmit information in a univocal manner as regulated by certain standards such as the representation scale, orientation, size, dimensions, thickness of lines, etc. In this context, the interpretation of the drawing is oriented towards an unequivocal direction for the identification of the graphic message, while the interpretation given by the public may be released according to the different branches of research (design, historical, etc.). An enhancement strategy adopted in the management of architectural archives is the communication, with multiple strategies and outcomes, however made difficult by the nature of the archive itself, often inaccessible or difficult to reach, where the documents could not be always consulted, risking the loss.

Architectural digital archives, as well as they provide research systems that could be queried to find the documents, are managed by methodologies useful to the safeguard and enhancement. The knowledge of the archive is the first key moment of safeguarding: for their constitution, archives are the containers of information not to be easily grouped without a direct survey of the institutions dedicated to the fruition. The first step to the knowledge is the catalog consulting; this initial research, even if it is remotely conducted, usually provides the descriptive metadata of the objects. Although this information is essential to the identification, often documents cannot be remotely accessed, because of their lacking digitization. In the case that the architectural drawings are inserted into the digital databases, the files displaying turns out to be quite unsuited to the research and study of the document, because the file of the image display, rarely downloadable, has a low-resolution, preventing the reading of the graphical information. In addition to the inadequate understanding of the graphic message, the lack of the dimensional scale prevents the recognition of the drawing dimen-

sions and the presence of the written measures may be relatively useful if they are insufficiently visible. These considerations emerge during the web navigation. The scanned drawings could be sometimes downloaded; however, also in this case the resolution remains most often inadequate. Despite the innovative technologies related to the scanning offer ever more efficient performance, the architectural drawing inserted in the database remains an icon that could be partially zoomed in. These limits are attributable to several factors: if the images are bound by strict regulations of copyright which force the user to a superficial study of the drawing, on the other hand the costs of digitization of whole architectural archives are still too high. The transition from the analog source to the digital product by the scanning produces files with different mass-diffusion standard extensions (.pdf,.tiff,.jpeg,.png), thus promoting the dissemination of the material to a wide public.

Photographic digital reproduction can be considered an operation of preventive conservation with two purposes: to avoid the damage resulting from the direct consultation and an unsuitable handling of the drawings, and to create security copies available for the production of “packages” for traveling exhibits, in place of the exposure of the original drawings. This procedure was adopted and promoted in Italy by the Centro Archivi, which has defined three levels of photographic definition of archival drawings: 300 dpi in full scale TIFF as facsimile, with the metric and chromatic scale; 150 dpi-1062 pixels JPG, 72 dpi-510 pixels JPG (Pesce, 2009).

Another factor inducing the institutions dealing with the safeguard and enhancement of the contemporary architectural archives to redefine the preservation and communication strategies is the presence of documents in digital format, extremely fragile from the point of view of the permanent preservation. Their use is bound by the characteristics of digital information (authenticity of the original digital file as a copy on the web, maintaining the functionality of the design produced by the interactive software). In this sense,

international conferences are promoted as the place and opportunity to debate about the hybrid archives management, trying to develop projects related to the archives digitization and their fruition in the virtual environment (Peyceré & Wierre, 2008). In Europe, the European Archives Group is active in the dissemination of the archival heritage, even the specific one about contemporary architecture. In that sense, a portal has been created for the research in all the European archives, called Archives Portal Europe, in which cultural heritage related to the architects and engineers' archives is expected as one of the topics. One aspect that EAG would advance is the collaboration with Europeana, for the archival documentation made accessible and searchable in the multidisciplinary context of cultural heritage conveyed by archives, museums and audiovisual online collections. At the moment, Europeana is in an expansion phase with regard to the disclosure of the contemporary architectural archives.

The case study of the Italian archival heritage, characterized at the same time by a large amount of materials (due to its unique historical and architectural tradition) and an accentuated fragmentation, lends itself to interesting in-depth analysis. Because of the extreme fragmentation characterizing the Italian heritage, widely diffused throughout the territory and therefore managed according to a polycentric model, other enhancement strategies have been adopted on the premise of making archival resources accessible, historically contextualized, to a wider public (Maiello, 2013). One approach useful to such disclosure is the creation and management of archival networks.

The exhibition of drawings from the architectural archives is, finally, another way of promotion and divulgation of such fragile heritage. The physical exposure, however, is not always practicable especially for preservative reasons (extreme fragility of the works, impossibility of their mobilization into other spaces). For this motive, the online virtual exhibitions fulfill effectively this purpose, allowing the physical safeguard of the drawings and the active participation of visi-

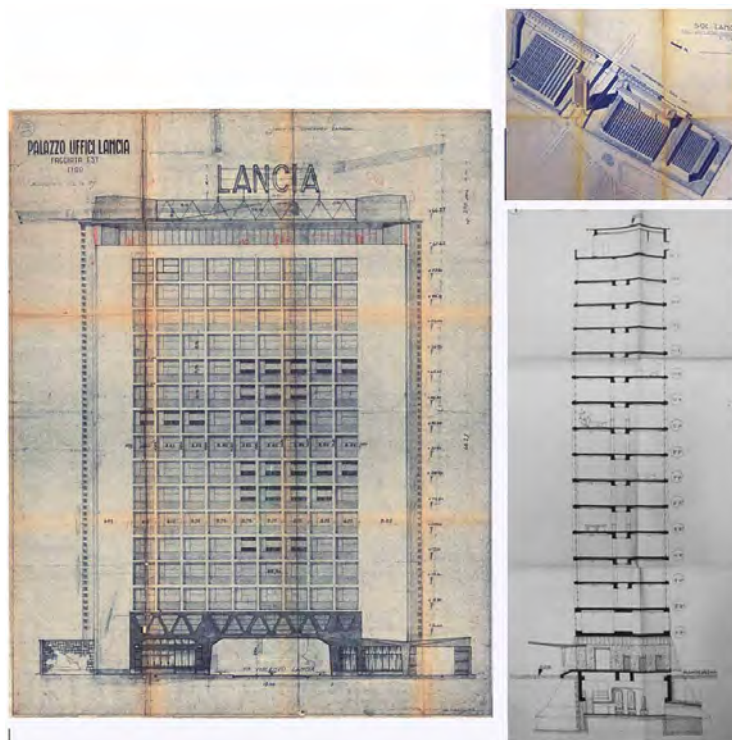
tors in a multimedia and interactive exhibition. In this context, the technological development in the telecommunications and computer graphics field has made possible new teaching experience approaches in the cultural heritage fruition, using the Web or mobile devices to display also digital products such as animations, applications for immersive experience of virtual and augmented reality (Christmas & Saccoccio, 2012). In the case of the archives of contemporary architecture, digital objects can be scanned drawings or three-dimensionally rendered ones, linked by a theme, made accessible thanks to technologies currently available through an architecture of system designed to offer the users engaging experiences. The result is the creation of dynamic products, updatable also by the public involved in the resources exploration according to free interpretative readings. The on-line virtual exhibition is also developed according to thematic itineraries that are paths that allow the public an extremely independent approach in

relation to the space-time location and the origin of the documents. Actually, a kit dedicated to the creation of online virtual exhibitions exploiting the open source software has been developed, taking advantages of the Web potentialities to make the invisible visible and available this sector of cultural heritage risking the oblivion, such as the extremely scattered architectural drawings in public and private archives.

SOLUTIONS AND RECOMMENDATIONS

From the point of view of the architectural representation research, the authors believe that the knowledge data, resulting from the direct documentation analysis, needs to be interpreted, represented and communicated (Spallone & Paluan, 2015).

Figure 1. Lancia Office Building, 1954, elevation and cross section (Archivio Rosani, Laboratorio Beni Culturali - Politecnico di Torino)



The digital model could be an effective database for collecting and synthesizing these analyses.

The extreme care in implementing 3D reconstructions is the final result of a new attention paid by the stakeholders to the communication capability of information technologies, both as for the large audience and the narrow scientific community. The simplest language is required to the producers of these models (De Francesco & D'Andrea, 2008), in order to share and diffuse information in a readily understandable form to a wide-ranging spectrum of public. Thanks to these models, the era of graphic papers that only specialists can decipher is over.

The authors are carrying out a research project in which they propose methodologies and techniques that can actively contribute to widen the knowledge and the understanding of Archival

Drawings, intended as Cultural Heritage, and support their sharing, in the firm conviction that safeguard and fruition should go together.

The analytic activities start from reconstruction, through the graphical analysis and the re-drawing of the archival drawings, possibly supplemented by other documentary sources, which can enrich the mainly geometric data arising from technical drawings, and could provide information on colors, building materials and techniques and characteristics of the original surrounding environment.

During this representation phase, the re-drawing of plans, elevations and sections involves the check of projective correspondences, aimed to three-dimensional synthesis, and the possible need to interpret inconsistencies and blanks. In this way the act of modeling may enhance knowl-

Figure 2. L'Oréal-Saipo plants in Settimo Torinese, 1959, perspective view, plans, elevations and cross-section (Archivio Rosani, Laboratorio Beni Culturali - Politecnico di Torino)

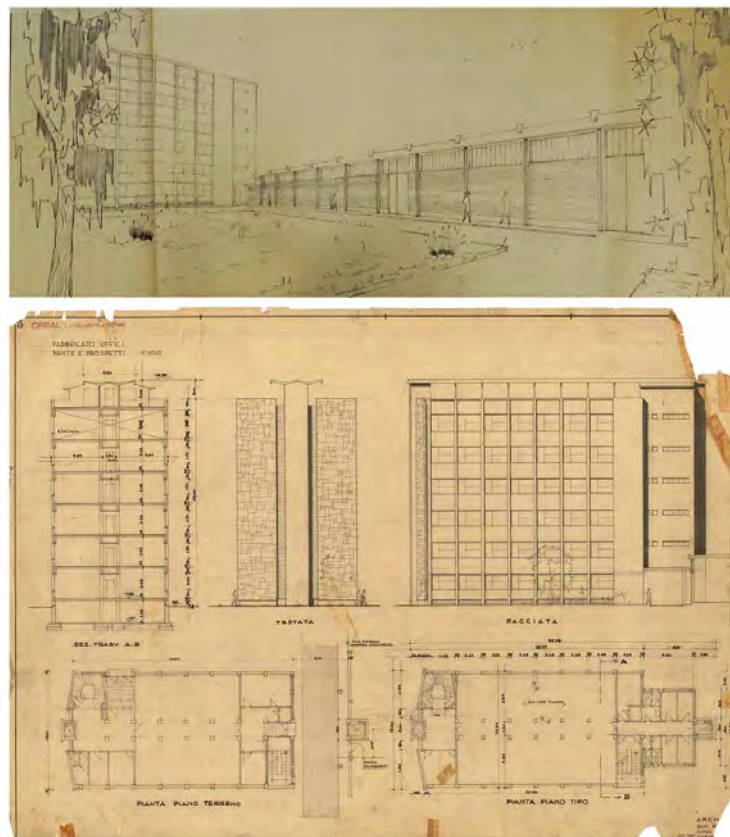


Figure 3. L'Oreal-Monsavon plants in Aulnay-sous-Bois, 1958, perspective view and sections (Archivio Rosani, Laboratorio Beni Culturali - Politecnico di Torino)

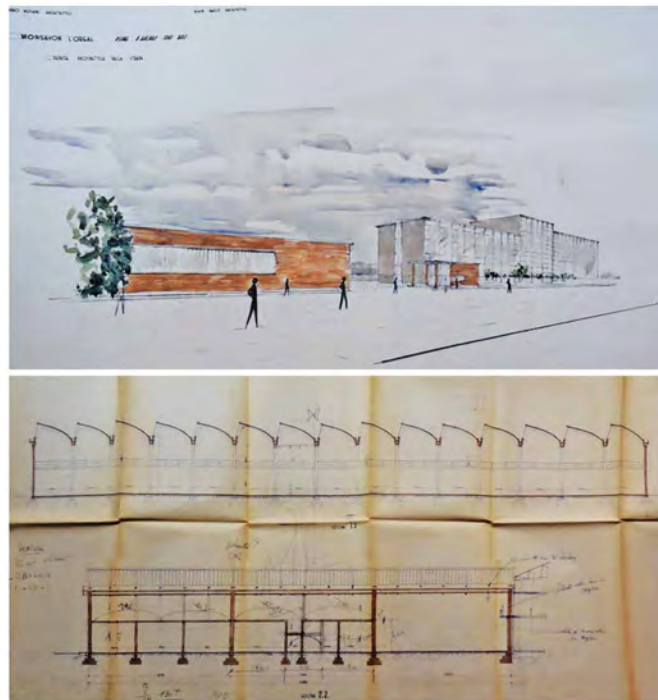
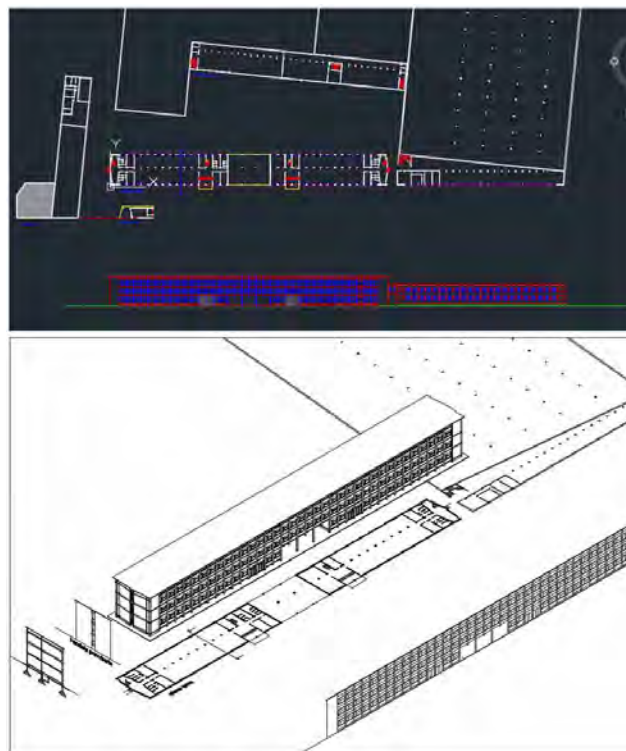


Figure 4. L'Oreal-Monsavon plants in Aulnay-sous-Bois, digital reconstruction (Digital drawings and 3D model by Francesca Paluan)



edge acquisition, not only for others, but for the modeler as well (Vandevyvere, Neuckermans, & De Jonge, 2005).

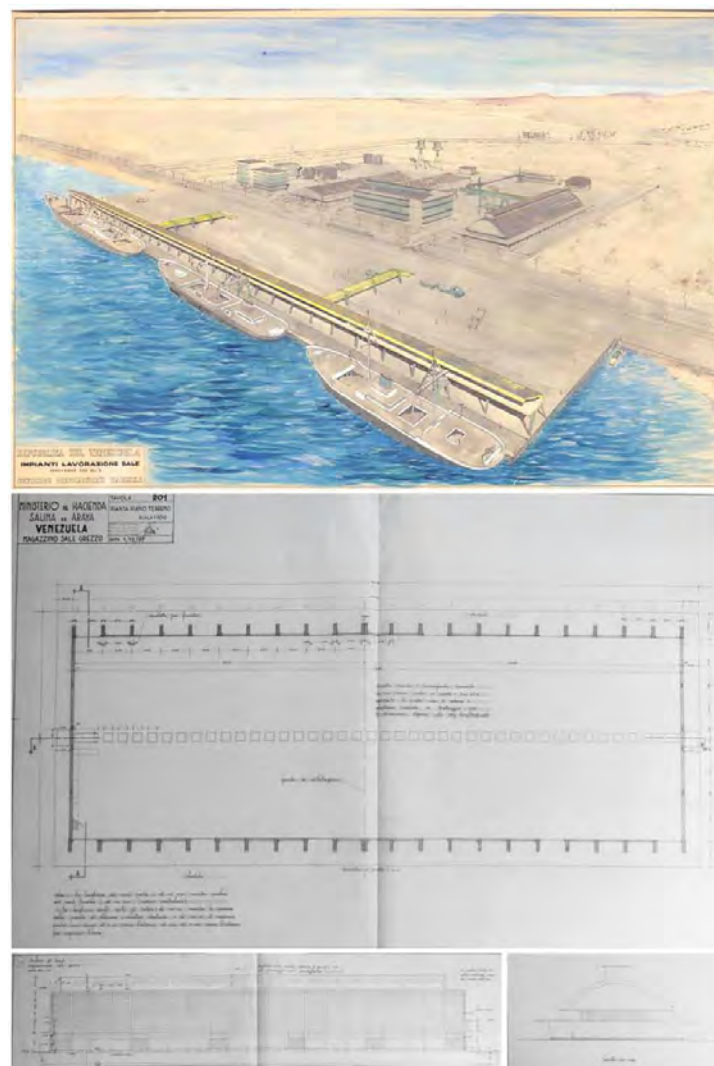
Regarding the next stages of modeling, the drawings created are the basis of the 'blue-prints technique', in which the two-dimensional drawings which write-up archive drawings in vector format are placed on orthogonal planes in the space of the digital work, so they can directly infer the geometric and dimensional references.

This methodology has been applied to Rosani Industrial Architectural Studio's archive, since the Fifties mainly involved in projects for major

and medium-sized industries based in Italy and abroad, with the collaboration of several masters of contemporary architecture, such as Giò Ponti on the design of Lancia Office Building in Turin, Carlo Mollino on the complex INA CASA in corso Sebastopoli and Pierluigi Nervi on the plants of L'Oreal in Settimo Torinese.

Rosani's archive collects a historical and cultural heritage today in danger of dispersion and forgetfulness: consultation is reserved for few experts, a non-specialist potential public cannot immediately understand the data, the planned buildings are earmarked mainly for the industrial

Figure 5. Salina de Araya, Venezuela, 1957, perspective view, plan, elevation and cross-section



use, so they are difficultly communicable because they represent an economic and social logic little different from the Cultural Heritage is generally subjected to. The hypothesis of a physical exhibition of the drawings considered the most exemplar would be a temporary event that will surely enhance the archive, but then it could not guarantee a return of interest from the public.

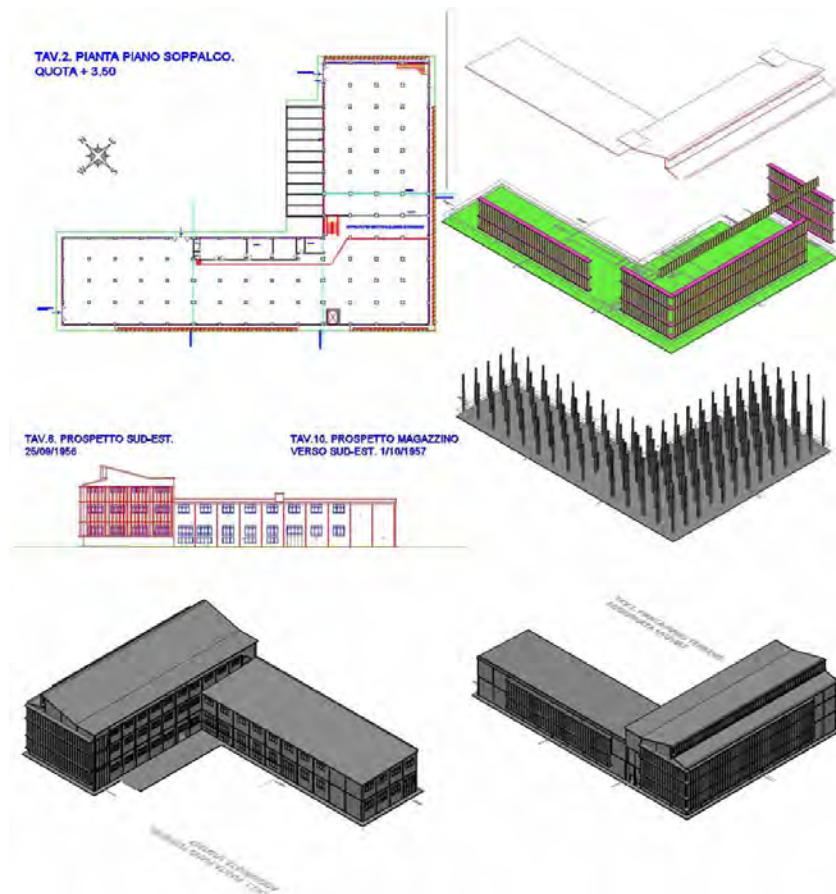
These reasons let hypothesize and, if possible, apply the strategies considered the most suitable for the enhancement and the safeguard, through the representation techniques and methodologies for the interpretation and communication of the projects. These objectives, under development, will be the basis of a meta-project Rosani's archive will be made ready for.

The magnitude of the archival collection obliged to focus only on some of the most significant projects by Nino Rosani. The selection was conducted according to certain parameters: the exemplarity of the plans; the existence of design variants; the inaccessibility of the existing building (because it is still in use), and the importance of transformations and expansions on the structure over time.

FUTURE RESEARCH DIRECTIONS

The process launched by the more advanced ICT about enhancement strategies of Cultural Heritage through communicative plans, is also applicable to

Figure 6. Salina de Araya, Venezuela, 1957, digital reconstruction (Digital drawings and 3D model by Francesca Paluan)





the archival heritage of contemporary architecture, partially digital born and designed to exploit the interactivity and multitasking from the beginning (Garagnani, 2010). Digitization is a subject of debate among the various institutions, so that drawing unequivocal conclusions that could be applied to the different areas of expertise is still impossible. The need to address this issue arises from the awareness that about the 80% of digital documents created in the nineties will be (or has been) lost due to the rapid technological evolution whose preservation problems have not been solved yet. The digitization of the architectural archival heritage is penalized by the high initial costs and software and hardware maintenance for data management, but it is a potential storage medium that ensures the enhancement broadening the range of public, running as security storage of information. The use of ICT is not necessarily limited to the communication: its propensity as a means of education facilitates the learning thanks to the immediacy of message delivery and the simultaneous establishment of hyperlinks.

CONCLUSION

Architectural digital archives have been established to prevent the risk of their dispersion and oblivion: the fragility of such cultural heritage depends on the low visibility except for a highly specialized field. Digitization is therefore an approach that first tries to make recognizable this hidden heritage, making it available to an increasingly connected public. Breaking down the barriers of space through navigation and remote access to the data and the temporal ones through the devices running in real time, the possibility to view documents is finally satisfied, although such documents are not always equipped with the information to a more specific analysis regarding the iconographic message.

REFERENCES

- Bartolomei, C. (2011). Strutturazione dell'A.I.M. (archivio Informativo Multimediale) per il progetto. *DisegnareCon*, 7(4), 188–197.
- De Francesco, G., & D'Andrea, A. (2008). Standards and guidelines for quality digital cultural three-dimensional content creation. *Proceedings of 14th VSMM 2008. Digital Heritage*, 229–233.
- Domenichini, R., & Tonicello, A. (2004). *Il disegno di architettura. Guida alla descrizione*. Padova, Italy: Il Poligrafo.
- Garagnani, S. (2010). Multimedialità ipertestuale per banche dati di progetto. In S. Brusaporci (Ed.), *Sistemi Informativi integrati per la tutela, la conservazione e la valorizzazione del patrimonio architettonico e urbano* (pp. 199–209). Rome, Italy: Gangemi Editore.
- Guenther, R., & Radebaugh, J. (2004). *Understanding Metadata*. Bethesda, MD: NISO Press. Retrieved April 2, 2014 from <http://www.niso.org/publications/press/UnderstandingMetadata.pdf>
- Maiello, R. (2013). Politiche e legislazione dell'Unione Europea per la digitalizzazione del patrimonio culturale. *DigItalia*, 8(2), 9–23.
- Mazzini, S., & Ricci, F. (2011). Linked archival authority data. Una sperimentazione sui linked open data nel sistema regionale informativo IBC Archivi. *DigItalia*, 6(2), 144–149.
- Natale, M. T., & Saccoccio, R. (2012). MOVIO – kit per la realizzazione di mostre virtuali online. *DigItalia*, 7(2), 138–143. Retrieved April 5, 2015 from <http://digitalia.sbn.it/article/view/591/426>
- Pesce, D. (2009). Progetti di digitalizzazione relativi agli archivi del MAXXI architettura. *DigItalia*, 4(2), 89–96.
- Peyceré, D., & Wierre, F. (2008). *Architecture et archives numériques: l'architecture à l'ère numérique: un enjeu de mémoire*. Gollion, France: Infolio.

Spallone, R., & Paluan, F. (2015). Contemporary design drawings as cultural heritage: interpretation and communication. Towards a digital archive of Rosani's industrial projects. In Gambardella, C. (Ed.), *Proceedings of XIII International Forum Le Vie dei Mercanti. Heritage and Technology*. (pp. 283-292). Napoli, Italy: La Scuola di Pitagora Editrice.

Vandevyvere, H., Neuckermans, H., & De Jonge, K. (2005). Digital historical reconstruction: case studies of an interdisciplinary task. In *Proceedings of EuropIA.10. Augmented Heritage*.

ADDITIONAL READING

Albisinni, P., & De Carlo, L. (Eds.). (2011). *Architettura disegno modello. Verso un archivio digitale dell'opera di maestri del XX secolo*. Roma, Italy: Gangemi.

Bruno, A. Jr, & Spallone, R. (2015). Cultural Heritage conservation and communication by digital modeling tools. Case studies: minor architecture of the Thirties in the Turin area. In I. S. P. R. S. Annals (Ed.), *II-5/W3* (pp. 25-32). Göttingen, Germany: Copernicus GmbH. doi:10.5194/isprsannals-II-5-W3-25-2015

Irace, F., & Ciagà, G. L. (Eds.). (2013). *Design & Cultural Heritage. Archivio Animato / Animated Archive*. Milano, Italy: Electa.

Maldonado, T. (2005). *Reale e virtuale*. Milano, Italy: Feltrinelli.

Manovich, L. (2001). *The Language of New Media*. Cambridge, Massachusetts: MIT Press.

Migliari, R. (2008). *Prospettiva dinamica interattiva. La tecnologia dei videogiochi per l'esplorazione di modelli 3D di architettura*. Roma, Italy: Edizioni Kappa.

Paluan, F. (in press). Telling architecture, Giving voice to archives. *Proceedings of the International Conference Cultural Heritage. Present challenges and future perspectives*.

Spallone, R. (2015). Reconstruction, modeling, animation and digital fabrication of 'architectures on paper'. Two ideal houses by Carlo Mollino. *SCIRES-IT*, 5(1), 101-114.

Terenzoni, E. (2012). L'archivio in rete: digitale e virtuale tra criticità e valorizzazione, il portale tematico degli architetti nel sistema archivistico nazionale. *DigItalia*, 7 (2), 67-83. Retrieved April 5, 2015 from <http://digitalia.sbn.it/article/view/581/411>

KEY TERMS AND DEFINITIONS

Architectural Document: The first definition was formulated in the conference "Towards standards for Architectural Archives" in Washington DC in 1982, with the participation of the International Confederation of Architectural Museums, the International Council of Museums, the International Council of Monuments and Sites, the Union International des Architects and UNESCO. It "designates any documentary material and anything annexed to it being correlated to the history, the theory and the practice of Architecture and the domains connected, whichever are the supports and the physical features [...] created or received by public or private entities during their activities and [...] collected, wherever it comes from".

Archive: It could be defined as a collection of historical documents or records containing information about and relations with a person, a group of people, institution, place; it is also the place where such documents are kept. It is a Cultural Heritage to be preserved and enhanced in order to safeguard the memory it is the depository of.



Database: A data archive about the same subject or a set of subjects connected one to each other, managed by dedicated software. The data are defined as metadata such as they identify and describe the resources and giving location information.

Descriptive Unit: A component of a finding aid to be used for searching. It is usually one record – or a group of records – within a fond, which is described by at least title, date of creation and reference code.

Digital Object: A digital representation of archival materials. It could be an image, text, sound, video, 3D. It could be also the original object when born digital.

Fond: An aggregation of documents that originates from the same source.

Intelligent Systems and Creative Industries for Cultural Heritage: Important and integral parts of systems that support product design for the Cultural Heritage knowledge, fruition and divulgation. The Marcus Institute for Digital Educational in the Arts (MIDEA) illustrated and identified the potentialities of the emerging technologies for the learning in the Horizon Report in 2010, focusing on the area of the mobile and social media, augmented and virtual reality, localization systems, the semantic web and the action computing.

Massive Digital Libraries (MDLs)

Andrew Philip Weiss

California State University – Northridge, USA

INTRODUCTION

To provide a clearer framework for analyzing the growth of digital libraries, Weiss and James have proposed the term Massive Digital Libraries (MDLs), which is based on the size, scope and increasing scalability of digitized book collections. Such MDLs rival the size, breadth, and depth of a physical library's print holdings, and often reach a scale seen among library consortia collections. (Weiss and James, 2013a, 2013b, 2014, 2015; Weiss, 2016)

The root of the concept begins in late 2004 when Google made its “resounding announcement” to digitize millions of the world’s books—including works still under copyright protection—and to place them *all* online. (Jeanneney, 2005) Jean-Noel Jeanneney, head of Bibliothèque nationale de France at the time, interpreted Google’s planned project as a wake-up call for European countries. Failure to catch up to the American company, he argued, would result in significant problems for non-American organizations.

Twelve years on, it is hard to imagine that Google’s desire to create an online digital library on such a large scale should have come as such a shock. Yet at the time Google caused significant hand-wringing and soul-searching among institutions traditionally charged with producing or preserving cultural artifacts. (Jeanneney; Venkatraman, 2009) In retrospect, the controversy seems almost quaint in comparison to the current crop of issues – especially the current “disruptions” of established economic models by Uber/Lyft, Facebook, Twitter, Spotify, Snapchat, e-readers, et al. and the encroachments on civil rights via electronic digital surveillance and other intrusions of privacy.

DOI: 10.4018/978-1-5225-2255-3.ch454

A number of mass-digitization projects have grown in the wake of Google’s announcement, including the *HathiTrust*, *Internet Archive*, *Digital Public Library of America (DPLA)*, *California Digital Library*, *Texas Digital Library*, *Gallica*, and *Europeana*. These projects each transcend their roots as localized digital libraries and have simultaneously adapted to and altered the digital landscape. These various MDLs have allowed for and contributed to the ascendancy of our current mass-digitization online culture.

This chapter will describe the characteristics of Massive Digital Libraries (MDLs) and outline their impact upon contemporary information science issues, especially with regard to digital collection metadata, copyright and the diversity of the source collections. Traditionally, libraries have been created to serve particular communities defined by geography, intellectual discipline, or specific end users. However, MDLs in their current trajectories promise—for better *and* for worse—to transcend such limits.

BACKGROUND: DEFINING MASSIVE DIGITAL LIBRARIES

Defining Criteria

Massive Digital Libraries (MDLs) describes a specific class of digital libraries that correspond to the size of a traditional, large brick-and-mortar library. Although other disciplines have discussed digital libraries and archives in terms of computer science, such as in the Very Large Digital Library (VLDL) movement, none have framed the discussion in terms of the principles of library science

or the services and content access provided by an actual, working library. (VLDL, 2011)

The following list of characteristics has been proposed to help define MDLs:

1. *Collection size: surpasses 500,000 texts; prime MDLs comprise tens of millions of texts;*
2. *Acquisitions, collection development & copyright: numerous partnering members contribute content regardless of author or copyright holder permissions and regardless of end-user needs;*
3. *Content type: mass-digitized print books; the resulting searchable digital corpus of texts becomes as important as the individual titles;*
4. *Collection diversity: diversity is dependent upon self-chosen partner members, which can reflect distortions or biases inherent to the source collections;*
5. *Content access: varying degrees of open access exist within MDLs; content is searchable through single, uniform interfaces (search engines & portals) representing all the collections as members of a single entity regardless of source;*
6. *Metadata: Metadata is gathered and aggregated from multiple sources, with a reliance on new digital description schema;*
7. *Content / digital preservation: consortium members provide long-term digital preservation strategies at local levels as well as “in the cloud”.*

These criteria and their attendant issues, though not necessarily unique to digital libraries, require different approaches when dealing with a Massive Digital Library. The issues involved with aggregating millions of previously published print materials into one uniform, yet decentralized, conceptual and online digital space become more complex as size increases. It is important to differentiate MDLs from their smaller counterparts as they are more difficult to police and analyze, especially

with metadata uniformity, copyright compliance and rights ownership. The larger the institution or system the more unwieldy and slow-to-change it may become. It is therefore important to remain cognizant of how these MDLs approach the common characteristics of books in ways that stretch the boundaries of the original print medium. Print books, for better or worse, remain static during their lifetimes, changing only when new editions are created. Yet e-books could change into ongoing “works in progress”; they can be altered with relative ease. Identifying MDLs as unique entities would allow scholars and researchers to both utilize and safeguard these newly-created digital corpuses.

REPRESENTATIVE MDLS

The following is a look at several representative MDLs and their defining characteristics. This section will also help readers get a sense of how MDLs stack up against each other. Criteria described for each MDL include their starting dates, *modus operandi*, estimated sizes of the collections, subjects collected, the number of partner institutions, and the languages covered in their collections.

American-Based, Anglo-Centric MDLs

Google Books (25 Million Texts)

The *Google Books* project started in 2004. Envisioned by Google co-founder Sergei Brin as the replacement of a library’s card catalog, *Google Books* currently holds approximately 25 million volumes within its collections, though exact figures have been difficult to verify.

Google Books is characterized in part by its extremely robust and flexible search engine. When searching in this MDL, a large amount of set results will often appear. Mistakes in the search are often auto-corrected. Their stated goal



is the digitization of all the books in the world, a number that reaches about 129,864,880. (Jackson, 2010) They work with authors directly through their Google Books Partner Program and with public organizations through their Google Books Library Project, which includes about 22 public and academic institutions and consortia from several countries, including Japan, Germany, England, and Spain. (Sato, 2010) Overall collection diversity is represented with a large number of languages represented and accessible, though the bulk of the source material in the digitized corpus is in English.

The *Google Books* corpus is open to anyone able to use a computer that is connected to the Internet. However, due to copyright restrictions and the threat of litigation, any materials falling outside the public domain have been closed to full-text viewing. Google provides users with three levels of accessibility to the texts beyond the basic metadata record: 1) Snippet View, which is a few sentences of the book, 2) Partial view, which is often several to dozens of pages, and 3) Full view, which is the complete text. Additionally, Google has provided access to the data sets from the digitized corpus for researchers to utilize. One of the most promising tools developed for this text mining is the Google Ngram viewer.

HathiTrust (14.6 Million Texts)

HathiTrust started in 2008. It currently holds 14.6 million volumes (representing 7.3 million unique titles) within its collections; 5.6 million volumes (38% of the total) are in the public domain. The *HathiTrust's* main goal is to meet the needs of the members of its partnering institutions. These partners include nearly 72 world-class academic research and national libraries. Notable partners include the Library of Congress and the California Digital Library. Collection development is focused on gathering digitized books and journals from partner institutions and in-house digital initiatives and those books digitized in part by *Google Books* and the *Internet Archive*.

As one might expect, English is the main language of the archive. Roughly 3.9 million of their books, representing 49% of the unique holdings, are in English. The second largest represented group is German, followed by French, Spanish, Russian, Chinese, and Japanese. Given the large Spanish-speaking population in the United States, the small and disproportionate amount of Spanish-language materials is surprising. However, given that the audience is college students and faculty, German and French (second and third respectively) have long represented the cutting-edge of scholarship in more disciplines than Spanish. (HathiTrust, 2016) This will be discussed in greater detail below.-

HathiTrust is characterized by a robust search engine and extremely well-organized metadata and records management. Where *Google Books* has issues with the retrieval of specific dates and other types of metadata, *HathiTrust* performs information retrieval and document display with superior results. The system allows, however, only two types of views to the non-member user: 1) a limited view, which provides only search results to retrieve word counts from queries (i.e. full-text search), and 2) full-view, which provides the full text – but this is reserved primarily for materials in the public domain or those with specific author-provided permissions for open access.

Access is open to anyone able to use a computer connected to the Internet, but partnering institutions and their members benefit from extra services such as content storage, data preservation and copyright clearance/usage consultation. Although the current size of the collection is smaller in comparison to *Google Books*, the ability for one to find information more easily should make the *HathiTrust* the go-to MDL for *all* libraries and their patrons.

Internet Archive (10.1+ Million Texts)

The Internet Archive is also a unique addition to the Massive Digital Library category. On one hand their main focus appears to be on archiving the

Internet via its *Wayback Machine*. One of their stated goals, in fact, is to create archival artifacts of the web itself. This is important as the Internet may replace many of the print media sources available, including books. Unlike *Google Books* or *HathiTrust*, which focus heavily on collection sizes, *Internet Archive*'s primary concern has been less focused on the number of titles acquired. Instead, the *Internet Archive* gathers and approves only those materials in the public domain or that have author permissions. As a result, their print-derived digital collections are smaller but completely accessible.

Their advantage is also in providing links to several mass-digitization projects and allowing these projects to exist under their own names as well as within the Internet Archive web pages. Projects such as the *Biodiversity Heritage Library*, *Project Gutenberg*, the *Universal Library* and *Children's Library* are worthy open access book digitization initiatives aimed at providing people with free access to resources, all falling under the *Internet Archive* umbrella.

Their difference is also in one of scope, with their emphasis on audio recordings, video, electronic video gaming, and other unique multi-media collections. Additionally, the *Internet Archive* also provides access to about 9,500 audio recordings of books, including literature and poetry.

Other MDLs of Note

In recent years several other U.S.-based MDLs of significance have begun to appear. The most well-known of these, *The Digital Public Library of America* (DPLA) is also one of the newest. Due to aggressive partnership and digital hub development, its growth since October 2010 has been exceedingly quick. It currently holds nearly 14 million digital items, a huge increase from the 2.5 million at its initial launch. The collections specialize in a range of materials from books to digital archives. The scope ranges from materials for students and teachers at the K-12 level, and

for universities up to the graduate-student level. The main types of partnering institutions are broken down as follows: public libraries, 23.38%; university libraries, 22.19%; historical societies, 11.33%; museums 11.33%; archives, 7.50%. It is likely one of the major up-and-coming MDLs and has, it appears, learned from the experiences of its predecessors. (DPLA, 2014)

Additionally, the *California Digital Library* and *Texas Digital Library* continue to grow. Although they serve more localized, state-based interests, their open online accessibility ensures widespread impact. Many of CDL's digital book collections appear as subsets within Google Books, the HathiTrust, DPLA, and Internet Archive. (CDL, 2016) The Texas Digital Library's collections focus on large-scale institutional repository development and issues inherent to scholarly communication, especially ETDs, data management planning and archiving, digital preservation and digital objects. (TDL, 2016)

Non-Anglo-Centric MDLs of Note

Europeana (53+ Million Items / 3 Million Texts)

The *Europeana* digital library aggregates multi-media content from more than 130 institutions across Europe, including video, image, text and digital artifacts. It is a multi-lingual, multi-cultural endeavor. There appears not to be one dominant language in the project. With a huge collection of over 3 million texts the digital library should be considered a leading MDL in the field. Their emphasis on interoperability and a strong metadata schema allows users to refine searches and hone in on more relevant search results easily. The main focus for the library is to ensure that digital materials of numerous cultures remain freely accessible to all. The project advocates the strongest pro-Public Domain stance of nearly all the MDLs. It is a vital and important MDL as a result.



Gallica (2.5+ Million Items / 680 Thousand Texts)

Despite its relatively smaller book collection size, *Gallica* is noteworthy for several reasons. One, it is a primarily France-centric and Franco-phone Massive Digital Library. Its partners include many universities and museums in France but also the National Library of Brazil and the Library of Congress. There are books offered in English, but it is a small amount compared to other MDLs and dwarfed, obviously, by the size of the French language collection. *Gallica* is also, in many ways, an MDL counterpoint to *Google Books*, with a very different philosophy and approach to the digital library.

MDLs: MAJOR ISSUES AND CONTROVERSIES

MDLs, Mass Aggregation and the Accuracy and Veracity of Results

Google Books' policy for developing their Massive Digital Library appears to be very simple: *digitize everything*. Such ambition to collect all of the world's knowledge is not unheard of in the history of libraries, yet these are time-consuming and costly efforts requiring extreme amounts of funding and personnel to do them adequate justice. However, it is as ambitious as it is unrealistic. It is possible complete digitization may never happen as new books are still being published in print form yearly. Furthermore, not all books are available with digital counterparts or can even be scanned accurately and adequately. Some works made with high-quality materials, odd-sized or hand-crafted paper, or three-dimensional forms are preferable to many readers as solid objects and not as two-dimensional digital copies. Some countries such as Japan have traditional binding procedures that are unique and not easily transferable to digital formats.

The estimation of 129 million books may also be flawed. Relying on ISBN numbers is not accurate as many countries did not adopt ISBNs until the 1960s or 1970s. The differentiation between serials, multi-volume works, multiple editions and the concept of "books" is unclear as well. Does Google consider the 51 volumes of the Harvard classics as one book, with one record as in a library catalog or the HathiTrust, or are they 51 books? Does the second edition of volume 13 of the Harvard Classics really warrant digitization if the first edition is available? A "digitize everything" approach does not suitably answer these questions and an assumption of universality becomes a disservice to all.

Endeavoring to digitize all the books is not so much a collection policy as it is an avoidance of one on a massive scale. The results are akin to a typical Google keyword search: a mixture of high and low culture with only relevance ranking, which can be manipulated by both black and white hat Search Engine Optimization (SEO), to determine whether it is a satisfactory result. One concrete example of this issue has been the inability for the Google Books search engine to accurately retrieve and coherently display single works of multiple volumes. A search of the Harvard Classics results in a jumbled non-sequential retrieval of titles. Unless one had known before-hand that *Harvard Classics* contains 51 volumes, one might never get an accurate accounting of the texts from Google's set result.

At the same time, digitizing all content will also include books of questionable natures, including the literatures of hate, pornography, and other provocative or even dangerous materials. Yet, despite their intention to digitize everything, it is unclear due to Google's secrecy whether they have digitized materials that are hateful or offensive to various groups.

One must be wary of this approach as well, because not only will books be missed - especially those not held by Google's Partners - but also a lot of unnecessary, spurious, or poorly written

materials will be added to the collection. This in turn dilutes the rationale for collecting something.

Additionally, the accuracy of the metadata needed to track and provide information retrieval in meaningful human-readable ways is also at risk. James and Weiss (2012) found a high rate (36%) of errors in their study of Google Books' metadata. Other scholars have been pointing out the "train wreck" that is Google Books' metadata for quite some time. (Duguid, 2007; Nunberg, 2009;)

Weiss and James have also examined rates of error among metadata and in scanning results. James (2010), Weiss and James (2013a, 2013b, 2014, 2015), and Weiss (2016) have found significant evidence of errors in Google books.

All of these issues, from responsible collection and content aggregation to the accurate retrieval of information, including both reputable and appropriate sources as well as accurate information for specific bibliographic units, are intertwined. Accuracy and veracity of the content goes with its ability to be perceived as trustworthy by the MDL's users.

MDLs and Copyright

It would be an understatement to say that MDLs have had some impact on copyright law. MDLs can be seen as logical extensions of the Internet's early promise of the information superhighway and the democratization of knowledge. Yet there has been intense backlash from all sides against MDLs. Indeed, the two largest of the Massive Digital Libraries, HathiTrust and *Google Books*, were saddled with high-profile lawsuits, both initiated by the Author's Guild, from the mid-2000s until recently.

By November, 2013, Google's long legal fight eased somewhat with the ruling in Google's favor by Judge Denny Chin. (Lee, 2013; Chant 2013) By 2014 appeals were also found in favor of the HathiTrust over the Authors Guild. (Chant 2014) By April 2016 Google's case had been resolved with the United States Second Circuit ruling in fall 2015 in favor of the tech giant, and the U.S.

Supreme Court declining to examine the case further. (Hurley, 2016)

Looking at it more objectively, though, Google's problems stemmed from the proprietary nature of their endeavor. If they were to abandon some of the profiteering or perhaps alter the purpose of their program, it might have been seen in a better light by the Authors Guild. Despite Google's controversial approach, however, the U.S. Supreme Court considered the project to be within the guidelines of fair use, especially regarding the ultimately transformative use of the source material. (Hurley)

HathiTrust, in contrast, weathered the copyright lawsuit storm quite well and remains poised to develop as the leader of MDLs in terms of copyright compliance and providing accessibility to the disabled and access to public domain works. The result of their lawsuit showed that the Fair Use defense remains a strong one as long as it is coupled with such factors as restricting access to copyrighted work and remaining non-profit in nature. Their digitized corpus was also ruled Fair Use because of the importance the *HathiTrust* placed upon providing works with greater accessibility for the disabled. The Fair Use defense also made it clear that the *scope* of a project, which the Author's Guild had criticized, is irrelevant. Fair Use holds, regardless of the number of books digitized.

Among other MDLs, *Europeana's* efforts in advocating for a strong and robust public domain make it a worthy cause to follow. (Fallon) It is hoped by the author that they continue to fight for strong and clear open access of cultural content disambiguated from the profit motives of the corporations that tend to own the lion's share of copyrighted material.

The Promises and Pitfalls of Diversity

Google Books' partners have included institutions in Spain, Germany, France, and one university in Japan. (Given the huge number of books, languages and cultures in the world, one must consider deeply

whether such limited partnerships are feasible and scalable. In order to have complete universal coverage, one would need to add partners from everywhere, including institutions devoted to sub-communities and sub-cultures within larger countries. But despite such ambitions to create a digital corpus of everything printed, it is likely not a realistic outcome.

Furthermore, the act of scanning itself causes distinct issues. As Thylstrup writes, once a decision is made to digitize content, scanning becomes a political act rife with “distinct subpolitical processes.” (2014) In addition, researchers have found that the contents of the Google Book corpus do not represent an unbiased sampling. Pechenick et al. argue that “the evolution of the corpus throughout the 1900s is increasingly dominated by scientific publications rather than popular works. We have shown that even the first data set specifically labeled as fiction appears to be saturated with medical literature.” (2015) Aside from the problem of misrepresenting common collections, or transforming the boundaries of traditional print forms, the selection of texts appears to be skewed in favor of scientific titles. Weiss and James find missing public domain books in the Google corpus and a low representation of Spanish and Japanese language texts. (2013b, 2015) The diversity of content in the *Google Books* project as well as any of the other MDLs, therefore, remains at the mercy of the partnerships they have created.

For the *HathiTrust*, similar issues exist. The number of partners is large, but still not comprehensive. According to their online statistics, the percentage of English language books remains at about 50%. Of the top ten languages, 60% are Indo-European, with one, Latin, being a dead language. Two are East-Asian Languages (Chinese / Japanese) and one is a Slavic language (Russian). (HathiTrust, 2016) By world population and the number of cultural artifact created by various cultures – especially those that are older than the United States by millennia – the collections are disproportionately skewed toward the English language. Again, like the *Google Books* project

-- and in part because many of the *HathiTrust* partners as well as digitized books are shared with *Google Books*--the issue of the source materials calls into question the ability of MDLs to create diverse collections.

For Public Domain works, this proportion skews even further to English - about 60 percent of titles in the public domain are English with 40 percent left for any other language. Generally, this reflects the needs of the faculty and students at the academic institutions currently partnering with the *HathiTrust*. It is quite telling that the second and third largest collections of works are German and French, while Spanish, the second most-spoken language in the United States, trails at a distant 4th. Only 4.48% of the HathiTrust online collection is in Spanish, yet 36 million people (roughly 12% of the population) in the United States speak Spanish as their primary language. In contrast, only about 1.3 million (or.41%) speak French and only about 1.1 million (about.35%) speak German, yet the HathiTrust collections include 7% and 9.2% of these languages respectively. (US Census Bureau, 2011)

It is clear, then, that subject matter diversity could be better addressed by even the MDL exemplars. It is hoped that as these digital libraries develop over time, the gaps found in their coverage will adapt to the new, at-large audiences that exist beyond the physical boundaries of the locations they represent.

FUTURE RESEARCH DIRECTIONS: IMPACT OF THE DIGITIZED BOOK CORPUS

Humanities and the Ngram Viewer

One important impact that MDLs can have on current scholarship is the mining of the immense data sets generated from the mass digitization process. The Ngram viewer, designed by John Orwant and Will Brockman at Google, provides data visualizations for the frequency of queried

terms. This is a helpful tool for not only historians, writers, artists, and social scientists, but also libraries. (Google N-gram, 2016) Michel, et al. (2011) have shown the impact that research in this area, which they have called “Culturonomics,” can have on scholarship and textual analysis. Additionally, one can track historical events in the news or the ideas and concepts that accompany them. It is truly a useful tool for those interested in the history of ideas and how events shape them.

In the example shown in the figure above, a search involving the terms <Hiroshima> (in blue) and <Nagasaki> (in red) was conducted by the author. The search results show the frequency with which the words appeared in the English language from the period of time from 1800 to 2000. For the first 140 years, Nagasaki was the more mentioned of the two cities, likely a result of it being the only port open to foreign trade during the closed-country *sakoku* period in Japan. During the 1850s-1860s, Nagasaki also began to surge, likely a result of the United States’ aggressive policies to “open” up the country to international trade. Again a surge occurred between 1900 and 1910, reflecting its status as a major hub for coal

production and the Japanese Navy’s winter port for the Russian Asiatic fleet.

Hiroshima during this same period of time is not mentioned very much in comparison, trailing at a very distant second. However, during the years after the bombing of Pearl Harbor, Hiroshima becomes ever more frequently mentioned. Only by 1944, a year prior to the atomic bombings of both cities, do Hiroshima and Nagasaki reach parity in terms of frequency. After 1945 until the present day as a result of being the first city to suffer an atomic bomb attack, Hiroshima has been mentioned the most.

It is a useful tool for students to get a better sense of the context of historical events and an understanding of how notoriety waxes and wanes over time. It becomes clear then that MDLs can offer more than just digital versions of books and get at assisting not only scholars but also interested lay-persons, librarians, and students. The ability to search for concepts across a large sample of materials will revolutionize the humanities, the history of science, and even science itself, as people begin to see the patterns in knowledge development and information retrieval.

Figure 1. Ngram viewer showing frequency of the names Hiroshima (blue) and Nagasaki (red) from 1800 to 2000. Note the significant change starting in the early 1940s and spiking post-WWII



CONCLUSION

The Massive Digital Libraries covered in this chapter (*Google Books*, *HathiTrust*, *Internet Archive*, *Europeana*, *Gallica*) show striking individuality in their approaches to mass-digitization and the mass-distribution of digitized content. Yet many of their defining characteristics—from size to content aggregation, to development of consortiums and to content diversity and general accessibility—remain nonetheless similar. Their central missions to share the content of print books that may be locked away in print on physical book shelves are also admirable. Varying degrees of scope, subject matter and audience involvement shape these massive collections in individual ways, making each MDL unique.

Overall when looking at the main examples of MDLs currently in existence, one gets the sense that ambitions for a “unified digital front” for all types of materials is the impetus driving their development. This ambition impacts brick-and-mortar libraries in that they may have to scramble to accommodate as well as take advantage of these growing initiatives.

The issues raised by MDLs speak directly to the issues at the heart of mass-digitization culture. Veracity, accuracy and diversity get at the essential quality of any information society: trustworthiness. If one is able to trust the mass aggregated collections, then MDLs will provide an essential service to the pluralistic and digitally democratic societies from which they have sprung. Additionally, the issues of copyright are providing a lesson to all on Fair Use and the development of new open access/open source culture dependent upon the alteration and recombination of previously created works. The new tools and applications of MDLs, such as the Ngram Viewer, despite some of the issues of representation existing at the heart of the Google Books corpus, promise to surpass the individual books in the MDLs’ corpuses. MDLs will ultimately provide users with enhanced approaches to existing disciplines as well as the promise of new transformations.

REFERENCES

- California Digital Library (CDL). (2016). *Homepage*. Retrieved 9/15/16 from: <http://www.cdlib.org/>
- Chant, I. (2013). Lawsuit against Google Books dismissed. *Library Journal*, 138(20), 14.
- Chant, I. (2014). Academic: Appeals court upholds HathiTrust verdict. *Library Journal*, 590(12), 15.
- Digital Public Library of America (DPLA). (2014). *DPLA Strategic Plan: 2015 through 2017*. Retrieved 9/15/16 from: https://dp.la/info/wp-content/uploads/2015/01/DPLA-StrategicPlan_2015-2017-Jan7.pdf
- Duguid, P. (2007). Inheritance and loss? A brief survey of Google Books. *First Monday*, 12. doi:10.5210/fm.v12i8.1972
- Fallon, J. (n.d.). Europeana and the Public Domain. *Europeana Professional*. Retrieved from <http://pro.europeana.eu/public-domain-content>
- Google. (2013). Google Ngram Viewer. *Google Books*. Retrieved from <https://books.google.com/ngrams>
- HathiTrust. (2016). *HathiTrust Languages*. Retrieved 7/18/16 from: http://www.hathitrust.org/visualizations_languages
- Hurley, L. (2016). *Reuters. Supreme Court rejects challenge to Google book-scanning project*. Retrieved from <http://www.reuters.com/article/us-usa-court-google-idUSKCN0XF1O3>
- Jackson, J. (2010, August 6). Google: 129 Million Different Books Have Been Published. *PC World*. Retrieved from IDG News Service: http://www.pcworld.com/article/202803/google_129_million_different_books_have_been_published.html
- James, R. (2010). An assessment of the legibility of Google Books. *Journal of Access Services*, 7(4), 223–228. doi:10.1080/15367967.2010.503486

James, R., & Weiss, A. (2012). An assessment of Google Books Metadata. *Journal of Library Metadata*, 12(Iss: 1), 15–22. doi:10.1080/19386389.2012.652566

Jeanneney, J. (2007). *Google and the myth of universal knowledge: a view from Europe*. Chicago: University of Chicago Press.

Lee, T. (2013, November 14). Google Books ruling is a huge victory for online innovation. *Washington Post*. Retrieved from <http://www.washingtonpost.com/blogs/the-switch/wp/2013/11/14/google-books-ruling-is-a-huge-victory-for-online-innovation/>

Michel, J., Shen, Y., Aiden, A., Veres, A., Gray, M., Pickett, J. P., & Aiden, E. L. et al. (2011). Quantitative analysis of culture using millions of digitized books. *Science*, 331(6014), 176–182. doi:10.1126/science.1199644 PMID:21163965

Nunberg, G. (2009a). Google Books: The Metadata Mess. *Google Book Settlement Conference*. UC Berkeley. Retrieved from <http://people.ischool.berkeley.edu/~nunberg/GBook/GoogleBookMetadataSh.pdf>

Nunberg, G. (2009b). Google's Books search: A disaster for scholars. *Chronicle of Higher Education*. Retrieved from <http://chronicle.com/article/Googles-Book-Search-A/48245/>

OCLC adds Google Books, Hathitrust record to WorldCat. (2010). *Advanced Technology Libraries*, 39(5), 4.

Pechenick, E. A., Danforth, C. M., & Dodds, P. S. (2015). Characterizing the Google Books Corpus: Strong Limits to Inferences of Socio-Cultural and Linguistic Evolution. *PLoS ONE*, 10(10), e0137041. doi:10.1371/journal.pone.0137041 PMID:26445406

Sato, Y. (2010). Keio gijuku daigaku ni okeru Google-Library-Project no chosakuken chosa nitsuite. *Media Net*, 17, 51-53. Retrieved 9/15/16 from: <http://www.lib.keio.ac.jp/publication/medianet/article/017/01700500.html>

Smith, E. (2013, November 15). Google Books scores big legal victory, authors [sic] guild plans appeal. *Associations Now*. Retrieved from <http://associationsnow.com/2013/11/google-books-scores-big-legal-victory-authors-guild-plans-appeal/>

Texas Digital Library (TDL). (2016). *Home page*. Retrieved 9/15/16 from: <https://www.tdl.org/>

Thylstrup, N. B. (2014). *The Politics of Mass Digitization*. Københavns Universitet, Det Humanistiske Fakultet.

U.S. Census Bureau. (2011). *Language Use*. U.S. Census Bureau. Retrieved from <http://www.census.gov/hhes/socdemo/language/>

Venkatraman, A. (2009). Europe must seize lead on digital books, urges reading. *Information World Review*, (260), 1-1. Retrieved from <http://search.proquest.com/docview/199357636?accountid=7285>

VLDL. (2011). Home page. *Fourth Workshop on Very Large Digital Libraries*. Retrieved from <http://www.delos.info/vldl2011/>

Weiss, A. (2016). Examining Massive Digital Libraries (MDLs) and their impact on reference services. *The Reference Librarian*, 57(4), 286–306. doi:10.1080/02763877.2016.1145614

Weiss, A., & James, R. (2013a). Assessing the coverage of Hawaiian and Pacific books in the Google Books digitization project. OCLC Systems & Services. *International Digital Library Perspectives*, 29(1), 13–21. doi:10.1108/10650751311294519

Weiss, A., & James, R. (2013b). An examination of Massive Digital Libraries' coverage of Spanish language materials: Issues of multi-lingual accessibility in a decentralized, mass-digitized world. Kyoto, Japan. *2013 International Conference on Culture and Computing*, 10–14. doi:10.1109/CultureComputing.2013.10



Weiss, A., & James, R. (2014). *Using Massive digital libraries: A LITA Guide*. Chicago, IL: Neal-Schuman Publishing.

Weiss, A., & James, R. (2015). Comparing the access to and legibility of Japanese language texts in Massive Digital Libraries. *2015 International Conference on Culture and Computing*. doi:10.1109/Culture.and.Computing.2015.51

KEY TERMS AND DEFINITIONS

Digital Corpus: The full-text of millions of books digitized by the MDLs. These provide opportunities for scholars to examine the frequency of terms as they appear in the print corpus.

Digital Humanities (DH): A branch of the humanities incorporating digital search, digitized texts, encoding, and other strategies to move previously print-bound materials and scholarship into computer science related analysis.

Fair Use: Defined by some as the breathing room to allow for the freedom of expression, it is a strong limitation of the extent of copyright law. It is determined by examining four factors: 1) the purpose and character of the use; 2) the nature of the copyrighted work; 3) the amount

and substantiality of the portion taken; and 4) the effect of the use upon the potential market.

Mass-Digitization: The practice of quickly and thoroughly digitizing items on a large scale. In the case of MDLs these efforts involve scanning millions of books with multiple institutional partners across national boundaries.

Massive Digital Libraries (MDLs): Term adopted to describe the mass-digitization of printed books and the mass-aggregation of their metadata into online, full-text-searchable digital collections; some component of open access and use of public domain works defines their collections; diffuse and diverse target end-user groups are also characteristic of MDLs.

Ngram Viewer: A tool developed by Google to aid in the visualization of the digital corpus of books. It plots frequencies of terms across a graph, helping users understand how common a word or concept was in the corpus at a specific moment in history.

Public Domain: Works that are no longer under copyright protection. In the United States these works tend to be those published before January 1, 1923 and unpublished works created prior to 1893. U.S. Government publications are public domain.

Mission, Tools, and Ongoing Developments in the So.Re.Com. “A.S. de Rosa” @-library



Annamaria Silvana de Rosa
Sapienza University of Rome, Italy

INTRODUCTION

Knowledge Domain

Originated from Serge Moscovici's¹ *Opera Prima* (1961), Social Representations theory (S.R.), originally European, is currently a multilingual worldwide field with a substantial body of literature in social psychology and other social science. Studies of “what” people know – and “how” this lay people knowledge relates to the social groups to which they belong and are shaped by the polyphonic traditional and new media system – are concerned with the social construction of everyday knowledge and the related socially situated practices. The “objects” studied have a strong societal impact in the social spheres.

This *supra-disciplinary field* is characterized by a *great consistency in terms of epistemological and theoretical inspiration* and a rich diversity both:

- From the *paradigmatic point of view* (“structural,” “social positioning,” “monographic,” “anthropological,” “dialogical,” “modelling” approaches);
- In terms of *methodological research designs* (qualitative, monographic, anthropological, experimental, descriptive, structural, visual-figurative, multi-methodological, etc.);
- From the *thematic point of view*
- With respect to the *applied contexts and domains of expert and lay knowledge*

production and transmission: education, health, economics, environment, tourism, politics, organizations, media industry, etc.

The goal of this contribution is to present background, mission, ongoing developments and future research directions of the *So.Re.Com. “A.S. de Rosa” @-library: a digital platform integrating scientific documentation, networking and training purposes* in the field of *Social Representations and Communication* (So.Re.Com.).

BACKGROUND

Born from a personal bibliographic inventory, initially including almost 500 references in the field of *Social Representations and Communication*, the *So.Re.Com. “A.S. de Rosa” @-library* (de Rosa, 2015c) has been transferred into an ad-hoc designed relational database implemented in 1996 on the *European/International Joint PhD in Social Representations and Communication* website (de Rosa, 2000, 2001, 2004 a, 2004 b, 2006; 2008, 2009 a, 2009 b, 2010 a, 2010 b, 2010 c; 2011, 2014, 2015a, 2015b, 2015c, 2016a; de Rosa & Picone, 2007).

It is built on an on-going multi-year project aimed at an *empirical meta-theoretical analysis of the whole literature on Social Representations* (de Rosa, 2013a, 2013b, 2016b; de Rosa, Dryjanska, Bocci, 2017a, 2017b). Guided by the main goal of *evaluating the impact of the scientific production driven by the Social Representations theory in the*

social arena faced with social demand, the aim is to take stock of the scientific field developed in more than 50 years by conducting an empirical meta-theoretical analysis of the literature on Social Representations, mapping the development of different paradigms, the related research methods, the thematic areas and their impact on the various applied fields within the multi-generational community of scientists and across different geo-cultural contexts. This research program is led by de Rosa, creator and program director of both the *European/International Joint PhD in S.R. & C.* (<http://www.europhd.eu>) and the *So.Re.Com. THEmatic NETwork* (<http://www.europhd.eu/SoReComTHEmaticNETwork>), a *network of networks* that promotes co-operation between academic, professional research and commercial institutions and facilitates the dissemination of scientific results in the field of Social Representations (see Figures 1 and 2).

Located at Sapienza University of Rome Sapienza - the largest university in Europe with a patrimony of 155 libraries and a recent investment in digitalising contents² - the state-of-the-art of the *Social Representations and Communication Research Centre and Multimedia Lab* includes the

integrated specialized physical and *So.Re.Com. “A.S. de Rosa” @-Library*. It is the nerve centre of the *European/International Joint PhD programme in Social Representations and Communication*, the training structure of the wider *So.Re.Com. THEmatic NETwork*, dedicated to research and training needs for both doctoral candidates and highly experienced researchers, individually or in cross-national research teams.

Recently (2012) transferred from its historical location opposite Campidoglio to another larger location in the very heart of Rome at the top of Aventino hill, overlooking the capital’ archaeological-architectonic center, it guarantees technical support for training and management activities, including real-time interactive exchanges with partner institutions around the globe.

Full-time staff is available to provide assistance to users.

MISSION

A research environment “*based on*” and “*aimed at*” integrating *scientific documentation, networking and training purposes*.

Figure 1. The So.Re.Com. Joint Innovative Doctoral Program institutional network



Figure 2. The So.Re.Com. THEmatic NETwork at a glance on global scenario



The So.Re.Com. “A.S. de Rosa” @-library is a multipurpose digital environment of interrelated relational databases conceived in the logic of the semantic web, including multiple web interfaces aimed to integrate *documentation services*, with *networking* and *research training*.

Once the ongoing migration into an open access platform will be completed, the So.Re.Com. “A.S. de Rosa” @-library will answer the increasing demand for multi-lingual access to multi-format documentation in Social Representations and Communication, also from users in disadvantaged areas in the world where information and training are difficult to obtain. The ultimate goal is to make it accessible to anyone, anytime, anywhere, therefore disseminating resources otherwise dispersed and fragmented in several universities, public and private research centers and governmental institutions.

To accomplish these goals, its creator and project leader is coordinating a large team of human resources with different degree of expertise since many years, in order to produce, collect and meta-analyze scientific documentation published in the field all over the world, not exclusively by the So.Re.Com. Thematic Network members.

This @-library supports access to multiple repositories and web-tools, designed by A.S. de Rosa and implemented on the EuroPhD website, in order to integrate the three main pillars (Figures 3, 4, and 5).

Documentation

- A *comprehensive bibliographical repository* of the literature on Social Representations, currently including more than 10000 references. This growing collection of bibliographic entries includes not only the classic bibliographic details, but also the author's and co-authors Institution affiliation and Country/Continent, the bibliometric indexes (Impact Factor and SJR) and other useful information for geo-mapping the diffusion and impact of the theory over the world, by decades and thematic areas and for visualizing the intra and inter-institutional authors' collaborations by city, countries, continents.
- A *meta-theoretical analyzed repository* of the literature on SR, currently including more than 4000 meta-analyzed articles or

Figure 3. The So.Re.Com. “A.S. de Rosa” bibliographical repositories and @-Library



book chapters using the on-line grid for the “meta-theoretical analyses” and the guidelines developed by A.S. de Rosa (1994, last version updated January 2014).

- An *intelligent @-Library* holding a rich documentation of almost 10000 PDF texts, videos, audio-interviews and multi-media scientific and training materials in the field of S.R.
- Advanced search engines hyper-linked with all the three above-mentioned repositories (also using as ontologies the categories of the very detailed “meta-theoretical analysis” grid) and for the full text search on the PDFs.

Networking

- More than 3000 members of the worldwide *SoReCom Thematic Network* on-line community, including personal and institutional information and scientific profiles.
- News and calendar of scientific events.
- Online co-operative research tools.
- On-line web-conference management system for participant registration, submission and review of abstracts, power point presentations and papers, interactive feedback system for editing and publishing.
- SoReCom THEmatic NETwork @-NEWS.

- Newly designed management service for “*Virtual meetings with authors or book presentations*”.

Research Training

- Multi-point web-videoconference interactive system, also sharing applications and files.
- Developing research trainees’ skills in meta-theoretical analysis of S.R. literature and advanced use of bibliographic *repositories*.
- Distance tutoring and co-tutoring based on personalised access to research reports and on-line evaluation of trainees’ research reports at the initial, intermediate and final stages.
- Video-lectures in streaming and web-video-interviews with the protagonists in the scientific field.
- Web-tools allowing research trainees and professors to provide “ratings,” i.e. feedback on the virtual training activities and those involving physical mobility, etc.
- Dynamic timeline for research training activities management and credit accumulation sheet.
- Personalised inventory from the application to the degree and beyond.
- Alumni community and career development.

Figure 4. @-learning Kit at disposal of the So.Re.Com. European/International Joint Ph.D. research trainees



@-learning Kit

The complex package of multimedia tools to be developed and disseminated via Internet includes:

- a full bibliographic inventory
- a meta-analyzed inventory
- an advanced search engine
- a specialized e-library in SR & C.
- digitalized studio produced videos
- magisterial lectures produced in real time (videoconferences)
- on-line interviews with leading experts
- European Ph.D. web-auditorium
- on-line workshops among participants and academics
- a "virtual campus" SR & C scientific community online database.

© All contents and images are protected by the European/International Joint PhD in S.R. & C. copyright

Figure 5. Distance tutoring and co-tutoring tools based on personalised access to research reports and on-line evaluation of trainees' research reports



Quality Control and On-Line Monitoring System

Assessment tools used in the Open distance learning system are of fundamental assistance in continuously monitoring the quality of training.

The quality of the scientific and didactic content is evaluated by the Program Director in co-operation with the the tutors. This is accomplished using evaluative tools and co-tutoring, including distance co-tutoring that continuously monitors the student training via the European doctorate's dedicated web site.

After monitoring the work of previous years, among the other quality control procedures, the Core Executive Committee decided to impose a more structured "time constraint" on the development of early stage researcher's programs.

This offers tutors on-line evaluation tools for various versions of the research trainees' projects:

- ♣ Initial short version
- ♣ Intermediate extended version
- ♣ Final full report and short version

These tools have been installed in the web site for every registered research trainee <http://www.europhd.net>. They are considered confidential and, by using a password, are accessible for consultation only by the didactic triad of the 3 tutors (1 national + 2 foreign tutors), the Programme Director and the research trainee him or herself.

In addition, the procedures for presentation and evaluation of dissertations before the Final Jury have been organized, structured and appropriately tested in the course of the first dissertation defenses.

© All contents and images are protected by the European/International Joint Ph.D in S.R. & C. copyright

RESEARCH DIRECTIONS: THE FUTURE IS ONGOING!

Coherently with the planned developments (see de Rosa, 2015b) *from the content point of view*, we are pursuing the envisaged progress:

- An increased critical mass of content and related metadata of the *So.Re.Com. "A.S. de Rosa" @-library* by 2017 - thanks to the approval by the European Commission of the 2013-2017 contract *SoReComJoint-IDP* People-Innovative Doctoral Program positioned at the highest ranked (98,80/100 mark) among 1175 proposals. Built on the sound experience of both the *European/International Joint PhD in Social Representations and Communication*, approved by the EC since 1992, and the *So.Re.Com. THEmatic NETwork*, the *SoReComJoint-IDP* project is coordinated by Sapienza University of Rome and includes 13 universities, 2 private companies, 1 national research centers with cooperation with private sector in 8 European and 6 extra-European countries (Austria, Czech Republic, France, Italy, Romania, Spain, Sweden, Switzerland, Argentina, Brazil, Canada, China, Mexico, United States). This large and diversified network ensures a wide range of complementary paradigmatic, methodological and thematic options offered by international and cross-sectorial research teams within and outside European borders. The associate partners perform vital tasks such as co-tutoring activities, hosting Early Stage Researchers (ESR) for secondments and providing valuable training modules during the project's foreseen scientific events.
- The *So.Re.Com.Joint-IDP* will contribute to disseminating also the meta-theoretical knowledge produced by a new team of 13 ESR – recruited in 2014 - who will continue to work together until 2017, by mediated access to the best comprehen-

sive specialized @-library worldwide, on the on-going multi-faceted project led by prof. A.S. de Rosa (the *European PhD Lab* director) on top of previous multiyear work carried out with the long term purpose of writing a *Biography of the Social Representations Theory*. This unified research project framework has been articulated into 13 precise objectives concerning the specific research lines to be developed by each recruited ESR, matching their own interests and area of prior expertise. The ESR cooperate in meta-theoretical analyses and contribute to the development of the specialized *So.Re.Com. "A.S. de Rosa" @-library*. The integrated research framework provide an integrated training opportunity to maximize the cooperative and joint expertise of both the host institution and associated partners, doing so from the theoretical, methodological and applied perspective of the researches conducted in more than 50-year development and diffusion of this scientific field, with a view to monitoring and exploiting its added value to respond to social demand in the multiple contexts of the social arena: from economics to the environment, politics, health and community issues, science communication, etc. The need for developing both core and wider employment related skills is driven by the goal of making research training of greater relevance for a wider variety of careers both in and outside academia, internationally and globally more attractive.

- Driven by the synergic goal of an integrated development of the documentation tool "of" and "for" other researches (*So.Re.Com. "A.S. de Rosa" @-library*) and the training activities of the International doctoral program - among other multi-year series of activities, the *Winter Session International Lab Meeting* are specifically aimed at training in the *meta-theoretical analysis* of the specialized literature on



S.R.: why and how to map the development of different paradigms, the related research methods, the thematic areas, their impact on the various applied fields and the new emerging research directions within the multi-generational community of scientists and across different geo-cultural contexts (http://www.europhd.eu/html_onda02/07/28.01.00.00.shtml). The program includes:

- *Key lecture* by the research leader, aimed at presenting the current status of the Meta-Theoretical analysis of complete body of the S.R. literature.
 - *Presentations by So.Re.Com. Joint-IDP expert post-docs* and advanced doctoral trainees involving a *multi-generational community of senior and junior scientists* that ranges from visiting post-docs to Euro PhD Alumni (including former Marie Curie Fellows), already well trained in the *meta-theoretical analysis* of the specialized literature on Social Representations.
 - *Training the ESR* for developing advanced bibliographic on-line search skills via access to other digital libraries, aimed at developing the large specialized full bibliographic *repository* and the *So.Re.Com. "A.S. de Rosa" @-library*.
 - *Presenting guidelines* for integrating bibliographic resources in the physical and digital SoReCom specialized libraries.
 - *Training the ESR in the meta-theoretical analysis grid* and tools of the meta-theoretical analysis *repository* including the following tasks and methodology. Following guidelines and procedures given by the project leader prof. A.S. de Rosa and a specific training, the specific ESR's contribution consist:
- To detect and retrieve publications not yet inserted in the So.Re.Com. "A.S. de Rosa" @-Library, and meta-analyze the whole body of the literature on Social Representations produced according to the "specific individual research focus" over the world and not only in the geo-cultural context/continent of the host institution for mobility stage;
 - To fill new information and upload the related PDF files into the specialized So.Re.Com. "A.S.de Rosa" @-Library;
 - To carry out statistical analyses on the complete meta-analysed body (extracted from the SoReCom "A.S. de Rosa" @-Library after the integration of the new information);
 - To produce video-interviews – according to an outline designed by the project leader - with the key scientists who contribute to the development of the paradigmatic approaches to Social Representations;
 - To fully describe, comment and visualize the results using de Rosa's Guidelines for the geo-cultural mapping in the PhD thesis in English (in the book format and short article).
- Illustration of the *double quality control system*.
 - *Co-ordination and planning session* for the organization of weekly individual and group ESR small meetings on the transversal common tools of the overall research program and the specific goals of each of the research focuses.

- The complementary training stages at 13 host institutions in Europe (Austria, Czech Republic, France, Romania, Spain, Switzerland), in North and Latino America (Canada, Argentina, Brazil, Mexico) and in China, from September 1 to November 30 2015, has also offered the ESR opportunities for networking with the associated partners and for further integrating virtual and physical international mobility.
- The supra-disciplinary, multi-methodological and thematic possibilities offered by the training network are enhanced the wide range of world-class visiting researchers invited to participate in the program’s intensive training “stages,” multi-point videoconferences, tutoring and co-tutoring for individuals and small groups, and various on-line training tools.

Also the planned developments *from the technical point of view*, are currently ongoing as envisaged:

1. Ongoing transfer into a new web-platform from *web-object* to *Drupal*, expanding access via new web interfaces (see Figures 6 and 7).
2. To implement a management service innovation enabling users to *virtually meet authors of the @-Library’s content by jointly coordinated on demand meetings* between early stage researchers and famous authors belonging to the SoReCom scientific community.
3. To pursue future integration in the larger Sapienza Digital Library and in the European Digital Library, improving accessibility, use and dissemination, while keeping our autonomy in the content management and quality control system to update data and meta-data.

Figure 6. The new home page of the European/International Joint-PhD in Social Representations and Communication



Figure 7. The new web interface of the Basic Search for the bibliographical repository of the So.Re.Com. “A.S. de Rosa” @-library

Therefore the developments of the *So.Re.Com. “A.S. de Rosa” @-library* from the content and technical point of views – supported by the approval of the SoReCom IDP People project – has facilitated by and will contribute to improve interaction among the three pillars of the *European/International Joint PhD in Social Representations and Communication Research Centre and Multimedia Lab*: documentation, training activities and networking.

CONCLUSION

The unique combination of a *European/International Joint doctoral program and a scientific network including academic and extra-academic partners* – once again approved by the European Commission as the top project 2013 People Innovative Doctoral Programs (*So.Re.Com.Joint-IDP*) – is an integrated physical and virtual campus

where initial research trainees and experts from academia and the public and private sectors can come together to learn, share, create and disseminate knowledge. We offer a progressive system for training early stage researchers through active research. Research training takes place in international environments, both academic and non-academic, with specifically designed intensive didactic “stages” in multilingual and multicultural settings. This includes access to the specialized physical and @-library in Social Representations (the *So.Re.Com. “A.S. de Rosa” @-library*), meta-theoretical analysis of the literature on Social Representations and Communication, tutoring and co-tutoring for research projects, active participation in International Lab meetings and International Summer Schools, advanced training in multi-methodological qualitative and quantitative design, software for statistical data analysis and results visualization, e-learning activities via interactive video-conferences, training in comple-

mentary skills, scientific networking, and active participation in international conferences, etc.

This integrated face-to-face and open virtual space, accessible from anywhere in Europe or the world, not only meet the needs of its users by providing easy access to publications and meta-data and many other kinds of digital multi-format documentation (printed, pictures, audio-videos), but also enables them to quickly become content providers and co-producers. The multigenerational nature of this interactive knowledge community guarantees cross-fertilization of ideas between senior scientists and early research trainees, as well as between experts from within and outside of academic circles, including policy makers interested in research on societal issues and their relevance in social contexts. Therefore, open access to the integrated physical and virtual *SoReCom THEmatic NETWORK* infrastructure will allow the members of this knowledge community to continue to grow. The newly designed web-services for enlarging interactive facilities of the *So.Re.Com. "A.S. de Rosa" @-library* not only gives research trainees access to documentation, but in the physical and virtual space of this infrastructure also they can also meet authors and each other to discuss new publications and critical research areas.

In conclusion the *So.Re.Com. "A.S. de Rosa" @-library* provides *an integrated digital tool for scientific documentation, networking and training purposes in the supra-disciplinary field of Social Representations and Communication*, thus stimulating a rich and ever-growing research and learning environment that is advantageous to research team members, trainees and experts alike.

Comparative analyses - aimed at answering to research questions related to the worldwide dissemination of the Social Representations and its positioning compared other theories and approaches of social sciences – can be based on almost “big data” and “meta-data”, filed in our *SoReCom "A.S. de Rosa" @-library* repositories. In fact this is the most comprehensive library specialised in Social Representation, compared to the universe of the total literature of this scientific

field. Therefore we may define our data sources as almost “big data”, if we assume that “There is no one definition of Big Data. Thought about in simple terms, Big Data involves datasets that are far larger than those traditionally examined in journals like this one. Yet there has always been considerable variation in the size of datasets, ranging from small experimental studies to large samples involving census or polling data. Size alone is therefore an insufficient descriptor” (Parks, 2014, 255).

Just to make an example, in order to examine the bibliometric impact of the literature inspired by the Social Representation theory compared to its global dissemination, we have used data concerning authors’ countries and institutional affiliations, the bibliometric indexes (Impact Factor and SJR), the language of the publications, the thematic choice, the networks of inter-institutional collaboration within and between countries and continents (in order to detect who is working with whom on what and in which country...), etc. The implications of the “*impact of the impact*” has been discussed in other publications in the light of the critical debate, which still animates the community of scientists, stimulating meta-reflexive discussion and view exchanges among the members of our scientific community on the preferable publishing options and collaborative strategies in the current editorial and academic scenario (de Rosa, A.S. 2015c, de Rosa, A.S. Dryjanska, L. Bocci, E. 2015, in press). These publications also illustrate an example of the prototypical use of the *SoReCom "A.S. de Rosa" @-library* as a *research tool*, rather than simply an *online repository for scientific documentation*, so accomplishing one of its mission.

REFERENCES

de Rosa, A. S. (2000). Distance Training of European doctoral training students in meta-theoretical analysis of fully-researched bibliographic corpus. *Proceedings of the Wanderstudent*, 95–98.



de Rosa, A. S. (2001). Sistema di co-operazione e formazione a distanza implementato sul sito web dell'European doctoral programme on Social Representations and Communication: verso l'attivazione di una "intelligent virtual library". In M. O. Belardinelli (Ed.), *Valentini Day* (pp. 99–109). Roma: Kappa.

de Rosa, A. S. (2004a). An idea that became an Institution: the European Doctorate on Social Representations and Communication. *Marie Curie Fellowship Association Annals*. Retrieved on April 10, 2009 from <http://www.mariecurie.org/annals/index.html>

de Rosa, A. S. (2004b). Institutional recognition and didactic & training accreditation within three different scenario's (E.U., Ministries, Universities): The case of the European PhD on Social Representations and Communication. In J. Van den Branden, C. Moreau, A.S de Rosa, A. Lapal-lainen, & A. Opsomer (Eds.), *C-EVU Final report contribution WG4 ACCREDITATION*. Retrieved from <http://www.cevu.org>

de Rosa, A. S. (2006). Social Representation and Communication Thematic Network: A case study for monitoring the development of a scientific community. In I. Labhrainn, C. McDonald Legg, D. Schneckenberg, & J. Wildt (Eds.), *The Challenge of eCompetence in Academic Staff Development* (pp. 49-57). Galway: CELT, NUI.

de Rosa, A. S. (2008). New forms of international cooperation in doctoral training: internationalization and international doctorate: One goal, two distinct models. *Higher Education in Europe*, 33(1), 3–25. doi:10.1080/03797720802228084

de Rosa, A. S. (2009a). Promoting Joint European/International Doctorates in a global scenario: Opening the European PhD on Social Representations and Communication to the World. In M. Gaebel, L. Purser, B. Wächter, & L. Wilson (Eds.), *Internationalisation of European Higher Education* (pp. 1–35). Stuttgart: RAABE; Retrieved from http://www.euophd.eu/html/doc/eua_aca_hndbk_2009_derosa.pdf

de Rosa, A. S. (2009b). The So.Re.Com. Euro-PhD World: Combining Partnerships with Innovation. In A. Klucznik-Tòrò, A. Csépe, & D. Kwiatkowska-Ciotucha (Eds.), *Higher Education Partnership, Innovation* (pp. 203–212). Budapest: Publikon Publisher/IDResearch Ltd.

de Rosa, A. S. (2010a). The Joint European/International Doctorate on Social Representations and Communication: an experience anticipating the Bologna process. *Proceedings of the 2nd International Seminar on Joint Degrees*. Retrieved on April, 202010 from <https://www.joiman.eu/activities/Lists/DevelopmentActivities/DispForm.aspx?ID=16&Source=https%3A%2F%2Fwww%2Ejoiman%2Eeu%2Factivities%2Fdefault%2Easpx>

de Rosa, A. S. (2010b). Internationalisation of collaborative doctorates and Joint International Doctorates: requirements and guidelines for each distinct model. *Higher Education Partnership and Innovation*, 8, 219-242. Retrieved from http://www.euophd.eu/html/doc/derosa_ihepi2010.pdf

de Rosa, A. S. (2010c). The Joint European/International Doctorate: a strategic tool for enhancing the worldwide institutional collaboration for research training. *IIE Networker International Education in Europe*, 28-29.

de Rosa, A. S. (2011). De Nouvelles Formes de Coopération internationale dans la formation doctorale: Internationalisation et doctorat international: un objectif, deux modèles distincts. In V. Baidenko (Ed.), *The Bologna Process: resume of the decade* (pp. 357–363). Moscow: Academic Press.

de Rosa, A. S. (2013a). Taking stock: a theory with more than half a century of history. In A. S. de Rosa (Ed.), *Social Representations in the "Social Arena"* (pp. 1–63). New York: Routledge.

de Rosa, A. S. (2013b). Research fields in social representations: snapshot views from a meta-theoretical analysis. In A. S. de Rosa (Ed.), *Social Representations in the "Social Arena"* (pp. 89–124). New York: Routledge.

de Rosa, A. S. (2014) *The So.Re.Com. “A.S. de Rosa” @-Library: A Multi-Purpose Web-Platform in the supra-disciplinary field of Social Representations and Communication*. Valencia: INTED2014 Publications. Retrieved from http://library.iated.org/?search_text=publication%3AINTED2014&adv_title=&rpp=25&adv_authors=&adv_keywords=&orderby=page&refined_text=de+Rosa

de Rosa, A. S. (2015a) The European/International Joint Phd In Social Representations and Communication: a Pioneer Case of Triple “I” (International, Interdisciplinary, Intersectoral) Networked Joint Doctorate. In *INTED 2015 Proceedings* (pp. 5449-5460). Madrid: INTED Publications. Retrieved from <http://library.iated.org/publications/INTED2015>

de Rosa, A. S. (2015b). The *So.Re.Com. “A.S. de Rosa” @-library*: a digital tool for integrating scientific documentation, networking and training purposes in the supra-disciplinary field of Social Representations and Communication. In M. Khosrow-Pour (Ed.), *Encyclopedia of Information Science and Technology* (pp. 4938–4949). Hershey, PA: IGI Global; doi:10.4018/978-1-4666-5888-2.ch487

de Rosa, A. S. (2015c) The Use of Big-Data and Meta-Data from the So.Re.Com A.S. de Rosa @-Library for Geo-Mapping the Social Representation Theory’s Diffusion over the World and its Bibliometric Impact. In *INTED 2015 Proceedings* (pp. 5410-5425). Madrid: INTED Publications. <http://library.iated.org/publications/INTED2015>

de Rosa, A. S. (2016a). The European/International Joint PhD in Social Representations and Communication: a triple “I” networked joint doctorate. In *Book of Proceedings of the 2nd International Conference on Development in Doctoral Education and Training* (Oxford, UK, 30th -31st March 2015) (D. Halliday & G. Clarke Eds.), pp. 47-60. London: Epigeum Oxford University Press.

de Rosa, A. S. (2016b). Mise en réseau scientifique et cartographie de la dissémination de la théorie des représentations sociales et son impact à l’ère de la culture bibliométrique. In G. Lo Monaco, S. Delouvée, & P. Rateaux (Eds.), *Les représentations sociales* (pp. 51–68). Editions de Boeck, Belgique.

de Rosa, A.S., Dryjanska, L., & Bocci, E. (2015). *Chi lavora con chi, su che cosa e dove? Network analisi e Geo-Mapping delle collaborazioni inter-istituzionali tra ricercatori in diversi paesi e continenti per una epidemiologia della produzione scientifica sulle rappresentazioni sociali nel mondo*. Simposio a cura di Ida Galli, Le rappresentazioni del sociale, Congresso Nazionale A.I.P. della Sezione di Psicologia Sociale.

de Rosa, A. S., Dryjanska, L., & Bocci, E. (2017a). Profiling authors based on their participation in Academic Social Networks INTED2017, (Valencia, SPAIN, 6-8th of March, 2017). In *Inted 2017 Proceedings*, (pp. 1061-1072). Madrid: INTED Publications.

de Rosa, A. S., Dryjanska, L., & Bocci, E. (2017b). The impact of the impact: Meta-Data Mining from the SoReCom “A.S. de Rosa” @ Library. In M. Khosrow-Pour (Ed.), *Encyclopedia of Information Science and Technology* (4th ed.). IGI Global.

de Rosa, A. S., & Picone, M. (2007). The European Ph.D. on Social Representations and Communication: Integrating Virtual and Physical Mobility via the European Ph.D. Web-Auditorium. In A. Boonen & W. Van Petegem (Eds.), *European networking and learning for the future* (pp. 49–57). Antwerp, Belgium: Garant.

Di Iorio, A., Schaerf, M., Guercio, M., Ortolani, S., & Bertazzo, M. (2013). *A digital infrastructure for trustworthiness*. Paper presented at the IX IRCDL, Italian Research Conference on Digital Libraries. Retrieved on August 10, 2013 from <http://www.dis.uniroma1.it/-ircdl13/?q=node/43>

Magarotto, A. (2013). *Sapienza libraries and project Google Books*. Paper presented at the IX IRCDL, Italian Research Conference on Digital Libraries. Retrieved on August 10, 2013 from <http://www.dis.uniroma1.it/-ircdl13/?q=node/43>

Parks, M. R. (2014). Big Data in Communication Research: Its Contents and Discontents. *Journal of Communication*, 64(2), 355–360. doi:10.1111/jcom.12090

ADDITIONAL READING

de Rosa, A. S. (1992). Thematic perspectives and epistemic principles in developmental Social Cognition and Social Representation. The meaning of a developmental approach to the investigation of. S.R. In M. von Cranach, W. Doise, & G. Mugny (Eds.), *Social representations and the social bases of knowledge* (pp. 120–143). Lewiston, N.Y.: Hogrofe & Huber Publishers.

de Rosa, A. S. (1994). From theory to meta-theory in S.R.: The lines of argument of a theoretical-methodological debate. *Social Sciences Information. Information Sur les Sciences Sociales*, 33(2), 273–304. doi:10.1177/053901894033002008

de Rosa, A. S. (2002). Le besoin d’une “théorie de la méthode”. In C. Garnier (Ed.), *Les formes de la pensée sociale* (pp. 151–187). Paris: P.U.F.

de Rosa, A. S. (2006). The boomerang effect of radicalism in Discursive Psychology: A critical overview of the controversy with the Social Representations Theory. *Journal for the Theory of Social Behaviour*, 36(2), 161–201. doi:10.1111/j.1468-5914.2006.00302.x

de Rosa, A. S. (2006). Por qué es importante?” Notas inspiradas en una mirada reflexiva a la teoría de las representaciones sociales. In S. Valencia Abundiz (Ed.), *Representaciones sociales. Alteridad, epistemología y movimientos sociales* (pp. 79–173). México: Universidad de Guadalajara, Centro Universitario de Ciencias de la Salud.

de Rosa, A. S. (Ed.). (2008). Special Issue “Looking at the History of Social Psychology and Social Representations: Snapshot views from two sides of the Atlantic,”. *Rassegna di Psicologia*, 2, 217.

de Rosa, A. S. (2010). Mythe, science et représentations sociales, In D. Jodelet and E. Coelho Paredes, (Eds.), *Pensée mytique et représentations sociales*, (pp. 85-124) Paris: L’Harmattan.

de Rosa, A. S. (2011) 1961-1976: a meta-theoretical analysis of the two editions of the “Psychanalyse, son image et son public,” in C. Howarth, N. Kalampalikis, P. Castro (Eds.) A half century of social representations: discussion on some recommended papers, *Special Issue, Papers on Social Representations*, 20 (2), retrieved September 15, 2013 from: http://www.psych.lse.ac.uk/psr/PSR2011/PSR2011_Issue_2.htm

de Rosa, A. S. (2011) 50 anos depois: a ‘Psychanalyse, son image et son public’ na era do Facebook. In A. Almeida, Z. Trindade, F. Santos (Eds.) *Teoria das Representações sociais - 50 anos*, (pp. 491-561) Technopolitik e Centro Moscovici, Brasília: MEMORES/UERJ, REPSO/UFES, LABINT/UFPE.

de Rosa, A. S. (2012). La psicoanalisi, la sua immagine, il suo pubblico: 1961-2011. Compire 50 anni nell’era dei social networks. In I. Galli (Ed.), *Cinquant’anni di Rappresentazioni sociali. Bilanci e prospettive di una Teoria in continuo divenire* (pp. 59–101). Milano: Edizioni Unicopli.

de Rosa, A. S. (Ed.). (2013). *Social Representations in the “Social Arena”*. New York, London: Routledge.

de Rosa, A. S., & d’Ambrosio, M. (2003). An empirical example of the comparison between multiple correspondence analysis and space analysis: The diffusion of the social representations theory through the institutional context of scientific communication. In S. Levy & D. Elizur (Eds.), *Facet Theory. Towards Cumulative Social Science* (pp. 73–86). Ljubljana: Faculty of Arts, Center for Educational Development.



de Rosa, A. S., & Farr, R. (2001). Icon and symbol: Two sides of the coin in the Investigation of Social Representations. In F. Buschini & N. Kalampalikis (Eds.), *Penser la vie, le social, la nature. Mélanges en hommage à Serge Moscovici* (pp. 237–256). Paris: Les Editions de la Maison des Sciences de l'Homme.

Deaux, K., & Philogène, G. (Eds.), *Representations of the Social*. Oxford: Blackwell.

Farr, R. M., & Moscovici, S. (Eds.). (1984). *Social Representations*. Cambridge, UK: Cambridge University Press.

Howarth, C., Kalampalikis, N., & Castro, P. (Eds.). A half century of social representations: discussion on some recommended papers, *Special Issue, Papers on Social Representations*, 20 (2), retrieved September 15, 2013 from: http://www.psych.lse.ac.uk/psr/PSR2011/PSR2011_Issue_2.htm

Jodelet, D. (Ed.). (1989). *Les représentations sociales*. Paris: Presses Universitaires de France.

Jodelet, D. (2009). Rappresentazioni e scienze sociali: incontri e rapporti reciproci. In A. Palmonari & F. Emiliani (Eds.), *Paradigmi delle Rappresentazioni Sociali* (pp. 253–282). Bologna: Il Mulino.

Laurens, S., & Roussiau, N. (Eds.). *La mémoire sociale. Identités et Représentations Sociales* (pp. 119–137). Rennes: Les Presses Universitaires de Rennes (PUR).

Markova, I. (2008). The Epistemological Significance of the Theory of social representations. *Journal for the Theory of Social Behaviour*, 38(4), 461–48. doi:10.1111/j.1468-5914.2008.00382.x

Moreira Parades, A. S., & Camargo Vizey, B. (Eds.). (2007). *Contribuições para a teoria e o método de estudos das representações sociais*. Joao Pessoa: Editora Universitaria UFPB.

Moscovici, S. (1961/1976). *La psychanalyse, son image et son public*. Paris: Presses Universitaires de France. Engl. Transl. (G. Duveen, Ed.) *Psychanalysis: Its Image and Its Public*, Cambridge: Polity Press, 2008. Italian Transl. (A.S. de Rosa, Ed.) *La psicoanalisi, la sua immagine, il suo pubblico*, Milano: Edizioni Unicopli.

Moscovici, S. (2000). *Social Representations: Explorations in Social Psychology*. Cambridge, UK: Polity Cambridge Press.

Moscovici, S. (2012). Saggio sulle rappresentazioni sociali e le minoranze etniche. In I. Galli (Ed.), *Cinquant'anni di Rappresentazioni sociali. Bilanci e prospettive di una Teoria in continuo divenire* (pp. 147–168). Milano: Edizioni Unicopli.

Moscovici, S. (2013). *Le scandale de la pensée sociale*. Paris: Editions EHESS.

Palmonari, A., & Emiliani, F. (2009) (Eds.) *Paradigmi delle Rappresentazioni Sociali*, Bologna: Il Mulino

Sammut, G., Andreouli, E., Gaskell, G., & Valsiner, J. (Eds.). (2015). *The Cambridge Handbook of Social Representations*. Cambridge: Cambridge University Press. doi:10.1017/CBO9781107323650

KEY TERMS AND DEFINITIONS

European/International Joint PhD in Social Representations and Communication and So.Re.Com. Joint-IDP: Built on the experience of both the European PhD on Social Representations and Communication, awarding a recognized joint degree since 1996 (<http://www.europhd.eu>), and the EU approved So.Re.Com. THEMATIC NETWORK, this international joint doctoral programme, co-ordinated by Sapienza University of Rome (Italy), includes 8 universities, 2 private companies and 1 public research institute in 8 European countries (AT, CH, CZ, ES, FR, IT, RO, SE) and 6 universities in United States, Canada,

Brazil, Argentina, Mexico and China. The goal is to provide doctoral training in Social Representations and Communication, a supra-disciplinary research area of the social sciences that studies the social construction of everyday knowledge in social spheres and media, disseminating European excellence beyond the EU boundaries and attracting the best Early Stage Researchers from abroad. Structured into transnational teams by common research area and complementary multi-methodological approaches, this So.Re.Com.Joint-IDP guarantees a well-tested training structure including: a) an innovative integrated physical and virtual campus, where world-class academic scientists, internationally recognized experts, experienced researchers and ESR cooperate face-to face and online “for” and “by” research; b) multiple supervision via tutoring and co-tutoring by at least three tutors in different countries; c) individual mobility for ESR at research centers for secondments; d) collective international mobility of trainees and teaching staff during International Summer Schools and Lab meetings; e) learning by doing (including transferable skills) in academic and non-academic settings; f) worldwide access to common web platform, as tool for documentation, networking, training and monitoring trainees’ progress; g) high tech infrastructure and lab facilities; h) quality evaluation system; i) officialisation of the joint degree; l) active integration in the world-wide SoReCom THEmatic NETwork; m) enhancement of career prospects both in and outside academia.

Social Representations: Social Representations and Communication is a supra-disciplinary area of the social sciences, representing a unifying meta-theoretical perspective, in particular from the optic of Social Psychology, inspired by the Social Representations Theory, one of the most important theories of the social construction of knowledge and its relation to socially situated

practices, founded by Serge Moscovici in 1961. This supra-disciplinary field is characterized by a great consistency in terms of epistemological and theoretical inspiration and its rich diversity both from the paradigmatic point of view, in terms of methodological approaches, from the thematic point perspective and with respect to the applied contexts and domains of expert and lay knowledge production and transmission: education, health, economics, environment, tourism, politics, organizations, media industry, etc.

So.Re.Com. “A.S. de Rosa” @-library: The So.Re.Com. “A.S. de Rosa” @-library is a multi-form digital environment of integrated relational repositories conceived in the logic of the semantic web, including a series of web interfaces aimed to integrate documentation services (bibliographic repository, meta-theoretical analysis repository, intelligent @-Library) with networking (interactive web-videoconference, on-line So.Re.Com. virtual community) and research training (European/International Joint PhD “Virtual Campus”: video-courses in streaming, distant tutoring and co-tutoring, on-line trainee evaluation, timeline management, etc.).

ENDNOTES

- ¹ Honorary program director of the European/International Joint PhD in Social Representations and Communication until his death (November 15, 2014): <http://www.europhd.eu>
- ² About the on-going development of *Sapienza Digital Library* and the related project Google Books see: Di Iorio, A. Schaerf, M. Guercio, M. Ortolani, S. and Bertazzo. M. 2013; Magarotto. A. 2013; see also: http://w3.uniroma1.it/biblioteche/risorse_elettroniche.htm



Social Media Applications as Effective Service Delivery Tools for Librarians

Ihuoma Sandra Babatope

Delta State College of Physical Education, Nigeria

INTRODUCTION

Social networking refers to a process of relationship building among a group with a common interest (Suraweera, Kahingala, Batepola, Punchihewa, Senevirathne and Kahandwaarachchi, 2011). The use of social media for information communication, information sharing, for pleasure, for dating, work related purposes and so on, is growing rapidly, therefore, their importance in academic libraries cannot be over emphasized. Academic libraries are libraries established in higher institutions (universities, polytechnics and colleges of education) by their parent institutions for the primary purpose of teaching, learning and research which therefore forms an integral part of the institution's educational system.. With the exponential growth of the use of social media such as the Facebook, Friendster, MySpace, Flickr, Instagram, LinkedIn YouTube, Twitter, Blogs, Wikis, Delicious, Online groups/forums,, Library Thing, Ning and so on, it has also become inevitable that academic librarians around the world and especially in Nigeria, must learn to use these tools to be able to keep their ever growing and sophisticated patrons and also remain relevant in the society.

Adeleye (2015) noted that these sites can be oriented towards work-related contexts (LinkedIn.com), romantic relationship initiation (Friendster.com), connecting those with shared interests such as music or politics (MySpace.com, Twitter), or the undergraduates population (Facebook.com). Social media such as Facebook was used initially for social discussions, over time it evolved into

the grouping of individuals into specific groups. Professional groups started to spring up and within time the library profession had its own group with the sole purpose of sharing ideas and gathering first hand information regarding the profession.

Beyond this, the different academic libraries in the country have felt the need to move with the times. The emergence of social media in academic libraries has changed the way information is delivered to the library patrons. Libraries have started to use these tools to interact with their patrons on real time. In fact, given the present economic scenario in Nigeria, where library budgets have been constantly on the decline, the social media have become a means for serving our patrons in a more specialized, interactive, and value added way without incurring undue expenses.

Hence, it has become very important for academic libraries to embrace these social media for effective service delivery to their library users and also to improve the library profession tremendously in Nigeria. Also, the previous top-bottom approach to service delivery where the library would pass information down to its patron without feedback will no longer suffice for our ever growing clients. Library users have been yearning to be a part of the services rendered to them, in essence being able to dictate what they need thereby making for a more interactive service delivery which will foster a two way communication pathway and provide the opportunity for more involvement. Undoubtedly, Nigeria which is still suffering from dwindling economy, are faced with diverse challenges amidst the opportunities derived from the intrusion of social media.

Consequently, a gap exists in literatures on the intrusion of social media in academic libraries in Nigeria. Hence, this chapter sought to bridge this gap by investigating the intrusion of social media in academic libraries in Nigeria as well as their opportunities and challenges for effective service delivery to their library users. Therefore this chapter will sought to identify the various types of social networking applications in academic libraries in Nigeria and how librarians use these social media for effective service delivery to their users, the opportunities accrued in using these applications, the various constraints in using these social networking applications in academic libraries, recommendations and the future trends will be succinctly discussed in this chapter.

BACKGROUND

Social media which are also known as social networks are tools (through the registration of profiles with their platforms) that help promote the sharing of information on real time basis, and these tools are dependent on modern digital communication like the Internet and the www (Ekuoye, 2015). In addition, social media are seen as computer-mediated tools that allow people to create, share or exchange information, career interests, ideas, and pictures/videos in virtual communities and networks (Buettner, 2015). However, social media was broadly defined to refer to ‘the many relatively inexpensive and widely accessible electronic tools that enable anyone to publish and access information, collaborate on a common effort, or build relationships’ (Murthy, 2013).

In the past, most of these relationships existed mainly among family members, colleagues, friends, church members, and others who relied mainly on face-to-face communication. Today the story is no longer the same as most of these relationships can be carried out in virtual space using social networking sites which makes it possible for people to interact on the Internet. Online

social networks can therefore, be described as relationships between groups of people carried out in virtual space.

The history of online social networks dates back to 1997 with the appearance of SixDegrees.com and there are over 100 websites which provide social networking functionalities (Boyd and Ellison, 2007). Social networking sites are two-way transparent communication that encourages a feedback mechanism; connecting people with shared interest. In the social network site, the user is a participant, a co-creator, and a builder of knowledge. The dynamic nature of this technology enables users to have an open access to knowledge and contribute local content on the social network space. This has being made possible by the tremendous growth and increase in the use of social media across the world. However, the developing countries are still lagging behind compared to their counterparts in the developed countries. This is confirmed in figure 1, which is social media use around the world.

According to Adeleye (2015), “Social media and its elements in Nigeria isn’t really as new as some people believe it is, social media started in Nigeria as a concept many years ago, but has evolved into sophisticated technology. The concept of social media invariably can be dated back to the use of the analog telephone for social interaction in the 1990s.

The emergence of social media in Nigeria has become very prevalent in every sector of the country for example in education, agriculture, aviation, health, finance, politics, entertainment and to mention but a few. Digital in Nigeria, a snapshot of the country key digital statistical indicators from Wikipedia internet live stats (2015) shows that; the total population of Nigeria is 183.5million (Urbanisation 50%) with active internet users of 70.3million (38%), active social media accounts of 13.6million (7%), mobile connections of 138.0million (75%) and active mobile social accounts of 12.4million (7%).



Figure 1. Social Media Use

Source: Wearesocial.net



MAIN FOCUS OF THE ARTICLE

Issues, Controversies, Problems

Online Social Networks and Academic Libraries

The use of online social network applications is still gaining popularity in academic libraries especially in Nigeria. Although there are hundreds of websites which provide online social network functionalities, but not all of these sites are likely to offer the same degree of usefulness to academic libraries.

Many researchers from the field of library and information science are beginning to show interest in online social networks. Cook and Wiebrands (2010) researched into librarians' use of online social networks for current awareness. The study revealed that curiosity was the major reason why librarians use social networking sites.

Differing viewpoint by Charnigo & Barnett-Ellis (2007), in a survey of 126 academic librarians about their knowledge of and attitude toward Facebook, revealed that 54% of the librarians

surveyed were overwhelmingly aware of Facebook but were not involved because they did not believe that there was an academic purpose for Facebook.

However, coming down to Africa, in Kenya, Gichora and Kwanya (2015) in their study revealed that Web 2.0 tools were widely used in most academic libraries in Kenya. In fact, all the sampled libraries used one or more Web 2.0 tools. It is also evident from the findings that the use of Web 2.0 tools in academic libraries in Kenya has had a positive impact on the effectiveness and efficiency of services they provide.

Also in a study carried out on the Use of Facebook in Information Service Delivery in Academic Libraries in Niger Delta, Nigeria by Achugbue and Ogbomo (2015), they observed that the entire 15 academic libraries which represents 100% of the population of the study has no Facebook application in their libraries.

Allison (2012) noted that Social media forms part of the rich environment of scholarship within which researchers work in the 21st century, and a librarian who does not take the time to grasp these changing conceptions is doing herself and her patrons a disservice.

Social Media Apps Used in Libraries in Nigeria

Social media applications allow librarians to adopt a new role by placing themselves into a social realm with users. By reading blogs, group postings, and message boards, the librarian becomes an active participant, who is able to anticipate and advise patrons as needs arise. Linking to patron profiles also keeps the library within the consciousness of users, potentially increasing interaction (Courtney, 2007).

With the enormous popularity of Web 2.0's social networking platforms, libraries of all types have embraced them as a method of promoting themselves within their communities. Some of these Social media applications gaining popularity and popularly used by librarians in academic libraries in Nigeria for effective information service delivery to their users include:

Facebook

Carlson (2010) opined that facebook was originally developed by Zuckerberg, Moskovitz, Hughes & Saverin in 2004 at Harvard University in order to provide the students with a place in which they could keep in contact with their classmates and share study-related information. One of the primary uses of Facebook by academic libraries is to market the library with a library fan page for understanding the library users' information needs.

Facebook statistics which was updated October 2015 shows also that at 1.49 billion, Facebook has more monthly active users than WhatsApp (500 million), Twitter (284 million) and Instagram (200 million)—combined (facebook, 2015). This shows that Facebook continues to reign in popularity over other social media channels.

It is most popular now also in academic libraries in Nigeria because it is librarian- friendly, with many applications like JSTOR search, World Cat, and much more. Academic Libraries in Nigeria use Facebook, mainly to advertise the library services push out announcements to library users, post

photos/videos, share ideas, provide chat reference, and have a presence in the social network without actually visiting the library.



MySpace

MySpace, which is popularly known as a “place for friends”, is defined as “an online community that lets you meet your friends.” MySpace interface provides a number of networking tools in a simple and relatively self-explanatory interface, including facilities that enable users to send messages back and forth to each other, blog, post bulletins to friends, leave comments for other MySpace users, and instant messages.

In academic libraries in Nigeria, library users have also taken advantage of this site to post, customize catalog search tools, and blog features to improve their presence. Many libraries have seized upon the networking opportunities MySpace affords, in building their own profiles and connecting with users. However, MySpace is not as popular as facebook in academic libraries in Nigeria.

Ning (www.ning.com)

According to Educause Learning Initiative (2008), Ning is an online service that allows users to create their own social networks to join, and participate in other networks through Interaction by engaging your community directly (messaging, wall posts, blogs, and so on). Although Ning does not provide usage statistics or numbers of registered users, the company does highlight the number of networks on its platform.

Ning provides an opportunity for students to create their own social networks and learn how to cultivate and sustain a community of users that might resemble professional contacts and relationships. Librarians can get connected with users, other libraries, library associations, and more for professional growth opportunities. The Nigerian Library Association (NLA) utilizes this platform to discuss topical issues among the members.

Blogs

Blogging is an easy process for publishing ideas on the web to get the reactions of other users of the web. Libraries all over the world are using blogs for the easy dissemination of information to the targeted users. Blogs can be used to communicate library events, publicize information resources, train staff, and offer subject-related reference services (Han and Liu, 2010).

The academic libraries in Nigeria by librarians have enormous use of blogs which includes: discussion groups, writing of articles, news on topical/current issues, book reviews, research purposes, and links to library catalogue.

Wikis

Wikis are free online encyclopedia that gives a background knowledge and definition of concepts. It offers a platform for users to access, edit and contribute to content. This is a collaborative web page for developing web content. A wiki is a collaborative web site which can be directly edited by anyone with access to it.

Studies have shown that librarians use wikis to collaborate with librarians in other libraries, and that few libraries use wikis to collaborate with patrons (Kim and Abbas, 2010). In academic libraries in Nigeria, wikis are used by librarians for knowledge/ information sharing, collaboration purposes with other librarians and research purposes (multidisciplinary subject coverage).

LinkedIn

LinkedIn (www.linkedin.com) is a social networking site for professionals and career connections that can help you promote your small business. LinkedIn dominates the professional social network segment. It is one of the oldest having started in 2002. Librarians can get patrons connected with specialists in their particular field of interest via LinkedIn. Librarians can use this platform to render specialized services such as Selective Dissemination of Information (SDI).

LinkedIn, in the opinion of Marshall (2009), stands out as a worthwhile social networking destination for library professionals. Through introductions made by online colleagues, librarians can continually expand their professional contacts using LinkedIn. LinkedIn is gradually gaining popularity in academic libraries in Nigeria, because librarians and users are now building a good professional network with LinkedIn. They can also share professional interests by forming a group.

Twitter

Another social networking site gaining popularity is Twitter, a micro blogging application, to keep staff and patrons updated on daily activities. According to Olaniran (2014), twitter is an online social networking and micro blogging service that enables users to send and read “tweets”, which are text messages limited to 140 characters. Registered users can read and post tweets but unregistered users can only read them. It was created by Jack Dorsey, Evan Williams, and Biz Stone in 2006. Users access Twitter through the website interface, SMS, or mobile device apps. It is now one of the ten most visited websites, and has been described as “the SMS of the Internet”.

In a study, Perceptions and acceptance of librarians towards using Facebook and Twitter to promote library services in Oyo State, Nigeria by Omobolanle (2015); the paper revealed that majority of the libraries in Oyo state, Nigeria, have Facebook profiles, though minority have Twitter accounts with high perception and acceptance of using Facebook and Twitter to promote library services. Also a research by Olaniran (2014) showed that the students of the University of Jos use Twitter for informational, academic and social purposes. Librarians in Nigeria use this platform to give users first hand information. Marshall (2009) identified Twitter as a tool for staying current. Twitter allows users to send instant messages to social network via short messages, also known as micro-blogs.

Discussion Group/ Online Forum

The first library online professional group was the Nigeria Library Association (NLA) forum (nla-online-forum). Librarians from all institutions in Nigeria enroll in this group by subscribing to it. This allows the librarians to exchange information on professional opportunities, professional events and new technologies within the Library and Information science field. This national group has continued to grow in leaps and bounds.

YouTube

YouTube provides the user with the ability to save favourite videos, create a playlist, rate videos, make comments about them, and share them with other users (Buckley, 2008). In academic libraries in Nigeria, events such as important highlights of inaugural lectures, talks, conferences, seminars and workshops are disseminated via the YouTube.

Flickr

Flickr is known largely as a photo sharing web site, it also allows users to post video. Academic Librarians in Nigeria are using this tool to share library photos, create discussion groups, distribute new images of library collections, and a virtual tour of the library. Cover page of new arrivals of both books and journals can be disseminated to users via Flickr. It can also be used to enlighten users on topical issues such as the different pictures of emblems of the political parties in Nigeria; on-going elections in Nigeria, many Public and Academic libraries put this to great use.

Library Thing

Library Thing is a social networking tool that enables users to catalog their books. This is done through the library OPAC, once an account is created, a list of books with ISBNs is sent to Library Thing which sends back a piece of code which is pasted into the footer of the Library OPAC. Aca-

ademic libraries in Nigeria are utilizing this to send a list of current publications to their library users.



Really Simply Syndication (RSS)

Xu (2007) found that blogs, IM and RSS were the main Web 2.0 applications which have been used extensively by academic libraries of New York State and Long Island in the USA. Libraries have already been making use of this new internet tool for disseminating library news and current alerts. Some of the Nigerian academic libraries major purpose of providing RSS news feeds in their library web sites includes: to disseminate library events, new arrivals, exhibitions, catalogue search, publishing library news and bulletins.

Instant Messaging (IM)

Instant Messaging (IM) is also a very useful tool which helps library professionals to provide library services. Some academic libraries in Nigeria currently use IM mainly to provide online reference services. "Ask a librarian" services are provided by instant messengers all over the world. Reference questions are answered immediately without the need to go to a reference desk. Librarians and patrons both appreciate the fact that instant messaging is faster than traditional chat services, and librarians also like the fact that the software can be downloaded for free (Houghton and Schmidt, 2005).

Opportunities for Libraries

The different sectors in the country have adopted the use of social media due to certain features which makes it more unique and accessible than the traditional media. Such features include and are not limited to affordability, availability, timeliness, current / easy updates, immediate feedback and flexibility.

In a study carried out by Onuoha (2013), it established that librarians in Nigeria actively seek means of developing themselves professionally.

Although various forms of social media applications exist, the valid empirical evidence showed that librarians in Nigeria use mostly online discussion forums, online social networks such as Facebook, Wikis and blogs for professional development, while social bookmarking tools, virtual world, podcasts and photo sharing sites are used minimally. Social networking presents some important opportunities to libraries which include marketing of library services and reference services (Ezeani & Igwesi, 2012).

However, some opportunities accrued to the libraries through social media are:

1. **User-Friendly:** The social media apps are user friendly because the users can fully participate, collaborate and interact with the librarians without restrictions and are allowed to make contributions as well.
2. **Reference Services:** Quick dissemination of information is attained by responding to the library user's complaints, queries, comments and feedbacks without physically visiting the library.
3. **Advertising / Marketing of Library Services:** Libraries can now publicize their library events, advertise new library resources / new arrivals, library orientation, exhibitions/ displays, book fair, library bulletins /news, seminars, talks and so on, are all made available on these platforms to promote the library services to their library patrons.
4. **Affordable Price:** With the emergence of social media, delivering information to the library users is now affordable notwithstanding the economic situation of the country which has affected the library too. According to Njoroge and Kang'ette (2013) social media is now a means of cheaper communication that is instant and fast.
5. **Enhances Information / Knowledge Sharing:** Librarians now provide and share more current and up-to-date information to their library users.
6. **Socialization:** Social media has increased the discussions, interactions and connections between librarians and their library users, librarians and other librarians/other staff of the school community including other professionals in different sectors, in order to keep abreast with new developments.

Problems Associated with Use of Online Social Media

Nigeria has several factors that militate against the use of Information Communication Technology (ICT) generally. According to Gbaje, (2007) the erratic nature of electricity supply in the country will not support the successful implementation of the new information trends like social networking in libraries in this age of modernization. Some of the major challenges faced by the intrusion of social media in academic libraries in Nigeria are as follows:

1. **Unsteady Power Supply:** This is assumed to be the greatest challenge faced by every sector in the nation and the academic libraries are not exempted. This has made most libraries to fail in the effective service delivery to their library users, as most libraries cannot afford the cost of running a generator daily.
2. **Poor Funding:** Without proper funding of the libraries and adequate budget allocation to libraries, they will not also be able to incorporate the social media into their various libraries.
3. **Inadequate Training:** Most librarians are not trained in these new modern technologies and are most times not sponsored by their institution to attend training sessions to update themselves. According to Kemrajh (2013), librarians are not given adequate time to embrace the technologies and social media is not given a priority in library work.
4. **Resistance to Change:** Most librarians still prefer to carry out their library operations using the traditional method than using the

modern methods. Some of their reasons are due to traditional beliefs (technologies are for the white man), negative attitudes, lack of computer self efficacy, technophobia and so on.

5. **Lack of Privacy Policy:** Most of the libraries have not developed or implemented Social Media privacy policies in their libraries hence copyrights, privacy, and other legal rules are not complied with. This can pose a security risk to the users if perpetrators take advantage of their personal information displayed on the social networking websites.
6. **Academic Fraud:** According to Adeleye (2015), academic fraud is also part of the negative side of social media since its inception in Nigeria, as many people copy another person's work online without acknowledging the real owner of the work, otherwise called plagiarism.
7. **Addictive Behaviours:** Many studies have shown that the extensive use of social media can actually cause addiction to users. Most users especially the undergraduates are victims of this behaviour.
8. **Social Violence:** There has been a rise in social violence activities in Nigeria. Some of the social violence witnessed in Nigeria from 2008 to 2015 has resulted in the death of the victims. Others were kidnapped, raped and molested through people they met on social networks.

All these and many more limit the full incorporation of social media apps in academic libraries in Nigeria for effective service delivery to their library users.

SOLUTIONS AND RECOMMENDATIONS

With respect to the foregoing the following recommendations were put forward:

1. Provision of stable power supply will encourage and facilitate the effective use of social media applications.
2. There is need for social media to be given priority by training librarians to in social media / networking applications.
3. The budgetary allocation to libraries should be upgraded to enable libraries afford the necessary facilities for social media applications.



FUTURE RESEARCH DIRECTIONS

The emergence of social media applications due to modern technologies in the society has brought about its intrusion in academic libraries all over the world and also in Nigeria. Future research directions should be towards the privacy policy issues in using the social media/ networking applications in academic libraries in Nigeria to reduce the social violence rate. The educational purposes of social media in embedded librarianship, teaching and learning should be looked into as well. Also entrepreneurship in using social media in academic libraries is required to be addressed.

CONCLUSION

This chapter examined issues pertinent to the intrusion of online social media in academic libraries in Nigeria. Indeed, the benefits of using social media applications in libraries cannot be overemphasized; it is obvious from existing literatures that there is no clear indication of a positive use of social media applications in performing library services by librarians in Nigeria. However, where there is the intention to use these media factors such as; unsteady power supply, poor funding, lack of training and so on appears to constitute aberrations. As such, there is a need for pro-active measures to encourage the use of social media applications in academic libraries in Nigeria.

REFERENCES

- Achugbue, I. E., & Ogbomo, M. O. (2015). The Use of Facebook in Information Service Delivery in Academic Libraries in Niger Delta, Nigeria. *Scholars Journal of Economics, Business and Management*. Available Online: <http://saspjournals.com/sjebm>
- Adeleye, K. (2015). Evolution of Social Media in Nigeria. *High Flyers Magazine*. Retrieved from hfmagazineonline.com
- Alison, H. (2012). Review of "Social Networking Tools for Academic Libraries". *Collaborative Librarianship*, 4(4), 190–191.
- Boyd, D., & Ellison, N. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1), 210–230. doi:10.1111/j.1083-6101.2007.00393.x
- Buckley, B. (2008). SueTube: Web 2.0 and copyright infringement. *The Columbia Journal of Law and the Arts*, 31(2), 235–263.
- Buettner, R. (2016). Getting a Job via Career-oriented Social Networking Sites: The Weakness of Ties. *49th Annual Hawaii International Conference on System Sciences*. Kauai, HI: IEEE. doi:10.1109/HICSS.2016.272
- Carlson, N. (2010). At Last — The Full Story Of How Facebook Was Founded. *Business Insider*.
- Charnigo, L., & Barnett-Ellis, P. (2007). Checking out Facebook.com: The impact of a digital trend on academic libraries. *Information Technology and Libraries*, 26(1), 23–34. doi:10.6017/ital.v26i1.3286
- Cook, S., & Wiebrands, C. (2010). *Keeping up: strategic use of online social networks for librarian current awareness*. Retrieved August 8, 2015 from: http://www.vala.org.au/vala2010/papers2010/VALA2010_78_Cook_Final.pdf
- Courtney, N. (2007). *Library 2.0 and beyond: Innovative technologies and tomorrow's user*. Westport, CT: Libraries Unlimited.
- Educause Learning Initiative. (2008). Retrieved from www.educause.edu/eli
- Ekuoye, O. (2015). *Justifying social media for Nigerian libraries and librarians*. Paper presented at the annual general meeting and conference of Edo state chapter of Nigerian library association at the ultra modern conference hall of Igbinedion University, Okada, Edo State.
- Ezeani, C. N., & Igwesi, U. (2012). Using Social Media for Dynamic Library Service Delivery: The Nigeria Experience. *Library Philosophy and Practice*. Retrieved August 8, 2015 from: <http://digitalcommons.unl.edu/libphilprac/814>
- Facebook. (2015). *Statistics Facebook*. Retrieved August 8, 2015 from: <http://www.facebook.com/press/info.php?statistics>
- Gbaje, E. S. (2007). *Implementing a National Virtual Library for High Institutions in Nigeria*. Retrieved August 8, 2015 from: http://libres.curtin.edu.au/libres17n2/Gbaje_2007_07_30a_Ess%20&%20Op_final.pdf
- Gichora, F. G., & Kwanya, T. (2015). The impact of Web 2.0 tools on academic libraries in Kenya. *International Journal of Library and Information Science*, 7(2), 21–26. doi:10.5897/IJLIS2014.0518
- Han, Z., & Liu, Y. Q. (2010). Web 2.0 applications in top Chinese university libraries. *Library Hi Tech*, 28(1), 41–62. doi:10.1108/07378831011026689
- Houghton, S., & Schmidt, A. (2005). *Web-based chat vs instant messaging: who wins?*. Academic Press.
- Kemrajh, M. (2013). Social media and academic libraries – is this a good fit? *Library space*.
- Kim, Y., & Abbas, J. (2010). Adoption of library 2.0 functionalities by academic libraries and users: a knowledge management perspective. *Journal of Academic Librarianship*, 36(3), 211–218.



Njoroge, G. G., & Kang'ette, P. (2013). *University libraries and social media: the postmodern Library, Kenyatta university*. Retrieved from http://www.0277.ch/ojs/index.php/cdrr_0277/articles/view/14/53

Marshall, B. (2009). Social networking strategies for professionals. *Computers in Libraries*, 29(9), 29–31.

Murthy, D. (2013). *Twitter: Social Communication in the Twitter Age*. Cambridge, UK: Polity.

Suraweera, Kahingala, Batepola, Punchihewa, Senevirathne, & Kahandwaarachchi. (2011). Value of social networking in libraries and information organizations in Asia and Oceania. *World Library and Information Congress, 76th IFLA General Conference and Assembly*.

Wikipedia internet live stats. (2015). Retrieved from internetlvestats.com

Xu, C. (2007). *The academic library meets Web 2.0: application & implication*. Retrieved August 8, 2015 from: http://eprints.rclis.org/12108/1/The_Academic_Library__Meets_Web_

Olaniran, S. (2014). Social media and changing communication Patterns among students: An analysis of Twitter Use by university of jos students. *Covenant Journal of Communication*, 2(1).

Omobolanle, S. F. (2015). Perceptions and acceptance of librarians towards using Facebook and Twitter to promote library services in Oyo State, Nigeria. *The Electronic Library*, 33(5), 870–882. doi:10.1108/EL-04-2014-0066

Onuoha, U. D. (2013). Librarians' Use of Social Media for Professional Development in Nigeria. *Information Management and Business Review*, 5(3), 136–143.

ADDITIONAL READING

Dankowski, T. (2013). How Libraries Are Using Social Media. *American Libraries*, 44(5), 38–41.

Dickson, A., & Holley, R. P. (2010). Social networking in academic libraries: The possibilities and the concerns. *New Library World*, 111(11/12), 468–479. doi:10.1108/03074801011094840

Swanson, T. (2012). *Managing Social Media in Libraries: Finding Collaboration, Coordination, and Focus*. Oxford: Chandos Publishing. doi:10.1533/9781780633770

Tella, A. (2013). Internet Policy Issues and Digital Libraries Management of Intellectual Property. In S. Thanuskodi (Ed.), *Challenges of Academic Library Management in Developing Countries* (pp. 272–284). USA: IGI Global Publishers. doi:10.4018/978-1-4666-4070-2.ch019

Tella, A., & Akinboro, E. O. (2015). The Impact of Social Media to Library Services in Digital Environment. In A. Tella (Ed.), *Social Media Strategies for Dynamic Library Service Development* (pp. 279–295). USA: IGI Global Publishers. doi:10.4018/978-1-4666-7415-8.ch015

KEY TERMS AND DEFINITIONS

Academic Libraries: These are libraries established in higher institutions like universities, polytechnics and colleges of education.

Information Delivery: This is the act of conveying information from one person to the other or from one place to the other.

Library Services: These are the various services rendered in the library to their users.

Nigeria: This is the most populous country in Africa otherwise known as giant of Africa.

Social Media: Online networking tool for interaction, communication and sharing of information among group of people.

Social Networking: This is the platform for people to connect, interact and discuss with one other.

Socialization: The act of interaction, information or knowledge sharing among group of people.

Web 2.0 From Evolution to Revolutionary Impact in Library and Information Centers

Zahid Ashraf Wani

University of Kashmir, India

Tazeem Zainab

University of Kashmir, India

Shabir Hussain

University of Kashmir, India

INTRODUCTION

Technological revolution is affecting different facets of social life, although there is some disagreement in the research regarding the positive and negative influences. The Internet allows greater flexibility in working hours and location, especially with the spread of unmetered high-speed connections. The World Wide Web is most significant technology after the advent of printing press. It is a global set of large number of resources interrelated by hyperlinks. Social Networks, RSS Feeds, Wikis are a part of Web 2.0 which is most advanced form of the web. It can be well articulated as the shift from simply being a website and a search engine to a shared networking space that drives work, research, education, entertainment and social activities, which essentially all people do. Web 2.0 applications are socially rich and community building is the core of these applications. A key element of the technology is that it lets people to create, share, collaborate & communicate. The Web 2.0 differs from ordinary websites as it does not require any web design or publishing skills to participate, making it easy for people to create and publish or communicate their work to the world. The nature of this technology makes it an easy and popular way to communicate information to either a select group of people or to a much wider audience.

The Web 2.0 paradigm has gained substantial momentum in the last decade. The influence of Web 2.0 principles and technologies has fuelled an explosion of information and media content on the Web, and individual and corporate adoption of the technologies continues to rise. In this milieu it is imperative to understand and learn about the present and evolving web 2.0 tools and their application in different walks of life. The proposed study made an endeavour to dig deep into the genesis, development and application of various Web 2.0 tools in library and information science.

The Internet allows greater flexibility in work, especially with the spread of high-speed connections. Internet-based technologies nowadays are playing a significant role in the way that societies communicate with the each other. It's a fact that the technologies allow information to be readily exchanged without geographic and time barriers. Many people use the terms Internet and World Wide Web synonymously. The World Wide Web is a global set of large number of resources interrelated by hyperlinks. It provides a mesh of services including email, social networking and Web 2.0. The Web 2.0 is not mere a website and a search engine but it furnishes a platform for research, education, entertainment and social activities, which essentially all people do. It also provides the services like blogs, wikis, and multimedia sharing services, content syndication, podcast-

ing and content tagging services. Applications of Web technology have been in use for years now, although new features and capabilities are being added on a regular basis. It is remarkable to see that many of these recent technologies are concatenations, i.e. they make use of existing services. World Wide Web (WWW) provided the base for Web 2.0 applications to create a new communication environment (Linh, 2008). It is a second wave that covers web tools and services. Davis (2005) describes Web 2.0 as an attitude not a technology and Birdsall (2007) believes that Web 2.0 is a social movement. So the Web 2.0 applications differ from Web 1.0 applications by their frequency of usage too. As the Web 2.0 applications are socially rich and community building is the core of these applications, so their usage increases many times than Web 1.0 static websites/applications. In 2004, Tim O'Reilly, the founder of O'Reilly Media (www.oreilly.com), used the term web 2.0 to describe the significant shift in how software developers and users were using the web. One of the characteristics of Web 2.0 web sites is that people go there to do something – as contrasted with Web 1.0 “brochure ware” sites that people came to primarily for information. Unlike in the past when it took months or years to implement new software, today we can download, set up and use powerful applications without the intervention of a technology expert. We can create our own website, manage huge databases, and stream rich media, the credit for all of this goes to the evolution of web. The users get attracted by these kind of experiences while they are added to the library environment. As per Dye (2007) libraries often impressed and get encouraged to use Web 2.0 technologies. These libraries that have adopted, implemented and embraced Web 2.0 technology are referred to have “Library 2.0” competences. As per Maness (2006) Library 2.0 offers various services like virtual reference services, databases, catalogue tagging, and downloadable media for home users. As Houghton-John (2005) defines it, Library 2.0 “simply means making a library’s space (virtual and physical) more interactive,

collaborative, and driven by community needs.” Library 2.0 has numerous façades reflecting the distinctive paths of user involvement that Web 2.0 facilitates. These façades comprise social networks; wikis; social bookmarks; blogs; photographs and audio; podcasting; videos; presentations; images; bibliographic reference managers; content syndication; messaging; video conferencing; chat, and much more.

BACKGROUND

Web 2.0 enable us to identify the various types of connections online environment avails to the users. Three key aspects of Web 2.0 have been identified in this regard, which are *Interpersonal Computing*, *Web Services*, and *Software as a Service*. Interpersonal Computing is about using the online technology to connect people to each other in social networks or business teams. The key Interpersonal Computing aspects of Web 2.0 are as SNS, Blogs, Wikis and online videos. All of these sites allow anybody to access as well as to add content. It allows users to leave messages and comments and also to exchange digital media in form of photos and videos etc. *Web Services* are components of online functionality that can be plugged together like the kind of digital letter in order to create integrated online offering or mashup. The final key aspect of Web 2.0 is ‘*Software as a Service*’ (*SaaS*). This involves application functionality being offered directly over the internet. In turn used data and application can then be accessed through internet enabled computing device (Barnatt, 2012). In addition to this, Web 2.0 can be described in three parts. (i) Rich Internet application (RIA) — defines the experience brought from desktop to browser whether it is from a graphical point of view or usability point of view. Some buzzwords related to RIA are Ajax and Flash. (ii) Web-oriented architecture (WOA) — is a key piece in Web 2.0, which defines how Web 2.0 applications expose their functionality so that other applications can



leverage and integrate the functionality providing a set of much richer applications. Examples are feeds, RSS, Web Services, mash-ups. (iii) Social Web — defines how Web 2.0 tends to interact much more with the end user and make the end-user an integral part (Web 2.0, 2013). Further, there are a number of Web-based services and applications that demonstrate the foundations of the Web 2.0 concept, and they are already being used to a certain extent in libraries. These are not really technologies as such, but services (or user processes) built using the building blocks of the technologies and open standards that underpin the Internet and the Web. The introduction of Web 2.0 tools into library mesh sites inspires collaboration or communiqué between customers and the library that drives afar than physical services increasing the prospective of library facilities to patrons situated afar from the corporeal organization. Libraries have evolved in terms of the standards and technology used specifically. For example The Young Adult Library Services Association's (YALSA) Public Library states that *"Young Adults Deserve the Best supports the use of social networking and Web 2.0 services like blogs and podcasts in the classroom and library"* (Husid, 2010). Linh's (2008) study examined the actual use of Web 2.0 at Australasian university indicates that the web 2.0 applications are well used in the university and further revealed that most popular applications were blogs, wikis, RSS, messaging, social networking, and social tagging. A similar study by Han and Liu (2010); Si, Shi, and Chen (2011) while conducting the similar study of Chinese university libraries, had yielded nearly identical findings.

WEB 2.0 SERVICE AND THEIR APPLICATION IN LIBRARIES

The prominent Web 2.0 services and applications are wikis, multimedia sharing services, content syndication, podcasting and content tagging services etc.

Blog

The term blog originally come from the phrase "web-log", which refers to a simple webpage containing paragraphs of opinion, information, personal diary entries, or links arranged in a chronological order with the most recent entry first in the style of an online journal (Doctorow, 2002). Blogs, which represent the earliest form of Social Media, are special types of websites that usually display date-stamped entries in reverse chronological order. Blogs have a variety of formats and might include the user expressing their opinion about a topic or documenting activities. Blogs are interactive in the sense that other users could provide comments on the information posted by the blog author. The educational applications of blogs include researching, tracking, interpreting, and evaluating blogs for political commentary (multiple perspectives), cultural events, business, or other news and for examining changes over time (Alexander, 2006). A blog takes in a user-friendly commentary format with separate postings rather than a threaded discussion board (Richardson, 2006). Blogs are personal Web diaries, where users can offer their ideas, experiences and opinions on any topic. Technorati (Internet search engine for searching blogs) currently counts around 112 million blogs worldwide with many more being added every day (Wright, 2013). Blog entries may include video and other rich media depending on the blogging software or service that is used. Blogs are now also employed by libraries for giving due publicity to their new activities, events and services and as such acts as a platform to get instant user feedback from cross section of the population. Rogers (2015) in his study found that 96% of the selected library websites featured an RSS application. In fact, the most common RSS application was employed for new book arrivals, book appraises, whereby libraries displayed book jackets or abstracts of the afresh titles. This way libraries can plug the holes and achieve excellence in their endeavours..

Wikis

The word Wiki originates from a Hawaiian term which means “quick” or “Super-fast”. As popularly known a wiki refers to a web site that anybody can edit. Knowledge in the Web 2.0 environment is built directly by the users. Some of the most popular tools for building shared knowledge are so-called wikis. Wikipedia’s popular success has meant that the concept of the wiki, as a collaborative tool that facilitates the production of a group work, is widely understood. (Ebersbach, & Glaser, 2006). The first wiki was launched by Howard G Cunningham whose wiki site was named as WikiWikiWeb on March 25, 1996. Wagner and Bolloju (2004) are of the view that as wikis are informal technology, they are effectively used for temporary problems with dispersed knowledge foundations. In the context of libraries, Wikis can aid reference librarians fill knowledge management requirement. It can be used as a ready reference database for FAQs. It can also be used as knowledge repository (Cohen, 2005). Wikis can also be used as cooperative booths to help accomplish knowledge for specific projects or teams in library reference services (Frumkin, 2005). “WikiRef” created by Reference Librarians of Butler University and “Biz Wiki” at Ohio University Libraries envisioned to be a joint review of reference resources (Kille, 2005).

RSS

RSS stands for Really Simple Syndication or Rich Site Summary. It is a technology which has brought about a significant advance in the fundamental architecture of the web (O’Reilly, 2005). It allows you to keep up-to-date with news and information from your favorite Web sites. RSS is a family of Web-feed formats that support the syndication of Web content. (This means making information and data on a Web site available to other Web sites.) The feeds notify people when new material is added. The feeds can provide text-only summaries, full pages of content, audio

(podcasting), or video. The feed receiver benefits from access to fresh and timely material that can be used on his or her own Web site to make it more interesting to visitors. The feed sender benefits from the increased exposure and traffic the feeds draw (Stiefvater, n.d.). RSS feeds is age old alerting service offered by libraries in new form and format. Libraries in digital environment can offer CAS and SDI services to the users using RSS feeds. Varnum(2006) provides a detail about how libraries use RSS feeds for patron use. *BlogBridge: Library (BBL)* is an application based on RSS which when mounted in a library’s system and attached with the social network of the library, will permit customers to take a solo, personalized, personal library page that groups all the library material to their research (Maness, 2006). RSS feed if facilitated to scientists and professionals can help them to do better in their respective areas.

Podcasts

A combination of the words “iPod” and “broadcast,” podcasts are digital media files distributed over the Internet and listened to on a portable media player. A related term is “vodcast,” which describes podcasts that incorporate video. Podcasts were originally conceived as a way for people to create their own radio shows without needing a recording studio or transmission network. They evolved into a means of recording and distributing speeches, classes and training sessions, and public safety messages. Podcasters and vodcasters record their audio and video sessions, edit them, and upload them to a feed. Listeners and viewers use a “pod catcher” service such as iTunes to search for and subscribe to one or more feeds. When a new podcast or vodcast is released on the feed, the audio or video file will download automatically to the subscribers’ iPod, computer, or other device so that they can listen to or view it at their leisure (Stiefvater, N.D). Publishing audio and video recordings of lecture materials and class discussion on a podcast is a popular way of in-



formation sharing among the new generation of students, teachers and social workers. Originally podcasts were called audio blogs and were the beginning efforts of adding audio streams to early blogs (Felix & Stolarz, 2006). Libraries being the centres to give neutral and non-judgemental platform to users can offer users virtual space to host their podcasts/vodcasts. Since most of the podcast/vodcast publishers and users are youth therefore, this move if taken has the tendency to make library once again nerve centre of knowledge sharing organisations.

Tagging and Social Bookmarking

A tag is a keyword that is added to a digital object (e.g. a website, picture or video clip) to describe it, but not as part of a formal classification system. Application of tagging was initiated in Joshua Schacter's del.icio.us website, who launched the 'social bookmarking' phenomenon. Social bookmarking systems have many features in common. They facilitate users to generate lists of 'bookmarks' or 'favorites', to store these centrally on a distant service and to share them with other clients. The idea of tagging has been expanded to include what are called *tag clouds*: groups of tags (*tag sets*) from a number of different users of a tagging service, which collects information about the frequency with which particular tags are used. This frequency information is often exhibited graphically as a cloud in which tags with higher frequency of use are displayed in larger text. Social bookmarking services such as Diigo or del.icio.us enable users to store lists of Internet resources they find useful and make them accessible to others with similar interests (Ktitraci, 2009). 'Social bookmarking' is the trend which was first motivated by the popularity of tagging sites such as del.icio.us (Anderson, 2007). Literature shows that the use of Social Bookmarking sites is not up to mark in the library environment. Maness (2006) suggests that tagging, or social bookmarking unambiguously, are "excellent resource discovery tools" which aid libraries to locate and share the

resources to its patrons. Therefore libraries have to play a proactive role to educate and equip users about the impact and benefits of these wonderful tools so that they can make most of this service.

Instant Messaging (IM)

Instant messaging (IM) is a collection of technologies that create the possibility of real-time text-based communication between two or more participants over the internet or some form of internal network/intranet. IM is an integral tool for reference services in terms of chat services. It could improve the timeliness of user interaction and the user's initiative, help libraries investigate user's requirement clearly and concisely in time, which is helpful to answer the questions (Si, Shi & Chen, 2011). Instant Messaging or "chatting" is a popular method of exchanging text messages in real time. Popular IM applications include WhatsApp, Facebook Messenger, AOL's Instant Messenger (AIM), Microsoft's Messenger, Google Talk (Hangouts) etc. IM (including text messaging) is fast replacing e-mail, especially among students. Libraries in present times can exploit IM technology as the real-time reference service to its users. Since the present generation is more tech-savvy particularly choosing Instant Messaging as the mode of communication, libraries can play an imperative protagonist to deliver information to its community via this web 2.0 tool. Maness (2006) states that the instant messaging reference services can be encouraged when certain consumer information seeking behaviors are perceived. It helps Librarian to know when users are vanished, and will offer instantaneous, prompt and real-time assistance.

Social Networking Sites

Another application of web 2.0 that has facilitated the timely and fast interaction among people in a more convenient way is Social networking sites. These sites displays user profiles, social links, and a variety of added services and enables users to share

views, thoughts, activities, events, and interests within their individual networks over the internet through e-mailing and instant messaging (Hossain & Aydin, 2011). Social networking sites act as a channel via which the user can share personal multimedia like audio and video clips, images, geographical location, contacts etc. It further allows users to restrict viewing of their personal information by means of privacy settings. (Wikis, Blogs & Web 2.0 Technology, 2008). Networking Sites like Facebook, Twitter, MySpace and LinkedIn facilitates online communication in between friends, colleagues and even strangers (Ktitraci, 2009b). Social networking sites allow users some control over the look and feel of their pages and can be customized with add-ins and “gadgets,” small software applications that provide information or amusement. Social network is the most prominent tool used to connect with the user community and libraries are no exception to it. In the present digital age, libraries can provide services and content to its users by providing information using social networking sites like Facebook. These networks can also be used to promote the libraries on wide platform. It provides a podium to libraries whereby information can be shared anytime and anywhere. *LibraryThing*, a library-based social network facilitates users to catalog their books, how librarians endorse reading to users. *LibraryThing* allows users, to recommend books to one another by seeing other’s collections. It also allows them to connect asynchronously, blog, and “tag” their books (Maness, 2006). SNS were found to be the third common used Web 2.0 feature with 55 libraries having one or more account

File-Sharing Sites

File-sharing sites let users upload files, photos, and videos for others to view and download. Users can also create an individual profile and list their favourite photos or videos (Wikis, Blogs & Web 2.0 Technology, 2008). They usually include feedback tools that allow others to rate and comment. The most popular are Flickr and Photo bucket for shar-

ing images and YouTube for sharing videos. YouTube is the most widespread video sharing site in the world. It enables users to share and watch their own, or downloaded, videos, rate or comment on them. Flickr is a platform for uploading, sharing and storing pictures. PodcastAlley is also other file-sharing site where users can search for and listen to podcasts uploaded on the internet (Wright, N.D b). Since file-sharing sites are considered to be the present day digital libraries, they can be particularly used for e-book requests which helps in collection development of libraries.

Mash-Ups

A mashup is a term used to refer to a website or webpage that combines data and services from various sources on the Web (Murugesan, 2007). A mashup, in web development, is a web page, or web application, that uses and combines data, presentation or functionality from two or more sources to create new services. Mashups can be divided into seven categories: mapping, search, mobile, messaging, sports, shopping, and movies. More than 40% of mashups are mapping mashups (van der Vlist, 2006). A mash-up is a Web application that combines data from more than one source into a single integrated tool. An example is the use of cartographic data from Google Maps to add location information to real estate data. This creates a new and distinct Web service that was not originally provided by either source (Wright, n.d.). There are many types of mashup, such as business mashups, consumer mashups, and data mashups. Map mashup is the most popular among them. Libraries in present times have lots of opportunities to use mashups to deliver and help deliver through their websites, offering new and interesting services to the users. Library 2.0 is itself a mashup. It is a hybrid of wikis, instant messaging, content aggregators streaming media, blogs and social networks. It lets the user to edit OPAC data and metadata, IM chats with librarians, protects the user’s tags, wiki entries with additional users and the user is capable to control access to their



profile; borrow and lend tags, and a massive user-driven catalog is generated and mashed with the traditional catalog (Maness, 2006). For example Cambridge Libraries have used map mashup on their website to locate their library.

FUTURE RESEARCH DIRECTIONS

Web is fast-evolving from simple Web to Web 2.0 to Web 3.0. This evolution is fascinating users and the expectations are increasing. The momentum must be maintained by the Web developers so that in coming years with the addition of virtual reality and semantics, the web would be whole new thing for professionals. In this milieu it would be in the fitness of things to understand user requirements and perceptions and how library professionals can mould it to their advantage. The manipulation and timely application of evolving web would keep us in focus and as such a profession sought after. Therefore the possible future areas would be Web 2.0 services and usability, Web 3.0, role and implications for libraries, application and utility of virtual reality in library and information centres. Further research could be directed on the use of Web 2.0 on academic libraries, which would offer a wider depiction of Web 2.0's acceptance at a whole.

CONCLUSION

Web 2.0 provides users with a plethora of opportunities for social interaction and collaboration. In a very large extent, Web 2.0 has greatly modified the way people using the Internet. People use web 2.0 applications to pool their knowledge. Web 2.0 not only emphasizes effectiveness and usefulness of its applications, but a major focus is on the social utility and entertainment. Web 2.0 provides an interface to users in terms of social interaction such as Facebook, entertainment such as YouTube and communication tools such as Blog. That is why Web 2.0 becomes more and more popular

in the Internet. Libraries form an essential part of academic institutions. Their role transcends the mere function of depositories of published information sources. Academic libraries have traditionally enabled and facilitated the exchange and growth of information, knowledge and culture among teachers, students and the general public. In this sense libraries represent a focal point of academic life and as such serve also a societal purpose of bringing together people around common themes. This purpose is nowadays enhanced and facilitated by the use of technology and, in recent times, by the so-called Web 2.0 (Maness, 2006). Although libraries select to "wait and see", and thus accept Web 2.0 somewhat sluggishly or even unenthusiastically, it is vibrant that "Library 2.0" is not a transitory trend. Libraries are indicting these applications and features to do their job, but in innovative ways: meeting patrons' needs. Thus, this service-based, and is attractive and instrumental to libraries with users that live are at excessive topographical distance from the library itself. A contemporary academic librarians are re-imagining library collection development. Instead of purchasing books, librarians are leveraging Web 2.0 applications for knowledge production and management by facilitating collaborative open access e-publication alternatives to textbooks for student and research use.

REFERENCES

- Alexander, B. (2006). Web2.0: A new way of innovation for teaching and learning. *EDUCASE Review*, 41(2), 32-44. Available at <http://www.educause.edu/ero/article>
- Anderson. (2007). *What is Web 2.0?: Ideas, Technologies and Implications for Education*. Available at <http://www.jisc.ac.uk/media/documents/techwatch/tsW070/co/oen.odt>
- Barnatt, C. (2012). *Web 2.0*. Available at <http://explainingcomputers.com/web2.html>



- Birdsall, W. F. (2007). Web 2.0 as a social movement. *Webology*, 4(2). Available at: <http://www.webology.ir/2007/v4n2/a40.html>
- Boateng, F., & Quan Liu, Y. (2014). Web 2.0 applications usage and trends in top US academic libraries. *Library Hi Tech*, 32(1), 120–138. doi:10.1108/LHT-07-2013-0093
- Cohen, S. M. (2005). Wiki while you work. *Public Libraries*, 44(4), 208–209.
- Davis, I. (2005). *Library 2.0 – the next wave of the field*. Available at <http://internetalchemy.org/2005/07/talis-web-20-and-all-that>
- Doctorow, C. (2002). *Essential blogging*. Available at http://books.google.co.in/books/about/Essential_blogging.html?id=ffdSAAAAMAAJ&redir_esc=y
- Dye, J. (2007). Collaboration 2.0: Make the Web your workspace. *Econtent*, 30(1), 32–36. Available at <http://www.econtentmag.com/Articles/Editorial/Feature/Collaboration-20-Make-The-Web-Your-Workspace-18978.htm>
- Ebersbach, A., & Glaser, M. R. (2006). *Wiki: Web Collaboration*. Springer-Verlag.
- Felix, L., & Stolarz, D. (2006). *Hands-On Guide to Video Blogging and Podcasting: Emerging Media Tools for Business Communication*. Focal Press.
- Frumkin, J. (2005). The wiki and the digital library. *OCLC Systems & Services: International Digital Library Perspectives*, 21(1), 18–22. DOI: 10.1108/10650750510578109
- Han, Z., & Liu, Y. (2010). Web 2.0 applications in top Chinese university libraries. *Library Hi Tech*, 28(1), 41–62. doi:10.1108/07378831011026689
- Hossain, M. M., & Aydin, H. (2011). Web 2.0 Technologies in Teaching-Learning Multiculturalism: Information Technology Based Higher. *Education + Training*, 355–362. Available at http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=5480096&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxppls%2Fabs_all.jsp%3Farnumber%3D5480096
- Houghton-John, S. (2005). *Library 2.0 discussion: Michael squared*. Librarian in Black (blog). Available at http://librarianinblack.typepad.com/librarianinblack/2005/12/library_20_disc.html
- Husid, W. (2010). Criteria for evaluating Web 2.0 and online technologies for use in the curriculum. *Current Studies in Librarianship*, 30(1/2), 27–36. Available at <http://connection.ebscohost.com/c/articles/66930166/criteria-evaluating-web-2-0-online-technologies-use-curriculum>
- Jhonson, D. (2005). *AJAX: Dawn of a new developer: The latest tools and technologies for AJAX developers*. Available at http://www.javaworld.com/javaworld/juv-10-2005/jw_1017_ajax
- Kille, A. (2006). Wikis in the workplace: how wikis can help manage knowledge in library reference services. *Libres*, 15(2). Available at: https://www.ischool.utexas.edu/~i385q/archive/kille_wikis_reference_services.pdf
- Kittraci, A. (2009). *All About Computers: Conquering Technophobia Web 2.0. Explained*. Available at http://www.kittraci.wikispaces.com/file/view/web_2.0.pdf
- Linh, N. (2008). A survey of the application of Web 2.0 in Australasian university libraries. *Library Hi Tech*, 26(4), 630–653. doi:10.1108/07378830810920950
- Maness, J. M. (2006). Library 2.0 theory: Web 2.0 and its implications for libraries. *Webology*, 3(2). Available at <http://www.webology.ir/2006/v3n2/a25.html>
- Murugesan, S. (2007). *Understanding Web 2.0*. The IEEE Computer Society. Available at http://91-592-722.wiki.uml.edu/file/view/understanding_web_20.pdf
- O'Reilly, T. (2005). *What is Web 2.0. Design patterns and business models for the next generation of software*. Available at <http://oreilly.com/web2/archive/what-isweb-20.htm>

Richardson, W. (2006). Blogs, Wikis, Podcasts and other powerful resources for classrooms. Thousand Oaks, CA: Crown Press. Available at http://apsportal.nl/_userdata/File/documenten/Microsoft%20Word%20%20Samenvatting%20Blogs%20wikis%20podcasts.pdf

Rogers, K. (2015). Academic and Public Libraries' Use of Web 2.0 Applications and Services in Mississippi. *SLIS Connecting*, 4(1). doi:10.18785/slis.0401.08

Si, L., Shi, R., & Chen, B. (2011). An investigation and analysis of the application of Web 2.0 in Chinese university libraries. *The Electronic Library*, 29(5), 651 – 668. DOI: 10.1108/117514971111142655

Stiefvater, E. (n.d.). *Youth Impact: A learners Resource*. Available at http://www.nationalservices.org/files/youth_imact_vol2.pdfUIS

Van der Vlist, E. (2007). *Professional Web 2.0 Programming*. Wrox/Wileypublishing.

Varnum, K. (2006). *RSS4Lib: Innovative ways libraries use RSS*. Available at <http://blogs.fletcher.tufts.edu/rss4lib/>

Wagner, C., & Bolloju, N. (2004). Supporting knowledge management in organizations with conversational technologies: Discussion forums, weblogs, and wikis. *Journal of Database Management*, 16(2), 1–8. Available at <http://www.citeulike.org/group/1990/article/821708>

Walia, P., & Gupta, M. (2012). Application of Web 2.0 tools by national libraries. *Webology*, 9(2), 13–13. Available at www.webology.org/2012/v9n2/a99.html

Web 2.0. (2013). Available at [http://uis.edu/colrs/learning/technologies/web 2.0.html](http://uis.edu/colrs/learning/technologies/web%202.0.html)

Wikis, Blogs & Web 2.0 Technology. (2008). Available at <http://www.unimelb.edu.au/copyright/information/guides/wikisblogsweb2blue.pdf>

Wright, S. (2013). *New Communication behaviours in a web 2.0 world changes, Challenges and opportunities in the era of the Information Revolution*. Available at http://enterprise.alcatel-lacent.com/private/active_docs/communication%20Behaviour%20In%20Web%200%20World_ALU.pdf

ADDITIONAL READING

Boateng, F., & Quan Liu, Y. (2014). Web 2.0 applications usage and trends in top US academic libraries. *Library Hi Tech*, 32(1), 120–138. doi:10.1108/LHT-07-2013-0093

Ceri, S. (2003). *Designing data-intensive Web applications*. San Francisco, Calif: Morgan Kaufmann.

Crawford, W. (2014). *Successful social networking in public libraries*. Chicago: ALA Editions, an imprint of the American Library Association.

Crowther, J. H., & Trott, B. (2004). *Partnering with purpose: A guide to strategic partnership development for libraries and other organizations*. Westport, Conn: Libraries Unlimited.

Farkas, M. G. (2007). *Social software in libraries: Building collaboration, communication, and community Online*. Medford, N.J: Information Today.

Hastings, R. (2010). *Microblogging and livestreaming in libraries*. London: Facet Pub.

Matthews, G., & Walton, G. (2012). *University libraries and space in the digital world*. Farnham, Surrey, England: Ashgate.

Morris, A. (1992). *The application of expert systems in libraries and information centres*. London: Bowker-Saur. doi:10.1515/9783110977806

Parkes, D., & Walton, G. (2010). *Web 2.0 and libraries: Impacts, technologies and trends*. Oxford: Chandos Pub. doi:10.1533/9781780631851

Primary Research Group. (2011). *Survey of use of RFID in libraries*. New York: Primary Research Group.

Riggs, D. E. (1989). *Creativity, innovation, and entrepreneurship in libraries*. New York: Haworth Press.

Sonja, S. M. B. Z. D. K. (2016). Information Literacy in participatory environments. *The Journal of Documentation*, 72(2), 247–264. doi:10.1108/JD-06-2015-0072

Stewart, J., Procter, R., Williams, R., & Poschen, M. (2013). The role of academic publishers in shaping the development of Web 2.0 services for scholarly communication. *New Media & Society*, 15(3), 413–432. doi:10.1177/1461444812465141

Swanson, T. A. (2012). *Managing social media in libraries: Finding collaboration, coordination and focus*. Oxford: Chandos Pub. doi:10.1533/9781780633770

Varnum, K. J., & Kroski, E. (2012). *Drupal in Libraries: (THE TECH SET® #14)*. Chicago: ALA Editions.

Weiner, R. G., Scott, R. W., Nyberg, A. K., Fee, W. T., & Goldsmith, F. (2010). *Graphic novels and comics in libraries and archives: Essays on readers, research, history and cataloging*. Jefferson, N.C: McFarland & Co.

White, H. S. (1992). *Ethical dilemmas in libraries: A collection of case studies*. New York: G.K. Hall.



KEY TERMS AND DEFINITIONS

Hyperlinks: A hyperlink is a reference to the information that a user can simply click and find other source of information.

Library: Collection of sources of information so that it can be made accessible to the defined group or community.

Social Web: Web services that provides a platform to users to remain socially connected.

Web 2.0: Web services that allow users to find, add, and manipulate the content available online.

Category L

Logistics and Supply Chain Management

Barcodes vs. RFID and Its Continued Success in Manufacturing and Services



Amber A. Smith-Ditizio
Texas Woman's University, USA

Alan D. Smith
Robert Morris University, USA

INTRODUCTION

Versatile Nature of the Ubiquitous Barcode

A barcode is a machine-readable code in the form of numbers and a pattern of parallel lines of varying widths, printed on a product (Sirico, 2010). Barcodes are essentially a type of symbolic language that uses printed horizontal strips of vertical bars used for identifying specifications. An accompanying scanning device reads the barcode by moving a beam across the symbol (McCathie, 2004). Barcodes can come in one-dimension (1D), 2-dimension (2D), or three-dimension (3D). Barcode systems are used in routine business from the manufacturing and ordering process to the parking lot after purchased. Joseph Woodland and Bernard Silver invented the barcode in 1949. Silver overheard a conversation of a supermarket executive from a food chain called Food Fair. The general rationale was to have an automatic system for capturing product information (Seideman, 2015).

Barcode technologies have had a relatively long history with almost every type of business that keeps track of goods and services. Point-of-sale (POS) is probably the most common use for the barcode (Chen & Dubinsky, 2003). Customers typically view barcodes on almost every item that is available for purchase within a retail environment. The cashier uses the barcode to scan the items into the computer system and retrieve the amount the customer will owe for that particular

item. Barcodes in manufacturing help in inventory control, quality control, and help measure productivity throughout the manufacturing process. Barcodes are used for packaging, tracking time and attendance of employees, and measurement of proficiency (Jain, Benyoucef, & Deshmukh, 2008; Kamhawi, 2008; Kay, 2007; Kearney, 2005; Kennedy & Widener, 2008; Koong & Lin, 2007). Evaluating the efficiency of barcode systems routinely reduce costs while improving quality, on-time performance, and reducing errors.

BACKGROUND

The future need for barcodes will probably be increasing due to its universal acceptance and easy-to-use with low costs of producing barcodes (e.g., it is usually included in the cost of printing labels or packaging materials on products) (Cowles, Kiecker, & Little, 2002; Davis, 1989; Devaraj, Fan, & Kohli, 2002). As the population grows, so does the need to make workplaces and their environments more efficient and safer. Barcodes are employed in supply chain management (SCM) applications in the healthcare industry. The use of barcodes in healthcare facilities allows hospitals to save space and reduce overstock by ordering the supplies needed daily (Harrop, Das, & Holland, 2016). In the retail setting, barcodes are becoming a part of smartphones with apps. The consumer can shop and scan items through the convenience of their phone. This easily enables

the retailer to track consumers' spending. A majority of manufacturing companies are turning to the barcode system as well. The use of barcodes in manufacturing facilities assists in making the manufacturing process faster and more efficient. However, with barcodes, as with all IT-intensive technologies, have important advantages and disadvantages to consider, as illustrated in Table 1.

Barcodes are extremely easy-to-use and generally require little to no training for users. The time saved by scanning a barcode is substantial to the amount of time that would be spent manually entering product information for one or one's entire inventory. Two-dimensional (2D) barcodes reduce error and read accurately when scanned. Linear or one-dimensional barcodes are limited in their accuracy since the scanner and barcode must be properly aligned for the information to be captured. Barcodes are generally the most cost efficient method of technology for coding information, but associated equipment such as POS or inventory tracking software and scanners, etc., can quickly increase the initial costs, including replacement costs. Security on barcodes is not good, as most information coded in barcodes readily found.

Undoubtedly, there are other issues that need to be generally discussed [e.g., total quality management (TQM), lean manufacturing, just-in-time (JIT) manufacturing, waste management, change management as well as numerous other practices that are comparable to the strategy of adopting AIDC] and how the results of this can

effect whether or not the AIDC-based technologies would be successful. As evident from recent headlines, the need to maintain the privacy and ethics of personally identifiable identifiers (PII), and some companies have been accused of prying and collecting data that were always consented. In a recent study, experts believe that the use of RFID will increase from 12 million in 2011 to billions by 2021 applications Irani, Gunasekaran, & Dwivedi, 2010; Smith & Rupp, 2013; Ustundag, 2010; Visich, Li, Khumawala, & Reyes, 2009). Even if the pros outnumber the cons, it still does not answer the question if management will take the long-term and choose to take advantage of the opportunity. Such technologies have been improving at an exponential rate and many studies (Bhat, 2008; Baxter & Hirschhauser, 2004; Biswas & Sarker, 2008; Browning & Heath, 2009; Cavaleri, 2008; Chan & Kumar, 2009; Devaraj, et al., 2002; Drejer & Riis, 2000; Gaukler, Özer, & Hausman, 2008; Grewal, 2008) have suggested that result of implementing AIDC with lean technologies will have positive outcomes. This is especially associated with the expectation that costs will drop for both the manufacturers and subsequently the consumer.

Advantages of RFID over Barcodes

The general advantages of RFID-related technology only make sense where it provides advantages over barcodes (Green & Khermouch, 2005). Some of the important advantages that RFID has over barcodes include the following (Harrop, 2010; Smith, 2005a, 2008):

1. Speed, since a RFID reader can read tags faster than a typically barcode scanner can scan barcodes. For example, RFID readers designed for supply chain operations can perform up to 1,500 read operations per second;
2. Read distance, as it is common to read RFID tags at distance of at least three meters between the tag and the antenna and the read field is three-dimensional;

Table 1. Selected advantages and disadvantages of barcoding as a part of identification technologies

Advantages	Disadvantages
Inventory Control (through tracking and essentially error-free readings)	Costs (associated equipment and replacement costs)
Accuracy (2D, 3D Barcodes), Error reduction	Accuracy (especially with linear barcode)
Time savings	Limitations by line of sight scanning
Easy-to-use	Security issues, as they are easily read

3. Simultaneous as opposed to sequential scanning, since RFID readers can identify multiple tags within the read field;
4. Non-line of sight, as radio waves can penetrate most materials depending on the frequency utilized;
5. Durability, since RFID tags can operate in extremely rugged working conditions and can be encased in plastic or even embedded directly into the products manufactured.

However, barcodes can be easily printed or applied directly on the products it is designed to identify as extreme accurate read rates and very low costs.

Inventory and Operational Decisions and Cost Analysis

Inventory-related challenges include ensuring stock arrives at the right place in production at the right time, that the correct quantity is delivered as promised, and inventory shrinkage is reduced. By enabling operations managers to automatically and accurately track the flow of crucial materials through the facility, RFID tags can streamline processes and reduce costs, leading to a leaner organization (Hu, Wang, Fetch, & Bidanda, 2008; Jain, Benyoucef, & Deshmukh, 2008; Kennedy & Widener, 2008; Scherrer-Rathje, Boyle, & Deflorin, 2009; Smith & Offodile, 2007, 2008, 2009).

In order to be successful with utilizing RFID technology, a manufacturing company will need to take a risk and jump at the opportunity before the competitive advantage is no longer available. Once this happens, efficiency and effectiveness will remain the result of the implementation of the product, however, there will not be an opportunity to be ahead of competition, only to be at the equivalent level of production as them. When analyzing the efficiency affects the AIDC-related technologies would have on a production process in the manufacturing industry, there is one main goal: improving the supply chain and hopefully

reorganize the structure of the company involved. The end result will be lowered costs and eliminate any unnecessary products or processes involved in production. The value of such technologies is that it adds to a company in its ability to achieve these levels of efficiencies.



CASE STUDIES OF BARCODES IN USE

Wal-Mart Applications

A well-known retailer that relies heavily on the use of barcodes is Wal-Mart. Wal-Mart uses an integrated barcode system in conjunction with a JIT inventory system to aid in their inventory and tracking management (Burgin, personal note, 2015). Wal-Mart uses barcodes and RFID tags in 937 Stores, 2,527 supercenter, 138 neighborhood markets, 129 distribution centers, and 566 Sam's Club stores. In a typical year, Wal-Mart will spend US\$6.4 million in readers to scan barcodes and RFID's in order to keep inventory and employees updated in their computer systems at all times (Chen, 2015).

During an interview with a Wal-Mart Assistant Manager at Clarksville, TN, Burgin explained the route a barcode takes from warehouse or vendor to the consumer's home. The JIT-based inventory system starts at the checkout of a product by a consumer. Suppose a customer purchases a bottle of Tide laundry soap. When the bar code is swiped across the register, the inventory system automatically does several functions. First, it makes a list to be pulled from the back room for the midnight shift; this is called a pic list. Then, the computer generates an order list based on what items sold that day. If there were 5 Tide laundry soaps sold today then the system would tell the midnight associate to pick 5 Tide laundry soaps to restock the floor. When the bin in the back is empty, the computer automatically reorders the allotted standby amount. The JIT inventory system

reduces the amount of space needed to store extra product. It allows the store to receive 3 inventory trucks a week. This keeps the shelves fully stocked.

Once a truck arrives at Wal-Mart, the receiving manager will either accept or decline the truck. When the truck is accepted, the inventory of what is on hand in the store will automatically update as it passes through scanners at the back door. This process saves time, and eliminated two jobs when installed. The scanners are able to read each barcode as it passes through the backdoor and updates the inventory. The scanner at the back door and the high frequency tags that Wal-Mart uses are a little costlier than regular barcodes. The rewards from the use of the scanner and high frequency tags outweigh the initial costs of the new system. Once a truck is accepted and product is in the store, the store employees move all the needed stock to the floor to be processed and stored. After the inventory is stored, the customers have the opportunity to pick it up and purchase it. At the cash register, the barcode is scanned again to record the sale of each item for the consumer and the retailer. The consumer gets the barcode put in to numeric form and on a receipt. The retailer after purchase uses the barcode to allow the inventory system to update the on-hand in store amount and order more if the amount left in the store is below the shelf cap.

There is a label on every shelf in Wal-Mart. Each label has a barcode. When an associate scans the barcode it tells them how many of each product go on each shelf, how many rows, how many in each row, and how many are in the store. Employees can use a Telxon gun to see where each item in the store goes and gives the employee shelf number and quantity in the store. This function of the barcode helps keep Wal-Mart uniform and keeps the associates on track for stocking. At the end of every shelf, there is a barcode. This barcode is scanned by department managers to see how many of each shelf is either on the floor in use or in the back stock room. Having the ability to know how many shelves a store has prevents the store from overspending on hardware for the store. Wal-

Mart uses barcodes for the independent vendors. Each vendor has their own unique barcode for their use. When a vendor restocks their endcaps and shelves the vendor must put the amount of inventory added to the store manually. This allows enhanced accountability to be on the vendor side.

The barcode system is used on soft lines and grocery departments, as well in the pharmacy and deli departments. Wal-Mart has a barcode for every department in their store. The pharmacy's JIT inventory system works similar to the rest of the store. When a prescription is filled the back computer automatically places an order to replace the consumed medicine for that order. The only difference is when pharmacy trucks are accepted at the store. The computer does not automatically update when entering the store. A pharmacy manager must manually input all orders when receiving them to ensure accuracy, and sign for the medicine.

The JIT inventory system has many great advantages. It is able to help save money and labor in their stores. The system gives Wal-Mart the ability to know what is on hand at all times. This reduces waste and storage cost. A down side of the just-in-time inventory system is that it does not take into account for theft. To account for this downfall, Wal-Mart takes a hard count of their entire inventory every three months to reconcile the inventory and starts the counts over. Another down side to the just-in-time inventory system is that it does not take into account for promotions or sales. Managers must look forward 6 weeks at a time to manually override the products they know will be going on sale (Burgin, personal note, 2015).

Wal-Mart generally prints barcodes on the bottom of receipts given to customers at the end of the checkout process. Wal-Mart recently added a value service to help customers save money called Savings Catcher® app. After check out, the customer can scan the barcode at the bottom of the receipt and the app will search to see if any of the items the customer purchased are cheaper at any other stores. If the company finds an item

is cheaper at a different store they will put the difference on an eGift card for the customer to use.

Wal-Mart does not use barcodes for inventory of merchandise only. Wal-Mart uses barcodes on their associates name badges too. Employees clock in and out with their assigned barcodes and find out what are their assigned daily tasks. Wal-Mart is able to track the movement of an employee from the barcode on their nametag. The barcode on the nametag stores all the payroll information, assignments, and length of time spend on each register. The use of the barcode system for employees reduced the HR department from a staff of 6 to 10 per store to 2 per store; a substantial cost savings motive for the company (Burgin, personal note, 2015).

Stone, Rudolph and Henry (SRH)

Another company, quite smaller than Wal-Mart, is Stone, Rudolph & Henry PLC; a medium-sized local accounting firm located in Clarksville, TN that uses barcodes in the day-to-day internal operations of the company. SRH uses barcodes to route documents using specific software by the name document. Documentation completed through purchased software that acts as a paperless storage and organizational management system. This feature allows all documents and files saved in specific folders distinguished by categories, subcategories, year, etc. The company has 4 commercial copier machines that are capable of printing, scanning, faxing, etc. These machines are refurbished and all repair/maintenance work is outsourced to Ricoh. SRH has a small technology department that is responsible for handling all of the internal operations of the business as well as providing external support to clients and

other businesses. The machines are connected to the local network and use a SMB service. This service allows for specific folders or paths to be programmed into the machine to route documents to specific, highly used, locations. This reduces the time spent scanning documents to a single location, then having to find the document, rename it and move it (Buse, personal note, 2015).

From the perspective of using the barcode routing sheet, the process is very straightforward. One gathers what needs to be scanned, puts the routing sheet on top of it, and feeds it into the tray on the copier machine. From the perspective of creating the routing sheet and knowing how it works is entirely more intricate. First, all routing sheets are created individually. One must select a client, choose to add a document, and fill out all the necessary fields. This includes creating the file name; once this is done, a routing sheet is generated. This sheet has a barcode on the top of the page and specifies all of the information used to create it, including the name and number associated with the client. The barcode numbers are five or six digits long and are randomly generated, in no particular order. Figure 1 illustrates this basic information flow using workflows of product routing sheets.

Once ready to scan and inserted onto the tray on the copy machine, the user must select a destination. In this case, the user would select the document button and press start. The document is scanned and starts the chain of reactions to process it to its final destination, which is coded in the barcode. The first place the document routed is to the scan server. Then optical character recognition (OCR) software executed for text recognition and or search. This takes the barcode, deletes the top routing sheet, changes the file name, currently in

Figure 1. Typical workflow path of the routing sheet using barcodes



a 10-20-digit number coded in the barcode, and renames the PDF as the barcode number. Next, the PDF is dropped in the export folder. CCH App Server and Document Watcher software then checks the document software to find the corresponding barcode number. This is the documents intended destination. The software applies the appropriate settings and specifications to the PDF, and transfers it to where it belongs. This process saves the company a great deal of time and money while allowing the staff to continue their work, uninterrupted. Not only are barcodes being used in industries across the country, they are being utilized across the globe.

CONTINUED SUCCESS OF BARCODES

Barcodes for European IDs

Europe has used barcodes on the back of personal identification (ID) cards for some, which is a requirement for most citizens depending on their country of residence. A person's basic information that can easily be found on the front of their identification card such as their home address, eye color, and weight among other personal factors are coded in the barcode. The type of barcode used is easily accessible by the simplest of scanners making it obtainable by anyone in the industries allowed to access this information. These industries include airport personnel, cab drivers, shops, entertainment and bars, where people live, and other places people spend money. Many identification cards have either barcodes or mag strips, while some contain both. Both of these methods are capable of coding information so that the proper authority can access it ("National ID cards," 2015). However, there are differences in size of information, cost, and security that should be taken into account when deciding which technology to use. Barcodes generally cost between US\$0.01-0.10 per barcode, while mag strips can range from US\$0.01-0.40. In terms of security, barcodes range from poor to

moderate while mag strips range from poor good. Therefore, depending on what type of information the customer are coding, whether it is common data or more personal information one would like to keep private, one should choose the technology best suited for their needs.

Barcodes for Domestic Identification Purposes

In the U.S., barcodes are typically used on selected citizen's identification cards. These cards, more commonly are referred to as driver's licenses. However, both do exist. The information coded in the barcode is similar to that of European countries including: the individual's full name, address, date of birth, physical attributes, medical impairments and donor information ("About the 2D barcode ...," 2015). Private information that is not included on the front of either card, such as social security number (SSN), face recognition template or digital fingerprint are included in the barcode. Since identification cards are not used the same in the United States as they are in Europe, less industries have access to scan the barcode, but the barcode is still as easy to decode. At least 39 states use barcodes on their IDs. The databases are independent and not shared, which is a gap in the system. There is speculation of using smart chips in identification cards in the future ("The future of RFID," 2016). This would include all of the coded information mentioned previously in addition to additional biometric identifiers. In order for this to be effective, all states would need to share a common database to ensure accurate and up-to-date information (Sammer, 2011).

Barcodes in Healthcare

Having proper healthcare is essential to human life. The simplest of mistakes made in a healthcare setting, such as a hospital, can cause someone a lot of pain and suffering. Sometimes a minor mistake, whether it is due to inaccurate labeling or performing the wrong surgery, can lead to patient

deaths. With the amount of patient safety challenges facing modern healthcare it is difficult for the industry to manage information with AIDC-embedded technologies. A constant struggle for the healthcare industry is tracking materials, inventory obstacles, and patient validation throughout day-to-day operations (Swaminathan, Chernew, & Scanlon, 2008). The use of barcodes provides the healthcare industry with an easy-to-use and cost effective solution to many of these issues (“About the 2D barcode ...,” 2015).

Barcoding can be used as much or as little as desired. The implementation of barcodes would be most beneficial to the healthcare industry if applied at a high level. Applying barcodes to multiple items within a facility will allow for the most accuracy. Barcodes can be applied to basically anything that can have a sticker attached to it. Some examples in a healthcare facility include: medication, vials (blood samples), patient wristbands at admittance, urine samples, all documents in patient charts, surgical instruments, personal belongings, and much more. Having these barcodes allows doctors, nurses, and other authorized personnel the ability to easily access information, as well as add new information about the patient or item into the system. Coupling this simple technology with RFID can have substantial benefits in the healthcare field (Seidman, Brockman, & Lewis, 2010).

Barcode wristbands, which include precise patient information, are usually generated at the point of admission and updated consistently based on the needs of the patient. Healthcare providers are able to use a mobile printer to create a barcode wristband at the patient’s bedside, which allows medication administered in a timely fashion by scanning the patient’s wristband and instantly record the patient’s medical record (Torres, 2012). Barcode technology provides certainty that the accurate treatment administered to the correct patient. Barcodes are used for workflow changes in hospitals and other healthcare practices. For instance, if a patient is transferred to another wing or department, once the barcode is scanned, the system is automatically updated. Then, when it

is time to administer medicine or for the doctor to visit, he/she can scan the barcode associated with the chart and is directed to exactly where the patient is. This reduces confusion and chaos that is brought on by the lack of communication that transpires between staff. Barcode technology allows practitioners to see all previous notations and medical information tethered to the individual and unique barcode, which is linked to each patient. Through this, staff can easily access records and see where a patient is in terms the medication cycle (Torres, 2012).

The use of barcoding in the healthcare profession has many benefits. Some of these benefits include inventory control, patient validation, and tracking materials. These benefits can be directly linked to patients, records, and overall safety. Barcodes are a key tactic to guarantee proper patient identification at the bedside (Degaspari, 2011). It is extremely important that such technology validate that the patient being treated is the right patient with the correct dosage of medication. When nurses scan a barcode, they can confirm that the item they are about to give or use on a patient is the appropriate item ordered by the doctor. They can view if they used the right item with the correct patient by scanning the barcode on the patient’s wristband and then scanning the barcode on the item. Barcode readers support the receiving and storage functions, as counting and reordering automatic, and an information system within the machine performs the calculations (Burbano, Saka, Rardin, & Rossetti, 2009).

Every healthcare facility must deal with the time-consuming process of controlling inventory. Using barcodes can speed up, as well as manage; this process is designed to automatically reorder materials when their inventory reaches a certain level. Barcode scanning systems can be useful for inventories no matter where the location and can help assure that the right supplies are available when needed. Materials are constantly traveling through a healthcare facility, from one point to the next, and it is easy to lose track of an item when it is constantly on the move. For this reason, along

it is essential to have a barcode system in these facilities. Barcodes can track an item to a specific patient and identify the physician who used it with or on the patient.

Like any other profession, healthcare services can suffer from operational problems; and, medical errors cost both lives and money (Southard, 2012). Barcode technology helps reduce these errors while not damaging the doctor/patient relationship. According to Degaspari (2011), barcodes and related technologies are not intended to replace human interaction with the patient, but they are redundant identifiers that are more fail-safe than many other approaches.

As previously mentioned, barcode scanners are used for medication inventory management and tracking, as well as for patient/drug identification/verification when administering medications at the point of care to prevent errors and increase patient safety. Anti-microbial products are particularly important in these activities. Scanners are useful for tracking bar-coded medical equipment inventories (McBride, 2012).

Advancement in technology reduces the chance for error, but as stated previously, errors cannot be completely eliminated. 2-dimensional barcodes can withstand a higher degree of damage before they are rendered useless. The question whether 2-dimensional barcodes are reliable or not, seems to be in the favor of it being reliable. The software needed to scan the barcode can be downloaded on smart phones, Android cell phones and the Apple iPhone.

The future of 2-dimensional barcoding in healthcare is very promising. The healthcare industry has already caught on to the benefits it would provide. Many hospitals are transitioning from linear to 2-dimensional barcode technology. The barcode historically has had success in other industries; many predict it will have a similar outcome in the healthcare industry. Many, doctors, nurses and patients alike, welcome the change it would provide though it is not possible to completely eliminate all dangers in the healthcare field. Two-dimensional barcoding will help prevent many mishaps and solve some

major problems. Its major competitor would be RFID, but as explained earlier in this chapter, 2-dimensional barcoding has the advantage since it is cheaper than RFID and better fit for tasks in which an object is moving and the person is not.

CONCLUSION

Barcodes are machine-readable codes; usually placed on a product, that have been around for over 50 years. Barcodes can be used in a variety of industries for an unlimited number of uses. Generally, barcodes exist in a linear or 2-dimensional format. They are the best choice of coding information in terms of costs, but not security. Barcodes have the ability to improve any process, and can be an asset to the largest and smallest of businesses. Barcodes are making more information more easily accessible, as discussed in connection with identification cards. Lastly, the use of two-dimensional barcodes in healthcare will be very beneficial and cost effective for the future of the industry. In hospitals especially, the benefits barcodes provide could possibly save a patient's life.

FUTURE RESEARCH DIRECTIONS

Consequentially, the best option a manufacturing company can choose is to implement AIDC technologies in combination with other strategies such as total quality management (TQM) and barcodes as a backup technology to reduce errors. The efficiency and effectiveness are good goals for using any intensive AIDC-intense technology when developing any new ideas for a production process. The authors made this an important focus of the study, which was very important. Then the two factors were linked to manufacturing processes such as supply chain management, which proved to aide in the research. The implementation of such technologies ultimately is still negotiable, but there are universal advantages of both types of identification, with barcodes well entrenched.

REFERENCES

- About the 2D barcode and your driver's license. (2015). Retrieved April 19, 2015 from http://archive.turbulence.org/Works/swipe/barcode_faq.html
- Baxter, L. F., & Hirschhauser, C. (2004). Reification and representation in the implementation of quality improvement programmes. *International Journal of Operations & Production Management*, 24(2), 207–224. doi:10.1108/01443570410514894
- Bhat, S. (2008). The effect of ordering policies for a manufacturing cell changing to lean production. *Proceedings - Institution of Mechanical Engineers*, 222(B11), 1551–1560. doi:10.1243/09544054JEM1216
- Biswas, P., & Sarker, B. R. (2008). Optimal batch quantity models for a lean production system with in-cycle rework and scrap. *International Journal of Production Research*, 46(23), 6585–6610. doi:10.1080/00207540802230330
- Browning, T. R., & Heath, R. D. (2009). Reconceptualizing the effects of lean on production costs with evidence from the F-22 program. *Journal of Operations Management*, 27(1), 23–35. doi:10.1016/j.jom.2008.03.009
- Burbano, A., Saka, B., Rardin, R., & Rossetti, M. (2009). Technology assessment for an inventory management process in a hospital unit. *IIE Annual Conference. Proceedings*, 791–796. Retrieved April 17, 2015 from <http://search.proquest.com.ezproxy.lib.apsu.edu/docview/192455998?accountid=8437>
- Cavaleri, S. A. (2008). Are learning organizations pragmatic? *The Learning Organization*, 15(6), 474–481. doi:10.1108/09696470810907383
- Chan, F. T. S., & Kumar, V. (2009). Performance optimization of a legality inspired supply chain model: A CFGTSA algorithm based approach. *International Journal of Production Research*, 47(3), 777–791. doi:10.1080/00207540600844068
- Chen, Y. (2015). *The Wal-Mart RFID Initiative*. Retrieved April 20, 2015 from <http://www.winmec.ucla.edu/rfid/course/2008s/The%20Walmart%20Initiative.pdf>
- Chen, Z., & Dubinsky, A. J. (2003). A conceptual model of perceived customer value in e-commerce: A preliminary investigation. *Psychology and Marketing*, 20(4), 323–347. doi:10.1002/mar.10076
- Collins, J. (2005). School studies RFID's effect on Wal-Mart. *RFID Journal*. Retrieved February 22, 2016, from <http://www.rfidjournal.com/article/articleview/1514/1/9/>
- Collins, J. (2006). RFID Implementation is an art. *RFID Journal*. Retrieved February 18, 2016, from <http://www.rfidjournal.com/article/articleview/2427/1/1/>
- Cowles, D. L., Kiecker, P., & Little, M. W. (2002). Using key informant insights as a foundation for e-retailing theory development. *Journal of Business Research*, 55(8), 629–636. doi:10.1016/S0148-2963(00)00203-4
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use and user acceptance of information technology. *Management Information Systems Quarterly*, 13(3), 319–340. doi:10.2307/249008
- Degaspari, J. (2011). Keeping track: Barcodes and RFID tags make inroads in hospitals. *Healthcare Informatics*, 28(3), 44–47. PMID:21446265
- Devaraj, S., Fan, M., & Kohli, R. (2002). Antecedents of B2C channel satisfaction and preference: Validating e-commerce metrics. *Information Systems Research*, 13(3), 316–333. doi:10.1287/isre.13.3.316.77
- Drejer, A., & Riis, J.O. (2000). New dimensions of competence development in industrial enterprises. *International Journal of Manufacturing Technology and Management*, 2(1/7), 660–882.

- Gaukler, G. M., Özer, O., & Hausman, W. H. (2008). Order progress information: Improved dynamic emergency ordering policies. *Production and Operations Management*, 17(6), 599–614. doi:10.3401/poms.1080.0066
- Green, H., & Khermouch, G. (2005). Bar codes better watch their backs. *BusinessWeek*. Retrieved January 8, 2016, from http://www.businessweek.com/magazine/content/03_28/b3841063.htm
- Grewal, C. (2008). An initiative to implement lean manufacturing using value stream mapping in a small company. *International Journal of Manufacturing Technology and Management*, 15(3-4), 404–421. doi:10.1504/IJMTM.2008.020176
- Harrop, P. (2010). *HF RFID-the great leap forward*. Retrieved February 4, 2010, from <http://www.rfidwizards.com.html>
- Harrop, P., Das, R., & Holland, G. (2016). *RFID in healthcare: 2016-2019*. Retrieved February 4, 2010, from http://www.idtechex.com/research/reports/rfid_for_healthcare_and_pharmaceuticals_2016_2019_000146.asp
- Hu, G., Wang, L., Fetch, S., & Bidanda, B. (2008). A multi-objective model for project portfolio selection to implement lean and Six Sigma concepts. *International Journal of Production Research*, 46(23), 6611–6648. doi:10.1080/00207540802230363
- Irani, Z., Gunasekaran, A., & Dwivedi, Y. K. (2010). Radio frequency identification (RFID): Research trends and framework. *International Journal of Production Research*, 48(9), 2485–2511. doi:10.1080/00207540903564900
- Jain, V., Benyoucef, L., & Deshmukh, S. G. (2008). Whats the buzz about moving from lean to agile integrated supply chains? A fuzzy intelligent agent-based approach. *International Journal of Production Research*, 46(23), 6649–6678. doi:10.1080/00207540802230462
- Kamhawi, E. M. (2008). System characteristics, perceived benefits, individual differences and use intentions: A survey of decision support tools of ERP systems. *Information Resources Management Journal*, 21(4), 66–83. doi:10.4018/irmj.2008100104
- Kay, M. (2007). Closing the RFID knowledge gap in apparel and footwear. *RFID Journal*. Retrieved February 18, 2016, from <http://www.rfidjournal.com/article/articleview/3267/1/82/>
- Kearney, A. T. (2005). *RFID will bring great benefits for retailers*. Retrieved February 24, 2016, from http://retailindustry.about.com/cs/it_rfid/a/bl_atk111003.htm
- Kennedy, F. A., & Widener, S. K. (2008). A control framework: Insights from evidence on lean accounting. *Management Accounting Research*, 19(4), 301–319. doi:10.1016/j.mar.2008.01.001
- Koong, L., & Lin, C. (2007). Evaluating the decision to adopt RFID systems using Analytic Hierarchy Process. *Journal of American Academy of Business*, 11(1), 72–77.
- McBride, M. (2012). Healthcare gadgets go high-tech. *Medical Economics*, 89(23), 52–54, 56–58. PMID:23516910
- McCathie, L. (2004). *The advantages and disadvantages of barcodes and radio frequency identification in supply chain management*. Retrieved April 21, 2015 from <http://www.ro.uow.edu.au/cgi/viewcontent.cgi?article=1009&context=thesisinfo>
- National ID cards. (2015). Retrieved April 19, 2015 from <http://www.barcodesinc.com/articles/national-id-cards.htm>
- Sammer, J. (2011). Swipe cards reduce errors but cost payers more to produce. *Managed Healthcare Executive*, 21(1), 5.



- Scherrer-Rathje, M., Boyle, T. A., & Deflorin, P. (2016). Lean, take two! Reflections from the second attempt at lean implementation. *Business Horizons*, 52(1), 79–85. doi:10.1016/j.bushor.2008.08.004
- Seideman, T. (2015). *Barcodes sweep the world*. Retrieved April 22, 2015 from http://www.barcoding.com/information/barcode_history.shtml
- Seidman, S. J., Brockman, R., Lewis, B. M., Guag, J., Shein, M. J., Clement, W. J., & Huntwork, D. et al. (2010). In vitro tests reveal sample radio frequency identification readers inducing clinically significant electromagnetic interference to implantable pacemakers and implantable cardioverter-defibrillators. *Heart Rhythm: The Official Journal of the Heart Rhythm Society*, 7(1), 99–107. doi:10.1016/j.hrthm.2009.09.071 PMID:20129290
- Sirico, L. (2010). *An RFID primer for manufacturers*. Retrieved February 4, 2010, from <http://www.rfidwizards.com.html>
- Smith, A. D. (2005a). Reverse logistics and their affects on CRM and online behavior. *VINE: The Journal of Information and Knowledge Management*, 35(3), 166–181. doi:10.1108/03055720510634216
- Smith, A. D. (2005b). Exploring the inherent benefits of RFID and automated self-serve Checkouts in a B2C environment. *International Journal of Business Information Systems*, 1(1-2), 149–183. doi:10.1504/IJBIS.2005.007405
- Smith, A. D. (2008). Evolution and acceptability of medical applications of RFID implants among early users of technology. *Health Marketing Quarterly*, 24(1-2), 121–155. doi:10.1080/07359680802125980 PMID:19042524
- Smith, A. D., & Offodile, O. F. (2007). Exploring forecasting and project management characteristics of supply chain management. *International Journal of Logistics and Supply Management*, 3(2), 174–214.
- Smith, A. D., & Offodile, O. F. (2008). Gauging web portals impact on supply chain management concepts. *International Journal of Logistics and Management*, 4(2), 184–206.
- Smith, A. D., & Offodile, O. F. (2009). The perceived importance of major RFID-related technology initiatives among retail store managers. *International Journal of Services and Operations Management*, 5(4), 520–547. doi:10.1504/IJSOM.2009.024583
- Smith, A. D., & Rupp, W. T. (2013). Data quality and knowledge/information management in service operations management: Regional supermarket case study. *International Journal of Knowledge-Based Organizations*, 3(3), 35–52. doi:10.4018/ijkb.2013070103
- Southard, P. B., Chandra, C., & Kumar, S. (2012). RFID in healthcare: A six sigma DMAIC and simulation case study. *International Journal of Health Care Quality Assurance*, 25(4), 291–321. doi:10.1108/09526861211221491 PMID:22755482
- Swaminathan, S., Chernew, M., & Scanlon, D. P. (2008). Persistence of HMO performance measures. *Health Services Research*, 43(6), 2033–2051. doi:10.1111/j.1475-6773.2008.00890.x PMID:18783460
- The future of RFID. (2016). *RFID Journal*. Retrieved January 8, 2016, from <http://www.rfidjournal.com/article/view/7268>
- Ustundag, A. (2010). Evaluating RFID investment on a supply chain using tagging cost sharing factor. *International Journal of Production Research*, 48(9), 2549–2562. doi:10.1080/00207540903564926
- Visich, J. K., Li, S., Khumawala, B. M., & Reyes, P. M. (2009). Empirical evidence of RFID impacts on supply chain performance. *International Journal of Operations & Production Management*, 29(12), 1290–1315. doi:10.1108/01443570911006009

ADDITIONAL READING

Okoniewska, B., Graham, A., Gavrilova, M., Wah, D., Gilgen, J., Coke, J., & Ghali, W. A. et al. (2010). Multidimensional evaluation of a radio frequency identification Wi-Fi location tracking system in an acute-care hospital setting. *Journal of the American Medical Informatics Association*, 19(4), 674–679. doi:10.1136/ami-ajnl-2011-000560 PMID:22298566

Seidman, S. J., Brockman, R., Lewis, B. M., Guag, J., Shein, M. J., Clement, W. J., & Huntwork, D. et al. (2010). In vitro tests reveal sample radiofrequency identification readers inducing clinically significant electromagnetic interference to implantable pacemakers and implantable cardioverter-defibrillators. *Heart Rhythm*, 7(1), 99–107. doi:10.1016/j.hrthm.2009.09.071 PMID:20129290

Taylor, N. P. (2013). Study: Use of 2D barcodes could save healthcare \$300M. Newton: Questex Media Group LLC. Retrieved April 18, 2015 from <http://search.proquest.com.ezproxy.lib.apsu.edu/docview/1467652358?accountid=8437>

Torres, A. (2012). The role of barcode technology in patient safety and identification. [Retrieved April 17, 2015 from <http://www.healthmgttech.com/online-only/the-role-of-barcode-technology-in-patient-safety-and-identification.php>

Using Georgia's driver's license and ID card. (2015). [Online]. Retrieved April 20, 2015 from http://www.dds.ga.gov/docs/news/brochur-barcode%20explanation_gt.pdf

Walker, T. (2013). Pilot program explores use of 2D barcodes in vaccine record-keeping. *Formulary (Cleveland, Ohio)*, 48(9), 281.

Zelbst, P., Green, K. W., Sower, V. E., & Reyes, P. M. (2012). Impact of RFID on manufacturing effectiveness and efficiency. *International Journal of Operations & Production Management*, 32(3), 329–350. doi:10.1108/01443571211212600

KEY TERMS AND DEFINITIONS

Barcoding Technology: A long-term and very reliable type of AIDC technology, it is known for its very accurate and economical approaches to identity products and machine-readable information from a variety of manufactured goods and services. Most barcodes use a type of standardized bars and spacing coding or symbology, certified by an international standards body (GS1 System). This system provides for the universal global acceptance of many types of barcodes designed for a variety of shipping and identification applications. Example barcode formats that are in common use today include EAN/UPC, GS1 DataBar, GS1-128, ITF-14, GS1 DataMatrix, GS1 QR Code and Composite Components.

Personally Identifiable Information (PII): Personally identifiable information (PII) refers to all information, including personal financial and healthcare information that is traceable to an individual customer or patient.

RFID-Embedded Technologies: RFID technologies are types of automatic data capture techniques that use a combination of active and passive senders and receivers to collect and store codified information for further uses. The implementation of such technologies should lead to improved managerial and/or supply chain performance. On the surface, there appears to be few drawbacks to implementing such technology into a production process, assuming it enhances performance and improves output of the product. The main issues surrounding the RFID applications are whether the initial costs and labor required to utilize this technology are worth it, and will result in a positive outcome of revenues.

Becoming Smart, Innovative, and Socially Responsible in Supply Chain Collaboration



Goknur Arzu Akyuz

University of Turkish Aeronautical Association, Turkey

Guner Gursoy

Okan University, Turkey

INTRODUCTION

Collaboration is rapidly emerging as an essential part of doing successful business (British Retail Consortium 2014). It has become the key to effectiveness, agility and competitiveness within the global dynamics of digital economy where competition is no longer between organizations but between supply chains (SCs), (Kim, 2006; Trkman et al., 2010; Akyuz & Gursoy, 2010, Cao & Zang, 2013; Lehoux, D'Mourse & Langevin, 2013). Because of various concerns; including rising resource costs, financial shocks, disruption in SCs, changing consumption patterns, emerging new business models and environmental issues; organisations are forced to manage an ever-changing business context. Such systemic challenges are evidently driven by multiple factors, and they cannot be resolved by any single organisation. Collaboration, two or more organisations working together to address common problems or develop opportunities, has the potential to create an overarching change for long-term sustainability. Hence, enterprises are turning to collaboration more and more to address problems too complex to deal with on their own (British Retail Consortium, 2014).

Collaboration in the context of SC is an amorphous and meta concept that has been interpreted in many different ways by both organizations and individuals. Academic definitions mainly focus on the business-to-business (B2B) internet-based technologies; while practical definitions have a wider scope (Wang, 2006).

Arshinder & Deshmukh (2008) list collaboration definitions in their study as: (a) joint planning, joint product development, mutual exchange of information and integrated information systems, cross coordination on several levels in the companies on the network, long-term cooperation and fair sharing of risks and benefits, (b) two or more independent companies working jointly to plan and execute supply chain operations with greater success than when acting in isolation, (c) a win-win arrangement to provide improved business success for both parties, (d) a strategic response to the challenges that arise from the dependencies.

Obviously, the concept is multi-dimensional, going far beyond simple transactional integrity among systems, and involving strategic-level exchange of information and decision making. It is also well-proven to be directly related with various ideas such as SC cooperation, integrity and visibility. In this context, both coordination and integrity refer to tight process couplings among SC partners. The term integration means the unified control of different processes, putting more emphasis on central control and ownership (Cao & Zang 2011), and collaboration puts more emphasis on governance through relationship.

The related literature provides sound support for the benefits accruing from collaboration; for the positive correlation with SC performance, and critical SC capabilities such as agility and flexibility (Akyuz & Gursoy, 2010; Sanders, 2007; Arshinder & Desmukh, 2008, Cao & Zang 2013, Wiengarten *et al.* 2013; Kim & Nettesine, 2013,

Cao & Zang 2011). When organisations come together, they can combine their resources, knowledge, insights, creativity and collective leverage to create radical change in critical business areas. Undoubtedly, process of learning to collaborate may take time; it may seem complex and unfamiliar and require passion, commitment and investment; but the achievements can be significant (British Retail Consortium 2014).

In this chapter, the confusion, interchangeable and ambiguous use of collaboration terminology is explored via literature taxonomy, and a collaboration maturity model is introduced. In the next section, the relationships and precedence among collaboration-related terminologies, as well as existing maturity models are discussed by highlighting the ambiguities and interchangeable use. Then, the concepts of smartness, innovation and corporate social responsibility (CSR) are discussed as the key themes for the current understanding of SC collaboration. Motivated by:

1. The lack of consensus on terminology;
2. Maturity stages of the existing maturity models; and
3. The critical importance of smartness, innovation and CSR, a conceptual model is developed based on the maturity model offered in Akyuz, Gursoy & Celebi (2014).

The conceptual maturity model:

1. Provides a mapping of the model stages onto various SC processes;
2. Offers process-based, staged and precise descriptions of chain-level evolution of collaboration;
3. Covers advanced forms of system-level collaboration; and
4. Explicitly treats smartness, innovation and CSR in relation to CSR.

Background - Terminology and Existing Maturity Models

Extant literature related with SC collaboration highlights that it is closely interrelated with the terms: communication, cooperation, coordination, integrity, partnership, visibility, trust and synchronization. The terms of “cooperation”, “coordination”, “collaboration” and “integrity” are interchangeably used and sometimes refer to different evolutionary stages along a continuum of dependency among SC partners. Akyuz, Gursoy & Celebi (2014) provides a comprehensive discussion on this confusion in terminology.

Thompson & Sanders (1998) put forward the continuum of ‘competition→ cooperation→ collaboration →coalescence’, considering the term “*coalescence*” as the highest level of integrity and joining forces. Kim et al. (2004) refer to the Speakman et al. (1998), which treat “cooperation”, “coordination” and “collaboration” as three different stages to define the transition to “collaboration”. In their classification:

1. Cooperation refers to long term contracts;
2. Coordination is associated with information linkages; and
3. Collaboration is associated with joint planning, integration and sharing.

The distinction of “intra- and inter-organizational coordination” is emphasized while discussing the opportunities of the internet-based information systems (Akyuz & Rehan, 2009; Chen & Chen, 2005; Arshinder & Deshmukh, 2008; Kelle & Akbulut, 2005). This transition emphasizes that ability to cooperate leads to coordination, which in turn evolves into the collaboration across SC partners. “Joint planning” appears as a critical ability which determines collaboration, and the term “integrity” is highlighted.

Based on Christopher (2005) and Werner (2008), Ivanov & Sokolov (2010) define the organizational levels as ‘open market negotiations→cooperation→communication/integration→coordination→collaboration’. In this categorization, collaboration is again treated as a broader and enhanced concept than cooperation, integration and coordination. However, it is striking that communication is used interchangeably with integration, the idea which the authors do not agree and also contradicts with Speakman et al. (1998). The authors hold the opinion that the integrity concept is far beyond “linkages” and “communication channels”.

Based on Cooks & Delattre (2001), Hoppe (2001) uses the term “coordination” and discusses four stages of coordination as ‘conventional’, ‘integrated’, ‘collaborative’ and ‘optimally coordinated’. The term “coordination” used is defined in generic, with a broader coverage, and the terms “integrity” and “collaboration” are used as “stages” along the evolution. Stages are defined starting from a firm-centric, conventional structure to first “externally integrated” supply chain-centric, and then strategically collaborating structure. It should be noted that “optimal coordination” is mentioned at a later stage, leaving the term “chain-centric collaboration” at an earlier one. This is quite controversial because collaboration appears as a “lower” phase of optimal coordination, contradicting with both Speakman *et al.* (1998), and Ivanov & Sokolov (2010). In this context, the term “optimal” is also controversial, since it raises question of the construction form of optimality.

Ayers (2010) mentions a 3-level classification with respect to the “degree of collaboration”. Collaboration is treated as a higher level than coordination, involving process collaboration, joint planning and scheduling and win-win partnership across the network. Ayers (2010) also refers to the collaboration roadmap developed by collaboration technical team working on the Supply Chain (SCOR) model. In this roadmap, levels are defined

in the order of increasing intensity of information exchange and uncertainty as follows:



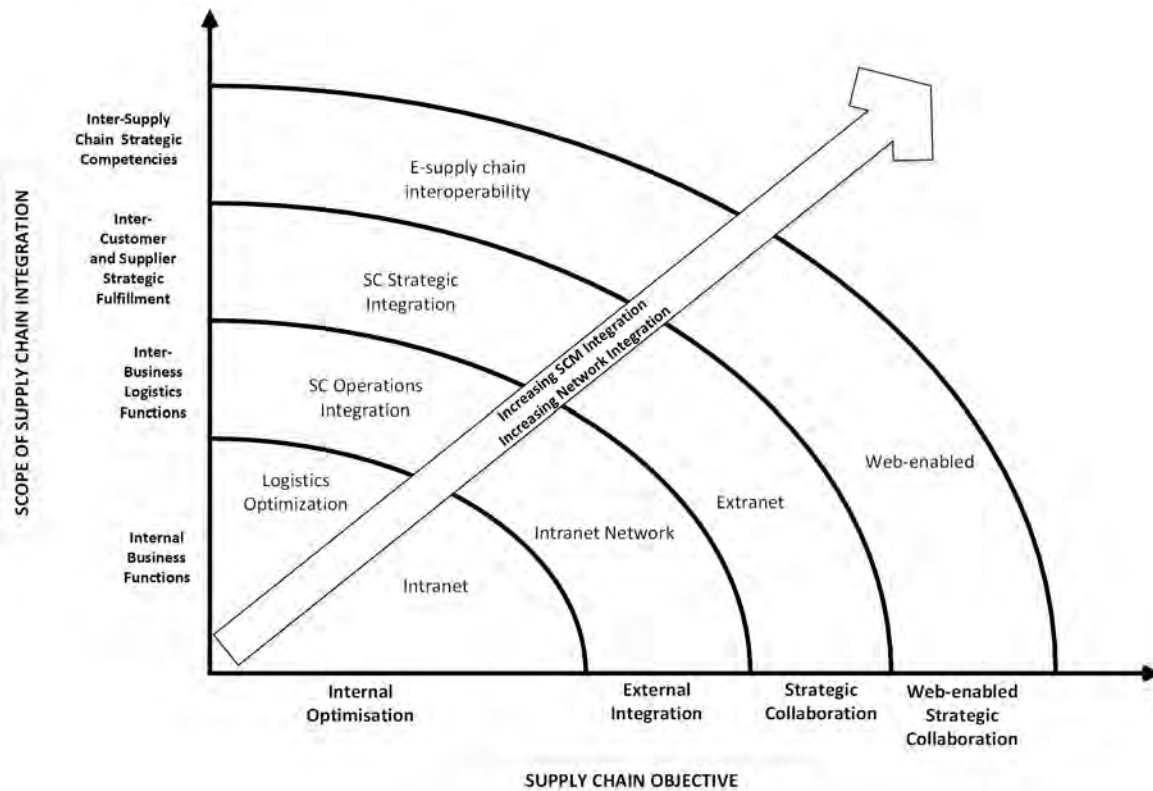
1. **Data Exchange Collaboration:** Where partners exchange information as required (one-way or two-ways), mainly to complete day-to-day transactions;
2. **Cooperative Collaboration:** Where partners share systems and tools so that all have access to information simultaneously;
3. **Cognitive Collaboration:** As the highest level requiring “joint, concurrent intellectual and cognitive activity” between partners, and embracing information sharing for joint decision-making. This roadmap considers “collaboration” as the main construct, treating coordination as a prior level to cooperation, and classifying them again as “degrees of collaboration”.

Jespersen & T.Skjott-Larsen (2005) discuss the SC integration levels in terms of complexity and SC relations. The proposed relationship defines the levels of ‘internal integration’, ‘coopetation’ and ‘synchronization’. This classification emphasizes the transition from functional to networked structures along with increased SC complexity. Although “integrated business processes” refer to external integrity, the terminology “external integration” does not appear explicitly in their study.

The “Integration” concept is also emphasized in Campbell & Sankaran, 2005; Xu et al., 2005; and Muffato & Payaro (2004) as part of the staged evolution of an organization. Emmet and Crocker (2006) prefer stages of “functional silos/ internal integration/ external integration”, based on the scope of integrity.

It is important to note that the distinction of “internal” versus “external” integration is frequently encountered in literature. Internal integration refers to “the degree to which a firm can structure its organizational practices, procedures and behaviors into manageable processes”, mainly involving data and information system integration

Figure 1. Supply chain objectives versus supply chain integration (Based on Ross, 2003)



through the use of Enterprise Resources Planning (ERP) systems, cross-functional cooperation and elimination of the functional silos (Zhao et al., 2011). External integration refers to the “*degree to which a firm can partner with its key supply chain members to structure their inter-organizational strategies, practices, procedures and behaviors into collaborative, synchronized and manageable processes*” (Zhao et al., 2011), information sharing, synchronized planning and joint problem resolution being the main themes. Ross (2003) discusses SC integrity using the “internal” versus “external” distinction along with extending level of network integrity and interoperability, leading to the e-supply chain formation, as depicted in Figure 1.

In this regard, various other literature items fully support the idea of ‘internal integrity being the necessary prerequisite for external integration’ (Rosenzweig et al., 2003; Stevens, 1989; Vickery et al., 2003; Akyuz & Rehan, 2009; Akyuz & Gursay,

2010). They support the transition from functional, internally-focused structures to extended architectures that can communicate and collaborate beyond the enterprise. Verdecho et al. (2011) emphasize this idea mentioning that “*if the collaboration is to succeed, internal collaboration has to be aligned with external collaboration*”. Clearly defined and integrated intra-enterprise process definitions become vital for being able to move to higher levels of integrity. Thus, standardizing, streamlining and coordinating the internal functions properly represent a critical stage of evolution (Akyuz & Rehan, 2009). Integrity of the internal processes are ensured as the isolated internal silos of functional information are broken down and enterprise-wide applications are efficiently put in place. Only after this stage that external integrity considerations and collaboration with other SC partners can be in question. Consequently, internal collaboration becomes a prerequisite for successful collaboration among partners. Zhao et al. (2011)



Table 1. Comparison table for various maturity models

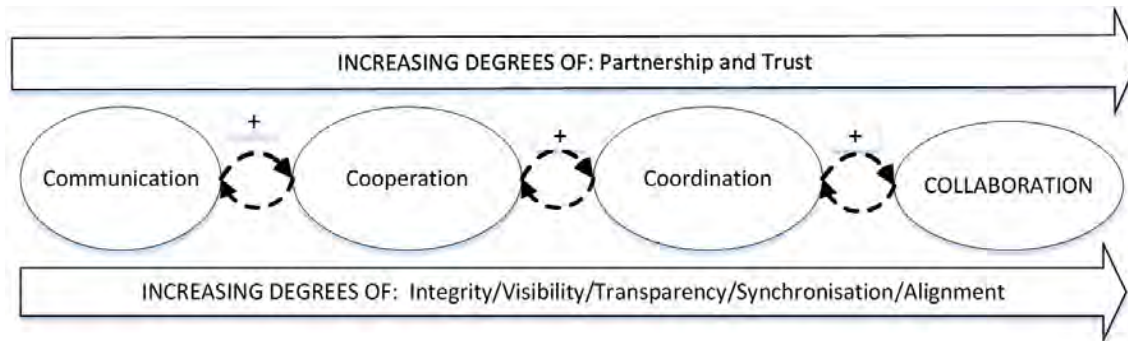
	Maturity Levels	Description
Computer Sciences Corp. (CSC) Framework based on (Poirier and Quinn 2003)	• Functional integration • Logistics integration • Inter-organizational processes integration • Collaborative initiatives • Complete join, collaboration	Defines maturity levels in terms of scope of integration/collaboration , starting from functional integration and extending to complete join.
Business Orientation Maturity Model by Lockamy & Mc Cormack (2004)	• Ad-hoc • Defined • Linked • Integrated • Extended	• A SCOR-based model • Starting with the ill-defined processes within an enterprise, stages are defined in terms of degree and extend of collaboration for various processes. • Highest level of integrity corresponds to the 'extended' level (multi-firm networks are strongly coupled by trust and mutual dependency). • Developed as a common language for benchmarking purposes, this model is critical as being SCOR-compliant, providing a grade scale for classifying the company into one of these classes.
Capability Maturity model integration- CMMI www.sei.cmu.edu/cmmi	• Initial • Managed • Defined • Quantitatively managed • Optimizing	• Ability to include quantitative approaches is defined as a maturity level. • Contains a set of best practices and roadmap.
Classification by Bendoly and Jacobs (2005)	• Fundamentals • Cross-functional teams • Integrated enterprise • Extended supply chain • Supply chain communities	Defines a 5-stage evolution model extending to 'SC communities'.
PMMM (Kerzner Project Management Maturity model) (Kerzner 2001, Ayers 2010).	• Common language • Common processes • Singular methodology • Benchmarking • Continuous improvement	5 stages are defined in terms of organizational awareness, management support, process discipline and motivation.
Pache and Spalanzani (2007)	• Intra-organizational maturity • Inter-organizational maturity • Extended, intraorganizational maturity • Multi-chain maturity • Societal maturity	5 stages are defined with a focus from intra to societal maturity.
SCOR Maturity, based on Cohen and Russel (2005).	• Functional integration • Internal integration • External integration • Intercompany integration	Four stages, extending from functional integration to intercompany integration.

(Based on Akyuz, Gursoy & Celebi, 2014)

emphasize the consistent findings in this line, provides empirical support arguing that internal integration enables external integration, and highlights that: “*organizations must first develop internal integration capabilities through system, data and process-integration before they can engage in any meaningful external integration*”.

Consequently, this is a staged evolution and collaboration-related maturity models present in the literature try to classify and define various maturity levels in terms of the terminology “integration” and “collaboration”. Table 1 shows a comparison table for different maturity models presents in the literature clearly indicates and

*Figure 2. Dependence and evolutionary sequence among different concepts
(Based on Akyuz, Gursoy & Celebi, 2014)*



compares the main stages used in these maturity models.

In this table, CSC (Computer Sciences Corporation) framework prefers the terminology “integration” and “collaboration” for defining stages. Business Orientation maturity model by McCormack & Lockamy (2004) use the wordings “linked”, “integrated” and “extended”. It is evident that collaboration is generally treated at a higher level than integrity. Although different naming and terminology is present, in all of the above-mentioned literature items a shift is observed from functionally-oriented and internally-focused structures to extended architectures in which communication and collaboration goes beyond the enterprise.

Collaborative Planning, Forecasting and Replenishment (CPFR) model by Voluntary Inter-industry Commerce Standards (VICS) Association is another key model related to SC collaboration (VICS, 2004, Seifert, 2003). The model is not cited as a “maturity model”, but rather it offers a process-based philosophy, and represents the ability to integrate planning, forecasting and replenishment-related activities across partners. It refers to tight couplings for the entire material management cycle, and utilizes the term “collaboration”. Its scope covers whole set of planning, forecasting and replenishment activities and it represents a significant stage for the integrity and restructuring of external relationships across partners.

All of the above discussion highlights that there is no clear consensus on the use of the terminology “coordination”, “collaboration”, “integration” and “cooperation” in literature. Apparently, developments in the SC domain lead to the search for and the creation of new terminology. This inconsistency is inevitably reflected in the naming and definitions of the maturity stages and further complicates the matters. Figure 2 clarifies the terminology and describe the dependence and evolutionary sequence among the concepts.

In this diagram:

- Communication forms the base for obtaining a certain level of “cooperation”, resulting in “coordination”. This is in compliance with the transition steps of Spekman et al. (1998);
- Continued “cooperation” and “coordination” evolve into the ability to “collaborate” at various processes. Meanwhile, visibility, transparency, synchronization and integrity among the partner systems continue to increase;
- Integrity covers architectural, procedural and organizational integrity both within and across the partners. This is in compliance with “internal integrity”/ “external integrity” distinction as well as accepting “internal integrity as a prerequisite to external integrity”. As external integrity in-

creases, higher levels of collaboration are achieved among actors;

- Visibility, generally used as synonymously with transparency, also increases in parallel with the level of integrity and synchronization. As partner systems start to operate in a more and more integral and synchronous manner, better visibility for entire set of processes is obtained;
- Transition is associated with increasing degree of partnership and trust.

SMART, INNOVATIVE AND SOCIALLY RESPONSIVE COLLABORATION

In the light of this basic background and evolutionary process described in the previous section, it is evident that integration and collaboration are the key themes characterizing the recent SCM practices. In SCM context, it is observed that being smart, innovative and socially responsible in collaborative activities are essential for success. This section discusses these three main themes in relation to SC collaboration.

Smart Collaboration

With all the dramatic advances and opportunities offered by technology, SCs turned into digitized smart value networks (Accenture 2014, Robinson 2015). We are witnessing the Industrial Revolution 4.0, which is characterized by:

1. Ubiquitous connection of people, things and machines;
2. The increasing digitization and interconnection of products, value chains and business models;
3. Integrated analysis and use of data with increasing transparency (Strategy & PwC 2014).

This is a totally new level of organization, control and visibility for the entire value chain

over the lifecycle of the products, revolutionizing the product portfolios, performance relationships, and resulting in innovative business processes. Piecemeal digitization of SC elements is seen to be counterproductive, and re-imagining the SC as an integrated, smart and value-creating supply network is essential. Digital technology renders the traditional SC models obsolete, facilitates an unprecedented degree of integration and collaboration across partners, and enables the handling of processes across SC partners in a smart and autonomous manner (Accenture 2014). Various recent concepts such as mobile technology, wearable technology, robot technology, autonomous vehicles, unmanned aerial vehicles, drones, automated cranes, low-cost sensors, advanced manufacturing systems, cloud computing, big data, 3D printing, Industry 4.0, and Internet of Things (DHL Report 2014, Deloitte 2015, Raconteur.net, Strategy & PwC (2014), Dimensional Research 2015, Robinson 2015) elevate the opportunities of integration and collaboration across partners to a totally new dimension. Today's SCM talks about smart factories, smart buildings and smart mobility with build-in intelligence, real-time capabilities providing end-to-end visibility. All logistics activities are handled by more and more intelligent systems (Raconteur 2015, Robinson 2015). People, trucks, containers and vehicles become increasingly autonomous and intelligent, working in perfect synchronization. 3D printing continues to be a hotly debated topic, with the prediction that as 3D printing starts to become a supplementary way to manufacture parts, a new complexity will be added to logistics activities. It would be necessary to smoothly integrate and orchestrate the logistics flows of traditionally manufactured parts and parts produced by 3D printing. This undoubtedly offers tremendous opportunities as well as challenges for the entire logistics cycle, completely reshaping the supply chains. All these technological opportunities make the SC more responsive and more demand-driven (Robinson 2015), and consequently, we are witnessing connected, complex, dynamic, digitized,

smart and scalable value creation networks (Kelly & Marchese 2015).

Within this technology-intensive character, enterprise application integration platforms of proven software vendors assume a vital role. These software suites provide seamless, end-to-end SC integration and visibility across partners for all the processes, including all logistics activities. In this regard, cloud technology offers tremendous opportunities for SC partners to access such platforms on the cloud without a prohibitive initial investment. This makes these software suites accessible even for small and medium sized enterprises, providing them with superb visibility and business analytics.

Consequently, the ideas of inter-enterprise integration and collaboration turned into smarter systems connectivity with the key capabilities of built-in business intelligence and analytics, autonomy, and integrated planning and analysis (Dimensional Research 2015, Strategy & PwC 2014). Hence, SCs will transform into smart, collaborative and value-creating networks with Industry 4.0 transformation.

Innovative Collaboration

Together with all these unprecedented technological advances, another key aspect for the success of the entire SC is to ensure that partners can develop innovative products, services and business models. Hence, collaborative innovation is the next big idea that allows players across the value chains to participate in the emergence of new collaborative business models. Based on solid foundations of intrapreneurship, collaborative innovation is the engine of modern and agile organizations capable of creating new capacity and pioneering radical new ideas (World Economic Forum 2015). Just like being a central theme to material management activities, collaboration is also central to innovation, since:

1. It increases the chances of associations between ideas resulting in innovative combinations;

2. Collaborative feedback accelerates the inevitable iterations;
3. It results in more connections to people to push a good idea forward;
4. Teams provide the energy and help overcome resistance; and
5. Collaboration helps ideas reach implementation.

Hence, across-partner collaboration on product design and development evolves into joint innovation creation ability in time. Undoubtedly, this ability is a competitive advantage in the harsh, global and chain-level competition continuously demanding faster product development cycles and new product ideas (Dance, 2008). World Economic Forum (2015) argues that companies that can manage collaborative innovation creates value by:

- Acknowledging the value of collaboration at strategic level across partners;
- Explicitly defining the objectives, openly declaring the intent and ensuring executive commitment at multiple levels from multiple partners;
- Defining the value and resources to be contributed by and brought to partners;
- Creating joint awareness and enabling joint management of the potential innovation and risks related with new product development;
- Identifying in advance at which stages and with which protections sensitive commercial information could be revealed to a partner;
- Ensuring that collaboration objectives and processes are aligned for maximum mutual gain.

Socially Responsible Collaboration

The concepts of sustainability, environmental sensitivity and corporate social responsibility (CSR) also became the key themes gaining utmost importance in SC collaboration. In this context,

Corporate Social Responsibility is the continuing commitment by business to contribute to economic development while improving the quality of life of the workforce and their families as well as of the community and society at large (The World Business Council for Sustainable Development). As it is evident from this definition, CSR is a broad concept which involves multiple stakeholders; and multiple issues. It is directly related with strategy, sustainability, environmental sensitivity, business ethics, value creation, service quality, customer retention and brand image of an enterprise (Wu 2006, Quarshie, Salmi and Leuschner 2016, Whitley and Willcoks 2011). More and more enterprises start to initiate various CSR programs encompassing a wide variety of subjects such as community involvement and development, consumer issues, fair operating practices and environmental protection, among others. Many concepts related with economic, environmental and social concerns are being treated under the umbrella term “CSR”. Thus, information exchange and integrating efforts across the SC partners towards meaningful CSR initiatives will create a chain-level competitive advantage (GCI-Capgemini 2016).

As a result, the ideas of smartness, innovation and social responsibility have to be addressed in any recent model related with SC collaboration concept.

COLLABORATION MATURITY MODEL

In the light of the previous two sections, main focus of this study is to develop a conceptual maturity model for SC collaboration. This section introduces a conceptual and 8-staged maturity model for defining the evolutionary phases of collaboration. This model is an enhanced version of the maturity model given in Akyuz, Gursoy & Celebi (2014), addressing the smartness, innovation and CSR concepts as well.

The model described incorporates: detailed description of the collaboration phases, system

complexity, scope of collaboration and matching of the stages with the spectrum of “communication → cooperation → coordination → collaboration” (given in Table 2), and mapping of different SC processes onto stages (provided in the Figure 3) by incrementally adding new ones at each stage.

The maturity model has the following characteristics:

- It considers the transition: “functional → enterprise → SC → SC network”, representing the move from functional silos within the enterprises to SC networks with complex interactions across partners;
- It is in compliance with:
 - The distinction “internal” versus “external” integration, and the necessity to obtain internal integrity as the prerequisite of external integrity;
 - The general trends observed in the maturity frameworks;
 - The presence of ERP systems as a foundation for improving integrity levels;
 - The idea of extended enterprise;
- It addresses advanced and strategic forms of collaboration. Various systemic efforts such as collaborative quality management, collaborative performance and risk management are integrated in this model structure;
- Very recent concepts such as collaborative innovation, collaborative CSR and smart collaboration are embedded in the model.

In this model, material management processes are considered as the “core processes” on which further ERP functionality is built to achieve internal integrity. The model considers complete ERP functionality at “*Internal integration level 2 (stage 2)*” by covering the processes such as costing, human resources and budgeting. The rest of the levels are externally oriented and supply chain-centric. Existence of an ERP backbone at this level ensures the achievement of a degree of

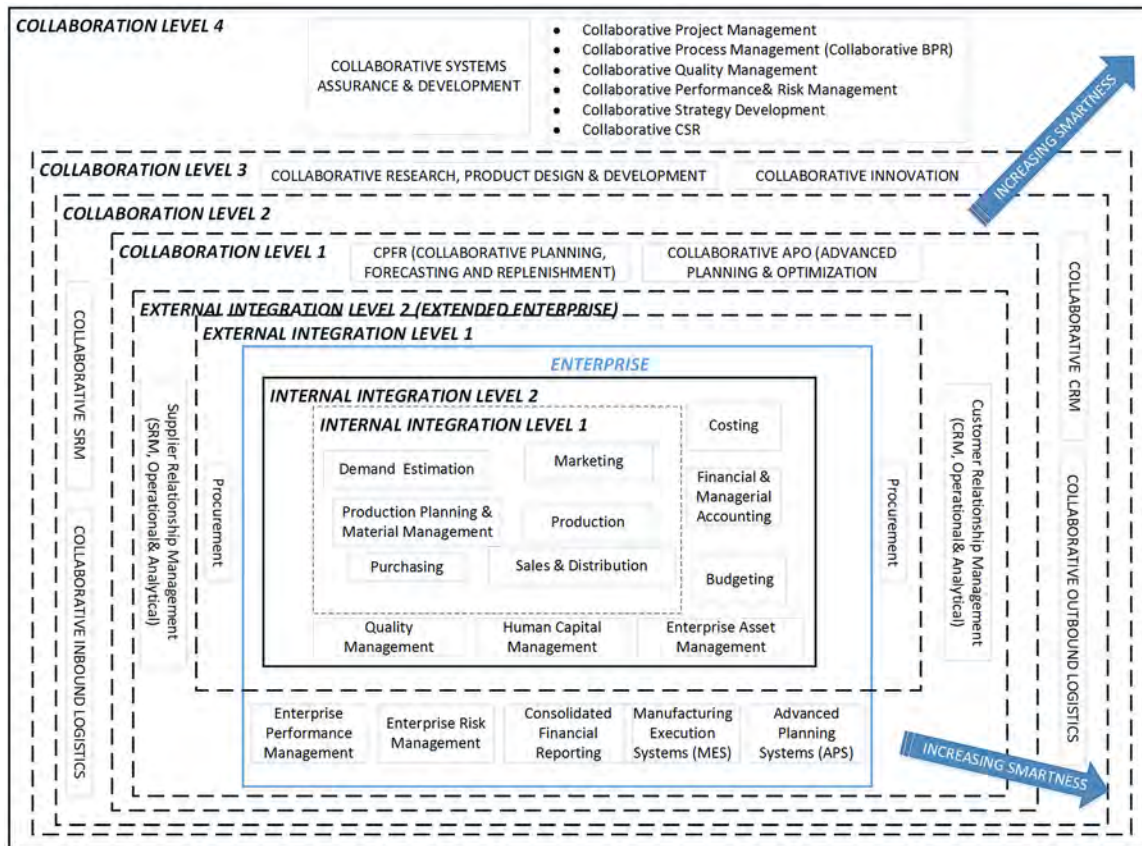


Table 2. Maturity model phases

Stages	Collaboration Level	Complexity	Scope of Collaboration	Level of Interaction Internal External	
1	Internal Integration Level 1	Functional	• Intra-departmental level. • Material management oriented. • Includes capacity planning. • Closed loop planning and decision cycle. • Represents the integration of purchasing, planning, inventory management and production.	Cooperate	Communicate
2	Internal Integration Level 2	Enterprise	• Intra-enterprise level. • Closed loop management of entire set of enterprise resources. • Includes the integration of costing, accounting, budgeting, quality management and HR processes.	Coordinate	Cooperate
3	External Integration Level 1	Supply Chain	• Inter-enterprise level including both suppliers and customers. • Material Management and Logistics oriented. • Initial ERP add-ons such as e-procurement. • Initial B2B integrity.	Coordinate	Cooperate
4	External Integration Level 2 (Extended Enterprise)	Supply Chain	• Inter-enterprise level including both suppliers and customers. • Extended view of the enterprise, extended ERP (Further ERP add-ons- SRM and CRM, mainly operational and analytical). • Further internal integrity with MES (Manufacturing Execution Systems) and APS (Advanced Planning and Scheduling) extensions. • Strategic enterprise management with corporate performance and risk measurement, consolidated financial reporting. • The concept of “smartness” starts to develop.	Collaborate	Coordinate
5	Collaboration Level 1	Supply Network	• Inter-enterprise level over a network. • Long term partnerships & visibility for inbound & outbound logistics • Collaborative planning, forecasting and replenishment. • Further capabilities like APO. • Joint problem solving, decision making, dispute resolution & action taking. • Increasing smart decisions across partners.	Collaboration	Collaboration
6	Collaboration Level 2	Supply Network	• Inter-enterprise level over a network. • Collaborative CRM and Collaborative SRM. • Strategic collaboration with customers, suppliers and logistics partners. • More strategic business intelligence and smarter decision making in SRM and CRM.	Collaboration	Collaboration
7	Collaboration Level 3	Supply Network	• Inter-enterprise level over a network. • Collaborative Product Design and Development. • This concept also includes the collaborative innovation. Ability to create innovative solutions • Sharing of mission-critical design data and innovative ideas	Collaboration	Collaboration
8	Collaboration Level 4	Supply Network	• Inter-enterprise level over a network. • Scope of collaboration extended to systems assurance and development, including • Collaborative Performance Management, • Collaborative Risk Management, • Collaborative Project Management, • Collaborative Quality Assurance, • Collaborative Strategy development, • Collaboration in Sustainability efforts & various CSR activities.	Collaboration	Collaboration
	Coalescence	Supply Network	• Perfect unification as a whole. • Perfect join of the forces. • Total integration of the systems.	Extreme Collaboration	Extreme Collaboration
	Virtual Network	Supply Network	• Perfect reliance on IT infrastructure. • Temporary structures with minimal or no existent resources.	Extreme Collaboration	Extreme Collaboration

(Based on Akyuz, Gursoy & Celebi, 2014)

Figure 3. Mapping of the supply chain processes to maturity stages
(Based on Akyuz, Gursay & Celebi, 2014)



“internal integrity” before any further systematic external integration efforts.

The model details are explained for the subsequent levels as follows:

- Procurement-oriented ERP add-on applications (such as e-procurement, e-invoicing, e-fulfillment) and initial B2B integrity are positioned in the model at “*External Integration Level 1*” (*stage 3*). This overrides the transaction-oriented and restricted understanding of e-procurement, and places the procurement function at the heart of the external integration;
- *External Integration Level 2* (*stage 4*) is characterized by increasing internal and external integrity, with further add-ons such as Manufacturing Execution Systems

(MES), Advanced Planning and Scheduling (APS), Supplier Relationship Management (SRM) and Customer Relationship Management (CRM). Operational as well as analytical SRM and CRM functionalities are involved. The ideas of “smartness” and “business intelligence” start to develop at this stage. Various smart technologies to capture instant data from manufacturing systems, customer side and supplier side are integrated with the backbone ERP systems. On-line, real-time business analytics enhance managerial decision making;

- *Collaboration Level 1* (*stage 5*) is defined to be totally compliant with Collaborative Planning, Forecasting and Replenishment (CPFR) model, and represents the joint planning, forecasting and replenishment

ability by taking joint resource constraints into consideration. This level is a critical stage for the network. Partners learn how to engage in joint problem solving, dispute resolution and exception handling across the complete planning, forecasting and order fulfilment cycle;

- *Collaboration Level 2 (stage 6)* provides strategic collaboration at SRM and CRM functions. Long term strategic partnership establishment is critical at this stage. Further business intelligence and smarter decisions for the customer and supplier sides are enabled;
- *Collaboration Level 3 (stage 7)*, extends the collaboration scope beyond CPFR to cover collaborative research, product design and development, and collaborative innovation. At this stage, all the design, customer and supplier-related data and information, as well as related plans and forecasts, are shared. Supply chain partners are able to create innovative ideas. This is compatible with the collaborative innovation idea offered in Whitney and Willcocks (2011). Undoubtedly, this level of transparency requires significant level of mutual trust among partners;
- *Collaboration Level 4 (stage 8)* represents a highly strategic degree of collaboration and alignment among partner systems for collaborative systems assurance and development. This level involves joint efforts for the areas such as collaborative quality management, collaborative performance and risk management. This necessitates sharing of strategic, system-level and confidential information and knowledge, which is clearly far beyond sharing forecasts, plans and replenishment data among partners. Mission-critical design data, joint risk assessment and performance measurement metrics and evaluations are some examples. At this stage, partners gain the “joint systems development” ability and

engage in joint business process reengineering. They also engage in collaborative strategy development, and reach a high level of strategic alignment for their long-term partnerships and for sustained collaboration. Efforts at this stage has a systems assurance and sustainability perspective. Collaboration scope is thus further broadened to encompass strategic corporate social responsibility (CSR) activities.

In this ordering, various applications like e-procurement, e-invoicing, SRM and CRM are treated at stages 3 and 4, in line with the well-supported “ERP add-ons” assertion in literature (Bendoly & Jacobs, 2005). The procurement relations are the first and the most fundamental external links to be initiated before any other add-ons, as commonly supported by the literature (Akyuz & Rehan, 2009; Leonard & Cronan, 2005; Kothari et al., 2007; Chang, Tsai & Hsu, 2013). Therefore, e-procurement (at stage 3) is listed before the other add-ons.

It is evident that the scope of collaboration defined in CPFR is beyond the capabilities of ERP add-ons as well as the operational and analytical levels of SRM and CRM. This broader scope demands systematic jointness and dispute resolution for the overall planning, forecasting and replenishment cycle (Seifert, 2003; Wang, 2006). Thus, stage 5 of the proposed model appears as a higher stage in compliance with CPFR. Capabilities mentioned at stages 6, 7 and 8 represent a collaboration scope beyond CPFR cycle, defining strategic collaborations at SRM, CRM, system and product development (Akyuz, 2011; Jayaram & Pathak 2013), joint BPR, performance and risk management areas.

Evidently, stages 4-8 demand strategic levels of information exchange, reporting and business analytics. Big data concerns (data warehousing, big data analytics, data security, and authority and authentication issues among multiple infrastructures) are increasingly amplified along these stages, starting mainly at stage 4.



Stage definitions “coalescence” and “virtual network” are included in the model as two extremes along the evolution.

“Coalescence” is the ultimate unification of the network partners and total integration of the systems. It is literally “sharing a common faith” among the partners, and “ability of acting as a single organism when needed”. This understanding is in compliance with the assertions of Thompson & Sanders (1998), positioning the term at the extreme along the path of evolution.

“Virtual network” is another extreme, generally representing temporary and rather volatile structures established on an ad-hoc basis with minimal or no level of physical resources. It widens the collaboration and business opportunities by relying on IT infrastructure and rapid reconfiguration capabilities. A general consensus is observed in the literature as to the temporary nature and reliance on IT for this term (Camarinha-Matos et al., 2009; Noran, 2009; Poirier & Quinn, 2003). Although they represent very dynamic and agile structures, the authors hold the opinion that the transient nature with minimal physical existence conflicts with the idea of long-term collaboration in business processes. Such a structure also neglects the soft aspects of the collaboration, such as corporate culture and mutual trust. With these characteristics, it is not highly probable that the scope of collaboration will be extended to collaborative product design and development, collaborative performance, risk and quality assurance for this type of organizational structure. Therefore, virtual enterprises not necessarily reach collaboration levels 2, 3 and 4 in terms of scope, or not necessarily lead to coalescence. Because of these characteristics, it is preferred to treat the term as an “extreme” stage in the model.

The term “value chain” is not included in our model. The concept of value creation for all the SC partners is inherently embedded throughout all stages to a higher or lesser extent. Value is being created starting even at the early stages of SC formation. For this reason, it is preferred not to use the term “value chain” to refer to a certain

specific maturity level. Consequently, collaboration level 4 (collaborative systems assurance and development) represents one of the highest levels to be reached in this structure.

The following parallelisms can be established between the stages of the proposed maturity model and the literature: (a) They coincide with the assertions of Cooks & Delattre (2001), as well as Jespersen & Skjott-Larsen (2005). (b) The “cognitive collaboration” stage of the roadmap from transactional to cognitive collaboration mentioned in Ayers (2010) corresponds to level 8 of the proposed model. (c) “Extended” stage (level 4) is compatible with the classification of Bendoly & Jacobs (2005).

Contribution and Strengths

The following issues stand out as the major contributions and strengths of the developed maturity model:

- The concepts of communication, cooperation, coordination, collaboration and integrity are jointly treated, clarified and tied up to a staged maturity model. The clarification of the concepts is represented in the Figure 2. We intend to contribute to the literature by eliminating the confusion in terminology, by establishing a common understanding and a common ground for future discussions in the SC field;
- Clear maturity stages are defined in compliance with the ideas of “beyond-ERP”, “ERP add-ons”, “internal integrity first”, “extended enterprise”, and the general trend followed by the existing maturity models. The proposed model provides more comprehensive and extended stage definitions when compared with the models existing in the literature. These refinements and definitions are significant contributions to SC research. The model contains 8 evolutionary and 2 extreme levels. Stages after collaboration level 1 (Stage 5) go beyond

the planning and material management oriented scope of CPFR. Hence, CPFR is extended after stage 5. This extended region is undefined in the maturity models given in literature, hence it is a contribution. The model provides the treatment of two extreme stages, virtual networks and coalescence, within the maturity model context as well. Since these recent terminologies in the literature are not treated systematically in relation to a staged maturity model, this is again a contribution;

- The model contains precise mapping of the maturity stages onto SC processes. Detailed definitions of the main processes which are incrementally added at each stage to the collaboration scope are provided in our model. The Figure 3 (mapping diagram), only Business Orientation Maturity Model by Lockamy & McCormack (2004) attempts to establish relationship with SC processes via SCOR methodology. However, its process coverage is limited with the processes included in SCOR model. The other models do not provide any relationship with the maturity stages and the SC processes. Hence, the proposed model is a contribution by establishing a precise mapping between the SC functions and the maturity levels;
- Collaboration is treated at all processes in the proposed model, including collaborative design and development, collaborative performance and risk management, collaborative project management, collaborative quality management and collaborative CSR. Process coverage of the existing models are mainly material management oriented, and they do not specifically include the processes mentioned at stages 6, 7 and 8, which are recent, advanced and more strategic forms of collaboration. This provides full process coverage in terms of collaboration scope. This is another contribution to SC literature, since existing

maturity models do not have such a broad and systematic coverage related with these advanced and strategic forms of collaboration. Hence, collaboration is treated at all inter-temporal dimensions (operational, tactical and strategic) for all processes in this model;

- Recent concepts of smartness, collaborative innovation and collaborative CSR are treated in this model.

Collaboration Roadmap

With the above-stated contributions and strengths, maturity stages and the evolutionary path, the model provides a managerial roadmap as well as a sound base for managerial decision making from a variety of perspectives:

1. **For Determining Collaboration Maturity Level of a Single Enterprise with Respect to the Levels Defined:** When supported by real-world field applications, stage definitions can be utilized as a base for critical decision making. As-is status of the enterprise can be determined with respect to these defined levels via questionnaire-based field studies and structured meetings designed in compatibility with the stage definitions of the maturity model. A standardized grading scheme and thresholds for the maturity levels can be determined for an enterprise. When the enterprise score is compared with the threshold levels, positioning of an enterprise can be established. In this regard, a variety of scales and different scorings can be utilized in connection with defined maturity levels, such as “low/fairly low/medium/fairly high/high/very high” or “non-existent/chaotic/repeatable/defined/managed/continuously improving” to give us an overall enterprise grade. Fuzzy approaches can also be applicable during this process. Thus, the developed model provides an open structure for the application of a variety of grading

schemes. Applying the measurement instrument systematically to multiple enterprises will eventually lead to the determination of valid thresholds and improved grading schemes. Determination of current positioning with respect to the maturity levels is a critical feedback to managers, since it shows the future direction along the roadmap and the generic path for achieving ultimate collaboration. It also reveals managerial development opportunities for higher levels of collaboration;

2. **For Obtaining a Collaboration Maturity Map for a Number of Enterprises:** If a similar survey-based approach is applied across enterprises, positioning of multiple enterprises with respect to the suggested maturity levels will also be possible. This will lead to a maturity map and a multi-enterprise view, showing clustering patterns and the general positioning of the multiple enterprises around the defined maturity levels. Such a picture will reveal the average maturity behaviour across multiple enterprises. Significant gaps in terms of collaboration maturity levels will be identified, allowing comparison and benchmarking across a number of enterprises. Thus, these stage definitions can be a sound base for benchmarking purposes. Within-network, within-sector, cross-sector, within-country and cross-country comparisons are possible to achieve collaboration maturity. From collaboration perspective, this is a vital network-level feedback, since:
 - a. Partners in need of radical improvement in specific processes can be determined;
 - b. Partners causing hindrance for the overall network collaboration can be identified;
 - c. Main processes in which enterprises fail to collaborate can be determined.

This feedback will enable joint managerial action decisions for the overall well-being of the

network, create an awareness of collaboration, and foster continuous improvement across the network.

L

FUTURE RESEARCH DIRECTIONS

The proposed conceptual model supports further SC research in two aspects:

1. Clearly defined and standardized stage definitions of the model, and mapping of each stage with the processes will establish a common understanding for collaboration maturity among supply chain researchers,
2. In its conceptual form, the model provides an open, flexible structure and a sound base for a number of future research items.

A multitude of survey-based field research can be designed based on the model. Questionnaire-based research studies can be designed at individual enterprise level to determine the current collaboration maturity. The model is generic and it is open for the development of different grading scales and survey instruments. Sector-based customizations and within-sector standardization can be made during the field studies for these measurement scales and survey instruments. Within-network, within-sector, cross-sector, within-country and cross-country comparative studies can also be designed to determine the maturity levels based on this model. Hence, model is valuable to SC researchers for their future studies.

As the field research accumulates along this direction, simplifications or refinements may be suggested for the stage definitions in the long run. Development and standardization of better quantitative measurement instruments, industry-specific customizations for both the survey instruments and maturity level transition thresholds will also follow. Thus, the model can evolve into a practical managerial maturity level assessment and a benchmarking tool along with the accumulation of field research. This assessment will support managers in determining the current maturity

levels and deciding within-enterprise, within-industry and cross-industry managerial actions to improve the SC collaboration.

The model can also be used as a base for the future development of collaboration and trust-related metrics. This study has also shown that future research is still justified in the SC domain regarding the following issues:

- Efforts for clarification, standardization and simplification of ever-increasing terminology;
- Development of causal dependency frameworks among various SC constructs; and
- Empirical studies to investigate the relationships among various constructs.

CONCLUSION

This paper first highlighted the lack of consensus and interchangeable use of the collaboration-related terminologies, and a clarification is provided by means of a dependence and evolutionary sequence diagram. Afterwards, the importance of being smart, innovative and socially responsible in collaborative supply chain relationships is comprehensively discussed in today's Industry 4.0 age. Then, a conceptual maturity model is developed for collaboration. The proposed model is a contribution to SC research by establishing a standardized, conceptual base, and offering a systematic, process-oriented big picture for multi-partner SC maturity.

The model is differentiated from the others by:

1. Providing full process coverage;
2. Extending CPFR model;
3. Defining clear maturity stages for advanced, strategic types of collaboration (like collaborative performance and risk management and collaborative systems assurance and development); and
4. Treating relatively recent terminology like "virtual networks" and "coalescence".

The model is also unique in embedding the recent concepts of smartness, innovation and CSR into a maturity model in relation to multi-partner collaboration.

It is expected that collaborating with Industry 4.0 applications, SCs will become integrated web of networks and increasingly aligned by gaining even more strategic character. In this regard, SC collaboration maturity has to be regarded as a continuously evolving issue with broadening collaboration scopes, and as a transformation from operational to strategic orientation. Within this broadened scope, enterprises have to demonstrate their ability to collaborate in larger scopes and in multiple modes, including joint innovation creation and joint CSR programs to create smart solutions.

REFERENCES

- Accenture. (2014). *The digital supply network: A new paradigm for supply chain management*. Retrieved June 15, 2016, from https://www.accenture.com/t20150708T025455__w_/fr-fr/_acnmedia/Accenture/Conversion-Assets/DotCom/Documents/Local/fr-fr/PDF_5/Accenture-Digital-Supply-Network-New-Standard-Modern-Supply-Chain-Management.pdf
- Akyuz, G. A. (2011). Collaborative quality management. In *Supply Chain Management: Pathways for Research and Practice* (pp. 43-56). InTech Open Access Publishing.
- Akyuz, G. A., & Gursoy, G. (2010). Taxonomy of collaboration in supply chain management. In *VIII. International Logistics and Supply Chain Congress* (pp.31-44). Maltepe University.
- Akyuz, G. A., Gursoy, G., & Celebi, N. (2014). Supply Chain Collaboration: A Conceptual Maturity Model. In J. Wang (Ed.), *Encyclopedia of Business Analytics and Optimization* (Vol. 5, pp. 2333-2349). doi:10.4018/978-1-4666-5202-6.ch210



- Akyuz, G. A., & Rehan, M. (2009). Requirements for forming an e-supply chain. *International Journal of Production Research*, 47(12), 3265–3287. doi:10.1080/00207540701802460
- Arshinder, A. K., & Deshmukh, S. G. (2008). Supply chain coordination: Perspectives, empirical studies and research directions. *International Journal of Production Economics*, 115(2), 316–335. doi:10.1016/j.ijpe.2008.05.011
- Ayers, J.B. (2010). *Supply Chain Project Management: A Structured, Collaborative and Measurable Approach*. CRC Press, Taylor and Francis Group.
- Bendoly, E., & Jacobs, F.R. (2005). *Strategic ERP Extensions and Use*. Stanford University Press.
- British Retail Consortium. (2014). *Collaboration Guidebook: Initiating, building and managing collaborations for long-term success*. Retrieved May 30, 2016 from <https://www.forumforthefuture.org/sites/default/files/files/Collaboration%20Guidebook.pdf>
- Camarinha-Matos, L. M. (2009). Leveraging knowledge for innovation for collaborative networks. *10th IFIP WG 5.5 Working Conference on Virtual Enterprises*. PRO-VE 2009, Thessaloniki, Greece.
- Campbell, J., & Sankaran, J. (2005). An inductive framework for enhancing supply chain integration. *International Journal of Production Research*, 43(16), 3321–3351. doi:10.1080/00207540500095852
- Cao, M., & Zhang, Q. (2011). Supply chain collaboration: Impact on collaborative advantage and firm performance. *Journal of Operations Management*, 29(3), 163–180. doi:10.1016/j.jom.2010.12.008
- Cao, M., & Zhang, Q. (2013). *Supply Chain Collaboration: Role of Interorganizational Systems, Trust and Collaborative Culture*. London: Springer-Verlag. doi:10.1007/978-1-4471-4591-2
- Chang, H. H., Tsai, Y., & Hsu, C. (2013). E-procurement and Supply Chain Performance. *Supply Chain Management: An International Journal*, 18(1), 34–51. doi:10.1108/13598541311293168
- Chen, T., & Chen, J. (2005). Optimizing supply chain collaboration based on joint replenishment and channel coordination. *Transportation Research Part E, Logistics and Transportation Review*, 41(4), 261–285. doi:10.1016/j.tre.2004.06.003
- Christopher, M. (2005). *Logistics and supply chain management: Creating value-adding networks*. Dorchester, UK: Financial Times Prentice Hall.
- CMMI. (n.d.). *Capability Maturity Model integration*. Software Engineering Institute, Carnegie Mellon. Retrieved July 10, 2012 from www.sei.cmu.edu/cmmi
- Cohen, S. & Russel, J. (2005). *Strategic Supply Chain Management*. McGraw-Hill.
- Consulting, D. (2015). *Supply Chain Talent of the Future: Findings from the third annual supply chain survey*. Retrieved June 10, 2016 from <https://www2.deloitte.com/content/dam/Deloitte/us/Documents/strategy/us-operations-supply-chain-talent-of-the-future-042815.pdf>
- Cookson, C., & Delattre, A. (2001). Collaborative manufacturing: A new supply chain opportunity. *Supply Chain Management Review Global*, (January-February), 10–14.
- Dance, J. (2008). *Five reasons why collaboration contributes to innovation*. Retrieved June 5, 2016 from <http://www.freshconsulting.com/5-reasons-why-collaboration-contributes-to-innovation/>
- DHL Trend Report. (2015). *Internet of things in logistics*. Retrieved May 20, 2016 from http://www.dpdhl.com/content/dam/dpdhl/presse/pdf/2015/DHLTrendReport_Internet_of_things.pdf

- DHL Trend Research & Cisco Consulting Services. (2014). *Self-driving vehicles in Logistics*. Retrieved June 5, 2016 from http://www.delivering-tomorrow.de/wp-content/uploads/2015/08/dhl_self_driving_vehicles.pdf
- Dimensional Research. (2015). *The state of supply chains in 2015: A survey of electronic manufacturing*. Retrieved May 15, 2016 from <http://cdn2.hubspot.net/hubfs/362383/JabilSupplyChainReport-v2.pdf?t=1430396680694&submissionGUID=45925539-8fe4-4efa-a04b-46bac932481d>
- Emmet, S., & Crocker, B. (2006). *The Relationship-driven Supply Chain; creating a culture of collaboration throughout the chain*. Gower Publishing Company.
- GCI & Capgemini. (2016). *Future supply chain 2016*. Retrieved May 15, 2016 from http://supplychainmagazine.fr/TOUTE-INFO/ETUDES/GCI_Capgemini-SC2016.pdf
- Hoppe, R. M. (2001). *Outlining a future of supply chain management: coordinated supply networks* (MS Thesis). MIT.
- Ivanov, B., & Sokolov, A. (2010). *Adaptive Supply Chain Management*. London, UK: Springer-Verlag Limited. doi:10.1007/978-1-84882-952-7
- Jayaram, J., & Pathak, S. (2013). A Holistic view of Supply Chain Integration in Collaborative Supply Chains. *International Journal of Production Research*, 51(7), 1958–1972. doi:10.1080/00207543.2012.700130
- Jespersen, B. D., & Skjott-Larsen, T. (2005). *Supply Chain Management in Theory and Practice*. Copenhagen: Copenhagen Business School Press.
- Kelle, P., & Akbulut, A. (2005). The role of ERP tools in supply chain information sharing, cooperation, and cost optimization. *International Journal of Production Economics*, 93-94(1), 41–52. doi:10.1016/j.ijpe.2004.06.004
- Kelly, E., & Marchese, K. (2015). *Supply chains and value webs*. Retrieved May 25, 2016 from <http://dupress.com/articles/supply-chains-to-value-webs-business-trends/>
- Kerzner, H. (2001). *Project Management: A Systems Approach to Planning, Scheduling, and Controlling*. New York: John Wiley and Sons.
- Kim, C. S. (2004). *Techniques to model the supply chain in an extended enterprise: state-of-the-art review*. VIVACE Consortium. VIVACE WP2.5/UNOTT/T/04021-1.0, 1-69. Retrieved Nov 30, 2011 from http://www.vivaceproject.com/content/engine/tmscee_full.pdf
- Kim, D. S. (2006). Process chain: A new paradigm of collaborative commerce and synchronized supply chain. *Business Horizons*, 49(5), 359–367. doi:10.1016/j.bushor.2005.11.004
- Kim, S., & Nettesine, S. (2013). Collaborative Cost Reduction and Component Procurement under Information Assymmetry. *Management Science*, 59(1), 189–206. doi:10.1287/mnsc.1120.1573
- Kothari, T., Hu, C., & Roehl, W. (2007). Adopting e-procurement technology in a chain hotel: An exploratory case study. *International Journal of Hospitality Management*, 26(4), 886–898. doi:10.1016/j.ijhm.2006.01.005
- Lehoux, N., D'Mourse, S. & Langevin, A. (2013). Inter-firm collaborations and supply chain coordination: review of key elements and case study. *Production Planning & Control: Management of Operations*. DOI: 10.1080/09537287.2013.771413
- Leonard, N. K., & Cronan, T. P. (2005). Electronic supply chain management: Does the customer benefit? *International Journal of Integrated Supply Management*, 1(13), 335–343. doi:10.1504/IJISM.2005.005954
- McCormack, K., & Lockamy, A. (2004). The development of a supply chain management process maturity model using the concepts of business process orientation. *Supply Chain Management: an International Journal*, 9(4), 272–278.



- Muffato, M., & Payaro, A. (2004). Integration of web-based procurement and fulfillment: A comparison of case studies. *International Journal of Information Management*, 24(4), 295–311. doi:10.1016/j.ijinfomgt.2004.04.008
- Noran, O. (2009). A decision support framework for collaborative networks. *International Journal of Production Research*, 47(17), 4813–4832. doi:10.1080/00207540902847355
- Pache, G., & Spalanzani. (2007). *La Gestion Des Chaînes Logistiques Multi-Acteurs: Perspectives Stratégiques*. Presses universitaires de Grenoble.
- Poirier, C.C. & Quinn F.J. (2003, September-October). A survey of supply chain progress. *Supply Chain Management Review*.
- Quarcshie, A. M., Salmi, A., & Leuschner, R. (2016). Sustainability and corporate social responsibility in supply chains: The state of research in supply chain management and business ethics journals. *Journal of Purchasing and Supply Management*, 22(2), 82–97. doi:10.1016/j.pursup.2015.11.001
- Raconteur. (2015). *Supply Chain 2015*. Retrieved May 25, 2016 from <http://raconteur.net/supply-chain-2015>
- Robinson, A. (2015). *4 Key Trends to Watch in Smart Manufacturing and Supply Chain*. Retrieved May 20, 2016 from <http://cerasis.com/2015/10/09/smart-manufacturing/>
- Rosenzweig, E. D., Roth, A. V., & Dean, J. W. Jr. (2003). The influence of an integration strategy on competitive capabilities and business performance: An exploratory study of consumer products manufacturers. *Journal of Operations Management*, 21(4), 437–456. doi:10.1016/S0272-6963(03)00037-8
- Ross, D. (2003). *Introduction to e-Supply Chain Management*. St. Lucie Press.
- Sanders, N. D. (2007). An empirical study of the impact of e-business technologies on organizational collaboration and performance. *Journal of Operations Management*, 25(6), 1332–1347. doi:10.1016/j.jom.2007.01.008
- Seifert, D. (2003). *Collaborative planning, forecasting and replenishment: how to create a supply chain competitive advantage*. USA: AMACOM.
- Spekman, R. E., Kamauff, J. W., & Myhr, N. (1998). An empirical investigation into supply chain management. *International Journal of Physical and Logistics Management*, 28(8), 630–650.
- Stevens, J. (1989). Integrating the supply chain. *International Journal of Physical Distribution & Materials Management*, 19(8), 3–8. doi:10.1108/EUM00000000000329
- Strategy & PwC. (2014). *Industry 4.0. Opportunities and challenges of the industrial internet*. Retrieved May 20, 2016 from <http://www.strategyand.pwc.com/media/file/Industry-4-0.pdf>
- Thompson, P., & Sanders, S. (1998). Partnering continuum. *Journal of Management Engineering*, 14(5), 73–78. doi:10.1061/(ASCE)0742-597X(1998)14:5(73)
- Trkman, P., McCormack, K., Oliveire, M. P. V., & Ladeira, M. B. (2010). The impact of business analytics on supply chain performance. *Decision Support Systems*, 49(3), 318–327. doi:10.1016/j.dss.2010.03.007
- Verdecho, M. J. (2011). *A Review of factors influencing collaborative relationships. In Adaptation and Value Creating Collaborative Networks: IFIP Advances in Information and Communication Technology* (Vol. 362, pp. 535–542). Springer. doi:10.1007/978-3-642-23330-2_58
- Vickery, S. K., Jayaram, J., Droge, C., & Calantone, R. (2003). The effects of an integrative supply chain strategy on customer service and financial performance: An analysis of direct vs. indirect relationships. *Journal of Operations Management*, 21(5), 523–539. doi:10.1016/j.jom.2003.02.002

VICS. (2004). *Collaborative Planning, Forecasting and Replenishment: an overview*(CPFR®). Voluntary Inter-industry Commerce Standards (VICS) Association. Retrieved from www.vics.org

Wang, Y. (2006). CPFR, E-collaboration-a literature analysis. *Intelligent Production and Machine Systems, 2nd I*PROMS Virtual International Conference*, (pp. 132-137). Elsevier.

Werner, H. (2008). *Supply Chain Management: Grundlagen, Strategien, Instrumente und Controlling*, 3. Wiesbaden: Gabler Verlag.

Whitley, E. A., & Willcocks, L. P. (2011). Achieving step-change in outsourcing maturity: toward collaborative innovation. *MIS Quarterly Executive*, 10(3), 95-109.

Wiengarten, F., Humphreys, P., McKittrick, A., & Fynes, B. (2013). Investigating the impact of e-business applications on supply chain collaboration in the German automotive industry. *International Journal of Operations & Production Management*, 33(1-2), 25–48. doi:10.1108/01443571311288039

World Business Council for Sustainable Development. (2016). *Corporate Social Responsibility*. Retrieved May 20, 2016 from <http://www.wbcsd.org/work-program/business-role/previous-work/corporate-social-responsibility.aspx>

World Economic Forum. (2015). *Collaborative Innovation Transforming Business, Driving Growth*. Retrieved June 10, 2016 from http://www3.weforum.org/docs/WEF_Collaborative_Innovation_report_2015.pdf

Wu, M. L. (2006). Corporate social performance, corporate financial performance, and firm size: A meta-analysis. *Journal of American Academy of Business*, 8(1), 163–171.

Xu, H., Ristic, M., Besant, B., & Pradoux, C. (2005). A web-based system for manufacturing co-ordination in complex supply chains. *International Journal of Production Research*, 43(10), 2049–2070. doi:10.1080/00207540500031865

Zhao, X., Huo, B., Selen, W., & Yeung, J. H. Y. (2011). The impact of internal integration and relationship commitment on external integration. *Journal of Operations Management*, 29(1-2), 17–32. doi:10.1016/j.jom.2010.04.004

ADDITIONAL READING

Altekar, R. V. (2005). *Supply Chain Management: Concepts and Cases*. New Delhi: Prentice-Hall of India Pvt. Ltd.

Blanchard, D. (2010). *Supply Chain Management Best Practices*. USA: John Wiley and Sons.

Burton, T. T., & Boeder, S. M. (2003). *The Lean Extended Enterprise: Moving beyond the four walls to value stream excellence*. USA: Ross Publishing Series.

Cohen, S. & Russel, J. (2013). *Strategic Supply Chain Management. The five disciplines for top performance*. USA: McGraw-Hill

Davis, E., & Spekman, R. (2003). *The Extended Enterprise: Gaining competitive advantage through collaborative supply chains*. USA: Prentice Hall.

Ireland, R.K. & Crum, C. (2005). *Supply Chain Collaboration: How to implement CPFR and other best collaborative practices*. APICS, USA: J. Ross Publishing.

Poirier, C.Q. & Morgan, F.S. (2009). *Diagnosing Greatness: Ten Traits of the Best Supply Chains*. FL, USA: J.Ross Publishing Inc.

Sanders, N. (2011). *Supply Chain Management: A Global Perspective*. USA: John Wiley and Sons.

Schwab, C. (2016). *The Fourth Industrial Revolution*. Switzerland: World Economic Forum.

Sehgal, V. (2011). *Supply Chain as Strategic Asset: The key to reaching business goals*. USA: Wiley.

Stadtler, H., & Kilger, C. (2008). *Supply Chain Management and Advanced Planning: Concepts, Models and Case Studies*. Berlin: Springer. doi:10.1007/978-3-540-74512-9

Unhelkar, B. et al.. (2010). *Collaborative Business Process Engineering and Global Organizations*. New York: USA. Hershey: Business Science Reference. doi:10.4018/978-1-60566-689-1

KEY TERMS AND DEFINITIONS

Collaboration: Ability of two or more partners to engage in joint activities at strategic, tactical and operational levels to plan and execute SC operations towards a win-win agreement with greater success than when acting in isolation.

Collaborative Innovation: Process of managing the interaction and collaboration of multiple partners to deliver new solutions within a business ecosystem.

Corporate Social Responsibility (CSR): Continuing commitment by business to contribute to economic development while improving the quality of life of the workforce and their families as well as of the community and society at large.

Enterprise Resources Planning (ERP) System: Integrated management information system towards planning, execution and control of all enterprise resources, including money and human resources.

External Integrity: The degree to which a firm can structure its organisational practices, procedures and behaviours into manageable processes across SC partners.

Internal Integrity: The degree to which a firm can structure its organisational practices, procedures and behaviours into manageable processes within the enterprise.

Maturity Model: Staged model describing different evolutionary levels towards improvement and better capabilities.

Smart Supply Chain: Instrumented, interconnected and intelligent SCs with unprecedented levels of integrity and collaboration across multiple SC partners to generate flawless information flow and decision making.

Supply: Chain (SC): Complex networked system of organizations, people, technology, activities, information and resources for moving a product or service from suppliers to customers. Includes channel partners, which can be suppliers, intermediaries, third-party service providers and customers, engaging in planning and management of all activities involved in sourcing, procurement, conversion, and all logistics management activities.

Visibility: Ability to trace all resources (including all physical items, data, information and knowledge) across entire processes and partners to provide a coherent view with a single version of truth across the network.



Concept and Practices of Cyber Supply Chain in Manufacturing Context

Anisha Banu Dawood Gani

Universiti Sains Malaysia, Malaysia

Yudi Fernando

Universiti Malaysia Pahang, Malaysia

INTRODUCTION

In an effort to improve business agility and responsiveness to changing market requirements, many manufacturing companies are decentralizing non-value adding activities by outsourcing. The anticipated benefits of outsourcing are to improve profitability and operating efficiency (Gonzalez, Gasco & Llopis, 2005), reduce capital investment (Lynch 2004), improve business focus (Baldwin, Irani & Love, 2001, Weerakkody, Currie and Ekanayake, 2003), enhance flexibility (Jennings 2002; Lynch 2004), and to gain a competitive advantage (Clott, 2004). This widely adopted business model only solves manufacturer's perennial priorities around cost, growth, risk sharing, and supply chain efficiency. However, achieving the appropriate level of supplier visibility is key, along with investment into greater technology enablement to have an agile and transparent supply chain.

KPMG International's Global Manufacturing Outlook (GMO) Report (2015) shows that while manufacturers are concerned about supplier performance and capacity, the visibility into supplier organizations remains surprisingly low. According to the report, only 14% of respondents (out of 386) claim to have complete supplier visibility into Tier 1 (manufacturers' direct suppliers), Tier 2 (Tier 1 supplier's supplier) and beyond. One of the primary condition for a supply chain to be effective is by integrating the members of the network to ensure an undisrupted information flow. Add transparency to the equation and it in

its entirety becomes a requirement that is achievable by digitally integrating the supply chain, or in the course of this chapter, otherwise termed as the cyber supply chain (CSC).

The advancements in technology has been pivotal in transforming the physical supply chain into an integrated CSC. CSC provides tremendous advances in efficiency and effectiveness. (Linton, Boyson, & Aje, 2014). The benefits of digital integration and collaborations are realized through low cost but rapid transmission of information's which facilitates joint planning and reaction to events, in real-time, by all stakeholders. When integrated for manufacturing, data and information technologies bring intelligence and insight that enables fact backed-up decisions.

This chapter is intended to address this deficiency by comprehensively examining the concept and practices of CSC in a manufacturing context. It is also aimed to provide awareness into the nature of CSC risks and recommend best practices in implementing CSC. Due to the limited literatures pertaining to CSC specifically, this chapter has applied general supply chain management literatures to CSC context while duly differentiating the uniqueness, where applicable.

BACKGROUND

Manufacturing is one of the leading industries of the global economy and is projected to undergo a major transformation driven by power of digita-

lization. Manufacturing industry's share of GDP has remained stable over the last 40 years. According to United Nations Industrial Development Organization (UNIDO), the total contribution of the entire manufacturing sector to GDP, measured as manufacturing value added (MVA), reached an all-time high of \$9,228 billion in 2014. The MVA of developing and emerging industrial economies (DE IEs) for the same year increased 2.4 times from 2000, while their GDP doubled (UNIDO, 2015). Inarguably, manufacturing remains a key driving force of overall economic growth globally.

UNIDO Director General, Mr. Li Yong recognized that the technological change shall become one of the main drivers of long-term growth for the industry:

"In the coming decades, radical innovations such as the mobile internet, the Internet of Things and cloud computing are likely to revolutionize production processes and enhance living standards, particularly in developing countries", he envisioned.

These rising trends are promising evidence that the digitalization of manufacturing industry will dramatically transform the way companies operate in many areas – from Research & Development (R&D) efficiency and faster product launches to supply chain improvements, better operations services and more efficient processes. Oliver Wyman, one of the leading management consultancies in the world, predicts that the global annual margin impact of digital industry across discrete manufacturing in 2030 could be an estimated US\$1.4 trillion (Oliver Wyman, 2016).

As these trends play out in a growing number of manufacturing sectors, the incumbents should concentrate on product innovations and core competencies. Non-core competency activities should be outsourced to suppliers who have the specialization and technological advantage to offer scalability. Therefore, the need for integration among supply chain partners becomes eminent

for a seamless and dynamic collaboration. Consequently, CSC network too can be seen growing in tandem to the manufacturing sectors growth. Thus, it is not far-fetched to conclude that, organization that has a strong and secure CSC are likely to develop a competitive advantage among industry players.

Many researchers have published isolated literatures on the benefits of integrated supply chain (Richey, Roath, Whipple & Fawcett, 2010; Ajmera & Cook 2009; Gimenez & Ventura, 2005; Sabath & Whipple, 2004) its security risks (Boyson, 2014; Elzarka, 2013; Ma, Chen, Meng, & Yi, 2014; Micheli, Mogre, & Perego, 2014; Tummala & Schoenherr, 2011) and integration challenges (Awad & Nassar, 2010; Carter, Monczka, Robert, & Ragatz, Gary, 2009; Pagell, 2004). However, there is no evidence on the concept of CSC and its practices being covered holistically nor has it been attributed specifically to a manufacturing industry. Thus, this paper will answer research question "what is cyber supply chain?" and "how does cybercrime affect supply chain?"

LITERATURE REVIEW

There is a lack of clear definitions and understanding of supply chain management (SCM) in general (New, 1997; Lummus, Krumwilde & Vokurka, 2001; Mentzer et al., 2001; Kauffman, 2002). This problem remains despite a considerable growth in the number of published articles dealing with the topic since the mid-1980s (Stock and Boyer, 2009) due to the unavailability of a good SCM definition (Kathawala & Abdou, 2003). Furthermore, there is no widely accepted definition of related topics such as supply chain integration (SCI) and supply chain collaboration. (Näslund & Hulthen, 2012). However, Mentzer et al. (2001) proposed a broad definition that covers issues related to SCM which can be used to explain any specific discipline such as SCI and CSC. The definition suggested that SCM can be defined as the systemic, strategic



coordination of the traditional business functions and the tactics across these business functions within a particular company and across businesses within the supply chain, for the purposes of improving the long-term performance of the individual companies and the supply chain as a whole. In lieu of this, this chapter introduces the term ‘cyber supply chain’ in defining a digitally integrated and transparent supply chain.

The term ‘cyber supply chain’ is non-existent in literatures on its own to be formally understood as integrated supply chain. Boyson et al. (2009) used the term CSC to represent ‘the entire set of key actors involved with or using cyber infrastructure’. Often many authors conjoin CSC with ‘security’ or ‘risk management’ instead. In fact, the introduction of CSC in this chapter is made by borrowing the term from ‘Cyber Supply Chain Risk Management’ (CSCRM). CSCRM is a new integrative discipline combining elements of cyber security (Boyson, 2014).

Scholars have often used terms like Supply Chain Integration (SCI), Supply Chain Management (SCM), e-Supply Chain Management (e-SCM), Supply Chain Management Integration (SCMI) and other related terms to define an integrated supply chain. However, since the prefix ‘cyber’ describes a control mechanism for information processing in an organization, it is befitting that the very definition of cyber supply chain management implies a governed diversity of hardware and software and system integration activities (Boyson, 2014).

In the context of this chapter, the term CSC is defined as ‘an end-to-end integration of critical supply chain processes between channel partners over secured and intricate digital network with controlled level of transparency’. There are four underlying elements that encompasses this CSC definition: digital network, process integration, security and transparency, which the authors believe to be the backbone of a successful supply chain integration. Thus, CSC is a combination of technology and supply chain management, which

can be further explained by understanding the stages and types of integration.

STAGES AND TYPES OF INTEGRATION

Stages of Integration

Boyson et al. (2009) identified four stages in supply chain integration approach:

Stage 1: Integration activities occur in isolation within own function/department

Stage 2: Internal integration among independent supply chain functions

Stage 3: Integration with external stakeholders, where technology and processes are extended to key customers and suppliers

Stage 4: Cross-enterprise collaboration and optimization with external stakeholders in joint decision making, with real time performance improvement

The CSC approach for the industry incumbent should (ideally) be a Stage 4 where the need to integrate goes beyond integrating to ensure seamless delivery, into having information flow visibility and transparency in a regulated and controlled manner. That is, proprietary information and classified design details must be vigilantly secured from open access yet the customer’s needs and adequate visibility into material demands must be made known so that it drives the requirement and subsequently the fulfilment as desired.

A relatively small sized company, however, may not necessarily undertake a major leap to integrate. Rather it is strategically viable to adopt Stage 1 or Stage 2 approach, depending on their financial capabilities, as these integration software and process are costly, time consuming and resource dependent. Therefore, for a new entrants or Small and Medium Enterprises’ (SME), the initial stages would be a safer starting point.

Types of Integration

In CSC, upstream (supplier) and downstream (customer) integration has emerged as an important element of manufacturing strategy. As discussed, global competition too, demands for the need for an integration among suppliers, internal processes, and customer to facilitate information flow of inbound and outbound flow of information and goods or services. (Won, Kwon, Severance, Lee, & Severance, 2013). Accordingly, Won et.al (2013) proposed three critical types of integrations: with supplier, within the organization and with customer.

Customer Integration

- Concerns with planning, implementing, and evaluating successful relationships between providers (supplier) and recipients (customer) either upstream or downstream of supply chain.
- Is a medium where customers can interact virtually using a secure, uninterrupted systems and be given option to obtain information or place orders ability to real time delivery progress status.
- Must consist of a customer relationships management (CRM) focus on both internal and external customers.

Supplier Integration

- Deals with strategic linkages with suppliers. For example, involving suppliers as early in the product's lifecycle phase in order to ensure advance sourcing and supply of required components to build product.
- Concerns developing a rapid response order processing system that assures reliable delivery and exchanging real time information.

Internal Integration

- Revolves around easy access to key operational data from the integrated informa-

tion system linking to various departments in an organization, accessing to real time inventory information status, utilizing a computer-based planning system between departments.



CRITICAL SUCCESS FACTORS AND ENABLERS

Kumar, Singh and Shankar (2015) have studied numerous literatures and identified 13 critical success factors for an integrated supply chain implementation, namely: top management commitment, development of effective SCM strategy, devoted resources for supply chain, logistics synchronization, use of modern technologies, information sharing with SC members, forecasting of demand on point of sale (POS), trust development in SC partners, developing just in time (JIT) capabilities in system, development of reliable suppliers, higher flexibility in production system, focus on core strengths and long-term vision for survival and growth.

In addition, this chapter has also reviewed four key enablers to a successful supply chain integration from 2009 CAPS research: i) Communication and E-Systems, ii) Organization Structure and People, iii) Metrics and iv) Trust. However, recognizing a need to have a current and redefined enablers, this chapter present six critical success factors for CSC integration by blending previous research with current literatures and business needs:

1. Integration and Information Technology

A secured technology to collect, store and disseminate real time data and information. Equally important are the process and procedures that supplements and encourages use of that data for decision making. Real-time information exchange with suppliers and with customers will create an opportunity where optimization can take place. The keyword here is a secured network that ensures the integrity of the data transmitted across and within supply chains.

2. Cooperation to Collaboration

Willingness to collaborate among supply chain partners both internally and externally are achievable by building trust. Utilization of mutually beneficial contracts or agreement would foster a trustworthy relationship among stakeholders. Manufacturers must involve and consider their suppliers and key stakeholder's ability and limitations in implementing the integration vision and in subsequent decision makings. Early involvement by key players promotes project ownership by those supplier (Nito, 2005)

3. Organization- People, Process, Policies and Culture (3P1C)

Simpson (2010) and Goertzel (2010) have both identified that people, process and technology constitutes the IT enabled supply chain. Boyson (2009) added policy developers as one of the key actors in cyber infrastructure. Adding a collaborative culture of exchanging information openly to reduce uncertainty (Christopher & Lee, 2004), this chapter introduces a 3P1C acronym to represent organizational readiness and receptiveness to adopt CSC as one of its implementation success factors.

4. Customer-Centric Metrics and Incentives

Structured incentives for drivers that facilitate management of shared risk should be developed across the supply chain. Rewards for cost savings and development of ideas that leads to improved products and/or higher cost savings should be offered based on proven and traceable metrics. Incentive-based payment is important (Clegg, Pitsis, Rura-Polley & Marosszeky 2002; Bayliss, Cheung, Suen, & Wong, 2004; Crespín-Mazet & Ghauri, 2007) since it distributes benefits evenly and signals that collaboration is legitimate and desired (Kadefors, 2004)

5. Access Authentication and Authorization

Controlled authorization on who is eligible to access classified and confidential data, and records about who has accessed the data, needs to be managed. This management capability is essential for secure operations and for robust detection and mitigation in the event of a security attack.

6. Transparency

Frameworks to mitigate supply chain risk must promote transparency by all parties. Vendors will need to provide an appropriate degree of transparency into their business processes, and visibility into the controls they use to ensure their product and service development and operations are secure.

CYBER SUPPLY CHAIN BENEFITS

The benefits of an integrated, secure and efficient cyber supply chain to nations and manufacturers are manifold. They ultimately lead to gaining competitive advantage, solving supply and demand related problems, improving supply chain efficiency, lowering operational cost, promoting partnerships across the supply chains and ultimately fulfilling and satisfying the needs of customers. National Institute of Standards and Technology (NIST) under the purview of US Department of Commerce has identified seven possible benefits achievable with CSC:

CYBER SUPPLY CHAIN (CSC) RISKS

In order to answer the second research question, a review on the potential cyber supply chain risks and the measure to secure it need to be looked at. The benefits of information flow in an integrated supply chain are not gained without risks (Smith, Watson, Baker, & Pokorski II, 2007). Access to information also provides an opportunity to cripple the supply chain if misused. The International Multilateral Partnership Against Cyber Threats (IMPACT), an alliance of 152 countries,

has recognized that “cyber threats and attacks are able to strike from virtually anywhere in the world, potentially causing catastrophic social and economic harm” (IMPACT, 2014). The flexibility, scalability, and efficiency of the very technology that enables information sharing between external supply chain partners has also created additional points of access to an organization’s proprietary information, increasing the risk of cybercrime.

Cybercrime is “any crime that is facilitated or committed using a computer, network, or hardware device” (Gordon & Ford, 2006). Common types of cybercrime include denial of services, web hacking and defacement, malware, spam and phishing (Wolden, Valverde, & Talla, 2015), online scams and fraud, (information) and identity theft, attacks on computer systems and illegal or prohibited online content (ACORN, 2016). Any breakdown or security breach in the information system at any point of integration would possibly be detrimental for the organizations reputation.

Manufacturers adopting CSC now must take into consideration the many types of threat scenarios when assessing their cybersecurity. In the past, manufacturers had to worry about cyber-attacks as a form of industrial espionage and thefts. Today however, cyber-attacks can include attempts to take control of production networks and infrastructure for blackmail purposes, infecting plants and equipment with malware to attack end customers, indirect attacks on critical infrastructure via controllers, etc. (Oliver Wyman, 2016).

Confluence of trends and technologies is expected to revolutionize manufacturing industry. Industry 4.0 (the digitization of the manufacturing sector) and Internet of Things (increased machine-to-machine communication) are example of technological advance that has probability to give rise to more threats. Therefore, manufacturers must be aware that when a data is shared between two companies, there is always a potential risk, particularly if one side of the connection is less secure or if a third party is handling the data. Moreover, the data interfaces and connections

Table 1. Benefits of cyber supply chain

Benefits	Details
1. Visibility	Real time visibility across the supply chain
2. Data storage	Untampered, electronic chain of custody of data
3. Traceability	Track and trace products movements and status online
4. Integrated value chain performance	analyze and model across the enterprise to address global performance objectives and take advantage of untapped opportunities for optimization
5. Optimization	Enterprise interoperability sets the stage for collecting and analyzing information and data on processes
6. Predictive analysis	Real-time analytics and modeling to form insights from supply chain interactions
7. Reduced Security and Legacy System Risk	secure environment to share critical data replacing out-of-date systems and equipment

between Business-to-Business (B2B) and vendor applications can themselves become points of cyber vulnerability. A successful breach of any one component could endanger the operation and security of other flows and result in system-wide failure. Lee and Whang (2000) consider achieving the critical balance between integration and security to be among the most challenging and relevant contemporary topics. Thus, it is imperative for manufacturing companies to implement protective measures to secure its CSC network without compromising its data integrity and speed to market opportunity.

PROTECTING CSC NETWORK

Cybercrime is ranked as a top ten risk to the global economy (Oliver Wyman, 2016), evidently indicating that the severity of the cyber threats are not to be taken lightly. No business is immune to the attacks, and thus, must invest its time, resource and



money to secure its CSC network. World Economic Forum (WEF) forecasts that delays in adopting cyber security capabilities could result in \$3 trillion loss in economic value by 2020 (KPMG, 2014). Manufacturers' must adopt a mix of measures to combat cyber threats and improve its supply chain resilience, in the areas of technology, organization, governance, and culture, and integrate them into the company's established risk management processes. These can be but not limited to having redundant backup systems, multiple-stage access for credentials, and ongoing threat monitoring to fend off the expanding list of malware attacks.

Some of the methods that manufacturers can use to address cyber threats are by:

1. **Regular Software Patches:** Patch widely used software on regular basis, systematically.
2. **Controlling Access threshold:** Govern and restrict access to sensitive data by monitoring which users have access to data and what they are doing with it.
3. **Periodic Internal Assessment:** Conduct regular audit to check vulnerability, limit internet usages, restrict external storage device and other media recording ports.
4. **Cyber Education:** Educating internal and external stakeholders on importance of cybersecurity.
5. **Security Certification:** Obtain certification from recognized, external auditors to ensure compliance to security standards as a mean to prove credibility.

Manufacturers must note that despite implementing the recommended steps above, it would still be insufficient to protect the organization without their Tier 1 and Tier 2 suppliers being secured as well. This can be achieved by surveying vendors' security practices across a supply chain network and applying security research, developing the IT capabilities needed to better examine and predict potential threats.

SOLUTIONS AND RECOMMENDATION

The significance of making an informed decision concerning supply chain activities cannot be undermined. In order for firms to obtain real time data, they must increase their visibility into their supply chain networks. Thus, firms are strongly recommended to invest in having a CSC to have a 'single version of truth' pertaining their business. The following measures are recommended to mitigate risks associated with CSC:

1. **Perform Risk Assessment:** Identify the types and storage of the data that is valuable to the organization, and establish protection and recovery plan.
2. **Establish Security Plan:** Define how assets would be protected, how threats can be detected and how to respond to security breaches.
3. **Hire Security Detectives:** Setup job role and hire right sets of people who are able to execute the security plan.
4. **Invest in Cybersecurity Solution:** To assist with threat detection and data encryption for a resilient supply chain.

The success factors of these security hygiene's rely on engaging and involving stakeholders to equally adopt the security practices.

DISCUSSION AND FUTURE RESEARCH DIRECTIONS

Today's global supply chains rely upon fast, accurate and secure dissemination of data among supply chain partners. The need for businesses to integrate does not appear to be slowing down with the industrial revolution predicted. However, CSC is often researched from an individual firm or a specific industry's perspective in separation.



Would businesses be more receptive to adopting CSC and cybersecurity solutions if they are guided and provided necessary startup and post-startup support? It would be fascinating to conceptualize firms' receptiveness to CSC and cybercrime by exploring the idea of setting up a national level cyber-resource center that governs cyber ecosystem by providing businesses, government and academia the sources, training and consultancy required to identify, mitigate and protect themselves from being a cybercrime victim. The findings might change the security landscape of a country.

CONCLUSION

Technological advancements are drastically transforming manufacturing industry, which in turn is driving the need for cyber supply chain (CSC). CSC enables manufacturing companies to reap numerous benefits from cost savings, efficiency improvement to ultimately gaining competitive advantage. There are three types of supply chain integration: with customer, supplier and internal and six critical success factors for implementation; i) integration and information technology, ii) cooperation to collaboration, iii) 3PIC (people, process, policies and culture, iv) customer-centric metrics and incentives, v) access authentication and authorization and vi) transparency. However, CSC is not gained without risks, therefore manufacturing companies must undertake several steps to protect their supply chain network, for example, by restricting access to sensitive information, performing regular software patches, conducting periodic audit, creating awareness among its stakeholders and obtaining security certification to imply compliance. Manufacturing companies must have top down management's support to turn cybersecurity as a culture and to allocate funds to invest in cybersecurity solutions. None of these measures would be fully successful without stakeholders participating in adopting the security hygiene practices.

REFERENCES

- Ajmera, A., & Cook, J. (2009). A multi-phase framework for supply chain integration. *SAM Advanced Management Journal*, 74(1), 37.
- Australian Cybercrime Online Reporting Network (ACORN). (2016). Retrieved 21 March 2016, from <http://www.acorn.gov.au/what-is-cybercrime/>
- Awad, D. (2010). Supply chain integration: definition and challenges. *Management and Technology*, 1(1).
- Baldwin, L. P., Irani, Z., & Love, P. E. (2001). Outsourcing information systems: Drawing lessons from a banking case study. *European Journal of Information Systems*, 10(1), 15–24. doi:10.1057/palgrave.ejis.3000372
- Bayliss, R., Cheung, S. O., Suen, H. C., & Wong, S. P. (2004). Effective partnering tools in construction: A case study on MTRC TKE contract 604 in Hong Kong. *International Journal of Project Management*, 22(3), 253–263. doi:10.1016/S0263-7863(03)00069-3
- Bitzenis, A., & Nito, E. (2005). Obstacles to entrepreneurship in a transition business environment: The case of Albania. *Journal of Small Business and Enterprise Development*, 12(4), 564–578. doi:10.1108/14626000510628234
- Boyson, S. (2014). Cyber supply chain risk management: Revolutionizing the strategic control of critical IT systems. *Technovation*, 34(7), 342–353. doi:10.1016/j.technovation.2014.02.001
- Boyson, S., Corsi, T., & Rossman, H. (2009). *Building a cyber supply chain assurance reference model*. Science Applications International Corporation. SAIC.
- CAPS Research. (2009). *Supply Chain Integration: Challenges and Good Practices*. Retrieved 21 March 2016, from <http://www.som.cranfield.ac.uk/som/dinamic-content/media/research/scrip/capsreport.pdf>

- Carter, P. L., Monczka, R. M., Ragatz, G. L., & Jennings, P. L. (2009). *Supply chain integration: Challenges and good practices*. Institute for Supply Management and WP Carey School of Business at Arizona State University.
- Christopher, M., & Lee, H. (2004). Mitigating supply chain risk through improved confidence. *International Journal of Physical Distribution & Logistics Management*, 34(5), 388–396. doi:10.1108/09600030410545436
- Clegg, S. R., Pitsis, T. S., Rura-Polley, T., & Marosszeky, M. (2002). Governmentality matters: Designing an alliance culture of inter-organizational collaboration for managing projects. *Organization Studies*, 23(3), 317–337. doi:10.1177/0170840602233001
- Clott, C. B. (2004). Perspectives on global outsourcing and the changing nature of work. *Business and Society Review*, 109(2), 153–170. doi:10.1111/j.0045-3609.2004.00189.x
- Crespin-Mazet, F., & Ghauri, P. (2007). Co-development as a marketing strategy in the construction industry. *Industrial Marketing Management*, 36(2), 158–172. doi:10.1016/j.indmarman.2006.02.011
- Elzarka, S. M. (2013). Supply chain risk management: The lessons learned from the Egyptian revolution 2011. *International Journal of Logistics Research and Applications*, 16(6), 482–492. doi:10.1080/13675567.2013.846307
- Gimenez, C., & Ventura, E. (2005). Logistics-production, logistics-marketing and external integration: Their impact on performance. *International Journal of Operations & Production Management*, 25(1), 20–38. doi:10.1108/01443570510572222
- Goertzel, K. M. (2010). Supply Chain Risk Management and the Software Supply Chain. *OWASP AppSec DC*.
- Gonzalez, R., Gasco, J., & Llopis, J. (2005). Information systems outsourcing reasons in the largest Spanish firms. *International Journal of Information Management*, 25(2), 117–136. doi:10.1016/j.ijinfomgt.2004.10.002
- Gordon, S., & Ford, R. (2006). On the definition and classification of cybercrime. *Journal in Computer Virology*, 2(1), 13–20. doi:10.1007/s11416-006-0015-z
- International Multilateral Partnership against Cyber Threats (IMPACT). (2014). *What we stand for*. Retrieved 21 March 2016, from <http://www.impact-alliance.org/aboutus/mission-&-vision.html>
- Jennings, D. (2002). Strategic sourcing: Benefits, problems and a contextual model. *Management Decision*, 40(1), 26–34. doi:10.1108/00251740210413334
- Kadefors, A. (2004). Trust in project relationships—inside the black box. *International Journal of Project Management*, 22(3), 175–182. doi:10.1016/S0263-7863(03)00031-0
- Kathawala, Y., & Abdou, K. (2003). Supply chain evaluation in the service industry: A framework development compared to manufacturing. *Managerial Auditing Journal*, 18(2), 140–149. doi:10.1108/02686900310455137
- Kauffman, R. G. (2002). Supply Management: What's in a Name? Or, Do We Know Who We Are?. *Journal of Supply Chain Management*, 38(3), 46–50.
- KPMG-Global Manufacturing Outlook. (2014). Retrieved 21 March 2016, from <https://www.kpmg.com/SG/en/IssuesAndInsights/ArticlesPublications/Documents/Advisory-CS-Cyber-Risks-in-Emerging-Markets.pdf>
- KPMG-Global Manufacturing Outlook. (2015). Retrieved 21 March 2016, from <https://www.kpmg.com/CN/en/IssuesAndInsights/ArticlesPublications/Documents/Global-Manufacturing-Outlook-O-201506.pdf>



- Kumar, R., Singh, R. K., & Shankar, R. (2015). Critical success factors for implementation of supply chain management in Indian small and medium enterprises and their impact on performance. *IIMB Management Review*, 27(2), 92–104. doi:10.1016/j.iimb.2015.03.001
- Linton, J. D., Boyson, S., & Aje, J. (2014). The challenge of cyber supply chain security to research and practice—An introduction. *Technovation*, 34(7), 339–341. doi:10.1016/j.technovation.2014.05.001
- Lummus, R. R., Krumwiede, D. W., & Vokurka, R. J. (2001). The relationship of logistics to supply chain management: Developing a common industry definition. *Industrial Management & Data Systems*, 101(8), 426–432. doi:10.1108/02635570110406730
- Lynch, C. F. (2004). Why outsource?. *Supply Chain Management Review*, 8(7), 44–51.
- Ma, Y. F., Chen, L. W., Meng, S. D., & Yi, C. Q. (2014). A study on the risk control of supply chain under the background of globalization. *Journal of Industrial and Production Engineering*, 31(4), 221–228. doi:10.1080/21681015.2014.934304
- Mentzer, J. T., DeWitt, W., Keebler, J. S., Min, S., Nix, N. W., Smith, C. D., & Zacharia, Z. G. (2001). Defining supply chain management. *Journal of Business logistics*, 22(2), 1–25.
- Micheli, G. J., Mogre, R., & Perego, A. (2014). How to choose mitigation measures for supply chain risks. *International Journal of Production Research*, 52(1), 117–129. doi:10.1080/00207543.2013.828170
- Näslund, D., & Hulthen, H. (2012). Supply chain management integration: A critical analysis. *Benchmarking: An International Journal*, 19(4/5), 481–501. doi:10.1108/14635771211257963
- National Institute of Standards and Technology (NIST). (n.d.). *Best Practices in Cyber Supply Chain Risk Management*. Retrieved 21 March 2016, from http://www.nist.gov/itl/csd/upload/NIST_USRP-SmartManu-Cyber-SCRM-Case-Study.pdf
- New, S. J. (1997). The scope of supply chain management research. *Supply Chain Management: An International Journal*, 2(1), 15–22. doi:10.1108/13598549710156321
- Pagell, M. (2004). Understanding the factors that enable and inhibit the integration of operations, purchasing and logistics. *Journal of Operations Management*, 22(5), 459–487. doi:10.1016/j.jom.2004.05.008
- Richey, R. G. Jr, Roath, A. S., Whipple, J. M., & Fawcett, S. E. (2010). Exploring a governance theory of supply chain management: Barriers and facilitators to integration. *Journal of Business Logistics*, 31(1), 237–256. doi:10.1002/j.2158-1592.2010.tb00137.x
- Sabath, R., & Whipple, J. M. (2004). Using the customer/product action matrix to enhance internal collaboration. *Journal of Business Logistics*, 25(2), 1–19. doi:10.1002/j.2158-1592.2004.tb00179.x
- Simpson, S., Baldini, D., Bits, G., Dillard, D., Fagan, C., Minnis, B., & Reddy, D. (2010, June). Software Integrity Controls—An Assurance-Based Approach to Minimizing Risks in the Software Supply Chain. *SAFECode*.
- Smith, G. E., Watson, K. J., Baker, W. H., & Pokorski Ii, J. A. (2007). A critical balance: Collaboration and security in the IT-enabled supply chain. *International Journal of Production Research*, 45(11), 2595–2613. doi:10.1080/00207540601020544
- Stock, J. R., & Boyer, S. L. (2009). Developing a consensus definition of supply chain management: A qualitative study. *International Journal of Physical Distribution & Logistics Management*, 39(8), 690–711. doi:10.1108/09600030910996323

Storch, T. (2011). Toward a Trusted Supply Chain: A Risk Based Approach to Managing Software Integrity. *Microsoft Corporation Paper*, 7-9.

Tummala, R., & Schoenherr, T. (2011). Assessing and managing risks using the supply chain risk management process (SCRMP). *Supply Chain Management: An International Journal*, 16(6), 474–483. doi:10.1108/13598541111171165

UNIDO. (2015). *Industrial Development Report 2016: The Role of Technology and Innovation in Inclusive and Sustainable Industrial Development*. Retrieved 21 March 2016, from <http://www.unido.org/resources/publications/flagship-publications/industrial-development-report-series/industrial-development-report-2016.html>

Weerakkody, V., Currie, W. L., & Ekanayake, Y. (2003). Re-engineering business processes through application service providers: Challenges, issues and complexities. *Business Process Management Journal*, 9(6), 776–794. doi:10.1108/14637150310506693

Wolden, M., Valverde, R., & Talla, M. (2015). The effectiveness of COBIT 5 Information Security Framework for reducing Cyber Attacks on Supply Chain Management System. *IFAC-PapersOnLine*, 48(3), 1846–1852. doi:10.1016/j.ifacol.2015.06.355

Won Lee, C., Kwon, I. W. G., & Severance, D. (2007). Relationship between supply chain performance and degree of linkage among supplier, internal integration, and customer. *Supply Chain Management: An International journal*, 12(6), 444-452.

Wyman, O. (2016). *Perspectives on Manufacturing Industries*. Retrieved 21 March 2016, from <http://www.oliverwyman.com/insights/publications/2016/feb/perspectives-on-manufacturing-industries-2016-1.html>

ADDITIONAL READING

Cappelli, D., Moore, A., Trzeciak, R., & Shimeall, T. J. (2009). Common sense guide to prevention and detection of insider threats 3rd edition–version 3.1. Published by CERT, Software Engineering Institute, Carnegie Mellon University, <http://www.cert.org>

Das, S. K., Kant, K., & Zhang, N. (2012). *Handbook on securing cyber-physical critical infrastructure*. Elsevier.

KEY TERMS AND DEFINITIONS

Cyber Supply Chain: An end to end integration of supply chain over secured and intricate digital network.

Cybercrime: Any crime that is facilitated or committed using a computer, network, or hardware device.

Cybersecurity: Mechanism to fend off cyber crime.

Digitalization: Use of digital technologies to change a business model.

Risk Management: Procedure to avoid and minimize risks.

Security: Free from danger and threats.

Supply Chain Integration: Coordination within a supply chain using integrated system.

The Concept of Modularity in the Context of IS Project Outsourcing



Shahzada Benazeer
University of Antwerp, Belgium

Philip Huysmans
University of Antwerp, Belgium

Peter De Bruyn
University of Antwerp, Belgium

Jan Verelst
University of Antwerp, Belgium

INTRODUCTION

Due to the globalization of the world economy and advancement in information and communication (ICT) technologies, information systems (IS) outsourcing became a very common practice in developed and in emerging economies. Over 94% of 'Fortune 500' companies are outsourcing at least one major business function (Modarress, Ansari, & Thies, 2014). Despite IS outsourcing's popularity, failure rates are high. Premature contract terminations and frequent dissatisfaction with IS outsourcing results are commonly encountered. Many IS outsourcing failures are not even reported publicly due to the fear of negative consequences from the market and the stakeholders. Empirical research has attempted to quantify the high probability of IS outsourcing failures. For instance, a joint study in 2000 by Oxford University's Institute of Information Management and the University of Missouri, which tracked 29 major IS outsourcing projects over eight years, reported that more than 35 percent failed (Gay & Essinger, 2000). A second study by Martineau and Shumway (2009), which compared data from 2004, 2006 and 2009, concluded that by 2009, the percentage of failed IS outsourcing projects (i.e., cancelled prior to completion or delivered and never used), being

44%, had decreased but was still (too) high. In the same study, the percentage of challenged IS projects (late, over budget, and/or with less than the required features and functions) also decreased by 2009 but still amounted to 32% in 2009. It is implied that the success rate in 2009, being 24%, did increase over time but was still low (Ciric & Rakovic, 2010). In this paper the authors are illustrating an alternative and new approach, based on the concept of modularity, which may supplement other approaches in order to significantly reduce the high failure rate in outsourcing. The newness of the proposed perspective is clear from Fixson, Ro, and Liker (2005, p.167) who state,

in the past, modularity and outsourcing were investigated predominantly in separate research communities but more recently, however, a new research stream has emerged that links these two topics together.

Modularity is defined as a property of a complex system, whereby the system is decomposed into several subsystems (i.e., modules). Modules are the decomposed and nearly independent parts of a larger system; the modules are less complex than the larger system. The decomposition of a larger system into modules allows breaking apart

or splitting up the complexities into smaller pieces (modules). The modules are still (mutually) compatible as, together, they work as a whole towards a common goal. The modules' compatibility logically follows from the adoption of specific 'design rules' (i.e., a chosen design option; see later) using an interface as connector (Langlois, 2002). The interface should be 'well-defined' (i.e., leaving no room for ambiguities) as it describes the input a module requires to perform its (functional) task, and the output it provides to its external environment (which includes the other modules embedded in the system). The amount of dependent 'design parameters' between different modules determines their coupling. Good modular design requires low coupling as low coupling enables design parameters of modules to remain stable, that is any change in the design parameter of one module has no or limited impact on the design parameters of other modules. A good modular design therefore exhibits the following two properties: (a) if the design of one module needs to be changed, the change will have no or only a limited impact on the design of other modules, and (b) the function of one module can be studied more or less in isolation from the functionality of the rest of the system. In order for necessary intermodular dependencies to work adequately, all intermodular dependencies should be clearly defined / made explicit *ex ante*, that is no hidden dependencies are allowed for (Mannaert & Verelst, 2009). If all intermodular dependencies in a system are clearly defined, a set of prescriptive rules is obtained, which all modules of the system need to adhere to. This set of rules is referred as the 'modular architecture'.

BACKGROUND

A pertinent question deals with how IS outsourcing failure may be avoided. So far, literature includes many suggestions ('remedies') offered by both scholars and practitioners. Peterson and Carco (1998) suggested to streamline operations and 'fix the problem' before outsourcing an IS.

Various remedies were introduced; the interested reader is referred to:

1. Lambert, Emmelhainz, and Gardner (1999) who introduced their '*Partnership Model*';
2. Greaver (1999) who formulated '*seven steps to successful outsourcing*'; and
3. Rottman (2008) who elaborate on the importance of 'knowledge transfer'; and
4. Harris, Herron and Iwanicki (2008) who stressed the importance of a high quality 'service level agreement' (SLA).

Despite the introduction of such remedies, the empirical research referred to above continues to attest to the high failure rate of IS outsourcing projects. It seems these remedies turned out to be partially successful at best. Partial successfulness may be due to the fact that most of those remedies focus on treating the symptoms (i.e., the consequences) rather than the causes of ineffective IS outsourcing. As a consequence, it is plausible that some important outsourcing risks were not identified *ex ante*. The desire to be better equipped to identify outsourcing risks *ex ante* has fueled the study presented in this paper. As a matter of fact, this paper presents an alternative, new perspective on how to avoid IS project outsourcing failures. The new perspective goes beyond the higher-mentioned remedies, and explores the concept of modularity in relation to IS outsourcing projects. Such a new perspective, which originates from the system sciences, may add value to outsourcing risk analyses. The added value comes from new or deeper insights which are instrumental in defining an exhaustive list of required mitigating actions, which is key for a successful completion of IS outsourcing projects. To illustrate the new perspective, a reanalysis of a failed (actual) outsourcing case is discussed. The paper shows in detail how poorly designed modular structures at the project level could have been identified *ex ante*. This approach is similar to the application of modularity to other context, such as requirements engineering (Verelst et al., 2013).

CASE INTRODUCTION

The selected case deals with a vendor organization, referred to as 'Aries', and a customer organization, referred to as 'Taurus'. Aries was regarded as a competent and reputed service provider and Taurus was a public sector university. The IS outsourcing project was rather small and simple. Taurus needed a web-based portal for academics' records management (e.g., new applicants' data, students' data, exams' data etc.). Taurus formed a team (hereafter 'focal team') of specialized professionals who had to deal with all aspects of the IS project and communicated to the Aries' team. After encountering many problems and extending project duration for 12 months, the project was categorized as a failure by the both parties. Taurus was unsatisfied with the delivered solution and did not use the new web-based portal. Aries was asking for more time and money to complete the IS project according to the demands of Taurus.

METHODOLOGY

This study relied on

1. An in-depth literature review and
2. A re-analysis of the IS outsourcing project (the case) using inference logic.

The in-depth literature review consisted of a systematic, step-by-step approach to collect, synthesize, analyse, and highlight the relevance of the modularity concept in the context of IS outsourcing. The literature review started with a systematic procedure (i.e., framing questions, identifying relevant work, assessing the quality, summarizing the evidence, and finally interpreting the findings) to identify and collect relevant papers which may be instrumental in achieving the study objectives (Webster & Watson, 2002). Next, data (i.e., paper texts) were collected and

analyzed thoroughly. For a literature review to be adequate (e.g., 'effective'; see Bandara, Miskon, & Fiel, 2011) 'key concepts' to capture and report on are to be formulated a priori (and revised during the literature review). The a priori list of key concepts included: '*modular systems*', '*modularity*', '*coupling*', '*dependencies*', '*IS outsourcing*', and some other, related concepts. Through an in-depth analysis of textual data, the nature of the relationship between the concept of modularity (e.g., dependencies, and coupling) in the context of IS outsourcing has been explored. Key findings from the literature review were supplemented with findings obtained from reanalyzing the case (introduced above). Benbasat, Goldstein, and Mead (1987) justified the use of 'case study analyses as it facilitates capturing practitioners' knowledge and developing theories from such knowledge. For this case analysis, a qualitative research approach was considered most suitable as this case study deals with a contemporary phenomenon (e.g. modularity, IS outsourcing) where the boundaries between the phenomenon and context are not clearly evident (Yin, 2011). The case selection criteria for this paper are the following:

1. Sufficient information should be available to explore the applicability of the concept of modularity;
2. The IS outsourcing project should have been presented and discussed already in the literature in order to explicitly illustrate how the outcome could be different if the concept of modularity were relied on *ex ante*.

The case study research efforts were aimed at better understanding the extent to which practitioners (working on an outsourced IS project) design fine grained modular structures. In addition, the re-analysis of the case may also illustrate how fundamental concepts such as modularity can be applied in practice.



FINDINGS: ANALYZING IS OUTSOURCING FAILURE

Analyzing the case, it has been revealed that some unfortunate decisions had been taken by the vendor and the customer that might have led to the failure of the IS outsourcing project. Concerning the service level agreement (SLA), Nauman and colleagues (2009) stated that “A weak legal environment in the developing countries ... does not allow the development team to challenge the client on the basis of Service Level Agreement” (p. 273). Even though the statement is a valid, it does not facilitate an understanding of why (some) unfortunate decisions were made, or why the problematic consequences occurred. Put differently, such an analysis does not look for the root causes of the observed project failure. In order to investigate these root causes, it is worthwhile to re-analyze the SLA from a modularity perspective. In re-analyzing the case, some deficiencies regarding the SLA were identified. For instance, incongruent with modularity, the SLA contained ‘hidden dependencies’ as referred to in the introduction. Apart from the SLA, other violations against the principles of modularity were found as well. Further on in the text it is illustrated how the implications of these violations of modularity have been identified by Nauman and colleagues (2009) to lead to the IS outsourcing project failure. Consequently, these violations can aid during the identification of the root causes for project failure in IS outsourcing projects. The violations against modularity -- and therefore root causes of failure -- are typically identified *ex post*. However, the formulation of an analysis framework based on the modularity-perspective could aid in detecting these violations *ex ante*. The re-analysis of the case therefore follows three recurring steps. First, to adequately identify modularity aspects in a certain part of the case, a subsection “Identifying the modular structure” is used. Second, this paper describes which requirements modularity prescribes with regard to the identified aspects

under a subsection “modularity requirement”. Third, evidence from the original case description is presented which illustrates how these modularity requirements are violated under a subsection “Assessing the modularity requirements”. Obviously, any violation of modular design principles may, at least partially, contribute to the failure of IS outsourcing project.

1. Analysis #1: Service Level Agreement (SLA) consisting of undefined or hidden dependencies

In the introduction it was stated that modules should interact with one another through the interface. To function adequately, the interaction between modules should be exhaustively and unambiguously documented in the interface. As far as the SLA is concerned, responsibilities of each module, rights of each module, and the relationships between modules are to be described in detail. In the context of IS outsourcing project, the SLA essentially provides an interface between the vendor (team Aries) and the customer (the focal team of Taurus).

a. Identifying the modular structure

Based on the paragraph just above, team Aries and the focal team of Taurus are conceived as modules and the SLA serves as the interface connecting both teams.

b. Modularity requirement

As explained in the introduction, team Taurus is only allowed to ask team Aries for services which are described in the SLA. Any service asked for which is not described in the SLA constitutes a hidden dependency, which may result in unwanted outcomes, namely coupling and/or CEs. As mentioned earlier, in a good modular design all intermodular dependencies are made explicit, that is no hidden dependencies are allowed for.



The explicitness of intermodular dependencies is now referred to as modularity requirement #1.

c. Assessing the modularity requirements

After many hurdles and delays, the out-sourced IS project completed the ‘testing phase’. As agreed in the SLA, after the testing phase, the Taurus team leader would start to lead the implementation of the IS project. But within three months the Taurus focal team discontinued implementing the project and asked the Aries team to appoint an expert who would supervise the entire implementation process, and to train the end users in using the new IS system. As such services were not part of the SLA, which was agreed upon by the both parties, the Aries team declined to offer such services without getting a financial benefit. In turn, the Taurus team was inflexible and insisted on receiving the services for free. As a result, both teams eventually turned away from the IS project.

The following excerpt shows that the SLA consisted of hidden (i.e., undefined) dependencies:

Head of Department of Computer Science started to lead the team to implement the project. However, the project implementation came to a standstill when the client organization desired deputation of full time experts by the customer organization to supervise the implementation which included training of the end users to use the system and subsequently adopt it. Vendor expressed their inability to depute an expert without charging further expenditure to customer (Nauman et al., 2009, p.271).

Team Taurus asking for unanticipated support from team Aries indicated an undefined intermodular dependency. The SLA specifying the dependencies between team Taurus and Aries is, therefore, said to be poorly defined and contains hidden dependencies. As such, requirement #1 was not met.

2. **Analysis #2:** No Standard Procedures and Definitions

A good analysis of IS system requirements (and thus a well-designed IS system) is critical to successful IS project outsourcing. Adequate analysis of IS system requirements should result in a well-designed IS system. As far as the Taurus organization is concerned, a single IS should serve all the departments of this organization with a standardized operating procedure. The specification of IS system requirements in such a context becomes challenging when different departments currently use different operating procedures for similar tasks.

a. Identifying the modular structure

Different departments of the Taurus organization are conceived as representing different modules. Different departments followed different procedures for the same tasks, and for some particular tasks there were no standard operating procedures. Using modularity terminology, the different departments were using different design options. In this part of the analysis the focus is on the ‘design options’ aspect of the concept of modularity.

b. Modularity requirements

As explained above, a (well-chosen) ‘design option’ (characterizing standardized operating procedures) should become a ‘design rule’. Knowing the design rule(s) chosen by organization Taurus, team Aries is able to conduct, with a high degree of reliability, an IS system’s requirements analysis. When adequately designed, the IS system should be able to serve all the common processes in the different Taurus departments. An unattractive alternative is not to decide on the design rule(s), and design a customized IS for each Taurus department. Obviously, both the cost and complexity of the unattractive alternative is

higher. As far as analysis #2 is concerned, the ex ante requirement for the “design option” aspect of the concept of modularity is as follows: “The Taurus organization should select a design rule from the existing design options or, alternatively, a new design option to serve all the common business processes throughout the departments. All departments should adhere to the design rule(s) selected”. Adherence to design rules is referred to as modularity requirement #2.

c. Assessing the modularity requirements

Team Aries was described as a “very competent and well reputed provider as it was one of the leading independent companies working as a business unit of a large and reputed international company” (Nauman et al., 2009, p. 269). Despite being deemed very competent and reputed, team Aries realized too late (during the testing phase) that for some processes Taurus organization had no standard operating procedures, and for the same business processes, different departments were following different operating procedures (Nauman et al., 2009).

Nauman et al. (2009) also diagnosed Taurus’ organizational processes to be inadequate to embrace automation and needed re-engineering. According to Nauman and colleagues this inadequacy was one of the main reasons why this IS outsourcing project failed (Nauman et al., 2009). Just one example of a faulty data definition is: “the pattern [i.e., the structure] of student registration numbers varied in different departments. Such anomalies caused some requirements changes, even at the later phases, and delayed the implementation” (Nauman et al., 2009, p. 271). Due to different operating procedures adopted by different departments the project was delayed. The above excerpts indicate that Taurus organization failed to select a design rule from a range of alternative design options. As the Taurus organization failed to decide on a design rule it is clear that modularity requirement #2 was not met.

3. **Analysis #3:** As the people changed, the mindset changed and the vision about the project also changed

Change is inevitable within organizations and accommodating change poses a challenge. It was observed that the Taurus team composition changed three times over the course of the project (fig.1). In analysis #3, non-technical root causes of failure are dealt with. The team’s composition can be interpreted and explained in terms of modular structures (Huysmans et al., 2014). As Terlouw (2011, p. viii) states, “modules can comprise humans and/or software systems”. In addition, Dietz (2006) proposed a method to identify modular actor role structures (Huysmans et al., 2014) and thereby asserts that “an enterprise is constituted by the activities of actor roles, which are elementary chunks of authority and responsibility, fulfilled by subjects” (Dietz, 2006, p. 81). As being said, the Taurus team composition changed; some members were replaced by new members. To minimize the impact of changes within the Taurus team, management appointed a software engineer to lead the Taurus team, and expected him to coordinate with the Aries team, up till the end of the project. Unfortunately, almost at the end of the project, during the IS testing phase, this team leader also left for another job and the head of the department of computer science became the new team leader.

Table 1. Different compositions of the Taurus focal team

Team Composition	
1	The Taurus focal team comprises senior faculty members from different departments; the team leader was one ‘Head of department’ (1 st team leader).
2	Due to some routine and policy decisions some team members were replaced by new team members.
3	A software engineer was appointed as new team leader (2 nd team leader).
4	(During the IS testing phase) the team leader left for another job and was replaced by the ‘Head of the computer science department’ (3 rd team leader).

Each change of Taurus' team composition caused some kind of hindrance to the project. Since the Taurus team was previously considered as a module, the next step of the analysis concentrates on the sub-modular level. This illustrates the usage of modularity as a hierarchical concept.

a. Identifying the modular structure

The module 'Taurus organization', as a first level of modularity, can be further subdivided in two (hierarchically nested) levels of submodules. A first submodular level refers to the Taurus 'focal team', and the second submodular level refers to an individual team member, for instance a member of the Taurus 'focal team'. In analysis #3, the Taurus organization and the Taurus 'focal team' are considered to represent a module and sub-module, respectively. The analysis is at sub-modular level, and the analysis focuses on the 'substitution' operator which is part of the modularity concept (i.e., a modular operator).

b. Modularity requirements

In a well-designed modular IS, applying the modular operator 'substitution' should not impact the existing structure negatively. Terlouw (2011, p. viii) asserts that "the modular operators are the actions that may change existing structures in a well-defined way in order to enhance the efficiency of the system". Huysmans, et. al, (2014, p. 4418) suggest that "substituting an older version of module with the newer version should ameliorate the overall performance of the system". The substitution modular operator can be applied successfully and relatively easily if all module versions adhere to the same interface and no undocumented intermodular dependencies are present. If the interface is changed, or the dependencies of the modules are not made explicit, the application of the substitution modular operator is not without risk; one risks that applying the substitution operator disrupts the working of the system, and may trigger ripple effects.

When considering a team as a modular structure, the interface includes the boundary of authority. Consequently, these boundaries should be made explicit and should not change when the module is substituted. Restricting or freezing the boundaries of each module will facilitate to identify any hidden dependencies between the modules and also safeguard the integrity of the modular architecture. A modular architecture describes as "a predefined set of prescriptive rules which all modules of the system need to adhere to" (Huysmans, et al., 2014, p. 4418). This definition clearly indicates that once the modular architecture is designed, the modules (members of the focal team) are obliged to adhere to the predefined rules. The modules should remain within their defined boundary and should not overstep it. So, the boundary of the modules of focal team should be defined clearly and the integrity of the modular architecture should remain unchanged. This condition is considered as modularity requirement #3.

c. Assessing the modularity requirements

In this case it has been observed that the composition of the sub-module "*focal team of Taurus*" has changed three times (fig.1). Using modularity terminology, this means that the sub-module has been substituted three times. Initially, the leader of the focal team only acted as a facilitator in order to communicate effectively and efficiently with the vendor 'Aries'. Therefore, the focal team leader has no authority to make any changes to the project conditions which were already agreed upon by all the stakeholders. According to the concept of modularity, modules are supposed to be free of any hidden dependencies and it does not threaten the integrity of the modular architecture.

In reality, this was obviously not the case. Whenever another version of the focal team was put in place, already established agreements between both parties (i.e. requirements specification was developed and agreed upon by the customer and the vendor) tended to change. Substituting sub-module "*focal team of Taurus*" with newer

versions was negatively affecting the efficiency of the project which can be observed by the following excerpts: “Changes at the organizational level ... led to some new requirements emerging from nowhere and caused frequent changes in the old requirements...The new members of the focal team were not clear about the scope and objectives of the project... Due to this kind of divisive environment, a huge time was lost in the advancement of the project” (Nauman et al., 2009, p. 270). The above excerpts illustrate that applying the modular operator “substitution” resulted in chaos and the modules did not remain within their defined (allocated) boundary. This led to the loss of integrity of the modular architecture. Therefore, in this situation and in the context of this case, it can be concluded that modularity requirement #3 was not met.

FUTURE RESEARCH DIRECTIONS

This is the 2nd IS outsourcing case analysed by the authors using the lens of modularity. These two findings may not be enough to generalize but

in the future, additional case analyses might lead to theory building and a set of hypotheses can be derived from it. In future studies, the newly emerged hypotheses can be tested quantitatively.

CONCLUSION, CONTRIBUTIONS, AND LIMITATIONS

IS project outsourcing became an important business strategy for many organizations. Unfortunately, despite academics’ and practitioners’ efforts made, failure rates of such projects remain far too high. This paper goes beyond the remedies proposed by academics and practitioners, which were partially effective at best, and offered a new perspective based on the concept of modularity. To the best of the authors knowledge, the concept of modularity has never been explored in the context of IS project outsourcing failure. Based on the information from a documented case study (i.e., an outsourced IS project that failed) and reliance on this new perspective this paper shows that specific factors that have contributed to IS project failure could have been identified ex ante.

Table 2. Summary of the findings

	Analyses 1	Analyses 2	Analyses 3
Modular Structure	The Aries team (vendor) and the focal team of Taurus (customer) are conceived as modules.	Different departments of the Taurus organization are conceived as modules.	The Taurus organization is conceived as a module and the Taurus focal team is conceived as submodule. In analogy, every member of the Taurus focal team is conceived as (sub) submodule. This analysis is at the sub-module level.
Modularity Aspects	Dependencies (hidden)	Design options	Modular operator “substitution”
Modularity Requirements (ex ante)	All intermodular dependencies should be made explicit, no hidden dependencies are allowed.	The Taurus organization should set a design rule to serve all the common business processes throughout the departments. All departments should adhere to the design rules.	The substitution operator can be applied successfully and relatively easily if the intermodular dependencies are explicit (i.e. no hidden dependencies are allowed)
Violations	All intermodular dependencies between two modules (Aries and the focal team of Taurus) were not identified ex ante. The SLA was poorly defined and consisted of undefined or hidden dependencies.	The Taurus organization had no standard operating procedures; different departments employed different operating procedures.	Substituting the older version of submodule Taurus focal team with a newer version negatively impacted the efficiency of the IS project. Applying the modular operator “substitution” was very difficult and most likely unsuccessful.

Identified factors related to a poor formulation of system requirements and violations of design principles underlying the modularity concept. This study is of practical use as insights in these factors that contributed to IS project failure are relevant to practitioners (e.g., project managers, software developers). The study is also relevant from a scientific point of view as the concept of modularity (in relation to IS project outsourcing) has been discussed in the scientific literature (see, for instance, Fixson, Ro, & Liker, 2005, p.167) but most of those literature are descriptive in nature. The current study adds to this scientific literature by going a step further, more specifically, the current study succeeded in illustrating how outsourcing projects can reap the benefit by applying ex ante the concept of modularity.

ACKNOWLEDGMENT

The authors are indebted to Prof. Dr. Alain De Beuckelaer (Institute for Management Research, Radboud University, Nijmegen, Netherlands) due to his contribution in improving the quality of this paper.

REFERENCES

- Bandara, W., Miskon, S., & Fielt, E. (2011). A systematic, tool-supported method for conducting literature reviews in information systems. *Proceedings of the 19th European Conference on Information Systems (ECIS 2011)*.
- Ciric, Z., & Rakovic, L. (2010). Change Management in Information System Development and Implementation Projects. *Management Information Systems*, 5(2), 23–28.
- Dietz, J. L. G. (2006). *Enterprise Ontology: Theory and Methodology*. New York: Springer. doi:10.1007/3-540-33149-2
- Fixson, S., Ro, Y & Liker, J. (2005). Modularity and Outsourcing: Who drives whom? A Study of Generational Sequences in the U.S. Automotive Cockpit Industry. *International Journal of Automotive Technology and Management*, 5(2), 166-183.
- Gay, C. L., & Essinger, J. (2000). *Inside Outsourcing: The insider's guide to managing strategic sourcing*. London, UK: Brealey.
- Greaver, M. F. II. (1999). *Strategic Outsourcing: A Structured Approach to Outsourcing Decisions and Initiatives*. New York: Amacom.
- Harris, M. D., Herron, D. E., & Iwanicki, S. (2008). *The Business Value of IT: Managing Risks, Optimizing Performance and Measuring Results*. CRC Press. doi:10.1201/9781420064759
- Huysmans, P., De Bruyn, P., Benazeer, S., De Beuckelaer, A., De Haes, S., & Verelst, J. (2014). On the Relevance of the Modularity Concept for Understanding Outsourcing Risk Factors. In *Proceedings of the 47th Hawaii International Conference on System Sciences* (pp. 4416-4425). IEEE. doi:10.1109/HICSS.2014.544
- Huysmans, P., & Verelst, J. (2013). Towards an engineering-based research approach for enterprise architecture: lessons learned from normalized systems theory. In *Advanced Information Systems Engineering Workshop: Proceedings of the 25th International Conference on Advanced Information Systems, Engineering (CAiSE)*, (pp. 58-72). Berlin, Germany: Springer. doi:10.1007/978-3-642-38490-5_5
- Kotlarsky, J., Oshri, I., & Willcocks, L. P. (2011). *The Handbook of Global Outsourcing and Offshoring*. Palgrave Macmillan.
- Lambert, D. M., Emmelhainz, M. A., & Gardner, J. T. (1999). Building Successful Logistics Partnerships. *Journal of Business Logistics*, 20(1), 165–181.

- Langlois, R. N. (2002). Modularity in technology and organization. *Journal of Economic Behavior & Organization*, 49(1), 19–37. doi:10.1016/S0167-2681(02)00056-2
- Mannaert, H., & Verelst, J. (2009). *Normalized Systems: Re-creating Information Technology Based on Laws for Software Evolvability*. Antwerp, Belgium: Koppa Publications.
- Martineau, M., & Shumway, A. (2009, July 1). Beat the Odds: Making your IT projects a success. *New Dawn Technologies*. Retrieved from <http://www.newdawn.tech.com/webinar/beattheodds.pdf>
- Modarress, B., Ansari, A., & Thies, E. (2014). The Impact of Technology Transfer through Foreign Direct Investment in Developing Nations: A Case Study in the United Arab Emirates. *International Journal of Economics & Finance*, 6(7), 108–126. doi:10.5539/ijef.v6n7p108
- Nauman, A. B., Aziz, R., & Ishaq, A. F. M. (2009). Information System Development Failure and Complexity: A Case Study. In M. G. Hinter (Ed.), *Selected Reading on Strategic Information Systems* (pp. 251–275). IGI Publications. doi:10.4018/978-1-60566-090-5.ch017
- Peterson, B. L., & Carco, D. M. (1998). *The Smart Way to Buy Information Technology: How to Maximize Value and Avoid Costly Pitfalls*. New York: AMACOM.
- Rottman, J. W. (2008). Successful Knowledge Transfer within Offshore Supplier Networks: A Case Study Exploring Social Capital in Strategic Alliances. *Journal of Information Technology*, 23(1), 31–43. doi:10.1057/palgrave.jit.2000127
- Schilling, M. A. (2000). Towards a General Modular Systems Theory and its Application to Inter-Firm Product Modularity. *Academy of Management Review*, 25(2), 312–334.
- Simon, H. A. (1962). The Architecture of Complexity. *Proceedings of the American Philosophical Society*, 106(6), 467–482.
- Terlouw, L. I. (2011). *Modularization and Specification of Service-Oriented Systems*. Delft Technical University. Delft, The Netherlands: Gildeprint.
- Verelst, J., Silva, A. R., Mannaert, H., Ferreira, D. A., & Huysmans, P. (2013, May). *Identifying combinatorial effects in requirements engineering*. Paper presented at the Advances in Enterprise Engineering VII: Third Enterprise Engineering Working Conference (EEWC'13), Luxembourg. doi:10.1007/978-3-642-38117-1_7
- Webster, J., & Watson, R. T. (2002). Analyzing the Past to Prepare for the Future: Writing a Literature Review. *Management Information Systems Quarterly*, 26(2), xiii–xxiii.

KEY TERMS AND DEFINITIONS

Coupling: Coupling is a measure of the dependencies between modules.

Design Rule: A well-chosen ‘design option’ (characterizing standardized operating procedures) should become a ‘design rule’. Compatibility among modules is ensured by “design rules”. Sometime modules are designed independently but at the end these modules should work as a whole. Standard dimensions and parameters or specifications (i.e., dimensions, tolerances, functionality, dependency etc.) constitute design rules.

Interface: An interface is a common boundary where direct contact between two modules occurs and where these two modules communicate with each other.

Modular Dependency: Dependency is the degree to which a module relies on other modules in order to function.

Modularity: Modularity is defined as a property of a complex system, whereby the system is decomposed into several subsystems (i.e., modules).

Module: A module is a unit whose internal parts are powerfully connected among themselves and relatively weakly connected to other units.

Developing Global Supply Chain Manager for Business Expansion



Puspita Wulansari

Telkom University, Indonesia

Yudi Fernando

Universiti Malaysia Pahang, Malaysia

INTRODUCTION

Globalization makes market becomes more competitive since nowadays the problem of geographical boundaries which hinders marketing of products and services can be overcome more easily. With the help of technological advancements, such as the internet and logistics, products and services can be marketed and consumed in any part of the world rather easily. Customers can freely choose the products and services based on their preferences on price, quality, or mode of acquisition (Morgan, 2003).

To win the competition in the global market, it is not enough for a company to operate on local level alone. Kiessling (2013) and Flint (2004) explain that with the global market, the products and services provided by MNCs can reach closer to customer value from all parts of the world. Compared to selling the product locally, global market shares are indeed much bigger. It has to operate globally, sell their products and services internationally. As an example, three brands which are produced by local companies but are already known as an international product, Nestle Coffee, Levi's jeans, and Carrefour, have to compete with superstores such as Wal-Mart (Hammond and Grosse, 2003). Local companies have to think globally and slowly transform themselves to become Multinational Corporates (MNCs) (Hammond and Grosse, 2003; De Chiara and Spane, 2011). Multinational companies operate in multiple countries, and they move products,

services, cultural styles, and economic power from one country to the others (Hammond and Grosse, 2003). For example, a German multinational engineering and electronics company, Robert Bosch GmbH, has established more than 350 subsidiaries across over 60 countries. Unilever, a London-based multinational consumer goods company, has created subsidiaries in at least 90 countries (Nguyen et al., 2015).

One of many ways for MNCs to be competitive is that the firm has to produce products and services which meet customers' preferences, which are: having high quality with affordable price (Griffith, 2006). MNCs which operate in multiple countries can get 4 location advantages by utilizing different characteristic from each geography, which are: resources, efficiency, market, and strategic asset (Dunning, 1998; De Chiara and Spane, 2011). MNCs have to think and find another alternative of raw materials with affordable cost from other countries. They also have to consider ways to shorten their time cycle to deliver products to customers and to minimize logistics cost. In addition, as an effort in terms of resources and efficiency, they have to find operating locations which have lower labor and resource cost (De Chiara and Spane, 2011; Christopher et al., 2006).

The needs to think globally and act locally can be materialized by utilizing supporting concept of global supply management. The global supply chain in MNCs will force companies to pay attention to supply chain matrix to achieve cost

reduction, better time utilization in production line, faster shipment, cheaper labors and satisfaction (Flint, 2004; Mollenkopf et al., 2010; Manuj and Mentzer, 2008).

The dynamics in business environment and buildup of the global market force the parties in supply chain to be equipped with different skills and competencies to facilitate them to manage the larger scope of global supply chain successfully (Hohenstein et al., 2014). A global supply chain manager's role is to handle the workflow of the materials and products across domestic and global markets. Capabilities and competencies to manage resources globally play a key role to enhance supply chain management (Keller & Ozment, 2009).

Earlier studies show that the effectiveness of supply chain management operations are disrupted by unclear goal setting implementation from an organization, and this can be improved through better human resources (Gowen & Tallon, 2003). According to Ellinger and Ellinger (2014); and Kiesling et al. (2004), supply chain management role has been changed dramatically in terms of required skill set, knowledge, and capabilities for them to be able to connect and run a global operation.

The importance of a global manager's role to support global supply chain management in MNCs is very critical. Earlier researches in understanding the importance of human resources in supply chain management have been conducted, but a specific research in global supply chain management has not yet been conducted—to be specific, on *building* global supply chain management in MNCs—only a few scholars conduct this before. Within this chapter, we will deepen the analysis on the competency required by a global supply chain manager in relation to his role to support business expansion globally, based on the previous study which has been published in the academic journals.

This chapter is structured as follows. The next section presents a short background on the growing importance of global supply chain managers. In the following section, in the main focus of this

chapter, previous researches about global supply chain management and its intersections with human resources management will be discussed. After the problem has been identified, the next chapter is designed to digest further the solution and recommendation, followed by future research direction, and to be closed with a conclusion.

BACKGROUND

In the beginning, supply chain management was only established in local companies. With globalization coming along the way, companies need to develop strategies related to supply chain management. Global supply chain management will allow company to utilize benefiting factors of varied location globally, for examples: level product and process technology expertise, labor force capabilities, input factors cost, local tax rates, and the capabilities offshore vendors (Cohen and Malik, 1997).

Global supply chain management will deal with varied interests from business management and supplier globally rather than locally, thus it requires manpower which can supervise effectively all of the activities of global supply chain and can collaborate internationally (Harvey et al., 2013). Involvement of human resources on the success implementation of global supply chain management is an interesting topic which has not been studied widely by researchers. It is well known that human resources can boost company's performance through strategic management of the knowledge, skill and HR capabilities. Human resources managements in general are series of activities such as recruitment, selection, training, performance management, compensation, labor relations to manage staff of the organization (Ivancevich, 2003). With the global world development, MNCs are required to manage human resources management in much more sophisticated ways. Compared to local companies, they have to ensure the recruited staff is the most talented staff

with the right knowledge, skill set and motivation to be positioned across company branch which may be located far away from where they belong to (Hohenstein et al., 2013).

Having a global manager who can establish good relations is considered to be one of the most important benefits in this competitive global market. As a result, companies do not consider customer and supply chain as a result from business action but rather as an asset which is needed to be developed and optimized (Harvey et al., 2013). Harvey and Richey (2001) explains that the perspective based on the competencies explicitly can overcome the dynamics of the global environment by understanding since the beginning the competencies owned by a global supply chain management during the recruitment process and followed by development after onboard. Competencies development is expected to enable the global supply chain management to define the strategic objective. By identifying the importance of global leadership competencies, corporates might be able to shorten the long period of time for developing global leaders. Thus, identifying global leadership competency model must be an urgent priority for the company so that they can adapt and survive in a changing global environment (Kim and McLean, 2015).

For further understanding, below are definitions for key processes on this chapter, such as supply chain management, global supply chain management, and human resource management, as they are defined by the experts.

Supply Chain Management

According to Mentzer et al. (2001), supply chain management is defined “as a systemic, strategic coordination of the traditional business functions and the tactics across these business functions within a particular company and across businesses within the supply chain, for the purpose of improving the long-term performance of the individual companies and the supply chain as a whole.”

Global Supply Chain Management



Definition of global supply chain management is “a network of facilities, activities, and social relationship that performs a multitude of integrated value-adding function in a global context (Mabert and Venkataramana, 1998 in Harvey and Richey, 2001). Or can be defined beyond the range of functional and relational activities undertaken by the downstream/upstream stakeholders in the supply chain (Harvey and Richey, 2001).

Human Resource Management

According to Ivancevich (2003), human resource management is “the function performed in organizations that facilitates the most effective use of people (employees) to achieve organizational and individual goals.” Meanwhile Dessler (2013) stated that human resource management is “the process of acquiring, training, appraisal, and compensating employees, attending to their labor relations, health and safety, and also fairness concerns.” Human resource management activities consists of numerous activities, which are:

1. Equal employment opportunity (EEO) compliance,
2. job analysis,
3. human resource planning,
4. employee recruitment, selection, motivation, and orientation,
5. performance evaluation and compensation,
6. training and development,
7. labor relations,
8. Safety, health, and wellness (Ivancevich, 2003).

Global Supply Chain Process

With development and expansion made to various country, role of supply chain management and logistic become much more critical factor for companies. MNC need to drive the supply

chain on their international network, where they not only required to manage the distribution of product and service globally, but also to focus on organizational learning by looking through the competitive advantage from the competitor (Kiessling). Therefore, MNC can operate their business as multinational player, where part of the global supply chain process are international manufacturing plants, purchasing and sells on global markets, multinational staffed, executive boards, and multicultural personnel (Hulsmann et al., 2008).

However, implementing the global supply chain is not simple. There are numbers of obstacles need to be faced and solved by companies, such as cultural economical, language and legal differences in various countries. The complexity and coordination involved in management, and geographical distance from one area to another not only costing on transportation cost, but also on inventory cost (Mollenkopt et al., 2010; Meixell and Gargeya, 2005; Kiessling et al., 2004).

In order running the Global Supply Chain Management process, it takes a huge effort. It needs human resource with certain skill set and competencies, by then they can manage to process from downstream to upstream from various countries. In the same time, companies need to have the capabilities to develop global supply chain managers. In order to develop global supply chain managers, it is important to understand the required competency for talents to be able to execute his role successfully in the global market.

Scholar has been conducting the research to understand the role of human resource management and global supply chain management as well as the competencies needed to become a global supply chain manager.

More specifically, Harvey and Richey (2001) mentioned that global supply chain managers need to possess;

1. Input competencies, i.e. capital, labor, and physical assets.

2. Managerial competencies, i.e. social managerial knowledge, informal business networks both internal and external.
3. Transformation-based competencies, management competencies to execute the task, to be in the proper position in the market place, and adaptability in the new organization environment.

They also discussed multiple intelligence for world-wide supply chain managers, which are;

1. Cognitive intelligence;
2. Emotional intelligence;
3. Political intelligence;
4. Cultural/social intelligence;
5. Organizational intelligence;
6. Network intelligence;
7. Innovative intelligence; and
8. Intuitive intelligence.

Meanwhile in further research in 2013, Harvey stated that managers who able to become world-wide supply chain managers are those glopatriate type, because he will be able to live for 10 years in rural areas far from their country.

Prajogo and Sohal (2013) tried to understand the competencies and skillset required by SC managers to manage global supply chain effectively. Communication and team works are the most critical competencies and skillset required by supply chain managers in future challenge. Supply chain managers have to own multiple skills to overcome various challenges in the futures, including cross cultural communication and the broad knowledge of technology.

Harvey (2013), Hohenstein, and Hartmann (2014) constitute an urgent demand to establish a global mindset in implementing the global supply chain in MNCs. The importance of global perspective is open mindedness and understanding of cultural diversity as a basis for global supply chain managers to be able to work in the midst of increasing globalization and growing target

market. Additionally Kiessling et al. (2004) adds that intangible asset is more needed compared to tangible assets for future strategic global success. The intangible assets that must be retained by the managers are global awareness, corporate strategy, cultural empathy, cross-cultural team building, global negotiation skill, and self-assurance.

Competencies are required by global supply chain manager according to researches that have been conducted. It is clear that the specific need related with global supply chain manager was not detailed. Study conducted by Harvey and Richey is already mentioned about the competencies of global supply chain manager, unfortunately the further detailed analysis or the scientific research using the quantitative was not available. Therefore, it is difficult to measure the requirement and competencies level. Prajogo and Sohal (2014) discussed the detail about the competency, but the studied competencies was combined with skill set, which are two different things. Therefore, further study is required to understand the required competency of global managers in managing the SCM in MNCs. Followed with empirical test to prove the hypothesis, so we can understand the most important competency required by global supply chain managers.

SOLUTIONS AND RECOMMENDATIONS

Based on the details above, it can be concluded that there is limited availability of study on the relation between HRM and SCM in the global company (global supply chain management). To be specific, competencies and skillset that are required to develop a global supply chain manager. The existing study results are limited to part of HRM – skill knowledge and abilities, selection, training and development, performance, and also education and teaching – while the other HRM activities related with recruitment, compensation,

labor relation, health and fair concern also global mindset were not studied in detail.

Studies that have been conducted merely done by the literature review or method qualitative, while the empirical studies are still very limited. The only quantitative study that has been conducted was done by Pajogo and Sohal (2014), but stop on the descriptive stage while the hypothesis testing with the quantitative complex has not yet been conducted.

To be global managers and MNCs are certainly not an easy job, it requires certain skill and competency, glopatriate, and global mindset, with that managers can adapt well in various cultures where companies operate. In the other hand, the management must ensure the compensation and the benefit package for managers to freely take the decision responsibly as global managers.

FUTURE RESEARCH DIRECTIONS

Based on discussion in this chapter, competencies that is required by global supply chain managers have been defined. However, deeper study is still need to be conducted on required competencies and skill set required by global supply chain managers.

Previous studies that have been conducted merely focus on literature review or qualitative study, there is no study yet which asses the required competency by empirical study or quantitative.

CONCLUSION

In this globalization era, companies are required to be able to compete in global level, in order to do that – companies have to produce products/ services with high quality and affordable price that affordable and companies operate under global perspective to gain cheaper raw materials, produce with lower cost, cheap labor, and distribute the products/services to all over part of the world with



competitive logistic cost and on time delivery. This can be happened only with the existence of global supply chain management which managed by global supply chain managers equipped with specific skillset, knowledge, competencies, abilities, training, development, and series of HR activities that.. Critical success factor for MNCs relies on the tangible and intangible asset, but the intangible asset contribution is much bigger.

Global supply chain managers are required to have competencies and skillset. Based on the details analysis on this chapter, we define several critical competencies of managerial competencies such as social knowledge, communication, and team works, and global awareness. Transformation-based competencies, and capability to adapt in various different environments.

REFERENCES

- Christopher, M., Peck, H., & Towill, D. (2006). A taxonomy for selecting global supply chain strategies. *The International Journal of Logistics Management*, 17(2), 277–287. doi:10.1108/09574090610689998
- Cohen, M. A., & Mallik, S. (1997). Global supply chains: Research and applications. *Production and Operations Management*, 6(3), 193–210. doi:10.1111/j.1937-5956.1997.tb00426.x
- De Chiara, A., & Russo Spena, T. (2011). CSR strategy in multinational firms: Focus on human resources, suppliers and community. *Journal of Global Responsibility*, 2(1), 60–74. doi:10.1108/20412561111128528
- Ellinger, , A., & Ellinger, A. (2013). Leveraging human resource development expertise to improve supply chain managers skills and competencies. *European Journal of Training and Development*, 38(1/2), 118–135. doi:10.1108/EJTD-09-2013-0093
- Flint, D. J. (2004). Strategic marketing in global supply chains: Four challenges. *Industrial Marketing Management*, 33(1), 45–50. doi:10.1016/j.indmarman.2003.08.009
- Flynn, B. B., Huo, B., & Zhao, X. (2010). The impact of supply chain integration on performance: A contingency and configuration approach. *Journal of Operations Management*, 28(1), 58–71. doi:10.1016/j.jom.2009.06.001
- Gowen Iii, C. R., & Tallon, W. J. (2003). Enhancing supply chain practices through human resource management. *Journal of Management Development*, 22(1), 32–44. doi:10.1108/02621710310454842
- Griffith, D. A. (2006). Human capital in the supply chain of global firms. *Organizational Dynamics*, 35(3), 251–263. doi:10.1016/j.orgdyn.2006.05.004
- Hammond, C., & Grosse, R. (2003). Rich man, poor man: Resources on globalization. *RSR. Reference Services Review*, 31(3), 285–295. doi:10.1108/00907320310486881
- Harvey, M., Fisher, R., McPhail, R., & Moeller, M. (2013). Aligning global organizations human capital needs and global supply-chain strategies. *Asia Pacific Journal of Human Resources*, 51(1), 4–21. doi:10.1111/j.1744-7941.2012.00054.x
- Harvey, M. G., & Richey, R. G. (2001). Global supply chain management: The selection of globally competent managers. *Journal of International Management*, 7(2), 105–128. doi:10.1016/S1075-4253(01)00040-0
- Hohenstein, N. O., Feisel, E., & Hartmann, E. (2014). Human resource management issues in supply chain management research: A systematic literature review from 1998 to 2014. *International Journal of Physical Distribution & Logistics Management*, 44(6), 434–463. doi:10.1108/IJPDLM-06-2013-0175



Hülsmann, M., Grapp, J., & Li, Y. (2008). Strategic adaptivity in global supply chains—competitive advantage by autonomous cooperation. *International Journal of Production Economics*, 114(1), 14–26. doi:10.1016/j.ijpe.2007.09.009

Ivancevich, J. M. (2003). *Human resource management* (9th ed.). Mc Graw Hill.

Keller, S. B., & Ozment, J. (2009). Research on personnel issues published in leading logistics journals: What we know and don't know. *The International Journal of Logistics Management*, 20(3), 378–407. doi:10.1108/09574090911002832

Kiessling, T., Harvey, M., & Garrison, G. (2004). The importance of boundary-spanners in global supply chains and logistics management in the 21st century. *Journal of Global Marketing*, 17(4), 93–115. doi:10.1300/J042v17n04_06

Kim, J., & McLean, G. N. (2015). An integrative framework for global leadership competency: Levels and dimensions. *Human Resource Development International*, 18(3), 235–258. doi:10.1080/13678868.2014.1003721

Manuj, I., & Mentzer, J. T. (2008). Global supply chain risk management strategies. *International Journal of Physical Distribution & Logistics Management*, 38(3), 192–223. doi:10.1108/09600030810866986

Meixell, M. J., & Gargeya, V. B. (2005). Global supply chain design: A literature review and critique. *Transportation Research Part E, Logistics and Transportation Review*, 41(6), 531–550. doi:10.1016/j.tre.2005.06.003

Mentzer, J. T., DeWitt, W., Keebler, J. S., Min, S., Nix, N. W., Smith, C. D., & Zacharia, Z. G. (2001). Defining supply chain management. *Journal of Business Logistics*, 22(2), 1–25.

Mollenkopf, D., Stolze, H., Tate, W. L., & Ueltschy, M. (2010). Green, lean, and global supply chains. *International Journal of Physical Distribution & Logistics Management*, 40(1/2), 14–41. doi:10.1108/09600031011018028

Morgan, H. (2003). The challenges of a global marketplace. In M. Goldsmith, R. Gandossy, & M. Effron (Eds.), *Human resources in the 21st century*. New York: Wiley.

Motwani, J., Larson, L., & Ahuja, S. (1998). Managing a global supply chain partnership. *Logistics Information Management*, 11(6), 349–354. doi:10.1108/09576059810242453

Nguyen, P., Felfe, J., Fookien, I., & Ngoc, H. T. (2015, August). Commitment to a parent company and a local operation: A comparison between local employees and Western expatriates working for multinational companies. In *Evidence-based HRM: a Global Forum for Empirical Scholarship* (Vol. 3, No. 2, pp. 181–204). Emerald Group Publishing Limited. doi:10.1108/EBHRM-05-2013-0013

Prajogo, D., & Sohal, A. (2013). Supply chain professionals: A study of competencies, use of technologies, and future challenges. *International Journal of Operations & Production Management*, 33(11/12), 1532–1554. doi:10.1108/IJOPM-08-2010-0228

ADDITIONAL READING

Bücker, J., & Poutsma, E. (2010). Global management competencies: A theoretical foundation. *Journal of Managerial Psychology*, 25(8), 829–844. doi:10.1108/02683941011089116

Cappellen, T., & Janssens, M. (2008). Global managers career competencies. *Career Development International*, 13(6), 514–537. doi:10.1108/13620430810901679

Fisher, R., McPhail, R., You, E., & Ash, M. (2014). Using social media to recruit global supply chain managers. *International Journal of Physical Distribution & Logistics Management*, 44(8/9), 635–645. doi:10.1108/IJPDLM-07-2013-0179

Gabrielsson, P., Gabrielsson, M., Darling, J., & Luostarinen, R. (2006). Globalizing international: Product strategies of ICT manufacturers. *International Marketing Review*, 23(6), 650–671. doi:10.1108/02651330610712157

Jeou-Shyan, H., Hsuan, H., Chih-Hsing, L., Lin, L., & Chang-Yen, T. (2011). Competency analysis of top managers in the Taiwanese hotel industry. *International Journal of Hospitality Management*, 30(4), 1044–1054. doi:10.1016/j.ijhm.2011.03.012

Jokinen, T. (2005). Global leadership competencies: A review and discussion. *Journal of European Industrial Training*, 29(3), 199–216. doi:10.1108/03090590510591085

Joshi, A., & Lazarova, M. (2005). Do “global” teams need “global” leaders? Identifying leadership competencies in multinational teams. *Managing multinational teams: Global perspectives*, 18, 281–302.

KEY TERMS AND DEFINITIONS

Globalization: A network of contemporaneous events, options, and constraints, which requires the development of a systematic concept of supply chain strategy development and implementation.

Supply Chain Management: The integration of all the activities and processes associated with the flow of goods and information from the raw materials stage to the end consumer of the product/service.

Global Supply Chain Management: Network of facilities, activities, and social relationships that performs a multitude of integrated value-adding function in a global context.

Global Manager: A “leader” which is a group member whose influence on group attitudes, performance, or decision making greatly exceeds that of the average member of the group.

Multinational Company: A mix of competitive and comparative advantage, global advantage as the basis rather than geographical locations of the organization’s activities and the integration of activities across locations.

Discrete Event Simulation in Inventory Management



Linh Nguyen Khanh Duong

Auckland University of Technology, New Zealand

Lincoln C. Wood

University of Otago, New Zealand & Curtin University, Australia

INTRODUCTION

Successful supply chain management aims to deliver the right products at the right time to the right place and in the right condition (Deniz, Scheller-Wolf, & Karaesmen, 2004). This is not a simple task, and many factors influence the success of the supply chain; perishability and substitutability characteristics of inventory are key attributes that cannot be ignored. Once produced, perishable products have a finite shelf life. When expired, they are either partially or wholly valueless. Perishability affects many industries (e.g., fresh food and chemicals). The more time that perishable inventory is in storage, the less time it is available for sale to customers. The combination of these factors often requires simulation models to be developed to understand the behavior of the system as the parameters change.

Product substitution is a possibility when considering multiple products. Research indicates that an alternative product is willingly chosen by customers if the preferred one is out of stock, and product substitution is important to companies (Chen, Feng, Kebliş, & Xu, 2015). Research shows that consumer-driven substitution due to product stock-outs common in the grocery industry (Bijvank & Vis, 2011). In a recent study, the Grocery Manufacturers of America estimated that approximately 70% of consumers who find a particular item is stocked out on the first occurrence will happily purchase another product (Grocery Manufacturers Association, 2015). van Donselaar,

Van Woensel, Broekmeulen, and Fransoo (2006) analyzed these types of situations and suggested that accounting for substitution while establishing inventory control policies could lead to a reduction in waste.

Holding inventory is necessary for a firm to fulfill customer orders; however, holding inventory also incurs holding cost (e.g., providing material storage and insurance). Each product has a holding cost applied to the average inventory level over a specified period, a selling price, and a cost per unit of stock. Managers must decide on the replenishment time and replenishment quantity for each item within product subcategory, to maximize expected profits under uncertain demand while minimizing the instances of running out of inventory (i.e., a 'stock out').

Determining the appropriate replenishment policy that will maximize profit under probabilistic consumer demand is known as stochastic optimization. In stochastic situations, it becomes difficult to formulate models accommodating so many factors. According to Lucas, Kelton, Sánchez, Sanchez, and Anderson (2015), a discrete-event simulation methodology is suitable to capture the dynamics of this problem. Discrete-event simulation involves modeling a system and where a specific event triggers a change in the state of the system. Such simulation allows tracking of specific items of inventory (e.g., when an item 'expires' it would trigger an event and a change in the system state); this is a necessary precondition that makes this type of simulation more appropri-

ate than continuous simulation for the modeling of substitutable and perishable inventory systems. Such simulations can also be used in the evaluation of new IT improvements that can be used to improve collaboration practices over the supply chain (Cannella, Framinan, & Barbosa-Póvoa, 2014). Simulation can incorporate stochasticity and complexity while providing detailed output for further analysis and optimization work.

BACKGROUND

In general, there are four types of perishable products: food items (e.g., meat, vegetables, dairy products, and beverages), medical/pharmaceuticals (e.g., vaccines, blood, and drugs), plants, and industrial/other (e.g., paint and chemicals). Each type may have several categories, subcategories, and product variants. For example, milk products can be divided into powdered milk or ready-to-drink milk products. Ready-to-drink milk products can be further divided into yogurt or drinks; then, by flavors and sizes.

Managers must decide the inventory level for perishable products to ensure customers have all desired products at the right time at minimal disposal cost. Managers want to provide the highest customer service level at the lowest cost. Some of these costs that play a significant role in perishable inventory management include ordering cost, holding cost, disposal cost and shortage/backlog cost. Nahmias (2011) gave a comprehensive review of the perishable inventory theory. There is two primary inventory management approaches, periodic review (monitoring inventory levels at fixed intervals) and continuous review (monitoring inventory levels continually). Periodic review has been widely used for a long time due to the relative simplicity of application; it involves the amount of inventory at a particular point in time and reviewing the inventory on a regular basis after that. For example, a store-person may count the inventory once per week and then calculate how much more inventory is required in the next

order. Contemporary approaches include vendor-managed inventory (VMI) and collaborative planning forecasting, and replenishment (CPFR) processes to improve multi-tier inventory management (Alftan, Kaipia, Loikkanen, & Spens, 2015).

A single period perishable inventory model the 'newsvendor' model; named for the 'newsboy' that would need to buy a stack of newspapers before shifting them to another location for sale to customers. The entrepreneurial newsvendor had to calculate the likely number of newspapers required; if too many were purchased, they would be worthless at the end of the day; if too few were purchased, they would lose potential sales (Nahmias, 2011). In real stochastic environments, where newsvendor models are most useful, managers must pay attention to planning models where there is greater age dependency in demand, consumers are sensitive to these factors, the products have a short shelf-life, and there is a risk of quality issues (Amorim, Costa, & Almada-Lobo, 2013).

To expand market share, many producers/distributors tend to provide more product lines, and many retailers opt to offer new choices for consumers by introducing new brands, more flavors or sizes of existing products, or new products with various attributes. Consequently, this creates additional managerial challenges as policies must account the effects of substitution and perishability. In consumer-driven substitution, the customer's willingness to substitute during a stock out is a major factor; in contrast, decision-maker-driven substitution involves a managerial decision to substitute a given product with a different variant of the product (Broekmeulen, Fransoo, Van Woensel, & van Donselaar, 2007). Many of these factors may be difficult to capture and 'price' in a meaningful way relative to other factors, often requiring some judgment when developing a model that seeks to minimize a total cost function. In contrast, many managerial situations require consideration of a broader range of metrics and KPIs (Cannella, Barbosa-Póvoa, Framinan, & Relvas, 2013; Cannella et al., 2014). This would also enable managers to include metrics that are

inherently more familiar to them and which are more comprehensible (e.g., average inventory levels). Having a relationship between metrics that are commonly used would make replenishment and inventory management models valuable over a wider range of decision-makers. Understanding the components to the model, and how changes to particular metrics they are familiar with will help the organization to achieve wider objectives, will help to influence behaviors at the divisional level, and allow improved alignment over multi-division metrics. Managers will also be able to rapidly and naturally adapt their management style if the replenishment policies are based on meaningful metrics.

Perishability relates to the deterioration of a product over time and is often associated with fresh food; a cut of meat will quickly ‘go bad’ if it is exposed to warmth and sunlight, creating food safety issues for consumers. Globalization has created a big change in consumers’ behavior as it has led to many people to seek superior quality foods, leading to increased trade in food products as some are produced in specialized locations. Supply chain management has become increasingly important as the volumes of perishable products from developing countries (e.g., Vietnam and Mexico) have increased; in particular, export volumes are expected to increase substantially when the Trans-Pacific Partnership is enacted (Executive office of the President, 2015).

A report from the United Nations shows that one-third of food, globally, is lost as waste; the solution to ensuring sufficient food for people around the world is to reduce food waste, rather than focusing on producing more (Scialabba, 2014). Another example is the loss almost blood platelet units (used in blood transfusions) that expire without being transfused - these expired units impose a significant financial burden on health services. Dynamic pricing strategies, influencing consumer behaviors, can improve overall profitability (Li, Zhang, & Tang, 2015); this can often lead to more comprehensive considerations of how the stock is managed and sold, with a focus on maximizing revenue rather than minimizing costs.

Therefore, the application of perishable inventory models provides higher customer service levels and lower waste.

The determination of replenishment and ordering policies to achieve requirements of perishable products is crucial but complex when there are stochastic demands. In such complex situation, stochastic simulation is an effective modeling tool. This chapter illustrates how simulation methods can be applied to the perishable inventory management system and outlines the advantages of using such simulation methods and provides future research opportunities.

THE USE OF SIMULATION IN CREATING POLICIES FOR THE MANAGEMENT OF PERISHABLE INVENTORY

Discrete-event simulation models can be developed to compare the performance of multiple systems. Through advanced search techniques in simulation, feasible solutions can be evaluated to obtain the best solution. This type of solution is not guaranteed to be optimal; however, the longer the search process is run, the higher the probability will be that the obtained result is optimal.

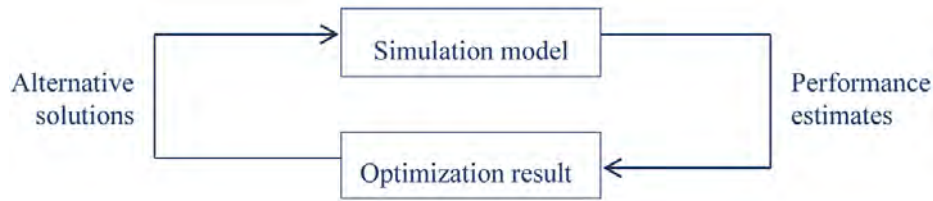
These models must be validated against real businesses where perishability is important (e.g., retailers, supermarkets). This critical step ensures that the model ‘works’ correctly and produces output that matches the output from the real business (Kelton & Sadowski, 2009); such validation provides confidence that the model will then prove to be useful when the parameters are changed.

The first step of a ‘simulation-optimization’ model is the design of simulation phase. There are four major events that occur in inventory management systems that work with perishable and substitutable products (Myers, 2009):

- **Update Inventory Age Event:** Here, the system will update inventory age for all products, update expired products, and schedule inventory aging in the next period.



Figure 1. Simulation: Optimization model



- **Order Event:** Here, the system will set order priority, update total unit order, schedule delivery, and schedule the next order.
- **Generate Demand Event:** Here, the system generates demand for each product and schedules the arrival time for orders in the next period.
- **Sales Event:** Here, in relation to perishability and substitution problem, this event establishes the preferred product and ranks substitute product to determine what sale is made, based on inventory availability.

After major events are defined, a selection of values of the parameters is determined for a range of operation conditions. Then the simulation - optimization model is run multiple times, and outcomes are recorded (Figure 1). The performance statistics are collected after a warm-up period determined by Welch's graphical procedure (Law & Kelton, 2000). The goal is to understand the impact of input parameters such as shelf-life, lead time, and substitution rules on the results. The optimization occurs by varying the parameters and then determining which parameters provide superior outcomes. These values of parameters can then be used by real managers in a real inventory system to achieve desired results.

The simulation - optimization model is designed to find the optimal solution based on the problem of choosing some factors that help to achieve the final goal. For example, the model designed by Myers (2009) aims to maximize the total profit for a retailer under customers' demand uncertainty. The optimal solution is based on choosing the 'order up to' level for each product.

Assumptions

This type of inventory model must account for a range of assumptions that closely mirror the situations found in real inventory systems. These include:

- **Shelf-Life:** The products are assumed to have a constant shelf-life.
- **Demand:** To use a simulation - optimization approach, using a discrete distribution is preferred.
- **The Profit Margin:** It is assumed that the profit margin for each product is the same although the price and unit cost can vary for each product.
- The replenishment lead time is deterministic.
- All products arrive into the system as 'new' products.
- Substitution is considered to be consumer-driven. For example, the customer considers to buy an apple with first preference is A brand, the second preference is B brand. The customer will look for fresh A brand first. If there is no fresh A brand, the customer will choose among four alternatives below:
 - Substitute with older A brand product if it is available.
 - Substitute with fresh B brand product if it is available.
 - Substitute with old B brand if no A brand is available, and no fresh B brand is available.
 - Decide not to buy either A or B brand product.

Model Verification and Validation

Verification ensures that the model is working properly and the logic of the model is correct. Validation then ensures that the model is an accurate representation of the real system (Sargent, 2013). The results of the model are compared with the actual system results from the past. Then the model results are discussed with managers to validate the model. Where discrepancies are observed critical errors in the model can be corrected. When a model is considered to be verified and validated, it can be used to help generate solutions to the managerial problems.

Model Results

The results from the simulation provide insight into the system performance. It shows how a simulation model helps to analyze the system's performance and establish solutions that will contribute to improving inventory system performance. The model can be run numerous times with a range of parameters. Each time some performance measures are calculated. These are collected over the numerous iterations of the model and are presented in graph and tabular forms. This presentation allows identification of the most appropriate parameters to use when managing the real inventory system; the best parameters will usually minimize or maximize certain performance measures. As an example, with perishable items, simulations would help us to establish what parameters are best able to provide outcomes that minimize the reduction in value of the inventory while maximizing sales due to substitutability.

This approach has been well-established as an appropriate method to manage inventory. Myers (2009) used a simulation model to analyze an inventory system with perishable and substitution, using the periodic review system. Myers (2009) found two essential components in the optimal order up to levels of the two products. The first element is the 'aggregate order up to' level of all the goods, which is converted to an

aggregate service level. The second element is 'allocation order up to' level for each product. The results were then used to develop a heuristic for establishing an inventory policy for perishability and substitution; heuristic techniques allow a problem to be rapidly solved but may provide a slightly non-optimal outcome. When age-based substitution level increases from low to high, the aggregate service level is increased no matter if the product-based substitution level is low or high.

Beshara, El-Kilany, and Galal (2012) used simulation methods to evaluate and analyze the performance of a replenishment model of agri-food supply chain for highly perishable products. They evaluated the system performance using the average purchasing cost/day, the average cost of waste/day, the average cost of selling in open market/day, the average holding cost, the average transportation cost/day, and the average of total cost/day. They established that the holding cost is the greatest cost, and the performance can be increased by decreasing waste and the amount of perished products at facilities. Including substitution flexibility can provide a significant improvement in costs, particularly by decreasing ordering costs (Seyedhoseini, Rashid, Kamalpour, & Zangeneh, 2015).

SOLUTIONS AND RECOMMENDATIONS

Simulation remains a powerful tool to understand and develop management solutions applicable to complex systems. Simulation models can accommodate many assumptions, especially relating to short-shelf-life products for which experimentation is expensive and risky. From detailed studies of a particular subsystem or wider studies of entire supply chains, simulation enhances our ability to create effective models of systems. Therefore, many scholars have turned to simulations to solve problems in these types of complex systems with stochastic demand and many input factors. They have also recognized that simulation allows sensi-

tive analysis of the effects of factors such as product based substitution, age-based substitution, lead time, shelf-life of products, cost structures, cost parameters, and some products on the simulated optimal solutions.

Myers (2009) used a simulation-optimization approach to analyzing inventory management systems that account for perishability and substitution. This approach allowed identification and evaluation of the impact of key factors in determining optimal policies. A heuristic approach for determining inventory policy for two product variants was developed after this.

Beshara et al. (2012) used simulation results to propose different scenarios to study the effect of order quantity on total cost. Their results were also used to analyze the contribution of each cost parameter (e.g., waste cost and holding cost) on total cost and the relationship between total cost and safety stock level. Another approach is to evaluate non-financial metrics and find a solution that balances the needs of multiple stakeholders (Duong, Wood, & Wang, 2015). Merely optimizing a single, combined financial metric has limitations, that can be addressed by evaluating a range of factors. As an example, many functional managers may need performance that is not effectively represented by using a total cost measure, but which may be captured by examining performance over a range of metrics (Cannella et al., 2013): Order Rate, Fill Rate, Average Inventory, Inventory Variance, and the Work in Progress (WIP) Variance ratio.

Simulation models provide a suitable method for developing new inventory management systems and approaches that will see wide application in current supply chain solutions. This use of simulation is an important use of information technology to solve significant management problems in the contemporary era.

FUTURE RESEARCH DIRECTIONS

Future research can extend models to account for questions that remained unanswered; e.g., optimal

ordering policies for fixed life perishables under continuous review (Nahmias, 2011). Such work would also extend Myers (2009)'s work that focused only on revenue as a performance measure and used a periodic review approach rather than a (more realistic) continuous review approach. The proposed future research would be based on more realistic assumptions (i.e., assumptions closer to those observed in real situations) than have been used in past studies, particularly emphasizing other decision-making criteria other than just revenue while adopting a continuous review approach. Cross-functional considerations and non-financial metric may be more easily assessed using simulation approaches (Duong et al., 2015). Real-time information (such as provided by radio frequency identification (RFID) technology coupled with enterprise resource planning (ERP) systems should be included as this is an area where research lags behind industry practice (Karaesmen, Scheller-Wolf, & Deniz, 2011, p. 429). Data from item-level tagging provides greater ability to manage these situations (Thiesse & Buckel, 2015). Future research can examine real-time information used in the context of issuing, pricing, and disposal strategies to extend current state-of-the-art inventory management systems (Lee, Mu, Shen, & Dessouky, 2014).

Future research will provide significant benefits for managers and society. Dynamic inventory management methods for perishable inventory should enable managers to reduce the incidence of stock-outs while improving other criteria. This research should reduce waste while enhancing business operations. Such outcomes are important globally as key inventory categories that suffer from perishability include pharmaceuticals and foodstuffs, which are of crucial importance in developing countries. The future research will help to reduce waste in these categories and ensure wider availability to the required products by many more people. By continuously improving inventory systems waste will be decreased, and the products will be more easily delivered to more customers. This may require consideration of

different KPIs that are more functionally based, indicating that simulation might enable development of more meaningful inventory management systems (Duong, Wood, & Wang, in press).

CONCLUSION

This chapter focused on solving problems relating to the management of perishable inventory due to the significant benefits globally that can be secured by reducing wastes accruing from deterioration. Effective ordering and management policies for perishability products can reduce the losses. Perishability and substitution inventory management systems with stochastic demands are very complex, making them most amenable to analysis using simulation as the use of analytical tools to model and analyze such complex systems remains impractical. Simulation enables models to be developed that can accommodate significant complexity and stochastic variation while generating a range of data that enables key relationships to be more clearly understood. The simulation data can be used to derive appropriate replenishment and inventory management policies that will enable organizations to reach any of a range of different objectives.

Research shows that discrete event simulation models are preferred for evaluating such complex inventory systems. This chapter showed that simulation can be used as a tool to determine optimal inventory policies accounting for perishability and substitution, particularly accounting for a range of non-financial metrics that functional managers may be required to use. The use of simulation is supported with optimization tools and simulation software and will prove increasingly important in scenarios, given the propensity of consumers to willingly substitute one product for another, increasing the complexity of management decisions. These models enable managers to understand complex interactions between perishability parameters and substitution (either driven by consumers or the managers) and the replenishment policy parameters.

REFERENCES

- Alftan, A., Kaipia, R., Loikkanen, L., & Spens, K. (2015). Centralised grocery supply chain planning: Improved exception management. *International Journal of Physical Distribution & Logistics Management*, 45(3), 237–259. doi:10.1108/IJP-DLM-02-2014-0017
- Amorim, P., Costa, A. M., & Almada-Lobo, B. (2013). Influence of consumer purchasing behaviour on the production planning of perishable food. *OR-Spektrum*, 36(3), 669–692. doi:10.1007/s00291-013-0324-9
- Beshara, S., El-Kilany, K. S., & Galal, N. M. (2012). *Simulation of agri-food supply chains*. World Academy of Science, Engineering and Technology.
- Bijvank, M., & Vis, I. F. A. (2011). Lost-sales inventory theory: A review. *European Journal of Operational Research*, 215(1), 1–13. doi:10.1016/j.ejor.2011.02.004
- Broekmeulen, R. A. C. M., Fransoo, J. C., Van Woensel, T., & van Donselaar, K. H. (2007). Consumer responses to shelf-out-of-stocks of perishable products. *International Journal of Physical Distribution & Logistics Management*, 37(9), 704–718. doi:10.1108/09600030710840822
- Cannella, S., Barbosa-Póvoa, A. P., Framinan, J. M., & Relvas, S. (2013). Metrics for bullwhip effect analysis. *The Journal of the Operational Research Society*, 64(1), 1–16. doi:10.1057/jors.2011.139
- Cannella, S., Framinan, J. M., & Barbosa-Póvoa, A. (2014). An IT-enabled supply chain model: A simulation study. *International Journal of Systems Science*, 45(11), 2327–2341. doi:10.1080/00207721.2013.769072
- Chen, X., Feng, Y., Kebblis, M. F., & Xu, J. (2015). Optimal inventory policy for two substitutable products with customer service objectives. *European Journal of Operational Research*, 246(1), 76–85. doi:10.1016/j.ejor.2015.04.033



- Deniz, B., Scheller-Wolf, A., & Karaesmen, I. Z. (2004). *Managing inventories of perishable goods: The effect of substitution*: GSIA Working Paper# 2004-E55, Carnegie-Mellon University, Pittsburgh, PA.
- Duong, L. N. K., Wood, L. C., & Wang, W. Y. C. (2015). A multi-criteria inventory management system for perishable & substitutable products. *Procedia Manufacturing*, 2, 66–76. doi:10.1016/j.promfg.2015.07.012
- Duong, L. N. K., Wood, L. C., & Wang, W. Y. C. (in press). A review and reflection on inventory management of perishable products in a single-echelon model. *International Journal of Operational Research*.
- Executive office of the President. (2015). *Trans-Pacific Partnership: Summary of U.S. objectives*. Retrieved from <https://ustr.gov/tpp/Summary-of-US-objectives>
- Grocery Manufacturers Association. (2015). *Solving the out-of-stock problem: A FMI/GMA trading partner alliance report*. Retrieved from http://www.gmaonline.org/file-manager/About/15032FMIN_TPA_OutofStock_v51.pdf
- Karaesmen, I. Z., Scheller-Wolf, A., & Deniz, B. (2011). Managing perishable and aging inventories: Review and future research directions. In K. G. Kempf, P. Keskinocak, & R. Uzsoy (Eds.), *Planning production and inventories in the extended enterprise* (Vol. 151, pp. 393–436). New York: Springer; doi:10.1007/978-1-4419-6485-4_15
- Kelton, W. D., & Sadowski, R. P. (2009). *Simulation with ARENA* (5th ed.). New York: McGraw-Hill.
- Law, A. M., & Kelton, W. D. (2000). *Simulation modeling and analysis*. Boston: McGraw-Hill.
- Lee, Y.-M., Mu, S., Shen, Z., & Dessouky, M. (2014). Issuing for perishable inventory management with a minimum inventory volume constraint. *Computers & Industrial Engineering*, 76(1), 280–291. doi:10.1016/j.cie.2014.08.007
- Li, S., Zhang, J., & Tang, W. (2015). Joint dynamic pricing and inventory control policy for a stochastic inventory system with perishable products. *International Journal of Production Research*, 53(10), 2937–2950. doi:10.1080/00207543.2014.961206
- Lucas, T. W., Kelton, W. D., Sánchez, P. J., Sanchez, S. M., & Anderson, B. L. (2015). Changing the paradigm: Simulation, now a method of first resort. *Naval Research Logistics*, 62(4), 293–303. doi:10.1002/nav.21628
- Myers, B. R. (2009). *Analysis of an inventory system with product perishability and substitution: A simulation-optimization approach* (Doctoral dissertation). Available from ABI/INFORM Complete database. (UMI No. 3374737)
- Nahmias, S. (Ed.). (2011). *Perishable inventory systems* (Vol. 160). New York: Springer; doi:10.1007/978-1-4419-7999-5
- Sargent, R. G. (2013). Verification and validation of simulation models. *Journal of Simulation*, 7(1), 12–24. doi:10.1057/jos.2012.20
- Scialabba, N. E.-H. (2014). *Food wastage footprint*. Presented at the meeting of the 29th FAO Regional Conference for Europe, Bucharest, Romania. Retrieved from http://www.fao.org/fileadmin/templates/nr/sustainability_pathways/docs/FWF2_FAO_final_02042014.pdf
- Seyedhoseini, S. M., Rashid, R., Kamalpour, I., & Zangeneh, E. (2015). Application of queuing theory in inventory systems with substitution flexibility. *Journal of Industrial Engineering International*, 11(1), 37–44. doi:10.1007/s40092-015-0099-5
- Thiesse, F., & Buckel, T. (2015). A comparison of RFID-based shelf replenishment policies in retail stores under suboptimal read rates. *International Journal of Production Economics*, 159, 126–136. doi:10.1016/j.ijpe.2014.09.002

van Donselaar, K. H., Van Woensel, T., Broekmeulen, R. A. C. M., & Fransoo, J. C. (2006). Inventory control of perishables in supermarkets. *International Journal of Production Economics*, 104(2), 462–472. doi:10.1016/j.ijpe.2004.10.019

ADDITIONAL READING

Alizadeh, M., Eskandari, H., & Sajadifar, S. M. (2014). A modified (S - 1, S) inventory system for deteriorating items with poisson demand and non-zero lead time. *Applied Mathematical Modelling*, 38(2), 699–711. doi:10.1016/j.apm.2013.07.014

Baron, O., Berman, O., & Perry, D. (2010). Continuous review inventory models for perishable items ordered in batches. *Mathematical Methods of Operations Research*, 72(2), 217–247. doi:10.1007/s00186-010-0318-1

Civelek, I., Karaesmen, I., & Scheller-Wolf, A. (2015). Blood platelet inventory management with protection levels. *European Journal of Operational Research*, 243(3), 826–838. doi:10.1016/j.ejor.2015.01.023

Ignaciuk, P. (2012). Dead-time compensation in continuous-review perishable inventory systems with multiple supply alternatives. *Journal of Process Control*, 22(5), 915–924. doi:10.1016/j.jprocont.2012.03.006

Kouki, C., Jemaï, Z., & Minner, S. (2015). A lost sales (r, Q) inventory control model for perishables with fixed lifetime and lead time. *International Journal of Production Economics*, 168(1), 143–157. doi:10.1016/j.ijpe.2015.06.010

Kouki, C., & Jouini, O. (2015). On the effect of lifetime variability on the performance of inventory systems. *International Journal of Production Economics*, 167(1), 23–34. doi:10.1016/j.ijpe.2015.05.007

Kumar, R. S., & Elango, C. (2010). Markov decision processes for service facility systems with perishable inventory. *International Journal of Computers and Applications*, 9(4), 14–17. doi:10.5120/1375-1852

Lian, Z., Liu, X., & Zhao, N. (2009). A perishable inventory model with Markovian renewal demands. *International Journal of Production Economics*, 121(1), 176–182. doi:10.1016/j.ijpe.2009.04.026

Pahl, J., & Voß, S. (2014). Integrating deterioration and lifetime constraints in production and supply chain planning: A survey. *European Journal of Operational Research*, 238(3), 654–674. doi:10.1016/j.ejor.2014.01.060

Panda, S., Saha, S., & Basu, M. (2013). Optimal pricing and lot-sizing for perishable inventory with price and time dependent ramp-type demand. *International Journal of Systems Science*, 44(1), 127–138. doi:10.1080/00207721.2011.598956

Shin, H., Park, S., Lee, E., & Benton, W. C. (2015). A classification of the literature on the planning of substitutable products. *European Journal of Operational Research*, 246(3), 686–699. doi:10.1016/j.ejor.2015.04.013

Wensing, T. SpringerLink ebooks, B., & Economics. (2011). Periodic review inventory systems (Vol. 651). Berlin: Springer.

Yadavalli, V. S. S., Sundar, D. K., & Udayabaskaran, S. (2015). Two substitutable perishable product disaster inventory systems. *Annals of Operations Research*, 224(1), 1–18. doi:10.1007/s10479-014-1783-3

Zhang, L., Wang, J., & You, J. (2015). Consumer environmental awareness and channel coordination with two substitutable products. *European Journal of Operational Research*, 241(1), 63–73. doi:10.1016/j.ejor.2014.07.043



Zhang, R.-Q., Zhang, L.-K., Zhou, W.-H., Saigal, R., & Wang, H.-W. (2014). The multi-item news-vendor model with cross-selling and the solution when demand is jointly normally distributed. *European Journal of Operational Research*, 236(1), 147–159. doi:10.1016/j.ejor.2014.01.006

KEY TERMS AND DEFINITIONS

Continuous Review System: A system where the stock level of each product is calculated each time a product is moved in or moves out the systems in real-time, triggering an order for more stock when the inventory level falls below a particular re-order point.

Optimization: A procedure to make a system or a process as efficient as possible. Optimization usually requires mathematical techniques to find the best possible solutions to a problem. In inventory management, optimization often relates to calculating how much inventory to hold (to meet customer demand) while holding as little as possible (to minimize the costs of holding inventory).

Periodic Review System: A typical inventory system where the inventory level is reviewed at

a regular time intervals (e.g., once a week), after that the decision is made as to how much to order to bring the inventory level up to a given amount.

Perishable Product: A product with short shelf life or one that easily deteriorates. These items include fresh foods, dairy products, and pharmaceuticals. This period of shelf life complicates the inventory management as they must be processed and move through the supply chain for sale to customers before they perish and lose either part of their value or their entire value.

Product Substitution: A possibility when considering management of multiple products. It is a case that a customer will substitute their preferred product with another with similar characteristics. This is particularly true with ‘commodity’ style products (e.g., milk, where two brands may have very similar characteristics of milk product).

Radio Frequency Identification (RFID): A system using radio frequencies to transfer data from tags on items to readers that are connected to other computer systems. This allows automatic identification and tracking products.

Simulation: A method to imitate the operations in real world system over time. The simulation requires a model that closely represents or replicates key characteristics of the physical system.

E-Business Supply Chains Drivers, Metrics, and ERP Integration



Jean C. Essila

Northern Michigan University, USA

INTRODUCTION

The world of business is witnessing the emergence of electronic supply chains (e-SC or e-supply chain). The Internet has radically changed the nature of supply chains at its core (Sambasivan, Mohamed, & Nandan, 2009). In fact, as more business is performed via the Internet, e-SCs are becoming an integral part of supply chain management (SCM; Gunasekaran, Patel, & Tirtiroglu, 2001; Sambasivan et al., 2009). Businesses are becoming supply chain-sensitive organizations. With business via the Internet requiring different fulfillment approaches, traditional drivers of regular supply chains are no longer adequate for explaining how and to what extent e-SC performance is driven (Sambasivan et al., 2009). E-commerce offers consumers more buying options than traditional business. Buyers can instantly compare prices, product attributes, and delivery parameters. As a result, customers have become increasingly demanding. The task of supply chain professionals is more complicated than ever because e-SCs rely on ERP software. This situation often leads to unsatisfied customers, which can force companies to close their doors.

SCM is responsible for ensuring that customer needs are satisfied in a cost-effective manner. To do this, they formulate strategies, allocate resources, organize activities, and assess performance. Effective and efficient SCM resource allocation is contingent upon a proper understanding and interpretation of its performance drivers (Stock & Boyer, 2009). However, managers often operate on the basis of their own experience and commonly-used methodologies that do not always result

in the desired level of performance. Therefore, a framework to leverage actual drivers of e-SC performance is needed (Caputo, Cucchiella, Fratocchi, Pelagagge, & Scacchia, 2004).

In fact, e-SCs involve partners that are linked by Internet technology in broad networks where customers, retailers, distributors, manufacturers, and suppliers are connected (Fliedner, 2003; Lightfoot, & Harris, 2003; Williams, Esper, & Ozment, 2002). Within and across the networks, actors collect, process, store, and disseminate information on materials, goods, funds, and services. e-SCs are composed of many-to-many connections, while relationships in traditional supply chains are characterized by one-to-one connections. A dramatic revision of current SCM techniques is needed (Caputo et al., 2004). Therefore, understanding e-SCM performance drivers and their integration with ERP becomes a necessity for any SCM professional. Based on the literature survey, little attention has been devoted to SCM performance driver evaluation despite the high volume of ongoing research in the field (Gunasekaran et al., 2001; Sambasivan et al., 2009).

This article discusses the performance drivers of eSCs, metrics for measuring efficiency, and their integration with ERPs. Considering the fact that e-SCs are becoming an integral part of the extended enterprise (Sambasivan et al., 2009), the first section of this article introduces a model of the traditional supply chain for both manufacturing and service systems. In addition, it assesses the logistical and cross-functional performance drivers of supply chains (Chopra & Meindel, 2010; Olver, Lant, Plant, Majeste, & Kursh, 2010). The section concludes with a brief comparison between physical product and information flow.

The second section examines e-SC structures and performance metrics for capturing and gauging system effectiveness and efficiency. It also explains the corresponding measures for the implementation of each metric. It should be noted that e-SC metrics and their corresponding measures are effective ways for managers to ensure that the supply chain is achieving the expected benefits (Riggins & Mitra, 2001).

The role of information technology has shifted from passive enabler to high-performing processes that directly impact the organization's performance. Because e-SC performance requires integration (Smart, 2008), the third section discusses the integration of ERPs into e-SCs to enhance their performance. In this last section, the author discusses the effectiveness and efficiency benefits of ERP for e-SCs (Sambasivan et al., 2009).

SCM is among the most important factors to organizational success (Gunasekaran et al., 2001). There are many benefits of e-SCs that are quantifiable and others that are not (Singh & Byrne, 2005). Effective SCM can enhance competitiveness and increase profitability. Nevertheless, SCM professionals and other actors must understand the factors that undergird driver performance in order to achieve a competitive advantage.

BACKGROUND

According to Stevenson (2012), supply chains are sequences of organizations involved in the production of a good and/or the provision of a service. The author argues that organizations generally consist of facilities, functions, or units, and carry out production or service provision activities. Therefore, their facilities, functions, and activities involved in the production or service provision are integral parts of supply chains. Facilities may include operating units, such as factories, storage facilities such as warehouses, processing centers, distributions centers, and even offices since information is manipulated to trigger, move, and track products and services within and

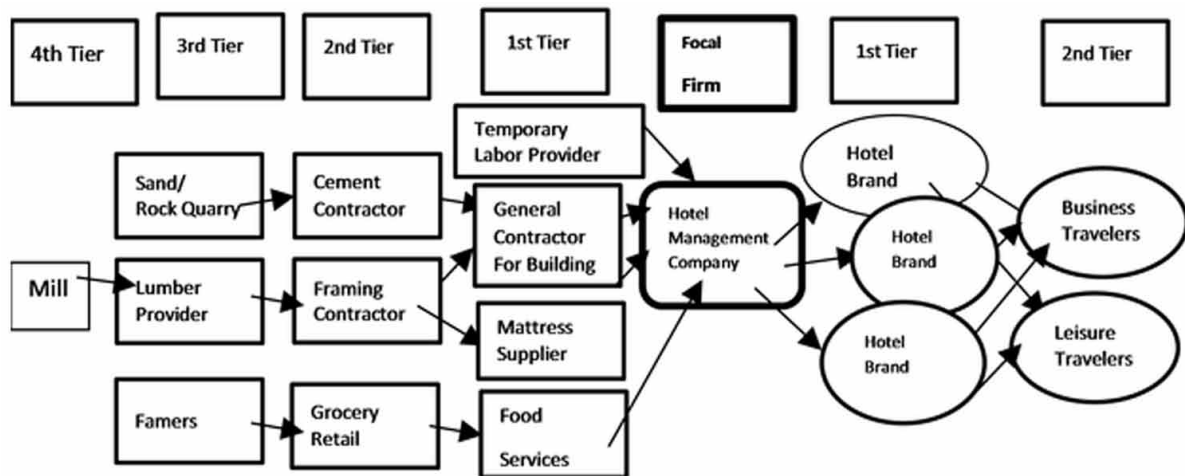
throughout the supply chain. As a result, every product is made through its supply chain within and across multiple sequential organizations. The main conceptual dilemmas have always been: (1) What activities and components of an organization should it include in its given supply chain? (2) Where are the boundaries of SCM territory?

The consensus today seems to be that the supply chain begins with the original suppliers of raw materials. This is followed by the production in an operating unit, the storage in processing centers and warehouses. It ends with the delivery of the finished item to the user. Therefore, supply chains existed since the creation of the first good. However, developing a consensus definition of a supply chain has not been an easy task for academicians and practitioners.

Stock and Boyer (2009) examined 173 unique definitions of SCM from systematic reviews of the entire decision science field in an effort to frame the adoption of a consensus definition of SCM. The definitions were considered unique because they added at least one new element that differed from existing definitions. This inquiry was performed based on the assumption that in the absence of a uniform agreed upon definition, it would be impossible to advance the SCM theory and practice (Stock & Boyer, 2009). The study identified three themes (activities, benefits, and constituents) associated with the definition of SCM. Six sub-themes were also identified. The theme titled "benefits" consisted of three sub-themes: value added, efficiency created, and customer satisfaction. These three sub-themes accounted for 47%, 35%, and 28%, respectively. The theme titled "activities" was credited with physical (materials), services, finances, and information flows counting as its first sub-theme, with networks of internal and external relationships as its second sub-theme. However, there continues to be a variety of theories on what is a supply chain and how SCM should be defined (Mentzer et al., 2001).

With the advent of the Internet and its business applications, the world has witnessed a new

Figure 1. Simple service supply chain (adapted from Fawcett et al., 2007) (Source: Essila, 2015)



type of supply chain, the e-SC. This added more confusion to the conceptualizations on SCM. Confusions upon definitions of SCM existed in both the academic and practitioner circles (New, 1997; Tan, 2001). This article is based on the existing literature in an attempt to frame a comprehensive understanding of the concept of an e-SC, its performance drivers, accompanying metrics, and integration with ERPs. A framework for structuring e-SC drivers helps to achieve a strategic fit between the supply chain strategy and the organization competitive strategy (Chopra & Meindl, 2010).

TRADITIONAL SUPPLY CHAINS

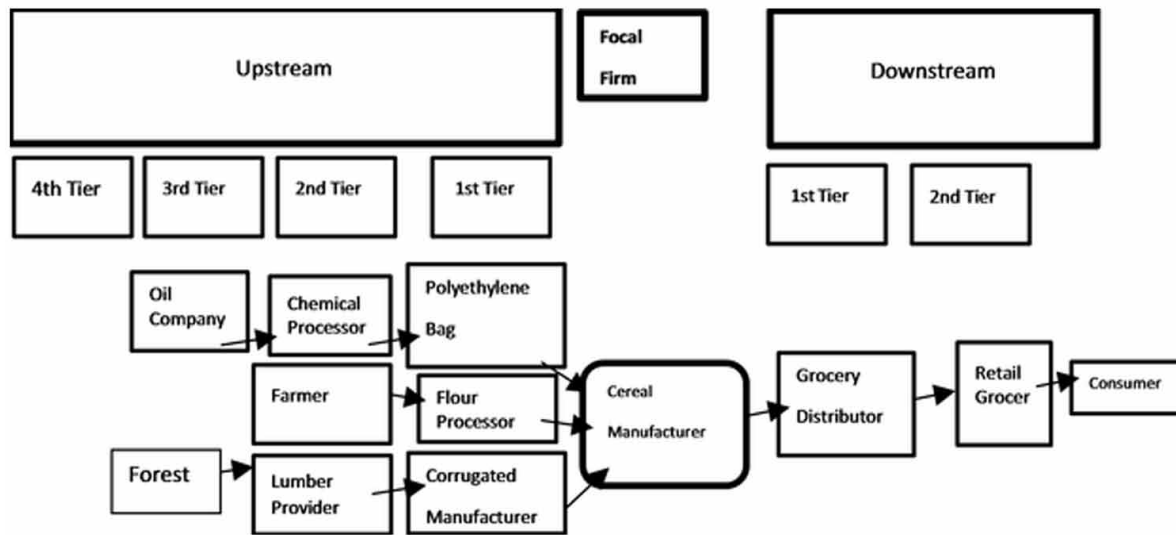
Models and Drivers

- Manufacturing and Service Supply Chain:** As mentioned, the Stock and Boyer (2009) study that compiled 173 unique definitions of SCM proposed a new consensus and encompassing definition of SCM. SCM is the management of networks of relationships within and throughout organizations consisting of suppliers of raw materials, procurement,

production units, logistics, and marketing that facilitate bidirectional flow of goods, services, funds, and information from the original supplier to the end-user with the benefits of adding value and maximizing profit while increasing customer satisfaction (Stock & Boyer, 2009). E-SCs focus on the management of the information flow, managing technology and information processes to optimize the flow of goods, finances, and materials to achieve customer satisfaction and competitive advantage (Srinivasan, 2010);

- Model of Manufacturing and Service Supply Chain:** Models are selected and used as simplified representations of reality. Models facilitate understanding of phenomenon or concepts. This section introduces models of traditional supply chains for both the manufacturing and service sectors. To illustrate supply chains, Fawcett, Ellram, and Ogden (2007) suggested the following simple models of manufacturing and service supply chains:
 - Simple model of a hotel service supply chain;
 - Simple model of a manufacturing supply chain.

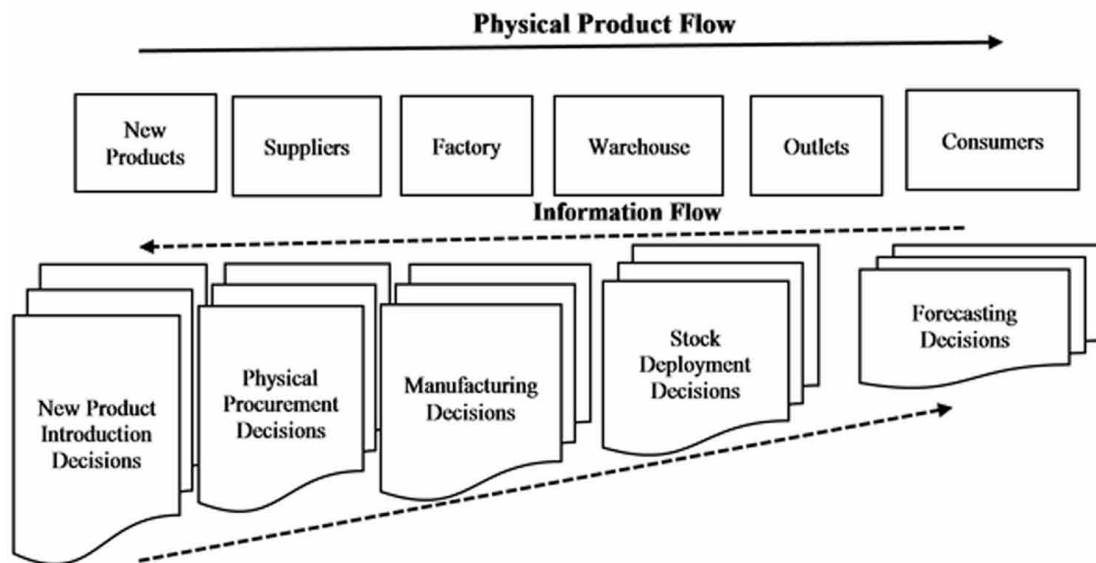
Figure 2. Simple manufacturing supply chain (adapted from Fawcett et al., 2007) (Source: Essila, 2015)



As depicted in Figures 1 and 2, models of supply chains are viewed from the central ring of the chain (also known as the focal firm). The focal firm is the departing point to both the left and the right. Partner firms to the left of the focal firm form the upstream side of the supply chain. Those to the right form the downstream side of the supply chain. Upstream partners are also known as tier “n” suppliers. Most definitions in the literature describe supply chains as the flow of materials, information, services, and funds from the suppliers of suppliers all the way down to the final consumers. The flow is symbolized by arrows as depicted in Figures 1 and 2. Some authors argue that the essence of SCM is to manage the flow of goods and information from the point of departure to the point of consummation, implying a unidirectional flow (Arthur, 1991; Zsidisin, Jun, & Adams, 2000). In fact, information, materials, services, and funds all follow a forward and backward path through the supply chain (Svenson, 2002; Towill, Childerhouse, & Disney, 2001). Svenson explained that the demand information flows upstream from the consumers; materials, funds, and services flow downstream in the supply chain:

- Drivers of Traditional Supply Chains:** According to Olver et al. (2010), supply chain performance is driven by both logistic and cross-functional factors. Olver et al. (2010) argue that analyzing supply chain drivers helps to understand how organizations can improve their supply chain productivity to become more responsive and efficient. Logistics drivers are manufacturing and supporting facilities, inventories, and transportation. Supporting facilities are physical locations needed to provide a service (Fitzsimmons & Fitzsimmons, 2011). Hospital buildings, doctors’ offices, airplanes, or classrooms are examples of supporting facilities. Cross-functional drivers of supply chains are sourcing, pricing, and information. Information affects and is directly affected by all other drivers. It is, therefore, the biggest driver of a supply chain (Olver et al., 2010). Figure 3 depicts both physical and information flow in the supply chain.

Figure 3. Physical and information flow in the supply chain (adapted from Sherer, 2005) (Source: Sherer, 2005)



E-SCs VS. TRADITIONAL SUPPLY CHAINS

Structures, Drivers, and Measures

- e-SC:** e-SCs encompass partners that are connected by Internet technology in broad networks where customers, retailers, distributors, manufacturers, and suppliers are all linked (Fliedner, 2003; Lightfoot & Harris, 2003; Williams et al., 2002). Within and across the networks, actors collect, process, store, and disseminate information on materials, goods, funds, and services. e-SCs are composed of many-to-many connections while relationships in a traditional supply chain are characterized by one-to-one connections. A dramatic revision of current SCM techniques is needed (Caputo et al., 2004). The primary purpose of e-SCs is to help organizations increase effectiveness, improve efficiency, and achieve strategic benefits, such as revenue increase and competitive advantage, by establishing customer loyalty (Riggins & Mitra, 2001; Van Hooft & Stegwee, 2001). Next, the researcher introduces a structure of a typical e-SC;
 - e-SC Structures:** Wheatley (1999) predicted that in the realm of business, the amount of investment on brick and mortar facilities will be replaced by investment in technology. An increase in investment puts more pressure on e-SC managers to achieve high performance in order to yield a higher return on investment. Traditional arm's length, horizontal, and vertical integrated structures are no longer appropriate for e-SC since they connect organizations. Unlike traditional supply chains, "the electronic supply chain is round in form" (Williams et al., 2001, p. 708). The traditional supply chain goal is to increase responsiveness and efficiency through long-term stability. Traditional supply chains usually build relationships that can achieve a long-term benefit among partners. On the other hand, e-SCs are designed to adapt to the dynamic environment of an e-business. Therefore, they are flexible in nature. Supply chains shape the organizational structure for future performance enhance-

ments. As depicted in Figure 4, firms are connected through information technology. The focal firm is represented by a central ring. Other partners in the e-SC are represented by outer rings;

- **Drivers and Metrics of Traditional Supply Chains:** Unlike e-SCs, traditional supply chain design initiatives are generally viewed as top to bottom processes similar to the classical waterfall-type model in systems engineering. The examination of the six drivers, along with their corresponding metrics and accompanying measures, reveals the need for more stable relationships with partners. The advent of globalization and the Internet has dramatically changed the paradigms. Supply chain leaders need to adjust their strategies and structure the six drivers of traditional supply chains depicted (see Table 2) in order to achieve a high level of responsiveness while minimizing logistics costs to the entire supply chain;
- **Drivers and Metrics of e-SC:** As mentioned, drivers of traditional supply chains are facilities, transportation, inventory, information, sourcing, and pricing. They are classified as either logistical or cross-functional performance drivers (Olver et al., 2010). Unlike traditional supply chains, factors that affect e-SC performance are the

adapted organizational structure, which effectively manages relationships among network partners, the *managerial criteria*, and the e-SC *critical activities* (Caputo et al., 2004). Critical activities can, therefore, be viewed as logistical drivers. The organizational structure and managerial criteria can be viewed as cross-functional drivers of the e-SC factor equation.

Due to their unique characteristics, e-SCs need new metrics to adapt to challenges using innovative solutions that take into account the new dynamic environment (Barnes & Hinton, 2007). Little attention has been given to supply chain metrics and their accompanying measures (Sambasivan et al., 2009). Effective SCM requires an optimal balance between responsiveness and efficiency. Because each driver interacts with the others, supply chain managers must effectively combine the logistical and cross-functional drivers to avoid conflicting goals (Olver et al., 2010). From a practical perspective, the working technique is to increase positive impacts while minimizing the negative effects on the entire logistics system. A strategic alignment between the supply chain strategy and the organization competitive strategy is also important. Since competitive strategies support corporate strategies, the necessary strategic alignment becomes a central nerve for e-SC performance.

Figure 4. The e-SC structure (adapted from Williams et al., 2002) (Source: Williams et al., 2002)

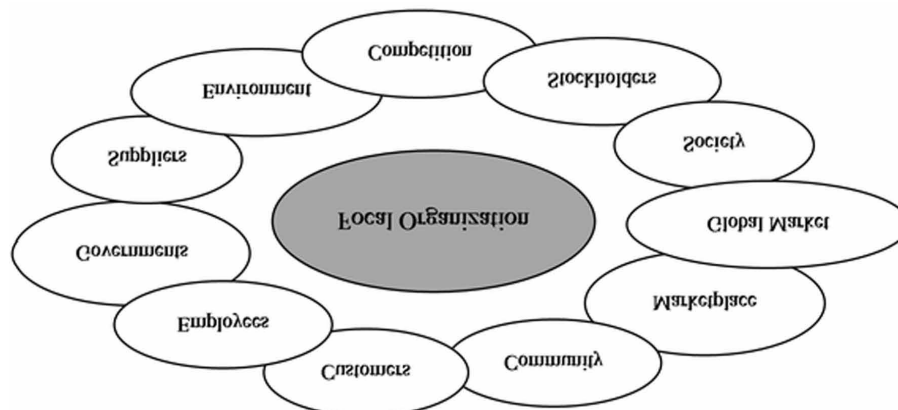


Table 1. *e-SC metrics and their accompanying measures (developed based on Caputo et al., 2004; Sambasivan et al., 2009)*

Driver Type	Drivers	Metrics	Measures
• Logistical	• Adapted organizational culture	• Web-enabled service	• Data access time • System response time • Data transmission speed between business-to-business applications • Number of system verification steps • Traffic volume per page and/or site
• Cross-functional	• Managerial criteria	• Transaction reliability	• Number of transaction errors • Number of backlog transactions • Number of partners accessing the system • Cost per transaction • ERP interface cycle time
• Cross-functional	• Managerial criteria	• Cost	• Administrative cost • Administrative cost
• Cross-functional	• Critical activities	• Time	• Administrative time • Number of stages in the purchasing cycle • Purchasing lead time • Procurement cycle time • Procurement response time
• Logistical	• Adapted organizational culture	• E-Response	• Reliability of mail service • Completed number of transactions per period (day, week, or month) • Pending number of transactions per period (day, week, or month)
• Cross-functional	• Adapted organizational culture	• E-Invoice process	• Number of steps involved • Dispute resolution completion time
• Logistical	• Adapted organizational culture	• E-Payment	• Payment time • Reconciliation time
• Cross-functional	• Managerial criteria	• E-Document	• Data accuracy • Data reliability

In SCM, metrics track and gauge performance of each driver, as well as the impact resulting from its interaction with the others. Sambasivan et al. (2009) suggested six practical metrics for e-SCs: Web-enabled service, data reliability, time and cost, e-response, invoice presentation and payment, and e-document management. e-SC metrics capture performance both at the aggregated and organizational level. The aggregated level measures the performance of the entire supply chain, encompassing all partners to the supply chain. The organizational level focuses on individual organizational performance, including the process and the performer levels. The accompanying measures for each performance metric are summarized in Table 1. It provides a summary of the three e-SC drivers, in addition to their metrics and measures.

Information is the key driver of both traditional and e-SCs because it serves as the connector to all drivers (Olver et al., 2010). Information affects and is affected by all supply chain drivers. It triggers logistical and cross-functional processes associated with supply chain performance factors. For example, information on customer demand is the basis for the aggregate plan that establishes operation capacity. The operation capacity and capability dictate the rules for establishing schedules for specific products or groups of similar items. At this planning stage, basic strategies are formulated to meet demand using a chase demand or level capacity strategy.

At every stage in the supply chain process, bi-direction flows of information play a critical role. As a leading performance driver, informa-

Table 2. Drivers, metrics, and measures of traditional supply chains (developed based on Olver et al., 2010)

Driver Type	Drivers	Metrics	Measures
• Logistical	• Facilities	• Design capacity • Effective capacity • Capacity cushion • Utilization • Efficiency • Product variability	• Production per unit • Production cycle time • Flow time • Flow time efficiency • Production service level • Average batch volume
• Logistical	• Inventory	• Inventory cycle • Safety stock • Product availability • Obsolescence	• Average inventory • Inventory turns • Average replenishment lot size • Rate of obsolete inventory
• Logistical	• Transportation	• Network design • Shipment • Transportation mode selection process	• Inbound cost per period • Outbound cost per period • Inbound average shipment size • Outbound average shipment size • Volume per mode
• Cross-functional	• Information	• Demand forecasting accuracy • Process design type (push vs. pull) • Sharing and coordination • Availability • Accessibility • Enabling technology (ERP, EDI, RFID, and SCM) • Encryption	• Mean absolute deviation (MAD) • Mean squared error (MSE) • Forecasting horizon • Frequency of update • Information velocity • Transactions per users per enabling technology • Reliability of enabling technologies
• Cross-functional	• Sourcing	• In-house/outsource rate • Supplier selection • Purchasing/Procurement • Quality of supply	• Average purchasing price • Purchasing range • On-time deliveries to total deliveries • Receipts defect rate (RDD)
• Cross-functional	• Pricing	• Price menu • Profit	• Range of sale price • Average sale price • Incremental fixed/variable cost per unit

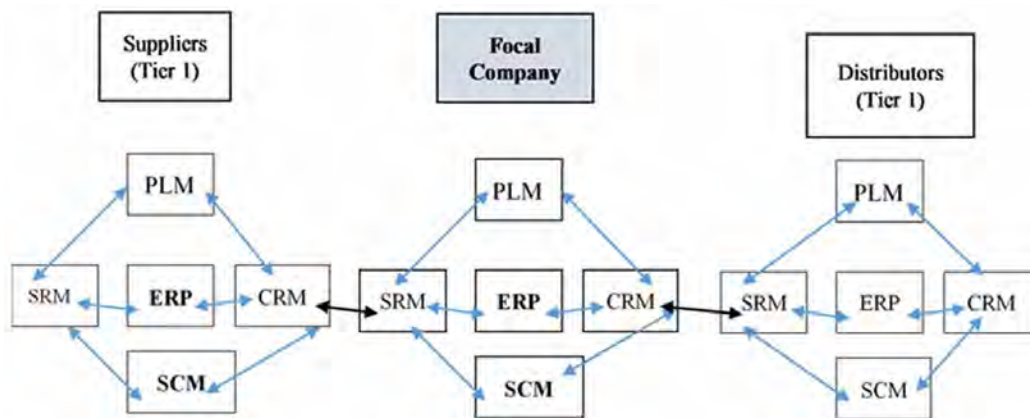
tion sets the stage for potential productivity gains for both logistical and cross-functional processes within, across, and throughout the entire supply chain. Information is a cross-functional driver of a supply chain (Oliver et al., 2010). As depicted in Table 2, information is a major supply chain driver of traditional and e-SCs. All supply chains often rely on enabling technology such as ERP, electronic data interchange (EDI), radio frequency identification (RFID), and SCM software. For e-SCs, enabling technology is a necessity. Such technology enables organizations to collect, store, process, and disseminate information within and across the entire supply chain. Next, the article will explain the necessary integration between e-SCs and ERP.

E-SC INTEGRATION WITH ERP

ERP Integration

ERPs, the third generation of enterprise systems (ESs), began with material requirements planning (MRP) in 1964. This was followed by manufacturing resource planning (MRP II) in 1983. ERPs are currently the most complex version of ESs to integrate both the functional and cross-functional process in the organization. ERPs have MRP as their core system. ERPs support all business function operations and transactions, linking them in a uniform platform to make information available in real-time across

Figure 5. Enterprise application suite in the supply chain (adapted from Magal & Word, 2011) (Source: Magal & Word, 2011)



the organization's entire supply chain. ERPs track the information within and across organizations. The Internet provides a broad visibility of the information. Supply chain managers utilize information provided by ERPs as the basis for making informed decisions.

A close look at an ES application suite reveals the necessity to integrate ERPs with SCM software. As depicted in Figure 5, the five basic components of a simple ES are the product life cycle management (PLM), the supplier relationship management (SRM), the ERP, the SCM, and the customer relationship management (CRM). ERPs are the epicenter of ESs because they integrate and connect the other ES components.

ERP systems are configurable ESs designed to integrate processes and information within and across organizations (Kumar & Van Hillegersberg, 2000). ERPs allow organizations to automate repetitive tasks, underlining transactions and processes. If properly used, SCM can be improved by integrating ERPs. Organizations that are best placed to succeed are those that have implemented an adequate business infrastructure utilizing ERP capabilities (Srinivasan, 2010). ERPs help to increase velocity in the supply chain fulfillment process.

FUTURE RESEARCH DIRECTIONS

This article identified variables relevant to the measurement of an effective e-SC. A similar study with more focus on the relationship among e-SC performance drivers that would allow a moderate (or even excessive) interference with a purpose of establishing a more elaborate causal relationship would provide more insight into the underlying construct. A longitudinal field experiment on the subject would also derive more detailed information about performance drivers of e-SCs.

Research could also be directed toward testing the hypothesis about the quality of enabled-technology used in an e-SC, as well as its effect on both efficiency and responsiveness. Suggested questions and ideas include:

1. Does full use of technology in the e-SC lead to greater effectiveness and a higher return to stakeholders?
2. What factors influence how supply chain professionals feel about the full use of technological supply chain capability?
3. If an entirely automated e-SC system was possible, would it be more effective?
4. Research could look more closely at traditional supply chains and technologically-

assisted e-SCs to elaborate a classification of technologically-assisted e-SC systems with a measure of the extent or magnitude of technology involved.

CONCLUSION

With the advent of globalization and the emergence of e-business, e-SCs are imperative. Organizations now compete using marketing, operations, and supply chain functions. Companies are no longer competing as stand-alone entities in today's business dynamic environment (Lambert & Cooper, 2000).

This article reviewed the six performance drivers of traditional supply chains based on the Olver et al. (2010) framework: facilities, inventory, transportation, information, sourcing, and pricing. The first three are referred to as logistics drivers because they are directly responsible for moving materials, funds, information, and service throughout supply chain pipelines. Information, sourcing, and pricing are known as cross-functional drivers of supply chains because they involve many business functions and processes. Traditional supply chain drivers aim to build long-term relationships among partners. They involve one-to-one relationship types. Early supply chains were arm's length and horizontally and vertically integrated in nature.

The emergence of the Internet and e-business calls for new types of supply chains—the e-SC. Unlike traditional supply chains, e-SCs are round in nature (Williams et al., 2001), and involve a many-to-many relationship among its partners. An e-SC initiative is designed to adapt to a dynamic environment. Therefore, they strive for flexibility and adaptability. e-SCs also require more enabled-technology than their traditional supply chain counterparts. As a result, performance drivers of traditional supply chains are no longer appropriate when explaining e-SC performance. Also, business via the Internet requires different fulfillment approaches, making traditional drivers of regular

supply chains no longer adequate for explaining how and to what extent e-SC performance is driven (Sambasivan et al., 2009).

A dramatic revision of current SCM techniques is needed (Caputo et al., 2004). Based on the framework suggested by Caputo et al. (2004), this article introduced three e-SC performance drivers: the adapted organizational structure for effectively managing many-to-many relationship types among network partners, the managerial criteria, and the e-SC critical activities. Enabled-technology employed by e-SCs are specific ESs such as ERP. ERPs are the third generation of ESs that began with MRP and MRP II. Effective integration of ERPs into e-SCs can significantly improve responsiveness and yield a high return on investment. Understanding drivers of e-SCs helps managers to better structure drivers to achieve their desired level of service at a minimal cost while increasing customer satisfaction and competitive advantage.

REFERENCES

- Arthur, L. D. (1991). *Logistics in service industry*. Oak Brook, IL: The Pennsylvania State University for the Council of Logistics Management.
- Barnes, D., & Hinton, M. (2007). Searching for e-business performance measurement systems. *The Electronic Journal Information Systems Evaluation*, 10(1), 1–8.
- Caputo, A. C., Cucchiella, F., Fratocchi, L., Pelagagge, P. M., & Scacchia, F. (2004). Analysis and evaluation of e-supply chain performances. *Industrial Management & Data Systems*, 104(7), 546–557.
- Chopra, S., & Meindel, P. (2010). *Supply chain management: Strategy, planning, and operations* (4th ed.). Upper Saddle River, NJ: Prentice Hall.
- Essila, J. C. (2015). *Effectiveness of logistics management* (Unpublished Doctoral dissertation). California InterContinental University, Irvine, CA.



- Fawcett, S., Ellram, L. M., & Ogden, J. A. (2007). *Supply chain management from vision to implementation*. Upper Saddle River, NJ: Prentice Hall.
- Fitzsimmons, J. A., & Fitzsimmons, M. J. (2011). *Service management: Operations, strategy, and information technology* (7th ed.). New York, NY: McGraw-Hill Irwin.
- Flidner, G. (2003). CPFR: An emerging supply chain tool. *Industrial Management & Data Systems*, 103(1), 14–21. doi:10.1108/02635570310456850
- Gunasekaran, A., Patel, C., & Tirtiroglu, E. (2001). Performance measures and metrics in a supply chain environment. *International Journal of Operations & Production Management*, 21(1/2), 71–87. doi:10.1108/01443570110358468
- Kumar, K., & Van Hillegersberg, J. (2000). ERP experiences and evolution. *Communications of the ACM*, 34(4), 22–26. doi:10.1145/332051.332063
- Lambert, D., & Cooper, M. (2000). Issues in supply chain management. *Industrial Marketing Management*, 29(1), 65–83. doi:10.1016/S0019-8501(99)00113-3
- Lightfoot, W., & Harris, J. R. (2003). The effect of the Internet in industrial channels: An industry example. *Industrial Management & Data Systems*, 103(2), 78–84. doi:10.1108/02635570310463401
- Magal, S. R., & Word, J. (2011). *Integrated business processes with ERP systems*. Hoboken, NJ: John Wiley & Sons, Inc.
- Mentzer, J. T., Dewitt, W., Kleeber, J. S., Min, S., Nix, N. W., Smith, C. D., & Zacharia, Z. G. (2001). Defining supply chain management. *Journal of Business Logistics*, 22(2), 1–25. doi:10.1002/j.2158-1592.2001.tb00001.x
- New, S. (1997). The scope of supply chain management research. *Supply Chain Management: An International Journal*, 2(1), 15–22. doi:10.1108/13598549710156321
- Olver, J. M., Lant, T. K., Plant, R., Majeste, K. D., & Kursh, S. R. (2010). *Supply chain management*. Upper Saddle River, NJ: Pearson Learning Solutions.
- Riggins, F. J., & Mitra, S. (2001). *A framework for developing e-business metrics through functionality interaction*. Retrieved from <http://www.tarrani.net/kate/docs/FrmwkDevelopingEBizMetrics.pdf>
- Sambasivan, M., Mohamed, Z. A., & Nandan, T. (2009). Performance measures and metrics for e-supply chains. *Journal of Enterprise Information Management*, 22(3), 346–360. doi:10.1108/17410390910949751
- Sherer, S. (2005). From supply-chain management to value network advocacy: Implications for e-supply chains. *Supply Chain Management*, 10(2), 77–83. doi:10.1108/13598540510589151
- Singh, M., & Byrne, J. (2005). Performance evaluation of e-business in Australia. *The Electronic Journal Information Systems Evaluation*, 8(1), 71–80.
- Smart, A. (2008). E-Business supply chain integration. *International Journal of Enterprise Information Management*, 21(3), 227–246. doi:10.1108/17410390810866619
- Srinivasan, M. (2010, October). E-business and ERP: A conceptual framework toward the business transformation to an integrated e-supply chain. *International Journal of Enterprise Information Systems*, 6(4), 1–19. doi:10.4018/jeis.2010100101
- Stevenson, W. J. (2012). *Operations management* (11th ed.). New York, NY: McGraw-Hill/Irwin.
- Stock, J. R., & Boyer, S. L. (2009). Developing a consensus definition of supply chain management: A qualitative study. *International Journal of Physical Distribution & Logistics Management*, 39(8), 690–711. doi:10.1108/09600030910996323
- Svenson, G. (2002). The theoretical foundation of supply chain management: A functionalist theory of marketing. *International Journal of Physical Distribution & Logistics Management*, 32(9), 734–754. doi:10.1108/09600030210452422
- Tan, K. (2001). A framework of supply chain management literature. *European Journal of Purchasing & Supply Chain Management*, 7(1), 39–48. doi:10.1016/S0969-7012(00)00020-4

Towill, D., Childerhouse, P., & Disney, S. (2000). Speeding up the progress curve towards effective supply chain management. *Supply Chain Management: An International Journal*, 5(3), 122–130. doi:10.1108/13598540010338866

Van Hooft, F. P., & Stegwee, R. A. (2001). E-business strategy: How to benefit from a hype. *Logistics Information Management*, 14(1/2), 44–53. doi:10.1108/09576050110360223

Wheatley, L. R. (1999). Electronic data interchange (EDI): A study of usage and adoption within marketing and logistics channels. *Transportation Journal*, 34(2), 37–45.

Williams, L. R., Esper, T. E., & Ozment, J. (2002). The electronic supply chain. Its impact on the current and future structure of strategic alliances, partnerships, and logistics leadership. *International Journal of Physical Distribution & Logistics Management*, 32(8), 703–719. doi:10.1108/09600030210444935

Zsidisin, G., Jun, M., & Adams, L. (2000). The relationship between information technology and service quality in the dual direction supply chain. *International Journal of Service Industry Management*, 11(4), 312–328. doi:10.1108/09564230010355359

ADDITIONAL READING

Gibson, B., Mentzer, J., & Cook, R. (2005). Supply chain management: Pursuit of consensus definition. *Journal of Business Logistics*, 26(2), 17–25. doi:10.1002/j.2158-1592.2005.tb00203.x

Gosain, et al.. (2005). Coordination for flexibility in e-business supply chains. *Journal of Management Information Systems*, 21(3), 7–45.

Kumar, K. (2001). Technology for supporting supply chain management. *Introduction Communications of the ACM*, 44(6), 58–6. doi:10.1145/376134.376165

Valverde, R., & Saadé, R. G. (2015). The effect of E-supply chain management systems in the North American electronic manufacturing services industry. *Journal of Theoretical and Applied Electronic Commerce Research*, 10(1), 79–98. doi:10.4067/S0718-18762015000100007

KET TERMS AND DEFINITIONS

Logistics: That part of SCM responsible for moving materials, funds, information, and services from the point of origin to the point of consumption.

Logistics Capabilities: Ability to employ logistics capacities to achieve the desired level of responsiveness at the lowest cost possible.

Logistics Capacities: This includes physical assets, such as buildings, plants, factories, manufacturing centers, processing centers, distribution centers, and warehouses, utilities, human resources, computers, cars, trucks, trains, aircrafts, ships, materials and goods, and supporting information.

Logistics Protocol: Sequence in which logistics activities are to be performed according to specific mathematical models.

Strategy: Set of actions designed and employed to achieve goals.

Supply Chain Facilities: Operating units (such as factories), storage facilities (such as warehouses), processing centers, distributions centers, and even offices since information is manipulated to trigger, move, and track products and services within and throughout the supply chain.

Supporting Facility: Supply chain facility necessary before a service can be provided (i.e., hospitals building, classrooms, airplanes, etc.)

Ecological Performance as a New Metric to Measure Green Supply Chain Practices



June Poh Kim Tam

Universiti Sains Malaysia, Malaysia

Yudi Fernando

Universiti Malaysia Pahang, Malaysia

INTRODUCTION

Sustainable development has become a focus nowadays by governments, companies, communities, and even individuals. Manufacturing industry is the first to receive the brunt of criticism as manufacturing activities are assumed to be the prime culprit in creating environmental issues (Gunasekaran & Spalanzani, 2012). Massive consumption of natural resources and energy speed up the resources depletion rate which could scarify the needs of future generation for today's economic developments. Waste generation from manufacturing activities could create negative impacts like various pollutions to environment which subsequently hinder the operation of eco-systems (Cordoba & Veshagh, 2013).

Manufacturing companies are moving towards green oriented approaches in mitigating environmental issues. Green supply chain management is one of the approaches to address the relationships between supply chain management and natural resources management (Simao et al., 2016). According to Srivastava (2005), he defined green supply chain management as integration of environment thinking into supply chain management. Green supply chain initiatives included product design, materials sourcing and selection, manufacturing processes, logistics and distributions and product end of life management. Hawken et al. (1999) predicted a new industrial revolution that will be based on efficient use of resources. Today, the prediction has come true as eco-efficiency

based strategy is now widely adopted by business organizations, especially manufacturing industry.

Manufacturing companies need a comprehensive evaluation system to measure the outcomes of the green supply chain activities that implemented. Anbumozhi and Kanda (2005) found that most of the developing countries remained to be the traditional command-and-control or end of pipe solutions which is very passive approach to reduce negative environmental impacts after they are created. This is because business benefit is the first priority to be considered in adoption of green supply chain practices. Balance between environment and economic is important to a company survival.

BACKGROUND

World Business Council for Sustainable Development (2000) defined eco-efficiency as delivery of competitively priced goods and services that satisfy human needs and bring quality of life while progressively reducing ecological impacts and resource intensity throughout the life cycle to a level at least in line with Earth's estimated carrying capacity. In short, eco-efficiency means creating more value with less impact. It is a concept that combines both economic and environment management. Managers will be motivated to engage in environmental management if the green supply chain initiatives can achieve both environment and economics performance at the same time.

DOI: 10.4018/978-1-5225-2255-3.ch465

Current environment performance focuses more on environment impact intensity and resource use intensity by the firms. Although there is positive linkage between environment performance and economics performance, however it varies between firms. Ecological performance could bring more solid outcome measurement of green supply chain initiatives at market level (Boons & Wagner, 2009). According to Hart (1995), a company can build competitive advantages via natural environment management. Besides being cost effective, integration of natural system into business core value can help a company to have continuous improvement in aspects like quality and flexibility which can preempt competitors and appears as market leader in the manufacturing industry.

Therefore, the scope of performance measurement should be extended to wider indicators other than monetary factors. Besides cost reduction, Eltayeb, Zailani, & Ramayah (2011) suggested operational performance can be investigated from the perspective of quality, flexibility and delivery. In business context, organization has to be responsive to customer requirement in cost reduction, flexibility and quality product. Nowadays, supply chain responsiveness is not limited to business needs but environment needs as well. Supply chain responsiveness has to be designed

with green concept incorporated which efficient to both economic and environment. The benefits brought by green supply chain initiatives are not constraint to the focal company manufacturing activities. Instead, it can be extended into wider scope so that more and more parties can enjoy the benefits.

An exploration in ecological performance should be carried out so that a more holistic performance measurement is developed to assist managers in decision making of adoption of green supply chain initiatives. This study is aimed to discuss the concept of ecological performance which extended from environmental performance.

ENVIRONMENT PERFORMANCE

Commonly, manufacturing companies measure environment performance in term of environment impact intensity that caused by human activities and resource use intensity (Bran et al., 2011). Environmental impact intensity covers emissions to water intensities, emissions to air intensities and GHG emissions intensities. While resource use intensity covers water intensity, energy intensity and materials intensity in a process. Table 1 shows the indicators used to measure environment performance in monetary form by various sectors.

Table 1. Indicators of environment performance in monetary measurement (Department for Environment, Food and Rural Affairs of UK, 2006)

Indicators	Resource-Use Intensity	Environmental Impact Intensity
Economy-wide	Water intensity (m ³ /GDP) Energy intensity (J/GDP) Land use intensity (km ² /GDP) Materials intensity (DMI/GDP)	Emission to water intensities (t/GDP) Emission to air intensities (t/GDP) GHG emissions intensities (t/GDP)
Manufacturing	Water intensity (m ³ /GDP) Energy intensity (J/GDP) Materials intensity (DMI/GDP)	CO ₂ intensity (t/GDP) BOD intensity (t/GDP) Solid waste intensity (t/GDP)
Industry	Water intensity (m ³ /GDP) Energy intensity (J/GDP) Materials intensity (DMI/GDP)	CO ₂ intensity (t/GDP) Solid waste intensity (t/GDP)
Agriculture	Water intensity (m ³ /GDP) Energy intensity (J/GDP) Land use intensity (km ² /GDP) Materials intensity (DMI/GDP)	CO ₂ intensity (t/GDP) CH ₄ intensity (t/GDP)

Table 2. Content analysis of environment performance related articles

Author	Title	Environment Performance Variables	Findings
Chien (2014)	Influences of green supply chain management practices on organizational sustainable performance	- Reduction of emissions - Reduction of hazardous materials consumption - Efficient energy and resource use - Decrease of accident frequency	Environment performance contributed to economics performance via cost reduction in material and energy use, waste and emissions management, environment management
Wang et al. (2015)	How information technology influences environmental performance: Empirical evidence from China	- Complying with environmental regulations - Preventing and mitigating environmental crises - Training employees on environmental issues	Environmental performance contributed to economics performance via efficiency of resource and energy
Kudroli (2014)	Green supply chain management and environmental sustainability – a comparative study on global and Indian perspective	- Reduction of power and paper consumption - Reduction of hazardous materials consumption - Reduction of waste and investment recovery	Environment performance contributed to economics performance via cost reduction in hazardous material and energy use, waste management, inventory management
Jackson et al. (2016)	Examining the impact of design for environment and the mediating effect of quality management innovation on firm performance	- Reduction of air emissions - Reduction of waste water - Reduction of solid waste - Reduction of hazardous materials consumption - Decrease of accident frequency	Environment performance contributed to economics performance via cost reduction in material and energy use, waste and emissions management, environment management
Laari et al. (2016)	Firm performance and customer driven green supply chain management	- Carbon dioxide emissions - Waste minimization - Reduction of energy consumption - Reduction of water consumption - Reduction of hazardous materials consumption	Environment performance contributed to economics performance via cost reduction in material and energy use, waste and emissions management,
Azevedo et al. (2012)	The influence of eco-innovation supply chain practices on business eco-efficiency	- Energy consumption - Business waste	Environmental performance contributed to economics performance via efficiency of resource and energy and waste minimization
Rosen & Kishawy (2012)	Sustainable manufacturing and design: Concepts, Practices and Needs	- VOC emissions	Environment performance contributed to economics performance via cost reduction in VOC emissions management

There are a large number of research about the relationship between environment performance and economics performance. Table 2 summarized some articles by authors from different countries to investigate if there is similarity on the findings. The outcomes may be influenced by other factors like culture, rules and regulations of the individual country.

Based on the content analysis of above articles, environment performance has positive influence towards company operational and financial performance. A good environment performance will benefit company by costs avoidance such

as reduction of materials and energy cost, waste management cost and treatment cost of emissions (Boons & Wagner, 2009). Although it is a positive linkage between environment performance and economics performance, environment performance is not holistic to measure the outcomes of green supply chain initiatives. The effectiveness is constrained to firm profit maximization and pollutions prevention and constrained to the individual firm context. Furthermore, there is considerably variation between companies. Zhu et al. (2007) claimed that different manufacturing industry types display different levels of GSCM outcomes.

Achievement of economic performance through cost reduction in waste and emissions treatment is viewed as end-of-pipe approach which may result in extra environmental investments. This would increase marginal cost as production cost is increased now (Boons & Wagner, 2009). Prevention and restoration efforts are costly and they go against the basic principle of economics. Managers can choose to control pollutions or maximize profits however the former can be only achieved at the expense of the latter. This implies that investors is hardly to invest in a company that is likely to be profitable and environmental oriented at the same time (Boons & Wagner, 2009).

Such notion should change as a company in fact can be profitable and at the same time protects its environment. It is depending on which system level and what kind of research objectives both the economic and environment performance are assessed at. Economic performance is the most important motivator why managers seek to implement green supply chain initiatives (Zhu et al., 2013). Besides costs, it has to be something else that can promote the company in order to outperform its competitors. Therefore, manufacturing companies need a performance measurement with more dimensions that can satisfy both economic and environment needs.

Nevertheless, the environment performance indicators still relevant at firm level as the cost reduction in environment performance still can motivate managers to continue look for opportunities within firm in justifying the business case that related to environmental management.

ECOLOGICAL PERFORMANCE

Ecologists concerns the natural resources that are available throughout an environment. The main focus of ecologist is to study the way how different ecosystems run, and the impact that inhabitants have on the amount of natural resources available. Business activities especially manufacturing industry can pose significant threat to environment

like CO₂ emissions, solid waste and ecosystems like excessive consumption in resource, water and land (Thoo et al., 2015).

Ecological performance is an extended version of environment performance with intention to cover more dimensions in measuring the outcomes of green supply chain initiatives. The concept of reduction of resource consumption and waste generation is not limited to focal company manufacturing activities but its strengths can be extended into other areas which benefits more people, either directly or indirectly. Ecological performance aims to measure green supply chain initiatives outcomes at market level and consumption system throughout the product life cycle.

Eco-efficiency is designed in all stages of product life cycle from materials extraction through production, consumption, recycling, and eventually waste disposal (Rosen and Kishawy, 2012). The reduction of ecological impacts should occur at every stage of product life cycle and the resource efficiency and waste minimization should perform up during the product usage even at the end life of the product. The continuity of eco-efficiency from manufacturer have to transfer to user and the reduction of ecological impacts should widely dispersed rather than individual firm.

A market driven system encompasses the firms that competing for the same customer base (Johnstone, 2007). Besides the two performance indicators discussed in above, manufacturing companies can build up more competitiveness through green supply chain initiatives. Continuous improvement in aspects of quality and flexibility would increase company's capabilities in responding to market dynamic changes. High responsiveness could help company to preempt competitors thus appears to be order winner in market.

Quality

The scope of quality is wide. It covers not only the product or service quality but also supplier quality, production quality, workforce quality as well as environment quality (Baumann & Genou-

laz, 2014). In traditional supply chain, evaluation of quality is mainly focus on product or service quality. Customers seek quality as the ability of the product or service to meet their expectations (Vachon & Klassen, 2008). With the increasing awareness of environment management, customers are now demanding more than the product quality. They are interested in knowing the materials used to build the product, would the production activities cause any negative impacts to environment, and what impacts that will product cause during its usage and what kind of the disposal needed when product is end of life. Life quality is equally important here.

Quality is not limited to focal company activities in supply chain. It is can be extended to suppliers and distributors as they are important players in quality management as a whole. This is crucial especially in the efforts of environment quality management. A company is not able to achieve sustainability without collaboration with its partners in the supply chain. Therefore, the scope quality management should be revised to include stakeholders that involved in the supply chain. The stakeholders could be suppliers, distributors, forwarders as well as employees.

A quality product is produced from a quality production. A production which is stable and within control always gives high quality on manufactured goods or service provided. The product meets the specifications and gives satisfaction to end users. Quality product can easily gain customer loyalty for long term business opportunity. Quality is a field that is less common in term of performance however it is worthy to extend the study in this area as it is a competency to company to compete with competitors (Vachon and Klassen, 2008).

Flexibility

Flexibility refers to how flexible a company can work in different ways, at different times, or in different places when it is necessary (Cambridge

Business Dictionary, 2016). Flexibility covers supplier flexibility, supply flexibility, production flexibility and delivery flexibility (Baumann & Genoulaz, 2014). Flexibility is important in a responsive supply chain. A flexible company has ability to react to customer change and deliver to customer in the soonest time.

Since always, researchers claimed that efficiency supply chain strategy and responsive supply chain strategies are used differently based on the product characteristic, functional and innovative (Fisher, 1997; Christopher & Towill, 2001). To be flexible, companies will keep high inventory in raw materials or finished goods so that they are ready if there is any changes or additional demand from customers. However, this is contrast to green philosophy. Besides materials, companies have to ensure the technology and processes used are flexible enough to cater the business uncertainty. It is a big challenge to company to have efficiency and flexibility exist concurrently.

Efficiency promotes to keep low inventory as low as possible, and try to avoid frequent machine conversion that could cause material and energy wastage. Flexibility is contradicted to concept of efficiency. It is really not easy for companies to balance between flexibility and efficiency especially when environment management comes into picture. Companies need a comprehensive consideration in decision making in order to sustain in both economic and environment developments. There is no one best single supply chain strategy that can meet customer's requirement. Therefore, leagile supply chain strategy was developed by Mason-Jones at al., (2000) by combining some elements of both strategies to fill the gaps. Leagile is using make-to-stock strategy/lean for high volume and stable demand products while make-to-order/agile strategy for everything else. This approach increases the flexibility in production capacity to meet demand upside or unexpected requirement. It helps company to achieve eco-efficiency without impacting company responsiveness.

Consumption System

Resource and energy consumption covers total ecosystems and cycles that impact earth conditions (Esty and Winston, 2006). Worldwide population continues growing and estimated to reach 8 billion by 2024 (United Nations, 2004). As at 2015, population has made a breakthrough of 7.3 billion. The population rapid growth rate increased the demand for commodities. The consumption volume and patterns influence heavily the ecological sustainability. As stated in World Wildlife Fund: Living Planet Report 2000, total global consumption of natural resources has risen by fifty percent since 1970, while earth's natural wealth has decreased by over thirty percent. This was due to the population growth and increase in consumption of many parts of the world. The humanity burden on earth is increasing however it was not corresponding increase in earth's bounty of natural resources.

Industry ecology has emerged with intention to reduce the ecological impacts by manufacturing industry. Industry ecology systematically examines local, regional, and global uses and flows of materials, and energy in products, processes, industrial sectors, and economies (Lowe, 2001). It metaphors manufacturing systems as natural systems to have capabilities to reuse the waste generated from one process as fresh input for another process. Hence, fewer resources extracted and minimal waste generated. This contributed to ecological sustainability.

Resources efficiency means the amount of the goods that come as a consequence of the spending of item resources (UNEP, 2010). This strategy measures the productive utilization and decrease of resources extraction from nature (Cordoba & Veshagh, 2013). The upgrade of 3R (reduce, reuse and recycle) concept to 6R (reduce, reuse, recycle, redesign, recover and remanufacturing) helps to enhance resources efficiency.

The purpose of introduction of 6R (reduce, reuse, recycle, redesign, remanufacture and refurbish) concept into a product life cycle is

to achieve a condition of a perpetual material flow, which help on minimization of that product ecological footprint (Jawahir, 2006). Besides, 6R concept simplifies and prolongs the product life cycle which could provide satisfaction to environmental oriented customers which may later translate to customer loyalty and better financial performance. 6R applies to product from external like customer return and distribution return as well as from internal like rework, scrap and end of life (Shaharudin et al., 2015).

In the past, the consumed products will be disposed of at their end of useful life. This cradle to grave product life cycle system has shifted to cradle to cradle type when organizations realized that proper return product management potentially generate higher profitability and provide opportunity for sustainable development (Shaharudin et al., 2015). Reverse logistics involve taking back the used or returned products for the purpose of reuse, remanufacture, recover, redesign, recycling and safe disposal of the product or materials (Eltayeb et al., 2011). The 6R concept is the most broadly used strategy in reverse logistics. Reverse logistics is applicable internally and externally. The waste substance that created during manufacturing process can be transformed into usable items via internal reverse logistic. The products that collected back from external could be unwanted used products from customers, end-of-life products and defective products from customers.

A firm that concerns about ecological impacts and the benefits of stakeholders is a firm that possesses specific capabilities which make the firm to be more competitive than its competitors. The ecological performance not only enhance economic performance in term of cost reduction and environmental impact minimization, the most important with the competitiveness built up the firm could progress and continue to sustain in market. Firm can build up the green image and create customer loyalty for sustainable developments.

The below table lists out parameters of environment performance that currently in use vs the ecological performance that proposed to be used

Table 3. Comparison between ecological performance and environment performance (Boons & Wagner, 2006)

Item	Environment Performance	Ecological Performance
Performance Variable	- Environment impact intensity - Resource use intensity	- Environment impact intensity - Resource use intensity - Quality (product quality, process quality, workforce quality) - Flexibility (delivery flexibility, production flexibility, supply flexibility) - Consumption system (energy use intensity during usage, recyclability)
System Level	Firm	Market
Boundaries	Individual firm	Firms that competing for same customer base
Linkage to economic performance	Profit maximization via resource consumption and waste reduction	Build up competitiveness to outperform the competitors. To be market leader in term of cost effective, quality, flexibility with leading capabilities
Motivation to Managers	High resource and production efficiency will enhance operational performance and financial performance	Proactive and value added to boost up company's competitiveness in market. Enhancement on company green image and create customer loyalty

in future. The ecological performance matrix apparently covers wider scope and more persuasive to manufacturers in adopting environmental strategies into business activities.

SOLUTIONS AND RECOMMENDATIONS

According to Bey et al. (2013), the main barriers that stop manufacturing companies from implementing environmental strategies are due to a lack of information on environmental impacts, expert knowledge, and allocated resources like manpower and time. These signify that a company would need to invest before the environmental strategies are implemented. Due to the pressure from customers and legislations, many manufacturing companies had started to engage business activities with sustainability efforts. Managers need more information and a justifiable performance measurement metrics to help them to make decisions in implementing the sustainable initiatives (Szekely & Knirsch, 2005).

Manufacturing companies have to ensure a balance tradeoff between environment and economic. They will be more proactive in adopt-

ing environmental strategies if the initiatives implemented can benefit them in wider scope. Besides reduction in environmental impacts and materials and energy consumption, there are other dimensions that can be included into performance measurement metrics.

A holistic performance metric should contain all three elements of sustainable developments: ecological, economic, and social. Economic metrics can be gross margin, return on investment, cost savings, and customer satisfaction. Ecological metrics can be CO₂ emissions, waste output, resources and energy consumption, and volume of pollutants. Social metrics can be employee satisfaction, community initiatives and ratings by publicity.

FUTURE RESEARCH DIRECTIONS

The ecological performance should be proliferated to developing countries. Economic performance is the most important motivator in encouraging companies to adopt environmental strategies. Anbumozhi and Kanda (2005) found that most of the developing countries remained to be the traditional command-and-control or end of pipe

solutions which is very passive approach to reduce negative environmental impacts after they are created. This is because business benefit is the first priority to be considered in adoption of green supply chain practices. Companies are preferable of conventional technologies and products that have been tried and tested. Besides that, the adoption of green supply chain management practices also subject to affinity for short term rather than long term operational benefits among development stakeholders (Ankrah et al., 2013).

CONCLUSION

The dimensions of performance measurement is subject to the important level of issues that managers would deal with. If the issue is limited at firm level, then information needed by managers would be simple thus less performance measurement dimensions are needed. If the consideration is at market level, then more performance measurement dimensions are needed in helping managers to make decision which is effective for overall business and ecological outcomes. A company has to continuously build up competitiveness to outperform the competitors in order to be market leader in term of cost effective, quality, flexibility with leading capabilities. Enhancement on company green image and create customer loyalty are part of the company business strategic management. Thus, ecological performance should be explored as a new metric to measure green supply chain practices.

REFERENCES

Anbumozhi, V., & Kanda, Y. (2005). *Greening the production and supply chains in Asia: is there a role for voluntarily initiatives?* IGES Kansai Research Center Discussion Paper, KRC-2005, No. 6E. Retrieved on 4 March 2016 from <http://www.iges.or.jp>

Ankrah, N. A., Manu, E., Hammond, F. N., Awuah, K. G. B., & Tannahill, K. (2013). Beyond sustainable buildings: Eco-efficiency to eco-effectiveness through cradle to cradle design. Sustainable Building Conference at Coventry University.

Azevedo, S., Cudney, E. A., Grilo, A., Carvalho, H., & Cruz-Machado, V. (2012). *The influence of eco-innovation supply chain practices on business eco-efficiency* (Unpublished doctoral dissertation). University of Beira Interior, Missouri University of Science and Technology, Universidade Nova de Lisboa.

Baumann, E. C., & Genoulaz, V. B. (2014). A framework for sustainable performance assessment of supply chain management practices. *Computers & Industrial Engineering*, 76, 138–147. doi:10.1016/j.cie.2014.07.029

Bey, N., Hauschild, M. Z., McAloone, T. C. (2013). Drivers and barriers for implementation of environmental strategies in manufacturing companies. *CIRP Annals - Manufacturing Technology*, 62, 43–46.

Boons, F., & Wagner, M. (2009). Accessing the relationship between economic and ecological performance: Distinguish system levels and the role of innovation. *Ecological Economics*, 68(7), 1908–1914. doi:10.1016/j.ecolecon.2009.02.012

Bran, F., Radulescu, C. V., & Ioan, I. (2011). Measures of Environmental Performance. *Review of International Comparative Management*, 12(5), 893–899.

Cambridge Business Dictionary. (n.d.). Retrieved March, 18, 2016 from <http://dictionary.cambridge.org/dictionary/business-english/>

Chien, M. K. (2014). Influences of green supply chain management practices on organizational sustainable performance. *International Journal of Environmental Monitoring and Protection*, 1(1), 12–23.



- Christopher, M., & Towill, D. R. (2000). Supply chain migration from lean and functional to agile and customized. *Supply Chain Management: An International Journal*, 5(4), 206–213. doi:10.1108/13598540010347334
- Cordoba, D. B., & Veshagh, A. (2013). *Managing eco design and sustainable manufacturing*. 20th CIRP International Conference on Life Cycle Engineering, Singapore.
- Eltayeb, T. K., Zailani, S., & Ramayah, T. (2011). Green supply chain initiatives among certified companies in Malaysia and environmental sustainability: Investigating the outcomes. *Resources, Conservation and Recycling*, 55(5), 495–506. doi:10.1016/j.resconrec.2010.09.003
- Esty, D. C., & Winston, A. S. (2006). *Green to gold: how smart companies use environmental strategy to innovate, create value and build competitive advantage*. Yale University Press.
- Fisher, M. L. (1997). What is the right supply chain for your product? *Harvard Business Review*, 75(2), 105–116.
- Gunasekaran, A., & Spalanzani, A. (2012). Sustainability of manufacturing and services: Investigations for research and applications. *International Journal of Production Economics*, 140(1), 35–47. doi:10.1016/j.ijpe.2011.05.011
- Hart, S. L. (1995). A natural-resource-based view of the firm. *Academy of Management Review*, 20(4), 986–1014.
- Hawken, P., Amory, B., Lovins, L., & Lovins, H. (1999). *Natural capitalism: Creating the next industrial revolution*. Boston, MA: Little, Brown and Co.
- Jackson, S. A., Remani, V. G., Mishra, R., & Napier, R. (2016). Examining the impact of design for environment and the mediating effect of quality management innovation on firm performance. *International Journal of Production Economics*, 173, 142–152. doi:10.1016/j.ijpe.2015.12.009
- Jawahir, I. S., Dillon, O. W., Rouch, K. E., Joshi, K. J., Venkatachalam, A., & Jaafar, I. H. (2006). *Total life-cycle consideration in product design for sustainability: A framework for comprehensive evaluation*. 10th International Research/Expert Conference. Barcelona-Lloret de Mar, Spain.
- Johnstone, N. (2007). *Environmental policies and corporate behavior*. OECD/Edward Elgar.
- Kudroli, K. (2014). Green supply chain management and environmental sustainability – a comparative study on global and India perspective. *International Journal of Conceptions on Management and Social Science*, 2(2), 2357–2787.
- Laari, S., Toyli, J., Solakivi, T., & Ojala, L. (2016). Firm performance and customer-driven green supply chain management. *Journal of Cleaner Production*, 112, 1960–1970. doi:10.1016/j.jclepro.2015.06.150
- Lowe, E. (2001). *The Eco-Industrial Park handbook for Asian developing countries*. Retrieved on 4 March 2016 from <http://www.indigodev.com>
- Mason-Jones, R., Naylor, B., & Towill, D. R. (2000). Lean, agile or le-agile? Matching your supply chain to market place. *International Journal of Production Research*, 38(17), 4061–4070. doi:10.1080/00207540050204920
- Rosen, M. A., & Kishawy, H. A. (2012). Sustainable manufacturing and design: Concepts, practices and needs. *Journal of Sustainability*, 4(12), 154–174. doi:10.3390/su4020154
- Shaharudin, M. R., Govindan, K., Zailani, S., & Tan, K. C. (2015). Managing product returns to achieve supply chain sustainability: An exploratory study and research propositions. *Journal of Cleaner Production*, 101, 1–15. doi:10.1016/j.jclepro.2015.03.074
- Simao, L. E., Goncalves, M. B., & Rodriguez, C. M. T. (2016). An approach to assess logistics and ecological supply chain performance using postponement strategies. *Ecological Indicators*, 63, 398–408. doi:10.1016/j.ecolind.2015.10.048

Srivastava, P. K., Kulshreshtha, K., Mohanty, C. S., Pushpangadan, P., & Sing, A. (2005). Stakeholder-based SWOT analysis for successful municipal solid waste management in Lucknow India. *Waste Management (New York, N.Y.)*, 25(5), 531–537. doi:10.1016/j.wasman.2004.08.010 PMID:15925762

Szekely, F., & Knirsch, M. (2005). Responsible leadership and corporate social responsibility: Metrics for sustainable performance. *European Management Journal*, 23(6), 628–647. doi:10.1016/j.emj.2005.10.009

Thoo, A. C., Huam, H. T., & Sulaiman, Z. (2015). Green Supply Chain Management, Environmental Collaboration and Sustainability Performance. *Procedia CIRP*, 26, 695–699. doi:10.1016/j.procir.2014.07.035

United Nation Development Report. (2004). Retrieved on 4 March 2016 from <http://www.un.org>

United Nations Environment Programme Report. (2010). Retrieved on 4 March 2016 from <http://www.unep.org/resourceefficiency>

Vachon, S., & Klassen, R. D. (2006). Extending green practices across the supply chain: The impact of upstream and downstream integration. *International Journal of Operations & Production Management*, 26(7), 795–821. doi:10.1108/01443570610672248

Wang, Y., Chen, Y., & Benitez-Amado, J. (2015). How information technology influences environment performance: Empirical evidence from China. *International Journal of Information Management*, 35(2), 160–170. doi:10.1016/j.ijinfomgt.2014.11.005

Zhu, Q., Sarkis, J., & Lai, K. (2007). Green supply chain management: Pressures, practices and performance within the Chinese automobile industry. *Journal of Cleaner Production*, 15(11-12), 1041–1052. doi:10.1016/j.jclepro.2006.05.021

Zhu, Q., Sarkis, J., & Lai, K. (2013). Institutional-based antecedents and performance outcomes of internal and external green supply chain management practices. *Journal of Purchase Supply Management*, 19(2), 106–117. doi:10.1016/j.pursup.2012.12.001

KEY TERMS AND DEFINITIONS

Company Competitiveness: The ability of a company to preempt the competitors in the same market and continue to be market leader.

Eco-Efficiency: A strategy that emphasize on resource and energy optimization and waste minimization. It advocates balance development between economic and ecology.

Ecological Performance: A measurement metric that used to measure and numerically marking the impacts towards ecological and economic sustainability by a series of human activities.

Environmental Performance: A measurement metric that used to quantify the environmental impacts intensity caused by anthropogenic activities.

Green Supply Chain Management: Supply chain management that integrated environmental element into the flow from upper stream to downstream. Green supply chain activities including green purchasing, green manufacturing, green logistics, green distribution and reverse logistics.

Industry Ecology: A study about manufacturing activities and their impacts towards ecosystems and ecological sustainability. Its intention is to find ways to mitigate the issues aroused by manufacturing activities.

Resource Efficiency: A cradle to cradle approach in managing the useful life of extracted materials. The materials can be recycled to different forms to extend its perpetuality.

Sustainable development: Developments that are balance between economic, environmental and social. Economic developments should not built upon the sacrifices of human and other living things on the earth planet.

E-Commerce in Logistics and Supply Chain Management



Yasanur Kayikci

Turkish-German University, Turkey

INTRODUCTION

The rapid development of computing services and information and communication technologies (ICT) in the 1990s has dramatically transformed the way how organizations use logistics and supply chain operations for achieving competitive advantage. This has led to the emergence of the e-commerce context (Turban, et al., 2007; Harris et al., 2015). Gaining a competitive advantage in e-commerce means finding a balance of the right item price, customer service and delivery time, therefore e-commerce includes innovative and highly scalable e-commerce platforms (Chen et al., 2012). The impact of e-commerce covers both business-to-business (B2B) and business-to-consumer (B2C) transactions. Simply to point out, e-commerce is conducted through the Internet computer network products, trading, payment and services and other economic activities, and also it integrates all inter-company and intra-company functions, meaning that the three flows namely, physical flow, financial flow and information flow of the supply chain are affected by e-commerce.

The digital revolution has driven the development of e-commerce. Early applications can be traced back to the 1960s, but e-commerce only became popular in the 1990s due to the widespread use of the Internet. E-commerce uses communications such as EDI, POS, electronic ordering system, electronic mail, file transfer, video conferencing, workflow or interaction with a remote computer to improve efficiency of logistics and supply chain activities, and to enhance the entire supply chain management decision capacity. E-commerce has also become associated with solutions and

tools such as portals, e-marketplaces, e-auctions or virtual inventory. Logistics and supply chain operations have been affected irrevocably by e-commerce solutions. But the possible major changes could take a lot longer to happen. For companies, the overall goal of e-commerce includes (Yang, 2012): to help companies develop global sales network and logistics infrastructure which support efficient online order fulfillment; to provide enterprises with information and data of various business activities (i.e. online order documentation), production and sales information to solve the difficult problem of collection; to reduce market entry link that help companies open up the market to minimize the circulation of goods; to reduce business cost of sales and minimize the transaction costs of goods; both for trading and online commodity trading to facilitate negotiations; supports online purchasing including secure online payment system; to provide the most reliable quality assurance; to provide customers the most convenient means for retrieval. Supply Chain Management (SCM) is based on customer demand, providing a common focus on the related products or services, enterprise information resources, software products based on Internet technology as a tool to manage, the entire channel which all relevant processes to maximize net value added, and improve the efficiency and effectiveness of a platform. In this chapter, the evolution of logistics and supply chain and trends of e-commerce in logistics and supply chains are described. Then the discussion on the integrated e-commerce and e-commerce logistics are given and finally, it ends with future research directions and conclusion.

DOI: 10.4018/978-1-5225-2255-3.ch466

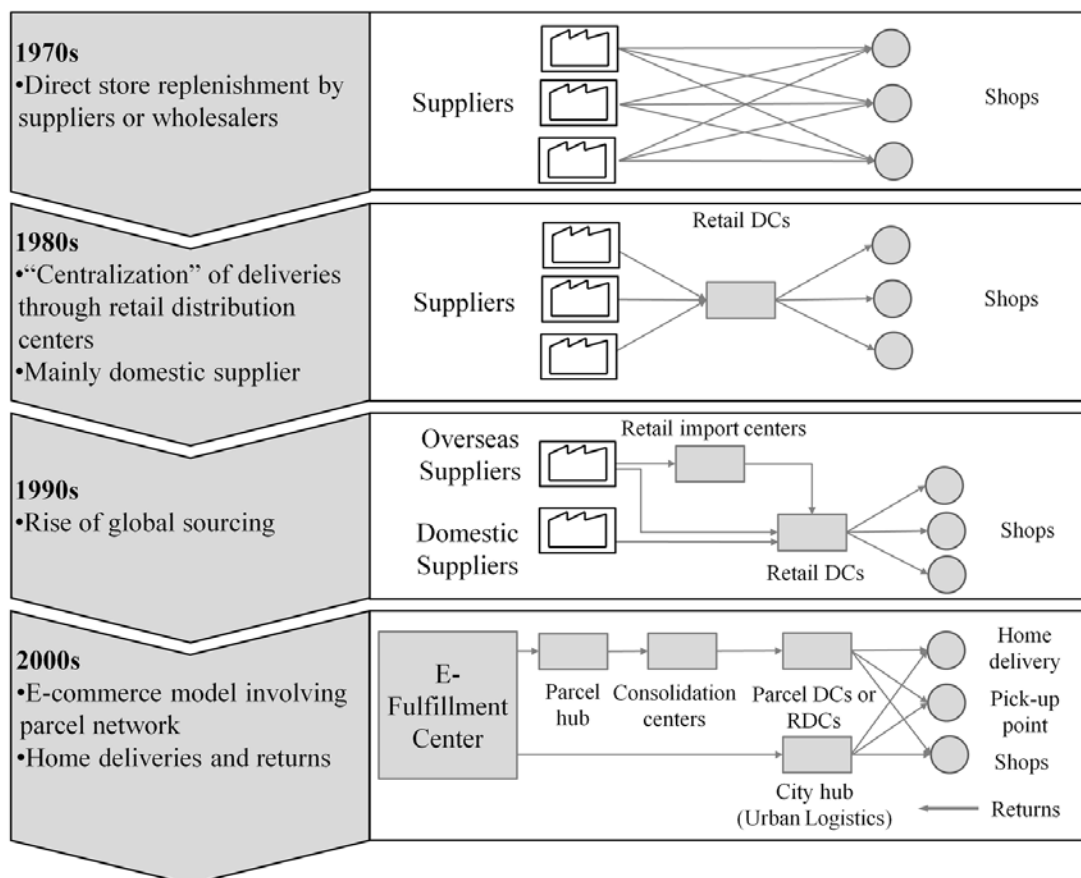
BACKGROUND

The rapid growth of e-commerce is being related to economic development and often been cited as a driver of economic growth. E-commerce is also touted as a powerful medium through which developing economies can exploit the potential of global markets. Today, e-commerce in developed economies represents the latest leader of change in physical distribution network and SCM, whereas in developing economies, e-commerce market is still in its infancy but it is expanding and becoming integral part of commercial life. This field evolved substantially over the past four decades and continues to grow today. In the early days, low prices and convenience were the driving forces for e-commerce, but today, e-commerce retailers (also known as e-tailers) are catering to

every kind of shopping experience, service and product. Presently, it remains the case that as the gross revenue earned from e-commerce is increasing year on year from nascence to critical mass, most transport service providers (TSP) or retailers, particularly multi- and omni-channel retailers, are only starting to realize what this will bring to their retail logistics and physical distribution network infrastructures and what role they can have within. So much as, today e-commerce has become a complementary retail touch point for the traditional brick-and-mortar retail.

The new e-commerce technologies have the potential to offer greater efficiencies and transparency and have led to significant changes within supply chains, with alternative ways of doing business, improved visibility and changes to distribution channels, including new interme-

Figure 1. The evolution of logistics and supply chain



diaries. From a distribution property perspective, the evolution of logistics and supply chain has passed through various phases and very broadly in the timeline as follows (JLL, 2013) (see Figure 1): most retail stores were replenished by direct deliveries from wholesalers and suppliers in the 1970s. The distribution centers (DC) emerged and retailers started to centralize store deliveries through new DCs that they controlled in the 1980s. Global sourcing took off in the 1990s, with many retailers developing import centers to receive and process mostly containerized imports. From around 2000, e-commerce began to rapidly expand with Internet-only (pure-play) retailers leading the way in establishing giant e-fulfillment distribution networks. In 2000, the business concepts like information sharing driving the extensive and intensive cooperation, collaborative warehousing, collaborative urban (city) as well as non-urban distribution are merged and implemented in logistics and supply chain operations (Capgemini, 2008). According to a study in developed economies the growth of online retail has been stronger in sectors such as fashion, electrical and ICT goods (Robinson, 2014). In the past times and models, there weren't many carrier options and purchased items were often distributed via a parcel, postal or freight network. However, e-commerce logistics models have created new demands.

There are four distinct types of logistics functions being setup (Robinson, 2014): An *e-fulfillment* center is a warehouse at which goods are stocked and chosen at item level. Today, there are numerous e-fulfillment centers, which are operated either by a retailer or a TSP such as eKart for Flipkart, JaVAS for Jabong and ATS for Amazon etc., often operating 24/7. Key facility configuration includes size (also called as big-box space which ranges from 50,000 to more than 100,000 m²) as well as features that contribute to more efficient picking, order fulfillment and throughput, such as accessibility for delivery vans, access to a strong employee base and cross-dock configuration. A *parcel hub/consolidation* center is where orders are sorted

according to geographical location, often by post or zip code, so they can be delivered to the relevant *parcel delivery* center for final delivery to the customer's home or designated collection point. Parcel delivery center handles the last mile delivery to the customer. *Seamlessly integrated technology* is where shopping carts connect via WEBXML, API or some other connection to a TMS, so customers get the exact price quote of shipping of larger items. Quotes are often more suited for less than truckload, as these technology products for logistics, such as a TMS, must accomplish along with the shopping cart for better management. These technology products for logistics feature: the ability to organize and track shipment according to transport modes; online order documentation and status; online dispatch invoice and documentation such as freight bills of lading and invoices; auto reminder for payments; seamless interface with existing ERP or SCM system; online alerts for critical information via mobile or text, information systems reports on delivery history, past data analysis, etc. Types of e-commerce logistics systems based on these considerations ensure the following benefits to transport users (e.g., shippers, receivers) and TSPs (e.g., carriers): improved communication, transparency into the supply chain, improved customer satisfaction, distribution and logistics optimization, cost reduction, improvement in efficiency and on-time delivery.

As online retail grows further, delivering order to customers quicker will become an increasing competitive advantage. As a result, the logistic facilities will encourage some retailers to set up their own networks of local depots - either to cross-dock items shipped from larger e-fulfillment centers or to ship certain fast moving products direct to customers. In this emerging model, e-fulfillment blends with urban logistics, as these facilities will be mainly based around the major population centers where online sales densities are highest. For example, in the U.S., Amazon opened smaller scale distribution facilities to offer same-day delivery services.

TRENDS OF E-COMMERCE IN LOGISTICS AND SUPPLY CHAINS

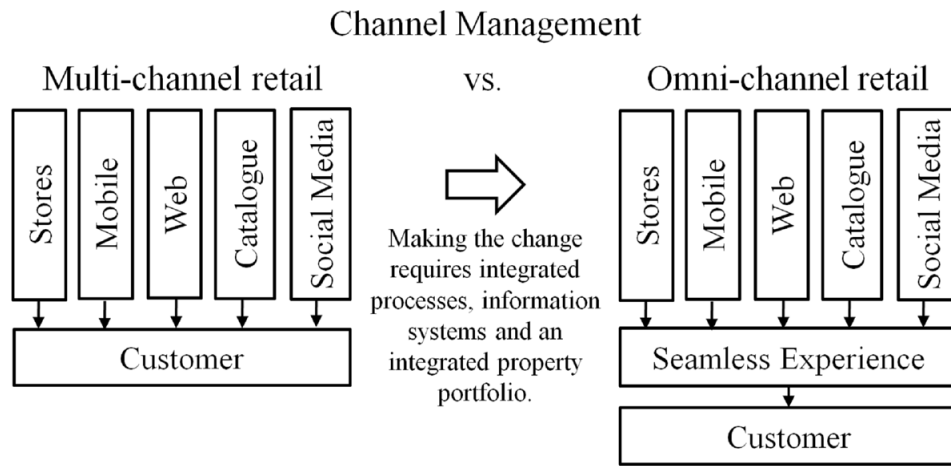
The e-commerce context is changing at a steady rate in both B2B and B2C, although it only represents 10% of all new leasing around the world, this percentage has doubled during the past three years, therefore companies have to anticipate strategies to manage these changes in their favor. Recognizing online presence as one of the biggest opportunity to expand and grow, many physical retailers have started to establish and increase their online presence. An e-commerce platform not only facilitates a transaction over the web, but it supports the creation and continuing development of an online relationship. The emerging trends of enabling e-commerce systems have been subject of discussion in recent years.

- **Cloud Computing:** There are four basic service configurations which are Software-as-a-Service (SaaS), Platform-as-a-Service (PaaS), Infrastructure-as-a-Service (IaaS) and Supply Chain-as-a-Service (SCaaS). In SaaS, the cloud user is able to access a software application hosted by an external provider over the Internet. Such as Google Apps, Salesforce. PaaS model describes a virtual platform which involves the use of an externally provided infrastructure to host the application like Apprenda, Google App Engine. IaaS provides virtualized computing resources over the Internet such as Amazon Web Services, Cisco Metapod, Google Compute Engine. SCaaS describes a framework for service oriented SCM such as OpenPMF SCaaS, HCL SCaS.
- **Mobile Applications:** Customer shopping habits have changed driven in part by increased use of mobile devices. Continued growth in tablets, smartphones, handheld devices use boosts the amount of time customers spend online – and they're more likely to shop there. With the proliferation of the Internet into daily life and the falling

data prices, customers access mobile applications conveniently to shop from any location like at airports, traffic signals, cafes or waiting in the line. This ease of ordering through mobile has further fueled impulse purchases. Companies are investing more resources for optimization of their mobile sites to enable customers to have seamless shopping experience across platforms.

- **Personalization for Customer Engagement:** Personalizing customer experience is a key customer engagement tool and is expected to drive e-commerce growth. With increase trend towards online shopping, brands are becoming more sophisticated in their ability to collect information about customers and using this data to deliver personalize recommendations, tailored offers and delivering differentiated experiences each time.
- **Big Data:** The Big Data has arguably been generated primarily from web and e-commerce communities. Significant market transformations have been accomplished by leading e-commerce vendors such as Amazon and Ebay through innovative and highly scalable e-commerce platforms and product recommender systems. Major internet companies such as Google, Amazon and Facebook continue to lead the development of web analytics, cloud computing and social media platforms.
- **Social Networks:** Social media has shown to be an effective platform for customers to discuss and voice opinions about products and brands, as well as for business to communicate with customers (DHL, 2015). Social media plays an important role. There are four different segments where marketing channels fit along the customer path to purchase: awareness, consideration, intent, and decision. Despite not being a direct sales channel, social media can support e-commerce logistics and supply chains in the ways (O'Leary, 2012):

Figure 2. Paradigm change in e-commerce



Demand information: using the percentage of likes and dislikes from social media exchanges as a measure of interest of customers in a product, thus providing some potential insight into customer demand. *Supplier choice:* Data mining of media such as Linked-In can provide important data to help infer information about whether a company will continue in business. *Logistics:* Information gathered over time about particular highways can help develop a better understanding of likely highway conditions. Effectively, intuition can be generated about routes and the likelihood of problems being encountered when moving goods over those routes

INTEGRATED E-COMMERCE

The evolution of multiple shipping options allows retailers today more control over their e-commerce logistics and supply chains. Today's customers have the ability to purchase whatever, whenever, however, wherever and at the price they want, putting them in total control of the market. With social media, mobile and e-commerce on the rise, omni-channel fulfillment is becoming increasingly important for retailers. Companies

in B2B and B2C commerce must be flexible and as omni-present as possible in order to remain competitive in this shop-anytime-anywhere world. As seen today, this trend is starting to displace the traditional role of parcel operators. Now the e-commerce challenge has shifting to finding a way of synchronizing and standardizing the business processes to achieve real time access and insight of the inventory movement. As predicted, retailers in developed markets are experiencing to shift from multi-channel retail to omni-channel retail. Paradigm change in e-commerce is depicted in Figure 2. Omni-channel retailers are managing their channels in an integrated way that offers customers a seamless experience, however they choose to shop. With omni-channel, a retailer may fulfill orders from local stores rather than warehouses, ultimately blurring the distinction between the two e-fulfillment centers.

With extensive number of sales channels, multiple warehouses, and dozens of suppliers, the risk of misplaced orders is much higher than ever before. In order to respond to this challenge, order fulfillment technologies have also helped integrate the front-end and back-end of online retail. Automated software and real-time fulfillment data transform the back-end process now in a collaborative effort. The alignment of important touch-points in the supply chain has

reduced inefficiencies and has helped identify redundant processes. There are even robots that pick inventory and move it around the warehouse (Robinson, 2014).

- **Multi-Channel:** Customers interact with retailers across the independently managed channels (e.g. social media, web and emails). Retailers with this approach are adopting two or more channels to engage their customers, however, they are not necessarily focused on delivering a seamless/consistent message across multiple touchpoint. Furthermore, these programs do not necessarily factor in optimizing customer experience based on different devices (e.g. smartphones, tablets and laptops).
- **Omni-Channel:** This business is diligent to ensure that customers receive the same experience and message through different channels and devices involved within their interactions with the retailer (Kourimsky & van den Berk, 2014). Thus, omni-channel has seamless integration with company software systems like ERP, WMS, warehouse control system, distributed order management (Hobkirk, 2016). Typically, omni-channel customers require their orders to be delivered within either same day or next day to their location of choice, whether the order is distributed to their home or can be picked up in a store or another location, moreover this type of customers expect to be informed of every step and any delays which may occur along the delivery process. Similarly, retailers have a holistic view into inventory in omni-channel environment, allowing them to fulfill orders from any location or ship inventory to other locations where inventory is needed or sells better. Omni-channel DCs share a common inventory pool and seamlessly combine both e-commerce and traditional store distribution channels in order to meet demand regardless of which

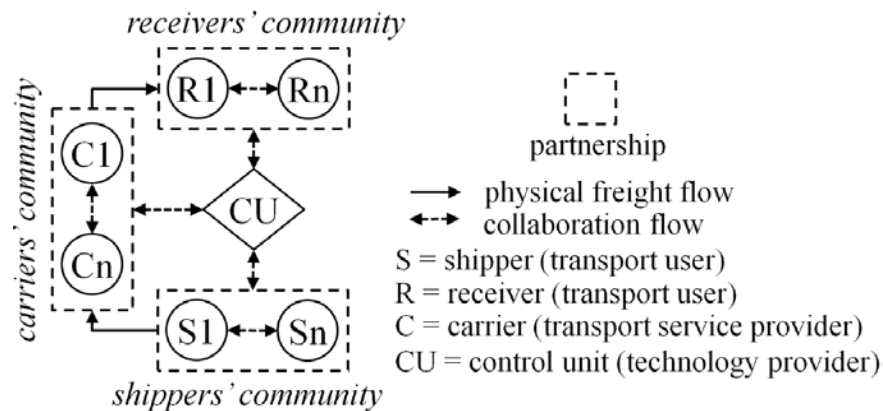
channel it comes from, whether customers are shopping online from a desktop or mobile device, by telephone or offline in a brick and mortar store. Basically, the goal of omni-channel is to give retailers an end-to-end view based on aggregated information from suppliers to customers, orders, and inventory.

In the past, retailers built two types of DCs, one to manage store fulfillment, whereas another to manage purely e-commerce. Omni-channel DCs seamlessly combine both e-commerce and traditional store distribution channels, omni-channel fulfillment strategies must be done right and quickly. Omni-channel improves the customer service through real-time integration with POS, web storefront and call center solutions, furthermore it increases flexibility and reduced costs through common carrier, small package and private/dedicated fleet delivery planning and optimization. It improves delivery control through transport execution capabilities including tendering, multi-carrier parcel manifesting, fleet dispatching, routing, proof of delivery and freight auditing, additionally transport operations are streamlined through carrier connectivity and global logistics network. To meet the challenges, retailers must completely redesign their distribution processes and add automation equipment to their DCs. In order to support the fulfillment promise, shippers must have technology and infrastructure in place to ensure sufficient response time between when customer places an order and when fulfillment center dispatches it. Additionally, significant capital investment is required in material handling, conveyor sortation and controls, optimized racking systems and lift equipment, inventory management software, and picking/packing technology.

E-COMMERCE LOGISTICS

In the context of logistics, e-commerce platforms can be termed e-commerce logistics platforms.

Figure 3. Overview of e-commerce logistics platform with consortia partners



E-commerce logistics is the use of web-based technologies to support the material acquisition, warehousing, and transportation processes. It enables distribution to couple routing optimization with inventory-tracking information. This platform is an inter-organizational system (IOS) that links transport users and TSPs together for the purpose of collaboration or trading. It enables online interactive transport exchanges in terms of transport supply and transport demand in order to match freight capacity with available shipments (Christiaanse, 2005; Kale et al., 2007; Wang et al., 2007). For example, internet-based freight acquisitions enable spot buying of trucking capacity. TSPs offer virtual logistics services by integrating and optimizing distribution resources. A company may even consider collaboration with its competitors to improve its supply chain. Utilizing the web has largely reduced the complexity and cost of implementation and the integration of IOSs. Incorporating the concept of cloud computing services, e-commerce platforms are emerging quickly as a viable alternative to large-scale client-server solutions (Kayikci & Zsifkovits, 2013; Wang et al., 2007), especially where they are hosted by the technology provider. Traditionally, transport software programs (e.g., TimoCom, TransMATCH) are developed for specific platforms, such as Windows, Linux, or Mac OS, where parties are advertising freights (cargoes) and looking for empty trucks. Today,

platform developers build web-based (e.g., www.timocom.com) as well as mobile-based (e.g., TimoCom transportbarometer) applications, as communication tool in supply chain that run on Internet and on mobile, that are completely independent of the user's actual computer operating system. E-commerce platforms offer powerful online solutions for transport collaboration, where TSPs easily search online for freight or post any spare capacity in order to increase load factor for return trip. Similarly, transport users offer freight for transportation or search for suitable vehicles.

Today companies (e.g. Shipwire) access end-to-end orchestration and fulfillment capabilities on web-based enterprise fulfillment platforms (DHL, 2015). These make it easy to access specific services and grow into new markets. Customers can choose their preferred transport option and this is factored into the total cost of the product (e.g. FreightOS). E-commerce logistics platforms offer three different types of business activities (Kaplan & Sawhney, 2000):

- *Aggregators* create a business community where TSPs offer freight capacities with predefined prices (e-catalog) for transport users on a website (e.g., Freefreight search).
- *Auctions* create market and reduce empty trip and promote the load factor, where transport users can bid simultaneously

for the TSP's freight capacity according to multiple real-time auction systems (e.g., FReight Exchange for International Transport-ONlinE).

- *Exchanges* create stable online market, where selected transport users and TSPs trade freight according to a fixed set of rules and at constantly changing prices (e.g., Transplace); this Expedia-like platform for freight providing instant, online real-time quoting for all legs of shipment. Shippers can generate and book instant, door-to-door quotes directly online, using their own internal negotiated rates or rates shared online by carriers or forwarders.

Characteristics of an E-Commerce Logistics Platform

- **Platform Community:** The network for transport exchanges allows transport users and TSPs to create *communities* to better manage their transport needs and resources (Kayikci & Zsifkovits, 2013). Figure 3 depicts an overview of e-commerce logistics platform with various consortia partners. If a transport coalition is composed of pre-selected members on *vertical plane*, it is called then *private community*; most of the time this community is closed and specific requirements should be fulfilled to enter into this community. If a transport coalition is composed of only receivers, then it is called *receivers' community* where the platform primarily serves interest of receivers; likewise, if it is composed of only shippers, which is called *shippers' community* and the platform serves the interest of shippers; if only carriers are collaborate, then it is *carriers' community* where the platform primarily serves interest of carriers. If shippers, receivers and carriers integrate together on the *lateral plane*, it is denoted as a *neutral community* which has generally heterarchical structure and serves interest of all participants and is often managed by a group of consortium. That means, there is no dominant community which rules the coalition, therefore

an independent platform usually tries to be of direct use for shippers, receivers and carriers. *Neutral community*, where the aforementioned three communities collaborate each other and their communications are supported by a neutral technology provider which is called as control unit (CU). This community is characterized by a high level of transport collaboration. It is web-based and hosted platform and enables high level information sharing and joint activities.

- **Platform Strategy:** E-commerce platforms are divided into two types: open and closed platforms. *Open platform* allows transport parties to use their services with no barriers to entry whereas *closed platform* tends to be geared towards the needs of particular shippers, receiver and carriers (Wang et. al, 2007). Closed platform denotes private community in other words, semi-private or private networks where the all participants know each other and they share information on shipping requirements. Most of the transport users prefer to collaborate in a private network rather than in an open network, however close platforms are less structured and often contain the opinion of large companies (Chen et al. 2012), whereas open platforms acquire bigger negotiation capability and offer more freight matching opportunities than closed platforms. However, the open platforms are not suitable in high level of transport collaboration (Cruijssen et al., 2007; Wang et. al, 2007). Especially strategic partnering level requires closed platform strategy.
- **Platform Sourcing:** The platforms also differ in how total freight load capacity is bought. Generally, transport users either engage in *systematic sourcing* or in *spot sourcing* (Kaplan & Sawhney, 2000). First, transport users can acquire freight capacity in a truck through fixed-commitment contracts and negotiated market prices

with TSPs. Especially major shippers and forwarders use *systematic sourcing* for acquiring available freight capacity (Kayikci et al., 2014). The *closed private platforms* exemplify this approach. Second, in *spot sourcing*, transport user's goal is to fulfill an immediate need at the lowest possible cost, whereas TSP's goal is to utilize truck space at the highest possible profitability. The *open source auction platforms* elucidate this approach. Spot transactions are being accompanied by development of a short-term, loose relationship among transport users and TSPs.

- **Platform Capability:** Development of an online platform includes creation and management of an online storefront, shopping cart management, PCI compliance, personalization transaction management, settlement and product visualization – enable organizations to build basic B2B or B2C online stores. Some specific capabilities in electronic communications are determined (Hajdul, 2014) that are essential for e-commerce platforms with the collaboration of transport users and TSPs: web-interfaces, event management / alerting functionality based on (re-)calculated routes/plans, reporting, transport/purchase order management at company level, monitoring of the performed task, digital map, route optimization, invoicing, fleet management, freight exchange, real-time monitoring of the performed tasks, transport/purchase order management at group of independent companies level, verification of business partners, coordination of transport orders and resources of independent companies, optimization of truck loading process, share of savings among group of cooperating companies, load tracking, deliveries, support of existing communication standards, settlement and reports. E-commerce

platforms can also include personalization/preference profiling, multichannel selling, site/product search, search engine optimization, customer community management or participation, integration with social media, and mobile stores.



FUTURE RESEARCH DIRECTIONS

Customer buying behavior is evolving faster than logistics and supply chains can adapt. The rapidly changing delivery expectations of e-commerce customers, driven by trends towards click & collect and next day delivery, mean that operating an effective e-commerce supply chain and fulfilment operation presents a whole new set of challenges. To meet these new expectations, e-tailers need to find new solutions at every stage, from demand forecasting, through to inventory management, warehousing, technology integration and distribution. A strong mobile presence, in-stock merchandise, and fast delivery are important keys to e-commerce success, and a strong argument for top-notch logistics solutions. Today, supply chains are segmented by channels, which results in duplication of inventory and infrastructure. Only few companies (such as Vargo) are beginning to offer omni-channel DCs, so that separate channel spaces within the same warehouse and different automation solutions are used in order to easily respond the changing supply chain demands of B2B and B2C channels. Unfortunately, there is still limited research associated with how e-commerce in terms of new trends has been and can be used in the logistics and supply chains. From this perspective, many opportunities for future research might be opening for researchers. These new research areas might be big data analytics in logistics and SCM, adapting industry 4.0 in logistics and supply chain, setting the strategies for omni-channel logistics, omni-channel integration as well as designing omni-channel DCs.

CONCLUSION

Expanding ICT has broken down barriers to make it easier to start and grow a business online in every industry and logistics and SCM continue to be pushed in new direction by e-commerce. The new trends in e-commerce like cloud computing, mobile applications, big data, social networks can offer great opportunities for companies in order to redefine their corporate strategies and to redesign their business models. The percentage of e-commerce sales is growing continuously and the increased volumes are putting unprecedented demands on the supply chains of e-commerce and omni-channel retailers. In this chapter, a broad overview of current e-commerce applications in logistics and SCM is given. It gives researchers and practitioners better understanding the role of e-commerce in logistics and SCM in order to expand business and provide competitive advantages.

REFERENCES

- Capgemini. (2008). *Future supply chain 2016*. White Paper, Capgemini Research.
- Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *Management Information Systems Quarterly*, 36(4), 1165–1188.
- Christiaan, E. (2005). Performance benefits through Integration hub. *Communications of the ACM*, 48(4), 95–100. doi:10.1145/1053291.1053294
- Crujssen, F., Dullaert, W., & Fleuren, H. (2007). Horizontal cooperation in transport and logistics: A literature review. *Transportation Journal*, 46(3), 22–39.
- DHL. (2015). *Omni-channel logistics: A DHL perspective on implications and use cases for the logistics industry*. White Paper, DHL Trend Research.
- Hajdul, M. (2014). Virtual co-opetition in transport: T-Scale platform case study. *Procedia: Social and Behavioral Sciences*, 111, 761–769. doi:10.1016/j.sbspro.2014.01.110
- Harris, I., Wang, Y., & Wang, H. (2015). ICT in multimodal transport and technological trends Unleashing potential for the future. *International Journal of Production Economics*, 159, 88–103. doi:10.1016/j.ijpe.2014.09.005
- Hobkirk, I. (2016). Key distribution strategies of top omni-channel retailers. *Logistics Management*, 1, 48–53.
- JLL. (2013). *E-commerce booms triggers transformation in retail logistics*. White Paper, Jones Lang Lasalle.
- Kale, R., Evers, R. T., & Dresner, M. E. (2007). Analyzing private communities on Internet-based collaborative transportation networks. *Transportation Research Part E, Logistics and Transportation Review*, 43(1), 21–38. doi:10.1016/j.tre.2005.07.004
- Kaplan, S., & Sawhney, M. (2000). E-hubs: The new B2B market places. *Harvard Business Review*, 97–103. PMID:11183982
- Kayikci, Y., Stix, V., LeBlanc, L. J., & Bartolacci, M. R. (2014). A novel application of a hybrid delphi-analytic hierarchy process (AHP) technique: Identifying key success factors in the strategic alignment of collaborative heterarchical transportation networks for supply chains. *International Journal of Applied Logistics*, 5(1), 54–77. doi:10.4018/ijal.2014010104
- Kayikci, Y., & Zsifkovits, H. (2013). Successful ICT integration in transport collaboration. In T. Blecker, W. Kersten, & C. M. Ringle (Eds.), *Pioneering solutions in supply chain performance management* (pp. 201–220). Köln: Josef Eul Verlag.

Kourimsky, H., & van den Berk, M. (2014). *The impact of omni-channel commerce on supply chains: How to make sure you effectively deliver products that meet the customer's expectations*. White Paper, Intelligence.

O'Leary, D. E. (2012). *Developing trust and relationships in the supply chain using social media*. Retrieved December 12, 2015, from <http://www.europeanbusinessreview.com/?p=2559>

Robinson, A. (2014). *E-commerce logistics: The evolution of logistics and supply chains from direct to store models to e-commerce*. Retrieved August 14, 2015, from <http://cerasis.com/2014/04/30/e-commerce-logistics/>

Turban, E., Leidner, D., McLean, E., & Wetherbe, J. (2007). *Information technology for management: transforming business in the digital economy*. New York: John Wiley & Sons.

Wang, Y., Potter, A., & Naim, M. (2007). Electronic marketplaces for tailored logistics. *Industrial Management & Data Systems*, 107(8), 1170–1187. doi:10.1108/02635570710822804

Yang, G. (2012). Relationships between e-commerce and supply chain management. In K. Haenakon (Ed.), *Advances in technology and management: Advances in intelligent and soft computing* (pp. 653–658). Berlin: Springer. doi:10.1007/978-3-642-29637-6_87



KEY TERMS AND DEFINITIONS

E-Commerce: E-commerce refers to conducting business communication and transactions over networks and through computer technology or buying and selling goods and services (also known as e-business), and transferring funds through digital communications.

E-Tailer: An e-tailer is a retailer that primarily uses the Internet as a medium for customers to shop for goods and/or services provided.

Omni-Channel: Omni-channel involves providing a seamless, consistent and complementary customer shopping experience across all interaction channels like in store, on the web, and on mobile devices.

Exploring Drivers of Closed Loop Supply Chain in Malaysian Automotive Industry

Fadzlina Mohd Fadzil

Universiti Sains Malaysia, Malaysia

Yudi Fernando

Universiti Malaysia Pahang, Malaysia

INTRODUCTION

With the emergence of industry, people have become voracious pursuit of profit, in which they affect the environment and social wellness. Issues related to environmental and social destruction such as pollution, carbon footprint, ozone depletion, deforestation and waste management problem had threatened the environment and human being on this day. These environmental issues have caused resurgence and challenges in finding a solution to this problem, namely sustainability.

The concept of sustainability has been introduced by the World Commission on Environment and Development (1987), has given a change to the business world today. The issue of sustainability has brought changes to the terms of a business venture of the company's goals, operations, manufacturing, supply chain, administration and marketing. The concept of sustainability has rejuvenated the world of supply chain where green practices in supply chain can reduce environmental impacts. The green supply chain management is defined as the integrated supply chain management where it revolves around the green purchasing ranging from suppliers, manufacturers, users, and so the process will revert to reverse logistics, which will close the loop process of the supply chain (Q Zhu et al., 2005).

BACKGROUND

This study takes the approach of a closed supply chain in the automotive industry in Malaysia. automotive industrial sector in Malaysia has been developing since Malaysia has its own national car industry and is seen as encouraging growth in the national economy to this day. Starting in 2015, the Malaysian government has imposed regulations to the automotive industry to perform activities 3Rs reduce, reuse and recycle. This has made the automotive supply chain to the supply chain is closed. Therefore, this study examined the main factors that make the automotive industry players to engage in the closed loop supply chain.

CLOSED LOOP SUPPLY CHAIN

Closed-loop supply chain is defined by Guide & Van Wassenhove (2009), as a supply chain management which can maximize the creation of value through the entire life cycle of a product and also the dynamics of the recovery of the species and the number of returns in a short period of time. The maximizing value creation process and also the dynamics of the recovery is obtained through the design, control and operation of the system.

The closed loop supply chain combines both forward and reverse supply chain where these green

practices able to minimize the industrial waste and also reduce the environmental impact. The practice of zero-waste which includes recycling, reuse, reduce, remanufacturing, refurbish and repair in closed loop supply chain has generate a process of cradle-to-cradle where it can generate values not only in terms of profitable values to the shareholder and also to its stakeholders. Shaharudin, Govindan, Zailani, & Tan (2015) stated that when the organization could identify the benefits of all types of product returns, then the combination of design for forward and reverse supply channels and an environmentally friendly way to dispose the product; it will then give the organization the ability to achieve sustainability in business.

Forward Supply Chain, Reverse Supply Chain and Reverse Logistics

Traditionally, supply chain management is revolving on the forward supply chain. For the forward supply chain, it starts from the raw materials purchases from the supplier to the manufacturer, manufacturing and assembling operations, distribution centre, then to the distributors and retailers, and ultimately to consumers (Souza, 2013). However, the traditional supply chain ignored the elements of environmental conservation and end of life of the product is considered as waste. As for the closed loop supply, it will change or renew the end-of-life phase of a product where the old product will be refurbished or remanufactured. Closed-loop supply chain will help in reducing the amount of industrial waste disposal of non-biodegradable and hazardous material and also reducing the shortage of landfill capacity for industrial materials disposal. Recently, the automotive industry starts to adopt this green practice in order to minimize the environmental impact. Among the product returns received in closed loop supply chain is the commercial returns, repair or warranty returns, leasing, end-of-use returns and end-of-life returns.

The reverse supply chain is established to manage the return of products from the consumer to the manufacturer and also from the manufacturer to the supplier (Schenkel, Krikke, Caniëls, & der Laan, 2015). According to Günther, Kannegiesser, & Autenrieb (2014), the process of closing the loop in the supply chain or reverse supply chain, it is including the important context of sustainable supply chain management; the reuse, refurbishment or recycling of goods. By closing the loop, it involves the final stages of life-cycle of a product where the product will go through processes of recycling, remanufacturing, reclamation and also reverse logistics (Sarkis, 2012).

In the closed loop supply chain system, the reverse supply chain will collect the product return from customers and returned to the original manufacturer in order for improvement and also enhances by refurbishment, remanufacturing, repair or recycle; either certain parts or whole parts of the product. This provided benefits for reverse supply chain and has been an emphasis on the manufacturing industry to shift towards convergence of the strategy to increase the product life cycle, in which they not only provide good value for but as well as integrated supply chain management strategies (Genovese, Acquaye, Figueroa, & Koh, 2015).

Closed loop supply chain could focus on product return and repair or improvement of products either by component or the entire product. Reverse logistics plays a key role in the implementation of closed-loop supply chain. The process of reverse logistics includes planning, implementing, controlling efficiency, cost effective flow of raw materials, the process of purchasing or inventory, finished goods as well as information related to the value or disposal of waste properly (Aravendan & Panneerselvam, 2014). The reverse logistics that have been effectively implemented could ensure the success of the implementation of closed-loop supply chain, in which they could reduce dumping in the inventory and distribution costs could be salvaged. The closed loop supply chain can

also increase the level of customer satisfaction as it recapture the economic value of the returned product, where the process of improvements, re-cycle, refurbishment or remanufacturing able to produce products at a lower cost (C. K. M. Lee & Lam, 2012). By adopting automotive closed loop supply chain management, the firms managed to reduce manufacturing costs of parts, reduce user costs, reduce carbon footprint released, reducing the rate of disposal of goods and pollution.

DRIVERS OF CLOSED LOOP SUPPLY CHAIN IN AUTOMOTIVE INDUSTRY

The closed loop supply chain management for Malaysian automotive industry requires drivers that could encourage more manufacturers and automotive suppliers to participate to adopt green supply chain and to ensure the success implementation of the closed loop supply chain so that it could address the problems associated with economic, environmental preservation and social development. The identified drivers closed loop supply chain could help in reducing the environmental impact and concerned on social benefits are listed below.

Customer Awareness

Customers nowadays conscious on the product benefits, services, value added products product effectiveness, cost and also conscious on environmental. Customers that have high awareness of the environmental issues will only buy products manufactured by firms that do green practices. High consumer awareness on the importance of protecting the environment could increase the level of participation of manufacturers and suppliers to do closed loop supply chain.

The closed-loop supply chain also covers business-to-customer and business-to-business customers such as the automotive manufacturer

and suppliers. Customers nowadays are not only focus on the price and quality of a product, but also sensitive to the choice of the products produced by manufacturers who are ethical and concerned on environmental issues. Therefore, if the manufacturer has a high awareness of environmental issues, the manufacturer will put a condition on the supplier to perform closed loop supply chain where product return undergoes recycled or re-manufacturing; so that the product is not directly disposed. This makes suppliers facing off with competition and also aware of the importance to practice green supply chain management. Awareness about sustainability is growing among consumers have been demanding automotive OEMs to do closed loop supply chain by recalling the defect products from their suppliers for new cycle or life of a product. (Go et al., 2011).

Akdoğan & Coşkun (2012), explains that in compliance with the law and also the creation of a responsible image in taking care of the environment, are the motivation behind the implication of closed loop supply chain. By practicing closed loop, it makes the firms has compliance with the law and this will make customer realize that the firm is really taking care on environment. This also could raise the level of awareness of the customer to return the product and get better options. Green supply chain practices will guarantee better quality service and also seen as a positive image of the company in addressing environmental issues.

The increase awareness on eco-friendly manufacturing solutions has made manufacturers including suppliers to establish green initiatives that are environmentally friendly. The industry players are urged to take a more green operation and green strategies in order to minimize negative impact on the environment (Xu et al., 2013). Evangelista (2014), stated that when customers have a high awareness of environmental issues, then it will help the organization to do green practices in achieving sustainability. This shows that high awareness of environmental preservation will increase the number of suppliers and

manufacturers to implement closed loop supply chain to do more recycling, remanufacturing or product improvements.

For manufacturers who have a high level of awareness about preserving the environment, they may educate their suppliers in matters related to the activities of environmental management. This will make the manufacturers and suppliers to have mutual understanding of the importance to preserve the environment and the wellness of society (Tarig K. Eltayeb, Zailani, & Ramayah, 2011). Through education on sustainability, suppliers will be aware of the importance of carrying out the green supply chain as an effort to minimize the impact on the environment. It could assess the level of awareness of the supplier of the products they produce, whether the product conducted in an environmentally manner and how goods will be dispose after the lifecycle reaches end. Suppliers can also understand the needs of customers who want quality products without adversely affect to the environment.

Governmental Legislation

As environmental issues have been a concern across the country, Malaysia also emphasized this issue mainly in terms of environmental protection and public wellness. In 1974, the environmental act has been established and this act has been amended several times include 18 sets of regulations related to clean air, waste and industrial assessment. The national environmental department stated that the manufacturing industry is a major contributor to pollution and destruction of natural resources. Therefore, it is important for Malaysia to achieve a balance between living standards and environmental protection (Zailani et al., 2015).

The awareness among stakeholders on the issues related to preservation of the environment has led to the establishments of environmental legislation and regulations in order to plan and design approaches that are appropriate to balance the development and conservation (Mondéjar-

Jiménez et al., 2015). Due to the increase in awareness of the importance of protecting the environment and also the implementation of the regulation by the government, then this has created conditions in which the parties involved in the industry and policy makers to look at the impact of supply chain activities on the environment (Fareeduddin et al., 2015).

Law plays an important role for the implementation of closed-loop supply chain. Starting in 2012, the government of Malaysia has ordered all party manufacturers and automotive suppliers to adopt a closed loop supply chain namely with 3Rs: Reduce, Reuse and Recycle practices in order to reduce the impact on the environment and also to reduce dumping in landfills. By imposing laws, acts and regulations relating to environmental preservation, then the parties involved in industries such as supplier and manufacturers could comply with regulations and this will force the parties to take green initiatives to mitigate the impact on the environment. The authorities will monitor and imposed penalty if there are any of those who violate the rules or performing their businesses unethically.

According to Eltayeb et al., (2011), in response to comply with the rules imposed by the government, including regulatory institutions, organizations are expected to take the approach through the purchase of green as additional adherence to the rules that have been imposed on them. In addition, the firm will also try to avoid potential excessive cost, any uncertainty and legal liability relating to environmental regulations. The regulatory institutions will monitor the activities of the operations carried out by the firm to ensure that every operation carried out does not affect the environment. Incentives are also offered to firms that comply with regulations. Even though in Malaysia there are no specific implementation rules relating to the nature of the industry which requires reverse logistics activities, but the government has begun to steer the organization to adopt green practices implemented by the automotive industry in western

countries. The government has offered incentives such as rebates and tax exemption to the firms that volunteered to do green practices.

Environmental Concern

The raising responsibility in preserving the environment has led companies began to realize the importance of performing green supply chain management. To ensure that sustainability can be achieved for a long time, the upstream and downstream is an important factor in playing a role in improving supply chain performance while satisfying customer needs (Ageron, Gunasekaran, & Spalanzani, 2012). This can be done by linking the channel of suppliers in the closed loop supply chain to collaborate in order to make the sustainable initiative in supply chain.

There are seven environmental requirements and economic aspects of sustainability practices stated by Diabat, Kannan, & Mathiyazhagan, (2014). The requirements are consisting of reduction in waste and emissions; reduction in energy intensity of goods and services; use of renewable and sustainable energy resources; maximum use and re-use of recycled components and materials; measurement and assessment of business impact on ecosystems; standard measures for evaluating sustainability performance; and environmental consciousness pervading the culture of an organization.

To address issues related to the environment and returns a profit, the firm needs a business model that is comprehensive and viable to develop strategies that effectively green supply chain (Das & Rao Posinasetti, 2015). Organizations that have a high awareness in environmental management could affect the participation of the supplier's also the success of green initiatives in the supply chain. Firms can make green initiatives as a corporate program and innovation from suppliers to ensure increased participation in achieving sustainability (Caniëls, Gehrsitz, & Semeijn, 2013). Therefore, many firms have realized the importance of working collaboratively with partner suppliers to

produce products that are environmentally friendly and can be recycled where it would improve their sustainability performance. In addition, the firm also create awareness, educate and convince consumers (Gopal & Thakkar, 2015).

The good and functioning conditions parts such as engines, transmissions, starter motors, alternators, water pumps, doors and bumpers could be reuse on other vehicles of the same model. Materials from batteries, catalytic converters, tyres and plastics are extracted and recycled into new products. Fluids such as engine oil, coolant, and gasoline are carefully collected and stored in double-walled tanks or secondary containment prior to being reused or recycled. Therefore, if the end of the life cycle of these components is only allowed for the disposal, then it will increase the rate of pollution. By perform closed-loop supply chain, the end life cycle of these components can converted into another new life cycle product.

Social Responsibilities

Traditional supply chain is more subjected to the economic perspective. However, the closed loop chain has revolved around the impact of industry on the environment and also social wellness. Social responsibility, ethical funds and eco-efficiency has shown that ongoing efforts to achieve sustainable social development (Diabat et al., 2014). The practice of social responsibility is regarded as important in which it is an effective solution to achieve such sustainability could help to deal with problems related to social development and environmental issues (Habidin et al., 2015).

The aspect of the closed loop supply chain is not only subject to cost, quality and financial performance. Issues of social development and preservation of the environment has attracted many parties such as customers, investors, regulators, government bodies and non-governmental organizations and the public to co-evaluate the importance of green supply chain management or closed loop supply chain. However, social issues are often neglected. Environmental issues such as

carbon emissions are more highly regarded while social impacts is ignored (Klassen & Vereecke, 2012). However, social issues, health and welfare of local residents should also take precedence over environmental issues. This is because by being concern on social issue, then the firm has provided assurance that the operations or activities carried out by them does not affect the welfare and health of future generations. This also educate the public that the current generation and future generations is equally important to achieve sustainability.

Waste Management

The traditional open supply chain focuses on player chain suppliers, manufacturers, distributors, retailers and end users. Usually the products will be disposed after its end of life. This increases the number of waste for disposal and reduce the value of the product. In the closed-loop supply chain, the product that has come to the end of life is could be converted from waste into other product. Therefore, it makes the cradle-to-grave business has turned to a cradle-to-cradle opportunity and makes the experience of the old product that undergoes improvement process could be reused.

The growth and industrial development in Malaysia has resulted in an increase in the production of industrial waste. The environmental management practices are very important aspect in a variety of industrial applications in all countries around the world. Efforts to reduce the disposal of waste in the manufacturing industry will improve environmental management. For the automotive industry, most automotive manufacturers have implemented environmental management to improve the operating performance of the company (Fadly Habidin et al., 2015).

Tarig Khidir Eltayeb & Zailani (2011), stated that a lot of efforts and expenses have been undertaken by the authorities in the management of solid waste disposal. The raising awareness of protecting the environment in many countries

have led the efforts to recycle and convert end of life products to new products. In Malaysia, the approach to end-of-life of a product will mostly end up in landfills or through the combustion process. This will result in damage to the land drainage and also carbon and other harmful gas emissions. Through the implementation of National Automotive Policy in 2014, all the suppliers and manufacturers are instructed to take green initiatives such as 3R. Therefore, the implementation of the closed-loop supply chain activities is able to reduce waste from industry.

Therefore, by performing closed-loop supply chain so it could help reduce the problems associated with landfill and also reduce the level of carbon emissions. Returned product will undergo a process of recycling, remanufacturing or improvements in production which could be produce at a lower cost. This would bring benefits for firms and consumers. Thus, the automotive industry in Malaysia could do refurbishment or repair of car components, engines, internal and external parts or electronic parts from product returns and it could be sold to the consumers at lower price. This green initiative will reduce the dumping of car components that have reached the final life cycle stages in the landfill.

FUTURE RESEARCH DIRECTIONS

There are many areas of closed-loop supply chain in the automotive industry to be explore. The closed loop supply chain for automotive industry in Malaysia is an emerging trend and have a promising future. For research related to closed loop supply chain in the automotive, it is suggested that researchers to study the effectiveness of drivers in influencing suppliers and manufacturers to perform closed-loop supply chain. With this, the performance of closed-loop supply chain could be evaluated.



CONCLUSION

This study focuses on the management of closed-loop supply chain and identify the drivers that can drive closed-loop the supply chain management for the Malaysian automotive industry. The implementation of a closed the supply chain management for the automotive industry in Malaysia is expected to help the country and also the automotive sector to achieve sustainability. In addition, it also could reduce the environmental issues that related to the automotive supply chain. The consumer awareness is essential for carrying out the closed loop supply chain, where it affects the market demand. The implementation of laws or regulations will ensure that the suppliers as well as manufacturers to comply with the rules relating to the use and disposal of products is carried out according to proper procedures, to ensure the well-being of nature and community is maintained. The organizations that are aware of the importance of protecting the environment shows that the organization is no longer focused on profit alone, even green initiatives and practices could reduce the environmental impact. Through environmental awareness, the organization will be more socially responsible, where the global community health and welfare are not jeopardized. While, waste management is the organization which manages the industrial waste in a way that is more environmentally friendly. Drivers described in this study were identified to be the catalyst for the automotive industry in Malaysia in promoting the growth of the automotive closed loop supply chain, contributing to excellence in the automotive supply chain as well as the achievement for a better sustainability.

REFERENCES

- Ageron, B., Gunasekaran, A., & Spalanzani, A. (2012). Sustainable supply management: An empirical study. *International Journal of Production Economics*, 140(1), 168–182. doi:10.1016/j.ijpe.2011.04.007
- Akdoğan, M. Ş., & Coşkun, A. (2012). Drivers of Reverse Logistics Activities: An Empirical Investigation. *Procedia: Social and Behavioral Sciences*, 58, 1640–1649. doi:10.1016/j.sbspro.2012.09.1130
- Aravendan, M., & Panneerselvam, R. (2014). Literature Review on Network Design Problems in Closed Loop and Reverse Supply Chains. *Intelligent Information Management*, 06(03), 104–117. doi:10.4236/iim.2014.63012
- Caniëls, M. C. J., Gehrsitz, M. H., & Semeijn, J. (2013). Participation of suppliers in greening supply chains: An empirical analysis of German automotive suppliers. *Journal of Purchasing and Supply Management*, 19(3), 134–143. doi:10.1016/j.pursup.2013.02.005
- Das, K., & Rao Posinasetti, N. (2015). Addressing environmental concerns in closed loop supply chain design and planning. *International Journal of Production Economics*, 163, 34–47. doi:10.1016/j.ijpe.2015.02.012
- Diabat, A., Kannan, D., & Mathiyazhagan, K. (2014). Analysis of enablers for implementation of sustainable supply chain management – A textile case. *Journal of Cleaner Production*, 83, 391–403. doi:10.1016/j.jclepro.2014.06.081
- Eltayeb, T. K., Zailani, S., & Ramayah, T. (2011). Green supply chain initiatives among certified companies in Malaysia and environmental sustainability: Investigating the outcomes. *Resources, Conservation and Recycling*, 55(5), 495–506. doi:10.1016/j.resconrec.2010.09.003
- Eltayeb, T. K., & Zailani, S. H. M. (2011). Drivers on the reverse logistics: Evidence from Malaysian certified companies. *International Journal of Logistics Systems and Management*, 10(4), 375. doi:10.1504/IJLSM.2011.043101
- Evangelista, P. (2014). Environmental sustainability practices in the transport and logistics service industry: An exploratory case study investigation. *Research in Transportation Business & Management*, 12, 63–72. doi:10.1016/j.rtbm.2014.10.002



- Fadly Habidin, N., Fadzlin Mohd Zubir, A., Mohd Fuzi, N., Azrin MdLatip, N., & Nor Azhari Azman, M. (2015). Sustainable Performance Measures for Malaysian Automotive Industry. *World Applied Sciences Journal*, 33(6), 1017–1024. doi:10.5829/idosi.wasj.2015.33.06.257
- Fareeduddin, M., Hassan, A., Syed, M. N., & Selim, S. Z. (2015). The Impact of Carbon Policies on Closed-loop Supply Chain Network Design. *Procedia CIRP*, 26, 335–340. doi:10.1016/j.procir.2014.07.042
- Genovese, A., Acquaye, A. a., Figueroa, A., & Koh, S. C. L. (2015). Sustainable supply chain management and the transition towards a circular economy: Evidence and some applications. *Omega*, 1–14. doi:10.1016/j.omega.2015.05.015
- Go, T. F., Wahab, D., Rahman, M. N. A., Ramli, R., & Azhari, C. H. (2011). Disassemblability of end-of-life vehicle: A critical review of evaluation methods. *Journal of Cleaner Production*, 19(13), 1536–1546. doi:10.1016/j.jclepro.2011.05.003
- Gopal, P. R. C., & Thakkar, J. (2015). Sustainable supply chain practices: An empirical investigation on Indian automobile industry. *Production Planning and Control*, 7287(January), 1–16. doi:10.1080/09537287.2015.1060368
- Govindan, K., Soleimani, H., & Kannan, D. (2014). Reverse logistics and closed-loop supply chain: A comprehensive review to explore the future. *European Journal of Operational Research*, 240(3), 603–626. doi:10.1016/j.ejor.2014.07.012
- Guide, V. D. R. Jr, & Van Wassenhove, L. N. (2009). OR FORUM The Evolution of Closed-Loop Supply Chain Research. *Operations Research*, 57(1), 10–18. doi:10.1287/opre.1080.0628
- Guide, V. D. R. J., Harrison, T. P., & Van Wassenhove, L. N. (2003). The Challenge of Closed-Loop Supply Chains. *Interfaces*, 33(6), 3–6. doi:10.1287/inte.33.6.3.25182
- Günther, H.-O., Kannegiesser, M., & Autenrieb, N. (2014). The role of electric vehicles for supply chain sustainability in the automotive industry. *Journal of Cleaner Production*, 90, 220–233. doi:10.1016/j.jclepro.2014.11.058
- Klassen, R. D., & Vereecke, A. (2012). Social issues in supply chains: Capabilities link responsibility, risk (opportunity), and performance. *International Journal of Production Economics*, 140(1), 103–115. doi:10.1016/j.ijpe.2012.01.021
- Lee, C. K. M., & Lam, J. S. L. (2012). Managing reverse logistics to enhance sustainability of industrial marketing. *Industrial Marketing Management*, 41(4), 589–598. doi:10.1016/j.indmarman.2012.04.006
- Lee, S. M., Tae Kim, S., Choi, D., Kim, S. T., & Choi, D. (2012). Green supply chain management and organizational performance. *Industrial Management & Data Systems*, 112(8), 1148–1180. doi:10.1108/02635571211264609
- Maiti, T., & Giri, B. C. (2014). A closed loop supply chain under retail price and product quality dependent demand. *Journal of Manufacturing Systems*. doi:10.1016/j.jmsy.2014.09.009
- Mondéjar-Jiménez, J., Segarra-Oña, M., Peiró-Signes, Á., Payá-Martínez, A. M., & Sáez-Martínez, F. J. (2015). Segmentation of the Spanish automotive industry with respect to the environmental orientation of firms: Towards an ad-hoc vertical policy to promote eco-innovation. *Journal of Cleaner Production*, 86, 238–244. doi:10.1016/j.jclepro.2014.08.034
- Paksoy, T., Bektaş, T., & Özceylan, E. (2011). Operational and environmental performance measures in a multi-product closed-loop supply chain. *Transportation Research Part E, Logistics and Transportation Review*, 47(4), 532–546. doi:10.1016/j.tre.2010.12.001

Sarkis, J. (2012). A boundaries and flows perspective of green supply chain management. *Supply Chain Management: An International Journal*, 17(2), 202–216. doi:10.1108/13598541211212924

Schenkel, M., Krikke, H., Caniëls, M. C. J., & der Laan, E. (2015). Creating integral value for stakeholders in closed loop supply chains. *Journal of Purchasing and Supply Management*, 21(3), 155–166. doi:10.1016/j.pursup.2015.04.003

Shaharudin, M. R., Govindan, K., Zailani, S., & Tan, K. C. (2015). Managing product returns to achieve supply chain sustainability: An exploratory study and research propositions. *Journal of Cleaner Production*, 101, 1–15. doi:10.1016/j.jclepro.2015.03.074

Souza, G. (2013). Closed-loop supply chains: A critical review, and future research*. *Decision Sciences*, 44(1), 7–38. doi:10.1111/j.1540-5915.2012.00394.x

WCED. (1987a). Our Common Future (The Brundtland Report). *Medicine and War* (Vol. 4). <http://doi.org/10.1080/07488008808408783>

WCED. (1987b). Report of the World Commission on Environment and Development: Our Common Future (The Brundtland Report). *Medicine, Conflict, and Survival*, 4.

Xu, L., Mathiyazhagan, K., Govindan, K., Noorul Haq, A., Ramachandran, N. V., & Ashokkumar, A. (2013). Multiple comparative studies of Green Supply Chain Management: Pressures analysis. *Resources, Conservation and Recycling*, 78, 26–35. doi:10.1016/j.resconrec.2013.05.005

Zailani, S., Shaharudin, M. R., Govindasamy, V., Ismail, M., Fatin, S., & Shaikh, A. (2015). *The Eco-Efficiency Practices of the Sustainable Packaging and its Effect towards Sustainable Supply Chain Performance*. Academic Press.

Zhu, Q., Sarkis, J., & Geng, Y. (2005). Green supply chain management in China: Pressures, practices and performance. *International Journal of Operations & Production Management*, 25(5), 449–468. doi:10.1108/01443570510593148

Zhu, Q., Sarkis, J., & Lai, K. (2008). Green supply chain management implications for closing the loop.. *Transportation Research Part E, Logistics and Transportation Review*, 44(1), 1–18. doi:10.1016/j.tre.2006.06.003

KEY TERMS AND DEFINITIONS

Automotive Original Equipment Manufacturers: Automotive original equipment manufacturers (OEMs) assemble single components to final vehicles that are sold via a dealer-network to business and private customers.

Closed Loop Supply Chain Management: Closed loop supply chain management is the design, control and operation of a system to maximize value creation over the entire life cycle of a product with dynamic recovery of value from different types and volumes of returns over time (V. D. R. J. Guide, Harrison, & Wassenhove, 2003).

Green Supply Chain Management: Green supply chain management or GSCM has emerged as an important new archetype for enterprises to achieve profit and market share objectives by lowering their environmental risks and impacts and while raising their ecological efficiency (Qinghua Zhu, Sarkis, & Lai, 2008).

Reverse Logistics: Reverse logistics are the process of planning, implementing, and controlling the efficient, cost effective flow of raw materials, in-process inventory, finished goods and related information from the point of consumption to the point of origin for the purpose of recapturing value or proper disposal (Govindan, Soleimani, & Kannan, 2014).

Supply Chain Activities: Supply chain activities is a set of activities that includes purchasing, manufacturing, logistics, distribution, marketing, that perform the function of delivering value to end customer (Paksoy, Bektaş, & Özceylan, 2011).

Supply Chain Management: Supply chain management is the integration of key business processes from end users through suppliers that provides product, services and information that add value for customers and stakeholders (S. M. Lee, Tae Kim, Choi, Kim, & Choi, 2012).

Sustainable Development: Sustainable development is the development that meets the

needs the present without compromising ability of future generations to meet their own needs (WCED, 1987b).

Value Creation: The creation of value is the core purpose and central process of economic exchange. Traditional models of value creation focus on the firm's output and price. Value is created by the firm in the form of a good, and this valuable good is exchanged in the marketplace for money (or possibly other goods). Value creation refers to traditional supply chain objectives, customer satisfaction and cost reduction, as well as environmental goals (Schenkel et al., 2015).



From Business-to-Business Software Startup to SAP's Acquisition

John Wang

Montclair State University, USA

Jeffrey Hsu

Fairleigh Dickinson University, USA

Sylvain Jaume

Saint Peter's University, USA

INTRODUCTION

Nearly one trillion in commerce was contributed by more than two million buyers and sellers around the world last year. Organizations tapped into the comprehensive source-to-pay solutions and business network offered by SAP Ariba. They digitized their supply chains and efficiently managed their trading relationships (SAP Ariba, 2016).

Ariba, Inc. was born during the dot-com bubble, at that time a start-up with great potential along with countless other e-commerce companies. The shake-out that occurred later on among a plethora of competing businesses, confirmed the solidity of Ariba's focus and business model while competitors like Commerce One, could not stay in business. However, the interesting evolution of Ariba, from a pioneer to a sufferer to a survivor, has taught us much about survival in the competitive business to business (B2B) software industry. Ariba's software would help many companies save money on their procurements, and control expenses in numerous areas aside from payroll. Ariba promised to help companies improve their bottom line, and many of Ariba's clients today hold positions on the coveted Fortune 100 list. Going public in 1999, Ariba's stock price at one time reached \$259 per share. At the time, Ariba was still getting their feet wet; they had not yet made a profit. In 2000 their stock reached \$168.75, but a negative turn in the economy lay just ahead. Many companies

began to cut back on investments and in just 9 months Ariba lost 95% of its value. This would be disastrous to any company, and would be the downfall of most, but Ariba made some critical key decisions that helped continue its leadership in the B2B e-commerce world (Clarke & Pucihar, 2013; Schumann et al., 2014). They are one of the few companies to not only survive the burst of the dot-com bubble, and but to thrive, and to this day remains a successful company. Ariba was purchased by the German software firm SAP for \$4.3 billion in 2012 (Jones, 2012).

BACKGROUND

E-commerce was in many ways revolutionized by Ariba Inc., a leading independent company in the sphere of B2B commerce network providers. Since the time of its founding, the company has been evolving constantly, in cooperation with leading companies in the industry, in order to deliver e-commerce platform products to its customers/clients. Its value chain model was successful in helping it to develop business relationships further than expected, the results of which made it a top 40 Fortune 500 company.

Ariba has overcome many obstacles, including lawsuits, changing customer requirements, and organizational restructuring, however still managed to remain a leader in its specific niche

DOI: 10.4018/978-1-5225-2255-3.ch468

industry area. The firm has done so by delivering solutions and services that meet customers' expectations, and have been able to cope with intense competition by keeping up with today's technologies as well as developing solutions for tomorrow. Ariba was founded in Sunnyvale CA, in September 1996, by seven men, one of the most influential being Steven Krach. Krach's early career accomplishments included being one of the youngest vice presidents of General Motors (Ariba, 2008). The challenges of procurement that he encountered there became the impetus for the development of Ariba. Through the use of the Internet and B2B e-commerce, Krach and his associates brainstormed and came up with the idea of automating the purchasing of commonly used supplies and services. While this may seem to be a rather simple concept, it turned out to be an insight and area that held huge demand and potential.

After three months of intensive research, which included meeting with 60 Fortune 500 companies, Ariba had a prototype developed and ready for their initial marketing campaign. Having signed software licensing deals with Cisco Systems, Advanced Micro Devices and Octel Communications, prior to software completion, the pieces were put into place for the launch of their product. While Krach and Ariba were innovative in their concepts and efforts, they were not the only pioneers in the area, Commerce One, Oracle, I2, and PeopleSoft, Inc. were competitors with similar, but not necessarily identical, offerings.

The objective was to become a leader in this niche industry with the means and resources to provide procurement software and network consulting services, which would allow corporations to manage and automate their spending more effectively and efficiently. The management of these would include essentially all non-payroll expenses associated with running a business. Since access to data was one critical component, Ariba offered their clients real-time access over the Internet. These applications were used in conjunction with the *Ariba Supplier Network*

to purchase goods and services. Ariba is a firm that is customer driven, and as such offered full support, including technical support, implementation, training, and consulting programs and packages. E-payment and service agreements were made with American Express and Bank of America. All of these were considered large and bold undertakings for a young startup company at that time. In June of 1999, Ariba went public at a modest \$23 per share, however traded as high as \$259 per share at times later that year (Schneider and Bruton, 2008; Haksoz and Seshadri, 2007). This was a phenomenal success for a three year old company which had yet to turn a profit at that time, since it benefited from being a "first mover" in the business. However, as mentioned earlier, other Internet start-up companies were beginning to offer similar software and services. Over time, smaller companies began offering websites that provided a place to manage procurement, some with lower costs and fees. Facing challenges in the market, Ariba was to be faced with difficult challenges and had to make major decisions in order to stay in business.

Ariba finally saw a profit of \$10 million in December of 2000, which also included the completion of a few *acquisitions*. Soon after, in 2001, the economy began to weaken in a downward spiral and Ariba's stock plummeted 95%, making a business overhaul necessary. Ariba decided to take drastic cost-cutting measures, cutting about a third of their staff. Because of their specialized and niche product line, their business was able to continue and survive the setbacks faced by other Internet software companies. Krach resigned as CEO in 2001, but stayed on as Chairman and appointed a CEO that would later cost the company much money and negative publicity. The bursting of the dot-com "bubble" marked the beginning of a relatively mild yet rather lengthy early 2000s recession (Marshall, 2001; Sahay, 2007). In time, Ariba, along with the rest of the B2B business community ran into two big problems. First, the brick-and-mortar Old Economy was stable and could adjust more readily to economic

downturns. Second, companies were interested in saving transaction fees by using alternate means instead of using costly B2B networks (Cerquides, López-Sánchez, Reyes-Moro, & Rodríguez-Aguilar, 2007).

Overtime, however, Ariba persisted and would once again regain its position as a leader in the B2B procurement industry. The firm made adjustments where necessary to continue to deliver the goods to their clients, without sacrificing their own bottom line. According to the current CEO, Bob Calderoni, Ariba is well positioned in the spend management market and will continue to grow in the current tough global economy.

The following sections explore the internal (*adaptations to a competitive environment, acquire to advance, consulting adds value, emphasis on the customer*) and external (*severe competition, high-priced software, regulator's investigation, unhappy customers*) factors that affect the company's struggles and challenges.

INTERNAL FACTORS

Adaptations to a Competitive Environment

At the height of the e-procurement frenzy, two companies dominated the B2B space: Commerce One and Ariba. Due to the near-collapse of the original B2B procurement model, both companies sought new niches. Commerce One moved towards web services in an attempt to seek viable markets. Ariba, meanwhile, emphasized enterprise spending management (Kinsey, 2004). Ariba strongly believed that a software firm's role first and foremost is to be a software tool provider. As the B2B world divided into industry sponsored exchanges and independent marketplaces, Ariba avoided making inroads into involvement in managing its customers' exchanges. In contrast to this, Commerce One believed that software makers had to do more than simply provide software tools. They believe the right approach was to form strategic

partnerships with its customers and to help manage their online marketplaces (Anderson, Opie, & Watton, 2003; Bannan, 2008). It also directed its customers towards an international trading network in order to build critical mass and facilitate e-commerce between them.

As a new CEO, Calderoni monitored the external environment, where a fundamental shift in the marketplace existed, and responded promptly to adjust the company's product offering. He believed that B2B e-business had a direct and indirect impact on all functional areas, and that these linkages with a company's supply chain system were critical. In effect, Ariba was changing its focus from e-procurement, to offering products that can increase customer satisfaction by solving a variety of "spend-related" problems faced by corporations (Tadeschi, 2008).

With the concept of division of labor as a microeconomic view, Calderoni added a purchasing system, general ledger, and field system into Ariba's line of products. The added features in the company's products were favored by Ariba's existing customers in the automotive, chemical, and manufacturing industries due to the ease of system comparability. The need to transfer data from legacy systems enabled these customers to remove outdated and inaccurate data from their systems and which also helped to improve relationships with their customers.

Acquire to Advance

According to Krach, a major component of Ariba's business model is partnering followed by organic growth and acquisition, and so the firm continued to follow this basic approach to help ensure the firm's success. Ariba acquired companies that had the technology and resources they needed to survive, instead of taking the time to develop them in-house. Mr. Calderoni implemented an aggressive acquisition strategy that significantly expanded Ariba's technology offerings and service capabilities, and positioned the company as a recognized leader in its market. One goal was

to secure top Fortune 10 companies and Global 500 companies as customers (Efendi et al., 2013).

In 2004, Ariba acquired Alliente Inc. and FreeMarkets Inc. to link their spend management software with its existing capacities as a B2B procurement hub. The acquisition of FreeMarkets increased Ariba's offerings by providing global supply management software and services. This acquisition also positioned Ariba as a serious contender in the automotive industry, adding General Motors Corp., Daimler-Chrysler AG and Ford Motor Co. to their customer base. By acquiring Alliente, Ariba expanded its spend management and procurement capabilities to include a procurement out-sourcing provider (Hosford, 2007).

In December 2007, Ariba announced that it had completed the acquisition of Procuri, Inc. a privately held provider of on-demand supply management solutions, rounding out Ariba's offerings that help companies automate the procurement process. According to Ariba CEO Bob Calderoni, more than 70% of Procuri's 300 customers have under \$5 billion in revenues. As a result, this deal also gave Ariba greater access to midmarket customers (Anonymous, 2007). These strategic acquisitions helped Ariba to grow its line of products, as well as expand its customer base.

Consulting Adds Value

Calderoni believed that Ariba had the ability to not only survive, but thrive, by expanding beyond software that focused mainly on transactions, to encompassing and supporting additional facets of the buying process. Calderoni hired hundreds of consultants to advise companies on how to procure goods and services while incurring lower costs, using Ariba's software. Although consulting is generally viewed to be less profitable than selling software, Calderoni predicted he can successfully combine the two as an integrated set of offerings. While consultants coach Ariba's clients on how to use the software effectively, Ariba's clients can also interact and use Internet-based purchasing

systems to help them buy direct materials that are core to their company's manufacturing processes.

In order to extend Ariba's consulting services, which in 2004 accounted for nearly half of the company's \$323 million in sales, Ariba made consultants available via email and phone for a fraction of the price that is charged to the larger companies who require dedicated consultants through in site consultation visits.

Emphasis on the Customer

In 2001, investors were looking for a change in leadership at Ariba after the firm missed revenue and earnings projections by a wide margin. Ariba moved Keith Krach out of the CEO position, filling the post with the company's President and COO, Larry Mueller. Mueller entered the position with a new strategy: to halt the company's current plans to enter new markets, and instead opting to add new features, including electronic payment and invoicing, to its existing e-procurement and auction applications and offerings. Mueller heightened the focus on improving e-procurement applications by making heavy investments in existing e-procurement and sourcing platforms; and building technology around the key interactions that enterprises have with trading partners.

Mueller remained focused on bolstering Ariba's role as a traditional B2B transaction platform. Ariba announced plans to invest heavily in its *Ariba Commerce Services Network* and its network-centric applications, including *Network Connect*, which allowed non-Ariba customers to use the Ariba services network and conduct business or procure products and services. The company also focused its development, sales, and marketing staff to focus on specific industries. According to Mueller, "Customer ROI is the focus." A focus on international expansion boosted revenue from outside the US from 10% in the third quarter 2000 to 25% in the same period for 2001 (Purdum, 2007). Ariba was trying to rebuild its fortunes while the public marketplaces that used its technologies were struggling – an indication to

L

some economists that the industry may not have been ready for e-commerce.

Since joining the company in 2000 as chairman and CEO of Ariba, Mr. Calderoni has successfully transformed Ariba from a narrowly focused e-procurement vendor to a comprehensive spend management solutions provider that companies of all sizes rely on to transform the way they do business globally. Also, Ariba announced the availability of Chinese and Korean translations. Under Calderoni's leadership, Ariba has led the way in developing and delivering innovative solutions that combine technology, commodity expertise, and services to help companies streamline the procurement process and drive bottom-line results.

EXTERNAL FACTORS

Severe Competition

Simply being a dot-com business survivor, however, would not ensure its continued existence and profitability, and Ariba was at risk of losing business to the likes of other competitors such as SAP and Oracle. SAP, a German enterprise resource planning software maker, joined this market and signed on with Hewlett-Packard for a product called mySAP.com e-business solutions. In addition, it built a marketplace for chemical and pharmaceutical firms by educating them on mySAP.com, and with the result of installing a large SAP user base among Fortune 500 companies.

Nevertheless, Oracle had already anticipated a shift in the market and made plans to capitalize on it. Right now, the procurement sector is dominated by leading software companies Ariba and Commerce One. But as the slowdown in the U.S economy continues, Oracle was hoping that the opportunity for companies like Ariba would start to shrink as users look to more established ones, like Oracle, for an all-around e-commerce package (Arora, Greenwald, Kannan, Karthik, & Krishnan, 2007). Ariba had in fact provided Oracle

with an opportunity to gain market share when it cut a third of its workforce and announced reduced earnings during economic recent downturns.

Ariba recognized that to remain competitive, it had to address the problem of hidden costs associated with the products they sell, in addition to the price they charged for the software itself within the supply chain, especially when the product was in the later stage of its cycle. When a company does not pay much attention to the hidden costs of new software implementation, it can creep up and well-intentioned efforts can be result in financial penalties (Angeles & Nath, 2007; Brown, 2008).

By August 2008, the market for supply chain management (SCM) software market had grown. Worldwide spending on SCM solutions reached \$6 billion in 2007, which was up 17.6% from 2006. SCM Technologies was well-positioned to address the economic realities facing worldwide markets where costs were skyrocketing while competition and customer demands were intensifying (Eschinger, 2008). A number of the SCM solution vendors are merging, and expanding their capabilities within the realm of supply chain technologies. In comparison, Ariba's 2007 revenue was \$160.3 million, which significantly trailed behind Oracle and SAP, who reported \$955.2 million and \$1,334.4 million in revenue, respectively, showing that the threat of these products cutting into Ariba's bottom line is a real one (Orme & Etzkorn, 2007).

High-Priced Software

Without a doubt, e-procurement is rising substantially among the nation's largest 500 companies. Well-financed corporations are willing to invest in Internet software and technology that can reduce the inefficiency associated with the purchasing and buying processes. The use of this software can help companies to track spending and make sure they purchase products in accordance with contracts they have negotiated with suppliers. In fact, businesses that spend billions each year on supplies can often save tens of millions in costs

by implementing such technology. However, it's only the large firms that can devote the time and money to installing such systems, which frequently required that suppliers link to such systems as well (Janita & Miranda, 2013). Since the software is generally expensive and can be complicated to install on the customer's system, for small- and medium-sized businesses facing an uncertain economy, investments of this magnitude are can be difficult to justify.

Ariba took advantage of this situation, and in 2005 announced a strategy to sell its software and services to smaller companies on an on-demand basis, so they can buy supplies more efficiently online, as well. Ariba reshaped its software system so its customers can plug into Ariba's software through the Internet instead of installing it on their own systems. One major benefit of this approach, Mr. Calderoni said, is that Ariba can sell software to procurement managers and others in charge of spending, without involving the company's information technology staff.

Regulator's Investigation

In early 2003, Ariba came under investigation by the Security and Exchange Commissions (SEC). The reason for the investigation was linked to Ariba's accounting errors, doubtful partner deals, and questionable e-payments items including chartered airplanes. Among the specific allegations were that Ariba failed to record a \$10 million payment from Chairman Keith Krach to former chief executive Larry Mueller as an expense. The restatements are unusual because the chairman -- not the company -- covered the expense (Lau and Wang, 2007). Then three weeks later, Ariba decided to do the same for \$1.2 million in chartered jet services that was considered as Krach's compensation to Mueller, who subsequently left Ariba in July 2001. The problem was that United States laws and regulations require that payments by a principal holder to executives be treated as expenses paid on behalf of the company.

In addition, Ariba reported an additional \$7.5 million in expenses. In 2000, Ariba acquired TradingDynamic Inc., Tradex Technologies Inc., and SupplierMarket.com, and it reclassified stock options, or goodwill, that it gave to employees of these companies as a compensation expense. So by combining all of these expenses, the results were 18.7 million of added expenses. Ariba was aware that the regulator had begun an informal inquiry into its accounting practices after the firm said it would restate its earnings for 10 quarters.

Unhappy Customers

There have been some bumps in the road, including some public relations debacles. Ariba was subject to bad publicity after sending out automated emails to mid-size suppliers announcing their accounts had been upgraded to Premier level status. The email listed premier supplier benefits as well. However, the email also informed them that as a Premier Member, they were now required to pay associated annual fees. Many of the small and mid-size companies viewed this as a marketing ploy and felt they should not be made to pay for the upgrade without prior consultation or the ability to decide whether an upgrade was desired (Eschinger, 2008).

FUTURE RESEARCH DIRECTIONS

Ariba has thrived on making strategic, yet controversial decisions throughout the years. Having gone through ups and downs during the years, they still managed to largely maintain a leadership position even though markets and needs have evolved and changed. Overall, they have been able to provide customers with superb services and innovative products, and reliability is what keeps old customers, and helps to bring in new customers.

According to Jin and Yu (2015), procurement auctions have gained significant momentum over



the last two decades. So it remains to be seen whether Ariba keep its promise of:

Using the Ariba Network - anytime, anywhere - you can connect with your trading partners to buy smarter, sell faster, and manage your cash more efficiently and effectively than ever before? (Ariba, 2014)

CONCLUSION

Ariba achieved recognition for its *Supplier Network*, an e-business center where millions of buyers and suppliers can electronically transact business online, with the goal of more efficient procurement. In addition, Ariba developed spend management software, and took this kind of network mainstream. Ariba's application-driven strategy of an open platform, hosted and implemented by partners with vertical-market domain expertise, has helped to secure them a dominant position in the B2B e-commerce application marketplace.

It is interesting to note that while many firms who came up around the time of Ariba and offered competing solutions, Ariba is one of the few that has survived and thrived, using continually new strategies to stay in business, providing a better customer experience, and utilizing advances in technology and ideas that no other companies dared to try. Ariba has been, and remains, a pioneer and a leader in an exceedingly competitive and changing marketplace. Due to increased globalization and deregulation, for a company to succeed, it needed to maintain leverage over the impact of competitive forces. Ariba has done a great job in setting themselves apart from their competitors by strategically aligning themselves with their partners, and also expanding its service offerings to a wider range of customers. Ariba's software offerings help companies focus on profitability, and together with its wide array of service options, are customizable for both larger and smaller companies. The firm's products are concrete and

customizable, depending on a client's needs. In addition, the product offerings are offered on a common platform, allowing information to travel accurately and quickly through the supply chain.

REFERENCES

- Anderson, J., Opie, G., & Watton, J. (2003). The corporate spend agenda preview. *Business Strategy Review*, 14(2), 5–7. doi:10.1111/1467-8616.00252
- Angeles, R., & Nath, R. (2007). Business-to-business e-procurement: Success factors and challenges to implementation. *Supply Chain Management*, 12(2), 104–115. doi:10.1108/13598540710737299
- Anonymous,. (2007, July). Moving forward by working backwards. *Preview Healthcare Purchasing News*, 31(7), 64–66.
- SAP Ariba. (2016). *SAP Ariba drives nearly one trillion in commerce*. bizwire.c66479468.
- Ariba. (2015). *The Ariba network*. Retrieved from <http://www.ariba.com/>
- Ariba, Inc. (2008). Preview contract management. *Resource Guide*, (Supplement), 62.
- Arora, A., Greenwald, K., Karthik, R., & Krishnan, R. (2007). Effects of information-revelation policies under market-structure uncertainty. *Management Science*, 53(8), 1234–1248. doi:10.1287/mnsc.1060.0688
- Bannan, K.J. (2008). Ariba creates controls for who sends marketing e-mails when. *B to B Chicago*, 93(10), 22.
- Brown, D. (2008). It is good to be green; environmentally friendly credentials are influencing business outsourcing decisions. *Strategic Outsourcing: An International Journal*, 1(1), 87–95. doi:10.1108/17538290810857501

Cerquides, J., López-Sánchez, M., Reyes-Moro, A., & Rodríguez-Aguilar, J. A. (2007). Enabling assisted strategic negotiations in actual-world procurement scenarios. *Electronic Commerce Research*, 7(3-4), 189–220. doi:10.1007/s10660-007-9007-4

Clarke, R., & Pucihar, A. (2013). Electronic interaction research 1988 - 2012 through the lens of the Bled eConference. *Electronic Markets*, 23(4), 271–283. doi:10.1007/s12525-013-0144-4

Efendi, J., & Kinney, M.R., & Smith, L.M. (2013). Marketing supply chain using b2b buy-side e-commerce systems: Does adoption impact financial performance? *Academy of Marketing Studies Journal*, 17(2), 57–83.

Eschinger, C. (2008, August). No recession in the supply chain arena. *Industry Week*, 257(8), 56.

Haksoz, C., & Seshadri, S. (2007). Supply chain operations in the presence of a spot market: A review with discussion. *The Journal of the Operational Research Society*, 58(11), 1412–1429. doi:10.1057/palgrave.jors.2602401

Hosford, C. (2007). Ariba pins turnaround hopes on keeping in touch with customers. *B to B Chicago*, 92(1), 18.

Janita, M. S., & Miranda, F. J. (2013). Exploring service quality dimensions in b2b e-marketplaces. *Journal of Electronic Commerce Research*, 14(4), 363–386.

Jin, M., & Yu, A. J. (2015). Procurement Auctions and supply chain performance. *International Journal of Production Economics*, 162, 192–200. doi:10.1016/j.ijpe.2015.01.013

Jones, S. D. (2012, May 22). *SAP to pay \$4.3 billion for cloud firm Ariba*. Retrieved from <http://www.wsj.com/articles/SB10001424052702304019404577420511203619228>

Kinsey, E. P. (2004, Fall). Where has AU the common sense gone? *Mid-American Journal of Business*, 19(2), 7–11.

Lau, K.-K., & Wang, Z. (2007). Software component models. *IEEE Transactions on Software Engineering*, 33(10), 709–724. doi:10.1109/TSE.2007.70726

Marshall, J. (2001, May). A foundation to build on. *Preview by Financial Executive*, 17(3), 7–13.

Orme, A. M., & Etzkorn, L. H. (2007). A parallel methodology for reduction of coupling in distributed business-to-business e-commerce transactions. *Journal of Electronic Commerce in Organizations*, 5(3), 52–67. doi:10.4018/jeco.2007070104

Purdum, T. (2007). Online marketplace evolution: In their infancy, many online marketplaces kept their focus too narrow. Those that saw the big picture survived. *Industry Week*, 356(1), 18–19.

Sahay, A. (2007, Summer). How dynamic pricing leads to higher profits. *MIT Sloan Management Review, Cambridge*, 48(4), 53.

Schneider, G. P., & Bruton, C. M. (2008). Using portal web sites in the purchasing function. *Allied Academies International Conference. Academy of Information and Management Sciences. Proceedings*, 12(2), 42–45.

Schumann, P., Scott, T., Kalinowski, J., Kaliski, J., & Pragman, C. (2014). Noble Systems Inc.: Inside the mind of an entrepreneur. *Journal of Business Case Studies*, 10(3), 263. doi:10.19030/jbcs.v10i3.8716

Tadeschi, B. (2008, November 14). A software strategy helps the little guy buy smarter. *The New York Times*, p. C9.

ADDITIONAL READING

Angeles, R., & Nath, R. (2007). Business-to-business e-procurement: Success factors and challenges to implementation. *Supply Chain Management*, 12(2), 104–115. doi:10.1108/13598540710737299



- Cerquides, J., López-Sánchez, M., Reyes-Moro, A., & Rodríguez-Aguilar, J. A. (2007). Enabling assisted strategic negotiations in actual-world procurement scenarios. *Electronic Commerce Research*, 7(3-4), 189–221. doi:10.1007/s10660-007-9007-4
- Dai, Q. Z., & Kauffman, R. (2006). To be or not to B2B: Evaluating managerial choices for e-procurement channel adoption. *Information Technology Management*, 7(2), 109–130. doi:10.1007/s10799-006-8103-9
- Degwekar, S., Lam, H., & Su, S. Y. W. (2007). Constraint-based brokering (CBB) for publishing and discovery of web services. *Electronic Commerce Research*, 7(1), 45–57. doi:10.1007/s10660-006-0062-z
- Elmaghraby, W. (2007). Auctions within E-Sourcing events. *Production and Operations Management*, 16(4), 409–422. doi:10.1111/j.1937-5956.2007.tb00269.x
- Haksoz, C., & Seshadri, S. (2007). Supply chain operations in the presence of a spot market: A review with discussion. *The Journal of the Operational Research Society*, 58(11), 1412–1429. doi:10.1057/palgrave.jors.2602401
- Hartley, J., Lane, M. D., & Duplaga, E. A. (2006). Exploring the barriers to the adoption of e-auctions for sourcing. *International Journal of Operations & Production Management*, 26(1/ 2), 202–221. doi:10.1108/01443570610641675
- Holmqvist, M., & Pessi, K. (2006). Agility through scenario development and continuous implementation: A global aftermarket logistics case. *European Journal of Information Systems*, 15(2), 146–158. doi:10.1057/palgrave.ejis.3000602
- Huang, G., Zhao, J. B., & Chen, X. (2007). Do It Yourself (DIY) Portales for developing e-business solutions for small and medium enterprises. *Journal of Manufacturing Technology Management*, 18(1), 72–89. doi:10.1108/17410380710717652
- Lau, K., & Wang, Z. (2007). Software component models. *IEEE Transactions on Software Engineering*, 33(10), 709–724. doi:10.1109/TSE.2007.70726
- Orme, A., & Etzkorn, L. (2007). A parallel methodology for reduction of coupling in distributed business-to-business e-commerce transactions. *Journal of Electronic Commerce in Organizations*, 5(3), 52–67. doi:10.4018/jeco.2007070104
- Raja, J., & Velmurgan, S. (2008). E-payments: Problems and prospects. *Journal of Internet Banking and Commerce*, 13(1), 1–18.
- Ratnasingam, P. (2007). A risk-control framework for e-marketplace participation: The findings of seven cases. *Information Management & Computer Security*, 15(2), 149–166. doi:10.1108/09685220710748029
- Rothkopf, M., & Whinston, A. (2007). On e-auctions for procurement operations. *Production and Operations Management*, 16(4), 404–408. doi:10.1111/j.1937-5956.2007.tb00268.x
- Schneider, G., & Bruton, C. (2008). Using portal web sites in the purchasing function allied. *Academies International Conference. Academy of Information and Management Sciences*, 12(2), 42–46.
- Sia, S., & Soh, C. (2007). An assessment of package-organisation misalignment: Institutional and ontological structures. *European Journal of Information Systems*, 16(5), 568–583. doi:10.1057/palgrave.ejis.3000700
- Smart, A. (2008). E-business and supply chain integration. *Journal of Enterprise Information Management*, 21(3), 227–246. doi:10.1108/17410390810866619

KEY TERMS AND DEFINITIONS

Acquisition: Acquiring control of a corporation, called a target, by a stock purchase or exchange, either hostile or friendly.

Business to Business (B2B): The exchange of products, services, or information between businesses.

E-Commerce (Electronic Commerce or EC): The buying and selling of goods and services over the Internet.

E-Payments (Online Payments, Electronic Payments, Internet Payments, Web Payments, and ePayment): An electronic payment made via

a web browser for goods and services using credit or debit cards or other related payment means.

E-Procurement: A system utilizing Internet technology to streamline the purchases of goods and products in order to reduce costs.

Supplier Network: A network of screened and qualified small-scale producers and committed medium-sized suppliers.

Supply Chain Management (SCM): The process of strategically managing flows of goods, services and knowledge, along with relationships within and among organizations, to realize greater economic value.

L

An Integrated Approach to Supply Chain Simulation

Nenad Stefanovic

University of Kragujevac, Serbia

Bozidar Radenkovic

University of Belgrade, Serbia

INTRODUCTION

Simulation can be a valuable tool for supply chain analysis, planning, optimization, evaluation, and risk management. Computer simulation and simulation models can be used to model intricate supply chains close to real systems, execute those models, and observe system behavior. The goal of simulation is to evaluate existing supply chain configurations, as well as to aid in design of the new supply chains. Supply chain simulation matters both supply chain design and supply chain control. In other words, it helps resolve different supply chain management (SCM) problems which can be grouped into the following categories:

- Infrastructure configuration that implies defining of the manufacturers, distribution centers, wholesalers, retailers and their locations (nodes).
- Planning and design of supply networks.
- Defining strategy related to processes at the nodes.
- Coordination between processes and activities with the purpose of their alignment and fulfillment of performance goals on global supply chain level.
- Information integration so that processes can exchange all necessary information.
- Risk mitigation in supply chain implementation both at the strategic and process levels.

- Supply chain validation through performance measurement which involves defining metrics at different supply chain levels.
- Validation through performance monitoring which involves defining metrics at different supply chain levels.
- Optimization by means of what-if analysis and application of best practices for improvement.

Modeling and simulation approach is the only practical option for exploring performance of the complex business networks such as supply chains. Furthermore, the modeling and simulation approach facilitates the design of the new supply chain configuration and policies, as well as the redesign of existing systems (Thierry et al., 2010). There are different supply chain modeling methods and types of simulation. Supply chain simulation can be performed with different goals. This can be analysis of supply chain dynamics, structure, or cash flow, risk mitigation, design, learning, etc.

In this chapter, the modelling approach and the simulation software which has integrative features capable to unify these various needs for different user groups are presented.

The background section gives definitions and explanations of key terms and concepts, as well as literature review with main contributions related to supply chain simulation. Simulation software integrates different models, entities and modules that function cooperatively. Its construction and

functioning principles are described and examples of supply chain simulations are given. The key contributions and benefits of the presented simulation solution are presented. Finally, the main future research directions and opportunities are examined.

BACKGROUND

By examining well designed simulation models, organizations can reinforce their decisions regarding supply chain processes. They can study and analyze effects of different supply chain initiatives and improvement programs through sensitivity analysis (such as what-if or goal seek) before investing huge amount of money or disrupting their operations.

Computer simulation and simulation models can be used to model intricate supply networks close to real systems, execute those models, and observe system behavior.

The main advantages of the supply network computer simulation are (Stefanovic et al., 2009):

- The simulation is relatively clear and flexible.
- It can be used for analysis of the complex real systems such as supply networks.
- With the simulation, it is possible to include real-world influences, for example uncertainty factor in demand or lead time.
- “Time compression” is possible. Effects of a certain business policy over a long period of time (months, years), can be obtained in a short time.
- The simulation enables “what-if” analysis. Managers can test the results of different decisions.
- The simulation does not interrupt real systems. For example, experimenting with different supply network configurations can be done without disruptions and significant investment.

- With the simulation, the effects of the individual components, parameters and variables can be studied at the global level.
 - The main disadvantages of the supply network computer simulation can be summarized as:
- Quality simulation models can be expensive and time consuming to develop and validate.
- This is “modify-try” approach. Typically, it does not generate optimal solutions.
- It is necessary to model and define all relevant data in order to produce valid results which can be very difficult in complex supply network scenarios.
 - Simulation models can be classified into several categories and based on more than a few ways. Kleijnen (2005) classifies simulation models into four types:
- **Spreadsheet Simulation:** Used in corporate modeling usually by managers;
- **System Dynamics (SD):** Organizations are viewed as systems with six types of flows: materials, goods, personnel, money, orders, and information;
- **Discrete-Event Dynamic Systems (DEDS) Simulation:** Incorporates uncertainties and is usually part of the ERP systems;
- **Business Games:** Interactive simulation where managers operate within the simulated ‘world’.

During the past few decades, a number of papers have been published that deals with different aspect of supply chain modeling and simulation.

Some of them are purely mathematical models using the linear or mixed integer programming (Ettl et al., 2000), which are not fully applicable in a complex and volatile business environment. Those mathematical models have quite limited modeling power because they are based on simplifying assumptions and typically do not support



uncertainty (Buckley & An, 2005). Some authors proposed a hybrid approach combining analytics and simulation methods for solving production–distribution problems in supply networks (Lee & Kim, 2000).

Tunali et al. (2011) applied a hybrid model which combines mathematical modeling and simulation in a produce-to-order manufacturing company. First, a mathematical model is used to generate production and distribution plans while minimizing costs. Based on these plans, optimal scheduling decisions are made. Finally, the dynamic and stochastic simulation is used to assess these decisions.

Li et al. (2012) used stochastic network mathematical model for analyzing the ordering dynamics and for identifying the key links in ordering networks. Based on these models and using simulation, effect of different ordering policies on economic performance of multi-stage supply chain are evaluated. The goal is to produce a decision support system for designing effective strategies to improve supply chain performance. The use of simulation for understanding the performance of supply chain collaboration is relatively a new approach and this can be used by companies that are interested in collaboration without having to invest a huge sum of money in establishing the actual collaboration (Ramanathan, 2014). Cigolini et al. (2014) used discrete event simulation along with scenario design technique and statistical analysis for studying how supply chain configuration and management decisions impact supply chain performance. The results showed that information sharing and coordination have very positive impact on supply chain performance.

Shukla et al. (2010), introduced a hybrid approach that incorporates simulation to a comprehensive supply chain network, Taguchi method to quickly determine a robust area and optimize qualitative factors, regression analysis and Psycho-clonal algorithm which refines the result obtained by Taguchi orthogonal array. The goal is better understanding the dynamic relationships of different factors in supply chains, decreased

influence of uncertainty, and determination of production capacities. Sahay and Ierapetritou (2013) proposed a hybrid approach which combines a mathematical programming model with an agent-based simulation model and uses them in an iterative framework. The optimization model is used to guide the decisions toward an optimal allocation of resources given the realistic supply chain representation given by the simulation.

The lack of a standard way for modeling and analysis of supply networks resulted in a variety of methods for developing supply network simulation models. SCOR (Supply Chain Operations Reference) model, as a widely accepted industry standard can serve as a good starting point for designing supply network simulation software and frameworks. The e-SCOR modeling and simulation environment (Barnett & Miller, 2000) is based on the SCOR model and built upon a discrete event simulator. Herrmann & Pundoor (2003) presented the supply chain simulation framework that follows the SCOR model and that integrates discrete event simulation and spreadsheets. Because of the complexity and low usability of the distributed simulation models, a new approach based on ontologies to integrate several supply chain views and models, which captures the required distributed knowledge to build simulation models, is proposed. Persson et al. (2009) presented modeling of supply chain structure based on the SCOR model and application of discrete simulation. They used the ARENA commercial simulation software and limited number of the SCOR processes.

Agent-based distributed simulation can be very useful for modelling and analyzing such complex adaptive systems as supply chain networks. Long (2015) proposed a methodology for distributed supply chain network modelling and simulation by means of integration of agent-based distributed simulation and an improved SCOR model. This approach enables accelerated flexible supply chain modeling and reuse of simulation models. Nikolopolou and Ierapetritou (2012) proposed an agent-based hybrid simulation optimization

approach that addresses the problem to the operational/tactical level of the supply chain management. The main criterion is cost minimization. One of the constraints of this approach is that it does not take into account stochastic nature of supply chains.

Thanks to their main characteristic (autonomy, interactivity, and reactivity), agents and multi-agent systems (MAS) are well-suited for modeling and simulating the dynamic behavior of supply networks (Swaminathan, 1998). These approaches are used for simulating different business scenarios such as harbor supply chain management (Yi et al., 2002), warehouse inventory planning and materials handling (Ito & Abadi, 2002). In multi-agent systems, business entities are represented as agents, and the involved information and material flows with proposed coordination method for collaborating among agents. Labarthe et al. (2007) proposed an agent modeling framework for the modeling and simulation of supply chains. The software presented makes use of collaboration between cognitive and reactive software agents. Chatfield et al. (2007) introduced the order-centric conceptual architecture that combines simulation formalisms to represent supply chain infrastructure (nodes, actions, policies, etc.). It consists of the knowledge base implemented as a specific XML dictionary called Supply Chain Markup Language (SCML), and the Java-based simulation engine used to transform the SCML files and to run the simulation (Chatfield et al., 2004).

Jiang et al. (2009) considered dynamic inventory control and non-stationary customer demand in order to achieve a preset customer level. The case-based reinforcement learning method is used. They used multi-agent simulation to analyze different inventory replenishment policies of a simplified two-echelon supply chain. Limitations of the simulation model are the number of nodes, as well as the number of goals.

Another example of the XML-based simulation is Simulation Markup Language (SML), an open-source tool (Kilgore, 2001), and Manufacturing Information Model for Simulation (MIMS) (Qiao & Riddick, 2004) that enables design of the platform-neutral supply chain model that can be further transformed to suite the schema of the particular simulation software tool.

Ding et al. (2005) based on analysis of existing literature noted that existing models lack general applicability and stressed the need for a general approach for both supply chain modeling and its evaluation by simulation. They proposed discrete-event simulation software made of a number of different building blocks (enterprises, suppliers, manufacturers, distributors, etc.)

Other papers are based on the “beer game” method, most often used in experiments that consider small number of elements for analysis in relation to certain limited number of influence factors, such as POS (point-of-sale) information and customer demand pattern (Fayez et al., 2005).

Due to supply chain complexity (more and more people use the expression: supply network) a great deal of existing simulation approaches and solutions deals with only particular SCM segments, presume assumptions that make simulation models less realistic and usable for practical application and with reduced opportunities for validation of derived results. They usually take into consideration only a single product, whereas in real systems there is a concurrency (multiple products) for resources. Also, there are considerable limitations in terms of experimenting with various supply chain structures and policies.

The following section provides an overview of computer-based supply chain simulation which is based on discrete-event simulation, process approach and intelligent system for event management. The proposed methodology and software solution enable fulfillment of the goals described in this section.



SUPPLY CHAIN MODELING AND SIMULATION

Characteristics of the Simulation Software

Supply chain simulation software solution should fulfill the following requirements:

- It should provide intuitive and straightforward modeling environment.
- To be intuitive and simplified for the end-users.
- It should enable modeling of any supply chain structure regardless of its complexity in terms of number of levels, nodes at each level, products, business policies, etc.
- It shouldn't be constrained by number of different elements such as products, participants or relationships.
- Ability to store models and simulation results in database.
- Simulation models should be close to corresponding real-world systems – they should embody various constraints, technologies and organization.
- It should enable performance measurement through appropriate metrics and at different levels.
- Beside sensitivity analysis, it should enable design of new supply chains, as well as learning by experimenting.
- It should be useful for wide range of users – researchers, business people, educators, consultants, etc.

Supply chain complexity represents the main obstacle in fulfilling these requirements. Solution presented in this chapter is based on process approach, system decomposition into modules, SCM Metamodel, application of intelligent decision mechanisms and expert systems, as well as application of standards.

Process Approach

Many of today's well-known business models and standard initiatives such as SCOR (Supply Chain Operations Reference), ODETTE SCMo (Supply Chain Monitoring), RosettaNet, and ebXML are based on process approach.

SCOR model represents a universal approach to supply chain management that can be applied in diverse business domains. Supply Chain Operations Reference (SCOR) model integrates the well-known concepts of business process reengineering, benchmarking, and process measurement into a cross-functional framework, and represents an industry standard. This standardization movement facilitates the deliverance of business content for the supply chain simulation model.

A process reference model contains (SCC, 2010):

- Standard descriptions of management processes.
- A framework of relationships among the standard processes.
- Standard metrics to measure process performance.
- Management practices that produce best-in-class performance.
- Standard alignment to features and functionality.

According to SCOR, supply chain management consists of the following integrated processes: Plan (P), Source (S), Make (M), Deliver (D), and Return (R) – from the suppliers' supplier to the customers' customer, and all aligned with a company's operational strategy, material, work, and information flows. Additionally, SCOR includes a series of Enable (E) elements for each of the processes. Enable elements focus on information policy and relationships to enable the planning and execution of supply chain activities (Bolstorff & Rosenbaum, 2003). SCOR contains three levels

of process detail. The top level defines scope and content of the supply network. This is where the bases for the performance targets are set. At the second level, the intersection of process and process types forms process categories. The third level consists of the actual process elements and their process flows. It is this bottom level of the SCOR model that is the most significant to the simulation. Process elements are the most granular detail of the SCOR model and serve as the first level of process decomposition. Each process element has the following attributes: process ID, process name, standard definition, performance metrics, best practices (where applicable), and finally, inputs and outputs.

Software Solution for Supply Chain Simulation

Simulation software that is introduced covers a wide range of simulation categories in the supply chain domain. Additionally, it meets a series of additional conditions.

One of the most important is the closeness of the simulation models with real supply chain systems. Supply networks have complex structure in terms of number of nodes and levels in which they are in. At nodes, multitude of different processes with large number of transactions are perform. Furthermore, final products with complex bill of materials structure are often represented. The real systems exists in an environment with various constraints such as financial, operational, warehousing, and transportation capacities, as well as those related to energy and human resources.

The next group of conditions is related to modeling and simulation. Modeling is performed in a way close to the business user who can completely dedicate in solving SCM problems. Metrics harmonized with existing standards assists in interpretation of results.

The basic components of the simulation software (Stefanovic & Stefanovic, 2015) are further described.

Information system for supply chain simulation consists of the following interconnected modules within the supply network Metamodel: SCOR (Supply Chain Operations Reference) model, information system entity submodel, business environment submodel, constraints submodel, simulation module, module for assessing and improving the supply network instances and training (education) submodel. The Metamodel contains the complete SCOR model structure.

Entities of the information system are designed in such a way that enables modeling of large number of real supply networks in various business domains.

Business environment submodel represents setting in which supply network entities exist. It is expressed by the available resources such as employed capacities, financial capital, human resources, materials, and energy.

The constraints submodel represents constraints for each process and it is expressed by the amount of engaged resources. Constraints can be technical, technological, financial, and organizational (Stefanovic & Stefanovic, 2009).

The simulation submodel integrates supply network structure, characteristics of processes in nodes, and entities that correspond to the simulation methodology. Process characteristics are associated with SCOR best practices, but besides that they also include other process implementations from theory and practice. All these forms the process library.

The simulation submodel is also responsible for realization of simulation methodology. The methodology is based on discrete event simulation and process approach. Its most important part is specific expert system which manages process interactions. The structure of the expert system consists of the knowledge base and the conclusion engine. The basis for the process interaction is the specialized three-phase rule (Stefanovic & Stefanovic, 2008).

The module for assessing and improving the supply network instances includes an intelligent

system that is rule-based and consists of several steps. The first step identifies a set of metrics that do not meet predefined goals. The second set of rules establishes relationship between the set of metrics obtained at the first level of rules and corresponding processes at each location. The third set of rules establishes a connection between the resulting set of processes and best practices. The fourth set of rules relates best practices with the performance attributes and their characteristic. Performance attribute characteristic for a given best practice shows whether the performance attribute increases or decreases. This way, from the set of best practices and for all processes on all locations, appropriate best practices whose value is increases are selected (Stefanovic & Stefanovic, 2013).

Simulation process involves modeling, simulation execution, as well as reporting and analysis of the results. Modeling process includes modeling of supply network structure (structure submodel), process modeling (process submodel), resource modeling (resource submodel), and constraints modeling (constraints submodel). Structure submodel encompasses supply network nodes (suppliers, manufacturers, distribution centers, wholesalers, retailers, etc.) and their relationships. Process submodel is composed of processes at each node and their connections. Processes correlate to the SCOR Level 3 processes. Within the submodel, characteristics for each process are determined. Modeling is carried out until the level of the single product, which also contributes to the realness of the simulation model.

Simulation execution comprises a set of activities that carry out the simulation. Important elements for simulation execution are also external events, generation of initial conditions and waiting queues. External events can be customer demand or delivery of the purchased materials. They trigger other events in the system. External events can be generated in one of the three ways: by Monte Carlo method, by generating random values, or they might be predefined. Initial conditions are starting inventory of products and materials

(semiproducts), parameters of specific reorder policy (i.e. order quantity, reorder level, lead time), customer backorders, unfulfilled supplier orders, number of work shift, etc. Waiting queues occur when a certain process should start, but at that moment resources are engaged for another product for which the executing has already begun (Stefanovic & Stefanovic, 2013).

Reporting of simulation results include a set of information like values of SCOR metrics, fluctuation of inventory levels at each location and for each product, cost overview, profit, waiting lines, etc.

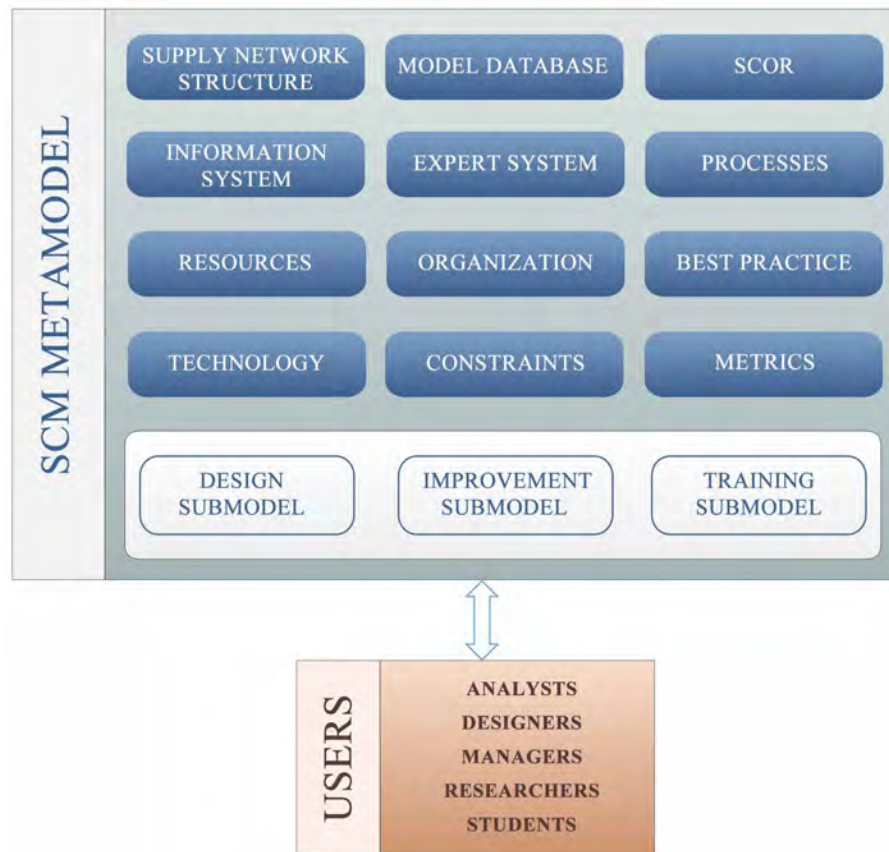
It can be noted that the simulation method and the software solution are quite complex and that they integrate several components into the unique system (Figure 1).

This gives possibility to perform a wide spectrum of supply network simulations, as well as convenience for different groups of users. Integration can be seen from the Metamodel itself. It contains the information system with entities that can cover wide range of supply network configurations. The Metamodel integrates supply network structure, processes, resources, constraints, technologies, organization, best practice and metrics. Furthermore, many of the established methods for particular processes are included in the simulation model. For example, different inventory replenishment models, models for solving transportation problems, reception of materials, warehousing, etc. Also, this way it is possible to test these models and to make comparisons. Evidently, it is possible to model some less complex systems, such as pure production or distribution system, or to test specific mathematical model for inventory replenishment by selecting appropriate processes.

Examples of Supply Chain Simulation

- **Reducing Risks of Implementing Supply Networks at Different Levels (Strategic, Tactical and Operational):** Every experimenting with supply networks in real

Figure 1. Architecture of the information system for supply network modeling and simulation



environment is very risky and can lead to huge losses and disruptions. In this regard, simulation offers many opportunities to completely avoid or to reduce mistakes and wrong decisions. It is very crucial how supply network is going to be organized, integrated and managed. Also, each process could have many forms which also can be tested through simulation (Stefanovic & Stefanovic, 2013).

- **Analysis of System Dynamics Through Uncovering Relationships and Interdependence Between Processes:** System dynamics manifests in many forms. For instance, different mathematical models for inventory management (i.e. FQS-Fixed Quantity System or FPS-Fixed Period System) contained within the purchasing process are tested in distribution systems.

They are combined with various characteristics of other processes: order fulfillment, shipment, receiving, etc. (Stefanovic & Stefanovic, 2008). Another example is testing of various supply network structures for different versions of replenishment and order fulfillment processes (Stefanovic & Stefanovic, 2009).

- **Infrastructure Configuration which Implies Arrangement of Producers, Suppliers, Distribution Centers, Wholesalers, Retailers and Their Locations (Nodes):** For single complex supply network different structures can be configured. This includes number of suppliers, location strategy, number of distribution centers or retailers, etc. Each of these configurations can be evaluated before making final decisions.

- **Defining Process Strategy at the Nodes and Coordination Between Processes with the Purpose of Their Alignment and Achieving Goals at the Global Supply Network Level:** Single process (or node) optimization does not necessarily mean the best results at the global level. It is necessary to align process characteristics in order to maximize results at the supply network level. For example, in certain cases the FQS system will produce better results, and in other the FPS system will perform better (Stefanovic & Stefanovic, 2011).
- **Analyzing Effects of Information Sharing:** It is possible to simulate different degrees of information sharing (visibility) among processes and supply network nodes and to observe effects of information sharing and to make decisions about what information should be shared and between what processes. This could significantly reduce the bullwhip effect and improve coordination of supply network activities.
- **Evaluation of Supply Network Through Performance Monitoring which Involves Defining the Metrics at Different Supply Network Levels:** Performance measurement is carried out according to the SCOR metrics (Stefanovic 2014), and also with other metrics available from the literature such as profit, stock level, capacity utilization, etc. (Stefanovic 2015).
- **Supply Network Design:** Besides supply network testing, methodology for designing supply networks has been also defined. It utilizes simulation for validation of the design. The process includes defining the supply network structure, processes in nodes and their relationships, process characteristics and resources that are engaged, according to the project goals. Additionally, constraints can be taken into consideration (Stefanovic & Stefanovic, 2013).
- **Training and Learning About Supply Chain Management:** Simulation software enables learning at different levels. First, it provides studying the SCOR model. All the standard information from the SCOR model can be obtained and also to make intelligent search. For example, users can obtain answers to questions such as: What processes influence cost reduction in supply chain? Besides this static learning, interactive learning based on simulation is also possible. The most useful case is using the simulation for understanding SCM processes and characteristics such as uncertainty and bullwhip effect.
- **Cash Flow Simulation:** Cash flow simulation is one of the future research projects. The simulation software includes necessary elements for experimenting in this domain, that is, it encompasses processes and metrics related to cash flow. These elements are integrated with other processes and metrics, so the cash flow is not analyzed separately. Cash flow simulation is integral part of the overall supply network simulation. Cash flow is observed from a point of taking actions for mitigating risk of running out of cash in certain period. This way, it is possible to analyze cash flow in relation to various variables like initial stock or purchase price.

Based on the described simulation scenarios it can be concluded that besides integrativity of the software itself, there is integrativity in terms of problem that can be analyzed and solved. In other words, by creating models for specific goals, it is also possible to obtain the complete set of results for other business needs.

Presented simulation software fulfils main requirements for efficient and effective supply chain simulation which is demonstrated in several research results (Stefanovic & Stefanovic, 2008; Stefanovic et al., 2009; Stefanovic & Stefanovic, 2013). It is flexible enough to be applied in vari-



ous supply chain networks regardless of the size, structure, complexity and dynamics. Here are some of the application examples:

- Measuring influence of different supply chain process types, their relationships and realization methods at the global level. These involve various replenishment models, order processing, shipment, initial stock and lead time. Some of the metrics include costs, profit and serviceability.
- Analysis and assessment of supply chain structures in terms of different inventory management models. Performance measures are based on the SCO metrics.
- Analysis and assessment of supply chain performance in situation where there is a concurrent usage of production capacities and other resources which results in waiting queues. Besides standard SCOR-based metrics, the metric which shows the waiting duration of certain products and components during the process execution.
- Supply chain design and reengineering using an intelligent modeling and simulation system.

On the other hand, the software can be used by different user groups such as supply chain architects, managers, people working in logistics or operation research, as well as students.

FUTURE RESEARCH DIRECTIONS

As supply chain management is becoming more and more critical for businesses, methods, techniques and tools for managing and analyzing these complex business networks will be topic of extensive research and development in years to come. Having in mind successful applications of simulation tools and models so far, simulation will definitely be important segment of this research. The future research trends can be categorized in the following areas:

- Supply chain modeling
- Standardization
- Integration of simulation models
- Intelligent analysis and optimization

Regarding the presented simulation approach, future research directions will be directed toward two main segments: first segment relates to experimenting, and the other relates to data analysis and mining. The new experiments will be focused on analysis of supply chain financials (costs and cash-flow), as well as supply chain information exchange and coordination (i.e. bullwhip effect). Considering that simulation software enables efficient modeling and simulation it is possible to create many models which are stored in the model library along with simulation results. These data sets can be used in conjunction with data mining tools in order to extract new knowledge, find hidden patterns and predict future.

CONCLUSION

Supply chain management in today's business climate characterized by high uncertainty, globalization of business, outsourcing, shorter product life-cycles, and high customer expectations is extremely challenging. Companies are now competing as parts of a supply chains and there is a growing need for the global planning and management of the distributed and intertwined supply chain processes.

High complexity of supply networks and the need for applying logistic principles (global approach, effectiveness, and efficiency) have led to an array of significant research domains that are directed toward improving the supply network performance. Many of this research efforts are based on simulation.

That is why the goal was set to develop the methodology and the software solution that would encompass as many of the important domains of supply network simulation application. This resulted in integrative simulation software capable

to support various simulation needs. It can be used for modeling, execution and analysis of simulation results. Modular structure of the software enables modifications and further extensions with the purpose of supporting future research projects related to SCM.

Simulation can be used to more scientifically decide how to modify processes and improve supply chain operations. Newer change-decision models incorporate simulation into the analysis. Applied within the context of experience, simulation can help organizations to analyze different supply chain scenarios, reduce costs, manage risk, diminish problems and ensure market share.

This chapter deals with supply chain simulation, provides literature review and presents unique approach for supply chain modeling and simulation. This approach offers:

- **Flexibility:** Any supply chain configuration can be modeled.
- **Extensibility:** Model database and business logic can be extended.
- **Expressiveness:** Models embody all important characteristics.
- **Authenticity:** Simulation models and execution are close to real systems.

The developed software solution enables straightforward and efficient modeling of supply chain structure and dynamics, as well as experimenting through simulation execution and analysis of results.

REFERENCES

- Barnett, W. M., & Miller, J. C. (2000). Analysis of the virtual enterprise using distributed supply chain modeling and simulation: An application of e-SCOR. In *Proceedings of the 2000 Winter Simulation Conference*, (pp. 352–355). doi:10.1109/WSC.2000.899738
- Bolstorff, P., & Rosenbaum, R. (2003). *Supply chain excellence: a handbook for dramatic improvement using the SCOR model*. New York: Amacom.
- Buckley, S., & An, C. (2005). Supply Chain Simulation. In *Supply Chain Management on Demand* (pp. 17-35). Berlin: Springer Berlin Heidelberg. doi:10.1007/3-540-27354-9_2
- Chatfield, D. C., Hayya, J. C., & Harrison, T. P. (2007). A multi-formalism architecture for agent-based, order-centric supply chain simulation. *Simulation Modelling Practice and Theory*, 15(2), 153–174. doi:10.1016/j.simpat.2006.09.018
- Cigolini, R., Pero, M., Rossi, T., & Sianesi, A. (2014). Linking supply chain configuration to supply chain performance: A discrete event simulation model. *Simulation Modelling Practice and Theory*, 40, 1–11. doi:10.1016/j.simpat.2013.08.002
- Ding, H., Benyoucef, L., & Xie, X. (2005). A Modeling and Simulation Framework for Supply Chain Design. In A. Dolgui, J. Soldek, & O. Zaikin (Eds.), *Supply Chain Optimisation: Product/Process Design, Facility Location and Flow Control*. Springer Science + Business Media, Inc. doi:10.1007/0-387-23581-7_16
- Ettl, M., Feigin, G., Lin, G., & Yao, D. (2000). A supply network model with base-stock control and service requirements. *Operations Research*, 48(2), 216–232. doi:10.1287/opre.48.2.216.12376
- Herrmann, J., Lin, E., & Pundoor, G. (2003). Supply chain simulation modeling using the supply chain operations reference model. In *Proceedings of the ASME Design Engineering Technical Conference*. Chicago: ASME. doi:10.1115/DETC2003/CIE-48220
- Ito, T., & Abadi, J. M. (2002). Agent-based material handling and inventory planning in warehouse. *Journal of Intelligent Manufacturing*, 13(3), 201–210. doi:10.1023/A:1015786822825



- Jiang, C., & Sheng, Z. (2009). Case-based reinforcement learning for dynamic inventory control in a multi-agent supply-chain system. *Expert Systems with Applications*, 36(3), 6520–6526. doi:10.1016/j.eswa.2008.07.036
- Kilgore, R. A. (2001). Open source simulation modelling language (SML). In *Proceedings of the 2001 Winter Simulation Conference* (pp. 607–613). IEEE. doi:10.1109/WSC.2001.977345
- Kleijnen, J. P. C. (2005). Supply chain simulation tools and techniques: A survey. *International Journal of Simulation and Process Modelling*, 1(1-2), 82–89.
- Labarthe, O., Espinasse, B., Ferrarini, A., & Montreuil, B. (2007). Toward a methodological framework for agent-based modelling and simulation of supply chains in a mass customization context. *Simulation Modelling Practice and Theory*, 15(2), 113–136. doi:10.1016/j.simpat.2006.09.014
- Lee, H. Y., & Kim, H. S. (2000). Optimal production–distribution planning in supply chain management using a hybrid simulation-analytic approach. In *Proceedings of the 2000 Winter Simulation Conference* (pp. 1252–1259).
- Li, C., & Liu, S. (2012). A stochastic network model for ordering analysis in multi-stage supply chain systems. *Simulation Modelling Practice and Theory*, 22, 92–108. doi:10.1016/j.simpat.2011.12.001
- Long, Q. (2014). Distributed supply chain network modelling and simulation: Integration of agent-based distributed simulation and improved SCOR model. *International Journal of Production Research*, 52(23), 6899–6917. doi:10.1080/00207543.2014.910623
- Nikolopoulou, A., & Ierapetritou, M. G. (2012). Hybrid simulation based optimization approach for supply chain management. *Computers & Chemical Engineering*, 47, 183–193. doi:10.1016/j.compchemeng.2012.06.045
- Persson, F., & Araldi, M. (2009). The development of a dynamic supply chain analysis tool— Integration of SCOR and discrete event simulation. *International Journal of Production Economics*, 121(2), 574–583. doi:10.1016/j.ijpe.2006.12.064
- Qiao, G., & Riddick, F. (2004). Modelling information for manufacturing-oriented supply-chain simulations. In *Proceedings of the 2004 Winter Simulation Conference* (pp. 1184–1188). IEEE.
- Ramanathan, U. (2014). Performance of supply chain collaboration – A simulation study. *Expert Systems with Applications*, 41(1), 210–220. doi:10.1016/j.eswa.2013.07.022
- Sahay, N., & Ierapetritou, M. (2013). Supply chain management using an optimization driven simulation approach. *AIChE Journal. American Institute of Chemical Engineers*, 50(12), 4612–4626. doi:10.1002/aic.14226
- SCC. (2010). *SCOR v9.0*. Washington, DC: Supply Chain Council, Inc.
- Shukla, S. K., Tiwari, M. K., Wan, H.-D., & Shankar, R. (2010). Optimization of the supply chain network: Simulation, Taguchi, and Psychoclonal algorithm embedded approach. *Computers & Industrial Engineering*, 58(1), 29–39. doi:10.1016/j.cie.2009.07.016
- Stefanovic, D., & Stefanovic, N. (2008). Methodology for modeling and analysis of supply networks. *Journal of Intelligent Manufacturing*, 19(4), 485–503. doi:10.1007/s10845-008-0098-0
- Stefanovic, D., Stefanovic, N., & Radenkovic, B. (2009). Supply network modelling and simulation methodology. *Simulation Modelling Practice and Theory*, 17(4), 743–766. doi:10.1016/j.simpat.2009.01.001
- Stefanovic, N. (2014). Proactive Supply Chain Performance Management with Predictive Analytics. *The Scientific World Journal*, Article ID 528917. doi:10.1155/2014/528917

Stefanovic, N. (2015). Collaborative predictive business intelligence model for spare parts inventory replenishment. *Computer Science and Information Systems*, 12(3), 911–930. doi:10.2298/CSIS141101034S

Stefanovic, N., & Stefanovic, D. (2004). Supply chain modelling and simulation. *JISA INFO*, 5, 116–120.

Stefanovic, N., & Stefanovic, D. (2013). Integrated and interactive software solution for knowledge-based supply network design. *Computer Systems Science and Engineering*, 28(1), 5–23.

Stefanovic, N., & Stefanovic, D. (2015). Supply Chain Simulation. In M. Khosrow-Pour (Ed.), *Encyclopedia of Information Science and Technology* (3rd ed.; pp. 5271–5282). Hershey, PA: Information Science Reference; doi:10.4018/978-1-4666-5888-2.ch521

Stefanovic, N., Stefanovic, D., & Radenkovic, B. (2011). Integrated Supply Chain Intelligence through Collaborative Planning, Analytics and Monitoring. In I. Mahdavi, S. Mohebbi, & N. Cho (Eds.), *Electronic Supply Network Coordination in Intelligent and Dynamic Environments: Modeling and Implementation* (pp. 43–92). Hershey, PA: Business Science Reference; doi:10.4018/978-1-60566-808-6.ch003

Swaminathan, M. J., Haas, A. W., Smith, F. S., & Sadeh, M. N. (1998). Modelling supply chain dynamics: A multiagent approach. *Decision Sciences*, 29(3), 607–632. doi:10.1111/j.1540-5915.1998.tb01356.x

Thierry, C., Bel, G., & Thomas, A. (2010). The Role of Modeling and Simulation in Supply Chain. *Management SCS M&S Magazine*, 1(4), 1–8.

Tunali, S., Ozfirat, P. M., & Ay, G. (2011). Setting order promising times in a supply chain network using hybrid simulation-analytical approach: An industrial case study. *Simulation Modelling Practice and Theory*, 19(9), 1967–1982. doi:10.1016/j.simpat.2011.04.014

Yi, W. D., Kim, H. S., & Kim, H. N. (2002). Combined modelling with multi-agent systems and simulation: its application to harbor supply chain management. In *Proceedings of the 35th Hawaii International Conference on System Sciences* (pp. 1615–1624). IEEE.

KEY TERMS AND DEFINITIONS

Business Process: A business process is a complete set of end-to-end activities that together create value for the customer.

Knowledge Base: A specialized database that serves as repository for information and knowledge management.

Metamodel: A conceptual model which defines concepts, relationships, and semantics and enables creation of concrete models.

Metric: A measure of particular process characteristic.

Model: A model is a simplified representation of a system at some particular point in time or space intended to promote understanding of the real system.

Performance Measurement: A process for collecting and reporting information regarding the performance of an individual, group or organizations. It can involve looking at process/strategies in place, as well as whether outcomes are in line with what was intended or should have been achieved.

Simulation: Simulation is defined as the process of creating a model of an existing or proposed system in order to identify and understand those factors which control the system and/or to predict the future behavior of the system.

Supply Network: Dynamic, interconnected and collaborative group of companies working jointly on planning, management and execution of cross-company business processes spanning from the first tier suppliers to the end-customers.

Latest Advances on Benders Decomposition



Antonios Fragkogios
University of Thessaly, Greece

Georgios K. D. Saharidis
University of Thessaly, Greece

INTRODUCTION

Nowadays, the increased demand for energy and goods around the world has made the business environment more and more complex concerning the design and operation of power systems, supply chain and transportation networks, production scheduling of factories etc. This complexity has made real case studies increasingly difficult to solve due to the large-scale mathematical models constructed. These models do not only depend on the number of parameters, decision variables and constraints. Frequently, even if these features are moderate, optimization problems could still be considered as large-scale due to complicating structures making their solution hard to get.

Information Science and Technology together with Operations Research and Mathematical Programming can show the way towards better solutions in all fields of economy. Nowadays, the evolution of computers combined with mathematical decomposition techniques can solve real size mathematical models and optimize the operation of real-life systems by e.g. avoiding an unpleasant outcome, decreasing operations cost or increasing the overall profit.

The mathematical decomposition techniques aim at decomposing the initial large-scale mathematical model into two or more sub-models based on mathematical programming theorems, as for example, the duality theorem. In principle, the mathematical decomposition techniques do not depend on the case study, but on the structure

of the developed mathematical model, which describes the system under study. Mathematical decomposition techniques are:

- Benders decomposition (Benders, 1962),
- Dantzig-Wolfe decomposition (Dantzig & Wolfe, 1960),
- Lagrangian relaxation,
- Lagrangian decomposition and
- Cross decomposition.

The main goal of this chapter is to present the Benders decomposition method, identify its weaknesses and display all the recent research made to tackle them.

In the next section, a very brief literature review on the method is provided. Following, the method is presented and explained together with its main weaknesses. The same section includes the latest advances from the literature, which address these weaknesses. The next section, future and emerging trends on Benders decomposition are discussed. Finally, the chapter is concluded at the final section.

BACKGROUND

Benders decomposition technique was introduced by (Benders, 1962) for linear problems with coupling integer variables. Although the algorithm was initially used for block-decomposed large-scale optimization problem, lately the algorithm

Table 1. Block-decomposable structure of a problem suitable for Benders Decomposition approach

$A =$	a_{11}	a_{12}	a_{13}	...	0	0	0	...	a_{1j-2}	a_{1j-1}	a_{1j}
	a_{21}	a_{22}	a_{23}	...	0	0	0	...	a_{2j-2}	a_{2j-1}	a_{2j}
	...	...	...	...	...	...	...	...	...	...	...
	0	0	0	...	a_{nk-1}	a_{nk}	a_{nk+1}	...	a_{nj-2}	a_{nj-1}	a_{nj}
	0	0	0	...	a_{n+1k-1}	a_{n+1k}	a_{n+1k+1}	...	a_{n+1j-2}	a_{n+1j-1}	a_{n+1j}
	...	...	...	...	...	...	...	...	...	...	...
	0	0	0	...	0	0	0	...	a_{i-1j-2}	a_{i-1j-1}	a_{i-1j}
	0	0	0	...	0	0	0	...	a_{ij-2}	a_{ij-1}	a_{ij}
	...	...	...	...	...	...	...	...	...	...	...
	0	0	0	...	0	0	0	...	a_{i-1j-2}	a_{i-1j-1}	a_{i-1j}
	0	0	0	...	0	0	0	...	a_{ij-2}	a_{ij-1}	a_{ij}

has been applied also in other types of problems. The Benders method has been applied successfully to stochastic programming (Watkins, McKinney, Lasdon, Nielsen, & Martin, 2000) for the solution of multi-scenario, two-stage and multistage stochastic problems where the scenarios are split to sub-scenarios and studied separately decreasing the amount of data taken under consideration in each optimization step; to global optimization problems (Zhu & Kuno, 2003) to address the non-convexity nature of big data; and more recently to build mathematical models (Ierapetriou & Saharidis, 2009), where a large amount of experimental data are available for model building.

BENDERS DECOMPOSITION APPROACH

Presentation of the method

The Benders decomposition (Benders, 1962) technique is based on the idea of exploiting decomposable structure of a given problem so that its solution can be converted into the solution of several smaller sub-problems. Table 1 shows this block-decomposable structure, where A is the matrix of the constraints coefficients.

For the construction of the subproblems, certain variables of the original problem are considered to be complicating. By fixing them, the original problem is decomposed into the Relaxed Master Problem (RMP), which contains only the complicating variables, and the Primal Sub-Problem (PSP), which contains the rest of the variables. Thus, RMP is a relaxation of the original problem and is expected to provide the optimal solution after the addition of a number of constraints (i.e. inequalities).

Without loss of generality, one could consider the following linear problem:

Original Problem (OP):

$$\text{Min } c^T x + d^T y$$

st.

$$Ax + By \leq b$$

$$x \in \mathbb{R}_+^n, y \in \mathbb{Z}_+^q$$

where $c \in \mathbb{R}^n, d \in \mathbb{R}^q, b \in \mathbb{R}^m$, A and B are $m \times n$ and $m \times q$ matrices respectively.

Assuming vector y contains a number of decision variables, which are considered to be complicating, the decision variables are partitioned into two sets x and y and the OP decomposes into the following problems:

Primal Slave Problem (PSP):

$$\begin{aligned} \text{Min } f(x) &= c^T x + d^T \bar{y} \\ \text{st.} \\ Ax &\leq b - B\bar{y} \\ x &\in \mathbb{R}_+^n \end{aligned}$$

Where $\bar{y}$ is the fixed values of the complicating variables, obtained by solving RMP.

Relaxed Master Problem (RMP):

$$\begin{aligned} \text{Min } F(y, z) &= z \\ \text{st.} \\ v^{iT}(b - By) &\leq 0 \\ u^{jT}(b - By) + d^T y - z &\leq 0 \\ z &\geq 0, y \in Z_+^q \end{aligned} \left. \vphantom{\begin{aligned} \text{Min } F(y, z) = z \\ \text{st.} \\ v^{iT}(b - By) \leq 0 \\ u^{jT}(b - By) + d^T y - z \leq 0 \\ z \geq 0, y \in Z_+^q \end{aligned}} \right\} \text{Benders Cuts}$$

where v^i is the vector that corresponds to the extreme ray i of the dual of PSP and u^j is the vector that corresponds to the extreme point j . An extreme ray (point) is a ray (point) that cannot be expressed as a linear combination of any other rays (points).

The dual of PSP has the following form:

Dual Slave Problem (DSP):

$$\begin{aligned} \text{Min } f'(u) &= u^T(b - B\bar{y}) \\ \text{st.} \\ A^T u &\geq c \\ u &\in \mathbb{R}^m \end{aligned}$$

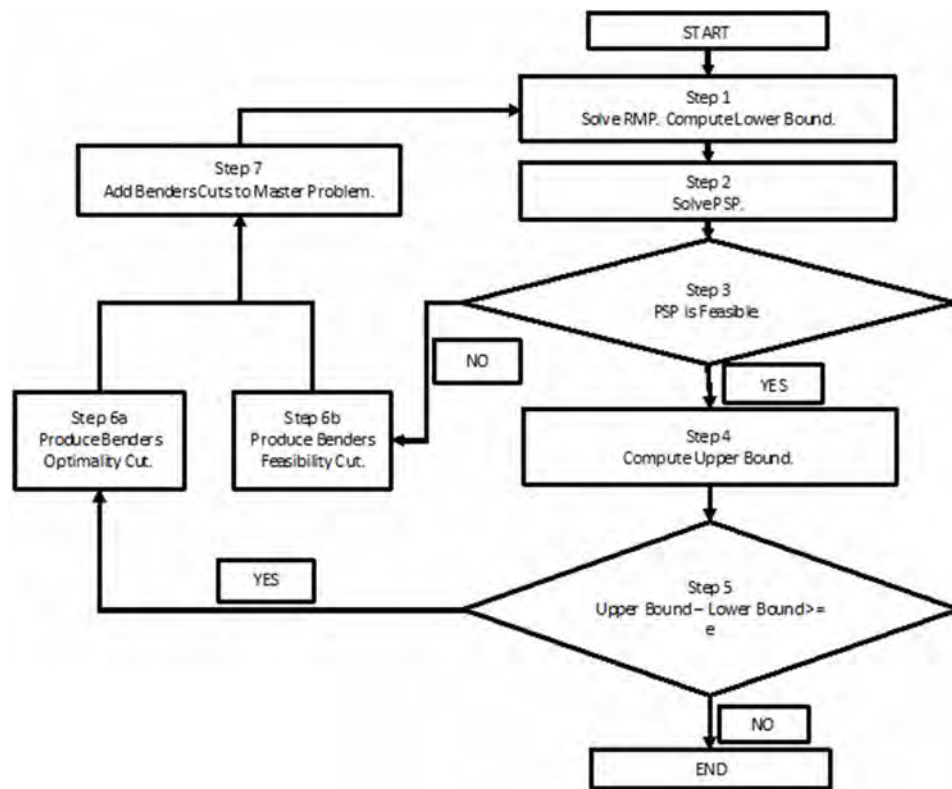
The Benders decomposition algorithm is an iterative algorithm. In each iteration, the RMP is solved and its solution (the values of the complicating variables) is transferred to the PSP. Afterwards, the PSP is solved. If the PSP is infeasible (DSP is unbounded), a *Benders feasibility cut* is created ($v^{iT}(b - By) \leq 0$), where v^i is the vector that corresponds to the extreme ray i of DSP. The feasibility cut is then added to the RMP. If the PSP is feasible (DSP is bounded), a *Benders*

optimality cut is created ($u^{jT}(b - By) + d^T y - z \leq 0$), where u^j is the vector that corresponds to extreme point j of DSP. The optimality cut is then added to the RMP. Thus, in each iteration RMP incorporates new constraints, so more information about the solution space of the OP. Afterwards, the enriched RMP is solved again and the procedure is repeated.

In each iteration in the case of minimization (-maximization), the RMP gives a valid lower (-upper) bound as a relaxation of the original problem and the SP gives a valid upper (-lower) bound as a restriction of the original problem. Thus, the convergence criterion for the termination of the Benders algorithm is defined based on the difference between the minimum upper (-lower) bound and the current lower (-upper) bound. When this difference is lower than a small number or equal to zero the algorithm stops, otherwise it continues. In Figure 1, Benders decomposition algorithm is described using a flowchart.

One could observe that during the iterations of the algorithm only the objective function of DSP is updated, as the constraints are not modified. This means that the solution space of DSP is always the same. If all possible Benders cuts were produced using all the extreme points and extreme rays of the DSP solution space, then the resulting augmented RMP is an equivalent version of OP and its solution yields the same optimal solution as the OP. The total number of Benders cuts, which equals the number of extreme points and extreme rays of DSP is, generally, enormous. However, it is known that at the optimum the number of RMP active constraints will never exceed the number of RMP decision variables (Saharidis, Minoux, & Ierapetritou, 2010). The main idea of the Benders algorithm is based on the observation that the algorithm will converge, satisfying the optimality condition before the addition of all Benders cuts. Note that the finite convergence of the algorithm results from the fact that DSP has a finite number of extreme points and extreme rays.

Figure 1. Flowchart of the Benders Decomposition Algorithm



Main Weaknesses of Benders Decomposition Algorithm

Benders decomposition has been applied in a lot of cases with notably good results (Conejo, Castillo, Minguez, & Garcia-Bertrand, 2006). However, in some cases the straightforward application of the classical Benders algorithm does not lead to fast convergence (Saharidis, Minoux, & Ierapetritou, 2010). A lot of publications exist to support the fact that Benders algorithm has not been uniformly successful in all the applications and has some deficiencies.

The main issues associated with slow convergence of the Benders algorithm (Saharidis & Ierapetritou, 2010) are:

1. The solution of resulted sub-problems and
2. The quality of the cuts produced in each iteration

Solution of Sub-Problems

As discussed in the previous section, during the application of the classical Benders algorithm, two problems are solved in each iteration, i.e. the RMP and the PSP. In many cases, the PSP and RMP are hard to be solved to optimality. The most frequent problem is the solution of the RMP which is usually an integer problem and should be solved in every iteration of the Benders algorithm.

For tackling this difficulty, (McDaniel & Devine, 1977) proposed the generation of cuts by relaxing the integrality constraints of the RMP. The integer restrictions of RMP are dropped for some iterations of the algorithm and in order to converge to optimality they are reinserted. The presented results appeared to be promising, although in some cases the classical algorithm can still be more efficient.

(Cote & Laughton, 1984) proposed three heuristic approaches to solve the RMP using the obtained solution to generate the Benders cuts from the PSP. The application of their three approaches has demonstrated very good results.

The weak lower bounds (in case of minimization) obtained by the RMP is another reason leading to slow convergence of the Benders decomposition algorithm, due to the many iterations (thus many solutions of RMP) needed to reach convergence. The development and introduction of valid inequalities to the RMP, eliminating a priori a number of infeasible solutions, is a successful strategy to lessen the number of iterations and, thus speed-up the Benders decomposition algorithm producing stronger lower bounds. (Cordeau, Soumis, & Desrosiers, 2000) added three sets of valid inequalities to RMP accelerating the convergence to the optimal solution of the locomotive and car assignment problem. (Cordeau, Pasin, & Solomon, 2006) introduced a new formulation of the logistics network design problem, where the Benders decomposition algorithm is applied and a series of valid inequalities are presented as the part of the solution methodology. (Andreas & Smith, 2009) presented a series of valid inequalities for the sub-tour elimination in order to ensure the existence of a feasible solution, each time that the PSP is resolved. (Tang, Jiang, & Saharidis, 2013), in order to further narrow the solution space of the initial RMP, propose some valid inequalities for the logistics network design problem with capacity expansions of existing warehouses. These valid inequalities combine all binary variables of the RMP problem to speed up the convergence to the optimal solution by no longer generating feasibility Benders cuts. (Saharidis, Minoux, & Dallery, 2009) have developed a series of valid inequalities for the optimal scheduling of crude oil in a refinery network, where Benders decomposition was applied in order to solve large instances of the problem.

The main issue concerning the method of initializing the RMP is that the generated valid inequalities are case-study-dependent. For this

reason, (Saharidis, Boile, & Theofanis, 2011) presented a series of valid inequalities which constitutes the first attempt to generalize the previously developed valid inequalities and extend applicability to the general fixed-charged network design problem. The presented numerical examples show that the initial lower bound of RMP using the valid inequalities increased significantly compared to the classical Benders algorithm resulting to considerable reduction in the CPU time.

A contemporary and very promising approach for the warm starting of the RMP is proposed by (Naoum-Sawaya & Elhedhli, 2013). The authors present an interior-point branch-and-cut algorithm for structured integer programs based on Benders decomposition and the analytic center cutting plane method (ACCPM).

They show that the ACCPM based Benders cuts are both pareto-optimal and valid for any node of the branch-and-bound tree and are used to warm-start the solution of the nodes after branching. The proposed ACCPM-Benders-branch-and-cut algorithm (ABBC) is compared to the classical Kelley's like Benders-branch-and-cut (KBBC) algorithm and the classical Benders decomposition with and without pareto-optimal cuts. The computational results show that ABBC outperforms all other approaches.

(Fakhri & Ghatee, 2013) propose a modified Benders algorithm for solving a preemptive multi-objective mixed-integer program. In fact, the authors make certain changes on the RMP and, specifically, alter its objective function and add a certain constraint. This method seems to be very promising in terms of reducing the number of iterations needed to reach convergence.

Quality of Benders Cuts

The second issue regarding the efficiency of Benders implementation is the quality of the Benders cuts. As mentioned in the previous section, every extreme point and extreme ray of the solution space of the DSP gives a valid cut for the RMP.

The maximum number of iterations of Benders algorithm is equal to the total number of extreme points and rays of the solution space of the DSP.

Prior to the execution of the Benders algorithm, it is known that it can converge in only one iteration if all possible Benders cuts were produced from the first iteration (full master problem). However, in this case the resulting full master problem would normally be more complicating than the original problem (Saharidis, Minoux, & Ierapetritou, 2010). Moreover, if we could define in advance which of the extreme points and rays correspond to the active constraints of the final RMP, then the Benders algorithm could converge in one iteration after the addition of these Benders cuts. Although this is practically impossible, a multi-generation of cuts strategy could be developed and applied in order to approximate this technique. Thus, generating additional constraints in every iteration could improve the convergence rate of the Benders algorithm. In order to be more efficient, this type of strategy should keep a balance between the number of extra cuts generated in each iteration and the size of the resulted RMP.

(Magnanti & Wong, 1981) are the first to propose generating more than one cut in each iteration to accelerate the Benders decomposition algorithm. The proposed multi-generation of cuts procedure uses Pareto-optimal cuts added to RMP together with the classical Benders cut. A cut is defined as Pareto-optimal if no other cut dominates it. The results obtained by their approach showed a significant improvement in the convergence of the algorithm. Later, (Papadakos, 2008) proved that it is not necessary to use a core point of the RMP solution space in order to produce a Pareto cut improving the convergence rate of the algorithm.

(Tang, Jiang, & Saharidis, 2013) have developed a strategy which produces Pareto cuts with high density, covering a high number of RMP decision variables. A high density cut as defined by (Saharidis, Minoux, & Ierapetritou, 2010) is a cut which covers a large number of decision variables of the RMP problem, resulting to a better convergence behavior of the algorithm due to the

restriction of the RMP solution space towards all its directions. The idea of the new method referred to as High Density Pareto (HDP) cut generation is based on (Papadakos, 2008) result. HDP (Tang, Jiang, & Saharidis, 2013) produces an additional cut which covers a large number of RMP decision variables using one of the multi-optimal solutions generated by the dual sub-problem.

Recently, (Sherali & Lunday, 2013) proposed an algorithmic strategy that utilizes a preemptively small perturbation of the right-hand-side of the PSP to generate maximal non-dominated Benders cuts, as well as a complimentary strategy that generates an additional cut in each iteration via an alternative emphasis on decision variable weights. The suggested technique represents a fundamental improvement, as it reduces the number of linear programs to be solved in each sub-problem iteration and that it does not require the identification of a core point within the RMP.

(Rei, Cordeau, Gendreau, & Soriano, 2009) showed how Benders feasibility cuts can be strengthened or replaced by local branching constraints. The main idea behind local branching is to divide the feasible region of a problem into smaller sub-regions and find the best solution (or at least a good feasible solution) in each one of them. By applying local branching throughout the solution process, one can simultaneously improve both the lower and upper bounds of the algorithm. To assess the performance of the idea, computational experiments were performed on a series of network design problems illustrating its benefits.

(Fischetti, Salvagnin, & Zanette, 2010) propose a new Benders cut selection criterion, that leads to a more clever choice of the separated cuts, in particular when both feasibility and optimality violated cuts exist. The authors propose a unified framework that generates a single type of cut, based on the fact that the primal SP can always be reformulated as an infeasible problem, i.e., the dual SP can be transformed into an unbounded problem. The computational results show the increased efficiency of the method compared to the classical Benders algorithm.

(Saharidis, Minoux, & Ierapetritou, 2010) presented a new multi-generation cut method suitable for case studies where the Benders cuts produced are most often low density cuts. A low density cut is a cut involving a small number of decision variables of the RMP, therefore its contribution to strengthening the RMP tends to be limited. The addition of a single low density cut does not restrict significantly the solution space of the RMP, thus leading to increased number of iterations and CPU solution time. The proposed strategy uses the concept of the Covering Cut Bundle (CCB). The CCB algorithm generates in each iteration a bundle of low density cuts by an auxiliary problem using the values obtained by the solution of the last RMP. Note that, in general, a bundle of low density cuts is more desirable for acceleration of the algorithm than a cut corresponding to the sum of these low density cuts, which is a high density cut itself. The produced bundle of cuts involves more decision variables of the RMP, thus restricting its feasible solution space and leading to faster convergence of the algorithm.

In (Azad, Saharidis, Davoudpour, Malekly, & Yektamarian, 2013), the authors present a method referred to as the maximum density cut (MDC) generation, which is based on the idea to generate a cut to cover the remaining decision variables which are not covered by CCB. For this reason, the authors suggest the generation of a cuts bundle, where only a certain number of decision variables is covered and the rest are covered by the MDC method. Furthermore, in (Saharidis & Ierapetritou, 2013), the MDC method is evolved and categorized into primal-MDC and dual-MDC. Its implementation as a standalone procedure accelerates significantly the classical Benders algorithm, but is outperformed by CCB. The combination of the two methods (CCB-MDC), however, leads to faster convergence than the standalone application of any of them.

A new multi-generation cuts algorithm is presented in (Saharidis & Ierapetritou, 2010) to improve the efficiency of the Benders decomposi-

tion approach for the cases where optimality cuts are difficult to be achieved within the iterations of the algorithm. This strategy is referred to as maximum feasible sub-system (MFS) cut generation strategy. In each iteration of the Benders algorithm, whenever the PSP is infeasible, the maximum feasible sub-system of the PSP is defined relaxing the remaining infeasible constraints with minimum changes and an additional cut is generated improving the RMP's objective function. Thus, each time the Benders algorithm produces a feasibility cut, an extra MFS cut is produced. The numerical results show a significant reduction of the solution time and of the total number of iterations, confirming the efficiency of the MFS cuts.

Infeasible PSPs have lately been addressed by (Martins de Sa, de Camargo, & de Miranda, 2013), as well, who introduce a new Benders cut selection scheme. Their method is based on the idea of retrieving cost information even when infeasible PSPs have to be addressed and extends the work of (Papadakos, 2008) by selecting a suitable value for the weight λ rendering a feasibility and also a Pareto-optimal cut from the infeasible PSPs. The new scheme outperforms other modern variants of the Benders algorithm (Papadakos, 2008), (Fischetti, Salvagnin, & Zanette, 2010).

(Jenabi, Fatemi Ghomi, Torabi, & Hosseini, 2015) address the power expansion planning problem and improved the convergence of the classical Benders algorithm by proposing three groups of additional cuts and adding them to the RMP. The first set of cuts, namely *valid inequalities* (VI), restricts the RMP from first iteration improving the quality of lower bound obtained. Second set of cuts are obtained by disaggregation of classical Benders cuts leading to *multi-generation of cuts*. For the third method, *high-density cuts* (HDC) are constructed by a combination of two well-known approaches, i.e. Pareto-optimal cut and MDC generation strategies, proposed by (Magnanti & Wong, 1981) and (Azad, Saharidis, Davoudpour, Malekly, & Yektamarian, 2013), respectively. All methods are tested on computational results and it is concluded that generating a bundle of relatively

low density cuts shows better performance than generating a single high density cut.

(Verstichel, Kinable, De Causmaecker, & Vanden Berghe, 2015) propose an exact Combinatorial Benders' decomposition to the lock scheduling problem. When an infeasible subproblem is encountered, one or more Combinatorial Benders cuts are generated and added to the master problem. The experimental results show that the proposed Combinatorial Benders' decomposition is very effective for the lock scheduling problem.

Concerning stochastic programming, (You & Grossman, 2013) present a multicut version of the Benders decomposition method for solving two-stage stochastic linear programming problems, including stochastic mixed integer programs with only continuous recourse variables. The main idea is to add one cut per realization of uncertainty to the master problem in each iteration, that is, as many Benders cuts as the number of scenarios. Computational studies show that significant savings in CPU time can be achieved by using the proposed multicut algorithm. (Oliveira, Grossman, & Hamacher, 2014) address the solution of a two-stage stochastic programming model, as well, and present the development of acceleration techniques for the stochastic Benders decomposition. The authors introduce the *dynamically updated near maximal Benders cuts* to strengthen the cuts generated and two additional acceleration techniques to improve the quality of the solution obtained from the MP by using *primal heuristics* and addressing the distance between two successive solutions (*trust region*). Their method of outperforms methods described by (Papadakos, 2008) and (Sherali & Lunday, 2013). Additionally, (Keyvanshokoo, Ryan, & Kabir, 2016) address a two-stage stochastic problem and introduce a hybrid robust-stochastic approach. To solve a combinatorial problem, an accelerated stochastic Benders decomposition algorithm is proposed. Two groups of valid inequalities are added to RMP, in order to substantially improve the lower bound, and Pareto-optimal cuts (Papadakos, 2008) are also applied to further accelerate convergence. Experiments prove the efficiency of their proposed approach.

Finally, (Vatsa & Jayaswal, 2016) introduce a novel model for the multi-period maximal covering location problem with server uncertainty and apply a modified Benders decomposition for its solution. The authors suggest several refinements to the Benders method, like heuristics for the subproblem and cut strengthening methods, which drastically reduce the computational time needed to solve problem instances.

FUTURE TRENDS

In order to further accelerate the Benders decomposition, a new strategy to explore is parallel optimization, since the solution of PSP and auxiliaries' problems (for the generation of multi-cuts) could be solved at the same time as the RMP. Moreover, one could investigate the fluctuation of the upper bound (in case of minimization) and how it is related to the convergence rate of the algorithm.

CONCLUSION

In this book chapter, the Benders decomposition algorithm is presented. The main drawbacks of this decomposition algorithm are put in focus. The chapter provides all theoretical results existing in the literature which have been developed in order to overpass these issues and accelerate the convergence of the classical Benders algorithm.

REFERENCES

- Andreas, A. K., & Smith, C. J. (2009). Decomposition Algorithm for the Design of an Nonsimultaneous Capacitated Evacuation Tree Network. *Networks*, 53(2), 91–103. doi:10.1002/net.20278
- Azad, N., Saharidis, G. K., Davoudpour, H., Malekly, H., & Yektamaran, S. (2013). Strategies for protecting supply chain networks against facility and transportation disruptions: An improved Benders decomposition approach. *Annals of Operations Research*, 210(1), 125–163. doi:10.1007/s10479-012-1146-x



- Benders, J. F. (1962). Partitioning procedures for solving mixed-variables programming problems. *Numerische Mathematik*, 4(1), 238–252. doi:10.1007/BF01386316
- Conejo, A. J., Castillo, E., Minguez, R., & Garcia-Bertrand, R. (2006). *Decomposition Techniques in Mathematical Programming: Engineering and Science Applications*. Berlin: Springer.
- Cordeau, J.-F., Pasin, F., & Solomon, M. M. (2006). An Integrated Model for Logistics Network Design. *Annals of Operations Research*, 144(1), 59–82. doi:10.1007/s10479-006-0001-3
- Cordeau, J.-F., Soumis, F., & Desrosiers, J. (2000). A Benders Decomposition Approach for the Locomotive and Car Assignment Problem. *Transportation Science*, 34(2), 133–149. doi:10.1287/trsc.34.2.133.12308
- Cote, G., & Laughton, M. A. (1984). Large-scale mixed integer programming: Benders-type heuristics. *European Journal of Operational Research*, 16(3), 327–333. doi:10.1016/0377-2217(84)90287-X
- Dantzig, G. B., & Wolfe, P. (1960). Decomposition Principle for Linear Programs. *Journal of Operations Research*, 8(1), 101–111. doi:10.1287/opre.8.1.101
- Fakhri, A., & Ghatee, M. (2013). Solution of preemptive multi-objective network design problems applying Benders decomposition method. *Annals of Operations Research*, 210(1), 295–307. doi:10.1007/s10479-013-1353-0
- Fischetti, M., Salvagnin, D., & Zanette, A. (2010). A note on the selection of Benders cuts. *Mathematical Programming*, 124(1), 175–182. doi:10.1007/s10107-010-0365-7
- Ierapetriou, M. G., & Saharidis, G. K. (2009). Resolution method for mixed integer bi-level linear problems based on decomposition technique. *Journal of Global Optimization*, 44(1), 29–51. doi:10.1007/s10898-008-9291-0
- Jenabi, M., Fatemi Ghomi, S., Torabi, S., & Hosseini, S. (2015). Acceleration strategies of Benders decomposition for the security constraints power system expansion planning. *Annals of Operations Research*, 235(1), 337–369. doi:10.1007/s10479-015-1983-5
- Keyvanshokoo, E., Ryan, S. M., & Kabir, E. (2016). Hybrid robust and stochastic optimization for closed-loop supply chain network design using accelerated Benders decomposition. *European Journal of Operational Research*, 249(1), 76–92. doi:10.1016/j.ejor.2015.08.028
- Magnanti, T., & Wong, R. (1981). Accelerating Benders decomposition algorithmic enhancement and model selection criteria. *Operations Research*, 29(3), 464–484. doi:10.1287/opre.29.3.464
- Martins de Sa, E., de Camargo, R., & de Miranda, G. (2013). An improved Benders decomposition algorithm for the tree of hubs location problem. *European Journal of Operational Research*, 226(2), 185–202. doi:10.1016/j.ejor.2012.10.051
- McDaniel, D., & Devine, M. (1977). A modified Benders partitioning algorithm for mixed integer programming. *Management Science*, 24(3), 312–319. doi:10.1287/mnsc.24.3.312
- Naoum-Sawaya, J., & Elhedhli, S. (2013). An interior-point Benders based branch-and-cut algorithm for mixed integer programs. *Annals of Operations Research*, 210(1), 33–55. doi:10.1007/s10479-010-0806-y
- Oliveira, F., Grossman, I. E., & Hamacher, S. (2014). Accelerating Benders stochastic decomposition for the optimization under uncertainty of the petroleum product supply chain. *Computers & Operations Research*, 49, 47–58. doi:10.1016/j.cor.2014.03.021
- Papadakis, N. (2008). Practical enhancements to the Magnanti-Wong method. *Operations Research Letters*, 36(4), 444–449. doi:10.1016/j.orl.2008.01.005

- Rei, W., Cordeau, J. F., Gendreau, M., & Soriano, P. (2009). Accelerating Benders decomposition by local branching. *INFORMS Journal on Computing*, 21(1), 333–345. doi:10.1287/ijoc.1080.0296
- Saharidis, G., Boile, M., & Theofanis, S. (2011). Initialization of the Benders master problem using valid inequalities applied to fixed-charge network problems. *Expert Systems with Applications*, 38(6), 6627–6636. doi:10.1016/j.eswa.2010.11.075
- Saharidis, G., & Ierapetriou, M. G. (2013). Speed-up Benders decomposition using maximum density cut (MDC) generation. *Annals of Operations Research*, 210(1), 101–123. doi:10.1007/s10479-012-1237-8
- Saharidis, G. K., & Ierapetriou, M. G. (2010). Improving Benders decomposition using maximum feasible subsystem (MFS) cut generation strategy. *Computers & Chemical Engineering*, 34(8), 1237–1245. doi:10.1016/j.compchemeng.2009.10.002
- Saharidis, G. K., Minoux, M., & Dallery, Y. (2009). Scheduling of loading and unloading of crude oil in a refinery using event-based discrete time formulation. *Computers & Chemical Engineering*, 33(8), 1413–1426. doi:10.1016/j.compchemeng.2009.02.005
- Saharidis, G. K., Minoux, M., & Ierapetriou, M. G. (2010). Accelerating Benders method using covering cut bundle generation. *International Transactions in Operational Research*, 17(2), 221–237. doi:10.1111/j.1475-3995.2009.00706.x
- Sherali, H. D., & Lunday, B. J. (2013). On generating maximal nondominated Benders cuts. *Annals of Operations Research*, 210(1), 57–72. doi:10.1007/s10479-011-0883-6
- Tang, L., Jiang, W., & Saharidis, G. K. (2013). An improved Benders decomposition algorithm for the logistics facility location problem with capacity extensions. *Annals of Operations Research*, 210(1), 165–190. doi:10.1007/s10479-011-1050-9
- Vatsa, A., & Jayaswal, S. (2016). A new formulation and Benders decomposition for the multi-period maximal covering facility location problem with server uncertainty. *European Journal of Operational Research*, 251(2), 404–418. doi:10.1016/j.ejor.2015.11.025
- Verstichel, J., Kinable, J., De Causmaecker, P., & Vanden Berghe, G. (2015). A Combinatorial Benders decomposition for the lock scheduling problem. *Computers & Operations Research*, 54, 117–128. doi:10.1016/j.cor.2014.09.007
- Watkins, D. W., McKinney, D. C., Lasdon, L. S., Nielsen, S. S., & Martin, Q. W. (2000). A scenario-based stochastic programming model for water supplies from the highland lakes. *International Transactions in Operational Research*, 7(3), 211–230. doi:10.1111/j.1475-3995.2000.tb00195.x
- You, F., & Grossman, I. E. (2013). Multicut Benders decomposition algorithm for process supply chain planning under uncertainty. *Annals of Operations Research*, 210(1), 191–211. doi:10.1007/s10479-011-0974-4
- Zhu, Y., & Kuno, T. (2003). Global Optimization of Nonconvex MINLP by a Hybrid Branch-and-Bound and Revised General Benders Decomposition Approach. *Industrial & Engineering Chemistry Research*, 42(3), 528–539. doi:10.1021/ie0200813

KEY TERMS AND DEFINITIONS

Benders Decomposition: Benders decomposition method is a classical approach for combinatorial optimization problems, based on the idea of partition and constraint generation.

Block-Decomposed Mathematical Model: A mathematical model which can be decomposed into sub-models.

Decomposition Technique: Solution method in which the main idea is to decompose the problem into sub-problems which are simpler to be solved.

Low Density Benders Cut: A cut involving a small number of decision variables of RMP therefore with low contribution to strengthening the RMP.

Multi-Generation of Cuts: The generation of more than one constraint in each iteration of Benders algorithm.

Quality of Benders Cut: A measurement of how efficient Benders cut is, affecting the Benders algorithm convergence.

Valid Inequality: An inequality which significantly restricts the solution space of a problem without excluding the optimal solution and leading to improve the convergence of a solution approach.



Lean Logistics of the Transportation of Fresh Fruit Bunches (FFB) in the Palm Oil Industry

Cheah Cheng Teik

Universiti Sains Malaysia, Malaysia

Yudi Fernando

Universiti Malaysia Pahang, Malaysia

INTRODUCTION

Variation in the production process is the cause of quality problems, (Zylstra, 2012). This is the manner of transportation of Fresh Fruit Bunches (FFB) in the palm oil industry. It can be an awesome responsibility that only gets more difficult and complicated as the estate gets larger with more hectare and the distance from mill gets farther. Despite the long distances, time zones and other hurdles involved, it is necessary for estates to reduce cost and inventory. The FFB and loose fruits must be delivered to the mill same day it is harvested and collected. Zero inventory would definitely reduce cost and increase productivity but how to go about it is an industrial problem in this important industry. The problem all the way is not just forecast accuracy but all variability, (Zylstra, 2012). Applying lean approach to distribution would compartmentalize and reduce variability such that replenishment become much easier and planning can be more effective. The ideal model is to reduce contact (damage to FFB), reduce time to consumer (deliver FFB to mill on the day of harvest) and completely eliminate as much cost as possible (efficient and effective transport of FFB).

BACKGROUND

The second top producer of palm oil after Indonesia is Malaysia. In 2014, the planted area

is 5.39 million hectares. It has increased 3.1% against 5.23 million hectares compared to year 2013. Sabah is the largest oil palm planted state with 1.51 million hectare or 28% of total oil palm planted area. In the palm oil industry, Malaysia is successful in terms of producing and marketing the palm oil, palm kernel oil and their derivative products. There are also huge amounts of palm oil wastes generated by the industry. This included oil palm shells, mesocarp fibre and empty fruit bunches from the mills, also oil palm fronds and oil palm trunks from the field during replanting.

The oil palm is five to ten times more productive in terms of oil yield than all other oil-bearing crops. It is in the best position to meet the growing global demand for oil. It is versatile and there are demands for palm oil in the non-food sector. Oil palm also contributed in the energy sector. Its biomass and biogas can be used as fuel for boilers. Palm oil is used as feedstock to produce biodiesel by blending palm diesel with petroleum diesel. This lowers the burning of fossil fuel in the world. Statistics Malaysia Palm Oil Board (MPOB).

Misguided anti-palm oil campaigns have been a talk of the town and the burning issues in business. This has been a constant challenge for the palm oil industry. During the 1980's the campaign were on health issues and Malaysia gather international independent experts to defend it.

Today the lobbyist turned their attacks to environmental and sustainability issues and MPOB, Ministry of Plantation Industries and Commodities

(MPIC) had continued to address them. Since oil palm is grown in one of the most ecologically sensitive areas of the world, it is imperative to grow it in a sustainable manner so that people, planet and profit are all part of the equation.

Bad things can happen to good companies that fail to take a broad view of accountability, (Savitz, 2012). Certifying Malaysia Sustainable Palm Oil (MSPO) and producing Sustainable Palm Oil is important for the country's business to promote export and sales to compete in the world market.

Even a well run company with good intentions and with a proud history of business and philanthropic achievement can stumble and fall when the principles of sustainability is ignored, (Savitz, 2012). The four areas in commitment to sustainability are land use, water, climate change and farmer livelihoods. As land is finite and human population is growing rapidly, it is important to ensure that the use of land is sustainable.

For market development, MPOB provides oil palm industry players with relevant economics and statistics. The important markets for our Malaysian palm oil are China, European Union, India, Pakistan, USA and West Asian countries. Currently, 1.29 million hectares of palm plantations in Malaysia are Round Table Sustainable Palm Oil (RSPO) certified and these contribute to 4.8 million tonnes of certified sustainable palm oil, (Datuk Amar Douglas Uggah Embas, Plantation Industries and Commodities Minister, 2015).

A look at the world market which concern palm oil, Malaysia must work hard to compete with an added advantage to promote sales of our palm oil and to start with encouraging MSPO and carrying out relevant research to further improve the industry for the estate, mill and refinery. A very important area which need focus is the transportation of FFB, the raw material for palm oil mills. Along with marketing come logistics aspects, which are as important to create differentiation advantages and competitiveness of retail supply chains (Lorentz and Lounela, 2011). The brand value of a company consists of intangible assets such as reputation and customer loyalty.

Oil palm is the most important agriculture crop and palm oil industry is the pillar of Malaysian agriculture and main stay of rural economy. It is an engine of growth contributing about one-third of the agriculture Gross Domestic Product (GDP). The industry provides employment to more than 610,000 people including some 200,000 small holders. (Statistics MPOB).

MPOB has developed a sustainability certification scheme Malaysia Sustainable Palm Oil (MSPO). The government has approved this on 21st March 2014 and this scheme is officially implemented as of January 2015. This is an alternative to currently available certification schemes like International Sustainability and Carbon Certification (ISCC), Roundtable on Sustainable Biomaterials (RSB), Indonesian Sustainable Palm Oil (ISPO), and Roundtable on Sustainable Palm Oil (RSPO). Innovation is pivotal to the success of the palm oil industry. As for now, there are 9 estates, 3 independent small holders and 6 mills certified MSPO. The benefits of certification are to broaden market access with the increasing demand for sustainable palm oil. It will enhance organizational image as it is seen as social responsibility. This further facilitates access to the growing "green market". The Malaysian palm oil industry strives to strike a balance between social, environmental and economic needs of country and people.

Issues and Problems

The research gap and problem statement is found no attempt done to make transport of FFB efficient and effective, integrating lean logistics with triple bottom line to address sustainability. The existing systems ignore new competitive realities that require a strong focus on such issues as product and process quality, customer satisfaction, and cross-functional integration, (Fawcett and Cooper, 1998).

Technology is a solution for the depleting agriculture land and the need to boost production because of rise in consumption. Malaysia Palm



Oil Board (MPOB) developed superior oil palm planting material with excellent oil yield potential and the breakthrough were published in highly acclaimed journal, *Nature*. Now we have the good planting material but the problem of transporting the produce (FFB) from estate to mill remain a grey area to be seriously researched.

The performance measures in five broad areas of logistics are transportation, warehousing, inventory control, order processing and logistics administration. A performance measure consist of efficiency and effectiveness, (Mentzer, Konrad, and Ponsford, 1991). The major transportation measures are labour, cost, equipment, energy and transit time. The different vehicle types and sizes and different physical sizes of goods loaded can cause loading time variances. All these factors are important for the transportation of FFB and there must be zero inventory.

A sustainable company manages its risk and maximise its opportunities by identifying key non-financial stakeholders and engaging them in matters of mutual interest, (Savitz, 2012). They are required to have strong values, great personal integrity, and a willingness to make the tough call. There has to be openness and transparency with multiple constituents.

What motivated this study is because the world population is always in need of a steady supply of vegetable oils and fats. The efficient transportation of fresh fruit bunches (FFB) to mill has become relevance to major challenges in the field. Heavy loading of FFB vehicle is seen daily and coupled with very poor condition roads is not only unhealthy but also post danger to the community. In related efforts only some short distance roads were tar to attract FFB suppliers to send FFB to mill but do not take into consideration issues of sustainability.

The physical nature of a product substantially effects almost every expect of logistics and distribution systems. It has to do with packing materials, handling, storage and transportation. The structure and cost of a distribution system of a given product are directly affected by the

products particular characteristics of volume to weight ratio, value to weight ratio, substitutability and special characteristics, (Farahani et al. 2011). The FFB is a special product and cannot be left in the field after harvesting. It must be delivered to mill on the same day.

In brief the objectives of the research is to innovatively triangulating the FFB transport process of efficiency and effectiveness and the assessment of logistics performance.

Customer service is the output of logistics because logistics involves delivering the right product to the right customer at the right place, at the right time with the right cost and quality. Where FFB is concern, the right quality is at time of harvest, the right time is on the same day of harvest and the right place is the palm oil mill. Cost would depend on the estate manger to plan his routing well since the estate is a large area with many hectare, his logistics network design must be carefully work out to effectively reduce the distance the FFB lorry need to travel to all collection centres before reaching mill.

In brief the research questions is what is the most important scope for logistics performance to be sustainable in FFB transportation.

SIGNIFICANT OF STUDY

Transportation plays a major role and is the most important element in logistics because of its considerable cost. It accounts for between 1/3 to 2/3 of total logistics cost and is about 10 to 20% of a product price. The significance of logistics has evolved from a more passive and cost minimization oriented activity to a key success factor for firm competitiveness, (Spillan, McGinnis, George, Yi 2013). It has become an integral part of a firm's strategic planning process. Researchers argued that the global manufacturing strategies alone may not be effective if not supported by successful logistics/supply-chain management strategies. Basic logistics decisions mainly deal with physical movement of the products. It must at the same

time be created from a unique set of operational and dynamic capabilities. It is the combination of both which is a source for sustainable competitive advantage, (Eisenhardt and Martin, 2000).

UNDERLYING THEORY

An important theme of performance is that of efficiency versus effectiveness. Achabal et al. (1984) argued that efficiency deals with the allocation of resources across alternative and it is about minimizing inputs given a level of output. Effectiveness is concerned with determining the strategy to maximize revenue given a level of input. This will search for the most effective use of resources assuming resources are used efficiently, (Sheth and Sisodia, 2002). Gronroos and Ojasalo, (2004) also posit that there are two sides to performance where efficiency (the cost-effective use of resources) and effectiveness (the revenue-generating capability). In the conceptual framework lean logistics to triple bottom line shows efficiency. Triple bottom line to sustainable logistics performance shows effectiveness. This shows the two sides to performance for lean logistics to sustainable logistics performance. Continuous performance measurement matrices in terms of efficiency and effectiveness are proved to be appropriate methods for continuous evaluation of lean performance, (Govindan, Azevedo, Carvalho, Cruz-Machado, 2015).

Logistics salience positively impacts logistics innovativeness and logistics service differentiation. Logistics innovativeness and logistics service differentiation positively influence logistics performance. This is in accordance to the resource based view of the firm which states that resources lead to capabilities which leads to performance. It highlighted the importance of logistics within firms and indicated that the function must be made salient throughout the firm to further capitalize on the benefits of logistics which include

enhanced logistics capabilities and their eventual impact on logistics performance. (Ralston, Grawe, Daugherty, 2013).



RESEARCH THEORETICAL FRAMEWORK

Path 1 investigates the relationship between lean logistics and sustainable logistics performance in the transport of FFB. No studies have been conducted to relate these two constructs, the work can use different measurement variables for lean logistics and sustainable logistics performance. In the study is considered logistics challenges in transport of FFB or a lean logistics construct that could be applied in the estate environment. Therefore, a new estate measurement construct for lean logistics was developed to study its impact on sustainable logistics performance. Lean is chosen because of locality, the place is very remote and also considering the knowledge standard of the staff and workers in the estate, it is only proper to start with lean implementation.

Path 2 investigates the relationship between triple bottom line and sustainable logistics performance in transport of FFB. The study is expected to show the estates can reduce cost and increase customer service by integrating with triple bottom line functions. The study want to show there are empirical links been established between the integration of triple bottom line and sustainable logistics performance as proposed in this study's conceptual model and since there are none to-date in estate literature.

Finally, Path 3 (dotted lines) investigates the indirect effect of triple bottom line on the relationship between lean logistics and sustainable logistics performance. Estates that integrate triple bottom line in transport of FFB may enhance their lean logistics and this should result in improved logistics performance and sustainability.

Figure 1.

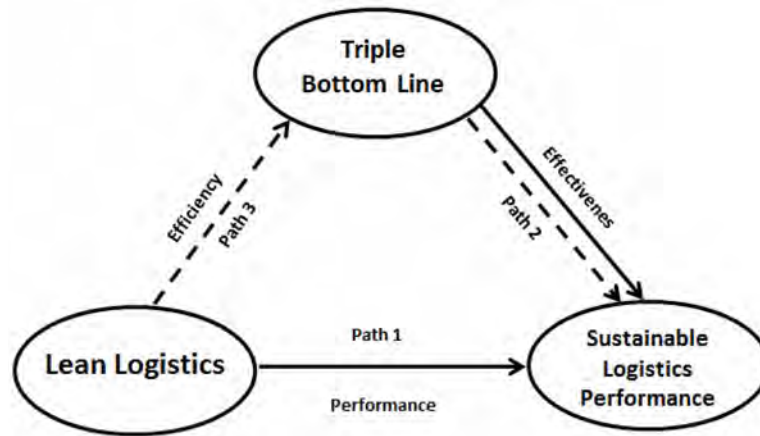
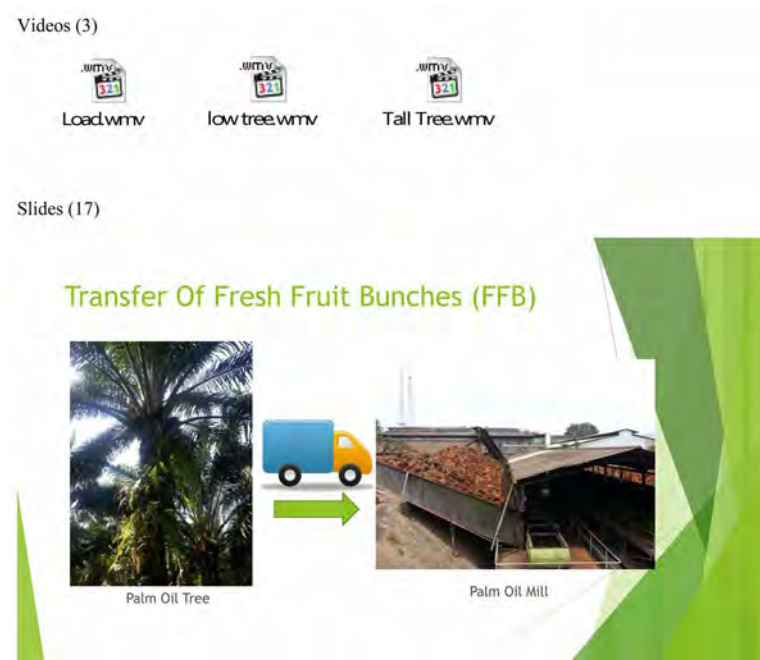


Figure 2.



THE CONCEPTUAL FRAMEWORK

Literature Review

MPOB and partners identify genetic secrets to improve palm oil yield. With the identification of the VIR gene palm growers can begin to replace their nigrescens palms with the virescens variety and can eventually increase the efficiency of har-

vest and oil yield. Such advancements to enable increased yields on existing planted areas can help increase the sustainability of palm cultivation by reducing pressure on valuable tropical forest.

The integration of corporate social responsibility and sustainability principles in logistics, (Miao, Cai and Xu 2012), pay special attention on the environmental pillar of sustainability for pollution or fuel consumption. Therefore logis-

tics transport traditionally operating on very low profit margins, can consider initiatives aimed at reduction of environmental and social impacts as a means of improving their economic performance, (Maja, Piecyk and Björklund 2015). This is the case of lorry and other vehicle transporting FFB as can be seen from the pictures at appendix.

Transportation is the absolute means to transfer products and the basic concept of lean philosophy is waste. The distribution function is responsible for successful customer service and is at the same time under pressure to reduce cost and inventory, (Zylstra, 2012).

Modern logistics has more to do than just management of transport and storage of physical goods. They have been extended to the coordination of all phases identified in the course of supply, production and sale of a company and its relations with the rest of the environment in which it operates, (Pulina and Timpanaro, 2012).

Policymakers has globally recognized that the logistics sector is a key pillar for development, (Arvis, Saslavsky, Ojala, Shepherd, Busch, Raj, 2014). In economic growth and competitiveness it is therefore necessary to improve on logistics performance and to make it sustainable. Logistics performance refers to cost, time, and complexity in accomplishing import and export activities and is statistically significantly related to the volume of bilateral trade. (Hausman, Lee, Subramanian 2013).

We measure everything move or not move because it still cost us to deal with it, (Bowersox et al. 1995). The palm estates cannot afford the FFB to seat. The FFB and loose fruits must be collected and send to mill the same day of harvest. The cross-functional nature of logistics makes it vital to every corporate strategy, particularly to actions and policies aimed at ensuring environmental and social sustainability of operations, (Maja et. al. 2015). Logistical capabilities at the company level play a key role in integrating global supply chains, (Wiengarten, Pagell, Ahmed, Gimenez 2014). In the palm estates there is a need to innovate

efficiency and effectiveness for transportation of FFB and to address sustainability.

The TBL dimensions are commonly known as the 3Ps which is profit, planet and people. It is necessary to define and measure the sales growth, cost efficiency, profitability, high yields, flexible for change in economics to address sustainability and logistics performance. Logistics plays a key role for industries and reveals a critical function designed and managed to pursue economic goals. (Digiesi, Mossa, Mummolo 2012).

Logistics service providers pursued environmental objectives by involving customers in their operations to perform logistics activities and achieve environmental performance, (Lun, Lai, Wong, Cheng 2015). It is in the logistic operations where most organizations can and do implement green supply chain strategies, (Byrne, Ryan, Heavey 2013).

Environmental performance of logistics is often consistent with the bottom-line impacts. LSPs traditionally operating on very low profit margins, can consider initiatives aimed at reduction of environmental and social impacts as a means of improving their economic performance, (Maja et. al. 2015),

Quality and production logistics have traditionally been investigated using independent approaches. Contributions have focused on their mutual inter-relation showing that benefits can be achieved from an integrated analysis. (Cagliano, Mustafa, Rafele, Zenezini 2014).

Awareness of integration's competitive potential is insufficient to mobilize resources and mitigate resistance to collaboration. Commitment is a super ordinate enabler, (Henry Jin, Fawcett, Fawcett 2013). Integrating sustainability into firms requires action that exceeds organizational boundaries. (Seuring, Gold 2013).

A determinant of logistics performance is the supply chain alignment. Leadership has partial mediation effect between supply chain alignment and human performance, (Dubey, Singh, Gupta 2015). The demands and expectations of



stakeholders must be an input for implementing sustainable practices, (Golicic and Smith, 2013).

Peter Oberhofer and Maria Dieplinger (2014), defined sustainability in a broad sense including social, ecological and economic aspects aiming at a viable integration of a company in the long run.

The dimensions of sustainability are economic prosperity, environmental stewardship and social responsibility or profit, planet and people, (Willard, B. 2012). In practicing sustainability, companies can increase shareholder value by properly managing risk, anticipating regulatory action and processing new markets, (Savitz, A. 2012). To meet today's and tomorrow's societal sustainability challenges, (Quist and Tukker 2013), pointed out that sustainable innovation, learning and collaboration are inherently interlinked and therefore all must be integrated with a long term focus upon sustainable societal transformation which will require behavioural, cultural and structural changes to society. A fully sustainable supply chain will develop a practical approach towards sustainability supply chain performance measurement and management, the vein of improving supply chains and increasing recycling, (Schaltegger, Burritt 2014). Environmental sustainability has become a concern that develop initiatives towards environmental sustainability with a focus on companies involved in logistics and transportation processes, (Colicchia, Marchet, Melacini, Perotti 2013).

SOLUTIONS AND RECOMMENDATIONS

The reasons are discussed as to why it is important for lean logistics to integrate with triple bottom line and address sustainability. Based on a review of the site visits, transportation of FFB from estate to mill is an area of logistics performance where lean must be implemented. Some recommendations for the successful implementation of lean in the logistics performance of value chains are provided. Such practice will increase OER (Oil

Extraction Rate) and align with MSPO certification to enhance the country's CPO sales.

IMPLICATIONS ON FUTURE RESEARCH DIRECTIONS

Logistics Industry Trends in the years to come is reckoned to strengthen further by factors facilitating the effective functioning of the industry itself, integration, confederation, technology, legislation and globalization. Ehrhart (2013), has argued that the future logistics model company should change towards sustainable innovations which would create business opportunities. Much more search is required for a better insight into the existing strengths of logistics clusters.

No identified study has explored lean logistics and triple bottom line on sustainable logistics performance for transportation of FFB in the palm oil industry. This paper is the first of its kind which determines the state of sustainability within the field of logistics performance and identifies areas and sets the agenda for future research in this field. This paper takes a novel approach to look at the "Lean logistics and Triple Bottom Line" from the Sustainable Logistics Performance perspective by proposing a preliminary research model based on existing site conditions and empirical data. Future research direction to better the transportation must take into consideration locality.

CONCLUSION

To survive today's businesses, it must be steered through a transition from an old and dangerously dysfunctional model to a far better one that will operate in harmony with nature that strive in a carbon contained world and put down the threats of global climate disruption, species extinction, resource depletion and environmental degradation. Required to develop business model that is sustainable, (Ray Anderson, confessions of a radical industrialist). The transportation of Fresh Fruit



Bunches (FFB) in the palm oil industry required this transition to become modern in implementing lean logistics and integrating with triple bottom line to address sustainability.

The contributions of this paper include

1. Innovatively triangulating the FFB transport process of efficiency and effectiveness and the assessment of logistics performance.
2. Implementing lean logistics to investigate factors that affect logistics performance
3. Integrating with triple bottom line to address sustainable decision making.
4. As a result, management could use the managerial decision-making matrix to improve strategies.

REFERENCES

- Achabal, D. D. (1984). Issues and perspectives on retail productivity. *Journal of Retailing*, 60(3), 107–127.
- Anderson & White. (2009). *Confessions of a Radical Industrialist: How My Company and I Transformed Our Purpose*. Sparked Innovation, and Grew Profits - By Respecting the Earth.
- Arvis, Saslavsky, Ojala, Shepherd, Busch, & Rai. (2014). *Connecting to Compete 2014: Trade Logistics in the Global Economy--The Logistics Performance Index and Its Indicators*. Academic Press.
- Baron & Kenny. (1986). The moderator-mediator variable distinction in social psychological research. *Journal of Personality and Social Psychology*, 51(5), 1173–1182. PMID:3806354
- Bowersox, D. J., Calantone, R. J., Clinton, S. R., Closs, D. J., Cooper, M. B., Droge, C. L., & Schmitz, J. M. et al. (1995). *World Class Logistics: The Challenge of Managing Continuous Change*. Oak Brook, IL: Council of Logistics Management.
- Byrne, Ryan, & Heavey. (2013). Sustainable logistics: A literature review and exploratory study of Irish based manufacturing organizations. *Int. J. Eng. Technol. Innov*, 3(3), 200–213.
- Cagliano, Mustafa, Rafele, & Zenezini. (2014). Logistics Performance Measurement for Sustainability in the Fast Fashion Industry. *Next Generation Supply Chains: Trends and Opportunities*, 113.
- Colicchia, M., Marchet, G., Melacini, M., & Perotti, S. (2013). Building environmental sustainability: Empirical evidence from Logistics Service Providers. *Journal of Cleaner Production*, 59, 197–209. doi:10.1016/j.jclepro.2013.06.057
- Digiesi, M., Mossa, G., & Mummolo, G. (2012). A loss factor based approach for sustainable logistics. *Production Planning and Control*, 23(2-3), 160–170. doi:10.1080/09537287.2011.591636
- Dubey, S., Singh, T., & Gupta, O. K. (2015). Impact of Agility, Adaptability and Alignment on Humanitarian Logistics Performance: Mediating Effect of Leadership. *Global Business Review*, 16(5), 1–20. doi:10.1177/0972150915591463
- Ehrhart. (2013). *Delivering Tomorrow, Towards Sustainable Logistics. How Business Innovation and Green Demand drive a Carbon-Efficient Industry*. Academic Press.
- Eisenhardt, & Martin, (2000), Dynamic capabilities: what are they?. *Strategic Management Journal*, 21(10-11), 1105-1121.
- Esmemr, S., Ceti, I. B., & Tuna, O. (2010). A simulation for optimum terminal truck number in a turkish port based on lean and green concept. *The Asian Journal of Shipping and Logistics*, 26(2), 277–296. doi:10.1016/S2092-5212(10)80006-9
- Farahani, R. Z., Rezapour, S., & Kardar, L. (2011). *Logistics operations and management: concepts and models*. Elsevier.

- Fawcett, S. E., & Cooper, M. B. (1998). Logistics Performance Measurement and Customer Success. *Industrial Marketing Management*, 27(4), 341–357. doi:10.1016/S0019-8501(97)00078-3
- Golicic, & Smith. (2013). A Meta-Analysis of Environmentally Sustainable Supply Chain Management Practices and Firm Performance. *Journal of Supply Chain Management*, 49(2), 78–95.
- Govindan, A., Azevedo, S. G., Carvalho, H., & Cruz-Machado, V. (2015). Lean, green and resilient practices influence on supply chain performance: Interpretive structural modeling approach. *International Journal of Environmental Science and Technology*, 12(1), 15–34. doi:10.1007/s13762-013-0409-7
- Gronroos & Ojasalo. (2004). *Service productivity, towards a conceptualization of the formation of inputs into economic results in services*. Academic Press.
- Hall, T. J. (2011). The triple bottom line: What is it and how does it work? *Indiana Business Review*, 86(1), 4.
- Hausman, L., Lee, H. L., & Subramanian, U. (2013). The impact of logistics performance on trade. *Production and Operations Management*, 22(2), 236–252. doi:10.1111/j.1937-5956.2011.01312.x
- Holmbeck. (1997). Toward terminological, conceptual, and statistical clarity in the study of mediators and moderators: Examples from the child-clinical and pediatric psychology literatures. *Journal of Consulting and Clinical Psychology*, 65(4), 599–610.
- Jin, H. (2013). Awareness is not enough: Commitment and performance implications of supply chain integration. *International Journal of Physical Distribution & Logistics Management*, 43(3), 205–230. doi:10.1108/IJPDLM-10-2011-0169
- Lorentz, H., & Lounela, J. (2011). Retailer supply chain capability assessment in Russia. *International Journal of Retail & Distribution Management*, 39(9), 682–701. doi:10.1108/09590551111159341
- Lun, L. (2015). Greening propensity and performance implications for logistics service providers. *Transportation Research Part E, Logistics and Transportation Review*, 74, 50–62. doi:10.1016/j.tre.2014.10.002
- Maja, I. (2015). Logistics service providers and corporate social responsibility: Sustainability reporting in the logistics industry. *International Journal of Physical Distribution & Logistics Management*, 45(5), 459–485. doi:10.1108/IJPDLM-08-2013-0228
- Mentzer, J., Konrad, T., & Ponsford, B. (1991). An Efficiency/Effectiveness Approach to Logistics Performance Analysis. *Journal of Business Logistics*, 12(1), 33.
- Miao, C., Cai, S., & Xu, D. (2012). Exploring the Antecedents of Logistics Social Responsibility: A Focus on Chinese Firms. *International Journal of Production Economics*, 140(1), 18–27. doi:10.1016/j.ijpe.2011.05.030
- Oberhofer, P., & Dieplinger, M. (2014). Sustainability in the Transport and Logistics Sector: Lacking Environmental Measures. *Business Strategy and the Environment*, 23(4), 236–253. doi:10.1002/bse.1769
- Pulina, P., & Timpanaro, G. (2012). Ethics, Sustainability and Logistics in Agricultural and Agrifood Economics Research. *Italian Journal of Agronomy*, 7(3), e33. doi:10.4081/ija.2012.e33
- Quist, J., & Tukker, A. (2013). Knowledge collaboration and learning for sustainable innovation and consumption: Introduction to the ERSCP portion of this special volume. *Journal of Cleaner Production*, 48, 167–175. doi:10.1016/j.jclepro.2013.03.051



Ralston, G., Grawe, S. J., & Daugherty, P. J. (2013). Logistics salience impact on logistics capabilities and performance. *The International Journal of Logistics Management*, 24(2), 136–152. doi:10.1108/IJLM-10-2012-0113

Savitz, A. (2012). *The triple bottom line: How today's best-run companies are achieving economic, social and environmental success--and how you can too*. John Wiley & Sons.

Schaltegger & Burritt. (2014). Measuring and managing sustainability performance of supply chains: Review and sustainability supply chain management framework. *Supply Chain Management: An International Journal*, 19(3), 232–241.

Seuring & Gold. (2013). Sustainability management beyond corporate boundaries: from stakeholders to performance. *Journal of Cleaner Production*, 56, 1–6.

Sheth & Sisodia. (2002). *Does marketing need reform, fresh perspectives on the future*. Academic Press.

Spillian, McGinnis, & Yi. (2013). A comparison of the effect of logistic strategy and logistics integration on firm competitiveness in the USA and China. *The International Journal of Logistics Management*, 24(2), 153 – 179.

Wiengarten, P., Pagell, M., Ahmed, M. U., & Gimenez, C. (2014). Do a countrys logistical capabilities moderate the external integration performance relationship? *Journal of Operations Management*, 32(1), 51–63. doi:10.1016/j.jom.2013.07.001

Willard, B. (2012). *The new sustainability advantage: seven business case benefits of a triple bottom line*. New Society Publishers.

Zylstra, K. D. (2012). *Lean distribution: applying lean manufacturing to distribution, logistics, and supply chain*. John Wiley & Sons. doi:10.1002/9781119201427

ADDITIONAL READING

Current, J., & Marsh, M. (1993). Multiobjective transportation network design and routing problems: Taxonomy and annotation. *European Journal of Operational Research*, 65(1), 4–19. doi:10.1016/0377-2217(93)90140-I

McCarthy, P. S. (2001). *Transportation economics: theory and practice: a case study approach*. Wiley-Blackwell.

Nagurney, A. (2000). *Sustainable transportation networks*. Books.

Tzeng, G. H., & Shiao, T. A. (1987). Energy conservation strategies in urban transportation. *Energy Syst. Policy*; (United States), 11(1).

KEY TERMS AND DEFINITIONS

Effectiveness: It is the degree to which objectives are achieved and the extent to which targeted problems are solved and is determined without reference to costs. Effectiveness means “doing the right thing” to address sustainability.

Efficiency: It is a measurable concept that can be determined by determining the ratio of useful output to total input. It minimizes the waste of resources such as loose fruits collection, energy and time, while successfully achieving the desired output. Efficiency means “doing things right”.

Fresh Fruit Bunches (FFB): It is the raw material for palm oil mills. The oil palm tree (*Elaeis guineensis*) originates from West Africa and was developed into an agricultural crop in Malaysia. The fruit from the tree is processed and has two main products crude palm oil and palm kernel.

Lean: This is zero inventory on FFB and loose fruits. The core idea is to maximize FFB value while minimizing waste. It means creating more value for mills with fewer resources.

Logistics: This is the routing network for transportation of FFB in the estates. Logistics is the management of the flow of FFB between the estate and the mill in order to meet.

OER: This is the Oil Extraction Rate. The extraction efficiency is affected by the percentage of oil content in FFB at the time of harvest. It is the amount of oil the mill can extract from it's process to the 100 tonne of the FFB being processed.

Performance: This is the efficient and effective means to address sustainability. The accomplishment of sustainable logistics performance measured against preset known standards of lean logistics and triple bottom line.

Sustainable: In transporting FFB it must have minimal negative impact on the global or local environment, community, society, and economy. It must have green and triple bottom line concept.

Transportation: This is the absolute means to transfer products of FFB. The act of taking the FFB from estate to mill through land and water.

Triple Bottom Line: The dimensions are profit, planet and people. 3BL is an accounting framework with three parts: social, environmental (or ecological) and financial. Lean logistics is integrated with 3BL in a framework to evaluate their performance in a broader perspective to create greater business value in addressing sustainability.

Major Techniques and Current Developments of Supply Chain Process Modelling



Henry Xu

The University of Queensland, Australia

Rena Agrey

The University of Queensland, Australia

INTRODUCTION

While there are many definitions of supply chains, a supply chain can be defined as a set of inter-linked business processes that add value from the perspective of end customers in terms of products and/or services (Christopher, 2011). Value to the customer includes good quality, a fair price and fast and accurate (i.e., correct items in full quantity on time) delivery (Russell & Taylor, 2014). This means that a supply chain can be regarded as an overarching business process which covers a full range of activities from raw material supply, manufacturing of components and finished products, to distribution of the finished products to end customers (Russell & Taylor, 2014). In other words, a supply chain is an extended process covering all activities from the point of origin to the point of consumption. Supply chain process modelling (SCPM) is essential for understanding, design or reengineering of a supply chain (Borstorff & Rosenbaum, 2012). Meanwhile, SCPM presents many challenges due to a large number of companies involved in the supply chain and the complexity in communication and interactions between them (Brown, Recker & West, 2011).

In this chapter, we first offer some background information on supply chain processes and SCPM. Then this is followed by critical discussion of the frequently used standards/techniques for SCPM. Finally, the chapter is concluded with discussion of future trends and research directions.

BACKGROUND

Supply chain management (SCM) has received increasing attention in the literature since the early 80's (Ellram & Cooper, 2014). While there are many definitions of SCM, it can be defined as the integration of key business processes covering a wide range of activities from the original supplier to the end customer that add value to stakeholders in terms of products, services and information (Cooper, Lambert, & Pagh, 1997; Mentzer et al., 2001). Many researchers support this process view of the supply chain and argue that it helps to reduce supply chain costs and enhance customer satisfaction (Trkman, Stemberger, Jaklic, & Groznik, 2007).

There are a number of models or frameworks for a holistic understanding of supply chain processes (Ellram & Cooper, 2014). The supply chain operations reference (SCOR) model developed by the Supply Chain Council (SCC) identifies six core processes in a supply chain: plan, source, make, deliver, return and enable (SCC, 2012). Meanwhile, the Global Supply Chain Forum (GSCF) pinpoints eight key business processes in a supply chain that work at both the strategic and operational levels for the improvement of internal and external results (Croxtton, Garcia-Dastugue, Lambert, & Rogers, 2001). Recently, Xu, Koh, and Parker (2009) propose a research framework that consists of seven business processes for manufacturing coordination in a complex supply network.

DOI: 10.4018/978-1-5225-2255-3.ch472

Though these supply chain models or frameworks have different purposes and scopes, the business processes proposed in these models or frameworks need to be analysed and implemented in practice through modelling.

Clearly, SCPM is closely related to business process modelling (BPM) as a supply chain process can be viewed an extended business process. BPM is often used as a platform for achieving a common understanding of a business process (Aguilar-Saven, 2004). It has been broadly recognised that good process models are essential for business process re-engineering (BPR) (or improvement) or for information system development. Generally, BPM can be regarded as a methodology which employs software systems to analyse, design and improve business processes so that a company's business performance in terms of productivity and profits can be improved (Bae & Seo, 2007; Trkman et al., 2007).

In the body of literature on BPM, it is a common view that depending on the purpose of modelling, one should choose an appropriate technique for a particular modelling task (Aguilar-Saven, 2004). This is because there are many tools and techniques currently available in the market from basic tools (e.g. flowcharting) to more complex techniques (e.g., Petri Nets) (Recker, Rosemann, Indulska, & Green, 2009). Generally, the techniques commonly used for BPM can be categorised into two classes: (1) descriptive (or pragmatic) approaches (e.g., the SCOR model), and (2) formal (or rigorous) approaches (e.g., simulation) (Aguilar-Saven, 2004; Ryan & Heavey, 2006). As a supply chain usually consists of a number of business processes, many techniques for BPM can be used for SCPM.

Though BPM can be regarded as the predecessor of SCPM, it poses more challenges than traditional BPM. This is because a supply chain is often a boundary-spanning process that includes at least a company's immediate suppliers and customers. Therefore, it is important to have a holistic view of the overall supply chain process, including inbound logistics, internal operations and outbound logistics (Min & Zhou, 2002). The

SCOR model is best known for modelling an overall supply chain in a structured and hierarchical way (Kasi, 2005).

MAIN FOCUS

Currently there are a wide variety of standards and techniques for SCPM. Leading standards such as SCOR and unified modelling language (UML) have been well recognized and adopted by researchers, practitioners and system developers for BPM and/or SCPM. In this chapter, we focus on those techniques that have been most frequently adopted for SCPM. Specifically, we will focus on four descriptive approaches (i.e. SCOR, IDEF, UML and BPMN) and two formal approaches (i.e. Petri Nets and simulation) for SCPM.

Descriptive Approaches

Supply Chain Operations Reference Model (SCOR)

SCC released the SCOR 1.0 model in 1996 that can be used to describe supply chain operational processes for the purposes of (re)-design, (re)-engineering or improvement. The SCOR model is broadly recognised as the first cross-industry framework that can be used to measure, evaluate and improve supply chain performance. As a reference model, SCOR provides a standardized terminology and a glossary of indicators chain members can use to achieve a common understanding of supply chain processes and to evaluate their performance (Persson, 2011). By comparing the current versus the future state of a process, SCOR empowers BPR. It also allows a company to benchmark their processes against best performers' processes in the industry and to apply best practices for specific business processes. In summary, the SCOR model is built on the three important mechanisms for improving business performance: reengineering, benchmarking and best practices (Lambert, Garcia-Dastugue, & Croxton, 2005).

Table 1. Three levels of process details in the SCOR model (source: SCC, 2012)

	Level		Examples	Comments
	#	Description		
Within scope of SCOR	1	Process Types (Scope)	Plan, Source, Make, Deliver, Return and Enable	Level-1 defines scope and content of a supply chain. At level-1 the basis-of-competition performance targets for a supply chain are set.
	2	Process Categories (Configuration)	Make-to-Stock, Make-to- Order, Engineer-to-Order Defective Products, MRO Products, Excess Products	Level-2 defines the operations strategy. At level-2 the process capabilities for a supply chain are set. (Make-to-Stock, Make-to-Order)
	3	Process Elements (Steps)	- Schedule Deliveries - Receive Product - Verify Product - Transfer Product - Authorize Payment	Level-3 defines the configuration of individual processes. At level-3 the ability to execute is set. At level-3 the focus is on the right: - Processes - Inputs and Outputs - Process performance - Practices - Technology capabilities - Skills of staff
Not in scope	4	Activities (Implementation)	Industry-, company-, location- and/or technology specific steps	Level-4 describes the activities performed within the supply chain. Companies implement industry-, company-, and/or location-specific processes and practices to achieve required performance

SCOR classifies a supply chain process into four hierarchical levels: process types, process categories, process elements and implementations as illustrated in Figure 1. Note that the implementation level is outside of the current scope of SCOR and thus is not covered by it. Level 1

defines the business scope and sets performance targets through the use of six basic processes: plan, source, make, deliver, return and enable (as shown in Figure 1). Level 2 decomposes the first five basic processes into 26 process categories, and each is labelled as planning, or execution or

Figure 1. Six major management processes in the SCOR model (source: SCC, 2012)



enabling. By choosing from these core process categories, a company can configure its own SCOR model. Level 3 identifies the specific process elements such as scheduling and production within a given configuration.

In brief, SCOR enables detailed analysis of a company's processes at three levels: Level 1 defines the business scope; Level 2 configures the supply chain; and Level 3 provides the detailed process elements and performance attributes (Li, Su, & Chen, 2011). The latest 11th version of the SCOR model provides greater guidance to supply chain managers with regards to the enabling processes, best practices and cost metrics such that they better manage data, assets, networks and performance (SCC, 2012). Thus, SCOR is considered by many practitioners and researchers as a 'supply chain excellence' approach for enhancing a company's competitive advantage (Bolstorff & Rosenbaum, 2012).

Nevertheless, SCOR has a number of limitations. First, the *return* process cannot be applied to the service industry directly as return of a service does not make business sense (Barnard & Mollaghasemi, 2006). Second, though SCOR is an excellent tool for assessing supply chain performance, it does not have the capability for solving problems (Husby & Swartwood, 2011). Third, managers tend to adopt SCOR at the high level but fail to implement it at a more detailed level (Husby & Swartwood, 2011). Finally, SCOR poses some special challenges with respect to its implementation in developing countries due to such constraints as shortage of qualified employees and inexperience of people in using new or sophisticated technologies (Georgise, Thoben, & Seifert, 2011).

Unified Modelling Language (UML)

Unified modelling language (UML) was used as a standard by the Object Management Group (OMG) in 1997, and has been managed by OMG since then. The current version of UML (i.e., v2.5) was officially released by OMG in June 2015. The

popularity of UML is evidenced by its international status as the ISO/IEC 19501:2005 standard (ISO, 2016). In practice, UML has become one of the most frequently adopted standards for modelling object-oriented information systems in the software industry (Xu, 2010).

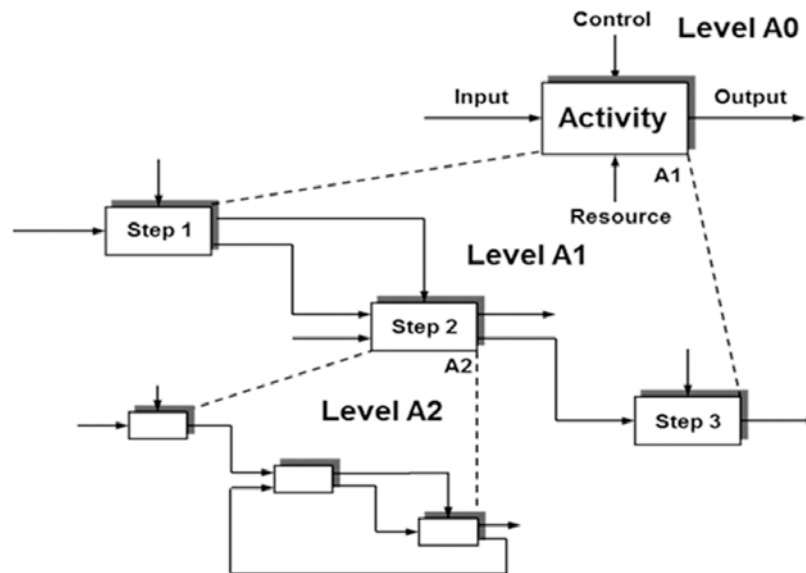
UML's notations are derived from three object-oriented methods: object modelling technique (OMT), object-oriented software engineering (OOSE), and Booch method. A key feature of UML is that it enables multiple views (e.g., static vs. dynamic view) of the process or system that is modelled (Stefanovic & Stefanovic, 2008). UML v2.5 defines 14 official diagrams which include activity diagram, object diagram, use case diagram, sequence diagram, etc. These diagrams can be classified into two categories: structural (i.e., the components of a system), and behavioural (i.e., how a system evolves over time or reacts to changes) (Stefanovic & Stefanovic, 2008). For instance, software developers often use sequence diagram to present the interactions between enterprise information systems and users in a time sequence (Xu, 2010).

In summary, UML can be used to specify, visualize and document business processes for BPR or information system development (Rumbaugh, Jacobson, & Booch, 1999). It has achieved success in modelling large and complex systems by using an engineering-based approach (Aguilar-Saven, 2004). Nevertheless, UML has certain weaknesses. Ryu and Yucsan (2007) suggest that UML lacks the ability to describe dynamic and collaborative processes where multiple participants are involved. Another shortcoming of UML is lack of formality needed for simulation and quantitative analysis (Baresi, 2001). To counter these weaknesses, UML has been used with other methods such as IDEF and the more formal approach of Petri Nets (Ruy & Yucsan, 2007).

The IDEF Family

IDEF was originally from the integrated computer-aided manufacturing (ICAM) project which was

Figure 2. Hierarchical structure of IDEF0 modelling approach



carried out by the US Air Force in the mid-1970s (Aguilar-Saven, 2004). IDEF denotes 16 modelling tools from IDEF0 to IDEF14 plus IDEF1X that can be used for system analysis and process design. IDEF0 defines the activities and their requirements in a system or organization. In alignment with the top-down design approach, IDEF0 uses a hierarchical approach for functional decomposition as illustrated in Figure 2.

The model starts by representing the whole system as a single box at the highest level labelled A0. The A0 box can be decomposed into a set of activities at level A1. This process repeats until the system is described in the necessary detail. In this hierarchical structure, each activity (or process step) is represented as a box. Inputs are information and/or materials which are transformed by the activity to outputs. Controls are the information and/or materials used to constrain the activity. Resources are people, facilities or systems that perform or provide the energy to the activity. From the user's perspective, IDEF0 uses simple rules and visual vocabulary for describing a system or business processes, making it easy to learn and visually appealing.

Among these 15 notations (IDEF0 – IDEF14 plus IDEF1X), IDEF0 is the most popular one on the market for a number of reasons. First, IDEF0 is visually appealing to users with its easy-to-understand vocabulary for description of a business process or system. Second, IDEF0 has simple yet strict rules making it easy to learn and understand. Furthermore, IDEF0's functional decomposition provides process views at both macro and micro levels of an organization. This makes it suitable for analysis of business processes in a supply chain system (Xu, 2015). Meanwhile, IDEF3 is also popular for BPM or SCPM as it provides the basis for process modelling. Nonetheless, IDEF0 and IDEF3 have different focuses on BPM. IDEF0, often called 'functional modelling', defines the activities that need to be conducted in a process, while IDEF3, known as 'process modelling', depicts how a process works (Ryu & Yucesan, 2007).

Nonetheless, both IDEF0 and IDEF3 have limitations. For example, people tend to interpret IDEF0 models as a series of activities since they may be presented as "boxes" in a left-to-right sequence (Aguilar-Saven, 2004). IDEF3 is ap-

propriate for delineating a static process but has limitations for describing a dynamic process (Ryu & Yucesan, 2007). Also, IDEF3 has limitations in terms of tracking all participants involved in a collaboration process. This limitation of IDEF3 renders it unsuitable for modelling collaborative business processes in a supply chain context where chain members interact dynamically with each other over time (Ryu & Yucesan, 2007).

Business Process Modelling and Notation (BPMN)

Business process management initiative (BPMI) released the first version of business process modelling and notation (i.e., BPMN v1.0) in 2004. Since BPMI and Object Management Group (OMG) merged in 2005, BPMN has been maintained by OMG. Hence BPMN is the latest member in the family of BPM tools. BPMN aims to provide a notation that is readily understandable by the business analysts creating the initial drafts, the technical developers responsible for implementing those processes, and all the business users managing and monitoring the processes (White, 2006). To this end, BPMN constructs are differentiated into two sets: a set of core graphical elements and an extended set of specialised elements (Recker, 2011). The core set was designed as a basic notation that is used to describe the essence of business processes, while the extended set supports complex process scenarios such as workflow specification (Recker, 2010).

BPMN has evolved significantly since its inception in 2004 as a graphical enrichment to the formal approach of UML (Dumas, La Rosa, Mendling, Reijers, 2013). As the latest version of BPMN, v2.0 was released by OMG in 2011. Enhancement of BPMN's functionality has been focused on quality (Moreno-Montes de Oca, Snoeck, Reijers, & Rodriguez-Morffi, 2015), process tailoring (Pillat, Oliveira, Alencar, & Cowan, 2015) and time (Arevalo, Escalona, Ramos, & Dominguez-Munoz, 2016). In its current version, BPMN defines five types of

modelling elements: flow objects, connecting objects, data objects, swimlanes and artifacts (OMG, 2011). Flow objects (e.g., activities) are the basic elements for creating BPMN models. Connecting objects are used for connecting flow objects through various types of arrows. Data objects show how data or documents are used and updated during the process flow. Swimlanes are used to group activities into categories for differentiating functional responsibilities. Artifacts can be added to a BPMN model to illustrate related information such as processed data. For further information on BPMN v2.0, reader can refer to OMG (2011).

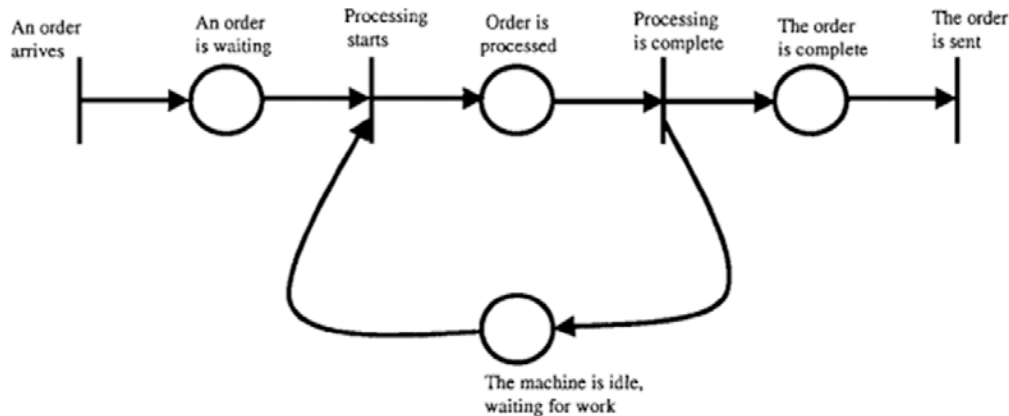
Since its first release, BPMN has quickly attracted attention from practitioners and academics (Recker, 2010). To a large extent, BPMN can be considered as the de facto standard for BPM (Recker, 2010; Yousfi, Bauer, Saidi, & Dey, 2016). Nonetheless, BPMN has a number of deficits. Particularly, BPMN does not provide sufficient support for users to specify business rules in their BPMN models (Recker, 2010). Similarly, BPMN has inadequate support for capturing process decomposition (Recker, 2011). This deficiency makes BPMN unsuitable for modelling of complex supply networks where process decomposition is essential for analysis of supply chain processes at different levels.

Formal Approaches

Petri Nets

Petri Nets was named after Dr. Carl Adam Petri who was credited as the developer of this modelling technique in the 1960s (Wang, 2007). Its main strength lies in its ability of combining graphical representations of system dynamics with mathematical analysis (David & Alla, 1994). Petri Nets can be used to describe and analyse concurrent systems where many entities (e.g. people) share the same resource (e.g. aeroplane). A classic Petri Net comprises two types of nodes as shown in Figure 4:

Figure 3. An example of Petri Net (Peterson, 1981)



1. 'Places' (graphically represented by circles,
2. And 'transitions' (graphically represented by bars).

Places (circles) and transitions (bars) are connected by 'input arcs' (from a place to a transition) and 'output arcs' (from a transition to a place). Figure 3 illustrates an example of Petri Nets.

According to the literature, a classic Petri Net has some drawbacks. First, it lacks data concepts which often results in oversized models. Second, a classic Petri Net lacks hierarchy concepts, which makes it unsuitable for building a large model (Aguilar-Saven, 2004). Therefore, a classic Petri Nets is unable to model complex, high-level business processes (Leymann & Altenhuber, 1994). As a result, classic Petri Nets have been extended in three aspects (i.e., time, colour and hierarchy) to enhance their capabilities. These extensions are often referred to as high-level Petri Nets (van der Aalst & van Hee, 1996).

To overcome the drawbacks of classic Petri Nets as mentioned above, coloured Petri Nets (CPNs) incorporates both data structuring and hierarchical decomposition into classic Petri Nets without compromising qualities of the original Petri Nets (Ramadan, Elmongui, & Hassan, 2011). These characteristics of CPNs make them suitable for modelling large complex systems such as supply chain networks that have a number of business processes (Aguilar-Saven,

2004; Landeghem & Bobeanu, 2002). Among the various versions of Petri Nets, CPNs are most well-known to researchers in the field of BPM or SCPM (Ramadan et al., 2011). Nevertheless, CPNs still have limitations in modelling collaborative business processes. Though coloured tokens can be used to differentiate various participants, the complexity of a model will increase significantly when a new participant is added to the existing model (Ryu & Yucesan, 2007).

Simulation

Like UML, simulation is another engineering-based technique for modelling. It is one of the most popular modelling techniques in operation management (Pannirselvam, Ferguson, Ash, & Siferd, 1999). There are two reasons for the popularity of simulation. First, analytical approaches (e.g., mathematical equations) tend to be inadequate in modelling complex systems (e.g., manufacturing systems) or business processes (e.g., supply networks). Second, complex systems such as flexible manufacturing systems need to be modelled and evaluated early at the design stage such that design parameters (e.g., the number of pallets) can be optimised for a given design scenario. Often simulation is used for evaluation of a complex system with low costs prior to the actual implementation of the system.

As can be seen above, the main purpose of simulation is to obtain knowledge about the behaviour and performance of a real-world system in a cost-efficient way (Maria, 1997). By using the simulation results of a system, it is expected that management can make informed decisions with respect to design or operation of the system under assessment. To achieve this, it is essential to build a simulation model that has an adequate level of similarity to the real-world system such that the simulation results can be trusted as true representations of the real system (Giaglis, 2001).

There are various forms of simulation such as discrete-event simulation (DES), system dynamics, Monte Carlo simulation, etc. Recent studies indicate that DES is still the most popular type of simulation in practical applications (Jahangirian, Eldabi, Naseer, Stergioulas, & Young, 2010). The main purpose of DES is to understand the behaviour of a system or evaluate different design or execution strategies of the system with the constraints imposed by a set of criteria. To this end, it is essential to design a model of the system under assessment and experiment with the built model (Shannon, 1975). As supply chain processes are usually treated as discrete-event systems, DES is one of the popular tools for SCPM (Ryan & Heavey, 2006). Nevertheless, DES has weakness in stakeholder engagement mainly due to the time and efforts required for data collection (Jahangirian et al., 2010).

Compared with analytical modelling approaches such as queuing theory, simulation is much more capable in terms of mapping complex business systems (Thierry, Bel, & Thomas, 2010). In some circumstances, simulation is the only practical modelling approach that can be used to study the behaviour of real complex systems such as supply networks (Stefanovic & Stefanovic, 2008). Yet, often there are some important issues that need to be addressed when the simulation approach is chosen for system evaluation. For example, for a large complex system, one of the key issues is how to verify and validate the simulation model of the system (Maria, 1997).

FUTURE TRENDS

The leading modelling standards discussed in this chapter (i.e., SCOR model, IDEF, UML and BPMN) are well recognized and adopted by practitioners and researchers. Among them, BPMN has gained a quickly rising popularity as the de facto standard for BPM (ISO/IEC 19510:2013) since it has been widely supported by commercial modelling tools (e.g., IBM Websphere) and integrated into universities' curriculum (Recker, 2010; Arevalo et al., 2016).

To make BPMN more applicable and powerful in functionality, OMG has developed some other notations which are complementary with BPMN. For example, the decision model notation (DMN) is developed to complement BPMN in an XML representation and to create a standardized means to automate decisions in processes (OMG, 2016). Besides, the case management model notation (CMMN) can be used to incorporate knowledge workers decisions into a process's meta-model (OMG, 2016).

On the other hand, BPMN has some major drawbacks. As mentioned earlier, BPMN has limited support for process decomposition that is vital for modelling large-scale business processes such as supply networks (Recker, 2010). Further, the current version of BPMN (v2.0) does not support ubiquitous computing where one person operates multiple computers (Yousfi et al., 2016). Therefore, it is anticipated that BPMN will be enhanced or extended to overcome those limitations in the future.

From the SCM perspective, the main objective of SCPM is to assist integration and collaboration of business processes both within individual chain members and across their boundaries. To this end, the literature shows that there is an increasing trend of combining different modelling techniques that complement each other (Stefanovic, Stefanovic, & Radenkovic, 2009). For example, Ryu & Yukesan (2007) developed a new modelling method called collaborative process modelling (CPM). This new method combines the analytical power of Petri



Nets, the simplicity of IDEF3 and the representation capability of UML.

Furthermore, recent literature review shows that researchers tend to use a hybrid modelling approach for simulation where two or more techniques are used in an integrated way (Jahangirian et al., 2010). For instance, Stefanovic et al. (2009) proposed a methodology for modelling structure and dynamics of complex supply networks by combining DES and SCOR. Another line of promising research is to create a merged model that subsumes a collection of process models (La Rosa, Dumas, Uba, & Dijkman, 2013).

CONCLUSION

Though supply chains can be generally regarded as business processes, they are much broader than normal business processes as they are boundary-spanning processes from raw materials, to component/part manufacturing, assembly, the distribution channel and to retail involving many participants. From the process perspective, the SCOR model is the most comprehensive standard for SCPM as it covers major processes, metrics, inputs, outputs and best practices. Meanwhile, to overcome the complexity of supply networks (e.g., in the aerospace or automotive industry), computer simulation (particularly DES) can be used as a valuable tool to analyse and evaluate the supply networks in terms of process design, operations or improvement in a quantitative way.

Based on an extensive review of the relevant literature on BPM and SCPM, this chapter offers a summarized description and critical discussion of the prominent standards and techniques for SCPM, including SCOR, IDEF, UML, BPMN, Petri Nets and simulation. It is expected that these standards or modelling techniques will evolve over time and will become more supportive to SCPM in the future. After all, the main objective of SCPM is to promote process integration and collaboration both within individual companies and between business partners along the supply chain.

REFERENCES

- Aguilar-Saven, R. S. (2004). Business process modelling: Review and framework. *International Journal of Production Economics*, 90(2), 129–149. doi:10.1016/S0925-5273(03)00102-6
- Arevalo, C., Escalona, M. J., Ramos, I., & Dominguez-Munoz, D. (2016). A metamodel to integrate business processes time perspective in BPMN 2.0. *Information and Software Technology*, 77, 17–33. doi:10.1016/j.infsof.2016.05.004
- Bae, H., & Seo, Y. (2007). BPM-based integration of supply chain process modelling, executing and monitoring. *International Journal of Production Research*, 45(11), 2545–2566. doi:10.1080/00207540601020502
- Baresi, L., Pezze, M., & Taentzer, G. (2001). Improving UML with Petri Nets. *Electronic Notes in Theoretical Computer Science*, 44(4), 1–13. doi:10.1016/S1571-0661(05)00186-6
- Barnard, J., & Mollaghasemi, M. (2006). *A Multi-view Framework for Defining the Services Supply Chain Using Object Oriented Methodology*. Retrieved from ProQuest Dissertations and Theses.
- Bolstorff, P., & Rosenbaum, R. (2012). *Supply Chain Excellence - A Handbook for Dramatic Improvement Using the SCOR Model* (3rd ed.). New York, NY: AMACOM – Book Division of the American Management Association.
- Brown, R., Recker, J., & West, S. (2011). Using virtual worlds for collaborative business process modelling. *Business Process Management Journal*, 17(3), 546–564. doi:10.1108/14637151111136414
- Christopher, M. (2011). *Logistics & supply chain management* (4th ed.). Harlow, UK: Financial Times Prentice Hall.
- Cooper, M. C., Lambert, D. M., & Pagh, J. D. (1997). Supply chain management: More than a new name for logistics. *The International Journal of Logistics Management*, 8(1), 1–14. doi:10.1108/09574099710805556

- Croxton, K. L., Garcia-Dastugue, S. J., Lambert, D. M., & Rogers, D. S. (2001). The supply chain management processes. *The International Journal of Logistics Management*, 12(2), 13–36. doi:10.1108/09574090110806271
- David, R., & Alla, H. (1994). Petri Nets for modelling of dynamic systems - A survey. *Automatica*, 30(2), 175–202. doi:10.1016/0005-1098(94)90024-8
- Dumas, M., La Rosa, M., Mendling, J., & Reijers, H. A. (2013). *Fundamentals of Business Process Management*. Dordrecht: Springer. doi:10.1007/978-3-642-33143-5
- Ellram, L. M., & Cooper, M. C. (2014). Supply chain management: Its all about the journey, not the destination. *Journal of Supply Chain Management*, 50(1), 8–20. doi:10.1111/jscm.12043
- Georgise, F. B., Thoben, K. D., & Seifert, M. (2011). Supply chain modelling and improving manufacturing industry in developing countries: A research agenda. *World Academy of Science, Engineering and Technology*, 60, 1998–2003.
- Giaglis, G. M. (2001). A taxonomy of business process modelling and information systems modelling techniques. *International Journal of Flexible Manufacturing Systems*, 1(2), 209–228. doi:10.1023/A:1011139719773
- Husby, P., & Swartwood, D. (2011). Supply chain operations reference (SCOR) model: Strengths and limitations. *Encyclopedia of Supply Chain Management*, 1265-1267.
- ISO. (2016). *ISO/IEC 19501:2005 – Information technology – Open Distributed Processing – Unified Modelling Language (UML) Version 1.4.2*. Retrieved July 25, 2016 from http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=32620
- Jahangirian, M., Eldabi, T., Naseer, A., Stergioulas, L. K., & Young, T. (2010). Simulation in manufacturing and business: A review. *European Journal of Operational Research*, 203(1), 1–13. doi:10.1016/j.ejor.2009.06.004
- Kasi, V. (2005). Systemic assessment of SCOR for modelling supply chains. *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*, 87. doi:10.1109/HICSS.2005.574
- La Rosa, M., Dumas, M., Uba, R., & Dijkman, R. (2013). Business process model merging: An approach to business process consolidation. *ACM Transactions on Software Engineering and Methodology*, 22(2), 1–42. doi:10.1145/2430545.2430547
- Lambert, D. M., Garcia-Dastugue, S. J., & Croxton, K. L. (2005). An evaluation of process-oriented supply chain management frameworks. *Journal of Business Logistics*, 26(1), 25–51. doi:10.1002/j.2158-1592.2005.tb00193.x
- Landeghem, R., & Bobeanu, C. V. (2002). *Formal modelling of supply chain: an incremental approach using Petri Nets*. Paper presented at the 14th European Simulations Symposium and Exhibition, Dresden, Germany.
- Leymann, F., & Altenhuber, W. (1994). Managing business processes as an information resource. *IBM Systems Journal*, 33(2), 326–348. doi:10.1147/sj.332.0326
- Li, L., Su, Q., & Chen, X. (2011). Ensuring supply chain quality performance through applying the SCOR model. *International Journal of Production Research*, 49(1), 33–57. doi:10.1080/00207543.2010.508934
- Maria, A. (1997). Introduction to modelling and simulation. In S. Andradottir, K.J. Healy, D.H. Withers, & B.L. Nelson (Eds.), *Proceedings of the 1997 Winter Simulation Conference*.



- Mentzer, J. T., DeWitt, W., Keebler, J. S., Min, S., Nix, N. W., Smith, C. D., & Zacharia, Z. G. (2001). Defining supply chain management. *Journal of Business Logistics*, 22(2), 1–26. doi:10.1002/j.2158-1592.2001.tb00001.x
- Min, H., & Zhou, G. (2002). Supply chain modelling: Past, present and future. *Computers & Industrial Engineering*, 43(1-2), 231–249. doi:10.1016/S0360-8352(02)00066-9
- Moreno-Montes de Oca, I., Snoeck, M., Reijers, H. A., & Rodriguez-Morffi, A. (2015). A systematic literature review of studies on business process modelling quality. *Information and Software Technology*, 58, 187–205. doi:10.1016/j.infsof.2014.07.011
- Object Management Group Inc. (2016) *OMG: Decision Model and Notation (DMN) Version 1.1*. Retrieved July 28, 2016 from <http://www.omg.org/spec/DMN/1.1/PDF/>
- Object Management Group (OMG). (2011). *Business Process Model and Notation 2.0, Technical Report*. Washington, DC: Object Management Group.
- Object Management Group (OMG). (2015). *OMG: Case Management Model and Notation Version 1.1*. Retrieved July 28, 2016 from <http://www.omg.org/spec/CMMN/1.1/Beta/>
- Pannirselvam, G. P., Ferguson, L. A., Ash, R. C., & Siferd, S. P. (1999). Operations management research: An update for the 1990s. *Journal of Operations Management*, 18(1), 95–112. doi:10.1016/S0272-6963(99)00009-1
- Persson, F. (2011). SCOR template - A simulation based dynamic supply chain analysis tool. *International Journal of Production Economics*, 131(1), 288–294. doi:10.1016/j.ijpe.2010.09.029
- Peterson, J. L. (1981). *Petri Net Theory and the Modelling of Systems*. Englewood Cliffs, NJ: Prentice-Hall.
- Pillat, R. M., Oliveira, T. C., Alencar, P. S. C., & Cowan, D. D. (2015). BPMNt: A BPMN extension for specifying software process tailoring. *Information and Software Technology*, 57, 95–115. doi:10.1016/j.infsof.2014.09.004
- Ramadan, M., Elmongui, H. G., & Hassan, R. (2011). *BPMN formalisation using coloured Petri Nets*. Paper presented at the Annual International Conference on Software Engineering & Applications, Singapore.
- Recker, J. (2010). Opportunities and constraints: The current struggle with BPMN. *Business Process Management Journal*, 16(1), 181–201. doi:10.1108/14637151011018001
- Recker, J. (2011). Evaluations of Process Modelling Grammars: Vol. 71. *Lecture Notes in Business Information Processing*. Berlin: Springer Berlin Heidelberg.
- Recker, J., Rosemann, M., Indulska, M., & Green, P. (2009). Business process modelling: A comparative analysis. *Journal of the Association for Information Systems*, 10(4), 333–363.
- Rumbaugh, J., Iacason, I., & Booch, G. (1999). *The Unified Modelling Language Reference Manual*. Reading, MA: Addison-Wesley.
- Russell, R., & Taylor, B. (2014). *Operations and Supply Chain Management* (8th ed.). Hoboken, NJ: John Wiley & Sons.
- Ryan, J., & Heavey, C. (2006). Process modelling for simulation. *Computers in Industry*, 57(5), 437–450. doi:10.1016/j.compind.2006.02.002
- Ryu, K., & Yukesan, E. (2007). CPM: A collaborative process modelling for cooperative manufacturers. *Advanced Engineering Informatics*, 21(2), 231–239. doi:10.1016/j.aei.2006.05.003
- Shannon, E. R. (1975). *Systems simulation: The art and science*. Englewood Cliffs, NJ: Prentice-Hall, Inc.

- Stefanovic, D., & Stefanovic, N. (2008). Methodology for modelling and analysis of supply networks. *Journal of Intelligent Manufacturing*, 19(4), 485–503. doi:10.1007/s10845-008-0098-0
- Stefanovic, D., Stefanovic, N., & Radenkovic, B. (2009). Supply network modelling and simulation technology. *Simulation Modelling Practice and Theory*, 17(4), 743–766. doi:10.1016/j.simpat.2009.01.001
- Supply Chain Council (SCC). (2012). *SCOR: Supply Chain Operations Reference Model Version 11.0*. Retrieved from www.supply-chain.org
- Thierry, C., Bel, G. & Thomas, A. (2010, October). The role of modelling and simulation in supply chain management. *SCS M&S Magazine*, 1-8.
- Trkman, P., Stemberger, I., Jaklic, J., & Groznik, A. (2007). Process approach to supply chain integration. *Supply Chain Management: An International Journal*, 12(2), 115–128. doi:10.1108/13598540710737307
- van der Aalst, W. M. P., & van Hee, K. M. (1996). Business process redesign: A Petri-Net-based approach. *Computers in Industry*, 29(1-2), 15–26. doi:10.1016/0166-3615(95)00051-8
- Wang, J. (2007). Petri Nets for dynamic event-driven system modelling. In Fishwick, P. (Ed.) *Handbook of Dynamic System Modeling*. CRC Press.
- White, S. A. (2006). *Introduction to BPMN*. Retrieved May 18, 2016, from http://www.omg.org/news/meetings/workshops/soa-bpm-mda-2006/00-T4_White.pdf
- Xu, H. (2010). A web-based system for proactive management of supply exceptions. *Journal of Manufacturing Systems*, 29(2/3), 91–101. doi:10.1016/j.jmsy.2010.11.003
- Xu, H. (2015). Business processes for enhancing coordination in supply networks. *International Journal of Information Systems and Supply Chain Management*, 8(2), 14–30.
- Xu, H., Koh, L., & Parker, D. (2009). Business processes inter-operation for supply network coordination. *International Journal of Production Economics*, 122(1), 188–199. doi:10.1016/j.ijpe.2009.05.015
- Yousfi, A., Bauer, C., Saidi, R., & Dey, A. K. (2016). uBPMN: A BPMN extension for modelling ubiquitous business processes. *Information and Software Technology*, 74, 55–68. doi:10.1016/j.infsof.2016.02.002

KEY TERMS AND DEFINITIONS

Benchmarking: A performance management practice by which a company compares its own business processes and performance metrics to those of the best in the industry.

Business Process Modelling: The activity of representing a company's business processes so that the current process can be performed in a more effective and/or efficient way leading to improved business performance.

Business Process Reengineering: A method by which business performance can be dramatically improved through rethinking and redesign of processes.

External Processes: The business processes that cannot be executed without involvement of an external business partner such as the customer relationship management process.

Internal Processes: The business processes that can be executed without involvement of an external business partner such as the internal production process.



Monte Carlo Simulation: A computerized simulation technique which is usually used for analyzing the behaviour of a system or a process involving uncertainties.

Object-Oriented: System design or analysis, or software program that use objects.

Process Integration: An approach of attaining close alignment and seamless coordination of internal and/or external business processes.

System Dynamics: A method through which the dynamic behaviour of a complex system over time can be better understood by taking into account internal feedback and time delays.

Measuring Low Carbon Supply Chain

Muhammad Shabir Shaharudin
Universiti Sains Malaysia, Malaysia

Yudi Fernando
Universiti Malaysia Pahang, Malaysia

INTRODUCTION

Environmental management is an important issue that has attract numerous scholars and practitioners especially the issue of climate change. Recently, scholars and practitioners are concerned with raising greenhouse gasses (GHGs) especially carbon emission and its impact on climate change (Palak, Eksioğlu, & Geunes, 2014). GHGs are dangerous but the world governments have taken initiative to control the release and used of these gasses except Carbon Dioxide. Carbon emissions are not specifically produce by manufacturer but almost by all activities. Thus, climate change mitigation strategy, which carbon emission as the main threat has been regard as very important. The mitigation strategy includes carbon taxes, carbon trading, clean development and joint implementation that can be found in Kyoto Protocol mechanism. The mitigation strategies for carbon emissions are adopted by government by introducing government regulation (Choudhary, Sarkar, Settur, & Tiwari, 2015). On the other hand, firms also need to adhere to government regulation and adopting carbon reduction strategies to meet their customer requirements and to remain competitive (Seuring, 2013). To do so, firms are expected to increase their competitiveness through Low Carbon Supply Chain Management (LCSCM) by practicing Carbon Footprint (CF). The definition of CF is the recording of carbon emissions throughout the life cycle of a product (Wiedmann & Minx, 2007).

In order to record carbon emissions, the concept of LCSCM need to be understood. As proposed

by Böttcher & Müller (2015), LCSCM consists of three groups of practices. The first is development of low carbon products in all product development process by assessing and reducing carbon emissions as well as achieving eco-design. The second practice is low carbon production process where monitoring and reduction of carbon emissions should apply in all production process until the final goods. The third practice focuses on low carbon logistics because transportation activities contribute significantly to increasing of carbon emissions. These practices are helping firms to develop carbon mitigation strategy and choose the best strategy to be implemented. As a result, firms now are able to identify and measure carbon emissions at operational and management settings.

BACKGROUND

As LCSCM is recently growing in importance, scholars and practitioners have started to put more interest in measuring carbon emission. In the literature, there are many scholars focusing on carbon emission measurement methods but lack of empirical data (Lee, 2011). Scholars then shift their focus in developing measurement methods and model to help practitioners to record and measure their carbon emissions using data generation (Bai, Sarkis, Wei, & Koh, 2012; Schmidt, 2009). However, there are also a few studies done by scholars using real firm or industrial data (Burnett, 2011; Cadarso, Gómez, López, & Tobarra, 2016; Lee, 2012; Nakajima, Kimura, & Wagner,

2014; Palak et al., 2014; Pelletier, Ibarburu, & Xin, 2013; Rahman & Khondaker, 2012; Tsai, Lin, Hwang, & Huang, 2014; Yusuf et al., 2013; Zhao, Neighbour, Han, McGuire, & Deutz, 2012; Zubelzu & Álvarez, 2015). Yet, the difficulty to get real data have seen many scholars using qualitative studies to measure LCSCM (Dadhich, Genovese, Kumar, & Acquaye, 2014; Dayaratne & Gunawardana, 2015; Gopalakrishnan, Yusuf, Musa, Abubakar, & Ambursa, 2012; Lee, 2012; Pueyo, 2013).

Difficulties in getting data also complicates the study of LCSCM. For example, in the study of Bai et al. (2012), environmental performance measurement does not give any information that can be implemented by practitioners. Another important gap in the literature is that scholars are divided between management and engineering focus when measuring LCSCM. There are a few scholars in management focus where their interest of studies are contributing to top management firm, policymakers' regulation and using management theories (de Sousa Jabbour, Jabbour, Latan, Teixeira, & de Oliveira, 2015; Fernando, Shaharudin, & Wahid, 2016; Gunasekaran & Spalanzani, 2012; Jabbour, Neto, Gobbo, de Souza Ribeiro, & de Sousa Jabbour, 2015; Lee, 2011; Matthews, Hendrickson, & Weber, 2008; Sarkis, Zhu, & Lai, 2011; Seuring, 2013; Shaharudin & Fernando, 2015; Zhu, Sarkis, & Lai, 2013). On the other hand, engineering focus scholars measure LCSCM using engineering theories such as Game Theory (Zhao et al., 2012), Chaos Theory (Stapleton, Hanna, & Ross, 2006), Swift Even Flow Theory (Seuring, 2009), Graph Theory (Wagner & Neshat, 2010), Neo-Institutional Theory (Genus & Mafakheri, 2014) and Fuzzy Set Theory (Erol, Sencer, & Sari, 2011; Liu, Perng, & Ho, 2013; Wang, Chan, Yee, & Diaz-Rainey, 2012) to name a few. This chapter is interested in looking at previous studies on LCSCM on management focus as it provides rich information to both scholars and practitioners.

The objectives of this study is twofold: 1) outcome of LCSCM and 2) measurement of LCSCM.

The significance of the study is that this chapter is contributing to LCSCM management focus literature. In model development, through extensive literature review, this chapter will provide evidence that LCSCM is an outcome that firms are pursuing. On the other hand, the literature review of LCSCM will provide evidence of previous studies method of CF in the supply chain.

As this chapter deals with literature review, the chapter will adopt and adapt content analysis method employed by Seuring (2013) and Stechemesser and Guenther (2012) with some modifications. The chapter has no classical structure but will be structured as follows: the method of content analysis with some descriptive information of previous studies, findings of the content analysis, discussion of the findings and conclusions.

METHODOLOGY

This chapter has followed a systematic review approach to build knowledge of the topic (Seuring, 2013; Stechemesser & Guenther, 2012). There are several steps taken to begin literature review analysis. The steps are as follows:

Step 1: *Selecting research questions, databases and appropriate search terms.* In this step, developing a research question to begin searching for journal articles is important. In this chapter, the research questions are “what is LCSCM?” and “what is the measurement of LCSCM?” Then, the databases used are: Elsevier (www.elsevier.com) and Emerald (www.emerald.com). Next is the search terms used to find articles from those databases. Even though this chapter is focusing on LCSCM, other related supply chain keywords should not be neglected because based on early readings, these keywords are used widely to explain LCSCM. Thus, the relevant search terms are:



- Carbon footprint supply chain management
- Low carbon supply chain management
- Environmental management supply chain
- Renewable energy supply chain management
- Green supply chain management

Step 2: *Applying practical screening criteria (Filter 1).* This step requires filter management from those databases. Filters are used to select only relevant publications. In this chapter, the criteria includes latest five years publications to ensure the relevancy of the topic and findings and only journal articles will be taken into consideration. In addition, only business management or supply chain and operations clusters journals are taken. Some of the examples include:

- Journal of Cleaner Production
- International Journal of Production Economics
- Industrial Marketing Management
- Renewable and Sustainable Energy Reviews
- Resources, Conservation and Recycling
- Ecological Economics
- Energy Policy
- Journal of Environmental Management
- European Journal of Operational Research
- Journal of Operations Management
- Industrial Marketing Management

Step 3: *Sorting and generating theme for the study (Filter 2).* After the completion of step one and two, as many as 910 journal articles were recorded. These articles are then managed by reference management software to help with sorting and generating theme of the study. Theme generation is to identify all keywords used by scholars in regards to LCSCM. The

management software used for this chapter is Qiqqa (www.qiqqa.com). Other than managing references, Qiqqa also is used to import missing metadata. For example, the title, year, author, journal publication, issue and page number of the article need to be checked and corrected. This software also used to remove any duplicate articles due to bulk downloads from several databases. The advantages of using Qiqqa is that it can provide another layer of filter to the current filtered articles. This will enable user to extract specific information and select most appropriate articles. Below is the step taken using Qiqqa to apply Filter 2 for 910 journal articles:

- Adding all 910 journal articles into Qiqqa software
- Using BibTeX Sniffer, a function in Qiqqa to identify missing metadata and information so that keywords of every journal articles can be identified correctly and used for the next step
- Qiqqa software will generate tags and autotags based on all journal articles
- Presenting Filter 2 results (see Table 1)
- Re-apply filters for measurement of LCSCM. Due to the first objective of this chapter is to investigate LCSCM as an outcome, this step of re-apply filters for measurement is to specifically answer second objective of this chapter.

Step 4: *Extracting relevant information.* In the final step, reference management software will be used to see pattern and identify quality articles for further investigation. Then, manual analysis of articles is required to extract information for the purpose of the study.

Table 1. Filter 2 using Qiqqa software

Criteria	Filter
Tags	Supply chain and operations management OR Supply chain capabilities OR Supply chain carbon footprint OR Supply chain analysis OR Supply chain changes OR Supply chain collaboration OR Supply chain competency OR Supply chain configuration OR Supply chain coordination OR Supply chain cycle time OR Supply chain design OR Supply chain disruption OR Supply chain dynamics OR Supply chain financial link OR Supply chain fit OR Supply chain flexibility OR Supply chain governance OR Supply chain implementation OR Supply chain information integration OR Supply chain key trends OR Supply chain knowledge management OR Supply chain leadership OR supply chain learning OR Supply chain learning loops OR Supply chain management (SCM) OR Supply chain management practices OR Supply chain management strategy OR supply chain maturity OR Supply chain model OR Supply chain modeling OR Supply chain network OR Supply chain network design OR supply chain network optimization OR Supply chain network optimization OR supply chain operations reference OR Supply chain optimization OR Supply chain partners OR Supply chain partnership OR Supply chain portfolio OR Supply chain practices OR Supply chain quality OR supply chain relational capital OR Supply chain responsiveness OR Supply chain risk assessment OR Supply chain risk management OR Supply chain strategy OR Supply chain sustainability trajectories OR Supply chain vulnerability
Autotags	Supply chain performance OR supply chain planning OR supply chain environmental OR supply chain management OR supply chain management practices OR supply chains OR supply chain sustainability OR supply chain strategy OR supply chain risk management OR supply chain relationship management OR supply chain practices OR supply chain partnership OR supply chain partners OR supply chain optimization OR supply chain operations OR supply chain network design OR supply chain networks OR supply chain leadership OR supply chain integration OR supply chain information OR supply chain governance OR supply chain flexibility OR supply chain disruptions OR supply chain design OR supply chain coordination OR supply chain configuration OR supply chain capabilities OR supply chain analysis OR supply chain agility Measurement OR measuring sustainability OR measuring performance
Publication	Elsevier OR Elsevier B.V. OR Elsevier Inc. OR Elsevier Ltd. OR Emerald Group Publishing Limited
Year	2016 OR 2015 OR 2014 OR 2013 OR 2012 OR 2011

FINDINGS

This section is structured according to review protocol such as information on author, year, publication and focus of the paper. Out of 910 journal articles (after Filter 1) that have been downloaded, only 246 journal articles (after Filter 2) are identified as important in Qiqqa that meet all search criteria.

In this chapter, for the purpose of review protocol (descriptive information), all 910 journal articles will be used while for author, country and focus will used 246 journal articles as basis. This is because 246 is more specific and smaller number that can give narrow or focus answer while the bigger number (910) will be used to capture general information of LCSCM. The last part of protocol, which is measurement of LCSCM needed manual analysis of 246 journal articles.

By using filter function performing inclusion and exclusion of criteria, descriptive analysis based on Table 2 protocol is able to be collected. Noted

that these articles have been given filters in step two and step three as discussed in methodology section. This section will start with bibliographic analysis for LCSCM and then analysis specifically for measurement. This is because the objectives of this study is twofold: 1) outcome of LCSCM and 2) measurement of LCSCM.

Table 2. Example Review protocol

Bibliographic Data	
Tag and Autotag	What is the highest keyword used in the publication?
Author	Who has the highest publication?
Country	Which highest country is subject of the publication?
Year	What year is the highest publication?
Publication	Which database has highest publication?
Focus of measurement	What is the main measurement of LCSCM?

Bibliographic Analysis

Qiqqa software has helped to identify keywords frequently used in the journal articles. Tags are given by author or journal publication based on keywords or key terms proposed by author or publisher. Based on 910 journal articles, renewable energy has the highest number (75) followed by supply chain management (73), LCSCM (69), GSCM (17). These are direct keywords that can be found in the journal articles and it shows that indeed low carbon studies are increasing in popularity in the literature. Renewable energy study which is a practice contributing towards LCSCM is an important study for future research to be undertaken by scholar. On the other hand, the Autotag of Qiqqa, after learning the pattern amongst 910 journal articles identify environmental management (497) and low carbon (257) as highest used keywords throughout all journal articles.

LCSCM studies has garner many interest from scholars both engineering and management approach. It is thought that engineering focus scholars will top the list since these scholars published more in LCSCM literature. However, interestingly management focus scholars has published journal articles on LCSCM more than engineering focus scholars. Top scholars are Joseph Sarkis (17), Kannan Govindan (17), Angappa Gunasekaran (13), Qinghua Zhu (11), Stefan Seuring (9) and Chunguang Bai (8).

The other interesting fact is the country where the studies being conducted. The findings are China (32), India (11), Malaysia (9), UAE (6), Australia (4), Brazil (4), Iran (3), Turkey (3), South Korea (2) and the UK (2). It is surprising to see Asian countries topping the list due to availability of data and LCSCM awareness are stronger in developed countries especially in Europe and USA. However, this information should not be taken literally because nowadays many scholars are able to conduct research almost everywhere in the world and collaborate with many scholars around the world. Also, some of the journal articles do not mentioned any country of study

due to type of article is either conceptual or using data generation software. Nevertheless, those top countries are giving good signal showing those countries are increasing knowledge of LCSCM.

The journal articles by year is also important because it can show a pattern of latest trend or future research direction. Out of 910 journal articles on LCSCM, the year 2014 had seen many journal articles been published in terms of LCSCM (227) while in the year 2015, there are 165 journal articles. On the other hand, the year 2012 and 2013 has 153 and 148 journal articles on LCSCM respectively. Looking at year 2016, even though it is not even half a year there are already 21 journal articles on LCSCM. This shows the strength of LCSCM studies in recent years.

The other part of analysis aim at the highest publication databases between Elsevier and Emerald. According to the analysis, Elsevier has the highest number of journal articles on LCSCM with 663 journal articles and Emerald database have 247. However, the number of management focus articles mostly are available in Emerald database but many top scholars have their publication in Elsevier database because it holds several reputable journal publications.

Next, the focus of study captured from specific 246 journal articles are renewable energy, manufacturing, climate change, GSCM, supplier, risk management, carbon emissions and LCSCM, performance and environmental management. These keywords are important for scholars. It is because all of these articles embedded these concepts and published in reputable journal publications. Thus, for scholars to publish in good journal publications, scholars should incorporated these keywords in their study.

Lastly, the second objective of this chapter is measurement of LCSCM. Surprisingly, there are only 442 journal articles on measurement keyword and out of that number, only 22 journal articles are on LCSCM measurement or LCSCM performance measurement. This findings confirm the gap in the literature. By manually navigates through these articles, the measurement of LC-

SCM is indeed depending on the underlining theory and type of measurement. Most journal articles on performance measurement are using Game Theory to measure LCSCM while the most influential management theory used for LCSCM measurement are Institutional Theory and Complexity Theory. The type of measurement used in LCSCM literature are mathematical or formula based studies with auto generated dataset and the most common method used is qualitative case study. This finding however is also raised by Lee (2011) who suggested more qualitative research should be undertaken to gather empirical results. This study on the other hand is looking at the measurement perspective. Since there are plenty measurement studies using qualitative methods, this chapter has identified qualitative method as top measurement of LCSCM.

DISCUSSION, FUTURE RESEARCH, AND RECOMMENDATIONS

Based on findings of this content analysis, the direction of LCSCM is clear. The first direction is that since renewable energy, SCM and GSCM are practices of LCSCM and among most used keywords, it define LCSCM as an outcome of the study. Scholars are shifting their attention from GSCM because it is heavily research and this has confirm LCSCM as the focus of the study or dependent variable. The second direction is scholars should embedded renewable energy and LCSCM in their study because it is the future trends in the literature.

There are two groups of scholars for LCSCM namely engineering focus and management focus. Even though engineering focus articles are plenty in the literature but management focus has established as more significant. This is due to several management focus scholars such as Sarkis, Govindan, Gunasekaran, Zhu and Seuring are publishing in reputable journal publications. In addition, in the business management and sup-

ply chain and operations clusters, management focus or approach is more appropriate because firms decision maker are top management that has little understanding of technical terms. While engineering focus scholars are also publishing journal articles with practical implications, it is more on process or technical approach. Thus, it is recommended for future research scholars to adopt management theory in their studies to ensure contribution to practitioner and policymakers are strong and to be able to publish in reputable journal publications.

This chapter only look into Elsevier and Emerald databases, which suggests that future research to include other databases. On the other hand, looking at the focus of the study out of 246 journal articles, LCSCM, risk management and manufacturing are all related to one another. These keywords suggest that LCSCM should be studied in the context of manufacturing industry and LCSCM is part of risk management strategy for firms. Perhaps, practitioners should consider LCSCM in implementation of risk management at firm.

In terms of contribution of this chapter to practitioner, by understanding the challenge in gathering real data for CF, firms should support the climate change mitigation strategy through public disclosure of carbon emissions data. In addition, practitioner should take note that the trend of renewable energy has already begun replacing fossil energy for power generation. On top of that, LCSCM has been an outcome that firms pursuing. Thus, practitioner need to start practicing LCSCM and GSCM to stay competitive.

An alarming finding is that there are only 22 journal articles on LCSCM measurement. This is in line with research gaps where scholars are finding it difficult to collect data of carbon emissions (Henri & Journeault, 2010; Lee, 2012; Matthews et al., 2008; Nakajima et al., 2014). As for the question “what is the measurement of LCSCM?” with current literature used in this chapter, management approach is more fruitful



and the use of qualitative study is more favorable because mathematical analysis mostly has not been tested with real data yet. The future research should then look at “how to measure LCSCM?” using qualitative study.

CONCLUSION

In summary, it can be confirmed that LCSCM is an outcome of study and the measurement of LCSCM depends on the type of study (theory used and method used in the study). Therefore, more study on LCSCM measurement is needed to generalize and standardize the content analysis. Finally, this chapter would like to highlight the importance of the findings especially the focus key terms such as renewable energy, LCSCM, GSCM and risk management. These are the main theme in the literature of LCSCM. In addition, the impact of limited literature on LCSCM measurement will definitely slow the adoption of LCSCM among practitioners and scholars.

REFERENCES

- Bai, C., Sarkis, J., Wei, X., & Koh, L. (2012). Evaluating ecological sustainable performance measures for supply chain management. *Supply Chain Management: An International Journal*, 17(1), 78–92. doi:10.1108/13598541211212221
- Böttcher, C. F., & Müller, M. (2015). Drivers, Practices and Outcomes of Low-carbon Operations: Approaches of German Automotive Suppliers to Cutting Carbon Emissions. *Business Strategy and the Environment*, 24(6), 477–498. doi:10.1002/bse.1832
- Burnett, M. S. (2011). Carbon Management in Tourism: Mitigating the Impacts on Climate Change. *Greenhouse Gas Measurement and Management*, 1(3–4), 216–217. doi:10.1080/20430779.2011.637672
- Cadarso, M., Gómez, N., López, L. A., & Tobarra, M. (2016). Calculating tourism's carbon footprint: Measuring the impact of investments. *Journal of Cleaner Production*, 111, 529–537. doi:10.1016/j.jclepro.2014.09.019
- Choudhary, A., Sarkar, S., Settur, S., & Tiwari, M. K. (2015). A carbon market sensitive optimization model for integrated forward-reverse logistics. *International Journal of Production Economics*, 164, 433–444. doi:10.1016/j.ijpe.2014.08.015
- Dadhich, P., Genovese, A., Kumar, N., & Acquaye, A. (2014). Developing sustainable supply chains in the UK construction industry: A case study. *International Journal of Production Economics*, 164, 271–284. doi:10.1016/j.ijpe.2014.12.012
- Dayaratne, S. P., & Gunawardana, K. D. (2015). Carbon footprint reduction: A critical study of rubber production in small and medium scale enterprises in Sri Lanka. *Journal of Cleaner Production*, 103, 87–103. doi:10.1016/j.jclepro.2014.09.101
- de Sousa Jabbour, A. B. L., Jabbour, C. J. C., Latan, H., Teixeira, A. A., & de Oliveira, J. H. C. (2015). Reprint of Quality management, environmental management maturity, green supply chain practices and green performance of Brazilian companies with ISO 14001 certification: Direct and indirect effects.. *Transportation Research Part E, Logistics and Transportation Review*, 74, 139–151. doi:10.1016/j.tre.2014.12.011
- Erol, I., Sencer, S., & Sari, R. (2011). A new fuzzy multi-criteria framework for measuring sustainability performance of a supply chain. *Ecological Economics*, 70(6), 1088–1100. doi:10.1016/j.ecolecon.2011.01.001
- Fernando, Y., Shaharudin, M. S., & Wahid, N. A. (2016). Eco-innovation practices: A case study of green furniture manufacturers in Indonesia. *International Journal of Services and Operations Management*, 23(1), 43–58. doi:10.1504/IJSOM.2016.073289

- Genus, A., & Mafakheri, F. (2014). A neo-institutional perspective of supply chains and energy security: Bioenergy in the UK. *Applied Energy*, 123, 307–315. doi:10.1016/j.apenergy.2014.01.084
- Gopalakrishnan, K., Yusuf, Y. Y., Musa, A., Abubakar, T., & Ambursa, H. M. (2012). Sustainable supply chain management: A case study of British Aerospace (BAe) Systems. *International Journal of Production Economics*, 140(1), 193–203. doi:10.1016/j.ijpe.2012.01.003
- Gunasekaran, A., & Spalanzani, A. (2012). Sustainability of manufacturing and services: Investigations for research and applications. *International Journal of Production Economics*, 140(1), 35–47. doi:10.1016/j.ijpe.2011.05.011
- Henri, J.-F., & Journeault, M. (2010). Eco-control: The influence of management control systems on environmental and economic performance. *Accounting, Organizations and Society*, 35(1), 63–80. doi:10.1016/j.aos.2009.02.001
- Jabbour, C. J. C., Neto, A. S., Gobbo, J. A., de Souza Ribeiro, M., & de Sousa Jabbour, A. B. L. (2015). Eco-innovations in more sustainable supply chains for a low-carbon economy: A multiple case study of human critical success factors in Brazilian leading companies. *International Journal of Production Economics*, 164, 245–257. doi:10.1016/j.ijpe.2014.11.015
- Lee, K. H. (2011). Integrating carbon footprint into supply chain management: The case of Hyundai Motor Company (HMC) in the automobile industry. *Journal of Cleaner Production*, 19(11), 1216–1223. doi:10.1016/j.jclepro.2011.03.010
- Lee, K. H. (2012). Carbon accounting for supply chain management in the automobile industry. *Journal of Cleaner Production*, 36, 83–93. doi:10.1016/j.jclepro.2012.02.023
- Liu, S. Y., Perng, Y. H., & Ho, Y. F. (2013). The effect of renewable energy application on Taiwan buildings: What are the challenges and strategies for solar energy exploitation? *Renewable & Sustainable Energy Reviews*, 28, 92–106. doi:10.1016/j.rser.2013.07.018
- Matthews, H. S., Hendrickson, C. T., & Weber, C. L. (2008). The Importance of Carbon Footprint Estimation Boundaries. *Environmental Science & Technology*, 42(16), 5839–5842. doi:10.1021/es703112w PMID:18767634
- Nakajima, M., Kimura, A., & Wagner, B. (2014). Introduction of material flow cost accounting (MFCA) to the supply chain: A questionnaire study on the challenges of constructing a low-carbon supply chain to promote resource efficiency. *Journal of Cleaner Production*.
- Palak, G., Ekşioğlu, S. D., & Geunes, J. (2014). Analyzing the impacts of carbon regulatory mechanisms on supplier and mode selection decisions: An application to a biofuel supply chain. *International Journal of Production Economics*, 154, 198–216. doi:10.1016/j.ijpe.2014.04.019
- Pelletier, N., Ibarburu, M., & Xin, H. (2013). A carbon footprint analysis of egg production and processing supply chains in the Midwestern United States. *Journal of Cleaner Production*, 54, 108–114. doi:10.1016/j.jclepro.2013.04.041
- Pueyo, A. (2013). Enabling frameworks for low-carbon technology transfer to small emerging economies: Analysis of ten case studies in Chile. *Energy Policy*, 53, 370–380. doi:10.1016/j.enpol.2012.10.078
- Rahman, S. M., & Khondaker, N. (2012). Mitigation measures to reduce greenhouse gas emissions and enhance carbon capture and storage in Saudi Arabia. *Renewable & Sustainable Energy Reviews*, 16(5), 2446–2460. doi:10.1016/j.rser.2011.12.003

- Sarkis, J., Zhu, Q. H., & Lai, K. H. (2011). An organizational theoretic review of green supply chain management literature. *International Journal of Production Economics*, 130(1), 1–15. doi:10.1016/j.ijpe.2010.11.010
- Schmidt, M. (2009). Carbon accounting and carbon footprint – more than just diced results? *International Journal of Climate Change Strategies and Management*, 1(1), 19–30. doi:10.1108/17568690910934372
- Seuring, S. (2009). The product-relationship-matrix as framework for strategic supply chain design based on operations theory. *International Journal of Production Economics*, 120(1), 221–232. doi:10.1016/j.ijpe.2008.07.021
- Seuring, S. (2013). A review of modeling approaches for sustainable supply chain management. *Decision Support Systems*, 54(4), 1513–1520. doi:10.1016/j.dss.2012.05.053
- Seuring, S., & Müller, M. (2008). From a literature review to a conceptual framework for sustainable supply chain management. *Journal of Cleaner Production*, 16(15), 1699–1710. doi:10.1016/j.jclepro.2008.04.020
- Shaharudin, M. S., & Fernando, Y. (2015). Low carbon footprint: The supply chain agenda in Malaysian manufacturing firms. *Promoting Sustainable Practices through Energy Engineering and Asset Management*.
- Stapleton, D., Hanna, J. B., & Ross, J. R. (2006). Enhancing supply chain solutions with the application of chaos theory. *Supply Chain Management: An International Journal*, 11(2), 108–114. doi:10.1108/13598540610652483
- Stechemesser, K., & Guenther, E. (2012). Carbon accounting: A systematic literature review. *Journal of Cleaner Production*, 36, 17–38. doi:10.1016/j.jclepro.2012.02.021
- Tsai, K. T., Lin, T. P., Hwang, R. L., & Huang, Y. J. (2014). Carbon dioxide emissions generated by energy consumption of hotels and homestay facilities in Taiwan. *Tourism Management*, 42, 13–21. doi:10.1016/j.tourman.2013.08.017
- Wagner, S. M., & Neshat, N. (2010). Assessing the vulnerability of supply chains using graph theory. *International Journal of Production Economics*, 126(1), 121–129. doi:10.1016/j.ijpe.2009.10.007
- Wang, X., Chan, H. K., Yee, R. W. Y., & Diaz-Rainey, I. (2012). A two-stage fuzzy-AHP model for risk assessment of implementing green initiatives in the fashion supply chain. *International Journal of Production Economics*, 135(2), 595–606. doi:10.1016/j.ijpe.2011.03.021
- Wiedmann, T., & Minx, J. (2007). A definition of 'carbon footprint'. *Ecological Economics Research Trends*, 2, 55–65.
- Yusuf, Y. Y., Gunasekaran, A., Musa, A., El-Berishy, N. M., Abubakar, T., & Ambursa, H. M. (2013). The UK oil and gas supply chains: An empirical analysis of adoption of sustainable measures and performance outcomes. *International Journal of Production Economics*, 146(2), 501–514. doi:10.1016/j.ijpe.2012.09.021
- Zhao, R., Neighbour, G., Han, J., McGuire, M., & Deutz, P. (2012). Using game theory to describe strategy selection for environmental risk and carbon emissions reduction in the green supply chain. *Journal of Loss Prevention in the Process Industries*, 25(6), 927–936. doi:10.1016/j.jlp.2012.05.004
- Zhu, Q., Sarkis, J., & Lai, K. (2013). Institutional-based antecedents and performance outcomes of internal and external green supply chain management practices. *Journal of Purchasing and Supply Management*, 19(2), 106–117. doi:10.1016/j.pursup.2012.12.001

Zubelzu, S., & Álvarez, R. (2015). Urban planning and industry in Spain: A novel methodology for calculating industrial carbon footprints. *Energy Policy*, 83, 57–68. doi:10.1016/j.enpol.2015.03.025

Stechemesser, K., & Guenther, E. (2012). Carbon accounting: A systematic literature review. *Journal of Cleaner Production*, 36, 17–38. doi:10.1016/j.jclepro.2012.02.021



ADDITIONAL READING

Böttcher, C. F., & Müller, M. (2015). Drivers, Practices and Outcomes of Low-carbon Operations: Approaches of German Automotive Suppliers to Cutting Carbon Emissions. *Business Strategy and the Environment*, 24(6), 477–498. doi:10.1002/bse.1832

Sarkis, J., Zhu, Q. H., & Lai, K. H. (2011). An organizational theoretic review of green supply chain management literature. *International Journal of Production Economics*, 130(1), 1–15. doi:10.1016/j.ijpe.2010.11.010

Seuring, S. (2013). A review of modeling approaches for sustainable supply chain management. *Decision Support Systems*, 54(4), 1513–1520. doi:10.1016/j.dss.2012.05.053

Seuring, S., & Müller, M. (2008). From a literature review to a conceptual framework for sustainable supply chain management. *Journal of Cleaner Production*, 16(15), 1699–1710. doi:10.1016/j.jclepro.2008.04.020

KEY TERMS AND DEFINITIONS

Carbon Footprint: Carbon footprint is a term used to record carbon emissions by individual or a firm in order to see the impact of Greenhouse Gasses (GHGs).

Content Analysis: A robust and structured analysis of the literature.

Green Supply Chain Management: Managing and monitoring supply chain activities in accordance with environmental management practices.

Low Carbon Supply Chain Management Measurement: Method used to analyze the carbon emissions after it has been recorded.

Low Carbon Supply Chain Management: Elimination or reduction of waste along the processes in firms and members of supply chain while monitoring and reducing carbon emissions throughout the supply chain.

Missing Part of Halal Supply Chain Management

Ratih Hendayani

Telkom University, Indonesia

Yudi Fernando

Universiti Malaysia Pahang, Malaysia

INTRODUCTION

The awareness about halal products in the world, especially about halal foods, is increasing. Not only 1.57 billion Muslims in the world spreading over 57 Islamic countries (Halal Trade Globally, 2015), but many non-Muslims also regard halal products, halal foods especially, as a trend of a healthy lifestyle. This is an opportunity for certified halal foods manufacturers to make their halal products as a competitive advantage, of which the other food companies do not possess. Regarding the production processes, the label “halal foods” must not be a mere proclamation made by the manufacturers. Halal food producers must start managing their supply chain, particularly to decide and inspect on where they obtain the raw halal material and to ensure the optimal cleanliness and hygiene in the distribution, to ascertain that the raw materials or processed foods are free of contamination which may cause the products to become not halal. In fact, according to Zulfakar et al. (2014), with the complexity of the food trade in the world, especially in terms of distributing halal products throughout the world, consumers of halal food products are still not sure whether the products produced are completely in accordance with the principles of halal or the Sharia laws, whether or not the food is contaminated when distributed to the whole world.

To avoid problems of distribution with a long peak time, according to Omar and Jafar (2011), the company needs to ensure all supply chain net-

works comply with halal procedures in the entire supply chain globally. Currently, the market for halal products is widely distributed throughout the world. This leads to increased demand for halal products, not only from the Islamic countries but also from the non-Muslims. Hence the demand for halal products not only from Muslims but also from the others, causes halal products to become globally renowned. This phenomenon stresses the need for the halal supply chain. Supply chain on halal products, especially halal foods, means that every process and partner in the supply chain should be able to maintain the halal status of food products so that they can be guaranteed to be absolutely halal by the laws.

Given the importance of managing supply chains in halal products, many studies have discussed halal supply chain managements, such as in Tieman (2011), Tieman (2015), Alqudsi (2014), Zulfakar et al. (2014), Zulfakar (2011), Zailani et al. (2010), Ngah et al. (2014), Tieman et al. (2012); Aziz and Sulaiman (2014), Talib et al. (2013); Muhammad et al. (2009); Zailani et al. (2015), and there are still some other studies which are more focused on supply chain managements for halal foods. In addition, not all of these studies discuss halal supply chain management as a whole thus they are unable to capture the complexity of halal supply chain itself. Some just take the perspective of the consumer or distribution or simply in terms of the logistics of its course. The studies that discuss halal supply chain managements still have shortcomings in

translating the overall supply chain management and supply chain management to separate its role in logistics management. In gaining integrated halal products, logistics management is part of the supply chain management. Companies should be able to implement the concept of halal in the whole value chain system such as in procurement phase, processing, packaging, labeling, transporting, finance and logistics of food with the complete process. Unfortunately, the concept of halal logistics lacks focus and it is not enhanced by other following researchers who discuss halal supply chain managements so that there is still a lost concept on halal supply chain management (Kamaruddin et al., 2012). The lost concept is discussed in other studies, namely finance.

This chapter aims to explore the missing part of halal supply chain management, which previously was not defined as part of halal supply chain. Ensuring products' halal status, from the material sources until the products are bought by the consumers, has to be maintained with proper management, especially in the global food company. Halal supply chain is required to meet the demand from halal industry, and at the same time it should be aligned with the requirements of Sharia laws.

BACKGROUND

Every Muslim is required to consume halal products by its religion. Consequently, the growing Muslim population in the world means that the need for halal products continues to increase. The inter-religious tolerance and social relationships which are created between Muslims and non-Muslims allow the latter to start recognizing and realizing the benefits of halal products. These non-Muslims grasp the great benefits of halal products, in terms of hygiene, health, and quality. Consequently, the request for halal products arises, not only from Muslim consumers but also from non-Muslim consumers worldwide. In fact, according to Alqudsi (2014), the level of awareness

on the consumption of halal products and halal law of consumers around the world is increasing even from major non-Muslims populated countries, such as Australia. They are willing to pay more for halal products. Awareness of halal products needs to be accompanied with the knowledge of the meaning of halal and of the procedure about how the halal status of a product is maintained so that consumers are confident when consuming it. According to Rezai et al. (2012), we cannot simply guarantee that our products are halal only by avoiding its contacting with the haram material when the consumers buy our products. We have to go from the starting point to ensure that the product is not contaminated by haram materials or breaking any of the halal rules, from raw materials suppliers of products included in the manufacturing process to the distribution.

According to Tarmizi et al. (2014), one of the ways to avoid contamination in the process of transportation and distribution is the halal logistics, which involves logistics as a part of a supply chain involving in businesses, such as suppliers, manufacturers, distributors, and retailers, who will work together to meet needs for raw materials and distribute them to consumers. In addition to keeping out of contamination, halal logistics also plays a role to ensure that proper information on halal products is well received by consumers. It means that in order to obtain halal status, accordance of any detail logistic processes with Sharia principles is required, such as the procurement, distribution, raw materials, up through the supply chain, all must in conformity with the Sharia laws.

Tarmizi et al. (2014), Tieman (2013), and Kamaruddin et al. (2012), explain that halal logistics is a part of the supply chain, mainly a part of supply chain management of halal product which has a process that is in accordance with the general principles of Islam or the Sharia. The focus of this chapter is halal supply chain due to the larger scope of halal logistics and management that could have a direct impact on halal product itself that is distributed with the involvement of the supply chain actors, and also the process steps

in the supply chain companies. Moreover, based on Bahrudin, Ilyas and Village (2011) in Zulfakar et al. (2014), mention that the problems that can arise when the raw material of halal products obtained globally is that not all of the countries supplying raw materials is a Muslim country or under the supervision of Sharia law and Muslim rule, so that the raw materials which are derived from non-Muslim countries would be questionable in terms of its halal status. On the other hand, the demand for halal products from around the world is constantly increasing. So, in order to overcome doubts about the halal status of the products—the ones which gather its raw materials from non-Muslim countries—a comprehensive approach to halal supply chain management is required. It shall be well-managed and able to be applied globally in order to comply with the Sharia law, which is important for Muslim consumers. The goal of the halal supply chain management is to ensure that customers' satisfaction is achieved and that the halal status of food products remains intact throughout the entire supply chain processes. Zulfakar et al. (2014) also conclude that the integrity of halal food products must be protected by all means and all necessary steps should be taken by all parties involved in the supply chain to avoid cross-contamination that would cause the product to become non-halal or haram. So it takes an integrated management that can be applied appropriately and in accordance with the provisions of the Sharia, that is halal supply chain management.

Since halal supply chain management is very important for halal products industry, the application of halal supply chain management should be supported by the development of theory and study on halal supply chain management topic. However, literature in the field of supply chain management of halal foods is still limited. The review in Zulfakar et al. (2011) on halal supply chain only includes 41 sources related to the halal industry, starting from the procedure of halal manual to Malaysian halal standard, articles from newspapers, magazines, journals, brochures, slide presentations,

and halal-related relevant authoritative websites in Malaysia. The main limitation of the review is that they only focused on identifying processes, information flows, and the actors involved in the process of halal certification. Unfortunately, the discussion of halal supply chain management that is focused on the identification of supply chain process for halal products leaves a process that should be present in the process and can be found in the meaning and scope of a supply chain.

The meaning and scope of a supply chain can be observed from some definitions of supply chain management. According to Heizer and Render (2014), supply chain management is the management of raw materials into goods in process or semi-finished goods and finished goods that send the product to consumers through the distribution system. These activities include the traditional purchasing function, plus other important related activities. Meanwhile, according to Chopra (2013), supply chain management is the management of the flow of information, products, funds, between and fellow stage supply chain to improve the supply chain advantage. Council of Supply Chain Management Professionals (CSCMP) (2016) also states that supply chain management is the planning and management of all activities involved in procuring and activities of logistics that involve coordination and cooperation with the actors involved in the supply chain and integrate the management of supply and demand between companies that have the scope to support the coordination between the processes and activities in marketing, sales, product design, financial, and information technology. An overview of the halal supply chain management means that every definition and scope of supply chain management is in accordance with the principles of halal.

However, when viewed from the previous literature regarding halal supply chain management, such as Syazwan et al (2015), which already reviews some concept of halal supply chain management of other researchers such as Tieman (2013), elaborates the definition of halal supply chain as a management process of purchasing, distribution

and material handling, raw materials, which are then stored on a good supply of food or not, and the flow of information and administration of an organization involving the supply chain that can adapt to the general principles of Sharia. However, defining principles about halal supply chain in Syazwan et al. (2015) are still unclear because each researcher has a different interpretation.

These unclear halal supply chain principles are a result of there being a missing part of halal supply chain management definition from previous researchers. So, this chapter will discuss an advance definition and scope of halal supply chain through elaborations on the meaning of supply chain with the principle of Sharia as a whole, not just by definition alone, but its scope as well.

ISLAMIC FINANCING SYSTEM TO SUPPORT OPERATIONS OF HALAL SUPPLY CHAIN

Halal products have become the emerging issue of the world as predicted to be the next global market, as evidenced by the increasing demand for halal products and public awareness in the world about the benefits of halal products that are not only limited to a religious understanding or belief of one religion alone. Halal products are known to the world community as products that have the image of clean, healthy, and appropriate according to Sharia quality assurance procedures. The focus then is to guarantee quality, that it is clean, healthy and certainly must meet the criteria of the Sharia, that it is still maintained carefully when the procurement, distribution, and processing is done globally, which allows all of the processes carried out in different countries. Moreover, we have to focus on how products' halal status can be guaranteed or how to monitor the products so that they remain guaranteed halal even when the activities of logistics and supply chain actors may come from states which are not Muslim countries, or worse, from the places which even do not know the word "halal." In fact, some countries

which supply the raw material of halal products are countries known for its Muslim population as the minority, such as Thailand and American countries which supply a lot of raw materials for halal products in Indonesia, the largest Muslim country in the world.

To overcome this, companies can apply a particular supply chain management which is in accordance with the provisions of Islamic laws or the Sharia laws, as the solution of how the availability of halal products can be maintained and how they can be distributed throughout the world—with supervision from the source to the consumer is fixed in accordance with Sharia principles—that is called the halal supply chain management. The need of halal supply chain knowledge becomes one of the critical things that halal product industry should have to run their supply chain operation system, because without the knowledge, halal product industry does not have the guidelines on whether or not the companies' supply chain is already in conformity with the Sharia or Islamic law. So that, several studies on halal supply chain management are already conducted by some researcher with different scopes and focuses, such as Tieman (2011) that plans the basic requirements of halal food supply chain through in-depth interviews with qualitative, exploratory study. The study employs a frame of which the dependent variable is direct contact with haram, risk perception, and the independent variables are haram products, product characteristics (dry vs. wet product, bulk vs. together product), market characteristics (schools that had a different idea and local fatwa). The weakness of this study is that to measure the perceptions of customers, concrete tools for designing effective halal chain supply is needed, not only the perceptions of customers. Zailani et al. (2010), reveals that the industry needs to be provided with the appropriate training or motivational programs on issues of halal standard. The members of the industry, therefore, can be more innovative in improving their operations and more competitive in the world food trade. However, this model requires a com-

L

prehensive assessment of the suitability of existing supply chain. Also, Zulfakar et al. (2014) explain about several factors in improving the integrity of the food supply chain and provide a conceptual framework about this particular problem by using a conceptual paper through relational analysis. This study comes up with the idea that the factors in the supply chain of halal foods, such as certified halal, halal standard, traceability, halal, halal dedicated assets, trust and commitment among members of the supply chain, should be given serious attention to ensure the needs and welfare of consumers, that the halal foods can be satisfying for them and are well protected. Syazwan et al. (2015) have made a comprehensive literature review and its result is the critical success factors, which are: government support, transportation planning, information technology, HRM, collaborative relationship, halal certification, and halal traceability.

Based on a short overview of several previous researches above that discuss halal supply chain management, it can be seen that most of them focus on technical aspects during processes of halal supply chain and management, such as Sharia-based procurement, halal distribution, halal inventory, halal handling, halal product from raw to finished, halal information, halal documentation, halal logistics, halal value adding, halal supply chain members. They do not include other important parts of supply chain management. If we look again to supply chain management concept, it is defined as:

a set of integrated activities to manage upstream and downstream relationships with suppliers, manufacturers, distributors, retailers, and customers in order to deliver superior customer value at less cost, efficient and effective distribution, purchasing and procurement and other supply chain activities with a continuous flow of information and to deliver customer service and gaining competitiveness. (Verma and Seth, 2011)

Moreover, Chopra and Meindl (2013) define supply chain management as the management of

flows between and among supply chain stages to maximize total supply chain profitability, of which the supply chain flows are information, product, and funds. Thus, based on supply chain management definition, it evolves not only information and product but also a fund that has not been disclosed in the previous study of halal supply chain management.

Although the concept of Halal logistics on Kamaruddin et al. (2012) also includes companies financing system, most of other studies that specifically discussed Halal supply chain, finance is not included in Halal supply chain definition. Although in the definition and scope of the supply chain of Chopra or CSCMP, finance is mentioned. Wuttke et al. (2013) whose study focused on financial supply chain, because he thinks companies' managements often focus solely on the flow of products and information only, but less attention to cash flows involved in the process of such management. Wuttke et al. (2013) opinions are based on an understanding of supply chain management of Mentzer et al. (2001), which states that supply chain management focus on the coordination and cooperation of several stakeholders in the company to improve the flow of products, information, and finance throughout the supply chain. Given the importance of the involvement of finance in the management of the supply chain, in order to obtain an integrated supply chain Halal and fully covers all the important elements in Halal Supply chain principles.

Halal financial flow means that Islamic finance refers to the financial system of banking in accordance with Islamic law or sharia. According to Catherine (2015),

...the underlying principles governing Islamic finance mutually-risk and profit sharing between parties, guarantees of justice based on standards of Shariah compliance checks underlying Sharia-based business activities and assets. This encourages entrepreneurship, trade and commerce, but prohibit activities that involve Riba (interest), Maisir (gambling), and Gharar (speculative trad-

ing). There is a consensus among all Islamic law prohibits business operations involving interest and gambling, and they are in the production, distribution, promotion and sale of goods or services of non-Halal as alcohol and pork, and entertainment facilities immoral such as brothels, pubs, massage parlors and discotheques.

The role of Halal financial principles in Halal supply chain will be one of the important things in Halal product industry, beside integrated operation and coordination through supply chain management that comply with Islamic law. Halal product industry also can completed the Halal brand of Halal products with Halal investment, Halal cost, and Halal fund. Halal financial gives assurance to customers. Furthermore, Surah Al-Baqarah: 275 in Holy Quran states that "...but Allah has permitted trade and has forbidden interest (*Riba*)..." It states that in Islam, if a transaction using *riba* then it is forbidden or not Halal (Haram). If the transaction is Haram, automatically the product Halalness also becomes questionable, as stated in hadist Ibn Masood (Radiyallaahu Anhu) that narrates Rasulullah Sallallaahu Alaihi Wa Sallam, "...who acquires haram wealth and gives charity from it, and it is not accepted from him. If he spends from it, he does not have any blessing (*barakah*) in it..." (Musnad Ahmad, Sharh As-sunnah; Mishkaatul Masaabih pg. 242, Qadeemi). Besides to fulfil the Islamic law requirement for Halal product, it also can completed the Halal supply chain definition as one package because if the financial system inside Halal supply chain does not comply with Islamic law, it is not *barakah* and doubttable. Islamic financial system is very important to support and improve operations of Halal supply chain.

SOLUTIONS AND RECOMMENDATIONS

Based on the study, as well as the exposure of the background and the main focus, the problem

is the loss of an important part of supply chain management that is not discussed in previous research of Halal supply chain management studies which is very important for Halal supply chain operation. The concept of Halal that has not been addressed in previous studies are finance or acquisition of funds and cash flow, which all must comply with Sharia or Sharia finance. Based on the discussions in main focus, Islamic financial is the important things to make sure all investment, cash flow, and fund complies with Islamic law. Because of the very important role of Islamic financial that can influence the Halalness of the product, Islamic Financial should be applied and included in Halal supply chain. If the finance is not Halal then the supply chain operation cannot called a Halal supply chain. So that, to complete the "Halal" in Halal supply chain and advance it operations, then it is recommended to include Islamic financial system in Halal supply chain. When the entire stream in supply chain, such as the flow of products through Halal logistic, the flow of information through the involvement of all players in the supply chains, and the flow of finance related to cash flow, and every operation involved in the supply chain management can be applied to the Halal supply chain management, then it will create an integrated system that is perfect for Halal products industry. The new meaning of Halal supply chain management that already involving Islamic financing can be defined as:

...planning and supply chain management ranging from Sharia procurement, distribution Sharia, processing of Halal products through the logistics Halal, Halal certification process, maintenance of Halal products and Halal certification by monitoring the flow of Halal products, information-based Islamic and Halal finance involves cooperation among supply chain actors who have good cooperation with the principles of Islamic cooperation.

This new definition can be a guidelines for new perspective about Halal supply chain management.



FUTURE RESEARCH DIRECTIONS

This chapter highlights Islamic financing that should be included in Halal supply chain and bring new definition that more completed and more integrated for Halal product operation requirement. But this concept still need to be assessed by the quantitative study and to be tested by various Halal product in Muslim and also non-Muslim country. Not all Muslims and, especially, non-Muslims are aware about Islamic financing system as well as Halal food. Future study can research about Muslims and non-Muslims perception about Islamic financing in Halal supply chain. Not only the consumer, but supply chain partner and the authority also should be aware about Islamic financing system when operating the Halal supply chain.

CONCLUSION

As pressure from consumers' demands of Halal products, supply chain should focus on Halal industry. Given the importance of Halal supply chain management, it needs to be studied and developed. And this chapter increase of the significance definition of Halal supply chain management by adding a financial component that does not exist in the previous study, so the idea of Halal supply chain definition in this chapter is:

the planning and management of the supply chain from the Sharia procurement, Sharia distribution, processing Halal products through Halal logistics, Halal certification process, Halal product maintenance and Halal certification by monitoring the flow of Halal products, Sharia-based information, and Halal financing involve cooperation among supply chain actors who have good cooperation with the principles of Islamic cooperation.

REFERENCES

- Ab Talib, M. S., Abdul Hamid, A. B., & Zulfakar, M. H. (2015). Halal supply chain critical success factors: A literature review. *Journal of Islamic Marketing*, 6(1), 44–71. doi:10.1108/JIMA-07-2013-0049
- Ab Talib, M. S., Rubin, L., & Zhengyi, V. K. (2013). Qualitative research on critical issues in halal logistics. *Journal of Emerging Economies and Islamic Research*, 1(2).
- Alqudsi, S. G. (2014). Awareness and demand for 100% halal supply chain meat products. *Procedia: Social and Behavioral Sciences*, 130, 167–178. doi:10.1016/j.sbspro.2014.04.021
- As-sahabah. (2016). *Effect of Haram Food and Haram Earnings*. Retrieved March 21, 2016 from <http://www.as-sahabah.org/index.php/islamic-library/54-haram-food-and-haram-earnings>
- Aziz, N. A., & Sulaiman, S. S. (2014). Role of the Local Authority in Issuing License for Halal Certified Premise in the City of Shah Alam. *Procedia: Social and Behavioral Sciences*, 121, 133–143. doi:10.1016/j.sbspro.2014.01.1115
- Chopra, S., & Meindl, P. (2013). *Supply Chain Management: Strategy, Planning, and Operation* (5th ed.). Harlow, UK: Pearson Education.
- CSCMP. (2016). *Definition of Supply Chain Management*. Retrieved March 1, 2016 from <https://cscmp.org/supply-chain-management-definitions>
- Halal trade globally. (2015). *Halal Business Potential In Indonesia*. Retrieved November 9, 2015 from <http://www.dagangHalal.com/HalalInfo/HalalArticlesDtl.aspx?id=218>
- Heizer, J., & Barry, R. (2014). *Operation Management* (11th ed.). Pearson Education, Inc.



- Ho, C. S. (2015). International comparison of Shariah compliance screening standards. *International Journal of Islamic and Middle Eastern Finance and Management*, 8(2), 222–245. doi:10.1108/IMEFM-07-2014-0065
- Kamaruddin, R., Iberahim, H., & Shabudin, A. (2012). Willingness to pay for halal logistics: The lifestyle choice. *Procedia: Social and Behavioral Sciences*, 50, 722–729. doi:10.1016/j.sbspro.2012.08.075
- Mentzer, J. T., DeWitt, W., Keebler, J. S., Min, S., Nix, N. W., Smith, C. D., & Zacharia, Z. G. (2001). Defining supply chain management. *Journal of Business Logistics*, 22(2), 1–25.
- Muhammad, N. M. N., Isa, F. M., & Kifli, B. C. (2009). Positioning Malaysia as halal-hub: Integration role of supply chain strategy and halal assurance system. *Asian Social Science*, 5(7), 44.
- Ngah, A. H., Zainuddin, Y., & Thurasamy, R. (2014). Adoption of Halal Supply Chain among Malaysian Halal Manufacturers: An Exploratory Study. *Procedia: Social and Behavioral Sciences*, 129, 388–395. doi:10.1016/j.sbspro.2014.03.692
- Omar, E. N., & Jaafar, H. S. (2011). Halal supply chain in the food industry-A conceptual model. In *Business* (pp. 384–389). Engineering and Industrial Applications. doi:10.1109/ISBEIA.2011.6088842
- Praharsi, Y., & Purnomo, H. D. (2015). Designing e-supply chain collaborative management: A case study in total food solution company. *Jurnal Ilmiah Teknik Industri*, 2(3).
- Rezai, G., Mohamed, Z., & Nasir Shamsudin, M. (2012). Non-Muslim consumers understanding of Halal principles in Malaysia. *Journal of Islamic Marketing*, 3(1), 35–46. doi:10.1108/17590831211206572
- Tarmizi, H. A., Kamarulzaman, N. H., Latiff, I. A., & Rahman, A. A. (2014). Factors influencing readiness towards halal logistics among food-based logistics players in Malaysia. *UMK Procedia*, 1, 42–49. doi:10.1016/j.umkpro.2014.07.006
- Tieman, M. (2011). The application of Halal in supply chain management: In-depth interviews. *Journal of Islamic Marketing*, 2(2), 186–195. doi:10.1108/17590831111139893
- Tieman, M. (2013). Establishing the principles in halal logistics. *Journal of Emerging Economies and Islamic Research*, 1(1), 1–13.
- Tieman, M. (2015). Halal clusters. *Journal of Islamic Marketing*, 6(1), 2–21. doi:10.1108/JIMA-05-2014-0034
- Tieman, M., van der Vorst, J. G., & Che Ghazali, M. (2012). Principles in halal supply chain management. *Journal of Islamic Marketing*, 3(3), 217–243. doi:10.1108/17590831211259727
- Verma, A., & Seth, N. (2011). A conceptual framework for supply chain competitiveness. *International Journal of Humanities and Social Science*, 6(1), 5–10.
- Wuttke, D. A., Blome, C., & Henke, M. (2013). Focusing the financial flow of supply chains: An empirical investigation of financial supply chain management. *International Journal of Production Economics*, 145(2), 773–789. doi:10.1016/j.ijpe.2013.05.031
- Zailani, S., Arrifin, Z., Abd Wahid, N., Othman, R., & Fernando, Y. (2010). Halal traceability and halal tracking systems in strengthening halal food supply chains for food industry in Malaysia (a review). *Journal of Food Technology*, 8(3), 74–81. doi:10.3923/jftech.2010.74.81

Zailani, S., Kanapathy, K., Iranmanesh, M., & Tieman, M. (2015). Drivers of halal orientation strategy among halal food firms. *British Food Journal*, 117(8), 2143–2160. doi:10.1108/BFJ-01-2015-0027

Zulfakar, M. H., Jie, F., & Chan, C. (2012, June). *Halal food supply chain integrity: from a literature review to a conceptual framework*. In *10th ANZAM Operations, Supply Chain and Services Management Symposium* (pp. 1-23).

Zulfakar, M. H., Marhani, M. A., & Mohamed, S. A. (2014). Conceptual Framework on Halal Supply Chain Integrity Enhancement. *Procedia: Social and Behavioral Sciences*, 121, 58–67. doi:10.1016/j.sbspro.2014.01.1108

ADDITIONAL READING

Chopra, S., & Meindl, P. (2013). *Supply Chain Management: Strategy, Planning, and Operation* (5th ed.). Harlow: Pearson Education.

Heizer, J., & Barry, R. (2014). *Operation Management* (11th ed.). New Jersey: Pearson Education, Inc.

Ho, C. S. (2015). International comparison of Shariah compliance screening standards. *International Journal of Islamic and Middle Eastern Finance and Management*, 8(2), 222–245. doi:10.1108/IMEFM-07-2014-0065

Tieman, M., van der Vorst, J. G., & Che Ghazali, M. (2012). Principles in halal supply chain management. *Journal of Islamic Marketing*, 3(3), 217–243. doi:10.1108/17590831211259727

Wuttke, D. A., Blome, C., & Henke, M. (2013). Focusing the financial flow of supply chains: An empirical investigation of financial supply chain management. *International Journal of Production Economics*, 145(2), 773–789. doi:10.1016/j.ijpe.2013.05.031

KEY TERMS AND DEFINITIONS

Halal Supply Chain: Planning and supply chain management ranging from Sharia procurement, distribution Sharia, processing of Halal products through the logistics Halal, Halal certification process, maintenance of Halal products, and Halal certification by monitoring the flow of Halal products, information-based Islamic, and Halal finance involves cooperation among supply chain actors who have good cooperation with the principles of Islamic cooperation.

Halal: Every things that allowed by the Islam laws.

Islamic Financing: According to Catherine (2015) the underlying principles governing Islamic finance mutually-risk and profit sharing between parties, guarantees of justice based on standards of Sharia compliance checks underlying Sharia-based business activities and assets. This encourages entrepreneurship, trade and commerce, but prohibit activities that involve “Riba” (interest), “Maisir” (gambling) and “Gharar” (speculative trading). There is a consensus among all Islamic law prohibits business operations involving interest and gambling, and they are in the production, distribution, promotion and sale of goods or services of non-Halal as alcohol and pork, and entertainment facilities immoral such as brothels, pubs, massage parlors and discotheques.

Supply Chain Flow: (2013) The flow of information, products, funds, between supply chains to improve the supply chain advantages.

Supply Chain Management: (2011) A set of integrated activities to manage upstream and downstream relationships with suppliers, manufacturers, distributors, retailers and customers in order to deliver superior customer value at lesser cost, efficient, and effective distribution, purchasing and procurement, and other supply chain activities with a continuous flow of information and to deliver customer service and gaining competitiveness. Verma and Seth.

Notions of Maritime Green Supply Chain Management



Fairuz Jasmi

Universiti Sains Malaysia, Malaysia

Yudi Fernando

Universiti Malaysia Pahang, Malaysia

INTRODUCTION

Intercontinental trade relies heavily on maritime transportations to carry various cargoes for catalyzing global import-export trade. Roughly about 80% of international trade by volume and over 70% by value is carried by maritime operations globally (Cheng, Farahani, Lai, & Sarkis, 2015). As a vital component of life-line trade for various manufacturing companies all over the world, maritime supply chain have established new opportunities and unforeseeable challenges. The challenges faced by maritime supply chain dampened in the increased rate of climate change and global warming during the precedent decade (Lai, Lun, Wong, & Cheng, 2011). In this respect, carbon dioxide (CO²) and other greenhouse gases (GHGs) are emitted through the burning of fossil fuels from maritime transportation are responsible for a host of global environmental concerns. Playing the critical role as a major transportation and intermediary to assist trade flows in the global supply chain (Wong, Lai, & Cheng, 2009; Yang, Marlow, & Lu, 2009), many issues have been raised in operational context of maritime supply chain especially on environmental issues. As such, environmental protection has been extensively discussed by stakeholders, businesses, authorities as well as political leaders globally (e.g., Boykoff & Yulsman, 2013; Boykoff, 2009; Revkin, 2009; Rosenthal, 2009). In scholarly field, there has also been a surge in researches committed to address the related issues of environmental issues (e.g., Lee

& Kim, 2011; Ostrom, 2008; Wong, Lai, Shang, Lu, & Leung, 2012). The tightened environmental regulations imposed by International Maritime Organization (IMO) to the maritime sectors also contribute to this rising green trend (Sulaiman, Akmar, & Michel, 2013). As a result of this 'green' pressure, numerous maritime organizations have gradually begun to react to environmental concerns by embracing green supply chain management (GSCM) and other sustainability concept in their supply chain operations (Lai et al., 2011).

To the author's limited knowledge, relatively a small amount of studies have discussed GSCM dimension in the context of the maritime supply chain as well as sustainable notion of maritime green supply chain management (MGSCM). Preceding studies have stressed on tangible aspect of maritime operation such as sewage pollution, air pollution, and greenhouse gas (GHG) emissions (Cariou, 2011; Corbett, Wang, & Winebrake, 2009; Giziakis & Christodoulou, 2012; Hoffmann, Eide, & Endresen, 2012; Lirn, Lin, & Shang, 2014) but no intangible aspect in term of organizational management perspective of internal or external organizational capabilities can be found. If there is, it is literally a new domain in maritime supply chain and it is sensible to fill the gap in the literature by examining the GSCM dimensions that can be conceptualize towards defining the notion of MGSCM. Accordingly, the purpose of this paper is not to investigate empirically the dimension of MGSCM but to identify key GSM capability factors and examine the accountability of the

GSCM concept to be aligned with MGSCM's notion of sustainability in the context of maritime supply chain. Thus, in the subsequent chapters, we would discuss on preceding literature with regard of GSCM definition and theory as well as MGSCM notion conceptualized from GSCM to support this paper.

BACKGROUND

GSCM as a Novel Concept of Sustainability

The issue of sustainability in the context of supply chain management (SCM) has been discussed using various terms in the prior literatures. A number of modern literature reviews on GSCM and sustainability have been published in this respect e.g., (Abbasi & Nilsson, 2012; Carter & Rogers, 2008; Gimenez, Sierra, & Rodon, 2012; Sarkis, Zhu, & Lai, 2011; Seuring & Muller, 2008). Early sustainability practices and concepts tended to stress on environmental impacts but, gradually many latest studies adopted the triple bottom line impact that includes environment, economic, and social dimension. Although similar to the concept of SCM, the boundary of GSCM concept is dependent on the supply chain concurrently with the product; however, adding the "green" and "sustainable" component to SCM involves further aspect of addressing the influence and conceptual relationships between SCM and the environment. Consequently, GSCM distinguish the disproportionate of supply chain processes towards environmental impact in an organization. GSCM practices in this sense, are acting as an environmental and operational structure of improvement to catalyze an operational planning that numerous of organizations nowadays are promoted to tackle (Rao & Holt, 2000). Accordingly, a diverse of definitions of GSCM present in the literature. The subsequent paragraph sum ups some of the definitions of GSCM that have emerged in

the recent literature that can be linked towards this paper's conceptualization of MGSCM concept. For examples GSCM can be interpreted as:

1. Network management of sustainable supply chain (Cruz & Matsypura, 2009);
2. Corporate social responsibility network in sustainable supply and demand (Cruz & Matsypura, 2009; Kovacs, 2004);
3. Environmental management of supply chain (Sharfman, Shaft, & Anex, 2009);
4. Green procurement and green purchasing (Günther & Scheibe, 2006; Min & Galle, 1997);
5. Environmental purchasing (Zsidisin & Siferd, 2001);
6. Green and environmental logistics (Murphy & Poist, 2000); and
7. Sustainability in supply chains (Bai & Sarkis, 2013; Linton, Klassen, & Jayaraman, 2007).

Based on GSCM definition mentioned, GSCM is a resolute effort throughout the organization and is more than plainly putting some green dimensions in place, but rather a consistent and holistic enhancement and improvement of the environmental dimension in all levels of internal and external organizational management perspectives (Davies and Hochman, 2007). GSCM in general can also be used practically in any industrial and business context within the management structural framework. The research on sustainability issues in supply chain has been conducted by numerous scholars in the past regarding the theory and practical contribution in various industrial backgrounds such as manufacturing field (Zhu & Sarkis, 2004), product innovation (Lee & Kim, 2011), organizational management (Alfred & Adam, 2009), automotive (Luthra, Kumar, Kumar, & Haleem, 2011; Olugu & Wong, 2011) and other fields. Therefore, GSCM can also falls within the purview of increasing literature on maritime perspective of supply chain and operational management outlooks.

MGSCM as a Concept of Sustainable Maritime Supply Chain

In the perspective of GSCM in maritime supply chain, various literatures emphasized sustainability within maritime operations in diverse terms. Psaraftis (2016) used the term green maritime logistics which defines as an effort to achieve adequate environmental performance in the maritime supply chain, while at the same time fulfilling traditional economic performance dimension. While Cheng, Zanjirani Farahani, Lai and Sarkis (2015) defined sustainable maritime supply chain as sustainable maritime SC which means the integration of maritime organizational units (ports, shipping companies, etc.) in a supply chain system and organization of materials (container, bulk and general cargoes), information, and monetary flows in order to accomplish customer needs while at the same time improving the economic competitiveness through conformity with regulations to control social and environmental impacts. The conceptual term “green shipping practice” (GSP) in defining sustainability has also becoming prominently used in the context of sustainable maritime supply chain to address this rising issue. As Lun et al. (2013) stated, GSP capability focuses mainly on enhancing efficiency in operational condition as well as reducing harmful impact of supply chain actions towards environment through internal and external GSCM capabilities. While others defined sustainability in maritime supply chain as business approaches and actions that meet the current and future needs of the maritime sector and its stakeholders while protecting and sustaining human and natural resources (AAPA, 2007). Additionally, Lai et al. (2011) proposed a theoretical framework with a few propositions to define sustainable maritime practices in the maritime context. They identified and defined six dimensions of sustainable maritime operations, which include organizational policy and procedure, maritime documentation, maritime equipment, shipper cooperation, maritime materials, and shipping design and compliance. Other researchers

have also used the term green management to link sustainability in maritime supply chain in prior literature. Zhu, Sarkis and Lai (2007) deposited that environmentally sustainable management, or the so-called ‘green management’, has also become known as an imperative management topic for organizations to attain profitable operation and market share while at the same time commit towards environmental protection. According Shrivastava (1995) organization’s improvement in competitive positions and performance could be increased by adopting green management practice (GMP) and concurrently decrease the harmful effects of their operations on the natural environment at the same time. Furthermore, sustainability in maritime supply chain in additional literature review encompasses the need to understand the basic conceptual ground of green supply chain management (GSCM) to be integrated in maritime industry. Finally, it is beyond this modest scope of paper to elaborate much on other researches on GSCM, but it is understandable at this point that incorporating environmental concerns into management approaches has become progressively significant for organization to achieve competitive advantage and lessen the environmental impact. Thus, in this paper, we concluded that the GSCM concept can be extended into the term maritime green supply chain management (MGSCM) that refers the integration of GSCM in maritime context; and define it as integrating environmental concerns into the inter-organizational practices of SCM in maritime context to ease the flow of operation in maritime supply chain system.

Conceptualizing MGSCM via GSCM Concept

In this section, we concluded that the dimensions of sustainability of MGSCM are needed to be drawn for successful completion of GSCM. Thus based of understanding from previous chapter, we defined MGSCM’s notion in this paper based on five dimensions of GSCM, that is, green information and communication system (GICS), green value added



logistic service (GVALS), green supply chain integration practice (GSIP), shipping design and compliance (SDC) and green financial flow (GFF) taken from GSCM literatures. We concluded that the conceptual notion of each MGSCM dimension can be presented and defined as:

1. **Green Information and Communication Systems (GICS):** Defined as the systematic application of sustainability in various processes of IT and communication management in order to reduce environmental impact (Kehoe & Boughton, 2001; Prajogo & Olhager, 2012; Swaminathan & Tayur, 2003). According to Preuss (2002), GSCM can only become a strategic organizational capability if it could simultaneously monitor manufacturing, purchasing and selling process collectively together. The benefit of implementing the information system integration technology in maritime supply chain helps the organizations to synchronize supply chain process and close collaboration with partners through the integration of data related information to the production schedules, managing item in inventory, demand forecasting, as well as managing the production quality (Li, Yang, Sun, & Sohal, 2009). It is also wise to approach the information flow of GICS in the perspective of just in time (JIT) context within the sector. According to Swafford, Ghosh and Murthy (2008) information technology supports the efficiency of information flow in term of integrated JIT schedules and set up the effective information links that help to reduce the shipment discrepancies time. The advantages of this system of GICS would provide the maritime industry with better management mechanism of storing, accessing, sharing as well as analyzing the information needed in real time and with high efficiency.
2. **Green Value Added Logistic Services (GVALS):** Defined as the systematic application of sustainability in various processes

value added logistic in supply chain to reduce ecological impact that encompasses diverse dimensions of emotional element, social and functional value (Sweeney & Soutar, 2001). From broader perspective, GVALS can be considered as a component of GSCM that intend at mixing together environmental philosophy into conventional SCM and logistics management through product design (in production phase), supplier assortment and selection as well as material sourcing, inbound and outbound shipping, manufacturing processes, waste reduction control, product wrapping and packaging, distribution to end customers, and reuse and end-of-life product returns for recycling (Beamon, 1999; Linton et al., 2007; Srivastava, 2007). GVALS practices like the employment of green and recycled packaging materials as well as green packaging designs and techniques, could largely diminish packaging waste and cost associated using conventional method (Hung Lau, 2011) if extended properly into maritime operations. From maritime logistic perspective, GVALS practices such as optimization and consolidation of orders, schedules as well as routes taken in shipping operation could reduce the delivery frequency and ultimately cut the fuel consumption (Rao, Grenoble, & Young, 1991). Finally, the organization could implement the GVALS by extending the scope towards the employment of fuel-efficient transportations as well as alternative energy sources to decrease environmental impact (Yusuf et al., 2013).

3. **Green Supply Chain Integration Practices (GSIP):** Defined as the systematic approach of integrating sustainability in various processes of supply chain in order to reduce impact to environment and gain efficiency to the extent of intra-organizationally or inter-organizationally processes to achieve significance value to the final customer demand with cost effectiveness (Frohlich

& Westbrook, 2001; Van Der Vaart & Van Donk, 2004). Handfield, Walton, Sroufe, & Melnyk (2002) suggested that throughout the value chain process, environmental sustainability efforts should be integrated accordingly to enhance the value creation with supply chain partners. As Vasileiou and Morris (2006) added that integration process in GSCM such as friendly processes, products, and services requires a combination and unification efforts by all component members of supply chain to avoid sub-optimization at the partner level.

4. **Shipping Design and Compliance (SDC):** Defined as the systematic approach of sustainability in various processes of shipping design as well as conformity with sustainable compliance in order to decrease impact to environment and gain efficiency (Chang, 2012). SDC emphasis a preventive focus on maritime technologies and design to mitigate the environmental harms caused in maritime operations (Lai, Wong, Lun, & Cheng, 2013). The dimensions of compliance in SDC involve energy saving measures, shipping equipment reuse, recycling, and recovery, targeting to mitigate ecological harm due to maritime activities. SDC is valuable for improving operational efficiency through waste elimination in the supply chain processes (Hart and Ahuja, 1996), increasing financial savings through reusing of shipping equipment and reducing redundant processes through efficient waste management. SDC in maritime context helps to manage and prevent emissions and effluents in compliance with environmental regulations for general performance gains.

5. **Green Financial Flow (GFF):** Defined as the systematic approach of sustainability in various processes of financial management and accounting in order to reduce impact to environment and improve sustainability in a long term (Cairns, 2000). It includes the commitment from management level,

the identification of monetary measurement to measure traditional private internal costs that directly affect the green bottom line of the balance sheet. In this sense, the cost would be the direct costs, such as raw materials and labours, which are attributed to a production items or department and indirect costs, or overheads, such as rent, administration, depreciation, fuel and power. Above all, externalities factor such as social, operational, economic and environmental costs that impact the external environment of organization operation must also be taken into account in green financial flow. For MGSCM capabilities, green financial flow consists of incorporating monetary savings from new cleaner technologies consequential in lower pollution and better health, new markets and using alternative of raw materials or operation processes in maritime supply chain system. Using a framework of green accounting, green financial flow in MGSCM would mean that investment decisions in managing supply chain are made by comparing the overall private and social costs against the private and social benefits. In this sense, via a lifecycle assessment, maritime companies can make decision based on environmental impacts calculation at each stage of supply chain system, from production, transshipment and final disposal of product or services to enhance green performance.

FUTURE RESEARCH DIRECTIONS

Although many studies (e.g., case studies and survey-based empirical methods) have been carried out in GSCM context, they have not completely dealt with each and every aspect of MGSCM. In this regards, there are some areas around MGSCM dimensions that still require further empirical study to conceptualize the notion of MGSCM in this paper. Therefore, future detailed empirical investigation is needed to be carried out



in maritime supply chain domain to truly define the sustainable notion of MGSCM based on the growing literature of GSCM dimensions. Future studies should be included other GSCM practices and dimensions in exploring accountability of the whole concept of GSCM empirically to support the sustainable notion of MGSCM. We are optimistic that the novel MGSCM dimensions can be adopted and materialized in the future. At such, for future studies, the conceptual notion of MGSCM dimensions can also be extended into performance related study that focusing on outcome measurement of economic, environmental, operational and social performance.

CONCLUSION

With regard to the rising global awareness of environmental protection, businesses have employed their GSCM to improve their core competitive advantage. MGSCM derived from GSCM capabilities is a progressively widely-diffused notion that is seeking to improve environmental performance in maritime supply chain context. MGSCM practices, which can be viewed as a cross-organizational and closed loop capability that reduces the ecological impact of maritime operation without sacrificing quality, cost, reliability, performance or energy utilization efficiency. We assumed that, by engaging in MGSCM-related knowledge and capability. Not only does reduction of pollutants and decrease of cost; it also gives the organization involved in improved sustainable competitiveness. Even though no empirical evidence presented in this paper, we are positive that MGSCM dimensions of GICS, GVALS, GSIP, SDC and GFF can be translated into sustainable notion of MGSCM based on general understanding of GSCM literatures. This is due to the fact that the key themes that came out of the MGSCM literature derived from GSCM over the last decades are more or less the same holistic concept of greening the product design, material management, manufacturing process, distribution

and marketing. By the assessing the MGSCM notion of organizational capabilities, this paper has establish new theoretical ground that contributed significance knowledge towards increasing body of literatures in green management.

REFERENCES

- AAPA (American Association of Port Authorities). (2007). *D-11: Embracing the concept of sustainability as a standard practice for ports and the association*. Retrieved from http://aapa.files.cms-plus.com/PDFs/sustainability_resolutions.pdf
- Abbasi, M., & Nilsson, F. (2012, March). Themes and challenges in making freight Transport sustainable: A logistics service providers' perspective. In *53rd Annual Transportation Research Forum*, (No. 207087). Transportation Research Forum.
- Alfred, A. M., & Adam, R. F. (2009). Green management matters regardless. *The Academy of Management Perspectives*, 23(3), 17–26. doi:10.5465/AMP.2009.43479261
- Bai, C., & Sarkis, J. (2013). Flexibility in reverse logistics: A framework and evaluation approach. *Journal of Cleaner Production*, 47, 306–318. doi:10.1016/j.jclepro.2013.01.005
- Beamon, B. M. (1999). Designing the green supply chain. *Logistics Information Management*, 12(4), 332–342.
- Boykoff, M. T. (2009). We speak for the trees: Media reporting on the environment. *Annual Review of Environment and Resources*, 34(1), 431–457. doi:10.1146/annurev.enviro.051308.084254
- Boykoff, M. T., & Yulsman, T. (2013). Political economy, media, and climate change: Sinews of modern life. *Wiley Interdisciplinary Reviews: Climate Change*, 4(5), 359–371. doi:10.1002/wcc.233



- Cairns, R. D. (2000). Sustainability accounting and green accounting. *Environment and Development Economics*, 5(01), 49–54. doi:10.1017/S1355770X00000048
- Cariou, P. (2011). Is slow steaming a sustainable means of reducing CO₂ emissions from container shipping? *Transportation Research Part D, Transport and Environment*, 16(3), 260–264. doi:10.1016/j.trd.2010.12.005
- Carter, C. R., & Rogers, D. S. (2008). A framework of sustainable supply chain management: Moving toward new theory. *International Journal of Physical Distribution & Logistics Management*, 38(5), 360–387. doi:10.1108/09600030810882816
- Chang, C. C. (2012). Marine energy consumption, national economic activity, and greenhouse gas emissions from international shipping. *Energy Policy*, 41, 843–848. doi:10.1016/j.enpol.2011.11.066
- Cheng, T. C. E., Zanjirani Farahani, R., Lai, K. H., & Sarkis, J. (2015). Sustainability in maritime supply chains: Challenges and opportunities for theory and practice. *Transportation Research Part E, Logistics and Transportation Review*, 78, 1–2. doi:10.1016/j.tre.2015.03.007
- Corbett, J. J., Wang, H., & Winebrake, J. J. (2009). The effectiveness and costs of speed reductions on emissions from international shipping. *Transportation Research Part D, Transport and Environment*, 14(8), 593–598. doi:10.1016/j.trd.2009.08.005
- Cruz, J. M., & Matsypura, D. (2009). Supply chain networks with corporate social responsibility through integrated environmental decision-making. *International Journal of Production Research*, 47(3), 621–648. doi:10.1080/00207540701513901
- Davies, J., & Hochman, S. (2007). The greening of the supply chain. *Supply Chain Management Review*, 11(5), 13–14.
- Frohlich, M. T., & Westbrook, R. (2001). Arcs of integration: An international study of supply chain strategies. *Journal of Operations Management*, 19(2), 185–200. doi:10.1016/S0272-6963(00)00055-3
- Gimenez, C., Sierra, V., & Rodon, J. (2012). Sustainable operations: Their impact on the triple bottom line. *International Journal of Production Economics*, 140(1), 149–159. doi:10.1016/j.ijpe.2012.01.035
- Giziakis, C., & Christodoulou, A. (2012). Environmental awareness and practice concerning maritime air emissions: The case of the Greek shipping industry. *Maritime Policy & Management*, 39(3), 353–368. doi:10.1080/03088839.2012.671543
- Grove, S. J., Fisk, R. P., Pickett, G. M., & Kangun, N. (1996). Going green in the service sector: Social responsibility issues, implications and implementation. *European Journal of Marketing*, 30(5), 56–66. doi:10.1108/03090569610118777
- Günther, E., & Scheibe, L. (2006). The hurdle analysis. A self-evaluation tool for municipalities to identify, analyse and overcome hurdles to green procurement. *Corporate Social Responsibility and Environmental Management*, 13(2), 61–77. doi:10.1002/csr.92
- Handfield, R., Walton, S. V., Sroufe, R., & Melnyk, S. A. (2002). Applying environmental criteria to supplier assessment: A study in the application of the Analytical Hierarchy Process. *European Journal of Operational Research*, 141(1), 70–87. doi:10.1016/S0377-2217(01)00261-2
- Hart, S. L., & Ahuja, G. (1996). Does it pay to be green? An empirical examination of the relationship between emission reduction and firm performance. *Business Strategy and the Environment*, 5(1), 30–37. doi:10.1002/(SICI)1099-0836(199603)5:1<30::AID-BSE38>3.0.CO;2-Q

- Hart, S. L., & Ahuja, G. (1996). Does it pay to be green? An empirical examination of the relationship between emission reduction and firm performance. *Business Strategy and the Environment*, 5(1), 30–37. doi:10.1002/(SICI)1099-0836(199603)5:1<30::AID-BSE38>3.0.CO;2-Q
- Hoffmann, P. N., Eide, M. S., & Endresen, Ø. (2012). Effect of proposed CO2 emission reduction scenarios on capital expenditure. *Maritime Policy & Management*, 39(4), 443–460. doi:10.1080/03088839.2012.690081
- Hung Lau, K. (2011). Benchmarking green logistics performance with a composite index. *Benchmarking: An International Journal*, 18(6), 873–896. doi:10.1108/14635771111180743
- Isaksson, K., Björklund, M., Evangelista, P., & Høge-Brodin, M. (2011). The Challenge and Adoption of Green Initiatives for Transport and Logistics Service Providers. In *The 16th Annual LRN Conference* (pp. 1–10).
- Kehoe, D., & Boughton, N. (2001). Internet based supply chain management: A classification of approaches to manufacturing planning and control. *International Journal of Operations & Production Management*, 21(4), 516–525. doi:10.1108/01443570110381417
- Kovacs, G. (2004). Framing a demand network for sustainability. *Progress in Industrial Ecology, an International Journal*, 1(4), 397–410.
- Lai, K. H., Lun, V. Y., Wong, C. W., & Cheng, T. C. E. (2011). Green shipping practices in the shipping industry: Conceptualization, adoption, and implications. *Resources, Conservation and Recycling*, 55(6), 631–638. doi:10.1016/j.resconrec.2010.12.004
- Lai, K. H., Wong, C. W., Lun, Y. V., & Cheng, T. C. E. (2013). Shipping design for compliance and the performance contingencies for shipping firms. *Transportation Research Part E, Logistics and Transportation Review*, 55, 74–83. doi:10.1016/j.tre.2013.03.004
- Lee, K. H., & Kim, J. W. (2011). Integrating suppliers into green product innovation development: An empirical case study in the semiconductor industry. *Business Strategy and the Environment*, 20(8), 527–538. doi:10.1002/bse.714
- Lee, S. M., Tae Kim, S., & Choi, D. (2012). Green supply chain management and organizational performance. *Industrial Management & Data Systems*, 112(8), 1148–1180. doi:10.1108/02635571211264609
- Li, G., Yang, H., Sun, L., & Sohal, A. S. (2009). The impact of IT implementation on supply chain integration and performance. *International Journal of Production Economics*, 120(1), 125–138. doi:10.1016/j.ijpe.2008.07.017
- Lirn, T. C., Lin, H. W., & Shang, K. C. (2014). Green shipping management capability and firm performance in the container shipping industry. *Maritime Policy & Management*, 41(2), 159–175. doi:10.1080/03088839.2013.819132
- Lun, Y. V., Lai, K. H., & Cheng, T. E. (2013). An evaluation of green shipping networks to minimize external cost in the Pearl River Delta region. *Technological Forecasting and Social Change*, 80(2), 320–328. doi:10.1016/j.techfore.2012.08.014
- Luthra, S., Kumar, V., Kumar, S., & Haleem, A. (2011). Barriers to implement green supply chain management in automobile industry using interpretive structural modeling technique: An Indian perspective. *Journal of Industrial Engineering and Management*, 4(2), 231–257. doi:10.3926/jiem.2011.v4n2.p231-257
- Min, H., & Galle, W. P. (1997). Green purchasing strategies: Trends and implications. *Journal of Supply Chain Management*, 33(3), 10.
- Murphy, P. R., & Poist, R. F. (2000). Green logistics strategies: an analysis of usage patterns. *Transportation Journal*, 5–16.

- Olugu, E. U., & Wong, K. Y. (2011). Evaluation of green supply chain management practices in the Malaysian automotive industry. *International Journal of Services and Operations Management*, 9(2), 245–258. doi:10.1504/IJSOM.2011.040710
- Ostrom, E. (2008). The challenge of common-pool resources. *Environment: Science and Policy for Sustainable Development*, 50(4), 8–21. doi:10.3200/ENVT.50.4.8-21
- Prajogo, D., & Olhager, J. (2012). Supply chain integration and performance: The effects of long-term relationships, information technology and sharing, and logistics integration. *International Journal of Production Economics*, 135(1), 514–522. doi:10.1016/j.ijpe.2011.09.001
- Preuss, L. (2002). Green light for greener supply. *Business Ethics (Oxford, England)*, 11(4), 308–317. doi:10.1111/1467-8608.00290
- Psaraftis, H. N. (2016). Green Maritime Logistics: The Quest for Win-win Solutions. *Transportation Research Procedia*, 14, 133–142. doi:10.1016/j.trpro.2016.05.049
- Rao, K., & Grenoble, W. L. IV. (1991). Modelling the effects of traffic congestion on JIT. *International Journal of Physical Distribution & Logistics Management*, 21(2), 3–9. doi:10.1108/09600039110005178
- Rao, P., & Holt, D. (2005). Do green supply chains lead to competitiveness and economic performance? *International Journal of Operations & Production Management*, 25(9), 898–916. doi:10.1108/01443570510613956
- Revkin, A. C. (2009). *Disaster Hot Spots on a Crowding Planet*. Academic Press.
- Rosenthal, E. (2009). Biggest obstacle to global climate deal may be how to pay for it. *International herald tribune. The New York Times*.
- Sarkis, J., Zhu, Q., & Lai, K. H. (2011). An organizational theoretic review of green supply chain management literature. *International Journal of Production Economics*, 130(1), 1–15. doi:10.1016/j.ijpe.2010.11.010
- Seuring, S., & Müller, M. (2008). From a literature review to a conceptual framework for sustainable supply chain management. *Journal of Cleaner Production*, 16(15), 1699–1710. doi:10.1016/j.jclepro.2008.04.020
- Sharfman, M. P., Shaft, T. M., & Anex, R. P. Jr. (2009). The road to cooperative supply-chain environmental management: Trust and uncertainty among pro-active firms. *Business Strategy and the Environment*, 18(1), 1–13. doi:10.1002/bse.580
- Shrivastava, P. (1995). Environmental technologies and competitive advantage. *Strategic Management Journal*, 16(S1), 183–200. doi:10.1002/smj.4250160923
- Sulaiman, O. O., Akmar, N. I., Michel, B., Stark, N., Azman, H., & Kader, A. S. A. (2013). Pilot Study for Quantification of Emissions of Green House Gas for Decision Support towards International Maritime Organization (IMO) Rule Making. *Global Journal of Science Frontier Research*, 13(1).
- Swafford, P. M., Ghosh, S., & Murthy, N. (2008). Achieving supply chain agility through IT integration and flexibility. *International Journal of Production Economics*, 116(2), 288–297. doi:10.1016/j.ijpe.2008.09.002
- Swaminathan, J. M., & Tayur, S. R. (2003). Tactical planning models for supply chain management. *Handbooks in Operations Research and Management Science*, 11, 423–454. doi:10.1016/S0927-0507(03)11008-0

Sweeney, J. C., & Soutar, G. N. (2001). Consumer perceived value: The development of a multiple item scale. *Journal of Retailing*, 77(2), 203–220. doi:10.1016/S0022-4359(01)00041-0

Van der Vaart, T., & Van Donk, D. P. (2004). Buyer focus: Evaluation of a new concept for supply chain integration. *International Journal of Production Economics*, 92(1), 21–30. doi:10.1016/j.ijpe.2003.10.002

Vasileiou, K., & Morris, J. (2006). The sustainability of the supply chain for fresh potatoes in Britain. *Supply Chain Management: An International Journal*, 11(4), 317–327. doi:10.1108/13598540610671761

Wong, C. W., Lai, K. H., & Cheng, T. C. E. (2009). Complementarities and alignment of information systems management and supply chain management. *International Journal of Shipping and Transport Logistics*, 1(2), 156–171. doi:10.1504/IJSTL.2009.024493

Wong, C. W., Lai, K. H., Shang, K. C., Lu, C. S., & Leung, T. K. P. (2012). Green operations and the moderating role of environmental management capability of suppliers on manufacturing firm performance. *International Journal of Production Economics*, 140(1), 283–294. doi:10.1016/j.ijpe.2011.08.031

Yang, C. C., Marlow, P. B., & Lu, C. S. (2009). Knowledge management enablers in liner shipping. *Transportation Research Part E, Logistics and Transportation Review*, 45(6), 893–903. doi:10.1016/j.tre.2009.05.003

Yusuf, Y. Y., Gunasekaran, A., Musa, A., El-Berishy, N. M., Abubakar, T., & Ambursa, H. M. (2013). The UK oil and gas supply chains: An empirical analysis of adoption of sustainable measures and performance outcomes. *International Journal of Production Economics*, 146(2), 501–514. doi:10.1016/j.ijpe.2012.09.021

Zhu, Q., & Sarkis, J. (2004). Relationships between operational practices and performance among early adopters of green supply chain management practices in Chinese manufacturing enterprises. *Journal of Operations Management*, 22(3), 265–289. doi:10.1016/j.jom.2004.01.005

Zhu, Q., Sarkis, J., & Lai, K. H. (2007). Initiatives and outcomes of green supply chain management implementation by Chinese manufacturers. *Journal of Environmental Management*, 85(1), 179–189. doi:10.1016/j.jenvman.2006.09.003 PMID:17084502

Zsidisin, G. A., & Siferd, S. P. (2001). Environmental purchasing: A framework for theory development. *European Journal of Purchasing & Supply Management*, 7(1), 61–73. doi:10.1016/S0969-7012(00)00007-1

ADDITIONAL READING

Davarzani, H., Fahimnia, B., Bell, M. G., & Sarkis, J. (2015). A Review of the Literature of Green Ports and Maritime Logistics. In *Green Logistics and Transportation* (pp. 149–158). Springer International Publishing. doi:10.1007/978-3-319-17181-4_9

Esmemr, S., Ceti, I. B., & Tuna, O. (2010). A simulation for optimum terminal truck number in a Turkish port based on lean and green concept. *The Asian Journal of Shipping and Logistics*, 26(2), 277–296. doi:10.1016/S2092-5212(10)80006-9

Gunasekaran, A., Subramanian, N., & Rahman, S. (2015). Green supply chain collaboration and incentives: Current trends and future directions. *Transportation Research Part E, Logistics and Transportation Review*, 74, 1–10. doi:10.1016/j.tre.2015.01.002

Jafarzadeh, S., & Utne, I. B. (2014). A framework to bridge the energy efficiency gap in shipping. *Energy*, 69, 603–612. doi:10.1016/j.energy.2014.03.056

Jepsen, M. K., Løfstedt, B., Plum, C. E., Pisinger, D., & Sigurd, M. M. (2011). A path based model for a green liner shipping network design problem. *Proceedings of the IAENG, ICOR, 2011*.

Lam, J. S. L. (2015). Designing a sustainable maritime supply chain: A hybrid QFD–ANP approach. *Transportation Research Part E, Logistics and Transportation Review*, 78, 70–81. doi:10.1016/j.tre.2014.10.003

Lirn, T. C., Jim Wu, Y. C., & Chen, Y. J. (2013). Green performance criteria for sustainable ports in Asia. *International Journal of Physical Distribution & Logistics Management*, 43(5/6), 427–451. doi:10.1108/IJPDLM-04-2012-0134

Rigot-Muller, P., Lalwani, C., Mangan, J., Gregory, O., & Gibbs, D. (2013). Optimising end-to-end maritime supply chains: A carbon footprint perspective. *The International Journal of Logistics Management*, 24(3), 407–425. doi:10.1108/IJLM-01-2013-0002

KEY TERMS AND DEFINITIONS



Green Supply Chain Management (GSCM):

Is about assurance to all usual and traditional supply chain deliveries while integrating the social and environmental concern in every phase of supply chain system to ultimately diminish or lessen product's ecological footprint.

Maritime Green Supply Chain Management (MGSCM): Defined as integrating environmental concerns and practices into the inter-organizational practices of SCM in maritime context to ease the flow of operation in maritime supply chain system.

Maritime Supply Chain: Defined as the movement of cargoes and related support service involving two substantial locations using maritime and land transportations.

Supply Chain Management (SCM): Defined as managing flow of materials, information, and finances as they move in a process from supplier to manufacturer to wholesaler to retailer to consumer.

Offshoring IT

Susan Cockrell

Austin Peay State University, USA

Terry Stringer Damron

Austin Peay State University, USA

Amye M. Melton

Austin Peay State University, USA

Alan D. Smith

Robert Morris University, USA

INTRODUCTION

Outsourcing as an Academic Disciple

Outsourcing as may be described as procuring goods or services needed by a business under contract with an external supplier. If one were to assess this definition in a literal sense then any form of trade could be viewed as outsourcing. We have narrowed our research to focus on information technology (IT) outsourcing. The relatively short history of IT seems to correlate with the defined term, which was only added to some standard dictionaries since 1979. This fact would indicate a brief period of existence before the inclusion. Previous research (Hamidi, Farahmand, Sajjadi, & Nygard, 2012; Kumar, Shankar, & Yadav, 2011; Mathirajan, Manoj, & Ramachandran, 2011; More & Babu, 2012) generally supports this historic trend, but a hard date indicating the birth of outsourcing probably does not truly exist. One of the earliest and most prominent example of IT outsourcing is with Electronic Data Systems (EDS) in 1962, which resulted in a very slow adaption.

The following chapter will look at outsourcing through several different lenses. Initially, some of the historical and current trends that exist in outsourcing IT, followed by some of the major services have been and are being utilized

by firms looking to outsource, outside tax incentives such as tax inversions. Next, an assessment of the benefits and risks of engaging in an IT outsourcing relationship will be discussed. In this section, some of the classical reasoning behind management decisions to outsource and the risks that managers should be aware of when entering into this relationship. Finally, this research effort returns to EDS and others to profile applications of IT outsourcing before moving on to look at the possibilities in the future of outsourcing.

The focus on process efficiency and cost control efforts has encouraged firms to look for creative means to structure their business efforts, especially in, customer service, IT, and lean operations (Beldona & Tsatsoulis, 2010; Bhamu, Khandelwal, & Sangwan, 2013). Outsourcing has become a common practice, especially in the area of information systems/technologies (IS/IT) among large and medium size enterprises (MSE) globally. For example, Spanish firms (up to 83.6%) outsource at least some IS function nationally (Gonzalez, Gasco, & Llopis, 2010). Although a growing trend, there has been limited research on how IT outsourcing has changed the traditional characteristics of this industry and the attitudes of existing IT professionals on the phenomenon over a period of time.

The aim of the present research is to specifically evaluate how IT outsourcing has affected

the following areas of research (i.e., the time required for IS managers to perform their jobs; the characteristics of the IS manager's working environment or post; the type of knowledge and skills required to carry out the IS manager's professional activities. A number of researchers (Pettersson & Segerstedt, 2011; Pradhananga, Hanaoka, & Sattayaprasert, 2011; Von Haartman, 2012) have investigated IT within an industrial focus, as compared to the service sector, which is differentiated as general services and IT-intensive (i.e., financial institutions, insurance companies, the tourism sector, attorneys, supply chain integration and collaboration, to name a few). The operations manager, on-the-other-hand, generally emphasizes the strategic consequences of job satisfaction and productivity as a function of industrial outsourcing.

This is because of rapid technological change, high investments firms have made in IT and the great interest in outsourcing have made these executives the target of many surveys. (Gonzalez, et al., 2010)

The results from this study suggest a weak association between IS outsourcing and job performance, work post, and required knowledge and skills (Willcocks, Hindle, Feeny, & Lacity, 2004). In relation to job performance, the influence that outsourcing has had on the amount of time IS managers dedicate to their various tasks (Gonzales, et al., 2010). Outsourcing has generally had little direct influence on the time IS managers need to carry out their activities, although a redistribution of that time has taken place. Indeed, it is easy to see that more time is needed for external relations. Conversely, less time is needed for internal relations because the number of tasks coordinated within the firm itself and within the actual department is considerably smaller. The potential impacts that outsourcing places on an IS manager's working post is generally positive, as many managers feel that their job has acquired more added value because they have

fewer internal tasks and believe that their work is more valuable. As a result, their employment satisfaction level increases and typically have a higher degree of autonomy. Outsourcing does not mean that the IS manager needs knowledge and skills to a lesser extent; if anything, the IS manager needs them more in order to negotiate good contracts. Except for finance-related knowledge (Jain & Kohli, 2009; Kanninen, Piche, & Mikkonen, 2009; Karjaluo, Jarvenpaa, & Kauppi, (2009), IS outsourcing requires a greater amount of knowledge, as IS managers help provide a communication bridge between their organization and the supplier it terms of negotiations. IT knowledge cannot be neglected since awareness of the services and possibilities provided by IT is essential if managers are to know what is being contracted and what the chances for improvement are needed. Of course, there is always the potential to exploit an outsourced IT service provider as a strategic project partner.

When studying the effect of position outsourcing on this industry in major core areas, there is general consensus that IS outsourcing requires more managerial skills, but less time strategy-related activities; more time systems management, better relations and staff management, sharper tactical management techniques and supportive of more positive work environment (Daim, Basoglu, & Tanoglu, 2010; Daramola, Oladipupo, & Musa, 2010; Dominic, Goh, Wong, & Chen, 2010). Obviously, smaller firms in terms of number of workers and sales volume generally have smaller IS departments while focus for IS managers in the larger enterprises was to improve their performance rather than reduce their workload (Gonzales, et al., 2010). It was found that the service sector and the IT-intensive service sector are the most sophisticated in terms of their IS, which is why they exploit the advantages of outsourcing either to liberate work or to enrich the job, but they never see outsourcing as an "enemy" that might impoverish and damage the IS manager's working post (Gonzales, et al., 2010).



Outsourcing Trends

The growing movement toward outsourcing IT departments reveals several trends with IT-intensive industries (Kim & Tadisina, 2010; Marthandan & Tang, 2010). Many companies are utilizing a global sourcing approach to offshoring IT. This global model involves a number of geographically dispersed groups of IT professionals working closely together and allows project work to continue through many time zones. An internal team works to set requirements and specifications, a third-party team in the same time zone works on development, and offshore personnel perform a variety of functions at a fraction of domestic costs. The 24/7 approach is working especially well for companies who invest heavily in new project developments. A globalization model allows companies to respond quickly to opportunities. The added time and risk of hiring people for new development work is less of a consideration and the initial design work is typically less costly with this type of arrangement. An onshore/offshore outsourcing model is being utilized by many companies who allocate most of their standardized and routine activities. This practice enables IT departments to better utilize in-house development teams to focus on expanding capabilities and achieving better results from their systems.

Multisourcing involves utilizing multiple IT service providers in many countries to achieve goals. Companies try to find the best skills at the best price regardless of function or geographic location. EDS, for example, in their Best Shore service program, work can be performed at over 16 IT facilities in 11 countries (King, 2003). A broader range of services is now available throughout the IT outsourcing landscape. One of the newer segments offered is business process outsourcing (BPO). Offshore providers have typically expanded their services to include human resource, accounting and finance capabilities. Interestingly, such mandate industry specific services (e.g., processing for meter reading in the utility companies) have emerged in a trend

that is referred to as “the industrialization of IT.” The ability of an outsourcing partner to process standardized transactions on a global basis is being applied to many industries including credit card and airline electronic transactions.

Another expansion of services, IT-service providers are even branching off into consulting and training. India-based Wipro, offers consulting services that focuses on training software development (King, 2003). A recent emerging trend in IT outsourcing is offshore handoffs. This involves large mature offshore outsourcing vendors, especially in India, subcontracting work to lower cost markets like China and Vietnam. This allows outsourcing vendors to be able to perform services in a way that will enable them to achieve the greatest economies of scale and, thereby, provide the lowest cost to the user. Infrastructure services such as help desk and desktop support are fast growing segments of outsourcing. An increasingly vital function within this segment is network monitoring. Maintaining a secure environment for both the company and its customers is critical to the IT outsourcing relationship (Jensen, 2004).

Undoubtedly, offshoring of IT-intensive services has increased demand for IT management professionals. The need for IT professionals with project management skills is important to the success of outsourcing projects. Many emerging opportunities in IT are for professionals who can work with customers to translate business requirements into technical specifications and work with the outsourcer on implementation. IT professionals, in an outsourcing environment, should have managerial expertise, significant business process knowledge and be skilled in contract negotiating, compliance monitoring as well as accounting and finance (Jain & Kohli, 2009; Kanninen, et al., 2009; Karjalainen, et al., 2009).

This reshuffling of more traditional IT roles is a trend companies should address when designing their outsourcing strategies. Having quality in-house personnel with the abilities to interact effectively with outsourcing partners is integral in fostering successful offshore relationships. The

need for IT professionals to broaden their skills into management is an important trend to consider for human resource departments and individuals in the domestic IT field. Generally, outsourcing has enabled the virtual corporation, a lean organization that performs the functions of planning, collaboration and money management (Browning & Heath, 2009; Cavaleri, 2008; Coates, 2005). The outsourcing of IT is just another inevitability in a highly competitive global environment.

BACKGROUND

Benefits of IT Outsourcing

The economic and political climate in the U.S. is constantly changing, both in terms of governmental support for economic development and closing loop holes for tax inversions (i.e., moving corporate headquarters overseas to take advantage of lower corporate taxes). The U.S. Federal Reserve's recent announcement that it will extend Operation Twist by buying an extra US\$267 billion of long-term Treasury bonds over the next six months (Project Syndicate, 2012). Even though this governmental action is to help the economic recovery, after expanding the monetary policy, the level of GDP is still just about 1% higher than it was nearly several years ago. On September 16, 2011, the U.S. enacted landmark patent legislation. As a result of the Leahy-Smith America Invents Act (AIA) the patent law of the United State will experience the most widespread changes since the initial patent status was first passed in 1790. (Gene, 2012) The primary reason organizations state for outsourcing IT services is improved financial well-being. In general, firms seek to reduce costs, up to 50% or more, and to free up additional capital that can be invested in other segments; overall outsourcing IT has several financial benefits. One benefit of outsourcing IT is that firms are able to concentrate on their competitive advantage and value adding activities. In recent years the specialized firms that offer IT outsourcing services have been able

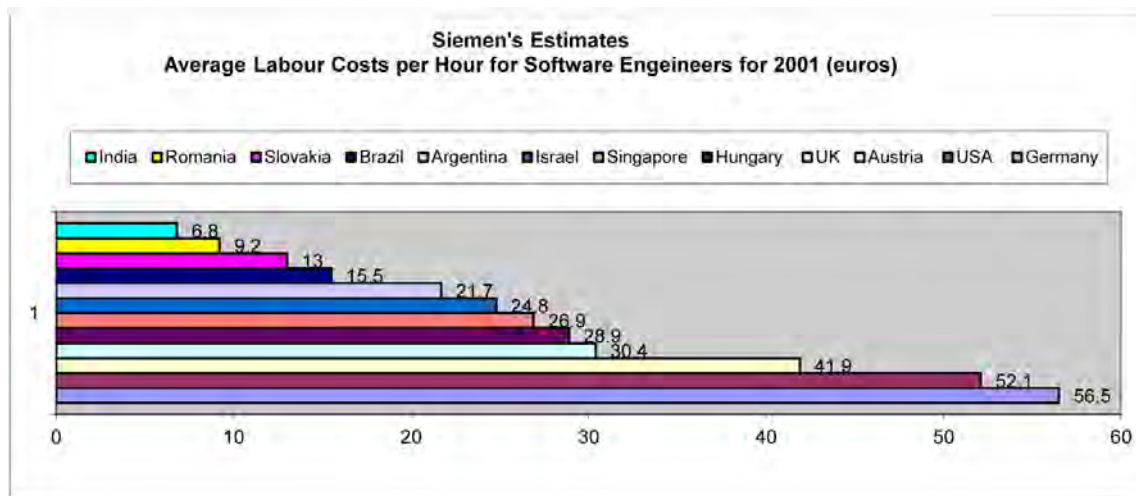
to increase their capacity and intelligence in the IT field. These firms are now offering a higher quality of service to their customers. A second benefit of outsourcing IT is that the practice provides a company with a global presence that can be beneficial in today's expanding business environment. Firms open themselves up to multiple opportunities by electing to outsource IT (Younies, Barhem, & Hsu, (2007). Firms are able to cut operating costs, shift their focus onto their core activities, provide a higher quality of service to their employees and customers, and establish a presence in the global business environment.

Research estimates that U.S. companies would save up to US\$11 billion in 2004 by outsourcing IT to India. The U.S. banking industry alone has saved US\$6 billion to US\$8 billion annually by outsourcing IT to India. In addition to these savings, it has been reported that India would purchase US\$3 billion worth of high tech equipment from the U.S. due to the outsourcing contracts (Pfannenstien & Tsai, 2004). Management can realize significant financial gains through the sale of industry specific equipment required to perform IT services at established standards. Cost savings and new sales are translated into debt reduction and/or expansion by practicing firms (Hall & Liedtka, 2005).

Outsourcing IT has been shown to stimulate growth in the U.S. economy contrary to popular belief. The above-mentioned corporate benefits have had a direct correlation on price for consumers. India's National Association of Software and Services Companies, Nasscom, commissioned a report by Evaluserve, which found that for every US\$100 worth of work sent abroad by U.S. companies, US\$130 to US\$145 would be reinvested in the U.S. economy. In a study performed by the Institute for International Economics in Washington, Catherine L. Mann found that because of outsourcing, IT hardware is 10 to 30% less expensive (Erber and Sayed-Ahmed, p.110). "A McKinsey study found that two thirds of the economic benefit from sending jobs offshore flows back to the



Figure 1. Offshoring and outsourcing of IT industry, adapted from Erber and Sayed-Ahmed (2005)



U.S. in the form of lower prices” (Pfannenstien & Tsai, 2004, p.78).

Financial benefits are felt locally as firms are utilizing the fiscal freedom they receive from outsourcing IT. Companies can leverage the benefits that outsourcing provides when negotiating contracts with firms within their own country. For example, hourly rates in Asia and other emerging markets are anywhere from 30 to 75% lower than they are in the U.S. In Figure 1 Siemen’s has compared the costs of software engineers in different countries. Based on the information provided in Figure 1, companies from the U.S. and Germany can threaten to go to India or Romania if they feel they are not getting a good enough deal at home for IT services.

Initially, cost savings and financial benefits were the primary purpose of outsourcing IT. However outsourcing IT has evolved into one of the most preferred techniques for restructuring a company. Outsourcing in general has allowed companies to shift their focus and concentration onto their core value adding activities. Outsourcing has provided corporations the opportunity to reduce sunk costs in fixed investments. Initially, outsourcing work and payments to lower cost countries can reduce fixed R&D outlays. The “substitution of fixed costs by variable costs gives the outsourcing company greater flexibility in

reallocating its capital” (Erber & Sayed-Ahmed, 2005, p.102). A study in the McKinsey Quarterly found that the U.S. economy has benefited from IT outsourcing because the money saved by outsourcing IT creates new jobs, or as the study puts it redeployed labor (Erber & Sayed-Ahmed, 2005).

The importance of the IT department within a company cannot be ignored. As Hall and Liedtka (2005, p. 200) stated, “A successful IT function contributes significantly to the operating efficiency of a firm.” In 2000 a reported survey found that 25 of the 36 major firms that have achieved level five maturity of the Software Capability Maturity Model (CMM) were located in India. The other 11 were located in the U.S, but were suppliers to the government and were not available to the commercial software market (Pfannenstien & Tsai, 2004). China has been known for years for the “China price” in manufacturing and now India has the “India price” in IT (King, 2005). India has proved that they are highly skilled in the IT field and are able to offer a cheaper rate than most domestic firms as confirmed by the fact that India contains approximately 70% of the level five CMMs in the world.

In today’s business environment having a global presence is a necessity for a firm to survive; and outsourcing provides that presence to

companies. Certain companies have come to the realization that allocating resources offshore through outsourcing is necessary for them to maintain competitiveness in a global economy (Kanninen, Piche, & Mikkonen, 2009; Karjalainen, Jarvenpaa, & Kauppi, 2009; Pfannestien & Tsai, 2004). It is believed that if a company resists the temptation to outsource now they will be hurting their ability to compete globally in the future. Whether it is IT or another related activity, outsourcing provides a company with knowledge and insight into other cultures, which can play a significant role in future decisions and determining future successes or failures.

Risks of Outsourcing IT

There are several risks that company's face today when attempting to outsource information technology to a vendor company. One of most obvious major risk is capability; or how well the vendor can effectively achieve the goals and objectives of the client. Outsourcing projects often have a geographical separation between the two companies, which can cause communication problems. The vendor company might lose focus of the actual goals and objectives are of the client company. A number of researchers have discussed this phenomenon and have suggested that complexity and chaos can easily cause activities to stray from requirements or specifications, resulting in product or service failure. Some of this failure results from expanding or deviating from the original concept, goal, or purpose of the project to be outsourced. Activities such as change requests and the prioritization of the changes may cause the scope of the operations for both companies to alter significantly (Rasolinezhad, 2009; Sakkthivel, 2009). If the scope of the project changes then it is possible for the vendor to lose focus on the overall objectives. The consequences for losing focus of the client's objectives could be disastrous for everyone involved.

Another important risk that is associated with outsourcing is the inability for the client company to maintain security of their technology information (Smith, Smith, Gopalakrishna-Remani, & Offodile, 2012; Smith, 2005; Smith & Lias, (2005). The client should be aware of the ability of the vendor to develop technology that could compromise the security of their information. International laws that regulate the sharing of information make it possible for the vendor company to take sensitive information and provide it to their clients' competitors. Several foreign countries do not have laws that prohibit such sharing of confidential information. This could cost the client an unnecessary amount of problems. For example, the client company could suffer a multitude of losses from customer trust to legal fees. Ultimately, it is important for the client company to maintain their own information security.

A vendor firm could possibly be performing technology outsourcing for industry competitors. Client information could be used against them or leaked to their competitors. Vendors may become capable of integrating forward; developing from a partner to a competitor. Once a vendor is able to procure technology they could enter the market and use it to take business from the client. Management should be willing to take these risks into consideration. Typically firms entering into an outsourcing relationship require that the vendor sign a non-competitive agreement that would prevent them from using their information in any way other than to perform IT services for the client. It can be difficult for companies to enforce breach of contract penalties or fines because of certain international regulations.

Companies engaging in outsourcing should take into account the possibility that the vendor will lose control over project. Loss of control by the vendor could result in complete chaos within the client firm. Parker and Russell (2004) suggested that this possibility can be reduced through "clear responsibilities and clear decision processes and procedures, combined with the formal, regulatory

controls and the more informal social controls can prevent process failures” (p. 60). It is important for managers of the client company to provide specific guidelines and objectives for the vendor to follow to prevent this chaos from occurring. By providing specific guidelines management can predict how the vendor will handle the technology process.

Companies looking to outsource IT face the risk that a vendor lacks experience in dealing with cutting edge information technology. Because there are low barriers to entry for new companies to perform outsourcing activities it is easy for a vendor with little experience to compete for large outsourcing contracts. Appropriate research by management is required to ensure that the vendor can handle the contract. “Even though a firm competes in the field of information technology, there is no guarantee that they are qualified to manage an outsourcing contract” (Sullivan & Ngwenyama, 2005, p.76). Management should verify that the vendor can perform the specific task that they want done. Even though a vendor has experience performing IT services it does not mean that the vendor has experience in what the client wants done. The research process should be in depth enough to reveal the number of experienced and qualified employees that the vendor has employed. A large outsourcing requirement would overburden a vendor, which has only a limited number of experienced employees to perform the service. It is essential for management to investigate these issues before awarding an outsourcing contract.

Client/vendor compatibility is another risk that executives should consider when collaborating on an outsourcing project. “For organizations to work together, it is imperative that there be some degree of cultural fit and compatibility” (Piachaud, 2005, p. 42). Compatibility can occur on several different levels. First, the cultures of both companies play an important role in the success of the project. If the culture of the client and vendor are not similar it is possible to lose focus on what the vendor is trying to accomplish. Second is the service attitude of the client company versus that of the vendor. If the client company is used to providing exceptional

service to their customers and the vendor is not capable of providing that same service then the client could potentially lose customers through poor service. The third level of compatibility is the long-term strategic fit of both companies. The vendor company should be willing to adapt and change according to how the client changes their long-term strategic goals. Since projects can span over several years it is possible that the strategic plans of the client may change throughout the duration of the contract. The vendor needs to be capable of changing along with their client.

Another important risk factor that is associated with outsourcing is the cost of making the decision to have a vendor provide information technology that you cannot provide on your own (Karlsson, 2003; Riis, Johansen, Waehrens, & Englyst, 2007). Typically, many companies outsource to save money. However, there are transaction costs that the outsourcing company needs to be aware of before making this decision. Transaction costs can be anything that is associated with making the outsourcing project happen. “Transaction costs can include setup costs, redeployment or relocation costs, sales tax on equipment or purchases, and leasing costs” (Bahli & Rivard, 2003, p.214). It is difficult to put a dollar value on some of these transaction costs, which makes it a very important aspect for management to consider before making any decisions.

Management plays the most important role in reducing the risks associated with outsourcing any type of project. It is clearly the managers of the client company’s responsibility to effectively motivate and communicate with their employees and with the vendor company. It is management’s fundamental responsibility to understand how the outsourcing process and set up the correct agreement with the vendor. Management needs to understand what can happen, how likely it will occur, and what will be the consequences if it happens when dealing with project risk. Ultimately the executives of the client company are going to be held responsible for any loss that occurs. They are going to have to explain to their shareholders

why outsourcing of their information technology failed; causing them to lose money and potentially customers.

MAIN FOCUS OF THE ARTICLE

Transaction Processing

Companies have been outsourcing information technology as far back as 1962, when EDS, started a technology revolution that is still going strong today. IT and outsourcing is a US\$680 billion global industry. EDS started as a US\$1000 investment and has grown into a US\$20+ billion technology services industry leader. Their IT services have grown to include systems management, systems integration, centralized transaction processing and as the world's largest private digital voice, data and video network. In 1963 Frito Lay became the first firm to sign a five-year agreement with EDS in facilities management. By signing this long-term agreement, EDS was given a more financially secure future. Over the next several years EDS entered into several contracts with medical firms who were outsourcing data processing of medical records. Among the firms signing with EDS were Medicaid in Texas and Blue Shield of California. EDS continues to maintain a long-term relationship with Blue Shield to become a US\$1 million-a-month contract.

In 1975 EDS began looking beyond the United States for contracts in outsourcing. Saudi Arabia, Singapore, Iran, and Pakistan were among the first countries where EDS outsourced information technology. Over the next several years, EDS continued to perform outsourcing for several companies. In 1986 EDS was sold to General Motors. GM expanded the operations of EDS to include projects in Latin America, France, and Japan. In 1987 (on EDS's 25th anniversary) this US\$1000 investment company became a US\$4.4 billion revenue company with operations in computers and communications throughout the world.

The 1990s showed more of the same with companies like Continental Airlines Holding Inc. and Saab in Sweden signing contracts for computer and communications facilities and systems management, respectively. These contracts brought in billions of dollars for long-term operations. 1992 revenues were US\$8.2 billion and EDS was working with GE MasterCard, Blue Cross Blue Shield, and purchasing firms in Germany in the IT services industry. In 2000 EDS won a government contract for information technology to develop and manage the Navy Marine Corps Intranet (NMCI).

EDS has over 65,000 servers and supports 2.5 million desktops globally. It has 350 million customer relationships and interacts with 2.4 billion customers in 48 languages. As its website states, EDS provides the best solutions for executives to maximize return on their IT investments. Its annual revenues today are over US\$20 billion. Currently EDS is the second largest information technology provider in the world.

Over the years several other firms have entered into the industry of information technology outsourcing. Cisco Systems was started in 1984 and is considered a worldwide leader in networking of the Internet. In 2002 Cisco Systems and SBC signed a 3-½ year contract that would lead to many outsourcing projects between the two firms. Cisco provides the networking equipment and SBC assists clients through the construction and management of networks for other companies in the telephone industry (Berniker, p.1). In 2004-2005 Cisco and EDS joined forces to develop a new model to help customers manage information technology resources. Cisco is providing networking expertise to gain new customers and insight into outsourcing and data center markets along with EDS's global presence (Waltner, 2005).

IBM is another company who uses strategic outsourcing services to manage information technology systems. IBM does this through transfer of IT employees and IT assets from companies. In order to gain the trust of companies, IBM works

with the customers through case studies, executive reports and videos to give them an understanding of all that is involved in outsourcing. IBM's employees are strategic partners with clients – they work with the customers to provide tools, processes and methods that best suit each customers' needs. In 2002, for example, IBM Global Services, the world's largest information technology services and consulting provider made over US\$36 billion in revenues. In 2003, for example, IBM was the leader in the European market of IT outsourcing, signing 11 of the 100 top outsourcing deals on the continent. By managing the customers' IT applications and business processes through IBM, customers have become more effective, productive, and saved money. Undoubtedly, it is the potential for the IT services vendor to be a project partner or collaborator with the procuring organization, in the form of a strategic contracting alliance. In this case, goods and services may be traded as "in-kind" transactions, with potential win-win cost savings. Risks still apply of course.

Outsourcing allows firms to structure their focuses on core products with the knowledge that the non-core projects are well managed. Other outsourcing companies include UNISYS, which uses IT outsourcing (ITO), business process outsourcing (BPO), and technology support and maintenance services to deliver information technology to its customers. Hp Invent offers strategic outsourcing with over 800 contracts globally. Acxiom provides IT outsourcing services to Fortune 1000 companies, who gain benefits in cost advantages and technology to gain competitive advantage. SAP, a German-based outsourcing company, offers solutions, services, and partnerships to ensure that business process outsourcing (BPO) can give the best results. The main theme behind all of these companies is the following: reduce risks, lower costs, and improve process quality.

Outsourcing Prospects

By most transactional accounts and processing evidence, outsourcing IT services appears to be a major operational decision supported by most global corporations. Companies can no longer ignore the benefits of outsourcing IT as they compete in an increasing competitive global market. McKinsey and Co suggested that in 2008 IT outsourcing was over a US\$57 billion annual export industry and that over 4 million people will be employed in India through international outsourcing. Forrester predicted that IT jobs overseas will rise from 27,000 in 2000 to over 109,000 in 2005 (Pfannenstien & Tsai, 2004). And U.S. companies have only started to explore outsourcing IT services on a global level since the early 1990s. Companies continue to learn about the process and are adapting to the problems they have experienced with outsourcing IT services as the process is still in the early stages of development.

In recent years companies have learned through experiences to shorten contract length with suppliers and include certain clauses in the contracts; such as performance measures. As companies continue to outsource IT services, they will have to continue to adapt to the problems they have experienced. For example King suggested that companies will have to require that their IT employees have the softer skills of business in the future. Traditionally, some IT employees lack the people skills or softer skills in business (Karls-son, 2003; Riis, et al., 2007). However, as King (2005) suggested, many companies will need to require their employees to have the softer skills of business. The development of the softer skills associated with business provide employees with the ability to interact with others, enhances negotiation techniques, and develops trust between partners, all of which will be critical if companies are going to grow on a global level.

FUTURE RESEARCH DIRECTIONS

There has been many research studies on the advantages and a disadvantages associated with outsourcing/offshoring as previously detailed in this chapter. According to Wee, Peng, and Wee (2010), the capital or employee size of a corporation does not affect whether or not an organization decides to outsource. Therefore, outsourcing and offshoring activities are not just for large corporations with huge cash reserves, it has filtered down to many players, regardless of size and influence. The authors the present study have found this statistic interesting and an issue to take into consideration the level of capital their respective organizations before making a decision. As more research matures in this area and more applications of financial and operational analytics come to play in discussing the ultimate benefits and costs of outsourcing/offshoring activities to a firm, industry, and/or nation, the future holds great promise for this area as an academic discipline. Undoubtedly, there are many different outsourcing techniques vary from different types of industries. For instance, automobile industries utilize different outsourcing strategies than the computer industry. The concepts of supplier collaboration and integration issues alone would be great areas to research. As previous discussed, there is a risk for supplier partners not to be generally satisfied with their related organizations as they feel that management may put an over emphasis on profit. Keeping their supplier satisfaction is a major key in creating a good outsourcing practice and beyond the scope of the present study.

CONCLUSION

Research in the present study has demonstrated that outsourcing in its various forms can be successfully applied to service jobs as well as manufacturing. The traditional thought is that

manufacturing has moved overseas to save in labor costs, but the more traditional service firms could not make that transition. However, the recent trend in IT-related offshoring/outsourcing has laid that claim to rest. In fact, it is possible and that any industry in the future can be and probably will be outsourced. IT outsourcing has recently become a lightning rod for controversy, but our research indicates that outsourcing has become an integral part of the domestic economy. Not only domestically, global firms now deeply depend upon the cost saving attributes of outsourcing/offshoring. Consumers have come to expect the cheaper products and services that have been made available through outsourcing. Therefore, as many cost sensitive consumers globally would gladly state, we welcome the further development of outsourcing strategies.

ACKNOWLEDGMENT

The authors wish to acknowledge the valuable contributions of the reviewers for their input into the final chapter.

REFERENCES

- Bahli, B., & Rivard, S. (2003). The information technology outsourcing risk: A transaction cost and Agency Theory based perspective. *Journal of Information Technology*, 18(3), 211–221. doi:10.1080/0268396032000130214
- Beasley, M., Bradford, M., & Pagach, D. (2004). Outsourcing? At your own risk. *Strategic Finance*, 86(1), 22–29.
- Beldona, S., & Tsatsoulis, C. (2010). Identifying buyers with similar seller rating models and using their opinions to choose sellers in electronic markets. *International Journal of Information and Decision Sciences*, 2(1), 1–16. doi:10.1504/IJIDS.2010.029901



- Berniker, M. (2002). *Cisco, SBC ink deal for gear and outsourcing*. Retrieved February 13, 2015 from <http://www.internetnews.com/infra/article.php/1558401/Cisco+SBC+Ink+Deal+for+Gear+and+Outsourcing.htm>
- Bhamu, J., Khandelwal, A., & Sangwan, K. S. (2013). Lean manufacturing implementation in an automated production line: A case study. *International Journal of Services and Operations Management*, 15(4), 411–429. doi:10.1504/IJSOM.2013.054883
- Browning, T. R., & Heath, R. D. (2009). Reconceptualizing the effects of lean on production costs with evidence from the F-22 program. *Journal of Operations Management*, 27(1), 23–35. doi:10.1016/j.jom.2008.03.009
- Cavaleri, S. A. (2008). Are learning organizations pragmatic? *The Learning Organization*, 15(6), 474–481. doi:10.1108/09696470810907383
- Clott, C. B. (2004). Perspectives on global outsourcing and the changing nature of work. *Business and Society Review*, 109(2), 153–170. doi:10.1111/j.0045-3609.2004.00189.x
- Coates, J. (2005). Technology Trends. *Research Technology Management*, 48(5), 7–9.
- Daim, T., Basoglu, N., & Tanoglu, I. (2010). A critical assessment of information technology adoption: Technical, organisational and personal perspectives. *International Journal of Business Information Systems*, 6(3), 315–335. doi:10.1504/IJBIS.2010.035048
- Daramola, J. O., Oladipupo, O. O., & Musa, A. G. (2010). A fuzzy expert system (FES) tool for online personnel recruitments. *International Journal of Business Information Systems*, 6(4), 444–462. doi:10.1504/IJBIS.2010.035741
- Dominic, P. D. D., Goh, K. N., Wong, D., & Chen, Y. Y. (2010). The importance of service quality for competitive advantage – with special reference to industrial product. *International Journal of Business Information Systems*, 6(3), 378–397. doi:10.1504/IJBIS.2010.035051
- Erber, G., & Sayed-Ahmed, A. (2005). Offshore outsourcing: A global shift in the present IT Industry. *InterEconomics*, 40(2), 100–112. doi:10.1007/s10272-005-0141-8
- Gonzalez, R., Gasco, J., & Llopis, J. (2010). Outsourcing and information systems managers: An empirical study. *Industrial Management & Data Systems*, 110(3), 325–350. doi:10.1108/02635571011030015
- Hall, J. A., & Liedtka, S. L. (2005). Financial performance, CEO compensation, and large scale information technology outsourcing decisions. *Journal of Management Information Systems*, 22(1), 193–221.
- Hamidi, M., Farahmand, K., Sajjadi, S. R., & Nygard, K. E. (2012). A hybrid GRASP-tabu search metaheuristic for a four-layer location-routing problem. *International Journal of Logistics Systems and Management*, 12(3), 267–287. doi:10.1504/IJLSM.2012.047602
- Jain, V., & Kohli, S. (2009). Exploring the security of e-banking systems: Questions of theft, fraud, jurisdiction and the shifting sands of time. *International Journal of Electronic Finance*, 3(4), 339–352. doi:10.1504/IJEF.2009.028975
- Jensen, B. K. (2004). The expert opinion. *Journal of Information Technology*, 6(4), 51–54.
- Kanniainen, J., Piche, R., & Mikkonen, T. (2009). Use of distributed computing in derivative pricing. *International Journal of Electronic Finance*, 3(3), 270–283. doi:10.1504/IJEF.2009.027850



- Karjaluoto, H., Jarvenpaa, L., & Kauppi, V. (2009). Antecedents of online banking satisfaction and loyalty: Empirical evidence from Finland. *International Journal of Electronic Finance*, 3(3), 253–269. doi:10.1504/IJEF.2009.027849
- Karlsson, C. (2003). The development of industrial networks, challenges to operations management in an extraprise. *International Journal of Operations & Production Management*, 23(1), 44–61. doi:10.1108/01443570310453253
- Kim, E., & Tadisina, S. (2010). A model of customers initial trust in unknown online retailers: An empirical study. *International Journal of Business Information Systems*, 6(4), 419–443. doi:10.1504/IJBIS.2010.035740
- King, J. (2003). IT's global itinerary. *Computer-world*, 37(37), 26–27.
- King, W. R. (2005). Outsourcing and offshoring: The new paradigm? *Journal of Global Information Technology Management*, 8(2), 1–4. doi:10.1080/1097198X.2005.10856393
- Kumar, P., Shankar, R., & Yadav, S. S. (2011). Global supplier selection and order allocation using FQFD and MOLP. *International Journal of Logistics Systems and Management*, 9(1), 43–68. doi:10.1504/IJLSM.2011.040059
- Marthandan, G., & Tang, C. M. (2010). Information systems evaluation: An ongoing measure. *International Journal of Business Information Systems*, 6(3), 336–353. doi:10.1504/IJBIS.2010.035049
- Mathirajan, M., Manoj, K., & Ramachandran, V. (2011). A design of distribution network and development of efficient distribution policy. *International Journal of Logistics Systems and Supply Management*, 9(1), 108–137. doi:10.1504/IJLSM.2011.040062
- More, D., & Babu, A. S. (2012). Benchmarking supply chain flexibility using data envelopment analysis. *International Journal of Logistics Systems and Management*, 12(3), 288–317. doi:10.1504/IJLSM.2012.047603
- Parker, D. W., & Russell, K. A. (2004). Outsourcing and inter/intra supply chain dynamics: Strategic management issues. *Journal of Supply Chain Management*, 40(4), 56–69. doi:10.1111/j.1745-493X.2004.tb00178.x
- Pettersson, A. I., & Segerstedt, A. (2011). Performance measurements in supply chains within Swedish industry. *International Journal of Logistics Systems and Management*, 9(1), 69–88. doi:10.1504/IJLSM.2011.040060
- Pfannenstien, L., & Tsai, R. J. (2004). Offshore outsourcing: Current and future effects on American industry. *Information Systems Management*, 21(4), 72–80. doi:10.1201/1078/44705.21.4.20040901/84190.9
- Piachaud, B. (2005). Outsourcing technology. *Research Technology Management*, 48(3), 40–46.
- Pradhananga, R., Hanaoka, S., & Sattayaprasert, W. (2011). Optimisation model for hazardous material transport routing in Thailand. *International Journal of Logistics Systems and Management*, 9(1), 22–42. doi:10.1504/IJLSM.2011.040058
- Rasolinezhad, E. (2009). Evaluating electronic banking systems in developing nations through Analytic Hierarchy Process model: A case study. *International Journal of Electronic Finance*, 3(4), 325–338. doi:10.1504/IJEF.2009.028974
- Riis, J. O., Johansen, J., Waehrens, B. V., & Englyst, L. (2007). Strategic roles of manufacturing. *Journal of Manufacturing Technology Management*, 18(8), 933–948. doi:10.1108/17410380710828262

- Sakkthivel, A. M. (2009). Impact of demographics on online buying behaviour towards different products. *International Journal of Electronic Finance*, 3(3), 284–296. doi:10.1504/IJEF.2009.027851
- Smith, A. A., Smith, A. D., Gopalakrishna-Remani, V., & Offodile, O. F. (2012). Gender preferences concerning identity theft and technology-related issues with sports e-ticketing. *International Journal of Services and Operations Management*, 13(3), 279–309. doi:10.1504/IJSOM.2012.049706
- Smith, A. D. (2005). Identify theft as a threat to CRM and e-commerce. *Electronic Government: An International Journal*, 2(2), 219–246. doi:10.1504/EG.2005.007096
- Smith, A. D., & Lias, A. R. (2005). Identity theft and e-fraud as critical CRM concerns. *International Journal of Enterprise Information Systems*, 1(2), 17–36. doi:10.4018/jeis.2005040102
- Sullivan, W., & Ngwenyama, O. K. (2005). How are public sector organizations managing IS outsourcing risks? An analysis of outsourcing guidelines from three jurisdictions. *Journal of Computer Information Systems*, 45(3), 73–87.
- Willcocks, L., Hindle, J., Feeny, D., & Lacity, M. (2004). IT and business process outsourcing: The knowledge potential. *Information Systems Management*, 21(3), 7–15. doi:10.1201/1078/44432.21.3.20040601/82471.2
- Younies, H., Barhem, B., & Hsu, C. E. (2007). A review of the adoption of Just-In-Time method and its effect on efficiency. *Public Administration and Management*, 12(2), 25–46.
- Fink, D., & Shoeib, A. (2003). Action: The most critical phase in outsourcing information technology. *Logistics Information Management*, 16(5), 302–311. doi:10.1108/09576050310499309
- Kleim, R. (2004). Managing the risks of offshore IT development projects. *Information Systems Management*, 21(3), 22–27. doi:10.1201/1078/44432.21.3.20040601/82473.4
- Leavy, B. (2004). Outsourcing strategies: Opportunities and risks. *Strategy and Leadership*, 32(6), 20–25. doi:10.1108/10878570410568875
- Rao, M. T. (2004). Key issues for global IT sourcing: Country and individual factors. *Information Systems Management*, 21(3), 16–21. doi:10.1201/1078/44432.21.3.20040601/82472.3
- Taylor, T. (2005). In defense of outsourcing. *The Cato Journal*, 25(2), 367–377.
- Von Haartman, R. (2012). Beyond Fishers product-supply chain matrix: Illustrating the actual impact of technological maturity on supply chain design. *International Journal of Logistics Systems and Management*, 12(3), 318–333. doi:10.1504/IJLSM.2012.047604
- Wee, H.-M., Peng, S.-Y., & Wee, P. K. (2010). Modelling of outsourcing decisions in global supply chains: An empirical study on supplier management performance with different outsourcing strategies. *International Journal of Production Research*, 48(7), 2081–2094. doi:10.1080/00207540802644852
- Whitten, D. (2004). User Information satisfaction scale reduction: Application in an IT outsourcing environment. *Journal of Computer Information Systems*, 45(2), 17–26.

ADDITIONAL READING

- Bonifazi, C., Desouza, K. C., & Power, M. (2004). The ten outsourcing traps to avoid. *The Journal of Business Strategy*, 25(2), 37–42. doi:10.1108/02756660410525399

KEY TERMS AND DEFINITIONS

Business Process Outsourcing (BPO): BPO is a generic term that has determine meaning in different contexts. It may be considered a type of

offshoring/outsourcing that deals primarily with the contracting of the operational aspects, management, duties, and legal and ethical responsibilities of a specific business process to third-party service providers.

Multisourcing: Refers to a very basic type of outsourcing approach in which IT operations and technology infrastructure are typically subcontracted to multiple vendors as opposed to a single or small set of vendors or service providers.

Outsourcing: It is a broad-based term that describes the procuring of goods or services that are needed by a business under contract with an external supplier. A literal definition would involve any form of trade done outside the firm as outsourcing.

Offshore Handoffs: When the process of formally charging an offshore or outsourcer with the contractual obligations, the host organization has to clearly identify what the responsibilities and expectations of onshore and offshore communication links and interactions. Such activities ensure that clearly defined service levels for each

functional area are outlined in the detail that will allow for better monitoring and updates among all locations.

Offshoring: Offshore and its related outsourcing describes the practice of contracting to outside vendors in another country, in which the client company has no direct ownership. This low cost, expansion in number and increased capability of global vendors has many important benefits and risks associated with such practices.

Service Quality: A long-term strategy that tries to differentiate a firm's products or services by assessing of how well a delivered customer service meets or exceeds a client's expectations of that good or service.

Tax Inversion: A tax inversion is basically a strategy where a domestic firm establishes or buys another company in a country with a lower corporate tax rate. Once this relocation of global headquarters is completed, the newly founded firm can avoid paying the higher tax rate in the formal host country, even if most of its sales and profit are derived in another country.

L

Pricing Based on Real-Time Analysis of Forklift Utilization Using RFID in Warehouse Management

Numan Celebi

Sakarya University, Turkey

Kübra Savaş

Istanbul University, Turkey

Ihsan Hakan Selvi

Sakarya University, Turkey

INTRODUCTION

For the purpose of meeting customer requirements logistics management is defined as the process of effective and efficient flow and storing, bringing under control and planning of movement of all kinds of product, service and information flow from starting point of the raw material to the final point of product's consumption in the supply chain (Cooper et al., 1997). The firms that are active in logistics field which has presently become a very important sector have been operating within difficult market conditions and intense competition environment. In order to maintain their existence under these conditions logistics firms must present their products and services to the market and their customers faster and in a more economical way than their competitors. In order to decrease supply time duration and provide fast response to demand changes that takes place in task durations in logistics, sector product information need to be obtained in real time. For this reason, firms need application techniques that decrease the costs and increase the efficiency and performance in logistics sector. Radio frequency Identification (RFID) is an automatic identification technology use of which has been gradually increasing in supply chain management.

Operations in logistics management are basically formed by the two processes; product and information mobility. When products are on the move which is during transportation, distribution and handling activities process product related information need to be rapidly identified and passed onto related departments, as well. The warehouses, which are used for product stocking in logistic, are places where product handling and physical movements are carried out intensively. Besides warehouse management is an important task in logistic from the point of bringing down the costs and reaching customer satisfaction. Today system control and management of warehouses that contain many products is difficult without having automatic identification. For example barcode technology that was started to be used in inventory controlling activities after 1990s has decreased erroneous and insufficient data entry that can take place due to human mistakes and allowed significant efficiency and performance increase. At present however, the RFID has superior advantages compared to barcode technology. So this technology allows for easily tracking of business processes, decreasing of labor costs and increasing of effectiveness of logistics and supply chain (Xiao et al., 2007).

Figure 1. RFID System Components



This article aims at design and application of a tracking system with RFID in response to the problem of pricing of activities of forklifts used in handling operations in logistics warehouse management. The rest of the study has been organized as follows: In the second section basic definitions and general information for RFID have been briefly given. In the third section the literature review encompassing the studies about RFID has been presented. In the fourth section problems dealt with in the application have been identified. In the fifth section structure of the designed RFID system has been explained. In the case study, pricing of movements of forklifts via designed system has been explained by an example. In the fifth section results of the study have been given and useful findings that could be drawn from system have been identified. The study has been ended with the conclusion section.

BACKGROUND

RFID is an object identification technology that uses radio frequency without any help from a human. Although this technology was found out in 1950s its real development and application have been widely spread after 1990 (Roberts, 2006). Basically RFID system is made up of a microchip (tag) around which there is wrapped antenna and a reader. Data and energy transfers are facilitated without having a physical contact between microchip and the reader (Zhang et al.,

2006). The most distinguishing advantage of the RFID systems is that unlike barcode systems its technology does not require a contact and line sight (Delen et al., 2007). As it is seen in Figure 1 an RFID system is made up of items like RFID tag, antenna, reader, controller and software that will facilitate the communications between this hardware as well as a database that will keep the data coming out as a result of this (Maloni and DeWolf, 2006).

When the literature is analyzed it is seen that RFID related studies have been conducted in two groups as academic and practical studies (Sarac et al., 2010). These studies have concentrated on warehouse related system components like RFID tag, reader and antenna. Ukkonen et al. (2004) have proposed a passive RFID antenna for metallic objects. In order to measure tag performance used in RFID technology Flores et al. (2005) have presented a simple method which was based on experimental and laboratory measurements.

The studies in recent years however have concentrated on RFID's applications in various fields such as logistics management, library and factory automation. In their wide-ranging research that identifies present day RFID applications Zhu et al. (2012) have presented a detailed analysis related to fields where RFID was used. Wang et al. (2007) have made the design of dynamic RFID based supply chain control and management system for construction industry. They have applied their proposed model to a factory that manufactures high-tech materials for construction industry in

Taiwan. For a pilot project initiated by Taiwan government Hong et al. (2011) have developed a financially feasible food tracking system with RFID. They have tried the system they developed at a markets chain company in Taiwan through a case study. In order to track the owned medical assets Öztekin et al. (2010) have shown that it is possible to apply RFID system in hospitals and to set up an effective and efficient medical asset tracking system. By conducting a simulation study LegoRobot and LabVIEW Elshayeb et al. (2010) proved that in all phases of supply chains of small and medium sized firms in retail sector products can be identified in real time and in a more effective way by RFID technology. In their article Björk et al. (2011) have explained the new RFID technology and tracking solutions that they developed for tree and forestry production industry. Hou et al. (2011) have developed a RFID based shopping tracking algorithm that would solve the problem customers experience to learn the price and to find the product particularly at large markets. Aubert (2011) has conducted a study particularly related to control and tracking with RFID of implanted devices on people with biological functional deficiencies.

While many of the previous studies have concentrated on tracking and identification of RFID tag information, today there is an increasing demand for automatic recognition systems with RFID as a solution to the problems emerging in internal and external processes of logistics enterprises. In their article Ngai et al. (2007) have presented findings of the case study they conducted at a container warehouse to track the places of containers and stackers. In order to transfer container and stacker vehicle information from RFID reader to main computer system they developed a prototype RFID system that was integrated to mobile commerce. Ergen et al. (2007) have made a study that facilitated controlling of supply chain management of the prefabricated materials used in construction sector as well as in job-order sector with the use of RFID and they achieved tracking of stocking, transportation and

assembly of the products. Dabbene et al. (2016) used RFID for food traceability and reported their implementation results in a particular food supply chain such as meat, cheese, fish, vegetables and plants. Chongwatpol and Sharda, (2013) assessed the RFID deployment to a manufacturing service provider company for tracing components on its production line. After the RFID deployment, the cycle time, machine utilizations, and penalty costs are significantly improved when it is compared to traditional approach. Feng et. al (2013) developed a cattle/beef traceability system that integrated RFID technology with PDA and barcode printer to provide accurate traceability information. They evaluated and optimized their system in a sample supply chain. Velandia et. al (2016) used RFID for automatic identification system in a crankshaft manufacturing domain. In their study they investigated the feasibility of implementing RFID system for manufacturing and assembly of crankshafts. As a result when the previous studies in the literature are analyzed, specifically there is no prior research that attempted to understand the adoption of RFID technologies in warehouse management for pricing of forklift operations. This study attempts to fill this gap.

RFID AND WAREHOUSE MANAGEMENT

Warehousing operations that require effective product mobility have an important place in logistics processes and they are the most costly processes in warehouse management. Besides, with respect to product features and structure of the sector requirements of the firms for warehousing operations change and it affects warehousing costs as well. Therefore for a successful warehouse management, when stocking the goods and transferring them from the warehouse to the required places, saving time, energy and cost is necessary. Nowadays strategic role of the warehouse has been changing, too. Previously, warehouses were used only to keep the goods but now product variety

and number of competitors are high in market, so warehouses are being operated as consolidation points that add strategic values to logistics and supply chain. Freese (2000) has stated this changing role of warehouse with the statement “Operational costs have been increasing faster than outputs of the process”. A warehouse management system provides the information necessary to manage and control the flow in the warehouse from accepting the goods to delivery. Donath et al. (2002) have pointed out that facilitating the traceability of products and instruments by using automatic definition systems like RFID particularly in warehouse management as a very important subject matter that should be taken into account.

The activities like loading, unloading and transfer that take place in warehouse management processes are called handling activities. These activities are the activities that do not add value to the final product but are the essential ones for the continuation of warehousing operations. In order to carry out these activities vehicles such as forklift (stacking), order picker and load carrier are needed. Heragu (2008) and Raman et al. (2009) have stated that forklifts are the most important instruments that allow fast and low cost handling of materials in warehouse management. Effective and efficient use of forklifts in handling activities affects warehouse management performance significantly. By using RFID for forklift tracking system, the cost of business processes can be reduced. In other words, by tracking forklift movements, real time, accurate and up to date information that is needed by the managers to determine warehouse costs can be obtained. Logistics firms can decrease their operational costs and increase service quality by this way.

PROBLEM STATEMENT

For this study, a real problem is obtained from a logistic firm. It is the leader of Turkish logistic sector. In addition to domestic and international transportation services, it has also been providing

warehousing services to its customers. The firm has been performing these services as follows: It has been renting out certain areas of the warehouses it owns to the firms that want to get warehousing service. The firms renting the warehouse space have been buying all types of operational level logistics services that are needed during warehousing from the logistic firm as well.

In practice, forklifts are rented on daily basis, not hourly basis because of economic reasons. So the logistic firm wants to measure and invoice the handlings service it provides to tenants. To put it another way, it aims to determine whether logistics services performed for two different warehouse spaces with the same size rented out to two different firms at the same unit price are the same or not from the cost perspective. In addition to unit space price the firm was determining the price of logistics services it would render to customers intuitively without making any measurements only using daily price of forklifts (for 24 hours). This situation resulted the fact that two firms renting the same size warehouse to pay the same amount of warehousing price provided that one of the firms is getting more logistics service than the other. As a result, the warehouse managers have decided to change this unfair pricing policy.

After the meetings held with warehouse managers' authors of this article have stated that this problem could be solved by a vehicle tracking system with RFID. For this purpose, firstly the pricing policy that the logistics firm would provide to its customers in warehousing field has been changed in a way that it takes the two criteria (fixed and variable costs) into account. As it is known from economics, if we assume that price quantity relation for a good is linear, the total cost is stated as the sum of fixed cost and variable cost as shown below:

$$y = a + bx \quad (1)$$

Where a shows fixed cost which means monthly or annual rent of the warehouse, x indicates the production amount which means forklift total operation time, b shows the cost per



Table 1. Equipment specification

Equipment	Brand	Specification	Quantity
Fixed Reader-UHF (Motorola FX7400)	Motorola	FX7400 UHF	2
Antenna	LAIRD	S8658 PL UHF (Dairesel polarizasyonlu)	4
Active RFID tag	Motorola	Confidex Survivor UHF	4
Computer for web and database server	Dell	With user interface	1

unit produced, in other words unit (hourly/daily) working cost of the forklift. Firstly, warehouse rent which constitutes the fixed cost is calculated as below:

Fixed cost = rented warehouse space x rent price per square meter x rental duration (2)

In the second step forklift movements that take place during products are stored in the warehouse and constituent variable costs (bx) need to be priced. The type of the forklift used in handling operation changes with respect to characteristics of raw material, intermediate good or product to be carried. Besides hourly working prices of forklifts differ with respect to forklift type and forklift carrying capacity. Accordingly, with respect to forklift's time spent for logistics services and used forklift's hourly price the variable cost is calculated as follows:

Variable Cost= Forklift operation time (hour) x forklift's hourly working price (3)

As a result, by real time tracking the forklifts with RFID system the following formula has been developed to determine price of operations.

$$TC = a + \sum_{i=1}^n F_j \left(\sum_j^m t_{ij}^{exit_time} - t_{ij}^{entrance_time} \right) \quad (4)$$

In this formula TC denotes total costs

$a = s \cdot r \cdot p$; (' s ' rented warehouse space, ' r ' warehouse's rent per square meter, ' p ' rental period) show the fixed costs;

F_j forklift's hourly working price
 $j = 1, 2, \dots, m$ denotes forklift number used in carrying operation.
 $t_{ij}^{entrance_time}$ j . forklift's entrance time to i . warehouse

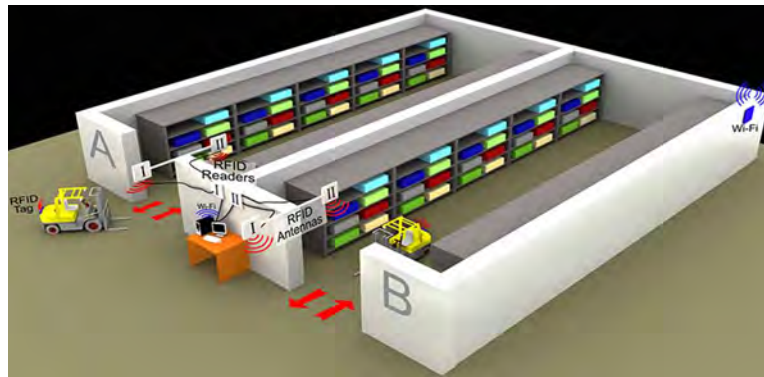
$t_{ij}^{exit_time}$ j . forklift's exit time from i . warehouse

$\sum_{i=1}^n F_j \left(\sum_j^m t_{ij}^{exit_time} - t_{ij}^{entrance_time} \right)$ formula states variable costs, $i = 1, 2, \dots, n$ denotes warehouse number

INTRODUCTION TO DEVELOPED FORKLIFT TRACKING SYSTEM

Application of forklift tracking system that was developed as a solution for the problem defined in the previous section has been made at the logistics firm's facility in Istanbul. For the experimental design the sections where products belonging to small kitchen appliances, house cleaning, personal and baby care products company (section A) and professional cleaning and sanitation systems support firm (section B) were selected. Forklift tracking system is made up of RFID UHF tags compatible with EPC code, reader, input/output antennas, a computer with developed forklift tracking software and a database system. Information about the hardware used in system design is presented in Table 1 and schematic representation of the used hardware in the warehouse is presented in Figure 2.

Figure 2. Schematic visualization of warehouses where application has been made



The active RFID tag used in tracking system makes real time information transfer to RFID reader about the location where the forklift is at. RFID tag in-built to forklifts sends a signal every 5 seconds for the location it is at. This signal is processed by RFID reader and with the help of developed software, entrance to and exit from warehouse information is recorded to the database on date and hour basis according to the signal data coming from the reader forklift. This information is later used in costing of operation times of forklifts used in the warehouse. In the photos

given in Figure 3 and Figure 4 two different views of the warehouse where forklift tracking system with RFID was applied is given. As seen from these photos antenna-I and antenna-II have been erected 110 cm above from the ground. Forklifts' entrance to and exit from warehouses are being controlled by these antennas. For this purpose, antenna number I in sections A and B is addressed as entrance antenna and antenna number II is addressed as the exit antenna. Similar design used in section A is made for section B section, as well.

Figure 3. System that was set up in application



Figure 4. RFID antennas used in application



RFID reader however was placed 90 cm above from the ground between entrance-exit antennas and it was connected to the computer system in the local network (Figure 5).

Four forklifts were used in the experimental study and RFID UHF tags were erected at the back of each forklift's driver seat as seen in Figure 6 and association between forklift and tag was done as shown in Table 2.

Figure 5. RFID UHF reader used in the application



FORKLIFT TRACKING SOFTWARE

A software has been developed for forklift movement tracking. The forklift entrance-exit tracking algorithm used in the software functions as follows: Entrance or exit decision to either section A or section B of any forklift is given through sensing by both RFID antennas (Antenna I and Antenna II) of forklift tag. Any forklift's entrance decision

Figure 6. RFID tag used in application



Table 2. Association between tag and forklift

Forklift Tag ID	Forklift Serial Number	Forklift License Plate
01	14829	34 BITS 34
02	14825	34 BITS 35
03	16947	34 BITS 36
04	14826	34 BITS 37

to any warehouse section is given upon sensing of RFID tag's incoming signal firstly from 1st antenna then from 2nd antenna by RFID reader. Exit decision is just the opposite to entrance operation. The flow chart showing the functioning of the algorithm is presented in Figure 7.

In order to transfer the information coming from RFID tag in real time a wireless network was installed between the RFID reader and the computer system. The data coming from RFID readers are transmitted to a database which is used for storing the forklift movements.

CASE STUDY: PRICING BASED ON REAL-TIME ANALYSIS OF FORKLIFT UTILIZATION

In this section the system designed for price determination is explained by a numerical example. For the numerical example, records of the forklift movements taking place at Section A and Section

Figure 7. The flowchart of the forklift tracking algorithm

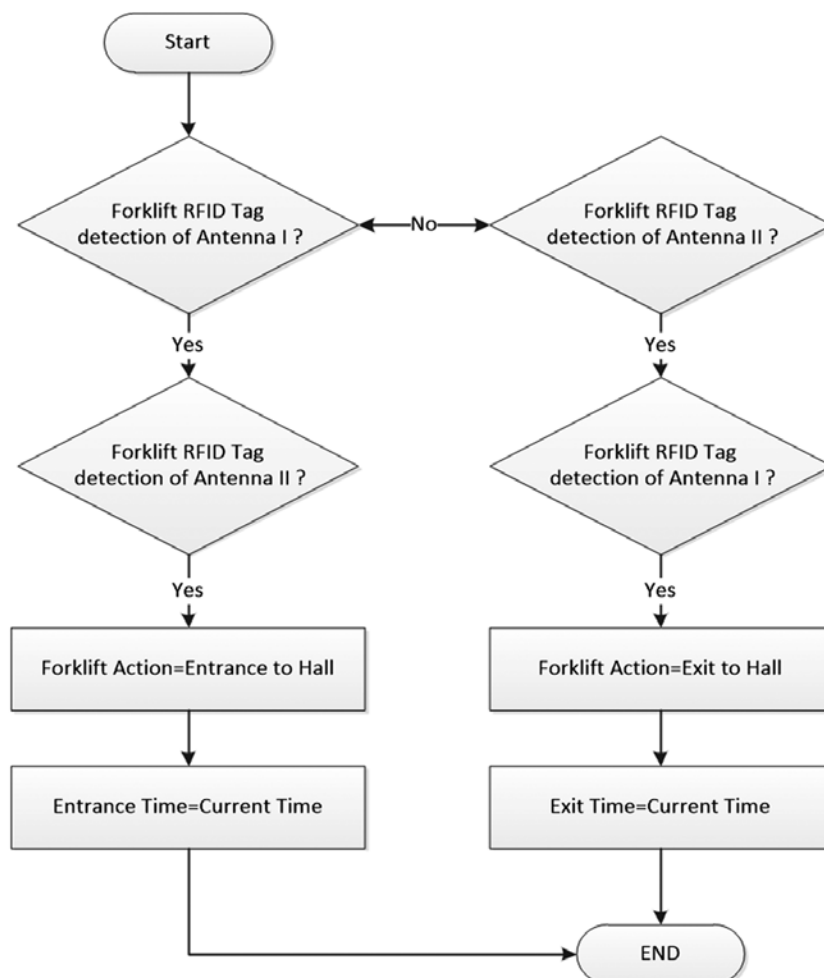


Table 3. Rented warehouse information

	Size of Rented Area (m ²)	Price of Per Square Meter (\$/m ²)	Rental Period (Month)
Tenant A	200	20	1
Tenant B	200	20	1

Table 4. Forklift types used and leasing price

Forklift Type	Daily Price of Forklifts (For 24 Hours)
I	\$85.00
II	\$100.00
III	\$65.00
IV	\$25.00

B during one month (for the sake of simplicity rent period has been taken as one month) have been taken into account. Firstly the warehouse information table that consists of information about space rented by firms (in Table 3) and daily leasing price of the forklifts (in Table 4) have been obtained as follows:

Then, the detail information about the warehouse, the forklift and the duration it was in operation was obtained from database records and required tables are created. For example, the forklift operation tables that show forklift type, and total working time used in handling operations for sections A and B (tenants warehouses) is shown in Table 5. The second column shows that the tenant firms rented the forklift types in same duration days. But, the tenant firms just used the forklifts in different operation times. The last

two columns show that total real working time of the forklifts in handling operations measured by RFID system for firm A and firm B respectively.

Finally, by using Table 4, Table 5 and by the help of equation 4 the fixed and variable costs have been calculated. *Fixed cost*: As it can be seen from Table 3 both Section A and Section B firms have rented the same size warehouse space (200 m²). In this case fixed costs are the same for both firms. *Variable costs*: It changes with respect to forklift operation duration and types of the forklift used. Therefore monthly total rent costs of two different firms renting same size warehouse space become different depending on the forklift operation costs. Before and after the RFID forklift tracking system, the handling operations costs were calculated and the results are shown in Table 6.

Table 5. Forklift types used in section A and B for operation times

Forklift Type	Total Rented Time (Days)	Section A	Section B
		Total Operating Time Measured With RFID Tracking System (Minutes)	Total Operating Time Measured With RFID Tracking System (Minutes)
I	5	3600 $\cong$ 2.5 (days)	2400 $\cong$ 1.66 (days)
II	6	4500 $\cong$ 3.125 (days)	3800 $\cong$ 2.63 (days)
III	4	2200 $\cong$ 1.527 (days)	3000 $\cong$ 2.08 (days)
IV	10	6700 $\cong$ 4.652 (days)	7500 $\cong$ 5.20 (days)

Table 6. Total cost calculation table for tenants

	Before RFID		After RFID	
	Fixed Costs	Variable Costs	Fixed Costs	Variable Costs
Tenant A	\$4000	\$1535	\$4000	\$740.5
Tenant B	\$4000	\$1535	\$4000	\$669.3

Figure 8. Forklift tracking report

Vehicle Tracking Report

34 BITS 34

<

As it can be seen from this example operation costs of the firm renting warehouse A has turned out to be higher than that of firm B. Despite the fact that fixed costs are the same for both companies total costs that should be paid differ. The logistic firm which rents its depots will be able to present to its customer more economic offers. This situation will give advantage to the logistic firm according to its competitor.

CONCLUSION

This section describes the computational experiments to evaluate the performance of the proposed system. This study showed two important results

for warehouse managers: Firstly, operations that forklifts perform in warehouses can be tracked in real time by RFID. Besides, from any point within the scope of local network it is possible to reach the network and thanks to web based software it is possible to monitor forklift movements easily. The report samples taken for this purpose are presented in Figure 8 below. In these reports it is possible to see entrance and exit information related to forklifts and it is easy to calculate forklift's operation time depending on this information. For example in the entrance-exit report of forklift with license plate "34 BITS 37" whose Tag ID is 4 there is no information. From this report it is understood that forklift number 4 was not used for any operation.

Its second importance is that it allowed accurate pricing of forklift operations taking place in different warehouses. As it can be seen from the case study forklift operation costs in warehouse A is greater than that of warehouse B. Therefore monthly total rent cost of two different firms renting the same unit space at the same rent price changes with respect to forklift operations. In other words total warehouse rent cost differs depending on variable costs.

FUTURE RESEARCH DIRECTIONS

This paper was aimed at determination of pricing for forklift per utilization time. We developed a forklift tracking method and introduced a well-designed RFID system. To the best of our knowledge, this was the first reported application of RFID to solve the problem under consideration. To determine the forklift operation costs in real time we presented a forklift tracking principle which works with RFID. According to this system the logistics firm provided with a solution to the problem of determining forklift utilization costs for the warehouses rented out to different firms. Besides, thanks to developed forklift tracking software and by the help of reports obtained from data base records warehouse manager can make cost analysis more easily. This study which was conducted in an experimental way has a flexible structure to be used in other units of the firm by making only hardware investments. Experimental results show that RFID, in addition to product tracking in retail and logistics management, can also be successfully used in solving specific problems that the firms have. Preliminary results on the forklift operation cost determination reveal that the proposed system is also effective, yet certain modifications are needed to make it more suitable for the different case. In the future, it is needed studies to take into account the subject for these purposes.

REFERENCES

- Aubert, H. (2011). RFID technology for human implant devices. *Comptes Rendus Physique*, 12(7), 675–683. doi:10.1016/j.crhy.2011.06.004
- Björk, A., Erlandsson, M., Häkli, J., Jaakkola, K., Nilsson, Å., Nummila, K., & Sirkka, A. et al. (2011). Monitoring environmental performance of the forestry supply chain using RFID. *Computers in Industry*, 62(8-9), 830–841. doi:10.1016/j.compind.2011.08.001
- Chongwatpol, J., & Sharda, R. (2013). RFID-enabled track and traceability in job-shop scheduling environment. *European Journal of Operational Research*, 227(3), 453–463. doi:10.1016/j.ejor.2013.01.009
- Cooper, M., Lambert, D. M., & Pagh, J. D. (1997). Supply chain management: More than a new name for logistics. *The International Journal of Logistics Management*, 8(1), 1–14. doi:10.1108/09574099710805556
- Dabbene, F., Gay, P., & Tortia, C. (2016). Radio-Frequency Identification Usage in Food Traceability. In M. Espiñeira & F. J. Santaclara (Eds.), *Advances in Food Traceability Techniques and Technologies Improving Quality Throughout the Food Chain, A volume in Woodhead Publishing Series in Food Science* (pp. 67–89). Elsevier.
- Delen, D., Hardgrave, B. C., & Sharda, R. (2007). RFID for Better Supply-Chain Management through Enhanced Information Visibility. *Production and Operations Management*, 16(5), 613–624. doi:10.1111/j.1937-5956.2007.tb00284.x
- Donath, B., Mazel, J., Dubin, C., & Patterson, P. (2002). *The IOMA Handbook of Logistics and Inventory Management*. New York: John Wiley & Sons, Inc.
- Elshayeb, S. A., Hasnan, K. B., & Yen, C. Y. (2010). Improving Supply Chain Traceability Using RFID Technology. *International Journal of Network and Mobile Technologies*, 1(1), 22–27.



- Ergen, E., Akinci, B., & Sacks, R. (2007). Life-Cycle Data Management of Engineered-to-Order Components Using Radio Frequency Identification. *Advanced Engineering Informatics*, 21(4), 356–366. doi:10.1016/j.aei.2006.09.004
- Feng, J., Fu, Z., Wang, Z., Xu, M., & Zhang, X. (2013). Development and evaluation on a RFID-based tracibility system for cattle/beef quality safety in China. *Food Control*, 31(2), 314–325. doi:10.1016/j.foodcont.2012.10.016
- Flores, J. L. M., Srikant, S. S., Sareen, B., & Vagga, A. (2005). Performance of RFID tags in near and far field. In *Proceedings of 2005 IEEE International Conference on Personal Wireless Communications* (353-357). IEEE. doi:10.1109/ICPWC.2005.1431365
- Freese, T. L. (2000). *Warehouse layout and design*. Principal of Freese & Associates Inc. Retrieved from http://www.freeseinc.com/images/Warehouse_Layout_and_Design
- Heragu, S. S. (2008). Material handling System. In *Logistics Engineering Handbook* (Vol. 11, pp. 1–26). New York: CRC Press.
- Hong, I. H., Dang, F. Jr, Tsai, Y. H., Liu, C. H., Lee, W. T., Wang, M. L., & Chen, P. C. (2011). An RFID application in the food supply chain: A case study of convenience stores in Taiwan. *Journal of Food Engineering*, 106, 119–126. doi:10.1201/9781420088571.ch7
- Hou, J. L., & Chen, T. G. (2011). An RFID-based Shopping Service System for retailers. *Advanced Engineering Informatics*, 25(1), 103–115. doi:10.1016/j.aei.2010.04.003
- Maloni, M., & DeWolf, F. (2006). *Understanding Radio Frequency Identification (RFID) and Its Impact on the Supply Chain*. Retrieved August, 2016, from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.89.4103&rep=rep1&type=pdf>
- Ngaia, E. W. T., Chengb, T. C. E., Auc, S., & Laib, K. H. (2007). Mobile commerce integrated with RFID technology in a container depot. *Decision Support Systems*, 43(1), 62–76. doi:10.1016/j.dss.2005.05.006
- Öztekin, A., Pajouh, F. M., Delen, D., & Swim, L. K. (2010). An RFID network design methodology for asset tracking in healthcare. *Decision Support Systems*, 49(1), 100–109. doi:10.1016/j.dss.2010.01.007
- Raman, D., Nagalingam, S. V., Gurd, B. W., & Lin, G. C. I. (2009). Quantity of Material Handling Equipment—A queuing Theory Based Approach. *Robotics and Computer-integrated Manufacturing*, 25(2), 348–357. doi:10.1016/j.rcim.2008.01.004
- Roberts, C. M. (2006). Radio frequency identification (RFID). *Computers & Security*, 25(1), 18–26. doi:10.1016/j.cose.2005.12.003
- Sarac, A., Absi, N., & Dauzere-Peres, S. (2010). A literature review on the impact of RFID technologies on supply chain management. *International Journal of Production Economics*, 128(1), 77–95. doi:10.1016/j.ijpe.2010.07.039
- Ukkonen, L., Engel, D., Sydanheimo, L., & Kivikoski, M. (2004). Planar wire-type inverted-F RFID tag antenna mountable on metallic objects. In *Proceedings of IEEE International Symposium on Antennas and Propagation Society International Symposium* (vol. 1, pp. 101-104), IEEE. doi:10.1109/APS.2004.1329563
- Velandia, D. M. S., Kaur, N., Whittow, W. G., Conway, P. P., & West, A. A. (2016). Towards industrial internet of things: Crankshaft monitoring, traceability and tracking using RFID. *Robotics and Computer-integrated Manufacturing*, 41, 66–77. doi:10.1016/j.rcim.2016.02.004

Wang, L. C., Lin, Y. C., & Lin, P. H. (2007). Dynamic mobile RFID-based supply chain control and management system in construction. *Advanced Engineering Informatics*, 21(4), 377–390. doi:10.1016/j.aei.2006.09.003

Xiao, Y., Yu, S., Wu, K., Ni, Q., Janecek, C., & Nordstad, J. (2007). Radio frequency identification: Technologies, applications and research issues. *Wireless Commun. Mobile Comput.*, 7(4), 457–472. doi:10.1002/wcm.365

Zhang, X., Hu, T., Janz, B. D., & Gillenson, M. L. (2006). Radio frequency identification: the initiator of a domino effect. In *Proceedings of 9th Southern Association for Information Systems Conference* (vol. 1, pp. 191–196).

Zhu, X., Samar, K., Mukhopadhyay, B., & Kurata, H. (2012). A review of RFID technology and its managerial applications in different industries. *Journal of Engineering and Technology Management*, 29(1), 152–167. doi:10.1016/j.jengtecman.2011.09.011

ADDITIONAL READING

Dabbene, F., Gay, P., & Tortia, C. (2016). Radio-Frequency Identification Usage in Food Traceability. In M. Espiñeira & F. J. Santaclara (Eds.), *Advances in Food Traceability Techniques and Technologies Improving Quality Throughout the Food Chain, A volume in Woodhead Publishing Series in Food Science* (pp. 67–89). Elsevier.

Dzeng, R.-J., Lin, C.-W., & Hsiao, F.-Y. (2014). Chong-Wey Lin, Fan-Yi Hsiao Application of RFID tracking to the optimization of function-space assignment in buildings. *Automation in Construction*, 40, 68–83. doi:10.1016/j.aut-con.2013.12.011

Hwang, Y.-M., Moon, J., & Yoo, S. (2015). Developing A RFID-based Food Traceability System in Korea Ginseng Industry: Focused on the Business Process Reengineering. *International Journal of Control and Automation*, 8(4), 397–406. doi:10.14257/ijca.2015.8.4.36

Ramanathan, R., Ramanathan, U., & Lok, W. L. K. (2014). Adoption of RFID technologies in UK logistics: Moderating roles of size, barcode experience and government support. *Expert Systems with Applications*, 41(1), 230–236. doi:10.1016/j.eswa.2013.07.024

KEY TERMS AND DEFINITIONS

Forklift: It is equipment that can load and upload items for shelving.

RFID: It is a technological device which can be used for item identification and traceability.

Tag: A small piece or part that is attached to an item body.

Warehouse: It is a facility where items are stored in it.

Warehouse Management: It covers many processes in it to be able to control all items, equipments and activities.

Profit Maximizing Network Modeling With Inventory and Capacity Considerations



Tan Miller

Rider University, USA

Renato de Matta

University of Iowa, USA

INTRODUCTION

Developing integrated strategic, tactical and operational manufacturing and distribution plans for the global supply chain of a large, international firm represents a formidable planning, as well as organizational undertaking. Moreover, to develop and execute plans that are not only *integrated*, but which *maximize profits* on a global basis presents a challenge of far greater magnitude. The use of advanced optimization modeling based analytics can generate keen insights for management decisions regarding sourcing, production, distribution, inventory and demand management on supply chain networks. This includes scenario planning analyses of complex strategic trade-offs such as the optimal balance between inventory levels and reserve manufacturing capacity on a network. The use of these techniques can bring clarity to the complex decisions that make integrated manufacturing and distribution planning both difficult and important (Shapiro 2010).

For purposes of this chapter, we will define “optimization modeling based analytics” as the utilization of mathematical optimization models to provide decision support for supply chain network decisions and management (i.e., models employing linear, mixed integer and nonlinear programming and related heuristic algorithms). Thus, optimization based modeling techniques represent an important component of the overall set of analytic decision support tools that can help facilitate efficient and effective supply chain network planning and management.

In summary, the objectives of this chapter include the following:

1. To review the role of optimization modeling based analytics in supporting a firm’s supply chain planning and management activities,
2. To discuss how mathematical optimization models with profit maximizing objective functions fit into a hierarchical framework for a firm’s supply chain network planning and scheduling processes,
3. To illustrate how optimization models can support key strategic network design decisions such as the appropriate balance between inventory and manufacturing capacity investment,
4. To review why optimization modeling based analytics will continue to play an increasingly important role in supply chain network decision support and management.

BACKGROUND

Supply chain management mathematical optimization models are the optimal tools for analyzing complex supply chain management problems (Shapiro 2010). In this chapter, we will focus on “deterministic” mathematical optimization models where a model solution is driven by an exogenously given (i.e., pre-determined) forecast. We note that in industry practice, the vast majority of optimization models employed are deterministic.

Practitioners typically address the potential limitations of using a single, fixed forecast by running their optimization planning models under multiple forecast scenarios, where often probabilities are assigned to each scenario. This approach alleviates the potential limitations of developing a planning solution based upon just one, deterministic forecast (Shapiro, 2010). Examples of potential uses of advanced optimization based analytics are as follows: in planning, the analysis of data to predict market trends of products and production capacity requirements; in sourcing, the use of an agent-based procurement system; and in delivering products, the applications of business analytics in logistics management to bring the products to the markets more efficiently (Trkman et al. 2010).

The methodologies and technology to support integrated profit maximization planning are well established, and the required resources are not exorbitant. In fact, the use of mathematical optimization to support logistics and supply chain management practice can be found in such industries as the oil and chemical industry as early as the 1960s (Miller, 2002). Yet despite early pioneering work in certain industries, this remains an area of terrific opportunity for many firms because surprisingly few employ these techniques in their current strategic and tactical planning processes.

There are several reasons for the underutilization of optimization techniques in supply chain planning. Briefly, factors that have contributed to this relatively slow uptake include:

1. Until the last 15 to 20 years, employing these techniques required that a firm have employees with advanced operations research skills (- note this is no longer a requirement because of commercial software advances)
2. Many senior managers remain relatively unfamiliar and/or uncomfortable with the true benefits of this mathematical technique
3. Many senior managers believe they can make effective network planning decisions without investing resources in optimization tools (- this is related to factor 2 above).

MAIN FOCUS OF THE ARTICLE

A well-constructed and maintained decision support system (DSS) supported by advanced optimization analytics represents a critical tool to facilitate coordinated supply planning and execution. A supply chain planning DSS typically has numerous components (see Liberatore and Nydick, 1998 and 2003, and Liberatore and Miller, 2012) and will vary significantly by firm. For global manufacturing and distribution planning, the inclusion of mathematical optimization models within a DSS structure can represent the difference between a DSS that provides information and a DSS that truly facilitates integrated planning and execution.

Specifically, advanced analytics for global manufacturing and distribution that utilize optimization models which consider all of a firm's manufacturing facilities, distribution locations, transportation routes and modes, and customer demands can facilitate the development, implementation, and maintenance of globally coordinated plans and schedules. A cost minimization methodology represents the most common optimization modeling approach for integrated manufacturing and distribution planning. Particularly in single country domestic applications or in environments where tax rates in local regions do not represent major decision factors, a cost minimization methodology can help formulate an effective integrated plan. However, in planning scenarios where there exist differing local country tax structures, tax rates and intra-company transfer pricing options, cost minimization methodologies generally cannot identify profit-maximizing global production and distribution plans. Instead, profit maximization models that explicitly evaluate decisions such as where to incur tax liabilities and how to set intra-company prices are required to develop the optimal, integrated global manufacturing and distribution plan. An example may help to clarify this point.

Consider a firm that has a global three echelon supply chain consisting of:

1. An intermediate (i.e., work-in-process, WIP) plant located in country A,
2. A finishing plant located in country B, and
3. Finished goods warehouses located in 20 markets (countries) where it sells its finished goods produced by country B.

This global firm will have to establish the price at which country A will sell its WIP product to country B, and also the price at which country B sells its finished goods products to the 20 markets that sell its products to the external customers in their respective local countries. The cumulative net profits after taxes and exchange rate impacts of all these local country profits yields the overall global net profit of the firm. A global profit maximization methodology facilitates formulating an optimization model which will determine for the firm what prices should be charged by countries A and B to their internal customers (i.e., to country B and to the 20 markets, respectively) to maximize overall global firm profits. In developing its solution, a profit maximization methodology can directly evaluate the tax impact in each country, as well as the impact of global exchange rates. A cost minimization methodology, however, cannot identify as precise nor as optimal a solution because this methodology does not allow the model to determine intra-firm (i.e., inter-country) prices endogenously within the model. The profit maximization methodology, however, does allow the internal prices charged by the firm's affiliates (i.e., different countries) to each other to be determined endogenously. Therefore, taxes on local country profits, as well as exchange rate impacts, are considered in determining a global profit maximizing solution. This contrasts with solutions identified as optimal by a cost minimization methodology which may not be optimal when the true tax implications of a plan are considered. See de Matta and Miller (2015) for an extensive discussion of the benefits of the strategic profit maximization modeling approach. Additionally, illustrations of tactical level decision models

which account for taxes on local country profits and exchange rate uncertainties can be found in Kouvelis (1999), Sceller-Wolf and Tayur (1999), Kazaz et al. (2005), Shunko and Garvirnneni (2007), Villegas and Ouenniche (2008), Hammami et al. (2009), Nickel et al. (2012), and Hammami and Frien (2013).

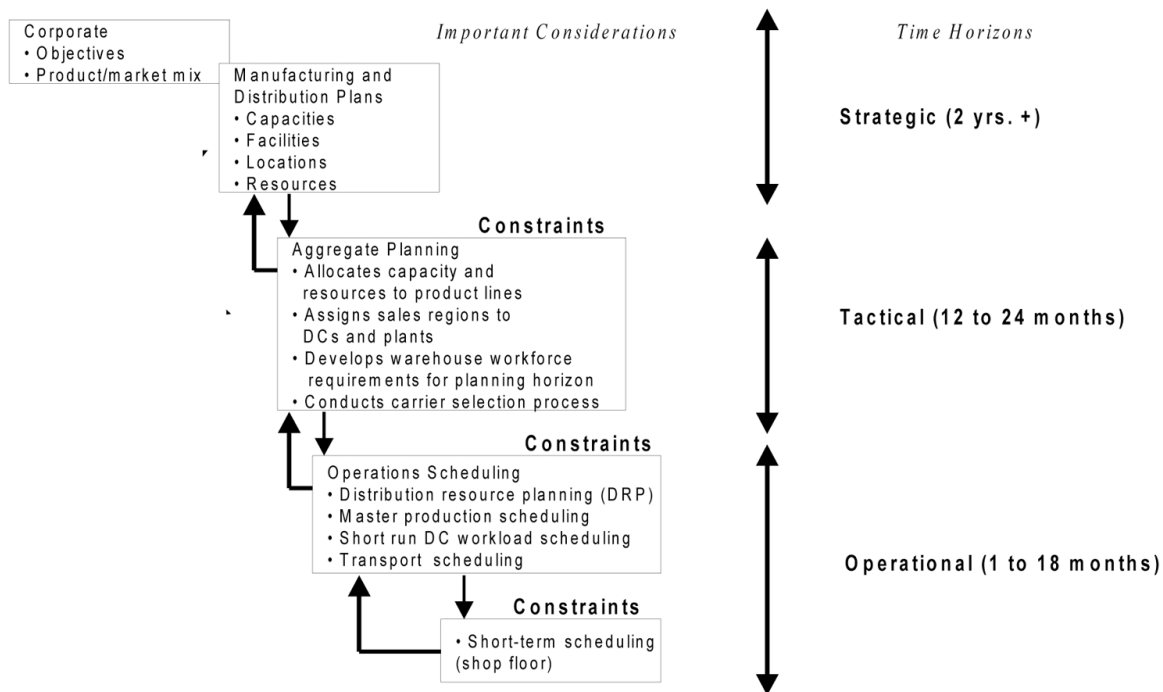
Hierarchical Supply Chain Planning

To provide additional perspective, we briefly review how our global profit maximization models fit into the overall supply chain planning and scheduling activities of a firm. To facilitate this perspective, we consider a well-known hierarchical production and distribution planning paradigm. Briefly, the hierarchical production planning process (see for example Hax and Meal 1975, Liberatore and Miller 1985, de Matta and Miller 1993, and Miller 2002) stratifies manufacturing and distribution planning and scheduling activities into three levels: strategic, tactical (which includes annual) and operational. As Figure 1 illustrates, the analytics we describe in this article support the strategic planning level, and can also easily be modified to support tactical and operational planning. Next we describe three global profit maximization models. The first model is used to provide guidance on constructing a supply chain. The second model generates an integrated, detailed global production and distribution plan at the production line, plant and distribution center level. The third model facilitates scenario planning exercises to determine the optimal balance between inventory and manufacturing capacity investment on a supply chain network.

Formation of a New Supply Chain

The process of establishing a new manufacturing and distribution network or significantly modifying an existing supply chain represents a major challenge. The global dimensions of 21st century business where sourcing, manufacturing and

Figure 1. The hierarchical production and distribution planning framework



marketing options often exist worldwide compound the complexity of this problem. Consider the perspective of a firm that wishes to establish a new manufacturing and distribution network to source, make and deliver a set of products to an international marketplace. To weigh its options, the firm first evaluates a set of potential supply chain alternatives from a very high level, strategic vantage point. In the supply chain selection process, the firm will consider all traditional supply chain operations factors such as sourcing, manufacturing and distribution costs and capacities, inventory investment requirements, and all freight, duty and tax costs. Additionally, however, the firm wants to incorporate the potential impact of exchange rates and transfer prices between different entities and countries on this network. By evaluating all pertinent operational costs and capacities, as well as exchange rate and transfer pricing considerations, the firm can establish a “profit maximizing” rather than a “cost minimizing” global manufacturing and distribution network.

SOLUTIONS AND RECOMMENDATIONS

A Strategic Manufacturing and Distribution Model

A manufacturing and distribution network (MDN) model provides guidance on constructing a supply chain from a wide assortment of sourcing, manufacturing, and country alternatives. The model allows a firm to develop a set of major production activities (echelons) that will collectively comprise a complete supply chain from original sourcing locations to final market demand points. The firm can propose as many alternative production activities as desired. Once the firm identifies all potential production activities, the mathematical model generates the optimal supply chain to satisfy the market demand forecast and maximize the firm’s profits (i.e., the model creates a comprehensive sourcing, manufacturing and delivery network). Plans generated by the

model include a description of how the selected production activities collectively link together to form this supply chain, as well as the projected output levels of, and flows between all links on the network. The model also considers the fixed cost of providing support services, for example, any access roads and loading or unloading capabilities that a firm may have to develop. Additionally, the model projects the optimal transfer prices between the firm's affiliates in individual countries as they ship the intermediate products and finished goods across the supply chain to the ultimate market destinations. Through transfer prices, the firm can allocate global profits to these affiliates in producing countries. The model's profit maximization methodology incorporates the impact of exchange rates on the global profit of the firm as calculated at its home office (i.e., country of registry). Alternatively the MDN model can also support analyses of the impact of major revisions to an existing supply chain.

The purpose of the MDN model is to provide preliminary high level planning recommendations. Thus, reflecting this high level strategic planning role, it does not explicitly model individual production lines at potential plants. Instead, the MDN model treats manufacturing capabilities as production activities that have a total output capacity, as well as fixed and variable costs associated with this capacity. These production activities can be pieced together to form a complete supply chain. Figure 1 illustrates this model's strategic role and how it fits into the overall manufacturing and distribution planning framework.

A Strategic and Tactical Planning Model

Once a firm determines the exact construction of its supply chain (using the MDN model), the firm would need to utilize more detailed planning models to support additional strategic and tactical planning activities. The strategic and tactical planning (STP) model described in Miller and De Matta (2008 & 2009) explicitly models production

lines (both their individual costs and capacities), as well as inventory levels by product family and location. The STP model generates an integrated global production and distribution plan at the production line, plant and distribution center level. Thus, a firm can employ the STP model in its more detailed strategic and tactical planning exercises. In contrast, the role of the MDN model precedes the role of the STP model. To summarize, the MDN model assists in the process of developing new or significantly revised supply chain networks, and then other models that capture the network's operations in more detail takeover to assist the more mature strategic and tactical planning processes.

How Does the Global Profit Maximization Model Work?

The STP profit maximization model evaluates the trade-offs between factors such as production costs and capacities, distribution costs, taxes, local country profits and exchange rates. These evaluations and trade-offs facilitate the development of an integrated global production and distribution plan that maximizes global profits. Figure 2 illustrates this process for a network consisting of a single or multiple echelons of plants, a single echelon of distribution centers (DCs), and external customers. In this example, a global firm owns plants in a number of countries. It also owns DCs in a number of countries, and it has external customers in each local country where it sells its products. Note that the firm may not have plants and/or DCs in every country where it generates sales (i.e., the firm may export directly from one country to external customers in another country).

Enhanced Contingency Planning: The Manufacturing Capacity – Inventory Investment Trade-off Model

We now extend our base STP model to provide a firm with an additional strategic planning capabil-



ity; namely, the ability to evaluate the potential costs and benefits of “extra” investments in manufacturing capacity and/or inventory – beyond the normal level required to meet forecast demand fulfillment goals. An important dimension of contingency planning for a firm is deciding how much, if any, reserve manufacturing capacity and/or reserve inventory to maintain on its network. For example, in its strategic planning process, assume a firm has a “most likely scenario” forecast of demand on its network for the next five years. By “most likely” let’s say a forecast to which the firm assigns a 70% to 80% probability of materializing. Assume also that this firm estimates there is a 10% to 20% probability that actual demand over the next 5 years will exceed its “most likely” forecast by 50% (i.e., this is the firm’s “high side scenario” forecast). From a contingency planning perspective, this firm must decide how prepared it wishes to be to meet the “high side” forecast, which it estimates has at best a 20% chance of occurring.

The manufacturing capacity–inventory investment (MCII) model allows the firm to determine the optimal facility location, manufacturing capacity, and inventory strategies that maximize profits while providing sufficient capacity to meet this high side forecast – should the firm decide to plan for this scenario. Importantly, the model also calculates the incremental costs to the firm of building this reserve manufacturing capacity and/or extra inventory. Thus, the MCII model allows the firm to analyze the impact on profits and costs of alternative manufacturing and inventory strategies that will provide the firm with different levels of readiness to meet alternative potential future demand scenarios.

To facilitate this contingency planning, we introduce the concept of “safety margin”. We define safety margin as the sum of:

1. Total unused manufacturing capacity (in units), and
2. Total reserve inventory (in units) above the normal level of total safety stock required to assure a defined fill rate (i.e., customer service level).

Using the MCII model, a firm’s planners can conduct contingency planning exercises where they evaluate the impact of different levels and mixes of planned safety margins on the firm’s costs, profits and network design.

“Total unused manufacturing capacity” is simply the difference between the total manufacturing capacity available on a network and the total manufacturing capacity actually used. This unused capacity represents “reserve capacity” that a firm could utilize should actual demand exceed forecasts in the future. Thus, this reserve capacity offers the firm a safety margin to protect against future high side demand scenarios. However, this reserve capacity comes with a cost, and exposes the firm to future risks. Specifically, if the firm invests in sufficient plant and equipment to meet a future high side forecast, and if that forecast does not materialize, the firm’s future profits will be depressed by the ongoing higher than necessary annual fixed costs associated with this extra unused capacity.

“Total reserve inventory” is the difference between the total inventory and the planned total safety stock inventory on a firm’s network. In the MCII model solution, this difference can range from zero to the defined upper bound on total inventory on the network. If this difference has a value of 0, this denotes that a firm will not plan to carry reserve inventory, beyond safety stock requirements, on its network. A positive differential, however, indicates that a firm will maintain reserve inventory levels, above normal safety stock requirements, in preparation for potential future demand levels that may exceed forecasts (e.g., exceed the “most likely” scenario forecast).

To summarize, in designing their manufacturing and distribution networks, firms must proactively plan for a wide range of contingencies including the possibilities that future demand may significantly exceed forecasts, and that unexpected facility disruptions and outages may occur. Two major strategies that can protect a firm against these contingencies are to build reserve manufacturing capacity and/or to build reserve inventory. Both

of these strategies require substantial incremental investments which may or may not be prudent. The MCII model provides a firm with insightful, quantitative decision support to evaluate these two strategies across a broad spectrum of real world scenarios.

Firm Organization and Accounting

In these models, the global firm operates in many countries. It has a “home” country of registry (e.g., the United States) and “local affiliates” in all other countries where it has a presence. Presence may take the form of a number of different entities (e.g., a plant; a DC; a sales office and sales force; or an autonomous operating unit with manufacturing, distribution, sales, finance and all other attributes of a stand-alone company).

Plants of local affiliates/countries (and plants of the home country) ship finished products to distribution centers of local affiliates around the world. Distribution centers receive finished goods from plants and then subsequently ship products to the firm’s external customers. This flow from the DC to the external customer represents the actual sale by the firm to its paying customers.

As Figure 2 illustrates, the price or value of a unit of product increases as it moves through the network. A plant in a local country ships (sells) a product to another local affiliate (country) at an agreed upon price. This price, at the minimum, includes the costs of production and distribution, as well as a markup on the production and distribution costs of the producing plant (country). This price, which also typically includes additional cost components such as administrative overhead, represents the transfer or intra-company price that a plant charges the local affiliate to which it delivers the product.

After it acquires the product at an intra-company transfer price, a local affiliate then sells this product to an external customer at the local selling price. The difference between an affiliate’s acquisition price and its selling price consists of the distribution, sales, and other costs of the lo-

cal affiliate, as well as a markup or profit margin that the local affiliate adds to its acquisition and delivery costs.

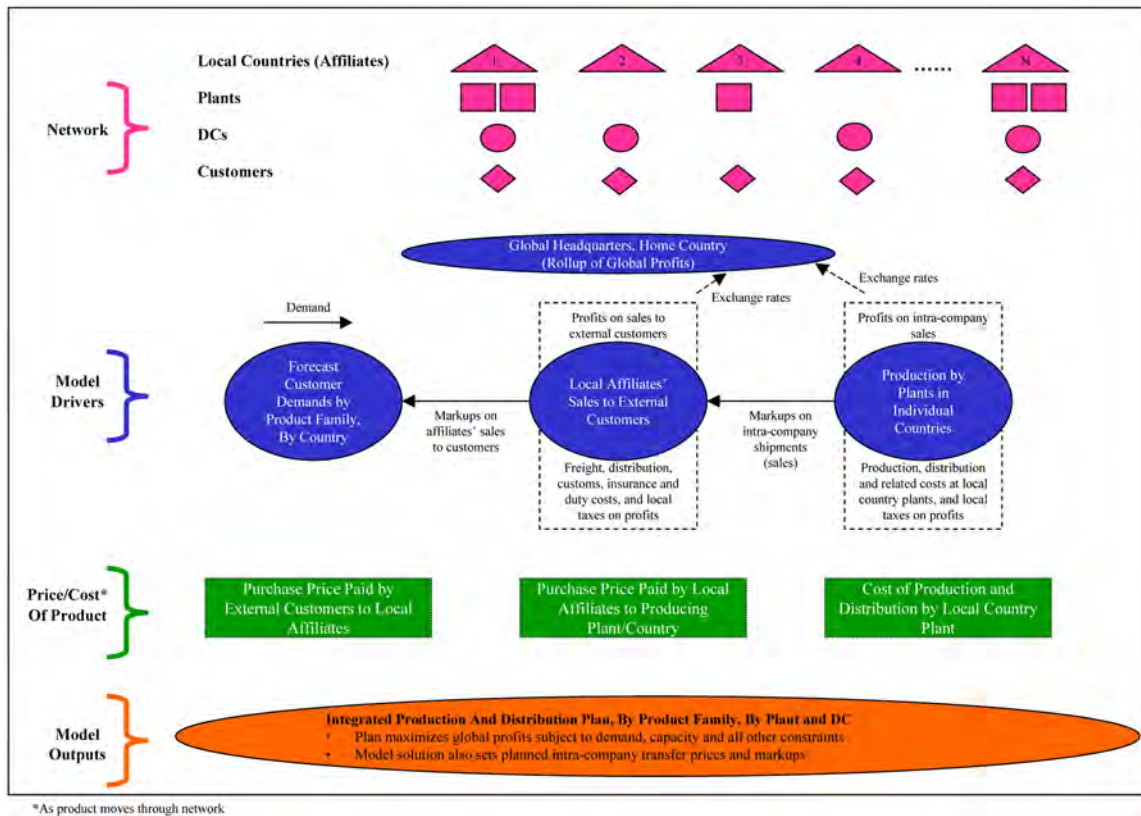
In summary, the global profit maximization model illustrated in Figure 2 evaluates and accounts for the incremental cost or value added to a product from its cost as finished goods inventory at a plant to its selling price to the firm’s external customer. In the model, the combination of all costs accounted for after production through delivery to the external customer, (i.e., all costs added to the base production cost of a unit of finished goods inventory), plus the markups on costs recommended by the model, increase the product cost of a unit of finished goods inventory at a plant to its ultimate selling price to the local external customer. The selling price to the external customer is an exogenous, pre-determined input to the model.

There are two markups on cost that create profit margins and profits. These markups occur at the point of shipment (sale) from the producing country to the receiving local affiliate, and at the point of shipment (sale) from the local affiliate to the external customer. The model recognizes and evaluates the local taxes paid on the profits recorded by the plant (country) selling to another local affiliate, and on the profits generated by the local affiliate selling to the external customer. The model utilizes the appropriate exchange rates between local affiliates and the home country of the firm to project the total global profit of the firm. Note also that this modeling methodology can explicitly account for cost factors such as freight duties, insurance on freight, and other customs charges.

Global Operations Plan Developed By The Models

The MDN and STP models generate an integrated global operations plan that maximizes the firm’s global after tax profits subject to satisfying inventory, production capacity and customer demand constraints. The models develop profit-maximiz-

Figure 2. Basic network representation of our global profit maximization models



ing solutions that take into account manufacturing costs, distribution costs including customs and duties, taxes and exchange rate impacts. The MCII model facilitates additional scenario planning to evaluate alternative integrated inventory and manufacturing investment strategies. The plans of these models maximize the after tax profits of plants plus the after tax profits of markets which are based on the product landed costs. The MDN output is a supply chain network formed from a wide assortment of sourcing, manufacturing, and country alternatives. The STP output includes a production plan for each plant by production line. The MCII output develops an additional contingency plan to address higher than forecast demand levels and/or unexpected major supply disruptions. Additionally, all three models create a global shipping and sourcing plan that indicates plant to market to customer assignments. Finally, model solutions develop recommended intra-

company transfer prices (i.e., markups) between plants and markets, and between markets and customers taking into consideration local tax rates and foreign exchange currency impacts.

We should note that a multinational firm often has internal policies regarding the range of markup percentages it will allow on transactions between its plants and local markets around the world. In many cases, local country regulations also heavily influence these corporate policies. The methodology described herein can easily account for these types of constraints. For more details, and for the actual model formulations and solution procedures, interested readers are referred to Miller and de Matta (2008) and (2009), de Matta and Miller (2015), and de Matta (2016). (See also Goetchalckx et al. 2002 for similar models.) Additionally, for descriptions of implementations of strategic and tactical network optimization models employing both profit maximization and

cost minimization methodologies, the reader is referred to Arntzen et. al. (1995), Gupta et. al. (2002), and Miller (2009).

Notes On Model Data And Other Requirements

The data requirements for the models presented in this chapter will primarily reside within a firm. Typical data elements needed would include transportation costs between all origins and destinations, all pertinent production costs and capacities, and all distribution center costs and capacities. In addition to these types of internal firm data elements, external data requirements would include all pertinent country taxes and duties, as well as exchange rate information. The reader interested in more detailed descriptions of data requirements for these types of models is referred to Miller and De Matta (2008) and Miller (2002).

The models described in this chapter can be run on standard desktop and laptop pcs. From a software perspective, to solve the more complex global profit maximization models such as formulated in Miller and De Matta (2008) and De Matta and Miller (2009), it is necessary to employ specialized mathematical programming techniques such as the generalized Benders decomposition approach (Benders, 1962). For example De Matta and Miller utilized a canned optimization package (CPLEX 8.0 Users Manual, 2002) and a customized computer algorithm in C++ to build and solve profit maximization formulations. It is important to note, however, that there also exist user friendly commercial optimization packages which can easily solve a variety of large scale profit maximization problems.

FUTURE RESEARCH DIRECTIONS

The use of profit maximization optimization models to support business analytics remains limited. The limited use of this methodology by private industry stems primarily from lack of acceptance

rather than from any algorithmic limitations. Algorithms and heuristics to solve large scale linear and non-linear profit maximization formulations abound. What is missing is the general acceptance by supply chain executives that profit maximizing network optimization models represent an indispensable planning decision support tool for *strategic and tactical* supply chain planning. Many executives continue to think of these models as exotic, resource consuming luxuries for which simpler, easier to understand methodologies (e.g., spreadsheet based analyses) represent adequate, less mysterious substitutes. In recent years, however, the use and acceptance of optimization based decision support methodologies have made significant inroads in *operational* activities such as transportation and production scheduling. In the future, we anticipate this expanded utilization of optimization based *operational* DSS tools to help fuel an increased use of optimization models in *strategic and tactical* network planning also. Additionally, Sherman (2013) has recently suggested that global optimization models will also gain popularity in the future as a supply chain risk assessment decision support tool. Thus, we expect a moderate, but steady adoption of optimization based network planning models, including profit maximization models, to occur in private industry in the coming years.

CONCLUSION

In conclusion, there exist numerous advanced analytics applications for hierarchical production planning that can utilize these global profit maximization models. We will briefly highlight several applications which address the strategic planning horizon, and also discuss how advanced business analytics can be employed for tactical planning purposes.

- The MDN model provides an integrated production and distribution framework for developing global supply chain plans



for a 2 year or longer planning horizon. Typically, planners would run the MDN model once a year to review the sourcing, capacity and delivery requirements of a firm. In this strategic application, the model can assist the firm to form or enhance a supply chain to make and deliver a new set of products to an international marketplace. This model can also support strategic decisions such as potential changes in production strategies using existing plants. For example, one can evaluate decisions regarding potential shifts in production of major products between plants.

- The STP model provides an integrated production and distribution framework for developing global supply chain plans for a 12 to 24 month planning horizon. Typically, planners would run this model once a quarter or every six months to update production plans for each plant and sourcing plans that recommend plant-to-warehouse to country (customer) assignments. In this tactical application, the model assists in determining the global plan that maximizes profit using existing supply chain infrastructure.
- Both the MDN and STP models provide excellent sensitivity analysis capabilities for strategic and tactical planning. Specifically, planners can employ these models to evaluate the impact of potential changes in exchange rates and local country taxes on medium and long-term plans. Examples include the following. At the tactical level, once an initial global plan is developed one can run the model with alternative exchange rate assumptions to determine how robust the recommended production plan is. Do production and sourcing decisions remain relatively constant over a wide range of exchange rates, or do plans vary significantly with relatively minor changes

in exchange relationships? At the strategic level, one can evaluate the potential impact of exchange rates or changes in local tax policies on long-run plant and distribution center location decisions.

- With production and distribution activities assigned to geographically dispersed facilities to form a supply chain, inventory and capacity contingency plans to reduce the risk associated with higher than forecast demands or unexpected disruptions to the supply chain can be developed using the MCII model (Handfield, 2007).
- One can also develop plans to address factors such as the impact of duty drawback opportunities and “local content” type production and distribution regulations.

These few examples provide a brief glimpse of the many potential applications that these planning models can facilitate.

In this chapter, we have stressed that global profit maximization models can address both the strategic and tactical planning levels. We believe that models for advanced business analytics such as described herein can play an important role in guiding plans for the overall operations of large-scale manufacturing and distribution networks at these longer term horizons. There are relatively few “reported” implementations of long run, strategic or tactical global profit maximization models for large-scale networks. Thus, we believe that the implementation of global profit maximization models represents a potentially significant unrealized opportunity worthy of serious consideration by many firms. This methodology remains heavily underutilized in private industry. Further, as previously noted, the investment and resources required to develop this capability within a firm are relatively modest, and the potential return on investment is extremely high.

REFERENCES

- Arntzen, B., Brown, G., Harrison, T., & Trafton, L. (1995). Global supply chain management at digital equipment corporation. *Interfaces*, 25(10), 69–93. doi:10.1287/inte.25.1.69
- Cplex 8.0 Users Manual. (2002). ILOG S.A. France.
- de Matta, R. (2016). *Contingency planning during the formation of a supply chain*. *Annals of Operations Research*. doi:10.1007/s10479-015-2085-0
- de Matta, R., & Miller, T. (1993). A Note On the growth of a production planning system: A case study in evolution. *Interfaces*, 23(4), 116–122. doi:10.1287/inte.23.4.116
- de Matta, R., & Miller, T. (2015). Formation of a strategic manufacturing and distribution network with transfer prices. *European Journal of Operational Research*, 241(2), 435–448. doi:10.1016/j.ejor.2014.08.042
- Geoffrion, A. (1972). Generalized benders decomposition. *Journal of Optimization Theory and Applications*, 10(4), 237–260. doi:10.1007/BF00934810
- Goetchalckx, V., & Dogan, K. (2002). Modeling and design of global logistics systems: A review of integrated strategic and tactical models and design algorithms. *European Journal of Operational Research*, 143(1), 1–18. doi:10.1016/S0377-2217(02)00142-X
- Gupta, V., Peters, E., Miller, T., & Blyden, K. (2002). Implementing a distribution-network decision-support system at Pfizer/Warner-Lambert. *Interfaces*, 32(4), 28–45. doi:10.1287/inte.32.4.28.54
- Hammami, R., & Frein, Y. (2013). A scenario-based stochastic model for supplier selection in global context with multiple buyers, currency fluctuation uncertainties, and price discounts. *European Journal of Operational Research*, 233(1), 159–168. doi:10.1016/j.ejor.2013.08.020
- Hammami, R., Frein, Y., & Hadj-Alouane, A. B. (2009). A strategic-tactical model for the supply chain design in the delocalization context: Mathematical formulation and a case study. *International Journal of Production Economics*, 122(1), 351–365. doi:10.1016/j.ijpe.2009.06.030
- Handfield, R. (2007). Reducing the impact of disruptions to the supply chain”, *SAScom Magazine*. April, 33-41.
- Hax, A., & Meal, H. (1975). Hierarchical integration of production planning and scheduling. *Studies in the Management Sciences*. North Holland.
- Kazaz, B., Dada, M., & Moskowit, H. (2005). *Global production planning under exchange-rate uncertainty, Technical Report*. Purdue University.
- Kouvelis, P. (1999). Global sourcing strategies under exchange rate uncertainty. Kluwer Academic Publishers.
- Liberatore, M., & Miller, T. (1985). A hierarchical production planning system. *Interfaces*, 15(4), 1–11. doi:10.1287/inte.15.4.1
- Liberatore, M., & Miller, T. (2012). *Supply chain planning: practical frameworks for superior performance*. New York, NY: Business Expert Press. doi:10.4128/9781606493175
- Liberatore, M., & Nydick, R. (1998). *Decision technology for business application*. Villanova, PA: Primus Custom Publishing.
- Liberatore, M., & Nydick, R. (2003). *Decision technology modeling. software and applications*. New York: Wiley & Sons.
- Miller, T. (2002). *Hierarchical operations and supply chain management*. London: Springer-Verlag. doi:10.1007/978-1-4471-0305-9
- Miller, T. (2009). *Notes on using optimization and DSS techniques to support supply chain and logistics operations*. In A. W. Chaovalitwongse, K. C. Furman, & P. M. Pardalos (Eds.), *Optimization and Logistics Challenges in the Enterprise* (pp. 191–210). New York: Springer.

Miller, T., & de Matta, R. (2008). A global supply chain profit maximization and transfer pricing model. *Journal of Business Logistics*, 29(1), 175–200. doi:10.1002/j.2158-1592.2008.tb00074.x

Miller, T., & de Matta, R. (2009). Maximize profit for your strategic global manufacturing and distribution planning process. *Outsourced Logistics*, 30-34.

Nickel, S., Saldanha-da-Gama, F., & Ziegler, H. P. (2012). A Multi-Stage Stochastic Supply Network Design Problem with Financial Decisions and Risk Management. *Omega*, 40(5), 511–524. doi:10.1016/j.omega.2011.09.006

Scheller-Wolf, A., & Tayur, S. (1999). Managing supply chains in emerging markets. Kluwer Academic Publishers.

Shapiro, J. (2010). Advanced analytics for sales and operations planning. *Analytics*. Retrieved from <http://www.analyticsmagazine.com>

Sherman, R. J. (2013). *Supply chain transformation*. New York: John Wiley and Sons.

Shunko, M., & Gavirneni, S. (2007). Role of transfer prices in global supply chains with random demands. *Journal of Industrial and Management Optimization*, 3(1), 99–117. doi:10.3934/jimo.2007.3.99

Trkman, P., McCormack, K., Oliveira, M., & Ladeira, M. (2010). The impact of business analytics on supply chain performance. *Decision Support Systems*, 49(3), 318–327. doi:10.1016/j.dss.2010.03.007

Villegas, F., & Ouenniche, J. (2008). A general unconstrained model for transfer pricing in multinational supply chains. *European Journal of Operational Research*, 187(3), 829–856. doi:10.1016/j.ejor.2006.04.048

ADDITIONAL READING

Arntzen, B., Brown, G., Harrison, T., & Trafton, L. (1995). Global supply chain management at digital equipment corporation. *Interfaces*, 25(10), 69–93. doi:10.1287/inte.25.1.69

Cavinato, J., & Mathis, J. (2010). Financing the global supply chain: Growing need for management action. *Thunderbird International Business Review*, 52(6), 467–474. doi:10.1002/tie.20373

de Matta, R., & Miller, T. (2015). Formation of a strategic manufacturing and distribution network with transfer prices. *European Journal of Operational Research*, 241(2), 435–448. doi:10.1016/j.ejor.2014.08.042

Goetchalckx, V., & Dogan, K. (2002). Modeling and design of global logistics systems: A review of integrated strategic and tactical models and design algorithms. *European Journal of Operational Research*, 143(1), 1–18. doi:10.1016/S0377-2217(02)00142-X

Göx, R., & Schiller, U. (2007). An economic perspective on transfer pricing. In C. Chapman, A. Hopwood, & M. Shields (Eds.), *Handbook of management accounting research* (Vol. 2, pp. 673–693). Amsterdam: Elsevier.

Lakhal, S. (2006). An operational profit sharing and transfer pricing model for network-manufacturing companies. *European Journal of Operational Research*, 175(1), 543–565. doi:10.1016/j.ejor.2005.05.024

Meixell, M., & Gargeya, V. (2005). Global supply chain design: A literature review and critique. *Transportation Research Part E, Logistics and Transportation Review*, 41(6), 531–550. doi:10.1016/j.tre.2005.06.003

Miller, T. (2002). *Hierarchical operations and supply chain planning*. London: Springer-Verlag. doi:10.1007/978-1-4471-0305-9

Miller, T., & de Matta, R. (2008). A global supply chain profit maximization and transfer pricing model. *Journal of Business Logistics*, 29(1), 175–200. doi:10.1002/j.2158-1592.2008.tb00074.x

Vidal, C., & Goetschalckx, M. (2001). A global supply chain model with transfer pricing and transportation cost allocation. *European Journal of Operational Research*, 129(1), 134–158. doi:10.1016/S0377-2217(99)00431-2

KEY TERMS AND DEFINITIONS

Decision Support Systems: Planning and scheduling tools that utilize an organization's data warehouses and other data flows to generate inputs to management decision-making.

Hierarchical Planning Time Horizons: Consists of the strategic (long run), tactical (medium term) and operational (short run) planning and scheduling horizons.

Hierarchical Supply Chain Planning: Represents an approach and a philosophy towards the organization, planning and scheduling of supply

chain activities which facilitates alignment between all short run, medium term and long run activities.

Operational Planning Horizon: Encompasses all planning and scheduling activities related to an organization's plans and schedules from the next one to eighteen months, with a primary focus on the nearer term.

Strategic Planning Horizon: Encompasses all planning activities related to an organization's plans for two years to as far into the future as the organization develops formal plans.

Supply Chain Optimization Models: Planning and scheduling models based on the mathematical techniques of linear, mixed integer and non-linear programming.

Tactical Planning Horizon: Encompasses all planning activities related to an organization's plans for the next 12 to 24 months, and includes the annual budgeting process.

Transfer Pricing: The price charged by one affiliate (i.e., subsidiary) of a multi-national firm to a second affiliate of the firm for goods and/or services produced and delivered to the second affiliate by the first affiliate.



Radio Frequency Identification Systems Within a Lean Supply Chain in a Global Environment

Alan D. Smith

Robert Morris University, USA

Terry Stringer Damron

Austin Peay State University, USA

Susan Cockrell

Austin Peay State University, USA

Amye M. Melton

Austin Peay State University, USA

INTRODUCTION

Supply Chain Management (SCM)

The ultimate goal of supply chain management (SCM) is to effectively and efficiently manage the movement of material or products through the supply chain cycle (Mateen & More, 2013; Park & Min, 2013). Such a lofty goal is accomplished only through the creation of strong partnerships with suppliers and customers, reduction of waste, and maintenance of product visibility to insure both quantity and quality meet predetermined requirements (Boerner, 2010a, 2010b). As SCM has matured as a discipline, so has the implementation of some tools to assist in achieving these goals, while still maintaining a green or sustainable presence (Berthon, Critenden, Desautels, & Pitt, 2010; Beamon, 1999). One such tool is the application of a lean production mode, which can assist in the achievement of waste reduction goals throughout the supply chain, especially when strategic partners apply the same principles. Another valuable tool is the utilization of large corporate IT-infrastructures and analytical data warehouses (Ha, Park, Lee, & Park, 2013; Mehrjerdi, 2009; Smith, 2015). As an added enhancement to a lean

supply chain, the use of real time data from these IT-infrastructures gives firms better details of current information and product movement. The use of real-time data sharing from both intra-firm and external partners is ideal for more accurate forecasting and reliable model development.

A supply chain is the sequence of activities to get materials or products to end-users. A firm's supply chain includes all steps from raw material to final product, as well as assurance of customer satisfaction (Smith, 2010, 2015; Smith & Minutolo, 2014). In the past, most firms have been concerned only with the work completed within their shop. Unfortunately, this has led to a rather disjointed supply chain characterized by higher inventory levels, poor forecasting ability, and limited visibility or products as a whole. The use of supply chain management as a means to remain competitive in a global economy is an essential component for financial success. In today's global business environment, effective SCM is crucial to enhancing business performance. It is a collaborative effort on the part of management to position and run the supply chain in the most efficient and effective ways possible. This effort covers activities from product development, sourcing, production, and logistics, as well as the

DOI: 10.4018/978-1-5225-2255-3.ch479

information systems needed to coordinate these activities.

Creating strong partnership with a select supplier base can allow for better communication and product knowledge (Browning & Heath, 2009; Cavaleri, 2008). Aligning firms with likeminded suppliers is a key activity of SCM. With the use of a select supply base, firms are better able to share pertinent data across supply chain members. Above all, this leads to better quality and reduced waste.

Understanding Lean Elements

Lean is a paradigm advocated as the foundation for a competitive SCM. Lean methodologies are well known in the corporate world, but sometimes hard to define due to the many aspects associated with successful implementation. A simple definition could be doing more with less; however, that definition underestimates the transformative power within the supply chain. As provided in Cabral and Cruz-Machado (2012), the term 'lean' implies a series of activities or solutions to eliminate waste, reduce non-value added (NVA) operations, and improve the value added (VA) process. Lean production is often connected with 'zero inventory' and just-in-time (JIT) approaches.

Lean principles were initially made popular by the Japanese manufacturing industry and most notably the Toyota Production System (TPS). The goal of lean is to simply balance the flow of materials with customer needs while maintaining optimal cost, quality, and customer service. A number of common elements define the scope of lean, helping management create a competitive firm. The implementation of lean practices does not guarantee success, but includes a primary goal of waste reduction through removal of NVA processes, reduction of excess inventory, limitation of material movement, minimization of production steps, management of scrap, and reduction of reject and rework (Carvalho, Cruz-Machado, & Tavares, 2012; Chan & Kumar, 2009; Grewal, 2008; Jain, Benyoucef, & Deshmukh, 2008;). Through worker-management relationships, a

firm can engage in a continual assessment of processes, methods, and materials for the value added to saleable products.

Purpose of Present Study

This chapter will look at the impact of using Radio Frequency Identification (RFID) systems within a lean supply chain, reviewing current research papers using a qualitative business case study approach (Baxter & Jack, 2008). Managers need to be able to justify an augmentation of lean practice and real time data with the use of RFID to improve SCMs ability to improve the goals of reducing waste and improving product visibility throughout the supply chain.

BACKGROUND

Lean Elements in Managing Supply Chains

Managing supply chain relationships by working with suppliers to meet the goal of waste reduction are essential functions for successful businesses. Firms often keep safety stock to adjust for inconsistent timing and quality of a supplier. Likewise, operations may have to manage work-in-process inventories for their own processing issues. Holding stocks of finished goods for longer periods of time before shipping to customers to avoid stock outs and maintain customer service levels adds to total costs. Unfortunately, all of these added inventories add little to no value to the company, while costing a considerable amount (Hu, Wang, Fetch, & Bidanda, 2008). Working with suppliers and customers can reduce shipment sizes while reducing inventory. Likewise, having the right inventory positioned nearer to the right customers assists in the firm's outbound shipments costs.

Lean layouts keep with the concepts of waste reduction (Drejer & Riis, 2000), minimizing the movement of workers, customers, and materials to create a more meaningful production and



distribution flow. The goal is to place people and products logically such that parts and people are positioned in the places they are needed. Visibility is central to this; line-of-sight should not be blocked. Additionally, the addition of quality checks at processing centers can assist in the reduction of rework and rejects.

Inventory and setup time reduction pull from production and distribution elements. If there are supplier or customer issues, larger inventories are likely. If there are setup time issues, layout issues are probably at hand. In this case, the need to reduce non-value added activities is part of issue identification. Properly aligning supplier and customer expectations will assist in the reduction of inventories, while updating layouts will allow for the needed flow of material to ramp up startup times.

A considerable amount of communication networking is required to properly address small-batch scheduling (Bhat, 2008; Baxter & Hirschhauser, 2004; Biswas & Sarker, 2008). The idea focuses on the ordering and production of products based on smaller, more frequent shipments. Understanding and attempting to mitigate all the SCM issues that arise is critical during this time. A main goal is minimal inventory and smaller lot sizes, at zero levels of inventories. Moving products effectively is extremely important and having the right product at the right time is even more crucial. Therefore, the need for a dedicated supplier base and a lean layout will prove effective in establishing supplier collaboration and integration policies and practices. The ultimate financial payoff of such a system is reduced cost of managing bloated inventories.

The previous components and their integration can be summed up with the goal of continuous improvement. In the case of lean layout, there is always room to improve a processing center's position or material location (Dasci, & Laporte, 2005). Reducing inventory and setup time have their own issues and any problems must be quickly resolved as suppliers and workers must continuously adjust to the faster pace of business. Once

proper communication channels are established and mutual trust and benefits are created, firms should seek to implement smaller inventories or improve the levels of current inventories. Likewise, always pushing for improved quality at lower costs requires continuous improvement strategies. Whether it is a bad stock from a supplier or product that does not meet requirements, continually monitoring and improving quality is an essential element in a lean production environment. This situation becomes even more complex in an outsourcing/offshoring environment (Wee, Peng, & Wee, 2010; Whitten, 2004).

Offshore outsourcing is the practice of contracting to outside vendors in another country, in which the client company has no direct ownership. The low-cost approach, expansion in number and increased capability of offshore outsourcing vendors in India, China, the Philippines and others, presents the opportunity to assess the hierarchical organization of functions such as software development, call centers and accounting. Since the late 1990s, offshoring as a major type of outsourcing has included ever more sophisticated finance and accounting functions, beyond that of simple data input and transactions. Offshore finance and accounting outsourcing (FAO) is becoming an increasingly attractive option for many companies, offering a means of gaining access to scarce skills, cutting costs, and achieving competitiveness (Hall & Liedtka, 2005; Nicholson & Aman, 2012).

Despite the rapid growth of the FAO industry, particularly in India, the industry has faced challenges in staff attrition. Attrition contributes to high annual turnover rates, shortages of qualified entry-level manpower, shortages of capable middle-managers, and rising wage costs. Attrition is costly to organizations, especially when those leaving their jobs are highly talented. Attrition is costly to both the provider, which must hire and train replacement employees; and also can be problematic for clients, due to the potential disruption to service when an outsourced employee leaves. Excessive attrition increases the risk of diluting service quality, affects business development ef-

forts, and impacts client relationships through the lean supply chain. Although issues of attrition are prevalent in the FAO industry, only a few studies have sought to examine how institutions influence the way firms control attrition.

The last major element of the lean production system is perhaps the most critical. Worker empowerment/commitment is required to achieve most of the previously discussed elements, as evident from the previous discussion of the FAO industry. Managers must show a strong support for a lean production system, giving their subordinates the needed resources, skills training, time, and tools required to identify process problems and bring the appropriate change to address the issues. Managers must create an environment in which the employee can bring up issues. Successful implementations of lean is difficult and many companies do fail by not being able to stay on track with the smaller faster shipments and keeping up with the continual improvements. Even after implementing a lean production system, a company can fail to engage in continuous improvement which can lead to waste and, ultimately, hinder the ability to continue in a lean production environment, with poor financial accountability (Kennedy & Widener, 2008).

Real-Time Data in SCM

SCM can take advantage of the IT-infrastructure to gather pertinent data on the firm, suppliers, products, and the market. Firms can employ many systems, including Enterprise Resource Planning (ERP) systems, large data warehouses, Business Intelligence (BI), and Customer Relation Management (CRM) systems (Smith & Offodile, 2007). Using data gathered from these systems, firms can develop a supply chain management system that offers a great advantage over the competition. Using large datasets of both historical and present data, SCM can improve forecasts and understanding of demand of current products. Likewise, using real-time point-of-sale data, SCM can better understand current customer demands. Predicting future development needs, SCM can better meet the demands of the customer.

Real-time data can be shared between the firm and strategic partners to advance SCM operations. When both supplier and customers have a better visibility of product and demand they can more effectively work together to meet the changing demand of the supply chain. Real-time data sharing can help to reduce the need for safety stock and work-in-progress inventories. Real-time data improves the tracking of materials and products both in the firms and throughout processing centers or external supply chain activities. Real-time data gives SCM a good level of visibility and improves the data used in forecasting, while providing a level of automation.

Barcode vs. Radio Frequency Identification Systems (RFID)

The current standard in inventory management is the barcode, the simplest tool in the automatic identification and data capture arsenal. Scanning a barcode provides the item's identification in the least expensive yet relatively secure manner. The scanner connects to backend databases of the IT-infrastructure and locates the item identification characteristics, linking the scanner to any pertinent information on the item it scanned. Such scanning can be done by an employee or an automated system and helps SCM track many attributes of the item such as weight, dimension, stage, and movement history. Automatic identification technology is essential for any production setting, as knowing where the material is located and minimizing non-value activities are part of the lean manufacturing structure.

Barcodes, however, have some limitations. One of the most noteworthy is that they require line-of-sight. In order to scan a barcode, the item labeled must be situated so that the scanner can read it. This requirement results in a loss of time and labor of an individual or system. A scanner can scan only one barcode at a time, meaning many labeled items must be properly aligned and scanned. Similarly, barcodes can represent

only the data that was printed. A typical barcode is read only and the amount of data that can be stored is very limited, although recent innovations in technologies, such as three-dimensional barcodes, have expanded greatly the amount of information that can be stored and accessed cheaply.

Most barcodes use standardized symbology that is certified by an international standards body (GS1 System). This system provides for the universal global acceptance of many types of barcodes designed for a variety of shipping and identification applications. (Example barcode formats in common use today include EAN/UPC, GS1 DataBar, GS1-128, ITF-14, GS1 DataMatrix, GS1 QR Code, and Composite Component.) However, a labeled package can only be what it was first labeled as. This information can be linked to back-end databases, but each firm could have its own interpretation of the item and, therefore, it is harder to share among elements of the whole supply chain. Since there is no way to reprogram a barcode if a package is mislabeled then a new label must be generated and applied to the item. While these issues may seem minor, occurrences throughout the supply add up to significant waste. Even with these limitations, an IT infrastructure has been built around the barcode, making it costly for a firm to switch to a new system; however, there are motivations to push SCM to a newer form of inventory management system, namely RFID.

As far back as June 2003, for example, Wal-Mart mandated that its top 100 suppliers place RFID tags on pallets shipped to their locations in the Dallas, Texas region (Delen, Hardgrave, & Sharda, 2007). This was motivation not only for suppliers of Wal-Mart to start looking into the use of RFID, but SCM as a whole. Shortly afterward, the Department of Defense, Target, and Alberston's followed with mandates of their own. With a surge of use by these dominant companies, RFID technology is starting to be explored as the replacement to the barcode.

RFID Technology

The basic system of RFID is composed of a tag, readers, and middleware (Kwok & Wu, 2009). The tag is placed on a pallet, container, or item and when passed by the readers are read. The tag itself consists of a chip and antenna. The two higher level types of RFID tags are passive and active tags. Passive tags have no power supply, require no maintenance, and can potentially have an indefinite operational life. These can be made small enough to be an adhesive label of its own but derives its power to operate from the electromagnetic field generated by the reader as it communicates with the reader. Active tags consists of a battery that powers the microchip and broadcasts the signal to an RFID reader. Table 1 illustrates basic data that RFID systems can typical capture about inventoried items. The reader is used to send a signal to the RFID tag activating it and allowing the transfer of data to the reader.

The use of RFID-related systems has a number of important and significant advantages over barcodes. RFID does not require line-of-sight. Tags may be read at varying distances. Further, RFID tags can be read in bulk. While this simplifies the task of reading larger areas or containers into inventory, it is important to note data accuracy issues exist when reading larger numbers of items.

Another advantage is the read-write capability of the tag, which contains information and passes it along to the distribution centers and retailer. A key to this is the use of Electronic Product Code (EPC). This is a universal identifier that gives a unique identity to the item to which the tag is affixed. The identity is made to be unique so that each object is identifiable within all of the objects. EPC creates uniqueness that is made to last over the area of the object. This, along with the ability of the RFID to capture, store and send data to reader, gives SCM more details of how and item may move among the warehouse or data center as well as throughout the SC. Table 1 demonstrates the use of RFID to monitor the

Table 1. Typical RFID-related systems data capture

Location	EPC	Date/Time	Reader
DC 123	0023800.341813.5000000024	08-04-05 23:15	Inbound
DC 123	0023800.341813.5000000024	08-09-05 7:54	Conveyor
DC 123	0023800.341813.5000000024	08-09-05 8:23	Outbound
ST 987	0023800.341813.5000000024	08-09-05 20:31	Inbound
ST 987	0023800.341813.5000000024	08-09-05 20:54	Sales floor
ST 987	0023800.341813.5000000024	08-10-05 1:10	Sales floor
ST 987	0023800.341813.5000000024	08-10-05 1:12	Backroom
ST 987	0023800.341813.5000000024	08-11-05 15:01	Sales floor
ST 987	0023800.341813.5000000024	08-11-05 15:47	Sales floor
ST 987	0023800.341813.5000000024	08-11-05 15:49	Box crusher

movement of an item as it moves from the truck to throughout the data center on its way to end of life. Insight such as this shows the real value RFID can add to SCM. These data can easily be shared with partners to help them better understand the movement of the product. The added benefits of this product visibility are a key reason why SCM is starting to adopt the idea of RFID over the conventional barcode.

MAIN FOCUS OF THE ARTICLE: BUSINESS CASE FOR RFID IN SCM TO IMPROVE LEAN OBJECTIVES

RFID Advantages

RFID presents several advantages over use of the barcode, tracking data about the item(s) to which it is attached and creating improved real-time product visibility. This is especially useful to lean firms, which—despite adherence to the principle of waste reduction and work to identify waste—do not have an easy way to visualize all the movement taking place in a given supply chain. Real-time data tracking via RFID makes this possible.

As RFID-tagged items move and tags are read, the RFID tag keeps data which can then be gathered and analyzed. With this technology, firms can find bottlenecks in the supply chain, develop real-time

solutions, and manage the work (Mehrjerdi, 2009). A virtual and tractable supply chain offers great advantage, as models can be used to track product and personnel issues and improve layout to create a better supply chain flow. The real-time data can be integrated into the firm's ERP and point of sales (POS) systems. A case study of this was completed using the textile industry:

RFID Integrated With ERP Systems

To apply RFID technology within a supply chain, a series of RFID gateways are installed in a series of stages of the textile chain. The RFID gateway acts as a check-in and check-out system, and synchronizes the in/out status to ERP in real time. RFID integrated with POS systems. For international retails, POS systems are commonly used to calculate shipping amounts and stock balance. An RFID gateway installed in a retail shop helps the shopkeeper check-in new stocks. A mobile RFID handheld device can be used for stock-taking (Kwok & Wu 2009). Having this type of 360-degree view of product movement from raw material to item at end of life creates a new model for managing inventory. Companies can greatly improve their current processes by reducing the number of human touch points of the products in their systems (Delen et al., 2007). As Chongwatpol and Sharda (2013) found, out of the

major activities that create waste, inventory level, overproduction, waiting time, processing time, cycle time, and backlog orders, RFID-embedded technology outperforms barcodes in reducing levels in all but overproduction.

A model can be implemented so that rather than having an employee physically inventory an item a reader at the loading dock can read a pallet. Dependent on the material, certain automated processes can be triggered. Likewise, when an RFID streams data and items at a retail store start to get low a shipment can be processed at the warehouse for a shipment. Streaming real-time can help firms better understand the movement of material in the supply chain and better prepare their inventory levels for market demands. As a fundamental part of lean is waste reduction, RFID certainly has a value-adding effect when it comes to finding new avenues to cut waste out of the supply chain.

Probably one of the most successful uses of an RFID is between strategic partners. The barcode did not have the ability to pass data along to the next stage in the SC. With RFID there is a new way for data to be interchanged between all members of the SC. With the standardization of EPC all companies can use the data as items are read. This can provide all members of the supply chain with invaluable information for forecasting of market demand and modeling of the current supply chain to best determine where more improvements can be made.

CONCLUSION

Overall, SCM has many tools, especially lean techniques, to use for improving the effectiveness of the supply chain. The goals are always to cut cost and find improvement to add value to the firm. A successful implementation of lean practice is a good step to continually improving the supply chain processes and workflow. The use of real-time data analysis can provide insight into the workings of the firm and market especially

when this data is shared amongst strategic partners. Emerging technology of RFID system for the transfer of data in real-time from within the firm and at large in the supply chain give SCM new ways of exploring the workings of their supply chain while simultaneously assisting their strategic partners so that all may better forecast and meet the demand of the market.

FUTURE RESEARCH DIRECTIONS

There is little doubt that RFID technology provides many benefits to companies (e.g., improved inventory control, less cost and shrinkage, greater security and authentication), but initial investments in a comprehensive system is still costly. Management must always be aware of what level does the cost of the system outweigh the benefits. Based on previous studies already cited, a RFID system used to identify a pallet of products for a company can provide a savings for that level of goods, but on an individual item the cost of the RFID system could outweigh the benefits. Therefore, additional research needs to focus if the members of the supply chain are able and willing to share the cost of RFID will this increase the benefit to the each member of the supply chain and who would benefit more. Ultimately, it may be the non-technology factors (e.g., accounting on operational processes, inventory forecasting accuracy, actual demand, case size, available shelf and warehouse spaces), that aid management in balancing the costs/benefits in any RFID embedded system and its maintenance in the strategic decision to implement such systems.

ACKNOWLEDGMENT

The authors wish to thank most heartedly for the valuable contributions by the reviewers for their input into the final paper. Peer reviewing and editing are commonly tedious and thankless tasks.

REFERENCES

- Baxter, L. F., & Hirschhauser, C. (2004). Reification and representation in the implementation of quality improvement programmes. *International Journal of Operations & Production Management*, 24(2), 207–224. doi:10.1108/01443570410514894
- Baxter, P., & Jack, S. (2008). Qualitative case study methodology: Study design and implementation for novice researchers. *Qualitative Report*, 13(4), 544–559.
- Beamon, B. M. (1999). Designing the green supply chain. *Logistics Information Management*, 12(4), 332–342. doi:10.1108/09576059910284159
- Berthon, P., Critenden, V., Desautels, P., & Pitt, L. (2010). Get the most out of green IT. *Industrial Management (Des Plaines)*, 52(5), 14–18.
- Bhat, S. (2008). The effect of ordering policies for a manufacturing cell changing to lean production. *Proceedings - Institution of Mechanical Engineers*, 222(11B11), 1551–1560. doi:10.1243/09544054JEM1216
- Biswas, P., & Sarker, B. R. (2008). Optimal batch quantity models for a lean production system with in-cycle rework and scrap. *International Journal of Production Research*, 46(23), 6585–6610. doi:10.1080/00207540802230330
- Boerner, H. (2010a). Sustainability rises to top of strategy-setting for growing number of corporate leaders. *Corporate Finance Review*, 15(1), 32–34.
- Boerner, H. (2010b). Sustainability and corporate responsibility: II. Boards and executives respond to stakeholders. *Corporate Finance Review*, 15(2), 32–37.
- Browning, T. R., & Heath, R. D. (2009). Reconceptualizing the effects of lean on production costs with evidence from the F-22 program. *Journal of Operations Management*, 27(1), 23–35. doi:10.1016/j.jom.2008.03.009
- Cabral, I. G. A., & Cruz-Machado, V. (2012). A decision-making model for Lean, Agile, Resilient and Green supply chain management. *International Journal of Production Research*, 50(17), 4830–4845. doi:10.1080/00207543.2012.657970
- Carvalho, H., Cruz-Machado, V., & Tavares, J. G. (2012). A mapping framework for assessing supply chain resilience. *International Journal of Logistics Systems and Management*, 12(3), 354–373. doi:10.1504/IJLSM.2012.047606
- Cavaleri, S. A. (2008). Are learning organizations pragmatic? *The Learning Organization*, 15(6), 474–481. doi:10.1108/09696470810907383
- Chan, F. T. S., & Kumar, V. (2009). Performance optimization of a legality inspired supply chain model: A CFGTSA algorithm based approach. *International Journal of Production Research*, 47(3), 777–791. doi:10.1080/00207540600844068
- Chongwatpol, J., & Sharda, R. (2013). Achieving lean objectives through RFID: A simulation-based assessment. *Decision Sciences*, 44(2), 239–266. doi:10.1111/deci.12007
- Dasci, A., & Laporte, G. (2005). A continuous model for multistore competitive location. *Operations Research*, 53(2), 263–280. doi:10.1287/opre.1040.0175
- Delen, D., Hardgrave, B. C., & Sharda, R. (2007). RFID for better supply-chain management through enhanced information visibility. *Production and Operations Management*, 16(5), 613–624. doi:10.1111/j.1937-5956.2007.tb00284.x
- Drejer, A., & Riis, J.O. (2000). New dimensions of competence development in industrial enterprises. *International Journal of Manufacturing Technology and Management*, 2(1/7), 660–882.
- Grewal, C. (2008). An initiative to implement lean manufacturing using value stream mapping in a small company. *International Journal of Manufacturing Technology and Management*, 15(3/4), 404–421. doi:10.1504/IJMTM.2008.020176

- Ha, O., Park, M., Lee, K., & Park, D. (2013). RFID application in the food-beverage industry: Identifying decision making factors and evaluating SCM efficiency. *KSCE Journal of Civil Engineering*, 17(7), 1773–1781. doi:10.1007/s12205-013-0297-x
- Hall, J. A., & Liedtka, S. L. (2005). Financial performance, CEO compensation, and large scale information technology outsourcing decisions. *Journal of Management Information Systems*, 22(1), 193–221.
- Hu, G., Wang, L., Fetch, S., & Bidanda, B. (2008). A multi-objective model for project portfolio selection to implement lean and Six Sigma concepts. *International Journal of Production Research*, 46(23), 6611–6648. doi:10.1080/00207540802230363
- Jain, V., Benyoucef, L., & Deshmukh, S. G. (2008). Whats the buzz about moving from lean to agile integrated supply chains? A fuzzy intelligent agent-based approach. *International Journal of Production Research*, 46(23), 6649–6678. doi:10.1080/00207540802230462
- Kennedy, F. A., & Widener, S. K. (2008). A control framework: Insights from evidence on lean accounting. *Management Accounting Research*, 19(4), 301–319. doi:10.1016/j.mar.2008.01.001
- Kwok, S., & Wu, K. W. (2009). RFID-based intra-supply chain in textile industry. *Industrial Management & Data Systems*, 109(9), 1166–1178. doi:10.1108/02635570911002252
- Mateen, A., & More, D. (2013). Applying TOC thinking process tools in managing challenges of supply chain finance: A case study. *International Journal of Services and Operations Management*, 15(4), 389–410. doi:10.1504/IJSOM.2013.054882
- Mehrjerdi, Y. (2009). RFID-enabled supply chain systems with computer simulation. *Assembly Automation*, 29(2), 174–183. doi:10.1108/01445150910945624
- Nicholson, B., & Aman, A. (2012). Managing attrition in offshore finance and accounting outsourcing: Exploring the interplay of competing institutional logics. *Strategic Outsourcing: An International Journal*, 5(3), 232–247. doi:10.1108/17538291211291765
- Park, B.-N., & Min, H. (2013). Global supply chain barriers of foreign subsidiaries: The case of Korean expatriate manufacturers in China. *International Journal of Services and Operations Management*, 15(1), 67–78. doi:10.1504/IJSOM.2013.050562
- Smith, A. D. (2010). Balancing internal supply chain logistics: A comparative analysis of manufacturing and service firm operations. *International Journal of Logistics Systems and Supply Management*, 3(2), 145–166.
- Smith, A. D. (2015). Process Improvements in Supply Chain Operations: Multi-Firm Case Studies. In S. Puranam (Ed.), *Handbook of Research on Organizational Transformations through Big Data Analytics. A volume in the Advances in Business Information Systems and Analytics (ABISA) Book Series* (pp. 240–261). doi:10.4018/978-1-4666-7272-7.ch016
- Smith, A. D., & Minutolo, M. C. (2014). Green supply chain acceptability and internal stakeholder concerns. *International Journal of Logistics Systems and Management*, 19(4), 464–490. doi:10.1504/IJLSM.2014.065666
- Smith, A. D., & Offodile, O. F. (2007). Exploring forecasting and project management characteristics of supply chain management. *International Journal of Logistics and Supply Management*, 3(2), 174–214.
- Wee, H.-M., Peng, S.-Y., & Wee, P. K. (2010). Modelling of outsourcing decisions in global supply chains: An empirical study on supplier management performance with different outsourcing strategies. *International Journal of Production Research*, 48(7), 2081–2094. doi:10.1080/00207540802644852

Whitten, D. (2004). User Information satisfaction scale reduction: Application in an IT outsourcing environment. *Journal of Computer Information Systems*, 45(2), 17–26.

ADDITIONAL READING

Aldaihani, M. M., & Darwish, M. A. (2013). Optimal production and inventory decisions for supply chains with one producer and multiple newsvendors. *International Journal of Services and Operations Management*, 15(4), 430–448. doi:10.1504/IJSOM.2013.054884

Azadeh, A., Gholizadeh, H., & Jeihoonian, M. (2013). A multi-objective optimisation model for university course timetabling problem using a mixed integer dynamic non-linear programming. *International Journal of Services and Operations Management*, 15(4), 467–481. doi:10.1504/IJSOM.2013.054886

Bhamu, J., Khandelwal, A., & Sangwan, K. S. (2013). Lean manufacturing implementation in an automated production line: A case study. *International Journal of Services and Operations Management*, 15(4), 411–429. doi:10.1504/IJSOM.2013.054883

Chen, C. C., Wu, J., Su, Y. S., & Yang, S. C. (2008). Key drivers for the continued use of RFID technology in the emergency room. *Management Research News*, 31(4), 273–288. doi:10.1108/01409170810851348

Dutta, A., Lee, H. L., & Whang, S. (2007). RFID and operations management: Technology, value and incentives. *Production and Operations Management*, 16(5), 646–655. doi:10.1111/j.1937-5956.2007.tb00286.x

Fisher, J. A., & Monahan, T. (2008). Tracking the social dimensions of RFID systems in hospitals. *International Journal of Medical Informatics*, 77(3), 176–183. doi:10.1016/j.ijmedinf.2007.04.010 PMID:17544841

Fumi, A., Scarabotti, L., & Schiraldi, M. M. (2013). The effect of slot-code optimisation on travel times in common unit-load warehouses. *International Journal of Services and Operations Management*, 15(4), 507–527. doi:10.1504/IJSOM.2013.054925

Helo, P., Anussornnitisarn, P., & Phusavat, K. (2008). Expectation and reality in ERP implementation: Consultant and solution provider perspective. *Industrial Management & Data Systems*, 108(8), 1045–1158. doi:10.1108/02635570810904604

Hu, G., Wang, L., Fetch, S., & Bidanda, B. (2008). A multi-objective model for project portfolio selection to implement lean and Six Sigma concepts. *International Journal of Production Research*, 46(23), 6611–6648. doi:10.1080/00207540802230363

Ifinedo, P., & Nahar, N. (2009). Interactions between contingency, organizational IT factors, and ERP success. *Industrial Management & Data Systems*, 109(1), 118–126. doi:10.1108/02635570910926627

Johansson, B., & Sudzina, F. (2008). ERP systems and open source: An initial review and some implications for SMEs. *Journal of Enterprise Information Management*, 21(6), 649–659. doi:10.1108/17410390810911230

Jones, E., Riley, M., Franca, R., & Reigle, S. (2007). Case study: The engineering economics of RFID in specialized manufacturing. *The Engineering Economist*, 52(3), 285–303. doi:10.1080/00137910701503951

Kamhawi, E. M. (2008). System characteristics, perceived benefits, individual differences and use intentions: A survey of decision support tools of ERP systems. *Information Resources Management Journal*, 21(4), 66–83. doi:10.4018/irmj.2008100104

Koong, L., & Lin, C. (2007). Evaluating the decision to adopt RFID systems using analytic hierarchy process. *Journal of American Academy of Business*, 11(1), 72–77.



Pacciarelli, D., DAriano, A., & Scotto, M. (2011). Applying RFID in warehouse operations of an Italian courier express company. *NETNOMICS: Economic Research and Electronic Networking*, 12(3), 209–222. doi:10.1007/s11066-011-9059-4

Smith, A. A., Smith, A. D., & Baker, D. J. (2011). Inventory management shrinkage and employee anti-theft approaches. *International Journal of Electronic Finance*, 5(3), 209–234. doi:10.1504/IJEF.2011.041337

Smith, A. D. (2011). Component part quality assurance concerns and standards: Comparison of world-class manufacturers. *Benchmarking: An International Journal*, 18(1), 128–148. doi:10.1108/14635771111109850

Smith, A. D., & Rupp, W. T. (2013). Data quality and knowledge/information management in service operations management: Regional supermarket case study. *International Journal of Knowledge-Based Organizations*, 3(3), 35–52. doi:10.4018/ijkbo.2013070103

Ustundag, A. (2010). Evaluating RFID investment on a supply chain using tagging cost sharing factor. *International Journal of Production Research*, 48(9), 2549–2562. doi:10.1080/00207540903564926

Visich, J. K., Li, S., Khumawala, B. M., & Reyes, P. M. (2009). Empirical evidence of RFID impacts on supply chain performance. *International Journal of Operations & Production Management*, 29(12), 1290–1315. doi:10.1108/01443570911006009

Wyld, D. C. (2006). RFID 101: The next big thing for management. *Management Research News*, 29(4), 154–173. doi:10.1108/01409170610665022

Yao, A., & Carlson, J. (1999). The impact of real-time data communication on inventory management. *International Journal of Production Economics*, 59(1), 213–219. doi:10.1016/S0925-5273(98)00234-5

Zang, C., & Fan, Y. (2007). Complex event processing in enterprise information systems based on RFID. *Enterprise Information Systems*, 1(1), 3–23. doi:10.1080/17517570601092127

KEY TERMS AND DEFINITIONS

Barcoding Technology: A long-term and very reliable type of AIDC technology, it is known for its very accurate and economical approaches to identity products and machine readable information from a variety of manufactured goods and services. Most barcodes use a type of standardized bars and spacing coding or symbology that is certified by an international standards body (GS1 System). This system provides for the universal global acceptance of many types of barcodes designed for a variety of shipping and identification applications. Example barcode formats that are in common use today include EAN/UPC, GS1 DataBar, GS1-128, ITF-14, GS1 DataMatrix, GS1 QR Code and Composite Components.

Electronic Product Code (EPC): EPC is a universal identifier that gives a unique identity to an item a RFID tag is affixed to. The identity is made to be unique so that each object is identifiable within all of the objects' field.

RFID-Embedded Technologies: RFID technologies are types of automatic data capture techniques that use a combination of active and passive senders and receivers to collect and store codified information for further uses. The implementation of such technologies should lead to improved managerial and/or supply chain performance. On the surface, there appears to be few drawbacks to implementing such technology into a production process, assuming it enhances performance and improves output of the product. The main issues surrounding the RFID applications are whether the initial costs and labor required to utilize this technology are worth it, and will result in a positive outcome of revenues.

Reflections of the 1Malaysia Supply Chain (1MSC)



Munira Halili

Universiti Sains Malaysia, Malaysia

Latifah Naina Mohamed

Universiti Sains Malaysia, Malaysia

Yudi Fernando

Universiti Malaysia Pahang, Malaysia

INTRODUCTION

Malaysia aspires to achieve a developed, high income nation status by the year 2020 (National Economic Advisory Council, 2009). Gearing towards that direction, Malaysian Prime Minister introduced a New Economic Model (NEM) intended to liberalize the economy. Among the eight Strategic Reform Initiatives (SRI) proposed in NEM is the need to have better integration between products and between production centers, so that economy of scales can be created and ensuring sustainability of growth economically, environmentally and socially. Recognizing the importance of a collaborative and integrated supply chain to seamlessly produce an output that meets a market demand and that is socially conscious, Sheng (2010) introduced a new concept called 1Malaysia Supply Chain (1MSC) in 2010. It consists of three strategic directions: Focus, Simplify and Integrate – basically all that is needed to integrate to achieve higher productivity, economies of scale and be sustainable focused.

BACKGROUND

Manufacturing industry is the biggest contributor to Malaysia's economic growth. According to a report from Malaysia External Trade Development Corporation (MATRADE), the Electrical &

Electronics (E&E) industry is one of the leading industries, contributing 24.5 per cent to the manufacturing sector in the Malaysia's Gross Domestic Product (GDP). In 2014, Malaysia's exports of E&E products was valued at RM231.23 billion, with 49.2 per cent share of manufactured goods exports and 32.9 per cent share of Malaysia's total exports.

Although the manufacturing industry has flourished, there are several challenges that the industry faces such as rising operational cost, scalability and rapid technology advancements. In addition, there is fierce competition between rival supply chain networks in the industry (Lo & Power, 2010). Currently, manufacturing industries are resorting to outsourcing the non-value added activities to contract manufacturers or suppliers. Most manufacturing companies have adopted to integrate their supply chain in order to obtain visibility into the supplier's deliverables and to detect possible issues upfront. While with supply chain integration (SCI) the operational cost can be lowered to an extent, it still does not address the gap of limited number of capable suppliers in the market to support the growing industrial needs. Thus, collaborative relationship with suppliers is needed (Stock, Boyer & Harmon, 2010).

Lack of technological expertise in a region have also becomes a bottleneck that hinders manufacturer's speed to market an innovation. These challenges raise a pressing need to develop

an effective and efficient SCI in a supply chain network (Huang, Yen & Liu, 2014). A successful SCI depends on information, coordination, and organizational linkage (Lee, 2000). However, the empirical results of SCI in the literature have been inconsistent. There is little evidence that integration provide improvements in effectiveness and efficiency (Fawcett & Magnan, 2002; Bagchi, Chun Ha & Skjoett-Larsen, 2005; Fabbe-Costes & Jahre, 2007). This is due to lack of clear definition of SCI and variation of definitions and interpretation of the concept (Fabbe-Costes & Jahre, 2007; Van der Vaart & Van Donk, 2008). Thus, scholars need to develop a model encompassing factors of SCI to benefit both practitioners and scholars (Zhang, Gunasekaran & Wang, 2015). To address this issue, this chapter took the challenge and proposes nationwide SCI by adopting the 1Malaysia concept.

The 1Malaysia concept was introduced by Malaysia's Prime Minister in year 2010 which comprises of eight values as its key motivators. Culture of Excellence, Perseverance, Humility, Acceptance, Loyalty, Meritocracy, Education and Integrity are the motivators behind the implementation of this concept. It focuses on national unity and ethnic tolerance which aims to bring the multicultural and multiethnic population as one. During its design and introductory phase, the concept was unclear to many and was considered as an alienated topic. This concept has now been progressively acknowledged and Malaysian has been utilizing the facilities that 1Malaysia offers such as the 1Malaysia clinics and the Kedai Rakyat 1Malaysia (KR1M). The driving force behind the realization of this concept is the strong foundation of its supply chain which integrates different parties as one.

The 1Malaysia concept coupled with SCI forms the proposed nationwide SCI called the 1Malaysia Supply Chain (1MSC). The outcome of 1MSC is to achieve supply chain performance and social performance, in which SCI failed to reach consistent results (Cousins & Menguc, 2006; Chen, Mattioda & Daugherty, 2007; Sezen, 2008; Flynn,

Huo & Zhao, 2010). This chapter shall explore a similar integration which has been successfully implemented in Malaysia through the KR1M initiative (component of 1MSC).

This chapter is organized by first explaining the overview of SCI concept and its challenges that would be relevant to the 1MSC concept. Then, the chapter discusses elements in the conceptual view of 1MSC and its contributions. Last, this chapter provides recommendations, future research directions and conclusion.

SUPPLY CHAIN INTEGRATION (SCI)

As 1MSC concept is derived from SCI, it is important for reader to fully understand the concept of SCI. Nowadays, the concept of companies versus companies has been replaced with supply chain versus supply chain competitiveness strategy, however only a small number of companies have implemented a fully integrated supply chain (Georgise, Thoben, & Seifert, 2014). The 1MSC concept involves collaboration and coordination within a supply chain, adopting a similar integration concept with the SCI. SCI is a topic that often being discussed in the supply chain management (SCM) literature (Power, 2005; Guan & Rehme, 2012) and has been widely studied by researchers such as Danese & Romano (2011), Zhang et al. (2015) and Yunus & Tadisina (2016). Firms require integration of all product flow procedures in order for SCM to be effective and efficient (Huang et al., 2014); hence collaboration among the stakeholders plays a vital role in ensuring smooth activities for the firms. SCI is beneficial to firms as it provides operational flexibility to respond to dynamic global needs, actions of competitors and changes in customer demand. Collaborating with supplier for instance enables firms to order materials as needed and response to customer's needs and requirements.

As defined by Flynn, Huo, & Zhao (2010), SCI focuses on strategic collaboration and process integration by looking at the extent to which a

manufacturer deliberately works together with its supply chain partners in dealing with internal or external practices and procedures. Leuschner, Rogers & Charvet (2013) on the other hand defined SCI as the scope and strength of linkages in supply chain processes across firms. SCI has its own characteristics such as having the basic objective and goal in meeting customers' needs, systematic and well defined procedure, and quick and ongoing communications between stakeholders.

According to Awad & Nassar (2010a), the integration of technology, people, business and processes is critical for survival and competitiveness in the current era, and its importance goes not only within a firm but also across extended enterprises. Hence, a good relationship with stakeholders requires firms to adopt SCI as a strategic tool and sharing of information. Information sharing refers to the degree to which critical and copyrighted information is communicated to other supply chain partner (Mohr & Spekman, 1994). Flynn et al. (2010), Gimenez, (2006) & Van der Vaart & Van Donk (2008) concluded that information sharing is part of an integration measurement in almost all surveys conducted on SCI. Thus, it is important not to neglect this driver as its importance has been proven in previous studies. According to Yu, Yan and Edwin (2001), information sharing is important to reduce uncertainties and supply chain performance can be achieved by changing decentralized supply chain to centralized method. This view is seconded by Awad & Nassar (2010a) that the most significant form of SCI was information sharing. The ability to facilitate, coordinates, and integrates the flow of information across the supply chain helps supply chain members to establish partnerships for better performance.

Although SCI brings benefit to the firms, it possesses great amount of challenges as well. Integration requires alignment between all stakeholders in order to achieve improved performance and service. In order to do so, knowledge sharing is deemed important and the sharing could be done through setting up common system. Technical

capabilities have its own challenges, especially in term of relation between SCM system and internal business strategy.

According to Awad & Nassar (2010b), certain challenges related to SCI are due to differences in firm's strategy theory and concepts, different ways of satisfying the needs and wants of customers from competitors, handling of logistics arrangement differently, difficulty in maintaining standards due to differences in elements such as the size and culture and importance of determining supplier's capability. Thus, the development of 1MSC concept aims to addresses these challenges in its implementation.

1 MALAYSIA SUPPLY CHAIN (1MSC)

1MSC is a term coined by Sheng (2010) and Zailani (2011). However, it has not been extensively researched and relatively unknown to both local and global scholars. This is because 1MSC is viewed as political agenda and lack of public awareness. Leveraging the concept of SCI, this notion goes beyond solely focusing on economic goals to also include sustainable development that encompasses fulfilling social and environmental objectives. According to Zailani (2011), the 1MSC aims at achieving targeted economic growth while at the same time contributing towards equally disseminated value-added productions and services among all players in the supply chain.

The 1MSC's strategy is to "Simplify, Focus and Integrate" that is, the supply chain network collaborated (integrate) and moved towards meeting the needs of all its stakeholders and parallelly providing platform to support business to improve value chains (simplify), finding new market niches as well as improving sustainable practices (focus) (Zailani, 2011). To do so, the Small Medium Enterprises (SMEs) have to be integrated into the larger global supply chain (with larger firms) and not dependent on subsidies in order to succeed in the long-run. Malaysia focusing on productivity-led and private led-growth, growth through clusters



and encourage technologically capable firms and industries and simplification of government structure, regulatory framework and a much-reduced role of government in business are required to ensure the successful implementation of overall national economic strategy (Sheng, 2010).

According to Sheng (2010), 1MSC would assist in achieving the NEM's goal by:

1. Integrating silos and production chains to be able to compete globally.
2. Using international standard as a reference to connect unrelated and overlapping policies, agencies towards achieving national objectives.
3. Achieving cluster efficiencies.
4. Directing attention to corruption and rent-seeking that hinders efficiency.
5. Developing SMEs to be part of global supply chain.
6. Providing feedback mechanisms from private sector to make supply chains more efficient and/or responsive.
7. Educating and re-training system to ensure labor force meets market needs and benefits from greater value creation.

In order to achieve these objectives and making 1MSC a success, there rise needs to have a single governing body to oversee and facilitate the integration. This body would be accountable to develop policies, guidelines and setup mechanism to bring all relevant stakeholders of the supply chain together. The authors proposed that this governing body be termed as National Development Incubator (NDI) and be setup as a government entity. This is the main driver for 1MSC in Malaysia that differs from current SCI concept.

NATIONAL DEVELOPMENT INCUBATOR (NDI)

The unique indicator that differentiates 1MSC from the conventional SCI is the introduction of NDI. NDI is a governing body that oversees the integra-

tion of all key players. NDI should develop unique and tailored responsibility for each of the players in the 1MSC namely i) Supplier, ii) Manufacturer, iii) Distribution Channel and iv) KR1M. Each of said elements is important to NDI. The NDI is targeted to address the challenges in SCI's implementation.

Supplier Vendor Program

In current supply chain models, individual firm focuses in selecting their own suppliers. Due to individual search, it resulted in a limited number of suppliers for each firm and focuses on small circle of network. The essential concept for the 1MSC is to introduce the supplier vendor program where selection of qualified pool of suppliers is done upfront by the NDI in order to address supplier's issue. According to Li & Zabinsky (2011), selecting the correct supplier has turned into a vital issue for firms. Each of the suppliers has its own strength and must fulfill the standard that has been imposed; however it is important for a firm to determine which suppliers are best suited. If it is not carried out as carefully and systematically as needed, there is an increased risk that the chosen suppliers will fail to perform effectively.

The common supply chain process which begins with limited number of suppliers, market dominating assembler and distributor is no longer relevant in the 1MSC context. Typically, the multiple local suppliers provide support either in goods or services to the manufacturers namely the SMEs or Multi-National Corporations (MNCs), however adhering to certain requirements or restrictions would impose some challenge to them and that is the issue that 1MSC expected to solve.

Ekici (2013) has outlined the four main criteria for suppliers' selection. Similarly, NDI have been adopting this criterion in the context of 1MSC:

- Quality of goods or services.
- Availability of goods or services which takes into account of supplier's geographic location.

- Supplier service and responsiveness which includes flexibility of supplier to respond to request.
- Cost efficiency by assessing price offered by supplier.

In addition to the above, assessing certifications from suppliers such as the ISO14000 (environmental management), which provide a guideline for firms that need to systematize and improve their environmental management efforts, the right attitude that supplier possesses to sustain the relationship, as well as periodically assessment through audit and monitoring actions are key elements in ensuring a successful supplier relationship (Brammer, Hoejmoose & Millington, 2011; Gimenez & Tachizawa, 2012).

Typically, supplier is expected to produce quality products at reasonable prices; however additional requirement that differentiate supplier selection in the 1MSC concept is the value added capabilities possess by the suppliers. Key differentiator of good supplier is the service that it provides. In the 1MSC context, all suppliers are all linked tightly under the NDI umbrella and deploy collaborative and strategic partnerships.

Cross-sourcing is an ideal solution to ensure no disruption of flow in the chain. According to Eltantawy, Giunipero, & Handfield (2014), cross sourcing can be achieved by having a pool of qualified supplier with same capabilities but each focusing on different segments, if anything would happen to supplier A for instance, supplier B has the capability to produce the same products and vice versa. All these potentially address supplier's competence requirement in SCI implementation.

Manufacturer

Manufacturers are required to register with the NDI in order for these firms to utilize the facilities that 1MSC has to offer as well as registering with government bodies such as the Ministry of Domestic Trade, Cooperative and Consumerism

(MDTCC). Manufacturing context in the 1MSC covers the SMEs and the MNCs. An ideal vision of the 1MSC is for the NDI to develop notable SMEs which have the competences to produce goods with high quality and globally exported. Where "Simplify, Focus and Integrate" is concerned, an intervention by the government in providing incentives, training and supports to the local SMEs is an important aspect for this concept to be realized as the initiative by the government will be able to assist the SMEs to compete both locally and globally (Zhang et al., 2015). It is also responsible to offer facilities and services to the players in the integrated chain such as training, equipment installation, SCM training and courses, IT support, retail and technical expertise. This is aligned to the need of education and re-training system that Sheng (2010) highlighted which will be benefiting the SMEs.

The MNCs is part of the integration as well where the MNCs provide job opportunities to locals. As a result, SMEs produces high level assembly and supplies the parts to MNCs, who has the ability either to produce its own finished good or enables the products to operate with its parts. In return, the Malaysian Government is offering tax relief to MNCs that wishes to make Malaysia a principal hub for their regional operations as well as other incentives for development in undeveloped areas, incentives for industrial area management and capital allowance to increase automation in labor-intensive industries (Mahalingam, 2015). Under NDI's supervision, business operation standards would potentially be maintained across different industries.

For products related to food and beverages (F&B) for instance, the firms must obtain certification from Department of Islamic Development Malaysia (JAKIM) and Ministry of Health (MOH) to ensure that the products sold are safe for consumption or usage, or the food is halal and certified by JAKIM. Firms must also adhere to regulation and audit as it validates that the process performed matches the stated procedure.

Distribution Channel

NDI is responsible to integrate distribution channel, where firms engage in logistics, warehousing and communication. The ideal concept in 1MSC context is to have a selection of logistics providers. Applying the integration concept, logistics activities need to be conducted with systematic approach within an interdepartmental context and closely integrated with external bodies such as shipping and freight agencies. A company's logistics strategy needs to be closely aligned with the supply chain's overall strategy. According to Ajmera & Cook (2009), outsourcing the overall SCM function to a designated third party, called fourth-party logistics (4PL) could provide greater expertise and a broader perspective as it serves a large number of customers across different supply chains. 4PL differs from 3PL as it is often a separate entity from the clients, acts as the point of contact between many logistics providers and its customers, as well its supply chain are managed by the 4PL. The integration could increase the effectiveness of information sharing with available advanced information technology (IT) infrastructure. This would address the challenge in handling logistics and ensure customers' needs are fulfilled accordingly.

Kedai Rakyat 1Malaysia (KR1M)

Another player under the NDI is the KR1M which operates in mini market format providing various basic necessities at low prices. The objective of KR1M is ensuring prices of goods are controlled and reducing the monopolization of global manufactured products. Referencing the 1MSC concept, in the context of supplier selection, NDI is responsible to provide list of SMEs that are able to manufacture products for KR1M. Those individuals or firms that would like to be part of the KR1M operation, has to register with government agency which is the MDTCC. Furthermore, since KR1M offers staple goods and basic necessities, it is compulsory for KR1M owners to obtain ap-

proval from JAKIM and MOH. This is to ensure the quality and safety of the products. The process between KR1M and other firms selling goods and services are not similar. The conventional firms do not need to obtain approval from government for operations. However, since the aim of KR1M is to offer cheaper and quality products to Malaysian consumer, it is understandable for KR1M to obtain these approvals. Cheaper is somehow associated with lack of quality and low safety standards. Thus, KR1M is a unique element of 1MSC.

BENEFITS OF 1MALAYSIA SUPPLY CHAIN (1MSC)

As outlined by Zailani (2011), the objective of 1MSC is to reach sustainable development be it for society, economic or environment. Currently, it is not compulsory for firms to model their business towards sustainable development. However, the acceptance of society and government encourages firms to think about environment, social as well as economic contributions. A sustainably IMSC concept is aimed to achieve the following:

Society

Satisfied customer is an important determinant of repurchase intention (Liao, Palvia, & Chen, 2009) and customer loyalty (Eggert & Ulaga, 2002). In the context of 1MSC, KR1M is expected to supply high quality products with reasonable price resulting in customer satisfaction. The players in 1MSC join forces to meet social needs by empowering consumers in Malaysia through opening its door to society to operate their own business. KR1M also offers society to open KR1M all around Malaysia to ensure everyone is able to purchase quality and cheaper products. Furthermore, KR1M, SMEs and MNCs are into the same network and would possibly able to penetrate international market through exports and adhering to regulations to fulfill international standards. Another unique factors that 1MSC able to offer is the contribution

to the humanitarian aid. Since this notion integrates all players in the supply chain such as the government agencies and distribution channels, 1MSC is expected to pull its players to be able to contribute back to the Malaysian government such as during flood. KR1M for instance will be able to contribute by providing and distributing much needed relief items such as food supplies, hygiene kits and blankets to affected communities.

Economy

The concern of increasing income inequality in Malaysia has become topical. Since the New Economic Policy in 1971, the Malaysian government has regularly executed development strategies by “growth with equity” resulted in the remarkable growth and structural revolution of the Malaysian economy and a substantial reduction in income disparity, however, these strategies which have successfully lessened the poverty rates in the rural sector; but has worsened the income inequality a few decades later (Saiful Nathan & Mohamad, 2016). Since 1MSC focuses in providing opportunity to SMEs, this notion is expected to contribute to economy growth by solving income disparity among society. The KR1M effort to open opportunity nationwide for the people to be part of the network chain is one way to reduce the unemployment issue. By reducing the gap and enabling firms to invest back into the economy through consumption, Malaysia economy will be in a much better shape.

Environment

In the context of environment, the 1MSC is estimated to enforce the need for ISO certification for the players. According to Murray (1999), when a firm gains ISO 14001 certification, it is one way to demonstrate to its customers, suppliers, competitors and the regulators that it is serious about environmental stewardship. With NDI overseeing the 1MSC processes, firms are expected to adhere

to the standards and procedure that the government that sets in.



SOLUTIONS AND RECOMMENDATIONS

The previous sections suggest the concept of 1MSC lies at the NDI's formation. Hence, this chapter recommends the government to build NDI to be resourceful in supporting the 1MSC concept. With NDI as the backbone of 1MSC notion, Malaysian government could implement 1MSC concept to a specific industry to test its effectiveness before proliferating to other industry. In order to be effective, once NDI is established it must be a policymaker to define the guidelines and incentives and enforcement has to be made by the authority.

Another recommendation is to create cluster based economic activities which gathers the experts in the same industry. With the success of KR1M, another recommendation is to create an ecosystem for entrepreneurship with the appropriate skilled labor and talent. This would encourage more individuals to become entrepreneur, hence contributing to the sustainability development.

The success of 1MSC concept can be geared towards the implementation of global SCI as well as achieving supply chain performance. The goal is to deploy one integrated, but globally dispersed, value chain or network that serves the entire worldwide business. With the introduction of 1MSC, firms in Malaysia can compete globally, hence supporting the notion of supply chain versus supply chain. Hence, integration between local and global players depends on the successful factors of 1MSC.

FUTURE RESEARCH DIRECTIONS

There is a need for further research to capture the characteristics of successful 1MSC implementa-

tion to other industry. This is to test if the concept is able to be replicated and accepted globally. It is also worthwhile to study the factors determining the level of 1MSC implementation. The concept of 1MSC is not extensively explored, thus any gaps has to be filled with further research. The impact of 1MSC on business performance and challenges associated with it could also be researched and explored further.

CONCLUSION

In current economic situation, firms create and integrate their own network. Implementation of such individual integration by the respective firm causes an issue where the firm may not have adequate knowledge, resource and funds to face global challenges. The 1MSC addresses this issue by having a nationwide integration and governs by the NDI. With this setup in place it is hoped that the firm will have adequate skills and resources which would enable firms to produce and maximize firms' output. The increment in individual output will increase the nation's GDP. This ensures sustainability in terms of social, environment and economic which relates back to 1MSC objectives and achieve the overall NEM objectives.

REFERENCES

- Ajmera, A., & Cook, J. (2009). A multi-phase framework for supply chain integration. *SAM Advanced Management Journal*, 74(1), 37.
- Awad, H. A., & Nassar, M. O. (2010a). Supply chain integration: definition and challenges. *Management and Technology*, 1(1).
- Awad, H. A., & Nassar, M. O. (2010b). A broader view of the supply chain integration challenges. *International Journal of Innovation. Management and Technology*, 1(1), 51.
- Bagchi, P. K., Chun Ha, B., Skjoett-Larsen, T., & Boege Soerensen, L. (2005). Supply chain integration: A European survey. *The International Journal of Logistics Management*, 16(2), 275–294. doi:10.1108/09574090510634557
- Brammer, S., Hoejmoose, S., & Millington, A. (2011). *Managing sustainable global supply chain: framework and best practices*. Academic Press.
- Chen, H., Mattioda, D. D., & Daugherty, P. J. (2007). Firm-wide integration and firm performance. *The International Journal of Logistics Management*, 18(1), 5–21. doi:10.1108/09574090710748144
- Cousins, P. D., & Menguc, B. (2006). The implications of socialization and integration in supply chain management. *Journal of Operations Management*, 24(5), 604–620. doi:10.1016/j.jom.2005.09.001
- Danese, P., & Romano, P. (2011). Supply chain integration and efficiency performance: A study on the interactions between customer and supplier integration. *Supply Chain Management: An International Journal*, 16(4), 220–230. doi:10.1108/13598541111139044
- Eggert, A., & Ulaga, W. (2002). Customer perceived value: A substitute for satisfaction in business markets? *Journal of Business and Industrial Marketing*, 17(2/3), 107–118. doi:10.1108/08858620210419754
- Ekici, A. (2013). An improved model for supplier selection under capacity constraint and multiple criteria. *International Journal of Production Economics*, 141(2), 574–581. doi:10.1016/j.ijpe.2012.09.013
- Eltantawy, R., Giunipero, L., & Handfield, R. (2014). Strategic sourcing managements mindset: Strategic sourcing orientation and its implications. *International Journal of Physical Distribution & Logistics Management*, 44(10), 768–795. doi:10.1108/IJPDLM-02-2014-0045



- Fabbe-Costes, N., & Jahre, M. (2007). Supply chain integration improves performance: The Emperors new suit? *International Journal of Physical Distribution & Logistics Management*, 37(10), 835–855. doi:10.1108/09600030710848941
- Fawcett, S. E., & Magnan, G. M. (2002). The rhetoric and reality of supply chain integration. *International Journal of Physical Distribution & Logistics Management*, 32(5), 339–361. doi:10.1108/09600030210436222
- Flynn, B. B., Huo, B., & Zhao, X. (2010). The impact of supply chain integration on performance: A contingency and configuration approach. *Journal of Operations Management*, 28(1), 58–71. doi:10.1016/j.jom.2009.06.001
- Georgise, F. B., Thoben, K. D., & Seifert, M. (2014). Supply Chain Integration in the Manufacturing Firms in Developing Country: An Ethiopian Case Study. *Journal of Industrial Engineering*.
- Gimenez, C. (2006). Logistics integration processes in the food industry. *International Journal of Physical Distribution & Logistics Management*, 36(3), 231–249. doi:10.1108/09600030610661813
- Gimenez, C., & Tachizawa, E. M. (2012). Extending sustainability to suppliers: A systematic literature review. *Supply Chain Management: An International Journal*, 17(5), 531–543. doi:10.1108/13598541211258591
- Guan, W., & Rehme, J. (2012). Vertical integration in supply chains: Driving forces and consequences for a manufacturers downstream integration. *Supply Chain Management: An International Journal*, 17(2), 187–201. doi:10.1108/13598541211212915
- Huang, M. C., Yen, G. F., & Liu, T. C. (2014). Reexamining supply chain integration and the suppliers performance relationships under uncertainty. *Supply Chain Management: An International Journal*, 19(1), 64–78. doi:10.1108/SCM-04-2013-0114
- Lee, H. L. (2000). Creating value through supply chain integration. *Supply Chain Management Review*, 4(4), 30–36.
- Leuschner, R., Rogers, D. S., & Charvet, F. F. (2013). A Meta-Analysis of Supply Chain Integration and Firm Performance. *Journal of Supply Chain Management*, 49(2), 34–57. doi:10.1111/jscm.12013
- Li, L., & Zabinsky, Z. B. (2011). Incorporating uncertainty into a supplier selection problem. *International Journal of Production Economics*, 134(2), 344–356. doi:10.1016/j.ijpe.2009.11.007
- Liao, C., Palvia, P., & Chen, J. L. (2009). Information technology adoption behavior life cycle: Toward a Technology Continuance Theory (TCT). *International Journal of Information Management*, 29(4), 309–320. doi:10.1016/j.ijinfomgt.2009.03.004
- Lo, S. M., & Power, D. (2010). An empirical investigation of the relationship between product nature and supply chain strategy. *Supply Chain Management: An International Journal*, 15(2), 139–153. doi:10.1108/13598541011028741
- Mahalingam, E. (2015). *Tax relief for MNCs that make Malaysia a principal hub*. Retrieved online on March 5, 2016 from <http://www.thestar.com.my/business/business-news/2015/04/07/tax-relief-for-mncs-that-make-malaysia-a-principal-hub/>
- Mohr, J., & Spekman, R. (1994). Characteristics of partnership success: Partnership attributes, communication behavior, and conflict resolution techniques. *Strategic Management Journal*, 15(2), 135–152. doi:10.1002/smj.4250150205
- Murray, P. C. (1999). Inching toward environmental regulatory reform—ISO 14000: Much ado about nothing or a reinvention tool? *American Business Law Journal*, 37(1), 35–72. doi:10.1111/j.1744-1714.1999.tb00268.x

- National Economic Advisory Council. (2009). *New Economic Model For Malaysia Part 1*. Retrieved online on Mar 16, 2016 from <http://www.epu.gov.my/epu-theme/pdf/nem.pdf>
- Power, D. (2005). Supply chain management integration and implementation: a literature review. *Supply Chain Management: An International Journal*, 10(4), 252–263.
- Saiful Nathan, S. B., & Mohamad, M. R. (2016). Distributional effects of non-farm incomes in a Malaysian Rice Bowl. *International Journal of Social Economics*, 43(2), 205–220. doi:10.1108/IJSE-09-2013-0200
- Sezen, B. (2008). Relative effects of design, integration and information sharing on supply chain performance. *Supply Chain Management: An International Journal*, 13(3), 233–240. doi:10.1108/13598540810871271
- Sheng, A. (2010). *NEM and the 1Malaysia Supply Chain The Way Forward for Malaysia to Transform into a High Income Nation*. Retrieved online on Oct 5, 2015 from <http://www.mca.org.my/English/Download%20Doc/2010/14082010Chinese%20Economy%20Congress/1.2Andrew%20Sheng.MCA%20Presentation%2014%20August%202010%20Final%20Draft.pdf>
- Stock, J. R., Boyer, S. L., & Harmon, T. (2010). Research opportunities in supply chain management. *Journal of the Academy of Marketing Science*, 38(1), 32–41. doi:10.1007/s11747-009-0136-2
- Van der Vaart, T., & van Donk, D. P. (2008). A critical review of survey-based research in supply chain integration. *International Journal of Production Economics*, 111(1), 42–55. doi:10.1016/j.ijpe.2006.10.011
- Yu, Z., Yan, H., & Edwin Cheng, T. C. (2001). Benefits of information sharing with supply chain partnerships. *Industrial Management & Data Systems*, 101(3), 114–121. doi:10.1108/02635570110386625
- Yunus, E. N., & Tadisina, S. K. (2016). Drivers of supply chain integration and the role of organizational culture: Empirical evidence from Indonesia. *Business Process Management Journal*, 22(1), 89–115. doi:10.1108/BPMJ-12-2014-0127
- Zailani, S. (2011). *1Malaysia Supply Chain: From myth to reality*. Retrieved online on December 23, 2015 from [https://sukaweb.penang.gov.my/pkn/akhbar.nsf/981d6336478572e5482575e6002c4e9f/b86aa9ff616becca482578e200133d74/\\$FILE/scan0006.pdf](https://sukaweb.penang.gov.my/pkn/akhbar.nsf/981d6336478572e5482575e6002c4e9f/b86aa9ff616becca482578e200133d74/$FILE/scan0006.pdf)
- Zhang, C., Gunasekaran, A., & Wang, W. Y. C. (2015). A comprehensive model for supply chain integration. Benchmarking. *International Journal (Toronto, Ont.)*, 22(6), 1141–1157.

ADDITIONAL READING

- AlSagheer, A., & Ahli, M. (2011). Impact of supply chain integration on business performance and its challenges. *The International Business & Economics Research Journal (Online)*, 10(12), 79.
- Beck, R., Koenig, W., & Weitzel, T. (2002). Promises and pitfalls of SME integration. BLED 2002 Proceedings, 34.

KEY TERMS AND DEFINITIONS

1 Malaysia Supply Chain: Collaborated and moved towards meeting the needs of all its stakeholders and providing platform to support business to improve value chains and improving sustainable practices.

Kedai Rakyat 1 Malaysia: Mini market format providing various basic necessities at low prices.

National Development Incubator: Governing body that oversees the integration of all key players.

New Economic Model: Malaysian plan to increase per capita income.

Supply Chain Integration: Information, coordination, and organizational linkage across supply chain activities among partners.

Supply Chain Performance: Outcome of supply chain activities among partners.

Sustainable Development: Achieving goal for social, economy and environment.



A Review of Advances in Supply Chain Intelligence

Nenad Stefanovic

University of Kragujevac, Serbia

Danijela Milosevic

University of Kragujevac, Serbia

INTRODUCTION

As the market pressures are forcing supply chain organizations to adapt to new business models, collaboration, integration and information sharing are becoming even more critical for the ultimate success. Supply chains are experiencing a major structural shift as more organizations rely on a community of partners to perform complex supply chain processes.

While supply chains are growing increasingly complex, from linear arrangements to interconnected, multi-echelon, collaborative networks of companies, there is much more information that needs to be stored and analyzed than there was just a few years ago.

Supply chains are complex systems with silos of information that is very difficult to integrate and analyze. The best way to effectively analyze these disparate systems is the use of business intelligence (BI). The ability to make, and then to process, the right decision at the right time in collaboration with the right partners is the definition of the successful use of BI (Stefanovic & Stefanovic, 2009).

During the past two decades companies have made large investments in supply chain management (SCM) information systems in order to improve their businesses. However, these systems usually provide only transaction-based functionality and mostly maintain operational view of the business. They lack sophisticated analytical capabilities required to provide an integrated view of the supply chain (Baars et al., 2014).

Supply Chain Intelligence (SCI) is relatively new initiative that provides the capability to improve supply chain performance by utilizing sophisticated analytical tools and collaborative decision making (Haydock, 2003). SCI takes broader, multidimensional view of supply chain in which, using patterns and rules, meaningful information about the data can be discovered. Supply chain intelligence reveals opportunities to reduce costs and stimulate revenue growth and it enables companies to understand the entire supply chain from the customer's perspective (Stefanovic et al. 2007).

Nevertheless, companies that implemented some kind of enterprise business intelligence systems still face many challenges related to data integration, storage and processing, as well as data velocity, volume and variety. Additional issues include lack of predictive intelligence features, mobile analytics and self-service business intelligence capabilities.

In this chapter the latest supply chain management issues, and the drivers for the implementation of the business intelligence and performance measurement systems are discussed. Review of the most important literature and research findings provides condensed view of the existing state of the art. Additionally, the latest software technologies and tools, and their impact on different supply chain areas such as collaboration, integration, and analytics are described.

Furthermore, the integrated supply chain intelligence system that enables creation of pervasive

analytical systems for collaborative planning, monitoring and management of the supply network is described. This includes architecture, main components, technologies and tools. The SCI system is robust and cloud-based, capable to handle big data analytical tasks. Its flexible and multilayered architecture enables creation of adaptive supply chain intelligence systems by composing various analytical software components, services and tools. Finally, the main trends and advanced information technologies that will shape the future SCI systems are introduced.

BACKGROUND

Today, there are variety of business initiatives and technologies such as joint planning and execution, business intelligence, performance management (PM), data mining and alerting that can be used for more efficient supply chain management.

By applying the concepts of business intelligence to data from SCM systems, SCI technologies seek to provide strategic information to decision makers (Reddy, 2004). Information categories range from what-if scenarios for reconfiguring key functions in sourcing, manufacturing, and distribution to measuring the ability of a supply chain to produce cost-effective products. Table 1 summarizes the main differences between the SCM and SCI systems (Russom, 2010).

What truly differentiates SCI from BI is the ability to collect and aggregate data across the value chain. Data is then analyzed and the results distributed to all parties along that chain, regardless of location.

SCI technologies promise to extract and generate meaningful information for decision makers from the enormous amounts of data generated and captured by SCM systems. SCI complements supply chain planning because BI applications provide incremental benefits while a business lays the foundation for more sophisticated tools and related business process changes.

Table 1. Comparison of SCM and SCI

Supply Chain Management	Supply Chain Intelligence
Largely about managing the procurement and production links of the supply chain	Provides a broad view of an entire supply chain to reveal full product and component life cycle
Transactional	Analytical
Tactical decision making	Strategical, tactical and even operational decision making
Helps reduce costs through improved operational efficiency	Reveals opportunities for cost reduction, but also stimulates revenue growth
Usually just the SCM application's data (as a vertical stovepipe)	Integrates supplier, manufacturing and product data; possibly syndicated data, too (horizontal)
Records one state of data representing "now"	Keeps a historic record and makes predictions
Assists in material and production planning	What-if forecasting based on historic data
Quantifies cost of some materials	Enables an understanding of total cost
Shows today's yield but cannot explain influences on it, thus provides no help for improvements	Drills into yield figures to reveal what caused the performance level so it can be improved
Simple reporting	Collaborative environment with personalizable monitoring of metrics

The primary source systems for BI are the internal operational systems, while SCI integrates data from partner and supplier information systems. What truly differentiates SCI from BI is the ability to collect and aggregate data across the value chain. Data is then analyzed and the results distributed to all parties along that chain, regardless of location.

Several factors are driving the demand for BI within the SCM domain. First, organizations want more granular visibility to their SCM processes so they can manage and control them more effectively. They also want to identify negative trends in supply chain performance, and identify root causes as early as possible to take corrective action. And they need to conduct 'what-if' analyses to evaluate the service and cost trade-offs of different supply



chain strategies and tactics (Partridge, 2013). The findings suggest the existence of a statistically significant relationship between analytical capabilities and performance (Trkman et al., 2010).

Business intelligence within the supply chain improves internal efficiencies and accountability while saving time and eliminating costs with metrics-driven decision-making and change management. It allows companies to enable more predictable business performance by putting actionable information into the hands of key decision makers (Dowse, 2015).

Most BI implementations are plagued by barriers that limit the strategic value of BI as a building block of the real-time supply chain. Most organizations have experienced the following pitfalls (Lim et al., 2013).

- There is no well-established enterprise architecture for business intelligence.
- Almost all of these tools are based on client/server architectures rather than the Web.
- Data access and the exclusive reliance on a data warehouse.

Having in mind all these drawbacks, as well as large expectation of the BI systems, a new generation of BI systems called BI 2.0 was introduced. The main characteristic of the BI 2.0 system are (Taylor, 2007):

BI 2.0 Integrates Events: BI systems perform event monitoring and react according to predefined business rules, which are prerequisites for the real-time supply network and automatic alerting.

BI 2.0 Delivers Insight: It is not only enough to deliver information via reports, but also to provide business users with additional analytics and instructions for next steps.

BI 2.0 Delivers Action Without Intervention: Based on the embedded expert system, corrective action without human intervention can be taken.

BI 2.0 is Real-Time: Probably the single most agreed feature - all of this must be in real-time or close to it.

BI 2.0 is Based on SOA and Web 2.0: Like most things new BI 2.0 is expected to be service-based and Web 2.0 technologies are expected to impact the user interface for BI 2.0.

In order to achieve required supply chain agility and adaptivity, it is necessary to use intelligent technologies and tools which enable monitoring and evaluation of supply chain performance. Supply chain requires that member companies have the means to assess the performance of the overall supply chain to meet the requirements of the end customer. In addition, it is necessary to be able to assess the relative contribution of individual member companies within the supply chain. This requires a performance measurement system that can operate not only at several different levels but also can link or integrate the efforts of these different levels to meet the objectives of the supply chain.

SCM, has received a remarkable attention from both academia and industry since the last decade, however, there is still lack of integration between SCM systems and performance management systems. The huge majority of performance measurement models and frameworks has focused on single organizations or cover specific type of performance such as financial. There are several performance measurement approaches specifically designed for the supply chain management domain (Chan et Al., 2006). Companies have to measure performance at strategic, tactical and operational levels with metrics dealing with sourcing, making, delivering and customer services (Gunasekaran, 2004).

The main goal of supply chain PM systems is business process optimization through monitoring and analysis of Key Performance Indicators (KPIs). These performance measures enable supply chain companies to align processes and activities with strategic objectives. Key performance indicators are normally backward-looking

because they are based on historical information and often do not help in forecasting future events or performance (Lappide, 2010). Predictive metrics would make possible to predict future problems in supply chain operations and to enable proactive evaluation and improvement actions in advance. Statistical modelling and data mining under the guise of predictive analytics have become critical building blocks in setting the new KPI standard for leading indicators (Gunasekaran et al., 2016).

There has been lot of research efforts during the last decade that deal with various aspects of supply chain performance management. Shepherd and Gunter provide taxonomy of supply chain performance measurement systems and metrics with critical review of the contemporary literature (Shepherd & Gunter, 2006). The study shows that despite substantial advances in the literature in the last decade, there are still important topics related to supply chain performance management that did not receive adequate attention, such as process modeling, data integration, software support, and forecasting. Estampe et al. (2013) provide analysis of various supply chain performance models by stressing their specific characteristics and applicability in different contexts, so that decision-makers can evaluate and apply models and metrics that best suit their needs. Akyuz and Erkan (2010) also provide a critical literature review on supply chain performance measurement. The results show that performance measurement in the new era is still an open area of research. This particularly refers to framework development, collaboration, agility, flexibility, and IT support systems. Gopal and Thakkar (2012) contribute a comprehensive review of supply chain performance measurement systems. The analysis shows that in spite of considerable evidence from the literature, there is a large scope for research to address critical issues in supply chain performance measurement, including: metrics, benchmarking, integration, business intelligence, and collaborative decision making.

The resolution to these problems can be integration of different BI and PM tools and technologies into the supply network collaboration infrastruc-

ture based on Internet technologies. This means packaging analytical and PM applications into corporate portal that business users can access online and then allowing them to share not just the static output, but also the actual dynamic analytical experience through online collaboration.

In the following section, a comprehensive supply chain intelligence model and its realization through a specialized SCI web portal which acts as a unique gateway to different information sources and provide services for supply network coordination, planning, analysis, monitoring and data mining is presented (Stefanovic, 2011). The model and the supporting software solution are developed having in mind some drawbacks of the existing systems, as well as characteristics of the future SCI systems.

SUPPLY CHAIN INTELLIGENCE MODEL

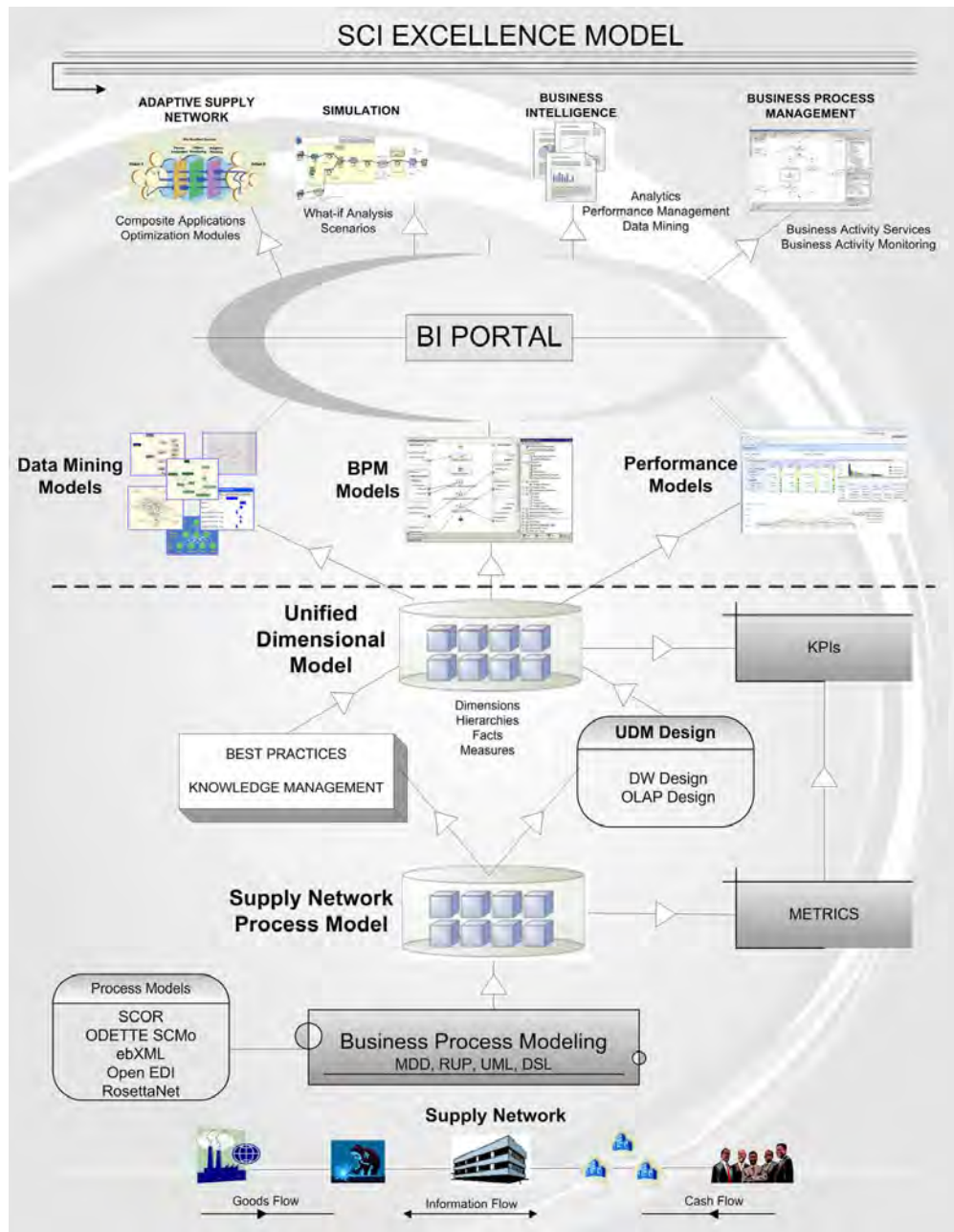
The core elements of the SCI model are process model with metrics, data warehouse and SCI web portal. The process modeling method and the data warehouse dimensional modeling approach, as well as the layered service-oriented composite web application introduce a new breed of robust and flexible SCI systems that can be relatively easily implemented, customized and personalized. This way, previously separate models, tools and technologies are seamlessly integrated into the complete SCI system.

The integrated SCI model should ultimately provide all the main capabilities required for the successful supply chain management (Figure 1).

BI web portal holds the central place in the model, as it comprises a portfolio of analytic, collaboration and communication services that connect people, information, processes, and systems both inside and outside the corporate firewall.

Supply chain planning subsystem provides a consistent and timely view into a business by using data from existing systems like data warehousing (DW) or On-Line Analytical Processing (OLAP)

Figure 1. Capabilities of the SCI System



cubes, without changing those systems. It provides centrally managed data, business metadata, business rules, business logic, views, forms, report definitions, process control, and security.

Data mining techniques can be applied to many supply chain applications, answering various types of business questions, such as demand forecasting,

inventory prediction (Stefanovic, 2015), customer and product segmentation, risk management, etc. They can be used also to perform key performance indicators (KPI) predictions in order to take proactive actions (Stefanovic, 2014).

Analytics and reporting subsystem should provide unique tools for decision makers to explore

Figure 2. Supply Chain Intelligence Model



large quantities of data in a way that provides rapid insights. These can include advanced analytic capabilities ranging from multidimensional “slice and dice” to decomposition trees, drill-across, root-cause analysis, self-service report models and information on-demand.

In order to fully realize these SCI capabilities, we need a supply network-wide business excellence model that will provide consistent framework for establishing, modelling, managing, measuring, and improving supply network processes. Figure 2 illustrates the overall supply chain intelligence model.

The proposed SCI model unifies processes, methodologies and tools into a single business solution. SCI model has been developed in such a way to seamlessly integrate within overall BI and collaboration framework (Stefanovic et al., 2011).

The SCI model is process-centric, metrics-based and modular. It introduces the new supply network and data warehouse modeling approaches, as well as layered application architecture which enables creation of composite BI systems.

In contrast to most other BI models that usually approach supply chain from the functional perspective or treat only internal processes, our approach uses the standardized supply chain process model and metrics. The novel process modeling method

enables modeling of any supply network regardless of the number of companies, process types and relationships. Process, metrics, interactions, flows and other elements (for example, resources and constraints) can be modeled at different levels of detail, thus serving as a basis for dimensional modeling, KPI design, simulation and business process management.

The dimensional modeling method offers best practices and design guidelines for developing main data warehouse elements. The unified dimensional model (UDM) serves as an extra layer that isolates business logic and logically unites heterogeneous data sources (such as relational databases, data marts, spreadsheets, web services, etc.) into the single data resource. Also, comparing to traditional DW approaches, the unified dimensional modeling approach allows design of additional DW elements such as complex calculation, hierarchies, measure groups, and KPIs. All the DW elements can be reused as templates since the definitions are stored as XML files. Analytical models can be built in both, traditional multidimensional OLAP and in-memory tabular modes, depending on the business requirements and priorities. This architecture of the model provides scalability and the performance is improved by using the in-memory data store.

The data warehouse approach allows the single version of the truth because of the integrated and consolidated data repository. Server-based definition of KPIs offers centralized metadata management and enables creation of various front-end BI and analytical applications. Data warehouse acts as an integrated data source upon which different BI elements are developed such as planning models, performance monitoring systems, reporting and data mining models.

SCI Web Portal

The main problem with traditional BI tools is problem with delivering information to end-users or other systems. They usually require special client tools, support only certain data sources (for example ERP systems) and they are difficult or impossible to customize or integrate. That is why our SCI model introduces specialized web portals for analytics, monitoring, planning, data mining and simulation. In order to fulfill these functions, BI portal is built as a layered service-oriented application which utilizes the newest web technologies for the special BI tasks.

The concept is based on service-oriented architecture, models and templates, which should enable creation of the new breed of SCI solutions that connects business processes, data, applications, and users into a single integrated system.

The idea is to define BI service-oriented architecture which provides a standard and flexible approach for integration of business processes, analytics, information, business transactions and collaboration tools. This approach enables separation of business processes and applications, and creation of data, application, and collaboration services on demand and based on events. These services can be further arranged into composite BI applications for particular business processes and activities.

In order to overcome the shortcomings of the existing BI and PM client tools, the specialized web portal that enables supply chain users to monitor business processes, collaborate and take

actions is designed (Stefanovic & Stefanovic, 2011). It has been successfully implemented as a pilot project in an automotive company. Automotive supply chain is typically very complex, with many organizations, intertwined processes, and multitude of users with different requirements.

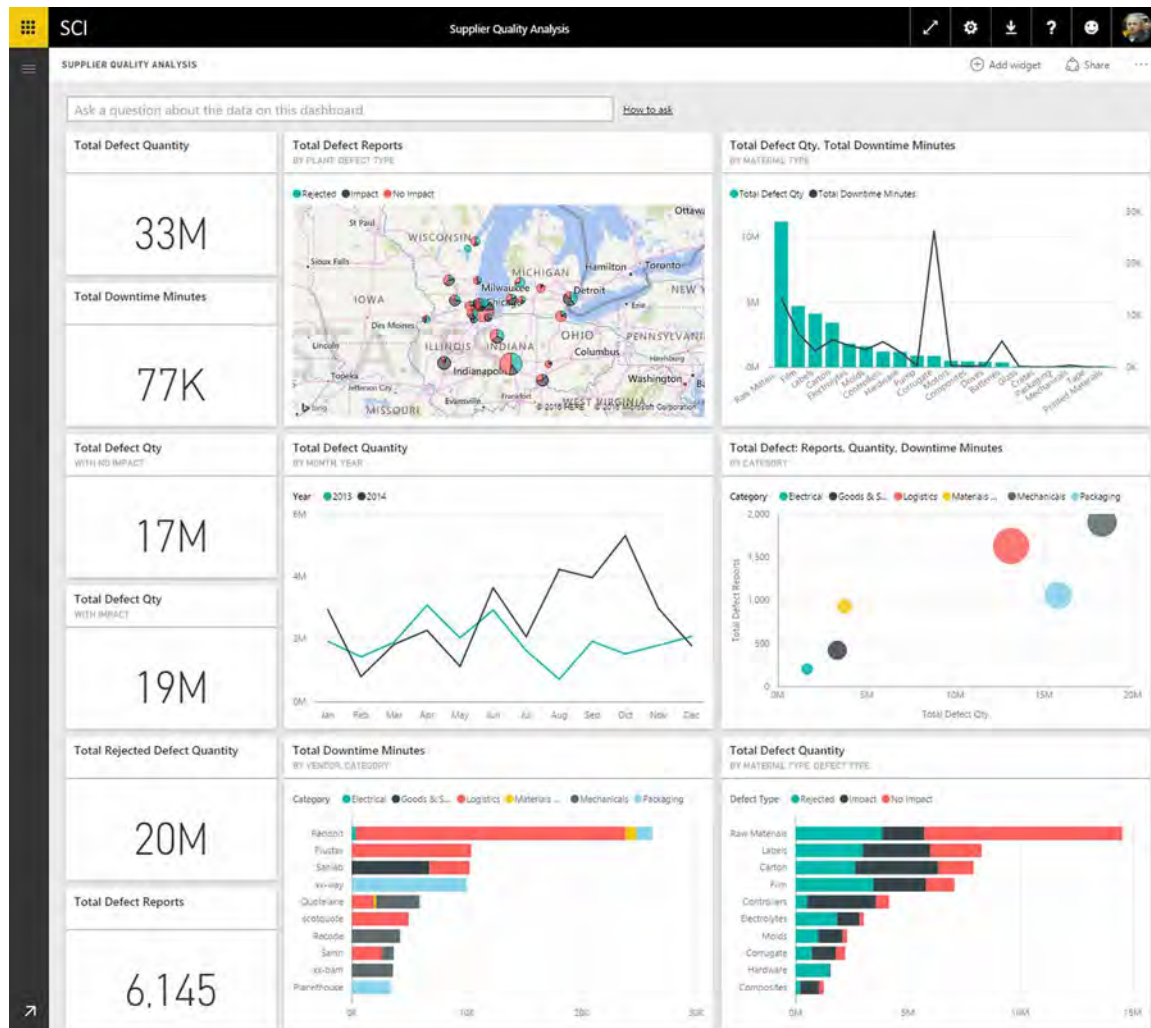
The portal represents a single point of access to all relevant information in a personalized and secured manner. Its composite and service-oriented architecture enables inclusion of different PM components and tools (KPIs, dashboards, scorecards, strategy maps, reports, etc.). PM elements can be personalized and adjusted, and information can be filtered just by using a web browser. PM elements can be defined within the portal and also embedded from the external source (OLAP, another application, spreadsheets) via web services. This information is presented through special analytical web parts. The portal itself can be a provider (via web services or RSS-Really Simple Syndication) to other applications.

BI portal is a composite web application made of certain elements. Elements can be viewed at four distinct layers: presentation layer, collaboration layer, application layer and data layer. The portal is modular and each module can be modified and customized and also new modules can be added, thus assembling composite BI applications that fit specific user needs.

Modular architecture of the portal enables creation of various BI mashups tailored to specific business needs. Web parts, pages and also the complete portals can be saved as templates and reused many times with minimum effort.

Figure 3 shows the analytical dashboard page with different web parts that display key performance indicators, charts and graphs. Each web part can have different data source such as OLAP KPI, spreadsheets, report server, web portal database data feeds, etc. Since each data source can reside at different supply chain company, there is a specialized data gateway that acts as a bridge, providing quick and secure data transfer and refresh between the analytical service and on-premises data sources. Such dashboards and

Figure 3. SCI Dashboard Page



reports can be embedded in SCI portals along with other BI web parts thus composing flexible and customized analytical solutions.

A system of templates and taxonomies enables creation of a number of hierarchical portals. For instance, it is possible to create a global supply network BI portal, several company-level BI subportals, and many division-level subportals.

Additionally, the portal supports events and automatic alerting. Users can subscribe to specific documents or keywords and KPIs, to be notified when metrics are updated or new intelligence becomes available. They can also use other features, such as planning, enterprise search, subscription,

and routing functions, to work with team members on a single item (i.e. scorecard, KPI, etc.), and to automate collaborative performance management processes. The portal also provides fine-grained authentication and authorization.

On the other hand, software system architecture enables relatively easy and fast customization and extensibility. For example, existing KPI components can be easily reused many times, and only needed items can be added to web sites and pages. Also, a new (custom) lower-level metrics can be defined in the SCM Metamodel and realized in the data warehouse, thus allowing company-specific and industry-specific performance measurement.

Self-service BI is enabled through the incorporation of particular analytical models that enable end-users to create various reports and queries without additional intervention of the IT staff.

All these capabilities make the proposed SCI model and the software system extremely flexible, and applicable in various supply chain scenarios and different industries. Being based on the SCM Metamodel and standardized metrics, ensures general applicability (in various business domains) (Stefanovic & Stefanovic, 2008).

By implementing the presented BI portal it is possible to deliver insights to the right person, any time, via any device and in a secure manner. Personalization and filtering capabilities enable end users to access information relevant to them. All this features allow supply chain partners to bring more informed decision collaboratively.

FUTURE RESEARCH DIRECTIONS

As financial analysts have pointed out (Gartner, 2009), the BI software segment is a bright spot in the software industry and continues to grow even during the world economic crisis. BI is ranked as the highest technology priority among CIOs. On the other side, SCM market also continues to grow and the number of companies who adopted and implemented SCM concepts and systems constantly increases. These trends will certainly boost the demand for the SCI set of processes, technologies and tools.

Besides holding a huge market, these technologies have been the subject of intense academic research and study. As an interdisciplinary science field, SCI will certainly be one of the very active research topics in the years to come, since it comprises various operations management disciplines and IT domains like business intelligence, big data, collaboration, service-oriented architecture, cloud technologies and web portals.

In the next few years, research and development in BI domain is expected to rise. The main

advances can be grouped into the following categories (Gartner 2016):

- Infrastructure (cloud BI, security and user administration and data source connectivity).
- Data management (governance and meta-data management, self-contained extraction, transformation and loading (ETL), and data storage, self-service data preparation).
- Analysis and content creation (Embedded advanced analytics, analytic dashboards, interactive visual exploration, mobile exploration and authoring).
- Sharing of findings (embedding analytic content, publishing analytic content, collaboration and social BI).

In this era of Big Data, even while BI 2.0 is still maturing, BI 3.0 is on the way. BI 3.0 provide high performance analytics with in-memory technologies, and put emphasize on social networking, collaboration, as well as ubiquitous and assistive analytics. These new and potentially revolutionary technologies present a great opportunity for research and development for both industry and academia (Schoenherr & Speier-Pero, 2015).

CONCLUSION

To succeed in a competitive marketplace, an agile supply chain requires SCI system to quickly anticipate, adapt, and react to changing business conditions. SCI systems provide sustainable success in a dynamic environment by empowering business users at all levels of the supply chain and enabling them to use actionable, real-time information.

Many of forward-thinking supply networks have realized that business intelligence networks are a first step to information consolidation and gaining visibility over the value chain. With BI networks, businesses can share information with customers, suppliers, and partners.

The presented SCI model fuses all relevant elements such as people, business processes, metrics, KPIs and BI tools into a single comprehensive system. The underlying architecture supports the complete cycle of BI/PM processes like ETL, DW, OLAP and reporting. The unified data model helps to establish a single version of the truth on supply network performance, which is quantifiable and understood by all entities in the supply network.

SCI web portal is the unifying component that provides integrated analytical information and services, and also fosters collaborative decision making. The portal offers the following benefits:

- Real-time supply network monitoring and alerting.
- Flexibility, personalization and customization.
- Integration with existing transactional, BI and collaboration systems.
- Built-in knowledge and best practices through predefined web parts and templates.

The proposed layered SCI architecture allows construction of the next-generation loosely-coupled SCI networks that are built as BI web services. This will enable collaborative, efficient, responsive, and adaptive supply chains.

REFERENCES

- Akyuz, A. G., & Erkan, E. T. (2010). Supply chain performance measurement: A literature review. *International Journal of Production Research*, 48(17), 5137–5155. doi:10.1080/00207540903089536
- Baars, H., Felden, C., & Gluchowski, P. (2014). Shaping the Next Incarnation of Business Intelligence. *Bus Inf Syst Eng*, 6(11). doi:10.1007/s12599-013-0307-z
- Chan, F. T. S., Chan, H. K., & Qi, H. J. (2006). A review of performance measurement systems for supply chain management. *International Journal of Business Performance Management*, 8(2-3), 110–131. doi:10.1504/IJBPM.2006.009032
- Curt. (2002). *Supply Chain Intelligence: Applying Business Intelligence to Enhance Operational Efficiencies*. Wipro.
- Dowse, S. (2015). *Why Supply Chains Need Business Intelligence*. Retrieved September 23, 2015, from <http://www.inboundlogistics.com/cms/article/why-supply-chains-need-business-intelligence/>
- Estampea, D., Lamouri, S., Paris, J.-L., & Brahim-Djelloul, S. (2013). A framework for analysing supply chain performance evaluation models. *International Journal of Production Economics*, 142(2), 247–258. doi:10.1016/j.ijpe.2010.11.024
- Gartner. (2009). *Predicts 2009: Business Intelligence and Performance Management Will Deliver Greater Business Value*. Retrieved August 30, 2015, from http://www.gartner.com/DisplayDocument?ref=g_search&id=842713&subref=simplesearch
- Gopal, P. R. C., & Thakkar, J. (2012). A review on supply chain performance measures and metrics: 20002011. *International Journal of Productivity and Performance Management*, 61(5), 518–547. doi:10.1108/17410401211232957
- Gunasekaran, A., Papadopoulos, T., Dubey, R., Wamba, S., Childe, S. J., Hazen, B., & Akter, S. (2016). Big data and predictive analytics for supply chain and organizational performance. *Journal of Business Research*. doi:10.1016/j.jbusres.2016.08.004
- Gunasekaran, A., Patel, C., & McGaughey, R. E. (2004). A framework for supply chain performance measurement. *International Journal of Production Economics*, 87(3), 333–347. doi:10.1016/j.ijpe.2003.08.003

- Haydock, P. M. (2003). Supply Chain Intelligence. *ASCET*, 5, 15–21.
- Lappide, L. (2010). Predictive Metrics. *The Journal of Business Forecasting*, 29(2), 1–23.
- Lim, P., Chen, H., & Chen, G. (2013). Business Intelligence and Analytics: Research Directions. *ACM Trans. Manage. Inf. Syst.*, 3(4), Article 17. doi:10.1145/2407740.2407741
- Parenteau, J., Sallam, R. L., Howson, C., Tapadinhas, J., Schlegel, K., & Oestreich, T. W. (2016). *Magic Quadrant for Business Intelligence and Analytics Platforms*. Gartner, Inc. Retrieved February 30, 2016, from <https://www.gartner.com/doc/reprints?id=1-2XXET8P&ct=160204>
- Partridge, A. R. (2013). *Business Intelligence in the Supply Chain*. Retrieved November 5, 2015, from <http://www.inboundlogistics.com/cms/article/business-intelligence-in-the-supply-chain/>
- Reddy, R. (2004). *Supply Chain Intelligence - Know what to expect and how best to achieve it*. Retrieved March 25, 2015, from http://www.intelligententerprise.com/030513/608infosc1_1.jhtml
- Russom, P. (2010). Increasing Manufacturing Performance Through Supply Chain Intelligence. *DM Review Magazine*. Retrieved January 5, 2016, from <http://www.dmreview.com/toc.cfm?issueid=1033>
- Schoenherr, T., & Speier-Pero, C. (2015). Data Science, Predictive Analytics, and Big Data in Supply Chain Management: Current State and Future Potential. *J Bus Logist*, 36(1), 120–132. doi:10.1111/jbl.12082
- Shepherd, C., & Gunter, H. (2006). Measuring supply chain performance: Current research and future directions. *International Journal of Productivity and Performance Management*, 55(3–4), 242–258. doi:10.1108/17410400610653219
- Stefanovic, D., & Stefanovic, N. (2008). Methodology for Modelling and Analysis of Supply Networks. *Journal of Intelligent Manufacturing*, 19(4), 485–503. doi:10.1007/s10845-008-0098-0
- Stefanovic, N. (2014). Proactive Supply Chain Performance Management with Predictive Analytics. *The Scientific World Journal*, Article ID 528917. doi:10.1155/2014/528917
- Stefanovic, N. (2015). Collaborative predictive business intelligence model for spare parts inventory replenishment. *Computer Science and Information Systems*, 12(3), 911–930. doi:10.2298/CSIS141101034S
- Stefanovic, N., Radenkovic, B., & Stefanovic, D. (2007). Supply Chain Intelligence. In D. T. Pham, E. E. Eldukhri, & A. J. Soroka (Eds.), *Intelligent Production Machines and Systems*. Whittles Publishing.
- Stefanovic, N., & Stefanovic, D. (2009). Supply Chain Business Intelligence – Technologies, Issues and Trends. In M. Bramer (Ed.), *Artificial Intelligence: An International Perspective, LNAI 5640* (pp. 217–245). Berlin: Springer. doi:10.1007/978-3-642-03226-4_12
- Stefanovic, N., & Stefanovic, D. (2011). Supply Chain Performance Measurement System Based on Scorecards and Web Portals. *Computer Science and Information Systems*, 8(1), 167–192. doi:10.2298/CSIS090608018S
- Stefanovic, N., Stefanovic, D., & Radenkovic, B. (2011). Integrated Supply Chain Intelligence through Collaborative Planning, Analytics and Monitoring. In S. Mohhebi, I. Mahdavi, & N. Cho (Eds.), *Electronic Supply Network Coordination in Intelligent and Dynamic Environment: Modeling and Implementation* (pp. 43–92). Hershey, PA: IGI Global. doi:10.4018/978-1-60566-808-6.ch003
- Taylor, J. (2007). *Business Intelligence 2.0 and Enterprise Decision Management*. Retrieved January 8, 2016, from <http://edmblog.com>
- Trkman, P., McCormack, K., de Oliveira, M. P. V., & Ladeira, M. B. (2010). The impact of business analytics on supply chain performance. *Decision Support Systems*, 49(3), 318–327. doi:10.1016/j.dss.2010.03.007

ADDITIONAL READING

Cecere, L.M. (2014). *Supply Chain Metrics that Matter*. John Wiley & Sons.

Linoff, G. S., & Berry, M. J. A. (2011). *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management* (3rd ed.). John Wiley & Sons, Inc.

Wang, J. (2014). *Encyclopedia of Business Analytics and Optimization*. IGI Global. doi:10.4018/978-1-4666-5202-6

KEY TERMS AND DEFINITIONS

BI Portal: A specialized web portal that provide a range of analytical services and which enables collaborative decision-making.

Business Intelligence: A set of processes, technologies and tools comprising data warehousing, On-Line Analytical Processing, and information delivery in order to turn data into information and information into knowledge.

Data Mining: Set of knowledge discovery techniques for intelligent data analysis in order to find hidden patterns and associations, devise rules and make predictions.

Data Warehouse: An integrated data repository system for extracting, loading and transforming data into information.

Performance Monitoring: Methodology to optimize the execution of business strategy through the real-time and proactive performance measurement and management of the supply network that consists of a set of integrated, closed-loop technologies and tools such as scorecards, key performance indicators, and business activity monitoring.

Supply Chain Intelligence: Process of integrating and presenting supply chain information in order to provide collaborative planning, monitoring, measurement, analysis and management of the supply network.

Supply Chain: Dynamic, interconnected and collaborative group of companies working jointly on planning, management and execution of cross-company business processes spanning from the first tier suppliers to the end-customers.

A Review of Supply Chain Risk Management in Agribusiness Industry

Sri Widiyanesti

Telkom University, Indonesia

Yudi Fernando

Universiti Malaysia Pahang, Malaysia

INTRODUCTION

There will always be risk in business. Nowadays, to be able to run business and succeed, companies should be able to identify risk factors and minimize the risks involved. The biggest challenge now is not only identifying and minimizing risk factor in the company's internal, but the risk factors in the company's supply chain network, in both upstream and downstream activity. Companies that are not able to handle the risks that arise in their business will inevitably suffer losses.

Facing increasing global competition and the fact that supply chains become more complex, there is a possibility that increasing the performance of supply chains will be difficult to reach, mainly because of the failure risk of supply chains (Tumala and Schoenherr 2011).

Companies that will able to survive is those resilient enough in managing risks. Risks that would be faced by companies are very diverse, depending on the company's business type. For example, the iron industries in India have risks and problems such as; short life cycle and season ability, supplier inefficiency, design and technological changes, union strike, fluctuations of demand, lack of ownership, capacity and information, inventories, corporate strategy and goal, implications, and customer uncertainty (Pradhan and Srikanta, 2014).

Indonesia is the largest archipelagic country in the world and agricultural lands are widely available in Indonesia. Furthermore, most of

Indonesians are working in agricultural sector. Agricultural sector is a critical industry to support daily life and food consumptions of Indonesian population. With such conditions and demands, Indonesia should be the world's agricultural source, but, ironically, it is a frequent occurrence that Indonesian depends on imported agricultural products including nine basic commodities rather than produce its own demands for food source.

Supply chain of agricultural commodities become a major concern of Indonesian scholars and practitioners. Many companies engaged in this industry unable to survive in the long term. This is an unfortunate condition because Indonesia has abundant natural resources and plenty of commodity producers.

The development of industry and services sectors have a significant impact on the amount of agricultural land that is converted industrial and residential areas (Trisnasari et al., 2010). Indonesia is facing a decline in agricultural land area of about 2.65% (Pertanian.go.id, 2014). Agricultural land in Indonesia are switching functions, fields and rice paddies are being converted to housing and offices. This decline in agricultural area contributed to the decline of food production, which led to Indonesia importing agricultural products.

Aside from the reduced agricultural production, agriculture is no longer a major contributor to Indonesia GDP. In 2012 to 2014, the percentage of agricultural GDP decreased if compared with trade and processing industry (BPS, 2015). Despite being forecasted that in 2017 Indonesia

will become self-sufficient in food production (Gera, 2015), it cannot be achieved if agricultural production has declined.

Indonesia has abundant natural resources adequate for agriculture, but unfortunately agriculture cannot be a source of revenue for Indonesia. The government must begin to focus on agriculture, because Indonesia has the resources, community livelihood, and agricultural fields. Besides agriculture is the primary food source for Indonesia, and a major needs worldwide.

Indonesia archipelago is unique compared to other agribusiness countries, such as Thailand. The problem faced by Indonesia is more complex, especially problems in the distribution and transportation sector. Besides Indonesia is experiencing other problems that hinder the development of agribusiness in Indonesia. With the current condition of Indonesia, farmers face a variety of business risks. Companies that will able to survive is those which resilient in managing risks. Risks that would be faced by companies are very diverse, depending on the company's business type.

The literature review was a synthesized and limited articles published on supply chain risk management in agribusiness industry. It mostly has been discussed regarding supply chain risk management in manufacturing industry. The lack of studies to investigate the risk factors in agribusiness will make this study a contribution to the supply chain literature. This is because the managing risk in supply chain in agribusiness industry could be different compare to manufacturing industry.

BACKGROUND

Supply chain management is an overall coordination of supply chain activities including the improvement of customer value (Heizer and Render, 2015).

Supply chain management will improve companies' performance. In supply chain management, companies will get challenges from risks that pos-

sibly would be faced by the company. Companies must be able to know the risks, therefore, could interfere in business resilience.

Risks that arise in the supply chain can negatively impact supply chain performance. Companies must have a strategy to manage the risks, so that effectiveness can be achieved (Giannakis and louis, 2010). This can be evidenced by an example of a toy company that faced lawsuits and must withdraw the toy because of the supply chain failure due to a risk emergence in the supply chain (Story, 2007; Tumala & Schoenherr, 2011). For a manufacturing company producing pencils, supply chain risks that may arise are demand, delay, disruption, inventory, manufacturing, capacity, supply, system, sovereign, and transportation problem (Tumala and Schoenherr, 2011).

According to Thun and Hoenig (2009), supply chain risk is divided into external and internal risks. Internal supply chain risks consist of supplier failure, malfunction of IT systems, supplier quality problems, change in customer demand, machine breakdowns, increasing raw material prices, transportation failure, disruption of the delivery chain, and technological change. While the external supply chain risks, for example, in manufacturing firms in German are accident, strike, war, natural disaster, import restriction, oil crisis, terrorist attacks, and increasing customs duty.

The definition of demand risk in supply chain is the order fulfilment errors, inaccurate forecasts due to longer lead times, product variety, swings demand, seasonality, short life cycle, and small customer based on information distortion due to sales promotions and incentives, lack of supply chain visibility, and exaggeration of demand during product shortage (Tumala and Schoenherr, 2011). While other researchers assert that demand fluctuations and customer demand uncertainty is also included in the demand risk in supply chain (Pradhan and Srikanta 2014).

Delay risks occur due to excessive handling in border crossing or change in transportation mode, port capacity and congestion, custom clearances

at ports, and transportation breakdowns (Tumala and Shoenherr, 2011).

Disruption risks include natural disaster, terrorism and wars, labor disputes, single source of supply, capacity and responsiveness of alternate supplier (Tumala and Shoenherr, 2011). While inventory risks include a cost of holding inventories, demand and supply uncertainty, rate of product obsolescence, and supplier fulfilment (Tumala and Shoenherr, 2011). Pradhan and Routray (2014) mentions several categories of risk that may occur in the supply chain in the iron industry in India which is short life cycle, capacity and information, season ability and, inventory.

Manufacturing (process) breakdown risks include poor quality (ANSI or other compliance standards), lower process yields, higher product cost, and design changes (Tumala and Shoenherr, 2011).

Physical plant (capacity) risks include the lack of capacity flexibility and capacity cost (Tumala and Shoenherr, 2011).

Supply (procurement) risks include delivery performance, supplier fulfilment errors, selection of wrong partners, high capacity utilization supply source, inflexibility of supply source, poor quality or process yield at supply source, supplier bankruptcy, rate of exchange, percentage of a key component or raw material procured from a single source (Tumala and Shoenherr, 2011). Pradhan and Srikanta (2014) mentions a number of categories of supply risks that may occur in the supply chain in the iron industry in India which include supplier inefficiency and early supplier involvement. Also the bigger risks in supply companies are characteristics of supplier and supply disruption.

System risks include information infrastructure breakdowns, lack of effective system integration or extensive system networking, and lack of compatibility in IT platforms among supply chain partners (Tumala and Shoenherr, 2011). Another study states that design and technological changes are included in system risks (Pradhan and Routroy, 2014).

Sovereign risks include regional instability, communication difficulties, government regulations, loss of control, and intellectual property breaches (Tumala and Shoenherr, 2011). Meanwhile according to Pradhan and Srikanta (2014), included in sovereign risks are corporate strategy/goal and union strikes.

Transportation risk include paperwork and scheduling, port strikes, delay at ports due to port capacity, late deliveries, higher costs of transportation that depends on transportation mode chosen. (Tumala and Shoenherr, 2011).

There are three classifications of risks in supply chains, the first is the risk in information security. Information security risks can be defined as the risks associated with sharing information with external agents. The second is a supplier opportunism which means suppliers act in their own interest by deliberately withholding or distorting information. The third risk is corporate social responsibilities, which occurs when suppliers do not pay attention to social and environmental demands (Spekman and Davis, 2004; Ellegaard, 2008).

According to David L. Olson and Wu Desheng (2011) there are several categories of risk that may occur in the supply chain of China's outsourcing strategy which are accounting, asset impairment risks, country risks, competitive risks, customer risks, downside risks, financial risks, interaction, legal risks, product risks, regulatory risks, reputation risks, shared risks, and supplier risks.

According to Telman and McCormack (2009) companies' risks are divided into two categories, namely continuous risk and discrete risk. Continuous risk is a potential event that raises cost continuously and relatively easy to predict. Discrete risks is a high-impact events which can be classified as terrorism, the spread of diseases, and natural disaster (Faisal et al, 2006. Telman and McCormack, 2009). Discrete risks are hard to be predicted and measure (Telman and McCormack, 2009).

According to the layout and Revoredo (2013), the risks of supply chains in the livestock industry

are production risks, market risks, and institutional risks.

According to Punniyamoorthy, Thamaraiselvan, and Manikandan (2013) some of the risks of supply chain in the steel mining industry are supply side risks, manufacturing side risks, demand side risks, logistics side risks, information risk and environment risks.

According Tang (2006) supply chain risks are divided into two; operational risks and disruption risks. The operational risks include uncertain customer demand, uncertain supply, and the uncertain costs. While disruption risk is a major disruption caused by both natural and man-made disasters, such as earthquakes, floods, hurricanes, terrorist attack, and economic crisis.

According to Aqlan and Lam (2014), there are nine categories of risks: supplier risk, customer risk, process and control risk, technology risk, product risk, occupational risk, culture risk, transportation risk, and commodity risk.

There are plenty of research on supply chain risk management already. Most of those studies are targeting the manufacturing industry, while we rarely see a study on this matter which is targeting the agribusiness industry. The lack of studies to investigate the risk factors in agribusiness makes this study contribute to enrich the literature on supply chain risk management.

MAIN FOCUS OF THE ARTICLE

Issues, Controversies, Problems

There are problems in the agricultural supply chain which is appeared to be the impact of the condition of commodities in Indonesia. Some of the problems are frequent fluctuation in commodity prices, commodity shortages, rises in seed or feed prices, unbalanced supply and demand, expensive distribution and production costs, the tendency to import raw material commodities, and natural disasters.

Commodity prices in Indonesia are frequently going up and down. It is not uncommon for commodity prices to change—not on a yearly or a semester basis—but on a daily basis. The price fluctuations are tough to manage, even by the Indonesian government. For example, the price of tomatoes initially valued at IDR 9,000 per kilogram; then it suddenly decreased by 44.44% reaching the price of IDR 4,000 per kilogram. For curly red chili, the prices dropped from IDR 80,000 per kilogram to IDR 25,000 per kilogram. As the result, farmers who have invested a tremendous amount of funds suffered a massive loss because of this situation (Wiyoso, 2014). At first, when the farmers have just started planting, the commodity prices tend to be stable. However, when the harvest time comes, the prices dropped significantly. The absence of planting date management in turns creates an unbalanced demand and supply in the market.

As mentioned before, Indonesia has an unbalanced supply and demand on its commodities. For example, one area supply for certain commodities that they produce exceeds demand which makes the price of that commodity, in that particular location, very cheap. Meanwhile, in another area supply fall short on demand, but the price of the commodity is very high. For example, the price of potatoes are variable in every city in Indonesia (BPS, 2015). In cities located in Java region such as Bandung, Cilacap, and Malang, the price of potato is under IDR 10,000. Meanwhile, outside Java the prices tend to be higher than IDR 10,000. These differences were the result of the expensive distribution cost in Indonesia.

There are many factors contributed to the expensive distribution costs in Indonesia, one of them is the transportation fee. The distribution cost in Indonesia is very expensive, even shipping a package from Jakarta to Hamburg (11,000 km) is cheaper than shipping from Jakarta to Padang (1,000 km). Distributing products to areas outside Java cannot be done directly between cities due to inadequate local infrastructures like port and

airports. It means that before the goods are shipped to the destination, it has to be sent to Jakarta first. In addition to problems of coordination between the central government and local governments. In Southeast Asia, Indonesia is one of the countries with the highest transportation costs. Thailand's logistic costs lower by 7 percent of the GDP, while Indonesia's logistic cost is 27 percent of GDP (Suara Cargo, 2015).

In addition to the high distribution costs, agribusiness in Indonesia is also influenced by the cost of production. High production costs are the result of the expensive fertilizer. The government tries to solve this problem by giving subsidies for fertilizers. Shortages of fertilizer are often the case in Indonesia; farmers can obtain the fertilizer, but the price is very high. The fertilizer that cost about IDR 90,000 per kg is sold to the farmers at the price of IDR 125,000 per kg. Apparently, many middlemen pile on the subsidized fertilizer that has been given by the government which forces the farmers to buy the fertilizer with a high price (Mashita, 2015).

The middlemen not only interferes in the distribution of raw material (fertilizer), but also in the distribution of the finished products. It also happens to onion farmers, who complained about the agents who always ask for a commission from the sale of one hundred pounds of onions. Fees paid to the agent is approximately 15% of the total sales (Rofiq, 2015). Such incident increases the selling price of onions in the market and lowers the profit for the farmers because of the high production costs. The middlemen also meddle in the rice industry. Middlemen are reportedly buying rice from farmers and sell it out of town for a higher price (Widiyanto, 2015). This phenomenon triggers the staple crops shortage, even in areas with a granary.

Unfortunately, the government fails to control the middlemen. They had a secure network and even protected by corrupt government officials. Besides the middleman, scarcity of commodities is also caused by the business owner. In most cases, they are reluctant to sell their products to

the market so that they can get a higher selling price. This incident not only occurred in agriculture industry but also in other nine basic commodities industry. For example in beef, the government predicts that there are an adequate amount of cattle at the local farm. Therefore, in the third quarter, the government decided to lower the amount of imported meat. Then, the government is asking the local cattle farmer to sell the cow to the market. However, the farmers are reluctant to sell the cow because usually the policy is issued right before the Eid al-Adha. As the result, some farmers chose to temporarily stop selling his cow in order to get a higher price. The beef meat prices in the market usually cost about IDR 80,000 / kg, but the delay on the sale causes the price to escalate to IDR 130,000 per kg. An increase in beef prices occurred almost throughout Indonesia, including in the area which is the center of livestock such as Madura, Bali, and NTT (Zamzami, 2015).

Increased prices of nine basic commodities in Indonesia happens because of the tendency to import the raw material. One of the examples is soybeans, which is the raw material of tofu. The fluctuation in the value of Indonesian Rupiah against the US Dollar causes the price of these commodities fluctuates. The exchange rate of the Indonesian Rupiah in the last four years is very volatile. In 2011, the exchange rate of the Indonesian Rupiah was at the level of IDR 8,958 per US Dollar, and in 2015 it was at IDR 14,467 per US Dollar. It appears that from 2011 to 2015 the exchange rate of the Indonesian Rupiah against the US dollar increased by 64.55. It eventually leads to a significant increase in the price of imported raw materials and the selling price of commodities like tofu (Aimon and Satrianto, 2014).

Natural disasters are also the cause of shortages in nine basic commodities. For instance, the eruption of Mount Sinabung in North Sumatra led to the scarcity of vegetables in the year of 2013. Supplies of whole vegetables in the nation reportedly decreased from 30% to 50%. The eruption of Mount Sinabung inevitably halted the farmers' activity because they have to evacuate to shelters.

As the result, farmers temporarily stop growing vegetables which lead to a vegetable shortage in Indonesia (Medan Bisnis, 2015).

In addition to volcanic eruptions, floods may also cause a shortage in commodities. Because of the flood, the farmers often have to face a devastating crop failure. For example, in the year of 2014, floods soak the rice fields in the area of Cirebon, West Java. The flood in this area was caused by high waves in the ocean located in the north of Java (Rachman, 2014). On February 2015, Banten suffered a flood that drowned the rice fields in the area, and the floodwaters itself does not recede within a week. Approximately 3000 hectares of land were flooded, and this may interfere with the availability of food in Banten (Deslatama, 2015). Besides that, flood often occurs in the capital city of Indonesia, Jakarta. The disaster left the economic and trading activities in the country paralyzed. If Jakarta floods, it will affect the distribution of the commodity in Indonesia, because the distribution is always centered in Jakarta. The losses suffer because of the floods that hit Jakarta is approximately 30 percent of the total distribution losses for the whole of Indonesia (Syahputri, 2013).

SOLUTIONS AND RECOMMENDATIONS

The problems exist in agribusiness industry in Indonesia are very complex. That explains why there are plenty of companies engaged in this industry that failed to survive. The fact is contradictive with the natural state of Indonesia with its abundant natural resources and plenty of commodities producer. These problems occur because companies are not ready to face the risks which exist in the agribusiness industry.

In Indonesia's agribusiness industry, the price is unstable. The price of commodities constantly fluctuates, and the price also varies from each region in Indonesia. Companies should be able to

deal with this problem. This problem may cause the company to suffer so many losses since the price may be lower than the cost of production. The imbalance between demand and supply can fluctuate price, especially in areas with commodity shortage. This is what we called as demand and supply risk. The lack of inventory may also cause scarcity of commodities in a region; this is because some of the commodities that are limited to particular areas. Agribusiness companies can avoid this if they can manage inventory risk properly.

Supply risk also occurs in the scarcity of manure as a raw material for the process of planting crops. Beside that there is also a shortage of the raw materials (e.g. the scarcity of soybeans as one of the raw materials to produce tempeh and tofu).

Shortages and prices fluctuation in Indonesia happen because of the middlemen who meddle in the distribution process. Apparently, the government finds it difficult to deal with these agents, even though they significantly affect farmers and the market. If the government wants a more stable business, then the government should create a rule that may minimize the sovereign risks. The emergence of powerful middlemen in agribusiness, which is the result of the poorly-managed transactions system, may pose a system risk to agribusiness companies.

The scarcity of commodities is also caused by natural disasters such as earthquakes, floods, and volcanic eruptions. These natural disasters may cause disruption risk in the agribusiness industry. The natural disaster itself is the hardest risk to predict and to avoid. Floods occurred in the capital city of Jakarta often causes delays in delivery and caused shortages in some areas. Risk arising from the flood is called delay risk.

The increased price of commodities in Indonesia is also caused by the tendency to import some of the raw materials from foreign countries. The unstable exchange rate in the transaction may give rise to procurement risk in agribusiness.

High prices of commodities in Indonesia are also the result of high distribution costs in

Indonesia and the delivery system in Indonesia. Commodities sent from an area outside the Java region must go through the capital city because the regions outside of Java do not have adequate port and airport. In this case, the risk faced by the agribusiness company is the transportation risk. Transportation risk, if cannot be anticipated, will lead to increased production costs. This centralizes distribution system makes it longer and harder for companies to receive the commodities. Therefore, it would bring delay risk to the agribusiness companies.

Based on interviews with various companies engage in agribusiness, it appears that there are several risk factors that they often face, which are demand risks, supply risk, delay risks, disruption risks, inventory risks, procurement risks, system risks, sovereign risks, and transportation risks.

SUGGESTION FOR FUTURE RESEARCH

Further research is suggested to discuss how supply chain risk management manages to minimize the risk factors of agribusiness in Indonesia. It can also discuss the dominant factor in Indonesia's agribusiness industry and what steps should be taken by the government to develop agribusiness in Indonesia. In addition to this, it is also recommended to study another type of supply chain risk management which will suit other countries, since Indonesia has unique characteristics compared to other nations in term of agribusiness.

CONCLUSION

This paper concludes that there are nine risk factors occur in the agribusiness industry in Indonesia, which are demand risk, supply risk, delay risk, disruption risk, inventory risk, procurement risk, system risk, sovereign risk, and transportation risk. To survive in this industry, companies which

engage in agribusiness should be able to identify risk factors that may occur in the industry and learn how to deal with them. Therefore, companies need to apply the supply chain risk management to minimize those risks.

REFERENCES

- Aimon, H., & Satrianto, A. (2014). Prospek Konsumsi dan Impor Kedelai di Indonesia tahun 2015 - 2020. *Jurnal Kajian Ekonomi*, 5(3).
- Aqlan, F., & Lam, S. S. (2014). A Fuzzy-based integrated framework for supply chain risk assessment. *International Journal of Production Economics*, 161, 54–63. doi:10.1016/j.ijpe.2014.11.013
- Cargo, S. (2015). *Mahalnya ongkos kirim akibat permasalahan infrastruktur*. Retrieved 20 October 2015 from <http://suaracargo.com/2015/03/28/mahalnya-ongkos-kirim-akibat-permasalahan-infrastruktur/>
- Delastama, Y. (2015). *Ribuan hektar sawah di Banten terancam gagal panen akibat banjir*. Retrieved 29 February 2016 from <http://bisnis.liputan6.com/read/2175091/ribuan-hektar-sawah-di-banten-terancam-gagal-panen-akibat-banjir>
- Ellegaard, T. (2008). Supply risk management in a small company perspective. *Supply Chain Management: An International Journal*, 13(6), 425–434. doi:10.1108/13598540810905688
- Faisal, M. N., & Shankar, B. R. (2006). Mapping supply chains on risk and customer sensitivity dimensions. *Industrial Management & Data Systems*, 106(6), 878–895. doi:10.1108/02635570610671533
- Gera, I. (2015). *Indonesia Swasembada Pangan Penuh 2017*. Retrieved 20 October 2015 from <http://www.voaindonesia.com/content/mentan-indonesia-swasembada-pangan-penuh-2017/2705270.html>

- Giannakis, M., & Louis, M. (2010). A Multi-agent based framework for supply chain risk management. *Journal of Purchasing and Supply Management*, 17(1), 23–31. doi:10.1016/j.pur-sup.2010.05.001
- Heizer, J., & Render, B. (2015). *Manajemen Operasi: Manajemen Kebelangsungan dan Rantai Pasokan, edisi 11*. Jakarta: Salemba Empat.
- Lambert, D. M. (2008). Supply Chain. *Management Review*, 8(6).
- Lockami, I. I. I. P., & McCormack, K. (2012). Modeling supplier risks using Bayesian networks. *Industrial Management & Data Systems*, 112(2), 313–333. doi:10.1108/02635571211204317
- Mashita, N. (2015). *Musim tanam petani menjerit karena harga pupuk melambung*. Retrieved from <http://www.lensaIndonesia.com/2015/01/30/musim-tanam-petani-menjerit-karena-harga-pupuk-melambung.html>
- Medan bisns. (2013). *Sayuran ditanah karo mulai langka*. Retrieved 20 October 2015 from http://www.medanbisnisdaily.com/news/read/2013/09/17/51107/sayuran-dari-karo-mulai-langka/#.Vg_kMs-hfIU
- Olson, D. L., & Wu, D. D. (2010). A review of enterprise risk management in supply chain. *Kybernet*, 39(5), 694–706. doi:10.1108/03684921011043198
- Pradhan, K., & Srikanta, S. R. (2014). Analyzing the supply chain risk issues for an Indian manufacturing company. *Journal of Advances in Management Research*, 11(2), 144–162.
- Punniyamoorthy, A., Thamaraiselvan, N., & Manikandan, L. (2013). Assesment of supply chain risk: scale development and validation. *Benchmarking: An International Journal*, 20(1), 79–105.
- Rachman, T. (2014). *Sawah kena banjir, petani Cirebon terancam gagal panen*. Retrieved 29 February 2016 from <http://nasional.republika.co.id/berita/nasional/jawa-barat-nasional/14/01/17/mziulr-sawah-kena-banjir-petani-cirebon-terancam-gagal-panen>
- Ridha, N. (2015). *Penduduk Indonesia dan Mata Pencahariannya*. Retrieved from https://www.academia.edu/9936388/Penduduk_Indonesia_dan_Mata_Pencahariannya
- Rofiq, M. (2015). *Temuan Mentan soal permainan tengkulak ke petani bawang*. Retrieved 20 October 2015 from <http://finance.detik.com/read/2015/06/24/124856/2950898/4/ini-temuan-mentan-soal-permainan-tengkulak-ke-petani-bawang>
- Syahputri, E. (2013). *Bisnis ritel rugi Rp. 460 miliar akibat banjir*. Retrieved 29 February 2016 from <http://www.antaraneews.com/berita/353540/bisnis-ritel-rugi-rp460-miliar-akibat-banjir>
- Tang, C. S. (2006). Perspective in supply chain risk management. *International Journal of Production Economics*, 103(2), 451–488. doi:10.1016/j.ijpe.2005.12.006
- Thun, J. H., & Hoenig, D. (2009). An empirical analysis of supply chain risk management in the german automotive industry. *International Journal of Production Economics*, 131(1), 242–249. doi:10.1016/j.ijpe.2009.10.010
- Trianasari, D., Sebayang, A. F., & Haryatiningsih, R. (2015). Keputusan Rumah Tangga Petani dalam Alih Fungsi Lahan Pertanian di Desa Bumi Wangi Kecamatan Ciparay Kabupaten Bandung. *Prosiding Ilmu Ekonomi*, 1(2), 25–33.
- Trkman, P., & McCormack, K. (2009). Supply chain risk in turbulent environments – A conceptual model for managing supply chain network list. *International Journal of Production Economics*, 119(2), 247–258. doi:10.1016/j.ijpe.2009.03.002

Tummala, R., & Schoenherr, T. (2011). Assessing and managing risks using the Supply Chain Risk Management Process (SCRMP). *Supply Chain Management: An International Journal*, 16(6), 474–483. doi:10.1108/13598541111171165

Tuncel, G., & Alpan, G. (2009). Risk assessment and management for supply chain networks: A case study. *Computers in Industry*, 61(3), 250–259. doi:10.1016/j.compind.2009.09.008

Widiyanto, D. (2015). *Beras Melambung Permainan Tengkulak*. Retrieved 20 October 2015 from <http://krjogja.com/read/249983/beras-melambung-permainan-tengkulak.kr>

Wiyoso, N. (2015). *Price Tomato and Chilli Red Precisely Down*. Retrieved 20 October 2015 from <http://disperindag.jabarprov.go.id/news/detail/eksternal/2015/01/23:10:54508c75c8507a2ae5223dfd2faeb98122>

Zamzami, F. (2015). *The increase in price of beef is not fair*. Retrieved 20 October 2015 from <http://www.republika.co.id/berita/koran/publik/15/08/08/nsrere1-kenaikan-harga-daging-sapi-tidak-wajar>

ADDITIONAL READING

Jin, Y., & Turvey, C. G. (2002). Hedging financial and business risks in agriculture with commodity-linked loans. *Agricultural Finance Review*, 62(1), 41–57. doi:10.1108/00214870280001128

Leat, P., & Revoredo-Giha, C. (2013). Risk and resilience in agri-food supply chains: The case of the ASDA PorkLink supply chain in Scotland. *Supply Chain Management: An International Journal*, 18(2), 219–231. doi:10.1108/13598541311318845

Maurer, K. (2014). Where is the risk? Is agricultural banking really more difficult than other sectors? In *Finance for Food* (pp. 139–165). Springer Berlin Heidelberg. doi:10.1007/978-3-642-54034-9_7

Rao, S., & Goldsby, T. J. (2009). Supply chain risks: A review and typology. *The International Journal of Logistics Management*, 20(1), 97–123. doi:10.1108/09574090910954864

Tang, O., & Musa, S. N. (2011). Identifying risk issues and research advancements in supply chain risk management. *International Journal of Production Economics*, 133(1), 25–34. doi:10.1016/j.ijpe.2010.06.013

Thun, J. H., & Hoenig, D. (2011). An empirical analysis of supply chain risk management in the German automotive industry. *International Journal of Production Economics*, 131(1), 242–249. doi:10.1016/j.ijpe.2009.10.010

KEY TERMS AND DEFINITIONS

Demand and Supply Risk: Demand risk is result from disruption emerging from downstream supply chain operator an supply risk is result from disruption emerging from upstream supply chain.

Disruption Risk: Disruption risk is risk which arise from natural disaster, such as weather disruption, or man made ones such as economic crises.

Sovereign Risk: Sovereign risk is risk which come from Regional instability, Communication difficulties, Government regulations, Loss of control, and Intellectual property breaches.

Transportation Risk: Transportation risk is risk which come from Port strikes, Delay at ports due to port capacity, Late deliveries, Higher costs of transportation, and Depends on transportation mode chosen.

The Role of Emerging Information Technologies for Supporting Supply Chain Management



Zlatko Nedelko

University of Maribor, Slovenia

Vojko Potocan

University of Maribor, Slovenia

INTRODUCTION

Over decades of rapid development of information technology and especially internet it has become apparent that information technology has been significantly influenced all areas of our life and business (Ayeh, 2008; Nedelko & Cirnu, 2008; Stucky & Weiss, 2008). In terms of its impact on business sphere an important viewpoint is related to the supply chains, which has become the center of attention in business literature. This is due to the globalization and the fact that each organization belongs at least to one, in business practice often more, supply chains or networks (Cooper et al., 1997; Beamon, 1998; Lummus & Vokurka, 1999; Christopher, 2011).

A basic definition defines supply chain as integrity of all parties involved, directly or indirectly, in fulfilling a customer's needs. The supply chain includes suppliers, manufacturers, customers, distributors, warehouses, retailers (Chopra & Meindl, 2013, p. 13). In short, supply chain encompasses all entities that are involved in flow of raw materials, through production to the end customer. An important aspect of supply chain management is the exchange of information among partners in the supply chain, which important influence success and agility of the supply chain.

In terms of the exchange of information and especially collaboration among entities in the supply chain, a plethora of solutions, based on modern information technology, are aimed to

enhance the working of entities in the supply chain and enable better exchange of information, goods, services, and collaboration across different entities in the supply chain. Different information technology solutions play an important role in supporting supply chain management in today's organizations and its supply chains. From the viewpoint of solutions, that are based on information technology, at the pedestal are solutions that support customer relationship management (Chen & Popovich, 2003; Tamosiuniene & Jasilioniene, 2007; Kumar, 2010). Besides, that in a contemporary environment, an important viewpoint for enhancing supply chain management, relates to the utilization of emerging solutions, which can be used in order to improve supply chain management in the future.

The main purpose of this chapter is to examine the role, importance and actual usage of emerging information technology solutions that support and have potential to enhance supply chain management in practice (Chen & Popovich, 2003; Rigby & Bilodeau, 2011). The role of information technology in supply chain management is in this paper discussed through the lenses of several selected emerging management tools, like collaborative innovations, corporate blogs, radio-frequency identification, and shared services centers. For a more comprehensive discussion, results for several selected management tools utilization are reported for selected worldwide regions, like North America, Latin America, Asia Pacific, Europe and

DOI: 10.4018/978-1-5225-2255-3.ch483

example two emerging economies from Central and East Europe.

BACKGROUND

There are many possible ways to examine the role of information technology in supply chains (Beamon, 1998; Simchi-Levi et al., 2009; Christopher, 2011; Chopra & Meindl, 2013). In this paper, we adopted an approach that is based on the typical development phases of the supply chain. In that framework we outline typical development phases and the role of information technology in each of these phases. Next, in line with the aims of our paper, we put our focus on examination of the role, importance and actual usage of information technology solutions that support supply chain management and customer relationship management in organizations worldwide.

Supply Chain

Supply chain presents ambitious and strategically significant concept, which can be defined as “managing the entire chain of raw material supply, manufacture, assembly and distribution to the end customer” (Heitzer & Render, 2003; Murphy & Wood, 2004; Christopher, 2011). The supply chain is the most developed integrated concept, but by its use, the organization meets some open dilemmas such as: 1) what sort of connections exist among the part of supply chain, 2) what is the role (meaning) of different units (e.g. parts) in the entire supply chain, and 3) how can we optimize the parts of the entirety (to form structure) to reach “optimal results” of common work.

Due to the crucial importance of cooperation and especially information exchange among entities involved in supply chain, information technology plays a crucial role since it enhances or facilitates exchange of information among entities in the supply chain. In order to emphasize the role of information technology in supply chain, we first outline basic integrated concepts of supply chain

that will serve us for a presentation of the steadily increasing role of information technology, through different phases of supply chain development.

There are many different ways in which the linkage involved in the flow of materials and services can be integrated or grouped together (Rushton et al., 2001; Heitzer & Render, 2003; Slack et al., 2006). Four main concepts will be presented here. These have focused attention on managing across the traditional functional areas of purchasing operation and physical distribution. They are material management, merchandising and logistics and supply chain management (SCM) (Cohen & Roussel, 2004; Blanchard, 2006; Daft, 2007).

During the last twenty years or even broader, more ambitious and strategically significant concept has emerged - SCM. SCM is a broader and a strategically more significant concept which includes the entire SC from the supply of raw material, through manufacture, assembly and distribution to the end customer.

The Role of Information Technology in Supply Chains

In last decade information technology has had significantly influence on business processes throughout the entire supply chain (Gunasekaran & Ngai, 2004; Manthou et al., 2004). A very simple definition of information technology utilized in supply chain defines it as a set of hardware and software that enable entities involved in the supply chain to collect and analyze data and information, that are foundation for their decisions (Chopra & Meindl, 2013, p. 500).

Organizations involved in supply chains use different information technology solutions that are aimed to support either internal processes in organizations or solutions that are designed to support collaboration between entities involved in the entire supply chain (Li et al., 2001; Gunasekaran & Ngai, 2004; Simchi-Levi et al., 2009, p. 407). In the frame of collaboration between the focal organization and other entities in the supply chain, information technology represents

an important support for managing supplier and customer relationships (Chopra & Meindl, 2013, p. 503). Using different information technology solutions, that support collaboration and especially exchange of information between all entities in the supply chain, represent an important success factor. Commonly used information technology solutions are enterprise resource planning, decision support systems, artificial intelligence, management information systems (Gunasekaran & Ngai, 2005; Simchi-Levi et al., 2009, p. 407).

From a view point of supply chain two most important milestones were the introduction of a computer hardware and software and in the mid-1990s internet (Sahin & Robinson, 2002; Murphy & Wood, 2004, p. 9; Chen et al., 2012). According to the different phases of supply chain development, we can conclude, that in first and second phase (i.e., materials management, merchandising, physical distribution) there had been a very basic support of information technology, due to its non-existence or very low level of information technology development in that time.

During the phase of supply chain consolidation (in 1980s) introduction of computers, importantly enhanced management of large amounts of data and information in organizations. Advancement in computer hardware, software and hardware's capacities have allowed single members of the supply chain to make faster, more informed, and more accurate decisions, required for their success (Murphy & Wood, 2004, p. 9). Computers were used for local work in frame of single organizations. In these circumstances, computers represents an important tool for management of the rapidly increasing amount of data and information.

In the phase of functional integration of the supply chain (in 1990s), where there was strong emphasis on linking different organizations (i.e. inter-organizational integration), the internet had played a crucial role in enabling a new way of collaboration between supply chain partners (Li et al., 2001; Christopher, 2011, p. 144). The early to mid-1990s witnessed a growing recognition that there could be value in coordination of the

various business functions not only within single organizations, but across organizations as well (Nigel, 1996; Rushton et al., 2001; Slack et al., 2006; Christopher, 2011). Functional integration was enabled and/or supported largely by the development of the internet, which makes connections and collaboration among geographically dispersed entities in supply chain possible. During the late 1980s and early 1990s, until the development of sophisticated information technology, information technology, especially computers and beginnings of the internet, importantly support collaboration among entities in the supply chain.

Today's, in the frame of supply chain management concept, supply chains are based on well developed information technology. Thus, computers, internet, and network infrastructure represent a backbone of the modern integrated supply chain (Rodrigue, 1998; Manthou et al., 2004; Wang & Chan, 2010). Information technology in today's supply chains, thus has become a prerequisite for success and represents a powerful tool for improving supply chain effectiveness and efficiency.

Usage of computers, coupled with the emergence of the internet and advancement in network equipment, has significantly changed the way of exchanging information and collaboration among entities in the supply chain in the last decade (Murphy & Wood, 2004, p. 41; Chopra & Meindl, 2013). Leading organizations very soon recognized that the key to successful supply chain is usage of different information technology solution that supports their work in frame of the supply chain (Manthou et al., 2004; Subramani, 2004; Christopher, 2011, p. 144).

In terms of supporting processes throughout the entire supply chain, we can find in practice of supply chain plethora of information technology solutions, which support supply chain management. Those solutions, for instance enable better management of customer relationship management and management of supplies, acquiring of knowledge in organizations, etc. (Simchi-Levi et al., 2009). On the other hand, there are also some new emerging solutions, which are based on in-

formation technology and importantly contribute to the enhancement of supply chain management activities, like collaborative innovations, corporate blogs, radio-frequency identification, and shared services centers (Pal & Kapur, 2009; Meidute et al., 2012; Dabic et al., 2013). For the selected emerging solutions in our focus, we provide following short definitions.

Shared service centers reduce costs, since one or several operations are consolidated into shared operation and used by multiple divisions, organizations, users. Based on information technology platforms, centralized services are available to all potential users (Rigby & Bilodeau, 2011). Radio frequency identification (RFID), enable wireless identification of objects and allows tracking the movement of the material flow within the supply chain (Meidute et al., 2012). Collaborative innovations enable innovation processes and generation of innovations in all areas, like product and service improvement, processes improvement, with the support of information technology platforms to easier gather ideas. An important part of collaborative innovation is an architecture for supporting the participation of organizational members in collaborative innovation processes. In that frame are often emphasized corporate blogs as a way in which organizational members participate in collaborative innovations (Gloor, 2006; Pal & Kapur, 2009; Nedelko & Potocan, 2013).

Information Technology Solutions for Supporting Supply Chain Management

In this section we will outline the actual state of management practices utilization that are used in supply chains in order to support supply chain management in selected worldwide areas. Among a plethora of used practices, we will focus on the solutions that are based on information technology. An overview of different management solutions utilization in selected areas is outlined in Table 1 (Potocan & Dabić, 2012; Dabic et al., 2013).

Above outlined results in table 1 reveals different patterns for top used management tools in different worldwide regions. The customer relationship management tool is among the top used in well developed areas of North America and Europe, as well as organizations in Asia-Pacific region. Turning to the catching up economies like Slovenia and economies in Latin America, it is evident that in these economies are in the forefront other tools, whereas for instance customer relationship management is placed at sixth place in Slovenia.

Differences in patterns of customer relationship management use worldwide can be attributed to the different levels of the country's development and especially specific characteristics of organizations operating within different economic and cultural settings. Thus, well developed economies have already recognized the importance of managing and especially maintaining good relationships with their customers (Kumar, 2010). On the other hand, organizations in catching up economies have also recognized the importance of good relationships with their customers, but they are still preoccupied with issues related to the optimization of their business, like outsourcing activities, looking for core competencies, managing and capturing knowledge, implementation of total quality concepts, etc. (Potočan & Mulej, 2007; Potocan et al., 2012; Dabic et al., 2013).

DISCUSSION AND PRACTICAL IMPLICATIONS

With the emergence of information technology, especially the internet, there has been a rapid increase in terms of exchanging information between entities in the supply chain based on different information technology solutions. Thus, current supply chain management solutions, which are based on modern information technology and adjacent network equipment, enable seamless exchange information across organizational borders,

Table 1. Management solutions usage in selected worldwide areas^a

Management Tool	Use of Management Tools - Rank							
	Global 2006	Global* 2008	North America	EU 15	Asia-Pacific	Latin America	Croatia	Slovenia
Strategic Planning	1	2	1	1	2	1	8	8
Customer Relationship Management	2	4	3	4	1	9	4	6 (t)
Customer Segmentation	3	7	6	2 (t)	3	3 (t)	5	11
Benchmarking	4	1	2	2 (t)	9 (t)	2	2	2
Mission and Vision Statements	5	3	4	7	5 (t)	5	1	6 (t)
Core Competencies	6	9	5	5 (t)	4	10	3	3
Outsourcing	7	5	8	5 (t)	7 (t)	3 (t)	6	1
Business Process Reengineering	8	8	10 (t)	10 (t)	5 (t)	14 (t)	10 (t)	10
Scenario and Contingency Planning	9	13	9	8	14	7	10 (t)	12
Knowledge Management	10	14	12	10 (t)	7 (t)	14 (t)	7	4
Strategic Alliances	11	11	7	9	13	13	20	17
Balanced Scorecard	12	6	13 (t)	13	12	11 (t)	13	9
Supply Chain Management	13	12	13 (t)	14 (t)	11	11 (t)	12	15
Growth Strategies Tools	14	16	10 (t)	10 (t)	15 (t)	6	19	18
Total Quality Management	15	17	18 (t)	14 (t)	9 (t)	8	9	5
Shared Service Centres	16	15	18 (t)	19	15 (t)	16	14	23 (t)
Lean Operations	17	-	15	17 (t)	18 (t)	19	25	23 (t)
Collaborative Innovation	18	22	16	20	18 (t)	18	17	19
Loyalty Management	19	24	21	17 (t)	17	20 (t)	15	13
Mergers and Acquisitions	20	10	17	14 (t)	22	17	22 (t)	14
Six Sigma	21	-	22	23	20	20 (t)	22 (t)	20
Off shoring	22	-	18 (t)	21	24	24	24	25
Consumer Ethnography	23	-	23 (t)	22	21	22	18	21
Corporate Blogs	24	-	23 (t)	24	23	25	16	16
Radio Frequency Identification	25	-	25	25	25	23	21	22

^aNote: Data for Global average (GL) 2006 and 2008, North America (NA) 2006, European Union 15 (EU) 2006, Asia Pacific (AP) 2006, and Latin America (LA) 2006 are calculated upon results from management practices utilization research (Rigby & Bilodeau, 2009; Rigby, 2011). Data for Slovenia and Croatia are adopted from research of management practices in Central and East Europe economies (Potocan et al., 2012; Dabic et al., 2013) Abbreviation t stands for tight result. (-) Stand for not applicable.

between entities in the supply chain (Rodrigue, 1998; Chopra & Meindl, 2013).

According to outline cognitions, it is evident that plethora of management solutions supports supply chain management in practice. Among those several solutions are based on information technology. Among those solutions, that are based on information technology, we can emphasize some solutions as emerging, due to their novelty

and low level of utilization for supporting supply chain management. Based on above outlined cognitions, we will focus on several emerging solutions, out of 25 top used, like collaborative innovating, corporate blogs, radio-frequency identification, and shared services centers. Throughout discussion, we will outline potential and possible barriers of several emerging solutions, for supporting supply chain management.

The main advantage of collaborative innovations is to support innovative processes and generation of ideas, suggestions, and potential innovations and innovations for enhancing products, services, processes, leadership style, organizational methods, etc. With the help of information technology, plethora of ideas can be generated, filtered and reused again in getting the most innovative ideas. Organizational members don't have to be collocated; rather they can participate via corporate blogs, as a mean for enabling the participation. A possible drawback can be related to the huge amount of gathered ideas and needed time for filtration of the ideas and further development of these ideas. Another possible drawback it related to the ethical issues, regarding ownership of the ideas developed and possible free-riding of participants.

The key features of RFID utilization in supply chains are related to the reduction of needed time to handle and identification of goods arriving or leaving in the organization. Another key benefit of using RFID is high speed of identification and tracking of the objects. Utilization of RFID among partners in supply chain enables to scan and record all goods by all members in the supply chain using the same platform and/or approach. With such approach, supply chain members can overcome problems with time consuming identification of goods arriving and leaving organizations as well as problems with different systems for identification of goods (e.g. different bar codes, incompatible systems). The main drawback is associated with the costs of labeling and standardizing products and/or goods identification tags in order to ensure compatibility between all members of the supply chain. In order to implement RFID throughout the entire supply chain, members also need to have appropriate equipment to generate, read and modify RFID identifications.

Shared service centers contribute to the reduction of costs and can speed up the processes in supply chains, by, for instance centralizing and especially standardizing the process or several processes. For instance shared service centers are

successfully used and organizations send positive feedback about its utilization. Commonly are used in the areas of finance and customer services and customer support. An important feature of shared service centers is that entities involved in the supply chain can establish shared service centers, which enable support for all members.

With the implementation of shared service centers entities in the supply chain will have a common platform, which will enable easier collaboration between entities in the supply chain and will diminish the impact of barriers of possible technology incompatibles. Thus, all suppliers will have a common platform for communication with focal company and especially for managing orders. Currently, in business practices companies often encounter problems of incompatibility of information platforms for managing supplies and orders. Suppliers often need interfaces, which increases costs and require plenty of work with suppliers. A special case is also small supplier, which do not have large sums of money to invest in expensive enterprise resource planning systems, that will be compatible with larger of the focal company.

Thus establishing of a common platform by focal company will resolve these problems of expensive software with smaller suppliers as well the processes will be standardized and simplified for employees in focal company, since all suppliers will have the same interface and common platform for ordering. Even more convenient is utilization of such approach in case of vendor managed inventory systems, with which suppliers manage inventories of focal company.

A drawback can be related to the process of standardizing this service for all entities in the supply chain. Another consideration can be related to the suppliers or distributions, which are part of several supply chains. In such cases, they will need to have their own system for instance for keeping records, while they will face differing platforms when collaborating with different supply chains.

Slovenia is a small economy, which is importantly dependent upon export, thus many organizations are members of different supply chains or

networks across the globe, especially in Europe. Thus, low level of supply chain management information technology solutions utilization in Slovenian organizations, importantly hinders the flow of information exchange between partners in the supply chain. In that framework, some Slovenian organizations complain that they have troubles to enter into “larger supply chains” that intensively use supply chain management solutions, like automated data exchange, vendor managed inventory, etc., for collaboration with their partners. Reason lies in the fact, that members and/or focal organization in supply chain does not practice anymore “the old model of collaboration between supply chain partners”, like an electronic exchange of order via email, using faxes, email, etc.

In that frame, emerging solutions represent an important consideration for suppliers in the larger supply chain. Again, despite huge potential of emerging solutions for enhancing and optimizing supply chain management, there is still a plethora of open issues for smaller players in supply chains. Most important hindering factor is associated with adoption of “rules of focal company”. This problem magnifies when an organization is a member of several supply chains. For instance, like having different platforms for managing of supplies, different requirements in frame of RFID utilization, etc.

In terms of limitations, most significant are: (1) considering the limited number of management tools, (2) focus only on several emerging management solutions, and (3) discussing the role of information technology in supply chain management, only through the lenses of several selected emerging management solutions, based on information technology.

FUTURE RESEARCH DIRECTIONS

Building on the limitations the main directions of future research are: (1) including higher number of management tools in analysis, to more comprehensively evaluate usage of tools that are

supported with information and communication technology, (2) examine also other potential tools, which are based on information technology and could enhance supply chain management, and (3) discussing the role of information technology in supply chain management, by taking into consideration also key management tools used in supply chain in order to obtain more comprehensive picture about emerging solutions.

CONCLUSION

This chapter outlines the potentials of emerging information technology solutions to support and optimize supply chain management in practices. In that framework are recognized main potential of several selected emerging management solutions, like collaborative innovations, corporate blogs, radio-frequency identification, and shared services centers. Analysis of potentials and drawbacks reveal that in the long run those solutions will become standardized solutions in backing up supply chain management activities. It is also evident that joint actions of focal company as well as other participants in the supply chain will be needed in order to yield all possible benefits. For instance, organizations acting as suppliers can strive towards for common platforms for managing supplies in order to diminish problems, but focal company must provide adequate support - i.e. provide common platform and all necessary steps. On the other hand, the situation is, inversely, when for instance focal company want to introduce RFID. In such case supplier needs to adapt to the requirements of focal company.

REFERENCES

Ayeh, J. K. (2008). Information communications technology and global education: The challenges of the african virtual university learning centres in ghana. *Information Development*, 24(4), 266–273. doi:10.1177/02666666908098070



- Beamon, B. M. (1998). Supply chain design and analysis: Models and methods. *International Journal of Production Economics*, 55(3), 281–294. doi:10.1016/S0925-5273(98)00079-6
- Blanchard, D. (2006). *Supply chain management best practices*. New York, NY: Wiley.
- Chen, I. J., & Popovich, K. (2003). Understanding customer relationship management (crm): People, process and technology. *Business Process Management Journal*, 9(5), 672–688. doi:10.1108/14637150310496758
- Chen, Y. D., Xiao, X., Ren, H. J., Wang, Y. X., & Asme. (2012). *Research on application of information technology in supply chain management*. New York, NY: Amer Soc Mechanical Engineers.
- Chopra, S., & Meindl, P. (2013). *Supply chain management: Strategy, planning and operation*. Harlow, UK: Pearson.
- Christopher, M. (2011). *Logistics & supply chain management*. Harlow, UK: Pearson.
- Cohen, S., & Roussel, J. (2004). *Strategic supply chain management*. New York, NY: McGraw-Hill.
- Cooper, M., Lambert, D., & Pagh, J. (1997). Supply chain management. *The International Journal of Logistics Management*, 8(1), 1–14. doi:10.1108/09574099710805556
- Dabic, M., Potocan, V., Nedelko, Z., & Morgan, T. R. (2013). Exploring the use of 25 leading business practices in transitioning market supply chains. *International Journal of Physical Distribution & Logistics Management*, 43(10), 833–851. doi:10.1108/IJPDLM-10-2012-0325
- Daft, R. L. (2007). *Understanding the theory and desing of organizations*. Mason, OH: South-Western.
- Gloor, A. (2006). *Swarm creativity. Competitive advantage through collaborative innovation networks*. Oxford, UK: Oxford University Press. doi:10.1093/acprof:oso/9780195304121.001.0001
- Gunasekaran, A., & Ngai, E. W. T. (2004). Information systems in supply chain integration and management. *European Journal of Operational Research*, 159(2), 269–295. doi:10.1016/j.ejor.2003.08.016
- Gunasekaran, A., & Ngai, E. W. T. (2005). Build-to-order supply chain management: A literature review and framework for development. *Journal of Operations Management*, 23(5), 423–451. doi:10.1016/j.jom.2004.10.005
- Heitzer, J., & Render, B. (2003). *Operations management*. New York: Prentice Hall.
- Kumar, V. (2010). *Customer relationship management*. New York, NY: John Wiley & Sons, Ltd. doi:10.1002/9781444316568.wiem01015
- Li, B. G., Riley, M. W., Qi, E., & Zhu, X. W. (2001). *Information technology in supply chain management*. Beijing, China: Publishing House Electronics Industry.
- Lummus, R. R., & Vokurka, R. J. (1999). Defining supply chain management: A historical perspective and practical guidelines. *Industrial Management & Data Systems*, 99(1-2), 11–17. doi:10.1108/02635579910243851
- Manthou, V., Vlachopoulou, M., & Folinas, D. (2004). Virtual e-chain (vec) model for supply chain collaboration. *International Journal of Production Economics*, 87(3), 241–250. doi:10.1016/S0925-5273(03)00218-4
- Meidute, I., Davidaviciene, V., & Raudeliuniene, J. (2012). Research on applying radio frequency identification technology at lithuanian companies. In R. Ginevicius, A. V. Rutkauskas & J. Stankeviciene (Eds.), *7th international scientific conference business and management 2012* (pp. 989-996). Vilnius, Lithuania: Vilnius Gediminas Technical Univ Press, Technika. doi:10.3846/bm.2012.127
- Murphy, P. R. J., & Wood, D. F. (2004). *Contemporary logistics*. Upper Saddle River, NJ: Pearson Education.



- Nedelko, Z., & Cirnu, C. E. (2008). Distance education and videoconferencing. In M. Vlada, G. Albeanu, & D. Popovici (Eds.), *Proceedings of the 3rd international conference on virtual learning* (pp. 389-396).
- Nedelko, Z., & Potocan, V. (2013). The role of management innovativeness in modern organizations. *Journal of Enterprising Communities*, 7(1), 36–49. doi:10.1108/17506201311315590
- Nigel, S. (1996). *The blackwell encyclopedic dictionary of operations management*. Cambridge, UK: Blackwell Publishers.
- Pal, S. K., & Kapur, V. (2009). *The gradual adoption & maturity of corporate blogs*. Norristown: Int Business Information Management Assoc-Ibima.
- Potocan, V., & Dabić, M. (2012). *Management tools in business environment*. Maribor: Faculty of Economics and Business.
- Potočan, V., & Mulej, M. (2007). *Transition into innovative enterprise*. Maribor, Slovenia: Faculty of Economics and Business.
- Potocan, V., Nedelko, Z., & Mulej, M. (2012). Influence of organizational factors on management tools usage in slovenian organizations. *Inzinerine Ekonomika-Engineering Economics*, 23(3), 291–300.
- Rigby, D. K. (2011, May 5). *Management tools 2011: An executive's guide*. Retrieved from <http://www.bain.com>
- Rigby, D. K., & Bilodeau, B. (2009, April 12). *Management tools and trends 2009*. Retrieved from <http://www.bain.com>
- Rigby, D. K., & Bilodeau, B. (2011, May 5). *Management tools and trends 2011*. Retrieved from <http://www.bain.com>
- Rodrigue, J. P. (1998, November 2). *Evolution of logistical integration*. Retrieved from <http://people.hofstra.edu/geotrans/eng/ch5en/conc5en/evolutionlogistics.html>
- Rushton, A., Oxley, J., & Croucher, P. (2001). *The handbook of logistics and distribution management*. London: K. Page.
- Sahin, F., & Robinson, E. P. (2002). Flow coordination and information sharing in supply chains: Review, implications, and directions for future research. *Decision Sciences*, 33(4), 505–536. doi:10.1111/j.1540-5915.2002.tb01654.x
- Simchi-Levi, D., Kaminsky, P., & Simichi-Levi, E. (2009). *Designing and managing the supply chain: Concepts, strategies and case studies*. Boston, MA: McGraw Hill.
- Slack, N., Chambers, S., Johnson, R., & Betts, A. (2006). *Operations and process management*. Harlow, UK: Prentice Hall.
- Stucky, W., & Weiss, P. (2008). E-skills and standards - prerequisites to fully exploit the potential of ict in europe. In R. Kaschek, C. Kop, C. Steinberger, & G. Fliedl (Eds.), *Information systems and e-business technologies* (Vol. 5, pp. 306–317). doi:10.1007/978-3-540-78942-0_30
- Subramani, M. (2004). How do suppliers benefit from information technology use in supply chain relationships? *Management Information Systems Quarterly*, 28(1), 45–73.
- Tamosiuniene, R., & Jasilioniene, R. (2007). Customer relationship management as business strategy appliance: Theoretical and practical dimensions. *Journal of Business Economics and Management*, 8(1), 69–78.

Wang, W. Y. C., & Chan, H. K. (2010). Virtual organization for supply chain integration: Two cases in the textile and fashion retailing industry. *International Journal of Production Economics*, 127(2), 333–342. doi:10.1016/j.ijpe.2009.08.006

ADDITIONAL READING

Arntzen, B. C., Brown, G. G., Harrison, T. P., & Trafton, L. L. (1995). Global supply chain management at digital-equipment-corporation. *Interfaces*, 25(1), 69–93. doi:10.1287/inte.25.1.69

Bagdoniene, L., & Hopeniene, R. (2013). Interdependence and trust in business relationships development: The differences of lithuanian professional services providers' and clients' approaches. *Inzinerine Ekonomika-Engineering Economics*, 24(2), 160–168.

Browne, J., & Zhang, J. (1999). Extended and virtual enterprises - similarities and differences. *International Journal of Agile Management Systems*, 1(1), 30–36. doi:10.1108/14654659910266691

Certo, S. C., & Certo, S. T. (2009). *Modern management: Concepts and skills*. New Jersey. Upper Saddle River, NJ: Pearson Education.

Croxton, K. L., García-Dastugue, S. J., Lambert, D. M., & Rogers, D. S. (2001). The supply chain management processes. *International Journal of Logistics Management*, 12(2), 13–36. doi:10.1108/09574090110806271

Davenport, T. H., & Short, J. E. (1990). The new industrial-engineering - information technology and business process redesign. *Sloan Management Review*, 31(4), 11–27.

Davidow, W. H., & Malone, M. S. (1993). *The virtual corporation: Structuring and revitalizing the corporation for the 21st century*. New York, NY: HarperBusiness.

Dyck, R., & Mulej, M. (1998). *Self-transformation of the forgotten four-fifth*. Dubuque, IA: Kendall/Hunt.

Eckhardt, J., & Rantala, J. (2012). The role of intelligent logistics centres in a multimodal and cost-effective transport system. In P. Papaioannou (Ed.), *Transport research arena 2012* (Vol. 48, pp. 612–621). doi:10.1016/j.sbspro.2012.06.1039

Fisher, C., & Lovell, A. (2006). *Business ethics and values: Individual, corporate and international perspectives*. New York: Prentice Hall.

Lai, H.-M., Chen, C.-P., & Chang, Y.-F. (2014). Determinants of knowledge seeking in professional virtual communities. *Behaviour & Information Technology*, 33(5), 522–535. doi:10.1080/0144929X.2013.864709

Laudon, K. C., & Laudon, J. P. (2002). *Management information systems*. New Jersey, NJ: Prentice-Hall.

Nedelko, Z., Potocan, V., & Dabić, M. (2015). Current and future use of management tools. *E a M. Ekonomie a Management*, 18(1), 28–45. doi:10.15240/tul/001/2015-1-003

Potocan, V., Mulej, M., & Cancer, V. (2008). Influence, of values, culture, ethics and norms on economic results: Case of slovenia. *Drustvena Istrazivanja*, 17(3), 373–395.

Saafein, O., & Shaykhian, G. A. (2014). Factors affecting virtual team performance in telecommunication support environment. *Telematics and Informatics*, 31(3), 459–462. doi:10.1016/j.tele.2013.10.004

Sahay, B. S., Gupta, J. N. D., & Mohan, R. (2006). Managing supply chains for competitiveness: The indian scenario. *Supply Chain Management-an International Journal*, 11(1), 15–24. doi:10.1108/13598540610642439

Sarkis, J., Talluri, S., & Gunasekaran, A. (2007). A strategic model for agile virtual enterprise partner selection. *International Journal of Operations & Production Management*, 27(11), 1213–1234. doi:10.1108/01443570710830601

Schalk, R., & Curseu, P. L. (2010). Cooperation in organizations. *Journal of Managerial Psychology*, 25(5), 453–459. doi:10.1108/02683941011048364

Thomas, D. J., & Griffin, P. M. (1996). Coordinated supply chain management. *European Journal of Operational Research*, 94(1), 1–15. doi:10.1016/0377-2217(96)00098-7

KEY TERMS AND DEFINITIONS

Collaborative Innovations: Enable innovation processes and generation of innovations in all areas, like product and service improvement, processes improvement, with the support of information technology platforms to easier gather ideas.

Customer Relationship Management Solution: Is an information technology supported tool, which enables organizations to collect and manage large amounts of data about customers.

Corporate Blogs: Are defined as a way in which organizational members participate in collaborative actions, like collaborative innovations.

Emerging Solutions: Is an information technology supported tool, which support information exchange, collaboration and processes in supply chains. Typically, these solutions are novel and less used in business practice. Most commonly emphasized emerging solutions are collaborative innovating, corporate blogs, radio-frequency identification, and shared services centers.

Information Technology Solutions: Are different types or applications of information technology that support processes in organizations. In terms of supply chain, information technology solutions are aimed to support customer and supplier relations, as well as internal processes in single organizations involved in the supply chain.

Supply Chain Management Solution: Is an information technology supported tool, which enables synchronization of the efforts of all parties involved in supply chain, from suppliers, through manufacturers, distributors, dealers, to end customers.

Radio-Frequency Identification: Enables wireless identification of objects and allows tracking the movement of the material flow within the supply chain.

Shared Services Centers: Reduce costs, since one or several operations are consolidated into shared operation and used by multiple divisions, organizations, users, etc.



Samsung Company and an Analysis of Supplier–Side Supply Chain Management and IT Applications

Amber A. Smith-Ditizio

Texas Woman's University, USA

Alan D. Smith

Robert Morris University, USA

INTRODUCTION

Supply Chain Management

Supply chain management (SCM) is the oversight of materials, information, and finances as they move in a process from supplier to manufacturer to wholesaler to retailer to consumer. SCM involves coordinating and integrating these flows both within and among companies. It is said that the ultimate goal of any effective SCM-related system is to reduce inventory (with the assumption that products are available when needed). As a solution for successful SCM, sophisticated software systems with web-based interfaces are competing with web-based application service providers (ASP) who promise to provide part or all of the SCM service for companies who rent their service (Von Haartman, 2012).

The discipline of SCM includes the active management of a company's supply chain activities to maximize customer value and achieve a sustainable competitive advantage. It represents a conscious effort by the supply chain firms to develop and run supply chains in the most effective & efficient ways possible. Supply chain activities cover everything from product development, sourcing, production, and logistics, as well as the information systems needed to coordinate these activities (Pradhananga, Hanaoka, & Satayaprasert, 2011).

The organizations that make up the supply chain are linked together through physical flows and information flows. Physical flows involve the transformation, movement, and storage of goods and materials (Idris, Rahman, Hassan, Aminudin, & Alolayyan, 2013; Ketikidis, Hayes, Lazuras, Gunasekaran, & Koh, 2013; Mateen & More, 2013; Varaprasad, Sridharan, & Unnithan, (2013). They are the most visible piece of the supply chain. But just as important are information flows. Information flows allow the various supply chain partners to coordinate their long-term plans, and to control the day-to-day flow of goods and material up and down the supply chain. SCM flows can be divided into the product flow, the information flow, and the finances flow (Basu & Nair, 2012; Brito & Botter, 2012; Bulcsu, 2011). The product flow includes the movement of goods from a supplier to a customer, as well as any customer returns or service needs. The information flow involves transmitting orders and updating the status of delivery. The financial flow consists of credit terms, payment schedules, and consignment and title ownership arrangements.

SCM Basics

There are basic components that make up SCM: plan, source, make, deliver, and return (Sprovieri, 2008; Smith, 2011; Summers & Scherpereel, 2008). Planning is the strategic portion of SCM,

as companies need a strategy for managing all the resources that go toward meeting customer demand for their product or service. A big major portion SCM planning is developing a set of metrics to monitor the supply chain so that it is efficient, costs less and delivers high quality and value to customers. Next follows source where companies must choose suppliers to deliver the goods and services they need to create their product. Therefore, supply chain managers must develop a set of pricing, delivery and payment processes with suppliers and create metrics for monitoring and improving the relationships. And then, SCM managers can put together processes for managing their goods and services inventory, including receiving and verifying shipments, transferring them to the manufacturing facilities and authorizing supplier payments.

Making is the manufacturing step. Supply chain managers schedule the activities necessary for production, testing, packaging and preparation for delivery (Tari & Sabater, 2004; Tiwari, Turner, & Sackett, 2007; Varaprasad, et al., 2013). This is the most metric-intensive portion of the supply chain, where companies are able to measure quality levels, production output and worker productivity. Delivering is the part that many SCM insiders refer to as logistics, where companies coordinate the receipt of orders from customers, develop a network of warehouses, pick carriers to get products to customers and set up an invoicing system to receive payments. Returning can be a problematic part of the supply chain for many companies. Supply chain planners have to create a responsive and flexible network for receiving defective and excess products back from their customers and supporting customers who have problems with delivered products.

There are several types of types of SCM software that deals with planning applications and execution applications (Sodhi & Lee, 2007). Planning applications use advanced algorithms to determine the best way to fill an order. Execution applications track the physical status of goods, the management of materials, and financial informa-

tion involving all parties. Some SCM applications are based on open data models that support the sharing of data both inside and outside the enterprise (e.g., this is called the extended enterprise, and includes key suppliers, manufacturers, and end customers of a specific company). This shared data may reside in diverse database systems, or data warehouses, at several different sites and companies.

By sharing this data upstream (e.g., with a company's suppliers) and downstream (e.g., with a company's clients), SCM applications have the potential to improve the time-to-market of products, reduce costs, and allow all parties in the supply chain to better manage current resources and plan for future needs. Increasing numbers of companies are turning to websites and web-based applications as part of the SCM solution (Paksoy & Cavlak, 2011; Pettersson & Segerstedt, 2011). A number of major websites offer procurement marketplaces where manufacturers can trade and even make auction bids with suppliers.

BACKGROUND

Case Study: Samsung Electronics Company

Background and History

Samsung Electronics Co., Ltd. is a world leader in digital media and digital convergence technologies industries. Samsung operates in over 50 countries worldwide with its main headquarters being located in city of Suwon, South Korea. President of Samsung Group is Mr Lee Kun-Hee who has held the role since the passing of his father who founded Samsung, Lee Byung-Chul. Samsung's name can be broken down into "sam" meaning three and "sung" meaning star in Korean." Samsung Electronics Co., Ltd. is mainly engaged in the production of consumer electronic products. It operates in three business divisions: consumer electronics (CE) division, which involves in the

color televisions (CTVs), monitors, printers, air conditioners, refrigerators, laundry machines and others; information technology & mobile communications (IM) division, which involves in the production of computers, handheld phones (HHPs), network systems, digital cameras and others, as well as device solutions (DM) division, which is divided into semiconductor and display business parts, providing dynamic random access memories (DRAMs), flashes, thin film transistor-liquid crystal displays (TFT-LCDs) and others. The Company distributes its products within domestic market and to overseas markets (Sodhi & Lee, 2007). Much of Samsung Electronics Company's manufacturing has since moved outside of Korea into emerging countries such as China and other countries in ASEAN such as Vietnam, the Philippines, and India. Some of the reasons for this transfer of labor overseas are to capitalize on the low cost as well as the raw materials available there.

Samsung Electronics was created in 1969 as a division of the mammoth Korean chaebol Samsung Group. The unit was established as a means of getting Samsung into the burgeoning television and consumer electronics industry. The division's first product was a small and simple black-and-white television that it began selling in the early 1970s. From that product, Samsung Electronics gradually developed a diverse line of consumer electronics that it first sold domestically, and later began exporting. The company also began branching out into color televisions, and later into a variety of consumer electronics and appliances. By the 1980s, Samsung was manufacturing, shipping, and selling a wide range of appliances and electronic products throughout the world.

During the 1970s and 1980s, Samsung Group created a number of electronics-related divisions, several of which were later grouped into a single entity known as Samsung Electronics Co., Ltd. Samsung Electron Devices Co. manufactured picture tubes, display monitors, and related parts. Samsung Electro-Mechanics Co. made VHF and UHF tuners, condensers, speakers, and other gear. Samsung Corning Co. produced television glass

bulbs, computer displays, and other components. Finally, Samsung Semiconductor & Telecommunications Co. represented Samsung in the high-tech microchip industry. Rapid growth in those industries, combined with savvy management, allowed the combined Samsung Electronics Co., Ltd., to become Samsung Group's chief subsidiary by the end of the 1980s (Sodhi & Lee, 2007).

Between 1977 and 1987, Samsung Group's annual revenues surged from US\$1.3 billion to US\$24 billion (e.g., about 20% of South Korea's entire GNP). Much of that growth was attributable to Samsung Electronics. Samsung's rapid rise and technical achievements put the company in the spotlight in the semiconductor industry. Its 4-megabit chip, in fact, had made it the leading global producer of DRAM chips by early 1995. Furthermore, Samsung was increasing its investment in development still further, as evidenced by a US\$2.5 billion outlay to develop a 64-megabit DRAM chip by 1998. In December of 1995, development on the world's first 1-gigabit synchronous DRAM chip was also in the works. Exports for the year increased to more than US\$10 billion.

Samsung's positive results continued in 2000 as the firm secured US\$26 billion in sales and US\$4.7 billion in net profits. Memory chips accounted for 38% of sales, information and telecommunications equipment secured 22%, digital media took 27%, and home appliances accounted for 8% of sales. The company looked to strategic partnerships, research and development, and growth, to maintain its positive financial record. In March 2001, it teamed up with Dell Computer Corporation in a US\$16 billion technology and research and development agreement. In addition, the company was selected by China to provide CDMA cellular phone networks in its four major cities.

Sales figures did decrease slightly in the first part of 2001, however, due to a slowdown in the personal computer market, an oversupply of LCDs, and a slowdown in the cellular market. Nevertheless, Samsung management continued to focus on remaining a leader in the electronics industry.



Samsung has made many innovations over the years in the electronics space often being the first one to do so. Some of these innovations include the following:

- Developed world's first 128MB Synchronous DRAM and 128MB Flash memory.
- Developed the world's first 64Mb DRAM, 256 Mb DRAM, and 1Gb DRAM.
- Developed the world's first 60-nano 8GB NAND memory chip.
- Developed the worlds' first 50nm 1G DRAM.
- Developed the world's first 30nm-class 64Gb NAND Flash™ memory.
- Developed the world's first 2Gb 50 NANO ("Fortune Global Top Company ...," 2016).

Mission and Strategy

Samsung's corporate mission statement and business philosophy: To fuel Samsung's aspirations to create new technologies that inspire the world, Samsung boldly invests billions in R&D and future technologies. Samsung operates 34 R&D centers globally, with 69,000 employees devoted to R&D. In 2013, Samsung invested USD 13.6 billion (KRW 14.3 trillion) in R&D and registered 4,676 new patents in the U.S. alone, ranking Samsung Electronics as the second-highest patent winner in the U.S. for the eighth consecutive year. Samsung's intellectual property portfolio today includes more than 100,000 patents ("Fortune Global Top Company ...," 2016).

Recognizing that great consumer experiences require a thoughtful integration of both hardware and software, Samsung Electronics has increased global investment to fuel software innovation, including embracing creative talent from both inside and outside the company. Samsung also strives to bring greater value to consumers' lives through its "Make It Meaningful" design philosophy, and today it employs some 1,500 designers across our global design centers in six countries.

The employees of Samsung Electronics are passionate about creating a better future for people and bringing about positive change in the communities in which we operate. Through Samsung Hope for Children programs, Samsung continues to pursue the goal of providing improved health, education and training opportunities to people around the world. Samsung also places a high priority on environmental sustainability, and, through their PlanetFirst™ initiative, Samsung are applying an innovative technology to achieve eco-friendly development throughout the product life cycle and to reduce our environmental footprint.

Samsung's plan for the future is driven by a deep understanding of people, cultures, and the impact of technologies that drive growth and create markets. To foster future growth, Samsung Electronics is pushing toward a 2020 corporate vision to "Inspire the World, Create the Future," and achieve annual revenue of USD 400 billion while cementing its reputation as one of the world's most innovative and admired companies ("Fortune Global Top Company ...," 2016).

SCM Approach

Since the electronics market changes quickly, Samsung need to be able to create a flexible production plan. The Samsung Group of companies have large, complex, global supply chains in most of the products it manufactures and makes extensive use of SCM solutions and process innovations to support and improve its operations. Most notably, at Samsung, advanced planning and scheduling (APS) systems have been adopted since the 1990s and have brought the company many successes in terms of operational excellence (Khanna, Song, & Lee, 2011). APS is especially well-suited to environments where simpler planning methods cannot adequately address complex trade-offs between competing priorities. Production scheduling is intrinsically very difficult due to the (approximately) factorial dependence of the size of the solution space on the number of items/products to be manufactured. APS systems

function as resources for efficient planning and scheduling. This type of system tracks costs based on the activities that are responsible for driving costs in the production of manufactured goods. It also allocates raw materials and production capacity optimally to balance demand and plant capacity. APS systems utilize where plant capacity is constrained, many different products are produced in each facility, or when products require a large number of components. To achieve the economy of scale, Samsung standardize processes and parts and the majority of components are produced in Korea.

Despite the extensive use of SCM solutions and process innovation to improve global business operation, in 2004, the company still felt that its supply chain operations had significant room for improvement. In the early 1990s, the Group's senior management decided to capitalize on the potential synergy between SCM and six-sigma metrics (Yang, Choi, Park, Suh, & Chae, 2007). Six-sigma is a set of techniques and tools for process improvement (Idris, et al., 2013; Ketikidis, et al., 2013; Park & Min, 2013; Rajapakshe, Dawande, & Sriskandarajah, 2013; Rajeev, Rajagopal, & Mercado, 2013). Six-Sigma is a disciplined, data-driven approach and methodology for eliminating defects (e.g. driving toward six standard deviations between the mean and the nearest specification limit) in any process – from manufacturing to transactional and from product to service. Samsung's SCM Business Team (SBT) researched six-sigma approaches at General Electric (GE), DuPont and Honeywell to get perspectives on how other companies have innovatively applied six-sigma to similar needs. Later, they create the methodology based on GE approach called DMAEV (e.g., define, measure, analyze, enable and verify).

Samsung was ranked 7th in a respected analyst's ranking of the global top 25 companies in supply chain excellence (Yang, et al., 2007). There are various approaches and systems available for process innovation. Six sigma and SCM-based metrics are among those techniques aiming for

process and quality improvement, and synchronization of a company's value chain, from inbound logistics to sales and customer services (Carvalho, Cruz-Machado, & Tavares, 2012; Hamidi, Farahmand, Sajjadi, & Nygard, 2012; Kumar, Shankar, & Yadav, 2011; Mathirajan, Manoj, & Ramachandran, 2011; More & Babu, 2012). At Samsung, SCM and six-sigma have been two important enablers for the group's management innovation and growth.

However, management realized that there is significant room for improvement in its SCM operation. Thus, the effort has been synthesizing SCM and six-sigma and developing a unique six-sigma based methodology to improve its SCM operation. Samsung's effort and investment has turned out to be fruitful. Their SCM six-sigma program has produced highly qualified and talented SCM specialists, who are currently training the methodology to other members in their organizations and leading SCM projects. SCM projects are being prepared and conducted in a more disciplined way and their outcomes are continuously monitored and shared through Samsung's repository for six-sigma. The company's endeavor for global optimum is continuing and SCM and six-sigma metrics are expected to play an increasingly enabling role.

Samsung also have implemented short cycle time and low inventory (SLIM) methodology in its electronics manufacturing processes. SLIM is a set of methodologies and scheduling applications for managing cycle time in semiconductor manufacturing. SLIM includes methodology for calculating target cycle times and target WIP levels for individual manufacturing steps, heuristic algorithms for factory floor scheduling, and optimization-based capacity analysis (Leachman, Kang, & Lin, 2002). Between 1996 and 1999, Samsung Electronics Corp., Ltd., implemented SLIM in its semiconductor manufacturing facilities (Leachman, et al., 2002). It reduced manufacturing cycle times to fabricate dynamic random access memory devices from more than 80 days to less than 30. Considering the decline of sell-

ing prices for dynamic random access memory devices, SLIM enabled Samsung to capture an additional US\$1 billion in sales revenue compared to the revenue it would have realized had cycle times not been reduced (Leachman, et al., 2002).

Samsung MIS Applications

In the current ever-developing business world, Samsung has employed various information technology (IT) applications in order to better manage its supply chain. Perhaps, the initial challenge was to develop a model for managing the supply chain that addressed Samsung's specific business requirements. The company relied heavily on existing technology while working internally to develop its own custom application set based on the observations of successful systems. The result is a SCM system with end-to-end visibility into every segment of the supply chain. Samsung annually manages over 50,000 employees across 42 global research facilities as well as a global network of 286,000 employees in more than 80 countries. Because it currently contains no brick and mortar locations getting the products from the manufacturing factory to the retail shelf is the biggest concern of Samsung's which can be a daunting task with so many employees and locations needing to coordinate seamlessly. In order to manage such a vast amount of goods, Samsung implemented wireless local area network (LAN), radio-frequency identification (RFID), and developed custom applications for internally managing the data.

Wireless LAN

Wireless local area network or (WLAN) is a communications network that provides connectivity to wireless devices within a limited geographic area. Wi-Fi is the universal standard for wireless networks. In the office, Wi-Fi networks are adjuncts to the wired networks. At home, a Wi-Fi

network can serve as the only network since all laptops and many printers come with Wi-Fi built in, and Wi-Fi can be added to desktop computers via USB. Wi-Fi is achieved with a wireless base station, called access points. Its antennas transmit and receive a radio frequency within a range of 30-150 ft. through walls and other non-metal barriers. Since all wireless and wired computers are interconnected, they can exchange data with each other for backup and file sharing.

An enterprise-class WLAN employs a large number of individual access points to broadcast the signal to a wide area. The access points have more features than home or small office WLAN equipment, such as better security, authentication, remote management, and tools to help integrate with existing networks. These access points have a larger coverage area than home or small office equipment, and are designed to work together to cover a much larger area. This equipment can adhere to the 802.11a, b, g, or n standard, or to security-refining standards, such as 802.1x and WPA2.

True efficient business environments depend on a reliable and fast network to meet the rising demand for bandwidth. In order to support enterprise mobility trends like BYOD, the Internet of Things (IoT) and digital consumption, Samsung's WLAN solution provides an improved end-user experience with noticeably faster network connectivity. Higher network speeds allow users to get on and off the network quickly, which conserves batteries and maximizes airtime for all devices ("Samsung Smart LAN," 2015).

Unlike many competitors, Samsung's WLAN technology is designed to function similarly to its own LTE solutions, fairly and efficiently optimizing the distribution of airtime, which leads to higher throughput, capacity, and an overall improved experience. Additionally, Samsung Wireless LAN simplifies network administration for IT managers, allowing for faster and more effective device management and service ("Samsung Smart LAN," 2015).

RFID Applications

RFID tags are part of intelligent automatic identification and data capture technologies (AIDC) that can communicate to a networked system to track every product that you put in a customer's shopping cart. RFID tags, a technology once limited to tracking cattle, are tracking consumer products worldwide. Many manufacturers use the tags to track the location of each product they make from the time it's made until it is pulled off the shelf and tossed in a shopping cart (Tari & Sabater, 2004; Tiwari, et al., 2007; Varaprasad, et al., 2013). RFID tags are an improvement over bar codes because the tags have read and write capabilities. Data stored on RFID tags can be changed, updated and locked. Some stores that have begun using RFID tags have found that the technology offers a better way to track merchandise for stocking and marketing purposes. Through RFID tags, stores can see how quickly the products leave the shelves and which shoppers are buying them.

Samsung have joined with DHL and Media-market to pilot an RFID-enabled smart packaging concept designed to improve supply chain efficiency for high-value electronics goods in Europe (Johnson, 2013). The supply chain for consumer electronics products like TVs, digital cameras, video games and mobile phones is very complex. Often the flow of merchandise comes from multiple sources including manufacturers and assemblers. Goods are then passed off to logistics providers for packaging, transport and delivery to warehouses of retail venues, where they are then encoded before being displayed on store shelves. Adding to the supply chain complexity is the wide variety of very valuable items shipped, many with very small dimensions. These characteristics make electronic products much more susceptible to potential vulnerabilities throughout the supply chain such as theft and loss.

RFID traceability begins when DHL receives products from Samsung at two warehouse locations that serve as high-value storage areas. In these restricted areas, products are inspected, wrapped

and stocked. Larger products like televisions are stocked in a traditional bulk area. RFID tags are activated by connecting the tag's serial number to the product's serial number. From thermo-retraction, to stockpiling, picking and shipment preparation, the tags enable the electronics products to be monitored at each step (Johnson, 2013).

The pallets pass through a door that is equipped with a scanner and RFID antennae that monitor the exit of goods from the high-value storage area which can read up to 300 tags per pallet (Johnson, 2013). The operator verifies the exiting process from the storage area through a monitor that reports the quantity and the relative shipment. After it leaves the storage area, the pallet is placed on the exit platform with its untagged counterparts. Each product is traced and identified from the picking process, to the loading phase onto the vehicle and to the Saturn retail venue.

Upon arrival at the retail venue, the shipment passes through a gateway equipped with scanner and RFID antennae that automatically monitors the incoming merchandise. RFID also allows Samsung to reduce stock in transit and to have better visibility into sell out information. In addition, products arrive at the retail venue more quickly and make it to the store shelves faster, a crucial element for electronics retailers and manufacturers that battle product obsolescence. RFID-tagged products enable the possibility to reduce self-service product stock out and the possibility of obsolescence, a very relevant occurrence as far as notebooks are concerned. RFID has also allowed the retailers to free up employees from inventory tasks and be put to use making sales.

IT Applications

MySingle™ is an intranet service created in house by Samsung in order for employees to communicate worldwide. An intranet is a network based on TCP/IP protocols belonging to an organization, usually a corporation, accessible only by the organization's members, employees, or others with authorization. An intranet's websites and

operations look and act just like any other Web sites, but the firewall surrounding an intranet fends off unauthorized access. Like the Internet itself, intranets are used to share information. Secure intranets are now the fastest-growing segment of the Internet because they are much less expensive to build and manage than private networks based on proprietary protocols.

Samsung's intranet was created by their subsidiary SDS. Samsung SDS provides information technology services which include consulting services, technical services, and outsourcing services. SDS is expanding its business area by investing on R&D and emerging IT technologies such as U-city, RFID and Engineering Outsourcing. SDS provides its IT services by operating 11 offices and data centers in 11 countries.

The e-mail system is a web-client first devised in 2000 by Samsung subsidiary Samsung Data Systems and went live across the company's entire network in 2001. The previous system was not that effective in communicating with each other. It is reported that a total of 425,000 employees of Samsung make 486 searches for employee information such as telephone numbers and e-mail addresses a day on average, which takes considerable time since the system has required many clicks and an exact name to find information on other employees.

Samsung, therefore, has enhanced the effectiveness of the intranet system. Aside from a name, searching the name of position or team can enable the person they are looking for to be located. The system is also set up to work with foreign based employees as well. Before, they used to be required to type in an exact full name of a Korean worker which is quite hard for them to do. Now with the renewed system, employees worldwide search for a person they want to communicate with without typing in their full name.

"Provided that it takes 20 seconds for a staff member to search for another member, it would 27,000 hours, to search for 486 employees," said Yoon-Jung (Personal Note 1, 2015), Samsung Portal UX Team Leader. "Thanks to the renewed

system, the speed is growing much faster than before, which makes workers much more satisfied by the progress," she added. In 2016, the leading Korean company plans to expand the streamlined system into other systems such as mail receiver search, document approver search and business card search. The system will also be morphed into a Microsoft-based platform as Samsung looks to reinforce real-time communications with 400,000 Samsung employees worldwide and to enhance their business performance by sophisticating its intranet system. Samsung Group will pilot-run an MS-based work system and will decide when to apply the system across the board based on the outcome of the pilot run. Yoon-Jung (Personal Note 1, 2015) from Samsung Group said, "There were issues with having effective and smooth communication with overseas employees under old system. So by creating everything on one platform and integrating everyone we hope to create better cohesion and unity among Samsung employees."

Managerial Application of IT Technologies to SCM

To optimize the functionality of their supply chain, Samsung has implemented existing technologies as well as providing customized tools for greater visibility throughout the supply chain. The development and combination of these technologies, while difficult in the beginning, have accounted for value added to the business. With the vast amount of manufactured goods that need to be transported daily across vast distances, Samsung has employed several systems to assist in accurately managing data collection. Two noteworthy technologies include wireless local area network (LAN) and RFID-embedded technologies.

The wireless LAN (WLAN) allows for information sharing within a given area. The system is hampered by its limited range along with the need to be within the specified region in order for this system to work efficiently and effectively. Because of the expanding technology and innovations to come, the company is poised to not only utilize

but provide the tools for other enterprises who wish to use WLAN in their SCM systems. By providing the WLAN and devices that operate on the network, Samsung is able to streamline their products to create an efficiency that can lead to a competitive advantage.

FUTURE RESEARCH DIRECTIONS

In addition to using wireless LANs, Samsung is piloting programs which use RFID-embedded technologies to further enhance its inventory controls. Currently with other systems barcodes are specifically adapted for a certain distribution center and are not universally recognized throughout the organization. The use of RFID is still in its infancy with Samsung, but the consensus is that eventually, if implemented accordingly the RFID could universalize the management of inventory. RFID-tagged products enable the possibility to reduce self-service product stock out and the possibility of obsolescence, a very relevant occurrence as far as notebooks are concerned. RFID has also allowed the retailers to free up employees from inventory tasks and be put to use making sales.

MySingle™ is simply an intranet knowledge center which allows each employee the ability to submit anything they deem useful to any aspect of the business. From a supply chain stance, it provides employees at any point along the supply chain access to vital information and if necessary, to join forces with other employees on work projects. The payoff is that employees who are greater informed generally make better decisions, this further adds value to the business.

CONCLUSION

As the emphasis on six-sigma quality, SCM metrics, and IT-technologies continues to grow as companies place more emphasis on cost effective logistics, Samsung is well placed to adapt to changing needs and or tastes of its customers

more rapidly. By utilizing available technologies Samsung has continued to foster positive change internally. Samsung has seen its brand grow to be famous globally for its service and design, its commitment of its employees, and reliable product innovations.

Many companies look to Samsung for their own SCM systems and IT applications for inspiration and assistance in their own enterprises. Samsung takes on a responsible approach to enterprise and global citizenship as well as pioneering globalization with its many partners and customers. Management at the company is willing to take their company in the global economy in a progressive direction for the next generation. Ultimately, the acquisition and execution best business practices and processes previously mentioned are done for the purpose of achieving the company mission of creating superior products and services that contribute to a global society.

REFERENCES

- Basu, P., & Nair, S. K. (2012). Supply chain finance enabled early pay: Unlocking trapped value in B2B logistics. *International Journal of Logistics Systems and Management*, 12(3), 334–353. doi:10.1504/IJLSM.2012.047605
- Brito, T. B., & Botter, R. C. (2012). Feasibility analysis of a global logistics hub in Panama. *International Journal of Logistics Systems and Management*, 12(3), 247–266. doi:10.1504/IJLSM.2012.047601
- Bulcsu, S. (2011). The process of liberalising the rail freight transport markets in the EU: The case of Hungary. *International Journal of Logistics Systems and Management*, 9(1), 89–107. doi:10.1504/IJLSM.2011.040061
- Carvalho, H., Cruz-Machado, V., & Tavares, J. G. (2012). A mapping framework for assessing supply chain resilience. *International Journal of Logistics Systems and Management*, 12(3), 354–373. doi:10.1504/IJLSM.2012.047606

Fortune Global Top Company. Samsung: Samsung History Timeline. (2016). Retrieved March 21, 2016 from http://www.samsung.com/us/about-samsung/samsung_group/history/

Hamidi, M., Farahmand, K., Sajjadi, S. R., & Nygard, K. E. (2012). A hybrid GRASP-tabu search metaheuristic for a four-layer location-routing problem. *International Journal of Logistics Systems and Management*, 12(3), 267–287. doi:10.1504/IJLSM.2012.047602

Huh, V. (2014). *Samsung Streamlines Intranet System, My Single*. Retrieved May 6, 2015, from <http://koreabizwire.com/samsung-streamlines-intranet-system-my-single/19423>

Idris, F., Rahman, N. M. N. A., Hassan, M. E. N., Aminudin, N., & Alolayyan, M. N. (2013). Validating dimensions of operations flexibility in Malaysian service firms: A confirmatory factor analytics approach. *International Journal of Services and Operations Management*, 15(1), 79–84. doi:10.1504/IJSOM.2013.050563

Johnson, J. (2013). *Sony Samsung and DHL pilot item level tagging for high-value goods*. Retrieved May 4, 2015 from <http://www.rfid24-7.com/2013/02/11/sony-samsung-dhl-pilot-item-level-tagging-for-high-value-electronics/>

Ketikidis, P. H., Hayes, O. P., Lazuras, L., Gunasekaran, A., & Koh, S. C. L. (2013). Environmental practices and performance and their relationships among Kosovo construction companies: A framework for analysis in transition economies. *International Journal of Services and Operations Management*, 15(1), 115–130. doi:10.1504/IJSOM.2013.050565

Khanna, T., Song, J., & Lee, K. (2011). *The Globe: The paradox of Samsung's rise*. Retrieved May 6, 2015, from <https://hbr.org/2011/07/the-globe-the-paradox-of-samsungs-rise>

Kumar, P., Shankar, R., & Yadav, S. S. (2011). Global supplier selection and order allocation using FQFD and MOLP. *International Journal of Logistics Systems and Management*, 9(1), 43–68. doi:10.1504/IJLSM.2011.040059

Leachman, R., Kang, J., & Lin, V. (2002). SLIM: Short cycle time and low inventory in manufacturing at Samsung Electronics. *Interfaces*, 32(1), 61–77. doi:10.1287/inte.32.1.61.15

Mateen, A., & More, D. (2013). Applying TOC thinking process tools in managing challenges of supply chain finance: A case study. *International Journal of Services and Operations Management*, 15(4), 389–410. doi:10.1504/IJSOM.2013.054882

Mathirajan, M., Manoj, K., & Ramachandran, V. (2011). A design of distribution network and development of efficient distribution policy. *International Journal of Logistics Systems and Supply Management*, 9(1), 108–137. doi:10.1504/IJLSM.2011.040062

More, D., & Babu, A. S. (2012). Benchmarking supply chain flexibility using data envelopment analysis. *International Journal of Logistics Systems and Management*, 12(3), 288–317. doi:10.1504/IJLSM.2012.047603

Paksoy, T., & Cavlak, E. B. (2011). Development and optimisation of a new linear programming model for production/distribution network of an edible vegetable oils manufacturer. *International Journal of Logistics Systems and Management*, 9(1), 1–21. doi:10.1504/IJLSM.2011.040057

Park, B.-N., & Min, H. (2013). Global supply chain barriers of foreign subsidiaries: The case of Korean expatriate manufacturers in China. *International Journal of Services and Operations Management*, 15(1), 67–78. doi:10.1504/IJSOM.2013.050562



- Pettersson, A. I., & Segerstedt, A. (2011). Performance measurements in supply chains within Swedish industry. *International Journal of Logistics Systems and Management*, 9(1), 69–88. doi:10.1504/IJLSM.2011.040060
- Pradhananga, R., Hanaoka, S., & Sattayaprasert, W. (2011). Optimisation model for hazardous material transport routing in Thailand. *International Journal of Logistics Systems and Management*, 9(1), 22–42. doi:10.1504/IJLSM.2011.040058
- Rajapakshe, T., Dawande, M., & Sriskandarajah, C. (2013). On the trade-off between remanufacturing and recycling. *International Journal of Services and Operations Management*, 15(1), 1–53. doi:10.1504/IJSOM.2013.050560
- Rajeev, V. (2013). Impact of service co-creation on performance of firms: The mediating role of market oriented strategies. *International Journal of Services and Operations Management*, 15(4), 449–466. doi:10.1504/IJSOM.2013.054885
- Samsung Smart, L. A. N. (2015.). Retrieved May 6, 2015, from http://www.samsung.com/us/system/b2b/resource/2015/03/29/150325_2015_Samsung_
- Smith, A. D. (2011). Component part quality assurance concerns and standards: Comparison of world-class manufacturers. *Benchmarking: An International Journal*, 18(1), 128–148. doi:10.1108/14635771111109850
- Sodhi, M., & Lee, S. (2007). An analysis of sources of risk in the consumer electronics industry. *The Journal of the Operational Research Society*, 64(1), 1430–1439. doi:10.1057/palgrave.jors.2602410
- Sprovieri, J. (2008). A Modest Increase. *Assembly*, 51(13), 22–41.
- Summers, G. J., & Scherpereel, C. M. (2008). Decision making in product development: Are you outside-in or inside-out? *Management Decision*, 46(9), 1299–1314. doi:10.1108/00251740810911957
- Tari, J. J., & Sabater, V. (2004). Quality tools and techniques: Are they necessary for quality management? *International Journal of Production Economics*, 92(3), 267–280. doi:10.1016/j.ijpe.2003.10.018
- Tiwari, A., Turner, C., & Sackett, P. (2007). A framework for implementing cost and quality practices within manufacturing. *Journal of Manufacturing Technology Management*, 18(6), 731–760. doi:10.1108/17410380710763886
- Varaprasad, G., Sridharan, R., & Unnithan, A. B. (2013). Internet banking adoption in a developing country - an empirical study. *International Journal of Services and Operations Management*, 15(1), 54–66. doi:10.1504/IJSOM.2013.050561
- Von Haartman, R. (2012). Beyond Fishers product-supply chain matrix: Illustrating the actual impact of technological maturity on supply chain design. *International Journal of Logistics Systems and Management*, 12(3), 318–333. doi:10.1504/IJLSM.2012.047604
- Yang, H., Choi, B., Park, H., Suh, M., & Chae, B. (2007). Supply chain management six sigma: A management innovation methodology at the Samsung Group. *Supply Chain Management: An International Journal*, 12(2), 88–95. doi:10.1108/13598540710737271

ADDITIONAL READING

- Carvalho, H., Cruz-Machado, V., & Tavares, J. G. (2012). A mapping framework for assessing supply chain resilience. *International Journal of Logistics Systems and Management*, 12(3), 354–373. doi:10.1504/IJLSM.2012.047606
- Hamidi, M., Farahmand, K., Sajjadi, S. R., & Nygard, K. E. (2012). A hybrid GRASP-tabu search metaheuristic for a four-layer location-routing problem. *International Journal of Logistics Systems and Management*, 12(3), 267–287. doi:10.1504/IJLSM.2012.047602

Haried, P., & Ramamurthy, K. K. (2009). Evaluating the success in international sourcing of information technology projects: The need for a relational client-vendor approach. *Project Management Journal*, 40(3), 56–71. doi:10.1002/pmj.20113

Kumar, P., Shankar, R., & Yadav, S. S. (2011). Global supplier selection and order allocation using FQFD and MOLP. *International Journal of Logistics Systems and Management*, 9(1), 43–68. doi:10.1504/IJLSM.2011.040059

Mathirajan, M., Manoj, K., & Ramachandran, V. (2011). A design of distribution network and development of efficient distribution policy. *International Journal of Logistics Systems and Supply Management*, 9(1), 108–137. doi:10.1504/IJLSM.2011.040062

More, D., & Babu, A. S. (2012). Benchmarking supply chain flexibility using data envelopment analysis. *International Journal of Logistics Systems and Management*, 12(3), 288–317. doi:10.1504/IJLSM.2012.047603

Paksoy, T., & Cavlak, E. B. (2011). Development and optimisation of a new linear programming model for production/distribution network of an edible vegetable oils manufacturer. *International Journal of Logistics Systems and Management*, 9(1), 1–21. doi:10.1504/IJLSM.2011.040057

Pettersson, A. I., & Segerstedt, A. (2011). Performance measurements in supply chains within Swedish industry. *International Journal of Logistics Systems and Management*, 9(1), 69–88. doi:10.1504/IJLSM.2011.040060

Von Haartman, R. (2012). Beyond Fishers product-supply chain matrix: Illustrating the actual impact of technological maturity on supply chain design. *International Journal of Logistics Systems and Management*, 12(3), 318–333. doi:10.1504/IJLSM.2012.047604

KEY TERMS AND DEFINITIONS



Automatic Identification and Data Capture Technologies (AIDC): Types of AIDC-related technologies to leave the human element out of the data collection and storage functions of information derived from manufacturing, integrated through the manufacturing process, type of authentication concerns, and/or e-security strategies, and relationship links to customer profiles. Typical types of AIDC include, bar-coding, RFID, magnetic strips, touch memory, and smart cards.

Internet of Things (IoT): IoT is a term that is applied to the general network of physical connected objects (e.g., services, vehicles, buildings and other items) that have mutually connected and enabled hardware, software, sensors, and network connectivity that allows communication and sharing of data and its protocols.

Operations Efficiency: Improving efficiency and reducing waste is a major challenge for hospitals and other patient care facilities looking to lower the cost of providing healthcare services. Far and away the largest contributor to operational costs in this industry is patient care activities. Since most clinical decisions involve managing products and medical supplies, finding ways to more efficiently manage supply chain activities can have a big impact on overall operational performance.

RFID-Embedded Technologies: RFID technologies are types of automatic data capture techniques that use a combination of active and passive senders and receivers to collect and store codified information for further uses. The implementation of such technologies should lead to improved managerial and/or supply chain performance. On the surface, there appears to be few drawbacks to implementing such technology into a production process, assuming it enhances performance and improves output of the product. The main issues surrounding the RFID applications are whether the initial costs and labor required to utilize this technology are worth it, and will result in a positive outcome of revenues.

Short Cycle Time and Low Inventory (SLIM) Methodology: In its electronics manufacturing processes, SLIM is a set of methodologies and scheduling applications for managing cycle time in semiconductor manufacturing. SLIM includes methodology for calculating target cycle times and target WIP levels for individual manufacturing steps, heuristic algorithms for factory floor scheduling, and optimization-based capacity analysis.

Supply Chain Management/Performance: In basic terms, supply chain is the system of organizations, people, activities, information and resources involved in moving a product or service

from supplier to customer. The configuration and management of supply chain operations is a key way companies obtain and maintain a competitive advantage. The typical manufacturing supply chain begins with raw material suppliers, or inputs. The next link in the chain is the manufacturing, or transformation step; followed the distribution, or localization step. Finally, the finished product or service is purchased by customers as outputs. Service and Manufacturing managers need to know the impact of supply on their organization's purchasing and logistics processes. However, supply chain performance and its metrics are difficult to develop and actually measure.

Simulating Complex Supply Chain Relationships Using Copulas



Krishnamurty Muralidhar

University of Oklahoma, USA

Rathindra Sarathy

Oklahoma State University, USA

INTRODUCTION

Supply chains involve complex interactions between different aspects of the research question and it is extremely difficult (if not impossible) to derive closed-form analytical solutions, leaving simulation as the only available alternative. The use of simulation in functional areas of business is well documented and includes accounting, finance, marketing, and other areas (Pegden et al., 1995). The analysis of supply chain systems is no exception. Recently, Waller and Fawcett (2012) observed that "...as simulation methods and technology are advanced, metaheuristics exist to find near optimal solutions, and powerful optimization is now in the hands of nearly any manager, let us not just model these simple systems, let us model the realistic, complex systems—in for a penny, in for a pound." The statement "When all else fails, there is simulation!" (Evers and Wan, 2012) is by no means an exaggeration.

The Council of Supply Chain Management Professionals defines supply chain management as encompassing "the planning and management of all activities involved in sourcing and procurement, conversion, and all logistics management activities. Importantly, it also includes coordination and collaboration with channel partners, which can be suppliers, intermediaries, third-party service providers, and customers. In essence, supply chain management integrates supply and demand management within and across companies." When modeling this complex system, simulation may be

the best alternative as the analytical tool (Waller and Fawcett, 2012). Swaminathan et al. (1998) also strongly argue for the use of simulation in supply chains.

This is because complex interactions between different entities and the multitiered structure of supply chains make it difficult to utilize closed-form analytical solutions. Benchmarking solutions provide insights into current trends but are not prescriptive. This leaves simulation as the only viable platform for detailed analysis for alternative solutions. Evers and Wan (2012) offer an excellent analogy in support of simulation. Griffis et al (2012), Rogers et al (2012), and Bartolacci et al (2012) also provide convincing evidence of the power of simulation as a methodological tool.

Constructing and modeling supply chains in practice, however, presents many challenges. Supply chains often involve "one or more families of related products", involving different entities or organizations that have different objectives, that are "highly interdependent when it comes to improving the performance of the supply chain in terms of objectives such as on-time delivery, quality assurance, and cost minimization." (Swaminathan et al., 1998). Hence, in order to effectively analyze supply chains using simulation, it is necessary to model the interdependence between different aspects of the model. Such interdependence may manifest itself in the demand for products, in the delivery of raw materials, in set up times, etc. Often, distribution of demand, delivery times, and set up times are not normally distributed, and must

be described using non-normal distributions such as the Gamma, Weibull, Beta, and other similar skewed distributions (Burgin, 1975; Wagner et al, 2009). Furthermore, it is possible that the demand for a given product has a certain marginal distribution (say Gamma), while that of another has a completely different distribution (say Weibull), but the two demand distributions are related. Thus, in order to analyze the performance and management of supply chains and similar complex systems, the ability to simulate random variates whose marginal distribution and dependence are specified is a necessary pre-requisite.

Currently, we do not have the ability even to completely specify such joint distributions, and the generation of such related random variables is extremely difficult. One possible approach for modeling the joint distribution of a set of variables with non-normal marginal distributions is to use copulas. The use of copulas for modeling complex supply chain relationships is not new. It has been previously used for modeling supplier default dependencies (Wagner et al, 2009) and to estimate the joint distribution of the disruption vector (Masihtehrani, 2011). The objective of this study is to demonstrate the viability of a method based on copulas for generating related random variates with specified marginal distributions and dependence structure for modeling supply chain inter-dependencies.

BACKGROUND

The ease with which random variates can be generated has improved considerably in recent years. This improvement can be attributed to improved hardware and software that is currently available (compared to just 15-20 years ago). Now, even standard desktop software such as Excel offer the ability to perform basic simulation without any other software. Software such as Crystal Ball and @Risk offer the ability to analyze complex decision analysis scenarios, and software such as Arena, GPSS, and ProModel offer the ability

to simulate complex systems. One of the crucial aspects of any simulation software is the ability to generate random variates from specified distributions in order to investigate the impact of changes in assumptions on the results of the analysis. Discussion on generating random variates can be found in practically all simulation textbooks. Readers interested in the technical details of algorithms are referred to the book by Fishman (1996), while those interested in the application of simulation are referred to the book by Law (2007).

While most textbooks offer a discussion of the generation of univariate forms of normal and non-normal random variates, discussion of multivariate random variates is invariably limited to the normal distribution. The generation of multivariate normal distribution is well understood and documented and many algorithms are available (see for example Fishman, 1996). Discussion of multivariate non-normal distributions, however, is another matter. Few algorithms for the generations of multivariate non-normal distributions are available due to the general complexity of non-normal multivariate distributions.

A few joint distributions for the bivariate Gamma and other non-normal distributions such as the Weibull and Exponential have been proposed (Kotz et al., 2000). Note that in these cases, it is assumed that the marginal distribution of the individual variables comes from the same family (such as Gamma, Weibull, etc.). Even then, these distributions are extremely complex. There are very few cases where the joint distribution of a set of variables whose marginal distributions are arbitrary (do not come from the same family of distributions) can even be specified. For example, consider a simple case where the marginal distribution of demand for a set of four products is given by the Gamma, the Uniform, the Normal, and Beta distributions, and that the demand for these four products are related. We currently do not even have the ability to specify the joint distribution of demand for the four products with the characteristics above. When the joint distribution is not specified, then it is impossible to derive



the conditional distribution or to generate related random variables for this specification.

One recent method for generating related non-normal random variates (Headrick and Sawilowsky, 1999) is based on the approach proposed by Fleishman (1978). This method generates a random variate from a normal distribution, and uses a transformation to create a new variate with a specified level of skewness and kurtosis. This method was later extended to generate multivariate non-normal random deviates (Headrick and Sawilowsky, 1999; Vale and Maurelli, 1983). However, this method is also not capable of producing all possible combinations of skewness and kurtosis and hence it is limited in its application. Further, the marginal distribution of random variates generated in this manner do not represent the true marginal distribution, and lack the ability to describe the probability density function (pdf) and cumulative density function (cdf). This implies that it would be impossible to derive the percentiles of the population from which these random variates had been generated (Tadikamalla, 1980).

In the context of operations management research, the inability to generate random variates from a specified marginal distribution that can be completely characterized by their pdf and cdf represents a serious shortcoming of this method. In many cases, the researcher may wish to change the parameters of the specified distribution to investigate the impact on the outcome. In addition, the inability to derive the percentiles for specified populations would make it very difficult to analyze issues such as service levels, etc. Hence, there is a need for methods that are capable of generating related random variates, with a specified dependence structure, and whose marginal distributions are non-normal (or a combination of normal and non-normal distributions). And in order for such a procedure to be useful, it must also be easy to implement. In the following section we describe the concept of copulas that can be used for this purpose.

The Copula Approach for Generating Random Variates

The copula approach is based on Sklar's theorem (Sklar, 1959) which states that for random variables $(A_1, A_2, \dots, A_k)$ with marginal cumulative distribution function (CDF) $F_i(a_i)$ ($i=1,2,\dots,k$), the joint CDF can be written as:

$$F(a_1, a_2, \dots, a_k) = C[F_1(a_1), F_2(a_2), \dots, F_k(a_k)] \quad (1)$$

where $C[F_1(a_1), F_2(a_2), \dots, F_k(a_k)]$ is a joint *copula* CDF with uniform marginal distributions. When $F_1(a_1), F_2(a_2), \dots, F_k(a_k)$ are continuous, the copula is also unique. In addition, the joint probability density function $f(a_1, a_2, \dots, a_k)$ can be written in product form as:

$$f(a_1, a_2, \dots, a_k) = \prod_{i=1}^k f_i(a_i) \times c[u_1, u_2, \dots, u_k] \quad (2)$$

where c is called the *copula* density, $f_i(a_i)$ are the marginal densities of A_i , and u_i is a uniform random variable. The joint density as shown in (2) also provides the ability to derive the conditional density of one or more of the random variables with respect to the other variables. For continuous distributions, the copula allows us to separately specify the marginal distribution of $(a_1, a_2, \dots, a_k)$ and the relationships between them, with the relationships being represented by the copula (Clemen and Reilly, 1999; Lindskog, 2000; Nelsen, 1991; Schweizer, 1991).

The primary applications of copulas have been to combine arbitrary marginal distributions into specified joint distributions that exhibit certain specified dependence or joint behavior. Nelsen (1995) and Schweizer (1991) serve as good introduction to the theory and application of copulas. A wide variety of copula functions have been investigated for combining non-normal distributions, both discrete and continuous. The selection

of a specific copula function depends on the specific problem under consideration. Some copula functions are well suited for describing bivariate distributions, but not for describing multivariate distributions. The specific characteristics of the joint distribution also vary depending on the specific type of copula selected. For our purpose, namely, generating random variates with specified marginal distributions and relationships, the multivariate normal copula is an effective approach.

The normal copula parameterized with product moment correlation matrix ρ can be written as:

$$C_{\rho}(u) = \phi_{\rho}^k \left(\phi^{-1}(u_1), \phi^{-1}(u_2), \dots, \phi^{-1}(u_k) \right) \quad (3)$$

where ϕ_{ρ}^k represents the joint CDF of a k -variate standard multivariate normal distribution with correlation matrix ρ , ϕ^{-1} represents the inverse of the CDF of the univariate standard normal function, and $(u_1, u_2, \dots, u_k)$ are uniform deviates. Now if we consider $(u_1, u_2, \dots, u_k)$ as:

$$[u_1 = F_1(a_1), u_2 = F_2(a_2), \dots, u_k = F_k(a_k)] \quad (4)$$

then the normal copula above can be re-written as:

$$C_{\rho}(u) = \phi_{\rho}^k \left(\phi^{-1}(F_1(a_1)), \phi^{-1}(F_2(a_2)), \dots, \phi^{-1}(F_k(a_k)) \right) \quad (5)$$

Thus, we are able to express the marginal distributions $F_i(a_i)$ of the original variables in the form of a multivariate normal copula. The next step in the process is to identify the relationship structure that results when the normal copula is used.

Traditionally, pair-wise product moment (or linear) correlations between the variables are commonly used to express relationships between random variables. The pervasive use of the linear correlation as a measure of relationship (or dependence) between variables can be attributed to its ease of use and the traditional assumption of multivariate normality. When the population is assumed to be multivariate normal, it is well

known that the relationship between the individual variables in the distribution is characterized by a linear relationship, and hence, the dependence structure between the variables is completely characterized by the pair-wise correlation matrix (Kotz et al., 2000). This however, is not true for multivariate non-normal distributions.

When the population is best described by a multivariate non-normal distribution the relationship between variables may be nonlinear in nature. Using linear correlation to measure the relationship between non-normal variables may result in an incorrect measure of the true dependence between the variables (Lancaster, 1982). In such cases, rank-based measures of dependence such as the Spearman's rank order correlation or Kendall's Tau can be used to measure dependence (Clemen and Reilly, 1999; Lancaster, 1982; Nelsen, 1991; Schweizer and Wolff, 1991). Unlike linear correlation, Spearman's rank order correlation and Kendall's Tau possess the desirable property that they are independent of the marginal distributions of the variables. These rank-based measures can be used as the index of dependence for many families of copulas (Clemen and Reilly, 1999). In this study we use Spearman's rank order correlation matrix to measure the relationship between the variables. The use of Kendall's Tau in place of Spearman's rank order correlation will provide similar results.

There is a direct relationship between linear correlation and Spearman's rank order correlation for the multivariate normal distribution. Let $R=(r_{ij})$ represent the $(k \times k)$ matrix of Spearman's rank order correlation and let $\rho=(\rho_{ij})$ represent the corresponding matrix of product moment correlation. Then, each element ρ_{ij} in ρ can be written as (Kruskal, 1958):

$$\rho_{ij} = 2 \text{Sin} \left(\frac{\pi r_{ij}}{6} \right) \quad (6)$$

Thus, if Spearman's rank order correlation matrix of the original variables is available, we can completely specify the multivariate normal

copula function described in (3). Once the copula function is specified, it becomes possible to generate related multivariate random variates with specified marginal distributions and dependence structure described by R(Clemen and Reilly 1999, Lindskog 2000).

The exact process of generating the multivariate random variables is as follows:

1. Identify the marginal distribution of $(A_1, A_2, \dots, A_k)$.
2. Construct $\mathbf{R}$, the Spearman's rank order correlation matrix of $(A_1, A_2, \dots, A_k)$.
3. Compute ρ using $\mathbf{R}$ as shown in equation (6).
4. Generate a row vector $\mathbf{x}$ of size k from a standard multivariate normal distribution with correlation structure ρ .
5. Convert each element in $\mathbf{x}$ into a uniform random deviate $u_i = \Phi(x_i)$ ($i=1, 2, \dots, k$).
6. Convert each u_i into the specified marginal distribution by using $F_i^{-1}(u_i)$ ($i = 1, 2, \dots, k$).
7. Repeat steps 4-6 for more values.

The values so generated will have marginal distributions $(F_1, F_2, \dots, F_k)$ respectively for $(A_1, A_2, \dots, A_k)$, the rank order correlation matrix specified in $\mathbf{R}$, and a joint distribution specified by the multivariate normal copula in (3). Note that each of the steps in the above process can now be performed by traditional methods: (1) Most simulation software (and Excel) provide the ability to compute the density and inverse of a diverse set of univariate marginal distributions and (2) The generation of the row vector $\mathbf{x}$ with specified correlation structure ρ can be performed by using Cholesky decomposition as described in Fishman (1986).

The copula based procedure described above is similar to the NORTA procedure that is described in Law (2007). The difference between the two procedures lies in the fact that the use of the Normal Copula approach provides a theoretical basis for specifying the dependence structure. One of the

difficult aspects of using the NORTA procedure is the ability to specify the correlation structure for the multivariate normal distribution that, when transformed, will result in the desired correlation structure for the transformed variables. In many cases, this may require extensive computation and solving non-linear equations of the order of $k \times k - 1$, where k represents the number of variables (Chen, 2001). By contrast, the copula based approach solves this problem by using rank order correlation and both the multivariate normal and transformed variables have the same rank order correlation, eliminating the need for the initial computations.

Finally, and perhaps most importantly, the copula based approach provides theoretical basis for describing multivariate data whose marginal distributions are not normal and exhibit certain dependence characteristics, while the NORTA procedure is intended only to generate random vectors. In this study, we have used the copula approach for generating random variables. In addition, it may be possible to use copulas for analyzing other situations in supply chains as well. Copulas have been used for modeling data from accidents (Yi and Bier, 1998), actuarial evaluations (Frees, et al. 1996), expert opinions (Jouini and Clemen, 1996), financial returns (Clemen and Reilly, 1999), and other situations. Hence, it is possible that in addition to providing a method for simulation related random variables, copulas may also offer the ability to model other complex situations in supply chains.

SOLUTIONS AND RECOMMENDATIONS

Illustrative Applications of the Copula Approach

In this section, we provide two illustrations of the application of the procedure based on multivariate normal copulas described above. To highlight



the ease of implementation, all computations were performed in Excel using Visual Basic on a standard laptop.

Generating Correlated Demand with Arbitrary Marginals

Assume that a researcher is interested in generating demand from a multivariate distribution with five variables, with the following characteristics:

- $A_1 \sim$ Gamma with shape parameter = 1 and scale parameter = 1 (Exponential);
- $A_2 \sim$ Gamma with shape parameter = 2 and scale parameter = 1;
- $A_3 \sim$ Beta with shape parameter = 2, scale parameter = 3;
- $A_4 \sim$ Beta with shape parameter = 3, scale parameter = 2;
- $A_5 \sim$ Normal distribution with mean = 0 and variance = 1.

These five distributions were selected to illustrate the ability of the copula-based procedure to generate widely varying marginal distributions. The first variable has a heavily (positively) skewed exponential distribution, the second variable comes from the same family as the first but has less skewness, the third variable comes from a different family (Beta) and is positively skewed (but the skewness is not as high as variable 1), the fourth variable also has a Beta distribution, but with negative skewness, and the last variable has a normal distribution. The generation of these vari-

ables, *individually and independently* is a simple matter. However, assume that the researcher wishes to generate random values that are correlated with rank order correlation matrix R provided in Table 1 (left panel). In this case, the researcher will have to use the copula based method that we have described in the previous section.

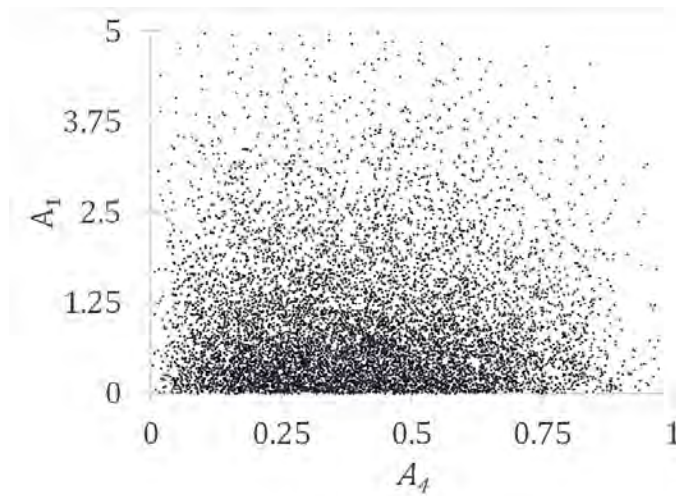
Using the parameters above, 10,000 values were generated in Excel using the copula procedure. Table 1 also provides the rank order correlation of the actual generated values (right panel). Comparing the two sets of correlations in Table 1 (specified and actual), it is easy to see that the copula procedure is able to maintain the rank order correlation of the generated values to be practically the same as the specified rank order correlations. Thus, the results of this illustration verify the fact that the copula-based procedure is capable of generating random variates with specified marginal distributions and specified relationships.

In the above analysis, we did not include the results when the variables are generated independently since, by definition, the variables would be uncorrelated. The following analysis shows the impact of the difference between the two procedures. Figure 1 provides a scatter plot of variables A_1 and A_4 generated independently and Figure 2 the scatter plot for the same variables generated using the copula approach. From Figure 2, it is evident that the copula based approach is very effective in preserving the strong positive relationship between the two variables; when the value of one variable is low, the value of the other variable is

Table 1. Specified and actual rank order correlation of demand data

Specified					Variable	Actual				
A_1	A_2	A_3	A_4	A_5		A_1	A_2	A_3	A_4	A_5
1.00					A_1	1.00				
0.50	1.00				A_2	0.51	1.00			
0.60	0.10	1.00			A_3	0.60	0.12	1.00		
0.80	0.20	0.70	1.00		A_4	0.81	0.21	0.70	1.00	
0.50	0.30	0.80	0.60	1.00	A_5	0.50	0.32	0.80	0.61	1.00

Figure 1. Scatter plots of variables 1 and 4 generated independently



also low and vice versa. But Figure 1 shows that if the values are generated independently, then no such relationship is exhibited; the value of one variable is independent of the other.

Table 2 provides the conditional probabilities of $(A_1|A_4)$ for the independent (left panel) and copula based (right panel) approaches. For comparison purposes, the middle panel provide the marginal probability of A_1 . As expected, when the variables are generated independently, the conditional probabilities of $(A_1|A_4)$ are almost identical to that of

the marginal probability of A_1 . By contrast, with the copula approach (right panel), we observe that when $A_4 < 0.25$, the probability of $A_1 > 2.5$ is practically zero. Similarly when $A_4 > 0.50$, the probability of $A_1 < 2.5$ is practically zero. When simulation supply chains, preserving these types of special relationships among variables that have a non-normal distribution can be critical to perform an accurate evaluation of the system. The copula based approach for generating related non-normal variable facilitates such evaluation.

Figure 2. Scatter plot of variables 1 and 4 generated using the Copula approach

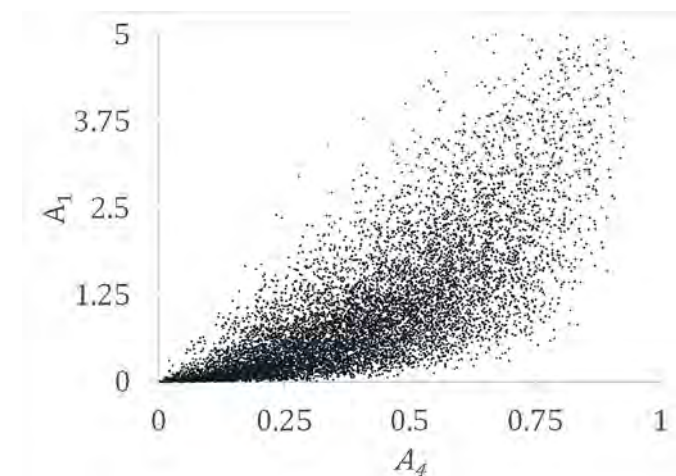


Table 2. Conditional probabilities of $(A_1|A_4)$

	Independent				Marginal A_1	Copula Based			
	$A_4 \leq 0.25$	$0.25 < A_4 \leq 0.50$	$0.50 < A_4 \leq 0.75$	$A_4 > 0.75$		$A_4 \leq 0.25$	$0.25 < A_4 \leq 0.50$	$0.50 < A_4 \leq 0.75$	$A_4 > 0.75$
$A_1 \leq 1.25$	72%	72%	71%	73%	72%	99%	83%	38%	5%
$1.25 < A_1 \leq 2.5$	20%	20%	20%	20%	20%	1%	16%	45%	28%
$2.5 < A_1 \leq 3.75$	6%	6%	6%	5%	6%	0%	1%	14%	34%
$A_1 > 3.75$	2%	2%	2%	2%	2%	0%	0%	3%	34%

Simulation of an Inventory System

In order to illustrate the importance of considering relationships between variables when modeling systems, we consider a simple inventory problem consisting of four products with related demand. Please note that it is not our objective in this illustration to evaluate the efficacy of the inventory policy, rather to show the difference in results that could occur when correlated demand is used compared to when uncorrelated (independent) demand is used. It is likely that in supply chain management research, the models will be more complex. The characteristics of the four products are provided in Table 3 as well as the specified rank order correlation. The following assumptions are made about the inventory system: (a) Inventory

review is performed every 10 days, (b) Delivery lead time is a constant 10 days, and (c) There is no safety stock.

The order quantity is determined as Order Quantity = (Expected Demand during Protection Interval – Inventory on Hand), where Protection Interval = Lead Time + Time between Reviews (Markland et al., 1998). The expected demand during lead-time was computed using the population mean values from which the variates were generated (see Table 2). The inventory at the beginning of the simulation was initialized at ten days expected demand (average demand during time between reviews).

The above system was simulated for a period of 10,000 days. Two sets of demand were generated, one using the copula-based related demand and

Table 3. Characteristics of product demand

Distributional Characteristics				
	Product 1	Product 2	Product 3	Product 4
Marginal Distribution	Gamma	Gamma	Gamma	Gamma
Shape Parameter	1.00	0.80	2.00	1.50
Scale Parameter	500	500	200	400
Mean	500	400	400	600
Standard Deviation	500	447	283	490
Coefficient of Variation	1	1.12	0.71	0.82
Rank Order Correlation				
	Product 1	Product 2	Product 3	Product 4
Product 1	1.00			
Product 2	0.80	1.00		
Product 3	0.70	0.75	1.00	
Product 4	0.90	0.80	0.80	1.00

another where the demand for each product was generated independent of the other products, but using the same population parameters described above. The focus of the analysis was on inventory shortages. Hence, on any day where there was a shortage, the ending inventory for all the products was recorded. Table 4 provides a summary of shortages by product and provides the number and percentage of days a given product was short for both independent and correlated demand. The results for both types of demand are very similar. This is to be anticipated since this summary provides the results of shortages based on *the marginal distribution* of the demand for individual products that is the same for both correlated and independent demand. However, as discussed below, relying on this information alone can be very misleading.

Table 5 provides a summary of shortages *by the number of products*. Table 5 shows the total number of days and the percentage of days in which shortages occurred by number of products (0 to 4). The first row of the table (zero products with shortage) provides the percentage of days that

there was no shortage. The difference between the two results is significant, with independent demand showing that no shortages for only 59% of the days, while correlated demand indicates that there were no shortages in 72% of the days. This large disparity can be attributed directly to the fact that correlated demand accounts for the inherent and strong relationship between the product demand patterns. Hence, when there is (no) shortage in one product, there is a high probability that the other products will also have (no) shortage. Hence, the total number of days in which there is a shortage of products is much less than independent demand would indicate. In this case, independent demand significantly over-estimates the number of days of shortage.

The differences in the individual probability of a stock-out for a given number of products are also quite different. For instance, the probability that all four products will be short on a given day is 7% according to correlated demand, while with independent demand the probability is only 1%. In this case, independent demand seriously under-estimates the probability of having shortage

Table 4. Summary of shortages by individual product

	Correlated Demand		Independent Demand	
	Number of Days	Percentage of Days	Number of Days	Percentage Days
Product 1	1827	18%	1759	18%
Product 2	1876	19%	1713	17%
Product 3	1335	13%	1355	14%
Product 4	1627	16%	1649	16%

Table 5. Summary of shortages by number of products

Number of Products Exhibiting Shortage	Correlated Demand		Independent Demand	
	Number of Days	Percentage Days	Number of Days	Percentage of Days
0	7207	72%	5888	59%
1	940	9%	2365	24%
2	556	6%	1217	12%
3	575	6%	443	4%
4	722	7%	87	1%

in all products. In both cases, using the incorrect demand structure when simulating the inventory system leads to incorrect estimation of shortages. The results provided in Table 5 for correlated demand are more representative of the true nature of the system. Without using correlated demand, the conclusions that would be reached regarding the system would be quite different. These differences in the results have important implications for inventory policy. Thus, the use of the copula method for generating related random variates with arbitrary marginal distributions allows the decision maker to model the system more accurately.

IMPLICATIONS FOR PRACTICE AND RESEARCH

The copula-based method for generating related multivariate non-normal random deviates using copulas provides the decision maker the ability to specify (a) The marginal distributions of each of the random variables, and (b) The relationship between them (expressed as rank order correlations). The ability to specify each of these independently and explicitly provides the decision maker with practically unlimited flexibility in evaluating the impact of changes to the system parameters on outcomes. For example, it allows the decision maker to investigate the impact of changing the demand distribution of one product from (say) Gamma to (say) Lognormal, as well as to change the parameters of these distributions. In practice, this will facilitate the analysis of related components accurately and evaluate the sensitivity of the output to distributional assumptions.

It should be noted that in order to use the methodology proposed in this study, it is necessary that the relationship between the variables be captured adequately by the pair-wise rank order correlation matrix. Such a requirement is generally satisfied by random variables that have monotonic (strictly increasing or decreasing) relationships. However, if the relationship between two vari-

ables is non-monotonic (U-shaped or $\cap$ -shaped, for example), then neither rank order nor product moment correlation will accurately capture such a relationship (Clemen and Reilly, 1999). Under such conditions, it may be necessary to consider either transforming the variables in question or consider alternatives to the copula.

FUTURE RESEARCH DIRECTIONS

In this study, we have used the normal (or Gaussian) copula to generate the random variates. The normal copula is perhaps the best known and the easiest to implement of the alternative copula models available. But alternative forms of copulas may be better suited for modeling related variables in some scenarios (Wagner et al 2009). The investigation of the implementation of alternative copula models in a simulation setting presents an opportunity for future research.

CONCLUSION

Supply chain systems are more complex than the traditional manufacturing systems since they typically involve multiple products, vendors, production lines, and locations. These multiple products may have related demands, the multiple vendors may have related lead times, and the different production lines may have related set up times. And more importantly, most of these quantities also tend to have skewed marginal distributions. Hence, in order to accurately model these systems, it is necessary to have the ability to generate related random deviates whose marginal distributions may be non-normal. In this study, we propose a simple, effective, and easy to implement procedure based on the multivariate normal copula for generating related multivariate non-normal random deviates to effectively model, analyze, and evaluate supply chain systems with complex relationship among variables using simulation.

REFERENCES

- Bartolacci, M. R., LeBlanc, L. J., Kayikci, Y., & Grossman, T. A. (2012). Optimization Modeling for Logistics: Options and Implementations. *Journal of Business Logistics*, 33(2), 118–127. doi:10.1111/j.0000-0000.2012.01044.x
- Burgin, T. A. (1975). The Gamma Distribution and Inventory Control. *Operational Research Quarterly*, 23(3), 507–525.
- Chen, H. (2001). Initialization for NORTA: Generation of Random Vectors with Specified Marginals and Correlations. *INFORMS Journal on Computing*, 13(4), 312–331. doi:10.1287/ijoc.13.4.312.9736
- Clemen, R. T., & Reilly, T. (1999). Correlations and Copulas for Decision and Risk Analysis. *Management Science*, 45(2), 208–224. doi:10.1287/mnsc.45.2.208
- Evers, P. T., & Wan, X. (2012). Systems Analysis Using Simulation. *Journal of Business Logistics*, 33(2), 80–89. doi:10.1111/j.0000-0000.2012.01041.x
- Fishman, G. S. (1996). *Monte Carlo: Concepts, Algorithms, and Applications*. New York: Springer-Verlag. doi:10.1007/978-1-4757-2553-7
- Griffis, S. E., Bell, J. E., & Closs, D. J. (2012). Metaheuristics in Logistics and Supply Chain Management. *Journal of Business Logistics*, 33(2), 90–106. doi:10.1111/j.0000-0000.2012.01042.x
- Headrick, T. C., & Sawilowsky, S. S. (1999). Simulating Correlated Multivariate Nonnormal Distributions: Extending the Fleishman Power Method. *Psychometrika*, 64(1), 25–35. doi:10.1007/BF02294317
- Jouini, M. N., & Clemen, R. T. (1996). Copula Models for Aggregating Expert Opinions. *Operations Research*, 44(3), 444–457. doi:10.1287/opre.44.3.444
- Kotz, S., Balakrishnan, N., & Johnson, N. L. (2000). Continuous Multivariate Distributions: Vol. 1. Models and Applications. John Wiley & Sons, Inc.
- Kruskal, W. (1958). Ordinal Measures of Association. *Journal of the American Statistical Association*, 53(4), 814–861. doi:10.1080/01621459.1958.10501481
- Lancaster, H. O. (1982). Dependence, Measures and Indices of. In *Encyclopedia of Statistical Sciences* (Vol. 2). New York: John Wiley & Sons.
- Law, A. M. (2007). *Simulation Modeling and Analysis*. New York: McGraw Hill.
- Lindskog, F. (2000). *Modelling Dependence with Copulas* (Master's Thesis). Swiss Federal Institute of Technology, Zurich, Switzerland.
- Markland, R. E., Vickery, S. K., & Davis, R. A. (1998). *Operations Management: Concepts in Manufacturing and Services*. Cincinnati, OH: South-Western College Publishing.
- Masihtehrani, B. (2011). *Stochastic Analysis of Disruption in Supply Chain* (Dissertation). Penn State University, State College, PA.
- Nelsen, R. B. (1991). Copulas and Association. In G. Dall'Aglio, S. Kotz, & G. Salinetti (Eds.), *Advances in Probability Distributions with Given Marginals* (pp. 51–74). Dordrecht, Netherlands: Kluwer. doi:10.1007/978-94-011-3466-8_3
- Nelsen, R. B. (1995). Copulas, Characterization, Correlation and Counterexamples. *Mathematics Magazine*, 68(1), 193–198. doi:10.2307/2691415
- Pegden, C. D., Shannon, R. E., & Sadowski, R. P. (1995). *Introduction to Simulation using SIMAN*. New York: McGraw Hill Inc.
- Rogers, D. S., Melamed, B., & Lembke, R. S. (2012). Modeling and Analysis of Reverse Logistics. *Journal of Business Logistics*, 33(2), 107–117. doi:10.1111/j.0000-0000.2012.01043.x

- Schweizer, B. (1991). Thirty Years of Copulas. In G. Dall'Aglio, S. Kotz, & G. Salinetti (Eds.), *Advances in Probability Distributions with Given Marginals* (pp. 13–50). Dordrecht, Netherlands: Kluwer. doi:10.1007/978-94-011-3466-8_2
- Schweizer, B., & Wolff, E. F. (1981). On Non-parametric Measures of Dependence for Random Variables. *Annals of Statistics*, 9(4), 879–885. doi:10.1214/aos/1176345528
- Sklar, A. (1959). Fonctions de Répartition à n dimensions et Leurs Mages. *Publications de l'Institut Statistique de l'Université de Paris*, 8, 229–231.
- Swaminathan, J. M., Smith, S. F., & Sadeh, N. M. (1998). Modeling Supply Chain Dynamics: A Multiagent Approach. *Decision Sciences*, 29(3), 607–632. doi:10.1111/j.1540-5915.1998.tb01356.x
- Tadikamalla, P. R. (1980). On Simulating Non-normal Distributions. *Psychometrika*, 45(2), 273–279. doi:10.1007/BF02294081
- Vale, D. C., & Maurelli, C. A. (1983). Simulating Multivariate Nonnormal Distributions. *Psychometrika*, 48(3), 465–471. doi:10.1007/BF02293687
- Wagner, S. M., Bode, C., & Koziol, P. (2009). Supplier Default Dependencies: Empirical Evidence from the Automotive Industry. *European Journal of Operational Research*, 199(1), 150–161. doi:10.1016/j.ejor.2008.11.012
- Waller, M. A., & Fawcett, S. E. (2012). Mathematical Modeling in Logistics: In for a Penny, in for a Pound. *Journal of Business Logistics*, 33(2), 78–79. doi:10.1111/j.0000-0000.2012.01040.x
- Waller, M. A., & Fawcett, S. E. (2013). Data Science, Predictive Analytics, and Big Data: A Revolution That Will Transform Supply Chain Design and Management. *Journal of Business Logistics*, 34(2), 77–84. doi:10.1111/jbl.12010

ADDITIONAL READING

- Joe, H. (2015). *Dependence Modeling with Copulas*. Boca Raton, FL: CRC Press.
- Nelsen, R. B. (1999). *An Introduction to Copulas*. New York: Springer-Verlag. doi:10.1007/978-1-4757-3076-0

Using RFID and Barcode Technologies to Improve Operations Efficiency Within the Supply Chain



Amber A. Smith-Ditizio
Texas Woman's University, USA

Alan D. Smith
Robert Morris University, USA

INTRODUCTION

Management Throughout the Value Chain

Managers in today's business environment find themselves not only focused on internal processes such as production efficiency and employee relations, but also acknowledging the need to increasingly focus on management throughout the value chain. A number of researchers (Anderson & Dekker, 2009; Smith & Minutolo, 2014; Smith & Offodile, 2007; Wee, Peng, & Wee, 2010; Whitten, 2004) have identified some of the complex components of the supply chain to include global suppliers, contract manufacturers, company-owned product/service centers, third-party logistics providers, and transportation providers. Each relationship in the supply chain poses unique management opportunities and challenges. It is important for firms to understand the existing tools and strategies in the operations management field as well as the developing technologies and practices to better control costs and enhance efficiency in the supply chain. These same researchers frequently cite that the management of many companies are not operationally efficient (e.g., it can be seen from the cost of production is larger than needed, generating waste and non-value activities, and lack of vendor collaboration and integration, to name a few related issues). Proper management of their supply chains can

dramatically improve such efficiencies. In the following sections, a number of studies suggest that the use of RFID in supply chain management (SCM) can improve operational efficiency. This chapter briefly discusses the contribution of knowledge regarding the use of RFID in the operational efficiency of the company.

In order to stay competitive in the marketplace, firms must do whatever is necessary to gain an edge and create value for their product. The desired end goal is to get the right product to the right customer at the right price at the right time; to accomplish this chain of events firms must search for methods to achieve a competitive edge. In this increasingly global economy, competition for buyers can be cut-throat because of the easy access to the international market and global materials because of the internet. Also, because of the need for speed, space issues, and timing, supply chain agreements frequently expect the raw materials to arrive at the right warehouse or facility at the right time, not a week before or later. Needing to stop production because a part is missing is disastrously expensive; also loss of goods or theft is frequently a concern with companies. So the issue is raised, how can a company make improvements in the supply chain to be more efficient? Value is created for the client when the item arrives to the client on the right day. When an item is late or there are stock-outs there is the risk of the bullwhip effect occurring, when the client over-orders because they fear they might have

DOI: 10.4018/978-1-5225-2255-3.ch486

their order rationed. By tracking the orders and products with barcode and/or RFID technologies (Ha, Park, Lee, & Park, 2013; Kwok & Wu, 2009; Mateen & More, 2013; Mehrjerdi, 2009; Roberti, 2014), operational management have the enhanced ability to notice if there are inefficiencies within the organization and address them quickly, and if the capability allows, the client can see that they will receive their order exactly when they wished in the quantity that they desired. This technology can add value all throughout the supply chain, to include the warehouse, distribution, and returns management areas.

BACKGROUND

Importance of Supply Chain Management (SCM)

Supply chain management (SCM) is one of the critical decision areas that a firm must focus on in order to compete in the ever-changing world of business. As a number of authors (Berthon, Critenden, Desautels, & Pitt, 2010; Bhat, 2008; Biswas & Sarker, 2008; Browning & Heath, 2009; Cabral & Cruz-Machado, 2012; Cavaleri, 2008; Carvalho, Cruz-Machado, & Tavares, 2012; Chan & Kumar, 2009; Chongwatpol & Sharda, 2013) have pointed out, lean firms are charged with minimizing costs and increasing efficiencies from the beginning to the end of the value chain. Firms must take a much broader view of operations management that extends far beyond their factory or office walls. Decisions involving the entire value chain must be made and can be extremely difficult. Firms that are able to successfully manage the key areas of sourcing, supplier selection, and the design of supply relationships will often realize a competitive advantage. Research and observation has revealed that this advantage can come in the form of low costs, high productivity, quality, customer responsiveness, and innovation (Park & Min, 2013; Smith, 2010, 2015).

The various components of SCM will determine the cost structure relating to upstream operations and play a large role in risk mitigation. Supplier selection is one of the core decisions of structural cost management. Firms must decide whether these upstream operations should be executed internally or whether they should be outsourced to a supplier. The article makes reference to the “make-buy-or-ally” decision that management faces. A *make* decision would entail a firm integrating vertically and obtaining components from internal units. A *buy* decision is the decision to purchase inputs from another firm in an “arms-length” contract. An *ally* decision is a mixture of the two previous decisions whereby a buyer and seller remain separate entities but collaborate in a mutually-beneficial, long-term relationship. The appropriate supplier choice would be the option that minimizes production and transaction costs (Drejer & Riis, 2000).

Designing the buyer/supplier relationship involves contractual management controls that govern aspects such as specifying authority to make supply decisions, performance requirements, and rewards for achieving performance goals or sanctions for non-performance (Boerner, 2010; Hu, Wang, Fetch, & Bidanda, 2008; Jain, Benyoucef, & Deshmukh, 2008; Kennedy & Widener, 2008). In addition, informal controls such as performance feedback processes, joint budgeting/forecasting processes, and co-location of employees are developed to enhance the buyer/supplier relationship. Formal legal contracts as well as behavioral norms are established to improve the efficiency and effectiveness of the relationships. Enhanced uses of automatic identification and tracking technologies help reduce error and potential litigation issues in the supply chain.

Purpose of Present Chapter

This chapter will look at the impact of using RFID-embedded systems within a lean supply chain environment via a qualitative business case

study approach (Baxter & Jack, 2008). Managers need to be able to justify an augmentation of lean practice and real-time data with the use of RFID to improve supply chain management (SCM) and its ability to improve the goals of reducing waste and improving product visibility throughout the supply chain. By using a qualitative business case study approach, four local NE Ohio companies will be addressed to inspect how to make their supply chains more efficient with the use of RFID and barcode technologies, common examples of automatic identification and data capture (AIDC) technologies. Some of the companies have global implications, while others are more local in nature, but all have superior corporate reputations in NE Ohio; as well as extensively use RFID and barcode technologies to make their businesses more competitive and reduce time, materials, and labor waste. This chapter also discusses how these regionally headquartered companies strategically and operationally leverage their RFID-embedded technologies to create value for other organizations.

CASE 1: GOODYEAR TIRE AND RUBBER COMPANY

The Goodyear Tire and Rubber Company is headquartered in Akron, Ohio and is one of the world's largest tire companies. The company employs approximately 73,000 people, manufactures products in 53 facilities in 22 countries around the world. Supply chain efficiency has been a stated goal of any manufacturing based company, but when the company is relatively large and global in its reach, spanning many countries, supply chain efficiency becomes even more necessary.

Goodyear has been experimenting with RFID-related technologies since 1984, and began their first field trial with 3000 tires ("Goodyear's RFID technology ...," 2014). Since then they have improved upon these trials, embedding their patented RFID tags in their tires to eliminate errors in identification of tires in the forward and reverse

supply chain. Management at Goodyear generally acknowledges that RFID has a multitude of uses in the supply chain, to include tire production, warehousing, and sales and service. Goodyear is somewhat unique in that it is important that the process be fast, efficient, and seamless as possible in sending the tires to the right group, with the ability for the group to return the tires quickly. The company traditionally leases its tires to the racing participants; this leasing program has several benefits: The overall cost of the tire is lowered because old tires can be refurbished to be like-new, and there is less environmental impact in a continued buildup of discarded tires. RFID aids in this process, as the chip is embedded within the tire so that even if the tire is scuffed to the point that the numbers on the sidewall are unreadable, the chip eliminates any chance of user error in identifying a tire for refurbishment. At the beginning and end of each race, thousands of tires need to be moved from the facility to the race and back again, RFID ensures that each team on each end receives the right tires.

What started out as a method of tire identification, it expanded to include information about specific tire data to improve vehicle performance. In 2012, Goodyear explained company plans of making tires more intelligent to the point that they are feeding back real-time information to the driver. This feature added value to the tires for the end-product user, creating a need for the product, creating more demand for these services in the supply chain. An accurate log of each tire's activity was hard to promise prior to RFID, now the company can without a doubt state the chain of each tire's life.

Goodyear has used this technology to gain a competitive edge in the truck market. Using RFID in tires, the tires can eliminate the need to read sidewall markings on tires, instead using a scanner the information can be read and instantly recorded. If stolen, the tires can be traced easier and tire management and tire security is improved. Goodyear and the truck fleet owner can link into the Goodyear FleetOnlineSolutions™

fleet management system to easily locate service providers if needed, the service providers can see contract details and tire information easily. The fleet management system takes care of invoicing and stock control and ensures that customers are charged according to their agreements and are invoiced accordingly.

CASE 2: SHERWIN-WILLIAMS, INC.

Sherwin-Williams, headquartered in Cleveland, OH, was receiving feedback from its clients that bar codes were needed to help with inventory. In order to maintain these relationships, Sherwin-Williams implemented a custom print-apply bar code labeling system (“Are bar code labels a ...,” 2016). To increase efficiency in the supply chain, retailers needed bar code labels on the cartons the products were shipped in. This ensures a quick scan with a reader to be sure the carton goes to the right place, without the need to slow down the process with a manual read. The process is complicated for the company, at least in part due to the condition of carrying the products different retailers were demanding different information on the labels depending on what was needed to make their supply chains more efficient. These bar codes include product information, description codes, part numbers and more for each of the customer’s requirements, which was necessary to move products through the that retailer’s warehouse and distribution systems quickly and easily.

The company has orders that are not consistent so it can be a challenge trying to balance everything to ensure that long wait times aren’t happening, stock-outs weren’t occurring, and product quality remained high. Typically, orders would range from as few as 2000 cans to 300,000 cans or more. To comply with the retailer’s requirements Sherwin-Williams had to install parts of the bar-code labeling system on five lines of its manufacturing facility, with a linked PC network that could access all of the print/apply systems simultaneously.

In order to make sure the right information is on the barcode, at each of the five stations of the facility an operator enters the order number, which tells the manufacturing line what information goes into that manufacturing line. The linking back to the main computer systems, both fixed and mobile devices, ensures the right product is used at the right time, with the batch and data numbers to link back to the right process, and all of this information is attached to the barcode that is printed and placed on the cartons.

CASE 3: SURGERE COMPANY

Surgere is a small business based out of Green, OOH that focuses on supply chain optimization by improving and reducing packaging, and using efficient packaging to make the supply chain more efficient. Surgere’s clients include Best Buy, Timken, and Whirlpool. Surgere uses RFID-related technologies that are linked to computer programs to log and track packages, using real time information to know where a package is and be able to tell when a package is delayed or missing, if applicable (“SRFID - container tracking and management ...,” 2016). Knowing about issues occurring simultaneously allows management the means to handle most issues quickly and be proactive in order to not waste additional time. In the background, the linked software collects and organizes the data to interpret and dwell on turn times and analyze data in order to work towards full fleet utilization. The ongoing value of RFID-embedded systems includes asset management, loss prevention, improved productivity, analytics, and cost reduction.

RFID-related technologies usually constitute improvements on barcode technology, primarily because instead of an employee scanning one barcode at a time, RFID readers can quickly process hundreds of tags at a time. Wireless transmission of Surgere’s tag information logs and tracks information and provides real time information about shipment details. The reusable containers and

racks reduce waste and can be tracked throughout the supply chain process so that management can see the trends in the supply chain and look for opportunities to improve. By automating the process and adding RFID tags, using electronic product codes (EPCs), allowing universal identifiers to give unique identities to items with RFID tags; ultimately, productivity is improved and delays may be avoided.

CASE 4: THE KENNEDY GROUP

The Kennedy Group is located in Willoughby, OH and is noted as a supply chain optimization company that specializes in RFID technology. The Kennedy Group has many different types of tags depending on the nature of the business, to include multi-use or single-use tags. Different industries have different needs, such as a bakery likely needs reusable trays and racks because those have a high costs and a 30% loss rate. Higher cost items likely need constant monitoring to be sure that the items are exactly located in the supply chain where they should be. Retail food distribution has unique needs, such as a need to link a product back to a shipping container should a recall happen in the future. The Kennedy Group can aid in RFID-related implementation needs, especially in the industries of bakery, retail food distribution, manufacturing, industrial, and warehouse and logistics management. Food distribution sensors may need the ability to track temperature of a package, whereas warehouse and logistics management needs the ability for forklift reader integration. A firm only wishes to pay for the services that the specific firm needs, there is no value in paying for additional unnecessary technology that would add too much cost onto the product; a classic value-added proposition common in business operations.

MANAGERIAL IMPLICATIONS



Related AIDC-based technologies, most notably RFID and barcode technologies are invaluable in the operation and supervision of a company's supply chain. However, management typically has a difficult choice for a firm's management is trying to decide the system that will create the most value for the organization. As noted above, Sherwin-Williams was successful using barcodes and barcode readers to track the products throughout the supply chain. Goodyear decided to go with a more advanced system of RFID, likely because the cost of tires, especially racing tires, is much greater than the cost of cans of paint. The Kennedy Group and Surgere are two companies that can meet the needs of the companies using RFID or even barcode technology.

The basic cost/benefit issue is, then, how should a firm decide which method is best? Barcodes are universally accepted, simple, and low cost. Scanners are widely available and affordable. Barcodes are everywhere in the supply chain, but due to the physical dimensions there may be limitations on the amount of information that can be placed on a barcode. For retail applications, a barcode may only note the product type and the manufacturer.

Barcodes require a direct line of sight to be read, and can only be read one at a time which can be labor intensive. The barcode must be on the outside of the package to be read, which means there is the risk that the barcode becomes damaged and unreadable. Barcodes are easily counterfeited, but they do come at a very low cost, they can be made for a fraction of a cent or even a few pennies. Barcodes are best for closed loop supply chains and process manufacturing such as Sherwin-Williams where liquids and metals are used. Liquids and metals cause interference with certain radio frequencies, so RFID may not be effective for these industries.

RFID is more expensive yet has dynamic data that can read, write, update and even ac-

tivate the next step in a chain of events. RFID can track lifespans of a product or even track temperatures of products. Some RFID units can be reprogrammed for repeated use, or they may be single use only. Active data RFID include a battery and will automatically send data back to the programs; passive RFID is without a battery and only submits information when the product is within range of a reader. Unlike barcode readers that require a line of sight, RFID readers only require being within 20-40ft. and the readers can read through packages, therefore reading hundreds of chips at a time (Chongwatpol & Sharda, 2013; Delen, Hardgrave, & Sharda, 2007). Active RFID tags are expensive, ranging from US\$5-50 per unit. This is the option that Goodyear has chosen for the racing tires because of the need of being able to constantly access the tires information and location. Single trip and multi trip passive RFID chips are cheaper, ranging from 10 cents each to US\$4 ("RFID solutions," 2016). These are likely the options chosen for the majority of items at retailers in order to keep costs down, yet still have enough visibility and tracking options in order to view the supply chain process and look for areas of opportunity. Primarily due to the ease of readability, RFID is often the preferred choice in a complex supply chain, especially one with high volume transactions. Especially for organizations looking to track products at the unique item level, RFID outperforms barcodes especially when considering labor costs, loss, and theft.

Management of each firm must consider their unique needs and react accordingly. A supply chain management tracking system that Sherwin-Williams uses would not be effective for Goodyear, which demonstrates why each company reacts differently for their own supply chain management. A major issue is always cost, because paint is a product that has a high amount of competition, Sherwin-Williams has to be sure that any method that they use in their supply chain will be able to be absorbed into the cost of the product or with low additional added cost. For racing tires, the value isn't completely in the price though cost in a minor

factor, the value for the racing consumer is added benefits. Both companies likely wish to monitor the location of the product in the supply chain, but Goodyear's clients would likely be willing to pay an additional price for RFID tracking on the field, which Goodyear can also use this RFID technology to track the tires within the supply chain. As mentioned, Goodyear's services are unique in that there is a lot of forward and reverse logistics occurring, because of the increased risk of loss and theft an active RFID tag is almost necessary to keep the process flowing smoothly.

FUTURE RESEARCH DIRECTIONS

Improving efficiency in the supply chain by using bar code and/or RFID technology is a valuable option for all companies, although some smaller companies may prefer to continue to manually track their operations within the supply chain. Typically, suppliers track their products using mainly barcodes, but that is where the tracking stops until the USPS tracking begins. Intermediaries or warehouses repackage these products in customized fashion to suit the individual needs of retailers and/or manufacturers. Within each firm's supply chain, much of the inventory control is manually in smaller firms and done automatically to some degree in larger firms. Obviously, there is usually some level of confusion as items flow from the ordering phase, to the processing phase, to the shipment phase, to the receiving phase by the ultimate buyer. Both large and small organizations must decide if investing in barcode and/or RFID technology internally will make the efficiency/profit tradeoff a viable option. The case study described in this chapter, hopefully, demonstrates a company's efficiency through increased productivity, reduced failure, profit, as well as other types of improvement. Ultimately, all organizations need a better system than the clipboards and filing cabinet systems as the products flow through the various phases of the firm's supply chain.

CONCLUSION

Manual tracking of goods throughout the supply chain is rarely effective except in very small businesses. As the business grows larger, barcode and/or RFID technology is needed to reduce the possibility of stock-outs, the bullwhip effect, reduce loss, theft, and various other complications of the supply chain. Due to the various costs involved, a business needs to evaluate what system would work best in order to make their own supply chain effective, and then use the services of The Kennedy Group or Surgere to improve the flow of goods in the organization. In the last few years, more and more companies are integrating RFID or barcode technology into their planning, since it has been proven to significantly impact supply chain performance. Substantial benefits can be gained by implementing RFID-embedded technologies to improve identification and improve the movement of products, shipments, and assets.

REFERENCES

- Anderson, S., & Dekker, H. (2009). Strategic cost management in supply chains, Part 1: Structural cost management. *Accounting Horizons*, 23(2), 201–220. doi:10.2308/acch.2009.23.2.201
- Are bar code labels a valuable customer service? Ask Sherwin-Williams. (2016). Retrieved March 18, 2016 from <http://www.weberpackaging.com/case-study-sherwin-williams.html>
- Baxter, L. F., & Hirschhauser, C. (2004). Reification and representation in the implementation of quality improvement programmes. *International Journal of Operations & Production Management*, 24(2), 207–224. doi:10.1108/01443570410514894
- Baxter, P., & Jack, S. (2008). Qualitative case study methodology: Study design and implementation for novice researchers. *Qualitative Report*, 13(4), 544–559.
- Beamon, B. M. (1999). Designing the green supply chain. *Logistics Information Management*, 12(4), 332–342. doi:10.1108/09576059910284159
- Berthon, P., Critenden, V., Desautels, P., & Pitt, L. (2010). Get the most out of green IT. *Industrial Management (Des Plaines)*, 52(5), 14–18.
- Bhat, S. (2008). The effect of ordering policies for a manufacturing cell changing to lean production. *Proceedings - Institution of Mechanical Engineers*, 222(11B11), 1551–1560. doi:10.1243/09544054JEM1216
- Biswas, P., & Sarker, B. R. (2008). Optimal batch quantity models for a lean production system with in-cycle rework and scrap. *International Journal of Production Research*, 46(23), 6585–6610. doi:10.1080/00207540802230330
- Boerner, H. (2010). Sustainability rises to top of strategy-setting for growing number of corporate leaders. *Corporate Finance Review*, 15(1), 32–34.
- Browning, T. R., & Heath, R. D. (2009). Reconceptualizing the effects of lean on production costs with evidence from the F-22 program. *Journal of Operations Management*, 27(1), 23–35. doi:10.1016/j.jom.2008.03.009
- Cabral, I. G. A., & Cruz-Machado, V. (2012). A decision-making model for Lean, Agile, Resilient and Green supply chain management. *International Journal of Production Research*, 50(17), 4830–4845. doi:10.1080/00207543.2012.657970
- Carvalho, H., Cruz-Machado, V., & Tavares, J. G. (2012). A mapping framework for assessing supply chain resilience. *International Journal of Logistics Systems and Management*, 12(3), 354–373. doi:10.1504/IJLSM.2012.047606
- Cavaleri, S. A. (2008). Are learning organizations pragmatic? *The Learning Organization*, 15(6), 474–481. doi:10.1108/09696470810907383



- Chan, F. T. S., & Kumar, V. (2009). Performance optimization of a legality inspired supply chain model: A CFGTSA algorithm based approach. *International Journal of Production Research*, 47(3), 777–791. doi:10.1080/00207540600844068
- Chongwatpol, J., & Sharda, R. (2013). Achieving lean objectives through RFID: A simulation-based assessment. *Decision Sciences*, 44(2), 239–266. doi:10.1111/deci.12007
- Delen, D., Hardgrave, B. C., & Sharda, R. (2007). RFID for better supply-chain management through enhanced information visibility. *Production and Operations Management*, 16(5), 613–624. doi:10.1111/j.1937-5956.2007.tb00284.x
- Drejer, A., & Riis, J. O. (2000). New dimensions of competence development in industrial enterprises. *International Journal of Manufacturing Technology and Management*, 2(1/7), 660–882.
- Goodyear's RFID technology has high-profile success. (2014). Retrieved December 2, 2015, from <http://www.prnewswire.com/news-releases/goodyears-rfid-technology-has-high-profile-success-55708912.html>
- Grewal, C. (2008). An initiative to implement lean manufacturing using value stream mapping in a small company. *International Journal of Manufacturing Technology and Management*, 15(3/4), 404–421. doi:10.1504/IJMTM.2008.020176
- Ha, O., Park, M., Lee, K., & Park, D. (2013). RFID application in the food-beverage industry: Identifying decision making factors and evaluating SCM efficiency. *KSCE Journal of Civil Engineering*, 17(7), 1773–1781. doi:10.1007/s12205-013-0297-x
- Hu, G., Wang, L., Fetch, S., & Bidanda, B. (2008). A multi-objective model for project portfolio selection to implement lean and Six Sigma concepts. *International Journal of Production Research*, 46(23), 6611–6648. doi:10.1080/00207540802230363
- Jain, V., Benyoucef, L., & Deshmukh, S. G. (2008). Whats the buzz about moving from lean to agile integrated supply chains? A fuzzy intelligent agent-based approach. *International Journal of Production Research*, 46(23), 6649–6678. doi:10.1080/00207540802230462
- Kennedy, F. A., & Widener, S. K. (2008). A control framework: Insights from evidence on lean accounting. *Management Accounting Research*, 19(4), 301–319. doi:10.1016/j.mar.2008.01.001
- Kwok, S., & Wu, K. W. (2009). RFID-based intra-supply chain in textile industry. *Industrial Management & Data Systems*, 109(9), 1166–1178. doi:10.1108/02635570911002252
- Mateen, A., & More, D. (2013). Applying TOC thinking process tools in managing challenges of supply chain finance: A case study. *International Journal of Services and Operations Management*, 15(4), 389–410. doi:10.1504/IJSOM.2013.054882
- Mehrjerdi, Y. (2009). RFID-enabled supply chain systems with computer simulation. *Assembly Automation*, 29(2), 174–183. doi:10.1108/01445150910945624
- Park, B.-N., & Min, H. (2013). Global supply chain barriers of foreign subsidiaries: The case of Korean expatriate manufacturers in China. *International Journal of Services and Operations Management*, 15(1), 67–78. doi:10.1504/IJSOM.2013.050562
- RFID Solutions. (2016). Retrieved March 18, 2016 from <http://www.kennedygrp.com/rfid-solutions>
- Roberti, M. (2014). *How RFID could save retailers \$42 billion annually*. Retrieved December 2, 2015, from <http://www.rfidjournal.com/articles/view?12410>
- Smith, A. D. (2010). Balancing internal supply chain logistics: A comparative analysis of manufacturing and service firm operations. *International Journal of Logistics Systems and Supply Management*, 3(2), 145–166.

Smith, A. D. (2015). Process Improvements in Supply Chain Operations: Multi-Firm Case Studies. In S. Puranam (Ed.), *Handbook of Research on Organizational Transformations through Big Data Analytics. A volume in the Advances in Business Information Systems and Analytics (ABISA) Book Series* (pp. 240–261). doi:10.4018/978-1-4666-7272-7.ch016

Smith, A. D., & Minutolo, M. C. (2014). Green supply chain acceptability and internal stakeholder concerns. *International Journal of Logistics Systems and Management*, 19(4), 464–490. doi:10.1504/IJLSM.2014.065666

Smith, A. D., & Offodile, O. F. (2007). Exploring forecasting and project management characteristics of supply chain management. *International Journal of Logistics and Supply Management*, 3(2), 174–214.

SRFID - container tracking and management - Surgere. (2016). Retrieved March 18, 2016 from <http://www.surgere.com/srfid/>

Wee, H.-M., Peng, S.-Y., & Wee, P. K. (2010). Modelling of outsourcing decisions in global supply chains: An empirical study on supplier management performance with different outsourcing strategies. *International Journal of Production Research*, 48(7), 2081–2094. doi:10.1080/00207540802644852

Whitten, D. (2004). User Information satisfaction scale reduction: Application in an IT outsourcing environment. *Journal of Computer Information Systems*, 45(2), 17–26.

ADDITIONAL READING

Al-Sakran, H. (2013). Agent and radio frequency identification based architecture for supermarket information system. *Journal of Computer Science*, 9(6), 699–707. doi:10.3844/jcssp.2013.699.707

Chao, C., Yang, J. M., & Jen, W. (2007). Determining technology trends and forecasts of RFID by a historical review and bibliometric analysis from 1991 to 2005. *Technovation*, 27(5), 268–279. doi:10.1016/j.technovation.2006.09.003

Chien, H.-Y., Yang, C.-C., Wu, T.-C., & Lee, C.-F. (2011). Two RFID based solutions to enhance inpatient medication safety. *Journal of Medical Systems*, 35(3), 369–375. doi:10.1007/s10916-009-9373-7 PMID:20703553

Condea, C., Thiessse, F., & Fleisch, E. (2012). RFID-enabled shelf replenishment with backroom monitoring in retail stores. *Decision Support Systems*, 52(4), 839–849. doi:10.1016/j.dss.2011.11.018

Devaraj, S., Fan, M., & Kohli, R. (2002). Antecedents of B2C channel satisfaction and preference: Validating e-commerce metrics. *Information Systems Research*, 13(3), 316–333. doi:10.1287/isre.13.3.316.77

Dutta, A., Lee, H. L., & Whang, S. (2007). RFID and operations management: Technology, value and incentives. *Production and Operations Management*, 16(5), 646–655. doi:10.1111/j.1937-5956.2007.tb00286.x

Ferguson, R. B. (2006). RFID loses reception: high tag costs are still putting the kibosh on returns on investment. *e-Week*, 23(10), 11–12.

Fisher, J. A., & Monahan, T. (2008). Tracking the social dimensions of RFID systems in hospitals. *International Journal of Medical Informatics*, 77(3), 176–183. doi:10.1016/j.ijmedinf.2007.04.010 PMID:17544841

Green, H., & Khermouch, G. (2005). Barcodes better watch their backs. *BusinessWeek*. Retrieved December 8, 2015, from http://www.businessweek.com/magazine/content/03_28/b3841063.htm



Min, D., & Yih, Y. (2011). Fuzzy logic-based approach to detecting a passive RFID tag in an outpatient clinic. *Journal of Medical Systems*, 35(3), 423–432. doi:10.1007/s10916-009-9377-3 PMID:20703549

Ohashi, K., Ota, S., Ohno-Machado, L., & Tanaka, H. (2008). Comparison of RFID systems for tracking clinical interventions at the bedside. In: *AMIA Annual Symposium Proceedings*, pp 525-529.

Smith, A. D. (2012). Case studies of RFID-related applications in the healthcare and voice-recognition industries. *International Journal of Knowledge-Based Organizations*, 2(3), 48–63. doi:10.4018/ijkbo.2012070103

Smith, A. D. (2013a). Quality and technology-based management initiatives within a hospital environment: Benchmarking case study. *International Journal of Procurement Management*, 6(6), 621–648. doi:10.1504/IJPM.2013.056746

Smith, A. D. (2013b). Competitive uses of information and knowledge management tools: Case study of supplier-side management. *International Journal of Knowledge-Based Organizations*, 3(1), 71–87. doi:10.4018/ijkbo.2013010105

Ting, S. L., Kwok, S. K., Tsang, A. H. C., & Lee, W. B. (2009). Critical elements and lessons learnt from the implementation of an RFID-enabled healthcare management system in a medical organization. *Journal of Medical Systems*, 35(4), 657–669. doi:10.1007/s10916-009-9403-5 PMID:20703523

van der Togt, R., Jan van Lieshout, E., Hensbroek, R., Beinat, E., Binnekade, J. M., & Bakker, P. J. M. (2008). Electromagnetic interference from radio frequency identification inducing potentially hazardous incidents in critical care medical equipment. *Journal of the American Medical Association*, 299(24), 2884–2890. doi:10.1001/jama.299.24.2884 PMID:18577733

Zang, C., & Fan, Y. (2007). Complex event processing in enterprise information systems based on RFID. *Enterprise Information Systems*, 1(1), 3–23. doi:10.1080/17517570601092127

KEY TERMS AND DEFINITIONS

Automatic Identification and Data Capture Technologies (AIDC): Types of AIDC-related technologies to leave the human element out of the data collection and storage functions of information derived from manufacturing, integrated through the manufacturing process, types of authentication concerns and/or e-security strategies, and relationship links to customer profiles. Typical types of AIDC include, bar-coding, RFID, magnetic strips, touch memory, and smart cards.

Barcoding Technology: A long-term and very reliable type of AIDC technology, it is known for its very accurate and economical approaches to identity products and machine readable information from a variety of manufactured goods and services. Most barcodes use a type of standardized bars and spacing coding or symbology that is certified by an international standards body (GS1 System). This system provides for the universal global acceptance of many types of barcodes designed for a variety of shipping and identification applications. Example barcode formats that are in common use today include EAN/UPC, GS1 DataBar, GS1-128, ITF-14, GS1 DataMatrix, GS1 QR Code and Composite Components.

Electronic Product Code (EPC): EPC is a universal identifier that gives a unique identity to an item a RFID tag is affixed to. The identity is made to be unique so that each object is identifiable within the objects' field.

Operations Efficiency: Improving efficiency and reducing waste is a major challenge for hospitals and other patient care facilities looking to lower the cost of providing healthcare services. Far and away the largest contributor to operational

costs in this industry is patient care activities. Since most clinical decisions involve managing products and medical supplies, finding ways to more efficiently manage supply chain activities can have a big impact on overall operational performance.

RFID-Embedded Technologies: RFID technologies are types of automatic data capture techniques that use a combination of active and passive senders and receivers to collect and store codified information for further uses. The implementation

of such technologies should lead to improved managerial and/or supply chain performance. On the surface, there appears to be few drawbacks to implementing such technology into a production process, assuming it enhances performance and improves output of the product. The main issues surrounding the RFID applications are whether the initial costs and labor required to utilize this technology are worth it, and will result in a positive outcome of revenues.

L

Category M

Management Science

The Business Transformation Framework for Managers in Transformation Projects

M**Antoine Trad***IBITSM, Switzerland***Damir Kalpić***University of Zagreb, Croatia*

INTRODUCTION

A decisive business decision in the business transformation of a traditional business environment into an automated business environment is the profile of the business transformation manager (BTM), who should be supported by a holistic framework (Trad & Kalpić, 2001; Trad & Kalpić, 2014a). The BTM's profile and the needed data modelling skills are essential for managing data models' in business transformations. This research chapter and the related research publications deal with business transformation projects (BTP) complexity as well as the support for the BTM's selection and the underlined BTP architecture. The proposed framework promotes the needed business transformation data architecture and modelling skills to insure success: 1) artefacts; 2) components; 3) architecture; and 4) modelling concepts.

The success of a business transformation project (BTP) depends on how an enterprise architecture, data architecture and modelling activities are synchronized (IMD, 2015).

That is why the implementation of such BTPs requires significant knowledge of data architecture and modelling techniques. The author has based his research on many credible research sources of information like the Gartner Inc. and many others. The main fact is that only a small percentage of business organizations successfully terminate innovation-related BTPs; another important fact

is that business environments, which have a good data architecture and modelling concept, will gain a substantial business advantage (Tidd, 2006; Tidd & Bessant, 2009).

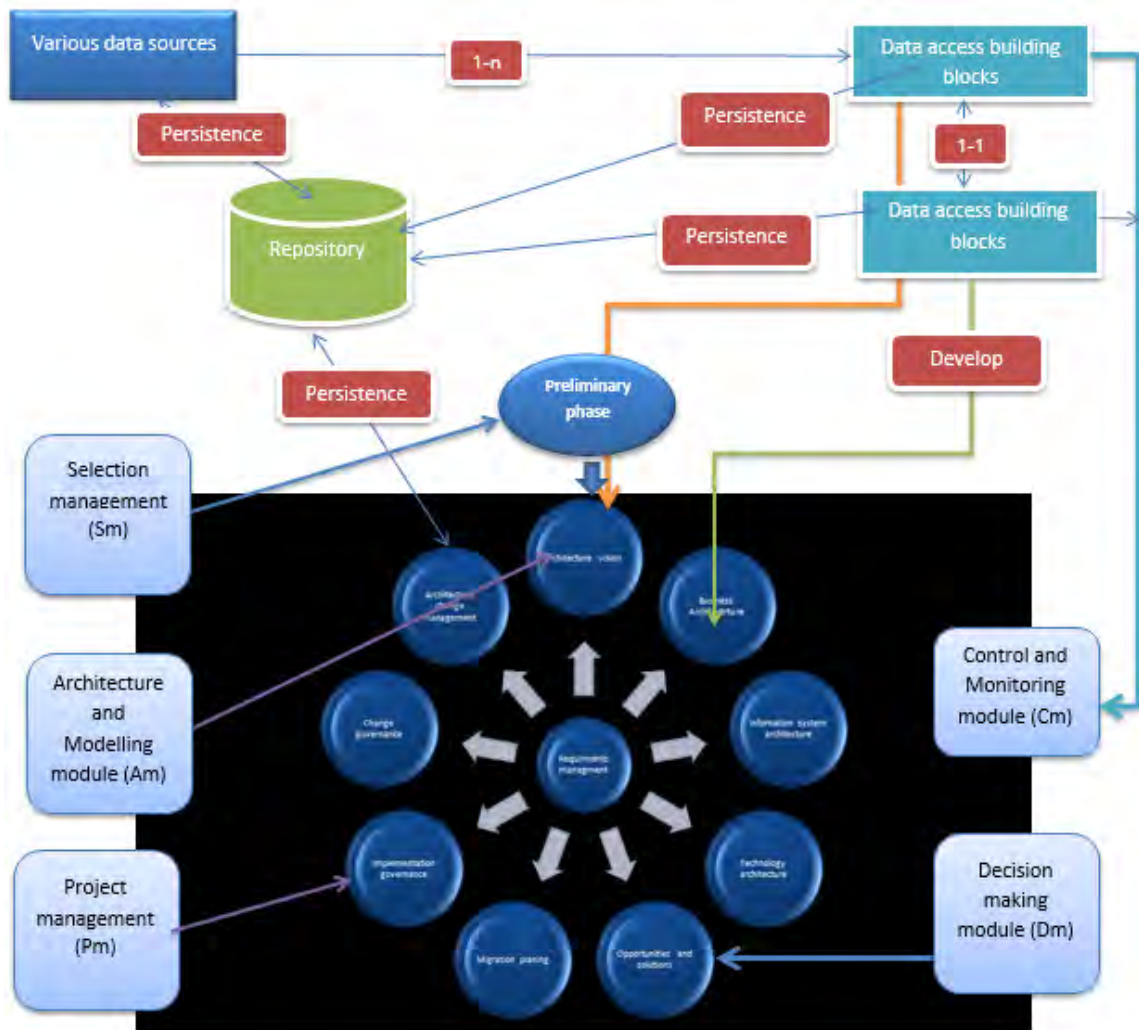
The data architecture and modelling module is a part of the Selection management, Architecture-modelling, Control-monitoring, Decision-making, Training management and Project management Framework (SmAmCmDmTmPmF, for simplification in further text the term *Environment* will be used), that supports the BTP's activities. As shown in Figure 1, the data architecture and modelling concept interacts with all the enterprise's architecture phases, using the data building blocks or the holistic brick (Trad & Kalpić, 2014a).

BACKGROUND

What is an architecture framework and more specifically, what is a data architecture and modelling module or concept? In general an architecture framework (The Open Group, 2011):

1. Is a foundational model, or set of classes/entities, which can be integrated in various architectures;
2. Describes a method for modelling a systemic view of the business enterprise in terms of a set of data building blocks, it should also show how the data building blocks and code building blocks collaborate;

Figure 1. Enterprise architecture cycles and the data access building blocks
Trad, 2015a; Trad, 2015b.



3. Contains a set of tools and provides a common vocabulary;
4. Includes a list of recommended standards and compliant products that can be used to implement the building blocks;
5. Includes a data architecture and modelling concept that refers to various techniques for the integration of different data models and data sources. Where this concept is based on data building blocks.

The global research topic's and final research question (hypothesis #1-1) is: "Which business transformation manager's characteristics and which type of support should be assured for the implementation phase of a business transformation project?" The targeted business domain is any business environment that uses: 1) internet technologies; and 2) frequent transformation iterations. For this phase of research the sub-question (or hypothesis #2-3) is: "What is the impact of the data architecture and modelling concept on

enterprise architecture and business transformation projects?”

MAIN FOCUS OF THE ARTICLE

In this research phase the author target the data architecture and modelling concept's usage that is a part of the Architecture module (Am), and he tries to prove that such a concept can be built on a loosely coupled architecture (Trad, 2015a; Trad, 2015b).

Data Management and Complexity

Chaotic and complex data architecture and modelling concepts can cause failures, where business environments have complex data architectures, as shown in Figure 2 (IBM, 2015a).

RESEARCH METHODOLOGY AND DESIGN

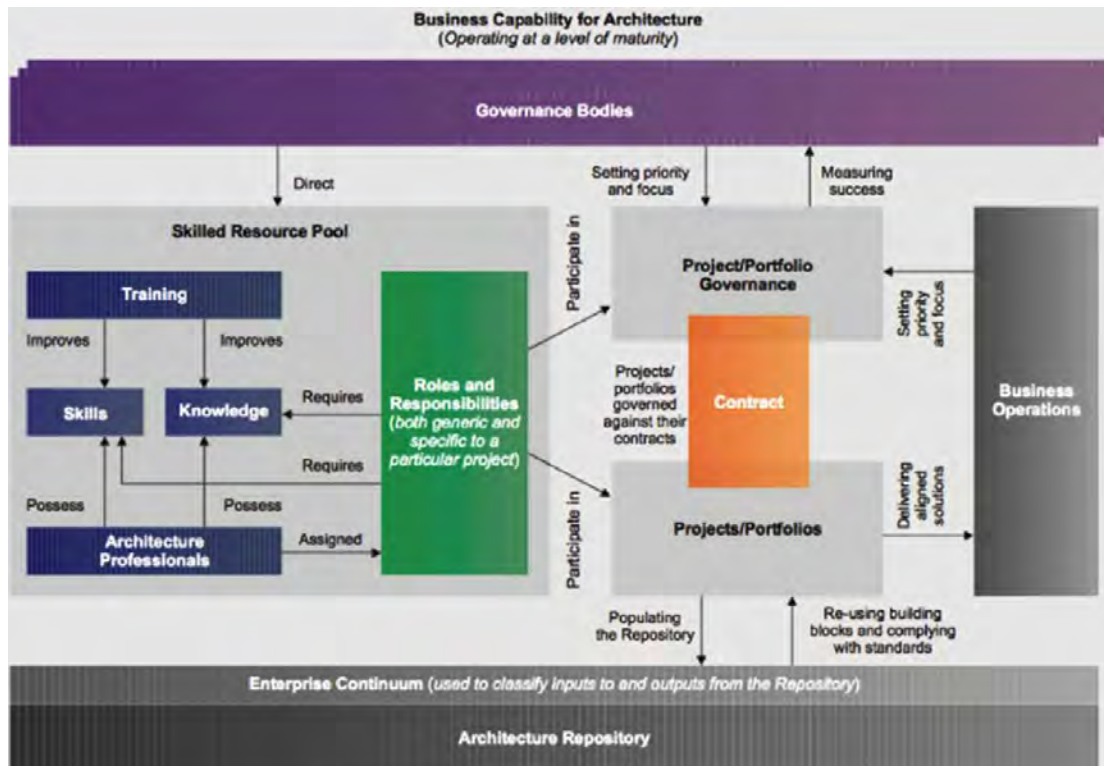
As shown in Figure 3, the *Environment's* research methodology focuses on the impacts of the mechanistic enterprise's architecture integration and uses a mixed hyper-heuristics based methodology (Vella, Corne, & Murphy, 2009). This methodology offers the possibility to select a BTM and also offers guidance for the BTP's management (Trad & Kalpić, 2001); it proposes also a data architecture and modelling concept to be used by BTPs (Trad, 2013a).

Business transformation research projects and business engineering research projects use a proof of concept or prototyping to establish the: 1) viability; 2) technical issues; and 3) needed recommendations (Camarinha-Matos, 2012).

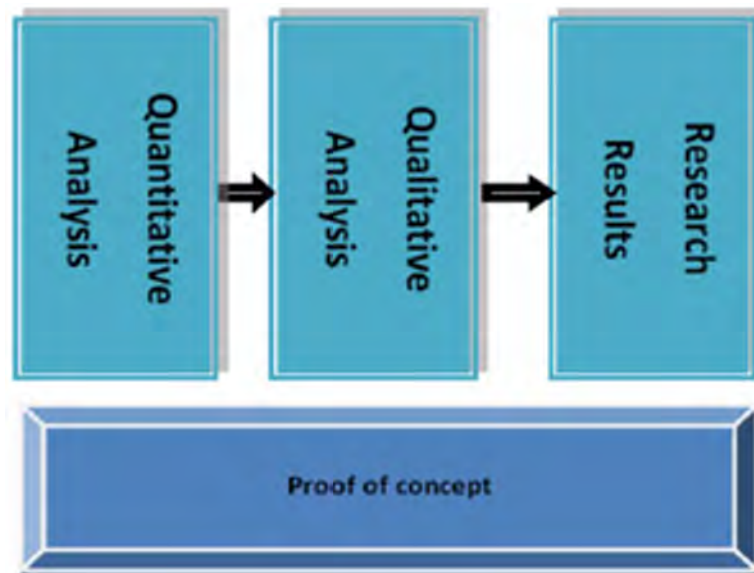
The research's sub-project's proof of concept is used to prove the feasibility of the data architecture

Figure 2. Structure of an enterprise architecture document and its relation to the data architecture and modelling concept

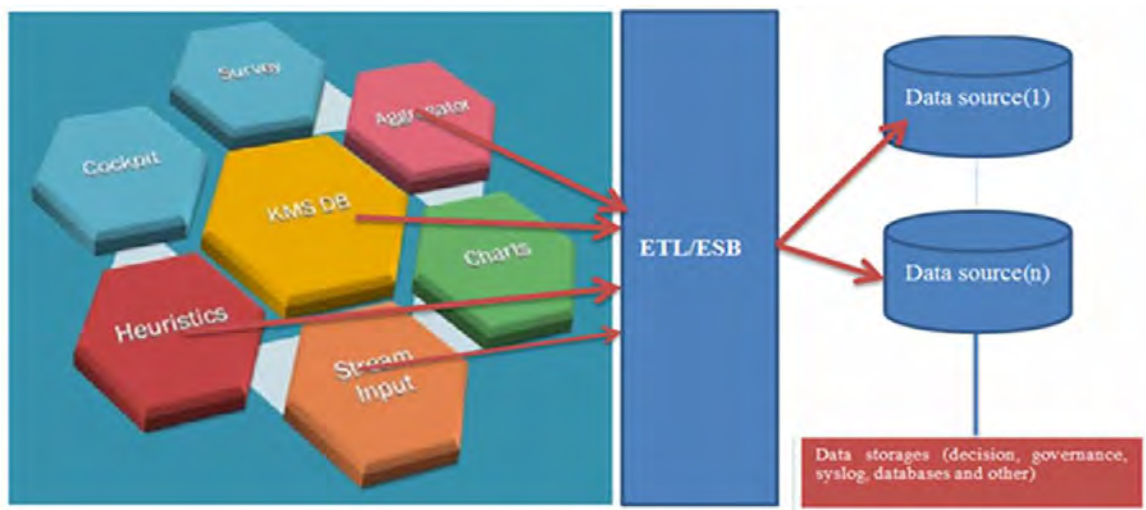
The Open Group, 2011; Trad, 2015a; Trad, 2015b.



*Figure 3. The mixed method flow diagram
Trad & Kalpić, 2013a.*



*Figure 4. The proof of concept's overall diagram
Trad, 2015a; Trad, 2015b.*



and modelling concept, as shown in Figure 4. The proof of concept is based on a transaction management case that uses data class diagram, as shown in Figure 5. In general data packages are defined in the preliminary phase and then calibrated in the requirement and business architecture phases. At the business analysis level, the data class diagram permits the definition of the fundamental BTP's data artefacts.

BUSINESS DATA ACCESS ARCHITECTURE AND AGILITY

Enterprise Architecture Principles

Enterprise architecture principles are divided into four categories (The Open Group, 2015a):

1. Business architecture principles.

2. Data architecture principles.
3. Application architecture principles.
4. Technology architecture principles.

The BTM must define the BTP's data architecture principles for various data sources that can be composed of the following characteristics (The Open Group, 2015a):

1. Data sources are shared in real-time.
2. Data sources are accessible in the intranet and extranet.
3. Data sources are of good quality and are trusted.
4. Data architecture and modelling have a common vocabulary.
5. Data security is implemented in all data sources.
6. Business processes that use data sources have to be standardized.
7. Data source management is managed by a defined role.
8. Data sources must be protected from unauthorized access and modification.

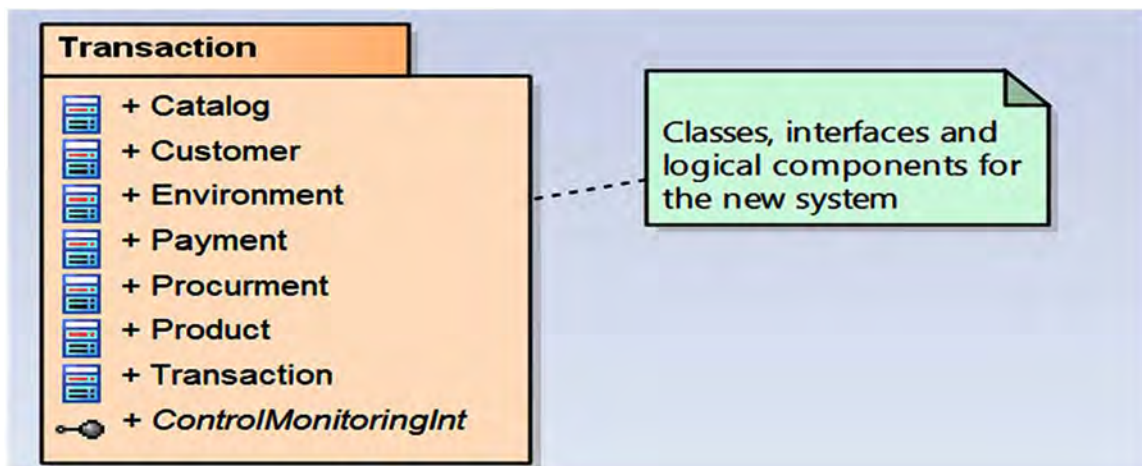
In this chapter the author focuses on the data architecture principles that have an important influence on all other architecture principles.

Data Strategy Principles

The data architecture and modelling strategy should be based on well-defined principles, like (IBM, 2015a):

1. Data have to be easy to enter and must be accurate.
2. Data collections are based on data modelling concepts.
3. Data must be available to business personnel.
4. Business processes for data management have to be automated.
5. Data should not be redundant.
6. Data have to be recorded accurately and must be auditable.
7. The cost of data collection procedures have to be optimized.
8. A data architecture and modelling concept must adopt standards for common data models.

Figure 5. The proof of concept's class diagram package
Trad, 2015a; Trad, 2015b.



The Role of Data Standards

The *Environment* uses data technologies and existing data standards that include: 1) relational database standards; 2) vertical industry documents' standards; 3) business process standards; 4) data governance standards; 5) business services standards; 6) extensible mark-up language standards; 7) object oriented and object relational modelling standards; and 8) resources description standards. Standardization for inter-operability can be achieved by using an Enterprise Service Bus (ESB) and its Extract Transform Load (ETL) module (Tamr, 2014). The evolution of technology standards, like the service oriented architecture standards, have enabled data architecture and modelling concepts to become more receptive to development and integration with various standards. Standardized BTPs have to be inter-operable and their focus must be on their: 1) data models; 2) data architectures; 3) modelling concepts; and 4) data monitoring platforms. Regardless of the business domain, executive management understands the immense need for agility and the integration of data architecture and modelling concepts. The *Environment's* data architecture and modelling concept helps in establishing a unique architecture (Trad, 2015b).

UNIFYING THE DATA SOURCES AND ENTERPRISE DATA NORMALIZATION

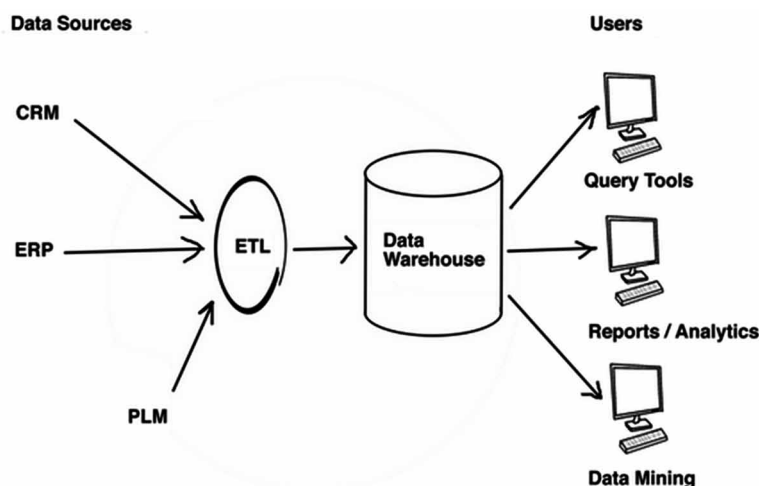
Enterprise Service Bus and Enterprise Application Integration

BTPs must use Enterprise Service Buses (ESB) to glue the various data sources of the business environment, through the use of the technology stack and data connectors, which permit a holistic data services' management (Erl, 2008; Erl, 2009)

Extraction, Transform and Load

Extract Transform and Load (ETL) processes are defined as accessing data stored in various locations and transforming them in order to enable their unification, quality or normalization. The *Environment's* data architecture and modelling concept, proposes the separation of data processing activities; that enables data services to access data without bothering about various data sources' complexities. The *Environment's* data architecture and modelling concept insures: 1) intra (or extra) data transparency; 2) managing accessibility; and 3) data quality control (Tamr, 2014). As shown in Figure 6, the ETL processes are responsible for the

Figure 6. Various data sources
Tamr, 2014.



access of data from heterogeneous data sources. The challenge lays in the real-time data transformation and normalization processes (Trujillo & Luján-Mora, 2003).

Data Transformation Models

The Meta-Model

This research sub-project's proof of concept uses a meta-model to show the relationships between the BTP's different types of data models and data sources. This BTP's meta-model contains descriptions on how data sources are handled in the newly transformed business system's and of the types of used data structures like (IBM, 2015a):

1. The physical structure uses physical classes and their attributes to model the data schema. Such class models use the enterprise application integration hub to transform data sets.
2. The establishment of relationships between different physical data classes.
3. Data transformation rules and mappings at the class attribute level.
4. Interface drivers, constraints and timing rules for interaction diagrams.
5. Global mapping technique and object oriented relations.

Normalization and Inter-Operability

A BTP data architecture and modelling concept must be based on various standards that are all in turn based on the eXtensible Markup Language (XML) technologies; XML enables ontology-mediated interoperability that is based on data schemas. Schemas enable the alignment by defining common data structures. Data architecture and modelling conceptual schemas are used in the mediation module that describes the semantics of data sources and destinations. Data architecture and modelling conceptual normalisation is a three layered approach for information representation,

as shown in Figure 7, the three layers for data integration include: 1) syntax checking; 2) data modelling; and 3) semantics control. The proof of concept uses the normalisation conversion modules that include a set of tables to automate the data translation process (Fodor, Dell'Erba, Ricci, Spada, & Werthner, 2002).

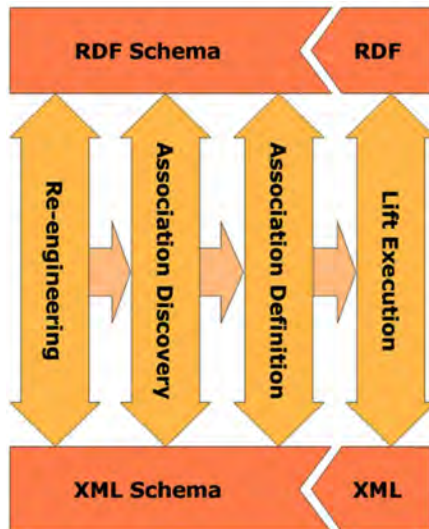
The Resources Description Framework's (RDF) database system links open data structures and is based on semantic technology (Curé & Blin, 2014). The role of metadata is important due to evolution of business, technology and the Web; where XML is the standard for data interchange, but XML schemas lack the capacity to express semantics and cannot formulate the document's structure. Currently there is a lack of semantic interoperability regarding current metadata standards. In the use of semantic web technologies, ontologies are produced to describe the semantics of a particular metadata format. Unfortunately XML data that have a specific metadata format cannot work with an ontology that forces the conversion of XML data to RDF instances. A generic approach can be implemented to transform XML data into RDF instances in an ontology-compliant manner. A mapping document supports the links between an XML Schema and Web Ontology Language (OWL) ontology (Van Deursen, Poppe, Martens, Mannens, & Walle, 2008).

DATA ARCHITECTURE AND MODELLING

The BTP needs enterprise, solution, technical and infrastructure architectures to support an enterprise information system's architecture that has many interrelated aspects, including applications, hardware, networks, business processes, technology choices, and data architecture. As shown in Figure 8, the data architecture is a layered set of models which provide a solid foundation for strategic initiatives such as (IBM, 2015a):

Figure 7. Stages of conceptual normalisation process

Fodor, Dell’Erba, Ricci, Spada, & Werthner, 2002.



1. The data architecture and modelling strategy that defines the business’s principles and objectives for the transformed data.
2. Business processes’ transformation.
3. Decisions on the BTP’s actions.
4. Integration, data storage management and reporting initiatives.

The Class Diagram

In data architecture and modelling, the main data class diagrams, as shown in Figure 9, are to be defined at the initial phase and then calibrated in the requirement and business architecture phases. At the management levels, the data class diagram permits the essential BTP’s data architecture and modelling concepts to be designed, without taking into account the enterprise’s complexities and resistances. This managerial or business architecture level enables the BTP team to concentrate on the business aspects (TOGAF Modelling, 2015b).

Mapping Concepts

The data architecture and modelling are based on the “1:1” mapping concept, this concept helps in achieving the assertion of the BTP’s data models (Kabzeva, Niemann, Müller, & Steinmetz, 2010).

Data Solution Blocks

As shown in Figure 11, the *Environment* uses the data architecture and modelling concept that includes the pattern on how to integrate data solution blocks which are instances of the data building blocks (Tripathi, 2011).

Figure 8. Data architecture and modelling is a layered set of models
IBM, 2015a.

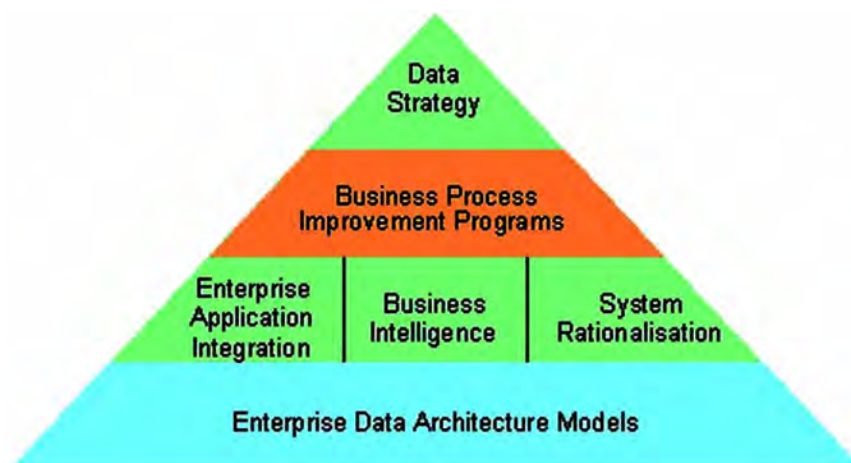
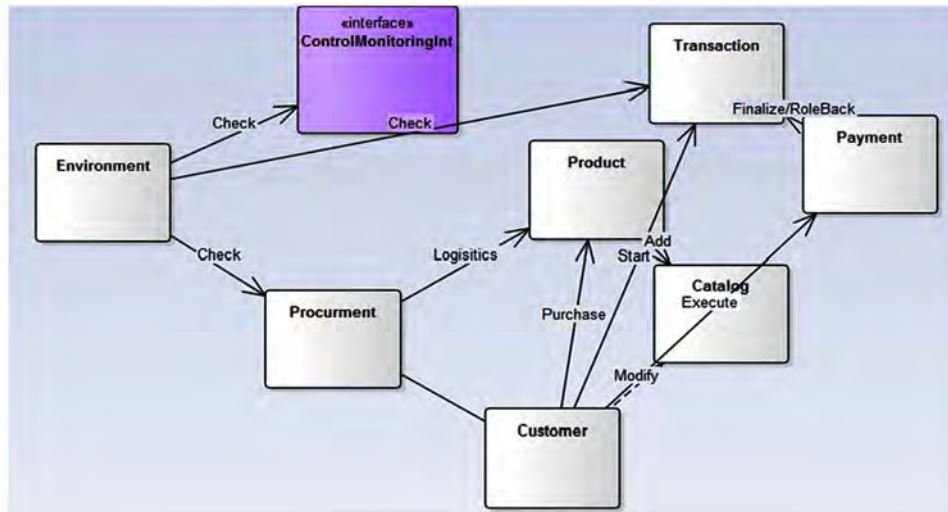


Figure 9. The proof of concept's class diagram
Trad, 2015a; Trad, 2015b.



Transactions Data Modelling

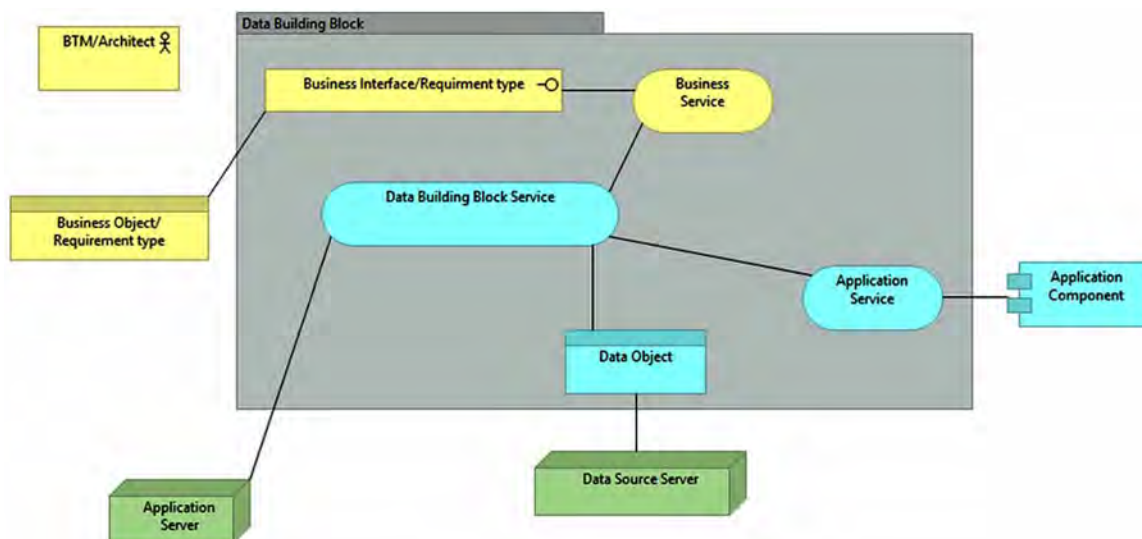
The BTP must put important considerations on technical assertion, governing, control, access management and monitoring of business transactions that are based on the Model View Controller (MVC) pattern, as shown in Figure 12. Business transactions use the following standards: 1) the policy framework (Oracle, 2015a); and 2) busi-

ness messaging that is needed for domain-specific integration (Oracle, 2015b; OASIS, 2009).

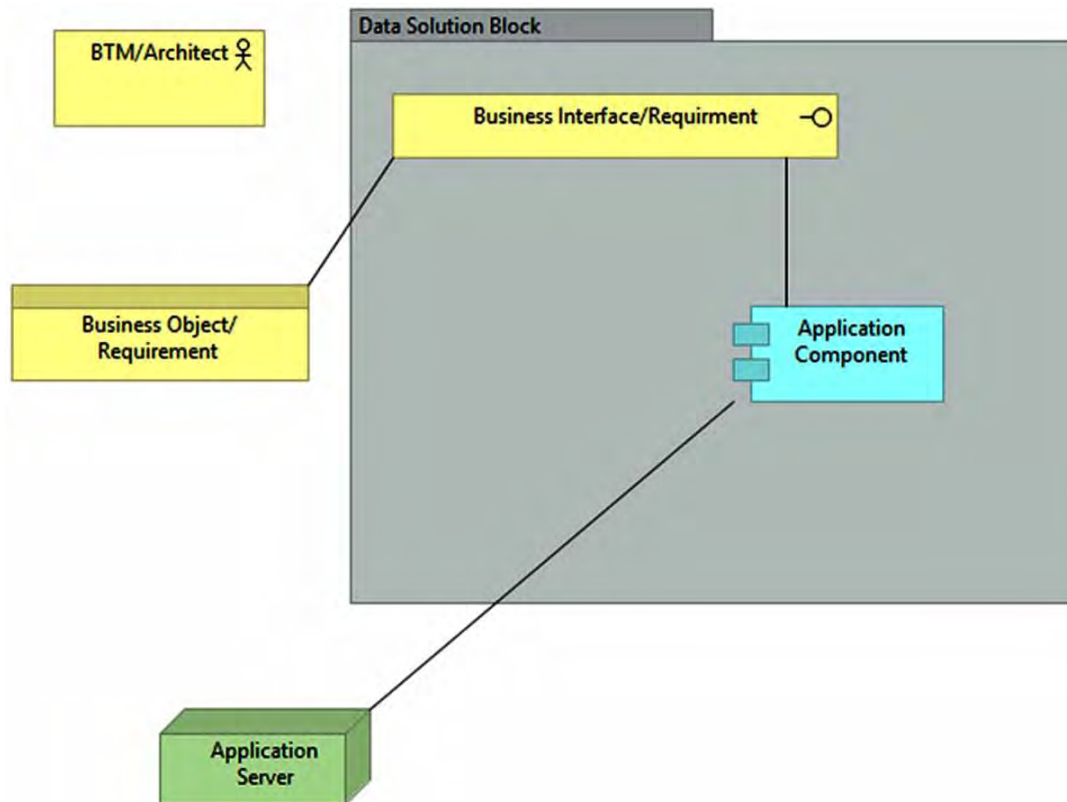
Architecture Development Method's Integration

The usage of the architecture development method's phases in the BTP's data architecture and modelling is done in the following phases (The Open Group, 2014a):

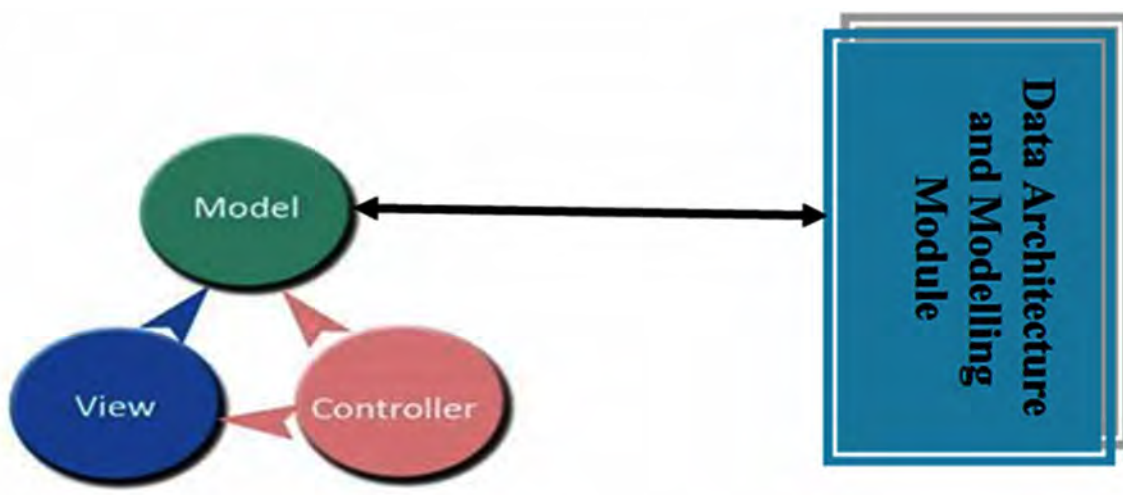
Figure 10. The proof of concept's data building block diagram
Trad, 2015a; Trad, 2015b.



*Figure 11. The proof of concept's data solution block diagram
Trad, 2015a; Trad, 2015b.*



*Figure 12. The proof of concept's model view controller diagram
Trad, 2015a; Trad, 2015b.*



1. **Vision Phase:** In this phase the concept of the basic set of data building blocks is defined.
2. **Requirement's Management Phase:** Insures that any changes to the BTP's requirements are handled in the appropriate data architecture and modelling building blocks (Kabzeva, Niemann, Müller, & Steinmetz, 2010).
3. **Business Architecture Phase:** Receives the active data building block instance(s) and enhances its contents.
4. **Information System's Phase:** Receives the active data building block instance and enhances the data architecture and modelling to: 1) integrate the detailed project plans; 2) the design of the needed real artefacts; 3) the analysis of the needed real artefacts; and 4) the implementation and testing of the needed real artefacts.
5. **Change Governance Phase:** Receives the active data building block instance and enhances the data architecture and modelling logging interfaces for the: 1) project plans changes; 2) modifications of the architecture vision; etc.

Data Modelling

A data model is needed to describe the structure and relationships among a set of data managed in the BTP and stored in a data management system. The most common is the relational model that is often represented in entity-relationship diagrams or class diagrams; these diagrams basically show data entities and their relationships. The data architecture and modelling concept for a BTP can be seen as a data architectural view (Merson, 2009); where the modelling process incorporates various conceptual views.

Data Dissemination View

The purposed data dissemination diagram shows the relationship within the proof of concept: 1) data

entities; 2) business data services; and 3) application components/services. This diagram as shown in Figure 13, presents the logical data entities to be physically managed/accesses by application components or services. That permits a flexible architecture and modelling of the data sources. This diagram includes data services that abstract various data sources (TOGAF Modelling, 2015a).

Data Lifecycle View

The proof of concept's data lifecycle diagram manages the business processes' data throughout their lifecycle, from architecture, modelling, usage to disposal, within the constraints of the business process. The data are loosely coupled with business processes and that allows the definition of common data models enabling an effective data resources' sharing. Each data model must have its status that contains the transitions and the gaps between phases. A phase's state must contain a stable data model, where the data architecture and modelling lifecycle enables a qualitative design of the related business entities. The resultant data models must be atomic or simple in order to (re) engineer the BTP's business processes (TOGAF Modelling, 2015c).

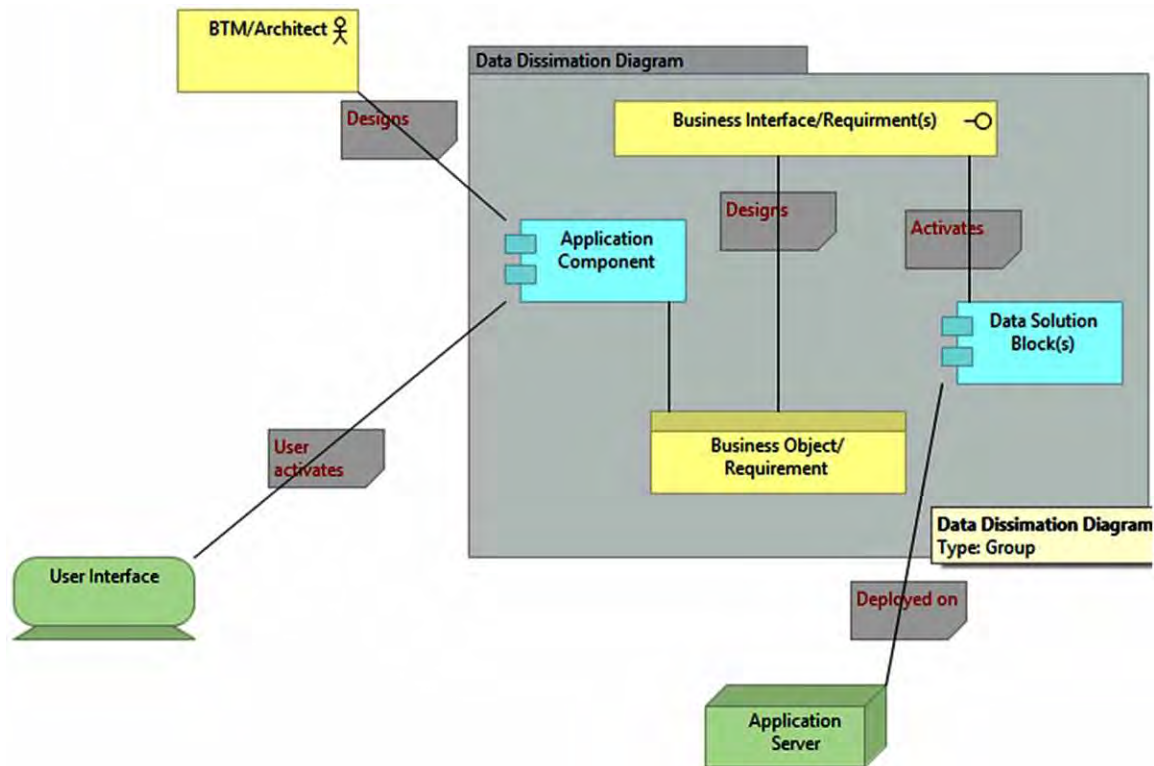
Data Migration Diagram

The proof of concept's data migration diagram presents the flow of data between the business systems' various data sources; it serves as a view for the BTP's data auditing and traceability map. The data migration diagram is enhanced through BTP's iterations (TOGAF Modelling, 2015d). A mapping table was developed to show the relationships between various data points (source, intermediate or destination), as shown in Table 1.

Data Security Diagram

Design of data architecture and modelling are the most important activities in the BTP and the transformed business data are a very important

*Figure 13. The proof of concept's data dissemination view
Trad, 2015a; Trad, 2015b.*



asset for the enterprise; that implies that the transformed business data security is an ultimate goal, as shown in Table 2. The proof of concept's data security diagram presents the BTP's actors' roles for transformed business data. A matrix is used to present the mapping relation between the role

and the transformed data entity. The data security diagram includes all external accesses which may have access to the business system's data (TOGAF Modelling, 2015e).

Enterprise Data Architecture Models

The proof of concept's data entities or classes are the basis for a variety of enterprise's data models, as shown in Figure 13 (IBM, 2015a), that contains the following technical domains:

1. Data quality;
2. Interoperability;
3. Logical data model.

Table 1. The names of the entities that are mapped from the source to the destination media

Source Class/Entity/Object	Destination Class/Entity/Object
Catalog	dCatalog
Customer	aCustomer
Environment	dEnvironment
Payment	dPayment
Procurement	dProcurement
Product	dProduct
Transaction	dTransaction

Trad, 2015a; Trad, 2015b.

Table 2. The actor data access rights mapped to the type of domain

Actor	Marketing	Finance	...
Catalog	CRUD	_R_	
Customer	_RU_	_R_	
Environment	CRUD	CRUD	
Payment	_R_	CRUD	
Procurement	CRUD	_R_	
Product	CRUD	_RU_	
Transaction	_R_	CRUD	

Trad, 2015a; Trad, 2015b.

Data Architecture and Modelling Exchange

Inputs

The inputs for this phase are: 1) Data principles; 2) Request for architecture work; 3) Statement of architecture work; 3) Architecture vision; 4) Relevant technical requirements; 4) Gap analysis; 5) Baseline business architecture; 6) Target business architecture; 7) Baseline data architecture; 7) Target data architecture; and 8) Re-usable building blocks.

Outputs

The outputs for this phase are: 1) Statement of architecture work; 2) Baseline data architecture; 3) Validated data principles; 4) Target data architecture; 5) Business data model; 6) Logical data model; 7) Data management process models; 8) Data entity/business function matrix; 9) Data interoperability requirements; and 10) Viewpoints addressing key stakeholder concerns.

Business Transformation Management's Data

Data Architecture and Modelling Gap Analysis

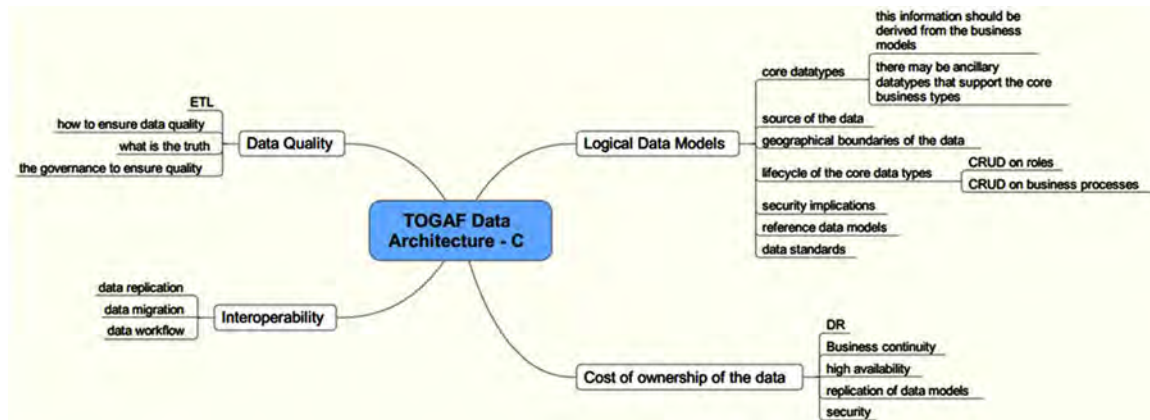
The data architecture and modelling concept's requirements must be tracked for the evolution

of the BTP's architecture development cycle(s). Data architecture and modelling concept's routine report must contain: the statuses, achievements and impact analysis. The impact analysis gaps present where the data architecture and modelling concept needs to be changed and which data building (and solution blocks) to be included. Changes in the data architecture and modelling concept implies a change in the data building and solution blocks. A key step in validating a data architecture and modelling concept, is to verify the artefacts and to find the need for change. Types of data architecture and modelling concept's gaps' need for change are (The Open Group, 2015b): 1) data not found by the data services; 2) wrong data is delivered; 3) data unreachable; 4) data was not correctly stored; 5) data cannot be used; 6) bad data relationships; and 7) those that include the stakeholder's concerns.

THE PROOF OF CONCEPT

The *Environment* serves as the proof of concept's template and is based resources from The Open Group's Architecture Framework (TOGAF) (Trad & Kalpić, 2014b). The *Environment* contains five modules which can be launched independently; these components are used to tune the characteristics of the BTM's profile and to support the BTP's implementation phase. The *Environment's* are interconnected and have a pool of critical success factors. The data architecture and modelling

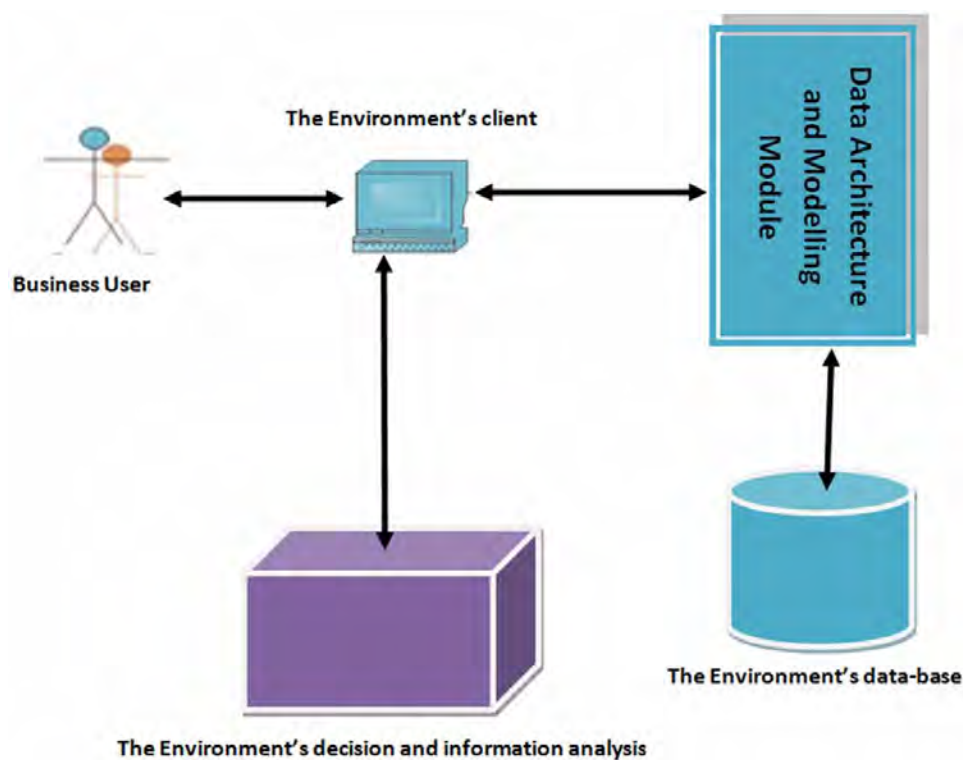
Figure 14. The proof of concept's data enterprise models
The Open Group, 2014X.



concept can be used in a real world system and this feature is the research's major business and managerial benefit (Wes, 2001). The *Environment's* data architecture and modelling concept structures, as shown in Figure 14, serve many functional domains and in this proof of concept

the functional domain are information analysis and decision making of a business transaction's log; where the data that results from business transactions are logged and used in the data analysis process, as shown in Figure 15.

Figure 15. The proof of concept's data interface view
Trad, 2015a; Trad, 2015b.



Information Analysis

The data architecture and modelling concept proof of concept took into account the integration of information analysis processes that are used by a virtual user who simulates a business analyst (Trad, 2015c). In this proof of concept a spreadsheet was built to use various data sources that fed the prototyped business transactions system, as shown in Figure 15.

Layers

The prototyped business transaction is based on the proof of concept's class model, as shown in Figure 16.

As shown in Figure 17, the data architecture and modelling proof of concept's layers are:

1. The client layer that contains the following packages based on: Microsoft Excel, Flat Files Interfaces and a Client web interface (Trad, 2015c).
2. The services layer that contains a data services hub.
3. The data management layer that contains the following packages: an entity relational

model, an extensible mark-up language transformer and a NoSQL database interface.

4. The control-monitoring layer contains a generic logger interface.
5. The platform layer that contains the following packages: an enterprise service bus, an object database connector, and a java database connector.

SOLUTIONS AND RECOMMENDATIONS

This chapter's and its proof of concept's list of the most important technical and managerial recommendations:

1. The BTMs should be supported with a set of data architecture and modelling concepts and tools that can be easily integrated with the defined data architecture principles (Trad & Kalpić, 2014b; IBM, 2015a).
2. The role of data standards is important; today there are many standards concerning data architecture and modelling concepts. The BTM must propose a solution on how

Figure 16. The proof of concept's transaction view
Trad, 2015a; Trad, 2015b.

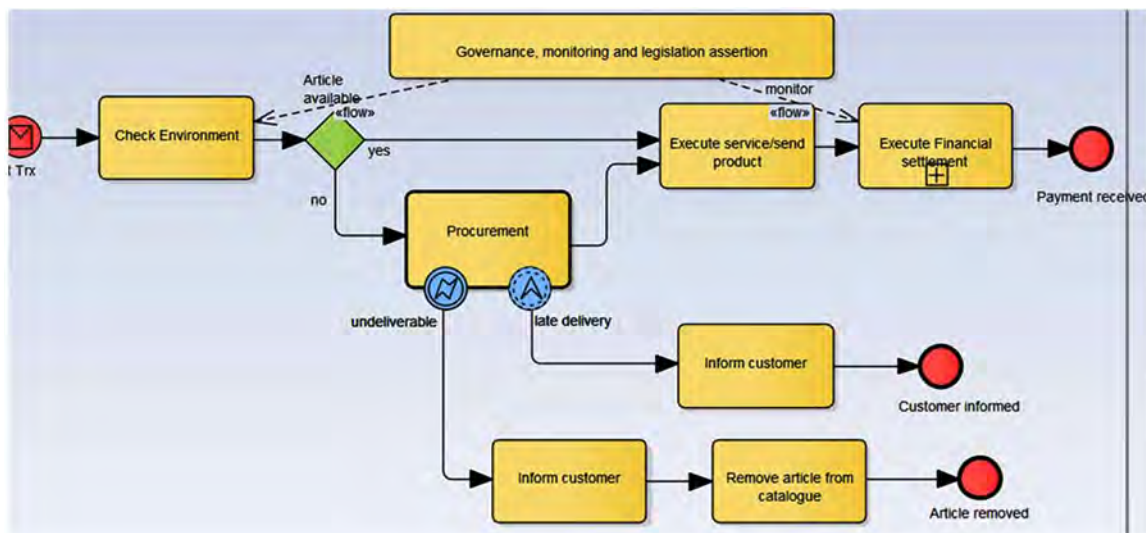
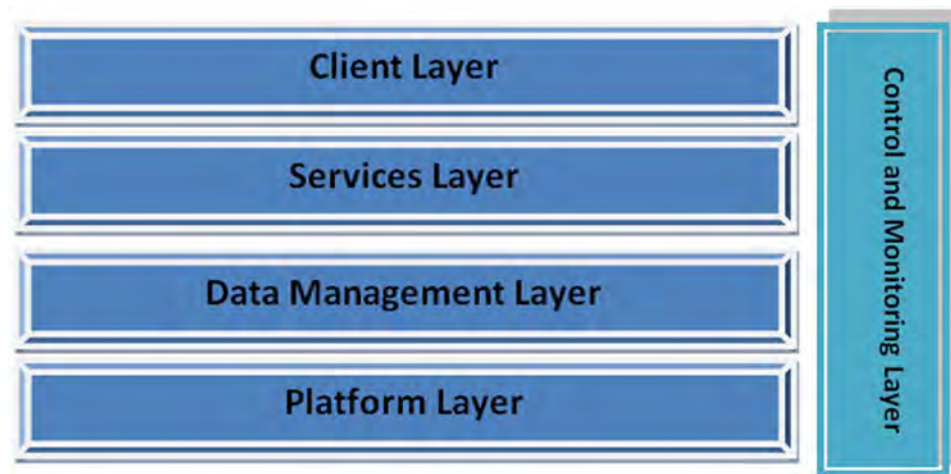


Figure 17. The proof of concept's layers view
Trad, 2015a; Trad, 2015b.



- to integrate these various data standards in the BTP.
3. An enterprise data service bus and enterprise application integration concept must be defined for the BTP. That would glue the various data sources of the business environment.
4. The meta-model and the corresponding “1:1” mapping approach can be used to manage an agile business transformation process (OASIS, 2014).
5. Unbundling through the use of business data services makes decision making well integrated in the context of a system that applies a holistic approach (Daellenbach, McNickle, & Dye, 2012).
6. The data services and data building blocks must unify the implementation and usage of data models.
7. Define the data architecture and modelling concept’s default critical success factors.
8. Implement an Extract, Transform and Load (ETL) process to access data from various locations to enable their unification and normalization.
9. Data transformation models need a meta-model to show the relationships between various types of data models and data sources.
10. Define a data normalization and interoperability paradigm that must be based on various standards.
11. Define a business process and a data architecture and modelling integration paradigm to enhance the business environment’s knowledge.
12. Develop an agile Model View Controller (MVC) pattern that is dynamically created for each business transaction. This MVC uses a data building block and its instance
13. Define the architecture development method’s integration with the proposed concept where the data building and solution blocks are the basic artefacts that circulate through the architecture development method.
14. Define business architecture integration paradigm that is needed to manage the implementation of the data building blocks (Trad & Kalpić, 2014b; The Open Group, 2014a).
15. Define conceptual views that can be built to simplify the application of the data architecture and modelling concept. Such views can also be used to simplify the enterprise’s data-model that links various data sources.

16. Create basic class diagrams for the data architecture and modelling concept(s) to be the central artefact that is defined in the initial phase and then calibrated in all other phases.
17. Define the usage of business transactions' data models in various class diagrams.

CONCLUSION

This is another chapter in a long series of articles (and chapters) related to the *Environment's* research framework, which is based on action research mixed method and is mainly motivated by high failure rates in business transformation projects (Wes, 2001). The proposed *Environment's* data architecture and modelling concept are result of a research and prototype development process. The requirement is to present the needed data architecture and modelling skills for a BTP's data concept that are needed to finalize the implementation phase of a BTP (The Open Group, 2014d).

FUTURE RESEARCH DIRECTIONS

The *Environment* future research will focus on a neural network based decision making module, which use a micro services concept.

REFERENCES

- Camarinha-Matos, L. M. (2012). *Scientific research methodologies and techniques- Unit 2: Scientific method. Unit 2: Scientific methodology. PhD program in electrical and computer engineering*. Uninova.
- Curé, & Blin. (2014). *RDF Database Systems, 1st Edition-Triples Storage and SPARQL Query Processing*. Morgan Kaufmann. Retrieved from <http://store.elsevier.com/RDF-Database-Systems/Olivier-Cur%C3%A9/isbn-9780127999579/>
- Daellenbach, H., McNickle, D., & Dye, Sh. (2012). *Management Science. Decision-making through systems thinking* (2nd ed.). Plagrave Macmillian.
- Enterprise Architecture Developments. (2004). Available online at: <http://www.enterprise-architecture.info/>
- Fodor, O., Dell'Erba, M., Ricci, F., Spada, A., & Werthner, H. (2002). *Conceptual Normalisation of XML Data for Interoperability in Tourism. Commerce and Tourism Research Laboratory ITC-irst*. University of Trento.
- IBM. (2015). *Modelling the enterprise data architecture*. IBM.
- IMD. (2015). *IMD business school and Cisco join forces on digital business transformation*. IMD.
- Kabzeva, A., Niemann, M., Müller, P., & Steinmetz, P. (2010). Applying TOGAF to Define and Govern a Service-oriented Architecture in a Large-scale "research & development" (R&D). *Proceedings of the Sixteenth Americas Conference on Information Systems*.
- Merson, P. (2009). *Data Model as an Architectural View. Technical note. CMU/SEI-2009-TN-024*. Research, Technology, and System Solutions.
- OASIS. (2009). *OASIS Web Services Reliable Messaging (WS-ReliableMessaging) Version 1.2*. OASIS Standard. Retrieved from <http://docs.oasis-open.org/ws-rx/wsrn/200702/wsrn-1.2-spec-os.html>
- OASIS. (2014). *ISO/IEC and OASIS Collaborate on E-Business Standards-Standards Groups Increase Cross-Participation to Enhance Interoperability*. Retrieved from <https://www.oasis-open.org/news/pr/isoiec-and-oasis-collaborate-on-e-business-standards>
- Oracle. (2015a). *Using WS-Policy in ALSB Proxy and Business Services*. Oracle. Retrieved from http://docs.oracle.com/cd/E11036_01/alsb30/security/ws_policy.html

- Oracle. (2015b). *About Web Services Policy*. Oracle. Retrieved from http://docs.oracle.com/cd/E11036_01/alsb30/security/ws_policy.html
- Tamr. (2014). *The Evolution of ETL*. Tamr. Retrieved from <http://www.tamr.com/evolution-etl/>
- The Open Group. (2011). *Open Group Standard-TOGAF® Guide, Version 9.1*. The Open Group.
- The Open Group. (2014a). *The Open Group's Architecture Framework-Building blocks-Module 13*. Retrieved from www.open-group.com/TOGAF
- The Open Group. (2015a). *Architecture Principles-Data Principles*. Retrieved from <http://pubs.opengroup.org/architecture/TOGAF9-doc/arch/chap23.html>
- The Open Group. (2015b). *Phase C: Information Systems Architectures - Data Architecture*. The Open Group.
- Tidd, J. (2006). *From Knowledge Management to Strategic Competence* (2nd ed.). London, UK: Imperial College. doi:10.1142/p439
- Tidd, J., & Bessant, J. (2009). *Managing Innovation, Integrating Technological, Market and Organizational Change* (4th ed.). Wiley.
- TOGAF Modelling. (2015a). *Data dissemination view*. TOGAF Modelling. Retrieved from <http://www.TOGAF-modelling.org/models/data-architecture-menu/data-dissemination-diagrams-menu.html>
- TOGAF Modelling. (2015b). *Class diagram view*. TOGAF Modelling. Retrieved from <http://www.TOGAF-modelling.org/models/data-architecture-menu/class-diagrams-menu.html>
- TOGAF Modelling. (2015c). *Data lifecycle diagram*. TOGAF Modelling. Retrieved from <http://www.TOGAF-modelling.org/models/data-architecture-menu/data-lifecycle-diagrams-menu.html>
- TOGAF Modelling. (2015d). *Data migration diagram*. TOGAF Modelling. Retrieved from <http://www.TOGAF-modelling.org/models/data-architecture-menu/data-migration-diagrams-menu.html>
- TOGAF Modelling. (2015e). *Data security diagram*. TOGAF Modelling. Retrieved from <http://www.TOGAF-modelling.org/models/data-architecture-menu/19-data-security-diagrams.html>
- Trad, A. (2015a). *The Transformation Framework's Resources Library*. IBISTM.
- Trad, A. (2015b). *The Transformation Framework Proof of Concept*. IBISTM.
- Trad, A. (2015c). "Data Analysis Courseware". *Laboratory works*. IBISTM.
- Trad, A., & Kalpic, D. (2001). *Building an extensible markup language (XML) based Object Mapping System (OMS)*. IEEE.
- Trad, A., & Kalpić, D. (2013). *The Selection and Training framework (STF) for Managers in Business Innovation Transformation Projects - The Literature Review*. IEEE 2013. Centeris.
- Trad, A., & Kalpić, D. (2014). *The Selection, and Training Framework selection and training framework (STF) for Manager's in Business Innovation Transformation Projects – Educational Recommendations*. EDEN.
- Trad, A., & Kalpić, D. (2015a). *The (e)Business Transformation Framework for (e)commerce Architecture-Modelling Projects. Encyclopedia of E-Commerce Development, Implementation, and Management*. IGI-Global.
- Trad, A., & Kalpić, D. (2015b). *The (e)Business Transformation Framework for (e)commerce Control and Monitoring Pattern. Encyclopedia of E-Commerce Development, Implementation, and Management*. IGI-Global.

Tripathi, G. (2011). *SOA with TOGAF*. Retrieved from <http://www.theserverside.com/tip/SOA-with-TOGAF>

Trujillo, J., & Luj'an-Mora, S. (2003). *A UML Based Approach for Modelling ETL Processes in Data Warehouses*. Dept. de Lenguajes y Sistemas Informáticos. Universidad de Alicante.

Van Deursen, D., Poppe, C., Martens, G., Mannens, E., & Walle, R. (2008). *XML to RDF Conversion: A Generic Approach. Automated solutions for Cross Media Content and Multi-channel Distribution*. Dept. of Electron. & Inf. Syst. - Multimedia Lab., Ghent Univ. - IBBT, AXMEDIS '08. IEEE.

Vella, A., Corne, D., & Murphy, C. (2009). *Hyper-heuristic decision tree induction*. Sch. of MACS, Heriot-Watt Univ., Edinburgh, UK. Nature & Biologically Inspired Computing, 2009. NaBIC 2009. World Congress.

Wes, W. (2001). Beyond grounded theory: The use of a heuristic approach to qualitative research. In *Counselling and Psychotherapy Research: Linking research with practice*. Taylor and Francis.

ADDITIONAL READING

Farhoomand, A. (2004). *Managing Business Engineering transformation*. Palgrave Macmillan.

Farhoomand, A., & Tsang, S. (2006). *Global Corporate Social Responsibility Vs. Local Legal Compliance-A Case of Internet Censorship in China*. Hong Kong: Asia Case Research Center, University of Hong Kong.

KEY TERMS AND DEFINITIONS

Business Engineering: The development and implementation of business solutions using business models to business processes and organizational structure.

CSFs: Can be used to manage the statuses and gaps in various project plans and gives the Projects the capacity to proactively and automatically recognize erroneous building blocks and to just-in-time reschedule the Project plan(s).

Data Modelling: The process of developing data models for the business transformation process.

Enterprise Data Architecture: A data strategy that defines the objectives of the business transformation process's data collection management.

Environment: This research's framework.

Manager: Business Transformation Manager.

Contemporary Leadership Development in Kazakhstan

Gainiya Tazhina

University of International Business, Kazakhstan

Judith Parker

Teachers College, Columbia University, USA

Arslan Ivashov

Kazakh Ablai Khan University of International Relations and World Languages, USA

INTRODUCTION

Since the independence of Kazakhstan, 25 years ago, universities in the country have developed the system of educating and training managers and leaders for all spheres of the economy. Obviously the world-wide processes such, as globalization, integration, information technologies in higher education as well have their impact on the formation of leadership. As in many other countries in order to educate new managers and leaders, traditionally western models of human resource development were translated and adopted. This article is the continuation of the series of social-psychological researches in the Republic of Kazakhstan (RK). The co-authors began their collaboration in spring semesters of 2009 and 2010, when Dr. Tazhina was undergoing an internship at the Columbia University and together with Dr. Parker developed a joint international research project. Later, in 2012, Dr. Ivashov joined the research when he had internship at the NYU.

BACKGROUND

The Traditional Culture of Kazakhstan and Contemporaneousness/Globalization

Traditional Kazakh culture has historically been determined by the nomadic lifestyle and harsh

living conditions of this vast territory and became a regulatory mechanism for daily life. Clan membership, a clear age hierarchy and inter-familial relationships were a priority. Elders were at the top of the hierarchy and the family was and still is in charge of resolving social and psychological conflicts. (Masanov, 1998). Later Akshalova (2002) clarified, that “from outset, it should be asserted that within the Kazakhstani culture, there is respect to elders and seniors and hence, all the more, father leadership should be encouraged and practiced in small business management. After all, it is a benevolent leadership style that benefits all parties concerned – the leaders, the people and the organizations”.

For this reason, from the point of view of the identity formation and managerial education, it is important to consider the impact of globalization and a joint conflict between western mind-set and traditional values (Tazhina, 2010). The latter and a number of other social phenomena led to risen demand for applied psychology, which takes form of social-psychological training of skills such as leadership, effective communication, motivation, etc.

While individual leadership is important, the context of the organization can enhance or inhibit the leadership of its members. Dorfman and House (2004) report that at the first GLOBE research conference in 1994, there was a consensus of the 54 researchers from 38 countries (of which Kazakhstan was one) about a definition of

DOI: 10.4018/978-1-5225-2255-3.ch488

organizational leadership. “the ability of an individual to influence, motivate, and enable others to contribute toward the effectiveness and success of the organizations of which they are members” (p. 56). GLOBE is international project, which is aimed to seek answers to questions like: in what ways can community culture and organizational culture affect behavior of leaders/managers and company’s/organization’s development, commenting on the effectiveness of both?

Contemporarily, Kazakhstan is actively involved in the processes of globalization and integration. The following are just some examples of such activity: OSCE Chairmanship of the RK in 2010 and the largest international exhibition of the decade, EXPO, is to be held in the capital of Kazakhstan in 2017. In this context the necessity of studying leadership and its development in each segment of the society and economy is evident.

Leadership Issues

The American historian David McCullough (2008) emphasizes that “We need leaders and not just political leaders. We need leaders in every field, in every institution, in all kinds of situations. We need to educate our young people to be leaders” (p. 45). He asserts that “the great leadership lessons don’t change” (p. 45) and cites numerous historical figures that he considers demonstrate these fundamental qualities.

In an effort to analyze leadership itself, numerous authors and researchers have delineated types of leadership. Lojeski and Reilly (2008) begin with noting that the most “traditional view of leadership is that of a manager: one who monitors, controls, and, more importantly, rewards desired behavior and punishes undesired behavior. However, they label several other types of leaders. They note that “Charismatic leaders are people who can make the emotional connection”. (p. 124). “Transformational leadership includes four kinds of behavior: 1. Idealized influence refers to actions that demonstrate vision, values, and beliefs and creates a sense of identification with the leader among

followers. 2. Individualized consideration involves coaching and encouraging and also promoting each individual’s belief that they can be successful. 3. Inspirational motivation involves creating a clear and appealing vision and serving as a model for desired behavior. 4. Intellectual stimulation involves making the team aware of problems and bringing new ideas so that followers become engaged in finding solutions. Transformational leaders also create an emotional connection with their followers so that their behaviors contribute to followers’ sense of ownership and commitment to the team’s goals and tasks” (p. 124-125).

However, putting these types of leaders into the context of a technology rich society, context, Lojeski, & Reilly (2008) note that transactional, charismatic, and transformational views of leadership were developed before the Digital Age, when organizations and teams were likely to be collocated and more culturally homogeneous. They suggest that today’s organizations need the ambassadorial leader. “The ambassadorial leader is a boundary spanner. Like an ambassador, the leader must span geographic, cultural, and organizational boundaries and foster trusting relationships among disparate groups of people” (p. 130). Wiseman and McKeown (2010) categorize leaders as multipliers and diminishers. They explain that “some leaders make us better and smarter. They bring out our intelligence”; some “focus on their own intelligence and their resolve to be the smartest person in the room had a diminishing effect on everyone else”. Their “research showed, that most of us fall along this spectrum and have the ability to move toward the side of Multiplier” (p. 29).

Culture and Leadership

Leadership is inextricably intertwined with both national and organizational culture. Hofstede (2001) explains the relationship between national culture and leadership by suggesting that “ideas about leadership reflect the dominant culture of a country. Asking people to describe the qualities of a good leader is in fact another way of asking

them to describe their culture. “The leader is a culture hero, in the sense of a model for behavior” (p. 388). Schein (1996) notes, that “leadership is the fundamental process by which organizational cultures are formed and changed” (p. ix). However, House & Javidan ((2004) emphasize that “we are just beginning to understand how culture influences leadership and organizational processes” (p. 9). They note that “given the increasing globalization of industrial organizations and the growing interdependencies among nations, the need for a better understanding of cultural influences on leadership and organizational practices has never been greater” (p. 10).

Dorfman and House (2004) cite numerous studies that

... document the astonishing diversity of organizational practices worldwide, of which many are acceptable and considered effective in one country and ineffective in a neighboring country. (p. 52)

They also provide examples of non-western conceptualizations of leadership. However, they continue by suggesting that this is similar to “task-oriented” and “support-oriented” leadership functions in Western leadership theories.

Interesting conclusions were made by Grachev (1999) and Ardichvili & Kuchinke (2002) about the corporate practices (“as they are”) in former USSR countries, which “take marginal place in international system of coordinates in the majority of parameters of GLOBE”. Therefore, this confirms the unbalanced and frequently unpredictable form of leadership/management style inside a company in these countries when working with foreign management.

Parker (1997) adds an example of global educational practices in the technical community at 3M. She notes that the growth of business in the Asia region in the 1990’s demanded educational programs in leadership skills. A number of learning activities were developed by her in partnership with consultants in Europe and Asia to meet these needs. The objective of their partnership was to

maintain sensitivity to and balance between the Western culture of the Minnesota based company and the cultures of the Asian countries.

Emphasizing the importance of local culture, Tazhina (2010) reported “it is vital to take into account traditional culture in training of managers. Only local trainers can understand the culture and mentality of the people living in their country and provide the most effective training” (p.2).

While attention to cultural differences is important, Dorfman & House (2004) also suggest that

... there is a substantial amount of cross-national convergence of management practices, values, and beliefs. This occurs as a result of interactions among organizations engaged in cross-border trading, and wide-spread proliferation of management education programs that reflect Western assumptions, values, and practices. (p. 53)

They cite Hofstede’s studies from the early 1970’s and their replication over the decades as evidence that change in these fundamental cultural values appear to be very slow and quite resistant to convergence forces.

Issues of national culture are complicated by the superposition of the culture of the organization itself. Schein (2004) explains that

... leaders externalize their own assumptions and embed them gradually/consistently in the mission, goals, structures, and working procedures of the group. Whether we call these basic assumptions the guiding beliefs, the theories-in-use, the mental models, the basic principles, they become major elements of the emerging culture of the organization. (pp. 406-407)

However, one additional influencer over this time has been the introduction of technology at varying rates in different cultures. Schein (2004) noted that

The introduction of any new technology into an organization, or society can be seen as a culture

change problem. Occupations typically build their practices, values, and basic self-image around their underlying technology. (p. 36)

As new technology is introduced into learning and leadership additional adaptations will need to be made. Schein (1985) also noted that

We should of course recognize that many people resist new technologies because they correctly sense that their cultural assumptions are being challenged and threatened. Technological changes not only disrupt behavioral patterns but force one to change underlying assumptions. (p. 288)

This will be an essential skill for the development of leadership in any country or organization.

CURRENT PERSPECTIVES

The initial research by Parker & Tazhina (2011) looked at leadership models and career orientations in Kazakhstan based on the selection of 60 graduate students, working in various industries. The research was redirected in 2014 involving 189 graduate students. That time the selection included 117 females and 72 males aged 21 to 34, who were managers in various companies. Additionally, 60% of the students are operational level managers. 15% of the participants are middle level managers and lead HDR, marketing, supply, and customer service departments. 25% of the participants are entrepreneurs. In order to gather the data, three instruments were chosen: questionnaire on career anchors inventory, leadership style test, questionnaire on interpersonal diagnosis of personality.

Career Anchors Inventory

The career anchors inventory was developed by Edgar Schein (1985), who explains that the anchor was not only an indication of “the different things people are looking for in their career” (p. 248) but referred more strongly to “those things that

the person would not give up if forced to make a choice” (p.250). The inventory provides individuals with information about their primary areas of competence, motives, and career values. We use Russian version of inventory by Chiker (2006). Schein (1996) originally identified 5 anchors and then added 3 anchors. Chiker (p.105) divided the anchor security/stability into two and identified these as security/stability of job and security/stability of residence. The 9 career anchors are described below:

- **Technical/Functional Competence:** Like being challenged, using personal skill to meet the challenge, and becoming the expert.
- **General Managerial Competence:** Like problem-solving, dealing with other people and responsibility.
- **Autonomy/Independence:** Need to work alone, under their own rules and at their own pace.
- **Security/Stability of Job:** Seek security and stability in their jobs. They dislike personal risk, and often identify with their organization.
- **Security/Stability of Residence:** Try to stay in the same place of residence with minimum relocations and business trips. Moving for these people is unacceptable and uncomfortable.
- **Entrepreneurial Creativity:** Like to be creative, inventive and run their own businesses. They will share the workload but need ownership and are easily bored. They consider wealth a sign of success.
- **Service/Dedication to a Cause:** Driven by how they can help others.
- **Pure Challenge:** Need constant difficult problems that they can solve. They will change jobs if current one gets boring so their career can be varied.
- **Lifestyle:** Integrate life and work with focus on personal passions.

Table 1. Career anchors inventory

Career Anchor		Number of Times This Was the Dominant Anchor for a Student, %		Mean of Career Orientation	Standard Deviation
		2010 N=60	2014 N=189	2014	2014
Technical/Functional competence		1,3	1,6	5,3	0,8
General Managerial competence		21	17	6,8	0,7
Autonomy/Independence		12	7,9	6,7	1,1
Security/ Stability	of job	1,3	32	7,3	0,5
	of residence	2,6	4,8	4,4	0,7
Entrepreneurial Creativity		29	11	6,7	1,0
Service/Dedication to a cause		15	11	6,9	1,1
Pure Challenge		5,1	3,2	6,0	0,9
Lifestyle		13	11	7,1	0,5

The individual student scores for each of the 9 career anchors were calculated. The highest score indicated the preferred anchor. The scores were compiled to develop a frequency distribution for each anchor by percentage in 2010 and 2014, means and their standard deviation are given in Table 1.

Of particular interest are the extreme high and low numbers. In 2010 the two highest anchors, entrepreneurial creativity and general managerial competence might be attributed to the fact that the sample consisted of students in finance. In 2014 the majority of the students preferred the security/stability of job. This indicator is 25 times higher than the figure in 2010. This could possibly be due to the economic crisis in Kazakhstan, when the factor of employment became crucial for an individual. The figures for career anchors are 1.2-2.6 times lower compared to 2010's indicators excepting technical/functional competence and security/stability of residence. However, It might be considered if the lowest scores in 2010/2014, technical/functional competence, security/stability of residence, and pure challenge might be characteristics consistent with the national culture. Considering that even the oldest of these graduate students had lived through the shift from the dominance by the Soviet Union

to an independent Kazakhstan, to globalization/internalization being their personal best, needing these competences for a career. These findings are in sharp contrast to Schein's (1996) findings in the 1970's and 1980's in which in USA "roughly 25 percent of ... populations anchored in 'general management,' 25 percent in 'technical/functional competence,' 10 percent each in 'autonomy' and 'security' and the rest spread across the remaining anchors" (p. 81).

Leadership Questionnaire

The leadership questionnaire is based on the Blake & Mouton Leadership inventory which rates people on their concern with the task and people. Blake & Mouton developed a matrix on which the scores are plotted in order to identify one of 5 leadership behaviors.

- **Impoverished (Low Task and Low People Score):** Exerts and expects minimal effort or concern for either staff satisfaction or work targets.
- **Country Club (Low Task and High People Score):** Attentive to staff's relationships and the work culture but at the expense of production results.

- **Middle of the Road (Politician):** A compromiser who wants to maintain the status quo.
- **Authoritarian (High Task and Low People Score):** Focus on achievement. People are only a commodity used to achieve results.
- **Team Leader (High Task and High People Score):** Achieves high work performance by inspiring involvement and leading people to become dedicated to the organizations goals.

The people and task preference scores were calculated for the students that completed this inventory. A frequency percentage distribution of these styles is shown in Table 2. The preferred leadership style was authoritarian. It is constant phenomena for both measurements. Moreover, combining categories indicate that in 2010 only 11,4% of participants scored in the “high people” range compared to 39,4% who scored in the “high task” range. In 2014, the “high people” range increased up to 18,9% in contrast to the “high task” range that enlarged to 46,5% of participants.

While both the samples of 2010 and 2014 are not sufficiently large to draw conclusions, details indicated some interesting facts. Authoritarian style is the most prominent in the selection. The majority of the graduate students prefer this style and its 2010/2014 indicators are relatively similar. Further analysis has shown that number of managers with Team, Country Club and Middle of the Road styles has risen by 1.8-2.2-3.2 times respectively. The prevailing styles, Authoritarian and Middle of the Road, are relatively similar for both genders. In 2014 the number of students with Impoverished style has dropped 1.5 times compared to 2010.

Interpersonal Diagnosis of Personality

To conduct the interpersonal diagnosis of personality (IDP) amongst managers the Leary's (1957) approach was chosen, but in its Russian-language modification by Sobchik (2004). IDP was used for managers to draw their own portrait as a leader. A respondent is asked to decide whether 128 statements represent his/her Real Self.

Table 2. Leadership questionnaire frequency distribution

Leadership Style	Indicator	Frequency, %		Mean and Standard Deviation, 2014	
		2010	2014	Concern with People	Concern with the Task
Country Club	High People/Low Task	1,6	3,2	7,0±1,4	3,5±0,7
Country Club/Team	High People	3,2	3,7		
Country Club/Impoverished	Low Task	11,5	6,9		
Impoverished	Low Task/ Low People	18,0	12	2,2±0,9	3,0±1,0
Impoverished/ Authoritarian	Low People	19,7	6,7		
Authoritarian	Low People/ High Task	27,9	26	2,5 ±0,8	6,4±1,4
Team/Authoritarian	High Task	4,9	8,5		
Team	High Task/ High People	6,6	12	6,8±0,6	6,8±1,2
Middle of the Road		6,6	21	5,0±1,1	5,1±0,9

Characteristics that don't go beyond 8 points are inherent to well-balanced individuals. Characteristics that exceed 8 points indicate the accentuation of attributes, that are described by a given octant. 9-12 points imply a high severity of the behavior, even leading up to an extreme.

In order to demonstrate the main social orientations Leary (p. 214) has developed a provisional diagram/circle divided into sectors. In this circle the horizontal and vertical axes are marked with 4 orientations: dominance-submission, friendliness-hostility. In turn, these sectors are divided into 8 more particular cases. The sum of points in each orientation is converted into an index, where the vertical (dominance-submission) and horizontal (friendliness-hostility) axes are dominating. The interval between deduced indicators and center of the circle points out to the adaptability or extremity of interpersonal behavior.

Leary (p. 168) has identified the following types of interpersonal behavior. The modified titles by Sobchik (p. 122) are shown in brackets.

1. Authoritarian (imperious-leading),
2. Egoistic (independent-dominating),
3. Aggressive (straightforward-aggressive),
4. Suspicious (untrusting-skeptical),
5. Submissive (meekly-shy),
6. Dependent (dependent-obedient),
7. Friendly (cooperating-conventional),
8. Altruistic (responsibly-generous).

All styles of interpersonal behavior are present amongst participating managers, with the exception of severely prominent maladaptive variation of untrusting-skeptical style. The means of interpersonal behavior styles were calculated for each leadership style (Table 3).

It can be seen that managers with Authoritarian and Country Club leadership styles are relatively balanced in their interpersonal interactions, rather than managers with other leadership styles. Managers with Impoverished and Middle of the Road leadership styles are extremely authoritarian. And managers with Team leadership style are the most contradictory with polar extreme tendencies towards authoritarianism and altruism. Based on this information Figure 1 demonstrates the circles of interpersonal behavior for five leadership styles.

Additionally, friendliness and dominance indices were calculated as well as the prevailing type of attitude to the peers. According to Leary (p. 174), the following formulae allow deducing dominance index (vector V) and friendliness index (vector G):

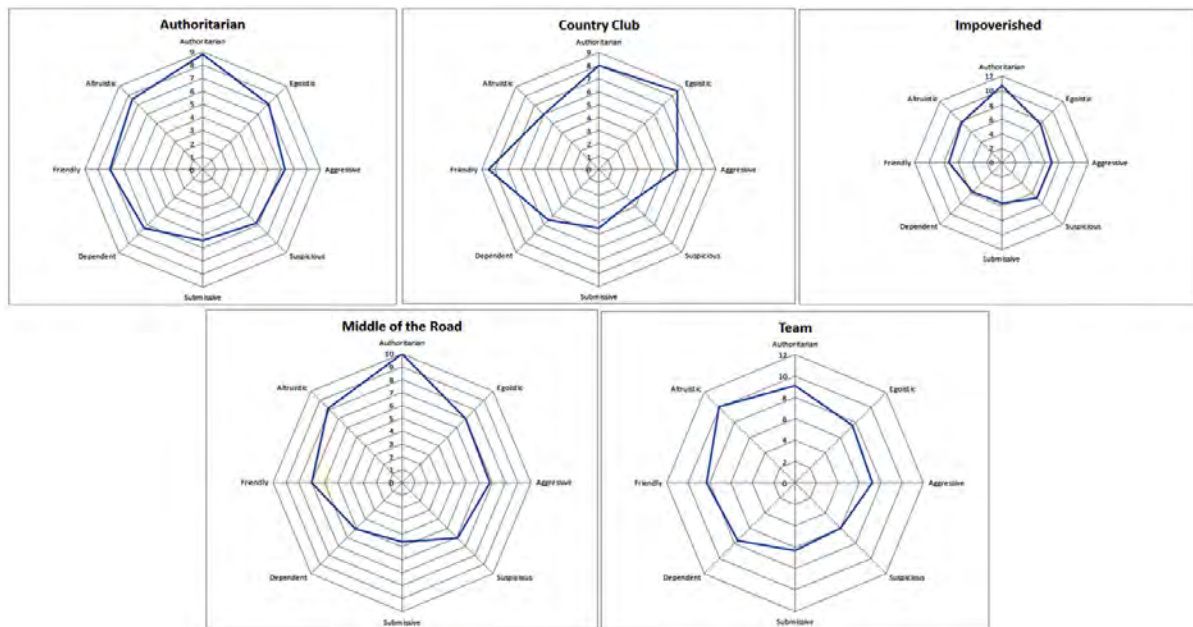
- Dominance $V = (I - V) + 0,7 \times (VIII + II - VI)$
- Friendliness $G = (VII - III) + 0,7 \times (VIII + VII - II)$

A result, that has a positive or negative direction incline, reveals the prevailing tendencies. According to Sobchik (p. 128), given recipes are

Table 3. Means of interpersonal styles by Leary for each leadership style by Blake & Mouton

Leadership Style by Blake & Mouton	Interpersonal Styles by Leary							
	Authoritarian	Egoistic	Aggressive	Suspicious	Submissive	Dependent	Friendly	Altruistic
Authoritarian	8,8	7,1	6,3	5,8	5,4	6,3	7,1	7,6
Country Club	8,0	8,5	6,0	3,5	4,5	5,5	8,5	6,0
Impoverished	10,7	7,6	6,9	6,9	5,6	5,7	7,3	7,8
Middle of the Road	10,0	7,0	6,8	6,1	4,6	5,1	7,0	8,1
Team	9,1	7,6	7,2	6	6,3	7,6	8,3	10,1

Figure 1. The Leary Circles of interpersonal behavior for five leadership styles



especially informative when generalizing data for a particular group, rather than a single individual (Table 4).

For each leadership style the prevailing tendency of IDP is dominance. Dominance index surpasses the friendliness index 2.5-9.5 times. Interestingly, the highest dominance figures are observed for impoverished and team leadership styles.

Table 4. Dominance and friendliness indices in accordance with IDP methodology vs leadership styles by Blake-Mouton

Leadership Style by Blake & Mouton	Prevailing Tendencies of Interpersonal Behaviour	
	Dominance	Friendliness
Authoritarian	9,35	3,21
Country Club	9,8	2,31
Impoverished	11,89	-0,3
Team	12,27	0,3
Middle of the Road	9,87	3,97

CORRELATIONS

SPSS statistical software was used to calculate the Spearman's rho correlations between each of measured scales and other numbers of the other instrument (Table 5).

There is a significant positive correlation between MT participants and those with security/stability of job (SEJ), entrepreneurial creative (EC), and technical/functional (TF) career anchors. There are also notable positive correlations between MP participants and those with lifestyle; and negative correlations with entrepreneurial creative career anchors. The similar correlations were found out in 2010th sample.

Within significant positive correlations between interpersonal relations and career anchor categories, the highest correspondences are between dominant/aggressive/friendly interpersonal relations and general managerial competence career anchor; between competitive interpersonal relations and security/stability of job (SEJ); between altruistic interpersonal relations and entrepreneurial creative career anchors.

Table 5. Correlations

	MP	MT	Authoritarian	Egoistic	Aggressive	Suspicious	Submissive	Dependent	Friendly	Altruistic
TF		.356*	.391*							
GM			.615**	.459*	.455**		.307**		.510**	
AU								-.489*		
SEJ		.410**		.591*						
SER								.391*		
EC	-.351*	.396**	.440**							.409**
SV										
PC					.387*					
LS	.301**						.309**		.483*	.375*

TF: Technical/Functional competence

GM: General Managerial competence

AU: Autonomy/Independence

SEJ: Security/Stability of job

SER: Security/Stability of residence

EC: Entrepreneurial Creativity

SV: Service/Dedication to a cause

PC: Pure Challenge

LS: Lifestyle

MP: People oriented management style

MT: Task oriented management style

*Correlation is significant at the.05 level

**Correlation is significant at the.01 level

It is remarkable that dependent interpersonal relations correlate negatively with autonomy/independence, and positively with security/stability of residence career anchors.

FUTURE RESEARCH DIRECTIONS

The obtained data about managers of different levels complement the information on leadership in Kazakhstan. Tazhina (2010) provides specific recommendations for the future within the area of applied psychology and leadership. She notes that “While Kazakhstan has the necessary resources; it needs fundamental development in psychological research”. Also Low (2013) examined leadership in relation to small business leadership and management in Kazakhstan. “The business owners/leaders or entrepreneurs are to extend their personal influence so that there can be positive effects with results attained for organizations. In

fact, there have been very few studies on leadership in small businesses in the Republic of Kazakhstan; hence the need for such study” (p.225).

Technology and e-learning will certainly be a part of future impacting leadership practice and possibilities. Fedoroff (2012) comments that “today, countries all over the world aspire to have high-skill, high-value economies, to become ‘knowledge societies’ and ‘knowledge economies.’ The extraordinary value of knowledge is, of course, that there are no limits to its growth or the value it can generate. Going beyond local benefit, it is increasingly important to create a global knowledge society” (p.503).

Tazhina & Parker (2013) surveyed the role of distance learning technologies in preparing global leaders in Kazakhstan. They noted some characteristics unique to the country, but also some aspects shared with other countries globally. “Kazakhstan is approaching its future with bold steps in the use of technology to educate its

workforce. The model used at UIB for distributing the knowledge of its faculty from the campus in Almaty to 5 regional centers around the country may well become a model for other universities. And the impact of the skills learned in a distance learning environment will influence the future of the nation and its workforce” (p.480).

Raising leaders in Kazakhstan is the issue of a day. Thus, there was held the International Scientific-Practical Conference for young scholars “Intellectual Leadership: New Stage in Development of Kazakhstan” by the Academy of Public Administration in February of 2016. Researchers in their early career, PhD and master students have discussed how to develop creative activity, principles of the professional state, priorities of the new economic and foreign policy of Kazakhstan in the context of globalization during breakout sessions.

CONCLUSION

This article is the continuation of series on social-psychological research in Kazakhstan. The obtained data is of a great interest and partially agrees with the existing few researches on leadership in Kazakhstan. Low (p.226) identified father leadership in Kazakhstan. “It can be said that Kazakhstani people tend to expect that employers provide transportation and this is a standard practice in many banks and most manufacturing companies. The father (leader) can be strict (task-oriented, and in this way, things get done) and may discipline yet he is caring (relationship-based)”.

In this research young graduate students, who were taught about western leadership theories, using the latest information technologies, and working as managers of various levels, still tend to choose traditional leadership values. This is evident when team leaders can simultaneously be highly dominant/authoritarian and altruistic. As Low (p.226) stated “as the country develops,

people will acquire more skills and further education, and with adequate numbers of skilled business executives, more will be exposed to individualism and the *laissez-faire* style. Nonetheless,...father leadership, coupled with concern for the staff's welfare, can in a way, be seen as a benevolent form of leadership; and that is gaining currency especially among the more educated Kazakhstanis”.

Ardichvili & Kuchinke (2002) compared leadership styles and cultural values in their cross-cultural study of over 4,000 managerial and non-managerial employees in several countries including Kazakhstan. “Any attempts to develop leadership training programs ..., or to conduct organization development interventions based on Western ... theories, should be tailored to each country's specificity, and the ... Republics should not be treated as homogeneous in their management and work cultures.

REFERENCES

- Akshalova, B. (2002). *Kazakh Traditions and Ways*. Almaty: Dyke-Press.
- Ardichvili, A., & Kuchinke, K. (2002). Leadership styles and cultural values among managers and subordinates: A comparative study of four countries of the former Soviet Union, Germany, and the US. *Human Resource Development International*, 1(5), 99–117. doi:10.1080/13678860110046225
- Chiker, V. (2006). *Psychological diagnostics of the organization and staff*. SPb: Speech.
- Dorfman, P., & House, R. (2004). Cultural influences on organizational leadership: literature review, theoretical rationale, and GLOBE project goals. In R. House, P. Hanges, M. Javidan, P. Dorfman, & V. Gupta (Eds.), *Culture, leadership, and organizations: The GLOBE study of 62 societies* (pp. 51–73). Thousand Oaks, CA: Sage Publications.

- Fedoroff, N. (2012). The global knowledge society. *Journal of Science*, 335(6068), 503. doi:10.1126/science.1219498 PMID:22301283
- Grachev, M. (1999). Management in the international system of coordinates. *Economic Strategy*, 2, 19-32.
- Hofstede, G. (2001). *Culture's consequences: comparing values, behaviors, institutions, and organizations across nations*. Thousand Oaks, CA: Sage Publications.
- House, R., & Javidan, M. (2004). Overview of GLOBE. In R. House, P. Hanges, M. Javidan, P. Dorfman, & V. Gupta (Eds.), *Culture, leadership, and organizations: The GLOBE study of 62 societies* (pp. 9–28). Thousand Oaks, CA: Sage Publications.
- Leary, T. (1957). *Interpersonal Diagnosis of Personality*. The Ronald Press Company.
- Lojeski, K., & Reilly, R. (2008). *Transforming Leadership and Innovation in the Globally Integrated Enterprise*. Hoboken, NJ: John Wiley & Sons.
- Low, P. K. C. (2013). Father Leadership in Kazakhstan. In *Leading Successfully in Asia* (pp. 225-234). Berlin: Springer-Verlag. doi:10.1007/978-3-642-31287-8_8
- Masanov, N. (1998). *The role of clans in Kazakhstan today*. Retrieved September 03, 2010, from <http://www.jamestown.org>
- McCullough, D. (2008). Timeless leadership. *Harvard Business Review*, 86(5), 45–49. PMID:18411965
- Parker, J. (1997, September). *Open learning without frontiers – open learning across International boundaries*. Paper presented at the Open Learning for Business Conference, Birmingham, UK.
- Schein, E. (1996). Career anchors revisited: Implications for career development in the 21st century. *The Academy of Management Executive*, 10(4), 80–88.
- Schein, E. (2004). *Organizational culture and leadership*. San Francisco: Jossey-Bass Publishers.
- Sobchik, L. (2004). The theory of the leading trends as a basis for psychodiagnostic research methodology. *Moscow Psychological Journal*, 3, Retrieved April, 2004, from http://magazine.mospsy.ru/nomer3/sob_03.shtml
- Tazhina, G. (2010). Applied Psychology in Kazakhstan. In *Proceedings of The Scientific Congress “Kazakhstan in Globe: Socio-Economic and Cultural Aspects”* (pp. 124-129). Almaty: UIB Press.
- Tazhina, G., & Parker, J. (2013) Preparing Global Leaders in Kazakhstan through Distance Learning Technologies. In *Proceedings of the 10th KIMEP International Research Conference “Central Asia in the Global Community: Challenges and Opportunities”* (pp. 477-481). Almaty-Astana: Academpress.
- Wiseman, L., & McKeown, G. (2010). *Multipliers: How the best leaders make everyone smarter*. Harper Collins Publishers.

ADDITIONAL READING

Intellectual Leadership. New Stage in Development of Kazakhstan (2016). Conference for young scientists supported by Academy of Public Administration under the President of the Kazakhstan and the EU-funded project. Retrieved February 26, 2016, from <http://pa-academy.kz/index.php/en/home/1964-intellectual-leadership-new-stage-in-development-of-kazakhstan>

Nantz, K. (Ed.). (2014). *Business Education in Central Asia: Best Practices in Integrative Study and Teaching*. U.S.-Central Asia Education Foundation.

KEY TERMS AND DEFINITIONS

Career Anchor: “A person’s career anchor is his or her self-concept consisting of 1) self-perceived talents and abilities, 2) basic values, and most important, 3) the evolved sense of motives and needs as they pertain to the career”. (Schein, 1996, p. 80).

Cultural Dimensions: The characteristics of a national culture usually associated with those dimensions defined in Hofstede’s work.

Culture: The shared attitudes, beliefs, practices and values that characterize a group.

Leadership: The ability of an individual or group to promote the ideas and to guide the vision of a group or organization.

National Culture: The culture associated with a geographical/political region and its inhabitants.

Organizational Culture: The culture attributed to an organization and usually associated with the values of the founders and leaders.

Technology: The application of scientific advances usually involving electronic advances.

M

Empirical Verification of the Performance Measurement System

Aleksander Janeš

University of Primorska, Slovenia

INTRODUCTION

Ebm-papst Group is Germany's most sustainable company which has a global sales network and is worldwide innovation leader in electric fans and motors. The Ydria Motors LL Company (YM) is a member of the ebm-papst group and is organized as a competence center that produces and develops machines, appliances and AC and EC electric motors and exports as much as 98% of its production. The company is wholly owned by the ebm-papst group (Ydria Motors, 2015).

Practical experiences with the Performance Measurement System (PMS) show that it is difficult to determine transparent relations between the different perspectives. However, the implemented PMS doesn't enable the identification of all information on the relations (i.e. causalities) between process Key Performance Indicators (KPIs). In this manner company doesn't have transparent evaluation of resource inputs in efficiency of the implemented model in the management system (Janeš & Dolinšek, 2010; Waruhiu, 2014). In the light of its development, the company has to resolve the issues to outline its future organic growth. Thus, the YM Company has already faced the challenge of setting the KPIs, which are the central theme of represented research.

Primarily, the represented research is focused on the layout of the balanced scorecard (BSC) i.e. it's causal relations, using qualitative and quantitative methods. In this case study, the company's business performance sustainability understanding, which is based on comprehensive data tests and semi-structured interviews with managers, contributed to the selection of Engle-Granger

two-step method for assessment of the error correction mechanism between the KPIs.

BACKGROUND

One of the main purposes of implementing PMS is to communicate strategy throughout the organization. Among the number of PMS approaches, a dominant position was achieved by Kaplan and Norton's BSC; it has emerged as a new synthesis between the traditional financial-accounting system and efforts to achieve long-term competitive ability. The BSC system considers the traditional financial KPIs as well as leading KPIs of future performance. In this way, it provides key information about the activities of managers. The BSC is a PMS theorized by Kaplan and Norton which was first created as a performance measurement tool, which has later evolved into a PMS, and has subsequently become a comprehensive strategic management system (Barnabè, 2011; Janeš, 2014).

Measurement of the organization's performance represents a good practice and is an integral part of the organization's management in the accomplishment of its strategies and objectives. The BSC and associated KPIs are treated in practice as a process that supports the reviewing and changes of the measurement system in relation to changes in the business environment of the organization.

Given the framework of the strategy map, which consists of four perspectives, and within them a large number of related strategic objectives, the added value of the business processes is manifested in the form of chains of cause-effect relations ranging from nonfinancial and quantifiable KPIs in the

learning and growth perspective via processes and customers to a finance perspective (Waruhiu, 2014; Kaplan, 2012, p. 543). Cause-effect relations can be understood as a set of hypotheses that are taken to meet the strategic objectives. A strategy map is a diagram that describes how an organization creates value by connecting strategic objectives that are in explicit cause-effect relations with each.

Approach of setting practical objectives is being performed through a series of approximations. Keeping the maximization of the rate of return as the stakeholder's long term value, companies usually develop a number of subsidiary objectives which contribute in different ways to improvement in the return and which are also measurable in business practice. A company which meets high performance in most of its objectives will enhance its long-term rate of return. But it cannot be proved that the result will be a maximum possible overall return. Namely the problem of reconciling claims of conflicting objectives is common across all sectors. In business practice company's objectives are what the company wants from its key stakeholders. Essential nature of company objectives makes a logical connection between what stakeholders want from a company, i.e. "strategic factors", and what the company wants from its stakeholders, i.e. "organization objectives". Both sides together provide real strategy framework. For example one of the stakeholder groups represents customers from which company wants to achieve profitable revenue. If an objective is an expression of what an organization wants from a key stakeholder, it must involve behavior by a stakeholder and, a desired behavioral outcome (Kenny, 2012, pp. 42-43). Once developed, strategic objectives will shape an organization's strategies effectively, leading to sustainable success. It's important to remember that targets on organization objectives shape strategies; they also provide a means of assessing the effectiveness of those strategies. Namely, without clear and quantified objectives, any strategy can be set (Kenny, 2012).

One of the main areas that both the relevant literature and Kaplan and Norton themselves

identified as being critical, is related to the identification and assessment of causal relations which are essential within the BSC (Barnabè, 2011; Barnabè & Busco, 2012; Kaplan, 2012; Nørreklit, 2000; Waruhiu, 2014, p. 119). In this context, the causal relations have been in the center of survey interest, because they are providing better relations model between the four BSC perspectives (Bukh & Malmi, 2005, p. 96).

CASE DESCRIPTION AND DISCUSSION

Methodology

The research project of empirical verification of the PMS was performed as a single case study of modeling the BSC system for a manufacturing company and founded on the complementary use of quantitative and qualitative methods (Gummeson, 2000, p. 88). For this purpose, were obtained and used KPIs that were already monitored by the company on a monthly basis and ranges from 2005M01 to 2011M12. Since the exact legalities between the observed KPIs were not known, the information contained in the time series of observed KPIs were expressly applied. Based on the selected Engle-Granger two-step method, a framework for the layout of the BSC including EVIEWS statistical software was defined (Engle & Granger, 1987). The central part of the case study is the part in which the empirical evaluation of the BSC's layout with the Engle-Granger two-step method was performed. Before starting with the Error Correction Model (ECM) modeling, several basic statistic tests of the KPIs' time series, i.e. stationarity and cointegration, were performed in order to exclude the possibility of false regression (Engle & Granger, 1987; Granger, 2004; Hassouneh, Serra, & Bojnec, 2015).

The main objective of the research aimed at developing a quantitative approach that would be complementary with a qualitative approach of the BSC layout. For this purpose a single equation

ECM was developed with the Engle–Granger (1987) two-step method. In the first step, the cointegration was evaluated. The linear combination of observed KPIs represents a co-integration vector that could be stationary. If the cointegrating vector is unique, then the ordinary least squares (LS) method can be applied to estimate the relation among KPIs. In this case, the residuals from LS regression equation can be used in place of the error correction term (u_{t-i}) to proceed with the estimation of the short-term dynamic model. To this end, the initial model of KPIs was set; the regression was calculated by the LS method and the residual values of the potential co-integrating vector were saved. The residual values were tested for stationarity, which is needed for the stability of the final model but it also confirms co-integration relations between the KPIs.

In the second step, the ECMs were set up using the seasonal differences and different time lags of KPIs and residuals. According to Engle and Granger (1987), when the variables are cointegrated, there must be an error correction mechanism that describes the short-run dynamics of the cointegrated variables towards their long-run equilibrium values. In the presentation of the ECMs (Table 1), u_{t-i} is a lagged error correction term that measures the speed of adjustment to long-run equilibrium. In order to restore equilibrium, the sign of the coefficient of the, u_{t-i} is expected to be negative (Engle & Granger, 1987; Janeš & Faganel, 2013).

Application of the Engle-Granger two-step method on the KPIs was performed with the ECM approach and short and long term effects between them were estimated.

A strategic map of the YM Company that contains the causal relations between the KPIs has been set and confirmed with the executive management. The method of semi-structured interviews was supplemented with the participation and observation of the researcher (Angrosino & Mays de Pérez, 2000; Bititci, Mendibil, Nudurupati, Garengo, & Turner, 2006; Bukh & Malmi, 2005; Gosselin, 2005; Janeš, 2014; Kaplan, 2012).

Design of the Strategic Map

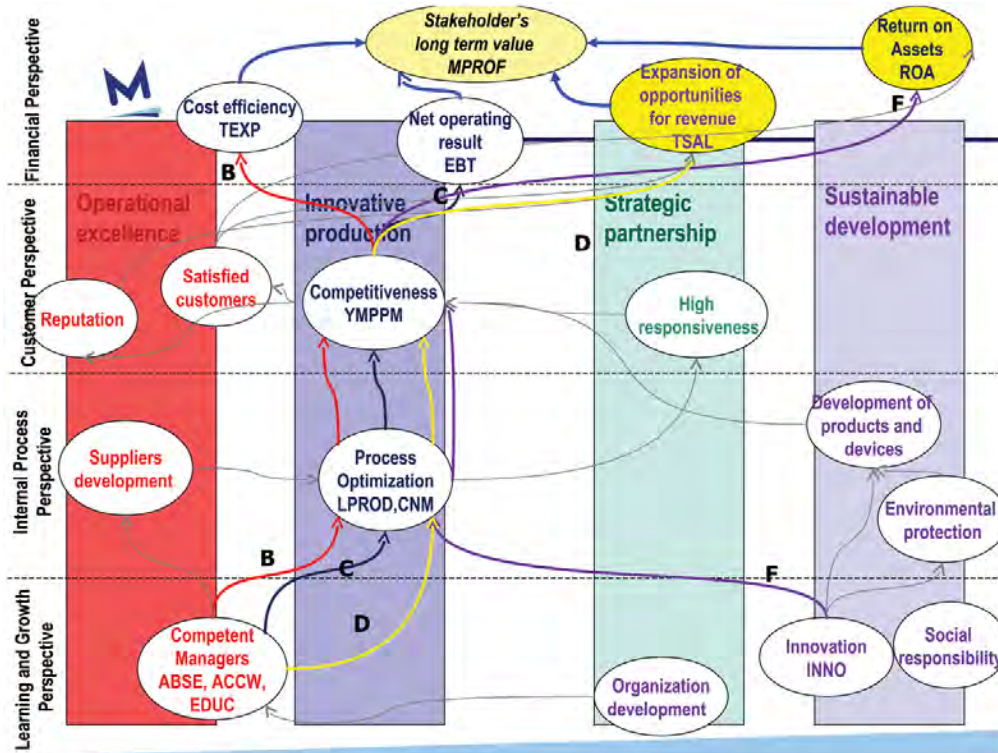
The method of semi-structured interviews in the form of workshops has enabled the qualitative analysis and modeling of the layout of the BSC. It was designed after a preliminary representation of the BSC concept to the executive management and with the use of prepared open questions, which directed the workshops. The strategic map has made it possible for the consensus between the executive managers and communication of chosen strategy in the range of “a story on one page”.

The cause-effect relations that are of interest for modeling and those that were modeled with the Engle-Granger method are clearly outlined in the strategic map (Figure 1). The latter are represented by the capital letters B, C, D and F. Relations that are not marked are those that we could not take into account due to different frequency of measurement.

The BSC model was designed with the involvement and consensus of the executive management in eight workshops, in which the following conclusions were reached:

- With the introduction of the BSC concept to the top management, the management has become familiar with the process of BSC modeling. The ultimate goal of the workshops was a designed BSC system, in which cause-effect relations were supported by an Engle-Granger method.
- With regard to the content of a destination statement, consensus was achieved, consisting of a single page and divided into four parts: financial expectations and expectations of the owner, strategic partnerships, business processes and sustainable development of the company. The destination statement, in a condensed form, represents a desired state of the company in the next three years from the viewpoint of executive management.
- Strategic themes represent decomposition of the overall strategy, because they con-

Figure 1. YM strategic map
Source: Adapted from Janeš, 2012.



tain essential parts, they define business processes that add value for stakeholders and enable the classification of strategic objectives (Figure 1).

$$\Delta MPROF_t = \alpha_0 + \sum_{i=1}^n \alpha_1 \Delta ABSE_{t-i} + \sum_{i=1}^n \alpha_2 \Delta ACCW_{t-i} + \sum_{i=1}^n \alpha_3 \Delta LPROD_{t-i} + \sum_{i=1}^n \alpha_4 \Delta YMPPM_{t-i} + \alpha_5 u_{t-i} + \varepsilon_t \quad (1)$$

Modeling of the Strategic Map

Due to the limited space, only some of the findings from the ECM analysis are represented. The use of the ECM approach can be seen on the case of the monthly profit (MPROF). From the Equation 1 it is evident that MPROF of the company depends only on internal KPIs. This is because the YM Company acts as a contract manufacturer. The production price of products is based on costs and is determined by the YM Company, whereas the market price of products is determined by the owner.

Since the monthly data were available, 12th time lag of the KPI's was included in the initial model- Equation 1.

The structure of the lags was evaluated with the criteria for determining the order of time lags, as are autocorrelation function, Q-statistics and correlograms. By testing various time lags of KPIs on the right side of the equation and the reduction of autocorrelation, respectively, monitoring the Durbin-Watson statistics, the final ECM was identified. The results from Table 1 can be interpreted as follows: the optimization of the MPROF is influenced by the MPROF from previous periods, absenteeism (ABSE), accidents at work (ACCW), labor productivity (LPROD), and the reduced parts per million (YMPPM). Short-term changes in the independent KPIs disrupt a balance between indicators of the ECM and cause the change of

Table 1. Comparison between the ECMs

Dependent Indicators	Monthly Profit (MPROF)*	Total Expenses (TEXP)	Earnings before Taxes (EBT)	Total Sales (TSAL)*	Return on Assets (ROA)
u_{t-i}	-0.995906	-1.004654	-1.000248	-0.983107	-0.999435
R ²	0.995342	0.998118	0.998787	0.973861	0.999990
Adj.R ²	0.994913	0.997945	0.998693	0.971824	0.999990
Akaike	17.08468	16.23047	15.60983	13.23195	-5.422925
Schwarz	17.31619	16.45181	15.81240	13.43452	-5.249295
Hannan-Quinn	17.17774	16.31858	15.69126	13.31338	-5.353127
Durbin-Watson	2.190392	2.055059	2.171403	2.350735	2.130595
Johansen test (sig.0.05)	1cointeg.eq.	1cointeg.eq.	1cointeg.eq.	2cointeg.eq.	3cointeg.eq.

Janež, 2012.

Note: The least squares method is used. The data sample ranges from 2005M01 to 2011M12. In analyses, the following KPIs were used: ABSE, ACCW, cost of net-material (CNM), earnings before taxes (EBT), education (EDUC), innovation (INNO), LPROD, MPROF, return on assets (ROA), total expenses (TEXP), total sales (TSAL) and YMPPM. * Asterisk means that there is a dependent indicator on the right side of the equation.

indicators to be corrected in the next period. The error correction mechanism of the MPROF in the next period again tends to long-term equilibrium.

The ECM result represented in Table 1 confirms the presence of an error correction mechanism, which indicates a relatively stable model of short-term dynamics of the factors that affect the MPROF. The final ECM for the MPROF reflects Granger's causality caused by the independent KPIs.

All the final ECM models from Table 1 were designed with KPIs according to the labeled causal relations between KPIs that are represented in Figure 1. From the results shown in Table 1, it is evident that all u_{t-i} have an expected negative sign.

From a comparison of the results of all models, is seen a high adjusted coefficient of determination (Adj. R²), which affects the bias reduction in the estimated co-integrated vector. In addition, the Durbin-Watson statistics, which in all cases have a value of about two, means that the impact of autocorrelation was significantly reduced.

Moreover, in all ECMs there are minimum values of Akaike, Schwarz and Hannan-Quinn

criteria, which confirm the suitability of selected lags in the models.

The assessment and identification of the ECMs was performed using the Engle-Granger method based on a unique cointegration equation (Johansen's cointegration test in Table 1).

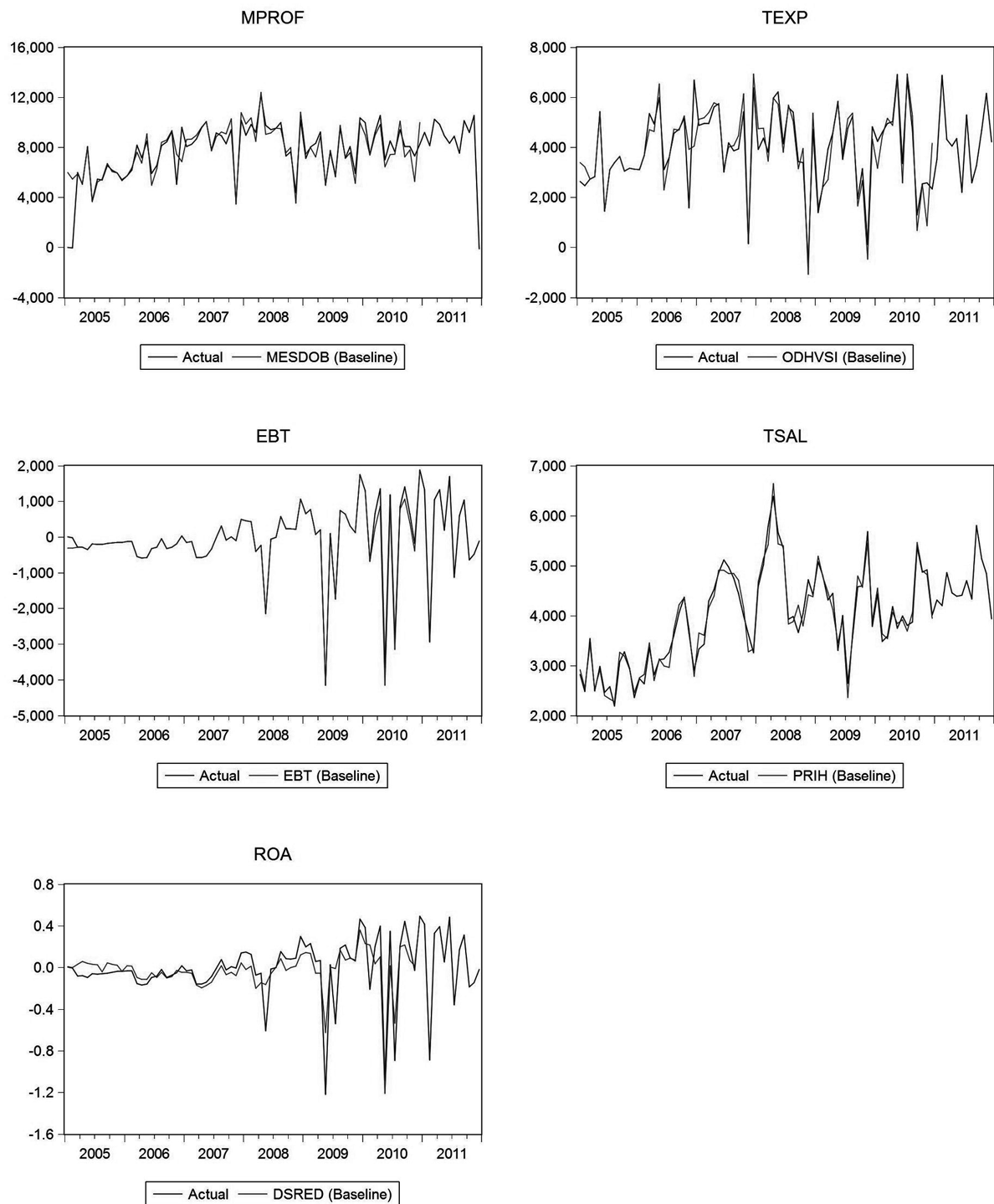
A number of ECM residuals tests regarding autocorrelation were undertaken; Q-stat and LM tests have shown the presence of some autocorrelation, but without significant influence on the final ECM models.

Tests for the presence of residual heteroscedasticity showed that it is mainly present in the ECM of TSAL. In other models, there is no heteroscedasticity or it is present to a non-significant extent.

Chow's test of coefficients stability showed that from 2009 to 2010, when Slovenia was already in the economic crisis (Verbič & Črnigoj, 2014), the coefficients of equations were stable. In the case of a ROA model, the test showed the existence of a breakpoint and in the case of TEXP the model stability was not confirmed for the 2010. For this reason, Quandt-Andrew's test of stability was performed, which has shown the existence of a breakpoint in model for TEXP. Therefore, the

Figure 2. Static-deterministic forecast simulation for one period ahead

Source: Adapted from Janeš, 2012.



Chow Forecast Test was conducted which showed the stability of the coefficients of the indicators in the period from 2009 to 2010, with the exception of the model of EBT and in the case of a TEXP model, stability was confirmed only for the period between 2006 and 2009. In the end, the recursive estimation of coefficients was carried out. The latter showed that in all cases ECMs coefficients are stable in the period from 2008 to 2010.

Based on the results of all performed tests, was ascertained the presence of a certain proportion of autocorrelation, heteroscedasticity and some instability in the coefficients of the individual models. However, based on the number of performed tests and their statistical significance, the ECM models were suitable for the merger in a linear forecasting simulation model.

Identification of the Linear Model

The identity of the linear model is that for which it is expected to prevail in the application in real cases. ECM equations are derived from statistical estimation procedures, while the identity is built on the basis of grounded relations between KPIs. With testing the forecasting capability the criteria of model adequacy are fulfilled as the forecasts are compared to the actual values of the KPI's time series. The linear model describes the linear relations between stakeholder's long-term value-MPROF KPI and the final ECMs (Equation 2).

$$MPROF = TEXP + EBT + TSAL + ROA \quad (2)$$

In order to test the linear model of MPROF, a static-deterministic forecast simulation for one period ahead was performed. In the case of static-deterministic forecast simulation the model in the selected period was relatively reliable with some minor to medium deviations. Therefore, it can be asserted that the linear model of MPROF is relatively suitable for short-term forecasts with a static-deterministic simulation. That was also confirmed with the Theil's coefficient of inequality, which represents the assessment of the

forecasting capacity and indicates the adequacy of the models. The results of the test are the following: for the MPROF it is equal to 0.108529, for the TEXP it is equal to 0.060058, for the EBT it is equal to 0.257273, for the TSAL it is equal to 0.077669 and for the ROA it is equal to 0.108529. It was noticed that in the long run, in 2011, the model no longer follows the actual values, which is a consequence of a relatively short period of data sample; therefore the simulation is no longer adapted to the actual values (Figure 2).

To fulfill the purpose of empirical verification of the BSC system, a long-term stochastic-dynamic forecast simulation was also performed. Based on the simulation results, it was found that the long-term stochastic-dynamic forecast simulation model for many periods in advance is not adequate. In addition, the Theil's coefficient of inequality test is the following: for the MPROF it is equal to 0.236994 for TEXP is equal to 0.091690, for the EBT is equal to 0.257173, for the TSAL it is equal to 0.544545 and for the ROA it is equal to 0.832250. A relatively large standard error, which is present in most of the observed indicators, is certainly due to the proportion of heteroscedasticity, but they do not affect the stability of coefficients in ECMs (Figure 3).

The comparison between the static-deterministic and stochastic-dynamic forecast simulations showed that the first simulation is more adequate. Econometric models that use time series are the most reliable for short-term forecasts, but this does not affect their validity (Pindyck & Rubinfeld, 1998; Janeš, 2014). By setting up the final ECMs for the financial KPIs (Table 1) and the static-deterministic/stochastic-dynamic forecast simulation (Equation 2), the layout of the BSC was empirically verified and confirmed.

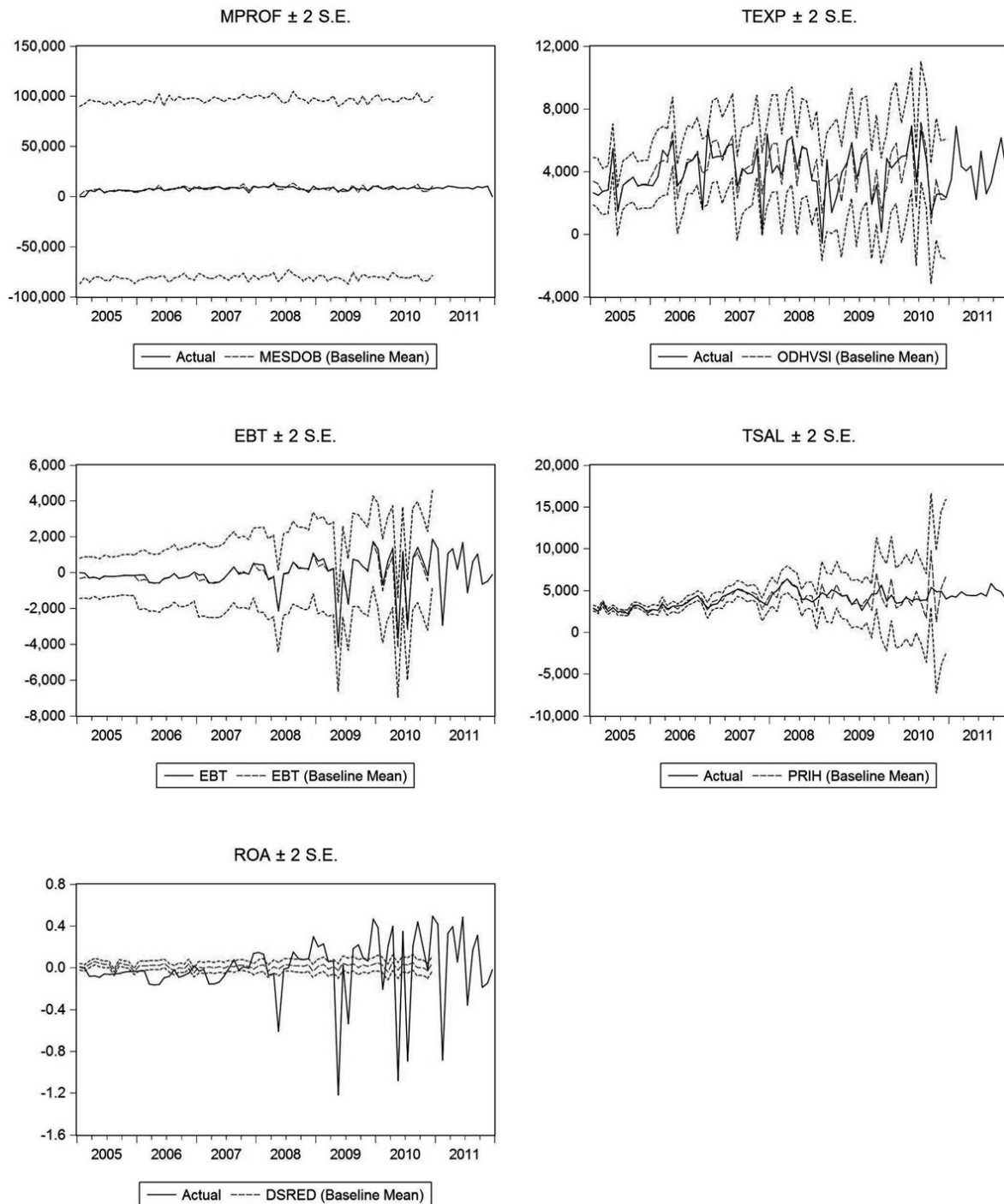
SOLUTIONS AND RECOMMENDATIONS

One of the major challenges for competitiveness and the opportunity to improve business processes was the transition to the new SAP ERP business

Figure 3. Stochastic-dynamic forecast simulation for many periods in advance

Source: Adapted from Janeš, 2012.

M



information system in 2012. For the SAP ERP project it is a necessity to have well defined processes and their respective KPIs. Because well-defined processes represent the background for descriptions of “the rules” with which SAP’s programmers can efficiently and effectively write programming code for modules of the whole information system.

In order to support these endeavors, the project of empirical verification of the PMS has been started in about the same period in 2011.

The company’s executive management was interested in the project because the establishment of a BSC begins with an overview of the strategy and is not based on actual outputs of business processes (Kaplan & Norton, 2004; Kaplan, 2012; Thakkar, Deshmukh, Gupta, & Shankar, 2007). Therefore, KPIs should be directly related to the strategy of the organization; it is mandatory to choose them on the basis of the strategic objectives of the organization (Hoque & Chia, 2012). The BSC and associated KPIs are treated in practice as a process that supports the reviewing and changes of the PMS in relation to changes in the business environment of the organization, and its success depends on the compliance with the qualitative and quantitative perspectives. It is necessary to clarify why, what and how often is needed to measure before the decision of how to measure. Managers were first asked about what they want to achieve, what their objectives are and how they can describe them. Historically, the processes in the YM Company were investigated mainly as qualitative. Previous research of business processes and PMSs was predominantly performed with the data within a short period of time. They are not taking into account that some activities have an immediate impact, while the impact of some others is manifested only after a certain time (Banker, Potter & Srinivasan, 2000; Faganel & Trnavčević, 2012; Kasperskaya & Tayles, 2013; Thakkar et al., 2007). It is important to emphasize that longitudinal and dynamic research for developing theories in this area is very rare (Janeš, 2013; Kaplan, 2012; Waruhiu, 2014).

Kaplan and Norton’s BSC provide a comprehensive framework that translates the strategy objectives of the organization into a coherent set of KPIs. As a rule, the KPIs are determined based on previous experience and through regular reviews. Where appropriate, an expanded range of KPIs may be included or some of them may also be phased out. With this approach, it is possible to find measurable perspectives of the current business processes, because the selected KPIs are likely to be important for determining the outcome of business processes in the future.

FUTURE RESEARCH DIRECTIONS

The concept of causality in the BCS is not extensively explained and seems unclear. In particular, in emphasizing that improvements in one area of the BSC lead to improvements in others, Kaplan (2009, 2012) and Kaplan and Norton (2004), have identified complex causal relations, subsequently connecting many KPIs across the four perspectives (Barnabè, 2011; Waruhiu, 2014, pp. 119-120).

The analysis of many qualitative and quantitative research studies about the implementation of KPIs, performed globally, showed the generally favorable influence of the KPI’s on the fulfillment of the sustainable strategy. Specifically, a recent trend in the performance evaluation of companies is the increasing emphasis on intangible measures and non-financial perspectives (Chareonsuk & Chansa-ngavej, 2010; Jerman, Kavčič, & Korošec, 2012; Kaplan, 2012). The completion of qualitative research with statistical tools holds great potential. In cases where companies have time series data related to their KPIs with which to validate the causal relations from the strategy map, proper statistical procedures could be applied (Bukh & Malmi, 2005; Janeš, 2013, 2014; Kaplan, 2012). However, it must be emphasized that the generalization of research findings had to be limited to only one manufacturing company. Based on the results it is recommended that further research is oriented on expanding the BSC on other entities within the ebm-papst group as well as at the top

strategic management level and to use the causal relations to forecast the future trajectory of the strategy (Kaplan, 2012). In addition, the use of nonlinear methods for the ECM modeling should also be considered. Further research in the form of additional case studies, i.e. companies from all sectors that operate independently on the markets, is definitely recommended, because the key issue about the KPIs causal relations is whether they really contribute to the benefits of the business processes exploitation and strategy fulfillment (Bukh & Malmi, 2005; Janeš, 2013; Janeš & Faganel, 2013).

CONCLUSION

One of the main areas that the relevant literature, BSC users, and Kaplan and Norton identified as being critical issues to be further developed is related to the assessment of causal relations which are essential within the BSC. The literature and practice has identified the necessity to further define the concept of causality within the layout of the BSC, in the direction of relying on specific quantitative tools needed to convert the BSC into a mathematical model (Janeš, 2012, 2014; Janeš & Faganel, 2013; Kaplan, 2012).

In this case study, the company's business performance sustainability understanding contributed to the selection of Engle-Granger two-step method for the assessment of the error correction mechanism between the KPIs. Therefore, based on the reviewed literature on ECM, the main characteristics of this approach have been depicted and, in particular, the potential of using ECM for exploring the concept of causality in BSC have been stressed. Consequently, a deeper understanding about the relations between financial and non-financial KPIs has been gained (Barnabè & Busco, 2012).

In addition, the data reveal an impact of global crisis on the company captured in a way that ap-

pears as not being substantial. A key finding in this regard is that the development of the BSC, supported by the Engle-Granger method, makes a substantial contribution to the causality explanation of the BSC system.

REFERENCES

- Angrosino, M. V., & Mays de Pérez, K. A. (2000). Rethinking observation. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 673–702). Thousand Oaks, CA: Sage.
- Banker, R. D., Potter, G., & Srinivasan, D. (2000). An empirical investigation of an incentive plan that includes nonfinancial performance measures. *Accounting Review*, 75(1), 65–92. doi:10.2308/accr.2000.75.1.65
- Barnabè, F. (2011). A system dynamics-based Balanced Scorecard to support strategic decision making Insights from a case study. *International Journal of Productivity and Performance Management*, 60(5), 446–473. doi:10.1108/17410401111140383
- Barnabè, F., & Busco, C. (2012). The causal relationships between performance drivers and outcomes. *Journal of Accounting & Organizational Change*, 8(4), 528–538. doi:10.1108/18325911211273518
- Bititci, U. S., Mendibil, K., Nudurupati, S., Garengo, P., & Turner, T. (2006). Dynamics of performance measurement and organisational culture. *International Journal of Operations & Production Management*, 26(12), 1325–1350. doi:10.1108/01443570610710579
- Bukh, P. N., & Malmi, T. (2005). *Re-examining the cause-and-effect principle of the balanced scorecard*. Retrieved August 12, 2012, from <http://www.costingforum.com/files/pdf.../BukhMalmi.pdf>

- Chareonsuk, C., & Chansa-ngavej, C. (2010). Intangible asset management framework: An empirical evidence. *Industrial Management & Data Systems*, 110(7), 1094–1112. doi:10.1108/02635571011069121
- Engle, R. F., & Granger, C. W. J. (1987). Co-integration and Error Correction: Representation, Estimation, and Testing. *Econometrica*, 55(2), 251–276. doi:10.2307/1913236
- Faganel, A., & Trnavčević, A. (2012). Sustainable natural and cultural heritage tourism in protected areas: case study. *Annales: Series historia et sociologia*, 22(2), 589–600.
- Gosselin, M. (2005). An empirical study of performance measurement in manufacturing firms. *International Journal of Productivity and Performance Management*, 54(5/6), 419–437. doi:10.1108/17410400510604566
- Granger, C. W. J. (2004). Time series analysis, cointegration, and applications. *The American Economic Review*, 94(3), 421–425. doi:10.1257/0002828041464669
- Gummesson, E. (2000). *Qualitative methods in management research*. Thousand Oaks, CA: Sage.
- Hassouneh, I., Serra, T., & Bojnec, Š. (2015). Nonlinearities in the Slovenian apple price transmission. *British Food Journal*, 117(1), 461–478. doi:10.1108/BFJ-03-2014-0109
- Hoque, Z., & Chia, M. (2012). Competitive forces and the levers of control framework in a manufacturing setting. *Qualitative Research in Accounting & Management*, 9(2), 123–145. doi:10.1108/11766091211240351
- Janeš, A. (2012). *Development of a model for empirical verification of the balanced scorecard* (Unpublished doctoral dissertation). University of Primorska, Koper, Slovenia.
- Janeš, A. (2013). Managing sustainability with the error correction model. *International Journal of Sustainable Economy*, 5(2), 125–139. doi:10.1504/IJSE.2013.052966
- Janeš, A. (2014). Empirical Verification of the Balanced Scorecard. *Industrial Management & Data Systems*, 114(2), 203–219. doi:10.1108/IMDS-04-2013-0195
- Janeš, A., & Dolinšek, S. (2010). Do we need a new compass for the journey through the global crisis? *Journal of Industrial Engineering and Management*, 3(2), 255–293. doi:10.3926/jiem.2010.v3n2.p255-293
- Janeš, A., & Faganel, A. (2013). Instruments and methods for the integration of company's strategic goals and key performance indicators. *Kybernetes*, 42(6), 928–942. doi:10.1108/K-08-2012-0022
- Jerman, M., Kavčič, S., & Korošec, B. (2012). The importance of the management accounting system in the decision making process: Empirical evidence from Slovenia. *African Journal of Business Management*, 6(24), 7252–7260.
- Kaplan, R. S. (2009). Conceptual foundations of the balanced scorecard. In C. Chapman, A. Hopwood, & M. Shields (Eds.), *Handbook of Management Accounting Research* (pp. 1–36). Oxford, UK: Elsevier. doi:10.1016/S1751-3243(07)03003-9
- Kaplan, R. S. (2012). The balanced scorecard: Comments on balanced scorecard commentaries. *Journal of Accounting & Organizational Change*, 8(4), 539–545. doi:10.1108/18325911211273527
- Kaplan, R. S., & Norton, D. P. (2004). *Strategy Maps: Converting intangible assets into tangible outcomes*. Boston: Harvard Business School.
- Kasperskaya, Y., & Tayles, M. (2013). The role of causal links in performance measurement models. *Managerial Auditing Journal*, 28(5), 426–443. doi:10.1108/02686901311327209

Kenny, G. (2012). From the stakeholder viewpoint: Designing measurable objectives. *The Journal of Business Strategy*, 33(6), 40–46. doi:10.1108/02756661211281507

Nørreklit, H. (2000). The balance on the balanced scorecard a critical analysis of some of its assumptions. *Management Accounting Research*, 11(1), 65–88. doi:10.1006/mare.1999.0121

Pindyck, R. S., & Rubinfeld, D. L. (1998). *Economic models and economic forecasts*. Singapore: McGraw-Hill.

Thakkar, J., Deshmukh, S. G., Gupta, A. D., & Shankar, R. (2007). Development of a balanced scorecard: An integrated approach of ISM and ANP. *International Journal of Productivity and Performance Management*, 56(1), 25–59. doi:10.1108/17410400710717073

Verbič, M., & Črnigoj, M. (2014). Corporate Investment and Corporate Taxation During the Economic Crisis in Slovenia. *Eastern European Economics*, 52(6), 32–56. doi:10.1080/00128775.2014.1003511

Waruhiu, H. (2014). Rebalancing the Balanced Scorecard: A Sequel to Kaplan and Norton. *European Journal of Business and Management*, 6(29), 116–124.

Ydria Motors, Y. M. (2015). *Introduction: Company Policy*. Retrieved April 15, 2015, from <http://www.ydria-motors.si/eng/introduction/179>

KEY TERMS AND DEFINITIONS

AC and EC Motors: Alternating Current motors and Electronically Commutated motors.

BSC: BSC is a managerial tool of four perspectives, which focuses attention of management on only a few most important KPIs.

Cointegration: Means that series have a common trend. If a pair of variables is cointegrated then it is likely that, they have a common factor and that one of the variables causes the other variable.

Error Correction Model: When the KPIs are cointegrated, there must be an ECM that describes the short run dynamics of the cointegrated variables towards their long run equilibrium values.

Error Correction Term: Measures the speed of adjustment to long run equilibrium between the KPIs.

KPIs: The most important performance indicators of PMS by which can be followed progress towards achievement of strategic goals.

SAP ERP: Systems, Applications, and Products in Data Processing for Enterprise Resource Planning.

Lack of Characteristics Management Causing Biggest Projects Failure

Loredana Arana

University of Phoenix, USA

INTRODUCTION

You might have the right amount of resources, the right amount of skill sets, the right amount of money and the perfect project plan. However, what you thought it was going to be an easy project will turn into a total disaster if you are not managing your resources characteristics or personalities. Once you are aware you have a problem, you now have to switch to damage control mode and waste valuable time, instead of concentrating on the core objectives of your project. This is byproduct of not properly assessing your resources correctly. Projects in themselves are complicated to manage; so dealing with difficult people can add more stress and create barriers that can hinder the progress. I don't think your resources plan on giving you a hard time, but like the old adage says, *the path to a project nightmare is littered with good intentions*. Project success is not only about balancing the triple constraints (cost, time and scope); it is also about controlling all the unknowns, including your resources. A team is only established when everyone pulls together and focuses on the project objectives. Your role as a project manager is key to insuring successful outcomes. A poor project manager fails to bring out the best in his team, but a good one can produce a winning combination, even with some individuals that may perhaps seem to lack ability to do so. Leadership is a critical and essential component for any project management team. To better understand the impact of project

manager, it is helpful to consider the following definitions of project manager:

- A project manager is usually a person who is responsible to run a project from the beginning to the end. It includes organization, implementation and control of their resources including people. (Techpedia, 2013).
- Project managers need to be leaders which is the process of influencing others to understand and agree about what needs to be done and how it can be done effectively, and the process of facilitating individual and collective efforts to accomplish shared objectives (Yukl, 2002).

You need to lead which involves the ability to influence people to take actions toward completing a goal or project. Projects contain a number of components with the main three being scope, cost, and time. For the project team to effectively meet scope, cost, and time goals, one must appreciate the impact of positive leadership. It is up to the project manager to manage issues related to scope, cost, and time, as well as to lead the team to successful completion of these goals and the project as a whole. A project manager manages their project staff by developing a positive environment.

We are all very familiar with the term "lead by example". Team members may already have negative characteristics, but if the leader displays

negative traits, other team members will follow to parallel behavior as the leader. However, these types of project manager that don't lead by example introduce a whole other issues and motivational scheme in which I am not addressing in this chapter. Any individual who forms part of a team will also make a decision to commit to the team's goals; this is why they need to be sure they understand the goals. Teams normally are complicated, it is helpful for the members to ask questions, clarify any misunderstanding and get to know their project manager.

BACKGROUND

Company survival depends on how well they adapt to changing technologies, diversity of their workforce and market situation. So it is no wonder that project management is the key to guide the company in the right direction. A great deal is being written about project management addressing things like budget, time and scope, but not a lot of emphasis on how to handle resource and their characteristics. The success of any project is not only attained by following a project methodology but by adhering to and effectively using the tools and resources available to you. Past experiences have shown that diverse resources can dramatically impact cost management as well as the overall timeline. It is important to recognize resource aspects that can cause disruptions in a project and how to mitigate and manage them more effectively to reach desired outcomes. Human resources can have a negative impact to projects. Project manager need to be prepared to put challenges on permanent lockdown. When looking back on previous unsuccessful projects, it was not the methodology that was deficient but in the way that the group came together as a team. It is also critical to analyze the different personality types and how they can influence the outcomes. Early analysis and observation of each personality involved can help to isolate potential pitfalls and help to address the issues immediately. It is often

better to deal with potential problems at the start that could delay timelines, and make the necessary adjustments for the betterments of the individual, team and the project.

M

PROJECT MANAGER COMMON MISTAKE STORIES

As it is, project managers are faced with great deal of pressure from tight timeline and budget constrains. So to put more pressure they now have to deal with the dilemma of working with team members that don't even report directly to them. How can you hold them accountable for not following through with their work? We have been told that as a project manager it is normal to go through conflict and friction with your team while it is undergoing a self-adjustment phase. Although this is a normal process, it's important as a project manager to analyze each personality and make sure they don't disrupt the project. We are responsible as project managers to smooth the progress of creating conformity within our project team in multiple ways. How will you make your team members accountable for completing task if they don't get along with you or any of the other team members? You might be thinking, "Do I need to polish up on my babysitting skills?" The answer is yes you do! In my vast experience as a project manager, I have encountered team members showing bad behavior, which make them hard to manage at times. So what are basic babysitting skills? It is managing behavior problems and understanding the psychology of how different personalities types function

A big issue is project managers that are indecisive, they keep changing their mind, and they like to move the team in new directions all the time. They are reacting to everything that is received and adjusting to new feedback, while changing directions at the drop of a hat. These are project managers which think they have everything under control by doing these unpredictable moves, can only damage team productivity and hamper moral.

This in turn will only drive the team to bad behavior as method of survival. Project managers should not take side or side with complainers, if they do; they automatically join the drama club. This club will be drawn you into endless and worthless arguments that will take you away from your real focus, your project. They want to demolish your project, hinder any progress made, or devour your funds, thus prolonging your timeline and converting your project scope into what is commonly referred to project scope creep.

ESCAPING PROJECT CATAclysms

Project Manager Utmost Common Oversights

Many reference materials and experienced project managers can tell you that there will be occurrences with your project team members where they will face disagreements through out all the phases of the project, during the decisions making sessions and even after post implementation. How about when you set goals and other forces change them? One of the forces is the people around you; they usually want to sway the goals you set. Most of the time your team member may feel that they think they know what is best for project, or perhaps they have their own agenda and want to take different paths. Obviously, it is critical to curate excellent relationships with your team members. Unclear statements or expectations are another common mistake. We assume that everyone understands them and they know what is expected of them, right? Wrong! I have seen project managers over communicate with their teams with vague goals and expectations that the team can easily misinterpret.

Project managers are part of the project team so they can also be the cause of the drama. Putting a self-centered project manager in charge can discourage his or her team members especially when they try to make any recommendations. It is true that the project manager is in charge and

can make the final call, but doesn't mean that no one else can have an opinion. This can be risky proposition which can lead to serious communication barriers while undermining the very nature of teamwork which normally should include healthy dose feedback from all team members. This personality only showcases a project manager egotistical characteristic, which will develop poor team morale and less desire to participate in the project.

Obviously, mistakes will occur as it is what make us grow and became great project managers. However, if we can avoid them and be on the lookout for bad team member traits, we can concentrate on the work at hand. It is up to you to spot these team member pitfalls and deal with them quickly and in diplomatic fashion. You need to address team members' mistakes fast to minimize disruption and prevent losing control, because there is no known cure for fixing people's personalities or traits, but you will need to know how to navigate through them.

How to Overcome Your Teams Destructive Characteristics

The first steps in project management are for project manager to develop general basic skills. These skills can be acquired through education and experience. They are five skills that are considered general project manager skills very important to develop: leading, communicating, negotiating, problem solving and influencing.

Next you need to understand that it is human nature to have a tendency to classify people we come across by putting an assortment of labels. Many books written by great philosophers and writers such as Machiavelli or Plato, to name a few, have provided us a lot of information regarding people's characteristics. They all have great insights, but the truth be told that no matter how we do it we need to assess all of our interactions with people to understand their characteristics so we can manage them and control the risks. At work, we have a job to do and a role to perform.

Figure 1. General project management skills



Behaviors are influenced by what sometime it is expected of us, not the way we would like to behave. Personalities have a tendency to exercise a strong influence on work behavior (Barrick & Mount, 1993). However, people can revert back to what they want to do. No matter how exigent, confrontational, or negative the difficult person's behaviors never take it personally. You only need to know that it is best to assess these bad behaviors at the start of the project so that you can control them before they get out of hand.

By recognizing some of the difficult personalities and reacting to each effective tactics, you can develop to your advantage instead of disadvantage. Although, it said that meetings are opportunities to bring the team together and to make sure everyone is on the same page, they can be a total loss if you can't handle your difficult team members in effective ways. So, how do you handle them? It is frequently difficult if you're assigned team members that don't directly report to you. But there is still hope, you can find methods to control them and keep your project on the right track. We can take a closer look at a few of what we called our "usual suspects" types which we've all come across in multiple projects. Although you may have to do a bit of political maneuvering to get what is needed for effective collaboration from the broader organization, it is well worth the trouble.

If a person or team is too sensitive, fewer decisions will be confidently made.

Resource Detrimental Characteristics

It is true that we cannot always use cookie cutter methods with all of our resources. However, below are some of the different personality traits that I have encountered. I also like to put people on simple categories to simplify the assessment that usually helps control my resources:

- Trespasser – Entering without Permission:** These types of team members want to be in everything.. They will critique everything you say, do or even think about. You will need to learn to involve them at the beginning and make sure you can avoid intrusions during the lifecycle of your project.
- Bully – Dominating Over Everyone:** These types of team members are never happy; the only thing in their mind is to destroy your project. Everything for them is an emergency and they require a lot of attention. They like to cause turmoil, while taking everyone out in their path.
- Time Bomb – Exploding Without Notice:** These types of team members can start ca-

Figure 2. Negative personalities



tastrophe without notice. They are the ones that can bring down your communication lines with the other team members. You need to learn how to diffuse them.

- **Watchtower – Watching Every Steps You Take:** These types of team members will judge every thing you do and will find something wrong in anything you do or say. You need to find a way to engage these types and let their characteristics work to your advantages.
- **Split Personality – Double the Drama:** These types don't know what they want. They are known for lack of ownership and will go where the wind blows. You will need to try to build confidence so you can cut down on vacillation.
- **Phantom – Hidden Trouble:** These types of team members are nowhere to be found. They hardly make your meetings or want to be involved in your project. Unfortunately, these are the people that you need to make important decisions or your project cannot move on. They seem to be always busy for you and will rarely attend. You will learn how to maximize the time they will give you.

These types of characteristics may seem hard to control, but how can you control them? Simple you will need to be sure that they are included early in the game. It is best to communicate often. I like to follow the Lewin's change model which is an easy and self-explanatory for managing change as well as people. Lewin teaches us the three different stages of change, which is a useful methodology for any project since most of them are constantly evolving. The first step is to create the inspiration to change which Lewin calls it as the unfreeze step. The next step progresses through the change process by advancing to an efficient communications and allow people to embrace new ways of functioning with change. When you accomplish a stage of stability this is called refreeze, this is essential in order to create the assurance from which to get on the next unavoidable change (Lewin, 1951).

Our team members' personalities are a display of their characteristics. Personalities have a lot variation, no one single personality is perfect. We need to accept the fact that personalities are complex. Sometimes a person with bad behavior may not know they have this behavior. Therefore, the person doesn't adapt well to changes. You will need to have a strategic plan that includes com-

munication, methods to stop rumors and control your team's toxic behavior. How to accomplish this? Sure you can try to use one the most commonly used theories, methodologies and other tools, which are wonderful. However, the simplest way to accomplish this is to just keeping it simple.

I have found in my experience as a project manager that following steps are often better than complicating things; the following are the PM TEAM steps that I found to be helpful:

- **Put it in Writing:** You will need to formalize your expectations in writing in order to serve as a constant reminder; people tend to forget whether it is on purpose or not.
- **Make a Deal:** Be specific with your team on the type of dedication you required by them. You need to tell them the objective, the results and what you are trying to achieve at the specific time and additional details such as cost and resources, do not be vague.
- **Talk to Negotiate:** Sometimes there is a need for you and your team member's to negotiate timeframe. You will need to be flexible, but not too much. For the most part, you will need to be able to meet people's requests.
- **Explicit:** All your requests will need to be very clear, it will make easier for the person to provide you with their estimate and commitments. You need to keep in mind that just because you put it in writing and made it explicit doesn't make it and command, you will still need to monitor and control. Your team needs you to provide them with specific goals; they don't need to know how you are going to measure the tasks, but also the results.
- **Accountability:** You will need to be a role model and live up to what you promise. If you are a transparent project manager you will avoid misunderstanding or gossip. Accountability also implies to possess the liability and power to do something and totally acknowledge the risks of the results of those actions.

Figure 3. Simple steps to control negative resources



- **Minimize:** You can minimize the pressure that your team will experience through out the project by motivating them and have a check and balance system. This is to check your team temperature and see how things are going.

The best advice is for a project manager is to stay calm, be straight and show the way. Most project managers will rely on their basic management skills, which are leading, negotiating, problem solving, influence and communication. These are good skills; however you will need to develop an aptitude assessment to recognize the characteristics of people that can ruin your project. It is necessary to spot the bad behavior and change it to your own advantage. The project managers' job is not to squash the bad personalities, but it is to use them for the success of the project. You also need to know how to foster a healthy environment for the team members that are different and to help them to come collectively and carry out.

Generally, it will be necessary for a project manager to change communication methods with each individual, as you know what they say; "everyone is different and have different needs". An example is if you have a touchy team member you will need to be direct and factual when talking to them. You need to also assess their skills not and not just the bad behavior so you can match them with the appropriate tasks.

FUTURE RESEARCH DIRECTIONS

Understanding Team Formation

Even thou you will have your project team for a short-term, it can take time to gel them into one. This typically follows familiar phases; the project team travels to become a unified team with a same goal. Bruce Tuckman, Psychologist, was the one that developed the concept of "forming, storming, norming, and performing" back in 1965. There is a path that a team must follow to reach formation

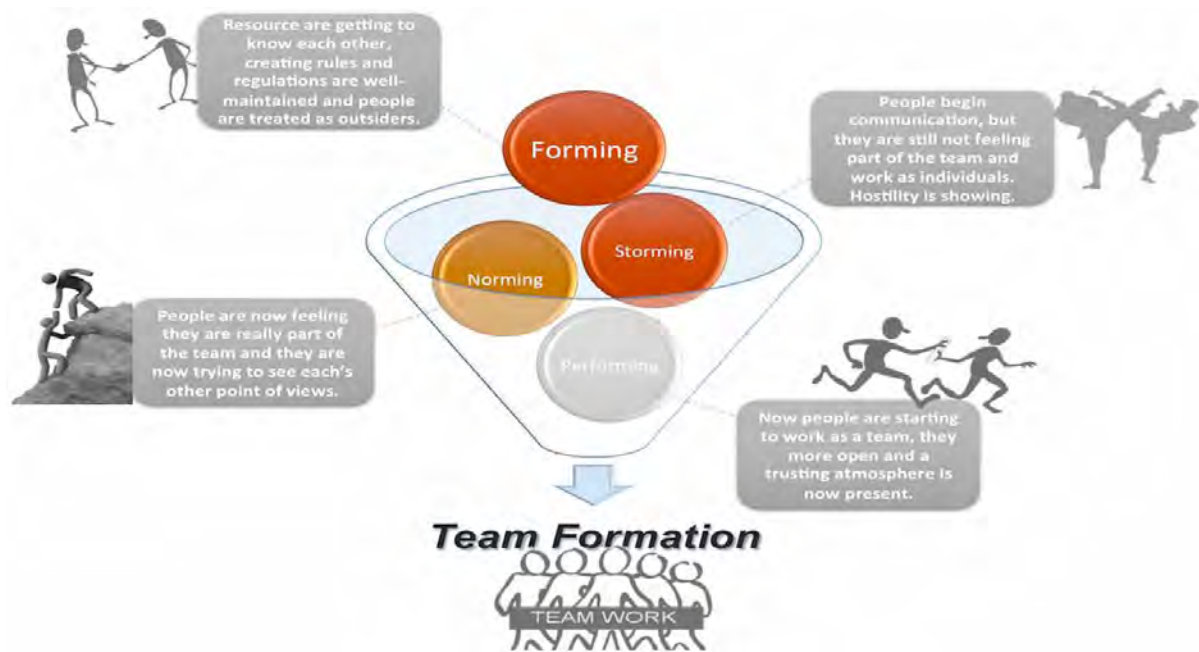
and become a true team. These are great concepts and will work for the most part, but it is important to study human personalities and characteristics as much it is as important to manage and adopt tactics to manage them.

Recommendation for futures research direction is to start examines the up-and-coming movement regarding team project manager dealing from different point of views such as networking, strategic interpersonal commitment methods. Team project managers find themselves on the take-off point providing last arrangements to jump ahead of organization "best practices". You should learn more about how project managers recognize, choose, act, and achieve their goals. (Lord & Brown 2004). While a great deal regarding team development has been written, people still need to learn a lot. Organizations try to employee teams and promote teamwork, yet there is very limited understanding on how to really develop a teams and control resources. The only way to know is through research, lesson learned and one own experience.

A team typically follows a different pathways or stages when seeking to become a productive team, this is described by psychiatrist M. Scott Peck in his book *The Different Drum: Community Making and Peace*. He mentions how a team must go through chaos and denial as part of the transformation. Therefore, project managers or leaders should not be in denial themselves by avoiding these chaos and ignoring the fact that some personalities will be the natural disaster of your project. The chaos that Peck is referring to is the differences of opinion and viewpoint that exist between the members in the group. Project manager or leader will be the target of outbreaks and this is when they need to understand overall dynamic, it is no time to give up in despair, swearing that the team of people will not be able to work together successfully.

The quality of your project team is frequently an important success factor. You will need to be sure you don't let them feel like they are your hostages. Because project team members who want

Figure 4. Team formation quick overview



to leave and are held against their will, they will make sure your project is a failure. It's healthier to allow them to leave, this will avoid for them to bring down the rest of your project team into their bad behavior pit. I would rather have quality than mediocrity as a team.

CONCLUSION

It is important to select the right team members. Some project managers will be lucky to have this coming to them naturally; others will need to work at it. It is about trying to familiarize yourself with your team members, their forte, and connecting their knowledge and experience with the project at hand. This is a difficult task and will require for you to have a smile on your face, even when you do not feel like it. You will need to connect and talk with your team on a regular basis. Besides getting to know your team's bad behavior, it is important to develop great relationships. You will be surprised how your team will want to join forces with to complete the project on time and within

budget if you curate and foster a perfect balance. Projects will need to be guided in a manner that creates agreeable outcomes while at the same time mitigating the risks and/or erecting barriers. We as project managers will need to paint a picture of encouragement and inspiration for our team's abilities for them to identify with that vision.

In conclusion, team can be complex because of the diversity of team members even if you are well organized. It is more complex in organizations that have a matrix environment. Project team members in a matrix environment have various allegiances and if the team is not well organized, they can be dangerous to the success of the project. You can make your team more interconnected and triumphant by identify the types of characteristics your team has as well as assessing their needs as individuals, so that your methods can be properly executed with the desired results. The key is to identify potential drawbacks that may impede good team dynamics to flourish, address team members conflicting personalities or characteristics early in the game. You can definitely develop a more productive and successful team by doing so.

REFERENCES

- Barrick, M. R., & Mount, M. K. (1993). Autonomy as a moderator of the relationships between the big five personality dimensions and job performance. *The Journal of Applied Psychology*, 78(1), 111–118. doi:10.1037/0021-9010.78.1.111
- Janssen, C. (2013). *Project Manager (PM)*. Retrieved November 14, 2013, from <http://www.techopedia.com/definition/677/project-manager-pm>
- Lewin, K. (1947). Frontiers in group dynamics. *Human Relations*, 1(1), 5–12. doi:10.1177/001872674700100103
- Lewin, K. (1951). *Field Theory in Social Science*. New York: Harper and Row.
- Lewis, J. P. (1999). *The Project Manager's Desk Reference: A Comprehensive Guide to Project Planning, Scheduling, Evaluation, and System*. New York: McGraw-Hill.
- Lord, R. G., & Brown, B. R. (2004). *Leadership Processes and Follower Self-Identity*. Hillsdale, NJ: Erlbaum.
- Project Management Institute. (2008). *A Guide to the Project Management Body of Knowledge (PMBOK)*. Newton Square, PA: PMI Publishing.
- Scott Peck, M. (1998). *The Different Drum: Community Making and Peace*. New York: Simon & Schuster.
- Yukl, G. (2002). *Leadership in organizations* (5th ed.). Upper Saddle River, NJ: Prentice Hall.
- Engler, B. (2008). *Personality Theories: An Introduction*. Cengage Learning; Retrieved November 12, 2013.
- Federal Glass Ceiling Commission. (1995) *A Solid Investment: Making Full Use of the Nation's Human Capital*. Washington, DC: U.S. Department of Labor, 6.
- Fehr, B., Samsom, D., & Paulhus, D. L. (1992). The construct of Machiavellianism: Twenty years later. In C. D. Spielberger & J. N. Butcher (Eds.), *Advances in personality assessment* (Vol. 9, pp. 77–116). Hillsdale, NJ: Erlbaum.
- Gobeli, D. H., & Larson, E. W. (1990). Project Management Problems. *Engineering Management Journal*, 2(2), 31–36. doi:10.1080/10429247.1990.11414571
- Gustafson, S. B., & Ritzer, D. R. (1995). The dark side of normal: A psychopathy-linked pattern called aberrant self-promotion. *European Journal of Personality*, 9(3), 147–183. doi:10.1002/per.2410090302
- Hamburger, D. H. (1990). The Project Manager: Risk Taker and Contingency Planner. *Project Management Journal*, 21(4), 11–16.
- Hjelle, L. A., & Ziegler, D. J. (1992). *Personality theories: basic assumptions, research, and applications*. McGraw-Hill; Retrieved November 12, 2013.
- Hofstede, G. (2001). *Culture's Consequences: Comparing Values, Behaviors, Institutions, and Organizations Across Nations*. Thousand Oaks, CA: Sage.
- John, O. P., Robins, R. W. & Pervin, L. A. (2010). *Handbook of Personality, Third Edition: Theory and Research*. Guilford Press. Retrieved November 12, 2013.
- Keyton, J. (2002). *Communicating in groups: Building relationships for effective decision making* (2nd ed.). Boston: McGraw-Hill.

ADDITIONAL READING

- Christie, R., & Geis, F. L. (1970). *Studies in Machiavellianism*. New York: Academic Press.
- Drucker, P. F. (1977). *People and performance*. New York: Harper's College Press.

Kharbanda, O. P., & Pinto, J. K. (1996). *What Made Gertie Gallop: Learning from Project Failures*. New York: Von Nostrand Reinhold.

Kowalski, R. M. (2001). *Behaving badly: Aversive behaviors in interpersonal relationships*. Washington, DC: American Psychological Association. doi:10.1037/10365-000

Lewis, J. P. (2000). *Project Planning, Scheduling and Controlling* (3rd ed.). Burr Ridge, IL: McGraw-Hill.

McHoskey, J. W., Worzel, W., & Szyarto, C. (1998). *Machiavellianism and psychopathy*. *Journal of Personality and Social Psychology*, 74, 192–210.

Mischel, W. (1999). *An Introduction to Personality*. John Wiley & Sons Incorporated; Retrieved November 13, 2013.

Pinto, J. K., & Slevin, D. P. (1988). Critical Success Factors Across the Project Life Cycle. *Project Management Journal*, 19(3), 72.

Porter, M. (1985). *Competitive advantage: Creating and sustaining superior performance*. New York: The Free Press.

Pretty, J. N., Guijut, I., Thompson, J., & Scoones, I. (1995). *A Trainer's Guide for Participatory Learning and Action*. London, UK: International Institute for Environment and Development.

Ruben, B. D. (1988). *Communication and Human Behavior* (2nd ed.). New York, NY: Macmillan Publishing Co.

Ryckman, R. M. (2007). *Theories of Personality*. Cengage Learning; Retrieved November 13, 2013.

Takaki, R. (1993). *A Different Mirror. A History of Multicultural America*. Boston: Little, Brown.

KEY TERMS AND DEFINITIONS

Characteristics: A personal quality, attribute, or trait.

Diversity: Considerable exceptionality of each individual, our difference from each other's.

Forming: Introduction phase of team members, everyone is getting to know each other.

Management: The process of controlling things or people.

Methodology: Set of practices perfection by someone else to guide you to do certain things.

Norming: Productivity phase, now team is concentrating on working toward project goals.

Performing: Team matures and is settling their expectations.

Storming: Phase where team is learning how to handle conflict.

Theory: It is a set of guidelines developed to explain an observable fact.

Triple Constraint: These are important elements (budget, scope and time) that need to be controlled efficiently for a triumphant completion of a project.

Making Sense of IS Project Stories

Darren Dalcher

University of Hertfordshire, UK

INTRODUCTION

Stories of failure make a compelling read, however, researchers with a keen interest in information systems failures are faced with a double challenge: Not only is it difficult to obtain intimate details about the circumstances surrounding such failures, but there is also a dearth of information about the type of methods and approaches that can be utilised to collect, describe and disseminate such information. The purpose of this chapter is to highlight some of the available approaches and to clarify and enhance the methodological underpinning that is available to researchers interested in investigating and documenting phenomena in context-rich and dynamic environments. The chapter begins by framing IS project failures in context, before highlighting the role of forensic failure investigation and the typical tools employed in gathering information. It encourages a move from case studies to case histories to capture the essence, dynamics and complexities of failure stories. It concludes by introducing a new range of antenarrative approaches that represent future developments in the study of IS failures, enabling a richer interpretation of linked factors that underpin IS failures.

BACKGROUND

The popular computing literature is awash with stories of IS project failures and their adverse impacts on individuals, organisations, and societal infrastructure. Indeed, contemporary software development practice is still characterised by runaway projects, late delivery, exceeded budgets,

reduced functionality and questionable quality that often translate into cancellations, reduced scope and significant re-work cycles (Dalcher, 1994). The net result is an accumulation of waste typically measured in financial terms. For example, in 1995 failed US projects cost \$81 billion, with an additional \$59 billion of overspend, totalling \$140 billion (Standish 2000). Jones contended that the average US cancelled project was a year late, having consumed 200% of its expected budget at the point of cancellation (Jones 1994). MacManus and Wood-Harper (2007) reported that the cost of software project failure across the European Union in 2004 was €142 billion. More recently, a McKinsey-Oxford survey of more than 5,400 software projects revealed that half of all projects significantly fail on budgetary assessment, while 17 per cent of projects actually threaten the very existence of the company, with the average project running 45 per cent over budget and seven per cent behind schedule, while delivering 56 per cent less functionality than predicted (Bloch et al., 2013).

IS failure investigations start with extensive attempts to collate relevant evidence. However, in most cases the researcher is exposed to specific information post-hoc, i.e. once the failure is well established and well publicised and the participants have had a chance to rationalise their version of the story. Most of the available sources are therefore already in place and will have been set up by agencies other than the researcher.

The purpose of a forensic investigation is to explain a given failure by using available information and evidence. The term Forensic is derived from the Latin 'Forensis', which is to do with making public. Forensic Science is the applied use of a body of knowledge or practice in deter-

mining the cause of death. Nowadays extended to include any skilled investigation into how a crime was perpetrated. Forensic systems engineering is the post-mortem analysis and study of project, product, artefact or service shortfalls and failures, which aims to uncover systemic and underpinning causes (Dalcher, 1994). The work involves a detailed investigation of a product or service, the underpinning project, its environment, decisions taken, politics, human errors and the relationship between subsystems. The work draws upon a multidisciplinary body of knowledge and assesses the project from several directions and viewpoints. The aim of forensic analysis is to improve the understanding of failures, their background and the dynamics that lead to them. The concept of systems is a central tool for understanding the delicate relationships and their implications in the overall project environment.

Forensic systems engineering is primarily concerned with documentary analysis and (post-event) interviews in an effort to ascertain responsibility lines, causal links and background information. The primary mode of dissemination of findings, conclusions and lessons is through the publication of case study reports focusing on specific failures. However, there are limited research methods to explore the dynamic and fragmented nature of complex failure situations. Lyytinen and Hirschheim (1987) noted that more qualitative research methods were needed for IS failure research as well as more extensive case studies that explored problems in more detail and viewed solution arrangements in light of what transpired. The same methods also need to account for group issues and cultural implications. Sadly, thirty years on, the same constraints in terms of methods are still in evidence.

DESCRIBING FAILURE

Making sense of IS failures retrospectively is difficult. In general, there is very little objective quantitative failure information that can be relied

upon. Instead, interpretation requires understanding of and engagement with the wider context. Indeed, a specific feature of failure is the unique interaction between the system, the participants, their perspectives, complexity and technology (Perrow, 1984). Lyytinen and Hirschheim (1987) pointed out that failure is a multifaceted phenomenon of immense complexity with multiple causes and perspectives. Research into failures often ignores the complex and important role of social arrangement embedded in the actual context. This is often due to the quantitative nature of such research.

Understanding the interactions that lead to failures likewise requires a humanistic stance that is outside the conventional positivist norm to capture the real diversity, contention and complexity embedded in real life. Forensic analysis thus relies on utilising qualitative approaches to obtain a richer understanding of failure phenomena in terms of action and interaction.

The fact that a failure phenomenon is being investigated, suggests that attention has already been drawn to the complexities, breakdowns and messy interactions that such a situation entails (i.e. the investigation is problem-driven). Many such inquiries deal with subjective accounts including impressions, perceptions and memories. The aim of the researcher is to increase in a systemic way the understanding of a situation, yet do so from a position that takes in the complexity of the entire situation and incorporates the different perspectives and perceptions of the stakeholders involved.

Overall, the purpose of a failure research method is to enable the researcher to make sense of the complexity of detail and the complexity of interaction and chart the contributory role of different causes and issues in the build up to failure. However, the armoury of research methods in this domain is often limited to case studies.

The term “case study” is an umbrella term used in different contexts to mean different things that include a wide range of evidence capture and analysis procedures. Yin (2013, p.14) defines the scope of a case study as follows:

A case study is an empirical inquiry that:

- *Investigates a contemporary phenomenon in depth and within its real-life context, especially when*
- *The boundaries between phenomenon and context may not be clearly evident.*

A case study can be viewed as a way of establishing valid and reliable evidence for the research process as well as presenting findings which result from research (Remenyi, 1998). According to Schramm (1971) the case study tries to illuminate a decision or a set of decisions and in particular emphasise why they were taken, how they were implemented and with what results. A case study is likely to contain a detailed and in-depth analysis of a phenomenon of interest in context; in our case, the failure scenario. Table 1 summarises some of the main advantages of using case studies.

The general aim of the case study approach is to understand phenomena in terms of issues in the original problem context by providing the mechanism for conducting an in-depth exploration. They often result from the decision to focus an enquiry around an instance or an incident, as

they are principally concerned with the interaction of factors and events. The combination of a variety of sources offers a richer perspective which also benefits from the availability of a variety and multiplicity of methods that can be used to obtain new insights about this single instance. A Case study allows the researcher to concentrate on specific instances in their natural setting and thereby attempt to identify the interacting perceptions, issues and processes at work, ultimately resulting in in-depth understanding. Crucially, the focus on a single incident thus enables the study of the particularity and complexity of a case, thereby coming to understand the activity within important circumstances (Stake, 1999) whilst including multiple sources of evidence and differing perspectives.

There are a number of general objections that are associated with the use of case studies (see Table 2). However, one must recognise that case studies are more likely to be used retrospectively rather than as an on-going perspective (especially from a failure point-of-view), as researchers are unlikely to know the potential for useful results and interest from the outset and may have difficulty in negotiating access to the location. Indeed, Yin (2013) concedes that using case studies for research purposes remains a most challenging endeavour.

Table 1. Main advantages of using case studies

• Ability to identify and focus on issues • Richness of detail • Multiple perspectives • Multiple sources and types of data • Multiple explanations (no absolute truth) • Cross disciplinary remit • Ability to recognise and minimise inherent complexity • Ability to handle conflict, disparity and disagreement • Ability to show interactions • Ability to observe emerging patterns • Opportunity to focus on the particular • Ability to gain real insight and understanding of a situation • Conducted in real-life (natural) setting • Encompasses original problem context • Ability to deal with interpretations • Features intensive analysis • Can extend the boundaries to include aspects of wider system environment • Can be accumulated to form an archive of cases • Can be strengthened and expanded with longitudinal features • Often retold in story format, which is more accessible to practitioners
--

Table 2. Main objections to the use of case studies

• Sometimes viewed as soft data (but some argue it is hard research) • Biases inherent in accepting views and perceptions • Questions about generalisability of findings (especially from a single case), but it is possible to build a library of such cases • Issues regarding objectivity of approach and perceived lack of rigour • Negotiating access to settings • Boundaries are difficult to define; but this could also be a strength! • Mainly retrospective • Sometimes viewed as likely to take too long and result in massive documentation • The observer effect • Reliability of conclusions • There is little control over events, but this may also be a strength
--

Comprehensiveness of coverage is not necessarily a requirement. The richness of detail can be controlled through the careful placement of systems boundaries and consideration of the wider system environment that is relevant to the specific phenomenon under study. Case studies can be utilised as a source of understanding, which is tolerant of ambiguity, paradox and contradiction. A case study is viewed as interpretative when events in the real world are observed and then an effort takes place to make sense of what was observed, i.e. when one tries to make sense of a failure from the perspectives of participants. They also offer the potential for generating alternative explanations from the different stakeholder perspectives, thereby allowing the researcher to highlight contradictions, conflicts and misunderstandings.

FROM CASE STUDIES TO CASE HISTORIES

The generally liberal use of the term *case study* requires a tighter definition of its meaning in failure research. While there may be a tradition of using case studies within the IS community, this is perhaps more often borrowed from the MBA culture than as a result of self conscious effort to adopt them as a research approach (Walsham, 1995; Cornford, 1996). Indeed, the case study is typically used more in its capacity as a teaching tool than as a research tool. The shift to studying the impact of issues within the organisational context renders case studies particularly useful for investigating failure scenarios. However, the use of the term often leads to some confusion.

Moreover, one of the major complications in failure investigations is in relating causes to effects through extended time horizons (Dalcher, 2000). The implications of actions may not be witnessed for years, or even generations. Delays between making a decision and observing the result distort the causal link between the two. As a result, people tend to associate a different level of severity to events occurring following a delay. The perceived severity is thus diminished with

the length of the delay further complicating the task of identifying patterns and interactions that contributed to a given failure. Failure researchers are thus required to provide adequate historical accounts of the interaction between actions, perceptions and the passage of time.

Case studies have typically been used to explore issues in the present and the past and comprise of ethnographic studies, single case studies and comparative case studies, as well as, action research, evaluative, exploratory, explanatory and descriptive case studies. In our experience there is a need to add the failure case study as a special example of a case study focusing primarily on the background, context, perception, interactions and patterns, especially as the failure investigation is likely to take place after the (failure) event. We thus propose the use of the label case histories to refer to the retrospective and specialised historical research studies focusing on failure incidents.

The time dimension (sequencing) is critical to understanding interactions and identifying their impacts when stories are constructed. Case histories are concerned with providing the background and context that are required to endow words and events with additional meaning. Background refers to previous history of the system itself, while context refers to interactions with the environment. As failures are time- and place-dependent, the case history framework enables readers to obtain an understanding of the intimate context surrounding the main event. The primary tool available to the community is the Case Histories of failures (derived from the use of the case study method). These represent a detailed historical description and analysis of actual processes from a relevant perspective. Their value is in tracing decisions (and recorded rationale) to their eventual outcomes by utilising techniques borrowed from decision analysis and systems engineering. Indeed, the historical description and presentation of a chronology of events infused with meaning, intention and understanding are based on the recognition that real life is ambiguous, conflicting and complex.

Case histories thus contain observations, feelings and descriptions. They can be used to

construct, share, dispute and confirm meanings, interpretations and scenarios in the context of real events (See for example, Dalcher, 2004; 2007). Rather than simply highlight a chronicled sequence of happenings, they convey a story encompassing a specific perspective, focus, and possibly some inevitable biases. The interpretation plays a key part in transmutating the chronicle into a meaningful story with plot, coherence and purpose. However, constructing a convincing narrative of a complex story with competing meanings, alternative perspectives and inherent prejudices is a challenge in itself.

FUTURE TRENDS: EMERGING NARRATIVES

Failures, in common with other activities that take place in organisations, are based on stories. The verbal medium is crucial to understanding behaviour within organisations and systems, and researchers are thus required to collect stories, grounded in practice, about what takes place (Gabriel, 2000; Simmons 2007). Gargiulo (2005) further asserts that effective organisational communication and learning is dependent upon stories: Listening to them is critical to the success of the organisation. Understanding failures often entails the retrospective untangling of complicated webs of actions and events and emergent interaction patterns. Failure storytelling can thus be understood as a combination of narrative recounting of empirical events with the purposeful unlocking of meaningful patterns, or a plot.

Historically, story telling has been an acceptable form of conveying and sharing ideas, norms, values, experience and knowledge of context. It plays a key role in communicating the cultural, moral or historical context to the listener. Indeed, Arendt, (1958) argued that the chief characteristic of human life is that it is always full of events, which ultimately can be told as a story. There are even strong claims that the narrative is the main mode of human knowledge (Bruner, 1990), as well

as the main mode of communication, learning and thinking (Fisher, 1987; Gargiulo, 2005; Denning, 2011). Moreover, children are often initiated into culture (and its boundaries) through the medium of story telling, offering models for emulation or avoidance.

In practice, the essence of any good case study revolves around the ability to generate an effective storyline, normally with a unique style, plot or perspective. In a large case, a general theme can be obtained from selected excerpts weaved together to illustrate a particular story. This is particularly useful when the researcher is trying to portray a personal account of a participant, a stakeholder or an observer in an incident, accident or failure. The implication is that the need to address personal aspects of interaction and story is fulfilled by the development of a research-valid narrative. Indeed, Remenyi et al. (1998) contend that a story, or a narrative description, is valid if the resulting narrative adds some knowledge. Furthermore, White (1973) describes a story as 'the process of selection and arrangement of data from the unprocessed historical record in the interest of rendering the record more comprehensible to an audience of a particular kind' by inserting a sense of perspective and purpose.

Storytelling can endow listeners with different meanings as stories can be understood in multiple ways. Narratives are neither discovered, nor found: they are constructed. Understanding IS failures is therefore more complicated than the discovery of a simplistic chronology of events as stories are crystallised through infusion with meaning and context. Narrative inquiry is evolving into an acceptable research approach in its own right in the social sciences and in management research circles (Gabriel, 2000; Boje, 2001; Czarniawska, 2004; Boje, 2011; Boje, 2014) as the story format provides a powerful way of knowing and linking disparate accounts and perspectives. When different accounts are combined, the emerging story line benefits from the richness of multifaceted insights.

Developing a narrative requires plot as well as coherence as a story is made out of events and

the plot mediates between the events and the story (Boje, 2001; Carr, 2001; Kearney, 2002). The narrative can thus become a powerful mechanism for eliciting and sharing experience in a meaningful way through intimate reflection. In failure stories, the plot often emanates from the actions and perceptions of participants emerging out of the flux of events, in (direct) contradiction with expectations. The storyteller is concerned with the perspective and purpose of participants as well as with the plausibility of the emerging plot. The combination of plot, purpose and perspective dictates the selection of elements, the filling in of links and the removal of 'irrelevant' noise.

Post-modern interpretation contends that most real life stories are fragmented, non-linear, discontinuous, multivariate and incoherent. This has already been highlighted as a feature of failure stories. Such stories also tend to be dynamic, polyphonic (multi-voiced) and collectively produced as they occur in asymmetrical, random and turbulent environments full of tensions and ambiguities. The stories are not plotted as such and they appear to flow, emerge and network offering complex clustering of events, emergent phenomena, causes, interventions, interferences and effects. Moreover, the accounts are often subjective, counter-intuitive and contradictory. This leads to interacting, and conflicting webs of narratives, characterised by coincidences, predicaments and crises.

Generally, stories appear to be improperly told, as a story is an 'ante' state of affairs existing previously to a carefully constructed narrative (Boje, 2001). The antenarrative, or the 'real' story, is the fragmented, messy and dynamic, multi-vocal, multi-plotted, multi-version and complex tale. Indeed, modern story-tellers look for new ways and mediums for weaving and depicting a multi-vocal reality, as exemplified by Mike Finggis's digitally shot film *Time's Arrow*, where the screen is split in four to allow for four separate perspectives and sub-stories that occasionally intersect or overlap. In the tradition of post-modern inquiry, a real life researcher is often faced with fragments rather than a whole story to tell; and many of the fragments

may reflect contrary versions of reality. This is potentially more acute when the accounts attempt to justify roles of participants in the lead-up to disaster or failure. It would also appear from past analysis that there are hierarchies of stories and stories that exist within, or interact with other stories. Using the terminology provided by Boje, the purpose of narrative methods is to take a complex situation characterised by collective (yet often conflicting) memory and an antenarrative and construct the plot and coherence that can be used to narrate and guide the story of interest.

The reality in failure stories is of multi-stranded stories of experiences and reactions that lack collective consensus. Indeed the discipline of decision-making has also recognised that making choices is about forming and selecting interpretations from a mosaic of possibilities (March, 1994; Weick, 1995). Not surprisingly, disasters, or traumatic stories, are hard to narrate, understand and justify. Stories have three basic properties: time, place and mind (Boje, 2001) which interact and build up as the story evolves. In forensic case histories, these are further clarified through the identification of the background and context, which clarify and justify the interpretation in the context of the emerging phenomena.

Boje (2001; 2014) and Kearney (2002) contend that the current view is of sequential single voice stories and implies excessive reliance on the hypothetical-deductive approach (akin to simplistic causal pairings). Reality emerges as fragmented retrospectives. Imposing a meaning is insufficient and inadequate, as actors need to find their own voice and make collective sense of a situation. The answer is not to develop Harvard type case studies but to rewrite stories as polyvocal tapestries enabling different perceptions, voices and interpretations to exist, thereby explaining webs of actions and interactions. What is new in this approach is the antenarrative reading which enables narrative analysis methods to be supplemented by antenarrative methods, allowing previously fragmented and personal storytelling to be interpreted as a unified whole. This focus offers

alternative discourse analysis strategies that can be applied where qualitative story analyses can help to assess subjective, yet ‘insightful’ knowledge in order to construct negotiated, ‘true’ and significant understanding of complex interactions (see for example, Drevin, 2011).

As for the longer-term future, good stories can also benefit from pictures, sound and added depth. Once we have mastered the techniques of telling complex, modern stories, we need to focus on composing that information. Even the most gripping story needs to be made attractive and believable. Textual information needs additional support not only in ‘emplotting’ and in maintaining coherence and perspective, but also in ascertaining the plausibility of constructed stories and in differentiating between noise and narrative. Developing improved techniques for organising or visualising knowledge (such as Net maps) can therefore help in untangling some of the fragmented strands and in making the stories more realistic and understandable, as well as ultimately more appealing, convincing and memorable.

CONCLUSION

IS failures are a familiar aspect of the reality and practice of software development and operation. Yet, attempts to make sense of their causes are limited by the paucity of methods capable of providing the needed insights. Improving the ability to analyse the systemic causes and their dynamic nature relies on the development of case histories that support historical research and retrospective sensemaking into the context of failure.

Stories provide a powerful research tool that can be used to reflect, share and make sense. With the benefit of hindsight it is possible to reconstruct a systematic re-telling of events that have led to a failure. The combination of case histories with narrative descriptions therefore offers richer insights. The narrated structure provides an explanation as to how and why failures occur. The purpose of the structure is to make sense of the rich tapestry of interactions and connections by

following an identified storyline that chronicles and links relevant issues within the environment. Engaging with the worlds of others through the medium of a story enables a deeper and richer reflection. Indeed, recounted life may prise open perspectives that would have been inaccessible using ordinary methods and thinking arrangements. Moreover, failure tends to highlight missing and incorrect assumptions and faulty defensive mechanisms and can therefore serve as a pretext to updating the frame of reference or the context for understanding as the listeners learn to construct a meaning and make sense of a highly multifaceted, multilayered and complex situation.

REFERENCES

- Arendt, H. (1958). *The Human Condition*. Chicago: University of Chicago Press.
- Bloch, M., Blumberg, S., & Laartz, J. (2013). Delivering Large-Scale IT Projects on Time, on Budget and on Value. *McKinsey on Finance*, 45, 28–35.
- Boje, D. M. (2001). *Narrative Methods for Organisational & Communication Research*. London: Sage.
- Boje, D. M. (2011). *Storytelling and the Future of Organizations: An Antenarrative Handbook*. New York: Taylor and Francis.
- Boje, D. M. (2014). *Storytelling Organizational Practices*. London: Routledge.
- Bruner, J. (1990). *Acts of Meaning*. Cambridge, MA: Harvard University Press.
- Carr, D. (2001). Narrative and the Real World: An argument for Continuity. In G. Roberts (Ed.), *The History and Narrative Reader* (pp. 143–156). London: Routledge.
- Cornford, T., & Smithson, S. (1996). *Project Research in Information Systems: A Student's Guide*. Basingstoke, UK: Macmillan. doi:10.1007/978-1-349-13863-0

- Czarniawska, B. (2004). *Narratives in Social Science Research*. London: Sage. doi:10.4135/9781849209502
- Dalcher, D. (1994). Falling down is part of Growing up; the Study of Failure and the Software Engineering Community, *Proceedings of 7th SEI Education in Software Engineering Conference*. New York: Springer-Verlag. doi:10.1007/BFb0017636
- Dalcher, D. (2000). Feedback, Planning and Control – A Dynamic Relationship. *FEAST 2000*, 34-38.
- Dalcher, D. (2004). Still Waiting? Computerisation of Ambulance Despatch Systems. *Annals of Cases on Information Technology*, 6(4), 440–456.
- Dalcher, D. (2007). Why the Pilot Cannot be Blamed: A Cautionary Note About Excessive Reliance on Technology. *International Journal on Risk Assessment and Management*, 7(3), 350–366. doi:10.1504/IJRAM.2007.011988
- Denning, S. (2011). *The Leader's Guide to Storytelling: Mastering the Art and Discipline of Business Narrative*. San Francisco: Jossey-Bass.
- Drevin, L., & Dalcher, D. (2011). Antenarrative and Narrative: The Experience of Actors Involved in the Development and Use of Information Systems. In *Storytelling and the Future of Organizations: An Antenarrative Handbook* (pp. 148–162). New York: Taylor and Francis.
- Fisher, W. R. (1987). *Human Communication as Narration: Towards a Philosophy of Reason, Value and Action*. Columbia, SC: University of South Carolina Press.
- Gabriel, Y. (2000). *Storytelling in Organizations: Facts, Fictions and Fantasies*. Oxford, UK: Oxford University Press. doi:10.1093/acprof:oso/9780198290957.001.0001
- Gargiulo, T. L. (2005). *The Strategic Use of Stories in Organizational Communication and Learning*. New York: M. E. Sharpe.
- Jones, C. (1994). *Assessment and Control of Software Risks*. Englewood Cliffs, NJ: Prentice-Hall.
- Kearney, R. (2002). *On Stories*. London: Routledge.
- Lyytinen, K., & Hirschheim, R. (1987). Information Systems Failures: A Survey and Classification of the Empirical Literature. *Oxford Surveys in Information Technology*, 4, 257–309.
- March, J. G. (1994). *A Primer on Decision Making*. New York: Free Press.
- McManus, J., & Wood-Harper, T. (2007). Understanding the sources of information systems project failure. *Management Services*, 51(3), 38–43.
- Perrow, C. (1984). *Normal Accidents, Living with High-Risk Technologies*. New York: Basic Books.
- Remenyi, J. (1998). *Doing Research in Business and Management: An Introduction to Process and Method*. London: Sage. doi:10.4135/9781446280416
- Schramm, W. (1971). *Notes on Case Studies of Instructional Media Projects*. Working paper for the Academy for Educational Development, Washington, DC.
- Simmons, A. (2007). *Who Ever Tells the Best Story Wins*. New York: AMACOM.
- Stake, R. E. (1995). *The Art of Case Study Research*. Thousand Oaks, CA: Sage.
- Standish Group. (2000). *Chaos 2000*. Standish: Author.
- Walsham, G. (1993). *Interpreting Information Systems in Organizations*. Chichester, UK: Wiley.
- Weick, K. E. (1995). *Sensemaking in Organizations*. Thousand Oaks, CA: Sage Publications.
- White, H. (1973). *Metahistory*. Baltimore, MD: The John Hopkins University Press.
- Yin, R. K. (2013). *Case Study Research: Design and Methods*. Newbury Park, CA: Sage.

ADDITIONAL READING

- Andrews, M. (2013). *Doing Narrative Research*. London: Sage. doi:10.4135/9781526402271
- Boje, D. M. (2011). *Storytelling and the Future of Organizations: An Antenarrative Handbook*. New York: Taylor and Francis.
- Boje, D. M. (2014). *Storytelling Organizational Practices*. London: Routledge.
- Bold, C. (2011). *Using Narrative in Research*. London: Sage.
- Clandinin, J. (2007). *Handbook of Narrative Inquiry*. London: Sage. doi:10.4135/9781452226552
- Clandinin, J. (2013). *Engaging in Narrative Inquiry*. London: Routledge.
- Cobley, P. (2013). *Narrative*. London: Routledge.
- Frank, A. (2010). *Letting Stories Breathe: A Socio-narratology*. Chicago: University of Chicago Press. doi:10.7208/chicago/9780226260143.001.0001
- Holstein, J. (2012). *Varieties of Narrative Analysis*. London: Sage. doi:10.4135/9781506335117
- Kim, J.-H. (2015). *Understanding Narrative Inquiry*. London: Sage.
- Matthews, R., & Wacker, W. (2008). *What's Your Story?* Upper Saddle River: Pearson.
- Riessman, C. K. (2008). *Narrative Methods for the Human Sciences*. London: Sage.
- Squire, C. (2014). *What is Narrative Research?* London: Bloomsbury.

KEY TERMS AND DEFINITIONS

Antenarrative: The fragmented and messy and dynamic stories of real life in their original context before a clear narrative is developed to explain away a certain aspect.

Case History: Specialised retrospective historical research focusing on failure incidents. Case histories emphasise the background and context that can help in untangling relationships and causes thus making sense of the events leading to the failure.

Case Study: Investigation of phenomena in naturalistic setting, conducted in order to enable an in depth analysis of that phenomena.

Challenged Projects: Partially successful completed and approved projects which are late, over budget, and have fewer features and functions than originally specified. The degree of challenge depends on the way constraints are applied and interpreted within the organisation and the priorities and tolerance available in any particular area.

Failed Projects: Projects that are: cancelled before completion, are never implemented or are scrapped following installation. May also apply to projects that involve significant litigation.

Forensic Systems Engineering: Post-mortem analysis and study of failed or challenged IT processes, products, artefacts, services or projects aimed at uncovering the systemic causes, dynamics and relationships that contributed to the shortfalls or failures.

Storytelling: A method of communicating and sharing ideas, experiences and knowledge in a specific context.

The Measurement and Recognition of Intellectual Capital in the Process of Accounting Convergence Trends and Patterns

M**Ionica Oncioiu***Titu Maiorescu University, Romania*

INTRODUCTION

The value of intangible assets has been the main focus in the debates between international professionals and business world for many decades. Today, this concept interferes with the international process of assessment, but also with the convergence of accounting. The need to determine the value of these assets lies in the more virulent criticism brought to the traditional accounting system, placed face to face with an increasing vision of financial assets of a company. In an uncertain world with imperfect and incomplete markets (financial crisis), no particular measurement objective should be regarded as having a monopoly, and different measurements should be regarded as complementing one another.

The problem is how to choose the best of them in the context of satisfying accounting information users' requirements, on the one hand and complying with accounting principles and fundamentals, on the other hand.

Basically, because of different research directions, so confusion can arise regarding the use of terms: intellectual capital, intangible assets, knowledge assets (knowledge assets). Thus, the term intangible assets is used especially in financial accounting, the term is used in active knowledge economy and the term intellectual capital / human capital management is used in particular.

Sometimes the term "intellectual capital" is considered synonymous with the term "intangible assets". OECD definition of the distinction made

however, by locating intellectual capital as a subset of intangible assets rather than intangible assets of the company. As a result, there are intangible elements of nature intangible assets are not logically part of intellectual capital. For example, reputation may be the result of judicious use of intellectual capital but is not part of it.

It is an intellectual potential, consisting of knowledge in different forms, which has the ability to make technological and management processes in a number of operational items, assets, creating value, to be integrated the final products both material and immaterial of a company.

Since historically, the distinction between intangible assets and intellectual capital was initially very vague. Producers and users of financial statements consider that in the measurement and presentation of accounting information, the most widely used is the historical cost, although it has some weaknesses. This is usually combined with other bases of evaluation. Moreover, the tendency is to use current cost accounting in response to failure based on historical cost model to solve problems connected with non-cash effect of changes in asset prices.

Accountants, in their turn, reflect accounting estimates only in registers and certainly not the value generated by the presentation of materials. It is a clear dichotomy between the historical cost accounting and fair value in accounting measurement, which creates a productive tension in discussions related to economy based on knowledge. However, it is more likely for a certain period of

time to develop both paradigms in parallel and then become convergent, thus solving the paradox of intangible asset valuation.

BACKGROUND

In recent decades there was an idea almost unanimous on the growing importance of intellectual capital. It is said more often today „intellectual capital is the engine of the new economy” or „knowledge and information” is the most valuable commodity. The new economy is recognized as an economy in which companies value lies in the knowledge and skills of managers / employees rather than tangible assets value.

Also, many economists have stressed the importance of intellectual capital as a production factor in comparison with traditional assets. There are two theories on economic growth: the theory of neo-classical growth theory exogenous and endogenous growth. Intangibles reflect different endogenous model.

According to endogenous growth model of production is defined as a function dependent accumulation of immaterial (intangible). Over time, permanent growth is possible only if the intangible capital stock changes virtually forever.

Management problem that intangible asset is not new, and the problem of „intellectual capital” is the last line developed in knowledge management.

The Norwegian Government has funded a research project to develop a model of skill capital of the company, he was later involved in an ISO-type certification process that included intellectual capital.

In all these experimental projects prevailing spirit of participation rather than a theoretical development, but current area of interest in intellectual capital research includes: the creation and management of intellectual capital, understanding how best to measure intellectual capital.

Scientists Pierre and Martory (2000) are indicating that “the scientific literature and official texts do not have clear definitions for the notion

of intangible asset. Most authors prefer to make a simple list of all the main components while accounting official texts define this vague notion. In fact, the general definitions quickly become useless due to their very little analytical and more precise definitions prove to be too limited to particular cases”.

Regarding the practical need of these measurements it can be supported by an international study on measuring the importance of intellectual capital in the businesses, led by Arthur Andersen in 1998 and held within the 368 companies in Europe, North America and Asia. The results of this study can be summarized as follows:

- The majority of respondents are convinced that reports on intellectual capital will increase quantitatively;
- Almost three quarters of respondents have already tried the organizations they lead, at least two non-financial measurement models;
- The majority believes that the assessment firm’s intellectual capital can be increased based on organizational performance;

Similar studies have been done so in 1998 a study conducted by Watherhouse and Svendsen (Measuring What People Know) and conducted on a sample of 114 top managers of the largest Canadian companies found that measuring intellectual capital is considered as a basic strategic condition and the company’s management should constantly reported, as often as other non-financial measurements such as innovation, product quality, customer relationships with investors and business partners, community relations, the effects company activities on the environment. Of all these non-financial measurements, the respondents were considered the least satisfied in terms of indicators that capture the firm’s intellectual capital.

However, in the ‘90s, new methodologies for measuring expression and intellectual capital have been developed and applied, such as “Value Platform” – developed by Dow Chemical – “Intangible

Assets Monitor” from Celemi, “Value Scheme” from Skandia, etc.

According to Sveiby (2010) a variety of metrics have been created to try to evaluate levels of intellectual capital in firms, both bottom-up (assessing individual employee knowledge and adding it all up) and top-down (calculating intangible assets as a remainder from financial statements). Preferences vary, but just note that measurement is a topic that has been explored in some depth and that dozens of potential choices are available to firms looking to evaluate and then better manage their own intellectual capital.

In general these models support the idea that intellectual capital components can be summarized in three groups: human capital, organizational capital and customer capital generated, the goal being to maximize the value of the company. A new emerging research trend seeks to apply the concept of intellectual capital to the field of entrepreneurship (Diego Matricano, 2016).

CHALLENGES ACCOUNTING FOR INTELLECTUAL CAPITAL: STAKES, ADVANTAGES AND LIMITS

Intellectual capital, measured by education and training at work, he has developed. This happened the importance of intangible assets for the success of companies has increased significantly and traditional economic concepts and management proved not to be adequate. Thus, the latest economic and management studies have focused on developing tools to facilitate a better understanding and representation of intangible.

The main issue in terms of measuring systems is that it is not possible to measure social phenomena with scientific precision. Thus it tends, in the absence of appropriate tools to use traditional methods based on the figures - financial indicators or remove the real cause of the phenomenon that caused the occurrence of these intangible assets.

Measure aims to discover opportunities for cost or value creation, which are otherwise hidden in

traditional accounts. How much does the replacement of staff? What is the lesson learned when company staff interaction occurs with customers? What is the lost opportunity to create value where there is inadequate management systems knowledge? These are some questions to answer which is given by measuring intangible.

To meet new challenges, many companies, especially in Northern European countries, have focused attention on the concept of intellectual capital and the adoption of tools for classification and identification of new determinants of value. Accordingly, they began to be used and linguistic terms such as “Intangibles Report” or “Intellectual Capital Report” (designate different models of identification and presentation of intangible assets).

Frustrations expressed by the studies mentioned on the measurement of intellectual capital are interesting and suggest, on the one hand, the need for tangible measurement of intangible assets in general and particularly intellectual capital, and on the other hand, that the attempt has been made now leaves room for improvement.

Considers that, before presenting in detail the different methods of reporting and measuring intangible are used so far, it is necessary to present a general classification will serve as a basis for analysis and will facility understanding of the specificity of each method. Sveiby proposed classification is presented in 2001 and is based primarily on the dichotomy that characterizes these methods and models: orientation atomistic/holistic orientation (e.g., models that measure and/or individual is intangible that models that measure and/or is intangibles overall available to organizations). Then, starting from this basic distinction is a distinction between methods of monetary and non-monetary:

- *Monetary methods are useful in mergers, acquisitions of new companies listing on stock exchanges, comparisons made between companies in the same field of activity, moreover, the methods money have*

been developed based on accepted accounting rules so that they can be easily learned by accountants;

- *Non-monetary methods is the determination of innovation and performance management company* also may be noted that these methods have the advantage that it does not express the intangible assets in financial terms, so that they can be applied at any level of organization in non-profit, public companies, and/or organizations with social objectives.

According to these differences, results in a matrix with four categories of methods:

Cell A - Holistic and Monetary Methods: A method includes using financial models and, on the one hand provides a general representation of the intellectual capital of organization supporting a determination of their money and, on the other hand, because of the nature their very general and brief no intangible enable administration for the purposes of creating value.

Cell B - Atomistic and Monetary Methods: Using monetary methods include but are applied to individual intangible assets. However, although these methods focus on individual assets do not provide a clear vision of the organization in total intangible assets of total assets.

Cell C - Atomistic and Non-Monetary Methods: Includes non-monetary atomistic methods, among them, being found most important innovations thousand and more in line with the "atmosphere" knowledge economy. But, and this presents some weaknesses, namely that it does not allow comparisons and that are not always able to accurately reflect the value of individual intangible assets and, therefore, often based on approximations.

Cell D - Holistic Methods, Non-Monetary: It is observed that the literature does not provide such methods of measurement yet eloquent.

Another classification of the methods used so far was made all the Sveiby, based on ideas suggested by Luthy (1998) and Williams (2001), is as follows:

- **Direct Assessment of Intellectual Capital Methods (DIC):** Estimating the value in cash of intangible assets by determining the various components. After these components are identified, they can be measured directly, either individually or with a compound coefficient.
- **Market Capitalization Methods (MCM):** Calculate the difference between a company's market capitalization and shareholder equity value as the value of its intellectual capital or intangible assets.
- **Method Returns on Assets (ROA):** Average pre-tax earnings of a company for a period of time are divided by average assets of the company's tangible. The result is ROA - return on assets of the company which then compares the average of its industry. The differences multiplied by the average of tangible assets of the company to calculate the average annual earnings from intangibles. Dividing the average earnings above the cost average capital company or the interest rate, you can estimate the value of intangible assets or intellectual capital thereof.
- **Method Scores (Scorecard - SC):** Establish various components of intangible assets or intellectual capital and generate indicators or indices that are reported as scores (scorecards) or graphics. These methods are similar methods in that DIC is based on identifying components of intangible assets, but unlike, they do not consider the need for a financial estimate of the value of intangible assets.

Methods offer different advantages. Thus the actual amount that provides evaluations (\$), such as ROA and MCM, are useful in cases of merger and

acquisition and stock assessments. May be used to compare companies within the same industry and are useful to illustrate the financial value of intangible assets, feature that tends to draw attention to managers. Because they are based on counting rules long established and recognized them in different national laws or regulations can be easily communicated to the accounting profession. The disadvantage is the fact that all the terms, making money, they can be superficial.

DIC and SC methods advantages are that it can create a more complex organization in terms of health than financial metrics and ROA and MCM can be applied easily to any level of organization. Measurement is made closer to the event and so reporting can be more accurate and faster than financial measurements. Their disadvantage is that the indicators are contextual and must be adapted to each organization and each objective, which makes comparison difficult. Also, the methods are new and are not easily accepted by the companies and managers who are accustomed to seeing it only in purely financial.

From the perspective of global experience in evaluating intangibles, we can remember the following methods:

- Skandia Navigator, is probably the most well-known and analyzed intellectual capital measurement model that literature offers to date. Thanks to its innovative qualities it is considered a kind of archetype for the class of models for non-monetary atomist measuring. Among the advantages of this model are: it initiated a series of efforts toward abandoning traditional ideas about the determinants of value in organizations and bringing to the fore the possibility of measuring intangible assets; it highlights the crucial role of customer relationships in business value creation and shows how the nature of the relationship between companies and their customers changed; for the first time a great importance is attached to the contribution that can bring renewal and development of the organization's value;
- The Intangible Assets Monitor is a model that can be applied to all companies and its indicators are chosen according to the characteristics of the company, industry, competitive variables, company's history, organizational culture and company management. According to the model, intangible income can be divided into: strengthening the company's image, strengthening the organization and enhancing its competence;
- The Balanced Scorecard model that Kaplan and Norton have designed does not present a general system of indicators to measure intellectual capital, the argument being that a general system of indicators could leave out certain aspects of each company related to its objectives, processes and the context in which it operates. Traditional indicators provide information about the past performance of the company, while non-financial indicators can measure, for example, customer satisfaction, corporate innovative activities and improvements to various processes of the company. Combining the two types of indicators allows managers to obtain a balanced image on company performance. Braam (2001) points out that diffusion of innovative practices of management or management accounting, such as the Balanced Scorecard, can bring a discrepancy between the "tag" of the implemented concept and its contents. Authors such as Guthrie and Dumay (2015) have shown that this model can be used for evaluating intellectual capital. Comparison made between Balance Scorecard and intellectual capital valuation methods reveal other similarities such as recognition strategy as an essential component of performance management need to find a concise and fully expressing the company's situation, attention to intangible assets and knowledge. In other words, it is possible that some organizations only claim that they use the Balanced Scorecard, while

the implementation substance ignores the considered defining characteristics of Balanced Scorecard (connecting indicators to the strategy, causal relationships between indicators etc.). Therefore, to assess whether the Balanced Scorecard is used in practice is dependent on defining and clarifying the concept, ie the unequivocal identification of the characteristics that define a Balanced Scorecard;

- Lev's Knowledge Capital Formula Method for measuring knowledge capital is based on the traditional method for calculating goodwill. One difference is that, in Lev's model, additional revenues are replaced by residual income attributable to intangible assets. Based on these critical observations, in the scientific literature, the Value Chain Scoreboard has appeared, which is an atomistic method of measuring intangible assets which are pending completion;
- Criticism of the method Tobin's Q considered the volatility of the courses on the capital market, often beyond the control of management, which directly affect the market value of the intellectual capital of the company, and focused on the fact that this way of calculating the intellectual capital is not very robust. However, in his work, Stewart argues, by exemplifying, that both market value and book value may be diminished. Another weakness of the method is that the indicator "Q" may not be determined by the value of intangible assets;
- The Economic Value Added model was initially developed as an internal measure of performance used by managers to help them make decisions on investments. In practice, it has been demonstrated that an investment project which causes higher income than the net cost of its funding and thus generates profit, does not necessarily lead to an increased company value;

- Technology Broker stated that it is itself an intellectual asset for organizations. But the main weakness of the model is considered to be related to the possibility of transforming the qualitative results in monetary results. For example, using market-based approach is likely to not be identified on the market any information on prices of many of the elements that form the intellectual capital. Income-based approach suffers from the bias of the estimates cash-flow;
- The Weighted Patents Citation model estimates a "technological factor" to identify the impact of research-development efforts leading to the creation of intellectual property and uses indicators such as research-development expenses / sales, number of patents, income / research-development expenses, etc. "The evaluation process of patents" is an effort that involves teamwork, in which specialists interact in marketing and research-development to decide the viability of abandoning or continuing the research-development process.

Based on this overview of the different methods of measuring intellectual capital of companies, we can say that the literature and practice offer several models, each with its strengths and weaknesses. Moreover, we can see that some of the intellectual capital valuation models are built on the same principles and using the same indicators, but they are named differently. For example, Skandia Navigator uses the term "human capital", Technology Broker uses the name "people-based activities" (human-centered assets) and Intangible Assets Monitor uses "staff competence" to denote the same concept.

Andriessen and Stam (2004) on the other hand, have introduced significant changes, particularly with respect to organization. They break down each of the other IC groups between assets, investments, and effects. This approach is intuitively appealing as it provides a reading on the current stock of

intellectual capital, its impact, and its potential for the future.

There are other versions, some adding categories, but most scholars stick with the three basic categories (and virtually all research includes the three in some form) (e.g. Andreou et al., 2007). If knowledge assets are the single most valuable asset available to a firm seeking to establish sustainable competitive advantage in the marketplace, per the resource-based/knowledge-based theory of the firm, there are clear implications for corporate governance. As a result, at the intersection of intellectual capital and corporate governance is an interest in disclosure. The art of managing knowledge assets is complex and, as previously noted, the payoff may be long in coming and difficult to tie directly to investment in knowledge. While the specifics disclosure are still early in development (Abeysekera, 2006), there have been some moves toward standardization. Managers recognize the value of showing investment and interest in developing intangibles, and voluntary disclosure rectifies issues such as information agency (Cerbiom and Parboncui, 2007). Guthrie and Dumay (2015) reveals that interest in public sector and IC is alive and well and the major difference between public sector and mainstream IC research is a complete lack of normative research.

This demonstrates that the field is still at an early stage and no one wants to give up their own nomenclature borrowing aspects of the work of others, as long as it was shown that one method is generally valid. For this reason, an important aspect of research is to develop a common set of definitions and indicators that can be easily generalized.

An optimistic attitude regarding the chances of success of a development project will lead to the capitalization of development costs, with implications for the size of the result, the capitalization for the year. In future exercises, expenses recording with amortization will impact the result in the sense of decreasing it. If the enterprise has the interest payment of development costs in the

profit and loss (expenditure) it will claim that at least one of the conditions stipulated by IAS 38 para. 43 is not met.

One criticism of the procedures above concerns that they contain a certain „dose” of creative accounting with the character of subjectivity with the realization that they change the presentation of accounts, but are nonetheless the result of an option, when selected from among more methods, which makes us say that they are within the accounting regulations.

Referring to development costs, for example, their positioning on the boundary between evaluation and accountability requires the existence of a bias caused by handling interest income. If the enterprise has a policy aimed at transferring these expenses in the profit and loss (expenditure) account, we can claim that at least one of the requirements for classification as an intangible asset development costs - which would affect the balance - is not met.

The Fourth Directive states that these expenses must be charged to the profit and loss, unless national law does not provide for their capitalization. These provisions were acquired differently by the EU countries, in Germany is forbidden to capitalize these expenses. France is the opposite, considering them capitalizable and depreciable in 5 years. In Belgium, Italy and the Netherlands, the capitalization of the costs of research is questionable, the trend being the exclusion of these expenses from the asset. Development costs can both be capitalized and amortized within 5 years (Belgium and Italy) or during their useful life (the Netherlands). In Spain and the UK, research costs are non-capitalizable under certain conditions and can be amortized for a period of 5 years, respectively during the useful life.

All the measures against “fraudulent” accounting practices must have the purpose to ensure a capital accounting information on qualitative intellectual capital that faithfully represents reality and to allow decisions to ensure an optimal allocation between participants in value creation.

FUTURE RESEARCH DIRECTIONS

Therefore, their representation in situations of intellectual capital value as intangible asset, to a value as close as possible to the real one, is essential for users since they ground their decisions based on its level.

Moreover, the success of any company depends not least on whether the intellectual capital has recognized the important role of an examination over several years of the development of intangible values both in the construction of the house of intangible values and the annual evaluation, and if it has implemented this knowledge consistently in everyday decisions.

The author believes that due to the high influence, both positively and negatively, of the intellectual capital on its value is necessary for conducting more detailed studies of diagnostic analysis to identify and measure them, and also to determine a set of evaluation methods that leads to a correct result, close to the market.

Of course, the appeal to intangible resources in order to create value remains a significant challenge for specialists because the country's competitive advantage should be interpreted in relation to the skills, talent and intellectual capital stock. Future research efforts of SMEs should turn those scales into operational parameters and promote the development of empirical methodological studies based on dialogue and the dissemination of knowledge and experience in interdisciplinary research teams.

CONCLUSION

Companies have recognized the importance of intellectual capital as a key factor for success, looking for ways to measure and control this factor. Financial performance measurement is not sufficient that reflects how the profit is due to

intellectual capital. In the knowledge economy, wealth is created through a combination of services and products that generate value. Each company has its unique set of knowledge, skills, values and solutions, all intangible resources. Monitoring and management of these resources can be made with value creation efficiency analysis because it is a priority for managers to obtain and maintain productivity and efficiency in the enterprise.

Another aspect that can be observed is related to the fact that some researchers have developed theories based on reflection the results of case studies, others have developed some indicators that have applied in practice (as in the case of Skandia, for example) but without testing their validity in advance and others have developed sets of indicators were not tested. For this reason another challenge in the study of intellectual capital is the need for researchers to base their conclusions on several empirical researches. And, to make conclusive statements about intellectual capital, we need to consider that a coherent field theory has positive effects in multiple areas such as accounting, human resources, finance, strategic management, etc.

Very important is that those who investigate aspects of intellectual capital to set realistic goals. As with many challenges, such as efforts to conceptualization of the field, trying to measure intellectual capital; setting causal relations - effect, demonstrating the accuracy of proxies, it happens that many companies have an intellectual capital management system its use in some individual indicators whose relevance is not fully justified.

It is therefore important to determine and establish relevant indicators, the group of them in specific categories (quantitative or qualitative aspects related to individual company's business - e.g. employees - or macro issues) and not I just put together, and also to assign each category a specific weight in calculating the final result.

REFERENCES

- Abernethy, M., Bianchi, P., Labory, S., Lev, S., Wyatt, A., & Zambon, S. (2003). *Study on the Measurement and Reporting of Intangible Assets and Associated Reporting Practices*. Prepared for the Commission of the European Communities – Enterprise Directorate General Contract, contract n. FIF.20010720.
- Abeysekera, I. (2006). The project of intellectual capital disclosure: Researching the research. *Journal of Intellectual Capital*, 7(1), 61–77. doi:10.1108/14691930610639778
- Andreou, A., Green, A., & Stankosky, M. (2007). A framework of intangible valuation areas and antecedents. *Journal of Intellectual Capital*, 8(1), 52–75. doi:10.1108/14691930710715060
- Andriessen, D. G., & Stam, C. D. (2004). *The Intellectual Capital of the European Union*. Diemen, Netherlands: Centre for Research in Intellectual Capital.
- Anson, W. (2008). *The intangible assets handbook: maximizing value from intangible assets*. Washington, DC: American Bar Association, Section of Business Law.
- Bontis, N., Dragonetti, N. C., Jacobsen, K., & Roos, G. (1999). The Knowledge Toolbox: A review of the tools available to measure and manage intangible resources. *European Management Journal*, 17(4), 391–401. doi:10.1016/S0263-2373(99)00019-5
- Braam, G., & Nijssen, E. J. (2001). *Antecedents of the Adoption of the Balanced Scorecard*. Working Paper. University of Nijmegen.
- Campisi, D. (2006). *The measurement of intangible assets: a method to assess knowledge productivity*. 2006 Business & Economics Society International Conference, Worcester, MA.
- Cerbioni, F., & Parbonetti, A. (2007). Exploring the effects of corporate governance on intellectual capital disclosure: An analysis of European biotechnology companies. *European Accounting Review*, 16(4), 791–826. doi:10.1080/09638180701707011
- Cravera, A., Maglione, M., & Ruggieri, R. (2001). *The Valuation of Intellectual Capital*. Il Sole 24 Ore.
- Doppegieter J., Zoller, M. (2008). *Adjusting economic value added for intangible assets*. Bruchsal: School of Business Administration.
- Epingard, P. (1999). *L'investissement immateriel, coeur d'une economie fondee sur la savoir*. Paris: Ed. CNRS.
- Guthrie, J., & Dumay, J. (2015). New frontiers in the use of intellectual capital in the public sector. *Journal of Intellectual Capital*, 16(2), 258–266. doi:10.1108/JIC-02-2015-0017
- Kaplan, R. S., & Norton, D. P. (1992). The Balanced Scorecard – Measures that Drive Performances. *Harvard Business Review*, 71–79. PMID:10119714
- Kwansa, F. A., Mayo, C., & Demirciftci, T. (2008). The importance of intangible assets: trends and patterns. In *Handbook of hospitality strategic management* (pp. 181–202). Amsterdam: Butterworth-Heinemann. doi:10.1016/B978-0-08-045079-7.00008-9
- Lev, B. (2001). *Intangibles: Management, Measurement, and Reporting*. Washington, DC: Brookings Institute Press.
- Lynn, L. E. (1999). *The Management of Intellectual Capital: The issues and the practice*. Hamilton, Ontario: Society of Management Accountants of Canada.
- Martory, B., & Verdier, F. (2000). Comment traiter le goodwill? Pratique d'une theorie, theorie d'une pratique. *Comptabilite, Controle, Audit*, 6(2).

Matricano, D. (2016). The impact of intellectual capital on start-up expectations. *Journal of Intellectual Capital*, 17(4), 654–674. doi:10.1108/JIC-04-2016-0040

Roos, J., Roos, G., Dragonetti, N.C., & Edvinsson, L. (1997). *Intellectual Capital: Navigating in the New Business Landscape*. Macmillan.

Sveiby, K. E. (2010). *Methods for Measuring Intangible Assets*. Retrieved: [www.sveiby.com.au/Intangible Methods.htm](http://www.sveiby.com.au/IntangibleMethods.htm)

KEY TERMS AND DEFINITIONS

Assets: As future economic benefits controlled by a particular business entity and also as economic resources which are owned by a business.

Estimate: Opinion regarding the value.

Intangible: Represent immaterial assets.

Valuation: The process of determining the value of an asset.

Project Control Using a Bayesian Approach



Franco Caron

Politecnico di Milano, Italy

INTRODUCTION

Forecasting is a critical activity in project management: relying upon sound estimates to complete, the project manager can steer the ongoing project in order to meet specific time and cost objectives (Dvir and Lechler, 2004). Moreover, foresight is needed to avoid constantly being forced to manage emergencies, since emergency implies a reactive action. Without anticipation there can be no rationale in making a decision and what we can do is just to be adaptable to changing circumstances.

Planning and forecasting are strictly intertwined both in the early stage when the project baseline must be determined and throughout the entire project life cycle when the project baseline has to be followed (Hogarth and Makridakis, 1981). Forecasting feeds (re)planning (corrective measures are taken based on forecast) and (re)planning feeds forecasting (corrective measures will influence the future).

In the project control process the role of the Estimate to Complete (ETC) is critical, since the information drawn from the ETC, compared with the project baseline, should highlight the need for and the type of corrective action that may improve the project performance. In fact, ETC is the base for any effective corrective action. This approach to project control corresponds to a *feed-forward* type control loop (Anbari, 2003; Christensen, 1996), since the forecast informs present-day decisions.

From a recent survey (Morrow, 2011), analyzing the data of more than 300 global mega-projects, it appeared that in 2010 65% of the industrial projects with a minimum budget of 1 billion US dollars did not succeed in meeting the objectives of cost, duration and quality. However, it remains an open question whether these failures are due to

a poor performance during the execution stage or to a lack of forecasting accuracy during the planning and control process. In the former case, both positive and negative deviations from the project baseline should be expected. On the contrary, a systematic overrun in terms of cost and/or time may be easier explained as a weakness in the forecasting process since the beginning of the project. As a consequence, the forecasting process plays a critical role.

To explain the accuracy of the forecasting process, some considerations must be developed about the knowledge sources feeding the process, the forecasting techniques to be applied and the mitigating measures taken in order to avoid possible biases affecting the forecasting process.

As shown in Figure 1, at a given time of the project duration, i.e. the time now (TN), a certain amount of the work will be already completed (Work Completed, WC), while the rest of the work will be ahead, corresponding to the Work Remaining (WR). The cost and time performance related to the Work Completed will be known, while a forecast will have to be developed for the WR.

It should be noted that both the *accuracy of the forecast about WR* and the *impact of the corrective actions* that may be implemented based on the forecast depend on the progress of the project at the Time Now. The effectiveness of the corrective actions is greater in the early stages of the project execution and progressively diminishes while progress increases: in fact, as progress increases, the degrees of freedom available to steer the project tend to reduce progressively. On the other hand, the capability to forecast the project final duration and the final cost follows an opposite trend. In fact, at an early time in the execution phase, the knowledge available to the decision maker is

DOI: 10.4018/978-1-5225-2255-3.ch493

Figure 1. Estimation at completion at time now (internal view)



scant and rapidly evolving; therefore, the capability to provide a *reliable forecast* is jeopardized, particularly if the forecast is based only upon the knowledge related to the ongoing project.

Drawing on a set of case studies (Caron et al., 2013a; Caron et al. 2013b, Caron et al., 2016), this paper proposes a Bayesian approach to determine the estimate to complete for a project. The paper has a twofold objective:

- To identify all the available knowledge in order to improve the forecasting process;
- To develop a Bayesian approach in order to integrate in a formal and rigorous way the diverse knowledge sources.

The second section analyzes the different knowledge sources available; the third section addresses the issue of possible bias during the forecasting process; the fourth section introduces the Earned Value Management approach frequently used to determine the estimate at completion for a project, both in terms of cost and time; the fifth section introduces the general structure of a Bayesian model and eventually some results are given stemming from the application of the model to some projects in the oil and gas industry.

BACKGROUND

As mentioned above, all the knowledge available should be used in order to address the planning and

control process for a complex project (Caron, 2014; Reich et al., 2014; Schindler and Eppler, 2003).

In general, the knowledge available to the project team may be classified in two ways: explicit/ tacit and internal/external. Explicit external knowledge corresponds to data records about projects completed in the past, including measures of forecasting capability in terms of the difference between estimated and actual final cost and duration. Taking into consideration past experience should mitigate possible “optimistic” bias in estimating future project performance (Lovallo and Kahneman, 2003). Explicit internal knowledge corresponds to data records concerning the work completed WC, allowing for an evaluation of project performance at Time Now. Tacit external knowledge concerns the similarities between the current project and some past projects in order to transfer past data to the current project. Tacit internal project is about possible events/situations affecting the project’s work remaining.

In fact, relying only on data records related to WC while developing a forecast could be misleading, since it would be similar to driving a car whilst looking just in the rear view mirror, so making it impossible to dodge the obstacles that may lie on the route ahead. It should be noted that the contribution deriving from the data records to the estimate at completion may be considered reliable just for the near future and losing progressively accurateness if a more extended horizon is considered. Since the reliability of the estimation based just on data records rapidly decreases, so the

contribution deriving from the subjective information expressed through experts' opinions becomes predominant, particularly at the project outset.

In general, the specific contribution given by tacit knowledge i.e. by experts about the future development of the project, may concern:

- Hindsight, i.e. monitoring drivers explaining the project development during WC, and presumably affecting also WR, e.g. answering the question what kind of plausible drivers may have generated the actual development of the project till the Time Now and how they will also influence the future? (e.g. drivers such as schedule aggressiveness, degree of engineering completeness, owner involvement, turnover in project leadership, anomalous low bid from subcontractors, unsatisfied stakeholders, new technology, project team integration, project team staffing, etc.).
- Insight, i.e. making sense of the weak signals that anticipate the emerging situations possibly affecting project performance (e.g. weak signals such as frequent scope changes, permits delay, engineering sequence not aligned with construction, high rate of rework in construction, missing data in design process, etc.) (Williams et al., 2012; Haji-Kazemi et al., 2013).
- Foresight, i.e. anticipating certain/uncertain events or conditions affecting project performance during WR which may originate both internally and externally to the project. Certain events may include planned corrective actions or contractual constraints, while uncertain events, i.e. risks, may arise both in terms of threats (e.g. adverse weather conditions) and opportunities (e.g. more efficient solutions deriving from suppliers' collaboration).

The final objective of this chapter is to develop a Bayesian approach integrating the knowledge contribution deriving from different sources, i.e.

data records related to the ongoing project, experts' judgment about the expected performance during WR and data records deriving from similar projects completed in the past.

FORECASTING BIAS

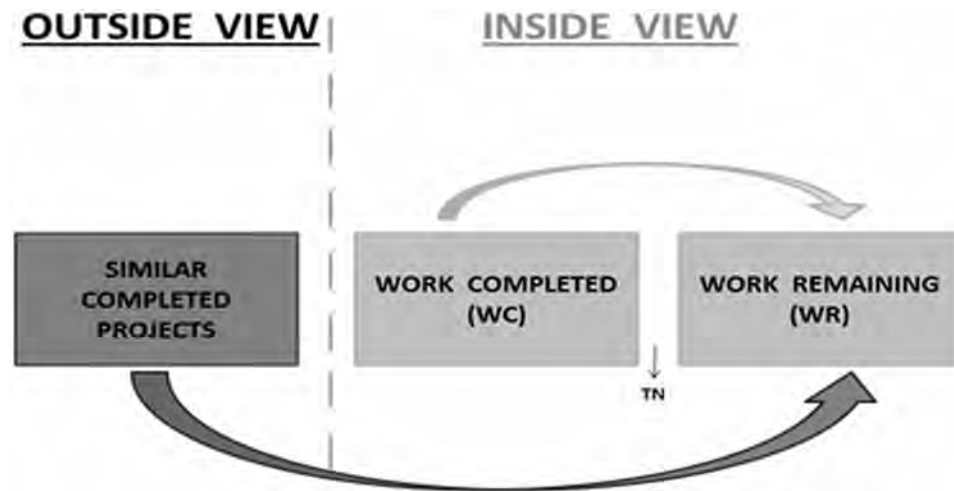
The knowledge contribution given by external knowledge related to similar projects completed in the past allows for stabilizing the forecast throughout the project life cycle, by making the forecast less sensitive to performance variability in particular at the project outset, and mitigating any possible bias effect on the forecast.

As for forecasting bias, Kahneman's and Tversky's studies (1979) showed that a major source of forecasting failure, which influences the accuracy of final cost and duration estimates, is linked to an exclusively "*internal*" view of the project, i.e. based only on knowledge deriving from inside the ongoing project. As a consequence, the research focus has moved to the psychological and political factors affecting the project planning process (Lovallo & Kahneman, 2003), and, in particular, two main sources of planning failure have been pointed out (Flyvbjerg, 2006; Flyvbjerg, 2009).

Firstly, the *cognitive bias* (availability, representativeness, anchoring and adjustment, etc.) normally generating over-optimism, i.e. the common attitude to assess future projects with greater optimism than justified by the actual previous experience. Secondly, the *strategic and political pressures*, that may typically emerge during proposal preparation. Indeed, the approval of a project pre-supposes a competition involving different proposals, which often causes a voluntary underestimation of cost and duration by the project proposers in order to make their own proposal as attractive as possible.

Even more so, the need emerges to exploit all the available knowledge during the planning process, in order to minimize any bias effect. In fact, as shown in Figure 2, the traditional control process often focuses only on data related to the

Figure 2. Internal and external view



current project, corresponding to an exclusively “internal” view (Flyvbjerg, 2006). An integration is needed between the “internal” and the “external” view, where the latter is based on knowledge related to the past experience (see Figure 2) (Flyvbjerg, 2006).

In fact, the current project can be viewed as belonging to a cluster of similar projects completed in the past. Note that the selection of the cluster of similar projects is basically subjective since it depends on the similarity criteria adopted (Green & Armstrong, 2007). Some cases, in fact, may express strong ambiguity. For example, if a company has to estimate the cost of an investment in a new technology and in an unfamiliar technological domain, should it take into account the set of highly innovative projects developed in different technological domains or the set of barely innovative projects but belonging to the same technological domain? Neither the former nor the latter option may be the best solution but both should be considered and weighted (Kahneman & Tversky, 1979). Besides similarity criteria, the subjective assessment should also consider the trade-off between using a large number of past projects, leading to the risk of including projects substantially different from the current one, and a small number of projects, leading to the risk of losing statistic significance. Also the transfer-

ability of past data to the current project and the possible presence of outliers in the distribution should be critically evaluated.

EARNED VALUE MANAGEMENT SYSTEM

The future is something having its seeds in the present (and in the past) (Kuosa, 2012). This is the basis for the forecasting process. In general, different approaches are available in order to implement the forecasting process during project control (trend analysis, pattern analysis, simulation, etc.). In particular, in project management Trend Analysis has become very popular. Trend Analysis is based on the extrapolation into the future, i.e. into the WR, of the actual trend at Time Now of a performance index, e.g. Cost Performance Index (CPI) and Schedule Performance Index (SPI), related to WC. This is the traditional approach applied in the Earned Value Management, representing a very popular methodology used to develop an estimate at completion for the project (Fleming and Koppelman, 2006).

As Project Management Institute (2013) stated, the main processes involved in project management are: initiating, planning, executing, monitoring, controlling and closing. In particular, Earned

Value Management (EVM) represents an effective way of addressing the project control process. EVM is an efficient performance measurement and reporting technique for estimating cost and time at completion (Project Management Institute, 2013; Marshall et al., 2008). The following basic parameters are used in EVM, where TN indicates Time Now, i.e. the time along the project life cycle at which the control process is implemented:

- Planned Value (PV), the budget cost of work scheduled at TN;
- Earned Value (EV), the budget cost of work completed at TN;
- Actual Cost (AC), the actual cost of work completed at TN.

EVM was improved by Lipke (2002_a; 2000_b; 2003), who introduced the concept of Earned Schedule (ES) for obtaining a measure of the schedule performance index based on time units and overcoming the flaws associated with a Schedule Performance Index SPI defined as the ratio between EV and PV, both of them expressed in monetary terms. Earned Schedule is the time at which the EV value achieved at TN should have been obtained according to the project baseline. The new Schedule Performance Index SPI(t) at TN, defined as the ratio between ES and TN, represents a more effective approach, since it avoids the problem of the convergence of the EV and PV values toward BAC, i.e. Budget at Completion, as the project reaches completion (Lipke, 2006_a; Lipke, 2006_b).

The above three parameters and the ES, all of them evaluated at TN, allow for the calculation of a set of indices and variances at TN. The most important of these are:

- Cost Performance Index:

$$CPI = EV / AC$$

- Cost Variance:

$$CV = EV - AC$$

- Schedule Performance Index:

$$SPI_{(t)} = ES / TN;$$

- Schedule Variance:

$$SV_{(t)} = ES - TN$$

Variances CV and SV(t) summarize the project's past performance during WC, while indexes CPI and SPI(t) may be used in order to extrapolate the current trend into the future and estimate the future performance during WR (Anbari, 2003).

Many Estimate at Completion formulas have been proposed during almost 50 years of EVM application but none of them has proved to be always more accurate than any other one (Christensen, 1993). In the basic approach, the estimations of final cost (i.e. EAC) and final duration (i.e. TAC, Time at Completion) are based on the following equations:

$$EAC = AC + (BAC - EV) / CPI_f \quad (1)$$

where:

AC = Actual Cost at TN;

BAC = Budget at Completion;

EV = Earned Value at TN;

CPI_f = Cost Performance Index estimated for the work remaining (WR).

$$TAC = TN + (PAC - ES) / SPI_f \quad (2)$$

where:

TN = Time Now;

PAC = Planned at Completion, i.e. the planned duration of the project;

ES = Earned Schedule at Time Now;

SPI_f = Schedule Performance Index estimated for the work remaining (WR).

It should be noted that future performance may significantly differ from past performance. The new performance indices CPI_f and $SPI(t)_f$ have

been introduced in equations 1 and 2 with reference to the Work Remaining and may consider a possible evolution of the project different from the expected, in particular in terms of performance. While the generic indices CPI and SPI(t) are related to the overall WC, CPI_f and $SPI_{(t)f}$ will be related to the overall WR. In fact, relying only on past performance while developing a forecast could be misleading.

Both equations 1 and 2 indicate that the values assigned to the performance indexes CPI_f and SPI_f play a critical role in order to obtain an accurate estimate of the final cost and duration. As a consequence, forecasting capability can be improved by utilizing all the available knowledge about the performance indexes CPI_f and SPI_f (Liu & Zu, 2007; Goodwin, 2005). The Bayesian approach allows for integrating in a rigorous way the knowledge content stemming from diverse knowledge sources.

BAYESIAN APPROACH

The Bayes Theorem represents a rigorous and formal approach allowing for an update of a prior estimate, which expresses the experts' preliminary opinion, by means of the data gathered in the field. For instance, the project team may assume a prior estimate of the final budget overrun, based on subjective expectations about the development of the ongoing project, and this prior estimate may be updated based on the actual performance of the current project at Time Now (Caron et al., 2013). It should be noted that the two knowledge sources should be considered independently, so avoiding any anchoring bias. In this way knowledge stemming from two different sources, data records and expert's judgment, can be integrated. In a Bayesian framework, the experts' preliminary opinions, are an example of use of subjective probability, the only probability concept applicable to non-repetitive processes such as the projects. Subjective probability is defined as the degree of belief in

the occurrence of an event, by a given person at a given time and with a given set of information. It should be noted that increasing the level of knowledge available may modify the value of the subjective probability assigned to a future event (De Finetti, 1937; Caron and al., 2013). While the metaphor of 'frequency based' probability is the dice throwing (i.e. a repetitive process), the metaphor of the subjective probability is the bet (i.e. a unique process). In general, we can assume that any proactive action involving the future is similar to a bet. In fact, De Finetti defined probability as a price. Let's consider a gamble where if the event E occurs the better wins 1, if it doesn't happen the better wins 0. How much is the better prepared to pay for accepting the gamble? That price corresponds to the subjective probability value associated to the event E. The probability value corresponds to the degree of belief about event E occurring.

The essence of Bayesian inference is using probability to describe our state of knowledge about some event or parameter of interest. A prior distribution (based on expert's tacit knowledge) is updated by means of experimental observations (data records) collected during the execution process, in order to obtain a posterior distribution, integrating both knowledge types.

For instance, the project team may assume a prior distribution of the final budget overrun, based on subjective expectations about the development of the current project, and this prior distribution may be updated based on the actual performance of the current project until Time Now (Caron et al., 2013).

Hence, Bayesian approaches to formulate valid forecasts have been proposed in literature (Palomo et al., 2006; Gardoni et al., 2007; Kim and Reinschmidt, 2009). The adoption of a Bayesian approach enables the elaboration of probabilistic time and cost estimates at completion based on multiple knowledge sources. This statement is easier to understand if the formulation of the Bayesian theorem is considered:

$$f(\mu | y_1, y_2, \dots, y_n) \propto L(\mu; y_1, y_2, \dots, y_n) \cdot f(\mu) \quad (3)$$

Eq. (3) consists of three main components:

- $f(\mu)$ is the prior distribution of the parameter of inference, μ : it summarizes the initial subjective opinion detained by the decision-maker about the probability density function of the parameter μ ;
- $L(\mu; y_1, y_2, \dots, y_n)$ is the likelihood function, obtained after the collection of the n experimental data $y_1, y_2, \dots, y_n$;
- $f(\mu | y_1, y_2, \dots, y_n)$ is the posterior distribution, i.e. the distribution that expresses the knowledge acquired about the parameter μ after updating the initial subjective judgments with the experimental data.

CASE STUDIES

Firstly, the Bayesian approach requires the definition of the inference parameter, for instance the performance indices CPI_f and $SPI(t)_f$ in order to forecast the final variance between budget cost and actual cost (or between scheduled duration and actual duration). Secondly, it requires the definition of the prior and likelihood distributions in order to obtain the posterior distribution. The prior distribution is based on the elicitation of the experts' tacit knowledge. It may be integrated by the data records related to projects completed in the past. The likelihood distribution is based on data records related to the ongoing project. The Bayesian approach allows for the integration of the diverse knowledge sources. The prior distribution updated by the likelihood function gives the posterior distribution. It should be noted that the Bayesian approach doesn't offer a point estimate, such as in the traditional EVM approach, but a distribution estimate, so increasing the knowledge available for the project team. Moving from the

prior distribution to the posterior distribution the degree of belief in the estimate increases, since a larger knowledge content has been used. Assuming that a standard distribution, e.g. a normal distribution, has been used, this improvement corresponds to a reduction of the standard deviation of the posterior distribution compared with the prior one.

An advanced Bayesian model has been applied to a set of large engineering projects in the oil and gas industry (Caron et al. 2013a; Caron et al. 2013b; Caron et al. 2016) in order to forecast the actual final duration and the actual final cost. The model utilizes three different knowledge sources: data records related to the ongoing project (indicating the actual performance trend during WC), experts' judgment about the ongoing project (looking forward to foreseeable events during WR) and data records about similar projects developed in the past (stabilizing the forecast during the project life cycle and mitigating a possible bias).

Assuming the final cost (or duration) as the inference parameter, the sequence of steps required to implement the model are:

- Determine the distribution of experts' opinions;
- Determine the distribution of similar past projects;
- Define a suitable value for the weight of the two knowledge sources included in the prior distribution;
- Determine the prior distribution;
- Determine the likelihood function;
- Obtain the posterior distribution.

The prior distribution is a summary of the initial knowledge available to the decision-maker for the forecasting process. In this model, the project team can rely on two main information sources: experts' opinions about possible events occurring in WR and data records related to similar projects completed in the past. By integrating these two components, a starting assumption about the distribution of the inference parameter (e.g. budget overrun or completion delay) can be formulated.

In order to translate these concepts in Bayesian terms, firstly two distributions have to be defined: one representing the experts' opinions and one the cluster of similar past projects. Note that the two components of the prior distribution have a subjective nature since the similarity criteria applied to past projects and the forecast of future event are based on experts' judgment.

A unique distribution has to be derived by the integration of these two components. In mathematical terms, the integration can be obtained by the concept of *mixture*: assigning a value ranging from 0 to 1 to a weight ε , the prior distribution is obtained as the weighted sum of the distribution of experts' opinions and the distribution of the cluster of similar projects:

$$\Pi(\cdot) = \varepsilon \Pi_E(\cdot) + (1 - \varepsilon) \Pi_S(\cdot) \quad (8)$$

In Equation (8):

- $\cdot$ is a generic point of the prior distribution;
- $\Pi_E(\cdot)$ is the evaluation of the distribution of experts' opinions in a generic point;
- $\Pi_S(\cdot)$ is the evaluation of the distribution of the cluster of similar projects in a generic point;
- ε is the weight.

Adjusting the value assigned to the weight ε it is possible to give more or less weighting to one of the two informative sources. When ε increases, the weight of the experts' distribution increases, while the weight of the distribution of similar projects decreases. Consequently, the prior distribution is characterized by a high level of flexibility, since it can be adapted to the project features.

As for the likelihood function, the experimental data collection will be represented by a single observation of the performance index at the Time Now. The posterior function is a mixture of the posterior distribution of similar projects and the posterior distribution of experts' opinions through the posterior weight ε^* .

Table 1. Comparison of the mean squared error

Mean Squared Error	Bayesian Model	EVM
MSE Cost	$1.23 \cdot 10^{-2}$	0.1683
MSE Time	$9.3 \cdot 10^{-3}$	$3.55 \cdot 10^{-2}$

In order to compare the Bayesian approach and the traditional EVM approach, a set of large projects related to the Oil&Gas industry have been analyzed in detail. These projects were characterized by a high level of uncertainty and complexity. After the projects reached completion, the forecasting error related to the Bayesian model were compared with the one that would have been obtained if traditional EVM forecasting formulas had been applied. The result was that the Bayesian approach allowed for a better forecasting accuracy independently from the type of project (on shore, off shore, subsea) and in particular since the early stage of the project when the information is scant and the potential impact of decisions is very high.

Eventually, it was possible to check the error made by the Bayesian model and to compare it with the one that would have been made if traditional EVM forecasting formulas had been applied. Table 1 shows the mean squared error made by the Bayesian Model and by EVM in the three Oil and Gas cases.

It is clear that the Bayesian model is more accurate: its mean squared error is ten times lower, both for cost and time performance, due to the larger amount of information used, even though the model was tested on very complex and risky projects.

FUTURE RESEARCH DIRECTIONS

Two problems have to be addressed: the elicitation of tacit knowledge by interviewing the experts, possibly avoiding cognitive and relational bias, and the extension of the application of the model to other industries.

CONCLUSION

Drawing on a set of case studies, this paper proposes a Bayesian approach to determine the estimate to complete, both in terms of cost and time, for a complex project. The paper focuses on the knowledge sources necessary to feed the forecasting process and the mitigating measures taken in order to avoid possible estimating biases affecting the process.

The Bayesian approach allows for increasing the forecasting accuracy, mainly in those fields where projects are denoted by a high degree of complexity and uncertainty and in particular in the early phase of the project when knowledge available is scant and rapidly evolving.

Eventually, a Bayesian model has been proposed allowing for integrating a multiple set of knowledge sources. The project team may assume a prior estimate of the final budget overrun, based both on subjective expectations about the current project and data records related to similar past projects, and this prior estimate may be updated based on the actual performance of the current project at Time Now, obtaining a high level of accuracy in forecasting the final cost and duration of the project.

REFERENCES

- Anbari, F. (2003). Earned Value: Project management method and extension. *Project Management Journal*, 34, 12–23.
- Caron, F. (2014). *Managing the continuum: certainty, uncertainty, unpredictability in large engineering projects*. Springer.
- Caron, F., Ruggeri, F., & Borgarucci, C. (2013b). Bayesian integration of internal and external views in forecasting project performance. *Journal of Modern Project Management*, 1(2), 112–121.
- Caron, F., Ruggeri, F., & Merli, A. (2013a). A Bayesian approach to improve estimate at completion in Earned Value Management. *Project Management Journal*, 44(1), 1, 3–16. doi:10.1002/pmj.21303
- Caron, F., Ruggeri, F., & Pierini, B. (n.d.). A Bayesian approach to improving Estimate at Complete. *International Journal of Project Management*.
- Christensen, D. (1996). Project Advocacy and the estimate at completion problem. *Journal of Cost Analysis and Management*, 35–60.
- De Finetti, B. (1937). La prévision: Ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré*, 7(1), 1–68.
- Dvir, D., & Lechler, T. (2004). Plans are nothing, changing plans is everything: The impact of changes on project success. *Research Policy*, 33(1), 1–15. doi:10.1016/j.respol.2003.04.001
- Fleming, Q. W., & Koppelman, J. (2006). *Earned value project management* (3rd ed.). Newton Square, PA: Project Management Institute.
- Flyvbjerg, B. (2006). From Nobel prize to project management: Getting risk right. *Project Management Journal*, 37, 5–15.
- Flyvbjerg, B. (2009). Survival of the un-fittest: Why the worst infrastructure gets built – and what we can do about it. *Oxford Review of Economic Policy*, 25(3), 344–367. doi:10.1093/oxrep/grp024
- Gardoni, P., Reinschmidt, K. F., & Kumar, R. (2007). A probabilistic framework for Bayesian adaptive forecast. *Computer-Aided Civil and Infrastructure Engineering*, 22(3), 182–196. doi:10.1111/j.1467-8667.2007.00478.x
- Goodwin, P. (2005). How to integrate management judgment with statistical forecasts. *Foresight*, 1, 8–12.

- Green, K. C., & Armstrong, J. S. (2007). Structured analogies for forecasting. *International Journal of Forecasting*, 23(3), 365–367. doi:10.1016/j.ijforecast.2007.05.005
- Haji-Kazemi, S., Andersen, B., & Krane, H. P. (2013). A review on possible approaches for detecting early warning signs in projects. *Project Management Journal*, 44(5), 55–69. doi:10.1002/pmj.21360
- Hogarth, R. M., & Makridakis, S. (1981). Forecasting and planning: An evaluation. *Management Science*, 27(2), 115–138. doi:10.1287/mnsc.27.2.115
- Kahneman, D., & Tversky, A. (1979). Intuitive prediction: Biases and corrective procedures. *TIMS Studies in Management Science*, 12, 313–327.
- Kim, B., & Reinschmidt, K. F. (2009). Probabilistic forecasting of project duration using Bayesian inference and the beta distribution. *Journal of Construction Engineering and Management*, 135(3), 178–186. doi:10.1061/(ASCE)0733-9364(2009)135:3(178)
- Kuosa, T. (2012). *The evolution of strategic foresight*. Gower.
- Lipke, W. (2002a, December). A study of the normality of Earned Value Management indicators. *The Measurable News*, 1-16.
- Lipke, W. (2002b). Statistical process control of project performance. *Crosstalk*, 13(March), 16–20.
- Lipke, W. (2003, Summer). Schedule is different. *The Measurable News*, 31-34.
- Lipke, W. (2006a). Earned Schedule Leads to Improved Forecasting. *Proceedings of the 3rd international conference on project management (PROMAC 2006)*.
- Lipke, W. (2006b, Winter). Statistical methods applied to EVM...the next frontier. *The Measurable News*, 18-30.
- Liu, L., & Zhu, K. (2007). Improving cost estimates of construction projects using phased cost factors. *Journal of Construction Engineering and Management*, 133(1).
- Lovall, D., & Kahneman, D. (2003). Delusion of Success: How optimism undermines executives' decisions. *Harvard Business Review*, 81, 56–63. PMID:12858711
- Merrow, E. W. (2011). Oil industry megaprojects: our recent track record. *Offshore Technology Conference*. doi:10.4043/21858-MS
- Palomo, J., Ruggeri, F., Rios Insua, D., Cagno, E., Caron, F., & Mancini, M. (2006). On Bayesian forecasting of procurement delays: A case study. *Applied Stochastic Models in Business and Industry*, 22(2), 181–192. doi:10.1002/asmb.627
- Reich, B. H., Gemino, A., & Sauer, C. (2014). How knowledge management impacts performance in projects: An empirical study. *International Journal of Project Management*, 32(4), 590–602. doi:10.1016/j.ijproman.2013.09.004
- Schindler, M., & Eppler, M. J. (2003). Harvesting project knowledge: A review of project learning methods and success factors. *International Journal of Project Management*, 21(3), 219–228. doi:10.1016/S0263-7863(02)00096-0
- Williams, T., Klakegg, O. J., Walker, D. H. T., Andersen, B., & Magnussen, O. M. (2012). Identifying and Acting on Early Warning Signs in Complex Projects. *Project Management Journal*, 43(2), 37–53. doi:10.1002/pmj.21259

ADDITIONAL READING

Ansoff, H. I. (1975). Managing strategic surprise by response to weak signals. *California Management Review*, 17(2), 21–23. doi:10.2307/41164635

Caron, F. (2016), Project Management. A Bayesian approach, International Society Business Industrial Statistics, ISBIS 2016 Meeting, 8-10 June 2016, Barcelona

Flyvbjerg, B., Holm, M.S. and Buhl, S. (2002), Underestimating costs in public works projects: error or lie? *Journal of the American Planning Association*, 68, 3 (Summer), 279-295.

Kolltveit, B. J., Karlsen, J. T., & Gronhaug, K. (2004). Exploiting opportunities in uncertainty during the early project phase [ASCE]. *Journal of Management Engineering*, (October): 2004.

Marshall, R. A., Ruiz, P., & Bredillet, C. N. (2008). Earned value management insights using inferential statistics. *International Journal of Managing Projects in Business*, 1(2), 288–294. doi:10.1108/17538370810866386

Project Management Institute. (2013). *A Guide to the Project Management Body of Knowledge* (5th ed.). Newtown Square.

Weick, K. E. (1995). *Sense-making in organizations*. Thousand Oaks: SAGE.

Williams, T., Samset, K., & Sunnevag, K. J. (Eds.). (2009). *Making essential choices with scant information*. Palgrave Macmillan. doi:10.1057/9780230236837

KEY TERMS AND DEFINITIONS

Bayes Theorem: Represents a rigorous approach to update a prior distribution, which expresses the experts' preliminary opinion, through the experimental data gathered on the field.

Earned Value Management System (EVMS): Management system aiming at an integrated control of project cost and schedule based on the concept of earned value i.e. budget cost of work performed at Time Now.

Estimate to Complete (ETC): Effort required to complete the project, in terms of cost and time

Monte Carlo Simulation: “What if” analysis of the future project scenarios, provided a mathematical/logical model of the project implemented on a computer.

Project Control: Project management process aiming at identifying and implementing possible corrective actions based on the expected performance of the project.

Subjective Probability: The degree of belief in the occurrence of an event, by a given person at a given time and with a given set of information.

Trend Analysis: Linear extrapolation to the work remaining of the data records related to indices of project performance, e.g. productivity, during the execution of the work completed in order to estimate cost and time to complete the project.

Shaping Mega–Science Projects and Practical Steps for Success

Phil Crosby

Curtin University, Australia

INTRODUCTION

Success and failure in projects is a topic frequently discussed among project management (PM) practitioners. Public funded mega-scale projects especially are scrutinised for performance by funders, users, and the popular press. While a good number of notable mega-projects are delivered within acceptable parameters of time, budget and scope, many large complex projects - especially those underpinned by, or delivering, new technology - too often fail in one or more success dimension (Hartman and Ashrafi, 2004; Ellis, 2008; MoD, 2009; Brouwer, 2011; Flyvbjerg and Budzier, 2014). Perhaps of most concern is that we don't seem to be learning. Large complex projects continue to underperform despite increased availability of systemic, disciplined PM approaches, training, and internet based resources (Archibald, 2003; Flyvbjerg, Bruzelius, & Rothengatter, 2003).

Much has been written regarding project performance, and the literature contains casework and empirical studies of tens, and sometimes hundreds, of projects in an effort to distil factors governing their success or failure (e.g. Pinto and Slevin, 1989; Müller and Turner, 2007; Ika, 2009). Flyvbjerg et al. (2003) identify a 'megaproject performance paradox' that, put simply, means that despite increasing opportunities to learn by experience, project risks remain unacknowledged or unaddressed by stakeholders, and that project performance continues to disappoint.

However, the focus of this chapter is not on causes of failure, but factors underpinning success. The objective is to bring together key findings from the author's research and casework, augmented

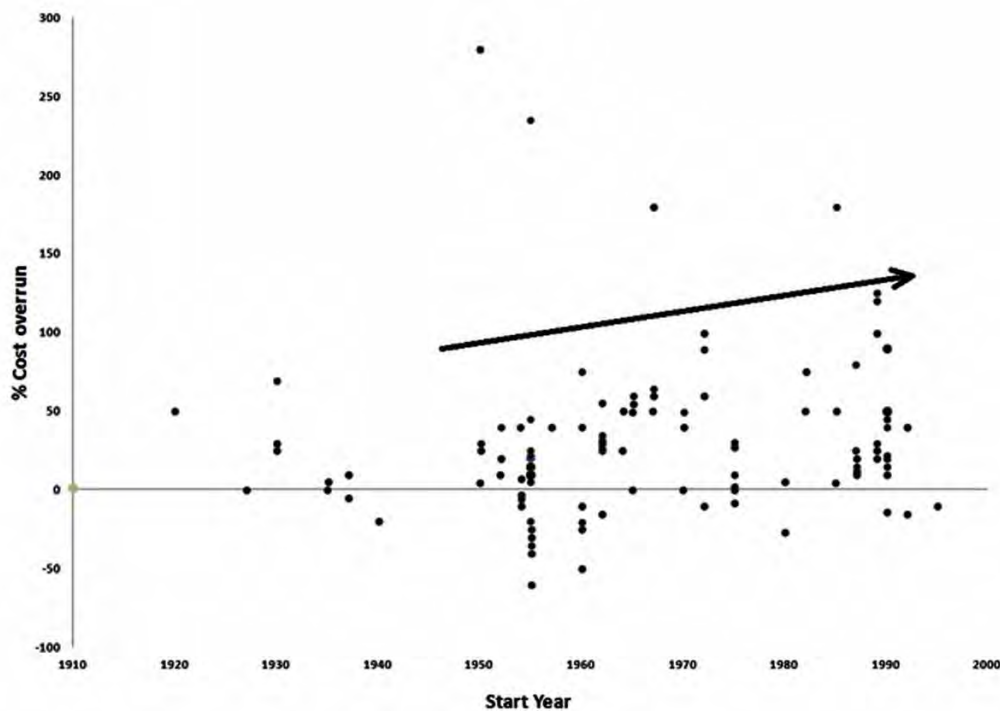
by recent reports and lessons learned, to identify strategic activities and/or actions at the project formation phase that show strong correlation to successful project outcomes. In this chapter, mega-projects are generally defined as having hundreds of millions or even billion dollar budgets, time-frames measured in several years, and often attracting public and/or political attention. Such projects generally involve a significant information technology (IT) or software component, application of leading edge science/engineering technologies, and complexities that test traditional, rational PM methodologies.

BACKGROUND

As high-technology (high-tech) projects have grown in size, cost and risk, so has the challenge in realising success. Between the 1960's and 1980's project success emphasised delivery against the "iron triangle" (time, cost, scope). By year 2000, success criteria had expanded to include client satisfaction and stakeholder benefits. The 21st century has seen the focus broaden to embrace business success and strategic objectives (Ika, 2009).

Systematic project management emerged in the 1950s with the Program Evaluation and Review Technique (PERT), and the Critical Path Method (CPM). These methodologies continued to proliferate through the 1960s and 1970s; later becoming computerised. By 1990, PM was effectively professionalised and managed through hierarchical organisational structures, along with their attending bureaucracies, linear mode planning tools, and standardised forms of project review.

Figure 1. Project cost control performance over last 90 years
Adapted from Flyvbjerg et al., 2003.



The application of skills and techniques to meet the demands of increasing complexity and the parameters by which modern project success is measured, has lagged. Whereas moderately scaled high-tech projects can be managed using traditional PM methods and tools, the reported poor performance of many mega-projects is compelling evidence that lessons are not being learned, and that advanced PM theory and practice is not being applied (Turner, 2004; Cooke-Davies, 2002; Grün, 2004; Shenhar and Dvir, 2007). An example of this is illustrated in data by Flyvbjerg et al., (2003). See Figure 1.

The topic of mega-project management, with its inherent new scale challenges, is receiving attention by researchers, with a growing awareness of the importance of front-end planning. Difficulties with dependence on early stage risk assessment amid uncertainty are examined by Flyvbjerg et al. (2003); Bakker, Cambre, Korlaar, & Raab (2010); and Geraldi, Lee-Kelley, & Kutsch (2010). Project shaping as a management craft

is investigated by Smith and Winter (2010) who show clear links to project success, while project shaping as a competitive advantage is addressed by Miller and Lessard (2000). Blanchard (1990) and Cook-Davies (2002) each discuss the people aspects of new projects and the pivotal role of management, while work by Jani (2010) asserts that self-efficacy enables resilience in complex IT project teams. Crosby's in-depth study (2012a) reveals new attitudinal and conditional factors for shaping of complex projects specifically. Early stage critical success factors (CSF) are noted by Elenbaas (2000) who signals the crucial conditions and complex environments within project start-ups. Much referenced PM authors Shenhar and Dvir (2007) emphasise the need for early tailoring of project success measures and dimensions. Engwall and Westling (2001) explore assumptions around linear project processes and the limiting effects of articulating imperfect knowledge at project start-up. Weston (2007), Fellows and Alexander (2010), and Fisher (2010) each touch

Table 1. List of case study projects and organisations¹

Project Acronym	Location	Description
ALMA	Northern Chile	Radio telescope array of ~66 dishes.
ASKAP	Mid-West of Western Australia	Radio telescope array of 36 dishes, and precursor for the SKA project.
ATCA	Northern NSW, Australia	Radio telescope array of 6 dishes completed in 1988.
HIPER	Site not yet decided.	High power laser-driven fusion experiment.
ILC (DESY) XFEL (DESY)	Site not yet decided. Hamburg, Germany	Dual opposing linear colliders of super high power. X-Ray high power free electron laser
ITER	Provence, France.	Thermonuclear experimental reactor.
LHC	Beneath the French-Swiss border.	Large Hadron Collider – a gigantic particle accelerator.
LIDAR	Davis Station, Antarctica.	Light detection and ranging instrument configured to probe the mesosphere.
LOFAR	Centered in Northern Netherlands	Radio telescope consisting of thousands of omni-directional dipole antennas.
MeerKAT	Northern Cape of South Africa.	Radio telescope array of 7 dishes (to be expanded to ~64), and an SKA precursor.
OPAL	South of Sydney, Australia.	20 Mega-watt open-pool research reactor.
SKA	Split between Africa and Australasia.	Giant radio telescope with up to 1 million square metres of collecting area.
SYNCH (Australian Synchrotron)	Melbourne, Australia.	A particle accelerator accommodating 30 beamlines.
TOPSAT	RAL (UK). TOPSAT is still in earth orbit	A micro-satellite with advanced, down-looking, imaging cameras.
VISTA	Northern Chile	A visible and infra-red survey telescope.

on early stage risk of immature technologies. An extensive statistical study of industrial mega-projects by Merrow (2011) compellingly points to Front End Loading (FEL) as the most critical phase for driving project success.

Complementing the literature sources for this chapter are data and narratives gathered by the author from 16 large scientific facility projects in Europe, Chile and Australia. The projects were mostly selected from contemporary big astro, nuclear, and particle physics machines. Each project is characterised by having specialised infrastructure supporting new technology, >US\$100 million budget, and a challenging IT requirement. The author observed first-hand the planning and execution processes, and formal and ad-hoc meetings, both in the project office and the worksite. At each project, between five and ten formal interviews with project directors and managers

were held, each loosely structured to gather the ‘lived experiences’ of project practitioners. The fieldwork study list is shown in Table 1. Further insights were gained at relevant conferences and workshops including NASA’s annual 2009 and 2010 Project Management Challenge event.

RESEARCH FOCUS AND FINDINGS

A consistent view expressed in both the literature, and in fieldwork narratives, is that project success is not indeterminate by nature. By undertaking certain activities, coupled with setting of particular policies and launch conditions at the front-end; a project can be shaped for robustness and success (Archibald (2003). This shaping phase has a profound effect on project stabilisation, and the challenges will be different for each mega-project.

Table 2. Selected success drivers by relative importance

ID ² Ref	Success Driver	All Project Ranking	High-Tech Ranking
B	Clear project definition, requirements framed, goals, objectives, scope, success criteria, and project mission defined; sound business case established.	23.72	19.53
E	Project baseline, realistic (benchmarked) estimates, project phasing, effective project performance reviews, mission assurance systems, and measurement standards.	7.85	8.96
H	PM/Organisational understanding & competence in complex project management (beyond PMBOK level).	3.17	3.37
J	Aligned perceptions of realistic project goals & success (management and teams); sense of urgency instilled.	2.31	2.37
K	Effective stakeholder partnerships/alliances including processes for funding, contracting, and communications).	2.31	2.37
N	Understanding & continuous management of known and unknown risk exposure; maintenance and visibility of risk register.	0.91	1.13
P	Effective means of learning from analogous experience, and continuous improvement environment	0.66	0.78

Derived from findings in Crosby (2012b).

Merrow (2011) holds that turbulence and disruption are more likely a feature of poor shaping than of project management.

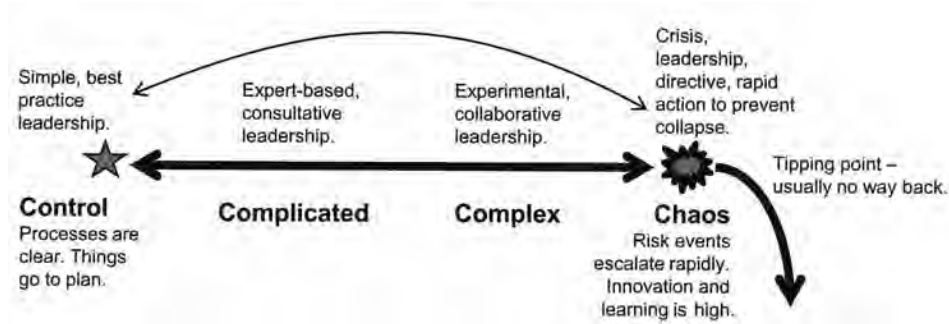
The author's previous meta-study (Crosby, 2012b) draws on mega-project casework and literature material from the previous four decades, to ask: what are the key strategic areas that show strong correlation to project success? The full method and results are not repeated here; suffice to say that this scientific study comprised rigorous casework using Yin's (2009) methodological approach to validating complex phenomena, which was contrasted against findings from 29 studies encompassing 2,820 projects. A sub-set of 20 studies (928 cases) were classed as 'high-tech' (although the author finds success drivers equally applicable in general projects). The Analytical Hierarchy Process (AHP) was then applied to reveal a ranked set of success drivers, followed by consistency tests to check fidelity. Several of the higher scoring factors are directly applicable at the planning, formation (FEL) and project approval stages. Table 1 shows the comparative ranking of selected general, and high-tech, project

success drivers shown to impact strongly on early project shaping.

The ranked project success drivers reveal valuable insights for improved front-end stage PM. Overall, a relatively small number of key project aspects are demonstrated to impact primarily on the chances of success. Most significantly, implementation of competent project control systems and processes, and a clearly defined project mission are shown to be twice as important as the next ranked driver (mature project information systems). These considerations are important throughout the project but it is clearly necessary to test the intent, strength, and understanding of the project objectives at the conceptual/approval stage. Second, the overall rankings show the importance of 'softer' indicators such as stakeholder engagement and expectations management, and instilling the right amount of urgency.

In the following sections, I expand on five topics that show positive correlation between the casework and the meta-data studies when examining the importance of shaping at mega-project front-end stages.

Figure 2. The continuum between control and chaos
Remington, 2011 p. 18.



Managing Complexity

The technical, logistical, and environmental hurdles that face complex mega-projects are daunting, even for seasoned practitioners. Complex systems theory describes degrees of complexity ranging from highly complicated to almost chaotic (see Figure 2). In such systems, it is not intuitively obvious how the impacts on a complex project may affect the outcomes.

System complexity tends to grow in response to ever more demanding science and technology applications, especially in relation to performance and reliability. For large high-tech projects that involve a site requirement (e.g. nuclear plants, particle colliders, etc.) institutional factors add further complexity, as well as stakeholder interests (Helmsman, 2009). The International Complex Project Management Task Force (ICCPM, 2011, p. 18) cites complexity as, “embodying uncertainty, ambiguity, dynamic interfaces and significant political or external influences. Such projects often [run] longer than the lifecycle of the technologies involved.”

The ‘players’ too shape the character and dimensions of the project. While the project office may hold the status of project integrator, the prime contractors too must be capable of complex PM. For cutting-edge programs, the system integration activities alone pose serious challenges in terms of industrial skills and capabilities. Much of this how-to knowledge is not codified, and casework

verifies that particular projects are much more than just complicated and difficult; they have special characteristics that pose extraordinary management challenges not always recognised at the outset by institutional agencies (ITER, 2009; ILC, 2009; ALMA, 2014; ASKAP, 2014; SKA, 2014).

Geraldi et al. (2011) assert that the use of tools for complexity analysis increases understanding of how to manage complexity at project start, e.g. helping to formulate the business case. The well-known Helmsman Complexity Scale (Helmsman, 2009) offers a useful comparator tool empirically derived from research across multiple industry sectors and underpinned by scientific and technical rigour. The scale clearly places global, high-tech mega-projects at the highest levels of complexity.

The International Centre for Complex Project Management (ICCPM) offers a more scientific Project Complexity Assessment (PCA) tool designed to create broad awareness at the early stages of projects of an organisation’s capacity to deliver; thus informing the likely success of the project. Then there are the mathematical models. Through his equations and extrapolations, Sessions (2009) shows an almost linear relationship of complexity to project early failure rates. Thomas and Mog (1997) investigate complexity as a function of the interactions between system components over time. Their approach informs technology readiness and programmatic risk, but is a non-trivial exercise in mathematics.

Table 3. Complexity elements in mega-projects

Technical	Organisational	Environmental
No. of goals	Project duration	No. of stakeholders
Clarity of goals	Size in capital expenditure	Stakeholder perspectives
Size of scope	Size in engineering hrs	Political influence
Uncertainties in scope	Size of project team	Internal support
Quality requirements	Size of site area	Required local content
No. of tasks	No. of locations	Interference
Variety of tasks	Resources, skills availability	Climate and weather
Dependencies between tasks	Experience in parties	Remoteness of site
Uncertainty in methods	Health, safety & environmental awareness	Experience in the country
Interrelations between technical processes	Compatibility between methods & tools	Dependencies on stakeholders
Conflicting Standards	Contract types	Strategic pressure
Innovation level	No. different nations	Project stability
Experience level	No. different languages	Level of competition
Technical risks	Cooperation JV partner	Environmental risks
Technical dependencies	Interfaces between disciplines	
	No. of financial resources	
	Trusted relationships	
	Organisational risks	

Adapted from Bosch-Rekvelde et al., (2010).

More palatable to PM practitioners, Bosch-Rekvelde, Jongkind, Mooi, Bakker, & Verbraeck, (2010) present a readily applied TOE (technical, organisational, environmental) framework to group (though not evaluate) the elements of project complexity, especially at the start-up phase (see Table. 3). Practitioners interviewed readily agreed that just having a shared understanding of the complexity dimensions within the project team aids comprehension and decision-making.

While the number and scale of complexity factors is shown to be an indicator of project execution difficulty, the fieldwork interviewees also mentioned the positive impacts of aspects of team diversity, spreading of risk, and varied sources of experience (ALMA, 2007, ATCA, 2009, XFEL, 2009, LOFAR, 2009). These findings accord well with an Australian study of defence projects (Helmsman, 2009). Although a basic knowledge of complexity theory is required of PMs, access to and implementation of practical tools and guides

to address complex project problems are also needed (Remington and Pollack, 2009)

Success Criteria and Setting the Project Goal(s)

Success criteria, comprising critical success factors (CSF), should not be confused with success *metrics* which can be set later. Success factors may be quantitative or qualitative, and expressed as goal statements, core and non-core benefits, and/or longer term objectives (SKA, 2014). Several casework projects placed user-community satisfaction far ahead of meeting budgetary or time constraints (ITER, 2009; LIDAR, 2009; XFEL, 2009; ASKAP, 2011).

Tolerance in meeting project success criteria will vary between projects. Stakeholders too, have different views of success, and diverse positions in terms of interest, influence, attitude, and requirements (British Computer Society, 2006).

Each project's set of performance measures is thus unique, and cannot be considered in a generic sense, even when a new project is modeled on a proven design (SYNCH, 2009).

In seeking to formulate a new approach to successful project execution, Shenhar and Dvir (2007) argue for a holistic technique when setting success criteria. Their model involves at least five fundamental dimensions (or 'soft metrics'):

- Project efficiency,
- **Impact on the Customer:** Meeting requirements and achieving customer satisfaction, benefits, and loyalty,
- **Impact on the Team:** Satisfaction, retention, and personal growth,
- **Business Results:** Return on investment, market share, and growth,
- **Preparation for the Future:** New technologies, new markets, and new capabilities.

The casework found that a considerable benefit of early declaration of success criteria is their use in front-end phases where project re-shaping, or even termination, is cheapest (Cooper, 2006).

Procurement (Contracting) Processes

Whatever the final cost reality, all complex engineering mega-projects commence life with a forecast budget, often into the billions of dollars. Such investment levels create expectations of financial competence including professional management of procurement. Despite this, the acquisition of goods and services for projects has traditionally been seen as largely administrative or operational rather than having strategic importance (Jaakkola, 2004; Crosby, 2011). However aspects other than purely transactional show a maturing of approach to procurement echoed in several of

Table 4. Nine procurement success topics and their associated Key Success Actions

Topic	Key Success Action
The procurement office/function	No later than the pre-contractual phase, establish the procurement office structure, resources, roles and responsibilities, and information management systems.
Procurement strategy, policy and planning	Formally approve procurement principles /strategies, policies, and plans, and document these within a concise approved and controlled Project Procurement Plan.
Informed contracting with industry	Obtain full understanding of global capability 'scouting' information, and employ appropriate approaches and procurement instruments with terms and conditions supporting project success measures.
Contracting models	Establish the permitted contracting model(s), aligned with the legal entity. Develop relationships with principal contractors and key components of the supply chain.
Ensuring competitiveness	Ensure competitiveness in contracting, through carefully planned pricing strategies, and a broad 'value for money' approach.
Procurement specifications	Apply purchasing specifications that are appropriate to the goods or services required, flexible for optimal outcomes, and developed with input from industry.
Procurement risk	Understand and actively manage industrial relationships and procurement risk, including the critical early stage engagement phases. Site readiness plans should include industry capacity and capability confirmations.
Tender evaluations and contractor selection	Evaluate and select project contractors using a fair and defensible process, and focussed on criteria weighted in favour of supplier reputation and investment, and mission success parameters.
The purchaser-supplier relationship	Implement fair and transparent procurement processes, and where obviously desirable, exploit ongoing strategic partnerships. Allow for evidence-based supplier compliance and quality inspections.

Crosby, 2011.

the case studies (ALMA, 2008; LOFAR, 2009; LHC, 2009; OPAL, 2009).

Rendleman and Faulconer (2011) demonstrate the importance of closing the gaps between available technologies and customer needs before beginning an acquisition so as to shape programs for success. Schill (1979) points out the strategic liaison role played by procurement to ensure that suppliers remain competitive and at the forefront of technology, as well as realise better cost and product performance. Moreover, he assigns strategic responsibility for materials and technology identification – a key tactic identified in the procurement scouting activity developed for the SKA telescope project (Crosby, 2010).

Investigations into effective mega-science project procurement (Crosby, 2011) reveal nine specific topics of attention. These are listed with associated Key Success Actions in Table 4.

A common theme among the Key Success Action is relationships, with the strong implication that traditional open tender procurement methods (EoI, RFTs, RFQ, etc.) are not always the best basis for effective acquisition in complex mega-projects. While not proposing a wholesale shift from ‘quote-supply’ processes, the often taxing experiences from the case studies (ALMA, 2009; ITER, 2009; LHC, 2009; ASKAP, 2012;) uphold the potential overall advantages of alliance-style contracting in the form of cooperative relationships.

A strategic partnership is much more than a motherhood catchphrase, and is fundamentally about trust; based on competence, interdependence, and integrity. This means cooperative and transparent relationships with shared risk/reward, and in their highest form, without competitive bidding in the traditional sense. Although more cautious, Merrow’s (2011) deep study of industrial megaprojects also supports alliances as viable industry participation models, while emphasising that each mega-project must decide its own most suitable procurement approach. Merrow’s findings concur with the author’s casework in that around half of all contracting was of the institutionally mandated ‘lump-sum’ type, even though such

arrangements tended to be less cost-effective and rarely reflected the final price paid.

Resilience-Building

Continued high-tech project failure has directed renewed investigations into early factors that help drive project resilience and robustness. Crosby (2012a) for example, identifies nine key early prerequisites for complex mega-project resilience. These are grouped into two types:

- **Attitudinal:** (Realism, fact based, a lessons-learned culture, ambiguity/uncertainty tolerant); and
- **Conditional:** (A defined mission, clear reporting and decision structures, strong information control, preparedness for unknown risk, external environmental awareness, and mission assurance).

Of these, three factors considered central to the project shaping process:

1. Project Realism

Large complex projects are naturally instigated in the enthusiastic belief by the protagonist(s) that they can be executed on time and on budget, and achieve their promised outcomes. Too often this optimism bias leads to underestimation of the difficulties (Grün, 2004). A UK Government Guide (UK Treasury, 2003 p. 86) specifically links the cause of over-optimism in project costs to, “poor definition of the scope and objectives of projects... due to poor identification of stakeholder requirements, resulting in the omission of costs during project costing”.

Evidence of the optimism ‘curse’ is not hard to find. A radio astronomy White Paper (Fisher, 2010) describes overly optimistic cost and time estimates as the most frequent cause of project de-scoping. A UK defence report (MoD, 2009 p. 5) lamenting a £205 million cost increase over 20 projects admitted that, “on far too many proj-

ects, the Department is over-optimistic and sets unachievable cost, time and performance objectives”. Butts and Linton (2009) who investigated NASA’s cost estimation performance, conclude that undershooting cost and schedule projections is common practice, borne out by NASA Director, George Morrow’s warning (NASA, 2009 p. 31) that, “a major pitfall is being overly optimistic early in the project lifecycle”. Over-optimistic budget and schedule expectations inhibit project success, and are especially likely on projects with institutional difficulties. IT projects also appear to be especially vulnerable (Verner and Cerpa, 2005; Mieritz, 2012); Flyvbjerg & Budzier, 2014).

Optimism phenomena were observed in the casework, most obvious within ASKAP, ITER, and the SKA. In each instance both schedule and cost were (and continue to be) stretched as the challenge of settling the design and transitioning through construction phases tests each deadline, prompts revised baselines, and raises fresh funding questions. Concerns over realism in estimates were rarely conveyed above PM level in organisations.

How then might we recalibrate our thinking? Blanchard (1990) claims objectivity may be the only defence and suggests that, as a balancing process, project proponents should also develop a case for *not* proceeding. Flyvbjerg et al. (2003) argue strongly for bona-fide cost estimates accompanied by risk analyses of future cost regimes. Flyvbjerg (2013) strongly promotes taking an ‘outside view’ through the Reference Class Forecasting technique. Further advice for dealing with over-optimism in complex projects is contained in a brief, but pithy, guide from the UK National Audit Office (NAO, 2013).

2. Lessons Learned

The benefit of advice from like-projects seems axiomatic, yet research for this chapter found little evidence of planned effort for this. None of the case study projects budgeted for even perfunctory enquiries of similar projects. Learning from one project to the next is strongly advocated in the

practice literature, though academic studies find that few organisations manage this systematically to any depth, or in a useful format (Verner and Evancho, 2005; Atkinson, Crawford, and Ward, 2006). Smith and Winter (2010) directly link competent front-end management with project success while promoting the idea of deliberately exploring insights and implications flowing from past projects.

When pressed by the author, interviewees agreed that simply assuming that lessons from previous projects have tacitly informed future projects is insufficient. Organisations must take steps to accumulate and disseminate the learning before it dissipates. In other words, knowledge gained ‘ad-hoc’ will not naturally lead to retention of project insights; a shortcoming dubbed as ‘project amnesia’.

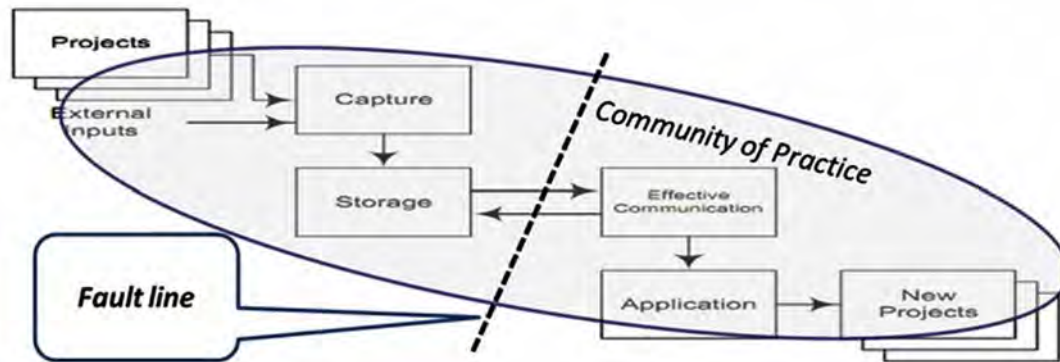
Case research found that the majority of organisations that do undertake a lessons-learned activity do so through meetings or workshops - activities that Williams (2008) shows to be well correlated with increased perceptions of success and markedly increased PM competency. Examples in successful project organisations show that the ‘fault-line’ of failed knowledge transference can be effectively bridged by ‘communities of practice’ as stimulants to successful cross-project learning (Fig. 3)

Innovative organisations are more likely to use corporate training, searchable databases, and transfer lessons-learned into organisational procedures. Often missing (yet easily implemented) is a policy requirement to revisit lessons from previous projects as a prelude to new project planning.

3. Mission-Assurance

One phenomena of complex project execution that surfaced sporadically among the case studies (ALMA, 2009; SYNCH, 2009) yet is more overt within industry and aerospace organisations was that of mission-assurance. In essence, it translates to an unwavering sense of purpose in making mission success the highest priority at all

Figure 3. Showing how communities of practice bridge the project knowledge fault-line



levels of the project. The adoption of a mission-assurance mindset is revealed most compellingly within NASA where the post-2001 integration of a formal Mission and Safety Assurance function correlated to a lift in mission success rate from 90% to almost 95% (Crosby, 2015).

The author's research supports the appointment a dedicated Mission Assurance person or group at the project concept stage. This function (part auditor, part advisor, part 'devil's advocate') should be positioned outside of mainstream project delivery, yet close enough to have ready participation in critical testing, meetings, and reviews, and with access to top level project management. This crucial role is not about meddling or being an undercover agent, but principally one of questioning and checking that activities, deviations and changes pose no unrecoverable threat to project execution and performance.

Beyond Traditional Risk Management

Conventional (PMBOK style) risk matrices used to self-score risk intensity by rating likelihood and consequences are common PM artefacts and were observed in all the case studies. Such tools appear to at least identify a view of known risk exposures at project start. However review and/or update of resulting risk registers as the project unfolds, or the systematic application of mitigating controls, was rare. But perhaps this lack of appli-

cation matters little, since a growing number of researchers consider that such arbitrarily chosen risk ratings are mostly worthless, due largely to the effects of cognitive bias and human variance. In their study of IT project success, de Bakker, Boonstra, & Wortmann, (2010 p. 499) conclude, "risks for which there is no classical or statistical probability distribution cannot be managed by means of the [rational] risk management process".

The inability or reluctance to plan for non-rational, unexpected massive negative (or even positive) impacts is investigated deeply by Taleb (2010) who classifies such events as 'Black Swans'. Taleb forcefully argues that the tendency for humans to scale and smooth predictions on assumed or unfounded (i.e. guessed) data guarantees erroneous forecasts. The author's observations within the high-tech mega-project environment confirm these tendencies, with highly rational scientist and engineers lacking skill when dealing with complex programmatic and social matters. Green (2012) adds the crucial point that project planners should not be fooled by the statistically insignificant frequency of Black Swans, but should instead pay close attention to the potential catastrophic consequences.

Geraldi et al. (2010) support observed data that unexpected events will inevitably emerge in large complex projects; likely surfacing from the external project environment, e.g. political turmoil, supply chain failure, exchange rate fluctuations, or even weather events. Miller and Lessard (2000)

report an average of five unexpected events in the formative stages of projects, with some confronting as many as 12. Two of the case studies (ALMA telescope, Australia Synchrotron) experienced such misfortunes.

How might complex projects be better shaped in terms of risk? First, it is prudent to invest in preparedness not prediction. Green (2012) promotes ‘uncertainty spotting’ skills sensitive to the early weak signals of threats and assumptions through active awareness – especially geopolitical and economic developments. In their study of tools for complex projects, Remington and Pollack (2008) discuss non-linear risk events requiring early decision-making so that even if managers cannot directly predict the source or possible ramifications, an ‘atmosphere of preparedness develops’.

FUTURE RESEARCH DIRECTIONS

Future research could usefully further investigate reasons for project success. Whereas there are clearly many single-point causes for failure, project triumph can only occur when all barriers are surmounted, and the project not only meets its programmatic goals, but also satisfies its customers and critics. It is project success, not failure, which offers data richness.

The author’s research suggests that the world of military contracting holds further learning in the area of mega-science project management, especially programs that integrate, and continually upgrade technologies. Future research in this genre should derive useful lessons regarding through-life project support, and joint agency execution of high-reliability programs.

Lastly, while a large amount of ‘know-how’ around PM is codified in guidebooks and articles, theory development and discussion tends to reside in academic circles, such as in Operations Research (e.g. Gerald, Maylor, & Williams, 2011). This means that epistemic insight into mega-project delivery often remains unseen by

working project managers. Inclusion of this chapter in the Encyclopedia of Information Science and Technology represents a proactive move by the publisher (IGI Global) to bring new and useful wisdom to project practitioners– something that can only be encouraged.

CONCLUSION

This chapter described the potential for specific, and often less obvious, factors that influence success in large and complex engineering/science projects (though probably equally effective in general mega-projects). Five topic areas were developed as being especially important at the front-end stages, leading to the following conclusions:

- Traditional PM skills and techniques are insufficient to properly plan or manage the complex dimensions of high-technology mega-projects. However wisdom exists to master and shape complex projects if we seek it.
- A detailed goal set (or mission) must be clearly articulated, including hard and soft definitions of success.
- Large high-tech projects are more than just complicated and difficult. An assessment of project complexity is essential in order to effectively plan for resources, cost, and risk.
- Procurement is strategically important to success, and not merely an administration function.
- Building project resilience requires curbing optimism, formally researching lessons-learned and implementing a mission-assurance function early.
- Project meta-risk and preparedness for the unknown is required. This includes identifying boundary threats at project outset, and shielding from the consequences of ‘Black Swans’.

- New findings from this research strongly argue that it is not enough just to attend to each of the factors described above, but rather they all need to be tackled together for synergistic results.

REFERENCES

- Archibald, R. D. (2003). *Managing high-technology programs and projects*. John Wiley & Sons Inc.
- Atkinson, R., Crawford, L., & Ward, S. (2006). Fundamental uncertainties in projects and the scope of project management. *International Journal of Project Management*, 24(8), 687–988. doi:10.1016/j.ijproman.2006.09.011
- Bakker, R. M., Cambre, B., Korlaar, L., & Raab, J. (2010). Managing the project learning paradox: A set-theoretic approach toward project knowledge transfer. *International Journal of Project Management*, 29(5), 494–503. doi:10.1016/j.ijproman.2010.06.002
- Blanchard, F. L. (1990). *Engineering project management*. New York: Marcel Dekker.
- Booz Allen Hamilton. (2012). *Improving Complex Facility Construction Projects by Using an Owner's Paladin*. Author.
- Bosch-Rekveltdt, M., Jongkind, Y., Mooi, H., Bakker, H., & Verbraeck, A. (2010). Grasping project complexity in large engineering projects: The TOE (Technical, Organizational and Environmental) framework. *International Journal of Project Management*, 29(6), 728–739. doi:10.1016/j.ijproman.2010.07.008
- British Computer Society. (2006). *Case study of successful complex IT projects*. Lancaster University Management School.
- Brouwer, G. E. (2011). *Own motion investigation into ICT-enabled projects*. Victorian Ombudsman.
- Butts, G., & Linton, K. (2009, April). *Joint confidence level paradox – a history of denial*. Paper presented at the NASA Cost Estimating Symposium, Galveston, TX.
- Cooke-Davies, T. J. (2002). The real success factors on projects. *International Journal of Project Management*, 20(3), 185–190. doi:10.1016/S0263-7863(01)00067-9
- Cooper, R. G. (2006). Managing technology development projects. *Research Technology Management*, 49(6), 23–31.
- Crosby, P. (2010). *Capability assessment model – ver 0.9*. (Unpublished report). SKA Program Development Office, Manchester University, UK.
- Crosby, P. (2011). *Procurement strategies enabling success in high-technology mega-projects: Preparatory work for the SKA* (Unpublished report). SKA Program Development Office, Manchester University, UK.
- Crosby, P. (2012a). Building resilience in large high-technology projects: Front end conditioning for success. *International Journal of Information Technology Project Management*, 4(3), 21–40. doi:10.4018/jitpm.2012100102
- Crosby, P. (2012b). Key success drivers: Meta-study findings applicable to large high-technology projects. *International Journal of Information Technology Project Management*, 2(3), 1–20. doi:10.4018/jitpm.2012040101
- Crosby, P. (2015). *NASA space missions: the effect of 'faster, better, cheaper' and 'mission assurance' on launch success rates*. Manuscript in preparation.
- de Bakker, K., Boonstra, A., & Wortmann, H. (2010). Does risk management contribute to IT project success? A meta-analysis of empirical evidence. *International Journal of Project Management*, 28(5), 493–503. doi:10.1016/j.ijproman.2009.07.002

- Elenbaas, B. (2000). Staging a project – are you setting your project up for success? In *Proceedings of the Project Management Institute Annual Seminars & Symposium*. Houston, TX: PMI.
- Ellis, K. (2008). *Business analysis benchmark: the impact of business requirements on the success of technology projects*. IAG Consulting.
- Engwall, M., & Westling, G. (2001). *The peripety of projects* (Unpublished Working Paper 21). Fenix Research Program & Stockholm School of Economics, Sweden.
- Fellows, J. D., & Alexander, J. K. (Eds.). (2010). *National Research Council workshop report: Decadal science strategy surveys*. Retrieved October 1, 2010, from <http://www.nap.edu/catalog/11894.html>
- Fisher, J. R. (2010). *Astro2010 technology development white paper: large instrument development for radio astronomy*. Observatory Technical Council, National Radio Astronomy Observatory.
- Flyvbjerg, B. (2013). Quality control and due diligence in PM – getting decisions right by taking the outside view. *International Journal of Project Management*, 31(5), 760–774. doi:10.1016/j.ijproman.2012.10.007
- Flyvbjerg, B., Bruzelius, N., & Rothengatter, W. (2003). *Megaprojects and Risk*. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9781107050891
- Flyvbjerg, B., & Budzier, A. (2014). *Why your IT project may be riskier than you think*. Retrieved February 9, 2014, from <http://hbr.org/2011/09/why-your-it-project-may-be-riskier-than-you-think/>
- Geraldi, J., Maylor, H., & Williams, T. (2011). Now, let's make it really complex (complicated): A systematic review of the complexities of projects. *International Journal of Operations & Production Management*, 31(9), 966–990. doi:10.1108/01443571111165848
- Geraldi, J. G., Lee-Kelley, L., & Kutsch, E. (2010). The Titanic sunk, so what? Project manager response to unexpected events. *International Journal of Project Management*, 28(6), 547–558. doi:10.1016/j.ijproman.2009.10.008
- Green, N. (2012). *Keys to success in managing a black swan event*. Aon Risk Solutions White Paper. Retrieved January 17, 2014, from http://www.aon.com/risk-services/thought-leadership/report_black_swan_event_whitepaper.jsp
- Grün, O. (2004). *Taming giant projects (management of multi-organisation enterprises)*. Berlin, Germany: Springer-Verlag. doi:10.1007/978-3-540-24818-7
- Hartman, F., & Ashrafi, R. (2004). Development of the SMART project planning framework. *International Journal of Project Management*, 22(6), 499–510. doi:10.1016/j.ijproman.2003.12.003
- Helmsman. (2009). *A comparison of project complexity between defence and other sectors*. The Helmsman Institute. Retrieved September 28, 2011, from <http://helmsman-institute.com/sites/institute/PDF/HelmsmanDefenceProject-ComplexityReport.pdf>
- ICCPM. (2011). *Complex project management: Global Perspectives and the strategic agenda to 2025* (Unpublished Task Force Report). Canberra, Australia: International Centre for Complex Project Management.
- Ika, L. A. (2009). Project success as a topic in project management journals. *Project Management Journal*, 40(4), 6–19. doi:10.1002/pmj.20137
- Jaakkola, K. (2004). Successful project management - Art, science and culture: A way to successful and strategic contract management. *Proceedings of Nordnet - International Project Management Conference*.

- Jani, A. (2010). Escalation of commitment in troubled IT projects: Influence of project risk factors and self-efficacy on the perception of risk and the commitment to a failing project. *International Journal of Project Management*, 29(7), 934–945. doi:10.1016/j.ijproman.2010.08.004
- Marrow, E. W. (2011). *Industrial megaprojects: concepts, strategies, and practices for success*. John Wiley & Son.
- Mieritz, L. (2012). *Why IT projects fail – a Gartner Survey*. Retrieved January 15, 2015, from <http://thisiswhatgoodlookslike.com/2012/06/10/gartner-survey-shows-why-projects-fail/>
- Miller, R., & Lessard, D. (2000). *The strategic management of large engineering projects*. Massachusetts Institute of Technology.
- MoD. (2009). Major projects report 2008. London, UK: House of Commons Public Accounts Committee.
- Müller, R., & Turner, R. (2007). The influence of project managers on project success criteria and project success by type of project. *European Management Journal*, 25(4), 289–309. doi:10.1016/j.emj.2007.06.003
- NAO. (2013). *Over-optimism in government projects*. National Audit Office.
- NASA. (2009, Fall). *ASK Magazine: The NASA source for project management and engineering excellence*, (36).
- Pinto, J. K., & Slevin, D. P. (1989). Critical success factors in R&D projects. *Research Technology Management*, 32(31), 31–35.
- Remington, K. (2011). *Leading complex projects*. Gower Publishing Ltd.
- Remington, K., & Pollack, J. (2008). Project management to Run: are there special tools for complex projects? *Proceedings of 22nd IPMA World Congress*.
- Remington, K., & Pollack, J. (2009). *Tools for complex projects*. Gower Publishing.
- Rendleman, J. D., & Faulconer, J. W. (2011). Escaping the space acquisition death spiral. *High Frontier Journal*, 7(4).
- Schill, R. L. (1979). Key issues in procurement strategy in advanced technology organisations. *International Journal of Physical Distribution & Logistics Management*, 9(6), 297–309. doi:10.1108/eb014452
- Sessions, R. (2009). *The IT complexity crisis: danger and opportunity*. ObjectWatch. Retrieved April 26, 2010 from http://www.objectwatch.com/white_papers.htm#ITComplexity
- Shenhar, A., & Dvir, D. (2007). *Reinventing project management: the diamond approach to successful growth and innovation*. Harvard Business School Press.
- Smith, C., & Winter, W. (2010). The craft of project shaping. *International Journal of Managing Projects in Business*, 3(1), 46–60. doi:10.1108/17538371011014026
- Taleb, N. N. (2010). *The black swan: the impact of the highly improbable* (2nd ed.). Random House.
- Thomas, L. D., & Mog, R. A. (1997). A quantitative metric of system development complexity. *Proceedings of INCOSE International Symposium*. doi:10.1002/j.2334-5837.1997.tb02243.x
- UK Treasury. (2003). *Supplementary Green Book Guidance: Optimism Bias*. London, UK: UK Treasury.
- Turner, J. R. (2004). Five conditions for project success. *International Journal of Project Management*, 22(5), 349–350. doi:10.1016/j.ijproman.2004.02.001
- Verner, J., & Evancho, W. M. (2005). In-house software development: What project management practices lead to success. *IEEE Software Magazine*.

Verner, J. M., & Cerpa, N. (2005). Australian software development: what software project management practices lead to success? *Proceedings of the 2005 Australian Software Engineering Conference (ASWEC'05)*. doi:10.1109/ASWEC.2005.14

Weston, J. (Ed.). (2007). *The conspiracy of optimism* (Unpublished report from RUSI Acquisition Focus Group). Royal United Service Institute.

Williams, T. (2008). How do organisations learn lessons from projects – and do they? *IEEE Transactions on Engineering Management*, 55(2), 248–266. doi:10.1109/TEM.2007.912920

Yin, R. K. (2009). *Case study research: design and methods* (4th ed.). SAGE Publications.

ADDITIONAL READING

Crosby, P. (2013). *Success in Large High Technology Projects: What Really Works*. Australia: ICCPM eBook Publications.

Dalcher, D. (Ed.). (2014). *Advances in project management: narrated journeys in uncharted territory*. UK: Gower.

Davies, J. (2014). Contracting for success in complex projects. In *Proceedings of the IACCM Australasia Forum*. Melbourne, Australia.

Standish Group International. (2014). *Big Bang Boom*. Research Report, Retrieved January 3, 2016 from https://www.standishgroup.com/sample_research_files/BigBangBoom.pdf

Wingate, L. M. (2015). *Project management for research and development: guiding innovation for positive R&D outcomes*. USA: CRC Press.

KEY TERMS AND DEFINITIONS

Black Swan: An event that comes as a surprise, has a major effect, and is often inappropriately rationalised with the benefit of hindsight.

Books of Knowledge: Standards for project management most commonly issued by professional Institutes. They contain the basic processes that project management experts agree are necessary for most projects in most environments.

Communities of Practice: A group of people who share a craft and/or a profession and created specifically with the goal of gaining and sharing knowledge related to their field.

Optimism Bias: Manifests itself in projects as an underestimation of the time, costs and risks to delivery and the overestimation of the benefits.

Procurement: Procurement is the acquisition of goods, services or works from an external source. It should include activities ranging from market scouting through to post-contract reviews.

Project Lifecycle: Refers to a series of activities which are necessary to fulfill project goals or objective, including; planning and concept design, commencement, execution and review, final review and learning.

Resilience: The ability of a project to readily resume from unexpected events, threats or actions.

ENDNOTES

¹ Where the case study facilities are referenced in this chapter, they appear as their acronym, and visit date (e.g. ALMA, 2009).

² Identifiers relate to the full list of success drivers in Crosby (2012b).

Staying Ahead in Business Through Innovation

M

N. Raghavendra Rao

FINAIT Consultancy Services, India

INTRODUCTION

Many business enterprises rush into generating ideas to become more innovative before clearly identifying the problem in their enterprise. If the enterprises expect their employees to come up with ideas that would work, the employees need to know the business problems faced by their organization. The employees must have access to the information pertaining to the problems of their organization. It is the practice of many enterprises that their organization's information is considered as important and is also kept as secret. It is meant for the management's view only. The environment in enterprises must be friendly instead of closed. Then only the employees will associate themselves with their organization. This would facilitate them to integrate creativity and strategic thinking. Creativity is certainly a part of innovation but innovation is much more (Tocci, Chesbrough, & Piller, 2016).

It is observed executives in enterprises are under the impression that innovation is associated with new products, new technology, and research development activity. This view is too narrow in the presence of business scenario. Innovation can be applied broadly across all aspects of an enterprise. They are such as existing products, services, business processes and business models.

Peter Drucker in his article on the discipline of innovation says that innovation is real work. It should be managed like any other corporate function. Further he explains that it does not mean the same as other business activities. He stresses that innovation is the work of knowing rather than doing. He argues innovative business

ideas come from methodically analyzing seven areas of opportunity. The areas are 1) Unexpected occurrences 2) In Congruities 3) Process Needs 4) Industry and Market Changes 5) Demographic Changes 6) Changes in Perception and 7) New knowledge. He groups the above areas of opportunity are under within a company or industry and outside a company in its social and intellectual environment. The first four areas fall under the group one and the balance three fall under group two. The glass is half full and the glass is half empty are the descriptions of some phenomena, but have vastly different meanings. Changing a manager's perception of a glass is from half full to half empty opens up big innovative opportunities.

BACKGROUND

Globalization has created new trends such as market consolidation, Vertical Market strategies and mergers in the business world. These trends have forced many business enterprises to adapt themselves to these new phenomena (Estrim, 2009). Now it has become a necessity for many business enterprises to look for "New Approach" which can help them to face new realities in business. The most talked about and discussed about the "New Approach" in the business world is "Innovation Management.

History of Innovation

Innovation has always been part of mankind. Since the discovery of fire by rubbing two stones together, human beings have been innovating.

DOI: 10.4018/978-1-5225-2255-3.ch495

Innovation is probably the oldest known process. In other words innovation is an extension of a person's creativity. Human beings have been using their inner skills to create many new things to help mankind.

Historical Background of Knowledge Management

The concept of knowledge management is not new. The focus and approach has been changing over a period of time. One may observe from the literature on the history of information and knowledge that the basic patterns of behavior remained constant over a period of centuries concerning the role of information and knowledge. History provides many examples. Emperors in the olden days always surrounded themselves with advisors who were scholars first and politicians second. Roman emperors like the ancient Greeks consulted educated priests to gain an insight into the possible future. Indian Kings seemed to be concerned with creation of knowledge among people by allocating places for schools and libraries. Julius Caesar used innovative construction methods and advancements in armaments to achieve military success. Napoleon made use of the advancements in artillery to defeat his opponents. Cornelius Vanderbilt had taken advantage of the advancements in rail road wheels and brakes to create vast commercial empire. J.P.Morgan made a huge fortune and revolutionized the financial system by developing modern investment banking practices. Tipu Sultan used war rocket in some of the battles he fought against the British soldiers. This earned universal fame as "Innovator of the World's first war Rocket". It was his victories against the British forces that ranked Tipu Sultan among the few Indian rulers who have defeated the British.

In the pre industrial era agriculture was the basis of nation's economy then the concentration was to learn more about farming. In the post industrial era manufacturing became the basis of nation's economy then the concentration was to

learn more about manufacturing techniques. In the present globalization scenario the focus is towards adapting innovation and process speed in business enterprises. Knowledge base in a business enterprise is an extension of knowledge sharing systems. Business success lies in converting the information into knowledge. In today's business scenario the use of sophisticated knowledge base depends on the skill with which executives in business enterprises arrive at their findings from their analysis for framing strategy for their organization. Further business executives can draw insights from the large information stored in their systems. Conclusions drawn from these analyses will help the business executives to involve in innovation.

Indian and Chinese Thinking

Jugard is a colloquial Hindi word that roughly translates as "Innovative Fix" (Radjou, Prabhu & Ahuja, 2012). It is an improvised solution born from ingenuity and cleverness. Jugard is quite simple. It is a unique way of thinking and acting in response to challenges. It is the gutsy art of spotting opportunities in the most adverse circumstances. Further it is making use of available resources fully improvising solutions by simple means.

Juggard is practiced by almost all Indians in their daily lives to make the most of what they have. Juggard applications include finding new uses for everyday objects such as empty soft drink or pickle bottles reused as containers for water, spices or lentils. When a truck cannot be repaired, diesel engine is removed and used as a cart. Juggard happens in an emergency situation. It is not a planned activity. Chinese believe in today's world the primary resource is knowledge and innovation. This resource comes from one's brain. If one's brain retires, then there is no more life (Sunted, 2011). The level of innovative thought will be one of the major economic drivers for China. Many global organizations have already established offices in China to capture both the creative kinds of its people and the tremendous market share.

TRENDS IN MANAGEMENT CONCEPTS

One can find from the analysis of trends in management in every decade. New concepts have been emerging. The moment it emerges, it is considered as “Next Big Thing” or “Next Important Concept”. In 1950’s “Brain Storming” Concept has become popular. It was considered to be a very important factor in involving the employees in business enterprises to generate new ideas. In 1960’s and 1970’s, sensitivity groups have played an important role in business enterprises. The group leaders of this group have been considered to be “Gurus”. They have been to lead group managers. Then it has been practice, “Gurus” would make brief opening remarks and they would wait for response from the group managers. Then the “Gurus” would facilitate them to express and try to clarify. It was a hope that “Inner Manager” would emerge. It was expected that more research would follow on their deliberations. The era 80’s was considered as the age of quality management. During this period quality control, quality circles, six sigma and other related approaches were followed for enhancing workers performance. The quality related approaches have been continuing for a long time. In the recent past, the need for knowledge management has been felt very much to globalization. In the present era the concept of “Innovation” is becoming ultimate goal of enterprises to face the complex global economy (Vint, 2009).

Invention, Innovation, and Creativity

The terms invention and innovation are used interchangeably. Actually these are separate and distinct concepts. Invention means designing and creating something which has never been made before. The classic examples are invention of light bulb by Thomas Edison, and Invention of telephones by Alexander Graham Bell. Louis Pasteur came up with vaccines. Innovation deals

with bringing in new methods and ideas resulting in required changes in an enterprise. In the present business context, innovation is taking interesting ideas and transforming them into usable solutions (Vangundy, 2008).

Creativity can be defined as discovery of a new idea or connection. Many ideas already exist somewhere in the World. Creativity is to identify these ideas and connect them in a new way. But innovation is much more than just identifying and connecting these creative ideas. It is about strategy and action bringing values to the enterprise through the implementation of these creative and strategic ideas. Innovation can therefore be interpreted as the profitable implementation of strategic creativity.

Example for Invention and Innovation

John Shepherd Barron reached his bank a little late and thought that why there could not be a system to get cash on the lines of chocolate vending machine. This thought gave him an idea to invent the first automatic cash dispensing machine now known as ATM. John had also thought of the four digits PIN Number. Creation of ATM can be considered as a classic example for invention and innovation. Many more features are being added to ATM by other people. Invention of ATM has given a scope for adding additional features in ATM. The enhancement of additional features can be considered as innovation.

Types of Innovation

Innovation can be applied broadly across all aspects of an enterprise such as existing product, services, processes and business models (Prahalad & Masshelkar, 2011). Innovation can be classified under three categories. They are 1) Efficiency Innovation 2) Evolutionary Innovation and 3) Revolutionary Innovation.

Efficiency Innovation

Efficiency Innovation talks about the changes in the existing processes in enterprises. The management is generally happy for small changes made in the existing processes and procedures in their organization. This approach is usually to cut costs, reduce cycle time or designing a simple format of a report to the top management of an enterprise.

Evolutionary Innovation

Evolutionary Innovation concentrates on ideas that represent something distinctly new and better. An example under this category is the introduction of internet banking. This has changed banking business from the fixed hours to banking at any hour. The other example under this type is advertisement in tabloids. Advertisements make the publishers of tabloids possible to distribute news free of charge with the revenue from the advertisements. The business models such as B2C, B2B, and B2G have been providing a scope to innovate improvements and expansion in enterprises.

Revolutionary Innovation

Revolutionary Innovation focuses on something radically new. This may change the existing structure of an enterprise and market place. McDonalds approach to fast foods changed the restaurant business. Disney's unique characters and interactive theme parks changed the entertainment business. Internet has brought a change in business scenario. It will continue to change in the industry sectors also.

Need for Innovation

The innovative spirit among the inventors is responsible for inventing the car, the telephone, television, and the internet. The world is benefited from manifestations of the innovative spirit. One may wonder what makes the need for innovation

different in the present day context (Chesbrough, 2011). Business World is in the middle of significant transition. Three factors are playing important role in the present transition period. The three factors are information technology, the expanding world, and demanding customers.

Need for Clear Focus

The important aspect that is required in an enterprise is the type of innovation it wants to pursue (Dastur, Sawant, & Bawa, 2012). There will be conflicting views among the executives in enterprises. Many times ideas suggested by the employees are not encouraged for fear of the unknown once they are presented. Enterprises generally support very small impact of efficiency innovation. It is very important at the beginning of any project to discuss the type of innovation the enterprise is expecting. Decision has to be taken in light of the available resources such as the core competencies and culture of the enterprise. Innovation momentum for an enterprise is like a snow ball. Snow ball gains in size and power as it rolls down the hills. Similarly an enterprise gains power as more and more employees adopt an innovative attitude and succeed in finding distinctly new and better ideas.

Enterprises cannot afford to depend on one person or one department to focus on the future. They need to have every employee in every area of an enterprise to be on the lookout of innovative ideas. It is the tendency of the employees to continue old and comfortable way of doing things. Sometimes it is a tough proposition for the employees to drop their emotional attachment to the way they are presently working. Enterprises need an innovation road map and detailed action plans. Innovation road map is required to enable enterprises in the creation of innovation centric organization. Innovation road map should address knowledge and skills of employees for innovation environment of an enterprise (Dyer, Halgregeresen, & Christensen, 2011).

Knowledge Harvesting

Knowledge harvesting is an integrated set of processes that allow the often hidden insight of human expertise in business enterprises to be captured. Then it is converted into a specific actionable know-how that can be transferred to others. The actionable know-how will be stored in an information system which will help thereby to create organizational value and enhance in decision making processes, record facts about critical decisions. Further it will facilitate to create Meta knowledge about actual working adapted to changes (Lester & Piore, 2010).

The word “Harvesting” generally applies to agriculture. It refers to the practice of increasing the yield of cultivable land. In the same way organizational databases are considered as equivalent to cultivable land. In the context of organizational environment, the employees “Wisdom” can be considered as “Manure”. The employees wisdom will help the data in the system to be converted into useable information.

Case Illustration

3G textile mill is one of the leading textile mills in India. This textile mill has been in the business over five decades. The unique aspect of this mill is that is being managed by a management consisting of president, senior president, and vice president who belong to three different generations. Their approach and decisions are based on their experiences, business insight and education. All the three in the top management team keep updating themselves new technologies in their textile business besides keeping them aware of new trends in information technology and market trends. They are assisted by a team of executives who have rich experience in the respective field of textile industry.

3G textile mill has been manufacturing cloth material for western and eastern ladies requirements. They also manufacture shirting, and suiting materials besides producing readymade garments

for men and women. They enjoyed major global market share for their products till the globalization policy followed by many countries. It is because of changing scenario in the global market, their market share is affected. Consequently their profits have been fluctuating. The top management along with their team of executives have analyzed that they have cost advantage over their competitors. They have edge over in the process technology and size of business enterprise. They have come to the conclusion, after the analysis is that their weak link is in the fashion technology. They hired the services of domain experts in the field of textiles who are familiar with the tastes of people in European countries and Eastern countries. They expected that their guidance will help them to improve their global market share. The domain experts have their offices in Paris and Singapore.

Cloth Material

Cloth material is designed by the domain experts from their experience based on the needs of the consumer requirements such as color combination of dyes with the cotton mix. The domain experts have made use of the virtual reality concept in designing the cloth material according to the latest trends in the European and Eastern countries.

Readymade Garments

The domain experts have taken into consideration the latest trends in the fashion in the European and Eastern Countries. The domain experts designed and created garments on the basis of real world requirements of the two countries (Gallo, 2012). Simulated versions of readymade garments are carried out in the computer systems under the cloud computing environment. The concept of virtual reality helped them to look from the real world situation. Resources in cloud computing environment have made it possible for the domain experts to create designs for the requirements of 3G textile mill.

Management Team of 3G Textile Mill

The management team along with their technical executives in India is able to view the design of cloth material and design of readymade garments on their system. They have suggested the changes wherever they felt they are required.

Social Media and 3G Textile Mill

The management team at 3G Textile Mill has realized that the market place is about brands and consumers speaking to each other. Further brands have been losing control over the kind of conversation that is taking place about the brand. The internet has moved the control of brands into consumers' hands and that is posing new challenges for brand owners. It has become a necessity for brand owners either need to have more people speaking positively about the brands or have fewer people speaking negatively. Social media plays a crucial role in both the cases. They have realized the need for exclusive social media sites for business activities.

Marketing Wing

The management team has felt that the concept of social media is to be made use of for their business. A sample cloth material and garments have been designed by the domain experts with their experience, knowledge and also with the inputs from the executives of the 3G Textiles. The concepts of virtual reality and multimedia are applied in creating a material with the different combination of colors. While designing the cloth material, the types of dyes available in the market are also kept in mind.

The marketing department of the mills has created an exclusive social media channel for the prospective customers to express their opinions of the proposed design of a cloth material and garments. The proposed design of their products is uploaded in the site. Vendors of dyes have been

informed to visit the social media site to get an idea the types of dyes are to be supplied once the proposed design is finalized.

Interactions with the prospective customers on the social media site have facilitated them to understand their needs and tastes. This has helped the mills to design a cloth material and garments from the data and information from the social media site created exclusively by the marketing department (Chang & Taylor, 2016).

This Case Illustration gives an overview of the approach followed by the textile mill to design a cloth material for ladies making use of the concepts such as virtual reality and multimedia. The feedback from the prospective customers on the social media channel facilitated them to finalize the design of the cloth for production. Further vendors who supply dyes and raw cotton are able to get the required information for supplying the materials as per the requirements of a new design. Customers' tastes are becoming more homogeneous around the globe. The feedback on social media channel facilitates to innovative approach (Chiporni, Chiesa, & Frattini, 2011).

Innovative Approach

It is clear from 3G textile mill's innovative approach that new products can be designed and developed by the domain experts who are away from the Indian soil. This is because the domain experts' involvement of the process of continuous exchange of ideas among the members of virtual team dispersed across the globe. This process helps in generating alternative ideas by taking inputs from different sources and structuring through virtual reality applications. This model provides an idea for the creation of global innovative approach. It facilitates to structure the work flow by visualizing the various phases of the development of a product.

In today's knowledge rich environment, 3G textile mill has proved that enterprises can no longer afford to rely entirely on their ideas to

advance their business. It may be noted that they have not restricted their innovative approach to a single path market. As a result the traditional model of innovation which has been largely internally focused or a “Closed One” has become obsolete. Emerging in its place is a new paradigm “open Innovation”. This strategically leverages internal and external sources of ideas and takes them to market through multiple paths. Global enterprises can take advantage of unique knowledge and resources wherever they are located. Information and communication technology has created virtualization in activities and ways of working. The approach of 3G textile mill explains in adopting innovative approach. Further it applies the concept of “Mind Invoking” (Govindarajan, & Christrimble, 2011).

Approach in Case Illustration

3G textiles felt for the need for innovation. They have a clear focus in their innovative approach. They have applied revolutionary innovation method in designing and developing a product. They have involved their executives along with the domain experts in the innovative approach followed by them. They have made use of the concepts such as virtual reality, cloud computing and knowledge harvesting in their product designs. It is interesting to note that the management team consists of three different generation people. They are ready to adopt innovative approach to face the challenges in the present globalization scenario. It is apt to note the observations of Linda A H, Greg B, Emily T, & Kent L, (2014) in their book that though each of our leaders and their firms differed in key ways all leaders paid particular attention to making sure their organizations were able to collaborate, engage in discovery driven learning, and make integrative decisions. In the case of 3G Textiles the top management team, executives in their organizations, domain experts, vendors, and the prospective customers worked as

a team to apply revolutionary innovation approach (Gupta, 2009).

Future Trends

The future of innovation is bright. It may be said there is no future without innovation. Adopting new methods and designing of products will ultimately be overtaken by new innovative approach. Innovation is the topic for twenty first century business. It is considered to be a primary source of competitive differentiation. Further it will be the most important driver of growth at corporate level and even at country level.

CONCLUSION

The unique approach to innovative thinking involves the integration of creative thinking, strategic thinking and transformational thinking. The role of innovation in any economy is to be addressed to knowledge performance, skills, innovation environment, and progressive thinking of communities. It is apt to recall Peter Drucker's observation on seven basic sources of innovation. They are 1) The surprise of unexpected success or failures, 2) Inconsistencies when things do not add up according to conventional wisdom, 3) Desperation when there is a crying need for a better way, 4) Outdated industries or processes over due for change, 5) Life styles or demographic changes, 6) Attitude changes of customers, and 7) Discovery where new knowledge or capabilities promote new opportunities. The industry landscape is changing because of the content in social media sites. This will be helpful for enterprises to plan their business activities. It is not about east or west or developed markets versus developing markets. We have become more of a connected world. Ideas can come from anywhere.

REFERENCES

- Chang, W., & Taylor, S. A. (2016). The Effectiveness of Customer Participation in New Product Development a Meta-Analysis. *Journal of Marketing*.
- Chesbrough, H. (2011). *Innovation: Open Services Innovation*. A Wiley Imprint.
- Chiporni, D., Chiesa, V., & Frattini, F. (2011). The Open Innovation Journey: How Firms Dynamically Implement the Emerging Innovation Management Paradigm. Elsevier.
- Dastur, P. A., Sawant, S., & Bawa, B. (2012). *Primer to Disruptive Innovation: Disruptive Innovation what every business should know*. Mumbai: Embassy Book Distributors.
- Dyer, J., Halgregeresen, & Christensen, C. M. (2011). Putting the innovator's data into practice: The Innovator's DNA mastering the five skills of disruptive innovators. Harvard Press.
- Estrim, J. (2009). *The capacity for change; closing the innovation gap-reigniting the spark of creativity in a global economy*. Academic Press.
- Gallo, C. (2012). *Connect Your Brand, The Power of Four Square-7 Innovative Ways to Get Your to Check in Wherever They Are*. New Delhi: Tata McGraw Hill Education Private Limited.
- Govindarajan, & Christrimble. (2010). *Introduction: The other side of Innovation*. Harvard Business Review Press.
- Gupta, P. (2009). *Business Innovation-A Comprehensive Approach to Institutionalize Business Innovation*. New Delhi: S Chand & Company Limited.
- Lester, R. K., & Piore, M. J. (2010). *Conversion, Interpretation and ambiguity, Innovation: Innovation the missing dimension*. Harvard University Press.
- Linda, A. H., Greg, B., Emily, T., & Kent, L. (2014). *Collective Genius: The Art of Practice of Leading Innovation*. Harvard Business Press.
- Prahalad, C. K., & Mashelkar. (2011). *Innovation Holy Grail: Inspiring & Executing Innovation*. Harvard Business Review Press.
- Radjou, N., Prabhu, J., & Ahuja, S. (2012). Jugard: A break through growth strategy: Jugard Innovation. Noida: Random House Publishers India Private Ltd.
- Sunted. (2011). *Inside the Chinese Business Mind – A Tactical Guide for Managers*. Mumbai: Jaico Publishing House.
- Tocci, C. L., Chesbrough, H., & Piller, F. (2016). When do Firms Undertake Open Collaborative Activities: Introduction to the Special Selection on Open Innovation and Open Business Models: Industrial and Corporate. Oxford University Press.
- Vangundy, B. (2008). *Getting to Innovation*. New Delhi: PHI Learning Private Limited.
- Vint, V. K. (2009). *Global Economy Strategy; the Global Emerging Market-Strategic Management and Economics*. New York: Routledge.

ADDITIONAL READING

- Aduah, A. (2009). *Disruptive Technologies as new games; Strategic Competitive advantage*. New York: Routledge.
- Chesbrough, H. (2006). *Why Business Models need to Open up; Open Business Models – How to thrive in the new Innovation Landscape*. Massachusetts: Harvard Business School Press.
- Christensen, C. M., Kaufman, S. P., & Shih, W. C. (2011). *Innovation killers how financial tools destroy your capacity to do new things Harvard Business Review on inspiring & executing innovation*. Boston: Harvard Business Review Press.

KEY TERMS AND DEFINITIONS

Cloud Computing: A Computing Model that makes information technology resources such as servers, Middleware, and applications available over the Internet as services to business Organizations in a self service manager.

Collaborative Technology: Various concepts in information and communication technology facilitate to work jointly in developing application software.

Core Team: A team consisting of domain experts to perform multi-Disciplinary activities.

Globalization: The process by which business starts operating on a global scale.

Knowledge Harvesting: Integrated set of processes that allow the often hidden insight of human expertise to be captured. Then it is converted into a specific actionable know-how that can be transferred to others.

Social Media: Use web technology to quickly disseminate knowledge and information to huge number of users. Face book, Twitter, Blog, and other similar social networking site are collectively referred to social media.

Virtual Reality: A system in which images that look like real three dimensional objects are created by a computer system.

M

Sustainable Competitive Advantage With the Balanced Scorecard Approach

Jorge Gomes

ISEG, Universidade de Lisboa, Portugal

Mário José Batista Romão

ISEG, Universidade de Lisboa, Portugal

INTRODUCTION

Why are some firms more successful than others? This question has been intensely debated by strategic management researchers over the last thirty years. Competitive advantage is recognised as being the major cause for explaining top organizational performance and is a fundamental goal of academic strategic management studies. Recently, there has been an increasing amount of empirical research on the subject of competitive advantage (Ray et al., 2004; Newbert, 2008) and about distinguishing competitive advantage from organisational performance (Powell, 2001). The relevance of competitive advantage is not simply determined by external factors, but also by those internal sources that have been considered critical for successful organisations. Porter (1985) says that competitive advantage is at the heart of organisational performance in the competitive business environment, and that the core of this view is that in order to achieve competitive advantage, firms should systematically provide added value to customers relative to the competition. Peteraf (1993) defined competitive advantage as being sustained performance above normal returns, and Barney (2002) claims that superior performance is obtained through the value generation of internal resources usage. The research of Wernerfelt (1984), Rumelt (1984), Barney (1986), Dierickx and Cool (1989), Amit and Schoemaker (1993) and Peteraf (1993) have all been recognised as a reference for the study of sustainable competitive

advantage (SCA), based on the resource-based view of the firm approach (RBV). According to this view, a firm's endowment of resources is that which gives it a sustainable competitive advantage. RBV highlighted the relevance of intangible resources as a crucial factor for SCA. Intangible assets, such as intellectual property, knowledge and skills of employees or relationships with our customers, are all sources of competitive advantages and long-term financial success, which are both increasingly important for organisations today (Kaplan & Norton, 1992, 1996, 2000). The competitive market environment and dealing with intangible assets have both become the main reason for SCA today for organisations, and they are needed to support organisations' strategy in reliable frameworks that measure strategy implementation, align business challenges with different internal activities, and include the management of IS/IT strategy, as well as other initiatives. The most globally-recognised management support system for fulfilling these organisational performance challenges is the Balanced Scorecard (BSC).

BACKGROUND

Sustainable Competitive Advantage

SCA is an organisation's ability to carry out the set of necessary steps for achieving lower costs than the competition in an efficient and unique way, creating differentiated value for buyers

Figure 1. How resources and capabilities combine to create differentiation



(Porter, 1985). SCA is the aim of every organisational strategy and it can be achieved in numerous ways. The rationale behind Porter's view (Porter, 1985) is that a higher performance is correlated with competitive advantage, and that achieving an advantage will certainly result in superior performance (Reed & DeFillipi, 1990). According to this approach, the organisational strategy definition depends on the external environment context. With the advent of the RBV approach, researchers focus mainly on the internal capabilities of the organisation, considering the external factors as background (Pralahad & Hamel, 1990; Spanos & Lioukas, 2001). The ultimate organisational goal is to achieve a superior return on capital through the identification, development, protection and allocation of resources and capabilities, and thus supply the organisation with a SCA (Amit & Schoemaker, 1993). RBV focusses on the development of resources and capabilities, supported by the belief that the set of resources in companies are heterogeneous, valuable and scarce (Collis & Montgomery, 1995; Hamel, 1994; Prahalad & Hamel, 1990), which are recognised as being a source of SCA (Barney, 1991; Barney & Wright, 1998; Wright, McMahan, & McWilliams, 1994). This approach assumes that firms can only create a SCA if they have superior resources, in conjunction with organisational capabilities, and that this combination is the best entry barrier for competition (Barney, 1991) (Figure 1).

Barney (1991) claims that for creating differentiating advantage, firm's resources and capabilities

must be valuable, rare, imperfectly imitable, and non-substitutable (VRIN). Intangible resources are the main drivers of performance sustainability across firms, and are usually tacit, difficult to codify (Kogut & Zander, 1992; Conner & Prahalad, 1996), and hard to acquire or develop, replicate, accumulate (Itami, 1987; Winter, 1987), and to be imitated by others (Rumelt, 1986; Dierickx & Cool, 1989; Nelson, 1991). It is this difficulty of imitation which makes them valuable and a SCA for a company (Lippman & Rumelt, 1982; Hall, 1993). Although subject to some criticism, some RBV authors (Peteraf & Barney, 2003; Amit & Schoemaker, 1993; Mahoney & Pandian, 1992; Conner, 1991; Barney, 1991; Wernerfelt, 1984) recognise the existence of bridges between both approaches and their complementarity in explaining the sources of firm performance (Foss, 1996). Teece, Pisano, and Shuen (1997), however, hold the opinion that an organisation has to establish dynamic capabilities to ensure its SCA. The dynamic capabilities view (DCV), which is supported by the RBV approach, notes that the dynamic capabilities of an organisation need to transform their resources into competitive advantages in order to generate competitive advantage. DCV emphasises the organisations' adaptability to the environment and their capability of creating and absorbing new knowledge, which is one important basis of its dynamic capabilities (Zollo & Winter, 2002).

According to the knowledge-based view (KBV), knowledge seems to be the most critical input and value source for a firms' process activi-

ties (Grant, 1996a), as all inputs are established on the foundation of knowledge. Knowledge is the most important resource that a firm possesses to create added value (Grant, 1996a), and it also has to possess features which are unique, difficult to transfer, and hard to replicate (Kogut & Zander, 1992; Spender, 1996; Grant, 1996a; Grant, 1996b), all of these being features which are recognised as being the sources of SCA.

The lack a solid operational definition for SCA and knowledge about how a network environment can enhance the core competencies of a firm are some of the obstacles facing the further development of the SCA theory (Hoffman, 2000).

BALANCED SCORECARD

The BSC concept assumes that competitive advantages are generated not only by factors that can be quantified, but that they are also derived from intangible assets, such as intellectual property, knowledge and skills of employees or customer relationships, which must be measured and managed. Traditional performance measures are insufficient to gauge performance and guide organisations in today's rapidly changing and complex economic environment. BSC recognises the rise of intangible assets in value creation, and the limitations of financial measurements in guiding organisations to the future. BSC is an alternative approach that balances financial and operational measures and which allows an organisation to control its organisational performance in a multidimensional concept. As defined by Kaplan and Norton (1996), BSC translates an organisation's mission and strategy into a comprehensive group of performance measures, providing the framework for a strategic measurement system, using a balanced set of measures for both financial and non-financial performance, which are linked in a cause-and-effect way, which is organised into four perspectives.

The Balanced Scorecard concept evolved over time. It was initially proposed as a performance

measurement system, and it then developed to being a more comprehensive performance management system. Today it is a strategic management system for managing organisational performance (Brudan, 2008; Andersen et al., 2001). Kaplan and Norton (2001) introduced five principles to keep the focus of management processes on organisational strategy:

1. **Strategy Translation Into Operational Terms:** A good strategy implementation needs a good understanding and internal adoption.
2. **Organisation Alignment With the Strategy:** Organisational performance is more than just the sum of its parts; all individual strategies should be linked and integrated into a single strategy.
3. **Daily Tasks Performed and Aligned With the Strategy:** BSC is the communication tool used by executives and managers to reach their employees.
4. **Strategy as a Continuous Process:** strategy should always be linked to the budgeting process.
5. **Executive Leadership Should Mobilise Change:** Executives and managers have a crucial role to play in the mobilisation process.

The BSC approach contributes to enhance firm's performance and SCA, by enabling four main elements that make the difference in the strategic management and learning when compared with other frameworks (Kaplan and Norton, 2007) (Figure 2), namely the following:

1. Clarification of the vision and the mission for all employees within the organisation;
2. The role of communication as an integration factor of all individual business units' efforts to meet the objectives;
3. Focus on the scorecard's importance as a tool that enables the revised strategy.

4. Focus on strategic feedback, by incorporating professional knowledge about the changes in the competitive environment.

Recently developments have taken BSC to being an *Execution Premium system* (Kaplan and Norton, 2008), which enhances the integration of strategy into a company's operations. This new BSC version is based on a continuous improvement approach and includes six phases:

1. Developing strategy based on the internal context, the external context, and the existing strategy.
2. Planning the strategy by developing Strategy Maps.
3. Aligning all the initiatives with the organisational strategy.
4. **Planning Operations:** Budget and strategy should be linked.
5. **Monitoring and Learning:** Both the strategy and the operations are reviewed in meetings.

6. **Testing and Adapting:** Checking if the strategy is being carried out and making the necessary modifications.

The emergence of Strategy allows organisations to describe and communicate their strategies as a way of identifying the cause-and-effect relationships between different objectives in a visual way, describing the strategy that should be followed, measured and actively managed (Kaplan & Norton, 2000). How the organisation will change is described by the strategy map. The strategy map frames its balanced scorecard in order to support the strategy and to explain from various perspectives what will be different and how the organisation will change. More importantly, strategy maps explain the underlying drivers of change. Currently, the strategy map and the cause-and-effect attract considerable interest among practitioners and firms that are developing their BSC based on the strategy map (Kasurinen, 2002).

Later writings suggested that the BSC may also facilitate the management of intangibles and serve as the basis of managing competing stakeholder

Figure 2. Managing strategy
Adapted from Kaplan & Norton, 1996a.



objectives (Kaplan & Norton, 2001, 2004). Despite its worldwide success in many organisations, BSC has drawn criticism, particularly from an extensive range of academics. The most important criticisms relate to the following:

- The causality relationships between the areas of measurement in the BSC (Nørrekelit, 2003; Akkermans & Van Oorscht, 2005).
- It neglects the dimension of time (Nørrekelit, 2003).
- BSC's reliance on a limited number of measures (Akkermans and Van Oorscht, 2005; Hudson, Smart, & Bourne, 2001; Mårtensson, 2008).
- The lack of integration between top-level and operational levels' measures (Akkermans and Van Oorscht, 2005; Hudson, Smart, & Bourne, 2001).
- An excessive emphasis on those internal processes that address external processes (Akkermans and Van Oorscht, 2005; Nørrekelit, 2003).
- The BSC concept, without modifications, is not effective enough to contribute to corporate sustainability (Bieker, 2003; Akkermans and Van Oorscht, 2005).

FUTURE RESEARCH DIRECTIONS

Sustainability Balanced Scorecard (SBSC)

Innovative business performance measures are mainly based on non-financial measures, such as operational measures, which can be used by managers during their decision-making process. Scholars of the field of sustainability management have identified BSC as being a promising management tool for the development of performance measures that combine sustainability and management in an integrated framework. The concept called the Sustainability Balanced Scorecards (SBSC) was developed with the goal

of the integration of non-monetary factors related to the current environmental and social challenges (Bieker & Waxenberger 2002; Epstein & Wisner, 2001; Figge et al., 2001, 2002a, 2002b; Schaltegger 2004; Schaltegger & Wagner 2006a, 2006b). Improving on the classical BSC by adding environmental and social components, represents an important step in promoting the concept of corporate social responsibility. The inclusion of sustainable development dimensions requires the extension of existing BSC perspectives, as well as a set of indicators. BSC has the ability to integrate intangible and qualitative aspects, but needs further development in order to become an integrated system for the management of corporate sustainability. In this research context, different approaches have been developed in order to enable SBSC to incorporate the main factors of sustainability in a measurement and performance management system (Bieker & Waxenberger 2002; Epstein & Wisner 2001; Figge et al., 2001, 2002a, 2002b, 2003; Schaltegger & Wagner 2006a).

Conceptually, sustainability management using BSC seeks to address the problem of corporate contributions to sustainability in an integrative way. It posits that in order for companies to be able to contribute to sustainable development, it is desirable that corporate performance improves simultaneously in all three of the dimensions of sustainability – economic, environmental and social. However, the practical use of BSC in the field of sustainability is a relatively new concept when compared with other systems. The research into how environmental issues can be integrated into the core business process through the extension of BSC is still at a very early stage. Results will open the door for future studies to consider the role of BSC as a performance management system which is able to include both social and environmental issues.

Balanced Scorecard in Healthcare

The review of the literature by Banchieri et al., (2011) found that more than 30% of papers

produced during the last years were applied to the healthcare industry. The implementation of Balanced Scorecard in the healthcare sector experienced many common factors when compared with its implementation in other sectors, although some challenges unique to healthcare exist. This is not surprising, as it is acknowledged that healthcare is one of the most complex industries (Marr & Creelman, 2011). One reason for the relatively slow assimilation of BSC in healthcare organisations is that they have traditionally relied heavily on the use of non-financial statistics, and therefore most healthcare organisations believed that they already possessed tools similar to Balanced Scorecard (Kocakülâh & Austill, 2007). In a comprehensive review, Zelman et al., (2003) show that Balanced Scorecard has been introduced across all areas related to healthcare. It is not only used for strategic management at an organisational level, but it is also used as an evaluation tool for: health programmes, the quality of care, improvement projects and also measuring performance in many hospitals (Zelman et al., 2003).

Inamdar et al., (2002) found that the main reason for its extensive implementation was a planned response to external forces, namely increasing financial pressures that motivated organisations to search for more effective and relevant strategic management tools than those which they were currently using.

Several authors have highlighted the benefits of using Balanced Scorecard for healthcare:

- The clarification of strategy and a framework for decision-making (Inamdar et al., 2002);
- Focus and alignment throughout all levels of the organisation (Chang et al., 2008);
- Quality improvements in patient care and outcomes (Aidemark & Funck 2009);
- Facilities a move from measuring performance to managing performance (Gottlieb, 2008);

- Enhances focus on customer service (Kocakülâh & Austill, 2007).

CONCLUSION

Value creation is the ultimate goal of an organisation. In the 1970s, Porter recommended that, in order to perceive the market position, one needs to look at the five competitive forces that affecting a company and its ability to earn above-average profits. The RBV approach suggests building a company around its core competencies. In the post-industrial world, the benefit from most traditional SCA strategies may already have been, or is about to be, eliminated. The competition can easily achieve economies of scale, and process improvements can easily be copied, although customer loyalty is harder to maintain as customers are constantly confronted by new products and services. Due to the diversity of challenges faced by organisations today, globalisation and multi-cultural factors keep pace with technological developments in an increasingly demanding and competitive environment, and thus it is unlikely that a simple qualitative approach can be universally applied, no matter how full it is of insight. In most industries, some firms are more profitable than others, regardless of whether the average profitability of the sector is high or low. Same organisations possess something special that is hard to imitate and which allows them to outperform their rivals. These unique and special skills and assets are known as resources, and they are the sources for SCA. In Porter's view, sustainability is achieved when the advantage resists the erosion of concurrency.

Corporate social responsibility and sustainable development are both issues that have recently become greatly exploited, on account of their importance for business success and the competitiveness of organisations.

SBSC helps to address different environmental and social aspects that are related to their relevance for the implementation of strategy and their execution at a business unit or company level

Future research should refine certain areas in order to give more insight into the following: strategies of successful leadership teams, success factors that provide flexibility, innovation and creativity, development of more trustworthy customer relationships, strategies development for new technologies, and human capital management.

REFERENCES

- Aidemark, L., & Funck, E. (2009). Measurement and health care management. *Financial Accountability & Management*, 25(2), 253–276. doi:10.1111/j.1468-0408.2009.00476.x
- Akkermans, H., & Van Oorscht, K. (2005). Relevance assumed: A case study of balanced scorecard development using system dynamics. *The Journal of the Operational Research Society*, 56(8), 931–941. doi:10.1057/palgrave.jors.2601923
- Amit, R., & Schoemaker, P. J. H. (1993). Strategic assets and organizational rent. *Strategic Management Journal*, 14(1), 33–46. doi:10.1002/smj.4250140105
- Andersen, H., Cobbold, I., & Lawrie, G. (2001). Balanced Scorecard implementation in SME: Reflections on Literature and Practice. *Proceeding of 4th SME International Conference*.
- Banchieri, L., Planas, F., & Rebull, M. (2011). What has been said, and what remains to be said, about the balanced scorecard? *Zb. rad. Ekon. fak. Rij.*, 29(1), 155–192.
- Barney, J. B. (1986). Organizational culture: Can it be a source of sustained competitive advantage? *Academy of Management Review*, 11(3), 656–665. doi:10.5465/AMR.1986.4306261
- Barney, J. B. (1991). Firm resources and sustained competitive advantage. *Journal of Management*, 17(1), 99–120. doi:10.1177/014920639101700108
- Barney, J. B. (2002). *Gaining and Sustaining Competitive Advantage* (2nd ed.). Reading, MA: Addison-Wesley.
- Barney, J. B., & Wright, P. M. (1998). On becoming a strategic player: The role of human resources in gaining competitive advantage. *Human Resource Management*, 37(1), 31–46. doi:10.1002/(SICI)1099-050X(199821)37:1<31::AID-HRM4>3.0.CO;2-W
- Bieker, T. (2003). Sustainability Management with the Balanced Scorecard. *Proceedings of 5th International Summer Academy on Technology Studies*, 17–34.
- Bieker, T., & Waxenberger, B. (2002). Sustainability Balanced Scorecard and Business Ethics. *proceedings of 10th International Conference of the Greening of Industry Network*.
- Brudan, A. (2010). Desired State of Evolution – An integrating management tool. *Proceedings of the 2010 ANZAM Conference*.
- Chang, W., Tung, Y., Huang, C., & Yang, M. (2008). Performance improvement after implementing the Balanced Scorecard: A large hospitals experience in Taiwan. *Total Quality Management & Business Excellence*, 19(11), 1143–1154. doi:10.1080/14783360802323560
- Collis, D., & Montgomery, C. (1995). Competing on resources: Strategy in the 1990s. *Harvard Business Review*, 76(3), 118–128.
- Conner, K. R. (1991). A Historical Comparison of Resource-Based Theory and Five Schools of Thought Within Industrial Organization Economics: Do We Have a New Theory of the Firm? *Journal of Management*, 17(1), 121–154. doi:10.1177/014920639101700109

- Conner, K. R., & Prahalad, C. K. (1996). A resource-based theory of the firm: Knowledge versus Opportunism. *Organization Science*, 7(5), 477–501. doi:10.1287/orsc.7.5.477
- Dierickx, I., & Cool, K. (1989). Asset Stock Accumulation and the Sustainability of Competitive Advantage. *Management Science*, 35(12), 1504–1511. doi:10.1287/mnsc.35.12.1504
- Epstein, M., & Wisner, P. (2001). Using a Balanced Scorecard to Implement Sustainability. *Environmental Quality Management*, 11(2), 1–10. doi:10.1002/tqem.1300
- Figge, F., Hahn, T., Schaltegger, S., & Wagner, M. (2001). Sustainability Balanced Scorecard. Value-Based Sustainability Management with the Balanced Scorecard. Lüneburg: Center for Sustainability Management (CSM).
- Figge, F., Hahn, T., Schaltegger, S., & Wagner, M. (2002a). Development of a Sustainability Balanced Scorecard: Translating Strategy into Value-Based Sustainability Management. *Journal of the Asia Pacific Centre for Environmental Accountability*, 8(1), 3–16.
- Figge, F., Hahn, T., Schaltegger, S., & Wagner, M. (2002b). The Sustainability Balanced Scorecard - Linking Sustainability management to Business Strategy. *Business Strategy and the Environment*, 11(5), 269–284. doi:10.1002/bse.339
- Figge, F., Hahn, T., Schaltegger, S., & Wagner, M. (2003). The Sustainability Balanced Scorecard as a Framework to Link Environmental Management Accounting with Strategic Management. In M. Bennett, P. Rikhardsson, & S. Schaltegger (Eds.), *Environmental Management Accounting – Purpose and Progress* (pp. 17–40). Dordrecht: Kluwer Academic Publishers. doi:10.1007/978-94-010-0197-7_2
- Foss, N. (1996). Research in Strategy, Economics, and Michael Porter. *Journal of Management Studies*, 33(1), 1–24. doi:10.1111/j.1467-6486.1996.tb00796.x
- Gottlieb, G. (2008). Moving from performance measurement to strategy management at Brigham and Women's / Faulkner Hospital. *Balanced Scorecard Report*, 10(2), 11–13.
- Grant, R. M. (1996a). Toward a knowledge-based theory of the firm. *Strategic Management Journal*, 17(special issue), 109–122. doi:10.1002/smj.4250171110
- Grant, R. M. (1996b). Prospering in dynamically-competitive environments: Organizational capability as knowledge integration. *Organization Science*, 7(4), 375–387. doi:10.1287/orsc.7.4.375
- Hall, R. (1993). A framework linking intangible resources and capabilities to sustainable advantage. *Strategic Management Journal*, 14(8), 607–618. doi:10.1002/smj.4250140804
- Hamel, G. (1994). The concept of core competence. In G. Hamel & A. Heene (Eds.), *Competence based competition*. New York: John Wiley.
- Hoffman, N. P. (2000). An Examination of the Sustainable Competitive Advantage Concept: Past, Present and Future. *Academy of Marketing Science Review*, 4.
- Hudson, M., Smart, A., & Bourne, M. (2001). Theory and practice in SME performance measurement systems. *International Journal of Operations & Production Management*, 21(8), 1096–1115. doi:10.1108/EUM0000000005587
- Inamdar, N., Kaplan, R., & Reynolds, K. (2002). Applying the balanced scorecard in healthcare provider organizations. *Journal of Healthcare Management*, 47(3), 179–195. PMID:12055900
- Itami, H., & Roehl, T. W. (1987). *Mobilizing Invisible Assets*. Cambridge, MA: Harvard University Press.
- Kaplan, R. S., & Norton, D. P. (1992). The Balanced Scorecard: Measures that Drive Performance. *Harvard Business Review*, (January-February), 71–79. PMID:10119714

- Kaplan, R. S., & Norton, D. P. (1996a). *The Balanced Scorecard: Translating Strategies into Action*. Boston, MA: Harvard Business School Press.
- Kaplan, R. S., & Norton, D. P. (2000). Having Trouble with Your Strategy? Then Map It. *Harvard Business Review*, 75(5), 167–176. PMID:11143152
- Kaplan, R. S., & Norton, D. P. (2001). *The Strategy-Focused Organization: how Balanced Scorecard Companies Thrive in the New Business Environment*. Boston, MA: Harvard Business School Press.
- Kaplan, R. S., & Norton, D. P. (2004). Measuring the Strategic Readiness of Intangible Assets. *Harvard Business Review*, 82, 52–63. PMID:14971269
- Kaplan, R. S., & Norton, D. P. (2007, July-August). *Using the Balanced Scorecard as a Strategic Management System*. *Harvard Business Review*.
- Kaplan, R. S., & Norton, D. P. (2008). *Execution Premium: Linking Strategy to Operations for Competitive Advantage*. Boston: Harvard Business Press.
- Kasurinen, T. (2002). *Conceptualizing the encoding process related to institutionalization in organizations: from key performance indicator scorecard to a strategic balanced scorecard*. Helsinki School of Economics, working paper, W-331.
- Kocakülâh, M., & Austill, D. (2007). Balanced Scorecard Application in the Health Care Industry: A Case Study. *Journal of Health Care Finance*, 34(1), 72–99. PMID:18972988
- Kogut, B., & Zander, U. (1992). Knowledge of the firm, combinative capabilities, and the replication of technology. *Organization Science*, 3(3), 383–397. doi:10.1287/orsc.3.3.383
- Lippman, S. A., & Rumelt, R. P. (1982). Uncertain imitability: An analysis of inter firm deficiency under competition. *The Bell Journal of Economics*, 13(2), 418–438. doi:10.2307/3003464
- Mahoney, J., & Pandian, R. (1992). The Resource-Based View within the Conversation of Strategic Management. *Strategic Management Journal*, 13(5), 363–380. doi:10.1002/smj.4250130505
- Marr, B., & Creelman, J. (2011). *More with Less: Maximising Value in the Public Sector*. London: Palgrave Macmillan. doi:10.1057/9780230300408
- Mårtensson, M. (2009). Recounting counting and accounting: From political arithmetic to measuring intangibles and back. *Critical Perspectives on Accounting*, 20(7), 835–846. doi:10.1016/j.cpa.2008.09.006
- Nelson, R. R. (1991). Why do firms differ, and how does it matter. *Strategic Management Journal*, 12(S2), 61–74. doi:10.1002/smj.4250121006
- Newbert, S. (2008). Value rareness, competitive advantage, and performance: A conceptual-level empirical investigation of the resource-based view of the firm. *Strategic Management Journal*, 29(7), 745–768. doi:10.1002/smj.686
- NØrreklit, H. (2003). The Balance on Balanced Scorecard: A critical analysis of its assumptions. *Management Accounting Research*, 11(1), 65–88.
- Peteraf, M. A. (1993). The Cornerstones of Competitive Advantage: A Resource-Based View. *Strategic Management Journal*, 14(3), 179–191. doi:10.1002/smj.4250140303
- Peteraf, M. A., & Barney, J. (2003). Unraveling the Resource-Based Tangle. *Managerial and Decision Economics*, 24(4), 309–323. doi:10.1002/mde.1126
- Porter, M. (1985). *Competitive Advantage: Creating and Sustaining Superior Performance*. New York: Free Press.
- Powell, T. (2001). Competitive advantage: Logical and philosophical considerations. *Strategic Management Journal*, 22(9), 875–888. doi:10.1002/smj.173

- Prahalad, C. K., & Hamel, G. (1990, May-June). *The Core Competence of the Corporation*. *Harvard Business Review*.
- Ray, G., Barney, J., & Muhanna, W. A. (2004). Capabilities, business processes, and competitive advantage: Choosing the dependent variable in empirical tests of the resource-based view. *Strategic Management Journal*, 25(1), 23–37. doi:10.1002/smj.366
- Reed, R., & DeFilippi, R. (1990). Causal Ambiguity, Barriers to Imitation and Sustainable Competitive Advantage. *Academy of Management Review*, 15(1), 88–102. Retrieved from <http://www.jstor.org/stable/258107>
- Rumelt, R. P. (1984). Towards a strategic theory of the firm. In R. B. Lamb (Ed.), *Competitive Strategic Management* (pp. 566–570). Englewood Cliffs, NJ: Prentice-Hall.
- Rumelt, R. P. (1986). *Strategy, Structure, and Economic Performance*. Boston, MA: Harvard Business School Press.
- Schaltegger, S. (2004). *Sustainability Accounting and Reporting*. *Centre for Sustainability Management (CSM), University of Lueneburg* (pp. 2–3). Germany: Newsletter.
- Schaltegger, S., & Wagner, M. (2006a). Integrative Management of Sustainability Performance, Measurement and Reporting. *International Journal of Accounting, Auditing and Performance Evaluation*, 3(1), 1–19. doi:10.1504/IJAPE.2006.010098
- Schaltegger, S., & Wagner, M. (2006b). *Managing the Business Case for Sustainability: The Integration of Social, Environmental and Economic Performance*. Sheffield, UK: Greenleaf Publishing.
- Spanos, Y., & Lioukas, S. (2001). An examination into the causal logic of rent generation: Contrasting Porters competitive strategy framework and the resource-based perspective. *Strategic Management Journal*, 22(10), 907–934. doi:10.1002/smj.174
- Spender, J. C. (1996). Making Knowledge the Basis of a Dynamic Theory of the Firm. *Strategic Management Journal*, 17(S2), 45–62. doi:10.1002/smj.4250171106
- Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal*, 18(7), 509–533. doi:10.1002/(SICI)1097-0266(199708)18:7<509::AID-SMJ882>3.0.CO;2-Z
- Wernerfelt, B. (1984). A Resource-based View of the Firm. *Strategic Management Journal*, 5(2), 171–180. doi:10.1002/smj.4250050207
- Winter, S. G. (1987). Knowledge and Competence as Strategic Assets. In D. J. Teece (Ed.), *The competitive Challenge: Strategies for Industrial Innovation and Renewal* (pp. 159–178). Cambridge, MA: Ballinger.
- Wright, P. M., McMahan, G. C., & McWilliams, A. (1994). Human resources and sustained competitive advantage: A resource-based perspective. *International Journal of Human Resource Management*, 5(2), 301–326. doi:10.1080/095851994000000020
- Zelman, W., Pink, G., & Matthias, C. (2003). Use of the Balanced Scorecard in health care. *Journal of Health Care Finance*, 29(4), 1–16. PMID:12908650
- Zollo, M., & Winter, S. G. (2002). Deliberate learning and the evolution of dynamic capabilities. *Organization Science*, 13(3), 339–351. doi:10.1287/orsc.13.3.339.2780

ADDITIONAL READING

- Bharadwaj, A. S. (2000). A Resource Based Perspective on Information Technology and Firm Performance; An Empirical Investigation. *Management Information Systems Quarterly*, 24(1), 169–196. doi:10.2307/3250983
- Braam, G. J. M., & Nijssen, E. J. (2004). Performance effects of using Balanced Scorecard: A note on Dutch experience. *Long Range Planning*, 37(4), 335–349. doi:10.1016/j.lrp.2004.04.007
- Collis, D. J. (1994). Research note: How valuable are organizational capabilities? *Strategic Management Journal*, 15(S1), 143–152. doi:10.1002/smj.4250150910
- Cool, K., & Schendel, D. (1998). Performance differences among strategic group members. *Strategic Management Journal*, 9(3), 207–223. doi:10.1002/smj.4250090302
- Fiol, M. (1991). Managing Culture as a Competitive Resource: An Identity-Based View of Sustainable Competitive Advantage. *Journal of Management*, 17(1), 191–211. doi:10.1177/014920639101700112
- Hamel, G., & Prahalad, C. K. (1994). *Competing for the future*. New York: McGraw Hill.
- Heeks, R. (2006). Health Information Systems: Failure, success and improvisation. *International Journal of Medical Informatics*, 75(2), 125–137. doi:10.1016/j.ijmedinf.2005.07.024 PMID:16112893
- Helfat, C. E., & Peteraf, M. A. (2009). Understanding dynamic capabilities progress along a developmental path. *Strategic Organization*, 7(1), 91–102. doi:10.1177/1476127008100133
- Hoskisson, R. E., Hitt, M. A., Wan, W. P., & Yiu, D. (1999). Theory and Research in Strategic Management: Swings of a Pendulum. *Journal of Management*, 25(3), 417–456. doi:10.1177/014920639902500307
- Kaplan, R. S., & Norton, D. P. (2006). *Alignment: Using the balanced scorecard to create corporate synergies*. Boston, Mass: Harvard Business Press.
- Mintzberg, H., & Lampel, J. (1999). Reflecting on the Strategy Process. *Sloan Management Review*, 40(3), 21–30.
- Nonaka, I. (1994). A Dynamic Theory of Organizational Knowledge. *Organization Science*, 5(1), 14–37. <http://www.jstor.org/stable/2635068> doi:10.1287/orsc.5.1.14
- Oliver, C. (1997). Sustainable Competitive Advantage: Combining Institutional and Resource –Based View. *Strategic Management Journal*, 18(9), 697–713. doi:10.1002/(SICI)1097-0266(199710)18:9<697::AID-SMJ909>3.0.CO;2-C
- Porter, M. (1980). *Competitive Strategy: Techniques for Analyzing Industries and Competitors*. New York: Free Press.
- Porter, M. E. (1979). Competitive Forces Shape Strategy. *Harvard Business Review*, (March–April): 137–145. PMID:18271320
- Porter, M. E. (1985). *How Information gives you Competitive Advantage*. *Harvard Business Review*. July–August.
- Rahimi, B., & Vimarlund, V. (2007). Methods to evaluate health information systems in health care settings: A literature review. *Journal of Medical Systems*, 31(5), 397–432. doi:10.1007/s10916-007-9082-z PMID:17918694
- Rumelt, R. P. (1991). How much does industry matter? *Strategic Management Journal*, 12(3), 167–185. doi:10.1002/smj.4250120302
- Teece, D. J. (2010). Business models, Business Strategy and Innovation. *Long Range Planning*, 34(2-3), 172–194. doi:10.1016/j.lrp.2009.07.003

KEY TERMS AND DEFINITIONS

Competitive Advantage: An advantage that a firm has over its competitors, allowing it to generate greater margins and/or retain more customers than its competitors.

Dynamic Capabilities: Higher order strategic processes that integrate, recombine, and generate new technological and marketing capabilities, which in turn increase firm performance.

Knowledge-Based View: Considers knowledge as being the most strategically significant resource of the firm.

Porter's View: An approach that focusses on the positioning of the organisation's business or brand in the marketplace to the best advantage.

Resource-Based View: Supports that resources possessed by organisations allow the obtaining of competitive advantage, and lead to superior long-term performance.

Sustainable Competitive Advantage: Company assets, attributes, or abilities that are difficult to duplicate or exceed; and to provide a superior or favourable long-term position over competitors.

M

Transformational Leadership for Academic Libraries in Nigeria

Violet E. Ikolo

Delta State University Library, Nigeria

INTRODUCTION

Rapid changes brought about by the information age have required that libraries re-invent themselves in order to maintain their relevance in the academic community. Specifically, libraries have had to contend with changes such as dramatic increases in the cost of library materials while budgets continue to decrease; new digital formats and communication technologies and improvements in information availability (Weiner, 2003).

The university librarian, who is the head in every academic library is required to be the executive administrator, in charge of the planning, administrative and budgetary functions of the library. However, times have changed and there is an increased call for leaders who possess, communicate, and implement a vision for transformed and improved academic libraries. It is believed that the approach the university librarian adopt in directing, guiding and controlling the staff under him/her determines the pace of progress in their libraries. However, not all leadership styles are suitable for the task of provoking or fostering innovations within the academic library.

Objectives

It is the objective of this chapter to:

1. Explore the current changes evident in academic libraries;
2. Define transformational leadership;
3. Identify the practices of transformational leaders;
4. Focus on three areas where transformational leadership can apply in academic libraries in Nigeria and;
5. Suggestions on how the future transformational leaders in librarianship can be nurtured.

BACKGROUND

It is a well-known fact that the traditional mission of academic libraries has from the beginning been to select, collect, preserve and create access to intellectual resources that can promote academic studies and research. "As a result, it occupies a central and primary place in the university, because it supports all the functions of a university with respect to teaching, research, extension of the frontiers of knowledge and the transmission to posterity the learning culture of the present and past" (Lippincott, 2005).

However, the demand and expectation placed on academic libraries have changed over time. Reasons being that technology has changed the expectations of library patrons; today, people expect to be able to find and access information from wherever they are. Libraries are now hubs of technology while offering services that have become extremely complex and sophisticated (Afolabi, 2014). Bazillion (2001) stated that electronic information sources have quickly established a formidable presence in academic libraries.

Furthermore, libraries have to implement increasingly complex solutions that involve distributed networking and access to remote information resources. The use of information technology in libraries has tremendously increased due to its

enhanced user satisfaction, cost effectiveness, faster and simpler programs, rapid communicative interaction and easier operational procedures (Storey, 1995). Another typical change that has evolved in today's academic library is in the form of library access. A library card today gives more than just access to books and periodicals at the local library; it gives access to the world just about any location. It also gives access to computing and networking technologies, in all their various forms. Some latest technologies used in academic libraries include internet access and web access to library resources which are an integral part of almost all library services (Rendon, 2014).

In the light of the glaring proliferation of information and information technologies, Germano (2011) pointed out that library managers now grapple with the demands of an increasingly complex and change-laden information landscape which offers patrons seemingly unlimited choices. Most libraries now risk becoming competitively-disadvantaged when compared to other technologically-mediated, user-defined information retrieval choices that are newly available. Critical to this seeming shift in the academic library's identity is leadership.

Townley (2003) defined leadership as mobilizing people to get things done. It involves motivating people and managing resources. Martin (2015) believes that leadership is the link between "intention and outcome. Yukl (1998) gave a comprehensive definition of leadership as the process whereby an individual member in a group or organization influences the interpretation of events, the choices of objectives, strategies, the organization of activities, the maintenance of cooperative relationships, the development of skills and cooperation from the people outside the group or organization. Library leadership has traditionally focused on structural forms of leadership by creating organization structures that support organizational goals (Schwartz, 1997). More recently, the focus is now on the library leader who is uniquely positioned to understand the value of the library and how the library serves

its parent institution while preserving and supporting foundational library values that enable that service (Hinchliffe, 2011).

In the light of the foregoing, it can be surmised that the level of effectiveness of a library leader can be measured by the value of services rendered to users, and the productivity of the library staff, which in turn relies on the leadership style employed, (Ogbah, 2013); hence, Obiwuru, Okwu, Akpa and Nwankwere (2011) noted that leadership styles are predictors to leadership effectiveness whereby leadership style in an organization is one of the factors that play significant role in enhancing or retarding the interest and commitment of the individuals in the organization.

TRANSFORMATIONAL LEADERSHIP

Inspired by Burns' (1978) notion of the transformational leader, Bass (1985) extended and developed transformational leadership theory and placed it in the context of work organizations. Transformational leaders are those who have a strong sense of direction (vision), which they communicate in inspiring ways. They are charismatic and elicit emotional responses and trust from followers. In addition, transformational leaders are said to raise follower self-confidence and self-efficacy (Shamir, House, & Arthur, 1993). Transformational leadership is a leadership style that seeks positive transformations "in those who follow" (Geib & Swenson, 2013). Simola, Barling, & Turner (2012) defined transformational leadership as a type of leadership in which interactions among interested parties are organized "around a collective purpose" in such a way that "transforms, motivates, and enhances the actions and ethical aspirations of the followers."

Apart from seeking to create positive change in their followers, a transformational leader enhances the morale, performance, and motivation of the employees, inspires change driven by a strong purpose, and is able to create a culture of trust and innovation within the organization (Tedford

& Lock, 2012). The following are the characteristics of transformational leaders as proposed by (Johannsen, 2014):

- **Keep Their Ego in Check:** Advocates the importance for the leader to keep their ego under control and not let it interfere with the best interest of their team or the organization.
- **Self-Management:** Ability to achieve results through others and avoidance of obsessional and workaholic.
- **Make Difficult Decisions:** Requires the courage to sometimes make hard decisions with a clear focus on the values, vision, objectives, and goals of the organization.
- **Inspirational:** The ability to motivate others to rise to the occasion is a key attribute that help to achieve the organization's goals.
- **Entertain New Ideas:** The ability to open up or be receptive to new ideas.
- **Adaptability:** The ability to adapt to change, new situations, and seek creative ways to respond to the dynamic business environment.
- **Proactive:** The ability to think outside the box and take an active role in growing the organization.
- **Lead With Vision:** The ability to have a set of realistic and achievable vision and then communicate this vision effectively, as well as inspire a sense of commitment and purpose.

Transformational Leadership Practices

Going by the glaring technologically characterized library environment, transformational leadership can be referred to as the act of getting extraordinary things done; often in extraordinary times. The Urban Libraries Council (2012) stated that the essence of leadership is establishing a vision and translating it into reality. While focusing on the

library leader's vision for the library, it is imperative to identify five practices that transformational leaders use to lead their followers as spelt out by Kouzes and Posner (2007). They are:

- **Modeling the Way:** Simply put, this involves leading by example. Townley (2003) states that it has to do with earning the right and respect to lead through direct individual involvement and action. How does a transformational leader model the way? They make sure that the followers know the leader's guiding principles and can use them to create effective plans. Also, character and credibility are essential to successful leadership modelling. According to Urban Libraries Council (2012) library leaders are stewards of the library's reputation as a trusted, safe, reliable, and welcoming community resource. The second step in modeling the way is affirmation of values through actions. A leader personifies their values by spending their time and attention addressing what is important to those values. They ask purposeful questions and seek feedback. And leaders teach others how to model the way by creating opportunities to discuss values. Dawson and Lewis (2007) advised that transformational leaders should earn respect and trust by demonstrating the behaviors sought from others, to be reliable and to demonstrate.
- **Inspiring a Shared Vision:** Vision in leadership is the process of forming a mental image in order to set goals, make plans, and solve problems that guide the organization into the future. This is vital when leading people. Vision is all about translating dreams to reality. By articulating a vision, the leader takes the risk that the vision may not happen and may result in discredit. Research has shown that a clearly articulated vision can draw even more commitment from people than charisma

(Urban Libraries Council, 2012). The leader must be able to see new possibilities and how they fit into the library, e.g. new technologies, convergence with other sectors such as education, the ability to anticipate users need and recognize obsolescence of a service as well as an information bearing resource before the users do.

- **Challenging the Process:** This third practice of transformational leadership has to do with courage. The courage rely solely on professional instincts not knowing exactly how things will turn out. To be able to challenge the process, a leader needs confidence and optimism. According to Stark (2012) “an organization cannot rise above its level of leadership.” This means that leaders’ attitudes set the tone for those who follow them. People will not follow leaders who lack confidence in their ability to survive a negative situation and improve the condition of the organization. Library employees want to work for a leader who truly believes that the tomorrow of the library where they work is going to be even better than it was today, and that the leader is taking the necessary actions to turn those beliefs into a reality.
- **Enabling Others to Act:** This has to do with team work, corporation among colleagues and co-library staff. Enabling others is making it possible for others to do good work through development, and trust. Enabling others also involves encouraging people to work together in learning communities. Transformational leaders create situations which lead to organizational and personal growth. They give room for other to make their own decisions and assume responsibility for their decisions. As observed by Kouzes and Posner (2007) “Leadership is not a solo act; it is a team effort.” In a world that is doing more with less, competitive strategies lose to collaborative strategies. A transformational

leader fosters collaborations by creating a climate of trust, facilitating positive interdependence and supporting interpersonal relations. He also dispels and discourages a climate of gossip and envy by verifying all information before acting on them. The transformational librarian encourages creativity and the birthing of new ideas in the library operations by carefully weighing the various opinions, inputs and suggestions from staff in the library, he/she doesn’t see himself as having a monopoly of knowledge. On the other hand, the reward for positive performance could be opportunity for capacity building through further studies, trainings and re-training, attending conferences, etc. The reward system imbedded in the transactional leadership style also encourages healthy competition among staff of the library (Segun-Adeniran, 2015).

- **Encouraging the Heart:** Encouraging the heart is about supporting individuals and groups to achieve their vision (Kouzes & Posner, 2007). Transformational leaders are cheerleaders for the people in their organizations. They acknowledge the contributions of individuals and groups. And they celebrate the victories in the organization. According to Townley (2003) there has been far too little use of this leadership practice in academic libraries. Motivation of library staff can be a good way for a transformational leader to encourage them. Research has conclusively shown that engaged employees are more productive and satisfied (Maxwell, 2013). A transformational leader encourages his followers to take part in job training exercises. The followers then become well equipped and have the necessary skills and abilities to do their jobs while helping the library achieve its vision. Transformational leadership also includes looking for leadership potential in others. By developing leadership skills the

transformational librarian creates an environment where other librarians can continue success in the long term, which is a true measure of any grate leader.

AREAS TRANSFORMATIONAL LEADERSHIP CAN APPLY IN ACADEMIC LIBRARIES IN NIGERIA

While decrying the fact that transformational leadership has not been widely adopted and practiced in libraries, there are potentials for academic libraries in Nigeria as their leaders seek to create increased relevance within the life of the academic community. It is perhaps in this regard that it is eminent to point out the areas that Nigerian academic library leaders need to focus more in the transformational leadership discuss:

- **Vision:** Transformational leadership is a vision propelled type of leadership. More than anything else, academic libraries in Nigeria require visionary and innovation driven leaders. A primary part of a transformational leader's vision for the academic library must be focused on identifying renovations that would encourage better use of the library. According to Bill Gates's quote on transformation –“you need to see your service is obsolete before the customer does. The only companies that succeed will be those who obsolete their own products before someone else does.” This means that library leaders must have the ability to anticipate the library users need and recognize obsolescence before the users do. Also, the ability to recognize that something new is needed instead of more refinement of what exists already. Library leaders don't have to wait for users to stop using library services or for them to register complaints to know that there is need to improve services.
- **Courage:** This is one quality that academic libraries in Nigeria desperately need in their leaders. It is of little wonder that.” Matthews (2002) stated that “leadership of tomorrow's academic libraries is not for the faint of heart. The library leader must have the courage to face and convince management of Nigerian tertiary institutions of issues pertaining to innovations in the library. In the face of the globalization, academic libraries in Nigeria are confronted with a unique set of problems. Like the financial resources with which to meet increased demands for access to both print and electronic resources, library staffs requirements of a new set of skills that will enable them survive and practice in the new digitalized collections. There is also

Academic libraries need transformational leaders whose confidence and optimism will not deprive them from entertaining new possibilities and suggestions from others. For instance, at the Z. Smith Reynolds Library at Wake Forest University, Tedford, and Lock (2012) stated that the library leader wanted to keep the library open for twenty-four hours a day during exam week (a practice initiated at the Undergraduate Library at Wayne State some years before), which required additional staffing and shifting staff hours. First, the transformational library leader modeled the way, volunteering to stay extra hours and issuing a call for four other volunteers to stay overnight. Thirty people stepped forward. Such an overwhelming response epitomized the alignment between change and the longstanding service culture of the organization and how Sutton's modeling the requested behavior affected the ability to make the change. Again, “Wake the Library” was repeated for the exam weeks of Fall 2006 and Spring 2007, providing the necessary data to recommend a permanent change to remain open twenty-four hours a day for five days a week.

the issue of the acquisition of electronic resources, along with its attendant problem of adequate bandwidth and the cost of keeping technology current. Again, there is the constant reminder of the problem of inadequate power supply without which none of the a 21st century academic library services can succeed.

According to Okiy (2010), of all the different types of libraries in Nigeria, only university libraries have a clearly defined policy of funding, because the National Universities Commission (NUC) has a statutory provision that at least 10% of the recurrent annual budget of the parent institutions should be allocated to the library. It is however sad to note that most University libraries still face the challenge of financial constraint. It is regrettable that such monies are not forthcoming as most university administrations tend to flout or fail to comply with this provision (Fabunmi, 2004; Okiy, 2010).

There seems to be less confidence on the part of library leaders about stirring things up with their parent organizations, this can result in the postponement of timely transformation or a loss of opportunity that puts the whole academic library at the risk of losing its significance in the academic community. Hence, transformational leaders require the guts to demand for what belongs to the library in the first place, which might involve being adamant and refusing to back down.

- **Communication:** The ability to communicate well is very germane to fruitful relationship and success of a leader. Giving relevant and timely information as well as the expectation of feedbacks from stakeholders are desirable (Fadehan & Oluwasanmi, 2010). One of the most important areas where communication may be affecting academic libraries in Nigeria is between the library and the campus community. Most often the campus community has little or no idea of the library's value in

their educational mission, the services offered and the overall support the academic library can provide to them. By demonstrating and communicating value exercises, transformational library leaders can influence the academic community and the parent organization.

Furthermore, the leader will have to recognize and tap into the problem-solving skills of all staff as well as external experts. This can only be pulled off with good communication skills of the library leader. Effective channels of communication such as holding of sectional meetings or talking with other library staff on a one-on-one basis can be employed. Also, the library leader can harness positive outcomes and performance indicators, and with the data collected can clearly articulate the value of the library to the parent institution. In addition, Sweeney (1994) stressed the fact that the transformational library leader must also play the role of a superb communicator who listens well, speaks and writes well. This leader must articulate concepts that are new and foreign to staff, users, administrators, and others in an appealing and rational manner.

NURTURING FUTURE TRANSFORMATIONAL LEADERS IN LIBRARIANSHIP

It is a well-known fact that you cannot give what you do not have, similarly, it is believed that the most successful leader will have developed his or her leadership skills over a long period of time. Hence, for future transformational library leaders to develop the following suggestions could be considered:

- **The Leader as a Mentor:** Perhaps this is the most important means of nurturing future transformation librarians. There are a lot of experienced and seasoned library leaders in Nigerian academic librar-

ies, however, not many of them practice the transformational leadership style. The few who do and have results to show must identify talented new individuals and even older library leaders and help them develop the qualities necessary for success in an environment of ambiguity, and providing productive learning experiences that build on existing strengths and minimize weaknesses.

- **Internship Programs:** The two major bodies of librarianship in Nigeria also has a part to play in the issue of nurturing future transformational leaders for the libraries. The Nigerian Library Association (NLA) and the Librarians Registration Council of Nigeria (LRCN) can organize leadership apprenticeship programs. Programs that can afford younger librarians opportunities to periods of internship with a particular transformational leader so that they can learn informally about what it actually takes to lead the modern day academic library. The NLA or LRCN can also use this sought of program to inspire younger librarians who excel at national or local competitions organized by the bodies. These governing bodies can also confer awards on libraries and library leaders in recognition of outstanding leadership in academic libraries.
- **Formal Transformation Leadership Training at Library Schools:** Here, in the library schools more than anywhere else is the awareness that the required skills to excel by students who are potential library leaders are constantly changing. This means the skills needed by library leaders are changing. Bass (1990) posits that transformational leadership skills can be learned, and plenty of methodologies to can also be found in literature. Curriculums of library schools should be upgraded especially at the postgraduate levels on the ideal qualities required of library leaders,

knowledge of these qualities can inform how well LIS programs prepare the future leaders of the field. Examples of such leadership courses on Leadership Programs developed by LIS schools and libraries to foster leadership skills in MLIS students and practicing librarians in the U.S. are The Library Leadership and Management's (LLAMA) Development division has created an annual seminar called the Leadership Development Seminar, which is held each year during ALA's Midwinter meeting (Considine, 2011).

- **Building a Knowledge Transmission Infrastructure:** The Nigerian society is well known for its oral traditions. Records of transformational solutions to innovative problems can be kept by the library leader for future reference and for upcoming leaders. Thus the use of more unobtrusive methods can be applied where leaders should be observed in their work place, while performing their day-to-day activities, when the knowledge needed to learn and perform tasks within a complicated process is written, updated dynamically, and shared using the technology infrastructure, then the knowledge needed to perform a job is not only in someone's head, but also well documented and stored.
- **Leader Followers / Followers Leader:** Today's leaders were once followers, similarly, today's follower can be tomorrow's leader. Academic librarians in Nigeria must be ready to invest in themselves. Librarians should put more efforts into understudying their leaders so that they too can be better prepared to take over the mantle of leadership in their libraries.

CONCLUSION

It is obvious from the foregoing that the traditional mission of academic libraries will never change,

however the means of achieving this mission must change. In order for Nigerian academic libraries to continue to build and sustain their position as active, trusted, and responsive agents of change in the academic community, they need transformational leaders who will move their libraries forward with vision, courage, purpose, and determination. The academic library of tomorrow needs a leader who is will not be afraid of the changes that are eminent in their libraries, but will encourage the followers to think, speak and act. A leader who is willing to take risks with innovations in information technologies in order to better information service delivery in libraries and one who will be bold to face management with convincing arguments in favor of the academic library, the librarians working in it and everything it stands for.

FUTURE RESEARCH DIRECTIONS

Although this chapter has attempted to shed light on important aspects of innovation in academic libraries, transformational leadership practices and the areas that Nigerian academic library leaders need to focus on, a major limitation is that the literature did not cover a qualitative or quantitative perspective of transformational leadership practices in Nigerian academic libraries. Further research on transformational leadership in academic libraries should cover areas such as the important leadership profiles, including personal competencies, which might facilitate or inhibit innovation in academic libraries. Empirical studies should be conducted on the effects of transformational leadership on staff productivity in academic libraries in Nigeria, as well as a study on the direct effect this leadership style will have the innovativeness of academic libraries given the unique circumstances these libraries face as regards availability of funds.

REFERENCES

- Afolabi, M. O. (2014, June). *The library: The academic life wire of a university system*. Lecture delivered at the Federal University Oye-Ekiti University main auditorium, Oye-Ekiti. Retrieved from [Http://repository.fuoye.edu.ng/bitstream/123456789/192/1/the_library_the_academic_life_wire.pdf](http://repository.fuoye.edu.ng/bitstream/123456789/192/1/the_library_the_academic_life_wire.pdf)
- Awan, M. R., Mahmood, K., & Idrees, H. (2014). Leadership style, culture and commitment: an analytical study of University Libraries in Pakistan. *Library Philosophy and Practice*, 1-7. Retrieved from <http://digitalcommons.unl.edu/libphilprac/1269/>
- Bass, B. M. (1985). *Leadership and performance beyond expectations*. New York: Free Press.
- Bass, B. M. (1990). From transactional to transformational leadership: Learning to share the vision. *Organizational Dynamics*, 18(3), 19-31. doi:10.1016/0090-2616(90)90061-S
- Bazillion, R. J. (2001). Academic libraries in the digital revolution: Libraries in the midst of revolution need new ways of thinking about their mission. *Educause Quarterly Viewpoint*, (1), 51-55. Retrieved from <https://net.educause.edu/ir/library/pdf/EQM0119.pdf>
- Burns, J. M. (1978). *Leadership*. New York: Harper & Row.
- Considine, S. (2011). Leadership development seminar. *Library Leadership & Management*, 25(4), 1-2.
- Davis, K., & Newstorm, J. W. (1985). *Human behavior at work: Organizational behavior* (7th ed.). New York: McGraw-Hill Book Company.
- Dawson, D., & Lewis, S. (2007). *Transformational leadership and its implications for the library profession*. Retrieved from www.lianza.org.nz/sites/.../dawson_lewis_transformational_leadership.pdf

- Fabunmi, B. A. (2004). Planning the university libraries for effective customer services in Nigeria. In E. C. Madu (Ed.), *Technology for information management and service: Modern libraries and Information centers on developing countries* (pp. 147–158). Ibadan: Evi-Coleman.
- Fadehan, O. A., & Oluwasanmi, H. (2010). Leadership in the library and Information environment. *Ozean Journal of Social Sciences*, 3(1), 91-100. Retrieved from http://www.ozelacademy.com/OJSS_v3n1_8.pdf
- Fatokun, J. O., Salaam, M. O., & Ajegbomogun, F. O. (2010). The influence of leadership style on the performance of subordinates in Nigerian libraries. *Library Philosophy and Practice*, 1-13. Retrieved from <http://digitalcommons.unl.edu/libphilprac/126>
- Geib, P., & Swenson, J. (2013). China: Transformational leadership for policy and product innovation. *Advances in Management*, 6(5), 3–10.
- Germano, M. A. (2010). Leadership style and organizational impact. *Library Work Life*. Retrieved from <http://ala-apa.org/newsletter/2010/06/08/spotlight>
- Germano, M. A. (2011). Library leadership that creates and sustains innovation. *Library Leadership and Management*, 25(1), 1–16.
- Hinchliffe, L. J. (2011). Understanding, demonstrating and communicating value: the leadership and management challenge—vision 2020: Innovative policies, services and tools. *Management and Marketing Section with Academic and Research Libraries*. Retrieved from <http://conference.ifla.org/ifla77>
- Johannsen, M. (2014). *Twenty leadership styles*. Retrieved from <https://www.legacee.com/types-of-leadership-styles/>
- Kouzes, J. M., & Posner, B. Z. (2007). *The Leadership challenge* (4th ed.). San Francisco: Jossey-Bass.
- Lippincott, J. K. (2005). Net generation students and libraries. In D. G. Oblinger & J. L. Oblinger (eds.), *Educating the net generation*. Retrieved from <http://educause.edu/ir/library/pdf/pub7101m.pdf>
- Martin, J. (2015). Leadership in academic libraries: Exploratory research on the use of transformational and transactional leadership styles. *Leadership in Academic Libraries*, 391-397.
- Matthews, G. (2002). The user's perspective: A personal view. In M. Melling & J. Little (Eds.), *Building a Successful Customer-Service Culture: A Guide for Library and Information Managers*. London: Facet Publishing.
- Maxwell, J. (2013). *What is transformational leadership?* The John Maxwell Team. Retrieved from jmsalesteam.s3.amazonaws.com/JoinJohnLive-Worksheet.pdf
- Obiwuru, T. C., Okwu, A. T., Akpa, V. O., & Nwankwere, I. A. (2011). Effects of leadership style on organizational performance: A survey of selected small Scale Enterprises in Ikosi-Ketu Council development area of Lagos State, Nigeria. *Australian Journal of Business and Management Research*, 1(7), 100-111. Retrieved from http://www.ajbmr.com/articlepdf/ajbmr_17_16i1n7a11.pdf
- Odumeru, J. A., & Ogbonna, I. G. (2013). Transformational vs. transactional leadership theories: Evidence in literature. *International Review of Management and Business Research*, 2(2), 335–361. Retrieved from www.irmbrjournal.com
- Ogbah, E. L. (2013). Leadership Style and Organizational Commitment of Workers in Some Selected Academic Libraries in Delta State. *International Journal of Academic Research in Business and Social Sciences*, 3(7), 110–118. doi:10.6007/IJARBSS/v3-i7/13

- Okiy, R. B. (2010). Towards accelerated development of academic library services in Nigeria for national development in the 21st Century. *Library Philosophy and Practice*. Retrieved from <http://unllib.unl.edu/LPP/>
- Rendon, F. (2014). *How innovation and technology are shaping libraries of today*. Retrieved from http://www.huffingtonpost.com/frankie-rendon/how-innovation-and-techno_b_5244601.html
- Riggo, R. E. (2012). *What is charisma and charismatic leadership? Is charisma born or made? What makes leaders charismatic?* Retrieved from <https://www.psychologytoday.com/blog/cutting-edge-leadership/201210/what-is-charisma-and-charismatic-leadership>
- Schwartz, C. A. (1997). Restructuring academic libraries: adjusting to technological change. In C. A. Schwartz (Ed.), *Restructuring academic libraries: organizational development in the wake of technological change* (pp. 1–30). Chicago: Association of College and Research Libraries.
- Segun-Adeniran, C. D. (2015). Leadership styles and job productivity of university library staff: interrogating the Nexus. *Library Philosophy and Practice*. Retrieved from <http://digitalcommons.unl.edu/libphilprac/1269>
- Shamir, B., House, R. J., & Arthur, M. B. (1993). The motivational effects of charismatic leadership: A self-concept based theory. *Organization Science*, 4(4), 577–594. doi:10.1287/orsc.4.4.577
- Simola, S., Barling, J., & Turner, N. (2012). Transformational leadership and leaders mode of care reasoning. *Journal of Business Ethics*, 108(2), 229–237. doi:10.1007/s10551-011-1080-x
- Stark, P. B. (2012). *The secret to inspiring others? Optimism*. Retrieved from <http://www.peterstark.com/inspiring-employees-optimism/>
- Storey, C. (1995). The impact of information technology on management and organization: the case of Hong Kong Polytechnic University Library. *Library Management*, 16(2), 22–28. Retrieved from <http://www.emeraldinsight.com/doi/abs/10.1108/01435129510081442>
- Tedford, R., & Lock, M. B. (2012). *How transformational leadership translates into recognized excellence in academic libraries*. Retrieved from https://wakespace.lib.wfu.edu/.../Chapter17_Tedford_Corbett_and_Lock
- Townley, C. T. (2003). *The innovation challenge: Transformational leadership in technological university libraries*. Retrieved from docs.lib.purdue.edu/cgi/viewcontent.cgi?filename=0&article
- Urban Libraries Council. (2012). *Leadership brief: Library leaders owning leadership*. Retrieved from www.urbanlibraries.org
- Weiner, S. G. (2003). Resistance to change in libraries: Application of communication theories. *Libraries and the Academy*, 3(1), 69–78. Retrieved from http://web.simmons.edu/~weigle/INNOVATION/Weiner_Resistance%20to%20Change.pdf
- Yukl, G. A. (1998). *Leadership in organizations* (4th ed.). Upper Saddle River, NJ: Prentice-Hall.

KEY TERMS AND DEFINITIONS

Academic Libraries: These include all libraries found in universities in Nigeria.

Library Leaders: University librarians in charge of academic libraries.

Transformational Leadership: The act of getting extraordinary things done for the academic library.

Category M

Marketing

Are Social Marketing Investments Used as a Tool for Voluntary Reporting or Disclosure?

M**Tugba Ucma Uysal***Mugla Sitki Kocman University, Turkey***Ganite Kurt***Gazi University, Turkey***Ali Naci Karabulut***Mugla Sitki Kocman University, Turkey*

INTRODUCTION

The need to respond to conditions that change with the effect of social transformation in line with the globalization phenomenon and developments in information technologies, have resulted with the development of new corporate behavioral patterns of enterprises. It is possible to observe these behavioral patterns both in terms of reporting activities and also in terms of marketing activities. The sustainability and corporate governance concepts that are being applied in businesses in an increasing manner stand for the cumulative efforts for accountability or a convergence. Because, as long as sustainable reporting of businesses are questioned, the meaning and nature of accountability are manifested. This in return is associated with the sense of transparency of businesses (Kolk, 2008) and even stands for the voluntary reporting of non-financial information.

On the other hand, if the situation is assessed in terms of marketing activities, it is possible to observe a similar evaluation to the assessment made in terms of reporting. In other words, the accountability of the business concept is once more in front of our eyes. Because, a number of issues such as air pollution, deterioration of public health, wasting of natural resources, economic stagnancy, inflation, unbalanced income distribution etc. that

are experienced due to increasing population at the global level and volume of production result with the questioning of businesses and modern marketing approaches by the people (Clifford, 1998; Drucker, 1974; Doğaner, 2011; Torlak, 2006; Yükselen, 2014). Furthermore, people have become more aware of illegal and unethical behavior in financial and legal presentations (Salimbene & Ferrera, 1996) believing that both stakeholders and non-stakeholders of organizations have the right of information on specific social agenda (Miles & Munilla, 2004).

There are a number of definitions for what sustainability reports stand for of which the importance is being emphasized at many points. But to this date, there is no globally accepted definition of corporate social responsibility (CSR) or sustainability reporting. However, a common European understanding of what CSR means has emerged on the basis of the European Commission definition of CSR as “*a concept whereby companies integrate social and environmental concerns in their business operations and in their interaction with their stakeholders on a voluntary basis*” (The Council and the European economic and social committee 2006, Dilling, 2010). Or these reports are being expressed as follows in the general sense and their relation with CSR concept is being underlined in the definition.

Sustainability reporting can be considered as synonymous with other terms for non-financial reporting; triple bottom line reporting, CSR reporting, and more. It is also an intrinsic element of integrated reporting; a more recent development that combines the analysis of financial and non-financial performance. (<https://www.globalreporting.org>)

Sustainability reporting encompasses a process that gives way to the forming of a relation between the enterprise and the society (Porter & Kramer, 2006). Voluntary reporting and disclosure are thus being used as a reporting process that reduces the amount of asymmetrical information between the business and stakeholders (Verrechia, 2001) and is being supported by means of a set of social sanctions. Social problems that emerge in the global scale have given way to the development of new social standards for the sustainability of organizations (Drucker, 1974: 337-338; Doğaner, 2011: 27; Torlak & Barca, 2012: 21; Wasik, 1996). The business community has to yield reasonable profits for stockholders and owners on one hand, while supplying decent products at reasonable prices, honest advertisements, fair treatments and profitable opportunities to employees, customers, suppliers and competitors with a strong sense of responsibility to the communities it inhabits and serves on the other hand (Gini, 2005; Pava, 1998). A distinct marketing discipline has emerged since the early 1970's which is called as social marketing which is more comprehensive than modern marketing and which refers primarily to efforts focused on this kind of social issues (Lee & Kotler, 2011; Doğaner, 2011). Therefore social marketers are primarily concerned with social issues and financial consequences of their actions only to the extent that they want to insure their organization's financial viability (Bloom & Novelli, 1981).

Investments in social marketing which were deemed as implementations that bring forth additional costs to organizations at the outset can actually be used as efficient tools in terms of voluntary reporting and disclosure of organiza-

tions. In this context, this chapter aims to find out if social marketing investments are used for voluntary reporting or disclosure by examining the websites of the companies which are ranking in the Turkey ISE 100 Index with Structuring Content Analysis method. Investigating the only way of voluntary reporting is the limitation of this study even though websites are the primarily disclosure instruments for the companies.

THEORETICAL BACKGROUND

According to the accounting literature, it is observed that these studies coincide with early 1990's (Lamberton, 2005). The measurement of social effects of these costs in particular and the reflection of these effects to reporting constitute the starting point of social accounting implementations (Aras & Crowther, 2009). Taking a look at the studies within the scope of said literature, it is observed that the relation between corporate social responsibility and financial performance is being focused upon in particular. Global Reporting Initiative (GRI) was established in 1999 to publish CSR reporting standards which is the sole system aiming to fill the gap in related literature and a model was presented to the US financial reporting system. GRI provides a standardized approach as an aid to reporting of sustainable information. GRI has also turned into a framework with the highest level of acknowledgment worldwide for voluntary reporting.

GRI is effective for the reporting of corporate social responsibility projects or voluntary reporting activities of businesses besides marketing strategies. One of the reasons for this is the fact that CSR is an important communication tool for stakeholders to have access to necessary information. Another reason is listed as the influence of CSR on the reputation management and brand protection activities of enterprises (Nikolaeva & Bicho, 2011). Important information is provided for social or environmental disclosure from variants obtained from markets. For instance, it has

been determined upon studies conducted by Gray et. Al. (2001) that corporate characteristics of businesses have directly affected the social disclosure levels of these companies. Kolk (2003) had at first scrutinized the sustainability reporting trends of companies listed under Fortune Global 250 and had later on made research on the reporting levels of multi-national companies within the framework of accountability, corporate governance and understanding of sustainability in 2008. Hedberg and Malmberg (2003) on the other hand made research checking if the GRI and CSR reports made any contribution to the reliability of Swedish companies. Gray (2006) made a research checking if sustainability reporting of companies had any effect on the development of corporate behavior. Brown et. al. (2009) studied the effects of the sustainability reports prepared within the framework of the GRI on the institutionalization of companies. It has been found if sufficient information can be obtained on corporate sustainability levels when corporate sustainability reporting is made within the framework of GRI upon studies conducted by Isaksson and Steimle (2009). Also, a critical analysis was made in studies conducted by Fonseca (2010) checking if sustainability reporting made by companies in the mining sector has contributed to the reliability of companies besides external trust.

According to the marketing literature, several studies can be traced since the 1960's on issues that are named as social marketing in present day but the origins of the term social marketing can be traced to Kotler and Zaltman's classic 1971 article titled "Social Marketing: An Approach to Planned Social Change" (Andreasen, A.R., 1994). In this article, authors were examining the applicability of marketing concepts and techniques to the promotion of social objectives and they were defining the social marketing as: "the design, implementation, and control of programs calculated to influence the acceptability of social ideas and involving considerations of product planning, pricing, communication, distribution, and marketing research". Kotler, et. al (2002), defined

social marketing again as "the use of marketing principles and techniques to influence a target audience to voluntarily accept, reject, modify, or abandon a behavior for the benefit of individuals, groups, or society as a whole".

The first argumentation on the social marketing approach in terms of sustainability was started by marketing scholars in the 1970's with Ecological Marketing (Henion & Kinnear, 1976) and The Ecologically Concerned Consumer (Kardash, 1976). Fisk (1973) made a contribution to the approach of Peattie and Peattie (2009) with the "Theory of Responsible Consumption". McKenzie and Mahr (2000) make the proposition of developing social marketing programs based on the society to achieve a sustainable set of behavior for the sake of the organization. Andreasen (2002) explains how the social marketing approach can be used as an active marketing tool according to social changes in the target audience. Peattie and Peattie (2009) focuses on the fact that the social marketing approach can be used as a tool in reducing consumption levels targeting a sustainable life.

MAIN FOCUS OF THE ARTICLE

Only one previous study (Kurt & Ucma Uysal, 2014) examined sustainability reporting for voluntary reporting or disclosure in Turkey. It found that sustainability reporting can be best framework for voluntary reporting and disclosure in Turkey. The aim of the present study is to use the methodologies of previous studies (Brown et al., 2009; Kolk, 2003; Kolk, 2004; Kolk, 2008; Simnett et al., 2009; Kurt & Ucma Uysal, 2014) to explore social marketing investments can be used as a tool for voluntary reporting or disclosure in Turkish companies in capital markets. The voluntary disclosure levels of companies listed on the ISE 100 index are thus being clarified revealing the importance attributed by businesses to sustainable growth. At this point, emphasis is made on social marketing investments which are perceived by businesses as net increases in costs while evaluation is made on social goals

set forth by companies in terms of sustainability. The Structuring Content Analysis which was used by Isaksson and Steimle (2009) has been utilized to scrutinize sustainability reporting by Turkish companies which is one of the non-financial reporting forms. Theoretical analyses could be made by means of the type of analysis, regarding businesses that publish sustainability reports in diverse aspects and those which do not publish sustainability reports but that provide information on web pages under such title. Main assessment criteria which we have selected in line with Kolk's studies (2008) on which we have focused during the analysis and criteria in relation with social marketing investments (Steurer et al., 2008; Lee et al., 2010; Revelli & Viviani, 2015) are listed as follows:

- Separate sustainability report,
- Social, ethics and environmental reports or politics,
- Financial reports with integrated sustainability information,
- Code of ethics,
- Separate sustainability department/unit,
- Separate section on corporate governance in the report,
- External verification/assurance of sustainability reports or sustainability information,
- Sustainability content on web page and in CEO Statement,
- Socially responsible investing and social marketing and social goal.

Separate Sustainability Report

The scope of the concept of reporting has become quite extensive in the present day. Businesses have attributed more importance to social, ethical and environmental issues upon pooling of sustainability and corporate governance concepts within a single report and thus giving start to such reporting means, while the understanding of sustainability has been more inclusive also encompassing control and accountability issues

(Kolk, 2008). It has been observed that only 31 companies among those listed in the ISE 100 index have published reports titled as sustainability reports besides separate financial reports. Having taken a look at the reports of the 31 companies that have published such reports, it has been observed that each report includes disclosures on the sustainability approach of the company and the Board of Directors and the importance of such an understanding has been emphasized in terms of sustainable growth. The reports also include statements regarding the importance of value attributed to customers besides corporate governance, environmental and social responsibility projects and ethical values, and value drivers. Also the web pages of companies that have published separate reports include disclosures on issues that are listed in the sustainability report under the heading of corporate social responsibility. The other 69 companies that are listed in the ISE 100 index as of the third quarter of 2015 do not publish separate sustainability reports. It is observed that a total of 17 companies listed in the ISE 100 index published separate sustainability reports at the end of 2013, upon evaluation made together with the results obtained in previous studies (Kurt & Ucma Uysal, 2014). An increase has been observed in the number of companies that have published separate sustainability reports upon studies conducted 1.5 years later. These findings display that there has been an improvement in the reporting level of non-financial information in Turkey which develops within the framework of an understanding of corporate governance and which is expressed as integrated reporting.

Social, Ethics, and Environmental Reports/Politics

There has been an increase in recent years in the number of companies that make reporting on their environmental and social performance. It has been revealed upon research made by KPMG (2002) that sustainability reports are being prepared to encompass economic, social, ethical and

environmental issues (O'Dwyer & Owen, 2005). These reports generally encompass projects and policies implemented by businesses within the scope of their trust indices. Taking a look at the matter for companies in the ISE 100 index, it is observed that reports published by 70 companies that make separate reporting on sustainability as mentioned above, include information on said issues. Under these circumstances, it is possible to claim that 70 percent of businesses listed in the ISE 100 index either publish their environmental and social reports or policies which are deemed as non-financial reports, in their web pages or include such information in sustainability reports which are published separately.

Financial Reports with Integrated Sustainability Information

Integrated reporting stands for reporting made by companies also defining their human, intellectual, natural and social assets in addition to their traditional financial assets and their production potential. Legal and structural arrangements on integrated reporting are being put into life in the US in particular. For instance, Sustainable Accounting Standards Board (SASB) brings forth related arrangements on integrated reporting and disclosure for US companies in the industrial sector by means of new standards set forth (Investor Responsibility Research Center Institute Report, 2013). The main purpose of the presentation of financial information in studies conducted for the standardization of financial reporting worldwide is to provide necessary information to facilitate decision-making processes of all internal and external users of information on the related matter. It is being observed that a number of multi-national companies provide information on their social and environmental performances and their understanding of sustainability in the present day, to cover the needs of information of their internal and external stakeholders (Ballou et. al., 2006; Simnett et. al., 2009). Businesses are

disclosing their activities within the framework of an understanding of sustainability (Kolk, 2004) to display their transparency in particular and to build trust in their financial performance. Annual activity reports of businesses in the ISE 100 index have been accessed primarily to conduct such a study and information has been sought in activity reports on sustainability and perpetuation of the company. Such information has been found in 74 businesses within the scope of studies. It is possible to claim that serious developments have been experienced in the integrated reporting approaches of businesses compared to results obtained in previous studies (Kurt & Ucma Uysal, 2014). This is being explained by means of the development of capital markets in particular and stands as a proof that companies that are traded at the ISE are seriously prepared for such issues in the national scale.

Code of Ethics

It is observed that the presence of ethical codes that are generalized following the auditing scandals in particular and which are set forth within the own bodies of organizations, are being presented as a part of the socialization phenomenon in which businesses display ethical behavior, when assessed in terms of theory and practice (Somers, 2001). For this reason, businesses publish a set of ethical codes somehow within their own bodies and provide information on their ethical behavior as an organization for internal and external information seekers. It has been determined that 70 percent of companies within the scope of ISE 100 index have ethical codes published in their web pages. The set of ethical codes set forth by businesses are also of importance in terms of their social marketing approaches. Because, there are a number studies (Eagle, 2008; Eagle et al., 2013) that focus upon the importance of ethical codes defined by businesses for a sound understanding of social marketing.

Separate Sustainability Department/Unit

It has been studied to check if businesses have a separate sustainability department within the framework of their understanding of sustainability. It has been determined that none of the companies have a sustainability department which has been formed separately within the body of the organization. Kolk (2008) determined upon studies conducted on multinational companies that there are separate sustainability departments or units in a part of multinational companies (Ford, Shell, BP, etc.). However, such a department does not exist within the bodies of companies listed in the ISE 100 index no matter if they publish sustainability reports or not. However, it has been found out that 8 companies out of those in question have an organizational unit similar to a social responsibility committee. Such data could only be accessed through organizational charts that are entered in the web pages of the businesses in question.

Separate Section on Corporate Governance in the Report

Checking if there is a separate heading on corporate governance within the reports published by companies within the scope of said studies that publish sustainability reports separately, has been included within the scope of studies in question. It has been determined that detailed sections regarding a corporate governance approach are included in the report by 30 companies that publish sustainability reports. It is observed that corporate governance and corporate social responsibility issues are taken up as a whole in the sustainability reports of 18 companies. Businesses have set forth sustainability strategies within the framework of a corporate governance approach in terms of relations with stakeholders in particular and have expressed these strategies on basis of corporate social responsibility and accountability. Integrated reporting develops with an understanding of corporate governance for each business as a bottom

line (Smith, 2014). However, such a relation is underlined only by 30 businesses within the scope of said studies, whereas no mention is being made to such a relation in other cases.

External Verification/Assurance of Sustainability Reports or Sustainability Information

We have expressed in previous sections that reporting of social, environmental and economic performances of businesses under the heading of sustainability or the presentation of these over their web pages are being considered as voluntary reporting tools within the framework of Global Reporting Initiative. Building of external trust by companies by means of non-financial reporting regarding information users and stakeholders or verification of policies in selected issues are of utmost importance in terms of the reliability of information (Fonseca, 2010; Matisoff et. al., 2013). At this point, companies make use of various tools such as the Carbon Disclosure Project (CDP), and quality and occupational safety certificates serving to external verification and assurance. It is observed that 45 percent of companies within the scope of said studies make use of external verification and assurance in issues of sustainability reporting and sustainability in the general sense.

Sustainability Content on Web Page and in CEO Statement

It has been checked within the scope of studies if there are any disclosures made in web pages of companies and the reports published by the Board of Directors on sustainability issues. According to findings, there is content regarding sustainability issues in the web pages of 60 percent of businesses. Also, according to findings and upon a contents analysis of speeches delivered by CEO's in reports published by the Board of Directors and annual activity reports and it has been determined that statements on topics such as sustainable development, sustainable growth,

the environment and sustainable development attributing importance to stakeholders have been the most frequent expressions during the speeches delivered by CEO's of 63 percent of companies.

Socially Responsible Investing, Social Marketing, and Social Goal

The relation between the corporate social responsibility and financial performance of businesses is generally being discussed in terms of socially responsible investing (SRI) (Hill et al., 2007; Matloff and Chaillou, 2013). SRI is being described as follows by Ciocchetti (2007) in the most general sense:

SRI, refers to financial contributions into investment vehicles designed to combine the traditional investment philosophy favoring profit-maximization with a values-based component seeking non-financial benefits. Such non-financial benefits are often referred to as social returns.

As will be seen from the definition, SRI is a concept which results with costs in the short run as a values-based component to seek non-financial benefits but which positively effects the long term profitability of the company and its perpetuation. This in return constitutes a reference for social marketing which is more extensive than the traditional marketing approach of the business. Because, a number of issues such as air pollution, deterioration of public health, wasting of natural resources, economic stagnancy, inflation, unbalanced income distribution etc. that are experienced due to increasing population at the global level and volume of production result with the questioning of businesses and modern marketing approaches by the people (Drucker, 1974; Doğaner, 2011; Torlak, 2006; Yükselen, 2014). Furthermore, people have become more aware of illegal and unethical behavior in financial and legal presentations (Salimbene & Ferrera, 1996) believing that both stakeholders and non-stakeholders of organizations have the right of

information on specific social agenda (Miles & Munilla, 2004) SRI and social marketing targets emerge as tools which are voluntarily executed by businesses to express themselves to the public in a better way by means of such an interactive relation. According to study results, 58 percent of businesses listed in the ISE 100 index disclose their social responsibility investments and social marketing activities on their web pages. Taking a look at social targets set forth by businesses at this point, it is observed that the environment, education, culture & arts, sports, health and safety issues top the list in a descending manner.

FUTURE RESEARCH DIRECTIONS

Marketing and reporting activities which have become more complicated upon developments in the business world are actually outstanding phenomena that affect decision-making processes for those that implement information. The reporting of non-financial information in particular without any legal sanctions and its publishing each year in accordance with social expectations also affect the long-term profitability and brand management processes of organizations in a direct manner.

In this regard, it is obvious that, in the near future, companies will become more transparent structures that will be more willing to give detailed information on their activities. It is without a doubt that the information technologies will be the principal tools to fulfill companies' such needs. It is expected that this chapter will contribute to the scholarly literature researching on the use of information technologies as a voluntary means of communication between companies and the public, thus, constitute a starting point for future researchers.

CONCLUSION

Voluntary reporting and disclosure concepts that constitute the basis of corporate social responsibil-

ity are of utmost importance for businesses wishing to catch a competitive advantage in present day conditions and particularly within the global economic system. Such an approach shapes the marketing approach of quite a number of businesses besides including the reporting of non-financial information. Main reasons for this can be listed as the fact that sustainability reports provide answers to diverse information users besides being used as a marketing communication tool, its contribution to brand management and its effects on the development of market based intellectual property (Nikolaeva & Bicho, 2011). Each of the reasons expressed above results with the avoiding of traditional marketing approaches of businesses giving way to attributing importance to social marketing and social responsibility investments which are expressed as an understanding of modern marketing. Taking pioneering studies as basis, it is being checked if businesses listed in the ISE 100 index as disclosed in the last quarter of 2015 make use of their social responsibility investments as a tool in terms of voluntary reporting.

Websites of the companies which are ranking in the Turkey ISE 100 Index had selected as research case and examined with Structuring Content Analysis method for this purpose. Under this limitation, it is worth pointing out that businesses execute these activities with a voluntary perspective due to the lack of a legal infrastructure in Turkey regarding sustainability reporting in particular. Assessment criteria used in the previous study has been taken up once more for the contents analysis during present studies and it has been observed that each criterion displayed a change compared to the previous year. Besides these, the fact that 58 percent of businesses attribute importance to SRI and social marketing activities and furthermore publish these on their web pages besides sustainability reports is to be mentioned to underline the importance of the whole matter in Turkey.

REFERENCES

- Andreasen, A. R. (1994). Social marketing: Its definition and domain. *Journal of Public Policy & Marketing*, 13(1), 108–114.
- Andreasen, A. R. (2002). Marketing social marketing in the social change marketplace. *Journal of Public Policy & Marketing*, 21(1), 3–13. doi:10.1509/jppm.21.1.3.17602
- Aras, G., & Crowther, D. (2009). Corporate sustainability reporting: A study in disingenuity? *Journal of Business Ethics*, 87(S1), 279–288. doi:10.1007/s10551-008-9806-0
- Ballou, B., Heitger, D. L., & Landes, C. E. (2006). The future of corporate sustainability reporting: A rapidly growing assurance opportunity. *Journal of Accountancy*. Retrieved December 15, 2015, from <http://www.journalofaccountancy.com/NR/exeres/999D3E96-0A53-4030-AFB4-B51FA2113371.htm>
- Bloom, P. N., & Novelli, W. D. (1981). Problems and Challenges in Social Marketing. *Journal of Marketing*, 45(2), 79–88. doi:10.2307/1251667 PMID:12280283
- Brown, H. S., Jong, M., & Levy, D. L. (2009). Building institutions based on information disclosure: Lessons from GRI's sustainability reporting. *Journal of Cleaner Production*, 17(6), 571–580. doi:10.1016/j.jclepro.2008.12.009
- Ciocchetti, C. A. (2007). *Socially responsible investing, from the selected works of Corey A Ciocchetti*. Retrieved December 10, 2015, from <http://works.bepress.com>
- Clifford, M. (1998). Economics, politics, and environmental protection. In M. Clifford (Ed.), *Environmental Crime: Enforcement, Policy, and Social Responsibility* (pp. 57–72). Gaithersburg, MD: Aspen Publication.

- Dilling, P. F. A. (2010). Sustainability reporting in a global context: What are the characteristics of corporations that provide high quality sustainability reports – an empirical analysis. *International Business & Economics Research Journal*, 9(1), 19–30.
- Doğaner, M. (2011). Pazarlama kavramı ve pazarlamanın gelişimi. In Pazarlama İlkeleri (pp. 11–39). İstanbul: Lisans Publication
- Drucker, P. F. (1974). *Management: Tasks, Responsibilities, Practices*. New York: Harper & Row Publication.
- Eagle, L. (2008). Social Marketing Ethics. London: Report for National Social Marketing Centre.
- Eagle, L., Dahl, S., Hill, S., Bird, S., Spotswood, F., & Tapp, A. (2013). Social Marketing. Harlow, UK: Pearson Publication.
- Fisk, G. (1973). Criteria for a theory of responsible consumption. *Journal of Marketing*, 37(2), 24–31. doi:10.2307/1250047
- Fonseca, A. (2010). How credible are mining corporations sustainability reports? A critical analysis of external assurance under the requirements of the international council on mining and metals. *Corporate Social Responsibility and Environmental Management*, 17(6), 355–370. doi:10.1002/csr.230
- Gini, A. (2005). *Case Studies in Business Ethics* (5th ed.). Pearson, Prentice Hall.
- Gray, R. (2006). Does sustainability reporting improve corporate behaviour? Wrong question? Right time? *Accounting and Business Research*, 36(1), 65–88.
- Gray, R. H., Javad, M., Power, D. M., & Sinclair, C. D. (2001). Social and environmental disclosure and corporate characteristics: A research note and extension. *Journal of Business Finance & Accounting*, 28(3), 327–356. doi:10.1111/1468-5957.00376
- GRI Organisation Web Site. (n.d.). Retrieved 13 December 2015, from <https://www.globalreporting.org>
- Hedberg, C. J., & Malmberg, F. (2003). The global reporting initiative and corporate sustainability reporting in Swedish Companies. *Corporate Social Responsibility and Environmental Management*, 10(3), 153–164. doi:10.1002/csr.38
- Henion, K. E., & Kinnear, T. C. (1976). *Ecological marketing*. American Marketing Association's 1st National Workshop on Ecological Marketing, Chicago, IL.
- Hill, R. P., Ainscough, T., Shank, T., & Manulang, T. (2007). Corporate social responsibility and socially responsible investing: A global perspective. *Journal of Business Ethics*, 70(2), 165–174. doi:10.1007/s10551-006-9103-8
- Investor Responsibility Research Center Institute (IRRCI) Report. (2013). *Integrated financial and sustainability reporting in the United States*. Retrieved December 9, 2015 from http://irrcinstitute.org/pdf/FINAL_Integrated_Financial_Sustain_Report_April_2013.pdf
- Isaksson, R., & Steimle, U. (2009). What does GRI reporting tell us about reporting corporate sustainability? *The TQM Journal*, 21(2), 168–181. doi:10.1108/17542730910938155
- Kardash, W. J. (1976). *Corporate responsibility and the quality of life: developing the ecologically concerned consumer*. In Ecological marketing: American Marketing Association's 1st National Workshop on Ecological Marketing, Chicago, IL.
- Kolk, A. (2003). Trends in sustainability reporting by the Fortune Global 250. *Business Strategy and the Environment*, 12(5), 279–291. doi:10.1002/bse.370
- Kolk, A. (2004). A decade of sustainability reporting: Developments and significance. *International Journal of Environment and Sustainable Development*, 3(1), 51–64. doi:10.1504/IJESD.2004.004688

- Kolk, A. (2008). Sustainability, accountability and corporate governance: Exploring multinationals reporting practices. *Business Strategy and the Environment*, 17(1), 1–15. doi:10.1002/bse.511
- Kotler, P., Roberto, N., & Lee, N. (2002). *Social marketing: Improving the quality of life* (2nd ed.). Sage Publications.
- KPMG International Survey of Corporate Sustainability Reporting. (2002). Retrieved December 13, 2015 from <http://www.gppi.net/fileadmin/gppi/KPMG2002.pdf>
- Kurt, G., & Ucma Uysal, T. (2014). Sustainability reporting framework for voluntary reporting or disclosure in Turkey. In M. Khosrow-Pour (Ed.), *Encyclopedia of Information Science and Technology* (3rd ed.; pp. 48–56). Hershey, PA: IGI Global.
- Lamberton, G. (2005). Sustainability Accounting – a brief history and conceptual framework. *Accounting Forum*, 29(1), 7–26. doi:10.1016/j.accfor.2004.11.001
- Lee, D. D., Humphrey, J. E., Benson, K. L., & Ahn, J. Y. K. (2010). Socially responsible investment fund performance: The impact of screening intensity. *Accounting and Finance*, 50(2), 351–370. doi:10.1111/j.1467-629X.2009.00336.x
- Lee, N. R., & Kotler, P. (2011). *Social marketing: Influencing behaviors for good* (4th ed.). Sage Publications.
- Matisoff, D. C., Noonan, D. S., & O'Brien, J. J. (2013). Convergence in environmental reporting: Assessing the carbon disclosure project. *Business Strategy and the Environment*, 22(5), 285–305. doi:10.1002/bse.1741
- Matloff, R., & Chaillou, J. H. (2013). *Socially responsible investing, in nonprofit investment and development solutions: A guide to strategies and solutions for thriving in today's economy*. Hoboken, NJ: John Wiley & Sons Inc. doi:10.1002/9781118650837
- McKenzie-Mohr, D. (2000). Promoting sustainable behavior: An introduction to community-based social marketing. *The Journal of Social Issues*, 56(3), 543–554. doi:10.1111/0022-4537.00183
- Miles, M. P., & Munilla, L. S. (2004). The potential impact of social accountability certification on marketing: A short note. *Journal of Business Ethics*, 50(1), 1–11. doi:10.1023/B:BUSI.0000021139.94166.74
- Nikolaeva, R., & Bicho, M. (2011). The role of institutional and reputational factors in the voluntary adoption of corporate social responsibility reporting standards. *Journal of the Academy of Marketing Science*, 39(1), 136–157. doi:10.1007/s11747-010-0214-5
- ODwyer, B., & Owen, D. L. (2005). Assurance statement practice in environmental, social and sustainability reporting: A critical evaluation. *The British Accounting Review*, 37(2), 205–229. doi:10.1016/j.bar.2005.01.005
- Pava, M. L. (1998). Corporate social responsibilities: yesterday and tomorrow. In R. F. Duska (Ed.), *Leadership and Business Ethics* (pp. 151–174). Kluwer Academic Publishers. doi:10.1007/978-0-585-27624-3_14
- Peattie, K., & Peattie, S. (2009). Social marketing: A pathway to consumption reduction? *Journal of Business Research*, 62(2), 260–268. doi:10.1016/j.jbusres.2008.01.033
- Porter, M. E., & Kramer, M. R. (2006). Strategy & society – the link between competitive advantage and corporate social responsibility. *Harvard Business Review*, 84(12), 78–92. PMID:17183795
- Revelli, C., & Viviani, J. L. (2015). Financial performance of socially responsible investing (SRI): What have we learned? A meta-analysis. *Business Ethics (Oxford, England)*, 24(2), 158–185. doi:10.1111/beer.12076

Salimbene, F. P., & Ferrera, G. R. (1996). Fiduciary responsibility and the duty to account for clients' funds. In E. S. Petry, J. B. Kamm, W. M. Hoffman, & R. E. Frederick (Eds.), *The Ethics of Accounting and Finance: Trust, Responsibility and Control* (pp. 28–41). Westport, CT: Quorum Books.

Simnett, R., Vanstraelen, A., & Chua, W. F. (2009). Assurance on sustainability reports: An international comparison. *Accounting Review*, 84(3), 937–967. doi:10.2308/accr.2009.84.3.937

Smith, S. S. (2014). Integrated reporting, corporate governance, and the future of the accounting function. *International Journal of Business and Social Science*, 5(10), 58–63.

Somers, M. J. (2001). Ethical codes of conduct and organizational context: A study of the relationship between codes of conduct, employee behavior and organizational values. *Journal of Business Ethics*, 30(2), 185–195. doi:10.1023/A:1006457810654

Steurer, R., Margula, S., & Martinuzzi, A. (2008). *Socially responsible investment in EU Member States: Overview of government initiatives and SRI experts' expectations towards governments*. Retrieved December 11 2015, from <http://www.sustainability.eu/csr-policies>

Torlak, Ö. (2006). Pazarlama ahlakı: sosyal sorumluluklar ekseninde pazarlama kararları ve tüketici davranışlarının analizi. İstanbul: Beta Publication.

Torlak, Ö., & Barca, M. (2012). Pazar dinamiklerindeki değişim, stratejik düşünme ve pazarlama düşüncesinin gelişimi. In *Pazarlama Stratejileri: Yönetel Bir Yaklaşım* (pp. 1-24). İstanbul: Beta Publication.

Verrechia, R. E. (2001). Essays on disclosure. *Journal of Accounting and Economics*, 32(1–3), 97–180. doi:10.1016/S0165-4101(01)00025-8

Wasik, J. F. (1996). *Green Marketing & Management: A Global Perspective*. Cambridge, MA: Blackwell Business.

Yükselen, C. (2014). Pazarlama: İlkeler – Yönetim – Örnek Olaylar. Ankara: Detay Publication.

KEY TERMS AND DEFINITIONS

Corporate Social Responsibility: A concept describing the voluntary actions of an organization for fulfilling their responsibilities towards third parties providing resources for the community and the organization.

Non-Financial Reporting: Represents voluntary provision of non-financial information by an organization for the information users constituting its external environment, including mainly its stakeholders, shareholders, investors, and partners, so that acting transparently.

Social Marketing: Is a marketing concept which seeks to influence behaviors not to benefit the company, but to benefit the society in general.

Sustainability Reporting: Represents provision of non-financial information by an organization to internal and external information users in line with the main objectives of development, in addition to an organization's transparency in terms of its performance and as well as its management approach.

Voluntary Disclosure: Represents sharing nonfinancial information, including ethical, environmental, and social issues, by an organization with internal and external information users without any legal obligation.

The Impact of Artificial Intelligence and Virtual Personal Assistants on Marketing

Christina L. McDowell Marinchak
University of Alaska Anchorage, USA

Edward Forrest
University of Alaska Anchorage, USA

Bogdan Hoanca
University of Alaska Anchorage, USA

INTRODUCTION

As marketers and consumers simultaneously adopt Artificial Intelligence (AI) services and applications, the dynamic of the process of exchange between the buyer and seller in the marketplace is being fundamentally altered. Clearly, the Internet and social media have radically altered not only the roles and relationships, but also the differential advantages held by marketer and consumers in the marketing game, as consumers can now find (via Google) and buy (via Amazon and more recently via Facebook) anything, anytime, anywhere on the recommendation of friends and family (via FaceBook) or other customers (via Yelp) and communicate globally the transaction in 140 characters or less (via Twitter). Traditionally, marketing has been something done *to* the consumer. Products were pushed. Communication was one-way and intrusive. Commercial messages remained mostly brief, designed as much to entertain as to inform and often provide the consumer with little more than a catch phrase in the guise of a slogan or jingle. With the advent of the Internet and emergence of social media, the communication pattern of the marketing game expanded from strictly “outbound” with the marketer doing all the talking, to include “inbound,” wherein the consumer searches, contacts, and transacts with the marketer. Marketers no longer can simply buy consumers’ interest; it must be earned. Consum-

ers increasingly rely on each other for product and service information and recommendations. Rather than the customer being targeted by the seller, the seller seeks out the buyer. However, as significant and swift this change has been (given social media is but a decade old), arguably the greatest change is still to come with the advent of AI. As marketers and consumers begin to fully take advantage of what AI agents have to offer, the nature and process of the facilitating exchange in marketplace will dramatically change. As AI agents increase the knowledge and insight marketers have regarding consumers’ product preferences and buying patterns, those same agents decrease the consumers’ susceptibility to the marketer’s message and manipulation. As AI agents search, evaluate, recommend, and finalize purchases, the actual consumer may actually be removed from the final purchasing decision.

This entry will review the state of the art in AI, with a particular focus on applications in marketing. Based on the current capabilities of AI in marketing, the author’s explore the new rules of engagement. Rather than simply targeting consumers, the marketing effort will also be directed at the algorithms controlling the consumers’ virtual personal assistants (VPAs). Instead of exploiting human desires and weakness, marketing will need to focus on meeting the user’s actual needs. The level of customer satisfaction will be even more critical as mar-

DOI: 10.4018/978-1-5225-2255-3.ch499

keting will need to focus on establishing and maintaining a reputation in competition with those of similar offerings in the marketplace. This entry concludes with thoughts on the long term implications, exploring the role of customer trust in the adoption of AI agents, the security requirements for agents and the ethical implications of access to such agents.

BACKGROUND: APPLICATIONS OF AI BY MARKETERS

The term Artificial Intelligence (AI) was first coined in 1956, and has evolved over a half century of steady progress in building computers that can manipulate symbols (both logical and linguistic). In the mid 1990's, with the commercialization and diffusion of the Internet, exponential growth in computing power and advanced software sophistication fully functional intelligent agents have come to pass, and the term "intelligent agent" has been formally defined:

An intelligent agent is software that assists people and acts on their behalf. Intelligent agents work by allowing people to delegate work that they could have done to the agent software. Agents can, just as assistants can, automate repetitive tasks, remember things you forgot, intelligently summarize complex data, learn from you, and even make recommendations to you. (Gilbert, 1997)

Whether termed an agent or assistant or defined as artificial or virtual, these emerging applications are an amalgamation of many technologies – voice recognition, linguistic sciences and natural language processing, machine learning, cognitive computing, analytical and predictive statistics, neural networking, and deep-learning algorithms.

Marketers are currently utilizing an array of advanced AI platforms, services and technologies. The scope of activities and functions that AI is already being applied includes:

- Basic market research (i.e., consumer behavior and psycho-analytics),
- Media and creative strategy (i.e., media planning, message design, and pricing),
- Customer service, sales, and relationship management.

Basic Market Research

AI has already enabled marketers to understand qualitative input from surveys, in addition to the quantitative survey results. Using tools from companies like Luminoso Technologies and Kanjoya Inc., marketers can automatically analyze text input on surveys and social media platforms to understand meaning as well as the emotions of the subjects (Rusli, 2014). Marketing firms currently employ AI agents, particularly exploiting an ability of AI to simultaneously enhance both intelligence and emotion (Mceleny, 2015). With the ability to analyze structured and unstructured data from credit cards, sales databases, social networks, location data, and web pages, marketers can ascertain trends in consumer preferences and sentiments as well as calculate sales probabilities.

AI has been applied to the real time analysis of human emotions. Advanced image processing has matured to the point where emotion recognition can be done either via a webcam or using a camera in a physical setting (and emotion detection in verbal interactions has an even longer history) (Eaton, 2012). Using AI techniques, the user's reaction to a product or to an advertisement can be interpreted, and the stimulus can be modified accordingly. For example, a viewer displaying positive emotions towards the product can already be cajoled into completing a transaction (or can be shown additional positive messages to encourage the transaction), while for negative reactions the viewer can be offered more explanations (if the emotion detected is confusion or a blank stare), a different set of options (if positive and negative emotions are displayed) or a different offering altogether (if strong negative emotions are encountered).

Indeed, AI systems such as IBM's Watson has the ability to discern consumers' personalities through *Psycholinguistic Analysis* of the language a consumer uses when posting on social media sites and predict major events that are likely to happen in a consumer's life. Watson does not need to know the social media accounts of a consumer beforehand, rather the technology is able to identify a consumer's social media usage by sifting through what a consumer posts via a company database (Russel, 2014). Thus, AI is changing the scope of basic market research, opening up a new challenge for marketers to move beyond traditional strategies and tactics for identifying consumer behavior.

Media and Creative Strategy

Advancements in AI enhanced research and analytic applications allow marketer's to find, read, and analyze a consumer's every search, post, tweet, pin, video message, product evaluation, credit card purchase, media consumption preference and/or any and every other bit of information that can be gleaned in the infosphere about the consumer. Companies have already started to mine social media to understand emotions and preferences as expressed on platforms (i.e., Bluefin Labs) (See Arndt, 2011).

Artificial intelligence is being used to create custom commercial messages for individual consumers based on their user profile and media behavior. Accordingly, campaign materials are modified in real-time, to maximize consumer interest and message retention. AI agents can adjust photos of human faces to be more memorable, attractive and friendlier (Khosla, Bainbridge, Torralba, & Oliva, 2013) as well as popular online (Khosla, Sarma, & Hamid, 2014). Similar technologies could be used in real time to optimize the light and presentation of a product in an ad, based on aggregated user reactions to the ad. "Dynamic Creative Optimization" technology is also being used by marketers to customize promotions and pricing. However, marketers should remain cog-

nizant of the risk of adjusting "prices on the fly" (i.e., Amazon) (See Gunderson, 2012).

The predictions is AI driven "Dynamic Creative Optimization" could be responsible for as much as 40–50 percent of all promotional messaging and advertising within the next ten years. As proclaimed by the media and communications agency PHD, inevitably "Algorithmic-driven creative and knowledge-driven targeting will open up the potential to move beyond segmentation and focus on individual level messaging—so each person has a communication targeted specifically at them based on their unique situation (what they know, what they like, category purchase propensity, the brand purchase propensity, where they are in the purchase process" (PHD, 2015).

Customer Service, Sales, and Relationship Management

The most recent and perhaps the most advanced applications of AI technology by marketers is evident in the areas of customer service, sales assistance, and relationship management. In May and August of 2015, Amazon and Facebook respectively introduced their AI enhanced customer assistance agents "Alexa" and "M." As described by one author, "buying items on Amazon would now be even easier than pressing a button with 'Amazon's Choice,' a new feature for the Internet-connected "Amazon Echo" speaker that will allow the unit to automatically order items in response to a voice command ... like Apple's Siri ... You "wake it up" with your buzzword, ask it to set reminders, play streaming music, and look up simple Internet queries (Beres, 2015).

Facebook's foray into the AI concierge market comes in the guise of "M." As described by Hackman (2015), M is

... powered by artificial intelligence, trained and supervised, and designed to help answer questions like 'Where is the best place to go hiking? Can you help me order flowers for my Mom's birthday?' Like Apple's Siri, Google Now, and Microsoft's

Cortana, Facebook's M "largely answers fact-based questions and performs simple tasks confined to a computing device. (Hackman, 2015)

Although, M goes

... a step further, making intelligent decisions on your behalf. As of now, however, Facebook is not tapping into its social graph to help answer those questions ... It's also not clear whether M will actually make a purchase for you, or simply point you to a service or website for you to take care of yourself. (Hackman, 2015)

Sophisticated AI applications in customer service have been developed by IBM. Envisioned as an intelligent technology whose services can apply to almost any industry, but especially those that receive many customer service inquiries, such as retailing, banking, insurance and telecommunications, IBM offers the Engagement Advisor (Lohr, 2013). IBM's Watson Engagement Advisor has the ability to scan, analyze, and interpret over 200 million pages of structured and unstructured data in a matter of seconds and subsequently enter into a natural language conversation based on its interpretation of and conclusions drawn from its analysis (Schiff, 2013). For any purchase the consumer is considering all that is needed is a quick query of Watson: Watson what is the best brand? What is the best price?

With the ability to analyze the personality of the consumers, categorize their lifestyle and buying style, extrapolate all their online information and communications, the marketer can now formulate the optimal marketing message for the right person at the right time and sell, cross-sell and up-sell their products.

FUTURE RESEARCH DIRECTIONS: THE FUTURE OF VIRTUAL ASSISTANTS

Sophisticated as current systems may be, the next generation of AI agents represent a paradigm shift

in AI marketing applications. Rooted in cognitive computing, advanced sensor technology, and natural language interfaces, the next iteration of the AI marketing agent is just beginning to emerge, and it fulfills the promise of that all-knowing and all-powerful entity that can obey and fulfill the consumer's ultimate command – "Tell me which brand is best to buy and why!"—transforming marketing.

Cognitive computing will be the ultimate marketing game changer, ushering in "the smart machine era" that (short of the singularity envisioned by Ray Kurzweil) "will be the most disruptive in the history of IT" (Austin, 2013). Today, Apple's Siri, Google Now, and Microsoft's Cortana are familiar companions to smartphone users. The consumer's unspoken thoughts are now being confirmed by the virtual personal assistant (VPA), which is designed to serve/guide them through everyday decisions. VPA's will soon be everywhere, and, yet, nowhere in the life of a consumer—a gateway to the Internet and other people like no other (Brauer, 2014). Already traces of its trajectory exist from mobile and wearable, to 'hearable' and perhaps one-day 'implantable' wearables, hearables, trackables, and even scent (Brauer, 2014). No doubt, consumers will initially want to determine when and where to synch their physiology, decision-making and emotions with a VPA, but, then real-time synchronization will just be easier and more efficient (Brauer, 2014).

In addition to knowing a consumer's tastes, needs, and desires, the VPA knows all past purchases, current preferences and future plans. Moreover, the VPA has a 360 degree view of the marketplace and can provide a ranking and recommendation of all competitive and comparative products. Rather than pushing a single brand, as a marketer's bot, the VPA more inclusively and objectively evaluates an entire evoked set of brands and proffers a recommendation aligned with and in the best interests of its master. Perhaps, in the long run the VPA will enable the consumer to gain the upper hand in the marketing game.

NEW RULES: NEW ROLES

As the marketer is increasingly disintermediated, it will become increasingly important for the marketing industry to recognize the primary role that the consumer's VPA plays in the marketing game and to adjust marketing efforts accordingly. Continuing to invoke the old moves in the new game will not only prove ineffective, but counterproductive in gaining consumers' trust and patronage.

However it remains that all too often one finds the application of AI agents has primarily served marketing professionals to augment and extend their traditional marketing toolset. It is fairly common practice to use data driven advertising, where click-through rate or revenue generated are used as a metric to identify which one of two ads is more effective (Ayres, 2008). Using AI techniques, ads are adjusted in real-time to make them more memorable or possibly more effective at converting views into revenue (Khosla, Bainbridge, Torralba, & Oliva, 2013). Google AdWords has advertisers bid against each other to auction space for ads on websites, based on the content of the page and the interests of the viewer, as reflected in the web cookies on his or her machine. Ostensibly, this leads to better, targeted ads. In the end, the consumer may not want fewer better, targeted, designed advertisements, but rather *no* advertisement.

Over the past decade, advertisements have become more and more intrusive. At one time, being selected to participate in a survey or a focus group made a person feel elite; however, today's consumer is constantly battling online ads (Arana, 2015). Such intrusion has spawned a variety of ad-blockers that attempt to alleviate users' frustration. Over the last year alone, ad-blockers in the United States jumped 48 percent – 16 percent of U.S. online users now try to block ads and 200 million people use ad-blocking software a month (Arana, 2015).

The claim is the rise of ad-blocking software is due to the increase use of mobile devices

when going online – smaller screen makes ads more intrusive. However, “not everyone in the ‘ecosystem’ users is inclined to see it that way” (Arana, 2015). Users are no longer receiving “‘sensible’ ads – generally defined as those limited to text or those appearing alongside, but not on top, of content” (Arana, 2015), which most are willing to support. Instead, most ads are intrusive, covering content, lacking a way to bypass them and blocking the user's screen for a fixed amount of time, up to tens of seconds in some cases. The reason for such an intrusive approach is the loss of focus of attention on user experience.

This increasing realization that the influence and impact of the marketer's commercial campaign is on the wane is further exacerbated by the fact that it may eventually be dismissed altogether when the consumer's purchase cycle is fully maintained and conducted by a consumer's VPA. As VPAs become available they will increasingly serve as the primary product purchasing platform for consumers. Accordingly, corporate marketers will need to address the VPA directly, instead, or in addition to the human target.

New Target: Market More to the Algorithm than the Consumer

In the emerging AI driven marketplace if you want to reach/influence the consumer you will need to reach/influence the algorithm that is controlling their communication stream. As argued in *Sentience: The coming AI revolution and the implications for marketing*, if one accepts “the likely scenario that our VPAs will be tasked with carrying out many of our day-to-day tasks, including searching for options,” then “it's clear it is our VPAs that will be making decisions on which brands/ products we should see... A significant focus for marketing will be to influence the algorithm and ensure that the VPA is surfacing your brands/ products ahead of your competitors (PHD, 2015).

New Rule: Access to the Consumer by Permission Only—Relevancy and Reputation Required

As the consumer's commercial gatekeeper, the AI agent will screen every attempt to enter its client's communication stream for relevancy and reputation. Relevancy may be determined by such factors as the time, location, expressed needs and established tastes. Reputation may be measured by both objective and subjective industry and consumer evaluations as well as the consumer's own past experiences. What product/service ultimately qualifies for consideration will be based on "extremely specific purchasing and behavioral data" unique to each individual. If the product/service meets the qualifying criteria it remains that an admission fee (or bid price) may be charged to the marketer for entry in/on to the communication stream.

VPAs will actually make the decisions on when to pull down a brand/product and surface-up a suggestion to their user... the bid price and relevancy/quality score are most likely to determine which brand, products or services are suggested... Then it will report back to the advertising company which ads were actually shown so that the company can bill its clients... All... will be handled in nanoseconds by our VPAs, and served up without us knowing which brands or products narrowly missed out. (PHD, 2015)

New Role: Reputation Management and Enhancement

Within this new paradigm of AI arbitrated commercial messaging a key variable that marketers must measure, monitor and bolster their company's or (for consultants) their client's reputation. Indeed, it has been suggested that managing a product/service's reputation is going to become a defining function of leading marketing practitioners. Marketers that can bolster the positive reviews and customer recommendations "will impact the

quality score, which in turn increases likelihood of the brand being selected by the VPA and/ or reduces the bid price required" (PHD, 2015).

New Role: AI as Utility

As another example of where the VPA technologies are currently, Viv Labs is a company that has recently announced a much more powerful agent than Siri, Cortana or M. Known as Viv, this AI agent will be able to quickly learn a user's preferences, and will be able to carry out complex tasks and answer complex queries for the user. The company founders view Viv as a utility, something that users can plug into and use whenever needed, as they use electricity. Viv will be a global brain, learning a user's preferences, but learning generic skills from its interaction with other users. This approach will allow the agent to learn much faster than a personal agent who interacts only with its own user. Users of Viv might be surprised to try something new, only to realize that Viv has already been trained to accomplish that type of task based on its interactions with other users. (Levy, 2014)

A side effect of AI as utility will arise from the ability of the global brain to aggregate feedback from countless interactions. Rather than relying on just a subset of customers to report experiences on Yelp or TripAdvisor, the Viv brain will automatically aggregate all the customer experiences and make that aggregate evaluation a part of any prospective transaction. This type of automatic and more objective reputation building for merchants will prevent some of the astroturfing techniques that currently involve the selective editing of reviews or providing incentives for some customers to post biased reviews.

New Rule: Toss Out the Rule Book

There is the old adage that marketer's "make people buy things they do not need, with money they do not have, to impress people they do not like." Regularly, traditional marketing relies on impulse buying (preference for a product that

is immediately available over one that is less available), cognitive heuristics and limitations (preference for a simple choice over a more complex choice), sensory overload (responding to appetizing smells or to arousing imagery) and irrational decision-making (including the inability to carry out complex calculations in one's head). In contrast, VPAs will be closer to the *homo economicus* ideal, with essentially unlimited ability to uncover and compare choices, with the ability to follow the user's preferences and not be swayed by the commercial message, and with the power to negotiate with several merchants at the same time. If consumers increase their reliance on VPAs, this will greatly reduce the success of some of the marketing techniques that rely on human traits. Marketers will need to truly understand what customers want, as well as how to reach those customer needs via the user's VPA. To summarize the new rules, marketers will need to market to the VPA algorithm (not to the user's human weaknesses). In marketing to the algorithm, marketers will need to gain the user's VPA's permission via reputation and relevancy. Finally, marketers will need to be aware of the AI as utility, and to realize that AI agents will communicate and share information, just as customers do today, just much faster and more comprehensively.

CONCLUSION

The role of AI in marketing is only in its nascent stage and its impact is predicted to increase exponentially. This entry illustrated applications in marketing that will impact both corporate marketing systems as well as those on the consumer's side. Research on corporate marketing systems will need to focus on how to use AI to improve current marketing processes. A related, but important research area will involve the customer perceptions of these AI systems. Because AI agents will learn about consumers' habits and will be able to infer behaviors, trends and preferences that might be

rather private, corporations will need to establish privacy policies that will reassure customers. Additionally, the rise of VPAs will have a considerable impact on how customers make purchasing decisions. As rational decision-making agents, VPAs will not be susceptible to the same types of marketing messages currently aimed at humans. Marketers will need to understand how to position their products and how to target a mixed audience of both humans and VPAs. Future research will also need to consider how VPAs will be created and deployed to gain the trust of consumers. Finally, a social concern is that the rise of VPAs will lead for further social inequalities. Wealthier people, who will be more likely to afford a more expensive and capable agent, will have even more access to the best deals (via the toughest negotiating agents), the most up-to-date information (via the most far reaching agent searches), and the most efficient consumption processes. Despite the possible economic benefits of VPAs, an unwise application of the technology has the potential to further deepen the digital divide, and to increase the inequalities that plague society. Further research will need to uncover ways in which such an outcome can be avoided.

Considering long-term implications, exploring the role of customer trust in the adoption of AI agents, the security requirements for agents, and the ethical implications of access to such agents this entry explored the new rules of marketing engagement. AI agents are upon us. The new roles—empowered consumer—are here and represented by AI, which is not so much a prediction, but a reality.

REFERENCES

- Arana, G. (2015, Oct 19). Ad-Blocking Has Online Ad Industry On The Run. *The Huffington Post*.
- Arndt, R. Z. (2011, November 7). Bluefin mines social media to improve TV analytics. *Fast Company*.

Austin, T. (2013). *The Disruptive Era of Smart Machines Is Upon Us*. Gartner. Retrieved from <http://www.gartner.com/doc/2598620>

Ayres, I. (2008). *Super Crunchers: Why Thinking-By-Numbers is the New Way To Be Smart*. Bantam.

Beres, D. (2015, May 18). Buying items on Amazon is now even easier than pressing a button. *Huffington Post*.

Braurer, C. (2014, Oct 8). *Beyond Siri, Google Now and Cortana: What is the future of virtual assistants?* Retrieved from: <http://thenextweb.com/insider/2014/10/27/beyond-siri-google-now-cortana-future-virtual-assistants/>

Eaton, K. (2012, June 7). Does your phone know how happy you are? The emotion recognition industry comes giddily of age. *Fast Company*.

Gilbert, D. (1997). *Intelligent Agents: The Right Information at the Right Time*. IBM Corp.

Gunderson, L. (2012, May 11). Amazon's 'dynamic' prices get some static. *The Oregonian*.

Hackman, M. (2015, Aug 26). Facebook 'M' deploys army of virtual assistants to help fix your life. *MacWorld*.

Khosla, A., Bainbridge, W. A., Torralba, A., & Oliva, A. (2013). Modifying the Memorability of Face Photographs. *International Conference on Computer Vision 2013*. doi:10.1109/ICCV.2013.397

Khosla, A., Sarma, A. D., & Hamid, R. (2014). What Makes an Image Popular? *International World Wide Web Conference*.

Levy, S. (2014, Aug 14). Siri's Inventors Are Building a Radical New AI That Does Anything You Ask. *Wired*. Retrieved from <http://www.wired.com/2014/08/viv/>

Lohr, S. (2013, May 21). *I.B.M. Puts Watson to Work in Business*. Retrieved from New York Times blogs: http://bits.blogs.nytimes.com/2013/05/21/i-b-m-puts-watson-to-work-in-business/?_php=true&_type=blogs&_php=true&_type=blogs&_r=1

Mceleny, C. (2015, June 29). *Rise of the machines: what artificial intelligence means for marketing*. Retrieved from Marketing Magazine: <http://www.marketingmagazine.co.uk/article/1352953/rise-machines-artificial-intelligence-means-marketing>

PHD. (2015). AI-Enabled Marketing: The Implications for Marketing. In SENTIENCE: The coming AI revolution and the implications for marketing (pp. 833-834). PHD.

Rusli, E. (2014, April). Firms Use Artificial Intelligence to Tap Shoppers' Views: Analyzing Focus Groups, Surveys, Online Forums and Social Media Provides New Data. *The Wall Street Journal*.

Russell, K. (2014, February 4). IBM Is Using Watson To Psychoanalyze People From Their Tweets. *Business Insider*. Retrieved from <http://www.businessinsider.com/ibm-watson-psychoanalyzes-tweets-2014-2#ixzz2sPUJ6cOw>

Schiff, A. (2013, May 21). Just Ask Watson: IBM Unveils the Watson Engagement Advisor. *Direct Marketing News*. Retrieved from <http://www.dmnews.com/just-ask-watson-ibm-unveils-the-watson-engagement-advisor/article/294283/>

KEY TERMS AND DEFINITIONS

Artificial Intelligence: A computer system designed to exhibit features of human intelligence, including for example being able understand questions via natural language (human speech), solve

complex problems and present reasoning, and output answers using natural language.

IBM's Watson: Artificial intelligence system originally designed to showcase AI efforts at IBM by competing in the Jeopardy game. After winning Jeopardy in 2011 against the then-current champions, Watson was marketed by IBM for applications in medicine, customer service and general problem solving. Currently, Watson is being developed into a general purpose AI system that has been opened to a community of beta tester.

Intelligent Agent (IA): An artificial intelligence system, possibly endowed with human-like personality traits and with the ability to recall context and history of interactions in a manner that would facilitate future interactions with a human being.

Marketing Recommendation System (MRS): Computer system using AI or similar technology to make product, brand or feature

recommendations to customers. Some current systems have emerging abilities of this nature, but true AI systems are expected to provide insights much closer to those of expert sales staff.

Siri: Apple's personal assistant on the iPhone/iPad/iPod platform, able to understand and respond to basic questions using natural language using AI in the cloud. While Siri was one of the first AI agents in general use, its capabilities are rudimentary compared with those expected of true AI systems.

Virtual Personal Agent (VPA): Intelligent agent that is intended to assist a human user with purchases over time, developing a relationship and a context that optimizes the communication with the human user; for example, the VPA could remember the preferences of the human user and would be able to reorder items from a favorite brand or to suggest brands that would best meet.

Marketing and Marketing Plan for Information Services

M

Sérgio Maravilhas
Universidade Salvador, Brazil

INTRODUCTION

Marketing is a social and economic process through which individuals and groups meet their needs¹ and desires² by creating and exchanging products³ (and services) among themselves (Kotler, 2000).

Initially, marketing was practiced by companies in selling sectors of products and services and did not seem to fit the non commercial side. However, in recent decades almost all types of organizations adopted the methods and practices of marketing (Iacobucci, 2001), including information services.

To carry out a coherent program of dissemination of products / services developed by an organization, it is necessary to develop a program to properly plan all the actions and resources needed to achieve these objectives, and effectively control the deviations relative to them, allowing correcting the actions that do not get the expected results in order to constantly improve the system implemented (Mohr, J., Sengupta, S., & Slater, S., 2010).

For this it is necessary to devise a marketing plan that focuses on the planning of all activities and objectives we want to achieve and the means necessary to do so (Dibb et al., 2006; Jobber & Ellis-Chadwick, 2013; Kerin et al., 2006; Kotler, 2000).

The Marketing Plan is a strategic document that will serve to identify the position occupied in the market⁴, set goals and how they will be achieved, all the resources that are needed and expected results.

The purpose of this plan is to enable entrepreneurs to capitalize its resources, spreading the information on the goods and services they provide and channeling the profits derived from such use for their own funding.

BACKGROUND

According to Kotler (2001), Marketing is a social and economic process through which individuals and groups meet their needs and desires by creating and exchanging products (and services) among themselves.

We must consider that marketing has to do with the satisfaction of needs and that if several economic, technological and human efforts are made in order to satisfy those needs, ultimately, society as a whole will be the one who benefits the most (Dibb et al., 2006; Jobber & Ellis-Chadwick, 2013; Kerin et al., 2006; Kotler, 2000).

Initially, marketing was practiced by companies in sectors of products and services selling and did not seem to fit the organizations with non commercial side, like information services, with no intention of making a profit. However, in recent decades almost all types of organizations adopted the methods and practices of marketing. Examples of this practice are political parties, religious bodies, social, philanthropic and charitable entities, in a logic which was instituted to be designated as Social Marketing (Dibb et al., 2006; Jobber & Ellis-Chadwick, 2013; Kerin et al., 2006; Kotler, 2000).

The adherence to this practice of management should have in mind the services providing information, essential for the development of collaborative R&D and innovation, although not strictly commercially oriented and intended to make a profit through the products and services they provide (Batchelor, 1997).

Let us not forget that, according to Kotler (2004, p. 9)

The fact that most of these services are 'free' does not affect the characterization of the products. A product is something that has value to someone. Charge or not for your consumption is a trait tangential, unessential, the definition of its value. In fact, most of these social goods has its 'price', although this is stated not in the usual way.

Thus, by referring to nonprofit suppliers of this type of product and service we cannot talk about companies but organizations, nor speak of customers but publics. Rather than speak of sales we should be promoting behaviors and instead of return we should mention the achievement of objectives.

Knowing that the objectives can be non-financial, in this perspective we should opt for a broader definition: «Marketing is the set of methods and means that an organization has to promote the behaviors, conducive to the realization of their own objectives, in the public's that interest them. (Lindon et al., 2000, p. 30)

We will not discuss here if Marketing satisfies or creates needs because, in one way or another, in the end it will be a need (created or not) that will be satisfied.

According to Zachert and Williams (1986, p. 61) the basic idea of marketing is that the effective response to the needs and requirements of customers are the key to the company's success in the market, being the most successful organization the one that has a proactive attitude towards the constant challenges faced by their clients.

To know our market, and customer requirements contained therein, it's necessary to obtain relevant information about consumer habits and behaviors. We can therefore say that the need to obtain relevant information exists in every business. Seems that companies, especially small ones (which constitute the majority of businesses) have this latent need, but are not informed about what information will meet their needs.

Someone must take the initiative to help them, informing them and directing them to ways of obtaining and using the resources of information from various sources, in order to enable the best way to meet the needs identified in their publics and markets where the company operates.

Also according to Zachert and Williams (1986, p. 61) the organization must foster participation of its customers to allow the design and development of products and services that fit their needs and requirements. The secret, they say, lies in a personal and open communication with them.

MARKETING-MIX: THE FOUR 'P'S'

Although the commercialization plan, or Marketing-mix, encompasses many activities, in the early 60's Jerome McCarthy proposed a four 'Ps' Marketing-Mix.

These four 'P's' are:

1. Product;
2. Price;
3. Place;
4. Promotion.

They define what is the product/service that our publics really want, the price they are available to pay to have it, where and how can they find and buy it the way they prefer to do it, and how can we communicate with our publics, by which means and media, to let them know we have what they want (Walker Jr., O., Mullins, J., & Larreche, J.-C., 2008).

The framework of the four 'P's' demands (Dibb et al., 2006; Jobber & Ellis-Chadwick, 2013; Kerin et al., 2006; Kotler, 2000):

1. That the people in charge of marketing decide about the product / service and its characteristics;
2. Establish the price;
3. Decide how best to distribute the product;
4. Choose methods of promoting it.

According to Kotler (2001), at least two more 'Ps' should be included, which are extremely important in the global market: Politics and Public. For Kotler, this marketing mix is the set of controllable variables and their levels, which the company uses to create a defined positioning in the environment and to exert an influence on the market that it targets.

These four Marketing variables must direct to and fit our market because a good marketing strategy involves these marketing mix variables perfectly tuned with the fixed target market of our business. But for that to happen, it is imperative to know who our targets are. Therefore, knowing the market implies classify it, segment it, investigate it, to know the consumer and how he decides the process of obtaining goods and services, how often, and so forth.

MARKET SEGMENTATION

We know a market when we can answer the following questions (Iacobucci & Calder, 2003):

1. Who are our potential clients/publics?
2. What do they buy (in terms of products or services)?
3. When do they buy?
4. Who is involved in these purchases?
5. Why do they buy these products or services?
6. How do they buy?

Answering these questions allows us to differentiate the market where we intend to act.

Thus, the strategy should be directed to the segmentation of the market, which means rank the total market of a product or service in a number of homogeneous subsets such as the habits, needs and tastes of its components, calling up each of these subsets as a market segment (Walker Jr., Mullins, & Larreche, 2008, p. 133). Simply stated, the intention of segmentation strategy is to meet the needs of a customer group with appropriate products and services. "Market segmentation is the process of identifying distinct groups of buyers who might require separate products and/or marketing mixes, then creating a profile of each group" (Rea & Kerzner, 1997, p. 41).

The purpose of identifying these segments is to enable companies to better tailor their Marketing policies to all or some of these subsets.

Segmentation is a strategy that aims to address not the public as a whole, but the different groups in more specialized ways.

The three most important steps in the market segmentation process are:

1. Identify a homogeneous segment that differs from other segments;
2. Specify criteria that define the segment;
3. Determine segment size and potential (Walker Jr. et al., 2008, p. 133).

Kotler (2000, pp. 263 - 269) describes some bases, or variables, of segmentation:

1. **Geographic:** Dividing the market into different geographical units such as nations, states, regions, counties, cities, or neighborhoods;
2. **Demographic:** Dividing the market into groups on the basis of variables such as age, family size, family life cycle, gender, income, occupation, education, race, religion, generation, nationality, and social class;
3. **Psychographic:** Buyers are divided into different groups on the basis of lifestyle, personality, and values;
4. **Behavioral:** Buyers are divided into groups based on their knowledge of, attitude toward, use of, or response to product characteristics.

For these variables to work effectively we must combine them in order to identify smaller and better defined target groups.

The benefits of segmentation are (Walker Jr. et al., 2008, p. 133):

1. Recognizing the importance of each segment;
2. Detecting and analyzing opportunities offered by the market, assessing the degree to which the needs of each segment are covered by existing products or services;
3. Tailoring products and marketing policies to the tastes, needs and preferences of each segment;
4. Defining strategies and budgets based on the most reliable information for each specific segment and make sales predictions safer and sounder;
5. Allowing adaptation of advertising and media related to the characteristics and habits of each segment;

Organizing the best way to deliver the product / service to potential buyers, according to the consumption characteristics of each segment.

Depending on the segmentation performed, there are three possible strategies:

1. **Undifferentiated Strategy:** The same for all segments. Means lower costs but in return there is less adaptation to markets;
2. **Differentiated Strategy:** Different strategies and products for each market segment. The offer is best suited to the segments allowing better results. May involve higher costs but allows better achievement of the proposed objectives;
3. **Focused Strategy:** To focus on a particular sector, facilitating a better understanding of the target customers and hence greater loyalty and participation, even if the focus is on a more limited market.

Marketing strategies are based on market segmentation, meaning that the majority of products

and services do not address the entire market but specific groups of consumers.

The choice of the strategy will depend on several factors (Jobber & Ellis-Chadwick, 2013):

1. Resources available to the organization;
2. The characteristics of the product or service;
3. The product life cycle;
4. The homogeneity of the market.

Besides all these constraints to adopt a strategy, a policy of targeting can bring significant advantages.

If it is certain that all alternatives may have their exceptions sometimes affront the installed habit can assume a competitive advantage.

Segmenting (S) the market is the first step, targeting (T) the segment that we want to work with is the second. The third step is positioning (P), making STP a powerful tool in the marketer toolbox.

POSITIONING

This refers to the set of images through which the market 'sees' a company, product, person, or even an idea, and clearly identifies it (even among similar ones).

Positioning is one of the Marketing action lines more attractive, complex and key success factor in current competitive dynamics. It is normal to develop around the product or service and divide itself between repositioning of existing products or services and positioning of new products or services.

Positioning is the set of the salient and distinctive image that allow the public to place the product on the universe of similar products and distinguish it from others. In other words, it is a representation (or perception) simplified, reductive, comparative, and distinctive of the product. (Lindon et al., 2000, p. 139)

Positioning, for a permanent activation of the competitive dynamics, involves:

1. The competence, courtesy and credibility of people in contact with customers;
2. The written and visual forms, the work environment, and the symbols used;
3. The 'service' relation with the customers;
4. Interaction, features, style, design of products / services, with the foregoing;
5. The planned, organized, controlled way, but at the same time flexible and creative that each company chooses to act in the market.

In consumers' minds must prevail the 'name' of the company or brand, so distinct and unique to be remembered whenever the consumer needs to acquire the same type of goods or services. By establishing a preference for quality and satisfaction in previous acquisitions, it must become pronounced in his mind.

MARKETING AND COMPETITIVE ADVANTAGE

The organization should learn how to detect needs and adopt a proactive stance in relation to those who can assist in satisfying their needs (Dibb et al., 2006; Jobber & Ellis-Chadwick, 2013; Kerin et al., 2006; Kotler, 2000).

We cannot be waiting for them to seek us, because for various reasons this may never happen (Walker Jr., O., Mullins, J., & Larreche, J.-C., 2008).

It is imperative that we send information about our product line to the interested publics and, if it becomes necessary, a technical explanation to accompany it for companies less endowed in terms of intellectual capital⁵ so they can seize it and draw the appropriate lessons that enable them to design products or perform services able to provide them with competitive advantages (Dibb et al., 2006; Jobber & Ellis-Chadwick, 2013).

Consequently, the increase of more specialized and specific product development, for example, with more customized services (Dibb et al., 2006; Jobber & Ellis-Chadwick, 2013; Kerin et al., 2006; Kotler, 2000).

This idea of joining product with service gives rise to the notion of expanded product which is a form of differentiation of a product with value added that allows obtaining competitive advantages (Dibb et al., 2006; Jobber & Ellis-Chadwick, 2013; Kerin et al., 2006; Kotler, 2000).

The main sources of market opportunities are (Dibb et al., 2006; Jobber & Ellis-Chadwick, 2013; Kerin et al., 2006; Kotler, 2000):

1. Provide a scarce article not widely available;
2. Provide articles or existing services in a new or better way;
3. Provide a new product or service.

So, Marketing applied to improvements in the satisfaction of needs and desires can actually lead to a way of generating competitive advantage for businesses and, by doing that, improving the wellbeing of the society.

MARKETING PLAN

To be able to carry out a coherent program of dissemination of products / services sold by us, it is necessary to develop a program to properly plan all the actions and resources needed to achieve these objectives, and effectively control the deviations relative to them, allowing to correct the actions that do not get the expected results in order to constantly improve the system implemented (Dibb et al., 2006; Jobber & Ellis-Chadwick, 2013; Kerin et al., 2006; Kotler, 2000).

For this it is necessary to devise a marketing plan that focuses on the planning of all activities and objectives we want to achieve and the means necessary to do so⁶.

Let us not forget that "To fail to plan, is to plan to fail".

The strategy outlined here should last 3-5 years (Walker Jr. et al., 2008, p. 133). However, its evaluation (and, if necessary, reorganization) should be done annually.

In order to properly define the alternatives to follow, the steps for the preparation of the Plan are:

1. **Analysis of the Market (Commercial Research):** By which we collect and gather all the information needed to make decisions and to select alternatives, like for example:
 - a. Who will be our market (who are our customers - or publics -, their needs, motivations, etc.);
 - b. How to deliver the product or service (best way to distribute the product / service for our customers: specialty stores, Internet, large retail, etc.);
 - c. The price to pay for the product / service (payment forms more suitable for our customers, importance of price in obtaining the product or service, etc.);
 - d. Demographic data (number of customers in a given market, qualifications, salary, areas of interest, family life cycle, etc.);
 - e. Communication (best way to reach our service to potential stakeholders: Media, Newsletters, Social Media, etc.).
2. **Commercial Planning:** In relation to the process of marketing this will be the basis of its development and must be done according to the selection of commercial alternatives and information we've collected in point 1. It is here that the alternatives to adopt will be developed, planned and executed, and must specify the resources necessary for its effective realization (financial, human and material), as well as its temporal variable that must be annual but must take into account the strategy of the organization in the long term (3-5 years).
3. **Business Organization:** To plan for the work is not just its preparation but people

are also needed, appropriate to the function, to execute it. Also important is the organizational system to implement so the plan can work (customer service, administrative support, technical support, etc.).

4. **Commercial Control:** The Marketing Plan should have its objectives well developed and quantified so they can be monitored during its operation. This allows knowing if the objectives are being met and if adjustments need to be taken to improve it (Lindon, Lendrevie, Rodrigues, & Dionísio, 2000).

The marketing plan is a strategic document that will serve to identify the position occupied in the market, set goals and how they will be achieved, the resources that are needed and the expected results.

Being this plan a part of the organization Business Plan, its implementation should be proposed at the highest level (the Board) to be accepted without reservation by all stakeholders (Donnelly, Gibson, & Ivancevich, 2000; Rea, & Kerzner, 1997).

The purpose of the preparation of this Plan is to enable entrepreneurs to capitalize on its resources, spreading the information on the goods and services they provide and channeling the profits derived from such use for their own funding. At the same time, aims to allow increased competitiveness and economic growth to companies that use the information collected for the preparation of this Plan, enhancing competitive advantages.

If we think of a plan as the definition of what an organization hopes to achieve, how and when, and the act of planning as a systematic way of an organization seeking to control the future, we deduce easily that the planning benefits are:

1. Promote a systematic reflection about the future;
2. Enable a better coordination of the resources;
3. Establish performance goals, allowing to measure results;
4. Provide a rationale for decision making;

5. Improve the ability to face changes;
6. Promote the ability to identify market opportunities (adapted from Kerin et al., 2006).

FUTURE RESEARCH DIRECTIONS

Further research should be conducted to evaluate the benefits of introducing these concepts in the daily work of information services, in order to extract best practices that could be applied where needed. Using benchmarking, it is possible to improve what's not working and benefit from others experience and expertise, improving results.

CONCLUSION

Marketing activity is one of the most important management activities of modern societies (Dibb et al., 2006; Jobber & Ellis-Chadwick, 2013; Kerin et al., 2006; Kotler, 2000). To be able to satisfy consumer's needs and desires better than their competitors, product quality, extraordinary service and customer support are needed, giving customers and the society in general the best that is possible with the means at their disposition (Iacobucci, 2001).

With the advent of Internet and mobile technologies (Michael & Salter, 2006), supported by Social Media and Web 2.0 tools (Ryan & Jones, 2012), new marketing strategies are being shaped to adapt to these new media and its users (Iacobucci & Calder, 2003). Digital marketing, social media marketing, web marketing and other similar concepts are taking shape (Ryan & Jones, 2013). Nevertheless, although the technologies and digital platforms are evolving, making some of the work easier but making some parts even difficult, the basics of Marketing are the same: users needs and desires, segmentation, targeting, positioning, product, price, promotion and distribution (Ryan, D., & Jones, C., 2012, 2013) are always needed to perform, but now they are supported by new ways

of interacting with the company publics (Jobber, D., & Ellis-Chadwick, F., 2013).

Every company – new or already in the market – needs a marketing plan, included in their annual business plan (Donnelly et al., 2000). A start-up needs a marketing plan to attract investors and other shareholders and stakeholders. Companies already operating need it for the same reasons, but also to 'keep an eye on the shore' and not let competition get ahead with an advantage knowing always where they are heading and where they want to get in a certain period of time (Rea & Kerzner, 1997).

We hope to have contributed to a better understanding of what marketing is and what are the characteristics and advantages of developing a marketing plan for establishing a strategy to be followed by the company in order to achieve its objectives.

The marketing plan should not be seen as a static document, unaltered, but as a dynamic element that can and should be adapted and adjusted according to changes that may be detected in the markets in which the company develops its activity. The marketing plan should be seen as a road map.

This does not require us to follow any specific path, but it helps us to see what the best way to achieve our goals is. In case of a blockage in the first route, it easily allows us to adjust the path and set a new route.

REFERENCES

- Batchelor, B. (1997). Marketing the Information Service. In A. Scammell (Ed.), *Handbook of Special Librarianship and Information Work* (7^a ed.). Aslib.
- Dibb, S., Simkin, L., Pride, W., & Ferrell, O. (2006). *Marketing: Concepts and Strategies* (5^a ed.). Boston: Houghton Mifflin.
- Donnelly, J. H., Gibson, J. L., & Ivancevich, J. M. (2000). *Administração: Princípios de Gestão Empresarial*. Lisboa: McGraw-Hill.

Iacobucci, D. (2001). *Kellogg on Marketing: The Kellogg Marketing Faculty*. Wiley.

Iacobucci, D., & Calder, B. (2003). *Kellogg on Integrated Marketing: the Kellogg Marketing Faculty and the Faculty of Integrated Marketing Communications at the Medill School of Journalism*. Hoboken, NJ: Wiley.

Jobber, D., & Ellis-Chadwick, F. (2013). *Principles and Practice of Marketing* (7^a ed.). Berkshire, UK: McGraw-Hill.

Kerin, R., Hartley, S., Berkowitz, E., & Rudelius, W. (2006). *Marketing* (8^a ed.). McGraw-Hill.

Kotler, P. (2000). *Marketing Management: The Millenium Edition*. Prentice Hall.

Kotler, P. (2001). *Marketing para o Século XXI* (2^a ed.). Lisboa: Editorial Presença.

Lindon, D., Lendrevie, J., Rodrigues, J., & Dionísio, P. (2000). *Mercator 2000: Teoria e Prática do Marketing* (9^a ed.). D. Quixote.

Machuret, J., Deloche, D., & D'Amart, J. (1996). *Comerciator: Teoria e Práticas da Via da Qualidade nos Sistemas de Vendas* (1^a ed.). D. Quixote.

Michael, A., & Salter, B. (2006). *Mobile Marketing: Achieving Competitive Advantage through Wireless Technology* (1^a ed.). Oxford, UK: Butterworth-Heinemann.

Mohr, J., Sengupta, S., & Slater, S. (2010). *Marketing of High-Technology Products and Innovations* (3^a ed.). Pearson.

Rea, P., & Kerzner, H. (1997). *Strategic Planning: a Practical Guide*. Van Nostrand Reinhold.

Ryan, D., & Jones, C. (2012). *The Best Digital Marketing Campaigns in the World: Mastering the Art of Customer Engagement* (1^a ed.). Croydon: Kogan Page.

Ryan, D., & Jones, C. (2013). *Understanding Digital Marketing: Marketing Strategies for Engaging the Digital Generation* (2^a ed.). Croydon: Kogan Page.

Silva, R., & Neves, A. (2003). *Gestão de empresas na era do conhecimento*. Lisboa: Sílabo.

Walker, O., Jr., Mullins, J., & Larreche, J.-C. (2008). *Marketing Strategy: A Decision-Focused Approach* (6^a ed.). New York: McGraw-Hill.

Zachert, M., & Williams, R. (1986). Marketing Measures for Information Services. *Special Libraries*, (Spring), 61–70.

ADDITIONAL READING

Badoc, M. (2001). *O Marketing da Start-Up: Conceitos, princípios e métodos para lançar com êxito a sua empresa* (1^a ed.). Lisboa: Bertrand.

Bateson, J., & Hoffman, K. (1999). *Managing Services Marketing: Text and Readings* (4^a ed.). Orlando: Dryden Press.

Brown, E. (2012). *Working the Crowd: Social Media Marketing for Business* (2^a ed.). British Informatics Society.

Carrera, F. (2012). *Marketing Digital na Versão 2.0: o que não pode ignorar* (2^a ed.). Lisboa: Sílabo.

Cravens, D., & Piercy, N. (2008). *Marketing Estratégico* (8^a ed.). S. Paulo: McGraw-Hill.

Cronin, B. (1992). *The Marketing of Library and Information Services 2*. (1^a ed.). Glasgow: Aslib.

Davenport, L., & Cronin, B. (1987). Marketing Electronic Information. *Online Review*, 11(1), 39–47. doi:10.1108/eb024235

Ferreira, M., Reis, N., & Serra, F. (2009). *Marketing para Empreendedores e Pequenas Empresas*. Lisboa: Lidel.

Hortinha, J. (2002). *X-Marketing* (1^a ed.). Lisboa: Sílabo.

Kotler, P. (1986). MegaMarketing. *Harvard Business Review*, (March/April): 117–124.

Kotler, P. (2004). *O Conceito Genérico de Marketing*. *Jornal de Negócios*.

Kotler, P., Wong, V., Saunders, J., & Armstrong, G. (2005). *Principles of Marketing* (4^a ed.). Harlow: Pearson - Prentice Hall.

Lambin, J. (2004). *Marketing Estratégico* (4^a ed.). Lousã: McGraw-Hill.

Roxo, F. V. (1992). *Marketing para que te quero* (2^a ed. ed.): IAPMEI.

Sáez, E. (2002). *Marketing Concepts for Libraries and Information Services* (2^a ed.). Cornwall: Facet Publishing.

Salaün, J. (1992). *Marketing des Bibliothèques et des Centres de Documentation*. Paris: Éditions du Cercle de la Librairie.

Scott, D. (2008). *As Novas Regras de Marketing e Relações Públicas: Como usar os blogues, o podcasting, os media online e as notas informativas para chegar directamente aos consumidores*. Porto: Ideias de Ler.

Stapleton, J. (1990). *Marketing* (3^a ed.). Viseu: Presença.

Viana, C., & Hortinha, J. (2005). *Marketing Internacional* (2^a ed.). Lisboa: Sílabo.

West, C. (2002). Keeping Focused on Our Customers: Marketing as a Tool for Library and Information Managers. In M. Melling & J. Little (Eds.), *Building a Successful Customer-service Culture: A Guide for Library and Information Managers* (1^a ed.). Cornwall: Facet Publishing.

Wood's, M. (2007). *Essential Guide to Marketing Planning* (1^a ed.). Gosport: Prentice Hall/Financial Times (Pearson Education).

Zarrella, D. (2009). *The Social Media Marketing Book* (1^a ed.): O'Reilly Media.

Zarrella, D. (2013). *The Science of Marketing: When to Tweet, What to Post, How to Blog, and Other Proven Strategies*. John Wiley & Sons.

KEY TERMS AND DEFINITIONS

Competitors: A company in the same business or similar industry which offers a similar product or service. Competitors can lower the prices of products and services and gain a larger market share, reducing clients and lowering profits.

Market: The set of offers and searches for a product or service in a particular sector. It's the place where offer and demand get together.

Marketing Plan: A strategic document that will serve to identify the position occupied in the market, set goals and how they will be achieved, all the resources that are needed and expected results. The purpose of this Plan is to enable entrepreneurs to capitalize its resources, spreading the information on the goods and services they provide and channeling the profits derived from such use for their own funding.

Marketing: A social and economic process through which individuals and groups meet their needs and desires by creating and exchanging products and services among themselves. Nowadays, is one of the Marketing action lines more attractive, complex and key success factor in current competitive dynamics.

Positioning: Relates to the set of images through which the market and its customers 'sees' a company, product, person, or even an idea, and clearly identifies it, even among similar ones.

Segmentation: The process of identifying distinct groups of buyers who might require separate products and/or marketing mixes, then creating a profile of each group. Divide the total market of a product or service in a number of homogeneous subsets such as the habits, needs and tastes of its components, calling up each of these subsets as a market segment.

Strategy: The way an organization chooses to do his business in order to surpass competition and gain consumers preference. What is going to do, how, for whom, using what resources, and so forth. It's the plan to conduct business, maximize profits, allocate the needed resources, and stay in the market as long as possible. Includes foresee

what can happen, using forecasting methodologies and tools, in order to avoid problems or, if they happen, be prepared to minimize their effects.

ENDNOTES

- ¹ “The notion of need translates the attitude of the potential consumer that is waiting or looking for anything; this state of frustration will only be overcome with the possibility of (...) having the desired good or service” (Machuret, Deloche, & D’Amart, 1996, p. 39)
- ² Desire can be described as the outward expression of need or personal will to satisfy a need (Donnelly, Gibson, & Ivancevich, 2000)
- ³ The product is the good or service that meets one or more requirements to satisfy

the necessity or desire (Dibb, Simkin, Pride, & Ferrell, 2006; Jobber & Ellis-Chadwick, 2013; Kerin, Hartley, Berkowitz, & Rudelius, 2006; Kotler, Wong, Saunders, & Armstrong, 2005; Lindon, Lendrevie, Rodrigues, & Dionísio, 2000)

- ⁴ “The market is defined as the set of offers and searches for a product or service in a particular sector” (Machuret et al., 1996, p. 39)
- ⁵ “Intellectual Capital - refers to the capacity, skill, experience and also the formal knowledge of people working in the company” (Silva & Neves, 2003, p. 407)
- ⁶ According to Batchelor (1997, p. 373) there are three stages when performing the iterative marketing goods or services: “These stages are marketing strategy, service delivery and implementation of the marketing plan”.

A Neuroaesthetic Approach to the Search of Beauty From the Consumer's Perspective

M

Gemma García Ferrer
Rey Juan Carlos University, Spain

INTRODUCTION

Consumers pursue hedonism and beauty throughout the decision-making processes regarding purchases and consumption. Even though rationality and functionality are part of these processes, the role of emotions is also evident. The five senses (sensory marketing) will be essential in the aesthetic perception that consumers have in these processes.

Therefore, consumers will go to points of purchase which they consider attractive, such as emblematic shopping streets, shopping malls with magnificent architecture, well-decorated and well-furnished designer brand stores, window displays which show the most important features of a store without being necessary to go inside it, well-lit and tidy supermarkets and so on.

Consumers want the packaging of the products to satisfy their needs of making a gift to other people or to themselves. They also want the color combination to be attractive and to boost their mood or the mood of the person who they would like to please.

The product needs to be visually appealing—we can even think of something as simple as a piece of fruit, or something much more sophisticated, such as a cellular phone.

Advertising strategies (billboards, commercials...) need to stimulate this quest for beauty.

Marketing strategists have been aware of this reality for a long time. However, the new neuromarketing and neuroaesthetic techniques can be useful complements to understand the consumers' quest for beauty.

The objectives of this article are to analyze the aforementioned consumers' quest for beauty, by specifying neuromarketing's possibilities and techniques, and to define the concept of neuroaesthetics and its usefulness.

BACKGROUND

A good starting point is the research carried out by Wang, Cruthirds, Axxin & Guo, C. (2013) about the essential value of aesthetics in consumer marketing. According to these authors, there has been disagreement among scholars regarding this aspect throughout the years. Nevertheless, this concept is completely accepted nowadays.

Strategic marketing and specific actions on marketing variables are heading in this direction.

As a useful tool to obtain information throughout the marketing process (Ferrer G.G., 2012), market research uses conventional or non-conventional (more innovative) techniques to increase the aesthetic value of the items to be offered to the consumer.

Consumers will appreciate beauty and aesthetics from the moment when the purchasing process begins. Everything will be sensorial stimuli which consumers perceive in a unique and different way. The five senses will become essential and a company's success will depend, to a great extent, on the fact that the individual's sensory sensations are satisfactory.

The product and everything that surrounds it should be aesthetic and beautiful in the eyes of the consumers. Throughout the chapter, the value

of aesthetics in the point of purchase, packaging, advertising strategy and in the product itself will be analyzed.

As we will see throughout the chapter, non-conventional market research about neuromarketing and neuroaesthetics will obtain physiological and biological measurements as a sign of the individuals' feelings: how they perceive advertising, billboards, promotions, packaging, brands, the place of purchase or the product itself.

MAIN FOCUS OF THE ARTICLE

Aesthetics in the Points of Purchase

Concerning the purchase of convenience products, it is usual that consumers look for a good price. However, this is not the only factor that has an influence on the products' purchase. In fact, we can list several aspects which are directly related to a delightful aesthetic combination, such as lighting, well-organized corridors and shelves and a rich combination of smells (Aldana, Serpa, Ortega & Ochoa, 2010). These factors also apply to online shopping, as it should not only be user-friendly and reliable, but also visually appealing.

Something similar will happen with the purchase of products which are not frequently needed by consumers but are more related to personal feelings of self-satisfaction. Consumers will not only see all their wishes fulfilled with the product itself. As a matter of fact, their emotions (quests for pleasure and wish fulfillment) will expand to the act of "going shopping". Such act will be accompanied by a wide range of sensory stimuli offered by the point of purchase (music, smells, possibility of seeing and touching the product, enjoying a specific atmosphere, interacting with the salesperson...) Walking along a shopping street means coming into contact with the landmarks of a city. This does not only satisfy the consumers' wishes to purchase something, but it brings those consumers closer to the world of architecture,

culture, history and art. For this reason, sensory stimuli are not only produced by the business establishment.

The physical marketing channel plays a fundamental role in the consumption and purchasing experience, and it enables consumers to come into contact with the real atmosphere of the brand.

The most important brands have their own stores and extend the concept of brand to the business establishment itself. In this regard, single-brand stores will take special care of the brand image, and this will show in aspects such as window displays, decoration, furniture and the product arrangement.

The concept of flagship store constitutes a usual and powerful tool of communication, which is supported by geomarketing studies. Flagship stores are located in landmarks of big cities, as well as in heavy pedestrian traffic areas, popular with both city residents and tourists. These stores represent the cutting edge of innovation and creative design. The term "flagship" comes from nautical terminology and was used for the first time in the world of fashion to define stores "which can wave the flag with the brand image". These shops' area ranges between 1,500 and 2,000 square meters (approximately 16,000-21,000 square feet) and their architecture brings the observer closer to the world of culture, art and glamour, as flagship stores are often located in renovated and refurbished historical buildings. Their interior displays tasteful decoration, the latest (and some of the past) fashion collections and advanced technologies such as interactive presentations with augmented reality. Other types of services, which increase the experiential and hedonic value in the purchasing process, are offered in these stores. Such services include private areas, spas, art galleries, cafeterias or restaurants, live music, exhibitions, fashion shows, talks and tasting sessions, among other possibilities.

Pop up stores (Lassus & Freire, 2014) are ephemeral nature stores which open for a short period of time and which also belong to the luxury

market. Pop up stores are more than simple stores: they are also a show, something special and extravagant. Thanks to them, the brand image can be transmitted with a minimum investment risk and low costs. Their temporary locations are usually placed in city landmarks, aiming at large turnouts of people.

Regarding virtual stores of products which are not frequently purchased, the aesthetic combination is also extremely important.

Aesthetics in Packaging

Product packaging does not only play the roles of containing and protecting the product, being functional and looking for constant innovation. As a matter of fact, it is also part of the product design and it is a key element in its sale. Furthermore, we need to take into account the high-level competition in all sectors of the market (Stalman, 2005). The packaging and its careful handling is essential, not only because of the information that it must contain, but also because of the message shown in it, its shape and its colors, which need to arouse hedonic values and emotions.

There are different types of packaging: those of simple products like bottles of water, which need to transmit purity, or those which are more sophisticated, such as perfumes or cosmetic products. In the latter, consumers' aspirations are more related to improving their self-concept and are closely connected to personality.

With this more sophisticated type of packaging, consumers aim at making a gift to other people or to themselves. Therefore, wishes and illusions skyrocket and aesthetics becomes much more important (Martín & Martín, 2012). We just need to think about the gifts we buy for key occasions like Christmas or St. Valentine's Day.

One of the aspects which have a greatest influence on packaging aesthetics is color and its psychological implications. The effects of colors on the consumer are summarized below:

1. **White:** Together with black, it is located on the extreme point of the grey range. White can convey peace, happiness, activity, purity and innocence. White is the universal background of graphic communication.
2. **Black:** It symbolizes silence, mystery and sometimes certain malignity. It confers nobility and elegance, especially when it is bright.
3. **Grey:** It is neutral and passive. It symbolizes indecisiveness and lack of energy. It can also convey doubt and melancholy.
4. **White and Black:** Along with gold and silver, are the colors of prestige. Red or purple can have a similar effect.
5. **Yellow:** It is the brightest, warmest, the most blazing and expansive color. It is also lively, cheerful, exciting, emotional and impulsive. In addition, it is related to nature.
6. **Orange:** It has a cozy, warm, stimulating character and very positive and energetic dynamic qualities.
7. **Red:** It conveys vitality. It is the color of blood and passion. It expresses sensuality, and is the symbol of burning and overflowing passion, sexuality and eroticism.
8. **Blue:** It is part of the cool colors. It conveys harmony, friendship, loyalty, calm and serenity. This color is associated with the sky, the sea and the air. Light blue can suggest optimism.
9. **Violet (Mix of Red and Blue):** It is the color of lucidity and reflection. It is mystical and melancholic, and it could also represent introversion. When it gets closer to purple, it projects a feeling of majesty.
10. **Green:** It is the calmest and the most sedating color. It brings to mind vegetation, freshness and nature. It is associated with hope. It is reminiscent of the fall atmosphere.

Each packaging color can cause a different reaction.

Aesthetics in the Product Itself

Our introduction pointed out that the quest for aesthetics or beauty in a product occurred with items as simple as a piece of fruit. For example, consumers argue that one of the disadvantages of ecological products is their less attractive appearance. However, aesthetics is also sought with more sophisticated products, such as computers or cellular phones.

Furthermore, product design needs to trigger experiences, to please the consumer and to stimulate the five senses: taste, sound, smell, touch and sight. This is sensory marketing, which will act in one way or another, depending on the type of product (Abril, Avello & Manzano, 2011; Krishna & Schwarz, 2014).

When we purchase a product, our intention is to increase our happiness, improve our self-concept and make other people have a superb impression of ourselves. Beauty is pursued in all aspects surrounding the individual and everything that we purchase: furniture, clothing, technology, convenience products and so on.

Just as with packaging, shape and color also contribute to aesthetics (Blijlevens, Carbon, Mugge & Schoormans, 2012).

Aesthetics in Advertising Strategy

The way of making the product known, either in the point of purchase, as we have previously mentioned, or via conventional advertising, has to contribute to the beauty perception of the consumer.

There are simple examples which show this concept, perhaps to a greater extent in the case of advertising aimed at women, as they are more concerned about reinforcing their own self-concept: perfumes, cosmetic products, clothes and accessories.

Anyhow, a billboard or a commercial should appeal to beauty and be attractive for the consumer. As we shall see later, neuroaesthetics has been

applied in the cinema industry and in the same way, it can be applied to advertising.

Advertising is an everyday phenomenon, which obviously has, among others, the function of fostering consumption. However, advertising should also be understood as an art –we seek in it seduction, fantasy, and ultimately, creativity. On the one hand, it is extremely criticized, but on the other hand, it stays in the viewer's retina (González Valerio, 2007).

Neuromarketing, Its Techniques and Its Usefulness in the Understanding of Beauty Assessment by the Consumer

Conventional market research techniques (Malhotra, 2008) do not manage to understand the significance of the consumer's behavior. For this reason, further procedures belonging to neuroscience have been used for several years. Such procedures are known as neuromarketing. Among other authors we can mention the following: Al Pop & Iorga (2012); Braeutigam (2005); Braidot (2009); Hammou, Galib & Melloul (2013); Nakagawa, T. (2011); Senior, C. & Rippon, G. (2007); Stewart & Furse (1982).

Neuroscience and Marketing converge into Neuromarketing, whose scope is multidisciplinary: Economics, Marketing, Sociology, Anthropology and Psychology. Neuromarketing revolutionizes Market Research and its applications to the whole Marketing process (analytical marketing, strategic marketing and operational marketing). It enables managers to take decisions related to their product or service in a very effective way. It is applied by doing neurological, physiological and biological tests to individuals. Therefore, it is much more difficult that such individuals lie, like in conventional market research. These tests are normally done in a laboratory. Next, some of these procedures will be listed:

- **Eye Tracking:** It can measure eye movements of an individual and track their point

of gaze (Duchowski, 2002, 2007). It has several applications: visualization of an advertisement or a website (usability studies) or eye gaze tracking analysis in a supermarket shelf.

- **Pupilometer:** It measures the variations of the pupil diameter as a response to a stimulus, such as advertising. Eye Tracking and Pupilometer techniques can be used together. There are laboratories that combine the use of both approaches (Carbon, Hutzler & Minge, 2006).
- **Psychogalvanometer:** It measures sweat variations, also as a response to a stimulus.
- **Voice-Pitch Analysis:** It enables us to assess emotions, which is very important in marketing.
- Analysis of individuals' saliva when they are looking at images (Sánchez-Navarro, Maldonado, Martínez-Selva, Enguix & Ortiz, 2012).
- **Evoked Potential:** Electrical response of the brain to a sensory stimulus.
- Heart rate.
- **Facial Electromyography:** Movement of facial muscles.
- Eye blinking.
- Skin conductance.
- **Functional Magnetic Resonance Imaging (fMRI):** It is a functional neuroimaging procedure that shows images of brain areas which perform certain tasks.

There is an increasing number of multinationals which hire the services of companies and laboratories specialized in Neuromarketing.

This list gives us an idea of the possibilities of neuromarketing and the ethical debate around it. It is understood that there is a need for updated codes of ethics which regulate the use of these procedures, as well as their use along with conventional research methods (Murphy, E. R., Illes, J., & Reiner, P. B., 2008). What is evident is that any of the aforementioned procedures can

help us to get to know the consumer's appreciation of beauty.

Neuroaesthetics and its Usefulness in Understanding the Consumer's Appreciation of Beauty

One of the disciplines belonging to neuroscience is neuroaesthetics or neurobiology of beauty. It arises from the convergence between neuroscience, aesthetics and creativity and is supported by psychobiology. It can be applied to all kinds of art: painting, sculpture, music, dance, cinema... Neuroaesthetics (normally) measures how individuals perceive beauty by means of brain responses. For this reason, it can also measure how the consumer perceives the creative aspects of the marketing strategy. All the creative arts can benefit from this new science and in fact, marketing is one of them: luxury products, conventional products, billboards, commercials...

Semir Zeki is the pioneer of neuroaesthetics. In his research he measured several individuals' neural responses to different stimuli such as paintings or photographs, analyzing the increasing areas of the neural activity in regard to beauty perception (Kawabata and Zeki, 2004). According to Semir Zeki (2009), creativity is determined by the attempt to satisfy the dissatisfied brain concept. Permanent dissatisfaction is one of the most powerful ingredients for creativity.

Neuroaesthetics has even been taken to film analysis (Andreu Sánchez & Martín Pascual, 2011) and as we mentioned before, it can be applied to the field of advertising.

SOLUTIONS AND RECOMMENDATIONS

Consumers pursue pleasure, emotions, beauty and aesthetics in their decisions of purchase and consumption, starting from the moment when they check a website or when they decide to go to a shopping street or mall.

All the marketing strategy needs to head in this direction. Consumers fill up their existential emptiness with their purchases and pursue positive perceptions that make them feel better.

Sensory marketing aims at the consumers' five senses and facilitates this quest for beauty.

In order to analyze marketing strategies and their results, we can use not only conventional research techniques, but also neuroscience techniques (neuromarketing and neuroaesthetics).

FUTURE RESEARCH DIRECTIONS

Future research aims at improving consumers' aesthetic perception. Understanding the individual's physiological and brain mechanism as a response to stimuli is a key element in the success of marketing strategies.

Multidisciplinary studies –in this case, we can see the contribution of psychobiology- play an essential role in the understanding of individuals and their perception processes. Along with other science disciplines and with more conventional techniques belonging to them, it is considered relevant to analyze the impact of beauty on consumption and purchasing mechanisms.

CONCLUSION

Neuroscience disciplines (neuromarketing and neuroaesthetics) are key to understand physiological and brain mechanisms on how consumers process creativity in the case of marketing.

It can be stated that consumers pursue beauty throughout the whole purchasing and consumption process.

In the purchasing process, the consumer will appreciate and perceive the beauty of a marketing strategy through the five senses. Non-conventional techniques of neuromarketing and neuroaesthetics can turn out to be very efficient complements to

conventional research, since they are based on biological and physiological data.

Creativity and aesthetics are fundamental values which should be sought by marketing strategists and managers.

REFERENCES

- Abril, C., Avello, M., & Manzano, R. (2011). Marketing visual: amor a primera vista. *Harvard-Deusto Marketing & Ventas*, (103), 46-51.
- Al Pop, N., & Iorga, A. M. (2012). A new challenge for contemporary marketing - neuromarketing. *Management & Marketing*, 7(4), 631-644.
- Aldana, E. L., Serpa, J. J., Ortega, M. D., & Ochoa, S. (2010). *Estética Y Ambientación De Los Supermercados De La Ciudad De Cartagena*. Revista Jornadas de Investigación.
- Andreu Sánchez, C., & Martín Pascual, M. Á. (2011). *La intertextualidad como nueva herramienta neuroestética de análisis cinematográfico*. Nuevas Tendencias. doi:10.6035/978-84-87510-57-1.2011.81
- Blijlevens, J., Carbon, C. C., Mugge, R., & Schoormans, J. P. (2012). Aesthetic appraisal of product designs: Independent effects of typicality and arousal. *British Journal of Psychology*, 103(1), 44-57. doi:10.1111/j.2044-8295.2011.02038.x PMID:22229773
- Braeutigam, S. (2005). Neuroeconomics – from neural systems to economic behaviour. *Brain Research Bulletin*, 67(5), 355-360. doi:10.1016/j.brainresbull.2005.06.009 PMID:16216681
- Braidot, N. (2009). *Neuromarketing en acción*. Ediciones Granica, SA.
- Carbon, C. C., Hutzler, F., & Minge, M. (2006). Innovativeness in design investigated by eye movements and pupillometry. *Psychological Science*, 48(2), 173-186.

- De Lassus, C., & Freire, N. A. (2014). Access to the luxury brand myth in pop-up stores: A netnographic and semiotic analysis. *Journal of Retailing and Consumer Services*, 21(1), 61–68. doi:10.1016/j.jretconser.2013.08.005
- Duchowski, A. T. (2002). A breadth-first survey of eye-tracking applications. *Behavior Research Methods, Instruments, & Computers*, 34(4), 455–470. doi:10.3758/BF03195475 PMID:12564550
- Duchowski, A. T. (2007). *Eye tracking methodology: Theory and practice* (Vol. 373). London, UK: Springer.
- Ferrer, G. G. (2012). *Investigación comercial*. Esic Editorial.
- González, M. A. (2007). Apuntes para pensar la publicidad como fenómeno estético. *Estudios Filosóficos*, LVI(161), 1–23.
- Hammou, K. A., Galib, M. H., & Melloul, J. (2013). The contributions of neuromarketing in marketing research. *Journal of Management Research*, 5(4), 20–33. doi:10.5296/jmr.v5i4.4023
- Kawabata, H., & Zeki, S. (2004). Neural correlates of beauty. *Journal of Neurophysiology*, 91(4), 1699–1705. doi:10.1152/jn.00696.2003 PMID:15010496
- Krishna, A., & Schwarz, N. (2014). Sensory Marketing, Embodiment, and Grounded Cognition: Implications for Consumer Behavior. *Journal of Consumer Psychology*, 24(2), 159–168. doi:10.1016/j.jcps.2013.12.006
- Lee, N., Broderick, A. J., & Chamberlain, L. (2007). What is neuromarketing? A discussion and agenda for future research. *International Journal of Psychophysiology*, 63(2), 199–204. doi:10.1016/j.ijpsycho.2006.03.007 PMID:16769143
- Malhotra, N. K. (2008). *Marketing Research: An Applied Orientation*. Pearson Education India. doi:10.1108/S1548-6435(2008)4
- Martín, I. R., & Martín, L. R. (2012). Envolviendo sueños. Diseño de Packaging: Ética, estética y eficacia. *Creatividad y sociedad: revista de la Asociación para la Creatividad*, (19), 5.
- Murphy, E. R., Illes, J., & Reiner, P. B. (2008). Neuroethics of neuromarketing. *Journal of Consumer Behaviour*, 7(4-5), 293–302. doi:10.1002/cb.252
- Nakagawa, T. (2011). *The feasibility study of applied neuroscience for advertising and marketing research* (Unpublished doctoral dissertation). Massachusetts Institute of Technology.
- Ramachandran, V. S., & Hirstein, W. (1999). The science of art: A neurological theory of aesthetic experience. *Journal of Consciousness Studies*, 6(6-7), 15–51.
- Sánchez-Navarro, J. P., Maldonado, E. F., Martínez-Selva, J. M., Enguix, A., & Ortiz, C. (2012). Salivary alpha-amylase changes promoted by sustained exposure to affective pictures. *Psychophysiology*, 49(12), 1601–1609. doi:10.1111/j.1469-8986.2012.01475.x PMID:23078627
- Senior, C., & Rippon, G. (2007). Cognitive neuroscience: Contributions from psychophysiology. *International Journal of Psychophysiology*, 63(2), 135–137. doi:10.1016/j.ijpsycho.2006.03.017 PMID:16815580
- Stalman, A. (2005). Packaging, algo más que estética. *MK: Marketing+ Ventas*, (199), 40–41.
- Stewart, D. W., & Furse, D. H. (1982). Applying psychophysiological measures to marketing and advertising research problems. *Current Issues and Research in Advertising*, 5(1), 1–38.
- Wang, Y. J., Cruthirds, K. W., Axinn, C. N., & Guo, C. (2013). In Search of Aesthetics in Consumer Marketing: An Examination of Aesthetic Stimuli from the Philosophy of Art and the Psychology of Art. *Academy of Marketing Studies Journal*, 17(2), 37.

Zeki, S. (2009). *Splendors and miseries of the brain: Love, creativity and the quest for human happiness*. Chichester, UK: Wiley-Blackwell.

ADDITIONAL READING

Brown, S., Gao, X., Tisdelle, L., Eickhoff, S. B., & Liotti, M. (2011). Naturalizing aesthetics: Brain areas for aesthetic appraisal across sensory modalities. *NeuroImage*, 58(1), 250–258. doi:10.1016/j.neuroimage.2011.06.012 PMID:21699987

Chatterjee, A. (2011). Neuroaesthetics: A coming of age story. *Journal of Cognitive Neuroscience*, 23(1), 53–62. doi:10.1162/jocn.2010.21457 PMID:20175677

Krishna, A. (2012). An integrative review of sensory marketing: Engaging the senses to affect perception, judgment and behavior. *Journal of Consumer Psychology*, 22(3), 332–351. doi:10.1016/j.jcps.2011.08.003

Nalbantian, S. (2008). Neuroaesthetics: Neuroscientific theory and illustration from the arts. *Interdisciplinary Science Reviews*, 33(4), 357–368. doi:10.1179/174327908X392906

Penn, D. (2011). ¿Qué aporta la neurociencia a la investigación de mercados. *Research World*, 26, 66–67.

Ramachandran, V. S., & Hirstein, W. (1999). The science of art: A neurological theory of aesthetic experience. *Journal of Consciousness Studies*, 6(6-7), 15–51.

Vecchiato, G., Di Flumeri, G., Maglione, A. G., Cherubino, P., Kong, W., Trettel, A., & Babiloni, F. (2014, August). An electroencephalographic Peak Density Function to detect memorization during the observation of TV commercials. In *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE* (pp. 6969-6972). IEEE. doi:10.1109/EMBC.2014.6945231

Vecchiato, G., Kong, W., Maglione, A. G., & Wei, D. (2011). Understanding the impact of TV commercials: electrical neuroimaging. *IEEE pulse*, 3(3), 42-47.

KEY TERMS AND DEFINITIONS

Beauty: It is understood as something that gives pleasure and satisfaction to the consumer from the aesthetic point of view.

Creativity: Basis for art, originality and imagination.

Neuroaesthetics: Measurement of the neural response of an individual to beauty.

Neuromarketing: Application of neuroscience and its methods to marketing and in particular, to the study of emotions.

Psychobiology: Branch of psychology which measures biological and physiological responses of an individual.

Sensory Marketing: Use of the five senses to capture the attention of the consumer.

Strategic Marketing: Planning of the segmentation and positioning in the consumer's mind, in order to determine afterwards the marketing mix in the short term.

Social Media Use and Customer Engagement

M

Aurora Garrido-Moreno
University of Malaga, Spain

Nigel Lockett
University of Lancaster, UK

Víctor García-Morales
University of Granada, Spain

INTRODUCTION

Social Media has revolutionized customer relationships; opening a new array of possibilities and opportunities for firms. Social media tools are fundamentally changing the way we communicate, collaborate, consume, and create; and they represent one of the most transformative impacts of information technology on business (Aral et al., 2013). Platforms such as Facebook, Twitter, YouTube, Instagram, Pinterest, have millions of daily users and are said to capture the “wisdom of the crowd” (Luo et al., 2013). So firms all over the world are increasingly using these tools, integrating them into their communication strategies to engage customers in active conversations. In this new 2.0 environment, traditional Customer Relationship Management (CRM) strategies are morphing to Social CRM. The latest trend in CRM is to try to take advantage of Social Media, whose relational properties and characteristics are particularly suited to customer interactions (Harrigan et al., 2015). Social Media is all about engagement and collaboration. The proliferation of these platforms has changed the way firms communicate with customers, driving companies to implement a new variety of Customer Engagement strategies. New media offers companies multiple ways to reach consumers, communicate

with them, and measure their preferences and purchase-related behaviors (Hennig-Thurau et al., 2010). Based on this personalized knowledge, Social Media enables firms to reinvent customer relationships, by engaging customers in marketing campaigns, product development and innovation. Customer Engagement emerges as a core concept and implies customers’ involvement in a new type of relationship with the enterprise, characterized by co-creation of knowledge, offers, products and value (Giannakis-Bompolisa and Boutsouki, 2014). It can be considered an intimate long-term relationship with customers that results in different behavior such as word-of-mouth (WOM), blogging, writing reviews and recommendations or helping other customers (Van Doorm, 2010).

Social media platforms can facilitate business transformation, improving the management of customer relationships, brand assets, and business processes. However, there is a growing skepticism about its real efficacy, and there is a lack of evidence that confirms a return on investment for companies that have invested in this technology (Rishika et al., 2013). To justify the significant resources and investments in Social Media, executives need to develop and quantify Social Media metrics, such as Customer Engagement, to empirically examine if these tools are creating value for the firm (Luo et al., 2013). Social Media use is a

very recent phenomenon among firms at the real world and, consequently, research on the topic is in its early stages. Additional studies are needed to better understand how firms can leverage Social Media use to create business value (Braojos-Gómez et al., 2015). To shed light on the topic, in this study we seek to answer the following research questions: 1) What are the main Social Media tools that companies are using?; 2) What is the impact of using these tools in Customer Engagement, and how can we measure it? Both questions are of practical importance because companies are investing considerable time, money and effort in their implementation of those tools, but little is known about their benefits. To examine the topic, we conducted a multiple linear regression analysis based on a sample of 212 hotels in Spain.

The article is organized as follows. After a literature review, the empirical study and main results are described. The conclusions, along with some future research lines close the study.

BACKGROUND

Social Media Use

In the last decade, Social Media has emerged as a real ‘game changer’, putting the customer at the center of the organization (Smith and Zook, 2011). It could be argued that Social Media represents a revolutionary new trend for companies (Kaplan and Haenlein, 2010), and some even assume it as the biggest change since the industrial revolution. Social media has changed the way individuals interact with each other and is redefining the way firms connect with their customers (Rishika et al., 2013).

Deepening on the concept, Social Media refers to a set of online tools that supports social interaction between users, facilitating the creation and sharing of knowledge, and transforming monologue (company to customer) into dialog (Hansen et al., 2011). One of the most accepted definitions of Social Media is the one proposed by Kaplan

and Haenlein, (2010, p. 61), who consider it as “a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of User Generated Content”. Thus, Social Media are browser or mobile-based applications that allow users to easily create, edit, access and link to content and/or to other individuals (Cabiddu et al., 2014). Its content comprises text, pictures, videos, and networks and includes tools such as discussion forums, blogs, microblogging, texting, chat, social networking sites, wikis, photo and video-sharing sites, and review sites (Hennig-Thurau et al., 2010). All these Social Media tools support the democratization of knowledge and information, transforming individuals from mere content consumers into content producers (Berthon et al., 2012). Social media enables firms to converse about their offerings with customers interactively while also providing them the opportunity to share information about products, so it has provided firms with a new tool for customer engagement (Luo et al., 2013). Although those tools are not specifically designed for organizational purposes, companies are integrating their use into their CRM strategies in order to augment customer-firm interactions and reinvent customer conversations.

Social Media have become incredibly popular in recent years. Facebook is the largest Social network in the world. The number of daily active Facebook users has already reached 890 million by December 2014, an increase of 18% over the prior year (Facebook, 2015). User-generated content (UGC) has become a mass phenomenon, with Facebook, YouTube, Wikipedia, and Twitter all being listed among the Top 10 global websites, in number of visitors and page views (Alexa, 2015). This surge in popularity has extensively increased online user-generated content (UGC) or word of mouth (WOM) and hence, attracted marketers’ attention. Burson-Marsteller (2012) empirically examined Fortune Global 100 companies’ social media activity on Twitter, Facebook, YouTube, Google Plus and Pinterest. They observed how

the vast majority of companies (87%) were using at least one of the major social platforms to communicate with online stakeholders, mainly Twitter (82%) and Facebook (74%). The study shows also an interesting trend in the increase of content creation among the examined companies. Since 2010, it has been remarkably how those companies' use evolved from broadcasting to customer engagement with content creation. Companies are now integrating more original multimedia content to share with followers on Twitter, Facebook and YouTube (Burson-Marsteller, 2012).

Parallel to this growth in the business world, research on Social Media has accelerated in the last few years. Some areas of research are rapidly developing, whereas others are nascent or nonexistent (Aral et al., 2013). Despite this quick development in recent years, the study of Social Media is still considered to be in its initial stages (Braojos-Gómez et al., 2015). In this vein, different authors highlight that measuring the impact and value of Social Media on individuals and firms warrants a great deal of additional attention (Dong and Wu, 2015; Rapp et al., 2013). Despite the prevalent use of Social Media by consumers and marketers, empirical research investigating its economic value still lags (Goh et al., 2013). Prior marketing research suggests that recent metrics, such as Customer Engagement, are expected to have a stronger predictive relationship with firm value creation, but empirical evidence on the topic remains scarce (Cabiddu et al., 2014).

Customer Engagement

Customer Engagement has become one of the most important research topics for marketing scholars today (Van Doorn et al., 2010). In essence, it focuses on satisfying customers by providing superior value to build trust and commitment in long-term relationships. Engaged customers become partners who collaborate with sellers in the value-adding process to better satisfy their needs, as well as the needs of other customers (Sashi, 2012).

Given that the concept of Customer Engagement is relatively novel and is still in a developmental phase, there are differing and at times conflicting opinions regarding its conceptualization (Kumar et al., 2010). The most accepted conceptualization, which has been more cited in recent literature is the one proposed by Van Doorn et al. (2010). These authors define Customer Engagement as the *customers' behavioral manifestation toward a brand or firm, beyond purchase, resulting from motivational drivers. It include a vast array of behaviors including word-of-mouth (WOM) activity, recommendations, helping other customers, blogging, writing reviews, and even engaging in legal action* (Van Doorn et al., 2010, p. 253). Customer Engagement is considered as a behavior that is not limited to simple transactions, because engaged customers contribute to firms in many ways that are beyond direct purchases (e.g., WOM, new product ideas, etc.). Porter et al., (2011) proposed a behavioral definition of Customer Engagement and define it as a class of behaviors that reflects community members' demonstrated willingness to participate and cooperate with others in a way that creates value for themselves and for others—including the community sponsor. Customer Engagement emerges as a new core focus for firms in the current environment and entails an emotional involvement of customers to a new type of relationship with the firms, characterized by transparent communications and co-creation of knowledge, products and value.

Embracing an inclusive, behavioral definition of the concept, we considered that Customer Engagement includes: word-of-mouth (WOM) activity, browsing and consuming consumer-generated media contents, content contribution, active participation in online communities, user-generated reviews and recommendations (Cabiddu et al., 2014; Van Doorn et al., 2010). The interactive nature of Social Media with its ability to establish conversations among individuals and firms and involve customers in content generation and value creation has excited practitioners with its potential to better serve customers and

satisfy their needs (Sashi, 2012). For managers, achieving a high level of consumer engagement is viewed as desirable, because it may enhance not only a company's reputation and brand loyalty, but also customer purchase decisions (Dijkmans et al., 2015). The relational consequences of consumer engagement may include commitment, trust, consumers' emotional brand attachment and loyalty. Consequently, executives not only believe that high Customer Engagement is necessary for future growth, they also believe that low Customer Engagement is detrimental to success, both due to lost sales and negative WOM (Kumar et al., 2010).

Social Media use is now a golden opportunity for firms and it represents a cornerstone to start building Customer Engagement. In the following section presents an empirical study covering both topics.

EMPIRICAL STUDY

Methods and Data Collection

The hotel industry was chosen for conducting the empirical study because tourism is an important and extremely competitive sector, which is characterized by its innovative behavior (Orfila-Sintes and Mattsson, 2009). In addition, there is an overwhelming requirement for academic exploration into how hotels can exploit Social Media tools, for improving customer relationships (Jung et al., 2013). Hospitality is one of the most relevant activities within the tourism sector, which has itself become increasingly important, as a source of economic and social development. This study looks particularly at hotels within Spain; since it is considered a World's Top Tourism Destinations as top the rankings by both international arrivals and receipts. Following data from the World Tourism Organization, in 2014 Spain was the second country in tourism earnings worldwide (and first in Europe) with US\$ 65 billion in 2014, and was third in arrivals with 65 million overnight visitors (World Tourism Organization, 2015).

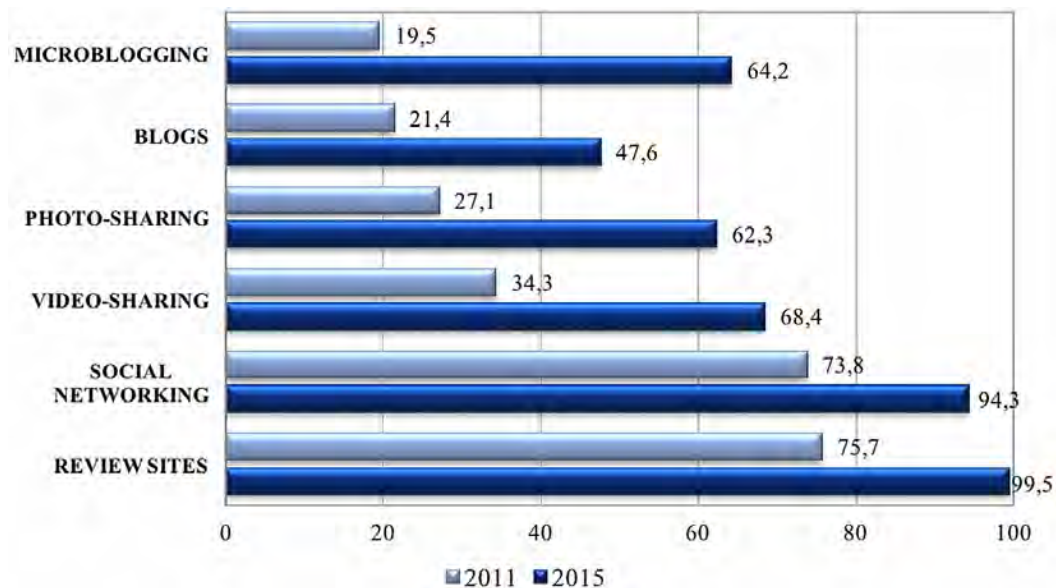
In order to empirically examine the phenomenon under study, a sample of 1000 hotel businesses was selected randomly from a database from Turespaña (Spain's Ministry of Tourism). The key-informant methodology was employed to collect the appropriate information, choosing the hotel managers as informants. Based on the literature review, we drew up a structured questionnaire and it was sent by mail to hotel general managers. We obtained a sample of 212 valid responses, which gave an approximate response rate of 21.2%. The possibility of non-response bias was checked by comparing the characteristics of the respondents to those of the original population sample. A series of chi-square and t-statistics revealed no significant differences between the respondents and the sample, between early and late respondents or based on the type or size of hotel. We also analyzed the non-existence of significant differences based in the geographical location of hotel.

Regarding the profile of the respondents, focusing on the size of the hotels, most of them were small hotels: 75.9% had less than 50 employees, and 24.1% were medium and large hotels. The majority of the sample has three (45.8%) and four stars (47.2%), and with regards to the ownership or property regime of the hotels, most of them were independent (67%) while the rest were integrated in chains.

Use of Social Media Tools

To examine the usage of Social Media tools in the studied hotels, the following tools were included in the questionnaire, as they proved to be the most used in the hotel sector (Sigala, 2011): blogs, microblogging (Twitter), social networking, video-sharing, photo-sharing and review sites. First at all, we examined the percentage of hotels who affirmed be using this tools for managing customer relations. As the authors conducted a previous study on the topic, also in the Spanish hotel sector, some years ago (Garrido-Moreno and Lockett, 2014), we compared longitudinally

Figure 1. Longitudinal analysis of the use of Social Media tools



how the use of those tools has evolved. Prior data was collected in 2011, based on a similar sample (N=210).

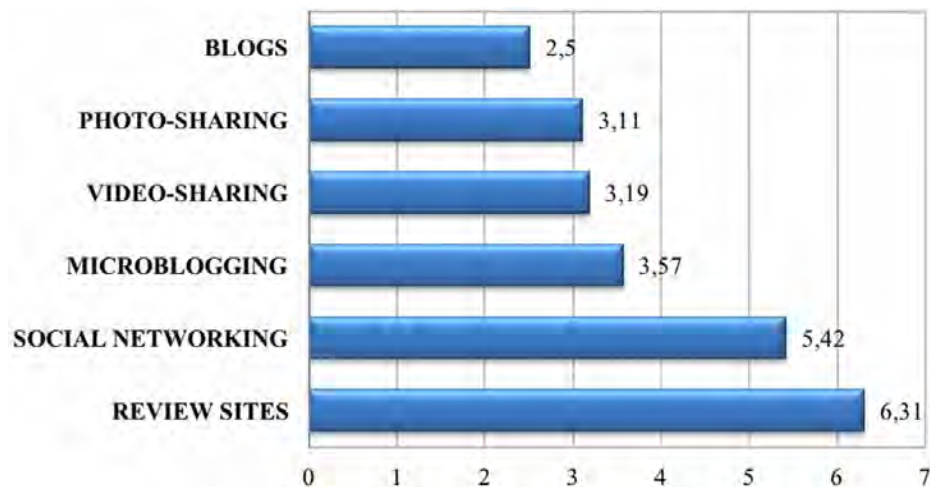
Figure 1 shows the use of these tools in hotels both in 2011 and 2015 (expressed in percentage), so we can observe their different level of penetration in the studied industry.

The most used tools in both periods were review sites (more than 99% of hotels reported using tools like TripAdvisor in 2015) and social networks. It was observed also how the use of most of the tools experienced a great increase among those years. Specifically, microblogging (Twitter) was the tools that experienced higher growth (from 19.5% of hotels to 64.2% in 2015). In order to know if the temporal differences between the use of those tools were relevant and statistically significant, the *Mann-Whitney U* test on difference between means was applied, as the data were not normal. We conducted these analyses for the six mentioned tools. As the bilateral critical level was lower than 0.05 ($p=0.00$ in all the cases), it was accepted that the difference of means was statistically significant in the use of all the tools examined: blogs, microblogging, social networking, video-sharing, photo-sharing and review sites.

Focusing on data obtained in 2015, we also asked hotels about their frequency of use of those Social Media tools, in order to obtain a more refined and accurate measure of their actual use of them. We included a specific question in the questionnaire and the range for measuring how often do they use each Social Media tool ranged from 1 (never) to 7 (every time). The average frequency of use of each tool is reflected in Figure 2.

We can see how review sites and social networks were, by far, the tools used with more frequency (their averages values were 6.41 and 5.32 respectively). So sampled hotels have integrated them into their daily activity in order to communicate with customers, gaining knowledge from them and establishing valuable conversations. Tools like microblogging, video-sharing and photo-sharing were used more occasionally and blogs were the less frequently used platform. If we compare the level of penetration of the different tools and their frequency of use, we found that results seem coherent. However, it is surprising to observe how microblogging, despite their high level of penetration (64.2% of hotels were using it in 2015); was not used as frequently as other tools (average frequency of use: 3.57 upon 7).

Figure 2. Frequency of the use of Social Media tools



Results show how the use of Social Media tools has become massive, especially in tourism firms, like hotels, where customers are making increasing use of the Internet as a tool to express their experiences with organizations, and potential consumers make purchasing decisions, on the basis of the information found on the Internet (Haro de Rosario et al., 2013).

We have observed how review sites, like TripAdvisor, have the higher level of penetration in this sector, and are the most frequently used tool. TripAdvisor-branded sites comprise the largest travel community in the world, with user-generated recommendations and international hotel reviews attracting more than 260 million users a month (Google, 2014). The system provides users with independent travel reviews and comments written from other members and provides a powerful platform for interaction and sharing experiences (Cabiddu et al., 2014). Consequently, our findings confirm that TripAdvisor handling is a strategic priority for the hotels analyzed.

Apart from review sites, the Social Media platforms that showed the highest frequency of use were social networkings sites and microblogging. These results are coherent with other studies conducted in the Spanish hotel sector (Escobar-Rodríguez and Carvajal-Trujillo, 2013; Haro de Rosario et al., 2013). The greater presence of

hotels in platforms like Facebook and Twitter is easy to understand, because they have the most users and provide a substantive part of the available online word-of-mouth (Dijkmans, 2015). The reported use of video and photo sharing services as Youtube and Flickr was also high (more than 60% of hotels used those platforms), as they allow a dynamic relationship with the users and a greater interaction with them (Escobar-Rodríguez and Carvajal-Trujillo, 2013).

Impact of Social Media use on Customer Engagement

To empirically test this impact, we conducted a multiple linear regression as estimation technique. To measure Customer Engagement, we introduce five items in the questionnaire based on previous studies (Van Doorn et al., 2010; Verleye et al., 2014), referring to: customer feedback; customer involvement in marketing decisions; WOM activity; customer recommendations and strategic relevance of WOM. A summative index of customer engagement was calculated as the average of all these items, and was considered the dependent variable. The frequency of use of the six mentioned tools were introduced as independent variables. Two control variables (firm size and age) were also included in the model.

Table 1. Descriptive statistics and correlations

		Mean	SD	1.	2.	3.	4.	5.	6.	7.	8.
1	Firm size	47.79	76.05								
2	Firm age	24.22	24.25	.208**							
3	Customer Engagement	5.39	1.20	.193**	-.064						
4	Frequency of use of Blogs	2.50	1.89	.166*	-.005	.182**					
5	Frequency of use of Microblogging	3.57	2.38	.136*	-.043	.240**	.501**				
6	Frequency of use of Social Networking	5.42	1.81	.207**	.015	.328**	.401**	.547**			
7	Frequency of use of Video-sharing	3.19	1.99	.232**	.007	.299**	.478**	.473**	.489**		
8	Frequency of use of Photo-sharing	3.11	2.10	.202**	-.067	.213**	.538**	.481**	.455**	.677**	
9	Frequency of use of Review Sites	6.31	1.13	.117	-.074	.362**	0.083	.132	0.118	.128	.121

Notes: N=212; * $p < 0.05$; ** $p < 0.01$ (two-tailed)

The descriptive statistics and correlations for the variables are reported in Table 1.

An assessment of multicollinearity was carried out through the computation of tolerance value and variable inflation factor (VIF). For all the variables, the tolerance values were well above

0.4 cutoff and the VIFs were substantially below 10 cutoff, indicating multicollinearity was not a problem.

Results of the regression (See Table 2) show how the use of three of the six SM tools examined directly influenced Customer Engagement. These

Table 2. Results of multiple regression analysis

Variables	Model 1	Model 2
Step 1: Control Variables		
Firm size	.216**	.101
Firm age	-.109	-.071
Step 2: Dependent Variables		
Frequency of use of Blogs		.001
Frequency of use of Microblogging		.021
Frequency of use of Social Networking		.208**
Frequency of use of Video-sharing		.174*
Frequency of use of Photo-sharing		-.073
Frequency of use of Review Sites		.304***
R ²	.049	.244
Adjusted R ²	.040	.214
ΔR^2	.049***	.174***
F	5.355**	8.18***

Notes: Regression coefficients are reported as standardized β values; [†] $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$.

variables were: social networking ($\beta=.208$, $p < .01$); video-sharing platforms ($\beta=.174$, $p < .05$) and review sites ($\beta=.304$, $p < .001$). Consequently, the use of these variables can be considered a relevant enabler of Customer Engagement. Surprisingly, blogs, microblogging and photo-sharing platforms were not directly related to customer engagement. F test on all adjusted R squared changes were significant; indicating that successive variables added in the regression models significantly improved the prediction for Customer Engagement.

Our findings support prior studies indicating that Social Media technologies are key enablers of Customer Engagement (Harrigan et al., 2015). Nevertheless, we examined the phenomenon in a more detailed way, analyzing the specific impact of the six more relevant Social Media tools. We observed that review sites were the one that showed the highest impact in Customer Engagement. These results are coherent with existing literature that indicates that platforms like Tripadvisor have become essential for managing online reputation, particularly in the hotel sector (Baka, 2016).

Similarly, Social Networking use proved to be determinant in fostering Customer Engagement. Nowadays, channels like Facebook, have changed marketing practices and are used for gaining an audience and communicating campaigns, offers and news (Baka, 2016). Our results are coherent with Dijkmans et al. (2015) and suggest that companies are increasingly trying to get consumers engaged in online discussions by including social networking sites like Facebook into their communication strategy.

Regarding Video-sharing platforms like Youtube, results also show that its use can be considered a determinant of Customer Engagement. YouTube focuses on video clips posted online by private and corporate users, and approximately more than two billion clips are watched daily (Mansson, 2011). Consequently, Youtube is a very dynamic platform that offers hotels multiple possibilities for engaging customers.

Finally, results show that the use of Microblogging, Photo-sharing sites and Blogs seems not to be directly related to Customer Engagement. Surprisingly, despite the increase in the level of use of Twitter, findings show that this use is not having a real impact in enhancing customer-engaged behavior. We have to point out that Twitter enables fast, brief conversations; and sometimes hotels are using it exclusively as a promotional tool (Jung et al., 2013), so it is important to keep content focused, occasional, and appropriate instead of a bombardment, which could be very harmful. It is interesting to note also that Twitter in comparison to Facebook has different user demographics, so hotel managers need to acknowledge this differences and develop exclusive strategies and customize content appropriately to each social network (Jung et al., 2013).

CONCLUSION AND FUTURE RESEARCH LINES

In the current 2.0 environment, Social Media tools have completely changed the way companies relate to the market, opening a wide range of new possibilities. Customers are more connected than ever and are using tools like Facebook, Twitter, Youtube, Flickr, and review sites such as Tripadvisor, on a daily basis. Traditional CRM have evolved towards Social CRM strategies, and companies have integrated Social Media tools into their marketing strategies to engage customers in open and active conversations, building emotional bonds with them. However, despite the eagerness on the part of firms to embrace Social Media, there is also much skepticism about its efficacy.

The present study focuses on the concept of Social Media use and Customer Engagement, exploring the level of use of these tools in a longitudinal way; and examining also how the use of these tools can create value for firms by fostering Customer Engagement.

Regarding the use of Social Media tools, we observed that the most popular tools employed by

the examined hotels were review sites and social networking. Hoteliers actively managed *TripAdvisor* reviews and acknowledged having an active presence on social networks, like *Facebook*. We also examined the phenomenon longitudinally (with data from 2011), and we observed a significant increase in the use of the six analyzed tools. Results of the regression show that three Social Media tools: review sites, social networking and video-sharing platforms appear as key enablers of Customer Engagement. Findings highlight how Social Media technologies are reinventing customer relationships and can be considered a basic enabler of customer engagement, supporting prior research (Harrigan et al., 2015).

This study has interesting implications for academics and practitioners. Given that Social CRM is a topic of practical relevance nowadays, the study provides useful implications for managers, as it shows not only which are the most used Social Media tools, but also which have a bigger impact on customer relationship, enabling customer-enabled behaviors. Based on this evidence, managers can readjust their Social Media strategies, promoting the use of the most effective tools. Regarding future research lines, this study proposes that it would be interesting to analyze the use of these tools in different sectors, comparing the differences between services and manufacturing companies. Secondly, it would be necessary to examine not only the impact of Social Media on Customer Engagement, but also its effect on organizational performance.

REFERENCES

- Alexa. (2015). *The top 500 sites on the web*. Retrieved October 28, 2015 from: <http://www.alexa.com/topsites>
- Aral, S., Dellarocas, C., & Godes, D. (2013). Introduction to the Special Issue—Social Media and Business Transformation: A Framework for Research. *Information Systems Research*, 24(1), 3–13. doi:10.1287/isre.1120.0470
- Baka, V. (2016). The becoming of user-generated reviews: Looking at the past to understand the future of managing reputation in the travel sector. *Tourism Management*, 53, 148–162. doi:10.1016/j.tourman.2015.09.004
- Berthon, P. R., Pitt, L. F., Plangger, K., & Shapiro, D. (2012). Marketing meets Web 2.0, social media, and creative consumers: Implications for international marketing strategy. *Business Horizons*, 55(3), 261–271. doi:10.1016/j.bushor.2012.01.007
- Braojos-Gomez, J., Benitez-Amado, J., & Llorens-Montes, F. J. (2015). How do small firms learn to develop a social media competence? *International Journal of Information Management*, 35(4), 443–458. doi:10.1016/j.ijinfomgt.2015.04.003
- Burson-Marsteller. (2012). *Global social media check-up 2012*. New York: Burson-Marsteller. Retrieved October 28, 2015 from <http://www.burson-marsteller.com/social/PressRelease.aspx>
- Cabiddu, F., De Carlo, M., & Piccoli, G. (2014). Social Media Affordances: Enabling Customer Engagement. *Annals of Tourism Research*, 48, 175–192. doi:10.1016/j.annals.2014.06.003
- Dijkmans, C., Kerkhof, P., & Beukeboom, C. J. (2015). A stage to engage: Social media use and corporate reputation. *Tourism Management*, 47, 58–67. doi:10.1016/j.tourman.2014.09.005
- Dong, J. Q., & Wu, W. (2015). Business value of social media technologies: Evidence from online user innovation communities. *The Journal of Strategic Information Systems*, 24(2), 113–127. doi:10.1016/j.jsis.2015.04.003
- Escobar-Rodríguez, T., & Carvajal-Trujillo, E. (2013). An evaluation of Spanish hotel websites: Informational vs. relational strategies. *International Journal of Hospitality Management*, 33, 228–239. doi:10.1016/j.ijhm.2012.08.008
- Facebook. (2015). *Facebook 2014 Annual Report*. Retrieved October 28, 2015, from: <http://investor.fb.com/annuals.cfm>

- Garrido-Moreno, A., & Lockett, N. (2014). Customer Relationship Management and Social Media use: in search of success. In M. Khosrow-Pour (Ed.), *Encyclopedia of Information Science and Technology*, Third Edition (pp. 1406-1414). IGI Global.
- Giannakis-Bompolis, C., & Boutsouki, C. (2014). Customer Relationship Management in the Era of Social Web and Social Customer: An Investigation of Customer Engagement in the Greek Retail Banking Sector. *Procedia: Social and Behavioral Sciences*, 148, 67–78. doi:10.1016/j.sbspro.2014.07.018
- Goh, K., Heng, C., & Lin, Z. (2013). Social Media Brand Community and Consumer Behavior: Quantifying the Relative Impact of User- and Marketer-Generated Content. *Information Systems Research*, 24(1), 88–107. doi:10.1287/isre.1120.0469
- Google. (2014). *Worldwide data. Google Analytics (Q1, 2014)*. Retrieved September 25, 2014, from: <http://www.google.com>
- Hansen, D. L., Shneiderman, B., & Smith, M. A. (2011). *Analysing Social Media Networks with NodeXL: Insights from a Connected World*. Burlington: Morgan Kaufmann.
- Haro de Rosario, A., Gálvez-Rodríguez, M. M., & Caba-Pérez, M. C. (2013). Development of social media and web 2.0 in the top hotel chains. *Tourism & Management Studies*, 9(1), 13–19.
- Harrigan, P., Soutar, G., Choudhury, M. M., & Lowe, M. (2015). Modelling CRM in a social media age. *Australasian Marketing Journal*, 23(1), 27–37. doi:10.1016/j.ausmj.2014.11.001
- Hennig-Thurau, T., Malhotra, E. C., Frieger, C., Gensler, S., Lobschat, L., Rangaswamy, A., & Skiera, B. (2010). The impact of new media on customer relationships. *Journal of Service Research*, 13(3), 311–330. doi:10.1177/1094670510375460
- Jung, T. H., Ineson, E. M., & Green, E. (2013). Online social networking: Relationship marketing in UK hotels. *Journal of Marketing Management*, 29(3-4), 3–4. doi:10.1080/0267257X.2012.732597
- Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of social media. *Business Horizons*, 53(1), 59–68. doi:10.1016/j.bushor.2009.09.003
- Kumar, V., Aksoy, L., Donders, B., Venkatesh, R., Wiesel, T., & Tillmanns, S. (2010). Undervalued or overvalued Customers: Capturing Total Customer Engagement Value. *Journal of Service Research*, 13(3), 297–310. doi:10.1177/1094670510375602
- Luo, X., Zhang, J., & Duan, W. (2013). Social Media and Firm Equity Value. *Information Systems Research*, 24(1), 146–163. doi:10.1287/isre.1120.0462
- Mansson, M. (2011). Mediatized Tourism. *Annals of Tourism Research*, 38(4), 1634–1652. doi:10.1016/j.annals.2011.02.008
- Orfila-Sintes, F., & Mattsson, J. (2009). Innovation behavior in the hotel industry. *Omega*, 37(2), 380–394. doi:10.1016/j.omega.2007.04.002
- Porter, C. E., Donthu, N., MacElroy, W. H., & Wydra, D. (2011). How to Foster and Sustain Engagement in Virtual Communities. *California Management Review*, 53(4), 80–110. doi:10.1525/cmr.2011.53.4.80
- Rapp, A., Beitelspacher, L. S., Grewal, D., & Hughes, D. E. (2013). Understanding social media effects across seller, retailer, and consumer interactions. *Journal of the Academy of Marketing Science*, 41(5), 547–566. doi:10.1007/s11747-013-0326-9
- Rishika, R., Kumar, A., Janakiraman, R., & Beza-wada, R. (2013). The Effect of Customers Social Media Participation on Customer Visit Frequency and Profitability: An Empirical Investigation. *Information Systems Research*, 24(1), 108–127. doi:10.1287/isre.1120.0460

Sashi, C. M. (2012). Customer engagement, buyer-seller relationships, and social media. *Management Decision*, 50(2), 253–272. doi:10.1108/00251741211203551

Sigala, M. (2011). eCRM 2.0 applications and trends: The use and perceptions of Greek tourism firms of social networks and intelligence. *Computers in Human Behavior*, 27(2), 655–661. doi:10.1016/j.chb.2010.03.007

Smith, P. R., & Zook, Z. (2011). *Marketing Communications*. London: Kogan Page Publishers.

Van Doorn, J., Lemon, K. N., Mittal, V., Naß, S., Pick, D., Pirner, P., & Verhoef, P. C. (2010). Customer Engagement Behavior: Theoretical Foundations and Research Directions. *Journal of Service Research*, 13(3), 253–266. doi:10.1177/1094670510375599

Verleye, K., Gemmel, P., & Rangarajan, D. (2014). Managing Engagement Behaviors in a Network of Customers and Stakeholders: Evidence From the Nursing Home Sector. *Journal of Service Research*, 17(1), 68–84. doi:10.1177/1094670513494015

World Tourism Organization. (2015). *UNWTO Tourism Highlights. 2015 Edition*. Retrieved October 29, 2015 from <http://www.e-unwto.org/doi/pdf/10.18111/9789284416899>

ADDITIONAL READING

Andzulis, J. M., Panagopoulos, N. G., & Rapp, A. (2012). A review of Social Media and Implications for the sales Process. *Journal of Personal Selling & Sales Management*, 32(3), 305–316. doi:10.2753/PSS0885-3134320302

Kietzmann, J. H., Silvestre, B. S., McCarthy, I. P., & Pitt, L. F. (2012). Unpacking the social media phenomenon: Towards a research agenda. *Journal of Public Affairs*, 12(2), 109–119. doi:10.1002/pa.1412

Leung, D., Law, R., Van Hoof, H., & Buhalis, D. (2013). Social Media in Tourism and Hospitality: A literature review. *Journal of Travel & Tourism Management*, 30(1-2), 3–22. doi:10.1080/10548408.2013.750919

KEY TERMS AND DEFINITIONS

Blogs: Websites owned and written by individuals who maintain regular commentaries, consisting of discrete entries (“posts”) displayed in reverse chronological order.

Customer Engagement: behavioral manifestation toward the brand or firm that goes beyond transactions. It involves the creation of a deeper, more meaningful connection between the company and the customer, and one that endures over time.

Microblogging: Social networking services that enable users to send and read very short messages, usually restricted by the number of characters (e.g., Twitter allows 140 characters per message).

Picture-Sharing Websites (e.g. Flickr) and Video-Sharing Websites (e.g. YouTube): Online platforms that allow users to store and share images and videos.

Review Sites: Websites on which reviews can be posted about people, businesses, products, or services (e.g., TripAdvisor).

Social Networking Sites: Online platforms on which users can find and add friends and contacts, send messages to friends, and update personal profiles (e.g. Facebook).

Category M

**Medical Education, Ethics, and
Law**

Comprehensive E–Learning Appraisal System

M**Jose Luis Monroy Anton***La Ribera University Hospital, Spain***Juan Vicente Izquierdo Soriano***La Ribera University Hospital, Spain***Maria Isabel Asensio Martinez***La Ribera University Hospital, Spain***Felix Buendia Garcia***Polytechnic University of Valencia, Spain*

INTRODUCTION

The healthcare sector in the XXI century presents a big technological development that covers a broad knowledge. All fields of medicine are deepening their knowledge, which increases the volume of material that must be handled by professionals in each specialty. This large volume of material should be taken into account by health professionals, because it contributes to a better quality of care.

The mode of transmission of this knowledge to professionals is usually through training courses in their own work areas (internal instruction) or outside the centers (external instruction). New technologies have entered the world of education, leading to changes in the way people teach and learn. The traditional way of teaching has been face-to-face classes; however, with rising technologies, virtual training via computers and virtual teachers are being implemented in some institutions. This change in the way of teaching also leads to changes in how to assess the knowledge gained through this method of learning.

The aim of this paper is to provide a small analysis of online training courses for health professionals, and deepen into an appraisal system developed to integrate different complementary

variables, and how they can be implemented as a method addressed to assess online courses in a more comprehensive way.

BACKGROUND

Health Training

Health training is a complex field; several areas of knowledge and practice make necessary an update in instructional and teaching methodologies (Schoonheim, 2014; Frenk, 2010; Horton, 2010).

There are many examples of instructional needs in the field of health. Many medical specialties (e.g. primary care, surgery, or oncology) require continuous training and up-to-date of professionals to provide the best patient care. Doctors, nurses, and technicians, all of them attend courses or educational programs in order to improve their knowledge or skills in each specialty.

One example is Evidence Based Medicine (EBM). It is based on the principle that medical decision making is achieved by integrating the best available evidence with clinical expertise and patient values. When teaching EBM it should integrate core knowledge with clinical practical

activities. This will lead to taught EBM with a variety of modes: lectures, tutorials, online, problem based or self-directed learning (Straus 2011; Del Mar, 2004).

Not only the health areas but also the professionals in medical specialties have troubles with their instruction/training. General practitioners (GPs) struggle with several barriers to the use of EBM related to insufficient knowledge and skills. They also find barriers related to their practice and patient population such as lack of time, patient-related factors or a lack of available evidence (Mayer 1999; Zwolsman 2013; Te Pas 2013).

Standalone teaching improves student knowledge, but not skills, attitudes or behavior in EBM. Non randomized clinical trials indicated that integrating teaching of EBM with clinical activities (blended learning) was associated with improvements across all four domains: knowledge, skills, attitudes and behavior (Coomarasamy 2004). In the results of randomized controlled trials, the authors concluded that any form of teaching, including lecture, tutorial, self-directed, online, problem-based, multidisciplinary, was associated with an increase in EBM competency (Ilic 2014).

A comprehensive meta-analysis recently conducted by Teachers College (Columbia University) (Means 2013), indicated that students in online classrooms had moderately better performances than those receiving instruction in traditional classrooms (Milic 2016).

In the Ilic's study students's perceptions were positive about using the BL approach in teaching and learning about EBM (Ilic 2015). They developed a three-step approach (i) self-directed learning through online multimedia presentations, (ii) discussion and activities in class, and (iii) application in practice. This method was positively received by students. Students involved in the BL approach found the content was useful, engaging and well-targeted to their level of competency and suggested that blended learning could be strengthened by introducing a journal club to small group activities early in the curriculum, providing an opportunity to learn in a group environment.

This could be similar to the “forums” created on e-learning platforms to debate issues or questions about different fields of knowledge.

Some other studies appreciate that e-learning alone will increase EBM knowledge but it is not effective in increasing EBM skills and changing EBM behavior. They recommended a blended learning course, because BL includes different learning or instructional methods, different delivery methods, different scheduling and different levels of guidance (Pankin 2012). An intensive blended learning course on EBM for GP trainers enables an increase in knowledge and skills that remains after four months. However, attitude and behavior towards EBM show no differences before and after the intervention (Te Pas 2015).

Training Methods

The classical training methods consist in classroom lectures with students and teachers together. Lecture is a simple and fast method to present the vast issues addressed to a high number of learners. Some disadvantages of this method are long periods of time, inactiveness or absence of the students, or fast forgetting of the subject. When students have control over the learning methodology, it is called a “learner-centered” model of teaching. The role of the teachers in this context is to facilitate knowledge acquisition (Bahner 2012; Ruiz 2006; Boulos 2006; Koops 2011; Cook 2010).

Traditional learning must be stepped at a specific time and place and is considered essential in building a sense of community. New technologies have grown in educational world, driving to changes in teaching methods (Kemp 2014; Conole 2008).

E-Learning refers to the educational system in which teachers and students are separated by physical distance, but the technology allows them to learn together: it is the online way of knowledge. This system has its own limitations: there is no human interaction and communication regarding to a face to face classroom (Thiele 2003; Twomey, 2004). Online education is growing rapidly, pro-

viding new and useful instructional elements in different professional fields (Slotnick, 2000; Cook 2008; Hugenholtz 2008; Zolfaghari 2007). It can not only improve suitability and effectiveness for individualized and collaborative learning, but also it can provide up-to-date information through the use of interactive multimedia (Choules 2007; Cook 2007; Yu 2006; Peng 2014; Wu 2010; Moreira 2015).

E-courses are becoming more popular because of the didactical advantages they offer, such as the possibility to adapt them to individual learning styles (Nilsson 2012) as well as the possibility to repeat courses (Sun 2008; Ellaway 2008). The logistic benefits, such as learning at any convenient time or anywhere in whatever place and using different devices, are also important advantages of e-learning (Wutoh 2004; Curran 2006; Choules 2007).

Blended education or blended learning (BL) combines two or more modalities that use other teaching methods than online courses (e-Learning), in addition to the traditional classroom lessons (Anderson 2005). Teaching styles are based on the concept that individuals differ with respect to what mode of study is most effective for them. Results of some trials show that students prefer an integrated learning methodology with different learning styles, and they are supporting the use of the BL approach (Pashler 2008; Grasl 2012; Lehmann 2013; Ilic 2015).

The *hybrid* instructional model is a blend of conventional face-to-face instruction and Web-based distance learning: a mix of both classroom face-to-face meetings and distance learning. Blended learning utilizes the best online tools to support a teacher led classroom. In blended learning scenarios, “face time” between students and teachers is not replaced by online course delivery. Hybrid learning focuses less on the technology and more on the most effective way to deliver a course to learners. In hybrid learning, a significant portion of the course takes place online. Blended and hybrid learning both will be very effective

because of the synergy that is created between different ways of studying.

Online and face-to-face traditional formats have their own respective strengths and weaknesses. Neither is better: they are complementary. This confirms the benefit of BL, which includes a combination of face-to-face and online learning methodologies integrated into a single course (Osguthorpe 2003; Garrison 2008). Research has suggested that well organized online classrooms are as effective and of comparable quality as traditional classrooms (Zebrack 2005; Cook 2008; Aggarwal 2011).

The main advantage of web-based education is its flexibility, allowing students to access content from several locations, at any time. It encourages self-management learning, with the ability to exchange links to related information (Baldwin 2001; Fieschi 2002; Harden 2002; Lewin 2009).

Appraisal Systems

The usual appraisal systems for official teaching and training are often based on objective tests checking knowledge acquisition, attendance control, delivery of practical works to improve clinical skills, and participation in classes, tutorials or seminars. Tests are mainly used as a way to assess students because there is evidence that they stimulate learning by increasing retention of the information (Roediger 2006; McDaniel 2007).

Evaluation methods of online courses should collect the objectives sought in each course or training. These objectives for online distance learning are represented by five areas:

1. Incorporation of learning principles and strategies.
2. Inclusion of real problems.
3. Inclusion of decision-making, or problem solving.
4. Inclusion of practice in evaluating viewpoints.
5. Development of interactive activities which drive to discussion and critical thinking.

To develop the methodology of on-line instruction, traditional appraisal systems may not be suitable, and some of them lose their objectivity. New assessment methods should be implemented, showing in a comprehensive manner and as objectively as possible, the development of students in the subject taught and their work progress (Casebeer 2004; Weston 2008; Freire 2012).

The Comprehensive Appraisal System of La Ribera University Hospital

La Ribera University Hospital (Alzira, Spain) periodically implements training courses for health-care professionals assigned to the corresponding Department, and addressed to all employees that integrate this Department.

It has been delivered online training courses for several years in the health field, based on the Campus Ribera, a corporate virtual platform implemented in our Department of Health, which has allowed the organization, to evaluate the methodology and performance of these courses (Monroy Anton 2011, 2013). In this field we have developed what we call the “*Comprehensive e-Learning Appraisal System*” (CeLAS), which aims to categorize and score each of the elements of learning/instruction, resulting in a final assessment for students very close to objectivity and fairness.

These training courses are developed electronically through the corporate platform *Campus Ribera* using a Moodle platform. Campus Ribera started in 2009 as a learning platform that allows optimizing the traditional training process by using online learning methods (Izquierdo et al, 2014). This virtual platform is managed from the institution, and teachers (tutors) are qualified professionals of the Department of Health.

After several years of developing different types of courses, the final assessment to obtain credits or certificates was very heterogeneous because the test systems were different: only the

connection to the course, meet a minimum number of input records online, make some assessment test, some drafting work, etc. None of them were perfect and contrast the existence of groups of students more interested and participatory compared to other more passive groups. Thus, the ultimate goal of the courses is distorted because the training of some professionals would not be that expected and motivation to achieve these courses would only be addressed to obtain certificates for merely passing the course.

To standardize somehow this deficiency and improve the quality of the course and the training of students it is designed a system of evaluation to attack different parts and elements, weighting each one according to the relative importance that teachers would give them, and resulting in a final note or score. We have called this system: “*Comprehensive e-Learning Appraisal System*” (CeLAS).

CeLAS was initially implemented in a blended course entitled “*Approach to the toxicity of treatments with external radiotherapy*”. It lasted 6 weeks, with a workload of 25 hours (mixed classroom and online) and it was accredited by the National Education Commission of Spain with 3.6 credits. They were accepted for the course seven students, based on the best ratio student/teacher parameters previously analyzed for our blended courses (Monroy Antón 2013).

Assessment Method

Three areas of training (blocks or sections) were proposed, with a system of weighted score for each one of a total of 100 points as it is displayed in Figure 1):

Area/Block I: Assessment of knowledge, representing 70% of the total score.

Area/Block II: Assessment of participation, representing 15% of the total score.

Area/Block III: Assessment of practices, representing 15% of the total score.

Figure 1. Areas of training



Final Course Certificate

For the final course certificate, students must achieve the following objectives:

1. Exceed 50% of the total points of the final evaluation.
2. Not obtain a score of zero in any of the three aforementioned evaluation blocks.

Methodology of Evaluation of Each Block

Evaluation of Block I

Assuming a total of 70 points which were obtained by performing knowledge test corresponding to each of the different thematic units (T.U.) of the course, and a final overall evaluation test (as it is shown in Figure 2).

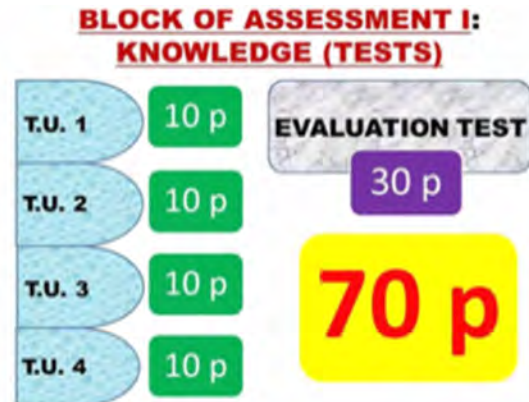
Evaluation of Block II

Weighted with a total of 15 points. They were obtained by the participations of students in specific forums for each thematic unit. These entries were evaluated according to the type of response issued by the students (Figure 3):

- 0 points = not answered;
 1 = unsatisfactory;
 2 = satisfactory;
 3 = above expectations.

The final grade of the evaluation block result of the sum of average scores in all forums, with a maximum possible score of 15 points.

Figure 2. Evaluation of knowledge



For the evaluation of forums, students must meet minimum requirements:

- It was mandatory to answer the specific forum “*Case Report*”. In the case that there was no participation in this forum, the rating would be zero in the entire block.
- To obtain the score greater than zero in at least three of the four forums on thematic units.

Evaluation of Block III

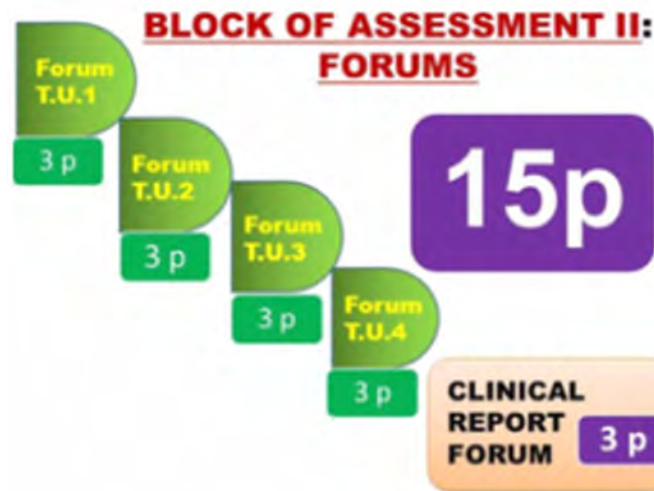
Weighted with a total of 15 points. Related to works proposed by teachers and delivered by students by electronic means, and a test with questions about the practical issues discussed during the course. The grading of the work was as follows (Figure 3):

- 0 points = not answered;
 1 = unsatisfactory;
 2 = satisfactory;
 3 = above expectations.

The final score of the block assessment result of the sum of average scores for all jobs, the maximum possible score was 15 points.

For these scores, minimum requirements had to be met:

Figure 3. Evaluation of participation (forums)



- The fulfillment of the test of practice was mandatory.
- Achieving greater than zero score in at least three of the four papers offered in the thematic units.

Students' Work Monitoring

To guide students on their progress throughout the course and the scores on each of the blocks and tasks, Moodle offers the possibility of gathering information through the *Gradebook* which monitors student's activities. The *Gradebook* in Campus Ribera allows tutors to organize and manage information about the students in online courses. For example, by displaying the scores obtained through objective tests or grading the delivery of assignments or forum participation. It also enables the aggregation of the different assessment sources in order to get a final grade.

FUTURE RESEARCH DIRECTIONS

The Global World leads to the use of new technologies for most of our daily activities. One of them is education. Training systems via e-Learning are gradually imposing traditional education methods,

allowing long distances formation of large population. However, this ability to communicate and learn knowledge should be assessed as objectively as possible, combining performance parameters and elements, so we must investigate in the development of new evaluation systems.

Recent studies suggest that a combination of face to face training with online instruction based on e-learning methods is more flexible than other methods (Garrison 2004).

The online learning paradigm offers valuable tools that could supplement or replace aspects of the traditional lecture-and-textbook-based approach to teaching and learning (Glazer, 2012).

Blended learning attempts to create an optimal learning environment by combining a variety of learning approaches (lecture, tutorial, online, problem-based and clinical) to deal with different learning styles and requirements. Some trials have evaluated the effectiveness of the BL model in medicine (Grasl 2012; Lehmann 2013; Masie 2002).

The asynchronous ability to exchange information, cost-saving, personalized learning, increased accessibility, ease of distribution and updating content are some examples of the advantages offered by the online classrooms. Studies and literature reviews show that technology based

learning improves student knowledge acquisition which requires abstract conceptualization and reflective observation, but adversely affects student ability to obtain knowledge requiring concrete experience (Hui 2007).

The meta-analysis performed by Liu (2016) concludes that blended learning has a large consistent positive effect on knowledge acquisition compared with other modalities (pure e-learning, face-to-face, or no intervention) and may be more effective than other ways of learning. They concluded that blended learning was very effective and educationally useful in health professions. Possible explanations could be:

1. Compared with traditional learning, blended learning allows students to review electronic materials as often as necessary. It likely enhances learning performance (Wu 2010).
2. Compared with e-learning, blended learning learners are less likely to experience feelings of isolation or reduced interest in the subject matter (Hara 2000; Maki 2000).

Analyzing each unit of the toxicity course, acquired knowledge of study matter improved the previous level that brought the students before the end of the teaching of these subjects. This was observed conducting tests before the study of the issues of each thematic unit and after teaching the same knowledge. The difference between the scores before and after training in other studies, led to increased knowledge scores in participants (Unal 2005; Koch 2005; Woo 2000).

Tests are considered the standard and principal ways of assessment for several courses. The implementation of a medical program with tests is developed in order to stimulate learning: after an initial contact with the learning material, being tested increases information retention more than reviewing that material again. Also the number of tests increases the effect of this system (Larsen 2008, 2009; Kerdijk 2013; Wood 2009; Karpicke 2008). Some institutions have created platforms

that employ test-enhanced methodologies to teach learning materials (Taveira 2014).

The activities that can help learners to apply knowledge from the courses are exercises, quizzes, self-assessment test; but also discussion cases, forums interactivity or many others. These are the items that the faculty of courses should evaluate.

Prospective trial conducted with third year medical undergraduate students attending the Faculty of Medicine of the University of Belgrade (Milic 2016) compared two methods of education: blended learning and traditional on-site. Students were asked to choose their learning method at the beginning of their course. The evaluation system was identical for both learning modalities:

1. Course activities throughout the year,
2. Written knowledge test,
3. Final exam.

On a similar way, Larson reported no significant differences among three education delivery methods (on-site, online, and blended learning) as measured by exam scores and final grades. The study concluded that online and blended methods are at least equivalent to the face-to-face one (Larson 2009).

As we can see, most work on evaluation of online courses use multichoice test, as used by us. A simple way is to establish a classification or numerical evaluation of knowledge. However, increasingly we turn to complete this evaluation with other activities, which should complement the results of the test.

Systematic or comprehensive appraisal is not common practice. Other ways for monitoring courses are observing participation and satisfaction. Appraisal should be driven not only by the test results but also other complementary issues such as skills or tasks proposals through problem-based methods (Kreber 2001; Wilkes 1999).

The future will probably be the implementation of courses tailored to the needs of students with online high probability. Evaluation systems

of these courses must be adapted to each of them, but mainly to the following facts: the group of students to whom it is addressed, the objectives pursued in the course, course duration, the scope for action of the students. Evaluation systems as developed in our hospital can serve as a model or basis for developing other more complete, complex, or just different, but adapted to the group of students that the course addresses.

CONCLUSION

In our experience of an online radiotherapy course, all students (100%) reached the minimum final score required in the course, fulfilling the objectives of each subject and evaluation block and obtaining the corresponding certificate. The results showed that students were getting better as performing the various blocks of the course which were arranged in a table of grades, so that students could see at any time their partial and overall marks knowing their score to overcome the required minimum for the final course certificate.

This “Comprehensive e-Learning Appraisal System” provides a fairly objective assessment method of learning, and can avoid or greatly reduce the subjectivity of the teacher assessment, as grades are displayed automatically, once students do course activities. It can be applied as a tool in the training courses in all fields of health care, and enables holistic assessment jointly with different variables of knowledge, weighting those considered most important in each specific course.

The grades or scores of the course can be accessed whenever by students and teachers by means of items designed specifically for each training activity (e.g. forums or wikis), which are integrated into the module of the course and have each of the graded sections and the corresponding score. It is a demanding system for the student, since it forces pursuing and constant work, reflecting on the final grade this degree of effort. These final

grades also serve to differentiate the use of the training received, positively influencing the professional careers of workers. It is an easy system to implement in courses with an online methodology; it does not require sophisticated virtual tools or complex calculations, since practically all scores of evaluable elements are provided directly from the various items of the learning platform.

REFERENCES

- Aggarwal, R., Gupte, N., Kass, N., Taylor, H., Ali, J., Bhan, A., & Bollinger, R. C. et al. (2011). A comparison of online versus on-site training in health research methodology: A randomized study. *BMC Medical Education*, 11(1), 37. doi:10.1186/1472-6920-11-37 PMID:21682858
- Anderson, W. L., Mitchell, S. M., & Osgood, M. P. (2005). Comparison of student performance in cooperative learning and traditional lecture-based biochemistry classes. *Biochemistry and Molecular Biology Education*, 33(6), 387–393. doi:10.1002/bmb.2005.49403306387 PMID:21638607
- Bahner, D. P., Adkins, E., Patel, N., Donley, C., Nagel, R., & Kman, N. E. (2012). How we use social media to supplement a novel curriculum in medical education. *Medical Teacher*, 34(6), 439–444. doi:10.3109/0142159X.2012.668245 PMID:22449268
- Baldwin, C. D., Gephart, D., & Maulitz, R. (2001). Collaborative planning of a Web-based learning resource for primary care education. *Academic Medicine*, 76(5), 549–550. doi:10.1097/00001888-200105000-00093 PMID:11346599
- Boulos, M. N. K., Maramba, I., & Wheeler, S. (2006). Wikis, blogs and podcasts: A new generation of Web-based tools for virtual collaborative clinical practice and education. *BMC Medical Education*, 6(1), 41. doi:10.1186/1472-6920-6-41 PMID:16911779

- Casebeer, L., Kristofco, R. E., Strasser, S., Reilly, M., Krishnamoorthy, P., Rabin, A., & Myers, L. et al. (2004). Standardizing evaluation of on-line continuing medical education: Physician knowledge, attitudes, and reflection on practice. *The Journal of Continuing Education in the Health Professions*, 24(2), 68–75. doi:10.1002/chp.1340240203 PMID:15279131
- Choules, A. P. (2007). The use of elearning in medical education: A review of the current situation. *Postgraduate Medical Journal*, 83(978), 212–216. doi:10.1136/pgmj.2006.054189 PMID:17403945
- Conole, G., de Laat, M., Dillon, T., & Darby, J. (2008). Disruptive technologies, pedagogical innovation: Whats new? Findings from an in-depth study of students use and perception of technology. *Computers & Education*, 50(2), 511–524. doi:10.1016/j.compedu.2007.09.009
- Cook, D. A. (2007). Web-based learning: Pros, cons and controversies. *Clinical Medicine*, 7(1), 37–42. doi:10.7861/clinmedicine.7-1-37 PMID:17348573
- Cook, D. A., Levinson, A. J., Garside, S., Dupras, D. M., Erwin, P. J., & Montori, V. M. (2008). Internet-based learning in the health professions: A meta-analysis. *Journal of the American Medical Association*, 300(10), 1181–1196. doi:10.1001/jama.300.10.1181 PMID:18780847
- Coomarasamy, A., & Khan, K. (2004). What is the evidence that postgraduate teaching in evidence based medicine changes anything? A systematic review. *BMJ (Clinical Research Ed.)*, 329(7473), 1017. doi:10.1136/bmj.329.7473.1017 PMID:15514348
- Curran, V., Lockyer, J., Sargeant, J., & Fleet, L. (2006). Evaluation of learning outcomes in Web-based continuing medical education. *Academic Medicine*, 81(Suppl), S30–S34. doi:10.1097/01.ACM.0000236509.32699.f5 PMID:17001130
- Da, C. (2010). Instructional design variations in internet-based learning for health professions education: A systematic review and meta-analysis. *Academic Medicine*, 85(5), 909–922. doi:10.1097/ACM.0b013e3181d6c319 PMID:20520049
- Del Mar, C., Glasziou, P., & Mayer, D. (2004). Teaching evidence based medicine. *BMJ (Clinical Research Ed.)*, 329(7473), 989–990. doi:10.1136/bmj.329.7473.989 PMID:15514319
- Ellaway, R., & Masters, K. (2008). AMEE Guide 32: e-Learning in medical education - Part 1: Learning, teaching and assessment. *Medical Teacher*, 30(5), 455–473. doi:10.1080/01421590802108331 PMID:18576185
- Fieschi, M., Soula, G., Giorgi, R., Gouvernet, J., Fieschi, D., Botti, G., & Berland, Y. et al. (2002). Experimenting with new paradigms for medical education and the emergence of a distance learning degree using the internet: Teaching evidence- based medicine. *Medical Informatics and the Internet in Medicine*, 27(1), 1–11. doi:10.1080/14639230110105301 PMID:12509118
- Freire, L. L., Arezes, P. M., & Campos, J. A. (2012). literature review about usability evaluation methods for e-learning platforms. *Work (Reading, Mass.)*, 41, 1038–1044. PMID:22316857
- Frenk, J., Chen, L., Bhutta, Z. A., Cohen, J., Crisp, N., Evans, T., & Zurayk, H. et al. (2010). Health professionals for a new century: Transforming education to strengthen health systems in an interdependent world. *Lancet*, 376(9756), 1923–1958. doi:10.1016/S0140-6736(10)61854-5 PMID:21112623
- Garrison, D., & Vaughan, N. (2008). Blended Learning. In *Higher Education: Framework, Principles, and Guidelines*. San Francisco, CA: Jossey-Bass.

- Garrison, D. R., & Kanuka, H. (2004). Blended learning: Uncovering its transformative potential in higher education. *The Internet and Higher Education*, 7(2), 95–105. doi:10.1016/j.ihe-duc.2004.02.001
- Glazer, F. S. (2012). *Blended Learning: Across the Disciplines, across the Academy*. Sterling, VA: Stylus.
- Grasl, M., Pokieser, P., Gleiss, A., Brandstaetter, J., Sigmund, T., & Erovic, B. et al.. (2012). A new blended learning concept for medical students in otolaryngology. *Archives of Otolaryngology Head & Neck Surgery*, 138(4), 358–366. doi:10.1001/archoto.2012.145 PMID:22508619
- Hara, N. (2000). Student distress in a Web-based distance education course. *Information Communication and Society*, 3(4), 557–579. doi:10.1080/13691180010002297
- Harden, R. M., & Hart, I. R. (2002). An international virtual medical school (IVIMEDS): The future for medical education? *Medical Teacher*, 24(3), 261–267. doi:10.1080/01421590220141008 PMID:12098412
- Horton, R. (2010). A new epoch for health professionals education. *Lancet*, 376(9756), 1875–1877. doi:10.1016/S0140-6736(10)62008-9 PMID:21112621
- Hugenholtz, N. I. R., De Croon, E. M., Smits, P. B., Van Dijk, F. J. H., & Nieuwenhuijsen, K. (2008). Effectiveness of e-learning in continuing medical education for occupational physicians. *Occupational Medicine*, 58(5), 370–372. doi:10.1093/occmed/kqn053 PMID:18495676
- Hui, W., Hu, P. J.-H., Clark, T. H. K., Tam, K. Y., & Milton, J. (2007). Technology-assisted learning: A longitudinal field study of knowledge category, learning effectiveness and satisfaction in language learning. *Journal of Computer Assisted Learning*, 24(3), 245–259. doi:10.1111/j.1365-2729.2007.00257.x
- Ilic, D., Bin Nordin, R., Glasziou, P. K., Tilson, J. K., & Villanueva, E. (2015). A randomised controlled trial of a blended learning education intervention for teaching evidence-based medicine. *BMC Medical Education*, 10(15), 39.
- Ilic, D., & Maloney, S. (2014). Methods of teaching medical trainees Evidence Based Medicine: A systematic review. *Medical Education*, 48(2), 124–135. doi:10.1111/medu.12288 PMID:24528395
- Izquierdo, J. V., Buendía, F., Ortega, J. L., & Taberner, E. (2014). E-Learning Experiences in La Ribera Health Department. *Journal of Information Technology Research*, 7(2), 7–23. doi:10.4018/jitr.2014040102
- Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science*, 319(5865), 966–968. doi:10.1126/science.1152408 PMID:18276894
- Kemp, N., & Grieve, R. (2014). Face-to-face or face-to-screen? Undergraduates opinions and test performance in classroom vs. online learning. *Frontiers in Psychology*, 5, 1278. doi:10.3389/fpsyg.2014.01278 PMID:25429276
- Kerdijk, W., Tio, R. A., Mulder, B. F., & Cohen-Schotanus, J. (2013, December). Cumulative assessment strategic choices to influence students study effort. *BMC Medical Education*, 13(1), 172. doi:10.1186/1472-6920-13-172 PMID:24370117
- Koch S, Townsend CD, Dooley KE. (2013). A case study comparison between web-based and traditional graduate level academic leadership instruction. *JAE 2005*, 46(4), 72-82.
- Koops, W., Van der Vleuten, C., De Leng, B., Oei, S. G., & Snoeckx, L. (2011). Computer-supported collaborative learning in the medical workplace: Students experiences on formative peer feedback of a critical appraisal of a topic paper. *Medical Teacher*, 33(6), e318–e323. doi:10.3109/0142159X.2011.575901 PMID:21609168

- Kreber, C., Brook, P., & Policy, E. (2001). Impact evaluation of educational development programmes. *The International Journal for Academic Development*, 6(2), 96–108. doi:10.1080/13601440110090749
- Larsen, D. P., Butler, A. C., & Roediger, H. L. III. (2008). Test-enhanced learning in medical education. *Medical Education*, 42(10), 959–966. doi:10.1111/j.1365-2923.2008.03124.x PMID:18823514
- Larsen, D. P., Butler, A. C., & Roediger, H. L. III. (2009). Repeated testing improves long-term retention relative to repeated study: A randomised controlled trial. *Medical Education*, 43(12), 1174–1181. doi:10.1111/j.1365-2923.2009.03518.x PMID:19930508
- Larson, D. K., & Sung, C. H. (2009). Comparing student performance: Online versus blended versus face-to-face. *Journal of Asynchronous Learning Networks*, 13, 31–42.
- Lehmann, R., Bosse, H., Simon, A., Nikendei, C., & Huwendiek, S. (2013). An innovative blended learning approach using virtual patients as preparation for skills laboratory training: Perceptions of students and tutors. *BMC Medical Education*, 13(1), 23. doi:10.1186/1472-6920-13-23 PMID:23402663
- Lewin, L. O., Singh, M., Bateman, B. L., & Glover, P. B. (2009). Improving education in primary care: Development of an online curriculum using the blended learning model. *BMC Medical Education*, 9(1), 33. doi:10.1186/1472-6920-9-33 PMID:19515243
- Liu Q, Peng W, Zhang F, Hu R, Li Y, Yan W. (2016). The Effectiveness of Blended Learning in Health Professions: Systematic Review and Meta-Analysis. *Journal of Medical Internet Research*, 18(1).
- Maki, R. H., Maki, W. S., Patterson, M., & Whitaker, P. D. (2000). Evaluation of a Web-based introductory psychology course: I. Learning and satisfaction in on-line versus lecture courses. *Behavior Research Methods, Instruments, & Computers*, 32(2), 230–239. doi:10.3758/BF03207788 PMID:10875167
- Masie, E. (2002). Blended learning: the magic is in the mix. In A. Rossett (Ed.), *The ASTD e-learning handbook*. New York: McGraw-Hill.
- Mayer, J., & Piterman, L. (1999). The attitudes of Australian GPs to evidence-based medicine: A focus group study. *Family Practice*, 16(6), 627–632. doi:10.1093/fampra/16.6.627 PMID:10625142
- McDaniel, M. A., Roediger, H. L., & McDermott, K. B. (2007). Generalizing test-enhanced learning from the laboratory to the classroom. *Psychonomic Bulletin & Review*, 14(2), 200–206. doi:10.3758/BF03194052 PMID:17694901
- Means, B., Toyama, Y., Murphy, R., & Baki, M. (2013). The Effectiveness of Online and Blended Learning: A Meta- Analysis of the Empirical Literature. *Teachers College Record*, 115, 1–47.
- Milic, N. M., Trajkovic, G. Z., Bukumiric, Z. M., Cirkovic, A., Nikolic, I. M., Milin, J. S., ... Stanisavljevic, D. M. (2016). Improving Education in Medical Statistics: Implementing a Blended Learning Model in the Existing Curriculum. *PLoS One*, 11(2).
- Monroy Anton, J. L., Izquierdo Soriano, J. V., & Buendía García, F. (2011). E-Learning for hospital personnel. A method for accessing scientific training. *EDULEARN11 Proceedings*, 4003–4006.
- Monroy Antón, JL; Izquierdo Soriano JV.; Ribes Llopis, L; Buendía García, F. (2013). *Análisis de parámetros cuantificables para el diseño de un perfil de cursos de formación online*. SEFM (19)/SEPR(14).

- Moreira, I. C., Ventura, S. R., Ramos, I., & Rodrigues, P. P. (2015). Development and assessment of an e-learning course on breast imaging for radiographers: A stratified randomized controlled trial. *Journal of Medical Internet Research*, 17(1), e3. doi:10.2196/jmir.3344 PMID:25560547
- Nilsson, M., Ostergren, J., Fors, U., Rickenlund, A., Jorfeldt, L., Caidahl, K., & Bolinder, G. (2012). Does individual learning styles influence the choice to use a web-based ECG learning programme in a blended learning setting? *BMC Medical Education*, 12(1), 5. doi:10.1186/1472-6920-12-5 PMID:22248183
- Osguthorpe, R., & Graham, C. (2003). Blended learning environments: Definitions and directions. *Revista Educación a Distancia*, 4, 227–233.
- Pankin, J., Roberts, J., & Savio, M. (2012). *Blended Learning at MIT*. White Paper. MIT Training Alignment Team.
- Pashler, H., McDaniel, M., Rohrer, D., & Bjork, R. (2008). Learning styles. *Concepts Evid Psychological Science in the Public Interest*, 9, 105–119. PMID:26162104
- Peng, Y., Wu, X., Atkins, S., Zwarentein, M., Zhu, M., Zhan, X. X., & Yan, W. R. et al. (2014). Internet-based health education in China: A content analysis of websites. *BMC Medical Education*, 14(1), 16. doi:10.1186/1472-6920-14-16 PMID:24467710
- Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning taking memory tests improves long-term retention. *Psychological Science*, 17(3), 249–255. doi:10.1111/j.1467-9280.2006.01693.x PMID:16507066
- Ruiz, J. G., Mintzer, M. J., & Leipzig, R. M. (2006). The impact of E-learning in medical education. *Academic Medicine*, 81(3), 207–212. doi:10.1097/00001888-200603000-00002 PMID:16501260
- Schoonheim, M., Heyden, R., & Wiecha, J. M. (2014). Use of a virtual world computer environment for international distance education: Lessons from a pilot project using Second Life. *BMC Medical Education*, 14(1), 36. doi:10.1186/1472-6920-14-36 PMID:24555833
- Slotnick, H. B. (2000). Physicians learning strategies. *Chest*, 118(2), 18S–23S. doi:10.1378/chest.118.2_suppl.18S PMID:10939995
- Sun P, Tsai RJ, Finger G, Chen Y, Yeh D. (2008). What drives a successful e-Learning? An empirical investigation of the critical factors influencing learner satisfaction. *Computers and Education*, 50, 1183–202.
- Taveira-Gomes, T., Saffarzadeh, A., Severo, M., Guimarães, J. M., & Ferreira, M. A. (2014). A novel collaborative e-learning platform for medical students-alert student. *BMC Medical Education*, 14(1), 143. doi:10.1186/1472-6920-14-143 PMID:25017028
- Te Pas, E., van Dijk, N., Bartelink, M. E. L., & Wieringa-De Waard, M. (2013). Factors influencing the EBM behaviour of GP trainers; a mixed method study. *Medical Teacher*, 35(3), e990–e997. doi:10.3109/0142159X.2012.733044 PMID:23102157
- Te Pas E, Wieringa-de Waard M, de Ruijter W, van Dijk N. (2015). Learning results of GP trainers in a blended learning course on EBM: A cohort study. *BMC Medical Education*, 12(15), 104.
- Thiele, J. E. (2003). Learning patterns of online students. *The Journal of Nursing Education*, 42(8), 364–366. PMID:12938899
- Twomey, A. (2004). Web-based teaching in nursing: Lessons from the literature. *Nurse Education Today*, 24(6), 452–458. doi:10.1016/j.nedt.2004.04.010 PMID:15312954
- Unal, Z. (2005). A Comparative Study On Learning Outcomes Of Web Based Vs. Classroom Based Instruction. *Journal of College Teaching & Learning*, 2(3), 1–6.

Weston, C. M., Sciamanna, C., & Nash, D. (2008). Evaluating online continuing medical education seminars: Evidence for improving clinical practices. *American Journal of Medical Quality*, 23(6), 475–483. doi:10.1177/1062860608325266 PMID:19001103

Wilkes, M., & Bligh, J. (1999). Evaluating educational interventions. *British Medical Journal*, 318(7193), 1269–1272. doi:10.1136/bmj.318.7193.1269 PMID:10231263

Woo, M. A., & Kimmick, J. V. (2000). Comparison of Internet versus lecture instructional methods for teaching nursing Research. *Journal of Professional Nursing*, 16(3), 132–139. doi:10.1053/PN.2000.5919 PMID:10860311

Wood, T. (2009). Assessment not only drives learning, it may also help learning. *Medical Education*, 43(1), 5–6. doi:10.1111/j.1365-2923.2008.03237.x PMID:19140992

Wu, J., Tennyson, R. D., & Hsia, T. (2010). A study of student satisfaction in a blended e-learning system environment. *Computers & Education*, 55(1), 155–164. doi:10.1016/j.compedu.2009.12.012

Wutoh, R., Austin Boren, S., & Balas, E. A. (2004). eLearning: A review of internet-based continuing medical education. *The Journal of Continuing Education in the Health Professions*, 24(1), 20–30. doi:10.1002/chp.1340240105 PMID:15069909

Yu, S., & Yang, K. (2006). Attitudes toward Web-based distance learning among public health nurses in Taiwan: A questionnaire survey. *International Journal of Nursing Studies*, 43(6), 767–774. doi:10.1016/j.ijnurstu.2005.09.005 PMID:16253261

Zebrack, J. R., Mitchell, J. L., Davids, S. L., & Simpson, D. E. (2005). Web-based curriculum. A practical and effective strategy for teaching womens health. *Journal of General Internal Medicine*, 20(1), 68–74. doi:10.1111/j.1525-1497.2005.40062.x PMID:15693931

Zolfaghari, M., Mehrdad, N., Parsa Yekta, Z., Salmani Barugh, N., & Bahrani, N. (2007). The effect of lecture and e-learning methods on learning mother and child health course in nursing students. *Indian Journal of Medical Education*, 7(1), 31–39.

Zwolsman, S. E., van Dijk, N., te Pas, E., & Wieringa-De Waard, M. (2013). Barriers to the use of evidence-based medicine: Knowledge and skills, attitude, and external factors. *Perspectives on Medical Education*, 2(1), 4–13. doi:10.1007/s40037-013-0039-2 PMID:23670651

KEY TERMS AND DEFINITIONS

Appraisal Systems: Evaluation systems for courses, attending different approaches and methodology of assessment.

Blended Learning: Combination of traditional face-to-face learning and distance (online) methods.

E-Learning: Education based in online courses. Educational system in which teachers and students are separated by physical distance, but the technology allows them to learn together.

Hybrid Learning: Finding the right mix for the student, out of all the possibilities in learning.

Flipping the Medical School Classroom

Kristina Kaljo

Medical College of Wisconsin, USA

Laura Jacques

Medical College of Wisconsin, USA

INTRODUCTION

Teaching and learning in medical schools and institutions is a multifaceted endeavor, especially as the United States' health care system continues to become more complex. Medical students need to develop a rich foundation of not only content knowledge, but also problem solving and technical clinical skills. However, due to various barriers such as inadequate or limited clinical placements, an unpredictable learning environment, and duty hour time restrictions (Feng et al, 2013; McLaughlin et al., 2014; Tainter et al, 2016), medical students may develop gaps in knowledge that could impact their future success. To circumvent these gaps, medical education programs must be innovative in how they deliver core content knowledge. Historically, lecture-based didactics permitted learners to be passive in the acquisition of knowledge, while teaching faculty were recognized as a "sage on the stage" (King, 1993). The role of "sage on the stage" is considered to be the keeper of knowledge and who considers students as empty vessels that need to be filled. This archaic method of interaction and learning is not appropriate in the preparation of today's 21st century medical student (O'Brien & Irby, 2013).

Educational opportunities should be provided in a manner that is learner-centered and innovative to further increase student engagement and deeper comprehension of the required medical knowledge (Irby, Cooke, & O'Brien, 2010; Prober and Heath, 2012). Suggestions from Irby, Cooke, and O'Brien

(2010) assert four goals that are imperative for the future of medical education:

1. Standardization of learning outcomes and individualization of the learning process,
2. Integration of formal knowledge and clinical experience,
3. Development of habits of inquiry and innovation, and
4. Focus on professional identity formation (pg. 5-6).

With these objectives in the forefront, educational theories and practices must be revisited, tailored, in order to improve and evolve current pedagogical practice.

Rapid advancements in technology, both of software and hardware, have encouraged e-learning to emerge as a key component in curricular development, especially in the field of medical education (Dankbaar & de Jong, 2014). Students can now access and interact with required readings, cases, models, videos, simulations, and collaborate with peers across the country or the world. Remarkably, e-learning has allowed educators to "mobilize" their teaching practice and provide content at any point in time, from any location (Jashapara & Tai, 2011). However, there are always challenges when attempting to develop and implement a new e-learning strategy while simultaneously integrating face-to-face contact. Thus, how does an educator integrate both e-learning and face-to-face learning in a balanced

and engaging way? This article synthesizes the method of flipped lectures or flipped learning and presents practical strategies to implement flipped learning within the unique medical education setting.

BACKGROUND

The flipped lecture, flipped learning, or inverted classroom originates from K-12 education (Bergmann & Sams, 2012) and provides students with the opportunity to interact with content prior to entering the classroom in a structured, organized way (see figure 1) (Barett, 2012; Mazur, 2009; O'Flaherty & Phillips, 2015; Lage, Platt, & Treglia, 2000). Flipped classrooms stemmed from the Khan Academy, conceived in 2006 by Salman Khan which provided access to 3200 videos and 350 practice exercises which were used by teachers for flipped classrooms (Bishop & Vergler, 2013).

From there, flipped lectures have steadily grown in popularity following a distinct curricular design. First, students participate in the asynchronous event, commonly recognized as an online component. This typically involves a module with videos, podcasts, interactive readings, reflective questions, and/or chatrooms (Prober & Health, 2012; Sharma et al., 2015). Then, during the face-to-face class time, or synchronous event, students participate in active learning and implementation of new knowledge. Flipped learning is *not* simply assigning homework or readings and then conducting teacher-led lectures. Rather, flipped learning incorporates a diverse amalgam of technology, literature, and inquiry to establish a critical knowledge base prior to engaging with the content on a higher level of learning. In turn, the learning environment becomes personalized, engaging, and content rich, thus improving the ways in which students apply abstract knowledge to solve realistic problems (Bergmann & Sams, 2012; Mann, 2011; Sharma et al., 2015).

Flipped lectures have been found as a successful method to engage students. In a study by Lew (2016) 46 medical students enrolled in an emergency medicine clerkship much preferred the flipped lecture to a more traditional lecture-based classroom. Furthermore, the students in this study remarked on the advantage of the interactivity and medical decision-making. Interestingly, Lindeman et al (2015) discovered with the implementation of flipped lectures, surgical students did not perform any better on the National Board of Medical Examiners (NBME) examination, but that students did give a higher satisfaction rating at the conclusion of their clerkship. Similarly, Liebert et al (2016) found comparable results, when investigating student NBME results from both the flipped lecture to those in the traditional classroom, there was no statistical difference. Notably, student satisfaction of the flipped lecture was higher than the traditional lecture-based classroom with individuals reporting a greater interest in surgical specialties after the clerkship.

Those who support flipped learning have found that this instructional method has a number of benefits (O'Flaherty & Phillips, 2015; Sharma et al., 2015). First, by implementing flipped learning, students learn at their own individualized pace, which is appropriate to their learning style and need. Additionally, students can take advantage of repetition and repeated exposure to the content by pausing or revisiting the material as needed (Pierce & Fox, 2012). Upon entering the physical classroom, students have a foundational knowledge of the required content. This is essential as students and educators now have the space to engage in robust discussion regarding complex cases, problems, and issues. This interactivity also provides for immediate feedback and evaluation (Sharma et al, 2015) to both student and educator regarding what additional knowledge gaps exist. Furthermore, this flipped model encourages students to take ownership of their learning as they are required to be prepared to actively engage to solve problems and address cases.

MAIN FOCUS OF THE ARTICLE

Issues, Controversies, Problems

The traditional United States medical school follows a 2 + 2 curriculum, two years of basic science education and 2 years of clinical clerkships. The clinical clerkships are a particularly unique learning experience as medical students are required to rotate through six or more medical specialties to develop an understanding of diagnosis and health management appropriate for each discipline. Each clerkship has its own entity and student experiences vary due to different clerkship objectives, clinical content, variability in care, and teaching faculty. There is however, a common expectation for clerkships to provide didactic, explicit educational experiences outside of the clinical setting. The Association for American Medical Colleges (2008) asserts that the aim of medical education is to “equip each student to develop the lifelong habit of skill learning and self-assessment, in order to further develop and maintain clinical performance expertise” (p. 5). To do so, students need to be involved in a multitude of hands-on learning experiences to further advance clinical skills, clinical problem solving, and medical knowledge.

The surgical clerkships, or surgical block which typically incorporates obstetrics and gynecology, surgery, and anesthesiology are unique as these are conducted with, “rapid transitions from one discipline to another, one patient population to another, one setting to another, and from one culture to another” (Irby, Cooke, & O’Brien, 2010, p. 84). As such, this fast-paced, transient learning environment becomes a revolving door, quickly pushing medical students through specialties (Bell, et al., 2008; Hammoud, et al., 2006). This revolving door becomes further exacerbated due to the limited time faculty have to plan, teach, mentor, observe, and evaluate medical students (Irby, Cooke, O’Brien, 2010). Thus, this environment further asserts the need for multifaceted, active opportunities which provide for a blended learning

experience. An environment where students learn independently to build a foundation of knowledge as well as interact with peers to engage in critical problem solving to further reinforce the necessary medical content.

Due to the various methods of content presentation, flipped learning has slowly gained popularity among specialties in medical education. Surgery, radiology, obstetrics and gynecology, intensive care, pharmacology, and emergency medicine (Belfi et al, 2015; Heitz et al, 2015; Khanova, et al., 2015; Liebert et al, 2016; Morgan et al., 2014; Tainter et al., 2016) have all explored the implementation of flipped learning with medical students. ‘

The research conducted on the radiology clerkship presented by Belfi et al (2015) employed flipped lectures as a method to address the reduced amount of teaching time from 20 hours to 12 hours. Medical students were given an outline of the topics and the available online modules called: *RadCasts*, *RadTutorials*, *RadGames*, and *Interactive Clinical Anatomy and Radiology Utilization Simulator (ICARUS)*. The *RadCasts* were short 10-15 minute PowerPoint presentations that reviewed the clinical approach to imaging with self-assessment questions students needed to complete. To review specific clinical conditions or a specific patient population, students accessed *RadTutorials*. *RadGames* were more interactive modules that students completed regarding chest diagnosis. The simulator *ICARUS* incorporated online simulation that encouraged students to use various problem-solving skills. The in-class sessions were mindfully planned, taking into consideration the prior knowledge students were constructing with the online components. Case-based teaching was implemented, with students encouraged to use an audience response system, a Jeopardy-type review game was facilitated to build on students’ collaborative skills, and additional face-to-face simulation workshops were facilitated. Overall, this pedagogical design was very successful two-fold, medical students achieved

higher posttests after the flipped classroom and students perceived the inclusion of such diverse teaching practices to be beneficial to their learning.

Of course, with any curricular change, challenges emerge which need to be addressed. “Practical considerations” when implementing flipped learning are documented by Tainter et al. (2016, pg. 4). First is the resistance to change of the traditional medical education framework. The argument being that medical education continues to prepare excellent physicians, why bother changing what has been in place for decades. The second deterring factor is of course, time and cost. Teaching faculty will need to make time outside of their clinical obligations to learn of the various technological resources and how those are implemented with learners. Third is a concern that students will not complete the online assignment(s) prior to meeting in the face-to-face setting. The fourth challenge is “resistance to active learning” (pg. 5), which embodies two constructs. First, there are students who prefer to be passive participants in the learning environment rather than engaging with peers during group discussion. Second, some students may be uncomfortable when teaching faculty put more ownership and responsibility on the learners during case-based or team-based learning experiences. Any one of these barriers may preclude teaching faculty from attempting or successfully integrating flipped learning into the curriculum.

SOLUTIONS AND RECOMMENDATIONS

To alleviate potential challenges in the field of medical education, implementing flipped learning opportunities provides greater flexibility for student and teaching faculty. Students have the ability to view and understand material prior to engaging with the content at a higher level in the classroom. By providing learning objectives online or virtually, teaching faculty have clearly delineated the expectations in the acquisition of

content. Rather than simply assign a chapter to read in a text, employing multiple modalities of technology allows students to further scaffold their learning (Vygotsky, 1978). This method encourages medical educators to plan wisely and outline the required learning goals and targets imperative for the student’s success. Furthermore, this process assures for greater transparency between the teaching faculty, the content, and the student as the requirements and learning objectives are clearly delineated through the selected material.

The structure of a flipped learning experience requires that the teaching faculty develop an online component for students to interact with. Detailed recommendations provided by the AAMC (2008) and research conducted by Moffett (2014) identify how various technological resources can be incorporated (see Table 2) to enhance the flipped learning environment. Integration of social networking, social media, podcasts, and simulations are a few of the examples teaching faculty may include to encourage students to construct an understanding of the required content. Sharma et al. (2014) emphasized the importance for faculty educators to also interact with the students online prior to the face-to-face activity. This encourages students to become more immersed in the content as they engage with peers and teachers. Finally, to integrate appropriate online content, teaching faculty who develop these learning experiences must keep the learning objectives and goals in mind to assure that students are receiving the necessary material throughout the entire experience. As such, an adequate timeline must be established to allow students enough time to navigate through the online material prior to entering the face-to-face classroom.

When students do enter the face-to-face classroom, teaching faculty must be prepared with advanced content and avoid simply reiterating the information already presented online. This may be a complex case that students must collaboratively solve. Team-based learning is another strategy for engagement in the classroom. This approach incorporates three steps that first, requires students

to complete a pretest independently. During the second step, students collaborate in teams to further conceptualize the material, and finally, at the conclusion of the class, students complete a posttest with their assigned group of peers. This process encompasses the strengths of various learning styles, specifically independent work to collaborative work. An additional strategy teaching faculty could employ is case-based learning (CBL) which encourages students to engage in self-directed learning. CBL integrates a particular and detailed scenario, known as the case which students need to critically evaluate (Gade & Chari, 2013). This type of strategy allows students to work independently or in a group solving exploratory questions.

Technology is a tremendous component of both the online, synchronous portion and the in-class asynchronous teaching. Traditionally, voiced over PowerPoint presentations or short video clips are used, however, there are a number of other online tools to make learning more dynamic (see Table 1).

The above table is not exhaustive of the current available resources applicable for the classroom. As technology and research advances are made, many more applications and online tools will be developed to improve and increase engagement in this blended classroom environment.

FUTURE RESEARCH DIRECTIONS

There is an ongoing need to bridge the gap between educational theory and practice in medical education institutions (Kim, et al., 2014). To accomplish this, additional research should be conducted to further investigate the effectiveness of flipped learning on the development of the student, especially in relation to the acquisition and retention of complex medical content and the implications on patient care. Little research exists regarding the retention and implementation of knowledge as medical students transition to different disciplines, into a residency program, or into their professional career. Flipped learning experiences are multifaceted and it is difficult to pinpoint specifically, which area impacts student learning.

A secondary research area is to investigate the ways in which teaching faculty develop their own professional pedagogical practice. By developing flipped learning experiences, faculty have the potential to also expand their own teaching practice by shifting control away from the “sage on the stage” approach, to instead empower rising physicians. This necessitates supplementary research on teaching faculty’s willingness and acceptance to integrate new technologies and strategies and,

Table 1. Examples of flipped lecture tools

Flipped Lecture Tool	Brief Description
ed.ted.com	Use YouTube to create private videos and embed quiz questions, links, and additional content.
Poll Everywhere	Online application for in-class response system. Can create graphs and real time tallies of who is responding and where misconceptions exist.
Loca Moda	Students can “post” a text message onto an online bulletin board to stimulate discussion.
Google Hangouts	Students can engage in an online, password protected chat room which saves discussions for future reference.
Kahoot	A free, online application that allows for in-class surveys and quizzes.
Articulate Storyline 2	A software program which allows the user to design interactive, online modules with questions, videos, short readings, and simulations.
Wikispaces	Free online tool to allow students and teachers to collaborate and discuss questions and online cases.
TEDTalks and TEDxTalks	Both TEDTalks and TEDxTalks have developed into a vast library of videos that reach a global audience to inspire critical thinking.

in turn, how these new learning experiences are perceived by students.

CONCLUSION

Flipped learning has generated interest and action for change in the education community, from K-12 through to medical education. With the advancements and increased accessibility of technology, students have the opportunity to engage with complex content through various modalities, addressing the diverse learning needs of today's medical student. This curricular process requires that medical educators are well versed in their pedagogical practice in order to design, organize, and implement a flipped learning experience.

REFERENCES

- Association of American Medical Colleges. (2008). *Recommendations for clinical skills curricula for undergraduate medical education*. Task Force on the Clinical Skills Education of Medical Students.
- Belfi, L. M., Bartolotta, R. J., Giambrone, A. E., Davi, C., & Min, R. J. (2016). Flipping the introductory clerkship in radiology: Impact on medical student performance and perceptions. *Academic Radiology*, 22(6), 794–801. doi:10.1016/j.acra.2014.11.003 PMID:25592027
- Bergmann, J. & Sams, A. (2012). *Flip your classroom. Reach every student in every class every day*. Academic Press.
- Bishop, J. L., & Verleger, M. A. (2013). The flipped classroom: a survey of the research. *120th ASEE Annual Conference and Exposition*, Atlanta, GA.
- Dankbaar, M. E., & de Jong, P. G. (2014). Technology for learning: How it has changed education. *Perspectives on Medical Education*, 3(4), 257–259. doi:10.1007/s40037-014-0141-0 PMID:25163662
- Feng, J. Y., Chang, Y. T., & Change, H. Y. (2013). Systematic review of effectiveness of situated e-learning on medical and nursing education. *Worldviews on Evidence-Based Nursing*, 10(2), 174–183. doi:10.1111/wvn.12005 PMID:23510119
- Heitz, C., Prusakowski, M., Willis, G., & Franck, C. (2015). Does the concept of the flipped classroom extend to the emergency medicine clinical clerkship? *Western Journal of Emergency Medicine*, 16(6), 851–855. doi:10.5811/westjem.2015.9.27256 PMID:26594277
- Jashapara, A., & Tai, W. C. (2011). Knowledge mobilization through e-learning systems: Understanding the mediating roles of self-efficacy and anxiety on perceptions of ease of use. *Information Systems Management*, 28(1), 71–83. doi:10.1080/10580530.2011.536115
- Khanova, J., McLaughlin, J. E., Rhoney, D. H., Roth, M. T., & Harris, S. (2015). Student perceptions of a flipped pharmacotherapy course. *American Journal of Pharmaceutical Education*, 79(9), 1–8. doi:10.5688/ajpe799140 PMID:26839429
- Kim, M. K., Kim, S. M., Khera, O., & Getman, J. (2014). The experience of three flipped classrooms in an urban university: An exploration of design principles. *The Internet and Higher Education*, 22, 37–50. doi:10.1016/j.iheduc.2014.04.003
- King, A. (•••). From sage on the stage to guide on the side. *College Teaching*, 93(41), 30–36.
- Lage, M. J., Platt, G. J., & Treglia, M. (2000). Inverting the classroom: A gateway to creating an inclusive learning environment. *Economic Instruction*, 31(1), 30–43.
- Liebert, C. A., Lin, D. T., Mazer, L. M., Bereksnyi, S., & Lau, J. N. (2016). Effectiveness of the surgery core clerkship flipped classroom: A prospective cohort trial. *American Journal of Surgery*, 21(2), 451–457. doi:10.1016/j.amjsurg.2015.10.004 PMID:26687962

Mazur, E. (2009). Farewell, lecture? *Science*, 322(5910), 50–51. doi:10.1126/science.1168927 PMID:19119207

McKeachie, W., & Svinicki, M. (Eds.). (2006). *McKeachie's Teaching Tips: Strategies, Research, and Theory for College and University Teachers*. Florence: Cengage Learning.

McLaughlin, J. E., Roth, M. T., Glatt, D. M., Gharkholonarehe, N., Davidson, C. A., Griffin, L. M., & Mumper, R. J. et al. (2014). The flipped classroom: A course redesign to foster learning and engagement in a health professions school. *Academic Medicine*, 89(2), 236–243. doi:10.1097/ACM.0000000000000086 PMID:24270916

Moffett, J. (2015). Twelve tips for flipping the classroom. *Medical Teacher*, 37(4), 331–336. doi:10.3109/0142159X.2014.943710 PMID:25154646

Morgan, H., Marzano, D., Lanham, M., Stein, T., Curran, D., & Hammoud, M. (2014). Preparing medical students for obstetrics and gynecology milestone level one: A description of a pilot curriculum. *Medical Education Online*, 19(0). doi:10.3402/meo.v19.25746 PMID:25430640

O'Brien, B. C., & Irby, D. M. (2013). Enacting the Carnegie Foundation call for reform of medical school and residency. *Teaching and Learning in Medicine*, 25(51), 51–58. PMID:24246101

Pierce, R., & Fox, J. (2012). Vodcasts and active-learning exercises in a flipped classroom model of a renal pharmacotherapy module. *American Journal of Pharmaceutical Education*, 76(10), 1–5. doi:10.5688/ajpe7610196 PMID:23275661

Prober, C. G., & Heath, C. (2001). Lecture halls without lectures - a proposal for medical education. *The New England Journal of Medicine*, 366(18), 1657–1659. doi:10.1056/NEJMp1202451 PMID:22551125

Sharma, N., Lau, C. S., Doherty, I., & Harbutt, D. (2015). How we flipped the medical classroom. *Medical Teacher*, 57(4), 327–330. doi:10.3109/0142159X.2014.923821 PMID:24934251

Tainter, C. R., Wong, N. L., Cudemus-Deseda, G. A., & Bittner, E. A. (2016). The flipped classroom model for teaching in the intensive care unit: Rational, practical considerations, and an example of successful implementation. *Journal of Intensive Care Medicine*. doi:10.1177/0885066616632156 PMID:26912409

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

ADDITIONAL READING

Beetham, H., & Sharpe, R. (2007). *Rethinking pedagogy for a digital age: Designing for 21st century*. New York, NY: Routledge.

Bell, B., & Gilbert, J. (1994). Teacher development as professional, personal, and social development. *Teaching and Teacher Education*, 10(5), 483–497. doi:10.1016/0742-051X(94)90002-7

Cochran-Smith, M. (2003). Learning and unlearning: The education of teacher educators. *Teaching and Teacher Education*, 19(1), 5–28. doi:10.1016/S0742-051X(02)00091-4

Dewey, J. (1916). *Democracy and education: An introduction to the philosophy of education*. New York, NY: Macmillan.

Gade, S., & Chari, S. (2013, December 01). Case-base learning in endocrine physiology: An approach toward self-directed learning and the development of soft skills in medical students. *Advances in Physiology Education*, 37(4), 356–360. doi:10.1152/advan.00076.2012 PMID:24292913

Gillois, P., Bosson, J. L., Genty, C., Vuillez, J. P., & Romanet, J. P. (2015). The impacts of blended learning design in first year medical studies. *Studies in Health Technology and Informatics*, 210, 607–611. PMID:25991220

- Gorres-Martens, B. K., Segovia, A. R., & Pfefer, M. T. (2016, March). Positive outcomes increase overtime with the implementation of a semiflipped teaching model. *Advances in Physiology Education*, 40(1), 32–37. doi:10.1152/advan.00034.2015 PMID:26847255
- Grossman, E., Grosseman, S., Azevedo, G. D., Figueiro-Filho, E. A., & McKinley, D. (2015). Flipped classroom on humanities: Medicine, narrative and art. *Medical Education*, 49(11), 1142. doi:10.1111/medu.12861 PMID:26494071
- Hill, H. C., Ball, D., & Schilling, S. G. (2008). Unpacking pedagogical content knowledge: Conceptualizing and measuring teachers' topic-specific knowledge of students. *Journal for Research in Mathematics Education*, 372–400.
- Leung, J. Y. C., Kumta, S. M., Jin, Y., & Yung, A. L. K. (2014). Short review of the flipped classroom approach. *Medical Education*, 48(11), 1127. doi:10.1111/medu.12576 PMID:25307666
- Lew, E. K. (2016). Creating a contemporary clerkship curriculum: The flipped classroom model in emergency medicine. *International Journal of Emergency Medicine*, 9(25), 1–5. PMID:27623956
- Liebert, C. A., Lin, D. T., Mazer, L. M., Bereknyei, S., & Lau, J. N. (2016). Effectiveness of the surgical core clerkship flipped classroom: A prospective cohort trial. *American Journal of Surgery*, 211(2), 451–457. doi:10.1016/j.amjsurg.2015.10.004 PMID:26687962
- Lindeman, B. M., Law, J. K., Lipsett, P. A., Arbella, T., Stem, M., & Lidor, A. O. (2015). A blended online curriculum in the basic surgery clerkship: A pilot study. *American Journal of Surgery*, 209(1), 145–151. doi:10.1016/j.amjsurg.2014.10.003 PMID:25466767
- Lochner, L., Wieser, H., Waldboth, S., & Mischo-Kelling, M. (2016). Combining traditional anatomy lectures with e-learning activities: How do students perceive their learning experiences? *International Journal of Medical Education*, 7, 69–74. doi:10.5116/ijme.56b5.0369 PMID:26897012
- McLaughlin, J. E., Roth, M. T., Glatt, D. M., Gharkholonarehe, N., Davidson, C. A., Griffin, L. M., & Mumper, R. J. et al. (2014). The flipped classroom: A course redesign to foster learning and engagement in a health professions school. *Academic Medicine*, 89(2), 236–243. doi:10.1097/ACM.0000000000000086 PMID:24270916
- Missildine, K., Fountain, R., Summers, L., & Gosselin, K. (2013). Flipping the classroom to improve student performance and satisfaction. *The Journal of Nursing Education*, 52(10), 597–599. doi:10.3928/01484834-20130919-03 PMID:24044386
- Ojennus, D. D. (2015). Assessment of learning gains in a flipped biochemistry classroom. *Biochemistry and Molecular Biology Education*, 44(1), 20–27. doi:10.1002/bmb.20926 PMID:26593859
- Whillier, S., & Lystad, R. P. (2015). No differences in grades or level of satisfaction in a flipped classroom for neuroanatomy. *The Journal of Chiropactic Education*, 29(2), 127–133. doi:10.7899/JCE-14-28 PMID:25902472
- Zappe, S., Leicht, R., Messner, J., Litzinger, T., & Lee, H. 2009. Flipping the classroom to explore active learning in a large undergraduate course. *Proc. 2009 Amer. Soc. Engin. Educ. Annu. Conf.*, Austin, TX.

KEY TERMS AND DEFINITIONS

Asynchronous e-Learning: The use of online learning resources to facilitate resources without constraints of time or place. Takes the form of discussion boards and blogs.

Flipped Lecture: Students engage with assigned content through online or mobile methods, then attend a face-to-face class to explore the content in more critical ways.

Pedagogy: The art or science of teaching.

Synchronous e-Learning: Online learning that takes place real-time from various remote

locations. Takes the form of videoconferencing and online chat rooms.

Theory of Active Learning: The belief that people learn best from experience and “doing”.

Transactional Distance Learning: Definition of distance education which implies the separation of teachers and learners (Moore, 2007).

Transformative Learning: How learners comprehend, validate, reflect on, and reformulate the meaning of their experience.

Zone of Proximal Development: A concept that informs how learners build on their learning to the ultimate goal of independence.

APPENDIX

M

Table 2. *Flipped lecture activities*

E-Learning Activities and Assessments	In-Class Activities and Assessments
• Online readings, interactive PDF • Reflective questions • Online survey • Online discussion • Pod cast • Videos • Online Simulation • Social networking (LinkedIn) • Social media (Twitter, Facebook)	• Use of audience response systems • Think Pair Share • Jig Saw activity • Case-study learning • Problem-based learning • Fishbowl discussion • Role play • Written reflections • Debate • Peer instruction • Problem-based learning • Simulation with standardized patient

McKeachie & Svinicki (2006); Moffett (2015).

Integrating Evidence-Based Practice in Athletic Training Through Online Learning

Brittany A. Vorndran

Seton Hall University, USA

Michelle Lee D'Abundo

Seton Hall University, USA

INTRODUCTION

Health care is always changing and as a result the associated health professions must adapt. As challenges persist such as rising health care costs, an aging population and the need to treat increasing numbers of people with chronic health conditions, the field of health care must continue to explore ways to deliver quality care while reducing costs. Like many other health care professions, athletic training has turned to evidence-based practice to assure that athletic trainers are trained to deliver the highest quality of care in the most efficient way to their patients. The transition to integrating evidence-based practice can be challenging and will require a massive diffusion of innovation throughout the field of athletic training.

Athletic trainers (ATs) are defined as “health care professionals who collaborate with physicians. The services provided by ATs comprise prevention, emergency care, clinical diagnosis, therapeutic intervention and rehabilitation of injuries and medical conditions. ATs work under the direction of physicians, as prescribed by state licensure statutes” (Profile of athletic trainers, 2014, para. 1). The profession of athletic training is constantly growing, and athletic trainers can be found in many different settings including high schools, colleges, working with the athletic teams, the military, and workplaces. Despite many employment opportunities, some people in the health care industry are unfamiliar with the responsibilities of the AT and it is apparent that

athletic training as a profession is still lagging behind other allied health professions (Hankemeier & Van Lunen, 2013a).

BACKGROUND

In order to change the perceptions of ATs to the public, the National Athletic Trainers' Association (NATA) is introducing the use of evidence-based practice (EBP) in both the education of new ATs, and in the clinical setting with athletic trainers already working in the field (Hankemeier & Van Lunen, 2013; Hankemeier et al., 2013; McCarty Hankemeier, Walter, Newton, & Van Lunen, 2013; Welch, Van Lunen, & Hankemeier, 2014b). There are many benefits to increasing the use of evidence-based medicine (EBM). One reason that is frequently discussed is improving both the image and recognition of athletic trainers as health care professionals and not personal trainers, physical education teachers, or the people carrying water bottles on the sideline of games. Other benefits to using EBM include improving the care provided to the patients and justifying third party reimbursement (McCarty et al., 2013; Welch et al., 2014a).

Using EBP can be broken down into five steps as shown by Sackett et al. (1996). These steps include: defining clinically relevant questions, searching for the best evidence, critically appraising the evidence, applying the evidence, and evaluating how effective evidence-based medicine was when put to use. While these steps

seem relatively simple when written out, most athletic training clinicians are not presently using EBP in their current treatment practice, but 98% of them were found to believe that it is important for the credibility of the profession (McCarty et al., 2013). Hankemeier et al. (2013b) showed that clinicians had a lower perceived importance score and a lower knowledge score when compared with post-professional educators. ATs who work in an athletic training education program have been more exposed to the notion of EBP and therefore are more knowledgeable regarding it and understand its importance more than individuals that only work in the clinical setting.

Since EBP is very new to the field of athletic training, practicing clinicians that are not recent graduates likely did not learn about EBP in the educational curriculum and would not be familiar with it unless the individual took initiative to learn the process independently. In a study completed by Hankemeier and Van Lunen (2013a) less than 20% of the surveyed clinicians had received any form of EBP training. Based on the information provided in this survey, very few clinicians have been trained to use EBP. These clinicians with no background in using the five steps would need some training in order to be able to effectively integrate EBP into daily clinical practice.

Disseminating knowledge about EBP to ATs is essential in this transition. Continuing education (CE) is an important method for educating athletic training clinicians on what EBP is and how to use. In order for CE to be truly effective, athletic trainers need to understand what modes are best for presenting this information to promote long term knowledge retention and knowledge translation. Popular ways that CE is presented is through online learning, in person lectures, discussions, hands on demonstrations, and mixed mode learning. Studying modes individually for these gains or comparing one type of CE with another will help to advance understanding on how EBP should be presented to maximize an increase in knowledge and use.

Distance education, e-learning, computer mediated, web-based, and online instruction are terms used to describe education delivered through computer-based technologies. Whatever term is chosen, such computer-based technologies are essential components of the preparation and continuing education of health professionals. One of the most frequently used modes of CE, particularly for learning EBP, is online learning. Web-based learning has become very popular due to its flexibility and ease. This type of learning can be done from anywhere and completed at the convenience of the individual taking the course (Militello, Gance-Cleveland, Aldrich, & Kamal, 2014).

Allied health care tends to use three models of online learning and instruction: blended learning, online learning, and continuing education (Stewart & Wright, 2004). Blended learning includes both face to face and online learning and instruction. Online learning is learning and instruction that is completely web-based. Continuing education for health professionals is offered in both blended and online formats. In fact, most health-related disciplines are using a combination of online and face-to-face learning in their programs.

From a theoretical standpoint, characteristics of an effective online learning experience have been outlined. Reeves and Reeves (2008) provided 10 dimensions to be considered when designing, implementing and evaluating an online class in health and social work. The model they described for health and social work education included 10 dimensions of interactive teaching and learning:

1. Pedagogical philosophy;
2. Learning theory;
3. Goal orientation;
4. Task orientation;
5. Source of motivation;
6. Teacher role;
7. Metacognitive support;
8. Collaborative learning support;
9. Cultural sensitivity; and
10. Structural flexibility.

In the discussion of developing online learning formats, best practices often discussed include attention to instructor and student interaction, application of active learning techniques, providing prompt feedback, and attending to diverse ways of learning. While the aforementioned guidelines make sense, they have not been vetted through research. This brings us to the next questions of whether web-based instruction can be as effective as face-to-face education. The answer is that we really do not know how effective face-to-face education is, because comprehensive assessment is rarely conducted of student learning for specific courses. Many factors play a role (i.e., faculty training and course design) in the effectiveness of any course. Most health professionals rely heavily on technology in their practice and thus, it seems appropriate for technology to play a role in the education of future health professionals. In fact, Hollis and Madill state (2006), "Evidence suggests that blending combinations of technologies with computer mediated learning enhances interaction and could address the higher order learning needs of professional programs such as occupational therapy" (p.61).

Through a review of research focused on computer-enhanced classroom-based teaching, online courses, online degree programs, and online continuing professional education, Hollis and Madill (2006) recommended that both online and face-to-face teaching and learning opportunities accommodate all types of students and faculty. Hollis and Madill (2006) emphasized the use of multiple technologies to enhance interaction between students, and between students and instructors. However, to optimize the effectiveness of online learning, they suggested appropriate preparation and support is required for both students and faculty. Again, these suggestions make sense, but also must be vetted through research designed to assess effective online learning experiences.

How is blended and online learning and education being used in health professions? Konin (2004) discussed how to enhance clinical education, like athletic training, through the use of web-

based programs. Like most health professionals, education and training includes some type of fieldwork, or internship placement. Information and assessments for clinical instructors can be delivered electronically through an online course management system such as Blackboard. Face-to-face class formats can be enriched through the availability of a multitude of supplemental online materials.

A study was completed with nursing clinical preceptors who worked with undergraduate nursing students. In this study the preceptors took part in a series of online CE course objectives. These objectives were geared toward the roles and expectations of a clinical preceptor. The results showed that there was a statistically significant gain of knowledge between the pre-test and post-test, showing that knowledge was gained from online learning (Zahner, Tipple, Rather, & Schendzielos, 2009). While this study showed that online learning can increase knowledge, it did not compare online learning with traditional learning methods such as live lectures.

There is research that has been done that compares online learning with traditional live lectures. One study completed by Phillips (2015) investigated the effects of teaching a lesson online versus the traditional in person lecture class. This study was completed using a class in a pharmacy program. Students were randomly assigned to either the control group of traditional lectures, or the experimental group of active online learning. The experimental group performed higher on a multiple choice quiz given immediately after the lesson, but there was no significant difference on the final exam scores between groups. The experimental group did express that the online class was enjoyable and felt that it should be used in conjunction with traditional lectures, but should not replace lectures altogether.

Online learning and instruction is also essential to the continuing education of health-related professionals. Williams (2006) conducted a meta-analysis of 25 studies assessing achievement in distance education in allied health and found that

distance education is as effective as face-to-face instruction. Student achievement outcomes were positively correlated with work experience and professional knowledge. Outcomes were also better for distance education experiences that included multiple learning and instructional strategies including, but not limited to, interaction and introspection.

Continuing education for ATs is offered in many different forms including taking college courses, publishing papers, presenting at conferences, attending in person meetings or education seminars, and online learning (Board of Certification, 2016). Studies have shown that online CE is frequently used because of the cost effectiveness and the ease of completing it individually (Armstrong & Weidner, 2011). December 2015 was the first reporting period that required athletic trainers to complete CE in an evidence-based practice category. During this period, ATs could choose from many different online options. These were typically formatted with a lesson that ended in a quiz to test knowledge. Passing the quiz was necessary for receiving the continuing education units. Because this reporting period ended so recently, there is no research yet available on its effectiveness. However, a study was completed by Welch et al. (2014a) that looked at the web-based modules pertaining to EBP.

In this study, there was an experimental group that had access to the continuing education learning modules, and a control group that did not. Both groups were given a pretest and a posttest to assess knowledge. The experimental group did significantly better than both their own pretest and the control group's scores on the pretest and the posttest. Learning was achieved in this study, but only short term knowledge gain was investigated. Whether this knowledge was retained in the long term from the online learning modules has not yet been concluded.

The experimental group in the previous study was also interviewed on their perceptions of evidence-based practice following the interven-

tion. Overall, the response was positive with the individuals all feeling that they had learned from the modules and had a better understanding of the concepts of EBP and how to use it (Welch et al., 2014b). However, the clinicians reported that the modules did not cause a change in clinical practice and that although they learned they did not change the way they made treatment decisions. Other qualitative studies have also been done regarding the attitudes toward EBP education. A topic that was emphasized was the need for constant repetition and increased exposure to evidence-based practice. These participants felt that the more exposure one has to EBP, the more comfortable they will become with it and the more likely they will become to use it (Welch et al., 2014c). A recent study by Keeley et al. (2016) found that participants felt that EBP was important, but overall they had not used EBP concepts at all in the eight weeks prior to completing the survey. The importance of EBP and short-term knowledge gain seem to be occurring, but there is still a gap between putting that knowledge to practice.

The reason that EBP is such a hot topic in the world of health care is to improve clinical outcomes, decrease cost, and increase patient satisfaction. To effectively implement EBP into athletic training, the athletic trainers must be using what they are learning through CE (Hankemeir & Van Lunen, 2013; Welch et al., 2014a; Welch et al., 2014b). More research needs to be done that helps to understand whether online, traditional, or a mixed mode of CE is most effective for knowledge gain and clinical use. Research that measures how well and individual is learning and whether that person is retaining the knowledge will help to make adjustments to the current situation regarding EBP. Comparing different ways of receiving the knowledge for effectiveness will also help to promote meaningful learning and knowledge translation. December of 2015 is the first time that ATs needed to report CEUs from an EBP category. It is likely that this will not be perfect on the first attempt. Continuing to understand the best ways

to increase the translation of knowledge in athletic training will help the profession to successfully incorporate EBP into daily practice.

Some studies have been completed that investigate the notion of using a mixed mode method of learning. This type of CE would include different methods of learning such as combining web-based learning with an in-person discussion, or hands on demonstration that compliments what was previously learned (Militello et al., 2014). The idea that mixing a passive lecture whether online or in person with an active discussion or demonstration works well with the idea that individuals learn differently. Furthermore, reinforcing a topic often promotes long term retention, comprehension, and may increase the chance that it will be implemented into clinical use (Schreiber, Perry, Downey, & Williamson, 2013; Zahner, et al., 2009).

A particular study that utilized a more comprehensive and multifaceted approach was completed by Schreiber et al. (2013). This study used a course that was geared for physical therapist (PTs) and met on three separate occasions in conjunction with an online community used to foster discussion. During the in-person meetings, there were lectures, hands on demonstrations, and time to talk about successes and barriers to implementing the clinical change learned. In the final session, all PTs were required to discuss how they had already used what was learned in the previous courses in the clinical setting with patients. They also needed to describe an EBP activity that they had undertaken in the clinical setting. This study showed that the participants enjoyed this mixed methods learning and showed improvement in knowledge and practice behaviors such as supporting colleagues' use of interventions and tests and measures that were related to the course content.

Studies are showing that multifaceted learning increases knowledge and the translation of knowledge into clinical practice (Schreiber et al., 2013; Zahner et al., 2009). In order for change to truly occur in the field of athletic training more research needs to be completed that investigates how effective the current CE learning methods are.

Short term knowledge gain is not enough to truly make a change. Whether long term knowledge gain is occurring is also important. If an individual is forgetting the course content immediately following the course then the information is not being relayed effectively (Phillips, 2015). This information must be something that the AT can comprehend and remember after a period of time. Even more importantly, however, is the idea of translating this knowledge in to clinical practice.

FUTURE RESEARCH DIRECTIONS

As previously mentioned, one of the benefits to online continuing education is that it can be completed at the convenience of the individual. This however, can also be a drawback. Each reporting period for continuing education is two years (Board of Certification, 2016). During that two-year period, athletic trainers can complete all of the CEUs at any time. This means that the CEUs can be completed two years early, or one day before the reporting period ends. It may be more beneficial to the profession if the EBP continuing education had benchmarks throughout the two year period that needed to be met. This would allow for continual exposure to the concepts and the idea of EBP throughout the reporting period, without it all being crammed into a few days or weeks and then forgotten about. Online learning would easily allow for this to happen with lessons and quizzes having due dates throughout the reporting period.

A more interactive online learning experience may be beneficial. Most of the online continuing education to this point includes a lesson followed by a quiz at the end (Welch et al., 2014a). This allows for an individual to pay minimal attention to the lecture itself and only focus on the quiz at the end. However, if there were checkpoints throughout the lecture where the individual had to answer a multiple choice question, or fill in the blank there would be more attention paid to the lecture. This may also increase interest in the topic. The quizzes that are used at the end

should also be evaluating a true understanding of the topics that are being presented and should be challenging enough that individuals are proving that they understand the information that was presented. Online continuing education is received positively by athletic trainers due to the cost and convenience, so making it effective is important to increasing the use of EBP (Armstrong, 2011).

Continuing education in EBP was provided through online courses, at sessions at the national and regional conferences, and at in-person seminars throughout the country. The way an individual athletic trainer received the CEUs was a personal decision. It is frequently shown that expense and convenience of continuing education play a very significant role in why it is selected (Cuppitt, 2001). Athletic trainers that work in a traditional setting such as a high school or college and are working unpredictable hours and weekends may be more inclined to complete the CEUs online due to convenience. Additionally, many ATs are paying for the continuing education independently and therefore will pick the most cost effective option. Unlike students in an athletic training program who will receive EBP information in a formal didactic setting that is constantly reiterated, certified ATs are only receiving this information when it is presented in the CE and through an additional education sought out individually (McCarty et al., 2013). Certified ATs may take longer to truly feel comfortable with the concept of EBP and may need additional reinforcement beyond the CE to promote a behavior change.

One of the major recurring themes relating to barriers of implementing EBM is that ATs are not comfortable with the process and therefore will not use it (Hankemeier & Van Lunen, 2013a). Many of the practicing athletic trainers were not taught how to use the process of EBP while in school resulting in not being comfortable reading scholarly articles and deciding what should be considered when making treatment decisions. This is one of the reasons that EBP is now in all athletic training education curriculums, but this does not help ATs that are already practicing. The

other major complaint is that the ATs working in many of the various clinical settings do not have time to go through the whole process of using EBP, particularly reading through lengthy articles to help with treatment decisions (McCarty et al., 2013; Welch, 2014a). Truly analyzing a peer-reviewed article to determine its value can be very time consuming. In a clinical environment where athletes and other patients may constantly be around and needing attention, it can be difficult to take the time to read articles and determine whether it has any relevant information that can be applied to the particular situation. Due to the time consuming nature, many ATs feel it is easier to just do what is comfortable or what has always been done because it allows patients to be seen and taken care of in a timelier manner (McCarty et al., 2013; Welch et al., 2014a). One study indicated that ATs stated that if the information in the articles were in a quicker, more streamlined medium there was a higher chance of reading it and utilizing it because it would be less time consuming (Welch et al., 2014a).

A possible way to help ATs understand scholarly articles is through the use of an online database. Since ATs understandably would be challenged with reading scholarly articles, it seems like asking researchers to provide EBP application summaries of their articles to an online open access database for ATs would be a good solution. The *Journal of Athletic Training* (JAT) is currently available to all members of the NATA. This journal comes out quarterly with monthly supplements and is available online. Currently ATs have access to the abstract, full text, and a PDF version. This is a great asset to ATs, but a frequent complaint is that using EBP and looking through research articles is too time consuming for the fast paced unpredictable nature of the job (McCarty et al., 2013). Since this journal is published for athletic trainers and typically by athletic trainers it would be beneficial if each author were asked to provide 3-5 bullet points that were directly related to applying the information in the article to practice. These bullet points could be added to the end of

the abstract and would help to streamline the information in the article which has been requested by practicing ATs (Welch et al., 2014a). The abstract and the EBP bullet points would allow ATs to quickly look to see whether an article is worth taking the time reading through and may allow clinical ATs to put more evidence into practice without spending valuable time reading through something to end up not using it.

If this were to become successful and widely used the NATA could also begin to provide research from other journals outside of JAT. While the JAT is a great source for information for athletic trainers, there are many other journals that also produce useful information from other allied health professions. It could be beneficial for ATs to have exposure to research that is being done in other field such as nursing and physical therapy. This exposure would broaden the knowledge of ATs and could increase interprofessional collaboration. Finding a way for ATs to be able to quickly determine whether an article is worth reading could be one way to increase the use of EBP and using bullet points that are provided with the abstract is one possible solution.

CONCLUSION

A major transition is occurring to achieve the implementation of EBP throughout the field of AT. In December of 2015, athletic training professionals were required to report ten of their fifty CEUs in an EBP category. This was the first time that athletic trainers needed a portion of CEUs to come from an EBP section. This means that all athletic trainers who would like to maintain certification have recently taken courses either online or in person that teach about EBP. The goal is that this will familiarize all ATs with EBP in hopes that it will begin to be used more. However, since this is the first reporting period that required the EBP category, there is little known about how effective adding CEUs in the field of EBP will be and whether knowledge will be gained and translated

into practice. Changing the way online continuing education is presented and completed by ATs may increase comprehension and understanding of EBP. This understanding may lead to using it in clinical practice. Having benchmark requirements for completing the CEUs throughout the two year period could lead to more frequent exposure to all ATs and may increase the understanding. This checkpoint process would prevent ATs from completing all CE at the beginning or end of the reporting period, and would instead spread it out throughout the two years allowing for increased exposure. One major complaint for using EBP in the clinic is a lack of comfort with how to use it (Hankemeier & Van Lunen, 2013a). Increasing frequency of exposure may help to make clinicians more comfortable with how to use EBP causing a change in practice. While there will no doubt be challenges associated with the implementation of EBP in athletic training, the use of online learning mentioned in this chapter may ease the transition for athletic trainers, managers, and organizations.

REFERENCES

- Armstrong, K. J., & Weidner, T. G. (2011). Preferences for and barriers to formal and informal athletic training continuing education activities. *Journal of Athletic Training, 46*(6), 680–687. PMID:22488195
- Athanassiou, N., McNett, J. M., & Harvey, C. (2003). Critical thinking in the management classroom: Blooms taxonomy as a learning tool. *Journal of Management Education, 27*(5), 533–555. doi:10.1177/1052562903252515
- Board of Certification. (2016). *Certification maintenance requirements*. Retrieved from <http://www.bocatc.org/ats/maintain-certification/continuing-education>
- Bridges, W. (2009). *Managing transitions: Making the most of the change* (3rd ed.). Philadelphia, PA: Perseus Books Group.

- Cuppert, M. M. (2001). Self-perceived continuing education needs of certified athletic trainers. *Journal of Athletic Training, 36*(4), 388–395. PMID:12937481
- Hankemeier, D. A., & Van Lunen, B. L. (2013). Perceptions of approved clinical instructors: Barriers in the implementation of evidence-based practice. *Journal of Athletic Training, 48*(3), 382–393. doi:10.4085/1062-6050-48.1.18 PMID:23675798
- Hankemeier, D. A., Walter, J. M., McCarty, C. W., Newton, E. J., Walker, S. E., Pribesh, S. L., & Van Lunen, B. L. et al. (2013). Use of evidence-based practice among athletic training educators, clinicians, and students, part 1: Perceived importance, knowledge, and confidence. *Journal of Athletic Training, 48*(3), 394–404. doi:10.4085/1062-6050-48.2.16 PMID:23675799
- Hertel, J. (2005). Research training for clinicians: The crucial link between evidence-based practice and third party reimbursement. *Journal of Athletic Training, 40*(2), 69–70. PMID:15970951
- Hollis, V., & Madill, H. (2006). Online learning: The potential for occupational therapy education. *Occupational Therapy International, 13*(2), 61–78. doi:10.1002/oti.209 PMID:16862957
- Keeley, K., Walker, S. E., Hankemeier, D. A., Martin, M., & Cappaert, T. A. (2016). Athletic trainers beliefs about and implementation of evidence-based practice. *Journal of Athletic Training, 51*(1), 35–46. doi:10.4085/1062-6050-51.2.11 PMID:26845629
- Konin, J. (2004). Using a Web-based program to enhance clinical education. *Athletic Therapy Today, 9*(1), 20–23. doi:10.1123/att.9.1.20
- McCarty, C. W., Hankemeier, D. A., Walter, J. M., Newton, E. J., & Van Lunen, B. L. (2013). Use of evidence-based practice among athletic training educators, clinicians, and students, part 2: Attitudes, beliefs, accessibility, and barriers. *Journal of Athletic Training, 48*(3), 405–415. doi:10.4085/1062-6050-48.2.19 PMID:23675800
- Militello, L. K., Gance-Cleveland, B., Aldrich, H., & Kamal, R. (2014). A methodological quality synthesis of systematic reviews on computer-mediated continuing education for healthcare providers. *Worldviews on Evidence-Based Nursing, 11*(3), 177–186. doi:10.1111/wvn.12041 PMID:24865984
- National Athletic Trainers' Association. (2014). *Education Overview*. Retrieved from <https://www.nata.org/about/athletic-training/education-overview>
- National Athletic Trainers' Association. (2014). *Profile of athletic trainers*. Retrieved from http://www.nata.org/sites/default/files/Athletic_Trainer_Profile.pdf
- National Athletic Trainers' Association. (2016). *Emerging Settings*. Retrieved from <https://www.nata.org/professional-interests/emerging-settings>
- Phillips, J. A. (2015). Replacing traditional live lectures with online learning modules: Effects on learning and student perceptions. *Currents in Pharmacy Teaching & Learning, 7*(6), 738–744. doi:10.1016/j.cptl.2015.08.009
- Reeves, P., & Reeves, T. (2008). Design considerations for online learning in health and social work education. *Learning in Health and Social Care, 7*(1), 46–58. doi:10.1111/j.1473-6861.2008.00170.x
- Sackett, D. L., Rosenberg, W. M., Muir Gray, J. A., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: What it is and what it isn't. *British Medical Journal, 312*(7023), 71–72. doi:10.1136/bmj.312.7023.71 PMID:8555924
- Schreiber, J., Perry, S., Downey, P., & Williamson, A. (2013). Implementation of an innovative continuing education program focused on translation of knowledge into clinical practice. *Journal, Physical Therapy Education, 27*(3), 63–71.
- Stewart, J., & Wright, V. (2004). E-learning in allied health care. *Athletic Therapy Today, 9*(1), 7–9. doi:10.1123/att.9.1.7

Welch, C. E., Hankemeier, D. A., Wyant, A. L., Hays, D. G., Pitney, W. A., & Van Lunen, B. L. (2014a). Future directions of evidence-based practice in athletic training: Perceived strategies to enhance the use of evidence-based practice. *Journal of Athletic Training, 49*(2), 234–244. doi:10.4085/1062-6050-49.2.15 PMID:24568230

Welch, C. E., Van Lunen, B. L., & Hankemeier, D. A. (2014b). An evidence-based practice educational intervention for athletic trainers: A randomized control trial. *Journal of Athletic Training, 49*(2), 210–219. doi:10.4085/1062-6050-49.2.13 PMID:24568228

Welch, C. E., Van Lunen, B. L., Hankemeier, D. A., Wyant, A. L., Mutchler, J. M., Pitney, W. A., & Hays, D. G. (2014c). Perceived outcomes of web-based modules designed to enhance athletic trainers knowledge of evidence-based practice. *Journal of Athletic Training, 49*(2), 220–233. doi:10.4085/1062-6050-49.2.14 PMID:24576306

Williams, S. (2006). The effectiveness of distance education in allied health science programs: A meta-analysis of outcomes. *American Journal of Distance Education, 20*(3), 127–141. doi:10.1207/s15389286ajde2003_2

Zahner, S. J., Tipple, S. M., Rather, M. L., & Schendzielos, C. (2009). Supporting nurse preceptors through online continuing education. *Journal of Continuing Education in Nursing, 40*(10), 468–474. doi:10.3928/00220124-20090923-01 PMID:19831329

Zidarov, D., Thomas, A., & Poissant, L. (2013). Knowledge translation in physical therapy: From theory to practice. *Disability and Rehabilitation, 35*(18), 1571–1577. doi:10.3109/09638288.2012.748841 PMID:23339718

ADDITIONAL READING

Hankemeier, D. A., Walter, J. M., McCarty, C. W., Newton, E. J., Walker, S. E., Pribesh, S. L., & Van Lunen, B. L. et al. (2013). Use of evidence-based practice among athletic training educators, clinicians, and students, part 1: Perceived importance, knowledge, and confidence. *Journal of Athletic Training, 48*(3), 394–404. doi:10.4085/1062-6050-48.2.16 PMID:23675799

McCarty, C. W., Hankemeier, D. A., Walter, J. M., Newton, E. J., & Van Lunen, B. L. (2013). Use of evidence-based practice among athletic training educators, clinicians, and students, part 2: Attitudes, beliefs, accessibility, and barriers. *Journal of Athletic Training, 48*(3), 405–415. doi:10.4085/1062-6050-48.2.19 PMID:23675800

Welch, C. E., Hankemeier, D. A., Wyant, A. L., Hays, D. G., Pitney, W. A., & Van Lunen, B. L. (2014a). Future directions of evidence-based practice in athletic training: Perceived strategies to enhance the use of evidence-based practice. *Journal of Athletic Training, 49*(2), 234–244. doi:10.4085/1062-6050-49.2.15 PMID:24568230

Welch, C. E., Van Lunen, B. L., & Hankemeier, D. A. (2014b). An evidence-based practice educational intervention for athletic trainers: A randomized control trial. *Journal of Athletic Training, 49*(2), 210–219. doi:10.4085/1062-6050-49.2.13 PMID:24568228

Welch, C. E., Van Lunen, B. L., Hankemeier, D. A., Wyant, A. L., Mutchler, J. M., Pitney, W. A., & Hays, D. G. (2014c). Perceived outcomes of web-based modules designed to enhance athletic trainers knowledge of evidence-based practice. *Journal of Athletic Training, 49*(2), 220–233. doi:10.4085/1062-6050-49.2.14 PMID:24576306

KEY TERMS AND DEFINITIONS

Allied Health: The segment of the workforce that delivers services involving the identification, evaluation and prevention of diseases and disorders; dietary and nutrition services; and rehabilitation and health systems management.

Athletic Trainers (ATs): Health care professionals who collaborate with physicians to provide prevention, emergency care, clinical diagnosis, therapeutic intervention and rehabilitation of injuries and medical conditions (Profile of athletic trainers, 2014, para. 1).

Continuing Education: Education provided for adults after they have left the formal education system, consisting typically of short or part-time courses.

Evidence-Based Practice: Evidence based medicine is the conscientious, explicit, and ju-

icious use of current best evidence in making decisions about the care of individual patients (Sackett et al., 1996).

Knowledge Translation: A dynamic and iterative process that includes synthesis, dissemination, exchange and ethically-sound application of knowledge to improve the health of patients, provide more effective health services and products, and strengthen the health care system (Graham et al., 2006).

National Athletic Trainers' Association (NATA): A professional membership association serving certified athletic trainers and others who support the athletic training profession in the United States.

Transitions: Change processes that involve a beginning, middle, and an end.

M

Integrating Web-Based Technologies into the Education and Training of Health Professionals

Michelle Lee D'Abundo
Seton Hall University USA

Cara Sidman
Arizona State University, USA

INTRODUCTION

With rising health care costs and the need to treat more and more people with chronic health conditions, many people wonder how healthcare can be advanced in America. Along with healthcare reform, education and training will play a significant role in meeting increased demands for health professionals. In this chapter, exploration of how online learning and instruction can be used to meet the educational needs for the healthcare industry is detailed.

Distance education, e-learning, computer mediated, web-based, and online instruction are terms used to describe education delivered through computer-based technologies. Whatever term is chosen, such computer-based technologies are essential components of the preparation and continuing education of health professionals. For many health professions, formal education, as well as continuing education, includes some type of web-based learning, with some programs completely online.

The American Journal of Health Promotion (2009) defined health promotion as:

Health Promotion is the art and science of helping people discover the synergies between their core passions and optimal health, enhancing their motivation to strive for optimal health, and supporting them in changing their lifestyle to move toward a state of optimal health. Optimal health is

a dynamic balance of physical, emotional, social, spiritual, and intellectual health. Lifestyle change can be facilitated through a combination of learning experiences that enhance awareness, increase motivation, and build skills and, most important, through the creation of opportunities that open access to environments that make positive health practices the easiest choice.

In this chapter, health promotion refers to all education and programming relating to health that is intended to promote positive health outcomes. Many health promotion interventions and programs are conducted online. In fact, the term e-health has been created to describe web-based health promotion initiatives. Health care professionals and health educators are using online instruction to teach patients. Higher education courses that promote lifelong wellness, including physical activity-based courses, are also using online learning formats.

The objectives of this chapter are: 1) to understand the benefits, quality indicators and use of online learning and instruction in health; 2) to discuss whether online learning and instruction is appropriate for training health professionals and for the delivery of health promotion; 3) to review an example of how blended and online instruction can be used for health promotion in higher education; 4) to discuss future trends for online learning and instruction in health.

DOI: 10.4018/978-1-5225-2255-3.ch506

BACKGROUND

In 2013, the Bureau of Labor Statistics reported there were more than 15.8 million jobs in the combined healthcare industries (Torpey, 2014). It is projected that over 4 million jobs will be added between 2012 and 2022, which is more than any other industry (Torpey, 2014). As the largest industry, the demands for preparation and continuing education of health-related professions comprise a significant percentage of adult education conducted in the United States.

Health care professionals are in demand throughout America, but are particularly needed in Health Professional Shortage Areas (HPSAs). According to the U.S. Department of Health and Human Services (2016), 8,200 practitioners are required to meet the need for primary care providers in HPSAs. In order to meet the demands for qualified health care professionals in America, institutions of higher education have sought cost efficient and effective methods of preparing and training health care providers. One way to address the high demand for education, especially for areas that do not have institutions of higher learning, is to begin to offer and expand distance education opportunities. The current use and continued expansion of online learning and instruction in higher education has caused health-related disciplines (and most other disciplines) to ask some difficult questions, which will be explored in this chapter.

SIGNIFICANT QUESTIONS AND ISSUES IN ONLINE LEARNING AND INSTRUCTION IN HEALTH

The health and wellness of individuals, communities, and society are largely dependent on health care providers and health promotion. Consequently, much discussion and debate has occurred as the delivery of adult education and training has evolved to include web-based learning and instruction in the field of health. Some of the significant questions and issues in the discussion about online learning and instruction include the

benefits, quality indicators, and the current use in the field of health.

One of most often asked questions is, “what are the benefits of online learning and instruction?” For potential students who are employed full-time, on active-duty in the military, have family obligations, or are geographically remote, online programs may be the only option for completing their education. According to Chaney et al. (2009), distance education programs benefit students and the university at large. For students, graduation time can be decreased through easier access to courses. Chaney believes distance education diversifies student populations as courses are accessible from any part of the country or world. For the university, strains on the built environment are resolved as no space is needed on campus. Costs are also reduced as all types of printing are eliminated when course materials and assessments are conducted online.

Online learning and instruction can also benefit the environment. Many campuses have struggled to resolve parking and air quality issues. Distance education reduces travel to and from campus, which significantly reduces student and faculty costs for travel, and the inevitable environmental effects associated with auto admissions.

Another question asked is, “what are the quality indicators of online learning and instruction for health-related professional education?” Like all disciplines, health professionals, faculty, and students have their opinions about what contributes to effective web-based learning and instruction. Reeves and Reeves (2008) provided 10 dimensions that should be considered during the instructional design, implementation and evaluation of online courses in health and social work. The model they describe for health and social work education includes 10 dimensions of interactive teaching and learning:

1. Pedagogical philosophy;
2. Learning theory;
3. Goal orientation;
4. Task orientation;

5. Source of motivation;
6. Teacher role;
7. Metacognitive support;
8. Collaborative learning support;
9. Cultural sensitivity; and
10. Structural flexibility.

While the description of these dimensions is beyond the scope of this chapter, overall they emphasize important pedagogical practices translated from the face-to-face to the online environment.

Chaney et al. (2007) described some similar quality indicators of distance education, with some additional ones specifically related to technology. These researchers agree with the importance of interaction, especially between the students and teacher. Other best practices include incorporating the use of active learning techniques, providing prompt feedback, and respecting diverse ways of learning. It is essential for any distance education program to correlate to the mission of the institution, implements the use of appropriate, and reliable multimedia tools, all while providing sufficient student and faculty support and resources. Instructors designing these courses need to provide a clear course structure and means for program evaluation and assessment so improvements can be made.

How is online learning and instruction used in the training of health professionals? Allied health care is using three models of e-learning: blended learning, online learning, and continuing education (Stewart & Wright, 2004). Blended learning includes both face to face and online learning and instruction. Online learning is learning and instruction that is completely web-based. Continuing education for health professionals is offered in both blended and online formats. In fact, most health-related disciplines are using a combination of online and face-to-face learning in their programs.

Perhaps the most salient question to ask is, “Is web-based instruction an effective mode of educating health professionals?” While this question is still in need of additional research evidence to

provide data-based answers, and many factors play a role (i.e., faculty training and course design), it certainly can be. Most health professionals rely heavily on technology in their practice and thus, it seems appropriate for technology to play a role in the education of future health professionals. In fact, Hollis and Madill stated (2006),

Evidence suggests that blending combinations of technologies with computer mediated learning enhances interaction and could address the higher order learning needs of professional programs such as occupational therapy. (p.61)

Through a review of research focused on computer-enhanced classroom-based teaching, online courses, online degree programs, and on-line continuing professional education, Hollis and Madill (2006) recommended that both online and face-to-face teaching and learning opportunities accommodate all types of students and faculty. Hollis and Madill (2006) emphasize the use of multiple technologies to enhance interaction between students, and between students and instructors. However, to optimize the effectiveness of online learning, they believe appropriate preparation and support is required for both students and faculty.

How is blended and online learning and education being used in health professions? Konin (2004) discussed how to enhance clinical education, like Athletic Training, through the use of web-based programs. Like most health professionals, education and training includes some type of fieldwork, or internship placement. Information and assessments for clinical instructors can be delivered electronically through an online course management system such as Blackboard. Face-to-face class formats can be enriched through the availability of a multitude of supplemental online materials. McLeod and Barbara (2005) reported, “The use of online chat rooms was seen to be beneficial to encouraging peer support while on fieldwork for allied health students” (p.276).

Web-based education has also been used to teach medical students. A web-based self-study

was developed to teach medical students how to help smokers quit and how to prevent others from starting to smoke. The curriculum helped medical students develop effective counseling skills in the prevention and treatment of smoking (Pederson, Blumenthal, Dever & McGrady, 2006). In another study, third- and fourth-year medical students completing a web-based women's health curriculum valued learning women's health, preferred this self-directed learning over lecture, scored highly on knowledge tests, and were involved in more high-level discussions of women's health with faculty (Zebrack et al., 2005).

Online learning and instruction is also essential to the continuing education of health-related professionals. Williams (2006) conducted a meta-analysis of 25 studies assessing achievement in distance education in allied health and found that distance education is as effective as face-to-face instruction. Student achievement outcomes were positively correlated with work experience and professional knowledge. Outcomes were also better for distance education experiences that included multiple learning and instructional strategies including, but not limited to, interaction and introspection.

WEB-BASED HEALTH PROMOTION

Researchers have examined online health promotion and found them to be effective methods of information delivery (Parlov, Cowdery & Hoerauf, 2004; Shegog et al., 2005). Wantland et al. (2006) conducted a meta-analysis of effects of patient knowledge and behavioral change outcomes in web-based and non-web-based interventions. In 16 of 17 web-based interventions reviewed, participants showed improved knowledge and behavioral outcomes. D'Abundo, Fiala, and Mariano (2010) found that an online learning module was an effective way to deliver information about smoking to college students in a general education fitness and wellness course. Some advantages of computer-based health promotion activities in-

clude individual tailoring, consistency of content, the incorporation of media, and availability of the program when a person is ready to make a change.

In order to meet the health promotion goals of many universities and colleges, university basic study curriculums include required hours in courses that promote life-long wellness. Many such courses are delivered in face-to-face formats and require participation in some type of physical activity. Like issues involved in the preparation of future health professionals, there are obstacles encountered in accommodating the diverse needs of learners in the university setting. The option of web-based delivery of such courses may help meet the demands of an ever-expanding student population.

RECOMMENDATIONS

For the practitioners and faculty that design, implement and evaluate training of future health professionals and health promotion interventions and programs, it may be difficult to envision how to develop blended and online courses/programs from existing traditional curriculums. There may also be concerns about how to ensure that objectives are being met in blended and online formats. The following are suggestions about developing blended and online instruction focused on health promotion in higher education. The information may be applied to the training of future health professionals as well health promotion activities outside of education.

1. **Start with Blended Courses Before Moving to Completely Online Formats:** For some health professionals and health faculty, it is difficult to decide how to adapt existing educational materials, activities and programs to web-based delivery. For both blended classrooms and completely online learning environments, there are a numerous online resources that are available for educational purposes. For example, PBS provides

much of their health-related programming on their website and includes educational materials for discussing the content. Bennett (2005) described the potential for using The Multimedia Educational Resource for Learning and Online Teaching (MERLOT) as an online resource linked to thousands of learning objects. Resources available include online learning materials, peer reviews, learning assignments, and user comments, all organized by discipline. Instructional technology consultants are also available to assist with the development of online learning materials and curriculums.

2. **When Moving Completely Online Go Beyond Content and Strive to Increase Interaction!** One criticism of online learning and instruction is that it limits both student-to-student and instructor-to-student interaction. Those of us that teach both online and face-to-face formats realize that interaction largely depends on the facilitation of class and the members of the class itself. In order to continue to raise awareness about web-based education, it is important to educate practitioners, students, and the public about the potential of online education.

One way to alleviate some of the concern about the lack of instructor and student interaction in online classes is to design strategies to promote interaction and build community among student and participants. Stow (2005) stated that in online courses, activities should include multiple types of interaction. The author recommends that the instructor communicate with the students in a timely manner. Stow suggested that methods like chat rooms, discussion, and feedback allow the students to become actively engaged in order to decrease the transactional distance.

One way to increase student interaction and build community in blended and online course formats is the establishment of "Course Buddies." The objective of using "Course Buddies" is to build community in blended and online courses

and in health promotion programs delivered in a web-based format. To begin the "Course Buddy" selection, instructors can ask students to post introductions that include three things that most people do not know about them and include a picture. The next step is to have students review each others' introductions and find three people that they have things in common with or they want to get to know better. Then instruct students to e-mail potential "Course Buddies" to exchange contact information or just communicate by using e-mail. The process of finding "Course Buddies" in itself helps build community and enables students to learn more about each other. Once "Course Buddies" are established, students can discuss activities, study for exams or work on assignments.

FUTURE TRENDS

The future for online learning and instruction in health seems promising as we strive to meet the demands of an ever-changing health care system with accessible training and continuing education opportunities. However, to further facilitate successful use of online delivery methods to train health professionals and to promote health, data-based research describing and analyzing its effectiveness is essential. Advocacy on the part of professionals in health-related disciplines is also critical for web-based education methods to be embraced. Part of such advocacy includes increasing awareness of the online opportunities that already exist and exploring the potential of alternate computer based methods of delivery. Those interested in advocating for online education may want to join The United States Distance Learning Association.

In order for online learning to continue to grow and become embraced, myths associated with this mode of instruction need to be dispelled. While it is undoubtedly serving a need, and the technologies available to interact with learners from a distance continue to evolve, professionals and educators with substantial reservations regarding its role

in teaching and learning remain. It is recognized that online learning is not for everyone, and it may not be the best method for teaching certain types of curriculum, it is certainly effective if taught based on sound instructional design principles and well trained instructors. A successful future that optimizes the use of online teaching and learning depends on the collaboration and communication among professionals regarding the most effective ways to meet learning objectives. Once those are established, technology use can be designed around those objectives, and not vice versa.

Another trend in online education is peer review. As online learning and instruction continues to expand in health-related disciplines, much discussion focuses on how to ensure the quality of web-based programming. Peer review has been suggested as a way to review and improve online course delivery as well as health promotion efforts. Quality Matters © (QM) (Quality Matters, 2016) is a peer-review process designed to certify the quality of online courses. The Quality Matters process uses a rubric to review eight broad standards in blended and online courses that include: learning objectives, assessment and measurement, resources and materials, learner engagement, course technology, learner support and accessibility. Many programs throughout the country have created their own versions of peer review or “online observations” to assess the quality of web-based curriculums.

CONCLUSION

As described in this chapter, online learning and instruction have become important components of formal and continuing education for health professionals and in health promotion programs. Curriculums that have previously been thought to be inappropriate for web-based delivery are now being offered online. Innovations in online learning and instruction have improved delivery of web-based education and have provided educational opportunities for many that were “unreach-

able” by traditional classrooms. As web-based educational opportunities continue to expand, it is our hope that more health professionals and the general public will learn about the efficacy and possibilities of online learning and instruction.

REFERENCES

- American Journal of Health Promotion. (2006). *Definition of health promotion 2.0: Embracing passion, enhancing motivation, recognizing dynamic balance, and creating opportunities*. Retrieved on October 26, 2009 from <http://www.healthpromotionjournal.com/publications/journal/en2009-09.htm>
- Bennett, J. (2005). Using MERLOT in the classroom. *Athletic Therapy Today*, 10(3), 37–39. doi:10.1123/att.10.3.37
- Chaney, B., Eddy, J., Dorman, S., Glessner, L., Green, B., & Lara-Alecio, R. (2009). A primer on quality indicators of distance education. *Health Promotion Practice*, 10(2), 222–231. doi:10.1177/1524839906298498 PMID:17925594
- D’Abundo, M. L., Fiala, K. A., & Marinaro, L. (2010). An online learning module focused on smoking education and prevention for college students: Implications for college health instructors and allied health professionals. *Journal of Allied Health*, 39(1), 43–48. PMID:20217006
- Hollis, V., & Madill, H. (2006). Online learning: The potential for occupational therapy education. *Occupational Therapy International*, 13(2), 61–78. doi:10.1002/oti.209 PMID:16862957
- Konin, J. (2004). Using a Web-based program to enhance clinical education. *Athletic Therapy Today*, 9(1), 20–23. doi:10.1123/att.9.1.20

McLeod, S., & Barbara, A. (2005). Online technology in rural health: Supporting students to overcome the tyranny of distance. *The Australian Journal of Rural Health, 13*(5), 276–28. doi:10.1111/j.1440-1584.2005.00717.x PMID:16171501

Pederson, L., Blumenthal, D., Dever, A., & McGrady, G. (2006). A web-based smoking cessation and prevention curriculum for medical students: Why, how, what, and what next. *Drug and Alcohol Review, 25*(1), 39–47. doi:10.1080/09595230500459503 PMID:16492576

Quality Matters. (2016). *Quality Matters in Online Learning*. Retrieved on from <http://www.qualitymatters.org/>

Reeves, P., & Reeves, T. (2008). Design considerations for online learning in health and social work education. *Learning in Health and Social Care, 7*(1), 46–58. doi:10.1111/j.1473-6861.2008.00170.x

Shegog, R., McAlister, A., Hu, S., Ford, K., Meshack, A., & Peters, R. (2005). Use of interactive health communication to affect smoking intentions in middle school students: A pilot test of the Headbutt risk assessment program. *American Journal of Health Promotion, 19*(5), 334–338. doi:10.4278/0890-1171-19.5.334 PMID:15895535

Stewart, J., & Wright, V. (2004). E-learning in allied health care. *Athletic Therapy Today, 9*(1), 7–9. doi:10.1123/att.9.1.7

Stow, R. (2005). Minimizing the distance in distance learning. *Athletic Therapy Today, 10*(2), 57–59. doi:10.1123/att.10.2.57

Torpey, E. (2014). Healthcare: Millions of jobs now and in the future. *Occupational Outlook Quarterly*. Retrieved from <http://www.bls.gov/careeroutlook/2014/spring/art03.pdf>

U.S. Department of Health and Human Services Health Resources and Service Administration. (2016). *Shortage Designation: Health Professional Shortage Areas & Medically Underserved Areas/Populations*. Retrieved from <http://www.hrsa.gov/shortage/>

Wantland, D., Portillo, C., Holzemer, W., Slaughter, R., & McGhee, E. (2004). The Effectiveness of web-based vs. non-web-based interventions: A meta-analysis of behavioral change outcomes. *Journal of Medical Internet Research, 6*(4), e40. doi:10.2196/jmir.6.4.e40 PMID:15631964

Williams, S. (2006). The effectiveness of distance education in allied health science programs: A meta-analysis of outcomes. *American Journal of Distance Education, 20*(3), 127–141. doi:10.1207/s15389286ajde2003_2

Zebrack, J., Mitchell, J., Davids, S., & Simpson, D. (2005). Web-based curriculum. *JGIM: Journal of General Internal Medicine, 20*(1), 68–74. doi:10.1111/j.1525-1497.2005.40062.x PMID:15693931

ADDITIONAL READING

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191–215. doi:10.1037/0033-295X.84.2.191 PMID:847061

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: W.H. Freeman & Company.

Bennett, J. (2005). Using MERLOT in the classroom. *Athletic Therapy Today, 10*(3), 37–39. doi:10.1123/att.10.3.37

- Campbell, M., Gibson, A., Richards, D., & Callery, P. (2008). Online vs. face-to-face discussion in a web-based research methods course for postgraduate nursing students: A quasi-experimental study. *International Journal of Nursing Studies*, 45(5), 750–759. doi:10.1016/j.ijnurstu.2006.12.011 PMID:17306272
- Chaney, B., Eddy, J., Dorman, S., Glessner, L., Green, B., & Lara-Alecio, R. (2009). A primer on quality indicators of distance education. *Health Promotion Practice*, 10(2), 222–231. http://search.ebscohost.com doi:10.1177/1524839906298498 PMID:17925594
- Deci, E. L., Ryan, R., Chaney, B., Eddy, J., Dorman, S., Glessner, L., & Lara-Alecio, R. et al. (2009). A primer on quality indicators of distance education. *Health Promotion Practice*, 10(2), 222–231. doi:10.1177/1524839906298498 PMID:17925594
- Ellery, J., McDermott, R., & Ellery, P. (2007). Computers as a formal continuing education tool: Moving beyond intention. *American Journal of Health Behavior*, 31(3), 312–322. doi:10.5993/AJHB.31.3.8 PMID:17402870
- Haskell, W. L., Lee, I.-M., Pate, R. R., Powell, K. E., Blair, S. N., Franklin, B. A., & Bauman, A. et al. (2007). Physical activity and public health: Updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. *Medicine and Science in Sports and Exercise*, 39(8), 1423–1434. doi:10.1249/mss.0b013e3180616b27 PMID:17762377
- Hollis, V., & Madill, H. (2006). Online learning: The potential for occupational therapy education. *Occupational Therapy International*, 13(2), 61–78. doi:10.1002/oti.209 PMID:16862957
- Hopper, K., & Johns, C. (2007). Educational technology integration and distance learning in respiratory care: Practices and attitudes. *Respiratory Care*, 52(11), 1510–1524. PMID:17971255
- Konin, J. (2004). Using a Web-based program to enhance clinical education. *Athletic Therapy Today*, 9(1), 20–23. doi:10.1123/att.9.1.20
- McLeod, S., & Barbara, A. (2005). Online technology in rural health: Supporting students to overcome the tyranny of distance. *The Australian Journal of Rural Health*, 13(5), 276–28. doi:10.1111/j.1440-1584.2005.00717.x PMID:16171501
- Parlove, A., Cowdery, J., & Hoerauf, S. (2004). Acceptability and appeal of a Web-based smoking prevention intervention for adolescents. *International Electronic Journal of Health Education*, 71–78.
- Pederson, L., Blumenthal, D., Dever, A., & McGrady, G. (2006). A web-based smoking cessation and prevention curriculum for medical students: Why, how, what, and what next. *Drug and Alcohol Review*, 25(1), 39–47. doi:10.1080/09595230500459503 PMID:16492576
- Reeves, P., & Reeves, T. (2008). Design considerations for online learning in health and social work education. *Learning in Health and Social Care*, 7(1), 46–58. doi:10.1111/j.1473-6861.2008.00170.x
- Schutt, M., & Hightower, B. (2009). Enhancing RN-to-BSN students information literacy skills through the use of instructional technology. *The Journal of Nursing Education*, 48(2), 101–105. doi:10.3928/01484834-20090201-11 PMID:19260403
- Shegog, R., McAlister, A., Hu, S., Ford, K., Meshack, A., & Peters, R. (2005). Use of interactive health communication to affect smoking intentions in middle school students: A pilot test of the Headbutt risk assessment program. *American Journal of Health Promotion*, 19(5), 334–338. doi:10.4278/0890-1171-19.5.334 PMID:15895535

Stewart, J., & Wright, V. (2004). E-learning in allied health care. *Athletic Therapy Today*, 9(1), 7–9. doi:10.1123/att.9.1.7

Stow, R. (2005). Minimizing the distance in distance learning. *Athletic Therapy Today*, 10(2), 57–59. doi:10.1123/att.10.2.57

Williams, S. (2006). The effectiveness of distance education in allied health science programs: A meta-analysis of outcomes. *American Journal of Distance Education*, 20(3), 127–141. doi:10.1207/s15389286ajde2003_2

Zebrack, J., Mitchell, J., Davids, S., & Simpson, D. (2005). Web-based curriculum. *JGIM: Journal of General Internal Medicine*, 20(1), 68–74. doi:10.1111/j.1525-1497.2005.40062.x PMID:15693931

KEY TERMS AND DEFINITIONS

Blended Learning and Instruction: Includes both face- to-face and online learning and instruction.

Course Buddies: An ice-breaker or introduction strategy used to build community in blended and online courses and in health promotion programs.

E-Health: Health promotion initiatives delivered online.

Face-To-Face Learning and Instruction: Course activities and instruction occur in the traditional classroom without the use of online learning and instruction.

Health Promotion: Refers to all education and programming relating to health that is intended to promote positive health outcomes.

Health-Related Professions: Professions that delivery medical care and health promotion including but not limited to nursing, occupational therapy, athletic training, respiratory therapy, physicians, physician assistants, social workers, and health educators.

Online Learning and Instruction: Learning and instruction that is completely web-based.

Interactivity in Distance Education and Computer–Aided Learning, With Medical Education Examples

M**D. John Doyle***Cleveland Clinic, USA***Patrick J. Fahy***Athabasca University, Canada*

INTRODUCTION

Webster's Revised Unabridged Dictionary defines interaction as: 1. intermediate action, 2. mutual or reciprocal action or influence; as, the interaction of the heart and lungs on each other. The *Oxford English Dictionary* defines interactivity as reciprocally active, allowing a two-way flow of information between source and user, responding to the user's input. This is the view advocated by many observers (Wagner, 1994; Wagner, 1997; Anderson, 2003). Notes Anderson: "Simply stated, interactions are reciprocal events that require at least two objects and two actions. Interactions occur when these objects and events mutually influence one another" (p. 8). The concept of "interaction" in education and training varies with subject domain (e.g., philosophy versus physics), with format (e.g., conventional versus distance education) and with other factors. Although the term is frequently overused and poorly understood, the notion of interactivity remains important and useful, and is especially topical given the availability of technologies designed to support advanced interactivity.

BACKGROUND

In his book *The Art of Interactive Design*, Crawford (2002) discusses the notion of interactivity in terms of a conversation: "a cyclic process in which two

actors alternately listen, think, and speak," and notes that the "quality of the interaction depends on the quality of each of the subtasks (listening, thinking, and speaking)." This model of interactivity avoids some of the pitfalls associated with some definitions offered in the past.

As noted above, educators place considerable importance on interactivity in learning (Anderson & Garrison, 1995; Fahy, 2005; Garrison & Cleveland-Innes, 2005; Lapadat 2007, Faghihi et al., 2016). Well-designed interactivity can help capture the learner's interest, has the potential to speed learning, and allows for continuous assessment of the degree to which the material is mastered. Technology (at least theoretically) can allow for high-quality interactivity by providing for frequent and relevant user feedback, by recognizing when students misunderstand a concept, and by providing aids such as animations or graphs that vary with user input. However, badly designed interactivity can also impede student progress, suppressing, or making more difficult, program flexibility and learner independence (Burbules & Callister, 2000; Faghihi et al., 2016) (*vide infra*).

The importance of interactivity in learning is illustrated by the fact that a number of journals address this topic. For instance, the *Journal of Interactive Learning Research* (JILR) publishes manuscripts dealing with "the underlying theory, design, implementation, effectiveness, and impact on education and training of the following interactive learning environments," including such varied

DOI: 10.4018/978-1-5225-2255-3.ch507

topics as “authoring systems, cognitive tools for learning computer-assisted language learning, computer-based assessment systems, computer-based training, computer-mediated communications, computer-supported collaborative learning, distributed learning environments, electronic performance support systems, interactive learning environments, interactive multimedia systems, interactive simulations and games, intelligent agents on the Internet, intelligent tutoring systems, microworlds, virtual reality based learning systems,” and anything else related to these.

In this review, we will discuss interaction in distance education and training, with an emphasis on medical education. Since educational technologies are often used to support interactivity, an important theme will be how technologies can support or impede interaction.

Interaction in Publication

Few would argue that published documents form a very important component in education. Stevan Harnard (1992) has emphasized that electronic publication provides a dimension of interactivity in document publication that is radically new. While online journals are still in evolution, and the form they will ultimately take – combining reader access with content credibility and publisher concerns – is still being determined, there can be little doubt that online journals are here to stay.

Interaction in Distance Education

One simple definition of “distance education” is that it is the delivery of instruction that does not require the student to be present in the same location as the instructor. What then is “interaction” in distance education? We suggest that it is a generic or broad term, covering all manner of notions of amplification, appraisal, clarification, commentary, communication, exploration, feedback, involvement and participation in the context of an educational exercise.

In more concrete terms, interactivity may be based on, or supported by, communication technologies such as the telephone, e-mail, instant messaging, or computer conferencing, as well technologies such as computer-based simulation. As such, it may be hard to define the concept of interaction precisely. Regardless, the notion of interaction should become clearer as the reader reads on.

Finally, it should be noted that although many individuals (e.g., Gilbert & Moore, 1998) use the terms “interaction” and “interactivity” interchangeably, some authors draw a distinction between the two. For example, Wagner (1994, 1997) defines “interaction” in terms of an interplay of behaviors in which individuals and groups influence each other, while “interactivity” is defined in terms of real-time characteristics of technological systems.

Importance of Interaction in Distance Education

Moore & Kearsley (1996) describe three types of interaction within distance education. McIsaac & Gunawardena (1996) add a fourth type, which takes into account the interaction that occurs when a learner must use educational technologies as part of the learning process. In these descriptions, students in a distance education “learning environment” are thus said to interact variously with the instructor, the content, the technology, and with other students.

In addition to these four forms of interaction, further distinctions must be made between instantaneous (“real time”, “synchronous”) and delayed (“asynchronous”) interaction, as well as between “same-place” and “different-place” interaction. Generally speaking, delayed and different-place interaction offer the student more flexibility and opportunities for thought and reflection, while immediate and same-place interaction, some would say, may allow for a greater sense of spontaneity, impulsiveness and even exhilaration.

With conventional classroom-based teaching, the focus is predominantly on student-teacher interaction, with student-content interaction also taking place, especially when out-of-classroom reading takes place. In the distance education setting, with the heavy use of self-study materials, the focus is usually on student-content interaction, with other forms of interaction occurring to varying degrees. Interaction with technology is also frequently important. Finally, recent research in collaborative and cooperative learning environments, especially those involving computer networks, has led to new insights into the nature of student-student (or more generally, peer-peer) interaction.

Note also that the above four-pronged approach to interactivity in education is not the only available model. In some educational domains alternate models may prove to be useful. For instance, we

have developed a rather different six-category approach to interaction, outlined in Table 1, which we believe works well to help explain the notion of interaction in the realm of distance education, using examples from medical education.

Learners and instructors will have preferences regarding interactivity, and some forms of education will rely on some forms of interactivity more than others. Regardless, the wise use of interactivity will help “engage” students and help ensure an enduring learning experience - learning enhanced by the potential individualization of the experience (Simonson et al., 2009).

Interactivity as a Driver of Change in Education Delivery

Interactivity requires a fundamental change in the manner in which education is delivered. Tapscott

Table 1. Types of interactivity in teaching and education, with examples focusing on medical education

[1] Person-to-person (student / teacher) interaction with both persons in direct verbal communication (“same place interaction”)
• Formal classroom or auditorium teaching (such as a question and answer period at the end of a lecture)
• Informal hallway or bedside discussions
[2] Person-to-person interaction (student / teacher) using remote communication technology (“different place interaction”)
• Telephone and audio conferencing
• Writing (paper, e-mail)
• Video conferencing
• Telemedicine (e.g., clinical discussion centering around a radiographic image)
[3] Paper-based interactivity [can also be “morphed” to a computer model]
• Interactive paper scenarios (If X is true then go to page Y otherwise go to page Z)
• Example 1: Patient with chest pain presents to the ER where Y = “admit patient to hospital”
• Example 2: Patient with abdominal pain presents to the ER where Y = “operate on patient [laparotomy]”
[4] Paper-based interactivity based on “developer pen” technology [can also be morphed to a computer model]
• Developer pen technology allows the student to see the result of a study or laboratory test once it has been “ordered”; based on the result of the selected investigation the student selects other clinical options (e.g., asks more questions, attempts to elicit a particular clinical finding, orders another test)
[5] Computer-based interactivity not using the World Wide Web. This usually involves using programming languages such as C++ or various older authoring platforms.
[6] Web-based interactivity. This usually involves using programming languages such as JavaScript or authoring platforms with Web support.

Note: further distinctions can to be made between immediate (“synchronous”, “real time”) and delayed (“asynchronous”) interaction.

(1998) has identified seven ways in which this change occurs:

- From linear to hypermedia learning.
- From instruction to construction and discovery.
- From teacher-centered to learner-centered education.
- From absorbing material to learning how to navigate and how to learn.
- From schooling to lifelong learning.
- From learning as torture to learning as fun.
- From the teacher as transmitter to the teacher as facilitator.

Interactivity and Adult Learning Theory

Why is interactivity important in education? One reason is that interactivity is particularly compatible with established psychological models of learning and the central tenets of adult learning theory (Simpson & Galbo, 1986). In particular, constructivist principles help account for the importance of interactivity in distance education. Constructivism is the view that knowledge is “constructed” by the learner by testing ideas, concepts and approaches based on existing knowledge and one’s actively acquired experiences, including the views of others, and that knowledge is not merely acquired passively. Constructivist theory holds that students learn best when they actively and collaboratively participate in problem-solving and critical thinking while involved in an appropriately formulated learning activity, and that this learning involves the integration of newly acquired knowledge with pre-existing intellectual constructs.

Experts in adult learning also argue that traditional didactic teaching merely encourages passive learning, instead of the development of higher order cognitive skills needed for true education. They point out that active involvement is essential for effective learning, and adults learn best, they argue, when one can draw on previous experi-

ence, using techniques such as group discussion, simulation exercises, and problem solving. That is, going beyond mere looking and listening motivates people to learn on their own, gives students the motivation to try out new ideas, and encourages them to critically examine issues that were once simply accepted passively.

Interaction Between Student and the Instructor

Interactivity in the traditional classroom is primarily between the student and the instructor, and technology usually plays a secondary role; students can ask questions and get immediate replies. Outside the classroom students often interact among themselves by verbal means (“Did you understand how he solved problem 6?”) and may later interact by telephone, instant messaging and e-mail, three means of interactivity also available to distance education students.

Students in some distance education programs view videos of the regular on-campus classes, which include interactions between the on-campus students and the instructor (usually in the form of questions), but obviously such viewers cannot have any synchronous interaction with the instructor at the time of the lesson, although they are still able to exchange words with both the instructor (and, possibly, an academic advisor) via telephone, fax, instant messaging and e-mail.

In many cases distance education learners have no direct interaction with the course instructor, and instead deal only with an on-site coordinator or tutor. Still, provided these individuals are knowledgeable in the materials being taught and have well-developed communication skills, this should not be a problem. Such individuals are especially helpful in settings where there is a high ratio of students to instructor and/or when direct access to the instructor is not practical.

In contrast to the setting where students have no direct interaction with the course instructor, online services also exist which offer instructor-

led courses where students can log into a message board for discussions with both the instructor and fellow students.

Interaction Between Student and Content

Many educational theorists would say that interaction between the student and the content is the heart of education, whether in the classroom or in the distance education setting (or, indeed, in the setting of self-education). In order to learn effectively, the content must be presented in a way that will motivate the student and inspire thought. Most often, the material is presented as text with some graphics, and a good writing style with good organization of the material are the key elements to producing a useful document. The technology used for the document (printed paper, PDF, HTML etc.) is less important than the quality of the writing, although poor formatting makes even good writing hard to read.

One interesting notion in this context is Bloom's "alterable variables," modifiable factors that Bloom found to enhance group learning to the level of tutoring (Bloom 1984). Examples include the use of reinforcement, supplying corrective feedback, and offering cues and explanations. It would appear that use of Bloom's alterable variables would be excellent educational elements that might be introduced into the process of developing educational resources.

Interaction Between and Among Students

Students have always been able to communicate with one another in traditional classrooms, sometimes by passing notes, but usually by meeting informally before and after class. Such social interactions are important to making new friends and getting exposed to new ideas. These interactions also help in carrying out collaborative projects.

It is interesting to note that interaction between students is a somewhat new phenomenon

in distance education, having started in a big way only since the development of the Web made computer-conferencing practical. Until relatively recently, distance education students had limited interaction with other students, their interactions being primarily with the instructor, the content, and the technology. However, the advent of Web-based computer conferencing and e-mail services has changed this. As a result of this and the development of groupware systems such as Lotus Notes, distance education is moving from highly individualized instruction (as in the correspondence education approach popular decades ago) to formats that encourage students to explore topics as a group and to undertake collaborative learning. This approach is also asynchronous, which gives learners the freedom to control their learning activities in terms of time, location and choice of collaborating partners. Other technologies, such as instant messaging or two-way video conferencing, also allow for communication among students, but are primarily synchronous in nature. By contrast, computer conferencing provides for asynchronous interaction, which fits much better into the hectic lives of many adult learners. Still, the nature of student-student interaction in distance education is fundamentally different from conventional educational settings.

Some students feel that the distance education format is often weak in the area of establishing "bonding" between classmates, who are not able to get to know one another as effortlessly as in a classroom. Students usually have no idea what their classmates look like, or what their voices sound like, thereby delaying the development of any "mental image" about the people in the class until much later, once many postings have been made and students get to know one another by means of written communications. This notion is known as "social presence" (Nippard & Murphy, 2007) which may be defined as "the degree to which participants are able to project themselves affectively within [a] medium" (Garrison 1997). Social presence affects student perceived learning (Richardson & Swan, 2003), may increase student

satisfaction in online learning (Newberry, 2001), and may lead to improved emotional satisfaction (Rourke et al., 1999).

Interaction Between Student and Technology

Until recent years, educational technology was not especially capable of maintaining a high degree of interactivity with a learner, in contrast, say, to the very high level of possible interaction between a teacher and a student discussing a philosophical matter, where a series of verbal exchanges may occur that involves a cognitive effort much more advanced than mere memorization or rote repetition.

This matter will likely change favorably as computer technology inextricably advances. In particular, advances in Artificial Intelligence, educational authoring systems and scripting languages such as JavaScript will likely have a direct impact on the evolution of computer-based interactivity. Still, computer-based learning is much more complex than working through a textbook, and many students who are uncomfortable with computers see no reason to deviate from a curriculum based on a well-written textbook, particularly if accompanied by study guides and solution manuals.

Many advocates of computer-mediated distance education draw attention to the helpful aspects offered by computer-based methods, and understate the kinds of communicative and technical difficulties learners may experience. When students use complicated equipment in their learning environment, it is important to ensure the technologies used help rather than hinder the learning experience. Students using unfamiliar technologies may experience frustration, anxiety, and confusion related to computer hardware problems, software problems at the level of the operating system or the various software applications, as well as difficulties with Web access. Files may mysteriously “disappear.” Attachments sent by e-mail may be “lost”. Random “discon-

nects” may frustrate the user. Unless students are moderately familiar with computers, substantial effort may be expended on technology-related issues rather than learning the intended materials.

Thus technologies that are awkward, unintuitive or present the learner with a steep learning curve may hinder interaction and even impede learning. Various means by which technology can impede interactivity include: haphazard, unintuitive or disorganized user interfaces; poor screen layout; poor use of fonts and icons; systems that respond too slowly to user inputs; or systems that crash frequently. While a detailed discussion of this matter is beyond the scope of this review, books by design gurus Jakob Nielsen (1994) and Donald Norman (1999), are recommended as a good place to explore the nature of good and bad user-interface design.

Studies show that technical problems can occasionally render the educational undertaking to be a substantially negative experience. For instance, Hara & Kling (1999) carried out a study of a web-based distance education course at an American university, and commented that distance learners may experience problems and frustrations that are not well documented in the literature. They noted that three interrelated sources accounted for most of the problems: “lack of prompt feedback, ambiguous instructions on the web, and technical problems.” They concluded that the frustrations the students experienced “inhibited their educational opportunity”. In a related paper Hara & Kling (2000) voiced similar concerns.

Interactivity in Medical Education

The Royal College of Physicians and Surgeons of Canada (RCPSC) has long taken an active role in the post-certification education of Canadian specialists through their Maintenance of Certification (MOC) program and its forerunners. Although they use the term “Continuing Professional Development” (CPD) in referring to this process, it is, in essence, simply another term for Continuing Medical Education (CME).

The MOC program requires that members complete 400 credit hours of acceptable CPD in a five-year period to maintain specialist certification. Among the most popular means of collecting credit is through “Section 1” educational activities (attending lectures, seminars, scientific meetings, journal clubs, etc.). Other acceptable activities involve “practice audits,” “personal learning projects,” and education via medical simulation.

The RCPSC feels strongly that Section 1 activities are of increased educational value where there is an opportunity for interaction between the “expert” speaker and the participants. This is often in the form of a question and discussion period that serves to further engage the listener, as well as to clarify some of the issues that may remain unanswered at the end of the formal part of the presentation.

The MOC program requires that “at least 25 per cent of the time of a CPD event should be allocated for interactive learning.” Otherwise, the program coordinators believe, the event takes place at the expense of audience involvement, a key ingredient in learning. This view is supported by a landmark meta-analysis of the effectiveness of formal CME by Davis et al. (1999), who showed that traditional didactic methods do not generally lead to a change in clinical practice, or to an improvement in patients’ health outcomes, whereas interactive techniques are more likely to lead to a change.

Simulation in Medical Education

In many respects, simulators offer the ultimate in interactivity and appear to have an excellent future in medical education. Computer-based simulators used in medical education fall into four general categories: (1) Screen-based text simulators, (2) Screen-based graphical simulators, (3) Manikin-based simulators, and (4) Virtual reality trainers.¹

Screen-based text simulators create verbal scenarios in which the users pick one of several responses and, based on the chosen response, a new text scenario is produced. For instance, in

a scenario involving a patient presenting with a severe headache, the user may be offered options such as prescribing an analgesic such as Tylenol or getting a CT-scan of the head. Based on the user’s choice, a new narrative is then generated and more management choices are offered. Being purely text-based, screen-based text simulators are relatively simple to construct, making them popular in early medical simulation efforts. However, since they are lacking in graphical elements, they are rarely used today.

Screen-based graphical simulators such as ‘Gasman’ (Philip, 1986; Philip, 2009) and ‘Body’ (Smith & Davidson, 2009) aim to recreate elements of reality in graphic form on a computer screen, often to elucidate the processes associated with drug administration. Usually, only a mouse interface is involved. While these simulations help one understand the conceptual theory underlying clinical practice, they usually do not confer practical skills. Their strength lies in an ability to help one understand abstract concepts while remaining portable and inexpensive.

Manikin-based simulators come in various levels of complexity, but are expensive. The advanced models include a physical model of the human body and provide continuous signals representing numerous physiological parameters. While some earlier systems required the intervention of a trainer to actively ‘stage manage’ the scenario in response to interventions, others make use of computer models of human physiology and pharmacology to automatically generate appropriate responses in the manikin. In contrast to screen-based simulations, these simulators appear to recreate enough elements of reality to allow the acquisition of skills that are transferable back to the clinical environment.

Virtual reality trainers are just beginning to gain popularity, especially amongst surgical trainees. They offer a transition from the two-dimensional world of the textbook to the three dimensional world of simulated patients. The principal problem so far has been the lack of reality modeling in terms of offering a tactile ‘feel’. However,

“haptic” solutions that provide feel and heft are being actively researched in centers, often using micro-engineering techniques.

Kinds of Interactivity in Medical Education

We have already noted that, as with all areas of education and training, interactivity plays an important role in medical education. While the RCPSC emphasizes means such as question and answer sessions to achieve this goal, many other methods can be imagined. One possible taxonomy of the kind of interaction in medical education is outlined in Table 1.

Social Media in Education

Social media such as Facebook can help serve to connect millions of learners, forming a virtual learning community where information, ideas, resources and suggestions are informally exchanged, offering community members ready access to knowledge, both from experts and other students. Some social media like Yahoo Answers focus on allowing users to submit questions and answer questions asked by others. As a way to encourage participation, members may get the chance to earn “points.” Users can also comment on answers, rate questions and answers, and offer suggestions to improve the service. However, such sites have sometimes been criticized as a problematic venue for serious education because of difficulties in readily deleting incorrect answers, and a point structure that rewards participation over accuracy (Liebenluft, 2007).

Massive Open Online Courses

The Khan Academy (www.khanacademy.org) is a popular non-profit educational website created in 2006 by Salman Khan. It has delivered millions of lessons in topics as varied as physics, chemistry, biology, mathematics, astronomy, economics and

history. The Khan Academy has both enthusiasts and critics. Enthusiasts support the academy for its ingenuity, its popularity, and its eradication of economic barriers to education. Others, however, are concerned that the videos lack interaction and active engagement.

Coursera (coursera.org) is an educational technology company founded by Andrew Ng and Daphne Koller. Rather than the mini-lecture approach taken by the Khan Academy, Coursera offers hundreds of free noncredit university level courses in computer science, engineering, mathematics, business, social sciences, humanities, and other fields. In the future, students may be take Coursera courses for university credit, at costs competitive with brick-and-mortar universities, via exams administered via an online proctoring service that allows students to be watched via webcam. Similar initiatives have also been launched by Education Portal (<http://education-portal.com/>) EDX (edxonline.org), Minerva Project (minervaproject.com) and Udacity (udacity.com). All are examples of Massive Open Online Courses (MOOCs).

Course Management Systems

Course management systems (sometimes also known as learning management systems) are software platforms that provide an instructor and students with a set of resources (course document storage, e-mail, discussion threads, videos, examination capabilities, etc.) that facilitate the development and delivery of online courses. Examples include Blackboard (www.blackboard.com), Moodle (moodle.org), Pearson LearningStudio (www.pearsonlearningsolutions.com/pearson-learning-studio/), OLAT (www.olat.org), and Sakai (www.sakaiproject.org). These systems are primarily intended for asynchronous content delivery; for the delivery of synchronous content, technologies such as videoconferencing and web conferencing are often used. Examples here include the use of Skype (www.skype.com),

Webex (www.webex.com) and GoToMeeting (www.gotomeeting.com) for the delivery of lectures and for virtual class discussions.

Interactive Learning Games

Interactive games are not only intrinsically engaging, but can also serve to help develop interpersonal and collaboration skills leadership, and promote creativity, critical-thinking, and problem-solving skills (Gee, 2003). Examples range from simple “typing tutor” games, where characters are falling down the screen and the learner must stop them before they reach the ground by pressing the corresponding key on the keyboard, to elaborate flight simulation games where one can learn the principles of flying an aircraft.

FUTURE RESEARCH DIRECTIONS

Although it is taken as a given in many circles that the addition of interactivity in the learning process is beneficial, the evidence for this viewpoint is not as strong as one might expect in medical education (v. Bridges (1992)). While use of interactivity offers a number of theoretical benefits (for instance, it is compatible with the constructivist model of learning), empirical proof remains rather limited.

Despite the obvious importance of interactivity in the educational process there are a number of questions that await definitive answers:

- Does interactivity really matter?
- How, if at all, does interactivity affect learners and learning?
- Do students prefer interactive learning?
- Do students perform better when exposed to interactive learning methods?

Perhaps the best available review of this topic in a medical context is a meta-analysis by Davis et al. (1999). In their study of a small number of “well-conducted” trials on methods of CME,

they noted that traditional didactic teaching sessions “do not appear to be effective” in changing physician behavior but noted that there is “some evidence that interactive CME sessions that enhance participant activity and provide the opportunity to practice skills can effect change in professional practice.”

CONCLUSION

Adult learning theory emphasizes the importance of interactivity in the educational process. In exploring the differences between traditional teaching methods and distance education, the issue of differences in the patterns and properties of interaction in the two settings figures prominently. Students in distance education interact variously with the instructor, the content, the technology, and with other students. Although four kinds of interactivity have been identified in the literature, other models specific to particular domains are possible. Interaction with technology will likely become increasingly important in computer-based distance education, but good design principles must be adhered to for successful outcomes.

REFERENCES

- Anderson, T. (2004). Toward a theory of online learning. In T. Anderson & F. Elloumi (Eds.), *Theory and Practice of Online Learning*. Athabasca University.
- Anderson, T., & Garrison, D. R. (1995). Transactional issues in distance education: The impact of design in audioteleconferencing. *American Journal of Distance Education*, 9(2), 27–45. doi:10.1080/08923649509526886
- Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, 13(6), 4–16. doi:10.3102/0013189X013006004

- Bridges, E. M. (1992). Problem-based learning: Background and rationale. In E.M. Bridges (Ed.), *Problem-based learning for administrators*. ERIC Documents: ED347617.
- Burbules, N., & Callister, T. (2000). Universities in transition: The promise and the challenge of new technologies. *Teachers College Record*, 102(2), 273–295. doi:10.1111/0161-4681.00056
- Crawford, C. (2002). *Understanding Interactivity*. San Francisco: No Starch Press.
- Davis, D., O'Brien, M. A., Freemantle, N., Wolf, F. M., Mazmanian, P., & Taylor-Vaisey, A. (1999). Impact of Formal Continuing Medical Education: Do Conferences, Workshops, Rounds, and Other Traditional Continuing Medical Education Activities Change Physician Behavior or Health Outcomes? *Journal of the American Medical Association*, 282(9), 867–874. doi:10.1001/jama.282.9.867
- Faghihi, S. A., Khankeh, H. R., Hosseini, S. J., Soltani Arabshahi, S. K., Faghih, Z., Parikh, S. V., & Shirazi, M. (2016). Improving continuing medical education by enhancing interactivity: Lessons from Iran. *J Adv Med Educ Prof*, 4(2), 54–63.
- Fahy, P. J. (2005). Two Methods for Assessing Critical Thinking in Computer-Mediated Communications (CMC) Transcripts. *International Journal of Instructional Technology and Distance Learning*, 2(3). Retrieved from http://www.itdl.org/journal/mar_05/article02.htm
- Garrison, D. R. (1997). Computer conferencing: The post industrial age of distance education. *Open Learning*, 12(2), 3–11. doi:10.1080/0268051970120202
- Garrison, D. R., & Cleveland-Innes, M. (2005). Facilitating cognitive presence in online learning: Interaction is not enough. *American Journal of Distance Education*, 19(3), 133–148. doi:10.1207/s15389286ajde1903_2
- Gee, J. (2003). *What Video Games Have to Teach Us About Learning and Literacy*. New York: Palgrave Macmillan.
- Gilbert, L., & Moore, D. R. (1998). Building interactivity into web courses: Tools for Social and Instructional Interaction. *Educational Technology*, 38, 29–35.
- Hara, N., & Kling, R. (1999). *Students' frustrations with a web-based distance education course: A taboo topic in the discourse*. Center for Social Informatics, Indiana University. Retrieved from http://rkcsi.indiana.edu/archive/CSI/WP/wp99_01-octoberB.html
- Hara, N., & Kling, R. (2000). *Students' distress with a web-based distance education course: An ethnographic study of participants' experiences*. Center for Social Informatics, Indiana University. Retrieved from <http://rkcsi.indiana.edu/archive/CSI/WP/wp00-01B.html>
- Harnad, S. (1992). *Interactive Publication: Extending American Physical Society's Discipline-Specific Model for Electronic Publishing*. Retrieved from <http://users.ecs.soton.ac.uk/harnad/Papers/Harnad/harnad92.interactivpub.html>
- Lapadat, J. C. (2007). Discourse devices used to establish community, increase coherence, and negotiate agreement in an online university course. *Journal of Distance Education*, 21(3), 59–92.
- Liebenluft, J. (2007). A Librarian's Worst Nightmare: Yahoo! Answers, where 120 million users can be wrong. *Slate*. Retrieved from http://www.slate.com/articles/technology/technology/2007/12/a_librarians_worst_nightmare.html
- McIsaac, M. S., & Gunawardena, C. N. (1996). In D. H. Jonassen (Ed.), *Distance Education, Handbook of research for educational communications and technology: a project of the association for educational communication* (pp. 403–437). New York: Simon & Schuster Macmillan.

Moore, M. G., & Kearsley, G. (1996). *Teaching and Tutoring, Distance education: A systems view*. Wadsworth Publishing Company.

Newberry, B. (2001). Raising student presence in online classes. *WebNet 2001: World Conference on the WWW and Internet Proceedings*.

Nielsen, J. (1994). *Usability Engineering*. San Francisco: Morgan Kaufmann.

Nippard, E., & Murphy, E. (2007). Social presence in the web-based synchronous secondary classroom. *Canadian Journal of Learning and Technology*, 33(1). Retrieved from <http://www.cjlt.ca/index.php/cjlt/article/view/24/22>

Norman, D. A. (1999). *The Invisible Computer*. Cambridge, MA: MIT Press.

Philip, J. H. (1986). Gas man - an example of goal oriented computer-assisted teaching which results in learning. *International Journal of Clinical Monitoring and Computing*, 3(3), 165–173. doi:10.1007/BF01716358

Philip, J. H. (2009). 'Gasman'-Understanding Anesthesia Uptake & Distribution [CD-ROM]. Chestnut Hill, MA: Med Man Simulations.

Richardson, C., & Swan, K. (2003). Examining social presence in online courses in relation to students' perceived learning and satisfaction. *Journal of Asynchronous Learning Networks*, 7, 68–88.

Rourke, L., Anderson, T., Garrison, D. R., & Archer, W. (1999). Assessing social presence in asynchronous, text-based computer conferencing. *Journal of Distance Education*, 14, 50–71.

Simonson, M., Smaldino, S., Albright, M., & Zvacek, S. (2009). *Teaching and learning at a distance: Foundations of distance education* (4th ed.). Upper Saddle River, NJ: Pearson Education, Inc.

Simpson, R. J., & Galbo, J. J. (1986). Interaction and learning: Theorizing on the art of teaching. *Interchange*, 17(4), 37–51. doi:10.1007/BF01807015

Smith, N. T., & Davidson, T. M. (2009). *Body Simulation-Anesthesia, Physiology and Pharmacology Simulation* [CD-ROM]. Point Roberts, WA: Advanced Simulation Corporation.

Tapscott, D. (1998). *Growing Up Digital: The Rise of the Net Generation*. New York: McGraw-Hill.

Wagner, E. D. (1994). In support of a functional definition of interaction. *American Journal of Distance Education*, 8(2), 6–26. doi:10.1080/08923649409526852

Wagner, E. D. (1997). Interactivity: From agents to outcomes. *New Directions for Teaching and Learning*, 71(71), 19–26. doi:10.1002/tl.7103

ADDITIONAL READING

Boettcher, J., & Conrad, R. (2016). *The Online Teaching Survival Guide: Simple and Practical Pedagogical Tips* (2nd ed.). New York: Jossey Bass.

Bogossian, F. E., Cooper, S. J., Cant, R., Porter, J., & Forbes, H. (2015). A trial of e-simulation of sudden patient deterioration (FIRST2ACT WEB) on student learning. *Nurse Education Today*, 35(10), e36–e42. doi:10.1016/j.nedt.2015.08.003

Cant, R. P., & Cooper, S. J. (2014). Simulation in the Internet age: The place of web-based simulation in nursing education. An integrative review. *Nurse Education Today*, 34(12), 1435–1442. doi:10.1016/j.nedt.2014.08.001

Du, S., Liu, Z., Liu, S., Yin, H., Xu, G., Zhang, H., & Wang, A. (2013). Web-based distance learning for nurse education: A systematic review. *International Nursing Review*, 60(2), 167–177. doi:10.1111/inr.12015

Faghihi, S. A., Khankeh, H. R., Hosseini, S. J., Soltani Arabshahi, S. K., Faghih, Z., Parikh, S. V., & Shirazi, M. (2016). Improving continuing medical education by enhancing interactivity: lessons from Iran. *J Adv Med Educ Prof*, 4, 54–63.

Jelacic, S., Bowdle, A., Togashi, K., & VonHoyer, P. (2013). The use of TEE simulation in teaching basic echocardiography skills to senior anesthesiology residents. *Journal of Cardiothoracic and Vascular Anesthesia*, 27(4), 670–675. doi:10.1053/j.jvca.2013.01.016

Mitchell, J. D., Mahmood, F., Wong, V., Bose, R., Nicolai, D. A., Wang, A., & Matyal, R. et al. (2015). Teaching concepts of transesophageal echocardiography via Web-based modules. *Journal of Cardiothoracic and Vascular Anesthesia*, 29(2), 402–409. doi:10.1053/j.jvca.2014.07.021

Noble, J. (2012). *Programming Interactivity* (2nd ed.). Newton, Mass.: O'Reilly.

Vai, M., & Sosulski, K. (2015). *Essentials of Online Course Design: A Standards-Based Guide* (2nd ed.). Oxford: Routledge.

KEY TERMS AND DEFINITIONS

Adult Learning Theory: Pertaining to the art and science of adult learning. Also known as andragogy.

Asynchronous Interaction: Interaction that does not take place synchronously and so is independent of time restrictions. Examples include messages sent via e-mail, SMS or messages posted to a discussion server. Contrasted to synchronous interaction, as exemplified by classroom teaching or holding a telephone conversation.

Course Management System: Software platforms that provide an instructor and students with resources (course document storage, e-mail,

discussion threads, videos, examination capabilities, etc.) for the delivery of online instruction.

Distance Education: The delivery of instruction that does not require the learner to be present in the same physical location as the instructor.

Electronic Publication: Also referred to as e-publishing or digital publishing or online publishing, electronic publication includes the digital publication of e-books, digital magazines, and journal articles, often in HTML: or PDF format.

Interaction: Mutual or reciprocal action or influence, as with the interaction between the moon and oceans causing tides.

Interactive Learning Environment: An environment that promotes learning via interactive processes, as exemplified by computer-based training or interactive computer simulation / games.

Interactivity: Reciprocal events that involve two or more objects and actions via mutual influences.

Massive Open Online Course (MOOC): An online course (often a university course) that offers unlimited participation, open access via the web, and automatic evaluation of the learner's progress.

Social Media: Websites and applications that enable users to create and share content (e.g., photographs) or to participate in social networking. Examples include Facebook, Twitter and MySpace.

ENDNOTE

- ¹ In the world of aviation, simple simulators are usually called “trainers”, with the term “simulator” being reserved for machines that provide graphics and/or full-motion replication of aircraft motions or actions.

Medical Equipment and Economic Determinants of Its Structure and Regulation in the Slovak Republic

M**Beáta Gavurová***Technical University of Košice, Slovakia***Viliam Kováč***Technical University of Košice, Slovakia***Michal Šoltés***Technical University of Košice, Slovakia*

INTRODUCTION

Healthy population is a significant component in the process of economic growth of the country, because the full economic potential can only be achieved health. The health sector is affected by the innovation process, which includes skilled workforce healthcare sector is one of the largest in the European Union. Research and development in the health sector has the potential to reach 0.3% of gross domestic product. healthcare sector is one of the largest in the European Union, representing around 10% of the European Union gross domestic product and employs one in 10 workers with a higher than average proportion of workers with tertiary education. Support for human health is an integral part of the objectives of smart and inclusive growth and Europe 2020 (European Commission, 2011). An innovative development is necessary to exist to ensure the sustainability of the sector, as well as to achieve inclusive growth in the context of demographic change. It is not related only with the growing health needs of the population, but also pressure to increase the efficiency of the health system.

BACKGROUND

Modern medical equipment and technologies enable efficient and rapid diagnosis and treatment of diseases, hence their deployment in the health system necessity (Rosina et al., 2014). In addition to the quantifiable economic savings also it happens to humane treatment, prevention of possible complications, and thus to the elimination of the additional costs in the event of subsequent medical care for patients. Timely and successful diagnostics will save unnecessary medical procedures according to experts who make up one third – analysis of the European Association of medical technology Eucomed. Modern medical technologies, modern approaches to diagnosis and treatment combined with early detection also allow to shorten hospitalisation time for some diseases from several weeks to a few days. Reducing the number of hospital admissions and inpatient treatment time and replace feature to make it by the day surgery has been a global trend lately (Gavurová & Šoltés, 2014). In addition to radical savings in the health system derived from therapeutic and diagnostic procedure in conjunc-

tion with application of the latest research findings and health technologies, the macroeconomic view on this issue is also important. Within the medical technology industry in Europe more than 500,000 people are employed, making it a production of medical technologies important element of the gross domestic product of individual European countries. Upgrading of medical technologies occurs not only to improve them, but also to their constant replacement of modern and innovative technologies in the short term. This is confirmed by the fact that, for instance in 2012 the European Patent Office received more than 10,000 patent applications in the field of medical technology and the number is increasing. It is related to the significant progress in research and development.

Despite the positive development of modern medical technology an important component in the process of successful diagnosis and treatment becomes their availability. Availability of advanced medical technologies in the Slovak Republic is not proportional and regionally balanced. There are various problems that have long-term systemic nature (Gavurová & Šoltés, 2014; Kneppo et al., 2014).

If we look at the situation abroad, the density of medical equipment in different countries varies considerably in the European countries – for instance in the case of devices for x-ray computed tomography and magnetic resonance imaging. The Greeks have ten times more magnetic resonance imaging devices for a person like the Hungarians, the Austrians three times more than the British. The Greeks have the most devices for x-ray computed tomography, there are three times more x-ray computed tomographs per person than the Dutch have or just like the Italians. The differences are evident not only among countries but also within countries. For instance, there are four-fold differences between trusts in the number of top medical equipment in the United Kingdom. If we conducted an international comparison of the other kinds of medical equipment in the country, we would have confirmed the existing significant differ-

ences between countries. Therefore, it is difficult to determine the optimal number in the country.

The primary objective of introducing and using of the variety of health technology is to improve the health of the population of the country. The Slovak Republic has still lacked standard procedures to govern the rational use of various diagnostic technologies. The negative consequences include unnecessary tests, which are also reflected in the economics of the Slovak healthcare system (Gavurová et al., 2014). There is no methodology to determine the optimal number of medical devices. Their introduction into medical practice is now accompanied with many obstacles. These are linked with the strategies of health insurance companies and with the waiting time for examinations -- waiting lists for procedures. Waiting, keeping the patient on the waiting list respectively, waiting time refer to the time between the placing of a patient on the waiting list and its disposal (Roberts et al., 2004). In relation to the effective healthcare, this is generally the time between approval and implementation of procedures (Mužik & Szalayová, 2013). According to the Organisation for Economic Co-operation and Development one of the reasons for the increase waiting lists has been a significant advance in surgical techniques and anaesthesia during the last two or three decades, what greatly have improved range, safety and efficacy of surgical procedures (Siciliani & Hurst, 2013). Many of the procedures have carried out at lower unit costs consequently a dramatic increase has occurred in demand for surgery, especially in effective – not acute – procedures, such as cataract surgery, hip surgery, or bypass in all the Organisation for Economic Co-operation and Development member countries. Waiting lists are a serious problem in these countries, including the Slovak Republic (Eriksson & Björnberg, 2009, Mužik & Szalayová, 2013). Long waiting times for efficient performances are typical for countries that combine health insurance with no or low deductibles and limited capacity of the patient (Siciliani, 2003). Fewer occurrences are

in countries with insurance schemes, in which choice of health insurance is allowed.

Waiting times for examinations cannot be reduced by introducing a greater number of devices in the diagnostic process, because, for example, many machines on magnetic resonance imaging or x-ray computed tomography examinations are currently used only a few hours per week (Mužík & Szalayová, 2013). A significant limiting factor for higher and more efficient use is mainly contractual limits by health insurance companies that this whole mechanism availability and use of appliances targeted manner.

ANALYSIS

Our analysis is based on the individual types of the medical equipment used in the healthcare facilities. We have applied categorisation of this equipment according to National Health Information Center (Národné centrum zdravotníckych informácií) of the Slovak Republic that differentiates twenty-nine types of the healthcare devices. We have focused not only on this categorisation, but also on the localisation of this medical equipment from a view of elementary administrative division in terms of the highest level of self-government in the Slovak Republic. There are eight self-governing regions, where the Banská Bystrica Self-governing Region, the Bratislava Self-governing Region, the Košice Self-governing Region, the Nitra Self-governing Region, the Prešov Self-governing Region, the Trenčín Self-governing Region, the Trnava Self-governing Region and the Žilina Self-governing Region belong.

Dataset

Our dataset contains data coming from seven years from 2008 to 2014. For each year every type of the medical equipment in counted according to its localisation in the particular self-governing regions.

At first, the basic look at the distribution of the individual equipment types is offered with total numbers of these types and in the separate self-governing regions.

Figure 1 displays localisation of the equipment types in the separate self-governing regions. The x axis shows order of all the equipment types recognised by the dataset from the National Centre of Health Information of the Slovak Republic and the y axis signifies the administrative division of the Slovak Republic at the level of the territory units in terms of the self-governing regions. Grey shading of the x axis tags represents a share of the number of equipment of that type in the total number of all the 29 equipment types on the x axis, respectively a share of the equipment localised in the individual self-governing regions in the total number of the equipment in the whole Slovak Republic on the y axis. Circles' diameter in the individual cells and its shading are determined by a share of the particular equipment in the total number of all the equipment in the peculiar self-governing region in a column way and also by a share of the particular equipment in the total number of all the devices of that type in the whole Slovak Republic in a row way. Empty cells express zero number of that type equipment (Bendixen, 2003).

At the beginning of the observed period in 2008 the most numerous type of medical equipment is monitoring device with 2423 pieces in the whole Slovak Republic. The Bratislava Self-governing Region has exactly 500 items, what expresses also the highest number of any type. On the opposite side there is positron tomograph, which is the least numerous type of the medical equipment. There are only 3 such devices in all the regions.

There are 8800 devices together in all the healthcare facilities. From 29 types, 9 of them are not localised in everyone region – the mostly absenting is positron tomograph in 5 regions and then brachytherapy apparatus in 3 regions.

At the end of the observed period in 2014 monitoring device remains the most numerous type. Its number was increased by 57 pieces to

Figure 1. Number of the equipment types in the self-governing regions in 2008

Source: Own elaboration compiled by the authors, 2016.

Legend for the x axis: 1 – angiograph; 2 – brachytherapy apparatus; 3 – bronchoscope; 4 – cystoscope; 5 – dialysis monitor; 6 – electrocardiograph; 7 – electroencephalograph; 8 – electromyography; 9 – endoscope, laparoscope, arthroscope; 10 – gamma camera; 11 – gastroscope, duodenoscope; 12 – isotope irradiator; 13 – colonoscope, sigmoidoscope, proctoscope; 14 – colposcope; 15 – cryogenic device; 16 – laryngoscope; 17 – laser; 18 – linear accelerator; 19 – lithotripter; 20 – magnetic resonance imaging device; 21 – mammograph; 22 – monitoring device; 23 – positron tomograph; 24 – x-ray; 25 – tomograph; 26 – ultraviolet and infrared emitter; 27 – ultrasound device; 28 – uretroscope; 29 – high-frequency device. Legend for the y axis: BB – the Banská Bystrica Self-governing Region, BA – the Bratislava Self-governing Region, KE – the Košice Self-governing Region, NR – the Nitra Self-governing Region, PO – the Prešov Self-governing Region, TN – the Trenčín Self-governing Region, TT – the Trnava Self-governing Region, ZA – the Žilina Self-governing Region.

Equipment	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	
Region																														
BB	5	2	13	5	46	187	19	9	63	4	31	4	21	5	2	7	45	2	4	4	7	309		97	12	1	112	3	28	1047
BA	8	1	13	21	457	288	18	15	102	13	54	3	41	16	8	20	60	5	6	7	19	500	2	145	14	6	248	18	55	2163
KE	9	4	26	15	20	265	12	15	115	10	43	6	39	6	8	19	35	2	4	10	13	431		120	9	2	131	11	79	1459
NR	1		22	11	30	111	9	10	53	2	14	4	20	3	2	6	16		5	5	6	202	1	103	6	4	114	5	22	787
PO	7		25	9	27	117	11	8	69	10	15	2	21	2	2	6	24	1	4	5	9	293		83	8	7	101	7	27	900
TN	3	1	11	10	15	129	12	2	41		12	1	20	3	1	2	13	2	2		5	119		61	4	5	73	3	21	571
TT			8	10	39	88	12	3	48	2	17		16		5	8	12		4	1	5	149		75	6		59	12	12	591
ZA	7	3	23	13	54	208	11	7	90	3	44	7	23	9	4	9	29	2	4	1	8	420		116	11	5	124	8	39	1282
	40	11	141	94	688	1393	104	69	581	44	230	27	201	44	32	77	234	14	33	33	72	2423	3	800	70	30	962	67	283	8800

Figure 2. Number of the equipment types in the self-governing regions in 2014

Source: Own elaboration compiled by the authors, 2016.

Legend: The same as for Figure 1.

Equipment	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	
Region																														
BB	5	1	9	7	53	196	11	5	46	1	28	1	21	6	6	6	27		2	2	8	312	1	61	10	7	145	1	30	1008
BA	5	1	18	17	100	283	15	14	113	7	54	3	36	16	19	24	43	6	5	9	22	404	3	108	15	18	278	16	68	1720
KE	7		16	4	151	233	10	11	77	8	29	3	25	2	6	17	25	3	1	2	11	292	1	65	8	10	131	4	36	1188
NR	3		19	8	109	157	9	8	46	2	16	3	14	6	2	8	19	1	3	4	10	185	2	76	9	3	116	6	42	886
PO	8	2	27	8	83	144	11	8	76	3	27	3	19	4	3	9	29	2	7	6	10	407		74	11	14	151	8	40	1194
TN	3		11	9	125	155	13	4	72		15		21	3	1	4	21	2	2	4	9	147		60	11	13	107	5	25	842
TT	1		7	13	87	113	12	9	65		21		17		6	7	17		5	4	8	214		61	7	7	90	21	21	813
ZA	5	3	20	11	108	243	14	5	110	4	36	6	20	9	6	9	33	3	5	5	11	519		96	13	14	170	11	52	1541
	37	7	127	77	816	1524	95	64	605	25	226	19	173	46	49	84	214	17	30	36	89	2480	7	601	84	86	1188	72	314	9192

the total number of 2480. The highest share belongs to the Žilina Self-Governing Region with 519 devices. The least numerous type a positron tomograph was doubled by a brachytherapy apparatus, which are represented by 7 items in all the regions together.

There are 9192 medical devices in all the Slovak healthcare facilities. There is to note that also these two equipment types – positron tomograph and brachytherapy apparatus – were missing in the largest number of the regions – in 4, while on the one hand positron tomograph was added

in 2 regions and on the other hand brachytherapy apparatus was set aside in 1 region.

Localisation of the Medical Equipment Types

To analyse distribution of the devices throughout the Slovak Republic, we quantify an equipment profile for each type of the equipment in 2014.

This profile performs as a share of that type equipment localised in the particular self-governing region of the total number of the equipment in the whole country.

Unsuitable localisation of the devices among the self-governing regions in 2014 is seen mainly in a case of the least numerous types that are mentioned above.

Table 1. Equipment profiles in 2014

	BB	BA	KE	NR	PO	TN	TT	ZA
1	0.1351	0.1351	0.1892	0.0811	0.2162	0.0811	0.027	0.1351
2	0.1429	0.1429	0	0	0.2857	0	0	0.4286
3	0.0709	0.1417	0.126	0.1496	0.2126	0.0866	0.0551	0.1575
4	0.0909	0.2208	0.0519	0.1039	0.1039	0.1169	0.1688	0.1429
5	0.065	0.1225	0.185	0.1336	0.1017	0.1532	0.1066	0.1324
6	0.1286	0.1857	0.1529	0.103	0.0945	0.1017	0.0741	0.1594
7	0.1158	0.1579	0.1053	0.0947	0.1158	0.1368	0.1263	0.1474
8	0.0781	0.2188	0.1719	0.125	0.125	0.0625	0.1406	0.0781
9	0.076	0.1868	0.1273	0.076	0.1256	0.119	0.1074	0.1818
10	0.04	0.28	0.32	0.08	0.12	0	0	0.16
11	0.1239	0.2389	0.1283	0.0708	0.1195	0.0664	0.0929	0.1593
12	0.0526	0.1579	0.1579	0.1579	0.1579	0	0	0.3158
13	0.1214	0.2081	0.1445	0.0809	0.1098	0.1214	0.0983	0.1156
14	0.1304	0.3478	0.0435	0.1304	0.087	0.0652	0	0.1957
15	0.1224	0.3878	0.1224	0.0408	0.0612	0.0204	0.1224	0.1224
16	0.0714	0.2857	0.2024	0.0952	0.1071	0.0476	0.0833	0.1071
17	0.1262	0.2009	0.1168	0.0888	0.1355	0.0981	0.0794	0.1542
18	0	0.3529	0.1765	0.0588	0.1176	0.1176	0	0.1765
19	0.0667	0.1667	0.0333	0.1	0.2333	0.0667	0.1667	0.1667
20	0.0556	0.25	0.0556	0.1111	0.1667	0.1111	0.1111	0.1389
21	0.0899	0.2472	0.1236	0.1124	0.1124	0.1011	0.0899	0.1236
22	0.1258	0.1629	0.1177	0.0746	0.1641	0.0593	0.0863	0.2093
23	0.1429	0.4286	0.1429	0.2857	0	0	0	0
24	0.1015	0.1797	0.1082	0.1265	0.1231	0.0998	0.1015	0.1597
25	0.119	0.1786	0.0952	0.1071	0.131	0.131	0.0833	0.1548
26	0.0814	0.2093	0.1163	0.0349	0.1628	0.1512	0.0814	0.1628
27	0.1221	0.234	0.1103	0.0976	0.1271	0.0901	0.0758	0.1431
28	0.0139	0.2222	0.0556	0.0833	0.1111	0.0694	0.2917	0.1528
29	0.0955	0.2166	0.1146	0.1338	0.1274	0.0796	0.0669	0.1656

Legend: The same as for Figure 1.

To examine profile of the self-governing regions from an equipment type point of view regional equipment profile is applied. It characterises the particular type of medical device in the light of all the self-governing regions. Regional equipment profile expresses a mean share of all the equipment profiles in each region.

Regional equipment profiles demonstrate average localisation of the equipment types according to the self-governing regions. The largest decrease all over the observed years appeared in a case of the Bratislava Self-governing Region, where regional equipment profile sank from 0.2458 in 2008 to 0.1871 in 2014. In terms of the absolute numbers of equipment it means fall from 2163 to 1720 devices. The largest year-over-year change happened to the Košice Self-governing Region, where in the last observed year was reduction from 0.1963 in 2013 to 0.1292, which in the absolute numbers mean loss of 881 devices – a 42.58% share of 2013. The Banská Bystrica Self-governing Region keeps its share on the same level roughly. All the other regions were slightly increasing their regional equipment profiles throughout the explored period with the highest year-over-year increment between 2013 and 2014.

Equipment Distribution

Distribution of the medical devices among the healthcare facilities have to take into account several parameters. From our point of view, we

look at the localisation at level of the highest tier self-government. Equipment spreading depends on financial capacity of individual healthcare facilities in the region, as there are public and private ones. Methods of their financing are considerably differentiated. Public healthcare facilities, whose debt steadily annually increases, have limited investment opportunities unlike other private healthcare facilities, which operate at a profit. It is difficult to pick up a need for proportional distribution. To get more detail analysis it is necessary to obtain information about investment strategies and financial possibilities of healthcare facilities and contracted volumes of medical services by health insurance companies.

There is to note that statement about oversizing or undersizing of numbers is based only on conclusion of our correspondence analysis.

Diagram marked as Figure 3 demonstrates localisation of the equipment types according to the individual self-governing regions regarding its share in these regions and its share among all the devices of that type too. Quadrilateral area of one cell represents a share of a number of the particular type equipment in the total number of all the equipment in the whole Slovak Republic or in the particular self-governing region. Blue shading expresses a situation, when a number of the particular type equipment is higher than it was in a case of their uniform distribution among all the healthcare facilities in the Slovak Republic. On the other hand, the red shading indicates a

Table 2. Regional equipment profiles

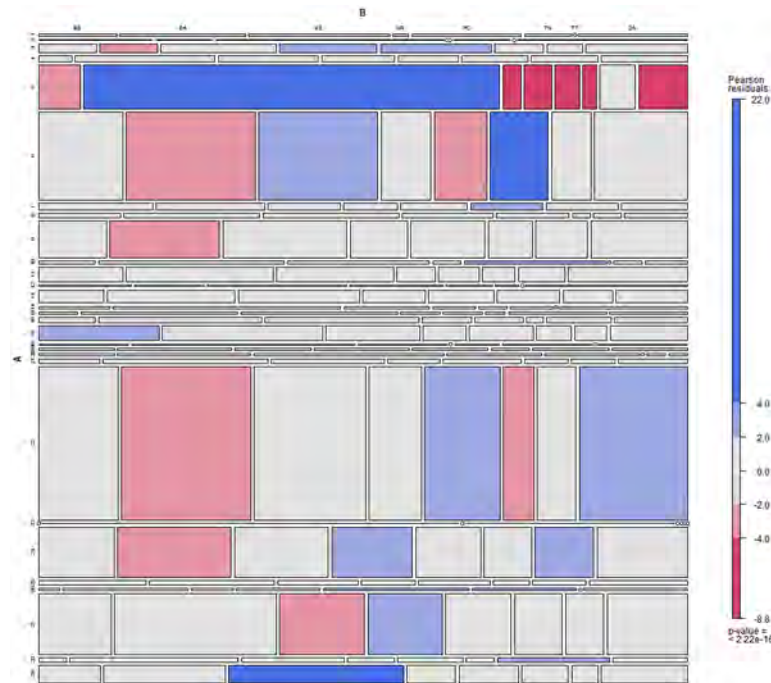
Year	BB	BA	KE	NR	PO	TN	TT	ZA
2008	0.119	0.2458	0.1658	0.0894	0.1023	0.0649	0.0672	0.1457
2009	0.1246	0.2118	0.1756	0.0875	0.1107	0.0719	0.0647	0.1533
2010	0.1178	0.2043	0.1939	0.0828	0.1092	0.0751	0.0696	0.1473
2011	0.117	0.2024	0.1978	0.0798	0.1091	0.0744	0.0684	0.151
2012	0.1187	0.2011	0.1956	0.0818	0.1093	0.073	0.0753	0.1452
2013	0.1173	0.1992	0.1963	0.0804	0.113	0.075	0.0747	0.1441
2014	0.1097	0.1871	0.1292	0.0964	0.1299	0.0916	0.0884	0.1676

Legend for the x axis: The same as for the y axis of Figure 1.

Figure 3. Equipment localisation in 2008

Source: Own elaboration compiled by the authors, 2016.

Legend: The same as for Figure 1.



state where there is a smaller number of that type equipment localised in the specific self-governing regions than it was in a case of uniform localisation.

The most unsuitably distributed device among the self-governing regions at the beginning of the observed period in 2008 is dialysis monitor. There were 457 monitors in the Bratislava Self-governing Region, where it reaches a dataset's maximum value of Pearson residual. It creates a share of 66.42% of this type total number and this can be considered oversized. Such a situation is reflected in the lowest Pearson residual in the five other self-governing regions – in the Košice Self-governing Region, the Nitra Self-governing Region, the Prešov Self-governing Region, the Trenčín Self-governing Region and in the Žilina Self-governing Region – where a number of dialysis monitors is undersized. There are still two extreme cases – electroencephalograph and high-frequency device. A number of electroencepha-

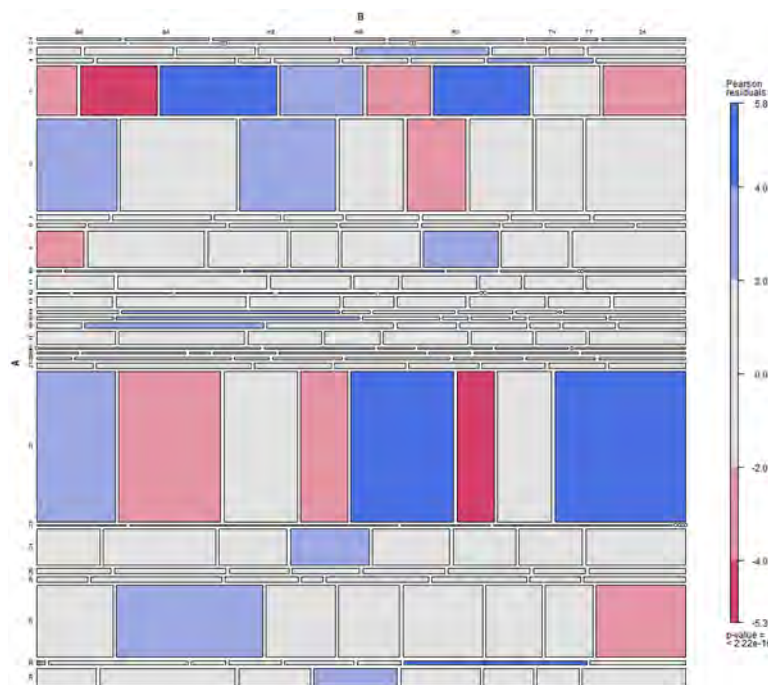
lographs in the Trenčín Self-governing Region is 129; although it is not a minimum value of all the regions, regarding the other equipment, it can be marked as undersized here. The second one extreme case is created by high-frequency device in the Košice Self-governing Region, where 79 devices are situated, what is a share of 27.92% of all that type devices in the Slovak Republic. Speciality of this case is the fact that no other region suffers from an undersized number of that type equipment. Such an instance only confirms that these analysis takes into account not only distribution of the equipment type among the regions, but also localisation of other equipment in the same region.

The situation in localisation of the equipment changed considerably over the explored period. In a case of dialysis monitor it turned the opposite way – its number decreased in the Bratislava Self-governing Region to 100, what represents a

Figure 4. Equipment localisation in 2014

Source: Own elaboration compiled by the authors, 2016.

Legend: The same as for Figure 1.



share of 12.25% of this device total number in the whole country, whilst in the Košice Self-governing Region its number increased to 151 with an 18.5% share in the Slovak Republic and this region is deliberated to be oversized by this device. Looking at the previous figure, there is another unsuitably localised medical device – monitoring device, whose situation does not change very significantly compared to the beginning of the observed period in 2008. Existing relations were only deepened more. Partially new situation occurred in cases of electromyograph and ultrasound device. Lack of the electromyographs in the Bratislava Self-governing Region in 2008 moved to Prešov Self-governing Region in 2014. In a case of ultrasound device, its lack in the Košice Self-governing Region in 2008 disappeared and raised in the Žilina Self-governing Region in 2014. On the other hand, its surplus from the Nitra Self-governing Region in 2008 moved to the Bratislava Self-governing Region in 2014.

Similarities in Equipment Localisation

State of medical equipment localisation among the healthcare facilities is very various in the Slovak Republic. To measure differences between situations in two equipment types we compute chi-squared distance.

The largest difference among the equipment types was in 2008 between brachytherapy apparatus and positron tomograph at level of 4.305. The largest similarity connects monitoring device with endoscope, laparoscope and arthroscope – their mutual distance is only 0.026. In 2014, the largest difference was between the same types as at the beginning of the observed period, but sank to 3.427. On the opposite side the two most similar devices stand – laser and ultrasound device that are separated from each other by only 0.008. In 2008, they lie at distance of 0.43. In the previous year, the most similar devices were in 2014 more distant from each other by 0.105.



Table 3. Chi-squared distances of equipment profiles in 2008 and 2014

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
1	0	0.567	0.307	0.42	1.476	0.191	0.449	0.295	0.194	0.274	0.238	0.377	0.249	0.349	0.596	0.371	0.221	0.334	0.515	0.466	0.198	0.147	2.982	0.342	0.268	0.516	0.278	0.684	0.153
2	1.041	0	0.974	1.001	2.54	0.559	0.991	0.871	0.674	1.325	0.524	0.477	0.761	0.765	0.95	0.787	0.754	0.802	1.21	1.138	0.873	0.659	4.305	0.876	0.811	1.662	0.928	1.367	0.457
3	0.127	1.107	0	0.222	1.908	0.224	0.323	0.187	0.123	0.498	0.342	0.278	0.171	0.635	0.552	0.392	0.361	0.815	0.173	0.382	0.228	0.168	2.602	0.13	0.232	0.366	0.206	0.527	0.239
4	0.608	1.702	0.409	0	1.158	0.11	0.184	0.244	0.087	0.55	0.186	0.509	0.05	0.386	0.222	0.186	0.281	0.49	0.112	0.548	0.093	0.157	2.064	0.066	0.186	0.476	0.088	0.17	0.201
5	0.376	1.923	0.258	0.24	0	1.193	1.394	1.24	1.332	1.151	1.055	1.935	1.224	0.653	1.166	1.028	0.975	0.844	1.316	1.596	0.932	1.174	1.094	1.28	1.178	1.778	0.93	1.046	1.398
6	0.213	1.365	0.21	0.218	0.129	0	0.125	0.146	0.037	0.452	0.05	0.288	0.029	0.227	0.271	0.147	0.081	0.277	0.175	0.436	0.061	0.048	2.363	0.063	0.071	0.574	0.059	0.333	0.073
7	0.344	1.525	0.261	0.074	0.095	0.084	0	0.325	0.153	0.658	0.226	0.601	0.118	0.529	0.416	0.34	0.158	0.47	0.13	0.67	0.188	0.198	2.623	0.1	0.089	0.652	0.151	0.351	0.337
8	0.38	1.861	0.28	0.184	0.21	0.177	0.186	0	0.115	0.291	0.189	0.266	0.131	0.368	0.347	0.146	0.149	0.608	0.17	0.108	0.105	0.099	1.889	0.123	0.162	0.703	0.101	0.426	0.129
9	0.277	1.266	0.204	0.121	0.115	0.069	0.043	0.182	0	0.354	0.075	0.283	0.031	0.344	0.207	0.111	0.133	0.46	0.109	0.349	0.051	0.026	2.453	0.036	0.068	0.547	0.073	0.239	0.076
10	0.472	1.618	0.612	1.115	0.787	0.524	0.917	0.583	0.664	0	0.458	0.816	0.407	0.632	0.507	0.315	0.356	0.725	0.525	0.266	0.231	0.302	2.436	0.488	0.422	0.833	0.385	0.516	0.358
11	0.232	1.16	0.244	0.187	0.275	0.058	0.131	0.143	0.08	0.485	0	0.274	0.119	0.178	0.207	0.107	0.091	0.339	0.23	0.498	0.105	0.045	2.569	0.117	0.078	0.801	0.111	0.294	0.123
12	0.501	0.746	0.373	0.997	0.81	0.534	0.833	0.787	0.612	0.469	0.538	0	0.39	0.429	0.716	0.478	0.383	0.796	0.468	0.54	0.438	0.23	2.658	0.314	0.336	0.857	0.356	0.893	0.277
13	0.249	1.566	0.256	0.146	0.117	0.038	0.045	0.134	0.055	0.648	0.068	0.786	0	0.335	0.235	0.135	0.141	0.374	0.119	0.363	0.039	0.079	2.245	0.043	0.107	0.498	0.052	0.253	0.085
14	0.543	1.198	0.464	0.574	0.682	0.299	0.526	0.594	0.438	0.662	0.28	0.499	0.411	0	0.583	0.374	0.218	0.178	0.337	0.743	0.251	0.239	1.76	0.367	0.314	0.744	0.205	0.651	0.325
15	0.734	1.722	0.768	0.392	0.726	0.372	0.492	0.312	0.414	0.63	0.188	0.994	0.341	0.432	0	0.054	0.392	0.772	0.302	0.53	0.219	0.267	2.46	0.266	0.323	1.219	0.32	0.078	0.264
16	0.316	1.633	0.314	0.349	0.311	0.15	0.302	0.095	0.207	0.254	0.099	0.888	0.156	0.377	0.192	0	0.205	0.57	0.229	0.292	0.096	0.127	2.164	0.171	0.198	0.991	0.171	0.156	0.122
17	0.162	1.166	0.151	0.175	0.179	0.028	0.07	0.17	0.049	0.583	0.031	0.543	0.036	0.285	0.35	0.162	0	0.283	0.224	0.408	0.098	0.075	2.211	0.143	0.049	0.723	0.086	0.43	0.2
18	0.491	1.503	0.482	0.691	0.519	0.371	0.561	0.582	0.351	0.38	0.369	0.652	0.391	0.361	0.572	0.293	0.366	0	0.739	0.979	0.329	0.402	2.432	0.525	0.433	0.704	0.341	0.818	0.409
19	0.497	1.149	0.278	0.226	0.506	0.455	0.29	0.314	0.271	1.166	0.321	0.823	0.394	0.807	0.677	0.519	0.301	0.906	0	0.448	0.147	0.148	2.058	0.033	0.1	0.617	0.097	0.214	0.32
20	0.368	1.356	0.175	0.103	0.243	0.195	0.122	0.184	0.095	0.837	0.15	0.712	0.141	0.391	0.436	0.255	0.109	0.405	0.159	0	0.307	0.348	2.132	0.413	0.474	1.041	0.38	0.681	0.267
21	0.277	1.511	0.191	0.134	0.144	0.05	0.087	0.097	0.066	0.557	0.057	0.626	0.036	0.279	0.288	0.096	0.04	0.306	0.344	0.077	0	0.055	2.019	0.083	0.101	0.51	0.037	0.219	0.103
22	0.159	0.789	0.158	0.278	0.325	0.113	0.172	0.255	0.105	0.565	0.062	0.376	0.164	0.378	0.417	0.24	0.061	0.495	0.236	0.195	0.16	0	2.331	0.066	0.043	0.57	0.063	0.304	0.1
23	1.527	3.427	1.36	1.498	1.454	1.119	1.527	1.087	1.54	1.158	1.177	1.479	1.251	0.759	1.053	0.882	1.236	1.34	1.991	1.359	1.002	1.521	0	2.156	2.375	2.52	1.789	2.285	2.522
24	0.257	1.331	0.135	0.11	0.115	0.044	0.047	0.125	0.046	0.674	0.074	0.534	0.06	0.341	0.414	0.186	0.032	0.448	0.252	0.082	0.035	0.099	1.188	0	0.055	0.533	0.044	0.542	0.175
25	0.23	1.309	0.158	0.145	0.121	0.049	0.035	0.221	0.049	0.776	0.103	0.644	0.046	0.34	0.498	0.265	0.026	0.42	0.318	0.096	0.056	0.127	1.329	0.028	0	0.642	0.072	0.306	0.22
26	0.255	1.228	0.266	0.248	0.223	0.165	0.118	0.346	0.064	0.776	0.165	0.809	0.103	0.508	0.545	0.327	0.101	0.305	0.365	0.142	0.149	0.188	1.827	0.163	0.094	0	0.457	1	0.693
27	0.196	1.233	0.168	0.175	0.209	0.035	0.097	0.152	0.072	0.547	0.026	0.543	0.044	0.222	0.286	0.125	0.008	0.332	0.326	0.099	0.022	0.087	1.069	0.038	0.043	0.134	0	0.292	0.168
28	1.417	2.422	1.071	0.315	0.813	0.901	0.598	0.516	0.62	1.783	0.725	1.699	0.753	1.489	0.752	0.843	0.82	1.472	0.422	0.57	0.716	0.822	2.438	0.663	0.825	0.875	0.823	0	0.422
29	0.212	1.204	0.106	0.219	0.195	0.052	0.142	0.161	0.095	0.494	0.068	0.375	0.104	0.203	0.378	0.135	0.041	0.329	0.33	0.109	0.039	0.097	0.976	0.031	0.066	0.206	0.028	0.863	0

Legend for the axes of Table 3 is the same as for the y axis of Figure 1. Legend for values of Table 3: Numbers written in the upper right half of the matrix belong to year 2008 and numbers displayed in the lower left half of the matrix belong to year 2014; main diagonal is the same for the both years because of its characteristic.

If we compare individual distances between the equipment types at the beginning and at the end of the observed period, we could find changes of similarities of the same equipment type pairs. There are two cases, when no change happened after the explored period – between angiograph and group of colonoscope, sigmoidoscope and proctoscope and then between laryngoscope and mammograph.

FUTURE RESEARCH DIRECTIONS

The main results of the executed analysis can be summarised subsequently. The most numerous healthcare equipment is monitoring device, dialysis monitor, electrocardiograph, and ultrasound device. The total number of the healthcare equipment was slightly increasing from 2008 to 2013, but in the last observed year 2014 fell down by 16.44% to 9192 devices. Some of the equipment types are not localised in each self-governing region, although during the explored period a few types were supplemented in the regions where there was lack of them.

Shares of a number of the equipment were throughout the observed period various. The Banská Bystrica Self-governing Region retained approximately the same shares, whilst share of the Bratislava Self-governing Region was sinking throughout the whole period from 24.58% down to 18.71%. In the Košice Self-governing Region, a share was raising from the beginning of the period to 2013, but then in the last year fell sharply from 19.92% to 18.71%. In the other self-governing regions, shares remained the same, but there was their small increase in the last observed year 2014.

Situation in the medical equipment types localisation was as in the successive lines. Dialysis monitor was oversized in the Bratislava Self-governing Region in 2008, but in 2014 already undersized. Oversizing of this equipment occurred also in the Košice Self-governing Region oversized in 2014. Another apparatus oversized in this region was high-frequency device in 2008. Unsuitable

localisation in 2008 with deepening of this state throughout the observed period occurred with monitoring device. Electromyograph appeared undersized in the Bratislava Self-governing Region in 2008 and in the Prešov Self-governing Region in 2014. Ultrasound device is a type of medical equipment with uneven distribution across the country too. Its number was undersized in the Košice Self-governing Region and in the Nitra Self-governing Region oversized in 2008, in the Žilina Self-governing Region undersized and in the Bratislava Self-governing Region oversized in 2014. All these numbers suggest there is missing systematic distribution of the medical equipment (Pharr, 2013; Otawová et al., 2015). This issue should be subject to discuss in future (Zavadiľ et al., 2016).

CONCLUSION

The development process and the mortality of global aging are processes that have impact on the system in the country and are limited by a number of determinants. In addition to health insurance system, expenses intended for healthcare, a network of healthcare providers and demand for them an important role is played in the quality of treatment processes and sophistication of medical technology. This is varied all over the different countries its level is influenced by the historical development of scientific and technical progress and the rate of implementation of research and development into practice. There is no long-term monitoring of the status of medical equipment in healthcare facilities in the Slovak Republic. Some of the healthcare facilities have still limited access to tomography or magnetic resonance imaging device. A critical condition occurs when in urgent cases – often in hospitals – patient wait for x-ray computed tomograph for several weeks. It is questionable whether the condition is caused by a lack or absence of financial resources from health insurance companies. At the present time, the Slovak economy health system is set so that

it does not allow natural reproduction of health technologies. Healthcare providers are not able to secure the financial resources to purchase new equipment during its operation life. It is therefore necessary to address this urgent problem and to deal with the availability and quality of instrumentation and to communicate at government level, the need to purchase diagnostic equipment for coordinating public health department, an issue to initiate setting of adequate normative that would force healthcare provider to modulate equipment in corresponding time period. An alternative to solve this problem would be the introduction of the process of financing healthcare technology from the state budget or by reimbursement from health insurance companies. Investments into radiological technology should be one of the most important indicators monitored within the Slovak health service strategic framework. Another issue is unsuitable distribution of medical equipment all over the Slovak regions, where are significant regional disparities between healthcare providers. They are not only in the field of localisation of the particular types of the medical equipment, but also in a way related to a share of the equipment in the individual self-governing regions of the Slovak Republic. These outputs should serve as a monitoring platform in evaluating accessibility and quality of healthcare technologies for use in this country. To obtain more detailed conclusions from the analysis we would need to have access to more data, which is the subject of our current research activities and cooperation with the institutions of the healthcare system.

REFERENCES

- Bendixen, M. (2003). A Practical Guide to the Use of Correspondence Analysis in Marketing Research. *The Marketing Bulletin*, (14), 1–15.
- Eriksson, D., & Björnberg, A. (2009). Canada Health Consumer Index 2009. *Frontier Centre for Public Policy Policy Series*, 61, 1–46.
- European Commission. (2011). *Rozhodnutie Európskeho parlamentu a Rady o ustanovení programu Zdravie pre rast, tretieho viacročného akčného programu EÚ v oblasti zdravia na obdobie rokov 2014 – 2020*. Brussels, Belgium: European Commission.
- Gavurová, B., & Šoltés, M. (2014). Identification of the Functionality Level of Day Surgery in Slovakia. *Ekonomický Casopis*, 62(10), 1031–1051.
- Gavurová, B., Šoltés, M., & Balloni, A. J. (2014). The Economic Importance of Using of ICT in the Health System (Ekonomický význam využívania informačno-komunikačných technológií v systéme zdravotníctva). *Ekonomický Casopis*, 62(1), 83–104.
- Gavurová, B., & Šoltés, V. (2014). The possibilities of day surgery system development within the health policy in Slovakia. *Health Economics Review*, 4(35), 1–12. PMID:24405520
- Kneppo, P., Ivlev, I., & Juříčková, I. (2014). *Health Technology Assessment*. Saarbrücken, Germany: Lambert Academic Publishing.
- Mužik, R., & Szalayová, A. (2013). *Časová dostupnosť vybraných efektívnych výkonov – analýza čakacích dôb*. Bratislava, Slovakia: Health Policy Institute.
- Otařová, R., Kamenský, V., Hasenöhrllová, P., & Rogalewicz, V. (2015). Cost-effectiveness studies on medical devices: Application in cardiology. *Lecture Notes in Computer Science: Vol. 9043. 3rd International Work Conference on Bioinformatics and Biomedical Engineering*, (pp. 163–174). Springer. doi:10.1007/978-3-319-16483-0_16
- Pharr, J. (2013). Accessible medical equipment for patients with disabilities in primary care clinics: Why is it lacking? *Disability and Health Journal*, 6(2), 124–132. doi:10.1016/j.dhjo.2012.11.002 PMID:23507163
- Roberts, M. J., Hsiao, W., Berman, P., & Reich, M. R. (2004). *Getting Health Reform Right*. Oxford, UK: Oxford University Press.

Rosina, J., Rogalewicz, V., Ivlev, I., & Kneppo, P. (2014). Health technology assessment for medical devices. *Lékař a technika*, 44(3), 23–36.

Siciliani, L., & Hurst, J. (2003). Explaining Waiting-Times Variations for Elective Surgery Across OECD Countries. *OECD Economic Studies*, 38(1), 95–123.

Zavadil, M., Rogalewicz, V., Kubátová, L., Matloňová, V., & Salačová, K. (2016). Hospital-based health technology assessment. *Casopis Lekaru Ceskych*, 155(5), 254–259. PMID:27696890

KEY TERMS AND DEFINITIONS

Efficiency: Known as effectiveness or productivity generally refers to the efficiency of the used resources and the utility obtained by their use. It is a ratio of inputs and outputs of an activity or a system.

Equipment: A device used in healthcare facilities to perform health care.

Health Care: A range of activities and measures leading to prolonged and sustained life of individuals, to improve quality of life and its protection, promotion, strengthening, improving and restoring health.

Health Policy: A guideline applied to health care system by state authorities responsible for health services.

Healthcare Facility: A place where medical equipment is used to perform health care.

Healthcare: A social set of professional knowledge institutions, establishments and authorities, workers and their corresponding activities, serving specifically to providing health care with an appropriate order to promote, preserve and restore health.

System of Health Care: Provision of health services. It is one component of health care and health policy determined by the state.

Use of Technology in Problem-Based Learning in Health Science

M**Indu Singh***Griffith University, Australia***Avinash Reddy Kundur***Griffith University, Australia***Yun-Mi Nguy***Griffith University, Australia*

INTRODUCTION

In current tertiary education environments, Information and Communication Technologies (ICT) has impacted on both the design and delivery of education to students with a strong focus on e-Learning (Clothey et al., October 2012). Interactive technology incorporating e-Learning has the ability to promote active learning, seeking information through accessing multiple modes of information and self-directed individually paced learning (Dede & Fontana 1995; Kahn 1993). Tertiary education faces a number of common challenges such as increasing class size, diverse cultural backgrounds, limited resources and teaching space facilities as well as concerns around the alignment of course curriculum for graduate capabilities (Gunn & Harper, 2007)(Smyth, Andrews, Caladine, & Bordujenko, 2011). To meet these demands, institutions such as Griffith University have had to employ existing technologies to increase learner autonomy. Through the introduction of e-learning activities, students have more flexible access to education and training, whilst avoiding increasing costs to campus infrastructure, laboratory staff and resources. This article will discuss the current use of technology to deliver various innovative forms of e-learning and blended problem based learning (PBL) as well

as some models that are being developed and trialled at Griffith University to train undergraduate Medical Laboratory Science (MLS) graduates for the global health workforce. Some of these technological models were designed and developed by the authors from Griffith University. These include digital slide microscopy, introduction of computerized cell differentiation counters, 3D virtual reality (a 360 degree panoramic view of a live pathology laboratory), demonstration videos and the use of electronic portfolios (e-portfolios). Throughout the 4 years of the degree program, simulated and real world case studies have been developed and scaffolded across the courses within the program both vertically across the year levels and horizontally across courses within same year level. This strategy of mix mode use of technology aligns with the cognitive learning principles identified by Swann (2013) suggesting that inclusion of multimedia requires learners to split their attention between different means of instructions and helps mentally integrate the learning.

Although digital microscopy has been widely used in laboratory based teaching for some time, its advantages have been limited due to its restricted access of images for students. With the recent development of Image Scope, this software platform has allowed students to view Aperio scanned slide images on multiple computers across campus and

off campus and has enabled students to access their dedicated course pathological slides without having to enter into a physical containment 2 (PC2) laboratory facility. Standardisation of training materials and extending the life capacity of rare and exotic case based slide sets that can degrade over time (such as loss of staining intensity due to long term storage) have also been achieved through the use of digital microscopy.

Simulation computer based morphology programs using Raspberry Pi mini computers have also enabled institutional laboratory facilities to incorporate electronic cell counters into haematology based practical sessions. This allows students to familiarise themselves with current cell counter technology found across private and public pathology laboratories. The computer system has been able to enhance educational learning and permits software development to be tailored to intuitional needs, without the burden of having to outlay large financial costs.

Furthermore, use of short demonstration videos showing different laboratory techniques have shown to improve student experience in understanding the techniques in detail. The demonstration videos had a significant impact through saving time and reducing the teaching load during laboratory sessions. Additionally encouraging students to develop their e-portfolio, a collection of their learning, accomplishments and experiences can greatly increase their employability. The use of e-portfolio has been recommended by a number of universities as it helps the student to develop their cognitive skills from self-directed learning. As the emergence of new technologies transform our traditional classrooms into virtual educational facilities, key challenges faced during the process, and issues yet to be addressed will also be highlighted in this article. As the demands for education and training continue to increase to meet the demands of current and future workforce, this article will focus heavily on how e-learning tools and blended PBL have impacted on MLS students studying at Griffith University and the challenges that were faced from both students and teachers.

BACKGROUND

Current electronic learning (e-learning) tools along with recent advances in Information and Communication Technologies (ICT) have significantly increased students access to evidence and problem based learning techniques (Clothey et al., October 2012). This has impacted on both designing and delivering education to health science students with a strong focus on e-learning (Gértrudix Barrio, Carmen Gálvez de la Cuesta, Álvarez García, & Valle, 2007; Kirkwood & Price, 2005; Salmon, 2011). Traditional teaching in classrooms accounted for conventional academic based didactic teaching, where investigator initiated and discipline based knowledge is taught in isolation. Today the face of learning and teaching is changing to context driven, problem based, work integrated, transdisciplinary and interprofessional knowledge, where the primary focus is on building an effective and efficient team to work for a short period of time on solving a specific problem from the real world (Gibbons, 1994)

E-learning technology can act as an active instigator for creating constructivist teaching methods and active learning strategies to enhance student knowledge and prepare them for the digital age. This mode of teaching can also enable the students to develop skills such as critical analysis and personal interpretation of knowledge (Cummings, Mason, Shelton, & Baur, 2015). E-learning has been a dynamic part of the educational landscape since the start of the 21st century and has been defined in various ways over time (Sangrà, Vlachopoulos, & Cabrera, 2012). Massy and Zemsky (2004) suggest three major categories of e-learning. The first category of e-learning described as distance education where courses are delivered almost entirely online via the internet. The second and third category of e-learning termed electronically mediated learning is where learning materials are digitally delivered through various technologies and e-learning is facilitated by transactions software. The software management systems organise and manage teach-

ing and learning schedules. In relevant terms for the purpose of this article, Ashwin (2006) also defines it as 'any form of the new technologies or applications in the service of learning or learner support' which includes any software, hardware, digital resources or technology to deliver and support education e-learning. It can also be replaced with mix mode learning where technology is used to replace some of the face to face teaching with technology mediated teaching. Even though there are many definitions for e-learning, it is evident that all e-learning forms may provide a learning opportunity for individuals to use various e-learning simulation and real case base learning tools, to produced more positive learning gains than traditional lecture methods (Cummings et al., 2015, Lencastre & Coutinho, 2015).

Griffith University Bachelor of Medical Laboratory Science program uses a combination of all the above means of e-learning to deliver a four year program preparing graduates for the global market. The early adoption of an integrated, automated, central and supported e-Learning environment allows reliable and ubiquitous delivery of services, in combination with other in-class learning tools. With increasing use of smart devices by students, learning-space technology is being delivered as software to the student's computer instead of hardware to the learning spaces. This enables Griffith University to encourage students to actively learn not only in the classroom but also at home or while travelling, and permits the use of new technologies in an economical way. Similarly health institutions using simulated 3D virtual videos can encourage and motivate not only students to practise critical thinking and reflective skills but to also allow new training methods for all staff types to revise and train on analysers that are not regularly available for training. Simulations can replace laboratory activities that can reduce the costs of lab resources such as time, equipment, supplies, staff expertise, and minimise health risks associated with dangerous activities (Cummings et al., 2015). In an edited book series by Clothey et al. (October 2012), a

study by Robyn Smyth, Trish Andrews, Richard Caladine, and Jason Bordujenko discuss the great potential for use rich media technology across a number of Australian universities in reducing costs and environmental footprints. They also found similar findings compiled from the National Centre of Science, Information and Communication Technology and Mathematics Education for Rural and Regional Australia (SiMERR) that the greatest benefits of using rich media technologies helped improved collaboraiton, convenient, reduced the ease of travelling, cost effective and promoted deeper discussions of course material (Clothey et al., October 2012).

E-learning can also promote inquiry based learning via exploration and can be individualised to match learners need and training requirements (Cummings et al., 2015). In a case study by Charl Wolhuter, Hennie J. Steyn, Elsa Mentz, and Ferdinand J. Potgieter in (Chapter 4) of Clothey et al. (October 2012), they found that the launch of blended leaning classes across South Afriacn universities, were found to be more positive than traditional face to face teaching structure as it allowed for flexible language choice, helped with the demands of increased enrolments and strained teaching resources.

MAIN FOCUS OF THE ARTICLE

The authors at Griffith University developed innovative, technology-based resources, to give students 24 hour access not only to PBL cases, lectures and published material but also digital technique videos and electronic haematology (e-Haem) Atlas which was developed by imaging our actual blood films of various disorders collected from Victoria, New South Wales and Queensland. Whole slide imaging (WSI) is a conventional way of digitally scanning slides in order to produce a virtual image (Pantanowitz, Farahani, & Parwani, 2015). WSI technology has been employed worldwide in pathology departments and has overcome barriers such as limited glass slide cases, faded

stained films stored over a number of years, use of slide sets at home or outside of a laboratory facility and allowed the standardisation of training sets for all students. With flexible slide accessibility, this complements student's morphology skills gained by actual microscopy in the laboratory, without the need for a demonstrator during students' voluntary self-study period. Suggestions were incorporated from peers to use videos of actual blood flow, platelet and WBC activity to help students understand the concepts easily. These teaching strategies build students' self-confidence, employability, and capacity to apply their skills and knowledge to a range of contexts not just within the Australian workforce but also relevant to the global workforce. These new innovative tools take the "classroom to the students" to enhance flexible learning where the software is available on every computer for students on campus as well as on their personal electronic devices.

For vertical and horizontal scaffolding of students learning across various courses an e-Portfolio is being rolled out to monitor and support student progress on and off campus; in collaboration with: Haematology Technical Staff, Curriculum, Assessment & Technology, Technical Solutions (Electronics), Network and Collaboration Services, Information Services, Corporate Technology Infrastructure and Griffith IT security teams. These innovative ways of using technology as a mainstream teaching tool has been recognized by peers from other institutions and countries resulting in multiple presentations about this curriculum design and teaching strategies within Australia and overseas, including as invited guest speaker.

Some of the techniques filmed at Griffith University were reviewed and discussed as part of the pre-laboratory session prior to each laboratory class. At this stage students were assessed by providing verbal answers and sometimes written responses on the worksheets provided for the case study covered in that particular practical class. Some of the questions asked were related

to the techniques they had viewed before physical demonstration of the techniques in the class. The assessment for Histology and Haematology courses includes mid-semester and end of semester practical exams. The impact of the videos on the routine laboratory and practical exam assessment were determined via a paper based survey and a formal question is included in the student evaluation of the courses.

A few of the methods filmed for MLS program were re-taken and edited after initial use, based on the industry and student feedback. Therefore some techniques videos which were implemented as part of teaching tool were complemented by incorporating still images of the same techniques provided to students. Responses to the modified teaching videos were very positive. Similar responses were also received from the teaching team, including demonstrators (experienced senior industry practicing scientists), as this allowed them time to focus efforts in improving students skills. Since the video of the whole set of the techniques was not available to any class at the start of the semester, official student evaluations could not be performed.

In 2014 all the videos formed a component of teaching material on the university's course website. It was ensured that each technique was made available for respective courses online at the start of semester and students were surveyed at the end of semester to obtain feedback on their experiences.

A current project underway in collaboration with staff of various departments at Griffith University is to develop and implement a Raspberry Pi2 (RPi) linked to old LCD monitors with built-in USB port to use as keyboard for the input. The RPi 2 systems are an inexpensive credit card sized computer that can be personalised to run an Ubuntu Linux operating system whereby programs such as the cell differential counter can be programmed and set up for individual use. RPi was founded by the not-for profit Raspberry Pi foundation in the United Kingdom for use in education to

allow children to familiarise themselves with computer hardware, programming and to enhance their computing learning experience (Ioannou, Ioannidis, Papadopoulos, & Tapeinos 2014). At Griffith University, instead of using this as a teaching tool only, we have incorporated the use of the RPi to program and run an electronic cell counter without having to outlay large costs for electronic cell counters available on the market. It also develops the skills of students who can familiarise themselves with the process of performing differentials using the computer based technology that is commonly found in industry pathology laboratories. By personalising the cell counter program we can allow students to perform white cell differential counts in class electronically and save these case base differentials on a file format which can be collated as part of e-portfolio for future review alongside digital images that can be accessed outside of the laboratory or of campus.

Primarily RPis will be used as an electronic cell counter device; however it has the potential to be used as minicomputer for the laboratory where it can allow the automatic connection of cameras to the microscopes networked in the teaching laboratories. This application of RPis will allow students to build their e-portfolio across all year levels as an evidence of competency of the skills gained during practical laboratory sessions in the classroom as well as in industry on clinical placement with endorsement by respective academic and industry supervisors. Upon graduation the students can then carry the same portfolio to job interviews.

E-portfolios are controlled by students for sharing certain folders with the academic and industry supervisors. This allows staff to monitor in real time the progress of the students and provide them with in time feedback and guidance instead of waiting until the end of semester and giving them final grades only. Also these portfolios can be carried by students across all 4 years of their study and they can use them to compile all their ongoing class work in other courses (they

develop folders/portfolios for example blood and tissue slides to identify the diagnostic features of the tissues and blood cells) in Haematology, Histology & Histopathology to gain multi skilled experience. Instead of carrying multiple folders with them to the clinical placement they can have all the material in one e-portfolio for future reference as well as an evidence of their competencies endorsed by academics and industry supervisors when they apply for jobs in workforce.

RPi in particular can be extremely beneficial for microscopy training in Haematology, Histology and Histopathology courses, with application in Clinical Biochemistry and Clinical Microbiology during first 3 years of the program and e-portfolio from 1st year through to 4th year. Assessment is an important driver for educational practice (Guskey 2003). E-portfolio can be used as an effective platform for formative assessments. With e-portfolio, students will have access that they control but can share files/folders with their academic supervisors as well as industry supervisors when on clinical placement. This will allow students to have all their study material across the courses in one single portfolio and will allow the supervisors to monitor their progress in real time to provide timely comments and feedback particularly in 1st year and 4th year where they have to keep a weekly reflective journal as part of their main assessment instead of end of semester final exam. Any work done by students on or off campus as part of an assessable critical reflection can be saved on their e-portfolio directly or through their external storage device. It will be very beneficial to have a user friendly system where students can save their reports and tissue/blood slides features to take with them to clinical placement and job interviews upon graduation, with their supervisors comments/feedback and endorsement of having met competencies.

Digital, role play, and simulations have become increasingly popular in higher education to provide a more authentic and engaging learning experience for students. There is extensive evidence

that ICT increases motivation, confidence and engagement (Blamire 2009). The virtual Reality 3D pathology experience project is underway. The purpose of this project is to show in the classroom and teaching laboratory the routine functioning, daily maintenance required and other basic details of the equipment routinely used by most laboratories worldwide but not practical or realistic for the universities to have these pieces of equipment in their schools or departments. This will allow students to have a better concept of automation in addition to all manual methods they perform at the university. With close attention to design, development and implementation of how to use and maintain various laboratory analysers' universities can increase the professional capabilities of students as well as new and experienced staff from rural and multi-skilled emergency laboratory health scientists that require this skill in the workplace.

By sharing the interactive 3D images of various instruments and explaining the principle behind each test, students can get a better understanding of daily maintenance and simple troubleshooting. Furthermore, by actually filming the instrument in action, as shown to new staff while training upon initial employment, it could help both the students and trainee scientists to gain a better understanding of the processes and acquire an overall representation of procedures prior to hands on experience.

Industry senior administrators, management and scientists-in-charge are excited about the idea and can see enormous benefit both for students work readiness and their new employees' standardized training. Utilisation of the facilities at the Gold Coast University Hospital was suggested as our students visit these labs in 1st year of the program and senior scientists can explain the whole process in each of the departments. Incorporating this virtual laboratory interactive sessions across various courses in 3rd year courses prior to clinical placement, students will be better prepared for placement and industry should find these graduates to be more work ready than other university graduate clinical scientists.

Issues, Controversies, Problems

There have been several issues that the teaching team came across during the development and incorporation of the e-learning technologies. Time was one of the major issues that the teaching and technical team had to face. Collaboration on projects can also be time consuming as different parties need to identify the project's objectives before identifying roles that are critical to driving the process. Development of the 3D virtual reality video was a collaborative project that required active participation from various departments such as the pathology team from the hospital, the teaching team from the University and the filming crews. Issues such as time constraints and communication between departments have been some of the major causes for the delay of the project. Furthermore the 3D virtual reality video is the representation of one of the most modern and fully automated laboratories however the equipment shown in the video cannot be universal as different laboratories depending on their geographical location, workload, patient population needs and financial capabilities may have different analysers performing the same tests. Similarly, the development of RPi's as cell counters has taken a lot of time to trial and program the software to ensure it is robust yet functional enough for student and staff use. Understanding how software programs work and what is required had to be communicated frequently between the software developer and haematology staff to perform certain functions and make the program user friendly.

The paradigm shift from face-to-face learning using mix mode learning is heavily based on technology and its consistency in delivery and accessibility are still institutional barriers that can decelerate learning potential. Accessibility of Aperio e-Atlas images can be improved by Image Scope, however data capacity and storage facilities of each slide image are yet to be resolved. Current limitations of WIS such as a lack of higher power magnification, image quality, digital slide storage and downloading massive file sizes over slow networks have proven to be an issue for

both students and staff. Institutions need to work alongside Information network Services (INS) to overcome online data storage and for students to be able to access and deliver data without having to use their own storage/ hard drive devices.

The main complications with the e-portfolio have been sharing their reflective journals with their respective academic and industry supervisors. Issues such as software incompatibility between various operating systems has inhibited the supervisors ability to monitor the student's progress in real-time. There has also been another issue, where by the assessment policy of university requires the same format of portfolio between students, however software incompatibility between operating systems does not allow that all the time.

Facility spacing and networking still remains an issue even with the adoption of RPi computers. RPi requires power outlets and networking ports in order to optimise cell counters and have the added features of printing and sending across real time e-portfolio entries. Laboratory facilities that do not support ethernet ports at every student station can propose a costly issue for institutions that would like to network their RPi. Alternatively by installing each RPi device using old monitors and PC2 compliant mouse and keyboards, can take up a reasonable amount of bench space that is predominantly taken up by other lab equipment such as water baths, microscopes and centrifuges.

Technology advances very quickly and to maintain or keep up to date with these changes requires constant updates and key accountabilities to continually support and upgrade systems, links and connections and can be a financial burden. With a range of advancements in e-learning technologies, studies have also shown that there is an increasing divide between the levels of access to the technological resources but also between the abilities of the user to benefit from these learning tools. And with an ever increasing access to IT, there has been a significant shift that has not only caused a digital divide but also a cognitive divide between students based on their level of literacy within the IT medium. One of the many

reasons for this divide can be the diversity in the rate by which technology is being incorporated into the early education system across different universities and countries (Cox, 2013).

SOLUTIONS AND RECOMMENDATIONS

Several strategies were devised to overcome the issues that were faced by the teaching team. Funding schemes can help start-up costs and take the load off elements and their financial burden. Sharing costs between a numbers of groups can also help reduce budget concerns and increase collaborative use of technology enhancing tools in delivering a number of courses. Action working parties have also been developed to help collaborative efforts to maintain good communication and keep everyone updated of their roles and responsibilities

For collaborative timetabling of the 3D virtual filming process, the filming crew was requested to start filming before the morning round of patient sample collection from the different wards because most of the start-up and maintenance on different equipment is performed early in the morning prior to early morning blood and sample collection in the industry. The student will be informed by annotation about the variation in equipment and changing technology that may lead to them working with different pieces of equipment. This will be reinforced by starting the presentation with an explanation of the same.

The students who are unable to share their reflective journals from the e-portfolio, due to different operating system or software variations with their respective supervisors will be requested to upload their assessment using Microsoft word as an attachment to e-portfolio. This is not an ideal solution but provides at least some support through occasional even though not regular feedback to students.

To combat limited laboratory space issues, the upgraded version of raspberry version Pi 3 which has a Wifi and Bluetooth capability as well as the

new Rpi touch screen monitors can provide us with a portable solution to ethernet cabling and reduce lab space usage. The implementation of a customized keyboard with copper or silver loaded plastic and antibacterial keys we can reduce the size and still be able to comply with laboratory health and safety standards.

As developments in technology continue to advance, processing powers of computers, cloud storage solutions and increases in data transfer speed will help ease the burden of slow data image transfers and increase the accessibility of slide images with better quality images. Similarly upgrading Aperio scanning imager to increase the number of slide sets being scanned can also reduce the predicament of time restricted technical staff.

Finally feedback from students and educators can help identify specific e-learning tools that are beneficial and how they can best be managed or improved.

FUTURE RESEARCH DIRECTIONS

Although there has been a significant raise in the usage of e-learning technology in various programs, the need to design and provide a more engaging and high quality content is crucial for creating an effective learning atmosphere for learners. This requires teaching and learning scholarship.

This article has discussed the innovative way in which a mixture of various technologies has been trialled. Some have been successful while others still require refining. However, it is also inevitable to face resistance to the implementation of e-learning technologies which requires changes to the traditional systems of educational institutions. It has been shown that the relationship between age and adoption of technology was mediated by cognitive abilities, computer self-efficacy, and computer anxiety (Czaja et al 2006). For this kind of change to be effective, evidence needs to be collected to demonstrate the favourable outcomes and the ease of implementation. This can be achieved

through survey based research showing successful incorporation of some of these technologies in teaching and learning of Health Science. The data collected from this research should also be used to convince the decision maker of universities and institutions to show the benefit of investing into multiple technologies for collaborative use.

CONCLUSION

In conclusion, e-learning is being increasingly embraced in tertiary education at numerous institutions and has gained approval from both the students and the teachers. Currently most of the e-learning tools are fragmented, incorporating just one or other technologies in class rooms. However, a standardised mix-mode of various online learning tools is required to cater for student from different socio-economic and educational backgrounds. At Griffith University the introduction of a mixed-mode of e-learning technologies including 3D virtual reality video, digital microscopy, Rpis, demonstration videos and e-portfolio has been demonstrated to be of effective educational value for students studying Medical Laboratory Science. In the future, studies should be performed to further evaluate the possible use of e-learning technologies in the education system and finding ways to convince the educators to embrace this mode of teaching and learning pedagogy. Finally, it is essential to standardise the use of education with maintenance of confidentiality and individuality to prepare graduates ready for global work force.

REFERENCES

- Ashwin, P. (2006). *Changing higher education: the development of learning and teaching*. New York: Routledge.
- Blamire, R. (2009). ICT Impact Data at Primary School Level: the STEPS approach. Assessing the effects of ICT in education, 199-211.

- Bradley, D., Noonan, P., Nugent, H., & Scales, B. (2008). *Review of Australian higher education: final report*. Canberra: Department of Education, Employment and Workplace Relations, Australia.
- Cela, K. L., Sicilia, M. A., & Sanchez, S. (2015). Social Networking analysis in E-Learning Environments: A preliminary Systematic Review. *Educational Psychology Review*, 27(1), 2219–2246. doi:10.1007/s10648-014-9276-0
- Clothey, R., Weidman, J. C., & Austin-Li, S. (Eds.). (2012). *Post-Secondary Education and Technology*. doi:10.1057/9781137037770
- Cox, M.J. (2013). Formal to informal learning with IT: Research challenges and issues for e-learning. *Journal of Computer Assisted Learning*, 29(1), 85–105. doi:10.1111/j.1365-2729.2012.00483.x
- Cummings, C., Mason, D., Shelton, K., & Bauer, K. (2014). Active learning assessment strategies for online and blended learning environments. In S. Keengwe & J. J. Agamba (Eds.), *Models for improving and optimizing online and blended learning in higher education* (pp. 58-82). Academic Press.
- Czaja, S. J., Charness, N., Fisk, A. D., Hertzog, C., Nair, S. N., Rogers, W. A., & Sharit, J. (2006). Factors predicting the use of technology: Findings from the centre for research and education on aging and technology enhancement (create). *Psychology and Aging*, 21(2), 333–352. doi:10.1037/0882-7974.21.2.333 PMID:16768579
- Dede, C., & Fontana, L. (1995). Transforming health education via new media. In L. Harris (Ed.), *Health and the new media: Technologies transforming personal and public health* (pp. 163-184). Academic Press.
- Gértrudix Barrio, M., Carmen Gálvez de la Cuesta, M. d., Álvarez García, S., & Valle, A. G. d. (2007). Design and Development of Digital Educational Content. In B. Fernández-Manjón, J. M. Sánchez-Pérez, J. A. Gómez-Pulido, M. A. Vega-Rodríguez, & J. Bravo-Rodríguez (Eds.), *Computers and Education: E-Learning, From Theory to Practice* (pp. 67-76). doi:10.1007/978-1-4020-4914-9_6
- Gibbons, M. (1994). The new production of knowledge: The dynamics of science and research in contemporary societies. *Contemporary Sociology*, 24(6).
- Gunn, C., & Harper, A. A. (2007). Using E-learning to Transform Large Class Teaching. In M. Bullen & D. P. Janes (Eds.), *Making the transition to E-Learning strategies and Issues*, (pp. 139-156). Academic Press.
- Guskey, T. R. (2003). How classroom assessments improve learning. *Educational Leadership*, 60(5), 6–11.
- Ioannou, N. K., Ioannidis, G. S., Papadopoulos, G. D., & Tapeinos, A. E. (2014). *A novel educational platform, based on the Raspberry-Pi: Optimised to assist the teaching and learning of younger students*. Academic Press.
- Kahn, G. (1993). Computer based patient education: A progress report. *M.D. Computing*, 10, 93–99. PMID:8469102
- Kirkwood, A., & Price, L. (2005). Learners and learning in the twenty-first century: What do we know about students attitudes towards and experiences of information and communication technologies that will help us design courses? *Studies in Higher Education*, 30(3), 257–274. doi:10.1080/03075070500095689
- Lencastre, J. A., & Coutinho, C. (2015). Blended Learning. In M. Khosrow-Pour (Ed.), *Encyclopedia of Information Science and Technology* (3rd ed.; pp. 1360–1368). doi:10.4018/978-1-4666-5888-2.ch129

Massy, W., & Zemsky, R. (2004). Thwarted Innovation? The Research Says What It Says. *Quarterly Review of Distance Education*, 5(4), xi–xii.

Pantanowitz, L., Farahani, N., & Parwani, A. (2015). Whole slide imaging in pathology: Advantages, limitations, and emerging perspectives. *Pathology and Laboratory Medicine International*, 7, 23–33. doi:10.2147/PLMI.S59826

Salmon, G. (2011). Flying not flapping: A strategic framework for e-learning and pedagogical innovation in higher education institutions. *Research in Learning Technology*, 13(3), 201–218. doi:10.1080/09687760500376439

Sangrà, A., Vlachopoulos, D., & Cabrera, N. (2012). Building an inclusive definition of e-learning: An approach to the conceptual framework. *International Review of Research in Open and Distance Learning*, 13(2), 145–159. doi:10.19173/irrodl.v13i2.1161

Smyth, R., Andrews, T., Caladine, R., & Bordujenko, J. (n.d.). *Rich Media Technologies: Opportunities in Australian Higher Education*. Academic Press.

Swann, W. (2013). The impact of applied cognitive learning theory on engagement with e-learning courseware. *Journal of Learning Design*, 6(1), 61–74. doi:10.5204/jld.v6i1.119

KEY TERMS AND DEFINITIONS

Cell Differentiation Counters: A counting tool used to count and tally different types of leukocytes.

Demonstration Video: A filmed video clip of a process or skill being performed for observation and review.

Digital Microscopy: The ability to view digital slides using a computer instead of a microscope.

E-Portfolio: A collection of electronic evidence assembled and managed by a user, usually on the Web. They are becoming standard practice for academics, students, and professionals and typically include examples of skills and achievements, as well as a reflective blog element.

Medical Laboratory Science: A four year undergraduate course in the field of medical science diagnostics and pathology.

Mix-Mode Learning: The use of technology to replace some of the face to face teaching with technology mediated learning tools.

Problem Based Learning (PBL): A student centred interactive learning pedagogy that challenges students to learn through engagement in a real life problem.

Raspberry Pi: A credit card sized computer that uses Linux Operating System.

Self-Directed Learning: Individuals perform their own study in an un-facilitated environment.

Simulations: A program or external setting that mimics an event or process, where participants can re-enact the process outside of a real-time event.

Tertiary Education: All post-secondary school education including but not limited to universities.

Category M

Medical Technologies

Defining and Characterising the Landscape of eHealth

Yvonne O'Connor

University College Cork, Ireland

Ciara Heavin

University College Cork, Ireland

INTRODUCTION

The need for reforming the delivery of healthcare services to accommodate the needs of modern societies is being experienced on a global scale. Government expenditure for the healthcare domain is increasing on a yearly basis (Bergamo, 2015). Population figures are growing with a reported increase of one billion people within the next 15 years, reaching 8.5 billion in 2030 (UN, 2015). Such a rapid increase in population is placing pressure on existing resources for healthcare (Keehan et al., 2015), thereby stressing the need for healthcare services to be reformed. Within the rapid rates of population growth, the fastest growing segment in the European population is the over 65 age group (Coale and Hoover, 2015), a group which utilises three to five times more healthcare services than their younger counterparts. Additionally, disease growth is on the increase as population rates incline. Despite the major progress in reducing death rates from various diseases, dramatic changes in global diseases have occurred whereby diseases have shifted to more lifestyle-related diseases (e.g. obesity, chronic obstructive pulmonary disease, type 2 diabetes and heart disease). Issues such as expenditure, population and disease growth enhance the need to reform the delivery of healthcare services. It is recognised that Information Systems (IS) is a critical success factor in the achievement of health care reform goals (Heeks, 2006) as the general use of IS globally is a powerful driver for change across

the healthcare industry. Advantages of successfully implementing IS within the healthcare domain are reported in literature. These include reduced time to diagnosis, improved access for patients to care, quality of life and patient satisfaction, enhanced healthcare provider decision making, increased accuracy in medical treatment and cost-savings (Bergamo, 2015).

Hospital Information Systems (HIS) were the main and earliest use of IS in healthcare (Reichert, 1975). Prior to 1960, electronic computers were not in general use in hospitals (Austin and Greene, 1978). From these early days HIS were defined as systems for “transmitting data among various personnel in different support units of the hospital” which ultimately seeks to “integrate all elements of information processing related to patient care into one system with a shared database” (Barnett and Zielstorff, 1974, p. 158). This definition of HIS has been used as a guide to defining the various eHealth technologies which subsequently emerged. However, since the 1970s HIS have evolved by incorporating other types of eHealth technologies, e.g. electronic medical records or electronic health records (EHR), electronic clinical decision support systems (CDSS), which has resulted in the varying definitions we observe in literature today.

Building from this, the objective of this paper is to define and characterise the landscape of eHealth technologies from an Information Systems (IS) perspective. The next section defines and characterises the concept of electronic Health (eHealth).

In doing so, various features and functionalities of eHealth technologies are reported. It is argued that the evolving nature of technological artefacts influence how the term eHealth is defined. In light of this, the concepts of eHealth integration and ontologies are subsequently discussed. Concluding this chapter is an overview of both academic and practitioner-orientated contributions and future eHealth research directions.

BACKGROUND

Although the term electronic health or ‘eHealth’ first emerged in the literature circa 1999-2000 (Eysenbach, 2001; Ahern et al. 2006) the use of IS within the healthcare system occurred well before then (Blum, 1984). According to Goldschmidt (2005) the first electronic health record was designed and deployed in the late 1960s-1970s, thirty to forty years before the term eHealth was coined. Since the turn of the 21st century the term eHealth has received attention in both academia and practice. Subsequently, eHealth has been defined by many people but a lack of evidence as to the existence of a comprehensive definition has emerged (Della Mea, 2001; Oh et al., 2005; Pagliari et al., 2005; Boogerd et al., 2015). To address this issue, journals such as “Journal of Medical Internet Research” published papers which explicitly sought to clarify what is meant by eHealth (cf. Oh et al., 2005; Pagliari et al., 2005) as part of its “What is eHealth” series leading to various reports about definitions concerning eHealth. For example for Ontario Hospital eHealth Council, eHealth is a consumer-centred model of health care where stakeholders collaborate, utilising ISs, including Internet technologies to manage health, arrange, deliver and account for care, and manage the health care system (Alvarez, 2002). The Internet clearly drives the development and adoption of eHealth applications; standing alone, it has the reach, the infrastructure, and the acceptance to achieve widespread change. Similar to this definition, Eng (2001) describes eHealth as the utilisation of emerging IS, especially the Internet, to improve or

enable health and healthcare. While many definitions for the term eHealth were proposed, it was the work of Della Mea (2001) which received the most attention as he described eHealth as a broad term that encompasses multiple domains (Boogerd et al., 2015). Over ten years later, and due to the rapidly changing field of eHealth, Boogerd et al., (2015) argues that a clear, uniform definition for eHealth still remains to be identified with many terms (such as Medicine 2.0, Health 2.0, mHealth) being used interchangeably.

For the purposes of this paper, eHealth is expressed as an umbrella term which encompasses three domains (i.e. business, clinical and consumer). According to DeLuca and Enmark (2000) business eHealth refers to systems focusing on financial and administrative transactions required to conduct the daily operations of healthcare whereas clinical eHealth refers to systems focusing on transactions which involve the collection, transmission and analysis of electronic health-related data by healthcare professionals. Consumer eHealth engage patients (i.e. consumers) in managing their own health care, to facilitate communication with providers and social networks, meeting their informational needs, making knowledgeable health decisions, using patient education resources, and promoting healthy lifestyles (Kreps and Neuhauser, 2010). Building from this, eHealth is defined as “*the use of any information technology by various stakeholders which transmits, stores, searches, and/or retrieves digitised data electronically for business, clinical and/or consumer purposes*” (adapted from De Luca and Enmark 2000; Della Mea, 2001). Further, the World Health Organisation (WHO) defines eHealth as “the transfer of health resources and health care by electronic means... encompasses three main areas: (1) The delivery of health information, for health professionals and health consumers, through the Internet and telecommunications; (2) Using the power of IT and e-commerce to improve public health services, e.g. through the education and training of health workers; (3) The use of e-commerce and e-business practices in health systems management. This

definition of eHealth is diagrammatically represented in Figure 1 and illustrates that eHealth (sometimes referred to in literature as Medicine 2.0 or Health 2.0) is an overarching term for the use of IT within the healthcare domain and comprises several subsections. Table 1 briefly synthesises fourteen eHealth subsections (i.e. technologies), associating them with at least one of the three domains. Note this is not an exhaustive list.

Information Systems (IS) research is typically characterised by the three pillars of People, Process and Technology (Earl, 1989). Figure 1 represents all three facets of IS identifying (1) the various stakeholders/*people* involved within the eHealth domain (i.e. government/policy makers, health insurers, software vendors/developers, patients/citizens, healthcare providers and NGOs); (2) business, clinical and consumer *processes* associated with eHealth and (3) sample *technologies* (i.e. desktop, mobile, wearable IT and 3D printing) used to collect, store, retrieve, and analyse medical-related data/information. Therefore, eHealth goes beyond technology to focus on further understanding the role of people and new models of care within the health care domain (Wyatt and Sullivan, 2005).

Table 2 presents some of the commonly referenced definitions by academics and practi-

Table 1. Various eHealth Technologies

eHealth Technologies	Domain
1. Clinical Decision Support System (CDSS)	Clinical
2. Computerised Physician Order Entry (CPOE)	Business
3. Critical Test Result Management (CTRM)	Clinical
4. Electronic Health Record (EHR)	Clinical/Consumer
5. Electronic Medical Billing (EMB)	Business
6. Electronic Medical Record (EMR)	Clinical/Consumer
7. ePrescription (eRx)	Clinical
8. Healthcare IS (HIS)	Clinical/Consumer
9. Medical and Patient Scheduling Software (MPSS)	Business
10. Picture Archiving and Communication System (PACS)	Clinical
11. Patient Portal (PP)	Consumer
12. Patient Health Records (PHR)	PHR
13. Real Time Location System (RTLS)	Business
14. Telehealth/Telemedicine (TH/TM)	Clinical/Consumer

tioners for well-known and widely used eHealth technologies. While a general consensus exists for defining some eHealth technologies there remains inconsistencies in defining EHR, EMR and PHR.

Figure 1. Overview of eHealth

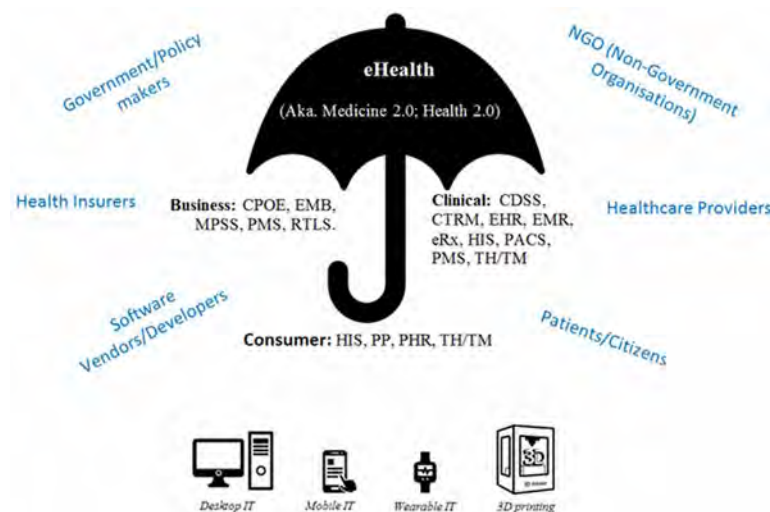


Table 2. Definitions of eHealth Technologies

Type	Definitions
CDSS	“any program designed to help healthcare professionals make clinical decisions” (Musen et al., 2001, p. 575) “any electronic system designed to aid directly in clinical decision making, in which characteristics of individual patients are used to generate patient-specific assessments or recommendations that are then presented to clinicians for consideration” (Lobach et al., 2012).
CPOE	“a process that allows a physician to enter medical orders directly and to manage the results of these orders” (Aarts et al., 2007, p. S4) “a computer application healthcare providers use to enter orders for medications, diagnostic tests, and ancillary services into a computer system. The EMR refers to a computerized record of patient data” (Dykstra et al., 2009, p.158)
CTRM	“also known as Critical Test Results Reporting, and Closed-Loop Reporting, is the software that handles a medical test result that has returned as critical to a patient’s health. It prevents the critical result from being lost in communication failures, improves patient safety, and documents the delivery of the results.” (MedicExchange, 2015)
EHR	“digitally stored healthcare information about an individual’s lifetime with the purpose of supporting continuity of care, education and research, and ensuring confidentiality at all times” (Lakovidis, 1998, p.106) “a paperless form of a patient’s medical record where HCPs [Healthcare providers] can enter and retrieve information utilizing a computerized system rather than writing in a paper chart” (Wallace et al., 2012, p.34)
EMB	.. allows interaction between payers/insurance companies and healthcare providers/organizations, managing payment for healthcare services (MedicExchange, 2015)
EMR	“an electronic health record that shares data only within its organization” (Bryne et al., 2010, p.637) “the health record used to gather, store, retrieve, and analyze medical information (Richards et al., 2012, p.121)” “synonymous with EHR” (Duncan, 2015, p.12)
eRx	“entering a prescription for a medication into an automated data entry system, and thereby generating a prescription electronically, instead of handwriting the prescription on paper” (Kilbridge, 2001, p.9) .. entire process of prescribing a medication in electronic format whereby the clinician prescribes medications which is then sent electronically to a pharmacy, and feedback comes back to the clinician when the patient collects the prescription (Cusack, 2008)
HIS	Includes all the elements that facilitate the capture, storage, processing, communication, security, and presentation of computer-based patient record information (Janczewski and Xinli Shi, 2002) “a comprehensive, integrated information system designed to manage the administrative, financial and clinical aspects of a hospital which encompasses paper-based information processing as well as computerised data processing” (Ismail et al., 2010, p.16).
MPSS	“provide intelligent scheduling system applications for healthcare practice, and let the care provider manage/schedule patient appointments through electronic preference cards” (MedicExchange, 2015)
PACS	..medical imaging technology which provides economical storage, rapid access to images from multiple modalities and simultaneous access to the same image at the one time (Choplin, 1992). “a computerised means of replacing the roles of conventional radiological film: images are acquired, stored, transmitted, and displayed digitally” (Strickland, 2000, p.82).
PP	... an internet-based interactive website for patients to communicate with their healthcare provider and with varied functions that gives them access to portions of their medical record and other services (Collmann and Cooper, 2007; Hess et al., 2006; Lin et al., 2005; Sorensen et al., 2009).
PHR	“provider-tethered applications that allow patients to access health information that is documented and managed by a health care institution” (Bourgeois et al., 2009, p.65). “is owned and controlled by the individual patient (or proxy), and may have information that is not contained in a medical record” (Emont, 2011, p.2).
RTLS	.. uses real time tracking to “enhance oversight on an array of hospital owned equipment/items. (MedicExchange, 2015).
TH/TM	“delivery of health care and sharing of medical knowledge over a distance using telecommunication system” (Thrall and Boland, 1998, p.145) “a service where clients’ vital signs are monitored remotely, at pre-set time intervals, using technology in their home” (Cartwright et al., 2011, p.6)

In order to fully appreciate how the eHealth technologies are defined (see Table 2), it is imperative to explore the characteristics (i.e. features and functionality) of the technologies. For the purpose of this study, we have focused on five key eHealth technologies namely, EHR, EMR, OHR, eRx and CDSS in Table 3. In order to identify the features and functionalities of the technologies the authors (i) reviewed three sources of the top ten eHealth technologies, (ii) identified five consistently named technologies across the lists, (iii) reviewed the selected technologies, and (iv) identified key features/functionalities. It is important to note that the core features and functions of eHealth technologies were captured as some technologies had 100+ features/functionalities.

Eng (2001) embraces 5 C's which seek to express eHealth functions and capabilities including: content, connectivity, community, commerce and care. Eysenbach (2001) identifies with 10 essential E's: efficiency, enhancing quality,

evidence-based, empowerment, encouragement, education, enabling, extending, ethics, and equity. Overtime, various features and functions have been incorporated within eHealth technologies which may account for the changing nature of defining eHealth technologies. Roberts and Grabowski (1990) refer to technological growth and argues that definitions for technological-related artefacts are not static but in fact evolving overtime. This may account for the varying definitions which have emerged for eHealth technologies over the last decade. It is further evident from Table 2 and from the definitions off the various eHealth technologies (Table 3) that there is a need to integrate the various features and functions of eHealth, as advocated by Eng since 2001.

Table 3. eHealth Technologies Features/Functionalities

eHealth Technologies	Basic Features	Basic Functions
EHR/EMR/PHR	• Patient management • Symptom management • Medication management • Note management • Results retrieval • Order entry • Report generator	• Identify and maintain a patient record • Manage patient demographics • Manage problem lists • Manage medication lists • Manage patient history • Manage clinical documents and notes • Capture external clinical documents • Present care plans, guidelines, and protocols • Manage guidelines, protocols and patient-specific care plans • Generate and record patient-specific instructions
eRx	• Patient management • Symptom management • Medication management • Order entry • Formulary management • Electronic Refills	• Identify and maintain a patient record • Manage patient demographics • Manage medication lists • Manage patient prescription history • Check drug interactions • Send and approve prescriptions
CDSS	• Patient management • Symptom management • Medication management • Reminder notification • Recommended diagnosis and treatment	• Manage patient demographics • Manage problem lists • Manage medication lists • Enhance diagnostic accuracy • Assist with clinical decision making • Improve patient safety

* Sources of Features/Functionalities

EHR/EMR/PHR: 1. Allscripts, 2. Athenaclinicals, 3. ClinicalWorks, 4. CareCloud, 5. PracticeFusion.

eRx: 1. Allscripts, 2. HealthFusion MediTouch, 3. NueMD, 4. PracticeFusion, 5. SureScripts.

CDSS: 1. EBMeDS, 2. IndiGO, 3. Mantis, 4. Persivia, 5. Zynx Health.

FUTURE RESEARCH DIRECTIONS

Integration has been the Holy Grail of Management Information Systems since the early days of computing in organizations. (Kumar and van Hillegersberg, 2000, p.23)

Integrated health systems are widely advocated to enhance performance in terms of quality and safety as a result of effective communication and standardised protocols (Suter et al., 2009). With the diffusion of the Internet and other information and telecommunication technologies, more integration has been enabled and demanded (Braa et al., 2007). In a healthcare context, Simoens and Scott (2005) argue that integration is integral to offering services from primary through tertiary care. The basis of integration is characterised by cooperation (O'Leary-Kelly and Flores, 2002), coordination, collaboration (Ettlie and Stoll, 1990; Kahn and Mentzer, 1998), information sharing (Pagell, 2004) and a fundamental shift away from managing individual functional processes, to managing integrated chains of processes (Akkermans et al., 1999). Building on this, it is therefore imperative that future eHealth developments move away from vertical silos to horizontal integrated systems (Adenuga et al., 2015) to ensure that a standardised delivery of care is provided (Simoens and Scott, 2005; Suter et al., 2009).

To facilitate a standardised and holistic delivery of health care services it is imperative that eHealth technologies are integrated across the three levels of delivery (Adenuga et al., 2015). These levels include primary care, secondary care, tertiary care (Reynolds, 1975). Primary care is defined as a set of universally accessible first-level contact, continuous, comprehensive, and coordinated care (Starfield, 1994) that promote health, prevent disease, and provide diagnostic, curative, rehabilitative, supportive, and palliative services (Lamarche et al., 2003). While, secondary care is a short term referral provided by informal consultations of secondary-care physicians with primary-care physicians (Starfield, 1994) to patients with symptomatic stages of a disease(s)

who require moderately specialised knowledge and technical resources for adequate treatment and/or rehabilitation (Reynolds, 1975). Often health services delivered at secondary care lead to referrals for tertiary care. According to Reynolds (1975), this function encompasses levels of diseases which are acutely threatening the health of the individual and require highly technical and specialised knowledge, facilities and personnel. Integration of eHealth technologies is important because of the highly interrelated nature of the functions that are performed across primary, secondary and territory levels in the provision of patient care (Austin and Greene, 1978).

Table 3 indicates that the eHealth technologies available within the healthcare system share common features and functionalities. In order for the seamless integration of these, eHealth developers need to embrace multiple ontologies (sometimes referred to as Medical Natural Language Processing) that can be combined to fully realise meaningful integration and analysis of data in the healthcare domain (Puri et al., 2011). An ontology, according to Stabb and Studer (2013), is a description (like a formal specification of a program) of concepts and relationships that can exist for an agent or a community of agents. The concept is important in terms of building a common language for the purpose of enabling knowledge sharing and reuse, which is integral for a standardised delivery of healthcare (Nardon and Moura, 2004).

The healthcare sector is criticised for the proliferation of terminologies across medical specialities, locales, health care facilities, and it also with clinical content (Shoniregun et al., 2010). Overcoming this challenge is reported to have numerous benefits for the healthcare sector. For example, the development and use of widely understood and accepted ontologies can build more powerful and more interoperable IS in healthcare (Kim and Chung, 2014) by facilitating the easy transfer and re(use) of patient-related data (Lucas and Bulbul, 2015). Moreover, it will promote a 'big data' environment which could be used to analyse trends/patterns in large data sets thereby

generating new insights in the medical domain (Dung and Kameyama, 2007; Bodenreider, 2008). Big data in healthcare denotes large and complex electronic health datasets which are perceived to be difficult (or impossible) to manage with traditional data management tools (Kołodziej et al., 2015).

According to Raghupathi and Raghupathi (2014, p. 2) 'big data' in a healthcare context amalgamates clinical data from CPOE and CDSS (physician's written notes and prescriptions, medical imaging, laboratory, pharmacy, insurance, and other administrative data); patient data in EHRs; machine generated/sensor data (e.g. vital sign monitoring), social media posts (e.g. Twitter) and web pages and less patient-specific information, including emergency care data, news feeds, and articles in medical journals. Therefore, integrated eHealth technologies with consistent ontologies will enable the creation of 'big data' for health care scenarios thus, enabling the discovery of associations and understanding of patterns and trends within the datasets for diagnosis, prognosis and treatment of illness and enhanced understanding of healthcare resource distribution (Arima, 2016).

CONCLUSION

There has been a surge of electronic health (eHealth) technologies encompassing a range of services available to various stakeholders within the health care system in both developed and developing countries. The advantages of successfully implementing eHealth technologies are reported in literature. Research indicates that these advantages significantly outweigh the heavily criticised costs associated with introducing such technological artefacts. As technology has evolved, the features and functionalities offered by eHealth technologies have grown dramatically leading to the development of many advanced technical solutions intended to serve various stakeholders (e.g. patients, clinicians, administrators). Additionally, the use of mobile telecommunications enables the use of eHealth at the point-of-care moving beyond its previously restricted desktop environments. Ex-

amples include Electronic Health/Patient Records, ePrescribing and Clinical Decision Support tools. As a result of this increased focus on eHealth, conflicting definitions for what constitutes the term eHealth have emerged in the literature. This lack of consensus is further inhibited by a dearth of research documenting the characteristics (i.e. features and functionalities) of such eHealth technologies. Given the nascent nature of eHealth research and the lack of definitional convergence, we define and characterise the landscape of eHealth technologies from an Information Systems (IS) perspective. Embracing this perspective provides a more detailed overview of the eHealth landscape. By examining existing literature and reviewing the market place, this paper reveals that there is a need to integrate the various features and functions of eHealth technologies thereby arguing that integration and interoperability is important for the growth of eHealth resulting in a more holistic approach to eHealth. Thus providing comprehensive coverage and up-to-date definitions of one of the most important issues, concepts, trends and technologies in information systems (IS) and information technology (IT); namely, eHealth.

There will be more changes in the future within the eHealth domain given how rapidly technology has evolved over the last two decades. With this in mind, eHealth definitions should be broad and flexible (i.e. modifiable with the changing components associated with new technologies). That being said, in defining eHealth it is essential that a common vocabulary (ontology) is developed and used to support the information needs of the various stakeholders, i.e. healthcare professionals and patients, while keeping standardisation and integration at the core of efficient and effective eHealth solutions.

Not only does this work contribute to the academic community, it also benefits practitioners in the decisions they make regarding eHealth technologies and the establishment of best practice in the form of eHealth standards and policies in the future.

REFERENCES

- Aarts, J., Ash, J., & Berg, M. (2007). Extending the understanding of computerized physician order entry: Implications for professional collaboration, workflow and quality of care. *International Journal of Medical Informatics*, 76, S4–S13. doi:10.1016/j.ijmedinf.2006.05.009 PMID:16798068
- Adenuga, O. A., Kekwaletswe, R. M., & Coleman, A. (2015). eHealth integration and interoperability issues: towards a solution through enterprise architecture. *Health Information Science and Systems*, 3(1), 1.
- Ahern, D. K., Kreslake, J. M., Phalen, J. M., & Bock, B. (2006). What is eHealth(6): Perspectives on the evolution of eHealth research. *Journal of Medical Internet Research*, 8(1). doi:10.2196/jmir.8.1.e4 PMID:16585029
- Akkermans, H., Bogerd, P., & Vos, B. (1999). Virtuous and vicious cycles on the road towards international supply chain management. *International Journal of Operations & Production Management*, 19(5/6), 565–581. doi:10.1108/01443579910260883
- Álvarez, R. (2002). The promise of e-Health—a Canadian perspective. *Ehealth International*, 1(1), 4. doi:10.1186/1476-3591-1-4 PMID:12459044
- Arima, H. (2016). Utilizing Big Data for Public Health. *Journal of Epidemiology*, 26(3), 105–105. doi:10.2188/jea.JE20160036 PMID:26947780
- Austin, C. J., & Greene, B. R. (1978). Hospital information systems: A current perspective. *Inquiry*, 15(2), 95–112. PMID:149767
- Barnett, G. O., & Zielstorff, R. N. (1977). Data Systems Can Enhance or Hinder Medical, Nursing Activities. *Hospitals*, 51(16), 157–161. PMID:914249
- Bergmo, T. S. (2015). How to Measure Costs and Benefits of eHealth Interventions: An Overview of Methods and Frameworks. *Journal of Medical Internet Research*, 17(11), e254. doi:10.2196/jmir.4521 PMID:26552360
- Blum, B. (1984). A framework for medical information science. *ACM SIGCSE Bulletin*, 16(1), 207–210. doi:10.1145/952980.808652
- Bodenreider, O. (2008). Biomedical Ontologies in Action: Role in Knowledge Management, Data Integration and Decision Support. *Yearbook of Medical Informatics*, 67–79. PMID:18660879
- Boogerd, E. A., Arts, T., Engelen, L. J., & van De Belt, T. H. (2015). “What Is eHealth”: Time for An Update? *JMIR Research Protocols*, 4(1).
- Bourgeois, F. C., Mandl, K. D., Shaw, D., Fleming, D., & Nigrin, D. J. (2009). *Mychildren's: integration of a personally controlled health record with a tethered patient portal for a pediatric and adolescent population*. Paper presented at the AMIA.
- Braa, J., Hanseth, O., Heywood, A., Mohammed, W., & Shaw, V. (2007). Developing health information systems in developing countries: The flexible standards strategy. *Management Information Systems Quarterly*, 31(2), 381–402.
- Byrne, C. M., Mercincavage, L. M., Pan, E. C., Vincent, A. G., Johnston, D. S., & Middleton, B. (2010). The value from investments in health information technology at the US Department of Veterans Affairs. *Health Affairs*, 29(4), 629–638. doi:10.1377/hlthaff.2010.0119 PMID:20368592
- Cartwright, C., Wade, R., & Shaw, K. (2011). *The Impact of Telehealth and Telecare on Clients of the Transition Care Program (TCP)*. Southern Cross University-Aged Services Learning & Research Collaboration. Available at: <http://www.cartwrightconsultingaustralia.com.au/library/The%20Impact%20of%20Telehealth%20and%20Telecare%20on%20Clients%20of%20the%20Transition%20Care%20Program%20Report%20-%20May%202011.pdf>

- Choplin, R., Boehme, J. M. II, & Maynard, C. D. (1992, January). Picture archiving and communication systems: An overview. *Radiographics*, 12(1), 127–129. doi:10.1148/radiographics.12.1.1734458 PMID:1734458
- Coale, A. J., & Hoover, E. M. (2015). *Population growth and economic development*. Princeton University Press.
- Collmann, J., & Cooper, T. (2007). Breaching the security of the Kaiser Permanente Internet patient portal: The organizational foundations of information security. *Journal of the American Medical Informatics Association*, 14(2), 239–243. doi:10.1197/jamia.M2195 PMID:17213500
- Cusack, C. M. (2008). Electronic health records and electronic prescribing: Promise and pitfalls. *Obstetrics and Gynecology Clinics of North America*, 35(1), 63–79. doi:10.1016/j.ogc.2007.12.010 PMID:18319129
- Della Mea, V. (2001). What is e-Health (2): The death of telemedicine? *Journal of Medical Internet Research*, 3(2), e22. doi:10.2196/jmir.3.2.e22 PMID:11720964
- DeLuca, J. M., & Enmark, R. (2000). E-health: The changing model of healthcare. *Frontiers of Health Services Management*, 17(1), 3. PMID:11184427
- Duncan, T. (2015). *An Examination of Physician Resistance Related to Electronic Medical Records Adoption*. Academic Press.
- Dung, T. Q., & Kameyama, W. (2007). *A proposal of ontology-based health care information extraction system: Vnhies*. Paper presented at the IEEE International Conference on Research, Innovation and Vision for the Future. doi:10.1109/RIVF.2007.369128
- Dykstra, R. H., Ash, J. S., Campbell, E., Sittig, D. F., Guappone, K., Carpenter, J.,... McMullen, C. (2009). *Persistent paper: the myth of “going paperless”*. Paper presented at the AMIA Annual Symposium.
- Earl, M. J. (1989). *Management Strategies for Information Technology*. Prentice Hall.
- Emont, S. (2011). *Measuring the impact of patient portals*. California Healthcare Foundation.
- Eng, T. R. (2001). *The eHealth landscape: A terrain map of emerging information and communication technologies in health and health care*. Princeton, NJ: The Robert Wood Johnson Foundation.
- Ettlie, J. E., & Stoll, H. W. (1990). *Managing the Design-Manufacturing Process*. New York. New York: McGraw-Hill.
- Eysenbach, G. (2001). What is e-health? *Journal of Medical Internet Research*, 3(2), e20. doi:10.2196/jmir.3.2.e20 PMID:11720962
- Goldschmidt, P. G. (2005). HIT and MIS: Implications of health information technology and medical information systems. *Communications of the ACM*, 48(10), 68–74. doi:10.1145/1089107.1089141
- Heeks, R. (2006). Health information systems: Failure, success and improvisation. *International Journal of Medical Informatics*, 75(2), 125–137. doi:10.1016/j.ijmedinf.2005.07.024 PMID:16112893
- Hess, R., Bryce, C. L., McTigue, K., Fitzgerald, K., Zickmund, S., Olshansky, E., & Fischer, G. (2006). The diabetes patient portal: Patient perspectives on structure and delivery. *Diabetes Spectrum*, 19(2), 106–110. doi:10.2337/diaspect.19.2.106
- Ismail, A., Jamil, A., Fareed, A. R., Abu Bakar, J. M., Saad, N. M., & Saadi, H. (2010). The implementation of hospital information system (HIS) in tertiary hospitals in Malaysia: A qualitative study. *Malaysian Journal of Public Health Medicine*, 10(2), 16–24.
- Janczewski, L., & Xinli Shi, F. (2002). Development of Information Security Baselines for Healthcare Information Systems in New Zealand. *Computers & Security*, 21(2), 172–192. doi:10.1016/S0167-4048(02)00212-2

- Kahn, K. B., & Mentzer, J. T. (1998). Marketings Integration with Other Departments. *Journal of Business Research*, 42(1), 53–62. doi:10.1016/S0148-2963(97)00068-4
- Keehan, S. P., Cuckler, G. A., Sisko, A. M., Madison, A. J., Smith, S. D., Stone, D. A., & Lizonitz, J. M. et al. (2015). National health expenditure projections, 2014–24: Spending growth faster than recent trends. *Health Affairs*, 34(8), 1407–1417. doi:10.1377/hlthaff.2015.0600 PMID:26220668
- Kilbridge, P., Gladysheva, K., Foundation, C. H., & Group, F. C. (2001). *E-prescribing*. California HealthCare Foundation.
- Kim, J., & Chung, K.-Y. (2014). Ontology-based healthcare context information model to implement ubiquitous environment. *Multimedia Tools and Applications*, 71(2), 873–888. doi:10.1007/s11042-011-0919-6
- Kołodziej, J., Correia, L., & Molina, J. M. (2015). *Intelligent Agents in Data-intensive Computing* (Vol. 14). Springer.
- Kreps, G. L., & Neuhauser, L. (2010). New directions in eHealth communication: Opportunities and challenges. *Patient Education and Counseling*, 78(3), 329–336. doi:10.1016/j.pec.2010.01.013 PMID:20202779
- Kumar, K., & Hilligersberg, J. (2000). ERP experience and evolution. *Communications of the ACM*, 43(4), 23–26.
- Lakovidis, I. (1998). Towards Personal Health Record: Current Situation, Obstacles and Trends in Implementation of Electronic Healthcare Record in Europe. *International Journal of Medical Informatics*, 52(1-3), 105–115. doi:10.1016/S1386-5056(98)00129-4 PMID:9848407
- Lamarche, P., Beaulier, M.-D., Pineault, R., Contandriopoulos, A.-P., Deis, D.-J., & Haggerty, J. (2003). Choices for Change: The Path for Restructuring Primary Healthcare Services in Canada. Canadian Health Services Research Foundation, New Brunswick Department of Health and Wellness, Saskatchewan Department of Health Ministère de la santé et des services sociaux du Québec, Health Canada.
- Lin, C.-T., Wittevrongel, L., Moore, L., Beaty, B. L., & Ross, S. E. (2005). An Internet-based patient-provider communication system: Randomized controlled trial. *Journal of Medical Internet Research*, 7(4), e47. doi:10.2196/jmir.7.4.e47 PMID:16236699
- Lobach, D., Sanders, G. D., Bright, T. J., Wong, A., Dhurjati, R., Bristow, E.,... Hasselblad, V. (2012). *Enabling health care decisionmaking through clinical decision support and knowledge management*. Academic Press.
- Lucas, J. D., & Bulbul, T. (2015). Ontology to Support Healthcare Facility Management. *Ontology in the AEC Industry*, 47.
- MedicExchange. (2015). Available at: <http://www.medicexchange.com/health-it.html>
- Musen, M. A., Shahar, Y., & Shortliffe, E. H. (2001) Clinical Decision Support Systems. In Medical Informatics. Computer Applications in Health Care and Biomedicine (2nd ed.). Springer-Verlag. doi:10.1007/978-0-387-21721-5_16
- Nardon, F. B., & Moura, L. A. (2004). Knowledge sharing and information integration in health-care using ontologies and deductive databases. *Medinfo*, 11(Pt 1), 62–66. PMID:15360775
- Oh, H., Rizo, C., Enkin, M., Jadad, A., Powell, J., & Pagliari, C. (2005). What is eHealth (3): A systematic review of published definitions. *Journal of Medical Internet Research*, 7(1). doi:10.2196/jmir.7.1.e1 PMID:15829471

- OLeary-Kelly, S. W., & Flores, B. E. (2002). The integration of manufacturing and marketing/sales decisions: Impact on organizational performance. *Journal of Operations Management*, 20(3), 221–240. doi:10.1016/S0272-6963(02)00005-0
- Pagell, M. (2004). Understanding the factors that enable and inhibit the integration of operations, purchasing and logistics. *Journal of Operations Management*, 22(5), 459–487. doi:10.1016/j.jom.2004.05.008
- Pagliari, C., Sloan, D., Gregor, P., Sullivan, F., Detmer, D., Kahan, J. P., & MacGillivray, S. et al. (2005). What is eHealth (4): A scoping exercise to map the field. *Journal of Medical Internet Research*, 7(1), e9. doi:10.2196/jmir.7.1.e9 PMID:15829481
- Puri, C. A., Gomadam, K., Jain, P., Yeh, P. Z., & Verma, K. (2011). *Multiple Ontologies in Healthcare Information Technology: Motivations and Recommendation for Ontology Mapping and Alignment*. Paper presented at the ICBO.
- Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. *Health Information Science and Systems*, 2(1), 3.
- Reichertz, P. (1975). *Hospital information systems, concept and implementation of the Medical System Hannover (MSH)*. Amsterdam: North Holland Publishing Company.
- Reynolds, R. E. (1975). Primary Care, Ambulatory Care, and Family Medicine: Overlapping but not synonymous. *Journal of Medical Education*, 50(9), 893–895. PMID:1152025
- Richards, R. J., Prybutok, V. R., & Ryan, S. D. (2012). Electronic medical records: Tools for competitive advantage. *International Journal of Quality and Service Sciences*, 4(2), 120–136. doi:10.1108/17566691211232873
- Roberts, K. H., & Grabowski, M. (1999). Organizations, technology and structuring. *Managing Organizations: Current Issues*, 159.
- Shoniregun, C. A., Dube, K., & Mtenzi, F. (2010). *Electronic healthcare information security* (Vol. 53). Springer Science & Business Media. doi:10.1007/978-0-387-84919-5
- Simoens, S., & Scott, A. (2005). Integrated Primary Care Organizations: To What Extent Is Integration Occurring and Why? *Health Services Management Research*, 18(1), 25–40. doi:10.1258/0951484053051951 PMID:15807978
- Sorensen, L., Shaw, R., & Casey, E. (2009). Patient portals: Survey of nursing informaticists. *Connecting Health and Humans. Proceedings of Nursing Informatics 2009*.
- Staab, S., & Studer, R. (2013). *Handbook on ontologies*. Springer Science & Business Media.
- Starfield, B. (1994). Is primary care essential? *Lancet*, 344(8930), 1129–1133. doi:10.1016/S0140-6736(94)90634-3 PMID:7934497
- Strickland, N. H. (2000). PACS (picture archiving and communication systems): Filmless radiology. *Archives of Disease in Childhood*, 83(1), 82–86. doi:10.1136/ad.83.1.82 PMID:10869010
- Suter, E., Oelke, N. D., Adair, C. E., & Armitage, G. D. (2009). Ten key principles for successful health systems integration. *Healthcare Quarterly (Toronto, Ont.)*, 13, 16.
- Thrall, J. H., & Boland, G. (1998). *Telemedicine in practice*. Paper presented at the Seminars in nuclear medicine.
- United Nations. (2015). *World Population Prospects The 2015 Revision*. Available at: http://esa.un.org/unpd/wpp/publications/files/key_findings_wpp_2015.pdf

Wallace, J. E., Friesen, S. P., White, D. E., Gilmour, J. G., & Lemaire, J. B. (2012). The Introduction of an Electronic Patient Care Information System and Health Care Providers' Job Stress: A Mixed-Methods Case Study. *Advancing Technologies and Intelligence in Healthcare and Clinical Environments Breakthroughs*, 33. World Health Organisation: E-Health. Trade, Foreign Policy, Diplomacy and Health. Available at: <http://www.who.int/trade/glossary/story021/en/>

Wyatt, J. C., & Sullivan, F. (2005). ABC of health informatics: eHealth and the future: Promise or peril? *British Medical Journal*, 331(7529), 1391–1393. doi:10.1136/bmj.331.7529.1391 PMID:16339252

KEY TERMS AND DEFINITIONS

eHealth Technologies: The umbrella term used to describe the wide variety of technologies, including internet based services, that support a health care system.

eHealth Technology Feature: Defines the purpose(s) of an eHealth system e.g. patient management, patient report generation.

eHealth Technology Functionality: An individual behaviour of an eHealth system e.g., identify and maintain patient record.

eHealth: The provision of patient-centric care and related services (i.e. people and processes) supported by technology.

Information Systems: The use of people, process and technology to get the right information to the right people at the right time.

Integrated Health Systems: Technologies that are characterised by their ability to cooperate, collaborate and communicate with other technologies.

Ontology: A set of agreed rules, usually in the form of a common vocabulary and relationships, used to enable the transfer of data/information/knowledge between two or more agents.

Kinect Applications in Healthcare

Roanna Lun

Cleveland State University, USA

Wenbing Zhao

Cleveland State University, USA

INTRODUCTION

Microsoft Kinect is one of the most popular inexpensive gadgets released in recent years. Kinect is equipped with a color camera, a depth camera, and a microphone array. The device allows users to interact with a computer via a natural user interface in terms of gestures or voice commands. So far, Microsoft has released three versions of Kinect sensors. The first two, typically referred to as Kinect for Xbox and Kinect for Windows, are rather similar except that Kinect for Windows is capable of doing near-mode tracking. Because both releases use the same depth sensing technology, they are collectively referred to as Kinect version 1, or Kinect v1 for short. The most recent Kinect sensor was released together with the new game console, Xbox One. Hence, it is sometimes referred to as Kinect for Xbox One. It uses a completely different depth sensing technology, which is based on time-of-flight measurement. As such, the new sensor is also referred to as Kinect version 2, or Kinect v2 for short. Kinect v2 has made drastic improvements over Kinect v1, including 1080p resolution for the color camera, and higher actual depth sensing resolution (the nominal resolution for the depth sensing resolu-

tion is actually lower than that for Kinect v1). Images of the two versions of Kinect sensors are shown in Figure 1.

The low-cost and the availability of Software Development Kit (SDK) for Microsoft Kinect \ has attracted many researchers to investigate its applications in many areas, particular in the healthcare realm. As the aging population rapidly grows in the United States, demands of healthcare services, especially physical therapy and rehabilitation services, have grown enormously in recent years. To meet the increasing demands and reduce the cost of services, providers are often looking for computers and other equipment that can assist them in providing services to patients in an affordable, convenient, and user-friendly environment. As a low-cost, portable, accurate, nonintrusive, and easily set up motion detecting sensor, Kinect enables researchers to develop computer-based vision control without using traditional input devices, e.g. mouse, keyboard, or joystick. This revolutionary technology makes it possible for Kinect to meet the challenge of providing high quality evaluations and interventions at an affordable price for healthcare services, as seen from the works surveyed in this chapter.

Figure 1. Two versions of Microsoft Kinect sensors. On the left is the Kinect v1. On the right is the Kinect v2.



DOI: 10.4018/978-1-5225-2255-3.ch511

BACKGROUND

Microsoft Kinect provides several streams of information to a software developer. The most common streams include: (1) A stream of 2D color image frames; (2) A stream of 3D depth image frames; and (3) A stream of 3D skeletal frames for at least one human subject in the view. A skeletal frame may contain the 3D position information for various number of joints. The availability of the skeletal frames has greatly facilitated Kinect application development because it frees the application developers from dealing with the complicated task of human pose estimation.

In Kinect v1, the depth of each pixel is calculated via a structured light method, which enables the use of a single infrared (IR) emitter and a single depth sensor to calculate the depth of each pixel. While this is a very clever scheme, the fidelity of the depth measurement is quite low because for the depth sensing to work perfectly, there has to be a

visible unique pattern for each pixel. Because there has to be some space between two adjacent dots as part of the structured light and this space has to be wide enough for the depth sensor to distinguish, only about 1 in every 20 pixels has a true depth measurement in typical situations and the depths for other pixels must be interpolated. Hence, the depth sensing resolution is actually significantly below the nominal 640x480 for Kinect v1. The depth-sensing technology used in Kinect v2 is completely different and the depth is calculated based on time of flight. The depth of each pixel can be calculated based on the phase shift between the emitted light and the redirected light.

Due to the maturity of the Kinect technology, developers can now publish Kinect v2 applications to the Windows store. Kinect v2 SDK also provides oial support for Unity, which is a development platform for 3D games. Finally, Kinect v2 SDK provides a tool to develop gesture recognition based on machine learning.

Table 1. A summary of Kinect applications in healthcare.

Application Domain	Specific Applications	References
Physical Therapy and Rehabilitation	Kinect-based physical rehabilitation systems	Chang et al., 2011; Rahman et al., 2013; Saini et al., 2012
	Virtual reality based games for balance training and upper body rehabilitation	Lange et al., 2011; Gotsis et al., 2012
	Kinect based game for Alzheimer patients	Cervantes et al., 2012
	A hand rehabilitation system	Metcalf et al., 2013
	Providing feedbacks on the quality of exercises	Velloso et al., 2013; Su, 2013
	A guidance and monitoring system for rehabilitation exercises	Zhao et al., 2014a
Clinical Environment	Kinect-based systems for high sterile operation rooms	Johnson et al., 2011; Gallo et al., 2011; Bigdelou et al., 2012
Fall Detection	Fall motion detection using Kinect depth images	Mastorakis & Makris, 2014
	Fall detection based on randomized decision forest algorithm	Bian et al., 2012; Girshick et al., 2011, Shotton et al., 2013
	A comparison study on using Kinect and mark-based systems for fall detection	Obdrzalek et al., 2012
	Fall detection and abnormal event detection on stairways	Parra-Dominguez et al., 2012
	Overcoming occlusions for fall detection	Rougier et al., 2011
	Use of two Kinects for fall detection	Zhang et al., 2012

APPLICATIONS REVIEW

In this section, we review the applications of the Kinect technology in a number of healthcare domains. A summary of the applications is provided in Table 1. The detailed description of the applications is given in individual subsections.

Physical Therapy and Rehabilitation Applications

Kinect has been widely used in physical therapy exercises, sports rehabilitation processes, and post-surgery training. Research has proved that Kinect can be a viable physical rehabilitation tool. Compared to traditional motion tracking and capturing programs that use reflective markers attached to users' bodies, the Kinect technology enables the development of solutions with noninvasive natural user interfaces and with real-time visual feedback to users.

Using Kinect's interactive interface could get patients more motivated and engaged in physical activities, particular for regular and repetitive movements in physical rehabilitation exercises. Chang, Chen, and Huang (2011) developed an interactive Kinect-based system to assist therapists in rehabilitating students in public school settings. The system is able to detect the student's joint position, and record movement data to allow therapists to review students' rehabilitation progress quickly. It also includes an interactive visual interface to enhance students' motivation, interest, and perseverance to engage in physical rehabilitation. The result has demonstrated that Kinect's natural interactive visual computer interface can significantly attract more users. In the study of real-time gesture recognition, Bigdelou, Benz, Schwarz, and Navab (2012) concluded that most users are fond of the dynamic interaction with system. Applications using Kinect technology are overwhelmingly regarded as friendly visual interaction programs.

A research group at the University of Southern California Institute for Creative Technologies

(ICT) has successfully applied the Kinect device to the study of virtual reality simulation technology for clinical purposes. Through extensive evaluation, assessment, and analysis, the researchers at the ICT have shown that Kinect technology can make a major contribution to the quality of traditional intervention with computer-based training games specializing in mental health therapy, motor skills rehabilitation, cognitive assessment, and clinical skills exercises. Lange et al. (2011) developed a virtual reality simulation game used for balance training of adults with neurological injury. Their colleagues Suma et al. (2011) developed the Flexible Action and Articulated Skeleton Toolkit, which is middleware that facilitates integration of full-body control with virtual reality applications and video games using Kinect.

Focusing on cognitive rehabilitation for Alzheimer patients, Cervantes, Vela, and Rodríguez (2012) developed a video game that replaces the traditional techniques used by therapists with the natural interaction Kinect device. The interactive body-motion controlled game increases patients' motivation to participate in exercise. An interactive rehabilitation system for disabled children was developed by Rahman et al. (2013). This application allows activities to be synchronized with on-line virtual counterparts. Patients can follow the exercise at home in the absence of the therapist because visual guidance for correctly performing the exercise is provided. Gotsis et al. (2012) demonstrated a mixed reality game for upper body exercise and rehabilitation. The exercise protocol was adopted from an evidence-based shoulder rehabilitation program for individuals with spinal cord injury.

Metcalf et al. (2013) studied an application for real-time monitoring and assessing hand movement based on Kinect. Hand rehabilitation is often needed for stroke patients' recovery. Medical professionals require detailed and accurate measurement of the hands in order to assess progress and functional recovery. The system is capable of measuring finger joint kinematics in real time for home-based clinical rehabilitation

programs. This system is designed in two steps: 1) identify anatomical landmarks and classification between grip types; 2) calculate hand joint angles from a kinematic model. The initial phase detects two different modes: spread hand mode and pincer grip mode. Then the landmarks go through validation procedures first by using ground truth estimation between reviewers and the associated algorithms to assess the accuracy of landmark definition, and second by joint angles compared to a laboratory-based gold standard motion capture system and validated kinematic measurement technique. And on the other side, patients require real-time feedback on their hand posture if they participate in a virtual environment rehabilitation gaming platform. Although this novel technique can robustly track and accurately measure the movements of the hand, it only provides a limited number of gestures that are only suitable for a specific rehabilitation application.

In order to reinforce the quality of home-based real-time rehabilitation exercise, Su (2013) proposed an approach using the Dynamic Time Warping (DTW) algorithm and Fuzzy Logic for ensuring that exercise is performed correctly. First it allows a patient to perform a prescribed exercise in the presence of a professional. Then the patient's exercise is recorded as a base for evaluation of exercise at home. The outcomes of the evaluation can be used as a reference for the patient to validate his/her exercise and to prevent adverse events. A summary report of the outcomes may also be uploaded to a cloud setting for physicians to monitor progress and adjust the prescription. The DTW algorithm is used to compare two sequences that have different time lengths in order to determine the similarity between the standard and the patient's exercise movements. For the rehabilitation exercise, the trajectory and speed mainly are based on experience and subjective evaluation. Therefore traditional logic theory is not practical for evaluating the quality of rehabilitation exercise by trajectory and speed. In this application, Fuzzy Logic is employed to build a fuzzy inference of a physician's subjective evaluation.

Saini et al. (2012) presented an approach to providing stroke patients with instruction and guidelines to perform rehabilitation exercises. A “watchdog” developed in this system provides the restriction that patients cannot exceed the limited angle, and ensures that a patient who is exercising at home will not exceed the prescribed level or perform for a longer period than necessary. A centralized system stores the entire patient's data that will be used to analyze the patient's movements and activities. The biofeedback system allows the user to give feedback to the therapist, and the therapist will evaluate the patient's movement and give further instruction or direction. The biofeedback is displayed for the patient as well as the therapist. The main benefit for biofeedback is that patients can directly communicate with the therapist for further instruction for improving their movements. This strategy is beneficial for stroke patients because patients will do exercises correctly.

Recently, Zhao et al. (2014a) developed a Kinect-based system that could be used at patient's home. In addition to providing visual guidance to patients regarding how to perform a prescribed rehabilitation exercise correctly, the system is capable of assessing the quality of the exercise in realtime based on a set of correctness rules (Zhao et al., 2014b).

Clinical Environment Applications

A key technical challenge using Kinect in clinical environment, e.g. in a medical operating room, is how to recognize the gestures interactively, accurately, robustly, and effectively. Touch-free interaction with a computer makes the Kinect device suitable for clinical environments in highly sterile conditions (Johnson et al., 2011). Their study showed that the touch-free motion recognition Kinect device provides an appealing opportunity to control medical images or image-guided devices without using the hands. Some researchers investigated gesture recognition systems based on the Kinect technology to address needs in medical surgical rooms.

Gallo, Placitelli, and Ciampi (2011) developed an open source interface using static postures and dynamic gesture recognition techniques to explore CT, MRI, or PET images in operating room. Utilizing hand and arm gestures, this application is able to execute basic tasks such as image selection, zooming, translating, rotating, and pointing; and some complex tasks, e.g. manual selection for controlling medical digital images effectively. The interaction follows the event – state – action paradigm and it is developed specifically for fast operation at video frame rates. The limitation of this interface is that the number of postures and gestures used for recognition is fixed. The extension of the gesture control interface to enable interaction with immersive medical imaging environments must be addressed in future development.

Bigdelou et al. (2012) applied Kinect 3D gesture-based interface technology to an intra-operative medical image viewer in a surgical environment. They used skeleton input to recognize the gesture type and the relative poses within that gesture. The recognized gesture and poses can be used for an interface for controlling a medical image viewer. The training phase is used to define and customize gestures from skeleton images, and during the interaction phase, the actual gesture and the relative pose within the gesture is estimated by discriminative kernel regression mapping. The quantitative experiments show this method can simultaneously recognize several gestures to a high degree of accuracy. Again, the number of gestures recognized is limited to 16 due to computational complexity for human body movement.

Fall Detection Applications

The Kinect device has been used as a real-time fall monitoring and detection system for a wide range of falls, including backward, forward, and sideways; on stairways; in occlusion places; in hospital wards; and others.

In a study of detecting human body falling motion, Mastorakis and Makris (2014) adopted the use of Kinect depth images instead of skeleton

data to robustly and efficiently track body motion. In their approach, a fall is detected by analyzing the 3D bounding box's width, height, and depth, i.e., the width, height, and depth of the human posture. The algorithm is provided by OpenNI. This approach outperforms the articulated model, which requires significantly more computational power. Since a human fall is a fast activity, a high frame rate in real-time systems is required to avoid missing detections. Furthermore, this algorithm has proved stable and can even detect movement when half of the subject's body is occluded by surrounding objects. Bian, Chau, and Magnenat-Thalmann (2012) presented an algorithm to improve the accuracy of skeleton extraction by rotating the human body trunk based on the pose in the previous frame.

Obdrzalek et al. (2012) presented a method that the accuracy of body tracking to a great extent depends on the type of poses being observed, the distance from the sensor and possible occlusions. They conducted a data acquisition process from captured human poses using the marker-based motion capture system developed by Microsoft Kinect. The result shows that Kinect works with much denser but less precise 3D data comparing to the marker-based system. They also observed that the markerless skeleton tracking of Kinect depends solely on dense depth information and thus frequently fails due to occlusions. For physical exercise of the elderly population, the Kinect skeleton tracking struggles with occluding body parts or objects in the scene, e.g. a chair. The occlusion is indeed an open issue for the use of Kinect technology in healthcare applications.

In the research work of Parra-Dominguez, Taati, and Mihailidis (2012), a method particularly to detect falls and other abnormal events on stairs was proposed. This work can automatically estimate the walking speed and extract a set of features that encode human motion during stairway descent.

Ni, Nguyen, and Moulin (2012) developed a computer vision based system to prevent potential falls in hospital ward environments. This system is

able to automatically detect the event of a patient getting out of a bed. The nursing staff is alerted immediately to provide assistance once the getting up event is detected. The detection algorithm combines multiple features from multiple modalities via an MKL framework to achieve high accuracy and efficiency.

Rougier et al. (2011) introduced an approach to address occlusion for detecting human body falls. The method is based on human centroid height relative to the ground and body velocity. With the help of computing 3D personal velocity just before the occlusion, this method can accurately detect falls by measuring human centroid height, because the vast majority of falls end on the ground or near the ground.

Zhang et al. (2012) proposed a system that employs two Kinect sensors to detect human falls. The statistical method is based on information about how the person moves during the last few frames. The major contribution of this research work is that it collects all training data from a specific viewpoint, and then collects all the test data from a different viewpoint. They believe the new protocol is a useful approach for measuring the robustness of the system by spatial displacement of camera images. Stone and Skubic (2012) developed a system for capturing variation of stride-to-stride gait in home environments for elderly adults. By measuring the changes in gait, falls can be predicted. If the motion of joints of a specific body subject is detected in an unusual time sequence, the prevention message is generated as a caution.

FUTURE TRENDS

The algorithms used for processing images obtained from the Kinect device play a very important role for human body motion analysis in healthcare applications. In the study of automatically tracking and detecting abnormal events, Parra-Dominguez, Taati, and Mihailidis (2012) evaluated different

binary classification algorithms. Velloso, Bulling, and Gellersen (2013) suggested exploring more complex curve analysis techniques.

Although some existing applications already provide feedback for physical rehabilitation, there is still a huge range of possibilities of how this feedback may be improved (Velloso, Bulling, & Gellersen, 2013). The potential improvements could be multi-modal interaction, closer loop between the expert and the user, automated assessment, and so forth (Zhao et al., 2015).

Enhancing the entertaining and amusing elements of the system are possible for future improvement (Chang et al., 2011). It would make rehabilitation more enjoyable if more peers (or users) are engaged in rehabilitation activities simultaneously. Game-based exercises could increase patients' motivation for therapy (Saini et al., 2012). Continuously exploring 3D interaction techniques in virtual environments (Cervantes et al., 2012) will be the focus in the next generation for patients involved in physical exercise programs.

Home-based remote healthcare service might become the next generation for out-patient physical therapy and rehabilitation. With the development of advanced 3D sensors and cloud gaming software (O'Connor, Davy, & Jennings, 2013), it is very possible that more home-based rehabilitation training exercises will be available online (Rahman et al., 2013) or in a cloud server, which would be more convenient for professional therapists wishing to instruct, monitor, and assess patients' activities.

Most existing applications using Kinect technology in healthcare only apply to human upper body parts. The limited viewing range of the Kinect sensor has made it difficult to detect and track full body movements. In a study of automatically tracking and detecting abnormal events on stairs in elderly living centers, Parra-Dominguez, Taati, and Mihailidis (2012) discovered that tracking hip joint points from the skeleton frame in Kinect is more stable than tracking feet and ankles, due to the fact that feet and ankles are sometimes partially occluded from Kinect view. On the other hand, ap-

plications in healthcare services require full body motion analysis. Therefore, full-body detection is a critical challenge for the future development of Kinect applications in healthcare.

One of the main drawbacks of using Kinect skeleton in healthcare applications is that Kinect uses a very non-anthropometric kinematic model with variable limb lengths (Obdrzalek et al. 2012). In general postures, the variability of the pose estimation is about 10 cm. Future algorithms should address occlusions and self-occlusions, unconventional body postures, and use of wheel-chairs or walkers.

CONCLUSION

We believe that the research and development on using Kinect technology in healthcare will gain more momentum. The demand of Kinect-based applications is high, due to Kinect's low cost and portability, and its accurate and robust motion detection capability. In this chapter, we have surveyed the current applications of using the Kinect technology in healthcare. Furthermore, we outlined a number of open research issues that could overcome the limitations of the current Kinect technology.

REFERENCES

- Bian, Z. P., Chau, L. P., & Magnenat-Thalmann, N. (2012, December). Fall detection based on skeleton extraction. In *Proceedings of the 11th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry* (pp. 91-94). ACM.
- Bigdelou, A., Benz, T., Schwarz, L., & Navab, N. (2012, March). Simultaneous categorical and spatio-temporal 3d gestures using Kinect. In *Proceedings of the IEEE Symposium on 3D User Interfaces* (pp. 53-60). IEEE. doi:10.1109/3DUI.2012.6184184
- Cervantes, J. C., Vela, F. L. G., & Rodríguez, P. P. (2012, October). Natural interaction techniques using Kinect. In *Proceedings of the 13th International Conference on Interacción Persona-Ordenador* (p. 14). ACM.
- Chang, Y. J., Chen, S. F., & Huang, J. D. (2011). A Kinect-based system for physical rehabilitation: A pilot study for young adults with motor disabilities. *Research in Developmental Disabilities*, 32(6), 2566–2570. doi:10.1016/j.ridd.2011.07.002 PMID:21784612
- Gallo, L., Placitelli, A. P., & Ciampi, M. (2011, June). Controller-free exploration of medical image data: Experiencing the Kinect. In *Proceedings of the 24th International Symposium Computer-Based Medical Systems* (pp. 1-6). IEEE. doi:10.1109/CBMS.2011.5999138
- Girshick, R., Shotton, J., Kohli, P., Criminisi, A., & Fitzgibbon, A. (2011, November). Efficient regression of general-activity human poses from depth images. In *IEEE International Conference on Computer Vision* (pp. 415-422). IEEE. doi:10.1109/ICCV.2011.6126270
- Gotsis, M., Lympouridis, V., Turpin, D., Tasse, A., Poulos, I. C., Tucker, D., & Jordan-Marsh, M. (2012, March). Mixed reality game prototypes for upper body exercise and rehabilitation. In *Proceedings of the IEEE Virtual Reality Workshops* (pp. 181-182). IEEE. doi:10.1109/VR.2012.6180940
- Johnson, R., O'Hara, K., Sellen, A., Cousins, C., & Criminisi, A. (2011, May). Exploring the potential for touchless interaction in image-guided interventional radiology. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 3323-3332). ACM. doi:10.1145/1978942.1979436
- Lange, B., Chang, C., Suma, B., & Newman, A. Rizzo, & Bolas, M. (2011). Development and evaluation of low cost game-based balance rehabilitation tool using the Microsoft Kinect sensor. *Annual International Conference of IEEE Engineering in Medicine and Biology Society, EMBC, vol. 2011*, (pp. 1831–4).

- Mastorakis, G., & Makris, D. (2014). Fall detection system using Kinects infrared sensor. *Journal of Real-Time Image Processing*, 9(4), 635–646. doi:10.1007/s11554-012-0246-9
- Metcalf, C. D., Robinson, R., Malpass, A. J., Bogle, T. P., Dell, T. A., Harris, C., & Demain, S. H. (2013, August). Markerless Motion Capture and Measurement of Hand Kinematics: Validation and Application to Home-Based Upper Limb Rehabilitation. *IEEE Transactions on Bio-Medical Engineering*, 60(8), 2184–2192. doi:10.1109/TBME.2013.2250286 PMID:23475333
- Ni, B., Nguyen, C. D., & Moulin, P. (2012, March). RGBD-camera based get-up event detection for hospital fall prevention. In *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference* (pp. 1405–1408). IEEE. doi:10.1109/ICASSP.2012.6287947
- O'Connor, C., Davy, A., & Jennings, B. (2013, February). Controlling the transfer of Kinect data to a cloud-hosted games platform. In *Proceeding of the 23rd ACM Workshop on Network and Operating Systems Support for Digital Audio and Video* (pp. 55–60). ACM. doi:10.1145/2460782.2460792
- Obdrzalek, S., Kurillo, G., Ofli, F., Bajcsy, R., Seto, E., Jimison, H., & Pavel, M. (2012, August). Accuracy and robustness of Kinect pose estimation in the context of coaching of elderly population. In *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE* (pp. 1188–1193). IEEE. doi:10.1109/EMBC.2012.6346149
- Parra-Dominguez, G. S., Taati, B., & Mihailidis, A. (2012, October). 3d human motion analysis to detect abnormal events on stairs. In *3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), 2012 Second International Conference* (pp. 97–103). IEEE.
- Rahman, A., Qamar, A. M., Ahmed, M. A., Rahman, M. A., & Basalamah, S. (2013). Multimedia interactive therapy environment for children having physical disabilities. *Proceedings of the 3rd ACM conference on International conference on multimedia retrieval*, 313–314. doi:10.1145/2461466.2461522
- Rougier, C., Auvinet, E., Rousseau, J., Mignotte, M., & Meunier, J. (2011). Fall detection from depth map video sequences. In *Lecture Notes in Computer Science: Vol. 6719. Toward Useful Services for Elderly and People with Disabilities* (pp. 121–128). Springer Berlin Heidelberg. doi:10.1007/978-3-642-21535-3_16
- Saini, S., Rambli, D. R. A., Sulaiman, S., Zakaria, M. N., Shukri, M., & Rohkmah, S. (2012, June). A low-cost game framework for a home-based stroke rehabilitation system. In *Computer & Information Science (ICCIS), 2012 International Conference (Vol. 1, pp. 55–60)*. IEEE. doi:10.1109/ICCISci.2012.6297212
- Shotton, J., Sharp, T., Kipman, A., Fitzgibbon, A., Finocchio, M., Blake, A., & Moore, R. (2013). Real-time human pose recognition in parts from single depth images. *Communications of the ACM*, 56(1), 116–124. doi:10.1145/2398356.2398381
- Stone, E. E., & Skubic, M. (2012, May). Passive, in-home gait measurement using an inexpensive depth camera: Initial results. In *Proceedings of the 6th International Conference on Pervasive Computing Technologies for Healthcare* (pp. 183–186). IEEE. doi:10.4108/icst.pervasive-health.2012.248731
- Su, C. J. (2013, August). Personal rehabilitation exercise assistant with Kinect and dynamic time warping. *International Journal of Information and Education Technology*, 3(4), 448–454. doi:10.7763/IJiet.2013.V3.316

Suma, E. A., Lange, B., Rizzo, A., Krum, D. M., & Bolas, M. (2011, March). Faast: The flexible action and articulated skeleton toolkit. In *Proceedings of the IEEE Virtual Reality Conference* (pp. 247-248). IEEE. doi:10.1109/VR.2011.5759491

Velloso, E., Bulling, A., & Gellersen, H. (2013). MotionMA: Motion Modelling and Analysis by Demonstration. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1309-1318). ACM. doi:10.1145/2470654.2466171

Zhang, Z., Liu, W., Metsis, V., & Athitsos, V. (2012, November). A viewpoint-independent statistical method for fall detection. In *Proceedings of the 21st International Conference on Pattern Recognition* (pp. 3626-3630). IEEE.

Zhao, W., Deborah, D. D., Reinthal, M., Ekelman, B., Goodman, G., & Niederriter, J. (2015, June). Privacy-Aware Human Motion Tracking with Realtime Haptic Feedback. In *Proceedings of the IEEE International Conference on Mobile Services* (pp. 446-453). IEEE. doi:10.1109/Mob-Serv.2015.67

Zhao, W., Feng, H., Lun, R., Espy, D. D., & Reinthal, M. (2014a, June). A Kinect-based rehabilitation exercise monitoring and guidance system. In *Proceedings of the 5th IEEE International Conference on Software Engineering and Service Science* (pp. 762-765). IEEE. doi:10.1109/ICSESS.2014.6933678

Zhao, W., Lun, R., Espy, D. D., & Ann Reinthal, M. (2014b). Realtime Motion Assessment For Rehabilitation Exercises: Integration Of Kinematic Modeling With Fuzzy Inference. *Journal of Artificial Intelligence and Soft Computing Research*, 4(4), 267-285. doi:10.1515/jaiscr-2015-0014

ADDITIONAL READING

Bannach, D., Lukowicz, P., & Amft, O. (2008). Rapid prototyping of activity recognition applications. *Pervasive Computing*, 7(2), 22-31. doi:10.1109/MPRV.2008.36

Chen, Y., Baran, M., Sundaram, H., & Rikakis, T. (2011). A low cost, adaptive mixed reality system for home-based stroke rehabilitation. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC*, pp.1827-1830.

Da Gama, A., Fallavollita, P., Teichrieb, V., & Navab, N. (2015). Motor rehabilitation using Kinect: A systematic review. *Games for Health Journal*, 4(2), 123-135. doi:10.1089/g4h.2014.0047 PMID:26181806

Levinger, P., Zeina, D., Teshome, A. K., Skinner, E., Begg, R., & Abbott, J. H. (2015). A real time biofeedback using Kinect and Wii to improve gait for post-total knee replacement rehabilitation: A case study report. *Disability and Rehabilitation. Assistive Technology*, 1-12. PMID:26336875

Lun, R., & Zhao, W. (2015). A survey of applications and human motion recognition with Microsoft Kinect. *International Journal of Pattern Recognition and Artificial Intelligence*, 29(5), 1555008. doi:10.1142/S0218001415550083

Moeslund, T. B., & Bagers, F. (1999). *Computer Vision-Based Human Motion Capture—A Survey*. Waltham, MA: Academic Press.

Myers, C., Rabiner, L., & Rosenberg, A. (1980). Performance tradeoffs in dynamic time warping algorithms for isolated word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 28(6), 623-635. doi:10.1109/TASSP.1980.1163491

Palacios-Navarro, G., García-Magariño, I., & Ramos-Lorente, P. (2015). A Kinect-Based System for Lower Limb Rehabilitation in Parkinsons Disease Patients: A Pilot Study. *Journal of Medical Systems*, 39(9), 1–10. doi:10.1007/s10916-015-0289-0 PMID:26265237

Zhao, W., Espy, D. D., Reinthal, M., & Feng, H. (2014, December). A feasibility study of using a single kinect sensor for rehabilitation exercises monitoring: A rule based approach. In *Proceedings of the IEEE Symposium on Computational Intelligence in Healthcare and e-health* (pp. 1-8). IEEE. doi:10.1109/CICARE.2014.7007827

Zhao, W., Lun, R., Espy, D. D., & Reinthal, M. (2014, December). Rule based realtime motion assessment for rehabilitation exercises. In *Proceedings of the IEEE Symposium on Computational Intelligence in Healthcare and e-health* (pp. 133-140). IEEE. doi:10.1109/CICARE.2014.7007845

KEY TERMS AND DEFINITIONS

Dynamic Time Warping (DTW): An algorithm for measuring similarity between two temporal sequences that may vary in time or speed. DTW calculates an optimal match between two given sequences (e.g. time series) with certain restrictions. The sequences are “warped” non-linearly in the time dimension to determine a

measure of their similarity independent of certain non-linear variations in the time dimension. This sequence alignment method is often used in time series classification.

Fall Detection: Detection of unexpected falls.

Fuzzy Logic: A form of many-valued logic. Fuzzy logic deals with reasoning that is approximate rather than fixed and exact. Compared to traditional true or false values, fuzzy logic variables may have a truth value that ranges in degree from 0 to 1. Fuzzy logic has been extended to handle the concept of partial truth, where the truth value may range between completely true and completely false.

Kinect SDK: A software developing kit released by Microsoft for the development of 3D sensing applications.

Randomized Decision Forest (RDF): An ensemble learning method for classification and regression that can operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes output by individual trees.

Rehabilitation Exercises: Exercises prescribed by a physical therapist to enhance the recovery of a patient after a surgery or certain diseases.

Virtual Reality: It refers to a computer-facilitated environment that simulates physical presence and allows the user to interact with the virtual world.

Neuroscience Technology and Interfaces for Speech, Language, and Musical Communication

Dionysios Politis

Aristotle University of Thessaloniki, Greece

Miltiadis Tsalighopoulos

Aristotle University of Thessaloniki, Greece

Georgios Kyriafinis

AHEPA University Hospital, Greece

INTRODUCTION

Recently, in a continuously changing environment with a wide variety of trendy mobile or wearable devices, in an increasingly demanding market for additional benefits in favor of the end user, the success of a product depends on its ability to adjust to the needs and desires of the consumer. So, new concepts have emerged in spectacular pageantries for the amazed public eyes.

Within this scenic approach, the concept of Interface Design and Usability has come to surface not only for smartphones and tablets, which are definitely powerful computer devices, but also for a variety of wearable or implanted pieces of apparatuses that are in position to perform many of the functions of computing machinery.

Inevitably, the issue of Usability comes to surface, as it measures perhaps more clearly than any other factor, the ability of a product not only to affect the body but also the mind of the user. This seemed to be the situation experienced thus far only with advanced medical devices and prosthetics that enabled monitoring of subtle neurological functionality. Indeed, for diagnostic or therapeutic purposes microensors, wearable monitoring systems and various imaging assemblages provide an enhanced telecommunication channel between the actual patient organs and the monitoring clinician (Figure 1).

DOI: 10.4018/978-1-5225-2255-3.ch512

The initial success of implanted or portable devices used for health and fitness reasons has widely promulgated mobile device applications of all types that take advantage of the affordable imaging modalities, the immediate reporting potential, the endoscopic representations of computer generated signals, the tactile sensing, the microrobotic precision, the motion tracking, the stereotactic functionality, the GPS positioning, the accelerometer measuring of shock, vibration or orientation, and many others.

As microprocessor signal pins interface with the internal neurological architecture of the human organism, they synchronize external events with the structure and functionality of the brain. The notion of Usability, extensively used in Human Computer Interaction (HCI), refers to the potency of a product that is massively deployed to end users to bring into action specific targetable derivatives under certain circumstances (Dix et al, 2004). These are measured in terms of:

1. Effectiveness, i.e. the degree to which interaction with specific computing machinery is successful in producing desirable outcomes for particular activities or purposes.
2. Efficiency, that is the design of interaction in such a way so that the ratio of useful work performed to the total efforts attempted is maximized as possible.

Figure 1. Wearable and implanted devices used for detecting and ameliorating oral and aural communication. Left: video Frenzel goggles that eliminate vision and provide simultaneous eye movement recording for the Dix-Hallpike maneuver that detects Benign Paroxysmal Positional Vertigo. Center: Otologic Carina™ cochlear implantation deployed within the inner ear. Right: EEG cap recording brain activity produced by electroencephalography.



3. Subjective Satisfaction, which results from the fulfillment of a user's expectations or needs while using components of electronic equipment, or the whole apparatus collectively.
4. A feature that recently has become a key element in the use of electronic devices is Learnability, which has to do with prior acquired knowledge or skill in using equipment of some kind, whether this ability has been developed by study, teaching, experience or sociability. The last factor has emerged as a new attribute of computing experience out of the personal relation with social networking.

Although the first three factors give a more mechanically driven approach to the evaluation processes of HCI, it is the fourth element that has gained momentum in terms of business intelligence. Indeed, Learnability is crucial for consumer electronics, since users flock around products that have a User Interface (UI) metaphor, a paradigm or an idiom they have previously used (Rogers et al., 2011).

By this attitude, they commit much less burden into their memory and as a result they prefer modules that have well-established UI paradigms, as is for instance the Android operating environment and the iOS paraphernalia. Indeed, the smartphone market is a classic example of an advanced and multi potent computing environment that is used

globally, for really sophisticated and sometimes unconventional tasks that demand no prior formal education or training received (Nielsen & Mack, 1994; Nielsen & Budiu, 2012).

This fourth factor, when bad Interaction Design is endorsed, tarnishes the acceptance of new media accounting for them the reputation of being

- Infuriating
- Confusing
- Aesthetically unpleasant

It is exactly this attitude towards Interaction Design that makes the commercial products that come up from the medical aids arena to deploy not only a new aesthetic approach but also to seek a different endurance and a smart communication that co-exists within the user habits. Indeed, as UIs become more efficient and effective, they promote their usage as Brain Computer Interfaces (BCIs) as well (Figure 2).

BACKGROUND

Although the general public's pictorial perception for a BCI is mingled between the image of an EEG cap (an electrode studded cap that reads brain activity) and a portable device like Oculus Rift, in recent proceedings a more subtle schematic view has been emerging with devices like

Figure 2. A series of trendy commercial devices with advanced HCI capabilities: Left, Google Glasses®. Center, the versatile iWatch™ calculating the heart rate and not only. Right, the Oculus Rift fully immersive audio-visual device in conjunction with hand held controllers. A new brave world for extreme Virtual Reality experiences.



Cochlear Implants that transform input activity to brain activity (Politis et al., 2014).

Generally speaking, there are two kinds of BCIs (Dix et al., 2004):

1. Invasive

This category of products, undoubtedly the flagship of Interfaces, produce overwhelming immersion into what brain activity projects. In audio matters, such devices offer state of the art multi-track digital input capabilities and thus they simulate and gradually replace the sound input of the outside world. It is obvious that the hardwiring of Figure 3, image on the right, indeed resembles the

beginnings of electromechanical Interfaces. Subtle operations and engineering methods take place as any malfunction or any unguided penetration may induct erroneously the electrode into the posterior semicircular.

2. Non-Invasive

Not so immersive, they need some training via biofeedback, and their bandwidth is rather limited. They act more like assistive devices and do not totally replace a handicapped organ or instrument.

The focus of this article lies on how commercial gadgets play the role of what instrumentation does to medical technology in terms of augmentation.

As body parts are failing to function properly and restoration to a healthy state seems inefficacious, prosthetics are used to rehabilitate vital bodily functions. In the same sense, gadgets readily available to the wide public can increase the data flow from the outside world to the brain and vice-versa. This consideration, albeit promoted with niche computing machinery, comprises the central vision for the already over \$1 billion Blue Brain Project (Markram, 2006) and the Human Brain Project in Europe (Waldrop, 2012) that promulgate medical informatics and neuromorphic computing. The equivalent project on the other side of the Atlantic is the US-financed Brain Initiative (2013). These approaches, which rely heavily on data analytics and simulations, attempt to promote brain-inspired Information Technology (Amunts et al., 2016). As a result, interdisciplinary trajectories are motivated linking Neuroscience, Brain Medicine and Computer Science in pursuit of the complex links amongst cerebral activity and behaviour (Markram et al., 2015; NSF, 2006).

At this point HCI emerges: it is the interdisciplinary field that correlates many disciplines like engineering, psychology, computer science, industrial design, ergonomics, artificial intelligence, and more recently, sociology, anthropology, art sciences, etc.

So it is the technological advances in these fields, as far as mobile, wearable and implanted devices are involved, that provide a new potential for the quality of the UI (Nielsen & Budiu, 2012). Also, a new workflow framework is introduced which further progresses the functionality of computing devices that support oral and aural communication. Under this prism, HCI becomes more central to the design and development of augmented reality machinery, investigating functionality that did not previously exist for the user or functionality that was not virtually usable.

In strict terms a Cochlear Implant (CI) is an electronic device that restores partial hearing to individuals with severe to profound hearing loss; these people normally would not benefit from a conventional hearing aid. Conventionally, as the

American Academy of Otolaryngology (2015) puts it “it is surgically implanted in the inner ear and activated by a device worn outside the ear. Unlike a hearing aid, it does not make sound louder or clearer. Instead, the device bypasses damaged parts of the auditory system and directly stimulates the nerve of hearing, allowing individuals who are profoundly hearing impaired to receive sound” (Figure 3).

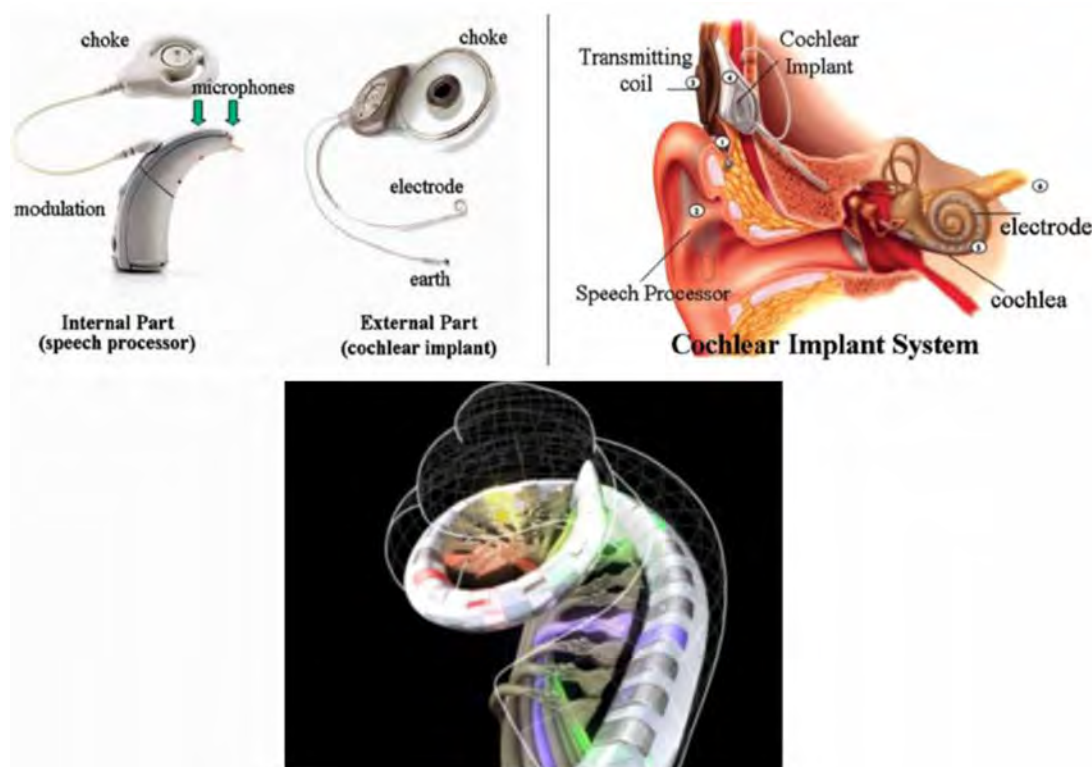
As stated earlier, the hardwiring of Figure 3, image far right, has a profound congruence with the origins of electromechanical Interfaces. Subtle operations and engineering methods take place as any malfunction or any unguided penetration may induct erroneously the electrode into the posterior semicircular canal damaging the facial nerve, or creating an unsupervised electrode contact within the vestibule. The implant consists of antenna, magnet, receiver stimulator in titanium casing, electrode array with electrode contacts, apical and basal, spaced at 2.4 mm intervals. The overall length of the implantable electrode array is 30 mm and its width 1.3 mm at the base and 0.8 mm at its apical end (Godara & Nikita, 2013).

Conclusively, a new generation of rather expensive gadgets transforms vigorous everyday activities: Google Glasses® consolidate with normal glasses, smart watches like iWatch™ display apart the others time and can dip into water, while contemporary fully immersive CIs disguise well with the scalp characteristics and allow their user even to swim. Either for medical purposes or for normal every day activities, computer aided infrastructures assists human senses to manipulate higher dimensional objects, which in mathematical terms leads to evading hyperspaces while in more pragmatic computer terms this directs users to virtual reality augmentations.

FROM PROSTHETICSTO BIONICS

Hearing disorders is the first pivotal point in disrupting the pipeline of proper oral and aural communication (Figure 4). They are quite common in

Figure 3. The Cochlear Implant pictorially: Left, the device, Cochlear's Nucleus Freedom™; Center: penetration and placement within the ear; Right, clinching the synapses of the audio nerve



modern societies not only due to congenital issues but also because diagnostic media are readily available. Hearing loss is perhaps the most common reason for visiting an ENT Department, both in the clinic and in the acute setting. As recent statistics reveal (Blackwell et al., 2014) approximately 15% of American adults (about 37.5 million) aged 18 and over, report some kind of hearing problem. 13% of the United States residents aged 12 or more (some 30 million people) have been diagnosed after standard hearing examinations to be experiencing hearing loss in both ears. More severely, a 2% of adults between 45 and 54 suffer from disabling hearing loss. As age increases, in the slot 55 to 64, this disability affects the 8.5% of the population. For those aged 65 to 74 the percentage rises to 25%, while for the ones who are 75 and older disabling hearing loss affects the 50% of that population group. According to these figures, hearing disorders impact adversely

the patient's quality of life by predominantly disrupting his oral-aural channel of communication (Stavrakas et al., 2016).

While hearing disorders directly affect the voicing mechanism, they have some smaller in extend negative influence on the way that musicians skillfully interact with their instruments in order to produce harmonic tones that bring to prominence beauty of form, harmony and emotion. Indeed, the distracted perception of musical sounds not only represses the ability of the performer to tacitly interact with his musical instrument, but is influencing the behavior of listeners who pervasively dip into oceans of digital music (Margounakis & Politis, 2011) supposedly delighting their ears. Actually, although the way that classically educated musician interact with computers has not changed in its essence the last 20-30 years, the general public has started to interact with computers, mobile devices and the

sound cloud in a two-way manner that enables computer aided synthesis and composition of new sounds, recording and mixing using smartphones and laptops, distribution through on-line libraries and repositories, (live) streaming services, and ubiquitous listening to music on smartphones and other portable music players (Tzanetakis, 2016).

Despite all these sweeping advances that raise the everyday audiophile to an enthusiastic producer, people who are hard of hearing lack the harmonic understanding or have impaired perception of natural tones struggle to participate in these social networks.

Hearing loss can be divided into two main categories, depending on the affected parts of sound transmission:

1. Conductive hearing loss results from any disruption in the passage of sound from the external ear to the oval window. It can be caused by pathologies involving the external and middle ear (external auditory meatus, tympanic membrane, ossicular chain).
2. Sensorineural hearing loss results from disruption of the passage of sound beyond the oval window. Such pathologies can be located to the auditory receptor cells of the cochlea and the eighth cranial nerve.

3. Mixed hearing loss represents a mixture of both conductive and sensorineural hearing loss.

Pure tone audiometry for producing audiograms is performed in an audiometric test room, where the subject's face should be clearly visible to the tester. When the test is observed from outside the audiometric test room the subject should be monitored through a window or a TV monitor. Excessive ambient noise can affect test results, thus it is recommended not to perform the test if the ambient noise is $>35\text{dB}$. Both ears are tested for air conduction firstly at 1000Hz and then at 2000Hz , 4000Hz , 8000Hz , 250Hz and 500Hz . A reference level of 0dB HL conventionally represents normal hearing across the entire frequency spectrum.

Some basic thresholds are the following:

- Threshold of hearing, 0 dB
- Whisper from 1m distance, 30 dB
- Normal conversation, 60 dB
- A shout, 90 dB
- Discomfort, 120 dB

The interpretation of the audiogram provides information not only for the quality of any potential

Figure 4. Left: the hearing process. Right: An Audiogram for normal hearing, in dB of Hearing Level versus frequency. air audiometry (o) and bone conducted audiometry (<) (Stavrakas et al., 2016)

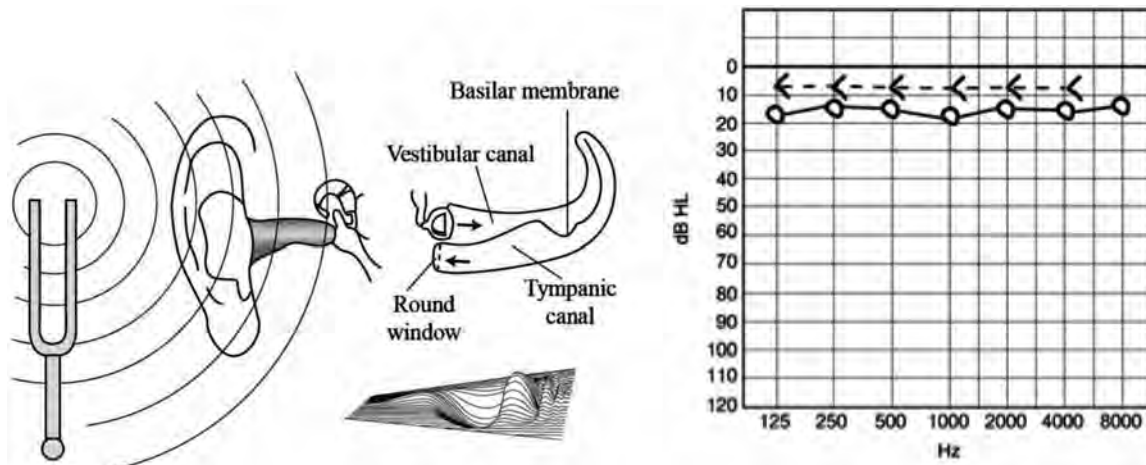
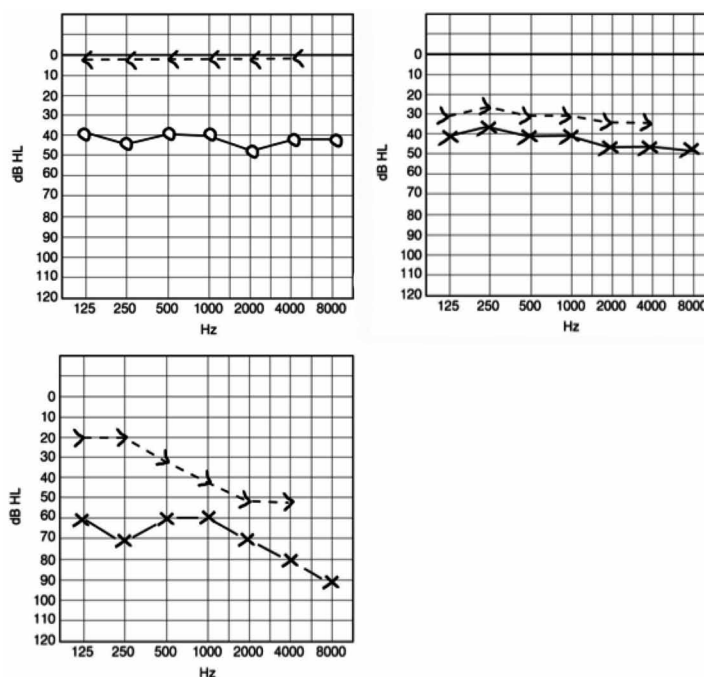


Figure 5. Left: Conductive hearing loss with an obvious air bone gap in its pure tone audiogram. Center: Sensorineural hearing loss audiogram. Right: Mixed hearing loss. All audiograms after Stavrakas et al., 2016.



hearing loss (conductive, sensorineural or mixed) but for the level of hearing loss as well. Generally, normal hearing is considered to be >20 dB, while mild hearing loss is >40 dB. Hearing thresholds and degrees of hearing loss are summarized pictorially in Figure 5 (Stavrakas et al., 2016).

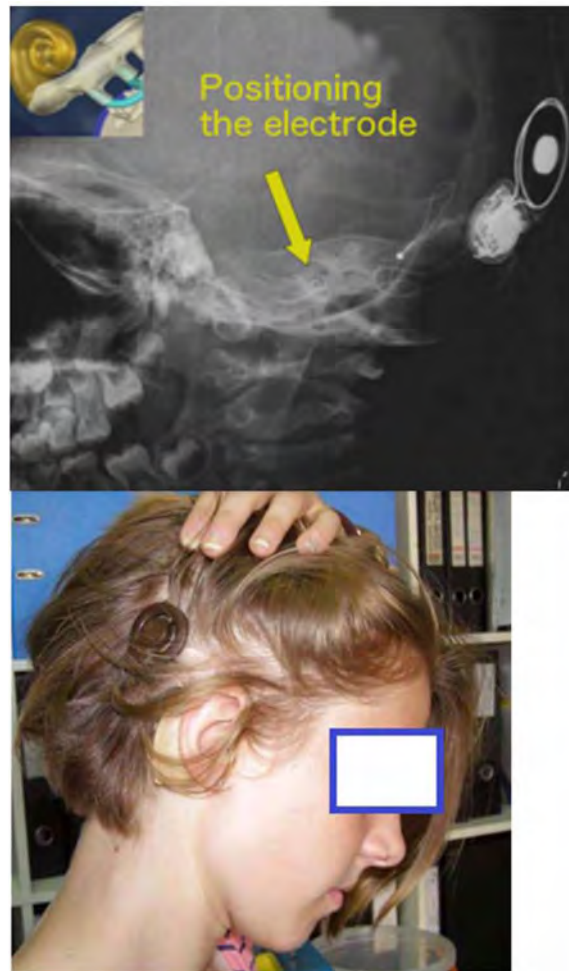
Unlike a hearing aid, CI does not make sound louder or clearer. Instead, the device bypasses damaged parts of the auditory system and directly stimulates the nerve of hearing, allowing individuals who are profoundly hearing impaired to receive sound” (Figure 6).

CIs are applicable for both children and adults with bilateral, severe to profound sensorineural hearing loss, who cannot take advantage of the use of powerful hearing aids and have not managed to improve their oral communication skills by a prescribed speech therapy. In this manner, the early stimulation of the acoustic Central Nervous System, especially in pre-school ages, may well lead to improved acoustic memory and sound discrimination.

Indications and preoperative requirements for cochlear implantation include a complete medical history and physical examination, medical valuation, audiologic examinations, CT and MRI scans to evaluate the cochlea and the auditory nerve, psychological tests, speech evaluation and enrolment in oral education program (Kyriafinis et al., 2007).

CIs do not restore normal hearing; they just provide a *representation* of sound. An inner ear problem, usually leads to sensorineural impairment or nerve deafness. In most situations, the hair cells are damaged and do not function. Even though various auditory nerve fibers may be intact and can transmit electrical impulses to the brain, these nerve fibers are unresponsive due to hair cell damage (American Academy of Otolaryngology, 2015). It is generally accepted that if severe sensorineural hearing loss cannot be put right with medical treatment, then the road towards cochlear implantation is wide open.

Figure 6. Applying a Cochlear Implant to a child at the AHEPA hospital, as seen with a CT head cross-section (left), and as it seen from outside (right), with an obvious attempt to camouflage the external parts of the implant.



FURTHER RESEARCH DIRECTIONS: ESTABLISHING A BCI MODEL FOR ORAL AND AURAL COMMUNICATION

It is obvious that cochlear implantation, along with highly immersive and invasive audiovisual and kinetic devices deploy a BCI that more or less revamps the Human Processor Model (Card et al., 1986) that was first established in mid 1980's by adding a blend of virtual reality sensing.

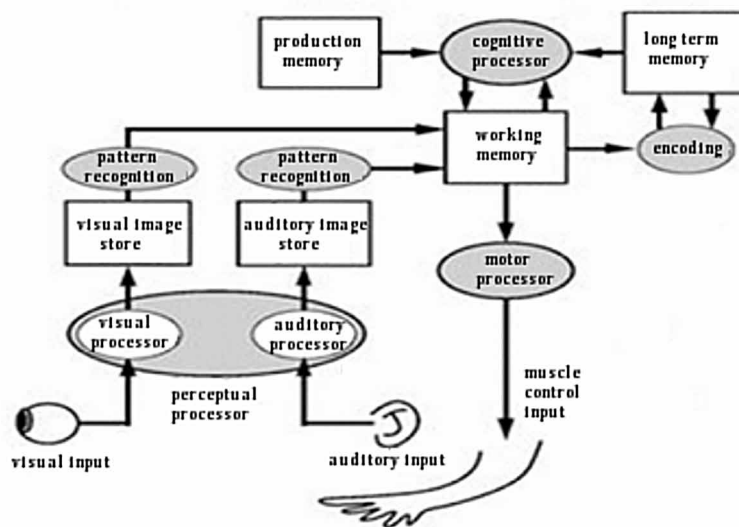
Indeed, the situation depicted in Figure 6 can be diagrammatically explained with the computer architecture module seen in figure 7.

The initial model proposed by Card, Moran and Newell, provided a workable simulation for analysing how human cognition responds in a HCI manner with the plethora of incoming signals that test the ability to see, hear, or become aware of something through the senses.

However, prosthetics and implanted devices offer a potential that did not exist in the original 1986 model, which could not envisage the possibility of desirable future characteristics (Pesaran et al., 2006), like an “audio processor” or a “graphics processor” apart from the Central Processing Unit.

This new approach propels theories that take into account special capacities that enable indi-

Figure 7. The diagram for the revamped Human Processor Model (cropped from <https://blombladivinden.wordpress.com>)



visuals to display exaggerated performances and skills, which cannot be explained in depth by the theories of intellectual performance based on a person's reasoning ability.

Prosthetics, bionics and augmentation devices allow subjects to readily experience an augmented virtual reality that gives them, although handicapped and impaired in normal physiological values, the possibility to experience the production of sensual impressions stimulated by "parts" of their body, or provoked by other senses.

For instance, a patient with a CI does not use the middle and the external part of his ear at all; instead, he is able to hear through an audio processor that establishes a sensorineural channel of communication between a computer device, the speech processor, its bionic extension, which is implanted within the skull, and a BCI, the place where the electrodes are conclusively settling and clinching the hair cells of the audio nerve (Figures 3 and 6).

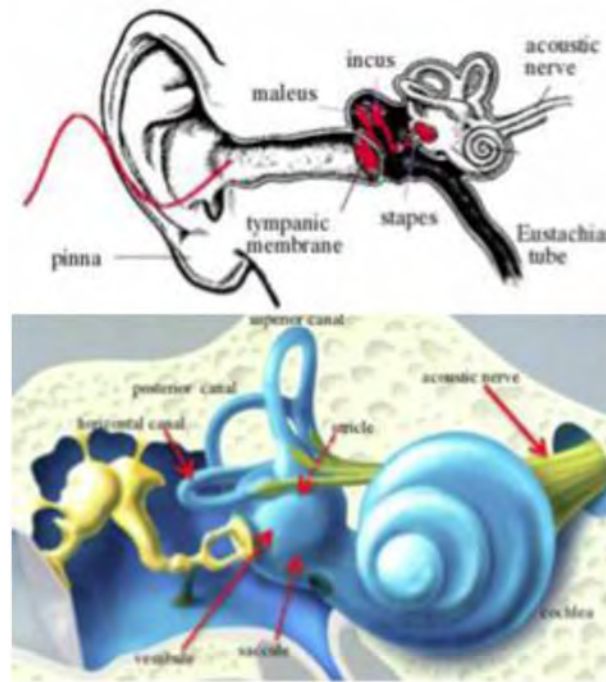
Indeed, all this arrangement is well described by the auditory processing channel seen in Figure 7. Furthermore, it may give the impression that the sense of hearing is completely replaced by a bionic part, and therefore the normal hearing ap-

paratus is not contributing to the bodily function of hearing.

Therefore, if the CI's battery is uncharged, or if the subject turns off the CI, then, he may not receive or process any auditory input. However, this is not the case. If, for instance, the subject is in a music hall and he is listening to dance music, which is highly rhythmic with strong subwoofer beats, he will not be able to hear the melody, but the strong beats via the ear's auditory canal will trigger with their vibrations the vestibulocochlear nerve, attesting the sense of balance. For a deaf subject, no sound or melody will be heard, since the bionic device is turned off, but, the strong vibrations will be sensed by the vestibule, sacule, superior canal and posterior canal, and therefore the subject will have a clear sense of the rhythm, allowing him to dance in accordance (Figure 8).

This ability, to have multiple representations of sounds, is stimulating the virtualization of our senses, by intermingling perceptual processors (which can be our senses or bionic devices) with the cognitive processor, which is our cerebral processing power, responsible for the integration of complex sensory and neural functions and the initiation and coordination of voluntary activity in the body.

Figure 8. Left, the hearing mechanism triggered in the outer and middle ear. Right, the structures that are triggered within the inner ear.



This approach can be somehow adequately incorporated with the theory of Multiple Intelligences (MIs).

The profiles of intelligences and their basic description defined by the MI theory are (Aleksic & Ivanovic, 2016):

- **Verbal/Linguistic:** Represents the primary means of communication amongst humans. It is reflected in symbolic thinking, language, reading, and writing.
- **Logical/Mathematical:** Is used for data processing, pattern recognition, working with numbers, geometric shapes.
- **Visual/Spatial:** Navigation, map making, visual arts, architecture, perspective.
- **Bodily/Kinesthetic:** Reflects the precise self-body motion control, non-verbal emotion expression, dance, fine hand-eye coordination.
- **Musical/Rhythmic:** Recognition and use of rhythmic and tonal patterns, recognition

of sound, speech and music instruments. It is used to interpret and create music.

- **Natural:** Recognizing patterns in nature, classification of objects and types of wildlife
- **Interpersonal:** The possibility of cooperation in small groups, communication with other people, individual's ability to recognize other people's intentions, mood, motivation, non-verbal signs.
- **Intrapersonal:** Recognizing own abilities, capacities, feelings, emotional reactions, self-reflection, and intuition.

Evidence of existence of a theoretical ninth (existential) intelligence has not been clearly established (Moran et al., 2006), so it will not be considered.

The theory of MIs has spread waves of enthusiasm in various educational circles since it promotes both an individualized approach and practical application in teaching. Researchers thus

recognize that by supporting multiple areas of activity they may enhance the students' mindset and better exploit it, especially amplifying its verbal and linguistic aptitude. MI theory increases learning efficiency by expediting instructors to direct further assistive education (Stagiopoulos et al., 2016).

Furthermore, it can be said that as sound or music experience is a complex procedure that involves multiple human sensors intermingling with virtual instruments or virtual worlds (Tzanetakis, 2016), that a more complex scheme should be used to represent logically, mentally and mathematically the multitude of HCI and BCI events.

Several systems have been proposed in various occasions, e.g. like that of Valbom & Macros (2003), Miranda & Wanderley (2006), Graham-Knight & Tzanetakis (2015), Politis et al. (2015)

that give an idea of the multitude of interaction with the world of music. New devices, new instrumentation and ample computer resources reshape the way that people either alone or en masse participate to the increasingly global music recreation (Figure 9).

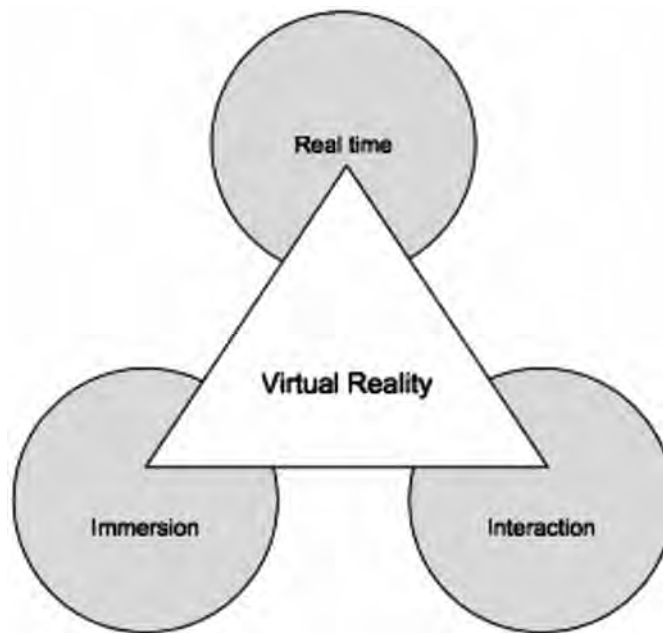
CONCLUSION

Playing a musical instrument, speaking properly a (foreign) language or being able to sing with the right tune are undoubtedly some of the most complex interactions for the convolution of oral and aural communication. All people have a lengthy experience of schooling for becoming able to handle properly their speech, language and musical communication skills. Furthermore,

Figure 9. Interacting with musical instruments. Left, a performer playing his guitar. Center, a virtual guitar with multitouch capabilities that can be played with an iPad. Right: A new concept for multisensor music reproduction – a “virtual” compact disc surface enabling the DJ to perform cueing effects with his hand.



Figure 10. The axes for developing augmented communication skills between prostheses, implants and wearable devices



the ability to manipulate the multiple aspects of linguistic ability (phonation, formation of speech, structural organization, etc) characterizes the overall competence of a subject's potential for his perceived general intelligence, at least for the schooling age.

A new perspective in the management of these capacities comes with the advent of devices that offer a bionic support and augmentation in proliferating oral and aural channels of communication.

Having already presented outstanding features in supporting communication with peers, they have commenced to demonstrate impressive features in their communication with the Internet, with digital technology equipment, with devices that make sound using music robotics and ultimately with wearable devices, portable devices or prosthetics that create an everyday virtual reality experience beyond expectation.

In any case, to manage the great expectations that have been created as far as handling, programming and interaction are involved, the three pivotal points exposed in Figure 10 should be adequately balanced.

REFERENCES

- Aleksić, V., & Ivanović, M. (2016). Psychometric evaluation of the reliability of IPVIS-OS multiple intelligences assessment instrument for early adolescents. *Journal of Educational Sciences and Psychology*, 6(68), 21-34.
- Amants, K., Ebell, C., Muller, J., Telefont, M., Knoll, A., & Lippert, T. (2016). The Human Brain Project: Creating a European Research Infrastructure to Decode the Human Brain. *Neuron*, 92(3), 574–581. doi:10.1016/j.neuron.2016.10.046 PMID:27809997
- American Academy of Otolaryngology — Head and Neck Surgery. (2015). *Cochlear Implants*. Author.
- Blackwell, D., Lucas, J., & Clarke, T. (2014). Summary Health Statistics for US Adults: National Health Interview Survey, 2012. *Vital and Health Statistics. Series 10, Data from the National Health Survey*, (260): 1–161. PMID:24819891

- Brain Initiative. (2013). *Fact Sheet*. The White House. Retrieved from <http://www.whitehouse.gov/the-press-office/2013/04/02/fact-sheet-brain-initiative>
- Card, S., Moran, T., & Newell, A. (1986). The Model Human Processor: An Engineering Model of Human Performance. In *Handbook of Perception and Human Performance: Cognitive Processes and Performance*, (vol. 2, pp. 1–35). Academic Press.
- Dix, A., Finlay, J., Abowd, G., & Beale, R. (2004). *Human Computer Interaction* (3rd ed.). Pearson-Prentice Hall.
- Godara, B., & Nikita, K. S. (Eds.). (2013). *Wireless Mobile Communication and Healthcare. Proceedings of the 3rd International Conference MobiHealth 2012*.
- Graham-Knight, K., & Tzanetakis, G. (2015). Adaptive Music Technology using the Kinect. *Proceedings of the 8th ACM International Conference on Pervasive Technologies Related to Assistive Environments. PETRA '15*. doi:10.1145/2769493.2769583
- Kyriafinis, G. (2005). *Cochlear Implantation*. Thessaloniki, Greece: Publish City.
- Kyriafinis, G., Vital, V., Psifidis, A., Constantinidis, J., & Nikolaou, A., Hitoglou-Antoniadou, M., & Kouloulas, A. (2007). Preoperative evaluation, surgical procedure, follow up and results of 150 cochlear implantations. *Hippokratia*, 11(2), 77–82. PMID:19582182
- Margounakis, D., & Politis, D. (2011). Music Libraries: How Users Interact with Music Stores and Repositories. In *E-Publishing and Legal Libraries – Legal and Organizational Issues*. IGI Global. doi:10.4018/978-1-60960-031-0.ch006
- Markram, H. (2006). The Blue Brain Project. *Nature Reviews. Neuroscience*, 7(2), 153–160. doi:10.1038/nrn1848 PMID:16429124
- Markram, H., Muller, E., Ramaswamy, S., Reimann, M. W., Abdellah, M., Sanchez, C. A., & Schürmann, F. et al. (2015). Reconstruction and Simulation of Neocortical Microcircuitry. *Cell*, 163(2), 456–492. doi:10.1016/j.cell.2015.09.029 PMID:26451489
- Miranda, E., & Wanderley, M. (2006). *New Digital Musical Instruments: Control and Interaction Beyond the Keyboard*. Middleton, WI: AR Editions.
- Moran, S., Kornhaber, M., & Gardner, H. (2006). Orchestrating multiple intelligences. *Educational Leadership*, 64(1), 22–27.
- National Science Foundation. (2006). A Report on Grand Challenges of Mind and Brain. NSF Publication.
- Nielsen, J., & Budiu, R. (2012). *Mobile Usability*. New Riders Press.
- Nielsen, J., & Mack, R. L. (Eds.). (1994). *Usability Inspection Methods*. New York: John Wiley & Sons. doi:10.1145/259963.260531
- Pesaran, B., Musallam, S., & Andersen, R. A. (2006). Cognitive Neural Prosthetics. *Current Biology*, 16(3), 77–80. doi:10.1016/j.cub.2006.01.043 PMID:16461265
- Politis, D., Piskas, G., Tsalighopoulos, M., & Kyriafinis, G. (2015). variPiano™: Visualizing Musical Diversity with a Differential Tuning Mobile Interface. *International Journal of Interactive Mobile Technologies*, 9(3).
- Politis, D., Tsalighopoulos, M., & Kyriafinis, G. (2014). Dialectic & Reconstructive Musicality: Stressing the Brain-Computer Interface. *Proceedings of the 2014 International Conference on Interactive Mobile Communication Technologies and Learning IMCL2014*. doi:10.1109/IMCTL.2014.7011126
- Rogers, Y., Sharp, H., & Preece, J. (2011). *Interaction Design: Beyond Human Computer Interaction* (3rd ed.). John Wiley & Sons.

Stagiopoulos, P., Politis, D., Piskas, G., Kyriafinis, G., & Tsalighopoulos, M. (2016). From Hearing Aids, Prostheses and Cochlear Implants to 'Bi-ionic' Feedback Phonation. *International Journal of Engineering Research and Applications*, 6(8).

Stavrakas, M., Kyriafinis, G., & Tsalighopoulos, M. (2016). Diagnosis and Evaluation of Hearing Loss. In D. Politis, M. Tsalighopoulos, & I. Iglezakis (Eds.), *Digital Tools for Computer Music Production and Distribution*. Hershey, PA: IGI Global. doi:10.4018/978-1-5225-0264-7.ch002

Tzanetakis, G. (2016). Natural Human-Computer Interaction with Musical Instruments. In D. Politis, M. Tsalighopoulos, & I. Iglezakis (Eds.), *Digital Tools for Computer Music Production and Distribution*. Hershey, PA: IGI Global. doi:10.4018/978-1-5225-0264-7.ch006

Valbom, L., & Macros, A. (2003). WAVE – AN Audio Virtual Environment. In *Proceedings of 2nd International Workshop on ICTs, Arts and Cultural Heritage - Digital Art Technologies, Applications & Policy*. Foundation of the Hellenic World, Cultural Centre (Hellenic Cosmos).

Waldrop, M. (2012). Computer Modelling: Brain in a box (H. Markram wants €1 billion to model the entire human brain. Skeptics don't think he should get it). *Nature*, 482(7386), 456–458. doi:10.1038/482456a PMID:22358809

ADDITIONAL READING

EmCAP. (2008). *Emergent Cognition through Active Perception*. FP6-IST project for Music Cognition (Music, Science and the Brain). Retrieved from <http://emcap.iaa.upf.edu>

Eshraghi, A., Nazarian, R., Telischi, F., Rajguru, S., Truy, E., & Gupta, C. (2012). The cochlear implant: Historical aspects and future prospects. *The Anatomical Record*, 295(11), 1967–1980. doi:10.1002/ar.22580 PMID:23044644

Odowichuk, G., Trail, S., Driessen, P., Nie, W., & Page, W. (2011). "Sensor Fusion: Towards a Fully Expressive 3D Music Control Interface", in *Proceedings of the Communications, Computers and Signal Processing 2011 IEEE Pacific Rim Conference (PacRim)*. doi:10.1109/PAC-RIM.2011.6033003

Wiley, M., & Kapur, A. (2009). "Multi-Laser Gestural Interface - Solutions for Cost-Effective and Open Source Controllers", in *Proceedings of the 2009 Conference on New Interfaces for Musical Expression NIME2009*.

Wilhelmsson, U. (2006). What is a Game Ego? (or How the Embodied Mind Plays a Role in Computer Game Environments). In M. Pivec (Ed.), *Affective and Emotional Aspects of Human-Computer Interaction*. The Netherlands: IOS Press.

Williamson, V. (2014). *You Are the Music: How Music Reveals what it Means to be Human*. UK: Icon Books Ltd.

KEY TERMS AND DEFINITIONS

BCI: The Brain Computer Interface research and development engulfs all the neuroprosthetics applications that focus in repairing, augmenting, and assisting human cognition or sensory-motor functions.

CT: Computed Tomography scans are special X-Ray tests that after being processed by computing machinery reconstruct detailed images of inner body structures and organs.

EEG: ElectroEncephaloGraphy is a monitoring methodology that gives concise measurements of the electrical activity of the brain while it performs cognitive tasks. Some 20 electrodes or more can be positioned on the skull to monitor small electric signals produced by brain functions. **GUI:** By the term Graphical User Interface is denoted the integral part of computing machinery that relies on graphic representations to carry out

commands. It also depicts visually the results of computer interaction establishing multimodal graphic communication between the smart device and the end user.

HCI: Human Computer Interaction is the scientific field that observes in depth how humans use all kinds of computing machinery, aiming to exploit the efficiency and effectiveness of their control potential by designing in cognitive and applicative terms new services.

MRI: Magnetic Resonance Imaging uses magnetic fields and radio waves to produce detailed images of the inner body. MRI may reveal problems that cannot be easily diagnosed by X-rays, ultrasounds or CT scans.

Senses: While computer research has put as goal to replicate the five senses, it becomes evident that machine perception can simulate, especially in gaming, some other sensations too, like vertigo, balance, equilibrium, acceleration, dizziness, magnetoception and so on.

UI: User Interface is the imminently responsive part of computable processes and structures. By it users can interact with a smart device.

VR: The synoptic term Virtual Reality refers to all the computer generated environments that create artificial worlds to feed our senses. Various devices like gloves, headsets, headphones, joysticks, eye cameras, 3D glasses, gamepads, motion controllers and similar offer to users simulated environments to “inhabit” and “wander in”.

Personalized Medicine



Sandip Bisui

Indian Institute of Technology (IIT) Kanpur, India

Subhas Chandra Misra

Indian Institute of Technology (IIT) Kanpur, India

INTRODUCTION

Electronic Medical Record (EMR) and Personalized Medicine (PM) systems utilize new strategies for genomic diagnosis to better deal with a patient's ailment or inclination towards a sickness. PM intends to accomplish ideal restorative results by helping physicians and patients pick the disease management approaches prone to work best in the setting of the patient's genomic profile and genetic data. Such methodologies might incorporate genetic screening programs that all the more precisely analyse sicknesses and their types offering physicians the right assistance with selecting the right treatment and most appropriate drugs best suited to that genomic group of people (PMC, 2010, PMC, 2014). EMR and PM systems consist of identifying nature and contribution of genes as well as different environmental factors and preparation of personalized drugs based upon the information derived. PM system can then facilitate disease prediction, prediction and treatment by determining whether an individual runs the risk of developing a disease. Thus PM system can develop early prevention strategies. It can dragonize disease faster. Therefore, treatment can be started early. Thereby, it is possible to prevent side effects that resulted from medicines given to a patient by trial and error method using the traditional system of treatment. Thus the main merits of PM system can be summarized as follows (PMC, 2007):

1. In this novel system, it is possible for the medical practitioners to prevent onset of different diseases more effectively.

2. It is more time and cost effective.
3. This treatment procedure bears the potential to reduce the possibility of adverse reaction of drugs.

However, it is to be borne in mind that for successful adoption of EMR and PM systems it is very important to first identify the critical challenges (Misra and Bisui, 2014), changes required (Misra et al., 2016), and trust, privacy and security concerns in the adoption. This act of identification will significantly add to the success of the personalized healthcare system management. As indicated above, the novelty of EMR and personalized healthcare lies in the fact that it makes use of information regarding environment, genes, proteins and clinical treatment of individual patients.

BACKGROUND AND MAIN FOCUS OF THE ARTICLE

Personalized Medicine is a recently developing idea in the modern health sector. Some of the personalized drugs have been now found by the researchers. However, reception of the Personalized Medicine and EMR idea still lacks practice implementations. Despite the fact that there are a several researches about on the genomic translation and genetic interpretation by Esvelt and Wang (2012), Dreyfuss (2012), Zamft et al. (2012), Cong et al. (2013) and Mali et al. (2013) advancement of exploration on various parts of Personalized Medicine is very insufficient. Bolouri (2010) has talked about a few issues

DOI: 10.4018/978-1-5225-2255-3.ch513

regarding personalized medicinal services. Eysenbach (2001) and Ahern et al. (2006) gave some insights about electronic healthcare. In a survey Jadad et al. (2005) and Eysenbach and Diepgen (2001) gave some direction towards the electronic medicinal services framework. There is an active promotion of e-health across the world. Several developed countries including Canada, United Kingdom and United States have a long-term plan to implement EMR. Taiwan has already adopted EMR and EMR Exchange. Canada had the goal to have EMR for all its citizens by the time of 2015. United Kingdom is building integrated IT infrastructure and systems to transmit health information safely and efficiently. United States has invested billions to encourage physicians and hospitals to adopt EMR and other applications of healthcare Information technology (HIT). Thus, the active promotion of EMR and EMR exchange is one of the important goals of health policy across the world (MRI, 2004). However, when it comes to developing countries, the success rate of the EMR adoption is not very high. Although Taiwan has implemented it, but physicians and clinics are still in a dilemma when it comes to practice implications of EMR and EMR exchange (Hwang et al., 2012, Chang et al., 2009, Chen et al., 2010). Countries like India, Netherlands and Australia are still trying to implement EMR and E-Health system. There are several challenges, trust issues, privacy and security concerns in this new technology adoption. We will try to discuss these issues further in this article.

SOLUTIONS AND RECOMMENDATIONS

While older people are more prone to drug adverse reactions because they are more likely to have multiple ailments and treatment of all the health problems need to be addressed simultaneously. Among younger generation people too, treatment by using personalized medicine is becoming increasingly attractive after the detec-

tion of the effect of the genetic background on the disease susceptibility and response to treatments. Pharmaceutical industries also indicate that the introduction of Personalized Medicine will ease the healthcare management to a significant extent. Personalized drugs are getting popular day by day. Personalized Medicine drugs will allow companies to provide a wider variety of drugs, each targeted to a smaller and more specific category of patients. In this way, adverse reaction of traditional drugs in patients' body can be minimised. Researchers have identified some issues regarding the implementation of Personalized Medicine. For any new technical adoption several managerial needs to be accessed before practice adoption. For this new healthcare paradigm we need to identify the feasibility of the adoption in the current scenario. The critical challenges those may hinder the adoption also needs to be known. The changes are required to the current practice must be identified. Success factors of the new medicare system will give the healthcare institutions an insight about what needs to be done for a successful adoption. Several trust issues may be associated with the consumer, privacy and security of the genomic database will also be an important concern. Therefore, one need to have an insight about all these issues before actually goes for a practice adoption. In this article we have tried to portray some key aspects regarding the above issues. A brief discussion has been made on the previous researches. A direction for future researches has also been given in this article.

Feasibility of the Personalized Medicine System in the Present Scenario

There are various reasons to think whether Personalized Medicine is really needed. Often many medical practitioners and patients feel that it is just a marketing hype (Bolouri 2010). However, it is to be kept in mind that the world population is growing rapidly and older people require better healthcare support. Importantly, older people have

varied health history, and therefore, have much diversified healthcare needs. They are also more prone to drug adverse reactions because they are more likely to have multiple ailments and treatment of all the health problems need to be addressed simultaneously. Among younger generation people too, treatment by using personalized medicine is becoming increasingly attractive after the detection of the effect of the genetic background on the disease susceptibility and response to treatments. Thus personalized medicine is needed by a large proportion of people.

Challenges Involved in the Adoption of Personalized Medicine

Misra and Bisui (2014) recently explored the challenges for the adoption of Personalized Medicine system within the healthcare management. The authors conjointly argue systematic and elaborate discussion on many vital aspects of personalized health care system. Within the same academic article, the authors have dwelt on variety of problems that provide hindrance to the implementation of this electronic healthcare system. The author analysed the crucial challenges from the perspective of medical physicians and common people. They have mentioned that there is a significant difference in both prospects. Some challenges include,

Genomic analysis for all individuals, High initial cost with no clear Return on Investment (ROI), Current healthcare work place doesn't have the training to administer Personalized Medicine, Present structure of medical record system doesn't have the infrastructure for Implementing Personalized Medicine, Unavailability of personalized medicine, Personalized Medicine concepts and tools have not yet been incorporated in medical school criteria, Government agreeing for the payment of the genomic tests, Complex cases where Personalized Drugs haven't yet discovered, Unawareness of doctors about personalized medicine, Ever increasing cost of health care, Insurance companies agreeing for the payment of the ge-

nomics tests, Categorization of different genomic groups. However, common people also identified Categorization of different genomic groups and Complex cases where Personalized Drugs have not yet discovered as significant challenges in the adoption of Personalized Medicine.

Changes Required While Adopting Personalized Medicine

Several scholars have showed that there are several challenges that need to be taken care of before adopting Personalized Medicine system. Misra et al. (2016) gave us a glimpse of the changes required for this new medicare system to adopt. Their study showed that changes from reactive to efficient medical care, from trial and error to right treatment for right person at right time, from narrow mind-set to open mindedness, from open information of patients to secure information, from less emphasis on IT infrastructure to more emphasis on IT infrastructure are some of the significant changes that are necessary for implementing the revolutionary medicare system of personalized Medicine. Their research gave us more useful and more accurate information necessary for implementing Personalized Medicine systems in the future. The results from their study serve as first step in the direction.

Factors Leading to Successful Adoption of Personalized Medicine

For any technical adoption identifying the critical success factors is a key step. Once the organizations are aware about the factors that could lead to successful adoption, the manager could prepare a plan or pathway to overcome the success factors. Misra and Bisui (2014) while identifying the critical challenges gave a hint of the factors that may have an impact in the success of the Personalized Medicine system. A study in the similar area indicated that data management, team work and composition, privacy and confidentiality, mind-set, return on investment, sufficient time, R&D

and alignment could be possible success factors. However, there is not any academic paper in which an attempt has been made to suggest the critical factors for the implementation of Personalized Medicine in healthcare management. Researchers could identify a multiple regression model for the consolidated dependent variable success, with the dependencies success factors as regressor.

Consumer Trust, Database Privacy and System Security in Personalized Medicine System

While receiving any new innovation the customer trust is dependably a noteworthy concern. In Personalized Medicine framework keeping up the protection and defending the security concerns will likewise assume a noteworthy part in the achievement of the adoption. As there will be a genomic database containing the patients' medicinal and individual data, therefore, keeping up the trust of the consumers will be a noteworthy test. There will likewise be numerous security concerns in regards to the information protection and utilization. Consumers and healthcare provider professionals would likewise be worried about their protection while utilizing this framework. Along these lines, there may be several trust issues could hinder or delay the success in the adoption. Before effectively receive this new medicare framework one needs to address all the trust issues that might influence the accomplishment of this progressive and novel medicare framework. Researchers are trying to identify the issues that may regulate the trust of consumers, however, as per our knowledge there are no academic literature available in this domain. We have taken this issue up and trying to identifying the moderating effect of trust concerns in the success factors of Personalized Medicine. As trust issues are mostly related to people and organization (Robert and Seibold, 2003). Therefore we took the success factors that are directly or indirectly related to people and organization. A brief of our study on the identification of trust issues and their moderating effect on people

related success factors is given bellow. Factors related to consumer and hospital system authorities are identified. These factors are the important modifiers of success. Next trust concerns of the consumers are listed. Finally the moderating effect has been checked using statistical methods. Along with trust privacy is also a major concern in this novel technology adoption. In Personalized Medicine system there will be a genomic database available in the web. This database is going to be the core of the adoption and it will contain all the genomic and personal information of the patients. Therefore, who are accessing the database will be a major concern for the patient' community. The privacy of the consumers will be directly in the hand of the database user. Therefore, common people will be very concern about their privacy. To prevent the privacy breach there will be a strong need of proper technology. Any technical glitch may leak the privacy of the consumer. Therefore, before adopting the Personalized Medicine system there is a strong need of research in the privacy concerns and their moderating effect on the success of Personalized Medicine. Similar to trust issues and privacy concerns, security will also be an important concern to taken care of. The implementation of Personalized Medicine system and E-healthcare platform is going to be very costly. Therefore, there will be high chance of security breach. Healthcare institutions and government should focus on developing policies and regulations to prevent the security breach. A proper research is needed before the practice implementation of Personalized Medicine.

FUTURE RESEARCH DIRECTIONS

Personalized Medicine system is surely to be an important enabler of the E-health system. It could shift the whole e-healthcare platform to a new level, where healthcare cost will be less, treatment time will be minimized and there will be no error in the treatment (PMC, 2007). However, Personalized Medicine system still lags research. There is not

much academic study in the implementation of this novel technology. Misra and Bisui (2014, 2016) and Misra et al. (2016) have provided some business models for the critical challenges, feasibility and changes of this system. However, more work needs to be done in the area of trust, privacy, security concerns of the novel technology. Shalhoub (2006), Hwang et al. (2012) have identified different trust, privacy and security concerns for the EMR adoption. Although there are some research works available in the area of EMR adoption and E-health implementation, however, current academic still lags works in Personalized Medicine implementation. We have given a glimpse about the possible research works that can be done in this area. Personalized Medicine system surely is an important asset/discovery for our society. Therefore, researchers must have to be focused in its implementation. Several countries including United States, Canada, Australia, India, Taiwan have taken the implementation of EMR and Personalized Medicine system seriously. Therefore, these research works could be very helpful to the administrators and healthcare care providers while adopting this new medicare technology.

CONCLUSION

The research works in the area of adoption of Personalized Medicine serve as a vital step towards an essential comprehension of the rising field of Personalized Medicare framework. The conceivable purpose behind these research works is to minimize the current healthcare cost and provide our society a better and efficient healthcare system. Personalized Medicine will be a vital enabler in this. Adoption of EMR and electronic healthcare system has broadened the pathway for Personalized Medicine. Several authors and organizations Misra and Bisui (2014, 2016), Misra et al. (2016) Hwang et al. (2012), NHMRC (2011), Ginsburg and McCarthy (2001), Ong et al. (2013), PMC (2007, 2010, 2014) have started working in the practical implications of the Personalized Medi-

cine. The outcome of the research works could serve as a noteworthy step towards the adoption of Personalized Medicine. We may hope that soon there will be an era of Personalized Medicine in the healthcare sector. Physicians, healthcare administrations and government could use these results to make new policy and regulations for the successful adoption of Personalized Medicine.

REFERENCES

- Ahern, D. K., Kreslake, J. M., Phalen, J. M., & Bock, B. (2006). What is e-health? Perspectives on the Evolution of eHealth Research. *Journal of Medical Internet Research*, 8. doi:10.2196/jmir.8.1.e4
- Bolouri, H. (2010). Personal Genomics and personalized Medicine. Imperial College Press.
- Chan, Y., et al. (2005). Information Privacy: Management, Marketplace, and Legal Challenges. *Communications of the Association for Information Systems*, (16), 270–298.
- Chang. (2009). Factors affecting cross-hospital exchange of Electronic Medical Records. *Information & Management*.
- Chen, Y.-H., Liu, C.-F., & Hwang, H.-G. (2010). Key factors affecting healthcare professionals to adopt knowledge management: The case of infection control departments of Taiwanese hospitals. *Expert Systems with Applications*.
- Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., & Zhang, F. et al. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science*, 339(6121), 819–823. doi:10.1126/science.1231143 PMID:23287718
- Davison, R. M. (2003). Information Privacy in a Globally Networked Society: Implications for IS Research. *Communications of the Association for Information Systems*, (12), 341–365.

- Dreyfuss, J. M., Levner, D., Galagan, J. E., Church, G. M., & Ramoni, M. F. (2012). How accurate can genetic predictions be? *BMC Genomics*, 13(1), 340. doi:10.1186/1471-2164-13-340 PMID:22827772
- Esvelt, K. M., & Wang, H. H. (2013). Genome-scale engineering for systems and synthetic biology. *Molecular Systems Biology*, 9(1), 641. doi:10.1038/msb.2012.66 PMID:23340847
- Eysenbach, G. (2001). What is e-health?. *Journal of Medical Internet Research*, 3(2), e20. doi:10.2196/jmir.3.2.e20 PMID:11720962
- Eysenbach, G., & Diepgen, T. L. (2001). The role of e-health informatics for evidence –based patient choice in the 21st century. *Clinics in Dermatology*, 19(1), 11–17. doi:10.1016/S0738-081X(00)00202-9 PMID:11369478
- Flavián, C., & Guinalú, M. (2006). Consumer trust, perceived security and privacy policy: Three basic elements of loyalty to a web site. *Industrial Management & Data Systems*, 106(5), 601–620. doi:10.1108/02635570610666403
- Galford, R., & Drapeau, A. S. (2003). The enemies of trust. *Harvard Business Review*, 81(2), 88–95. PMID:12577656
- Ginsburg, G. S., & McCarthy, J. J. (2001). Personalized medicine: Revolutionizing drug discovery and patient care. *Trends in Biotechnology*, 19(12), 491–496. doi:10.1016/S0167-7799(01)01814-5 PMID:11711191
- Hu, Q., West, R., & Smarandescu, L. (2015). The Role of Self-Control in Information Security Violations: Insights from a Cognitive Neuroscience Perspective. *Journal of Management Information Systems*, 31(4), 6–48. doi:10.1080/07421222.2014.1001255
- Hwang, Han, Kuo, & Liu. (2012). *The Differing Privacy Concerns Regarding Exchanging Electronic Medical Records of Internet Users in Taiwan*. Academic Press.
- Jadad, A., Oh, H., Rizo, C., Enkin, M., Powel, J., & Pagliari, C. (2005). What eHealth: A Systematic Review of Published Definitions. *Journal of Medical Internet Research*, 7.
- Jain, K. K. (2009). *Textbook of personalized medicine*. New York: Springer. doi:10.1007/978-1-4419-0769-1
- Karake Shalhoub, Z. (2006). Trust, privacy, and security in electronic business: The case of the GCC countries. *Information Management & Computer Security*, 14(3), 270–283. doi:10.1108/09685220610670413
- Katsikas, S. K., Lopez, J., & Pernul, G. (2005). Trust, privacy and security in e-business: Requirements and solutions. In *Advances in Informatics* (pp. 548–558). Springer Berlin Heidelberg. doi:10.1007/11573036_52
- Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., ... Church, G. M. (2013). RNA-guided human genome engineering via Cas9. *Science*, 339(6121), 823–826.
- McGuire, A. L., Fisher, R., Cusenza, P., Hudson, K., Rothstein, M. A., McGraw, D., & Henley, D. E. et al. (2008). Confidentiality, privacy, and security of genetic and genomic test information in electronic health records: Points to consider. *Genetics in Medicine*, 10(7), 495–499. doi:10.1097/GIM.0b013e31817a8aaa PMID:18580687
- Misra, S. C., & Bisui, S. (2014). Critical Challenges for Adopting Personalized Medicine System in Healthcare Management: Perspectives of Clinicians and Patients. *International Journal of E-Health and Medical Communications*, 5(2), 72–89. doi:10.4018/ijehmc.2014040104
- Misra, S. C., & Bisui, S. (2016). Feasibility of large scale implementation of personalized medicine in the current scenario. *International Journal of E-Health and Medical Communications*, 7(2), 30–49. doi:10.4018/IJEHMC.2016040103

- Misra, S. C., Bisui, S., & Fantazy, K. A. (2016). Identifying critical changes in adoption of Personalized Medicine (PM) in Healthcare Management. *International Journal of E-Health and Medical Communications*, 7(3), 1–15. doi:10.4018/IJEHMC.2016070101
- MRI. (2004). *Sixth Annual Survey of Electronic Health Record Trends and Usage for 2004*. Medical Record Institute. Retrieved from <http://www.me-drecinst.com/>
- NHMRC. (2011). *Clinical utility of personalised medicine, Information for health professionals 2011*. National Health and Medical Research Council, Australian Government. Retrieved from http://www.nhmrc.gov.au/_files_nhmrc/publications/attachments/ps0001_clinical_utility_personalised_medicine_feb_2011.pdf
- Ong, F. S., Kuo, J. Z., Wu, W.-C., Cheng, C.-Y., Blackwell, W. L. B., Taylor, B. L., & Wong, T. et al. (2013). Personalized medicine in ophthalmology: From pharmacogenetic biomarkers to therapeutic and dosage optimization. *Journal of Personalized Medicine*, 3(1), 40–69. doi:10.3390/jpm3010040 PMID:24624293
- PMC. (2007). *Personalized medicine: An introduction*. Retrieved from http://www.personalizedmedicinecoalition.org/sites/default/files/personalmed_backgrounder.pdf
- PMC. (2010). *Personalized medicine 101: The challenges*. Retrieved from <http://www.personalizedmedicinecoalition.org/about/about-personalized-medicine/personalized-medicine-101/challenges>
- PMC. (2014). *The case for Personalized Medicine*. Retrieved from http://www.personalizedmedicinecoalition.org/Userfiles/PMC-Corporate/file/pmc_the_case_for_personalized_medicine.pdf
- Smith, H. J., Milberg, S. J., & Burke, S. J. (1996). —Information Privacy: Measuring Individuals Concerns About Organizational Practices. *Management Information Systems Quarterly*, 2(20), 167–196. doi:10.2307/249477
- Zamft, B., Marblestone, A., Kording, K., Schmidt, D., Martin-Alarcon, D., Tyo, K., & Church, G. M. et al. (2012). Measuring Cation Dependent DNA Polymerase Fidelity Landscapes by Deep Sequencing. *PLoS ONE*, 7(8), e43876. doi:10.1371/journal.pone.0043876 PMID:22928047

Pervasive Mobile Health

Muhammad Anshari

Universiti Brunei Darussalam, Brunei

Mohammad Nabil Almunawar

Universiti Brunei Darussalam, Brunei

INTRODUCTION

Recent progress in mobile health (mHealth) is very encouraging as it allows healthcare services to be accessed through powerful smart mobile devices. Nowadays smart mobile devices such as smartphones and tablets support many features that can be utilized to access healthcare services conveniently. mHealth will indeed be a convenient experience to patients since there is currently a growing trend towards pervasive knowledge acquisition and management. Pervasive mHealth is an electronic health where healthcare services can be accessed regardless of time and place, and patients can continuously be connected to their healthcare services through their smart mobile devices.

According to Eysenbach (2011), healthcare services in mHealth are delivered through mobile devices. Pervasive mHealth provides services that go beyond traditional healthcare services. It provides various services accessible through smart mobile devices, including searching relevant health information, interactive discussion of healthcare matters with colleagues or friends, updating health related information from smart mobile devices, and seeking opinion through online healthcare services. Pervasive mHealth is becoming an increasingly important instrument in healthcare services for the public.

Most healthcare initiatives perceive patients as recipients of healthcare, where they hardly have roles in the process of healthcare activities and decision-making. For example, it is difficult for a patient to get access to his/her own health

record or to find out his/her health history. On the other hand, the advancement of mobile technology has given customer (patients) a greater role in the process including the ability to gain access and control of information flow that fits their personalized needs. In terms of healthcare activities, many smart electronic devices have enabled patients to check their own health condition without assistance from healthcare staffs such as checking blood pressure, food hygiene Apps, mobile thermometer, mobile glucose measurement, mobile hypertension tool, etc.

In addition, the wide adoption of smartphones and tablets as well as the emerging of their utilization in healthcare raises the possibility to engage patients, patients' families, and society to participate in their own healthcare processes. For instance, social networking, which is accessible through smartphones, can provide an avenue to identify any misuse or misunderstanding in consuming antibiotics, share health information, promote positive behavior change, disseminate valid information, and explore how such tools can be used to gather real-time health data (Scanfeld et al., 2010). Pervasive mHealth can offer healthcare providers a comprehensive perspective of patient's condition and thus provide an environment in achieving complex healthcare goal(s) such as building lasting relationships with patients. The main goal of this chapter is to accommodate the concept of pervasive mHealth scenario within pervasive knowledge acquisition in healthcare systems. We propose a pervasive mHealth model that extends to multi-channel knowledge resources in mHealth. The model is designed to establish

longitudinal healthcare services anytime and anywhere which in turn will raise health literacy of individuals and the community at large.

BACKGROUND

Currently almost every healthcare organization depends on information and communication technology (ICT) in every level of their activities. One of the most valued aspects of ICT is its capability to enhance processes or functions to improve consistency, accuracy, and efficiency. Similarly, efficient and effective healthcare services have become increasingly dependent on accurate and detailed clinical information, which is transferred from interrelated departments or even between organizations (Conrick, 2006). In addition, people's demands have changed because they have been empowered by the vast amount of information available and accessible to them through smart mobile devices, including mHealth related services (Anshari et al., 2013). For example, half of smartphone owners in US use their devices to get health information and one fifth of smartphone owners have health-related apps (Fox & Duggan, 2012). Despite some limitations of smartphones (Koushanfar et al., 2000), they can be used as services platform for mobile health information delivery, access and communication (Nkosi et al., 2010).

Rapid adoption of smartphones and smart mobile devices has removed the space, time and distance barriers for users, including the patients, to access information. In fact, smartphones facilitate the use of dynamic Web 2.0 and Apps allows intensive and immediate interactivity through social networks. In Web 2.0, users can collectively contribute contents and applications (O'Reilly, 2005). Web 2.0 is defined as a set of economic, social, and technology trend that collectively forms the basis for the next generation of Internet – a distinctive medium characterized by users' participation in using networks such as *Facebook*, *Twitter*, *Snapchat*, *Instagram* and *LinkedIn*, which have

indeed grown rapidly, facilitating peer-to-peer collaborations, participation and networking. The term social networks refer to the use of Web 2.0 that describes the social characteristics supporting the promotion of collaborative sharing. Web 2.0 is commonly associated with technologies such as weblogs (blogs), social bookmarking, wikis, podcasts, Really Simple Syndication (RSS) feeds (and other forms of many-to-many publishing), social software, and Web application programming interfaces (APIs) (Kristaly et al., 2008). Web-oriented applications and services use the Internet as a platform to provide unique features by engaging and providing users with rich interactive experiences (Rodrigues & Vaidya, 2010).

The widespread acceptance of the Web 2.0 services will affect the operation of future businesses. Greenberg (2009) explained these changes and that the customer's ownership of the conversation ultimately leads to organizations readjusting their business strategy. The fast adoption of smartphones has made customers or patients demand accessibility of health services through their smartphones. It is imperative for Health service providers to respond to these demands to sustain their businesses. In fact, health service providers can provide services that take advantage of smartphones' features or other mobile devices such as Web 2.0 features. The use of smart mobile devices with the Internet connection has permeated almost all aspects of our lives, creating pervasive computing.

Pervasive computing is associated with smart environments in which information and communication technologies employ intelligence and machine learning techniques to reason about, control and adapt to our physical surroundings. It focuses on sensing, interacting with and aiding humans at an individual and community level (Cook & Das, 2004). This means that pervasive computing should encompass every device worldwide that has built-in active and passive intelligence (Saha & Mukherjee, 2003). While mobile computing supports IT such as remote information access and adaptive applications, pervasive computing

extends this notion to provide computing and communication capabilities that are integrated with users (Weiser, 1991).

MOBILE HEALTH

Challenges and Opportunities

As a business entity, a healthcare provider needs to have a comparable standard of customer service with other business organizations. It is a fact that customer expectations in healthcare are high, creating serious challenges for healthcare providers as they have to make an exceptional impression on every customer. In a competitive market, poor service drives customers to switch from one healthcare provider to another (Anshari & Almunawar, 2011). The adoption of mHealth will be a natural consequence of the adoption of smartphones. Hence, healthcare services must be adjusted in such a way that people can access the service through their smartphones or other mobile devices.

Most recent healthcare initiatives perceive customers (patients) as recipients of health care as they do not have a significant role in the decision-making process of their health (Anshari & Almunawar, 2015a). However, the advancement of Web 2.0 offers patients to have a greater role in the decision-making process related to their health. This is because they can be provided with the ability to access and control some personal information that fits with their personalized needs, which allows them to actively participate in the decision-making process related to their health.

Web 2.0 tools have brought a possibility to extend e-health by enabling patients, patients' families, and the community at large to participate more actively in the process of health promotion and education through a social networking process (Almunawar, Anshari, & Younis, 2015). Embedding Web 2.0 in healthcare is a challenging task in order to provide a new meaning in building relationships between patients and healthcare

organizations within a social network platform. In addition, social network applications that exist in smartphones can be utilized to facilitate interactions between patients and their healthcare providers (Anshari & Almunawar, 2015b). Furthermore, the impact of Web 2.0 technologies for health continues to grow, and the term Medicine 2.0 has entered popular categorization. Hughes et al., (2008) mentioned that the terms of Medicine 2.0 and Health 2.0 were found to be similar and categorized into five major notable cores:

1. The participants involved (doctors, patients, etc.);
2. Its impact on both traditional and collaborative practices in medicine;
3. Its ability to provide personalized health care;
4. Its ability to promote on going medical education; and
5. Its associated methods and tool related issues, such as potential inaccuracy in end user-generated content.

However, discussion on Medicine 2.0 needs further investigation to map the relationship model of patient-patient interactions, and patient-healthcare provider interactions. This chapter proposes that the role of patients in mHealth services can be extended into an individual health actor, a social health agent through social networks, and a medical care partner.

Managing interactions in any business organization requires proper concepts, business process, tools, people, and strategies. Therefore, mHealth model must provide a mechanism that have the ability to broaden service beyond its usual practices. Three aspects of interaction in mHealth can be considered (Figure 1). First, *patient-self interactions* with the system, which is the ability of a patient as an individual to actively utilize mobile healthcare services provided by the healthcare provider. For instance, patient-self interaction enables the patient to view his/her medical records through his/her mobile phone,

or request for a consultation. Secondly, *patient-to-patient interactions*. This allows patients to interact among themselves, share information, concerns, and experiences. For example, the system facilitates social networks for patients with similar conditions to make conversation and share their experiences. Thirdly, *patient-to-staff interactions* provide online interaction between patient and healthcare staff.

Finally, the development of the Internet of Things (IoT), big data, and massive medical software applications (apps) has supported the generation of pervasive knowledge including mobile health devices. Mobile health devices have empowered patients to use care tools such as drug references, healthcare references, diagnostic aids, medical calculators, etc. Whenever mobile health devices connects with mHealth system it can generate information that enriches personal health records. For instance, a blood pressure tool measures a patient's blood pressure and instantly updates the record of his/her blood pressure in the mHealth system. The focus of pervasive mHealth is recognized in different areas of data and information management, including accessibility, collaboration, information utilization, integration, and process automation of mHealth data.

Figure 1. Interactions scenario in mHealth



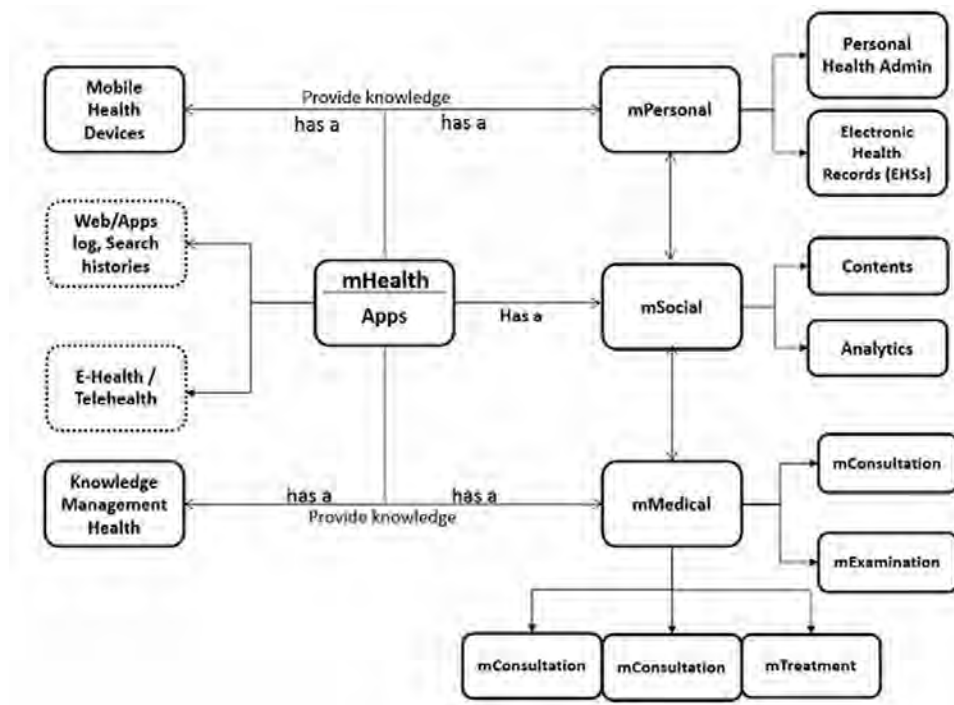
Solutions and Recommendations

Active role of patients through mobile services is delineated as the embodiment of future mHealth, with the hope to make the system more reliable, effective and efficient in addressing patients' demands. The model addresses multi levels interactions to provide online information access, consultations, information sharing, and other services through smart mobile devices. The model is expected to lay the perspective for pervasive knowledge management in which mHealth provides service for improving healthcare literacy and building relationships among all the parties involved (a healthcare organization, patients, and the community at large).

Understanding a business process helps healthcare organizations to deliver the strategic plan in providing proper mHealth services. The model combines social networks, mobile health devices, and knowledge health database (Figure 2). It is important to examine each business process as a layer of value to the patients. Patients place a value on these services according to the quality of outcome, service, and price. The value of each layer depends on how well they are performing. If the processes do not correspond to the strategy, the value will be significantly diminished. For example, the value of mHealth is reduced by any delayed responses to patients' queries or healthcare staffs' poor communication skills.

Figure 2 depicts the model of pervasive mHealth with features of social networks, pervasive knowledge management, and mobile health devices. Mobile health devices support personalized healthcare anywhere and anytime in delivering healthcare service to users. Social network has enabled patients to dynamically enrich their personalized information. Conventional systems have relied solely on static information provided by the service provider. However, they are now empowered with the ability to conduct conversations with other patients to share and exchange experiences through discussions in social networks.

Figure 2. Proposed model of pervasive mHealth



Pervasive mHealth offers patients a greater role in active healthcare processes by empowering them with the ability to control and generate flow of information that fits with their personalized needs. Pervasive knowledge in mHealth is proposed to enhance all actors in the system (patients, healthcare staffs, automate smart mobile devices, etc.) to generate content with relevant knowledge from various channels. The concept of pervasive knowledge in the model is to enable multi-channel knowledge that is available in the Internet. Knowledge health database can link information across different online sources to form a more comprehensive knowledge to support synchronous lecture, asynchronous study, online consultation, knowledge sharing in social networks.

In addition, the model accommodates the three dimensions of mHealth: *mPersonal*, *mSocial*, and *mMedical*. *mPersonal* refers to patient-self interaction. *mPersonal* derives functions of personal health activities that cover all personal health related activities in mHealth services as

an individual (patient) which can directly or indirectly affect their health conditions. Activities in this category may include personal Identity (ID), personal Habits of patient (HB), Exercise activities (EX), Spiritual and Emotional activities (SE), personal Health Plan (HP), personal Account information (AC), and so on. For instance, ID consists of personal information in the mHealth system such as name, address, phone no, email address, login ID, and password, where the information solely belongs to personal and a patient (patient-self) is in control of the data. HB stands for daily habits of the individual which may consist of eating, sleeping, and any other habits that can affect personal health. The EX is routine exercise activities of the individual that may be beneficial when they are recorded in the system. *mPersonal* will replace the conventional approach; instead of the health staffs entering personal information into the system, the customers themselves will do it on their own, empowering them.

Secondly, *mSocial* refers to managing patient to patient interactions through social networks.

The social networks refer to the social networks applications such as Facebook, Twitter, LinkedIn, and Instagram, or social networks that are operated, managed and maintained within the healthcare's infrastructure which patients may belong to. This is targeted towards patients/families within the same interest or health problem/ illness to converse between each other. For example, patients with diabetes will be motivated to share his/her experiences, learning, and knowledge with other diabetic patients. Constructive conversation and information from external or internal social networks should be captured as it is an excellent, innovative strategy to improve services. mSocial enable individuals and the community as social health agents for others in mHealth systems. mSocial accommodates the concept of social network within mHealth services. Some activities in this category include conversations (CS), chatting, updating status, forums, wikis, blog, knowledge management (KM), group knowledge and asking for specific services (RS). Nowadays, people constantly use social networks in their daily lives. Updating status in social networks trigger conversation among circles of friends. Bringing this scenario into mHealth services is both interesting and challenging. For instance, patients with the same condition like diabetes may share their experiences with other patients in social networks. Sharing in social networks may become a virtual supporting group that can enrich and strengthen their motivation to fight for better health. In fact, social networks also affect the relationship between the healthcare provider and customers (patients). Therefore, embedding mSocial in mHealth is imperative.

Finally, mMedical refers to the interactions between healthcare/medical staffs with patients. In fact, mMedical is the core activity of mHealth, because it is where patients raise medical concerns through mobile devices and are responded by dedicated medical staffs managing mHealth services. mMedical proposes that patients can be involved in the medical care processes and medical decision-making in the mHealth scenario. Any medical

queries from patients can be facilitated through the module available in mMedical. mMedical's module may consist of examination (XM), mAppointment (MA), mPrescription (MP), and mTreatment (MT). XM is online consultation between patients and medical staffs which can lead to the generation of electronic medical record (EMRs) for the patients. It is a common service in many e-health initiatives; however, when there is empowerment or participation from patients in these processes of medical activities, the result of consultation can be accessed by patients directly, anytime and anywhere. For instance, many healthcare providers prevent patients from accessing their EMR. In fact, those patients are not in the position to track their medical history themselves; whenever the patient needs consultation, the diagnosis will start from the beginning. mMedical should enable patients to have access over their medical records regarding their health status, condition, and history anywhere-anytime. The MP is another subset of EMR, another area in which customers need to be empowered. Empowerment in MP will speed up the process of managing prescription, and customers will know how to consume their medication appropriately, as they are able to access and learn anytime-anywhere. In summary, participation and empowerment in mMedical can shift the role of a patient from a recipient of care, to a partner of care. Access to knowledge and contents (EMR) will make patients more knowledgeable on their health conditions and issues.

FUTURE RESEARCH DIRECTIONS

The advancement of pervasive mHealth is expected to improve interactions of all actors. Those interactions range from patient-patient, patient-healthcare provider, patient-community, healthcare-community, and smart mobile health-patient. Meanwhile, there are a lot of opportunities for broad arrays of research in developing smart mobile health devices that will integrate all data into a single platform of mHealth. In addition,

empowerment in mHealth will continue to grow in three subdomains; *mPersonal*, *mMedical*, and *mSocial*. *mPersonal* can be the first module to be tested for deployment since the data generated is less complex, involving data on personal health habits. Secondly, *mSocial* encourages patients to share and support other patients with their knowledge and experiences. From an organizational perspective, *mSocial* can be an effective marketing tool to promote services to customers. *mSocial* can also be utilized as a place to capture ideas from customers. Further studies in developing listening tools will help healthcare organizations to better understand their customers' needs. Thirdly, *mMedical* is the core of the mHealth's system because critical data is stored in this module. It is important to decide which data can be shared with patients and which cannot. This issue needs to be carefully studied, including any related legal aspects. Although this study covers a broad-spectrum of issues across multiple domains, the four primary potential target groups for this study are:

1. Customers (patients / patients' family) who need to be informed, empowered, educated, and connected with others to encourage information or knowledge sharing.
2. Healthcare providers that are seeking innovative ways to improve their customer relationship, loyalty, trust, customer satisfaction, market leader, and any state of interactions.
3. Government and policy makers who seek innovative healthcare solutions through mHealth services with less budget for their citizens.
4. The researchers who need to fully understand the concept of healthcare business processes, technology, and the issues of each domain.

CONCLUSION

Pervasive knowledge is generated from various channels and tools. It can come from smart mobile

health devices, social networks, Internet of Things (IoT), and knowledge health database. Integrating all the resources will provide personalized service in pervasive mHealth scenario. Pervasive mHealth extends multi channels sources and interactions from all actors. Those actors are patients, healthcare providers, social networks, mobile health devices, and knowledge health database. Pervasive mHealth covers three domains of patients' empowerment to ensure interactivity and participation in the healthcare processes. Finally, accommodating the advantages of each model above will give the opportunity to improve the mHealth services that are expected to meet users' demand which is the concern of this study.

REFERENCES

- Almunawar, M.N., Anshari, M., & Younis, M. Z. (2015). Incorporating Customer Empowerment in Mobile Health. *Health Policy & Technology*. DOI:10.1016/j.hlpt.2015.08.008
- Anshari, M., & Almunawar, M. N. (2011). Evaluating CRM Implementation in Healthcare Organization. *2011 International Conference on Economics and Business Information*. IACSIT Press.
- Anshari, M., & Almunawar, M. N. (2015a). Designing Role of Online Health Educators in Healthcare Services. *Journal of Evidence-Informed Social Work*, 12(2), 220-236.
- Anshari, M., Almunwar, M. N., Wint, Z., Low, P. K. C., & Younis, M. (2013). Adopting Customers' Empowerment and Social Networks to Encourage Participation in E-Health Services. *Journal of Health Care Finance*. Aspen Publishers, 40(2), 17-42. PMID:24551960
- Conrick, M. (2006). *Health informatics: Transforming healthcare with technology*. Thomson Social Science Press.
- Cook, D., & Das, S. K. (2004). *Smart Environments: Technologies, Protocols and Applications*. Wiley. doi:10.1002/047168659X

Eysenbach, G., & Group, C. E. (2011). Consort-ehealth: improving and standardizing evaluation reports of Web-based and mobile health interventions. *Journal of Medical Internet Research*, 13(4). doi:10.2196/jmir.1923

Greenberg, P. (2009). CRM at the Speed of Light: Social CRM 2.0 Strategies, Tools, and Techniques for Engaging Your Customers (4th ed.). McGraw-Hill Osborne Media.

Hughes, B., Joshi, I., & Wareham, J. (2008). Health 2.0 and Medicine 2.0: Tensions and Controversies in the Field. *Journal of Medical Internet Research*, 10(3), e23. doi:10.2196/jmir.1056

Koushanfar, F., Potkonjak, M., Prabhu, V., & Rabaey, J. (2000). Processors for Mobile Applications. *Proceedings of the IEEE International Conference on Computer Design*.

Kristaly, D. M., Sisak, F., Truican, I., Moraru, S. A., & Sandu, F. (2008). Web 2.0 technologies in web application development. *Proceedings of the 1st International Conference on Pervasive Technologies Related to Assistive Environments*.

Marconi, J. (2002). *E-Health: Navigating the Internet for Health Information Healthcare Advocacy*. Healthcare Information and Management Systems Society. Retrieved from http://www.himss.org/content/files/ehealth_whitepaper.pdf

Nkosi, M. T., & Mekuria, F. (2010). Cloud computing for enhanced mobile health applications. In *Proceedings of the 2010 IEEE Second International Conference on Cloud Computing Technology and Science (CloudCom)* (pp. 629-633). IEEE. doi:10.1109/CloudCom.2010.31

Reinhold. (2006). Health information systems – past, present, future. *International Journal of Medical Informatics*, 75, 268-281.

Rodrigues, J. J. P. C., & Vaidya, B. (2010). *Health Information Systems: Concepts, Methodologies, Tools and Applications*. Hershey, PA: Medical Information Science Reference. doi:10.4018/978-1-60566-988-5

Saha, D., & Mukherjee, A. (2003). Pervasive computing: A paradigm for the 21st century. *Computer*, 36(3), 25–31. doi:10.1109/MC.2003.1185214

Scanfeld, D., Scanfeld, V., & Larson, E. L. (2010). Dissemination of health information through social networks: Twitter and antibiotics. *American Journal of Infection Control*, 38(3), 182–188. doi:10.1016/j.ajic.2009.11.004

Weiser, M. (1991). The computer for the twenty-first century. *Scientific American*, 265(3), 94–104. doi:10.1038/scientificamerican0991-94 PMID:1675486

ADDITIONAL READING

Almunawar, M. N., & Anshari, M. (2011, August). Improving Customer Service in Healthcare with CRM 2.0. *GTSF Business Review*, 1(2), 228–234.

Almunawar, M. N., Wint, Z., Low, K. C. P., & Anshari, M. (2012). Customer's Expectation of E-Health Systems in Brunei. *Journal of Health Care Finance. Aspen Publishers*, 38(4), 36–49.

Anshari, M., & Almunawar, M. N. (2012). Framework of Social Customer Relationship Management in E-Health Services. *Journal of Health Management*, 2012, 1–14. doi:10.5171/2012.766268

Anshari, M., & Almunawar, M. N. (2013). Shifting Role of Customers From Recipient to Partner of Care in Healthcare Organization. PhD Colloquium, The 13th ASEAN Graduate Business, Economic Program (AGBEP) Network, Yogyakarta, April 2, 2013, p. 11

Anshari, M., & Almunawar, M. N. (2015b). Tracking Future Path of Consumers Empowerment in E-Health. *International Journal of E-Health and Medical Communications*, 6(3), 62–75. doi:10.4018/IJEHMC.2015070104

- Anshari, M., Almunawar, M.N., Low, K.C.P. (2012). CRM 2.0 Within E-Health Systems: Towards Achieving Health Literacy and Customer Satisfaction, *Journal of Development Informatics*, Vol.1, No. 1.
- Anshari, M., Almunawar, M. N., Low, P. K. C., & Al-Mudimigh, A. S. (2013). Empowering Clients Through E-Health in Healthcare Services: Case Brunei. *International Quarterly of Community Health Education*, 33(2), 191–221. PMID:23661419
- Anshari, M., Almunawar, M. N., Low, P. K. C., & Wint, Z. (2012). Customer Empowerment in Healthcare Organisations Through CRM 2.0: Survey Results from Brunei Tracking a Future Path in E-Health Research. ASEAS – Austrian. *Journal of Southeast Asian Studies*, 5(1), 139–151.
- Austin, C., & Boxerman, S. (2003). *Information systems for healthcare management* (6th ed.). Chicago: Health Administration Press.
- Ball, M. J., & Lillis, J. (2001). E-health: transforming the physician/patient relationship. *International Journal of Medical Informatics*, 61(1), 1–10. PubMed doi:10.1016/S1386-5056(00)00130-1
- Berzina, I., & Bommel, P. V. (2009). *Behind the Internet Business Models; An E-Health Industry Case* (p. 83). Copenhagen Business School.
- Chaudhry, B., Wang, J., Wu, S., Maglione, M., Mojica, W., Roth, E., et al. (2006). Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. *Annals of Internal Medicine*, 144(2), 742–752. PubMed doi:10.7326/0003-4819-144-10-200605160-00125
- Chekuri, L., & Schumacher, F. (2012). *Health Systems: Improving with Information Technology*, Retrieved 10 September, 2012 from http://www.cwru.edu/med/epidbio/mph439/Medical_Informatics.htm
- Conrick, M. (2006). *Health informatics: Transforming healthcare with technology*. Thomson Social Science Press.
- Czepiel, J. A., Rosenberg, L. J., & Akerele, A. (1974). Perspectives on Consumer Satisfaction. In R. C. Curhan (Ed.). 1974 Combined Proceedings Series No: 36. American Marketing Association.
- Eysenbach, G. (2001). What is e-health? *Journal of Medical Internet Research*, 3(2), e20. doi:10.2196/jmir.3.2.e20 PMID:11720962
- Fox, S., & Duggan, M. (2012). Mobile health 2012. Pew Research Center's Internet x0026 American Life Project.
- Istepanian, R., Laxminarayan, S., & Pattichis, C. S. (2006). *M-health: Emerging mobile health systems. M-Health: Emerging Mobile Health Systems* (R. Istepanian, S. Laxminarayan, & C. S. Pattichis, Eds.). Berlin: Springer.

KEY TERMS AND DEFINITIONS

Electronic Health (e-Health): According to WHO (2005) is a cost effective and secure use of ICT in support of health and health related fields, including healthcare service, health surveillance, health literature and health education, knowledge and research.

Health Information System (HIS): Refers to any system that captures, stores, manages or transmits information related to the health of individuals or the activities of organizations that work within the health sector. This definition incorporates things such as district level routine information systems, disease surveillance systems, and includes laboratory information systems, hospital patient administration systems (PAS) and human resource management information systems (HRMIS).

Healthcare Business process: Refers to a set of activities which activity adds value to the customer. These are registration, patient care, discharge, marketing, and service.

Healthcare: Refers to the diagnosis, treatment, and prevention of disease, illness, injury, and other physical and mental impairments in humans. Practitioners in medicine, optometry, dentistry, nursing, pharmacy, allied health, and other care providers deliver health care. It refers to the work done in providing primary care, secondary care, and tertiary care, as well as in public health.

Mobile Health (mHealth): mHealth is an extension of electronic health (e-health) in which healthcare services can be accessed through smart mobile devices. The scope of mHealth is mobile personal (mPersonal), mobile social (mSocial), and mobile medical (mMedical).

Patient Empowerment: In healthcare is a principle to give patients more power in making

decisions related to their health and to give them required authority to access or produce information related to their health.

Pervasive Computing: Refers to an embedding microprocessor in device (mobile) to connect and communicate in the networks.

Pervasive Knowledge: Refers to a massive knowledge generated from multi sources of information such as social network, smart mobile devices, and structure knowledge management database.

Smartphone: Refers to a mobile phone built on a mobile operating system, with more advanced computing capabilities and connectivity than feature phones.

M

Radio Frequency Identification Technologies and Issues in Healthcare

Amber A. Smith-Ditizio

Texas Woman's University, USA

Alan D. Smith

Robert Morris University, USA

INTRODUCTION

Barcoding and RFID-Related Technologies in the Services Sector

Barcodes and related identification technologies have been used to control inventory and supply chain management (SCM) for some time, especially in retail and purchasing applications (Aldaihani & Darwish, 2013; Azadeh, Gholizadeh, & Jeihoonian, 2013; Bhamu, Khandelwal, & Sangwan, 2013). Barcoding equipment is fairly inexpensive and easy to use as compared to other AIDC technologies (Smith, 2011; Smith, Smith, & Baker, 2011; Smith & Rupp, 2013; Visich, Li, Khumawala, & Reyes, (2009; Wilson, 1995; Wyld, (2006). However, certain limitations create the need for a new approach to increase efficiency (Drejer & Riis, 2000; Dutta, Lee, & Whang, 2007; Fisher & Monahan, 2008; Fumi, Scarabotti, & Schiraldi, 2013). Barcodes are only accurate if items are continuously scanned in and out as they move along the supply system. An employee must ensure an item is scanned at each stage, or entry and exit point. An example of this would be a delivery driver that scans packages as a truck is loaded and scans the package at the delivery point. If the package is not scanned, the action is not accurately recorded. Barcodes require an employee to physically inspect the item for scanning purposes to ensure inventory accuracy and determine the product's location, while Radio Frequency Identification (RFID) conveniently

tracks products through radio waves, designed to improve operational efficiency.

RFID technologies helped Walmart to control its inventory and track product movements along its supply chain (Tarofdor, Marthandan, Mohan & Tarofdor, 2013). Furthermore, RFID has been used in the identification of stray pets, which is known as chipping and in transportation, in addition to sports. Experimentation with RFID in soccer may soon allow a visually impaired player to participate in the game by integration of computers and video cameras, along with an active RFID tag to signal to the player through a set of audio headphones (Zare, McMullen & McCune, 2014).

DISCUSSION

RFID in the Healthcare Industry

Anand and Wambaa (2013) performed a case study to evaluate how RFID can be applied in healthcare settings. The authors suggested that there are significant benefits for both patients and healthcare organizations in the application of such technology. Adoption in the healthcare industry is not widely spread because of the initial high cost associated with equipment and technology implementation. One argument against the investment emphasizes that the technology primarily focuses on tracking and implies that the costs outweigh the benefits (i.e., compared to the cheaper alternative of barcodes). However, Anand

et al. compared the evidence collected from 39 hospitals and analyzed the costs and benefits of RFID implementation over the long run. Their findings indicated that RFID-embedded systems can save time and reduce waste.

Hence, RFID-related technologies can help reduce costs associated with the medical supply chain (e.g., pharmaceuticals and medical supplies) or in patient tracking and management. For example, high-valued supplies, such as blood can be equipped with RFID tags that can ensure that the right type of blood is being administered to the right patient (Winters, 2010). When associated with patient tracking, RFID can be used to track patient history as well as hospital admittance, transfer, and discharge. Advantages also include reducing redundant work and time consuming processes by making certain that the correct blood is drawn from storage. This allows healthcare providers to concentrate on patient care and service (Dominic, Goh, Wong, & Chen, 2010; Tarofdor, et al., 2013). RFID-related technologies can also reduce errors and waste by assuring that the accurate drug or medical product, ordered from the appropriate supplier is used at the proper time for the correct patient in the suitable dose.

RFID is an acceptable way to automate and improve the tracking of medical products and pharmaceuticals. Patients gain from RFID-related technology because of the reduction in errors when healthcare providers are prescribing or fitting medical products to them. As suggested by a number of researchers (Smith, 2012, 2013a, 2013b, 2013c; Ting, Kwok, Tsang, & Lee, 2009; van der Togt, Jan van Lieshout, Hensbroek, Beinat, Binnekade, & Bakker, 2008; Wickboldt & Piramuthu; Zang, & Fan, 2007), the experimentation of RFID in a variety of fields [i.e., sports, SCM, operations management, lean manufacturing] may lead to new discoveries and other applications (Hu, Wang, Fetch, & Bidanda, 2008; Ifinedo & Nahar, 2009; Jain, Benyoucef, & Deshmukh, 2008; Johansson & Sudzina, 2008; Jones, Riley, Franca, & Reigle, 2007; Ketikidis, Hayes, Lazuras, Gunasekaran, & Koh, 2013; Koong & Lin, (2007).

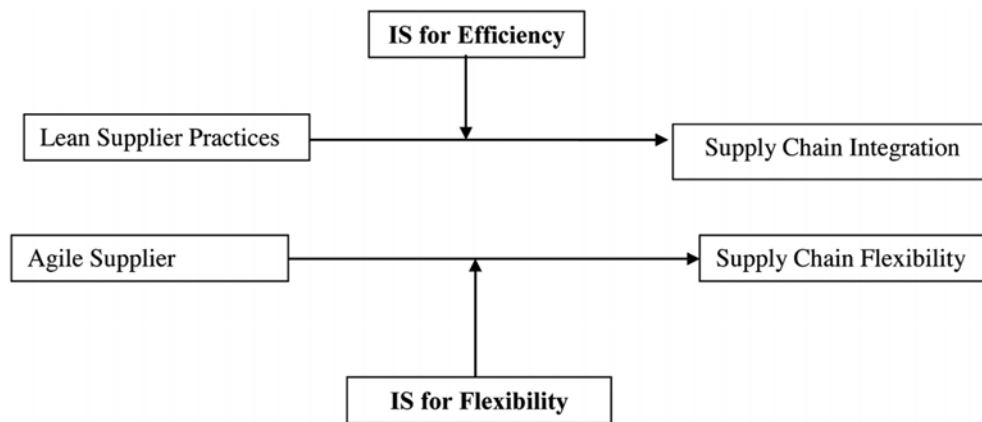
The application of RFID technology on the soccer field, for example, may also be applied to patients in the emergency room setting or soldiers on the battlefield; the range of active RFID tag can easily reach 300ft. and can detect direction and motion (Zare, et al., 2014). Emergency room specialists can analyze the information from RFID to track patient movements.

Implications for Medical Supply Chain Management

RFID utilizes radio waves to communicate, track, and trace products along the supply chain with the goal of helping to boost efficiency through enhanced data communication, counterfeit reduction, and drug quality insurance. There are many studies that point to the obvious operational efficiencies and reduction in waste and errors associated with the implementation of RFID-embedded systems (Aldaihani & Darwish, 2013; Azadeh, et al., 2013; Bhamu, Khandelwal, & Sangwan, 2013; Chen, Wu, Su, & Yang, 2008). For example, one main cost for any firm is a lack of knowledge of supply chain integration and how to manage a company's inventory, especially in promoting vendor-managed inventory system and the use of virtual asset trackers (VAT). A company does not want to order as many products as they can and expect it to empty its shelves instantly. By using certain IS applications [automatic identification and capture systems (AIDC) like barcodes, smart cards, touch memory, RFID] aligned with supplier practices, a company can have an improved supply chain performance well into the future ("Healthcare IT outsourcing ...," 2016).

Qrunfleh, Tarafdar, & Ragu-Nathan (2012) provided an empirical based study of how the alignment between supplier management practices and information systems strategy provides a positive association with supply chain integration goals. From their research, as illustrated in Figure 1, IS for efficiency refers to a strategy that is oriented toward operational support of intra- and inter-organizational efficiencies and that the

Figure 1. Using Information systems of increased supply chain efficiencies and flexibilities (partially adapted from Qrunfleh, et al., 2012)



ultimate goal of supplier management practices is to achieve better performance in the supply chain. Ultimately, healthcare organizations have asked their decision makers in strategic positions to evaluate hospitals' need to find ways to support IT and operational functions to deliver high-quality patient care and increase the proficiency of its operations and finances (Devaraj, Ow, & Kohli, 2013). Bottlenecks such as overages in time for surgery, emergency room inflow, and scheduling errors need to be reduced in order to achieve an even and swift process flow.

The above exhibit illustrates some of the interrelationships and interdependencies of supplier practices and different IS strategies that directly affect the supply chain performance of a company. Lean supplier practices consist of activities aimed at creating the most cost efficiency in the supply chain by reducing the inventory and focusing on improving the quality in the supply chain by eliminating waste.

Compared to the alternative of using barcodes, RFID has multiple advantages, which include enhanced traceability and transparency that result in increased operational efficiency (Wambaa, et al., 2013). Moreover, the exchange of data, access to real-time information, and "external and internal coordination of material flows" (Chircu, et al., 2014, p. 738) are key factors to improved efficiency within the supply chain. The collabora-

tion of the manufacturer, wholesaler, pharmacies, and hospitals adds value to the supply chain by reducing errors, and as a result, costs are also lowered. One of the many problems facing the pharmaceutical industry is the persistence of generic and counterfeit drugs as companies try to successfully deal with competition that strategically leverages lower prices. Proper use of RFID technology in supply chain integration issues may offer an alternative to solve these pricing issues. Although using barcodes is cheaper and easier to affix to products, they are not tamper-proof, and it is not possible to read them in bulk, which are two advantages of RFID labels. Ultimately, any transition to new technologies will incur costs. Combining RFID with electronic documents (i.e., e-pedigree) that summarize data about a drug is useful to all supply chain participants. Essentially, it is vital for any organization within any industry to operate as efficient as possible and to search for new ways on how to add value to its processes. RFID can be the solution.

In healthcare settings, RFID can be utilized to manage medication and materials for complex procedures that previously were time consuming and subject to human mismatching errors. One adoption decision was made by Integra's to overcome waste. The use of RFID saved US\$111,500 in lost and missing products and US\$223,000 in expired products (Anand & Wambaa, 2013). Healthcare

professionals, for example, may use RFID readers to access information about medications that can be displayed conveniently on a computer screen.

IMPLICATIONS IN THE HOSPITAL SETTING

Multiple Implications for RFID Technologies in Healthcare

There are multiple implications for RFID technologies in healthcare outside of SCM and supplier integration and performance (Lee & Shim, 2007). In hospitals' workflows in the areas of scheduling, admitting, discharging, and operating can be improved with the implementation of RFID. RFID tags planted in patient bracelets can accommodate information about the patient, such as his or her location, medical history, allergies, age, weight, blood results, diagnosis, living will, and religious preferences. Those bracelets can be utilized in different ways to increase patient safety, operate leaner, and offer improved service (Revere, et al., 2010). For example, RFID bracelets can reduce errors in the billing process. Portable scanners can be used by hospital personnel to scan the patient's bracelet and enter medications administered and services provided. Updated and stored in a database, this information is later used by the billing department to bill the patient. This accuracy leads to faster payments and improved customer satisfaction.

According to one recent study (Southard, Chandra & Kumar, 2012), medical errors lead to a significant amount of deaths. The number ranges between 44,000 and 98,000 per year. Medical errors result in a cost of about US\$2 billion annually. Information on RFID tags in patient bracelets can be used to reduce errors in the administering of drugs by comparing patient information on the bracelets with information on the medication. This reduces errors involving administering the wrong medication or the wrong dosage. Whenever medication is dispensed, the technology can be

utilized to manage inventory levels electronically. Another implication to increase patient safety involves the tracking of operating supplies and materials to avoid surgical tools being "left inside a patient" (Revere, et al., 2010, p.30).

The tracking of patients also results in the reduction of waiting time and an increased efficiency of capacity usage by optimizing bed turnover. As soon as a patient is being discharged from the hospital, an automatic alert can be sent out to housekeeping to prepare the room for a new patient (Revere, et al., 2010). Another improvement concerning time efficiency can be achieved by automatically syncing the computer displays of examination rooms with current patient information as soon as the patient walks through the door. These advances should help reduce miscommunication between physicians and nurses. As the results of the case study concluded, there existed quantitative evidence of the benefits of implementing RFID in healthcare processes by examining surgical processes (Southard, et al., 2012). In one example, it was estimated that nurses typically were spending an average of 15 to 20 mins. per shift trying to locate the proper equipment. Moreover, hospitals waste about US\$50,000 monthly on rental charges because of misplaced or lost equipment. Resulting in about US\$20,000 annually wasted on human resource costs to search for the lost equipment. Other than increased costs, further negative implications include medical equipment missing routine maintenance. In these situations, implementation of RFID can improve the coordination of hospital processes and enhance visibility.

One hospital spent only US\$35,000 on software and US\$50 per active RFID tag. Moreover, it used existing Wi-Fi infrastructure and did not acquire any readers (Southard, et al., 2012). The benefits, transparency and prevention of errors of RFID, outweigh the costs. Again, RFID has two main advantages over barcodes. RFID tags do not have to be physically close to the reader to be read and multiple "pieces of information" can be stored on the tag. This is useful in the supply chain of

pharmaceuticals when tracking and controlling information about temperature, movement, and tampering is important (DesRoches, et al., 2008; Ludwick & Doucette, 2009).

Usage, Application, and Adoption Issues

The usage and application of RFID tags are spreading rapidly and are increasingly spreading into the healthcare industry (DesRoches, et al., 2008; Smith, 2012). Perhaps, the first step in successful implementation of RFID is to perform an initial analysis and needs assessment of how supplier management practices and IS integration strategy tied together to help a company with their supply chain performance practices. The adoption of the RFID-related technology in the healthcare industry has historically lagged behind as compared to other industries in the manufacturing and retail industries, perhaps due in part to certain ethical dilemmas. Many leaders and decision makers of the healthcare industry have begun to realize the role that this technology can positively benefit the industry by maintaining focus on quality care, while doing so by meeting the pressures from competition and achieving business and performance goals (Smith, 2010). The introduction of information technology (IT) into the healthcare industry led to more efficient tasks and processes, a higher quality of healthcare services, and increased accessibility to healthcare providers (Jones, Bekeris, Nakhleh, Walsh, & Valenstein, 2009; Kjøs, Botten, & Romøren, 2008; Palmon & Sudit, 2009; Santouridis, Trivellas, & Reklitis, 2009; Sauerland, 2009; Scanlon, Swaminathan, Lee, & Chernew. 2008; Smith, Casey, Hurst, Fenton, & Scholefield, 2009).

The organizational level of the IT adoption research focuses on the understanding of the diffusion and adoption process. There have been numerous studies that have tried to include as many distinctive characteristics of context as possible in the development phase of organizational IT adoption theories (Domi-

nic, Goh, Wong, & Chen, 2010; Kapur, Gupta, Jha, & Goyal, 2010; Keramati & Behmanesh, 2010; Kim & Tadisina, 2010). Also, researchers have tried to identify the impacts of key categories (i.e., individual, technical, structural, environmental factors, task-related factors) in the adoption process. A number of the previously cited studies that have been done have found that the characteristics of organizations are very significant of the IT adoption process. Another force driving the decision to adopt the RFID process is the environmental factor, “environmental uncertainty, competitive pressure, industrial pressure, and government policy all serve as pressures on organizations (Lee & Shim, 2007, p.714).” Another concept used when deciding whether to adopt the RFID process is the concept of technology-push/ need-pull (TP/ NP); “the technology-push model is based on the view that a new scientific discovery will trigger events ending with diffusion or application of the discovery” (Lee & Shim, 2007, p. 714).

The likelihood of total adoption RFID-related technologies within healthcare organizations is unclear but, it is looking up. Research has showed that RFID is still in the early stages of adopting in the healthcare industry. In a survey done out of 126 organizations (Kumar, Swanson, & Tran, 2009), only 4 are currently using the RFID systems and almost half of those are still planning to evaluate their situation. Also, the results of the survey showed that the adoption rate of the RFID technology is expected to increase since 22 out of 23 that have evaluated RFID processes would adopt it. Applications of RFID-related technologies are inherently very valuable, but some managers feel it is at a great cost to integrate it into a business, at least during their tenure as manager. Healthcare organizations may reap important benefits from RFID application, since it automates the tracking of medical equipment, patients, medical supplies, nurses, and drug usage. These benefits should result in the increase of RFID technology to more than US\$2.1 billion in 2016 (“Healthcare IT outsourcing market ...,” 2014).

Challenges in RFID Application in Healthcare

Even though there are many benefits that should encourage the implementation of RFID technologies in healthcare organizations, the total cost consists of multiple factors that management has to consider before introducing the new system. Adoption and implementation, in addition to maintenance and the training of employees, incur costs. According to a recent case study, the cost of each aspect varies among supply chain partners (Chircu, et al. 2014). RFID-related applications in healthcare involve several benefits; however, there are several challenges that are faced with such a complicated system. Perhaps the most influential challenge are the various costs put into the system, in fact, the high costs of implementing the system into healthcare is what kept the system away from healthcare for so long and kept within commercial applications (Mehrjerdi, 2010, 2012; Smith, 2010).

The concern in the long-term investment is the chance of failure after the expenses have been incurred. It is more likely in RFID for healthcare, but the payoff in RFID is never guaranteed, which is a major concern in adoption in other fields (Erdem, Zeng, Zhou, Shi, & Wells, 2009; Wickboldt, et al., 2012). Mehrjerdi (2011) stated that products that are US \$15 or less may not be worth the costs or return on investment because of the unpromising performance a product would deliver at that low of a price. Infrastructure can also be complex which delays a process that requires patience that some may not think is worth it, since technology is constantly changing and systems are continuously being altered as more information is compiled.

Patient privacy and security is a major challenge in applying RFID to healthcare. Consumer protection programs such as CASPIAN are being implemented because of the possible invasion of privacy to patients (Kapoor, et al., 2009). Patient information is embedded deep within systems;

therefore, patients may be left with a sense of insecurity of their personal information and skeptical of buying into the system. Not only is patient information vulnerable, but also the data concerned with the organization must ensure safety (Wickboldt, et al., 2012). On-the-other-hand, there is a challenge faced with those in control of the RFID trust the system more than the patient, even when that should not be the case. A trusting relationship whether it be between a person and systems, such as patients and physicians trust in RFID renders themselves “vulnerable” (Entwistle & Quick, 2006). Once a healthcare provider has earned the trust from a patient, it is vital that the provider should work even harder to provide the patient with the safety he/he deserves (Entwistle, McCaughan, Watt, Birks, Hall, 2010). Those serving patients with RFID assistance must buy-in to the system; however, the trust must be rational and suitable. Certain situation information is most accurate when it comes from the patient, and if trust is placed in the wrong place, a situation where a patient can contribute the most effectively to their own safety, may be overlooked.

Such as an RFID “VeriMed” system can be thought to attain the “most accurate information” about a patient; however, if the system interpreted information incorrectly such as medications and allergies and the provider trusts the system more than the patient, they may not give it a second look (Monahan, et al. 2011, p.373). Even though a healthcare provider can be well educated and pertain all qualifications for their position there can be differences in the knowledge of the person and the system. Varied abbreviations could be used in a particular organization, and if a provider does not realize that a new system uses a different abbreviation for a medication, it could lead to an overdose or delivery of an incorrect medication (Entwistle, et al., 2006). Therefore, if a doctor trusts the technology to an extent that information within the system is never questioned, then there is the potential for serious medical errors (Fisher & Monahan, 2010).

CONCLUSION

Hopefully, after reading through this chapter, the reader will receive a better understanding of the role that an IS-based strategy, utilizing RFID, and various healthcare supplier management practices do play a significant role on the supply chain performance for a company. Healthcare organizations need to partner and integrate such technologies and medical suppliers in order to strategically leverage each other's strengths to improving the supply chain and better healthcare delivery and related outcomes. A supply chain manager may be very knowledgeable about operational and technological practices to help a company save money by inventory control and production planning, but with the help of RFID-embedded technologies, it has the potential to make the healthcare system operationally run more efficient and effective.

FUTURE DIRECTIONS

Overall the outlook for RFID acceptance is very positive when contemplating the use in supply chain management of medical supplies and pharmaceuticals. Pertinent patient information could prove to be the biggest obstacle to overcome in taking steps of implementation. Managers, healthcare providers, and IT professionals must keep patient concerns in mind when making decisions, conducting training, and processing information. Patients must also be open to the use of RFID in order to reduce medical errors. The reduction of these errors can be critical and life-saving. The use of these case studies will continue to provide evidence supporting the use of RFID in healthcare.

REFERENCES

- Acharyulu, G. (2007). RFID in the healthcare supply chain: Improving performance through greater visibility. *ICFAI Journal of Management Research*, 6(11), 32–45.
- Acharyulu, G. V. R. K. (2011). Information management in a healthcare system: Knowledge management perspective. *International Journal of Innovation. Management and Technology*, 2(6), 534.
- Aldaihani, M. M., & Darwish, M. A. (2013). Optimal production and inventory decisions for supply chains with one producer and multiple newsvendors. *International Journal of Services and Operations Management.*, 15(4), 430–448. doi:10.1504/IJSOM.2013.054884
- Anand, A., & Wambaa, S. F. (2013). Business value of RFID-enabled healthcare transformation projects. *Business Process Management Journal*, 19(1), 111–145. doi:10.1108/14637151311294895
- Azadeh, A., Gholizadeh, H., & Jeihoonian, M. (2013). A multi-objective optimisation model for university course timetabling problem using a mixed integer dynamic non-linear programming. *International Journal of Services and Operations Management.*, 15(4), 467–481. doi:10.1504/IJSOM.2013.054886
- Bhamu, J., Khandelwal, A., & Sangwan, K. S. (2013). Lean manufacturing implementation in an automated production line: A case study. *International Journal of Services and Operations Management.*, 15(4), 411–429. doi:10.1504/IJSOM.2013.054883
- Carr, A. S., Zhang, M., Kloppig, I., & Min, H. (2010). RFID technology: Implications for healthcare organizations. *American Journal of Business*, 25(2), 25–40. doi:10.1108/19355181201000008

- Chen, C. C., Wu, J., Su, Y. S., & Yang, S. C. (2008). Key drivers for the continued use of RFID technology in the emergency room. *Management Research News*, 31(4), 273–288. doi:10.1108/01409170810851348
- Chircu, A., Sultanow, E., & Saraswat, S. P. (2014). Healthcare RFID in Germany: An integrated pharmaceutical supply chain perspective. *Journal of Applied Business Research*, 30(3), 737. doi:10.19030/jabr.v30i3.8559
- de Vries, J., & Huijsman, R. (2011). Supply chain management in health services: An overview. *Supply Chain Management*, 16(3), 159–165. doi:10.1108/13598541111127146
- DesRoches, C. M., Campbell, E. G., Rao, S. R., Donelan, K., Ferris, T. G., Jha, A., & Blumenthal, D. et al. (2008). Electronic health records in ambulatory care: national survey of physicians. *The New England Journal of Medicine*, 359(1), 50–60. doi:10.1056/NEJMs0802005 PMID:18565855
- Devaraj, S., Ow, T. T., & Kohli, R. (2013). Examining the impact of information technology and patient flow on healthcare performance: A Theory of Swift and Even Flow (TSEF) perspective. *Journal of Operations Management*, 31(4), 181–192. doi:10.1016/j.jom.2013.03.001
- Dominic, P. D. D., Goh, K. N., Wong, D., & Chen, Y. Y. (2010). The importance of service quality for competitive advantage – with special reference to industrial product. *International Journal of Business Information Systems*, 6(3), 378–397. doi:10.1504/IJBIS.2010.035051
- Dutta, A., Lee, H. L., & Whang, S. (2007). RFID and operations management: Technology, value and incentives. *Production and Operations Management*, 16(5), 646–655. doi:10.1111/j.1937-5956.2007.tb00286.x
- Entwistle, V. A., McCaughan, D., Watt, I. S., Birks, Y., & Hall, J. (2010). Speaking up about safety concerns: Multi-setting qualitative study of patients' views and experiences. *Quality & Safety in Health Care*, 19(6), 1–7. PMID:21127092
- Entwistle, V. A., & Quick, O. (2006). Trust in the context of patient safety problems. *Journal of Health Organization and Management*, 20(5), 397–416. doi:10.1108/14777260610701786 PMID:17087402
- Fisher, J. A., & Monahan, T. (2008). Tracking the social dimensions of RFID systems in hospitals. *International Journal of Medical Informatics*, 77(3), 176–183. doi:10.1016/j.ijmedinf.2007.04.010 PMID:17544841
- Haugh, R. (2006). Supply chain management. *Hospitals & Health Networks*, 5(3), 20–24, 28. PMID:16615635
- Healthcare IT outsourcing market by application & industry - global forecast to 2018. (2014). *PR Newswire*. Retrieved March 8, 2016 from <http://search.proquest.com.ezproxy.lib.apsu.edu/docview/1512339674?accountid=8437>
- Hu, G., Wang, L., Fetch, S., & Bidanda, B. (2008). A multi-objective model for project portfolio selection to implement lean and Six Sigma concepts. *International Journal of Production Research*, 46(23), 6611–6648. doi:10.1080/00207540802230363
- Ifinedo, P., & Nahar, N. (2009). Interactions between contingency, organizational IT factors, and ERP success. *Industrial Management & Data Systems*, 109(1), 118–126. doi:10.1108/02635570910926627
- Jain, V., Benyoucef, L., & Deshmukh, S. G. (2008). What's the buzz about moving from lean to agile integrated supply chains? A fuzzy intelligent agent-based approach. *International Journal of Production Research*, 46(23), 6649–6678. doi:10.1080/00207540802230462

- Jones, B. A., Bekeris, L. G., Nakhleh, R. E., Walsh, M. K., & Valenstein, P. N. (2009). Physician satisfaction with clinical laboratory services: A College of American Pathologists Q-Probes study of 138 institutions. *Archives of Pathology & Laboratory Medicine*, 133(1), 38–43. PMID:19123734
- Kapur, P. K., Gupta, A., Jha, P. C., & Goyal, S. K. (2010). Software quality assurance using software reliability growth modelling: State of the art. *International Journal of Business Information Systems*, 6(4), 463–496. doi:10.1504/IJBIS.2010.035742
- Keramati, A., & Behmanesh, I. (2010). Assessing the impact of information technology on firm performance using canonical correlation analysis. *International Journal of Business Information Systems*, 6(4), 497–513. doi:10.1504/IJBIS.2010.035743
- Kim, E., & Tadisina, S. (2010). A model of customers initial trust in unknown online retailers: An empirical study. *International Journal of Business Information Systems*, 6(4), 419–443. doi:10.1504/IJBIS.2010.035740
- Kjøs, B. Ø., Botten, G., & Romøren, T. I. (2008). Quality improvement in a publicly provided long-term care system: The case of Norway. *International Journal for Quality in Health Care*, 20(6), 433–438. doi:10.1093/intqhc/mzn031 PMID:18653584
- Kumar, S., Swanson, E., & Tran, T. (2009). RFID in the healthcare supply chain: Usage and application. *International Journal of Health Care Quality Assurance*, 22(1), 67–81. doi:10.1108/09526860910927961 PMID:19284172
- Lee, C., & Shim, J. P. (2007). An exploratory study of radio frequency identification (RFID) adoption in the healthcare industry. *European Journal of Information Systems*, 16(6), 712–724. doi:10.1057/palgrave.ejis.3000716
- Ludwick, D. A., & Doucette, J. (2009). Adopting electronic medical records in primary care: Lessons learned from health information systems implementation experience in seven countries. *International Journal of Medical Informatics*, 78(1), 22–31. doi:10.1016/j.ijmedinf.2008.06.005 PMID:18644745
- Marthandan, G., & Tang, C. M. (2010). Information systems evaluation: An ongoing measure. *International Journal of Business Information Systems*, 6(3), 336–353. doi:10.1504/IJBIS.2010.035049
- Mehrjerdi, Y. Z. (2010). RFID-enabled healthcare systems: Risk-benefit analysis. *International Journal of Pharmaceutical and Healthcare Marketing*, 4(3), 282–300. doi:10.1108/17506121011076192
- Mehrjerdi, Y. Z. S. (2011). RFID and its benefits: A multiple case analysis. *Assembly Automation*, 31(3), 251–262. doi:10.1108/01445151111150596
- Palmon, D., & Sudit, E. F. (2009). Commercial insurance of financial disclosure: Auditors independence, and investors protection. *Group Decision and Negotiation*, 18(1), 27–40. doi:10.1007/s10726-008-9120-3
- Qrunfleh, S., Tarafdar, M., & Ragu-Nathan, T. (2012). Examining alignment between supplier management practices and information systems strategy. *Benchmarking*, 19(4), 604–617. doi:10.1108/14635771211258034
- Reiner, J., & Sullivan, M. (2005). RFID in healthcare. *Healthcare Purchasing News*, 29(6), 74–76.
- Revere, L., Black, K., & Zalila, F. (2010). RFIDs can improve the patient care supply chain. *Hospital Topics*, 88(1), 26–31. doi:10.1080/00185860903534315 PMID:20194108

- Santouridis, I., Trivellas, P., & Reklitis, P. (2009). Internet service quality and customer satisfaction: Examining Internet banking in Greece. *Total Quality Management and Business Excellence*, 20(2), 223–238. doi:10.1080/14783360802623084
- Sauerland, D. (2009). The legal framework for health care quality assurance in Germany. *Health Economics, Policy, and Law*, 4(1), 79–98. doi:10.1017/S1744133108004647 PMID:19099618
- Scanlon, D. P., Swaminathan, S., Lee, W., & Chernew, M. (2008). Does competition improve health care quality? *Health Services Research*, 43(6), 1931–1944. doi:10.1111/j.1475-6773.2008.00899.x PMID:18793214
- Smith, A. A., Smith, A. D., & Baker, D. J. (2011). Inventory management shrinkage and employee anti-theft approaches. *International Journal of Electronic Finance*, 5(3), 209–234. doi:10.1504/IJEF.2011.041337
- Smith, A. D. (2011). Component part quality assurance concerns and standards: Comparison of world-class manufacturers. *Benchmarking: An International Journal*, 18(1), 128–148. doi:10.1108/14635771111109850
- Smith, A. D. (2012). Operational efficiencies and ethical issues associated with microchip implants. *International Journal of Mobile Communications*, 10(3), 281–302. doi:10.1504/IJMC.2012.048113
- Smith, A. D. (2013a). Quality and technology-based management initiatives within a hospital environment: Benchmarking case study. *International Journal of Procurement Management*, 6(6), 621–648. doi:10.1504/IJPM.2013.056746
- Smith, A. D. (2013b). Competitive uses of information and knowledge management tools: Case study of supplier-side management. *International Journal of Knowledge-Based Organizations*, 3(1), 71–87. doi:10.4018/ijkbo.2013010105
- Smith, A. D. (2013c). Perceived value and strategic importance of supply chain options: Multi-firm case study. *International Journal of Logistics Systems and Supply Management*, 13(2), 244–267. doi:10.1504/IJLSM.2012.048938
- Smith, A. D., & Rupp, W. T. (2013). Data quality and knowledge/information management in service operations management: Regional supermarket case study. *International Journal of Knowledge-Based Organizations*, 3(3), 35–52. doi:10.4018/ijkbo.2013070103
- Smith, S., Casey, A., Hurst, K., Fenton, K., & Scholefield, H. (2009). Developing, testing and applying instruments for measuring rising dependency-acuity's impact on ward staffing and quality. *International Journal of Health Care Quality Assurance*, 22(1), 30–39. doi:10.1108/09526860910927934 PMID:19284169
- Southard, P. B., Chandra, C., & Kumar, S. (2012). RFID in healthcare: A six sigma DMAIC and simulation case study. *International Journal of Health Care Quality Assurance*, 25(4), 291–321. doi:10.1108/09526861211221491 PMID:22755482
- van der Togt, R., Jan van Lieshout, E., Hensbroek, R., Beinat, E., Binnekade, J. M., & Bakker, P. J. M. (2008). Electromagnetic interference from radio frequency identification inducing potentially hazardous incidents in critical care medical equipment. *Journal of the American Medical Association*, 299(24), 2884–2890. doi:10.1001/jama.299.24.2884 PMID:18577733
- Zare, L., McMullen, K. A., & McCune, C. G. (2014). *Design of an accessible and portable system for soccer players with visual impairments*. Association for Computing Machinery. Retrieved March 8, 2016 from <http://dl.acm.org/citation.cfm?id=2581357>

ADDITIONAL READING

Barlow, R. D. (2010). The missing links between supply chain data, EHR/EMR software. *Healthcare Purchasing News*, 34(3), 10-12, 14-15.

Smith, A. D. (2011). Operational effectiveness and quality assurance mechanisms with stochastic demand of blood supply: Blood bank case study. *International Journal of Electronic Healthcare*, 6(2-4), 174–191. doi:10.1504/IJEH.2011.044348 PMID:22189177

Smith, A. D. (2012). Case studies of RFID-related applications in the healthcare and voice-recognition industries. *International Journal of Knowledge-Based Organizations*, 2(3), 48–63. doi:10.4018/ijkbo.2012070103

Ting, S. L., Kwok, S. K., Tsang, A. H. C., & Lee, W. B. (2009). Critical elements and lessons learnt from the implementation of an RFID-enabled healthcare management system in a medical organization. *Journal of Medical Systems*, 35(4), 657–669. doi:10.1007/s10916-009-9403-5 PMID:20703523

Visich, J. K., Li, S., Khumawala, B. M., & Reyes, P. M. (2009). Empirical evidence of RFID impacts on supply chain performance. *International Journal of Operations & Production Management*, 29(12), 1290–1315. doi:10.1108/01443570911006009

Wambaa, A., Ananda, A., & Carter, L. (2013). A literature review of RFID-enabled healthcare applications and issues. *International Journal of Information Management*, 33(5), 875–891. doi:10.1016/j.ijinfomgt.2013.07.005

Wickboldt, A.-K., & Piramuthu, S. (2012). Patient safety through RFID: Vulnerabilities in recently proposed grouping protocols. *Journal of Medical Systems*, 36(2), 431–435. doi:10.1007/s10916-010-9487-y PMID:20703708

Wu, I., & Hu, Y. (2012). Examining knowledge management enabled performance for hospital professionals: A dynamic capability view and the mediating role of process capability. *Journal of the Association for Information Systems*, 13(12), 976–999.

Zang, C., & Fan, Y. (2007). Complex event processing in enterprise information systems based on RFID. *Enterprise Information Systems*, 1(1), 3–23. doi:10.1080/17517570601092127

KEY TERMS AND DEFINITIONS

Automatic Identification and Data Capture Technologies (AIDC): Types of AIDC-related technologies to leave the human element out of the data collection and storage functions of information derived from manufacturing, integrated through the manufacturing process, types of authentication concerns and/or e-security strategies, and relationship links to customer profiles.

Barcoding Technology: A long-term and very reliable type of AIDC technology, it is known for its very accurate and economical approaches to identity products and machine readable information from a variety of manufactured goods and services. Most barcodes use a type of standardized bars and spacing coding or symbology that is certified by an international standards body (GS1 System).

Ethical Dilemmas: There are a number of ethical theories that are appropriate for dealing with healthcare issues concerning automatic identification. These theories involve both individual and group behaviors that are grounded in moral philosophy, especially in the concepts of consequentialism or deontology ethical philosophy.

Healthcare Service Strategies: When it comes to the healthcare industry, an overall strategic goal is to provide affordable coverage and quality service to all citizens that need such services at affordable costs.

Operations Efficiency: Improving efficiency and reducing waste is a major challenge for hospitals and other patient care facilities looking to lower the cost of providing healthcare services.

RFID-Embedded Technologies: RFID technologies are types of automatic data capture techniques that use a combination of active and passive senders and receivers to collect and store codified information for further uses.

Supply Chain Management/Performance: In basic terms, supply chain is the system of organizations, people, activities, information and resources involved in moving a product or service from supplier to customer. The configuration and management of supply chain operations is a key way companies obtain and maintain a competitive advantage.

Vendor-Managed Inventory Systems (VMI): VMI-based systems are designed to transfer the control of inventory and its planning activities to a manufacturer or distributor in order to provide a beneficial relationship to promote a more transparent and seamless flow of goods and services at lower costs.

Virtual Asset Trackers (VAT): VAT is a type of application of active RFID tagging that allows healthcare practitioners to enhance security to patients, residents, and hospital staff while providing better patient and resident-care services.

M

Social Telerehabilitation

Gilberto Marzano

Rezekne Academy of Technologies, Latvia

INTRODUCTION

In a relatively short time, information and communication technologies (ICT) have spread worldwide, from defense and space exploration to large industrial applications, and to the worlds of commerce, education, and entertainment. ICT is changing people's daily lives, the way they work, buy, sell, and learn, and also the way that services are run in the healthcare sector. Consequently, *e-health*, *telehealth*, and *telemedicine* are now terms that are commonly used in this sphere, encompassing three main computer-assisted areas in healthcare, namely: clinical assessment, diagnosis, and therapy. E-health comprises the monitoring of patients' health, the promotion of good practices, and the prevention and treatment of health conditions by electronic means, as well as the provision of online access to literature and medical knowledge. One of the first reviews of telehealth is that provided by Winters (2002), who identified two major subsets of telehealth: telemedicine (i.e., delivery of clinical services) and telehealthcare (i.e., management of disability and health).

Nowadays, there is a progressive specialization within the field of e-health, and the specificity of the different branches appears to be increasingly sharp.

Telerehabilitation is a recent emerging field in telehealthcare area and this chapter presents social telerehabilitation, a new specialized sector of telerehabilitation.

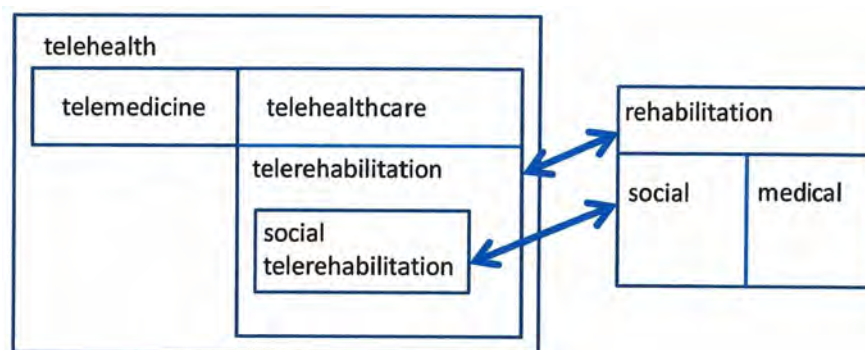
BACKGROUND

Before to deal with social telerehabilitation, the more general notions of rehabilitation and telerehabilitation ought to be introduced.

In fact, social telerehabilitation, which focuses on solving limitations and social issues associated with health conditions, represents a further specialization in telerehabilitation. Figure 1 shows the position of social telerehabilitation within the general framework of telehealth.

It has to be noted that social telerehabilitation results from a change in the concept of rehabilitation, which has seen a shift in view point away from a predominantly medical one, towards an

Figure 1. Social telerehabilitation within the telehealth general framework



DOI: 10.4018/978-1-5225-2255-3.ch516

increasingly complex one in which psychological and socio-cultural aspects are deemed to be of equal importance (Brown & Hughson, 1993; Wade & de Jong, 2000; Altman, Swick, Parrot & Malec, 2010; Karkou, Martinsone, Nazarova & Vaverniece, 2011).

The Notion of Rehabilitation

In general, rehabilitation encompasses any services or providers that are directed to the reduction of impairments, activity limitations, or social participation restrictions experienced by an individual (Figure 2). The World Health Organization describes rehabilitation for people with disabilities as being a process aimed at enabling them to reach and maintain their optimal physical, sensory, intellectual, psychological, and social functional levels.¹

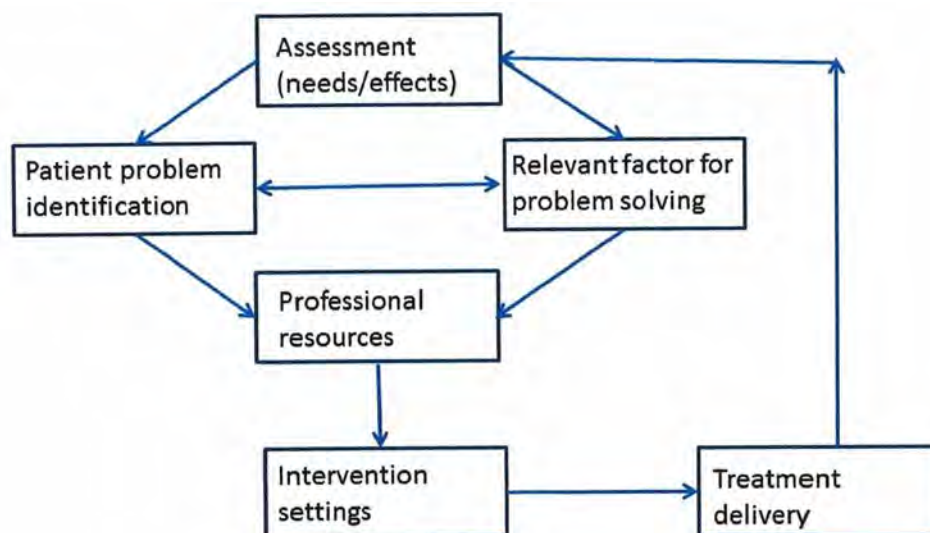
Generally, two different forms of rehabilitation can be distinguished: *medical* and *social*, or, in the latter case, also referred to as *vocational*.

Both forms are aimed at developing the functional and psychological abilities of the individual and, if necessary, his/her compensatory mechanisms, in order to enable him/her to attain self-dependence and lead an active life. The main

difference between them is that medical rehabilitation is more restricted to curative medicine, and involves intensively trained clinicians and different health professionals, especially physiotherapists, whilst social rehabilitation focuses more on the individuals' social sphere, and includes services such as rehabilitative nursing, occupational therapy, speech and language therapy, audiology, dietetics, prosthetics and orthotics, podiatry, art therapy, music therapy, and social work. However, although social rehabilitation is directed, for example, at overcoming barriers for people to access, maintain, or return to employment or other useful occupations (ILO, 2008; Kuoppala & Lamminpää, 2008), it substantially differs from strategies and policies aimed at promoting and developing the inclusion of people with disabilities in the work place. Offender rehabilitation (Robinson & Crow, 2009) and addict rehabilitation (Wilson *et al.*, 2012), for example, are considered to be forms of social rehabilitation that focus on the development of practices and programs for helping individuals to retake their place in society.

Many social scientists and policy makers share the opinion that community-based rehabilitation practices are a part of general medical care (Haig, 2013). As a consequence of this, it is not always

Figure 2. The rehabilitation process



possible to trace a line of demarcation between clinical and social rehabilitation, and sometimes there is an overlapping relation between these two forms of rehabilitation. A case in point is the *Report on Disability* (WRD) of June 9th 2011, published by the World Health Organization (WHO) and World Bank (WB), which outlined concrete actions that befit both medical and social rehabilitation.

Such concrete actions intended to help overcome barriers to the provision of rehabilitation services (WHO & WB, 2011) include:

- Increasing human resources for rehabilitation (training, retaining, and retention of rehabilitation personnel);
- Expanding and decentralizing service delivery;
- Increasing the use and affordability of technology and assistive devices.

It has been observed that the WRD report presents a challenge to medical rehabilitation (Bethge, von Groote, Giustini & Gutenbrunner, 2014) and, although it doesn't provide a fixed definition of rehabilitation, its description of the concept represents a valuable frame of reference that can help to foster a common understanding of the rehabilitation scope and rehabilitation services (Meyer, Gutenbrunner, Bickenbach, Cieza, Melvin & Stucki, 2011). In fact, many authors cite in their works the WRD definition of rehabilitation, that being: "a set of measures that assist individuals who experience, or are likely to experience, disability to achieve and maintain optimal functioning in interaction with their environments" (Meyer et al., 2014; Lynggaard, May, Beauchamp, Nielsen & Wittrup, 2014; Gibson, 2016).

It is important to note that the WRD definition recalls that of Article 26, entitled *Habilitation and Rehabilitation*, of the United Nations Convention on the Rights of Persons with Disabilities (CRPD, 2006):

*[...] appropriate measures, including through peer support, to enable persons with disabilities to attain and maintain their maximum independence, full physical, mental, social and vocational ability, and full inclusion and participation in all aspects of life.*²

The above definitions fit well with both medical and social rehabilitation, as well as with telerehabilitation that includes social telerehabilitation.

Telerehabilitation

The term "telerehabilitation" is a relatively new one, although it has been observed that the first experiment with remote medical interventions dates back to the 1880s, following the invention of the telephone (Scalvini, Vitacca, Paletta, Giordano & Balbi, 2004).

Until a few years ago, telerehabilitation applications were considered within the wider field of e-health, and only recently telerehabilitation has been assumed to have an autonomous, separate, and parallel identity under the telehealth umbrella, within telehealthcare (Schmeler, Schein, McCue & Betz, 2009).

Essentially, telerehabilitation comprises methods of delivering rehabilitation services using ICT to minimize the barriers of distance, time, and cost, and, more specifically, has been defined "as the application of telecommunication, remote sensing and operation technologies, and computing technologies to assist with the provision of medical rehabilitation services at a distance" (Cooper et al., 2001).

Social Telerehabilitation Definition and Scope

Both telerehabilitation and social telerehabilitation are grounded in the delivery of rehabilitation services through telecommunication networks, especially by means of the Internet. Likewise,

one can define social telerehabilitation as being the application of ICT to provide equitable access to social rehabilitation services, at a distance, to individuals who are geographically remote, and to those who are physically and economically disadvantaged.

The scope of social telerehabilitation is vast, since it faces challenges related to both medical and community care settings (Hill, 2010). Social telerehabilitation aims at helping people regain their psycho-physical functions and improving their daily quality of life through the use of ICT and mobile technologies (Reeder, Chung & Stevens-Lapsley, 2016; Kamwesiga, Tham & Guidetti, 2016). In this regard, research has underlined the potential and effectiveness for social media, mobile phones, and the Internet in general, to improve social rehabilitation processes.

Expanding and Decentralizing Rehabilitation Services

Telerehabilitation and social telerehabilitation share the same goal: to expand and decentralize rehabilitation service delivery.

In fact, both medical and social rehabilitation services are generally located in urban centers and health care structures, and this hinders their access by people with disabilities or those living in rural areas. In many cases, the burden of transportation, as well as the mobility assistance for persons with temporary or chronic diseases, represents a serious barrier that limits the advantages of rehabilitation services. Parking, distance, or transportation difficulties can significantly interfere with the receiving of treatments. In fact, the more frequent the treatments are, the more the mobility burden increases.

Furthermore, in the majority of cases, rehabilitation, either medical or social, involves multidisciplinary professional competence and expertise (Neumann et al., 2010; Stanos, 2012; Malladi, 2015). This means that the planning activity and the actual treatments involve teams of professionals (physiotherapists, psychologists,

social educators, etc.) who must work together in order to achieve common and complementary goals. Their activity is directed to resolving the patient problems, following an iterative, active, educational, and problem-solving process that focuses on the patient's behavior and adapts to it (Marzano, Lubkina & Rizakova, 2015). The most common goals of the rehabilitation process are:

- Maximizing the participation of the patient in his/her social setting;
- Minimizing the pain and distress experienced by the patient;
- Minimizing the distress of, and the stress on, the patient's family and caregivers.

Accordingly, telerehabilitation is particularly useful when the coordination of rehabilitation service delivery requires the improvement and expansion of the availability of coordinated multidisciplinary rehabilitation in an effective care framework.

Both social and medical telerehabilitation services are generally provided on an outpatient basis, usually at home and under the supervision of at-a-distance professionals. Experts check patient evolution through regular online meetings or by analyzing data from sensors and equipment. Depending on the telerehabilitation programs, there are many kinds of assistive devices whose cost is lowered by mass production, e.g. there are many very affordable teleconference systems, and mobile phones are used for remote data collection, control of equipment, and telemonitoring. Supervisors can help the patient to perform exercises during the day-to-day rehabilitation sessions, suggesting the correct execution.

In the last few years, many researchers and practitioners have argued the importance of social telerehabilitation services for patients, caregivers, and public institutions, for instance in regard to services for children suffering from physical handicaps and emotional disturbances (Nilsson & Nilsson, 2011b), alcoholics (Simpson, T. L., Kivlahan, Bush & McFall, 2005; Rose, Skelly,

Badger, Naylor & Helzer, 2012), and for patients with affective disorders, such as depression and manic-depression or bipolar disorder (Sachpazidis & Sakas, 2008; Acampora, Cook, Rashidi & Vasilakos, 2013).

Research has underlined the potential for social media, mobile phones, and the Internet in general, to improve mental and physical health, treat addictions, and also to help individuals experiencing homelessness (Quan, Joseph, Keller & Arch, 2011; McInnes, Fix, Solomon Petrakis, Sawh & Smelson, 2015). Furthermore, Artificial Intelligence is discovering new opportunities to overcome the limitations of the computational approach and, consequently, making possible new advanced healthcare applications, especially in the social-healthcare (Pravettoni, Folgieri & Lucchiari, 2015).

Application of Social Telerehabilitation

It has been underlined that social telerehabilitation tackles some issues that set it apart from the broader telehealthcare arena. For example, it has been observed that one challenge is that social rehabilitation is often provided across both medical and community settings, often with different funding structures and rehabilitation protocols in place (Scaffa & Reitz, 2013; Camicia, Black, Farrell, Waites, Wirt & Lutz, 2014).

From the analysis of the literature, three principal areas of interest emerge in terms of the development of social telerehabilitation services:

1. The psycho-physical area;
2. The social-communication area;
3. The behavioral area (e.g. behavioral addictions).

Applications are available for health education and counseling programs, many of them offered to myocardial infarction patients during and after hospitalization. During hospitalization, the inter-

ventions consist of individual counseling sessions and group health education sessions focusing on medication, healthy habits, and anxiety and depression. After discharge from hospital, patients are checked at home by telephone or via the Internet, and are encouraged to perform exercises and physical activity. Social telerehabilitation programs demonstrate their effectiveness in the areas of decreasing unhealthy eating habits, cessation of smoking, and treating anxiety and depression of patients following a stroke (Beatty, Fukuoka & Whooley, 2013; Varnfield et al., 2014).

A Typical Social Telerehabilitation Service for Elderly Who Need Physical Exercises

To better understand how social telerehabilitation processes operate, we can illustrate a typical example of a telerehabilitation service that concerns physical exercises for elderly people.

The first step is the identification of needs and the recognition of related problems: the social rehabilitators' team performs the evaluation of the elderly person conditions in order to verify his/her capability to benefit from the service. Professionals engaged in the rehabilitation program use teleconference and/or phone communication to discuss the case and define the rehabilitation program, and to generate its electronic workflow. Following this, the trainer demonstrates exercises to the patient, and answers his/her queries and clears his/her doubts. This step can be performed either at a distance or face-to-face. The trainer then conducts videoconferencing sessions for the progression of exercises, supervision of program completion, additional demonstrations, and consultations. As may be the case, exercises might include flexibility, strength, endurance, and other types of active motion activities. For videoconferencing, a professional videocamera and Skype software with a broadband internet facility can be used. The patient is regularly engaged in exercises and followed by the same trainer, who

will give him/her instructions, e. g. in order to adjust the video camera for each exercise, number of repetitions, etc.

Online delivery of social rehabilitation services encourages the rethinking, reorganization, optimization, and extension of the current social rehabilitation services that fall under face-to-face treatments and protocols. Of course, this can require the retraining of rehabilitators

In general, a social telerehabilitation system should provide patients and caregivers suitable mechanisms to interact with each other as if they were meeting face-to-face. Furthermore, patients should have at their disposal all the information about the schedule of the rehabilitation plan, a full description of the rehabilitation activities, and real-time feedback about their performance. For rehabilitators, the system should gather information about the fulfillment of the program, and the correctness of the patients' performance. This will allow a quick and ubiquitous assessment of the patients' evolution, and an adjustment of their rehabilitation program if necessary.

A social telerehabilitation system includes:

- Computers and other general-purpose devices (i.e. tablets, laptops, videocameras, and so on), for both patients and rehabilitators, which are used to run the applications.
- Different types of sensors used to monitor the patient's performance; the simplest of them being low-cost wearable motion capture devices (e.g. accelerometers that gather values of linear acceleration along three perpendicular axes (x, y, z)).
- Specific software application, e.g. for social-cognitive exercises, speech therapy, etc.

FUTURE RESEARCH DIRECTIONS

The idea of exploiting ICT in psychosocial care is not a novelty. Eliza, the computer program created by Weizenbaum between 1964 and 1966 (Weizen-

baum, 1966), can be considered to have been a primitive attempt at social telerehabilitation. It was a tongue-in-cheek application, which emulated the responses of a Rogerian non-directional psychotherapist during an initial psychiatric interview. Since then, much progress has been made to the point that social telerehabilitation now represents a topical field of research.

Almost 50 years have passed between the creation of the Eliza program and the appearance of the recent programs designed for rehabilitation services. The difference between the two is that Eliza was one of the first experimental programs in the field of natural language processing, whilst the new electronic applications are able to utilize the achievements made in natural language processing for communicating with patients. The focus is now on programming machines for empathy emulation and ethical therapeutic activities. One of the next future challenges is to provide robotic caregivers for social rehabilitation purposes.

In fact, thanks to the investigations that have started in the last fifteen years on different forms and roles of affection in virtual agents and robots (Scheutz, 2011), some researchers are now considering the employment of robots in social rehabilitation services (Tickle-Degnen, Scheutz & Arkin, 2014). The new robots could provide moral functionalities that enable them to establish therapeutic human relationships, for example in the context of occupational therapy and Parkinson's or Alzheimer's disease (Salichs, Encinar, Salichs, Castro-González & Malfaz, 2015). At the simplest level, a robot could ensure basic functional elements of interpersonal communication. On a more sophisticated level, they could act as *observers* that accurately detect and prioritize people, objectives, and context attributes relevant to reasoned and ethical therapeutic interaction (Arkin, Scheutz & Tickle-Degnen, 2014).

Furthermore, the applications of wearable technology to monitor older adults and subjects with chronic conditions at home and in community settings are multiplying, while the integration of wearable and ambient sensors is expected to make

Figure 3. Future caregiver robot

Future caregiver robot	
Key characteristics	
•	Embodiment
	Representation independent
	Situadness (social behavior)
	Emergent complexity (collective action of different agents)
	Real time responses

relevant progress (Patel, Park, Bonato, Chan & Rodgers, 2012).

Finally, behavior-based caregiver robots represent a social telerehabilitation new advanced research area that presumes interaction among a machine and a human patient. The keystone concepts behind the artificial behavior-based caregivers are showed in Figure 3.

However, at next future, we are left with some as yet unresolved questions concerning social telerehabilitation:

1. To what measure can social telerehabilitation replace traditional face-to-face social rehabilitation therapies?
2. Which are the priority sectors for intervention?
3. Is social telerehabilitation in these sectors really economical?
4. What is the educational effort and cost that would be required to diffuse knowledge about social telerehabilitation practices?

A broad, coordinated, research effort will be needed to provide answers to the above questions.

CONCLUSION

The use of computers to deliver traditional social rehabilitation services cannot be a sufficient condition for transforming them into effective social telerehabilitation services: to make these services really effective, it is often necessary to both rethink them in the light of the opportunities

offered by the new technologies and, at the same time, to carefully analyze the impact that they have on the persons involved in the process. In many cases it is worthwhile to verify the impact on the organization which delivers social telerehabilitation services, e.g. considering the requalification of personnel, the reliability of the computing services, the costs, and so on.

Internet-based applications can be realized not only for follow-up communications between caregiver-doctor-patient after a therapy, which is the most common use of ICT in the rehabilitation process, but also for improving the quality of the traditional rehabilitation therapies, or for realizing services that are entirely new.

The aim of developing new types of services is shared by social innovation. Another common point between social innovation and social telerehabilitation is that both entail cutting across organizational, sectoral, or disciplinary boundaries of expertise. Furthermore, social innovation looks towards technology as a means to create new social relationships between previously separate individuals and groups.

Social innovation and social economy are closely related: both pursue the same goal of creating solutions that can generate a positive impact, not only on the economic but also on the social development of people and their communities. The basis of social economy is the creation of services that focus on specific target groups who need to be sustained and assisted (e.g. services that complement core hospital staff, employment counseling, vocational training, education/rehabilitation for addict/misfit people, and so on), or that satisfy a general public need (e.g. nurseries, care services for the elderly, shelters for the homeless, integration for the disabled, and so on).

Social telerehabilitation can primarily be useful in the above-mentioned sectors, especially in the scope of social rehabilitation applications related to the labor sector. In this case, social telerehabilitation needs to be underpinned by an integrated approach and by a strong collaboration among different subjects, namely: rehabilitation

structures, employers, and workers. Education institutions play a crucial role in the implementation of social telerehabilitation services since the different subjects involved in them need to be educated. More generally, though, telerehabilitation and social innovation necessitate training and retraining activities for both the providers and users of services.³

REFERENCES

- Acampora, G., Cook, D. J., Rashidi, P., & Vasilakos, A. V. (2013). A survey on ambient intelligence in healthcare. *Proceedings of the IEEE*, 101(12), 2470–2494. doi:10.1109/JPROC.2013.2262913 PMID:24431472
- Altman, I. M., Swick, S., Parrot, D., & Malec, J. F. (2010). Effectiveness of community-based rehabilitation after traumatic brain injury for 489 program completers compared with those precipitously discharged. *Archives of Physical Medicine and Rehabilitation*, 91(11), 1697–1704. doi:10.1016/j.apmr.2010.08.001 PMID:21044714
- Arkin, R. C., Scheutz, M., & Tickle-Degnen, L. (2014, May). Preserving dignity in patient caregiver relationships using moral emotions and robots. *Ethics in Science, Technology and Engineering, 2014 IEEE International Symposium*, 1–5. doi:10.1109/ETHICS.2014.6893414
- Beatty, A. L., Fukuoka, Y., & Whooley, M. A. (2013). Using mobile technology for cardiac rehabilitation: a review and framework for development and evaluation. *Journal of the American Heart Association*, 2(6). Retrieved from <http://jaha.ahajournals.org/content/2/6/e000568.short>
- Bethge, M., von Groote, P., Giustini, A., & Gutenbrunner, C. (2014). The World Report on Disability: A challenge for rehabilitation medicine. *American Journal of Physical Medicine & Rehabilitation*, 93(1), S4–S11. doi:10.1097/PHM.000000000000016 PMID:24356081
- Brown, R. I., & Hughson, E. A. (1993). *Behavioral and social rehabilitation and training*. Captus Press.
- Camicia, M., Black, T., Farrell, J., Waites, K., Wirt, S., & Lutz, B. (2014). The essential role of the rehabilitation nurse in facilitating care transitions: A white paper by the Association of Rehabilitation Nurses. *Rehabilitation Nursing*, 39(1), 3–15. doi:10.1002/rnj.135 PMID:24402652
- Cooper, R. A., Fitzgerald, S. G., Boninger, M., Brienza, D. M., Shapcott, N., Cooper, R., & Flood, K. (2001). Telerehabilitation: Expanding access to rehabilitation expertise. *Proceedings of the IEEE*, 89(8), 1174–1193. doi:10.1109/5.940286
- Gibson, B. (2016). *Rehabilitation: A post-critical approach*. CRC Press. doi:10.1201/b19085
- Haig, A. J. (2013). Disability Policy Must Espouse Medical as well as Social Rehabilitation. *Social Inclusion*, 1(2), 136–138. doi:10.17645/si.v1i2.111
- Hill, A. J. (2010). *Report on the potential application of telerehabilitation to adult rehabilitation services in Scotland*. Edinburgh, UK: Scottish Government.
- ILO. (2008). *Employment (Disabled Persons) Convention (No. 159) and Recommendation (No. 168): United Nations Convention on the Rights of Persons with Disabilities*. Geneva: International Labor Office.
- Kamwesiga, J. T., Tham, K., & Guidetti, S. (2016). Experiences of using mobile phones in everyday life among persons with stroke and their families in Uganda – a qualitative study. *Disability and Rehabilitation*, 1–12. doi:10.3109/09638288.2016.1146354 PMID:26939597
- Karkou, V., Martinsone, K., Nazarova, N., & Vaverniece, I. (2011). Art therapy in the post-modern world: Findings from a comparative study across the UK, Russia and Latvia. *The Arts in Psychotherapy*, 38(2), 86–95. doi:10.1016/j.aip.2010.12.005

- Kuoppala, J., & Lamminpää, A. (2008). Rehabilitation and work ability: A systematic literature review. *Journal of Rehabilitation Medicine*, 40(10), 796–804. doi:10.2340/16501977-0270 PMID:19242615
- Lynggaard, V., May, O., Beauchamp, A., Nielsen, C. V., & Wittrup, I. (2014). LC-REHAB: Randomized trial assessing the effect of a new patient education method-learning and coping strategies–in cardiac rehabilitation. *BMC Cardiovascular Disorders*, 14(1), 2–9. doi:10.1186/1471-2261-14-186 PMID:24400648
- Malladi, N. (2015). Interdisciplinary Rehabilitation. *Physical Medicine and Rehabilitation Clinics of North America*, 26(2), 349–358. doi:10.1016/j.pmr.2014.12.008 PMID:25952069
- Marzano, G., Lubkina, V., & Rizakova, L. (2015, May). Delivering Social Telerehabilitation Services. In *Society, Integration, Education: Proceedings of the International Scientific Conference (Vol. 4, 457-467)*. doi:10.17770/sie2015vol4.411
- McInnes, D. K., Fix, G. M., Solomon, J. L., Petrakis, B. A., Sawh, L., & Smelson, D. A. (2015). Preliminary needs assessment of mobile technology use for healthcare among homeless veterans. *PeerJ*, 3, e1096. Retrieved from <https://doi.org/10.7717/peerj.1096>
- Meyer, T., Gutenbrunner, C., Bickenbach, J., Cieza, A., Melvin, J., & Stucki, G. (2011). Towards a conceptual description of rehabilitation as a health strategy. *Journal of Rehabilitation Medicine*, 43(9), 765–769. doi:10.2340/16501977-0865 PMID:21826389
- Meyer, T., Gutenbrunner, C., Kiekens, C., Skempes, D., Melvin, J. L., Schedler, K.,... Stucki, G. (2014). ISPRM Discussion Paper: Proposing a conceptual description of health-related rehabilitation services. *Journal of Rehabilitation Medicine*, 46(1), 1–6.
- Neumann, V., Gutenbrunner, C., Fialka-Moser, V., Christodoulou, N., Varela, E., Giustini, A., & Delarque, A. (2010). Interdisciplinary team working in physical and rehabilitation medicine. *Journal of Rehabilitation Medicine*, 42(1), 4–8. doi:10.2340/16501977-0483 PMID:20111837
- Nilsson, A. G., & Nilsson, T. H. (2011a). Towards a New Concept for Supporting Needy Children in Developing Countries–ICT Centres Integrated with Social Rehabilitation. In *Information Systems Development (pp. 437–448)*. Springer New York. doi:10.1007/978-1-4419-7355-9_37
- Nilsson, A. G., & Nilsson, T. H. (2011b, October). Global Development and ICT for Building Civil Societies in Developing Countries. *Proceedings of IDIA2011, the 5th International Development Informatics Conference on ICT for Development: People, Policy and Practice*, 26-28.
- Patel, S., Park, H., Bonato, P., Chan, L., & Rodgers, M. (2012). A review of wearable sensors and systems with application in rehabilitation. *Journal of Neuroengineering and Rehabilitation*, 9(1), 1. Retrieved from <http://jneuroengrehab.biomedcentral.com/articles/10.1186/1743-0003-9-21>
- Pravettoni, G., Folgieri, R., & Lucchiari, C. (2015). Cognitive science in telemedicine: from psychology to artificial intelligence. In *Tele-oncology (pp. 5–22)*. Springer International Publishing. doi:10.1007/978-3-319-16378-9_2
- Quan, X., Joseph, A., Keller, A., & Arch, M. (2011). *Designing safety-net clinics for innovative care delivery models*. The Center for Health Design. Retrieved from <http://www.lahealthaction.org/library/PDFDesigningClinicsInnovativeCare-DeliveryModels.pdf>
- Reeder, B., Chung, J., & Stevens-Lapsley, J. (2016). Current Telerehabilitation Research With Older Adults at Home: An Integrative Review. *Journal of Gerontological Nursing*, 42(10), 15–20. doi:10.3928/00989134-20160201-02 PMID:26870988

- Robinson, G., & Crow, I. D. (2009). Offender rehabilitation: Theory, research and practice. *Sage (Atlanta, Ga.)*.
- Rose, G. L., Skelly, J. M., Badger, G. J., Naylor, M. R., & Helzer, J. E. (2012). Interactive voice response for relapse prevention following cognitive-behavioral therapy for alcohol use disorders: A pilot study. *Psychological Services*, 9(2), 174–184. doi:10.1037/a0027606 PMID:22662731
- Sachpazidis, I., & Sakas, G. (2008, July). Medication intake assessment. In *Proceedings of the 1st international conference on Pervasive Technologies Related to Assistive Environments*. ACM.
- Salichs, M. A., Encinar, I. P., Salichs, E., Castro-González, Á., & Malfaz, M. (2016). Study of scenarios and technical requirements of a social assistive robot for alzheimers disease patients and their caregivers. *International Journal of Social Robotics*, 8(1), 85–102. doi:10.1007/s12369-015-0319-6
- Scaffa, M. E., & Reitz, S. M. (2013). *Occupational therapy community-based practice settings*. FA Davis.
- Scalvini, S., Vitacca, M., Paletta, L., Giordano, A., & Balbi, B. (2004). Telemedicine: A new frontier for effective healthcare services. *Monaldi Archives for Chest Disease*, 61(4), 226–233. doi:10.4081/monaldi.2004.686 PMID:15909613
- Scheutz, M. (2011). The Inherent Dangers of Unidirectional Emotional Bonds between Humans and Social Robots. In *Robot Ethics: The Ethical and Social Implications of Robotics*, 205–222.
- Schmeler, M. R., Schein, R. M., McCue, M., & Betz, K. (2009). Telerehabilitation clinical and vocational applications for assistive technology: research, opportunities, and challenges. *International Journal of Telerehabilitation*, 1(1), 59–72.
- Simpson, T. L., Kivlahan, D. R., Bush, K. R., & McFall, M. E. (2005). Telephone self-monitoring among alcohol use disorder patients in early recovery: A randomized study of feasibility and measurement reactivity. *Drug and Alcohol Dependence*, 79(2), 241–250. doi:10.1016/j.drugalcdep.2005.02.001 PMID:16002033
- Stanos, S. (2012). Focused review of interdisciplinary pain rehabilitation programs for chronic pain management. *Current Pain and Headache Reports*, 16(2), 147–152. doi:10.1007/s11916-012-0252-4 PMID:22427179
- Tickle-Degnen, L., Scheutz, M., & Arkin, R. C. (2014). *Collaborative Robots in Rehabilitation for Social Self-Management of Health*. Retrieved from <https://smartech.gatech.edu/bitstream/handle/1853/52674/Tickle14.pdf?sequence=1>
- Varnfield, M., Karunanithi, M., Lee, C. K., Honeyman, E., Arnold, D., Ding, H.,... Walters, D. L. (2014). Smartphone-based home care model for the improved use of cardiac rehabilitation in postmyocardial infarction patients: results from a randomized controlled trial. *Heart*. Retrieved from <http://heart.bmj.com/content/early/2014/06/27/heartjnl-2014-305783.full.pdf+html>
- Wade, D. T., & de Jong, B. A. (2000). Recent advances: Recent advances in rehabilitation. *BMJ: British Medical Journal*, 320(7246), 1385–1388. doi:10.1136/bmj.320.7246.1385 PMID:10818031
- Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1), 36–45. doi:10.1145/365153.365168
- Wilson, K., Halsey, A., Macpherson, H., Billington, J., Hill, S., Johnson, G., & Abbott, P. et al. (2012). The psycho-social rehabilitation of patients with alcohol-related brain damage in the community. *Alcohol and Alcoholism (Oxford, Oxfordshire)*, 47(3), 304–311. doi:10.1093/alcal/agr167 PMID:22278316
- Winters, J. M. (2002). Telerehabilitation research: Emerging opportunities. *Annual Review of Biomedical Engineering*, 4(1), 287–320. doi:10.1146/annurev.bioeng.4.112801.121923 PMID:12117760
- World Health Organisation & World Bank. (2011). *World Report on Dis-1. ability*. Geneva: WHO. Retrieved from http://www.who.int/disabilities/world_report/2011/report.pdf

ADDITIONAL READING

Fardoun, H. M., Penichet, V. M., & Alghazzawi, D. M. (Eds.). (2015). *ICTs for Improving Patients Rehabilitation Research Techniques: Second International Workshop, REHAB 2014, Oldenburg, Germany, May 20-23, 2014, Revised Selected Papers* (Vol. 515). Springer. doi:10.1007/978-3-662-48645-0

Kumar, S., & Cohn, E. R. (Eds.). (2012). *Telerehabilitation*. Springer Science & Business Media.

Lin, P., Abney, K., & Bekey, G. A. (2011). *Robot ethics: the ethical and social implications of robotics*. MIT press.

Mattingly, S. (Ed.). (2012). *Rehabilitation today*. Springer Science & Business Media.

KEY TERMS AND DEFINITIONS

Artificial Intelligence: Is a branch of computer science that aims to improve machine behaviour in tackling complex tasks that are specific of human beings.

Behavior-Based Robots: Are implemented following an approach that focuses on exhibiting in a robot complex-interactive behaviors in given environment.

Caregiver: Is someone who (paid or unpaid) takes care of another person who is either sick or disabled; a caregiver is most commonly used to address impairments related to old age, disability, a disease, or a mental disorder.

Occupational Therapy: Helps to solve the problems that interfere with an individual ability to do the things that are important to them; it can also prevent a problem or minimize its effects.

Problem Solving Process: Consists of using methods, in a methodical and organized manner, for finding solutions to problems.

Social Educator: Is an independent and recognized professional figure linked to the area of social professions; he supports and guides a person with developmental disabilities in making his or her own choices and implementing their decisions.

Wearable Technology: Consists in incorporating in clothing computer and advanced electronic technologies.

World Bank (WB): Is an international financial institution that provides loans to developing countries for capital programs.

World Health Organization (WHO): Is a specialized agency of the United Nations that is concerned with international public health.

ENDNOTES

- ¹ <http://www.who.int/topics/rehabilitation/en/>; last accessed 7.06.2016
- ² <http://www.un.org/disabilities/documents/convention/convoptprot-e.pdf>; last accessed 06.06.2016.
- ³ See the Latvian National Research Program. “Innovative solutions in social rehabilitation in Latvian schools in the context of inclusive education” (“Inovatīvi risinājumi sociālajā telerehabilitācijā Latvijas skolās iekļaujošās izglītības kontekstā – VPP INOSOCETEREHI”). VPP INOSOCETEREHI is a three-year multidisciplinary project (2015-2017) in the field of social telerehabilitation focused on social inclusion; it is being conducted by four Latvian Universities (Rezekne Technology Academy, Latvia University, Riga Technical University, and Liepaja University), and investigates the use of mobile technology in the scope of social rehabilitation (<http://telerehabilitation.lv/>)

A Validation Study of Rehabilitation Exercise Monitoring Using Kinect

M**Wenbing Zhao***Cleveland State University, USA***Deborah D. Espy***Cleveland State University, USA***Ann Reinthal***Cleveland State University, USA*

INTRODUCTION

In rehabilitative health care, a carefully designed physical exercise plan could be instrumental to the recovery of a patient provided that the patient. Exercise programs are prescribed to address specific problems, and are often individually tailored by a clinician due to the presence of co-morbidities and additional impairments. It is critical that the patient perform the proscribed program correctly and with adequate practice repetitions (in the range of thousands) (Kleim and Jones, 2008), otherwise, the exercise may be ineffective, or even dangerous (Escamilla et al., 2009; Tino & Hillis, 2010).

Correct adherence to supplemental home exercise is essential for safe, effective, and efficient care. The lack of correct feedback during independent in-home exercise is therefore a serious concern. The use of simple counting devices helps verify the exercise repetitions. However, such simple, commercially available devices cannot fully capture all the required movements beyond the most simple, such as counting steps or recording overall amounts of activity (Wagner et al., 2012; Yang & Hsu, 2010), and are, therefore not useful for most prescribed home exercises.

The release of the Microsoft Kinect sensor, which is equipped with a depth camera capable of measuring 3 dimensional positions of the objects in its view, has triggered tremendous interest in its use to monitor in-home physical therapy exercises

(Chang et al., 2013; Chang et al., 2012; Garcia et al., 2012; Gibson et al., 2012; Guerrero & Uribe-Quevedo, 2012; Huang, 2011; Zannatha et al., 2013; Pastor et al., 2012). A Kinect-based system could facilitate proper performance of the exercise or fitness program, increase patient accountability, allow the clinician to correct any errors in exercise performance, and allow program modification or advancement as needed. Hence, the Kinect sensor based system could potentially provide sufficient feedback and guidance to patients performing clinician prescribed in-home exercises, significantly minimizing costly and inconvenient trips to outpatient centers, and improving the effectiveness and outcomes of courses of treatment.

Many existing clinical trials with Kinect-based systems appear to have proceeded without comprehensive validation tests (Chang et al., 2013; Chang et al., 2012; Garcia et al., 2012; Gibson et al., 2012; Guerrero & Uribe-Quevedo, 2012; Huang, 2011; Zannatha et al., 2013; Zhao et al., 2014; Tamei et al., 2015; Ebert et al., 2015). Other studies have aimed to characterize the accuracy of the Kinect sensor; however, these validation studies have focused primarily on the movements within the frontal plane for a subset of the joints or segments (Clark et al., 2013; Obdrzalek et al., 2012; Mobini et al., 2013). In this article, we report our validation study on using a Kinect-based system for physical therapy exercise monitoring. Instead of comparing the joint positions or angles

DOI: 10.4018/978-1-5225-2255-3.ch517

formed by key segments with respect to a (usually far more expensive) reference system, we take a completely different approach by focusing on the feasibility of using such a system to assess the correctness rules for a few common exercises in physical therapy. The correctness rules are readily implementable in a computer program for real-time motion tracking and feedback.

Due to space limitation, we only present results for two exercises, namely hip abduction and toe touch. In the standing hip abduction exercise, one leg is moved into hip abduction without any additional sagittal plane hip flexion/extension or transverse plane hip rotation. The pelvis, knee, and ankle remain stationary. In the standing toe touch exercise, the trunk bends forward and the arms reach to touch the floor. Motion occurs primarily as sagittal plane spine and hip flexion with concurrent shoulder flexion. When done correctly, there is minimal movement of the elbows, wrists, knees and ankles. We show that, although the Kinect-based system is capable of assessing many correctness rules for these exercises, it fails in the presence of significant self-occlusion, especially for the toe-touch exercise.

BACKGROUND

Microsoft Kinect was initially released as an add-on device for the Xbox 360 game console. Kinect enables a person to interact with a game using gestures and voice commands via what is referred to as the Natural User Interface. In early 2011, Microsoft released an official driver for Kinect and a software development kit (SDK). The most useful feature of the SDK is skeletal tracking, via which, an application can receive pre-processed frames containing up to two skeletons each with 20 joints for Kinect v1, and up to four skeletons each with 26 joints for Kinect v2 (Lun and Zhao, 2015). Also very useful in the Microsoft Kinect SDK are its floor clipping application programming interfaces (APIs). Each skeleton frame includes a floor clipping plane vector containing

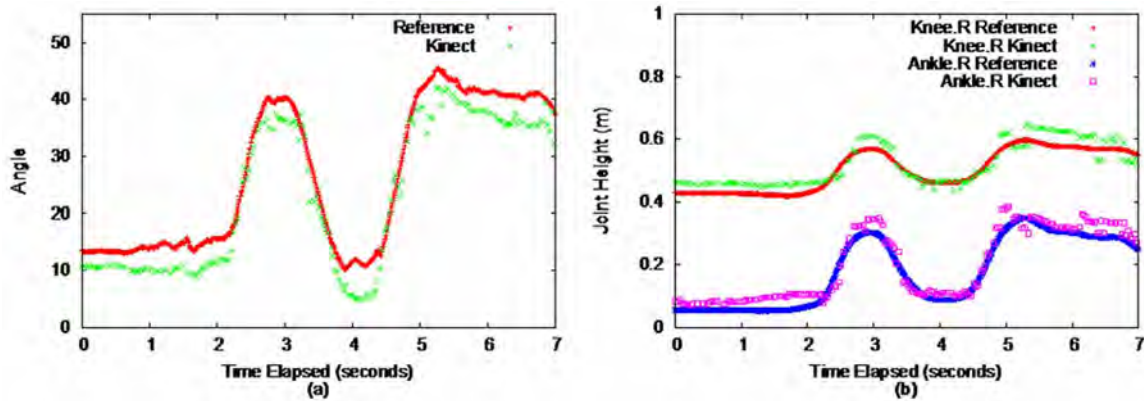
the coefficients of the floor plane equation. Based on the floor clipping plane vector, we can calculate the vertical height off the floor clipping plane of each joint. In our study, we exploited the APIs to assess the correctness of some of the exercises.

VALIDATION STUDY

To validate the use of Microsoft Kinect for rehabilitation monitoring, we compared the skeletal joint results obtained from Kinect to those obtained concurrently by a Cortex motion capture system (Motion Analysis Corp, Santa Rosa CA) (2010). The subjects wore a full Helen Hayes marker set and each exercise was recorded by 8 cameras. The Cortex motion results were used to establish the ground truth for the study according to its established accuracy.

Our study has gone through three phases. During the initial phase, we attempted to carry out a mechanical point-to-point comparison between the two systems based on coordinate transformation. This proved to be difficult because it would require the Kinect sensor to be placed in a known location and a known tilt angle to the Cortex system. In the second phase, we attempted to focus on the comparison of relative movement between joints and the angles formed between adjacent segments during the movement. This effort was partially successful because it did not require coordinate transformation. However, a comprehensive comparison was still difficult because the positions of the markers used in the Cortex system did not correspond strictly to the joint positions reported by the Kinect sensor. Figure 1 shows a partial comparison of results we obtained in the first two phases for the hip abduction study. By exploiting the floor clipping plane APIs, we managed to compare the height of the knee and ankle during the hip abduction motion (i.e., right knee and right ankle) reported by the Kinect sensor and that captured by the Cortex system. As shown in Figure 1(b), the Kinect results were fairly close to those captured by the Cortex system. Figure

Figure 1. A comparison of the Kinect measurement and that from the reference motion capture system in early stages of this study. (a) The angle formed by the two legs in hip abduction. (b) The height of the right (abducting) knee.



1(a) shows the angle formed by the two legs as reported by Kinect and the Cortex system. Again, we can see that the Kinect results were very close to those captured by the Cortex system.

Although the early results proved promising in using Kinect for therapeutic exercise monitoring, they were far from comprehensive. After all, the correct angle or the correct trajectories of particular joints do not necessarily mean an exercise is done correctly. We needed to perform a system-level assessment of Kinect's feasibility. We believe the most appropriate way to examine whether Kinect can be used to tell if a patient is doing an exercise correctly is based on a set of pre-defined correctness rules. Such rules establish a set of invariance for each exercise with appropriate error bounds. We argue that this approach is far more practical compared with the mechanical point-by-point based assessment:

- The point-by-point based approach would require: (1) a coordination systems translation between the Kinect coordinate system and that used by the reference motion capture system; and (2) the reference system joint positions derived from external marker placements to match those reported by the Kinect SDK. In our experiences, neither is trivial.
- Typically, prescribed therapeutic exercises allow a range of movement speeds and trajectories according to predefined correctness rules. As such, even if the joint positions reported by Kinect contain error, they might be adequate to assess whether or not the rules have been followed by the patient. Furthermore, some rules are more strict than others. A 5 cm error in Kinect measurement for a particular joint might not impact the assessment of the correctness for one exercise, but might blur the correctness boundary for another. Therefore, what matters ultimately is whether or not we can use a Kinect based system to distinguish correct exercises from wrong ones.
- Similarly, self-occlusion happens frequently during an exercise when a single Kinect camera is used. The positions for the occluded joints are estimated in Kinect. Hence, such joint positions are inevitably inaccurate. However, whether or not a large error for an occluded joint matters depends on the correctness rules for each exercise. The point-by-point based validation approach cannot give us a definitive answer as to whether or not Kinect is useful because such validation is without context.

In the following subsections, we report our experimental validation results for two typical physical therapy exercises, namely, hip abduction and toe touch.

Hip Abduction

For the hip abduction exercise, typically the following 7 rules are used to evaluate if a patient is doing the exercise correctly. Throughout, we assumed that the left leg is stationary and the right leg is the abducting leg:

Rule 1: Weight shift onto the left leg prior to abducting right leg.

Rule 2: The left leg is straight and leans slightly left after the weight shift and during right leg hip abduction.

Rule 3: Upper body shifts to the left side over the left leg after the weight shift and during right leg hip abduction.

Rule 4: Arms are held out in the frontal plane.

Rule 5: The right leg stays in frontal plane during hip abduction.

Rule 6: The right foot remains pointing forward during hip abduction.

Rule 7: The pelvis stays stable after weight shift and motion occurs only at the hip joint.

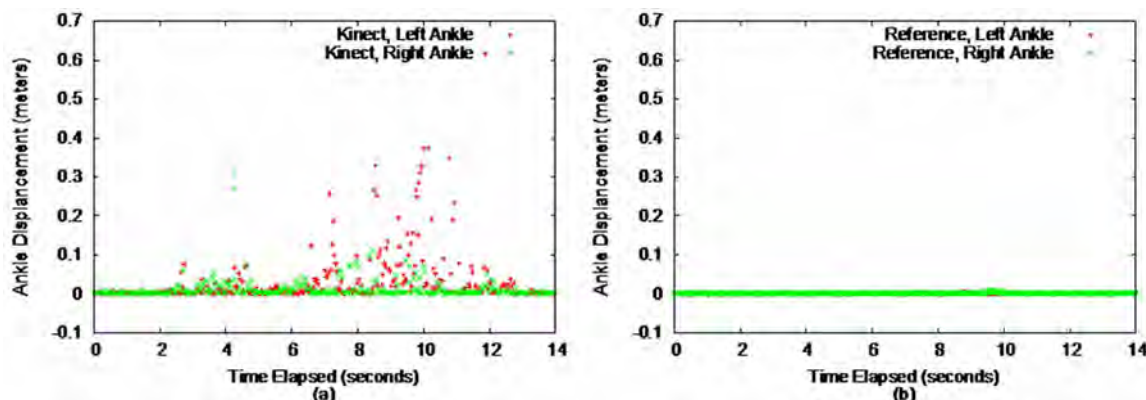
To assess the rule 1 conformance, we measured the angle formed by the following two vectors:

(1) a vector from the left ankle to the left hip joint *obtained in the initial pose*, and (2) a vector from the left ankle joint to the center of mass of the patient. Ideally, the angle should be reduced to 0 when the weight is completely shifted to the left leg. As can be seen in Figure 2, the angle calculated from the Kinect measurement clearly shows a gradual reduction from about 4.5 degrees to less than 1 degree, and it is very consistent with that measured by the Cortex system, although noticeably more noisy.

Rule 2 has two components: (1) the left leg must remain straight, and (2) the patient leans slightly left. Component (1) can be easily assessed by measuring the angle formed between the left knee to left hip segment and the left knee to the left ankle segment (referred to as left knee angle), and component (2) can be characterized by a leaning angle defined as the angle formed between the left hip to the left ankle and the vertical line from the left hip to the floor. Ideally, the left knee angle should remain at 180 degrees and the lean angle should change from 0 to a small value.

During our experiment, the subject performed 5 iterations of hip abduction with the first 3 iterations performed correctly and the last 2 intentionally violating rule 5 (the very last iteration is partially captured). As shown in Figure 3(a), the left knee angle as measured by Kinect stays within 170-180 degrees. The measurement result from the Cortex system reveals slightly larger variations

Figure 2. A comparison of the weight shift angle for hip abduction



(from about 160-180 degrees). The leaning angle indeed starts from about 0 degrees to an average of about 8 degrees for both the Kinect and Cortex measurement results as shown in Figure 3(b). However, the Kinect data are much noisier than the Cortex results. Nevertheless, the results showed that with appropriate error bounds, the Kinect based system can be used to assess rule 2 conformance.

Rule 3 is evaluated by measuring the angle formed between the shoulder center to hip center vector and the same vector at its initial position. This upper body shift angle should start at 0 and increase to a moderate angle. In between two iterations of the hip abduction, the angle should reduce to a smaller angle. As can be seen, this was indeed the case as measured by both the Kinect and the Cortex systems as shown in Figure 4. The maximum shift angle as measured by the Cortex system was clearly bigger than that measured by the Kinect based system, indicating a systematic error in the Kinect measurement. Fortunately, this systematic error is relatively small, about 2 degrees for correct iterations and as large as 6 degrees for incorrect iterations. We believe such small error is acceptable for assessing rule 3.

Rule 4 is concerned with the placement of the arms during hip abduction. There are two com-

ponents to the rule: (1) both arms must remain in the frontal plane, and (2) both arms must move out. To assess this rule, we must first establish the frontal plane, accomplished by using the following two vectors: (1) hip center to the left shoulder, and (2) hip center to the right shoulder. More specifically, the frontal plane was established by using the two vectors obtained in the *initial* T-pose. We calculated the out-of-frontal-plane angle of each arm. Ideally, the angle should remain 0. When an arm is moved out, it would form a 90 degree angle with respect to the torso (represented by a vector from shoulder center to the hip center). Due to the weight shift and leaning during abduction, the angle would oscillate around 90 degrees.

As shown in Figure 5(a), both arms remained within about 10 degrees from the frontal plane according to the Cortex system. As measured by the Kinect based system, the left arm indeed stayed within 10 degrees of the frontal plane; however, the right arm appeared to deviate from the frontal plane by as much as 25 degrees. This indicated that we would need a fairly generous error bound to assess rule 4 using the Kinect based system. As shown in Figure 5(b), the arms deviated from 90 degrees by as many as 10 degrees according to the Cortex system. While the arm-to-torso angles measured by the Kinect based system showed

Figure 3. A comparison of the left knee angle and leaning angle for hip abduction rule 2 assessment

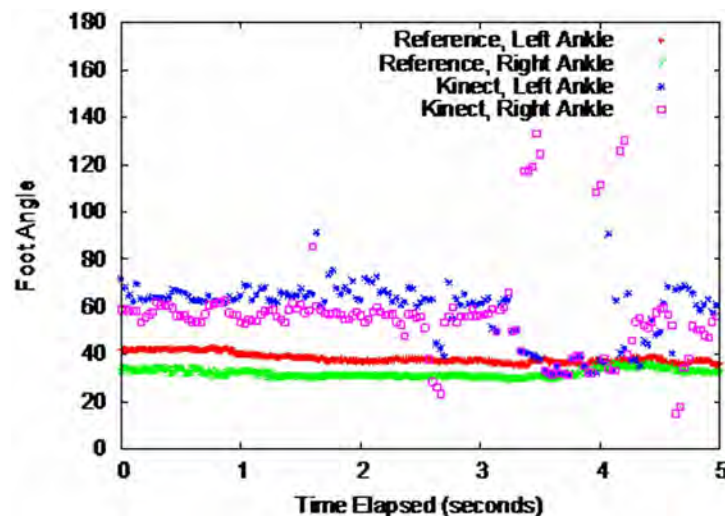


Figure 4. A comparison of the upper body shift angle for hip abduction hip 3 assessment

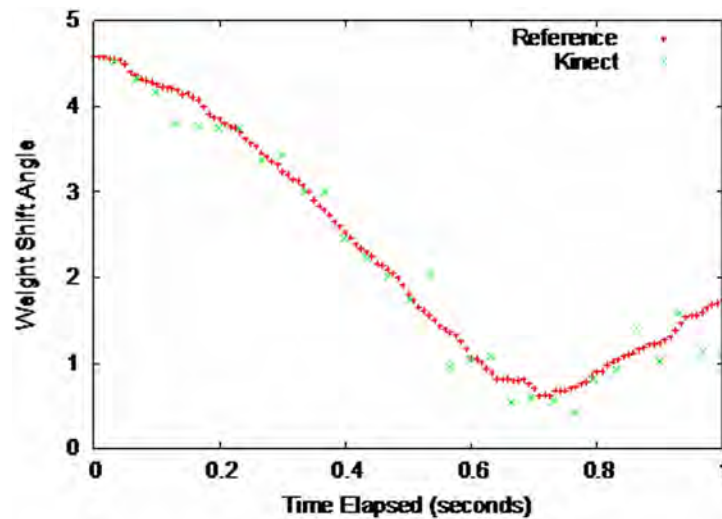
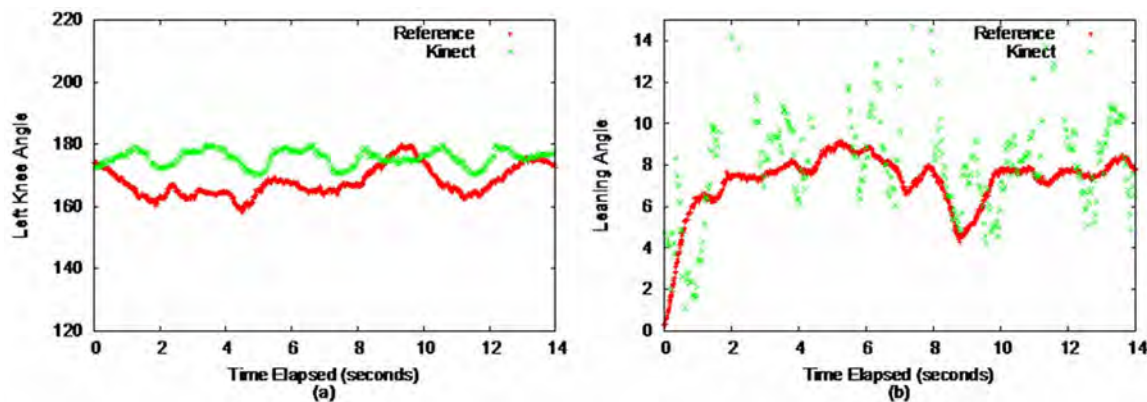


Figure 5. A comparison of the upper body shift angle for hip abduction rule 3 assessment



smaller deviations from 90 degrees, they were apparently systematically larger than those measured by the Cortex system by approximately 10 degrees. Nevertheless, this systematic error does not appear to limit the use of Kinect to assess the second component of rule 4 for hip abduction.

Rule 5 can be assessed very similarly to the first component of rule 4. We calculated the out-of-frontal-plane angle formed between the right leg and the frontal plane. For correct iterations, this angle should be near 0 degrees, and should have significant non-zero angle values for the incorrect iterations. As can be seen Figure 6, according to

the Cortex system, the out-of-frontal-plane angle for the right leg oscillated slightly around 5 degrees with an amplitude of about 5 degrees for the 3 correct iterations, and increased to over 30 degrees off the frontal plane for the incorrect iterations. The Kinect data showed that the oscillation was around 8 degrees (instead of Cortex's 5) for correct iterations and the amplitude about 8 degrees (instead of 5). This error in Kinect measurement can be easily overcome by setting slightly larger error bounds when assessing the hip abduction exercises in practice. For the wrong iteration, the highest angle captured by Kinect was very con-

sistent with that measured by the Cortex system at about 32 degrees. Hence, it is clear that Kinect is capable of detecting the violation of this rule.

For rule 6, we measured the angle formed between the ankle-to-foot vector and a forward-pointing vector. The forward-pointing vector was perpendicular to the frontal plane and can be easily derived from the cross product of the hip-center-to-left-shoulder vector and the hip-center-to-right-shoulder vector. If the foot is indeed pointing strictly forward, the angle should be 0.

As shown in Figure 7, the right foot is, in fact, not even close to 0 degrees. According to the Cortex system, the foot angles varied around 30 degrees. The Kinect data showed very large variation, between 20 and 70 degrees. Our experiments clearly show that Kinect is not capable of assessing this rule due to the inaccuracy of the measurement for the foot segment.

Rule 7 is not assessable by the Kinect based system because the current Kinect SDK does not provide APIs to capture the motion of the pelvis.

In summary, the majority of correctness rules can be assessed by a Kinect-based system. However, we note that rules for subtle movements, such as the foot alignment and pelvis movement, cannot be evaluated by Kinect. If these movements are critical for some physical therapy exercises,

either the Kinect's capability has to be expanded and improved, or other motion capture systems would need to be used.

Toe Touch

For the toe touch exercise, typically the following 4 rules are used to evaluate if a patient is doing the exercise correctly.

Rule 1: Both knees are extended, i.e., the legs are straight.

Rule 2: The torso should not rotate as the fingers reach for the floor.

Rule 3: Ankles should not move.

Rule 4: Feet are pointed forward.

We can assess rule 1 by measuring the angle formed by the upper leg and the lower leg. If the legs are straight, then the angle should be at 180 degrees. As shown in Figure 8, for a toe touch run with two iterations, according to the Cortex system, both legs were indeed fairly straight as the knee angle stayed very close to 180 degrees. Unfortunately, the angles calculated from the Kinect data were far off 180 degrees when the subject's torso occluded the legs when his fingers reached for the floor. This clearly shows that we

Figure 6. A comparison of the angle between the right leg and the frontal plane for hip abduction

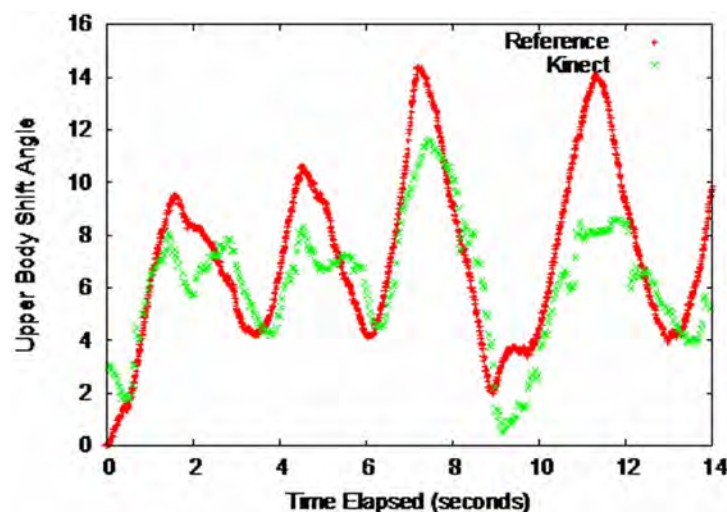
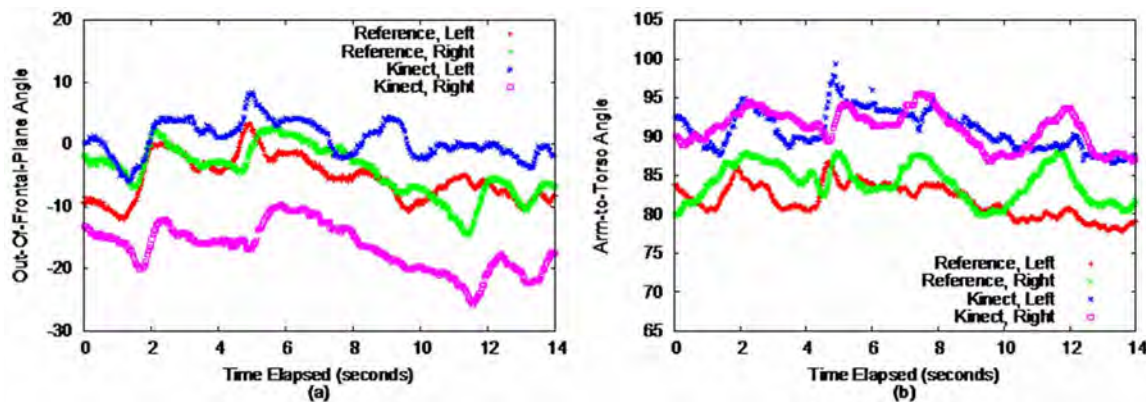


Figure 7. A comparison of the angle between the right leg and the frontal plane for hip abduction



cannot use Kinect to properly evaluate rule 1 conformance before the self-occlusion problem is resolved.

When the torso does not rotate during the movement, the left and right shoulder should move within the sagittal plane in parallel. Hence, the traveled distance from its initial position should be the same for the left or right shoulder joint. Therefore, we assessed the rule 2 conformance by measuring and comparing the distance travelled relative to the initial position (in T-pose). If the torso does not rotate, the distances for the left and right shoulders should be the same. As can be seen in Figure 9(a), this was the case for

a correct iteration of toe touch. For an iteration with torso rotation, the distances were different, as revealed in Figure 9(b).

Note that there were apparent discrepancies in the distance measurement results between the Kinect based system and the Cortex system. As shown in Figure 9, the discrepancy can be as large as 0.2 meters when the subject's fingers are reaching for the toe. This large measurement error in the Kinect system apparently will not prevent us from using Kinect to assess rule 2.

We can assess rule 3 by measuring the displacement of the ankles with respect to their initial positions. If the ankles do not move, the

Figure 8. A comparison of the knee angles for toe touch

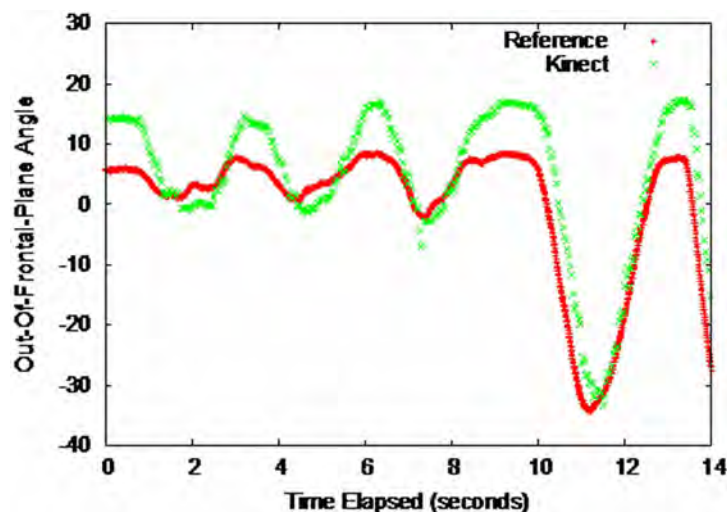
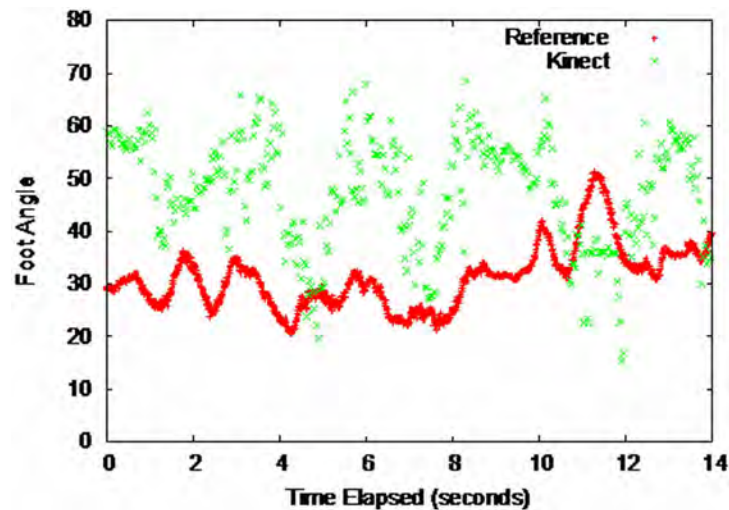


Figure 9. A comparison of the distances between the two shoulder joints with respect to their initial positions for toe touch



displacement should be close to 0. As can be seen from the Cortex system data shown in Figure 10, indeed this was the case. However, the Kinect data exhibited large displacement during toe touch, probably due to self-occlusion. This clearly shows that we cannot use Kinect to assess rule 3.

We can assess rule 4 by measuring the angle formed by the foot and a forward vector, in exactly the same way as that for hip abduction rule 6. If the feet are pointing forward, the angle should be

0. As shown in Figure 11, this was not the case as revealed by the Cortex system measurement (with an angle around 40 degrees). The angles measured by Kinect were significantly larger (about 60 degrees) and varied significantly in the presence of self-occlusion. Hence, we conclude that we cannot assess rule 4 using a Kinect based system.

In summary, only one of the four correctness rules can be assessed by a Kinect-based system for the toe touch exercise. Hence, we cannot use

Figure 10. A comparison of the ankle displacement for toe touch

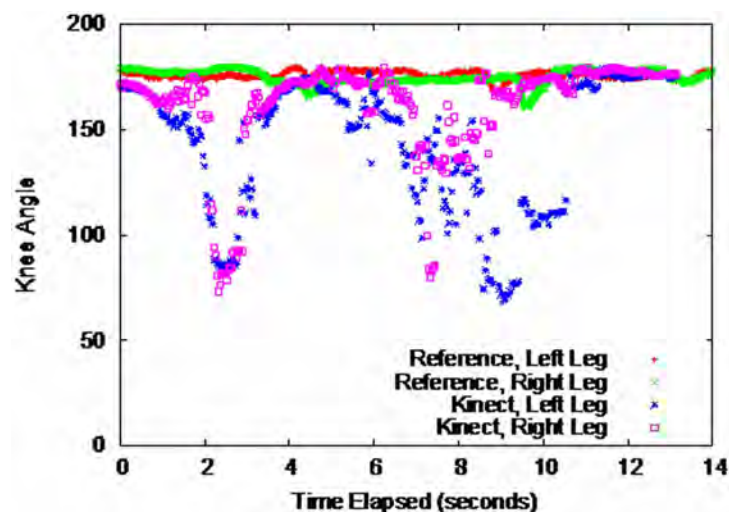
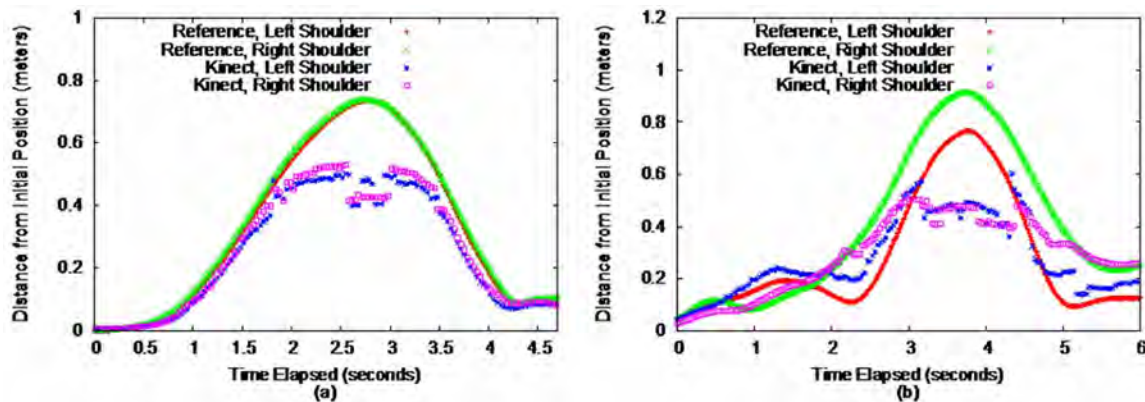


Figure 11. A comparison of the foot angle for toe touch



such a system to evaluate the correctness of the toe touch exercise and provide meaningful feedback to the patients before the self-occlusion problem is addressed by the Kinect research community.

FUTURE TRENDS

The current Kinect SDK has two obvious limitations: (1) it captures only 20 joints in Kinect v1 and 26 joints in Kinect v2, and (2) the joint positions often contain large errors in the presence of occlusion. As we have shown in our feasibility study results presented in the previous section, the limitation of the joint set prevents us from monitoring subtle pelvis movement, and the large errors in joint positions in the presence of self-occlusion make it impossible for us to use Kinect to evaluate some correctness rules, especially for exercises that involve self-occlusion predominately. There have been intensive research and development efforts to address the two issues (Han et al., 2013). For example, a Kinect based motion capture system for finger tracking was designed, implemented, and thoroughly validated in a very recent work (Metcalf et al., 2013). The self-occlusion problem has been tackled by using sophisticated machine learning methods with constraints (Shen et al., 2012), by incorporating the reliability estimation

on the selection of kinematically valid postures (Shum et al., 2013), and by using multiple Kinect sensors (Asteriadis et al., 2013).

CONCLUSION

In this article, we presented our work on a validation study of using Microsoft Kinect to monitor rehabilitation exercises. Differing from other validation efforts, we focused on a system-level assessment instead of the joint-level comparison with reference motion capture systems. We assessed the feasibility of using Kinect by examining the enforceability of a set of correctness rules defined for each exercise, which are invariances of each exercise and hence independent from the coordinate system used. This method is more advantageous in that (1) it does not require coordinate system transformation between those of the reference motion capture system and of the Kinect based system, (2) it does not require an exact match of the Kinect joints and the corresponding external marker placements or derived joint centers often used in reference motion capture systems, and (3) the correctness rules and their mapping for Kinect motion data analysis developed in this study are readily implementable for a real motion monitoring system for physical therapy.

REFERENCES

- Asteriadis, S., Chatzitofis, A., Zarpalas, D., Alexiadis, D. S., & Daras, P. (2013, June). Estimating human motion from multiple Kinect sensors. In *Proceedings of the 6th International Conference on Computer Vision/Computer Graphics Collaboration Techniques and Applications* (p. 3). ACM. doi:10.1145/2466715.2466727
- Chang, C. Y., Lange, B., Zhang, M., Koenig, S., Requejo, P., Somboon, N., & Rizzo, A. A. et al. (2012, May). Towards pervasive physical rehabilitation using Microsoft Kinect. In *Proceedings of the 6th International Conference on Pervasive Computing Technologies for Healthcare*, (pp. 159-162). IEEE. doi:10.4108/icst.pervasive-health.2012.248714
- Chang, Y. J., Han, W. Y., & Tsai, Y. C. (2013). A Kinect-based upper limb rehabilitation system to assist people with cerebral palsy. *Research in Developmental Disabilities*, 34(11), 3654–3659. doi:10.1016/j.ridd.2013.08.021 PMID:24012594
- Clark, R. A., Pua, Y. H., Bryant, A. L., & Hunt, M. A. (2013). Validity of the Microsoft Kinect for providing lateral trunk lean feedback during gait retraining. *Gait & Posture*, 38(4), 1064–1066. doi:10.1016/j.gaitpost.2013.03.029 PMID:23643880
- Ebert, D., Metsis, V., & Makedon, F. (2015, July). Development and evaluation of a unity-based, kinect-controlled avatar for physical rehabilitation. In *Proceedings of the 8th ACM International Conference on Pervasive Technologies Related to Assistive Environments* (p. 88). ACM. doi:10.1145/2769493.2769556
- Escamilla, E., Yamashiro, K., Paulos, L., & Andrews, J. (2009). Shoulder muscle activity and function in common shoulder rehabilitation exercises. *Sports Medicine (Auckland, N.Z.)*, 39(8), 663–685. doi:10.2165/00007256-200939080-00004 PMID:19769415
- Garcia, J. A., Navarro, K. F., Schoene, D., Smith, S. T., & Pisan, Y. (2012). Exergames for the elderly: Towards an embedded Kinect-based clinical test of falls risk. *Studies in Health Technology and Informatics*, 178, 51–57. PMID:22797019
- Gibson, S., McBride, S. J., McClelland, C., & Watson, M. (2012). A technological evaluation of the Microsoft Kinect for automated behavioural mapping at bed rest. *Studies in Health Technology and Informatics*, 188, 39–45. PMID:23823286
- Guerrero, C., & Uribe-Quevedo, A. (2012). Kinect-Based Posture Tracking for Correcting Positions during Exercise. *Studies in Health Technology and Informatics*, 184, 158–160. PMID:23400149
- Han, J., Shao, L., Xu, D., & Shotton, J. (2013). Enhanced Computer Vision with Microsoft Kinect Sensor: A Review. *IEEE Transactions on Cybernetics*, 43(5), 1318–1334. doi:10.1109/TCYB.2013.2265378 PMID:23807480
- Huang, J. D. (2011, October). Kinerehab: a kinect-based system for physical rehabilitation: a pilot study for young adults with motor disabilities. In *Proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility* (pp. 319-320). ACM. doi:10.1145/2049536.2049627
- Lun, R., & Zhao, W. (2015). A survey of applications and human motion recognition with Microsoft Kinect. *International Journal of Pattern Recognition and Artificial Intelligence*, 29(5), 1555008. doi:10.1142/S0218001415550083
- Metcalfe, C., Robinson, R., Malpass, A., Bogle, T., Dell, T., Harris, C., & Demain, S. (2013). Markerless motion capture and measurement of hand kinematics: Validation and application to home-based upper limb rehabilitation. *IEEE Transactions on Bio-Medical Engineering*, 60(8), 2184–2192. doi:10.1109/TBME.2013.2250286 PMID:23475333

- Mobini, A., Behzadipour, S., & Saadat Foumani, M. (2014). Accuracy of Kinects skeleton tracking for upper body rehabilitation applications. *Disability and Rehabilitation. Assistive Technology*, 9(4), 344–352. doi:10.3109/17483107.2013.805825 PMID:23786360
- Motion Analysis Corporation. (2010). *Cortex, version 2.1 reference manual*. Author.
- Obdrzalek, S., Kurillo, G., Ofli, F., Bajcsy, R., Seto, E., Jimison, H., & Pavel, M. (2012, August). Accuracy and robustness of Kinect pose estimation in the context of coaching of elderly population. *Proceedings of the annual international conference of the IEEE Engineering in medicine and biology society*, 1188-1193. doi:10.1109/EMBC.2012.6346149
- Obdrzalek, S., Kurillo, G., Ofli, F., Bajcsy, R., Seto, E., Jimison, H., & Pavel, M. (2012, August). Accuracy and robustness of Kinect pose estimation in the context of coaching of elderly population. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, (pp. 1188-1193). IEEE. doi:10.1109/EMBC.2012.6346149
- Pastor, I., Hayes, H. A., & Bamberg, S. J. (2012, August). A feasibility study of an upper limb rehabilitation system using kinect and computer games. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 1286-1289). IEEE. doi:10.1109/EMBC.2012.6346173
- Shen, W., Deng, K., Bai, X., Leyvand, T., Guo, B., & Tu, Z. (2012, June). Exemplar-based human action pose correction and tagging. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1784-1791). IEEE.
- Shum, H. P., Ho, E. S., Jiang, Y., & Takagi, S. (2013). Real-Time Posture Reconstruction for Microsoft Kinect. *Cybernetics. IEEE Transactions on*, 43(5), 1357–1369.
- Tamei, T., Orito, Y., Funaya, H., Ikeda, K., Okada, Y., & Shibata, T. (2015). Kinect-based posturography for in-home rehabilitation of balance disorders. *APSIPA Transactions on Signal and Information Processing*, 4, e17. doi:10.1017/ATSIP.2015.17
- Tino, D., & Hillis, C. (2010). The full can exercise as the recommended exercise for strengthening the supraspinatus while minimizing impingement. *Strength and Conditioning Journal*, 32(5), 33–35. doi:10.1519/SSC.0b013e3181d54721
- Wagner, F., Basran, J., & Dal Bello-Haas, V. (2012). A review of monitoring technology for use with older adults. *Journal of Geriatric Physical Therapy*, 35(1), 28–34. doi:10.1519/JPT.0b013e318224aa23 PMID:22189952
- Yang, C. C., & Hsu, Y. L. (2010). A review of accelerometry-based wearable motion detectors for physical activity monitoring. *Sensors (Basel, Switzerland)*, 10(8), 7772–7788. doi:10.3390/s100807772 PMID:22163626
- Zannatha, J. M. I., Tamayo, A. J. M., Sánchez, Á. D. G., Delgado, J. E. L., Cheu, L. E. R., & Arévalo, W. A. S. (2013). Development of a system based on 3D vision, interactive virtual environments, ergonomic signals and a humanoid for stroke rehabilitation. *Computer Methods and Programs in Biomedicine*, 112(2), 239–249. doi:10.1016/j.cmpb.2013.04.021 PMID:23827333
- Zhao, W., Feng, H., Lun, R., Espy, D. D., & Reinthal, M. (2014, June). A Kinect-based rehabilitation exercise monitoring and guidance system. In *Software Engineering and Service Science (ICSESS), 2014 5th IEEE International Conference on* (pp. 762-765). IEEE. doi:10.1109/ICSESS.2014.6933678

ADDITIONAL READING

- Abdur Rahman, M., Qamar, A. M., Ahmed, M. A., Ataur Rahman, M., & Basalamah, S. (2013, April). Multimedia interactive therapy environment for children having physical disabilities. In *Proceedings of the 3rd ACM conference on International conference on multimedia retrieval* (pp. 313-314). ACM. doi:10.1145/2461466.2461522
- Arici, T., Celebi, S., Aydin, A. S., & Temiz, T. T. (2014). Robust gesture recognition using feature pre-processing and weighted dynamic time warping. *Multimedia Tools and Applications*, 72(3), 3045–3062. doi:10.1007/s11042-013-1591-9
- Bó, A. P. L., Hayashibe, M., & Poignet, P. (2011, August). Joint angle estimation in rehabilitation with inertial sensors and its integration with Kinect. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, (pp. 3479-3483). IEEE. doi:10.1109/IEMBS.2011.6090940
- Brokaw, E. B., & Brewer, B. R. (2013). Development of the Home Arm Movement Stroke Training Environment for Rehabilitation (HAMSTER) and Evaluation by Clinicians. In LNCS, Virtual, Augmented and Mixed Reality, Systems and Applications, 8022 (pp. 22-31). Springer Berlin Heidelberg.
- Chang, Y. J., Chen, S. F., & Huang, J. D. (2011). A Kinect-based system for physical rehabilitation: A pilot study for young adults with motor disabilities. *Research in Developmental Disabilities*, 32(6), 2566–2570. doi:10.1016/j.ridd.2011.07.002 PMID:21784612
- Clark, R. A., Pua, Y. H., Fortin, K., Ritchie, C., Webster, K. E., Denehy, L., & Bryant, A. L. (2012). Validity of the Microsoft Kinect for assessment of postural control. *Gait & Posture*, 36(3), 372–377. doi:10.1016/j.gaitpost.2012.03.033 PMID:22633015
- Dukes, P. S., Hayes, A., Hodges, L. F., & Woodbury, M. (2013, March). Punching ducks for post-stroke neurorehabilitation: System design and initial exploratory feasibility study. In *Proceedings of the IEEE Symposium on 3D User Interfaces*, (pp. 47-54). IEEE.
- Ganesan, S., & Anthony, L. (2012, May). Using the kinect to encourage older adults to exercise: a prototype. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems* (pp. 2297-2302). ACM. doi:10.1145/2212776.2223792
- Hachaj, T., & Ogiela, M. R. (2013). Rule-based approach to recognizing human body poses and gestures in real time. *Multimedia Systems*, published online.
- LaBelle, K. (2011). Evaluation of Kinect Joint Tracking for Clinical and In-Home Stroke Rehabilitation Tools (Doctoral dissertation, Thesis, Indiana).
- Lange, B., Chang, C. Y., Suma, E., Newman, B., Rizzo, A. S., & Bolas, M. (2011, August). Development and evaluation of low cost game-based balance rehabilitation tool using the Microsoft Kinect sensor. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, (pp. 1831-1834). IEEE. doi:10.1109/IEMBS.2011.6090521
- Lange, B., Koenig, S., McConnell, E., Chang, C., Juang, R., Suma, E., & Rizzo, A. (2012, March). Interactive game-based rehabilitation using the Microsoft Kinect. In *Proceedings of the IEEE Virtual Reality Workshops* (pp. 171-172). IEEE. doi:10.1109/VR.2012.6180935
- Oszust, M., & Wysocki, M. (2013, June). Polish sign language words recognition with Kinect. In *Proceedings of the 6th International Conference on Human System Interaction* (pp. 219-226). IEEE. doi:10.1109/HSI.2013.6577826

Rafibakhsh, N., Lee, H. F., Gong, J., Ko, H. S., & Mohammadi, A. (2013). Applying Fuzzy Logic for Optimal Placement of XBOX Kinect Sensors for Industrial Applications. *Advanced Materials Research*, 628, 433–439. doi:10.4028/www.scientific.net/AMR.628.433

Staranowicz, A., & Mariottini, G. L. (2012, June). A comparative study of calibration methods for Kinect-style cameras. In *Proceedings of the 5th International Conference on Pervasive Technologies Related to Assistive Environments* (p. 49). ACM. doi:10.1145/2413097.2413159

Štrbac, M. D., Malešević, N. M., Čobeljić, R., & Schwirtlich, L. (2013). Feedback control of the forearm movement of tetraplegic patient based on Microsoft Kinect and multi-pad electrodes. *Journal of Automatic Control*, 21(1), 7–11. doi:10.2298/JAC1301007S

Su, C. (2013). Personal Rehabilitation Exercise Assistant with Kinect and Dynamic Time Warping. *International Journal of Information and Education Technology*, 3(4), 448–454. doi:10.7763/IJiet.2013.V3.316

Velloso, E., Bulling, A., & Gellersen, H. (2013, April). MotionMA: motion modelling and analysis by demonstration. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1309-1318). ACM. doi:10.1145/2470654.2466171

Zhao, W., Espy, D. D., Reinthal, M., Ekelman, B., Goodman, G., & Niederriter, J. (2015, June). Privacy-Aware Human Motion Tracking with Realtime Haptic Feedback. In *Mobile Services (MS), 2015 IEEE International Conference on* (pp. 446-453). IEEE. doi:10.1109/MobServ.2015.67

Zhao, W., Lun, R., Espy, D. D., & Ann Reinthal, M. (2014). Realtime Motion Assessment For Rehabilitation Exercises: Integration Of Kinematic Modeling With Fuzzy Inference. *Journal of Artificial Intelligence and Soft Computing Research*, 4(4), 267–285. doi:10.1515/jaiscr-2015-0014

KEY TERMS AND DEFINITIONS

Floor Clipping Plane: As part of the Microsoft Kinect SDK, each skeleton frame contains a set of parameters that can be used to determine the floor plane as seen by the Kinect sensor.

Frontal Plane: Using a three-dimensional (x-y-z) coordinate system, this is the plane dividing the front and back half of the body; movement in the frontal plane is typically viewed from an anterior or posterior perspective; an example of movement in this plane is hip abduction where the lower extremity is moved out to the side of the body from the midline.

Microsoft Kinect: A sensor produced by Microsoft that is equipped with a Webcam, a depth camera, and a microphone arrays. It is designed to facilitate a natural user interface.

Natural User Interface: A natural user interface enables users to interact with a computer or game console via gestures and voice commands instead of keyboard, mouse, or game controller.

Occlusion: An occlusion occurs when a joint or more is hidden from the camera. Self-occlusion means that one body part hides another part.

Sagittal Plane: Using a three-dimensional (x-y-z) coordinate system, this is the plane dividing the left and right sides of the body; movement in the sagittal plane is typically viewed from a lateral perspective; an example of movement in this plane is shoulder flexion where the upper extremity is moved from the side to up in front of the body.

Software Development Kit (SDK): A set of libraries that offer software developers the access of a set of application programming interfaces.

Virtual Reality as Distraction Technique for Pain Management in Children and Adolescents

M**Barbara Atzori***University of Florence, Italy***Hunter G. Hoffman***University of Washington, USA***Laura Vagnoli***Meyer Children's Hospital of Florence, Italy***Andrea Messeri***Meyer Children's Hospital of Florence, Italy***Rosapia Lauro Grotto***University of Florence, Italy*

INTRODUCTION

The problem: uncontrolled pain during medical procedures. For many medical procedures, patients remain awake during the procedure, they feel pain during the procedure, and they remember pain and unpleasantness they experienced, long after the procedure is over. These memories for the painful experience(s) can influence the patient's attitudes towards receiving healthcare, increases avoidance of medical visits that could help prevent medical problems. For example, a child who has a very painful or anxiety provoking dental procedure is more likely to become avoidant of dentists. The consequences of avoidance can be expensive. Many dental problems that could be easily prevented or corrected early on with regular checkups (e.g, filling a cavity), become much more expensive and painful problems, as the disease progresses untreated (e.g., root canal and tooth extraction) for patients who refuse to visit the dentist. Excessive pain during medical procedures is a worldwide problem, and affects a wide range of patient populations. For example, patients

with a wide range of medical problems often get intravenous (IV) ports inserted "incanulation". Once inserted, healing medications, nutrients and fluids can be administered into the patient's bloodstream through the IV port. But the initial implantation of the IV port (e.g., sticking the IV port into the patient's arm) is often surprisingly painful. The same painful port insertion procedure is used with cancer patients, burn patients, patients being prepared for surgery, and a long list of other medical problems. And IV port placement is just one of large number of medical procedures where excessive pain is common.

Some medical procedures sound fairly painless, but patients' fears and anxiety associated with the procedures can amplify the patient's negative experience. For example, with blood draws, a fairly large percentage of people faint during blood draws, despite the short duration of the procedure.

A number of psychological influences can amplify or increase how much pain patients experience during medical procedures. High anxiety can increase sensitivity to pain. Some patients

DOI: 10.4018/978-1-5225-2255-3.ch518

must receive medical procedures repeatedly. For example, children with large severe burn injuries receive several painful medical procedures per week, often daily wound cleaning and painful physical therapy skin stretching exercises, and the most painful segments of the wound care procedures often last 25 minutes or longer. If the patient has an extremely high pain during wound care, their memory for previous painful wound care sessions can increase their subjective experience of pain during subsequent wound care procedures, and make it harder to control the patient's pain using pain medications alone. There are concerns that repeated painful medical procedures can be made worse if the patient has pre-existing psychological problems such as depression or PTSD. These conditions can reduce the effectiveness of opioid analgesics. And repeated painful medical procedures during hospitalization may increase patients' risk of developing chronic pain or chronic PTSD. There has lately been a growing realization in the medical community for the need for more effective non-pharmacologic pain control during medical procedures. Use of higher doses of pain medications may increase analgesia, but also increase undesirable medical side effects of the pain medications such as nausea, delirium, constipation, urinary retention, sleeping through meals, and interference with sleeping at night. Diversion of pain medications is another concern. On the streets of the U.S.A., addiction to opioids has become a major medical problem in itself. According to the CDC, overdose deaths involving prescription opioids have quadrupled since 1999, and set a new record high in 2014 (CDC, 2016).

Because pain has a strong psychological component, in theory, psychological interventions can be used to help counteract some of the negative psychological influences. In other words, there is growing interest in using psychological techniques to help reduce pain. Typically, psychological techniques can be used in addition to traditional pain medications, adjunctively (Stinson, Yamada, Dickson, Lamba & Stevens, 2008; Koller & Goldman, 2012). Distraction is a familiar technique. Parents frequently distract children to help get

them to stop crying (using a rattle toy). Music is widely used to help patients during dental procedures. According to a Cochrane meta-analysis, listening to music resulted in a statistically significant reduction in pain, but on average, music only reduced the amount of pain by around 5% reduction (not a clinically meaningful reduction from listening to music). A stronger distraction is needed for acute pain during medical procedures.

Recently Virtual Reality (VR) has emerged as a new unusually effective distraction technique, allowing the user to be immersed in a computer-generated environment. Patients look into virtual reality goggles that block the patient from being able to watch their wound care, and earphones and computer generated sound effects and music from VR hinder the sounds of the real environment, (Hoffman et al., 2000). The user can interact with the virtual environments using a mouse or a head-tracking.

Hoffman and Patterson and colleagues co-originated the technique of using immersive virtual reality analgesia during acute procedural pain (Hoffman, 1998; Hoffman et al, 2000). Two adolescents with severe burns used immersive virtual reality during burn wound care: staple removals. Both children reported large reductions in their pain during wound care while using immersive virtual reality, compared to when the same children played a Nintendo video game during the same wound care sessions (Hoffman et al., 2000). Several controlled studies support VR efficacy in reducing pain but most of them investigated its effects on adults, in particular during procedural pain in burn patients (Hoffman, Patterson, Magula, Carrougner, Zeltzer, Dagadakis & Sharar, 2004; Maani, Hoffman, Morrow, Maiers Gaylord, McGhee & DeSocio, 2011) and during dental procedures (Bentsen, Svensson & Wenzel, 2001; Hoffman, Garcia-Palacios, Patterson, Jensen, Furness III & Ammons, 2001; Furman, Jasinevicius, Bissada, Victoroff, Skillicorn & Buchner, 2009; Wiederhold, Gao & Wiederhold, 2014).

The aim of the chapter is to review the use of VR for pediatric and adolescent pain management and its applicability in clinical settings. Between

December 2011 and September 2015 some of the most important psychological and medical databases were consulted (PubMed, ScienceDirect and Scopus), including studies in which VR was used with children and adolescents as a distraction technique for procedural pain management. Studies with only adults or that applied VR for non-procedural pain, such as chronic pain, or during non-painful procedures were excluded. Search terms included combinations of “virtual reality”, “distraction”, “analgesia”, “pain”, “children”, “adolescents”.

BACKGROUND

Although the mechanisms by which VR reduces pain are still under investigation, according to Hoffman et al., 2006, the logic of how VR reduces pain can be explained by applying the Attention Pain Theory by Eccleston and Crombez (1999). People need attention in order to feel pain. VR is very attention demanding, leaving less attentional resources available for the brain to process incoming pain signals. As a result, patients report lower pain levels during VR, they spend less time thinking about their pain, and report higher levels of fun while in VR. The essence of immersive VR is the patients’ illusion of going inside the 3D computer generated world, as if it is a place they are visiting. This illusion of “presence” or “feeling present” in the virtual world helps make VR unusually attention grabbing and unusually effective as distraction.

Slater and Wilbur (1997) made a distinction between the *immersiveness* of the VR system (e.g., the objectively measurable field of view of the VR goggles) and *presence*, the psychological illusion of going inside the computer generated world and feeling present in the virtual world. In laboratory studies, wide field of view VR goggles were significantly more effective at reducing pain than narrower field of view goggles ($p < .05$) (Hoffman, Seibel, Richards, Furness III, Patterson & Sharar, 2006). Moreover, another laboratory study with young adults (college students) showed

that an interactive VR system was significantly more effective at reducing pain than passive VR ($p < .005$) (Wender, Hoffman, Hunner, Seibel, Patterson & Sharar, 2009). In some studies, patients with a strong illusion of going inside the computer generated world were also the patients who reported greater reduction in their pain during VR, and patients who reported a less compelling illusion of going into the virtual world reported less pain reduction during VR (Hoffman, Doctor Patterson & Carrougner, 2000; Hoffman, Sharar, Coda, Everett, Ciol, Richards & Patterson 2004). Manipulating the objective *immersiveness* of the VR system has been shown to affect the magnitude of VR analgesia in predictable ways, even though the illusion of presence was not significantly higher for the high immersion VR system than for the low tech VR system ($p > .05$) (Hoffman et al., 2006; Wender et al., 2009). In other words, in some studies, increasing the *immersiveness* of the VR system significantly reduced pain, but did not significantly increase patients’ presence ratings.

VR effects on brain activity during painful stimuli have been demonstrated by Hoffman et al. (2004) in a study with adult healthy volunteers. Pain-related brain activity was measured using fMRI. During their brain scans, healthy volunteer participants received brief thermal pain stimulations during VR and during No VR. Participants reported large reductions in pain during VR, and the fMRIs showed large reductions in pain-related brain activity during VR. In another study, Hoffman et al. (2007) used pain ratings and function magnetic resonance imaging to measure how much pain participants reported experiencing, and measured pain-related brain activity in adults undergoing thermal pain stimulation and receiving opioid analgesia and/or VR distraction. Results indicated that opioids alone significantly reduced subjective pain unpleasantness ratings and significantly reduced pain related-brain activity in the insula and thalamus; VR alone significantly reduced both worst pain and pain unpleasantness and significantly reduced pain related-brain activity in the insula, thalamus and SS2, but the lowest pain scores and largest reductions in pain

related-brain activity were obtained when subjects received both opioid and VR distraction. These results support that notion that VR should be used adjunctively, in addition to traditional pain medications.

VIRTUAL REALITY ANALGESIA FOR CHILDREN AND ADOLESCENTS

Applicability and Effectiveness in Clinical Settings

VR has been used with children and adolescents between 4-19 years, mostly with burned children during bandage/dressing changes, wound cleaning or physical therapy (Hoffman 1998; Hoffman et al., 2000; Hoffman, Patterson, Carrougher & Sharar, 2001; Das, Grimmer, Sparnon, McRae & Thomas, 2005; Chan, Chung, Wong, Lien & Yang, 2007; Sharar, Carrougher, Nakamura, Hoffman, Blough & Patterson, 2007; Van Twillert, Bremer & Faber, 2007; Hoffman, Patterson, Seibel, Soltani, Jewett-Leahy & Sharar, 2008; Schmitt, Hoffman, Blough, Patterson, Jensen, Soltani, Carrougher, Nakamura & Sharar, 2011; Kipping, Rodger, Miller & Kimble, 2012; Faber, Patterson & Bremer, 2013; Hoffman, Meyer III, Ramirez, Roberts, Seibel, Atzori, Sharar & Patterson, 2014; Jeffs, Dorman, Brown, Files, Graves, Kirk, Meredith-Neve, Sanders, White & Swearingen, 2014; Hua, Yao, Zhang & Chen, 2014). Others evaluated VR analgesia for cancer with children undergoing venipuncture, lumbar puncture or port access (Sander Wint, Eshelman, Steele & Guzzetta, 2002; Gershon, Zimand, Lemos, Rothbaum & Hodges, 2003; Gershon, Zimand, Pickering, Rothbaum & Hodges, 2004; Wolitzky, Fivush, Zimand, Hodges & Rothbaum, 2005). Other studies evaluated VR effectiveness with a cerebral palsy patient during physical therapy (Steele, Grimmer, Thomas, Mulley, Fulton & Hoffman, 2003), with patients undergoing fluoride therapy and primary mandibular molar restoration (Aminabadi, Erfanparast, Sohrabi, Oskouei & Naghili, 2012) and with patients with

unknown disease during intravenous placement (Gold, Kim, Kant, Joseph & Rizzo, 2006).

VR reduced pain more than an interactive video game distraction during dressing change in a 16-years boy and in a 17-years boy (Hoffman et al., 2000), and during port access placement in a 8-years child (Gershon et al; 2003). In a pilot study, Gershon et al. (2004) randomly assigned children undergoing subcutaneous port access to three groups: VR distraction, PC screen-based distraction and no distraction. Nurses rated children in the VR and PC condition as experiencing less pain than those in control condition ($p < .05$), but no differences emerged from children reports. Similar results emerged from a study on burn children (Kipping et al., 2012).

Hua et al. (2015) found that children, their caregivers and nurse involved in the wound care reported significantly less pain levels during the dressing change in VR group compared to the standard distraction group ($p < .05$). In a randomized controlled trial study, Van Twillert et al. (2007) demonstrated that children undergoing dressing change reported significantly less pain during VR than during other forms of distraction ($p < .01$). Van Twillert et al., (2007) found that VR was somewhat superior to TV, but was not significantly more effective than watching TV.

Jeffs et al. (2014) reported that participants in the VR group reported significantly less procedural pain during VR compared to the passive distraction group (watching a movie on TV) and burn patients reported less procedural pain compared with standard care group ($p < .05$).

Seven studies also revealed significant differences between pain in the VR condition compared to a no distraction condition (Das et al., 2005; Wolitzky et al., 2005; Sharar et al., 2007; Hoffman et al., 2008; Schmitt et al., 2011; Aminabadi et al., 2012; Hoffman et al., 2014). VR continued to be effective when used for multiple treatments (Hoffman et al., 2001; Faber et al., 2013). Using relatively narrow (40 degree field of view) VR goggles, and a small sample size (ten children per group = 20 patients total), Gold et al. (2006) did not find a significant difference in pain reduction

between VR and control condition, but Gold et al.'s patients receiving VR distraction during intravenous placement were significantly more satisfied with their pain management than children in the control condition ($p < .01$). Steele et al. (2003) used VR with a 16-year-old patient with cerebral palsy undergoing physiotherapy: he reported that his pain was lower during adjunctive VR distraction than pharmacological analgesia alone.

VR Systems

The most frequently used software was "SnowWorld" an interactive virtual environment where the patients glide slowly through a 3D icy canyon and shoot snowballs at snowmen, igloos, penguins, woolly mammoths and flying fish. SnowWorld was specifically designed for burn procedural pain management and it was used also by adults in several studies about VR pain management (see reviews by Malloy & Milling, 2010; Li, Montaña, Chen & Gold, 2011). "The ice-cream factory" game, is another VR software created for children burn pain management, where subjects goal was to scare a fox away the factory by shooting it with ice-cream (Chan et al., 2007). Helmets and VR goggles used in studies varied in quality and in some cases the field of view was not mentioned in the publications.

SOLUTIONS AND RECOMMENDATIONS

On the basis of the evidence reported in the literature, VR analgesia appears to be effective for children and adolescents without side effects during dressing changes, physical therapy, lumbar puncture and port access placement. The most important medical contraindications for the use of VR regard the localization of the wound, if this doesn't allow the patient to wear the helmet, and a diagnosis of epilepsy, because the use of VR could expose children to possible epileptic seizure. These contraindications are the same as those found in

adult patients (Maani et al., 2011). In pediatric burn patients, adjunctive VR reduced pain more than a Video Game distraction. VR reduced pain more than watching an age-appropriate movie, more than other distraction techniques and adjunctive VR consistently reduced pain compared to standard pharmacological analgesia. However, we must consider that in four of these studies included both children and adults, the mean age of the sample was more than 19 and the generalizability of those results to children should be interpreted cautiously for those four studies (Hoffman et al., 2001; Van Twillert et al., 2007; Hoffman et al., 2008; Faber et al., 2013). Results exploring the use of VR with cancer patients indicate a significant VR analgesic effect during port access and a non-significant pain reduction during lumbar puncture. The lumbar puncture study (Sander Wint et al., 2002) was an early study using a passive, non-interactive VR system. Patients watched videos in a pair of low resolution VR glasses (virtual i-o glasses). VR effectiveness was likely limited by the lack of interactivity, and the narrow field of view. As mentioned earlier, Wender et al, 2009 showed that VR was significantly more effective at reducing pain when participants could actively interact with the virtual world, and Hoffman, Seibel et al (2006) showed that VR was significantly more effective at reducing pain when a wide field of view VR goggles were used (compared to narrow field of view goggles, such as those used by Sander Wint et al., 2002). Future studies using more immersive VR systems (e.g., with wide field of view VR goggles and that allow patients to actively interact with the virtual world) may get stronger results during lumbar punctures.

Because most studies didn't report helmet quality, a complete analysis on the role of this aspect wasn't possible. However, five of the six clinical studies reporting a significant pain reduction in VR condition compared to control condition, used helmet with a diagonal field of view of 50-80 degrees field of view (Hoffman et al., 2000; Hoffman et al., 2001; Sharar et al., 2007; Hoffman et al., 2008; Schmitt et al., 2011). Hua

et al, 2014 used a 40 degree FOV helmet. In two others, the helmet quality was unknown (Das et al., 2005; Aminabadi et al., 2012).

Hoffman et al. (2008) reported that pain levels were significantly lower when presence scores were higher. In contrast, Sharar et al, (2007), evaluating the differences in pain perception between adults and children/adolescents, indicated that, although children report higher levels of presence and realism than adults, there was no difference in pain reduction between the pediatric and adult age groups.

FUTURE RESEARCH DIRECTIONS

Until recently, the cost of a wide field of view VR system was tens of thousands of dollars more expensive than the narrow field of view VR goggles. In 2016/2017, the cost of a wide field of view VR system with approximately 90-110 degrees diagonal field of view, and very high resolution dropped in price from \$35,600 (not including head tracking) to approximately \$800 for one pair of wide field of view VR goggles. The \$800 helmet already includes head tracking as well as position tracked hand control input devices. There aren't any studies that describe the lifespan of a VR system, however, on the basis of our experience we can affirm that it can be considered a good investment: VR requires very low maintenance costs (mainly battery charging and cleaning gel) and, changing the VR software, it is possible to get a different VR environment and so a new distraction instrument. Moreover, most VR goggles can be adjusted on the base of inter-ocular distance and head dimensions, and it is suitable for children, adolescents and adults too.

Hoffman and colleagues recently developed a portable water-friendly VR system with a robot-like arm goggle holder (Maani et al., 2011). The custom made robot-like arm holds the goggles near the patient's eyes, and the VR goggles do not even touch the patient. This system uses the more expensive battery powered VR goggles by

NVIS MX90 VR goggles. Researchers have also been able to use a robot-like arm goggle holder with the new Oculus Rift Developers Kit Version 2 (DK2), for a patient undergoing physical therapy skin stretching exercises, with a patient who was able to tolerate having the goggles lightly touch his face (Hoffman et al., 2014).

In the U.S.A., the National Institutes of Health (NIC) are particularly interested in research on "novel therapies for pain treatment", with a specific attention to VR as demonstrated by the dozens of the funded projects using this distraction technique (Thomas, 2014). For the future, the availability of inexpensive but highly immersive VR goggles (e.g., Hoffman et al., 2014) will increase dissemination of VR studies on pain management also in other countries, including developing countries. As indicated by Hoffman et al. (2011), there is a need of more research on pediatric population, because the large numbers of children severely burned each year, but also for other kinds of painful procedures. Moreover, the importance of VR hardware and software tailored to the needs of the patient is essential to obtain a better analgesia. Recently Bidarra et al., (2013) reported that during dental procedures the patients have to reduce their movements, and created a specific software for children to use during dental procedures (Bidarra, Gambon, Kooij, Schutjes & Tzouvara, 2013). Similarly, SnowWorld, the first VR world specifically designed for pain control in burn patients, (www.vrpain.com) is specifically designed to encourage patients to sit still during wound care (e.g., as the wound care nurse uses tweezers to remove staples from an open burn wound).

CONCLUSION

More studies exclusively conducted on children (without mixed data between children and adults) are needed to evaluate the potential of VR analgesia for helping control pediatric pain in several medical settings.

In 2017, major companies such as Facebook, HTC VIVE, Sony, Magic Leap, Microsoft and others, are investing a total of several billions of dollars into developing immersive (and also augmented) virtual reality, and are mass marketing VR systems to the general public. Highly immersive VR systems are becoming more widely available, increasing the availability of immersive VR for pain control. More research using these highly immersive VR systems is needed.

ACKNOWLEDGMENT

This research was funded by NIH grants R01GM042725, R01AR054115, and R01-DA026438, Shriners Hospital for Children grant (award ID #71011-GAL, PI Walter Meyer), and by Effat University Research and Consultancy Institute, in Jeddah Saudi Arabia.

REFERENCES

- Aminabadi, N. A., Erfanparast, L., Sohrabi, A., Oskouei, S. G., & Naghili, A. (2012). The impact of virtual reality distraction on pain and anxiety during dental treatment in 4-6 year-old children: A randomized controlled clinical trial. *Journal of Dental Research, Dental Clinics, Dental Prospects.*, 6, 117–124. PMID:23277857
- Bentsen, B., Svensson, P., & Wenzel, A. (2001). Evaluation of effect of 3D video glasses on perceived pain and unpleasantness induced by restorative dental treatment. *European Journal of Pain (London, England)*, 5(4), 373–378. doi:10.1053/eujp.2001.0256 PMID:11743703
- Bidarra, R., Gambon, D., Kooij, R., Nagel, D., Schutjes, M., & Tziouvara, I. (2013). Gaming at the dentist's- serious game design for pain and discomfort distraction. In B. Schouten, S. Fedtke, T. Bekker, M. Schijven, & A. Gekker (Eds.), *Games for Health* (pp. 207–215). Wiesbaden, Germany: Springer. doi:10.1007/978-3-658-02897-8_16
- CDC. (2016). *Wide-ranging online data for epidemiologic research (WONDER)*. Atlanta, GA: CDC, National Center for Health Statistics. Available at <http://wonder.cdc.gov>. <http://www.cdc.gov/drugoverdose/data/overdose.html>
- Chan, E. A., Chung, J., Wong, T., Lien, A., & Yang, J. Y. (2007). Application of a virtual reality prototype for pain relief of pediatric burn in Taiwan. *Journal of Clinical Nursing*, 16(4), 786–793. doi:10.1111/j.1365-2702.2006.01719.x PMID:17402961
- Das, D. A., Grimmer, K. A., Sparnon, A. L., McRae, S. E. & Thomas, B. H. (2005). The efficacy of playing a virtual reality game in modulating pain for children with acute burn injuries: a randomized controlled trial. *BioMed Central Pediatrics*, 5.
- Eccleston, C., & Crombez, G. (1999). Pain demand attention: A cognitive-affective model of the interruptive function of pain. *Psychological Bulletin*, 125(3), 356–366. PMID:10349356
- Faber, A. W., Patterson, D. R., & Bremer, M. (2013). Repeated use of immersive virtual reality therapy to control pain during wound dressing changes in pediatric and adult burn patients. *Journal of Burn Care & Research; Official Publication of the American Burn Association*, 34(5), 563–568. doi:10.1097/BCR.0b013e3182777904 PMID:23970314
- Furman, E., Jasinevicius, T. R., Bissada, N. F., Victoroff, K. Z., Skillicorn, R., & Buchner, M. (2009). Virtual reality distraction for pain control during periodontal scaling and root planing procedures. *The Journal of the American Dental Association*, 140(12), 1508–1516. doi:10.14219/jada.archive.2009.0102 PMID:19955069

- Garrett, B., Taverner, T., Masinde, W., Gromala, D., Shaw, C., & Negraeff, M. (2014). A rapid evidence assessment of immersive virtual reality as an adjunct therapy in acute pain management in clinical practice. *The Clinical Journal of Pain*, 30(12), 1089–1098. doi:10.1097/AJP.0000000000000064 PMID:24535053
- Gershon, J., Zimand, E., Lemos, R., Rothbaum, B. O., & Hodges, L. (2003). Use of virtual reality as a distractor for painful procedures in a patient with pediatric cancer: A case study. *Cyberpsychology & Behavior*, 6(6), 657–661. doi:10.1089/109493103322725450 PMID:14756933
- Gershon, J., Zimand, E., Pickering, M., Rothbaum, B. O., & Hodges, L. (2004). A pilot and feasibility study of virtual reality as a distraction for children with cancer. *Journal of the American Academy of Child and Adolescent Psychiatry*, 43(10), 1243–1249. doi:10.1097/01.chi.0000135621.23145.05 PMID:15381891
- Gold, J. I., Kant, A. J., Kim, S. H., & Rizzo, A. (2005). Virtual anesthesia: The use of virtual reality for pain distraction during acute medical interventions. *Seminars in Anesthesia, Perioperative Medicine and Pain*, 24(4), 203–210. doi:10.1053/j.sane.2005.10.005
- Gold, J. I., Kim, S. H., Kant, A. J., Joseph, M. H., & Rizzo, A. (2006). Effectiveness of virtual reality for pediatric pain distraction during IV placement. *Cyberpsychology & Behavior*, 9(2), 207–212. doi:10.1089/cpb.2006.9.207 PMID:16640481
- Hoffman, H. G. (1998). Virtual Reality: A new tool for interdisciplinary Psychology research. *Cyberpsychology & Behavior*, 1(2), 195–200. doi:10.1089/cpb.1998.1.195
- Hoffman, H. G., Chambers, G. T., Meyer, W. J. III, Arceneaux, L. L., Russell, W. J., Seibel, E. J., & Patterson, D. R. et al. (2011). Virtual reality as an adjunctive non-pharmacologic analgesic for acute burn pain during medical procedures. *The Society of Behavioral Medicine*, 41(2), 183–191. doi:10.1007/s12160-010-9248-7 PMID:21264690
- Hoffman, H. G., Doctor, J. N., Patterson, D. R., Carroughe, G. J., & Furness, T. A. III. (2000). Virtual reality as an adjunctive pain control during burn wound care in adolescent patients. *Pain*, 85(1), 305–309. doi:10.1016/S0304-3959(99)00275-4 PMID:10692634
- Hoffman, H. G., Garcia-Palacios, A., Patterson, D. R., Jensen, M., Furness, T. A. III, & Ammons, W. F. Jr. (2001). The effectiveness of virtual reality for dental pain control: A case. *Cyberpsychology & Behavior*, 4(4), 527–535. doi:10.1089/109493101750527088 PMID:11708732
- Hoffman, H. G., Meyer, W. J. III, Ramirez, M., Roberts, L., Seibel, E. J., Atzori, B., & Patterson, D. R. et al. (2014). Feasibility of articulated arm mounted oculus rift virtual reality goggles for adjunctive pain control during occupational therapy in pediatric burn patients. *Cyberpsychology, Behavior, and Social Networking*, 1(6), 397–401. doi:10.1089/cyber.2014.0058 PMID:24892204
- Hoffman, H. G., Patterson, D. R., Carroughe, G. J., & Sharar, S. R. (2001). Effectiveness of virtual reality-based pain control with multiple treatments. *The Clinical Journal of Pain*, 17(3), 229–235. doi:10.1097/00002508-200109000-00007 PMID:11587113
- Hoffman, H. G., Patterson, D. R., Magula, J., Carroughe, G. J., Zeltzer, K., Dagadakis, S., & Sharar, S. R. (2004). Water-friendly reality pain control during wound care. *Journal of Clinical Psychology*, 60(2), 189–195. doi:10.1002/jclp.10244 PMID:14724926

- Hoffman, H. G., Patterson, D. R., Seibel, E., Soltani, M., Jewett-Leahy, L., & Sharar, S. R. (2008). Virtual reality pain control during burn wound debridement in the hydrotank. *The Clinical Journal of Pain*, 24(4), 299–304. doi:10.1097/AJP.0b013e318164d2cc PMID:18427228
- Hoffman, H. G., Richards, T. L., Coda, B., Bills, A. R., Blough, D., Richards, A. L., & Sharar, S. R. (2004). Modulation of thermal pain-related brain activity with virtual reality: Evidence from fMRI. *Neuro Report*, 15(8), 1245–8.
- Hoffman, H. G., Richards, T. L., Van Oostrom, T., Coda, B. A., Jensen, M. P., Blough, D. K., & Sharar, S. R. (2007). The analgesic effects of opioids and immersive virtual reality distraction: Evidence from subjective and functional brain imaging assessment. *Anesthesia and Analgesia*, 105(6), 1776–1783. doi:10.1213/01.ane.0000270205.45146.db PMID:18042882
- Hoffman, H. G., Seibel, E. J., Richards, T. L., Furness, T. A. III, Patterson, D. R., & Sharar, S. R. (2006). Virtual reality helmet display quality influences the magnitude of virtual reality analgesia. *The Journal of Pain*, 7(11), 843–850. doi:10.1016/j.jpain.2006.04.006 PMID:17074626
- Hoffman, H. G., Sharar, S. R., Coda, B., Everett, J. J., Ciol, M., Richards, T., & Patterson, D. R. (2004). Manipulating presence influences the magnitude of virtual reality analgesia. *Pain*, 111(1), 162–168. doi:10.1016/j.pain.2004.06.013 PMID:15327820
- Hua, Y., Qiu, R., Yao, W., Zhang, Q., & Chen, X. (2014). The effect of virtual reality distraction on pain relief during dressing changes in children with chronic wounds on lower limbs. *Pain Management Nursing*, 1–7. PMID:25972074
- Jeffs, D., Dorman, D., Brown, S., Files, A., Graves, T., Kirk, E., & Swearingen, C. et al. (2014). Effect of virtual reality on adolescent pain during wound care. *Journal of Burn Care & Research; Official Publication of the American Burn Association*, 35(5), 395–407. doi:10.1097/BCR.000000000000019 PMID:24823326
- Keller, D., & Goldman, R. D. (2012). Distraction techniques for children undergoing procedures: A critical review. *Journal of Pediatric Nursing*, 27(6), 652–681. doi:10.1016/j.pedn.2011.08.001 PMID:21925588
- Kipping, B., Rodger, S., Miller, K., & Kimble, R. M. (2012). Virtual reality for acute pain reduction in adolescents undergoing burn wound care: A prospective randomized controlled trial. *Burns*, 38(5), 650–657. doi:10.1016/j.burns.2011.11.010 PMID:22348801
- Li, A., Montañó, Z., Chen, V. J., & Gold, J. I. (2011). Virtual reality and pain management: Current trends and future directions. *National Institutes of Health Pain Management*, 1, 147–157. PMID:21779307
- Maani, C. V., Hoffman, H. G., Morrow, M., Maiers, A., Gaylord, K., McGhee, L. L., & DeSocio, P. A. (2011). Virtual reality pain control during burn wound debridement of combat-related burn injuries using robot-like arm mounted VR goggles. *The Journal of Trauma*, 71(supplement), S125–S130. doi:10.1097/TA.0b013e31822192e2 PMID:21795888
- Mahrer, N. E., & Gold, J. I. (2009). The use of Virtual Reality for pain control: A review. *Current Pain and Headache Reports*, 13(2), 100–109. doi:10.1007/s11916-009-0019-8 PMID:19272275
- Malloy, K. M., & Milling, L. S. (2010). The effectiveness of virtual reality distraction for pain reduction: A systematic review. *Clinical Psychology Review*, 30(8), 1011–1018. doi:10.1016/j.cpr.2010.07.001 PMID:20691523
- Sander Wint, S., Eshelman, D., Steele, J., & Guzzetta, C. E. (2002). Effects of distraction using virtual reality glasses during lumbar punctures in adolescents with cancer. *Oncology Nursing Forum*, 29(1), E8–E15. doi:10.1188/02.ONF.E8-E15 PMID:11845217

- Schmitt, Y. S., Hoffman, H. G., Blough, D. K., Patterson, D. R., Jensen, M. P., Soltani, M., & Sharar, S. R. et al. (2011). A randomized, controlled trial of immersive virtual reality analgesia, during physical therapy for pediatrics burns. *Burns*, 37(1), 61–68. doi:10.1016/j.burns.2010.07.007 PMID:20692769
- Sharar, S. R., Carrougner, G. J., Nakamura, D., Hoffman, H. G., Blough, D. K., & Patterson, D. R. (2007). Factors influencing the efficacy of virtual reality distraction analgesia during postburn physical therapy: Preliminary results from 3 ongoing studies. *Archives of Physical Medicine and Rehabilitation*, 88(12), S43–S49. doi:10.1016/j.apmr.2007.09.004 PMID:18036981
- Slater, M., & Wilbur, S. (1997). A framework for Immersive Virtual Environment (FIVE): Speculations on the role of presence in Virtual Environments. *Presence (Cambridge, Mass.)*, 6(6), 603–616. doi:10.1162/pres.1997.6.6.603
- Steele, E., Grimmer, K., Thomas, B., Mulley, B., Fulton, I., & Hoffman, H. G. (2003). Virtual reality as a pediatric pain modulation technique: A case study. *Cyberpsychology & Behavior*, 6(6), 633–638. doi:10.1089/109493103322725405 PMID:14756928
- Stinson, J., Yamada, J., Dickson, A., Lamba, J., & Stevens, B. (2008). Review of systematic reviews on acute procedural pain in children in the hospital setting. *Pain Research & Management*, 13(1), 51–57. doi:10.1155/2008/465891 PMID:18301816
- Taddio, A., & McMurtry, C. M. (2015). Psychological interventions for needle-related procedural pain and distress in children and adolescents. *Pediatrics & Child Health*, 20(4), 195–186. doi:10.1093/pch/20.4.195 PMID:26038636
- Thomas, D. A. (2014). Virtual Reality research continues to progress at the National Institutes of Health. *Cyberpsychology, Behavior, and Social Networking*, 17(6), 334. doi:10.1089/cyber.2014.1509 PMID:24892194
- Van Twillert, B., Bremer, M., & Faber, A. W. (2007). Computer-generated virtual reality to control pain and anxiety in pediatric and adult burn patients during wound dressing changes. *Journal of Burn Care & Research; Official Publication of the American Burn Association*, 28(5), 694–702. doi:10.1097/BCR.0B013E318148C96F PMID:17667488
- Wender, R., Hoffman, H. G., Hunner, H. H., Seibel, E. J., Patterson, D. R., & Sharar, S. R. (2009). Interactivity influences the magnitude of virtual reality analgesia. *Journal of Cyber Therapy and Rehabilitation*, 2, 27–33. PMID:20390047
- Wiederhold, B. K., Sobiomro, A., Riva, G., & Wiederhold, M. D. (2014). Future directions: Advances and implications of virtual environments designed for pain management. *Cyberpsychology, Behavior, and Social Networking*, 17(6), 414–422. doi:10.1089/cyber.2014.0197 PMID:24892206
- Wiederhold, M. D., Gao, K., & Wiederhold, B. K. (2014). Clinical use of virtual distraction system to reduce anxiety and pain in dental procedures. *Cyberpsychology, Behavior, and Social Networking*, 17(6), 359–365. doi:10.1089/cyber.2014.0203 PMID:24892198
- Wolitzky, K., Fivush, R., Zimand, L. H., Hodges, L., & Rothbaum, B. O. (2005). Effectiveness of virtual reality distraction during a painful medical procedure in pediatric oncology patients. *Psychology & Health*, 20(6), 817–824. doi:10.1080/14768320500143339

KEY TERMS AND DEFINITIONS

Attention: The cognitive process that let to select information from the environment in a specific moment.

Distraction: The action to draw of attention away from the primary task to another stimulus.

Pain: Unpleasant experience connected with an actual or potential tissue damage, or described in terms of such damage.

Presence: The subjective perception of being in a virtual environment and living a realistic experience.

Procedural Pain: An acute pain caused by medical-diagnostic procedures.

Snow World: The first virtual reality environment created for virtual reality analgesia with burn patients.

Virtual Environment: An artificial 3D world, that reproduces real objects or generates new ones, with which users can interact at different levels.

Virtual Reality Analgesia: Psychological technique based on distraction, used to reduce pain in adults and children.

M

Category M

Mobile and Wireless Computing

Biogeography–Based Optimization Applied to Wireless Communications Problems

M

Sotirios K. Goudos

Aristotle University of Thessaloniki, Greece

INTRODUCTION

The Evolutionary Algorithms (EAs) mimic behaviour of biological entities and they are inspired from Darwinian evolution in nature. The EAs have been extensively studied and applied to several real-world engineering problems. Biogeography-based optimization (BBO) (Simon, 2008) is a recently introduced evolutionary algorithm. BBO is based on mathematical models that describe how species migrate from one island to another, how new species arise, and how species become extinct. The way the problem solution is found is analogous to nature's way of distributing species. In the BBO approach there is a way of sharing information between solutions (Simon, 2008), similar to the other evolutionary algorithms such as GAs, DE, and PSO. This feature makes BBO suitable for the same types of problems that the other algorithms are used for, namely high-dimensional data. Additionally, BBO has some unique features, which are different from those found in the other evolutionary algorithms. For example, quite different from GAs, DE and PSO, the set of the BBO's solutions is maintained from one generation to the next and is improved using the migration model, where the emigration and immigration rates are determined by the fitness of each solution. These differences can make BBO outperform other algorithms (Simon, 2008). BBO has been applied successfully to several real world engineering problems (Ashrafinia, Pareek, Naem, & Lee, 2011; Bhattacharya & Chattopadhyay, 2010; Boussaïd, Chatterjee, Siarry, & Ahmed-Nacer, 2011; S. K. Goudos et al., 2012; Jamuna & Swarup, 2011; Kankanala, Srivastava, Srivas-

tava, & Schulz; Mandal, Bhattacharya, Tudu, & Chakraborty, 2011; Rathi, Agarwal, Sharma, & Jain, 2011; Silva, dos S Coelho, & Freire, 2010).

The purpose of this chapter is to briefly describe the BBO algorithm and present its application to antenna and wireless communications design problems. This chapter presents results from design cases that include patch antenna, linear antenna array, and a Partial Transmit Sequence (PTS) scheme for OFDM signals based on BBO. The chapter is supported with an adequate number of references. This chapter is subdivided into five sections. The "Background" Section presents the issues, problems, and trends with BBO. Then we briefly present the main BBO algorithm. In the next Section, we describe the design cases and present the numerical results. An outline of future research directions is provided in the following Section while in the "Conclusion" Section we conclude the chapter and discuss the advantages of using a BBO-based approach in the design and optimization of wireless systems and antennas. Finally, an "Additional Reading Section" gives a list of readings to provide the interested reader with useful sources in the field.

BACKGROUND

The mathematical models of Biogeography are based on the work of Robert MacArthur and Edward Wilson in the early 1960s. Using this model, they have been able to predict the number of species in a habitat. The habitat is an area that is geographically isolated from other habitats. The geographical areas that are well suited as

residences for biological species are said to have a high habitat suitability index (HSI). Therefore, every habitat is characterized by the HSI which depends on factors like rainfall, diversity of vegetation, diversity of topographic features, land area, and temperature. Each of the features that characterize habitability is known as suitability index variables (SIV). The SIVs are the independent variables while HSI is the dependent variable.

Similar to the other evolutionary algorithms (EAs) such as Genetic Algorithms (GAs), Particle Swarm Optimization (PSO), Artificial Bee Colony (ABC) optimization, in the BBO approach there is a way of sharing information between solutions (Simon, 2008). This feature makes BBO suitable for the same types of problems that the other algorithms are used for, namely high-dimensional data. Additionally, BBO has some unique features that are different from those found in the other evolutionary algorithms. For example, quite different from GAs, Ant Colony Optimization (ACO) and PSO, from one generation to the next the set of the BBO's solutions is maintained and improved using the migration model, where the emigration and immigration rates are determined by the fitness of each solution. BBO differs from PSO in the fact that PSO solutions do not change directly; the velocities change. The BBO solutions share directly their attributes using the migration models. The migration operator provides BBO with a good exploitation ability. These differences can make BBO outperform other algorithms (Ma, 2010; Ma, Simon, Fei, & Chen, 2013; Simon, 2008). It must be pointed out if PSO or ABC are constrained to discrete space then the next generation will not necessarily be discrete (Ma, et al., 2013). However, this is not true for BBO; if BBO is constrained to a discrete space then the next generation will also be discrete to the same space. As the authors in (Ma, et al., 2013) suggest this indicates that BBO could perform better than other EAs on combinatorial optimization problems, which makes BBO suitable for application to the PTS problem. The main computational cost of EAs is in the evaluation of the objective function. The BBO mechanism is

simple, like that of PSO and ABC. Therefore, for most problems, the computational cost of BBO and other EAs will be the same since it will be dominated by objective function evaluation (Ma, 2010). More details about the BBO algorithm can be found in (Ma, et al., 2013; Simon, 2008).

BIOGEOGRAPHY-BASED OPTIMIZATION ALGORITHM

Therefore, a solution to a D-dimensional problem can be represented as a vector of SIV variables:

$$[SIV_1, SIV_2, \dots, SIV_D]$$

which is a habitat or island. The value of HSI of a habitat is the value of the objective function that corresponds to that solution and it is found by:

$$HSI = F(\text{habitat}) = F(SIV_1, SIV_2, \dots, SIV_D)$$

Habitats with a high HSI are good solutions of the objective function, while poor solutions are those habitats with a low HSI. The habitats with high HSI are those that have large population and high emigration rate μ . For these habitats, the immigration rate λ is low. The immigration and emigration rates are functions of the rank of the given candidate solution. The rank of the given candidate solution represents the number of species in a habitat. These are given by:

$$\mu_k = E \left(\frac{k}{S_{\max}} \right), \lambda_k = I \left(1 - \frac{k}{S_{\max}} \right)$$

here I is the maximum possible immigration rate, E is the maximum possible emigration rate, k is the rank of the given candidate solution, and $S_{\max}$ is the maximum number of species (e.g. population size). The rank of the given candidate solution or the number of species is obtained by sorting

the solutions from most fit to least fit, according to the HSI value (e.g. fitness). That is the best candidate solution has a rank of $S_{\max}$ and the worst candidate solution has a rank of one. BBO uses both mutation and migration operators. The application of these operators to each SIV in each solution is decided probabilistically. In each generation, there is a probability $P_{\text{mod}} \in [0, 1]$ that each candidate solution will be modified by migration. P_{mod} is a user-defined parameter that is typically set to a value close to one, and is analogous to crossover probability in GAs. That is the migration operator in each generation is applied only if P_{mod} is greater than a random number generated from a uniform distribution from the interval $[0,1]$.

The migration for NP habitats can be described in Algorithm 1.

The X_i in Algorithm 1 is habitat i . The information sharing between habitats is accomplished

using the immigration and emigration rate. The λ_i is proportional to the probability that SIVs from neighboring habitats will migrate into habitat X_i . The μ_i is proportional to the probability that SIVs from habitat X_i will migrate into neighboring habitats.

The mutation rate m of a solution S is defined to be inversely proportional to the solution probability and it is given by:

$$m_s = m_{\max} \left(\frac{1 - P_s}{P_{\max}} \right)$$

where P_s is the probability that a habitat contains S species (e.g. the rank of the solution is S), $P_{\max}$ is the maximum P_s value over all $s \in [1, S_{\max}]$, and $m_{\max}$ is a user-defined parameter.

Mutation can be described using Algorithm 2.

The m_i in Algorithm 2 is the mutation rate of solution i . As with other evolutionary algorithms, BBO also incorporates elitism. This is implemented with a user-selected elitism parameter p . This means that the p best solutions remain from one generation to the other.

Algorithm 1. Habitat migration

1: for $i=1$ to NP
2: Select X_i with probability based on λ_i
3: if $rnd(0,1) < \lambda_i$ then
4: for $j=1$ to NP
5: Select X_j with probability based on μ_j
6: if $rnd(0,1) < \mu_j$ then
7: Randomly select a SIV σ from X_j
8: Replace a random SIV in X_i with σ
9: end if
10: end for
11: end if
12: end for

Algorithm 2. Habitat mutation

1: for $i=1$ to NP
2: Compute the probability P_i
3: Select SIV $X_i(j)$ with probability based on P_i
4: if $rnd(0,1) < m_i$ then
5: Replace $X_i(j)$ with a randomly generated SIV
6: end if
7: end for

APPLICATIONS, RESULTS, AND DISCUSSION

In this Section, we provide representative wireless communications design problems that are solved with the BBO algorithm. First, we present an example case of linear array synthesis by reducing the number elements required. The second example is that of an arbitrary-shaped patch antenna which can be used for satellite communications. The third example uses the BBO algorithm for addressing a common problem in wireless communication system, the reduction of peak to average power ratio (PAPR) of transmitted signals.

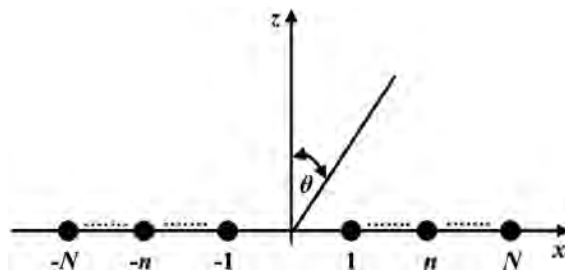
Reducing the Number of Elements in Linear Arrays using BBO

Several applications of modern wireless communications systems require radiation characteristics such as higher directivity that cannot be achieved by using a single antenna. The use of antenna arrays is the obvious solution for such a case. Among others, antenna arrays are employed for several radar and wireless communications applications in space and on earth. Their advantages include the possibility of fast scanning and precise control of the radiation pattern.

Synthesis of linear antenna arrays has been extensively studied in the last decades using several analytical or stochastic methods. The linear array design problem is multimodal and therefore requires the use of an optimization method that does not easily get trapped in local minima. The reduction of array elements plays an important role in cases where the cost, weight and size of antennas are limited. Therefore, a common goal in array synthesis is the generation of a desired pattern using the minimum number of elements. Additional constraints could include the minimum and maximum distance between two adjacent elements.

We assume a $2N$ -element linear array, which is symmetrically placed along the x -axis, see Figure 1.

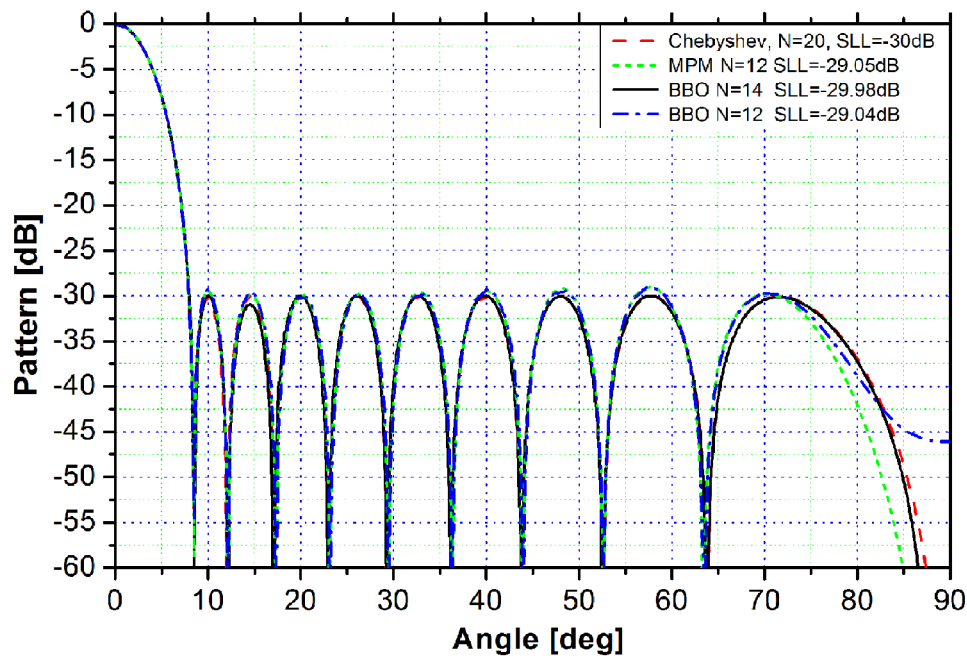
Figure 1. Geometry of the antenna array



We examine the synthesis of a pattern taken from the literature. We assume that $d_{\min} = 0.5\lambda$, and $d_{\max} = \lambda$, where $d_{\min}$, $d_{\max}$ is the minimum and maximum distance between two adjacent elements respectively.

The desired pattern in this case is that of a 20-element Chebyshev array uniformly spaced with $d = 0.5\lambda$, $SLL_d = -30$ dB, and $BW_d = 17.2^\circ$. The authors in (Yanhui, Zaiping, & Qing Huo, 2008) use a non-iterative synthesis method which is based on the matrix pencil method (MPM) (S. Yang, Y. Liu, & Liu, 2009; Yanhui, Qing Huo, & Zaiping, 2010; Yanhui, et al., 2008). We set the habitat modification probability, P_{mod} , to 1 and the maximum mutation rate, m_{max} , equal to 0.001. The maximum immigration rate I, and the maximum emigration rate E are both set both to one. The elitism parameter p is set to two. The population size and the number of generations is set to 100 and 2000 respectively. We run the algorithm for two cases with 12 and 14 elements respectively. Table 1 shows the radiation pattern characteristics. We notice that our results have improved the mean square error (MSE) compared with the results from (Yanhui, et al., 2008). Both patterns achieve the desired mainlobe beamwidth of 17.2° . The first pattern with $N=12$ slightly deviates from the desired with a sidelobe level (SLL) at -29.04 dB. The second synthesized pattern with $N=14$ matches closely the desired with $SLL = -29.98$ dB. The best reconstructed patterns are shown in Figure 2 and are compared with the example of (Yanhui, et al., 2008). This figure shows that the

Figure 2. Synthesis of a 20-element Chebyshev pattern with SLL=-30dB with $N=12$ and $N=14$ elements using BBO



performance difference between the BBO reconstructed patterns with $N=12$ and $N=14$ is due to the difference at angles greater than about 75 degrees. In order to achieve lower MSE values, two more elements are required.

Arbitrary-Shaped Dual Band Patch-Antenna Design using the BBO Algorithm

The use of microstrip patch antennas in wireless communication systems provides several advantages like low profile, low cost, and ease of fabrication. Patch antennas or arrays of such

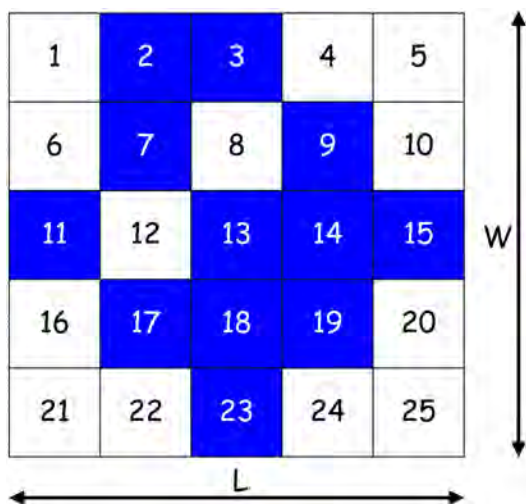
antennas in the Ku-band have practical applications in radars (Bilgic & Yegin, 2014), and satellite communications (Huayan et al., 2014; Sharma, Yadav, Kumar, Ranga, & Bhatnagar, 2011; Zhang, Mahe, & Razban, 2014).

The next design case is that of arbitrary-shaped dual band patch antenna. Figure 3 shows the patch geometry. It consists of a 5X5 grid where the blue indicates the existence of metal and the white indicates dielectric. The grid is numbered from 1 to 25 as it is shown in the figure. The patch antenna is fabricated on a dielectric substrate with $\epsilon_r = 3.38$ and 30mils thick. A coaxial feed is used, which is placed in the patch center. There-

Table 1. Comparative radiation pattern characteristics for the 20-element array design case

	Desired	MPM (Yanhui, et al., 2008) N=12	BBO N=12	BBO N=14
SLL (dB)	-30.00	-29.05	-29.04	-29.98
BW (deg)	17.20	17.20	17.20	17.20
MSE	—	7.907e-6	5.743e-6	1.299e-6

Figure 3. Arbitrary-shaped patch geometry. The blue boxes indicate presence of metal.



fore, the center grid 13 is always metal. The design goal is to minimize the S11 magnitude in two frequencies 12GHz and 14GHz. For this case, the population size is set to 10 vectors and the maximum number of iterations is set to 200. We run the BBO algorithm for 20 independent trials. The surface current distributions for the best-obtained design and the geometry are shown in Figure 4. Figure 5 presents the frequency response of the best designs. The BBO obtained antenna has a

bandwidth between 11.83 and 12.10 GHz and from 13.82GHz to 14.13GHz with $|S_{11}| < -10dB$.

A PTS Technique Based on BBO for PAPR Reduction of OFDM Signals

Orthogonal frequency division multiplexing (OFDM) is widely used in several high-bit-rate digital communication systems such as Digital Audio Broadcasting (DAB), Digital Video broadcasting (DVB) and wireless local area networks (Hara & Prasad, 1997; Van Nee & Prasad, 2000). OFDM systems still have several research challenging issues. A major drawback of OFDM signals is the high value of peak to average power ratio (PAPR). Thus, it is important to reduce the PAPR of OFDM signal in order to fully utilize the OFDM technical features. Partial transmit sequences (PTS) (Cimini Jr & Sollenberger, 2000), is a popular PAPR reduction method with good PAPR reduction performance. However, PTS requires an exhaustive search in order to find the optimal phase factors. Thus, the search complexity is high. Several methods have been published in the literature for PAPR reduction using PTS with low search complexity.

In an OFDM system, the high-rate data stream is split into N low-rate data streams that are si-

Figure 4. Simulated surface current distribution of the best design found by BBO (a) at 12 GHz, (b) at 14 GHz

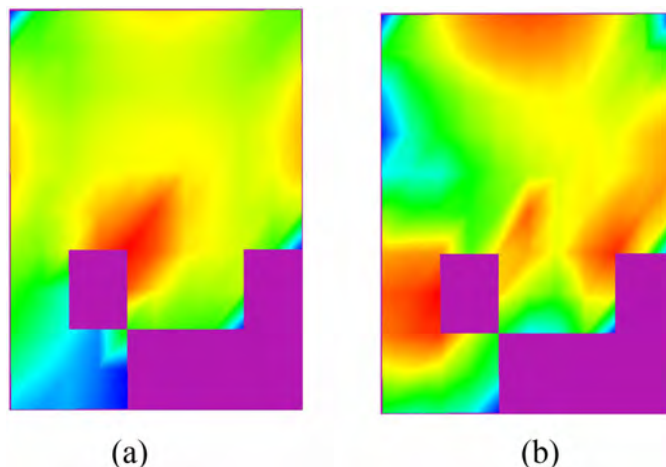
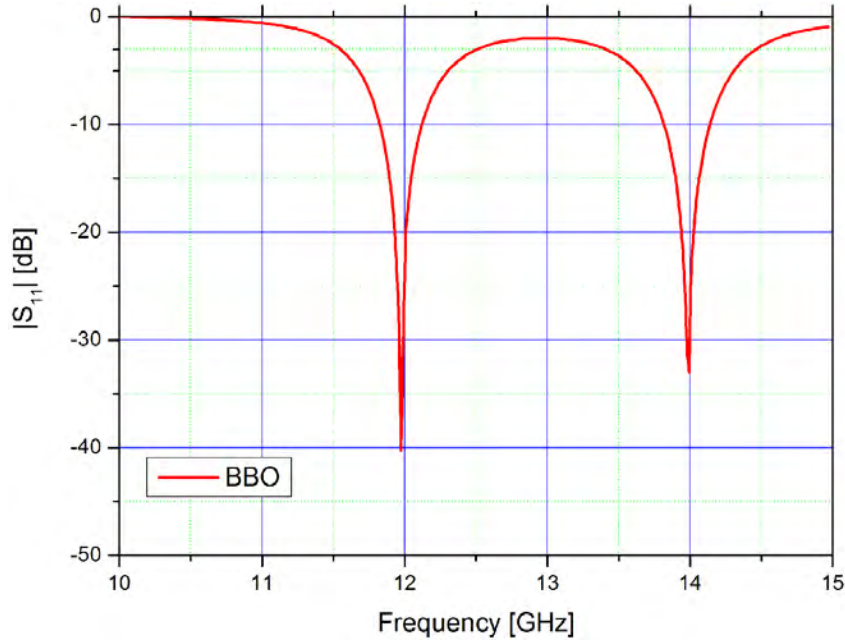


Figure 5. Simulated S11 curve for the best design found by BBO



multaneously transmitted using N subcarriers. The discrete-time signal of such a system is given by:

$$s_k = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} S_n e^{\frac{j2\pi nk}{LN}} \quad k = 0, 1, \dots, LN - 1$$

where L is the oversampling factor, $S = [S_0, S_1, \dots, S_{N-1}]^T$ is the input signal block. Each symbol is modulated by either phase-shift keying (PSK) or quadrature amplitude modulation (QAM).

In the PTS approach the S input data OFDM block is partitioned into M disjoint subblocks represented by the vector S_m $m = 1, 2, \dots, M - 1$ and oversampled by inserting $(L - 1)N$ zeros. Next, the subblocks are converted to time domain using LN point inverse fast fourier transform (IFFT). The PTS objective is to produce a weighted combination of the M subblocks using $b = [b_1, b_2, \dots, b_M]^T$ complex phase factors to minimize PAPR. In order to reduce the search complexity the phase factor possible values are limited to a finite set. The set of allowable phase factors is:

$$A = \left\{ e^{\frac{j2\pi n}{W}} \mid n = 0, 1, \dots, W - 1 \right\}$$

where W is the number of allowed phase factors. Therefore in case of M subblocks and W phase factors the total number of possible combinations is W^M . In order to reduce the search complexity we usually set fixed one phase factor.

The block diagram of the PTS technique is shown in Figure 6.

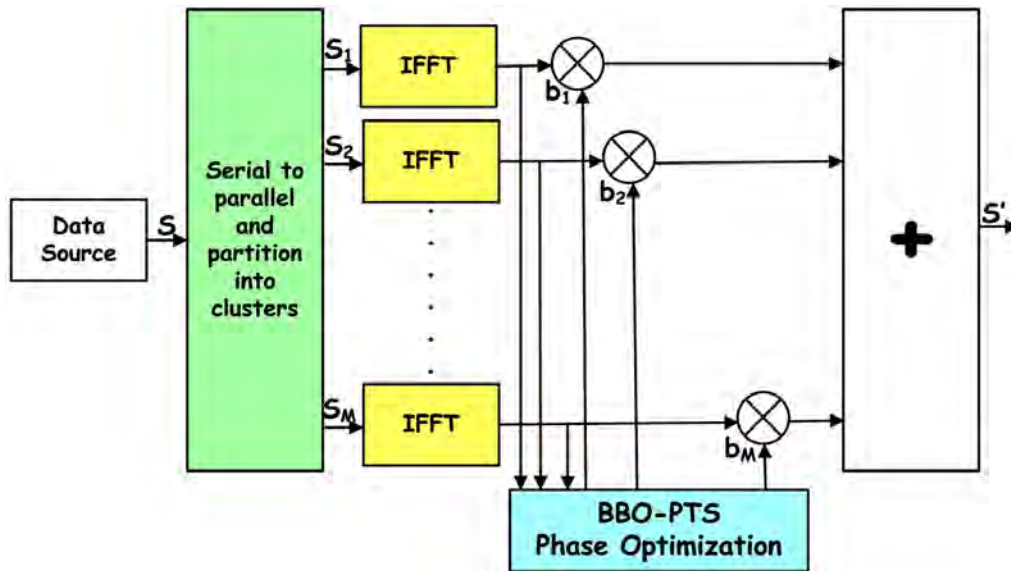
The optimization goal of the PTS scheme is to find the optimum phase combination for minimum PAPR. Thus, the objective function can be expressed as:

$$\text{Minimize: } F(b) = 10 \log_{10} \frac{\max_{0 \leq k \leq LN-1} |s'(b)|^2}{E[|s'(b)|^2]}$$

subject to:

$$b \in \left\{ e^{j\phi_m} \right\}^M \text{ where } \phi_m \in \left\{ \frac{2\pi l}{W} \mid l = 0, 1, \dots, W - 1 \right\}$$

Figure 6. Block diagram of the PTS technique



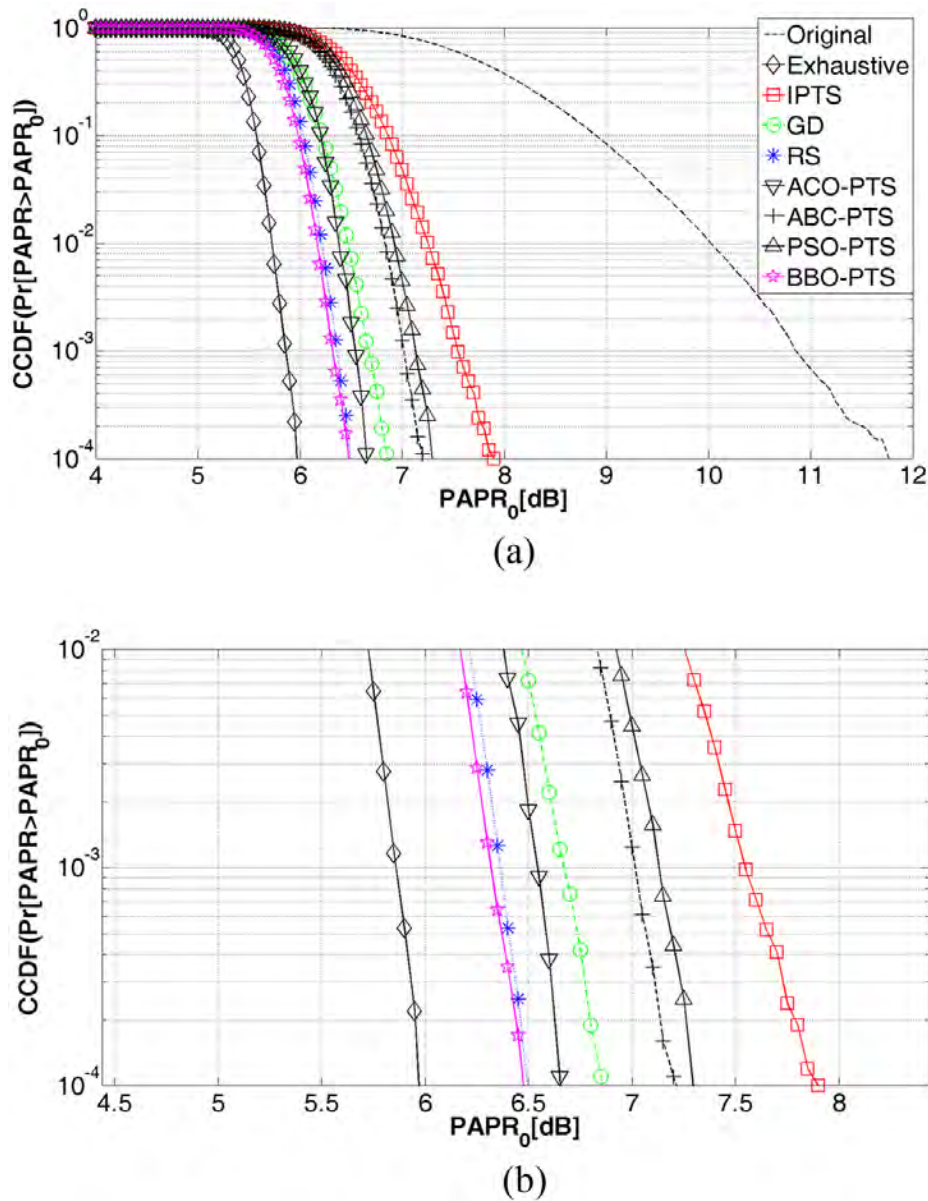
We have evaluated the objective function above using different evolutionary algorithms and methods found in the literature. We have used two main measurement criteria namely the complementary cumulative distribution function (CCDF) and the computational complexity. In all our simulations, 10E5 random OFDM blocks are generated. The transmitted signal is oversampled by a factor $L=4$. We consider 16-QAM modulation with $N=256$ sub-carriers which are divided into $M=16$ random subblocks. The phase factors for $W=2$ are selected. We consider the first phase factor to be fixed so the total number of unknown phase factors is $M-1$. Figures 7a and 7b show the comparison between the CCDF by BBO-PTS and other PTS reduction techniques. For $\Pr(PAPR > PAPR_0) = 10^{-3}$ the PAPR of the original OFDM transmitted signal is 10.84dB. For all evolutionary algorithms, the population size NP is set to 30 and the maximum number of generations G is set to 30. Thus, the computational complexity of BBO-PTS for this case is $NP \times G = 900$. The PAPR by the BBO-PTS is 6.32dB. The computational complexity of the exhaustive search is $W^{M-1} = 32768$ while the PAPR for this case is 5.86dB. The PAPR by the iterative

flipping algorithm for PTS (IPTS) (Cimini Jr & Sollenberger, 2000) is 7.55dB with search complexity $(M-1)W=30$. The PAPR by the gradient descent method (GD) (Han & Lee, 2004) with search complexity $C_{M-1}^r W^r I = C_{15}^2 2^2 3 = 1260$ is 6.96dB. We also use the Random Search (RS) (Cimini Jr & Sollenberger, 2000) method by selecting 900 random phase factors. The PAPR for this case is 6.37dB. The PAPR by Artificial Bee Colony (ABC) PTS (Y. Wang, Chen, & Tellambura, 2010), Particle Swarm Optimization (PSO) PTS (Wen, Lee, Huang, & Hung, 2008), and Ant Colony Optimization (ACO) PTS (Dorigo & Stutzle, 2004), is 7.01dB, 7.13dB, and 6.52dB, respectively.

FUTURE RESEARCH DIRECTIONS

The research domain of biogeography-based optimization is growing rapidly. Several extensions of the original algorithm have been emerged. Among others a trend in BBO research is to use different migration models like curve shapes (Ma & Simon, 2011a), and blended migration (Ma & Simon, 2011b). A performance comparison

Figure 7. a) PARP reduction performance comparison of BBO-PTS algorithm with other PTS schemes for $NP=30$, $G=30$. b) Zoom in view of PARP reduction performance comparison of BBO-PTS.



of different migration models is given in (Guo, Wang, & Wu, 2014) and in (Ma, 2010). Other BBO variants proposed include (Ma & Simon, 2011a), blended migration (Ma & Simon, 2011b), oppositional BBO (Roy & Mandal, 2011), real-coded BBO (Gong, Cai, Ling, & Li, 2010), chaotic BBO (Saremi, Mirjalili, & Lewis, 2014), and BBO with a polyphyletic migration operator (Xiong, Shi, & Duan, 2014).

BBO hybrids with other algorithms is also a growing research trend. BBO has been combined with Differential Evolution (DE) (Gong, Cai, & Ling, 2011), with evolutionary strategies (Du, Simon, & Ergezer, 2009), with Harmony Search (HS) algorithm (G. Wang et al., 2013), and local search methods (Feng et al., 2013). The application of multi-objective BBO algorithms (Costa E Silva, Coelho, & Lebensztajn, 2012; Jamuna

& Swarup, 2012; Li & Yin, 2013) to antenna and wireless design problems introduces new challenges regarding performance and computational cost. Recently a novel Multi-objective BBO algorithm has been applied to wireless networks design (Sotirios K. Goudos, Plets, Liu, Martens, & Joseph, 2015). Therefore, new research directions have to be explored.

CONCLUSION

In this chapter, we discussed and evaluated the application of BBO optimizers to antenna and wireless communications design problems. BBO is a recently introduced global stochastic optimizer that is based on mathematical models of species migration. Our results show the efficiency of this approach for antenna synthesis. In the design case presented here, the BBO results are better than those found in the literature. The BBO algorithm has been applied to the design of an arbitrary-shaped patch antenna. The obtained best design is quite efficient in both frequencies. Additionally, we have used BBO a new low complexity PTS scheme. The comparison with other existing PTS reduction methods shows that BBO-PTS performs better than other methods with the same computational complexity. Simulation results show that BBO-PTS is an efficient and powerful tool for PAPR reduction of OFDM signals.

Recent BBO algorithms and their application to wireless communications problems will be part of our future work. The BBO algorithms can be valuable tools for constrained single or multi-objective optimization problems in antennas and wireless communications.

REFERENCES

- Ashrafinia, S., Pareek, U., Naeem, M., & Lee, D. (2011). *Source and Relay Power Selection Using Biogeography-Based Optimization for Cognitive Radio Systems*. Paper presented at the 2011 IEEE Vehicular Technology Conference (VTC Fall).
- Bhattacharya, A., & Chattopadhyay, P. K. (2010). Biogeography-Based Optimization for Different Economic Load Dispatch Problems. *IEEE Transactions on Power Systems*, 25(2), 1064–1077. doi:10.1109/TPWRS.2009.2034525
- Bilgic, M. M., & Yegin, K. (2014). Wideband Offset Slot-Coupled Patch Antenna Array for X/Ku-Band Multimode Radars. *Antennas and Wireless Propagation Letters, IEEE*, 13, 157–160. doi:10.1109/LAWP.2013.2296911
- Boussaïd, I., Chatterjee, A., Siarry, P., & Ahmed-Nacer, M. (2011). Hybridizing biogeography-based optimization with differential evolution for optimal power allocation in wireless sensor networks. *IEEE Transactions on Vehicular Technology*, 60(5), 2347–2353. doi:10.1109/TVT.2011.2151215
- Cimini, L. J. Jr, & Sollenberger, N. R. (2000). Peak-to-average power ratio reduction of an OFDM signal using partial transmit sequences. *IEEE Communications Letters*, 4(3), 86–88. doi:10.1109/4234.831033
- Costa, E., Silva, M. D. A., Coelho, L. D. S., & Lebensztajn, L. (2012). Multiobjective biogeography-based optimization based on predator-prey approach. *IEEE Transactions on Magnetics*, 48(2), 951–954. doi:10.1109/TMAG.2011.2174205
- Dorigo, M., & Stutzle, T. (2004). *Ant Colony Optimization*. Cambridge, MA: The MIT Press.
- Du, D., Simon, D., & Ergezer, M. (2009). *Biogeography-based optimization combined with evolutionary strategy and immigration refusal*. Paper presented at the Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics. doi:10.1109/ICSMC.2009.5346055
- Feng, Q., Liu, S., Wu, Q., Tang, G., Zhang, H., & Chen, H. (2013). Modified biogeography-based optimization with local search mechanism. *Journal of Applied Mathematics*.

- Gong, W., Cai, Z., & Ling, C. X. (2011). DE/BBO: A hybrid differential evolution with biogeography-based optimization for global numerical optimization. *Soft Computing*, 15(4), 645–665. doi:10.1007/s00500-010-0591-1
- Gong, W., Cai, Z., Ling, C. X., & Li, H. (2010). A real-coded biogeography-based optimization with mutation. *Applied Mathematics and Computation*, 216(9), 2749–2758. doi:10.1016/j.amc.2010.03.123
- Goudos, S. K., Baltzis, K. B., Siakavara, K., Samaras, T., Vafiadis, E., & Sahalos, J. N. (2012). *Reducing the number of elements in linear arrays using biogeography-based optimization*. doi:10.1109/EuCAP.2012.6206314
- Goudos, S. K., Plets, D., Liu, N., Martens, L., & Joseph, W. (2015). A multi-objective approach to indoor wireless heterogeneous networks planning based on biogeography-based optimization. *Computer Networks*, 91, 564–576. doi:10.1016/j.comnet.2015.08.037
- Guo, W., Wang, L., & Wu, Q. (2014). An analysis of the migration rates for biogeography-based optimization. *Information Sciences*, 254, 111–140. doi:10.1016/j.ins.2013.07.018
- Han, S. H., & Lee, J. H. (2004). PAPR reduction of OFDM signals using a reduced complexity PTS technique. *IEEE Signal Processing Letters*, 11(11), 887–890. doi:10.1109/LSP.2004.833490
- Hara, S., & Prasad, R. (1997). Overview of multi-carrier CDMA. *IEEE Communications Magazine*, 35(12), 126–133. doi:10.1109/35.642841
- Huayan, J., Kuo-Sheng, C., Wenquan, C., Chih-Chun, C., Hua-Juan, L., & Quan, X. (2014). Differential-Fed Patch Antenna Arrays With Low Cross Polarization and Wide Bandwidths. *Antennas and Wireless Propagation Letters, IEEE*, 13, 1069–1072. doi:10.1109/LAWP.2014.2328352
- Jamuna, K., & Swarup, K. S. (2011). *Power system observability Using Biogeography Based Optimization*. Paper presented at the International Conference on Sustainable Energy and Intelligent Systems (SEISCON 2011). doi:10.1049/cp.2011.0394
- Jamuna, K., & Swarup, K. S. (2012). Multi-objective biogeography based optimization for optimal PMU placement. *Applied Soft Computing*, 12(5), 1503–1510. doi:10.1016/j.asoc.2011.12.020
- Kankanala, P., Srivastava, S. C., Srivastava, A. K., & Schulz, N. N. Optimal Control of Voltage and Power in a Multi-Zonal MVDC Shipboard Power System. *IEEE Transactions on Power Systems*, (99), 1-9.
- Li, X., & Yin, M. (2013). Multiobjective binary biogeography based optimization for feature selection using gene expression data. *IEEE Transactions on Nanobioscience*, 12(4), 343–353. doi:10.1109/TNB.2013.2294716 PMID:25003163
- Ma, H. (2010). An analysis of the equilibrium of migration models for biogeography-based optimization. *Information Sciences*, 180(18), 3444–3464. doi:10.1016/j.ins.2010.05.035
- Ma, H., & Simon, D. (2011a). Analysis of migration models of biogeography-based optimization using Markov theory. *Engineering Applications of Artificial Intelligence*, 24(6), 1052–1060. doi:10.1016/j.engappai.2011.04.012
- Ma, H., & Simon, D. (2011b). Blended biogeography-based optimization for constrained optimization. *Engineering Applications of Artificial Intelligence*, 24(3), 517–525. doi:10.1016/j.engappai.2010.08.005
- Ma, H., Simon, D., Fei, M., & Chen, Z. (2013). On the equivalences and differences of evolutionary algorithms. *Engineering Applications of Artificial Intelligence*, 26(10), 2397–2407. doi:10.1016/j.engappai.2013.05.002

- Mandal, K. K., Bhattacharya, B., Tudu, B., & Chakraborty, N. (2011). *A novel population-based optimization algorithm for optimal distribution capacitor planning*. Paper presented at the 2011 International Conference on Energy, Automation, and Signal (ICEAS). doi:10.1109/ICEAS.2011.6147075
- Rathi, A., Agarwal, A., Sharma, A., & Jain, P. (2011). *A new hybrid technique for solution of economic load dispatch problems based on Biogeography Based Optimization*. Paper presented at the TENCON 2011 - 2011 IEEE Region 10 Conference.
- Roy, P. K., & Mandal, D. (2011). Quasi-oppositional biogeography-based optimization for multi-objective optimal power flow. *Electric Power Components and Systems*, 40(2), 236–256. doi:10.1080/15325008.2011.629337
- Saremi, S., Mirjalili, S., & Lewis, A. (2014). Biogeography-based optimisation with chaos. *Neural Computing & Applications*, 25(5), 1077–1097. doi:10.1007/s00521-014-1597-x
- Sharma, M. M., Yadav, S., Kumar, A., Ranga, Y., & Bhatnagar, D. (2011). *Compact elliptical microstrip patch antenna with slotted ground for Ku-band applications*. Paper presented at the Applied Electromagnetics Conference (AEMC).
- Silva, M. A. C., dos S Coelho, L., & Freire, R. Z. (2010). *Biogeography-based Optimization approach based on Predator-Prey concepts applied to path planning of 3-DOF robot manipulator*. Paper presented at the 2010 IEEE Conference on Emerging Technologies and Factory Automation (ETFA). doi:10.1109/ETFA.2010.5641274
- Simon, D. (2008). Biogeography-Based Optimization. *IEEE Transactions on Evolutionary Computation*, 12(6), 702–713. doi:10.1109/TEVC.2008.919004
- Van Nee, R., & Prasad, R. (2000). *OFDM for wireless multimedia communications*. OFDM for Wireless Multimedia Communications.
- Wang, G., Guo, L., Duan, H., Wang, H., Liu, L., & Shao, M. (2013). Hybridizing harmony search with biogeography based optimization for global numerical optimization. *Journal of Computational and Theoretical Nanoscience*, 10(10), 2312–2322. doi:10.1166/jctn.2013.3207
- Wang, Y., Chen, W., & Tellambura, C. (2010). A PAPR reduction method based on artificial bee colony algorithm for OFDM signals. *IEEE Transactions on Wireless Communications*, 9(10), 2994–2999. doi:10.1109/TWC.2010.081610.100047
- Wen, J. H., Lee, S. H., Huang, Y. F., & Hung, H. L. (2008). A suboptimal PTS algorithm based on particle swarm optimization for PAPR reduction in OFDM systems. *EURASIP J. Wireless Commun. Networking*, 2008(14).
- Xiong, G., Shi, D., & Duan, X. (2014). Enhancing the performance of biogeography-based optimization using polyphyletic migration operator and orthogonal learning. *Computers & Operations Research*, 41(1), 125–139. doi:10.1016/j.cor.2013.07.021
- Yang, S., Liu, Y., & Liu, Q. H. (2009). Combined strategies based on matrix pencil method and tabu search algorithm to minimize elements of non-uniform antenna array. *Progress In Electromagnetics Research B*, 18, 259–277. doi:10.2528/PIERB09092408
- Yanhui, L., Qing Huo, L., & Zaiping, N. (2010). Reducing the Number of Elements in the Synthesis of Shaped-Beam Patterns by the Forward-Backward Matrix Pencil Method. *IEEE Transactions on Antennas and Propagation*, 58(2), 604–608. doi:10.1109/TAP.2009.2037709

Yanhui, L., Zaiping, N., & Qing Huo, L. (2008). Reducing the Number of Elements in a Linear Antenna Array by the Matrix Pencil Method. *IEEE Transactions on Antennas and Propagation*, 56(9), 2955–2962. doi:10.1109/TAP.2008.928801

Zhang, H., Mahe, Y., & Razban, T. (2014). Low-cost Ku-band dual-polarized and beam switchable cross-type antenna array for satellite communications. *Microwave and Optical Technology Letters*, 56(11), 2656–2659. doi:10.1002/mop.28670

ADDITIONAL READING

Ashrafinia, S., Pareek, U., Naeem, M., & Lee, D. (2011a, 11-15 April 2011). Biogeography-based optimization for joint relay assignment and power allocation in cognitive radio systems. Paper presented at the 2011 IEEE Symposium on Swarm Intelligence (SIS). doi:10.1109/SIS.2011.5952589

Ashrafinia, S., Pareek, U., Naeem, M., & Lee, D. (2011b, 5-8 Sept. 2011). Source and Relay Power Selection Using Biogeography-Based Optimization for Cognitive Radio Systems. Paper presented at the 2011 IEEE Vehicular Technology Conference (VTC Fall).

Bhattacharya, A., & Chattopadhyay, P. K. (2010a). Biogeography-Based Optimization for Different Economic Load Dispatch Problems. *IEEE Transactions on Power Systems*, 25(2), 1064–1077. doi:10.1109/TPWRS.2009.2034525

Bhattacharya, A., & Chattopadhyay, P. K. (2010b). Solving complex economic load dispatch problems using biogeography-based optimization. *Expert Systems with Applications*, 37(5), 3605–3615. doi:10.1016/j.eswa.2009.10.031

Bhattacharya, A., & Chattopadhyay, P. K. (2011). Hybrid differential evolution with biogeography-based optimization algorithm for solution of economic emission load dispatch problems. *Expert Systems with Applications*, 38(11), 14001–14010.

Boussaïd, I., Chatterjee, A., Siarry, P., & Ahmed-Nacer, M. (2011). Hybridizing biogeography-based optimization with differential evolution for optimal power allocation in wireless sensor networks. *IEEE Transactions on Vehicular Technology*, 60(5), 2347–2353. doi:10.1109/TVT.2011.2151215

Garg, V., & Deep, K. (2015). A state-of-the-art review of biogeography-based optimization.: Vol. 336. *Advances in Intelligent Systems and Computing* (pp. 529–545). Springer Verlag.

Jamuna, K., & Swarup, K. S. (2011, 20-22 July 2011). Power system observability Using Biogeography Based Optimization. Paper presented at the International Conference on Sustainable Energy and Intelligent Systems (SEISCON 2011). doi:10.1049/cp.2011.0394

Lohokare, M. R., Pattnaik, S. S., Panigrahi, B. K., & Das, S. (2013). Accelerated biogeography-based optimization with neighborhood search for optimization. *Applied Soft Computing*, 13(5), 2318–2342. doi:10.1016/j.asoc.2013.01.020

Pattnaik, S. S., Lohokare, M. R., & Devi, S. (2010). Enhanced biogeography-based optimization using modified clear duplicate operator. Paper presented at the Proceedings - 2010 2nd World Congress on Nature and Biologically Inspired Computing, NaBIC 2010. doi:10.1109/NABIC.2010.5716322

Rathi, A., Agarwal, A., Sharma, A., & Jain, P. (2011, 21-24 Nov. 2011). A new hybrid technique for solution of economic load dispatch problems based on Biogeography Based Optimization. Paper presented at the TENCON 2011 - 2011 IEEE Region 10 Conference.

Roy, P. K., Ghoshal, S. P., & Thakur, S. S. (2010). Biogeography based optimization for multi-constraint optimal power flow with emission and non-smooth cost function. *Expert Systems with Applications*, 37(12), 8221–8228. doi:10.1016/j.eswa.2010.05.064

Roy, P. K., & Mandal, D. (2011). Quasi-oppositional biogeography-based optimization for multi-objective optimal power flow. *Electric Power Components and Systems*, 40(2), 236–256. doi:10.1080/15325008.2011.629337

Sharaqa, A., & Dib, N. (2011, 6-8 Dec. 2011). Design of linear and circular antenna arrays using biogeography based optimization. Paper presented at the 2011 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT). doi:10.1109/AEECT.2011.6132503

Sharaqa, A., & Dib, N. (2014). Design of Linear and Elliptical Antenna Arrays Using Biogeography Based Optimization. *Arabian Journal for Science and Engineering*, 39(4), 2929–2939. doi:10.1007/s13369-013-0794-8

Silva, M. A. C., dos S Coelho, L., & Freire, R. Z. (2010, 13-16 Sept. 2010). Biogeography-based Optimization approach based on Predator-Prey concepts applied to path planning of 3-DOF robot manipulator. Paper presented at the 2010 IEEE Conference on Emerging Technologies and Factory Automation (ETFA). doi:10.1109/ETFA.2010.5641274

Simon, D., Omran, M. G. H., & Clerc, M. (2014). Linearized biogeography-based optimization with re-initialization and local search. *Information Sciences*, 267, 140–157. doi:10.1016/j.ins.2013.12.048

Singh, U., Kumar, H., & Kamal, T. S. (2010). Design of Yagi-Uda Antenna Using Biogeography Based Optimization. *IEEE Transactions on Antennas and Propagation*, 58(10), 3375–3379. doi:10.1109/TAP.2010.2055778

Singh, U., Kumar, H., & Kamal, T. S. (2010). Linear array synthesis using biogeography based optimization. *Progress In Electromagnetics Research M*, 11, 25–36. doi:10.2528/PIERM09120201

Wang, L., & Xu, Y. (2011). An effective hybrid biogeography-based optimization algorithm for parameter estimation of chaotic systems. *Expert Systems with Applications*, 38(12), 15103–15109. doi:10.1016/j.eswa.2011.05.011

Xu, Z. D., Mo, H. W., Xu, Z. J., & Zhang, J. (2013). *Multi-objective biogeography-based optimization algorithm for the control of robotic manipulator* (Vol. 659, pp. 59–63).

KEY TERMS AND DEFINITIONS

Genetic Algorithms: A stochastic population-based global optimization technique that mimics the process of natural evolution.

Mainlobe Beamwidth: In a radio antenna's radiation pattern, the main lobe, or main beam is the lobe containing the maximum power. The beamwidth of the main lobe is the width of the main lobe specified by the angles between the points on the side of the lobe where the power has fallen to zero.

Sidelobe Level (SLL): The ratio, usually expressed in decibels (dB), of the amplitude at the peak of the main lobe to the amplitude at the peak of a side lobe.

Orthogonal Frequency-Division Multiplexing (OFDM): A method of encoding digital data on multiple carrier frequencies.

Digital Audio Broadcasting (DAB): A digital radio technology for broadcasting radio stations.

Digital Video broadcasting (DVB): A suite of internationally accepted open standards for digital television.

Wireless Local Area Network (WLAN): A network in which a mobile user can connect to a local area network (LAN) through a wireless (radio) connection. Most modern WLANs are based on IEEE 802.11 standards, marketed under the Wi-Fi brand name.

BYOD (Bring Your Own Device), Mobile Technology Providers, and Its Impacts on Business/Education and Workplace/Learning Applications

M

Amber A. Smith-Ditizio
Texas Woman's University, USA

Alan D. Smith
Robert Morris University, USA

INTRODUCTION

Concepts of BYOD

Until recently, was the norm for the employee could only do company work on a company device. Technological advances are now assisting individual's access to company information and applications on personally owned devices. BYOD (i.e., bring your own device), is making significant advances in the business world with about 44% of employees already using it at work. Companies are having a hard time controlling the employees from bringing their devices into the office and believe that implementing a BYOD policy will help employees be more productive.

The concept of BYOD was introduced in 2009 and has grown rapidly in the past few years. As of 2015, 74% managers/professional workers stated that their organization is using or planning to use BYOD" and "just over a quarter of have ruled it out entirely (Hammond). According to Apple, using BYOD can boost mobile workforce. "Users connect their mobile devices to the network without assistance from IT, and get first-tier technical support through online reference documents and instructional videos" (Apple). Being able to use these mobile devices in a group setting can substantially improve work efficiency. The devices are more mobile and are easily functioned.

Another process in information systems that BYOD could be implemented with is cloud computing. When the business has a cloud server, the employees will be able to access company data and information from their mobile devices and desktop computers in the firm. Some businesses that already have BYOD are using their devices to connect to cloud services in order to save space on the firm's computers. Although cloud computing could be beneficial for the firms in terms of storage, some employees and administrators feel uneasy about cloud computing. Some of the data that needs to be stored is often confidential information about customers about the business. Allowing that information to be uploaded to the Internet and open for everyone to see could be hazardous for the firm in the long run. When employees bring their own device, they have to make sure that their devices are not jail broken or altered in anyway. This could cause data leakages within the company, which could later lead to lasting consequences. Anyone could get ahold of their information and leak data to the world. Ensuring that the firm's cloud computing is secured and protected is one of the biggest challenges in implementing it.

According to Qrunfleh, Tarafdar, & Ragu-Nathan (2012), one main cost for any firm is a lack of knowledge of supply chain integration and how to manage a company's inventory. A company

does not want to order as many products as they can and expect to sell items immediately and successfully. BYOD would allow the flexibility of all members in the supply chain to access the level of communication technology needed to develop long-term supplier integration. It is very important to understand how supplier management practices and IS strategies can help a business improve their supply chain performance. To improve supply chain performance, a business needs to adopt specific information systems that support the processes and activities of their supply chain. In order to have an effective and enhanced supply chain performance, IS that a focal firm deploys in its supply chain should support and enhance the objectives and goals of supplier management practices. IS can assist a company with inventory control and provide its suppliers with production planning, so that, the right amount of inventory is made and that not too much is in production. The alignment of these aspects can provide a company with a low-cost supply chain and can reduce its labor cost as well as inventory costs. By using certain IS application aligned with supplier practices a company can have an improved supply chain performance.

Outside the traditional business institutions, a number of academic institutions are looking to add BYOD policies. For example, some K-12 schools are supplying teachers with devices that can be an aid in learning and universities are allowing students to bring their own devices to add to the enrichment of the programs. BYOD-based technology and its applications have numerous benefits, but they can be difficult to implement. Obviously, as with any technological innovation, there will always be adoption considerations and issues (Bruner & Kumar, 2005; Burton-Jones & Hubona, 2005; Chau, 1996; Chen, Gillenson, & Sherrell, 2002). Such technological innovations usually result in lower operating costs for the business, but they often require the organization to adapt (Belmans, Mobley, Page, Uy, & Lambrette, 2012). Since there are so many constraints and situations that could inhibit the security of

the business, some businesses are hesitant to use BYOD in their business.

Growth of Mobile Devices and BYOD

Few companies have made as big of an impact on our society and more specifically world culture as Apple. It is many mobile device and applications offerings, which has significantly fueled the BYOD movement. Apple, once a company on the verge of fading into oblivion made an incredible comeback, using a combination of marketing techniques and devices unlike any of its competitors. Whether leaking information deliberately to generate a marketing campaign or creating commercials that emphasize its unique draw to the business professionals and millennials, Apple, like its competition in Microsoft and Samsung, can be considered an innovative marketing company. Many researchers believe that companies like Apple that have strategically leveraged electronic word-of-mouth (eWOM) and social media platforms to become a model of business practices over the last decade (Sinha & Thirumalai, 2011; Smith, 2012; Smith, Synowka, Clark, & Smith, 2012; Smock, Ellison, Lampe, & Wohn, 2011). Such companies as Apple have not only been turning out revolutionary products, but their marketing methods (e.g., specifically the product launches) have been brilliant. Because of this, they have had one of the fastest expanding market presences in recent memory over the last decade. Once a company that lagged far behind Microsoft in every important category, Apple has now carved out its own niche, cornering the market on hand-held electronic devices like iPods, iPhones, and iPads. The authors of this chapter firmly believe that they could be in a position to strategically leverage the BYOD movement to the next level with their product and accompanying security features to allow professionals to use sensitive corporate and customer data.

Apple excels at turning brand into market shares. When the iPod was first introduced, Apple made sure that it had its iTunes store in place. In

addition to marketing MP3 players, Apple made money from selling downloadable music. The original iTunes store sold music that could only play on iPod players, making it a proprietary marketing strategy. That basic strategy continued with the arrival of the iPhone and the iPad, which use proprietary software applications.

Once Apple succeeded in convincing buyers that the only MP3 player worth having was an iPod, it set its sights on the next innovation, the cell phone market. In order to begin the marketing campaign, rumors began surfacing online in late 2006 that Apple was planning on looking into getting into mobile phones. To make the phone as affordable as possible, the company had worked out an exclusive arrangement with AT&T so that the cost of the phone would be underwritten by the cell phone carrier. In addition, building on the concept of the iTunes store, Apple came out with an “App Store,” allowing users to download software directly onto their iPhones. As always, Apple promoted the iPhone experience as totally user friendly and unobtrusive, a definite part of the Apple marketing strategy. Apple products are aesthetically attractive and seem to have an edge that other products simply do not. This provides a customer experience created to garner customer loyalty and to build trust. Apple’s marketing strategy has always been to allow developers to profit from their creations, while retaining ultimate control in order to maintain the Apple standards. There is no reason to assume that the company can equally embrace the BYOD concept.

BACKGROUND

Growth of Mobile Devices and Apple Products

The margins for the iPods and iPhones are supposed to be in the region of 40% to 50%. However, the iPad pricing strategy was reversed. Apple has priced the iPad much below prices of competitive tablets and no other manufacturer has so far been

able to reach these prices. This means that Apple is following a penetration strategy for the iPad in contrast with the iPods and iPhones. Therefore, it is a fair question to ask how Apple can have a product that can be priced much lower than the competition. There are multiple reasons for the low price, but one of the most significant ones is relationships with components suppliers. Relationships, contracts, and future deals with major component suppliers in the technology business matter a lot. Component suppliers of high quality are few, leading to a limited manufacturing capacity for the OEMs in the marketplace. Therefore, the OEM who books the manufacturing capacity of these suppliers in advance will create shortages and barriers for other OEMs. Furthermore, the target market for technology devices is characterized by innovation, excitement and dynamism, which means that any supply disruption creating shortage of products in the retail market will have extremely adverse effects. Customers are excited by the prospect of using and owning high tech products and unavailability at any point in time can put them off.

Part of Apple’s success surely has to do with their widely publicized product launches, and good example of this is the iPad product launch. Apple launched its very first iPad campaign during the Oscars a few years ago. The campaign was centered on one key video asset: Meet iPad. Besides being the first ad campaign for the iPad, the campaign was special for another reason as it marked a shift in Apple’s video advertising strategy. Prior to this iPad spot, Apple had been trying to drive traffic to Apple.com. Meet iPad was the first ad campaign that Apple actively pushed in video by publishing it to YouTube. The campaign was a hit. It drove over 2.5 million views in a week and inspired over 100 video placements uploaded by audiences across the web, all which contributed to views for the campaign. This pair of videos, the typical 30-second TV ad and a long form featurette, has become standard practice for Apple with its product announcements. Both go onto Apple’s YouTube page almost immediately

after the product event. Apple has made it standard practice to run paid media on YouTube for the new assets. The paid media component of these video campaigns makes them predictable. This strategy of launching an ad alongside a long form featurette goes back to the iPad 2 launch, which drove over half the views for its introducing iPad 2 campaigns from its featurette. iPhone 4 and iPhone 4S took this approach. The featurettes for both of these iPhone launches have driven well over 10 million views.

The iPad represented an evolution for Apple from their specialization in personal computers to their specialization in hand-held computers. Its release was preceded by the iPhone, which revolutionized the market of mobile computing with its innovative touch screen system. During 2007, the year the iPhone was released, Apple CEO Steve Jobs had announced that Apple would no longer go by the moniker Apple Computers, and instead would be called Apple, Inc. This was representative of the shift to mobile devices that Jobs envisioned for Apple. Rumors about the iPad began circulating not long after the iPhone had hit the shelves and experienced tremendous success. It would later be admitted by Jobs that the iPad actually was developed before the iPhone, but work on the iPad was stopped when he realized all of the capabilities of the tablet could be packaged into a phone. It seems very possible that Apple leaks of information contributed to the rumors swirling around the iPad. This is part of the genius of Apple's marketing ploys surrounding the iPad product launch. They developed consumer interest through rumors and speculation long before the iPad was turned out to store shelves. By the time the iPad product launch occurred, it almost seemed anticlimactic. People had been discussing the tablet for what seemed an interminable amount of time, and Apple fans were waiting for the iPad like anxious kids on Christmas morning. The iPad product launch occurred on April 3, 2010. Pre-orders had begun mid-March, and even though Apple initially only offered the iPad at The Apple Store, they sold 300,000 iPads on the first day of

the product launch. A month later, they had sold 1,000,000. By the time of the product launch for the second generation of iPads, Apple had sold 15 million of the first generation. Since then, Apple has showcased more generations of the iPad, each of which experienced similar success to the first version. The device is now fully Wi-Fi and 3G capable, and has garnered world-wide success ("What you can learn from" 2012).

It is interesting to note that they remove old ad campaigns from the Net so that you can only view the newest ones. In fact, if one goes to YouTube to search for these original assets for Apple's past products, not much will be typically found. The company may have recognized the value of launching video campaigns for new products, but this doesn't mean that they've completely bought into a world where people are in control of their media and are free to choose to watch essentially whatever they want. One of the benefits of online video for advertisers is that videos and campaigns never die. Even when paid media dries up, the video assets live on. When advertisers launch new campaigns, the old assets typically get a bump in views from being reactivated. Unless the promoters kill them, and Apple has done just that. With every new product launch, Apple is effectively killing its old campaigns, making the assets private on YouTube. At a glance, this makes sense: you don't want people getting distracted from a new product with new messaging with an old product and old messaging.

Thus, when it come it comes to product launches and driving anticipation, Apple has learned from their past mistakes and modified successful product launches to ensure even greater launches in the future. In both promotions, Apple leaked info to create interest and curiosity, and then followed up with the formal introduction on TV and the net. Both campaigns appeal to an audience that utilizes the phone and the pad for business and pleasure. Apple is consistent with the video ad approach and the TV ads although how they presented each and the target market is slightly different. Both product launches for the

iPhone and iPad have been met with eager anticipation and impressive sales figures. However, Apple has managed to uniquely craft each product launch to best penetrate the market considering the circumstances in front of each new launch. The challenges associated with BYOD penetration and its potential dependence on Apple products have similar characteristics to the various product launches of Apple.

CASE STUDY

Intel and its Application of BYOD

As previously discussed, BYOD is when an individual brings their personally owned device into the workplace or classroom. These devices include iPhones, iPads, tablets and laptops. BYOD first started when Intel recognized an increase in employee devices being brought to work and connected to the corporate network. As noted by Compeau, Haggerty, and Chandrasekhar (2013), M. Harkins, the chief information officer at Intel wanted to extract value from the initiative and turn BYOD into a new source of competitive advantage. He forecasted the number of employee devices being brought into the workplace to triple in just one year and that more than half of his workforce would use their own device to do more than half of their job.

The trend of BYOD was causing problems with the IT professionals because they were concerned with information security within the company. An organization's immediate concern was protecting the outside devices from carrying out company information. The concerns of financial and law related issues are important, as employees were purchasing their own devices and using them to do work on the road and at the office. The practice was reducing the cost of Intel's device procurement, but increasing the cost of configuring and supporting the growing number of devices. The data of the company were extremely vulnerable and every employee hired by Intel after would

be subject to bringing their own device into the workplace. By 2009, management at Intel realized that smartphones were had extreme levels of functionality and could connected to the data center through the cloud. As noted by Compeau, et al., 2013, Harkins gathered input from employees who were bringing their own devices as well as from employees who were not and 100% of employees viewed that the company should be managing the security of personal devices, allowing employees the freedom to bring their own devices to work. Hence, it was decided that management would control the devices as long as the employees accepted the training that went along with being able to bring their devices to work and other BYOD-based initiatives.

Historically, the company had a relatively small percentage of computer notebooks were lost per year that were either misplaced or stolen. However, under the recently adopted BYOD initiative, this inventory shrinkage of computer hardware was no longer a problem. Since the company no longer needed to purchase computer hardware, the BYOD policy promoted employees to become more conscious over where their devices were because of their ownership. It was found that the BYOD policy was increasing productivity and it was lessening the costs of data plans.

Probably, management did not predict that a BYOD policy was an issue of technology, but more about the underlying parts of the corporation. Management was having trouble defining a workable BYOD policy because of the various legal, human resources, and accounting issues. There were much effort to define the various details about software licensing and compliance with the BYOD policy. It was found that such policies could not cover the devices as a whole so it could only be used as a guideline or framework for the company.

The policy was set up in a level system so that each item under the umbrella of the BYOD policy would be covered. As noted by Compeau, et al., 2013, most companies have found that multi-tier IT architecture provides the most security as well as

very profitable investments. In general, companies need to protect against direct cyberattacks and Advanced Persistent Threats (APT) when developing their IS support systems. The degree of security and ROI are a function of level of access to data and services, with associated degrees of IS commitments of resources. The lowest level of security and IS commitments would be those data that are publicly available (i.e., real-time stock price movements, general corporate information available on public list serves and computer financial databases). The next level of increasing security and IS commitments would be more personal and confidential applications (i.e., payroll, health insurance, human resources information, to name a few) that directly involved personally identifiable information (PII). At this level, management must actively factor the issues of privacy, even though the device was owned by the employee. The next level security and IS commitments are basic corporate data used in decision-making and is the permissive level of access to corporate data (i.e., calendar of events/meetings, contacts, e-mails). The next level of security and IS commitments is more intermediate and more specific to the applications that employee would need to completed their assigned tasks (i.e., applications pertaining to specific lines of business). The most exclusive level of security and IS commitments (referred to as the Managed Equivalent) would have the most permissive level of access and would be the most traumatic to the company's operations if a security breach would occur.

For a short-term, as previously outlined, Intel only wanted the activity of the business to be understood by a few employees in the organization. With improved CoIT (Consumerization of IT), the company's engagement with customers, vendors, and partners improved drastically. It was easier for stakeholders to do business with the company and it was easier for new customers to sign up for the company's offerings. The two main security risks in Intel were device and data. The security risk with devices were new to

Intel because before BYOD they could control how many devices were company property. Now, the employee was responsible for his or her own device. The hardware operated by Intel was made with built-in information security features including security settings, log-on procedures, authentication protocols, access controls, firewalls, and anti-malware software. The new BYOD-based policy would compromise the devices and trying to secure both unmanaged and managed devices could potentially wipe out the entire encryption.

Ultimately, global IDs (such as Google ID and Live ID) were becoming popular and employees were trying to have multiple global ID's on the same devices they did their work. Integrating the ID and the corporate directory was creating security risks and the passwords protecting Intel's cloud was not strong enough to protect the secret corporate data. Since sharing corporate data will continue to be a major strategic issue, managers will still be reluctant in many cases to allow employees to place corporate data onto their own device and cross tied such confidential data with the employees' home computer. Such easily downloadable data files may be vulnerable to employees' family's devices and could be unknowingly shred with cybercriminals and competitors. Therefore, country-wide security issues have become increasingly important issues as managers typically are concerned with data protection laws abroad, or their lack of. In general, domestic companies have legal requirements, under the U.S. Federal Rules of Civil Procedures (FRCP), to formally follow demands from the courts of law for documents in the event of litigation, as evident with the recent Apple's case involving the FBI's investigation concerning information on a company owned device in the hands of a employee whom was also a terrorist suspect. Hence, the Apple and the U.S. Justice Department latest positions of world's largest company against the federal government on the extraction of data from an iPhone used by a gunman in the San Bernardino, CA (Benner & Lichtblumarch, 2016) certainly highlights

the difficulty that could arise from any BYOD policy and its unintended policies associated with e-Discovery.

The challenges in e-Discovery are several, as briefly outline in Apple's case with the federal government, as the matter of ownership is extremely important. The actual iPhone was owned by the company where the suspected gunman was employed and has since giving their consent for the data to be recover for homeland service purposes. If the company does not own the devices and individual employees do, therefore personal ownership precludes intervention by IT management without consent. Of course, in the case with Apple, that was not the case. In the case of Intel, with BYOD devices used by 10,000+ employees is not the similar case, it does underscore the complexities of BYOD policies. Different categories characterize such complexities [e.g., data (corporate data verses non-corporate data; device-created data, versus application-created data; user created versus data-making retrieval)] are important issues that need strict oversight to make BYOD policies work. There is the important issue of location of actual data (e.g., located on corporate network, a server, a cloud, a telecom carrier, or the employee-owned device) is critical as well when responding to an e-Discovery request.

With all of the benefits and challenges, many companies need to be very careful in any strategic initiative designed to integrate BYOD within the workforce environment. Considerations include financial planning (e.g., purchasing devices to support the loads of unmanaged devices), security (e.g., to protect the company data and response in a timely fashion to legitimate e-discovery requests) all require requiring employees to sign appropriate disclosure forms. The company still has to manage the devices brought into their offices on a daily basis. With hiring new employees and new devices being upgraded more often, the IT department is on its toes when it comes to securing the data and managing all the devices. It is a constant task to keep up with and Intel will always be a step ahead.

Institutions of Higher Learning and applications of BYOD



With technology growing and textbooks becoming more expensive, technology in the classroom is becoming more popular. Teachers are using PowerPoint presentations to present on topics in class to students and many students are using tablets and computers to do simulation exercises on the topics. Learning is becoming more and more interactive through available and shared technology. When discussing the topic of BYOD and its role in innovative teaching strategies, it is critical for a business or school to look at the various advantages and disadvantages of implementing such a policy. The obvious benefits that employers or school boards have initially attractive are the cost savings of implementing a BYOD policy. BYOD can improve the user's retention rate of information. Listening to a professor lecture about supply chain management may be a difficult proposition if only theory was represented. However, if a podcast was followed by active links to videos and current topics by Internet news services, perhaps student will be more likely to retain and understand the information due to visual aids. Each student learns differently as some students learn through listening and others learn through doing. Showing a video or doing an activity can only increase the knowledge retained, especially if each student follow an adopted BYOD policy sponsored by their institution.

Students have the potential to substantial savings on textbooks if schools implemented a BYOD policy. They could buy the less expensive versions of the textbook online via eBooks and bring a laptop and/or tablet to class. This policy may improve the student's health because they do not have to carry heavy books and backpacks that would could result in spinal or back disorders. Due to the convenience of a mobile device, as Apple products previously discussed, it may enhance learning as tablets and smartphones are easily available and do not require elaborate or heavy storage. Hence, administration at these BYOD-

Table 1. Worldwide device shipments by segment, as of December 2013 (partially adapted from “Gartner says worldwide ...” 2014)

Device Type	2013	2014	2015
PC (Desk-Based & Notebook)	299,342	277,939	268,491
Tablet (Ultramobile)	179,531	263,450	324,565
Mobile Phone	1,804,334	1,893,425	1,964,788
Other Ultramobiles (Hybrid & Clamshell)	17,195	39,636	63,835
Total	2,300,402	2,474,451	2,621,678

Note: Numbers are in 1000s of units.

sponsored schools could create their own rules and regulations regarding the students bringing their own devices. The policy would state what the devices should be used for and if seen using them in any other way than for learning purposes, the student could be asked to put it away and that could be the punishment.

An obvious challenge to the BYOD movement is the potential high risk of security breaches (e.g., APT and direct cyberattacks). Corporate and educational organizations' operated Wi-Fi networks are usually secure and can only be accessed through proper login information. This situation usually creates little problems, since each worker/student and supervisor/teacher would need their own logon credentials in order to access the secure network. In particular, it is hoped that by securing the network, it will allow for businesses/schools to block certain websites so the employees/students cannot be distracted by gaming websites or social media. It is important to review the level of the business that needs to be done within the institution that way you can see if the policy works for its various stakeholders. Although it is not the stated purpose of this chapter to extend discussion from a more technical perspective on how to solve this problem, it is a major issue to be addressed. For example, one of the reviewers of this chapter stated “I see a big problems that our students at sport faculties are obsessed with mobile sport applications. They are analyzing sport activities of each other during the lectures and comment on their activities, etc.” Rauter, Doupona, and Fisher

(2000) suggested that mobile devices, coupled with the popular of smart phones and watches, allow athletes to train smarter. However, Al-Shuaibi, Subramaniam, and Mohd (2014) acknowledged that mobile applications, including sports, are becoming a major distraction to students/employees as part of an electronic distraction (commonly termed cyberdeviance) that may lead to losses in productivity. According to a recent study, the worldwide tablet market is forecasted to grow 47% in 2014, in part due to lower average selling prices attracting new users (“Gartner says worldwide ...” 2014). Ownership of mobile devices are rapidly expanding, as evident from the statistics displayed in Table 1. Management and education must take steps to properly leverage these developments in a positive, non-distractive direction.

FUTURE RESEARCH DIRECTIONS

The global workforce needs to that envision being part of the global economy, other than just being able to identify themselves as a small component in a social network map or a box on a bulky organizational chart. There needs to be a renewed sense of shared context, at least among team members and with the business itself, if managers are to stem the growth of “virtual distance.” As noted by Sobel-Lojeski (2015), virtual distance is a sense of psychological and emotional detachment that begins to grow little by little and unconsciously when most encounters and experiences are medi-

ated by screens on smart devices. It is frequently assumed that physical separation and/or time zone gaps are the root causes of virtual distance. Although physical distance between employees may add to the problem, there may be more subtle circumstances.

According to Sobel-Lojeski (2015), virtual distance model is made up of physical distance, operational distance, and affinity distance. Physical distance deals with geographic distance, while operational distance develops when there is a general lack of shared context (resulting in white or backboard noise like miscommunications or technical problems). Lastly, with affinity distance, it becomes more subtle divisions in the workplace that can prevent deeper relationships from occurring. These disruptions in the human moments that result in avoiding richer, longer-lasting relationships or meaningful mutuality may prevent the proper motivation within the work environment that would provide happier, more productive workers. It is the hope by the authors of this chapter that promoting appropriate BYOD-based policies should reduce virtual distance by allowing students/workers to have greater flexibility and voice in their learning/working environments.

CONCLUSION

BYOD is a phenomenon that is spreading across companies and education institutions in the global economy. BYOD can create valuable strategic opportunities and creation. If done properly, BYOD-based policies can operate on a secure data infrastructure and help motivate employees as well as students by reducing virtual distance and promoting more autonomous learning/working environments. Ultimately, such policies, it is hoped, would allow for individuals to work more independently and innovatively.

REFERENCES

- Al-Shuaibi, A., & Subramaniam, C., & Mohd, S. F. (2014). The Mediating Influence of Job Satisfaction on the Relationship between HR Practices and Cyberdeviance. *Journal of Marketing and Management*, 5(1), 105–119.
- Belmans, W., Mobley, B., Page, R., Uy, J., & Lambrette, U. (2012). *Enterprise Journey to the Clouds*. Cisco Systems White Paper. Retrieved March 11, 2016 from http://www.cisco.com/c/dam/en_us/about/ac79/docs/EJ-to-Clouds.pdf
- Benner, K., & Lichtblaumarch, E. (2016). Apple and Justice Dept. trade barbs in iPhone privacy case. *New York Times*. Retrieved March 16, 2016 from http://www.nytimes.com/2016/03/16/technology/apple-court-filing-iphone-case.html?_r=0
- Bruner, G. C. II, & Kumar, A. (2005). Explaining consumer acceptance of handheld internet devices. *Journal of Business Research*, 58(5), 553–558. doi:10.1016/j.jbusres.2003.08.002
- Burton-Jones, A., & Hubona, G. S. (2005). Individual differences and usage behavior: Revisiting a technology acceptance model assumption. *The Data Base for Advances in Information Systems*, 36(2), 58–77. doi:10.1145/1066149.1066155
- Chau, P. Y. K. (1996). An empirical assessment of a modified technology acceptance model. *Journal of Management Information Systems*, 13(2), 185–204. doi:10.1080/07421222.1996.11518128
- Chen, L., Gillenson, M. L., & Sherrell, D. L. (2002). Enticing online consumers: An extended technology acceptance perspective. *Information & Management*, 39(8), 705–719. doi:10.1016/S0378-7206(01)00127-6
- Compeau, J., Haggerty, N. R. D., & Chandrasekhar, R. (2013). Intel Corp. - Bring Your Own Device. *Harvard Business Review*. Retrieved January 1, 2016, from <https://hbr.org/product/intel-corp--bring-your-own-device/W13035-PDF-ENG>

Friedman, J., & Hoffman, D. (2008). Protecting data on mobile devices: A taxonomy of security threats to mobile computing and review of applicable defenses. *Information, Knowledge, Systems Management*, 7(1-2), 159–180.

Gartner says worldwide traditional PC, tablet, ultramobile and mobile phone shipments on pace to grow 7.6 percent in 2014. (2014). Retrieved April 27, 2016 from <http://www.gartner.com/newsroom/id/2645115>

Qrunfleh, S., Tarafdar, M., & Ragu-Nathan, T. (2012). Examining alignment between supplier management practices and information systems strategy. *Benchmarking: An International Journal*, 19(4), 604–617. doi:10.1108/14635771211258034

Rauter, S., Diupona Topic, M., & Fister, I., Jr. (2000). *Mobile sport applications can make our cycling more sociable*. Retrieved April 27, 2016 from <http://www.iztok-jr-fister.eu/static/publications/71.pdf>

Sinha, K. K., & Thirumalai, S. (2011). Customization of the online purchase process in electronic retailing and customer satisfaction: An online field study. *Journal of Operations Management*, 29(5), 477–487. doi:10.1016/j.jom.2010.11.009

Smith, A. A., Synowka, D. P., Clark, J. S., & Smith, A. D. (2012). Exploring the degree of commercialization associated with sports marketing activities. *International Journal of Services and Operations Management*, 11(2), 170–200. doi:10.1504/IJSOM.2012.045198

Smith, A. D. (2012). E-personalization and its tactical and beneficial relationships to e-tailing. *International Journal of Information Systems in the Service Sector*, 4(2), 48–71. doi:10.4018/jisss.2012040104

Smock, A. D., Ellison, N. B., Lampe, C., & Wohn, D. Y. (2011). Facebook as a toolkit: A uses and gratification approach to unbundling feature use. *Computers in Human Behavior*, 27(6), 2322–2329. doi:10.1016/j.chb.2011.07.011

Sobel-Lojeski, K. (2015). The subtle ways our screens are pushing us apart. *Harvard Business Review*. Retrieved March 12, 2016 from <https://hbr.org/2015/04/the-subtle-ways-our-screens-are-pushing-us-apart>

ADDITIONAL READING

Haried, P., & Ramamurthy, K. K. (2009). Evaluating the success in international sourcing of information technology projects: The need for a relational client-vendor approach. *Project Management Journal*, 40(3), 56–71. doi:10.1002/pmj.20113

Kapur, P. K., Gupta, A., Jha, P. C., & Goyal, S. K. (2010). Software quality assurance using software reliability growth modelling: State of the art. *International Journal of Business Information Systems*, 6(4), 463–496. doi:10.1504/IJBIS.2010.035742

Keramati, A., & Behmanesh, I. (2010). Assessing the impact of information technology on firm performance using canonical correlation analysis. *International Journal of Business Information Systems*, 6(4), 497–513. doi:10.1504/IJBIS.2010.035743

Smith, A. D. (2016). Search engine marketing and customer relationship management within a social media environment. *International Journal of Business Information Systems*, 22(2), 181–210. doi:10.1504/IJBIS.2016.076246

KEY TERMS AND DEFINITIONS

Advanced Persistent Threats (APT): Advanced Persistent Threats typically employs a number of phases and techniques to break into a network, avoid detection, and gather customers' or companies' information over an extended time period. Unfortunately, such well-planned and extensive strategy is difficult to detect, as it many

techniques and phases may be extremely complex and expensive to develop a counteroffensive that is successful.

BYOD: Types of AIDC-related technologies to leave the human element out of the data collection and storage functions of information derived from manufacturing, integrated through the manufacturing process, types of authentication concerns and/or e-security strategies, and relationship links to customer profiles. Typical types of AIDC include, bar-coding, RFID, magnetic strips, touch memory, and smart cards.

CoIT (Consumerization of IT): CoIT refers to the strategic use of IT systems in order for a company to enhance its engagement with customers, vendors, and partners. Mutually beneficial relationships are the targets, with enhanced customer service and firms' increased profitability.

Cyberdeviance: Cyberdeviance refers to an employee's use of the company's Internet access during work hours for non-work related purposes, and is a subset of counterproductive work behavior. Counterproductive work behavior refers to any activity of an employee with the intended or unintended consequence of harm to the organization.

Direct Cyberattack: Direct cyberattack causes data communications interception at the hands of a specific hacker who may have located a specific person or Wi-Fi area in order to perform such cybercrime. A type of direct attack is direct access, where the criminal gains physical access to either the computer or its network in order to compromise security by loading worms, Trojans, compromising data, etc.

e-Discovery: This term refers to all types of electronic information storage (e.g., e-mails, directories, documents, searches, etc.) that are either targets investigations in which the affected organization need to produce evidence of such transactions. In other cases, information have been mistakenly lost or deleted and managers or employees need to retrieve message-level data.

Global IDs: This term refers to identifications tied to software associated with social media platforms (e.g., Google ID and Live ID) as well as to more company-based application-wide URI that uniquely identifies a particular model. Global IDs as referred to in this chapter refers to simple identifier to reference a variety of classes of objects (i.e., payroll, operations planning, job scheduling). Global IDs typically reference a model object so that it can more easily locate a model and/or application at the time it needed to perform the specific task by employees/managers.

Personally Identifiable Information (PII): Personally identifiable information (PII) refers to all information, including personal financial and healthcare information that can be traced to an individual customer or patient.

Virtual Distance: Virtual distance is a psychological and emotional sense of detachment that may result from lack of human interactions or from the loss of the human moment. Students and employees that share cultural values, similarities in communication style, and attitudes toward work can directly mediate the effects of office isolation.

Cell Phone Conversation and Relative Crash Risk Update

Richard A. Young

Driving Safety Consulting, LLC, USA

INTRODUCTION

A few brief definitions of basic terms will facilitate understanding the research updates reviewed in this chapter. This chapter uses the same definitions of terms as in the corresponding article in Chapter 102 in the previous edition of this Encyclopedia (Young, 2015a), except for the following additions and enhancements.

- **Confidence Interval:** In this paper, the 95% Confidence Interval (abbreviated CI). “If the underlying statistical model is correct and there is no bias, a CI derived from a valid analysis will, over unlimited repetitions of the study, contain the true RR with a frequency no less than 95% of the time” (Porta, 2008, p. 49).
- **Demand Terms:** The metrics characterizing driver performance during secondary tasks can be grouped into the two orthogonal dimensions of *Physical* and *Cognitive* demand (Young, 2016a, 2016b). *Physical* demand concerns the aspects of secondary tasks which place demands on the eye glances, hands, or feet of the driver. *Cognitive* demand concerns the aspects of secondary tasks that place demands upon the attentional resources of the driver.
- **Driver Distraction:** “Driver distraction is the diversion of attention away from activities critical for safe driving toward a competing activity, which may result in insufficient or no attention to activities critical for safe driving” (Regan et al., 2011, p. 1776). As pointed out by Young (2015a, Section

6), a secondary task with a relative risk confidence interval that encompasses one, or that is entirely below one, is not a driver distraction according to this definition, because the secondary task does not impair activities critical for safe driving.

- **Epidemiological Terms:** *Case window.* In naturalistic driving studies, a short time period (often 6 seconds) near the time of the precipitating event, which in turn immediately precedes the time of the crash. *Control window.* A time period with the same duration as the case window, but during driving on some day before the crash, when there was no safety-related incident. *Homogeneous.* Assume the population under study is divided into two or more categories or strata (e.g., defined by exposure and confounder levels). The homogeneous assumption is that within each analysis subgroup, “the probability (risk) of an outcome event arising within a unit of person-time is identical for all person-time units in the stratum” (Rothman et al., 2008, pp. 239-240). That is, the effect is constant or uniform across strata. If so, the strata can be pooled or combined (e.g., crashes and near-crashes can be combined into one group). If the effect is not equal across strata, then epidemiologists say that the effect measure is *heterogeneous*, meaning that it is modified or varies across strata (Rothman et al., 2008, p. 63). Strata that are heterogeneous cannot properly be pooled to create a single estimate. Standard tests for homogeneity exist in epidemiol-

DOI: 10.4018/978-1-5225-2255-3.ch521

ogy and should always be used before combining strata. *Confirmation bias* is “a form of bias that may occur when evidence that supports one’s preconception is evaluated differently from evidence that challenges those conceptions” (Porta, 2008, p. 49). *Selection bias* refers to “a distortion in the estimate of the effect due to the manner in which subjects are selected for the study” (Porta, 2008, p. 225). An example of a potential reason for selection bias is if all drivers with a safety-critical event are chosen for the exposed group, and only at-fault drivers with a safety-critical event are chosen for the unexposed group (Young, 2013a). Another example of selection bias is if drivers engaged in a cell phone conversation accompanied by other secondary tasks are chosen for the exposed group, and drivers without any secondary tasks are chosen for the unexposed group (Young, 2017a).

- **Naturalistic Driving:** An example of non-experimental driving, as is real-world driving. Vehicles are specially equipped with video cameras that record the driver’s behavior, and other instruments such as inertial sensors that record the vehicle’s behavior. These measurements occur in real time, while the vehicles are driven in everyday fashion over a prolonged period, from months to several years. A naturalistic driving study (NDS) also allows for exact timing of crashes and calls, at least those using hand-held phones. Many naturalistic driving studies do not have audio recordings, only video, so determining whether a driver is engaging in a hands-free conversation, or just singing or talking to themselves, is difficult. Hence, only hand-held cell phone conversations are evaluated in some naturalistic driving studies.
- **Real-World Driving:** Another example of non-experimental driving, as is naturalistic driving. Real-world driving refers to driv-

ing a vehicle in an everyday manner, without experimental instructions or special instrumentation. In real-world driving, tasks such as engaging in a cell phone conversation that are secondary to primary driving, if performed at all, are performed at times and under traffic and environmental conditions chosen by the driver, and no special equipment beyond that installed at the time of purchase is attached to the vehicle. Young (2001) and Young and Schreiner (2009) are examples of real-world driving studies that relied upon GPS, crash sensors, and cell phone records to determine the exact times and locations of all crashes severe enough to deploy an airbag, along with all communications using an embedded wireless device (OnStar). Studies based instead on police crash reports or crash databases are also real-world studies, but they typically have inexact crash times, making the association with cell phone conversation or other secondary tasks difficult to establish with accuracy (Young, 2014a, Appendix A).

- **Relative Risk (RR) Estimate Terms:** A *crude* RR estimate refers to a measure in which the effects of differences in composition of the populations being compared (e.g., differences in age or sex distributions) have not been minimized by statistical methods. An *adjusted* RR estimate refers to a measure in which the effects of differences in composition of the populations being compared have been minimized by statistical methods (Porta, 2008, p. 4). A *corrected* RR estimate refers to a measure that has been corrected for errors made in the analysis method, or in the underlying data. *Odds Ratio* (OR) is used in some epidemiological studies, such as a case-control study, to estimate the Relative Risk (RR). The OR estimate closely approximates the RR if the event of interest is rare (such as a crash).

- **Talk Terms:** *Talk* is used throughout this chapter as a synonym for “cell phone conversation,” meaning talking or listening on a wireless cellular device while driving. Talk Alone refers to Talk as the sole task performed in a case or control window, without another secondary task being performed in that same window. Pure Talk refers to Talk Alone without any confounding effects from driver behavior errors or driver impairments. Talk in any of these types can take place via a hand-held portable phone, a hands-free portable phone, or a hands-free wireless device embedded in the vehicle (such as the OnStar device).
- **Task Terms:** *Primary tasks* are defined as the operational tasks of driving *per se*: namely, steering, pressing and releasing the accelerator, braking, and detecting and responding with an appropriate steering or braking maneuver to objects and events in the roadway. In vehicles with manual transmissions, primary tasks would also include pressing and releasing the clutch pedal and operating the gearshift lever. *Secondary tasks* are defined as tasks performed in a vehicle by a driver that are not related to the primary driving tasks.

RECENT SCIENTIFIC KNOWLEDGE ABOUT TALK RELATIVE CRASH RISK

Study A: 100-Car NDS Talk RR Estimates

100-Car corrected crude Talk RR estimates. Klauer et al. (2006) estimated a crude Talk RR of 1.29 (CI 0.93-1.90) for the 100-Car NDS. However, their analysis method had selection bias, because all drivers with a safety-critical event were chosen for the exposed group, and only at-fault drivers with a safety-critical event were chosen for the unexposed group (Young, 2013a).

Young (2013b, 2014a) re-analyzed the original 100-Car data and corrected this selection bias by including all exposed and unexposed crash/near-crash cases (regardless of fault). The Klauer et al. (2006) crude Talk RR estimate of 1.29 (CI 0.93-1.90) was thereby reduced to 0.78 (CI 0.56-1.06). Klauer et al. (2014) later used another correct method of eliminating selection bias by including only at-fault exposed and unexposed groups, giving rise to a crude Talk RR estimate of 0.76 (CI 0.51-1.13) for experienced adult drivers in the 100-Car crash/near-crash data. The corrected crash/near-crash crude Talk RR estimate (whether 0.76 or 0.78) for the 100-Car study data is still potentially biased upward by environmental variables (e.g., traffic, closeness to junction, weather, and time of day), and demographic variables. These variables were unmatched, or only partially matched, between case and control video clips in Klauer et al. (2006, 2014).

100-Car Talk RR estimates adjusted for driver and environmental conditions. Young (2013b, 2014a) reanalyzed the 100-Car case data (VTI, 2014), the same data that had been used by Klauer et al. (2006). However, Young used the baseline data from Klauer et al. (2010) that had been matched with the environmental and driver variables to each crash/near-crash case, rather than the unmatched baseline data that had been used by Klauer et al. (2006). Young (2013b, Table 9; Young, 2014a, Table 10) found that the Talk RR estimate declined by about 23%, from 0.78 to 0.56 (CI 0.41-0.76) (Study A) by using matched baseline data. The matched baseline data removes bias resulting from the driver variables because each participant acts as his or her own control. Matching by environmental conditions also reduces bias resulting from differences between case and control video clips in variables such as traffic density, closeness to junction, weather, and time of day.

Summary: The crude hand-held Talk RR estimate from the 100-Car study, when corrected for selection bias in the analysis methods, is below one, with a confidence interval that overlaps

one. The hand-held Talk RR, when adjusted for demographic and environmental variables using matched baseline data, falls to 0.56 (CI 0.41-0.76), with a confidence interval entirely below one, indicating a protective effect.

Study B: Fitch et al. (2013) NDS Talk RR Estimates

Fitch et al. (2013) performed a matched case-control study of all *safety-critical* events, defined as crashes (including curb strikes), near-crashes, and crash-relevant conflicts. This study estimated an Odds Ratio (OR) near one for integrated hands-free Talk; i.e., 0.71 (CI 0.30-1.66). They used a method suggested by Young (personal communication) of using the driving period before the crash or near-crash in the same trip. This method matches and therefore automatically adjusts for the driver (including demographics, genetics, personality, etc.), the approximate time of day, weather, and road type. Fitch et al.'s Talk OR estimates were all near one: integrated hands-free Talk (0.71, CI 0.30-1.66), portable hand-held Talk (0.79, CI 0.43-1.44), and portable hands-free Talk (0.73, CI 0.36-1.47). These three Talk ORs for the different wireless device types were homogeneous, so they could be pooled into one estimate for Talk, regardless of device, giving rise to a tighter confidence range. The pooled RR was 0.75 (CI 0.49-1.15) (Study B).

Summary: The Fitch et al. (2013) Talk RR estimates are all below one, with a confidence interval that overlaps one, for hands-free, hand-held, and embedded cellular phones.

Studies C-E: Real-World Talk RR Estimates

Young and Schreiner's (2009) real-world cohort study of three million personal conversations on a hands-free integrated wireless device observed only 14 crashes during Talk out of 2,316 airbag-deployment crashes. The study estimated a crude

Talk RR of 0.62 (CI 0.37-1.05) (Study C), when compared to baseline driving with no Talk.

The Redelmeier and Tibshirani (1997) and McEvoy et al. (2005) real-world case-crossover studies claimed adjusted Talk RR estimates near 4.0 for crashes severe enough to cause injury. However, these studies contained a major bias by assuming that driving took place during the entire 10-minute control windows on previous days (Young, 2012a, 2013b). Based on several large GPS studies, Young (2012a, 2013b) estimated that study participants in the case-crossover studies were likely in the vehicle during only 20% of the control window periods in which they recalled driving; that is, they drove only part-time during a given control window. The number of calls made during any given time interval is lower when not in the vehicle than in the vehicle (Young, 2014a). Hence the number of calls made during the control windows was biased low, artificially increasing the estimated RRs. Correcting this bias (Young, 2014a), and also misclassification bias from counting calls made after the crash, the Redelmeier and Tibshirani (1997) Talk RR estimate reduced to 0.61 (CI 0.38-0.98) (Study D), and the McEvoy et al. (2005) estimate reduced to 0.64 (CI 0.32-1.27) (Study E), now consistent with the Young and Schreiner (2009) real-world cohort study estimate of 0.62 (CI 0.37-1.05) (Study C).

Claims that the Redelmeier and Tibshirani and McEvoy et al. case-crossover studies pooled dialling or other cell phone tasks with Talk, and that these other tasks were what elevated the RRs (rather than the bias from assuming that driving was full-time in the control windows), are invalid (Young, 2015a, Section 4; Young, 2014d), because dialling, searching for the phone to answer it, etc., do not appear in cell phone records if interrupted by a crash.

Summary: The Talk RR results of three real-world driving studies, after correction for known biases, are consistent with the naturalistic study Talk RR results in Studies A and B. Namely, the risk of crashing during a cell phone conversation

is not greater than the risk of crashing when not engaged in a cell phone conversation.

Study F: SHRP2 NDS Talk RR Estimate

Dingus et al. (2016) estimated the crude OR for hand-held Talk as 2.2 (CI 1.6-3.1), using the SHRP2 (2014) NDS dataset. This crude Talk OR estimate is high compared to those in previous naturalistic studies analyzed by the same Institute and many of the same authors: The Klauer et al. (2014) revised 100-Car crude Talk OR estimate of 0.76 (CI 0.51-1.13) (1.13), and the Fitch et al. (2013) pooled NDS dataset adjusted estimate of 0.75 (CI 0.49-1.15) (see Study B). The Dingus et al. (2016) estimate is also inconsistent with the crude 100-Car NDS Talk OR estimate of 0.56 (CI 0.41-0.76) by Young (2013b, 2014a) (see Study A).

To resolve these discrepancies, Young (2017a) directly analyzed the SHRP2 database using the Insight query program (<https://insight.shrp2nds.us/>). There were 35 Talk cases for all crashes, except curb strikes were excluded following Dingus et al. (2016). Of these, drivers in 19 cases were exposed to Talk and additional secondary tasks, and only 16 were exposed to Talk Alone. There were 240 crashes during “model driving” (where there were no secondary tasks) so the odds in favor of Talk Alone exposure among cases is 16/240 or 0.0667. There were 541 Talk Alone controls in the balanced-sample baseline without crashes. Compared to the 9,415 “model driving” controls, the odds in favor of exposure among non-cases is 541/9415 or 0.0575. The Talk Alone OR estimate is then 1.16 (CI 0.65-1.94). This Talk OR estimate further reduces to 0.92 (CI 0.29-2.2) (Study F) after eliminating confounding by driver behavior errors (Young, 2017a). This estimate is less than half the Dingus et al. (2016) OR estimate, with the confidence range now bracketing one, rather than being above one.

The upward bias in Dingus et al. arises mainly from the 19 cases that Dingus et al. included in

their exposed category, which contained additional secondary tasks in the same exposure case window as Talk. Fourteen of these cases had one additional secondary task, and five had two additional, for a total of 24 additional secondary tasks counted with Talk, even more than the 19 instances of Talk itself. These additional secondary tasks typically had high OR estimates (when analysed alone without Talk): e.g., “Cell phone, Dialing hand-held” (2.8), “Looking at an object external to vehicle” (8.7), or “Moving object in vehicle” (62.8). This result means that the Dingus Talk OR estimate was upwardly biased due to selection bias: 19 of the 35 drivers selected for the exposed group had additional secondary tasks besides Talk in their case window, but the drivers selected for the unexposed group had no secondary tasks. Removing this selection bias and the confounding bias from driver behavior errors, the corrected SHRP2 Talk OR estimate was 0.92 (CI 0.29-2.2). This SHRP2 Talk OR estimate is now consistent with the 100-Car study OR estimate of 0.56 (Study A), Fitch et al.’s (2013) matched case-control NDS pooled OR estimate of 0.75 (Study B), and the real-world study RRs reviewed in Studies C-E.

Summary: A major bias in the analysis method used by Dingus et al. (2016) arose because their exposed group contained up to 2 additional secondary tasks in the same 6-s video clips as Talk, while their unexposed group excluded any secondary task in its 6-s video clips. The SHRP2 NDS Talk OR estimate, after eliminating this and another bias, is consistent with other naturalistic and real-world studies in finding that the risk of crashing during a cell phone conversation is not greater than the risk of crashing when not engaged in a cell phone conversation, or when compared to just driving with no secondary tasks.

Confusion of Prevalence With Causality

Young (2015a, Section 9) noted a common misunderstanding in the driving safety literature; namely,

that the presence of an activity co-occurring with crashes indicates that the activity itself is a contributing factor to crash risk.

For example, Dingus et al. (2006, p. xli), based on video observations of crashes and near-crashes in the 100-Car naturalistic driving study, stated that “78 percent of the crashes had ... inattention ... as a contributing factor.” However, the 100-car data show that drivers were “inattentive” (as defined by Dingus) more than 78% of the time during baseline driving (Young, 2015c). That is, baseline driving had a higher prevalence of inattention than crashes. Hence, the Dingus et al. (2006) statement that inattention overall is a “contributing factor” risk for crashes/near-crashes is not supported by their data.

In other words, as noted by Young (2015a, Section 9), it is not valid to conclude based solely on prevalence that a distraction or other risk factor is *per se* detrimental to driving safety. To estimate whether a factor is causal, coincidental, or protective for driving safety, the prevalence (or odds, or risk) of that factor during a safety critical event must be compared with its prevalence (or odds, or risk) during control periods when there is no safety critical event.

Dingus and colleagues unfortunately carry forward this incorrect assumption about prevalence in the Dingus et al. (2016) article on the SHRP2 data. After presenting a table solely of prevalence values in their Figure 1, they state, “Other important findings from Fig. 1 show that nearly three-quarters of the crashes (i.e., 73.7%) involved some type of error; 68.3% of crashes involved some type of observable distraction; and 54.5% of crashes involved both. These findings conclusively show the detrimental impact of distraction alone and in combination with a variety of other sources of error and impairment.” Whether distraction alone or other sources of error are “detrimental” to driving safety can only be determined after a comparison to baseline driving, not simply by using prevalence numbers during crashes.

Summary: It is incorrect to make claims that an activity is a contributing factor for safety-critical

events, or detrimental to driving safety, based on prevalence alone. An appropriate comparison must be made to baseline driving before such claims can be made.

META-ANALYSIS TALK RR ESTIMATE

Table 1 and Figure 1 (end of this chapter) summarize the individual Talk RR estimates discussed in Studies A-F. The individual Talk RR estimates are homogeneous (p for homogeneity equals 0.886), so a pooled estimate can be made, which is 0.63 (CI 0.52-0.76). The pooled RR confidence interval is entirely below 1, indicating that Talk reduces relative crash risk compared to Not Talk. A likely explanation driver self-regulation as discussed in the following section.

Driver Self-Regulation

Strayer et al. (2013, 2014, 2015a, 2015b, 2015c), based on an analysis of data from their experimental studies, claim that auditory-vocal tasks (e.g. Talk, or speech interactions with in-vehicle infotainment systems) cause major driver “impairments” at “catastrophic” levels. However, Young (2014b, 2014c; 2015b) showed that the assumption by Strayer et al. (2013, 2015a) that auditory-vocal tasks (e.g. Talk, or speech interactions with in-vehicle infotainment systems) cause major driver “impairments” at “catastrophic” levels are not supported by their own data. The Strayer et al. (2013, 2015a) data, upon close examination, show instead that drivers increase their following headway to a lead vehicle during Talk, and the resultant increased headway more than offsets the small 50 to 200 ms increase in response time (RT) from the effect of Talk on orienting attention (Young, 2014b, Figure 6; 2015b). That is, drivers self-regulate their driving behavior during Talk to offset the risk increases associated with the slight increases in RT associated with the attentional effects of Talk. As a result, drivers’ overall crash

risk during Talk does not increase, and may even decrease during real-world driving (Young, 2013b; 2014a; 2014b, Figure 5; 2015b).

Specifically, Young (2014 b, 2014c; 2015b) compared the time and distance headways to the response time increases during all auditory-vocal tasks (not just Talk) in the Strayer et al. (2013, 2015a) data. He found complete compensation, in that the headway times and distances increased sufficiently such that the increases in safety margin fully offset the RT increases in all the auditory-vocal tasks in the Strayer et al. (2013, 2015a) study. The increased headways were not recognized as compensation by Strayer et al. (2013, 2015a), but were instead counted as evidence of increased impairment. However, for many tasks, the drivers increased their headways even more than necessary to offset any RT increases arising from the attentional effects of cognitive load during auditory-vocal tasks, so that there would actually be a net safety benefit to the extent that the experimental results are predictive of real-world results. These and other “overcompensations,” if substantiated in naturalistic driving studies, would help explain why the corrected and adjusted relative risks from the studies discussed in Studies A-F have ORs that are not greater than one, and the pooled relative risk estimate for Talk while driving 0.63 (CI 0.52-0.76) from the meta-analysis shown in Table 1 and Figure 1, indicating a protective effect (i.e., crash reduction).

As mentioned in Young (2015a), a useful analogy is that older drivers have slower RTs compared to younger drivers (on average, RT decreases with age). However, older drivers self-regulate by driving less often during nighttime, bad weather, and poor road-surface conditions, and by being less hurried, alcohol intoxicated, or distracted by non-driving activities (Hakamies-Blomqvist, 1994). Because of such self-regulation, older drivers fully compensate or even overcompensate for their increased RTs, and hence are the safest drivers of any age group in terms of crash rate per driver, after controlling for low mileage bias by comparing the crash rates of drivers of different

ages that were matched for yearly driving distance (Langford et al., 2006). The crash rate decreases continuously with age, even though the fatality rate for the oldest drivers approximates that of the youngest drivers, but that is because of frailty, not because of an increased crash rate.

Young (2015a) further noted that the amount of the RT increase from these attentional effects is known to depend upon the mode of the secondary task. For example, visual-manual tasks typically give rise to slightly greater RT increases (about 300 to 450 ms) than do auditory-vocal tasks (about 50 to 200 ms), compared to “just driving” with no secondary task (Angell et al., 2006; Young, 2014b, Appendix, Fig. 6). In part, this is because visual-manual tasks are typically accompanied by increased eyes-off-road time, which also increases RTs to visual events in the roadway, adding to the RT increases from the attentional effects of cognitive load associated with visual-manual tasks. However, the RT effects of the attentional component of visual-manual tasks can be measured independently of the eyes-off-road effects, because RT increases also occur when using tactile, rather than visual, Detection-Response Task (DRT) stimuli (ISO17488, 2016; Young et al., 2013d). Indeed, a two-dimensional model (with orthogonal physical demand and cognitive demand dimensions) is required to explain driver performance effects of secondary tasks (Young and Angell, 2003; Young, 2012b; 2016a, 2016b). For visual-manual secondary tasks, the physical demand component is dominant, but the cognitive demand component is non-trivial, and larger than for auditory-vocal secondary tasks, at least for the task sets examined in Young (2016b), and the Crash Avoidance Metrics Partnership task set in Angell et al. (2006) – see Young (2014b). For auditory-vocal tasks alone, without visual-manual tasks or mixed-mode tasks, the cognitive demand dimension is dominant (Young, 2016b).

Almost every experimental study of cell phones and driving behavior to date shows increases in time and distance headways associated with Talk. For example, such increased time and distance

headways during Talk are evident in eight simulator studies by Strayer and colleagues (for citations see Refs. 11-18 in Young, 2014b), as well as Brookhuis et al. (1991), Jamson et al. (2004), Ranney et al. (2004), Kircher et al. (2004), and Yannis et al. (2010).

Young (2013b, 2014a, 2014b, 2014c, 2015a) further explained why self-regulation is not always observed or reported in experimental studies. The reason is that the drivers are required to perform the tasks during the experiment in order that the physical and/or the attentional effects of cognitive demand of those tasks can be measured. Thus in an experimental study, it is not possible to measure a driver's willingness or strategy used to perform the same task under real-world conditions, when they may delay or even choose not to perform a given task until the vehicle is stopped, or they otherwise self-regulate their driving behavior when performing such secondary tasks. Such self-regulatory/compensatory driving behavior generally is more evident in direct observations of drivers of instrumented vehicles in naturalistic driving studies, although self-regulatory driving behavior can be discovered even in some experimental studies as shown in this section and in Young (2014b, 2014c). Talk also decreases the RR estimate for drowsy driving in the 100-car NDS dataset (Young, 2013c), but this result should be confirmed in the SHRP2 (2014) NDS dataset. Secondary tasks also reduce crash risk because when drivers perform secondary tasks they make fewer driver behavior errors. For example, in the SHRP2 naturalistic driving study, performing secondary tasks reduced the OR estimates for speeding by about 50% compared to not performing secondary tasks (Young, 2017b).

Summary: Driver self-regulation compensates for the effects of Talk on attention while driving, and may account in part for the overall protective effect for Talk while driving found by the meta-analysis in the previous section (see Table 1 and Figure 1).

CONCLUSION

The results from six published naturalistic and real-world driving studies of passenger vehicles to date, after correction where needed for known errors and biases, concur that the risk of crashing during a cell phone conversation is not greater than the risk of crashing when not engaged in a cell phone conversation, or when just driving with no secondary task.

A meta-analysis of these six studies finds that their relative risk estimates are homogeneous, regardless of cell phone type, crash severity, driver sample, or epidemiological study type. The pooled Talk relative risk estimate is 0.63 (CI 0.52-0.76), indicating a protective effect of cell phone conversation, which most likely arises because of driver self-regulation.

FUTURE RESEARCH DIRECTIONS

The SHRP2 naturalistic driving study offers additional data on the different subtasks of cell phone usage besides conversation: dialing, texting, locating, answering, etc. The relative risks of these subtasks should also be estimated, and the results examined for homogeneity to see if the subtasks can be properly combined into an overall estimate of the relative risk for cell phone "usage." If the tasks are heterogeneous, the various subtasks in cell phone usage must be separately reported (and legislated if needed).

There is an opportunity for future research to do a case-crossover or matched case-control study of the SHRP2 data, to adjust for driver demographic factors, as well as roadway factors, weather, closeness to junction, and so forth. In addition, illegal and improper driver behaviors (as well as other secondary tasks) that co-occur with a secondary task must be adjusted for (Young, 2017b), to arrive at unbiased estimates of the relative risks of the various cell phone usage subtasks. Finally,

research should be done to determine whether cell phone conversation reduces the relative risk of drowsy driving in the SHRP2 study data, as it did in the 100-Car study data (Young, 2013c).

ACKNOWLEDGMENT

I thank Bruce Papazian and an anonymous reviewer for comments on an early draft of this chapter.

REFERENCES

- Angell, L., Auflick, J., Austria, P., Kochhar, D., Tijerina, L., Biever, W.,... Kiger, S. (2006). *Driver workload metrics task 2 final report* (DOT HS 810 635). Crash Avoidance Metrics Partnership (CAMP). Retrieved from <http://www.nhtsa.gov/DOT/NHTSA/NRD/Multimedia/PDFs/Crash%20Avoidance/Driver%20Distraction/Driver%20Workload%20Metrics%20Final%20Report.pdf>
- Brookhuis, K., de Vries, G., & de Waard, D. (1991). The effects of mobile telephoning on driving performance. *Accident; Analysis and Prevention*, 23(4), 309–316. doi:10.1016/0001-4575(91)90008-S PMID:1883470
- Dingus, T. A., Guo, F., Lee, S., Antin, J. F., Perez, M., Buchanan-King, M., & Hankey, J. (2016). Driver crash risk factors and prevalence evaluation using naturalistic driving data. [abstract]. *Proceedings of the National Academy of Sciences of the United States of America*, 113(10), 2636–2641. doi:10.1073/pnas.1513271113 PMID:26903657
- Dingus, T. A., Klauer, S. G., Neale, V. L., Petersen, A., Lee, S. E., Sudweeks, J.,... Knipling, R. R. (2006). The 100-car naturalistic driving study, Phase II—results of the 100-car field experiment. (DOT HS 810 593). Washington, DC: National Highway Traffic Safety Administration, Department of Transportation. Retrieved from <http://www.nhtsa.gov/DOT/NHTSA/NRD/Multimedia/PDFs/Crash%20Avoidance/2006/100CarMain.pdf>
- Fitch, G. A., Socolich, S. A., Guo, F., McClafferty, J., Fang, Y., Olson, R. L.,... Dingus, T. A. (2013, April). *The impact of hand-held and hands-free cell phone use on driving performance and safety-critical event risk* (Report No. DOT HS 811 757). Washington, DC: NHTSA. Retrieved from <http://www.nhtsa.gov/DOT/NHTSA/NVS/Crash%20Avoidance/Technical%20Publications/2013/811757.pdf>
- Hakamies-Blomqvist, L. (1994). Compensation in older drivers as reflected in their fatal accidents. *Accident; Analysis and Prevention*, 26(1), 107–112. doi:10.1016/0001-4575(94)90073-6 PMID:8110349
- ISO 17488:2016. (2016). *Road vehicles -- Transport information and control systems -- Detection-Response Task (DRT) for assessing attentional effects of cognitive load in driving*. 76 pages. Retrieved from http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=59887
- Jamson, A. H., Westerman, S. J., Hockey, G. R. J., & Carsten, O. M. J. (2004). Speech-based e-mail and driver behavior: Effects of an in-vehicle message system interface. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 46(4), 625-639. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/15709325>
- Kircher, A., Vogel, K., Törnros, J., Bolling, A., Nilsson, L., Patten, C.,... Ceci, R. (2004). *Mobile telephone simulator study*. Linköping, Sweden: Swedish National Road and Transport Research Institute (VTI). Retrieved from <http://www.vti.se/EPIBrowser/Publikationer%20-%20English/M969A.pdf>
- Klauer, S. G., Dingus, T. A., Neale, V. L., Sudweeks, J. D., & Ramsey, D. J. (2006). The impact of driver inattention on near-crash/crash risk: An analysis using the 100-Car naturalistic driving study data (Report No. DOT HS 810 594). Washington, DC: National Highway Traffic Safety Administration. Retrieved from <http://www.nhtsa.gov/DOT/NHTSA/NRD/Multimedia/PDFs/Crash%20Avoidance/Driver%20Distraction/810594.pdf>

- Klauer, S. G., Guo, F., Simons-Morton, B. G., Ouimet, M. C., Lee, S. E., & Dingus, T. A. (2014). Distracted driving and risk of road crashes among novice and experienced drivers. *The New England Journal of Medicine*, 370(1), 54–59. doi:10.1056/NEJMs1204142 PMID:24382065
- Klauer, S. G., Guo, F., Sudweeks, J. D., & Dingus, T. A. (2010). *An analysis of driver inattention using a case-crossover approach on 100-Car data: Final Report* (DOT HS 811 334). Virginia Tech Transportation Institute. Retrieved from <http://www.nhtsa.gov/DOT/NHTSA/NVS/Crash%20Avoidance/Technical%20Publications/2010/811334.pdf>
- Langford, J., Methorst, R., & Hakamies-Blomqvist, L. (2006). Older drivers do not have a high crash risk — A replication of low mileage bias. *Accident; Analysis and Prevention*, 38(3), 574–578. doi:10.1016/j.aap.2005.12.002 PMID:16426560
- McEvoy, S. P., Stevenson, M. R., McCartt, A. T., Woodward, M., Haworth, C., Palamara, P., & Cercarelli, R. (2005). Role of mobile phones in motor vehicle crashes resulting in hospital attendance: A case-crossover study. *BMJ (Clinical Research Ed.)*, 331(7514), 428–430. doi:10.1136/bmj.38537.397512.55 PMID:16012176
- Porta, M. (2008). *A Dictionary of Epidemiology* (5th ed.). New York: Oxford University Press.
- Ranney, T., Watson, G., Mazzae, E., Papelis, Y., Ahmad, O., & Wightman, J. (2004). Examination of the distraction effects of wireless phone interfaces using the National Advanced Driving Simulator-Preliminary report on a freeway study (DOT HS 809 737). Washington, DC: U.S. Department of Transportation, NHTSA. Retrieved from http://www.nhtsa.gov/DOT/NHTSA/NRD/Multimedia/PDFs/VRTC/ca/capubs/Wireless1F_FinalReport.pdf
- Redelmeier, D. A., & Tibshirani, R. J. (1997). Association between cellular-telephone calls and motor vehicle collisions. *The New England Journal of Medicine*, 336(7), 453–458. doi:10.1056/NEJM199702133360701 PMID:9017937
- Regan, M. A., Hallett, C., & Gordon, C. P. (2011). Driver distraction and driver inattention: Definition, relationship and taxonomy. *Accident; Analysis and Prevention*, 43(5), 1771–1781. doi:10.1016/j.aap.2011.04.008 PMID:21658505
- Rothman, K., Greenland, S., & Lash, T. (2008). *Modern epidemiology* (3rd ed.). Philadelphia, PA: Lippincott Williams & Wilkins.
- Rothman, K. J. (2015). *Episheet: Spreadsheets for the analysis of epidemiologic data*. Retrieved from <http://krothman.org/Episheet.xls>
- SHRP2. (2014). *The SHRP 2 Naturalistic Driving Study*. Retrieved from http://www.trb.org/StrategicHighwayResearchProgram2SHRP2/Pages/The_SHRP_2_Naturalistic_Driving%20Study_472.aspx
- Strayer, D. L., Cooper, J. M., Turrill, J., Coleman, J., Medeiros-Ward, N., & Biondi, F. (2013, June 12). *Measuring cognitive distraction in the automobile*. Washington, DC: AAA Foundation for Traffic Safety. Retrieved from <https://www.aaafoundation.org/sites/default/files/Measuring-CognitiveDistractions.pdf>
- Strayer, D. L., Cooper, J. M., Turrill, J., Coleman, J. R., & Hopman, R. J. (2015b). *Measuring cognitive distraction in the automobile III: A comparison of ten 2015 in-vehicle information systems*. Washington, DC: AAA Foundation for Traffic Safety. Retrieved from https://www.aaafoundation.org/sites/default/files/strayerIII_FINALREPORT.pdf
- Strayer, D. L., Cooper, J. M., Turrill, J., Coleman, J. R., & Hopman, R. J. (2015c). *The Smartphone and the driver's cognitive workload: A comparison of Apple, Google, and Microsoft's intelligent personal assistants*. Retrieved from https://www.aaafoundation.org/sites/default/files/strayer-IIIa_FINALREPORT.pdf

- Strayer, D. L., Turrill, J., Coleman, J. R., Ortiz, E. V., & Cooper, J. M. (2014, October). *Measuring cognitive distraction in the Automobile II: Assessing in-vehicle voice-based interactive technologies*. Washington, DC: AAA Foundation for Traffic Safety. Retrieved from https://www.aaafoundation.org/sites/default/files/Cog%20Distraction%20Phase%202%20FINAL%20FTS%20FORMAT_0.pdf
- Strayer, D. L., Turrill, J., Cooper, J. M., Coleman, J. R., Medeiros-Ward, N., & Biondi, F. (2015a). Assessing cognitive distraction in the automobile. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 57(8), 1300-1324. Retrieved from <http://hfs.sagepub.com/content/57/8/1300.abstract>
- VTTI. (2014). *VTTI Data Warehouse*. Retrieved from <http://www.access.vtti.vt.edu/>
- Yannis, G., Papadimitriou, E., Karekla, X., & Kontodima, E. (2010). Mobile phone use by young drivers: Effects on traffic speed and headways. *Transportation Planning and Technology*, 33(4), 385-394. Retrieved from https://www.researchgate.net/publication/245313548_Mobile_phone_use_by_young_drivers_effects_on_traffic_speed_and_headways
- Young, R., Hsieh, L., & Seaman, S. (2016b). The Dimensional Model of Driver Demand: Extension to auditory-vocal and mixed-mode tasks. *SAE International Journal of Transportation Safety*, 4(2016-01-1427), 72-106. Retrieved from <http://papers.sae.org/2016-01-1427/>
- Young, R., Hsieh, L., & Seaman, S. (2016b). The Dimensional Model of Driver Demand: Extension to Auditory-Vocal and Mixed-Mode Tasks. *SAE International Journal of Transportation Safety*, 4(2016-01-1427), 72-106. Retrieved from <http://papers.sae.org/2016-01-1427/>
- Young, R., Seaman, S., & Hsieh, L. (2016a). The Dimensional Model of Driver Demand: Visual-manual tasks. *SAE International Journal of Transportation Safety*, 4(2016-01-1423), 33-71. Retrieved from <http://papers.sae.org/2016-01-1423/>
- Young, R., Seaman, S., & Hsieh, L. (2016a). The Dimensional Model of Driver Demand: Visual-Manual Tasks. *SAE International Journal of Transportation Safety*, 4(2016-01-1423), 33-71. Retrieved from <http://papers.sae.org/2016-01-1423/>
- Young, R. A. (2001). *Association between embedded cellular phone calls and vehicle crashes involving airbag deployment*. Paper from the Driving Assessment 2001: The First International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, Aspen, CO. Retrieved from http://drivingassessment.uiowa.edu/DA2001/81_young-richard.pdf
- Young, R. A. (2012a). Cell phone use and crash risk: Evidence for positive bias. *Epidemiology (Cambridge, Mass.)*, 23(1), 116-118. doi:10.1097/EDE.0b013e31823b5efc PMID:22082996
- Young, R. A. (2012b). Event detection: The second dimension of driver performance for visual-manual tasks. *SAE Int. J. Passeng. Cars - Electron. Electr. Syst.*, 5(1). doi:10.4271/2012-01-0964
- Young, R. A. (2013a). *Naturalistic studies of driver distraction: effects of analysis methods on odds ratios and population attributable risk*. Paper from the 7th International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, Bolton Landing, NY. Retrieved from http://drivingassessment.uiowa.edu/sites/default/files/DA2013/Papers/077_Young_0.pdf

Young, R. A. (2013b). *Cell phone conversation and automobile crashes: Relative risk is near 1, not 4*. Paper from the 3rd International Conference on Driver Distraction and Inattention, Gothenburg, Sweden. Retrieved from <http://document.chalmers.se/download?docid=cf54630-edad-4476-b145-bd46fc08d9b7>

Young, R. A. (2013c). *Drowsy Driving Increases Severity of Safety-Critical Events and Is Decreased by Cell Phone Conversation*. Paper from the 3rd International Conference on Driver Distraction and Inattention, Gothenburg, Sweden. Retrieved from <http://document.chalmers.se/download?docid=19e9af22-8aec-4b5e-95d5-c24d9d286020>

Young, R. A. (2014a). An unbiased estimate of the relative crash risk of cell phone conversation while driving an automobile. *SAE Int. J. Trans. Safety*, 2(1), 46-66. Retrieved from <http://papers.sae.org/2014-01-0446/>

Young, R. A. (2014b). Self-regulation minimizes crash risk from attentional effects of cognitive load during auditory-vocal tasks. *SAE Int. J. Trans. Safety*, 2(1), 67-85. Retrieved from https://www.researchgate.net/publication/261472492_Self-Regulation_Minimizes_Crash_Risk_from_Attentional_Effects_of_Cognitive_Load_during_Auditory-Vocal_Tasks

Young, R. A. (2014c). *Self-regulation reduces crash risk from the attentional effects of cognitive load from auditory-vocal tasks*. Paper from the Transportation Research Board 93rd Annual Meeting, No. 14-1010. Retrieved from https://www.researchgate.net/publication/272153134_Self-Regulation_Reduces_Crash_Risk_from_the_Attentional_Effects_of_Cognitive_Load_from_Auditory-Vocal_Tasks

Young, R. A. (2014d). Distracted driving and crash risk. *The New England Journal of Medicine*, 370(16), 1564-1566. doi:10.1056/NEJMc1401241 PMID:24738682

Young, R. A. (2015a). Cell phone conversation and relative crash risk. In Y. Zheng (Ed.), *Encyclopedia of Mobile Phone Behavior* (pp. 1274-1306). Hershey, PA: IGI Global. Retrieved from <http://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/978-1-4666-8239-9.ch102>

Young, R. A. (2015b). Driver compensation: Impairment or improvement? *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 57(8), 1334-1338. Retrieved from <http://hfs.sagepub.com/content/57/8/1334.full.pdf+html>

Young, R. A. (2015c). *Revised odds ratio estimates of secondary tasks: A re-analysis of the 100-Car naturalistic driving study data*. SAE Technical Paper 2015-01-1387. Retrieved from <http://papers.sae.org/2015-01-1387/>

Young, R. A. (2017a). Removing biases from crash odds ratio estimates of secondary tasks: A new analysis of the SHRP 2 naturalistic driving study data. Paper #2017-01-1380 from the Society of Automotive Engineers, Detroit, MI USA. Retrieved from <http://papers.sae.org/2017-01-1380/>

Young, R. A. (2017b). Adjusted crash odds ratio estimates of driver behavior errors: A re-analysis of the SHRP 2 naturalistic driving study data. Paper from the Proceedings of Driving Assessment 2017: The 9th International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, Manchester Village, Vermont.

Young, R. A., & Angell, L. S. (2003). *The dimensions of driver performance during secondary manual tasks*. Paper from the Driving Assessment 2003: The Second International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, Park City, UT. Retrieved from http://drivingassessment.uiowa.edu/DA2003/pdf/25_Youngformat.pdf

Young, R. A., Hsieh, L., & Seaman, S. (2013d). *The Tactile Detection Response Task: Preliminary validation for measuring the attentional effects of cognitive load*. Paper from the 7th International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, Bolton Landing, NY. Retrieved from http://drivingassessment.uiowa.edu/sites/default/files/DA2013/Papers/012_Young_0.pdf

Young, R. A., & Schreiner, C. (2009). Real-world personal conversations using a hands-free embedded wireless device while driving: Effect on airbag-deployment crash rates. *Risk Analysis*, 29(2), 187–204. doi:10.1111/j.1539-6924.2008.01146.x PMID:19000076

ADDITIONAL READING

Young, R. A. (2012). Cell phone use and crash risk: Evidence for positive bias. *Epidemiology (Cambridge, Mass.)*, 23(1), 116–118. https://www.researchgate.net/publication/51797905_Cell_phone_use_and_crash_risk_evidence_for_positive_bias doi:10.1097/EDE.0b013e31823b5efc PMID:22082996

Young, R. A. (2013). *Cell phone conversation and automobile crashes: Relative risk is near 1, not 4*. Paper from the 3rd International Conference on Driver Distraction and Inattention, Gothenburg, Sweden. <http://document.chalmers.se/download?docid=cfd54630-edad-4476-b145-bd46fc08d9b7>

Young, R. A. (2015). Cell phone conversation and relative crash risk. In Y. Zheng (Ed.), *Encyclopedia of Mobile Phone Behavior* (pp. 1274–1306). Hershey, PA, USA: IGI Global; <http://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/978-1-4666-8239-9.ch102> doi:10.4018/978-1-4666-8239-9.ch102

Young, R. A. (2017). Removing biases from crash odds ratio estimates of secondary tasks: A new analysis of the SHRP 2 naturalistic driving study data. Paper #2017-01-1380 from the Society of Automotive Engineers, Detroit, MI USA. Retrieved from <http://papers.sae.org/2017-01-1380/>

Young, R. A., & Schreiner, C. (2009). Real-world personal conversations using a hands-free embedded wireless device while driving: Effect on airbag-deployment crash rates. *Risk Analysis*, 29(2), 187–204. https://www.researchgate.net/publication/23464587_Real-world_personal_conversations_using_a_hands-free_embedded_wireless_device_while_driving_effect_on_airbag-deployment_crash_rates?ev=prf_cit doi:10.1111/j.1539-6924.2008.01146.x PMID:19000076

KEY TERMS AND DEFINITIONS

Cell Phone Conversation: Talking/listening on a wireless device. Abbreviated in this chapter as Talk.

Cognitive Distraction: “Any epoch of cognitive loading that competes with activities necessary for safe driving” (Foley et al., 2013).

Cognitive Load: The loading of cognitive resources, “which are the alerting, executive, and orienting attentional networks singly or in combination, as well as the memory and representational systems (e.g., working and long-term) from which information may be retrieved and in which it may be held and operated upon” (Foley et al., 2013). Same as Cognitive demand (see Introduction).

Compensation: As used here, a synonym for self-regulation. For example, during a cell phone conversation, drivers tend to compensate for the 200-300 millisecond increase in their brake response time by increasing their headway times to a lead vehicle.

Crash Risk: The probability of crash occurrence.

Driver Distraction: “The diversion of attention away from activities critical for safe driving toward a competing activity, which may result in insufficient or no attention to activities critical for safe driving” (Regan et al., 2011, p. 1776).

Driver Attention: Attention is three entities, not one as commonly believed. The alerting, executive, and orienting attentional networks in the brain are separate networks that can be affected in opposite ways by cell phone tasks. For example, Talk improves alerting attention, while slightly diminishing orienting attention.

Relative Risk: As used in the current chapter, the crash risk of performing a secondary task

relative to the baseline risk when not performing that secondary task. Relative risk is used here as a generic term to refer to the risk ratio, the rate ratio, or the odds ratio, all of which are technical terms to estimate relative risk in different epidemiological study designs.

Secondary Task: Any task performed in a vehicle that is not related to the primary driving tasks.

Self-Regulation: A change in tactical driving behavior to compensate for adverse effects on safety from a secondary task. A synonym for compensation.

Talk RR: The relative risk of cell phone conversation while driving.

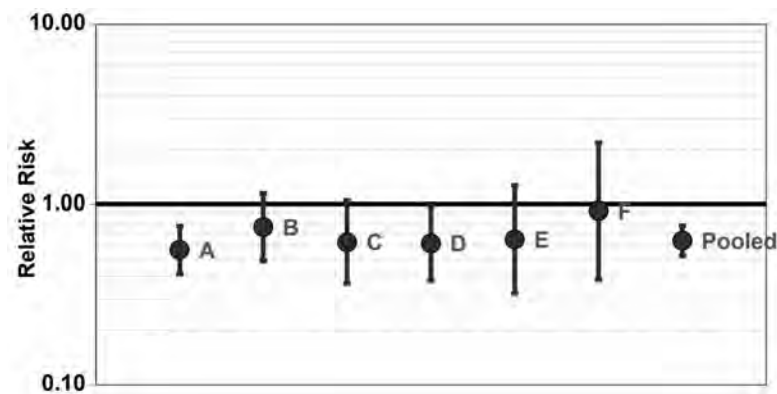
APPENDIX

Table 1. Studies discussed in main text

Study	Wireless Device	Dataset	RR Estimate	Crash Type	Study	RR	LL	UL
A	hand-held	100-Car NDS	adjusted	Crash & Near-Crash	Young (2013b, 2014a)	0.56	0.41	0.76
B	all	cell-phone NDS	adjusted	Safety-Critical	Fitch et al. (2013)	0.75	0.49	1.15
C	embedded	real-world	crude	Airbag Deployment	Young & Schreiner (2009)	0.62	0.37	1.05
D	hand-held	real-world	adjusted	Injury	Redelmeier & Tibshirani (1997)	0.61	0.38	0.98
E	hand-held	real-world	adjusted	Injury	McEvoy et al. (2005)	0.64	0.32	1.27
F	hand-held	SHRP2 NDS	crude	Moderate to Severe	Young (2017a)	0.92	0.29	2.20
A-F					pooled	0.63	0.52	0.76

Key: RR = Risk Ratio; LL = lower 95% confidence limit; UL = upper 95% confidence limit.

Figure 1. Plot of the six Talk relative risk estimates in Table 1 with confidence intervals. The pooled estimate is to the right. The pooled estimate and graph were done using Episheet (Rothman, 2015).



Comb Filters Characteristics and Current Applications

M

Miriam Guadalupe Cruz-Jimenez

Institute INAOE, Mexico

David Ernesto Troncoso Romero

CONACYT at ESCOM-IPN, Mexico

Gordana Jovanovic Dolecek

Institute INAOE, Mexico

INTRODUCTION

A comb filter is a Linear Time-Invariant (LTI) *digital filter*, where linear means that its output to a *scaled* sum of input *digital signals* is equal to the scaled sum of the outputs to every one of these input signals (i.e., the filter satisfies the superposition principle) and time-invariant means that, for any input signal that has a given *delay*, the output undergoes the same delay as the input (Antoniu, 2006). The name comb is derived by the fact that its *magnitude response* resembles the teeth of a comb. Since there are several filters having magnitude responses with such characteristic, the term comb filter is rather general. The duration of the *impulse response* of comb filters can be either finite or infinite, i.e., there are Finite Impulse Response (FIR) comb filters and Infinite Impulse Response (IIR) comb filters (Zölzer, 2008).

The simplest FIR comb filter has the following *transfer function*:

$$H_a(z) = 1 + z^{-M} \quad (1)$$

This filter adds to a signal a version of that signal delayed by M sample periods, and it is the basic building block to introduce echo effects in audio signals (Zölzer, 2008). Moreover, if the addition in (1) is replaced by a subtraction, the resulting comb filter is a useful building block

to remove DC and harmonics (Diniz, Da Silva & Neto, 2010). The unintentional delay of an audio signal due to the environment is also modeled as a comb filter (Toole, Shaw, Daigle & Stinson, 2001), and this effect may be undesirable in many cases. Similarly, a simple IIR comb filter has the following transfer function:

$$H_b(z) = 1/(1 - az^{-M}) \quad (2)$$

with $a < 1$. This filter is a basic building block to model and create reverberation effects or, in general, to artificially reproduce the acoustics of a room (Zölzer, 2008).

One of the most important comb filters for several Digital Signal Processing (DSP) applications is the one based in the FIR filter where all the samples of its impulse response have values equal to one. Unlike the aforementioned comb filters described by $H_a(z)$ in (1) and $H_b(z)$ in (2), this comb filter has a low-pass characteristic, which makes it useful to pass a baseband signal and remove unwanted high-frequency spectra (Milic, 2009). The rest of this chapter is dedicated to this particular filter, which will be referred as the comb filter hereafter. The main characteristics of the comb filter, as well as its advantages and disadvantages will be highlighted. Moreover, we will present the selected methods commonly used to decrease the disadvantages of the comb filters with minimum affectation of its advantages.

In the efficient implementation of the comb filter, a comb filter with transfer function based in (1) (just with the addition replaced by a subtraction) is employed. In order to avoid confusion, that filter will be referred hereafter as comb differentiator, since it is based on a simple first-order differentiator.

Background

Consider a simple FIR filter that has the following transfer function (Milic, 2009):

$$H(z) = \frac{1}{M} \sum_{k=0}^M z^{-k} \quad (3)$$

where M is the filter order. The coefficients that multiply the variable z are all equal to 1. Thus, the non-recursive implementation of this filter does not require *multipliers* for its coefficient's values. The impulse response of this filter is $h(k) = 1$ for $0 \leq k \leq M$ and 0 for other values of k . The scaling factor ($1/M$) is included to provide a normalized gain of 0 dB at frequency equal to zero.

The transfer function of the comb filter arises from expressing the transfer function given in (3) in recursive form as follows (Lyons, 2004):

$$H(z) = \frac{1}{M} \frac{1 - z^{-M}}{1 - z^{-1}} \quad (4)$$

This transfer function is the starting point for an efficient implementation of the filtering required in systems where the sampling rate is increased (upsampling) or decreased (downsampling). Both, the non-recursive structure to implement (3) and the recursive structure to implement (4) are shown in Figure 1.

In the context of digital filters, non-recursive means that the filtering structure does not require a feedback, whereas recursive means that the filtering structure has a feedback (i.e., the output depends on delayed versions of that output). Since the transfer function (3) and (4) are equivalent,

the comb filter is, in fact, the same. We observe from (4) that the recursive transfer function is the product of the terms $1/(1-z^{-1})$, which is an integrator, and $(1-z^{-M})$, which is a differentiator whose delay has been replaced by M delays. The frequency response of the comb filter is:

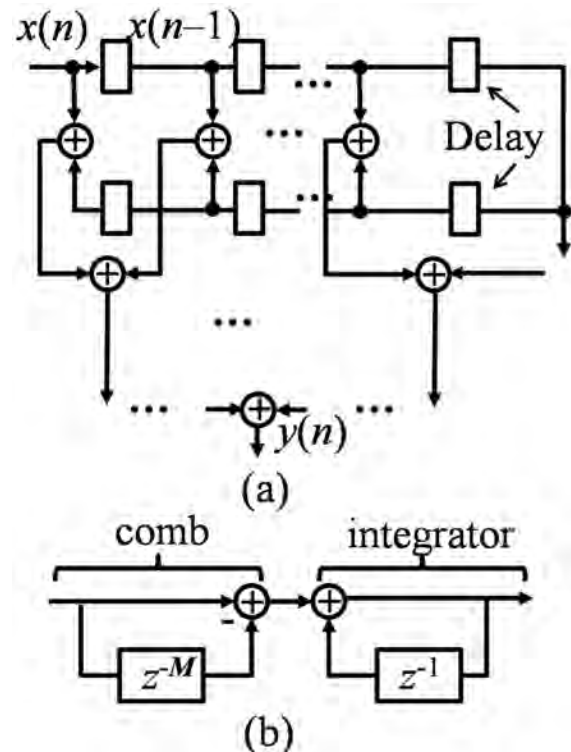
$$H(e^{j\omega}) = \frac{1}{M} \frac{\sin(\omega M / 2)}{\sin(\omega / 2)} e^{-j\omega(M-1)/2} \quad (5)$$

where ω is the normalized angular frequency.

The comb filter is very popular because it has the following advantages:

- It is a FIR recursive system, so it has linear-phase and guaranteed stability whenever the proper bus width is used (Lyons, 2004);

Figure 1. Filtering structures: (a) FIR comb filter in non-recursive structure and (b) recursive comb filter with an integration part and a comb differentiation part, where the delay z^{-1} has been replaced by z^{-M}



- It is multiplier-free and it only requires two additions (in fact they are subtractions, but *adders* and subtractors are considered with equal hardware complexity). Since *multipliers* are the most expensive elements in a filter, the comb filter can be used to design low-complexity multiplierless filters (Lyons, 2004);
- In the frequency range from $\omega=0$ to $\omega=\pi$, the zeros of the magnitude response of a comb filter are placed over the frequencies $\omega_k = 2\pi k/M$, for $k = 1, 2, \dots, \lfloor M/2 \rfloor$ ($\lfloor x \rfloor$ denotes the integer part of x). These zeros provide stopband regions with natural rejection to undesirable signals that are introduced into the band of interest when the sampling rate of such signals is decreased (an effect called aliasing) (Milic, 2009).

Due to the aforementioned characteristics in its magnitude response, the comb filter is commonly used as a multi-band filter, with one passband and several stopband regions. The passband, Ω_p , and the k -th stopband, $\Omega_{s,k}$, are given as follows:

$$\Omega_p = [0, \omega_p] \quad (6)$$

$$\Omega_{s,k} = [(2\pi k/M) - \omega_p, (2\pi k/M) + \omega_p] \quad (7)$$

where:

$$\omega_p \leq \pi/(2M) \quad (8)$$

$$k = 1, 2, \dots, \lfloor M/2 \rfloor \quad (9)$$

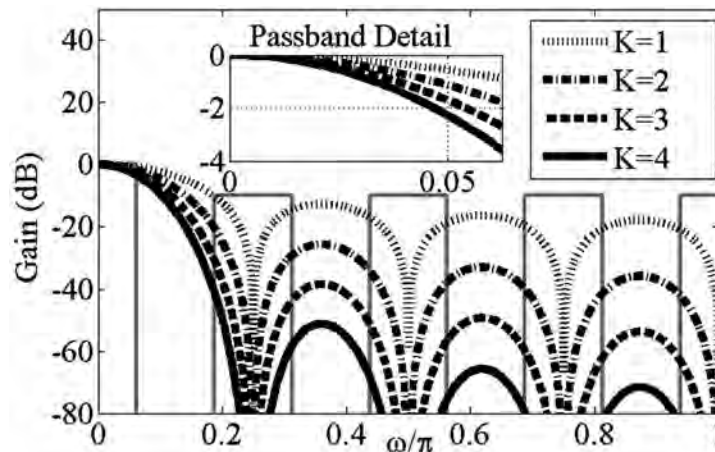
Nevertheless, the magnitude response of the comb filter is, in general, poor. These filters have the following disadvantages:

- The filter exhibits a passband droop that distorts the signals of interest;
- The useful stopband regions around the zeros of the comb filter are very narrow.

In order to improve the attenuation in the stopbands, K comb filters (commonly, the recursive blocks presented in Figure 1b) are used in cascade. However, this worsens the passband droop. As an example, Figure 2 illustrates the magnitude response of a comb filter with $M = 8$ and $K = 1, 2, 3$ and 4, where the bands of interest for the case $\omega_p = \pi/(2M)$ are enclosed into vertical rectangles. Note that the worst-case droop occurs in the right edge of the passband and the worst-case attenuation occurs in the left edge of the first stopband.

Active research has been developed in the last three decades to solve the aforementioned disad-

Figure 2. Magnitude responses of a comb filter with $M = 8$, $\omega_p = \pi/(2M)$ and $K = 1, 2, 3$ and 4



vantages. In the following section we detail the general objectives in this research line, along with two recent methods that are representative of it.

METHODS TO IMPROVE THE MAGNITUDE CHARACTERISTICS OF COMB FILTERS

Over the past three decades, research on comb filters has been carried out to improve the magnitude response characteristic of comb filters with the lowest increase in *computational complexity*. There are three different ways:

1. Improvement just in the passband region;
2. Improvement just in the stopband region; or
3. Improvement in both, passband and stopband regions.

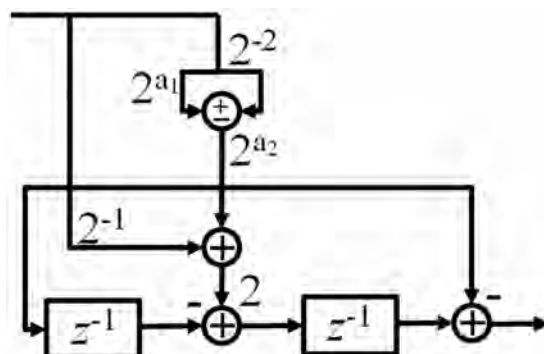
In the following we explain some of the most representative and recent methods in these categories.

Magnitude Response Improvement in the Passband with Compensators

A compensator is a low-order filter that has every delay element replaced by M delays. The magnitude response of this filter compensates for the droop of the comb filter in the passband because it exhibits a raise from 0 dB inversely proportional to the passband droop.

The proposal of (Romero & Dolecek, 2013) presents a simple method to design compensation filters based on the cascaded interconnection of second-order subfilters that, even though exhibit a droop, only need three additions. The shape of the droop is inverted by using some extra coefficients and, for the case of using only one of these subfilters, a simple coefficient is used, which needs only one addition to be implemented. Thus, the resulting compensator requires 4 additions and corrects in an effective way the passband droop of comb filters independently of both, the value of M

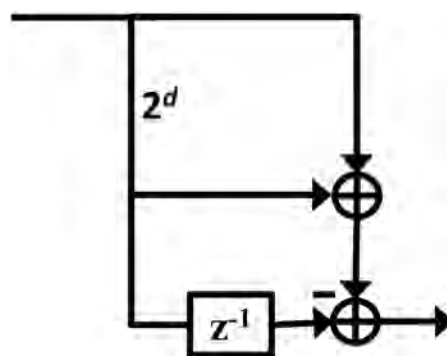
Figure 3. Simple compensation filter with four additions



and the bandwidth of the passband. This second-order compensation filter is shown in Figure 3, where the values a_1 and a_2 represent variable shifts that change depending on the bandwidth of the passband.

To our knowledge, the simplest compensator available in scientific literature has been recently proposed by (Romero, Molina & Dolecek, 2015), and it is a first-order filter that requires only 2 additions. It corrects the passband droop independently of the value of M , but it works better for a narrow bandwidth of the passband. This filter also presents a desirable low delay property at expenses of a slight phase distortion. This filter is shown in Figure 4, where d represents variable shifts that change depending on the bandwidth of the passband.

Figure 4. Simple compensation filter with two additions



As a simple example, consider the case of $M = 16$, $\omega_p = \pi/(4M)$ and $K = 4$. Figure 5 shows the magnitude responses of the original comb filter, the compensation filter and the compensated comb filter with the four-adders compensator. Figure 6 presents similar graphs for the two-adders compensator. Clearly the compensated comb filters have a much better passband characteristic, achieved with a low additional computational cost. Also note that the attenuation remains almost equal in the original comb filters and in the compensated comb filters.

Magnitude Response Improvement in the Stopband Region

The proposal of (Coleman, 2012) is based on the stopband attenuation improved by modifying the cascade of K recursive comb filters with the addition of some extra coefficients and delay elements. Since the worst-case attenuation of the comb filter occurs in the left-side edge of its first stopband, the additional complexity allows improving the attenuation in this stopband, whereas the attenuation in the other stopbands is slightly degraded. At the end, all the stopbands have nearly the same attenuation, which is higher than the worst-case attenuation obtained with K cascaded comb filters.

Therefore, for the same worst-case attenuation, this scheme needs less cascaded comb filters than the traditional approach. As a consequence, the overall filter uses a narrower bus width.

Coleman presented the mathematical framework that allows having a simple method to obtain the values of the additional coefficients using the N -th degree Chebyshev polynomials, $T_N(x)$, which can be easily calculated with the following recursive formula (Coleman, 2012):

$$T_0(x) = 1 \quad (10)$$

$$T_1(x) = x \quad (11)$$

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x) \quad (12)$$

With this, no intricate optimization method is needed. The value N indicates the number of cascaded comb filters, and the polynomial coefficients represent the extra complexity added to the N cascaded comb filters.

Let us consider a simple example where $M = 16$ and $\omega_p = \pi/(2M)$. Figure 7 shows the magnitude responses of the original comb filter with $K = 7$, and the modified comb filter with $N = 5$.

Figure 5. Magnitude responses of the comb filter, the compensation filter and the compensated comb filter (four-adders compensator), for $M = 16$, $\omega_p = \pi/(4M)$ and $K = 4$

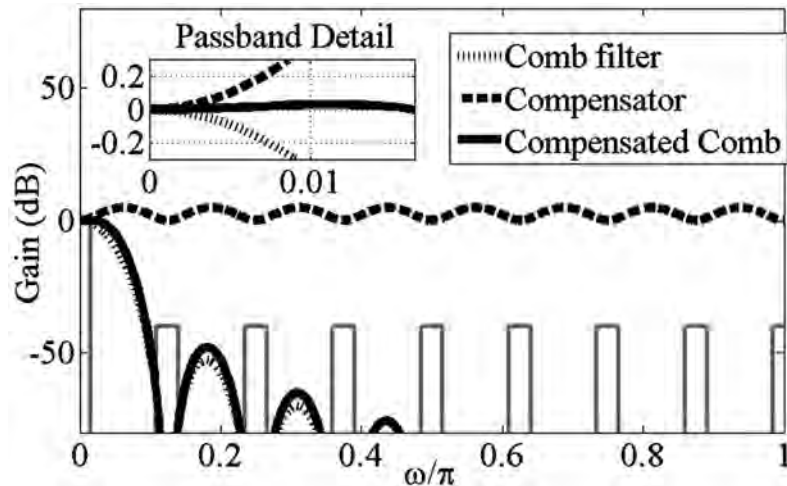
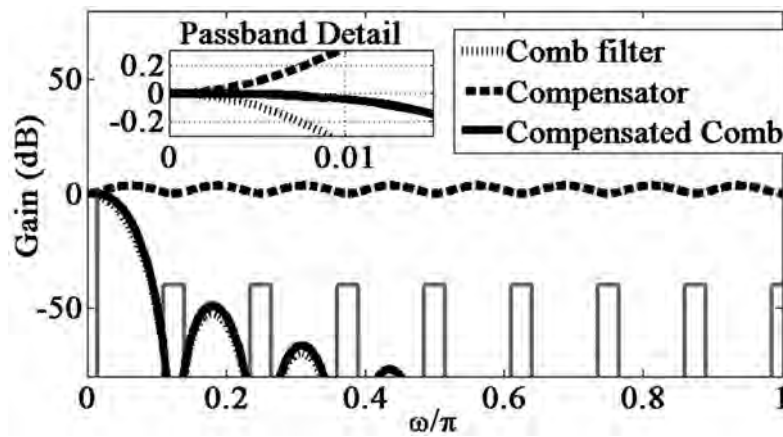


Figure 6. Magnitude responses of the comb filter, the compensation filter and the compensated comb filter (two-adders compensator), for $M = 16$, $\omega_p = \pi/(4M)$ and $K = 4$



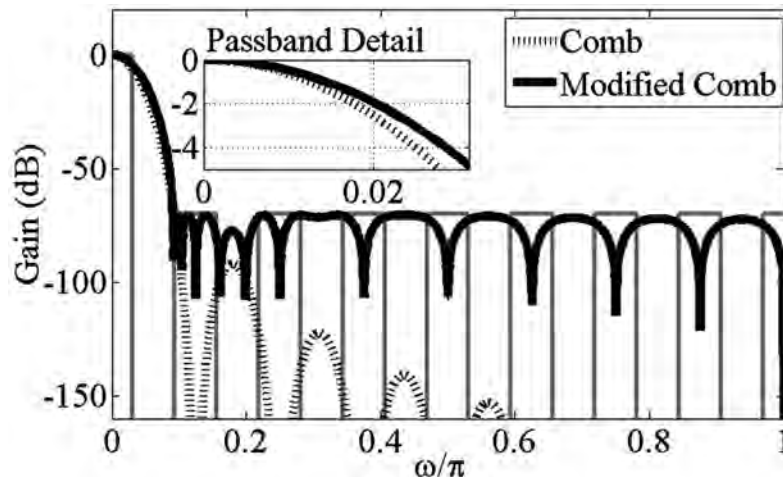
Note that both filters have 70 dB of worst-case attenuation. However, the modified comb filter uses 5 cascaded comb filters in comparison to the traditional comb filter where 7 cascaded comb filters are used.

Simultaneous Magnitude Response Improvement in Passband and Stopband

The sharpening technique permits simultaneous improvements of both, passband and stopband

characteristics. This technique is based on an Amplitude Change Function (ACF) which is a polynomial $P_{m,n}(x)$ that maps the amplitude x into an improved amplitude $y = P_{m,n}(x)$. The improvement in the amplitude near to the passband depends on m , the order of tangency of the ACF at the point $(x, y) = (1, 1)$ to a line with slope equal to zero. Similarly, the improvements in amplitudes near the stopband depend on the order of tangency of the ACF at the point $(x, y) = (0, 0)$ to a line with slope equal to zero, which is denoted as n . The polynomial approximation to the desired ACF also

Figure 7. Magnitude responses of the comb filter and the modified comb filter by (Coleman, 2012), for $M = 16$, $\omega_p = \pi/(2M)$, $K = 7$ and $N = 5$



is controlled by other two parameters, namely, σ , the slope of a line that passes over the point $(x, y) = (1, 1)$ and δ , the slope of another line that passes over the point $(x, y) = (0, 0)$. The constraints on the approximating polynomial $y = P_{\sigma, \delta, m, n}(x)$ are:

1. The n th-order tangency at $(x, y) = (0, 0)$ to the line of slope δ , i. e., $P_{\sigma, \delta, m, n}(x) \approx \delta x$, for $x \approx 0$;
2. The m th-order tangency at $(x, y) = (1, 1)$ to the line of slope σ , i. e., $P_{\sigma, \delta, m, n}(x) \approx \sigma(x - 1) + 1$, for $x \approx 1$.

A general formula was deduced by (Samadi, 2000) to obtain directly the desired amplitude change function from the design parameters. The formula is given by:

$$P_{\sigma, \delta, m, n}(x) = \delta x + \sum_{j=n+1}^R (\alpha_{j,0} - \sigma \alpha_{j,1} - \delta \alpha_{j,2}) x^j \quad (13)$$

where $R = n + m + 1$ and:

$$\alpha_{j,0} = \sum_{i=n+1}^j (-1)^{j-i} \binom{R}{j} \binom{j}{i}, \quad \alpha_{j,1} = \sum_{i=n+1}^j (-1)^{j-i} \binom{R}{j} \binom{j}{i} \left(1 - \frac{i}{R}\right), \quad \text{and} \quad \alpha_{j,2} = \sum_{i=n+1}^j (-1)^{j-i} \binom{R}{j} \binom{j}{i} \frac{i}{R} \quad (14)$$

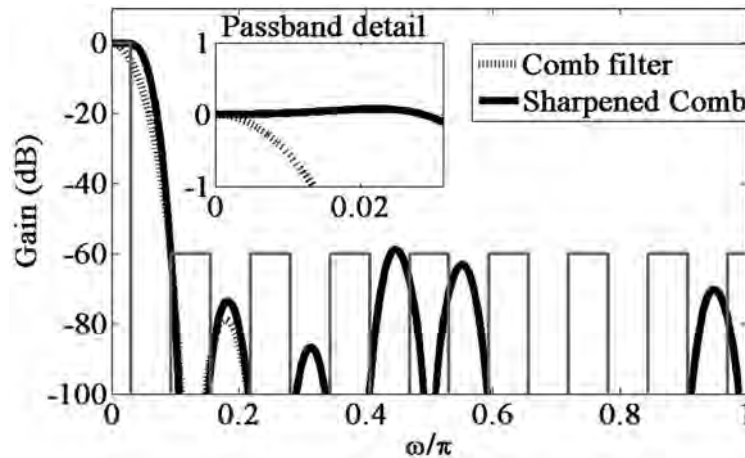
with:

$$\binom{n}{k} = \frac{n \times (n-1) \times \cdots \times 1}{[k \times (k-1) \times \cdots \times 1] \times [(n-k) \times (n-k-1) \times \cdots \times 1]} \quad (15)$$

The simultaneous improvement of the passband and stopband characteristics of a comb filter presented by (Jimenez & Dolecek, 2013) and (Jimenez & Dolecek, 2015) consists in splitting that filter into stages, and then using the aforementioned improved sharpening technique on the last stage comb filter, which is previously compensated. The sharpening polynomials are translated to cascaded compensated comb filters aided with low-complexity coefficients. The resulting architectures resemble that of Coleman's filters presented in the previous section, but a much better improvement in passband and stopband is achieved. Usually, this comes at the cost of a more complex architecture.

A simple example to illustrate the sharpening technique is given in Figure 8. The sharpening polynomial uses tangencies $m = 1$, $n = 3$, $\sigma = -1$ and $\delta = 0$ and it is compared with a comb filter where $M = 16$, and $K = 6$. The passband edge is $\omega_p = \pi/(2M)$. Note that both filters accomplish 60 dB, but the sharpened filter has a compensated passband droop.

Figure 8. Magnitude responses of the comb filter with $K = 7$ and the modified comb filter by (Jimenez & Dolecek, 2013), for $M = 16$ and $\omega_p = \pi/(2M)$



FUTURE RESEARCH DIRECTIONS

A comb filter is a basic building block with a simple and hardware-efficient structure. However, its disadvantages are its magnitude characteristics. One of the most common applications of the comb filters presented in this chapter is as anti-aliasing filters in decimation processes. In these cases, the filter is placed before the downsampling process. Since the filtering incur in higher power consumption as their sampling frequency operation grows, a traditional anti-aliasing filter may consume high power. The recursive structure of the comb filter is effective because its comb differentiator part can be moved to the low-rate section (in that case, only one delay z^{-1} is used, instead of the M delays shown in Figure 1b). However, the integrator part still works at high rate. Decreasing the computational complexity of these filters is a topic that has received much attention in current research (Molina, Dolecek & De La Rosa, 2014). Thus, one of the recent insights has been using the comb filter in its non-recursive form and decomposing it in subfilters that can work at the lower sampling rate section. This design scheme is generally known as polyphase decomposition (Milic, 2009).

The non-recursive form and polyphase decomposition of the comb filter has the disadvantage of demanding much chip area. Thus, a wise research path to be pursued is combining non-recursive and recursive structures of modified comb filters. In this way, the power and area demand can be lowered and, simultaneously, the modified comb filter can provide better magnitude characteristics than its traditional version.

CONCLUSION

The concept of comb filter was introduced in this chapter. We showed that comb filter is a general term that encloses several types of filters, including Finite Impulse Response (FIR) and Infinite

Impulse Response (IIR) systems. These filters share a common characteristic: their magnitude responses resemble the teeth of a comb. Additionally, the simplest FIR and IIR comb filters have been presented.

This chapter was mainly focused on illustrating the characteristics of the comb filter based on the moving average low-pass filter, which has unitary coefficients. Then, the efficient recursive form of this filter was presented, which requires only two addition operations. This structure represents the main advantage of the comb filter. However, its disadvantage is the poor characteristics of its magnitude response.

Four recent methods to improve the magnitude characteristic of comb filters have been introduced. The first two methods, based on the use of compensation filtering, correct the passband droop of the original comb filter. The third method uses Chebyshev polynomials to increase the worst-case attenuation of the original comb filter, whereas the fourth method uses sharpening polynomials to perform both improvements: droop correction and stopband attenuation. These methods have been illustrated with examples. The end of the chapter presents the future research directions on the design of comb-based FIR filters that are applied as anti-aliasing filters in decimation processes.

There are many existing methods that use comb filters for applications in, for example, audio, digital communications or analog to digital data conversion. In this chapter we have presented a concise introduction of comb filters, which can serve as a useful door to enter into the fascinating world of efficient processing of digital signals based on comb filtering.

REFERENCES

Antoniou, A. (2006). *Digital Signal Processing: Signals, Systems and Filters*. New York: McGraw-Hill.

Coleman, J. O. (2012). Chebyshev stopband for CIC decimation filters and CIC-implemented array tapers in 1D and 2D. *IEEE Trans. on Circuits and Systems I: Regular papers*, 59(12), 2956-2968.

Diniz, P. S. R., Da Silva, E. A. B., & Netto, S. L. (Eds.). (2010). *Digital signal processing system analysis and design*. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511781667

Jimenez, M. G. C., & Dolecek, G. (2013). Application of generalized sharpening technique for two-stage comb decimator filter design. *Elsevier Procedia Technology*, 7, 142-149. doi:10.1016/j.protcy.2013.04.018

Jimenez, M. G. C., & Dolecek, G. (2015). On compensated three-stages sharpened comb decimation filter. In W. Deng (Ed.), *Applied Engineering Sciences* (pp. 17-21). London, UK: CRC Press, Taylor and Francis group.

Lyons, R. G. (Ed.). (2004). *Understanding digital signal processing*. Prentice Hall.

Milic, L. (Ed.). (2009). *Multirate filtering for digital signal processing MATLAB applications*. Hershey, PA: IG Global. doi:10.4018/978-1-60566-178-0

Molina, G., Dolecek, G., & De La Rosa, J. M. (2014). An overview of decimation structures for efficient sigma-delta converters: trends, design issues and practical solutions. *Proceedings of IEEE International Symposium on Circuits and Systems*, 1592-1595.

Romero, D. E. T., & Dolecek, G. J. (2013). Application of amplitude transformation for compensation of comb decimation filters. *Electronics Letters*, 49(16), 985-987. doi:10.1049/el.2013.1492

Romero, D. E. T., Molina, G., & Dolecek, G. J. (2015). Simple two-adders CIC compensator. *Electronics Letters*, 51(13), 993-994. doi:10.1049/el.2015.0742

Samadi, S. (2000). Explicit formula for improved filter sharpening polynomial. *IEEE Transactions on Signal Processing*, 9(10), 2957-2959. doi:10.1109/78.869057

Toole, F. E., Shaw, E. A. G., Daigle, G. A., & Stinson, M. R. (2001). Sound Propagation. In J. Whitaker & K. Blair Benson (Ed.), *Standard handbook of audio and radio engineering*. McGraw Hill.

Zolzer, U. (Ed.). (2008). *Digital audio signal processing*. John Wiley & Sons Ltd. doi:10.1002/9780470680018

ADDITIONAL READING

Ahmed, N. Y., Ashour, M. A., & Nassar, A. M. (2008, March). *Power efficient polyphase decomposition comb decimator filter in multi-rate telecommunication receivers*. Paper presented at Mosharaka International Conference on Communications, Propagation and Electronics, Amman, Jordan. doi:10.1109/MICCPE.2008.4555757

Dolecek, G. J. (2009). Simple wide-band CIC compensator. *Electronics Letters*, 45(24), 1270-1272. doi:10.1049/el.2009.1860

Dolecek, G. J., & Laddomada, M. (2010). An economical class of droop-compensated generalized comb filters: Analysis and design. *IEEE Transactions on Circuits and Systems II, Express Briefs*, 57(4), 275-279. doi:10.1109/TCSII.2010.2043469

Dolecek, G. J., & Laddomada, M. (2010, March). *Comb-cosine prefilter based decimation filter*. Paper presented at IEEE International Conference on Industrial Technology, Viña del Mar, Valparaíso.

Isen, F. W. (2009). DSP for MATLAB and LabVIEW: Vol. III. *Digital Filter Design*. USA: Morgan & Claypool publishers.

Jack, K. (2007). *Video Demystified: A Handbook for the Digital Engineer*. Burlington, MA: Elsevier.

- Jia-Hui, L. Wang-heng, & Yu-hong, J. (2012, August). *One comb filter for lower power consumption*. Paper presented at International Conference on Computer Science and Information Processing (CSIP), Xi'an, Shaanxi.
- Karnati, N. R., Lee, K.-S., Carletta, J., & Vaillette, R. (2011, August). *A power-efficient polyphase sharpened CIC filter for sigma-delta ADCs*. Paper presented at in Proc. Int. Midwest Symp. on Circuits and Systems, Seoul, Korea. doi:10.1109/MWSCAS.2011.6026689
- Kim, S., Lee, W., Ahn, S., & Choi, S. (2006). Design of CIC roll-off compensation filter in W-CDMA digital IF receiver. *Digital Signal Processing*, 16(6), 846–854. doi:10.1016/j.dsp.2006.06.003
- Kwentus, A. Y., Jiang, Z., & Willson, A. N. (1997). Application of filter sharpening to cascaded integrator- comb decimation filters. *IEEE Transactions on Signal Processing*, 45(2), 457–467. doi:10.1109/78.554309
- Laddomada, M. (2007). Generalized comb decimation filters for sigma/delta A/D converters: Analysis and design. *IEEE Transactions on Circuits and Systems I*, 55(8), 2287–2299. doi:10.1109/TCSI.2008.920136
- Lentinen, V., & Renfors, M. (2007, February). *On CIC decimator variants-from shifting zeros to the sparse FIR-CIC structure*. Paper presented at 9th Int. Symp. on Signal Processing and Its applications, Sharjah.
- Liu, Q., & Gao, J. (2009, September). *Efficient comb decimator filter with sharpened magnitude response*. Paper presented at 5th. International Conference on Wireless Communications, Networking and Mobile Computing, Beijing, China.
- Liu, X., & Willson, A. N., Jr. (2010, June). *A 1.2 Gb/s recursive polyphase cascaded integrator-comb prefilter for high speed digital decimation filters in 0.18- μ m CMOS*. Paper presented at proceedings of the IEEE International Symposium on Circuits and Systems, Los Angeles, California.
- Molnar, G., & Vucic, M. (2011). Closed-Form of CIC compensator based on maximally flat error criterion. *IEEE Transactions on Circuits and Systems. II, Express Briefs*, 58(12), 926–930. doi:10.1109/TCSII.2011.2172522
- Newbold, R. (2013). *Practical Applications in Digital Signal Processing*. Ann Arbor, Michigan: Prentice Hall.
- Pecotic, M. G., Molnar, G., & Vucic, M. (2012, May). *Design of CIC compensators with SPT coefficients based on interval analysis*. Paper presented at in Proc. The 35th IEEE Int. Convention MIPRO, Opatija, Croatia.
- Pei, S., & Tseng, C. (1998). A comb filter design using fractional sample delay. *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, 45(5), 649–653. doi:10.1109/82.673650
- Rasmussen, J. L. (2012, November). *New method of FIR comb filtering*. Paper presented at Conference Record of the Forty Sixth Asilomar Conference on Signals, System and Computers, Pacific Grove, CA.
- Rudiger, K. (2002). *Decimation Lowpass filters for Sigma-Delta Modulator: A comparative study*. Germany: Diplomarbeiten Agentur diplom.de.
- Saramaki, T. (1983, September). *Efficient recursive digital filters for sample for sampling rate conversion*. Paper presented at proceedings of the IEEE International Symposium on Circuits and Systems.

Sarkar, G., Mallik, D., & Jatana, H. S. (2013, June). *Design of Multi-stage Cascaded Integrator Comb filter using single Adder and Subtractor*. Paper presented at International Conference on Microelectronics, Communication and Renewable Energy, Kanjirapally, Kerala. doi:10.1109/AICERA-ICMiCR.2013.6575959

Shaji, A., & Jayaraj, V. S. (2013, December). *FPGA implementation of an efficient cascaded integrator comb filter for narrow and wideband designs*. Paper presented at International Conference on Control Communication and Computing, Thiruvananthapuram. doi:10.1109/ICCC.2013.6731630

Shuni Chu, , & Burrus, C. Shuni Chu & Burrus. (1984). Multirate filter designs using comb filters. *IEEE Transactions on Circuits and Systems*, 31(11), 913–924. doi:10.1109/TCS.1984.1085447

Sugiura, Y. KAwamura, A., & Liguni, Y. (2012, December). *A comb filter with adaptive notch gain for periodic noise reduction*. Paper presented at Asia-Pacific Signal & Information Processing Association Annual Summit and Conference (APSIPA ASC), Holliwood, CA.

KEY TERMS AND DEFINITIONS

Adder: Is a component of a digital filter that produces at its output the sum of the values of the samples present at its inputs.

Complex Exponential Signal: A signal whose samples are complex numbers, where the real and imaginary parts of the samples form, respectively, a cosine wave and a sine wave, both with the same frequency.

Complex Function: A function whose values are complex numbers.

Complex Number: Is a number composed by the sum of a real number a and an imaginary number $i \times b$, where b is a real number and i is defined as the square root of -1 .

Computational Complexity: Is the number of arithmetic operations that a filter performs

to obtain an output sample. The computational complexity is proportional to both, the power consumption of the filter and the speed of operation. Thus, this complexity must be decreased in practical applications.

Delay: Is the retention of a sample of a digital signal during a sampling period T_s and the subsequent release of that sample at the next sampling period. This is done with a one-input one-output device called delay. In other words, for a signal $x(nT_s)$, its delayed version is $y(nT_s) = x(nT_s - T_s)$. In a block diagram, this element is represented as z^{-1} . A block of M delays is represented as z^{-M} .

Digital Filter: Is a system that receives a digital signal and returns another digital signal with modified characteristics.

Digital Frequency: Is a value that expresses how much of a cycle of a sinusoidal wave is represented by a sample. This value is denoted by f . For example, $f = 1/2$ means that 2 samples of a digital signal correspond to a sinusoidal cycle. If, instead of cycles we use radians, we obtain the angular digital frequency denoted as ω , which is equal to $2\pi f$.

Digital Signal: Is a sequence of numbers, so-called samples, where every sample lasts T_s seconds. The time T_s is called sampling period and the number produced in the n -th sampling period (i.e., after n times T_s seconds) is denoted as $x(nT_s)$. The values $x(nT_s)$ are constrained to belong to a finite set of possibilities (for example, in a 1-bit signal, these values are just 0 or 1). Additionally, it is usual to consider that $x(nT_s) = 0$ for negative values of n . For analysis purposes, it is usual to consider $T_s = 1$.

Frequency Response: Is the response of a Linear Time-Invariant (LTI) digital filter to a complex exponential signal with frequency $\omega = 2\pi f$. Generally, the frequency response is a complex function of ω . The frequency response is a periodic function of ω with period 2π . Thus, for analyzes it is necessary to describe the frequency response just in the interval of ω from $-\pi$ to π . When the samples of the impulse are real numbers, the frequency response is described just in the interval

of ω from 0 to π because of the symmetries of its magnitude and phase responses. The frequency response is usually denoted as $H(e^{j\omega})$.

Impulse Response: Is the response of a Linear Time-Invariant (LTI) digital filter to an input signal $x(nT_s)$ that has the value 1 for $n = 0$ and 0 for other values of n . The samples of the impulse response are usually denoted as $h(nT_s)$.

Magnitude Response: Is a function of the angular frequency ω where every value is obtained as the magnitude of the complex value of the frequency response in that frequency ω . If the value of the frequency response in ω is a complex number of the form $a(\omega) + i \times b(\omega)$, the magnitude of that number is given by $\{[a(\omega)]^2 + [b(\omega)]^2\}^{1/2}$. When the samples of the impulse are real numbers, the phase response is symmetric and hence it is described just in the interval of ω from 0 to π .

Multiplier: Is a component of a digital filter that takes the samples at its inputs and produces the multiplication of these values at its output. A multiplier is an expensive and power-consuming element, and therefore the efficient design of digital filters consists on reduce the number of required multipliers or even avoid them.

Phase Response: Is a function of the angular frequency ω where every value is obtained as the phase of the complex value of the frequency response in that frequency ω . If the value of the frequency response in ω is a complex number of the form $a(\omega) + i \times b(\omega)$, the phase of that number is given by $\arctan\{b(\omega)/a(\omega)\}$. When the samples of the impulse are real numbers, the phase response is anti-symmetric and hence it is described just in the interval of ω from 0 to π .

Scaling: Is the change of the range of values that a signal can take. For example, if a signal can take values from -1 to 1 and that signal is scaled by 5 , the scaled signal can take values from -5 to 5 .

Transfer Function: Is a mathematical representation of the relation of a Linear Time-Invariant (LTI) digital filter between its input and its output. The transfer function is expressed in terms of the variable z , which can take complex values. If the values for z are constrained to be complex numbers with unitary magnitude, the transfer function becomes the frequency response. The transfer function is usually denoted as $H(z)$.

Consumer Adoption of PC-Based/Mobile-Based Electronic Word-of-Mouth

M**Akinori Ono***Keio University, Japan***Mai Kikumori***Ritsumeikan University, Japan*

INTRODUCTION

Word of mouth (WOM) is one of the most fascinating topics for marketing scholars and practitioners (cf. Ono and Kikumori, 2015). WOM is defined as oral, person-to-person communication between a receiver and a communicator whom the receiver perceives as non-commercial, concerning a brand, a product or a service (Arndt, 1967, p.3). Previous studies on WOM have found that it can have a great impact on consumer attitude formation, purchase decisions, and even post-usage perceptions of a product or service (e.g., Katz & Lazarsfeld, 1955; Brown & Reingen, 1987; Herr, Kardes, & Kim, 1991; Bone, 1995).

Since the advent of the internet in the early 1990s, electronic word of mouth (eWOM) has become extremely important (Bickart & Schindler, 2001; Godes & Mayzlin, 2004). Consumers increasingly communicate product information with other consumers via new forms of communication, including e-mail, community sites, review sites, social networking sites, blogs, online discussion forums, and news groups (e.g., Goldsmith, 2006; Okazaki, 2008). Many scholars have found that eWOM messages about products and services can have a great impact on purchase decision-making (Cheung, Lee, & Rabjohn, 2008; Park & Kim, 2008; Schlosser, 2011; Sen & Lerman, 2007).

Moreover, with the rapid growth of mobile phone ownership, an increasing number of consumers can communicate with others and

send and receive product information through their mobile phones (cf. Shen, Wang, & Xiang, 2013). Communication via mobile phones in the form of e-mail, SMS, and text messages has the unique characteristic of ubiquitous connectivity, which enables consumers to exchange information anytime and anywhere (Lee, 2005; Okazaki, 2008). Therefore, mobile-based electronic word of mouth (mWOM) is now regarded as an important mobile marketing tool for offering consumers context-sensitive and time-critical recommendations (Okazaki, 2009).

Viral marketing campaigns, which are defined as distribution or communication that relies on consumers to transmit content via electronic communication techniques to other potential consumers in their social sphere and to animate these contacts to also transmit the content, have been a focus for pioneers of eWOM research (Wiedemann, 2007, p.53). Most research on PC-based viral marketing campaigns has qualitatively investigated how recipients of viral messages from marketers via friends and relatives adopt and send messages (Phelps, Lewis, Mobilio, Perry, & Raman, 2004). Several studies have also constructed and tested a causal model that explains why recipients adopt and send viral messages (Ho & Dempsey, 2008). In addition, some studies have focused on what determines consumers' intentions to participate in mobile viral marketing (Okazaki, 2008), and have tried to describe the entire consumer decision-making process for mobile viral marketing cam-

paings across multi stages (Wiedemann, Palka, & Pousttchi, 2008; Palka, Pousttchi, & Wiedemann, 2009; Pescher, 2014).

The determinants of eWOM adoption have also been investigated. Some eWOM studies have investigated what determines the intention to adopt eWOM messages on the Web (Cheung, Lee, & Rabjohn, 2008), whereas more recent studies have focused on how recipients of mWOM reviews from anonymous message senders evaluate and adopt the reviews (Shen, *et al.*, 2013; Wang, Shen, & Sun, 2013).

This chapter examines previous studies that have tried to answer why consumers adopt PC-based/mobile-based eWOM messages or participate in PC-based/mobile-based viral marketing campaigns. Our understanding can be enhanced by re-examining the findings of the current leading scholars.

BACKGROUND

Consumer Intention to Participate in PC-Based Viral Marketing

Viral Marketing is defined as marketing that ‘infects’ its customers with an advertising message, which passes from one customer to the next like a rampant flu virus (cf. Montgomery, 2001). Recently, marketing scholars have investigated various topics relating to viral marketing (e.g., Chiu, Hsieh, Kao, & Lee, 2007; Gangadharbatla & Lisa 2007; Hinz, Skiera, Barrot, & Becker, 2011; Huang, Lin, & Lin, 2009; Phelps, *et al.*, 2004). As pioneers in these studies, Dr. Phelps at the University of Alabama and his colleagues (Phelps, *et al.*, 2004) examined what the motives were that caused consumers to pass along the email marketing messages they had received. The results of in-depth interviews showed that the most common motivations were fun, including enjoyment and entertainment, and social connection, such as helping others and to communicate caring.

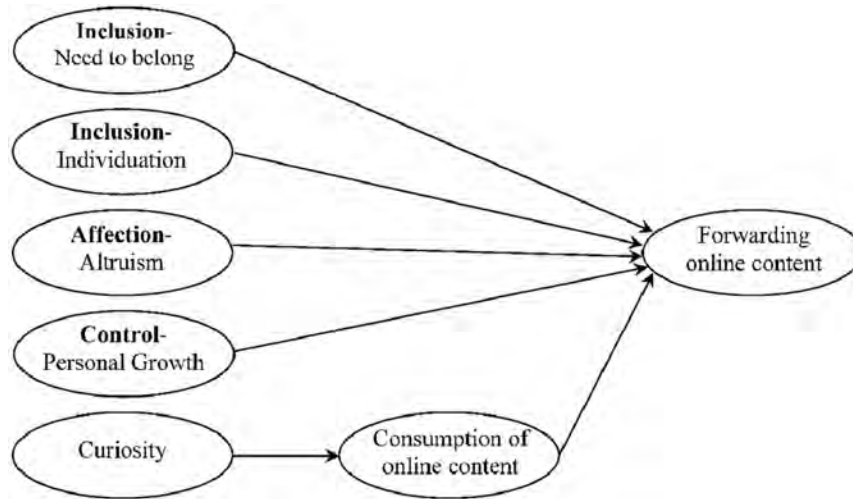
The earliest studies on viral marketing focused on the demographic, psychographic, and behavioral characteristics of mobile phone users engaged in viral marketing behavior (e.g., Chiu, Lee, & Chen, 2006; Dobeles, Lindgreen, Beverland, Vanhamme, & van Wijk, 2007; Phelps, *et al.*, 2004; Wiedemann, 2007; Wiedemann, *et al.*, 2008). As pioneers who addressed this gap, Dr. Ho at Simon Fraser University and Dr. Dempsey at Ryerson University (Ho & Dempsey, 2008) identified four motives involved in forwarding online content (the need to belong, individuation, altruism, and personal growth) based on the FIRO (fundamental interpersonal relations orientation) theory (Schutz, 1966), which identified three motives—inclusion, affection, and control. They proposed a causal model describing that these four motives, as well as the amount of online content consumed due to curiosity, affect the frequency of forwarding online content (Fig.1). The results of structural equation modeling showed that three of five direct motives—individuation, altruism, and the amount of online content consumed—had significant and positive impacts on the frequency of forwarding online content. However, the need to belong and curiosity were insignificant factors. Moreover, contrary to the proposed hypothesis, the impact of personal growth on the frequency of forwarding was negative.

Consumer Intention to Participate in Mobile Viral Marketing

The pioneering studies of consumer intention to participate in mobile viral marketing were conducted by Dr. Okazaki (Okazaki, 2008) at the Autonomous University of Madrid. He modeled causal relationships among user characteristics, perceived values, attitudes, intentions, and consumer behaviors in those who send mWOM messages and participate in mobile viral marketing. He modeled three exogenous variables (commitment to promoted brand, relationship with mobile devices, and group-person connectivity) as indirect determinants of the attitude-intention-behavior

Figure 1. Ho and Dempsey's model (2008)

Source: Ho & Dempsey (2008), p.1003



chain using two kinds of values (entertainment/hedonic and purposive/utilitarian) in the mobile viral marketing campaign (Fig.2). The proposed model was empirically tested using structural equation modeling, and the results showed that all paths are significant. Therefore, the proposed model was supported. In addition, Dr. Okazaki performed a multi-group analysis among users whose information sources were different from each other. The findings showed that only two

paths were statistically different between people who use e-mail magazines as an information source and those who pass-along e-mail. In contrast, all paths were statistically different between people who use pass-along e-mail as an information source and those who use direct referral.

Like Dr. Okazaki (2008), Dr. Yang and Dr. Zhou (Yang & Zhou, 2011) at Appalachian State University proposed a causal model describing the determinants of consumer intentions to participate

Figure 2. Okazaki's model (2008)

Source: Okazaki (2008), p.718.

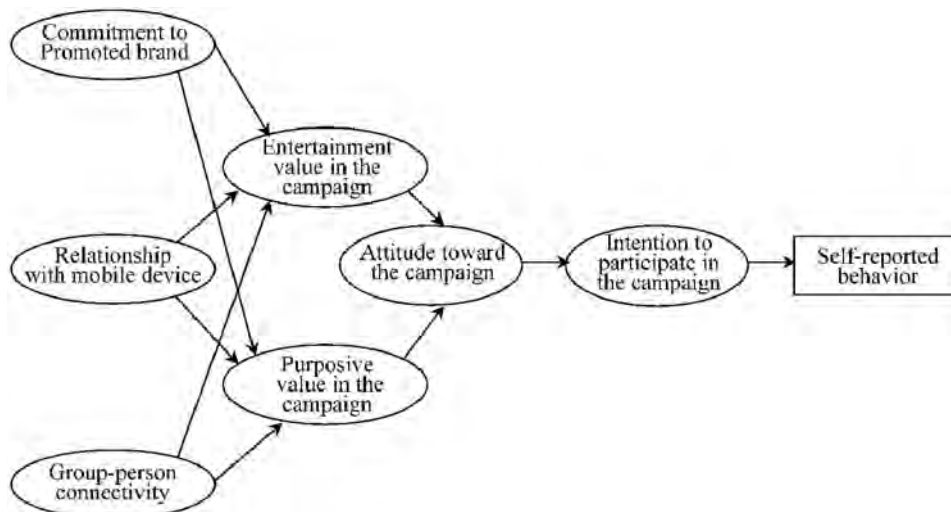
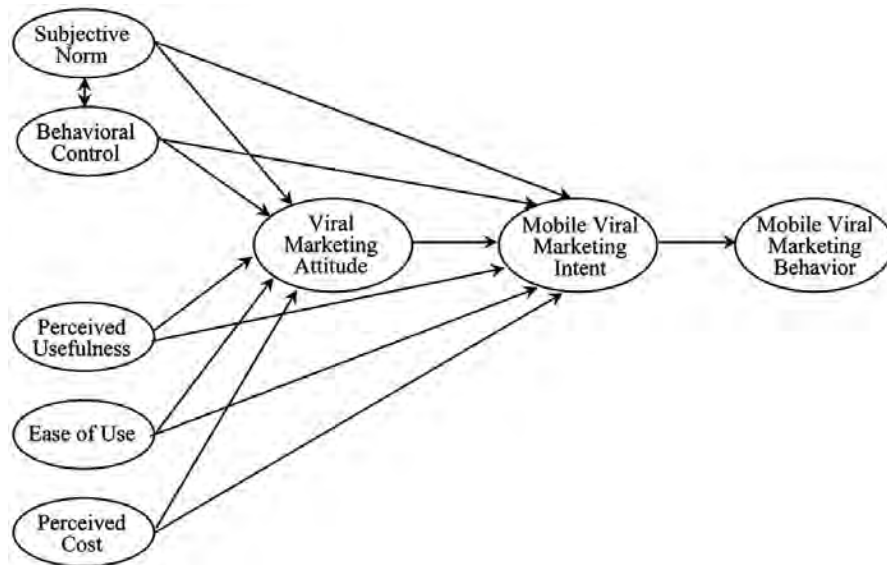


Figure 3. Yang & Zhou's model (2011)

Source: Yang & Zhou (2011), p.89.



in mobile viral marketing and their actual behavior with regard to forwarding mobile viral messages. Based on the theory of planned behavior (Ajzen, 1991), they modeled three determinants (attitude, subjective norm, and perceived behavioral control) of consumer intention toward forwarding mWOM messages and the frequency of forwarding mWOM messages in the context of mobile-based viral marketing. They also modeled three additional determinants—perceived usefulness, ease of use, and perceived cost (cf. Davis, 1989; Palka, *et al.*, 2009) (Fig.3). The results of structural equation modeling showed that three of the five proposed exogenous variables (subjective norm, behavioral control, and perceived cost) affected viral marketing attitude, intent, and behavior. In addition, perceived usefulness directly affected intent. Unfortunately, all other paths were insignificant.

Multi-Stage Consumer Referral Process in Mobile Viral Marketing

Dr. Wiedemann, Dr. Palka, and colleagues at the University of Augsburg proposed a customer funnel-type process model named the “basic model (or grounded theory) of mobile viral

marketing process” that described a consumer’s psychological flow. After a presentation at a conference (Wiedemann, *et al.*, 2008), they published a paper regarding their model (Palka, *et al.*, 2009). Whereas the researchers mentioned in the previous section constructed and tested causal models to explain why a consumer forwards viral marketing messages to friends and relatives, their studies focused on illustrating the entire consumer decision-making process, which consists of receiving, using, and forwarding a mobile viral marketing message (Fig.4). They developed not only the three-stage comprehensive process model but also detailed models. The receipt model, one of the three detailed models, was proposed to describe six elements of three kinds of conditions (perceived risk and trust as security-related conditions, sender recognition and perceptual affinity as social conditions, and self-efficacy and perceived cost as resource-based conditions) grounding a recipient’s intention to open the mobile viral marketing message (Fig.5).

Like Dr. Palka and colleagues, Dr. Pescher (Pescher, 2014) at Ludwig Maximilian University, Munich, modeled the entire consumer decision-making process for a mobile viral marketing

Figure 4. The basic model of a mobile viral marketing process

Source: Palka, et al. (2009), p.175

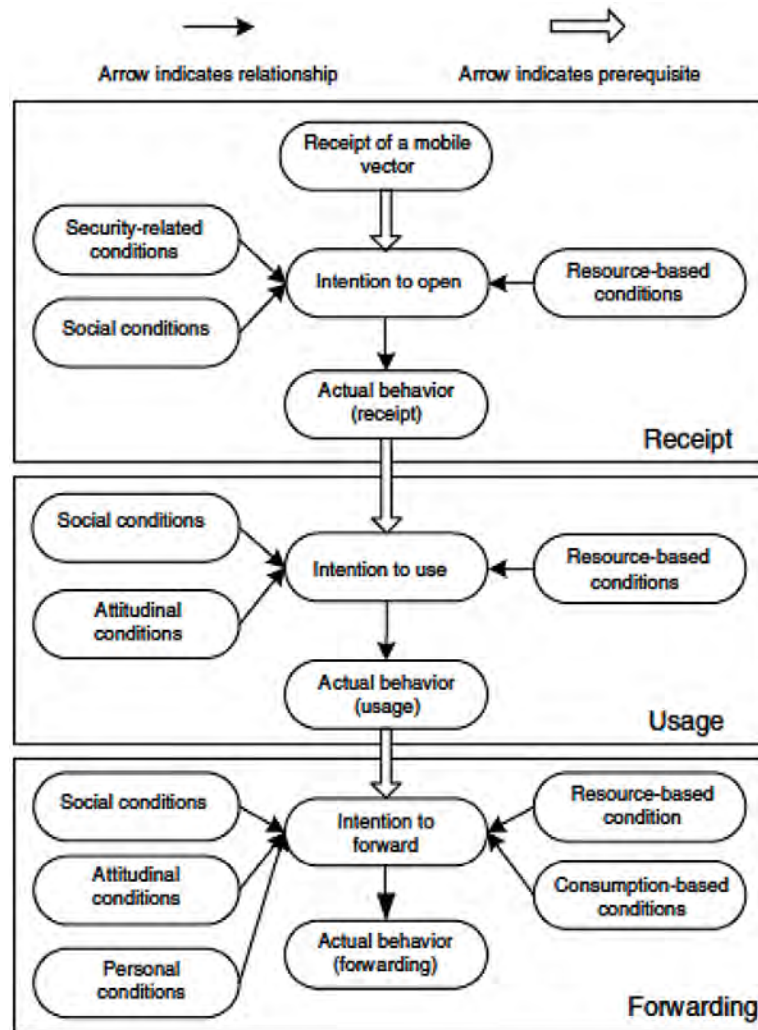


Figure 5. Receipt model

Source: Palka, et al. (2009), p.176

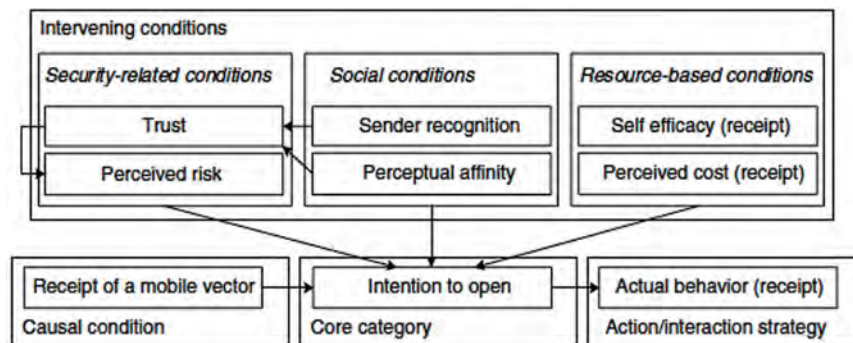
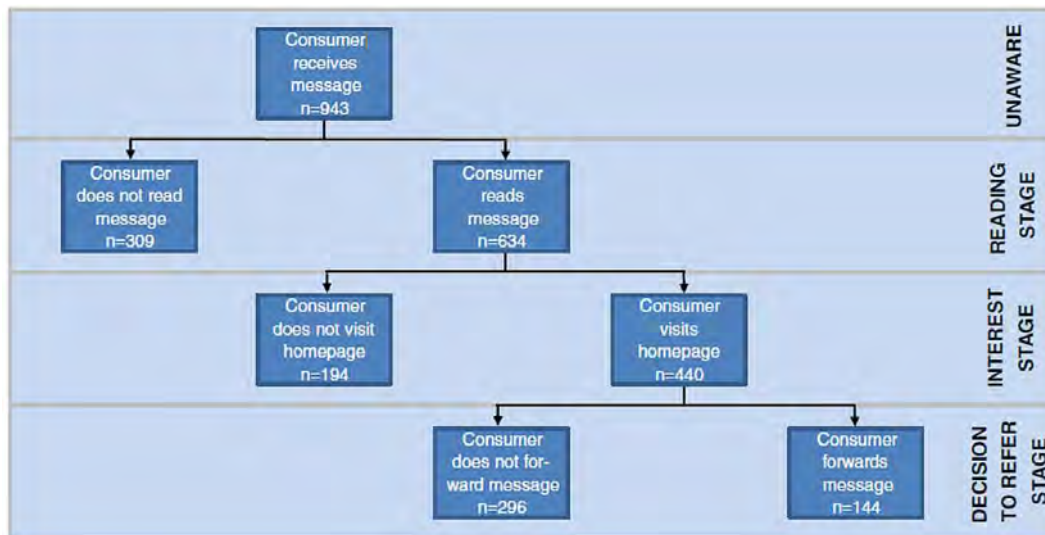


Figure 6. Pescher's process model (2014)

Source: Pescher (2014), p.49.



campaign via mobile phones across multiple stages (Fig.6). While Dr. Palka and colleagues modeled three stages of receipt, usage, and forwarding of mobile viral marketing messages, Dr. Pescher modeled reading, interest, and decision to refer, based on traditional consumer decision-making models (cf. Bettman, 1979). The results of sequential logit modeling showed that (1) the first stage (receipt and reading) was determined by entertainment value (Okazaki, 2008) and tie strength (Granovetter, 1973); (2) the second stage (interest and use) was determined by entertainment value and purposive value (Okazaki, 2008); and (3) the final stage (intention to forward) was determined by entertainment value, purposive value, tie strength, and usage intensity, which referred to the number of text messages sent via mobile phones per day (Steenkamp & Gielens, 2003).

CONCLUSION

Consumer Intention to Adopt eWOM Messages

Dr. Cheung at Hong Kong Baptist University and colleagues (Cheung, *et al.*, 2008) examined

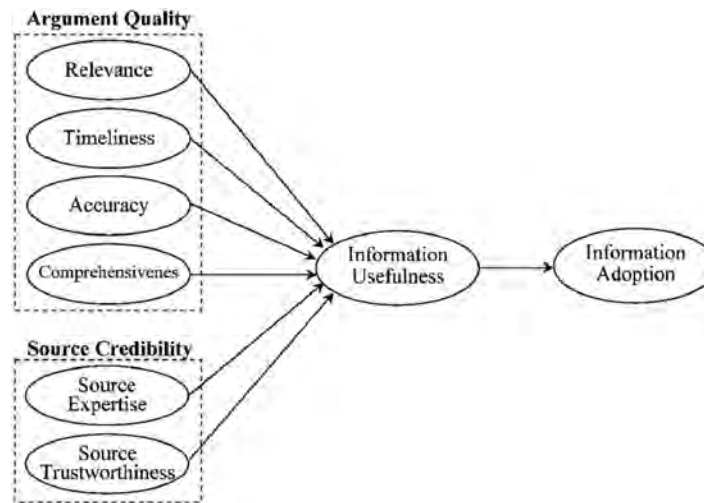
factors that determined consumer willingness to accept and adopt eWOM messages within online communities. Using dual-process theories (Sussman and Siegal, 2003; Bhattacharjee and Sanford 2006), they proposed a causal model describing four factors of argument quality and two factors of source credibility that affect perceived usefulness and, in turn, adoption of eWOM messages within online communities (Fig.7). The results of Partial Least Squares modelling showed that two of four argument quality factors (relevance and comprehensiveness) had positive impacts on information usefulness, which in turn positively affected information adoption. Unfortunately, all other paths were not statistically significant.

Consumer Intention to Adopt mWOM Applications and Messages

Several recent studies have focused on how recipients of mWOM reviews from anonymous message senders evaluate and adopt the reviews. Of these studies, Dr. Shen at Wuhan University, Dr. Wang at the University of Science and Technology of China, and a colleague (Shen, *et al.*, 2013) examined the effects of the mobile device system and information characteristics on consumer perceptions

Figure 7. Cheung, et al.'s model (2008)

Source: Cheung, et al. (2008), p.233



of ubiquitous decision support systems. Based on the IS success model (DeLone & McLean, 1992), the technology acceptance model (Davis, 1989), the information adoption model (Davis, 1989), and the human-computer interaction literature (Li & Yeh, 2010), Dr. Shen and colleagues proposed a causal model to identify factors influencing consumer intentions to use mWOM applications and adopt mWOM reviews (Fig.8).

The results of structural equation modeling showed that two explained variables—consumer

intentions to use the mobile review application (system adoption) and adopt mWOM reviews (information adoption)—were directly determined by system and information usefulness. System usefulness had a stronger impact on system adoption, whereas information usefulness had a stronger impact on information adoption. The results also showed that three out of five system characteristics (wireless network stability, physical appearance of mobile devices (Yun, Lee, Kim, & Kettinger, 2011), and design aesthetics of the

Figure 8. Wang, et al.'s model (2013)

Source: Wang, et al. (2013), p.1397.

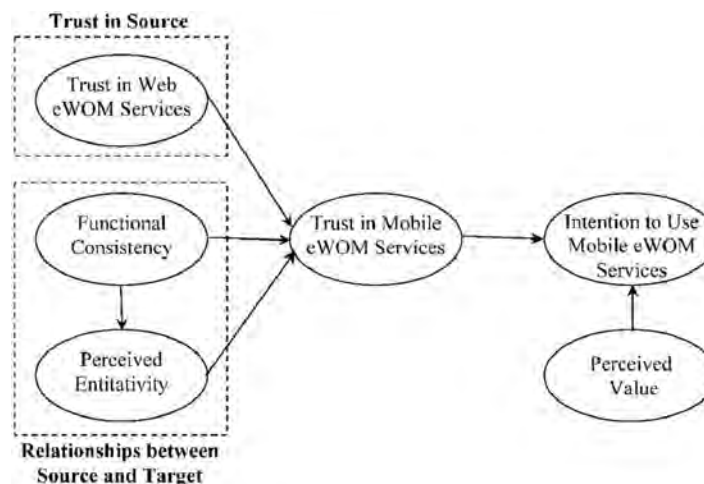
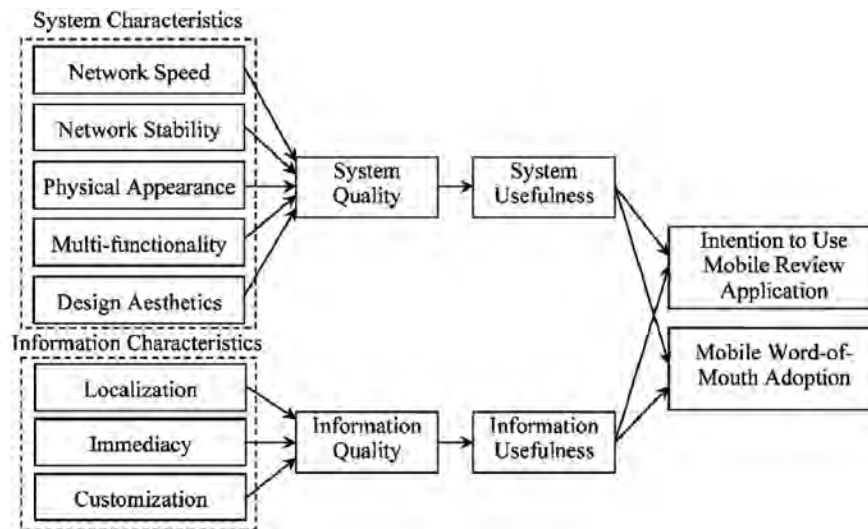


Figure 9. Shen, et al.'s model (2013)

Source: Shen, et al. (2013), p.45.



mobile application (Cyr, Head, & Ivanov, 2006)) significantly predicted system quality, which in turn determined system usefulness. In contrast, all three information characteristics (localization, immediacy, and customization of the information obtained from the system) were major predictors of information quality, which in turn determined information usefulness.

In another article, Dr. Shen, Dr. Wang, and a colleague (Wang, *et al.*, 2013) investigated the process of trust transfer between PC- and mobile-based eWOM services and its role as a determinant of intention to use the services. From the trust transfer perspective, they proposed a causal model explaining that the initial trust in PC-based eWOM services as well as two factors regarding the relationships between the source and target of trust transfer (functional consistency and perceived entitativity) affect trust in mWOM services, which in turn predicts user intentions to use the mWOM services (Fig.9). They also added the perceived value of mWOM services as a control variable to their model.

The results of structural equation modeling showed that consumer trust in PC-based eWOM services had a significant, positive impact on consumer trust in mobile-based eWOM services,

which in turn affected consumer intention to use the mobile-based eWOM services. These results suggest that consumer trust in PC-based eWOM services can be transferred to mobile-based eWOM services. The results also showed that functional consistency between PC-based and mobile-based eWOM services and perceived entitativity had a significant, positive impact on consumer trust in mobile-based eWOM services. Moreover, functional consistency enhances perceived entitativity. When users perceive that the relationship between two kinds of eWOM services is strong, they are more likely to trust in mWOM services, given their initial trust in PC-based eWOM services.

FUTURE RESEARCH DIRECTIONS

There is still room for future research on consumer adoption of eWOM/mWOM messages. First, previous studies have not compared eWOM to mWOM. Most studies have not even referred to possible differences between eWOM and mWOM. Future research should examine the influences of receiving a message on a desktop PC vs. a mobile device on consumer adoption of the message.

Second, research methods should be re-examined. For example, most previous studies on mWOM conducted PC-based surveys, not mobile-based surveys, to collect data. Furthermore, both survey sampling design and experimental design should be used to conduct empirical tests. In addition, most previous studies have investigated a single product category. Future research should consider differences among products in terms of consumer involvement and knowledge. Moreover, researchers should take into account that this is just the beginning of internet marketing. The findings of the previous research may change with the increasing prevalence of eWOM and the related marketing activities.

Finally, research on consumer adoption of eWOM should be expanded to include the choice of brand/store that is adopted. Several studies have already begun investigating the impact of adopted eWOM message characteristics, such as the ratio of positive to negative messages, on consumer attitudes toward the brand (cf. Kikumori & Ono, 2013). However, the impact of mWOM characteristics on consumer choice has not yet been sufficiently investigated. Future research should be conducted to determine the impact of mWOM as well as the differences between eWOM and mWOM (cf. Ono, Nakamura, Okuno, & Sumikawa, 2012).

REFERENCES

- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179–211. doi:10.1016/0749-5978(91)90020-T
- Arndt, J. (1967). *Word of mouth advertising: A review of the literature*. New York, NY: Advertising Research Foundation.
- Bettman, J. R. (1979). *An information processing theory of consumer choice*. Reading, MA: Addison-Wesley.
- Bhattacharjee, A., & Sanford, C. (2006). Influence processes for information technology acceptance: An elaboration likelihood model. *Management Information Systems Quarterly*, 30(4), 805–825.
- Bickart, B., & Schindler, R. M. (2001). Internet forums as influential sources of consumer information. *Journal of Interactive Marketing*, 15(2), 31–40. doi:10.1002/dir.1014
- Bone, P. F. (1995). Word-of-mouth effects on short-term and long-term product judgment. *Journal of Business Research*, 32(3), 213–223. doi:10.1016/0148-2963(94)00047-I
- Brown, J. J., & Reingen, P. H. (1987). Social ties and word-of-mouth referral behavior. *The Journal of Consumer Research*, 14(3), 350–362. doi:10.1086/209118
- Cheung, C. M. K., Lee, M. K. O., & Rabjohn, N. (2008). The impact of electronic word-of-mouth: The adoption of online opinions in online customer communities. *Internet Research*, 18(3), 229–247. doi:10.1108/10662240810883290
- Chiu, H. C., Hsieh, Y. C., Kao, Y. H., & Lee, M. (2007). The determinants of email receivers disseminating behaviors on the Internet. *Journal of Advertising Research*, 47(4), 524–534. doi:10.2501/S0021849907070547
- Chiu, H. C., Lee, M., & Chen, J. R. (2006). Viral marketing: A study of e-mail spreading behavior across gender. *Journal of Website Promotion*, 2(3/4), 17–30.
- Cyr, D., Head, M., & Ivanov, A. (2006). Design aesthetics leading to m-loyalty in mobile commerce. *Information & Management*, 43(8), 950–963. doi:10.1016/j.im.2006.08.009
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *Management Information Systems Quarterly*, 13(3), 319–340. doi:10.2307/249008

- DeLone, W. H., & McLean, E. R. (1992). Information systems success: The quest for the dependent variable. *Information Systems Research*, 3(1), 60–95. doi:10.1287/isre.3.1.60
- Dobele, A., Lindgreen, A., Beverland, M., Vanhamme, J., & van Wijk, R. (2007). Why pass on viral messages? Because they connect emotionally. *Business Horizons*, 50(4), 291–304. doi:10.1016/j.bushor.2007.01.004
- Gangadharbatla, H., & Lisa, J. (2007). EWOM: The effect of individual level factors on viral consumers' email pass along behavior. Paper presented at the American Academy of Advertising Conference, Burlington, VT.
- Godes, D., & Mayzlin, D. (2004). Using online conversations to study word of mouth communication. *Marketing Science*, 23(4), 545–560. doi:10.1287/mksc.1040.0071
- Goldsmith, R. E. (2006). Electronic word-of-mouth. In M. Khosrow-Pour (Ed.), *Encyclopedia of e-commerce, e-government and mobile marketing* (pp. 408–412). Hershey, PA: Idea Group Publishing. doi:10.4018/978-1-59140-799-7.ch067
- Granovetter, M. (1973). The strength of weak ties. *American Journal of Sociology*, 78(6), 1360–1380. doi:10.1086/225469
- Herr, P. M., Kardes, F. R., & Kim, J. (1991). Effects of word-of-mouth and product attribute information on persuasion: An accessibility-diagnostics perspective. *The Journal of Consumer Research*, 17(4), 454–462. doi:10.1086/208570
- Hinz, O., Skiera, B., Barrot, C., & Becker, J. (2011). Seeding strategies for viral marketing: An empirical comparison. *Journal of Marketing*, 75(6), 55–71. doi:10.1509/jm.10.0088
- Ho, J. Y. C., & Dempsey, M. (2008). Viral marketing: Motivations to forward online content. *Journal of Business Research*, 63(9/10), 1000–1006.
- Huang, C. C., Lin, T. C., & Lin, K. J. (2009). Factors affecting pass-along email intentions (PAEIs): Integrating the social capital and social cognition theories. *Electronic Commerce Research and Applications*, 8(3), 160–169. doi:10.1016/j.elerap.2008.11.001
- Katz, E., & Lazarsfeld, P. F. (1955). *Personal influence: The part played by people in the flow of mass communications*. New York, NY: Free Press.
- Kikumori, M., & Ono, A. (2013). *Paradoxical relationship between the amount of negative eWOM messages and positive consumer attitude*. Keio/Kyoto Joint Global COE Discussion Paper Series.
- Lee, T. (2005). The impact of perceptions of interactivity on customer trust and transaction intentions in mobile commerce. *Journal of Electronic Commerce Research*, 6(3), 165–180.
- Li, Y. M., & Yeh, Y. S. (2010). Increasing trust in mobile commerce through design aesthetics. *Computers in Human Behavior*, 26(4), 673–684. doi:10.1016/j.chb.2010.01.004
- Montgomery, A. L. (2001). Applying quantitative marketing techniques to the Internet. *Interfaces*, 31(2), 90–108. doi:10.1287/inte.31.2.90.10630
- Okazaki, S. (2008). Determinant factors of mobile-based word-of-mouth campaign referral among Japanese adolescents. *Psychology and Marketing*, 25(8), 714–731. doi:10.1002/mar.20235
- Okazaki, S. (2009). Social influence model and electronic word of mouth: PC versus mobile internet. *International Journal of Advertising*, 28(3), 439–472. doi:10.2501/S0265048709200692
- Ono, A., & Kikumori, M. (2015). Consumer adoption of mobile e-WOM messages. In *Encyclopedia of Mobile Phone Behavior*. IGI Global.
- Ono, A., Nakamura, A., Okuno, A., & Sumikawa, M. (2012). Consumer motivations in browsing online stores with mobile devices. *International Journal of Electronic Commerce*, 16(4), 153–177. doi:10.2753/JEC1086-4415160406

- Palka, W., Pousttchi, K., & Wiedemann, D. G. (2009). Mobile word-of-mouth: A grounded theory of mobile viral marketing. *Journal of Information Technology*, 24(2), 172–185. doi:10.1057/jit.2008.37
- Park, D. H., & Kim, S. (2008). The effects of consumer knowledge on message processing of electronic word-of mouth via online consumer reviews. *Electronic Commerce Research and Applications*, 7(4), 399–410. doi:10.1016/j.elerap.2007.12.001
- Pescher, C., Reichart, P., & Spann, M. (2014). Consumer decision-making processes in mobile viral marketing campaigns. *Journal of Interactive Marketing*, 28(1), 43–54. doi:10.1016/j.intmar.2013.08.001
- Phelps, J. E., Lewis, R., Mobilio, L., Perry, D., & Raman, N. (2004). Viral marketing or electronic word-of-mouth advertising: Examining consumer responses and motivations to pass along email. *Journal of Advertising Research*, 44(4), 333–348. doi:10.1017/S0021849904040371
- Schlosser, A. E. (2011). Can including pros and cons increase the helpfulness and persuasiveness of online reviews? The interactive effects of ratings and arguments. *Journal of Consumer Psychology*, 21(3), 226–239. doi:10.1016/j.jcps.2011.04.002
- Schutz, W. (1966). *FIRO: A Three Dimensional Theory of Interpersonal Behavior*. New York, NY: Holt, Rinehart, & Winston.
- Sen, S., & Lerman, D. (2007). Why are you telling me this? An examination into negative consumer reviews on the web. *Journal of Interactive Marketing*, 21(4), 76–94. doi:10.1002/dir.20090
- Shen, X., Wang, N., & Xiang, L. (2013). Unleash the power of mobile word-of-mouth: An empirical study of system and information characteristics in ubiquitous decision making. *Online Information Review*, 37(1), 42–60. doi:10.1108/14684521311311621
- Steenkamp, J. E. M., & Gielens, K. (2003). Consumer and market drivers of the trial probability of new consumer packaged goods. *The Journal of Consumer Research*, 30(3), 368–384. doi:10.1086/378615
- Stewart, K. J. (2003). Trust transfer on the world wide web. *Organization Science*, 14(1), 5–17. doi:10.1287/orsc.14.1.5.12810
- Sussman, S. W., & Siegal, W. S. (2003). Informational influence in organizations: An integrated approach to knowledge adoption. *Information Systems Research*, 14(1), 47–65. doi:10.1287/isre.14.1.47.14767
- Wang, N., Shen, X., & Sun, Y. (2013). Transition of electronic word-of-mouth services from web to mobile context: A trust transfer perspective. *Decision Support Systems*, 54(3), 1394–1403. doi:10.1016/j.dss.2012.12.015
- Wiedemann, D. G. (2007). Exploring the concept of mobile viral marketing through case study research. In B. König-Ries, F. Lehner, R. Malaka, & C. Türker (Eds.) *Proceedings of the 2nd Conference on Mobility and Mobile Information Systems* (pp.49-60). Bonn: Gesellschaft für Informatik.
- Wiedemann, D. G., Haunstetter, T., & Pousttchi, K. (2008). Analyzing the basic elements of mobile viral marketing-an empirical study. *Proceedings of the 7th International Conference on Mobile Business* (pp.75-85). Los Alamitos, CA: IEEE Computer Society. doi:10.1109/ICMB.2008.41
- Wiedemann, D. G., Palka, W., & Pousttchi, K. (2008). Understanding the determinants of mobile viral effects-towards a grounded theory of mobile viral marketing. *Proceedings of the 7th International Conference on Mobile Business* (pp. 323-333). Los Alamitos, CA: IEEE Computer Society. doi:10.1109/ICMB.2008.42
- Wright, A. A., & Lynch, J. G. Jr. (1995). Communication effects of advertising versus direct experience when both search and experience attributes are present. *The Journal of Consumer Research*, 21(4), 708–718. doi:10.1086/209429

Yang, H., & Zhou, L. (2011). Extending TPB and TAM to mobile viral marketing: An exploratory study on American young consumers mobile viral marketing attitude, intent and behavior. *Journal of Targeting, Measurement and Analysis for Marketing*, 19(2), 85–96. doi:10.1057/jt.2011.11

Yun, H., Lee, C. C., Kim, B. G., & Kettinger, W. J. (2011). What determines actual use of mobile web browsing services? A contextual study in Korea. *Communications of the Association for Information Systems*, 28(1), 313–328.

KEY TERMS AND DEFINITIONS

Word-of-Mouth (WOM): Any positive, neutral, or negative messages sent by potential, actual, or former customers about a product, company, or users, which is made traditionally in oral form or, in a broader sense, in written form via the Internet.

Electronic Word-of-Mouth (eWOM): WOM sent via the Internet, which is made available to

an unspecified number of people by using online review platforms, or to particular persons by using e-mail, short mail services (SMS), or social networking sites (SNS),

Mobile Electronic Word-of-Mouth (mWOM): eWOM sent and/or received using mobile devices, such as cellphones, smartphones, and tablet computers.

Consumer Intention to Adopt eWOM/mWOM Messages: The degree to which consumers intend to closely follow eWOM/mWOM messages and be motivated to take action.

Mobile Viral Marketing: The transmission of mobile advertising in the form of consumer referral of promoted content.

Consumer Referral Behavior: The processes by which consumers send WOM messages related to a viral marketing campaign to friends or acquaintances.

Mobile-Based eWOM Services: Online review platforms provided by a firm in the mobile context where consumers can exchange eWOM.

Context-Aware Personalization for Mobile Services

M

Abayomi Moradeyo Otebolaku
Liverpool John Moores University, UK

Maria Teresa Andrade
University of Porto, Portugal

INTRODUCTION

The last decade has witnessed an unprecedented proliferation of multimedia-enabled mobile devices and an escalation of online multimedia content. In fact, consumption of audiovisual content in diverse mobile platforms has risen exponentially and this trend is expected to continue in the next few years as mobile devices become more and more sophisticated. According to Gartner, a world leading information technology research and advisory company, *Smartphone* sales represented two-thirds of global *mobile phone* market in 2014 with over 1.2 billion units sold (Gartner, 2015). Thus, it appears, smartphones are gradually replacing our desktops as they increasingly become cheaper and more powerful (in terms of processing capability, network connectivity) with excellent multimedia processing support (Flora, 2010; Ricci, 2011). This development indicates global penetration and increasing acceptance of smartphones as the primary platform for information access and processing.

As mobile users go about their daily routines, they continuously browse the Web, seeking interesting content to consume, and occasionally also uploading their personal content. However, users encounter huge volume of content, which often does not match their preferences resulting in what we call *mobile information overload*. To address this problem, user preferences are usually learned to predict relevant content. However, user preferences are dynamic, changing as users move from

one place to another, performing different activities. Therefore, it is important to mine, learn and understand contexts in which users perform such activities. This contextual information can be used to filter, customize and deliver interesting content, the process generally referred to as context-aware personalization. Its main goal is to assist users to overcome mobile information overload by selecting from overwhelming set of potential choices, services, etc. that match user's context dependent content to target users, tailoring such services to user's tastes, network and device characteristics. Consequently, users can focus more on important activities, minimizing distractions and time, while consuming multimedia services.

Context-aware media personalization (CAMP) has been the focus of researchers addressing mobile multimedia information overload problem in the last years (Ricci, 2011; Otebolaku & Andrade, 2015a; Lee et. al, 2015). CAMP assists users to select content among a deluge of alternatives by considering users' preferences and contexts to improve their consumption experience. Existing solutions of this kind, however, are limited to static preferences. The traditional personalization solutions only consider user preferences without contexts such as location, activities, etc. Static solutions are not effective in mobile environments as users are increasingly mobile, moving from one location to another, engaging in diverse activities. For example, the type of content a user would consume at home would be, in most cases, different from those they would consume in the office

or at airports. Even those solutions that consider contexts have relied on explicit *contextual information*. For example, they usually ask users to provide their current contexts. Systems relying on this kind of static context information have not been able to address information overload problem. Thus, our contribution to CAMP is *Mobile Context-Aware Media Personalization* (MobCAMP), which we define as a special type of personalization that utilizes user's contexts and activities to select and adapt media content according to user's tastes and contextual situations. Recommendation algorithms currently are the most popular techniques for realizing personalization, whilst context-aware recommendations deal with providing suggestions to users when and where content are needed, context-aware personalization can be used to provide such suggestions when, where and how such content is required.

MobCAMP is proposed as an extension of our previous solution (Otebolaku and Andrade, 2015b), built on a number of existing technologies. First, rapid development in the field of mobile and telecommunication networks enabled ubiquitous communications whereby smartphone users can connect to the Web and consume services anywhere, anytime. With this development, users can access content such as news, music, movies, etc. at their convenience. Second, mobile devices now come with cheap, built-in sensors, enabling ubiquitous context sensing (Otebolaku & Andrade, 2016; Kwapisz et. al., 2010). Sensors such as thermometers, accelerometer (sensing movement), and GPS sensor (sensing location), etc. now ship with smartphones. Third, context-awareness has enabled the ability to deliver personalized information based on user's contextual situations. Information such as location, activity, time, weather, etc. can now be obtained readily in real-time from smartphones. Fourth, traditional recommendation systems have matured, and are helping users to find relevant information (Adomavicius et. al., 2005). These solutions can be explored to realize context-aware mobile multimedia personalization. Thus, MobCAMP builds on these core solutions,

using mobile user's preferences to suggest useful and interesting content, tailored to users' contextual situations.

Let us consider a scenario to illustrate this concept.

Carisa is an Electrical Engineering student of the University of Porto. She lives at the city outskirts, and catches a train every morning at 7:30 AM to the campus in time to attend 9:00 AM lecture. Before leaving home, while taking her breakfast, Carisa normally checks weather forecast. On her way walking to the train station she listens to country music, whereas on the train Carisa prefers to read breaking news. From the train stop, while walking, to the campus she profits from listening to her favorite music. However, when she is in the classroom she switches her mobile phone to silent mode. Then, on her way back home in the evening, she also enjoys listening to music except on Fridays when she eagerly seeks for good movies playing in the theaters near her home or in downtown Porto provided the weather forecast indicates a dry weekend.

In the presented scenario, it is important to capture *Carisa's* activities such as walking, time, 9:00 AM, preferably at a higher level, e.g. morning. *Carisa's* location should also be inferred from GPS or Wi-Fi. MobCAMP is designed to obtain this information dynamically and use it to suggest and customize relevant content to *Carisa*.

This chapter presents MobCAMP, a system for suggesting Web-based multimedia content to mobile users based on context recognition, contextual user profiling, and adaptive context-aware recommendation processes.

BACKGROUND

Traditional personalized recommendation systems aim at guiding users to the most relevant items but they do not take contextual information into account (Adomavicius et. al., 2005). Personalized

recommendation systems can build, manage, and represent information, customized to a particular user. Personalization grew mainly out of information retrieval system community to minimize information overload and to retain customers by anticipating their content needs. As succinctly put by Burke in (Burke, 2002), the criteria *individualizing*, *interesting*, and *useful* information distinguish personalized recommendation systems from retrieval systems. If we remove personalization from recommendation processes, we are left with only a set of popular items, which are independent of users' desired tastes. Therefore, on the one hand, in information retrieval systems such as search engines, the results that are relevant to a query usually do not change irrespective of the individual who issues such query or where and when it is issued. On the other hand, in traditional recommendation systems, the results of a query are usually a set of popular items.

The traditional personalized recommendation techniques can be broadly categorized into three (Adomavicius & Tuzhilin, 2005): 1) based on the opinions and preferences of other users, designated as Collaborative Filtering (CF) (Schafer et. al., 2007); 2) based on user's consumption history and the descriptions of available candidate items, referred to as Content Based Recommendation (CBF) (Pizzani & Billsus, 1997); and 3) based on a combination of CF and CBF denoted as Hybrid Recommenders (HR).

Personalized collaborative systems traditionally suggest items to target users based on how similar users have rated the same items. Formally, rating $R(u, i)$ is predicted based on rating $R(u', i)$ given to the same item i by user u' who is similar to user u (Adomavicius & Tuzhilin, 2005). CF is broadly categorized into 2: Memory-based and model-based collaborative systems. The former uses heuristics to make predictions based on the entire collection of previously rated items, whereas the latter uses collection of previously rated items to learn a model to provide recommendations for users. Because they load the entire dataset into system's memory for processing, memory-based

CF systems are computationally expensive, especially when a large number of users or items are involved. On the other hand, model-based CF requires less computational resources, in addition to having shorter response time. Memory-based CF, nevertheless, produces more accurate predictions than a model-based CF (Adomavicius & Tuzhilin, 2005). A good example of CF system is GroupLens (Resnick et. al., 1994).

Personalized content-based systems utilize user's preferences to analyze candidate items to identify similarity between a target user and candidate items (Adomavicius & Tuzhilin, 2005). Only items with a high degree of similarity to the target user's profile are recommended. A user profile describes preferences of the user, based on the history of the user's actions (Alberto, 2010). However, one major drawback of CBF is that of over specialization (Adomavicius & Tuzhilin, 2005). This means that a target user can be presented only with recommendations that are similar to those items he has rated in the past. From user's perspective, this method provides boring or stale recommendations. Users always want to try out new items based on their present environment and situations. Pizzani & Billsus (2007) presents a detailed survey of CBF systems.

Personalized hybrid recommendation systems use a combination of item ratings provided by user and ratings of items consumed by her so-called neighbors or friends in the past to compute recommendations. Combining CF and CBF techniques, HR was developed to address the weaknesses of both CBF and CF, such as over specialization problem of CBF and the new user problem of CF (Adomavicius & Tuzhilin, 2005). However, these personalization techniques' effectiveness has been limited by weaknesses (e.g. absence of rating information, and lack of contextual information) of the constituent techniques it incorporates (Yu et. al, 2007). Good examples of hybrid recommender systems can be found in a survey conducted by Burke R. (2002).

Recently, new approaches incorporating contextual information have emerged to address

weaknesses of traditional recommendation techniques (Adomavicius et. al., 2005). In traditional recommendation systems, *user preferences* are considered as static information. This assumption means that a user who likes a specific item in the past will always like the same item or similar items in the future. This assumption does not hold, especially in a mobile environment where user preferences change with contextual situations (Said et al., 2011). For example, a user who likes watching movies when at home every weekend, would most likely not indulge in watching movies when in the office. She would rather probably prefer to listen to music. Therefore, it is important to capture, in real-time, this contextual information and relate it to the preferences of users to suggest interesting content, which are relevant to user preferences in context.

This problem has been explored extensively by many researchers in recent time. Adomavicius et al. (2005) pioneered and proposed the incorporation of contextual information into recommender systems, using a multidimensional approach. Additionally, Chen (2005) proposed a context-aware collaborative recommendation technique that computes similarity between context where a target user has rated an item and his present context to recommend new items. It looks at what like-minded users have done in the past under similar contexts (e.g., location, weather) to suggest interesting items. Nevertheless, generating similarity between contexts can be unrealistic in some application domain, especially when ratings of items consumed in such contexts are not available (Vallet et al., 2007). Another important part of the system is user profiling, which summarizes consumption history of a user and her preferences. Generally, it refers to a collection of long or short-term data associated with a specific user or a group of users (Alberto, 2010). Existing user profiling approaches do not consider the relationship between consumption contexts and user preferences. Therefore, they do not support contextual recommendations. Said et. al. (2011) proposed a solution incorporating contexts into

user profile to describe user preferences according to the contextual situations where such preferences are expressed. This solution is one of the latest attempts to address this problem.

Most of those solutions, however, require users to regularly provide their contextual information, with an assumption that such contextual information is readily available. However, this is not a realistic solution in practice. For example, it is not convenient to ask a user to supply information such as “I am walking near home, wanting to watch a drama movie alone”. This puts extra burden on users, causing distraction and ultimately ineffectiveness. Users would prefer such system to automatically infer their contextual situations to suggest relevant content, using notifications without distraction. Additionally, existing user profile models use 2 dimensional (2D) and multi-dimensional (MD) approaches to relate user preferences to consumption history (Adomavicius et. al., 2005; Adomavicius & Kwon, 2007; Yu et. al., 2006). The difference between the two is that multi-dimensional approach uses multiple dimensions (User $\times$ Ratings $\times$ Contexts $\times$ Relevance) to relate user preferences and their consumption history, whereas a 2D approach uses only information about users and ratings (User $\times$ Ratings). Example of MD recommendation systems is presented by Gho et al. (2011) where location, device screens, network bandwidths, activity, mood, time, gender, language, etc., are used to deliver relevant multimedia contents to mobile users. A good example of 2D profiling approach is presented by Sarwar et. al. (2001), which uses only user information and item ratings to recommend items to users.

Unlike 2D-based user profiles that do not provide means to accommodate additional dimensions, such as contextual information, MD-based user profiling technique addresses this weakness. Nevertheless, current MD-based profiles rely on user ratings. In situations where ratings are not provided, this approach will suffer from poor recommendation quality. A solution that renders recommendations to mobile users with or without

ratings by incorporating contextual information into the user profile model is therefore desirable. Such profiles should be flexible enough to allow incorporation of contextual information into existing recommendation algorithms. We propose a contextual user profiling process to address the stated problems of existing user profile techniques. We have also proposed the extension of the traditional CBF, CF, and hybrid recommendation systems to produce context-aware CBF, context-aware CF and context-aware hybrid recommendation algorithms based on this new contextual user profile.

CONTEXT-AWARE MEDIA PERSONALIZATION FOR SMARTPHONE USERS

Assisting mobile users to make multimedia consumption choices is very crucial to the goal of anytime, anywhere information access. We designed MobCAMP by extending existing traditional personalized recommendation algorithms to suggest content, using contextual information. MobCAMP relies on a contextual user profile model, and a suit of context-aware recommendation services, namely collaborative, content and hybrid recommendations. The basic working concept of MobCAMP involves computing and tailoring personalized recommendations based on the contextual situations in which a target user or his friends have preferred certain content in the past. To achieve this, MobCAMP implements some basic components namely, context recognition model, contextual user profile model, multimedia profiling and context-aware recommendation algorithms.

Context Recognition

To capture contexts dynamically without asking users to manually provide them, we built a context recognition process (Otebolaku & Andrade, 2016), which can be ported on smartphones running so-

phisticated operating systems, such as Android. The recognition process uses events from the smartphone built-in sensors, such as accelerometer, orientation, and rotation sensors (to recognize user movement), GPS/WiFi (to recognize the user location), and the device clock (to capture time information). Generally, data produced by these sensors are low-level events that cannot be used directly by practical mobile applications (Kwapisz et. al., 2010). Therefore, MobCAMP relies on the context recognition process on mobile user's device to infer *high-level contexts* from sensor's low-level data. This process is achieved in four phases, which are smartphone sensor data capturing, data preprocessing, model building and context classification. In the data capturing phase, labeled high-level contexts such as user's activities [Walking, Sitting, Jogging, Driving, Running, etc.], including points of interest (POI) location information were collected using a mobile application running on the smartphone. The application collected 128 samples of data from each sensor's 3 axes, in continuous 3 seconds, with 64 samples from the previous 3 seconds overlapping the next 3 seconds. In the data preprocessing phase, outliers were removed (Lau and David, 2010). This is achieved by removing erroneous data segments from the beginning and ending of each collected data sample to minimize the influence of noise in the data. Simple time series features [range, mean and standard deviation] were then extracted from the remaining data samples.

In the model building and classification phases, extracted features in the last process is fed into a set of classification algorithms (Maninni, 2010) to build the context recognition models, which classify the features into activities and their corresponding locations, using leave-one-subject-out validation approach (Otebolaku & Andrade, 2013).

Contextual User Modeling and Profiling

The goal of user modeling or profiling is to capture not only interests or preferences of users but also

information about those things they do not like or want in specific contexts using either implicit or explicit processes (Otebolaku and Andrade, 2014a). We designed a dynamic and generic contextual user profile model, which summarizes user's content consumptions into a limited set of categories (Otebolaku and Andrade, 2014b). Categories are characterized by one or more genres, and genre is characterized by a number of properties. Additionally, it incorporates contextual dimension, associating one or more inferred contexts with each category-genre-property concept. We modeled the user profile as a four-level tree. The root of the tree is the user u_i with optional demographic information. The first level nodes correspond to category; the second level represents genre; the third level contains properties of a given category-genre. This level provides media item's context, characterizing at a finer detail, the consumed content and thus, user preferences. A limited set of properties is used for each genre to obtain a good compromise between sufficient degree of characterization of content and reasonable dimensions of the user profile. The leaf nodes provide information about contexts where user preferences are observed. Leaf nodes have three fields – type, weight, and lifetime, whereas all other nodes have only the type field. In the leaves, types represent the type of context. The weight provides information about number of times a user has consumed items of that *category-genre-property* in specific contexts, whereas lifetime defines the last time the user has consumed an item in a category. Each time the user starts consuming multimedia content, the system generates a contextual user profile vector V_c from the overall contextual user profile. After acquiring context data from the smartphone's sensors and inferring the current contextual situation of the user, the system uses this inferred data to filter the user profile based on context pre-filtering method (Adomavicius et al., 2005). Value of the recognized current context of use is compared with the leaves of the profile (context nodes) to identify the upper nodes that provide values for computing elements of V_c .

The *contextual user profile* is an important part of MobCAMP, which seamlessly integrates other components of the system. It also keeps track of the smartphone user's consumption history and contexts. The contextual user profile is updated using two approaches, explicit and implicit relevance feedbacks (Pazzani & Billsus, 2007). The former grants users the ability to modify values assigned to their preferences by the system. The implicit approach involves preference learning without involving users directly, but by updating their profiles, using information about user's response to recommended items.

Media Content Profiling

In addition to the contextual user profiling, it is necessary to create a media profile, represented by a vector V_M , for each candidate media item. This vector contains terms representing content of each candidate multimedia item (Otebolaku and Andrade, 2015a). To describe media items, the proposed system relies on the availability of semantic metadata using the MPEG-7 MDS description tools (MPEG7, 2013). For each media item, a vector V_M is initially created as an exact replica of V_C . Then, for every element of V_M , the system inspects the MPEG-7 metadata for a match. If it finds a match, it retains the intensity (relevance) of the matching element in V_M . Otherwise, it assigns zero to the element. The intensity measures interest generated by the user profile based on the frequency of media item's consumption in context.

Context-Aware Personalized Recommendations and Adaptations

Classifications of multimedia items in specific contexts, which can be performed in three phases, rely on the contextual user profile model. We extended three algorithms, namely (1) context-aware content based process. (2) Context-aware collaborative filtering process. (3) Context-aware content based collaborative process (CACBCP).

However, we will only introduce the third process. One common factor in these processes, however, is the *contextual user profile* (CUP) that relates user preferences to infer user preferences in any given contextual situation. The context recognition model infers the contextual information. These context-aware processes were extended with flexibility to allow non-contextual versions to execute in situations when contextual information might not be available. This would allow the system to switch between contextual and non-contextual user profiles. CACBCP first classifies users and multimedia content they have consumed in contexts similar to the target user's present context. Subsequently, it identifies every user who is similar to the target user by searching through their contextual profiles, looking for contexts that match the target user's present context. For every profile with a match, the intensity of corresponding category-genre-property nodes in the target user's profile is retrieved to form a user-multimedia content matrix. This matrix is then used to generate the similarity between each user's profile and the target user profile, using Pearson Correlation. After this calculation, the top n most similar users are then selected in decreasing order of the generated similarity values. The first n users, who have consumed content in the same or similar contexts to the target user's current context, with the highest similarity values are the so-called friends or neighbors of the target user. In the second phase, candidate media items for each neighbor are ranked by obtaining vectors V_c and V_m of the contextual user profile of every neighbor and candidate media profile vectors respectively. By applying the cosine distance correlation, the distance between these vectors is then generated.

In the third phase, using the cosine distance, the prediction for each of the contents for the target user, is generated using the Resnick prediction (Resnick et. al., 1994). Recommendation set is built by ordering candidate media items in descending order of magnitude of the computed prediction values. CACBCP also tracks the user's consumption behavior to improve the system, by

updating the profile whenever the user consumes any kind of media item.

MobCAMP Implementation

The implementation of MobCAMP consists of three primary components. The context recognition module that runs on the user's smartphone, monitoring, learning, and predicting user's contexts. This module gathers events from smartphone built-in sensors, pre-processing them, and inferring user's higher-level contexts from those events. Recognized context is then sent to CUP. CUP manages the preferences of users, by relating them to the user's present and past contexts (if any). It also learns the user's content consumption preferences and updates the profile accordingly. It stores user's consumption history and preferences according to the contexts in the profile repository. The third component, the personalized recommendation service, implements among others, CACBCP and a media content profiling process. The latter obtains content metadata from the Web, crawling and extracting useful information, whereas the former ranks and recommends content to the target user according to her contexts and preferences, and those of her friends. The adaptation process uses the network and device constraints as well as media item metadata to tailor the selected content accordingly. The context recognition process and recommendation client were implemented as an Android process and an application running on Galaxy S I9000 smartphone. The context-aware personalized recommendation service, the CUP, and the media profile service were implemented using Java EE components and exposed as REST web services interfaces (JavaEE, 2015), which can be accessed by clients running on the smartphones via Wi-Fi. Experiments conducted to evaluate the feasibility and the efficacies of the system produced encouraging results. For example, evaluations of the system using contextual information and without using contextual information show that recommendations with contextual information produce more relevant suggestions.

FUTURE RESEARCH DIRECTIONS

In order to address the weaknesses of personalized recommendation systems, many researchers are exploring CARS (Otebolaku & Andrade 2015a; Adomavicius et. al., 2011; Su et. al., 2010; Adomavicius et. al., 2005; Chen, 2005; Vallet et. al., 2007; Yu et. al. 2006). Nevertheless, present context-aware personalized recommendation techniques rely on rating provided by users in certain contexts. In practice, mobile users do not always provide ratings and sometimes, they consider this as distractions. Another challenge is that it is computationally expensive to compute recommendations for users who have similar tastes as the current user, especially when millions or even billions of users are involved. Tracking consumption history and knowing with whom users interact also require expensive computations. Recently, however, adoption of social networks, such as Twitter, Facebook, Google+, YouTube, etc. is providing a new platform to deal with this problem (Konstas, 2009). Personalization of mobile content using contextual recommendations will benefit from real-time social network data analytics based on user profile information, such as the user's online relationships, connections between other users, multimedia content sharing, and social tagging. Using social networks and contextual information as sources of data for recommendation systems, the expensive computations required by current similarity measures, especially when huge numbers of users and items are involved, will be greatly reduced. What would be required are friends of users on social networks, since it is believed that people have direct connections only with their friends, forming a very strong social influence. In addition, using social contexts to understand what a user needs, when and how he/she needs it as well as social and personalized engagements are important factors that will drive next generations of personalized recommendations for mobile users with relevant, timely, context-sensitive personalized yet non-invasive content deliveries. This important area

will become the center of the next generation of context-aware personalized systems. But, using social context sources is a challenging process considering privacy and security issues. Many social network users feel unsafe to share their preferences, contexts e.g. locations, activities, etc. with third party applications.

Nonetheless, we believe that social network based contextual information will improve the efficacy of context-aware recommendation systems, particularly those developed for mobile users. Nevertheless, researchers must study the implications of building context-aware personalized systems based on contextual information gleaned from social networks. Problematic issues of trust, privacy, and security, which are very critical, require important attention. There have been disturbing concerns by users on the issues of trust, privacy, and security, which could expose users to grave dangers and fraud as their data might be exposed to mischievous third parties. Users need the assurance that they are safe, and that their privacy is adequately guaranteed.

CONCLUSION

Smartphones have become integral parts of our daily routines because of their portability, incredible appeals and ability to process myriad multimedia content. With smartphones, we are now moving from the era of user's online routines to the era where online has become parts of user's routines. We are continuously seeking content that not only satisfies our contextual interests but also satisfies how we consume such content. However, with the huge volume of online-based content and constantly changing user preferences, it is difficult for users to find interestingly tailored content based on their contextual preferences. In this chapter, we have proposed a context-aware personalization system utilizing recommendation algorithms to assist mobile users to access relevant and interesting media content anywhere, anytime. It is not enough to suggest content to users when

and where they need the content but it is also equally important to suggest such content such that it is customized to the preferences of the users. One of the key components of the proposed system is the contextual user profiling/modeling. The contextual user profile has a wide scope, and can serve more than one context-aware filtering process. We have explained how to extend one of these processes, using contextual information, with the contextual user profile model serving a central role. We also discussed the context-aware content based collaborative technique and described its possible implementation. MobCAMP shows how context can be explored to improve recommendation efficacies, helping to filter out preferences that are not relevant in the present context and using it to adapt user preferred content. We have also argued that the future of context-aware multimedia personalization lies in using social network data with contextual information to glean user preferences for real-time contextual personalized recommendations.

REFERENCES

- Adomavicius, G., Mobasher, B., Ricci, F., & Tuzhilin, A. (2011). Context-Aware Recommender Systems. *AI Magazine*. doi:10.1609/aimag.v32i3.2364
- Adomavicius, G., Sankaranarayanan, R., Shen, S., & Tuzhilin, A. (2005). Incorporating contextual information in recommender system using a multidimensional approach. *ACM Transactions on Information Systems*, 23(1), 103–145. doi:10.1145/1055709.1055714
- Adomavicius, G., & Tuzhilin, A. (2005). Towards the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6), 734–749. doi:10.1109/TKDE.2005.99
- Alberto, P. (2010). *User-Centric Mobile Services: Context Provisioning and User Profiling*. Paper presented at the 11th Annual International Digital Government Research, Puebla, Mexico.
- Chen, A. V. (2005). *Context-Aware Collaborative Filtering System*. Paper presented at the International Workshop on Location and Context-Awareness, Portland, OR. doi:10.1007/11426646_23
- Flora, S. T. (2010). Introduction to Mobile Information Retrieval. *IEEE Intelligent Systems*, 25(1), 11–15. doi:10.1109/MIS.2010.22
- Gartner. (2015). *2013 Annual Smartphone sales*. Retrieved on March 7, 2014, from <http://www.gartner.com/newsroom/id/3115517>
- Goh, K. H., Tham, J. Y., Zhang, T., & Laako, T. (2011). *Context-Aware Scalable Multimedia Content Delivery Platform for Heterogeneous Mobile Devices*. Paper presented at the 3rd International Conference on Advances in Multimedia, Budapest, Hungary.
- Java, E. E. (2015) Retrieved on October 16, 2015 from <http://www.oracle.com/technetwork/java/javaee/overview/index.html>
- Kapteijn, M., Markopoulos, P., Ruyter, B., & Aarts, E. (2015). Personalizing persuasive technologies: Explicit and implicit personalization using persuasion profiles. *International Journal of Human-Computer Studies*, 77, 38–51. doi:10.1016/j.ijhcs.2015.01.004
- Konstas, I., Stathopoulos, V., & Jose, J. M. (2009). *On social networks and collaborative Recommendation*. Paper presented at 32nd international ACM SIGIR conference on Research and Development in Information Retrieval, New York, NY.
- Kwapisz, J., Weiss, G., & Moor S. (2010). Activity Recognition using Cell Phone Accelerometers. *ACM SIGKDD Explorations Newsletter*, 12(2).

- Lee, W.-P., & Tseng, G.-Y. (2015). Incorporating contextual information and collaborative filtering methods for multimedia recommendation in a mobile environment. *Multimedia Tools and Applications*, 2015. doi:10.1007/s11042-015-2915-8
- Mannini, A., & Sabatini, A. M. (2010). Machine Learning Methods for Classifying Human Physical Activity from Onbody Accelerometers. *Sensors (Basel, Switzerland)*, 10, 1154–1175. doi:10.3390/s100201154
- MPEG 7. (n.d.). Retrieved September 12, 2013 from: <http://mpeg.chiariglione.org/standards/mpeg-7/mpeg-7.htm>
- Otebolaku, A. M., & Andrade, M. T. (2013). *Recognizing High-Level Contexts from Smartphone Built-in Sensors for Mobile Media Content Recommendation*. Paper presented at the 14th IEEE International Conference on Mobile Data Management, Milan Italy. doi:10.1109/MDM.2013.84
- Otebolaku, A. M., & Andrade, M. T. (2014a). Context-Aware User Profiling and Multimedia Content Classification Smart Devices. In *28th IEEE International Conference on Advanced Information Networking and Applications Workshops (WAINA-2014)* (pp.560–565). IEEE. doi:10.1109/WAINA.2014.92
- Otebolaku, A. M., & Andrade, M. T. (2014b). A Context-Aware Framework for Media Recommendation on Smartphones. In L. De Strycker (Ed.), *ECUMCIT 2014* (pp. 87–108). Berlin: Springer. doi:10.1007/978-3-319-05440-7_8
- Otebolaku, A. M., & Andrade, M. T. (2015a). Context-Aware Media Recommendations for Smart Devices. *J Ambient Intell Humaniz Comput.*, 6(1), 13–36. doi:10.1007/s12652-014-0234-y
- Otebolaku, A. M., & Andrade, M. T. (2015b). *Context-Aware Multimedia Content Recommendations for Smartphone Users*. *Encyclopedia of Information Science and Technology* (3rd ed.). IGI Global; doi:10.4018/978-1-4666-5888-2
- Otebolaku, A. M., & Andrade, M. T. (2016). User context recognition using smartphone sensors and classification models. *Journal of Network and Computer Applications*, 66(C), 33–51. doi:10.1016/j.jnca.2016.03.013
- Pazzani, M., & Billsus, D. (2007). Content-based Recommendation Systems. In P. Brusilovsky, A. Kobsa, & W. Nejdl (Eds.), *The Adaptive Web* (pp. 325–341). Heidelberg, Germany: Springer-Verlag, LNCS. doi:10.1007/978-3-540-72079-9_10
- Ricci, F. (2011). Mobile Recommender Systems. *JIT Tourism*, 12(3), 205–231.
- Said, A., Luca, E. W. D., & Albayrak, S. (2011). *Inferring contextual user profiles—improving recommender performance*. Paper presented at the 3rd RecSys Workshop on Context-Aware Recommender Systems, Chicago, IL.
- Sarwar, B., Karypis, G., Konstan, J., & Reidl, J. (2001). *Item-based collaborative filtering recommendation algorithms*. Paper presented at the 10th international conference on World Wide Web, Hong Kong.
- Schafer, J. B., Frankowski, D., Herlocker, J., & Sen, S. (2007). Collaborative Recommender Systems. In P. Brusilovsky, A. Kobsa, & W. Nejdl (Eds.), *The Adaptive Web* (pp. 325–341). Heidelberg, Germany: Springer-Verlag, LNCS; doi:10.1007/978-3-540-72079-9_9
- Sian, L. L., & David, K. (2010). *Movement Recognition Using the Accelerometer in Smartphones*. Paper presented at the Future Network and Mobile Summit, Florence, Italy.
- Vallet, D., Fernandez, M., Castells, P., Mylonas, P., & Avrithis, Y. (2007). *Personalized information Retrieval in Context*. Paper presented at the 3rd International Workshop on Modeling Retrieval Context, Boston, MA.

Yu, Z., Zhou, X., Zhang, D., Chin, C. Y., Wang, X., & Men, J. (2006). Supporting Context-Aware Media Recommendations for Smart Phones. *IEEE Pervasive Computing*, 5(3), 68–75. doi:10.1109/MPRV.2006.61

ADDITIONAL READING

Abayomi Otebolaku. Context-Aware Personalization for Mobile Multimedia. PhD Thesis, Faculty of Engineering, University of Porto; 2015. Available from: <http://www.dart-europe.eu/full.php?id=1081749&ref=y>

Adomavicius, G., & Kwon, Y. (2007). New Recommendation Techniques for Multicriteria Rating Systems. *IEEE Intelligent Systems*, 22(3), 48–55. doi:10.1109/MIS.2007.58

Adomavicius, G., & Tuzhilin, A. (2011). Context-aware recommender systems. In F. Ricci, L. Rokach, B. Shapira, & P. Kantor (Eds.), *Recommender systems handbook* (pp. 217–256). Berlin: Springer; doi:10.1007/978-0-387-85820-3_7

Andrade, M. T., Dogan, S., Carreras, A., Barbosa, V., Arachchi, H. K., Delgado, J., & Kondo, A. M. (2012). Advanced delivery of sensitive multimedia content for better serving user expectations in virtual collaboration applications. *Multimedia Tools and Applications*, 58(3), 633–661. doi:10.1007/s11042-011-0749-6

Bobadilla, J., Ortega, F., Hernando, A., & Gutierrez, A. (2013). Recommender systems survey. *Knowledge-Based Systems*, 46, 109–132. doi:10.1016/j.knosys.2013.03.012

Burke, R. (2002). Hybrid Recommender Systems: Survey and Experiments. User Modeling and User-Adapted Interaction, 331–370. Doi: 10.1023/A:1021240730564.

Kaptein, M., Markopoulos, P., Ruyter, B., & Aarts, E. (2015). Personalizing persuasive technologies: Explicit and implicit personalization using persuasion profiles. *International Journal of Human-Computer Studies*, 77, 38–51. doi:10.1016/j.ijhcs.2015.01.004

Otebolaku, A. M., & Andrade, M. T. (2011). Context Representation for Context-Aware Mobile Multimedia Recommendation. In *Proceedings of the 15th IASTED International Conference on Internet and Multimedia Systems and Applications*. Washington, USA. doi:10.2316/P.2011.746-004

Otebolaku, A. M., & Andrade, M. T. (2014). Context-Aware Multimedia Content Recommendations for Smartphone Users. In 3rd Edition of Encyclopedia of information science and technology (pp. 5658–5666). doi:10.4018/978-1-4666-5888-2.ch559

Pessemer, T. D., Dooms, S., & Martens, L. (2013). *Context-aware recommendation through context and activity recognition in a mobile environment*. Multimed Tools Appl.

Su, J. H., Yeh, H. H., Yu, P. S., & Tseng, V. S. (2010). Music Recommendation Using Content and Context Information Mining. *IEEE Intelligent Systems*, 25(1), 16–26. doi:10.1109/MIS.2010.23

Xia, F., Asabere, N. Y., Ahmed, A. M., Li, J., & Kong, X. (2013). Mobile Multimedia Recommendation in Smart Communities: A Survey Access. *IEEE*, 1, 606–624.

Zhang, D., & Yu, Z. (2007). Spontaneous and Context-aware Media Recommendation in Heterogeneous Spaces. Paper presented at the IEEE 65th Semiannual Vehicular Technology Conference. Dublin, Ireland. doi:10.1109/VETECS.2007.67

KEY TERMS AND DEFINITIONS

Context Recognition: Is a process that identifies user's real-time contextual situations from sensory data, using pattern recognition, signal processing and machine learning algorithms.

Context-Aware Personalized Collaborative Recommendation: Is a personalized recommendation technique that suggests items, using items that other users have consumed in contexts similar to the target user's context.

Context-Aware Personalized Content-Based Collaborative Recommendation: Is a personalized recommendation technique that provides new items to a target user, using items that other users who are similar to the target user have consumed in the contexts similar to the target user's present context, including (if any) items that the target user has consumed in similar contexts.

Context-Aware Personalized Content-Based Recommendation: Is a personalized recommendation technique that provides suggestions to users, using items that such users have consumed in the past in contexts similar to their present contexts.

Context-Aware Personalized Recommendation: Is a technology that assists users to select

from overwhelming set of potential choices by matching and customizing available services to user's context dependent preferences. It supports the user by providing the right services in the right format at the right moments in order to improve user's experience.

Contextual User Profile: In context-aware personalized recommendations, a contextual user profile contains user's multimedia consumption preferences/interests in relation to contextual information such as location, optional personal information, user's device characteristics, and network information.

Smartphone: Is a mobile phone with sophisticated processing and computing capabilities, running an advanced operating system such as Android, Windows mobile, IOS, etc.

Smartphone Built-in Sensors: These are sensors including accelerometers, gyroscopes, GPSs, Wi-Fi, proximity sensors, etc. embedded in commercial smartphones.

Target User: Is a user whose profile is currently being processed by the recommendation system. It can also be considered as a user to whom recommendations are being provided.

Design of Compensators for Comb Decimation Filters

M

Gordana Jovanovic Dolecek
Institute INAOE Puebla, Mexico

INTRODUCTION

Decimation is the process of decreasing the sampling rate by an integer, called decimation factor. Decimation has applications in communications, audio signal processing, Sigma Delta Analog to Digital converters, among others. In order to prevent aliasing (unwanted replicas of the input signal), the signal must be previously filtered by a low pass filter, called decimation filter (Jovanovic Dolecek, 2003). The comb filter is the simplest decimation filter usually used in the first decimation stage (Hogenauer, 1981). This filter does not require multipliers, because all its coefficients are equal to unity. In order to achieve correct performance, the comb decimation filter should have a flat pass band of interest. However, comb magnitude characteristic has a droop in the pass band of interest which may deteriorate the decimated signal. The solution is to compensate for a comb pass band droop by an additional simple filter, called compensator. Different methods are proposed for compensator designs. The objective of this paper is to categorize and describe the most important methods, proposed so far, and to propose some future direction for the compensator designs.

BACKGROUND

The transfer function of comb filter is given by the following equation:

$$H(z) = \left[\frac{1}{M} \frac{1 - z^{-M}}{1 - z^{-1}} \right]^K \quad (1)$$

where M is the decimation factor and K is the order of the filter.

The magnitude response of the filter is given as:

$$|H(e^{j\omega})| = \left| \frac{1}{M} \frac{\sin(\omega M / 2)}{\sin(\omega / 2)} \right|^K \quad (2)$$

The comb pass band is defined by the pass band edge (Kwentus & Willson, 1997):

$$\omega_p = \pi / RM \quad (3)$$

where R is the decimation factor of the stage that follows the comb decimation stage.

For values $R < 4$, the pass band is considered as a wideband, and in an opposite case it is a narrowband.

As an example, Figure 1 shows the wide pass band zoom ($R=2$), of the magnitude response of comb filter with the decimation factor $M=12$ and an order equal to $K=3$. Note that the response is not flat and has a droop, which increases with the increase of the frequency ω . The inverse comb magnitude characteristic:

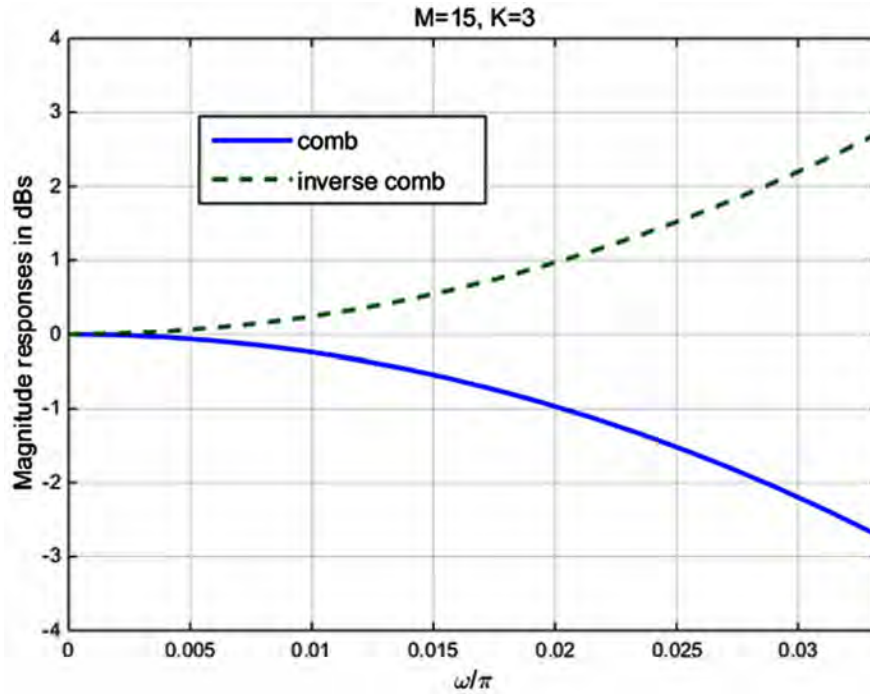
$$|H_i(e^{j\omega})| = \left| \frac{1}{H(e^{j\omega})} \right| = \left| \frac{M \sin(\omega / 2)}{\sin(\omega M / 2)} \right|^K \quad (4)$$

is also shown.

The product of the magnitude characteristics (2) and (4) results in unity:

$$|H(e^{j\omega})| |H_i(e^{j\omega})| = 1 \quad (5)$$

Figure 1. Magnitude and inverse magnitude characteristics of comb, $M=15$, $K=3$, $R=2$



Consequently, in order to get a flat comb magnitude characteristic it is necessary to cascade comb with a filter which has magnitude characteristic approximately equal to the inverse comb magnitude characteristic in the pass band. This filter is called a compensation filter. Denoting the magnitude characteristic of compensator as $|G(e^{j\omega})|$ it follows:

$$|G(e^{j\omega})| \approx |H_i(e^{j\omega})|, \text{ for } 0 \leq \omega \leq \omega_p \quad (6)$$

where ω_p is the pass band edge defined in (3).

Usually, compensation filter works at a low rate, i.e. after decimation. As a consequence, at high input rate, compensator is expanded by M .

The compensated comb is the cascade of comb and compensator. The corresponding transfer function at high input rate is:

$$H_{comp}(z) = H(z)G(z^M) \quad (7)$$

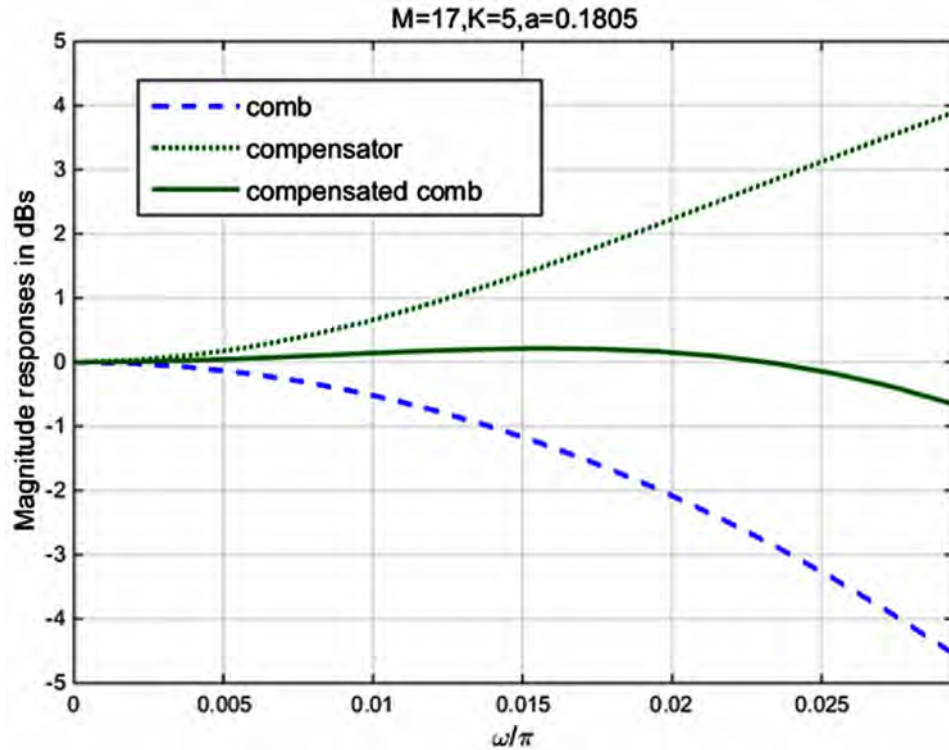
There are two principal reasons why an inverse comb filter cannot be a compensator: 1) Comb filter has zeros on the unit circle, becoming poles in the inverse comb filter, thus resulting in instability. 2) The resulting magnitude characteristic is equal unity for all frequencies. However, the unity magnitude characteristic is only required for the pass band.

Consequently, compensator should be designed to approximate inverse comb magnitude characteristic only in the pass band, and to no deteriorate comb stop band characteristic. Additionally, knowing that a comb filter is a very simple multiplier less filter, its compensator should be also a simple and desirable multiplier less filter.

REVIEW OF METHODS FOR DESIGN OF COMB COMPENSATORS

Different methods for design of comb compensators are advanced in the literature. The com-

Figure 2. Pass bands of comb, minimum error function-based compensator and compensated comb



compensator filters are usually FIR (Finite Impulse Response) filters due that FIR filters are always stable and can be designed to have a linear phase. The methods can be divided into two main groups: Methods requiring multipliers and Methods with multiplier less designs.

Methods Requiring Multipliers

1. **Minimum Error Function Method (Kim, Lee, Alm, & Choi, 2006):** The wideband compensator at low rate is defined by the following second order transfer function:

$$G(z) = \frac{-a}{(1-2a) + 1/(1-2a)z^{-1}} \quad (8)$$

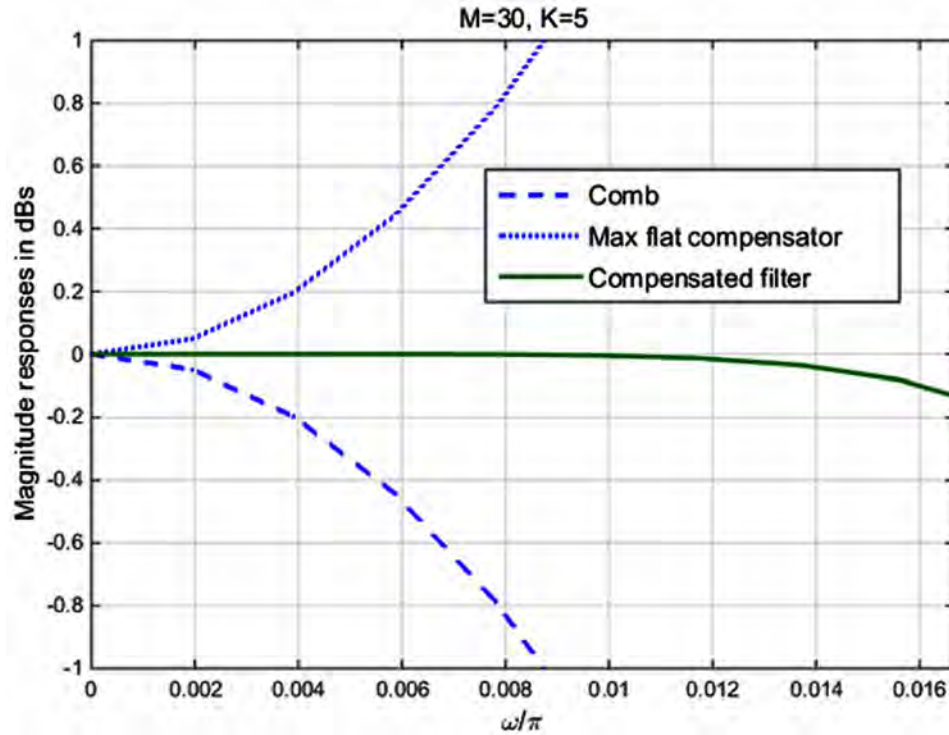
where a is parameter obtained by minimizing the error function for a given M and K . Taking into account the filter symmetry, total two multipliers

are required. The authors provide the values of the parameter a , obtained by minimizing the error function for values for $M=2$ to 17 and $K=5$. As an example Figure 2 shows the pass bands of the original comb, compensator, and the compensated comb taking $M=17$, and $K=5$ ($a=0.1805$). The pass band is defined in (3) taking $R=2$.

2. **Maximally Flat Error Criterion Method (Molnar & Vucic, 2011):** Wideband compensator is designed as a linear-phase FIR filter that approximates the response in (4) in a maximally flat sense. The compensator's coefficients (up to 18 coefficients) are obtained by solving a linear system of equations, depending on values of M and K . As an example, taking $M=30$ and $K=4$, the transfer function of the designed filter is obtained as:

$$G(z) = 0.0008z^{-1} - 0.0112z^{-2}$$

Figure 3. Pass bands of comb, maximally flat compensator and compensated comb



$$\begin{aligned}
 &+0.0800z^{-3}-0.3987z^{-4}+1.6582z^{-5} \\
 &-0.3987z^{-6}+0.0800z^{-7}-0.0112z^{-8} \\
 &+0008z^{-9}
 \end{aligned} \quad (9)$$

The compensator requires five multipliers and eight adders. The pass band zoom of the compensator magnitude characteristic is shown in Figure 3. The zooms of the pass band magnitude characteristics of comb and compensated comb are also shown. As a result of maximally flat design, the compensated comb has a flat magnitude characteristic especially in the lower part of the pass band, as can be seen in Figure 3.

A similar approach, based on maximally flat design, is proposed in (Fernandez-Vazquez & Jovanovic Dolecek, 2012), that includes also a multiplier less design.

Multiplier Less Methods

3. **Sinusoidal Magnitude Response Method (Jovanovic Dolecek & Mitra, 2008):** Second order narrowband compensation filter has the following transfer function:

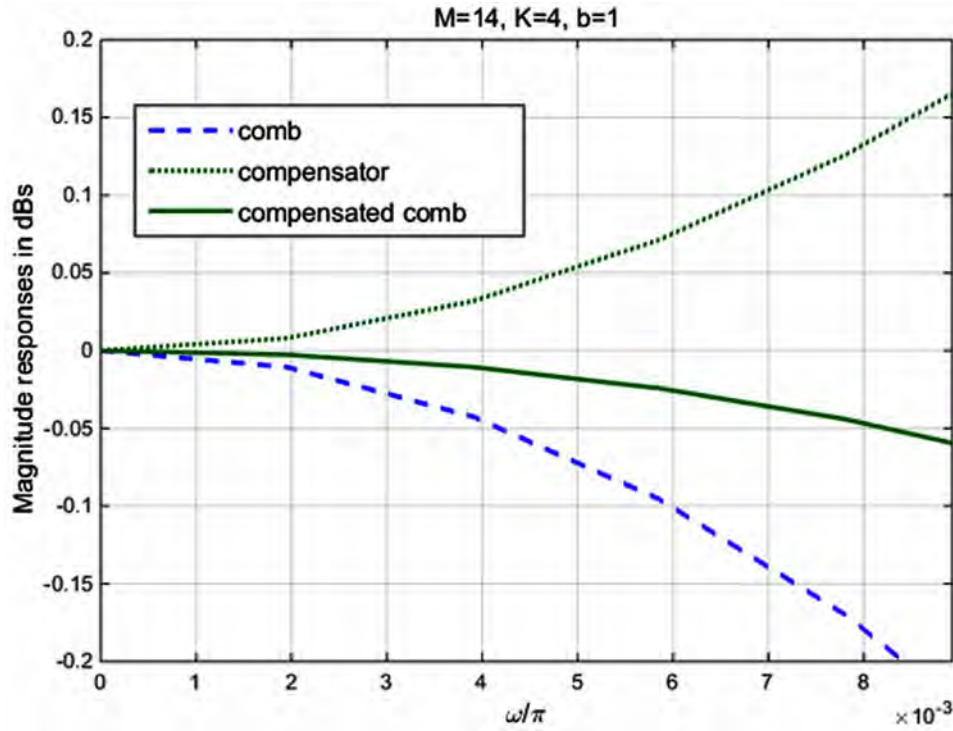
$$G(z) = -2^{-(b+2)} \left[1 - (2^{b+2} + 2)z^{-1} + z^{-2} \right] \quad (10)$$

where b is a parameter that depends on the comb parameter K . One can observe that the compensator does not depend on the decimation factor M .

The values of parameter b are: 2 for $K=2$, and 3; 1 for $K=4$; and 0 for $K=5$ and 6, taking $R=8$ in (3). The magnitude characteristic is in the sinusoidal form:

$$\left| G(e^{j\omega}) \right| = 1 + 2^{-b} \sin^2(\omega M / 2) \quad (11)$$

Figure 4. Pass bands of comb, narrowband compensator and compensated comb



As an example, using $M=14$, $K=3$, ($b=2$), Figure 4 shows the pass bands of comb, sinusoidal magnitude response compensator, and the compensated comb.

4. **Minimum Squared Error Method (Jovanovic Dolecek, 2009):** The wideband compensator, ($R=2$) proposed in (Jovanovic Dolecek, 2009) has the following transfer function:

$$G(z) = [G_1]^{K_1} \quad (12)$$

where:

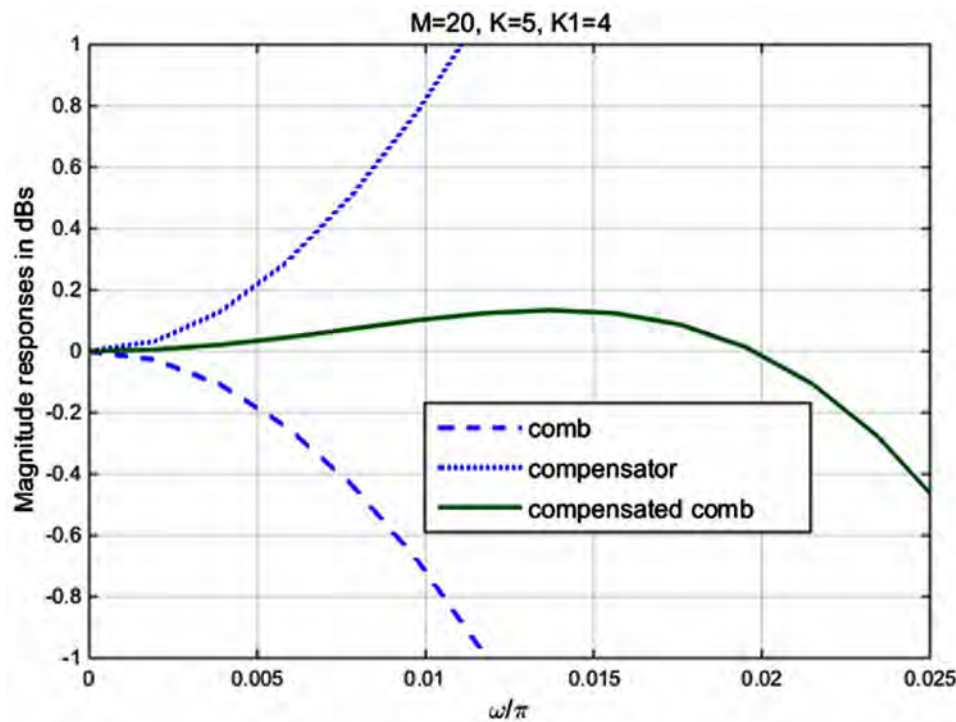
$$G_1(z) = -2^{-4} [z^{-1} - (2^4 + 2)z^{-2} + z^{-3}] \quad (13)$$

$$K_1 = \begin{cases} K & \text{for } 1 < K \leq 3 \\ K-1 & \text{for } K > 3 \end{cases} \quad (14)$$

The coefficients of the filter (13) are obtained using the condition that the compensator magnitude characteristic has the value 1 for $\omega=0$ and minimizing the squared error between the inverse comb characteristic and the compensator, in the pass band. Finally, the obtained coefficients are rounded using the rounding constant $r=2^{-6}$. The only parameter of design is the parameter K_1 , which depends only on the parameter K of the comb filter, as shown in (14). Note that filter (13) remains the same for all values of comb parameters M and K (see Figure 5).

5. **Min Max Absolute Error Function Method (Jovanovic Dolecek & Dolecek, 2010):** This method has a goal to decrease maximum absolute value of the pass band deviation in comparison with the method proposed in (Jovanovic Dolecek, 2009), and in same time decrease the number of the required adders. To this end, the proposed filter has the following transfer function:

Figure 5. Pass bands of comb, wideband minimum squared error compensator and compensated comb



$$G(z) = S \left[Bz^{-1} + Az^{-2} + Bz^{-3} \right] \quad (15)$$

where the parameters S , B , and A depend on the value of comb parameter K , and are obtained by minimizing the maximum absolute error function. The corresponding values are shown in Table 1.

Figure 6 shows the magnitude responses in the pass band, for $M=16$ and $K=4$, for comb, min max absolute error compensator, and compensated comb.

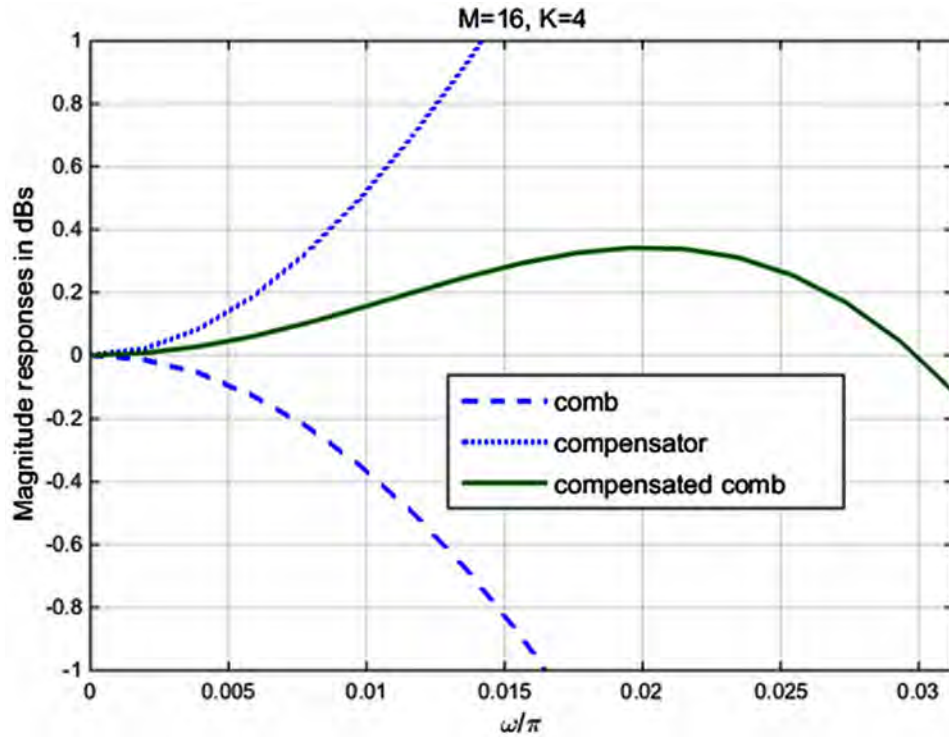
Table 1. Design parameters of min max-based compensator

K	S	B	A
1	2^{-4}	-1	2^4+2^1
2	2^{-3}	-1	2^3+2^1
3	2^{-4}	$-2 \cdot 2^0$	$2^4+2^2+2^1$
4	2^{-2}	-1	2^2+2^1
5	2^{-4}	$-2^2 \cdot 2^0$	$2^4+2^3+2^1$

6. **Interval Analysis Based Method (Pecotic, Molnar & Vucic, 2012):** The compensator's coefficients are obtained by optimization technique based on the interval analysis, and are expressed as the sums of powers of two (SPT), thus resulting in a multiplier less design. The filter coefficients in SPT forms are provided in a Table in terms of the values of K and the defined number of the compensator coefficients. The advantage of this method is that it provides the filter's coefficients not only for different values of K but also for different compensator orders. As an example, the decimation factor $M=32$ and $K=5$ are considered. In order to demonstrate the flexibility of the design two cases are elaborated, depending on the desired number of the filter coefficients.

In the first case the order of the filter is three, requiring 7 adders:

Figure 6. Pass bands of comb, wideband min max compensator and compensated comb



$$G(z) = g_0 + g_1 z^{-1} + g_2 z^{-2} \quad (16)$$

where:

$$\begin{aligned} g_0 &= g_2 = -2^{-2} - 2^{-4} - 2^{-6}; \\ g_1 &= 1 + 2^{-1} - 2^{-3} + 2^{-7} \end{aligned} \quad (17)$$

Figure 7 shows the pass bands of comb, interval analysis-based compensator, and compensated comb.

In the second case the compensator order is five and the filter has 14 adders:

$$G(z) = g_0 + g_1 z^{-1} + g_2 z^{-2} + g_3 z^{-3} + g_4 z^{-4} \quad (18)$$

where:

$$\begin{aligned} g_0 &= g_4 = 2^{-3} - 2^{-5} - 2^{-7} + 2^{-10}; \\ g_1 &= g_3 = -2^{-1} - 2^{-6} + 2^{-9} - 2^{-11}; \\ g_2 &= 2^{-3} - 2^{-5} - 2^{-7} + 2^{-10} \end{aligned} \quad (19)$$

Figure 8 illustrates the pass bands of comb, compensator and compensated filter.

As expected, compensator in the second case provides better comb compensation, due to its higher complexity.

7. Trigonometric Approach Method (Jovanovic Dolecek & Fernandez-Vazquez, 2014): The method is based on the sine-squared magnitude response:

$$|G(e^{j\omega})| = 1 + B \sin^2(\omega M / 2) \quad (20)$$

Table 2. Parameters of trigonometric-based compensator

The Values of Comb Parameter K	The Values of Compensator Parameter B
1	2^{-2}
2	2^{-1}
3	$2^{-1} + 2^{-2}$
4	2^0
5	$2^0 + 2^{-2}$

Figure 7. Pass bands of comb, interval analysis based-compensator (three coefficients), and compensated comb

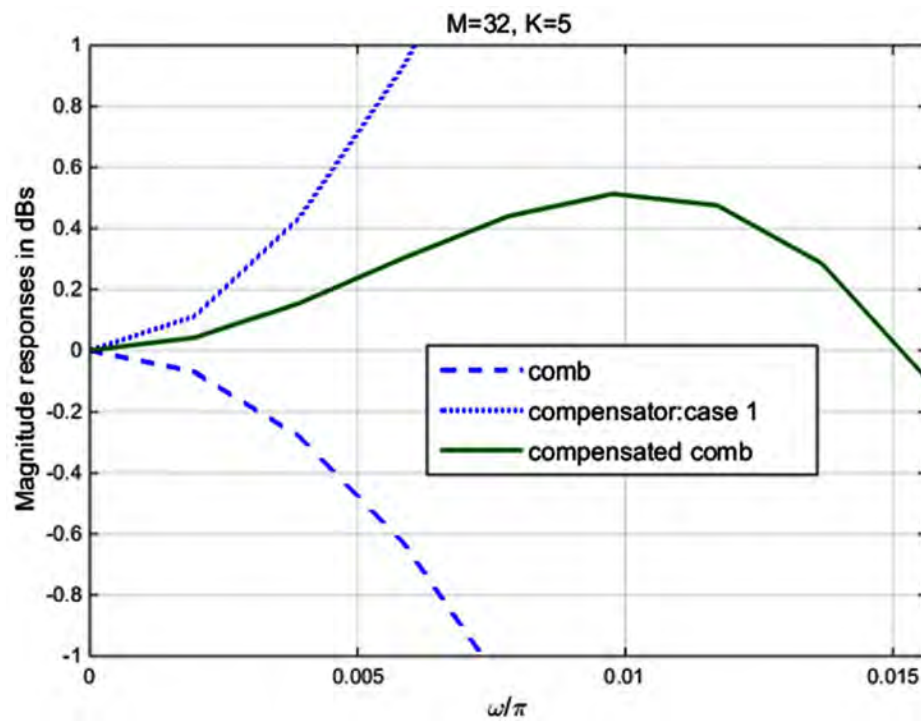
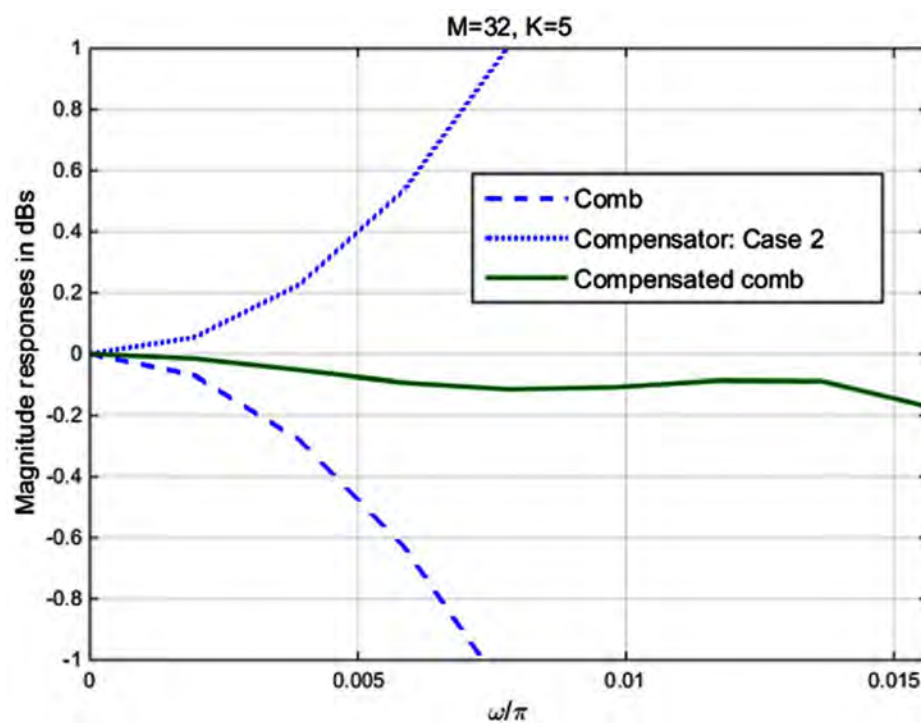


Figure 8. Pass bands of comb, interval analysis based-compensator (five coefficients), and compensated comb



where the values of compensator parameter B depend on the values of K , as shown in Table 2.

The filter has a low complexity requiring 3 adders for $K=1, 2, 4$ and 4 adders for $K=3$, and 5. As an example Figure 9 shows the magnitude responses in the pass band for the comb, compensation filter, and compensated comb, taking $M=18$ and $K=4$. (The parameter $B=2^0$).

8. **Amplitude Transformation Method (Troncoso Romero & Jovanovic Dolecek, 2013):** The method is based on the amplitude transformation of the cosine squared magnitude characteristic. The filter has low complexity, requiring only 4 adders. The filter coefficients in terms of K are obtained from the provided formula. As an example considering $M=25$ and $K=5$ the transfer function of the compensator is given as:

$$G(z) = (1 + 2^{-2})[2^{-2} + z^{-1}(1 - 2^{-1}) + z^{-2}2^{-2}] \quad (21)$$

Figure 10 shows the magnitude characteristics in the wide pass band ($R=2$) for comb, amplitude transformation-based compensator and the compensated comb.

9. **Two-Adders Method (Romero Troncoso, Molina Salgado & Jovanovic Dolecek, 2015):** The compensator proposed in (Romero Troncoso, Molina Salgado & Jovanovic Dolecek, 2015) is the simplest compensator requiring only two adders. The transfer function is given as:

$$G(z) = 1 + 2^{d-2} - 2^{d-2} z^{-1} \quad (22)$$

where d is an integer which depends on K .

The values for the parameter d are given for both: narrowband and wideband compensation. For a narrowband compensation the values of d are equal to -1 and 0 for the values of K equal to 2, 3, and 4, 5 respectively. Similarly, for a wideband case, $d=0$ for K equal to 2, and 3; $d=1$, for $K=4$, and 5. This filter requires only two adders and does

Figure 9. Pass bands of comb, trigonometric-based compensator, and compensated comb

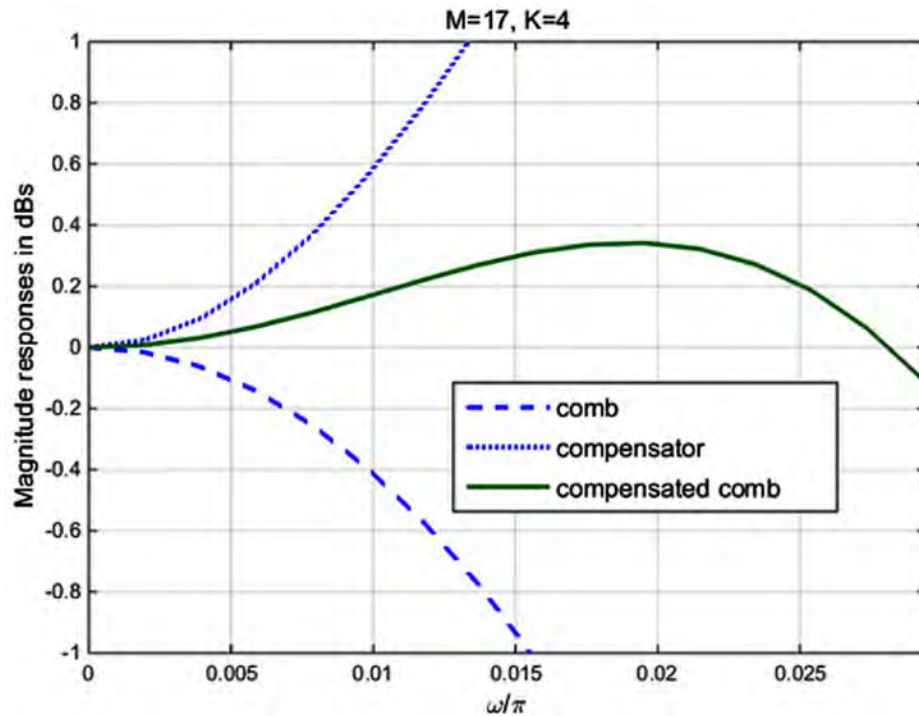
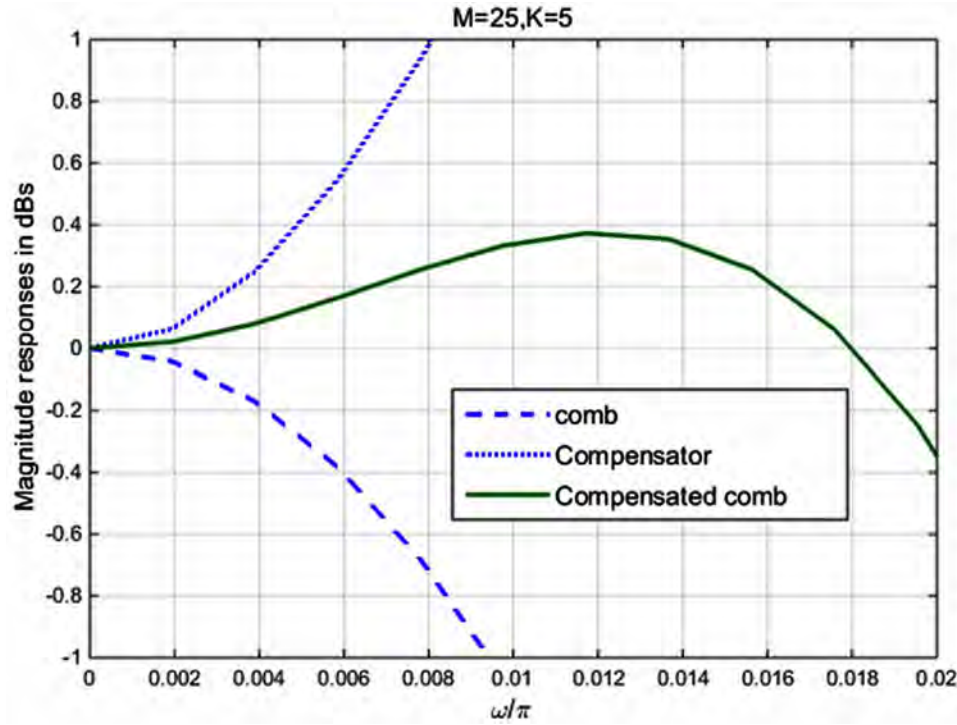


Figure 10. Pass bands of comb, amplitude transformation -based compensator, and compensated comb



not have the symmetry and consequently its phase is not linear. However, the phase is approximately linear in the pass band. As an example consider $M=10, K=4$ and narrowband compensation ($R=8$), taking $d=0$. Figure 11 shows the magnitude responses of comb, two adders based-compensator and compensated comb.

Next, the wideband case is considered taking the value of $d=1$. Figure 11 shows the corresponding magnitude responses.

It can be observed that the narrowband compensation is better than the wideband compensation. Another interest feature of this filter is that it has zero inside of unit circle. Knowing that the comb filter has all zeros on the unit circle, the compensated comb is practically a minimum-phase filter thus having a low group delay.

10. **Sharpening-Based Method (Jovanovic Dolecek & Harris):** The narrowband filter given in (10) is sharpened using the simplest sharpening polynomial $H_{sh}=2H-H^2$:

$$G(z) = 2H_c(z) - [H_c(z)]^2 \quad (23)$$

where:

$$H_c(z) = -2^{-(b+2)} [1 - (2^{b+2} + 2)z^{-1} + z^{-2}] H(z) \quad (24)$$

and the parameter b is given in (Jovanovic Dolecek & Mitra, 2008). In spite that the compensation filter from (Jovanovic Dolecek & Mitra, 2008) is a narrowband filter, using the sharpening, the wideband compensation is obtained, as shown in the following example.

As an example, consider $M=11$ and $K=2$. The parameter b is equal to 1, leading to the following transfer function:

$$\begin{aligned} H_c(z) &= -2^{-3} [1 - (2^3 + 2)z^{-1} + z^{-2}] \left[\frac{1}{11} \frac{1 - z^{-11}}{1 - z^{-1}} \right]^2 \\ &= -2^{-3} [1 - (2^3 + 2)z^{-1} + z^{-2}] \left[\frac{1}{11} \frac{1 - z^{-11}}{1 - z^{-1}} \right]^2 \end{aligned} \quad (25)$$

Figure 11. Pass bands of comb, two adders narrowband compensator, and compensated comb

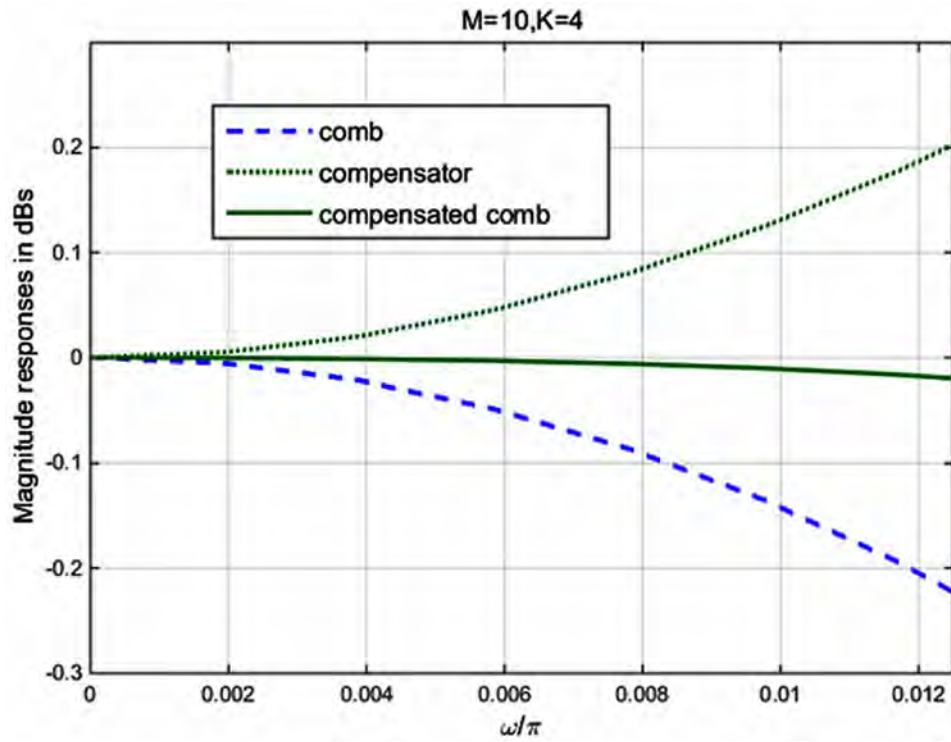


Figure 12. Pass bands of comb, two adders wideband compensator, and compensated comb

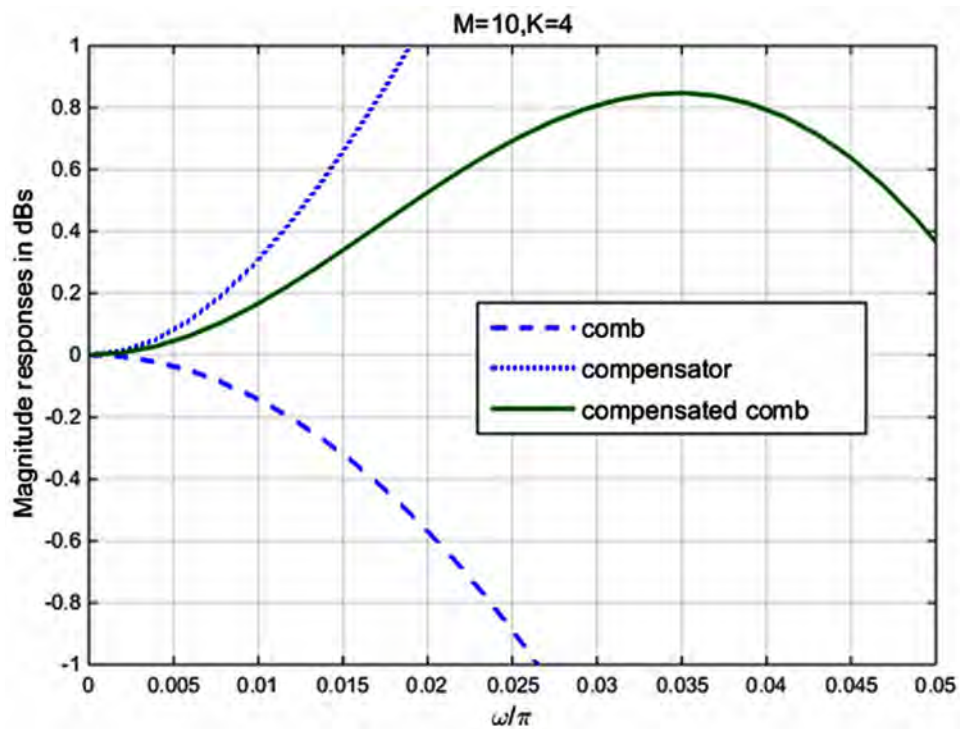


Figure 13. Pass bands of comb and sharpened narrowband compensator

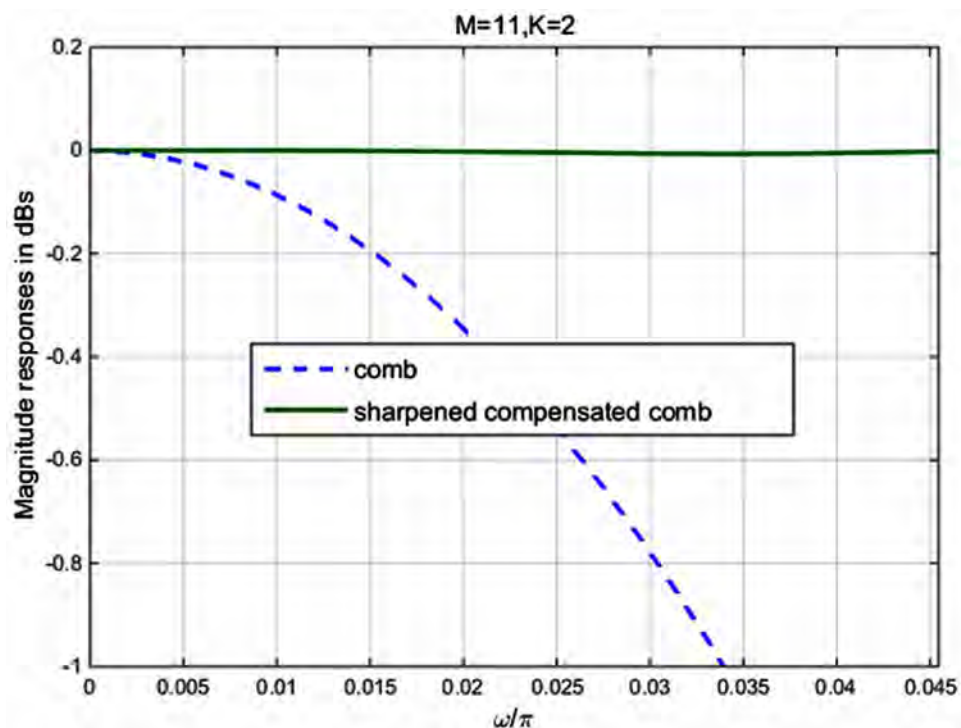


Figure 13 shows the pass bands of comb and the filter from (25). Note that there is perfect wideband compensation. However, the price is the increased complexity and deterioration of comb alias rejection.

FUTURE RESEARCH DIRECTIONS

The number of methods for comb pass band compensation is proposed so far. Some promising characteristics of the proposed compensators are for example multiplier less design, work at low rate, independency of the decimation factor, easy design. However, considering the future development of modern information technology and new designs of oversampled sigma delta modulators and software defined radio, where comb filters have important applications, the future trends are in the low power multiplier less designs, and reconfigurability. Additionally, knowing that the multiplier less compensators proposed so far

still have an absolute value of pass band deviation which is about 3 dBs, in future designs it is expected to decrease the maximum absolute value of the pass band deviation, and thus provide approximately flat pass band characteristic of the decimated comb, without a significant increase of filter complexity expressed in the number of adders, and without deterioration of a comb aliasing rejection capability.

Knowing that there are a number of methods, proposed so far, for improving a comb alias rejection, one possible direction is to investigate efficient compensator designs for modified combs with an improved alias rejection.

CONCLUSION

This article presents review of the methods for comb compensators designs. The simplest decimation filter is a comb filter. However, its magnitude characteristic in the pass band is not flat and has

a pass band droop, which may deteriorate a decimated signal. Therefore, it is important to improve a comb pass band magnitude characteristic with a simple filter, called compensator. Knowing that a comb filter is a very simple multiplier less filter, its compensator should be also a simple filter.

The compensators usually work at low rate, and do not depend on the decimation factor but only on the comb order K . All methods can be divided into two main groups, depending if need multipliers or not. Usually, better compensation is obtained with more complex compensators including multipliers and/or an increased number of adders. The methods may be also classified as methods for narrowband and wideband compensation. Some of proposed methods can be used for both: narrowband and wideband compensation, just changing the design parameters.

REFERENCES

- Fernandez-Vazquez, A., & Jovanovic Dolecek, G. (2012). Maximally flat CIC compensation filter: Design and multiplierless implementation. *IEEE Transactions on Circuits and Systems II, Express Briefs*, 54(2), 113–117. doi:10.1109/TCSII.2011.2180093
- Hogenauer, E. B. (1981). An economical class of digital filters for decimation and interpolation. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-29(2), 155–162. doi:10.1109/TASSP.1981.1163535
- Jovanovic Dolecek, G. (Ed.). (2003). *Multirate systems: Design and applications*. Hershey, PA: IIR Press.
- Jovanovic Dolecek, G. (2009). Simple wideband CIC compensator. *Electronics Letters*, 45(24), 1270–1272. doi:10.1049/el.2009.1860
- Jovanovic Dolecek, G., & Dolecek, L. (2010). Novel multiplierless wide-band CIC compensator. In *Proceedings of 2010 IEEE International Symposium on Circuits and Systems ISCAS 2010* (pp. 283–288). Paris, France: IEEEExplore.
- Jovanovic Dolecek, G., & Fernandez-Vazquez, A. (2014). Trigonometrical approach to design a simple wideband comb compensator. *International Journal of Electronics and Communications*, 68(5), 437–441. doi:10.1016/j.aue.2013.11.007
- Jovanovic Dolecek, G., & Harris, F. (2009). Design of wideband compensator filter for a digital IF receiver. *Digital Signal Processing*, 19(5), 827–837. doi:10.1016/j.dsp.2009.03.013
- Jovanovic Dolecek, G., & Mitra, S. K. (2008). Simple method for compensation of CIC decimation filter. *Electronics Letters*, 44(19), 1162–1163. doi:10.1049/el:20081603
- Kim, S., Lee, W. C., Alm, S., & Choi, S. (2006). Design of CIC roll-off compensation filter in a W-CDMA digital receiver. *Digital Signal Processing*, 16(6), 846–854. doi:10.1016/j.dsp.2006.06.003
- Kwentus, A., & Willson, A. Jr. (1997). Application of filter sharpening to cascaded integrator-comb decimation Filters. *IEEE Transactions on Signal Processing*, 45(2), 457–467. doi:10.1109/78.554309
- Molnar, G., & Vucic, M. (2011). Closed-form design of CIC compensators based on maximally flat error criterion. *IEEE Transactions on Circuits and Systems II, Express Briefs*, 58(12), 926–930. doi:10.1109/TCSII.2011.2172522
- Pecotic, M., Molnar, G., & Vucic, M. (2012). Design of CIC compensators with SPT coefficients based on interval analysis. In *Proceedings of 35th IEEE International. Convention MIPRO 2012* (pp. 123–128). Opatija, Croatia: IEEEExplore.

Troncoso Romero, D. E., & Jovanovic Dolecek, G. (2013). Application of amplitude transformation for compensation of comb decimation filters. *Electronics Letters*, 49(16), 985–987. doi:10.1049/el.2013.1492

Troncoso Romero, D. E., Molina Salgado, G., & Jovanovic Dolecek, G. (2015). Simple two-adders CIC compensator. *Electronics Letters*, 51(13), 993–994. doi:10.1049/el.2015.0742

ADDITIONAL READING

Harris, F. J. (2004). *Multirate signal processing for communications systems*. Prentice Hall.

Jovanovic Dolecek, G. (Ed.). (2001). *Multirate systems: Design and applications*. Hershey, PA, USA: IGI Global.

Jovanovic Dolecek, G., & Diaz Carmona, J. (2005). A new cascaded modified CIC-cosine decimation filter. In *Proceedings IEEE Symposium on Circuits and Systems, ISCAS 2005*. (pp. 3733–3736). Kobe, Japan: IEEEExplore. doi:10.1109/ISCAS.2005.1465441

Lyons, R. G. (2010). *Understanding digital signal processing* (3rd ed.). Prentice Hall.

Milic, L. (2009). *Multirate filtering for digital signal processing: MATLAB Applications*. Hershey, PA, USA: IGI Global. doi:10.4018/978-1-60566-178-0

KEY TERMS AND DEFINITIONS

Comb Filter: A simplest decimation filter which has all coefficients equal to unity.

Comb Compensation Filter (Compensator): The filter which has the magnitude characteristic which is an approximation of the inverse comb magnitude characteristic in the pass band. The filter is cascaded with the comb filter.

Comb Passband: The frequency band in which the decimated signal must be preserved. Ideally its characteristic is inverse of that of comb, thus resulting in approximately flat magnitude characteristic of the compensated comb in the pass band. The width of the pass band depends on the comb decimation factor, and the decimation factor of the stage which follows the comb decimation stage.

Decimation: Decreasing the sampling rate in a digital form by an integer.

Error Function: The difference of the desired and designed magnitude responses of compensated filter.

Multipliers: Implementation of coefficients of filter transfer function which cannot be presented as a power of two.

Multiplierless Design: The coefficients of filter are presented as powers of two which can be implemented as shifts and adders, thus avoiding multipliers.

An Empirical Study of Mobile/Handheld App Development Using Android Platforms

M**Wen-Chen Hu***University of North Dakota, USA***Naima Kaabouch***University of North Dakota, USA***Hung-Jen Yang***National Kaohsiung Normal University, Taiwan*

INTRODUCTION

When Apple launched its iPhones in 2007, they opened an App Store subsequently. They claimed the store has over 100 billion apps download, and has paid out more than \$30 billion to developers since its inception. There were more than 1.5 million apps available in 2015 (TechCrunch, 2015). The highly popular iPhones help the sales and development of applications. On the other hand, the large number of apps helps the sales of iPhones too. It is a win-win situation for both of the Apple, Inc. and app developers. Witnessing the success of the App Store, other mobile operating system providers realized they might be left behind if they did not have this kind of stores for their operating systems. They set up their own app stores immediately. Some of the major stores are given in Table 1 (Hu, 2016). The highly popular apps create great opportunities for IT companies and workers. However, traditional desktop programmers have problems switching to handheld programming because it requires a different approach from desktop programming (Kiely, 2001). This chapter introduces essential technologies for mobile/handheld computing, so more IT workers can join the mobile trend of computing.

Desktop application developers use standard tools or software (such as C++ and Java compil-

ers) to develop applications for different platforms (like Linux and Windows) with little or no changes. Unlike desktop application development, there are no widely accepted tools or software for mobile app development. For example, completely different approaches are required for Android and iOS app development. At the same time, mobile app development is much more complicated and platform-specific compared to desktop application development. This chapter introduces mobile app development by giving a simple Android application. Mobile developers can get a sense of mobile app development by reading this chapter and apply it to other platforms or further explore the Android app development.

The rest of this chapter is organized as follows. Section 2 gives background information about mobile/handheld computing, which includes a generic system structure of mobile handheld devices and client-side handheld computing. A variety of approaches is available for mobile app development. Sections 3 and 4 introduce the development using Android. Related tools and software for Android app development are introduced in Section 3. Section 4 explains a simple app development using Android. The example includes several XML and Java files and the Android Studio IDE (Integrated Development Environment) is used in the development. The final section summarizes this study.

Table 1. Major mobile application stores (Hu, 2016)

Company		Mobile Application Store		
Name	Major Mobile Products	Name	Major Operating Systems Supported	Launch Date
Apple Inc.	Smartphone Mobile operating system	App Store	iOS	07/10/2008
Open Handset Alliance (Alphabet)	Mobile operating system	Google Play	Android	10/22/2008
Microsoft	Mobile operating system	Windows Phone Apps	Windows Phone	10/06/2009
Research In Motion	Smartphone Mobile operating system	BlackBerry World	BlackBerry OS	04/01/2009
Samsung	Smartphone	Samsung Apps	Android Windows Phone	09/14/2009
LG	Smartphone	LG Smart World	Android Windows Phone WebOS	
GetJar	None	GetJar	Almost all	xx/xx/2004
Opera	Mobile browser	Opera Mobile Store	Almost all	02/16/2009
Sony	Smartphone	Apps	Android Windows Phone	02/xx/2004

BACKGROUND

Handheld computing is the use of handheld devices like smart cellular phones to perform wireless, mobile, handheld operations such as browsing the mobile Web and finding the nearby gas stations. This section discusses two handheld computing subjects: mobile handheld devices and client-side handheld computing.

Mobile Handheld Devices

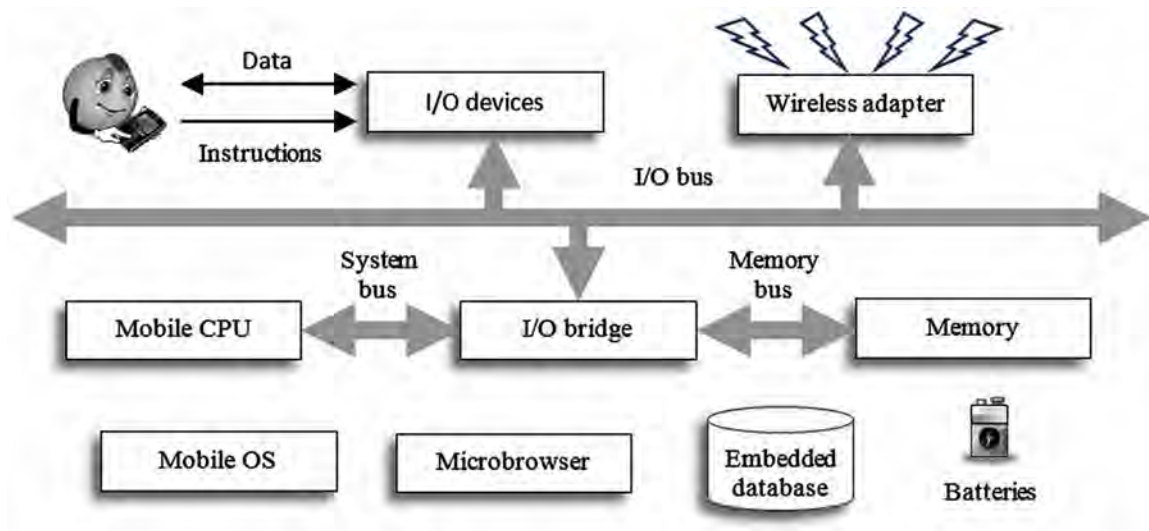
Mobile users perform mobile transactions using small wireless Internet-enabled devices, which come with several aliases such as handhelds, palms, PDAs, pocket PCs, and smartphones. To avoid any ambiguity, a general term, mobile handheld devices, is used in this article. Mobile handheld devices are small general-purpose, programmable, battery-powered computers, but they are different from desktop PCs, notebooks, or tablets due to the following two special features: small screen/body size and mobility.

Short battery life and limited network bandwidth, memory, processing power, and functional-

ity are additional features, but these problems are gradually being solved as the technologies improve and new methods are constantly being introduced. Though the Wi-Fi and 4G networks and above go some way toward addressing the problem of limited bandwidth, the wireless bandwidth is always far below the bandwidth of wired networks. The small screen/body size restricts the input speed and convenience. Figure 1 shows a typical system structure for handheld devices, which includes the following six major components (Hu, Zuo, Chen, & Yang, 2010):

- **Mobile Operating Systems:** Simply adapting desktop operating systems for handheld devices has proved to be futile. A mobile operating system needs a completely new architecture and different features to provide adequate services for handheld devices. A generalized mobile operating system structure can be visualized as a six-layer stack: (i) applications, (ii) GUI, (iii) API framework, (iv) multimedia, communication infrastructure, and security, (v) com-

Figure 1. A system structure of mobile handheld devices



puter kernel, power management, and real-time kernel, and (vi) hardware controller.

- Mobile Central Processing Units:** Handheld devices are becoming more sophisticated and efficient every day and mobile users are demanding more functionality from their devices. To achieve this advanced functionality, in addition to the obvious feature, low cost, today's mobile processors must have the following features: (i) high performance, (ii) low power consumption, (iii) multimedia capability, and (iv) real-time capability. The cores and architectures designed by Cambridge-based ARM Holdings Ltd. have dominated the mobile CPU market.
- Microbrowsers:** Microbrowsers are miniaturized versions of desktop browsers such as Netscape Navigator and Microsoft Internet Explorer. They provide graphical user interfaces that allow mobile users to interact with mobile apps.
- Input/Output Components:** Various I/O devices have been adopted by mobile handheld devices. The only major output device is the screen, but there are several popular input devices, among them: soft keyboards and touch screens.

- Memory:** Three types of memory are usually employed by handheld devices: (i) RAM, (ii) ROM, and (iii) flash memory. Hard disks provide much more storage capacity, but are more fragile.
- Batteries:** At present, rechargeable Lithium Ion batteries are the most common batteries used by handheld devices. However, the life of this kind of battery is short and the technology will not significantly improve unless and until manufacturers begin to switch to fuel cells, which may not happen for at least several years.

Synchronization connects handheld devices to desktop computers, notebooks, or peripherals to transfer or synchronize data. Without needing serial cables, many handheld devices now use either an infrared (IR) port or Bluetooth technology to send information to other devices.

Client-Side Handheld Computing

There are two kinds of handheld computing: server and client-side handheld computing. Server-side handheld computing includes the design and development of handheld software such as PHP programs that reside on servers. Client-side hand-

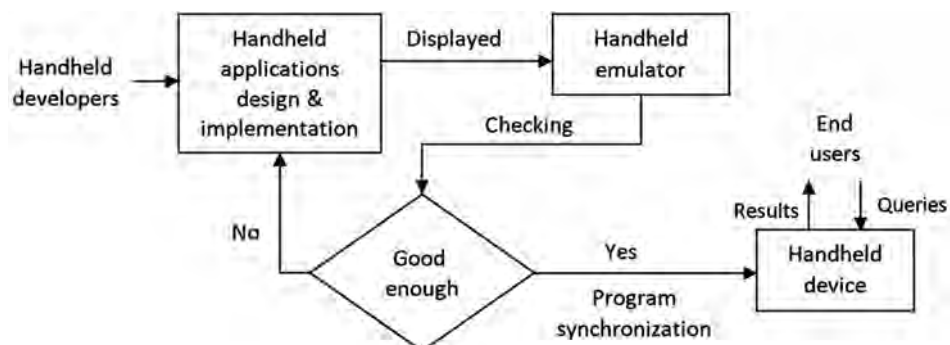
Table 2. The most popular four mobile operating systems and their features

Mobile OS		Android	iOS	Windows Phone	BlackBerry OS
Company/developer		Open Handheld Alliance	Apple	Microsoft	RIM
2015	Market share	1 st (81.6%)	2 nd (12.6%)	3 rd (1.9%)	4 th (0.3%)
	Millions of units shipped	1,160.2	179.8	16.7	4.3
2014	Market share	1 st (80.7%)	2 nd (15.4%)	4 th (2.8%)	3 rd (0.6%)
	Millions of units shipped	1,004.6	191.7	34.9	7.5
Development languages		Java	Objective C/C++	Visual C++	Java
Kernel type		Linux	Hybrid	Windows CE 6/7	Unix
IDEs, libraries, frameworks		Android SDK; Android Studio	iPhone SDK	Windows Phone SDK (works with Visual Studio)	BlackBerry JDE
Source model		Open	Closed (open for the core)	Closed	Closed
Initial release		2008	2007	2010	1999
Latest version as of March 2016		6.0.1 Marshmallow	9.0	8.1	7.1
Mobile application store		Google Play	App Store	Windows Phone Apps - Microsoft Store	BlackBerry World

held computing is the programming for handheld devices and it does not need the supports from server-side programs. Typical applications created by client-side handheld computing include (i) address books, (ii) video games, (iii) note pads, and (iv) to-do list. Various environments/operating systems/languages are available for client-side handheld programming. Table 2 lists the most popular four mobile operating systems (Hu, 2016; Google, n.d.a; Research In Motion, n.d.; Apple, Inc., n.d.; Microsoft, n.d.).

Each mobile operating system applies a different approach to accomplishing the development of mobile applications. Figure 2 shows a generic development cycle applied by them. Handheld emulators instead of the handhelds themselves are used for the development because of the convenience reason. The next section gives an example of Android application development. From the example, readers can get a sense of mobile app development, which is complicated and difficult compared to desktop application development.

Figure 2. A generic client-side handheld computing development cycle



ANDROID APPLICATION DEVELOPMENT ENVIRONMENT

There are a variety of mobile platforms and each platform has its unique approach for mobile app development. It is not possible to cover all of them in this chapter. This chapter focuses Android, which is the most popular mobile platform nowadays. Android is a software stack for mobile devices such as smartphones and tablet PCs (Open Handset Alliance, n.d.). It was developed by the Open Handset Alliance, a consortium of 84 hardware, software, and telecommunication companies led by Google devoted to advancing open standards for mobile devices. Google purchased the initial developer of the software, Android Inc., in 2005. Android includes an operating system, middle-ware, and key applications. The Android SDK (Software Development Kit) provides the tools and APIs necessary for developing applications on the Android platform using the Java programming language. Android uses the Dalvik virtual machine with just-in-time compilation to run compiled Java code. Android Studio download (Google, n.d.b) provides everything you need to start developing apps for Android, including the software such as Android Studio IDE, Android SDK tools, Android 6.0 (Marshmallow) Platform, Android 6.0 emulator system image with Google APIs, Android virtual device, and Intel Hardware Accelerated Manager (HAXM). This section introduces major components of Android and the next section will give an example of using these components to develop a small app.

Android Studio

In the past, Android app development in Eclipse with ADT (Android Development Tools) was highly recommended. Eclipse is a software development environment comprising an integrated development environment (IDE) and an extensible plug-in system and ADT is a plugin for the Eclipse IDE. Together they provide an integrated environment in which to build Android applications. Since

2014, Android Studio became the primary IDE for native Android application development. It is freely available under the Apache License 2.0. Android Studio is a powerful IDE and Figure 3 shows a screenshot of an instance of Android Studio, which includes the following major sections: Android logcat, component tree, Gradle console, (view) properties, project structure, and toolbar and tool buttons. Some of the features offered by Android Studio include Gradle-based build system, build variants and multiple apk (Android Application Package) file generation, code templates to help developers, layout editor with support for drag and drop theme editing, lint tools to catch problems, ProGuard and app-signing capabilities, and built-in support for Google Cloud Platform.

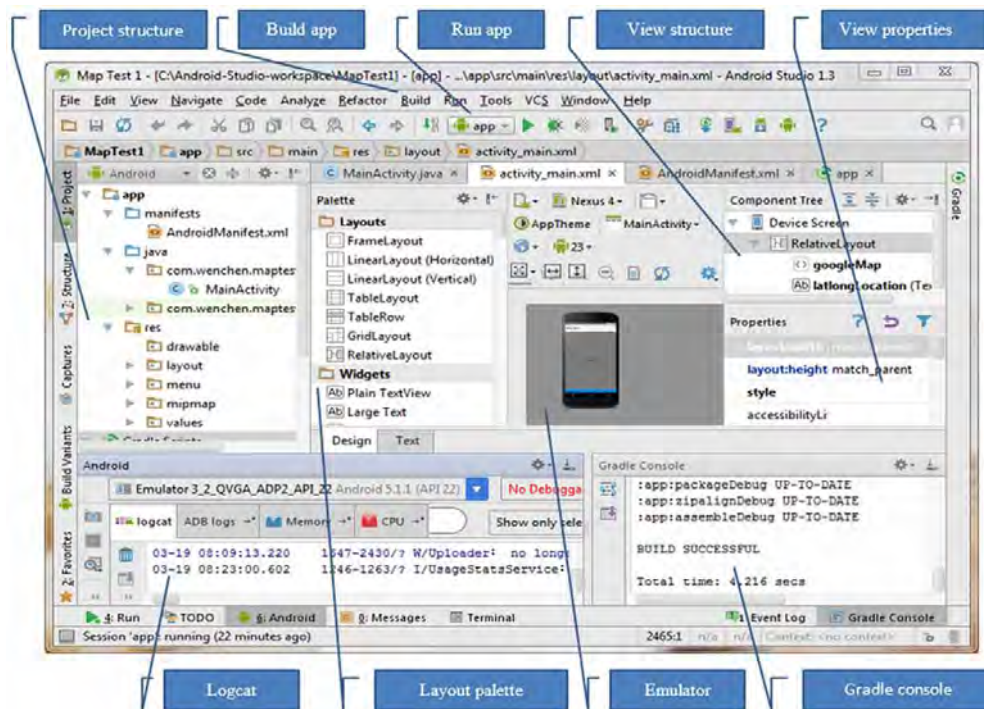
Android Virtual Devices (AVDs)

Mobile app development on the devices is not feasible because of many reasons such as the small size of devices and lack of powerful IDEs. In order to develop apps on desktop computers, a smartphone emulator is therefore needed. An AVD is an emulator that models an actual device by defining hardware and software options through AVD Manager:

- **A Hardware Profile:** Which defines the hardware features of the virtual device such as camera, network, emulated performance, and memory and storage,
- **A Mapping to a System Image:** Which defines what version of the Android platform will run on the virtual device such as the version of the standard Android platform and API level, and
- **Other Options:** Which specify the emulator skin you want to use with the AVD such as screen dimensions and resolution.

Figure 4 shows an instance of an AVD. Other than the AVD Manager, Android Device Monitor provides a graphical user interface for several Android application debugging and analysis tools.

Figure 3. A screenshot of an instance of Android Studio



Among many of its features, DDMS (Dalvik Debug Monitor Server), a debugging tool, provides port-forwarding services, screen capture on the device, thread and heap information on the device, logcat, process, and radio state information, incoming call and SMS spoofing, location data spoofing, and more.

SQLite, an Embedded RDBMS (Relational Database Management System)

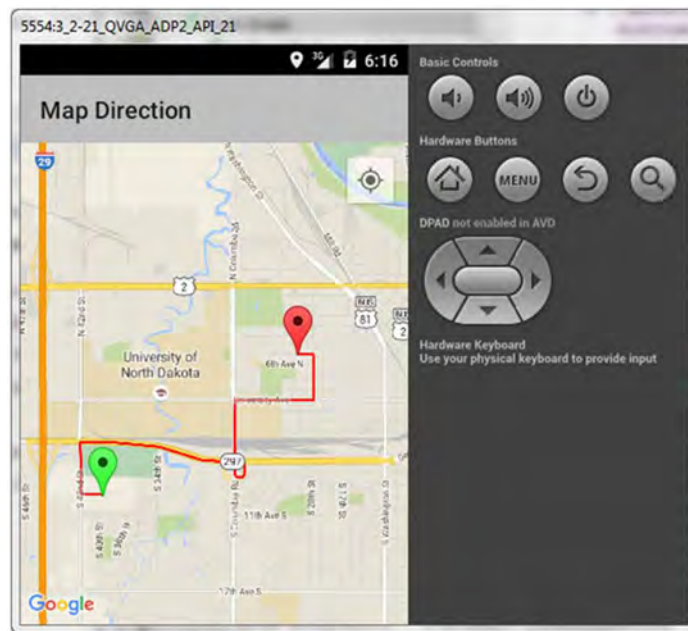
Embeddable databases are embedded within devices or application processes. They include the following features making them fast and ideal for handheld devices: (i) removing the overhead associated with a client-server configuration, (ii) light weight in the sense that they require little memory during run time, and (iii) written in compact code. SQLite (n.d.a) is an open source embeddable RDBMS available on every Android device. SQLite includes the following major features: (i) supporting most of the SQL-

92 standard, (ii) running on all major operating systems, (iii) having support for the major computer languages, and (iv) including multitasked read operations and sequential writes. The SQLite structure (n.d.b) consists of four components: (i) SQL Compiler, (ii) Core, (iii) Backend, and (iv) accessories. It can store data up to two TB with each database saved in a single disk using a B⁺ tree data structure. SQLite supports the following datatypes: NULL, INTEGER, REAL, TEXT, and BLOB. The SQL statements are compiled into assembly code executed on the SQLite virtual machine, Virtual Database Engine (VDBE). An SQLite database is private to the Android app that creates it. To share data with other apps, Android provides Content Provider that uses standardized ways to retrieve and manipulate the stored data.

Google Play Services

Google Play (Store) is a digital distribution platform for the apps developed by using Android platform and published through Google.

Figure 4. An Android emulator displaying a driving direction between two locations



In order to use many services provided by the Google and Android, Google Play services is required for app development and uses. It is used to update Google apps and apps from Google Play. Google Play services provides core functionality like authentication, synchronization, and user privacy setting. It is an APK (Android application package), a package file format used by the Android operating system for distribution and installation of mobile apps and middleware. APK files, a type of archive file, can be installed on Android powered devices just like installing software on PC. Google Play services allows Android apps to easily integrate with Google services like Google Maps. Apps may not work if Google Play services is not installed.

A SIMPLE ANDROID APP EXAMPLE

This section gives an example of mobile app development using Android (Google, n.d.c). This application is based on the default template when an Android project is created. It has two different screens and is able to move from one

screen to another screen. Figure 5 shows the result screenshots. Two activities/pages are included in this application:

- HelloWorldActivity.java, which is the start-up page and
- NextPageActivity.java, which is the other page.

Other than the text, each of the pages contains a button, which is used to display the other page. The following five components of this app will be detailed next:

- *Android manifest file (AndroidManifest.xml)*, which is used to glue everything in the app,
- *Layout XML code (main.xml and next.xml)*, which defines the UI elements, their properties, and their arrangement, and
- *Java source code (HelloWorldActivity.java and NextPageActivity.java)*, which drives everything. It is this code that is ultimately converted to Dalvik executable and runs the application.

Figure 5. Two screenshots of the Android application



Android Manifest File

In its root directory, each Android application must have an AndroidManifest.xml file, whose main missions are to explain what the application consists of, what its major building blocks are, what permissions it requires, etc. It must declare the four components of an application:

- **Activity:** Which represents a single screen with a user interface,
- **Service:** Which runs in the background to perform long-running operations or to perform work for remote processes,

- **Content Provider:** Which manages a shared set of application data stored in the file system, SQLite database, or any persistent storage, and
- **Broadcast Receiver:** Which responds to system-wide broadcast announcements.

For example, if the application wants to access the Internet, this file must state that it would like to use the related permission, as presented in Box 1.

Layout XML Code

The user interface for activities is defined via layouts. The layout defines the UI (user interface)

Box 1.

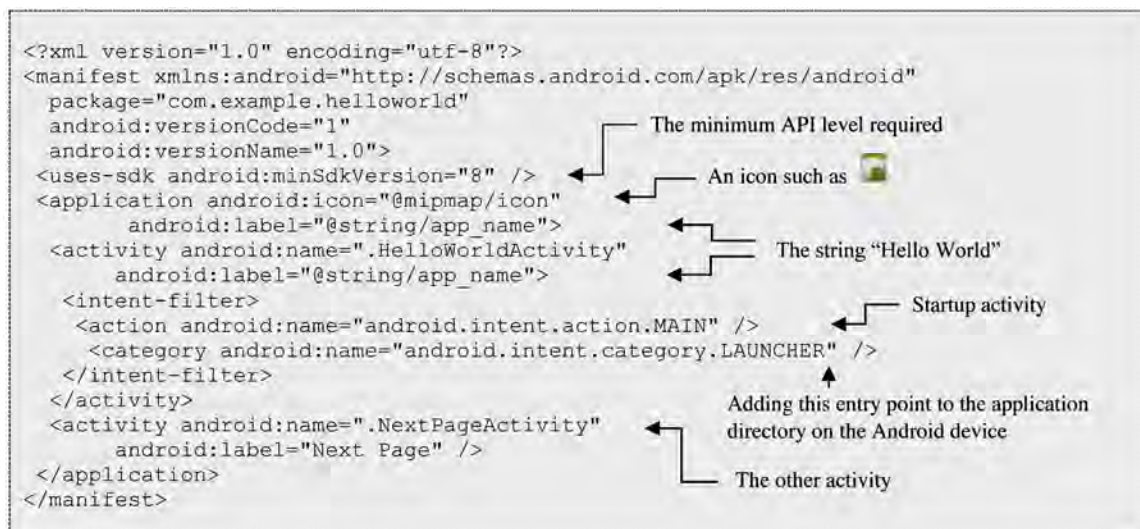


Figure 6. A screenshot of a layout generated by Java code



elements, their properties, and their arrangement. There are two ways, Java code or XML, to define the Android app layouts. The Java code is usually for the dynamic layouts (i.e., the layouts are unknown when coding) and the XML is normally for the static layouts (i.e., the layouts are known in advance).

- **Java Code:** Java code is typically used to generate the layout when the content is unknown until runtime; e.g., if the layout depends on content which is read from the Internet. An example of this is the next three statements whose screenshot is shown in Figure 6:

```
TextView tv = new TextView(this);
tv.setText("Hello, Android");
setContentView(tv);
```

which compose the string “Hello, Android” and display it.

- **XML:** XML based layouts are defined via a resource file in the folder `/res/layout`. This file specifies the view groups, views, their relationship, and their attributes for a

specific layout. Defining layouts via XML is usually the preferred way as this separates the programming logic from the layout definition. For example, the following two layouts are used in this application:

- Home page at `res/layout/main.xml` and
- Next page at `res/layout/next.xml`

The application loads the screen specified in `main.xml` by the following Java command in the `HelloWorldActivity.java`:

```
setContentView(R.layout.main);
```

The following code, presented in Box 2, shows the contents of the file `main.xml`, one of the two screens of this application.

Java Source Code

The Java code is what drives everything. This application includes two screens/activities/classes:

- `HelloWorldActivity.java`, which is the main page, and

Box 2.

```
<?xml version="1.0" encoding="utf-8"?>
<LinearLayout xmlns:android="http://schemas.android.com/apk/res/android"
    android:orientation="vertical"
    android:layout_width="fill_parent"
    android:layout_height="fill_parent" ← Taking up all space
    android:gravity="center"
    android:background="@android:color/white">
    <TextView
        android:layout_width="fill_parent"
        android:layout_height="wrap_content" ← Expanding only far enough to contain the values
        android:gravity="center"
        android:text="@string/hello" ← The string "Hello, World!\n"
        android:textSize="25sp"
        android:textStyle="bold"
        android:textColor="@android:color/black"
    />
    <Button android:id="@+id/next" ← The button's ID being "next"
        android:layout_width="wrap_content"
        android:layout_height="wrap_content"
        android:textSize="15sp"
        android:text="@string/button"
    />
</LinearLayout>
```

- NextPageActivity.java, which is the other page.

These two screens are similar. Each of the screens displays text and a button. The button is used to activate the other activity/class, which displays the other screen. Notice that the class is based on the Activity class. An Activity is a single application entity that is used to perform actions. An activity is not required to have a user interface, but usually does. The onCreate method is called by the Android system when the Activity starts—it is where all initialization and UI setup should be performed. The OnClickListener is an interface definition for a callback to be invoked when a view is clicked and the onClick method is activated when the button on the application interface is clicked. The code presented in Box 3 shows the file HelloWorldActivity.java (the NextPageActivity.java has similar code):

FUTURE RESEARCH DIRECTIONS

This chapter gives an empirical study of mobile app development, which is more difficult compared to the development of desktop applications. However, mobile IDEs (integrated development environments) are revised and improved constantly. Though it may not come anytime soon, mobile IDEs may be as powerful and convenient as desktop IDEs one day. Before the day is coming, mobile developers are forced to use the mobile IDEs available to them. This chapter helps readers to get ready for mobile app development. The authors will keep updating this article based on the up-to-date mobile IDE technology from time to time. At the same time, the developers also have to pay attention to mobile IDE development, so they could incorporate the newest technology to their app development.

Box 3.

M

```

package com.example.helloworld;

import android.app.Activity;
import android.content.Intent;
import android.os.Bundle;
import android.view.View;
import android.widget.Button;

public class HelloWorldActivity extends Activity {
    /** Called when the activity is first created */
    @Override
    public void onCreate( Bundle savedInstanceState ) {
        super.onCreate( savedInstanceState );
        setContentView( R.layout.main ); ← Displaying the screen of main.xml

        final Button button = (Button) findViewById( R.id.next );
        button.setOnClickListener( new View.OnClickListener() {
            public void onClick( View v ) { ← Called when the button is clicked
                /** Here i calls a new screen. */
                Intent i = new Intent(
                    HelloWorldActivity.this, NextPageActivity.class );
                startActivity( i ); ← Launching the activity NextPageActivity
            }
        } );
    }
}

```

CONCLUSION

Mobile application stores (or app stores) sell or provide mobile applications/services for handheld devices such as smartphones and tablet computers. The applications/services are not necessarily from the storeowners. Many of them are from the third parties such as independent developers. A wide variety of mobile applications is available on the stores. Popular applications include mobile games, mobile offices, mobile shopping, and music. Additionally, mobile commerce is a coming milestone after electronic commerce blossoming in the late 1990's. However, it is also commonly admitted that the development in this field is constrained. There are some considerable barriers waiting to be overcome. One of the barriers is most software engineers are not familiar with handheld programming, which is the programming for handheld devices. This chapter explains the design and implementation of mobile apps by detailing the two subjects: handheld computing and Android

application. Client-side handheld computing is to use handheld devices to perform handheld operations, which do not need the supports of server-side computing. Various environments/languages are available for client-side handheld computing and programming. The rising and demise of mobile operating systems happen quickly and frequently and are hard to predict. Three current mobile operating systems likely to succeed are

- *Android*, the most popular mobile operating system nowadays,
- *iOS*, from the prestigious company Apple, and
- *Windows Phone*, from the software behemoth Microsoft.

They apply different approaches to accomplishing the development of handheld applications and it is almost impossible to predict which approaches will dominate the client-side handheld computing in the future, as the Windows to desktop PCs. Most

client-side handheld programming languages are a version of either C/C++ or Java. This chapter introduces the essential technologies for mobile/handheld app development and gives a simple example of Android application development. The given example shows handheld programming is complicated and difficult compared to desktop programming. Many developers thus try to stay away from client-side programming by switching to web apps (Shankland, 2012). However, this approach has the disadvantage of slowness and requires Internet accesses.

REFERENCES

- Apple, Inc. (n.d.). *iOS 9—Apple*. Retrieved October 28, 2015, from <http://www.apple.com/ios/>
- Google. (n.d.a). *Android*. Retrieved December 14, 2015, from <http://www.android.com/>
- Google. (n.d.b). *Download Android Studio and SDK Tools - Android Developers*. Retrieved November 9, 2015, from <http://developer.android.com/sdk/>
- Google. (n.d.c). *Android Developers*. Retrieved February 04, 2016, from <http://developer.android.com/index.html>
- Hu, W.-C. (2016). *Handheld Computing Research*. Retrieved March 16, 2016, from <http://wenchen.cs.und.edu/handheldresearch/>
- Hu, W.-C., Yeh, J., Chu, H.-J., & Lee, C. (2005). Internet-enabled mobile handheld devices for mobile commerce. *Contemporary Management Research*, 1(1), 13–34. doi:10.7903/cmr.70
- Hu, W.-C., Zuo, Y., Chen, L., & Yang, H.-J. (2010, January). Contemporary issues in handheld computing research. *International Journal of Handheld Computing Research*, 1(1), 1–23. doi:10.4018/jhcr.2010090901
- Kiely, D. (2001). Wanted: Programmers for handheld devices. *IEEE Computer*, 34(5), 12–14. doi:10.1109/2.920605
- Microsoft. (n.d.). *Windows Phone*. Retrieved March 12, 2016, from <https://www.microsoft.com/en-us/windows/phones>
- Open Handset Alliance. (n.d.). *Open Handset Alliance*. Retrieved November 06, 2015, from <http://www.openhandsetalliance.com/>
- Research In Motion. (n.d.). *BlackBerry 10 OS*. Retrieved December 15, 2015, from <http://us.blackberry.com/software/smartphones/blackberry-10-os.html>
- Shankland, S. (2012, March 20). *Survey: Android Programmers Shifting toward Web Apps*. Retrieved March 14, 2015, from http://news.cnet.com/8301-30685_3-57400136-264/survey-android-programmers-shifting-toward-web-apps/
- SQLite. (n.d.a). *SQLite*. Retrieved October 12, 2016, from <http://sqlite.org/>
- SQLite. (n.d.b). *The Architecture of SQLite*. Retrieved November 28, 2015, from <https://www.sqlite.org/arch.html>
- TechCrunch. (2015, June 8). *iTunes App Store Passes 1.5M Apps, 100B Downloads, \$30B Paid to Developers to Date*. Retrieved August 30, 2015, from <http://techcrunch.com/2015/06/08/itunes-app-store-passes-1-5m-apps-100b-downloads-30b-paid-to-developers/>

ADDITIONAL READING

- Apple, Inc. (n.d.). *iOS Dev Center*. Retrieved November 30, 2013, from <https://developer.apple.com/devcenter/ios/index.action>
- Conway, J., Hillegass, A., & Keur, C. (2014, February 21). *iOS Programming: The Big Nerd Ranch Guide, 4th Edition*. Big Nerd Ranch Guides.
- Mednieks, Z., Dornin, L., Meike, G. B., & Nakamura, M. (2012). *Programming Android*. O'Reilly Media, Inc.

Microsoft. (n.d.). *Windows Phone Dev Center*. Retrieved December 08, 2013, from <https://dev.windowsphone.com/>

Phillips, B., Stewart, C., Hardy, B., & Marsicano, K. (2015, August 3). *Android Programming: The Big Nerd Ranch Guide, 2nd Edition*. Big Nerd Ranch Guides.

Research In Motion, Limited. (n.d.). *BlackBerry Developer*. Retrieved December 13, 2013, from <https://developer.blackberry.com/>

StackOverflow. (n.d.). *Newest Android Questions*. Retrieved March 14, 2016, from <http://stackoverflow.com/questions/tagged/android>

Whitechapel, A. (2012). *Windows Phone 7 Development Internals*. Microsoft Press.

KEY TERMS AND DEFINITIONS

Android Studio IDE (Integrated Development Environment): Since 2014, Android Studio became the primary IDE for native Android application development. It is free software under the Apache License 2.0 developed by Google. It includes the following features: intelligent code editor, code templates and GitHub integration, multi-screen app development, virtual devices for all shapes and sizes, and Android builds evolved, with Gradle.

Android: It is a software stack for mobile devices that includes an operating system, middleware, and key applications such as contacts. It is a project proposed by the Open Handset Alliance, a group of more than 84 technology and mobile companies including Google, Inc.

Apps: A mobile app (or application) is a kind of software designed to run on mobile handheld devices such as smartphones. Examples of apps

are calendars, video games, and short message services (SMS).

Client-Side Handheld Programming: It is design and development of handheld/mobile software such as Android programs that reside on the handheld/mobile devices.

iOS (Previously iPhone OS): It is a mobile operating system developed by Apple Inc. for its mobile/handheld devices such as iPhone and iPad. iOS is derived from OS X (used for Apple's computers such as iMac and MacBook) and both are based on the Unix. It includes four abstraction layers: (i) the Core OS layer, (ii) the Core Services layer, (iii) the Media layer, and (iv) the Cocoa Touch layer. iOS includes a user interface interacted with multi-touch gestures such as swipe, tap, and pinch, all of which have specific definitions in iOS.

Mobile Handheld Devices: They are small general-purpose, programmable, battery-powered computers, but they are different from desk- or lap- top computers mainly due to the following special features: (i) limited network bandwidth, (ii) small screen/body size, and (iii) mobility.

Mobile/Handheld Computing: It is to use handheld devices such as smartphones to perform wireless, mobile, handheld operations such as managing personal data, playing video games, and browsing the Internet.

Server-Side Handheld Programming: It is design and development of handheld/mobile software such as CGI programs that reside on the servers.

Windows Phone: Windows Phone is a mobile operating system developed by Microsoft and is aimed at the consumer market. It is derived from Windows Mobile platform, which focuses on the enterprise market instead. Windows Phone includes a new user interface of flat, colored live tiles and a laterally scrolling canvas for accommodating more tiles.

Enhancing the Mobile User Experience Through Colored Contrasts

Jean-Éric Pelet

ESCE International Business School, France

Basma Taieb

University of Cergy Pontoise, France

INTRODUCTION

In 1940, during the Second World War, the Cambridge zoologist Hugh Cott published a textbook about camouflage, warning coloration and mimicry. He introduced the notion of disruptive coloration, which was defined as contingent on background matching, suggesting

the effect of a disruptive pattern is greatly strengthened when some of its components closely match the background, while others differ strongly from it. Under these conditions, by the contrast of some tones and the blending of others, certain portions of the object fade out completely while others stand out emphatically (p.50).

As stated by Schaefer & Stobbe (2006), disruptive coloration and background matching are both techniques of camouflage that are often quoted as textbook examples of natural selection. For example, the experiments conducted by Cuthill *et al.* (2005) illustrated that artificially colored moths, which had been designed to match their background of oak trunks in terms of color and brightness, survived better if they sported highly contrasting patterns on the edge of their wings compared to moths with the same patterns inside their wings. Furthermore, this phenomenon of contrast concerns all natural species in various fields and even human beings, such as soldiers in the army or surgeons in their operating room.

For example, military camouflage patterns were worn in the army to protect personnel and equipment from observation by enemy forces. Green uniforms, especially, and later other drab colors were used depending on the field of war (in Vietnam khaki green colored uniforms were favored, whereas in Irak khaki yellow was privileged). Surgeons operate in completely green or blue operating rooms to allow them to focus on what is red, namely blood and organs. This can be explained by the physiology of the human eye, notably the retina, which is lined with photoreceptor cells known as rods and cones. The rods are stimulated by dim light. In parallel, cones run “at full throttle” in normal brightness. Because human beings have three types of cones with different levels of sensitiveness to red, green and blue - red-sensitive cones are the most numerous, followed by the green-sensitive, and finally the blue-sensitive cones - the red color captures the surgeon’s attention more easily than the other colors. Thus, having an operating room with a dominant green or blue color allows a surgeon’s eyes to focus on the red color of blood and therefore on the surgery itself. The same question of contrast occurs on a webpage such as an e-commerce webpage, enabling users to read more or less easily, thanks to a properly chosen contrast ratio between the foreground and background colors.

“A frustrating experience on a website hurts my opinion of the brand overall” says Miller (2012). The author adds that many customers of

a brand are disappointed in the company itself if the mobile experience doesn't meet their expectations. The mobile experience or "*Mobile User Experience*" (MUE) refers to the perception users have of a mobile product or service, such as an app, a social media or a website, for example an e-commerce website. This perception mainly leans on the sense of sight. The principal variable among the viewable features of the Cascade Style Sheet (CSS) used to write formatting instructions (rules) for websites is color. Indeed most of the information available on a website comes from what is viewable: in fact, 80% of the information processed by an Internet user's brain results from sight (Mattelart, 1996), making color the main variable to take into account in research on consumer behavior when using a handled device for shopping purposes. Bearing in mind that the background color is the first factor to become apparent when the download of a webpage is in progress (Gorn *et al.*, 2004), thinking about colors to improve the MUE and leverage the benefits of companies is fundamental for the Return on Investment (ROI) of an organization. Nevertheless, attention given to the background color itself and the various parameters that summarize it (such as hue, brightness and saturation) is not sufficient to improve the customer experience. In order to answer this question, the present research compares the design cues of a m-commerce website on its color contrast occurred between foreground and background. An experimental design was developed to investigate the effects of m-commerce website design on emotions and behavioral intention within the context of smartphone usage.

This chapter therefore presents a literature review on the importance of the colored contrast of mobile website interfaces. The methodology section then presents the website built for the experiment, followed by results aiming at highlighting the behavioral intentions derived from the presentation of positive and negative colored contrasts. We conclude the paper with a discussion, some limitations and suggestions for future research, which give rise to some managerial implications.

BACKGROUND

Definition of the Mobile User Experience (MUE)

The term "mobile commerce" (m-commerce) was introduced in 1997 by Kevin Duffey at the Global Mobile Commerce Forum. According to Duffey's definition, m-commerce is an area separated from the e-commerce (electronic commerce) market, which is characterized by the carrying out of all operations by means of wireless devices, mainly smartphones and tablets (Shaikh & Gupta, 2014). M-commerce is specifically mobile-phone access to a wide spectrum of services covering finance (mobile banking), entertainment (music, movies, e-publications), and information and localization services (Pralat, 2013). Among mobile consumers, another action is increasingly practiced, namely "showrooming", where people use smartphones in traditional shops to search for information about the products on the Internet, compare prices and, in cases where they find an attractive offer, make their purchases online instead of at the shop (Pralat, 2013). The key to improving the MUE of e-commerce websites visited on a laptop/desktop interface seems to lie partly in the contrast that occurs between the background and foreground colors (Pelet, 2014). However, until now the effects of the colored contrasts occasioned by the foreground and background colors of the CSS of a website on consumer behavior have not been studied.

Total retail sales are increasing slowly but steadily in the USA according to e-Marketer (2015). Although sales are growing most quickly in the digital market, e-commerce is expected to rise only slightly as a share of the total, from 7.2% this year to 9.8% by 2019. Mobile still accounts for 1.6% of all retail, and by 2019 its share is expected to be 2.7%. According to Gentle (2015), current m-commerce figures indicate that a major shift is taking place, since the percentage of smartphone owners that use their phones to access the Internet was seen to double between

2009 and 2013, from 31 percent to 63 percent (Duggan and Smith, 2013). Gentle adds that in 2014, over 52 percent of e-retail traffic over four days, between Thanksgiving (Thursday, November 24, 2016) and Cyber Monday (Monday, November 28, 2016) was mobile (Siwicki, 2014). The editor underlines that current studies show that 67 percent of consumers say they are more likely to buy from a mobile-friendly site.

Customers are five times more likely to visit a competitor after having a frustrating mobile experience (Gentle, 2015). Consumers will more readily buy online when the website meets their mobile needs. 'Mobile needs' refers to the optimization of the page user's view on their mobile screen. This is called the "landing page" and its main function lies in optimizing the content of a website for any screen size, from watches and huge TV screens, to the screens on mobiles and tablets. For web designers or those in charge of digital content, optimizing the landing page therefore means relying on certain practices. Not respecting these guidelines can work against the brand. When a website fails to follow the guidelines from key industry players such as Google, it is penalized in several areas, for instance SEO (Search Engine Optimization) and it may subsequently not appear on the first page of major search engines. Users may stay longer thanks to the ease-of-use of a website, making them more likely to buy a product or use a service. They might also quit if they do not see what they are looking for right away on the website. Users therefore quit a website for ease-of-use reasons; and this is why the mobile-friendly aspect of its interface remains a key issue in retaining users for as long as possible. It is also a reason why three quarters of respondents said they are more likely to revisit a mobile-friendly website, in a survey managed by Google (2011). A mobile-friendly site encompasses several aspects to go beyond simply what the landing page promises, by taking into account content and layout that are compatible with various devices and screens

of all sizes. This is all about "responsive design", and includes touch, gesture and voice activation controls. The biggest changes that have occurred between laptop and desktop screens and mobile devices include the exploitation of another sense in addition to sight and hearing, namely touch. However, this sense is activated through visual data, since users touch the screen depending on what they see or read on the webpage.

Reading on the screen of a smartphone in the street can sometimes prove to be extremely challenging, especially when the sun shines, and when legibility conditions are not very good. Brands that do not pay attention to the readability of the content of their m-commerce (mobile commerce) website may lose consumers' intentions to buy and revisit as well as their intentions to recommend the website. In this respect, the contrast ratio created by the foreground (text) and background colors of the website needs to be carefully designed. As already mentioned by Pelet (2014) regarding e-commerce, the color contrasts of a website enhance the memorization of information and the intention to purchase (Pelet & Papadopoulou, 2010, 2012). Especially in the context of web design, color serves as a supportive medium for the layout of information. Hence, it works as a contrasting base. Reading time is shorter when there is good contrast. Some studies show that a positive contrast obtained with a light text over a dark background provokes online visitors to abandon a website and puts users in a negative mood (Pelet, 2010). A positive contrast inflicts a visual strain accompanied by eye fatigue, contrary to a negative contrast (a dark text over a light background). This is why we choose to distinguish the background and foreground colors in the experiment we present below by using both negative and positive contrasts to conduct our measurements. The positive contrast is yellow on green, and the negative one green on yellow, following Hill & Scharff's (1997) and Pelet & Papadopoulou's (2012) findings.

Controlling the Contrast Ratio of a Mobile Website's Interface

The legibility of the content is of paramount importance for MUE. By increasing the color contrast when designing websites, the readability and legibility of the information presented are improved, which directly correlates with the information that is absorbed and retained by consumers (Chadwick-Dias *et al.*, 2007; Gao *et al.*, 2011; Pelet and Papadopoulou, 2012). Researchers have already worked on the contrast of web interfaces (see for example Scharff & Ahumada, 2002: authors offer measures that predict the effects of background contrast variations on text readability; July 2003: authors measured text readability for positive and negative contrast text on a plain background at two contrast levels, 30% and 45%; 2003: authors show that letter identification performance is better for negative contrast (black text on gray background) than positive contrast (white text on gray background); 2003a: authors found that word identification and word discriminability was affected by the backgrounds in the same way that the paragraph search performance was affected, but that letter identifiability on these two backgrounds was predicted by the metric. They have also found a significant improvement from including different contrast gains for positive and negative contrasts in the metric. Unfortunately, word readability is not necessarily simply related to letter identifiability and simple contrast measures according to authors; 2003b; 2005: authors measured text legibility for light text and dark text. For paragraph readability and letter identification, responses to light text were slower and less accurate for a given contrast.; 2008; 2009; Ahumada, Scharff & Watson, 2006; Scharff, Ahumada & Hill, 1999; Scharff, Hill and Ahumada, 2000; Zuffi *et al.*, 2006; 2007: authors confirm previous works and find that letter identification latencies are predicted by an asymmetric contrast metric). None of these research projects relate to mobility and external environment conditions, which affect the contrast of the screen.

Legibility is the degree to which glyphs (individual characters) in the text are understandable or recognizable, based on appearance. The legibility of a typeface is related to the characteristics inherent in its design, which relate to the ability to distinguish one letter from another. Readability is the ease with which text can be read and understood. It is influenced by line length, primary and secondary leading, justification, typestyle, kerning, tracking, point size, etc. Thus, readability is increased with legibility. Also, reading is more enjoyable due to the correlation with the improved legibility. Users should not be obliged to zoom when discovering content from an m-commerce website interface or typing on the screen. Actionable targets should be clickable without zooming. This means that the text size and the colored contrast between the text and the background color have to be strong enough to facilitate the readability of the content.

Among the few basic factors to consider when trying to improve MUE, accessibility proves to be essential to ensure that the content is available and visible on mobiles. It can be improved thanks to the “WCAG AA” Accessibility Standards. Web Content Accessibility Guidelines (WCAG) 2.0 define how to make Web content more accessible to people with disabilities.

“Following these guidelines will make content accessible to a wider range of people with disabilities, including blindness and low vision, deafness and hearing loss, learning disabilities, cognitive limitations, limited movement, speech disabilities, photosensitivity and combinations of these. Following these guidelines will also often make your Web content more usable to users in general” says W3C (2016).

These guidelines provide the necessary basics of any website, whether it will feature on smartphones, tablets, laptops, or desktop devices.

In order to improve MUE, speed of download turns out to be important too. As for e-commerce on large screens (laptop/desktop), a significant criticism that consumers raise about using the Internet

is that it often takes a long time to download web pages, according to Dennis (2001) in Gorn *et al.* (2014). One of the mobile user's primary goals is to obtain information as quickly as possible while completing online tasks on the go. This particular problem is only becoming worse as webpages become more and more technical, with features enabling more interactivity or the integration of rich content (video, animation), increasing the page weight and its downloading time. A proper color contrast enables users to find the information more quickly, diminishing the frustration of a slow downloading time. The initial feelings experienced by a user is crucial because during the first 50 milliseconds users decide whether or not to continue browsing the website (Lindgaard *et al.*, 2006). The aim of designers is therefore to provide interactive systems that are not only useful and usable, but also capable of transmitting positive emotions through their aesthetic characteristics. The factors that slow down mobile pages encompass the number of requests (loaded video, images, scripts, etc.) followed by the quantity of bytes (page size).

The MUE of an m-commerce website also benefits from its structure. If the main information users/consumers need to locate is at the top of the webpage, users do not have to scroll down and run the risk of starting to feel irritated by not discovering the content they expect to find. This region is also called "*above the fold*" by web designers. It is characterized by a thin zone directly at the top of any website. A single-column layout providing the priority content at the top in an easy-to-read way can be a solution to retain the engagement of users who do not expect to scroll down systematically when they look for a product or service.

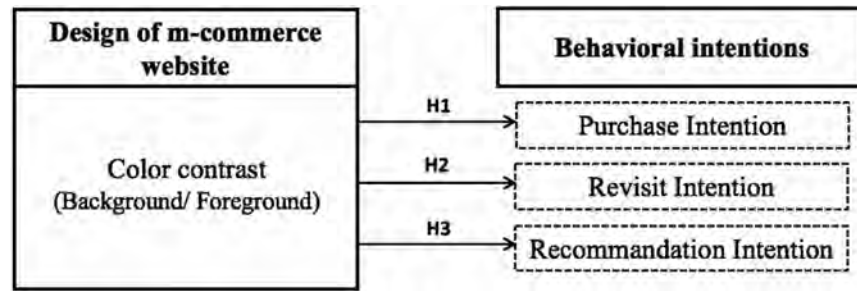
Diminishing the content to be read, whether with titles or paragraphs, also plays in favor of a greater MUE. Since less content to read also means

less tiredness, users may be sensitive to shorter texts provided on an m-commerce website, which preserve the principles of visual and cognitive ergonomics (Ling and Schaik, 2002). Leveraging creative copywriting and considering visual symbols such as icons to replace or reduce the copy represent interesting solutions, and this is why some actors are becoming key factors of success for successful interfaces. Some websites¹ provide toolkits for web designers in order to entice them to use vectors rather than images. Such actions help to diminish the size of a webpage. Indeed, these vectors give web designers scalable vector icons that can be instantly customized thanks to particular variables such as text size, color, drop shadow, and anything that can be done with the CSS.

Besides traditional landing page optimization fundamentals, web designers focusing on mobile tactics should concentrate on simple usability (reading/clicking), fast performance, minimal yet compelling action and easy-to-read content, thanks to a strong contrast between the foreground and background. Practitioners are advised to emphasize strong contrast between the foreground and background, especially for text messages. An important factor is appearance, particularly the proper choice of font and color contrast, which creates reliability and enhances e-service quality or perception if the website is user-friendly (Lowry *et al.*, 2014). Thus, we can hypothesize:

- H1:** The color contrast of the mobile website positively influences purchase intentions.
- H2:** The color contrast of the mobile website positively influences revisit intentions.
- H3:** The color contrast of the mobile website positively influences recommendation intentions.

Figure 1. Conceptual model



Following on from these hypotheses, we propose the following conceptual model:

METHODOLOGY

Conditions of the Survey, Stimuli and Data Collection

Respondents answered the questionnaire in the real-life, outdoor conditions of using a smartphone. For this reason, respondents were tested for color blindness using the Ishihara test² (Pelet, 2010). This guaranteed the validity of our sample's responses, by retaining only those people with perfect color vision. In order to control environmental variables, the ambient lighting³, humidity⁴ and temperature⁵ were also measured during each respondent's visit. These measurements enabled us to neutralize variables and were reported on each questionnaire of every respondent.

Two versions of a mobile website specializing in the sale of music CDs were especially designed for this experimental study: a site with a positive contrast (yellow text on a green background) and another with a negative contrast (green text on a yellow background). The experimental site contains 30 CDs available in 10 categories (3 CDs / category). For each CD, participants could see the CD cover, the album title, the artist's name and other details (music category, status (new or used), price, delivery time, etc.).

312 participants visited the site's two versions: 160 participants for the site with positive contrast

and 152 participants for the site with negative contrast. Each participant was required to view the details of at least two CDs of their choice and add them to their cart without making real purchases. After viewing at least two CDs, a link appeared asking them to respond to a questionnaire.

The questionnaire includes measures related to site design (5 items borrowed from Bressolles, 2006), intention to revisit (3 items of Mukherjee & Nath, 2007), purchase intent (4 items of Limayem & Rowe, 2006) and intention to recommend (3 items of Goyette *et al.*, 2010).

Validation of Measurement Scales

The measurement scales have a good internal consistency (Cronbach α and ρ of Jöreskog are above 0.7). The average variance extracted (AVE) by items is greater than 0.5. The variance extracted by each construct is greater than the squared correlations among constructs (Fornell & Larker, 1981), thus discriminant validity was supported. The indices for the measurement model also indicate a good fit (RMSEA = .08, RMR = .03, $\chi^2/df = 3.27$, $p < .01$).

Hypothesis Testing

The hypotheses were tested using the method of structural equations (AMOS). Their validity was verified for mobile websites with positive and negative contrast. The results are presented in Table 1.

Table 1. Test of the hypotheses

	Site With Negative Contrast (Green Text on a Yellow Background)				Site With Positive Contrast (Yellow Text on a Green Background)			
	Estim	S.E.	C.R.		Estim	S.E.	C.R.	
Revis<--- Desig	.10	.04	2.73**	H1 partially supported	.23	.03	7.00**	H1 supported
Buy<--- Desig	.02	.04	0.36		.20	.04	5.30**	
Recom<--- Desig	.16	.04	4.40**		.17	.03	5.21**	

Note: ** p <.01 ; * p <.05

Results show that the website's positive contrast has a significant impact on purchase, revisit and recommendation intentions. Unlike the site with positive contrast, the website's negative contrast does not influence purchase intent.

FUTURE RESEARCH DIRECTIONS

Our results show that the color contrast of a mobile commerce website, in particular a positive contrast, strongly influences the consumer's behavioral intentions in external conditions. It should also be noted that the consumer would be more inclined to buy from the mobile website with a positive contrast (in our case, a yellow text on a green background). On a mobile website, in the outdoor conditions of the experimentation, respondents have better behavioral intentions when the contrast ratio is a bright text over a dark background (positive contrast), which is the opposite of most actual mainstream websites, which use dark texts such as black over bright backgrounds such as white (negative contrast) (see figure 1):

In conditions where the sun shines and where it is often quite difficult to read the smartphone screen, users seem to prefer positive contrasts. Interestingly, this contrast exists in any smartphone as an "accessibility" option, in the "View" folder of the "Settings" section when pressing the "Negative Colors" button. However, it is not very easy to set up, especially when the lighting conditions are difficult, such as when the sun shines. An initial managerial implication of this research could therefore be to make this service available when the image sensor of the smartphone detects the sunshine and automatically switches it on when it reaches a certain level. This would enhance the legibility of the screen and protect users' eyes at the same time.

A second managerial implication that comes from this finding is that it could save the smartphone battery somewhat. It uses more energy to show a bright screen than a dark one. This is why a smartphone turns itself off relatively quickly when we do not touch its screen: in order to preserve the battery life, the more it shows a black screen, the less energy it uses. A website

Table 2. Comparison of Contrast polarity, External Environmental Condition and Interface used for the experimentation

Contrast Polarity	Site with negative contrast (Green text on a yellow background)	Site with positive contrast (Yellow text on a green background)
External Environmental Condition	Indoor conditions (laptop/desktop) Insert Image 2 (laptop_no_300dpi.tif) here	Outdoor conditions (mobile devices) Insert Image 3 (laptop_yes_300dpi.tif) here
Interface	Insert Image 4 (css0_Diana_300dpi.tif) here	Insert Image 5 (css7_Diana_300dpi.tif) here

that is predominantly made up of dark colors rather than bright colors may help to save the battery life and therefore preserve the environment: it becomes an “eco-friendly” strategy for the brands which work in this way and make users “Socially Conscious Consumers”, as depicted by Brooker (1976).

Our study contributes to a better understanding of consumer behavior surrounding mobile websites. Nevertheless, some limitations may be noted. Consideration of other variables such as feelings of privacy and perceptions of waiting time could provide more details about the purchasing behavior of people with sight problems. Maintaining constant levels of brightness and saturation, which are two components of color (along with hue), to study the overall effect of color on behavioral intentions without distinguishing the effect of text color from background color is one limitation of this research. Future studies are advised to consider other colors and to separately examine the effect of background color, and text color, as well as the interaction between the two. Moreover, there is still no agreement on whether human reactions to color are innate or learned and whether they can be conditioned by each individual’s social experience (Crozier, 1999). Cultural aspects must also be taken into consideration insofar as we perceive color differently according to our origins.

CONCLUSION

Whereas smartphones continue to offer more and more graphic possibilities thanks to their ever-growing power, enabling them to show more than 16 billion colors, our eyes enable us to distinguish only 8 billion (Chrisment *et al.*, 1994). Among these, it is easy to create contrast ratios that enable users suffering from color blindness, as well as other sight problems, to exploit free and easy-to-use tools (for a list of these, please visit <http://www.scoop.it/t/color>). As already described

by Pelet & Papadopoulou (2012), by taking into account the W3C and Web Accessibility Initiative (WAI) guidelines⁶, the use of color becomes more professional and the choices web designers make are more informed in terms of usability, as well as in terms of human-computer interaction in general.

ACKNOWLEDGMENT

The authors would like to thank reviewers from the « First International Conference on Advanced Research Methods and Analytics », CARMA 2016, Universitat Politècnica de València, València, Spain, 2016. DOI: <http://dx.doi.org/10.4995/CARMA2016.2016.3110> for their thorough review which helped to improve the research. The present chapter is an enhanced version of the paper published prior to the aforementioned conference, available at: <http://ocs.editorial.upv.es/index.php/CARMA/CARMA2016/paper/view/3110>

REFERENCES

- W3C. (2016). *Web Content Accessibility Guidelines (WCAG) 2.0*. Retrieved from the Internet at <https://www.w3.org/TR/WCAG20/>
- Ahumada, A. J., Scharff, L. F. V., & Watson, A. B. (2006). *What pattern does the visual system detect best?* European Conference on Visual Perception. *Perception*, 35, 59.
- Ahumada, A. J., Watson, A. B., & Scharff, L. V. (2006, November). Which Visual Images are Detected Best? *Abstracts of the Psychonomic Society*, 11, 37.
- Bressolles, G. (2006). La qualité de service électronique: NetQu@l Proposition d'une échelle de mesure appliquée aux sites marchands et effets modérateurs. *Recherche et Applications en Marketing*, 21(3), 19–45. doi:10.1177/076737010602100302

- Brooker, G. (1976). The self-actualizing socially conscious consumer. *The Journal of Consumer Research*, 3(2), 107–112. doi:10.1086/208658
- Chadwick-Dias, A., Bergel, M., & Tullis, T. S. (2007). Senior surfers 2.0: A re-examination of the older Web user and the dynamic web. In C. Stephanidis (Ed.), *Universal access in human computer interaction: Coping with diversity* (pp. 868–876). Berlin: Springer. Retrieved from <http://www.springerlink.com/content/108q65578604130w8/>
- Chrisment, A., Durchon, P., Lanthony, P., & Tavernier, I. (1994). *Communiquer par la couleur - Mesurer, Reproduire, Observer, Vivre la couleur*. Paris: 3C Conseil.
- Cott, H. B. (1940). Adaptive Coloration in Animals. Methuen.
- Crozier, W. R. (1999). The meanings of color: Preferences among hues. *Pigment and Resin Technology*, 28(1), 6–14. doi:10.1108/03699429910252315
- Cuthill, I. C., Stevens, M., Sheppard, J., Maddocks, T., Párraga, C. U., & Troscianko, T. S. (2005). Disruptive coloration and background pattern matching. *Nature*, 434, 72–74. doi: 10.1038/nature0331.
- Deeb, S. S., & Motulsky, A. G. (2015). *Red-Green Color Vision Defects*. Retrieved from the Internet at <http://www.ncbi.nlm.nih.gov/books/NBK1301/>
- Dennis, S. (2001). Almost 50 Percent of Online Purchases Aborted. *Computer User*. Available at http://snipurl.com/dennis_S_2000
- Duggan, M., & Smith, A. (2013). *Cell Internet Use 2013*. Retrieved from <http://www.pewinternet.org/2013/09/16/cell-internet-use-2013/>
- e-Marketer. (2015). *Commerce Snapshot: October 2015*. Retrieved from the Internet at www.emarketer.com
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *JMR, Journal of Marketing Research*, 18(1), 39–50. doi:10.2307/3151312
- Gao, Q., Ebert, D., Chen, X., & Ding, Y. (2011). Design of a Mobile Social Community Platform for Older Chinese People in Urban Areas. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 25(1), 66–89.
- Gentle, T. J. (2015). *Responsive Website Design Drives M-commerce*. Published on January 14, 2015 at <https://blogs.adobe.com/digitalmarketing/digital-marketing/responsive-website-design-drives-m-commerce/>
- Google. (2011). *The Mobile Movement: Understanding Smartphone Users*. Google/IPSOS OTX MediaCT. Retrieved from the Internet at http://www.gstatic.com/ads/research/en/2011_TheMobileMovement.pdf
- Gorn, G., Chattopadhyay, A., Sengupta, J., & Tripathi, S. (2004). Waiting for the web: How screen color affects time perception. *JMR, Journal of Marketing Research*, 41(May), 215–225. doi:10.1509/jmkr.41.2.215.28668
- Goyette, I., Ricard, L., Bergeron, J., & Marticotte, F. (2010). e-WOM Scale: word of mouth measurement scale for e-services context. *Canadian Journal of Administrative Sciences/Revue Canadienne des Sciences de l'Administration*, 27(1), 5–23.
- Hill, A., & Scharff, L. V. (1997). Readability of websites with various foreground/background color combinations, font types and word styles. *Proceedings of 11th National Conference in Undergraduate Research*, 2, 742–746.
- Kim, H., Choi, Y., & Lee, Y. (2015). Web atmospheric qualities in luxury fashion brand websites. *Journal of Fashion Marketing and Management*, 19(4), 384–401. doi:10.1108/JFMM-09-2013-0103
- Koufaris, M. (2002). Applying the technology acceptance model and flow theory to online consumer behavior. *Information Systems Research*, 13(2), 205–224. doi:10.1287/isre.13.2.205.83

- Li, M., Dong, Z. Y., & Chen, X. (2012). Factors influencing consumption experience of mobile commerce: A study from experiential view. *Internet Research*, 22(2), 120–141. doi:10.1108/10662241211214539
- Limayem, M., & Rowe, F. (2006). Comparaison des facteurs influençant les intentions d'achat à partir du Web à Hong Kong et en France: Influence sociale, risques et aversion pour la perte de contact. *Revue Française du Marketing*, 209(4/5), 25–48.
- Ling, J., & Schaik, P. V. (2002). The effect of text and background colour on visual search of web pages. *Displays*, 23(5), 223–230. doi:10.1016/S0141-9382(02)00041-0
- Lowry, P. J., Wilson, D. W., & Haig, W. L. (2014). A picture is worth a thousand words: Source Credibility Theory applied to logo and website design for heightened credibility and consumer trust. *International Journal of Human-Computer Interaction*, 30(1), 63–93. doi:10.1080/10447318.2013.839899
- Mattelart, A. (1996). *The Invention of Communication*. London: University Minnesota Press.
- Mayer, J. D., & Gaschke, N. Y. (1988). The experience and meta-experience of mood. *Journal of Personality and Social Psychology*, 55(1), 102–111. doi:10.1037/0022-3514.55.1.102 PMID:3418484
- Mazaheri, E., Richard, M. O., & Laroche, M. (2011). Online consumer behavior: Comparing Canadian and Chinese website visitors. *Journal of Business Research*, 64(9), 958–965. doi:10.1016/j.jbusres.2010.11.018
- Miller, M. (2012). 72% of Consumers Want Mobile-Friendly Sites: Google Research. Retrieved from the Internet July 17, 2016 at <https://searchenginewatch.com/sew/study/2208496/72-of-consumers-want-mobilefriendly-sites-google-research>
- Mukherjee, A., & Nath, P. (2007). Role of electronic trust in online retailing: A re-examination of the commitment-trust theory. *European Journal of Marketing*, 41(9/10), 1173–1202. doi:10.1108/03090560710773390
- Pelet, J.-É. (2010). Effets de la couleur des sites web marchands sur la mémorisation et sur l'intention d'achat. *Systèmes d'Information et Management*, 15(1), 97–131. doi:10.3917/sim.101.0097
- Pelet, J.-É. (2014). Investigating the Importance of Website Color Contrast in E-Commerce: Website Color Contrast in E-Commerce. In *Encyclopedia of Information Science and Technology* (3rd ed.). Hershey, PA: IGI Global.
- Pelet, J.-É., & Papadopoulou, P. (2012). The effect of colors of e-commerce websites on consumer mood, memorization and buying intention. *European Journal of Information Systems*, 21(4), 438–467. doi:10.1057/ejis.2012.17
- Prałat, E. (2013). *M-commerce – rozwój na świecie i w Polsce*. Retrieved from <http://www.ur.edu.pl>
- Schaefer, H. M., & Stobbe, N. (2006). Disruptive coloration provides camouflage independent of background matching. *Proceedings of the Royal Society B, Biological Sciences*. doi:10.1098/rspb.2006.3615
- Scharff, L., & Ahumada, A. (2008). Contrast polarity in letter identification. *Journal of Vision (Charlottesville, Va.)*, 8(6), 627. doi:10.1167/8.6.627
- Scharff, L., & Ahumada, A. (2009). A Contrast Polarity Search Effect in Letter Identification. *Journal of Vision (Charlottesville, Va.)*. doi:10.1167/9.8.1021
- Scharff, L. F. V., & Ahumada, A. J. Jr. (2002). Predicting the readability of transparent text. *Journal of Vision (Charlottesville, Va.)*, 2(9), 653–666. <http://journalofvision.org/2/9/7/> doi:10.1167/2.9.7 PMID:12678636

- Scharff, L. F. V., & Ahumada, A. J. (2002). *Using letter identifiability to predict readability of transparent text on textured backgrounds*. Optical Society of America Fall Vision Meeting, San Francisco, CA. doi:10.1167/2.10.137
- Scharff, L. F. V., & Ahumada, A. J. Jr. (2002). Predicting the readability of transparent text. *Journal of Vision (Charlottesville, Va.)*, 2(9), 653–666. doi:10.1167/2.9.7 PMID:12678636
- Scharff, L. F. V., & Ahumada, A. J. (2003a). Contrast measures for predicting text readability. In *Human Vision and Electronic Imaging VII*. SPIEProc. 5007, Paper 46.
- Scharff, L. F. V., & Ahumada, A. J., Jr. (2003b). Letter identification latencies are predicted by an asymmetric contrast metric. In *Vision Sciences Society Abstracts*. Academic Press.
- Scharff, L. F. V., Ahumada, A. J., Jr., & Hill, A. L. (1999) Discriminability measures for predicting readability. In *Human Vision and Electronic Imaging III*. SPIE.
- Scharff, L. F. V., Hill, A., & Ahumada, A. J. (2000). Discriminability measures for predicting readability of text on textured backgrounds. *Optics Express*, 6(4), 81–90. doi:10.1364/OE.6.000081 PMID:12238520
- Scharff, L. V., & Ahumada, A. J. (2003). Contrast measures for predicting text readability. In B. E. Rogowitz & T. N. Pappas (Eds.), *Human Vision and Electronic Imaging VII*. SPIEProc. doi:10.1117/12.479663
- Scharff, L. V., & Ahumada, A. J. (2003, November). Letter Identification: Contrast Polarity and Speed-Accuracy Trade-off Strategies. *Abstracts of the Psychonomic Society*, 8, 67.
- Scharff, L. V. & Ahumada, A. J. (2003). *Why is letter identification performance better for negative than positive contrast?* Berkeley Area Vision Research Day.
- Scharff, L. V., & Ahumada, A. J. (2003). *Letter identification latencies are predicted by an asymmetric contrast metric*. Vision Sciences Society Annual Meeting.
- Scharff, L. V., & Ahumada, A. J. (2005). *Understanding text polarity effects*, European Conference on Visual Perception. *Perception*, 34, 133.
- Scharff, L. V., & Ahumada, A. J. (2005). Why is light text harder to read than dark text? Vision Sciences Society Annual Meeting. *Journal of Vision (Charlottesville, Va.)*, 5(8), 812a. doi:10.1167/5.8.812
- Shaikh, A.A. & Gupta, G.K. (2014). M-Commerce Recommendation With Mobile Cloud Architecture. *International Journal of Application or Innovation in Engineering & Management*, 3(11), 347-352.
- Siwicki, B. (2014). *Thanksgiving through Cyber Monday: 'An m-commerce explosion'*. Retrieved from the Internet at <https://www.internetretailer.com/2014/12/04/mobile-commerce-thanksgiving-black-friday-cyber-monday/>
- Wu, C. S., Cheng, F. F., & Yen, D. C. (2008). The atmospheric factors of online storefront environment design: An empirical experiment in Taiwan. *Information & Management*, 45(7), 493–498. doi:10.1016/j.im.2008.07.004
- Zuffi, S., Beretta, G., & Brambilla, C. (2006). A Color Selection Tool for the Readability of Textual Information on Web pages. *Proc. IS&T/ SPIE, Internet Imaging VII, Electronic Imaging Conference*. doi:10.1117/12.651589
- Zuffi, S., Brambilla, C., Beretta, G., & Scala, P. (2007). Legibility and Contrast. *Proc. 14th International Conference on Image Analysis and Processing (ICIAP)*.

ADDITIONAL READING

Bartikowski, B., Taieb, B., & Chandon, J. L. (2016). Targeting without alienating on the internet: Ethnic majority and minority consumers. *Journal of Business Research*, 69(mars), 1082–1089. doi:10.1016/j.jbusres.2015.08.024

Pelet, J.-É. (2010). Effets de la couleur des sites web marchands sur la mémorisation et sur l'intention d'achat. *Systèmes d'Information et Management*, 15(1), 97–131. doi:10.3917/sim.101.0097

Pelet, J.-É. (2014). *M-Commerce* (C. Pearson, Ed.). Village Mondial.

Pelet, J.-É. (2014). Investigating the Importance of Website Color Contrast in E-Commerce: Website Color Contrast in E-Commerce. In Khosrow-Pour, M. (2015). *Encyclopedia of Information Science and Technology* (3rd ed.). (Vols. 1–10). Hershey, PA: IGI Global; <http://www.igi-global.com/chapter/investigating-the-importance-of-website-color-contrast-in-e-and-m-commerce/112647>, doi:10.4018/978-1-4666-5888-2

Pelet, J.-É., Ettis, S., & Cowart, K. (2016). Optimal Experience of Flow Enhanced by Telepresence: Evidence from Social Media use. *Information & Management*. doi:10.1016/j.im.2016.05.001

Pelet, J.-É., & Papadopoulou, P. (2011, July-September). The Effect of E-commerce Websites Colors on Customer Trust. *International Journal of E-Business Research*, 7(3), 1–18. doi:10.4018/jebr.2011070101

Pelet, J.-É., & Papadopoulou, P. (2011). Investigating the Effect of Color on Memorization and Trust in E-Learning: The Case of KMCMS.net (Knowledge Management and Content Management System). In O. Bak & N. Stair (Eds.), *Impact of E-Business Technologies on Public and Private Organizations: Industry Comparisons and Perspectives* (pp. 52–78). Hershey, PA: Business Science Reference; doi:10.4018/978-1-60960-501-8.ch004

Pelet, J.-É., & Papadopoulou, P. (2012). The effect of colors of e-commerce websites on consumer mood, memorization and buying intention, *European Journal of Information Systems*, Special Issue: “Information Research. Education and Policy in the Mediterranean Region”, 21, 438–467. published online 22 May 2012. doi:10.1057/ejis.2012.17

Pelet, J.-É., & Papadopoulou, P. (2013), *Handbook of Research on User Behavior in Ubiquitous Online Environments*, <http://www.igi-global.com/book/user-behavior-ubiquitous-online-environments/76724>, DOI: 10.4018/978-1-4666-4566-0

Pelet, J.-É., & Papadopoulou, P. (2013). Colored vs. Black Screens or How Color Can Favor Green e-Commerce. In A. Scupola (Ed.), *Mobile Opportunities and Applications for E-Service Innovations* (pp. 57–76). Hershey, PA: Information Science Reference. doi:10.4018/978-1-4666-2654-6.ch004

Pelet, J.-É., & Papadopoulou, P. (2013). The Effect of E-Commerce Websites' Colors on Customer Trust. In I. Lee (Ed.), *Mobile Applications and Knowledge Advancements in E-Business* (pp. 167–185). Hershey, PA: Business Science Reference. doi:10.4018/978-1-4666-1960-9.ch011

Pelet J.-É., Papadopoulou P. (2014). Investigating Social Media in M-Commerce, *International Journal of Technology and Human Interaction*, Special Issue on Social Media and Social CRM, 10 (4).

Pelet, J.-É., & Papadopoulou, P. (2014). Social media and m-commerce: An exploratory study. *International Journal of Internet Marketing and Advertising*, 9, 1. doi:10.1504/IJIMA.2015.068358

Pelet, J.-É., & Papadopoulou, P. (2015). Tablet and social media adoption in m-commerce: An exploratory study. *International Journal of Strategic Innovative Marketing*, 2, 1. doi:10.15556/IJSIM.02.01.004

KEY TERMS AND DEFINITIONS

Contrast: is the difference in luminance and/or color that makes an object (or its representation in an image or display) distinguishable. In visual perception of the real world, contrast is determined by the difference in the color and brightness of the object and other objects within the same field of view. Because the human visual system is more sensitive to contrast than absolute luminance, we can perceive the world similarly regardless of the huge changes in illumination over the day or from place to place. The maximum contrast of an image is the contrast ratio or dynamic range. Finally, contrast is the perceived difference in colors that are in close proximity to each other. Using contrast effectively not only differentiates a design from others, it is the essential ingredient that makes content accessible to every viewer, on a laptop or a handled device such as a smartphone or tablet.

ENDNOTES

¹ Websites such as “<http://fontawesome.io/>” enable the use of metaphoric icons.

- ² The Ishihara test is the most common color blindness test used today (Deeb & Motulsky, 2011). It consists of a number of plates, either 24 or 38, (these are known as the Ishihara plates), each containing a circle of dots which appear in random colors and sizes. Within the circle are dots forming a number which should be clearly visible to viewers with normal color vision and hard to see or invisible to viewers with defective color vision. To pass the test, participants should recognize the number in every plate.
- ³ An interval between 300 lux and 500 lux was fixed.
- ⁴ A maximum of 65% humidity was retained.
- ⁵ Data were collected when the temperature ranged from 19 to 29°C.
- ⁶ The Web Access Initiative (WAI) works with the W3C (World Wide Web Consortium) working groups to address and improve accessibility within specifications and W3C technologies.

Ethical Ambiguities in the Privacy Policies of Mobile Health and Fitness Applications

M**Devjani Sen***University of Ottawa, Canada***Rukhsana Ahmed***University of Ottawa, Canada*

INTRODUCTION

Mobile leisure, health, and wellness applications (apps) are ubiquitous. A recent study reveals that there are approximately 97,000 varieties of inexpensive and easy to use mobile health apps available in the market; at such a pace numbers are becoming outdated almost as soon as they are published (Privacy Clearinghouse, 2013). It is predicted that by 2017 half of the world's more than 3.4 billion smart phone users will have downloaded health and fitness apps (Comstock, 2013), which raises the question: what happens to the sensitive data consumers enter into these apps?

Indeed, a hot topic in both Canada and the U.S., concerns exactly what third parties, such as insurance companies, can legally do with personal data. American law dictates that health insurance companies cannot discriminate based on a history of illness. However, while data held by a health plan, health care provider, or lab may be protected by the federal Health Insurance Portability and Accountability Act (HIPAA), legal scholars warn that if a patient is going to upload health or wellness data to a mobile application (app), it may not be covered by those laws (Rogers, 2014; Whitman & Mattord, 2012). Such legal ambiguities have implications for Canadian users of health and wellness apps, because many of these devices are based in the U.S., with the data being stored on U.S. servers and thus they may not conform to privacy requirements (Akkad, 2013).

There are some other important concerns with privacy and security issues related to mobile health and fitness applications. For example, personal apps collect all sorts of personal information like name, email address, age, height, weight, and in some cases detailed health information. When using such apps, many users may trustfully log everything from diet to sleep patterns in the apps. By sharing such personal information end-users may make themselves targets to misuse of this information by unknown third parties. Moreover, according to Gralla et al. (2011), apps can gather the phone number and the unique ID number of each type of phone: the Unique Device Identifier (UDID) on an iPhone, the International Mobile Equipment Identity (IMEI) number on a BlackBerry, and (depending on the make) the IMEI or the Mobile Equipment Identifier (MEID) on an Android phone. In this way, personal information that apps gather about an end-user can be matched to these IDs, which means that ad networks can easily combine various pieces of information collected by multiple apps to build a sophisticated profile about a given end-user and thereby posing a major privacy risk to personal data. Therefore, uninformed decision making by end-users raises important concerns regarding the ethics around sharing personal data gathered from health and fitness apps to third parties. These concerns can be much graver when Martínez-Pérez and colleagues (2014), in a review of privacy and security in mobile health apps, found evidence of insecure handling of clinical and medical data.

DOI: 10.4018/978-1-5225-2255-3.ch528

To summarize, the issues raised above may be broken down to the following concerns:

1. Ownership and veracity of sensitive data shared on personal apps;
2. What end users really understand about the use of their data (what data is collected and the specifics of how it may be used);
3. The ethics of sharing end-users' personal information and sharing it with third-parties.

Despite the important role of informed consent in the creation of health and fitness mobile applications, the intersection of ethics and sharing of personal information is understudied and is an often-ignored topic during the creation of mobile apps. After reviewing the online privacy policies of a select set of mobile health and fitness apps, this chapter will conclude with a set of recommendations when designing privacy policies for the sharing of personal information collected from health and fitness apps.

BACKGROUND

Online privacy policies, which regulate the relationship between the user and the website with the purpose of limiting companies' legal liability during site use, are also used by users to inform their understanding of the ways personal data are treated by companies. Despite their importance to users, however, studies suggest that these policies are often ignored (Angulo, Fischer-Hübner, Westlund, & Pulls, 2012; Jensen & Potts, 2004; Kesan, Hayes, & Bashir, 2012; Tsai, Egelman, Cranor, & Acquisti, 2011). As pointed out by Steinfeld (2016), since agreeing to the terms of the policy is usually a prerequisite for subscribing to a website or a web service, users typically sign their consent almost automatically, so that these terms are rarely considered as reasons for joining or avoiding a given website.

Studies suggest that many apps do not have a privacy policy, or that apps do not grant users

access to and control over personal information before users download and/or after use apps (Privacy Rights Clearinghouse, 2013). Nevertheless, having a privacy policy or providing a link to it is not enough to safeguard end-users' right to data privacy. Studies have also identified a number of cognitive factors which limit the comprehension of existing privacy policies, including complexity, legal language, and length (Angulo et al., 2012; Milne & Culnan, 2004; Nissenbaum, 2011; Tsai et al., 2011), use of vague terms and concepts (Anton et al., 2003), as well as design issues such as format and font size (Milne & Culnan, 2004).

IDENTIFYING PRIVACY RISKS IN HEALTH AND FITNESS APPLICATION PRIVACY POLICIES

Clearly, if privacy policies are not being read, there is a need for more readable policies that will be better understood and will more effectively inform users decisions on whether it is prudent for them to join a given website service agreement. Indeed, given the varied educational levels and socio-economic backgrounds of online consumers that encompasses a wide range of the global population, it is critical that companies communicate effectively with customers through their online privacy policies (Wheatman & Ghiselli, 2014). With these considerations in mind, the purpose of the present chapter was to investigate the presence of important information and its presentation in plain language format in privacy policies of health and fitness applications. To this end, a close examination of the privacy policies of selected commonly used health and fitness applications was performed. Considering the limited scope of this chapter, the number of apps to examine was limited to four in order to properly execute a combined approach of using a checklist and discussing relevant excerpts from the selected privacy policies.

Health and fitness applications are application programs that offer health-related services on

portable devices such as smartphones and tablet PCs. The apps examined in this chapter included: 1) Fitbit, which records data such as the number of steps walked, heart rate, quality of sleep, steps climbed, and other personal metrics; 2) JEFIT, a free app that records and logs fitness exercises, tracks workouts and progress, enables access to a large exercise database; 3) Nike Training Club, which gives the user a personal trainer, with more than 60 custom-built workouts from a Nike Professional Trainer as well as enables user to sync their results to social media sites such as Facebook to show friends how training is going; and 4) MyFitnessPal, a free calorie counter smartphone application and website, which helps to track calorie intake for exercise and diet and employs gamification elements to engage and motivate users to achieve their goals. In order to ensure that the health and fitness apps examined comprised popular apps with a wide user base, the selection criteria for each app was set at a minimum of one million downloads. Care was also taken to ensure that: a) each app had an androgynous user base, having an equal appeal for male and female users

and b) two of the four apps in the sample included free apps considering such apps are expected to hold greater biases for selling material to third party sites. It is also important to note that the apps selected offer a diversity of core functions, including recording, logging and tracking personal metrics, diet, exercise and workout routines and progress, even working with certified personal trainers.

In order to assess the privacy policies of the selected four health and fitness apps, a checklist was used that comprised five privacy risk categories. These privacy risks were based on two questions: a) is important information missing to make informed decisions about the use of personal data? and b) is information being shared that might compromise the end-user's right to privacy of that information? The two authors of this chapter examined the online privacy policies of each selected app in search of important privacy risks. Based on their search, each author completed a separate checklist and compared their findings to reach an agreement of the presence or absence of each privacy risk category. Table 1 outlines the

Table 1. Checklist of privacy risks for Fitbit, JEFIT, Nike and MyFitnessPal mobile applications

Description of Privacy Risk	Fitbit	JEFIT	MyFitnessPal	Nike Training Club
Explicitly stated that under no circumstances will Personal Identifiers (PI) be shared	×	×	×	×
Use of cookies explicitly stated as well as what information is tracked and how it is used	√	√	√	√
Explicitly stated what sharing is done with third party sites and the security regarding links to other Sites	√	√	√	√
Location and use of personal data after account is deactivated	√	×	×	×
Ownership of data is stated clearly when device is used by international users	√	×	√	√

results of the final checklist assessing the occurrence of each privacy risk examined.

The following section will discuss in further detail the subsections from the checklist from the assessment of the privacy policies of the four health and fitness applications.

Type of Data Collected Explicitly Explained and it is Stated What is Done With Personally Identified (PI) Data

Regarding personal identifiers, JEFIT (2015) states:

The information we collect in order to create your JEFIT.com account includes data such as your name and email address. The email address may be used to contact you with matters regarding JEFIT, such as a newsletter, notifications of new messages from other JEFIT members, or an announcement about a new product or service. We also collect some personal fitness information so that we can provide you with a personalized plan for successful weight loss. JEFIT also collects other profile data including but not limited to: personal interests, geographical location, gender, and age in order to assist our members find and communicate with each other. JEFIT collects personal information about you when you sign up for a JEFIT account. This information includes your age, metric preferences, and other personal fitness details. JEFIT also stores the data you enter about workouts, exercise, and anything you choose to track and keep records of.

When referring to non-personally-identifiable information, the policy states:

JEFIT also logs non-personally-identifiable information including IP address, profile information, aggregate user data, and browser type, from users and visitors to the site. This data is used to manage the website, track usage and improve the website services. This non-personally-identifiable

information may be shared with third-parties to provide more relevant services and advertisements to members. User IP addresses are recorded for security and monitoring purposes.

From the above excerpts, it is not completely clear whether non-personally-identifiable information may be shared with third-parties. It is simply stated that personal information is stored and collected, without further explanation about the purpose of this storing and collecting. What does it really mean to use this data to provide the end-user with a personalized plan for successful weight loss? Does it mean sharing this information with third parties who can advise on how to best do this?

Further, in the section entitled Privacy Policy Change, JEFIT (2015) states:

When you use the JEFIT.com sites to provide personally identifiable information, you consent to the collection and use of that information in accordance with this privacy policy.

What does this statement convey? Again, the intended or planned use of the collected and stored personally identifiable information is not clearly communicated.

Finally, of the privacy policies of the four health and fitness applications explored, the most alarming ambiguity seems to be stated in the privacy policy of MyFitnessPal (2013):

Except as otherwise provided in this Privacy Policy, and when You otherwise give permission or under the following circumstances, MyFitnessPal will not share Your Personal Information with third parties.

However, what is stated in subsection User Content, under Terms and Use, appears to conflict with (interestingly located in a poor visibility area of 19th bullet) MyFitnessPal's (2013) explicit statement regarding the PI data that the company can do whatever it chooses with the end-users personal data:

“User Content” is any content, materials or information (e.g., any text, information, photos, images, video, and other content and material, including nutritional information contributed to the Food Database) that You upload or post to, or transmit, display, perform or distribute by means of, the Services, whether in connection with Your use of Website, Application, or through the use of any Third Party Websites or Third Party Services or otherwise. You hereby grant MyFitnessPal and its officers, directors, employees, agents, affiliates, representatives, service providers, partners, sublicensees, successors, and assigns (collectively, the “MyFitnessPal Parties”) a perpetual, fully paid-up, worldwide, sublicensable, irrevocable, assignable license to copy, distribute, publish, transmit, publicly display or perform, edit, modify, translate, reformat and otherwise use User Content in connection with the operation of the Services or any other similar services or related business, in any medium now existing or later devised, including without limitation in advertising and publicity. You further agree that this license includes the right for the MyFitnessPal Parties to publish, display or otherwise use and make available your User Content and possibly your name and/or any user name of yours in connection with their exercise of the license granted under this section. You agree to waive, and hereby waive, any claims arising from or relating to the exercise by the MyFitnessPal Parties of the rights granted under this section, including without limitation any claims relating to your rights of personal privacy and publicity.

Regarding personal identifiers, Fitbit (2016) states the following:

We will never sell your data, and will only share personally identifiable data when you direct us to (or under the circumstances outlined in our Privacy Policy).

Although this information appears to be clear, it might not be understood the same way by dif-

ferent users. For example, many people might understand the above to mean that if they do not give explicit consent their data will not be shared. However, it could also be that directing data to be shared might occur in other more covert ways that most end-users are not aware of.

While NIKE (2014) does state in what ways PIs may be shared with other parties, they also explicitly state:

However, like other companies, NIKE cannot guarantee 100% the security or confidentiality of the information you provide to us.

Use of Cookies, What Information is Tracked and How it is Used

Most privacy policies of the health and fitness applications examined defined what is meant by cookies and what they do (how they track information) and give details about how to disable cookies. However, many useful features of fitness apps warrant the use of entering in information that might be tracked by cookies. As Fitbit (2016) states:

When You Add Information to Your Account

You can customize your Fitbit experience by adding other types of information to your account, such as entering a food log or setting an alarm, personalizing your profile with photos, participating in discussion boards, or sending messages to your Fitbit friends. Whenever you add this type of data, we collect it and store it in your Fitbit account.

This statement raises the question of whether it is realistic for all cookies to be disabled. For example is it reasonable to expect a user to sign into a fitness app where they never enter information into food logs, set an alarm, participate in discussion boards, add photos to their profile, or interact with their friends who have this same app?

Sharing with Third Party Sites

On the whole, this section was well done by the four mobile application privacy policies examined. All apps stated that if end-users choose to share their information with Third Party Sites, they need to carefully read the privacy policies of these third party sites. This way, sharing of personal information with third party sites may likely happen with end-users' knowledge and based on their informed decision.

Location and Use of Personal Data After Account is De-Activated

With the exception of Fitbit, it was not possible to clearly locate information on what happens to personal information once the user de-activates his or her account. Fitbit (2016) states:

You can deactivate your Fitbit account by contacting Customer Support. When you do, data that can identify you will be removed from the Fitbit Service, including but not limited to your email, name, photo(s), friends list and links to sites such as Facebook and Twitter. Backup copies of this data will be removed from our server based upon an automated schedule, which means it may persist in our archive for a short period. Fitbit may continue to use your de-identified data after you deactivate your account.

This statement raises two questions: 1) what happens to PIs during the short period of time referred to? and 2) what is the length of this period of time? Clearly more attention needs to be given to these important issues in terms of clarity and completeness of the information provided.

Ownership of Data is Stated Clearly When Device is Used by International Users

Only Fitbit (2016) clearly stated that the U.S. Privacy Policies do not apply to the sharing of

personal data when using Fitbit devices outside the United States. Interestingly, although NIKE (2014) and MyFitnessPal (2013) also stated this information, it was missed entirely by one of the authors due to poor visibility. This may be an important concern for international use of such devices by travelers from the U.S. in addition to end-users who reside in other countries.

RECOMMENDATIONS

Based on the review of the online privacy policies of four mobile health and fitness applications, and reflecting on the analysis of ambiguities and privacy risks in the privacy policies of those apps, the following recommendations are offered. Although these recommendations should be taken into consideration when designing privacy policies for the sharing of personal information collected from health and fitness apps, they can also be extended to the creation of privacy policies of other mobile apps and web services such as in the areas of human resources, planning and project management, customer engagement, property management, education and training:

1. It is suggested that personally identifiable information be stated clearly; PI and non-PI information should be clearly distinguished. It is also important to clearly state what will never be done with PI data. Moreover, this information should be made highly visible (within the first page) of any Privacy Policy/Terms and Conditions;
2. It was found that cookies were clearly defined in the privacy policies of the apps examined. However, given that information is tracked and used by cookies, the kinds of activities that enable activation of cookies might need to be stated more clearly;
3. Generally, when the end-user enters a Third Party Site, its privacy policies apply. While it is recommended that this guideline be stated clearly, the need to read and exam-

ine these third party site privacy policies should be highlighted as well. Furthermore, if Contests and Giveaways comprise Third Party sites, this information also needs to be highlighted; at some part of a Contest or Giveaway advertisement, it clearly needs to be stated that the privacy policy of this site is now in effect (vs. the privacy policy of the original site);

4. It is advised that information regarding the location and use of personal data after an account is disabled be stated clearly in the privacy policies. Not doing so may suggest that all data is automatically deleted upon deactivation of an account. Hence, it is recommended that this information be clearly presented under a subheading such as What Happens When Your Account is De-Activated or something similar to this;
5. It is proposed that the rightful owner of data be more clearly stated when devices are used by international users and this information be stated in a clearly visible manner.

FUTURE RESEARCH DIRECTIONS

This encyclopedia provides a comprehensive summary of research-based chapters on copious aspects confronting the field of information science and technology. The rapid innovations and advances in Information and Communication Technologies (ICTs) and the growing rates at which these technologies are transforming lives around the world and shaping the future, demand incubation of inter/cross/transdisciplinary research agendas. On the one hand, technological innovations can significantly improve health care by enabling easier communication of electronic record of patient health information between different health care professionals (Mardis, 2013), reaching at-risk populations (Briones & Sundstrom, 2013), and so on. On the other hand, integrating ICTs into health care is complicated and it is a daunting process to create an ICT-driven

health care system that is accessible, reliable, safe, and thus effective (Avtgis & Polack, 2013). Future research should proceed with caution when considering the design of health and fitness applications on mobile devices to help safeguard privacy when sharing personal information on health and fitness apps.

As explored in this chapter, mobile health and fitness applications raise many pertinent yet unanswered questions for health care ethics. Many end-users may automatically consider health and fitness apps to be beneficial; but, there are complications that may arise from neglecting privacy and security concerns when sharing personal information on these apps. Overall, ICTs help facilitate access to health information and thus can help to increase awareness which translates to patient participation in their own health care. There is an app for just about any health, fitness, and weight-loss resolution. Amidst this myriad health and fitness apps that are permeating lives all around the globe, an important question of data privacy looms large: what happens when advertisers and other third parties share this information without the user's knowledge? Questions like this and others that remain largely unexplored and that have only recently been asked by scholars should guide future investigations. Future studies should continue to investigate the complexities at the intersection of ethics and information sharing on health and fitness apps in order to tap into the potential of mobile apps to improve health behavior and health care delivery.

Health and fitness mobile application development should integrate the end-user experience perspective. Future research should privilege the end-user voice in determining ethics, privacy, and security of personal data shared on these apps, and in co-designing guidelines for data protection. Surveys, focus groups, and interviews can help gather important information regarding end-user awareness and comprehension of data privacy. Such data could yield valuable information to aid in designing simple and user-friendly privacy policies. Researchers across disciplines should

continue to share good practices with regard to personal data use, sharing, and retention practices of health and fitness applications.

As the field of information science and technology continues to evolve, so do the available ICTs become integrated and may eventually be an integral part of the health care system. The end-users of mobile applications will have opportunities to influence developments in health information technologies that can lead to personalized tools to promote health and prevent disease. Yet, the intersections of ethics, privacy, security, and information sharing is understudied and is an often-ignored topic during the creation of mobile apps. In order to complement the research reported in this chapter, further exploration in the following areas is advocated for:

- A comprehensive review of privacy policies of the plethora of health and fitness applications that are available;
- Full investigation, from the perspectives of both app developers and end-users, of ethical ambiguities in privacy policies across health and fitness applications;
- An advancement of current understanding of how people with limited literacies – functional, health, digital, technology, and so on – use and apply health and fitness apps.

CONCLUSION

This chapter explored ambiguities in the privacy policies of selected health and fitness mobile applications. In doing so, it raises ethical and security concerns about sharing health and medical information online. Through its review of topics such as the ethics of information sharing on apps and online privacy policies, this chapter helped to advance knowledge on the dangers of sharing personal information on mobile apps, which can

jeopardize the rights to privacy of end-users of mobile apps.

Today's health care system greatly depends on information and communication technologies. These advancements facilitate communication between patients, doctors, and other health care professionals. Technology, with its many uses, continues to flourish as society continues to develop and expand. There are numerous benefits as well as negatives to using these technologies in the health care context.

Finally, this chapter calls for multidisciplinary and multisectoral research partnerships to best meet the challenges of today's information technologies. Health and fitness mobile applications will not address ethical and privacy concerns in and of itself. Researchers, practitioners, policy-makers, app developers, programmers, promoters, and users need to come together to debate and address these concerns.

REFERENCES

- Angulo, J., Fischer-Hübner, S., Weastlund, E., & Pulls, T. (2012). Towards usable privacy policy display and management. *Information Management & Computer Security*, 20(1), 4–17. doi:10.1108/09685221211219155
- Anton, A. I., Earp, J. B., & Carter, R. (2003). Precluding incongruous behavior by aligning software requirements with security and privacy policies. *Information and Software Technology*, 45(14), 967–977. doi:10.1016/S0950-5849(03)00095-8
- Avtgis, T. A., & Polack, E. P. (2013). Coming full circle in rural trauma: Chronicling the development and testing of communication systems in rural trauma networks. In R. Ahmed & B. R. Bates (Eds.), *Health communication and mass media: An integrated approach to policy and practice* (pp. 151–164). Farnham, UK: Gower Publishing Ltd.

- Briones, R. L., & Sundstrom, B. (2013). Reaching the unreachable: How eHealth and mobile health technologies impact at-risk populations. In R. Ahmed & B. R. Bates (Eds.), *Health communication and mass media: An integrated approach to policy and practice* (pp. 133–147). Farnham, UK: Gower Publishing Ltd.
- Comstock, J. (2013, March 14). *Report: 1.7B to download health apps by 2017*. Retrieved from <http://mobihealthnews.com/20814/report-1-7b-to-download-health-apps-by-2017>
- El Akkad, O. (2013, December 19). How insurers are turning to fitness apps to decide your health coverage. *The Globe and Mail*. Retrieved from <http://www.theglobeandmail.com/technology/how-insurers-are-turning-to-fitness-apps-to-decide-your-health-coverage/article16065068/>
- Fitbit. (2016). *Fitbit privacy policy*. Retrieved from <https://www.fitbit.com/ca/privacy>
- Gralla, P., Sacco, A., & Fass, R. (2011). *Smartphone apps: Is your privacy protected?* Retrieved from <http://www.computerworld.com/article/2509878/dataprivacy/smartphone-apps--is-your-privacy-protected-.html>
- JEFIT. (2015). *Privacy policy*. Retrieved from <https://www.jefit.com/privacy-policy/>
- Jensen, C., & Potts, C. (2004). Privacy policies as decision-making tools: An Evaluation of online privacy notices. *Proceedings of SIGCHI Conference on Human Factors in Computing Systems*, 471-478. doi:10.1145/985692.985752
- Kalyani, M. (2013). *Privacy Concerns with Mobile Health Apps*. Retrieved from <https://spideroak.com/privacypost/cloud-security/privacy-concerns-with-mobile-health-apps-draft/>
- Kesan, J., Hayes, C., & Bashir, M. (2012). Information privacy and data control in cloud computing: Consumer privacy choices, informed consent, and baseline protections to facilitate market transactions in the cloud. *Illinois Program in Law, Behavior and Social Science*, 11-20. Paper No. LBSS12-11. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2042638
- Mardis, N. (2013). Form patient-based records to patient-centered care: Reconfiguring health care systems for interoperable electronic health records. In R. Ahmed & B. R. Bates (Eds.), *Health communication and mass media: An integrated approach to policy and practice* (pp. 165–178). Farnham, UK: Gower Publishing Ltd.
- Martínez-Pérez, B., de la Torre-Díez, I., & Miguel López-Coronado, M. (2015). Privacy and security in mobile health apps: A review and recommendations. *Journal of Medical Systems*, 39(1), 181. doi:10.1007/s10916-014-0181-3 PMID:25486895
- Milne, G. R., & Culnan, M. J. (2004). Strategies for reducing online privacy risks: why consumers read (or don't read) online privacy notices. *Journal of Interactive Marketing*, 18(3), 15-29. 10.1002/dir.20009
- MyFitnessPal. (2013). Retrieved from http://www.myfitnesspal.com/account/terms_and_privacy?with_layout=true#privacy
- Nike. (2016). *Nike digital privacy policy*. Retrieved from https://helpall.nike.com/app/answers/detail/a_id/52863/~nike-digital-privacy-policy
- Nissenbaum, H. (2011). A contextual approach to privacy online. *Daedalus*, 140(4), 32-48. 10.1162/DAED_a_00113

Privacy Rights Clearinghouse. (2013). *Fact Sheet 39: Mobile health and fitness Apps: What are the privacy risks?* Retrieved from <https://www.privacyrights.org/mobile-health-and-fitness-apps-what-are-privacy-risks>

Rogers, J. (2014, September 16). *Expert: Apple's health push poses big privacy challenges*. Retrieved from <http://www.foxnews.com/tech/2014/09/16/expert-apples-health-push-poses-big-privacy-challenges/>

Steinfeld, N. (2016). "I agree to the terms and conditions": (How) do users read privacy policies online? An eye-tracking experiment. *Computers in Human Behavior*, 55(Part B), 992-1000.

Tsai, J. Y., Egelman, S., Cranor, L., & Acquisti, A. (2011). The effect of online privacy information on purchasing behavior: An experimental study. *Information Systems Research*, 22(2), 254-268. doi:10.1287/isre.1090.0260

Wheatman, S., & Ghiselli, M. (2014). *Privacy policies: How to communicate effectively with consumers*. Retrieved from https://iapp.org/media/pdf/knowledge_center/IAPP_KMPrivacy-Paper_FINAL.pdf

Whitman, M. E., & Mattord, H. J. (2012). Legal, ethical, and professional issues in information security. In *Principles of information security* (4th ed.; pp. 133-147). Boston, MA: Course Technology, Cengage Learning. Retrieved from http://www.cengage.com/resource_uploads/downloads/1111138214_259148.pdf

ADDITIONAL READING

Acquisti, A., & Grossklags, J. (2012). An online survey experiment on ambiguity and privacy. *Communications & Strategies*, 88, 19-39.

Al-Hakim, L. (2007). *Web mobile-based applications for healthcare management*. Hershey, NY: IGI Global. doi:10.4018/978-1-59140-658-7

Canadian Institutes of Health Research. (2003). *Secondary use of personal information in health research: Case studies, November 2002*. Retrieved from <http://www.cihr-irsc.gc.ca/e/1475.html>

Huckvale, K., Prieto, J. T., Tilney, M., Benghozi, P.-J., & Car, J. (2015). Unaddressed privacy risks in accredited health and wellness apps: A cross-sectional systematic assessment. *BMC Medicine*, 13(1), 214. doi:10.1186/s12916-015-0444-y PMID:26404673

Rowan, M., & Dehlinger, J. (2014). A privacy policy comparison of health and fitness related mobile applications. *Procedia Computer Science*, 37, 348-355. doi:10.1016/j.procs.2014.08.051

Russello, G., Dong, C., & Dualy, N. (2009). A framework for capturing patient consent in persuasive healthcare applications. In P. Olla & J. Tan (Eds.), *Mobile health solutions for biomedical applications* (pp. 216-230). Hershey, NY: IGI Global. doi:10.4018/978-1-60566-332-6.ch012

Silva, B. M., Rodrigues, J. J., Canelo, F., Lopes, I. C., & Zhou, L. (2013). A data encryption solution for mobile health apps in cooperation environments. *Journal of Medical Internet Research*, 15(4), e66. doi:10.2196/jmir.2498 PMID:23624056

KEY TERMS AND DEFINITIONS

Ethics: The critical examination of the advantages and disadvantages when deciding upon the correct conduct involving a moral issue.

Information Security: The process of protecting the availability, integrity and privacy of information.

Mobile Applications: A term used to describe Internet applications that run on portable devices such as smartphones and other mobile devices to make it easier for users to access the Internet.

Mobile Health and Fitness Applications: Application programs that offer health-related

services on portable devices such as smartphones and tablet computers.

Online Privacy Policy: A document, typically required by law, which regulates the relationship between the user and the website with the purpose of limiting companies' legal liability during site use.

Personal Identifiers Information (PIIs):

Comprises information, that when considered alone, or in combination with data from other sources, may contribute to distinguish (identify) an individual.

Privacy: An individual's right to control how and to what extent information about him or her may be shared and acted upon by others.

M

Exploring the Growth of Wireless Communications Systems and Challenges Facing 4G Networks

Amber A. Smith-Ditizio
Texas Woman's University, USA

Alan D. Smith
Robert Morris University, USA

INTRODUCTION

Wireless and Mobile Communications

In a traditional sense, wireless systems were considered as an auxiliary approach used in regions where it was difficult to build a connection by using typical wires and cables (Alderete & Gutiérrez, 2014). Firms operating in a competitive and global marketplace are relying more and more on the data that flows throughout their organization. As a means of enhancing supply chain management (SCM), firms are working to improve the integration level with suppliers and clients, materials and inventory regarding the production costs, influence of raw material prices, and the active request for changes by the client (Dharni, 2014; Gupta & Naqvi, 2014; Qrunfleh, Tarafdar, & Ragu-Nathan, 2012; Soon, Mahmood, Yin, Wan, Yuen, & Heng, 2015). Many organizations are seemingly more concerned with the identification of logistic issues as new opportunities arise. Even though changes in commercial activity are an important part of B2B (business-to-business) operations; for the most part, managers today do not comprehend insightful ways to deal with them. Many of the smaller firms that are integrated into a supply chain, lack the technological capacity to implement electronic operations of data processing, transmission of data, and reception of data (Qrunfleh, et al., 2012).

Originally, 1G-networks were introduced in the 1980s and were generally based on analog techniques. It built the basic structure of mobile communications and solved many fundamental problems. Some of these problems included cellular architecture adopting, multiplexing frequency band, roaming across domain, and non-interrupted communication. The only service of 1G was Speech Chat. Then, once implemented, 2G gained popular acceptance during the 1990s. It was based on digital signal processing techniques and regarded as a bridge from analog to digital technology. It introduced a new option of communication called text messaging, multimedia messaging or picture messaging, General Packet Radio Service, Wireless Application Protocol, Enhanced Data Rates for GSM Evolution, and Internet services. 2G's main contributions were the utilization of SIM (Subscriber Identity Module) cards and support capabilities for a large number of users (e.g., 2.5G was regarded as 3G services for 2G networks). This network extended the 2G network with data service and packet switching methods. 2.5G brought the Internet into mobile personal communications under the same networks with 2G, which was an important concept leading to different kinds and combinations of communications.

The 3G network, known as "the third generation mobile broadband data services", came with several incremental improvements in radio technology and command-and-control software.

DOI: 10.4018/978-1-5225-2255-3.ch529

Radio advancements of the 3G network are referred to as antenna techniques or coding/modulation schemes. According to Akintoye (2013, p. 43), “Several new radio techniques are employed to achieve high rates and low latencies.” These new radio techniques included Space Division Multiplexing (SDM) via Multiple Input/Multiple Output (MIMO), Space Time Coding (STC) using higher order of modulation, and encoding schemes, beam forming, beam directional control, and inter-cell interference techniques. These growths in wireless data services have placed higher demands on mobile wireless networks, and in response, wireless carriers are upgrading their networks to offer faster data rates (Latha & Suganthi, 2015; Sundarambal, Dhivya, & Anbalagan, 2010).

BACKGROUND

Mobile Systems and 4G

The use of mobile wireless data services continues to increase worldwide. Mobile systems focus on effortlessly integrating the current wireless technologies including global system for mobile communication, wireless local area networks, and Bluetooth. Ultimately, 4G systems support comprehensive and personalized services. These factors are essential in providing stable system performance and quality service. According to Akintoye (2013), the 4G platform as a mobile multimedia, universally available, global mobility support, integrated wireless, and e-customized service network system. It was designed to use a number of types of broadband wireless access communication systems and to work side-by-side cellular telephone systems.

The official 4G-based requirements are set and published by International Telecommunication Union (ITU). It specified a number 4G-based requirements that can be grouped into 5 types: data rates (i.e., cell spectral efficiency, peak spectral efficiency, cell-edge user spectral efficiency, mobility data rates), latency, handover interruption

time, RF bandwidth, and number of active voice-over-Internet (VoI) protocol users (Yang, 2012). A 4G cellular system must have target peak data rates of up to approximately 100 Mbit/s for high mobility, such as mobile access. For low mobility, one needs data rates up to approximately 1 Gbit/s, such as local wireless access. The higher 4G data rates are available in those areas in which 4G cellular service is provided. They should provide scalable bandwidths of up to at least 40 MHz. A 4G system is expected to provide a comprehensive and secure all-IP based solution where facilities such as IP telephony, ultra-broadband Internet access, gaming services, and high-definition television streamed multimedia may be provided to users. It is very helpful for any organizations' compliance with e-Discovery requests.

In 4G networks, users are joining the network through add mobile routers to the network infrastructure. Wherever the concentration of customers is in more in one area, additional routes are created, enabling further access to network capacity. Hence, network capacity and coverage has dynamically shifted to make available more appts and high-speed connections available to suit changing user patterns. This enhanced flexibility permits the network to vigorously and routinely balance capacity and increase network utilization. As with web-based technologies, 4G is playing a significant role in promoting social networking and social media uses.

Business Audience for 4G

4G wireless communications are used in many businesses throughout the global economy and its vast communication networks to enhance and improve productivity. Each of the 5 major U.S. network operators (e.g., Verizon Wireless: 142.8 million (Q2 2016), AT&T Mobility: 133.3 million (Q3 2016), T-Mobile US: 69.4 million (Q3 2016), Sprint Corporation: 60.2 million (Q3 2016), U.S. Cellular: 4.9 million (Q2 2016)] (“List of United States wireless...” 2016) offers a slightly different version of 4G wireless communication services

that provide significantly faster data speeds than its 3G networks. Verizon is vastly expanding its 4G network based on the Long Term Evolution (LTE) standard.

LTE is a standard for wireless communications for mobile phones based on packet switching. This is an entirely different approach from the circuit-switching approach employed in 3G telecommunications networks. Carriers must basically recreate their voice call networks to convert to the LTE standard. The Sprint version of 4G network is based on the WiMAX (Worldwide Interoperability for Microwave Access) standard. This standard is of a faster speed than traditionally offered 3G networks, but not as fast as networks based on LTE. WiMAX operates like Wi-Fi in many aspects, only over greater distances and at faster transmission speeds. Fewer WiMAX base stations are required to cover the same geographical area than when Wi-Fi technology is used. The AT&T and T-Mobile 4G networks are based on the HSPA+ (Evolved High Speed Packet Access) standard, with AT&T speedily converting from HSPA+ to the LTE standard. HSPA+ is a broadband telecommunications standard. These network providers rely heavily on 4G to operate their routine business operations in order to be profitable in this highly competitive environment.

Advantages of 4G

When 3G was introduced to the market, the technology did not have enough advantages compared to the 2G. Obviously, it was an improvement compared to 2G, but 3G did not bring consumers to transition fast from 2G to 3G wireless network technologies, perhaps due to speed and options-related improvements were not significant enough for consumers to invest money and change their devices. However, that was not the case when 4G wireless mobile communication networks were developed. 4G had way more advantages over previous networks such as 2G or 3G, so everyone wanted to adapt to the new technology and enjoy all the advantages it could offer.

An important advantage that 4G offered technology was user friendly. Another goal was to provide countless services to the users at relative high speeds. For example, making devices to recognize the user's voice and execute appropriate commands that allow users to achieve their goals more efficiently and in less time. Perhaps, one of the most significant advantages of 4G is that it provides opportunity to access the Internet basically from anywhere by using wireless devices. This advantage helps to provide enhanced user personalization opportunities. For example, when a user is accessing repository of data or other services, the user has the ability to set up service features that would be used the most.

Terminal and Network heterogeneity is another advantage provided by 4G wireless networks, and in other words, it means service personalization. This gives the opportunity to pick an access to the network between WiMAX, Wi-Fi, or Universal Mobile Telecommunications System. This 4G wireless network provides high performance. One of the most impressive facts about 4G is its download speeds when compared to 3G technologies. In general, 4G networks are about 260 times faster than previous generations (i.e., 3G wireless networks). Interoperability is one of other many advantages this technology offers, for example, it gives options for users to use all kind of services from multiple suppliers by using the same mobile device and nowadays this is an important thing to have because there are so many options and everyone would like to have access to those options.

One of the most impressive advantages is its Intelligent Networking. 4G wireless technologies are trying to use Wireless LAN and WAN designs to promote ease of use and widespread acceptance. Many businesses and educational institutions are moving more and more towards e-commerce and web-based standards (Chand, Raj, & Shankar, 2015; Elysee, 2015; Han, Ada, Sharman, Gray, & Simha, 2015; Latha & Suganthi, 2015; Shukai, Chaudhari, & Dash, 2010; Sundarambal, et al., 2010). So, in order to conduct business more effectively, Internet access is needed, so many

businesses, in order to attract more customers, decided to create Wi-Fi spots in their facilities and share it with their customers (Daim, Basoglu, & Tanoglu, 2010; Daramola, Oladipupo, & Musa, 2010; Dominic, Goh, Wong, & Chen, 2010). As evident in the wide-spread practice of Wi-Fi and 4Gspots in malls, restaurants, libraries, service centers, etc.; hence, not having the opportunity to get connected to Wi-Fi, or forced to use GPRS services can be a great disadvantage to businesses and other institutions in attracting customers. However, then the place with the Wi-Fi connection is reached by the user; mobile device would automatically switch to use this connection (LAN design) without any actions performed by the customer. This is called Intelligent Networking and is thought to be a significant advantage.

Network convergence is a major advantage that 4G can provide to its users. One of the goals of creating 4G wireless networks was to integrate it to the global IP based network, and the plan was to blend devices and services together. Since this innovation was getting popular, scalability was really important, and it was solved by using IPv6 addressing scheme, which can provide access to a large number of users who have wireless devices. This eliminates the need for Network Address Translation (NAT). Users of this technology are still increasing and scalability provides a big advantage. Those kinds of technologies, such as 2G or 3G, have huge effect on battery life; however, 4G is trying to minimize that problem and convert it into advantage. A possible solution that was proposed was to reduce battery usage by this technology to create shorter communication links. Finally, another advantage concerning 4G wireless technology is that does not generate significant expenses or burdens for users. They do not need to buy any additional accessories in order to make it more efficient. In-other-words, it is not a completely new system, but it is built on the existing networks, so it supports the use of 2G or 3G. This is one of the reasons why it is cheaper and does not require any additional spending for the users.

4G provides way more advantages compared to what 3G had to offer when it was released. The above listed advantages are the reasons why users are using more and more of 4G wireless services nowadays. 4G made it really convenient for the user to adopt this new technology from the beginning, made it more scalable, efficient, and most importantly, there is little additional costs associated with the adoption this new technology (Aeron, Kumar, & Janakiraman, 2010; Barra, Savage, & Tsay, 2010; Smith, 2013, 2014; Smith & Synowka, 2014; Smith & Rupp, 2013; Smith, Synowka, & Smith, 2014)..

Challenges Facing 4G Networks

There are a number of challenges 4G wireless network faces, such as security and privacy issues, quality of service, and complex architecture (Sinha & Thirumalai, 2011; Smock, Ellison, Lampe & Wohn, (2011). These are the challenges that should be overcome in order to provide more efficient and safer wireless network for the users. Security and privacy issues are really important in today's world because hackers are accessing users' private information, such as home address, credit card codes, and/or social security number. So, there are urgently needs to provide users with safe ways to send and receive information through the network. 4G networks may higher likelihood of security attacks due to its widespread adoption, so preventative measures and additional security levels created such as identification of the user and identification of IP addresses from where a user connects to the network or passwords should be safeguarded. These methods might help to protect the data, which is about to be transferred through the network.

4G wireless networks are trying to provide the highest quality of service, such as improved connectivity, faster speed, and enhanced accessibility. Unfortunately, one of the challenges it faces is connecting non-IP-based and IP based devices. These two devices are used for different purposes. Non-IP based devices are used for Voice-over Internet Protocol (VoIP), while IP

based devices are usually used for transferring data. 4G networks are trying to serve both devices, but creating mechanisms in order to do that effectively creates protocols and connectivity problems that must be addressed.

Another challenge facing 4G wireless network integration is its complex architecture. In order to reduce the operating cost or power consumption, networks should be able to be operational and functional in different working environments. Multi-mode devices were one of the suggested solutions in order to fix this problem because it allows for the device to adapt to various networks by using a software radio. Another challenge 4G faces is system discovery and selection because wireless devices are receiving all kinds of signals from various systems and after processing those signals its need to determine the most appropriate service provider in order to provide proper connections. Some potential solutions to connectivity problems might not be available in a timely manner due to their regulations, rules or location. However, the solution proposed to solve this challenge was system-initiated discoveries, and this program is able to find automatically the best software module. Lastly, supervising customers' accounts and sending bills to their 4G network made it more complicated and the main reason is heterogeneity of 4G networks.

Successful Application of 4G

Technology is continuously evolving and even over the past decade there have been drastic changes to accompany how individuals in society operate and communicate to each other. In years past, cellphones were made for solely making phone calls, while smart phones and tablets and their numerous applications were simply fantasy or science fiction. People not only rely on cellphones, tablets, and computers for making phone calls and sending e-mails, but they are used to do activities such as banking, send text messages, video chat with someone in a different location, shop online, and so much more.

4G networks are widely advertised and advanced the need for widespread wireless communication connectivity. One of its main features of 4G networks over previous networks is that it provides high-speed connections and it gives users the opportunity to utilize these networks outside of their personal space. Other advantages this network offers includes: intelligent networking, network convergence, scalability, lowers power consumption, and lower operational costs. Not only are individuals utilizing these services for personal day-to-day needs but businesses have become involved and are benefiting from the 4G network as well. One business in particular is the White House Utility District (WHUD), which is located in White House, TN.

The WHUD is one of the largest water and sewer utilities in Tennessee, serving a large population of 90,000 with over 30,000 connections within a 600-square-mile ("White House Utility District," 2016). The use of a 4G network has drastically improved communication, effectiveness, and the overall quality of how the WHUD operates and manages the communication needs of both its employees and their customers. An important benefit of 4G is that it allows maintenance crews to be alerted quickly, ultimately resulting in faster response times. This is due in part that essentially all of the employees of WHUD are provided with 4G phones and tablets. Other 4G network-based services offered that benefit customers include the ability to pay bills online, check account balances, establish or discontinue service with the company, update account information. For example, there are educational programs that are offered to inform customers about its efforts to ensure safe drinking water by logging on to their website using the 4G network.

WHUD has developed a way to better keep track of their employees, customer calls, and water outages. Each employee is equipped with a tablet that allows the customer to log into the WHUD application. This feature is currently available to the company's workers and management staff. Such applications allow direct communication within

the company from managers to workers. It connects the customers to the workers. As customers call the utility department to report problems, the operator begins to issue alerts to the crews in the field. The alert is then sent to the closest crew to location of the customer. Once the crews arrive at the location they are instructed to update the job information. Once the job information is updated managers know exactly what the crew has found and what machines will need to be sent to the job site. There are many benefits the employees themselves get from this application.

Benefits for employees include: being able to log job site information on to tablets, allowing workers to clock in and clock out. The use of the 4G network allow managers to inform each crew that their work schedule are and what each job involves. Before such 4G networks were available, miscommunications could often occur between management and employees in the field, since not all pertinent information could be logged in a timely manner. Unfortunately, without such networks, some employees might not be given full amount of pay because hours could not be logged while out on the field and there would little information related to time an employee was actually spent at the job site. With 4G, crews are immediately clocked in when the alert is sent through the application and they click to verify that they are on their way. Management was able to track their vehicles through this application as well. This tracking capability allowed managers to know where each truck is and who is driving them to ensure that employees were proper compensated.

WHUD is equipped with its own IT technicians. These technicians work primarily on the servers that transmit information from the call centers to the maintenance crews in order to better serve the community. Such applications through the 4G network allow multiple servers and hardware devices to connect the call centers, managers, and employees in the field. These technicians are on call 24/7 and need to make sure everyone is able to connect in order to allow fast response times.

Working through the 4G network allows great mobility to the infrastructure but requires a lot of upkeep. Technicians are required to update the servers and the application. Updates are required monthly unless any software bugs are found that hinder the use of the application. If bugs are found, the technicians immediately begin to solve the issues. Most issues are fixed within a day, but there has been a period of a 2-day down time. According to a current manager with WHUD, during this 2-day down time customers and employees realized how much the 4G networks and its apps help them in routine operations.

As suggested by interviews conducted by the authors of this chapter, this 2-day downtime was treacherous and possibly dangerous to their work crews. In a particular example, one of the interviewees stated that the 2-day down time was due to a storm that came through the Sumner County Area. This storm disconnected one of the servers for WHUD, eliminating the transmission of information to the employees. Employees could not contact managers. Calls centers were forced to hand-dial every members' work phone, call each crew to determine their locations, and finally put a crew in route to the issue. This task took up to 2 hours as compared to what would take only about 5 minutes when the 4G network was fully operational. The manager stated how relieved management and staff were when the first alert came back through the network. Technicians said that it was one of the worst outages they had seen. Since then there have been IT plans in place incase another outage like this were to happen. Customers benefit from the use of 4G networks being able to connect to their utility department while they complete other tasks outside their homes and offices.

WHUD provides customers with a user-friendly website. On this website, there is a link for frequently asked questions which not only provides responses to the questions, but it provides short videos to some of those questions. In particular, a couple of those frequently asked questions to this company that are accompanied with videos are:

“How do I shut my water off in my house in the case of an emergency?” and “How do I check for a leak?” Being able to learn this information can be essential for a customer. If a customer runs into a problem such as a water line bust in their house they can quickly connect to the WHUD website and turn their water off so they do not have to wait for a service crew. If a water pipe busts in a house it is only minutes before a lot of damage occurs.

Overall, 4G networks have proven to be beneficial to not only individuals but to businesses as well. Such networks offer faster services, the opportunity to build a bigger and better business, as well as the ability to make necessary advancements to keep up with society’s ever changing technology (Elysee, 2015; Latha & Suganthi, 2015)..

FUTURE RESEARCH DIRECTIONS

5G Networks

The future of wireless telecommunications networks may be difficult to predict, but its current growth is well established. Although there are not formal dates have been set yet for the development of Fifth Generation (5G) Networks yet, but a number of companies and organizations have begun establishing their own timelines to anticipate the growth of the impending transition within the industry. It is generally estimated that the timescale for the first 5G networks will be around 2020; however, the useful lifetime for 5G is likely to take time to develop, as it is aimed at providing general connectivity with many applications needing to remain in place for many years. Utility meters, for example remain in place for many years, and the utility companies will not take kindly to having to replace their meters more frequently to follow the cellular technologies.

It is important to understand that the impact of 5G will extend well beyond telecommunications. By connecting people, machines, and things on a massive scale, 5G will facilitate the delivery of personalized healthcare, optimize transportation

logistics, enhance worldwide access to culture and education, and quite possibly virtually revolutionize public services. Before exploring the future of telecommunications, many elements must be considered and the concluding pages of this chapter will cover topics such as standardization, development of new technology, and ultimately the release of much needed spectrum speeds. These elements, among others, will need to be managed closely throughout infrastructure and process development and expertly combined to enable system launch within the required timescale.

To expand upon elements needed to enable future 5G networks, one major consideration will be the release of frequency spectrum. Radio Access Network (RAN) is one wireless network in particular which uses radio frequencies to communicate between any remotely controlled machine (i.e., devices such as mobile phones or computers) and their core networks. RAN includes the same technologies used in base stations and is anticipated to take an estimated six months to reach agreements on the bulk of 5G network RAN requirements.

Globally, strict management will be needed to ensure commonality and reduce the interference between services, especially those operating internationally. Organizations such as the International Telecommunications Union (ITU) will be called on for oversight in this daunting task. Large international meetings called by the World Radio Conference (WRC) are held every four years, with the next scheduled to take place throughout November of 2015 in Geneva, Switzerland (“World Radio Communication Conference,” 2015).

At the World Radio Conference, the main focus for mobile telecommunications is centered on providing additional frequency spectrums for 4G services, seeing as work is not yet sufficiently advanced for determining allocations for 5G. However, by 2019 the WRC anticipates that 5G developments will be sufficient enough to sustain telecommunications initiatives. Although the earliest deployments for 5G may occur around 2020, these will be comparatively limited and its use not

as wide as it is anticipated to be later. Accordingly, the dates for spectrum release for 5G will grow progressively as 5G deployments increase, bandwidth requirements grow, and technology is made available to the consumer (“World Radio Communication Conference,” 2015).

Technology development started shortly after the first 4G deployments. Currently, the most promising advance is multiple input multiple output (MiMo) technology, which uses multiple small antennae to service each individual data stream. Samsung successfully tested advanced MiMo antennae with data speeds at 1 gigabyte per second in May, 2013. The timeline for the research phase can only be broadly bounded and work will be ongoing even after the system enters service, it is anticipated that the basic research and investigations will need to be complete by around 2016 to enable this to feed into the standardization process.

Standardization, just like in any industry, is a key element of the 5G development process whereas the timescales and dates of this activity are integral to the successful deployment of 5G. Standardization of 5G will involve several agencies, of which the 3rd Generation Partnership Project (3GPP), The European Telecommunications Standards Institute (ETSI), Next Generation Mobile Networks (NGMN), and Institute of Electrical and Electronics Engineers (IEEE) are a few. Why are these agencies so important in any standardization effort? As their collective efforts were used to establish the definition of the cellular standard and one of the most important authorities in this arena is 3GPP, so having their collaboration is essential for any standardization effort. Although, with 5G emerging as collaboration of technologies, it is likely to require a number of standards institutions to reciprocate the aggregation in order to make 5G widespread and acceptable to the global economy.

Cost Considerations of 5G

Undoubtedly, global communications reached a point of unprecedented change in both mobile and fixed networks. Each attempts to meet the insatiable demand for data that has been driven by smart phones and tablets. Currently, these devices use both 4G and Wi-Fi radio technologies to access services via both mobile and fixed network cores, but developments in the next generation of wireless communication networks are rapidly moving to commercialization.

Although, it is not possible to know how 5G-network technologies will affect mobile phone and data rates, it should have major impacts on the mobile communication industry. According to a recent article (“A look into the future ...,” 2014), the South Korea’s Minister of Engineering, Science and Technology estimated that private companies will probably spend over US\$300 billion in 5G infrastructure and technology. This coordinated effort is designed to 5G worldwide and the intensity of competition in the industry should be high. Data prices have continuously fallen, from approximately 46 cents per megabyte in 2008 to 2-3 cents in 2014. It is hoped that successful 5G networks would provide 50 gigabytes of data per person per month, a significant improvement in the current offerings of 4G. What will be different? Currently, there are no 5G phones and devices to use the new technology, although manufacturers like Samsung, LG, and HTC are experimenting with new designs. No one knows with certainty what features a 5G phone or device will have. Innovative mobile service providers, companies like Netflix, will have to decide whether they want to make their data available at 5G speeds, so there are many variables affecting the future of mobile technology.

CONCLUSION

Wireless networking has evolved quickly over a very short timeframe. Historically, 2G network focused on real-time audio, 3G focused on data, and 4G has been centered largely on providing relatively high-quality video. The specifics on future network developments are still unclear, and while the issue is mind-boggling and complex, 5G networks will probably usher in an entirely new era all about making connections. 4G wireless networks have forever revolutionized the way people communicate, and the challenges still facing service providers today will likely become advantages under 5G infrastructure. Global implementation is inevitable, with mobile networks estimated to support over 7 trillion connected devices beginning in just a decade. The mobile communications industry is turning towards future innovations that will forever change lifestyles.

REFERENCES

- A look into the future: 5G network speeds and when they'll arrive. (2014). Retrieved March 19, 2016 from <http://tech.co/look-future-5g-network-speeds-theyll-arrive-2014-03>
- Aeron, H., Kumar, A., & Janakiraman, M. (2010). Application of data mining techniques for customer lifetime value parameters: A review. *International Journal of Business Information Systems*, 6(4), 530–546. doi:10.1504/IJBIS.2010.035744
- Akintoye, S. B. (2013). Wireless mobile communication - a study of 4G technology. *Kuwait Chapter of the Arabian Journal of Business and Management Review*, 2(9), 42–52. doi:10.12816/0001232
- Alderete, M. V., & Gutiérrez, L. H. (2014). Drivers of information and communication technologies adoption in Colombian services firms. *International Journal of Business Information Systems*, 17(4), 373–397. doi:10.1504/IJBIS.2014.065553
- Barra, R. A., Savage, A., & Tsay, J. J. (2010). Equational zero vector databases, non-equational databases, and inherent internal control. *International Journal of Business Information Systems*, 6(3), 354–377. doi:10.1504/IJBIS.2010.035050
- Chand, M., Raj, T., & Shankar, R. (2015). A comparative study of multi criteria decision making approaches for risks assessment in supply chain. *International Journal of Business Information Systems*, 18(1), 67–84. doi:10.1504/IJBIS.2015.066128
- Daim, T., Basoglu, N., & Tanoglu, I. (2010). A critical assessment of information technology adoption: Technical, organisational and personal perspectives. *International Journal of Business Information Systems*, 6(3), 315–335. doi:10.1504/IJBIS.2010.035048
- Daramola, J. O., Oladipupo, O. O., & Musa, A. G. (2010). A fuzzy expert system (FES) tool for online personnel recruitments. *International Journal of Business Information Systems*, 6(4), 444–462. doi:10.1504/IJBIS.2010.035741
- Dharni, K. (2014). Exploring information system evaluation in Indian manufacturing sector. *International Journal of Business Information Systems*, 17(4), 453–468. doi:10.1504/IJBIS.2014.065564
- Dominic, P. D. D., Goh, K. N., Wong, D., & Chen, Y. Y. (2010). The importance of service quality for competitive advantage – with special reference to industrial product. *International Journal of Business Information Systems*, 6(3), 378–397. doi:10.1504/IJBIS.2010.035051
- Elysee, G. (2015). An empirical examination of a mediated model of strategic information systems planning success. *International Journal of Business Information Systems*, 18(1), 44–66. doi:10.1504/IJBIS.2015.066126
- Gupta, R., & Naqvi, S. K. (2014). A framework for applying critical success factors to ERP implementation projects. *International Journal of Business Information Systems*, 17(4), 469–490. doi:10.1504/IJBIS.2014.065565

- Han, W., Ada, S., Sharman, R., Gray, R. H., & Simha, A. (2015). Factors impacting the adoption of social network sites for emergency notification purposes in universities. *International Journal of Business Information Systems*, 18(1), 85–106. doi:10.1504/IJBIS.2015.066129
- Kapur, P. K., Gupta, A., Jha, P. C., & Goyal, S. K. (2010). Software quality assurance using software reliability growth modelling: State of the art. *International Journal of Business Information Systems*, 6(4), 463–496. doi:10.1504/IJBIS.2010.035742
- Keramati, A., & Behmanesh, I. (2010). Assessing the impact of information technology on firm performance using canonical correlation analysis. *International Journal of Business Information Systems*, 6(4), 497–513. doi:10.1504/IJBIS.2010.035743
- Kim, E., & Tadisina, S. (2010). A model of customers initial trust in unknown online retailers: An empirical study. *International Journal of Business Information Systems*, 6(4), 419–443. doi:10.1504/IJBIS.2010.035740
- Latha, T. J., & Suganthi, L. (2015). An empirical study on creating software product value in India - an analytic hierarchy process approach. *International Journal of Business Information Systems*, 18(1), 26–43. doi:10.1504/IJBIS.2015.066125
- List of United States wireless communications service providers. (2016). Retrieved October 27, 2016 from https://en.wikipedia.org/wiki/List_of_United_States_wireless_communications_service_providers
- Marthandan, G., & Tang, C. M. (2010). Information systems evaluation: An ongoing measure. *International Journal of Business Information Systems*, 6(3), 336–353. doi:10.1504/IJBIS.2010.035049
- Qrunfleh, S., Tarafdar, M., & Ragu-Nathan, T. (2012). Examining alignment between supplier management practices and information systems strategy. *Benchmarking: An International Journal*, 19(4), 604–617. doi:10.1108/14635771211258034
- Radu, I. O., Horațiu, S. C., Bogdan, B. L., & Mihai, G. (2013). Aspects regarding the role of information technologies in the assurance of supply chain management performance. *Annals of the University of Oradea. Economic Science Series*, 22(1), 1495–1504.
- Shukai, L., Chaudhari, N. S., & Dash, M. (2010). Selecting useful features for personal credit risk analysis. *International Journal of Business Information Systems*, 6(4), 419–443. doi:10.1504/IJBIS.2010.035745
- Sinha, K. K., & Thirumalai, S. (2011). Customization of the online purchase process in electronic retailing and customer satisfaction: An online field study. *Journal of Operations Management*, 29(5), 477–487. doi:10.1016/j.jom.2010.11.009
- Smith, A. A., Synowka, D. P., & Smith, A. D. (2014). E-commerce quality and adoptive elements of e-ticketing for entertainment and sporting. *International Journal of Business Information Systems*, 15(4), 450–487. doi:10.1504/IJBIS.2014.060397
- Smith, A. D. (2013). Online social networking and office environmental factors that affect worker productivity. *International Journal of Procurement Management*, 6(5), 578–608. doi:10.1504/IJPM.2013.056173
- Smith, A. D. (2014). Case Studies of RFID Practices for Competitive Inventory Management Systems. In J. Wang (Ed.), *Management Science, Logistics, and Operations Research: A volume in the Advances in Logistics, Operations, and Management Science (ALOMS) Book Series* (pp. 1–25). doi:10.4018/978-1-4666-4506-6.ch001
- Smith, A. D., & Rupp, W. T. (2013). Data quality and knowledge/information management in service operations management: Regional supermarket case study. *International Journal of Knowledge-Based Organizations*, 3(3), 35–52. doi:10.4018/ijkbo.2013070103

- Smith, A. D., & Synowka, D. P. (2014). Lean operations and SCM practices in manufacturing firms: Multi-firm case studies in HRM and visual-based metrics. *International Journal of Procurement Management*, 7(2), 183–200. doi:10.1504/IJPM.2014.059554
- Smock, A. D., Ellison, N. B., Lampe, C., & Wohn, D. Y. (2011). Facebook as a toolkit: A uses and gratification approach to unbundling feature use. *Computers in Human Behavior*, 27(6), 2322–2329. doi:10.1016/j.chb.2011.07.011
- Soon, J. N. P., Mahmood, A. K., Yin, C.-P., Wan, W.-S., Yuen, P.-K., & Heng, L.-E. (2015). Barebone cloud IaaS: Revitalisation disruptive technology. *International Journal of Business Information Systems*, 18(1), 26–43. doi:10.1504/IJBIS.2015.066130
- Sundarambal, M., Dhivya, M., & Anbalagan, P. (2010). Performance evaluation of bandwidth allocation in ATM networks. *International Journal of Business Information Systems*, 6(3), 398–417. doi:10.1504/IJBIS.2010.035052
- White House Utility District. (2016). Retrieved March 21, 2016 from http://whud.hcents.com/about_us.cgi
- World Radio Communication Conference 2015. (2015.). Retrieved April 29, 2015, from <http://www.itu.int/en/ITU-R/conferences/wrc/2015/Pages/default.aspx>
- Yang, S. C. (2012). Mobile applications and 4G wireless networks: A framework for analysis. *Campus-Wide Information Systems*, 29(5), 344–357. doi:10.1108/10650741211275107
- Alexopoulou, N., Nikolaidou, M., Kanellis, P., Mantzana, V., Anagnostopoulos, D., & Martakos, D. (2010). Infusing agility in business processes through an event-centric approach. *International Journal of Business Information Systems*, 6(1), 58–78. doi:10.1504/IJBIS.2010.034005
- Ali, M., El-Haddadeh, R., Eldabi, T., & Mansour, E. (2010). Simulation discounted cash flow valuation for Internet companies. *International Journal of Business Information Systems*, 6(1), 18–33. doi:10.1504/IJBIS.2010.034002
- Althonayan, A., & Sharif, A. M. (2010). Aligning business and technology strategy within the airline industry. *International Journal of Business Information Systems*, 6(1), 79–94. doi:10.1504/IJBIS.2010.034006
- Badii, A., & Thiemert, D. (2010). A framework towards a multi-modal fingerprinting scheme for multimedia assets. *International Journal of Business Information Systems*, 6(1), 133–149. doi:10.1504/IJBIS.2010.034009
- Chahal, K., & Eldabi, T. (2010). A multi-perspective comparison for selection between system dynamics and discrete event simulation. *International Journal of Business Information Systems*, 6(1), 4–17. doi:10.1504/IJBIS.2010.034001
- Ocass, A., & Fenech, T. (2003). Web retailing adoption: Exploring the nature of internet users Web retailing behaviour. *Journal of Retailing and Consumer Services*, 10(1), 81–94. doi:10.1016/S0969-6989(02)00004-8
- Oetega, B. H., Martinez, J. J., & Hoyos, M. J. M. (2006). Analysis of the moderating effect of industry on online behaviour. *Online Information Review*, 30(6), 681–698. doi:10.1108/14684520610716162

ADDITIONAL READING

Al-Karaghoul, W., & Fadare, E. B. (2010). The impact of information lifecycle management process in the Nigerian financial sector. *International Journal of Business Information Systems*, 6(1), 111–132. doi:10.1504/IJBIS.2010.034008

KEY TERMS AND DEFINITIONS

e-Discovery: This term refers to all types of electronic information storage (e.g., e-mails,

directories, documents, searches, etc.) that are either targets investigations in which the affected organization need to produce evidence of such transactions. In other cases, information have been mistakenly lost or deleted and managers or employees need to retrieve message-level data.

Multiple Input, Multiple Output (MiMo)

Technology: MiMo, or often referred to as antenna technology, is a term that refers to multiple input, multiple output, which is basically an antenna technology for wireless communications. This technology, multiple antennas are used at both the source (i.e., transmitter) and the destination (i.e., receiver) and the antenna arrays are combined to minimize errors and optimize data speed.

Long Term Evolution (LTE): LTE is a commonly accepted standard for wireless communications for mobile phones based on packet switching. This is an entirely different approach from the circuit-switching approach employed in 3G telecommunications networks.

Radio Access Network (RAN): RAN is one-wireless network in particular which uses radio

frequencies to communicate between any remotely controlled machine, typically mobile phones or computers, and their core networks.

Terminal and Network Heterogeneity:

The next generation wireless networks should be heterogeneous (e.g., consisting of multiple radio access technologies that simultaneously coexisting in the same geographical area). Such heterogeneous wireless networks should consist of mobile terminals of different capabilities (i.e., heterogeneous terminals) that can be used by customers to access network services. Typically, heterogeneous mobile terminals (e.g. single-mode, dual-mode, triple-mode, etc.) use call-blocking and call-dropping capabilities in a cooperative and shared wireless environment.

Worldwide Interoperability for Microwave Access (WiMAX) standard: WiMAX is a global wireless industry coalition created to foster the acceptance and advancement of IEEE 802.16 standards for broadband wireless access (BWA) networks.

Flying Adhoc Networks Concept and Challenges

Kuldeep Singh

Thapar University, India

Anil Kumar Verma

Thapar University, India

INTRODUCTION

In case of a calamitous event, when traditional communication methods are out of service or simply not available, in those situations mobile adhoc networks (MANETs) plays an important role in established communication. MANETs employ grounded mobile nodes which are capable to collect information with the help of sensor, camera and other devices. Nodes in MANETs communicate with each other through wireless communication without use of existing infrastructure. The data collected with nodes in MANETs are transmitted to the base station using a multi-hop path. These networks are economical and can be formed in quickly whenever required as they do not require existing infrastructure. MANETs have several application areas such as natural disaster, sensor networks, etc.. But there are some extreme situations (such as flooding, battlefield and rescue operations, etc.) where MANETs cannot be deployed. In those situations, flying ad-hoc networks (FANETs) can play vital role in established communication. FANET is a subclass of MANETs and made up of a swarm of small flying vehicles enable with camera, sensor and GPS system. Swarms of UAVs arrange themselves to communicate with large operational area using wireless network without any centralized device. According to muller (2012) UAVs communicate with each other locally, with base station and also interact with

their environment to get information. FANETs use different types of UAVs based on the various application areas. FANETs employ unmanned aerial vehicle (UAVs), UAV is an aircraft which flies without a pilot. The UAV can manage itself and fly based on preprogrammed flight plans or can be operated using complex dynamic automation systems which are versatile and flexible in implementation (Muller, 2012). FANETs use multi-UAVs to perform operations because of the limitations of a single UAV system such as limited surveillance capability, scalability and flexibility (Bekmezci, 2013). The multi-UAV system has various advantages over single UAVs such as:

- With more number of UAVs, tasks can be parallelized which in turn reduce the completion time of mission. This kind of behavior is very useful for search and rescue applications;
- In case of a single UAV system, whole mission collapse if UAV fails. In multi-UAV systems, nodes can distribute tasks among themselves. Fault tolerance of the network increases with multi-UAVs;
- In multi-UAV environment, heterogeneous UAVs can form a network. It is possible to use capabilities of other UAVs as and when required for task completion. So multi-UAV systems are very advantageous for critical applications.

FANET Applications

Due to various advantages and wide range of application areas FANETs are getting attention from the research community around the globe. Various applications of FANETs are explained and shown in Figure 1:

- **Military Services:** FANET are very useful in military services. Setting up the proper communication system is very difficult in military areas. So FANETs are used for information exchange among soldiers, military headquarters;
- **Security Purpose:** FANET is capable of receiving information quickly. It can be used to collect information for the security purpose of a delegate visiting to a place where no network infrastructure exists;
- **Calamity Administration:** FANET is useful when the existing communication system is damaged due to natural disaster like floods, earthquake, firing, etc. Gathering and sending information in such areas becomes difficult in such situation. FANET provides the necessary communication;
- **Search and Rescue Operations:** FANET can be used to provide a better way to do search and rescue operations such as rescue of hostages (Bekmezci, 2013). Sometimes in extreme situations, cellular networks get damaged. FANETs provide better rescue services in such conditions by sending periodic updates to other locations;
- **In Sensor Networks:** Different sensor devices can be used to collect data to do daily functions like weather forecasting, terrestrial movement tracking, etc. FANETs can approach to any remote location without difficulties;
- **Location Aware Services:** FANETs can be used in many services like forwarding calls to any location, travel guide for passengers, identify information regarding specific location;

Figure 1. FANET applications



- **City Architecture Planning:** FANETs are designed to focus emergency services and military applications, but FANETs can be used in the development of a city. FANETs can be used to develop city architecture. UAVs can help to plan and manage architectures of roads and buildings in a city based on its geographical location;
- **Law Enforcement:** UAVs capabilities can be utilized for law and order of a city. FANETs can provide surveillance facilities on the roads and the streets of a city in situations of riots and curfew, FANETs can collect video recordings and images to support police or soldiers to control the situation. Later images and videos can be used as evidence in the court against the accused.

BACKGROUND

The term FANETs is first introduced by the Michael Muller in 2012. Before FANETs, MANETs were the best way to communicate in emergency situations, as they can be established quickly, easy to deploy and economical. MANETs employ grounded moving nodes, which are capable to communicate each other without using pre-existing infrastructure. These networks are also known as grounded ad-hoc networks. MANETs have some limitations which force researchers find a better solution for those situations where MANETs cannot be useful. Researcher's wants to use the capability of UAV in ad-hoc network, so they developed single large UAV guided ad-hoc network. In this network, UAV collects the information and passes it to the grounded, moving nodes, which help grounded nodes to perform task efficiently. But single UAV based ad-hoc network cannot accomplish all the tasks by itself. Finally, research community came up with an idea of using multiple UAV to create ad-hoc networks. The idea behind the FANET is to overcome with the limitations of grounded ad-hoc network by

using the advantages of multiple UAVs together to perform tasks. In FANETs, a swarm of UAVs established communication link between them. They collect information and transfer it to the grounded nodes with the help of established links. Due to the wide range of FANETs applications, the research community is actively participating to formulate efficient infrastructure, routing and security mechanisms for FANETs.

FANET Concept

FANETs are basically adhoc network created by a group of UAVs, which is a new family of adhoc networks. Single UAV guided adhoc networks have some limitations as single UAV cannot collect information from the whole geographical area of the operation independently. So, multiple UAVs are used to form adhoc networks called FANET. As, FANETs uses multiple UAVs, so they improve scalability and flexibility ad-hoc networks. Each UAV is equipped with sensors, on-board monitor, GPS receiver and autopilot. The UAVs are fixed wing aircrafts with autopilot, capable of flying up to 12 m/s and up to 3 km/s in emergency situations. UAVs are available in different sizes for different applications. Different sizes of UAVs are: Micro UAVs, Low altitude and low endurance UAVs and high altitude and high endurance UAVs. The swarm of UAVs communicates among themselves and with the ground station. There are two types of communication:

- **UAV to UAV Communication:** In UAV to UAV communication, UAVs communicate with each other to complete the allocated task. Two UAVs can either directly communicate or through other UAVs. To communicate indirectly using multiple hops, a communication path is decided based on different routing strategies like reactive, proactive and hybrid routing. UAVs can have either short range or long range communication; it depends on requiring data transfer rate;

- **UAV to Ground Station Communication:** In UAV to ground station communication, UAVs connect to the ground stations to provide or receive information about various operations. Ground stations are basically known as base stations that instruct UAVs for various tasks.

Flying nodes have very high mobility and topological changes, so existing solutions of MANETs cannot be directly applied to FANETs. To formulate solutions for FANETs, either existing solutions can be modified or new techniques can be formulated. To establish communication between nodes, various routing protocols found in literature. These protocols are categorized as follows:

- **Proactive Routing Protocol:** In these protocols, each and every node in the network shares its routing information from its routing table at regular time interval which is used by the other nodes to identify the path for destination nodes and make the map of the whole network. The big advantage of these protocols is that these protocols take a very short time period to get the path to the destination. But, it costs very much bandwidth consumption to update the information within a short period of time thorough which it maintains maps of the whole network. There are several proposed algorithms under this category like WRP, DSDV and OLSR (Singh & Verma, 2014);
- **Reactive Routing Protocol:** Reactive routing protocols are on-demand protocols, they build path whenever required by network nodes. The Reactive protocols do not broadcast their routing table information in regular time interval. They broadcast their routing information only when it is needed. Therefore, they minimize the use of network bandwidth. But, due to the reactive nature, there is a disadvantage to

these types of routing algorithms, End to End delay of packet delivery is increased as compared to proactive protocols. They also take more time to select an immediate node to transfer the data packet because of the dynamic network topology. Reactive protocols are less likely to use in applications in a dynamic environment. Many algorithms are proposed under this category like AODV, DSR and ABR;

- **Hybrid Routing Protocol:** Hybrid protocols are a combination of both proactive and reactive protocols. These protocols are designed to minimize the overhead, which occurs in both types of protocols (reactive and proactive). But, the protocols under this category are not suitable for large networks having more than hundred nodes because of large overlapping of zones like ZRP.

FANET Design Considerations

Due to distinctive features of FANET, it requires different design considerations than traditional ad-hoc networks. Here we discuss major design considerations: adaptability, latency, mobility, UAV platform constraints, scalability and bandwidth requirement:

- **Adaptability:** FANET nodes are highly mobile in nature. Due to this behavior of nodes, they keep on changing their location (Lin, Sun, Li & Yang, 2012). The routes between UAVs keep changing and distance between them is also not constant. Another problem is UAV failure that decreases the number of UAVs in networks. So the FANET design should consider these UAV failures and frequent path changes. The overall performance of the network depends on the adaptability on these path changes and topology changes;
- **Latency:** Latency is considered as one of the main design requirement of any net-

work. Latency is basically an application dependent factor. For real time applications of FANET like search and rescue operations, military application latency should be low for transferring information. These are time bound applications. For non-time critical applications such as city architecture, planning etc. the latency factor can be less compromised (Rosati et al., 2013);

- **Mobility:** Mobility of UAVs plays an important role in the performance of FANETs. There are several movement patterns for an adhoc network node. Moving pattern can be for group or individual node. It is very crucial to identify that which moving pattern of UAVs will be suitable to complete task effectively and in time bound manner;
- **UAV Platform Constraints:** UAVs used in FANETs have limited payload capability, so they have certain constraints. Hardware weight affects the performance of UAVs, light payload increases the endurance. So, UAV with a lighter payload have the scope for additional resources such as sensors and other peripheral devices. Another limitation faced by the UAVs used in FANETs is space limitation;
- **Scalability:** Multi-UAVs can enhance the overall performance of network as compared to single UAV systems. Performance is enhanced by increasing the number of UAVs in the network. Higher the number of UAVs, faster will be the task completion and more reliable will be the network. Scalability is an important factor for time dependent applications;
- **Bandwidth Requirement:** In every FANET application, UAVs collect data from environment with the help of cameras and different sensors, collected data is transmitted to ground base through multipath route. Data collected in applications such as monitoring, rescue, surveillance and battlefield are mainly higher resolution video, images and audio (Rosati et al., 2014)..

Also captured data is crucial and required to be transmitted with strict delay bound. So FANETs require high bandwidth resource to complete operation efficiently within time bound.

FANET CHALLENGES

FANET is the member of MANET family. FANET faces some additional challenges along with already existing challenges in MANET family due to its high node speed, high topology changes and mobility models. Based on the finding of literature review, following are the areas identified that require significant research to be done (see Table 1):

- **Routing:** Routing in FANETs is different from other adhoc networks family. Node movement is relatively very high in FANETs. So the topology changes very frequently. One of the biggest challenges is to develop an efficient routing algorithm that not only able to work with high mobility nodes, but should be quick to update its routing table frequently as the topology changes (Sahingoz, 2014);
- **Security:** Ensuring confidentiality, availability and Integrity of information during the communication between UAV to UAV communication and UAV to ground node communication is one of the major issues faced by FANETs (Yap et al., 2015). Due to lack of physical security node compromise becomes easy in FANETs. Trust Management among nodes is another challenge due to high topology changes. Nodes join and leave the network very frequently. Secure routing is another point of concern in FANET;
- **Quality of Service (QoS):** In FANETs UAVs transmit data includes audio, video, images, text, GPS locations, etc. To transfer such data FANETs should have a good

quality of service with less delays and error rates (Yanmaz et al., 2010);

- **UAV Mobility and Placement:** The placement of UAVs at appropriate location is one of the major research concerns in FANETs. UAVs of different capacity and capability are used for different purpose. Mini-UAVs are meant for carrying fewer payloads, like a thermal camera, single radar, camera, image sensor, *etc.* So, this is an open challenge to optimize the UAV placement to diminish energy feeding when the retrieved information is taking more time;
- **Network Scalability:** Single UAV system can perform limited tasks. To perform tasks quickly and efficiently, collection of UAVs is required. This motivated the concept of multi-UAV based system (Saleem et al., 2015). Sometimes if during the task few UAVs are out of the network due to power or some technical reason than it is required to add some new UAVs in the network to complete tasks efficiently. FANET algorithms should be designed in such a way so that they will be flexible to add any number of UAVs;
- **Reliable Data Delivery:** FANET applications transfer very important information in different applications, which required to be delivered in time bound manner. So the reliability of the network should be very high. Reliability should be defined with the criticality of data.

FUTURE RESEARCH DIRECTIONS

FANETs are getting more attention due to their capabilities in various situations. These networks have the additional advantages of the traditional adhoc networks because they use flying agents to collect and transmit the information in the operation areas. Due to high speed of UAVs, these networks are highly dynamic and have high topology changes. To support data transition with

Table 1. Common attacks of FANET

FANET Security Layer	Attacks
Multilayer Attacks	DOS, Impersonation, Replay, MIMA
Physical Layer	Intervention, Jamming, Snooping
Data Link Layer	Disruption MAC (802.11), bandwidth stealing, traffic monitoring, WEP weakness
Network layer	Black Hole, Link Spoofing, Worm-hole, Route tracking, Message fabrication, Byzantine
Transport layer	Session hijacking, SYN Flooding
Application layer	Repudiation, data corruption

dynamicity of these networks, an efficient routing scheme is required. So, developing appropriate routing protocols is one of the inspiring areas of research. Another explorative field of research in FANETs is the appropriate size of flying node, its power consumption, range covered by the node and speed of flying node. Along with communication and infrastructure needs, security of FANETs is also grabbing attention of researchers because FANETs carry sensitive and confidential information of critical application area. Security of flying nodes, trust management within nodes and secure routing are significantly important areas of research in FANETs.

CONCLUSION

The capabilities and applications of FANETs are very prominent in complex areas. UAVs coordinate and communicate with each other to perform complex task efficiently and faster where ground communication does not work. The swarm of UAVs communicating with each other without the use of pre-existing infrastructure is termed as flying adhoc network. Aim of FANETs is to provide fast, reliable and secure communication services in the various application areas. Nodes

in FANETs are highly mobile in nature and topology changes very frequent. Managing UAVs in FANET is a very challenging task.

In this article, the basic concept of FANET, various applications and challenges are explained. This article aims to motivate the research community to come up with various solutions to the challenges faced in FANETs.

REFERENCES

- Bekmezci, I., Sahingoz, O. K., & Temel, Ş. (2013). Flying ad-hoc networks (fanets): A survey. *Ad Hoc Networks*, 11(3), 1254–1270. doi:10.1016/j.adhoc.2012.12.004
- Lin, L., Sun, Q., Li, J., & Yang, F. (2012). A novel geographic position mobility oriented routing strategy for UAVs. *Journal of Computer Information Systems*, 8(2), 709–716.
- Muller, M. (2012). Flying Adhoc Network. In *Proceedings of the 4th Seminar on Research Trends in Media Informatics*. Institute of Media Informatics, Ulm University.
- Rosati, S., Kruzelecki, K., Heitz, G., Floreano, D., & Rimoldi, B. (2014). *Dynamic Routing for Flying Ad Hoc Networks*. Academic Press.
- Rosati, S., Kruzelecki, K., Traynard, L., & Rimoldi, B. (2013, December). Speed-aware routing for UAV ad-hoc networks. In *Globecom Workshops (GC Wkshps)*, 2013 IEEE (pp. 1367–1373). IEEE. doi:10.1109/GLOCOMW.2013.6825185
- Sahingoz, O. K. (2014). Networking models in flying Ad-hoc networks (FANETs): Concepts and challenges. *Journal of Intelligent & Robotic Systems*, 74(1-2), 513–527. doi:10.1007/s10846-013-9959-7
- Saleem, Y., Rehmani, M. H., & Zeadally, S. (2015). Integration of Cognitive Radio Technology with unmanned aerial vehicles: Issues, opportunities, and future research challenges. *Journal of Network and Computer Applications*, 50, 15–31. doi:10.1016/j.jnca.2014.12.002
- Singh, K., & Verma, A. K. (2014, May). Applying OLSR routing in FANETs. In *Advanced Communication Control and Computing Technologies (ICACCCT), 2014 International Conference on* (pp. 1212–1215). IEEE. doi:10.1109/ICACCCT.2014.7019290
- Yanmaz, E., Costanzo, C., Bettstetter, C., & Elmenreich, W. (2010). A discrete stochastic process for coverage analysis of autonomous UAV networks. In *GLOBECOM Workshops (GC Wkshps)*, 2010 (pp. 1777–1782). IEEE. doi:10.1109/GLOCOMW.2010.5700247
- Yap, W. S., Liu, J. K., Tan, S. Y., & Goi, B. M. (2015). *On the security of a lightweight authentication and encryption scheme for mobile ad hoc network*. Security and Communication Networks.

ADDITIONAL READING

- Frew, E. W., & Brown, T. X. (2009). Networking issues for small unmanned aircraft systems. *Journal of Intelligent & Robotic Systems*, 54(1-3), 21–37. doi:10.1007/s10846-008-9253-2
- Khare, V. R., Wang, F. Z., Wu, S., Deng, Y., & Thompson, C. (2008, September). Ad-hoc network of unmanned aerial vehicle swarms for search & destroy tasks. In *Intelligent Systems, 2008. IS'08. 4th International IEEE Conference* (Vol. 1, pp. 6–65). IEEE. doi:10.1109/IS.2008.4670440
- Sahingoz, O. K. (2013, May). Mobile networking with UAVs: opportunities and challenges. In *Unmanned Aircraft Systems (ICUAS), 2013 International Conference on* (pp. 933–941). IEEE. doi:10.1109/ICUAS.2013.6564779
- Sharma, V., & Kumar, R. (2015). A Cooperative Network Framework for Multi-UAV Guided Ground AdHoc Networks. *Journal of Intelligent & Robotic Systems*, 77(3-4), 629–652. doi:10.1007/s10846-014-0091-0
- Zhou, L., & Haas, Z. J. (1999). Securing ad hoc networks. *IEEE Network*, 13(6), 24–30. doi:10.1109/65.806983

Zhou, Y., Cheng, N., Lu, N., & Shen, X. S. (2015). Multi-UAV-Aided Networks: Aerial-Ground Cooperative Vehicular Networking Architecture. *Vehicular Technology Magazine, IEEE*, 10(4), 36–44. doi:10.1109/MVT.2015.2481560

KEY TERMS AND DEFINITIONS

ADHOC Networks: An adhoc network is a local network that is created spontaneously to perform a specific task. Instead of coordinating

with base station, nodes forward packets with each other.

FANET: FANET is a subclass of MANETs and made up of a swarm of small flying vehicles enable with camera, sensor and GPS system.

MANET: A mobile ad hoc network (MANET) is self-organizing, infrastructure less network of mobile nodes connected without any wired connections.

UAV: An unmanned aerial vehicle (UAV) is an aircraft without any onboard human pilot. UAVs can be controlled by either ground pilots or by pre-programmed software.

M

Health Wearables Turn to Fashion

Lambert Spaanenborg
Comoray AB, Sweden

INTRODUCTION

In the early years of manned space exploration, it is observed that astronauts easily lose a sense of time. Consequently they forget to eat in time and are not constantly fit to operate. From the need to monitor the physical condition of astronauts during travel, it became customary to equip their clothing with health sensors (Soller, 2002). One can see this as an early instance of health tele-monitoring.

These measures originate from the customary monitoring of test pilots to find the effect of extreme flight situations on the human body. But in space travel, the monitoring was not just performed during a couple of minutes but routinely for weeks or even months. This has been extended to other situations where a human pilot is confronted with extreme, potentially life-endangering situations, such as car racing. In all such situations, the clothing can be rigged with a wired network of a fixed-design sensory network and connected to a single control room (Jafari, 2006).

With the advance of microelectronics, a potential consumer market is opening. Mobile Health (m-health) is a technical challenge at the crossroads between personal fitness and hospitalization costs. Lack of exercise and/or bad nutrition undermines health and brings diseases such as heart failure and diabetes that can be hardly cured once become easily noticeable. Ultimately hospitalization cannot be avoided. Such stays will be regular and lengthy while surgery will be part of the service.

Tele monitoring is currently tested to reduce the length of hospital stays and the number of check-up visits by providing post-surgery health monitoring at home over remote links. On the other hand, a proper life style can delay the advent of diseases or even preclude their coming. The

current research interest is to collect data on large groups of people to identify risk groups. Then persons with a high risk can be identified and will receive pre-screening. Consequently, organizing special heart failure and diabetes clinics reduce hospital costs.

People do fitness in order to stay in shape. Usually it requires a regular doctor visit to ensure that the training does not damage the person's health. Self-quantification allows the person to play doctor himself, making the health check available constantly. There are little more health-related products in the common house than the thermometer, scales, cough drops and the medicine cabinet. Tele-monitoring requires additional 2,500 Euro worth of equipment to be installed in the house of the patient. This price tag cannot be tolerated for the average household, but the insurance may pick up the tab for the approved patient. Usually health-monitoring equipment is found in clinics and fitness clubs, where many people can take advantage. The modern equivalent is the 'kiosk': a dedicated booth in public places in and outside hospitals that can be visited at will and provides equipment in a direct link to the hospital infrastructure.

M-health is the next step to bring affordable healthcare in the home, and to any place where a person goes. It can be attached as simple gadgets or be integrated as plaster, garment or ornament. But it will only be more than a consumer gadget, when it (a) gets medical quality information and (b) communicates that information to the medical world, if desired and necessary. Further into the future, but not that far away, is the caring home, where healthy dwellers are measured to ensure that any health-endangering conditions can be detected and brought to the attention of a physician in time.

DOI: 10.4018/978-1-5225-2255-3.ch531

According to the World Health Organization (WHO) such precautions can reduce the amount of visits to the hospital by 90% (WHO, 2011)!

In the following, the history of wearable devices is outlined, coming from m-health to e-fashion. Then the state-of-the-art of oximetry technology is discussed, leading to an overview of trends in consumer health technology, leading to conclusions on their role in personal life-style improvement.

BACKGROUND

The m-health arena has started from a replacement market. Here Apps are administrative systems where the end-user is logging available data as provided from other sources. Examples are body temperature from a thermometer, calories from the food wrapper and body weight from bathroom scales. Such products do not provide more than a simple agenda for logging and many products share a lack of enthusiasm in the market.

The arena is growing in interest. Here electronic devices are integrated into common products, like bracelets, plasters and chains to provide single measurements on the body. A number of them still require additional handling. For instance, the usual glucose meter requires a blood sample to be taken and prepared for a special portable device. There is a tendency to become less invasive using intraperitoneal sensing (vanDijk, 2015), but non-invasive would be even better. A nice example is the gradual miniaturization of the pulse oximeter, which has gradually decreased in size while moving from the hospital to the clinic, and which will likely become as common as the thermometer in the very near future.

The forerunner in the pursuit of personal health is the fitness market. Apps like RunKeeper and EndoMondo have received a lot of attention for the creation of social health networks. Gradually they will make inroads to electronically capture health data. The widely accepted Cooper test is based on the correlation between the VO₂max and

the distance covered in a 12 minute run (Cooper, 1969). Hence it indicates fitness by the ability of the body to use oxygen while running. This can nowadays be monitored with a strapped-on oximeter. Another requirement for fitness training is heart rate or ECG monitoring. Here we find an exponentially increasing amount of wearables with more or less success. Here continuous logging of an oximeter is an attractive alternative.

The typical health service needs first of all to provide a regular check like yearly done by the family doctor, i.e. looking at health in general. This can be seen as the Mc.Luhan effect on health provision, where instead of having the doctor come for a yearly visit, ICT will allow a regular check-up which can be viewed by the doctor at any time and any place. In other words, the players in the health service game are swapping places. The every Tom, Dick and Harry are not really concerned with health till they get sick. They do not want to be treated as patients and are therefore not susceptible to public information provision. When health gets their interest, it is usually through a desire to stay forever young. Alternatively, when they have been to the hospital, the main interest is to get back into shape.

Beyond the Thermometer

The exemplary device for electronification of health equipment is the pulse oximeter, originally created to support the anesthetist in his work. This device measures the degree of perfusion and thereby certifies the distribution of anesthetics. Then Aoyagi notes that a longer observation shows also the pulsating period of the blood pulse. This opens the door for non-invasive measuring more health parameters _ like cardiac functions showing atrial fibrillation and stress _ for everyday use through the single photoplethysmographic (PPG) signal. Wieringa (2007) re-calls the evolution of the thermometer into a common household product and then foretells that the modern oximeter will revolutionize personal healthcare in a similar way.

The early pulse oximeter is based on the comparison of reflected light in the infrared band (showing oxydized hemoglobine) and the red band (showing de-oxydized hemoglobine). As shown by Mendelson from the Worchester Institute of Technology over the last 2 decades, the same results can be found by looking at reflections in the green and the red band (Scully, 2012). Especially the green band is interesting, as it seems less sensitive to muscle movements in the body. The principle of operation transforms the intensities of the reflected light into a time series on which filtering and Fourier transform brings the energy-frequency map where the characteristic cardiac parameters can be found. It requires top-notch filter design to boost the signal enough while reducing the noise level.

The use of reflective light to achieve pulse oximeter functionality is largely caused by the light transport through the skin tissue, where the effective intensity signal is reduced by a factor 50 through scattering and dispersion. As Mendelson has demonstrated in a number of careful experiments, the smartphone image provides enough information to measure cardiac values that are less than 5% different from hospital lab equipment. But further tuning was required to allow for a real-time smartphone implementation. Current offers in the market (the heart rate, sleep and stress monitors) aim for real-time while compromising accuracy and functionality. The next generation will support nutrition, life style and living environment in their influence on the bodily condition from a single smartphone: Like a modern Internet connected thermometer!

Wearable

The ideal is non-invasive measurements and of course it has taken a long time to reach that ideal. The critical aspect of its breakthrough is the movement of the phone while measuring. Especially when forcibly trying to keep it still, every hand-held device will be subject to a degree of jitter. When the exposure time for measurement

exceeds a couple of seconds, other spastic movements will come into play. The mixed influence of interfaces between (largely watery) transmission media and sensor movement sets the lower limit to the achievable precision and accuracy. This is not new by nature. Single-point measurements are generally acknowledged as being impossible to stabilize. The commonly accepted solution is the multi-point measurement. The medical ECG measurement takes a minimum of 8 probes to achieve a reliable result, i.e. it can detect handling errors and comes with a useful error margin.

In search for the successor of the smart phone, the smart watch is on the horizon. Typically they are to be worn on the body, and are in competition with similarly wearable gadgets such as the earring or the tracking shoe. The early smart watch has limited health information, mostly limited to heart rate, and more fitness data collection capability. It has a large display to allow entering and reading data and a large size to optimize energy harvesting to increase battery lifetime.

The alternative of reflective vision technology uses the accelerator. Where movement distorts the PPG signal, the accelerator measures the motion directly and must be filtered to find the heart rate (Hernandez, 2015). The BioPhone claims a similar accuracy, even when the smart phone is handled loosely. As the outcome of the measurements depends on the body part used for sourcing the signal, they can also be used in combination with other accessories into a local health network.

The Body Area Network

Having garment equipped with a sensory system provides the basis for multiple probes. The hospital equipment uses these strobes to collect data centrally and continuously. By subsequent signal analysis, time correlation identifies the important features. In the garment, this spaghetti wiring is organised and woven into the fabric (Jafari, 2006). Moreover, the probes are intelligent by itself making for a distribution of functionality over a medical embedded network.

The Body-Area-Network (BAN) is not a new concept (Yang, 2006). In the usual sense, the sensors are connected with a director, a gateway that collects and analyzes data. The smart phone can take this role as director. At least it may bring the data from the typical short distance of the BAN to the large distance of the wireless telephone network. The added advantage is that all computationally, energetically and spatially intensive tasks can be performed on the phone, releasing such demands from the network.

A typical BAN example is the Clothing+ sportswear. It provides a number of sensors in a sport suit, which can be collected over a close-by mobile product, such as an E-Pad. It was invented to help polar adventurers to survive the extreme conditions under deep-winter conditions, and can easily be extended to the E-sports market where the amateur athlete's life may already be threatened under less extreme weather conditions. For instance, T.Ware adds support for tracking suits in general, but especially for people where the added sensors give a clear added experience, such as for autistic children.

MAIN FOCUS OF THE ARTICLE

The PPG technique was originally published by Hertzman in 1938 and became popular around 2000 for perfusion measurements to aid anesthetic monitoring. (Forstner (1988) gives a good breakdown of the time in between). The latest developments have confirmed that PPG brings accurate non-invasive acquisition of major physiological parameters. Where PPG signals are extracted from the absorption of light in blood, devices come in two different flavors. On thin body parts, such as the tongue or earlobe, light will be transmitted from one side and sensed on the other. When the body part is too thick, light source and sensor are at the same side. The latter, light reflective device is the most popular for reason of the ease of construction.

The earliest detection of decreased blood oxygenation is by the skin color shift towards blue (also named cyanosis). Together with a weakening pulse, this was known (especially to mountain climbers) as a life threatening condition that requires immediate attention. Hard to perform manually over a longer period, a number of opto-electronic devices have been attempted. The major breakthrough is the observation by Aoyagi in 1974 that by the pulsatile nature of blood flow through the arteries the contribution of arterial absorbance can be distinguished from tissue absorbance components. From there, pulse oximetry rapidly spread as a simple, bedside technique to monitor heart rate and arterial oxygen saturation.

Wieringa (2005) remarked that pulse oximetry has changed many clinical routines and made the analogy with the influence of thermometry three centuries earlier. From the historical development of thermal cameras in extension of the conventional thermometer, he claims room for remote sensing in photoplethysmography over the coming years. Literature shows that he may have been right.

Measuring by Reflective Light

The principle of the oximeter is based on the pulsatile nature of blood flow. The high frequency reaction of the light on the capillary blood shows as color, while the intensity changes in value with the flow. The flow of blood is then found as variations in the observed color and intensity. The quantification of the blob is done by arbitrarily averaging an arbitrary number of pixels. So we have two sources of information: the intensity and the color as energy and frequency in the spectrum. This seems to imply the use of Fast Fourier Transform (FFT) to determine the location of the pulses in the energy-frequency plot. The visual spectrum lies between 200 and 600 nm, i.e. in the TeraHertz region. The physiological parameters of direct interest are at 1 Hz (heart rate with the 1st harmonic at 2 Hz) and the respiration rate between 0.1 and 0.16 Hz (with harmonics at 0.2, 0.3 and 0.4 Hz).

These frequencies are relatively easy to determine, but are not so easily distinguished because they change considerably under the exercise regime. (When the heart rate goes up, also the respiration rate increases and comes around the resting values of the heart rate. This seems to preclude the use of fixed-band filters (Nakajima, 1996)). The classical oximeter approach is based on the Beer-Lambert law on absorption, which states that the absorption at a given wavelength is the product of the absorbency, the concentration and the path length. How this can be interpreted has been regularly subject of discussion. It seems that the absolute value needs calibration for every new person. Also the digital signal processing can cause 5 – 10% numerical accuracy issues when done improperly.

The clinical approach for raising the accuracy for the computation is the ‘ratio of ratios’. For instance, arterial oxygenation can be found from the ratio between peak-to-peak amplitudes at a minimal of two different wavelengths. Using a white LED as illumination source, the popular color for the PPG measurement is green. Jonathan (2010) claims that blue and red are also usable, but the signal strength is highly position specific. However, it usually pays to see the emergence of low blood oxygen levels (hypoxemia) ahead of time. For this, it is useful to look at the red and infrared bands. Scully (2012) uses also green for the heart rate, but the SpO₂ is computed from the red and the blue where the blue band is the reference for the infrared wavelength in the traditional pulse oximeter.

Pulse Oximetry

The classical oximeter means to determine the arterial oxygen saturation (SaO₂) by measuring differences in the visible and near infrared absorbance effects of fully oxygenated and deoxygenated arterial blood. They have largely replaced clinical blood gas analyzers that can only take intermittent values from blood samples. It provided a clear

innovation, freeing access to the machine from the operating room / critical care unit to enter any ward in the hospital.

The early oximeter is non-pulsatile using multiple wavelengths, where one (805 nm in the near infrared) is used as a reference to eliminate the effect of tissue absorption and skin pigmentation, while the other (650 nm in the blue band) determines the value. The earlobe was a popular location as it is easily accessible, while being painless and bloodless in the reference mode (Huei, 2011). The pulse oximeter exploits the pulsed nature of the photoplethysmogram to bring the wavelength needs to a single one.

Mendelson (1992) gives such a historical background together with an overview of practical problems, as felt in 1990. This overview lists also nicely what can be potentially achieved by technological improvements. Typically it mentions the measurement of carbon monoxide, as neglecting this effect leads to an overestimation of SaO₂ in smokers, lung patients or (more general) people with a limited supply of oxygen. It has been attempted to exploit the PPG technology through non-contact measures, say, through a camera in near infrared. As discussed in (Wu, 2003), it provides clear vision on venous issue, supporting the need to determine perfusion in a general sense. The paper mentions vein insufficiency and thrombosis. However, the strong disturbance of tissue reflection brings the arterial information to less than 2% of the total amplitude.

Over the past years, there is an increasing interest in non-invasive PPG measurement using commercial mobile devices, such as (Web) cameras and mobile telephones. The early literature claims the need for HD devices; the later assumes this to be always available. This shows a move away from the original anesthetic measurement (to provide feedback on the quality of the blood flow) to the cardiac monitor on a consumer product. Or, in terms of technology, it exemplifies a move from understanding the PPG waveform to implementing digital signal processing. In that sense Shelley is

still waiting for an answer (Shelley, 2007), but we can applaud the systematic research by Mendelson that opened many doors for further developments.

FUTURE RESEARCH DIRECTIONS

Along the historical development of the smart phone, it is reasonable to assume that the first steps in m-health will be to serve geo- and socio-positioning. Geo-positioning helps persons to be found; a function to position an object will probably require a keyboard and therefore a phone. Socio-positioning will identify persons of the same social community, for instance having the same original brand of clothing. Because of a lack of active functionality, this all appears like a gadget with little added value, especially compared to the smart phone.

The above distinction becomes more meaningful in sensory networks. A simple measurement will not only vary in time, but also in geo-location. Moreover, the measurable value will vary. For instance, blood pressure depends on the location on the body and on the position of the body. Hence, conclusive values do not result from averaging a series but require an intelligent model. A typical configuration appears in fashion where multiple sensors for redundant health factors are found in free network composition. Problems are even more complex with reference to socio-positioning.

The Future in Fashion

Fashion is clothing with the artful expression of individuality. It can be coarsely subdivided into couture, commercial and catalogue while stating the cost of acquisition, or into factory, office, leisure, athletic (and so on) while stating the current place and activity. The latter is interesting as m-health devices can measure either the body or the environment in its influence to the body. The former is a balance between luxury and commodity. Though m-health is targeted for the mass-market, Apple (with the i-watch) and Misfit

(Fossil since 2015) show that exclusive DESIGN is still part of reality.

Fashion is typically a physical interface. It guards the body against dust and cold. Wearing a heavy coat is not ideal to keep the sun out, while a raincoat will not give enough protection against the freezing rain that lashes the pedestrian. The importance of embedded sensors will be to activate the interface, i.e. to adapt the nature of the clothing to the potential danger coming from the outside. An early example is the 'climate dress' from Diffus. Similarly, the 'Aegis Parka' from Nieuwe Heren, a design bureau in Utrecht (The Netherlands), measures the common urban pollutants (NH₃, NO_x, alcohol, benzene, smoke and CO₂) and shows the relative threat by the illumination of an LED display.

The degree of an environmental threat is given by the apparent lack of protection. To determine that requires body measurements also. For instance, when the atmosphere contains a high degree of CO₂ and the blood of the person who wears this cloth has a high degree of CO₂ then there is a good reason to assume that the health of the person is in danger. When the body also shows a lack of balance, there is a good reason for alarms. An extensive discussion of the current offering and listing of active players is given in (Florea, 2012).

The Future in Intelligence

Where the body is a complex physical system, there are many ways to measure and model it. The preferred way for measurement is non-invasive, as many do not enjoy forced bodily access. This needs medical processes to be available outside the body, like through the skin, via a cavity or by fluids leaving the body. The use of substances naturally leaving the body, like hairs or urine, works similarly to where samples are taken out of the body. The modern alternative is the subdermal sensor implant, where likewise substances need to be delivered to the body. A suitable short-distance wireless connection

A mobile device brings application issues. The easy solution is to create a detachable add-on, such as a thimble, that can be ideally fixed at the right location. The alternative is an attachable product, like a watch or an earring, but this will usually require an additional phone to connect to the outside world. The advantage of the specialized part is that it is designed for a specific parameter, while the disadvantage is that it cannot be programmed for other parameters. This extends naturally to the “Internet of Things”.

In healthcare, security and dependability are required for all thinkable reasons. Hence the demands on the models seem almost crushing. However, functional redundancy alleviates (if not even removes) that burden. The key to all this is ‘deep learning’. A neural network extracts information by correlation; deep learning adds modularity and similar methods of structural composition that avoid the problems of unstructured, large networks. The immense variability of human bodies makes it hard to prepare the learning set.

CONCLUSION

The smart-watch is entering the health market with a fitness proposition, while the smart-phone is keeping strong for vital health. More specialized medical applications are looking for room in the gadget market. Such things become increasingly complex and fashionable.

A typical sexy device is the Vital Signs Camera, launched in early 2012 by Philips as app on an iPad. It measures heart rate from skin coloring and breathing from chest movements by the tablet’s front-facing camera. More experimental is the laser-based blood flow management, where model-based matching helps to achieve medical accuracy in single-probe ECG measurement. As pointed out by Spaanenburg (2010), this opens the door to ad-hoc sensor clouds where health sensors from the environment and those on the body co-operate to bring guidance for life style. Such developments enable personal health monitoring as an Internet of Wearables.

REFERENCES

- Cooper, K.H. (1969). *Aerobics*. Bantam Books.
- Dept, W. H. O. of Health Statistics. (2011). World Health Statistics 2011. Author.
- Florea, A. C. (2012). *Techno Fashion: A Cultural Revolution* (Master Thesis). University of Aarhus, Denmark.
- Forstner, K. (1988). Pulsoximetrie: Stand und Entwicklung der Technik. *Biomedizinische Technik*, 33 + 3, 6 – 9.
- Gregovski, M.J. (2012). Development and validation of a smartphone heart rate acquisition application for health promotion and wellness telehealth applications. *Int. Journal of Telemedicine and Applications*.
- Hernandez, J., McDuff, D. J., & Picard, R. W. (2015). BioPhone: Physiology Monitoring from Peripheral Smartphone Motions. *Proceedings 37th Int. IEEE Engineering in Medicine and Biology Society Conference*. doi:10.1109/EMBC.2015.7320048
- Huei, Y. J., Hwang, G.-H., & Jeong, D.-U. (2011). Designing a wearable earlobe hook-type photoplethysmography sensor for ubiquitous healthcare monitoring application. *Proceedings IEEE 6th Int. Conference on Computer Sciences and Convergence Information Technology IC-CIT*, 47 – 50.
- Jafari, R. (2006). *Medical Embedded Systems* (Ph.D. thesis). UCLA, Los Angeles, CA.
- Jonathan, E., & Leahy, M. (2010). Investigating a smartphone imaging unit for photoplethysmography. *Physiological Measurement*, 31(11), 79–83. doi:10.1088/0967-3334/31/11/N01 PMID:20871134
- Mendelson, Y. (1992). Pulse Oximetry: Theory and Applications for noninvasive monitoring. *Clinical Chemistry*, 38(9), 1601–1607. PMID:1525987

Nakajima, K., Tamura, T., & Miike, H. (1996). Monitoring of heart and respiratory rates by photoplethysmography using a digital filtering technique. *Medical Engineering & Physics*, 18(5), 365–372. doi:10.1016/1350-4533(95)00066-6 PMID:8818134

Poh, M.-Z., McDuff, D. J., & Picard, R. W. (2010). Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Optics Express*, 18(10).

Rubins, U., Miscuks, A., Rubenis, O., Ert, R., & Grabovskis, A. (2010). The analysis of blood flow changes under local anesthetic input using non-contact technique. *Proceedings 3rd Int. Conference on Biomedical Engineering and Informatics BMEI*, 601 – 604. doi:10.1109/BMEI.2010.5640023

Scully, C. G., Jinseok Lee, J., Meyer, J., Gorbach, A. M., Granquist-Fraser, D., Mendelson, Y., & Chon, K. H. (2012). Physiological Parameter Monitoring from Optical Recordings With a Mobile Phone. *IEEE Transactions on Bio-Medical Engineering*, 59(2), 303–306. doi:10.1109/TBME.2011.2163157 PMID:21803676

Shelley, K. H. (2007). Photoplethysmography: Beyond the calculation of arterial oxygen saturation and heart rate. *Anesthesia and Analgesia*, 105(6), S31–S36. doi:10.1213/01.ane.0000269512.82836.c9 PMID:18048895

Soller, B. R., Cabrera, M., Smith, S. M., & Sutton, J. P. (2002). Smart medical systems with application to nutrition and fitness in space. *Nutrition (Burbank, Los Angeles County, Calif.)*, 18(10), 930–936. doi:10.1016/S0899-9007(02)00897-3 PMID:12361790

Spaenenburg, L., & Spaenenburg, H. A. E. (2010). *Cloud Connectivity and Embedded Sensory Systems*. Boston, MA: Kluwer Press.

vanDijk, P. R. (2015). *Continuous intraperitoneal insulin infusion in the treatment of type 1 diabetes mellitus* (Ph.D. thesis). University of Groningen, The Netherlands.

Wieringa, F. P. (2007). *Pulsoxygrafie* (Ph.D. thesis). Erasmus University, Rotterdam, The Netherlands.

Wieringa, F. P., Mastik, F., & Steen, A. F. W. (2005). Contactless multiple wavelength photoplethysmographic imaging: A first step towards SpO2 camera technology. *Annals of Biomedical Engineering*, 33(8), 1034–1041. doi:10.1007/s10439-005-5763-2 PMID:16133912

Wu, T. (2003). PPGI: New Developments in noninvasive and contactless diagnosis of dermal perfusion using near infrared light. *Journal of the GCPD*, 7(1), 17–24.

Yang, G. Z. (Ed.). (2006). *Body sensor networks*. London: Springer Verlag. doi:10.1007/1-84628-484-8

ADDITIONAL READING

Buske, O., Neils, C., & Regnier, M. (2009). *HEARTBEAT: Design and Development of a headphone-mounted infrared heart-rate monitor*. Seattle, USA: University of Washington.

Dickson, C. (2012). *Heart Rate Artifact Suppression*. M.Sc. Thesis, Grand Valley State University, Michigan (USA).

Fox, S., & Duggan, M. (2013). *Tracking for Health*. Report Pew Research Center, Washington D.C. Hue, Y.J., Hwang, G.-H., and Jeong, D.-U. (2011). Designing a wearable earlobe hook-type photoplethysmography sensor for ubiquitous healthcare monitoring application. *Proceedings 6th Int. Conference on Computer Sciences and Convergence Information Technology*, 47-50.

Jo, J., Lee, Y.K., and Shin, H.S. (2013). Real-time analysis of heart rate variability for a mobile human emotion recognition system. *Recent Advances in Electrical and Computer Engineering*, 162-166.

Jung, J., Ha, K., Lee, J., Kim, Y., & Kim, D. (2008). Wireless Body Area Network in a Ubiquitous Healthcare System for Physiological Signal Monitoring and Health Consulting International. *Journal of Signal Processing, Image Processing and Pattern Recognition*, 1(1), 47–54.

Lee, Y. K., Jo, J., & Shin, H. S. (2013). Development and evaluation of a wristwatch-type photoplethysmography array sensor module. *IEEE Sensors Journal*, 13(5), 1459–1463. doi:10.1109/JSEN.2012.2235424

Poh, M.-Z. (2011). *Continuous Assessment of Epileptic Seizures with Wrist-worn Biosensors*. Ph.D. thesis, MIT, Cambridge (USA).

Poh, M. Z., Swenson, N. C., & Picard, R. W. (2009). Comfortable Sensor Wristband for Ambulatory Assessment of Electrodermal Activity. *1st Biennial Conference of the Society for Ambulatory Assessment*, Greifswald, Germany.

Poh, M.-Z., Swenson, N. C., & Picard, R. W. (2010). Motion tolerant magnetic earring sensor and wireless earpiece for wearable photoplethysmography. *IEEE Transactions on Information Technology in Biomedicine*, 14(3), 786–794. doi:10.1109/TITB.2010.2042607 PMID:20172836

Poh, M.-Z., Swenson, N. C., & Picard, R. W. (2010). A Wearable Sensor for Unobtrusive, Long-Term Assessment of Electrodermal Activity. *IEEE Transactions on Bio-Medical Engineering*, 57(5), 1243–1252. doi:10.1109/TBME.2009.2038487 PMID:20172811

Poon, C. C. Y., Liu, Q., Gao, H., Lin, W.-H., & Zhang, Y.-T. (2013). Wearable Intelligent Systems for E-Health. *Journal of Computing Science and Engineering*, 5(3), 246–256. doi:10.5626/JCSE.2011.5.3.246

Poon, C. C. Y., Liu, Q., Gao, H., Lin, W.-H., & Zhang, Y.-T. (2013). Wearable Intelligent Systems for E-Health. *Journal of Computing Science and Engineering*, 5(3), 246–256. doi:10.5626/JCSE.2011.5.3.246

Priyanka Pankaj Shah. (2008). *Photoplethysmograph-derived respiration*. New-Jersey, USA: New-Jersey Institute of Technology.

Renevey, P. et aleni (2001). Wrist-located pulse detection using IR signals, activity and nonlinear artefact cancellation. *Proceedings 23rd Int. Conference on Engineering in Medicine and Biology*, 3030.

Subramanya, K., Kamath, S., & Prasad, H. (2013). Development of a wearable device for monitoring electrodermal activity and predicting changes in different blood pressure indexes. *Proceedings 35th Int. IEEE Engineering in Medicine and Biology Society Conference*.

KEY TERMS AND DEFINITIONS

Body-Area-Network: A wireless network of wearable computing devices.

EHealth: The use of electronic processes and communication for healthcare practice.

mHealth: An abbreviation for mobile health, the use of mobile devices for medicine and public health.

Perfusion: The body process delivering capillary blood in its biological tissue.

Photoplethysmography: The use of a pulse oximeter to illuminates the skin and measure changes in light absorption.

PPG: Signal is the photoplethysmographical time-series that is extracted by an oximeter on basis of light reflection on the skin.

Pulse Oximetry: A non-invasive method for measuring a patient's O₂ saturation.

Quantified Self: A movement to incorporate data acquisition technology for health in aspects of a person's daily life.

Reflective Light: Visual light that is sensed from the reflection on the skin.

Smart Phone: A mobile phone with more advanced capability than basic feature phones, such as a vision sensor.

Smart Watch: A wristwatch that at least monitors cardiac effects.

Spectroscopy: Delivers the absorption frequency to indicate the presence and amount of matter.

Wearable Technology: Comprises clothing and accessories incorporating computer and advanced electronic technologies. Meant for the consumer market, it has a clear aesthetic, say fashionable, aspect.

M

Human Psychomotor Performance Under the Exposure to Mobile Phones–Like Electromagnetic Fields

Giuseppe Curcio

University of L'Aquila, Italy

INTRODUCTION

At the present time the great majority of human beings on the planet are directly or indirectly exposed to the electromagnetic fields (EMFs) emitted by mobile phones, base stations, as well as other types of wireless communication technologies. Moreover, to date for the first time ever it can be affirmed that there are more mobile devices in the world than living people: if we look at GSMA's real-time tracker (June 2016), the number of mobile devices is assessed to be more than 7.75 billion, while the total world population is estimated to be around 7.3 billion (US Census Bureau, June 2016). Such repentine increase in number of mobile phones has induced a growing interest toward possible health effects in human beings, bringing also the World Health Organization (WHO) to open a health topic on this issue.

Since a discrete amount of radiofrequency (RF) electromagnetic fields (EMFs) emitted by mobile phones is transmitted through the skull and can reach the brain, it is possible to hypothesize a physiological influence of these low level RF EMFs on human cerebral activity, and, consequently, a potential influence on cognitive and behavioural performance. In the last 20 years a number of studies have assessed several aspects of human performance: vigilance and attention (divided, selective, and focused), perception, learning, short-term and working memory, spatial and verbal recognition, decision making. All these studies have been repeatedly reviewed (e.g. Kwon and Hämäläinen, 2011) and metaanalyzed (e.g., Valentini et al., 2010). And the scientific interest

DOI: 10.4018/978-1-5225-2255-3.ch532

still continues, since that only in the last 5 years (2011-2015) several new papers have been published (Ofstedal et al., 2016). Thus, the present entry will focus on volunteer studies (i.e. experimental studies with volunteer human individuals) and is intended as an updating of a previous critical review of literature (Valentini and Curcio, 2015). Most experimental studies with RF-EMFs exposure have been conducted as laboratory studies. Within each laboratory study, the entry will report only cognitive and psychomotor effects of mobile phone-like EMFs. In different studies, each of these effects have been tested through different tests and tasks, both administered by a computer and in a paper-and-pencil version. Usually in this kind of literature measures of speed (i.e., the time needed to accomplish the requested activity) or accuracy (i.e., the number of correct responses to the task or, conversely, the number of errors or absence of response to a task) have been considered as dependent variables.

BACKGROUND

Most of the included studies entail mobile phone-like signals or base station-like signals as well as other types of EM signals. This review will focus on experimental provocation studies with human volunteers, most of them being carried out as laboratory studies. Only laboratory studies focused on cognitive and psychomotor effects of mobile phone-like EMFs emissions will be taken into consideration. Here, only studies published in the last 20 years and focusing on mobile phone-like

emissions will be considered as relevant. To this respect, we will also provide a qualitative overview of the most recent studies published up to 2015.

Pioneer attempt to study human psychomotor performance was undertaken by Koivisto et al. (2000) and Krause et al. (2000) at the University of Turku (Finland). Most current relevant contributions originate from several scholars distributed across different continents. Particular methodological improvements have been proposed by Curcio et al. (2004, 2008) at the Sapienza University of Rome (Italy) and by Regel et al. (2007 a,b) at the University of Zurich (Switzerland), while enlarged sample sizes have been recently studied by Keetley et al. (2006) and Hamblin et al. (2006) at Swinburne University (Australia). In addition, important attempts of replications were performed by Russo et al. (2006) at University of Essex (UK), and by Haaraala et al. (2007) at the University of Turku (Finland).

As said in a previous review (Valentini and Curcio, 2015) also in this case it is due that several studies reported here were aimed at investigating also brain neurophysiology (with Electroencephalography-EEG, Event-Related Potentials-ERPs, Event-Related Synchronization/Desynchronization-ERS/ERD, Transcranial Magnetic Stimulation-TMS, Magnetoencephalography-MEG) or metabolism (by means of Positron Emission Tomography-PET, functional Magnetic Resonance Imaging-fMRI, functional Near Infrared Spectroscopy-fNIRS); however, only concomitant cognitive and attentional measures will be presented and examined. In particular, we will discuss specifically those outcomes related to speed of motor responses (time needed to complete the trial). This choice is substantiated by the fact that most of studies reporting an effect of RF EMFs observed a significant reduction of reaction times during or after the active exposure, while only a few of them showed effects on accuracy, effects that disappeared when highly conservative post hoc statistical tests were applied. Moreover, the choice is also supported by the hypothesis that RF EMFs may act as modulators of cortical excit-

ability (e.g., Ferreri et al., 2006), which in turn would influence the most sensitive psychomotor function, i.e. motor reaction times to external sensory stimuli.

SCIENTIFIC KNOWLEDGE ABOUT MOBILE PHONE INFLUENCE ON HUMAN PSYCHOMOTOR PERFORMANCE

It is known that volunteer studies are experimental provocation studies with human volunteers; with respect to observational studies, they allow to accurately control some exposure factors that may influence the outcome (as, for example, the exposure characteristics). However, all experimental human studies aimed at exploring transient and relatively mild effects of MP exposure ever respect ethical aspects using exposures that do not exceed the maximal permissible levels defined by ICNIRP (1998) and currently under continuous revision. Considered these limitations, most experimental studies with RF exposure were laboratory studies, but some observational studies also do exist: in these limited cases exposures were administered within a real life situation (such as at work or home). Control of exposure and other potential confounding factors are usually a greater challenge in observational studies than in laboratory provocation studies.

Here, it is provided an updated outlook of laboratory research (up to year 2015). Literature selection criteria and studies endpoints are based on two recent scientific contributions (Valentini et al., 2010; Valentini and Curcio, 2015). In particular, when more than one sensory modality was tested in a single study, we selected only results from the auditory modality (i.e., auditory reaction times), thus concentrating on the effects of mobile phone-like EMFs on the auditory system. This criterion is justified by the rationale that the peri-auricular temporal area is the most exposed to the thermal effects of radiations (see for example Curcio et al., 2004), as also documented

by dosimetric studies (e.g., Cardis et al., 2008). Most of studies on human cognitive performance can be included in three categories, two of which related to the study of attention (focused attention and speed of processing, divided and sustained attention) and a third associated to the study of memory.

Finally, in the respect of methodological standard, to be assessed relevant and included in this review, studies had to mandatorily satisfy some qualitative criteria, such as: blinding of the study (i.e., at least a single-blind procedure); task trials/ conditions randomisation and/or counterbalancing of exposure sessions; detailed exposure characterization (e.g., main frequency and sub-frequencies of the EMFs, signal modulations, power density, SAR estimation).

Studies Dealing with Attention and Speed of Processing in Adults and Adolescents

In a single-blind, counterbalanced, cross-over design, Koivisto et al. (2000) investigated the effects of mobile phone exposure on response times to twelve different tasks. Forty-eight participants were exposed to both GSM and sham signal for 60 minutes, with the two sessions separated by 24 hours. The results indicated a significant reduction (9 ms) of response time in the simple reaction times task.

Although the RF signal and experimental protocol were akin to Koivisto et al.'s study, Haarala et al. (2003 a, b) extended and methodologically improved the previous experiment by using a double blind design, larger sample size, multicentre testing and some additional tasks. Sixty four participants were enrolled (32 in Finland and 32 in Sweden) and then exposed to a GSM and to a sham signal for about 65 minutes. In this case no significant effects were reported on speed measures in the simple reaction time task.

Curcio et al. (2004) investigated the time-course of RF-induced effects on cognitive performance at different tasks. Twenty volunteers

were randomly assigned to one of two groups receiving the exposure before or during the testing session; half of the participants in each group were exposed to a GSM and half to a sham signal for 45 minutes. Results indicated an improvement of response speed in real exposure (decrease of 47 ms in the simple-reaction time task), and a faster performance of participants exposed prior to the test than of those exposed during the test itself (with a gain of 85 ms). This latter result indicates that a minimal exposure time would be required to induce cognitive effects.

A subsequent study was aimed at investigating the effects of MP exposure on cognitive performance in 32 children (Haarala et al., 2005). Tasks were selected for assessing attention, vigilance (simple-, two- and choice- reaction time tasks, vigilance task), and memory (short-term memory task with four memory load conditions) under GSM or sham exposure. No statistically significant differences between conditions were observed on speed outcomes of cognitive functions.

Hamblin et al. (2006) performed an extension of a previous study (Hamblin et al., 2004, see below) trying to overcome some methodological limitations (small sample size, limited endpoints, multiple statistical comparisons). Thus, 120 participants were exposed to a GSM-like signal, in two separate sessions (sham/sham and a sham/active session). Half of the participants received exposure to the left temporal region, and half to the right one. In both sessions they were asked to attend an odd-ball task. Also in this case, no significant differences were reported on reaction times in the simple reaction time task as a consequence of the exposure to mobile phone signal.

In a singular effort to emulate a real-life exposure, Besset et al. (2005) planned a complex and long protocol of exposure. Fifty-five volunteers participated in a double blind study lasting 45 days (3 of baseline, 28 of exposure period, 14 of recovery): during this period they were exposed 2 hours per day, 5 days per week to a EMFs-on or EMFs-off signal. Four cognitive assessments were carried out: at baseline (day 2), during the

exposure period (day 17 and 32) and in the recovery phase (day 45). Despite the study was designed to reveal medium-term size effects, no impact of exposure on psychomotor performance was found.

Keetley et al. (2006) aimed at investigating the effect of exposure to GSM signal on some different cognitive tasks administered to a large sample of 120 volunteers, controlled for age, education, gender. They detected an impairment of performance in a simple reaction time task (digit span) as a consequence of exposure to GSM-like signal.

Similar to other studies which attempted to control for methodological limitations (small sample size, single blind design, type of exposure signal) of previously published studies, Russo et al. (2006) enrolled 168 volunteers to investigate the effect on attention of 35-40 min exposure to a pulsed (GSM-like), CW, and sham signal. The exposure was directed either to the right or to the left temporal region. Despite methodological improvements (double-blind, counterbalancing of conditions, cross-over design), no effect of exposure on measures of attention and speed processing was reported.

Independently by Russo and co-workers, Haarala et al. (2007) conducted a study with very similar set-up and methodology. In this case, 36 volunteers were exposed to three different conditions: pulsed (GSM), CW, and sham signal. Again, exposures were localized on both the left (45 min) and right (45 min) hemispheres. As in the Russo et al.'s study, they reported no effect of exposure to real signal on response times in a simple reaction time task.

In a study by Regel et al. (2007 a) participants were exposed to three different conditions: pulsed GSM, CW, and sham signal. The two RF signals used the same frequency (900 MHz) and specific absorption rate (SAR). The EMFs irradiation lasted for 30 min and was directed over the left ear. Results indicated no significant effect on response speed in a simple-reaction time task. The same group (Regel et al., 2007 b) aimed at investigating possible dose dependent effects of GSM signals on attention and memory tasks. Fif-

teen participants were exposed over their left ear for 30 min. Also in this case, the authors reported no significant difference between conditions in the simple reaction time task.

Curcio et al. (2008) attempted to test the possible cumulative effects of brief (15 min) and repeated (3 times) exposures in a single daily session. Using an exposure setting identical to the previous work (Curcio et al, 2004), 24 volunteers were exposed to both GSM and sham signal: immediately after, they were asked to complete two psychomotor tasks. No statistically significant difference arose as a consequence of exposure to the GSM radiation.

As an extension of Regel's studies, Schmid et al. (2012) exposed 30 participants to two differently modulated GSM signals (14 and 217 Hz, respectively) and to a sham condition: the exposure was carried out before sleep and lasted 30 min. Attention and working memory performance were assessed during the exposure. Results indicated no statistical evidence of an influence of EMFs signals on the attentional performance.

With respect to studies on adolescents, recently Loughran et al. (2013) carried out one of the first studies on young people. In this case the research aimed at investigating the potential sensitivity of teens to mobile phone-like RF EMFs irradiation. Participants underwent three experimental sessions in which they were exposed to mobile phone-like RF EMFs at two different intensities, and to a sham session. Cognitive tasks were performed during the exposure period. The authors reported no effect of exposure on performance in a simple reaction time task.

Studies Dealing with Divided and Sustained Attention in Adults and Adolescents

The large spectra of tasks used by Koivisto et al. (2000) included a subtraction and a vigilance task. They exposed 48 participants to GSM and to sham signal for 60 minutes: the two sessions were separated by 24 hours. Results indicated

a significant reduction of response time in the vigilance task (reduction of 25 ms).

Croft et al. (2002) investigated the influence of mobile phone exposure on neural functioning, including performance in an auditory discrimination task. Twenty-four participants were exposed to 20-minutes GSM or sham exposure. During the exposure, participants completed the discrimination task four times. No significant differences were observed on reaction times.

Lee et al. (2003) tested 78 university students performance in different tasks, including an auditory vigilance test. This task involves withholding pressing a key for targets that occur infrequently. Participants in the experimental group performed better on this test only after they had been exposed to the GSM EMFs, thus suggesting an improvement of attention in the participants.

In a study proposed by Curcio and coworkers (2004), still discussed before, also a choice reaction time task has been administered to participants: in this case shorter reaction times (reduction of 40 ms) have been observed as a function of MP-like exposure.

Interestingly, an examination of medium term effects had also null results. Indeed, as we already discussed, Besset et al. (2005) used both choice reaction time tasks and sustained attention tasks in their battery of daily tests. These tests confirmed the absence of exposure effects on response speed.

As briefly introduced before, Hamblin et al. (2004) studied the effects of mobile phone exposure on psychomotor performance during an auditory task. Here, 12 participants were exposed to a GSM and sham signal for a total duration of 60 minutes. The choice reaction time task required participants to respond to target auditory stimuli by pressing the mouse button as fast as possible. A statistically significant difference in reaction times between real and sham conditions was accounted for by a reduced speed under real exposure (i.e. an increase of 53 ms). However, as no correction for multiple comparisons was applied, any conclusive interpretation of the data is compromised.

As already discussed in the previous paragraph, Keetley et al. (2006) reported mixed results, with an unexpected impairment of choice-reaction times and sustained attention. However, the authors did not apply any correction for multiple comparisons.

As previously described, Russo et al. (2006) and Haarala et al. (2007), assessed the effects of GSM, CW (same frequency as the GSM-signal), and sham signal on a choice-reaction time task. No effects of exposure on measures of speed were found.

Unterlechner et al. (2008) carried out a double-blind crossover study on 40 volunteers, assessing the effects of Low and High UMTS signal compared to Sham exposure on divided attention and vigilance tasks. Also in this case, no significant differences were reported on response speed.

In a study mainly aimed at investigating EEG features during an auditory oddball paradigm, Stefanics et al. (2008) exposed 36 participants to a UMTS and sham signal for 20 minutes. Performance (accuracy index) was tested before and after exposure: also in this case, no statistically significant effects of exposure were reported.

As already anticipated, Schmid et al. (2012) exposed their volunteers to two differently modulated GSM signals (14 and 217 Hz, respectively) and to a sham condition for 30 min before sleep. Again, the attentional task (choice reaction time task) did not vary as a function of exposure to RF EMFs.

With respect to the youngest, Haarala et al. (2005) tested 32 children under both sham and MP-like exposure, using a choice reaction time task (similar to the one used in Curcio et al., 2004) and a vigilance task. No effects due to the irradiation were reported in the two measures. A further experiment on adolescents was carried out by Preece et al. (2005): these authors investigated the effect of EMFs by means of participant's response times: again, no effects were observed.

Studies Dealing with Memory in Adults and Adolescents

Krause et al. (2000) enrolled 16 participants to complete an auditory verbal memory task, under both a 30-min GSM and sham condition, and exposed them over the right posterior temporal region. They reported no significant effect on performance as a function of exposure. Yet, the authors attempted to systematically replicate this study with methodological improvements (Krause et al., 2004). The participants were asked to complete an auditory verbal memory task lasting about 60 min (30 min under real GSM signal and 30 min under sham exposure). Here, at variance with the previous study, the left posterior temporal region was irradiated. Anyhow, no significant effects between conditions were reported.

In addition, as outlined in the previous sections, Koivisto et al. (2000) included a subtraction task. Their results indicated a significant reduction of response time (mean reduction of 29 ms), indicating an improvement of memory performance as a consequence of irradiation.

The multicentre study (two separate laboratories in Finland and Sweden) by Haarala et al. (2004) also investigated the short term memory task with varying memory load (0-, 1-, 2- and 3-back). No significant effect was observed on memory performance as a function of exposure to the GSM signal.

Haarala et al. (2003 b) within a study on the effects of a GSM exposure on cerebral blood flow, assessed the working memory of their participants. The task was similar to the one used in previous studies by the same group (Krause et al., 2000). Behavioural data were recorded on 10 participants, each exposed to both GSM and sham signal for about 45 minutes. No significant effects were reported on memory functioning. Some years later, the same research group carried out a new study on the effects of mobile phone exposure on cerebral blood flow (Aalto et al., 2006). Behavioural data were obtained by a simple working memory task (1-back task) in 12 participants during a 51 min

exposure to a modified GSM and sham signal. As in the previous study, no effects on memory performance were observed.

Curcio et al. (2004) reported a significant effect in a subtraction task, with shorter reaction times (i.e. reduction of 40 ms under real with respect to sham exposure). However, the same type of task applied to a much larger sample size ($n=168$) revealed no significant effect on reaction times (Russo et al., 2006).

Besset et al. (2005) investigated memory according to several tasks. A verbal learning task required to recall the maximum number of words remembered of a 15 words list. A sixth delayed recall was performed 45 min after the fifth presentation. The working memory was assessed with the digit span backward, the spatial span backward, and the number letter sequencing. In the backward task the participant was asked to repeat the sequences in exactly reversed order. In the number letter sequencing task the participants were required to listen to an auditory presentation of alternating numbers and letters and to repeat them back. No effect of mobile phone exposure on the reaction time performance in these tasks was found.

Krause et al. (2007) aimed to assess the effects of continuous wave (CW) and pulse modulated EMFs on brain functioning, and the possible presence of differences between left and right side EMFs exposure. These authors carried out two different experiments, each one on a sample of 36 healthy males: one with the auditory verbal memory task and one with the visual sequential letter memory task. Each experiment included six exposure conditions: sham, continuous wave, and pulse modulated wave were irradiated on both left- and right-side of the head. No significant effects were reported on speed of memory performance.

In one of the previously discussed studies by Regel et al. (2007 a), also auditory working memory was investigated. The results indicated a significant increase of response speed in the 2-back and 3-back tasks during real exposure. These effects appeared to be greater during the pulsed

condition than during the CW and sham conditions. No significant differences were observed when applying the simplest (1-back) condition for working memory. As the authors stated, these results show a significant improvement in memory performance, albeit subtle. In the following study proposed by the same group (Regel et al., 2007 b) previous results were not confirmed, since an isolated trend to increase of speed in the 1-back task as a function of the increase of field intensity was reported.

Sauter et al. (2011) aimed to compare possible cognitive effects of GSM and WCDMA/3G UMTS signals. In this study a 9 days long duration (7h 15 min per day) exposure was directed to the head. Of the 9 days, 3 were dedicated to GSM, 3 to WCDMA/3G UMTS, and 3 to sham conditions; consecutive experimental days were separated by two weeks. A 0- and 2-back task was administered twice every experimental day during the exposure. No statistically relevant effect of exposure was reported.

Schmid et al. (2012), in their study assessing the effects of two differently modulated GSM signals (14 and 217 Hz, respectively) with respect to sham condition, reported a reduction in speed performance at 2-back task under 217 Hz pulse-modulated exposure.

Finally, Leung et al. (2011) were the first to investigate the differential effects of RF emitted by mobile phones of second (2G, namely GSM) and third generation (3G, namely UMTS) on three groups of volunteers: adolescents, young adults and elderly. Each participant was exposed to three exposure conditions (sham, 2G, 3G) during which an n-back task at varying cognitive load (1-, 2-, 3-back) was administered. No effects on response speed were observed.

FUTURE RESEARCH DIRECTIONS

Most of the research findings here reported were included in two meta-analyses. Valentini et al. (2010) reviewed papers published in the period

1999-2008. Of the 42 relevant studies, 24 met the inclusion criteria (presence of real and sham condition in human provocation/laboratory studies; at least single blinded experimental studies; acute exposure to GSM and UMTS signals; enrolment of healthy subjects; inclusion of cognitive tests used at least once in other studies and reporting speed measures; randomization and/or counterbalancing of exposure sessions/conditions; detailed exposure characterization; unbiased statistical analyses). Results overall point to a lack of short-term acute effects of GSM-EMFs exposure on reaction times. Yet, most of the poorly methodologically-statistically controlled studies reported a faster performance in the simple reaction time task (measure of attention), 2-back task and 3-back task (measures of working memory) following GSM-RF exposure. Such effects disappeared when effect size was controlled for, by excluding outlier papers characterized by highest heterogeneity. Moreover, some of these positive findings were reported by studies with methodological flaws, as for example studies providing only single blind procedure. In fact, the replication of the same study with a double blind procedure systematically brought about negative findings. Summarizing, based on Valentini et al. (2010) results, it can be concluded that GSM-EMFs exposure does not produce relevant and consistent effects at cognitive level.

Most recently, Barth et al. (2012) came to the same conclusions. The authors searched for papers published until February 2011: 29 papers were reported and 17 of them fulfilled the requirements and were included in the analysis. Despite less stringent inclusion criteria were applied, results indicated no consistent short-term effects of GSM on the 11 cognitive outcomes taken into consideration (speed of processing, focused and sustained attention, working memory). To date, some of these studies initially reporting an effect of mobile phone-like EMFs exposure were subsequently replicated with statistical and methodological improvements: in all these cases positive findings were not confirmed. It should be stressed

that also some of the studies reporting no effects of exposure suffered from weaknesses, e.g. few participants, fixed order of exposure conditions, tasks administered in fixed sequence. Yet, there is thus no reliable and consistent evidence that cognitive performance of healthy adults is influenced by RF exposures. Too few studies on cognition have been conducted with children and adolescents to allow a final conclusion for this group.

Since the available literature, future research should be focused on the medium- and long-term effects to repeated EMFs exposure. So far, no psychomotor and cognitive effect has been found with daily acute repeated exposures (three 15 min exposures each day; Curcio et al., 2008), or in medium-term exposure designs (more than 7 h per day, for non-consecutive 9 days; Sauter et al., 2011; 2 h daily exposure over 3 weeks; Besset et al., 2005). The existence of possible middle- and long-term effects of mobile phone-RF exposure remains unresolved and is a cause of concern, especially regarding people in a developmental period. At present, in fact, only three reliable peer reviewed studies have been published on children and teenagers, and in all no significant effects on memory or attention performance were reported (Haarala et al., 2005; Preece et al., 2005; Loughran et al., 2013). Thus, more research is needed on potentially sensitive populations (children, the elderly, neurological impaired patients, RF workers, etc.) and caution should be urged until robust findings will motivate the exclusion of any potential risk (WHO, International Project Research Agenda, 2010). All these observations and suggestions should also be taken into consideration by international organizations as WHO when developing the ongoing Environmental Health Criteria (EHC) for RF-EMFs.

CONCLUSION

The first studies on humans addressing cognitive functioning changes as a consequence of radiofrequency (RF) EMFs exposure, date back

to almost 20 years ago. The effects on human behaviour showed in those pioneering works indicated a somewhat improvement of performance under the exposure to the signal, compared with sham exposure. These first and striking results were not fully replicated by subsequent studies that were characterized by a more methodological robustness and attention to exposure aspects. In accordance with this view, latest reviews and metaanalyses have confirmed the paucity of evidence and the lack of reliability of psychomotor and cognitive effects of acute RF EMF exposure on human volunteers, particularly when assessed in well controlled laboratory settings. Thus, despite the public opinion about potential biologic effects of acute RF EMFs irradiation, it can be concluded that to date there is substantial lack of evidence about a negative influence of non-ionizing radiations on cognitive functioning in humans.

ACKNOWLEDGMENT

I wish to thank Elia Valentini (now at the Department of Psychology, University of Essex) a friend and colleague with which I worked for some years on this topic and that co-authored some previous critical reviews that are the basic and logical ground for the present contribution.

REFERENCES

- Aalto, S., Haarala, C., Bruck, A., Sipilä, H., Hämäläinen, H., & Rinne, J. O. (2006). Mobile phone affects cerebral blood flow in humans. *Journal of Cerebral Blood Flow and Metabolism*, 26(7), 885–890. doi:10.1038/sj.jcbfm.9600279 PMID:16495939
- Barth, A., Ponocny, I., Gnambs, T., & Winker, R. (2012). No effects of short-term exposure to mobile phone electromagnetic fields on human cognitive performance: A meta-analysis. *Bioelectromagnetics*, 33(2), 159–165. doi:10.1002/bem.20697 PMID:21853449

- Besset, A., Espa, F., Dauvilliers, Y., Billiard, M., & de Seze, R. (2005). No effect on cognitive function from daily mobile phone use. *Bioelectromagnetics*, 26(2), 102–108. doi:10.1002/bem.20053 PMID:15672372
- Cardis, E., Deltour, I., Mann, S., et al. (2008). Distribution of RF energy emitted by mobile phones in anatomical structures of the brain. *Physics in Medicine and Biology*, 53(11), 2771–83.
- Croft, R. J., Chandler, J. S., & Burgess, A. P. (2002). Acute mobile phone operation affects neural function in humans. *Clinical Neurophysiology*, 113(10), 1623–1632. doi:10.1016/S1388-2457(02)00215-8 PMID:12350439
- Curcio, G., Ferrara, M., De Gennaro, L., Cristiani, R., DInzeo, G., & Bertini, M. (2004). Time-course of electromagnetic field effects on human performance and tympanic temperature. *Neuroreport*, 15(1), 161–164. doi:10.1097/00001756-200401190-00031 PMID:15106850
- Curcio, G., Valentini, E., Moroni, F., Ferrara, M., De Gennaro, L., & Bertini, M. (2008). Psychomotor performance is not influenced by brief repeated exposures to mobile phones. *Bioelectromagnetics*, 29(3), 237–241. doi:10.1002/bem.20393 PMID:18163437
- Ferreri, F., Curcio, G., Pasqualetti, P., De Gennaro, L., Fini, R., & Rossini, P. M. (2006). Mobile phone emissions and human brain excitability. *Annals of Neurology*, 60(2), 188–196. doi:10.1002/ana.20906 PMID:16802289
- GSMA-Groupe Speciale Mobile Association. (2016). Retrieved from <https://gsmaintelligence.com/>
- Haarala C, Aalto S, Hautzel H, et al. (2003 b). Effects of a 902 MHz mobile phone on cerebral blood flow in humans: a PET study. *Neuroreport*, 14(16), 2019–23.
- Haarala, C., Bergman, M., Laine, M., Revonsuo, A., Koivisto, M., & Hämäläinen, H. (2005). Electromagnetic field emitted by 902 MHz mobile phones shows no effects on childrens cognitive function. *Bioelectromagnetics*, 7(S7), S144–S150. doi:10.1002/bem.20142 PMID:16059918
- Haarala, C., Bjornberg, L., Ek, M., Laine, M., Revonsuo, A., Koivisto, M., & Hämäläinen, H. (2003 a). Effect of a 902 MHz electromagnetic field emitted by mobile phones on human cognitive function: A replication study. *Bioelectromagnetics*, 24(4), 283–288. doi:10.1002/bem.10105 PMID:12696088
- Haarala, C., Ek, M., Bjornberg, L., Laine, M., Revonsuo, A., Koivisto, M., & Hämäläinen, H. (2004). 902 MHz mobile phone does not affect short term memory in humans. *Bioelectromagnetics*, 25(6), 452–456. doi:10.1002/bem.20014 PMID:15300731
- Haarala, C., Takio, F., Rintee, T., Laine, M., Koivisto, M., Revonsuo, A., & Hämäläinen, H. (2007). Pulsed and continuous wave mobile phone exposure over left versus right hemisphere: Effects on human cognitive function. *Bioelectromagnetics*, 28(4), 289–295. doi:10.1002/bem.20287 PMID:17203481
- Hamblin, D. L., Croft, R. J., Wood, A. W., Stough, C., & Spong, J. (2006). The sensitivity of human event-related potentials and reaction time to mobile phone emitted electromagnetic fields. *Bioelectromagnetics*, 27(4), 265–273. doi:10.1002/bem.20209 PMID:16437544
- Hamblin, D. L., Wood, A. W., Croft, R. J., & Stough, C. (2004). Examining the effects of electromagnetic fields emitted by GSM mobile phones on human event-related potentials and performance during an auditory task. *Clinical Neurophysiology*, 115(1), 171–178. doi:10.1016/S1388-2457(03)00313-4 PMID:14706485

ICNIRP-International Commission on Non-Ionizing Radiation Protection. Guidelines for limiting exposure to time-varying electric, magnetic, and electromagnetic fields (up to 300 GHz). (1998). Retrieved from <http://www.icnirp.org/en/activities/work-plan/details/work-plan-hf.html>

Keetley, V., Wood, A. W., Spong, J., & Stough, C. (2006). Neuropsychological sequelae of digital mobile phone exposure in humans. *Neuropsychologia*, 44(10), 1843–1848. doi:10.1016/j.neuropsychologia.2006.03.002 PMID:16616941

Koivisto, M., Revonsuo, A., Krause, C. M., Haarala, C., Sillanmäki, L., Laine, M., & Hämäläinen, H. (2000). Effects of 902 MHz electromagnetic field emitted by cellular telephones on response times in humans. *Neuroreport*, 11(2), 413–415. doi:10.1097/00001756-200002070-00038 PMID:10674497

Krause, C. M., Haarala, C., Sillanmäki, L., Koivisto, M., Alanko, K., Revonsuo, A., & Hämäläinen, H. et al. (2004). Effects of electromagnetic field emitted by cellular phones on the EEG during an auditory memory task: A double blind replication study. *Bioelectromagnetics*, 25(1), 33–40. doi:10.1002/bem.10143 PMID:14696051

Krause, C. M., Pesonen, M., Haarala Bjornberg, C., & Hämäläinen, H. (2007). Effects of pulsed and continuous wave 902MHz mobile phone exposure on brain oscillatory activity during cognitive processing. *Bioelectromagnetics*, 28(4), 296–308. doi:10.1002/bem.20300 PMID:17203478

Krause, C. M., Sillanmäki, L., Koivisto, M., Häggqvist, A., Saarela, C., Revonsuo, A., & Hämäläinen, H. et al. (2000). Effects of electromagnetic field emitted by cellular phones on the EEG during a memory task. *Neuroreport*, 11(4), 761–764. doi:10.1097/00001756-200003200-00021 PMID:10757515

Kwon, M. S., & Hämäläinen, H. (2011). Effects of mobile phone electromagnetic fields: Critical evaluation of behavioral and neurophysiological studies. *Bioelectromagnetics*, 32(4), 253–272. doi:10.1002/bem.20635 PMID:21452356

Lee, T. M., Lam, P. K., Yee, L. T., & Chan, C. C. H. (2003). The effect of the duration of exposure to the electromagnetic field emitted by mobile phones on human attention. *Neuroreport*, 14(10), 1361–1364. doi:10.1097/00001756-200307180-00016 PMID:12876474

Oftedal, G., Loughran, S., & Rubin, J. (2016). Review of human experimental RF EMF studies published in the period 2011-2015. *BioEM Congress*.

Preece, A.W., Goodfellow, S., Wright, M.G., et al. (2005). Effect of 902 MHz mobile phone transmission on cognitive function in children. *Bioelectromagnetics*, 7, S138 - 43.

Preece, A. W., Iwi, G., & Davies-Smith, A. (1999). Effect of a 915-MHz simulated mobile phone signal on cognitive function in man. *International Journal of Radiation Biology*, 75(4), 447–456. doi:10.1080/095530099140375 PMID:10331850

Regel, S. J., Gottselig, J. M., & Schuderer, J. (2007b). Pulsed radio frequency radiation affects cognitive performance and the waking electroencephalogram. *Neuroreport*, 18, 803–807. doi:10.1097/WNR.0b013e3280d9435e PMID:17471070

Regel, S. J., Tinguely, G., Schuderer, J., Adam, M., Kuster, N., Landolt, H.-P., & Achermann, P. (2007a). Pulsed radio-frequency electromagnetic fields: Dose-dependent effects on sleep, the sleep EEG and cognitive performance. *Journal of Sleep Research*, 16(3), 253–258. doi:10.1111/j.1365-2869.2007.00603.x PMID:17716273

Russo, R., Fox, E., Cinel, C., Boldini, A., Defeyter, M. A., Mirshekar-Syahkal, D., & Mehta, A. (2006). Does acute exposure to mobile phones affect human attention? *Bioelectromagnetics*, 27(3), 215–220. doi:10.1002/bem.20193 PMID:16304701

Sauter, C., Dorn, H., Bahr, A., Hansen, M.-L., Peter, A., Bajbouj, M., & Danker-Hopfe, H. (2011). Effects of exposure to electromagnetic fields emitted by GSM 900 and WCDMA mobile phones on cognitive function in young male subjects. *Bioelectromagnetics*, 32(3), 179–190. doi:10.1002/bem.20623 PMID:21365662

Schmid, M. R., Loughran, S. P., Regel, S. J., Murbach, M., Bratic Grunauer, A., Rusterholz, T., & Achermann, P. et al. (2012). Sleep EEG alterations: Effects of different pulse-modulated radio frequency electromagnetic fields. *Journal of Sleep Research*, 21(1), 50–58. doi:10.1111/j.1365-2869.2011.00918.x PMID:21489004

Stefanics, G., Thuróczy, G., Kellényi, L., & Hernádi, I. (2008). Effects of twenty-minute 3G mobile phone irradiation on event related potential components and early gamma synchronization in auditory oddball paradigm. *Neuroscience*, 157(2), 453–462. doi:10.1016/j.neuroscience.2008.08.066 PMID:18835335

Unterlechner, M., Sauter, C., Schmid, G., & Zeithofer, J. (2008). No effect of an UMTS mobile phone-like electromagnetic field of 1.97 GHz on human attention and reaction time. *Bioelectromagnetics*, 29(2), 145–153. doi:10.1002/bem.20374 PMID:17957702

US Census Bureau. (2016). Retrieved from <http://www.census.gov/popclock/>

Valentini, E., & Curcio, G. (2015). Mobile phones-like electromagnetic fields effects on human psychomotor performance. In Z. Yan (Ed.), *Encyclopedia of Mobile Phone Behavior*. Hershey, PA: IGI Global. doi:10.4018/978-1-4666-8239-9.ch057

WHO – World Health Organization. (2010). *International project research agenda. EMF research priorities, 2010*. Retrieved from http://whqlibdoc.who.int/publications/2010/9789241599948_eng.pdf?ua=1

WHO – World Health Organization. (n.d.). Retrieved from http://www.who.int/topics/electromagnetic_fields/en/

ADDITIONAL READING

Barth, A., Ponocny, I., Gnambs, T., & Winker, R. (2012). No effects of short-term exposure to mobile phone electromagnetic fields on human cognitive performance: A meta-analysis. *Bioelectromagnetics*, 33(2), 159–165. doi:10.1002/bem.20697 PMID:21853449

Barth, A., Winker, R., Ponocny-Seliger, E., Mayrhofer, W., Ponocny, I., Sauter, C., & Vana, N. (2008). A meta-analysis for neurobehavioural effects due to electromagnetic field exposure emitted by GSM mobile phones. *Occupational and Environmental Medicine*, 65(5), 342–346. doi:10.1136/oem.2006.031450 PMID:17928386

Kheifets, L., Repacholi, M., & Saunders, R. et al.. (2005). The sensitivity of children to electromagnetic fields. *Pediatrics*, 116(2), e303–e313. doi:10.1542/peds.2004-2541 PMID:16061584

Valberg, P. A., van Deventer, T. E., & Repacholi, M. H. (2007). Workgroup report: Base stations and wireless networks-radiofrequency (RF) exposures and health consequences. *Environmental Health Perspectives*, 115(3), 416–424. doi:10.1289/ehp.9633 PMID:17431492

Valentini, E., Curcio, G., Moroni, F., Ferrara, M., De Gennaro, L., & Bertini, M. (2007). Neurophysiological effects of mobile phone electromagnetic fields on humans: A comprehensive review. *Bioelectromagnetics*, 28(6), 415–432. doi:10.1002/bem.20323 PMID:17503518

Valentini, E., Ferrara, M., Presaghi, F., Gennaro, L. D., & Curcio, G. (2010). Systematic review and meta-analysis of psychomotor effects of mobile phone electromagnetic fields. *Occupational and Environmental Medicine*, 67(10), 708–716. doi:10.1136/oem.2009.047027 PMID:20837651

van Rongen, E., Croft, R., Juutilainen, J., Lagroye, I., Miyakoshi, J., Saunders, R., & Xu, Z. et al. (2009). Effects of radiofrequency electromagnetic fields on the human nervous system. *Journal of Toxicology and Environmental Health. Part B, Critical Reviews*, 12(8), 572–597. doi:10.1080/10937400903458940 PMID:20183535

WHO – World Health Organization. International project research agenda. EMF research priorities, 2010. http://whqlibdoc.who.int/publications/2010/9789241599948_eng.pdf?ua=1 (Retrieved April 26, 2014).

WHO – World Health Organization. The International EMF Project. Retrieved April 26, 2014, from <http://www.who.int/peh-emf/project/en/>

KEY TERMS AND DEFINITIONS



Attention: The process of directing psychomotor and cognitive resources on a task.

Auditory Effects: The result of the irradiation exposure on the auditory function.

Behavioral Effects: The result of the irradiation exposure on the motor responses.

Cognitive Effects: The result of the irradiation exposure on, more generally, any cognitive activity measured by laboratory tasks.

EMFs: Electromagnetic fields are physical fields produced by electrically charged objects. The magnetic counterpart of the field is determined by the movement of the charged object.

GSM: Global System for Mobile communication. It was one of the most widely used wireless telephony technologies before the advent of 3G protocols.

Memory: The process of storing and retrieving or recognizing information.

Radiofrequency: Electromagnetic fields have frequencies of radiation in time and space. These go from large wavelength and slow cycles such as 3 KHz to short wavelength and fast cycles such as 300 GHz.

Reaction Time: The time lapsing between a stimulus and a response.

Identification of Wireless Devices From Their Physical Layer Radio-Frequency Fingerprints

Gianmarco Baldini

European Commission – Joint Research Centre, Italy

Gary Steri

European Commission – Joint Research Centre, Italy

Raimondo Giuliani

European Commission – Joint Research Centre, Italy

INTRODUCTION

Extensive research has been performed in recent years for the identification of wireless devices from their radio frequency (RF) emissions both intentional or unintentional. The term “intentional” is used to identify the RF emissions generated by a wireless device to implement a specific wireless standard. For example, the RF emission of the uplink transmission of a mobile phone, which implements a specific wireless communication standard. The term “unintentional” instead, identifies the RF emissions not directly related to the services offered by the wireless device or the wireless standard, but which are generated due to the operation of the device. For example, the RF emission generated by the electronic components of the device. It is well known in literature that electronic devices can release RF emissions containing information on the operation of the device, thus disclosing sensitive information and generating a security threat. This was one of the primary drivers for the definition of the TEMPEST (Telecommunication and Electronic Material Protected from Emanating Spurious Transmission) standard. TEMPEST refers to the possibility of spying on information systems through leaking emanations, including unintentional radio or electrical signals, sounds, and vibrations (see Rohatgi,

P., (2009) for a description of the TEMPEST standard). This issue is also known as emission security (EMSEC), which is a subset of communications security (COMSEC). Since there is already a well-defined standard (i.e., TEMPEST) and a considerable amount of work in EMSEC and COMSEC for “unintentional” emissions, this chapter focuses only on the collection of fingerprints from “intentional” emissions, which are generated by the wireless device while performing its communication function or other services.

The main idea of identifying a wireless device through its RF emissions is that the electronic circuits and the RF components have specific characteristics determined by the production and manufacturing processes. These characteristics, which result in unique differences, can be used to distinguish a wireless device from another. The RF components can include filters, amplifiers, oscillators and other electronics, which are used to compose and transmit the RF signal. The differences on the electronic components are randomly generated, and are mainly due to imperfections in the material or the component itself. For example, the material can have impurities due to the presence of different substances or tiny differences in the soldering or casing of the amplifier, which have an impact in the generation of the RF emissions. These imperfections

appear as a subtle modification of the RF signal in space even if the wireless device generates a signal conformant to the standard. For example, a GSM mobile phone can transmit a RF signal with the modulation and range of frequencies defined in the GSM standard, but the physical imperfections will produce minor changes in the amplitude or phase of the signal, which can be collected and processed by a receiver. Note that these minor changes will be substantially the same from statistical point of view in every transmission of the signal and they can be used as a fingerprint of the wireless component (and, consequently, of the GSM mobile phone). As it will be described in the following sections of this article, machine learning or signal processing techniques can be applied to the collected RF signal to extract the imperfections and the related fingerprints.

RF fingerprints can have many applications if the level of identification accuracy is high (e.g., 80-90% or more). The possibility of identifying wireless devices from their RF fingerprints can be used for multi-factor authentication, where a wireless device can be authenticated not only on the basis of conventional cryptographic methods but also by processing the RF fingerprints. Another potential application is to fight against the distribution of counterfeit products. Counterfeit wireless devices (e.g., mobile phones) have electronic components of worst quality in comparison to the genuine ones (Tehranipour et al., 2015). For example, if a counterfeit phone has been built with low grade RF amplifiers in the uplink transmission chain, it will generate different RF fingerprints compared to an original one.

The structure of this article is the following: the *Background* section provides the main definitions and a literature review on this topic. The section *RF fingerprinting methodology* describes the main workflow for the collection and processing of the RF signals to generate the fingerprints. It also highlights the main outstanding challenges. The section *Solutions and recommendations* identifies and describes the potential approach

and techniques that can be used to address the outstanding challenges described in the previous section. In case gaps are identified, the section *Future research directions* describes future possible research developments. Finally, the *Conclusions* section concludes the chapter.

BACKGROUND

The possibility to identify wireless devices from the intrinsic characteristic of their components has grown in importance in the last 10-15 years, due to the improvement in the radio frequency sensors and receivers' capabilities for A/D processing, which have also led to a decrease in price. In fact, the possibility to extract valuable RF fingerprints is possible only when the collected RF signals (i.e., the observables) are digitized with a high degree of clock precision. Many papers describe the use of high-end spectrum analyzers or oscilloscopes to collect and process the RF signal. For example, in Reising et al., (2010), the authors have used a high grade spectrum analyzer to capture and process the signals from a GSM mobile phone. Then the signals were down-converted, digitized and stored as complex in-phase and quadrature (I-Q) components. From the IQ samples, the authors have extracted statistical features, which are representative of the imperfections in the communication path and are used to generate the needed fingerprints. A filter was used to remove bias and unwanted interference, since the latter could degrade or remove the fingerprints. In other words, the collection and processing algorithm would not be able to identify and distinguish the fingerprints in the signal, if the wireless interference had overpowered the fingerprints in the signal. Another important aspect is the need to remove the content related information from the signal, otherwise the fingerprints would be biased by information content. Using this approach only the non-content sections of the GSM bursts are useful for device identification. Figure 1 provides

a description of a GSM burst, where the content related parts are the payload in sections 2a and 2b, while the sections 1,3,4 are non-content related.

On the basis of the previous discussion on the content-related sections, it is clear that the processing of fingerprints mainly depends on the wireless standard or technologies implemented by the RF device. Not only each wireless standard has its own content and non-content parts but different parts of the signal could be more or less suitable for the extraction of the fingerprints. This approach, although standard-dependent, is quite general in scope as most communication systems with the noticeable exception of CDMA have repeatable information patterns such as the Wi-Fi preamble.

In both previous cases, the authors have used high-end receiver systems with a high resolution and A/D dynamic range. Other papers have used also high-end receivers like expensive spectrum analyzers and oscilloscopes as in Bihl et al., (2015) for ZigBee and Reising et al., (2012) for WiMax standard. The advantage of using high-end receivers is that the sampling rate and the dynamic range allow a very high level of granularity to retrieve and identify the fingerprints of the wireless devices. At the same time, the cost of this high-end receiver equipment is in the order of magnitude of thousands of dollars, which does not make practical the collection and processing of radio frequency fingerprinting for a variety of applications like multi-factor security, fight against counterfeiting and so on. To address this issue, some authors have recently investigated the feasibility of fingerprinting with relatively low-cost receivers. An analysis of the impact on the identification accuracy between low-end and high-end receivers was recently reported in Patel et al., (2014), where six different ZigBee devices of the same brand were analyzed using a high cost receiver (i.e., NI PXIe-108) and a low cost software defined radio (i.e., the USRP 2921 model). The cost ratio between the most expensive receiver and the low cost receiver is at least four. The results show a difference of accuracy approximately between 6% and 8% (details are

in Patel et al., (2014)) especially where Gaussian noise is present. Similar results were achieved by Rehman et al., (2012) with the 802.11a standard, where the high-end equipment is a PSA Agilent E4448A spectrum analyzer and the low-end are USRP receivers. We note that the cost ratio for the equipment used in Rehman et al., (2012) (i.e., at least a factor of 10) is even higher than in Patel et al., (2014). The differences in accuracy are also higher: between 10% and 20% even in presence of limited Gaussian noise. The conclusion drawn by these preliminary studies is that the receivers can significantly impact the identification and verification accuracy.

The extraction of the fingerprints is usually implemented by a) extracting statistical features from the signal in space or by b) correlation of the signals as described in Rehman et al., (2014), Danev et al., (2012) and Khanna (2015). The first category has been used by Reising et al., (2010), Reising et al., (2015) and others, where the used statistical features are variance, skewness and kurtosis either in time or frequency. The mutual signal correlation has been used in Sanchez et al., (2015) for DECT device identification. Other techniques like the Hilbert-Huang transform (HHT) can be used to generate various statistical features both in time and frequency, as shown in Zhang et al., (2016) for GSM device identification.

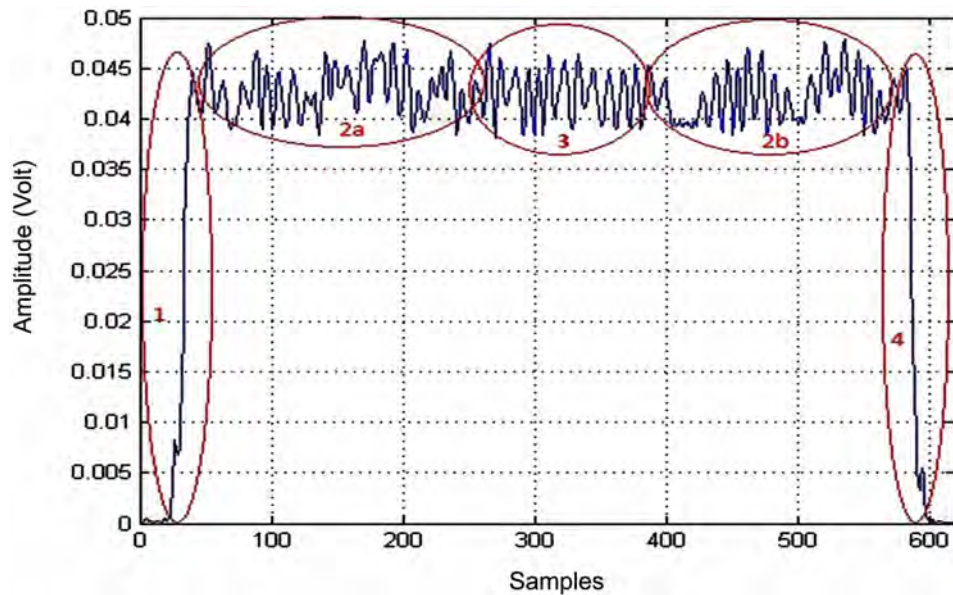
Some of the issues and challenges described in this paper have been identified also in recent surveys on the topic of RF device identification like, for example, in Khanna, (2015) and Xu et al., (2016).

RF FINGERPRINTING METHODOLOGY

Description of the Methodology

As described in several papers, the extraction of fingerprints of wireless devices can be based on the use of different set of features or the different communication standards. However, both

Figure 1. GSM burst with the content and non-content sections, which can be used for fingerprinting
Source: By the authors of this paper.



approaches have a common main methodology. Figure 2 depicts the workflow of the fingerprinting methodology, where the part in blue is common to all the fingerprint generation schemes, while the green and the orange flows represent the branches for the correlation-based and feature-based techniques respectively.

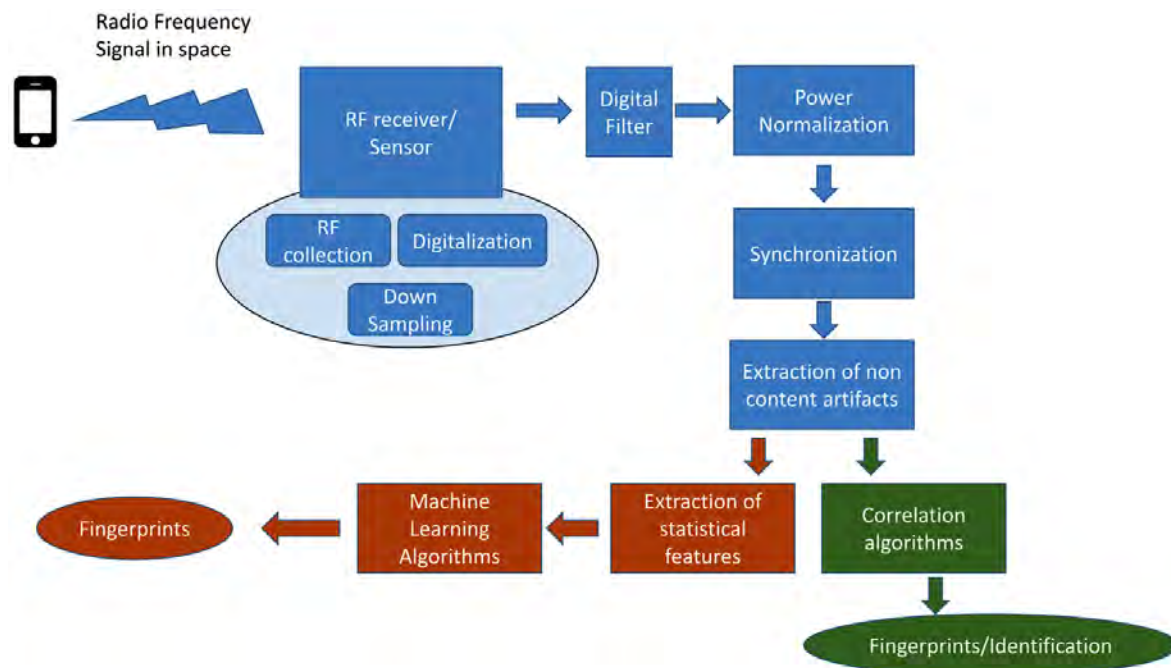
Each step is described as follows:

1. A wireless device (e.g., mobile phone) generates a radio frequency signal.
2. The signal in space is collected by a RF receiver/sensor (the terms receiver and sensor are used in literature with the same meaning). The RF receiver/sensor can be a high-end expensive spectrum analyzer or a low cost digitizer. This phase is critical to support the accuracy for identification and validation because the quality of the collected samples is usually proportional to the identification accuracy. The trade-off is between the cost of the RF receiver/sensor and its components and the level of accuracy. The RF receiver/sensor must provide the following functions described in the figure:
 - a) RF collection or capture of the RF signal in space through an antenna, an amplifier and other components, b) a down sampler to bring down the signal from the operational frequency to the baseband, and c) a digitizer like an A/D converter, which converts the analogic signal in digital samples.
3. A digital filter is usually applied to eliminate RF interference, which could have been captured together with the wanted RF signal.
4. The samples must be averaged and normalized to ensure that the application of the statistical features are coherent, otherwise the fingerprints would be biased.
5. In a similar way, the samples must be synchronized to ensure that the application of correlation or statistical features algorithms is executed on all the signal components at the same time. This step is general in scope and needs to be applied to the radio frequency signals that are composed by bursts such as WiFi, GSM, Bluetooth, etc. The bias introduced by lack of burst synchronization must be removed before the statistical features are applied.

6. At this point, two different techniques are applied (even if combined approaches are also possible). A first set of techniques (the green flow in Figure 2) is based on the application of correlation algorithms. For example, the GSM bursts of different GSM mobile phones are correlated to identify the bursts with higher similarity, which should be linked to a specific mobile phone. Another set of techniques (the orange flow in Figure 2) is based on the extraction of statistical features. For example, the value of variance, skewness or kurtosis are extracted from the GSM bursts to generate a collection of data sets, which can be used as fingerprints. The flows are described in the following steps respectively as flow A) and flow B).
7. (Flow A). The digital samples are correlated among them to evaluate their similarity on the basis of the consideration that digital samples of the same wireless device should be more correlated than the digital samples of different wireless devices. Various correlation algorithms can be used, including Pearson, Kendall or Spearman (see Hollander et al., (2013) for a description of the algorithms). Fingerprints are generated from the results of the correlation.
8. (Flow B). Statistical features are extracted from the digital samples to generate the fingerprints. Various statistical features can be used: from various moments of different orders like variance, kurtosis and skewness to the coefficients of the decomposition of the digital samples in different spectra or transformation (e.g., wavelets). The assumption demonstrated in literature is that the specific characteristics of the wireless devices can be captured to a different degree by the statistical features. The challenge is to identify the most appropriate or most representative features.
9. (Flow B). The statistical features extracted from the digital samples are processed through machine learning algorithms that can be both based on supervised or unsu-

Figure 2. Workflow for wireless devices identification

Source: By the authors of this paper.



pervised learning. In supervised learning, a training phase is performed. Then, the observables from a wireless device to be identified or validated are compared against the training set.

Removing Bias: Issues and Challenges

An important issue, which was highlighted by recent studies (Rehman et al., 2012), (Patel et al., 2014) and (Wang et al., 2016) is the influence of the receivers on the collection of the physical layer fingerprints and the issue of portability. When the signal is collected by the receiver, the digital samples do also contain the fingerprint of the receiver or they are biased by the specific limitations of the receiver. This phenomenon is more relevant in low quality and cost effective receivers because they introduce more bias. In particular, there is a strong difference between the fingerprints of high-end and low-end receivers as highlighted in Wang et al., (2016) and Patel et al., (2014). Even within receivers of the same class, the portability issue appears. As shown in Rehman et al., (2012), the fingerprints collected by one receiver cannot be directly used by another receiver for identification. In other words, if a receiver captures the observables and the tester uses this observable against a set of fingerprints generated with another receiver of the same class but different model, the identification process will usually fail. In Rehman et al., (2012), the authors have proven this aspect with a low cost USRP software defined radio. The identification accuracy will not be adequate in order to recognize the same wireless device on the basis of the fingerprints taken by the different receivers. This research problem still has to be solved. An example of this phenomenon can be seen in Figure 3, where the statistical features of Variance and Skewness are extracted from the RF observables of a single nRF51822 SoC RF module by using two different RF SDR USRP receivers (in green

and black). However, we can see clearly that the fingerprints are quite different among the different receivers even if the signals have been averaged and normalized.

Another relevant issue is that RF fingerprints may not be always collected in ideal conditions of Line of Sight (LOS) and absence of RF noise. While a reference library of RF fingerprints could be generated in ideal conditions, the RF observables for identification or validation could be impacted by the presence of attenuation and fading. Various authors have investigated the impact of Gaussian Noise (Reising et al., 2015) or Fading (Wang et al., 2016) and the result is obviously that accuracy will get worst in a proportional way to the presence of Gaussian Noise or Fading effects. An example of the impact of the attenuation on the identification accuracy is shown in Figure 4 for the identification of mobile phones supporting the GSM standard, where the X axis shows the Signal to Noise Ratio (SNR) expressed in dB, while on the Y axis it is reported the identification accuracy. The identification accuracy is calculated as the ratio of the sum of True Positive (TP) and True Negative (TN) over the total number of samples: $TP+TN/(TP+TN+FP+FN)$ where FP are the False Positives and FN are the False Negatives. The picture represents the identification of the accuracy between an Apple, a HTC, a Sony and a Samsung mobile phone, all supporting the GSM standard. The accuracy has been calculated using a Support Vector Machine (SVM) supervised learning algorithm.

Finally, either the wireless device or the receiver could be impacted by environmental conditions or lack of synchronization. While the clock drift of the wireless device could be itself an element to generate a physical fingerprint of the device, lack of synchronization in the receiver could affect the identification accuracy or the collection of the observables from the wireless device to build the fingerprints. Temperature could also influence the collection and generation of fingerprints even if there is a scarcity of studies on this aspect.

Figure 3. Fingerprints of the same device taken with two different receivers

Source: By the authors of this paper.

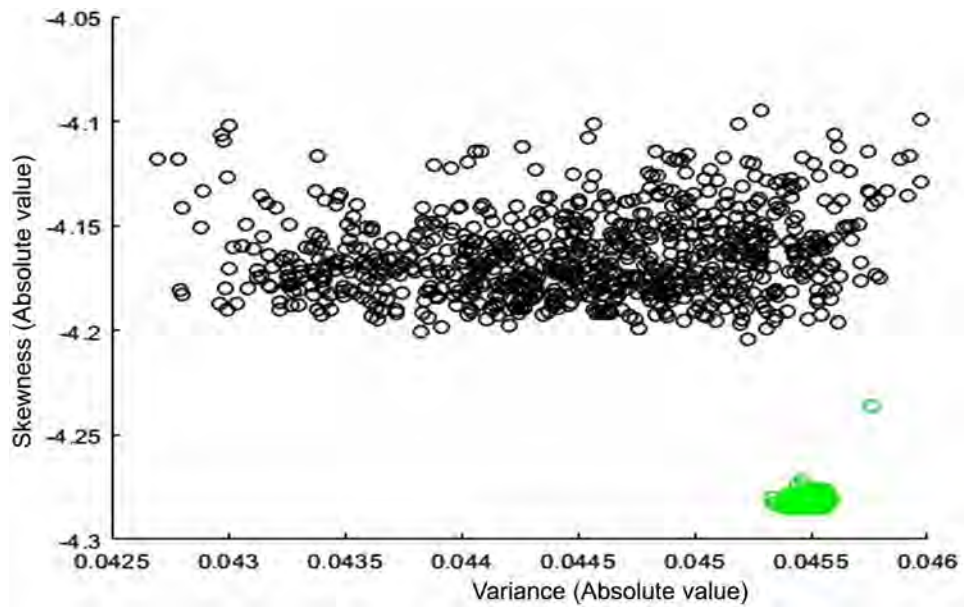
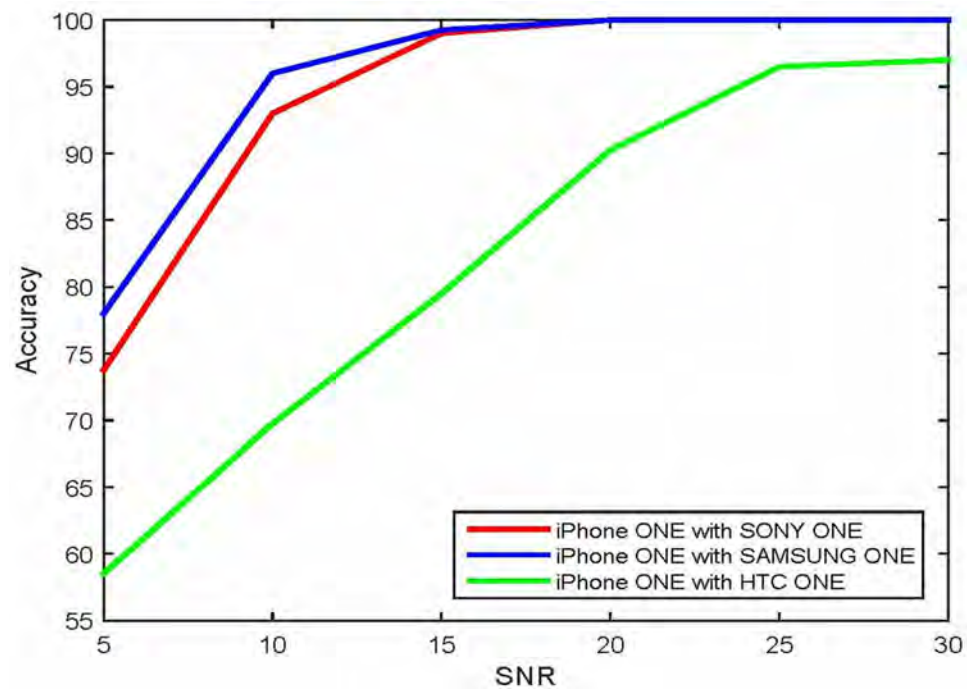


Figure 4. Impact of the attenuation on the identification accuracy

Source: By the authors of this paper.



SOLUTIONS AND RECOMMENDATIONS

In this section, we describe the potential solutions to address the issues described in the previous section. The issue of the bias introduced by the receiver in the collection and generation of the fingerprints of a wireless device could be theoretically mitigated in two ways:

1. By calibrating the receiver to remove the bias in the collection phase.
2. By identifying the receiver bias, recording it and applying it in the processing phase of the wireless device observables to remove it, so that the generation of the fingerprint is neutral against the receiver model.

This approach is not new and it is indeed applied in radio frequency measurements to remove the bias of the test equipment in measurement campaigns so that the test results are independent as much as possible from the bias of the test equipment and they are reproducible. Ideally, a reference signal, which is representative of the wireless communication standard supported by the wireless device, could be used to evaluate the bias of the receiver. The reference signal could be generated by a signal generator or similar test equipment. If such equipment is not available, a wireless device could be selected as a golden reference to evaluate the bias of the receivers. We note that this approach has not been studied in detail in literature even if it was only recently recommended by Wang et al., (2016).

The issue of non-ideal conditions for the collection of the fingerprints from wireless devices could be addressed by defining models for the attenuation and fading effects and applying them to the extraction of the fingerprints. The models must still be calibrated to the context where the observables are collected. In other words, the environment where the signal is collected should be evaluated and calibrated to identify which attenuation or fading model applies and what

are the parameters defining the model. This task may not be easy to achieve and it is obviously highly dependent on the context where the radio frequency measurements are taken. In another environment, the same models and parameters may not apply. An alternative approach is to select statistical features or correlation algorithms more resistant to attenuation and multi-path fading effects. A number of papers ((Reising et al., 2010), (Danev et al., 2012), (Wang et al., 2016), (Patel et al., 2015)), have investigated the impact of attenuation and multi-path fading on the identification accuracy. In particular, (Wang et al. 2016) have defined attenuation and fading models for fingerprinting and tested them against a specific wireless standard. In a similar way, the authors in Patel et al., (2015) have investigated the impact of attenuation (i.e., decreasing values of Signal to Noise Ratio) on the ZigBee wireless standard for different set of features and different algorithms.

Lack of synchronization or presence of clock drift in the receivers can be mitigated by synchronizing the receivers with Global Navigation Satellite Systems (GNSS) or by re-synchronization of the digital samples in a post processing phase. This latter approach may not be used when the fingerprint is the clock drift itself as described in Lanze et al., (2012) and Polak et al., (2015), then it is recommended to use the former approach whenever possible (e.g., when GNSS receivers and signal is available).

Finally, the dependency on the environment has still to be investigated by the research community. Even if many wireless devices are designed and built to provide stable performance in a wide range of temperatures (for the obvious reasons that they should support wireless communication in different context), there are limited studies about the performance of identification based on RF fingerprinting as pointed out in Rehman et al., (2014). The authors in Tekbas et al., (2004) have observed that a change in environmental conditions such as battery voltage or ambient temperature causes the statistical features vectors to spread over the feature space, resulting in classification

performance degradation. In the same reference (Tekbas 2004), it was experimentally shown that performance degradation resulting from changing environmental conditions can be compensated for by training the system with data collected in a range of battery-voltage and ambient temperature levels.

FUTURE RESEARCH DIRECTIONS

The future research directions are linked to the issues and solutions identified in the previous sections. New statistical features, which are resilient to multi-path fading effects, could be investigated to improve the identification accuracy in difficult propagation conditions. Here, the extensive work already done in the definition of signal processing algorithms to overcome multi-path fading in wireless communication or radar systems could be applied to fingerprinting. New statistical features could focus on the components of the radio frequency signal, which are less impacted by the multi-path fading effects.

The issue of portability of the fingerprints from one receiver to another remains a very important challenge to overcome if RF fingerprinting has to become a practical mean for identification or multi-factor authentication (Rehman et al., 2012), (Patel et al., 2014). In this context, researchers may investigate the most effective way to remove the fingerprints or bias of the receivers either by calibrating them or by comparing a set of RF fingerprints from golden references. Most of the work done so far in RF fingerprinting has been experimental: researchers have collected RF signals from wireless devices and processed them with algorithms and with a selection of different statistical features. The selection of the best algorithms or statistical features are mostly based on the experimental results. Future research directions should focus on the definition of theoretical models to identify the elements, which may increase identification accuracy and portability of the fingerprints. Since fingerprints are based on

subtle differences in the electronic components and materials, a joint effort between researchers in wireless propagation, machine learning and electronic materials could be beneficial.

CONCLUSION

This chapter describes the main techniques for the fingerprinting of wireless devices using their RF transmission. The focus is only on the collection of fingerprints from “intentional” emissions because there is already a consolidated literature for “unintentional” ones. Several articles have shown that for commercial systems such as WiFi, GSM, Bluetooth, and others, fingerprinting in ideal conditions can achieve high identification accuracy (e.g., 80-90% or more). This can be used for multi-factor authentication, where a wireless device can be authenticated not only on the basis of conventional cryptographic methods but also by processing the RF fingerprints. Another potential application is to fight against the distribution of counterfeit products. To support these applications, there are still some key challenges to overcome. This chapter tries to identify them in this context as well as providing possible approaches to solve them. Further research work is needed to investigate the portability issues between fingerprints taken using different receivers, as well as to identify and remove potential other sources of bias.

REFERENCES

Bihl, T. J., Bauer, K. W., Temple, M. A., & Ramsey, B. (2015, October). Dimensional reduction analysis for Physical Layer device fingerprints with application to ZigBee and Z-Wave devices. In *Proceedings of Military Communications Conference, IEEE MILCOM 2015* (pp. 360-365). Tampa, FL: IEEE. doi:10.1109/MILCOM.2015.7357469

- Danev, B., Zanetti, D., & Capkun, S. (2012). On physical-layer identification of wireless devices. *ACM Computing Surveys*, 45(1), 6. doi:10.1145/2379776.2379782
- Hollander, M., Wolfe, D. A., & Chicken, E. (2013). *Nonparametric statistical methods*. John Wiley & Sons.
- Khanna, V. K. (2015). Remote fingerprinting of mobile phones. *IEEE Wireless Communications*, 22(6), 106–113. doi:10.1109/MWC.2015.7368831
- Lanze, F., Panchenko, A., Braatz, B., & Zinnen, A. (2012, December). Clock skew based remote device fingerprinting demystified. In *Proceedings of the IEEE Global Communications Conference (GLOBECOM)*, (pp. 813-819). Anaheim, CA: IEEE doi:10.1109/GLOCOM.2012.6503213
- Patel, H., Temple, M. A., & Ramsey, B. W. (2014, October). Comparison of High-end and Low-end Receivers for RF-DNA Fingerprinting. In *Proceedings of 2014 IEEE Military Communications Conference (MILCOM)*, (pp. 24-29). Baltimore, MD: IEEE. doi:10.1109/MILCOM.2014.13
- Patel, H. J., Temple, M. A., & Baldwin, R. O. (2015). Improving ZigBee Device Network Authentication Using Ensemble Decision Tree Classifiers With Radio Frequency Distinct Native Attribute Fingerprinting. *IEEE Transactions on Reliability*, 64(1), 221–233. doi:10.1109/TR.2014.2372432
- Polak, A. C., & Goeckel, D. L. (2015). Wireless device identification based on RF oscillator imperfections. *IEEE Transactions on Information Forensics and Security*, 10(12), 2492–2501. doi:10.1109/TIFS.2015.2464778
- Rehman, S. U., Sowerby, K., & Coghill, C. (2012, September). Analysis of receiver front end on the performance of RF fingerprinting. In *Proceedings of the 2012 IEEE 23rd International Symposium on Personal, Indoor and Mobile Radio Communications-(PIMRC)* (pp. 2494-2499). Sidney, Australia: IEEE. doi:10.1109/PIMRC.2012.6362777
- Rehman, S. U., Sowerby, K. W., Alam, S., & Ardekani, I. (2014, October). Radio frequency fingerprinting and its challenges. In *Communications and Network Security (CNS), 2014 IEEE Conference on* (pp. 496-497). IEEE. doi:10.1109/CNS.2014.6997522
- Reising, D. R., Temple, M. A., & Jackson, J. A. (2015). Authorized and rogue device discrimination using dimensionally reduced RF-DNA fingerprints. *IEEE Transactions on Information Forensics and Security*, 10(6), 1180–1192. doi:10.1109/TIFS.2015.2400426
- Reising, D. R., Temple, M. A., & Mendenhall, M. J. (2010, April). Improving intra-cellular security using air monitoring with RF fingerprints. In *Proceedings of the 2010 IEEE Wireless Communication and Networking Conference (WCNC)*, (pp. 1-6). Sidney, Australia: IEEE. doi:10.1109/WCNC.2010.5506229
- Reising, D. R., Temple, M. A., & Oxley, M. E. (2012, January). Gabor-based RF-DNA fingerprinting for classifying 802.16 e WiMAX mobile subscribers. In *Proceedings of the International Conference on Computing, Networking and Communications (ICNC)*, (pp. 7-13). Maui, HI: IEEE. doi:10.1109/ICCNC.2012.6167534
- Rohatgi, P. (2009). Electromagnetic attacks and countermeasures. In *Cryptographic Engineering* (pp. 407-430). Springer US. doi:10.1007/978-0-387-71817-0_15
- Sanchez, I., Satta, R., Giuliani, R., & Baldini, G. (2015, May). Detection of DECT identity spoofing through Radio Frequency fingerprinting. In *Proceedings of the 38th – 2015- International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, (pp. 1296-1301). Opatija, Croatia: IEEE. doi:10.1109/MIPRO.2015.7160475
- Tehranipoor, M. M., Guin, U., & Forte, D. (2015). *Counterfeit Integrated Circuits*. Springer International Publishing.

Tekbas, O. H., Serinken, N., & Ureten, O. (2004). An experimental performance evaluation of a novel radio-transmitter identification system under diverse environmental conditions. *Canadian Journal of Electrical and Computer Engineering*, 29(3), 203–209. doi:10.1109/CJECE.2004.1532524

Wang, W., Sun, Z., Ren, K., Zhu, B., & Piao, S. (2015). Wireless Physical-Layer Identification: Modeling and Validation. *IEEE Transactions on Information Forensics and Security*, 11(9), 2091–2106.

Xu, Q., Zheng, R., Saad, W., & Han, Z. (2016). Device fingerprinting in wireless networks: Challenges and opportunities. *IEEE Communications Surveys and Tutorials*, 18(1), 94–104. doi:10.1109/COMST.2015.2476338

Yuan, Y., Huang, Z., Wu, H., & Wang, X. (2014). Specific emitter identification based on Hilbert-Huang transform-based time-frequency-energy distribution features. *IET Communications*, 8(13), 2404–2412. doi:10.1049/iet-com.2013.0865

ADDITIONAL READING

Chatfield, C. (2016). *The analysis of time series: an introduction*. CRC press.

Fulcher, B. D., & Jones, N. S. (2014). Highly comparative feature-based time-series classification. *IEEE Transactions on Knowledge and Data Engineering*, 26(12), 3026–3037. doi:10.1109/TKDE.2014.2316504

Goldsmith, A. (2005). *Wireless communications*. Cambridge university press. doi:10.1017/CBO9780511841224

Hagen, J. B. (2009). *Radio-frequency electronics: circuits and applications*. Cambridge University Press. doi:10.1017/CBO9780511626951

Joachims, T. (2002). *Learning to classify text using support vector machines: Methods, theory and algorithms*. Kluwer Academic Publishers. doi:10.1007/978-1-4615-0907-3

Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2012). *Foundations of machine learning*. MIT press.

Razavi, B., & Behzad, R. (1998). *RF microelectronics* (Vol. 1). New Jersey: Prentice Hall.

Rong, Z., & Rappaport, T. (1996). *Wireless communications*. Prentice Hall.

Ulversoy, T. (2010). Software defined radio: Challenges and opportunities. *IEEE Communications Surveys and Tutorials*, 12(4), 531–550. doi:10.1109/SURV.2010.032910.00019

KEY TERMS AND DEFINITIONS

Fingerprint: A specific characteristic of a wireless device.

Reference Library: Library of fingerprints of wireless devices.

RF Emissions: Emissions of wireless device in the radio frequency spectrum.

Software Defined Radio: A radio communication system or receiver where the functions to process the radio frequency signal are implemented in software rather than hardware.

Statistical Features: Features of a data set, which can be defined and calculated through statistical analysis.

Supervised Learning: The machine learning task of deducing a function from a set of labeled training data.

Unsupervised Learning: The machine learning task of deducing a function to describe hidden structure from a set of unlabeled data.

Wireless Device: An electronic device capable of transmitting and receiving information using radio frequency spectrum.

The Impact of Mobile Phones on Plastic Surgery and Burn Management

M

Maria Giaquinto-Cilliers

Kimberley Hospital Complex, South Africa

Tertius N. Potgieter

Kimberley Hospital Complex, South Africa

Gert Steyn

Kimberley Hospital Complex, South Africa

INTRODUCTION

Traditionally defined as “a specialized branch of surgery concerned with deformities and defects of the integument (the skin) and the underlying musculoskeletal framework” (Converse, 1997, p.3), plastic surgery nowadays combines traditional areas of expertise and principles with innovation; the plastic surgeon performs operations in diverse anatomical areas of the body, using creativity and innovation to restore the form and function affected by traumatic injuries, burns, congenital abnormalities, and losses caused by infection, tumours or cancer excision (Thorne, 2007). Subdisciplines of plastic surgery may include, amongst others: cosmetic or aesthetic surgery, burn surgery, craniofacial surgery, hand surgery, microsurgery and paediatric plastic surgery.

Although recognized as a specialty with extensive field of work, plastic surgeons are not always available in most hospitals around the world; besides, plastic surgeons are usually busy seeing patients in their offices, running between different hospitals and clinics, or operating, which makes them difficult to reach by emergency department doctors and in training young surgeons. There is an even greater shortage in developing countries, particularly in rural areas where, due to costs, distances to bigger centres must be considered

before the referral of patients who need assessment from the specialists. To address this issue and improve patient care through cost saving and time efficiency, an exponential growth for *telemedicine* (from the Greek prefix *tele* meaning ‘at a distance’ and the Latin word *meden* meaning ‘healing’) services in the field have been identified in recent years, varying from consultations by the telephone and image transfers with the purpose of diagnosis at a distance, to the complex technology of telesurgery (Grunwald, Krummel, & Sherman, 2014).

The new generation of smartphones are mobile phones with more advanced computing capability and connectivity. They boast higher resolution built-in digital cameras; larger crystal display screen; capacity to store data and images in archives; access to wireless data allowing online communication anywhere in the world and access to mobile applications – apps (Smartphone, 2014). These smartphones are equipped with the necessary technology to fulfil the prerequisites for a complete telemedicine system (‘m-health’), which by the World Health Organisation’s definition includes:

... the delivery of health care services, where distance is a critical factor, by all health care professionals using information and communica-

tion technologies for the exchange of valid information for diagnosis, treatment and prevention of disease and injuries, research and evaluation, and for the continuing education of health care providers, all in the interests of advancing the health of individuals and their communities. (World Health Organisation [WHO], 1998, as cited in WHO, 2010)

In the field of plastic surgery, patient's conditions are basically evaluated by visual inspection (including acute traumatic wounds and burns), and mobile phones may be used in remote, rural and resource-poor settings to provide essential information to specialists in larger and academic centres. Plastic surgeons are able to assess images and data and define management strategies required for those patients. Communication occurs not only between the plastic surgeon and the referring physician from other sites in the emergency setting, but also between plastic surgeons and patients (for continuity of care); nurses and plastic surgeons (for surgical care instructions); junior/trainee and supervisor/mentor plastic surgeons (for management instructions), and between senior plastic surgeons and their peers (for remote collaboration on diagnostically complex cases).

The availability of information in mobile technology has also allowed a growing market for education tools for plastic surgeons and their potential patients, and the use of social media for marketing's purpose.

In this chapter, the authors review the available literature on the use of mobile phones in the field of Plastic Surgery and Burns.

BACKGROUND

The majority of published studies debate telemedicine in plastic surgery using one of the two conventional systems, either the *store-and-forward (pre-recorded, asynchronous)* mode, which involves prior storage of still images or videos followed by transmission of the data (Buntic et al.,

1997; Roa, Gomes-Cia, Acha & Serrano, 1999; Baldwin & Langton, 2001; Pap, Lach & Upton, 2002; Jones, Milroy & Pickford, 2004; Wallace, Smith & Pickford, 2007; Varkey, Tan, Girotto, Tang, Liu & Chen, 2008; Wallace, Jones, Milroy & Pickford, 2008; Diver, Lewis & Gordon, 2009; Gray, Armfield & Smith, 2010; Trovato, Scholer, Vallejo, Buncke & Granick, 2011; Hoppe, Lee, Granick, & Scott, 2014) or the *real time (video conferencing, synchronous)* mode, which implies the interaction between patient, referring physician and specialist at the same time (Syed-Abdul, Scholl, Chen, Santos, Jian, Liou & Li, 2012).

Studies using mobile or cellular telemedicine, where portable devices with in-built camera capture digital images and computing and networking features allow direct and real time interaction, are still scarce but have shown promising results, particularly related to low costs associated with the use of this technology. Lam, Preketes, and Gates (2004), at Nepean Hospital in Australia; Tsai, Pong, Liang, Lin and Hsieh (2004) and Hsieh, Tsai, Yin, Chen, Yang and Jeng (2005), at Chang Gung Memorial Hospital in Taiwan; Saffle (2006), from the University of Utah, USA, and Shokrollahi, Sayed, Dickson, and Potokar (2007), from United Kingdom were amongst the earliest to report the successful use of mobile phones for teleconsultation in plastic surgery and burns.

Probably due to ethical and medico-legal issues related to the use of images and security concerns, and due to limitations to perform proper randomised controlled studies on the use of telemedicine, only a few more studies have been further published on the application of m-health for plastic surgery-related topics (Knobloch, Gohritz & Vogt, 2009; Knobloch, Rennekampff & Vogt, 2009; Engel, Huang, Tsao, Lin, Chou & Brey, 2011; Farber, Haik, Liran, Weissman and Winkler, 2011; Suliman, 2013; Hernandez, 2016. Wani, Rabah, AlFadil, Dewanjee & Najmi, 2013). Gardiner and Hartzell (2012), from Addenbrooke's Hospital in Cambridge, United Kingdom; and Wallace, Hussain, Khan and Wilson (2012), from Queen Elizabeth Hospital, Birmingham, United

Kingdom; Atiyeh, Dibo, and Janom (2014) from the University of Beirut, Lebanon performed a literature review of studies published on telemedicine for plastic surgery and burns, discussing technical and legal issues associated with this technology, including with the use of mobile phones.

More recently, plastic surgeons have been exploring the use of mobile technology for online education (m-Learning), research, and marketing. Recent publications have begun to explore smartphone applications more pertinent to plastic surgery (Amin, 2011; Amin, Edmonds, Abboudi & Sivakumar, 2012; Freshwater, 2011; Molajo, Vaiude & Benson, 2011; Workman & Gupta, 2013; Al-Hadithy & Ghosh, 2013; Valente, 2015

Australia, where Lam, (Lam, Preketes & Gates, 2004), who was undergoing training in Plastic and Reconstructive Surgery, used mobile phone photo-messaging for 27 cases of hand trauma (fractures, wounds, loss of fingertips, burns and infections) for correspondence between him and the consultant in the emergency department over a two-month period. He discussed the management with his senior consultant and sent the pictures, changing his strategy in four cases after checking out with his seniors. In this study, he recommended the use of mobile phone photo messaging into the clinical practice, stressing the low cost and ease of use of the available technology.

The same methodology was used by Tsai and colleagues (2004) at the Chang Gung Memorial Hospital, in Taiwan. A total of 60 patients and 82 extremity wounds were analysed for teleconsultation using transmission of 182 images through the remote camera phone. Despite the small dimension of the display screen, the quality of images was sufficient for identification of selected wound descriptors (gangrene, necrosis, erythema, and cellulitis) in the majority of the wounds, comparing the onsite evaluation with the remote assessment. The authors added as limitation of the study the restricted use of mobile telephone in some hospitals; the skills of the onsite photographer and intensity of circumstantial light in image acquisition interfering with quality of image acquired and further interpretation; and the need for a proper complete examination and transfer of information from the onsite examiner to the remote clinician for an adequate diagnostic and recommendation for management.

Hsiek and colleagues (2004) reviewed 45 mobile phone images of soft tissue injury of 81 digits, founding 76% sensitivity and 75% specificity for the remote diagnosis of the skin defect. In 25% of the cases reviewed, there was some discordance of opinions amongst the researchers involved which was explained by the low-resolution digital image and difficulties in the on-site physical examination, but the authors recommended the use of the mobile camera-phone associated with

CURRENT SCIENTIFIC KNOWLEDGE ON THE USE OF MOBILE PHONES IN PLASTIC SURGERY AND BURNS

1. Mobile Phones used for Teleconsultation in Acute Emergency Plastic Surgery

Plastic surgeons have always relied on photographs of their patients for comparison of pre and post operative results, for presentation in scientific meetings, or publication in peer-reviewed journals with educational purpose and dissemination of information within the scientific community. The technological upgrade from the former analogical cameras to the current digital cameras has resulted in better image resolution associated with better system for storage and filing of the images obtained. The advent of smartphones has lead to an even simpler and relatively less costly way to perform the same task, resulting in the substitution of the outdated and costly digital cameras for the practicality of the high-resolution, built-in camera-phones associated with all the technological package brought in with the most recent smartphones.

One of the former publications in this field is the study at Sydney's Nepean Hospital in

on-line verbal communication and review of the transmitted captured image by a senior specialist. They stressed the ease of use, low cost, high portability and mobility of this system, and the potential for future applications in telemedicine and telecare.

Farber and colleagues (2011), from the Sheba Medical Centre in Israel, used more advanced mobile phone technology to assess 57 patients; 58 multimedia consultation were performed between junior and senior plastic surgeons, using digital photograph from integrated phone camera sent via MMS (multimedia messages), video calls or video clips. The study suggested that high resolution images provided by this technology create an objective language that is not limited by the verbal skills of the examiner (junior doctor under training) nor by the listening comprehension of the senior doctor (specialist consultant). The transmission of videos was also an advantage, as there was no need for a specialized videoconferencing setting.

At the University of Heidelberg, Germany, Engel and colleagues (2011) monitored 103 consecutive free flaps performed and found that the remote smartphone photography assessment had a comparable accuracy rate (94.2%) as with in-person examination (98.7%), and shorter response time to re-exploration of the flaps, if needed (due to decrease in perfusion of the flaps).

Wani and colleagues (2013), from the King Fahad Medical City in Saudi Arabia assessed the efficacy of smartphone and its WhatsApp (Santa Clara, California), the instant short message service application for smartphones which allows creation of groups and sharing of real time text messages, pictures and videos, as a communication method amongst staff. The researchers found that this method was effective for patient care and academic endorsement, and suggested that it could be used for management of trauma and disaster to start pre-hospital management.

2. Mobile Phones used for Teleconsultation in the Assessment and Management of Burn Wound Injuries

Victims of burn injuries might require admission in burn centres or units, with highly specialized professionals participating in an interdisciplinary team. In parts of the world with few and relatively inaccessible burn centres, particularly in low-and middle-income countries, traditional telemedicine has shown a positive influence in burn injury management (Turk, Aydogan, Oguz, Tarim, Karakayali & Haberal, 2010; Bezuhly & Fish, 2012; Syed-Abdul et al., 2012). In military setting, traditional telemedicine has shown to improve the triage of burn patients from remote environments to rear-based specialists (McManus, Salinas, Morton, Lappan & Poropatich, 2008).

The severity of burn injuries is determined by the percentage of the total body surface (TBSA) injured, depth of the burn, age of the patient, the presence or absence of inhalation injury and whether other co-morbidities are present. The TBSA burned and the depth of the burn wound are assessed mainly by visual inspection, and this is influenced by the experience of the health professional who examines the victim. The appearance of the burn wound plays an important role in the selection of emergency treatment, need for surgery, type of dressing or the need for referral to the next level of care. The telephonic discussion between the referring facility and the burn expert may lack important clinical information, which may result in less than optimal treatment. Therefore, the inclusion of a digital/coloured image of the burn wound is helpful and almost mandatory for correct burns assessment: the receiving clinician will be able to validate any information given over the phone, as well as additional information may be captured from the images supplied, improving the communication between the parts involved in the patient's management (Wallace, Smith & Pickford, 2007; Giaquinto-Cilliers, 2013).

Saffle (2006), from the University of Utah, United States, acknowledged that the referral of inappropriately assessed burn cases to specialized centres will increase due to several factors. These factors include increased public awareness, the related liability adding pressure on less resourced settings and less experienced doctors with burn assessment and management. A single photograph of a burn wound taken and sent through a cell phone would minimize the unnecessary and expensive transportation of minor burn injuries and would thus also allow the feedback from the specialist centre supporting the decision to transport a significantly burnt patient for proper treatment.

Shokrollahi, et al. (2007), from the Welsh Burns Centre, United Kingdom, assessed 31 minor burns (less than 10% of TBSA) and showed that the extent and depth of the injuries could be accurately estimated with the use of a 1-mega-pixel camera integrated into a mobile phone. The authors also suggested that with the rate of technological advances, especially with mobile phones, this telemedicine modality might have exciting benefits in future times.

Knobloch, Rennekampff and Vogt (2009), from the Hannover Medical School, in Germany, discussed over digital images taken from two burn patients with a 5 mega pixel camera incorporated to a cell phone the immediacy and accuracy of the MMS transmitted images. The authors of this study stressed that a burn consultant, irrespective of his or her geographic position within the hospital, abroad or at home, after-office hours may be reached for discussion about the appropriate treatment to be given to a particular patient; this would increase patient's safety through the use of expertise consultation. Also, the JPEG images obtained by the phone camera could be easily stored in a hardware system for patient's documentation via the camera card, which is important for legal cases.

Wallace, Hussain, Khan and Wilson (2012) from Birmingham, United Kingdom, undertook a systematic literature review dating from 1993 till 2010 using visual images of 24 chosen articles.

The researchers discussed feasibility, clinical validation and cost effectiveness of the studies. They have confirmed that the current mobile phone technology provides resolution superior to the earliest telemedicine programmes and that interactive dynamic systems are now the technology, that is currently in mobile phones as single touch applications. These factors made them anticipate a growing number of case reports for the use of mobile phones for interactive burn assessment, although legal concerns and data protection still needs to be addressed. Furthermore, the researchers made a point that some static and video services provided encryption, which is needed for data protection (e.g. Skype™).

Atiyeh, Dibo and Janom (2014) reviewed the available literature about telemedicine and burn injuries however only two systematic reviews of the literature were found revealing a limited number of relevant published papers on this topic.

3. Mobile Phones Used as Other Resources by Plastic Surgeons

The use of mobile technology by healthcare professionals in the healthcare setting has gained popularity with the availability of several smartphone applications ("apps") specifically developed to aid clinicians from all medical specialties. As plastic surgery is per definition a speciality that must be up front with innovation and creativity, plastic surgeons have been encouraged to apply the latest technology offered by handheld devices.

Amin and colleagues (Amin, 2011; Amin, Edmonds, Abboudi & Sivakumar, 2012), from St. George's Hospital and Royal Free Hospital School of Medicine, United Kingdom, have pointed out that there are fewer applications currently on the market developed for plastic surgeons (senior and trainees), in comparison with other medical specialties. Freshwater (2011) and Molajo, Vaiude and Benson (2011) suggested that other non-specific apps (such as Body Mass Index- BMI- Calculator, updated information on medications, emergency guidelines, comprehensive medical calculators

and scores) might also be used by the plastic surgeon. Both groups suggested free and paid apps useful for the speciality, amongst them: anatomy apps, guidelines for open fractures of the lower limb (leg trauma), plastic surgery glossary, and apps from plastic surgery journals and scientific societies.

Although not specific for plastic surgeons, Mosa, Yoo and Sheets (2012), from the University of Missouri, USA, performed a systematic review of published articles discussing healthcare applications for smartphones to be used by healthcare professionals and patients.

Workman and Gupta (2013), from the Loma Linda University School of Medicine, reviewed the smartphone applications related to plastic surgery after an online search through all available markets (app stores) and classified the available applications (161 from the Apple's app store) in the following categories: practice or surgeon-based apps (54); academic-based apps (23); patient-based apps (10) and non-applicable apps (27). The majority of surgeon/practice-based apps involved opportunities for cosmetic surgeons to market their practice. Academic-based apps offered access to abstracts, papers, and other scientific content, although few specifically useful to plastic surgeons or opportunity for education. Patient-based apps provided education to patients in search of plastic surgery, while non-applicable apps usually involved creative photo morphing not important to plastic surgeons.

Al-Hadithy and Ghosh (2013), from the United Kingdom, also performed a review of the current use of smartphones and future developments for the practising and in-training plastic surgeons. The researchers presented a list of useful smartphone applications and mobile versions of websites (no need for apps download). They discussed the availability of podcasts (series of digital, audio or video files that are released episodically and downloaded via internet) and videos demonstrating online surgical procedures and clinical examinations through sites such as You-Tube, which are useful for in-training plastic surgeons

and the availability of plastic surgery textbooks in ebook versions reached through the smartphones.

A pilot study was launched in Brazil by Valente, D.S., Eifler, L.S., Carvalho, L.A. (2014) with the objective to demonstrate an experience of telemedicine in Plastic Surgery. The study entailed the real time streaming of a surgical procedure to 32 Plastic Surgeons and then to determine, whether or not, they would be able to perform it; 92,6% felt that they would be able to do the surgery demonstrated. Telemedicine thus provides additional access to education for Plastic Surgeons who seek medical education from more distant and isolated regions.

Dunne and Rawlins (2013) reviewed smartphone applications relevant to burn care and found it limited, and suggested the development of more applications educating on early burns management for the public and healthcare professionals.

There are currently a number of software applications for mobile phones with the purpose of supporting the management of burn injured patients. These include the BurnCare app by Pierre-Antoine Meley; BurnMed by Johns Hopkins Mobile medicine; LiAo Burns by Omesoft; uBurn by JAMB and MerseyBurns by St Helen and Knowsley Teaching Hospitals. With the exception of the latter two, very few of the applications available have been validated or described in scientific literature.

More recently Wallis, et al (2016) developed a smartphone application and cloud-based consultation system for burn injury emergency care, aiming to be different from the pre-existing mobile applications by providing actual consultation from a burn expert.

Grover (2013) added to the findings that the use of smartphones had the opportunity to reach, market to and educate potential and existing patients through non-practice related social media applications, such as Facebook, Twitter, YouTube, LinkedIn, Pinterest, and Tout. Through mobile phone use the plastic surgeon could update potential patients about the latest news, trends and topics in plastic surgery; invite patients to

upcoming educational seminars (particularly in cosmetic surgery) and post procedural videos and patient testimonials – a change in behaviour from a few years ago. This participation in social media is also discussed by Wheeler, Said, Prucz, Rodrich and Mathes (2011), Larson (2011) and Camp and Mills II (2012),

Even scientific and peer-reviewed journals for the speciality, professional societies and publishers in the field have joined the social media platform available in mobile phones through Facebook and Twitter (and foreseeing new platforms, such as Tumblr and Google Plus), to reach readers in a broader platform (plastic surgeons and ‘e-patients’) and interact with them; also, the new approach to social media has created a source for continuing medical education and best practices for the plastic surgery community/ world wide network (Rohrich & Weinstein, 2012).

Hughes, K. (2016) conducted a study with the goal to determine the feasibility of following up post-operative, plastic surgery patients, using only patient-taken images. The study highlighted the benefits for the management of patients who are separated from their surgeons due to geographic or financial constraints. This follow-up method ultimately led to a new management paradigm in the practices studied.

Other researchers have approached the use of mobile phone technology to create a ‘Telewound’ network for chronic wounds (Braun et al., 2005) and the home follow-up of patients undergoing ambulatory surgery (Ramos, Cerdan & Lopez, 2009). Even studies exploring Telewound care, and its use in home care settings with traditional telemedicine systems, have discussed the future migration of these systems to mobile phones devices, stimulated by prototypes of mobile systems in development that use spectral analysis to detect the presence of bacteria within a chronic wound (Santamaria & Kapp, 2013).

There is a general agreement that smartphone technology creates channels that are here to stay, and the behaviour of plastic surgeons has considerably changed since the conception of the first

smartphone. This behaviour has lead to the creation of new social media guidelines, addressing the benefits and drawbacks of online interaction, and proposing safeguards to avoid ethical risks (Kling, 2013).

Suliman M.T., from the King Khalid Civilian hospital in Saudi Arabia reported in 2013 that by using the WhatsApp service he was able to save patients’ travel expenses and treatment time by reducing the need for additional consultations after receiving images of a scar for review by a remotely located physician.

Aspects of consent amongst members of the Canadian Society of Plastic Surgeons were reviewed by Chan, N., Charette, J., Dumestre, DO. & Fraulin (2014). She determined that amongst 142 respondents, 75% believed that obtaining verbal consent before taking clinical photographs on a smart phone was sufficient to ensure the privacy rights of the patient were respected. Additionally, 72% of respondents felt that obtaining written consent for each patient prior to taking photographs was impractical. She concluded that smart phones are convenient and when password protected offer adequate security to ensure the privacy of the patient.

FUTURE RESEARCH DIRECTIONS

Traditional telemedicine systems using store-and-forward or real-time equipments are expensive and difficult to set up and require ongoing maintenance costs, while mobile phone messaging is comparatively cheaper and easier to use, with more portability and mobility. Limitations for the use of tele-technology include: poor infrastructure, limited equipment availability, and insufficient access to training and education for medical personnel, particularly in settings with poor resources as suggested by Lewis, Thomas, Wilson and Mbarika (2012).

Earlier studies have used relative ‘old’ mobile phone technology, where the quality of images (size and resolution) generated by small screen

probably showed much lower resolution than the current images offered by the new generation of smartphones; nevertheless, the results were promising. Definition even in small screens, and the use of mobile telemedicine will be more attractive to the specialty, with decrease in misinterpretation of the images. The transmission of videos showing full assessment of the function of the affected part of the body under examination (e.g., the hand movement) will add more value to future studies, although transmission time may vary according to the service provider. Also, 3D technology incorporated to cell phones will allow better assessment of wounds, and images will bring more information that is missed through the flat transmission currently obtained.

Currently, about 90% of handset sales worldwide are for devices powered by Google's Android (Samsung, HTC, Motorola, LG) and Apple's iOS mobile (iPhone) operating systems (Smartphone, 2014). Accessing the smartphones market stores now in search for applications for plastic surgery with terms such as *plastic surgery*, *cosmetic surgery*, *reconstructive surgery* and *burns* will display a crescent number of available apps compared with the previous published studies (Apple, 2014).

Patel, NG., Rozen, WM., Marsh, D. (2016) described the use of smartphone applications in the peri-operative management in microsurgical breast reconstruction in the United Kingdom. Though no single ideal plastic surgery application was found, multiple general applications were highlighted that could prove useful for use in microsurgical breast reconstruction. Applications were grouped into those used for efficient communication, storage, education and flap monitoring. Patel concluded that the future of smart devices lies in wearable devices able to monitor clinical condition and provide real time feedback where needed. Application development for the smart watch is seen as the start of this new revolution.

Future development of mobile applications for the specialty should also incorporate expert

knowledge of plastic surgeons to the development of specific software for the specialty (Amin, 2011; Workman & Gupta, 2013). Applications designed to collect data for research with the transfer of data electronically to a centrally based computer, described for other specialties (Tomlinson, 2009) should also allow research in certain fields of plastic surgery, particularly where epidemiological data is important, such as with burn injuries. The potential for using smartphones for data collection has been explored by Hundeshagen, Weissman, Farber, Winkler and Haik (2013).

WhatsApp (WhatsApp Inc., Mountain View, California, USA) is a cross platform, cost-effective mobile phone communication application facilitating the exchange of instant messages, pictures, videos and voice calls over the internet between two or more individuals simultaneously. The advantage of using a cheaper messaging platform is ease of access. As of 26 February 2016, one billion people around the world use WhatsApp.

The application allows in group text format multiple members to participate in a single ongoing discussion. This serves as an additional teaching platform as well as allowing input from multiple senior colleagues either within normal consultations, the operating room or after hour consultations. Fernandez-Valencia, JA., Egea, J., Pinyol, MC., Orench, M. (2015) explored the use of a WhatsApp group in aiding the coordination of a multidisciplinary surgical team. After eight months 830 messages were posted and constituted 177 conversations relating to operating room occupancy, operating schedule, prosthetic materials, patients' clinical issues, surgical information and miscellaneous topics. He concluded that instant messaging services such as that provided for by WhatsApp can be used to optimize patient care.

CONCLUSION

The integration of the mobile phone technology into wider practice will depend on clear regulations

set by official statutory bodies addressing patient's confidentiality and safer use of transmitted images and data. Medico legal issues, medical malpractice and security concerns must be addressed in a constructive way to enable integration of this tool into wider practice; through anonymising all data transferred, documenting the patient's consent, and keeping patient's confidentiality by an adequate and safe use of the images obtained (one phone available for the receiving unit), a par with the current regulations provided by the country's official organizations (Shokrollahi et al., 2007; Mars & Jack, 2010). Smartphones designed entirely around security, encryption and identity protection will address this issue in a near future. Smartphones, as well as standard mobile phones, may be stolen and security functions overridden allowing breach of confidentiality with the exposure of all data stored in the portable device.

There is a large potential clinical application for the use of mobile phones and their sophisticated resources as a telemedicine system in countries which lack human resources, distances between specialized centres are a real concern, and costs play an important factor in the health system (Giaquinto-Cilliers, 2013). The growing numbers of studies in recent years has shown the interest of plastic surgeons within this topic.

Above all, the validation of the increased value added by this 'cheaper' telemedicine system, as for the traditional system, must be constantly assessed by further researches, taking into consideration the acceptance and satisfaction of the patients, costs involved, infra-structure, and technical skills. Face-to-face interaction with the specialist is beneficial, and must always take precedence over other systems; but the current technology used to help where resources are scarce will definitively add value and possibly decrease costs in the initial management of the patient in need of plastic surgery.

REFERENCES

- Al-Hadithy, N., & Ghosh, S. (2013). Smartphones and the plastic surgeon. *Journal of Plastic, Reconstructive & Aesthetic Surgery; JPRAS*, 66(6), e155–e161. doi:10.1016/j.bjps.2013.02.014 PMID:23523169
- Amin, K. (2011). Smartphone applications for the plastic surgery trainee. *Journal of Plastic, Reconstructive & Aesthetic Surgery; JPRAS*, 64(9), 1255–1257. doi:10.1016/j.bjps.2011.03.026 PMID:21493171
- Amin, K., Edmonds, K., Abboudi, H., & Sivakumar, B. (2012). Smartphone applications for the plastic surgeon: An update of Google's android operating system. *Journal of Plastic, Reconstructive & Aesthetic Surgery; JPRAS*, 1440-1442. doi:10.1016/j.bjps.2012.04.043 PMID:22652289
- Apple. (2014). *iTunes Store*. Retrieved May 25, 2014, from <https://itunes.apple.com/us/genre/ios-medical>
- Atiyeh, B., Dibo, S. A., & Janom, H. H. (2014, June 30). Telemedicine and Burns: An overview. *Annals.*, 27(2), 87–93. PMID:26170782
- Baldwin, A. J., & Langton, S. G. (2001). Postoperative monitoring of flaps by digital camera and internet link. *The British Journal of Oral & Maxillofacial Surgery*, 39(2), 120–121. doi:10.1054/bjom.2000.0610 PMID:11286445
- Bezuhly, M., & Fish, J. S. (2012). Acute burn care. *Plastic and Reconstructive Surgery*, 130(2), e349–e358. doi:10.1097/PRS.0b013e318258d530 PMID:22842431
- Braun, R. P., Vecchietti, J. L., Thomas, L., Prins, C., French, L. E., & Gewirtzman, A. J. Salomon, D. Telemedical wound care using a new generation of mobile telephones. A feasibility study. *Archives of Dermatology*, 141(February), 254–258. PMID:15724023

- Buntic, R. F., Siko, P. P., Buncke, G. M., Ruebeck, D., Kind, G. M., & Buncke, H. J. (1997). Using the internet for rapid exchange of photographs and X-ray images to evaluate potential extremity replantation candidates. *Trauma*, 43(2), 342–344. doi:10.1097/00005373-199708000-00022 PMID:9291383
- Camp, S. M., & Mills, D. C. II. (2012). The marriage of plastic surgery and social media: A relationship to last a lifetime. *Aesthetic Surgery Journal*, 32(3), 349–351. doi:10.1177/1090820X12439862 PMID:22395327
- Chan, N., Charette, J., Dumestre, D. O., & Fraulin, F. O. G. (2016). Should Smart Phones be used for patient photography? *Plastic Surgery*, 24(1), 32–34. doi:10.4172/plastic-surgery.1000954 PMID:27054136
- Converse, J. M. (1997). Introduction to Plastic Surgery. In J. M. Converse (Ed.), *Reconstructive Plastic Surgery: Principles and procedures in correction, reconstruction and transplantation* (pp. 3–68). Philadelphia, PA: W.B. Saunders Company.
- Diver, A. J., Lewis, H., & Gordon, D. J. (2009). Telemedicine and trauma referrals- A plastic surgery pilot project. *The Ulster Medical Journal*, 78(2), 113–114. PMID:19568447
- Dunne, J. A., & Rawlins, J. M. (2013). *Mobile phone apps in burn care- A review*. Retrieved April 2, 2014, from <http://www.cdesign.com.au/anzba2013/posters/anzba2013asm1final00036.pdf>
- Engel, H., Huang, J. J., Tsao, C. K., Lin, C. Y., Chou, P. Y., Brey, E. M., & Cheng, M. H. et al. (2011). Remote real-time monitoring of free flaps via smartphone photography and 3G wireless internet: A prospective study evidencing diagnostic accuracy. *Microsurgery*, 31(8), 589–595. doi:10.1002/micr.20921 PMID:22072583
- Farber, N., Haik, J., Liran, A., Weissman, O., & Winkler, E. (2011). Third generation cellular multimedia teleconsultations in plastic surgery. *Journal of Telemedicine and Telecare*, 17(4), 199–202. doi:10.1258/jtt.2010.100604 PMID:21508079
- Fernandez-Valencia, J. A., Egea, J., Pinyol, M. C., Orench, M., Salas, E., & Gallart, X. (2015). WhatsApp Messenger for Team Co-ordination in Surgical Areas: Initial experience of a Hip Team in a Third Level Hospital. *International Journal of Advanced Joint Reconstruction*, 2(1), 23–25.
- Freshwater, M. F. (2011). iPhone and iPad applications for plastic surgeons. *Journal of Plastic, Reconstructive & Aesthetic Surgery; JPRAS*, 64(10), 1397–1399. doi:10.1016/j.bjps.2011.06.020 PMID:21911206
- Gardiner, S., & Hartzell, T. L. (2012). Telemedicine and plastic surgery: A review of its applications, limitations and legal pitfalls. *Journal of Plastic, Reconstructive & Aesthetic Surgery; JPRAS*, 65(3), e47–e53. doi:10.1016/j.bjps.2011.11.048 PMID:22178033
- Giaquinto-Cilliers, M. G. C. (2013). Telemedicine, mobile phones and burn wound Assessment and management: A valid resource for South Africa? *Wound Healing Southern Africa*, 6(2), 56–59.
- Gray, L. C., Armfield, N. R., & Smith, A. C. (2010). Telemedicine for wound care: Current practice and future potential. *Wound Practice & Research*, 18(4), 158–163.
- Grover, S. (2013). Commentary on Workman & Gupta (2013). *Aesthetic Surgery Journal*, 33(2), 281–282. doi:10.1177/1090820X12472339 PMID:23388647
- Grunwald, T., Krummel, T., & Sherman, R. (2014). Advanced technologies in plastic surgery: How new innovations can improve our training and practice. *Plastic and Reconstructive Surgery*, 114(6), 1556–1567. doi:10.1097/01.PRS.0000138242.60324.1D PMID:15509950
- Hoppe, I. C., Lee, B. A. Y., Granick, M. S., & Scott, S. S. (2014). Digital store and forward Imaging as a quality assessment tool for emergency plastic surgery consultations. *Eplasty*, 14. Retrieved from <http://www.eplasty.com> PMID:24501615

- Hsieh, C. H., Tsai, H. H., Yin, J. W., Chen, C. Y., Yang, J. C. S., & Jeng, S. F. (2004). Teleconsultation with the mobile camera-phone in digital soft-tissue injury: A feasibility study. *Plastic and Reconstructive Surgery*, 114(7), 1776–1782. doi:10.1097/01.PRS.0000142402.07896.21 PMID:15577348
- Hughes, K. (n.d.). *Article 1: Then Role of Telemedicine in the management of the cosmetic surgery patient*. Retrieved from <http://hughes-plasticsurgery.com>
- Hundeshagen, G., Weissman, O., Farber, N., Winkler, E., & Haik, J. (2013). Approaches to clinical research: Alternative use of Google docs survey function for mobile data collection using physicians smartphones. *Plastic and Reconstructive Surgery*, 131(1), e135–e136. doi:10.1097/PRS.0b013e318272a20a PMID:23271548
- Jones, S. M., Milroy, C., & Pickford, M. A. (2004). Telemedicine in acute plastic surgical trauma and burns. *Annals of the Royal College of Surgeons of England*, 86(4), 239–242. doi:10.1308/147870804344 PMID:15239862
- Kling, J. (2013). *New social media guidelines issued for physicians*. Retrieved April 29, 2014, from <http://www.medscape.com/viewarticle/782459>
- Knobloch, K., Gohritz, A., & Vogt, P. M. (2009). Cell phone-based multimedia messaging service in reconstructive microsurgery: A novel telemedicine application. *Plastic and Reconstructive Surgery*, 123(6), e220–e222. doi:10.1097/PRS.0b013e3181a3f53b PMID:19483557
- Knobloch, K., Rennekampff, H. O., & Vogt, P. M. (2009). Cell-phone based multimedia messaging service (MMS) and burn injuries. *Burns*, 35(8), 1191–1193. doi:10.1016/j.burns.2009.03.012 PMID:19577850
- Lam, T. K., Preketes, A., & Gates, R. (2004). Mobile phone photo messaging assisted communication in the assessment of hand trauma. *The Australian and New Zealand Journal of Surgery*, 74(7), 598–602. doi:10.1111/j.1445-2197.2004.03069.x PMID:15230801
- Larson, D. (2011). Commentary on Wheeler et al (2011). *Aesthetic Surgery Journal*, 31(4), 444. doi:10.1177/1090820X11405029 PMID:21551437
- Lewis, E. R., Thomas, C. A., Wilson, M. L., & Mbarika, W. A. (2012). Telemedicine in Acute-phase injury management: A review of practice and advancements. *Telemedicine Journal and e-Health*, 18(6), 434–445. doi:10.1089/tmj.2011.0199 PMID:22694296
- Mars, M., & Jack, C. (2010). Why is telemedicine a challenge to the regulators? *South African Journal of Bioethics and Law*, 3(2), 55–58.
- McManus, J., Salinas, J., Morton, M., Lappan, C., & Poropatich, R. (2008). Teleconsultation program for deployed soldiers and healthcare professionals in Remote and austere environments. *Prehospital and Disaster Medicine*, 23, 210–216. PMID:18702266
- Molajo, A., Vaiude, P., & Benson, A. (2011). 'iReconstruct': Clinically relevant plastic surgery apps. *Journal of Plastic, Reconstructive & Aesthetic Surgery*. doi:10.1016/j.bjps.2011.11.032
- Mosa, A. S. M., Yoo, I., & Sheets, L. (2012). A systematic review of healthcare applications for smartphones. *BMC Medical Informatics and Decision Making*, 12(1), 67. doi:10.1186/1472-6947-12-67 PMID:22781312
- Pap, S., Lach, E., & Upton, J. (2002). Telemedicine in plastic surgery: E-consult the attending surgeon. *Plastic and Reconstructive Surgery*, 110(2), 452–456. doi:10.1097/00006534-200208000-00012 PMID:12142659

- Patel, N. G., Rozen, W. M., Marsh, D., Chow, W. T. H., Vickers, T., Khan, K., & Hunter-Smith, D. J. et al. (2016). Ramakrishnan. Modern use of smartphone applications in the perioperative management in microsurgical breast reconstruction. *Gland Surgery*, 5(2), 150–157. PMID:27047783
- Ramos, C.M, Cerdan, M. T. & Lopez, R.S. (n.d.). Mobile phone-based telemedicine system for the home follow-up of patients undergoing ambulatory system. *Telemedicine & e-Health*, 15(6), 531-538. Doi: 10.1089/tmj.2009.0003
- Roa, L., Gomez-Cia, T., Acha, B., & Serrano, C. (1999). Digital imaging in remote diagnosis of burns. *Burns*, 25(7), 617–623. doi:10.1016/S0305-4179(99)00053-4 PMID:10563688
- Rohrich, R. J., & Weinstein, A. G. (2012). Connect with plastic surgery: Social media for good. *Plastic and Reconstructive Surgery*, 129(3), 789–792. doi:10.1097/PRS.0b013e3182402f85 PMID:22373982
- Saffle, J. R. (2006). Telemedicine for acute burn treatment: The time has come. *Journal of Telemedicine and Telecare*, 12(1), 1–3. doi:10.1258/135763306775321353 PMID:16438771
- Santamaria, N., & Kapp, S. (2013). TeleWound care – providing remote wound assessment and treatment in the home care setting: Current status and future directions. *Smart Homecare Technology & TeleHealth*, 1, 35–41. doi:10.2147/SHTT.S34353
- Shokrollahi, K., Sayed, M., Dickson, W., & Poto-kar, T. (2007). Mobile phones for the assessment of burns: We have the technology. *Emergency Medicine Journal*, 24(11), 753–755. doi:10.1136/emj.2007.046730 PMID:17954825
- Smartphone. (2014). In *Wikipedia, the free encyclopedia*. Retrieved May 27, 2014, from <http://en.wikipedia.org/wiki/Smartphone>
- Suliman, M. T. (2014). Sending photos through WhatsApp: A faster method for teleconsultation. *Journal of Local and Global Health Science*, 2.
- Syed-Abdul, S., Scholl, J., Chen, C.C., Santos, M.D.P.S., Jian, W.S., Liou, D.M., & Li, Y-C. (2012). Telemedicine utilization to support the management of the burns treatment involving patient pathways in both developed and developing countries: A case study. *Journal of Burns Care & Research*, 33(4), e207-e2012. Doi: 10.1097/BCR.0b013e318241b6b7
- Thorne, C. H. (2007). Techniques and principles in Plastic Surgery. In C. H. Thorne (Ed.), *Grabb & Smith's Plastic Surgery* (pp. 3–14). Philadelphia, PA: Lippincott Williams & Wilkins.
- Tomlinson, M., Solomon, W., Singh, Y., Doherty, T., Chopra, M., Ijumba, P., & Jackson, D. et al. (2009). The use of mobile phones as a data collection tool: A report from a household survey in South Africa. *BMC Medical Informatics and Decision Making*, 9(1), 51. doi:10.1186/1472-6947-9-51 PMID:20030813
- Trovato, M. J., Scholer, A. J., Vallejo, E., Buncke, G. M., & Granick, M. S. (2011). eConsultation in Plastic and Reconstructive Surgery. *Eplasty*, 11(November), e48. PMID:22140594
- Tsai, H. H., Pong, Y. P., Liang, C. C., Lin, P. Y., & Hsieh, C. H. (2004). Teleconsultation by using the mobile camera phone for remote management of the extremity wound: A pilot study. *Annals of Plastic Surgery*, 53(6), 584–587. doi:10.1097/01.sap.0000130703.45332.3c PMID:15602257
- Turk, E. K. E., Aydogan, C., Oguz, H., Tarim, A., Karakayali, H., & Haberal, M. (2010). Use of telemedicine in decision-making and burn follow-up: Initial experience from two burn units. *Burns*, 37, 415–419. doi:10.1016/j.burns.2010.10.004 PMID:21146313

Varkey, P., Tan, N. C., Girotto, R., Tang, W. R., Liu, Y. T., & Chen, H. C. (2008). A picture speaks a thousand words: The use of digital photography and the internet as a cost-effective tool in monitoring free flaps. *Annals of Plastic Surgery*, 60(1), 45–48. doi:10.1097/SAP.0b013e31805003df PMID:18281795

Valente, D.S., Eifler, L.S., Carvalho, L.A., Filho, G.A.P., Ribeiro, V.W., Padoin, A.V. (2015). Telemedicine and Plastic Surgery: A Pilot Study. *Plastic Surgery International*.

Wallace, D. L., Smith, R. W., & Pickford, M. A. (2007). A cohort study of acute plastic surgery trauma and burn referrals using telemedicine. *Journal of Telemedicine and Telecare*, 13(6), 282–287. doi:10.1258/135763307781644933 PMID:17785024

Wallace, D. L., Jones, S. M., Milroy, C., & Pickford, M. A. (2008). Telemedicine for acute plastic surgical trauma and burns. *Journal of Plastic, Reconstructive & Aesthetic Surgery; JPRAS*, 61(1), 31–36. doi:10.1016/j.bjps.2006.03.045 PMID:18068653

Wallace, D. L., Hussain, A., Khan, N., & Wilson, Y. T. (2012). A systematic review of the evidence for telemedicine in burn care: With a UK perspective. *Burns*, 38(4), 465–480. doi:10.1016/j.burns.2011.09.024 PMID:22078804

Wallis, L. A., Fleming, J., Hasselberg, M., Laflamme, L., & Ludin, J. (2016). A Smartphone App and Cloud-Based Consultation System for Burn Injury Emergency Care. *PLoS ONE*, 11(2), e0147253. doi:10.1371/journal.pone.0147253 PMID:26918631

Wani, S. A., Rabah, S. M., AlFadil, S., Dewanjee, N., & Najmi, Y. (2013). Efficacy of communication amongst staff members at plastic and reconstructive surgery section using smartphone and mobile WhatsApp. *Indian Journal of Plastic Surgery*, 46(3), 502–505. doi:10.4103/0970-0358.121990 PMID:24459338

WhatsApp Website. (n.d.). Retrieved from <http://blog.whatsapp.com/>

Wheeler, C. K., Said, H., Prucz, R., Rodrich, R. J., & Mathes, D. W. (2011). Social media in plastic surgery practices: Emerging trends in North America. *Aesthetic Surgery Journal*, 31(4), 435–441. doi:10.1177/1090820X11407483 PMID:21551435

Workman, A. D., & Gupta, S. C. (2013). A Plastic surgeons guide to applying smartphone technology in patient care. *Aesthetic Surgery Journal*, 33(2), 275–280. doi:10.1177/1090820X12472338 PMID:23335649

World Health Organisation. (2010). *Telemedicine: Opportunities and developments in Member States: Report on the second global survey on eHealth 2009*. Retrieved May 27, 2014, from http://www.who.int/goe/publications/goe_telemedicine_2010.pdf

ADDITIONAL READING

Gardiner, S., & Hartzell, T. L. (2012). Telemedicine and plastic surgery: A review of its applications, limitations and legal pitfalls. *Journal of Plastic, Reconstructive & Aesthetic Surgery; JPRAS*, 65(3), e47–e53. doi:10.1016/j.bjps.2011.11.048 PMID:22178033

Giaquinto-Cilliers, M. G. C. (2013). Telemedicine, mobile phones and burn wound Assessment and management: A valid resource for South Africa? *Wound Healing Southern Africa*, 6(2), 56–59.

Grunwald, T., Krummel, T., & Sherman, R. (2014). Advanced technologies in plastic surgery: How new innovations can improve our training and practice. *Plastic and Reconstructive Surgery*, 114(6), 1556–1567. doi:10.1097/01.PRS.0000138242.60324.1D PMID:15509950

Shokrollahi, K., Sayed, M., Dickson, W., & Poto-
kar, T. (2007). Mobile phones for the assessment
of burns: We have the technology. *Emergency
Medicine Journal*, 24(11), 753–755. doi:10.1136/
emj.2007.046730 PMID:17954825

Wallace, D. L., Hussain, A., Khan, N., & Wilson,
Y. T. (2012). A systematic review of the evidence
for telemedicine in burn care: With a UK per-
spective. *Burns*, 38(4), 465–480. doi:10.1016/j.
burns.2011.09.024 PMID:22078804

Workman, A. D., & Gupta, S. C. (2013). A Plastic
surgeons guide to applying smartphone technol-
ogy in patient care. *Aesthetic Surgery Journal*,
33(2), 275–280. doi:10.1177/1090820X12472338
PMID:23335649

KEY TERMS AND DEFINITIONS

Burns: Injury of the skin caused by heat,
electricity, chemicals, friction or radiation.

**Mobile Phone/s and Mobile Phone Technol-
ogy:** Handset phone that can make and receive
telephone calls while moving around a wide
geographic area.

Mobile Telemedicine: Use of mobile device-
s for consultation at a distance.

Photo Messaging: Transmission of pictures
through mobile phones.

Plastic Surgery: Specialized branch of sur-
gery that aims to restoring the form and function
of the skin and underlying structures affected by
traumatic injuries, burns, congenital abnormali-
ties, and losses caused by infection, tumours or
cancer excision.

Reconstructive Surgery: Sub discipline of
plastic surgery.

Smartphone(s): Mobile phones with more
advanced computing capability and connectivity
than basic feature phones.

Smartphone Application: Software pro-
gramme designed to run in smartphones and other
mobile devices.

Telemedicine: Delivery of health care services
at a distance by health care professionals using
information and communication technologies for
diagnosis, treatment and prevention of disease and
injuries, research and evaluation, and continuing
education of health care providers.

The Intersection of Religion and Mobile Technology

M**Wendi R. Bellar***Texas A&M University, USA***Kyong James Cho***Texas A&M University, USA***Heidi A. Campbell***Texas A&M University, USA*

INTRODUCTION

The proliferation of mobile technology throughout society has influenced many aspects of contemporary life, including religious practice. While much scholarship is focused on religious mobile applications (apps), earlier scholars also explored how different religious groups adopted, negotiated, or rejected mobile technology according to their core values and beliefs. Research on the intersection of mobile technology and religion falls under studies of “digital religion,” which seeks to understand the complex relationship between digital technology and religion (Campbell, Altenhofen, Bellar, & Cho, 2014). Broad themes explored within digital religion over the last 20 years include religious authority, identity, community, and authenticity (Campbell et al., 2014). These research themes help reveal the ways religious individuals and communities engage with both mobile technologies and digital culture as a whole.

This article is divided into three sections. First, a brief background on studies of mobile technology and digital religion provides necessary context for understanding the history and trajectory of religion and mobile studies. Next, three key areas of religion and mobile media research are explicated: how religious individuals and communities negotiate mobile media use, the ways in which religious mobile apps are being designed and used for religious practice, and the commodification

of religion through mobile media. Finally, the article concludes with a discussion of emerging trends within mobile media and religion, which suggests future research opportunities.

BACKGROUND

Mobile media have been described as “unanchored communication” which takes place “during transition” (S. Campbell, 2013, p. 10). It is important, however, to distinguish between laptop devices and mobile devices. While laptops can move from place-to-place, the proper infrastructures, from physical space to wireless internet connections, are required for use. Mobile devices, however, can be used anytime, anywhere. Additionally, the diffusion of mobile devices throughout the developed world raises concerns and inquiries about their social and cultural impacts (Smith, 2013), such as the positive and negative impacts mobile technology can have on religious communities.

In the US and other developing countries, mobile phone penetration has almost reached saturation point and smartphone adoption is increasing rapidly (“2 billion consumers”, 2014). Recent surveys also provide a clearer picture of mobile app use. Neilson (2015) found apps account for 89% of mobile phone usage, while 11% can be attributed to the mobile web. In the US, the average smartphone user has approximately 28 apps

DOI: 10.4018/978-1-5225-2255-3.ch535

downloaded on their devices and has used 11 of those apps within the last month (Google, 2015).

While religious mobile use is not reflected in these numbers, other studies point to religious engagement through mobile technology. For example, religious engagement online has been well documented since the 1980s (Campbell, 2010). Sixty-four percent of Americans reported searching for religious information or engaging in religious practice online (Hoover, Clark, & Rainie, 2004). The Barna Group reported that one in four Christians uses an e-reader or other mobile device for reading the Bible (2013). That same report showed that pastors' use of mobile technology to read and study the Bible tripled in two years to 44% (Barna, 2013). While it is clear mobile technology is being used for religious purposes, the impact this mobile technology use is having on religious users and communities is underresearched.

Numbers on the diffusion of mobile technology in general, and in religious contexts, highlight the need for careful study of the ways in which these technologies influence religion and how religious users engage mobile technology. This brief review of statistics implies the intersection of religion and technology follows similar patterns as mainstream adoption and use. Yet in their work digital religion scholars emphasize mediating factors such as specific beliefs, core values, and ritual practices that contribute to unique forms of engagement (Campbell, 2010). Therefore, digital religion seeks to understand these unique engagements, to illuminate the broader relationship between technology and religion as a whole.

Digital Religion

Digital religion is defined as a “framework for articulating the evolution of religious practices online which are linked to online and offline contexts simultaneously” (H. Campbell, 2013, p. 1) Research on mobile technology and religion is a subfield of digital religion in that: a) mobile technologies are providing a new way for prac-

ticing religion; and b) those religious practices happen concurrently in multiple online/mobile/offline contexts While early studies of religion and media contextualized the online and offline as separate spaces, digital religion argues for a “third space” in which the two are blurred (Hoover & Echchaibi, 2012). Mobile users and religious contexts occupy these two spaces simultaneously in everyday life.

Studies on religion and the internet began two decades ago (for more detailed accounts see additional reading suggestions). Religious authority, community, identity, and authenticity have been major themes within digital religion (see Campbell et al., 2014). The development of the field can be described as flowing through four waves: a) a focus on describing new opportunities online spaces provide for religious practice and the potential impact technology has on religious understanding (H. Campbell, 2013b, p. 8); b) an exploration of technologies and users practicing and engaging religion in online spaces through questions about religious identity, authenticity, and practice (H. Campbell, 2013b, p. 8); c) theoretical and interpretive research on how the “embeddedness” of the internet in everyday life influences religious digital practice” (H. Campbell, 2013b, p. 9-10); and d) a focus on new methodological modes, as well as development of typologies and categorizations for interpretive purposes (Campbell & Lövheim, 2011, p. 1092). Mobile technologies are a focus within this fourth and current wave.

KEY THEMES IN RELIGION AND MOBILE TECHNOLOGY

This section discusses the latest research in three key areas of religion and mobile technology studies: religious communities and individual engagement with mobile technology; the design, development, and use of religious mobile apps; and the commodification of religion related to mobile technology. Each section provides a broad overview of the current and research available pro-

viding important information toward identifying gaps and evaluating promising future directions.

Accepting, Rejecting, and Negotiating

While religious communities have historically evaluated new technologies (e.g., the printing press and electronic media) and their potential impacts on religious life, mobile technology's multiple modalities present a unique case. Access to information, media content, and communication in real time helps meet religious goals or needs, and also causes concern for potential negative outcomes. Generally, religious communities have three responses when evaluating new technology: accept and justify technological affordances as tools to meet religious requirements, reject and frame technological affordances as problematic or evil, or acknowledge both good and bad technological affordances and negotiate rules for proper use (Campbell, 2010).

When religious communities and members advocate for mobile technology, they frame use as a tool for religious communication and practice. Mobile device affordances in this case align with the religious community's values. For instance, the ease and speed of communicating with others, as well as using mobile technology seen as "cool and hip" resulted in a positive outlook for a young group of Javanese Muslims (Barendregt, 2009). Ease and speed of communication were also used as a justification for Pentecostal Christian Church members to embrace mobile use (Togara-sei, 2012). Filipino Christians used mobile SMS texting to communicate religious and inspirational messages, particularly around religious holidays (Roman, 2006). Similarly, mobile affordances allowed Jamaican church members to keep in touch with other members and coordinate church related events (Horst & Miller, 2006). These cases demonstrate how mobile technology affordances were framed as tools to help the community meet both religious goals and the need to stay in touch and coordinate religious meetings and rituals.

Religious leaders and members frame inappropriate mobile technology as harmful or even potentially evil. Access to pornography through mobile internet is of particular concern to many religious groups (Barendregt, 2009; Roman, 2006). In these instances, mobile affordances do not help meet religious objectives and undermine core religious values and beliefs (Campbell, 2010). A backlash against mobile use within the Muslim community pointed to the tension between religious beliefs and practices and unwanted shifts in cultural values. When this happens "...members are more likely to unite and speak out against the technology (Campbell, 2006b, p. 142).

Generally, religious communities discuss both positive and negative mobile affordances and negotiate rules to meet religious needs. The negotiation process leads communities to engage technology only in certain ways and contexts. For instance, using "correct" religious phrases to begin and end a text message is seen as proper etiquette within certain groups (Barendregt, 2009). Additionally, community members often negotiate their personal use of mobile technology outside the recommended rules given by religious authorities. Cohen et al. (2008) found young Haredi Jews maintained personal rules for mobile use in private, but adhered to the religious communities' rules of use in public (Cohen, Lemish, & Schejter, 2008, p. 98). Richardson and Pardun's (2015) study showed rules are created for specific religious contexts. Their survey of highly religious Baptist users in the southern US revealed individuals' excitement and confidence in their personal religious use of mobile technology, but concern for the possible distraction mobile devices cause in public worship.

While religious users in previous examples negotiated proper use of the technology for specific contexts, other communities decide to reshape the technology as whole to meet their religious needs. Campbell's (2010) religious social shaping of technology (RSST) framework allows researchers to explore this by focusing on a given group's history and traditions, as well as their core values and patterns. These factors played a role,

for example, in Ultra-Orthodox Jews' negotiation of the mobile phone. As originally designed, the mobile phone conflicted with Ultra-Orthodox Jews' unique social boundaries and needs, because it allowed access to problematic content through internet and texts. The reshaping of the mobile phone as a "kosher" phone blocked this content and was framed by religious leaders as a righteous symbol against secular, immoral forces. This reframing of the technology to fit within a more traditional framework made it acceptable to religious authorities.

Acceptance, rejection, and negotiation with specific mobile technologies depend upon their use, context, and alignment with religious communities' unique values and goals. Therefore, a group's history of media use, religious values, and patterns of practice play a vital role in their negotiated use of technology. Therefore, RSST offers a useful approach for scholars studying these phenomena.

Religious Mobile Apps

Religious mobile apps, defined as small software applications designed for use on mobile devices specifically for religious purposes, are the fastest growing research foci within digital religion. Research in this area falls into two categories: religious app design and development, and religious app design and use. Design and development studies generally use textual and content analysis of the apps themselves. Only one study within digital religion collected data directly from developers (Hutchings, 2014). Most studies within the last two years have focused on user intentions and motivations and user-experience design research used to test individual religious apps.

- **Religious App Design and Development:** Studies in this category investigate the apps themselves in order to gain insight about their type and scope, and also to interpret how religious app affordances translate into specific outcomes for users and com-

munities. Wagner (2013) and Campbell, Altenhofen, Bellar, and Cho (2014) created categories and typologies of religious apps, respectively. Wagner (2013) highlighted six categories of religious apps on iTunes, including prayer, ritual, sacred text, and religious social media apps. Her analysis also focused on how apps impact users' religious identities. She concluded the personal nature of religious apps as far as choice and use allows individuals to create their own understandings apart from the structure and rules of traditional religious authorities.

Wagner's (2013) categories helped lay the groundwork for Campbell, Altenhofen, Bellar, and Cho's (2014) systematic categorization of more than 400 religious apps available on iTunes, which resulted in a typology of 11 diverse categories defining different religious app use (Campbell et al., 2014, p.163). Their analysis resulted in two parent categories: apps oriented around religious practice, which help facilitate traditional religious practices or experiences like prayer, and apps embedded with religious content that provide religious information or materials, but do not necessarily facilitate religious practices or rituals. Subcategories for religious practice apps include sacred textual engagement, prayer, focus/meditation, devotional worship, and ritual. Secular app counterparts may be readily identified—e.g., music players that provide both religious and secular content. Subcategories for religious content apps included religious utilities, religious wisdom and leaders, religious media, religious games, religious apps for kids, and religious social media.

Other studies on app design and development analyzed how the affordances of the device and apps themselves influence religious practices and understandings. An earlier study by Torma and Teusner in 2011 explored the intersection of cultural and social values associated with the iPhone and religious values and beliefs. Their study narrowed in on affordances of connectivity,

audio-visual, and the private and personal nature found in app user interfaces and content. They concluded that for religious apps to be meaningful to users, developers needed to take advantage of audio and visual affordances to create interactivity and engagement. Wagner and Accardo (2014) similarly noted the personal and individualized nature of Buddhist apps and their connection to the westernization of Dharma. Their approach is reminiscent of technological determinism, which posits that the medium directly affects use: "... when we study Dharma embedded and communicated in the context of portable devices, it can indeed be shaped by its environment" (Wagner & Accardo, 2014, p. 137). However, they note user attitude will also affect religious app use and effectiveness. Other noted effects include mixing the profane with the sacred, and the fact that the isolating and personal nature of apps changes the value Buddhists place on communal practice. While this approach suggests a direct influence of apps on religious practices, current work focuses on how users negotiate and possibly reshape apps to avoid effects that do not align with their religious values.

As mentioned earlier, only one study collected data directly from app developers—Hutchings's (2014) case study of the Bible App developed by LifeChurch.tv, a mega church based in Oklahoma, USA. He concluded that app designers used specific features (i.e., social media and reading plans) to guide users' Bible reading in specific ways. Therefore, app developers and designers are either already traditional religious authorities, or become a type of religious authority through employing or not employing specific affordances in app design. This is in contrast to Wagner's (2013) conclusion that individual users will determine their personal practice outside the purview of religious authorities.

This overview shows scholars have studied religious apps and their affordances to either create app typologies or to link affordances with possible religious outcomes such as practice. Yet it is important to note an app's affordances

do not guarantee individuals will use the app in intended ways.

- **Religious App Design And Use:** Other researchers focus on religious app users in conjunction with, or in addition to, the apps themselves. One of the earliest apps studies designed an app for use in Muslim prayer and evaluated user engagement (Wyche, Caine, Davison, Patel, Arteaga, & Grinter, 2009). Similar to Torma and Teusner's (2011) conclusion that audio-visual cues are important for meaning-making in religious apps, Wyche et al. found Muslim users embraced the app because of its use of imagery in the app design (i.e., nature and the sun, simple pictures of mosques). Participant interviews revealed images—the placement of the sun on the background of the app, and images of mosques—made users feel connected to the larger Muslim community and the analogue understanding of prayer.

Bellar (2016) used the concepts of networked identity and networked community (Campbell, 2012) to analyze how religious app use influenced Evangelical Christian users' identities. Findings showed participants were more likely to choose religious apps that already aligned with their core religious practices, such as reading the Bible. When the app met religious expectations, it reinforced participants' religious identities, but their identities were negatively impacted when those expectations were not met. Second, findings showed participants used their religious networked community to choose religious apps. The private, personal nature of religious app use influenced the ways in which users were comfortable sharing with others.

Hutchings (2015) conducted an online survey of UK and US digital Bible readers and found five positive effects of app use: convenience, easy access, easier to study, engagement with other readers online, and more consistent reading

practices. Negative aspects of app use were: the sacred status of the Bible was diminished, verses and passages were more likely to be read out of context, and there was a “loss of meaningful relationship to the printed Bible” (Hutchings, 2015, p. 15). Similarly, Richardson and Pardun (2015) found that highly religious Baptist app users were concerned with convenience but worried about possible distractions app use in public worship cause. Additionally, their user survey and focus groups revealed a hesitation to replace the physical Bible with digital text, because the printed book was related to important memories and allowed tactile engagement through underlining, highlighting, and taking notes.

Lechebusch, Kowalewshi, Lidynia, and Ziefle (2015) compared Christian and Muslim adoption and use of religious apps in Germany. Their focus groups and survey data showed that Muslims were more likely than Christians to use religious apps. Both groups were interested in using apps with access to holiday calendars, religious quizzes, information on proper social app use, and information about religious history. Findings also showed Muslims were more likely than Christians to use digital prayer beads on a mobile app. Additional comparative studies are needed on different religions and regions to gain a more holistic picture of religious app adoption and use.

Finally, user-experience researchers have been involved in designing and testing apps for religious use. These studies highlight the need for more interdisciplinary research in this area. For example, five studies over the last two years used Human and Computer Interaction (HCI) approaches. These have focused on Muslim and some Christian engagement—use of Islamic apps by the elderly (Ahmed, Zainal, Razak, Adnan, & Osman, 2015), using accelerometer functions to monitor correct prayer postures for Muslims (Al-Ghannam, Kanjo, & Al-Dossari, 2016), intentions to use a pilgrimage app for Spain’s Camino de Santiago (Attunes & Amaro, 2016) evaluating appropriate design affordances using Near Field Communication for Muslims’ Hajj pilgrimage

(Mohandes, 2015), and development and use of a religious education app for Indonesian Muslim youth (Siadin, Mohamed, Adzmi, & Azhar, 2015). These studies highlight technological skills used to create engaging affordances for different religious users, and how creating user tests helps evaluate app success. Digital religion scholars would benefit from working with scholars in HCI in order to accurately understand design processes. Conversely, HCI scholars would benefit from collaboration with digital religion scholars in evaluating and understanding how religious groups’ core values and beliefs drive engagement with technology.

RELIGIOUS MEANING-MAKING IN MOBILE CULTURE

A third theme in mobile media and religion research has focused on the intersection of religion and consumer culture, or how people connect notions of religion and spirituality to mobile phone technology and culture. Fears of trivializing the sacred or falling into idolatry can cause a potential backlash against the commodification of religion (Barendregt, 2009; Togarasei, 2012). Consumer culture and religion are therefore often seen as antagonistic. Wagner & Accardo (2014) also noted the potential for downloads and advertisements to alter individuals’ perception of religious practice within capitalistic culture.

Another way these areas intersect is through the appropriation of religion into consumer culture, or selling products using religious elements. Robinson (2013) argues there is an intimate relationship between religion and technology. Technology companies such as Apple, for example, can become purveyors of spiritual ideals through design and marketing. Analysis of Apple advertisements showed that Steve Jobs’ spiritual-religious experiences informed his technological philosophy and subsequent approach to business and the design of Apple products, so that Apple technology is imbued with implicit religion.

It is not only developers who propagate religious devotion and meaning in their products. Users of Apple's Newton, a personal digital assistant (PDA) device that predated smartphones, rallied around the technology by creating a device narrative containing supernatural, religious, and magical themes (Muniz & Schau, 2005). Researchers assert that creating religious narratives around specific technologies points towards our impulse to believe in something outside of mundane reality, or, "it points to the inevitable intermingling of God and goods" (Muniz & Schau, 2005, p. 746).

Consumers, developers, and news media can co-create a cultural and religious symbolism that can build meaning around a particular technology. A good illustration is Gizmodo blogger, Brian Lam, who coined the term "Jesus Phone" in reference to the iPhone before the device even launched as way to frame it as revolutionary and even divinely inspired. Mac fans, critics, news media, and even Apple itself adopted this term to hype and critique the transformative power attributed to the device. Campbell & La Pastina (2007) assert consumers, bloggers, and critics appropriated the reference to Jesus to describe the transcendent nature of the iPhone that helped propel it into prominence in consumer culture. The Jesus Phone case study exemplifies the way technology like mobile phones can be imbued with greater cultural and religious meaning.

Mobile media, religion, and consumer culture intersect through the appropriation of religious elements by consumers and developers alike. Religious elements, such as references to Jesus or allusions to sacred narratives or themes, can act as a conduit for product creation or meaning-making. Because religious symbolism can be a deeply familiar resource from which to draw, mobile products can be imbued with implicitly sacred symbols, a fact that reveals how easily they can be mixed with religion in consumer culture.

EMERGING TRENDS AND FUTURE RESEARCH



Mobile media use by religious individuals and communities is a developing area within digital religion studies. This article highlights three primary ways scholars have approached the study of religion and mobile media: by looking at religious groups' and individuals' mobile phone use, by tracking religious app design and use, and by examining how the commodification of religion through mobile devices can manifest popular beliefs about religion. More research is needed in these areas to fully understand how mobile phones are impacting religious communities and users. Our observations lead us to note a number of important areas worthy of future research.

First, more refined theoretical understandings and methodologies are needed to illuminate religious engagement of mobile technology. Scholars can meet this need by applying theories and methods from communication studies, new media studies, HCI, and religious studies to mobile media and religion specifically, by developing and modifying theories based on research. Second, a more comprehensive overview of religious app development and use will aid further app development and act as a framework of study for future researchers. This includes the need to focus more on developers specifically, conduct more comparative studies of major and minor religions, and engage in interdisciplinary collaboration among researchers from media studies, HCI, and religious studies. Third, examining the appropriation of technology to reinforce or undermine traditional religious authority and structures will lead to more definitive answers about the impact mobile technology is having on religion. In this area it is important to examine access to information outside of traditional religious authority. Fourth, a further synthesis of the study of mobile media with broader religious activities will yield a more comprehensive overview of the use of mobile

media for religious purposes. Helpful starting points are key themes within the current study of digital religion work: ritual, identity, community, authority, authenticity, and religion (outlined in Campbell, 2013).

CONCLUSION

Investigating the relationship between religion and mobile media is an emerging area of study within digital religion. Research has yielded insight into how religious individuals and communities respond to mobile media, including what they see as positive and negative uses, and how/if groups create specific guidelines and rules for use; some preliminary pictures of the religious mobile landscape; and the connection between religion and consumer culture through mobile media. Because religion interacts with many social, political, economic, technological, and global factors, research demands continued theoretical development and nuanced study.

REFERENCES

- 2 billion consumers worldwide to get smart (phones) by 2016. (2014). *eMarketer*. Retrieved from: <http://www.emarketer.com/Article/2-Billion-Consumers-Worldwide-Smartphones-by-2016/1011694>
- Ahmad, N. A., Zainal, A., Razak, F. H. A., Adnan, W. A. W., & Osman, S. (2015). User experience evaluation of mobile spiritual applications for older people: An interview and observation study. *Journal of Theoretical & Applied Information Technology*, 72(1).
- Al-Ghannam, R., Kanjo, E., & Al-Dossari, H. (2015, October). Monitoring prayer using mobile phone accelerometer. In *International conference on cloud computing* (pp. 168-175). Daejeon, South Korea: Springer International Publishing.
- Antunes, A., & Amaro, S. (2016). Pilgrims' acceptance of a mobile app for the Camino de Santiago. In *Information and communication technologies in tourism 2016* (pp. 509-521). Bilbao, Spain: Springer International Publishing.
- Barendregt, B. (2009). Mobile religiosity in Indonesia: Mobilized Islam, Islamized mobility and the potential of Islamic techno nationalism. In E. Alampay (Ed.), *Living the Information Society in Asia* (pp. 73-92). Singapore: Institute of Southeast Asian Studies.
- Barna Group. (2013). How technology is changing millennial faith. *Millennials and Generations*. Retrieved from: <https://www.barna.org/barna-update/millennials/640-how-technology-is-changing-millennial-faith#.V3RVC5MrLow>
- Bellar, W. (2016). Private practice: Using digital diaries and interviews to understand Evangelical Christians choice and use of religious mobile applications. *New Media & Society*. doi:10.1177/1461444816649922
- Campbell, H. (2006a). Religion and the internet. *Communication Research Trends*, 25(1), 3-24.
- Campbell, H. (2006b). Texting the faith: Religious users and cell phone culture. In A. Kavoori & N. Archeaux (Eds.), *The cell phone reader. Essays in social transformation* (pp. 139-154). New York: Peter Lang.
- Campbell, H. (2007). What Hath God Wrought? Considering how religious communities culture (or kosher) the cell phone. *Continuum (Perth)*, 21(2), 191-203. doi:10.1080/10304310701269040
- Campbell, H. (2010). *When religion meets new media*. London: Routledge.
- Campbell, H. (2012). Understanding the relationship between religious practice online and offline in a networked society. *Journal of the American Academy of Religion*, 80(1), 64-93. doi:10.1093/jaarel/lfr074

- Campbell, H. (Ed.). (2013). *Digital religion: Understanding religious practices in new media worlds*. London: Routledge.
- Campbell, H., Altenhofen, B., Bellar, W., & Cho, K. (2014). There's a religious app for that!: A framework for studying religious mobile applications. *Mobile Media & Communication*, 2(2), 154–172. doi:10.1177/2050157914520846
- Campbell, H., & La Pastina, A. (2010). How the iPhone became divine: New media, religion and the intertextual circulation of meaning. *New Media & Society*, 12(7), 1191–1207. doi:10.1177/1461444810362204
- Campbell, H., & Lövheim, M. (2011). Rethinking the online-offline connection in the study of religion online. *Information Communication and Society*, 14(8), 189–207. doi:10.1080/1369118X.2011.597416
- Campbell, S. (2013). Mobile media and communication: A new field, or just a new journal? *Mobile Media & Communication*, 1(1), 8–13. doi:10.1177/2050157912459495
- Cheong, P. (2013). Authority. In H. Campbell (Ed.), *Digital religion* (pp. 72–87). London: Routledge.
- Cohen, A., Lemish, D., & Schejter, A. (2008). *The wonder phone in the land of miracles: Mobile telephony in Israel*. Hampton Press, Inc.
- Hoover, S., Clark, L., & Rainie, L. (2004). *Faith online: 64% of wired Americans have used the internet for spiritual or religious information*. Pew Internet and American Life Project. Retrieved from <http://www.pewInternet.org/reports/toc.asp?Report=119>
- Hoover, S., & Echchaibi, N. (2012). *The 'third spaces' of digital religion*. Paper presented at the Center for Media, Religion, and Culture, University of Colorado Boulder. Retrieved from <http://cmrc.colorado.edu/wp-content/uploads/2012/03/Third-Spaces-Essay-Draft-Final.pdf>
- Horst, H., & Miller, D. (2006). *The cell phone: An anthropology of communication*. Oxford, UK: Berg.
- Hutchings, T. (2014). Now the Bible Is an app 9. In K. Granholm, M. Moberg, & S. Sjö (Eds.), *Religion, media, and social change*. New York: Routledge.
- Hutchings, T. (2015). *E-reading and the Christian Bible*. Studies in Religion/Sciences Religieuses. doi:10.1177/0008429815610607
- Leckebusch, J., Kowalewski, S., Lidynia, C., & Ziefle, M. (2015, August). "Faith to go or devil's work"—Social media acceptance in taboo-related usage contexts. In *International Conference on Cross-Cultural Design* (pp. 34–45). Daejeon, South Korea: Springer International Publishing.
- Mohandes, M. A. (2015). Mobile technology for socio-religious events: A case study of NFC technology. *IEEE Technology and Society Magazine*, 34(1), 73–79. doi:10.1109/MTS.2015.2396122
- Muniz, A. Jr., & Schau, H. (2005). Religiosity in the abandoned Apple Newton brand community. *The Journal of Consumer Research*, 31(4), 737–747. doi:10.1086/426607
- Neilson. (2015). Tops of 2015: Digital. *Media and Entertainment*. Retrieved from <http://www.nielsen.com/us/en/insights/news/2015/tops-of-2015-digital.html>
- Richardson, K. B., & Pardun, C. J. (2015). The new scroll digital devices, Bible study and worship. *Journal of Media and Religion*, 14(1), 16–28. doi:10.1080/15348423.2015.1011984
- Robinson, B. (2013). *Appletopia: Media technology and the religious imagination of Steve Jobs*. Waco, TX: Baylor University Press.
- Roman, A. (2006). *Texting God: SMS & religion in the Philippines*. Paper presented at the 5th International Conference on Media, Religion and Culture, Stockholm, Sweden. Retrieved from <http://www.freinademetzcenter.org/pdf/Texting%20God%20SMS%20and%20Religion%20in%20the%20Philippines.pdf>

Saidin, A. Z., Mohamed, K. S., Adzmi, Z. H., & Azhar, N. W. (2015). Q-ibadah mobile application: A usability pilot testing. *Jurnal Teknologi*, 77(29).

Smith, R. (2013). So what? Why study mobile media and communication? *Mobile Media & Communication*, 1(1), 38–41. doi:10.1177/2050157912459498

Togarasei, L. (2012). Mediating the Gospel: Pentecostal Christianity and media technology in Botswana and Zimbabwe. *Journal of Contemporary Religion*, 27(2), 257–274. doi:10.1080/13537903.2012.675740

Torma, R., & Teusner, P. (2011). iReligion. *Studies in World Christianity*, 17(2), 137–155. doi:10.3366/swc.2011.0017

Wagner, R. (2013). You are what you install: Religious authenticity and identity in mobile apps. In H. Campbell (Ed.), *Digital religion: Understanding religious practices in new media worlds* (pp. 199–206). London: Routledge.

Wagner, R., & Accardo, C. (2014). 8 Buddhist apps. In G. P. Grieve & D. Veidlinger (Eds.), *Buddhism, the internet, and digital media: The Pixel in the Lotus*. Academic Press.

Wyche, S. P., Caine, K. E., Davison, B. K., Patel, S. N., Arteaga, M., & Grinter, R. E. (2009, April). Sacred imagery in techno-spiritual design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 55–58). ACM.

ADDITIONAL READING

Cheong, P., Fischer-Nielsen, P., Gelfgren, S., & Ess, C. (Eds.), *Digital religion, social media and culture*. New York: Peter Lang.

Cho, K. (2011). New media and religion: Observations on research. *Communication Research Trends*, 30(1), 4–22.

Helland, C. (2000). Online-religion/religion-online and virtual communities. In D. Cowan & J. Hadden (Eds.), *Religion on the internet. Research prospects and promises* (pp. 205–224). Amsterdam, London, New York: JAI Press.

Hojsgaard, M., & Warburg, M. (Eds.). (2005). *Religion and cyberspace*. London: Routledge.

Hoover, S. (2006). *Religion in the media age*. New York: Routledge.

Jensen, K. B. (2013). What's mobile in mobile communication? *Mobile Media & Communication*, 1(1), 26–31. doi:10.1177/2050157912459493

Lundby, K. (2013). Theoretical frameworks for approaching religion and new media. In H. Campbell (Ed.), *Digital Religion* (pp. 225–237). London: Routledge.

KEY TERMS AND DEFINITIONS

App Developers: Individuals, religious groups, and secular companies who design and produce mobile apps.

Apps: Small computer programs designed for use on mobile devices.

Commodification: Transferring religious information or practices into commodities to be bought and sold, and using religious narratives to sell secular goods and services.

Digital Religion: The study of the evolution of religious beliefs and practices at the intersections of online, mobile, and offline contexts.

Mobile Devices: Cell phones, smartphones, and tablets that provide the infrastructure to communicate and access content anytime, anywhere.

Religious Communities: A group of people who adhere to specific beliefs and practices that relate to the sacred.

User-Experience Design: A research approach that explores the interaction between design and use with the goal of increasing usability, accessibility, and pleasure.

Methods for Simultaneous Improvement of Comb Pass Band and Folding Bands

M

Gordana Jovanovic Dolecek
Institute INAOE Puebla, Mexico

INTRODUCTION

Decimation is a reduction of a sampling rate in a digital form by an integer, called decimation factor. It consists of two stages: filtering and down sampling. Practically the sampling rate is decreased only in the second stage, called down sampling (Harris, 2004). However, during this process, the replicas of the main signal spectrum occur. If the signal is not properly filtered before down sampling, the overlapping of the repeated replicas of the original spectrum, called aliasing, may occur. The aliasing may destroy the useful information of the decimated signal and must be eliminated by the filter which proceeds the down sampling, called decimation filter.

The most simple decimation filter is a comb filter which has all coefficients equal to unity, and consequently does not require multipliers, (Hogenauer, 1981). The transfer function of comb is given as:

$$H(z) = \left[\frac{1}{M} \sum_{k=0}^{M-1} z^{-k} \right]^K, \quad (1)$$

where M is the decimation factor, and K is order of the filter.

Note that all filter coefficients are equal to unity. The comb transfer function (1) is also known as non recursive form. An equivalent transfer function, also known as a recursive form, is given as:

$$H(z) = \left[\frac{1}{M} \frac{1 - z^{-M}}{1 - z^{-1}} \right]^K. \quad (2)$$

From the recursive form (2), a simple comb structure, called CIC (Cascaded-Integrator-Comb) filter, proposed by (Hogenauer, 1981), is obtained. It consists of the cascade of K integrators $1 / (1 - z^{-1})$ and K combs, $(1 - z^{-1})$ separated by a down sampler by M .

To preserve the decimated signal, the comb magnitude characteristic:

$$H(e^{j\omega}) = \left[\frac{1}{M} \frac{\sin(\omega M / 2)}{\sin(\omega / 2)} \right]^K, \quad (3)$$

must have a flat magnitude response in the pass band of interest, and high attenuation across the bands around the comb zeros, also called folding bands. However, comb magnitude characteristic exhibits a low attenuation in folding bands and a pass band droop in the pass band of interest, which may deteriorate the decimated signal. Increasing the order K , the attenuation in folding bands is improved, as shown in Figure 1, for $M=9$ and $K=1, 2, 3, 4$, and 5 .

However, the pass band droop is increased as illustrated in Figure 2, for the same parameters as in Figure 1.

Different methods are proposed for decreasing comb pass band droop, increasing the attenuation in the comb folding bands, and for both: simultaneous decreasing of comb pass band droop, and increasing folding band attenuation. The goal of this article is to present the principal approaches for the comb simultaneous pass band and stop band improvements.

Figure 1. Overall comb magnitude responses for $M=9$ and different values of K

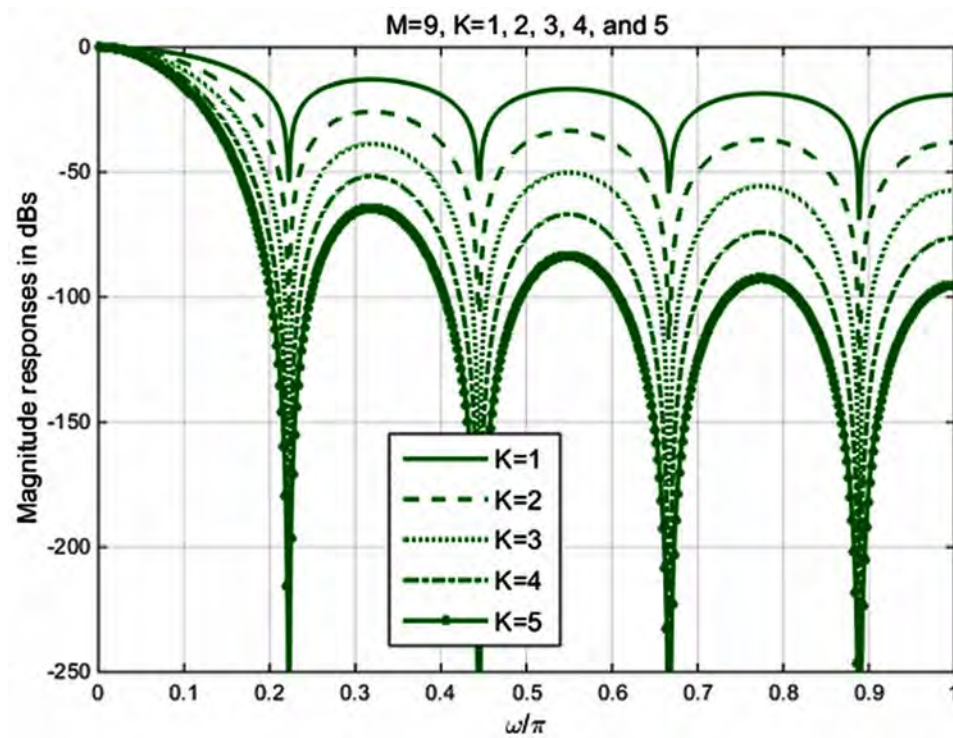
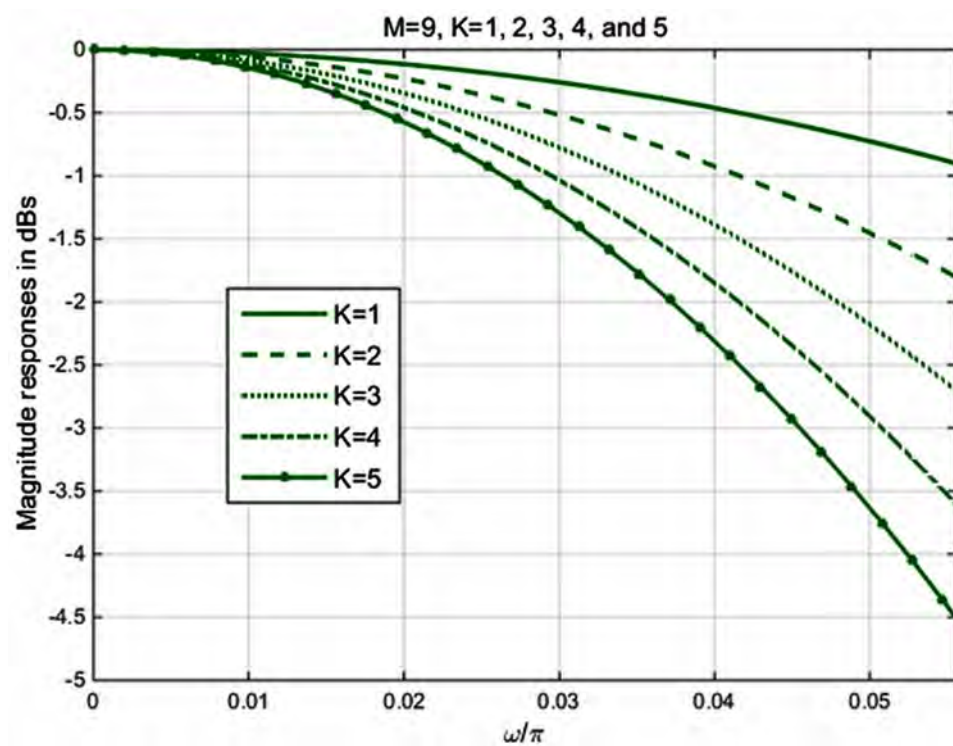


Figure 2. Pass band comb magnitude responses for $M=9$ and different values of K



BACKGROUND

First, the comb pass band and folding bands are defined. Comb pass band is determined by the pass band edge ω_p , which depends on the comb decimation factor M and the decimation factor R of the stage which follows the comb decimation stage (Kwentus & Willson, 1997):

$$\omega_p = \frac{\pi}{MR}. \quad (4)$$

To this end, the comb pass band is defined for frequencies ω :

$$0 \leq \omega \leq \omega_p. \quad (5)$$

Depending on the value R , the comb pass band can be wide band ($R \leq 4$), and narrow band ($R > 4$). The zeros of comb are defined at:

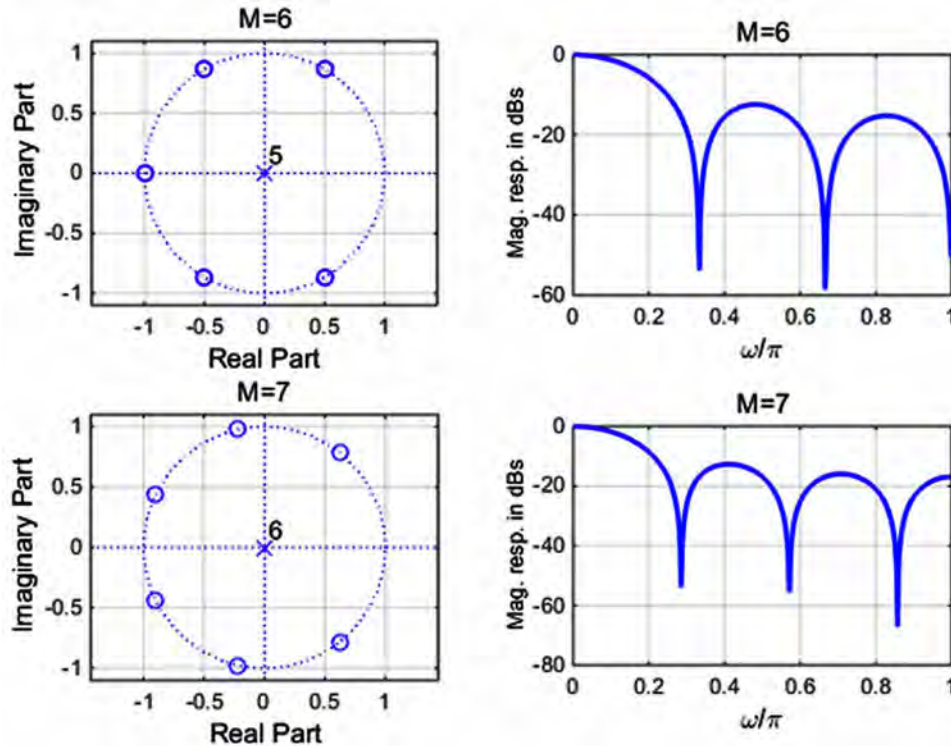
$$\frac{2\pi k}{M}, \quad k = \begin{cases} 1, \dots, M/2 & \text{for } M(\text{even}) \\ 1, \dots, (M-1)/2 & \text{for } M(\text{odd}) \end{cases} \quad (6)$$

The comb folding bands are defined as bands around comb zeros:

$$\frac{2\pi k}{M} - \omega_p \leq \omega \leq \frac{2\pi k}{M} + \omega_p, \quad k = \begin{cases} 1, \dots, M/2 & \text{for } M(\text{even}) \\ 1, \dots, (M-1)/2 & \text{for } M(\text{odd}) \end{cases} \quad (7)$$

As an example Figure 3 shows the comb zeros and the corresponding magnitude responses for values of $M=6$, (M is even), and $M=7$, (M is odd). For $M=6$, there are $6/2=3$ comb zeros for $0 \leq \omega \leq \pi$, and three folding bands around the frequencies: $2\pi/6$, $4\pi/6$, and $6\pi/6 = \pi$. Note that for M is even, there is always a folding band around π . Similarly, for $M=7$, there are three folding bands around the frequencies: $2\pi/7$, $4\pi/7$, and $6\pi/7$.

Figure 3. Comb zeros and magnitude responses for $M=6$ and $M=7$



DESCRIPTION OF METHODS

Classifications

Knowing that the comb filter is a very simple filter, the methods for improving its magnitude responses should also be simple, desirable multiplier less. The methods can be divided into three main groups: Methods based on sharpening; Methods based on simple corrector filters; Methods that combine alias rejection methods and compensator design methods. In continuation are presented all three groups of methods and illustrated with examples.

Sharpening-Based Methods

The filters sharpening technique (Kaiser & Hamming, 1977) can be used for simultaneous improvement of both pass band and stop band characteristics of a linear phase FIR digital filter. The technique uses amplitude change function (ACF). An ACF is a polynomial relationship of the form $H_{sh} = f(H)$ between the amplitudes of the overall, and the prototype filters, H_{sh} and H , respectively. The improvement in the pass band, near $H=1$, or in the stop-band, near $H=0$, depends on the order of tangencies m and n of the ACF at $H=1$ or at $H=0$.

The expressions proposed by Kaiser and Hamming for the m^{th} and n^{th} order tangencies of the ACF at $H=1$ and $H=0$, respectively, are given as,

$$H_0 = H^{n+1} \sum_{k=0}^m \frac{(n+k)!}{n!k!} (1-H)^k, \quad (8)$$

$$= H^{n+1} \sum_{k=0}^m C(n+k, k) (1-H)^k,$$

where $C(n+k, k)$ is the binomial coefficient.

The values of the ACF for some typical values of m and n are given in Table 1, where H_{sh} is transfer function of the sharpened filter and H is the transfer function of the prototype filter.

The polynomial in the first row is a simple cascade of filter with itself, which improves only the filter stop band. Similarly, the polynomial in the second row improves only the pass band. In

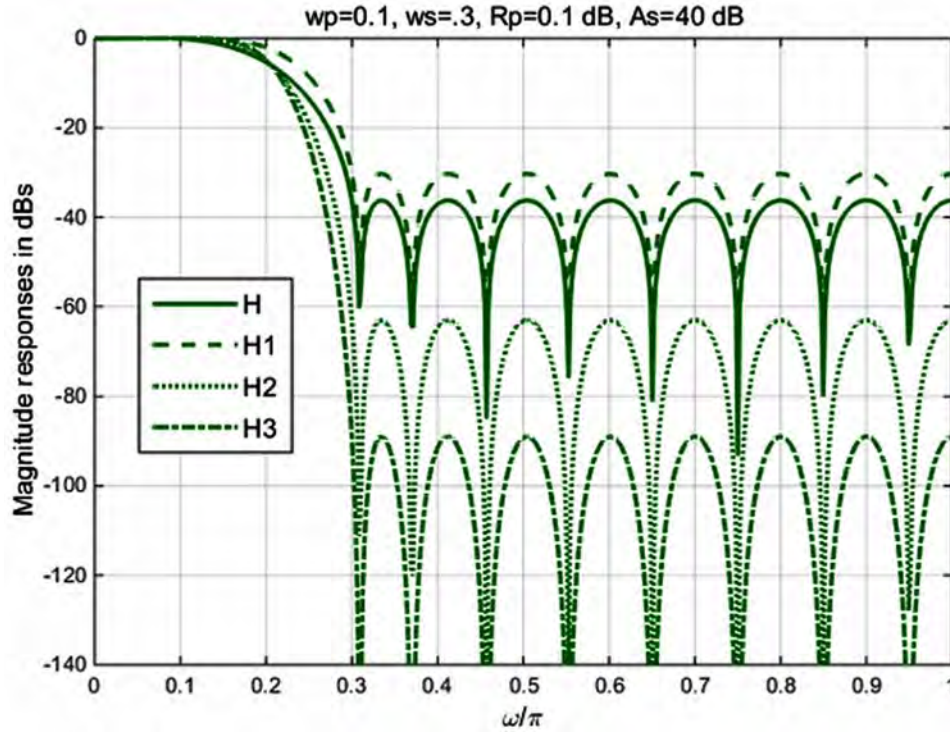
Table 1. ACF polynomials for $m = 1, 2, 3$ and $n=1, 2$, and 3

m	n	H_{sh}
0	1	H^2
1	0	$2H-H^2$
1	1	$3H^2-2H^3$
1	2	$4H^3-3H^4$
1	3	$5H^4-4H^5$
2	1	$3H^4-8H^5+6H^6$
2	2	$6H^5-15H^6+10H^7$
2	3	$10H^6-24H^7+15H^8$
3	1	$-4H^5+15H^6-20H^7+10H^8$
3	2	$-10H^6+36H^7-45H^8+20H^9$
3	3	$-20H^7+70H^8-84H^9+35H^{10}$

third row is presented the simplest polynomial which improves both, pass band and stop band. Increasing the values of m and n the polynomials become more complex. That is why here are presented the polynomials only for values $m, n=1, \dots, 3$.

To illustrate the use of sharpening polynomials, one simple filter is designed by Remez algorithm (Mitra, 2010), taking the following specification: pass band frequency $\omega_p 0.1\pi$; stopband frequency $\omega_s = 0.3\pi$; pass band ripple $R_p = 0.1$ dB, and attenuation in the stop band $A_s = -40$ dB. The designed filter is a prototype filter $H(z)$ and its magnitude response is shown in Figure 4 with the solid line. Next, three sharpening polynomials from Table 1, are applied to the designed filter, and the following sharpened filters are obtained: H_1 (for $m=0, n=1$); H_2 (for $m=1, n=1$); H_3 (for $m=2, n=2$). The corresponding overall magnitude responses are also shown in Figure 4. Similarly, Figure 5 shows the corresponding pass bands for the prototype filter and the sharpened filters H_1 , H_2 , and H_3 .

Note that, for $m=0, n=1$ (filter H_1), only pass band of the prototype filter is improved, while the stop band is deteriorated. Taking $m=n=1$, both pass band and stop band are improved in the sharpened filter H_1 . Similarly, for more complex polynomial, ($m=n=2$) the improvement of both bands in the

Figure 4. Overall magnitude responses for different sharpening polynomials, defined by m and n in Table 1

sharpened filter H_3 is better in comparison with sharpened filter H_2 with $m=n=1$.

Sharpening Comb Method (Kwentus & Willson, 1997)

To improve simultaneously the comb magnitude characteristic in pass band and folding bands, the authors proposed sharpening technique using the simplest sharpening polynomial which improves both the pass band and stop band, i.e. $m=n=1$. The sharpened comb transfer function is given as:

$$H_{sh}(z) = 3H^2(z) - 2H^3(z) = H^2(z)[3z^{-\tau} - 2H(z)], \quad (9)$$

where $H(z)$ is a comb transfer function given in (1) and (2), and τ is the delay which prevents possible nonlinearity of the phase, i.e. 3 must be placed just in the middle coefficient of $-2H(z)$ in (9). Consequently, the filter $2H(z)$ must have the first-type symmetry, i.e. have an odd order,

(Mitra, 2010). Note that sharpened comb works at high input rate, thus consuming a high power.

As an example, consider $M=10$ and $K=2$. The cascade of combs with $M=10$ has the first type symmetry. However, taking $K=3$, the second order symmetry is obtained and the formula (9) cannot be applied.

Knowing that the integers 3 and 2 can be presented as sum of power of two, which can be implemented by shifts and adders, the sharpened comb is a multiplier less filter.

The magnitude response of the sharpened comb is given as:

$$|H_{sh}(e^{j\omega})| = \left| 3 \left[\frac{1}{M} \frac{\sin(\omega M / 2)}{\sin(\omega / 2)} \right]^{2K} - 2 \left[\frac{1}{M} \frac{\sin(\omega M / 2)}{\sin(\omega / 2)} \right]^{3K} \right|. \quad (10)$$

Figure 6 compares the overall, and pass band magnitude responses in a wide band, ($R=2$) of the prototype comb filter ($M=10$, $K=2$), and the sharpened filter.

Figure 5. Pass band magnitude responses for different sharpening polynomials, defined by m and n in Table 1

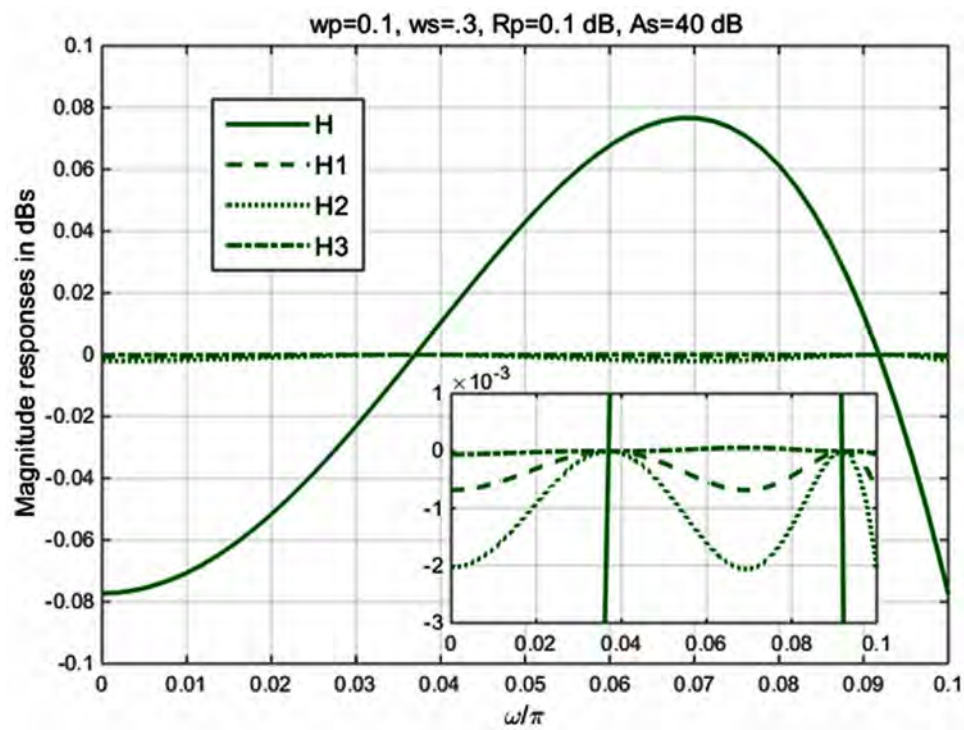
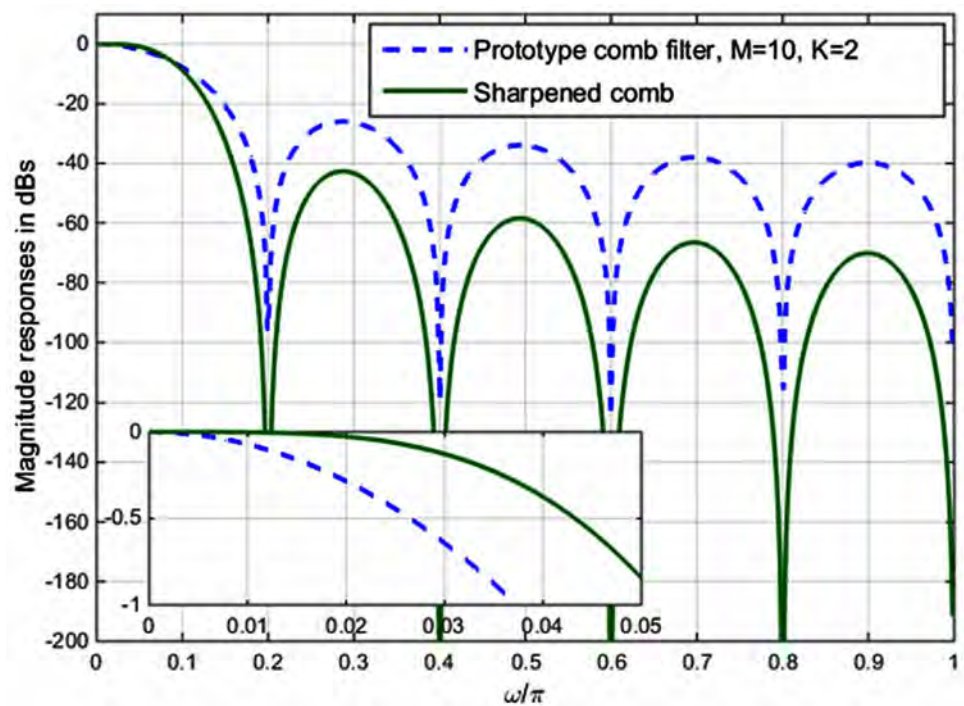


Figure 6. Overall magnitude responses and zoom in pass band magnitude responses for the prototype and sharpened combs



Note that the pass band droop is decreased in narrow band and attenuation in the folding bands is increased.

Different methods are proposed to improve the original sharpening method.

Two-Stage Sharpening (Jovanovic Dolecek & Mitra, 2005)

It is supposed that the decimation factor can be presented as a product of two integers:

$$M = M_1 M_2. \quad (11)$$

The decimation has two stages. At first stage is the comb with the following transfer function:

$$H_1(z) = \left[\frac{1}{M_2} \frac{1 - z^{-M_1}}{1 - z^{-1}} \right]^{K_1}. \quad (12)$$

Similarly, transfer function of the comb at the second stage is:

$$H_2(z) = \left[\frac{1}{M_2} \frac{1 - z^{-M_2}}{1 - z^{-1}} \right]^{K_2} \quad (13)$$

Next, in order to decrease power consumption, the sharpening is applied only on the second stage. The transfer function of the equivalent comb filter at high input rate is:

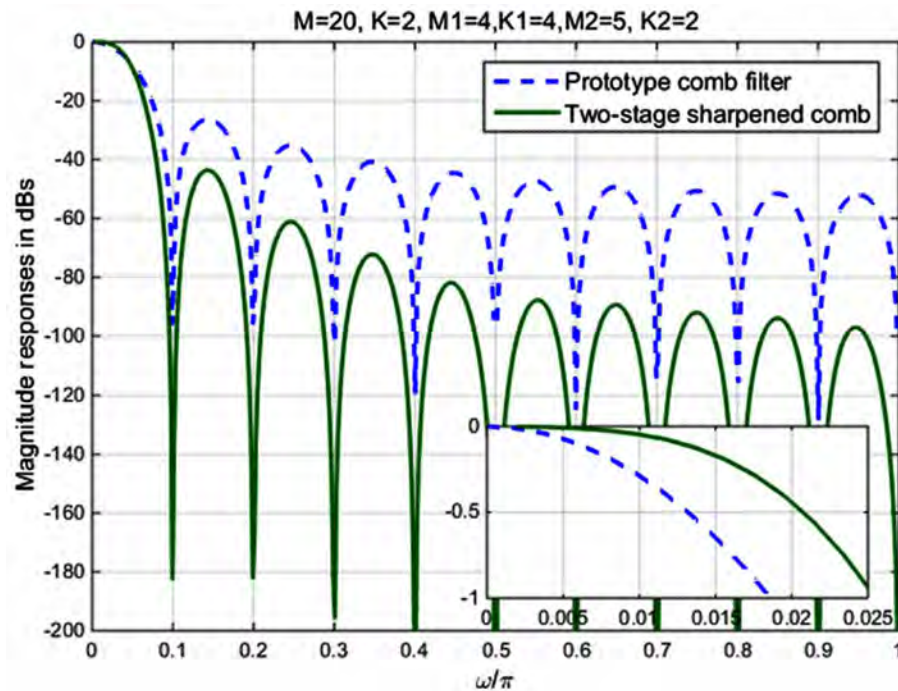
$$H_{sh}(z) = H_1(z) Sh\{H_2(z^{M_1})\}, \quad (14)$$

where $Sh\{\}$ means sharpening.

The corresponding magnitude response is:

$$|H_{sh}(e^{j\omega})| = \left[\frac{1}{M_1} \frac{\sin(\omega M_1 / 2)}{\sin(\omega / 2)} \right]^{K_1} \left\{ 3 \left[\frac{1}{M_2} \frac{\sin(\omega M / 2)}{\sin(\omega M_1 / 2)} \right]^{2K_2} - 2 \left[\frac{1}{M} \frac{\sin(\omega M / 2)}{\sin(\omega M_1 / 2)} \right]^{3K_2} \right\}. \quad (15)$$

Figure 7. Overall magnitude responses and zoom of pass band magnitude response for the prototype and two-stage sharpened combs



As an example consider $M=20$ and $M_1=4$ and $M_2=5$, $K_1=4$, $K_2=2$. Figure 7 shows the overall magnitude responses and pass band zooms ($R=2$) for the prototype comb ($M=20$, $K=2$), and the two-stage sharpened comb.

Different modifications of sharpening technique are proposed in literature with aim to improve magnitude characteristic o decrease complexity, like (Laddomada & Mondin, 2004), (Saravanan & Jayaprakasan, 2013).

Simple Corrector Filters-Based Methods

The methods use the simple corrector filters designed to simultaneously improve both: the pass band and stop band of comb filter.

Five Corrector Filters Method (Jovanovic Dolecek & Fernandez-Vazquez, 2013)

Five multiplier less corrector filters, for values of $K=1, \dots, 5$ are designed with aim to improve comb magnitude characteristic in both: pass band and folding bands. The choice of the corrector filter depends on the comb order K , taking values $K=1, \dots, 5$, and does not depend on the decimation factor. The corrector filter is used for even decimation factors and works at rate which is $M/2$ time lesser than high input rate. As a result, attenuation is improved only in odd folding bands. Additionally, a comb pass band droop is decreased such that the absolute maximum pass band deviation is less than 0.25 dB. The maximum number of adders is 16 for $K=5$, while the minimum is equal to 7, for $K=1$.

As an example, consider $M=12$ and $K=3$. The corrector filter has 12 adders and the following coefficients:

$$g(0) = g(11) = 2^0; g(1) = g(10) = 2^0; g(2) = g(9) = -2^1; g(3) = g(8) = -2^3; g(4) = g(7) = 2^0; g(5) = g(6) = (2^4 + 2^3). \quad (2)$$

$$= g(9) = -2^1; g(3) = g(8) = -2^3; g(4) = g(7) = 2^0;$$

$$g(5) = g(6) = (2^4 + 2^3). \quad (16)$$

The transfer function of the corrector filter is given as:

$$G(z) = \sum_{n=0}^{11} g(n)z^{-n}, \quad (17)$$

where the coefficients $g(n)$ are given in (16).

Figure 8 shows the overall and pass band magnitude responses for comb and corrected comb using the corrector filter (17). Note that the pass band droop is decreased, and the attenuation in folding bands is increased in all odd folding bands.

Methods that Combine Alias Rejection Methods and Compensator Design Methods

There are a number of methods combining the methods for alias rejection with the methods for compensator designs.

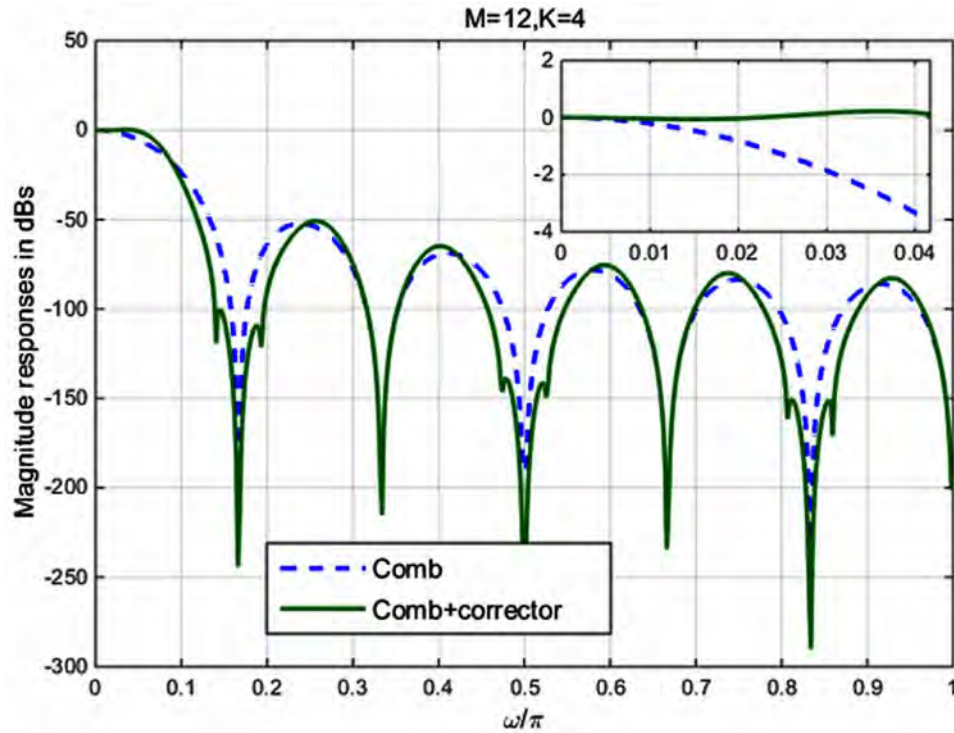
Two-Stage Combs with Narrowband Compensator Method (Jovanovic Dolecek & Mitra, 2010)

Two-stage combs with different orders, and narrowband compensator was proposed in (Jovanovic Dolecek & Mitra, 2010).

The overall transfer function at high input rate is:

$$H(z) = \left[\frac{1}{M_1} \frac{1 - z^{-M_1}}{1 - z^{-1}} \right]^{K_1} \left[\frac{1}{M_2} \frac{1 - z^{-M_2}}{1 - z^{-1}} \right]^{K_2} G(z^M), \quad (18)$$

where M_1, M_2 and K_1, K_2 are the decimation factors and comb orders of the first and second stages,

Figure 8. Overall and pass band magnitude responses for comb and corrected comb, $M=12$, $K=4$ 

respectively, and $G(z)$ is the compensator with the following transfer function:

$$G(z^M) = -2^{-(b+2)}[1 - (2^{b+2} + 2)z^{-M} + z^{-2M}], \quad (19)$$

where b is the compensator parameter which depends on K_1 and K_2 as shown in Table 2.

Additionally, the factor M_1 and M_2 are chosen to be close to each other in values with $M_1 \leq M_2$.

The frequency, in which the worst case attenuation A_c occurs, is defined as:

$$\omega_c = \frac{2\pi}{M} - \frac{\pi}{RM}, \quad (20)$$

where R is the decimation factor which follows the comb filter. Here is used $R=8$. The values for worst case attenuation are also shown in Table 2.

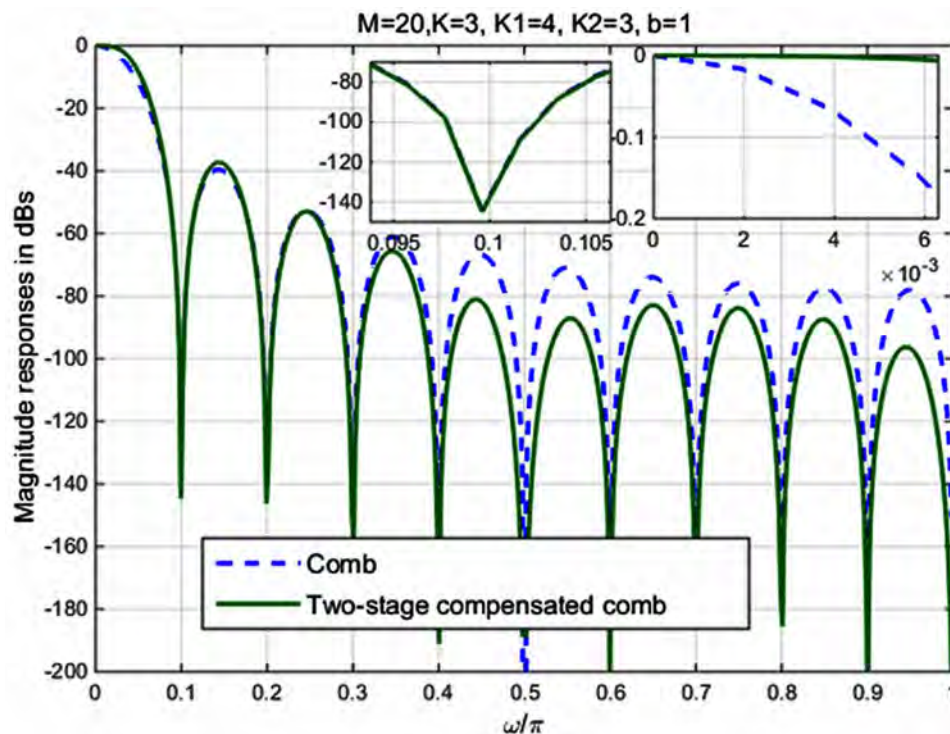
As an example, consider $M=20$ with $M_1=4$ and $M_2=5$. The desired minimum worst case attenuation is $A_c=-70$ dB. From Table 2 the following

parameters are chosen: $K_1=4$, $K_2=3$, $b=1$. Figure 9 shows the magnitude responses of comb, and two-stage compensated combs. The zoom in the first folding band is shown to confirm that the desired minimum worst case aliasing attenuation is satisfied. Similarly, the pass band zoom shows that the pass band droop is decreased in the two-stage comb filter.

Table 2. Parameters of design for given values of K_1 and K_2

(K_1, K_2)	b	Worst Case Attenuation A_c in dB
(2,2)	2	-46.5
(3,2)	1	-68.75
(4,3)	1	-92.25
(5,4)	0	-115
(6,4)	0	-139.34
(7,5)	0	-160
(8,6)	0	-184.186
(9,6)	-1	-205.75

Figure 9. Magnitude responses for comb and two-stage compensated combs



In continuation are mentioned some methods for improving pass band in the modified combs with an increased alias rejection. For example, in (Fernandez Vazquez & Jovanovic Dolecek, 2009) and (Fernandez Vazquez & Jovanovic Dolecek, 2011), are proposed novel methods for compensator design for Generalized comb filter (GCF).

In (Molina Salgado, Jovanovic Dolecek & de la Rosa, 2015) is combined simplest corrector filter from (Jovanovic Dolecek & Fernande Vazquez, 2013) with the sharpening technique. In (Garcia & Jovanovic Dolecek, 2014) the corrector filters from (Jovanovic Dolecek & Fernandez Vazquez, 2013) are used in the recursive comb structure with the decimation factor which is power of two.

Different sharpening techniques are also used in multistage combs, and with corresponding compensators like in: (Gogi & Kulkarni, 2014), (Jayaprakasan & Madheswaran, 2013), (Jeong, Min & Kim, 2011), (Molnar, Pecotic & Vucic

2013), (Nikolic & Lutovac 2011), (Sukhpreet Singh & Panag, 2015), Jovanovic Dolecek, G. (2010), (Jovanovic Dolecek & Mitra, 2003), among others.

SOLUTIONS AND RECOMMENDATIONS

Generally, there is a trade-off between the quality of the improvement and the overall filter complexity. Knowing that comb filter is a simple multiplier less filter, the filters which improve comb magnitude characteristic in the pass band and folding bands, must also be multiplier less filters. The most promising methods are based on simple corrector filters. However, the existing corrector filters require high number of adders. Additionally, attenuation in all folding bands is not increased.

FUTURE RESEARCH DIRECTIONS

Future designs must be dedicated to the increase of the quality of magnitude improvement (low absolute value of pass band deviation and high folding band attenuation) and the decrease of complexity expressed in number of adders. Additionally, special interest will be in a low power and low area designs.

CONCLUSION

This article presents an overview of the methods for simultaneous decreasing of comb pass band droop and increasing alias rejection in the folding bands. Three main groups of methods are considered: Method based on sharpening, Methods based on simple corrector filters, and Methods that combine the methods for improvement of alias rejection and design of compensators that also include the methods combined with sharpening technique. Methods based on sharpening usually use the simplest sharpening polynomials. The compensators designed for comb filters cannot be directly used for the modified comb filters with an improved alias rejection. To this end, either the special methods are developed, or, new compensator parameters are defined.

REFERENCES

- Fernandez-Vazquez, A., & Jovanovic Dolecek, G. (2009). A general method to design GCF compensation filter. *IEEE Transactions on Circuits and Systems. II, Express Briefs*, 56(5), 409–413. doi:10.1109/TCSII.2009.2019337
- Fernandez-Vazquez, A., & Jovanovic Dolecek, G. (2011). An L2 design of GCF compensation Filter. *Signal Processing*, 91(5), 1143–1149. doi:10.1016/j.sigpro.2010.11.001
- Garcia, R., & Jovanovic Dolecek, G. (2014). Application of corrector filters to improve magnitude response of comb decimation filter. In *Proceedings of 57th Int'l Midwest Symposium on Circuits & Systems*, (pp.699-702), College Station, TX: IEEEExplore. doi:10.1109/MWSCAS.2014.6908511
- Gogi, A., & Kulkarni, S. (2014). Performance enhancement of CIC filters. *International Journal of Electrical. Electronics and Data Communication*, 2(6), 73–77.
- Harris, F. J. (2004). *Multirate signal processing for communication system*. Prentice Hall.
- Hogenauer, E. (1981). An economical class of digital filters for decimation and interpolation. *IEEE Transactions Acoustic, Speech and Signal Processing*, (29),156-162.
- Jayaprakasan, V., & Madheswaran, M. (2013). Cascading sharpened CIC and polyphase FIR filter for decimation filter. In *Proceedings 2nd International Conference on Advances in Electrical and Electronics Engineering ICAEE'2013*, (pp. 148-153). Dubai, UAE: IEEEExplore.
- Jeong, C., Min, Y. J., & Kim, S. W. (2011). Double-sharpened decimation filter employing a pre-droop compensator for multistandard wireless applications. *ETRI Journal*, 33(2), 169–175. doi:10.4218/etrij.11.0110.0272
- Jovanovic Dolecek, G. (2010). Low power sharpened comb decimation filter. In *Proceedings of International Conference on Green Circuits and Systems ICGCS 2010* (pp. 226-229). Shanghai, China: IEEEExplore. doi:10.1109/ICGCS.2010.5543064
- Jovanovic Dolecek, G., & Fernandez-Vazquez, A. (2013). Novel droop-compensated comb decimation filter with improved alias rejections. *International Journal for Electronics and Communication*, 67(5), 387–396. doi:10.1016/j.aee.2012.10.010

- Jovanovic Dolecek, G., & Mitra, S. K. (2003). Efficient sharpening of CIC decimation filter. In *Proceedings 2003 International Conference on Acoustics, Speech, and Signal Processing ICASSP 2003* (vol. 4, pp.385-388). Hong Kong, China: IEEEExplore.
- Jovanovic Dolecek, G., & Mitra, S. K. (2005). A new two-stage sharpened comb decimator. *IEEE Transactions on Circuits and Systems. I, Regular Papers*, 52(7), 1416–1420.
- Jovanovic Dolecek, G., & Mitra, S. K. (2008). Simple method for compensation of CIC decimation filter. *Electronics Letters*, 44(9), 1162–1163. doi:10.1049/el:20081603
- Jovanovic Dolecek, G., & Mitra, S. K. (2010). Two-stage CIC-based decimator with improved characteristics. *IET Signal Processing*, 4(1), 22–29. doi:10.1049/iet-spr.2008.0160
- Kaiser, F., & Hamming, R. W. (1977). Sharpening the response of a symmetric nonrecursive filter by multiple use of the same filter. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 25(5), 415–422. doi:10.1109/TASSP.1977.1162980
- Kwentus, A., & Willson, A. Jr. (1997). Application of filter sharpening to cascaded integrator-comb decimation filters. *IEEE Transactions on Signal Processing*, 45(2), 457–467. doi:10.1109/78.554309
- Laddomada, M., & Mondin, M. (2004). Decimation schemes for Sigma-Delta A/D converters based on Kaiser and Hamming sharpened filters. *IEE Proceedings. Vision Image and Signal Processing*, 151(4), 287–296. doi:10.1049/ip-vis:20040477
- Mitra, S. K. (2010). *Digital signal processing*. N.Y: Mc Graw-Hill.
- Molina Salgado, G., Jovanovic Dolecek, G., & de la Rosa, J. M. (2015) Novel two-stage comb decimator with improved frequency characteristic. In *Proceedings Latin American Circuits and Systems Symposium LASCAS 2015*. (pp. 1-4). Montevideo, Uruguay: IEEEExplore. doi:10.1109/LASCAS.2015.7250430
- Molnar, G., Pecotic, M. G., & Vucic, M. (2013). Weighted least-squares design of sharpened CIC filters. In *Proceedings 36th International Convention on Information & Communication Technology Electronics & Microelectronics (MIPRO)*, (pp. 91-95). Opatija, Croatia: IEEEExplore.
- Nikolic, M., & Lutovac, M. (2011). Sharpening of the multistage modified comb filters. *Serbian Journal of Electrical Engineering*, 8(2), 281–291. doi:10.2298/SJEE1103281N
- Saravanan, P. D., & Jayaprakasan, V. (2013). Design and implementation of efficient CIC filter structure for decimation. *International Journal of Computers and Applications*, 65(14), 1–7.
- Sukhpreet Singh, S., & Panag, T. S. (2015). Design of modified three stage sharpened CIC filter for decimation. *International Journal of Innovative Research in Computer and Communication Engineering*, 3(5), 4662–4668.

ADDITIONAL READING

- Baig, M. A., Kumar, M., Mishra, V. K., & Aggarwal, A. (2016) Design of fourth order GCF compensation filter. Intelligent Systems Technologies and Applications (pp.283-291), In *Advances in Intelligent Systems and Computing (AISC)* (384): Springer. N.Y.
- Harris, F. J. (2004). *Multirate signal processing for communications systems*. New Jersey, USA: Prentice Hall.

Jovanovic Dolecek, G. (Ed.). (2001). *Multirate systems: Design and applications*. Hershey, PA, USA: IGI Global.

Lyons, R. G. (2010). *Understanding digital signal processing* (3rd ed.). Prentice Hall.

Milic, L. (2009). *Multirate filtering for digital signal processing: MATLAB Applications*. Hershey, PA, USA: IGI Global. doi:10.4018/978-1-60566-178-0

Molnar, G., Dudarin, A., & Vucic, M. (2016). Minimax design of multiplierless sharpened CIC filters based on interval analysis. In *Proceedings 39th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2016*. (pp.94-98). Opatija, Croatia: IEEEExplore. doi:10.1109/MIPRO.2016.7522118

Stosic, B. P., & Pavlovic, V. D. (2016). Design of new selective CIC filter functions with passband-droop compensation. *Electronics Letters*, 52(2), 115–117. doi:10.1049/el.2015.3333

Stosic, B. P., Pavlovic, V. D., & Milic, D. N. (2015). Novel selective CIC functions with improved magnitude response characteristic. In *Proceedings 23rd Telecommunications forum TELFOR 2015*, (pp.407-410). Belgrade, Serbia: IEEEExplore. doi:10.1109/TELFOR.2015.7377494

KEY TERMS AND DEFINITIONS

Comb Corrector Filters: The simple filters used to correct pass band droop and to increase attenuation in the folding bands of the comb filter.

Comb Filter: The filter with all coefficients equal unity. Usually used as a decimation filter at first stage of decimation.

Comb Folding Bands: The frequency bands of the comb filter where aliasing occurs, and consequently must have high attenuation to attenuate aliasing.

Comb Order: The number of cascaded combs.

Pass Band Droop: Ideally, filter passes all frequencies without attenuation, in the pass band. Consequently, the magnitude response in the pass band must be equal to unity. If values of magnitude characteristic are less than unity in the pass band, then there is a pass band droop.

Pass Band Edge: The frequency which defines the pass band.

Sharpening: Technique used for simultaneous improvement of pass band and stop band of a linear phase finite impulse response filter. The technique uses a polynomial relationship between the amplitudes of the sharpened and the prototype filters, H_{sh} and H , respectively. The improvement in the pass-band, and the stop-band, of the prototype filter depends on the order of tangencies m and n of the polynomial relationship at $H=1$ or at $H=0$, respectively.

Worst Case Aliasing: The aliasing in the left edge of the first folding band.

Micro to Macro Social Connectedness Through Mobile Phone Engagement

Dominic Mentor

Columbia University, USA

INTRODUCTION

This chapter concentrates on the relationship between mobile phone use and social connectedness, emphasizing the various ways in which people use mobile communication as a means to cultivate and maintain social connectedness. The chapter explores the possible connection and relevance of social connectedness through mobile communication on a personal, group and macro level. Individuals are constantly shifting their roles and identities as they use their mobile phones; from personal to group, social to political, philanthropic, and civic participation as they traverse these different spaces. Included is a review of studies on how various mobile engagements, like mobile messaging, voice, video, social media and texting, is used by individuals. All of which has been repurposed within some formal and informal organizations for a variety of service objectives, including mobile learning with tacit or unintentional social connectedness embedded. The nurturing of social connectedness through mobile engagement is essential in people's lives as it reveals their ongoing and affective exchanges and experiences with a single friend, family member(s), an event, cause, artifact or a group. The cultivation of social connectedness can be used as a proxy for socio-emotional goals that can feed strong perceptions or real associations for people of all ages. The conceivable positive influence of social connectedness through multimedia or textual mobile messaging could be used to create a sense of affiliation and has a perceived usefulness for personal, civic, as well as academic and work life engagement.

BACKGROUND

The advent in 2001 and maturation standard of Multimedia Messaging Services (MMS) in 2002, extended text messaging to include, photos, animations (gifs), audio and /or video clips (MMS London, 2016). WAP Mobile communication services like Mixit, WhatsApp, and Kakao, broke the mobile messaging hold from mobile service providers, offering agnostic, and globally free access to one's contacts and Social Media Network Services (SMNS) communities. With features that encompasses voice notes, synchronous voice and video calling, to time set message deletion as with Snap Chat, all of which presents the ability to communicate with others and can serve as a powerful means to foster a sense of connectedness. Lam, (2012), and Ling (2015) as well as Cumiskey & Ling (2015) offers examples of how the mobile phone hosts a way for improving team attitude, playful identity, and social cohesion. While other multimodal messages are sent during a live or televised event, cultivating co-presence (Cui, 2016; Schroeder, 2010). The agnostic mobile messaging services have become the communication method of choice for building, as well as maintaining social connectedness, but also aids how individuals and groups traverse the different spaces of their lives.

Various definitions are touted for the concept of social connectedness that range from defining it as a sense of belonging, affiliations or associations (Chayko, 2007; Flomenbaum, 2008; Visser, Dadlani, van Bel, and Yarosh, 2010). Wei and Lo (2006) cites Teixeira (1992) as defining social connectedness as maintaining "interpersonal, community, and general social ties" (p. 62). While

the Jamieson report (2007) refers to social connectedness as how people come together, interact, and network, indicating electronic communication as an all-important factor in establishing and maintaining social connectedness. Berkman, Glass, Brissette, and Seeman (2000) stated that it is widely recognized that social relationships and affiliation have powerful effects on physical and mental health (p. 843), while Wyn & Cuervo et al., (2005) saw “the emerging technologies as enhancing social connectedness through the internet by means of web-blogs, social websites such as MySpace, and through mobile phone technologies via Short Message Systems (SMS)” (Martino, 2007, pp. 6-7). Chayko (2007) not only provides an overview of technologically generated communities, but also discusses the sociological implications of mobile technology as it pertains to a rise in personal availability and, more importantly, societal cohesion (p. 373).

Register and Herman views the concept of social connectedness as forming the basis for all human existence and similarly to the 2007 New Zealand report, indicates the phenomenon of connectedness that brings quality of life (Register & Herman, 2010, p.53). Whereas Lee, Draper and Lee sees social connectedness as “an attribute of the self that reflects cognitions of enduring interpersonal closeness with the social world” (2001). Other researchers asked, if social media or video calling is the new face-to-face (F2F) or if social connectedness can be derived from SMNS like Facebook or other online interactions and be linked to overall well-being (Grieve, Indian, Witteveen, Tolan, & Marrington, 2013; Lundy, & Drouin, 2016). Additionally, Krastel, Basselier, and Ramaprasad, (2015) found an important use and role of social features in online music sites, mostly accessed via mobile, as positively enabling social connectedness. Research studies from medical health to educational environments, and psychology to public health, all point to the importance and benefits of social connectedness (Berkman, Glass, Brissette, & Seeman 2000; Richards, 2016). The medical field point out faster

recovery rates from patients who have strong social connectedness ties (Cheung, Sedikides, & Wildschut, 2016), while numerous cases of educational benefits, from Henderson & Guy, 2016; Naismith, Lonsdale, Vavoula, & Sharples, 2004; Nolan, Hendricks, & Towell, 2016; Yoo, Miyamoto, & Ryff, 2016 have also proven as successful for enhancing teaming and success in an academic context (Lam, 2012; Walton & Cohen, 2011). The latter examples signify the many ways social connectedness can be defined, or leveraged in different contexts.

SOCIAL CONNECTEDNESS: MOBILE COMMERCE, GAMING, AND HEALTH

Issues, Controversies, Problems

Even though a unanimous definition for social connectedness does not necessarily exist, delineations centralize around a sense of belonging (Richards, 2016; Satici, Uysal, & Deniz, 2016) in a virtual, real or abstract sense with either people or an artifact. Controversially, virtual connectedness has become the new face to face. Especially from mobile phone video. Whereas mobile messaging, including SMS/texting, as well as social media updates through mobile has increased dramatically. According to Dag Kittlaus, co-creator of Siri, SMNS and mobile messaging dominated people’s attention and time on their phones (2016). The cross-platform mobile phone messaging services like Snapchat, WeChat, Facebook Messenger, Line, Kik, Twitter or Viber, has overtaken texting and phone calls, and have gained a prominent place in the social and communication spheres (Statistica, 2014). These apps have similar features of being able to share texts, images, video and other multimedia artifacts. SMNS like Facebook, Google Hangout, or the Chinese WeChat, Weibo, RenRen or QQ have all been optimized for mobile access. WeChat has also been described as more than a portal for performing mobile commercial transactions,

ordering taxi services, booking and delivering geo-targeted cinema tickets and coupons, processing person-to-person money transfers, and the ability to make utility payments or even accessing city services (Chan, 2015, Russell, 2015).

Whereas Hansen, Shneiderman, and Smith in 2011 contended email as the lifeblood of communication, professionals from medical, educational or commercial arenas, have drawn their attention to the rise in mobile messaging, accessing information via SMNS as well as other services via mobile phones (Lee & Phang, 2015; Moreno-Munoz, Bellido-Outeirino, Siano, & Gomez-Nieto, 2016; Wu, 2016). While the proclivity to mobile messaging is captured in the aforementioned studies, mobile communication can also nurture a connectedness or a sense of affiliation. Similarly, Chris Messina credited with initiating the creator of Twitter's hashtag (#), indicates that he "can find a community of 100,000 people online, whereas a couple of years ago I would have felt like I was the only one." There is also a daily practice of routine maintenance of connectedness (Al-Failakawi, 2006; Chilwa, 2008; Deng, 2013; Horwitz, 2014; Mbah & Ogbonna, 2012) through updates, 'snaps' or snap stories', postings, 'likes', or 'pokes', with SMNS and Mobile messaging services recording more than 60 billion messages a day as compared to 20 billion SMS text messages (Goode, 2016).

As with online social media, or mobile gaming, mobile phone communication provides the user with a virtual presence and on-the-go accessibility. Pokemon Go blends the virtual and physical realm through low level, but effective augmented reality. Players are tasked to walk to landmarks so as to earn points or get to Pokemons. The virtual presence holds true for one-on-one mobile gaming like Words with Friends, or multiplayer cross platform games like Spaceteam, Minecraft: Pocket Edition, Real Racing 3, or Super Stickman Golf. In short, the benefits of mobile phone engagement, offer a new, interpersonal means to stay involved with others (Crang, & Mohamad, 2016; Freeman, Bardzell, & Bardzell, 2016).

Various studies (Kim, 2016; Kim, Wang, & Oh, 2016; Longhurst, 2016) indicate that an individual's identity can be made up of self-regulated mobile interactions. In these diverse ways, mobile communication aids *agency* in managing social networks, and grants instant access to users' personal or organizational affiliations, while strengthening bonds and community spirit (Konok, Gigler, Bereczky & Miklósi, 2016). With mobile messaging apps and SMNS'es, the mobile phone contacts' list has grown and expanded beyond the initial family and friendship circle to include ephemeral contacts, businesses or organizations.

Organizational mobile engagement is situated within a different domain of practice from individual messaging, like businesses, public sector organizations or health and education entities. This type of engagement also includes mobile communication in many forms beyond organizational promotions. Numerous sectors of society have shown a growing interest in migrating mobile phone engagement habits to their or the clients' advantage. In the health sector, appointment reminders or sexual health texting (Rana, van den Berg, Lamy, & Beckwith, 2016; McIver, Dyda, McNulty, Knight, Wand, & Guy, 2016), or monitoring health (Berrouiguet, Baca-García, Brandt, Walter, & Courtet, 2016; Hannan, Brooten, Page, Galindo, & Torres, 2016) including smoking cessation with mobile gaming (DeLaughter, Sadasivam, Kamberi, English, Seward, Chan, Volkman, Amante & Houston, 2016; Jamalian, Mezei, Levitan, Garber, Hammer, & Kinzer, 2012) have been researched. The medical health focus of the literature covers the same themes many an organization or person can benefit from as a whole. With reminders for appointments, taking medication, or offering paramedical support, discreet sexual health information services, shows the individual's ability to host connections to multiple environments. Although the aforementioned studies are different in terms of their purpose and focus, it is also the size of the organization, private or the number of participants involved that

actuate the basic set up, structure and power of the mobile engagement processes. These lessons are also relevant for any migration of use from personal messaging to institutional, promotional or wide scale mobile macro engagement.

From Micro to Macro Mobile Communication with a Cause

Macro level mobile engagement range from political campaigns, emergency responses or the revolutionary moments in northern Africa during the Spring of 2011. With continuous innovation, and creative repurposing of mobile apps and tools, a connectedness to socio-political issues abound around the globe. From the 2012 French election, 2016 USA presidential campaign to Obama's first election campaign when he announced his VP candidate via mobile communication (Baumgartner, Mackay, Morris, Otenyo, Powell, Smith, & Waite, 2010; Kenski, & Conway, 2016; Wells, Van Thomme, Maurer, Hanna, Pevehouse, Shah, & Bucy, 2016; West, 2012), are all examples of macro level mobile engagement. According to Baumgartner, et al (2010) and West (2012), Obama's campaign made use of texting and SMSN to identify and communicate with supporters, but also to raise money and coordinate virtual gatherings simultaneously across the country. According to West and the Mobile Economy Project (ibid.), mobile phones could expand mobilizing opportunities and increase citizen participation in civic issues, but could also cultivate government accountability.

In the words of the researcher John Donner "mobile phones accelerate urban metabolism" (2006), and in the case of the Tunisian and Egyptian revolutions, the use of mobile phones heightened the demonstrators' ability to coordinate protests as well as facilitate and disseminate the flow of information (Potts, 2011). With the rise of mobile phones, people are able to micro-coordinate (Ling, 2008, Mentor, 2014) more than just rendezvous' points or revolutions, but their daily activities from ordering food, car services,

mobile payments, or philanthropic support of a cause. Some of these affordances were re-modeled as a powerful macro-coordinating tool during various campaigns ranging from Black Lives Matter protests to elections. Now, with relative ease, a whole country's population can connect and mobilize to rally around a single agenda. Moving from one-to-one communication models, to one-to-many and many-to-many communication connectedness, like the mass protests and toppling of government examples mentioned earlier, there has been powerful utility in many modern day scenarios. For example, monitoring for free and fair elections in a crowdsourcing manner in any country with mobile phone infrastructure; from Kenya (Jeffery, 2011) to hopefully the USA.

Other examples of crowdsourcing through mobile phones are fast being adopted all around the world from crime reporting to live video broadcasts via mobile phone. Some examples include: South Africa's Corrupt Watch (2012), Kenya's Hatari Crowdmapping of various kinds of crime, mobile technology for crowdsourced agricultural development (Baumüller, 2013), Atlanta's MapATL for crime statistics, and New York City's DOT report a street maintenance reporting tool. Live mobile broadcast featured prominently in the global political arena in 2016 from legislative sit-ins in the USA protesting about gun violence (Folkenfalk, 2016; Peterson and Andrews, 2016) to the Turkish president debunking a coup and calling on the Turkish people to take to the streets in protest (Sherr, 2016). Live video recording and broadcasting with tools like Meerkat or Periscope, are used to document people's lives, documenting police brutality, a live show or kayaking down the Chicago river. One can but only hope that the continued rise of live mobile recordings and broadcasting will become part of police training, protocol or naturally evoke a sense of a panopticon in those police personnel who need it. These efforts rally the cognitive surplus (Shirky, 2010) of communities and unleashes the power of ordinary people by converting the mobile phone from a one-on-one communication

device to a one-to-many mass media tool. Issues of corruption, collating ideas and having people's voices heard by mobile means, impacts mobile civic engagement. People can report incidents which in turn builds consensus and connectedness to and with a cause. The new and unique ways that a previously private device can now be employed and affect change, is inspirational and speaks to the empowerment of people converted to participative citizens.

Problematically, many organizations and government entities have still not wholly embraced the wide spread adoption and practice of mobile service delivery. While bars and businesses, many educational institutions, live events or television shows invite you to interact with them via social media with photos or tweets, government departments are still neglecting to offer mobile friendly services or are still playing catch up in offering mobile first service designs.

SOLUTIONS AND RECOMMENDATIONS

There is a strong need for government agencies to partner with businesses, higher education and/or software companies and software engineers. This would cater for engaging with any of the aforementioned stakeholders' constituents, but also contribute to and promote social accountability. Many examples of multi-disciplinary efforts abound, with three examples to mention (a) a project that leveraged technology graduate students in New York City to build solutions or improvements in partnership with the New York Mayor's Office of Adult Education (Mentor and Murdoch, 2011), (b) The World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education (AACE, 2015) and the Obama administration's successful appeals to have software technology companies and individuals fix and improve the Healthcare web portal (Dawes, Vidasova, & Parkhimovich, 2016).

Given the mobile phone as the go-to, in-hand device for all of the aforementioned, this rapid movement has also informed a mobile first design framework (Van Esselstuyt, 2015). There certainly has been an increase in the adoption and adjustment to mobile first designs for websites, however, the need is still massive for a mobile first e-learning design approach (Curinga & Svaranos, 2016) as well as mobile first usability studies. The latter could aid and perhaps speed up the development of government and other organizational electronic engagement, which in turn could lead to better service offerings via mobile devices and improvements to quality of life offerings.

FUTURE RESEARCH DIRECTIONS

This chapter looked at the concept of social connectedness and the ways it may have been revealed in the studies referenced in this chapter. The categorization of the different mobile communication can be categorized into three groups of pairs, groups and macro levels. That categorization can be used to assist contextualizing the scope of micro or macro mobile communication and examine how the mobile engagement projects can be designed and delivered in different contexts.

The literature reviewed exemplifies that there is quantitative value in the number of mobile engagements. However, the socio-emotional aspects are prevalent, but cannot necessarily be quantified. What is lacking is a multifaceted mixed methods approach and analysis that looks at reported perceptions of socio-emotional value of social connectedness. Additionally, how trust circles have expanded on a previously considered private device via SMNS and if quality of life was truly impacted beyond just civic service delivery, but also through socio-emotional elements. Further research into the deeper causal relationships that exists between organizations and the individual participant is needed. Organizational and macro level mobile engagement needs to be interrogated

beyond commercial interest, but also how perceptions of social connectedness possible offer intangible support. Additionally, a delphi method could help establish a guide or definition for the concept of social connectedness which could offer further research direction with clarity, especially with the multitude of mobile devices and tools from which connectedness could be cultivated and maintained.

With the proliferation of mobile devices, apps and repurposing of handheld device tools, as well as the ability to collect big data through mobile engagement, there is a continued need for big data research studies. The quantitative approach could delve into granular or big data per user, group or cause, while the qualitative approach could offer insights via semi-structured individual or group interviews, focus groups, using a snowball method, or grounded theory. While quantitative analysis could be applied to all mobile communication or activity logs, short or long form surveys could be conducted prior to or after interviews to aid triangulation or crystallization of data. Mobile surveys could aid research input and design, if it takes a mobile first design and delivery approach, which could leverage a conditional method of posing questions, one at a time, and providing a question based on the previous response. Conditional surveys taken on a mobile device provides manageable bite size questions and response fields and could aid a deeper digging into specific aspects of the research area that needs to be investigated. However, completing a text survey could be influenced by the form factors and element interactivity of the mobile devices, due to physical constraints. The multitude of considerations for the varying form factors found in different geographic locations however could be overcome by a web responsive mobile first design.

As mobile digital geographies grow, become more complex and condensed at the same time, through SMNS and other mobile device activity, data visualization could be leveraged to offer snapshots of moments in time, or offer insights and analysis into social networks in a far quicker

manner than longitudinal studies. Another pictorial method that could be employed in conjunction with or separate from data visualization would be concepts maps for future mobile social connectedness research designs.

Additionally, Özçınar's 2016's research method of visualizing the evolution of mobile learning research through a co-citation analysis paradigm could be used as means to conduct and map research on social connectedness through users' mobile device use. However, capturing the research of human interaction through mobile computing systems, would not address gaps or needs of users. There would be a need to add to the co-citation approach with a specific focus on one element. Be it individual users, groups, causes or even specific devices, countries or regions within a country or the world. By zoning in on a specific focus, would offer data integrity to the investigations of mobile social interactions. Different countries have a multitude of service providers, competitive markets and costs that impact the use of mobile devices. Thus, research framing or foci will differ by country, and in the case of the USA and many countries, even by States or regions. Other demographic or genogram aspects could also be investigated to offer cross-sectional or comparative studies where factors such as access to mobile devices, usage according to age, family make-up, birth order or gender. In terms of information science, further research could focus on particular devices or a device, activity trends, or how social connectedness can be leveraged for education or exploited for mobile commerce.

CONCLUSION

Many more organizations have and want to dive into the use of mobile device engagement, (Chavin, Ginwala, & Spear, 2012; Muench, Stolk-Cooke, Van Morgenstern, Kuerbis, & Markle, 2014; Nielsen & Webb, 2011; Quin, 2013; West, 2014), but lack the theoretical and/or technical expertise. The size and shape of these organizations vary

and that bears upon the context and operational process of each organization's mobile intervention. The studies reviewed in this chapter show how mobile messaging increases social connectedness and improves people's quality of life. It must be noted that although life benefits and community affordances are possible because one's phone is always available; feelings of connectedness are not caused by the mobile phone itself. Fostering a sense of community is dependent on the reciprocity of communication initiated via the mobile phone, be it by voice calling, messaging or SMNS activity. Many of the processes and lessons learned from the literature reviewed, be it from mobile messaging research conducted within the health care settings or the political arena, could be applied to the workplace, educational, local city or federal governmental settings to offer socio-emotional support or services that could aid academic success and social well-being.

REFERENCES

- AACE. (2015, October). *E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education*. AACE.
- Baumgartner, J. C., Mackay, J. B., Morris, J. S., Otenyo, E. E., Powell, L., Smith, M. M.,... Waite, B. C. (2010). Communicator-in-chief: How Barack Obama used new media technology to win the White House. Lexington Books.
- Berrouguet, S., Baca-García, E., Brandt, S., Walter, M., & Courtet, P. (2016). Fundamentals for Future Mobile-Health (mHealth): A Systematic Review of Mobile Phone and Web-Based Text Messaging in Mental Health. *Journal of Medical Internet Research*, 18(6), e135. doi:10.2196/jmir.5066 PMID:27287668
- Chan, C. (2015, August 6). *When One App Rules Them All: The Case of WeChat and Mobile in China*. Retrieved August 06, 2016, from <http://a16z.com/2015/08/06/wechat-china-mobile-first/>
- Cheung, W. Y., Sedikides, C., & Wildschut, T. (2016). Induced nostalgia increases optimism (via social-connectedness and self-esteem) among individuals high, but not low, in trait nostalgia. *Personality and Individual Differences*, 90, 283–288. doi:10.1016/j.paid.2015.11.028
- Crang, M., & Mohamad, S. M. H. (2016). 17 Researching Virtual Communities. *Key Methods in Geography*, 270.
- Cui, D. (2016). Beyond connected presence: Multimedia mobile instant messaging in close relationship management. *Mobile Media & Communication*, 4(1), 19–36. doi:10.1177/2050157915583925
- Cumiskey, K. M., & Ling, R. (2015). The social psychology of mobile communication. *The Handbook of the Psychology of Communication Technology*, 228–246.
- Curinga, M. X., & Saravanos, A. (2016). Mobile First E-Learning. In D. Mentor (Ed.), *Handbook of Research on Mobile Learning in Contemporary Classrooms* (pp. 23–36). Hershey, PA: Information Science Reference; doi:10.4018/978-1-5225-0251-7.ch002
- Dawes, S. S., Vidasova, L., & Parkhimovich, O. (2016). Planning and designing open government data programs: An ecosystem approach. *Government Information Quarterly*, 33(1), 15–27. doi:10.1016/j.giq.2016.01.003
- DeLaughter, K. L., Sadasivam, R. S., Kamberi, A., English, T. M., Seward, G. L., Chan, S. W., Volkman, J.E., Amante, D. & Houston, T. K. (2016). Crave-Out: A Distraction/Motivation Mobile Game to Assist in Smoking Cessation. *JMIR Serious Games*, 4(1), e3.
- Folkenfalk, D. (2016, June 23). *When TV Cameras Shut Down, House Democrats Live-Stream Their Sit-In*. Retrieved July 17, 2016, from <http://www.npr.org/sections/alltechconsidered/2016/06/23/483205687/house-democrats-deliver-sit-in-via-digital-platforms>

- Freeman, G., Bardzell, J., & Bardzell, S. (2016, May). Revisiting Computer-Mediated Intimacy: In-Game Marriage and Dyadic Gameplay in Audition. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 4325-4336). ACM. doi:10.1145/2858036.2858484
- Goode, L. (2016, April 12). *Messenger and WhatsApp process 60 billion messages a day, three times more than SMS*. Retrieved June 19, 2016, from <http://www.theverge.com/2016/4/12/11415198/facebook-messenger-whatsapp-number-messages-vs-sms-f8-2016>
- Grieve, R., Indian, M., Witteveen, K., Tolan, G. A., & Marrington, J. (2013). Face-to-face or Facebook: Can social connectedness be derived online? *Computers in Human Behavior*, 29(3), 604–609. doi:10.1016/j.chb.2012.11.017
- Hannan, J., Brooten, D., Page, T., Galindo, A., & Torres, M. (2016). Low-Income First-Time Mothers Effects of APN Follow-up Using Mobile Technology on Maternal and Infant Outcomes. *Global Pediatric Health*, 3.
- Henderson, D. X., & Guy, B. (2016). Social connectedness and its implication on student–teacher relationships and suspension. *Preventing School Failure: Alternative Education for Children and Youth*, 1-7.
- Jamalian, A., Mezei, J., Levitan, P., Garber, A., Hammer, J., & Kinzer, C. K. (2012). The Lit2Quit Mobile App: Evoking Game-based Physiological Effects that Mimic Smoking. In *Proceedings, GLS 8.0 Games + Learning + Society Conference* (pp. 484-485). Madison, WI: ETC Press.
- Kenski, K., & Conway, B. A. (2016). Social Media and Elections. *Praeger Handbook of Political Campaigning in the United States*, 191.
- Kim, H. S. (2016). What drives you to check in on Facebook? Motivations, privacy concerns, and mobile phone involvement for location-based information sharing. *Computers in Human Behavior*, 54, 397–406. doi:10.1016/j.chb.2015.08.016
- Kim, J. (2012, June 28). *10 Examples of Mobile Health Around the World*. Retrieved July 10, 2016, from <http://www.networkworld.com/article/2286628/data-center/75802-10-Examples-of-Mobile-Health-Around-the-World.html#slide1-6>
- Kim, Y., Wang, Y., & Oh, J. (2016). Digital media use and social engagement: How social media and smartphone use influence social activities of college students. *Cyberpsychology, Behavior, and Social Networking*, 19(4), 264–269. doi:10.1089/cyber.2015.0408 PMID:26991638
- Konok, V., Gigler, D., Bereczky, B. M., & Miklósi, Á. (2016). Humans attachment to their mobile phones and its relationship with interpersonal attachment style. *Computers in Human Behavior*, 61, 537–547. doi:10.1016/j.chb.2016.03.062
- Krastel, Z., Bassellier, G., & Ramaprasad, J. (2015). Enabling Social Connectedness: The Role of Social Features on Online Music Sites. In *Conference on Information Systems and Technology (CIST)* (p. 2).
- Lam, C. (2012). The efficacy of text messaging to improve social connectedness and team attitude in student technical communication projects: An experimental study. *Journal of Business and Technical Communication*.
- Lee, R. M., Draper, M., & Lee, S. (2001). Social connectedness, dysfunctional interpersonal behaviors, and psychological distress: Testing a mediator model. *Journal of Counseling Psychology*, 48(3), 310-318.
- Lee, S. Y. T., & Phang, C. W. D. (2015). Leveraging social media for electronic commerce in Asia: Research areas and opportunities. *Electronic Commerce Research and Applications*, 14(3), 145–149. doi:10.1016/j.elerap.2015.02.001
- Longhurst, R. (2016). Mothering, digital media and emotional geographies in Hamilton, Aotearoa New Zealand. *Social & Cultural Geography*, 17(1), 120–139. doi:10.1080/14649365.2015.1059477

- Lundy, B. L., & Drouin, M. (2016). From social anxiety to interpersonal connectedness: Relationship building within face-to-face, phone and instant messaging mediums. *Computers in Human Behavior*, 54, 271–277. doi:10.1016/j.chb.2015.08.004
- McIver, R., Dyda, A., McNulty, A. M., Knight, V., Wand, H. C., & Guy, R. J. (2016). Text message reminders do not improve hepatitis B vaccination rates in an Australian sexual health setting. *Journal of the American Medical Informatics Association*, 23(e1), e88–e92. doi:10.1093/jamia/ocv145 PMID:26499103
- Mentor, D. (2014). Book Chapter Mobile phone use enhances social connectedness. In Z. Yan (Ed.), *Encyclopedia of Mobile Phone Behavior*. Hershey, PA: IGI Global.
- Mentor, D., & Murdoch, D. (2011, October). We Are New York Social Media for Adult Education. In *E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education* (Vol. 2011, No. 1, pp. 1499-1504).
- Moreno-Munoz, A., Bellido-Outeirino, F. J., Siano, P., & Gomez-Nieto, M. A. (2016). Mobile social media for smart grids customer engagement: Emerging trends and challenges. *Renewable & Sustainable Energy Reviews*, 53, 1611–1616. doi:10.1016/j.rser.2015.09.077
- Naslund, J. A., Aschbrenner, K. A., Marsch, L. A., & Bartels, S. J. (2016). The future of mental healthcare: Peer-to-peer support and social media. *Epidemiology and Psychiatric Sciences*, 25(02), 113–122. doi:10.1017/S2045796015001067 PMID:26744309
- Nolan, S., Hendricks, J., & Towell, A. (2016). Adolescent mothers' use of social networking sites creating positive mental health outcomes. *Australian Nursing and Midwifery Journal*, 23(11), 50. PMID:27530036
- Özçınar, H. (2016). Visualizing the Evolution of Mobile Learning Research. In D. Mentor (Ed.), *Handbook of Research on Mobile Learning in Contemporary Classrooms* (pp. 1-22). IGI Global.
- Peterson, K., & Andrews, N. (2016, June 22). *With TV Cameras Off, House Democrats Broadcast Their Own Sit-In*. Retrieved July 17, 2016, from <http://blogs.wsj.com/washwire/2016/06/22/house-democrats-stage-sit-in-protest-for-gun-bill-vote/>
- Piwek, L., & Joinson, A. (2016). What do they snapchat about? Patterns of use in time-limited instant messaging service. *Computers in Human Behavior*, 54, 358–367. doi:10.1016/j.chb.2015.08.026
- Rana, A. I., van den Berg, J. J., Lamy, E., & Beckwith, C. G. (2016). Using a Mobile Health Intervention to Support HIV Treatment Adherence and Retention Among Patients at Risk for Disengaging with Care. *AIDS Patient Care and STDs*, 30(4), 178–184. doi:10.1089/apc.2016.0025 PMID:27028183
- Richards, L. (2016). For Whom Money Matters Less: Social Connectedness as a Resilience Resource in the UK. *Social Indicators Research*, 125(2), 509–535. doi:10.1007/s11205-014-0858-5 PMID:26770011
- Rose, C. (2016, June 16). *Dag Kittlaus*. Retrieved June 19, 2016, from <https://charlierose.com/videos/28249>
- Russell, J. (2016, March 17). *Messaging app WeChat is becoming a mobile payment giant in China*. Retrieved August 07, 2016, from <https://techcrunch.com/2016/03/17/messaging-app-wechat-is-becoming-a-mobile-payment-giant-in-china/>
- Satici, S. A., Uysal, R., & Deniz, M. E. (2016). Linking social connectedness to loneliness: The mediating role of subjective happiness. *Personality and Individual Differences*, 97, 306–310. doi:10.1016/j.paid.2015.11.035

Schroeder, R. (2010). Mobile phones and the inexorable advance of multimodal connectedness. *New Media & Society*, 12(1), 75–90. doi:10.1177/1461444809355114

Sherr, I. (2016). The Smartphone Has Become the Center of Our Political Lives. *CNET*.

Statista. (2014, December 20). *Number of mobile messaging users worldwide 2014* | *The Statistics Portal*. Retrieved August 21, 2016, from <http://www.statista.com/statistics/369260/mobile-messenger-users/>

The History of Multimedia Messaging (MMS) - MMS London. (2016). Retrieved June 19, 2016, from <http://www.mmsworldlondon.com/history.htm>

Visser, T., Dadlani, P., van Bel, D., & Yarosh, S. (2010, April). Designing and evaluating affective aspects of sociable media to support social connectedness. In CHI'10 Extended Abstracts on Human Factors in Computing Systems (pp. 4437–4440). ACM. doi:10.1145/1753846.1754168

Walton, G. M., & Cohen, G. L. (2011). A brief social-belonging intervention improves academic and health outcomes of minority students. *Science*, 331(6023), 1447–1451. doi:10.1126/science.1198364 PMID:21415354

Weinswig, D. (2015). *The Changing Face of Mobile Payments*. Hong Kong: Fung Business Intelligence Center - Global Retail & Technology. Retrieved August 6, 2015, from <http://docplayer.net/15493532-The-changing-face-of-mobile-payments.html>

Wells, C., Van Thomme, J., Maurer, P., Hanna, A., Pevehouse, J., Shah, D. V., & Bucy, E. (2016). Coproduction or cooptation? Real-time spin and social media response during the 2012 French and US presidential debates. *French Politics*, 14(2), 206–233. doi:10.1057/fp.2016.4

West, D. (2012). M-Campaigning: Mobile Technology and Public Outreach. *Issues in Technology Innovation*, 15, 1–10.

Wu, L. (2016). Understanding the Impact of Media Engagement on the Perceived Value and Acceptance of Advertising within Mobile Social Networks. *Journal of Interactive Advertising*, 16(1), 59–73. doi:10.1080/15252019.2016.1160331

Yoo, J., Miyamoto, Y., & Ryff, C. D. (2016). Positive Affect, Social Connectedness, and Healthy Biomarkers in Japan and the US. *Emotion (Washington, D.C.)*, 27(Jun). doi:10.1037/emo0000200 PMID:27348497

ADDITIONAL READING

Alehegn, T., & Mentor, D. (2016). Mobile Learning for Social Change: Democratizing Education and Civic Engagement. *Handbook of Research on Mobile Learning in Contemporary Classrooms*, 363–377.

Anderson, K. E. (2016). Getting acquainted with social networks and apps: WhatsApp-ening with Mobile Instant Messaging? *Library Hi Tech News*, 33(6), 11–15. doi:10.1108/LHTN-07-2016-0032

Chadwick, A., & Stromer-Galley, J. (2016). Digital Media, Power, and Democracy in Parties and Election Campaigns Party Decline or Party Renewal? *The International Journal of Press/Politics*, 21(3), 283–293. doi:10.1177/1940161216646731

Chayko, M. (2012). Connecting: How we form social bonds and communities in the Internet age.

Kenski, K., & Conway, B. A. (2016). Social Media and Elections. *Praeger Handbook of Political Campaigning in the United States* [2 volumes], 191.

Okoli, C., & Pawlowski, S. D. (2004). The Delphi method as a research tool: An example, design considerations and applications. *Information & Management*, 42(1), 15–29. doi:10.1016/j.im.2003.11.002

Schroeder, R. (2010). Mobile phones and the inexorable advance of multimodal connectedness. *New Media & Society*, 12(1), 75–90. doi:10.1177/1461444809355114

Walton, G. M., & Cohen, G. L. (2011). A brief social-belonging intervention improves academic and health outcomes of minority students. *Science*, 331(6023), 1447–1451. doi:10.1126/science.1198364 PMID:21415354

KEY TERMS AND DEFINITIONS

Co-Presence: The practice of people engaging in real-time face-to-face social activities and being engrossed in the here and now of that social activity, but also being virtually present and via texting or mobile messaging with dual occupation of real and virtual space.

Connected Presence: Unconstrained connections from time and place, refers to people situated in a different physical location, but splitting their attention and connecting via virtual means with another location or person located elsewhere.

Mobile Crowdsourcing: Mobile crowdsourcing describes the collation of a large group of people's views and/or observations. These crowdsourcing activities are processed on mobile phones or other handheld mobile devices.

Mobile Learning: Also known as mLearning refers to creating, leveraging or designing learning opportunities via portable technologies from and/or for interpersonal and organizational mobile communication.

Mobile Messaging: Mobile messaging typically from a mobile phone or mobile device. Examples include SMS texting, Multimedia Message Service (MMS) to a host of wireless access protocol (WAP) mobile messaging services like WhatsApp, WeChat, Kakao, Kik, Line, etc. The latter being device agnostic, meaning they can be installed and used on many different types of mobile phone platforms.

Mobile Social Media: Social communication and networking where affiliated individuals with dis/similar interests converse and connect virtually and/or in-person with one another through their mobile phone, handheld and/or tablet device.

Social Connectedness: Concept often been used to characterize degrees of interpersonal trust, attachment security, social competency, and a sense of belonging, online and in real world.

Socio-Emotional: The concept of socio-emotional aspects refers to a person initiating, cultivating and responding to others, to form relationships with other people in their lives such as parents, relatives and friends. These socio-emotional aspects can be ephemeral or long-standing, when it comes to interactions with people. Typically with familiar, but not excluding unfamiliar people. It can span the spectrum of how people interact with others to how they manage communication or cope with adversity, stress, establish positive relationship and the capacity to develop relationships with peers, often times through communication.

Mobile Applications for Automatic Object Recognition

M**Danilo Avola***University of Udine, Italy***Gian Luca Foresti***University of Udine, Italy***Claudio Piciarelli***University of Udine, Italy***Marco Vernier***University of Udine, Italy***Luigi Cinque***Sapienza University, Italy*

INTRODUCTION

In recent years, the technological improvements of mobile devices in terms of computational capacity, embedded sensors, natural interaction and high-speed connection are enabling an ever-increasing number of designers to develop advanced mobile applications to be used in everyday life. Among these, the vision based applications for the Automatic Object Recognition (AOR) play a key role since enable users to interact with the world around them in innovative way that makes more productive and profitable their entertainment, learning and working activities. Within the introduced topic, objects can be physical entities of any type, from small sized objects (e.g., paintings, statues) up to medium sized objects (e.g., road signs, vehicles) or large sized objects (e.g., airplanes, buildings). The real step forward in moving the AOR from the common Computer Vision (CV) paradigm to that of the Mobile Vision (MV) is the ability to perform, with a mobile device, the recognition and classification of objects both in pervasive way (i.e., anywhere, anytime) and in indoor as well as outdoor environments. Up to about ten years ago, the only way to recognize an object of the

real world acquired by an RGB mobile camera in outdoor environments was the adoption of visual expedients (e.g., markers, barcodes) which, once recognized, contained all the required information of the target object. Although it is still an open issue, the recent advances in image quality, image feature extraction and image classifiers make possible the automatic recognition of an ever-increasing number of objects. Nowadays, there are many AOR applications that support a wide range of popular fields (e.g., culture, tourism, food). In addition, recently, an ever-increasing number of AOR applications aimed in supporting professional fields (e.g., forensic, architectural, engineering, medical) have been developed. The proposed chapter is divided into four sections. The first one, Background, introduces a taxonomy of the current mobile applications and explores the most recent works in AOR mobile applications highlighting crucial aspects of the object recognition algorithms, including the feature extraction processes and the implemented classifiers. The second one, MV Development Technologies, provides an overview of the current frameworks used to support mobile AOR applications. The third one, Future Research Trends, briefly discusses the

DOI: 10.4018/978-1-5225-2255-3.ch538

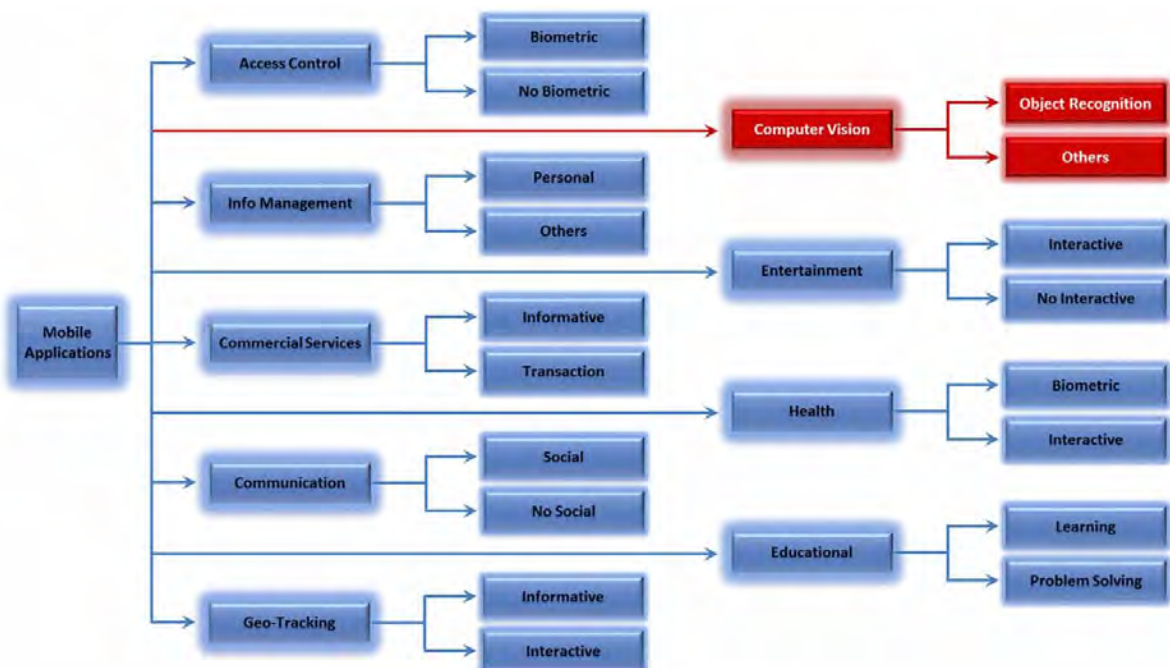
aims of the next generation of AOR applications. Finally, the last one, Conclusion, summarizes what has been reported in the present chapter.

BACKGROUND

In the last decade, mobile devices have had ongoing and growing technological advances. Currently these devices, even those of low cost, have a set of hardware features that make them comparable with a wide range of desktop processing units. In fact, these mobile devices, with particular reference to those of the latest generation, present a set of significant improvements, including:

- **Multi-Core Processor (MCP):** A single processor that contains several cores. This technology typical of common processing units (e.g., workstations, servers) allows mobile devices to rapidly process a large amount of data also improving the performance of running multiple applications.
- **Advanced Storage Capacity (ASC):** A large amount of internal memory and the possibility to adopt external memories (e.g., compact flash, memory stick). This technology allows mobile devices to support both complex data and bulky applications.
- **Mobile Sensing (MS):** The set of sensors embedded in a mobile device. This technology allows mobile devices to be equipped with a wide range of sensors, including: Red-Green-Blue (RGB) sensor (i.e., image camera), Global Position System (GPS) receiver (i.e., localization system), accelerometer sensor (i.e., proper acceleration “g-force”), gyroscope sensor (i.e., orientation system) and others. These sensors are used to acquire any type of information from both the external world (e.g., images, temperature, pressure) and user’s behavior (e.g., speed, locations, actions).
- **Natural User Interfaces (NUIs):** The set of natural interfaces to favor the Human-Computer Interaction (HCI) between users

Figure 1. A taxonomy of mobile applications



and mobile devices. This technology allows users to adopt human-oriented interfaces (e.g., speech recognition, touch-screen interaction) to manage data and devices.

- **Fast Internet Connections (FICs):** The set of information and communication technologies (ICTs) that allow devices to access to any type of resource, including: World Wide Web (WWW), cloud computing, dedicated networks and others.

In contrast to the common processing units (e.g., laptops, desktops), the mobile devices are versatile and pervasive. Moreover, they can be suitably used in indoor and outdoor environments. The introduced aspects have shifted the role of the mobile devices far beyond their original aim. Today, these latter are considered real tools to support each and every activity of the human life thus promoting an ever-increasing development of mobile applications in any kind of context including entertainment and tourism as well as learning and working. In Figure 1 a taxonomy of the current mobile applications is reported. Although more classes of the taxonomy can characterize any application, usually one of them can be considered the main one, that is, the class that more than others defines aims and scope of the application itself. Among the mobile applications, those based on vision play a key role since they allow users to expand one of their senses: the

sight. In fact, nowadays, the AOR applications support many critical tasks, including: forensic investigations, interior design and visual disabilities. Finally, it should be noted that currently the term mobile device covers a wide range of tools from the common smartphones up to the handheld tablets or intelligent glasses. In the next subsection, a selected set of AOR mobile applications is provided. For each of them two main aspects are highlighted: feature extraction process and the implemented classifier.

AOR Mobile Applications

The RGB camera is the most common sensor equipped within the current mobile devices. Its typical standard resolution for the acquisition of static photos is 12 megapixels (MP), while the acquisition of video is usually performed in 1080p (i.e., Full High Definition, FHD, defined as 1080 x 1920 pixels) with 30 frames per second (fps). Moreover, these cameras are often controlled by embedded software able to optimize their main features, including: image stabilization, sensor focus and image color balance. All these characteristics have a huge importance since each vision based mobile application works by processing one or more photos, otherwise a video in on-line or off-line mode. In any case, regardless of a specific task, the pipeline of an AOR mobile application can be generalized as shown in Figure 2.

Figure 2. Typical architecture of an AOR mobile application



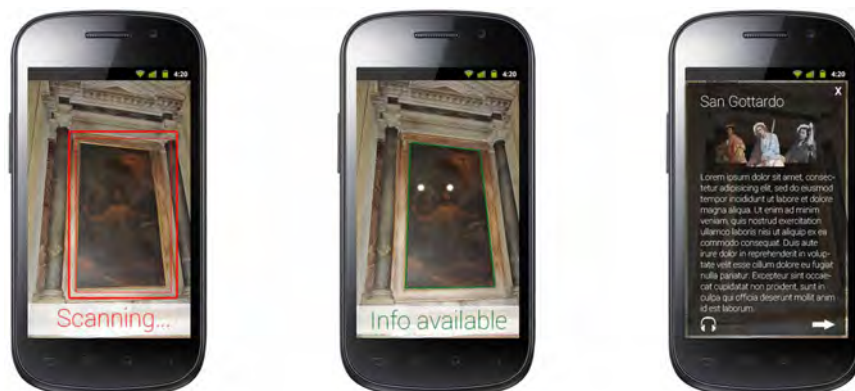
In general, the first step of an AOR mobile application is the definition of the database containing a template representation (i.e., feature descriptors) of each object to be recognized. The construction of the database is usually performed in monitored environment where all the technical aspects (e.g., lighting, contrast, resolution) are carefully treated. The database can be made available to the mobile application locally (i.e. embedded within the application) or remotely (e.g., via server or cloud). When the mobile device acquires a set of photos or a video from the real world containing the target objects, a description of them (i.e., features descriptors) is extracted. The last step of an AOR mobile application is the comparison (i.e., feature matching) between these descriptions and the ones previously stored within the database to detect the correct object and provide the linked information to the user. In Figure 3 an example of AOR mobile application for the recognition of paintings is reported. In particular, the Figure 3 highlights three main steps of the application running: scanning of a region of interest (left), recognition of the painting (middle) and, finally, the feedback provided by the application to the user (right).

A first remarkable AOR mobile application is presented in Saeed et al. (2013), where the authors describe a system to help blind and visually impaired people to better understand the environments in which they live and work thus improv-

ing the quality of their life. The proposed system has different strengths, including: high accuracy, speed, portability and low cost. The AOR mobile application designed by the authors adopts ORB as local feature detector and descriptor. They claim that the main reason behind this choice is that ORB is able to compute and match sets of feature descriptors very efficiently. Since the real-time performance is a mandatory requirement of their application, the authors have preferred ORB to other local feature detectors and descriptors (e.g., SIFT, SURF) which are more robust to viewing condition changes. The classifier of the proposed system has been implemented with three different variations. In the first one, the classifier adopts a similarity matching measure based on Hamming distance. In this context, the Hamming distance can be considered as the measure able to quantify the similarity of two key points in terms of shape and color properties. In the second one, the classifier adopts an outlier removal technique based on Nearest Neighbor Ratio. This technique tends to consider as inliers those keypoints (of a live image) whose Hamming distance from a reference keypoint (of a database image) is less with respect to a fixed threshold. However, regardless a specific classifier, the AOR mobile application described by the authors presents high levels of robustness and performance.

Another interesting and funny AOR mobile application is reported in Kawano et al. (2013),

Figure 3. An example of AOR mobile application for the recognition of paintings



where the authors detail a food recognition system able both to estimate calories and nutritious of foods and to record the user's eating habits. Also in this case, the system proposed by the authors has different strengths, including: high accuracy and speed. The authors have spent a large amount of time to study a suitable algorithm able to extract in real-time a set of key features from the target objects. Finally, they have chosen a combination between color histogram and Bag-of-SURF as feature detector and descriptor. In particular, for the extraction of a first set of keypoints the authors subdivide the target image into 9 blocks (i.e., 3 x 3 blocks) and extract a 64-bin RGB color histogram from each of them. Another set of keypoints is extracted from the image by using a SURF variation of the Bag-of-Words. This last is a technique initially developed to support the natural language processing where a document (i.e., a set of sentences) can be represented by a histogram (i.e., occurrences) of its words disregarding the grammar. In CV this algorithm has been expanded with the opportunity to consider the keyfeatures of an image instead of the words of a document. The authors have adopt a linear Support Vector Machine (SVM) as core of the classifier. In general, a linear SVM is an algorithm able to separate two sets of objects into their respective groups. In the context proposed by the authors, each SVM classifier is able to recognize a specific food category. Since they have tested the AOR application with 50 food categories, they have trained it with 50 linear SVM classifiers.

A very interesting and practical AOR mobile application is described in Chen et al. (2015), where the authors present a multi-purpose, continuous, real-time object recognition system. The application proposed by the authors presents a wide range of advanced features: it captures full-motion video, locates objects of interest, recognizes and labels them, and tracks them from frame to frame for the user. The authors claim that their algorithms for the object recognition entail significant computation, for this reason their system runs them on server machines. During the development,

the authors have taken care of several technical details to support the effectiveness and robustness of the application. Among these, an active cache of video frames has been implemented to overcome possible latency issues due to the communication between server and mobile device. In this project, the authors have mainly adopted SIFT and SURF as feature detectors and descriptors, while they have mainly used SVM as classifier. The proposed AOR mobile application has been tested on a wide set of objects (e.g., road signs) concerning different scenarios. The results have shown that the application achieves high precision and accuracy, moreover, its architecture is able to save bandwidth and resources. This system can be considered a step forward over the current state-of-the-art.

The authors of the work reported in Hong et al. (2012) present an AOR mobile application for the recognition of flowers. This work is of particular interest since the target objects are usually placed within natural environments whose background is affected by several critical aspects (e.g., light, surrounding objects) that make the feature extraction process from the foreground a hard task. Their system is based on a client-server architecture, where the client (e.g., smartphone) acquires flower images and sends them to a remote server that, after the processing step, provides to the user a feedback concerning their classification. The system proposed by the authors adopts a complex feature extraction process based on searching for the

contour of flowers by using color and edge information. The features extracted by means of these latter are compared with the images stored into the database by using a classifier based on color clustering and distance histogram matching. Like the previous selected works, also in this case the authors have obtained remarkable results. The application could be also customized for the recognition of other kinds of objects.

A different work is discussed in Jung (2012), where the author explains an AOR mobile application whose classification approach has been

specifically designed to be trained and recognized by a mobile device. The system has been developed with a dual purpose. The first one regards the opportunity to recognize a three-dimensional object on a two-dimensional photo, even when the object is shown from different perspectives and presents different shadows and reflections. The second one concerns the opportunity to perform the training phase, related to the introduction of new objects within the database, by using the mobile device. The proposed AOR application adopts a complex feature extraction process. A first part of the feature vector is composed by color, edge activity and frequency-based features which are obtained from a statistical analysis of local image properties. This makes the features robust against image variations, but also makes them blind towards the geometrical shapes of the objects themselves. To compensate this aspect, the rest of the feature vector is composed by interest point features that are computed by using the SURF algorithm. Also in this application, the author has used SVM as classifier thus obtaining excellent results.

Analyzing the literature, in the last decade, other considerable systems that share technical aspects with the previously discussed AOR mobile applications have been developed. In Hua et al. (2014), for example, the authors have designed an interesting recognition system that can work both with images and with arbitrary entities. While, the work reported in Ling et al. (2011) presents a robust, portable and expandable system for the automatic recognition of traffic signs. The authors of the work described in Quack et al. (2008) show a sophisticated system that combines the concepts of object recognition and Internet of Things (IoT). They present an AOR application that allows to request information on physical objects by taking a picture of them. Finally, the work detailed in Kumar et al. (2012) introduces a novel multi-frame object detection application for the mobile platform that is capable of object localization. The significance of their method is

its ability to bring evidence from minor changes in views to improve object detection.

A last work, Aletto et al. (2016), is now introduced to illustrate the generalization of the term mobile device. In this latter the authors propose to adopt the first person prospective and a wearable device to support a tourist during the exploration of a site. Their system is designed to retrieve architectural details of historical buildings and to provide feedback and information to the users. The main idea behind this project is that, in an unconstrained outdoor tour, a user may not be able to autonomously identify all the relevant details in the site. So having a smart embedded guide that sees what a user sees and can provide precise and contextual information when asked to could result in a more enjoyable and complete experience. The object recognition application proposed by the users is composed by a wearable board, a glass-mounted camera, a wireless audio output device and the visitor's mobile device. Using the latter as interface, the user can ask the system to acquire a frame from the wearable camera, process it, and retrieve all the details present in the scene. The user can then see a visual representation of the output on the device screen, where tapping each detail provides audio information regarding the selected detail. The current technology improvements are allowing to obtain devices (e.g., Google Glass) which integrate all-in-one the described interfaces.

With the aim to provide an overview of the just reported keypoint detectors and descriptors, Table 1 summarizes some basic information about SIFT, SURF and ORB (Isik et al. (2015)).

Analyzing the reported works, it is possible to observe the widespread usage of the SVM classifier. This fact is due to several features which are reported within the Table 2.

MV DEVELOPMENT TECHNOLOGIES

All the AOR mobile applications reported in the previous section have been developed by using

Table 1. Keypoint detectors and descriptors

Algorithm	Main Information
SIFT	• The Scale-Invariant Feature Transform (SIFT) is a keypoint detector and descriptor. It consists of four main stages (Lowe (2004)): ◦ Scale-Space Extrema Detection: The algorithm searches over all scales and image locations. It is implemented efficiently by using a difference-of-Gaussian function to identify potential interest points that are invariant to scale and orientation; ◦ Keypoint Localization: At each candidate location, a detailed model is fit to determine location and scale. Keypoints are selected based on measures of their stability; ◦ Orientation Assignment: One or more orientations are assigned to each keypoint location based on local image gradient directions. All future operations are performed on image data that has been transformed relative to the assigned orientation, scale, and location for each feature, thereby providing invariance to these transformations; ◦ Keypoint Descriptor: The local image gradients are measured at the selected scale in the region around each keypoint. These are transformed into a representation that allows for significant levels of local shape distortion and change in illumination.
SURF	• The Speeded Up Robust Features (SURF) is a keypoint detector and descriptor. Its pipeline is similar to that of SIFT, but includes some significant differences aimed to make SURF faster, in particular (Bay et al. (2008)): ◦ The algorithm computes the keypoints by using the Fast-Hessian Detector that bases on an approximation of the Hessian matrix for a given image point; ◦ The responses of Haar wavelets are used for orientation assignment, before the keypoint descriptor is formed from the wavelet responses in a certain surrounding of the keypoint.
ORB	• The Oriented FAST and Rotated BRIEF (ORB) is a keypoint detector and descriptor. It has been developed by the composition of FAST keypoint detector and BRIEF keypoint descriptor, where (Rublee et al. (2011)): ◦ Features from Accelerated Segment Test (FAST) is a corner detection algorithm whose main strength is the computational efficiency, while its main weakness is the inability to produce scale-invariant keypoints. However, due to its high speed performance FAST is very suitable for real-time image and video processing; ◦ Binary Robust Independent Elementary Features (BRIEF) is a feature point description algorithm that can be combined with arbitrary detectors. The main weakness of BRIEF is that it can not manage wide rotations of the feature points. • In a first step, ORB uses FAST to quickly find the keypoints of an image and adds an additional component to FAST to manage their scale. In a second step, ORB uses BRIEF to accurately describe these keypoints and adds an additional component to BRIEF to manage their rotation.

one or more computer vision frameworks. These last are useful tools that include a wide range of basic algorithms which can support the most common aspects of the implementation of any application, including: feature extraction, client-

server communication, classification process and others. Since an AOR mobile application can be developed both in embedded and in client-server mode, each framework can be potentially used with any kind of mobile device (e.g., Android,

Table 2. SVM classifier

Classifier	Main Information
SVM	• The Support Vector Machine (SVM) can be considered one of the most effective classifiers to support applications in real contexts. This is due to several technical features which are briefly reported below (Tong et al. (2002)): ◦ The basic form of the SVM is based on the concept that two classes of objects can be suitably separated by a linear function. This simple and effective method results robust enough to be used in a wide range of practical cases since in these latter it is quite easy to find such sets of discriminative features; ◦ Unlike other popular classifiers (e.g., HMMs, NNs) the SVM can be easily implemented in supervised or unsupervised mode according to the requirements of the specific application; ◦ The linear SVM can be extended to a non-linear classifier by expanding the plane of the solutions into a feature space by using a set of non-linear basic functions. These latter can be computed by means of a kernel representation whose solution is written as a weighted sum of simpler functions. In this way, the SVM can be used to classify more than two classes maintaining, at the same time, all the advantages of the linear case; ◦ In the last ten years, an interesting new trend has regarded the implementation of the classifiers by means of the Neural Trees (NTs), which have the same benefits of the SVM based classifiers but can be also used in complex environments (Foresti et al. (2004a), Foresti et al. (2004b)).

Table 3. Computer vision framework

Framework	Main Features	Notes
OpenCV	- It is composed by the following main modules: Core: It includes core functionalities such as the management of the data structures and macros; Image Processing: It includes image filtering, geometric image transformation, histograms; Video Analysis: It includes motion analysis, background analysis, object tracking; Features Framework: It includes feature detectors, feature extractors, feature descriptors, feature matchers, object categorization; Machine Learning: It includes statistical models, SVM, decision trees, boosting, neural networks.	- Current version 3.1 under a BSD license. - Written in optimized C/C++. - Desktop OS: - Windows, Linux, Android, MacOS; - FreeBSD, OpenBSD. - Mobile OS: - Android, Maemo, iOS. - Supported Modules: - CUDA, OpenCL. - Link: http://opencv.org/
BoofCV	- It is composed by the following main modules: Image Processing: It includes binary images, color space, image filters, image derivate; Feature: It includes associate interest points, edge detector, dense optical flow, fitting ellipses; Segmentation: Color segmentation, super-pixels, thresholding, watershed with seeds; Structure from Motion: Perspective-n-point, multi-view scene reconstruction.	- Current version 0.23 under an Apache 2.0 license. - Written in optimized Java. - Supported Modules: - Dedicated Kinect module; - Dedicated Android module; - Link: http://boofcv.org/
SimpleCV	- It is composed by the following main modules: Features Package: It includes feature extractors, morphology features, image filters, blob filters; Machine Learning Package: It includes SVM classifier, confusion matrix, KNN classifier, shape context classifier, temporal color tracker; Segmentation Package: It includes MOG segmentation, color and running segmentation; Tracking Package: It includes LK tracker, CAM shift tracker, MF tracker, SURF tracker.	- Current version 1.3 under a BSD license. - Written in optimized Python. - Desktop OS: - Windows, Linux, MacOS. - Mobile OS: - Android. - Required Modules: - OpenCV. - Link: http://simplecv.org/
EmguCV	- One of the main feature of the framework is the .NET wrapper for the OpenCV image processing thus including all its libraries, such as: - Core functionalities; - Image Processing; - High-Level GUI and Media I/O; - Video Analysis; - Camera Calibration and 3D Reconstruction.	- Current version 3.1.0 under dual license that includes also GNU GPL v3. - Desktop OS: - Windows, Linux, MacOS. - Mobile OS: - Android, Windows Phone, iOS - Wrapper Module for OpenCV. - Link: http://www.emgu.com/

The table shows only a little portion of the features of each framework (follow the link for the whole set).

iOS, Windows Phone). For this reason, Table 3 summarizes some main information of some of the most popular computer vision frameworks regardless of a specific Operative System (OS) mounted on a mobile device.

FUTURE RESEARCH DIRECTIONS

As presented in the previous section, in literature exists many significant mobile applications able to automatically recognize objects on a scene.

Mainly, they are based on robust recognition algorithms that include both feature extraction processes and classifiers. The current state-of-the-art of these algorithms has been designed by considering as input both images and video frames at standard resolution. Nevertheless, the last generation of mobile devices are equipped with more sophisticated image sensors able to capture static images with a higher resolution (e.g., 16MP) as well as videos in Ultra High-Definition (i.e., 3840 x 2160 pixels with 30/60 frame per second) or higher. This breakthrough in image and video

resolution will allow feature extraction algorithms to better distinguish objects into the scene. Moreover, also objects placed at long distance will be processed in more accurate way. It is worth noting that high resolution frames include a large number of pixels compared to those at a standard definition, thus are required high computational costs to process each of them. However, it should not be a significant issue since, usually, the evolution of the camera sensors follow the evolution of the processors mounted on mobile devices.

Another important feature that should be investigated in future researches regards the use of the IR cameras. These latter are particular image sensors through which the depth maps of persons and objects are built. The maps of these elements show the distance between them and the camera thus making the recognition process simpler with respect to the traditional approaches based on RGB cameras. One of the first mobile devices in this direction is Tango, a Google's project that adopts an advanced tablet equipped with an RGB-IR camera and combines 3D motion tracking with depth sensing to provide the tablet with the ability to know where it is and how it moves through space (source: <https://www.google.com/atap/project-tango/>). Finally, a last future proposal should be the analysis of the innovative cloud infrastructures (e.g., clustering networks) which can support both complex recognition applications and computational costs.

CONCLUSION

In this contribution, a selection of some of the most important AOR mobile applications has been presented. For each of them the feature extraction process and the classifier have been highlighted. The contribution has also reported some main features of some of the most popular computer vision frameworks used to support the develop-

ment of the mobile applications. Initially, the contribution has also provided a possible taxonomy of the mobile applications to point out the large amount of tasks that are currently supported by these applications. A brief analysis of the future research directions concludes the chapter. In this latter, the role of the next generation of mobile technologies is highlighted.

REFERENCES

- Alletto, S., Abati, D., Serra, G., & Cucchiara, R. (2016). Exploring architectural details through a wearable egocentric vision device. *Sensors (Basel, Switzerland)*, 16(2), 1–15. doi:10.3390/s16020237 PMID:26901197
- Bay, H., Ess, A., Tuytelaars, T., & Van Gool, L. (2008). Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding*, 110(3), 346–359. doi:10.1016/j.cviu.2007.09.014
- Chen, T. Y.-H., Ravindranath, L., Deng, S., Bahl, P., & Balakrishnan, H. (2015). Glimpse: Continuous, real-time object recognition on mobile devices. In *Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems – SenSys* (pp. 155–168). New York: ACM Publisher. doi:10.1145/2809695.2809711
- Foresti, G. L., & Dolso, T. (2004a). Adaptive high-order neural trees for pattern recognition. *IEEE Transactions on System. Man and Cybernetics – Part B*, 34(2), 988–966. doi:10.1109/TSMCB.2003.818538
- Foresti, G. L., Micheloni, C., & Snidaro, L. (2004b). Event classification for automatic visual-based surveillance of parking lots. In *Proceedings of the 17th International Conference on Pattern Recognition – ICPR* (pp. 314–317). Cambridge, UK: IEEE Publisher. doi:10.1109/ICPR.2004.1334530

- Hong, S. W., & Choi, L. (2012). Automatic recognition of flowers through color and edge based contour detection. In *Proceedings of the 3th International Conference on Image Processing Theory, Tools and Applications – IPTA* (pp. 141-146). Istanbul: IEEE Publisher. doi:10.1109/IPTA.2012.6469535
- Hua, X.-S., & Li, J. (2014). Tell me what. In *Proceedings of IEEE International Conference on Multimedia and Expo Workshops – ICMEW* (pp. 1-2). Chengdu: IEEE Publisher.
- Isik, S., & Ozkan, K. (2015). A comparative evaluation of well-known Feature detectors and descriptors. *International Journal of Applied Mathematics. Electronics and Computers*, 3(1), 1–6.
- Jung, K. (2012). Object recognition on mobile devices. In *Proceedings of the International Conference on Consumer Electronics – ICCE* (pp. 258-262). Berlin: IEEE Publisher.
- Kawano, Y., & Yanai, K. (2013). Real-time mobile food recognition system. In *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition Workshops – CVPRW* (pp. 1-7). Portland: IEEE Publisher.
- Kumar, S. S., Sun, M., & Savarese, S. (2012). Mobile object detection through client-server based vote transfer. In *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition – CVPR* (pp. 3290-3297). Providence, RI: IEEE Publisher. doi:10.1109/CVPR.2012.6248066
- Ling, W. H., & Seng, W. C. (2011). Traffic sign recognition model on mobile device. In *Proceedings of IEEE Symposium on Computers & Informatics – ISCI* (pp. 267-272). Kuala Lumpur: IEEE Publisher. doi:10.1109/ISCI.2011.5958925
- Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2), 91–110. doi:10.1023/B:VISI.0000029664.99615.94
- Quack, T., Bay, H., & Van-Gool, L. (2008). Object recognition for the internet of things. In *Proceedings of the first International Conference in Internet of Things – IOT* (pp. 230-246). Zurich: Springer Berlin Heidelberg. doi:10.1007/978-3-540-78731-0_15
- Rublee, E., Rabaud, V., Konolige, K., & Bradski, G. (2011). ORB: An efficient alternative to SIFT or SURF. In *Proceeding of the International Conference on Computer Vision – ICCV* (pp. 2564-2571). Washington, DC: IEEE Publisher. doi:10.1109/ICCV.2011.6126544
- Saeed, N. N., Salem, M. A. M., & Khamis, A. (2013). Android-based object recognition for the visually impaired. In *Proceedings of IEEE 20th International Conference on Electronics, Circuits, and Systems – ICECS* (pp. 645-648). Abu Dhabi, UAE: IEEE Publisher. doi:10.1109/ICECS.2013.6815497
- Tong, S., & Koller, D. (2002). Support vector machine active learning with applications to text classification. *Journal of Machine Learning Research*, 2(1), 45–66.

ADDITIONAL READING

- Akhtar, Z., Martinel, N., & Foresti, G. L. (2014). Mobile re-identification based on local feature analysis. In *Proceedings of the 8th ACM/IEEE International Conference on Distributed Smart Cameras – ICDSC* (pp. 1-6). New York, NY: ACM Publisher. doi:10.1145/2659021.2659037

Akhtar, Z., Micheloni, C., & Foresti, G. L. (2015). Biometric liveness detection: Challenges and open research opportunities. *IEEE Security & Privacy Magazine*, 13(5), 63–72. doi:10.1109/MSP.2015.116

Avola, D., Caschera, M. C., & Grifoni, P. (2006). Solving ambiguities for sketch-based interaction in mobile environments. In *Proceedings of the International Workshop On the Move to Meaningful Internet Systems – OTM Workshop* (pp. 904–915). Montpellier, France: Springer Verlag. doi:10.1007/11915034_115

Avola, D., & Paolozzi, S. (2009). Context Modelling Approaches for Mobile Systems. In I. Khalil (Ed.), *Handbook of Research on Mobile Multimedia* (2nd ed., pp. 364–378). Hershey, PA: Information Science Reference. doi:10.4018/978-1-60566-046-2.ch026

Borji, A., Cheng, M. M., Jiang, H., & Li, J. (2015). Salient object detection: A benchmark. *IEEE Transactions on Image Processing*, 24(12), 5706–5722. doi:10.1109/TIP.2015.2487833 PMID:26452281

Gupta, S., Arbeláez, P., Girshick, R., & Malik, J. (2015). Indoor scene understanding with RGB-D images: Bottom-up segmentation, object detection and semantic segmentation. *International Journal of Computer Vision*, 112(2), 133–149. doi:10.1007/s11263-014-0777-6

Martinel, N., Micheloni, C., & Foresti, G. L. (2013). Robust painting recognition and registration for mobile augmented reality. *IEEE Signal Processing Letters*, 20(11), 1024–1027. doi:10.1109/LSP.2013.2279014

Ruiz-Sarmiento, J. R., Galindo, C., & Gonzalez-Jimenez, J. (2015). Scene object recognition for mobile robots through semantic knowledge and probabilistic graphical models. *Expert Systems with Applications*, 42(22), 8805–8816. doi:10.1016/j.eswa.2015.07.033

KEY TERMS AND DEFINITIONS

Computer Vision (CV): The field of the computer science that deals with the analysis and understanding of images and videos. This field covers a wide range of practical and theoretical topics, including: image and video acquisition, feature extraction, image and video understanding, feature representation and feature matching. This field is also closely linked to a huge amount of scientific disciplines, including: artificial intelligence, machine learning and pattern recognition.

Feature Extraction (FE): The process for the extraction and description of salient points from images and videos. Any image, or frame of a video, can be represented by a set of keypoints (i.e., features) whose aim is to highlight and synthesize shapes, objects and properties contained within them. These keypoints are extracted from an image, by a detector algorithm, and, subsequently, they are exhaustively described, by a descriptor algorithm.

Feature Matching (FM): The process for the comparison of two sets of keypoints coming from two different images or video frames. The process compares the description of each keypoint of the first image (or frame) with each keypoint of the second image (or frame). A rank algorithm establishes a list of the best matching between them. This process is used to check the bi-univocal correspondence between the keypoints of two similar or overlapped images (or frames).

Global Position System (GPS): A radio navigation system that provides accurate positional data of an object equipped with a GPS receiver. The positional data is referred to the surface of the earth and include latitude, longitude and altitude along with the time. This basic information can be processed by the receiver to derive other dynamic data, such as: speed and acceleration. The working of the GPS system is due to a network of satellites deployed in the space that continually emit a measurable signal.

Human-Computer Interaction (HCI): The field of the computer science that deals with the interaction between users and computers. The aim of the HCI is the definition of interactive interfaces by which to guide any system (desktop or mobile) in respect to the usability principles.

IR Camera (IR-Cam): A camera equipped with an infrared (IR) technology (i.e., IR projector and IR sensor) through which the depth maps of persons and objects are built. The maps of these elements show the distance between them and the camera thus making the recognition process simpler with respect to the traditional approaches based on RGB cameras.

Mobile Vision (MV): The field of the computer science that deals with the analysis and understanding of images and videos focused on mobile devices (e.g., smartphones, tablets).

Natural User Interface (NUI): The field of the computer science that deals with the human-oriented interfaces. The term NUI highlights that these interfaces have to be invisible to the users. The NUI interfaces are designed to detect the natural actions of the human beings (e.g., hand movements, body poses, body gestures) and use them to interact with any kind of system.

Object Classification (OC): The field of the computer science that deals with the classification of objects. The classification is supported by a specific algorithm named classifier which can be implemented by using different theoretic principles, including: machine learning, statistical approaches, mathematical approaches, geometrical approaches and exhaustive computations.

Object Recognition (OR): The field of the computer science that deals with the recognition of objects. It consists of a set of algorithms, including: feature extraction, feature matching and object classification. OR regards both images and video frames.

RGB Camera (RGB-Cam): A camera equipped with a standard CMOS sensor through which the colored images of persons and objects are acquired. The acquisition of static photos is usually expressed in megapixels (e.g., 12MP, 16MP) that define the amount of pixels (i.e., length x height) that compose a photo. While, the acquisition of videos is usually expressed with explicative terms such as Full HD (i.e., 1080 x 1920 pixels with 30 frame per second) or Ultra HD (i.e., 3840 x 2160 pixels with 30/60 frame per second).

Mobile Apps Threats



Donovan Peter Chan Wai Loon
University of Malaya, Malaysia

Sameer Kumar
University of Malaya, Malaysia

INTRODUCTION

In recent times smart mobile devices have become ubiquitous. More than half of all the mobile phones are now smartphones, and this statistic does not take into account the other devices such as tablets that are operating on similar systems (Gates, Chen, Li, & Proctor, 2014). With the abundant usage of smartphones, the way we go around our daily lives has certainly been transformed. The smartphones of today are more like mini computers than mobile phones of a few years back. Essentially, smart phones are computers with additional hardware—namely, a Global System for Mobile Communications or GSM radio and a baseband processor to control it (Miller, 2011).

Although this is great, there is also a threat emerging. In 2012 alone, Google estimated that more than 400 million Android devices had commenced operations. Android devices have been adopted widely for both personal and business use (Wang, Sun, Wang, & Jing, 2015). From adults to children, beginners to experts, and in numerous different countries around the world, there is a diverse user base for mobile devices (Gates et al., 2014). Attackers are now targeting these devices in the same way computers have been targeted for a long time. The extensive usage of mobile devices poses new threats to privacy and security of our digital lives (Zhou & Jiang, 2012; Zhou, Zhang, Jiang, & Freeh, 2011). Email messages, contact lists, passwords and files are often stored

both locally and in the cloud. Illegal access to this private information by any unknown parties puts users at risk. Threats become even more dangerous as these devices may provide deep insights by integrating our digital to our daily lives. The GPS unit can pinpoint exact information of our whereabouts, while the microphone can record audio and the camera documents images (Khan, Xiang, Aalsalem, & Arshad, 2013). Moreover, mobile devices are frequently connected directly to monetary risks, via SMS authentication messages, as a means to validate financial transactions, or directly linked to bank account through a ‘digital wallet’ (Gates et al., 2014). Getting access would mean that any application (app) that operates on the devices has the potential to tap into and to provide certain details of information of the users.

The purpose of this article is to gauge the trends and challenges posed by malicious mobile applications. Specifically the authors look at the some of the current research to detect malicious applications and remedy for poor risk communications on Android-based devices.

The rest of the paper is organized as follows: In Section 2 we provide a background on mobile threats from an adapted threat model. Emerging mobile threats and an overview of mobile malware for Android and iOS is presented in Section 3 followed by an example of detecting malicious apps presented in Section 4. Future research directions are discussed in Section 5. Section 6 concludes the paper.

BACKGROUND

Understanding Mobile Threats

In order to offer a wide indication of threats facing mobile devices, it is first important to understand the objectives, reasons and distribution techniques of potential attacks. In this paper, we adapted a threat model from prior research by Delac, Silic, & Krolo (2011 p. 2-3) and divided into two main components: attack goals and attack paths. This model is further supported by a similar study in the same year by Leavit (2011 p. 11-13) and has similar descriptions of the main components.

Attack Goals

Attack goals are motives for penetrating mobile devices. The objectives may be hidden or destructive intents. Hidden attacks are executed while eluding a user's detection. Destructive attacks on the other hand, are meant to interfere with the usual function of the mobile device.

- **Collect Private or Personal Data:** Attackers usually target confidential information stored on the device. An effective attack may permit the attacker with capability to read SMS messages, emails to contact details and call history. Additionally, attackers may retrieve classified information by accessing applications stored on the device. Furthermore, once the mobile device has been compromised, attackers may use the hardware features to gather extra data from an individual's environments. For example, the attacker may use the microphone to record audio, the camera to take photos and pinpoint a user's exact location information through the GPS component.
- **Exploiting Processing Properties:** Attackers target mobile devices for the raw processing capabilities. Most modern mobile devices come with high-powered CPU and multi core processors. Coupled with high

speed Internet connectivity, mobile devices are appealing for malicious activities, such as distribution of botnets.

- **Destructive Activities:** These actions are meant to cause distress to the mobile users compared to benefiting the attackers. While easy to detect, attacks of this nature are designed at causing as much harm as possible. The attacks may vary from data corruption, draining the battery from the device, and generating huge network traffic (Delac et al., 2011). Eventually, by acquiring controls to key systems, attacks could even deactivate the device rendering it useless.

Attack Paths

Attack paths are routes provided by mobile platforms for the distribution of malevolent packages. Mobile malware is malicious software developed particularly to attack mobile devices. Most attacks done are often a combination of several variants of mobile malware.

- **Mobile Network Services:** Initially, attackers used cellular services such as SMS, MMS and voice calls. False URL links are sent via SMS to unsuspecting users. When a user clicks the link, it automatically opens a browser allowing the device to be prone to an attack. MMS messages facilitated the delivery of malicious content through hidden codes embedded within the message as users downloaded the content. Voice calls on the other hand allowed attackers to maintain contact with the users by portraying to be a legal entity such as a bank asking for sensitive information such as credit card details. This act is also known as "phishing".
- **Internet Access:** Most modern mobile devices have Internet access through the use of mobile networks such as 3G/4G or Wi-Fi networks. The chance for an attack rises

if the mobile device is connected to public Wi-Fi network. The major weakness is hotspot architecture with no encryption to protect transmitted data (Leavit, 2011).

- **Social Networking:** New Internet services such as social media platforms are also prime areas for the spreading of malicious contents. For example, an attacker may spread harmful links on social media by masking it with topics of interest. Users may be misled to download harmful content or diverted to unauthorized sites.
- **Bluetooth and USB:** Bluetooth is used to spread malware from one device to another. For attacks to occur, the intended mobile devices must first be within range to enable the attacker to connect and send the malicious content. Attacks using this method are limited because of the by default most devices are not discoverable and requires authentication from the users to transfer and physical download the file. Another widespread attack is done using USB. Generally, a USB connection is used to pair mobile devices with computers for the transfer of files. Malicious content embedded in certain software applications is then transferred from the computer into the mobile devices.

Emerging Threats to Mobile Platforms

According to a report by Kaspersky Network Security Data, Kaspersky Lab products had detected 2,961,727 infections in 2015. There were a total of 884,774 new modifications of malicious programs targeting mobile devices; three times higher than the previous year. This indicates that the mobile malware segment has advanced swiftly in the past few years. Cybercrimes may have evolved from being acts of solo hackers to becoming an underground commercial organization. There are numerous kinds of players involved in the mobile malware industry: virus writers, testers, interface

designers the malicious applications and the web pages they are circulated from, and mobile botnet owners (Chebyshev & Unuchek, 2014).

Among the top trends for mobile malware in 2015 as highlighted by (“Mobile malware evolution 2015 - Securelist”, 2016) are: -

- **Evolution of Malware Towards Monetization:** Mobile Trojans targeting user bank accounts continue to grow with 7,030 new mobile banking Trojans detected in 2015.
- **Cybercriminals Actively Using Phishing Windows to Conceal Legitimate Applications:** One of the most notable examples is the Trojan-SMS.AndroidOS.OpFake.cc variation that is able to imitate the interface of 100 legitimate banking and finance applications.
- **Growth in the Volume of Ransomware:** Trojan-Ransom applications doubled in 2015 from the previous year, while new the amount of new modifications had increased 3.5 times. Another key indicator is the amount of people attacked in 2015, which increase five times compared to the year before.
- **Malware in Official Stores:** In October 2015, Kaspersky detected several Trojans in the official Google Playstore that stole passwords from the Russian social network VKontakte. Attackers had published these Trojans 10 times in the official app store using different names. The number of estimated downloads of these Trojans was between 100, 000 and 500, 000.
- **Malware for iOS:** 2015 saw the number of malicious programs for iOS increase 2.1 times compared to 2014. Contrary to popular belief, this demonstrated that iOS is indeed vulnerable to malware. Attackers did not hack the App Store but posted a malicious version of Apple’s Xcode. This had been spread in China affecting one of the most popular applications; a free mes-

senger “WeChat” installed on more than 700 million users devices. A Trojan was also detected in Japan using phishing techniques to steal money from users.

Among the nations in the world, China and Nigeria topped the ranking of countries facing malware with 37% in both countries (Chebyshev & Unuchek, 2016). Most of the attacks were from Trojan families and could lead to financial loss through withdrawals from mobile accounts of unsuspecting users. In addition to that, percentages of recorded attacks recorded also depended with the number of users in the country.

Mobile Malware: Android and iOS

Both iOS and Android offer a public marketplace for consumer applications. Apple has the App Store and Android uses the Google Play Store respectively. The two however adopt considerably different methods to minimize malware on their devices.

iOS

In standard Apple manner, the App Store is securely controlled from top to bottom. Apple must first endorse an application before it can be included in the App Store. Apple does this on the device by implementing code signing. iPhones will not be able to load an application or library if it's without Apple's private encryption key signature. Therefore only Apple knows how closely it reviews iOS apps. According to Miller (2011), Apple cannot be assessing the applications all that cautiously considering the number of applications, but any sort of review will remove the most apparent malware. Should a piece of malware manage to go through the review and enter the App Store, and users detect it, Apple could delete it from the App Store and the devices it was installed in. Although some may argue that the App Store is bad for developers, it has an efficient block to malware. Once installed on the device, applications operate

within a sandbox that limits their actions (Miller, 2011). For instance no application can read the saved SMS (short message service) messages, or one application cannot access another's data. All applications share common sandbox rules, so any action any application needs is approved. One example is that all applications can easily access the Internet.

Android

In the Android Market, developers are able to load their applications directly without any prior review done. Although Android phones need applications to be signed, they only require a partial signature. Google primarily uses the signatures for recording purposes and not to control the codes. Due to the open nature of Android, users may download apps from any source, not just the Android Market. For malware prevention, Android employs crowd sourcing instead of using the top-down approach. Applications are rated and commented by users. Users can see how many other people have used an application and can report any malicious applications to Google. If sufficient users give negative feedback regarding an application, Google will delete it from the market and from the devices. Android users are recommended not to ever download an application that does not have thousand of downloads and have mostly good comments. There have been several malicious Android apps, but most of them mainly existed in other markets than the Android Market. The openness of Android permits developers' ease of use but also but also allows easy entrance by malware developers (Zhou, Wang, Zhou, & Jiang, 2012).

When the application is on the device, Android too employs the sandbox model, however the sandbox is specific for applications. Before installation, applications will prompt users about which permissions are needed. These permissions can then be approved or declined by users. Applications that are declined will not be installed. One advantage is that these sandboxes can be personalized for each application instead of Apple's

method of one-size-fits-all (Miller, 2011). For example, a Calculator application does not require Internet access and so it will not be granted that access. The drawback, however, is that, in this model, users will have to make their own security decisions. These choices made are bad than good. Moreover, the reasons users are trying to install those applications are because they want them in the first place. This creates the tendency to click through the screens without doing much check on the requirements.

SOLUTIONS AND RECOMMENDATIONS

Detecting Malicious Applications (Android Market)

There are several researches done in detecting malicious applications in both official and alternative Android Markets. For the purpose of this paper, we selected the model design of “Droid Ranger” as proposed by (Zhou et al., 2012). The study aimed at presenting a logical way to understand and detect malicious applications of the overall Android Markets. It was conducted in June 2011 over a period of two months and covered a total

of 204,040 applications. The researchers used the following approaches in detecting malware:

- Permission-Based Behavioral Footprinting:** The scheme filters applications based on inherent Android permissions of malware and matching those with malware specific behavior.
- Heuristics-Based Filtering Scheme:** Defines suspicious activity of applications such as suspicious attempts to retrieve and perform codes from untrusted websites. Determines whether the application is benign or malicious based on monitoring runtime execution.

Design

To thoroughly identify malicious applications in the existing Android Markets, the researchers used three definitive design goals. The goals were accuracy, efficiency and scalability. Accuracy is a normal requirement of detecting negative malicious applications in the current market places. Efficiency and scalability were more complex. As described by Zhou et al. (2012) there were 200, 000 applications that needed to be scanned. It took two weeks to scan the entire collection.

Figure 1. Overall architecture of DroidRanger
Zhou et al., 2012.

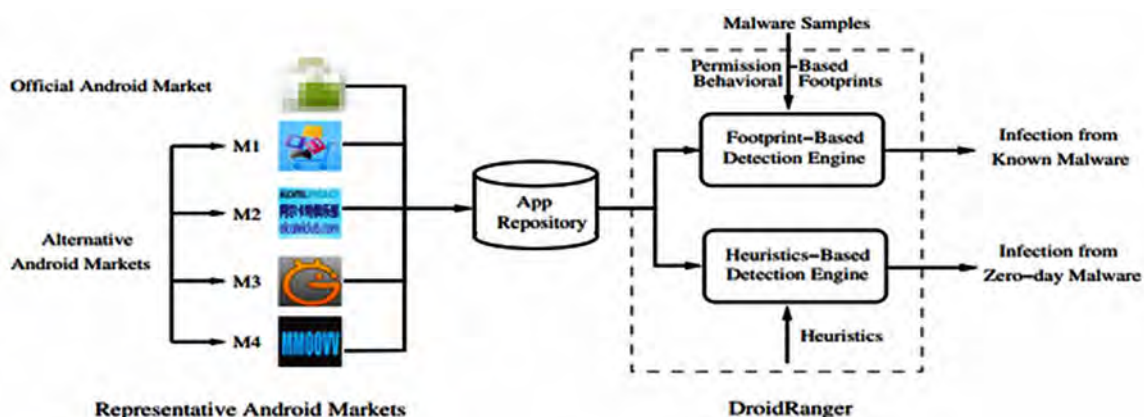


Figure 1 illustrates a summary of the approach. DroidRanger collects applications from current Android Markets and stores them into a local repository. DroidRanger then extracts the fundamental properties of each application and analyses them according to the two approaches mentioned earlier. For known malicious malware, the DroidRanger categorized indexed them with the applications itself. Unknown malwares on the other hand were classified under “Zero-day Malware. A central database was used for effective indexing and lookup.

The research yielded good results and made a number of contributions. The first contribution was the DroidRanger pioneered a systematic study dedicated to detecting malicious applications on Android markets. Secondly was the extensive study of 200,040 Android applications. The researchers proposed two different schemes that allowed for efficient scalable detection of both known and unknown malicious applications. Finally, when applied, the DroidRanger had managed to successfully detect 211 malicious applications. Reporting from the researchers resulted in all those applications being removed from the Android marketplace and perhaps saving thousands of users from being affected by malicious malware

Effective Risk Communication for Android Applications

Recent efforts have tried to remedy some of the issues in poor risk communication towards android applications. Among the enhancements proposed by Felt et al. (2012) were to modify permission category headers, emphasize on risk, enable customized permission lists, incorporate user reviews, reduce the number of permissions and rethink the timing of how and when permissions permitted. Lin et al. (2012) suggested a method using crowd-sourced expectations of which authorizations are considered sensible and display it at the applications main description page. Kelley, Cranor, and Sadeh (2013) presented an idea of “privacy facts” which carries out at high-level types of information

an application has to access and display it on the applications main description page. However the present risk communication tools of Android are insufficient and lacking in terms of effectiveness. Users are still inclined to disregard the permissions that an application accesses.

In this section, the authors look into the alternative approach to the problem as recommended by Gates et al. (2014). The aim of the approach is to assist users to understand security and privacy consequences, aiding them to make installation decisions and reduce the amount of area used to show the risk information. Adding a short risk rating for the applications would enable users to compare risk among applications that have related functions. Gates et al. (2014) deduced that the main reason why users often ignored current permission information was due to way in which the information was presented. Information was in a separate section and required users to have a substantial amount of technical understanding. Users would also need much more time interpreting the information, thus creating problems in comparing applications. One significant feature in the Android ecosystem is that users have a wide selection and options when choosing a mobile application. Given the knowledge that a certain application may be riskier than another, which has the same functionality, the user may tend to choose the less risky application. Application designers will then conform and follow the fundamental standards by only requesting basic permissions.

Peng et al. (2012) offered a probable way for producing a principled metric to rank an application risk centered on the set of permissions it requests. The way it is done is that any Android application available Google’s online market (Google Play Store) is ranked in a summary risk rating. The risks are categorized into different values ranging from very low, low, medium and high risk. When a user looks for an application, the summary risk rating will provide brief information helping the user make an evaluation before proceeding to download the application. This is different from the current approach, where

Table 1. Risk rating experiments

Experiment	Testing	Results
1. Adding a risk metric	- Tested the amount to which having a risk category affects users' choices. - The behavior of users when were provided with the summary risk was compared to when there was no summary risk information.	Tests confirmed that specifying risk summary triggers users preference towards applications with lower risk
2. Safety versus risk	- Compared ease of processing the rankings in terms of risk or safety - Used an in-lab choice-	Results showed that responses were faster in the safety condition than the risk condition
	Reaction task for which response time and accuracy were measured.	
3. Safety versus risk with user ratings	- Similar to Experiment 2. - Displayed risk information and user ratings together in application context.	Mixed results showing responses were faster in the safety condition than in the risk condition. The interaction was not significant.
4. Safety versus risk in context	Determine whether symbolic presentation of risk influenced choices in a normal environment	Test confirmed overall participants chose the safer/less-risky apps more often

Conducted by Gates et al. (2014).

risk information of an application often appears to the user in the final stage when the user has already chosen the application. Gates et al. (2014) conducted four different experiments in testing their approach.

Table 1 summarizes the tests conducted.

In Experiment 2 and 3, the researchers had to decide beforehand which structure to frame the categories; either more symbols signify greater risk or more symbols signify better safety (safety dimension). Since users generally associate “more stars with being better”, the framing dimension used was that of the safety dimension. After conducting all four experiments, the hypothesis that when a summary risk rating is presented in a user-friendly fashion, it will encourage users to choose applications with lower risk had proved to be true.

FUTURE RESEARCH DIRECTIONS

The increasing usage of mobile devices today for numerous tasks such as online banking makes malicious malware very attractive for attackers. Hackers can now use various methods to launch attacks and steal financial data or cause financial loss. Mobile Malware has now evolved into a

moneymaking industry for hackers at the expense of unsuspecting users. In many cases, new modifications of mobile threats are constantly being developed and combined with ultimate goal of reaping profits from illegal sources.

The research has significance in that explains the main objectives, methods, and types of mobile threats faced by users today. Although research on mobile threats is rising, there are currently limited studies on how to overcome it. Available literature on the matter is focused on within the technology savvy community rather than the general public. This study is an attempt to highlight the mobile threats as well as create awareness amongst users on its detrimental effects.

There are several limitations with this study. In this research, only a model of detecting malicious applications via risk communication had been discussed. This could be enhanced through other studies such as biometrical security for users, which would more secure when downloading application. This research also only highlighted preventions on one mobile platform that was Android due to its popularity for mobile devices. For future research direction, an analysis including alternate platforms such as iOS may provide a better insight and understanding towards this study.

CONCLUSION

Smartphones are gradually becoming more and more valuable in our everyday life. They keep people connected and no one is ever lost or unreachable. The cutting-edge technology and functionality of this gadget has made it an appealing target for malicious and invasive applications. Even though solid security measures are there for most mobile infrastructure, the part where these systems often fail is the dependence on the user to make decisions that affect the security of a device. There is also a need for tighter policy procedures, especially for non-regulated alternative marketplaces.

REFERENCES

- Ben-Asher, N., Kirschnick, N., Sieger, H., Meyer, J., Ben-Oved, A., & Möller, S. (2011, August). On the need for different security methods on mobile phones. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services* (pp. 465-473). ACM. doi:10.1145/2037373.2037442
- Bottazzi, G., Casalicchio, E., Cingolani, D., Marturana, F., & Piu, M. (2015, October). MP-Shield: A Framework for Phishing Detection in Mobile Devices. In *Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing (CIT/IUCC/DASC/PICOM), 2015 IEEE International Conference on* (pp. 1977-1983). IEEE. doi:10.1109/CIT/IUCC/DASC/PICOM.2015.293
- Chebyshev, V., & Unuchek, R. (2014). Mobile malware evolution: 2013. *Kaspersky Lab ZAO's SecureList*, 24.
- Delac, G., Silic, M., & Krolo, J. (2011, May). Emerging security threats for mobile platforms. In *MIPRO, 2011 Proceedings of the 34th International Convention* (pp. 1468-1473). IEEE.
- Felt, A. P., Ha, E., Egelman, S., Haney, A., Chin, E., & Wagner, D. (2012). *Android permissions: User attention, comprehension, and behavior*. Paper presented at the Eighth Symposium on Usable Privacy and Security. doi:10.1145/2335356.2335360
- Gates, C. S., Chen, J., Li, N., & Proctor, R. W. (2014). Effective risk communication for android apps. *Dependable and Secure Computing. IEEE Transactions on*, 11(3), 252-265.
- Kelley, P. G., Cranor, L. F., & Sadeh, N. (2013). *Privacy as part of the app decision-making process*. Paper presented at the SIGCHI Conference on Human Factors in Computing Systems. doi:10.1145/2470654.2466466
- Khan, W. Z., Xiang, Y., Aalsalem, M. Y., & Arshad, Q. (2013). Mobile phone sensing systems: A survey. *IEEE Communications Surveys and Tutorials*, 15(1), 402-427. doi:10.1109/SURV.2012.031412.00077
- Lin, J., Amini, S., Hong, J. I., Sadeh, N., Lindqvist, J., & Zhang, J. (2012). *Expectation and purpose: understanding users' mental models of mobile app privacy through crowdsourcing*. Paper presented at the 2012 ACM Conference on Ubiquitous Computing. doi:10.1145/2370216.2370290
- Miller, C. (2011). Mobile attacks and defense. *Security & Privacy, IEEE*, 9(4), 68-70. doi:10.1109/MSP.2011.85
- Peng, H., Gates, C., Sarma, B., Li, N., Qi, Y., Potharaju, R.,... Molloy, I. (2012). *Using probabilistic generative models for ranking risks of android apps*. Paper presented at the 2012 ACM conference on Computer and communications security. doi:10.1145/2382196.2382224

Wang, X., Sun, K., Wang, Y., & Jing, J. (2015). *DeepDroid: Dynamically Enforcing Enterprise Policy on Android Devices*. Paper presented at the NDSS. doi:10.14722/ndss.2015.23263

Zhou, Y., & Jiang, X. (2012). *Dissecting android malware: Characterization and evolution*. Paper presented at the Security and Privacy (SP), 2012 IEEE Symposium on. doi:10.1109/SP.2012.16

Zhou, Y., Wang, Z., Zhou, W., & Jiang, X. (2012). *Hey, You, Get Off of My Market: Detecting Malicious Apps in Official and Alternative Android Markets*. Paper presented at the NDSS.

Zhou, Y., Zhang, X., Jiang, X., & Freeh, V. W. (2011). *Taming information-stealing smartphone applications (on android)*. In *Trust and Trustworthy Computing* (pp. 93–107). Springer.

ADDITIONAL READING

Arslan, B., Gunduz, S., & Sagioglu, S. (2016). A review on mobile threats and machine learning based detection approaches. In *2016 4th International Symposium on Digital Forensic and Security (ISDFS)* (pp. 7-13). IEEE. doi:10.1109/ISDFS.2016.7473509

He, D., Chan, S., & Guizani, M. (2015). Mobile application security: Malware threats and defenses. *IEEE Wireless Communications*, 22(1), 138–144.

Kearns, G. S. (2016). Countering Mobile Device Threats: A Mobile Device Security Model. *Journal of Forensic & Investigative Accounting*, 8(1).

Kearns, G. S. (2016). Countering Mobile Device Threats: A Mobile Device Security Model. *Journal of Forensic & Investigative Accounting*, 8(1).

Li, Q., & Clark, G. (2013). Mobile security: A look ahead. *IEEE Security and Privacy*, 11(1), 78–81. doi:10.1109/MSP.2013.15

KEY TERMS AND DEFINITIONS

Botnets: A Botnet is a set of infected devices, which can be ordered and managed remotely by an attacker. Attacks may range from sending spam mail to committing DOS attacks (Leavitt 2011).

Mobile Malware: (McAfee 2016) describes Mobile Malware as malicious software specifically aimed at attacking mobile devices.

Phishing: Phishing as “a type of attack that communicates socially engineered messages to humans via electronic communication channels in order to persuade them to perform certain actions for the attacker’s benefit” (Bottazzi et al., 2015).

Ransomware: Ransomware is a type of malware that severely restricts access to a computer, device or file until the user pays a ransom (Kaspersky Securelist, 2015).

Trojan: Trojans can be used to collect confidential information or to install other malicious applications like worms or botnets. Trojans can be also used to commit phishing activities. (Sujithra & Padmavathi, 2012).

Worms: Worms are self-replicating malicious applications designed to spread autonomously to uninfected systems. (Delac, Silic, & Krolo 2011).

Mobile Technologies Impact on Economic Development in Sub-Saharan Africa

Adam Crossan

Letterkenny Institute of Technology, Ireland

Nigel McKelvey

Letterkenny Institute of Technology, Ireland

Kevin Curran

Ulster University, UK

INTRODUCTION

In today's world mobile technology has been used to help improve and enhance almost every sector, in this paper the author researches and discusses how it has helped benefit developing areas such as sub-Saharan Africa, the paper outlines the constraints of mobile technology in sub-Saharan Africa. This paper focuses mainly on how the use of mobile technology has improved the health and education standards in different areas. It outlines how mobile devices are being used to educate those all over sub-Saharan Africa on how to protect against diseases such as polio. This paper also details how mobile devices are being utilized to help provide better healthcare to those in African cities, through topping up smart cards with mobile devices for the distribution of clean water as well as the delivery of vouchers which can be exchanged for healthcare equipment. This report also discussed how mobile devices are being used to track and monitor the outbreak of malaria in sub-Saharan Africa and is applied to a system to help ensure health care facilities in areas with affected residents have the capacity to deal with those who have contracted the disease. Finally, this paper discusses the ethical effects of using the mobile devices and how they are affecting the residents of sub-Saharan Africa.

BACKGROUND

Mobile Technology is a sweeping term covering many different areas of technology. Many of the uses for this technology in developed countries include helping with general day to day operations for the user such as online banking, shopping, or social media applications. Mobile technology is a sector which is growing rapidly, and also merging with other sectors to help make use of the technologies full potential. One of the areas where mobile technology has been proven to be successful, is when using simple technology and mobile devices to help developing countries such as Africa (Aker & Mbiti, 2010). Many different organizations and companies are using mobile technology to accomplish huge feats for rural African villages. A leading success story has been the M-Pesa service launched in 2007 by Vodafone for Safaricom and Vodacom. This has generated mobile money revenues of more than \$300 million in 2004. This makes it the most successful mobile transfer service at this time. It has now expanded to Afghanistan, South Africa, India and Eastern Europe. It seems to have found its niche market. In Kenya alone, mobile money transactions totalled \$22+ billion in 2013. Other African countries such as Sudan, Somalia, Tanzania, and South Africa are also huge revenue generators for operators as well (BBC Africa, 2015).

Mobile Tech in Africa

In today's world when discussing developing countries, the most commonly mentioned country is Africa. Africa is known worldwide for its widespread poverty, levels of education, unsanitary living conditions, gang violence and high mortality, and birth rates. Many companies and organizations are now turning to mobile technology to help combat these issues facing those living in Africa. With the increase in advancements in technology, many organizations and companies are researching and looking into how some technologies could be adapted to help reach areas which are usually not easily accessible. A simple yet effective method at developing communication with these areas is through mobile devices. In recent years the ownership of mobile phones in Africa have increased rapidly, with the percentage of adults who own a phone rising from 64% in 2002 to now match the US with 89% in 2014 (Global, 2015). An infographic depicting this information can be seen in Figure 1. As shown in reports conducted by Pew Global (2015), the most popular activity for mobile phone users in Africa is sending text messages. The report shows that 80% adult cell phone owners use their phone mostly to send text messages as seen in Figure 2. SMS is arguably the most popular activity in this area due to the fact that many in this area only have ownership of a cell phone rather than smartphone, with a study conducted on Nigerian residents showing that although 89% of the population there owned a mobile phone only 27% owned a smartphone (Pew Global, 2015).

Constraints Affecting Mobile Devices in Africa

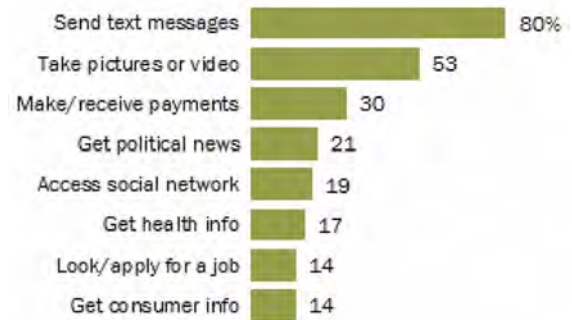
Africa is known for its lack of proper infrastructure, and industrialization. Technology isn't something that is commonly associated with Africa. The main contributing factor for this being that mostly all technology needs to be powered in some way. A study by The World Bank (2013), showing that

Figure 1. Percentage of what adults use their mobile devices for in sub-Saharan Africa
As stated by Pew Global (2015).

M

Texting Most Common Use of Cell Phones in Africa

Median adult cell phone owners who used a cell phone in the past 12 months to ...



Note: Median percentages across seven African countries.

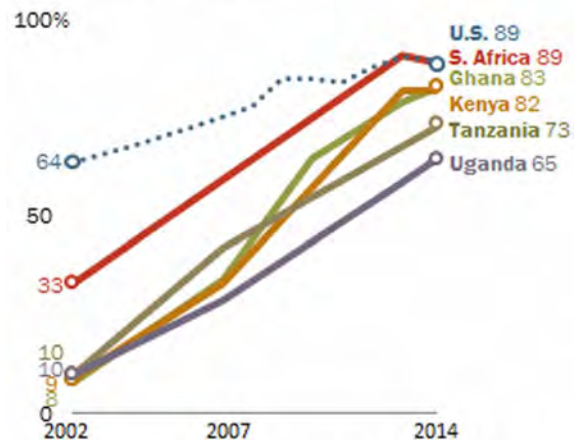
Source: Spring 2014 Global Attitudes survey. Q74a-h.

PEW RESEARCH CENTER

Figure 2. Infographic depicting the number of adults who own a mobile device globally
As stated by Pew Global (2015).

Cell Phone Ownership Surges in Africa

Adults who own a cell phone



Note: U.S. data from Pew Research Center surveys.

Source: Spring 2014 Global Attitudes survey. Q68.

PEW RESEARCH CENTER

76% of sub-Saharan Africa is without access to electricity, the possibility of being able to use technology within these areas decreases drastically. Another area of technology which is obstructing the growth of use of technology in Africa is connectivity. A study by iPass, (2014) shows that if growth in connectivity hotspot numbers continue to grow at the same rate as previous years then by 2018 there will be a WiFi beacon for every 20 people on earth, however only one beacon for every 400 Africans.

Education

The use of mobile technology has been hugely influential in the improvement of education throughout Africa. The need for helping improve education throughout sub-Saharan Africa has been a main priority for many organizations, one of which was Oxfam. Oxfam are a charity organization which help in delivering medical aid, supplies and help to a number of different countries including Africa (Oxfam, 2016). One particular project which Oxfam began was to help improve sub-Saharan resident's knowledge on different diseases which were quite common in the areas however easily preventable. Oxfam developed an SMS delivery system, which would send out four texts daily with educational information on the different diseases and how to prevent against them. This was incredibly simple as the only technology needed for the residents was a simple mobile phone which could send and receive SMS.

Health

Mobile technology is also being used to help better the health services in areas of sub-Saharan Africa. Organizations and companies such as UNICEF are using simple mobile technology to help teach about how to prevent from contracting diseases, help improve distribution of medical supplies and equipment to people who would otherwise not have access to the materials. They are also helping to monitor and ensure medical facilities contain the

stock levels needed to combat against diseases. Two of the main diseases technology had been used to battle the spread of was polio and malaria.

Polio

Alongside the project discussed earlier to help better educate areas of Africa about the polio and how it can be contracted, a project was also launched that would help provide the equipment and sanitary supplies needed to help prevent the spread of the disease to rural and hard to reach areas. The project would also work through SMS where the users would receive a SMS with a mVoucher. As stated by Textello (2016), the mVoucher or mobile voucher is a unique code, each code is created using the customers contact number, resulting in each customer and code being linked, this then reduces the problem of multiple claims, and users mVouchers being stolen. The user would then present their code to a registered trader, where they would then in return receive the specified package containing the items and supplies needed (Oxfam, 2015).

Malaria

To help combat against the threat of malaria in different areas of Africa another project also using SMS was created. This project was however to help survey the outbreak of the disease, and help get real time accurate updates on the impact of the disease. A system to combat malaria such as this was needed as malaria is still a large threat, particularly to children as indicated by the World Health Organization (2015), detailing that a child dies every two minutes in sub-Saharan due to malaria. Rather than simply just supplying information on the disease and vouchers for equipment to prevent from contracting the disease like done with the polio SMS system, the new system would "improve access to essential malaria medicine in areas of developing countries" (Novartis, 2014). The new SMS system created was called "SMS for life". SMS for life uses common technology

to help achieve its goals of eliminating medical facilities from depleting stock levels completely, increase access to essential medicines needed for those infected with the disease and to help reduce the amount of deaths caused by malaria.

SMS for life is a web based system which makes use of a SMS management tool, this then links with a web based application created for reporting. The system uses the SMS management tool to store the number of a health worker from each of the medical facilities registered in the area, SMS for life then uses a short code number that allows the health facility employee to receive an SMS with a data request, and send an additional SMS with the requested data counts completely free. The data sent from the health worker is then manipulated by the online reporting system, which makes it available for viewing online by a higher body, usually the District Health Management Team. The higher body view this information on the SMS for life website by password, this allows a secure method of monitoring the malaria surveillance in each medical facility in the area (Githinji, et al., 2014).

Water ATM's

In some areas in Africa, local governments have teamed with Grundfos, a Danish water engineering company to install “water ATM’s”. The water ATM’s work by have having a water dispenser installed with users having an account with the dispenser. Each user then receives a smart card, these smart cards and accounts are topped up through mobile phones. The water dispensers can then be activated by the smart card. The user simply scans or swipes the smart card to be dispensed clean, safe water (Wesangula, 2016).

This project has brought water to many different areas of rural Africa and has now began to be installed in more urban areas such as Nairobi (BBCAfrica, 2015). As stated by Mourdoukoutas (2015), there are still some issues with the systems however Rasoul Mikkelsen, Grundfos director for global partnerships believes that “he project

will have a huge positive impact on health. It is an affordable system with sustainable, good quality water,”.

As well as the projects discussed above there have also been systems developed using SMS to help monitor Zambia for measles outbreaks, allow HIV sufferers health workers to send information on patient status as well as inform their patients details of their anti-retroviral therapy (Aker & Mbiti, 2010).

Ethical Effects

It is clear to see that the results of the Polio projects mentioned above have been extremely successful. An evaluation report by Oxfam (2015) stated that the project had a reach of 104,358 residents receiving the text which in turn was estimated with a reach of 1 million people, this was a completion rate of 86%. These people are now much more educated on how to better protect themselves against Polio, giving not just those who received the informative SMS messages but also those who were in communication with them, a better standard of living and a greater opportunity at life. Also with the pairing of the mVouchers for sanitary supplies and equipment to prevent contraction, with the educational messages it gives an incentive, as it gave the participants of the scheme something physical alongside the messages, something that they can then use the information received to put into practice (Oxfam, 2015).

However, there were some issues surrounding the project. The mVoucher system feedback report shows that the amount of participants that received the voucher compared to the amount of vouchers collected varies greatly with 71.4% of those who received a voucher declining to redeem it (Oxfam, 2015). Oxfam distributed a household survey to each of those who participated in the project, the survey asked the participants if they had collected their supplies with the voucher and if not then why? The main reason given for not claiming the supplies was “Distance to distribution Centre”, this highlights how these projects

seem to be specifically targeted at more urban areas rather than the outer lying rural areas of sub-Saharan Africa which is usually in greater need of the aid. This can also be seen in the number of distributors in each district, when the project began it was outlined that each district was to have 3 traders and 15 distributors, however this did not happen in every district, because of the lack in distributors the resulting distribution was below expected capacity, with fewer distribution points (Oxfam, 2015).

Issues can also be seen with the water ATM's which have been implemented in certain cities throughout Africa. Although the ATM's have achieved their goal of delivering safe clean water to different parts of the city, they too can be seen to be focused on the urban areas. As discussed in an article by Mourdoukoutas (2015), it states that Mr. Gichuki, the managing director of the Nairobi Water and Sewerage Company, said that "the Nairobi Water and Sewerage Company has so far identified 611 more points to install the dispensers and aims to eventually install one every 100 meters throughout the city". Meaning the residents from more rural secluded areas are left having to travel further for clean water when those who live in the cities have multiple choices for where to receive their water from, as well as having to wait longer before the installation of a water ATM in a nearer location. This again shows the imbalance between the urban and rural parts of sub-Saharan Africa.

FUTURE RESEARCH DIRECTIONS

The effects of the increased use of mobile technology can be seen in many areas, and with numbers showing that Africa has had the fastest growing region in the years 2009 to 2014, and as stated in a report by GSMA intelligence (2014), is predicted to continue to grow at a rate of 7% annually until 2020. These figures show the level of growth expected for Africa however when looking at figures relating to each individual country a drastic imbalance can be seen. As shown in the report by

GSMA intelligence (2014), the breakdown of the number of subscribers to mobile phone providers show the harsh difference between the more urban countries and the more rural countries. For example, the percentage of subscribers in Mali is 70%, compared to the considerably smaller percentage of 22% in Ethiopia. This large margin shows the growing issue of the difference between these two areas, an issue which needs to be examined. As stated in a report by Aker & Mbiti (2010), more examination of the possible impacts that could occur is needed, the report suggested that "careful evaluations of mobile phone development projects are required for better understanding of the economic and social outcomes".

CONCLUSION

From researching and examining how mobile technology has helped improve sub-Saharan Africa, in the different areas of health and education, it is clear that the technology and the projects it was used in has greatly benefited the African residents. The educational system created to aid those at risk of polio helped residents learn cleaning procedures and how to use the sanitary items provided with the mVouchers also distributed through mobile devices also. The use of mobile devices has also helped service those suffering from malaria, as it ensures that those needing treatment's local health facility will always be within capacity, thanks to the "SMS for Life" system. Finally, the statistics outlining the difference from the rural and urban areas of sub-Sahara, show how the effect differs from region to region. To conclude the paper, it is evident from the information presented that mobile technology has greatly helped improve different areas of life such as education and health. However conversely the effects of the increased use of mobile devices in Africa has contributed to a poverty imbalance and perpetuating health inequalities between the urban and rural areas of sub-Saharan Africa, and its possible impacts is an area that needs to be greatly researched before any implementation of larger projects can begin.

REFERENCES

Aker, J. C., & Mbiti, I. M. (2010). Mobile Phones and Economic. *The Journal of Economic Perspectives*, 22–23.

BBCAfrica. (2015). *Kenya slum Mathare gets cheap water through ATMs*. Available at: <http://www.bbc.com/news/world-africa-33223922>

Githinji, S. (2014). Using mobile phone text messaging for malaria surveillance in rural Kenya. *Malaria Journal*, 13(107), 1–8. PMID:24642130

Global, P. (2015). *Cell Phone Ownership Surges in Africa*. Available at: <http://www.pewglobal.org/2015/04/15/cell-phones-in-africa-communication-lifeline/africa-phones-7/>

GSMA Intelligence. (2014). *The Mobile Economy - Sub-Saharan Africa*. London: GSMA.

iPass. (2014). Available at: <http://www.ipass.com/wifi-growth-map/index.html>

Mourdoukoutas, E. (2015). *ATM-operated water dispensers bring safe water to poor communities*. Available at: <http://www.un.org/africarenewal/magazine/december-2015/africa-wired>

Novartis. (2014). *Innovation home*. Available at: <http://malaria.novartis.com/innovation/sms-for-life/index.shtml>

Oxfam. (2015). *Oxfam evaluation report*. Oxford, UK: Oxfam GB.

Oxfam. (2016). *Our purpose and beliefs*. Available at: <https://www.oxfam.org/en/our-purpose-and-beliefs>

Pew Global. (2015). *Few Own Smartphones in Africa, But Cell Phones Common*. [Online] Available at: <http://www.pewglobal.org/2015/04/15/cell-phones-in-africa-communication-lifeline/africa-phones-5/>

Pew Global. (2015). *Cell Phones in Africa: Communication Lifeline*. Available at: <http://www.pewglobal.org/2015/04/15/cell-phones-in-africa-communication-lifeline/>

Textello. (2016). *mVoucher*. Available at: <http://www.textello.com/features/mvoucher/>

The World Bank. (2013). *The World Bank and Energy in Africa*. Available at: <http://web.worldbank.org/WBSITE/EXTERNAL/COUNTRIES/AFRICAEXT/0,contentMDK:21935594~pagePK:146736~piPK:146830~theSitePK:258644,00.html>

Wesangula, D. (2016). *Mail & Guardian*. Available at: <http://mg.co.za/article/2016-02-18-nairobi-slum-gets-water-from-atms>

World Health Organization. (2015). *African populations are extremely affected by malaria*. Available at: <http://www.afro.who.int/en/clusters-a-programmes/dpc/malaria/features/2287-10-facts-on-malaria-in-africa.html>

KEY TERMS AND DEFINITIONS

Advance Fee Scam (Nigeria 419): An advance fee fraud which involves a payment in advance with the promise of wealth, wills, gifts, employment or prizes. The scammer contacts their intended target via mail, phone, fax or email. Originally known as the Spanish Prisoner Letter it has been carried out since at least the 16th century via ordinary postal mail. Section 419 of the Nigerian Penal Code prohibits this type of activity and it has often referred to as ‘419 frauds’ or ‘Nigeria 419’.

Developing Countries: Often referred to as underdeveloped countries and they refer to countries where there is a less developed industrial base and they have a low Human Development Index (HDI). Developing countries differentiate from developed countries in that the people have a lower life expectancy, lower standard of education and the people of developing countries have a lower income and thus less money.

Ethical Standards: Principles which when followed will promote values such as trust, good behaviour, fairness and/or kindness. One consistent set of standards that all companies follow does

not exist but rather each company has the right to develop a set of standards which are meaningful to that particular organization.

Information and Communication Technology (ICT): An umbrella term which includes any communication device or application, encompassing: radio, television, cellular phones, computer and network hardware and software, satellite systems and so on, as well as the various services and applications. The term ICT however has no set definition as the concepts methods and applications involved in ICT are constantly changing almost on a daily basis.

Mobile Point of Care (MPOC): Involves the use of multitasking devices which support clinician use cases, improve productivity, streamline communication and bridge the gap between in-patient and outpatient care. This fast growing area provides an abundance of apps for mobile devices and will change the way healthcare is delivered providing the opportunity for remote monitoring, measuring drug response, tools for drug development and diagnostic assistance for point-of-care and management of chronic or infectious disease.

Mobile Technology: A technology which is used for the purpose of cellular communication. The standard mobile device has moved from being

a simple two-way pager to being a mobile phone. Mobile technology as the name implies refers to a technology which is portable. Examples of mobile IT devices include: laptops, netbook computers, tablets, smart phones, global positioning system (GPS) devices

Personal Data: Data relating to a living individual who is or can be identified from the data or using the data in conjunction with other information which is in or likely to come into the possession of the data controller.

Short Message Service (SMS): A simple text messaging service component found in phones, the Web, or in mobile communication systems. It adopts a series of standardized communications protocols which enables either fixed line or mobile phone devices to exchange short text messages. More commonly referred to using the term “SMS” it is used for both user activity and all types of short text messaging in many parts of the world.

Sub-Saharan Africa: Geographically refers to all the areas of the continent of Africa which lie south of the Sahara Desert. Some of the countries which fall into this category include countries such as Kenya, Ethiopia, Ghana, Nigeria and South Africa.

Mobile Virtual Reality to Enhance Subjective Well-Being



Federica Pallavicini

Università di Milano-Bicocca, Italy

Luca Morganti

Università di Milano-Bicocca, Italy

Barbara Diana

Università di Milano-Bicocca, Italy

Olivia Realdon

University of Milano-Bicocca, Italy

Valentino Zurloni

Università di Milano-Bicocca, Italy

Fabrizia Mantovani

Università di Milano-Bicocca, Italy

INTRODUCTION

Focusing on health-care, in the last few decades advanced technologies have become crucial keys in supporting subjective well-being (Botella et al., 2012; Riva et al., 2012). Among them, Virtual Reality (VR) - defined as three-dimensional, stereoscopic, interactive computer graphics — has been proven effective in promoting mental health at different levels. In particular, several research studies showed the efficacy of VR in stress management (Gaggioli et al., 2014; Pallavicini et al., 2013; Rizzo et al., 2012; Serino et al., 2014) and in the treatment of different disorders; these include phobias (Parsons & Rizzo, 2008; Rothbaum et al., 1995), anxiety (Meyerbröcker & Emmelkamp, 2010, 2011; Pallavicini et al., 2009) and eating disorders (Ferrer-García et al., 2009; Riva, 2005). This technology is also adopted in neuropsychology, for both the assessment and training of cognitive processes (Cipresso et al., 2014; Fordell et al., 2011; Raspelli et al., 2012).

Although the dramatic development in the field of VR systems, there are still important problems related to the use of this technology (Pallavicini et al., 2015; Proffitt & Lange, 2015). First, from a technological point of view, VR are not so easy to be used, requiring a specific training for the clinician and the patient. Secondly, from a clinical perspective, these technologies are not easy to be moved at patients' home, where the delivery of interactive exercises may be useful, especially when patients are provided with a rehabilitative training (Pallavicini et al., 2015).

Within this perspective, the growing availability, low-cost and easy-to-use of Mobile Virtual Reality (MVR) (i.e., the integration of VR system on mobile devices such as smartphone and tablet) represents a meaningful opportunity to support mental health interventions (Gaggioli et al., 2014; Gorini et al., 2010; Pallavicini et al., 2009). MVR, in particular, can offer to the community an innovative tool for the management, monitoring and delivery of exercises, that can

also be used in individuals ‘favorite environment’, their home (Pallavicini et al., 2015; Tong et al., 2015; Schroeder et al., 2013).

The present chapter will first briefly describe MVR, highlighting the specific features that characterize it. Then, a MVR design practice will be presented. In conclusion, the current application of MVR to enhance subjective well-being will be discussed, with the support of concrete examples and research studies-analysis.

BACKGROUND

VR is a high-end user-computer interface that involves real-time simulation and interactions through multiple sensorial channels (Burdea & Coiffet, 2003). Although many authors have defined VR essentially as a technology (Heim, 1998), more recent approaches (Riva et al., 2007) forward a more complex vision, considering VR as a human experience and underlining how “the essence of VR is the inclusive relationship between the participant and the virtual environment” (Fitzgerald & Riva, 2001).

VR can be presented in at least five ways:

- **Fully Immersive VR:** It consists of 3D simulation that allow participants to observe and interact with an environment through an available set of actions. With this type of solution, the user appears to be fully inserted in the computer-generated environment. This illusion is rendered by providing a Head Mounted Display (HMD) with 3D viewing and a system of head tracking, to guarantee the exact correspondence and coordination of user’s movements with the feedbacks of the environment;
- **Desktop VR:** It consists of computer-generated environments, which exist in 3D (even if they are shown on a 2D display). It uses a computer monitor as display to provide graphical interface for users. It is cost-effective when compared to the immersive

VR as it does not require any expensive hardware and software and is relatively easy to develop. Interaction with the interface can be made via mouse, joystick or typical VR peripherals such as Dataglove;

- **Augmented Reality:** The user’s view of the world is supplemented with virtual objects, usually to provide information about the real environment;
- **Mixed Reality:** Sometimes referred to as hybrid reality, refers to the merging of real and virtual worlds to produce new environments and visualizations where physical and digital objects co-exist and interact in real time;
- **CAVE:** The term stands for CAVE Automatic Virtual Environment and takes the form of a small room where a computer-generated world is projected on the walls. The projection is made on both front and sidewalls. This solution is particularly suitable for collective VR experience because it allows different people to share the same experience at the same time;
- **Telepresence:** Users can influence and operate in a world that is real but in a different location. The users can observe the current situation with remote cameras and achieve actions via robotic and electronic arms.

From about 2007, the year in which Apple presented the first model of iPhone and the mobile industry significantly changed - thanks to the new possibilities offered by smartphone and tablet to integrate VR, MVR has been risen.

MVR can be defined as “the integration of VR environments on mobile devices” (Pallavicini et al., 2015), and it can be divided into two sub-categories based on its technical features:

- **Screen MVR:** It uses the mobile screen as display to provide 3D virtual environments, even if they are shown on a 2D display;
- **Immersive MVR:** The user is completely immersed in the computer-generated en-

vironment by providing a Head Mounted Display (HMD) designed for mobile phones, like Samsung Gear and Google's Cardboard VR.

HUMAN COMPUTER INTERACTION AND DESIGN PRINCIPLES FOR MOBILE VIRTUAL REALITY

A central task in the design of MVR is to create scenarios and tasks that could be effective, engaging (and usable) within and across different contexts (e.g., clinic and home). Thus, an interdisciplinary approach including users, trainers, designers, psychologists and technology experts can define concrete guidelines, may support strategic choices and training design activities evaluation.

To this end, much can be learnt from Interaction Design Research (e.g., Stolterman, 2008), which offer general principles for Human Computer Interaction (HCI) or “the region of intersection between psychology and the social sciences, on the one hand, and computer science and technology, on the other” (Carroll, 1997). In other words, HCI is a science of design, aimed to understand and support human beings interacting with and through technology. Within HCI, a specific area is focused on VR systems, defined as “Human Virtual Environment Interaction” (HVEI) (Bainbridge, 2007; Stanney, Mourant, & Kennedy, 1998).

During the design of MVR scenarios, technology and users-related variables are equally important to be considered (Harrison et al., 2013). In fact, due to the close bond between the user and the system, is impossible to segregate human factors from design issues when striving to achieve the potential of MVR technology. It is the capabilities and limitations of the user that often will determine the effectiveness of virtual worlds.

- **Technology-Related Variables:**

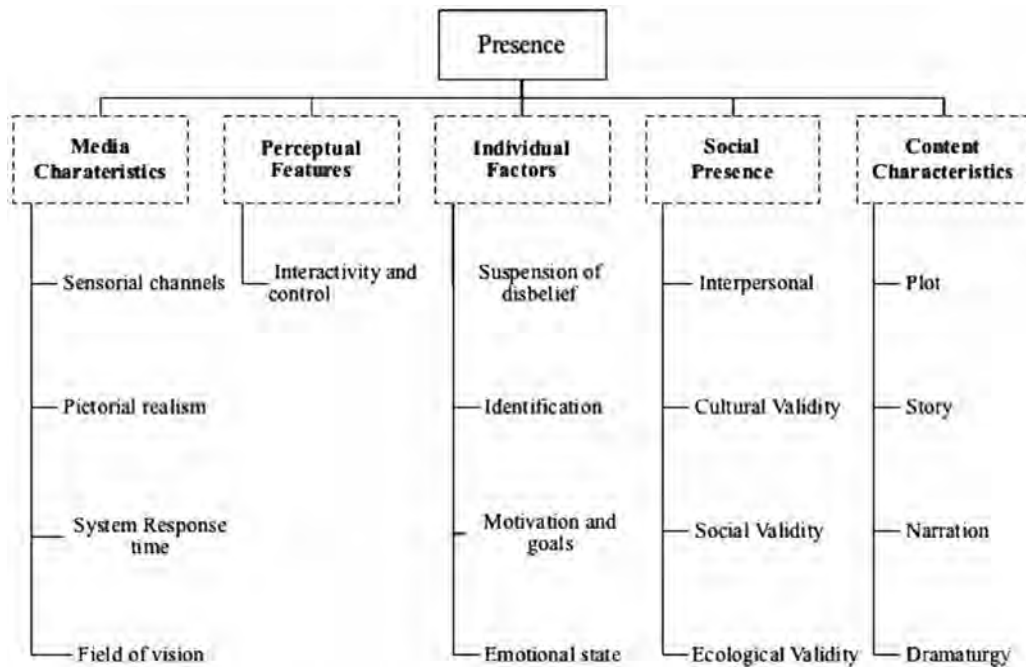
- **Fidelity:** This term refers to the extent to which the virtual environments emulate the real world (Alexander,

Brunyé, Sidman, & Weil, 2005). In other words, it can be defined as the objective level of sensory realism a VR system provides. Physical fidelity can be achieved and developed on different levels: visual, auditory and haptic. MVR systems vary in their degree of physical and functional fidelity based on technical feasibility, cost, and needs. Obviously, it is not possible to simulate the physical world in all its details and complexity.

- **Users-Related Variables:**

- **Usability:** Is defined as the extent to which a product can be used, by specified users, to achieve specified goals with effectiveness, efficiency and satisfaction, in a specified context of use. Usability is often defined as a quality attribute, which determines how the interface is easy to use by users. Following the definition, it is possible to assume that a good usability can be crucial to provide a good user experience. However, depending on the context, the system may have a good user experience without good usability, since several other areas such as aesthetics, fidelity, interactivity and presence are related (Garrett, 2010; Shneiderman, 2010; Sutcliffe & Kaur, 2000).
- **Sense of Presence:** Presence is widely accepted as the key concept to consider in any research involving human interaction with VR systems. Presence has been used as a global experiential quality metric to evaluate, develop, and optimize virtual environments and media systems (Riva, Waterworth, Waterworth, & Mantovani, 2011; Riva et al., 2007; Stanney et al., 1998). Factors connected with the possibility to develop the sense of presence experience by

Figure 1. Usability framework

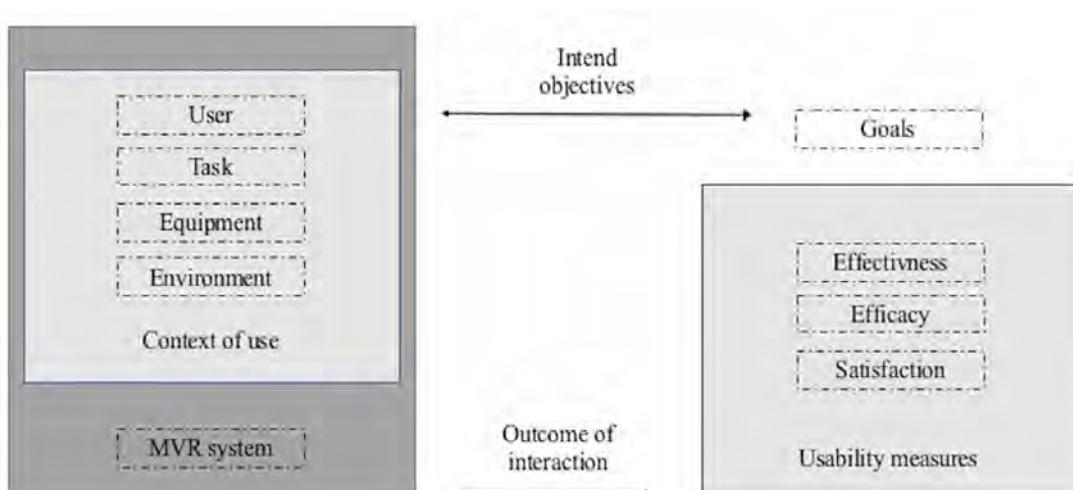


users can be included in five main areas: (1) media characteristics; (2) perceptual features; (3) individual factors; (4) content characteristics; (5) interpersonal, social and cultural context.

- **Cybersickness:** In some cases, users develop symptoms that are similar to

the common symptoms found when people get motion sick. Users who exhibit these type of motion sickness-like symptoms suffer from a malady called cybersickness. A number of symptoms can occur due to cybersickness (LaViola, 2000; Vinson, 2012).

Figure 2. Developing presence



Regarding the process of the design of MVR scenarios and tasks is a complex and iterative process that should include different phases that could be summarized as following (see Table 2):

1. **Comprehension Phase:** During this first stage, ‘what the MVR system must do’ (i.e., what features should have and how it should adapt to other elements) and requirements are defined. To identify the requirements is necessary to collect information about the
- final users, the other stakeholders (i.e., all the people who will be affected by the system), and the contexts in which the MVR will be used (i.e. Immersive VR versus Screen VR). The requirements to be defined include those: (1) non-functional (i.e., what the system must be able to do), and (2) functional (i.e., the reality of what is technically feasible).
2. **Design Phase:** It includes two different types of design activities, and in particular: (1) conceptual design, that is the design of

Table 1. Designing MVR for subjective well-being: steps and guidelines

Step	Description	Main Questions You Have to Ask Yourself
1	Define the goals	• What do I want to achieve? • Is my goal SMART (specific, measurable, achievable, relevant, time-bound)? • What are the skills I want to assess/train? • How can I define each one of these skills? • What is my target? • What is the impact this goal is going to have on my target?
2	Select objectives that can use MVR for attainment	• Why should I use MVR to assess/train these skills? • How can I assess/train these skills with a MVR system? • What are the competitive advantages if a MVR system is used?
3	Refine selection, choosing those that can use MVR	• Among all the skills I want to assess/train, which are the most suitable MVR? • Why do I have to choose them?
4	Perform substeps for each objective	• How can I achieve my goal? • Are there any other subordinate goals? • What are the tasks users need to complete in order to achieve each subgoal? • What are the steps required to achieve the goal?
5	Choose MVR software/hardware and/or delivery system	• Is fidelity important for my goals? • What kind of fidelity (physical, functional, equipment, psychological) do I have to enhance? • How can I enhance it? • What type of immersion and presence is required? • What kind of interaction is required?
6	Design and build the MVR scenarios	• What knowledge and skills are required to produce the required behavior given the information provided? • Do the color and texture matter? What are the critical appearance attributes? • What aspects of behavior have to be produced to generate interactions? • What sounds have to be produced and to what degree of fidelity?
7	Evaluate the MVR scenarios with a pilot group	• Can a pilot group achieve the main goal? • Which tools and measure can I use to evaluate MVR efficacy?
8	Modify the MVR scenarios according to evaluation. Repeat 7 & 8 until satisfactory	• Is the system usable? • Do users experience cybersickness? • Do users experience a sense of presence in the environment?
9	Evaluate the MVR scenarios with a target population	• Is the system effective? • Which tools and measure can I use to evaluate MVR efficacy? • What are the main variables I have to consider?
10	Modify according to results. Repeat 9 & 10 as needed	• Is the system usable? • Do users experience cybersickness? • Do users experience a sense of presence in the environment?

a system from the abstract point of view; (2) physical design, the concretization of ideas and the definition of operating details and interface design, with particular attention to ways of interaction (interaction design) and layout (aesthetic).

3. **Envisionment Phase:** At this stage, it is necessary to represent the design ideas to demonstrate the feasibility of ideas. The MVR project can be represented in different display modes including the creation of storyboards, prototypes, sketches (Davies, 2004).
4. **Evaluation Phase:** Any design and envisionment activities is followed by the evaluation phase, which has the purpose to verify the effectiveness of the MVR scenarios. There are several techniques, including the creation of specific scenarios of use (Sutcliffe & Kaur, 2000).
5. **Implementation Phase:** Development and validation of the MVR system. Essential to check that the entire system complies with the requirements laid down and to be able to launch and complete.

SOLUTIONS AND RECOMMENDATIONS

As already pointed out in chapter, MVR is a young technology, in particular as regards the Immersive MVR. For this reason, still few studies have tested the effectiveness of this technology for enhancing subjective well-being (e.g. Pallavicini et al., 2009, Gorini et al., 2010, Tong 2015). Below the MVR app tested so far for mental health will be presented.

- **Anxiety Disorders:** The first clinical interventions for subjective well-being using MVR was developed for the treatment of General Anxiety Disorder (GAD) (Gorini et al., 2010; Pallavicini et al., 2011). Patients were given a smartphone (HTC

Touch Pro T7272) in which they can experience a relaxing scenario, achieved by presenting virtual environments (i.e., campfire, waterfall, beach and a gazebo) for relaxation homework. Results show the efficacy of this mobile technology in the treatment of GAD and the advantage for patients to set a guided relaxing experience using a mobile device (Gorini et al., 2010; Pallavicini et al., 2009; Repetto et al., 2011).

- **Stress:** Positive Technology is a mobile application developed for stress management (Gaggioli et al., 2014). It included three different components: (1) Guided Relaxation, that consisted of four 3D interactive environments and six relaxing music, designed to support relaxations; (2) Biofeedback, in which heart rate data detected in real-time modified the size of virtual objects (fire, waterfall) in the virtual scenario; (3) Stress self-reporting and tracking, where the user can report stress level or let the app heart his/her heart rate values.
- **Pain:** A technological solution, integrating MVR within the theoretical perspective of mindfulness (a form of meditation teaching present-moment awareness) has been developed by Schroeder and colleagues (Schroeder et al., 2013). Unity 3D for iOS system was used to develop the software for a solution consisting of separate virtual cameras, tied to each of the two rendering viewports displayed on the phone. The result is a set of three scenarios: (1) the Mindful-Movement Environment, that recreates a mindful walking, with the related body perceptions within a calm garden and ambient sounds scenario; (2) the Engaging-Multiple-Senses Environment, which integrate also auditory, visual and haptic stimuli, directing the attention to present perceptions offered by the five senses; (3) the Lying-down Guided-rest Environment, to facilitates visualization exercises while

Table 2. MVR current applications to enhance subjective well-being

Area	Study	MVR App	Operating System	MVR Scenarios
Anxiety	• Pallavicini, 2009 • Gorini, 2010	INTREPID system, virtual environments for relaxation and biofeedback training	Windows Mobile	Four virtual environments (campfire beach, waterfall and gazebo).
Stress	Gaggioli, 2014	Positive Technology -App for relaxation and biofeedback	iOS (iPad 2/New iPad and iPhone 4/5)	Five environments (island landscapes for relaxation, with fire and waterfall for biofeedback).
Pain	Schroeder, 2013	Virtual scenarios for pain management with tactile body sensors	iOs (iPhone 4)	Three environments (mindful-movement, engaging-multiple- senses, lying-down guided-rest).
Pain	Tong, 2016	Cryoblast - First-person shooter game for self-pain management	Android system, played via Google Cardboard (Do It Yourself VR)	Six environments (six different types of cave as game scenarios).
Neglect	Pallavicini, 2015	NeglectApp - App for the assessment of the disease	iOs (iPad 2)	Three environments (Neglect App cancellation tests, Neglect App card dealing task).

lying down, detecting psychophysiological parameters. Recently, Tong (Tong et al., 2015) developed a MVR solution for self-managing of pain, called Cryoblast. This pain distraction game has been developed for Google's Cardboard and works on mobile phone as a first person shooter. The players are virtually turned into the character of the game that has to go through six environments (virtual caves) and two kinds of games (fighting enemies and collect items). Enemies are the cells responsible for pain that the player has to defeat, using analgesic drugs as weapons, shooting through the Cardboard magnetic button.

- **Neglect:** Recently Pallavicini et al. (Pallavicini et al., 2015) developed a MVR app for iOS system, called NeglectApp, for the assessment of Neglect. This app contains a series of test for the evaluation of the USN using interactive virtual environments. Specifically, the app includes two different categories of tasks: (1) Neglect App cancellation tests, which are designed to recreate the traditional paper-and-pencil cancellation test, played within the application in a three-dimensional version; (2) Neglect App card dealing task that recre-

ates a virtual environment in which the patient has to give three cards for a game to himself/herself, to each one of the three individuals seated and four in the middle of a table. Neglect App test were carried out with the aid of a stylus for touchscreen, on an iPad2. A scientific evaluation reported that the tests made with the app are equally effective to the traditional ones (Pallavicini et al., 2015).

MVR can play a significant role to support subjective well-being and mental health care. Through the growing availability of cheap and safe devices, MVR is becoming a mass consumer product, and moving to the home. This implies that MVR can provide innovative opportunities to design and to shape the continuity from the real to the virtual (and vice versa), in both training and assessment, in a variety of application domains (from the clinical, or the neuropsychological, to the learning one). It is not a matter of generating a seamless continuity from the real to the virtual. The aim is to combine simulations in the real with simulation in the virtual, in a serial fashion and in a way that allows researchers to assess the integrative process carried out by perceivers over time.

FUTURE RESEARCH DIRECTIONS

In the future, it would be interesting to develop MVR scenarios and tasks that exploit the full potential of this technology, using its unique features of navigation, realism and interactivity. Secondly, it could be useful to study the possibility offered by MVR in training and rehabilitation program for mental health-care. Such technology, in fact, could make patients' homework and tasks more engaging, and consequently more effective, than traditional ones. Potentially, moreover, data collected about exercises made on tablet (scores at each task, frequency of use, etc.) could be tracked over time and integrated with other patient information, providing important information to the clinician.

To this end, it will be important that future studies will be aimed to: (1) scientifically test the effectiveness of this innovative technology for personal empowerment and well-being; (2) investigate aspects of User Experience related to MVR, in order to define which are the unique aspects of this technology, compared to other traditional tools, and to provide useful guidelines for the design of this type of application.

CONCLUSION

A recent Goldman Sachs research report predicts the VR industry may soon become an \$80 billion market (\$45 billion in hardware and \$35 billion in software) by only 2025. Since MVR is still a new and developing space, analysts say the forecasted \$80 billion is likely to waver according to actual adoption rates among enterprise and consumer users.

In light of such steep growth rates and innovation, VR and MVR will probably create new industries and change existing ones, including the mental health-care industry. Designing usable and effective interactive MVR system in supporting subjective well-being in a more engaging way

than usual care programs - and consequently potentially more effective - will be the challenge to overcome in the next years.

Furthermore, efforts will be needed before MVR can become part of clinical practice; however, the growing availability of this easy-to-use and low-cost technology can be a great challenge to support subjective well-being and mental health care programs.

REFERENCES

- Alexander, A. L., Brunyé, T., Sidman, J., & Weil, S. A. (2005). *From Gaming to Training: A Review of Studies on Fidelity, Immersion, Presence, and Buy-in and Their Effects on Transfer in PC-Based Simulations and Games*. DARWARS Training Impact Group.
- Bainbridge, W. S. (2007). The scientific research potential of virtual worlds. *Science*, 317(5837), 472–476. doi:10.1126/science.1146930 PMID:17656715
- Botella, C., Riva, G., Gaggioli, A., Wiederhold, B. K., Alcaniz, M., & Baños, R. M. (2012). The Present and Future of Positive Technologies. *Cyberpsychology, Behavior, and Social Networking*, 15(2), 78–84. doi:10.1089/cyber.2011.0140 PMID:22149078
- Burdea, G., & Coiffet, P. (2003). *Virtual reality technology*. John Wiley.
- Carroll, J. M. (1997). Human-computer interaction: Psychology as a Science of Design. *Annual Review of Psychology*, 48(1), 61–83. doi:10.1146/annurev.psych.48.1.61 PMID:15012476
- Cipresso, P., Albani, G., Serino, S., Pedroli, E., Pallavicini, F., Mauro, A., & Riva, G. (2014). Virtual multiple errands test (VMET): a virtual reality-based tool to detect early executive functions deficit in Parkinson's disease. *Frontiers in Behavioral Neuroscience*, 8, 405.

- Davies, R. C. (2004). Adapting virtual reality for the participatory design of work environments. *Computer Supported Cooperative Work*, 13(1), 1–33. doi:10.1023/B:COSU.0000014985.12045.9c
- Ferrer-García, M., Gutiérrez-Maldonado, J., Caqueo-Úrizar, A., & Moreno, E. (2009). The validity of virtual environments for eliciting emotional responses in patients with eating disorders and in controls. *Behavior Modification*, 33(6), 830–854. PMID:19822737
- Fitzgerald, M., & Riva, G. (2001). *Telemedicine Glossary*. European Commission-DG INFSO.
- Fordell, H., Bodin, K., Bucht, G., & Malm, J. (2011). A virtual reality test battery for assessment and screening of spatial neglect. *Acta Neurologica Scandinavica*, 123(3), 167–174. doi:10.1111/j.1600-0404.2010.01390.x PMID:20569225
- Gaggioli, A., Cipresso, P., Serino, S., Campanaro, D. M., Pallavicini, F., Wiederhold, B. K., & Riva, G. (2014). Positive technology: A free mobile platform for the self-management of psychological stress. In *Annual Review of Cybertherapy and Telemedicine 2014: Positive Change: Connecting the Virtual and the Real* (Vol. 99, pp. 25–29).
- Gaggioli, A., Pallavicini, F., Morganti, L., Serino, S., Scaratti, C., Briguglio, M., & Riva, G. et al. (2014). Experiential virtual scenarios with real-time monitoring (interreality) for the management of psychological stress: A block randomized controlled trial. *Journal of Medical Internet Research*, 16(7), e167. doi:10.2196/jmir.3235 PMID:25004803
- Garrett, J. J. (2010). *The Elements of User Experience: User-Centered Design for the Web and Beyond*. New Riders Publishing.
- Gorini, A., Pallavicini, F., Algeri, D., Repetto, C., Gaggioli, A., & Riva, G. (2010). Virtual reality in the treatment of generalized anxiety disorders. *Studies in Health Technology and Informatics*, 154, 39–43. PMID:20543266
- Harrison, R., Flood, D., & Duce, D. (2013). Usability of mobile applications: Literature review and rationale for a new usability model. *Journal of Interaction Science*, 1(1), 1. doi:10.1186/2194-0827-1-1
- Heim, M. (1998). *Virtual Realism*. Oxford, UK: Oxford University Press.
- Meyerbröker, K., & Emmelkamp, P. M. G. (2010). Virtual reality exposure therapy in anxiety disorders: A systematic review of process-and-outcome studies. *Depression and Anxiety*, 27(10), 933–944. doi:10.1002/da.20734 PMID:20734361
- Meyerbröker, K., & Emmelkamp, P. M. G. (2011). *Virtual Reality Exposure Therapy for Anxiety Disorders: The State of the Art* (pp. 47–62). Springer Berlin Heidelberg.
- Mine, M. R. (1995). *Virtual Environment Interaction Techniques*. University of North Carolina at Chapel Hill.
- Pallavicini, F., Algeri, D., Repetto, C., Gorini, A., & Riva, G. (2009). Biofeedback, virtual reality and mobile phones in the treatment of generalized anxiety disorder (gad): A phase-2 controlled clinical trial. *Journal of Cyber Therapy and Rehabilitation*, 2(4), 315–327.
- Pallavicini, F., Gaggioli, A., Raspelli, S., Cipresso, P., Serino, S., Vigna, C., & Riva, G. et al. (2013). Interreality for the management and training of psychological stress: Study protocol for a randomized controlled trial. *Trials*, 14(1), 191. doi:10.1186/1745-6215-14-191 PMID:23806013
- Pallavicini, F., Pedrolì, E., & Serino, S., Dell'isola, A., Cipresso, P., Cisari, C., & Riva, G. (2015). Assessing unilateral spatial neglect using advanced technologies: The potentiality of mobile virtual reality. *Technology and Health Care*, 1, 1–13. PMID:26409526

- Pallavicini, F., Pedrolì, E., Serino, S., Dell'Isola, A., Cipresso, P., Cisari, C., & Riva, G. (2015). Assessing Unilateral Spatial Neglect using advanced technologies: The potentiality of mobile virtual reality. *Technology and Health Care*, 23(6), 795–807. doi:10.3233/THC-151039 PMID:26409526
- Parsons, T. D., & Rizzo, A. A. (2008). Affective outcomes of virtual reality exposure therapy for anxiety and specific phobias: A meta-analysis. *Journal of Behavior Therapy and Experimental Psychiatry*, 39(3), 250–261. doi:10.1016/j.jbtep.2007.07.007 PMID:17720136
- Proffitt, R., & Lange, B. (2015). Considerations in the Efficacy and Effectiveness of Virtual Reality Interventions for Stroke Rehabilitation: Moving the Field Forward. *Physical Therapy*, 95(3), 441–448. doi:10.2522/ptj.20130571 PMID:25343960
- Raspelli, S., Pallavicini, F., Carelli, L., Morganti, F., Pedrolì, E., Cipresso, P., & Riva, G. et al. (2012). Validating the Neuro VR-Based Virtual Version of the Multiple Errands Test: Preliminary Results. *Presence (Cambridge, Mass.)*, 21(1), 31–42. doi:10.1162/PRES_a_00077
- Repetto, C., Gaggioli, A., Pallavicini, F., Cipresso, P., Raspelli, S., & Riva, G. (2011). Virtual reality and mobile phones in the treatment of generalized anxiety disorders: A phase-2 clinical trial. *Personal and Ubiquitous Computing*, 253–260.
- Riva, G. (2005). Virtual reality in psychotherapy: review. *Cyberpsychology & Behavior: The Impact of the Internet, Multimedia and Virtual Reality on Behavior and Society*, 8(3), 220–30.
- Riva, G., Baños, R. M., Botella, C., Wiederhold, B. K., & Gaggioli, A. (2012). Positive Technology: Using Interactive Technologies to Promote Positive Functioning. *Cyberpsychology, Behavior, and Social Networking*, 15(2), 69–77. doi:10.1089/cyber.2011.0139 PMID:22149077
- Riva, G., Mantovani, F., Capideville, C. S., Prezi-osa, A., Morganti, F., Villani, D., & Alcañiz, M. et al. (2007). Affective interactions using virtual reality: The link between presence and emotions. *Cyberpsychology & Behavior*, 10(1), 45–56. doi:10.1089/cpb.2006.9993 PMID:17305448
- Riva, G., Waterworth, J. A., Waterworth, E. L., & Mantovani, F. (2011). From intention to action: The role of presence. *New Ideas in Psychology*, 29(1), 24–37. doi:10.1016/j.newideapsych.2009.11.002
- Rizzo, A. A., John, B., Williams, J., Newman, B., Koenig, S. T., Lange, B. S., & Buckwalter, J. G. (2012). Stress resilience in virtual environments: training combat relevant emotional coping skills using virtual reality. In *9th Conference on Disability, Virtual Reality & Associated Technologies* (pp. 229–237).
- Rothbaum, B. O., Hodges, L. F., Kooper, R., Opdyke, D., Williford, J. S., & North, M. (1995). Effectiveness of computer-generated (virtual reality) graded exposure in the treatment of acrophobia. *The American Journal of Psychiatry*, 152(4), 626–628. doi:10.1176/ajp.152.4.626 PMID:7694917
- Schroeder, D., Korsakov, F., Jolton, J., Keefe, F. J., Haley, A., & Keefe, D. F. (2013). Creating widely accessible spatial interfaces: Mobile VR for managing persistent pain. *IEEE Computer Graphics and Applications*, 33(3), 82–89. doi:10.1109/MCG.2013.38 PMID:24807994
- Serino, S., Triberti, S., Villani, D., Cipresso, P., Gaggioli, A., & Riva, G. (2014). Toward a validation of cyber-interventions for stress disorders based on stress inoculation training: A systematic review. *Virtual Reality (Waltham Cross)*, 18(1), 73–87. doi:10.1007/s10055-013-0237-6
- Shneiderman, B. (2010). *Designing the user interface : strategies for effective human-computer-interaction*. Addison Wesley Longman.

Stanney, K. M., Mourant, R. R., & Kennedy, R. S. (1998). Human Factors Issues in Virtual Environments: A Review of the Literature. *Presence (Cambridge, Mass.)*, 7(4), 327–351. doi:10.1162/105474698565767

Stolterman, E. (2008). The nature of design practice and implications for interaction design research. *International Journal of Design*, 2(1), 55–65.

Sutcliffe, A. G., & Kaur, K. D. (2000). Evaluating the usability of virtual reality user interfaces. *Behaviour & Information Technology*, 1(6), 415–426. doi:10.1080/014492900750052679

Tong, X., Gromala, D., Amin, A., & Choo, A. (2015). The Design of an Immersive Mobile Virtual Reality Serious Game in Cardboard Head-Mounted Display for Pain Management. In *International Symposium on Pervasive Computing Paradigms for Mental Health* (pp. 284–293). Springer International Publishing.

ADDITIONAL READING

Bartle, R. A. (2004). *Designing Virtual Worlds*. New Riders Publishing.

Benynon, D. (2014). *Designing interactive systems: a comprehensive guide to HCI, UX and interaction design*. Pearson.

Dalgarno, B., & Lee, M. J. (2010). What are the learning affordances of 3D virtual environments? *British Journal of Educational Technology*, 41(1), 10–32. doi:10.1111/j.1467-8535.2009.01038.x

Lyons, F. A. (2016). *U.S. Patent No. D751,072*. U.S. Patent and Trademark Office.

Morina, N., Ijntema, H., Meyerbröker, K., & Emmelkamp, P. M. (2015). Can virtual reality exposure therapy gains be generalized to real-life? A meta-analysis of studies applying behavioral assessments. *Behaviour Research and Therapy*, 74, 18–24. doi:10.1016/j.brat.2015.08.010 PMID:26355646

Parisi, T. (2015). *Learning Virtual Reality: Developing Immersive Experiences and Applications for Desktop, Web, and Mobile*. O'Reilly Media, Inc.

Villani, D., Cipresso, P., Gaggioli, A., & Riva, G. (2016). *Integrating Technology in Positive Psychology Practice*. IGI Global. doi:10.4018/978-1-4666-9986-1

KEY TERMS AND DEFINITIONS

Desktop VR: Computer-generated environments which exist in three dimensions (even if they are shown on a 2D display).

Fidelity: The extent to which the virtual environments emulate the real world.

Fully Immersive VR: Three-dimensional (3D) simulation that allow participants to observe and interact with an environment through a head mounted display with 3D viewing.

Immersive MVR: A MVR system in which the user is completely immersed in the computer generated environment by providing a Head Mounted Display (HMD) designed for mobile phones.

Mobile Virtual Reality (MVR): Integration of Virtual Reality system on mobile devices such as smartphone and tablet).

Screen MVR: A MVR system in which that uses the mobile screen as display to provide 3D virtual environments, even if they are shown on a 2D display.

Usability: The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use.

Novel Methods to Design Low-Complexity Digital Finite Impulse Response (FIR) Filters

David Ernesto Troncoso Romero
CONACYT at ESCOM-IPN, Mexico

Gordana Jovanovic Dolecek
Institute INAOE, Mexico

INTRODUCTION

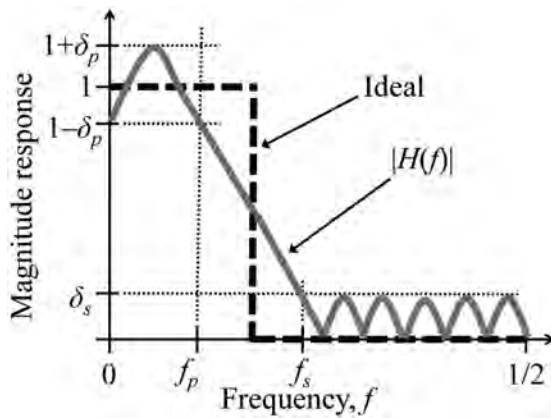
A discrete-time signal is a sequence of numbers, so-called samples. A sample lasts T_s seconds, then another sample appears during the next T_s seconds and so on, forming the sequence. The time T_s is called sampling period and the number produced in the n -th sampling period (i.e., after n times T_s seconds) is denoted as $x(nT_s)$. For analysis purposes it is usual to assume that $x(nT_s) = 0$ for negative values of n and, in general, the samples can be *complex numbers*. If so, the signal is a complex signal. In these analyses, the digital frequency f is a value that expresses how much of a cycle of a sinusoidal wave is represented by a sample. In general terms, discrete-time signals may contain information for different values of f .

A digital filter is a system that receives a discrete-time signal and returns another discrete-time signal with modified characteristics (Antoniu, 2006). Many digital filters exhibit the *Linear Time-Invariant* (LTI) property. When the input signal $x(nT_s)$ has the value 1 for $n = 0$ and 0 for other values of n , the output signal of a LTI digital filter is called impulse response. LTI digital filters are often classified by the duration of its impulse response as Infinite-duration Impulse Response (IIR) or Finite-duration Impulse Response (FIR) filters (Tan & Jiang, 2013). Additionally, a LTI digital filter is generally described by its frequency response, which is the response of the filter to a *complex exponential signal* with frequency $2\pi f$. The frequency response is a *complex function* of f

and is periodic over f , with a period of $1/T_s$. Because of this periodicity, the frequency response is just observed in the range of values from $f = -1/2T_s$ to $f = 1/2T_s$. As a complex function, the frequency response has two functions associated with it: its *magnitude response* and its *phase response*.

When the values of the impulse response are real, the magnitude response is symmetric and the phase response is anti-symmetric around $f = 0$ (Taylor, 2012). Thus, these responses are just observed in the range of values from $f = 0$ to $f = 1/2T_s$. Systems with real impulse response are used in a wide variety of applications, being one of the most common the traditional low-pass filter. This filter passes the frequency components of the input signal that range from 0 to f_p and rejects the frequency components of the signal that range from f_s to $1/2T_s$, with $0 < f_p < f_s < 1/2T_s$. The values f_p and f_s are the passband edge and the stopband edge frequencies, respectively, whereas the range from 0 to f_p is the passband, the range from f_s to $1/2T_s$ is the stopband and the difference $f_s - f_p$ is the transition band. The ideal values of the magnitude response in passband and stopband are, respectively, 1 and 0. For a realizable filter, an acceptable deviation from these ideal values must be specified. Usually, these deviations are represented as the numbers δ_p and δ_s , respectively. Figure 1 shows the specifications of the magnitude response of a low-pass filter, along with the ideal response and the actual magnitude response of the filter, denoted as $|H(f)|$. For analysis purposes it is usual to assume $T_s = 1$, i.e., $x(nT_s) = x(n)$.

Figure 1. Magnitude response specifications of a low-pass filter



Thus, in Figure 1 the range of interest of values f is from 0 to $1/2$.

Digital FIR filters play a central role in modern Digital Signal Processing (DSP) systems. These filters find extensive applications in communication systems, audio systems, instrumentation, image enhancement, processing of geophysical signals, processing of biological signals, among others (Antoniou, 2006). This is mainly due to the advantages of FIR filters, being the most important (Taylor, 2012):

- Absolute stability.
- The fact that FIR filters can be designed to have linear phase response.

The main disadvantage of conventional FIR filter designs is that they require, especially in applications demanding narrow transition bands, a high number of arithmetic operations and hardware components. This makes the implementation of FIR filters with narrow transition bands very costly (Taylor, 2012). However, several methods to design low-complexity FIR filters have been developed over the past three decades, approximately. One of the most popular approaches among them consists on designing an overall filter as an interconnection of simple subfilters, where some of these subfilters have a frequency response with one or more of its periodic shapes, so-called im-

ages, appearing into the range $0 < f < 1/2$ (in other words, the period of the frequency response of these subfilters is $1/2$ or smaller). These methods are classified as “frequency response masking methods”. This approach is adequate from a low power consumption perspective, since the subfilters are simple and the overall filter performs a lower number of computations in comparison to a direct design.

The design of low-complexity FIR filters is a challenging task since many aspects have to be taken into account. The efficient design of FIR digital filter needs a trade-off between very stringent design specifications, low power consumption, low area requirements, high speed of computations and low time and design effort. An acceptable design should balance the trade-off to a reasonable degree. Thus, in this chapter we introduce a review of methods employed to efficiently design low-complexity linear-phase FIR filters.

BACKGROUND

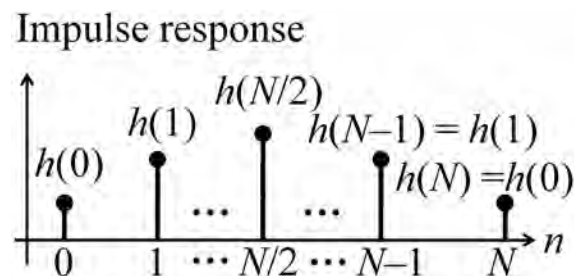
A FIR filter produces a signal $y(n)$ as follows,

$$y(n) = \sum_{k=0}^N h(k)x(n-k) \quad (1)$$

where the integer N is the filter order and the values $h(k)$, with $0 \leq k \leq N$, are the samples of the impulse response of the filter. A FIR filter has a linear phase if its impulse response satisfies either $h(k) = h(N-k)$, which is called condition of symmetry, or $h(k) = -h(N-k)$, which is called condition of anti-symmetry, for $0 \leq k \leq N$. Figure 2 illustrates the impulse response of a symmetric filter for N even.

The direct-form implementation of a symmetric filter for N even is presented in Figure 3. The filter consists of three types of digital circuits: *multipliers*, *adders* and *delays*. Among them, multipliers are the most critical blocks in terms of power consumption and speed of operation

Figure 2. Impulse response of an N -th order symmetric filter, with N even



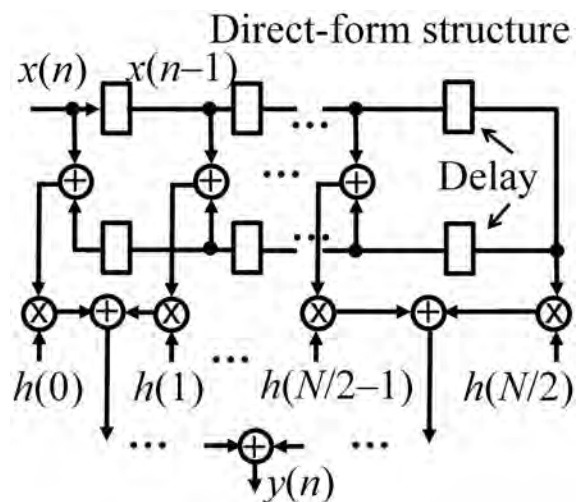
(Aksoy, Flores & Monteiro, 2014). The values of the impulse response of the FIR filter are physically implemented in the structure, and they are usually referred as coefficients of the filter. The direct-form implementation of a symmetric (or an anti-symmetric) filter requires approximately N adders, N delays and $N/2$ multipliers. The number of arithmetic operations that a filter performs to obtain an output sample is the computational complexity of the filter. For the aforementioned filter, this complexity is N additions and $N/2$ multiplications. Clearly, the complexity increases proportional to N . The computational complexity is proportional to both, the power consumption of the filter and the speed of operation. Thus, this complexity must be decreased in practical applications.

The minimum order of an optimum linear-phase FIR filter meeting the low-pass filter specifications $(1-\delta_p) \leq |H(f)| \leq (1+\delta_p)$ for $0 \leq f \leq f_p$ and $-\delta_s \leq |H(f)| \leq \delta_s$ for $f_s \leq f \leq 1/2$ (Figure 1) can be estimated as

$$N \approx \frac{-20 \log_{10}(\sqrt{\delta_p \delta_s}) - 13}{14.6(f_s - f_p)} \quad (2)$$

After estimating N for a set of specifications f_p , f_s , δ_p and δ_s , a direct design consists in finding the values $h(k)$, with $0 \leq k \leq N$, that make the filter to accomplish the desired specifications. This is usually done with the Parks McClellan algorithm (Taylor, 2012).

Figure 3. Direct-form implementation structure of an N -th order symmetric filter, with N even



From the above estimate, it is seen that as the transition bandwidth $f_s - f_p$ is made smaller, the filter order N increases inversely proportional to it. Therefore, the direct design method yields very costly filters in applications where the transition bandwidth is narrow. For these applications, specialized design methods must be employed.

EFFICIENT METHODS TO DESIGN FIR FILTERS WITH NARROW TRANSITION BANDS

A filter having every one of its delay elements replaced by M delays is called a filter expanded by M (M is a positive integer greater than 1). Expanding a filter by M is equivalent to introduce $M-1$ impulse response coefficients with value zero between each of its coefficients. Its frequency response is periodic over f with period given as $1/M$, where M is the expansion factor, and an expanded filter is so-called a periodic filter.

In the following we provide a review of some popular methods to design low-complexity linear-phase FIR filters having the specifications illustrated in Figure 1. All the methods are explained for the usual case of low-pass filters.

Interpolated Finite Impulse Response (IFIR) and Generalized IFIR methods

The Interpolated Finite Impulse Response (IFIR) method is considered pioneering in the use of periodic subfilters (Milic, 2009). This method uses two subfilters: a model filter and an interpolator filter. Their respective orders are N_m and N_i . Thus, the computational complexity of the IFIR filter is approximately $(N_m + N_i)/2$ multiplications and $N_m + N_i$ additions. The model filter is expanded by a factor M and generates the shape of the desired frequency response with (approximately) M times less complexity than a direct design. However, it introduces unwanted images of that frequency response in the interval $0 < f < 1/2$. The interpolator filter is cascaded to the model filter to remove these unwanted images. To take advantage of the IFIR method, the value M must guarantee that $(N_m + N_i) < N$ holds. The optimum value of M is approximately $1/[f_p + f_s + (f_s - f_p)^{1/2}]$ (Mehrnia & Wilson, 2004).

The generalized IFIR method (Mehrnia & Wilson, 2004) consists in applying the IFIR method to design the interpolator filter of an IFIR filter. The IFIR method is used again to design the resulting interpolator filter, and so on. The process is repeated K times in total. The final filter, called generalized IFIR filter, has $K+1$ model filters, expanded by the values $M_1, M_2, \dots, M_{K+1}$, cascaded with an interpolator. The orders of the model filters are $N_{m,1}, N_{m,2}, \dots, N_{m,K+1}$, and the order of the interpolator filter is N_i .

The value of K is usually found by exhaustive search, i.e., K is set equal to 1 and the resulting generalized IFIR filter is compared with an IFIR filter (where $K = 0$). If the generalized IFIR filter uses a lower number of arithmetic operations than its IFIR counterpart, K is increased by 1 and the result is compared with the generalized IFIR filter designed using the previous K . The value K keeps increasing by one while the computational complexity of the current design is less than that of the previous design. When the aforementioned

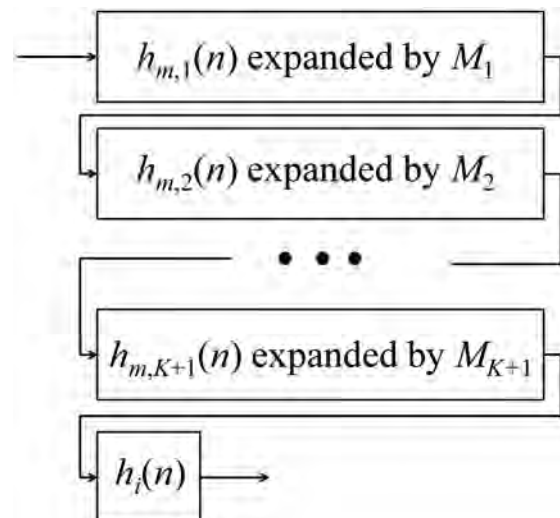
condition no longer holds, the resulting generalized IFIR filter is the one designed in the previous iteration. For a given K , the values $M_1, M_2, \dots, M_{K+1}$ are those that minimize the number of arithmetic operations in the overall filter. Figure 4 shows the structure of a generalized IFIR filter, where $h_{m,k}(n)$ represents the impulse response of the k -th model filter, and $h_i(n)$ represents the impulse response of the interpolator filter. Note that, if only a single model filter is used, the structure corresponds to a traditional IFIR filter.

Frequency Response Masking (FRM)

Let us consider a linear-phase FIR filter that has an impulse response $h(n)$, with $0 \leq n \leq N$, and an even order N , i.e., $N = 2l$ with l integer. Its complementary filter is another filter whose impulse response is $h_c(n) = 1 - h(n)$ for $n = l$ and $h_c(n) = -h(n)$ for other values of n that satisfy $0 \leq n \leq N$. The complementary filter passes the frequencies that its counterpart rejects with a passband deviation being the stopband deviation of its counterpart and vice versa (Milic, 2009).

A filter designed with the Frequency Response Masking (FRM) FRM method is called FRM filter. In the FRM method two filters are created: the filter F1 and the filter F2. F1 is formed by con-

Figure 4. Structure of a Generalized IFIR filter



necting in cascade a filter expanded by M , called expanded model filter, with a masking filter, called masking- m filter, which passes some of the frequency response images of the expanded model filter. The filter F2 is formed by connecting in cascade the complementary filter of the expanded model filter with another masking filter, called masking- c filter, which passes some of the frequency response images of the complementary filter. The filters F1 and F2 share the same input and their outputs are added to form the output of the overall FRM filter. Since knowing the impulse response of the model filter allows knowing the impulse response of its complementary counterpart, only three filters need to be designed in the FRM method: the model filter with impulse response $h_m(n)$ and order N_m (which must be even), its masking filter (masking- m filter) with impulse response $g_m(n)$ and order N_{Gm} and the masking filter of its complementary counterpart (masking- c filter), with impulse response $g_c(n)$ and order N_{Gc} . Figure 5 shows the structure of a FRM filter, where $D_1 = MN_m/2$ and $D_2 = |N_{Gm} - N_{Gc}|/2$. The structure in Figure 5(a) is chosen when $N_{Gc} > N_{Gm}$, otherwise we choose the structure in Figure 5(b).

The value M must guarantee that the sum $N_m + N_{Gm} + N_{Gc}$ is less than N , such that the computational complexity of the FRM filter results lower than the computational complexity of a directly

designed filter. The optimum value of M is near to $0.5[1/(f_s - f_p)]^{1/2}$.

Serial FRM

Much effort has been made to improve the FRM filter structure. One of the most successful methods was presented by (Wei & Lian, 2010), where the use of serial masking schemes was proposed. The insight of this approach is that the expanded model filter and its complementary counterpart do not need to use their own independent masking filters. Similar to the FRM method, three filters need to be designed: the model filter with impulse response $h_m(n)$ and order N_m and two masking filters, called masking-1 and masking-2 filters, whose impulse responses are, respectively, $g_1(n)$ and $g_2(n)$, and whose orders are, respectively, N_{G1} and N_{G2} . The orders N_m , N_{G1} and N_{G2} must be even.

The expanded model filter is first cascaded with the masking-1 filter expanded by an integer L . Then, the output of this connection is added with the delayed output of the complementary filter and finally the output of this addition is cascaded with the masking-2 filter. With this, a lower computational complexity is obtained in comparison to the traditional FRM method because the expanded masking-1 filter has few non-zero coefficients. Since in the resulting structure both

Figure 5. Structure of a FRM filter

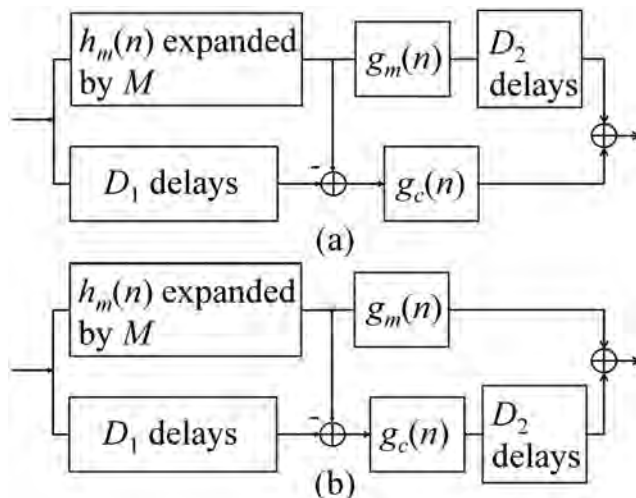
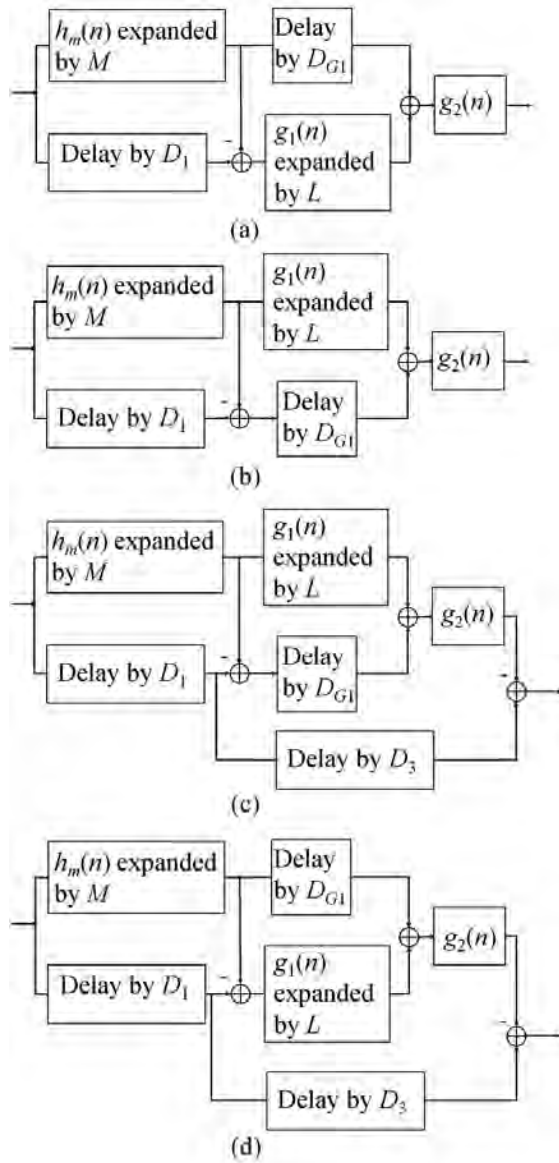
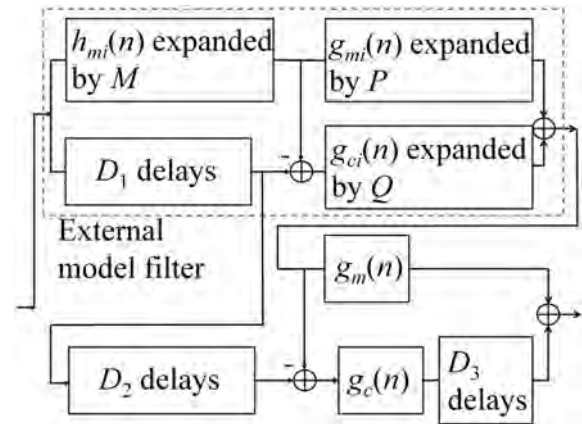


Figure 6. Structure of a Serial FRM filter



masking filters are serially connected, the scheme is called Serial FRM. A filter designed with this method is called serial FRM filter. Figure 6 shows the structures of serial FRM filters for these cases, where $D_{G1} = LN_{G1}/2$, $D_3 = LN_{G1}/2 + N_{G2}/2$. There are four design cases in the serial FRM method. Any structure in Figures 5(a) or 5(b) is used for $f_s < 1/4$. Otherwise, any structure in Figures 5(c) or 5(d) is employed.

 Figure 7. Structure of a Two-stage FRM filter with non-periodic frequency responses ($N_{Gm} > N_{Gc}$ is assumed)


Two-Stage FRM with Non-Periodic Frequency Responses

For cases with very narrow transition bands, designing the expanded model filter of an FRM filter as another FRM filter results in a convenient option. The method recently introduced by (Wei & Liu, 2013) and (Wei, Shaoguang & Ma, 2015) uses that approach.

Similar to the FRM method, the following subfilters are used: the model filter and its complementary counterpart, as well as the masking- m filter, with impulse response $g_m(n)$ and order N_{Gm} , and the masking- c filter, with impulse response $g_c(n)$ and order N_{Gc} . Nevertheless, here the model filter is composed by the following internal subfilters: a model filter expanded by M , its corresponding masking filter expanded by P , and the masking filter of the complementary model expanded by Q . By tuning the integers M , P and Q , an overall low-complexity design can be achieved. Figure 7 shows the structure previously described. The delays D_1 , D_2 and D_3 are obtained in a similar way as explained in the previous section. The name non-periodic is because the external model filter has a non-periodic frequency response.

Advantages of IFIR- and FRM-Based Filters

The advantages of an IFIR- or FRM-based filter in comparison to a directly designed filter are the following:

- The number of adders and multipliers required in the IFIR-/FRM-based filter is less than the number of adders and multipliers required by the directly designed filter.
- Since the computational complexity of the IFIR-/FRM-based filter is less than the computational complexity of the directly designed filter, the former has a lower power consumption.

The price to pay for these advantages is an increase in the number of required delay elements.

Simple Design Guidelines

Even though the decision of using either IFIR-based or FRM-based methods for a given design specification is not straightforward, in the following we present simple guidelines to make easier such choice.

- The IFIR-based methods are suitable for filters with narrow passband, i.e., where $f_s < 1/4$. However, if a filter with wide passband is required, a useful IFIR-based solution consists in designing the complementary counterpart of the desired filter. Then, the complementary version of the designed filter is implemented.
- A region of usefulness of the IFIR method in a plane of band-edge frequencies is presented in Figure 8. R1 is the region for low-pass filters with narrow passband, R2 is for high-pass filters with narrow passband, R3 is for high-pass filters with wide passband and R4 is for low-pass filters with wide passband. If the desired band-edge frequencies of a given filter lie into the re-

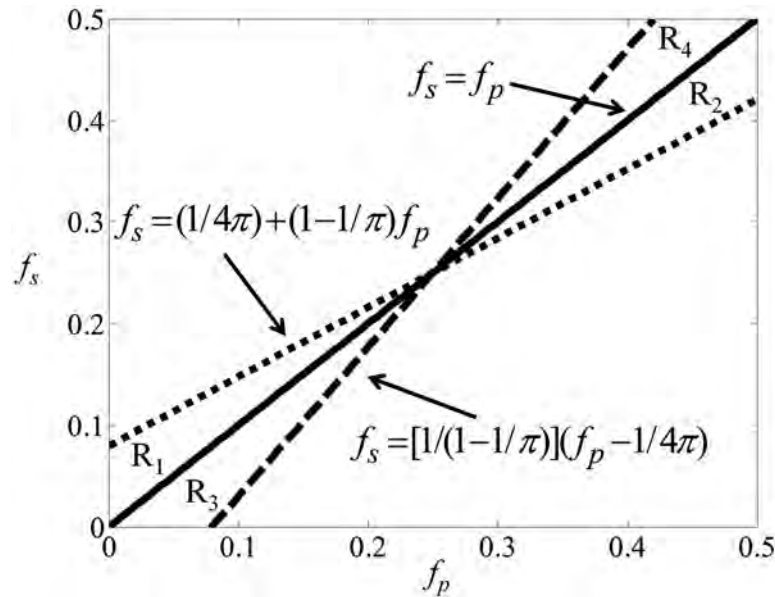
gion of usefulness, it is expected that the IFIR method provides a low-complexity solution.

- The FRM method was proven in (Wei & Liu, 2013) to be effective when the transition band of the desired filter, $f_s - f_p$, is less than $1/16$. Moreover, the FRM method is expected to be more effective than the IFIR method “if the square root of the normalized transition bandwidth is less than the arithmetic mean of the normalized passband edge and stopband edge” (Wei & Liu, 2013, pp.3270). In general, FRM-based methods should be applied when the band-edge frequencies do not lie into the region of usefulness of the IFIR method.
- For cases where $f_s = 1/2 - f_p$, neither the IFIR-based methods nor the Serial FRM method can be used. In this case, the traditional FRM method is better. If the passband and stopband deviations are equal, i.e., $\delta_p = \delta_s$, the method presented by (Lim, Yu & Saramaki, 2005) is particularly useful.

FUTURE RESEARCH DIRECTIONS

The efficient design of digital FIR filters is closely related to the emerging DSP technologies and applications. The hitherto methods are not capable of completely satisfying the plethora of conflicting design requirements like stringent design specifications, low power consumption, high speed of computations and low time and design effort, just to name a few. Therefore, it is important to pursue research towards more general methods to design low-complexity FIR filters, where subfilter-based techniques can be appropriately improved and then applied in a holistic point of view depending on the required application and the implementation platform. However, this task must be delimited taking into account the evolving technologies and their capacities, as well as the priorities in the design goals.

Figure 8. Region of usefulness in terms of band-edge frequencies for IFIR-based filters



Designs based on periodic subfilters can be effectively used, depending on the frequency response specifications, the clock rate available and the metrics for area, speed and power efficiency in different target technologies. However, keeping separated design methods for distinct applications and gradually migrating to most general methods whenever is possible is a wise decision. In this sense, the emerging design of sparse filters (filters where the number of non-zero coefficients is minimized, subject to a desired magnitude response specification) can be also investigated.

Recently, Field Programmable Gate Array (FPGA) platforms have gained a tremendous ground in the field of DSP and therefore this can be the target technology for future research. The current FPGA technologies include a much wider number of dedicated DSP resources that can be efficiently used for schemes based in periodic subfilters. The efficient integration of these resources to design low-complexity FIR filters with increasingly strict frequency response specifications is also an important research to be carried out.

CONCLUSION

In this chapter we have introduced the concept of digital filter as a system that modifies the characteristics of discrete-time signals. Particularly, Linear Time-Invariant (LTI) filters with Finite Impulse Response (FIR) characteristic have been addressed because of their advantages, namely, guaranteed stability and feasibility to be designed with linear phase. We have presented the characteristics of the impulse response of a FIR filter with linear phase, and it has been shown how to estimate the order of a linear-phase filter from its specifications of band-edge frequencies and passband/stopband magnitude deviations.

The main focus of this chapter has been the efficient design of linear-phase FIR filters. Since FIR filters with narrow transition band have a high computational complexity when they are implemented in direct-form, we have presented some of the most efficient methods to design linear-phase FIR filters with narrow transition band and low computational complexity. These methods are based on the use of periodic subfil-

ters. Particularly, the Interpolated Finite Impulse Response (IFIR) method and its generalized version, as well as the Frequency Response Masking (FRM) method and its recent modifications, have been concisely explained.

The end of the chapter presents simple guidelines to decide whether using IFIR-based or FRM-based methods for a given set of band-edge frequencies. Additionally, future research directions on the design of low-complexity FIR filters have been highlighted. Even though the set of existing methods for efficient design of FIR filters is vast, the present chapter offers a solid reference that explores two of the most useful methods to design low-complexity linear-phase FIR filters.

REFERENCES

Aksoy, L., Flores, P., & Monteiro, J. (2014). A tutorial on multiplierless design of FIR filters: Algorithms and architectures. *Circuits, Systems and Signal Processing*, 33(6), 1689–1719. doi:10.1007/s00034-013-9727-8

Antoniou, A. (2006). *Digital Signal Processing: Signals, Systems and Filters*. New York: McGraw-Hill.

Lim, Y. C., Yu, Y. J., & Saramaki, T. (2005). Optimum masking levels and coefficient sparseness for Hilbert transformers and half-band filters designed using the frequency response masking technique. *IEEE Transactions on Circuits and Systems I*, 52(11), 2444–2453. doi:10.1109/TCSI.2005.853518

Mehrnia, A., & Wilson, A. N. Jr. (2004). On Optimal IFIR filter design. *Proceedings of IEEE International Symposium on Circuits and Systems*, 3, 133–136.

Milic, L. (2009). *Multirate filtering for digital signal processing*. Hershey, PA: IGI-Global. doi:10.4018/978-1-60566-178-0

Tan, L., & Jiang, T. (2013). *Digital Signal Processing: Fundamentals and applications* (2nd ed.). Waltham, MA: Elsevier.

Taylor, F. J. (2012). *Digital Filters: Principles and applications with MATLAB*. Hoboken, NJ: John Wiley & Sons.

Wei, Y., & Lian, Y. (2010). Frequency-response masking filters based on serial masking schemes. *Circuits, Systems and Signal Processing*, 29(1), 7–24. doi:10.1007/s00034-009-9112-9

Wei, Y., & Liu, D. (2013). Improved design of frequency-response masking filters using band-edge shaping filter with non-periodical frequency response. *IEEE Transactions on Signal Processing*, 61(13), 3269–3278. doi:10.1109/TSP.2013.2259485

Wei, Y., Shaoguang, H., & Ma, X. (2015). Improved design of frequency-response masking filters using band-edge shaping filter with non-periodical frequency response. *IEEE Transactions on Signal Processing*, 61(13), 3269–3278. doi:10.1109/TSP.2013.2259485

ADDITIONAL READING

Alam, S. A., & Gustafsson, O. (2011). Implementation of Narrow-Band Frequency-Response Masking for Efficient Narrow Transition Band FIR Filters on FPGAs, *Proceedings of IEEE NORCHIP*, 1–4. doi:10.1109/NORCHIP.2011.6126702

Alam, S. A., & Gustafsson, O. (2011). Implementation of Time-multiplexed sparse periodic FIR filters for FRM on FPGAs, *Proceedings of IEEE International Symposium on Circuits and Systems*, 661–664. doi:10.1109/ISCAS.2011.5937652

Baran, T., Wei, D., & Oppenheim, A. V. (2010). Linear programming algorithms for sparse filter design. *IEEE Transactions on Signal Processing*, 58(3), 1605–1617. doi:10.1109/TSP.2009.2036471

- Jiang, A., Kwan, H. K., & Zhu, Y. (2012). Peak-Error-Constrained Sparse FIR Filter Design Using Iterative SOCP. *IEEE Transactions on Signal Processing*, 60(8), 4035–4044. doi:10.1109/TSP.2012.2199316
- Johansson, K., Gustafsson, O., DeBrunner, L. S., & Wanhammar, L. (2011). Minimum adder depth multiple constant multiplication algorithm for low power FIR filters. *Proceedings of IEEE International Symposium on Circuits and Systems*, 1439–1442. doi:10.1109/ISCAS.2011.5937844
- Kong, B. Y., & Park, I.-C. (2012). FIR Filter Synthesis Based on Interleaved Processing of Coefficient Generation and Multiplier-Block Synthesis. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 31(8), 1169–1179. doi:10.1109/TCAD.2012.2190826
- Laddomada, M., Romero, D. E. T., & Dolecek, G. J. (2011). Design of multiplierless decimation filters using an extended search of cyclotomic polynomials. *IEEE Transactions on Circuits and Systems II*, 58(2), 115–119. doi:10.1109/TCSII.2010.2104012
- Lehto, R., Saramaki, T., & Vainio, O. (2007). Synthesis of narrowband linear-phase FIR filters with piece-wise polynomial impulse response. *IEEE Transactions on Circuits and Systems I*, 54(10), 2262–2276. doi:10.1109/TCSI.2007.904591
- Lehto, R., Saramaki, T., & Vainio, O. (2010). Synthesis of wide-band linear-phase FIR filters with a piecewise-polynomial-sinusoidal impulse response. *Circuits Systems and Signal Processing*, 29(1), 25–50. doi:10.1007/s00034-009-9111-x
- Lehto, R., Tauren, T., & Vainio, O. (2011). Recursive FIR filter structures on FPGA. *Microprocessors and Microsystems*, 35(7), 595–602. doi:10.1016/j.micpro.2011.08.012
- Lian, Y., Yang, C. Z., & Lim, Y. C. (2007). Complexity reduction of FRM filters via multiplication-free prefilter-equalizer structures. *Proceedings of IEEE International Symposium on Circuits and Systems*, pp. 2958–2961. doi:10.1109/ISCAS.2007.377870
- Lu, W.-S., & Hinamoto, T. (2010). Digital filters with sparse coefficients. *Proceedings of IEEE International Symposium on Circuits and Systems*, 169–172.
- Lyons, R. (2007). Turbocharging Interpolated FIR filters. *IEEE Signal Processing Magazine*, 24(5), 140–143. doi:10.1109/MSP.2007.904739
- Sheikh, Z. U., & Gustafsson, O. (2010). Design of sparse non-periodic narrow-band and wide-band FRM-like FIR filters. *IEEE 1st International Conference on Green Circuits and Systems*, 279–282.
- Sheikh, Z. U., & Johansson, H. (2011). A Class of Wide-Band Linear-Phase FIR Differentiators Using a Two-Rate Approach and the Frequency-Response Masking Technique. *IEEE Transactions on Circuits and Systems I*, 58(8), 1827–1839. doi:10.1109/TCSI.2011.2107270
- Sheikh, Z. U., & Johansson, H. (2012). A Technique for Efficient Realization of Wide-Band FIR LTI Systems. *IEEE Transactions on Signal Processing*, 60(3), 1482–1486. doi:10.1109/TSP.2011.2178408
- Sheikh, Z. U., & Johansson, H. (2012). Efficient Wide-Band FIR LTI Systems Derived Via Multi-Rate Techniques and Sparse Bandpass Filters. *IEEE Transactions on Signal Processing*, 60(7), 3859–3863. doi:10.1109/TSP.2012.2194712
- Shen, T., & Lim, Y. C. (2011). Low-complexity frequency-response masking filters using modified structure based on serial masking. *19th European Signal Processing Conference EUSIPCO*, 1400–1404.
- Shi, D., & Yu, Y. J. (2010). Low-complexity linear-phase FIR filters in cascade form. *Proceedings of IEEE International Symposium on Circuits and Systems*, 177–180. doi:10.1109/ISCAS.2010.5538008
- Shi, D., & Yu, Y. J. (2011). Design of discrete-valued linear-phase FIR filters in cascade form. *IEEE Transactions on Circuits and Systems I*, 58(7), 1627–1636. doi:10.1109/TCSI.2011.2143250

Vinod, A. P., Lai, E., Maskell, D. L., & Meher, P. K. (2010). An improved common subexpression elimination method for reducing logic operators in FIR filter implementations without increasing logic depth. *Integration. The VLSI Journal*, 43(1), 124–135. doi:10.1016/j.vlsi.2009.07.001

Yao, C.-Y., Hsia, W.-C., & Ho, Y.-H. (2014). Designing Hardware-Efficient Fixed-Point FIR Filters in an Expanding Subexpression Space. *IEEE Transactions on Circuits and Systems*, 61(1), 202–212. doi:10.1109/TCSI.2013.2268551

Ye, W. B., & Yu, Y. J. (2013). Single stage and cascade design of high order multiplierless linear phase FIR filters using genetic algorithm. *IEEE Transactions on Circuits and Systems*, 60(11), 2987–2997. doi:10.1109/TCSI.2013.2256211

Yli-Kaakinen, J. & T. Saramaki. (2011). An efficient algorithm for the optimization of FIR filters synthesized using the multistage frequency-response masking approach. *Circuits, Systems and Signal Processing*, (30), 157-183.

Yu, Y. J., Shi, D., & Lim, Y. C. (2009). Design of extrapolated impulse response FIR filters with residual compensation in subexpression space. *IEEE Transactions on Circuits and Systems I*, 56(12), 2621–2633. doi:10.1109/TCSI.2009.2016165

KEY TERMS AND DEFINITIONS

Adder: A component of a digital filter that produces at its output the sum of the values of the samples present at its inputs.

Complex Exponential Signal: A signal whose samples are complex numbers, where the real and imaginary parts of the samples form, respectively, a cosine wave and a sine wave, both with the same frequency.

Complex Function: A function whose values are complex numbers.

Complex Number: A number composed by the sum of a real number a and an imaginary number $i \times b$, where b is a real number and i is defined as the square root of -1 .

Delay: A component of a digital filter that has one input and one output. The delay retains its input value during a sampling period T_s and releases it at the next sampling interval.

Lineal Time-Invariant (LTI): Let us consider a filter that has as input a scaled sum of arbitrary discrete-time signals. If the output of this filter is equal to the scaled sum of the outputs of the filter to every one of these arbitrary signals, the filter is considered linear. If additionally the input signal is delayed in any possible amount and the output signal undergoes the same delay as the input, the filter is also considered time-invariant. A filter is LTI if it has both, linear and time-invariant properties.

Magnitude Response: A function of the frequency f where every value is obtained as the magnitude of the complex value of the frequency response in that frequency f . If the value of the frequency response in f is a complex number of the form $a(f) + i \times b(f)$, the magnitude of that number is given by $\{[a(f)]^2 + [b(f)]^2\}^{1/2}$.

Multiplier: A component of a digital filter that takes the samples at its inputs and produces the multiplication of these values at its output.

Phase Response: A function of the frequency f where every value is obtained as the phase of the complex value of the frequency response in that frequency f . If the value of the frequency response in f is a complex number of the form $a(f) + i \times b(f)$, the phase of that number is given by $\arctan\{b(f)/a(f)\}$.

Open Source



Heidi Lee Schnackenberg
SUNY Plattsburgh, USA

INTRODUCTION

Open source is a philosophy and a methodology associated with free and collaborative, creation, modification and use of software applications and operating systems. The term “open source” was coined in 1998 at the “Open Source Summit” in Palo Alto, California in an effort to standardize both the idea and a word or phrase for it. Shortly thereafter, the “Open Source Initiative” (OSI) was formed and sought to promote the idea of free applications (specifically, free source code) to the public (Tiemann, 2009).

Open source is characterized by a particular creation process associated with the development of software and/or operating systems. It is protected by specific licenses which allow users and developers the ability to freely modify, upgrade, use and distribute the products as they wish (von Krogh & Spaeth, 2007). Although not relegated to a hierarchical structure, a project manager generally emerges and leads the initiative by keeping things organized and moving forward while volunteers offer their talents by designing, coding, debugging, beta testing, and utilizing the product (Crowston & Howison, 2006). Volunteers can also offer feedback and provide technical support, either in-person, via phone, or in online user forums. This process can end in the free proliferation of an open source product, which users can modify and improve according to their specific needs. Although open source products can have imperfections, the user isn’t paying for those deficiencies like they sometimes do with commercial products (Schnackenberg & Vega, 2010). At times, open source products can also lead to a profitable venture if the software is sold to a company and marketed commercially. O’Neill

(2012) notes that open source can be viewed as a change agent. It has altered the way software and operating systems are produced and marketed, and has even modified the way knowledge is created. In this way, open source has also redefined business and marketing models around the world (Bonaccorsi, Giannageli, & Rossi, 2006; von Krogh & Spaeth, 2007).

BACKGROUND

An excellent example of this is the way in which open source has made deep inroads into the realm of mobile technologies (Chao, 2011). While “free” applications for mobile devices – smart phones, cell phones, tablets, etc. – are available for download, they are not open source products. These “free” applications, or “apps” are merely trial or limited versions of more complete products that users are tempted to buy in order to get the most complete functionality (Schnackenberg, Vega, & Heymann, 2014). True open source mobile applications come complete to the user when downloaded, and their source code is available for altering (Syer, Nagappan, Hassan, & Adams, 2013). In fact, the operating system of the Android smart phone itself is open source (Butler, 2011). Presently, the number of open source mobile applications (apps) available to download on smart phones that assist the user in completing day-to-day activities is growing exponentially. It is likely that smart phone, cell phone, and tablet users all search for free, open source, “apps” before they download the ones that are commercially produced. Given that, it is clear that open source has indeed impacted mobile technologies and will continue to do so for the foreseeable future.

DOI: 10.4018/978-1-5225-2255-3.ch543

In many senses, open source is a prevalent common practice of the hacker community (Sarma & Matheus, 2015; Soderberg & Delfanti, 2015), and a strongly held belief of technological idealists (Lakhani & Wolf, 2007). It is perhaps a type of high-tech grassroots movement. Open source products and applications are a manifestation of these practices and beliefs from which lay people can benefit. In his book, *The Cathedral and the Bazaar*, Raymond (2005) famously likens the culture and the creation process of open source to a bazaar, or an open market, where everyone has some unique goods to offer that all interplay for the benefit of the whole. Raymond contrasts this to the very hierarchical way in which ancient cathedrals were built. Despite its philosophical underpinnings and model of operations running entirely counter to conventional wisdom about how knowledge is created and how traditional business models function, the open source movement has taken a strong foothold in our technological culture (Weber, 2005).

Issues, Controversies, Problems

Several challenges with open source and open source products exist with varying degrees of frustration for developers and users (Meeker, 2008). The following is a description of some of the most prevalent issues with open source:

- **Confusing Licensure:** For the most part, OSS is free to modify for individual user needs. *Copyleft* is a term commonly associated with open source. It is a form of licensing that makes use of copyright law to ensure that a software, and its subsequent upgrades, can always be freely modified by users (van Holst, 2013). Different open source products, or different versions of open source products, can carry varying restrictions and usage policies. It is difficult to tell which software you can modify in what ways because there is no consistency in the licensing policies. At times, it

could be risky to utilize certain software in a particular way, or combine other software for a specific function, or to even create an entirely different product, for fear of violating licensing restrictions. To read all of the varying agreement policies that a developer or group of developers can attach to open source software can be extremely time-consuming and frustrating;

- **Collaborative Creation Process Doesn't Necessarily Mean That the Product Will Be Better:** One of the great attractions of the open source movement is that products can be modified by a developer or group of developers (Tapscott & Williams, 2006). The idea of software creation being a community process is novel to the computer application industry. While most commercially produced software is highly propriety and the source code and creation process quite secretive, open source products are a highly collaborate event. Admittedly, the idea of making software free and open to the public is an enormous, and attractive, economic and philosophical shift. However, sometimes too many individuals on a project can make the process inefficient and the product inferior. This phenomena is commonly referred to as "Brook's Law" (Brooks, 1886) and is particularly applicable when it comes to working on source code. Computer coding is a very detail-oriented, careful skill and often if one programmer makes an error when coding, it is difficult for another programmer to find and fix it. In this way, a piece of software can become either unfriendly to use or at worst, dysfunctional. Indeed, sometimes adding more software developers to a project doesn't necessarily make the product better;
- **Open Source Products Are Complicated to Use:** Linux-based products were originally designed for people who understand the concept of what makes Linux work.

The purest form of Linux requires a solid understanding of computer programming. One of the reasons for this is that by using what is known as typed “command lines,” a user can gain direct access to a particular application without having to wait to locate the application, then for the “pretty” graphics to reveal the application, and finally for the application to open in order to work. For many people who are accustomed to working in Windows or Apple environments, the idea of having to type a command that will open even a simple word processing application doesn’t seem worth the trouble (especially since the user would need to troubleshoot his/her command if an error is encountered). Therefore, the Graphical User Interface, or GUI, is vital in winning over users of traditional operating systems. Currently, more and more Linux-based operating systems now have GUIs for those who want to try Linux without the fear of the “command line” and allow users to simply point-and-click in order to access any application;

- **Similarly, Open Source Products Can Often Be Difficult to Install:** This is because, unlike commercial products, users are given options as to where to install software or operating systems. Open source applications can be installed in a variety of places on a computer’s hard drive, an external hard drive, or even on a jump drive or small USB drive. Conversely, commercial software simply leads the user through a pre-determined set of directions where s/he simply needs to accept where the applications are being installed. Granted, there are some choices as to where to install commercial software on a computer, but these are fairly limited. However, many users prefer the easiest method of software installation and for some, this means the least amount of choices and decisions possible;
- **A Final Factor That Can Add to the Complexity of Using Open Source Products is That Not All Software Runs on Open Source Operating Systems:** Running a GNU/Linux-based operating system requires users to understand that the commercial or proprietary applications that would normally run on Windows or Apple operating systems will not always run on GNU/Linux. Some applications are simply not compatible with others and it may take some trial-and-error to find out which programs work with which operating systems. While this sort of investigation may seem adventurous to some users, to others it would feel frustrating and inefficient. (N.B. There is, in fact, a way to enable Windows applications to run on GNU/Linux operating systems, but the process is extremely complicated and not for the faint of heart);
- **No Formal Help/Support and Poor Documentation:** Perhaps the least attractive aspect of using open source products is the fact that there is no formal support or help if the user has technical or usability problems. With proprietary products, manuals virtually always accompany them and often a helpline is accessible for specific hours each day. Interestingly, the open source community has responded to this by creating open, online, user groups, forums, and even videos on how to use, modify, and/or repair open source software. While the price is clearly right, corporations may often use commercial products because they are promised 24-7 technical support, free upgrades (for a specified time period), or they are given the software with the purchase of hardware. These perks certainly make the use of commercially-produced software attractive, particularly the promise of consistent technical support. Fortunately, open-source applications have come a long way. Many of today’s open-source products have associated technical

support in some capacity, sustained by the open-source community. Therefore, when a user looks up tutorials on how to use a specific application, a long list of individualized sites that can satisfy a variety of needs will usually appear. Still, this more prevalent appearance of technical support for open source may not be enough for the business community, where time means money and answers need to be obtained yesterday.

SOLUTIONS AND RECOMMENDATIONS

Despite facing some significant challenges, the open source movement has made strides to improve or correct the problems that it faces. Organizations like Creative Commons (2015) have provided a standardized way for its members to share their innovations through user-friendly copyright licenses. Its staff and volunteers assist users with tools for attaching the licensures to their work so that it can be shared, modified and utilized in the spirit of the open source initiative. Groups like Creative Commons have recognized and responded the confusing copyright issues that can accompany open source so that more people can customize and use the products without frustration or worry.

Similarly, many researchers (Bonaccorsi, Giannageli, & Rossi, 2006; Casadesus-Masanell & Llanes, 2011; Crowston & Howison, 2006) have studied how the collaborative process occurs in open source development. As noted, often, “too many cooks in the kitchen” can create an open source product that is ineffective. The same phenomena can also create an inefficient process, despite the quality of the product. The open source community has recognized that global collaboration may not always be the best approach for software development and several solutions to the problem have been suggested and implemented. Most notably, some structure or framework has

begun exist when a project is proposed and source code is made public. Project managers are often attached to the task and provide some organization for the work process. Alternately, a mixed model approach is adopted where some segments of commercial software development are adopted and a somewhat hierarchical structure exists. Another aspect of the collaborative process may be that developers already working on a project reserve the right to invite others to participate or deny some that right. Regardless of the approach taken to the shared -development model with open source undertakings, in most cases, some sort of organizational structure has begun to exist.

FUTURE RESEARCH DIRECTIONS

Although a significant amount of research has been done on open source since its inception (Raasch, Lee, Spaeth, & Herstatt, 2013), there are still many aspects of the movement and its associated initiatives that have yet to be investigated. For instance, the fact that free software and free operating systems are so readily available, yet consumers still prefer to pay for commercial products (which may or may not be better), is still intriguing. Commercial products are often released with well-noted errors and problems, similar to open source applications, yet the public still prefers to purchase this merchandise when they can acquire something of similar quality for free. This phenomenon is particularly puzzling and would be an interesting area of research to pursue.

Another aspect of open source worth investigating is how individuals come to be informed about the products and applications. A notable problem with open source is its lack of notoriety. How the public finds out about open source would be valuable information that could aid in its proliferation and adoption. It would therefore be beneficial to the open source community to find out how the public comes to know about its initiatives.

Social networks and open source is a new area also worthy of future study. While the so-

cial structures and norms of working within the open community and/or an open source project have been investigated to some extent, the use of open applications for what we more commonly think of as social networking has not yet been examined. Sites such as Facebook, Instagram, Snapchat, etc. are very prevalently used by many people, and accounts are free, but the code is not open for users to modify. These websites are also for-profit, despite the fact that they are free to use (for the most part) for patrons. It would be intriguing to explore what would happen if these social networking sites opened their source codes and allowed users to modify them according to their own needs.

Finally, it might be interesting for the open source community to do some predictive research and look to the future to see where technological innovation is headed and where open source applications might be most useful and/or popular. Intriguingly, when the author first began writing about the open source movement, mobile technologies were in their infancy. Who could have (or should have) predicted that open source applications for cell phones, smart phones, tablets, etc., would be where open source would have its strongest foothold in present day? It may benefit the open source community to do some forward thinking and perhaps anticipate this kind of opportunity in the future.

CONCLUSION

Open source has, in one form or another, permeated almost every aspect of modern technologies. Particularly with the advent and proliferation of mobile technologies, and the corresponding free “apps” needed to make them usable, open source has firmly rooted itself in the present day functioning of global citizens. Free source code, enabling programmers to modify a product to specifically suit their needs, is an enduringly attractive phenomenon. The coordination of a group

of programmers working on the same application can be challenging, especially when there is no formal hierarchical structure, but that may also be part of the allure of open source. Although commercially produced software is still the popular choice for most of the public, open source is a stalwart second place and rapidly gaining traction. Whether becoming part of the movement, or simply remaining a spectator, the open source movement is certainly one to observe closely in years to come.

REFERENCES

- Bonaccorsi, A., Giannageli, S., & Rossi, C. (2006). Entry strategies under competing standards: Hybrid business models in the open source software industry. *Management Science*, 52(7), 1085–1098. doi:10.1287/mnsc.1060.0547
- Brooks, W.K. (1886). Report on the Stomatopoda collected by the H.M.S. Challenger during the years 1873-1876. *Challenger Zoological Reports, Part 45*.
- Butler, M. (2011). Android: Changing the mobile landscape. *Pervasive Computing, IEEE*, 10(1), 4–7. doi:10.1109/MPRV.2011.1
- Casadesus-Masanell, R., & Llanes, G. (2011). Mixed source. *Management Science*, 57(7), 1212–1230. doi:10.1287/mnsc.1110.1353
- Chao, L. (2011). *Open source mobile learning: Mobile Linux applications*. Hershey, PA: IGI Global. doi:10.4018/978-1-60960-613-8
- Comino, S., & Manenti, F. M. (2005). Government policies supporting open source software for the mass market. *Review of Industrial Organization*, 26(2), 217–240. doi:10.1007/s11151-004-7297-4
- Creative Commons. (2015). *Creative Commons*. Retrieved October 13, 2015 from <http://creativecommons.org>

- Crowston, K., & Howison, J. (2006). Hierarchy and centralization in free and open source software team communications. *Knowledge, Technology & Policy*, 18(4), 65–85. doi:10.1007/s12130-006-1004-8
- Lakhani, K., & Wolf, R. G. (2007). Why Hackers Do What They Do: Understanding Motivation and Effort in Free/Open Source Software Projects. In J. Feller, B. Fitzgerald, S. A. Hissam, & K. R. Lakhani (Eds.), *Perspectives on free and open source software* (pp. 3–22). Cambridge, MA: MIT Press.
- Meeker, H. J. (2008). *The open source alternative: Understanding risks and leveraging opportunities*. Hoboken, NJ: John Wiley & Sons, Inc.
- O'Neill, A. (2012). Open source software. In R. Chadwick (Ed.), *The Encyclopedia of Applied Ethics* (2nd ed.). Maryland Heights, MO: Elsevier Inc. doi:10.1016/B978-0-12-373932-2.00063-6
- Raasch, C., Lee, V., Spaeth, S., & Herstatt, C. (2013). The rise and fall of interdisciplinary research: The case of open source innovation. *Research Policy*, 42(2).
- Sarma, M., & Mattheus, T. (2015). Knowledge creation in virtual communities – exploring practices in open source software hacker communities. *The International Technology Management Review*, 5(2), 94–104. doi:10.2991/itm.2015.5.2.4
- Schnackenberg, H. L., & Vega, E. S. (2010). Open Source and Bridging the Digital Divide: A Case Study. In R. Luppigini & A. K. Haghi (Eds.), *Cases on Digital Technologies in Higher Education: Issues and Challenges* (pp. 24–33). Hershey, PA: IGI Global. doi:10.4018/978-1-61520-869-2.ch002
- Schnackenberg, H. L., Vega, E. S., & Heymann, M. J. (2014). Open Source and the Digital Divide. In M. Khosrow-Pour (Ed.), *Encyclopedia of Information Science and Technology*. Hershey, PA: IGI Global.
- Soderberg, J., & Delfanti, A. (2015). Hacking hacked! The life cycles of digital innovation. *Science, Technology & Human Values*, 40(5), 793–798. doi:10.1177/0162243915595091
- Syer, M. D., Nagappan, M., Hassan, A. E., & Adams, B. (2013, November). Revisiting prior empirical findings for mobile apps: An empirical case study on the 15 most popular open-source Android apps. *Proceedings of the Centre for Advanced Studies Conference*, 23, 283–297.
- Tapscott, D., & Williams, A. D. (2006). *Wikinomics: How Mass Collaboration Changes Everything*. New York, NY: Penguin Group.
- Tiemann, M. (2009). *History of the OSI*. Retrieved April 17, 2013 from <http://www.opensource.org/history>
- van Holst, W. (2013). Copyleft, -right and the case law on APIs on both sides of the Atlantic. *International Free and Open Source Software Review Law*, 5(1).
- von Krogh, G., & Spaeth, S. (2007). The open source software phenomenon: Characteristics that promote research. *The Journal of Strategic Information Systems*, 16(3), 236–253. doi:10.1016/j.jsis.2007.06.001
- Weber, S. (2005). *The success of open source*. Cambridge, MA: Harvard University Press.

ADDITIONAL READING

- Bonaccors, A., Giannangeli, S., & Rossi, C. (2006). The penguin has entered the building: The commercialization of open source software products. *Management Science*, 52(7), 1085–1098.
- Casadesus-Masanell, R., & Ghemawat, P. (2006). Dynamic mixed duopoly: A model motivated by Linux vs. Windows. *Management Science*, 52(7), 1072–1084. doi:10.1287/mnsc.1060.0548

- Chesbrough, H. W., & Appleyard, M. M. (2007). Open innovation and strategy. *California Management Review*, 50(1), 57–76. doi:10.2307/41166416
- Choi, C. J., Kim, S. W., & Yu, S. (2009). Global ethics of collective internet governance: Intrinsic motivation and open source software. *Journal of Business Ethics*, 90(4), 523–531. doi:10.1007/s10551-009-0057-5
- Costa, J. P. (2012). *Digital learning for all, now: A school leader's guide for 1:1 on a budget*. Thousand Oaks, CA: Corwin.
- DiBona, C., Stone, M., & Cooper, D. (2005). *Open sources 2.0: The continuing evolution*. Sebastopol, California: O'Reilly Media.
- Economides, N., & Katsamakos, E. (2006). Two-sided competition of proprietary vs. open source technology platforms and the implications for the software industry. *Management Science*, 52(7), 1057–1071. doi:10.1287/mnsc.1060.0549
- Fitzgerald, B. (2006). The transformation of open source software. *Management Information Systems Quarterly*, 30(3), 587–598.
- Fosfur, F., Giarratana, M. S., & Luzzi, A. (2008). The penguin has entered the building: The commercialization of open source software products. *Organization Science*, 19(2), 292–305. doi:10.1287/orsc.1070.0321
- Hars, A., & Ou, S. (2002). Working for free? Motivations for participating in open-source projects. *International Journal of Electronic Commerce*, 6(3), 25–39.
- Iqbal, J., Rasool, T., & Quadri, S. M. K. (2011). Open source systems and engineering: Strengths, weaknesses and prospects. *Trends in Information Management*, 7(2), 188–197.
- Lee, J., & Ware, B. (2002). *Open source development with lamp: Using Linux, Apache, MYSQL, PERL, and PHP*. Boston, MA: Addison-Wesley Professional.
- Lemley, M. A., & Shafir, Z. (2006). Who chooses open-source software? *The University of Chicago Law Review. University of Chicago. Law School*, 78(1), 139–164.
- Lerner, J., & Tirole, J. (2002). Some simple economics of open source. *The Journal of Industrial Economics*, 50(2), 197–234. doi:10.1111/1467-6451.00174
- Lerner, J., & Tirole, J. (2005). The scope of open source licensing. *Journal of Law Economics and Organization*, 21(1), 20–56. doi:10.1093/jleo/ewi002
- Moody, G. (2002). *Rebel code: Linux and the open source revolution*. New York, NY: Basic Books.
- Raymond, E. S. (2001). *The cathedral & the bazaar: Musings on Linux and open source by an accidental revolutionary*. Sebastopol, California: O'Reilly Media.
- Raymond, E. S. (2005). *The cathedral and the bazaar*. Snowball Publishing.
- Roberts, J. A., Hann, I., & Slaughter, S. A. (2006). Understanding the motivations, participation, and performance of open source software developers: A longitudinal study of the apache projects. *Management Science*, 52(7), 984–999. doi:10.1287/mnsc.1060.0554
- Sen, R. (2007). A strategic analysis of competition between open source and proprietary software. (2007). *Journal of Management Information Systems*, 24(1), 233–257. doi:10.2753/MIS0742-1222240107
- Stewart, K. J., & Gosain, S. (2006). The impact of ideology on effectiveness in open source software development teams. *Management Information Systems Quarterly*, 30(2), 291–314.
- von Krogh, G., & von Hippel, E. (2006). The promise of research on open source software. *Management Science*, 52(7), 975–983. doi:10.1287/mnsc.1060.0560

Weber, S. (2005). *The success of open source*. Cambridge, Massachusetts: Harvard University Press.

Williams van Rooij, S. (2009). Adopting open-source software applications in U.S. higher education: A cross-disciplinary review of the literature. *Review of Educational Research*, 79(2), 682–701. doi:10.3102/0034654308325691

KEY TERMS AND DEFINITIONS

Open Source: A philosophy and a methodology associated with free and collaborative, creation, modification and use of software applications and operating systems.

FOSS: Free Open Source Software.

FLOSS: Free and Legal Open Source Software.

Web 2.0: In the broadest sense, the second iteration of the World Wide Web (WWW), where web pages allow users to communicate, collaborate, and construct knowledge via social networks, video sharing sites, blogs, wikis, etc.

Intellectual Property: Ideas, creations, and inventions that are the exclusive property of the owners.

Linux: A Unix-compatible operating system that was collaboratively developed and is free to the public.

OLPC: One Laptop per Child.

Power Consumption in Wireless Access Networks

M**Vinod Kumar Mishra***B. T. Kumaon Institute of Technology, India***Pankaja Bisht***B. T. Kumaon Institute of Technology, India*

INTRODUCTION

Early wireless systems consisted of a base station with a high-power transmitter and served a large geographic area having less number of users. The systems were isolated from each other and only a few of them communicated with the public switched telephone networks. Today, the cellular systems consist of a cluster of base stations with low-power radio transmitters. Each base station serves a small cell within a large geographic area. The total number of users served is increased because of channel reuse and also larger frequency bandwidth. The cellular systems connected with each other via mobile switching center and directly access the public switched telephone networks. The most advertised advantage of wireless communication systems is that a mobile user can make a phone call anywhere and anytime.

The power consumption of ICT is approximately 4% of the annual energy production and it is expected that this number will double within the next 10 to 15 years. Internet traffic has a compound annual growth rate of 40%, which is an increase in two fold in every two years. The wireless world research forum (WWRF) has a vision of 7 trillion wireless devices serving 7 billion users by 2017 (Deruyck et al., (2011)). This indicates that the power consumption of wireless access networks become an important and crucial issue in the coming years. Earlier work also showed that the radio access network is a large contributor to CO₂ emissions. Particularly, the base stations

are responsible for roughly two-thirds of the CO₂ emissions of these radio access networks. NTT DoCoMo states that the daily energy consumption per customer is 0.83 Wh for a terminal and 120 Wh for the mobile network, which is a consumption ratio of terminal versus networks of about 1:150 (Jimaa et al., (2011)). The energy consumption of the terminals is thus negligible with respect to the energy consumption of the networks. Several studies indicate that within communication networks, the wireless access networks are high power consumers. Especially the base stations (BSs) are responsible for a significant part of the power consumption caused by wireless access networks. Therefore, it is clear that we should focus on the base stations in the wireless access networks in order to reduce the power consumption.

This chapter provides an overview and comparison of the power consumption of some of existing technologies of wireless access network and suggests the optimum technology for a specific area on the basis of power consumption per covered area. The recent significant literature and some frontier of this area are also included in this chapter.

BACKGROUND

The power consumption and energy efficiency of telecommunications networks become more and more important. For wireless access networks, most studies focus on energy-efficient deploy-

ments of wireless access networks. These studies shows that the user try to reduce the number of base stations in the network by switching them off during periods of low traffic which is a very promising method for energy savings. These studies in combination with a realistic power consumption model for base stations will enable to quantify the actual power savings in the network.

Jain, (2007), presented a description of the different Path Loss models. He gathered information from various IEEE and ITU standards. In Deruyck et al. (2014), authors presented the power consumption model for macrocell and microcell base stations and energy efficiency of three different wireless technologies (mobile WiMAX, LTE, and HSPA) has been compared. In Alshami et al. (2011), authors presented a comprehensive study of frequency analysis for different path loss on mobile WIMAX. They

discussed and implemented Okumura, Hata, Cost-231, Ericsson, Erceg, Walfish, Ecc-33, Lee and the simplified free space path loss models. In Deruyck et al. (2011), authors calculated the power consumption of base stations for mobile WiMAX, fixed WiMAX, UMTS, HSPA, and LTE. They introduced a new metric, the power consumption per covered area PCarea and investigated the influence of MIMO. In Deruyck et al. (2012), they categorized the base stations as macrocell, microcell, and femtocell base stations and compared the power consumption for each base station. In Sharma et al. (2012), they presented the calculation of path loss, link budgets etc, which are in the part of wireless system designing. Arif et al. (2013), presented a hybrid deployment model of a LTE-femtocell for a residential suburban scenario. In their model they placed the femtocell in a fix grid to

Algorithm 1. To calculate the total power consumption

Input: For all the technology Coverage Range R, PCmax, Wireless Technology to be used, Path Loss Models, and Total Power to be returned Pel. Declare Variables Ptot, Nsector, Ntx and R.

Output: Total Power Consumption.

Step 1: Start.

Step 2: Ptot = Calculate Total Power (Pel, Tech).

Step 3: Ptot $\leftarrow$ Ptot*1000 // to convert Power in milliwatt.

Step 4: Calculate the coverage range (d, Tech, Model).

Step5: PCmax= Ptot / (PI*R*R) // calculate power consumption per covered area.

Step 6: Return PCmax;

Step 7: If Nsector $\leftarrow$ 1, Ntx $\leftarrow$ 1 // for SISO system.
// the values of Pproc, Pamp, Ptrans, Pairco, Pmicro, Prect are same for all the technologies.

Step 8: If Nsector $\leftarrow$ 3, Ntx $\leftarrow$ 2// For MIMO system
// the values of Pproc, Pamp, Ptrans, Pairco, Pmicro, Prect are different for all technologies.

Step 9:End if

// defined procedure Calculate Total Power (Pel, Tech) to calculate total power for all the technologies.

Step 10: Pcomm $\leftarrow$ Pmicro+Pairco

Step 11: Pel $\leftarrow$ Nsector *(Ntx * (Pamp + Ptrans) + Pproc + Prect)+Pcomm

Step 12: Return Pel;

Step 13: Stop

Algorithm 2. To Calculate total coverage range (R) and the path loss (PLmax)

Input: Total Coverage Range d, Wireless technology to be used Tech, Path loss Model, PathLoss to be returned PLmax, Wireless Technology Tech. Declare Variables path loss(PLmax), frequency(f), Height of base station(Hbs) and mobile station, distance(d0), speed of light(C), shadowing margin(S), Correction factor, mobile station antenna gain(Gms), Declare Link budget variables Fade margin (Fm), Interference margin(Im), Transmitter output power(Ptx), Transmitter antenna gain(Gtx), Receiver sensitivity(Rs), Effective isotropic radiated power(EIRP), Connector loss(Lc), Mobile station antenna gain(Gms), Body loss(Blm).

Output: Coverage Range and Path Loss(PLmax).

```

Step 1: Start
Step 2: PLmax ← CAL_PATHLOSS (PL, Tech)
Step 3: if Tech ← Technologies (MWiMAX, FWiMAX, UMTS, HSPA, LTE)
Step 4: Set frequency (f) for all technologies
Step 5: determine coverage range (d) for macrocell base station using following models.
if(Model = Erceg C)
Elseif(Model =HATA)
Elseif(Model = ERCEG A)
Elseif(Model=ERCEG-B)
Elseif(Model =COST-231)
Elseif(Model=W-I)
Elseif Model=ITU-R
Step 8: Return d;
Step 9: Set Link Budget Parameters Gms, Fm, Im, Ptx, Gtx and Rs for all technologies put value from link budget table
// calculate Receiver sensitivity, where KTB is thermal noise, NF is noise figure, SINR is signal to noise ratio and IM is interference margin.
Step 10: Rs ← KTB+NF+SINR+IM-3
// Calculate Effective Isotropic Radiated Power.
Step 11: EIRP ← Ptx - Lc + Gtx.
// define procedure CAL_PATHLOSS (PL,Tech) to Calculate Total PathLoss.
Step 12: PLmax ← EIRP - Rs - Blm - Im - Fm - LC + Gms.
Step 13: Return PLmax;
Step 14: Stop.

```

calculate path loss for each femtocell. In Ho et al. (2013), they proposed the Quality of Service (QoS) supporting video streaming system with minimum data service cost over heterogeneous wireless networks. They determined that, one of the critical problems is power consumption at the mobile devices due to limited battery capability.

Although some researchers calculated the total power consumption and coverage range for two or three technologies using only one or two PathLoss models, but this seems to be missing in the literature that the calculation and comparison of the power consumptions of most popular wireless technologies using dif-

ferent Path Loss models (Erceg Type A, B and C, HATA model, COST-231 model, WI model and ITU-R model) at one place. In this chapter, we will provide calculation and comparisons of the total power consumption of most popular wireless technologies using different Path Loss models with coverage Range and power consumption per covered area. E.g. Long Term Evolution (LTE), mobile WiMAX (Worldwide Interoperability for Microwave Access), Fixed WiMAX, UMTS (Universal mobile telecommunication standard), HSPA (High Speed Packet Access) for SISO system and LTE, Mobile WiMAX, HSPA for 2x2 MIMO system using different path loss models.

MAIN FOCUS

In this section we describe the methodology to calculate the total power consumption of the base stations and power consumption per covered area for different wireless technologies. For this we use following two algorithms: first for the calculation of total power consumption and second to calculate the path loss and coverage area.

Detailed Description to Calculate the Power Consumption of the Base Stations

Power Consumption of a Base Station

In a base station, we typically find several power consuming components. Fig.1, 2 and 3 gives an overview of these components. The area covered by one base station is called a cell. Each cell is further divided into a number of sectors. Each sector is covered by a sector antenna.

Power Consumption of a Macrocell Base Station

In macrocell base station block diagram as depicted in Figure 1 shows that a macrocell base station consists of six power-consuming components.

- **Rectifier:** Converts alternating current (AC) to direct current (DC), also known as the AC-DC converter.
- **Digital Signal Processing:** Is concerned with the conversion of the signal to a sequence of bits or symbols and the processing of these signals.
- **Transceiver:** Is responsible for transmitting and receiving the signals.
- **Power Amplifier:** Converts the DC input power into a significant radio-frequency (RF) signal.
- **Air Conditioning:** Regulates the temperature in the base station cabin.
- **Microwave Link:** Is responsible for the communication with the backhaul network

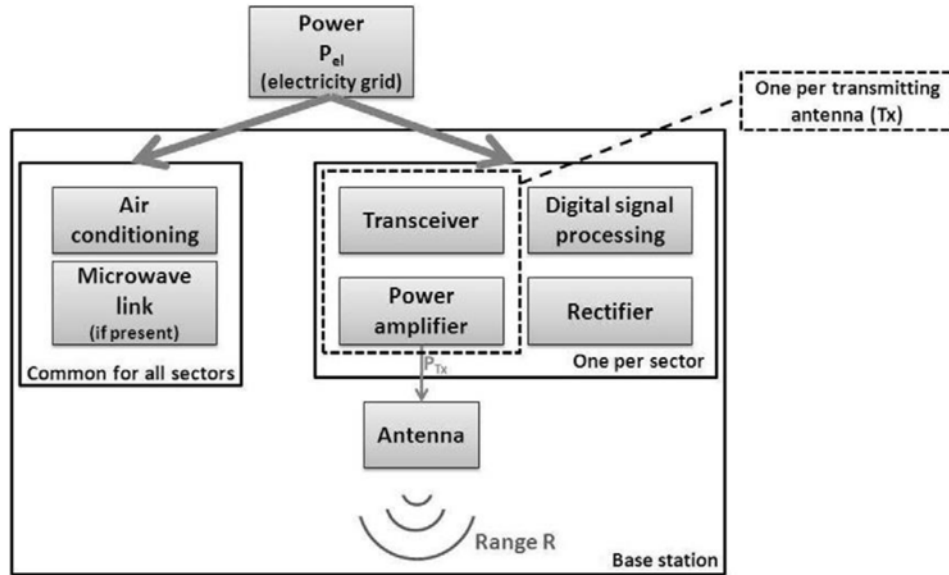
If we take the sum of all those components, we can determine the base station power consumption. However, in Figure 1, we see that some of the components are used multiple times. Each of the components of the base station has its own typical power consumption which is given in Table 1. The power consumption of the microwave link and the rectifier is assumed to be constant throughout time. The power consumption of the air conditioning is not influenced by the time but rather by the temperature inside and outside of the base station cabin. For the power amplifier, the maximum power consumption is indicated. The power consumption of the digital signal processing and transceiver are based on the confidential data retrieved from an operator. Important to remark is the results presented in this chapter depend on the values listed in Table 1.

Furthermore, the power consumption of the power amplifier is also dependent of the input power of the antenna P_{Tx} . The power consumption of the power amplifier is determined as follows:

$$P_{el/amp} = \frac{P_{Tx}}{n} \quad (1)$$

where n efficiency of the power amplifier is the ratio of RF output power $P_{out/amp}$ to the electrical input power $P_{el/amp}$ of the power amplifier :

Figure 1. Block diagram of the base station equipment (Deruyck et al. (2012))



$$n = \frac{P_{out} / \text{amp}}{P_{el} / \text{amp}} \quad (2)$$

Power Consumption of a Microcell Base Station

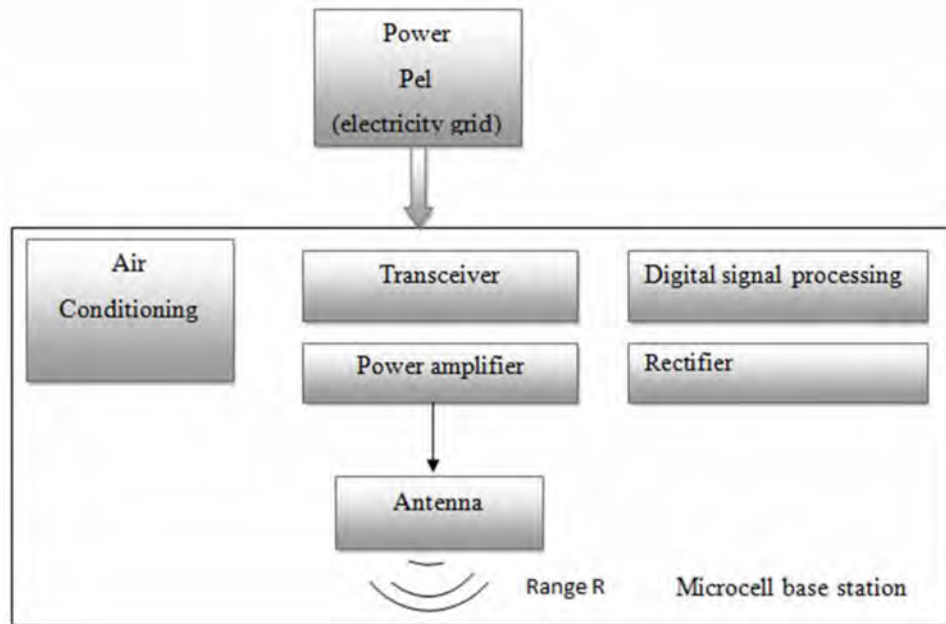
For the macrocell base station, mobile WiMAX and fixed WiMAX has a typical P_{Tx} of 35dBm while HSPA, UMTS and LTE have a P_{Tx} of 43 dBm. Once the power consumption of each component is known, the power consumption of macro cell base station P_{el}/macro (in Watt) can be calculated:

$$P_{el}/\text{macro} = n_{\text{sector}} * (n_{Tx} * (P_{el}/\text{amp} + P_{el}/\text{trans}) + P_{el}/\text{proc} + P_{el}/\text{rect}) + P_{el}/\text{micro} + P_{el}/\text{airco} \quad (3)$$

Where P_{el}/amp , P_{el}/trans , P_{el}/trans , P_{el}/proc , P_{el}/rect , P_{el}/micro , P_{el}/airco denotes the power consumptions (in Watt) of the power amplifier, the transceiver, the digital signal processing, the rectifier, the microwave link (if present), and the air conditioning respectively, N_{sector} is the no. of sectors, n_{Tx} is the number of transmitting antennas.

Figure 2 shows that the same components are used for a microcell base station as for a macrocell base station except the microwave. The microwave (backhauling) of a microcell base station is typically established through the overlaying macrocell base station. A microcell base station supports only one sector covered by one antenna. The power consumption of the components is shown in Table 1 and is similar to those for the macrocell base station except for the power amplifier and the air conditioning. As fewer devices are present in the base station's cabinet, a less powerful air conditioning with lower power consumption can be used. Note that the air conditioning is not always necessary in a microcell base station but, in this chapter we included air conditioning also. The power amplifier consumes also less than for the macrocell base station because the input power P_{Tx} of the antenna is 33dBm for all consider wireless technologies.

Figure 2. Block diagram of Microcell base station equipment (Deruyck et al. (2012))



Once the power consumption of each component is known, the power consumption of micro cell base station $P_{el}/micro$ (in Watt) can be calculated:

$$P_{el}/micro = P_{el}/amp + P_{el}/trans + P_{el}/proc + P_{el}/rect + P_{el}/airco \quad (4)$$

Where P_{el}/amp , $P_{el}/trans$, $P_{el}/trans$, $P_{el}/proc$, $P_{el}/rect$, $P_{el}/airco$ are denotes the power consumptions (in Watt) of the power amplifier, the transceiver, the digital signal processing, the rectifier and the air conditioning respectively.

Power Consumption of a Femtocell Base Station

The size of a femtocell base station is much smaller than that of a macrocell and microcell base station and is comparable to that of a Wi-Fi access point. Therefore, the power-consuming components are different from those of a macrocell and microcell base station as shown in Figure 3. The femtocell base station consists of a microprocessor, a FPGA (Field-Programmable Gate Array), a transceiver and a power amplifier. The microprocessor is

responsible for implementing and managing the standardized radio protocol stack and the baseband processing and also takes care of the communication with the backhaul network. The FPGA is responsible for a number of features such as data encryption, hardware authentication, etc. The power consumption of these components is also listed in Table 1.

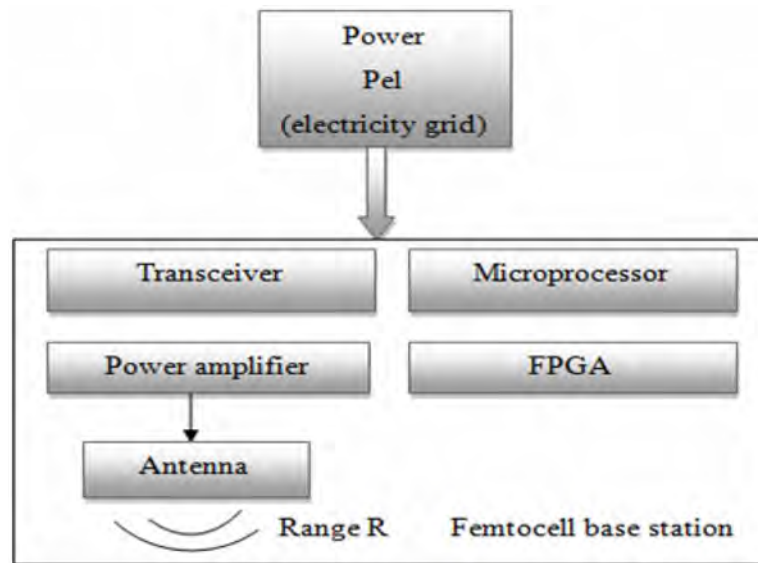
Once the power consumption of each component is known, the power consumption of femto cell base station $P_{el}/femto$ (in Watt) can be calculated:

$$P_{el}/femto = P_{el}/mp + P_{el}/FPGA + P_{el}/amp + P_{el}/trans \quad (5)$$

Where P_{el}/mp , $P_{el}/FPGA$, $P_{el}/trans$, P_{el}/amp are the power consumptions (in Watt) of the microprocessor, FPGA, transceiver, and power amplifier respectively.

We compare the power consumption with available data and measurements. For a 3-sector macrocell base station with one antenna per sector, P_{el} of 1673W is found with eq. (1) for UMTS, HSPA and LTE and P_{el} of 1279W for mobile WiMAX and fixed WiMAX. P_{el} for a

Figure 3. Block diagram of Femtocell base station equipment (Deruyck et al. (2012))



1-sector microcell base station with one antenna is 376W is found with eq. (2) for UMTS, HSPA, LTE, mobile WiMAX and fixed WiMAX. P_{el} for a 1-sector femtocell base station with one antenna is 12W is found with eq. (3) for UMTS, HSPA, LTE, mobile WiMAX and fixed WiMAX.

Power consumption values of the base station equipments are given as in Table 1. These values are received from datasheets of various

manufactures of network equipment's and from standards.

Calculate the Path Loss and Coverage Range

In calculation of total power consumption as described above, we know that how to calculate the power consumption P_{el} of the base station. It

Table 1. Power consumption of the macrocell microcell and femtocell base station components for the technologies considered (mobile WiMAX, fixed WiMAX, UMTS, HSPA and LTE)

Equipment	Parameter	Power Consumption (Base Station)		
		Macrocell	Microcell	Femtocell
Digital signal processing	$P_{el}/proc$, P_{el}/mp , $P_{el}/fpga$	100W	100W	7.9W (Microprocessor +FPGA)
Power amplifier (SISO)	N	12.8%	-	-
	P_{el}/amp (max)	24.7W($PTx=35$ dBm) 156W($PTx=43$ dBm)	16.6W ($PTx=33$ dBm)	2.4W ($PTx=33$ dBm)
Power amplifier (MIMO)	N	11.5%	-	-
	$P_{el}/amp(max)$	10.4W	-	-
Transceiver	$P_{el}/trans$	100W	100W	1.8W
Rectifier	$P_{el}/rect$	100W	100W	-
Air conditioning	$P_{el}/airco$	225W	60W	-
Microwave link	$P_{el}/micro$	80W	-	-

is dependent on the wireless range R covered by each base station. Now we calculate the Effective Isotropic Radiated Power (EIRP), Receiver Sensitivity (RS) and range. To calculate all these values a link budget has to be determined before.

Link Budget- a link budget is the most important entity for system-level design of wireless systems. It's a "budget" similar to an accountant's budget: the sum of gains and losses in various parts of the system has to result in satisfactory performance (typically output SNR or BER). The link budget includes all gains (and losses) from baseband input to baseband output (Deruyck et al. (2011)). Since values are gains/losses it is more convenient to add values expressed in dB instead of multiplying and dividing. The basic quantities included in a link budget are: transmitter output power, transmit antenna gain, path loss, receive antenna gain, fade margin and interference margin

Some typical link budget elements such as: transmitter feed line losses, transmit antenna gain, receiver antenna gain are also use to calculate the coverage range.

Path Loss: The path loss is the ratio of the transmitted power to the received power of the signal [4]. It is the reduction in power density of an electromagnetic wave. It includes all of the possible elements of loss associated with interactions between the propagating wave and any objects between the transmitting and receiving antennas. To determine the maximum allowable path loss PL_{max} , These parameters are retrieved from specifications and are typical values proposed by the operators themselves. The typical values allow making a fair comparison between the considered technologies. With the help of Path Loss we can determine the cell ranges.

After calculating the path loss for multiple technologies, it is very difficult to determine which one is the most energy efficient: one technology could have higher power consumption but also have higher ranges; another one could have a smaller range but also have lower power consumption. Therefore, the power consumption per covered

area is defined to quantify the power consumption and efficiency for different technologies:

$$PC_{area} = \frac{P_{el}}{AR^2} \quad (6)$$

Here P_{el} the power consumption of the entire base station (in watt) and R the covered range (in meters).

Simulation and Result Analysis

We used an open source simulator LTE-SIM, which is an open source framework for simulating different wireless networks. It supports C++ as the basic programming language. In this simulation, we place the base station outdoor in a suburban environment for macrocell and microcell and for femtocell base stations are placed in indoor environment. For macrocell with a transmitting antenna height is considered as 30 m and for microcell it is 6 m. Further we defined two configurations for the outdoor base stations: A basic reference configuration and extended configuration. In the basic reference configuration, one receiving and one transmitting antenna is considered, i.e., a SISO (Single input Single output) system. In the extended configuration both the base station and the receiver have multiple antennas, i.e., a MIMO (Multiple input Multiple output) system. The basic reference configuration is supported by all the technologies and for all macrocell, microcell and femtocell base stations, and the extended configuration is supported only by mobile WiMAX, HSPA and LTE for macrocell base station. The frequencies used for the link budget calculations of the different technologies are as following: 2.5 GHz for mobile WiMAX, 3.5GHz for fixed WiMAX, 2.1GHz for UMTS and HSPA, and 2.6 GHz for LTE. For the femtocell we used frequency 3.5GHz for all considered technologies. We used Erceg C as a reference model in our whole comparisons. The following Table 2 presents the parameters we apply in the simulation.

Table 2. Simulation parameters

Parameter	Value
Area type	Suburban, hilly area, indoor office
Number of sectors nsector	3 for macrocell and 1 for microcell and femtocell
Height of base station hbs	30 m for macrocell, 6 m for microcell
Height of mobile station hms	1.5 m for macrocell and microcell
Path loss model	Erceg type C, type A, type B, HATA, COST-231, W-I model, ITU-R P.1238
Shadowing margin	13.2dB for macrocell, 12.8dB for microcell

Coverage Range through Erceg C Path loss Model

Table 3 shows the simulation results using Erceg type C Path loss model for Suburban area. This table also describes the comparison of the wireless technologies (mobile WiMAX, HSPA and LTE) for SISO and 2*2 MIMO for macrocell base station.

Coverage Range Through HATA Path Loss Model

Here we compare the power consumption of wireless base station for mobile WiMAX, Fixed WiMAX, HSPA, UMTS and LTE access networks

for 1*1 SISO systems, for 2*2 MIMO system considered technologies is mobile WiMAX, HSPA and LTE. We used Path loss model HATA for calculating the range and area type is Sub urban area and used macrocell base station.

Coverage Range Through COST-231 Path Loss Model

Here we used COST-231 model for calculating the range, we compare the power consumption of wireless base station for mobile WiMAX, Fixed WiMAX, HSPA, UMTS and LTE access networks for 1*1 SISO system, for 2*2 MIMO system considered technologies is mobile WiMAX, HSPA and LTE. We used Suburban area and macrocell base station. COST-231 model is modified version of HATA model.

Coverage Ranges through Erceg Type A Path Loss Model

For calculating the range for hilly area terrain with heavy tree density we used Erceg type A path loss model. We compare the power consumption of wireless base station for mobile WiMAX, Fixed WiMAX, HSPA, UMTS and LTE access networks for 1*1 SISO system, for 2*2 MIMO system considered technologies are mobile WiMAX, HSPA and LTE. We used Hilly area and macro cell base station.

Table 3. Comparison of the wireless technologies for SISO and 2*2 MIMO

	Mobile WIMAX	HSPA	LTE
1*1 SISO			
Range [m]	301.307	221.129	347.637
Total Power Consumption[w]	1279	1673	1673
Total Power Consumption /Covered area[mW/m2]	4.484	10.890	4.406
2*2 MIMO			
Range [m]	434.629	318.083	498.104
Total Power Consumption [w]	1689	2669	2669
Total Power Consumption/Covered area[mW/m2]	2.846	8.396	3.424

*Table 4. Comparison of the wireless technologies for SISO and 2*2 MIMO*

	MWIMAX	FWIMAX	UMTS	HSPA	LTE
1*1 SISO					
Range [m]	302.714	321.781	435.139	219.624	354.456
Total Power Consumption[w]	1279	1279	1673	1673	1673
Total Power Consumption /Covered area[mW/m2]	4.442	3.931	2.312	11.040	4.238
2*2 MIMO					
Range [m]	464.495	-	-	335.898	539.639
Total Power Consumption [w]	1689	-	-	2669	2669
Total Power Consumption /Covered area[mW/m2]	2.4918	-	-	7.529	2.917

*Table 5. Comparison of the wireless technologies for SISO and 2*2 MIMO*

	MWIMAX	FWIMAX	UMTS	HSPA	LTE
1*1 SISO					
Range [m]	277.126	274.896	412.891	203.395	412.891
Total Power Consumption[w]	1279	1279	1673	1673	1673
Total Power Consumption /Covered area[mW/m2]	5.3011	5.387	3.123	12.262	5.139
2*2 MIMO					
Range [m]	425.231	-	-	318.724	490.057
Total Power Consumption [w]	1689	-	-	2669	2669
Total Power Consumption /Covered area[mW/m2]	2.973	-	-	8.363	3.537

*Table 6. Comparison of the wireless technologies for SISO and 2*2 MIMO*

	MWIMAX	FWIMAX	UMTS	HSPA	LTE
1*1 SISO					
Range [m]	272.406	302.769	345.148	208.865	307.997
Total Power Consumption[w]	1279	1279	1673	1673	1673
Total Power Consumption /Covered area[mW/m2]	5.486	4.441	4.470	12.207	5.613
2*2 MIMO					
Range [m]	373.093	-	-	285.378	419.415
Total Power Consumption [w]	1689	-	-	2669	2669
Total Power Consumption /Covered area[mW/m2]	3.862	-	-	10.431	4.829

Table 7. Comparison of the wireless technologies for SISO and 2*2 MIMO

	MWIMAX	FWIMAX	UMTS	HSPA	LTE
1*1 SISO					
Range [m]	299.914	336.743	388	224.165	343.12
Total Power Consumption[w]	1279	1279	1673	1673	1673
Total Power Consumption /Covered area[mW/m2]	4.526	3.590	3.524	10.597	4.523
2*2 MIMO					
Range [m]	423.362	-	-	315.602	481.302
Total Power Consumption [w]	1689	-	-	2669	2669
Total Power Consumption /Covered area[mW/m2]	2.995	-	-	8.529	3.667

Coverage Range through Erceg Type B Path Loss Model

For calculating the wireless range for hilly area terrain or flat terrain with heavy tree density we used Erceg type B path loss model. we compare the power consumption of wireless base station for mobile WiMAX, Fixed WiMAX, HSPA, UMTS and LTE access networks for 1*1 SISO system, for 2*2 MIMO system considered technologies is mobile WiMAX, HSPA and LTE.

Coverage Range through Walfish-Ikegami Model Path Loss

We used WI model for calculating the wireless range microcell base station, we compare the power consumption of wireless base station for mobile WiMAX, Fixed WiMAX, HSPA, UMTS and LTE access networks for 1*1 SISO system. We used Suburban area and microcell base station. In the simulation we used the following

data, i.e. number of sector is 1, and Transmitting antenna is SISO, considered area type suburban, base station is microcell and mobile station height 1.5 m, base station height 6 m, roof height is 2 m and number of floors is 3. For this microcell base station we calculate the total power consumption through equation (2).

Coverage Range through ITU-R P.1238 Path Loss Model

We used ITU-R model for calculating the wireless range in femtocell base station, we compare the power consumption of wireless base station for mobile WiMAX, Fixed WiMAX, HSPA, UMTS and LTE access networks for 1*1 SISO system. In the simulation we used the following data, i.e. number of sector is 1, and Transmitting antenna is SISO, considered area type indoor office, base station is femtocell and frequency is 3.5 GHz. For this femtocell base station we calculate the total power consumption through equation (3).

Table 8. Comparison of the wireless technologies for SISO

	Mwimax	Fwimax	UMTS	HSPA	LTE
1*1 SISO					
Range[m]	78.715	57.377	56.576	47.288	71.111
Total Power Consumption[w]	376	376	376	376	376
Total Power Consumption/Covered area[mW/m2]	19.316	36.354	37.391	53.522	23.667

Table 9. Comparison of the wireless technologies for SISO

	Mwimax	Fwimax	UMTS	HSPA	LTE
<i>1*1 SISO</i>					
Range[m]	58.720	54.152	36.33	30.365	54.877
Total Power Consumption[w]	12	12	12	12	12
Total Power Consumption/Covered area[mW/m2]	1.107	1.302	2.894	4.142	1.260

CURRENT AND FUTURE TRENDS

We considered only sub urban area while in the future the total power consumption can be calculated for rural and urban area and reader can also work on how to reduce the total power consumption for base stations?

CONCLUSION

This chapter described the total power consumption, coverage range and power consumption per covered area for the most popular wireless technologies and compared the power consumption of these technologies in the 5 MHz bandwidth for macrocell base station and area has considered as suburban type, if one input and one output (SISO) transmitting antenna is considered the result showed that UMTS is most energy efficient and its range is higher than other technologies and through HATA model UMTS gives higher range and less power consumption area. Fixed WiMAX is less energy efficient but its range is not high, LTE is more energy efficient than mobile WiMAX and its range are higher than mobile WiMAX, HSPA is less energy efficient. Furthermore the results of MIMO showed that Mobile WiMAX is most energy efficient than LTE and HSPA.

REFERENCES

Alshami, M., Arslan, T., Thompson, J., & Erdogan, A. T. (2011). Frequency analysis of path loss models on WiMAX. *Proceedings of the Computer Science and Electronic Engineering Conference*.

Arif, M., Yameen, I. M., & Matin, M. A. (2013). Femtocell Suburban Deployment in LTE Networks. *International Journal of Information and Electronics Engineering*, 3(2), 208–212.

Deruyck, M., Joseph, W., & Martens, L. (2014). Power consumption model for macrocell and microcell base stations. *Transactions on Emerging Telecommunications Technologies*, 25(3), 320–333. doi:10.1002/ett.2565

Deruyck, M., Tanghe, E., Joseph, W., & Martens, L. (2011). Modelling and optimization of power consumption in wireless access networks. *Computer Communications*, 34(17), 2036–2046. doi:10.1016/j.comcom.2011.03.008

Deruyck, M., Tanghe, E., Joseph, W., & Martens, L. (2011). Modelling the energy efficiency of microcell base stations. *Proceedings of the International Conference on Smart Grids, Green Communications and IT Energy-Aware Technologies (ENERGY-2011)*, 1-6.

Deruyck, M., Vereecken, W., Joseph, W., Lannoo, B., Pickavet, M., & Martens, L. (2012). Reducing the power consumption in wireless access networks: Overview and recommendations. *Progress in Electromagnetics Research*, 132, 255–274. doi:10.2528/PIER12061301

Ho, D., Park, Y., & Song, H. (2013). QoS-supporting video streaming system with minimum data service cost over heterogeneous wireless networks. *Journal of Visual Communication and Image Representation*, 24(8), 1293–1302. doi:10.1016/j.jvcir.2013.08.012

Jain, R. (2007). *WiMAX Forum, "Documentation, Technology Whitepapers"*. Available at WiMAX Forum.org: <http://www.wimaxforum.org/resources/documents>

Jimaa, S. (2011). LTE-A an Overview and Future Research Areas. *Proceedings of IEEE 7th international conference on wireless and mobile computing, networking and communications (WiMob)*, 395-399. doi:10.1109/WiMOB.2011.6085370

Sharma, P. K., & Singh, R. K. (2012). Cell coverage area and link budget calculations in GSM system. *International Journal of Modern Engineering Research*, 2, 170–176.

ADDITIONAL READING

Stallings, W. (2005). *Wireless Communications and Networks* (2nd ed.). Pearson Prentice Hall.

KEY TERMS AND DEFINITIONS

Effective Isotropic Radiated Power (EIRP):

It is the output power when a signal is concentrated into a smaller area by the antenna. It radiates power equally in all directions.

Enhanced Data Rates for Global Evolution

(EDGE): The enhanced data rate for global evolution (EDGE) is the new radio interface technology to boost network capacity and user data rates for GSM/GPRS networks. It has been standardized by ETSI and also in the United States as part of the IS-136 standards.

HSPA (High Speed Packet Access): The standard mainly designed to support high speed data rate in the uplink and downlink. HSPA falls under categories viz. HSDPA, HSUPA and HSPA+.

LTE Technology: A new network architecture, which is developed by the 3rd Generation Partnership Project (3GPP) and support radio access network (RAN). LTE constitutes the latest step towards the 4th generation (4G) of radio technologies designed to increase the capacity and speed of mobile communications.

Path Loss: The ratio of the transmitted power to the received power of the signal. It is the reduction in power density of an electromagnetic wave.

UMTS (Universal Mobile Telecommunication System): Developed by ETSI (European Telecommunications Standardization Institute) and operates in the 2.1 GHz band. UMTS uses W-CDMA (Wideband Code Division Multiple Access) as multiple access technique and has a maximum bit rate of only 3 Mbps. UMTS has been specified as an integrated solution for mobile voice and data.

Resource Management for Multimedia Services in Long Term Evaluation Networks

Vinod Kumar Mishra

B. T. Kumaon Institute of Technology, India

Tanuja Pathak

B. T. Kumaon Institute of Technology, India

INTRODUCTION

Wireless multimedia networks are becoming progressively popular, which provide the mobile phone user to accommodate of access to information and services at any time, any place and in any configuration with user desire and due to demand of high speed internet and growing wireless multimedia communication systems, future communication systems are expected to transmit large amount of multimedia traffic such as video, voice and text with a variety of Quality of Service (QoS). The important issue for future communication systems and upto certain extend in current communication system is the time and frequency being used for scarce resources and the requirement of QoS for multimedia applications which are sensitive in nature. Long Term Evolution (LTE) is one of the fastest growing technology which supports a variety of applications like video conferencing, video streaming, VoIP, file transfer, web browsing etc. In order to support multiple applications, radio resource management (RRM) procedure is one of the key design that has a role in improving the system performance. Long Term Evolution (LTE) establishes a significant milestone in the evolution of 3G systems towards the fourth generation (4G) technologies. The expanding utilization of complex multimedia services requires an advanced resource management mechanism in particular in wireless network environment. The mechanism that is responsible

for managing and controlling radio resources are known as resource management. The admission control is one of the resource management mechanisms, and task of admission control is to admit or reject the establishment request for the new session.

The performance targets promised by LTE, make it an ideal solution to accommodate the ever increasing demand for wireless broadband. LTE's promised performance targets were made possible due to improvements such as implied system access architecture and a fully IP-based platform. LTE has also great enhancements in its enabling radio technologies by introducing orthogonal frequency division multiplexing (OFDM) and advanced antenna technologies (Astély et al., 2009). In addition, LTE capabilities are further improved with enhanced Quality of Service (QoS) support for multiple data services, such as voice and other multimedia applications.

The main challenges in 3G, 4G and in next generation cellular networks are maintaining the quality of service (QoS) requirements for each class of the multi-service traffic in the presence of the limited radio resources (Ghaderi & Bouataba, 2006). The components that are in-charge of over-seeing and controlling the radio resources are known as radio resource management (RRM). The procurement of wireless multimedia services to mobile users with a certain quality of service, force stringent necessities on the configuration of radio resource systems. Admission control for

wireless system has been generally concentrated on under distinctive system architectures and systems heads approaches.

In this chapter, we introduce the fundamental idea of the admission control design and quality of service (QoS) procurement in wireless and mobile network. Provide, in brief, the call admission control (CAC) mechanisms, under a radio resource management (RRM) to support multimedia applications in the next generation wireless cellular networks with different QoS requirements prerequisites e.g. transfer speed (bandwidth), delay/jitter, and priority. CAC schemes enable the system to provide QoS for the new and also for the existing calls. Resource reservation scheme, such as Guard Channel, is utilized to hold the resources for certain high priority calls. On the other hand, network is required to exploit the resource sharing among traffic to accomplish better channel utilization. Acquiring a right harmony between the two restricting criteria is again a huge challenge. This schemes cover RRM outline for the channel-based wireless system, such as the time division multiple access (TDMA) and the frequency division multiple access (FDMA) systems, as well as the interference-based code division multiple access (CDMA) system. We focus on the implementation of the admission control algorithm in the long term evolution (LTE) network which enhancing session establishment success. For this we use the concept of service degradation (also known as media degradation path (MDP)) for handling multimedia multiclass services in admission control. The next generation network will be supported via various service classes with different QoS. Each service class call will treated differently according to the set criteria and adopted operating principle for the admission procedure. Since multimedia service may contain several media flows and user preference regarding flow importance may also vary so “user” and “service” related knowledge can be used at session initiation to specify alternative service configuration. We apply this knowledge in the context of admission control in non-MDP algorithm. In this approach

the algorithm allows the sessions to be admitted with lower quality configurations, in cases when there are not enough resources to admitting optimal (highest quality) configurations. Algorithm is implementing on LTE Simulator (LTE-Sim) while previously this is implemented in other simulators. The result of simulation of this approach shows that by using this approach in existing algorithm the session admission probability will increase over non-MDP algorithm.

BACKGROUND

Numerous previous research works have been published that allow higher priority for handover calls over new calls. Most of these proposed schemes are in light of the thought of “guard band,” where various channels are reserved for the restrictive utilization of handover calls. Although schemes based on guard bands are simple and capable of reducing the dropping probability, these schemes also result in reduced bandwidth utilization. To improve resource utilization of the guard channel algorithm, Ramjee et al. (1997) proposed fractional guard channel algorithm. The fractional guard channel algorithm decides an acknowledgment or a dismissal of new calls with choice probabilistically changed as the quantity of occupied channels. In case that channel is sufficient available, handoff calls are accepted, but new calls can be rejected. In this way, in the fractional guard channel algorithm, dropping probability of handoff calls may be smaller than that of new calls. Ghaderi and Boutaba (2006) also introduced guard channel based schemes, in their scheme guard channel algorithm reserves some channels among the total number of channels for handoff and calls admission control procedure. Guard channel algorithm is straightforward and its implementation is simple. In addition, the guard channel algorithm may decrease the use of physical resources as the quantity of holding channels increments. Wang et al. (2003), Wu (2005) and Bae et al. (2009) have introduced Quality of service related admission

control schemes. Wang et al. (2003) introduced a probability-based adaptive algorithm for call admission control in wireless networks, and do a little examination. This is based on probability to control the guard channels as indicated by controlled QoS measure. Wu (2005) gave an overview of the issues and strategies in QoS provisioning for wireless networks, specifically, the study in five sub-areas, network service models, traffic specification, packet scheduling for wireless transmission, call admission control in wireless networks, and wireless channel characterization while Bae et al. (2009) have proposed resource estimated CAC algorithm. The quantities of PRBs which are allocated to the requested call are estimated in this algorithm. The type of the requested service and current MCS level of UE (user equipments) are needed for deciding the quantity of required PRB (physical resource block). What's more, the resource-estimated CAC algorithm calculates available PRBs in view of the PRB utilization of on-going call measured by eNodeB.

Many researchers have pointed out that call admission schemes enormously impact the network performances, and in the meantime a straightforward fixed call admission scheme will bring about low system efficiency without a doubt. Qian et al. (2009) have proposed a novel RAC scheme for multiclass services in LTE systems which go for improving system capacity while in the meantime ensuring QoS requirements of every service class. To upgrade the system capacity, they adopt a service degradation scheme if there should arise an occurrence of resource limitations in our RAC scheme. Shen and Li (2009) have developed the admission control algorithm for wireless mesh network (WMNs) in light of the estimation of the accessible bandwidth in MAC layer, which support the QoS in WMN. The Admission control algorithm treats various traffic types in different ways. It gives the real time flows with admission decisions at the MAC layer and the non real-time flows with sending rates. Kim et al. (2009) proposed Semi-Adaptive Call Admission Control (SA-CAC) approach which

joins resource allocation and CAC scheme. As per his work the bandwidth of an adaptive real-time call can just change during setup or handover occurrence. To maximize the network revenue and keep up predefined QoS requirements, the QoS limitations are classified as two categories: long period constraints and instantaneous constraints.

Antonopoulos and Verikoukis (2010) have introduced a connection admission control (CAC) mechanism for IEEE 802.16 broadband wireless access standard. The issue of "occupied hour" in communications traffic variation is considered in this algorithm which gives the basis for the bandwidth reservation concept. WiMax infrastructure is mostly considered in this with the goal that it can be adjusted for long term evolution (LTE). This mechanism provides higher priority to VoIP calls contrasted with different types of traffic in the network.

Khanjari et al. (2010), Wang and Qiu (2013) and Chwdhury et al. (2013) introduced admission control scheme which follows the degradation concept, Khanjari et al. (2010) proposed an adaptive Call Admission Control (CAC) scheme for multi-class service wireless cellular networks. The proposed scheme utilizes Fairness policies for resource reallocation and allocation which consider the mechanism of call bandwidth degradation and bandwidth upgradation by utilizing the complete sharing approach of the accessible bandwidth among every traffic class, the scheme additionally applies call preemption as indicated by the class of call priorities and call reactivation at whatever point the traffic conditions are allowed. Wang and Qiu (2013) introduced a multi-service call admission control technique for LTE femtocell network. This strategy considers multiple classes of services with different QoS requirement and different bandwidth requirement. Wang considered the Guard Channel as a dynamic proportion rather than a fixed amount of Physical Resource Blocks (PRBs). Chwdhury et al. (2013) proposed a scheme of degradation if there should be an occurrence of scarce capacity, i.e. at the point when the accessible bandwidth is low, the

scheme releases some bandwidth from effectively accepted non real-time calls, as to suit new and for handover calls. More bandwidth is released to support handover calls bring over new calls.

Kim et al. (2013) presented a mobility aware call admission control (MA-CAC) with a handoff queue. The vehicle mobility governs the admission control policies. To keep up high channel use, MA-CAC introduced stop and moving stage: at the stop stage a handoff priority scheme with the guard channel concept is utilized and then again at the moving stage no guard channel reserved for handoff users. Williams and Asuquo (2014) introduced an effective model that can decrease call blocking probability and enhance the quality of service (QoS) gave to mobile users. The proposed model computes the steady state probability and resource occupancy distribution, traffic distribution, intra-cell and inter-cell interferences from mobile users. Recently Badri et al. (2015), Sepänen et al. (2014), Ivesic et al. (2013) and Ivesic et al. (2014) gave admission control mechanism for adoptive multimedia services.

On the basis of literature review, we conclude that the researchers addressed particular schemes, models and ways to deal with how to manage and keep up the quality of services (QoS) requirements of different multimedia services. A few schemes have connected to keep up a bandwidth reservation for handover call and new call, call blocking probability, handover call dropping probability and channel utilization. We observed that there is a need of the particular way to admit more session without dropping the call and maintain the quality of service and user preference in view of the current era of services. The MDP based admission control algorithm used to accomplish the objective of upgrade the session establishment executing on LTE Simulator.

MAIN FOCUS

Amid the most recent decades the wireless communication network clients have been quickly

expanded alongside their interest for new sight and sound services. The requirement for higher speed communication is as opposed to the rare spectrum resource distributed for wireless system in universal international organization. In this way, a fit resource management (RRM) is crucial to designate the current system asset among uniting clients, taking into examination their needs and separate needs us to give them the obliged QoS. RRM usefulness plan to enhance framework execution by boost the general framework limit in the wireless system protecting in the meantime QoS attributes of versatile clients. The admission control is one of the radio resource management mechanisms. The key idea of admission control is to ensure the QoS of individual connection by appropriately managing the network resources. The admission decision is based on not only the available network resources, but also the QoS requirements for requesting and ongoing users. Hence decision should be taken to consider multiple parameters such as the network characteristics, the service type, user mobility and the network condition. Long Term Evolution technology displays an exceptionally difficult multiuser issue: Several user equipment's (UEs) in the same geographic region require high data rates in a limited bandwidth with low latency. Numerous access techniques permit UEs to share the accessible bandwidth by allocating to each UE a fraction of the total system resources. The strong motivation past the Admission control schemes for admitting the session is the improvement of system performance by increasing the spectral efficiency of the wireless interface and thus improving the system capacity. A Call Admission Control (CAC) is an algorithm of decision-making that provides QoS in the network by restricting access to the network resources (Ghaderi & Boutaba, 2006).

According to the requested call type, CAC decides to accept or block the new call according to the network resources availability. At the point, when there are not sufficient resources to guarantee the call quality or to keep the active calls QoS (services already accepted-established),

CAC blocks the new call. Otherwise, the call is accepted.

The Quality of Service (QoS) and Quality of Experience (QoE) aware admission control (AC) determines whether a new radio bearer should be granted or denied access based on if required QoS of the new radio bearer will be fulfilled while guaranteeing the required QoS of the in progress sessions. Ivesic et al. (2013) proposed a degradation concept of knowledge (known as MDP) for maximizing the number of admitted session and to optimize the system capacity, they adopt a service degradation scheme in case of resource limitations. A multimedia service joins two or more media components, such as audio, video, text, data, 2D/3D graphics, Within a single session, the media components may be consolidated in distinctive ways, and need not be dynamic in the meantime their number, and in addition properties and prerequisites regarding network resources normally change. Multimedia services as a rule comprise of different media flows, the number and properties of which may change over the span of a continuous session. At the point when a session is initiated, AC mechanisms are invoked to make an admission decision based on operator policy, resource availability, service type, and user priority. Providing QoS insurance to different applications is a critical target in outlining the next-generation wireless networks. Different applications can have extremely different QoS necessities as far as data rates, delay bounds, and delay bound violation probabilities, among others chooses whether a connection request should be accepted or rejected, in view of the asked for QoS and the network status.

QoS/QoE-Aware Admission Control (MDP Based)

The collection of multiple feasible configurations of giving service concept is applied to the admission control problem. QoS Class Identifier (QCI) ensures bearer traffic is allocated appropriate QoS. Different bearer traffic requires different

QoS and therefore different QCI values. QCI classifies the different types of bearers into different classes, with each class having appropriate QoS parameters for the traffic type. Examples of the QoS parameters include Guaranteed Bit Rate (GBR) or non-Guaranteed Bit Rate (non-GBR), Priority Handling, Packet Delay Budget and Packet Error Loss rate. The QCI is a scalar that is used within the access network (namely the eNodeB) as a reference to node specific parameters that control packet forwarding treatment (e.g., scheduling weight, admission thresholds and link-layer protocol configuration). Here we assume that three service categories, which are referred to as low, medium, high and divide the available resources into zones pertaining to these categories. High class belongs to real time service flow with constant bit rate (CBR) such as voice and video phone service, medium service class is belong to real time service flow with variable bit rate (VBR) such as interactive multimedia and video streaming, and low class is belong to Non –real flow with VBR or UBR (Unspecified Bit Rate) such as email, file transfer service. The first zone is available to all sessions, the second to low service handoff sessions and all medium and high service sessions, the third to medium handoff sessions and all high service sessions and the fourth to high service handoff sessions only, Besides the zones, we also define the zone critical border B_0 as a limit that, when crossed by resource consumption, instructs the admission control algorithm to admit sessions with suboptimal configurations, For the session i the end of the zone that it belongs to is denoted by T_i and the area between the limits B_0 and T_i is defined as a critical area. The resource consumption is spans from left to right (from T_0 to F , w here F marks the end of the area with occupied resources and the beginning of the area with free resources), so that free resources are on the right. For each session, the zone critical border is calculated as a predefined percentage of the resources from the regarding zone, and the limits for different suboptimal configurations are created by dividing

the critical area into n equal intervals, where n is the number of configurations.

Pseudo Code for MDP Based Admission Control Algorithm (Ivesic et al., 2013)

```

For all the flows do
{
Determine the zone and set critical
border limit of the zone
Admission Control (flow, Zone limit,
occupied Resources, Bandwidth)
{
Initialize  $n \leftarrow$  Size of flow
Initialize the critical border limit
of QoS configuration B0 in available
Zone
B0.resize (n, zone)
Combine the available resources to
configure in each zone
While (!config.empty())
{
Zone_t zone =config.begin() -->first
For (availResource_iter = availRe-
sources.begin(); availResource_iter
!= availResources.end(); availRe-
source_iter++)
{
Zone_t QCI =qci;
If (QCI< B0 [v] && available resource
→ rsc <qci)
Select optimal configuration
Else if (QCI< T1) then
Select alternative config
}
}
Assign session
Else if (available resource == NULL)
Reject session
Return;
}
}

```

To evaluate the effectiveness of the Admission Control Algorithm we use the LTE Simulator. We implement this admission control algorithm in LTE-Sim, which is an open-source framework for simulating LTE networks developed by Giuseppe Piro, Luigi Alfredo Grieco, Gennaro Boggia, Francesco Capozzi, and Pietro Camarda. LTE Sim tool provides discrete event simulation of user defined Network. LTE-Sim supports several aspects of LTE networks from the application layer down to the physical layer, such as single-cell and multi-cell environments (Pietro et al. 2011), QoS management, multi-users environment, user mobility, CQI feedback, handover procedures, and frequency reuse techniques. Fundamental network nodes like UE, eNB, MME and S-GW are modeled in LTE-Sim, with the supports of the trace-based, Voice over IP (VoIP), Constant Bit Rate (CBR), and infinite-buffer traffic generators at the application layer and the management of data radio bearers.

For comparison we used a non-MDP based admission control algorithm which does not consider the concept of alternative configuration according to services and if occupied resources in the optimal configuration zone are equal to the limit of zone then reject the session.

Configuration

For the simulation, a cell of 10 MHz bandwidth and 2 GHz carrier frequency are modeled. The cell is assumed to consist of 50 RBs and K users and the users are simulated within 1 km radius.

The serving eNodeB has a fixed location at the centre of the cell and controls all the available RBs. These RBs are to be shared by all users within the cell. At each TTI(Transmit time interval), RBs can be assigned to a user. While other simulation parameters are simulation duration: 10.2 s, number of cells: 1, cell radius: 1 km, bandwidth: 10 MHz, scheduler: EXP rule, frame structure: frequency-division duplex, user speed: 3 km/h. In case of session arrival time, the admission

control module is invoked and initial resources are allocated if the session is admitted. The Zone limits for zone 1, zone 2 and zone 3 have been set to 85%, 90% and 95% of the available resources respectively. The limit of B_o (critical border limit of QoE configuration) has been set to 65% of the available resources within the given zone.

Result and Analysis

The simulation result indicated that, in the Case of applying non MDP based admission control the number of active sessions are near by 45%, and in the case of applying MDP based admission control the numbers of active sessions are near by 54%. The active bearers are close to 75% in non MDP based admission control while in MDP based admission control the number of active bearers are near by 90%. Thus in case of QoS /QoE admission control algorithm the admitted session and active bearers enhanced in comparison to non MDP admission control algorithm.

CURRENT AND FUTURE TRENDS

With the quick development of wireless networks and incredible achievement of Internet multimedia applications, the future wireless networks are imagined that they will not simply give a higher data rate to the mobile users, additionally to serve different mobile terminals ranging from just-voice phone (dumb phone) to smart phone, and to support heterogeneous applications with different quality of service (QoS) requirements in terms of data rates, delay and packet loss. Nevertheless, providing QoS or QoE in wireless networks is a challenging task while following the transmission condition of wireless channel which is dynamically changes over time.

We take note of that a present barrier of the LTE-Sim tools is an absence of backing for mapping service flow to numerous QCIs. We will likewise run longer simulations in LTE-Sim and

test the calculations in distinctive network conditions. The proposed admission control scheme can be simulated for a wireless cellular network of many cells. Wherein, an extended representation of physical mobility can be applied using mobility prediction techniques.

CONCLUSION

In this chapter, we provide an overview of the call admission control (CAC) mechanisms under a radio resource management (RRM) to support multimedia applications in the next generation wireless cellular networks with different QoS requirements. Implemented and analyzed a multi-service QoS/QoE admission control scheme for LTE network on LTE- simulator. This scheme considered various classes of services with diverse QoS necessity. The admission control system works on a flow premise as opposed to on classes, when another flow is entering the eNodeB the management system must choose whether there is sufficient ability to completely sustain the flow. The simulation result indicates the normal increment in admitting session and the quantity of bearers is marginally diminished in a few occasions. The detail discussion of recent significant literature and some frontier of this area has also included in this chapter.

REFERENCES

- Antonopoulos, A., & Verikoukis, C. (2010). Traffic-aware connection admission control scheme for broadband mobile systems. *IEEE Communications Letters*, 14(8), 719–721. doi:10.1109/LCOMM.2010.08.100652
- Astély, D., Dahlman, E., Furuskär, A., Jading, Y., Lindström, M., & Parkvall, S. (2009). LTE: The evolution of mobile broadband. *IEEE Communications Magazine*, 47(4), 44–51. doi:10.1109/MCOM.2009.4907406

- Badri, T. F. Z., Saadane, R., Mbarki, S., & Wahbi, M. (2015). Call Admission Control Scheme and Handover Management in LTE Femtocell-Macrocell Integrated Networks. *Computer and Information Science*, 8(1), 135–142. doi:10.5539/cis.v8n1p135
- Bae, S. J., Lee, J. J., Choi, B.-G., Kwon, S., & Chung, M. Y. (2009). A resource-estimated call admission control algorithm in 3GPP LTE system. In *Computational Science and Its Applications-ICCSA, (LNCS)* (Vol. 5593, pp. 250–260). Berlin, Germany: Springer. doi:10.1007/978-3-642-02457-3_22
- Chowdhury, M. Z., Jang, Y. M. & Haas Z. J. (2013). Call Admission Control Based on Adaptive Bandwidth Allocation for Wireless Networks. *IEEE/KICS Journal of Communications and Networks*, 15(1), 15-24.
- Ghaderi, M. & Boutaba, R. (2006). Call admission control in mobile cellular networks: a comprehensive survey. *Wireless Communications and Mobile Computing*, 6(1), 69-93.
- Ivesic, K., Lea, S.-K., & Matijasevic, M. (2013). Admission control for adaptive multimedia service based on user and service related knowledge. In *Proceeding of EUROCON*. Zagreb, Croatia: IEEE.
- Ivesic, K., Lea, S.-K., & Matijasevic, M. (2014). Cross-layer QoE-driven admission control and resource allocation for adaptive multimedia services in LTE. *Journal of Network and Computer Applications*, 46, 336–351. doi:10.1016/j.jnca.2014.09.010
- Khanjari, S. A., Arafeh, B., Day, K., & Alzeidi, N. (2010). An adaptive call admission control scheme with complete bandwidth sharing for multi-class traffic in wireless networks. *proceeding of International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, 265-269.
- Kim, T.-S., Yang, Y., Hou, J. C., & Krishnamurthy, S. V. (2009). Joint resource allocation and admission control in wireless mesh networks. *Proceeding of 7th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOPT)*, 1-10.
- Kim, Y., Ko, H., Pack, S., Lee, W., & Shen, X. (2013). Mobility-aware call admission control algorithm with handoff queue in mobile hotspots. *IEEE Transactions on Vehicular Technology*, 62(8), 3903–3912. doi:10.1109/TVT.2013.2245928
- Pietro, G., Grieco, L. A., Boggia, G., Capozzi, F., & Camarda, P. (2011). Simulating LTE cellular systems: An open-source framework. *IEEE Transactions on Vehicular Technology*, 60(2), 498–513. doi:10.1109/TVT.2010.2091660
- Qian, M., Huang, Y., Shi, J., Yuan, Y., Tian, L., & Dutkiewicz, E. (2009). A novel radio admission control scheme for multiclass services in LTE systems. *Proceedings of 7th IEEE Global Telecommunications Conference (GLOBECOM)*, 1-6. doi:10.1109/GLOCOM.2009.5426190
- Ramjee, R., Towsley, D., & Nagarajan, R. (1997). On optimal call admission control in cellular networks. *Wireless Networks*, 3(1), 29–41. doi:10.1023/A:1019172226345
- Seppänen, J., Varela, M., & Sgora, A. (2014). An autonomous QoE-driven network management framework. *Journal of Visual Communication and Image Representation*, 25(3), 565–577. doi:10.1016/j.jvcir.2013.11.010
- Shen, Q., & Li, P. (2009). Admission control based on available bandwidth estimation for wireless mesh networks. *IEEE Transactions on Vehicular Technology*, 58(5), 2519–2528. doi:10.1109/TVT.2008.2006680
- Wang, J., & Qiu, Y. (2013). A new call admission control strategy for LTE femtocell networks. *Proceeding of 2nd International Conference on Advances in Computer Science and Engineering*, 334-338.

Wang, X.-P., Zheng, J.-L., Zeng, W., & Zhang, G.-D. (2003). A probability-based adaptive algorithm for call admission control in wireless network. *Proceedings of International Conference on Computer Networks and Mobile Computing (ICCNMC 2003)*, 197-204.

Williams, E. E., & Asuquo, D. E. (2014). An Efficient Model For Reducing Soft Blocking Probability In Wireless Cellular Networks. *International Journal of Wireless & Mobile Networks*, 6(4), 85–99. doi:10.5121/ijwmn.2014.6407

Wu, D. (2005). QoS provisioning in wireless networks. *Wireless Communications and Mobile Computing*, 5(8), 957-970.

ADDITIONAL READING

Kandolcar, P., Das, M. K., & Mukherjee, S. (2013). Radio Resource Management – Radio Admission Control and Bearer Control. White Paper, T.C.S., India.

Tsiropoulos, G. I., Stratogiannis, D. G., & Tsiropoulou, E. E. (2010). Call Admission Control in Mobile and Wireless Networks. Mobile and Wireless Communications: Network layer and Circuit Level Design, INTECH Open Access Publisher, Croatia, E.U.

Zander, J. (1997). Radio resource management in future wireless networks: Requirements and limitations. *IEEE Communications Magazine*, 35(8), 30–36. doi:10.1109/35.606024

KEY TERMS AND DEFINITIONS

1G Technology: 1G refers to the first generation of wireless telephone technology; mobile telecommunications which was first introduced in 1980s and completed in early 1990s. It is based on analog system. It's Speed up to 2.4 kbps and Allows user to make voice calls in 1 country. AMPS

(Advance Mobile Phone System) were launched by the US & it was the 1G mobile systems.

2G Technology: 2G technology refers to the 2nd generation which is based on GSM. It was launched in Finland in the year 1991. 2G network use digital signals. Its data speed was up to 64kbps. Features Included in 2G technology is: It enables services such as text messages, picture messages and MMS (multimedia message). It provides better quality and capacity. 2G are the handsets we are using, with 2.5G having more capabilities.

3G Technology: 3G technology refer to third generation which was introduced in year 2000s. Data Transmission speed increased from 144kbps-2Mbps. It provide Superior voice quality and Good clarity in video conference E-mail, PDA, information surfing, on-line shopping / banking, games, etc. This was followed by 3.5G, 3G+ or turbo 3G enhancements based on the high-speed packet access (HSPA) family, allowing UMTS networks to have higher data transfer speeds and capacity.

4G Technology: The next generations of wireless technology that promises higher data rates and expanded multimedia services. Developed in 2010. Its Speed up to 100 Mbps. Provide any kind of service at any time as per user requirements, anywhere.

Radio Resource Management (RRM): The mechanisms that are responsible for managing and controlling radio resources are known as radio resource management (RRM).

Quality Of Service (QoS): QoS is characterized as “totality of characteristics of a telecommunications service that bear on its capacity to fulfill expressed and inferred needs of the user of the service.”

Quality of Experience (QoE): QoE is defined as “the general agreeableness of an application or service, as perceived subjectively by the end-user.

Resource Block (RB): A Resource Block (RB) is a time- and frequency resource that occupies 12 subcarriers ($12 \times 15 \text{ kHz} = 180 \text{ kHz}$) and one slot = 0.5 ms).

SMS & Civil Unrest

M

Innocent Chilwa

Covenant University OTA, Nigeria

INTRODUCTION

According to Epstein (e.d.) (2015), “the whole world is texting...” in virtually all communication contexts of human endeavor. And studies show that a high number of telecommunication users rely on the Short Message Service (SMS) or texting via Smartphone to send and receive messages on different topics related to the society (Ahn, 2011); politics (Shirky, 2011); education (Aziz et al, 2013); business (Torres & Conaway, 2014); and religion (Bell, 2006; Campbell, 2005; 2006a; Chilwa & Uba, 2015), among others. According to *Pew Research Center*, over 6 billion messages are sent everyday in the United States (*Forester.com*) and about 8.3 trillion text messages were sent globally in 2015 alone, which is almost 23 billion messages per day or almost 16 million messages per minute (*Portio Research*). On the average, 8 trillion text messages are sent every year world-wide (*Bloomberg*). With this amazing prevalence of texting in our daily living, it is clear that SMS plays a significant role in influencing behavior, which includes initiating; implementing and championing civil unrests (see Chilwa, 2016), which unfortunately have become part of our everyday experience.

This entry is an updated version of my previous paper published in the *Encyclopedia of Mobile Phone Behavior*, (vol.2) entitled “Text messaging in social protests,” published by IGI Global (pages 1024-1031, 2015), edited by Zheng Yan. The current version contains new information on texting behaviors of protesters at different socio-cultural and economic contexts. Hence, the literature section has been expanded to include latest research findings.

“Civil unrest” in the current study is broadly used to include all forms of civil disorder, especially those associated with or organized by social groups, trade unions, civil rights groups, occupy movements, or university students. This will include mass protests (violent or “peaceful”); social or political unrests, street demonstrations; sit-ins and occupy protests, or industrial actions and strikes. In “text messaging in social protests,” (2015), I highlighted the fact that protest behaviors vary depending on the social and political contexts, as well the motives and the degree of participation. Hence, there is a peaceful protest that involves mass protesters in a protest march. In this case, protesters generally carry placards with protest messages. Protesters also sometimes wear T-shirts with protest messages inscribed on them (see Chilwa & Ajiboye, 2016). Sometimes, some peaceful protests have turned violent, when protesters began to engage security agents in direct confrontations through open fights and other criminal activities such as vandalism of public property, looting and raping. In this case what started as “peace” protests and rallies have ended up resulting in loss of lives and property. In most cases, security agents have had to resort to the use of tear gas or water canon to disperse protesters. An example was the Ferguson unrest that involved peaceful protests and riots that began a day after the shooting of a black American teenager (Michael Brown) by a white police officer Darren Wilson on August 9, 2014 in Ferguson, Missouri. The riots led to the imposition of curfews in the area and the deployment of the riot squad to maintain order. The protests continued until November 24, 2014 after a grand jury did not indict police officer Wilson.

DOI: 10.4018/978-1-5225-2255-3.ch546

As highlighted above, protest behaviors also include industrial actions involving labour/trade unions, where members deliberately stay away from work. In some peculiar circumstances, hunger strikes have also been engaged by individual protesters or activists. Extreme cases have been the deliberate self-immolations of activists in protest of institutional injustice. An example is that of Mohamed Bouazizi, the Tunisian street vendor, who set himself ablaze on December 10, 2010 in protest of the confiscation of his goods, and constant harassment and humiliation by municipal council officials. His death triggered off the Tunisian revolution. Collective actions and social movements participating at the same time have been caused by the need to achieve change, or “revolution” that is believed to be desperately needed either gradually or immediately or spontaneously (see Chilwa, 2012). In recent times mass protests have enabled protesters to express their grievances over perceived deprivation, frustration, injustice or violation of fundamental human rights by governments or constituted authorities (Stekelenburg & Klandermans, 2010). The Ferguson riots for example raised questions and debates in the United States about the relationship between law enforcement officers and African Americans, the militarization of the police and the use of force doctrine in America.

BACKGROUND

The spread of social/political protests around the world has been attributed to the impact of new information and communication technologies (ICTs). People who are concerned about political events and developments in their countries are radically responding and mobilizing against oppressive regimes and systems and demanding for change (Smith 2010). According to Garret (2010) social media and ICTs such as *Twitter* and *Facebook* are currently changing the ways in which activists communicate, collaborate, and demonstrate. And text messaging has played im-

portant roles alongside these other social media, sometime earning government ban. Though these protests respond to very different socio-political and economic circumstances, they share the same social features (Gonzalez-Bailon et al, 2011).

Social movement theorists have established that during civil uprisings, the media become a site for the construction of meaning and a space for the discursive and symbolic contests (Gamson, 2004). The use of the new forms of ICTs, particularly, *Twitter*, *Facebook* and text messaging have been sites for framing contests and as mobilizing structures. Due to their anonymous and decentralized characteristics, these networks are hard to control by government authorities and rapidly emerging as primary tools in organizing protests especially among tech-savvy youths. These virtual networks are also able to disseminate injustice frames in the form of videos, cartoons, or slogans to local and international audiences (Sarfati, 2015).

Social media use in protests is a modern (new) development; this also suggests that research literature in this field is still emerging. Studies on social media roles in contemporary mass protests have centered on *Facebook* and *Twitter*, while mobile phone use has not been explicitly documented. However, before the current overwhelming popularity of *Facebook* and *Twitter* in the organization and implementation of offline protests, text messaging via Smartphones had been used to champion revolutionary actions (Hong, 2006). For instance, Ahrens (2001) had observed that cell-phone was used to coordinate the protest against the World Bank, by American University activists in 2000, and concluded that mobile phone is a weapon in the hands of the mobilizing people seeking social justice. Obadare (2005) also documented the GSM boycott in Nigeria in 2005 that forced mobile phone operators to “vigorously renew their commitment to corporate social responsibility” (p.24). In his study of “how texting helped fuel the anti-austerity protests roiling Europe,” from early 2015, Jacob Groshek, an assistant professor of emerging media in the Department of Communication, Boston University, outlines the success

Smartphone technology and texting in mobilizing protests across European union states (e.g. Brussels in Belgium where over 100,000 protesters took to the streets in early October, 2015 and the police had to deploy water cannon to disperse them). In Groshek's own words:

my research shows that one of the most basic forms of modern communication and simplest uses of a Smartphone, the text message, was crucial to mobilizing anti-austerity efforts across the European Union... What I found (using the regression modeling method) was that the volume of text messages sent in all countries was a positive predictor of protest activity, as were youth unemployment rates. What this means is that as the total number of text messages sent went up in one year, protest activity increased in the next year. The same was true for youth unemployment rates. More protests followed rising levels of joblessness among young people... (pages 1 & 2). The study focused on 24 European Union states.

The story of mobile phones in protest planning and implementation by activists ranges from the success of coordinated "flash mobbing" to "txtmobbing" and "swarming." A "flash mob" is defined in terms of a group of people brought together often by text messages or social media to a designated location at a particular time to perform an indicated action before dispersing; thus, texting and media technology enable flash mobbers/protesters to instantaneously communicate with one another and are empowered to immediately change venue such as train stations, parks, town squares, and in some cases challenge or evade authorities (Fitzpatrick, 2013). Similarly, a *TXMmob*, which was developed in collaboration with protest organizers, and has been applied by widespread adoption of SMS-enabled cell phones among activists and by evolutions in protest tactics, facilitated new social formations and modes of participation (Hirsch & Henry, 2005:1455). Texting has also been instrumental for coordinating "smart mobs" (Rheingold, 2002) or swarming.

The latter, which is recognized as a tactic of mobilizing actions by decentralized groups, is used in military parlance to describe a "dispersion of command among many small, autonomous units that are able to collectively attack an enemy from all directions" (Hirsch & Henry, *ibid*). This method was used to describe the June 18 (J18) actions that occurred around the world, which also paralyzed central London in 1999, as well as the demonstrations that shut down a World Trade Organization (WTO) meeting in Seattle in the same year (see Hirsch & Henry, 2005). Text messaging appeared as an ideal medium to coordinate these types of protests since cell phones are common and users would not immediately be identified as members of protest movements. Widespread cell phone use also meant that the protest networks could be expanded by thousands of activities without requiring the purchase of expensive radio technology. SMS is a technology that is widely adopted by many social sectors and relies on vibrant networks that are controlled by large companies (*ibid*).

In addition to texting, few studies have documented the use of Blackberry Instant Messaging (an extended system of mobile phone technology) in protests. Reporting the north London riots of 2011, Halliday (2011) notes that Blackberry handsets was used by about 37% of British teenagers. Blackberry Messenger (BBM) enables users to send one-to-many messages to their network contacts, who are connected by "BBMPINS." Unlike Facebook or Twitter, many BBM messages are not traceable by the authorities making BBM the most preferable by teens to spread gossip and mobilize for civil unrests. In the London protests for example, protesters had first gathered on Facebook with more than 7,500 fans to mourn and vow to avenge the murder of a Tottenham resident (Mark Duggan) and by 10.30 pm on Saturday (August 6), protesters had thronged the Tottenham High Road in a first public show of protest; this turned into a full blown uprising with widespread rioting and looting. Sheid (2013) further adds that BBM allows users to send free messages to individuals and to all their contacts simultaneously; the

belief remained that individuals were using the BlackBerry as a tool for coordinating activity, allowing rioters and looters to stay one step ahead of the police. This had warranted the call for the suspension of BBM by a Member of Parliament for Tottenham who identified BBM as “the reason that criminals were outfoxing the police force, noting that encryption prevented access for the police” (p.176).

OFFICIAL RESPONSES TO SMS IN SOCIAL PROTESTS

Recognizing the importance of information sharing during protests and the danger inherent in some actions considered as unlawful, some governments, the police and security agents have attempted to disrupt communication networks by interfering with activist mobile technology based media and by arresting activist group members. Two men, for instance, were arrested and charged for sending mobile text messages and using social media sites to recruit people for “violent” demonstrations in Sidney in September, 2012. The man was accused of planning an anti-Islamic protest and had sent texts and online messages to incite members of the public to participate in a Sidney protest and commit acts of violence (Ralston, 2012). Chittum (2013) similarly reports the case of the “temporal” interruption of mobile wireless service at train stations in San Francisco in 2011 by the Bay Area Rapid Transit (BART) after they suspected a flash mob protest following the killing of a man by the BART police. The action was intended to prevent protesters from taking advantage of mobile communication devices to coordinate surprise demonstrations. This tactics drew criticisms from several commentators who compared BART actions to what was typical of Middle East despots. Most people wondered whether interference with cell phone service does not amount to an unconstitutional infringement on citizens’ rights to free speech. Lackert (2013) agreed that while the BART station shutdown

was proper to protect public order, the unilateral action raised significant legal questions as to whether this was authorized under the US telecommunications law relating to passengers’ right to access telephone networks and the legality of a shutdown by a quasi-governmental authority (p. 577). Hence, smartphones and social networking platforms have been recognized globally as having heightened the provision of greater capability of people to organize and implement protests quickly via text, Twitter, and Facebook and non-Internet based cell phones. Countries where the Internet was shut down, as in Egypt, the only means of communication was through the existing telephone systems; this service was not easily or centrally controlled by Internet servers. This is why governments and institutions across the political spectrum have become alarmed at this development, and have struggled to give varying weights to public expression in relation to public order.

CURRENT SCIENTIFIC KNOWLEDGE

The first major work that articulates mobile phone’s collective actions is by Dr. Howard Rheingold, an expert and leading researcher in digital media at the University of California, Berkeley. His book: *Smart Mobs: The Next Social Revolution* (2002) is an important and comprehensive account of mobile phone use in contemporary activism. Rheingold explores the growing importance of “smart mobs” (also referred to as “mobile ad hoc networks”) in civil collective action. Citing the cases of the WTO (anti-globalization) protests in Seattle in 1999, and the “people power 11” protests in the Philippines that ousted Joseph Estrada in 2001, he suggests that the prevalence of smart phone and its influence is indeed responding to peoples’ need of cooperation amplified by ICTs (see also Mater, 2004; Monterde and Postill, 2013). Clay Shirky –a professor of New Media at New York University gives a detailed and interesting account of the power of text messaging in the Philippines protests of 2001. In his “the

political power of social media,” he recounts: “on January 17, 2001, during the impeachment trial of Philippine President Joseph Estrada, loyalists in the Philippine Congress voted to set aside key evidence against him. Less than two hours after the decision was announced, thousands of Filipinos, angry that their corrupt president might be let off the hook, converged on *Epifanio de los Santos Avenue*, a major crossroads in Manila. The protest was arranged, in part, by forwarded text messages reading, “Go 2 EDSA. Wear blk.” The crowd quickly swelled, and in the next few days, over a million people arrived, choking traffic in downtown Manila. The public’s ability to coordinate such a massive and rapid response - close to seven million text messages were sent that week - so alarmed the country’s legislators that they reversed course and allowed the evidence to be presented. Estrada’s fate was sealed; by January 20, he was gone. The event marked the first time that social media had helped force out a national leader. Estrada himself blamed ‘the text-messaging generation’ for his downfall...” (p.1). Similarly, Professor Shirky, in his landmark publication: *Here Comes Everybody: The Power of Organizing without Organizations* (2008), establishes the potentials of mobile technologies for collective action arguing that they promote the rise of new forms of mass action by reducing the users’ investment in time and money.

Monterde & Postill (2013) further gives interesting insights to the *uses of mobile phones for social protests by Spain’s indignados*, where text messages gave the Spaniards an “an alternative information channel” to the mainstream media, and like in Philippines, protesters “passed on” text messages that resulted in “unstoppable snowball effect,” (p.3). Jan H. Pierskalla at the German Institute of Global and Area Studies (GIGA) and Florian M. Hollenback at Duke University (2013) examine the roles of *technology and collective action and the effect of cell phone coverage on political violence in Africa* and argue that the spread of cell phone technology across Africa has transforming effects on the economic and political

sphere of the continent; and that the availability of cell phones allows political groups to overcome collective action problems more easily and improve in-group cooperation and coordination (p.1). A recent research has also tested the possible impact of mobile phones on voter information and participation, as well as its possible impact on fighting electoral fraud and corruption (see Aker, Collier, and Vincente 2011; Bailard 2009).

Beginning with the Arab spring, and social protests multiplying in western liberal democracies being confronted by financial crises and changes in welfare policies, more interesting studies are emerging with important findings on the effective roles of mobile telephony and web 2.0 in the protests. Kavanaugh et al (2011) give important insights to the functions of text messages in the Iran, Tunisian and Egyptian revolutionary protests, where SMS messages were not completely shut down (as in Iran 2006), because subscribers could easily be identified and possibly arrested. In Egypt, it was not technically difficult to disrupt mobile phone service as the Egyptian government did in one of the days of the protests; they simply ordered the Egyptian operator (ECMS) to shut off service, especially in the main demonstration areas of Cairo and Alexandria. Mobile phone signals were patchy and text messages inoperative (p.9). Tufekci & Wilson (2012) further provide quantitative documentation of the role of texting in information gathering and discussions of the Egyptian protests that ousted Hosni Mubarak. According to the study, about 82% of respondents used mobile phones to communicate about the protests in Egypt, with 62% texting; while 51% used Facebook. It further demonstrates that events in North Africa and the Arab world were being shaped by a new system of political communication which sets into sharp relief the importance of digitally mediated interpersonal communication, characterized by the increasingly interrelated use of satellite television, social media platforms and the widespread use of Internet-enabled cellphones capable of transmitting photos and video (p.375). This study further illustrates how elements of this

system contributed to lowering the costs of initiating and coordinating collective action to topple longstanding repressive regimes. It concludes that the high degree of adoption of digital communication in digital activism, illustrates how citizens overcome the potential risks of online activism in authoritarian regimes (Tufekci & Wilson, 2012).

Ever before the Egyptian uprising, Sean Jacobs and Diana Duarte writing for *Afronline news* had reported the “bread riots,” in Mozambique chronicling the power of SMS. According to the report, residents of the capital, Maputo, took to the streets having been prompted by text messages that told them to “enjoy the great day of the strike” and to “protest the increase in energy, water, mini-bus taxi and bread prices.” The government had insisted that the increases in price were not reversible. But days later, under pressure from rapidly SMS-organized protests, the government reconsidered their position and the old bread prices returned. And as a sign of its impact, the government, allegedly ordered cell phone service providers to briefly suspend text messaging for its users. Journalists that covered the protests as well as commented on social media were surprised at the power of SMS in a small African country where twice as many Mozambicans have cell phones as they have access to electricity. Opinion speculated about expanded opportunities for political engagement in developing countries (Jacobs & Duarte (2010:1).

Similarly the “occupy Nigeria” protests of January, 2012 was organized and implemented largely by the adoption of social media and text-messaging. The study concludes that with the growing Internet literacy and common use of mobile phones in Nigeria, many youths have also been increasingly aggressive in the use of social media and mobile telephony to engage in sociopolitical debates and critique government actions. Moreover, there is no ruling out the possibility that many Nigerian online activists are gradually more in touch with their counterparts in North Africa or elsewhere (Chiluwa, 2015).

Recounting, the Gezi Park uprising in Turkey in 2013, Yusuf Sarfati (2015), discovers that social

media and text messaging not only mobilized protesters, they also became “a tool for counter-mobilization” because “the well-organized supporters of Turkey’s governing JDP were quick in responding to the protesters and circulating counterframes to both national and international target audiences. This shows that social media can be a potential tool in organizing protests against the state, but can also be used by the state and its allies to diffuse or discredit protests,” (p.29). The Gezi Park protests began on May 30, 2013, when a group of environmentalists disagreed with the proposed uprooting of trees in Gezi Park at the city centre in Istanbul by the municipal authorities. The Municipal council began to implement its plan to burn down the park and build in its place a replica of an Ottoman era military barrack. Protesters were manhandled by the municipal police and tents of the camping protesters in Gezi Park were burnt down. When the images of wounded protesters were transmitted via social media and text messages, more enraged protesters gathered at the Taksim square adjacent to Gezi Park and in few days, the protest movement had been transformed into a national uprising against the government, mobilizing million all around Turkey. The riot police used tear gas, water cannons and pepper spray to disperse protesters (Sarfati, 2015, p.25). The use of texting and twitter as counter-mobilization strategy, especially for explaining the policy and activities of government is significant here, not only in Turkey but also in Nigeria, where the former President Jonathan and the current Senate President (Bukola Saraki), use twitter and sometimes, SMS to explain government positions and mobilize supports.

FUTURE RESEARCH DIRECTIONS

Shanthi Kalathil and Taylor Boas in their very popular and insightful book: *Open Networks, Closed Regimes* (2003) re-examine the use of the Internet and ICT-based communications in social protests, and argue that the impact of social media

and mobile phones towards effecting political change has actually been exaggerated. Based on an empirical assessment of evidence from China, Cuba, Singapore, Vietnam, Burma, the UAE, Saudi Arabia and Egypt, they contend that the Internet is indeed not a threat to authoritarian regimes. According to Morodov (2011) the assumed emancipator potential of new media technologies in social protests, actually strengthens the surveillance capabilities of repressive regimes. Shirky (2011) also doubts the effectiveness of information technology in organizing protests and argues that repressive governments are becoming better users of electronic tools to suppress dissents. These and a few other studies have brought to the fore the question: “does the internet and mobile telephony really achieve any significant political change?” In my opinion, this question should further provide future directions in research on the role of modern ICTs in sociopolitical change. In his epoch-making work: *The Net Delusion: the Dark Side of Internet Freedom* (2011), Evgeny Morodov – a visiting scholar at Stanford University, and a Schwartz fellow at the New America Foundation, argues that “cyber utopia” (i.e. the assumption that the Internet is liberator of the masses), is erroneous, because it leads to a conclusion that texting and social media can be used to bring down authoritarian governments and advance democracy. He advises that rather, the American people should shun their belief in cyber utopianism and acknowledge the downside of online communications. He believes that the public should come to terms with “net delusions” and implement a policy of cyber realism in order to achieve their political goals. While acknowledging the positive impact of social media in facilitating mobilization by reducing barriers to collective action, Bridwell (2013) further argues that online-based actions and mobilizations can work in non-democratic regimes. Hence, the advantages of social media and ICTs are especially suited to mobilizing in non-democracies. “In a democracy, the particular advantages of social media are more redundant, even if its tools offer improvements for mobiliz-

ers. Moreover, the negative qualities of social media cited by skeptics are more likely to hold in a democratic context” (p.3). Factors such as how gender and economics can affect access to new technologies should also be taken into consideration in assessing the utility of social media in protests (Christensen, 2011).

Arguably, texting and social media have been instrumental in mobilizing protesters and effecting political change in places like Tunisia, Egypt, and Libya. But the results are different in Bahrain, Syria, Turkey and Ukraine. In the Nigerian fuel subsidy protests, the protests failed to achieve their purpose because demonstrators were divided along ethnic lines. Some protesters also pledged their loyalty to different political parties; thus, while protesters were mobilized by the new ICTs, the offline protesters were not united in purpose. Citing the uprisings in Iran and Bangkok, Shirky (2011) also notes that the June 2009 protests by the Green Movement in Iran were a failure because, while the activists used every possible technology available to them to coordinate the protest against electoral misconduct, they were dispersed by a violent government crackdown. The Red Shirt uprising in Thailand in 2010 was similarly dealt with by the Thai government although protesters were well mobilized with the help of the social media. Many of the protesters even lost their lives.

CONCLUSION

Despite the skepticism among scholars about the effectiveness of ICT technology in achieving an impactful social change, a recent study by a women advocacy group, believes that an exclusive use of mobile phone and SMS will promote women activist efforts and reduce gang violence against women. In her position paper titled: “Mobilizing Change in Central America: Fostering Women’s Networks to Combat Gang Violence, (2016),” Emily S. Wasek of the *Institute for the Theory and Practice of International Relations at The College of William and Mary*, Virginia argues

that “despite women’s restricted movement within gang territories, cell phone technology can provide Central American women with a form of ‘immobile mobility’ that surpasses physical and cultural boundaries... With mobile phone technology, activists are able to more easily coordinate mass movements through the potentially viral effects of forwarding messages. Thus, access to mobile phones can help Central American women safely create anti-gang political violence.” She therefore called on the United States government to “encourage the development of women’s political networks through the use of mobile phone technology. In so doing, policymakers can both empower marginalized Central American women and enhance current policies aimed at disrupting gang activities” (Wasek, 2016, p.5).

It is clear therefore that texting via SMS has come to stay and have formed significant part of our daily living; and it is impossible to deny its effect and influence in persuasive communication. Whether immediate or on the long run, the results of activism and protests orchestrated by the social media and text messaging are significant. However, I still believe that it is important that researchers take a more cursory look at different political contexts and some unique nature of protests in order to properly assess the impact of the Internet and mobile telephone in social change

REFERENCES

- Ahn, J. (2011). The Effect of Social Network Sites on Adolescents: Social and Academic Development: Current Theories and Controversies. *Journal of the American Society for Information Science and Technology*, 62(8), 1435–1445. doi:10.1002/asi.21540
- Ahrens, F. (2001, April 20). For Activists Today: Its Marks, not Marx. *The Washington Post*.
- Aziz, S. (2013). The Impact of Texting/SMS Language on Academic Writing of Students: What do we need to panic about?. *Elixir Linguistics & Translation*, 55, 12884–12890.
- Bell, G. (2006). ‘No More SMS from Jesus: Ubicomp, Religion and Techno-spiritual Practices. In P. Dourish & A. Friday (Eds.), *Ubicomp 2006, LNCS 4206* (pp. 141–158). doi:10.1007/11853565_9
- Bridwell, J. M. (2013). *Twitter, Texting, and Street Demonstrations: Assessing Social Media’s Political Relevance for Citizen Empowerment*. Original Speech Transcript from the State Department. Retrieved from <http://www.state.gov/secretary/rm/2010/01/135519.htm>
- Campbell, H. (2005). *Exploring Religious Community Online*. New York: Peter Lang-Digital Formation Series.
- Campbell, H. (2006a). Religion and the Internet. *Communication Research Trends*, 25(1), 1–24.
- Chiluwa, I. (2012). Social Media Networks and the Discourse of Resistance: A Sociolinguistic CDA of Biafra Online Discourses. *Discourse & Society*, 23(3), 217–244. doi:10.1177/0957926511433478
- Chiluwa, I. (2015). ‘Occupy Nigeria 2012’: A Critical Analysis of Facebook Posts in the Fuel Subsidy Removal Protests. *Clinica: An Interdisciplinary Journal of Translation, Interpreting and Intercultural Communication*, 1(1), 47–69.
- Chiluwa, I., & Uba, E. (2015). Texting and Christian practice. In Z. Yan (Ed.), *Encyclopedia of Mobile Phone Behavior* (Vol. 1, pp. 369–376). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-8239-9.ch030
- Chiluwa, L. & Ajiboye. (2016). Discursive Pragmatics of T-shirt Inscriptions: Constructing the Self, Context and Social Aspirations. *Pragmatics & Society*, 7(3), 436–462.

- Chittum, T. (2013). 'Can you hear me now?' Cell Phone jamming and the tenth amendment. *Nevada Law Journal*, 13, 257–289.
- Christensen, C. (2011). Discourses of Technology and Liberation: State Aid to Net Activists in an Era of Twitter Revolution. *Communication Review*, 30(2), 233–253. doi:10.1080/10714421.2011.597263
- Gamson, W. A. (2004). Bystanders, Public Opinion, and the Media. In D. A. Snow, S. A. Soule, & H. Kriesti (Eds.), *The Blackwell Companion to Social Movements* (pp. 242–261). Malden, MA: Blackwell Publishing.
- Garrett, R. K. (2006). Protests in an Information Society: A Review of Literature on Social Movements and New ICTS. *Information Communication and Society*, 9(2), 202–224. doi:10.1080/13691180600630773
- Gonzalez-Bailon, S. (2011). *The Role of the Social Media in Protests*. University of Oxford. Retrieved from http://www.ox.ac.uk/media/news_stories/2011/111612.I
- Hong, C. (2006). New Political Tool: Text Messaging. *Christian Science Monitor*. Retrieved from http://usatoday30.usatoday.com/tech/news/2005-06-30-politics-text-tool_x.htm
- Jacobs, S. & Duarte, D. (2010, September 16). Protest in Mozambique: The Power of SMS. *Afronline.com*.
- Kalathil, S., & Boas, T. (2003). *Open Networks, Closed Regimes: the Impact of the Internet on Authoritarian Rule*. Washington, DC: Carnegie Endowment for International Peace.
- Kavanaugh, A., Yang, S., Sheetz, S. Li, L.T., & Fox, E.A. (2011). Between a Rock and a Cell Phone: Social Media Use during Mass Protests in Iran, Tunisia and Egypt. *ACM Trans. of CHI*, 1-10.
- Lackert, R. (2013). BART Cell Phone Service Shutdown: Time for a Virtual Forum. *Federal Communications Law Journal*, 64, 577–598.
- Mater, M. (2004). *Filipino Smart Mobs. The Emergence of Counter-publics in Asian Civil Society?* Paper presented at the 13th Annual Asian Media Information and Communication Centre Conference: Impact of New & Old Media on Development in Asia.
- Monterde, A., & Postill, J. (2013). Mobile Ensembles: The uses of Mobile Phones for Social Protest by Spain's Indignados. In G. Goggin & L. Hjorth (Eds.), *Routledge Companion to Mobile Media*. Routledge.
- Morozov, E. (2011). *The Net Delusion: The Dark Side of Internet Freedom*. New York: Public Affairs, Perseus Book.
- Obadare, E. (2005). *The GSM Boycott: Civil Society, Big Business and the state in Nigeria*. Civil Society Working Paper No. 23, 1-34. The London School of Economics and Political Science.
- Pierskalla, J., & Hollenbeck, F. (2013). Technology and Collective Action: The Effect of Cell Phone Coverage of Political Violence in Africa. *The American Political Science Review*, (May), 1–18.
- Ralston, N. (2012). *Two Charged over Mobile Protest Messages*. Retrieved from <http://www.smh.com.au/action/externalEmbeddedPlayer?id=d=26b81620349>
- Rheingold, H. (2002). *Smart Mobs: The Next Social Revolution*. Cambridge, MA: Perseus Books.
- Sarfaki, Y. (2015). Dynamics of Mobilization during Gezi Park Protests in Turkey. In I. Epstein (Ed.), *The Whole World is Texting: Youth Protest in the Information Age* (pp. 25–44). Rotterdam: Sense Publishers. doi:10.1007/978-94-6300-055-0_2

Scheid, M. (2013). Shutting Down Dissent: Global Responses to Mobile-Fueled 'Flash Mobs' and Political Protests. *I/S Journal of Law and Policy*, 9(1), 173–200.

Shirky, C. (2008). *Here Comes Everybody: The Power of Organizing Without Organizations*. New York: Penguin.

Shirky, C. (2011). The Political Power of Social Media: Technology, the Public Sphere and Political Change. *Foreign Affairs*, 90(1), 28–31.

Stekelenburg, J. & Klandermans, B. (2010). The Social Psychology of Protest. *Sociopedia.isa*.

Torres, J., & Conaway, R. (2014). Managerial Use of Text Messaging in International Organizations. *Global Advances in Business Communication*, 3(1), Article 5. Available at: <http://commons.emich.edu/gabc/vol3/iss1/5>

Tufekci, Z., & Wilson, C. (2012). Social Media and the Decision to Participate in Political Protest: Observations from Tahrir Square. *Journal of Communication*, 62(2), 363–379. doi:10.1111/j.1460-2466.2012.01629.x

ADDITIONAL READING

Florian, F. (2013), 'Cascades of Collective Actions? Analysing the Impact of Protest History and Social Media on Regime Change in the Context of the 2011 Uprising in Egypt and Syria' *CGP Working Papers*. Pp1-47

Halliday, J. (2011), 'London Riots: How BlackBerry Messenger Played a Key role.' *The Guardian*, August 8

Harley, D., Winn, S., Pemberton, S., & Wilcox, P. (2007). Using Texting to Support Students Transition to University'. *Innovations in Education and Teaching International*, 44(3), 229–241. doi:10.1080/14703290701486506

Tad, H., & John, H. (2005) 'TXTmob: Text Messaging For Protest Swarms' *CHI Conference*, April 2-7

Tagg, C. (2009) 'A Corpus Linguistics Study of SMS Text Messaging.' Unpublished PhD Thesis. The University of Birmingham.

Thurlow, C. (2003). 'Generation Txt? Exposing the Sociolinguistics of Young Peoples Text-Messaging. Discourse Analysis Online.

Thurlow, C., & Poff, M. (2009). The Language of Text-Messaging. In S. C. Herring, D. Stein, & T. Virtanen (Eds.), *Handbook of the Pragmatics of CMC*. Berlin, New York: Mouton de Gruyter.

Wasek, E. S. (2016). 'Mobilizing Change in Central America: Fostering Women's Networks to Combat Gang Violence.' *The Project on International Peace and Security*. Virginia: Institute for the Theory and Practice of International Relations. The College of William and Mary.

Yun, X. (2012). Chinese Use of Mobile Texting for Social Interactions: Cultural Implications in the Use of Communication Technology. *Intercultural Communication Studies*, 21(2), 131–150.

KEY TERMS AND DEFINITIONS

Computer-Mediated Communication (CMC): Communication that takes place through the use of electronic devices, usually by individuals who are connected either online or a network connection using social software and interact with each other via separate computers.

Mobile Phone (or Cell Phone): A device that can make and receive telephone calls over a radio link, while moving around a wide geographical area. This is possible by connecting to a cellular network provided by a mobile phone operator, allowing access to the public telephone network (see *Wikipedia*, 2013).

Mobile Technology: Technology used for cellular communications.

Protest: The expression of objection by protesters through words or actions or other forms of social behavior that rejects or resists certain government policies, or the existing sociopolitical structures.

Protester: Someone engaged in a protest.

SMS: 'Short message service' is text messaging service component of the mobile phone, web or mobile communication systems.

Social Media: The interaction among people in which they create, share, and/or exchange information and ideas in virtual communities and networks (see Ahlqvist et al 2008); or 'a group of internet-based applications that build on the ideological and technological foundations of *Web 2.0* and that allow the creation and exchange of usergenerated content (see Kaplan & Haenlein 2010).

M

A Survey of People Localization Techniques Utilizing Mobile Phones

Levent Bayındır

Ataturk University, Turkey

INTRODUCTION

With the ongoing diffusion of mobile computing and context-aware applications, knowledge of the current location of an individual can be leveraged in a number of different domains, from personal diaries and fitness-related applications to human behavior analysis and targeted advertising.

Mobile phones are routinely carried by their owner during daily activities, and the embedded sensors commonly found in modern devices can be used to capture data of interest from the area surrounding an individual; these data can then be processed to estimate the current location of the phone user.

Two different categories of localization studies can be identified: indoor/outdoor detection and indoor localization. In the first category, the aim is to differentiate between indoor and outdoor areas, while the second category comprises all the studies addressing the challenges encountered in identifying the location of a person in an indoor environment, where the majority of people spend most of their time, and where the technology typically used for outdoor localization (i.e. satellite-based positioning) is generally not available.

The capability to automatically differentiate between indoor and outdoor environments can be used for example to enable an efficient use of power-hungry sensors such as GPS receivers, possibly selecting alternative localization methods while indoors; applications that recognize and analyze human activities can leverage indoor/outdoor detection for improved accuracy, given that many activities can only be executed either in

indoor environments or outdoors; other examples of possible uses are automatic image annotation when a picture is taken with a smartphone, or adaptation of phone settings to the current environment type (e.g. increasing ringer and notification sound volume when outdoors).

A typical use of indoor localization techniques is for personal navigation, especially useful in large buildings such as shopping malls, hospitals, airport terminals, university campuses, or office buildings; more advanced uses are made possible by the proliferation of mobile apps that can leverage location information: for example, a museum visitor can access detailed information about the nearby items on display; a customer in a grocery store can search for a specific product and easily reach the aisle where the product is located; other applications allow precise tracking of people in need of care, such as children, elders or disabled people. In addition, human activity classification algorithms can greatly benefit from knowledge of the physical location of an individual, since in many cases location is highly correlated to the activity being carried out (e.g. sleeping is typically done in the bedroom, and cooking in the kitchen).

This article presents a review of past research works describing techniques for utilizing smartphone sensors to identify the environment where a smartphone user is located. The review focuses on studies where user location can be computed autonomously and continuously by a smartphone, without the need for an active involvement of the user, and where issues such as power consumption and dependence of sensor readings from the on-body position of the phone are addressed.

BACKGROUND

In the last few decades there has been an increasing interest in positioning technologies. The deployment of a number of satellites in the Earth's orbit enabled satellite-based positioning, whose main use case was vehicle navigation, but due to poor performance of this technology in indoor areas, indoor location methods have to rely on other means. The first indoor location techniques required carrying specialized devices and/or deploying ad-hoc hardware in the environment; then, the continuous enhancement of mobile phone sensing and computation capabilities, and the widespread deployment of infrastructure for wireless communication opened new frontiers for indoor localization; now, an increasing number of location-based services are made possible by different technologies for locating people in indoor environments.

Many pervasive computing applications are enabled or can be enhanced by knowledge of the current *user context*; while the exact definition of user context can vary between applications, physical location is an important piece of information in defining the context for many applications. Thus, methods for automatic localization of users can be considered as part of the more general issue of user context recognition (Hoseini-Tabatabaei, Gluhak, & Tafazolli, 2013).

Virtually all sensors and communication interfaces embedded in modern smartphones can be used for localization: receivers for wireless technologies such as GSM, GPS, Wi-Fi, Bluetooth, and even FM radio can detect and identify existing infrastructure such as cell towers or Wi-Fi access points; inertial and orientation sensors can detect physical movements and orientation changes of the phone, and thus their readings can be correlated to movements of the phone user; other sensors can capture characteristics of the surrounding environment from which useful information for localization can be extracted.

Two basic methods can be identified in most indoor localization applications: fingerprinting

and dead reckoning (Subbu, Zhang, Luo, & Vasilaikos, 2014). These two methods, which can be used alone or in combination, utilize sensor readings to detect local characteristics of the environment or movements of a mobile device, and are often complemented by additional techniques (Yang et al., 2015) to overcome their inherent limitations and improve localization performance: Shang, Hu, Gu, Wang, & Yu (2015) identified three categories of methods that fuse sensor readings with *spatial contexts*: map matching, landmark fusion, and spatial model-aided methods.

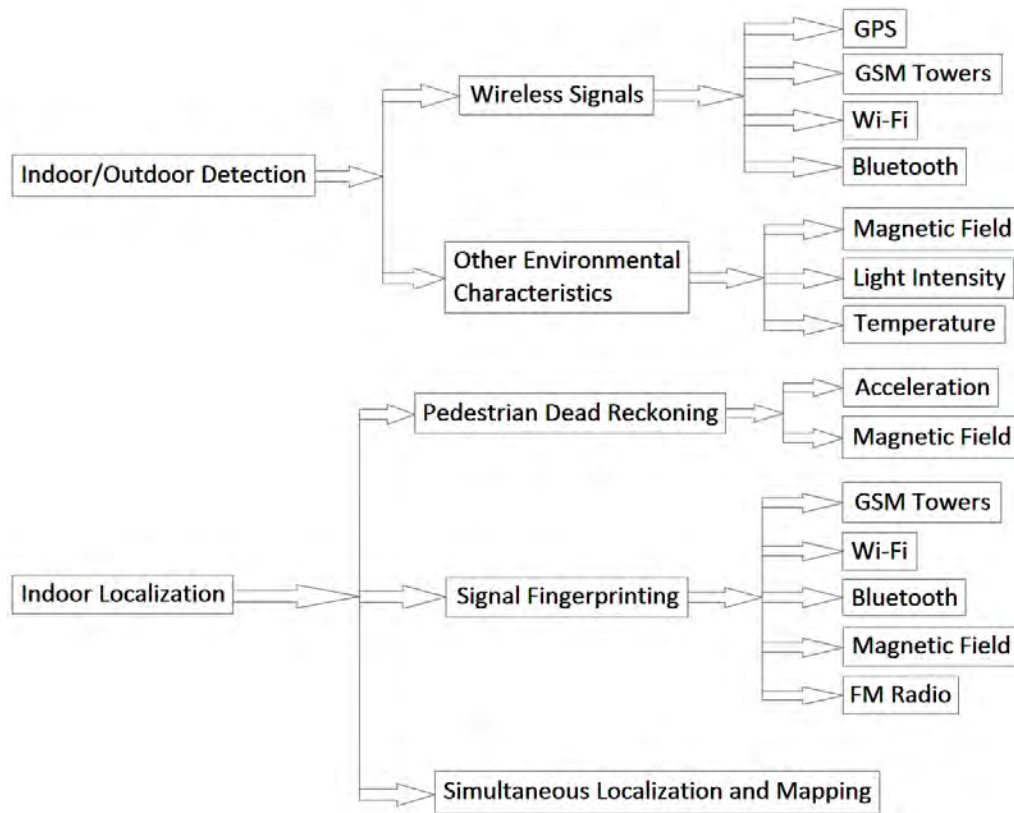
Figure 1 provides a taxonomy for the different studies on indoor/outdoor detection and indoor localization reviewed in this article.

INDOOR/OUTDOOR DETECTION

In a number of recent research studies dealing with environment detection, the focus is on discriminating between indoor and outdoor environments. Since all smartphones are equipped with cameras, among the different methods that can be implemented in a smartphone to detect the environment type there are vision-based methods that analyze images captured from the camera to extract features useful for scene understanding. These techniques, which have been studied extensively even before the advent of smartphones, will not be reviewed here, since when implemented in a smartphone they usually require active user involvement to capture the scenes of interest, and are not applicable in most situations where the mobile device is worn randomly by the user.

General unavailability of satellite coverage for navigation purposes in indoor areas can be used to infer indoor/outdoor location. For example, Hermann, Zappi, & Rosing (2012) and Ravindranath, Newport, Balakrishnan, & Madden (2011) used the inability to obtain a GPS location as a hint that the user is indoors; in Cho, Song, Park, & Hwang (2014), from identification of visible satellites and information on their current position in the Earth orbit, the directions on which there is a clear

Figure 1. Taxonomy of people localization techniques utilizing mobile phones



view of the sky, and conversely the directions on which the view of the sky is obstructed, are identified, and this information is used to classify the environment type as indoors or outdoors. Other researchers combined observation of GPS signals with other wireless signals that have different characteristics in different environment types: Pei et al. (2013) used the strength of Wi-Fi and GPS signals to derive a measure of the probability of being indoors or outdoors; Parviainen, Bojja, Collin, Leppänen, & Eronen (2014) extracted a set of features from Wi-Fi, GPS and Bluetooth signals, and used them as input to a binary indoor/outdoor classifier; Anagnostopoulos & Deriaz (2015) based their method on estimation of the accuracy of two complementary localization techniques (GPS-based and Bluetooth-based).

Since the GPS receiver is a major source of power consumption, many researchers studied methods to avoid utilizing this sensor in indoor

areas (in fact, this is one of the motivations for indoor/outdoor detection studies): for example, in the Wi-Fi-based indoor navigation system by Gallagher, Li, Dempster, & Rizos (2011), Wi-Fi networks near the entrances of a building are marked as “transition zones”, and the GPS receiver of the mobile phone is turned on only when located in such transition zones; Soria Morillo, Ortega Ramírez, Alvarez García, & Gonzalez-Abril (2012) developed a method to automatically learn the pattern of Wi-Fi signals observed when entering a building, so that when a similar pattern is observed at a later time the system can infer a transition to an outdoor area; Zou et al. (2016) deployed a set of fixed Bluetooth devices in semi-outdoor areas so that a mobile phone can infer a transition between indoors and outdoors when these devices are sensed by the Bluetooth receiver. In Chen, Sivaraman, Das, Ravindranath, & Balakrishnan (2015), Wi-Fi signal strength is

used as input to a binary classifier for indoor/outdoor detection.

Usually, measurement of wireless signals by a mobile phone incurs a relatively high cost in additional energy consumption (a notable exception is given by cell tower signals, because the relevant radio receiver in the phone is always powered on for proper phone operation); for this reason, some studies focused on utilizing for indoor/outdoor detection more battery-friendly sensors, which can capture other characteristics that can help in differentiating between indoor and outdoor environments. One of these characteristics is given by the magnetic field, which due to the presence of electric appliances and steel structures is typically subject to high variations inside buildings; other characteristics, such as light intensity and temperature, are purposely modified by people indoors in order to reach desired comfort levels. These ambient characteristics can be acquired by low-power sensors in modern smartphones, and as such can be used for power-efficient indoor/outdoor detection algorithms (Nakamura, Ono, Sekiya, Honda, & Takahashi, 2015; Radu, Katsikouli, Sarkar, & Marina, 2014; Zhou, Zheng, Z. Li, M. Li, & Shen, 2012).

INDOOR LOCALIZATION

Dead reckoning methods have been extensively studied for localization in indoor areas; more specifically, pedestrian dead reckoning (PDR) refers to dead reckoning techniques explicitly modeling movement of pedestrians as a sequence of discrete walking steps, and thus calculating the displacement of a pedestrian via step detection, heading estimation and step length estimation. While step detection can be implemented relatively easily in smartphones equipped with an accelerometer, heading estimation is generally more difficult, due to the fact that phone orientation can be different from user orientation, and the presence in most indoor settings of high magnetic field disturbances is a major source of noise in magnetometer readings;

some of the techniques that have been adopted to address this issue involve using the gyroscope to complement noisy compass readings, and applying Principal Component Analysis to the acceleration signal (Deng, Wang, Hu, & Wu, 2015; Mohssen, Momtaz, Aly, & Youssef, 2014), based on the fact that the walking direction usually corresponds to the direction of maximum acceleration variance.

Signal fingerprinting is a localization method that exploits the presence in the environment of electromagnetic signals that change depending on the physical location. The most used signals for this purpose are the signals emitted by Wi-Fi access points, which are ubiquitously used in modern buildings to provide wireless network connectivity. Practical application of fingerprinting is made difficult by factors such as an irregular signal propagation pattern in indoor environments and the dependence of signal reception in a mobile device from the on-body position and orientation of the device; in various studies, different methods have been envisaged to deal with these issues: for example, in Sánchez-Rodríguez, Hernández-Morera, Quinteiro, & Alonso-González (2015), Wi-Fi signal strength values in the fingerprint database are combined with phone orientation information. Another disadvantage of fingerprinting methods is the need to build the fingerprint database, often by means of time-consuming site surveys; to mitigate this problem, Bisio et al. (2014) suggested the use of modeling and analytical estimation of wireless signal propagation instead of site surveys.

In addition to Wi-Fi, other types of signals have been investigated for fingerprinting: for example, magnetic field fingerprinting takes advantage of the presence of magnetic field anomalies inside buildings (Carrillo, Moreno, Úbeda, & Skarmeta, 2015; Zhang, Luo, & Wu, 2014); GSM fingerprinting leverages the presence of the cell towers used for mobile communications (Li, Bigham, Bodanese, & Tokarchuk, 2013); FM radio broadcast signals are another source of information that lends itself to fingerprinting (Chen, Lymberopoulos, Liu, & Priyantha, 2012; Moghtadaiee & Dempster, 2014; Yoon, Lee, &

Rhee, 2013). All these alternative signals can be used alone or in combination with Wi-Fi to improve localization performance.

Some studies investigated the use of PDR and fingerprinting methods combined; the most used techniques for fusing the two methods are Kalman filtering (G. Chen et al., 2015; Z. Chen et al., 2015; Deng, Hu, Yu, & Na, 2015; Faragher & Harle, 2013) and particle filtering (Hilsenbeck, Bobkov, Schroth, Huitl, & Steinbach, 2014; Shu et al., 2015). Other researchers (He, Chan, Yu, & Liu, 2015; Li et al., 2015) found that fingerprinting can be combined and take advantage of simple step detection methods, without explicitly calculating the heading direction of a pedestrian.

An inherent disadvantage of using a relative positioning method such as PDR is its tendency to accumulate the localization error over time; map matching techniques can help to solve this problem by adjusting location estimates according to a map of the indoor environment. Various map matching algorithms have been investigated, such as graph-based methods (Shih & Lan, 2012), particle filtering (Li et al., 2012; Qian, Pei, Ma, Ying, & Liu, 2015; Shang, Gu, Hu, & Kealy, 2015), Hidden Markov Models (Zhou, Li, Mao, Tu, & Zhang, 2015) and Conditional Random Fields (Xiao, Wen, Markham, & Trigoni, 2014).

Fingerprinting and map matching methods require the presence of a map of the indoor environment, and this is a potential problem because it may not be convenient to create or have access to updated maps for any indoor environment a user visits. To address this issue, simultaneous localization and mapping (SLAM) methods have been developed. SLAM is based on the principle that local characteristics of the environment can be dynamically learned and recognized as a place is visited multiple times. Different types of such characteristics have been studied, and Wi-Fi signals are probably the most used information source: the sequence of Wi-Fi measurements while a user walks in an indoor environment can be recorded and compared with data stored from previous visits, allowing the system to build a

model of the indoor space and to learn the adjacency between different parts of a building (M. Zhou et al., 2015). Learning the indoor layout is often facilitated by using PDR (Luo, Hong, & Chan, 2014), which can improve the accuracy of estimating the spatial relationship between different Wi-Fi measurements. Instead of implementing a full dead reckoning algorithm, other researchers (Wu, Yang, & Liu, 2015) estimated (via step detection) only the distance walked between successive measurements, while in other studies (Jiang et al., 2013) a simple motion sensing algorithm is used.

SLAM methods often utilize landmarks of the environment to facilitate the mapping process: such landmarks can be detected as a consequence of a mobile user moving in specific locations in the environment, like the inertial-based anchor points in Alzantot and Youssef (2012), or can be manually placed objects such as NFC tags (Mirowski, Ho, Yi, & MacDonald, 2013); in Zhang, Subbu, Luo, & Wu (2015), landmarks can be detected in correspondence of turns in walking paths, and can also be added manually by users by annotating the indoor map. Analogously to what happens when people learn how to navigate a new place, if known landmarks are sensed in a localization application an imprecise location estimate can be corrected.

FUTURE RESEARCH DIRECTIONS

The multitude of indoor localization techniques developed in recent years is filling the gap between indoor and outdoor localization; with the widespread use of Wi-Fi networks and the possibility of automatically building and updating indoor maps via crowdsourcing, internet-based navigation (Zeinalipour-Yazti, Laoudias, Georgiou, & Chatzimilioudis, 2016) has the potential to enable seamless transition between indoor and outdoor environments with continuous localization services.

As mentioned earlier, physical location represents only part of the context where users are located, and other characteristics of the environ-

ment are often equally important in defining the context. Methods for automatic environment classification (for example, to identify places such as restaurants, pubs, cafes, offices, homes, streets, parks, etc.) and high-level semantics annotation are already being researched (Ali, Basalamah, & Youssef, 2014; Parviainen, Bojja, Collin, Lepänen, & Eronen, 2014), and it can be expected that these methods will be further improved in the near future, to provide pervasive computing applications a user context including high-level semantic information.

CONCLUSION

This article reviewed the state of the art of localization techniques used in applications that can run on mobile phones; more specifically, two areas of interest have been identified: studies aiming at differentiating between indoor and outdoor environments, and studies tackling the issue of localization in indoors areas, where satellite-based positioning services are generally unavailable. Due to the presence of a large body of research (especially on the topic of indoor localization), the review has been limited to recent publications, which build upon research conducted in previous years and exploit technological advances in modern smartphones.

REFERENCES

- Aly, H., Basalamah, A., & Youssef, M. (2014). Map++: a crowd-sensing system for automatic map semantics identification. In *Proceedings of the 11th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)* (pp. 546-554). New York, NY: IEEE.
- Alzantot, M., & Youssef, M. (2012). CrowdInside: automatic construction of indoor floorplans. In *Proceedings of the 20th International Conference on Advances in Geographic Information Systems* (pp. 99-108). New York, NY: ACM.
- Anagnostopoulos, G. G., & Deriaz, M. (2015). Automatic switching between indoor and outdoor position providers. In *Proceedings of the 2015 International Conference on Indoor Positioning and Indoor Navigation (IPIN)* (pp. 1-6). New York, NY: IEEE. doi:10.1109/IPIN.2015.7346948
- Bisio, I., Cerruti, M., Lavagetto, F., Marchese, M., Pastorino, M., Randazzo, A., & Sciarrone, A. (2014). A trainingless WiFi fingerprint positioning approach over mobile devices. *Antennas and Wireless Propagation Letters*, 13, 832–835. doi:10.1109/LAWP.2014.2316973
- Carrillo, D., Moreno, V., Úbeda, B., & Skarmeta, A. F. (2015). MagicFinger: 3D magnetic fingerprints for indoor location. *Sensors (Basel, Switzerland)*, 15(7), 17168–17194. doi:10.3390/s150717168 PMID:26184230
- Chen, G., Meng, X., Wang, Y., Zhang, Y., Tian, P., & Yang, H. (2015). Integrated WiFi/PDR/smartphone using an unscented Kalman filter algorithm for 3D indoor localization. *Sensors (Basel, Switzerland)*, 15(9), 24595–24614. doi:10.3390/s150924595 PMID:26404314
- Chen, T. Y. H., Sivaraman, A., Das, S., Ravindranath, L., & Balakrishnan, H. (2015). Designing a context-sensitive context detection service for mobile devices. Massachusetts Institute of Technology, Computer Science and Artificial Intelligence Laboratory Technical Report, Cambridge, MA.
- Chen, Y., Lymberopoulos, D., Liu, J., & Priyanka, B. (2012). FM-based indoor localization. In *Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services* (pp. 169-182). New York, NY: ACM.
- Chen, Z., Zou, H., Jiang, H., Zhu, Q., Soh, Y. C., & Xie, L. (2015). Fusion of WiFi, smartphone sensors and landmarks using the Kalman filter for indoor localization. *Sensors (Basel, Switzerland)*, 15(1), 715–732. doi:10.3390/s150100715 PMID:25569750

- Cho, H., Song, J., Park, H., & Hwang, C. (2014). Deterministic indoor detection from dispersions of GPS satellites on the celestial sphere. In *Proceedings of the 11th International Symposium on Location-Based Services*. New York, NY: Springer.
- Deng, Z. A., Hu, Y., Yu, J., & Na, Z. (2015). Extended Kalman filter for real time indoor localization by fusing WiFi and smartphone inertial sensors. *Micromachines*, 6(4), 523–543. doi:10.3390/mi6040523
- Deng, Z. A., Wang, G., Hu, Y., & Wu, D. (2015). Heading estimation for indoor pedestrian navigation using a smartphone in the pocket. *Sensors (Basel, Switzerland)*, 15(9), 21518–21536. doi:10.3390/s150921518 PMID:26343679
- Faragher, R., & Harle, R. (2013). SmartSLAM- an efficient smartphone indoor positioning system exploiting machine learning and opportunistic sensing. In *Proceedings of the 26th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS+ 2013)* (pp. 1-14). Manassas, VA: The Institute of Navigation.
- Gallagher, T., Li, B., Dempster, A. G., & Rizos, C. (2011). Power efficient indoor/outdoor positioning handover. *Proceedings of the 2011 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*.
- He, S., Chan, S. H. G., Yu, L., & Liu, N. (2015). Calibration-free fusion of step counter and wireless fingerprints for indoor localization. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (pp. 897-908). New York, NY: ACM. doi:10.1145/2750858.2804254
- Herrmann, R., Zappi, P., & Rosing, T. S. (2012). Context aware power management of mobile systems for sensing applications. In *Proceedings of 2012 International Conference on Information Processing in Sensor Networks*. New York, NY: ACM.
- Hilsenbeck, S., Bobkov, D., Schroth, G., Huitl, R., & Steinbach, E. (2014). Graph-based data fusion of pedometer and WiFi measurements for mobile indoor positioning. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (pp. 147-158). New York, NY: ACM. doi:10.1145/2632048.2636079
- Hoseini-Tabatabaei, S. A., Gluhak, A., & Tafazolli, R. (2013). A survey on smartphone-based systems for opportunistic user context recognition. *ACM Computing Surveys*, 45(3), 27. doi:10.1145/2480741.2480744
- Jiang, Y., Xiang, Y., Pan, X., Li, K., Lv, Q., Dick, R. P., & Hannigan, M. et al. (2013). Hallway based automatic indoor floorplan construction using room fingerprints. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (pp. 315-324). New York, NY: ACM. doi:10.1145/2493432.2493470
- Li, F., Zhao, C., Ding, G., Gong, J., Liu, C., & Zhao, F. (2012). A reliable and accurate indoor localization method using phone inertial sensors. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing* (pp. 421-430). New York, NY: ACM. doi:10.1145/2370216.2370280
- Li, K., Bigham, J., Bodanese, E. L., & Tokarchuk, L. (2013). Location estimation in large indoor multi-floor buildings using hybrid networks. In *Proceedings of the 2013 IEEE Wireless Communications and Networking Conference (WCNC)* (pp. 2137-2142). New York, NY: IEEE. doi:10.1109/WCNC.2013.6554893
- Li, Y., Zhuang, Y., Lan, H., Zhang, P., Niu, X., & El-Sheimy, N. (2015). WiFi-aided magnetic matching for indoor navigation with consumer portable devices. *Micromachines*, 6(6), 747–764. doi:10.3390/mi6060747

- Luo, C., Hong, H., & Chan, M. C. (2014). PiLoc: a self-calibrating participatory indoor localization system. In *Proceedings of the 13th International Symposium on Information Processing in Sensor Networks (IPSN-14)* (pp. 143-153). New York, NY: IEEE. doi:10.1109/IPSN.2014.6846748
- Mirowski, P., Ho, T. K., Yi, S., & MacDonald, M. (2013). SignalSLAM: Simultaneous localization and mapping with mixed WiFi, Bluetooth, LTE and magnetic signals. In *Proceedings of the 2013 International Conference on Indoor Positioning and Indoor Navigation (IPIN)* (pp. 1-10). New York, NY: IEEE. doi:10.1109/IPIN.2013.6817853
- Moghtadaiee, V., & Dempster, A. G. (2014). Indoor location fingerprinting using FM radio signals. *IEEE Transactions on Broadcasting*, 60(2), 336-346. doi:10.1109/TBC.2014.2322771
- Mohssen, N., Momtaz, R., Aly, H., & Youssef, M. (2014). It's the human that matters: Accurate user orientation estimation for mobile computing applications. In *Proceedings of the 11th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services* (pp. 70-79). Gent, Belgium: ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering). doi:10.4108/icst.mobiquitous.2014.257920
- Nakamura, Y., Ono, M., Sekiya, M., Honda, K., & Takahashi, O. (2015). Indoor/outdoor determination method using various sensors for the power saving of terminals in Geo-fencing. *Proceedings of the 2015 International Workshop on Informatics*.
- Parviainen, J., Bojja, J., Collin, J., Leppänen, J., & Eronen, A. (2014). Adaptive activity and environment recognition for mobile phones. *Sensors (Basel, Switzerland)*, 14(11), 20753-20778. doi:10.3390/s141120753 PMID:25372620
- Pei, L., Guinness, R., Chen, R., Liu, J., Kuusniemi, H., Chen, Y., & Kaistinen, J. et al. (2013). Human behavior cognition using smartphone sensors. *Sensors (Basel, Switzerland)*, 13(2), 1402-1424. doi:10.3390/s130201402 PMID:23348030
- Qian, J., Pei, L., Ma, J., Ying, R., & Liu, P. (2015). Vector graph assisted pedestrian dead reckoning using an unconstrained smartphone. *Sensors (Basel, Switzerland)*, 15(3), 5032-5057. doi:10.3390/s150305032 PMID:25738763
- Radu, V., Katsikouli, P., Sarkar, R., & Marina, M. K. (2014). A semi-supervised learning approach for robust indoor-outdoor detection with smartphones. In *Proceedings of the 12th ACM Conference on Embedded Network Sensor Systems* (pp. 280-294). New York, NY: ACM. doi:10.1145/2668332.2668347
- Ravindranath, L., Newport, C., Balakrishnan, H., & Madden, S. (2011). Improving wireless network performance using sensor hints. In *Proceedings of the 8th USENIX Conference on Networked Systems Design and Implementation* (pp. 21-21). Berkeley, CA: USENIX Association.
- Sánchez-Rodríguez, D., Hernández-Morera, P., Quinteiro, J. M., & Alonso-González, I. (2015). A low complexity system based on multiple weighted decision trees for indoor localization. *Sensors (Basel, Switzerland)*, 15(6), 14809-14829. doi:10.3390/s150614809 PMID:26110413
- Shang, J., Gu, F., Hu, X., & Kealy, A. (2015). APFiLoc: An infrastructure-free indoor localization method fusing smartphone inertial sensors, landmarks and map information. *Sensors (Basel, Switzerland)*, 15(10), 27251-27272. doi:10.3390/s151027251 PMID:26516858
- Shang, J., Hu, X., Gu, F., Wang, D., & Yu, S. (2015). *Improvement schemes for indoor mobile location estimation: a survey*. *Mathematical Problems in Engineering 2015*. New York, NY: Hindawi Publishing Corporation.
- Shih, W. Y., & Lan, K. C. (2012). Using smartphone with un-scaled map for indoor localization. *Proceedings of the IEEE 9th Annual Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON)*, 18-21.

- Shu, Y., Bo, C., Guobin, S., Zhao, C., Li, L., & Zhao, F. (2015). Magicol: Indoor localization using pervasive magnetic field and opportunistic WiFi sensing. *IEEE Journal on Selected Areas in Communications*, 33(7), 1443–1457. doi:10.1109/JSAC.2015.2430274
- Soria Morillo, L. M., Ortega Ramírez, J. A., Alvarez García, J. A., & Gonzalez-Abril, L. (2012). Outdoor exit detection using combined techniques to increase GPS efficiency. *Expert Systems with Applications*, 39(15), 12260–12267. doi:10.1016/j.eswa.2012.04.047
- Subbu, K., Zhang, C., Luo, J., & Vasilakos, A. (2014). Analysis and status quo of smartphone-based indoor localization systems. *Wireless Communications*, 21(4), 106–112. doi:10.1109/MWC.2014.6882302
- Wu, C., Yang, Z., & Liu, Y. (2015). Smartphones based crowdsourcing for indoor localization. *IEEE Transactions on Mobile Computing*, 14(2), 444–457. doi:10.1109/TMC.2014.2320254
- Xiao, Z., Wen, H., Markham, A., & Trigoni, N. (2014, April). Lightweight map matching for indoor localisation using conditional random fields. In *Proceedings of the 13th International Symposium on Information Processing in Sensor Networks, IPSN-14* (pp. 131-142). New York, NY: IEEE. doi:10.1109/IPSN.2014.6846747
- Yang, Z., Wu, C., Zhou, Z., Zhang, X., Wang, X., & Liu, Y. (2015). Mobility increases localizability: A survey on wireless indoor localization using inertial sensors. *ACM Computing Surveys*, 47(3), 54. doi:10.1145/2676430
- Yoon, S., Lee, K., & Rhee, I. (2013, June). FM-based indoor localization via automatic fingerprint DB construction and matching. In *Proceeding of the 11th Annual International Conference on Mobile Systems, Applications, and Services* (pp. 207-220). New York, NY: ACM. doi:10.1145/2462456.2464445
- Zeinalipour-Yazti, D., Laoudias, C., Georgiou, K., & Chatzimilioudis, G. (2016). *Internet-based indoor navigation services*. *IEEE Internet Computing*. New York, NY: IEEE.
- Zhang, C., Luo, J., & Wu, J. (2014, May). A dual-sensor enabled indoor localization system with crowdsensing spot survey. In *Proceedings of the 2014 IEEE International Conference on Distributed Computing in Sensor Systems (DCOSS)* (pp. 75-82). New York, NY: IEEE. doi:10.1109/DCOSS.2014.14
- Zhang, C., Subbu, K. P., Luo, J., & Wu, J. (2015). GROPING: Geomagnetism and crowdsensing powered indoor navigation. *IEEE Transactions on Mobile Computing*, 14(2), 387–400. doi:10.1109/TMC.2014.2319824
- Zhou, B., Li, Q., Mao, Q., Tu, W., & Zhang, X. (2015). Activity sequence-based indoor pedestrian localization using smartphones. *IEEE Transactions on Human-Machine Systems*, 45(5), 562–574. doi:10.1109/THMS.2014.2368092
- Zhou, M., Zhang, Q., Xu, K., Tian, Z., Wang, Y., & He, W. (2015). PRIMAL: Page rank-based indoor mapping and localization using gene-sequenced unlabeled WLAN received signal strength. *Sensors (Basel, Switzerland)*, 15(10), 24791–24817. doi:10.3390/s151024791 PMID:26404274
- Zhou, P., Zheng, Y., Li, Z., Li, M., & Shen, G. (2012). IODetector: a generic service for indoor outdoor detection. In *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems* (pp. 113-126). New York, NY: ACM. doi:10.1145/2426656.2426668
- Zou, H., Jiang, H., Luo, Y., Zhu, J., Lu, X., & Xie, L. (2016). BlueDetect: An iBeacon-enabled scheme for accurate and energy-efficient indoor-outdoor detection and seamless location-based service. *Sensors (Basel, Switzerland)*, 16(2), 268. doi:10.3390/s16020268 PMID:26907295

ADDITIONAL READING

Gu, Y., Lo, A., & Niemegeers, I. (2009). A survey of indoor positioning systems for wireless personal networks. *IEEE Communications Surveys and Tutorials*, 11(1), 13–32. doi:10.1109/SURV.2009.090103

Khan, W. Z., Xiang, Y., Aalsalem, M. Y., & Arshad, Q. (2013). Mobile phone sensing systems: A survey. *IEEE Communications Surveys and Tutorials*, 15(1), 402–427. doi:10.1109/SURV.2012.031412.00077

Krumm, J. (2016). *Ubiquitous computing fundamentals*. CRC Press.

Lane, N. D., Miluzzo, E., Lu, H., Peebles, D., Choudhury, T., & Campbell, A. T. (2010). A survey of mobile phone sensing. *IEEE Communications Magazine*, 48(9), 140–150. doi:10.1109/MCOM.2010.5560598

Macias, E., Suarez, A., & Lloret, J. (2013). Mobile sensing systems. *Sensors (Basel, Switzerland)*, 13(12), 17292–17321. doi:10.3390/s131217292 PMID:24351637

Srivastava, M., Abdelzaher, T., & Szymanski, B. (2012). Human-centric sensing. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 370(1958), 176–197.

Varshavsky, A. (2005). Are GSM phones THE solution for localization? In *Seventh IEEE Workshop on Mobile Computing Systems & Applications (WMCSA'06 Supplement)* (pp. 34–42). IEEE.

Zeimpekis, V., Giaglis, G. M., & Lekakos, G. (2002). A taxonomy of indoor and outdoor positioning techniques for mobile location services. *ACM SIGecom Exchanges*, 3(4), 19–27. doi:10.1145/844351.844355

KEY TERMS AND DEFINITIONS

M

Dead Reckoning: A localization method that consists of calculating the current location as the result of a displacement function applied to the previous location.

Fingerprinting: A localization method where the current location is determined by measuring the local characteristics of one or more signals (usually of electromagnetic nature) and comparing them with a set of values stored in a database. The database contains a map of signal values and their corresponding physical locations.

Inertial Measurement Unit (IMU): A sensing device that reacts to changes in its velocity (accelerometer) or orientation (gyroscope). When mounted on a rigid body, it allows measuring the acceleration or orientation of the body.

Landmark: A location characterized by specific values or patterns of sensor readings that can be used to uniquely identify the location during successive visits. Measured values can be specific to the local environment (e.g. electromagnetic signals), or related to how the location is accessed by the visitor (e.g. acceleration patterns observed when using a staircase).

Map Matching: The technique of matching location estimations coming from sensor readings with constraints or hints coming from a map of the environment.

Simultaneous Localization And Mapping (SLAM): A technique for localizing a person or object in an area and simultaneously building a map of the area by processing sensor reading data. Usually, the map is built and adjusted incrementally during multiple visits to the area, and localization accuracy improves during the map building process.

Step Detection: A method for automatically recognizing the walking steps of a pedestrian, usually by means of an IMU mounted on a device carried by the pedestrian.

Technological Innovation and Use in the Early Days of Camera Phone Photo Messaging

Jonathan Lillie

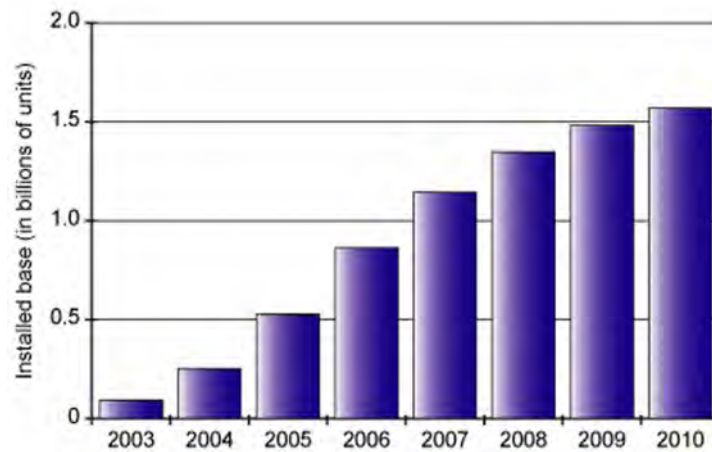
Loyola University, USA

INTRODUCTION

The introduction of picture taking and messaging capabilities to mobile phones is one of the four most significant and impactful innovations that mobile phones in general have offered to humanity in the last 25 years; the others being mobile telephony, text messaging, and mobile Internet access/services. Of course in making such a claim, one quickly runs into a morass of classification. After all, many mobile Internet behaviors are based in part on image sharing practices; and image sending piggybacks somewhat on the older texting platforms, both in terms of technology and behavior. The journey of camera phone use and image messaging – from early designs and strategies of phone manufactures to the wide diversity of practices that are observable today – is a classic example of technological innovation. And by “classic” I mean that innovation tends to never cease, especially with those technologies that are built on flexible platforms, or else where manufacturers have incentive to continually change the technology based on user innovation and practices. Such was the case when Nokia, and other handset companies, changed strategies (at least in part) after early user studies demonstrated that early camera phone adopters had different practices in mind than the companies did. Since that time, the innovation in mobile imaging practices has mostly been in the hands of user groups and social media companies.

This chapter presents a sketch of the early days of camera phone and MMS use. Since the main goal of the chapter is to review relevant literature, it does not address the contexts of camera phone use for any specific user group or region in any depth. However, rather than merely presenting the findings of the many excellent studies of camera phone users in the early and mid-2000s, I hope to also complement these studies, in part, by showing the role played by one of the main regional and global industry actors (Nokia) in shaping the technology and then responding to user trends and innovation. In the book *Everyday Innovators* (2004), which is largely devoted to research of user innovation with mobile phones, Alexandre Mallard of the Center for the Sociology of Innovation in Paris concludes that, “where users will behave creatively seems to be quite unpredictable!” (p. 42). Users, he continues, often reproduce the roles proscribed for them by industry, but they can also discover and champion new unintended uses, and enthusiast can sometimes even alter the hardware or software to essentially change what the technology can be used for. The studies reviewed in this chapter (focusing mostly on European and Asian countries) show that early camera phone users embraced the technology as a significantly enhanced form of the portable analog camera, which had a long history of cultural practice, as opposed to being more enthralled with photo messaging as industry had hoped.

Figure 1. Camera phone units as of 2005 and projection up to 2010



BACKGROUND

The Camera Phone and MMS: Industry Innovation and Marketing

The major mobile phone manufacturers started to market and sell their first camera phones in late 1999/early 2000. Just seven years later the number of camera phones surpassed the one billion mark (Figure 1). But the problem for the industry was not whether people would buy handsets, it was if people would use the new photo messaging features. Nokia along with many other European manufacturers and service providers wanted to create the next “killer app” for mobile technology, and were pushing for that to be a new standard called multimedia messaging service or MMS. Some industry players expected MMS to piggyback on the success of the SMS text-messaging platform. Over the previous decade Nokia had invested heavily in the design of SMS and the cellular platform that made it possible, the Global System for Mobile Communications (GSM). GSM was the second-generation, digital cellular system adopted by most European countries. The company’s success in the 1990s was due in part to its investment in SMS, which resulted in the technology’s seamless integration with its handsets. As

is well chronicled in the industry and academic literature, texting boomed in Europe propelled by 1990s youth culture. Though the industry did not anticipate the texting craze, it quickly responded “to capture these new uses and repackage them as fee-producing services” (Snowden, 2006, p. 112). So as the third generation (“3G”) cellular system was being developed to improve GSM with faster connections and enhanced services, Nokia (and other players) ensured that the new standard would include a photo/video messaging service commiserate with its plans for introducing camera phones by the end of the decade.

Nokia bet that new camera phone users would view MMS as SMS 2.0 (Loubser, 2005; Pohjola, O. P. and Kilkki, K., 2005), which they would not mind paying more money for due to the utility of sending and receiving multimedia content (Nokia, 1999). The company’s early market studies focused in on person-to-person and person-to-multiperson picture messaging, since these services were most like SMS and hence, they hoped, most likely to attract potential early adopters of MMS (Nokia, 1999). One of Nokia’s early user studies of MMS found that Internet downloads of pictures and other commercial media such as news alerts, travel info, and weather updates were also attractive to users and likely to be high revenue services. However,

Nokia's aggressive rollout of MMS phones in the early 2000s meant that many of these services, such as downloading commercial media from websites, were still in the development stage, while others like camera-to-camera photo messaging worked better on some service networks than on others.

One of the leading experts on the Nokia Company is Dan Steinbock of Shanghai Institute of Foreign Trade. In *The Nokia Revolution* (2001) Steinbock explains how Nokia also emerged as the global leader in mobile phone production in the 1990s because of its focus on innovation in marketing. When SMS-enabled phones started rolling out, the company applied Geoffrey Moore's theories of technology-adoption life cycles to new product designs and marketing campaigns aimed at targeted lifestyle groups such as "'poseurs,' 'trendsetters,' 'social contact seekers,' and 'high-fliers'" (Steinbock, 2001, p. 279). New downstream marketing innovations involved ad campaigns focused on these groups, artistic removable faceplates, monthly rollouts of new models, and new multi-model product categories. Nokia unveiled its first camera phone in 1999, which was a high-end model designed for business users. The company continued to help to develop the MMS technical standard while at the same time releasing mid-market camera phones like the 3650 model, its first camera phone sold in the United States in 2003. Having as many camera phones on the market as possible was essential if MMS was to take off on a large scale like SMS did. Commercials on TV and the company's websites depicted new MMS users discovering fun and creative uses of picture messaging with the latest handset models (see Lillie (2012) for analysis of Nokia's advertising during this period). Nokia had 28 percent of the global market by 2007, making it the largest manufacturer of camera phones (Taylor, 2007). By the end of the decade almost all Nokia models were equipped with cameras and MMS.

EARLY CAMERA PHONE USER STUDIES

A Nokia press release from 2003 illustrates the company's enthusiasm for MMS:

The depth of demand for MMS services is extremely good news for the industry ... Consumers really want MMS and, as operators are already discovering, people are as intrigued by and eager to use MMS as SMS.

However surveys of early use show that consumers were not as hyped for the new service. In the UK only 27 percent of camera phone owners had ever used MMS by 2003 (Continental Research, 2004). Across the Atlantic in the US, potential users were actually less interested in MMS in 2004 than in 2003 (Zelos Group, 2004). A mid-year 2006 US survey of camera phone imaging practices revealed phone-display picture viewing to be by far the most common use (72%) followed by saving pictures to a computer (33%), emailing pictures (31%), sent via MMS (28%), printed (13%), shared online (12%), and none of the above (17%) (Lyra Research, 2006). Studies in Asian countries also found that early uses tended to focus on the camera features as opposed to photo-sending. In Japan the first camera-embedded phone came out in 2000, but only four years later they were already a majority of all mobile phones (Okabe and Ito, 2006). Surveys of camera phone use in those early years showed that almost all users showed photos on the phone to others, while 50 percent reported "emailing" photos to friends and family (japan.internet.com, 2003). Just having an easy-to-use digital camera that was also handy for mobile telephony appeared to be what users were at first attracted to in the camera phone. Early on that simple but culturally and cognitively powerful add-on was the real killer app. Not MMS.

The early ethnographic studies of camera phone use confirmed that MMS was being used by some people for image sharing, mobile-to-mobile, but also showed that other uses of built-in cameras were far more common. One of the most frequently cited works from this early period is by Dr. Barbara Scifo (2005) of the University of Milan who applied focus groups and in-depth interviews to explore how teenagers and young adults in Italy were using camera phones. Based on the study, Scifo argued that the social uses of camera phone mostly fall into two categories. The first category is face-to-face interaction where fast viewing of saved photos added to conversation and the sharing of personal experience. The second category included the various uses of MMS. She argues that,

notwithstanding the [low] number of MMS messages that are being exchanged at present between those starting to use this new process of communication, there are already clear signals concerning the spread of a precise culture of communication and the birth of a new language (p. 5).

Three functional types of picture messaging were identified by Scifo. The first included MMS messages employed to induce an action by someone else. For example, I might send a photo – of myself with a pint at a local pub – to a friend as an invitation to join me. Informative messages, on the other hand, might show off my current vacation location or the first flower of spring in my front yard. Lastly, problem-solving messages, are used to reduce effort, costs, or solve immediate problems, such as taking a picture of the jacket I want for my birthday and sending it to my wife with the caption, “a not so subtle hint.” Scifo found that MMS messages are typically limited to conversations among friends and family, which is supported other studies in European and Asian countries (Kindberg, Spasojevic Fleck and Sellen, 2004; Okabe and Ito, 2006). “The camera-phone represents a new form of extension of one’s experience and memory,” Scifo concludes. “On the

socialization level, it represents a new communication resource that can be invested in one’s peer group and in one’s relationships” (p. 3).

In their review of the MMS-focused literature Kindberg, Spasojevic Fleck and Sellen note that the early studies sponsored by industry leaders, such as Hewlett-Packard and Nokia, observed how participants sent images “to increase or maintain group cohesion, express affection, support conversation, and tell stories” (2004, p. 1). A similar study by Battarbee and Kurvinen (2003) considered how users can develop creative personal styles in the way they use MMS, particularly in terms of the types of images sent and how these are used to communicate with friends and family. The authors note that although user innovation and experimentation is important in such creative communication, the designs of the handsets and services themselves are also vital.

“Technology enables these experiences, it defines what can and cannot be done,” the authors conclude. “Technology determines how simple and easy it is for people to co-presently collaborate in its use, and how in new situations, users negotiate new uses and meanings” (p. 7).

Most of the early camera phone user studies were conducted in European and Asian countries where – for a variety of cultural and commercial reasons – camera phones (and mobile phones in general) were adopted earlier and at a faster rate than in North America and other areas of the world. For studies in European countries see the work of experts such as Dr. Barbara Scifo in Italy (2005, 2009; Colombo and Scifo, 2005); Dr. Leslie Haddon in the UK (2005, 2007) of the London School of Economics; and in Finland Dr. Virpi Oksman (2005a, 2005b, 2006) of Tampere University of Technology and Dr. Ilpo Koskinen (2006, 2008) of the University of Helsinki. Researchers were also examining the uses of camera phones in the emerging Asian information societies of Japan and South Korea. One of the best book-length treatments from this time is Dr. Mizuku Ito, Dr. Misa Matsuda and Dr. Daisuke Okabe’s edited work *Personal, Portable, Pedestrian: Mobile Phones*

in Japanese Life (2005). In the introduction Dr. Mizuko Ito of the University of California at Irvine writes that the *keitai*, the nominal term for mobile phone in Japan, has become an organizing and socializing force in outside and routine settings for Japanese people vis-à-vis the more hypermediated indoor spaces of video and computer monitors.

In contrast to the cellular phone of the United States (defined by technical infrastructure), and the mobile phone of the United Kingdom (defined by the untethering of fixed location), the Japanese term keitai (roughly translated, "something you carry with you") references a somewhat different set of dimensions. A keitai is not so much about a new technical capability or freedom of motion but about a snug and intimate technosocial tethering, a personal device supporting communications that are a constant, lightweight, and mundane presence in everyday life. (p. 1, author's emphases)

In their ethnographic study of camera phone use in Japan, Ito and her colleague Dr. Daisuke Okabe of Keio University (2005) find the device's principle use as a networked portable camera, where photos can be 'stored,' 'shared,' and 'saved.' But the camera phone has some more immediate uses than its analog, camera-only predecessor since photos could be stored and then instantly displayed, i.e., "shared," with others in the same location or shared by posting to virtual communities. Most phones also came with cords to connect to computers so photos could be saved in personal archives and printed.

For studies of camera phone use in South Korea see Haddon and Kim, 2007; Hjorth, 2007; and Ok, 2005. Dr. Larissa Hjorth (2007) of RMIT University places Korean camera phone practices in the context of a country where media technology innovation and use is a source of national pride. "By scripting itself as a centre for digital, convergent technologies," she writes, "Korea is rewriting its history through the global currency of ICT production" (p. 228) In her ethnographic study of how a group of youth used camera phones

Hjorth found the content categories of photos taken to be somewhat similar to those reported by Scifo in Italy: "everyday, special occasions/places, friendship/family, self-portraiture, and favourites" (Hjorth, 2007, p. 234). She notes that for many of the study participants the photo taking, posting, ringtones, and wallpaper choices revolved around their involvement in romantic relationships. "For the youth of Korea, where most still live at home before getting married, these third spaces," between home and work or school, she observed, "operate as spaces to connect with other people" (p. 230).

The young adults studied by Hjorth rarely used MMS to send picture messages. Instead they posted photos to a virtual community, "Cyworld's mini-hompy." Hjorth interprets people's preference for using mini-hompy rather than MMS as due to "a broader shift in web 2.0 towards the user's choice of context informing the content and distribution of their images in a social network" (p. 236). In a 2005 study of youth using the mini-hompy community Hjorth and Kim conclude that "virtual connecting was always about the need and desires to be connected at various levels and never about substituting for face-to-face contact" (Hjorth, 2007, p. 230). Although this small study cannot be generalized, it offers another possible reason why MMS use for picture messaging did not quickly take off in many countries, since company's like Nokia believed that one of the most popular uses of MMS would be as a substitute for face-to-face communication (Lillie, 2012).

Some more recent studies have considered how the use of mobile phones as a whole, including imaging functions, is caught up in identity development and self-expression. The studies and edited anthologies by Dr. James Katz of Boston University are well known in this area. Katz and Sugiyama (2006), who surveyed students in the USA and Japan, concluded that, "young people use the mobile phone as a way of expressing their sense of self and perceive others through a 'fashion' lens" (p. 281). James Katz's edited book *Handbook of Mobile Communication Studies* (2008) is one

of the best collections in this field due to the large variety of country studies it includes, each focusing on the specifics of mobile phone use in one national or sub-national context. Another excellent book-length work is *Moving Cultures: Mobile Communication in Everyday Life* (2007) by Dr. Letizia Caronia of the University of Bologna and Dr. Andre Caron of the University of Montreal. One of their studies focused on how teens used mobiles to explore processes of identity making through public performance. Mobile performances, they argue, “work as a social grammar through which people are supposed to define themselves and those around them” (p. 28). In fact, long before the advent of camera phones, personal cameras have served a similar role. Social and personal practices of picture taking and viewing helps people to understand their relation to the world, not just the world itself. As one noted scholar put it, “photography, more than merely representing, has taught us a way of seeing...that...has transformed contemporary self-knowledge” (Lury, 1998, p. 3). Media archeologist Erkki Huhtamo (2004) astutely observed that the mobile phone was certainly not the first mobile media device, being predated by the Walkman and more importantly for this topic, the portable camera. As Ito and Okabe (2005) and other researchers have found, the popularity of camera phones are due mostly to their advances of the standard portable camera, which has been a common technology of families in developed countries for many decades.

Nokia of course was aware of the slow uptake of MMS and conducted its own research to find a solution to this problem. “Apparently, usage has not emerged from merely experimenting with multimedia messages to every day usage, as has happened with SMS,” suggested one Nokia research paper (Pohjola and Kilkki, 2005, p. 1). In these early years, manufacturers and carriers were also trying to fix problems of interoperability for MMS. This continued to be an issue because there was a wide disparity between the technical capabilities of different handsets and different network providers. An MMS sent from a person

using one network carrier might not be properly handled or decoded by a different carrier. Or the receiver’s handset might not be able to download or display some or all of the message. Lacking the ability to quickly fix problems such as handset usability, cellular coverage, or MMS message interoperability, Nokia addressed the lack of user interest in person-to-person picture messaging in 2003/4 by trying to better understand user experiences of MMS and then using this knowledge to design new marketing campaigns for its latest camera phone models. One of Nokia’s studies during this period was aptly titled, “What is wrong with Multimedia Messaging?” The study provided an end-user cost-benefit analysis based on the behaviorist assumption that, “End-users have various needs and wants that they can fulfill using an MMS service” (Pohjola and Kilkki, 2005, p. 1) and offers a conceptual model of how user experiences influence behaviors involved in a person-to-person MMS. The report notes that advertising can demonstrate the ideal user experience, but negative experiences of poor network coverage, high per-message costs, or confusing handset interfaces will unravel any gains made through marketing. Nokia’s subsequent advertising applied this model by demonstrating how people learn about and experience MMS and showing how this experience can fulfill their needs (Lillie, 2012).

M

MOBLOGGING AS MMS INNOVATION

It was around 2005 that the use of MMS to upload content to image-sharing websites began to garner more attention from users, service providers, and cultural critics. In particular, the practice of “moblogging” began to emerge among photography enthusiasts and programmers who realized that MMS was well suited for users to upload photos and videos from handsets to websites, since MMS media is already stored on a server before it is delivered when someone views the message on their phone (Loubser, 2005, p. 4). Moblogging was

pioneered by amateur innovators experimenting with SMS, WAP (the first 2G and 3G standard for connecting to mobile-optimized websites), and eventually MMS to post messages and eventually photos to their modified blogs (Goggin, 2005, p. 5). Dr. Gerard Goggin (2005, 2006, 2008) of the University of Sydney is one of the first researchers to study mobile blogs. He found that these early innovators have various motivations for posting photos to their blogs (2005). Some were happy to develop a new way to share aspects of their daily lives with friends and the public, while others hoped this new web-publishing innovation could offer more socio-political reach. Goggin has single authored or edited several noted titles in this area including *Cell Phone Culture* (2006), *Mobile Media* (2007) and *Global Mobile Media* (2011).

Industry took note of moblog innovation and began to develop online image services to go beyond simple photo archiving to offer, “more playful community activities, involving guessing and mocking games, and the introduction of tagging” (Nightengale, 2007, p. 290). To encourage users to upload images and engage in various web-based services and communities, manufacturers like Nokia started introducing handsets with non-removable image memory to replace early models that featured removable memory cards (Nightengale, p. 291). Goggin recounts how Nokia was the first industry player in Europe to create commercial moblog-like services for their customers, starting in 2005 with the launch of “LifeBlog” (2006, p. 159). Nokia’s Lifeblog was one of the company’s first efforts to encourage user practices that utilize MMS to upload photos to websites. Taking advantage of the way MMS already worked, any picture that had been sent phone-to-phone could be selected by the user for display on her Lifeblog page. In 2007 the company rolled its moblogging offerings into a more comprehensive suite of services dubbed the “Ovi Portal.” Ovi included the Nokia Music Store, Nokia Maps, mobile games, moblogs. Many of these services used MMS for uploading and downloading of content to and from handsets.

But 2007 is better known in the computer world for being the year Apple began selling its touch screen smartphone, the iPhone. Nokia had been an innovator in the smartphone category years earlier with its successful 9000 series. Although the company had the foresight to make all mobile handsets camera phones, it continued to envision smartphones as a niche business-class product. Smartphones meant users did not need MMS to add photos to their blogs, Flickr albums, or Facebook pages. Facebook began allowing photo uploads as early as 2005 and was an early innovator in using mobile web standards and developing mobile apps for emerging mobile operating systems like iOS and Android.

CONCLUSION

Of course eventually there were new killer apps in the mobile phone world to follow up on the success of SMS texting. But it was not traditional wireless carriers and handset manufacturers who were the hardware and software innovators. It was computer industry giants like Apple who championed smartphones, devices that were almost always connected to the internet and therefore allowed users to share camera-phone captured media and socialize via social media web sites that became killer apps of their own. MMS is still used today, especially in areas where smart phones are too expensive for the average user. But the face of photo sharing today is Instagram, Snapchat, Facebook, Flickr, Tumblr, and Twitter. The type of creative visual communication that Nokia envisioned and that studies such as Ito and Okabe (2006) and Scifo (2005) found practiced among some early users of MMS, is now commonly seen as conversations on Facebook through both “News Feeds” and private messaging. And even the goliath Facebook is losing its popularity among younger users who prefer to use apps like Instagram and Snapchat for photo sharing (CITE). Use of MMS in the early and mid 2000s might have even primed some users for how they could

communicate via smart phones with social media apps. Hjorth's finding that South Korean youth preferred posting photos to virtual communities rather than using MMS perhaps anticipated the possibility that sharing media predicated on social peer-group communication would eventually eclipse what Nokia and others hoped MMS would be.

A large portion of today's visual culture is created by industry, and it is important not to downplay the roll that marketing processes of transnational corporations like Nokia play in deciding which technologies are available, how people make meaning through them, and how people use them. Although mobile phone advertisements have rarely been studied, they are integral in introducing new media technologies to users. Two scholars who study the history and cultural implications of marketing, John Harms and Douglas Kellner (2003), argue that advertising has "multi-faceted social functions ranging from short range efforts to induce individuals to buy specific products to more long range functions that attempt to sell consumer capitalism as a way of life." However, as we see in the example of Nokia's heavy marketing of MMS, advertising can only give users suggestions for how to use technology. The use of the camera phone as a discrete, portable, digital camera with easy display of archived pictures as it turns out was the more intuitive and useful feature, following up on decades of amateur photography practices familiar to most people in developed countries. Writing about the cultural significance of amateur photography in the 20th century Susan Sontag observed in 1977 that, "[t]o photograph is to collect the world. It means putting oneself into a certain relation to the world that feels like knowledge – and, therefore, like power" (1977, p. 3-4). But as a cultural practice amateur photography rarely reached into the public sphere – except when appropriated by commercial media – before web-based visual practices starting with homepages and moblogs and now via social media. The so called 'social media revolution' would probably not be what it is today without the industry and

user innovations of camera phones and messaging systems like MMS.

All of this effort toward product development and marketing obviously has some impact, though certainly not total control, over how people use, identify, and make meaning of, and with, new commercial ICTs. As Snowden's study (2007) of SMS showed, texting itself was not some completely unexpected use of the mobile phone. Policy makers, engineers, and many companies such as Nokia worked very hard to bring SMS technology to the market. What was unexpected was how texting use boomed due to its massive uptake by teenagers and young adults as a form of everyday peer-group socialization. But no amount of advertising was able to entice people to use MMS picture messaging at the same frequency as they did with SMS. Thus the cultural life of a media technology traverses a complex web of user identities and appropriations, social meanings, institutional (often national) discourses, and industry design rooms.

FUTURE RESEARCH DIRECTIONS

Noted cultural theorist W.J.T. Mitchell (1994) describes a pivotal "pictorial turn" of contemporary culture in the 20th century: "a postlinguistic, post-semiotic rediscovery of the picture as a complex interplay between visibility, apparatus, institutions, discourse, bodies, and figurality" (p. 16). Certainly the camera phone seems to fit perfectly in an era of proliferating visual media, and in particular as part of the personal-to-public digital media revolution that started off with personal homepages in the early 1990s. However, it is important to note that before the camera phone, the mobile phone was still mostly an interpersonal communication device. And that is how MMS was intended to be used by Nokia and others. Today that is still what the standard is still largely used for, especially by those who do not have smartphones, which can directly use web-based services like Facebook and Twitter. And while camera phones might just seem

like a logical, intuitive step in the incessant march of a postmodern, visual-dominated culture, it is vital to remember that it is people as individuals, communities, corporations (and other groupings) who create that culture and its technologies; and therefore to study the cultural aspects of media technologies we must analyze as many of those groups (and their contexts) as possible.

REFERENCES

- Caronia, L., & Caron, A. H. (2004). Constructing a specific culture: Young peoples use of the mobile phone as a social performance. *Convergence (London)*, 10(2), 28–61. doi:10.1177/135485650401000204
- Caronia, L., & Caron, A. H. (2007). *Moving cultures: Mobile communication in everyday life*. Montréal: McGill-Queen's University Press.
- Continental Research. (2004). *Autumn mobile phone report*. London: Continental Research.
- David, G. (2010). Camera phone images, videos and live streaming: A contemporary visual trend. *Visual Studies*, 25(1), 89–98. doi:10.1080/14725861003607017
- Goggin, G. (2005). "Have fun and change the world": Moblogging, mobile phone culture and the Internet. Presented at BlogTalk Downunder, Sydney, Australia.
- Goggin, G. (2006). On mobile photography: Camera phones, moblogging, and new visual cultures. In G. Goggin (Ed.), *Cell phone culture: Mobile technology in everyday life* (pp. 143–161). London: Routledge.
- Goggin, G. (2008). Cultural studies of mobile communication. In J. Katz (Ed.), *Handbook of Mobile Communication Studies* (pp. 353–366). Cambridge, MA: MIT Press. doi:10.7551/mitpress/9780262113120.003.0026
- Goggin, G. (2011). *Global mobile media*. New York: Routledge.
- Gye, L. (2007). Picture this: The impact of mobile camera phones on personal photographic practices. *Continuum (Perth)*, 21(2), 279–288. doi:10.1080/10304310701269107
- Haddon, L., & Kim, S. D. (2007). Mobile phones and web-based social networking: Emerging practices in Korea with Cyworld. *The Journal of the Communications Network*, 6(1).
- Hjorth, L. (2007). Snapshots of almost contact: The rise of camera phone practices and a case study of Seoul, Korea. *Continuum (Perth)*, 21(2), 227–238. doi:10.1080/10304310701278140
- Huhtamo, E. (2004). *An Archaeology of Mobile Media*. Keynote Address, 12th ISEA Conference. Helsinki, Finland.
- Ito, M. (2006). Japanese media mixes and amateur cultural exchange. In D. Buckingham & W. Willet (Eds.), *Digital Generations*. Mahwah, NJ: Lawrence Erlbaum.
- Ito, M., & Okabe, D. (2005). *Intimate visual co-presence*. Paper presented at UbiComp, Tokyo, Japan. Retrieved from <http://www.itofisher.com/mito/>
- Ito, M., & Okabe, D. (2006). Everyday contexts of camera phone use: Steps towards technosocial ethnographic frameworks. In J. Höfllich & M. Hartmann (Eds.), *Mobile Communication in Everyday Life: An Ethnographic View*. Frank and Timme.
- Japan.internet.com. (2003). *65% of camera phone users report they use the camera function*. Retrieved from <http://japan.internet.com/research/20030602/1.html>
- Kindberg, T., Spasojevic, M., Fleck, R., & Sellen, A. (2004). How and why people use camera phones. *HP Consumer Applications and Systems Laboratory*. Available on: <http://www.hpl.hp.com/techreports/2004/HPL-2004-216.pdf>

Lee, D. (2006). Women's creation of camera phone culture. *Fibreculture*, 6. Retrieved from http://journal.fibreculture.org:issue6:issue6_donghoo.html

Lillie, J. (2012). Nokias MMS: A cultural analysis of mobile picture messaging. *New Media & Society*, 14(1), 80–97. doi:10.1177/1461444811410400

Loubser, M. (2005). User created content with MMS. *KTH School of Information and Communication Technology*. Available on: http://www.ict.kth.se/courses/IK2555/ExamplePapers//2G1330_Max_Loubser_MMS-20050731.pdf

Lury, C. (1998). *Prosthetic Culture: Photography, Memory, and Identity*. New York: Routledge. doi:10.4324/9780203425251

Mallard, A. (2006). Following the emergence of unpredictable uses? New stakes and tasks in scientific understanding of ICT uses. In L. Haddon, E. Mante, B. Sapio, K.H. Kommonen, L. Fortunati, & A. Kant (Eds.), *Everyday Innovators, Researching the Role of Users in Shaping ICTs*. Springer.

Nightingale, V. (2007). The cameraphone and online image sharing. *Continuum: Journal of Media & Culture*, 21(2), 289–301. doi:10.1080/10304310701278157

Nokia. (1999). *Early demand for Nokia multimedia message service: Market study of MMS*. Nokia. Available on: http://nds2.ir.nokia.com/NOKIA_COM_1/About_Nokia/Press/White_Papers/pdf_files/study_MMS.pdf

Ok, H. (2005). Cinema in you hand, cinema in the street: The aesthetics of convergence in Korean mobile (phone) cinema. *Proceedings of the conference 'Seeing, Understanding, Learning in the Mobile Age'*.

Okabe, D., & Ito, M. (2006). Everyday contexts of cameraphone use: Steps toward technosocial ethnographic frameworks. In J. Höfflich & M. Hartmann (Eds.), *Mobile Communication in Everyday Life: Ethnographic Views, Observations and Reflections*. Berlin: Frank and Timme.

Pohjola, O. P., & Kilkki, K. (2005). What is wrong in multimedia messaging? *Conference proceedings of the Sixth IEEE International Symposium*. doi:10.1109/WOWMOM.2005.110

Poropudas, T. (2002). Nokia spearheads MMS alone in CeBIT. *MobileMonday*. Available on: <http://www.mobilemonday.net/news/nokia-spearheads-mms-alone-in-cebit>

Scifo, B. (2005). The domestication of camera-phone and MMS communication: The early experiences of young Italians. In K. Nyíri (Ed.), *A Sense of Place: The Global and the Local in Mobile Communication* (pp. 363–374). Vienna: Passagen Verlag.

Snowden, C. (2006). Cstng a pwr4l spll: D evolshn f sms. In A. P. Kavoori & N. Arceneaux (Eds.), *The Cell Phone Reader: Essays in Social Transformation* (pp. 107–124). New York: Peter Lang Publishers.

Sontag, S. (1977). *On Photography*. New York: Farrar, Straus and Giroux.

Steinbock, D. (2001). *The Nokia Revolution: The Story of an Extraordinary Company that Transformed an Industry*. New York: American Management Association.

Taylor, C. (2007). 1/3 of world population to have camera phones by 2011, firm says. *Electronic News*, 53(27), 3.

Zelos Group. (2004). *US wireless services user survey: Females and youth drive interactive services*. Available on: <http://www.researchandmarkets.com/reports/74536>

ADDITIONAL READING

Caronia, L., & Caron, A. H. (2007). *Moving cultures: Mobile communication in everyday life*. Montréal: McGill-Queen's University Press.

David, G. (2010). Camera phone images, videos and live streaming: A contemporary visual trend. *Visual Studies*, 25(1), 89–98. doi:10.1080/14725861003607017

Goggin, G. (Ed.). (2006). *Cell phone culture: Mobile technology in everyday life*. London: Routledge.

Hjorth, L. (2007). Snapshots of almost contact: The rise of camera phone practices and a case study of Seoul, Korea. *Continuum (Perth)*, 21(2), 227–238. doi:10.1080/10304310701278140

Katz, J. (Ed.). (2008). *Handbook of Mobile Communication Studies*. Cambridge, Mass.: MIT Press. doi:10.7551/mitpress/9780262113120.001.0001

Nightingale, V. (2007). The cameraphone and online image sharing. *Continuum: Journal of Media & Culture*, 21(2), 289–301. doi:10.1080/10304310701278157

Snowden, C. (2006). Cstng a pwr4l spll: D evolshn f sms. In A. P. Kavoori & N. Arceneaux (Eds.), *The Cell Phone Reader: Essays in Social Transformation* (pp. 107–124). New York: Peter Lang Publishers.

Sontag, S. (1977). *On Photography*. New York: Farrar, Straus and Giroux.

KEY TERMS AND DEFINITIONS

Camera Phone: Any mobile phone with a built-in camera.

Cell Phone: A popular term used in many countries for mobile phones, arising out of the longer-term ‘cellular phone,’ indicating the devices’ uses of cellular radio systems for voice communication.

Cultural Studies: A mostly academic, scholarly movement in the humanities that argues for multi-disciplinary and intra-disciplinary approaches to studying cultural groups, with particular focus often on media use.

Mobile Phone Culture: A term used to describe how individuals and groups use mobile technology in their everyday lives.

Multimedia Messaging Service (MMS): A mobile phone service that allows the sending and receiving of messages combining text, photos, and short video clips.

New Media Studies: The study of how and why people use new media technologies, including a range of approaches in the social science and humanities.

Technological Innovation: A variety of scholars from different disciplinary backgrounds have studied how people and companies create new uses for new or existing technologies, and how these new innovations diffuse within populations.

Viterbi Decoder in Hardware

M

Mário Pereira Véstias

Instituto Politécnico de Lisboa, Portugal

INTRODUCTION

Trellis decoding is used to recover encoded information that was corrupted during transmission over a noisy channel. The Viterbi algorithm is the most well known trellis-based maximum likelihood decoding algorithm. The Viterbi algorithm is executed by a Viterbi decoder. Different hardware solutions may be considered to implement a Viterbi decoder with different design requirements in terms of area, performance, power consumption, among others. The most appropriate solution depends on the metric requirements of the application as well as on the target technology.

Properties of the Viterbi algorithm are used to simplify and improve the architecture of the Viterbi decoder. In particular, statistical properties of the Viterbi algorithm are used to design parallel Viterbi decoders with very high data decoding rates.

The article focuses on the implementation of a Viterbi decoder in hardware, including optimizations to improve the area and performance.

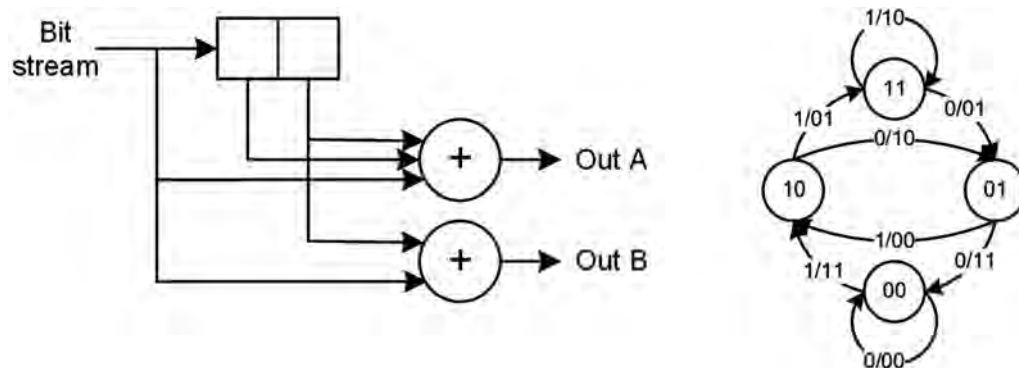
BACKGROUND

Generically, a communication system includes a channel encoder at the transmitter and a channel decoder at the receiver. Encoding the bit stream to be transmitted reduces the energy per data bit needed to achieve the same bit error rate (BER) over a noisy channel compared to a non-coded transmission. A better channel code needs lower energy to obtain a particular BER.

Convolutional Encoder

Two major classes of binary codes are block codes and convolution codes. In this article we are only concerned with convolution codes (P. Elias, 1955). Convolution codes operate over a continuous bit stream. A convolutional encoder is used to generate the encoded bitstream. A typical implementation consists on a shift register and a set of modulo-2 addition. The content of the shift register defines the state of the encoder (s bits), from which we obtain a finite state machine with $m = 2^s$ states. Figure 1 shows an example of a

Figure 1. Convolution encoder with rate 1/2



DOI: 10.4018/978-1-5225-2255-3.ch549

convolution encoder with rate 1/2 (rate 1/c generates c bits for each input bit) and its state machine.

An alternative way to represent the behavior of the encoder was proposed by Forney (G. D. Forney, Jr., 1967) designated *trellis representation*. A trellis is one of the most convenient ways to visualize the behavior of the decoding algorithms (see Figure 2).

Each node in the trellis represents a state of the FSM. After m time steps, the trellis is full and the branch pattern repeats indefinitely.

Most work on trellis decoding was dedicated to the codes with a rate of 1/2. This type of codes will provide the simplest trellis decoding since there are only two nodes leaving and entering a state.

Viterbi Algorithm

The Viterbi algorithm (VA) was proposed in 1967 by Andrew Viterbi (Viterbi, 1967). The algorithm decodes convolutional codes with an optimum non-sequential algorithm. Contrary to previous algorithmic solutions whose complexity increases exponentially with the length of the encoded sequence, the computational complexity of the Viterbi algorithm increases linearly with the length of the bit stream. The algorithm has three main steps: (1) branch metric calculation; (2) trellis calculation; and (3) traceback decoding.

The branch metric, bm , calculation determines the squared Euclidian distance, between the re-

ceived symbol, r_k , and the expected symbol associated with a branch of the trellis diagram, c_k , that is, r_k . At every time step, $|r_k - c_k|^2$ branch metrics are calculated since each state has two input branches.

The linear computation complexity of the VA is achieved by discarding unlikely branches in the trellis diagram using the Principle of Optimality (Bellman and Dreyfus, 1962). According to this, whenever two paths of the trellis diagram merge into the same state only the path with the best metric (most likely) must be kept. This path is designated *survivor path*. At each time step, this operation is applied to all paths and m survivor paths are obtained at each time step. Formally, a state metric, $sm(state)$ (see Figure 3), is updated as follows:

$$2 \times m$$

At each time step, the trellis calculation step obtains m decision symbols, one for each surviving branch. These symbols must be stored to help recover the symbols responsible for specific transitions during the traceback decoding step.

The trellis computation stops after a predetermined number of steps or after the decoding of the complete stream. From the m survivor paths, the final step (traceback decoding) starts in the state with the best metric and tracebacks through the

Figure 2. Trellis diagram of the encoder in Figure 1

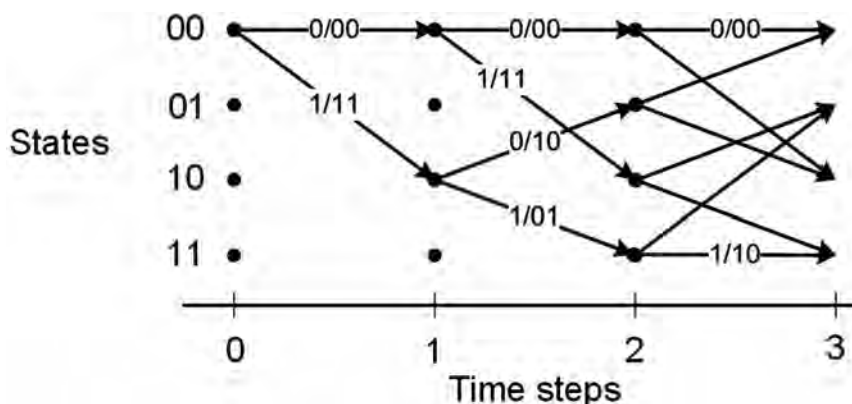
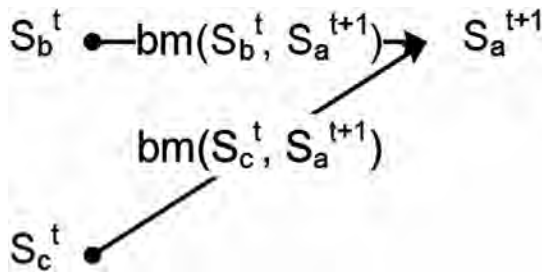


Figure 3. State metric update



path identified by the decision bits until reaching the first state and, consequently, obtain the most likely sequence of symbols.

The higher the number of steps before starting the traceback process, the higher the latency and the memory requirements. To reduce the latency and/or memory requirements the trellis process can be stopped after a pre-defined number of steps by, for example, considering extra bits that terminate the trellis in a defined state. The disadvantage of the method is the overhead associated with the introduction of bits with no information. Fortunately, the VA has one property that allows us to stop the trellis process in the middle of the decoding process and still recover the trellis path. Considering the m survivor paths at a step in the trellis diagram, when traced back, these paths merge into a single path (see Figure 4).

Formal estimations of L can be found in the literature (Anderson and Balachandran, 1989). However, in practice, the required value for L is found by simulation.

VITERBI DECODER IN HARDWARE

The Viterbi decoder implements the Viterbi algorithm with three functional units, one for each of the three steps of the VA (see Figure 5).

The branch metric unit (BMU) calculates the Euclidian distances associated with each branch for each input symbol. The Viterbi algorithm considers, by definition, that the received sequence is “hard”, that is, the decoded bits are either 0 or 1. Instead of hard decision, soft decision can be considered, where soft bits (several bits) represent real numbers. In practice, at least three bits are used to represent the expected and received symbols.

The branch metrics are then accumulated with the state metrics and the survivor paths are chosen. This is implemented with an add-compare-select unit (ACSU), which adds the metrics, compares the metrics entering a state and selects the best metric. The decisions, d , are stored in a memory and used by the traceback unit (TBU) to recover the symbols associated with the survivor path.

Figure 4. Traceback of survivor paths merging into a single path

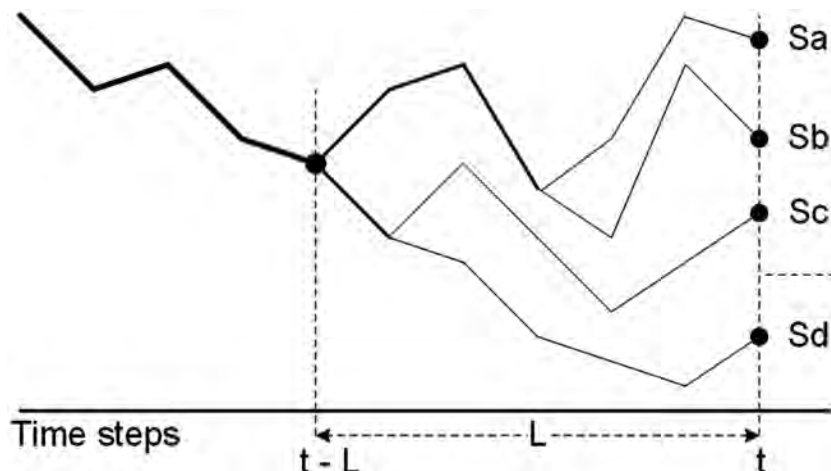
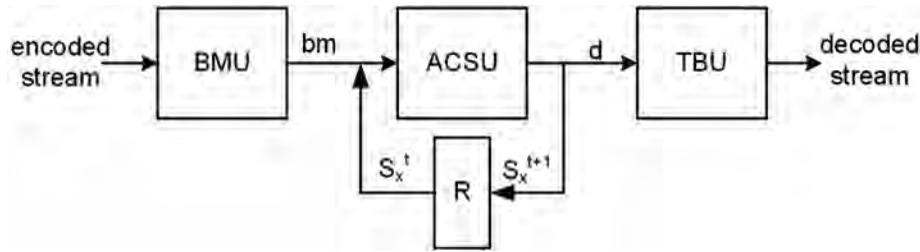


Figure 5. Viterbi decoder



In the following, code rates of $1/c$ are considered. This assumption not only simplifies the trellis decoding, since there are only two nodes leaving and entering a state, but also simplifies the implementation of the branch metric and the add-compare-select units.

Branch Metric Unit

The Euclidian distance between a received symbol, r , and the expected symbol associated with a branch of the trellis diagram, c , is $sm(S_a^{t+1}) = \min(S_b^t + bm(S_b^t, S_a^{t+1}), S_c^t + bm(S_c^t, S_a^{t+1}))$ applied to all dimensions, d . Formally,

$$|r - c|^2$$

In the equation, both squares, $b_m(r - c) = \sum_{n=0}^{d-1} (r_i^2 - 2r_i c_i + c_i^2)$ and r_i^2 contrib-

ute equally to the equation and thus can be neglected. So, the equation simplifies to:

$$c_i^2$$

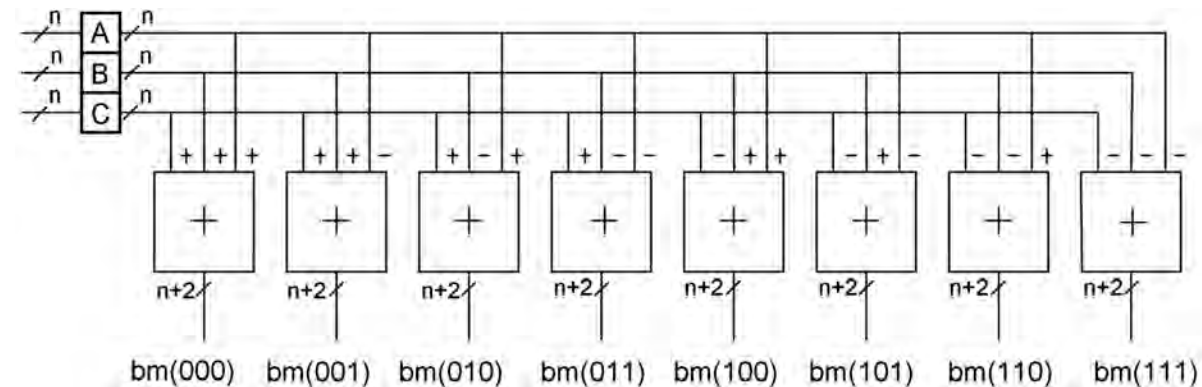
Considering noiseless symbols 0 and 1 of c represented by the symmetrical values -1 and +1, the branch metric reduces to add/subtraction of the received values

$$b_m(r - c) = -\sum_{n=0}^{d-1} (r_i c_i)$$

Therefore, only additions need to be implemented to calculate the branch metrics (see example in Figure 6 for $1/3$ code rate).

The unit can be reduced to half since $b_m(r - c) = -\sum_{n=0}^{d-1} \pm r_i$. For example, $b_m = -\overline{b_m}$. Both branch and complemented branch are gener-

Figure 6. Simplified branch metric implementation



assigning multiple add-compare-select operations to a single ACS unit (*folding factor*) (Shung, et al., 1993 – part I), (Shung, et al., 1993, – part II). The time-multiplexed implementation reduces the number of ACS units, but there is an increased design complexity relative to the controller and an increase in the number of registers. Also, the critical path increases due to extra multiplexers to route data to a single unit which reduces the expected linear increase operating frequency considering the folding factor.

When the encoded data is received in a continuous stream of bits, the path metric which is the accumulated value of the branch metrics, may overflow. To avoid overflow, the path metrics must be normalized. A very efficient technique that avoids overflow is the modulo normalization technique (Shung et al., 1992) that automatically normalizes the path metric. In modular arithmetic the metric, m_j , is replaced by the normalized metric $\hat{m}_j = 2^s$, where C is the maximum value represented by the number of bits used for defining the path metric. The differences in the survivor metrics remain unchanged using modular arithmetic provided that $\hat{m}_j = (m_j + \frac{C}{2})(\text{mod } C) - \frac{C}{2}$, where diff is the bound of the difference between survivor metrics.

Formally, the required number of bits for the state metric is given by:

$$\text{diff} \leq \frac{C}{2}$$

$$\text{W h e r e } S_{\text{bits}} = \lceil \log_2(2 \times \Delta_{\text{max}}) \rceil ,$$

$\Delta_{\text{max}} = \lambda_{\text{max}} \log_2 m = \text{maximum branch metric}$ and m is the number of states. For example, for 64 states, and code rate of 1/3, $\lambda_{\text{max}} = 18$. So, $S_{\text{bits}} = 8$. This normalization is implemented in hardware using a 2's complement adder.

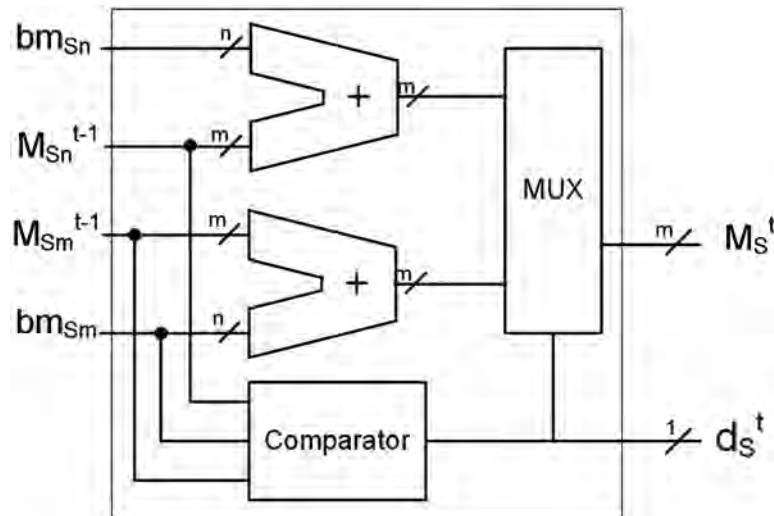
Another simplification rule following the metric calculation using modular arithmetic is the modified compare (Shung et al., 1992). The modified comparison rule applied to two normalized metrics with p bits λ_{max} and $\hat{m}_x = (mx_{p-1} + mx_{p-2} + \dots + mx_0)$ is given by:

$$\hat{m}_y = (my_{p-1} + my_{p-2} + \dots + my_0)$$

where $y(\cdot)$ denotes the unsigned comparison. In general, this comparator is smaller and faster than using a subtractor to implement comparison.

A few optimizations techniques have also been proposed to improve the performance of the

Figure 8. Parallel ACS unit implementation



ACS unit. For example, an ACS parallelization technique was proposed in (Véstias and Sarmiento, 2010) to reduce the critical path of the unit. The critical path of the ACS unit includes one adder, one comparator and one multiplexer. To reduce this critical path, the addition in the ACS unit is done in parallel with the comparison (see Figure 8).

The comparator determines the addition output with the lowest value according to the following equation

$$mx_{p-1} \oplus my_{p-1} \oplus y(mx_{p-2} + \dots + mx_0, my_{p-2} + \dots + my_0)$$

$$\text{Since } d_s^t = (m_{y \rightarrow x}^{t-1} - b_{m_z \rightarrow x}) - (m_{z \rightarrow x}^{t-1} - b_{m_z \rightarrow x})$$

$$b_{m_z \rightarrow x} = -\overline{b_{m_z \rightarrow x}}$$

Results indicate improvements of around 50% in the delay considering a *Field Programmable Gate Array* (FPGA) implementation.

Trace-Back Unit

The trace-back unit receives the accumulated metric values of all states, finds the state with the lowest value and returns the path ending in this state, that is, the sequence of bits that best matches the received sequence.

To find the state with the lowest metric all m metrics are compared. Several architectures with different area/performance tradeoffs can be considered to determine the maximum of a set of values. A fast approach is to use a tree of comparators having $\log_2 m$ levels, that is, $d_s^t = (m_{y \rightarrow x}^{t-1} - m_{z \rightarrow x}^{t-1} + 2 \times b_{m_z \rightarrow x})$ comparators at

the first level, $\frac{m}{2}$ at the second, and so on (or $\frac{m}{4}$ shared comparators). After $\log_2 m$ cycles, the state with the lowest metric is identified. The solution is fast but expensive in terms of hardware resources. An approach with small area is to con-

sider a recursive process where at each iteration the highest value found so far is compared to a next value. This sequential solution takes m cycles to complete, increasing considerably the latency and may become the bottleneck if the ACS step takes less than m cycles. Usually, an efficient solution is to consider a tree of iterative comparators (Véstias et al., 2012).

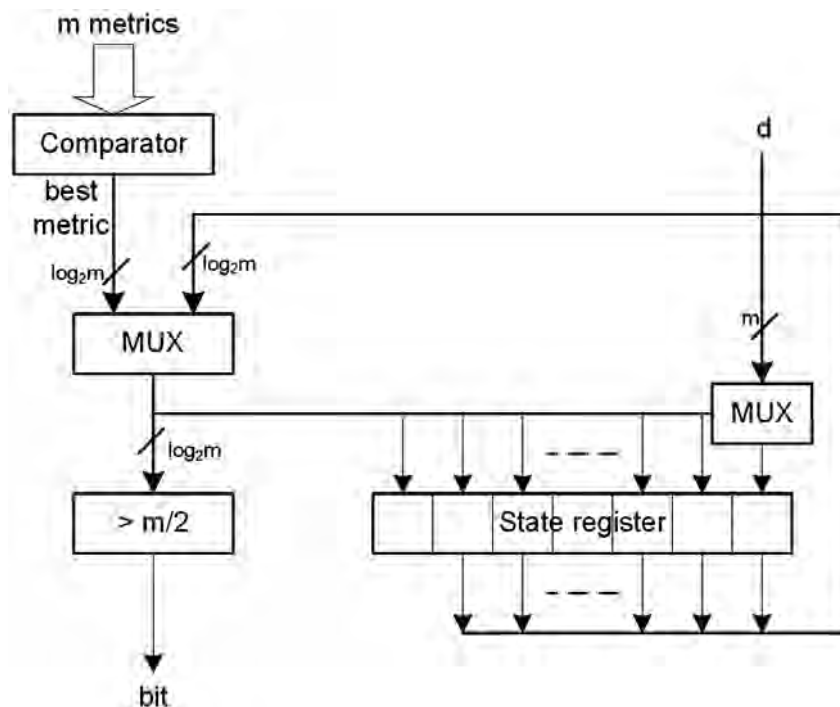
Given the state with the highest metric, the TB unit starts from this state and determines serially the values of the decoded bits ($b_0, b_1, b_2, b_3, \dots, b_{\text{TBL}-1}$), where TBL is the trace back length or the number of time steps in the trellis diagram before traceback. The circuit determines the backward path based on the decision bits found during the calculation of the paths. From a state and a decision bit the block finds the previous state. The process repeats iteratively until all decision bits are found. Two properties of the trellis diagram are used to find the correct path. One of the properties is that there are transitions

to states from 0 to $\frac{m}{2}$ if the input bit is 0 and to

states $\frac{m}{2} - 1$ to $\frac{m}{2}$ if the input bit is 1. This is easy to check since the most significant bit of the state is the input bit. The other useful property states that the preceding states are obtained by shifting the bits identifying the state one bit to the left and introducing 1 or 0. For example, the preceding states of 17 (010001) and 49 (110001) are 34 (100 010) and 35 (100 011). The correct previous state depends on the decision bit stored during the trellis decoding. The property results from the fact that the next state is obtained by shifting the state one bit to the right. Both properties are used to implement the TB unit (see Figure 9).

The state with the best metric is initially stored in the state register. Then the content of the state register evolves iteratively according to the stated property to find the previous state. The output block $m - 1$ implements the first property to determine the decoded bit.

Figure 9. Implementation of the TB unit

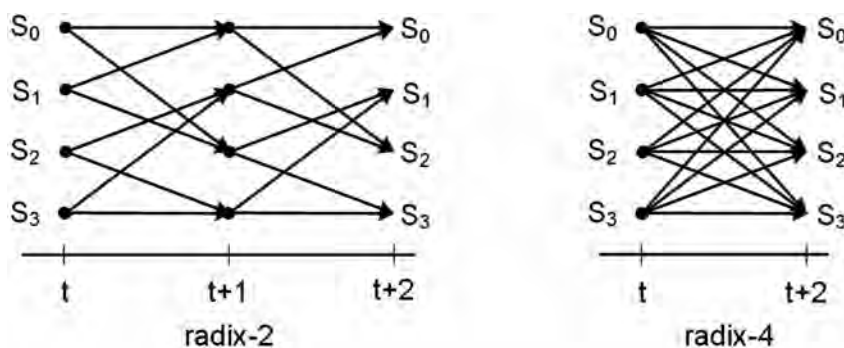


VITERBI DECODER PARALLELIZATION

The parallel VD approaches improve the decoding throughput by increasing the number of state updates at a single time step. A way to achieve this is to apply a lookahead technique of two or more levels (Fettweis, 1989). The look-ahead technique is a parallelization technique proposed in (Black & Meng, 2002). In a single cycle the technique

updates the state metrics of n consecutive steps and thus ideally increases the throughput n times, compared to a direct implementation. However, there is an area penalty proportional to n . To see how the technique works, let's consider a trellis diagram with states with two inputs and two outputs (radix-2) and collapse two steps into a single one producing a radix-4 trellis diagram (four branches entering and leaving each state) (see Figure 10).

Figure 10. A radix-2 trellis diagram transformed into a radix-4 trellis



We could think about higher radices. However, as concluded in (Black & Meng, 1992), radix-4 solutions are more area efficient than higher radices. Also, the critical path increases due to the extra operations of the ACS unit and the feedback interconnections. In (Black & Meng, 1992), a full-custom implementation of both radix-2 and radix-4 architectures have shown speed-ups of 1.7, lower than the theoretical 2 times speed-up.

Unbounded throughput using parallelization can be obtained with the minimized method (Fettweis & Meyr, 1991) and the sliding-block decoder (Black & Meng, 1997)). Following these techniques, parallel implementations of the VD can be implemented by decomposing the input data into $> \frac{m}{2}$ blocks which can be processed in parallel using k conventional Viterbi decoders. The complexity of the circuit increases linearly with the throughput. However, each block processing requires the initial state metrics to be known and this is only available after the previous block has finished processing. Three techniques have been proposed as a solution to this problem: state initialization (Dunn, 1978), interleaving (Lin & Messerschmitt, 1989) and sliding block (Tzu & Dunham, 1981). The state initialization technique forces the encoder to a known state at the start of each block. This adds complexity to the decoder since it needs block synchronization. The interleaving technique processes independently encoded sequences. However, this approach cannot be used when intersymbol interference is present. The sliding block decoder technique processes independent overlapping blocks simultaneously.

To acquire the initial states, two sync blocks are used at the beginning and at the end of the block. The method is based on two main observations. One is that the survivor paths from all possible starting states merge with high probability L iterations back into the trellis. The parameter is the survivor path length (SPL) and is typically $k \times M$ (Clark & Cain, 1981). The other is that when starting with unknown initial state metrics, the state metrics after K trellis iterations are independent of the initial metrics. The parameter is the synchronization length (SL) and is also typically $5 \times s = 5 \times \log_2 m$ (Viterbi & Omura, 1979).

Based on this characteristics, a state at time n can be decoded using only information from the interval $5 \times \log_2 m$ to $n - K$. Typically, $K = L$ is considered. The sliding block VD (SBVD) method applies the Viterbi algorithm to the interval $[n - M/2 - L, n + M/2 + L]$ to decode the interval $[n - M/2, n + M/2]$. To guarantee correct initial and final states, the method considers an initial synchronization block (SL) and a final trace-back block (SPL) without decoding for each parallel VD (see the behavior of each VD for the SBVD in Figure 11).

With this technique, we can use independent parallel VD to increase the decoding rate (see Figure 12).

The relative area efficiency is $n - L$. As M increases the method approaches the ideal efficiency of 1. However, the bigger the value of M , the bigger the latency and more storing resources are needed to save intermediate states. The decoding rate, DR, of an SBVD is:

Figure 11. Sliding block method for Viterbi decoder

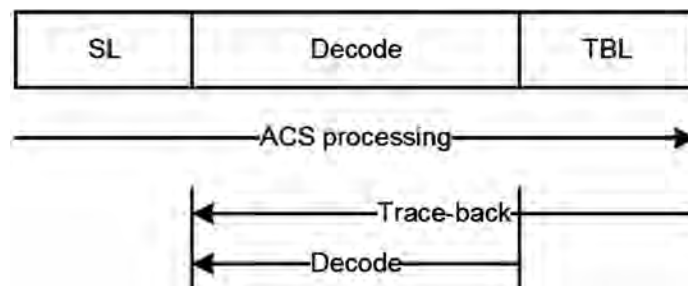
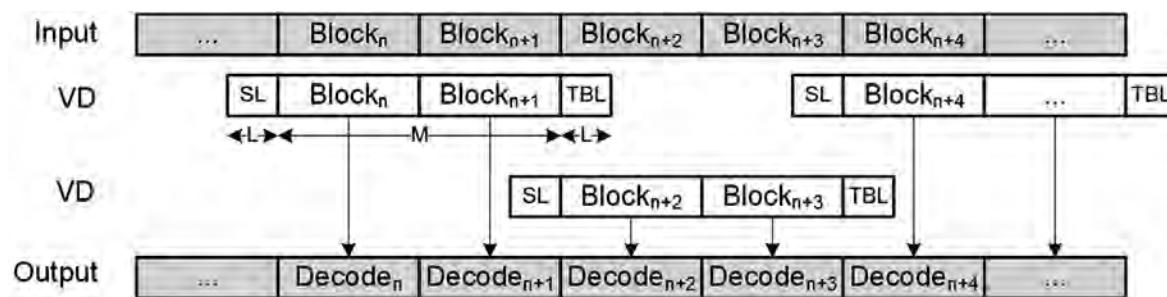


Figure 12. Sliding block method for Viterbi decoder



$$\frac{M}{M + 2L}$$

where nVD is the number of parallel VD and $drVD$ is the decoding rate of a single VD. The bit error rate of the decoding process of the VD depends on the length of the initial synchronization step, the length of the final trace-back length and of the trace-back decode length. All these parameters are usually found through system simulation.

FUTURE RESEARCH DIRECTIONS

Viterbi decoding implementations exist with different implementation tradeoffs including area, performance and power consumption. Parallel techniques allow any decoding rate just by increasing the number of parallel Viterbi decoders and thus an increase in the area and power consumption.

One of the characteristics of the Viterbi decoder is that of being the most resource-consuming compared to other decoding algorithms, but theoretically it does maximum likelihood decoding. However, there are some impediments to yield the ideal maximum likelihood due to several factors, including quantization of the input stream and finite trace back length.

A major challenge in the future of the Viterbi decoder is how to achieve high throughput with low power consumption. This is very important for portable devices.

CONCLUSION

The article describes the backgrounds of the Viterbi algorithm as a way to design a Viterbi decoder. The typical architecture of the Viterbi decoder is explained and implementations of each block of this architecture are provided, including optimizations to get faster or smaller implementations. Parallel solutions are also discussed which permits the design of high throughput Viterbi decoders.

The Viterbi algorithm was proposed in 1967 and since then many Viterbi decoder architectures were proposed. Different design requirements are appearing every day in different technologies, e.g., handheld devices, making the research about Viterbi decoder an actual topic of research.

REFERENCES

- Anderson, J. B., & Balachandran, K. (1989). Decision depths of convolutional codes. *IEEE Transactions on Information Theory*, 35(2), 455–459. doi:10.1109/18.32141
- Bellman, R. E., & Dreyfus, S. E. (1962). *Applied Dynamic Programming*. Princeton, NJ: Princeton University Press. doi:10.1515/9781400874651
- Black, P., & Meng, T. (2002). 140-Mb/s, 32 state, radix-4 Viterbi decoder. *IEEE Journal of Solid-State Circuits*, 27(12), 1877–1885. doi:10.1109/4.173118

- Black, P. J., & Meng, T. H.-Y. (1992). A 140-Mb/s, 32-state, radix-4 Viterbi decoder. *IEEE Journal of Solid-State Circuits*, 27(12), 1877–1885. doi:10.1109/4.173118
- Black, P. J., & Meng, T. H.-Y. (1997). A 1-Gb/s, four-state, sliding block Viterbi decoder. *IEEE Journal of Solid-State Circuits*, 32(6), 797–805. doi:10.1109/4.585246
- Chang, Y.-N., Suzuki, H., & Parhi, K. K. (2000). A 2-Mb/s 256-state 10-mW rate-1/3 Viterbi decoder. *IEEE Journal of Solid-State Circuits*, 35(6), 826–834. doi:10.1109/4.845186
- Clark, G. C., & Cain, J. B. (1981). *Error-Correction Coding for Digital Communications*. New York: Plenum. doi:10.1007/978-1-4899-2174-1_6
- Dunn, J. (1978). A 50 Mb/s Multiplexed Coding System for Shuttle Communications. *IEEE Transactions on Communications*, COM-26, 1877–1885.
- Elias, P. (1955). Coding for noisy channels. *IRE Convention Record*, 2, 4:37–47.
- Fettweis, G., & Meyr, H. (1989). Parallel Viterbi algorithm implementation: Breaking the ACS bottleneck. *IEEE Transactions on Communications*, 37(8), 785–790. doi:10.1109/26.31176
- Fettweis, G., & Meyr, H. (1991). High-speed parallel Viterbi decoding: Algorithm and VLSI-architecture. *IEEE Communications Magazine*, 29(5), 46–55. doi:10.1109/35.79382
- Forney, G. D., Jr. (1967). *Review of random tree codes*. Final Report Appendix A Contract NAS2-3737, NASA CR 73176. NASA Ames Research Center.
- Lin, C.-C., Shih, Y.-H., Chang, H.-C., & Lee, C.-Y. (2005). Design of a power reduction Viterbi decoder for WLAN applications. *IEEE Transactions on Circuits and Systems—Part I: Regular Papers*, 52(6), 1148–1156. doi:10.1109/TCSI.2005.849106
- Lin, H., & Messerschmitt, D. (1989). Algorithms and Architectures for Concurrent Viterbi Decoding. *Proceedings ICC'89*, 2, 836–840. doi:10.1109/ICC.1989.49807
- Shung, C. B., Lin, H.-D., Cypher, R., Siegel, P. H., & Thapar, H. K. (1993). Area efficient architectures for the Viterbi algorithm—Part I: Theory. *IEEE Transactions on Communications*, 41(4), 636–644. doi:10.1109/26.223789
- Shung, C. B., Lin, H.-D., Cypher, R., Siegel, P. H., & Thapar, H. K. (1993). Area efficient architectures for the Viterbi algorithm—Part II: Applications. *IEEE Transactions on Communications*, 41(5), 802–807. doi:10.1109/26.225495
- Shung, C. B., Siegel, P. H., Ungerboeck, G., Thapar, H. K., Trung, T. K., Shih, M. T., & Satorius, E. H. et al. (1992). VLSI architectures for Metric Normalization in the Viterbi algorithm design. *IEEE Transactions on Communications*, 40(3), 616–624.
- Tzu, K.-H., & Dunham, J. (1981). Sliding Block Decoding of Convolutional Codes. *IEEE Transactions on Communications*, COM-29(September), 1401–1403.
- Véstias, M., Neto, H., & Sarmiento, H. (2012). Efficient Design of Sliding Block Viterbi Decoders in FPGA. *Proceedings of IEEE International Conference on Field Programmable Logic Applications*. doi:10.1109/FPL.2012.6339215
- Véstias, M., & Sarmiento, H. (2010). Implementing High-Speed Viterbi Decoders for MB-OFDM on FPGA. *Proceedings of XXV Conference on Design of Circuits and Integrated Systems*.
- Viterbi, A. J. (1967). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, 13(2), 260–269. doi:10.1109/TIT.1967.1054010
- Viterbi, A. J., & Omura, J. K. (1979). *Principles of Digital Communication and Coding*. McGraw-Hill.

ADDITIONAL READING

Anders, M. A., Mathew, S. K., Hsu, S. K., Krishnamurthy, K., & Borkar, S. (2008). A 1.9 Gb/s 358 mW 16–256 State Reconfigurable Viterbi Accelerator in 90 nm CMOS. *IEEE Journal of Solid-State Circuits*, 43(1), 214–222. doi:10.1109/JSSC.2007.909336

Anderson, J. B. (2005). *Digital Transmission Engineering* (2nd ed.). Piscataway, NJ: IEEE Press. doi:10.1002/0471733660

Balivada, L. K., & Raju, K. P. (2012). *Optimization Techniques of Viterbi Algorithm: Performance Analysis of Different Algorithms*. LAPLAMBERT Academic Publishing.

Benedetto, S., & Biglieri, E. (2002). Principles of Digital Transmission: With Wireless Applications: Viterbi Algorithm. Springer US.

Chu, C.-Y., & Wu, A.-Y. (2012). Power-Efficient State Exchange Scheme for Low-Latency SMU Design of Viterbi Decoder. *Journal of Signal Processing Systems for Signal, Image, and Video Technology*, 68(2), 1–13. doi:10.1007/s11265-011-0603-0

Kamuf, M., Owall, V., & Anderson, J. (2008). Optimization and Implementation of a Viterbi Decoder Under Flexibility Constraints. *IEEE Transactions on Circuits and Systems I*, 55(8), 2411–2422. doi:10.1109/TCSI.2008.918148

Lin, M.-B. (2000). New Path History Management Circuits for Viterbi Decoders. *IEEE Transactions on Communications*, 48(10), 1605–1608. doi:10.1109/26.871381

Lin, S., Kasami, T., Fujiwara, T. & Fossorier, M. (1998). *Trellis and Trellis-Based Decoding Algorithms for Linear Block Codes*. Springer Science+Business Media, LLC.

KEY TERMS AND DEFINITIONS

Add-Compare-Select Unit (ACSU): A unit that accumulates the branch metrics of each branch along the trellis diagram.

Branch Metric Unit (BMU): A unit to calculate the Euclidian distances associated with each branch of the Trellis diagram for each input symbol.

Convolutional Codes: A type of error correcting code used in telecommunications.

Convolutional Encoder: A module used to generate convolutional codes.

Trace-Back Unit (TBU): A unit of the VD that finds the state with the lowest value and returns the path ending in this state.

Trellis Diagram: A type of graph where nodes represent states and each node is connected to at least one previous node (state) and one next node (state).

Viterbi Algorithm: An algorithm to decode convolutional codes with an optimum non-sequential algorithm. The computational complexity of the Viterbi algorithm increases linearly with the length of the bit stream. The algorithm has three main steps: branch metric calculation; trellis calculation; and traceback decoding.

Viterbi Decoder (VD): An implementation of the Viterbi algorithm consisting of three main functional units, one for each of the three steps of the Viterbi algorithm.

Wireless Implant Communications Using the Human Body

M

Assefa K. Teshome
Victoria University, Australia

Behailu Kibret
Victoria University, Australia

Daniel T. H. Lai
Victoria University, Australia

INTRODUCTION

Advancements in medical diagnostics and bio-sensing technology opened up a great venue for research in electronic medical implant. To improve accuracy and timeliness of diagnosis, electronic devices could be implanted inside human body to provide various real-time diagnostics information. However, effective technique for implant communication is still an open problem. Early efforts based on radio wave propagation are standardised as the Medical Implant Communication Services (MICS) for 402–405 MHz frequency range which was later adopted as Medical Devices Radio Communications Services (MedRadio) for 401–406 MHz frequency range (Hanna, 2009).

Currently, the radio-frequency (RF) implant wireless communication is enabled by utilizing small antennas that radiate radio waves inside the human body. As a bid to find alternative wireless implant communication mechanisms within the Wireless Body Area Network (WBAN), in this work, the authors explored two complementary techniques. The first uses galvanically coupled Intra-body communication (IBC) for implant-to-surface communication. IBC is a relatively new technique that uses the human body as a channel with communication frequencies not exceeding several MHz. The second technique uses the human body itself as an antenna by feeding an RF current into the tissues.

This chapter first examines a new analytical electromagnetic model that uses galvanically coupled IBC where the implant transmitter differentially injects current into the tissue via its anode and cathode electrodes. A wearable receiver on the surface of the skin samples the resulting potential difference using its two electrodes. Frequencies ranging from hundreds of kHz up to a few MHz are considered under quasi-static assumptions. The model is unified in the sense that it is based on multilayered ellipsoidal geometry that can be applied to any part of the body (i.e., head, torso, limbs etc.). It also effectively describes influences of tissue properties and geometry of the body part. The security and low power consumption of IBC are also apparent in this model. The path loss characterisation of IBC implants shows lower values compared to their MICS counterparts.

In addition, the chapter also elaborates on the scenario when the RF current is fed by a tiny toroidal inductor that is implanted and clamped around the tissues in the ankle. The frequency range of 1–70 MHz is considered, which includes the resonance frequency of the human body. Theoretical results show that the system exhibit broadband characteristics with a maximum gain of - 25 dB between 20 to 40 MHz, assuming an isotropic radiation from human body. However, for the case of the small toroidal inductors considered, the radiation resistance of the system is very small, which increases the power consumption.

DOI: 10.4018/978-1-5225-2255-3.ch550

BACKGROUND: IMPLANT COMMUNICATION IN THE WBAN ARCHITECTURE

To improve accuracy and timeliness of diagnosis, and hence improve quality of life, sophisticated actuators and biosensors are emerging for various diagnostic applications; for example, glucose sensors for continuous diabetes monitoring (Heo et al., 2013). In broad terms, implant communication technique explored in literature use radio wave propagation, magnetic induction and volume conduction (Poon, 2010), (Bjorninen et al., 2012), (Yang, 2006).

For implant communication it is important that the transmitter consumes small power to conserve battery life. The implant should also be miniaturized for a minimal invasive embedding. Besides, due to sensitive nature of medical data, security is a paramount requirement of implant communication. To achieve security either the signal needs to be encrypted at the transmitter or be confined to within the body detectable by as far as an on-body receiver. In the case of MedRadio based implant, the signal is radiated outside the human body; hence, requires all security features be implemented right at the transmitter which increases the transmitter complexity. Hence, the transmitter consumes large power and is difficult to miniaturise.

Implant communication scenario can be divided as implant-to-implant, implant-to-surface, and human body to external environment. The authors explored the use of galvanically coupled IBC for implant-to-implant and implant-to-surface scenario; and the use of the human body itself as an antenna for a possible scenario of communicating implant inside human body to external environment.

The general architecture of body area sensor networks, as shown in Figure 1, is that a link node wearable on the surface talks to and listens from the implanted and other surface mounted devices. It then combines and relays the signal to devices external to the body – mainly a monitoring or

controlling device away from the body. Another likely scenario is the possibility of two implants talking to each other; for example, a glucose sensor and an insulin pump. To reduce complexity and power consumption it is better to implement advanced security features at the link node rather than each individual implanted or on-body device. To avoid eavesdropping attempts to listen or talk to sensors in and on the body by any transceiver external to the body, the signal needs to be confined to within the body. For frequencies ranging from a few hundreds of kHz to tens of MHz, the human body hardly radiates radio waves. Thus, this band is suitable for body confined (intra-body) transmissions – implant-to-implant, implant-to-surface and surface-to-surface communications.

To communicate the signal to outside the body wirelessly, radio wave propagation of RF signals is required. Such is the case for the link node or possibly an individual implant desired to directly communicate with the outside environment. For this scenario, the authors explored using the human body as an antenna. Specifically, frequency ranging from 10 – 110 MHz was considered. If the human body is excited by an RF current in this frequency range, the human body radiates electromagnetic signals. Furthermore, the human body resonance frequency is found to be between 30 – 70 MHz which falls within the frequency of interest. Transmission power levels within human tissues considered in this chapter are in accordance with the safety recommendations of Specific Absorption Rate (SAR) levels by the International Commission on Non-Ionizing Radiation Protection (ICNIRP) (ICNIRP, 1994).

THE HUMAN BODY AS A CHANNEL FOR IMPLANT COMMUNICATIONS

The HBC uses an electric field communication (EFC) where the human body is effectively a volume conductor. It exploits the lossy dielectric nature of the conductive tissue layers to induce a current, and hence a potential distribution, as a

result of the electric field caused by the current injected by the transmitter electrode(s). Such a low frequency signal is expected to penetrate deeper into the tissue layers, thus requiring less power be detected by a receiver on the surface of skin. Moreover, due to a non-conducting free space outside the skin surface, the signal is confined to within the human body (i.e., inherently secure).

GALVANICALLY COUPLED IBC MODEL FOR IMPLANT COMMUNICATION

In literature, several modeling techniques have been used to model the human body as a channel. Some of these techniques include the two-terminal circuit model (Besio, 2006), (Hachisuka, 2006), four-terminal circuit model (Hachisuka, 2006), (Kibret et al., 2014), finite element approach (Wegmueller et al., 2006), (Amparo et al., 2014), surface electromagnetic propagation models (Ruiz, 2006), (Bae, 2012), nearfield electromagnetic models (Bae, 2012) and quasi-static electromagnetic models (Pun, 2011). Each model shows various

useful aspects of intra-body communication. However, the electromagnetic models used are based on simple geometries which limit their use to only specific regions of the body. On the other hand, circuit models are generally short-sighted in the sense that the relationship between channel variations and circuit components is not obvious in the transfer functions, albeit well characterized channels in fixed settings.

The analytical electromagnetic model presented on this chapter is motivated by geometries robust enough to capture tissue layer effects in a scalable way that can be applied to any part of the body and yet simple enough to guarantee analytical solutions. Based on this geometry, a mathematical model of the channel is derived to characterise the received signal as a function of the size of the transmitter, tissue layers of the body part, transmitter location, receiver location and electrode spacing.

Most body parts can be approximated by a variation of the ellipsoidal geometry as shown in Figure 2. For example, the human head can be modeled using an ellipsoid close to spherical symmetry. The torso can be modeled by a pro-

Figure 1. Implant communication the WBAN architecture

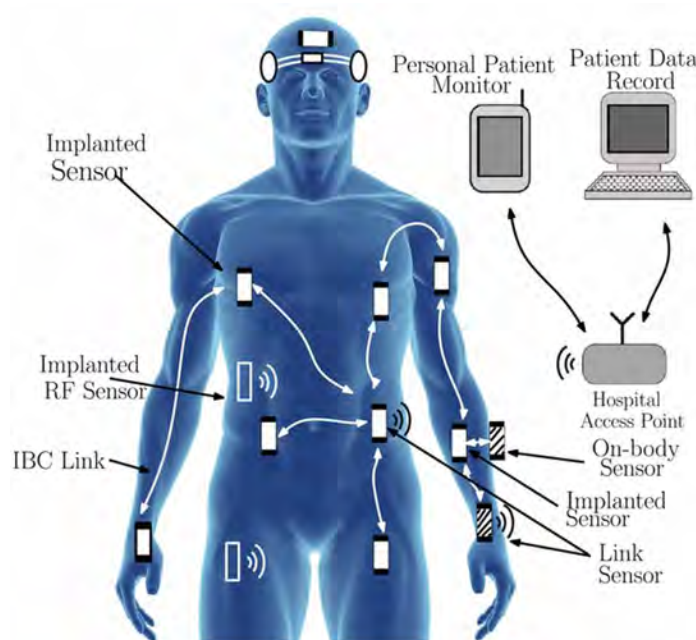
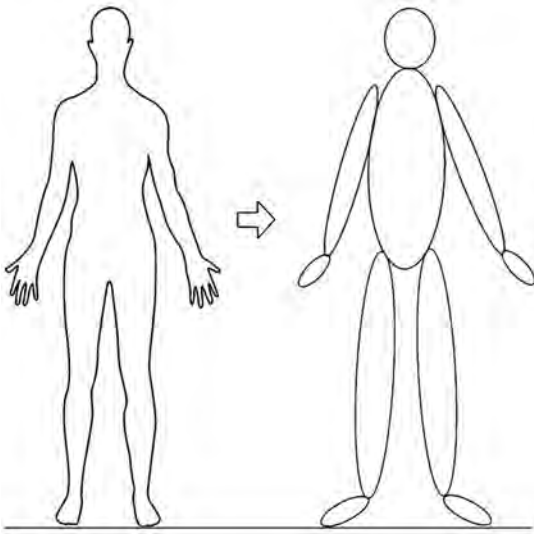


Figure 2. Ellipsoidal approximation of the human arm



late or oblate spheroid versions of the ellipsoidal geometry. Limbs can be modeled by an ellipsoid where a dominating semi-axis represents the limb length whereas the other two axes represent the larger and the shorter radii of the limb. When one of the semi-axes is large and the other two are comparable, the geometry resembles a cylinder which is often used in existing electromagnetic models e.g. (Pun et al., 2011), (Amparo et al., 2014).

Hence, an analytical model based on multilayered ellipsoidal representation of tissue layers can represent the various body parts by defining the semi-axes lengths. The ellipsoidal shells in the layers represent each tissue layer with varying thickness and complex permittivity.

VOLUME CONDUCTION THEORY FOR IMPLANT COMMUNICATION

Volume conduction can be defined as a transmission technique for electric field inside the volume of a lossy dielectric where an electric field is induced by a primary current source; and this field propagates to the receiver by means of an induction current induced in the conductive medium (Tenke

et al., 2012). An implant transmitter coupled galvanically can be envisaged as the primary current source inside any one of the layers of body tissue. The total current inside the volume (i.e., primary and induction currents added) creates an electric potential distribution inside and on the surface of the volume (Rutkove, 2007). The receiver could be either another implant or on-body device where its two electrodes are used to sample the potential difference between the two points of the body the electrodes are connected to. The transmission frequencies we consider are smaller than the high frequency band of the spectrum. From the conductivity and permittivity profiles of human body tissues at these small frequencies, electromagnetic signals in the body can be assumed to be quasi-static (Pun et al., 2011), (Plonsey et al., 1967). In quasi-static assumptions, the variations of electric and magnetic fields, denoted as $\mathbf{E}$ and $\mathbf{B}$ respectively, with time are negligible. Hence, the set of Maxwell's equation describing the fields inside the human body can be modified as given in (1).

$$\nabla \times \mathbf{E} = 0, \quad \nabla \times \mathbf{B} = \mu \mathbf{J}, \quad \nabla \cdot \mathbf{E} = 0, \quad \text{and} \quad \nabla \cdot \mathbf{B} = 0 \quad (1)$$

where μ is permeability of free space and $\mathbf{J}$ is the net current density inside the volume. Here, the electric field $\mathbf{E}$ is given as

$$\mathbf{E} = -\nabla V, \quad (2)$$

where V is the electric potential distribution. The current inside the volume is due to the conduction source current density $J^s(r) = Md(r - r_0)$ $J^s(r) = \mathfrak{M}\delta(r - r_0)$ and the induced current density $J^i(r) = \sigma(r)\mathbf{E}(r)$ where $\sigma(r)$ is the conductivity of the tissue layer at point r in space, $\mathfrak{M}$ is the electric dipole moment and r_0 is the point in space that is mid-way between the transmitter electrodes. Thus, the net current density $J(r)$ is given as

$$J(r) = \mathfrak{M}\delta(r - r_0) + \sigma(r)\mathbf{E}(r). \quad (3)$$

No current flows out of the human body due to the non-conducting medium (air) outside the body; hence, $J(r)$ is a divergent free current density (i.e., $\nabla \cdot J(r) = 0$). Taking the divergence of (3) and replacing (2) into (3), we can see that

$$\nabla \cdot \sigma(r) \nabla V(r) = \nabla \cdot \mathfrak{M} \delta(r - r_0). \quad (4)$$

Equation (4) is the governing equation for the electric potential which takes different forms for the parts of the volume; i.e., it takes the form of a Poisson's equation in the region (layer) that contains the source and takes the form of a Laplace's equation in the layers that do not contain the source.

POTENTIAL DISTRIBUTION: AN EXAMPLE SCENARIO

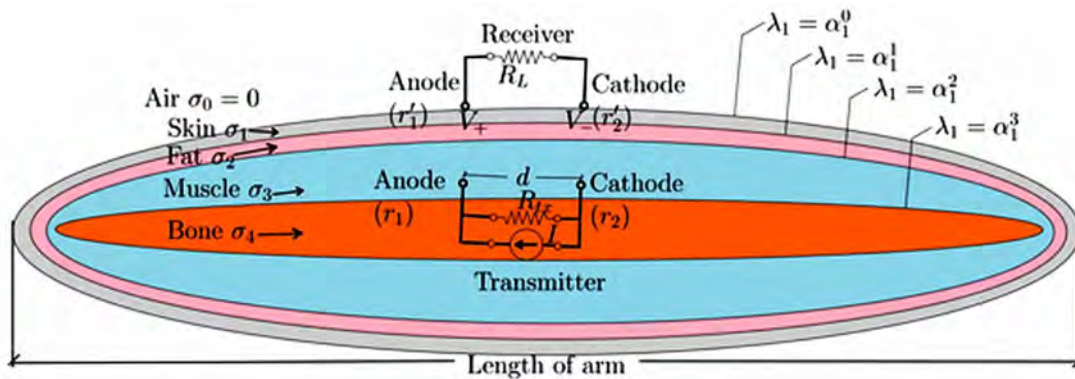
In this section, the solution for the potential V in Equation (4) is presented for an example scenario shown in Figure 3. The scenario considered is the case where a transmitter is implanted inside the muscle tissue of the human arm and the receiver is placed on the surface of the skin. In this case four tissue layers are considered; i.e., bone, muscle, fat and skin. The tissue layers are represented by multi-layered confocal ellipsoidal shells.

The appropriate coordinate system to solve for the potential is the Ellipsoidal coordinate system. The solution of (4) in Ellipsoidal coordinate systems can be simplified by using special functions called Lamé's functions of the first and second kind (Dechambre, 2000). The general form of the potential distribution inside the i^{th} layer is given by

$$V_i(\lambda_1, \lambda_2, \lambda_3) = \sum_{n=0}^{\infty} \sum_{p=1}^{2n+1} A_{(n,p)}^i \mathbb{E}_n^p(\lambda_1, \lambda_2, \lambda_3) + \sum_{n=0}^{\infty} \sum_{p=1}^{2n+1} B_{(n,p)}^i \mathbb{F}_n^p(\lambda_1, \lambda_2, \lambda_3), \alpha_1^i \leq \lambda_1 \leq \alpha_1^{i-1}, \quad (5)$$

where λ_1, λ_2 , and λ_3 are the ellipsoidal coordinate system, $\mathbb{E}_n^p$ and $\mathbb{F}_n^p$ are the Lamé's function of first and second kind, respectively, with degree p and order n . $A_{(n,p)}^i$ and $B_{(n,p)}^i$ are the coefficients corresponding to the i^{th} tissue layer and α_1^i is the dominating semi-axis length of the i^{th} tissue layer. Equation (5) is the general solution and specific solution for every scenario is provided by thoroughly specifying the coefficients $A_{(n,p)}^i$ and $B_{(n,p)}^i$ by applying Dirichlet and Numan boundary conditions at each layer.

Figure 3. Implant communication scenario where the transmitter is implanted inside the muscle tissue of the human arm and the receiver is mounted on the surface of the skin



PATH LOSS AND POTENTIAL DISTRIBUTION

To calculate the potential distribution and hence the path loss human arm model given in Figure 3 and Equation (5) are used. An arm of smallest semi-axis 43.5 mm is considered with tissue thicknesses of skin = 1.5 mm, fat = 8.5 mm, muscle = 27.5 mm, bone = 6 mm.

Consider the transmitter injecting an r.m.s. current of 1 mA with its electrodes spaced by 5 mm located along the major semi-axis of the arm at 6.9 mm into the muscle tissue from the muscle-bone interface. The maximum electric potential developed along the axis of the dipole as a function of radial distance from the center is shown in Figure 4. The received potential inside the body decreases at a slower rate inside the body and decreases at a faster rate once the signal leaves

Figure 4. Potential distribution with distance at different frequencies

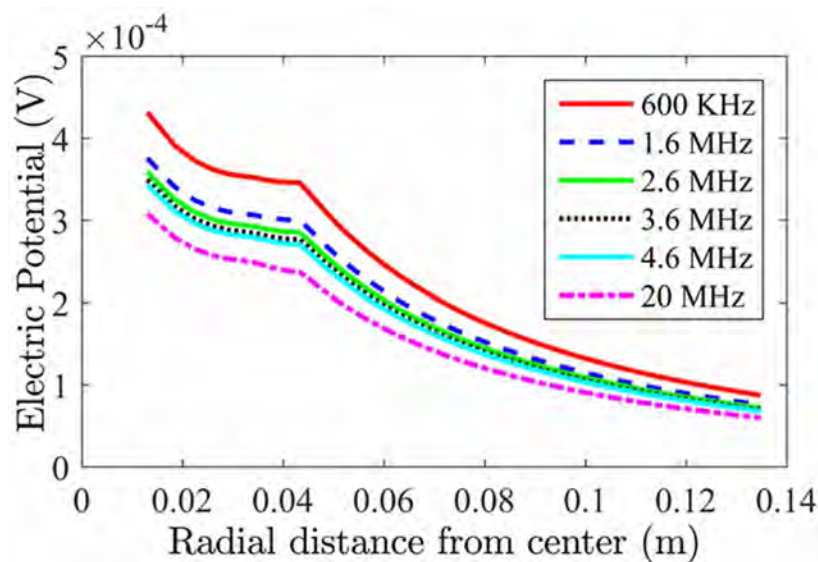


Figure 5. Path loss as function of distance at different frequencies

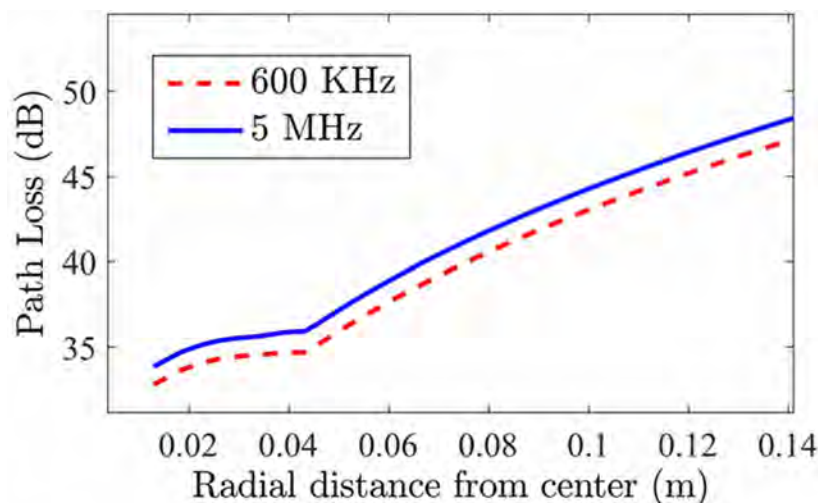
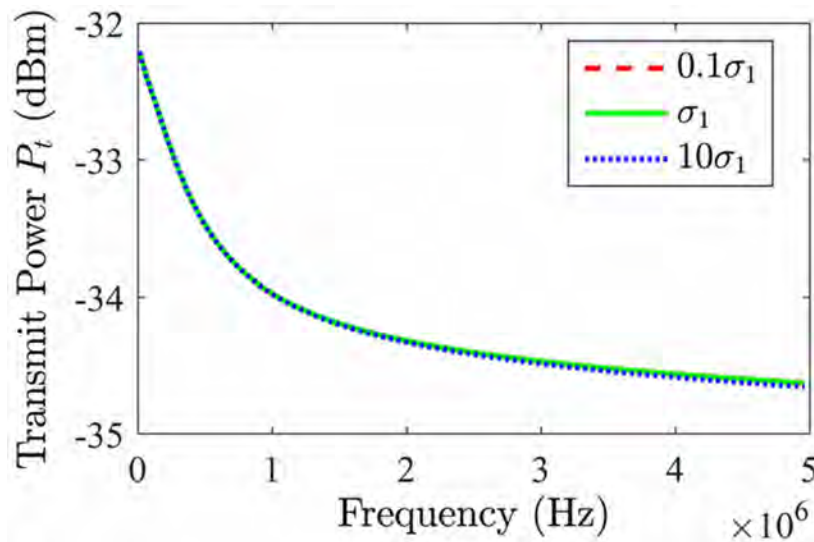


Figure 6. Transmitted power as a function of frequency for different skin conductivity



the body. This can also be seen in the path loss which increases at a large slope outside the skin (at 0.0435m from the center) as shown in Figure 5. The path loss at the surface of the skin is around 35 dB which is a lot smaller than the path loss reported for MICS based implants. The impedance of the tissue layer affects the amount of power transmitted. From Figure 6 the transmit power decreases with frequency which due to decreasing impedance with frequency. For a 1 mA r.m.s., the transmitted power is around -32 to -35 dBm; hence, the received power is around -67 dBm to -70 dBm. The received power is thus a lot larger than the average receiver sensitivity of -92.5 dBm required by IEEE 802.15.6 standard. The level of signals considered here are well within the ISNIRP guidelines.

THE HUMAN BODY ANTENNA: IN TRANSMISSION MODE

A remarkable historical attempt to use the human body as an antenna goes back to the 1970s, in a research undertaken by the US army, in an effort to design camouflaged and wearable radio transmitters that use the human body as antenna

(Ikraht et al., 1973). The objective was to replace the traditional infantry whip antennas, which were not suitable to be used in a jungle environment - being physical obstacles to the radio operator. The human body can be completely characterized as a transmitting or receiving antenna from the induced currents in the body. The problem of computing the induced currents can be simplified by considering a typical scenario of a vertically polarized plane wave illuminating a human subject standing on a highly conductive ground. For this specific case, it was assumed that the axial current induced by the vertically polarized electric field is dominant. Other characteristics of the human body as a receiving or transmitting antenna, such as the antenna impedance, radiation efficiency and reflection coefficient, can be derived from the expressions of the axial current.

This problem of characterising the human body as a transmitting antenna is addressed by the authors in their previous work (Kibret et al., 2015) where the human body is effectively characterised as a cylindrical monopole antenna; experimentally confirmed results show that, for a human subject of height 1.76 m and weight 73 kg, it was found that the human body behaves like a monopole antenna with resonance frequency of 40-60 MHz

when fed on the foot depending on the posture of the body. Thus, a transmitter inside the ankle inducing an axial current in the human can act as a feed to the human body antenna.

THE AXIAL CURRENT DENSITY IN THE ANKLE

Since the cross-section of the ankle consists of different tissues, the axial current density is not uniformly distributed.

The tissues have different dielectric properties; therefore, they interact differently to the axial RF electric field passing through them. Accordingly, the authors developed a model for the cross-section of the ankle based on a realistic anatomical atlas of the human body (Netter, 2014) with various tissues assigned specific colours as shown in Figure 7. The 4-Cole-Cole dispersions model was used to represent the dielectric properties of these tissues.

For the frequency range of interest, it was assumed that the axial electric field in the cross-section of the ankle is uniform. The total induced

Figure 7. The cross-section of ankle joint with tissues assigned different colours. 1-muscle, 2-fat, 3-connective tissue, 4-cancellous bone, 5-skin, 6-blood vessel, 7-tendon, 8-nerve, 9-blood, 10-cortical bone

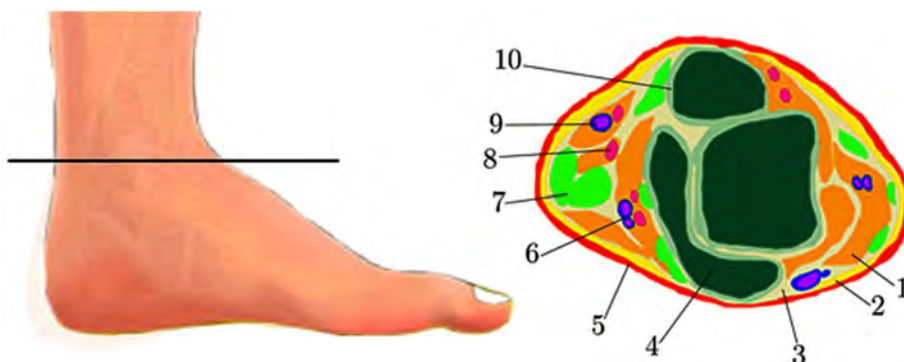
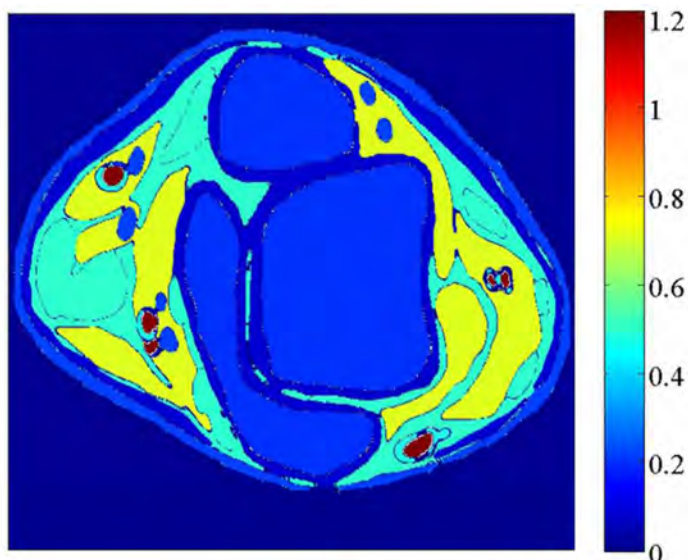


Figure 8. The magnitude of axial current density in the cross-section of the ankle joint with area of 0.0154m^2 for an incident electric field $E_0 = 1\text{ V/m r.m.s.}$



axial current I_z in the ankle, which includes both the conduction and displacement current in the cross-section of the ankle of area S , can be calculated from the axial current density distribution in the ankle J_z as

$$I_z = \int_S J_z ds = E_z \int_S \sigma_s^* ds \quad (6)$$

where E_z is the axial electric field that is taken to be uniform and σ_s^* is the complex conductivity of 'points' in the cross-section. Considering the cross-section of the ankle represented by the grayscale image, the expression in (6) can be approximated as given in (7)

$$I_z \simeq 2E_z \sum_{m=1}^M \sigma_m^* A_{pixel} \quad (7)$$

where σ_m^* is the complex conductivity of the tissue represented by the m^{th} grayscale pixel; M is the total number of pixels in the ankle cross-section and A_{pixel} is the area of a single pixel. A_{pixel} can be easily calculated by dividing the actual area of the ankle cross-section with M . Therefore, the axial current density J_z^n in the area represented by the n^{th} pixel can be approximated as

$$J_z^n = \sigma_n^* E_z = \frac{0.5 \sigma_n^* I_z}{A_{pixel} \sum_{m=1}^M \sigma_m^*} \quad (8)$$

Figure 8 shows the magnitude of the axial current density calculated using (8) for an incident electric field $E_0 = 1$ V/m r.m.s. illuminating the adult male human subject that was discussed in (Kibret et al., 2015). It can be seen that the highest current density occurred in the blood and muscle tissues, and the least occurring in the skin, fat, nerve, blood vessels, cortical bone and cancellous bone. Almost half of the current density in the muscle tissue exists in the connective tissues and tendons.

THE IMPLANTED TOROIDAL INDUCTOR

In the preceding section, we showed that a large axial current density is induced inside the muscle tissue. In order to collect this current for the purpose of implant wireless communication, a toroidal inductor that is clamped around a group of muscle fibers or a tendon in the ankle is assumed. Consider a toroid placed around the muscle tissue or tendons with the axis of toroid perpendicular to the cross-section of the ankle.

The toroid has inner radius r_1 , outer radius r_2 , and height h_t , as shown in Figure 9. An enameled

Figure 9. The half-section of the implanted toroidal inductor with ferrite core

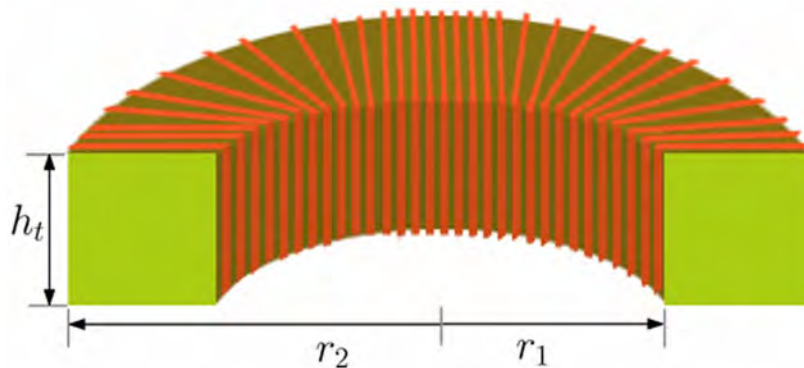
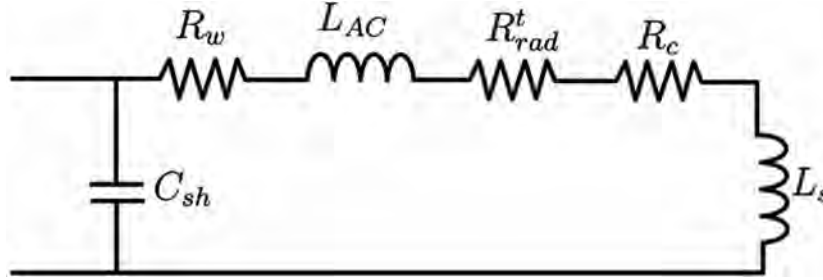


Figure 10. The equivalent circuit of the toroidal inductor. C_{sh} is the parasitic capacitance between turns; R_w represents the loss in winding wire; L_{AC} is the internal inductance of the wire; R_{rad}^t is the radiation resistance of the toroid; R_c represents the magnetic core loss; and L_s is the self-inductance.



round copper wire of radius r_{cop} is wound around a ferrite core making N number of turns. In order to reduce the effect of leakage inductance, the inner wall of the toroid is completely covered by the copper windings. Thus, the number of turns was derived as $N \approx \pi r_1 / (r_{oop} + t_e)$ where t_e is the thickness of the enamel layer of the copper wire. The equivalent circuit representation of the toroid is shown in Figure 10.

THE PERFORMANCE OF THE WHOLE SYSTEM IN TRANSMISSION MODE

The operation of the system is similar to that of a current transformer; the toroid acts like the primary winding while the human body was approximated as a single turn secondary, as shown in Figure 11. The field induced inside the toroid core induces axial current in the tissue along the axis of the toroid. The representation of the human body as a single turn secondary describes more accurately a toroidal inductor clamped to the exterior of the ankle. The equivalent circuit of the ankle implant in transmission mode was approximated as shown in Figure 12. As a loop antenna, a toroidal inductor by itself has poor radiation efficiency (Balanis, 2005); therefore, we ignored its radiation resistance R_{rad}^t . The approximation of this setup to an implanted toroidal inductor is based on the result from (9) that large axial current density exists in

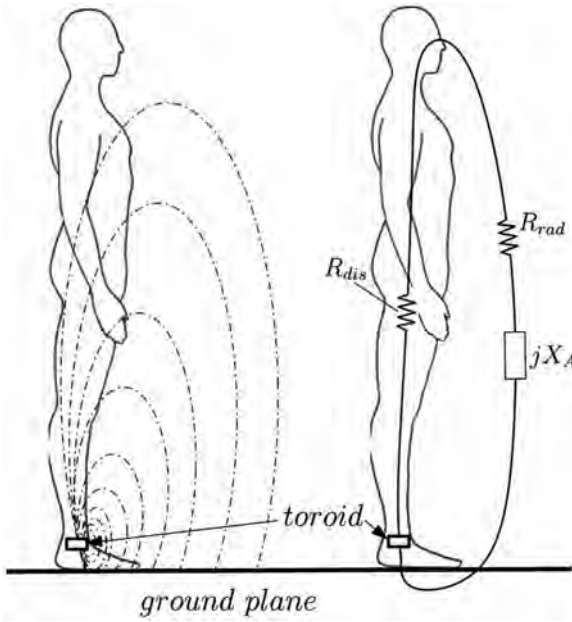
the muscle tissue which spans about 20% of the ankle joint cross-sectional area.

The equivalent circuit in Figure 12 consists of the equivalent circuit of the toroid in transmitting mode includes magnetic core effects, skin-effect and proximity-effect in the copper winding, and parasitic capacitance of the inductor

In Figure 12, Z_b is the sum of the impedances that represent the effects of the human body, the shoes, the ground and near-field coupling with nearby objects. Assuming the human subject is standing on a highly conductive ground plane and bare foot, Z_b can be approximated by the human body antenna impedance Z_A which is expressed in (Kibret et al., 2015) as the antenna impedance of the human body when the human subject is excited by a delta-gap electromotive force (emf) placed between the feet and the ground. The delta-gap emf was assumed to be equal to the emf induced by the implanted inductor, which is approximated by a current-probe feeding model.

Based on the analysis in (Kibret, 2015), Z_A can be expanded as $Z_A = R_{rad} + R_{dis} + jX_A$ where R_{rad} characterizes the power radiated P_{rad} ; R_{dis} characterizes the power dissipated inside the human body due to ohmic and dielectric loss of tissues; and X_A characterizes the near-field reactive power oscillating in the vicinity of the human body. From the expression of the resistance per unit length of the equivalent cylindrical antenna and from the definition of the radiation efficiency in (Kibret et al., 2015), the expression of R_{rad} and R_{dis} can be derived.

Figure 11. Visualisation of the lines of force in the human body forming the secondary loop



The circuit in Figure 12 can be simplified by transferring the secondary impedance to the primary as shown in Figure 13. Since the winding is assumed to completely cover the inner wall of the toroid, the magnetising inductance L_m was approximated by the self-inductance L_s of the toroid. The equivalent impedance seen by the transmitter Z_{eq} can be written as given in (9)

$$Z_{eq} = \frac{Z_s}{j\omega C_{sh} Z_s + 1} \quad (9)$$

where

$$Z_s = \frac{j\omega L_m (N^2 Z_A + R_c)}{j\omega L_m + N^2 Z_A + R_c} + R_c + R_w + j\omega L_{AC}.$$

Figure 12. The equivalent circuit of the implanted toroidal inductor. V_{in} is the input voltage of the transmitter; Z_o is the output impedance of the transmitter; L_m is the magnetising inductance referred to the primary side; and Z_B is the impedance of the human body and the surrounding

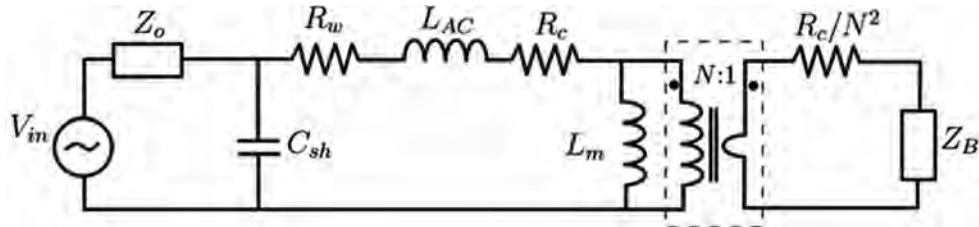
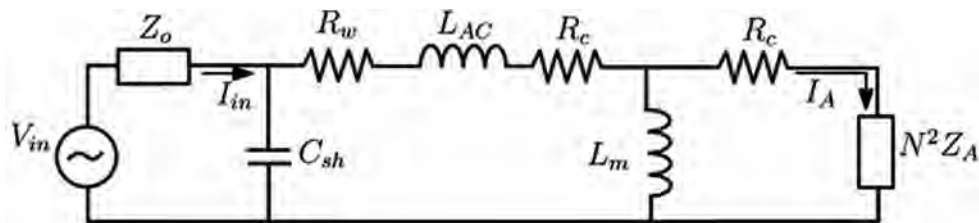


Figure 13. The simplified equivalent circuit of the implanted toroidal inductor in transmitting mode



Therefore, referring Figure 13, the radiation efficiency of the system η_r^t defined as the ratio of the radiated power P_{rad} to the input power P_{in} can be written as given in (10)

$$\eta_r^t = \frac{P_{rad}}{P_{in}} = \frac{|I_A|^2 N^2 R_{rad}}{|I_{in}|^2 \text{Re}(Z_{eq})} \quad (10)$$

The radiated power escapes into the surrounding using the human body as antenna. For simplicity, assuming the radiated power was emitted equally in all directions, the antenna gain of the system G can be approximated as

$$G(\text{dB}) = 10 \log_{10}(\eta_r^t). \quad (11)$$

The gain in the direction of the maximum emission can be approximated by that of a quarter-wave monopole antenna by adding 5.19 dBi to the value obtained from (11).

THE IMPLANTED TOROID INDUCTOR: AN ILLUSTRATIVE SCENARIO

Considering two implanted toroidal inductors with parameters shown in Table 1, the gain G calculated is shown in Figure 14. As can be seen in the table, the two toroids have a core thickness, $r_2 - r_1$ of 3 mm and a height h_t of 1 cm. The areas enclosed by the toroids are 0.64% and 2.16% of the cross-sectional area of the ankle joint considered in this study, which is indicated by the parameter A % in the table. The copper wire used was assumed to

be 20 AWG (American Wire Gauge) copper wire with enamel thickness of $0.25r_1$. It can be seen that the gain exhibits a broadband feature with the maximum occurring between 20 – 45MHz, which includes the resonance frequency of the human body. The resonance frequency of the human subject used for the two scenarios, which has height H_m and weight m given in Table 1, is near 40 MHz (Kibret et al., 2015).

The gain calculated using (11), was based on the assumption that only the human body impedance affects the current induced inside the human body. But, in reality, other factors, such as, the ground, shoes and nearby objects, affect the gain. When the impedance of 1 cm thick rubber shoes was added in series with the human body impedance Z_A , the calculated gain decreased by a maximum of 5 dB.

The given scenario can be further analysed by assuming an input current $|I_{in}|$ of constant magnitude. It was also found that more than 98.5% of the power was dissipated inside the magnetic core and a maximum of 1% of the power was absorbed inside the human body, with a negligible amount dissipating inside the wire, as shown in Figure 15. The radiated power was less than 0.5%. This suggests that the two important factors that determine the performance of the toroid are the parasitic capacitance and the magnetic core losses. Reducing these effects improves the efficiency of the system.

One way of reducing this capacitance is by increasing the separation of turns in the inductor. But this approach also introduces a leakage inductance that does not contribute to the axial current induced inside the enclosed tissue. As can be deduced from Figure 14, a toroid that

Table 1. Parameters of the system (Kibret et al., 2016)

Parameters	r_1 [mm]	r_2 [mm]	h_t [cm]	d [m]	r_{cop} [mm]	t_e [mm]	N	σ [S/m]	H_m [m]	m [Kg]	A%
Case 1	6.59	9.59	1	$\lambda/2\pi$	0.4	0.1	41	5.592×10^7	1.76	73	0.64
Case 2	11.28	14.292	1	$\lambda/2\pi$	0.4	0.1	70	5.592×10^7	1.76	73	2.16

Figure 14. The calculated gain for the two cases

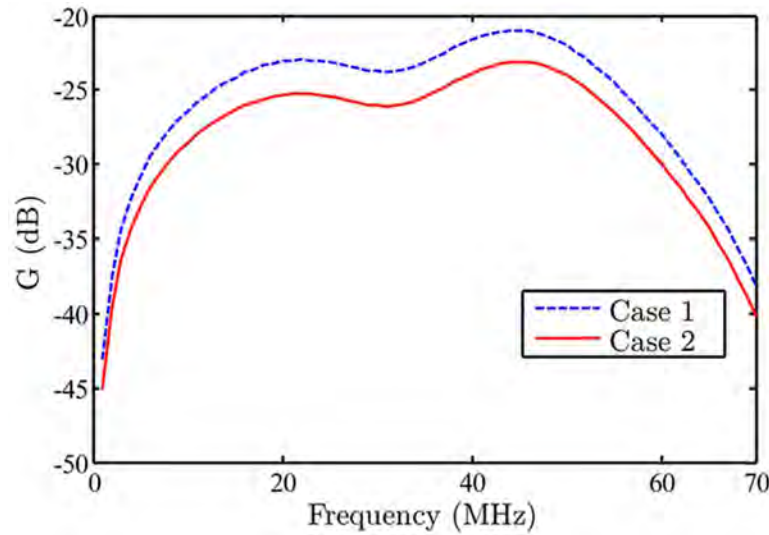
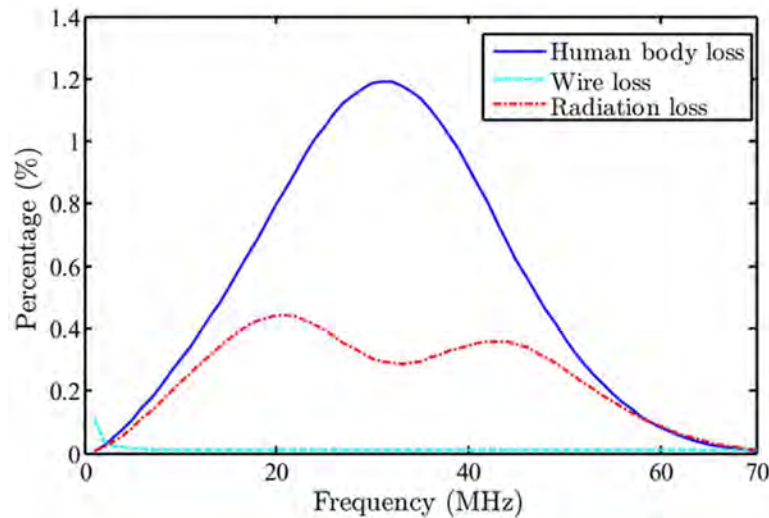


Figure 15. The power loss percentage in the human body, the wire and the radiated power



has a minimised leakage inductance and smaller number of turns has a better performance than a toroid with larger number of turns. Therefore, the design of an efficient system calls for an optimized choice of the smallest number of turns that result in a reduced leakage inductance.

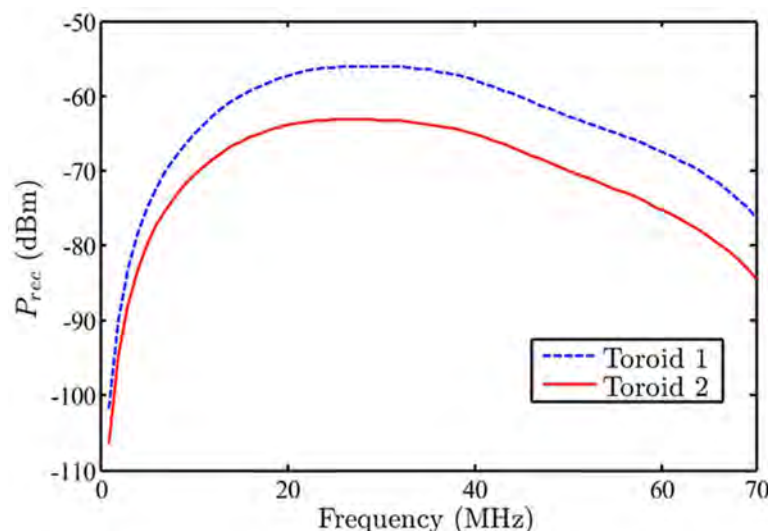
Assuming a transmitter of $EIRP=0$ dBm is located a distance of $d=\lambda/2\pi$ from the body, the power received P_{rec} on the load impedance $Z_L=50\Omega$ that is connected to the terminals of both inductors

is shown in Figure 16. As the received power is a function of the gain, the received power for toroid 1 is larger than that of toroid 2.

FUTURE RESEARCH DIRECTIONS

Experimental validations, as can be referred in (Teshome et al., 2016, and Kibret et al., 2015b), showed that the two techniques presented are not

Figure 16. The power loss percentage in the human body, the wire and the radiated power



only theoretical assertions, but also practical possibilities. As an immediate follow-up, the authors are working on transceivers that make use of the models presented in this chapter. The rising breakthrough in medical diagnostics and sensor technology demands more efficient implant transceivers. Most of existing implant-communication mechanisms are not tuned to this rising demand for medical implants and implant communications. Hence, low-power implant transmissions such as IBC and effective use of the human body as antenna needs further investigation in terms of transceiver coupling, powering systems and investigation of mitigation techniques for losses. The authors believe the results in this chapter could fill part of the knowledge gap in the field.

CONCLUSION

In this chapter two approaches are explored as alternative to existing implant communication techniques. A novel unified model for galvanically coupled IBC implant communication is presented. Analysis of the model shows that the technique is low power consuming, secure and offers a lower path loss compared to existing MICS based

implants for implant-to-implant and implant-to-surface setting. The second technique uses the human body itself as an antenna to radiate signals from the body to the outside environment. With proper mitigation of losses, the toroidal inductor provides a feasible feeding mechanism to utilise the body as an antenna.

REFERENCES

- Amparo Callejon, M., Reina-Tosina, J., Naranjo-Hernández, D., & Roa, L. M. (2014). Galvanic coupling transmission in intrabody communication: A finite element approach. *IEEE Transactions on Bio-Medical Engineering*, 61(3), 775–783. doi:10.1109/TBME.2013.2289946 PMID:24216629
- Bae, J., Cho, H., Song, K., Lee, H., & Yoo, H.-J. (2012). The signal transmission mechanism on the surface of human body for body channel communication. *IEEE Transactions on Microwave Theory and Techniques*, 60(3), 582–593. doi:10.1109/TMTT.2011.2178857
- Balanis, C. A. (2005). *Antenna Theory: Analysis and Design*. John Wiley & Sons.

- Besio, W., & Prasad, A. (2006). Analysis of skin-electrode impedance using concentric ring electrode. *EMBS'06. 28th Annual International Conference of the IEEE*, 6414–6417.
- Bjorninen, T., Muller, R., Ledochowitsch, P., Sydanheimo, L., Ukkonen, L., Maharbiz, M., & Rabaey, J. M. (2012). Design of wireless links to implanted brain-machine interface microelectronic systems. *IEEE Antennas and Wireless Propagation Letters*, 11(1), 1663–1666. doi:10.1109/LAWP.2013.2239252
- Dechambre, D. (2000). *Computation of ellipsoidal gravity field harmonics for small solar system bodies*. Electrical Engineering-Systems in The University of Michigan.
- Hachisuka, K., Terauchi, Y., Kishi, Y., Sasaki, K., Hirota, T., Hosaka, H., & Ito, K. et al. (2006). Simplified circuit modeling and fabrication of intrabody communication devices. *Sensors and Actuators. A, Physical*, 130(1), 322–330. doi:10.1016/j.sna.2006.04.044
- Hanna, S. (2009). Regulations and standards for wireless medical applications. *Proceedings of the 3rd international symposium on medical information and communication technology*, 23–26.
- Heo & Takeuchi. (2013). Towards smart tattoos: Implantable biosensors for continuous glucose monitoring. *Advanced Healthcare Materials*, 2(1), 43–56.
- Ikrath, K., Murphy, K. J., & Kennebeck, W. (1973). *Development of camouflaged body coupled radio transmitters*. DTIC Document.
- International Commission on Non-Ionizing Radiation Protection (ICNIRP). (1994). Guidance for limiting exposure to time-varying electric, magnetic, and electromagnetic fields (up to 300 GHz). *Health Physics*, 74(4), 494–522.
- Kibret, B., Seyedi, M., Lai, D. T. H., & Faulkner, M. (2014). Investigation of galvanic-coupled intra-body communication using the human body circuit model. *IEEE Journal of Biomedical and Health Informatics*, 18(4), 1196–1206. doi:10.1109/JBHI.2014.2301165 PMID:25014932
- Kibret, B., Teshome, A. K., & Lai, D. T. H. (2015). Characterizing the human body as a monopole antenna. *IEEE Transactions on Antennas and Propagation*, 63(10), 4384–4392. doi:10.1109/TAP.2015.2456955
- Kibret, B., Teshome, A. K., & Lai, D. T. H. (2016). *Analysis of the Human Body as an Antenna for Wireless Implant Communication*. *IEEE Trans Antennas Propagation*. doi:10.1109/TAP.2016.2526070
- Netter, F. H. (2014). *Atlas of Human Anatomy*. Philadelphia: Saunders Elsevier.
- Plonsey, R., & Heppner, D. B. (1967). Considerations of quasi-stationarity in electrophysiological systems. *The Bulletin of Mathematical Biophysics*, 29(4), 657–664. doi:10.1007/BF02476917 PMID:5582145
- Poon, A. S. Y., ODriscoll, S., & Meng, T. H. (2010). Optimal frequency for wireless power transmission into dispersive tissue. *IEEE Transactions on Antennas and Propagation*, 58(5), 1739–1750. doi:10.1109/TAP.2010.2044310
- Pun, S. H., Gao, Y. M., Mak, P., Vai, M., & Du, M. (2011). Quasistatic modeling of human limb for intra-body communications with experiments. *IEEE Transactions on Information Technology in Biomedicine*, 15(6), 870–876. doi:10.1109/TITB.2011.2161093 PMID:21724520
- Ruiz, Xu, & Shimamoto. (2006). Propagation characteristics of intra-body communications for body area networks. *3rd IEEE CCNC 2006*, 1(1), 509–513.

Rutkove. (2007). Introduction to volume conduction. *The Clinical Neurophysiology Primer*, 388(1), 43.

Tenke, C. E., & Kayser, J. (2012). Generator localization by current source density (CSD): Implications of volume conduction and field closure at intracranial and scalp resolutions. *Clinical Neurophysiology*, 123(12), 2328–2345. doi:10.1016/j.clinph.2012.06.005 PMID:22796039

Teshome, A. K., Kibret, B., & Lai, D. T. H. (2016). Galvanically Coupled Intra-body Communications for Medical Implants: A Unified Analytic Model. *IEEE Transactions on Antennas and Propagation*, 64(7), 2989–3002. doi:10.1109/TAP.2016.2559519

Wegmueller, M. S., Huclova, S., Froehlich, J., Oberle, M., Felber, N., Kuster, N., & Fichtner, W. (2006). Galvanic coupling enabling wireless implant communications. *IEEE Transactions on Instrumentation and Measurement*, 58(8), 2618–2625. doi:10.1109/TIM.2009.2015639

Yang, G. Z. (2006). *Body Sensor Networks*. London: Springer. doi:10.1007/1-84628-484-8

KEY TERMS AND DEFINITIONS

Antenna: A device designed to radiate or receive electromagnetic radio waves.

Electromagnetic Models: Analytical and/or numerical models that describe propagation of electromagnetic signals through a medium.

Galvanically Coupled IBC: A type of intra-body communication where electric current is differentially coupled to the body using the anode and cathode electrodes implanted or mounted on the surface of the human body.

Human Body Antenna: The usage of the human body as an antenna to transmit or receive electromagnetic radio waves.

Implant Communication: A communication scheme where an electrical signal is transmitted to or from an electronic device implanted inside the human body.

Intra-Body Communication (IBC): Electric field communication using the human body as channel.

Wireless Body Area Network (WBAN): Signal transmission network of transmitters and receivers in, on and around the vicinity of the human body using the human body as part of communication channel.

Category M

Mobile Learning

E-Collaborative Learning (e-CL)

Alexandros Xafopoulos

University College London, UK

INTRODUCTION

Electronic collaborative learning (e-CL) is a highly topical issue in today's world of worldwide online communication. It can be regarded as a type of learning focusing on a specific learning method, collaboration, and a specific collaborative learning medium or resource, electronic technology (e-technology). The objectives of this article are to clarify the highly debated and promising concept of e-CL and its related terms, to provide insight in its process, and to present current, especially promising, trends.

E-CL appears new to the present digital age when employing digital technology, but its roots are found back in history as regards e-technology (Kock, 2008). The concepts and practices of collaboration and learning were evident in several forms from the beginning of human history and life.

BACKGROUND

Learning and Knowledge

To approach the meaning of e-CL the concept of learning must be examined. First of all, it should be noted that in this article human learning is considered although it may occur among other creatures, such as animals. Second, the concept of learning is interpreted by learning theories connected with the concept of knowledge. The hereby adopted interpretation of the learning process is as knowledge building, or creation or construction, following Scardamalia and Bereiter (2006).

To further explore the concept of knowledge, it is worth mentioning that there is a widely accepted progressive distinction of informative elements

into data, information, knowledge, and wisdom. At a data level there is isolated and meaningless information. Data are transformed into information when they acquire meaning. Information is enriched into knowledge or intelligence when it acquires situated context. Finally, knowledge is transformed into wisdom when it acquires personalised pragmatic context.

What is also worth noting is that four basic and increasingly difficult and important types of knowledge can be considered. Declarative, or factual or plain, knowledge regarding what, where, and when, as an instance; procedural or skill knowledge regarding how; conditional, or conceptual or structural or contextual or competence, knowledge regarding the previous in an interrelated context and regarding why; and finally, metacognitive knowledge or metaknowledge regarding cognition and knowledge (Krathwohl, 2002). The first two types may be characterised lower-order knowledge whereas the last two, the most valued, higher-order knowledge.

Collaboration

In addition, e-CL incorporates collaboration, which should be considered as a partnership community process, that is, the collaborating members share a sense of belonging, inclusivity, trust, and reciprocity as well as a common thinking and performing area. Moreover, the term collaboration is different from cooperation, since the former implies the construction of shared, mental or material, products to achieve a shared, mental or material, goal, for instance, problem solving, whereas in the latter each team member undertakes one part of the whole responsibility, constructs a separate product, combined with other members'

DOI: 10.4018/978-1-5225-2255-3.ch551

products in a later stage, and achieves a subgoal (Arvaja, Häkkinen, & Kankaanranta, 2008; Ertl, 2008; Laurillard, 2012).

Technology

Another e-CL fundamental notion is technology, which in general refers to any intended, mental or material, product of a mind-bearing being. In certain contexts it may be used to denote the digital technology. In this study it is used in the sense of e-technology, which can be analogue or digital, but with a great focus on the digital aspect used for collaborative learning purposes, which is more often and widely applied.

e-CL

Taking the previous conceptual descriptions into account one attempt to define learning is as the knowledge building process with knowledge building goals or objectives, and, in that sense, collaborative learning is regarded as the communicative shared-knowledge building process with shared-knowledge building goals. Following these, e-CL is considered as the communicative shared-knowledge building process with shared-knowledge building goals using networked electronic devices (Kock, 2008; Laurillard, 2012). The constructed shared knowledge is also referred to by the term 'collective intelligence'.

What is more, some discriminations are to be made. First, this study is concentrated on formal and not non-formal or informal learning. According to Werquin formal learning implies intended process and preset goals whereas non-formal implies intended process but non-preset goals and informal implies non-intended process and non-preset goals (Cameron & Harrison, 2012). Second, collaboration in learning differs from each of communication, interaction, discussion, inquiry, and practice in learning, although it may include them, by being a process involving management, iterative development, and agreements

to produce shared knowledge and outcomes based on it (Laurillard, 2012).

Third, this article focuses on digital devices, although the networked electronic devices do not necessarily imply computers, desktop or portable, but they may also be interconnected analogue electronic devices, such as, telephones or cameras (Kock, 2008). When computer networks are employed e-CL may appear as computer-supported collaborative learning (CSCL) applies. Finally, in the e-CL process several learning elements are identified. These are here distinguished into human, supportive, and framework elements and are examined below.

HUMAN ELEMENTS

In the e-CL process there are several roles of human participants or partners (Good & Robertson, 2006). They are usually considered and presented as communities or groups, although in lacking or trivial learning environments they may be constituted by only one person or be merged with each other. In large projects the communities may consist of several subcommunities and even create a structured hierarchy of nested communities. Furthermore, participants may be members of more than one community creating overlapping communities (Masterman & Vogel, 2007).

For the normal operation of the communities intra- and inter-community collaboration is essential. The participants in e-CL are also called stakeholders and the communities may be online or virtual, offline, face-to-face or real, or blended. Two frequently used terms, related to the e-CL process communities, are the community of enquiry (CoE) (Garrison, Anderson, & Archer, 1999) and the community of practice (CoP) (Wenger, 1998), which use enquiry and practice to achieve their goals, respectively.

The human elements are classified into the following communities; the management, the design, the implementation, the learning, the technology,

and other supportive communities (Xafopoulos, 2016). The management community determines the learning mission and overall learning aims, which must be implemented and followed by the other communities. The design community, following the general management community guidelines, undertakes the coordination, analysis, planning, and design of the e-CL process. The main role of the implementation community is the realisation of the e-CL process in actual settings. The learning community iteratively collaborates in order to build shared knowledge and artefacts.

Finally, supportive communities offer support to the e-CL process and may include technologists, funders, suppliers, maintenance staff, researchers, psychologists, neuroscientists, sociologists, anthropologists, linguists, philosophers, engineers, pastors, and guardians. The e-CL process can also be affected by people unsystematically or unintentionally interacting with the participants.

SUPPORTIVE ELEMENTS (TPACK)

Introduction

In the learning process of e-CL we may identify certain supportive learning elements. These are the learning media, or resources or technologies, the pedagogy or andragogy, the learning methods, including the way of communication and collaboration, and the learning content. The supportive elements may be categorised as being originally created, or being an adapted version, or just selected unchanged from another learning source. Following the technological pedagogical content knowledge (TPACK or TPCK) model or framework (Mishra & Koehler, 2006) three supportive elements could be identified. Technology, pedagogy, including methodology, and content.

Technology: Media

Technology incorporates media, applications, environments, and infrastructure. The appropriate

technologies are identified, combined, and applied for each e-CL learning environment. The presentation of the content through its several sensory forms, called modalities, requires the presence and function of certain devices or hardware called media, which substantiate it (Bezemer & Kress, 2008). There are several kinds of hardware supporting e-CL, for instance, digital and analogue mass media, desktop, mobile, and handheld computers, and mobile phones and devices.

Additional learning supporting hardware includes computer peripherals, such as cameras, projectors, speakers, and microphones, and other advanced tools, such as, advanced computer-mediated or augmented or virtual reality (VR) tools, interactive whiteboards (IWBs), digital wallboards, interactive transformable objects, multi-sensory and multi-touch devices, which interact with the participants in a variety of sensory ways, as well as robots.

Technology: Applications

The modalities can also be presented with the assistance of software applications. Several types of applications are used in combination with collaborative learning, which transform it into e-CL (Xafopoulos, 2016). The collaboration can occur real time in the same or different virtual room of these applications, which can be separated into web or web-based and non-web. The former ones gain increasing attraction and are divided into three generations (Xafopoulos, 2016).

Another distinction is between multimedia and non-multimedia, as well as interactive and non-interactive applications. Multimedia applications integrate some of the content modalities, whereas non-multimedia only a single content modality. The multimedia and interactive applications are increasingly widespread. Another category increasingly used in learning are mobile applications. Their considerable advantage is that they are usable in several places, even when moving or travelling. Finally, two applications which have attracted special interest in the educational

community are agents and avatars. They both are virtual humanoids, or even other kinds of creatures, created by programmers (Xafopoulos, 2016).

Technology: Environments

The applications and media can be integrated into electronic learning environments or platforms, such as, portals, social networks, virtual learning environments (VLEs) or learning management systems (LMSs), personal learning environments (PLEs), intelligent tutoring systems (ITSs), learning design environments, and massive online open courses (MOOCs) (Xafopoulos, 2016). Significant aspects that some of these environments support are learning content management, learner engagement, analytics and dashboards, and learning process capture. These learning environments can further be integrated into virtual worlds.

Technology: Infrastructure

The several kinds of networks and networking equipment, wired or wireless, are further technological elements supporting the learning process. These could be considered synonyms with the term information and communications technologies (ICTs). Another related concept is cloud computing for e-learning, that is, networks of remote servers and applications providing centralised storage and online access to e-learning services and resources (Dong, Zheng, Yang, Li, & Qiao, 2009).

Finally, certain digital technologies are supportive of the previously mentioned main technologies in the e-CL process, such as, user interfaces (UIs), which are studied by human-computer interaction (HCI), databases, knowledge bases, and search engines. These together with networks may be referred to with the term infrastructure.

Pedagogy: Learner Modelling

The pedagogy examines and proposes mainly about the human, and especially learner, model-

ling, about the learning objectives or goals or outcomes, and about the learning theories. It also proposes about the other supportive elements and serves as a basis to methodology which can be considered as part of pedagogy.

The learner modelling refers to the process of construction of learner models, that is, knowledge representations about learners. If the human models are accessible by the ones they model, and possibly also by other participants, they are called open. They may also be editable, interactive, and cooperative. The access takes place through an appropriately developed interface, such as, skill meters or concept maps. The represented knowledge is both static, which is also called profile, and dynamic, and usually significantly qualitative. For the learner modelling several kinds of participant aspects must be investigated (Xafopoulos, 2016).

Pedagogy: Learning Objectives

As regards the learning objectives or outcomes they refer to the desired results of the learning procedure given the status of the learning elements. The objectives may be structured in several layers from the more general and long-term, for example, final or institutional objectives, to the more specific and short-term ones, such as, activity objectives.

The more specific objectives are expressed using actions, conditions, regarding the human and supportive elements except the pedagogy, and criteria, regarding the framework elements and correctness levels. The actions, and the objectives including them, can be distinguished into cognitive, affective, and psychomotor ones according to the domains described in Bloom's taxonomy of learning objectives. The cognitive objectives can be classified in the 24 cells of the taxonomy table described in (Krathwohl, 2002) with a six-column cognition horizontal dimension and a four-row knowledge vertical dimension. One kind of criteria are the SMART ones (Doran, 1981), that is, to be specific, measurable, achievable, relevant or interesting, and time-bound.

An example of a significant general objective in the learning process is that the learner should advance to reach the limits of their capabilities, acquire metalearning and metacognitive skills, that is learn how to learn and think, and become both an autonomous and a collaborative flexible learner.

Pedagogy: Learning Theories

The theories investigate the way that learning is achieved, adopt certain principles, and propose learning techniques. They are considered of great importance in the design of a learning process. The major classes of learning theories are behaviourism, cognitivism and its variations, constructivism and constructionism, and social theories, which expand on the two latter with several variations (Xafopoulos, 2016).

Although instances of behaviourism, cognitivism, cognitive constructivism and constructionism are also applied in the e-CL process the most influential learning theories underpinning e-CL are social theories, such as, constructivism, communal constructivism, social constructionism, and distributed and situated cognition due to their collaborative background (Holmes & Gardner, 2006). Nevertheless, a blended approach regarding the theories, that is, selecting, or even combining, more than one theory according to each special occasion and objective, is what is usually applied in practice.

Methodology: Introduction

The methodology, taking into account the other supportive elements, and mainly pedagogy and the affective status of the learners, examines and suggests about the learning policies, techniques, and activities (Xafopoulos, 2016). The policies regulate the learning process, for instance, the kind of communication and collaboration inside and among communities, the community size, the level of formality and the prioritization and leadership among stakeholders. Methodological

principles are called strategies. The techniques are proposed, by underpinning learning theories and policies, routes leading to learning, such as the flipped classroom. Barkley, Cross, and Major (2014) proffer a classification of collaborative techniques depending on their focus; for example, performing, problem-solving, discussing, writing, drawing or constructing, and playing.

Methodology: Activities

The activities are implementations of the techniques in a specific learning environment. They can be substantially classified into presenting, involving presentation from the learning environment to the learner, and performing, involving learner performance in the learning environment. Among the numerous other types of activities some important distinctions are between electronic and not, collaborative and not, and linguistic or not. As for the electronic activities they can be further distinguished into online and not. With regard to linguistic activities the presenting include reading and listening, and the performing writing and speaking. In respect of the non-linguistic the presenting activities include the five basic human senses and the performing the physical movements.

A further possible categorisation of the activities would be the one according to the revised Bloom's taxonomy of learning objectives, which could also be applied to characterise activities, especially the cognitive domain taxonomy using the taxonomy table (Krathwohl, 2002). According to the latter taxonomy there are six increasingly difficult cognitive skills. The three lowest, that is remember, understand, and apply, are characterised lower-order thinking skills (LOTS) while the three highest, analyse, evaluate, and create, higher-order thinking skills (HOTS). The latter are most valued in learning environments and a particularly interesting expansion of them are those including multimedia, multiple interaction, collaboration, and HOTS. These can be termed MICH activities.

It is also noteworthy that certain activities, such as, exploratory, may combine several of the previously mentioned types. Moreover, in order that the activities are successful they must include clear and assistive descriptions and instructions regarding the learning process and environment. Finally, support actions could also be considered a kind of activities, which are usually undertaken by the implementation community to support the learning community (Sloep, Hummel, & Manderveld, 2005).

Content

The learning content corresponds to an aspect or subdomain of a learning subject and may expand to whichever area, or domain or discipline or subject, of study. The subdomain may be a simple or complex one. It can be classified into the four kinds of knowledge (Krathwohl, 2002). Furthermore, it should be modelled using a domain knowledge representation scheme. The content creation process must follow certain pedagogical criteria and a decisive factor for it is the learning community knowledge.

The learning content may appear in certain modalities, that is, culturally-shaped perceivable content forms (Bezemer & Kress, 2008). They are text, audio, such as speech, music, or other kind of sound, image, either photograph or graphic, animation, and video. There are three main ways the modalities are perceived by learners corresponding to three human senses; the visual, the auditory, and the tactile or haptic.

To advance learning the design and implementation members must possess, on the one hand, state of the art domain knowledge and, on the other hand, the skill to transform and represent this knowledge to the learning level of the learning community members using modalities and technological media so that the latter community may be assisted to explore this information and construct domain knowledge itself.

FRAMEWORK ELEMENTS

In the e-CL process there are certain framework or contextual elements, which usually play a secondary role in the process. The most basic of them are first time, in terms of point, duration, frequency, and temporal context, second space, in terms of location, area, settings or physical environment, and spatial context, and third society, in terms of sociocultural environment and context. These elements should be modelled in the learning design.

The e-CL approaches can be classified in terms of the framework elements time and space as follows. Firstly, as regards space they can be implemented either in a learning, physical or virtual, room, or in a distance, which can be further discriminated into long distance and short distance ones, or both in a learning room and in a distance, either in parallel or in different occasions, called blended or hybrid. The latter are the most suggested course of action for effective learning.

Secondly, regarding time the e-CL approaches can be implemented either in the same time, called synchronous or live or real-time, or in different time, called asynchronous or on-demand, or both in the same and in different time, in parallel or in different occasions, which let be called semi-synchronous.

As regards society, environments where e-CL could be applied are enterprises, the industry, e-commerce, entertainment, the arts, journalism, engineering, e-science, medicine, lower, higher and special education, and research.

e-CL PROCESS

Process Modelling (ADDIE)

The elements previously mentioned participate in, support, and frame the e-CL process, respectively. This process has been modelled in several ways called learning models. One of the most

frequently followed model is ADDIE (analysis, design, development, implementation, evaluation) (Branch, 2009; Branch & Merrill, 2012), which has several variants (Xafopoulos, 2016). Its acronym comprises the first letters of its five phases. When the phases are implemented through e-technology they can be characterised as e-analysis, e-design, e-development, e-implementation, and e-evaluation.

ADDIE Phases

The five phases of ADDIE, although in general terms are ordered, are not strictly sequential, but there may be several directions among them, such as, iterative cycles or spiral, backwards, overpassing and others, resulting in a dynamically iterative process, which may also be longitudinal (Branch & Merrill, 2012). There is often the need to practice more than one, or even all five, phases concurrently (Branch, 2009).

Analysis

To begin with, the problem, domain, goal, and human elements are specified at a preparatory phase, followed by the analysis phase, where the human elements are analysed and a needs analysis or assessment is conducted, generally specifying the rest learning elements supporting the learning needs. The analyses conducted can be distinguished into qualitative and quantitative ones. The latter are referred to as learning analytics, including prediction, structure discovery, and relationship mining (Baker & Siemens, 2014). In the present focus on e-CL collaborative or social learning analytics are of special importance (Ferguson & Shum, 2012). A useful graphical representation of their results capturing the learning dynamics are learning sociograms.

Design

At the design phase, where the analysis phase is taken into account, there can be a distinction of three subphases. First, the modelling design sub-

phase, where human and community models or profiles, whose descriptive form is called persona, as well as, policies, general and specific objectives, and content are designed based on underpinning learning theories. Second, the implementation design subphase, where implementation issues related to communities, such as group formation, technologies, techniques, activities, and framework elements are designed.

Third, the assessment design subphase, where assessments, that is, evaluative criteria and activities of the learning procedure, as well as other kind of evaluations, are designed. These assessments should be aligned with the learning objectives and activities. This correspondence or alignment among objectives or outcomes, activities, and assessments is termed 'constructive alignment' (Biggs & Tang, 2011). A detailed formal definition of a learning design is called a prototype. Finally, it should be noted that the design phase and its subphases are not strictly sequential but usually follow a spiral form, being reflective.

Development and Implementation

At the development phase the designed elements are practically deployed and at the implementation phase the developed e-CL process is performed with specific learning elements. Analyses like the ones performed in the analysis phase are usually performed in the implementation phase with an evaluative instead of a preparatory purpose. When these analyses are employed continuously over a time period they are referred to as tracking. Their result may be used as feedback to readjust the learning environment.

Evaluation

At the evaluation phase several evaluations of the implementation are executed. The analyses performed in the analysis and implementation phase are evaluatively extended and complemented in the evaluation phase. The evaluation appears in three forms. Generic evaluation, with general

characterisations of the outcome; assessment, checking against determined assessment criteria; and feedback, offering explanatory assistive and improving comments or reports. An assessment example is learning performance assessment, that is, the estimation of the degree of learning outcomes achievement against the assessment criteria, often including marking schemes.

A further usual categorisation of evaluation, according to the time it is performed, is into the three types of diagnostic or placement, formative, and summative, which evaluate before, during, and after the learning process, respectively (Miller, Linn, & Gronlund, 2009). Diagnostic evaluation is discriminated in some contexts to express the diagnosis of difficulties during learning (Miller et al., 2009).

The evaluation lately often is proposed to include formative feedback to the communities and actually does, so that they identify learning difficulties and overcome them. The time and way of feedback, especially formative, is of great importance for supporting learning. For instance, timely personalised formative feedback enhances the learning experience.

Another commonly used classification of evaluation, according to the way it is performed, is into criterion-referenced, normative, and ipsative. The first evaluates a learner's, or even a learning group, performance against specific evaluation criteria, the second against the other learners' performance, and the third against the learner's previous performance.

Design Element Hierarchy

What is of utmost importance in the design process is the hierarchy of the elements of the learning process. In other words, determining the amount of design effort and care to be provided in the interests of each of the human, the supportive, and the framework elements. For example, according to the hierarchy the learning design approach may be implementation-community centred, learning-community centred, content centred,

outcome centred, method centred, or blended (Selwyn, 2011).

Another serious concern in the design process is whether, how, and how much each of the participants in the learning process will participate in the design process. A highly esteemed among designers learning design approach, involving learners at different stages of the design process of a learning environment and supporting e-CL, is participatory design (PD) and its collaborative variant collaborative participatory design (CPD) (Xafopoulos, 2016).

FUTURE DIRECTIONS

The current 21st century has been predominantly termed in the sociological literature as the age of late modernity, which is characterised by uncertainty and worldwide radical changes. Therefore, it is too challenging to definitely predict and present the future of e-CL, which, nevertheless, appears to be very promising.

The future trends that can be diagnosed given the current circumstances are web integration; worldwide open access; enhanced learning analytics; learner-centred and partnering design; interactive and experiential, multimedia, situated and immersive, deep and multicultural learning; e-inclusion and personalisation; frequent personalised formative feedback; virtual learning worlds and MICH activities; and augmented online collaboration. Another emerging direction is the consideration of the learning environment as a self-organised, adaptable, and evolving ecosystem including diverse collaborating or antagonising entities or elements (Dong et al., 2009).

CONCLUSION

This article explored the highly topical and key issue of e-CL in a holistic overview. First of all, a clarification of the term and context of e-CL was provided comparing it with similar concepts and

analysing relevant concepts. The terms learning, knowledge, collaboration, collaborative learning, technology, e-technology, and e-CL were explored. Second, the human elements of e-CL were examined, together with their roles and aspects in the learning process. These elements were classified into main communities and secondary supportive communities.

Third, the supportive learning elements were visited, with a focus on each of them. Initially, on the technology, with a representative presentation of its numerous type of media, applications, environments, and infrastructure. Consequently, the article focused on the pedagogy, with an examination of the learner modelling, the learning objectives, and the major learning theories identifying the e-CL supportive ones. Furthermore, the methodology was visited, with an emphasis on the activities. And finally, the learning content and its modalities were explored.

Fourth, the framework elements time, space, and society were described and a classification of e-CL approaches according to them was provided. Fifth, the e-CL process was examined following the ADDIE model, analysing its five phases, and also visiting the design element hierarchy. Finally, future directions of e-CL were considered. Throughout the article key and significant approaches, methods, and terms were pinpointed and concisely developed.

Concluding, in the light of this study e-CL appears as a promising field for research, design, innovation, as well as creative and effective implementation. It proffers some of the most contemporary opportunities and challenges in terms of learning, collaboration, and technology, and therefore requires rigorous research, study, and experience to leverage its facilities.

REFERENCES

- Arvaja, M., Häkkinen, P., & Kankaanranta, M. (2008). Collaborative Learning and Computer-Supported Collaborative Learning Environments. In J. Voogt & G. Knezek (Eds.), *International Handbook of Information Technology in Primary and Secondary Education* (Vol. 20, pp. 267–279). Springer, US. doi:10.1007/978-0-387-73315-9_16
- Baker, R., & Siemens, G. (2014). Educational Data Mining and Learning Analytics. In R. K. Sawyer (Ed.), *Cambridge handbook of the learning sciences* (2nd ed.; pp. 253–274). New York, NY: Cambridge University Press. doi:10.1017/CBO9781139519526.016
- Barkley, E. F., Cross, K. P., & Major, C. H. (2014). *Collaborative Learning Techniques: A Handbook for College Faculty* (2nd ed.). San Francisco, CA: John Wiley & Sons.
- Bezemer, J., & Kress, G. (2008). Writing in Multimodal Texts: A Social Semiotic Account of Designs for Learning. *Written Communication*, 25(2), 166–195. doi:10.1177/0741088307313177
- Biggs, J., & Tang, C. (2011). *Teaching for Quality Learning at University* (4th ed.). Maidenhead, UK: The Society for Research into Higher Education and Open University Press.
- Branch, R. M. (2009). *Instructional Design: The ADDIE Approach*. New York, NY: Springer. doi:10.1007/978-0-387-09506-6
- Branch, R. M., & Merrill, M. D. (2012). Characteristics of Instructional Design Models. In R. A. Reiser & J. V. Dempsey (Eds.), *Trends and Issues in Instructional Design and Technology* (3rd ed., pp. 8–16). Boston, MA: Pearson Education.

- Cameron, R., & Harrison, J. L. (2012). The inter-relatedness of formal, non-formal and informal learning: Evidence from labour market program participants. *Australian Journal of Adult Learning*, 52(2), 277–309.
- Dong, B., Zheng, Q., Yang, J., Li, H., & Qiao, M. (2009). *An E-learning Ecosystem Based on Cloud Computing Infrastructure*. Paper presented at the International Conference on Advanced Learning Technologies ICALT 2009, Riga, Latvia. doi:10.1109/ICALT.2009.21
- Doran, G. T. (1981). There's a SMART way to write management's goals and objectives. *Management Review*, 70(11), 35–36.
- Ertl, B. (2008). E-Collaborative Knowledge Construction. In N. Kock (Ed.), *Encyclopedia of E-Collaboration* (pp. 233–239). Hershey, PA: Information Science Reference. doi:10.4018/978-1-59904-000-4.ch036
- Ferguson, R., & Shum, S. B. (2012). *Social Learning Analytics: Five Approaches*. Paper presented at the Proceedings of the International Conference on Learning Analytics and Knowledge LAK 2012.
- Garrison, D. R., Anderson, T., & Archer, W. (1999). Critical Inquiry in a Text-Based Environment: Computer Conferencing in Higher Education. *The Internet and Higher Education*, 2(2), 87–105. doi:10.1016/S1096-7516(00)00016-6
- Good, J., & Robertson, J. (2006). CARSS: A framework for learner-centred design with children. *International Journal of Artificial Intelligence in Education*, 16(4), 381–413.
- Holmes, B., & Gardner, J. (2006). *E-learning: Concepts and Practice*. London, UK: Sage.
- Kock, N. (2008). A Basic Definition of E-Collaboration and its Underlying Concepts. In N. Kock (Ed.), *Encyclopedia of E-Collaboration* (pp. 48–53). Hershey, PA: Information Science Reference. doi:10.4018/978-1-59904-000-4.ch008
- Krathwohl, D. R. (2002). A Revision of Blooms Taxonomy: An Overview. *Theory into Practice*, 41(4), 212–218. doi:10.1207/s15430421tip4104_2
- Laurillard, D. (2012). *Teaching as a design science: Building Pedagogical Patterns for Learning and Technology*. New York, NY: Routledge.
- Masterman, L., & Vogel, M. (2007). Practices and processes of design for learning. In H. Beetham & R. Sharpe (Eds.), *Rethinking Pedagogy for a Digital Age: Designing and delivering elearning* (pp. 52–63). London, UK: Routledge.
- Miller, M. D., Linn, R. L., & Gronlund, N. E. (2009). *Measurement and Assessment in Teaching* (10th ed.). Upper Saddle River, NJ: Merrill.
- Mishra, P., & Koehler, M. (2006). Technological Pedagogical Content Knowledge: A Framework for Teacher Knowledge. *Teachers College Record*, 108(6), 1017–1054. doi:10.1111/j.1467-9620.2006.00684.x
- Scardamalia, M., & Bereiter, C. (2006). Knowledge Building: Theory, Pedagogy, and Technology. In R. K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (pp. 97–115). Cambridge, UK: Cambridge University Press.
- Selwyn, N. (2011). *Education and Technology: Key Issues and Debates*. London, UK: Continuum.
- Sloep, P., Hummel, H., & Manderveld, J. (2005). Basic Design Procedures for E-learning Courses. In R. Koper & C. Tattersall (Eds.), *Learning Design: A Handbook on Modelling and Delivering Networked Education and Training* (pp. 139–160). Berlin: Springer. doi:10.1007/3-540-27360-3_8
- Wenger, E. (1998). *Communities of Practice: Learning, Meaning, and Identity*. Cambridge, UK: Cambridge university press. doi:10.1017/CBO9780511803932

Xafopoulos, A. (2016). E-Collaborative Learning (e-CL): Overview and Proposals. In A. Kok & H. Lee (Eds.), *Cultural, Behavioral, and Social Considerations in Electronic Collaboration* (pp. 114–146). Hershey, PA: Business Science Reference. doi:10.4018/978-1-4666-9556-6.ch006

KEY TERMS AND DEFINITIONS

ADDIE Model: A widely used framework which models the evolution of a learning process following spirally, that is revisiting the previous, the five phases of analysis, design, development, implementation, and evaluation.

Collaborative Participatory Design (CPD): A learning design approach where the learners collaborate potentially in every phase of the dynamically iterative and reflective learning process informing its design.

Computer-Supported Collaborative Learning (CSCL): The communicative shared-knowl-

edge building process with shared-knowledge building goals supported by a computer network.

E-Collaborative Learning (E-CL): The communicative shared-knowledge building process with shared-knowledge building goals using networked electronic devices.

Learning Design: The design of a learning process in three spirally-followed phases called modelling design, implementation design, and assessment design.

Multimedia Multi-Interactive Collaborative Higher-Order Thinking Skills (MICH) Activities: Learning supporting activities using multiple modalities, being interactive in many ways, being collaborative, and employing analytical, evaluative, and synthetic skills.

Technological Pedagogical Content Knowledge (TPACK) Model: A framework arguing that a combination of three kinds of knowledge is needed by the design and implementation members; technological, pedagogical, including methodological, and content knowledge.

Learning With Mobile Devices



Helen Crompton

Old Dominion University, USA

John Traxler

University of Wolverhampton, UK

INTRODUCTION

Mobile devices are being used to extend learning in both the global north and the global south. The launch of the International Journal of Mobile and Blended Learning is one of several indicators that mobile learning globally is reaching a critical and sustainable momentum and identity. The past decade has seen a host of pilots and initiatives across sectors and across countries. Mobile learning is being used to extend pedagogies to develop new ways of learning more aligned to empirical understandings of how students learn. Mobile learning is also taking learning to individuals, communities and countries where access to learning was challenging or problematic.

Environmental factors have meant that this development has often been haphazard. As researchers try to examine these new learning opportunities they are faced with a set of methodologies that were developed for use with tethered technologies with learning and evaluation methods being static (Sharples, Sanchez, Milrad, & Vavoula, 2009). In mobile learning context can vary significantly. This has expanded the field of mobile learning to have researchers examining learning that happens in a formal typical educational setting, non-formal non-intended serendipitous environments (Crompton; 2013a), and informal atypical settings (e.g. museums and science centers: National Research Council, 2009). With learning that happens in personal contexts, the researcher has the added difficulty in collecting valid research data

without interfering with the learning happening in those contexts Beale's (2007). This is becoming increasingly difficult in recent years with the pervasive ubiquitous nature of the devices today.

As the scholarly understanding of mobile learning is still relatively new and emerging, the mobile learning community is also now faced with broader challenges of scale, durability, equity, embedding and blending in addition to the earlier and more specific challenges of pedagogy and technology, but these developments take place in the context of societies where mobile devices, systems and technologies have a far wider impact than just mobile learning as it is currently conceived.

BACKGROUND

The concept of learning with small portable computers was developed by Alan Kay in 1972. Since that early conception, scholars, such as Traxler, Sharples, and Soloway are the pioneering scholars who have paved the way to a better understanding of the philosophical, pedagogical, and conceptual underpinnings of mobile learning today. Kay began with the initial idea of a portable device for learning. Traxler, Sharples and colleagues have explored the emerging theoretical frameworks of mobile learning to provide us with a better understanding of this field. Soloway and Norris have focused their work on how the affordances of mobile learning can extend traditional classroom pedagogies.

Defining Mobile Learning

We need to define what we mean by ‘mobile learning’, not merely as a way of establishing a shared understanding but also as a way of exploring the evolution and direction of mobile learning and as a way of identifying the community of practitioners and researchers. In discussing how we define mobile learning we address many wider issues in terms of explaining, understanding and conceptualising it.

‘Mobile learning’ is certainly not merely the conjunction of ‘mobile’ and ‘learning’; it has always implicitly meant ‘mobile e-learning’ and its history and development have to be understood as both a continuation of ‘conventional’ e-learning and a reaction to this ‘conventional’ e-learning and its perceived inadequacies and limitations. Over the last ten or so years this ‘conventional’ e-learning has been exemplified technologically by the rise of virtual learning environments (VLEs) and the demise of computer assisted learning (CAL) ‘packages’, and pedagogically by the rise of social constructivist models of learning over the behaviourist ones, by the growth of the learning object approach, by expectations of ever increasing multi-media interactivity and of ever-increasing power, speed, functionality and bandwidth in networked PC platforms. These are some of the points of departure for mobile learning. They refer back to ‘conventional’ e-learning and perhaps this is the mark of early ‘mobile learning immigrants’ and not the mark of the growing number of ‘mobile learning natives’.

We have to recognise that attempts at identifying and defining mobile learning grow out of difference, out of attempts by emergent communities to separate themselves from some older and more established communities and move on from perceived inadequate practices. Interestingly, at the first mLearn conference in the spring of 2002, in Birmingham UK, a key-note speaker predicted that mobile learning would have a separate identity for perhaps five years before blending into general e-learning. This has still yet to happen and mobile

learning continues to gain identity and definition rather than lose it.

Irrespective of the exact definition, personal mobile and connected technologies, including handheld computers, personal digital assistants, camera phones, smartphones, graphing calculators, personal response systems, games consoles and personal media players, are ubiquitous in most parts of the world and have led to the development of ‘mobile learning’ as a distinctive but ill-defined entity (see for example the reviews by Cobcroft 2006, and Naismith et al. 2004).

Early approaches at defining mobile learning focused on technology, for example saying it was “any educational provision where the sole or dominant technologies are handheld or palmtop devices” (Traxler, 2005), or on the mobility of the technology, describing mobile learning as, “elearning through mobile computational devices: Palms, Windows CE machines, even your digital cell phone.” (Quinn, 2000). Another view of mobile learning says it involves: “Any sort of learning that happens when the learner is not at a fixed, predetermined location, or learning that happens when the learner takes advantage of learning opportunities offered by mobile technologies” (O’Malley et al., 2003), whilst Desmond Keegan took a similar position in 2005, saying that the focus should be on mobility and mobile learning should be restricted to learning on devices which a lady can carry in her handbag or a gentleman can carry in his pocket. He defined mobile learning as ‘the provision of education and training on PDAs/palmtops/handhelds, smartphones and mobile phones and the characteristics of mobile learning is that it uses devices:

- Which citizens are used to carrying everywhere with them,
- Which they regard as friendly and personal devices,
- Which are cheap and easy to use,
- Which they use constantly in all walks of life and in a variety of different settings, except education.” (Keegan, 2005:3)

The MoLeNET initiative, a £6m programme across the UK vocational sector, still takes this approach, defining mobile learning as, “exploitation of ubiquitous handheld hardware, wireless networking and mobile telephony to enhance and extend the reach of teaching and learning” (MoLeNET, 2007). These definitions were too technocentric and imprecise. The transience and diversity of the devices, systems and platforms means that these definitions are also highly unstable. They merely put mobile learning somewhere on e-learning’s spectrum of portability (ending perhaps in ubiquitous, pervasive and wearable learning).

Whilst these attempts at definition use specific technical attributes to consolidate a definition of mobile learning in order to help us reason about it, other technical attributes, notably connectivity, usability and latency, have the very opposite effect and disrupt the notion that there is such a thing as mobile learning as an artifact of mobile technologies.

The uncertainty about whether laptops and Tablets deliver mobile learning – because of the lack of spontaneity in carrying them and starting them up - illustrates the difficulty with this kind of definition and the emergence of the UMPC (Ultra-Mobile PC) and netbook formats and the low cost XO systems will further trouble this boundary (Crompton, 2014a). They do however hint at the underlying challenge, that of conceptualising mobile learning in a way that recognises its origins and practices in specific technological systems but is abstract enough to be durable and to act as a stable platform for theorising about education and learning.

Crompton (2013b) looked beyond just the technical attributes to develop a definition of mobile learning from consolidating what she described as the four central constructs to mobile learning: pedagogy, technological devices, context, and social interactions. This resulted in mobile learning being defined as

“learning across multiple contexts, through social and content interactions, using personal

electronic devices” (p. 4). This is the most accepted current definition of mobile learning at this time, but will undoubtedly be modified in the future within this rapidly changing field.

CURRENT SCIENTIFIC KNOWLEDGE IN MOBILE LEARNING

There have been a few pioneering scholars in the field of mobile learning (e.g., Kay, Traxler, Sharples, Soloway & Norris) who have extended the literature in mobile learning. In the 1970s, Kay had the idea of developing a small portable device that students could learn from. Such a device did not prove to be feasible at that time, but due to more recent technological advancements, the mobile learning of today bears a good resemblance to Kay’s initial ideas.

As mobile learning started moving into fruition, scholars, such as Traxler and Sharples have provided us with an evolving theoretical framework that has enabled us to use gain a better understanding of this new form of learning. Soloway and Norris have focused on how mobile learning is changing the way students learn. While, the work of these scholars has significant overlap, for the purpose of this chapter, these two strands (theories and pedagogy) have been dichotomized to provide a more in-depth look at each of these topics.

Kay’s Dynabook

In the 1970’s Kay (1972) created the concept model of the Dynabook, which he described as a personal computer for children of all ages. This personal computer would have a number of unique features:

Imagine having your own self-contained knowledge manipulator in a portable package the size and shape of an ordinary notebook. Suppose it had enough power to outrace your senses of sight and hearing, enough capacity to store for later retrieval thousands of page-equivalents of refer-

Table 1. Emergent categories from mobile learning

1	Technology-driven mobile learning – a specific technological innovation is deployed to demonstrate technical feasibility and pedagogic possibility, perhaps the iPhone and iPad.
2	Miniature but portable e-learning – mobile, wireless and handheld technologies are used to re-enact approaches and solutions found in ‘conventional’ e-learning, perhaps porting an established e-learning technology, such as the VLE or e-portfolio, onto mobile devices.
3	Connected classroom learning – the same technologies are used in a classroom setting to supported static collaborative learning, perhaps connected to other classroom technologies; personal response systems, graphing calculators, PDAs linked to interactive whiteboards etc.
4	Mobile training and performance support – the technologies are used to improve the productivity and efficiency of mobile workers by delivering information and support just-in-time and in context for their immediate priorities, roles and duties
5	Large-Scale Implementation – the deployment of mobile technologies at an institutional or departmental level to learn about organisational issues.
6	Inclusion, assistivity and diversity – using assorted mobile and wireless technologies to enhance wider educational access and participation, for example personal information management for students with dyslexia.
7	Informal, personalised, situated mobile learning – the same core technologies are enhanced with additional unique functionality, for example location-awareness or video-capture, and deployed to deliver educational experiences that would otherwise be difficult or impossible; for example, informal context-aware information in museum spaces.
8	Remote, rural and development mobile learning – the technologies are used to address environmental and infrastructural hurdles to delivering and supporting education where ‘conventional’ e-learning technologies would fail. This classification is not purely theoretical. It has implications for the objectives and methods of evaluation and for the techniques and objectives for implementation; it may also imply the presence or otherwise of different underlying models of pedagogy and learning.

ence materials, poems, letters, recipes, records, drawings, animations, musical scores, waveforms, dynamic simulations, and anything else you would like to remember and change. (Kay & Goldberg, 1977/2001, p. 167).

These ideas were a little beyond that time and the Dynabook was never created; however, those seeds of ideas were sown and mobile learning of today has surpassed Kay’s initial ideas of learning with a mobile device.

Theories of Mobile Learning

As mobile learning emerged from Kay’s ideas, Traxler, Sharples, and colleagues have further explored the theoretical underpinnings of their emerging field. Kukulska-Hulme and Traxler (2007) saw a number of emergent categories that come from reviewing the mobile learning literature, see Table 1.

These may be innovative or conservative technically or pedagogically by virtue of their place in the classification. Of course, this attempt to define

mobile learning by making instances – definition by denotation rather than by connotation as we tried earlier – is potentially problematic since in choosing the instances we create a circular definition but it nevertheless takes us a bit further forward. Niall Winters (2006) provides a similar taxonomy which gives us an additional perspective on what might characterise different types of mobile learning, saying, “Current perspectives on mobile learning generally fall into the following four broad categories: (1) Technocentric. This perspective dominates the literature; (2) Relationship to e-learning. This perspective characterises mobile learning as an extension of e-learning; (3) Augmenting formal education; and (4) Learner-centered.” This echoes our earlier points. Another classification of mobile learning that might help us towards a definition is due to Naismith et al. (2004) who suggest that mobile technologies can relate to six types of learning, or ‘categories of activity’, namely behaviourist, constructivist, situated, collaborative, informal/lifelong, and support/coordination. The mobile learning may be manifest in the following ways: For behaviourist-type

activity, it is the quick feedback or reinforcement element, facilitated by mobile devices, that is most notable; For constructivist activity, mobile devices enable immersive experiences such as those provided by mobile investigations or games; For situated activity, learners can take a mobile device out into an authentic context, or use it while moving around a context-aware environment in a specially equipped location such as a museum; For collaborative learning, mobile devices provide a handy additional means of communication and a portable means of electronic information sharing; For informal and lifelong learning, mobile devices accompany users in their everyday experiences and become a convenient source of information or means of communication that assists with learning, or records it on the go for future consultation; Support, or coordination of learning and resources, can be improved by the availability of mobile technologies at all times for monitoring attendance or progress, checking schedules and dates, reviewing and managing, activities that teachers and learners engage in at numerous times during the day.

An attempt by Sharples, Taylor and Vavoula (2005:4) suggested that a theory of mobile learning should be assessed against the following five criteria as in effect their perspective on the defining characteristics of mobile learning: (1) Is it significantly different from current theories of classroom, workplace or lifelong learning? (2) Does it account for the mobility of learners? (3) Does it cover both formal and informal learning? (4) Does it theorise learning as a constructive and social process? (5) Does it analyse learning as a personal and situated activity mediated by technology?

Ann Jones (Jones et al 2006) makes a similar contribution based on the motivational or affective aspects of mobile learning as defining characteristics. These are both important in themselves and often cited anecdotally as major factors behind decisions to deploy a mobile learning strategy. They are control (over goals), ownership, fun,

communication, learning-in-context, and continuity between contexts.

There have also been attempts to define mobile learning and thus to distinguish it from 'conventional' e-learning in terms of the learners' experiences. One view (Traxler, 2006) in looking at characterisations of mobile learning found in the literature finds words such as 'personal', 'spontaneous', 'disruptive' 'opportunistic', 'informal', 'pervasive', 'situated', 'private', 'context-aware', 'bite-sized' and 'portable'.

These are contrasted with words from the literature of 'conventional' e-learning such as 'structured', 'media-rich', 'broadband', 'interactive', 'intelligent' and 'usable'. We can use these to make a blurred distinction between mobile learning and 'conventional' e-learning. However, this distinction, based on the learners' experiences of the two different modes of learning, misses a greater distinction. 'Conventional' e-learning nearly always takes place in situations where the learners' time and space have been dedicated and committed to e-learning, facing their computer, sat with their back to the world, with e-learning taking centre-stage (Crompton, 2013a).

Mobile learning in the sense that we have been talking about it takes place woven into a host of daily tasks, places, groups, interactions and situations. The associations that learners generally have with these two technologies, the static and the mobile, must also be vastly different. This distinction based around learner experience is however not only blurred but in part is also only temporary.

Many of the virtues of 'conventional' e-learning are the virtues of the power of its technology (and the investment in it) and these virtues will be accessible to mobile devices too as market forces drive improvements in memory size, interface design, processor speed, battery life and connectivity bandwidth. Nevertheless, this approach underpins a definition of mobile learning in terms of the learners' experiences and an emphasis on 'ownership', informality, spontaneity, mobility

and context that will always be inaccessible to 'conventional' e-learning. We should add that the reported learner experience of mobile learning may depend on where the specific project fits into the earlier taxonomy.

The communities cohering around mobile learning may still feel the need for a theory of mobile learning as well as a definition, for example because of the ability of theory to define a research agenda or produce useful predictions and generalizations (although in a postmodern era, the role of theory as an informing construct is under threat). Such a theory may however be particularly problematic since mobile learning is an inherently 'noisy' phenomenon where context is everything and confounding variables abound, and if theory is something generated by abstracting upwards from practice and experience, then perhaps mobile learning has yet to reach the critical mass of experience and practice that justify such abstraction and has been too fragmented to justify transferable generalisations.

The work of Kuhn (1962) on the structure of intellectual change provides some insights into the role of 'theory' in relation to the professional activities of researchers (though not one without its critics) 'Conventional' e-learning has certainly gained credibility and status from the work of, for example, Laurillard (2002) and Salmon (2000) but there is currently insufficient work in mobile learning generally to underpin much theory building. Theories of 'conventional' e-learning rest on the experience of stable technology platforms; the dominant and enduring nature of Windows, QWERTY, IP, HTML and WWW means that theorising about 'conventional' e-learning can take place in a technology environment that is consistent, homogeneous and transparent – the technology no longer gets in the way. The technology platform upon which mobile learning theory might rest is by comparison volatile, inconsistent and haphazard and so must impede the work of understanding mobile learning itself. Mobile learning needs a 'theory of technology'. We could argue that the mobile learning community in look-

ing for theory is – to oversimplify – faced with three different options and dilemmas: (1) Import theory from 'conventional' e-learning and worry about transferability, (2) Develop theory *ab initio* locally and worry about validity, and (3) Subscribe to some much more general and abstract theory and worry about specificity and granularity.

Diana Laurillard's recent recognition of the impact of mobility and mobile technologies on the Conversational Framework (Laurillard, 2007) is an example of taking the first option. She discusses the possibilities of increasing interaction between the learner and the environment but also how problematic or unproductive this might be in informal learning or unsupervised learning (for example, in museum spaces) where a teacher is neither in a position to set appropriate tasks nor to provide meaningful feedback.

This is within more general remarks about the use of the Conversational Framework to support "a rigorous approach to working out how to support all the component learning activities, in remote locations, with learners guided only by the tasks set, the information available online, the characteristics of the world they are in, and peer support." This is a case of mobile learning looking to challenge and extend an accepted e-learning theory. The emerging theories of 'connectivism' (Siemens, 2004) and 'navigationism' (Brown, 2005) are nearer to the second option.

People are now learning "through communities of practice, personal networks, and through completion of work-related tasks" in an environment in which "know-how and know-what is being supplemented with know-where (the understanding of where to find knowledge needed)" (Siemens, 2005). Thirdly, it is fair to say that many of the more theoretically inclined members of the mobile learning community (see for example Sharples et al, 2005) subscribe to versions of Yrjö Engeström's 'Activity Theory' (1987) and this would be the most obvious example of the third option, an analysis of much purposive human activity. Engeström and his colleagues refers to Activity Theory as a "commonly accepted name

for a line of theorizing and research initiated by the founders of the cultural-historical school of Russian psychology.” whilst others (Er and Kay, 2005) say that the underlying philosophy of the theory is to explain human activity and behaviour.

Learning is analysed as a cultural-historical activity system, mediated by tools that constrain and support the learners in their goals of transforming their knowledge and skills. This is not an attempt to explain or assess Activity Theory but merely to position it as a broad and abstract account of more than just learning and technology. Returning to the issue of definition, Josie Taylor (2006) comes at it from a high level, seeing the question as whether ‘mobile learning’ signified a) learning mediated by mobile devices or mobility of learners (regardless of their devices), or b) mobility of content/resources in the sense that it can be accessed from anywhere.

In this account her audience preferred the broader concept of learning taking place in the ‘mobile age’, rather than the use of the narrower term ‘mobile learning’. Focusing on defining mobile learning in an age where actually nothing stays still is perhaps missing the point; the question, ‘what is mobile learning?’ must be replaced by the questions, ‘what is learning in a mobile age?’ or perhaps ‘what is mobile learning?’ Our societies are changing as mobile devices, systems and technologies become universally owned, accepted and used, and as a consequence the meaning and significance of learning are changing too. Perhaps ‘learning with mobile devices’ was adequate all along.

Extending Pedagogies

In the early 2000’s, Soloway and Norris (2003a, 2003b, & 2004) called for educators to see beyond 1:1 computing with laptops and to consider the use of personal computers that can fit into the palm of their students’ hands. These devices would revolutionise learning. This change is evident today as mobile learning is extending the boundaries of traditional pedagogies with learning that is

personalized, contextualized, and unrestricted by time and environment (Crompton, 2013a; 2014).

Mobile devices, and their technologies and systems, are eroding established notions of time as a common structure that had previously underpinned social organisation and the consensual understanding of the world. Time-keeping is being replaced by the ‘approx- meeting’ and the ‘multi-meeting’ (Plant, 2000), ‘socially negotiated time’ (Sørensen et al, 2002), the ‘microcoordination of everyday life’ alongside the ‘softening of schedules’ (Ling, 2004) afforded by mobile devices and Nyíri (2006:301) says, “with the mobile phone, time has become personalized”. Whereas previously our social and business relations had to be organized and synchronised by absolute clock time, now mobile technologies allow us to renegotiate meetings and events on-the-fly.

Mobile devices are also eroding physical place as a predominant attribute of space. It is being diluted by “absent presence” (Gergen, 2002), the phenomenon of physically co-located groups all connected online elsewhere—everyone in the room is online elsewhere - and “simultaneity of place” (Plant, 2002) created by mobile phones, a physical space and a virtual space of conversational interaction, and an extension of physical space, through the creation and juxtaposition of a mobile “social space”. Ironically, many conversations on the mobile phone, the device to demolish locatedness, start with, “I’m on the train”. Clearly we are still adjusting to the disembodied world of mobility.

Mobile devices are reconfiguring the relationships between spaces, public ones and private ones, and the ways in which these are penetrated by mobile virtual spaces. This is documented in the literature of mobilities, for example Plant (2000), Katz and Aakhus (2002), Ling (2004) and Brown et al (2004). Virtual space, and its tasks and relationships, that used be occupied by people sat down, monopolising their attention and partitioning them from the other people and the physical spaces around them moving, now moving amongst these other people and spaces and amongst other tasks and relationships.

This is accompanied by what goes on in those spaces; Cooper (2002) says that the private “is no longer conceivable as what goes on, discreetly, in the life of the individual away from the public domain, or as subsequently represented in individual consciousness”, Sheller and Urry (2003) argue “that massive changes are occurring in the nature of both public and private life and especially of the relations between them,” and Bull (2005) writing about the iPod says “The use of these mobile sound technologies informs us about how users attempt to ‘inhabit’ the spaces within which they move.

The use of these technologies appears to bind the disparate threads of much urban movement together, both ‘filling’ the spaces ‘in-between’ communication or meetings and structuring the spaces thus occupied.” Earlier work on the Sony Walkman came to similar conclusions, “the Walkman disturbed the boundaries between the public and private worlds” (Du Gay et al 1997: 115) Mobile devices are redefining discourse and conversation. Goffman (1971), for example, noted the phenomenon of ‘civil inattention’, where in certain situations it is customary not only to not speak to others but to avoid looking directly at others. This management of gaze is one way in which the boundary between public and private is negotiated and is now often a characteristic of creating a private space for mobile phone conversations in a public setting; a similar concept is the ‘tie-sign’, those signs that keep a face-to-face encounter live and ‘in play’ whilst servicing an interruption caused by a mobile phone call. The recipient of the call is obliged to “play out collusive gestures of impatience, derogation, and exasperation” according to Goffman. Murtagh (2002) describes a wide set of non-verbal actions and interactions with the mobile phone in public, and these are part of a wider transformation of discourse and social interaction as society engages with mobile technologies.

Alongside these evolving patterns of behaviour, mobile devices help communities and sub-cultures define themselves by affording new forms of

language, txt-speak being the obvious example of a language that helped its original users mark themselves out as different from non-users, usually their parents. Mobile devices are creating communities and groupings, sometimes transient and virtual ones, arguably at the expense of existing and traditional ones, captured in Howard Rheingold’s (2003) defining book. With these groupings come new norms, expectations, ethics and etiquettes (for example, see Ling (1997, 2004) for a discussion of ethics in a mobile context; and shifting ideas about the self and identity.)

Geser (2004:11) points out that, “the cell phone helps to stay permanently within the closed social field of familiar others: thus reinforcing a unified, coherent individual identity.” These are the contexts of ‘learning in a mobile age’. Clearly there is much here for educators and learning technologists to digest, from the changes in expectations and practices of interpersonal behaviour in the tutorial, seminar and lecture to the expectations of universities, schools and colleges in specifying the times and places to deliver education. Obviously, mobile devices, systems and technologies are also dramatically changing the economy, thus changing our ideas about artifacts and assets, and the jobs and organisations needed to create, distribute and trade them, and dramatically changing the nature of work itself. Educators must digest these too in their role preparing learners for employment.

Mobile devices, systems and technologies also have a direct and pervasive impact on knowledge itself, and how it is generated, transmitted, owned, valued and consumed in our societies. At the most superficial level, they do finally deliver on the ‘anywhere, anytime’ promise and apparently on other slogans too, namely ‘just-in-time’, ‘just-for-me’, ‘here-and-now’. These may be however less simple and benign than they seem. Firstly, knowledge is not an absolute. It has been argued that it is socially determined and socially constructed but it has also always been mediated by its container, its medium, its repository. Mike Sharples (2005:1) says, “Every era of technology has, to some extent, formed education in its own

image. That is not to argue for the technological determinism of education, but rather that there is a mutually productive convergence between main technological influences on a culture and the contemporary educational theories and practices.

The teacher is no longer seen as the holder of information but students are now provided with the opportunity to access this information for themselves (Norris & Soloway, 2011). In the era of mass print literacy, the textbook was the medium of instruction, and a prime goal of the education system was effective transmission of the canons of scholarship. During the computer era of the past fifty years, education has been re-conceptualised around the construction of knowledge through information processing, modelling and interaction. For the era of mobile technology, we may come to conceive of education as conversation in context, enabled by continual interaction through and with personal and mobile technology.”

The earliest formats, the book and the lecture, originally constrained knowledge to a linear format, the book having at least usually some facilities of graphics, review and organisation and more recently, computers provided web-based hyper-linked information. This was delivered with greater multi-media richness than books but in smaller chunks governed by the heuristics of usability and increased nonlinear navigational complexity. Mobile devices can now deliver information in far smaller chunks but with a vastly increased navigational overhead.

Clearly, these different formats must each have an effect on information and on knowledge in their different ways, on what is accessible and what is valued. With mobile devices, there is a concern that they serve up vast amounts of information and knowledge in small disconnected and trivial chunks. As T. S. Elliott (1934) said, “Where is the Life we have lost in living? Where is the wisdom we have lost in knowledge? Where is the knowledge we have lost in information?” Search engines and

knowledge bases can now serve up information that is uniquely customised to the user and their context, meaning their history, their location, their interests, their preferences and their environment.

Whilst this level of personalization seems attractive and desirable, there is also concern that knowledge and information become individualised, a ‘neo-liberal nightmare’ where each user exists in their own unique information world, fragmenting learners in a ‘fragmented society’, to use Bauman’s (2001) phrase in an accurate but narrower sense than he intended. User-generated content, meaning in user-generated knowledge and user-generated information, is widely available on mobile technologies. Google and Wikipedia, both now location-specific, are examples and they both allow learners control over what they learn, unmediated by any formal institutional learning.

They also allow learners to participate in creating learning through their contributions. This can take place through such systems as Wikipedia but most conspicuously with mobile technologies through the activity of citizen-journalism (Owen, 2005), where members of the public using camera-phones capture images of breaking news and post them straight onto shared file-spaces such as Flickr or YouTube. Journalism has been called the first draft of history and here we see it generated without the intervention of professional journalists or centralised and controlling organisations, perhaps from the perspectives of a mobile culture or particular mobile subcultures.

This generation of new knowledge intrudes a new protagonist into the debate and dichotomy between utilitarian and liberal views of education, and challenges the idea of a common curriculum or universal canon of accepted and useful knowledge that an education system must deliver. It challenges too formal learning, its institutions and its professionals, in their roles as society’s gate-keepers to learning and technology for disadvantaged individuals and communities.

CONCLUSION

This chapter puts the work and evolution of mobile learning into the broadest possible context and explores the significance of ideas about ‘learning in a mobile age’ in the context of the current development of mobile learning. The mobile learning community has an increasingly clear sense of its achievements and its direction but looking beyond the immediate community reveals a far more complex and changing situation. At this point, we can only sketch parts of the evolving picture, guess how society, its conception of learning and the role of mobile technologies in supporting that conception will fit together and wonder at the place of our current work. The challenge for the mobile learning community is the balance between facing inwards, to develop its work, and facing outwards, to understand the context and importance of that work.

FUTURE RESEARCH DIRECTIONS

In looking back at this chapter and attempting to revise it, the main sense and direction of it remains valid and true but a complementary and more critical account (Traxler 2010) sits alongside it. Taken together, these neatly encapsulate mobile learning research and its wider significance.

REFERENCES

- Ally, M. (2006). *mLearn2006 Book of Abstracts*. Available online at www.mlearn.org.za/CD/BOA_p.07.pdf
- Arnedillo-Sánchez, I., Sharples, M., & Vavoula, G. (2007). *Beyond Mobile Learning Workshop*. Dublin, Ireland: Trinity College Dublin Press.
- Beale, R. (2007). How to enhance the experience without interfering with it? In M. Sharples (Ed.), *Big Issues in Mobile Learning: Report of a Workshop by the Kaleidoscope Network of Excellence Mobile Learning Initiative*. Nottingham, UK: Learning Sciences Research Institute.
- Brown, B., & Green, N. (2002). *Wireless World—social and interactional aspects of the mobile age*. London: Springer.
- Brown, T. H. (2005). *Beyond constructivism: Exploring future learning paradigms*. Thames, New Zealand: Aries Publishing Company.
- Bull, M. (2005). No Dead Air! The iPod and the Culture of Mobile Listening. *Leisure Studies*, 24(4), 343–356. doi:10.1080/0261436052000330447
- Cobcroft, R. (2006). *Literature Review into Mobile Learning in a University Context*. Queensland University of Technology. Available: <http://eprints.qut.edu.au/archive/00004805>
- Cooper, G. (2002). The Mutable World: Social Theory in The Wireless World. In B. Brown, N. Green, & R. Harper (Eds.), *Wireless World: Social and Interactional Aspects of the Mobile World*. London: Springer. doi:10.1007/978-1-4471-0665-4_2
- Crompton, H. (2014). A diachronic overview of mobile learning: A shift toward student-centered pedagogies. In M. Ali & A. Tsinakos (Eds.), *Increasing access through mobile learning*. British Columbia, Canada: Commonwealth of Learning Press.
- Crompton, H. (2013a). Mobile learning: New approach, new theory. In Z. L. Berge & L. Y. Muilenburg (Eds.), *Handbook of mobile learning* (pp. 47–57). Florence, KY: Routledge.
- Crompton, H. (2013b). A historical overview of mobile learning: Toward learner-centered education. In Z. L. Berge & L. Y. Muilenburg (Eds.), *Handbook of mobile learning* (pp. 3–14). Florence, KY: Routledge.
- Donner, J. (2008). The Rules of Beeping: Exchanging Messages Via Intentional Missed Calls on Mobile Phones. *Journal of Computer-Mediated Communication*, 13(1), 1–22. doi:10.1111/j.1083-6101.2007.00383.x
- Du Gay, P., Hall, S., Janes, L., & Nequs, K. (1997). *Doing cultural studies: the story of the Sony Walkman*. London: Sage.

- Elliott, T. S. (1934). *The Rock*. Academic Press.
- Engeström, Y. (2003). Retrieved from <http://www.edu.helsinki.fi/activity/pages/chatanddwr/activitysystem/>
- Engeström, Y. (1987). *Learning by expanding: An activity-theoretical approach to developmental research*. Helsinki: Orienta-Konsultit.
- Engeström, Y., Miettinen, R., & Punamäki-Gitai, R. (1999) Perspectives on activity theory. In *International Congress for Research on Activity Theory*. New York: Cambridge University Press.
- Er, M., & Kay, R. (2005) Mobile technology adoption for mobile information systems: an activity theory perspective. In *International Conference on Mobile Business (ICMB'05)*. Sydney, Australia: IEEE. doi:10.1109/ICMB.2005.70
- Gergen, K. J. (2002). The challenge of absent presence. In J. E. Katz & M. A. Aakhus (Eds.), *Perpetual Contact. Mobile Communication, Private Talk, Public Performance* (pp. 227–241). Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511489471.018
- Geser, H. (2004). *Towards a Sociological Theory of the Mobile Phone*. University of Zurich.
- Goffman, E. (1971). *Relations in Public*. Harmondsworth, UK: Allen Lane.
- JISC. (2005). *Innovative Practice with e-Learning: A good practice guide to embedding mobile and wireless technologies into everyday practice*. Bristol: Joint Information Services Committee.
- Jones, A., Issroff, K., Scanlon, E., Clough, G., & McAndrew, P. (2006) Using mobile devices for learning in Informal Settings: Is it Motivating? *Proceedings of IADIS International conference Mobile Learning*.
- Katz, J. E., & Aakhus, M. (Eds.). (2002). *Perpetual Contact—Mobile Communications, Private Talk, Public Performance*. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511489471
- Kay, A. C. (1972, Aug). A personal computer for children of all ages. *Proceedings of the ACM National Conference*. Doi:10.1145/800193.1971922
- Keegan, D. (2005). The Incorporation of Mobile Learning into Mainstream Education and Training. *Proceedings of mLearn2005- 4th World Conference on mLearning*. <http://www.mlearn.org.za/CD/papers/keegan1.pdf>
- Kuhn, T. S. (1962). *The Structure of Scientific Revolutions*. Chicago: Univ. of Chicago Press.
- Kukulska-Hulme, A., Evans, D., & Traxler, J. (2005). *Landscape Study in Wireless and Mobile Learning in the Post-16 Sector, Technical Report*. Bristol: Joint Information Services Committee.
- Kukulska-Hulme, A., & Traxler, J. (2005). *Mobile Learning: a Handbook for Educators and Trainers*. London: Routledge.
- Kukulska-Hulme, A., & Traxler, J. (2007). Design for Mobile and Wireless Technologies. In *Rethinking Pedagogy for the Digital Age*. London: Routledge.
- Laurillard, D. (2002). *Rethinking University Teaching- A Conversational Framework for the Effective Use of learning Technology* (2nd ed.). London: Routledge; doi:10.4324/9780203304846
- Laurillard, D. (2007). Pedagogic forms of mobile learning: framing research questions. In N. Pachler (Ed.), *Mobile Learning - towards a research agenda* (pp. 153–177). London: Institute of Education, University of London.
- Ling, R. (1997). One can talk about mobile manners! The use of mobile telephones in appropriate situations. In L. Haddon (Ed.), *Communications on the Move: the experience of mobile telephony in the 1990s, COST 248 Report*. Farsta: Telia.
- Ling, R. (2004). *The Mobile Connection—the cell phone's impact on society*. San Francisco: Morgan Kaufmann Publishers.

- Metcalfe, D. S. (2006). *mLearning: Mobile Learning and Performance in the Palm of your Hand*. Amherst, MA: HRD Press.
- Murtagh, G. (2002). Seeing the “Rules”: Preliminary Observations of Action, Interaction and Mobile Phone Use. In B. Brown, N. Green, & R. Harper (Eds.), *Wireless World: Social and Interactional Aspects of the Mobile World*. London: Springer. doi:10.1007/978-1-4471-0665-4_6
- Naismith, L., Lonsdale, P., Vavoula, G., & Sharples, M. (2004). *Literature Review in Mobile Technologies and Learning*. Bristol: NESTA FutureLab.
- Nyiri, K. (2006). Time and Communication. In *Time and History: Proceeding of the 28 International Ludwig Wittgenstein Symposium*. Kirchberg am Wechsel.
- Norris, C., & Soloway, E. (2011). BYOD as the catalyst to transform Classrooms. *District Administration Magazine*. Retrieved from: <http://www.districtadministration.com/viewdepartment.aspx?deptid=66>
- Norris, C., & Soloway, E. (2004). Envisioning the handheld-centric classroom. *Journal of Educational Computing Research*, 30(4), 14–19. doi:10.2190/MBPJ-L35D-C4K6-AQB8
- Norris, C., & Soloway, E. (2003a). Handheld computers can solve the computer access problem and impact learning. *MACUL Newsletter*. Retrieved from <http://www.macul.org/newsletter/2003/summer2003/norrissoloway.html>
- Norris, C., & Soloway, E. (2003b). *The viable alternative: Handhelds. Why the personal computer of choice in K-12 ought to fit in a student's palm*. Arlington, VA: School Administrator, April, American Association of School Administrators.
- O'Malley, C., Vavoula, G., Glew, J., Taylor, J., Sharples, M., & Lefrere, P. (2003). *Guidelines for learning/teaching/tutoring in a mobile environment*. Mobilelearn project deliverable. Available from <http://www.mobilelearn.org/download/results/guidelines.pdf>
- Oliver, C. (2007) *mLearn2007 Conference Proceedings*. Available online at http://www.mlearn2007.org/files/mLearn_2007_Conference_Proceedings.pdf
- Plant, S. (2000): *On the Mobile. The Effects of Mobile Telephones on Social and Individual Life*. Retrieved from <http://www.motorola.com/mot/documents/0,1028,333,00.pdf>
- Quinn, C. (2000). mLearning: Mobile, Wireless, in your Pocket Learning. *LineZine*. Retrieved from <http://www.linezine.com/2.1/features/cqmmwiyp.htm>
- Rheingold, H. (2003). *Smart Mobs: The Next Social Revolution*. New York: Perseus Books.
- Salmon, G. (2000). E-moderating - The key to teaching and learning online. Kogan Page.
- Sharpe, R., Benfield, G., Roberts, G., & Francis, R. (2006). *The undergraduate experience of blended e-learning: A review of UK literature and practice undertaken for the Higher Education Academy*. Retrieved from www.heacademy.ac.uk/4884.htm
- Sharples, M. (2005). Learning As Conversation: Transforming Education in the Mobile Age. In *Proceedings of Conference on Seeing, Understanding, Learning in the Mobile Age* (pp. 147- 152). Budapest, Hungary: Academic Press.
- Sharples, M. (Ed.). (2006). *Big Issues in Mobile Learning*. Nottingham, UK: Kaleidoscope Network of Excellence, Mobile Learning Initiative.
- Sharples, M., Sánchez, I. A., Milrad, M., & Vavoula, G. (2009). Mobile learning: Small devices, big issues. In S. N. Blacheff (Ed.), *Technology-enhanced learning: Principles and products* (pp. 223–251). Berlin, Germany: Springer-Verlag. doi:10.1007/978-1-4020-9827-7_14
- Sharples, M., Taylor, J., & Vavoula, G. (2005). Towards a theory of mobile learning. *Proceedings of mLearn2005- 4th World Conference on mLearning*. Retrieved from <http://www.mlearn.org.za/CD/papers/Sharples%20Theory%20of%20Mobile.pdf>

Sheller, M. (2004). Mobile publics: Beyond the network perspective. *Environment and Planning. Society & Space*, 22, 39–52. doi:10.1068/d324t

Sheller, M., & Urry, J. (2003). Mobile Transformations of Public and Private Life. *Theory, Culture & Society*, 20(3), 107–125. doi:10.1177/02632764030203007

Sheller, M., & Urry, J. (2006). The new mobilities paradigm. *Environment & Planning A*, 38(2), 207–226. doi:10.1068/a37268

Siemens, G. (2004). Connectivism: A Learning Theory for the Digital Age. *elearnspace*. Retrieved from <http://www.elearnspace.org/Articles/connectivism.htm>

Siemens, G. (2005). Connectivism: A Learning Theory for the Digital Age. *International Journal of Instructional Technology & Distance Learning*. Retrieved from http://www.itdl.org/Journal/Jan_05/article01.htm

Sørensen, C., Mathiassen, L., & Kakihara, M. (2002). *Mobile Services: Functional Diversity and Overload*. Presented at New Perspectives On 21st-Century Communications, Budapest, Hungary.

Souza e Silva, A. (2006). From cyber to hybrid: Mobile technologies as interfaces of hybrid spaces. *Space and Culture*, 9(3), 261–278. doi:10.1177/1206331206289022

Tamminien, S., Oulasvirta, A., Toiskallio, K., & Kankainen, A. (2003). *Understanding mobile contexts*. Udine, Italy: Springer. doi:10.1007/978-3-540-45233-1_3

Taylor, J. (2006). What are appropriate methods for evaluating learning in mobile environments? Evaluating Mobile Learning. In M. Sharples (Ed.), *Big Issues in Mobile Learning*. Nottingham, UK: Kaleidoscope Network of Excellence, Mobile Learning Initiative.

Traxler, J. (2005). Mobile learning- it's here but what is it? *Interaction*, 9, 1.

Traxler, J. (2007). Defining, Discussing and Evaluating Mobile Education. *International Review of Research in Open and Distance Learning*. doi:10.19173/irrodl.v8i2.346

Traxler, J. (2008). Mobile Learning: Moving Through Philosophical Space. *Proceedings of IADIS Mobile Learning*.

Traxler, J. (2011). Mobile Learning - Starting in the Right Place, Going in the Right Direction? *International Journal of Mobile and Blended Learning*, 3(2).

Traxler, J., & Kukulska-Hulme, A. (2005). Evaluating Mobile Learning: Reflections on Current Practice. *Proceedings of mLearn2005: 4th World Conference on mLearning*. Available: http://www.mlearn.org.za/CD/BOA_p.65.pdf

Urry, J. (2000). *Sociology beyond societies: Mobilities for the 21st Century*. London: Routledge.

Winters, N. (2006). What is Mobile Learning? In M. Sharples (Ed.), *Big Issues in Mobile Learning*. Nottingham, UK: Kaleidoscope Network of Excellence, Mobile Learning Initiative.

ADDITIONAL READING

Berge, Z. L., & Muilenburg, L. Y. (Eds.). (2013). *Handbook of mobile learning*. Florence, KY: Routledge.

Churchill, D. & Notari M. P. (Eds.), *Proceedings of the International Mobile Learning Festival 2016: Mobile Learning, Emerging Learning Design & Learning 2.0*. Bangkok, Thailand. Retrieve from <http://imlf.mobi/publications/proceedings2016.pdf>

Crompton, H., & Burke, D. (2015). School culture for the mobile digital age [Special Issue]. *Medical Education*, 6(2), 208–223.

Ali, M., & Tsinakos, A. (Eds.). (2014). *Increasing access through mobile learning*. British Columbia, Canada: Commonwealth of Learning Press.

Traxler, J. (2011). Mobile Learning - Starting in the Right Place, Going in the Right Direction? *International Journal of Mobile and Blended Learning*, 3(2).

Traxler, J., & Wishart, J. (2011) Making Mobile Learning Work: Case Studies of Practice, Bristol: ESCAlate (HEA Education Subject Centre)

KEY TERMS AND DEFINITIONS

Mobile Learning: Mobile learning is defined as “learning across multiple contexts, through social and content interactions, using personal electronic devices” (Crompton, 2013, p. 4).

Pedagogies: Methods of teaching.

Personal Computers: Another term typically used to describe mobile devices.

Tethered Technologies: Digital Technologies that typically require a constant electrical source, such as desktop computers.

User Generated Content/Information: This term is used to describe content/information created by consumers or end –users, such as video, digital images, audio files, and blogs.

Mobile Game-Based Learning



Boaventura DaCosta

Solers Research Group, USA

Soonhwa Seok

Korea University, South Korea

Carolyn Kinsell

Solers Research Group, USA

INTRODUCTION

The use of video games to enhance learning has been the subject of fierce debate (Guillén-Nieto & Aleson-Carbonell, 2012). Nonetheless, experts have long strived to comprehend the draw of these games and their potential role in education. Thus, there is increasing interest in the application of video games in primary, secondary, and higher education; government; financial services; healthcare; hospitality; science and technology; telecommunications; and corporate and military training (Garris, Ahlers, & Driskell, 2002), to achieve a variety of learning outcomes (Kebritchi & Hirumi, 2008). To date, findings point to the instructional benefits of video games in the areas of computer science, geography, language, mathematics, photography, and science (DaCosta, Seok, & Kinsell, 2015).

Advancements in mobile computing have helped pave the way for new video game research (DaCosta & Seok, 2017a, 2017b) in different areas of interest, to include social development, intellectual activities (Spikol & Milrad, 2008), and general learning (Facer et al., 2004; Rogers & Price, 2006). The anytime-, anywhere-, and on-any-device characteristics of mobile technology are of particular importance because they offer new opportunities to research game-based learning (GBL) that is free from space and time restrictions (DaCosta et al., 2015). That is, mobile games have the potential to bridge the gap between the

classroom and the real world by placing students in authentic places and learning situations (Costabile et al., 2008). This has, in part, contributed to what is mounting research on what could be called *mobile game-based learning* (mGBL).

This chapter offers a review of the literature on mGBL. Extending the work of DaCosta et al. (2015), the discussion includes (a) a definition of mGBL in the context of game-based learning; (b) a review of the educational benefits of mobile games, building on what is known about video games and learning; and (c) identification of mobile game examples (called out in the literature) that have been experimented with and/or used to explore learning.

To ensure scholarly rigor, the research offered here was compiled in a staged approach similar to that of primary research (Cooper, 1998). The majority of the content comes from books, academic journals, and databases. While peer-reviewed materials were foremost, other resources were also used, including online articles, as supplemental information befitting their academic stature and to provide context regarding “practicing real world examples or an in-the-trenches view” (Kinsell, DaCosta, & Nasah, 2014, p. 161). Thus, this work presents the perspectives and findings of numerous scholars, practitioners, researchers, and experts in an effort to provide a well-rounded view of the educational possibilities of mobile games.

Throughout the chapter the term *video game* refers to a game played on a personal computer

(PC) or dedicated game device, such as a game console (e.g., Xbox®, PlayStation®, Wii®) or handheld game device (e.g., 3DS®, Vita®). The term *mobile game*, on the other hand, is used to refer to a video game played on a mobile device (e.g., cell phone, personal digital assistant [PDA], smartphone, tablet). Finally, this work is not intended to debate the practice of video games in education or the use of these games in classrooms, but rather to drive forward the study of mobile learning.

BACKGROUND

What Is Mobile Game-Based Learning?

While game-based learning is a fairly recognized notion, attempts at offering a thorough description have proven difficult (Perrotta, Featherstone, Aston, & Houghton, 2013). Many explanations often consist of definitions attempting to identify the key principles and mechanics involved (e.g., Perrotta et al., 2013). At its core, GBL is rooted in the belief that games can be used in the learning process. This means that GBL is less about the games themselves and more about the educational contribution they can make, with focus on the social dynamics involved and how to best use these games to enhance learning (Perrotta et al., 2013). Building upon this idea, mGBL may be understood as an extension of GBL, fixed in the belief that games played on a mobile device (e.g., cell phone, PDA, smartphone, tablet) can be used to enhance the learning experience.

What Do We Know About Mobile Game-Based Learning?

As mentioned, the educational value of video games has long been a topic of debate (Guillén-Nieto & Aleson-Carbonell, 2012). It should come as no surprise, therefore, that studies have reported mixed findings regarding the academic

impact of these games, with contradictory findings commonplace (Egenfeldt-Nielsen, 2006). One of the major factors contributing to the difficulty in offering detailed, consistent, and concrete evidence in support of the academic benefits of video games has been the nature of the studies themselves (Perrotta et al., 2013). For example, studies have varied in their aims, ranging from the impact of video games on learning outcomes measured in the contexts of academic achievement, cognitive performance, gender, attitudes toward learning using the games themselves, acquisition of knowledge and skills, motivation and classroom engagement, to types of learning and kinds of games. Studies have also examined the educational value of video games across different domains, such as civics and society, computer science, language, and mathematics (Perrotta et al., 2013). Further, diverse conceptual frameworks and learning theories have been adopted, encompassing behaviorism, cognitivism, and constructivism approaches (Egenfeldt-Nielsen, 2006). Finally, it has been noted that many of these studies are subject to methodological flaws and limitations (Perrotta et al., 2013), which, among other challenges, have included research bias, weak assessments, short exposure times, and lack of control groups (Egenfeldt-Nielsen, 2006).

Even with these mixed findings, wide range of study topics, and methodological flaws, the positive relationships between video games and learning (Egenfeldt-Nielsen, 2006; Perrotta et al., 2013) have been no less argued. It has been offered, for example, that video games are more flexible than other media, almost naturally supporting adaptive learning (del Blanco, Marchiori, Torrente, Martínez-Ortiz, & Fernández-Manjón, 2013). This has helped promote the idea that video games can play an important role in active learning, especially in terms of critical thinking skills, knowledge construction, collaboration, and effective use of and access to information and communication technology (Ellis, Heppell, Kirriemuir, Krotoski, & McFarlane, 2006). Moreover, because video games can provide immediate

feedback pertaining to actions and decisions (at times, within a complete and interactive virtual playing space), it has been proposed that ambient information can foster an immersive experience, stimulating interest in the game (Mitchell & Savill-Smith, 2004), and encouraging exploration and experimentation (Kirriemuir, 2002). This, in turn, can help promote authentic learning, allowing students to practice within a highly realistic but safe and risk-free environment (del Blanco et al., 2013). It has also been proposed that video games put the learner in a decision-making role by creating harder and harder challenges, allowing learning to occur through trial and error (Gee, 2007) or through the presentation of increasingly difficult challenges that are adapted to the learner's mastery of previously acquired knowledge and skills (Gentile & Gentile, 2008). Given this benefit, video games can be used to offer activities that are paced to the student's aptitude, thereby providing differentiated instruction (Paraskeva, Mysirlaki, & Papagianni, 2010). This allows knowledge and skills to be practiced to a mastery level, becoming automatized, and enables students to focus consciously on comprehending or applying new information (Gentile & Gentile, 2008).

These are not new ideas; acceptance that video games can be successfully used in learning dates back several decades. For instance, studies have suggested that video games can be used to (a) increase the retention of embedded instructional subject matter (Dempsey, Lucassen, Gilley, & Rasmussen, 1993; Jacobs & Dempsey, 1993; Pierfy, 1977; Ricci, Salas, & Cannon-Bowers, 1996) when teaching specific skills (Griffiths, 2002), to include algebra (Corbett, Koedinger, & Hadley, 2001), biology (Ybarrondo, 1984), photography (Abrams, 1986), and computer programming (Kahn, 1999); (b) develop higher-order cognitive skills (Wood & Stewart, 1987); and (c) work through personal problems (Bowman, 1982; Chaffin, Maxwell, & Thompson, 1982; Trinkaus, 1983). In addition, findings have shown improved performance on visual attention tasks (e.g., Feng, Spence, & Pratt, 2007; Green

& Bavelier, 2003, 2006), spatial abilities (e.g., Greenfield, Brannon, & Lohr, 1994), hand-eye coordination (e.g., Griffith, Voloschin, Gibb, & Bailey, 1983), reaction times (e.g., Goldstein et al., 1997; Orosy-Fildes & Allen, 1989; Yuji, 1996), and development of social skills among students with disabilities (e.g., Gaylord-Ross, Haring, Breen, & Pitts-Conway, 1984; Sedlak, Doyle, & Schloss, 1982).

Against this background, the body of knowledge on the educational value of video games continues to grow (Cogoi, Sangiorgi, & Shahin, 2006). This includes research focused on mobile computing. However, to date, research focused explicitly on GBL in the context of mobile games has predominantly centered on the technical capabilities of emerging mobile devices (e.g., Bell et al., 2006; Cheok, Sreekumar, Lei, & Thang, 2006; Göth, Häss, & Schwabe, 2004; Grant et al., 2007; Licoppe & Inada, 2006; Matyas et al., 2008; Naismith, Lonsdale, Vavoula, & Sharples, 2004; Schmitz, 2014; Sedano, Laine, Vinni, & Sutinen, 2007; Sotirou et al., 2008), with considerable attention placed on location-aware (i.e., Global Positioning System [GPS]) capabilities, in turn, offering new ways to embed learning in authentic settings. That is, digital content can be infused with the real world in an augmented reality (AR), blurring the boundaries between a virtual world and real life, resulting in a blended experience (De Freitas & Griffiths, 2008; Huizenga, Admiraal, Akkerman, & ten Dam, 2009; Montola, 2011). This mixed reality is incredibly important because it allows students to experience the real world, while at the same time, interact with virtual objects in a real space, in effect, taking students and learning out of the traditional classroom setting (Grant et al., 2007; Huizenga et al., 2009; Roschelle & Pea, 2002). In short, this offers educators new ways to help their students grasp relationships and connections (Yuen, Yaoyuneyong, & Jonhson, 2011). This mixed reality is also important because it is believed to provide students with a truly reflective experience (Price, Rogers, Scaife, Stanton, & Neale, 2003; Rogers, Scaife, Gabrielli, Smith,

& Harris, 2002; Sharples, Corlett, & Westmancott, 2002; Stanton, O'Malley, Ng, Fraser, & Benford, 2003). Altogether, these factors make mobile technology a well-matched platform for supporting learning activities (Klopfer, Squire, & Jenkins, 2002).

Examples of Mobile Game-Based Learning

DaCosta et al. (2015) identified mobile games found in the literature that have been used to integrate learning with aspects of the physical world. An expansion of their list is shown in Table 1. While not exhaustive, this compilation can be viewed as a starting point for those interested in examining mobile games that have been used to explore GBL in the context of mobile computing. Providing different experiences and learning contexts, these games demonstrate the educational potential of mobile computing. For example, some of these (a) demonstrate opportunities to align educational institutions more closely with local communities; (b) recreate historical events by placing learners in real-world situations and places as a way to immerse and motivate students, helping them learn about past people, locations, and events; or (c) show how place and time can be used to teach important concepts. The following are some specific examples.

In the augmented reality mobile game project *Mentira*, Holden and Sykes (2011) described how, using clues provided on an iPod Touch®, students practice their Spanish while solving the prohibition-era murder of Dionisio Silva by exploring various locations in the neighborhood of Los Griegos in Albuquerque, New Mexico. Along with moving learning outside of the classroom, the project sparked interest among residents, who were excited that students were going to be learning about their neighborhood. In another mobile game project, *Frequency 1550* (Huizenga et al., 2009), using GPS-capable smart video phones, students acquired historical knowledge of medieval Amsterdam by being placed in the city and

completing tasks at certain locations. Also moving learning outside of the classroom, this project demonstrated the use of authentic settings in the learning process. Colella (2000, 2002), in turn, showed how students can learn about the spread of infection using a classroom-based participatory simulation, in which they interacted with one another in a physical location using custom-built devices. Although the location was not necessarily of importance in this project (as it was in others), it demonstrated how concepts can be simulated with the aid of mobile devices in a physical space.

Other mobile games, found in the literature, are not necessarily educational or use location-based technology, but may be no less important to the discussion on mGBL. For example, the following games would not be considered pedagogical, but are unique in that they demonstrate the potential of mobile computing, with some demonstrating the use of time and place mechanics: *AlienRevolt* (de Souza e Silva, 2008), *BotFighters* (Sotamaa, 2002), *Can You See Me Now?* (Benford et al., 2006), *CitiTag* (Vogiazou & Raijmakers, 2004), *Epidemic Menace* (Lindt, Ohlenburg, Pankoke-Babatz, & Ghellal, 2007), *GeoCaching* (Groundspeak, 2016), *Human Pacman* (Cheok et al., 2004), *I like Frank* (Flintham, Benford, & Humble, n.d.), and *Uncle Roy All Around You* (Benford et al., 2004).

Whereas, the following games could be viewed from the perspective of mGBL, offering some level of instruction, but they do not necessarily take advantage of place-based technology or take advantage of this technology to the extent of some of the other games discussed thus far: *AKAMIA* (Ahmad & Rahman, 2014), *Eduventure* (Ferdinand, Müller, Ritschel, & Wechselberger, 2005), *GEIST* (Kretschmer, 2002), *iPlayCode* (Zhang & Lu, 2014), *MaCMoG* (Shiratuddin & Zaibon, 2001), *Mad City Mystery* (Squire & Jan, 2007), *Mobile Rhythm Learning System* (Wang & Lai, 2011), *The MOBO City* (Fotouhi-Ghazvini, Earnshaw, Robison, & Excell, 2009), *Weatherlings* (Sheldon et al., 2010), and *ZooQuest* (Veenhof, Sandberg, & Maris, 2012).

Table 1. Mobile games presented in the literature that have been used to integrate learning with aspects of the physical world

Title	Description	Purpose
<i>Blåtannkoden</i>	Players visit the Norwegian Oslo Norsk Telemuseum. Using their mobile devices, players photograph QR (quick response) codes, identifying solutions to riddles related to museum topics. Additional information about the topic is offered with correct answers, to include a letter. At the end of the game, letters form a secret word that can be used to get a free gift (Ceipidor, Medaglia, Perrone, De Marsico, & Di Romano, 2009).	Explore learning about history
<i>BuinZoo</i>	Players visit a zoo in central Chile and participate in a guided tour presented as a quiz game. In groups of four, players use a virtual map on a PDA to locate animals and digital resources to answer questions. Later in the classroom, players use a simulation to learn about evolution by preserving and developing three species of four classes (fish, amphibians, reptiles, and birds) (Sánchez, Salinas, & Sáenz, 2006).	Explore learning about species and classes
<i>Capture the Flag</i>	Players track flags in an AR environment that combines play with PC-based visualizations and mobile digital flags (Cheok, Sreekumar, Lei, & Thang, 2006).	Explore active game play
<i>CitiExporer</i>	Players take geo-referenced photos, localize geographic points of interest in the city of Bamberg, Germany, and Fujisawa, Japan, and categorize them semantically. A community of reviewers verifies that the information uploaded by the players (photo, text, GPS coordinates, and location category) is valid (Matyas et al., 2008).	Explore ability to interpret urban environments
<i>COLLAGE</i> (Collaborative Learning Platform Using Game-Like Enhancements)	Players engage in a board-like game on a site of educational interest using mobile phones, PDAs, and GPS capabilities to communicate with one another on site and in the classroom. Educators create mobile learning experiences to support their curriculum with the assistance of an implementation guide and web-based interface for creating and designing the game (Sotirou et al., 2008).	Explore context-dependent and location-based learning
<i>EcoMOBILE</i>	Visiting a pond, players work in pairs to make observations about organisms and classify them (producer, consumer, and decomposer). Players then collect water measurements using the TI NSpire® and environmental probes. They also participate in other activities, to include comparing observations and water measurements with other players. An AR program offers players instructions while at the pond. Later in the classroom, players compile all data and discuss results (Kamarainen et al., 2013).	Explore learning about physical ecosystems and water quality measurements
<i>Environmental Detectives</i>	Players take the role of environmental scientists tasked with investigating health concerns at a site linked to the release of toxins in the local water supply. Players experience a “virtual practicum” similar to working on an environmental research team (Klopper et al., 2002).	Explore effective investigation using quantitative and qualitative data
<i>Explore!</i> (Excursion game)	Players visit the archaeological site of Egnathia, an ancient city in the Apulia region of Italy, where they are given a case to solve. Role-playing as a Roman family, players work as a group using a mobile phone and a map of the site to find their way around, placing markers to solve the case (Ardito, Buono, Costabile, Lanzilotti, & Pederson, 2007; Costabile et al., 2008).	Explore learning about history
<i>Feeding Yoshi</i>	Players explore a city looking for virtual fruit in which to feed their character. Played in teams, players use HP iPAQ 2750s® and 4150s® to collect points by feeding the creatures. The game uses existing public and private wireless hotspots (Bell et al., 2006).	Explore ability to interpret urban environments
<i>Frequency 1550</i>	Players visit the historical city of Amsterdam. Using smart and GPS-enabled video phones and the Internet, players gain citizenship to the city by conducting exercises tied to the themes of labor, trade, religion, rules and government, local knowledge, and defense (Huizenga et al., 2009).	Explore learning about history; player engagement and motivation

continued on following page

Table 1. Continued

Title	Description	Purpose
<i>Massey Mobile Helper</i>	Players are given applicable contextual information (e.g., location of lecture halls and laboratory rooms, and lecture cancellations) in support of their daily school life as they traverse the Massey University campus. Using a PDA phone (HP-IPAQ 6700®) with a Bluetooth GPS device (GlobalSat BT-338®), the game locates players' positions on the university map to provide relevant content (Brown, Ryu, & Parsons, 2006).	Explore acclimating to university life
<i>Mentira</i>	Players learn Spanish in the Albuquerque neighborhood of Los Griegos. Using iPod Touches®, players must solve the prohibition-era murder of Dionisio Silva. Players collect clues and try to solve the murder by exploring various neighborhood locations and examine information provided on their mobile devices (Holden & Sykes, 2011).	Explore learning the Spanish language
<i>MobileGame</i>	Players are given tasks that must be completed at certain spots as they are led through various university locations. Tasks are comprised of one or two simple questions displayed on a PDA (Göth et al., 2004; Schwabe & Göth, 2005).	Explore learning through an orientation-based game
<i>MobileMath</i>	Players working in teams of two are sent to a physical location. Using GPS-enabled mobile devices and instructions, players must draw four vertex shapes (squares, rectangles, or parallelograms) to score points. Correct shapes are added to a virtual map coded by team color for all players to see (Wijers, Jonker, & Kerstens, 2008).	Explore learning mathematics
<i>MobiMissions</i>	Players create missions for other players. Players take a series of photographs and enter text to create missions, which are "dropped" from their mobile phones. Missions remain dropped until found and "picked up" by other players (Grant et al., 2007).	Explore collaborative experiences in a physical space
<i>Mogi</i>	Players collect virtual objects throughout the city of Tokyo. Using a GPS-enabled mobile phone, players catch objects, such as fruits or animals. Virtual objects are depicted on the mobile device when players are within 400 meters of an object (Licoppe & Inada, 2006).	Explore ability to rediscover a city
<i>REXplorer</i>	Players visit Regensburg, Germany. Players use a rented GPS-enabled "game device" with camera-based motion estimation to create encounters with historical figures. Players wave the device to "cast a spell" to awaken the spirit of figures associated with existing buildings and settings throughout the city (Ballagas et al., 2007).	Explore tourists' acclimation to surroundings; learning about history and culture
<i>Savannah</i>	Players explore the challenges experienced by lions on the Savannah. Using GPS tethered to iPAQ 5450® PDAs with headphones, players must balance costs and benefits as well as collaborate and negotiate with one another to meet game objectives (Facer et al., 2004).	Explore conceptual understanding of animal behavior
<i>SciMyst</i>	Players solve different "enigmas" or questions based on real-world information. Using Nokia N80® mobile phones, players either collaborate or play individually to answer questions (Sedano et al., 2007).	Explore collaboration
<i>Sick at South Shore Beach</i>	Players are presented with an environmental scenario and visit South Shore Beach, a mixed-recreational park in Milwaukee, to investigate a collection of illnesses that are linked to the beach. Players are presented with information about the scenario using AR. Using a GPS-enabled PDA, players use the information gathered (along with classroom research) to formulate, present, and defend their final hypothesis about the illnesses (Mathews, Holden, Jan, & Martin, 2008).	Exploring learning about disease related to bacteria found on sandy beaches
<i>Skattjakt</i>	Players collaborate to solve a mystery about a castle located on the Vaxjo University campus. The playing field is spread over six locations throughout the campus. Players find markers using text and audio clues provided on Nokia N70® mobile phones (Spikol & Milrad, 2009).	Explore learning about history; informal skills in map reading

continued on following page

Table 1. Continued

Title	Description	Purpose
<i>SupaFly</i>	Players create and interact through characters in a community-based soap opera. Interaction and character development take place using mobile phones through short message service (SMS) and a website (Jegers & Wiberg, 2006).	Explore social aspects of mobile game play
<i>TimeWarp</i>	Players visit Cologne, Germany. Based on the legend of the Heinzelmännchen, players explore the city to bring back the fabled elves. Equipped with a mobile AR system and a GPS-enabled handheld information device with Bluetooth, players search for relevant locations to interact with various elements to locate the elves (Herbst, Braun, McCall, & Broll, 2008).	Explore learning about history; ability to interpret urban environments
<i>Virus Game</i>	Players learn about the spread of infection using a classroom-based participatory simulation. Moving around a predetermined physical location, players meet one another face-to-face. A custom-built device worn by each player tracks interactions and determines if the infection spread (Colella, 2000, 2002).	Explore learning about the spread of disease

FUTURE RESEARCH DIRECTIONS

While literature on mGBL is mounting, as with video game research in general, findings to date are far from conclusive (Huizenga et al., 2009). DaCosta and colleagues (2015) concluded that much of the research thus far has focused on the technology. To fully appreciate the educational possibilities mobile computing has to offer, experts must look beyond individual devices and functionality (Naismith et al., 2004). That is, the need exists to more thoroughly examine the cognitive aspects of mGBL (Huizenga et al., 2009) if we are to understand how to best leverage mobile computing in educational contexts. As a result, future research should focus not only on the anytime-, anywhere-, and on-any-device characteristics of mobile technology, but pedagogy as well, to include the proper implementation of learning theories and strategies (DaCosta et al., 2015).

CONCLUSION

The huge spread of mobile computing is undeniable, with the draw of mobile gaming reaching

epic proportions. This popularity, along with the unique characteristics of mobile computing, has ignited mounting interest in the use of these games in educational contexts, to the extent that numerous examples can be found in the literature. Although much of the research to date has focused on the technology itself, recent studies of mGBL are promising. Future research should look beyond individual devices and functionality, and place greater importance on pedagogy, emphasizing the learning that can take place through the use of these games.

REFERENCES

- Abrams, A. (1986). *Effectiveness of interactive video in teaching basic photography skills*. Paper presented at the annual convention of the Association for Educational Communication and Technology, Las Vegas, NV.
- Ahmad, W. F. W., & Rahman, N. F. A. (2014). AKAMIA: Chemistry mobile game-based tutorial. *Proceedings from the 3rd International Conference on User Science and Engineering*, 221-226. doi:10.1109/IUSER.2014.7002706

- Ardito, C., Buono, P., Costabile, M. F., Lanzilotti, R., & Pederson, T. (2007). Mobile games to foster the learning of history at archaeological sites. *Proceedings from the IEEE Symposium on Visual Languages and Human-Centric Computing* (pp. 81-84). Los Alamitos, CA: IEEE Computer Society. doi:10.1109/VLHCC.2007.54
- Ballagas, R. A., Kratz, S. G., Borchers, J., Yu, E., Walz, S. P., & Fuhr, C., ... Tann, M. (2007). REXplorer: A mobile, pervasive spell-casting game for tourists. *Proceedings from the Conference on Human Factors in Computing Systems* (pp. 1929-1934). San Diego, CA: ACM Press. doi:10.1145/1240866.1240927
- Bell, M., Chalmers, M., Barkhuus, L., Hall, M., Sherwood, S., & Tennent, P. (2006). Interweaving mobile games with everyday life. *Proceedings of the SIGCHI Conference on Human Factors in Computing System* (pp. 417-426). Montreal, Canada: ACM Press. doi:10.1145/1124772.1124835
- Benford, S., Crabtree, A., Flintham, M., Drozd, A., Anastasi, R., Paxton, M., & Row-Farr, J. et al. (2006). Can You See Me Now?. *ACM Transactions on Computer-Human Interaction*, 13(1), 100–133. doi:10.1145/1143518.1143522
- Benford, S., Flintham, M., Drozd, A., Anastasi, R., Rowland, D., & Tandavanitj, N. ... Sutton, J. (2004). Uncle Roy All Around You: Implicating the city in a location-based performance. *Proceedings from Advanced Computer Entertainment 2004*. Ipswich, UK: ACM Press.
- Bowman, R. F. (1982). A Pac-Man theory of motivation. Tactical implications for classroom instruction. *Educational Technology*, 22(9), 14–17.
- Brown, R., Ryu, H., & Parsons, D. (2006). Mobile helper for university students: A design for a mobile learning environment. *Proceedings of the 18th Australia conference on Computer-Human Interaction: Design: Activities, Artefacts and Environments* (Vol. 206, pp. 297-300). Sydney, Australia: ACM Press. doi:10.1145/1228175.1228227
- Ceipidor, U. B., Medaglia, C. M., Perrone, A., De Marsico, M., & Di Romano, G. (2009). A museum mobile game for children using QR-codes. *Proceedings from the 8th International Conference on Interaction Design and Children* (pp. 282-283). Como, Italy: ACM Press. doi:10.1145/1551788.1551857
- Chaffin, J. D., Maxwell, B., & Thompson, B. (1982). ARC-ED curriculum: The application of video game formats to educational software. *Exceptional Children*, 49(2), 173–178.
- Cheok, A. D., Goh, K. H., Liu, W., Farbiz, F., Fong, S. W., Teo, S. L., & Yang, X. et al. (2004). Human Pacman: A mobile, wide-area entertainment system based on physical, social, and ubiquitous computing. *Personal and Ubiquitous Computing*, 8(1), 71–81. doi:10.1007/s00779-004-0267-x
- Cheok, A. D., Sreekumar, A., Lei, C., & Thang, L. (2006). Capture the flag: Mixed-reality social gaming with smart phones. *IEEE Pervasive Computing/IEEE Computer Society [and] IEEE Communications Society*, 5(2), 62–69. doi:10.1109/MPRV.2006.25
- Cogoi, C., Sangiorgi, D., & Shahin, K. (2006). mGBL – Mobile game-based learning: Perspectives and usage in learning and career guidance topics. *eLearning Papers*, 1(1), 1-6.
- Colella, V. (2000). Participatory simulations: Building collaborative understanding through immersive dynamic modeling. *Journal of the Learning Sciences*, 9(4), 471–500. doi:10.1207/S15327809JLS0904_4
- Colella, V. (2002). Participatory simulations: Building collaborative understanding through immersive dynamic modeling. In T. Koschmann, R. Hall, & N. Miyake (Eds.), *CSCL2: Carrying forward the conversation (computers, cognition, and work)* (pp. 357–391). Mahwah, NJ: Lawrence Erlbaum.
- Cooper, H. (1998). *Synthesizing research* (3rd ed.; Vol. 2). Thousand Oaks, CA: Sage Publications.

- Corbett, A. T., Koedinger, K. R., & Hadley, W. (2001). Cognitive tutors: From the research classroom to all classrooms. In P. S. Goodman (Ed.), *Technology-enhanced learning* (pp. 235–263). Mahwah, NJ: Lawrence Erlbaum.
- Costabile, M. F., De Angeli, A., Lanzilotti, R., Ardito, C., Buono, P., & Pederson, T. (2008). Explore! Possibilities and challenges of mobile learning. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 145–154). Florence, Italy: ACM Press. doi:10.1145/1357054.1357080
- DaCosta, B., & Seok, S. (2017a). Factors that explain adolescent and young adult mobile game play, part 1: A quantitative examination of the characteristics describing the casual player. In R. Zheng & M. Gardner (Eds.), *Handbook of research on serious games for educational applications* (pp. 320–339). Hershey, PA: IGI Global; doi:10.4018/978-1-5225-0513-6.ch015
- DaCosta, B., & Seok, S. (2017b). Factors that explain adolescent and young adult mobile game play, part 2: A quantitative examination of the casual player in the context of age and gender. In R. Zheng & M. Gardner (Eds.), *Handbook of research on serious games for educational applications* (pp. 340–365). Hershey, PA: IGI Global; doi:10.4018/978-1-5225-0513-6.ch016
- DaCosta, B., Seok, S., & Kinsell, C. (2015). Mobile games and learning. In Z. Yan (Ed.), *Encyclopedia of mobile phone behavior* (Vol. 1, pp. 46–60). Hershey, PA: IGI Global; doi:10.4018/978-1-4666-8239-9.ch004
- De Freitas, S., & Griffiths, M. (2008). The convergence of gaming practices with other media forms: What potential for learning? A review of the literature. *Learning, Media and Technology*, 33(1), 11–20. doi:10.1080/17439880701868796
- de Souza e Silva, A. (2008). Alien revolt (2005/2007): A case study of the first location-based mobile game in Brazil. *IEEE Technology and Society Magazine*, 27(1), 18–28. doi:10.1109/MTS.2008.918036
- del Blanco, A., Marchiori, E. J., Torrente, J., Martínez-Ortiz, I., & Fernández-Manjón, B. (2013). Using e-learning standards in educational video games. *Computer Standards & Interfaces*, 36(1), 178–187. doi:10.1016/j.csi.2013.06.002
- Dempsey, J., Lucassen, B., Gilley, W., & Rasmussen, K. (1993). Since Malones theory of intrinsically motivating instruction: Whats the score in the gaming literature? *Journal of Educational Technology Systems*, 22(2), 173–183. doi:10.2190/2TH7-5TXG-TAR7-T4V2
- Egenfeldt-Nielsen, S. (2006). Overview of research on the educational use of video games. *Digital Kompetanse*, 1(3), 184–213.
- Ellis, H., Heppell, S., Kirriemuir, J., Krotoski, A., & McFarlane, A. (2006). Unlimited learning. Computer and video games in the learning landscape. London, UK: ELSPA (Entertainment and Leisure Software Publishers Association).
- Facer, K., Joiner, R., Stanton, D., Reid, J., Hull, R., & Kirk, D. (2004). Savannah: Mobile gaming and learning? *Journal of Computer Assisted Learning*, 20(6), 399–409. doi:10.1111/j.1365-2729.2004.00105.x
- Feng, J., Spence, I., & Pratt, J. (2007). Playing an action video game reduces gender differences in spatial cognition. *Psychological Science*, 18(10), 850–855. doi:10.1111/j.1467-9280.2007.01990.x PMID:17894600

- Ferdinand, P., Müller, S., Ritschel, T., & Wechselberger, U. (2005). *The Eduventure – A new approach of digital game-based learning combining virtual and mobile augmented reality game episodes*. In H. Diener, S. Malo, & B. Urban (Eds.), *Game-based learning. Beiträge des Preconference Workshop der 3. Deutschen E-Learning Fachtagung Informatik (DeLFI 2005)* (pp. 79–90). Rostock: Fraunhofer IRB Verlag.
- Flintham, M., Benford, S., & Humble, J. (n.d.). *I like Frank: A mixed reality game for 3G phones*. Retrieved from https://www.blasttheory.co.uk/wp-content/uploads/2013/02/research_ilf_a_mixed_reality_game_for_3g_phones.pdf
- Fotouhi-Ghazvini, F., Earnshaw, R., Robison, D., & Excell, P. (2009). *The MOBO City: A mobile game package for Technical language learning*. *International Journal of Interactive Mobile Technologies*, 3(2), 19–24. doi:10.3991/ijim.v3i2.757
- Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, motivation, and learning: A research and practice model. *Simulation & Gaming*, 33(4), 441–467. doi:10.1177/1046878102238607
- Gaylord-Ross, R. J., Haring, T. G., Breen, C., & Pitts-Conway, V. (1984). The training and generalization of social interaction skills with autistic youth. *Journal of Applied Behavior Analysis*, 17(2), 229–247. doi:10.1901/jaba.1984.17-229 PMID:6735954
- Gee, J. P. (2007). *What video games have to teach us about learning and literacy* (2nd ed.). New York, NY: Palgrave.
- Gentile, D. A., & Gentile, J. R. (2008). Violent games as exemplary teachers. A conceptual analysis. *Journal of Youth and Adolescence*, 37(2), 127–141. doi:10.1007/s10964-007-9206-2
- Goldstein, J., Cajko, L., Oosterbroek, M., Michielsen, M., van Houten, O., & Salverda, F. (1997). Video games and elderly. *Social Behavior and Personality*, 25(4), 345–352. doi:10.2224/sbp.1997.25.4.345
- Göth, C., Häss, U.-P., & Schwabe, G. (2004). Requirements for mobile learning games shown on a mobile game prototype. In J. Attewell & C. Savill-Smith (Eds.), *Mobile learning anytime everywhere – A book of papers from MLEARN 2004* (pp. 95–100). Rome, Italy: Learning and skills development agency. doi:10.5167/uzh-61359
- Grant, L., Daanen, H., Benford, S., Hampshire, A., Drozd, A., & Greenhalgh, C. (2007). *MobiMissions: The game of missions for mobile phones. Proceedings from the ACM SIGGRAPH 2007 Educators Program*. New York, NY: ACM Press; doi:10.1145/1282040.1282053
- Green, C. S., & Bavelier, D. (2003). Action video game modified visual selective attention. *Nature*, 423(6939), 534–537. doi:10.1038/nature01647 PMID:12774121
- Green, C. S., & Bavelier, D. (2006). Enumeration versus multiple object tracking: The case of action video game players. *Cognition*, 101(1), 217–245. doi:10.1016/j.cognition.2005.10.004 PMID:16359652
- Greenfield, P. M., Brannon, C., & Lohr, D. (1994). Two-dimensional representation of movement through three-dimensional space: The role of video game expertise. *Journal of Applied Developmental Psychology*, 15(1), 87–103. doi:10.1016/0193-3973(94)90007-8
- Griffith, J. L., Voloschin, P., Gibb, G. D., & Bailey, J. R. (1983). Differences in eye-hand motor coordination of video-game users and non-users. *Perceptual and Motor Skills*, 57(1), 155–158. doi:10.2466/pms.1983.57.1.155 PMID:6622153
- Griffiths, M. (2002). The educational benefits of videogames. *Education for Health*, 20(3), 47–51.
- Groundspeak, Inc. (2016). *Geocaching: Join the world's largest treasure hunt*. Retrieved from <https://www.geocaching.com/play>

- Guillén-Nieto, V., & Aleson-Carbonell, M. (2012). Serious games and learning effectiveness: The case of *Its a Deal!* *Computers & Education*, 58(1), 435–448. doi:10.1016/j.compedu.2011.07.015
- Herbst, I., Braun, A.-K., McCall, R., & Broll, W. (2008). Timewarp: Interactive time travel with a mobile mixed reality game. *Proceedings of the 10th international conference on Human computer interaction with mobile devices and services* (pp. 235–244). Amsterdam, Netherlands: ACM Press. doi:10.1145/1409240.1409266
- Holden, C. L., & Sykes, J. M. (2011). Leveraging mobile games for place-based language learning. *International Journal of Game-Based Learning*, 1(2), 1–18. doi:10.4018/ijgbl.2011040101
- Huizenga, J., Admiraal, W., Akkerman, S., & ten Dam, G. (2009). Mobile game-based learning in secondary education: Engagement, motivation and learning in a mobile city game. *Journal of Computer Assisted Learning*, 25(4), 332–344. doi:10.1111/j.1365-2729.2009.00316.x
- Jacobs, J. W., & Dempsey, J. V. (1993). Simulation and gaming: Fidelity, feedback, and motivation. In J. V. Dempsey & G. C. Sales (Eds.), *Interactive instruction and feedback*. Englewood Hills, NJ: Educational Technology Publications.
- Jegers, A., & Wiberg, M. (2006). Pervasive gaming in the everyday world. *IEEE Pervasive Computing / IEEE Computer Society [and] IEEE Communications Society*, 5(1), 78–85. doi:10.1109/MPRV.2006.11
- Kahn, K. (1999). *A computer game to teach programming*. Paper presented at the National Educational Computing Conference, Atlantic City, NJ.
- Kamarainen, A. M., Metcalf, S., Grotzer, T., Browne, A., Mazzuca, D., Tutwiler, M. S., & Dede, C. (2013). Ecomobile: Integrating augmented reality and probeware with environmental education held trips. *Computers & Education*, 68, 545–556. doi:10.1016/j.compedu.2013.02.018
- Kebritchi, M., & Hirumi, A. (2008). Examining the pedagogical foundations of modern educational computer games. *Computers & Education*, 51(4), 1729–1743. doi:10.1016/j.compedu.2008.05.004
- Kinsell, C., DaCosta, B., & Nasah, A. (2014). Simulation games as interventions in the promotion of social skills development among children with autism spectrum disorders. In B. DaCosta & S. Seok (Eds.), *Assistive technology research, practice, and theory* (pp. 160–180). Hershey, PA: IGI Global; doi:10.4018/978-1-4666-5015-2.ch011
- Kirriemuir, J. (2002). *The relevance of games and gaming consoles to the higher and further education learning experience*. Techwatch Report TSW 02.01. Retrieved from http://www.jisc.ac.uk/uploaded_documents/tsw_02-01.rtf
- Klopfer, E., Squire, K., & Jenkins, H. (2002). Environmental detectives: PDAs as a window into a virtual simulated world. *Proceedings from the 1st IEEE International Workshop on Wireless and Mobile Technologies in Education* (pp. 95–98). Washington, DC: IEEE Computer Society. doi:10.1109/WMTE.2002.1039227
- Kretschmer, U. (2002). Using mobile systems to transmit location based information. *Proceedings of the ISPRS Commission III Symposium 2002, Photogrammetric Computer Vision PCV'02*, 129–133.
- Licoppe, C., & Inada, Y. (2006). Emergent uses of a location-aware mobile game: The interactional consequences of mediated encounters. *Mobilities*, 1(1), 39–61. doi:10.1080/17450100500489221
- Lindt, I., Ohlenburg, J., Pankoke-Babatz, U., & Ghellal, S. (2007). A report on the cross-media game epidemic menace. *Computers in Entertainment*, 5(1), 8. doi:10.1145/1236224.1236237

- Mathews, J., Holden, C., Jan, M.-F., & Martin, J. (2008). Sick at South Shore Beach: A place-based augmented reality game as a framework for building evidence-based arguments. *Proceedings of the 8th International Conference for the Learning Sciences* (pp. 89-90). International Society of the Learning Sciences.
- Matyas, S., Matyas, C., Schlieder, C., Kiefer, P., Mitarai, H., & Kamata, M. (2008). Designing location-based mobile games with a purpose – Collecting geospatial data with *CityExplorer*. *Proceedings of the 2008 International Conference on Advances in Computer Entertainment Technology* (pp. 244-247). Yokohama, Japan: ACME Press. doi:10.1145/1501750.1501806
- Mitchell, A., & Savill-Smith, C. (2004). The use of computer and video games for learning: A review of the literature. London, UK: Ultralab: Learning Skills and Development Agency.
- Montola, M. (2011). A ludological view on the pervasive mixed-reality game research paradigm. *Personal and Ubiquitous Computing*, 15(1), 3–12. doi:10.1007/s00779-010-0307-7
- Naismith, L., Lonsdale, P., Vavoula, G., & Sharples, M. (2004). Report 11: Literature review in mobile technologies and learning. *FutureLab*. Retrieved from http://www2.futurelab.org.uk/resources/documents/lit_reviews/Mobile_Review.pdf
- Orosy-Fildes, C., & Allen, R. W. (1989). Psychology of computer use: XII. Videogame play: Human reaction time to visual stimuli. *Perceptual and Motor Skills*, 69(1), 243–247. doi:10.2466/pms.1989.69.1.243 PMID:2928053
- Paraskeva, F., Mysirlaki, S., & Papagianni, A. (2010). Multiplayer online games as educational tools: Facing new challenges in learning. *Computers & Education*, 54(2), 498–505. doi:10.1016/j.compedu.2009.09.001
- Perrotta, C., Featherstone, G., Aston, H., & Houghton, E. (2013). *Game-based learning: Latest evidence and future directions (NFER Research Programme: Innovation in Education)*. Slough, UK: NFER.
- Pierfy, D. A. (1977). Comparative simulation game research: Stumbling blocks and stepping-stones. *Simulation & Games*, 8(2), 255–268. doi:10.1177/003755007782006
- Price, S., Rogers, Y., Scaife, M., Stanton, D., & Neale, H. (2003). Using tangibles to promote novel forms of playful learning. *Interacting with Computers*, 15(2), 169–185. doi:10.1016/S0953-5438(03)00006-7
- Ricci, K. E., Salas, E., & Cannon-Bowers, J. (1996). Do computer-based games facilitate knowledge acquisition and retention? *Military Psychology*, 8(4), 295–307. doi:10.1207/s15327876mp0804_3
- Rogers, Y., & Price, S. (2006). Using ubiquitous computing to extend and enhance learning experiences. In M. van 't Hooft & K. Swan (Eds.), *Ubiquitous computing in education: Invisible technology, visible impact* (pp. 329–347). Mahwah, NJ: Lawrence Erlbaum.
- Rogers, Y., Scaife, M., Gabrielli, S., Smith, H., & Harris, E. (2002). A conceptual framework for mixed reality environments: Designing novel learning activities for young children. *Presence (Cambridge, Mass.)*, 11(6), 677–686. doi:10.1162/105474602321050776
- Roschelle, J., & Pea, R. D. (2002). A walk on the WILD side: How wireless handhelds may change computer-supported collaborative learning. *Proceedings from the International Conference on Computer Support for Collaborative Learning 2002* (pp. 51-60). Mahwah, NJ: Lawrence Erlbaum. doi:10.1075/ijct.1.1.09ros
- Sánchez, J., Salinas, A., & Sáenz, M. (2006). Mobile game-based science learning. *Proceedings from the Distance Learning and the Internet Conference* (pp. 18-30). Tokyo, Japan: APRNet 2006.

- Schmitz, B. (2014). *Mobile games for learning: A pattern-based approach* (Unpublished master's thesis). Open Universiteit in the Netherlands at Welten Institute – Research Centre for Learning, Teaching and Technology (Dutch Research School for Information and Knowledge Systems Dissertation Series No. 2014-45).
- Schwabe, G., & Göth, C. (2005). Mobile learning with a mobile game: Design and motivational effects. *Journal of Computer Assisted Learning*, 21(3), 204–216. doi:10.1111/j.1365-2729.2005.00128.x
- Sedano, C. I., Laine, T. H., Vinni, M., & Sutinen, E. (2007). Where is the answer? – The importance of curiosity in pervasive mobile games. *Proceedings of the 2007 Conference on Future Play*, 46–53. doi:10.1145/1328202.1328211
- Sedlak, R. A., Doyle, M., & Schloss, P. (1982). Video games – A training and generalization demonstration with severely retarded adolescents. *Education and Training in Mental Retardation and Developmental Disabilities*, 17(4), 332–336.
- Sharples, M., Corlett, D., & Westmancott, O. (2002). The design and implementation of a mobile learning resource. *Personal and Ubiquitous Computing*, 6(3), 220–234. doi:10.1007/s007790200021
- Sheldon, J., Perry, J., Klopfer, E., Ong, J., Hsueh-Hua Chen, V., Tzuo, P. W., & Rosenheck, L. (2010). Weatherlings: A new approach to student learning using web-based mobile games. *Proceedings of the Fifth International Conference on the Foundations of Digital Games* (pp. 203–208). Monterey, CA: ACM Press. doi:10.1145/1822348.1822375
- Shiratuiddin, N., & Zaibon, S. B. (2001). Mobile game-based learning with local content and appealing characters. *Mobile Learning and Organisation*, 4(1), 55–82. doi:10.1504/IJMLO.2010.029954
- Sotamaa, O. (2002). All the world's a Botfighter stage: Notes on location-based multi-user gaming. *Proceedings from the Computer Games and Digital Cultures Conference* (pp. 35–44). Tampere, Finland: Tampere University Press.
- Sotirou, S., Chryssafidou, E., Owen, M., Mario, B., Zistler, E., Lohr, M., ... Spikol, D. (2008). *The COLLAGE project: Guide of good practice for mobile and game based learning*. Retrieved from http://www.ea.gr/ea/myfiles/File/publications/books/Collage_GGP2008.pdf
- Spikol, D., & Milrad, M. (2008). Physical activities and playful learning using mobile games. *Research and Practice in Technology Enhanced Learning*, 3(3), 275–295. doi:10.1142/S1793206808000562
- Squire, K. D., & Jan, M. (2007). Mad City Mystery: Developing scientific argumentation skills with a place-based augmented reality game on handheld computers. *Journal of Science Education and Technology*, 16(1), 5–29. doi:10.1007/s10956-006-9037-z
- Stanton, D., O'Malley, C., Ng, K. H., Fraser, M., & Benford, S. (2003). Situating historical events through mixed reality. In B. Wasson, S. Ludvigsen, & U. Hoppe (Eds.), *Designing for change in networked learning environments: Proceedings of the Conference on Computer Support for Collaborative Learning (CSCL2003)* (pp. 293–302). Amsterdam, Netherlands: Springer.
- Trinkaas, J. W. (1983). Arcade video games: An informal look. *Psychological Reports*, 52(2), 586–586. doi:10.2466/pr0.1983.52.2.586
- Veenhof, G., Sandberg, J. A. C., & Maris, M. G. (2012). *ZooQuest*: A mobile game-based learning application for fifth graders. In S. A. Cerri, W. J. Clancey, G. Papadourakis, & K. Panourgia (Eds.), *Proceedings of the 11th International Conference on Intelligent Tutoring Systems* (pp. 687–688). Heidelberg, Germany: Springer. doi:10.1007/978-3-642-30950-2_122

Vogiazou, Y., & Raijmakers, B. (2004). Urban space as a large-scale group playground. *Proceedings from the 6th Annual Ubiquitous Computing Conference*, 54-55.

Wang, C.-Y., & Lai, A.-F. (2011). Development of a Mobile Rhythm Learning System based on digital game-based learning companion. In M. Chang, W.-Y. H., M.-P. Chen, & W. Müller (Eds.), *Proceedings from the 6th International Conference on E-Learning and Games, Edutainment 2011* (pp. 92-100). doi:10.1007/978-3-642-23456-9_17

Wijers, M., Jonker, V., & Kerstens, K. (2008). Mobilemath: The phone, the game and the math. *Proceedings from the 2nd European Conference on Games-Based Learning* (pp. 507-516). Barcelona, Spain: Academic Conferences Ltd. doi:10.1007/s11858-010-0276-3

Wood, L. E., & Stewart, P. W. (1987). Improvement of practical reasoning skills with a computer game. *Journal of Computer-Based Instruction*, 14(2), 49-53.

Ybarrondo, B. A. (1984). *A study of the effectiveness of computer-assisted instruction in the high school biology classroom* (Master's thesis). Boise State University. (ERIC Document Reproduction Service No. ED 265 015)

Yuen, S., Yaoyuneyong, G., & Jonhson, E. (2011). Augmented reality: An overview and five directions for AR in education. *Journal of Educational Technology Development and Exchange*, 4(1), 119-140.

Yuji, H. (1996). Computer games and information-processing skills. *Perceptual and Motor Skills*, 83(2), 643-647. doi:10.2466/pms.1996.83.2.643 PMID:8902044

Zhang, J., & Lu, J. (2014). Using mobile serious games for learning programming. *Proceedings from the Fourth International Conference on Advanced Communications and Computation*. Retrieved from https://www.thinkmind.org/download.php?articleid=infocomp_2014_1_50_60029

ADDITIONAL READING

Avouris, N., & Yiannoutsou, N. (2012). A review of mobile location-based games for learning across physical and virtual spaces. *Journal of Universal Computer Science*, 18(15), 2120-2142. doi:10.3217/jucs-018-15-2120

Barlett, C. P., Anderson, C. A., & Swing, E. L. (2009). Video game effects – Confirmed, suspected, and speculative: A review of the evidence. *Simulation & Gaming*, 40(3), 377-403. doi:10.1177/1046878108327539

DaCosta, B., Seok, S., & Kinsell, C. (2015). Mobile games and learning. In Z. Yan (Ed.), *Encyclopedia of mobile phone behavior* (Vol. 1, pp. 46-60). Hershey, PA: IGI Global; doi:10.4018/978-1-4666-8239-9.ch004

Egenfeldt-Nielsen, S. (2006). Overview of research on the educational use of video games. *Digital Kompetanse*, 1(3), 184-213.

Ke, F. (2009). A qualitative meta-analysis of computer games as learning tools. In R. E. Furdig (Ed.), *Handbook of research on effective electronic gaming in education* (pp. 1-32). Hershey, PA: IGI Global; doi:10.4018/978-1-59904-808-6.ch001

Kirriemuir, J., & McFarlane, A. (2004). Literature review in games and learning. *A NESTA Futurelab Research report, Report 8*. Retrieved from <https://telearn.archives-ouvertes.fr/hal-00190453/document>

McClarty, K. L., Orr, A., Frey, P. M., Dolan, R. P., Vassileva, V., & McVay, A. (2012). A literature review of gaming in education: Research report. New York, NY: Pearson; Retrieved from http://researchnetwork.pearson.com/wp-content/uploads/Lit_Review_of_Gaming_in_Education.pdf

Mitchell, A., & Savill-Smith, C. (2004). The use of computer and video games for learning: A review of the literature. London, UK: Ultralab: Learning Skills and Development Agency.

Perrotta, C., Featherstone, G., Aston, H., & Houghton, E. (2013). *Game-based learning: Latest evidence and future directions (NFER Research Programme: Innovation in Education)*. Slough, UK: NFER.

KEY TERMS AND DEFINITIONS

Game-Based Learning (GBL): A notion rooted in the belief that games can be leveraged in the learning process.

Mobile Game-Based Learning (mGBL): A notion rooted in the belief that games played on a mobile device (e.g., cell phone, personal digital assistant, smartphone, tablet) can be leveraged in the learning process.

Mobile Games: Digitally based games mostly played on mobile devices (e.g., cell phone, personal digital assistant, smartphone, tablet). DaCosta, Seok, & Kinsell.

Video Games: Digitally based games typically played on PCs or dedicated gaming devices, such as a game consoles (e.g., Xbox®, PlayStation®, Wii®) or handheld game devices (e.g., 3DS®, Vita®). DaCosta, Seok, & Kinsell.

Mobile Game-Based Learning in STEM Subjects

Marcelo Leandro Eichler

Universidade Federal do Rio Grande do Sul, Brazil

Gabriela Trindade Perry

Universidade Federal do Rio Grande do Sul, Brazil

Ivana Lima Lucchesi

Universidade Federal do Rio Grande do Sul, Brazil

Thiago Troina Melendez

Instituto Federal Sul-Riograndense, Brazil

INTRODUCTION

The acronym STEM – Science, Technology, Engineering, and Mathematics has entered the agenda of educational policies. Particularly remarkable is the Horizon Report by Johnson et al. (2013), which identified the most notable and emerging issues, trends and digital technology challenges in STEM education for the coming years.

In today's schools, teaching about STEM may take place in many general education and career and technical education subject areas such as agriculture, science, health, technology and engineering, and family and consumer science. Learning about the attributes of STEM and how they connect can help promote innovation. At the collegiate level, STEM education encourages students to pursue STEM careers in order to meet the growing demand for trained professionals in these areas (Reeve, 2015).

The concept of STEM education is being discussed differently by different nations. Whereas some deem STEM education as the improved teaching of the separate STEM subjects, others believe STEM should be taught through an integrative subjects approach. Many believe it is a combination of both of these approaches. Different countries are exploring STEM due to

political and economic pressures and because some believe it is a means to improve the delivery of such knowledge (Ritz & Fan, 2015).

We acknowledge the strategic role of science and mathematics learning. But how to foster it? “Games” might be a promising answer. Games are successful at making people spend time trying to reach goals via a well structured set of rules - and STEM domains are also bound by well structured rules. Thus, if we could make games that would embed the rules of those domains into their gameplay, well, maybe our students would enjoy learning science and mathematics. Rapini (2012) points out that games are now being revisited as educational tools by several leading organizations, i.e.: MIT's Education Arcade and Games-to-Teach project; Woodrow Wilson Foundations' Serious Games Initiative; University of Wisconsin's Games Learning Society; the Federation of American Scientists; the Bill and Melinda Gates Foundation; and the U.S. Department of Education. Some of them - markedly those from MIT's Education Arcade project - have been developed with mobile devices in mind.

Regarding mobile device widespread use in STEM education, Johnson et al. (2013) indicated the relevance of games and gamification in STEM learning and teaching: a) discovery-based and

goal-oriented learning is inherent in educational games and this strategy offers opportunities for collaboration and development teamwork skills; b) educational games can be used to teach interdisciplinary concepts in many complex scientific issues in a more appealing way than traditional methods; c) simulations and serious games allow students to recreate difficult situations to try new answers or pose creative solutions.

And, lastly, our choice to narrow the focus of this paper down to mobile devices is due to the impressive influence they have over our lives and to the amount of time we spend on them. The revolution of mobile devices + wide internet access is so deep that it is making the very action of “making a phone call” obsolete. We have been using mobile devices for a myriad of tasks - so how can we use them for learning purposes? These are the concerns presented in this paper.

BACKGROUND

The development of mobile game-based learning has been seen as a new line of research and technological development in the field of educational technology, STEM education and game design.

These fields are rather new and intrinsically multidisciplinary, making it even more exciting. Regarding the use of mobile devices for teaching and conceptual learning, we can highlight the work of John Traxler. He is the Director of the Learning Lab at the University of Wolverhampton and a Founding Director of the International Association for Mobile Learning, among many other academic attributions related to m-learning.

Due to their role as pioneers on Multimedia Studies, we cite professor Richard Clark – from the University of Southern California – and professor Richard Mayer – from the University of California – since they are on both sides of the scale: Clark says instructional design is all that matters; Mayer says multimedia design is also a key for effectiveness. Professor Robert Kozma – currently (2016) emeritus director and principal

scientist of the Center for Technology in Learning – could be pointed out as a leading scholar. His numerous publications focus on the potential of IT in learning contexts, as well as Science and Mathematics Education (he is a multidisciplinary researcher).

And lastly, regarding Game Design, we would highly recommend our readers the work of Stefan Björk, Göteborg University, Sweden, who is certainly one of the leading scholars in the field of Game Design. Regarding Games and Affection studies, we would cite Boehner et al (2017) and Quigley (2014), who have been studying how the theoretical foundations of affection interact with the instruments used to measure it.

CURRENT SCIENTIFIC KNOWLEDGE IN SCIENCE AND MATHEMATICS LEARNING GAMES FOR MOBILE PLATFORMS

There is a growing interest in Mobile Game related research, whether strictly technological or applied in social contexts. We thought a nice way to connect this data and depict the current scenario would be to break the subject into pieces: Game Design; Affection and Play; Mobile Learning; Games for Learning; Science and Mathematics Education; and lastly, summarizing it in the Games for STEM Education section.

Game Design

The association is direct: people enjoy playing games, people love their mobile gadgets; if we want people to enjoy better learning performance and experiences, we should have them play mobile games! But before engaging with game development, it is important to not only acknowledge the technological options, but to also be aware of the design process and workflow. The reasoning for this is because it is most certainly teachers who will be asking for the development of a game, and they are likely to also be game designers (because

the development team do not master the subject domain). According to O'Neil et al (2005), Linehan et al (2011), and Brom et al (2011), while it seems apparent that games have the potential to function as valuable teaching tools, there is very little evidence that they 1) produce reliable, valid and long lasting educational outcomes, or 2) do so better than traditional education structures. Linehan et al (2011) highlight the importance of content, but you should not focus too heavily on it to the detriment of gameplay; it must be intrinsic to gameplay. A fictitious example of bad anchoring to content would be a game where you have to save a princess by answering correctly to algebraic problems. For these reasons, it might be worth understanding how games may be designed.

With many different types of games, target audiences and content demands, there is not a large number of scientific studies on this subject. Even when it comes to books on the game design process, plenty of different approaches can be found depending on game type and platform (mobile, digital and board games). As references of such variety, we cite Barwood's (2001) game design guidelines, patterns for game design (Kreimeier, 2002; Björk & Holopainen, 2005) and tools for supporting specific game design tasks (Cook, 2012; Hunicke et al. 2004; Brathwaite & Schreiber, 2008; Davidsson et al., 2004). Furthermore, we think we should mention two well known game design methods: Schell (2008) and Fullerton (2014). Both authors propose an iterative method based on play testing with prototypes of varying degrees of functionality. None of those references is specific for educational games or even mobile games, but they are all very enlightening, even for persons who have some experience.

Out of these, we would like to highlight the design approaches to game design found in Hunicke et al. (2004) and Björk & Holopainen (2005). The book *Patterns in Game Design* (Björk & Holopainen, 2005) is a comprehensive collection of more than 200 design patterns – descriptions of recurring interaction elements of gameplay that are shared by several games.

Each pattern has a name and a description, and also include the relation with other patterns and consequences/examples of use. It is a great way to become familiarized with game vocabulary and to recognize game design strategies that might be subconscious even to experienced game players. Hunicke et al. (2004) has a similar but broader approach, since it does not point to gameplay elements. They divide every game into three zones: rules, systems and fun, which, in turn, are linked to design counterparts: mechanics, gameplay and aesthetics. Mechanics create gameplays that produce aesthetic experiences.

Next we review studies focusing on the assessment of affective states while playing educational games. We consider this an important subject because affection might be the most prominent difference between games and all other pedagogical strategies (digital or not).

Affection and Play

A number of studies focus on the relation between affective states and gameplaying. In this section, we review some of those studies. With affection being a very broad subject that encompasses several areas (mainly psychology, neuroscience and artificial intelligence), we narrowed our search down to educational game evaluation. Those studies are mainly from the area of Human Computer Interaction, an applied science area. The authors of such papers usually restrict their experimental designs to one or two affective states, as it is difficult to identify and distinguish between such states. For a thorough analysis of theoretical views on affection, we recommend the review by Boehner et al. (2007) and Quigley (2014).

The studies we have reviewed relate to motivation and flow, and how these relate to game design. For example, Kapp (2012) and Clark & Rossiter (2008) investigated the relation between corrective feedback and motivation, while Brookes e Moseley (2012) and Wood et al. (2014) investigated how the possibility of failing and starting again can enhance motivation. We believe these

design elements are important because they favor hypothesis testing in the game.

Increasing motivation is also related to the perceived challenge posed by the game (Hanus & Fox, 2015; Sharek & Wiebe, 2014; Admiral et al, 2011). If it is challenging and the player succeeds, they can feel very proud of their achievements, otherwise, there might be feelings of helplessness and incompetence (Mc Gonigal, 2011), which, of course, game designers try to avoid. A well balanced challenge can lead to engagement, an affective state similar to flow but with less pronounced effects (in Jennett et al, 2008 the reader will find a comprehensive distinction between those states). Level design might be an important key to balancing challenge (Sharek & Wiebe, 2014; O'Brien & Toms, 2010; Pavlas et al, 2010). The flow state is deep involvement with the game, high concentration, and gradual loss of time awareness (O'Brien & Toms, 2010; Pavlas et al, 2010; Sharek & Wiebe, 2014; Shute et al, 2014). Breaking flow states leads to great frustration. Hsieh et al (2016) analyzed the relation between flow experiences and performance on students of elementary schooling level and saw that both variables work together.

Mobile Learning

The expression *mobile learning* may be associated with several meanings depending on the context it is inserted within, user objectives, geographical location, and even the era of reference. Traxler (2009) reports on how the different facets hinder an outline of a description as to what mobile learning would be. Following a global analysis that considered all the continents, the definition by Taylor (2006) stands out, according to which *mobile learning*, or as he deems more appropriate, *learning in the mobile age*, means learning through handheld equipment, or; learner mobility carrying their mobile devices, or; mobility of knowledge and/or resources accessible anywhere.

In that same sense, Sawaya & Putnam (2015) and Crompton (2013) adopt the *m-learning*

concept as “learning within multiple contexts, through social and content interactions making use of mobile, personal electronic equipment”. Therefore, they consider the use of handheld equipment (tablets and smartphones) in education in both senses. Hung & Zhang (2012) foresee a huge growth in the amount of publications in this field and provided a list of countries that are source of research on Mobile Learning (n = 119 papers) – Taiwan (27%), United States (15%), and South Korea (9%). Hwang & Tsai (2011) also analyzed this variable and found similar results: Taiwan (42%), United States (10%), and United Kingdom (13%).

There is some very remarkable data in the reviews by Wu et al (2012), Hwang & Tsai (2011) and Connolly et al (2012): the distribution of subjects that are the most frequent themes. We point out the relatively small number of Science and Mathematics related game studies, excluding Computer Science, which would naturally have more studies. For us it was a surprise to find such a high amount of studies related to Language and Linguistics.

Games for Learning

James Paul Gee (2003, 2008), a prominent advocate of Games for Learning, listed several properties of games that make them special tools for learning, e.g.: “the use of affordance matches between bodies or tools and worlds” and “using models to make learning from concrete experience more general and abstract”. Gee (2008, 2009) considers successful video games as highly engaging problem-solving spaces. Since they are often long, difficult and complex, they must be learned and mastered in effective ways. Gee & Morgridge (2007, p.1028) believe that video games make players think like scientists, with gameplay being built on a cycle typical of experimental science: “hypothesize, probe the world, get a reaction, reflect on the results, re-probe for better results”.

Regarding Games for Learning, we found two recent reviews and one somewhat older review.

One of the more recent reviews carried out by Connolly et al (2012) had each paper assigned a “quality score” – the lowest score possible was 5 and the highest 15. The mean rating of 8.56, and 70 papers (54% out of 129) were above average. Out of these high quality papers, 52% were about “entertainment”, 37% were “games for learning”, and 11% were “serious games”. Regarding game genre, from the 34 high quality papers about “Games for Learning” or “Serious Games”, 6% (2) were Action games; 73% (25) were Simulations; none were Mobile games.

The review by O’Neil et al (2005), albeit older, has been referred to by several previous studies. They analyzed 19 papers from 1990 to 2005 to support the hypothesis that positive findings might be attributed to instructional design, not to games per se (p.462). This hypothesis is in consonance with Clark’s “method-not-medium” hypothesis, which seems to have very low popularity nowadays, despite the strength of Clark’s arguments. Regarding this subject, the reader might want to follow Richard Clark and Robert Kozma’s debate. Robert Kozma is an eminent advocate of multimedia potentiality, especially for Science Learning. Some of the most interesting papers are Clark (1994); Kozma (1994); Clark & Feldon (2005).

Li & Tsai’s (2013) literature review found a gap between students’ possible learning experience with games and their learning outcomes being assessed: the games did not fully take advantage of the potential of genres and platforms. The authors cite examples of MMO (Massive Multiplayer On-line) games in which the students explored the world and solved tasks individually. Moreover, even when game designs were guided by different theories or models to provide a variety of learning opportunities, learning was still mainly focused on scientific and mathematical knowledge or concepts. We agree with Li & Tsai (2013) when they say the affordance of game environments to facilitate science learning has not yet been fully explored and requires further investigation.

Science and Mathematics Education

Nowadays (in 2016), it could be said that, within the Science and Mathematics teaching and learning context, we are experiencing a paradigm shift regarding the use of mobile technologies in school activities.

Borba et al. (2014) listed the four phases in the use of digital technologies in Mathematics Education, and the respective tools that may represent them: 1st phase – computer programming, constructionism and Seymour Papert’s LOGO language; 2nd phase – dynamic geometry software such as Cabri Géomètre and GeoGebra; 3rd phase – Internet and distance education; and the current 4th phase – fast Internet and mobile technologies. The 4th phase seems to retrieve aspects from the previous phases that may be better explored due to the technologies that were non-existent in their original context.

Borba & Villarreal (2005) state that knowledge is being produced collectively by human beings and technologies and coined the expression *humans-with-media* in opposition to the idea that pieces of knowledge are only built by persons, either in groups or individually.

Bekebrede et al. (2011) defined this generation of students as the *gamer generation* or *net generation*, who have their lifestyles and social interactions harnessed to the digital technologies they have lived with since childhood. Those students express a preference for active, collaborative, technology-rich learning, which is motivation to include digital games in teaching practices.

With emphasis on the teaching of Mathematics, Misfeldt & Gjedde (2015) consider three ratings for educational games: i) as a means to present some piece of knowledge that stands out in relation to other traditionally adopted means; ii) as an instrument that controls or directs a more complex activity; or iii) simply as an object that may be developed and analyzed by teachers and students.

Games for STEM Education

Masek et al (2012) indicate that digital games allow the handling of multiple representations associated with scientific knowledge. In this way, multimodal representations can be used to scaffold the construction of understanding, scientific explanations and reasoning. Girard et al (2013) suggest that the use of games also allows for the harmless simulation of many physical situations and natural phenomena that cannot be reproduced in real-world situations (such as ecological disasters, astronomical phenomena and emergencies). The Graz University of Technology produces several types of Mathematical applications for iPhones and iPads. Based on those products, Ebner (2015) lists four categories:

1. **Stand-Alone Learning Apps:** Applications that do not require registry or Web connection, allowing to be accessed anywhere. Pedagogically, they do not require outside instructions or other persons to be played, which hinders any learning assessment.
2. **Game-Based Learning Apps:** Feature some problem solving characteristics, such as variable answers and problematization of a contextualized situation, with added competitiveness. The increase in gamification has mitigated high programming costs.
3. **Collaborative Apps:** Seek to stimulate group activity and collaborative learning social interactions in addition to exploring the differentials of mobile devices through Bluetooth communication or WiFi connection.
4. **Learning Analytics Apps:** This category aims to aid teachers in verifying student performance by accessing individual data in each device, allowing for further statistical analysis.

For meaningful learning of STEM concepts to take place, it is important that teachers and app developers take into account three pedagogical

issues: the possibilities the mobile device allows; learning objectives; and types of activities that may be elaborated (Sawaya & Putnam, 2015).

1. **Possibilities:** Insert data and perform computing procedures such as complex calculations; - Access information on websites, audios videos, etc.; capture images, videos, audios and localization via GPS; communicate and share information and files for joint work; and create materials that may be promptly used by colleagues and teachers.
2. **Learning Objectives:** Construct new items of knowledge through problem solving within mathematics or contextualization; relate and apply mathematics in other areas of knowledge; make use of representations to model and interpret physical, social and mathematical phenomena.
3. **Types of Activities:** Presentation of new content; mathematical comprehension exercise tool; investigation activities; applications in different contexts; creation of diversified materials.

It is important to highlight that all above listed features already exist in schools. The advantage of mobile devices is the capability of aggregating and optimizing all possibilities in a single electronic apparatus. With emphasis on geometry teaching, location-based game MobileMath (Wijers et al., 2008; Wijers et al., 2010) is an ideal example of a mobile learning game that includes such requirements.

FUTURE RESEARCH DIRECTIONS

To open this section, we would like to rely on our experience as designers of Games for STEM Education (Eichler et al, 2003; Eichler et al, 2005; Eichler et al, 2011; Perry e Eichler, 2015; Perry et al., 2012; Perry & Schnaid, 2012) to point out possible causes for the gap Li & Tsai (2013) found between students' possible

learning experience with games and their actual assessed learning outcomes. We hypothesize that it is due to the nature of game development. Games are the most difficult kind of software to both design and develop – code and design requirements are usually unique, not shared among projects, and usually complex (Blow, 2004). The more sophisticated a game is, the more difficult (and expensive) it becomes to find skilled personnel. We see some possible ways to overcome this situation:

1. When the game is to be developed by educational institutions, money and time are the critical resources, in that order. Financial resources could be allocated for hiring professional game designers and developers. Subject domain knowledge, as a resource, is available in educational institutions, with educators having classroom experience, knowing what topics alumni find more difficult and the pedagogical strategies to approach those topics.
2. When the game is to be developed by educational institutions, financial and time resources could be allocated for personnel training — educators, researchers, technical staff and students. The use of financial resources would decrease vastly, but it would take some time to create a cohesive and mature team. Additionally, design and development could be restrained by the skill set of the team. This could be an interesting institutional approach, since equilibrium would be reached at some point, after which the ROI (Return On Investment) would be positive.
3. When the game is to be developed by studios, then the critical resource is knowledge of the subject domain. Studios could associate with educators to bridge this gap. In this case, it is important to give educators time to become familiar with the game development process.

In our experience, we have perceived that mobile technologies are already present in schools, but still face obstacles for their acceptance in the classroom. That is why prohibitive norms should be reviewed and activities carried out to motivate teachers to accept incorporating such resources in their school practices.

Traxler & Crompton (2015) highlight that Bring Your Own Device (BYOD) may be a good option for Web access at schools be broadened out to every environment and to all students. As part of this project, each one may use their own personal equipment in professional and schooling environments, leveraging the speed of wireless Internet. In addition to digital inclusion, this action would reduce equipment acquisition and maintenance costs and eliminate the need for an underused computing laboratory.

Following this idea, Borba & Lacerda (2015) suggest the One Cellular Phone per Student Program for Brazilian schools, being motivated by the modest advances observed in previous national projects, also taking into account the *humans-with-media* present in schools and the need to encourage continued training for the faculty who are in classrooms already and those in early training.

Lastly, we suggested in Perry & Eichler (2015) that one path to be taken for further qualification in digital educational games would be to encourage their development by teaching institutions, though that is still at an early stage and lacking an attractive visual design.

CONCLUSION

In the face of the above exposed, our conclusions point to the fact that we are taking the first steps in a digital game development process for teaching mathematics in the school environment and the acceptance of Smartphones as tools that add value to education. By uniting both trends, it is possible to direct game design to promote collaborative activities and go beyond the classroom.

REFERENCES

- Admiral, W., Huizenga, J., Akkerman, S., & ten Dam, G. (2011). The concept of flow in collaborative game-based learning. *Computers in Human Behavior*, 27(3), 1185–1194. doi:10.1016/j.chb.2010.12.013
- Barwood, H. (2001). Four of the Four Hundred. *GDC lecture*. Available at: <<http://www.finitearts.com/pages/400page.html>>
- Bekebrede, G., Warmelink, H. J. G., & Mayer, I. S. (2011). Reviewing the need for gaming in education to accommodate the net generation. *Computers & Education*, 57(2), 1521–1529. doi:10.1016/j.compedu.2011.02.010
- Björk, S., & Holopainen, J. (2005). *Patterns in Game Design*. Charles River Media.
- Blow, J. (2004). Game Development: Harder Than You Think. *Queue*, 1(10), 28–37. doi:10.1145/971564.971590
- Boehner, K., DePaula, R., Dourish, P., & Sengers, P. (2007). How emotion is made and measured. *International Journal of Human-Computer Studies*, 65(4), 275–291. doi:10.1016/j.ijhcs.2006.11.016
- Borba, M. C., & Lacerda, H. D. G. (2015). Políticas Públicas e Tecnologias Digitais: Um celular por aluno. *Educação Matemática: Pesquisa*, 17, 490–507.
- Borba, M. C., Scucuglia, R. R. S., & Gadanidis, G. (2014). *Fases das Tecnologias Digitais em Educação Matemática: sala de aula e internet em movimento. 1*. Belo Horizonte: Autêntica.
- Borba, M. C., & Villarreal, M. V. (2005). *Humans-With-Media and the Reorganization of Mathematical Thinking: information and communication technologies, modeling, experimentation and visualization*. New York: Springer.
- Brathwaite, B., & Schreiber, I. (2008). *Challenges for game designers*. Boston: Course Technology.
- Brom, C., Preuss, M., & Klement, D. (2011). Are educational computer micro-games engaging and effective for knowledge acquisition at high-schools? A quasi-experimental study. *Computers & Education*, 57(3), 1971–1988. doi:10.1016/j.compedu.2011.04.007
- Brookes, S., & Moseley, A. (2012). Authentic contextual games for learning. In N. Whitton & A. Moseley (Eds.), *Using games to enhance learning and teaching: A beginner's guide* (pp. 91–107). London: Routledge.
- Clark, M. C., & Rossinter, M. (2008). Narrative Learning in the Adult Classroom. In *AERC - Adult Education Research Conference*. Available in: <http://newprairiepress.org/aerc/2008/papers/13>
- Clark, R. E. (1994). Media will never influence learning. *Educational Technology Research and Development*, 42(2), 21–29. doi:10.1007/BF02299088
- Clark, R. E., & Feldon, D. F. (2005). Five common but questionable principles of multimedia learning. In R. E. Mayer (Ed.), *Cambridge Handbook of Multimedia Learning* (pp. 1–22). Cambridge University Press. doi:10.1017/CBO9780511816819.007
- Connolly, T. M., Boyle, E. A., MacArthur, E., Hailey, T., & Boyle, J. M. (2012). A systematic literature review of empirical evidence on computer games and serious games. *Computers & Education*, 59(2), 661–686. doi:10.1016/j.compedu.2012.03.004
- Cook, D. (n.d.). Loops and Arcs. *LOSTGARDEN*. Available at <http://www.lostgarden.com/2012/04/loops-and-arcs.html>

- Crompton, H. (2013). A historical overview of mobile learning: Toward learner-centered education. In *Handbook of mobile learning* (pp. 3–14). Routledge.
- Davidsson, O., Peitz, J., & Björk, S. (2004). *Game design patterns for mobile games. Project Rep.* Nokia Research Center.
- Ebner, M. (2015). Mobile Applications For Math Education – How Should They Be Done? In H. Crompton & J. Traxler (Eds.), *Mobile Learning and Mathematics. Foundations, Design, and Case Studies* (pp. 20–32). New York: Routledge.
- Eichler, M. L., Gonçalves, M. R., Silva, F. O. M., Junges, F., & Del Pino, J. (2003). Virtual learning environments designed in Brazil. *Educational Technology*, 43(6), 58–60.
- Eichler, M. L., Perry, G. T., & Fritsch, G. (2011). Xenubi: The development of a chemistry educational game for mobile phones. *Proceedings of the IADIS International Conference Game and Entertainment Technologies 2011, Part of the IADIS Multi Conference on Computer Science and Information Systems 2011, MCCSIS 2011* (pp. 151–153). Academic Press.
- Eichler, M. L., Xavier, P. R., Araújo, R. C., Forte, R. C., & Del Pino, J. C. (2005). Carbopolis: A Java Technology-Based Free Software for Environmental Education. *Journal of Computers in Mathematics and Science Teaching*, 24(1), 43–72.
- Fullerton, T. (2014) *Game Design Workshop: A Playcentric Approach to Creating Innovative Games*. Boca Raton, FL: CRC Press.
- Gee, J. P. (2008). Cats and Portals: Video Games, Learning, and Play. *American Journal of Play*, 1(2), 229–245.
- Gee, J. P. (2009). Deep Learning Properties of Good Digital Games: How Far Can They Go? Theories and Mechanisms: Serious Games for Learning (pp. 63–80). Academic Press.
- Gee, J. P., & Morgridge, T. (2007). Being a Lion and Being a Soldier Learning and Games. In Facer (Ed.), *Savannah: Mobile Gaming and Learning?* (pp. 1027–1041). Academic Press.
- Girard, C., Ecalle, J., & Magnan, A. (2013). Serious games as new educational tools: How effective are they? A meta-analysis of recent studies. *Journal of Computer Assisted Learning*, 29(3), 207–219. doi:10.1111/j.1365-2729.2012.00489.x
- Hanus, M. D., & Fox, J. (2015). Assessing the effects of gamification in the classroom: A longitudinal study on motivation, satisfaction, effort, and grades. *Computers & Education*, 80, 152–161. doi:10.1016/j.compedu.2014.08.019
- Hsieh, Y.-H., Lin, Y.-C., & Hou, H.-T. (2016). Exploring the role of flow experience, learning performance and potential behavior clusters in elementary students game-based learning. *Interactive Learning Environments*, 24(1), 178–193. doi:10.1080/10494820.2013.834827
- Hung, J.-L., & Zhang, K. (2012). Examining mobile learning trends 2003–2008: A categorical meta-trend analysis using text mining techniques. *Journal of Computing in Higher Education*, 24(1), 1–17. doi:10.1007/s12528-011-9044-9
- Hunick, R., Leblanc, M., & Zubek, R. (2004). MDA: A formal approach to game design and game research. *Proceedings of the Challenges in Games AI Workshop, 19th National Conference on Artificial Intelligence*.
- Hwang, G.-J., & Tsai, C.-C. (2011). Research trends in mobile and ubiquitous learning: A review of publications in selected journals from 2001 to 2010. *British Journal of Educational Technology*, 42(4), E65–E70. doi:10.1111/j.1467-8535.2011.01183.x
- Jennett, C., Cox, A. L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., & Walton, A. (2008). Measuring and defining the experience of immersion in games. *International Journal of Human-Computer Studies*, 66(9), 641–661. doi:10.1016/j.ijhcs.2008.04.004

- Johnson, L., Adams Becker, S., Estrada, V., & Martín, S. (2013). *Technology Outlook for STEM+ Education 2013-2018: An NMC Horizon Project Sector Analysis*. Austin, TX: The New Media Consortium.
- Kapp, K. M. (2012). What every chief learning officer needs to know about games and gamification for learning? Department of Instructional Technology and Institute for Interactive Technologies. Available in: karlkapp.com/category/gamification/
- Kozma, R. B. (1994). Kozma - 1994 - Will Media Influence Learning Reframing the Debate. *Educational Technology*, 42, 7–19.
- Kreimeier, B. (2002). The Case For Game Design Patterns. *Gamasutra*. Available at http://www.gamasutra.com/view/feature/132649/the_case_for_game_design_patterns.php
- Li, M.-C., & Tsai, C.-C. (2013). Game-based learning in science education: A review of relevant research. *Journal of Science Education and Technology*, 22(6), 877–898. doi:10.1007/s10956-013-9436-x
- Linehan, C., Kirman, B., Lawson, S., & Chan, G. G. (2011). Practical, Appropriate, Empirically-Validated Guidelines for Designing Educational Games. CHI 2011, Vancouver, Canada. doi:10.1145/1978942.1979229
- Masek, M., Murcia, M., & Morison, J. (2012). Getting serious with iPads: The intersection of game design and teaching principals. *Australian Educational Computing*, 27(2), 34–38.
- Mayer, R. (2001). *Multimedia Learning*. Cambridge Press. doi:10.1017/CBO9781139164603
- McGonigal, J. (2012). *Realidade em jogo: por que os games nos tornam melhores e como eles podem mudar o mundo*. Rio de Janeiro: Best Seller.
- Misfeldt, M., & Gjedde, L. (2015) What is a Game for Geometry Teaching: Creative, Embodied and Immersive Aspects. In *Proceedings of the 9th European Conference on Games Based Learning ECGBL 2015*. (pp. 378-386). Academic Conferences Limited.
- O'Brien, H. L., & Toms, E. G. (2010). The development and evaluation of a survey to measure user engagement. *Journal of the American Society for Information Science and Technology*, 61(1), 50–69. doi:10.1002/asi.21229
- O'Neil, H. F., Wainess, R., & Baker, E. L. (2005). Classification of learning outcomes: evidence from the computer games literature. *Curriculum Journal*, 16(4), 455-474.
- Pavlas, D., Heyne, K., Bedwell, W., Lazzara, E., & Salas, E. (2010). Game-based learning: The impact of flow state and videogame self-efficacy. *Proceedings of the Human Factors and Ergonomic Society 54th Annual Meeting*. doi:10.1177/154193121005402808
- Perry, G. T., & Eichler, M. L. (2015). Science Learning Games for Mobile Platforms. In *Encyclopedia of Mobile Phone Behavior* (pp. 562-574). IGI Global. doi:10.4018/978-1-4666-8239-9.ch047
- Perry, G. T., Kulpa, C. C., Pinheiro, E., & Eichler, M. L. (2012). Lessons from an Educational Game Usability Evaluation. *International Journal of Interactive Mobile Technologies*, 6(2).
- Perry, G. T., & Schnaid, F. (2012). A case study on the design of learning interfaces. *Computers & Education*, 59(2), 722–731. doi:10.1016/j.compedu.2012.03.009
- Quigley, K. S., Lindquist, K. A., & Barrett, L. F. (2014). Inducing and Measuring Emotion and Affect. In H. T. Reis & C. M. Judd (Eds.), *Handbook of Research Methods in Social and Personality Psychology* (pp. 220–252). Cambridge University Press.

- Rapini, S. (2012). *Beyond textbooks and lectures: Digital game-based learning in STEM subjects*. McLean, VA: Center for Excellence in Education.
- Reeve, E. M. (2015). STEM Thinking! *Technology and Engineering Teacher*, 75(4), 8–16.
- Ritz, J. M., & Fan, S.-C. (2015). STEM and Technology Education: International State-of-the-Art. *International Journal of Technology and Design Education*, 25(4), 429–451. doi:10.1007/s10798-014-9290-z
- Sawaya, S. F., & Putnam, R. T. (2015). Bridging the Gap: Using Mobile Devices to Connect Mathematics to Out-of-School Contexts. In H. Crompton & J. Traxler (Eds.), *Mobile Learning and Mathematics. Foundations, Design, and Case Studies* (pp. 9–19). New York: Routledge.
- Schell, J. (2008). *The Art of Game Design*. Burlington: Elsevier.
- Sharek, D., & Wiebe, E. (2015). Investigating real-time predictors of engagement: Implications for adaptive videogames and online training. *International Journal of Gaming and Computer-Mediated Simulations*, 7(1), 20–37. doi:10.4018/IJGCMS.2015010102
- Shute, V. J., Wang, L., Greiff, S., Zhao, W., & Moore, G. (2014). Gamification in education. *Computers in Human Behavior*, 63, 106–117. doi:10.1016/j.chb.2016.05.047
- Taylor, J. (2006). What are appropriate methods for evaluating learning in mobile environments? Evaluating Mobile Learning. In M. Sharples (Ed.), *Big Issues in Mobile Learning*. Nottingham, UK: Kaleidoscope Network of Excellence, Mobile Learning Initiative.
- Traxler, J. (2009). The Evolution of Mobile Learning. In G. Retta (Ed.), *The Evolution of Mobile Teaching and Learning*. Santa Rosa, CA: Informing Science Press.
- Traxler, J., & Crompton, H. (2015). Mobile Maths: Figuring Out What Comes Next. In H. Crompton & J. Traxler (Eds.), *Mobile Learning and Mathematics. Foundations, Design, and Case Studies* (pp. 223–232). New York: Routledge.
- Wijers, M., Jonker, V., & Drijvers, P. (2010). MobileMath: Exploring mathematics outside the classroom. *ZDM Mathematics Education*, 42(7), 789–799. doi:10.1007/s11858-010-0276-3
- Wijers, M., Jonker, V., & Kersten, K. (2008). MobileMath: The phone, the game and the math. In T. Connolly, & M. Stansfield (Eds.), *Proceedings of the European Conference on Game Based Learning*, (pp. 507–516). Reading: Academic Publishing Limited.
- Wood, L. C., Reiners, T., & Bastiaens, T. J. (2014). Design perspective on the role of advanced bots for self-guided learning. *International Journal of Technology Knowledge in Society*, 9(4), 187–199.
- Wu, W. H., Jim Wu, Y.-C., Chen, C.-Y., Kao, H.-Y., Lin, C.-H., & Huang, S.-H. (2012). Review of trends from mobile learning studies: A meta-analysis. *Computers & Education*, 59(2), 817–827. doi:10.1016/j.compedu.2012.03.016

ADDITIONAL READING

- Boehner, K., DePaula, R., Dourish, P., & Sengers, P. (2007). How emotion is made and measured. *International Journal of Human-Computer Studies*, 65(4), 275–291. doi:10.1016/j.ijhcs.2006.11.016
- Crompton, H., & Traxler, J. (2015). *Mobile Learning and Mathematics. Foundations, Design, and Case Studies*. New York, London: Routledge.
- Crompton, H., & Traxler, J. (2015). *Mobile Learning and STEM: Case Studies in Practice*. New York, London: Routledge.

Frohberg, D., Göth, C., & Schwabe, G. (2009). Mobile learning projects – a critical analysis of the state of the art. *Journal of Computer Assisted Learning*, 25(4), 307–331. doi:10.1111/j.1365-2729.2009.00315.x

Hunick, R., Leblanc, M., & Zubek, R. (2004). MDA: A formal approach to game design and game research. In: Proceedings of the Challenges. In: *Proceedings of the Challenges in Games AI Workshop, 19th National Conference on Artificial Intelligence*, San Jose (USA).

Li, M.-C., & Tsai, C.-C. (2013). Game-based learning in science education: A review of relevant research. *Journal of Science Education and Technology*, 22(6), 877–898. doi:10.1007/s10956-013-9436-x

Linehan, C., Kirman, B., Lawson, S., & Chan, G. G. (2011). Practical, Appropriate, Empirically-Validated Guidelines for Designing Educational Games. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1979-1988). New York: ACM. doi:10.1145/1978942.1979229

Quigley, K. S., Lindquist, K. A., & Barrett, L. F. (2014). Inducing and Measuring Emotion and Affect. In H. T. Reis & C. M. Judd (Eds.), *Handbook of Research Methods in Social and Personality Psychology* (pp. 220–252). Cambridge: University Press.

KEY TERMS AND DEFINITIONS

Game Design: Refers to the art of applying design and aesthetics to create a game to facilitate interaction between players (our learners) for either entertainment or educational purposes. This is not a scientific area of research, so the reader will find most of the information in books and on web portals such as GamaSutra. The information can be Design or Technology oriented (it will be easy to distinguish, even for the uninitiated reader).

Games and Affection: Refers to the study of measuring affective states while playing. The greatest challenge is to distinguish between states, to design instruments that can capture such states and to create experimental settings in which the player is affected solely by the game. Very tricky an interdisciplinary scientific area.

Games and Learning: Refers to a field of education research that studies what is learned by playing games, and how design principles, data and communities of game play can be used to develop new learning environments. An interdisciplinary scientific area which has grown since the early 2000s.

Mobile Learning: Refers to a way of learning across multiple contexts, through social and content interactions, using personal electronic devices. An interdisciplinary scientific area which has grown since the early 2000s.

STEM Education: Refers to the academic disciplines of science, technology, engineering and mathematics, the acronym is typically used when addressing education policy and curriculum choices in schools to improve science and technology development. A very strong and established scientific area since 1950.

Mobile Learning in and out of the K–12 Classroom

Pena L. Bedesem

Kent State University, USA

Tracy Arner

Kent State University, USA

INTRODUCTION

Today there is widespread ownership of mobile technology, including mobile phones, tablets, chromebooks among school-aged youth. To wit, researchers report that nearly 70 percent of 8- to 18-year-olds own their own mobile device. In parallel with increased accessibility and usage, researchers in various fields of education have begun to explore how mobile technology can impact teaching and learning. In this chapter, we discuss the work of pioneering scholars in the area of emerging technology in K-12 education. Next, we describe the current scientific knowledge on the ways in which mobile technologies are used by students in K-12 education. We then review important research that highlights the added effects of mobile technology to support non-traditional learners. Recommendations for future lines of investigation and further reading are provided.

BACKGROUND

The processes involved in student learning are now as never before. Delivery of information is more various, instantaneous, and itinerant and thus new ways of information delivery and instructional models are necessary to meet the needs of 21st Century learners. One method of meeting these needs is by using *mobile technologies*. Mobile technologies are digital devices that include tablets, personal digital devices, Chromebooks, and

mobile phones. The use of these technologies in the classroom is commonly referred to as *mobile learning*. This chapter will focus on the use of mobile technology to engage students, enhance instruction, and support non-traditional students including those with disabilities and students who are non-native English speakers.

In today's society mobile devices have become a ubiquitous technology. For example, 88% of adults (Zickuhr, 2011) and 77% of teens (Lenhart, 2012) own a mobile phone. The most recent Speak Up Survey by Project Tomorrow found that over 80% of secondary students have their own device to use for learning. Further, more and more schools are purchasing tablets, Chromebooks, and mobile media players to augment classroom instruction (Herold, 2016). In spite of the increasing availability of mobile technology in learning environments, deliberate integration into daily instruction, by teachers, is not as prevalent. Teachers cite numerous concerns with the use of technology including readiness, lack of training, classroom disruption, cheating, and access to inappropriate content as barriers to consistent integration into instruction (Kim, Kim, Lee, Spector, & DeMeester, 2012; O'Bannon & Thomas, 2014; Thomas & O'Bannon, 2013; Thomas, O'Bannon, & Bolton, 2013). Despite these barriers, the constantly increasing functionality and decreasing cost of mobile devices is making them both accessible and valuable learning tools both in and out of the classroom (Thomas, O'Bannon, & Bolton, 2013). Moreover,

DOI: 10.4018/978-1-5225-2255-3.ch555

researchers have shown increases in student motivation (Brown, 2008), and active participation (Kinsella, 2009) when technology is used in the classroom. Given the promise of mobile devices as instructional tools in K-12 education, it is important to discuss (a) pioneering work that has led to the use of mobile technologies, (b) stakeholder perceptions of using technology in the classroom, (c) how mobile technologies are being used, and (d) future research directions.

PIONEERING SCHOLARS

Marc Prensky, Mark van 't Hooft, and Liz Kolb were among the first educational researchers to advocate for the use of technology to meet the needs of the new learner in K-12 classrooms. While these researchers focus on different aspects of technology integration, collectively they account for important conceptual and empirical contributions that led to the acceptance of mobile devices as instructional tools. Instructional uses may include Internet research, communication, note taking, reading, organization, video-based instruction and more (Sansosti & Bedesem, 2015).

Marc Prensky is credited for coining the terms *digital natives* and *digital immigrants*. According to Prensky (2001), digital natives are of the generation born after 1980, whereas digital immigrants are of the generation born before 1980. Digital immigrants have to learn to adapt their environment and work at integrating technology into their daily lives. On the other hand, digital natives are characterized as immersed in technology including computers, videogames, digital music players, and mobile phones, and can seamlessly integrate technology into every aspect of their lives. Prensky posits that as a result of being exposed to a technologically rich environment and the sheer volume of students' interaction with it, today's students think and process information fundamentally different from digital immigrants. Thus, to stay engaged during learning activities

digital natives have come to expect the same level of technology integration in the classroom as in their personal lives (Martin, & Ertzberger, 2013).

Mark van 't Hooft has been active in the area of mobile learning since 2001, beginning with the Palm Education Pioneer (PEP) Project - the first large-scale evaluation of handheld mobile technology for education. Through the PEP Project, 175 K-12 classrooms throughout the United States were provided sets of handheld computers. An overwhelming majority (90%) of teachers who participated in the PEP Project indicated handheld computers are effective instructional tools that have the potential to have a positive impact on students' learning, and that they would continue to use handhelds in the future. In a later study, Swan, van 't Hooft, and Kratcoski (2005) examined students' use of mobile technologies and the effects on their motivation to learn and engagement in learning activities, as well as their teachers' support for the learning process. Their findings suggest that use of the mobile technologies resulted in increased student motivation, engagement, productivity, and quality of work. Teacher interviews also indicated using mobile technologies enhanced the learning process, especially in the area of writing. Moreover, several teachers noted improvements in the writing of students with disabilities and commented on the benefit of mobile technologies for lessening the performance gap between students with and without disabilities.

Liz Kolb's work explores methods teachers and students can use to integrate "everyday technologies" to which students have become accustomed, into the classroom. Specifically, Kolb investigates the integration of mobile phones into classroom instruction as a way to better connect how students use technology in everyday communication and the way they use technology in the classroom (Kolb, 2008). Mobile phones can be used, for example, to increase engagement, enhance instruction, and extend learning beyond the typical school day. In this area, Kolb has published two

books on how teachers can use mobile phones in the classroom for authentic purposes. Her first book, *Toys to Tools: Using Student Cell Phones in Education* (2008), explored ideas for authentic lessons, solutions for common concerns, and research to support the use of mobile phones in education. The purpose of her second book, *Cell Phones in the Classroom: A Practical Guide for Educators* (2011), is to “illustrate many practical ways to use cell phones in learning: classroom groups, individual projects, homework outside of the classroom, field trips, and as communication tools for teachers, students, and parents” (p. 1). In this book, Kolb presents 16 case studies to demonstrate how teachers use mobile phones in education and illuminates their experiences and activities. In addition to her books, Kolb also maintains a website (www.cellphonesinlearning.blogspot.com) with a host of resources including lesson plans that use student mobile phones and information about mobile safety education.

The work of Prensky, van ’t Hooft, and Kolb has significantly influenced research in the area of technology and its application in educational settings. Together, they have identified the new learner, examined the use of emerging technologies as instructional tools, and provided practical classroom applications for educators. They have also encouraged schools to keep up with the advances in technology and adopt programs such as Bring Your Own Device (BYOD) to reduce the cost of providing and maintaining equipment. Additionally, their work has led to research in how educators can use mobile phones in K-12 education to support all students, including those with disabilities.

CURRENT SCIENTIFIC KNOWLEDGE IN K-12 Education

In the following sections we discuss the perceptions of the use of technology in education, including the perceived benefits of and barriers to their

use. Next, we highlight how mobile technology is currently being used in the classroom. Finally, we make recommendations for future research and further reading.

Stakeholder Perceptions of Technology Integration

The increase in availability of mobile technology in the hands of students has led to an increase in classroom integration. The degree and method of integration vary considerably across teacher experience levels. Digital Natives, as defined by Prensky (2001), are now becoming pre-service teachers (Lei, 2009). In spite of the normality of technology in their daily lives, these pre-service teachers may not be prepared to be digital teachers. They exhibited strong positive feelings about technology in education but were not confident in their ability to use more complex technology in the classroom (Lei, 2009). In 2014, O’Bannon and Thomas focused on age as it relates to support for the use of mobile phones in the classroom. They found that teachers who were in age ranges 33-49 and teachers 32 and under were significantly more supportive of mobile phones in the classroom than teachers 50 and older. The depth and frequency of technology integration is also tied to teachers’ Technological Pedagogical and Content Knowledge (TPACK) and beliefs in the integration of technology (Kim et al., 2012).

Perceived Benefits of Mobile Technology

Researchers have also examined the perceived benefits of using mobile technology in the classroom. Teachers (pre-service and in-service), parents, and school administrators believe that some benefits of using technology in classrooms are increased student engagement and motivation as well as developing proficiency in 21st Century Skills such as communication and collaboration (Project Tomorrow, 2015; Thomas & O’Bannon,

2013; Thomas, O'Bannon, & Bolton, 2013). Three quarters of the 431,000 students surveyed by Project Tomorrow (2015) believe that using a tablet, phone, laptop or Chromebook is essential to their educational experience. Students indicate significant educational use both inside and outside the classroom walls with activities like emailing teachers with questions, watching instructional videos, and contributing to class blogs (Project Tomorrow, 2015). Teachers and administrators described an increase in the use of blended learning environments, accessed on a variety of mobile devices, as beneficial to student learning outcomes (Project Tomorrow, 2015). Pre-service teachers also have strong, positive beliefs regarding classroom mobile technology use citing an increase in access to information, extending learning beyond the classroom walls, increased student to teacher communication (Lei, 2009; Thomas & O'Bannon, 2013).

Perceived Barriers to Mobile Technology

Pre-service and in-service teachers have identified a variety of possible barriers to substantial integration of mobile technology. Kim et al., (2012) separate barriers into two categories; first order and second order. First order barriers consist of issues such as computer access, Internet, or teachers' technological pedagogical content knowledge. Second order barriers are related to teachers conceptual and epistemological beliefs. Classroom integration of technology is highly correlated with teachers epistemological beliefs specifically in lesson design, implementation and class discussion (Kim et al., 2012).

Teachers across studies identified classroom disruption, cheating, and access to inappropriate content as barriers to the use of technology in the classroom (O'Bannon & Thomas, 2014; Thomas & O'Bannon, 2013; Thomas, O'Bannon, & Bolton, 2013). In-service teachers also identified cyberbullying, sexting, and the negative impact of texting on student writing as additional

barriers (O'Bannon & Thomas, 2014; Thomas, O'Bannon, & Bolton, 2013). O'Bannon and Thomas (2014) also found that teachers over 50 years old perceived each barrier as being more problematic than teachers 33-49 and teachers 32 and under. According to Thomas and O'Bannon (2013), pre-service teachers were less concerned about access to inappropriate content, sexting, or cyberbullying than they were about classroom disruption and cheating. Lei (2009) found that in spite of the widespread access to and use of technology in the daily lives of pre-service teachers, they were apprehensive about how to successfully integrate mobile technology into instruction. Ally and Prieto-Blazquez (2014) suggest that teacher training programs be redesigned to prepare teachers to enter the technology-enhanced classroom. Teachers also indicated concern about future students becoming too dependent on technology and believe that some content is better taught without technology (Lei, 2009).

How Mobile Devices are Used in K-12 Education

A review of the literature revealed little empirical research on the use of mobile devices in K-12 education. Much of the research on the use of technology in education focuses on higher education and professional education settings. Institutions of higher education are implementing mobile technology to increase access and flexibility for students (Ally & Prieto-Blazquez, 2014). For example, researchers have examined the use of text messaging to support students' transition to the university setting (Harley, Winn, Pemberton, & Wilcox, 2007), the use of mobile phones as audience response systems (Tremblay, 2010), anonymous peer feedback (Rotsaert, Raes & Schellens, 2015) and perceptions of faculty and graduate students regarding mobile phone use in college classes (Burns & Lohenry, 2010).

Empirical research in K-12 education is gradually increasing as access to mobile devices and Internet connectivity become more common.

Current research shows that mobile technology holds great promise for educational use. One instructional use that has an increasing body of research is the adoption of simulations to provide authentic learning environments for students in science classrooms. A simulation is a computerized reproduction of a process or model of a system in which the user can manipulate inputs or steps in the process to analyze the varying outcomes (Rutten, van Joolingen, & van der Veen, 2012). Simulations may be used to enhance or replace traditional instruction or to enhance or replace laboratory experiments. In a review of literature on simulation use in the science classroom, Rutten et al., (2012) found that nearly all studies provided evidence of increased student comprehension and conceptual understanding.

Thomas and Orthober (2011) examined the use of text messaging in secondary classrooms. In their study, high school students in three different classes received school-related text messages from their teachers once daily. One teacher sent only reminders about assignments and the other teacher used text messages to communicate on a variety of school-related topics. The authors reported that an overwhelming majority of students (92%) found school-related text messages from their teachers helpful and 11% of the students used text messaging to ask their teacher questions and for help on homework. The instructors also expressed that they thought the text messages were helpful for students and noted that the students were better prepared for class. Results of this study demonstrate that mobile phones and text messaging have the potential to positively impact student learning and engagement.

Researchers have also begun to look at the use of mobile technology to support English Language Learners in elementary and middle school (Liu, Navarrete & Wivagg, 2014). Students and teachers were provided an iPod Touch to support teaching and learning over the course of one school year. Student devices were setup with applications and multimedia to support language acquisition such as creation of vocabulary study guides with the

Storykit application. Teachers reported that they were better able to differentiate instruction for the various levels of language acquisition. Students indicated the benefit of increased practice time and extended learning opportunities due to the 24/7 access to the device as well as increasing family engagement in the educational process (Liu et al., 2014).

Additionally, researchers have begun to examine how mobile phones can be used as *assistive technology* to support students with disabilities. Assistive technology is defined as any item, piece of equipment, or product system that is used to increase, maintain, or improve functional capabilities of individuals with disabilities (Individuals With Disabilities Education Improvement Act, 2004, I.A.602.1). Although the language used here to define assistive technology is broad, the majority of assistive technology in schools is typically packaged technology programs such as text-to-speech programs, word-prediction programs, and screen magnifiers (Edyburn, 2004). These packaged programs assist students with disabilities by compensating for a student's disability; however, they are costly, stigmatizing, and disconnected from the student's use of everyday technology (Bouck, Shurr, Tom, Jasper, Bassette, Miller, & Flanagan, 2012).

As mobile technology becomes more ubiquitous, manufacturers are increasing the number of assistive technologies that are native on mobile devices. Apple has incorporated intuitive tools to assist with vision impairment, hearing impairment, motor skills deficits, and learning and literacy impairments. Native accessibility tools include Dictation, Voice Command, VoiceOver, Speak Screen, Zoom, and Siri, a personal digital assistant that will respond to voice commands from the user (Apple, 2016). Google has also provided increased access for mobile technology users with tools like "Ok, Google" which will perform Internet searches via voice request from any mobile device and perform over 70 additional voice requests on devices with the Android operating system (Knoll, 2016).

Augmented Reality (AR) technology allows users to combine real and virtual objects through overlays of text, images, video, 3D models, or sounds (Bower, Howe, McCreddie, Robinson, & Grover, 2014). Educational applications of AR are emerging which allow educators to incorporate marker-based learning experiences around a specific content area such as space, microbiology, and language learning. Additionally, technologies are now available that allow teachers to create their own Augmented Reality, marker-based learning units. Students access the content using an app on their mobile device (Bower et al., 2014). Teachers are able to add game-based elements, digital storytelling, and physical activity to instructional units through the use of AR and GPS coordinates (Dunleavy, Dede & Mitchell, 2009). Students report an increase in motivation and improved attitudes about learning. Some research indicates improved learning outcomes as well (Bower et al., 2014).

Escobedo et al. (2012) used mobile phones equipped with an AR application to facilitate social skills practice in real life situations with students with autism. The authors developed an application that guided students with autism through six fundamental social skills of the Social Compass Curriculum (e.g., make eye contact, maintain appropriate personal boundaries, reply to conversation initiators, share interests with partners, disengage appropriately at the end of an interaction, identify potential communication partners). Escobedo and colleagues reported that the use of the mobile phones with the application positively impacted social interactions and increased the number of social interaction by 56.73%. Additionally, the authors noted that students learned how to use the mobile phone and application with minimal assistance, were excited to use it, and found it helpful.

Bedesem (2012) examined the use of mobile phones to change and support behavior in educational settings. Bedesem taught middle school students with mild to moderate disabilities (e.g., ADHD, LD) how to self-monitor their on-task behavior in an inclusive general education classroom.

For her study, Bedesem developed an intervention, called CellF-Monitoring that combined mobile phones with social networking to aid students in staying on task during instructional time. The author reported that the procedure produced positive changes in the on-task behavior of student participants. Additional findings indicate that the mobile phones did not seem to distract the students or their peers and that the teachers and students expressed an overall satisfaction with the CellF-Monitoring procedure.

FUTURE RESEARCH DIRECTIONS

Technology use in K-12 education is still an emerging literature base. There are articles, blogs, books, and websites dedicated to how educators can incorporate technology into instruction, assessment, feedback, and communication but empirical research on the subject is scarce. More research is needed on how the use of mobile technologies affect student engagement, motivation, differentiation and learning outcomes. Additionally, further exploration is needed on how schools can minimize barriers to the integration of technology as a regularly used tool for learning. Student access to technology has increased greatly through personally owned devices as well as technology provided by schools. Access to technology for learning is not available for all students and varies greatly between, public and private institutions as well as those separated demographically and geographically.

With respect to students with disabilities, technology holds great promise for increasing independence, learning and practicing social skills, and changing and supporting behavior. More evidence, however, is needed on the prevalence of mobile use among students with disabilities for instructional use as well as improving quality of life. Studies that report ownership and usage of technology among students do not report rates specific to students with disabilities. Prevalence rates specific to students with disabilities would

determine if technology ownership among students with disabilities is the same as their non-disabled peers. This is important because the rationale of using technology in the classroom that is often cited in the literature is the ubiquitous nature of the devices. However, if fewer students with disabilities have access to technology, compared to their non-disabled peers, then the current progress toward increasing technology integration in the classroom is widening the digital divide between students with and without disabilities.

CONCLUSION

Additional research is needed in the area of preparing pre-service teachers to enter the classroom as digital natives. In spite of frequent personal technology use, teacher candidates are not adequately prepared for seamless integration of technology when they enter the classroom. Pre-service teachers are comfortable with digital consumption but become less confident when working with more sophisticated technology for instructional use. Pre-service teacher programs need to include technology training integrated with the curriculum, pedagogy, and intervention training for the training to be effective. Technology only courses may not provide concrete instances of instructional use that pre-service teachers can adopt upon entering the classroom (Sansosti & Bedesem, 2015).

REFERENCES

- Accessibility - iOS - Apple. (n.d.). Retrieved June 27, 2016, from <http://www.apple.com/accessibility/ios/#hearing>
- Ally, M., & Prieto-Blazquez, J. (2014). What is the future of mobile learning in education? Mobile Learning Applications in Higher Education. *Revista de Universidad y Sociedad del Conocimiento (RUSC)*, 11(1), 142-151. <http://de.doi.org/10.7238/rusc.v11i1.2033>
- Bedesem, P. L. (2012). Using cell phone technology for self-monitoring procedures in inclusive settings. *Journal of Special Education Technology*, 27(4), 33-46. doi:10.1177/016264341202700403
- Bouck, E. C., Shurr, J. C., Tom, K., Jasper, A. D., Bassette, L., Miller, B., & Flanagan, S. M. (2012). Fit it with TAPE: Repurposing technology to be assistive technology for students with high-incidence disabilities. *Preventing School Failure*, 56(2), 121-128. doi:10.1080/1045988X.2011.603396
- Bower, M., Howe, C., Mccredie, N., Robinson, A., & Grover, D. (2014). Augmented reality in Education — Cases, places, and potentials. 2013 IEEE 63rd Annual Conference International Council for Education Media (ICEM). doi:10.1109/cicem.2013.6820176
- Digital Learning 24/7: Understanding technology-enhanced learning in the lives of today's students [Scholarly project]. (2015, April 1). Retrieved June 20, 2016, from <http://www.tomorrow.org/speakup/pdfs/SU14StudentReport.pdf>
- Dunleavy, M., Dede, C., & Mitchell, R. (2009). Affordances and limitations of immersive participatory augmented reality simulations for teaching and learning. *Journal of Science Education and Technology*, 18(1), 7-22. doi:10.1007/s10956-008-9119-1
- Edyburn, D. L. (2004). Rethinking assistive technology. *Special Education Technology Practice*, 5(4), 16-23.

- Escobedo, L., Nguyen, D. H., Boyd, L., Hirano, S., Rangel, A., Garcia-Rosas, D., & Hayes, G. et al. (2012, May). MOSOCO: a mobile assistive tool to support children with autism practicing social skills in real-life situations. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems* (pp. 2589-2598). ACM. doi:10.1145/2207676.2208649
- Harley, D., Winn, S., Pemberton, S., & Wilcox, P. (2007). Using *texting* to support students transition to university. *Innovations in Education and Teaching International*, 44(3), 229–241. doi:10.1080/14703290701486506
- Herold, B. (2016, February 5). *Technology in Education: An Overview*. Retrieved July 27, 2016, from <http://www.edweek.org/ew/issues/technology-in-education/>
- Individuals With Disabilities Education Improvement Act of 2004. 118 Stat. 2647
- Kim, C., Kim, M. K., Lee, C., Spector, J. M., & Demeester, K. (2013). Teacher beliefs and technology integration. *Teaching and Teacher Education*, 29, 76–85. doi:10.1016/j.tate.2012.08.005
- Kinsella, S. (2009). Many to one: Using the mobile phone to interact with large classes. *British Journal of Educational Technology*, 35(5), 956–958. doi:10.1111/j.1467-8535.2008.00888.x
- Knoll, M. (2016, April 24). *Complete list of 70 awesome 'Okay Google' voice commands*. Retrieved June 27, 2016, from <http://trendblog.net/list-of-google-now-voice-commands-infographic/>
- Kolb, L. (2008). *Toys to tools: Connecting student cell phones to education*. Eugene, OR: International Society for Technology in Education.
- Kolb, L. (2011). Adventures with cell phones. *Educational Leadership*, 68(5), 39–43.
- Kolb, L. (2011). *Cell phones in the classroom: A practical guide for educators*. Eugene, OR: International Society for Technology in Education.
- Lei, J. (2009). Digital natives as preservice teachers: What technology preparation is needed? *Journal of Computing in Teacher Education*, 25(3), 87–97.
- Lenhart, A. (2012). *Teens, smartphones and texting*. Pew Internet & American Life Project. Retrieved from <http://pewinternet.org/Reports/2012/Smartphone-Update-2012/Findings.aspx>
- Liu, M., Navarrete, C. C., & Wivagg, J. (2014). Potentials of Mobile Technology for K-12 Education: An Investigation of iPod touch Use for English Language Learners in the United States. *Journal of Educational Technology & Society*, 17(2), 115–126.
- Martin, F., & Ertzberger, J. (2013). Here and now mobile learning: An experimental study on the use mobile technology. *Computers & Education*, 68, 76–85. doi:10.1016/j.compedu.2013.04.021
- Nielsen, L., & Webb, W. (2011). *Teaching generation text: Using cell phones to enhance learning*. John Wiley & Sons.
- OBannon, B. W., & Thomas, K. (2014). Teacher perceptions of using mobile phones in the classroom: Age matters! *Computers & Education*, 74, 15–25. doi:10.1016/j.compedu.2014.01.006
- Prensky, M. (2001). Digital natives, digital immigrants. *On the Horizon*, 9(5), 1–6. doi:10.1108/10748120110424816
- Rideout, V. J., Foehr, U. G., & Roberts, D. F. (2010). *Generation M²: Media in the Lives of 8-to 18-Year-Olds*. Menlo Park, CA: Henry J. Kaiser Family Foundation.

Roblyer, M. D., & Doering, A. H. (2010). *Integrating educational technology into teaching* (5th ed.). Boston, MA: Allyn & Bacon/Pearson.

Rotsaert, T., Raes, A., & Schellens, T. (2015). *Anonymous peer assessment through mobile response technology in higher education: effects of interpersonal variables and students' preferred type of teacher guidance*. Presented at the American Educational Research Association (AERA 2015).

Rutten, N., van Joolingen, W. R., & van der Veen, J. T. (2012). The learning effects of computer simulations in science education. *Computers & Education*, 58(1), 136–153. doi:10.1016/j.compedu.2011.07.017

Sansosti, F. J., & Bedesem, P. L. (2015). The Use of Mobile Technologies for Students At-Risk or Identified with Behavioral Disorders within School-Based Contexts. *Recent Advances in Assistive Technologies to Support Children with Developmental Disorders*, 114.

Swan, K., van t Hooft, M., Kratcoski, A., & Unger, D. (2005). Uses and effects of mobile computing devices in K-8 classrooms. *Journal of Research on Technology in Education*, 38(1), 99–112. doi:10.1080/15391523.2005.10782451

Thomas, K., & Orthober, C. (2011). Using text-messaging in the secondary classroom. *American Secondary Education*, 39(2), 55–76.

Thomas, K. M., OBannon, B. W., & Bolton, N. (2013). Cell phones in the classroom: Teachers perspectives of inclusion, benefits, and barriers. *Computers in the Schools*, 30(4), 295–308. doi:10.1080/07380569.2013.844637

Tremblay, E. A. (2010). Educating the mobile generation – using personal cell phones as audience response systems in post-secondary science teaching. *Journal of Computers in Mathematics and Science Teaching*, 29(2), 217–227.

van 't Hooft, M., & Swan, K. (Eds.). (2006). *Ubiquitous computing in education: Invisible technology, visible impact*. Routledge.

Zickuhr, K. (2011). *Generations and gadgets*. Pew Internet & American Life Project. Retrieved from <http://pewinternet.org/Reports/2011/Generations-and-gadgets.aspx>

ADDITIONAL READING

Education Week Digital Directions. Trends and Advice for K-12 Technology Leaders. (n.d.). Retrieved July 27, 2016, from <http://www.edweek.org/dd/?intc=main-topnav>

Kolb, L. (2011). *Cell phones in the classroom: A practical guide for educators*. Eugene, OR: International Society for Technology in Education.

Kolb, L. (2013). *Help your child learn with cell phones and web 2.0*. Eugene, OR: International Society for Technology in Education.

Mobile Learning. 21st Century Tools for 21st Century Schools (n.d.). In Special Interest Group Mobile Learning Wiki: <http://sigml.iste.wikispaces.net/Main>

Pearson, P. D., Ferdig, R. E., Blomeyer, R. L. Jr, & Moran, J. (2005). *The Effects of Technology on Reading Performance in the Middle-School Grades: A Meta-Analysis With Recommendations for Policy*. Learning Point Associates/North Central Regional Educational Laboratory. NCREL.

Thomas, K. M., & McGee, C. D. (2012). The only thing we have to fear is...120 characters. *TechTrends*, 56(1), 19–33. doi:10.1007/s11528-011-0550-4

Walker, Z. (n.d.). The LAST BACKPACK GENERATION. Retrieved July 27, 2016, from <http://lastbackpack.com/>

KEY TERMS AND DEFINITIONS

Assistive Technology: Any type of equipment that can be used to increase, maintain, or improve functional capabilities of student with disabilities.

Augmented Reality: The blending of physical and digital worlds, often with a virtual overlay on a physical item.

Mobile Phones: A telephone that uses cellular towers so that it can be used anywhere.

Mobile Technology: Any portable electronic device that accepts, processes, and stores data at high speeds (e.g., smart phone, tablet computer).

Simulation: A technologically based reproduction of a real world event in which variables can be manipulated by the user.

Students with Disabilities: Students with disabilities are students who have been identified, assessed, and determined to be eligible for special education services under the Individuals with Disabilities Education Act.

M

A Psychological Perspective on Mobile Learning

Melody M. Terras

University of the West of Scotland, UK

Judith Ramsay

Manchester Metropolitan University, UK

INTRODUCTION

The mobile internet offers 24/7 access to a wide range of social, commercial and educational opportunities. If the educational opportunities are to be maximized then a full understanding of user skills, motivations, preferences and learning profile is required. We propose that this detailed understanding is best informed by adopting a psychological perspective on the mobile learning process. Therefore this article considers the insights that can be derived from viewing mobile learning through a psychological lens: in particular the psychological challenges that mobile learning presents and the importance of considering individual differences in learner skills and behaviours. We will also consider the insights that psychology offers in addressing these challenges, especially the importance of media literacy skills, the presentation of learning resources in a way that considers the limited cognitive resources of the learner and the importance of context and contextualized behavior. In so doing, we demonstrate the persuasive case for embedding psychologically driven insights in future multi-disciplinary research and practice.

BACKGROUND: APPROACHES TO MOBILE LEARNING

Rapid technological advancements are substantially influencing the educational landscape, of-

fering learners, providers and developers a range of new platforms for learning. One of the most exciting platforms is mobile learning. Mobile learning can be defined as the use of wirelessly connected devices to augment learning and teaching within and between different psychological, social, physical and temporal contexts. Mobile learning has recently been identified as one of the key facilitators of learning across the globe (UNESCO, 2013). If these learning opportunities are to be maximised then a detailed understanding of the determinants and influences on uptake, use and effectiveness is required. Just as educators have to understand how we learn in traditional contexts, we now have to understand how we learn using mobile devices. Psychology is well placed to offer detailed insight into the necessary skills, preferences and behaviours necessary for successful mobile learning to occur. In the following section we outline the different perspectives on mobile learning.

The evidence base for mobile learning is rapidly developing and reflects understanding of the nature, process and influences on mobile learning from a number of different, but often complementary, perspectives. One of the first and most influential approaches is that of Mike Sharples who advanced an early definition in his seminal paper on mobile lifelong learning: “a new genre of educational technology - personal (handheld or wearable) computer systems that support learning from any location throughout a lifetime” (Sharples, 2000, p.177). Sharples sub-

sequently developed guidelines for the design and delivery of mobile teaching and learning which emphasised how mobile learning is different from other forms of learning and how learners “artfully engage with their surroundings” (Sharples, Taylor and Vavoula, 2007, p.2), be observant of state of the art best practice and of the availability of personal technologies. This latter point was further emphasised by Vavoula and Sharples (2008) in their recognition of the distinguishing role of context for mobile learning, in particular for the need to evaluate learning in and across the contexts in which it occurs. More recently, mobile devices have been found to be effective in supporting enquiry-based learning both within and beyond formal scheduled learning contexts (Sharples et al, 2015). Roleplay-based games delivered via short messaging (SMS) have also proven instrumental in the development of the critical thinking skills of peer educators (Roy and Sharples, 2015).

The situated, contextual nature of mobile learning is also recognised by Norbert Pachler who in collaboration with Bachmair, and Cook (2010) takes an ecological stance on mobile learning by emphasising the importance of “understanding and knowing how to utilize our everyday life-worlds as learning spaces” (p. 6), thus highlighting the role of sociocultural influences upon learning behaviour. More recently, the importance of understanding the specific affordances of mobile learning and the role of mobile devices as “cultural transformational tools” (p. 938) has been emphasised in mobile learning design, in the quest to design for it appropriately (Bannan et al, 2016). The use of mobile technology and the learning opportunities it offers for young children within the context of the home has been examined by Marsh (2015) who investigated family practices concerning mobile technologies such as laptops, tablets, and mobile phones in and around the home and found that children typically use a mobile device such as a tablet alongside a parent, rather than alone. Research evidence indicates that parents play an important enabling role guiding and educating their children’s use of technology and may also

act as role models for its use. For example, parents’ actual use of mobile media (i.e. their behaviours), as opposed to their attitudes towards their child’s use is, in part, predictive of their child’s use (Lauricella et al. 2014; Terras, Yousaf and Ramsay, 2016). Such findings highlight the importance of taking an ecological, contextually-informed approach to the mobile technology use of learner of today but also of considering the learners of the future, as ecological contexts influence the development of skills, practices and attitudes.

A different perspective on mobile learning, which is more educational in emphasis, is offered by John Traxler who defined mobile learning as “the provision of education and training on PDAs/palmtops/handhelds, smartphones and mobile phones” (Traxler, 2009, p.2). Traxler’s focus is on formal learning and he observes that access to information when one is mobile is leading to a reconceptualization of what it means to be trained and what it means to be supported in doing one’s work. Traxler (2007) also makes the incisive point that “mobile” does not merely describe the manner in which one learns, rather it now heralds a reconceptualization of what it means to learn. The notion of a reconceptualization is a very important one for educators as it points to a fundamental shift in the nature of education that may, arguably, introduce new and as yet unidentified pedagogical and psychological challenges for learners. In the introduction to a recent special edition on mobile-assisted language learning, Traxler et al (2016) made the insightful observation that “Mobile learning might, in fact, not be the mobile aspect of learning so much as the educational aspect of mobility” (p. 1235). This insight lies at the heart of the growing appreciation that time, space, place and manner of activities (whether learning or other) need to be characterised and understood to ensure that they capture the rich and dynamic nature of human behaviour whilst mobile. Only by doing so will the educational potential be captured.

An influential approach on mobile learning is that of Agnes Kukulska-Hulme who emphasises

the evolving nature and use of mobile devices and user preferences: in particular the fact that mobile learning often occurs via devices that may not have been intended to support learning (e.g. smartphones) which reflects the natural and informal way in which mobile learning has developed (Kukulska-Hulme, 2007). Kukulska-Hulme and Jones (2011) have advocated the integration of an understanding of learner practices, learning contexts, and technologies within pedagogical approaches. Kukulska-Hulme (2012) also considers the question of learning being defined by the time and place within which it occurs and therefore argue for frameworks to organise these new ways of learning. Pedagogies of mobile learning should recognise and potentially capitalise on the dynamic, changing sociocultural context(s) of learning and the highly individual, situated and learner-initiated nature of the learning that takes place. They also need to be cognisant of the social communication and interpersonally constructed models of understanding and knowledge generation that characterise any Web 2.0 applications.

Whilst clearly reflective of the situated, ecological perspective taken by Pachler, the psychological nature of Kukulska-Hulme's arguments also complement well the more explicit psychological perspective taken by Terras and Ramsay, especially in the identification of contextual factors including the importance of the temporal dimension outlined below.

A Psychological Perspective on Mobile Learning

Terras and Ramsay (2012) have leveraged the potential of psychological concepts, theory and method to assist in mapping the challenges of the mobile learning landscape from the perspective of learner *behaviours* and the psychological infrastructure that underpins them. The five psychological challenges reflect well-established constraints that limit human cognitive functioning and behaviour: (1) memory is often context dependent, (2) information processing resources

are finite, (3) cognition is distributed in nature, (4) self-regulation and meta-cognition are essential for effective learning and (5) importance of recognising individual differences in learner skills and preferences.

This unique Psychological perspective not only differentiates their proposals from other researchers in the field but also offers a way to integrate research in the area. Terras and Ramsay argue that it is by conceptualising the issues raised by their contemporaries, in psychological form, that the field of mobile learning will be in a position to advance. The quest to understand the complex, embedded nature of mobile device use, along with an appreciation of the need to recognise both formal and informal styles of usage across changing contexts, offers a way to integrate the differing perspectives on mobile learning. The key issues raised by their peers in the field, such as transfer, personalisation, time and context (Kukulska-Hulme), the re-conceptualisation of what is meant by training (Traxler), the need for guidelines for mobile learning (Sharples and Vavoula), the need to evaluate mobile learning across contexts and understand how learners engage with their surroundings (Sharples), and the need to understand everyday worlds (Pachler) *can all be rendered in psychological form as challenges* to be addressed by the field. Terras and Ramsay argue that it is by laying bare the underlying, commonly-shared psychological mechanisms that underpin mobile learning across time and context that a sufficiently nuanced understanding of the nature of mobile learning as an emergent form of learning will be realised.

The Psychological Challenges of Mobile Learning

Just as educators have to understand how we learn in traditional contexts, it is now necessary to understand how we learn using mobile devices and Psychology is well placed to offer detailed insight into the necessary skills, preferences and behaviours necessary for successful mobile learn-

ing to occur. The following sections provide an overview of the challenges, the insights that a psychological approach can bring, and the solutions it offers.

As discussed above, Terras and Ramsay (2012) identified five major psychological challenges to mobile learning:

1. The challenge of context dependence: stimuli and cues that assist with the development and consolidation of learning exist within the very fabric of our environment. Perhaps most challenging of all, is the ever changing nature and intermittent presence of these contextual cues.
2. The challenge of finite information processing resources: the ability to hold information mentally so that it is ready to use – this is known as working memory capacity, and the ability to remember intentions i.e. future actions- prospective memory. Both extremely vulnerable to overload from contextual or environmental factors and this vulnerability may have a detrimental impact on mobile learning given its' situated nature.
3. The challenge of distributed cognition: cognition, namely, our mental models, plans and understandings of the world do not exist within the confines of our individual selves. Rather, for such models and plans to be of use, they are necessarily shared in various forms between individuals. Communication is at the heart of what it means to be human, and in communicating, we co-create commonly shared models of how the world is. These common understandings are not carbon copies, rather, they exist in versions within different individuals. Additionally, common understandings and representations of how the world is, are maintained not only across groups of individuals, but also across time. It is memory that allows us to do this, but it is not without its challenges. It has long been established that human memory is subject to very many influences. This, in

turn, introduces a further complex consideration, to the distributed nature of human cognition.

4. The challenge of self-regulation and meta-cognition. Self-regulation refers to the ability to be aware of one's own behavior, desires and impulses, and to manage them in a socially appropriate manner. This is done through a combination of introspection and evaluation of social norms, which requires an understanding of the needs and expectations of other people. Metacognition, namely thinking about thinking, is a higher-order form of cognitive control that allows the individual to manage their cognitive resources (their mental processes) to achieve better outcomes. Mobile learners need to become more self-aware and self-evaluative: they need to be sensitive to how their learning is affected by both internal factors such as their own cognitive skills and preferences and the influence of external factors such as the nature and content of formal and informal learning resources and the more general external context (Hu and Zhuang, 2011; Terras and Ramsay, 2012; 2014).
5. The challenge of accommodating learner differences. It has long been established in the field of Psychology that human beings are individually different. Whilst there exist patterns, likelihoods and tendencies that are associated with particular behaviours across large groups of people, it has nevertheless been established that learners are individually distinct. Indeed, learning whilst moving through individually-steered social, virtual and geographical contexts adds a significantly deeper complexity to the challenge of characterizing individual mobile learners.

The Spatio-Temporal Context of Mobile Learning

The changing spatio-temporal contexts of mobile access can present increased psychological de-

mands for learners. Therefore, Terras and Ramsay extended their conceptualisation of context to include a more detailed consideration of the temporal dimension of mobile learning and the temporal aspects of the five initial challenges (Terras and Ramsay, 2014a; 2014b). It is essential to explicitly consider the temporal aspects of mobile learning because (i) change in location occurs over time, and (ii) time may be experienced differently depending on the context. Given that time and locational change are often closely linked, consideration of the changing locations supported by mobile access must be situated within the wider temporal context. In order to do so, it is necessary to distinguish between the physical and subjective aspects of time during the mobile learning experience (Terras and Ramsay 2014a; Terras and Ramsay, 2014b). Physical or objective time, is how time is generally conceptualised: it is fixed, linear and is measured in standardised units such as seconds, hours and days. Psychological or subjective time is context dependent, non-linear and does not extend from the objective “past” to “present” (Zaky, 2012). The subjective experience of time is unique and is a product of cognitive, emotional or psycho-social influences (Droit-Volet and Gil, 2009). From a psychological perspective, consideration of the temporal dimension of mobile learning highlights the “importance of considering a learner’s experience of time, how contexts changes over time, the retrospective and prospective aspects of time, the importance of the sequencing of information over time and how information, knowledge and understanding are distributed across time” (Terras and Ramsay, 2014a, p. 114).

Mobile learners need to be aware of the spatio-temporal context. The concept of context-awareness is addressed in computer science as part of a system that keeps track of location (Dey, 2001); in a mobile context the best known example of a context-aware system is the mobile tourist guide. However, a more comprehensive and user-centred, rather than system centred perspective i.e. how mobile learners perceive context is required to inform good design and promote good mobile learning

skills. Five aspects of context have been identified: time, location, identity, activity and relationships to other entities and objects. Empirical data indicates that people are sensitive to the temporal aspects of contexts (Crane et al, 2011) but less is known about how this monitoring takes places, how it may vary across different contexts and how it may be influenced by the motivations and goals of the learner or the activities they are engaged in (Terras and Ramsay, 2014b).

Solutions and Recommendations: Psychological Insights and Solutions

Instructional Design

To date, psychological theory and research has offered immense insight into the design of effective e-learning resources by ensuring that instructional material is sensitive to the processing constraints of the human cognitive system thereby, avoiding cognitive overload. There are three types of cognitive load: intrinsic cognitive load, extraneous cognitive load, and germane cognitive load (Sweller, 1988). Intrinsic cognitive load reflects the processing necessary to undertake a task i.e. the processing load or task difficulty is inherent to the task. Germane load refers to the working memory resources that are required to deal with intrinsic cognitive load. Extraneous cognitive load is external to the task and can therefore be manipulated, ideally being kept to a minimum. Extraneous load is most often dependent on the format of the information presented e.g. describe a shape or show a picture of the shape. Therefore, instructional material should impose the lowest possible extraneous cognitive load in order to free our limited resources to process the content. Another influential load-based theory is the Cognitive Theory of Multimedia Learning which utilises a concept very similar to cognitive load, termed *Demands for Cognitive Processing* to highlight the dangers of overload (Mayer and Moreno (1988; 2003) by drawing upon a classic psychological theory concerning the nature and

structure of the cognitive system: Paivio's Dual Coding theory which posits dual information processing channels: one for visual and another for verbal information (Paivio, 1986). From a learning perspective, information can be acquired from both words and images, but our ability to process information in either visual or verbal format is limited, and if we become overloaded then we will fail to learn.

Cognitive overload is especially significant when designing mobile learning materials. Although designers can control the content of formal learning materials, the changing spatio-temporal contexts associated with mobile learning, place more control in the hands of the learner resulting in the possibility of overload, and this increased if they accessing informal learning materials which must be managed by the learner. More recently psychologically-based research has highlighted the importance of considering individual differences in a learner's cognitive profile, especially their working memory capacity. Problematic mobile device usage is associated with a lower ability to manage distractions, along with having poorer working memory capacity and weaker attentional control (Hadlington, 2015). Learners with low working memory capacity often achieve poorer outcomes than learners with high working memory capacity when they had to divide their attention across a mix of audio, text and images (Fenesi, Kramer & Kim, 2016).

Motivations for Use

Psychological theory and research has also provided considerable insight into the reasons and motivations for use of technology, both static and mobile through the Technology Acceptance Model (TAM) and its associated refinements. The Technology Acceptance Model (Davies, 1981) explains why people do, or in some instances do not, use technology by drawing upon well-established psychological constructs to understand and predict behaviour: The Theory of Reasoned Action and the Theory of Planned Behaviour (Fishbein and

Ajzen, 1975). The theory of Reasoned Action draws upon three psychological constructs: Behavioral Intentions, Attitude towards the behavior and the Subjective norm: behaviour is determined by the Intention to perform it, and this Intention is determined by a person's attitude toward the behaviour and the subjective norms of the behaviour. These proposals were further refined by the Theory of Planned Behavior (Ajzen, 1985) with the inclusion of the additional construct of Perceived Behavioral Control to extend the remit of the theory to include behavior under non-volitional behavior. The theory of planned behavior predicts that more positive Attitudes, supportive Subjective norms and high degrees of Perceived Control, the stronger the behavioral intention, therefore the behavior is more likely to occur. When applied to the uptake of mobile learning in higher education, 87.2% of the intention to use mobile learning was explained by the three key factors of Attitude, Subjective norm and Behavioral control, with Perceived behavioral control as the strongest predictor of the mobile learning uptake (Cheon, Lee, Crooks and Song, 2012). Using structural equation modelling Park et al (2012) examined the utility of a TAM based explanation of mobile learning that included the following components: independent (exogenous latent) variables were individual factors such as mobile learning self-efficacy, major subject studied, organisational factors (e.g. accessibility of the mobile learning system), and social factors reflecting the subjective norms. The outcome (endogenous) variables were perceived usefulness (PU), perceived ease of use (PE), mobile learning attitudes and behavioral intentions. The endogenous variables reflect three psychological domains: cognitive, affective and behavioral. The results indicate that behavioral intention and actual use is determined by four main groups of variables: individual context, system context, social context and organisational context. Clearly illustrating how the psychological constructs of the TAM model allows the prediction of direct and indirect influences on students' intentions to use mobile

learning (Park, Nam and Cha, 2012). Based on the TAM model Park et al make three practical recommendations: (1) Promote positive attitudes to mobile learning and emphasize its relevance as Attitudes we shown to have the largest direct effect on mobile learning behavior, social norms too were directly related to behavior. (2) Good wireless facilities and the provision of either devices or subsidies of mobile devices are essential as System Accessibility was another factor that directly influenced Behavioral Intention. (3) Provide support (e.g. mentoring) to enable students to use mobile learning. The model indicates that mobile learning self-efficacy affects both attitudes and perceived effort, with perceived effort influencing perceived usefulness which in turn affects attitudes. It is clear that psychologically driven theories offer a way not only to conceptualise and explore the challenges of mobile learning but also offers practical ways in which these challenges may be overcome.

Skills: How to Learn and Media Literacy

If mobile learning is to be effective, then learners must possess and apply the relevant skills: they must know how to learn using mobile technology and its associated learning resources. It should not be assumed that just because learners can use technology, especially social technology, that they will be effective mobile learners (Terras, Ramsay & Boyle, 2011). They may require explicit instruction on how to use technology to learn to help them overcome the specific psychological and practical challenges presented by mobile learning (Ramsay and Terras, 2015). The challenges presented by changing spatio-temporal context and the increased contextual awareness necessary have already been noted, and these challenges may be exacerbated by multitasking. In recent years, the increasing sophistication of wireless internet access means that technologies such as telephones and personal computers (PCs) have undergone a uni-function to multi-function

shift e.g. mobile telephones now not only support verbal communication, but also internet access and all the opportunities it affords such as email, video calling, social networking and the accessing of formal educational resources such as virtual learning environments (VLEs) and Massive Open Online Courses (MOOCs) as well as more informal user-generated resources typified by YouTube and Wikipedia. It is now possible to not only multitask across different devices, but also multitask on the same device. Hence learners may require increased sensitivity to their own multitasking ability, especially as people often overestimate their ability to multi-task (Sanbonmatsu et al. (2013), advice on how to allocate or inhibit attention across competing tasks, and instruction on how to use a mobile device which they generally use for social purposes to support learning. Frequently these situations may co-occur and they must be managed by the learner. For example, students frequently multitask on Facebook during self-directed study (Judd, 2014) and also in class which often result in poorer academic performance (Junco, 2012). Raising learners' awareness of their behaviour and its associated consequences may prove useful, as could instruction and use of active study and meta-cognitive strategies as well as 'technology breaks' (Rosen et al, 2013).

The ability to create, interpret and evaluate informational content is a key skill for effective learning. Various terms have been used to capture these skills, most notably media, digital or participatory literacy. The concept of media literacy is the most relevant, both now and in the foreseeable future to capture the skill set necessary for effective mobile learning as it is broad in focus (Ramsay & Terras, 2016). Media literacy is defined by the United Kingdom's telecommunications regulator as "the ability to use, understand and create media and communications in a variety of contexts" (Ofcom 2015, p. 19). Therefore, media literacy skills operate across a wide range of technology such as television, games, smartphones, internet and social media in a variety of work, social and educational settings all of which are relevant to

the mobile learner. Like print literacy, media literacy is generally viewed from two distinct perspectives: skill or social practice. An influential skills-based approach in higher education is that taken by the Joint Information Systems Committee (JISC) who identify a multi-component skill set under the umbrella term of Digital Literacy (JISC, 2014). The wide remit of the JISC approach highlights the complex nature of the skills required to benefit from technology-enhanced learning. In contrast, media literacy can be viewed as a social practice which emphasises how skills develop as individuals interact with their environment and how these interactions are dynamic and reciprocal (Livingstone, 2015). Considering skills in context encourages the consideration of contextual influences and individual differences. In order to fully understanding the skills necessary for effective mobile learning it may be more productive to adopt a blended approach that situates skills within their socio-ecological contexts (Ramsay & Terras, 2016).

FUTURE RESEARCH DIRECTIONS

Technological change and mobile learning provision is moving at a fast pace, so future research should be informed by technological advances, pedagogical theory and practice, psychological insights and to maximise potential of both research and practice, interdisciplinary working *to develop* an holistic understanding of the mobile learning experience.

One of most exciting advances with huge potential to enhance the mobile learning experience is “big data”: Data that can be gathered about learner behaviours including which type of information they access, how it is accessed, where, time spent on learning activities, the format or media that is used for different activities, achievement of learning outcomes, grades and similar information. This learning analytical data can be used to develop patterns of learner behaviour to provide insight

into learner behaviour and support formative reflection by learners and evaluations by providers.

Specifying the psychological challenges of mobile learning not only adds to our understanding of the process, but also helps identify how learners can be supported. The information discussed here highlights that learners may need to be taught how to learn using mobile technology and how to overcome both the practical and psychological challenges of learning in contexts that change across both space and time. The most effective intervention will be psychologically informed and support the psycho-social and cognitive factors that underpin necessary skills such as metacognition, self-regulation, self-efficacy and motivation as they allow students to govern their own learning.

CONCLUSION

The potential of mobile learning is rapidly being recognised and maximised as evidenced by the work of Sharples, Kukulska-Hulme, Terras and Ramsay, Traxler and Pachler and other colleagues across the globe. However, irrespective of the approach taken, it is essential that the psychological challenges that mobile learning presents are also recognised and addressed. Although many of these psychological challenges are attributable to limitations inherent in the human information processing system adopting a psychological perspective allows plausible and effective solutions to be developed. Cognitive load-based theories inform the design of learning materials and the Technology Acceptance Model helps identify factors that influence the uptake of mobile learning. Considering the experiential perspective of the mobile learner emphasises the importance of contextual awareness and the need to support the necessary skills to learn and monitor and control both external and internal contexts to ensure effective mobile learning. A detailed appreciation of individual learner skills and preferences will become increasingly important when mobile learn-

ing takes advantage of the opportunities afforded by Web 3.0 and the “semantic web”, where individualised profiles can be used to inform search results. Understanding the psychosocial profile of mobile learners will be essential for the provision of fluid and effective personalised learner experiences in the future. Psychologically-driven theory and research has contributed greatly to our understanding of the mobile learning experience to date and we expect it will continue to do so in the future.

REFERENCES

- Ajzen, I. (1985). From intentions to actions: A theory of planned behavior. In J. Kuhl & J. Beckman (Eds.), *Action-control: From cognition to behavior* (pp. 11–39). Heidelberg, Germany: Springer. doi:10.1007/978-3-642-69746-3_2
- Alrasheedi, M., & Capretz, L. F. (2013). An M-Learning Maturity Model for the Educational Sector. *The Sixth Conference of MIT's Learning International Networks Consortium (LINC) June 16th - 19th, 2013*. MIT.
- Bannan, B., Cook, J., & Pachler, N. (2016). Reconceptualizing design research in the age of mobile learning. *Interactive Learning Environments*, 24(5), 938–953. doi:10.1080/10494820.2015.1018911
- Bennett, S., Maton, K., & Kervin, L. (2008). The digital natives debate: A critical review of the evidence. *British Journal of Educational Technology*, 39(5), 775–786. doi:10.1111/j.1467-8535.2007.00793.x
- Bronfenbrenner, U., & Ceci, S. J. (1994). Nature-nurture reconceptualized in developmental perspective: A bioecological model. *Psychological Review*, 101(4), 568–586. doi:10.1037/0033-295X.101.4.568 PMID:7984707
- Cheon, J., Lee, S., Crooks, S. M., & Song, J. (2012). An investigation of mobile learning readiness in higher education based on the theory of planned behaviour. *Computers & Education*, 59(3), 1054–1064. doi:10.1016/j.compedu.2012.04.015
- Crane, L., Benachour, P., & Coulton, P. (2011). A user study of the spatial and temporal dimensions of context to support virtual learning environments. *Proceedings of the Mlearn 10th World Congress in Mobile and Contextual Learning*.
- Davis, F. D., Jr. (1986). *A technology acceptance model for empirically testing new end-user information systems: Theory and results* (Doctoral dissertation). Massachusetts Institute of Technology.
- Dey, A. K. (2001). Understanding and using context. *Personal and Ubiquitous Computing*, 5(1), 4–7. doi:10.1007/s007790170019
- Droit-Volet, S., & Gil, S. (2009). The time-emotion paradox. *Journal of Philosophical Transactions of the Royal Society, B (Biological Sciences)*, 364, 1943–1953.
- Fenesi, B., Kramer, E., & Kim, J. A. (2016). Split-Attention and Coherence Principles in Multimedia Instruction Can Rescue Performance for Learners with Lower Working Memory Capacity. *Applied Cognitive Psychology*, 30(5), 691–699. doi:10.1002/acp.3244
- Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research*. Reading, MA: Addison-Wesley.
- Hadlington, L. J. (2015). Cognitive failures in daily life: Exploring the link with Internet addiction and problematic mobile phone use. *Computers In Human Behaviour*, 51, 75–81. doi:10.1016/j.chb.2015.04.036
- Humphrey, W. S. (1988). Characterizing the software process: A maturity framework. *IEEE Software*, 5(2), 73–79. doi:10.1109/52.2014

- Humphrey, W. S. (1989). *Managing the Software Process. SEI series in software engineering*. Reading, MA: Addison-Wesley.
- JISC. (2014). *Developing digital literacies*. Accessed on 30 July 2016 from <https://www.jisc.ac.uk/guides/developing-digital-literacies>
- Jones, A., Gaved, M., Kukulska-Hulme, A., Scanlon, E., Pearson, C., Lamas, P., & Jones, J. et al. (2014). Creating Coherent Incidental Learning Journeys on Smartphones Using Feedback and Progress Indicators. *International Journal of Mobile and Blended Learning*, 6(4), 75–92. doi:10.4018/ijmbl.2014100105
- Judd, T. (2014). Making sense of multitasking: The role of Facebook. *Computers & Education*, 70, 194–202. doi:10.1016/j.compedu.2013.08.013
- Junco, R. (2012). In-class multitasking and academic performance. *Computers in Human Behavior*, 28(6), 2236–2243. doi:10.1016/j.chb.2012.06.031
- Kirschner, F., Paas, F., & Kirschner, P. A. (2009). A cognitive-load approach to collaborative learning: United brains for complex tasks. *Educational Psychology Review*, 21(1), 31–42. doi:10.1007/s10648-008-9095-2
- Kukulska-Hulme, A. (2007). Mobile Usability in Educational Contexts: What have we learnt? *International Review of Research in Open and Distance Learning*, 8(2), 1–16. doi:10.19173/irrodl.v8i2.356
- Kukulska-Hulme, A. (2012). Language learning defined by time and place: A framework for next generation designs. In J. E. D'Íaz-Vera (Ed.), *Left to My Own Devices: Learner Autonomy and Mobile Assisted Language Learning. Innovation and Leadership in English Language Teaching*, 6 (pp. 1–13). Bingley, UK: Emerald Group Publishing Limited. doi:10.1163/9781780526478_002
- Kukulska-Hulme, A. (2012). Language learning defined by time and place: A framework for next generation designs. In J. E. D'Íaz-Vera (Ed.), *Left to my own devices: Learner autonomy and mobile assisted language learning. innovation and leadership in English language teaching* (pp. 1–13). Bingley, UK: Emerald Group Publishing Limited; doi:10.1163/9781780526478_002
- Kukulska-Hulme, A. (2013). Mobile learners: who are they and who will they become? In Z. Berge & L. Y. Muilenburg (Eds.), *Handbook of Mobile Learning* (pp. 145–154). New York: Routledge.
- Kukulska-Hulme, A., & Jones, C. (2011). The next generation: design and the infrastructure for learning in a mobile and networked world. In A. D. Olofsson & J. Lindberg (Eds.), *Informed Design of Educational Technologies in Higher Education: Enhanced Learning and Teaching* (pp. 57–78). Hershey, PA: Information Science Reference (an Imprint of IGI Global).
- Lauricella, A. R., Wartella, E., & Rideout, V. J. (2015). Young children's screen time: The complex role of parent and child factors. *Journal of Applied Developmental Psychology*, 36, 11–17. doi:10.1016/j.appdev.2014.12.001
- Livingstone, S. (2015, April-June). From mass to social media? Advancing accounts of social change. *Social Media and Society*, 1–3.
- Marsh, J., Plowman, L., Yamada-Rice, D., Bishop, J. C., Lahmar, J., Scott, F., ... Winter, P. (2015). *Exploring Play and Creativity in Pre-Schoolers' Use of Apps: Final Project Report*. Retrieved May 17, 2016 from <http://www.techandplay.org>
- Mayer, R. E., & Moreno, R. (2003). Nine Ways to Reduce Cognitive Load in Multimedia Learning. *Educational Psychologist*, 38(1), 43–52. doi:10.1207/S15326985EP3801_6

- Ofcom. (2015). *Adults' media use and attitudes report*. Retrieved 30 July 2016 from <http://stakeholders.ofcom.org.uk/market-data-research/other/research-publications/adults/media-lit-10years/>
- Pachler, N., Bachmair, B., & Cook, J. (2010). *Mobile learning: structures, agency, practices*. New York: Springer. doi:10.1007/978-1-4419-0585-7
- Paivio, A. (1986). *Mental representations: a dual coding approach*. Oxford, UK: Oxford University Press.
- Park, S. Y. (2009). An Analysis of the Technology Acceptance Model in Understanding University Students' Behavioral Intention to Use e-Learning. *Journal of Educational Technology & Society*, 12(3), 150–162.
- Park, S. Y., Nam, M. W., & Cha, S. B. (2012). University students behavioral intention to use mobile learning: Evaluating the technology acceptance model. *British Journal of Educational Technology*, 43(4), 592–605. doi:10.1111/j.1467-8535.2011.01229.x
- Ramsay, J., & Terras, M. (2015, May-July). The Pendulum Swing of User Instruction and Interaction: Resurrection of How to Use Technology to Learn in the 21st Century. *E-Learning and Digital Media*, 12(12), 372–390. doi:10.1177/2042753015571827
- Ramsay, J., & Terras, M. (2016). Digital Learning and Media Literacy. In M. A. Peters (Ed.), *Encyclopaedia of Educational Philosophy and Theory*. Springer Science and Business Media Singapore; doi:10.1007/978-981-287-532-7_127-1
- Rosen, L. D., Carrier, L. M., & Cheever, N. A. (2013). Facebook and texting made me do it: Media-induced task-switching while studying. *Computers in Human Behavior*, 29(3), 948–958. doi:10.1016/j.chb.2012.12.001
- Roy, A., & Sharples, M. (2015). Mobile Game Based Learning: Can it enhance learning of marginalized peer educators? *International Journal of Mobile and Blended Learning*, 7(1), 1–13. doi:10.4018/ijmbl.2015010101
- Sanbonmatsu, D. M., Strayer, D. L., Medeiros-Ward, N., & Watson, J. M. (2013). Who multi-tasks and why? Multitasking ability, perceived multi-tasking ability, impulsivity, and sensation seeking. *PLoS ONE*, 8(1), e54402. doi:10.1371/journal.pone.0054402 PMID:23372720
- Sharples, M. (2000). The design of personal mobile technologies for lifelong learning. *Computers & Education*, 34(3-4), 177–193. doi:10.1016/S0360-1315(99)00044-5
- Sharples, M., Scanlon, E., Ainsworth, S., Anastopoulou, S., Collins, T., Crook, C., & O'Malley, C. et al. (2015). Personal Inquiry: Orchestrating Science Investigations Within and Beyond the Classroom. *Journal of the Learning Sciences*, 24(2), 308–341. doi:10.1080/10508406.2014.944642
- Sharples, M., Taylor, J., & Vavoula, G. (2007). A theory of learning for the mobile age. In R. Andrews & C. Haythornthwaite (Eds.), *The Sage Handbook of E-learning Research* (pp. 221–247). London, UK: Sage. doi:10.4135/9781848607859.n10
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285. doi:10.1207/s15516709cog1202_4
- Terras, M. M., & Ramsay, J. (2012). The five central psychological challenges facing effective mobile learning. *British Journal of Educational Technology*, 43(5), 707–832. doi:10.1111/j.1467-8535.2012.01362.x

Terras, M. M., & Ramsay, J. (2014a). A Psychological Perspective on the Temporal Dimensions of E-learning. *E-Learning and Digital Media*, 11(2), 108–123. doi:10.2304/elea.2014.11.2.108

Terras, M. M., & Ramsay, J. (2014b). E-learning, Mobility and Time: A Psychological Framework. In E. Barberà & P. Reimann (Eds.), *Assessment and Evaluation of Time Factors in Online Teaching and Learning* (pp. 63–90). Hershey, PA: IGI Global Publishing. doi:10.4018/978-1-4666-4651-3.ch003

Terras, M. M., Ramsay, J., & Boyle, E. (2011). A Psychological Perspective on Media Literacy. In Science Education in a rapidly changing world. Nova Science Publishers.

Traxler, J. (2007, June). Defining, Discussing, and Evaluating Mobile Learning: The moving finger writes and having writ.... *International Review of Research in Open and Distance Learning*, 8(2). doi:10.19173/irrodl.v8i2.346

Traxler, J. (2009). Learning in a mobile age. *International Journal of Mobile and Blended Learning*, 1(1), 1–12. doi:10.4018/jmbl.2009010101

Traxler, J., Barcena, E., & Laborda, J. G. (2016). Mobile Technology for Foreign Language Teaching: Building Bridges between Non-formal and Formal Scenarios. *Journal of Universal Computer Science*, 21(10), 1234–1247.

UNESCO. (2013). *The future of mobile learning: the implications for policy makers and planners*. Retrieved 30 July 2016 from unesdoc.unesco.org/images/0021/002196/219637e.pdf

Vavoula, G. N., Lefrere, P., O'Malley, C., Sharples, M., & Taylor, J. (2004). Producing guidelines for learning, teaching and tutoring in a mobile environment. In J. Roschelle, T. Chan, Y. Kinschuk, & Yand, S. J. H., (Eds.), *Proceedings of the 2nd IEEE International Workshop on Wireless and Mobile Technologies in Education (WMTE)* (pp.173-176). Los Alamitos, CA: Computer Society Press. doi:10.1109/WMTE.2004.1281378

Vavoula, G. N., & Sharples, M. (2008). Challenges in Evaluating Mobile Learning. In J. Traxler, B. Riordan, & C. Dennett (Eds.), *Proceedings of the mLearn 2008 Conference*. University of Wolverhampton.

Zaky, D. (2012). Experiencing time in daily life. *The Psychologist*, 25(8), 578–581.

Zhou, Y. (2012). Towards Capability Maturity Model of e-Learning Process. *Intelligent Information Management*, 4(4), 95–98. doi:10.4236/iim.2012.44015

ADDITIONAL READING

Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and under water. *British Journal of Psychology*, 66(3), 325–331. doi:10.1111/j.2044-8295.1975.tb01468.x

Hutchins, E. (1996). *Cognition in the wild*. Cambridge, MA: The MIT Press.

Jones, C. and Shao, B. (2011). The Net Generation and Digital Natives: Implications for Higher Education. *A literature review commissioned by the Higher Education Academy*.

Kukulska-Hulme, A. (2007). Mobile Usability in Educational Contexts: What have we learnt? *International Review of Research in Open and Distance Learning*, 8(2), 1–16. doi:10.19173/irrodl.v8i2.356

Mayer, R. E., & Moreno, R. (2003). Nine Ways to Reduce Cognitive Load in Multimedia Learning. *Educational Psychologist*, 38(1), 43–52. doi:10.1207/S15326985EP3801_6

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science*, 21(1), 8–14. doi:10.1177/0963721411429458 PMID:22773897

- Pachler, N., Bachmair, B., & Cook, J. (2010). *Mobile learning: structures, agency, practices*. New York: Springer. doi:10.1007/978-1-4419-0585-7
- Park, S. Y. (2009). An Analysis of the Technology Acceptance Model in Understanding University Students' Behavioral Intention to Use e-Learning. *Journal of Educational Technology & Society*, 12(3), 150–162.
- Sharples, M. (2000). The design of personal mobile technologies for lifelong learning. *Computers & Education*, 34(3-4), 177–193. doi:10.1016/S0360-1315(99)00044-5
- Sharples, M., Taylor, J., & Vavoula, G. (2007). A theory of learning for the mobile age. In R. Andrews & C. Haythornthwaite (Eds.), *The Sage Handbook of E-learning Research* (pp. 221–247). London, UK: Sage. doi:10.4135/9781848607859.n10
- Smith, S. M. (1988). Environmental context-dependent memory. In G. Davies (Ed.), *Memory in context* (pp. 13–31). New York: John Wiley & Sons Ltd.
- Suchman, L. A. (1987). *Plans and Situated Actions: The Problem of Human-Machine Communication*. Cambridge: Cambridge Press.
- Suchman, L. A. (2007). *Human-machine reconfigurations*. New York: Cambridge University Press.
- Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction*, 12(3), 185–233. doi:10.1207/s1532690xc1203_1
- Terras, M. M., & Ramsay, J. (2012). The five central psychological challenges facing effective mobile learning. [Wiley-Blackwell.]. *British Journal of Educational Technology*, 43(5), 707–832. doi:10.1111/j.1467-8535.2012.01362.x
- Terras, M. M., & Ramsay, J. (2014a). E-learning, Mobility and Time: A Psychological Framework. In E. Barberà & P. Reimann (Eds.), *Assessment and Evaluation of Time Factors in Online Teaching and Learning* (pp. 63–90). Hershey: IGI Global Publishing. doi:10.4018/978-1-4666-4651-3.ch003
- Terras, M. M., & Ramsay, J. (2014b). A Psychological Perspective on the Temporal Dimensions of E-learning. *Journal of E-Learning and Digital Media*, 11(2), 108–123. doi:10.2304/elea.2014.11.2.108
- Trafton, J. G., Altmann, E. M., & Brock, D. P. (2005). Huh, what was I doing? How people use environmental cues after an interruption. *Human Factors and Ergonomics Society Annual Meeting Proceedings*, 49, 468–472.
- Traxler, J. (2009). Learning in a mobile age. *International Journal of Mobile and Blended Learning*, 1(1), 1–12. doi:10.4018/jmbl.2009010101
- UNESCO. (2013). The future of mobile learning: the implications for policy makers and planners. Accessed 27/03/2014 from unesdoc.unesco.org/images/0021/002196/219637e.pdf
- Vygotsky, L. S. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.
- Zaky, D. (2012). Experiencing time in daily life. *The Psychologist*, 25(8), 578–581.
- Zhou, Y. (2012). Towards Capability Maturity Model of e-Learning Process. *Intelligent Information Management*, 4(4), 95–98. doi:10.4236/iim.2012.44015

KEY TERMS AND DEFINITIONS

Cognitive Constraints: Refer to the finite nature of human mental processing.

Context Awareness: To the ability of a human or a technology to characterise and exploit knowledge about its current environment.

Media Literacy: Ability to access, evaluate and create content effectively and responsibly via different media types.

Meta-Cognition: Thinking about thinking – the ability to control, manage and manipulate one's mental processes.

Mobile Learning: Mobile learning refers to use of handheld devices such as mobile phones

and tablets to access both formal and informal (e.g. general internet-based information) learning resources.

Psychological Challenges: Refers to the difficulties and obstacles that mobile learners encounter due to their finite mental processing ability.

Psychological Time: This is the subjective experience of time i.e. how time is perceived by the individual. It contrasts with objective or physical time that is conceptualised in standardised units such as seconds, hours and days.

Psychology: The scientific study of human behaviour.

M

The Role of Distance Education in Global Education

Kijpokin Kasemsap

Suan Sunandha Rajabhat University, Thailand

INTRODUCTION

Distance education mainly serves learners who cannot attend the face-to-face courses or programs for one or another reason (Ekmekçi, 2015). Learners stay at home or office and follow the course, do the assignments, and interact with each other and the teacher via the Internet (Ekmekçi, 2015). Distance education contributes to a large portion of the population has access to education and represents a possibility of meeting the new formative demands arising apace, with the adoption of a teaching format that is not restricted to classroom activities and mainly due to its low cost and its high degree of flexibility enables the approximation of the long distances and the democratization of access to education for learners seeking the better use of time, compatibility, and flexibility of schedules, geographic independence, among other learners (Rodrigues, Affonso, Quinelato, & Montiel, 2014).

Lifelong learning and knowledge management are practically directed at employment, career opportunities toward continuing professional education for both employed and unemployed people (Kasemsap, 2016a). Distance education develops in the field of technical education, which is connected with the characteristics of the educational material of humanities disciplines, difficult to formalize and, secondly, with the greater competence of pedagogic staff of technical education in the utilization of new information and communication technology (ICT) and modern software (Nurmukhametov, Temirova, & Bekzhanova, 2015). Distance education course formats can alter the

modes of information exchange and interpersonal interaction related to traditional course formats (Mackey & Freyberg, 2010). There has been an explosive growth in distance education which is rapidly transforming the traditional learning habits with the more recent ones utilizing the latest educational technology (Paksoy, 2015).

This article aims to bridge the gap in the literature on the thorough literature consolidation of distance education. The extensive literature of distance education provides a contribution to practitioners and researchers by describing the current issues and approaches in distance education to appeal to the different segments of distance education in global education.

BACKGROUND

For the last twenty years, many universities and educational institutions tried to find the new ways and methods in order to effectively utilize the Internet-based technologies in teaching and learning through the distance education initiatives (Altunisik, 2013). The increasing educational demand, time limitations, and geographical obstacles practically increase the demand of distance education (Ilgaz & Askar, 2013). With the development of Internet technologies, online distance education is becoming an important format for teaching, learning, and instruction (Li, 2007). Internet technologies are realized that they can employ in the training system and they are recognized as the educational facilitator in the distance education (Çelik, Ferikoglu, Kiraz, Albayrak, &

Ekiz, 2013). Distance education, having arisen at the end of the 20th century, enters the 21 century as one of the most effective systems of learners' training (Lenar, Artur, Ullubi, & Nailya, 2014).

Education in the 21st century requires the implementation of innovation in the teaching technology (Malganova & Rahkimova, 2015). In parallel with the recent developments, education has initiated to utilize the Internet media (Çelik et al., 2013). Distance education exploits the Internet intended for course supply for the students and educators in the different locations (Paksoy, 2015). Learners electronically interconnect with the educators in global education (Sims, Dobb, & Hand, 2002). Regarding distance education, course content can be presented through the recorded lectures, slide presentations, text-based documents, interactive online activities of various forms, audio files, video files, pod casts, wikis, live chat sessions, online discussions, live online presentations, videoconferencing, and other methods. Video cases can capture the complexity of the classroom context and provide the efficient way to expose the learners to the authenticity of the classroom (Bayram, 2013).

The emergence of worldwide communications networks and powerful computer technologies has redefined the concept of distance learning and the delivery of engineering education content (Bourne, Harris, & Mayadas, 2005). Distance education continues to have a significant impact on higher education (Hollis & Madill, 2006). With pressures to increase the access to higher education, colleges and universities have focused on increasing the number of online courses and educational programs offered (Meyer, 2014). Distance education can overcome the barriers to the learning (e.g., time limitation and geographical location) and provide the online communities of practice in circumstances where different factors make it difficult (Siri & Rui, 2015).

Distance education, with its growing availability and reduced intrinsic costs, becomes more prevalent in global education (Buxton &

De Muth, 2013). Distance education technologies are considered as the major factors promoting the development of network interaction for the educational institutions (Mozhaeva, 2014). Distance education is developed in order to meet the needs of learners who cannot participate in the face-to-face classes (Beldarrain, 2006). Many distance education courses have been created and disseminated by using the Internet or similar network systems in support of pedagogical communication concerning technical improvement to the undergraduate and postgraduate students (Rodrigues et al., 2014).

Distance education has been recognized as an effective pedagogical method and tool, and is broadly integrated into various types of teaching and learning strategies in higher education (Park, 2011). Regarding distance education, digital libraries comprise digital collections, services, and infrastructure to educationally support the lifelong learning, research, and conservation of the recorded knowledge (Kasemsap, 2016b). Continuing professional development is the training and education that continues throughout an individual's career in order to improve the skills and knowledge (Kasemsap, 2017a). Educational computer games can motivate students to develop the basic competencies and encourage challenging themselves to be better and learn the additional knowledge related to the important tasks (Kasemsap, 2017b).

Advances in multimedia and Internet technologies have generated the explosive growth in distance education (Hay, Peltier, & Drago, 2004). Online activities and educational tools (e.g., multimedia and discussion boards) increase the emotional engagement in distance education (Sun & Rueda, 2012). The Internet and multimedia, including the non-linear integration of video, audio, text, and graphics can provide the rich environment for case studies which promote the mental construction of knowledge toward integrating the modern technology into the curriculum (Cannings & Talley, 2002). Multimedia,

a computer-based application, the normal text, audio, images, graphics, animation, video, and the supported use of visual encouragement is the perspective of the user interface in distance education (Bourne et al., 2005).

ADVANCED ISSUES OF DISTANCE EDUCATION IN GLOBAL EDUCATION

This section emphasizes the overview of distance education and the importance of distance education in global education.

Overview of Distance Education

Distance education is defined as the learning process in which individuals have the opportunity for learning, independent from time and place, and in which various techniques are used in the learning activities (Kor, Aksoy, & Erbay, 2014). Distance education is defined as the process of extending learning, or delivering instructional resource-sharing opportunities, to locations away from the classroom, building or site by using video, audio, computers, multimedia communications, or some combination of these with other traditional delivery methods (O'Neil, Fisher, & Newbold, 2009). Web-based distance education has become an economically viable option, due to the near universal availability of computers in the workplace and home, and to technology improvements (Buxton & De Muth, 2013).

Distance education makes the learner effectively learn through the technological tools (Rashid, Khokhar, & Tahir, 2013). Eisinger (2000) stated that distance education is the planned learning experience toward encouraging the student's communication. The most widely used presentation systems in distance education are the printed materials, television broadcasting, broadcasting via microwave, and other electronic methods, such as computer conferencing, audioconferencing, and

videoconferencing (Keating, 2006). Despite long distance between the learning and teaching groups, the educational groups have the opportunity to learn in the virtual classes, similar to traditional classroom setting, through computers with Internet connection (Kor et al., 2014).

With the growing popularity of distance education opportunities, students' attitudes toward distance education is rapidly changing (Çelik & Uzunboylu, 2015). Service quality of distance education applications in the education sector can be investigated through five service quality determinants (i.e., tangibles, reliability, responsiveness, credibility, and empathy) (Dursun, Oskaybas, & Gokmen, 2013). Satisfaction with learning, level of control of the learning process, and study motivation for distance education are positively related to the students' preferences for the structured distance education (Katz, 2002). Online communication, electronic work submission, and online meeting capabilities mean that participants can complete the courses without attending the class in the university (Allworth, 2014).

Importance of Distance Education in Global Education

Distance education is a significant topic of discussion within institutions at all levels of education (Annetta & Shymansky, 2006). The rapid spread of distance education in the world not only offers learners and teachers great opportunities but also brings a lot of challenges (Ekmekçi, 2015). Distance education is the economic and interactive education perspective using information technology (Çelik et al., 2013). Distance education is among the significant fields for the application of educational technology (Ding, Niu, & Han, 2010). Activities in the field of distance education have shown a significant improvement in the world (Kor et al., 2014). Distance education students have opportunities to structure their own learning contexts as they engage with individuals in their everyday environment (Ferguson, 2010).

Distance education has become deeply embedded in the offerings of most community colleges (Milam, Voorhees, & Bedard-Voorhees, 2004). Distance education allocates importance to interaction, thus establishing several distance education environments (Ergül & Koç, 2013). Distance education relies heavily on self-learning and self-learning resources, although the interaction at the designated learning centers and online learning platforms are occasionally offered (Zhao, Chen, & Panda, 2014). Learning communities can provide learners with an educational environment conducive to the increased interactions and alleviate their feeling of isolation (Yuan & Kim, 2014). The learner-learner interaction concentrates on technology-enhanced learning using video communications which provide opportunities for more authentic online collaborative learning (Smyth, 2011).

In distance education, the most common interaction-enabling activities are forum and chat (Scarinci, 2015). The size of the demand for higher education both in terms of campus or open and distance education, depends on the employment opportunities of the graduates more than the offered training programs and the images of the educational institutions (Berberoglu, 2015). Distance education provides access to education with satellite, video, audio, graphics, computer, multimedia technology with the support of electronic devices (Çelik et al., 2013). Computer-computerized systems (e.g., laptops, tablets, and smartphones) enable the establishment of the educational programs from which the learners can get education by utilizing the Internet (Çelik & Uzunboylu, 2015). Educators normally use distance education for delivering the entire courses and the educational programs worldwide, or for providing the learning experiences for educational courses, staff development, and continuing education (Billings, 2007; Hyde & Murray, 2005).

In order to deliver the library instruction and promote the library resources to the distance education students, distance education librarians

often take advantage of computer applications to produce the audio/video and web-based tutorials (Tang, 2013). Distance education librarians also use the videoconferencing and the virtual classroom technology to teach the information literacy (Tang, 2013). In the concept of distance education, the remote controlled laboratories and the virtual laboratories provide an educational opportunity for the laboratory work by the remote control of the real laboratory equipment in global education (Stefanovic, 2013). Distance education can contribute to the health care professionals' practice knowledge (Moule, Ward, & Lockyer, 2010). Distance education significantly decreases the cost of academic health education at all levels (Mattheos, Schitteck, Attström, & Lyon, 2001). Distance education can take a variety of forms including in the health professional education concerning the simulations of clinical scenarios (Tait, Tait, Thornton, & Edwards 2008), the Internet-based discussion boards, and the role play (Nelson & Blenkin, 2007).

FUTURE RESEARCH DIRECTIONS

The classification of the extensive literature in the domains of distance education will provide the potential opportunities for future research. E-learning allows students to choose content and tools appropriate to their differing interests, needs, and skill levels (Kasemsap, 2016c). Web-based learning supports the open learning concept by providing students with the ability to connect to the educational resources when it is convenient for them, and allowing students to explore the educational resources in an effective order that suits their educational needs (Kasemsap, 2016d). Learning analytics can gather data, analyze data, generate reports, and enable interventions in the modern educational world (Kasemsap, 2016e).

Regarding mobile learning, the widespread adoption of mobile devices coupled with the powerful computational and communication capacity of convenient mobile devices promises

a fundamental shift in learning and knowledge work. Mobile learning often takes place outside a formal learning environment, and it tends to become personalized through users' personal mobile devices. Educational computer games educationally create learning experiences, provide personalized learning opportunities, and encourage a learning environment for authentic and relevant assessment in the game-based learning environment. Future research direction should broaden the perspectives in the promotion of e-learning, web-based learning, learning analytics, mobile learning, educational computer games, and distance education in the modern learning environments.

CONCLUSION

This article highlighted the overview of distance education and the importance of distance education in global education. Distance learning becomes an alternative to traditional classrooms. Students can benefit from the flexibility, convenience, effectiveness, and affordability that come with distance education. For students who do not have the time or money to attend the traditional schools, distance education can provide a path to higher education. Distance education can increase the opportunities for students across the globe, ensuring that students do not have to choose a school based on its location.

Because of the high level of flexibility that an online program offers to people all over the world, it is common to network internationally as a result of program participation. One of the benefits of distance learning is that there is a wide variety of materials that can meet everyone's learning preference, at least part of the time. Distance education enables students to work at their own pace in many circumstances, and leads to networking with an even wider variety of people than you might encounter at a local institution. Facilitating

distance education has the potential to improve educational performance and gain sustainable competitive advantage in global education.

REFERENCES

- Allworth, M. B. (2014). Postgraduate distance education in sheep health veterinary education. *Small Ruminant Research*, 118(1/3), 97–99. doi:10.1016/j.smallrumres.2013.12.013
- Altunisik, R. (2013). The role of lecturer related factors in students' perceptions and satisfaction in distance education. *Procedia: Social and Behavioral Sciences*, 106, 3075–3083. doi:10.1016/j.sbspro.2013.12.355
- Annetta, L. A., & Shymansky, J. A. (2006). Investigating science learning for rural elementary school teachers in a professional-development project through three distance-education strategies. *Journal of Research in Science Teaching*, 43(10), 1019–1039. doi:10.1002/tea.20106
- Bayram, L. (2013). Enhancing an online distance education course with video. *Procedia: Social and Behavioral Sciences*, 83, 463–467. doi:10.1016/j.sbspro.2013.06.091
- Beldarrain, Y. (2006). Distance education trends: Integrating new technologies to foster student interaction and collaboration. *Distance Education*, 27(2), 139–153. doi:10.1080/01587910600789498
- Berberoglu, B. (2015). Open and distance education programs of Anadolu University since the establishment. *Procedia: Social and Behavioral Sciences*, 174, 3358–3365. doi:10.1016/j.sbspro.2015.01.1004
- Billings, D. M. (2007). Distance education in nursing: 25 years and going strong. *Computers, Informatics, Nursing*, 25(3), 121–123. doi:10.1097/01.NCN.0000270044.67905.4a PMID:17496472

- Bourne, J., Harris, D., & Mayadas, F. (2005). On-line engineering education: Learning anywhere, anytime. *The Journal of Engineering Education*, 94(1), 131–146. doi:10.1002/j.2168-9830.2005.tb00834.x
- Buxton, E. C., & De Muth, J. E. (2013). Pharmacists' perceptions of a live continuing education program comparing distance learning versus local learning. *Research in Social & Administrative Pharmacy*, 9(2), 230–235. doi:10.1016/j.sapharm.2012.05.003 PMID:22835712
- Cannings, T. R., & Talley, S. (2002). Multimedia and online video case studies for pre-service teacher preparation. *Education and Information Technologies*, 7(4), 359–367. doi:10.1023/A:1020969723060
- Çelik, B., & Uzunboylu, H. (2015). High school students' attitudes towards distance education: Comparative study. *Procedia: Social and Behavioral Sciences*, 197, 292–297. doi:10.1016/j.sbspro.2015.07.139
- Çelik, F., Ferikoglu, A., Kiraz, A., Albayrak, Z., & Ekiz, H. (2013). An adaptation of circuit analysis course to distance education with content development. *Procedia: Social and Behavioral Sciences*, 103, 101–108. doi:10.1016/j.sbspro.2013.10.313
- Ding, X., Niu, J., & Han, Y. (2010). Research on distance education development in China. *British Journal of Educational Technology*, 41(4), 582–592. doi:10.1111/j.1467-8535.2010.01093.x
- Dursun, T., Oskaybas, K., & Gokmen, C. (2013). The quality of service of the distance education. *Procedia: Social and Behavioral Sciences*, 103, 1133–1151. doi:10.1016/j.sbspro.2013.10.441
- Eisinger, J. (2000). Education evolution. *Journal of Association Management*, 52(13), 52–59.
- Ekmekçi, E. (2015). Distance-education in foreign language teaching: Evaluations from the perspectives of freshman students. *Procedia: Social and Behavioral Sciences*, 176, 390–397. doi:10.1016/j.sbspro.2015.01.487
- Ergül, E., & Koç, M. (2013). The role of animated agents in web-based distance education. *Procedia: Social and Behavioral Sciences*, 83, 1016–1022. doi:10.1016/j.sbspro.2013.06.189
- Ferguson, R. (2010). Peer interaction: The experience of distance students at university level. *Journal of Computer Assisted Learning*, 26(6), 574–584. doi:10.1111/j.1365-2729.2010.00386.x
- Hay, A., Peltier, J. W., & Drago, W. A. (2004). Reflective learning and on-line management education: A comparison of traditional and on-line MBA students. *Strategic Change*, 13(2), 169–182. doi:10.1002/jsc.680
- Hollis, V., & Madill, H. (2006). Online learning: The potential for occupational therapy education. *Occupational Therapy International*, 13(2), 61–78. doi:10.1002/oti.209 PMID:16862957
- Hyde, A., & Murray, M. (2005). Nurses' experiences of distance education programmes. *Journal of Advanced Nursing*, 49(1), 87–95. doi:10.1111/j.1365-2648.2004.03267.x PMID:15610385
- Ilgaz, H., & Askar, P. (2013). The contribution of technology acceptance and community feeling to learner satisfaction in distance education. *Procedia: Social and Behavioral Sciences*, 106, 2671–2680. doi:10.1016/j.sbspro.2013.12.308
- Kantek, F. (2015). Distance education in nursing in Turkey. *Procedia: Social and Behavioral Sciences*, 116, 639–643. doi:10.1016/j.sbspro.2014.01.271
- Kasemsap, K. (2016a). The roles of lifelong learning and knowledge management in global higher education. In P. Ordóñez de Pablos & R. Tennyson (Eds.), *Impact of economic crisis on education and the next-generation workforce* (pp. 71–100). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-9455-2.ch004
- Kasemsap, K. (2016b). Mastering digital libraries in the digital age. In E. de Smet & S. Dhamdhare (Eds.), *E-discovery tools and applications in modern libraries* (pp. 275–305). Hershey, PA: IGI Global. doi:10.4018/978-1-5225-0474-0.ch015

- Kasemsap, K. (2016c). The roles of e-learning, organizational learning, and knowledge management in the learning organizations. In E. Railean, G. Walker, A. Elçi, & L. Jackson (Eds.), *Handbook of research on applied learning theory and design in modern education* (pp. 786–816). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-9634-1.ch039
- Kasemsap, K. (2016d). Exploring the role of web-based learning in global education. In M. Raisinghani (Ed.), *Revolutionizing education through web-based instruction* (pp. 202–224). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-9932-8.ch012
- Kasemsap, K. (2016e). The role of learning analytics in global higher education. In M. Anderson & C. Gavan (Eds.), *Developing effective educational experiences through learning analytics* (pp. 282–307). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-9983-0.ch012
- Kasemsap, K. (2017a). Encouraging continuing professional development and teacher professional development in global education. In R. Cintron, J. Samuel, & J. Hinson (Eds.), *Accelerated opportunity education models and practices* (pp. 168–202). Hershey, PA: IGI Global. doi:10.4018/978-1-5225-0528-0.ch008
- Kasemsap, K. (2017b). Mastering educational computer games, educational video games, and serious games in the digital age. In R. Alexandre Peixoto de Queirós & M. Pinto (Eds.), *Gamification-based e-learning strategies for computer programming education* (pp. 30–52). Hershey, PA: IGI Global. doi:10.4018/978-1-5225-1034-5.ch003
- Katz, Y. J. (2002). Attitudes affecting college students' preferences for distance learning. *Journal of Computer Assisted Learning*, 18(1), 2–9. doi:10.1046/j.0266-4909.2001.00202.x
- Keating, S. B. (2006). *Curriculum development and evaluation in nursing*. Philadelphia, PA: Lippincott Williams & Wilkins.
- Kor, H., Aksoy, H., & Erbay, H. (2014). Comparison of the proficiency level of the course materials (animations, videos, simulations, e-books) used in distance education. *Procedia: Social and Behavioral Sciences*, 141, 854–860. doi:10.1016/j.sbspro.2014.05.150
- Lenar, S., Artur, F., Ullubi, S., & Nailya, B. (2014). Problems and decision in the field of distance education. *Procedia: Social and Behavioral Sciences*, 131, 111–117. doi:10.1016/j.sbspro.2014.04.088
- Li, X. (2007). Intelligent agent-supported online education. *Decision Sciences Journal of Innovative Education*, 5(2), 311–331. doi:10.1111/j.1540-4609.2007.00143.x
- Mackey, K. R. M., & Freyberg, D. L. (2010). The effect of social presence on affective and cognitive learning in an international engineering course taught via distance learning. *The Journal of Engineering Education*, 99(1), 23–34. doi:10.1002/j.2168-9830.2010.tb01039.x
- Malganova, I., & Rahkimova, A. (2015). The usage of distance education practice of the world leading universities in Russian multiethnic region. *Procedia: Social and Behavioral Sciences*, 191, 2622–2625. doi:10.1016/j.sbspro.2015.04.392
- Mattheos, N., Schitteck, M., Attström, R., & Lyon, H. C. (2001). Distance learning in academic health education. *European Journal of Dental Education*, 5(2), 67–76. doi:10.1034/j.1600-0579.2001.005002067.x PMID:11683217
- Meyer, K. A. (2014). Student engagement in online learning: What works and why. *ASHE Higher Education Report*, 40(6), 1–114. doi:10.1002/aehe.20018
- Milam, J., Voorhees, R. A., & Bedard-Voorhees, A. (2004). Assessment of online education: Policies, practices, and recommendations. *New Directions for Community Colleges*, 2004(126), 73–85. doi:10.1002/cc.156

- Moule, P., Ward, R., & Lockyer, L. (2010). Nursing and healthcare students' experiences and use of e-learning in higher education. *Journal of Advanced Nursing*, 66(12), 2785–2795. doi:10.1111/j.1365-2648.2010.05453.x PMID:20946565
- Mozhaeva, G. (2014). Network interaction in distance education: Analysis of Russian experience. *Procedia: Social and Behavioral Sciences*, 152, 1124–1127. doi:10.1016/j.sbspro.2014.09.286
- Nelson, D. L., & Blenkin, C. (2007). The power of online role-play simulations: Technology in nursing education. *International Journal of Nursing Education Scholarship*, 4(1), 1–12. doi:10.2202/1548-923X.1263 PMID:17402925
- Nurmukhametov, N., Temirova, A., & Bekzhanova, T. (2015). The problems of development of distance education in Kazakhstan. *Procedia: Social and Behavioral Sciences*, 182, 15–19. doi:10.1016/j.sbspro.2015.04.729
- O'Neil, C. A., Fisher, C. A., & Newbold, S. K. (2004). *Developing an online course: Best practices for nurse educators*. New York, NY: Springer-Verlag.
- Paksoy, P. (2015). Use of global distance education program in educational goals of airline. *Procedia: Social and Behavioral Sciences*, 177, 372–378. doi:10.1016/j.sbspro.2015.02.365
- Park, J. Y. (2011). Designed education online: Learning delivery and evaluation. *International Journal of Art & Design Education*, 30(2), 176–187. doi:10.1111/j.1476-8070.2011.01689.x
- Rashid, M., Khokhar, A. G., & Tahir, A. A. (2013). Learner's attitude towards printed word in distance education. *The Journal of Educational Research*, 16(1), 106–111.
- Rodrigues, S. J., Affonso, S. A. B., Quinelato, E., & Montiel, J. M. (2014). Distance learning in undergraduate education: The challenges of building a collaborative environment. *Procedia: Social and Behavioral Sciences*, 116, 3499–3501. doi:10.1016/j.sbspro.2014.01.791
- Scarinci, A. L. (2015). Evaluating resources for scientific modelling in a distance education course. *Procedia: Social and Behavioral Sciences*, 167, 238–244. doi:10.1016/j.sbspro.2014.12.668
- Sims, R., Dobbs, G., & Hand, T. (2002). Enhancing quality in online learning: Scaffolded planning and design through proactive evaluation. *Distance Education*, 23(2), 135–148. doi:10.1080/0158791022000009169
- Siri, A., & Rui, M. (2015). Distance education for health professions' students. *Procedia: Social and Behavioral Sciences*, 174, 730–738. doi:10.1016/j.sbspro.2015.01.608
- Smyth, R. (2011). Enhancing learner–learner interaction using video communications in higher education: Implications from theorising about a new model. *British Journal of Educational Technology*, 42(1), 113–127. doi:10.1111/j.1467-8535.2009.00990.x
- Stefanovic, M. (2013). The objectives, architectures and effects of distance learning laboratories for industrial engineering education. *Computers & Education*, 69, 250–262. doi:10.1016/j.compedu.2013.07.011
- Strand, H., Fox-Young, S., Long, P., & Bogossian, F. (2013). A pilot project in distance education: Nurse practitioner students' experience of personal video capture technology as an assessment method of clinical skills. *Nurse Education Today*, 33(3), 253–257. doi:10.1016/j.nedt.2011.11.014 PMID:22154874
- Sun, J. C. Y., & Rueda, R. (2012). Situational interest, computer self-efficacy and self-regulation: Their impact on student engagement in distance education. *British Journal of Educational Technology*, 43(2), 191–204. doi:10.1111/j.1467-8535.2010.01157.x
- Tait, M., Tait, D., Thornton, F., & Edwards, M. (2008). Development and evaluation of a critical care e-learning scenario. *Nurse Education Today*, 28(8), 971–981. doi:10.1016/j.nedt.2008.05.016 PMID:18603339

Tang, Y. (2013). Distance education librarians in the United States: A study of job announcements. *Journal of Academic Librarianship*, 39(6), 500–505. doi:10.1016/j.acalib.2013.08.012

Yuan, J., & Kim, C. (2014). Guidelines for facilitating the development of learning communities in online courses. *Journal of Computer Assisted Learning*, 30(3), 220–232. doi:10.1111/jcal.12042

Zhao, H., Chen, L., & Panda, S. (2014). Self-regulated learning ability of Chinese distance learners. *British Journal of Educational Technology*, 45(5), 941–958. doi:10.1111/bjet.12118

ADDITIONAL READING

Abbott, L., Kreszock, M., Ochoa, L., & Purpur, G. (2013). Evolving technology, evolving roles for distance education librarians. In A. Sigal (Ed.), *Advancing library education: Technological innovation and instructional design* (pp. 197–212). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-3688-0.ch013

Akimova, O., Bobyрева, N., Palutina, O., & Pomortseva, N. (2015). Distance language education. *Procedia: Social and Behavioral Sciences*, 199, 348–356. doi:10.1016/j.sbspro.2015.07.558

Badiru, A. B., & Jones, R. R. (2012). A systems framework for distance learning in engineering graduate programs. *Systems Engineering*, 15(2), 191–202. doi:10.1002/sys.20202

Baran, E., Correia, A. P., & Thompson, A. (2011). Transforming online teaching practice: Critical analysis of the literature on the roles and competencies of online teachers. *Distance Education*, 32(3), 421–439. doi:10.1080/01587919.2011.610293

Çelen, F. K., Çelik, A., & Seferoğlu, S. S. (2013). Analysis of teachers' approaches to distance education. *Procedia: Social and Behavioral Sciences*, 83, 388–392. doi:10.1016/j.sbspro.2013.06.076

de Freitas, S. I., Morgan, J., & Gibson, D. (2015). Will MOOCs transform learning and teaching in higher education? Engagement and course retention in online learning provision. *British Journal of Educational Technology*, 46(3), 455–471. doi:10.1111/bjet.12268

Henckell, M., Kilburn, M., & Starrett, D. (2015). Components of a distance education evaluation system. In M. Khosrow-Pour (Ed.), *Encyclopedia of information science and technology* (3rd ed., pp. 2220–2228). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-5888-2.ch215

Hess, L. M. (2014). A begrudging, recalcitrant academic observes what she's learning: Distance learning in leadership formation. *Teaching Theology & Religion*, 17(3), 264–271. doi:10.1111/teth.12210

Hockridge, D. (2013). Challenges for educators using distance and online education to prepare students for relational professions. *Distance Education*, 34(2), 142–160. doi:10.1080/01587919.2013.793640

Holcomb, L. B., & Krüger-Ross, M. (2013). Enhancing social presence and communities of practice in distance education courses through social media. In B. Pătruț, M. Pătruț, & C. Cmeciu (Eds.), *Social media and the new academic environment: Pedagogical challenges* (pp. 92–114). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-2851-9.ch005

Jin, Q. (2011). *Distance education environments and emerging software systems: New technologies* (pp. 1–408). Hershey, PA: IGI Global. doi:10.4018/978-1-60960-539-1

Kaya, B. Y., Kaya, G., & Dağdeviren, M. (2014). A sample application of web based examination system for distance and formal education. *Procedia: Social and Behavioral Sciences*, 141, 1357–1362. doi:10.1016/j.sbspro.2014.05.234

- Kim, C., Park, S. W., & Cozart, J. (2014). Affective and motivational factors of learning in online mathematics courses. *British Journal of Educational Technology*, 45(1), 171–185. doi:10.1111/j.1467-8535.2012.01382.x
- Latchem, C., & Ryan, Y. (2013). Transnational distance education: Cultural and quality considerations. In J. Willems, B. Tynan, & R. James (Eds.), *Global challenges and perspectives in blended and distance learning* (pp. 55–72). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-3978-2.ch004
- Li, X. (2012). ICTs in Chinese distance higher education: Increased opportunities and continuous challenges. In S. Chhabra (Ed.), *ICTs for advancing rural communities and human development: Addressing the digital divide* (pp. 180–192). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-0047-8.ch012
- Lin, C., & Warschauer, M. (2015). Online foreign language education: What are the proficiency outcomes? *Modern Language Journal*, 99(2), 394–397. doi:10.1111/modl.12234_1
- Mashaw, B. (2012). A model for measuring effectiveness of an online course. *Decision Sciences Journal of Innovative Education*, 10(2), 189–221. doi:10.1111/j.1540-4609.2011.00340.x
- McCutcheon, K., Lohan, M., Traynor, M., & Martin, D. (2015). A systematic review evaluating the impact of online or blended learning vs. face-to-face learning of clinical skills in undergraduate nurse education. *Journal of Advanced Nursing*, 71(2), 255–270. doi:10.1111/jan.12509 PMID:25134985
- Ojo, B. A. (2011). Marketing e-learning and the challenges facing distance education in Africa. In U. Demiray & S. Sever (Eds.), *Marketing online education programs: Frameworks for promotion and communication* (pp. 235–247). Hershey, PA: IGI Global. doi:10.4018/978-1-60960-074-7.ch015
- Richardson, J. T. E. (2013). Approaches to studying across the adult life span: Evidence from distance education. *Learning and Individual Differences*, 26, 74–80. doi:10.1016/j.lindif.2013.04.012
- Rubio, F. (2015). Assessment of oral proficiency in online language courses: Beyond reinventing the wheel. *Modern Language Journal*, 99(2), 405–408. doi:10.1111/modl.12234_4
- van Deusen-Scholl, N. (2015). Assessing outcomes in online foreign language education: What are key measures for success? *Modern Language Journal*, 99(2), 398–400. doi:10.1111/modl.12234_2
- van Rooyen, A. (2015). Distance education accounting students' perceptions of social media integration. *Procedia: Social and Behavioral Sciences*, 176, 444–450. doi:10.1016/j.sbspro.2015.01.495
- Willems, J. (2013). Equity in distance education. In J. Willems, B. Tynan, & R. James (Eds.), *Global challenges and perspectives in blended and distance learning* (pp. 17–35). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-3978-2.ch002
- Xu, D., & Jaggars, S. S. (2013). The impact of online learning on students' course outcomes: Evidence from a large community and technical college system. *Economics of Education Review*, 37, 46–57. doi:10.1016/j.econedurev.2013.08.001
- Yuksekdag, B. B. (2015). The experts in design of distance nursing education. In G. Eby & T. Yuzer (Eds.), *Identification, evaluation, and perceptions of distance education experts* (pp. 106–120). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-8119-4.ch008
- Zengin, O., & Altindag, O. (2013). Social work education from distance. *Procedia: Social and Behavioral Sciences*, 106, 2426–2430. doi:10.1016/j.sbspro.2013.12.278

KEY TERMS AND DEFINITIONS

Distance Education: The learning that uses television, video tapes, computers, and Internet, instead of physical attendance at classes in a centralized facility.

Education: The wealth of knowledge acquired by an individual after studying the subject matters or experiencing life lessons that provide an understanding of something.

Information Technology: The science and activity of using computers and other electronic equipment to store and send the information.

Internet: The worldwide computer network that provides information on many subjects and enables users to exchange the messages.

Learning: The measurable and permanent change in behavior through experience, instruction, and study.

Multimedia: The integration of animation, audio, graphics, text, and full-motion video through computer hardware and software for education, entertainment, and training.

Technology: The use of scientific knowledge to solve the practical problems, especially in the industry and commerce.

Category M

Multimedia Technology

Adaptive Hypermedia Systems:

Ana Carolina Tomé Klock

Federal University of Rio Grande do Sul (UFRGS), Brazil

Isabela Gasparini

Santa Catarina State University (UDESC), Brazil

Marcelo Soares Pimenta

Federal University of Rio Grande do Sul (UFRGS), Brazil

José Palazzo M. de Oliveira

Federal University of Rio Grande do Sul (UFRGS), Brazil

INTRODUCTION

With the growth of access to technology, the limitation of the traditional hypermedia applications has become evident: it provides the same contents and links to all users, disregarding their profiles and individual characteristics. Considering a diverse user population, the traditional system is not able to be “all things to all people”, since students with different goals will need different explanations, readers with different interests will enjoy different books and visitors with different knowledge will avail different curiosities about a place (Brusilovsky, 2001). An attempt to bypass the negative effects of the “one-size-fits-all” approach is in providing these systems the ability to adapt themselves according to the needs of the individual user (Brusilovsky et al., 1998).

Based on that, Adaptive Hypermedia (AH) systems are all hypertext or hypermedia systems that reflect some features of the user (e.g., goals, preferences and knowledge) in a user model and modify their various visible aspects according to this model (Brusilovsky, 1996). When hypermedia records and stores learners’ preferences, individual needs and learning goals, using them throughout interaction in order to meet their personal needs and to promote a better learning, then hypermedia

becomes AH (Turel, 2015). To be considered an AH, the system should satisfy three criteria: being a hypermedia (or hypertext) system, having a user model and using this model to adapt the hypermedia (Brusilovsky, 1996).

In order to meet the first criterion, the system should store all information into small units (usually called nodes) containing text (hypertext) or other forms of information (hypermedia). These units are presented one per window, being interconnected by links that allow users to navigate in a predefined way (Akscyn et al., 1988). To meet the second criterion, it must have a model where data about the users (e.g., knowledge, interests, goals and individual traits) are stored (Brusilovsky & Millán, 2007). Finally, the AH system must adjust its presentation – content (defining the most relevant contents to each user and how to structure them) and interface (making a presentation more efficient and meaningful to the user) (Bunt et al., 2007) – and navigation (guiding the user in the hyperspace through the selection and presentation of the most important links in order to continue the navigation system) (Brusilovsky, 2007) based on the data to be stored in the user model.

There are different application areas to AH systems, e.g.: Educational Hypermedia, On-line Information, On-line Help, Information Retrieval,

Institutional Information, Personalized Views, etc. The AH educational systems goal is to create an instructionally sound and flexible environment that supports learning for students with a range of abilities, disabilities, interests, backgrounds, and other characteristics, enabling students to learn as much as possible on the available material (Shute & Zapata-Rivera, 2012).

The aim of this chapter is to present Adaptive Hypermedia Systems fundamentals and trends. To achieve it, section 2 introduces the background about user modeling. Section 3 describes some methods and techniques that allow adaptation. Section 4 illustrates some of the existing AH systems. Section 5 points out trends and future directions and Section 6 concludes this work.

BACKGROUND

User model is the representation of information about an individual user, which is essential to make the AH system behave differently according to each user. Based on this model, the AH system can prioritize the most relevant results of a search, manipulate links to facilitate the navigation and present the content adaptively. During the user modeling, the amount and nature of the sorted user information depends on the kind of adaptation effect the AH system has to deliver (i.e., content, presentation and/or navigation) (Brusilovsky & Millán, 2007). The most common stored data are described below based on Brusilovsky & Millán (2007) and Schiaffino & Amandi (2009).

- Knowledge about the application domain enables proper assistance or content adaptation. The users' knowledge is a changeable feature that can increase (i.e., learning) or decrease (i.e., forgetting) from time to time. The AH system must recognize this change and update the user model to provide a more reliable and assertive adaptation. The knowledge about each element can be represented in binary form (i.e., knowing or not) or in scalar form, either quantitative (e.g., from zero to ten) or qualitative (e.g., beginner, intermediate, expert).
- Interests can represent news, web pages, documents, work-related or hobbies-related topics that can be classified as short-term or long-term interests. These interests can be stored in several ways: through weighed vector of keywords (weights usually represent the relevance of each keyword), topic hierarchies (each node in the hierarchy represents a topic of interest, defined by a set of representative words) and topic ontology (an human-understandable, but machine-readable format).
- Goals and Tasks represent the immediate purpose or objective for a user's work within the system, usually being a learning goal in educational AH systems. The goal or task is almost always changing from session to session or even several times within one single session, being the most changeable user feature. Plan recognition is not trivial but it is a useful technique that helps discovering the user's intention by observing the set of input tasks and inferring the next (most probable) one and, hence, the user's goal. Goals or tasks can be represented in the user model with the same patterns of the knowledge, but also as a Bayesian network where nodes represent user tasks and arcs represent probabilistic dependencies between them.
- Background and Skills are a set of features related to the user's previous experience outside the core domain of the system (e.g., profession, work experience or language) and they are used most frequently for content adaptation. They are very similar to user's knowledge and they can be also stored as binary or scalar form, but their update is typically explicitly (requesting

users themselves or their superior – e.g., teacher, boss – to fulfill their background and skills).

- Behavior represents a repetitive pattern that can be used to assist the user. The modeled behavior type depends on the application domain: some systems help users to schedule their day (by learning rules and suggesting meetings based on them), buying things (by recording preferences) and learning (by recording the period of time that an action occurs), for instance.
- Interaction preferences refer to the users' interaction habits and preferences when they interact with an interface. The AH system should infer the preferred agent action (e.g., warning, suggestion or interruption of the work) and modality for different situations or contexts. For example, one student is creating a study group to review the content on Saturday, but another student involved probably will not attend because he never attended commitments on Saturday. Some students would prefer a simple warning, while others would want suggestions about an alternative date. To learn users' interaction preferences, the AH system must store information about the necessary situations and the intervention type.
- Individual traits define who the user is as an individual, including demographic information (e.g., sex, age, marital status, country, number of children), personality traits (e.g., introvert/extravert), cognitive styles (e.g., holist, serialist), cognitive factors (e.g., working memory capacity) and learning styles.
- Context involves any information that can be used to characterize the person, place or object which is relevant to the interaction between the user and the AH system, including the user and the system themselves. Context-aware systems provide rel-

evant information based on the context and the adaptation is guided by the user needs.

The process of acquiring information to fulfill the user model involves capturing the appropriate data, selecting the relevant information and interpreting the user interactions activities (Koch, 2001). This acquisition can be explicitly (through the input of data via forms provided for this purpose) or implicitly (observing actions). The problem of explicitly acquisition is that users are not willing to fill in long forms providing information about them and, if they do, it may not be the truth about them (intentionally or deliberately). Based on data stored in the user model, the AH system can perform the adaptation with many different methods and techniques.

METHODS AND TECHNIQUES

Many techniques and methods can be used to adapt hypermedia systems. While techniques are a part of the implementation level (being characterized by a specific kind of knowledge representation and by a specific adaptation algorithm), methods are part of the contextual level (being generalizations of one or more techniques) (Brusilovsky, 1996). This section describes some existing methods and techniques divided into three levels: content, presentation and navigation.

Content Adaptation

Content level enables the AH system to modify content, showing basic or detailed information about a topic according to users' knowledge, for example. The content adaptation aims to increase the system usability for a group of users with different knowledge, background and goals. This adaptation consists essentially in providing additional, comparative or alternative content, also hiding the irrelevant parts of content for each user (Gasparini, 2003). Content adaptation methods

can be divided into: Additional Content, Explanation Variant and Sorting, described below based on Brusilovsky (1996) and Kobsa et al. (2001).

- **Additional Content:** Shows only the relevant fragments based on goals, interests, preferences or levels of knowledge of each user. It is the most widely used method to adapt the content, being divided into:
 - **Additional Explanation:** Hides irrelevant fragments to the current level of knowledge of the user (i.e., inability to understand) or in agreement with user interest.
 - **Prerequisite Explanation:** Modifies the fragments based on prerequisites among related concepts and on user's knowledge. AH system considers that the user knows about the related concepts (i.e., prerequisites) if the user has already accessed each one.
 - **Comparative Explanation:** Presents a comparative explanation (defining similarities and differences) between the presented concept and related ones. This method is based on similarity links between concepts.
- **Explanation Variant or Variant Content:** Stores variants of the fragments and shows only the ones considered more appropriate for each user. This method is usually used when showing or hiding fragments is not enough, since users may need information fundamentally different.
- **Sorting:** Places the fragments in a hierarchical order and the most relevant fragments are displayed first, based on the user's background and knowledge.

Content adaptation techniques are described below based on Brusilovsky (1996), Kobsa et al. (2001) and Knutov et al. (2009):

- **Conditional Text:** Divides all information into fragments and each fragment is associated with one or more conditions repre-

sented in the user model (if conditions are satisfied, the content is presented). This is used to implement the Additional Content and Explanation Variant methods.

- **Adaptive Stretchtext:** Shows or hides all fragments related with a specific "hotword". It replaces the hotword with the fragments if they are important and relevant, based on the user's knowledge and preferences. Otherwise, fragments will be collapsed and the user will only see the hotword. This technique is used to implement the Additional Content method.
- **Page Variant:** Stores two or more variants of a page to present differently the same concept. This technique is used to implement the Explanation Variant method.
- **Fragment Variant:** Stores two or more variants of a fragment to be combined during the presentation of a single page, in a way that best suits the user model. This technique is considered more complex to implement than the Page Variant technique, since the page needs to be dynamically created with the predefined fragments. This technique is used to implement the Additional Content and Explanation Variant methods.
- **Fragment Coloring:** Presents the same content while highlights the fragments of information according to their classification (e.g., very relevant, relevant, less relevant and irrelevant) for each user. This technique is used when all users must view the same content (sorting method) because it has a relatively low adaptation.
- **Frame-Based Approach:** Represents all information about a particular concept in form of a frame. These slots of a frame can contain several explanation variants of the concept, links to other frames and examples that are shown, hidden and reordered based on the user model. This technique is used to implement the Additional Content, Explanation Variant and Sorting methods.

Presentation Adaptation

Presentation level enables the AH system to present more or less navigational functionalities based on user's knowledge, background and skills. The presentation adaptation aims to increase the system usability for users, by making a visual change (layout) or textual change (language) (Gasparini, 2003). Presentation adaptation methods can be divided into: Multi-language and Layout Variant, described below based on Koch (2001).

- **Multi-Language:** Changes the system language according to the user's preference, which may also depend on the context.
- **Layout Variant:** Creates several alternatives for required items of interface: color, font size, font style, size, maximum image size, text orientation.

Presentation adaptation techniques to implement the presentation adaptation methods are almost the same techniques used for adapt the content, except for the adaptive stretchtext and the fragment coloring. The Conditional Text, Page Variant, Fragment Variant and Frame-based Approach techniques can be used to implement Multi-language and Layout Variant methods. Beyond these techniques, there is the Styleguiding technique, which defines different styleguides for the same page, to implement the Layout Variant method (Koch, 2001).

Navigation Adaptation

Adaptive navigation helps users to find their paths in hyperspace by modifying the presentation of links (Brusilovsky, 1996). A link is a visible and "clickable" representation of related pages for which the user can navigate, being classified by its kind: Local Non-contextual Links, Contextual Links, Links from Index or Content Pages and Links on Local or Global Hyperspace Maps.

Navigation adaptation methods can be divided into: Local Guidance, Global Guidance, Local Orientation, Global Orientation and Personalized Views, described below based on Brusilovsky (1996).

- **Local Guidance:** Helps users to continue the navigation, suggesting the best link(s) from the current page, based on users' preferences, knowledge and background.
- **Global Guidance:** Helps users to find the shortest path to get to what they are looking for, through the suggestion of the most appropriate links from the current page, for example.
- **Local Orientation:** Helps users to understand the different possibilities of navigation and to follow the most appropriate link from a given position.
- **Global Orientation:** Helps users to understand the structure of the entire hyperspace and their position within this structure.
- **Personalized Views:** Generates and updates the hyperspace with a personalized view for the user, building a working interface in a personalized way through adaptation.

Navigation adaptation techniques can be divided into: Direct Guidance, Sorting of Links, Link Hiding, Link Removal, Link Disabling, Link Annotation and Map Adaptation, described below based on Brusilovsky (1996), De Bra & Calvi (1998) and Knutov et al. (2009).

- **Direct Guidance:** Chooses the best page to continue the navigation. This is the most flexible technique, since it is possible to recommend a link that is not directly related to the actual one.
- **Sorting of Links:** Classifies and orders the available links, placing the most relevant links first. Its usage can reduce the naviga-

tion time, but it also can confuse users who reminds the link location, since the AH system reorders the links on each access.

- **Link Hiding:** Restricts the navigation by hiding irrelevant links. This is one of the most used techniques to adapt the navigation because it can be used with all kinds of links described above, besides reducing the workload of the user.
- **Link Removal:** Restricts the navigation by removing irrelevant links, instead of hiding them. This technique is usually combined with the Sorting of Links technique, which defines the most relevant links to be shown while the others are removed.
- **Link Disabling:** Restricts the navigation by disabling the functionality irrelevant links (i.e., removing anchors of links). In this case, the link is still visible to the user, but clicking on it will have no effect.
- **Link Annotation:** Adds comments to the link, allowing users to know more about its actual state. This comment can be textual or visual (using icons, colors or different font size) and it can classify links in several ways (e.g., visited/not visited, very relevant/relevant/irrelevant, not ready to be learned/ready to be learned/learning/learned) instead of only two (relevant/not relevant).
- **Map Adaptation:** Is a combination of all other techniques listed above, being applied to a graphical presentation of the navigational structure (i.e., map). It can be used to adapt the map without changing its structure, while reducing the browsing space by elimination anchors and guiding users in order to prevent them to get lost in the hyperspace.

There are several applications that use these methods and techniques to perform the adaptation of the system, described in the following section.

APPLICATIONS

Focusing on Educational Hypermedia, one of the first web-based AH systems created was the ELM-ART (acronym for ELM-Adaptive Remote Training), which was originally implemented to help students in learning LISP programming language (Brusilovsky et al., 1996). Even with the positive results of several empirical studies, ELM-ART approach received few expansions since its original version had many limitations (e.g., it can only focus on beginners due the high complexity of advanced concepts compared with the benefits of the possible learning support and it allows only the authors to add new content and activities) (Weber & Brusilovsky, 2015). Another Educational AH system is Interbook, created as an attempt to refine and improve some features of ELM-ART. Interbook uses users' knowledge to provide adaptive guidance (via direct guidance) and adaptive navigation (via link annotation) (Brusilovsky et al., 1998).

One of the most famous AH system is AHA! (Adaptive Hypermedia Architecture), which is an open-source and web-based system originally created for educational area (De Bra et al., 2006) that likewise distilled some of the ELM-ART limitations. AHA! uses the AHAM reference model and some methods to adapt its content (e.g., additional content) and presentation (e.g., layout variant), in addition to some techniques to adapt the navigation (e.g., link hiding, link removal and link annotation) (De Bra & Calvi, 1998) (De Bra et al., 2003).

The AdaptWeb® system is an Educational AH system focused on distance education. The AdaptWeb® architecture is based on four modules: authoring, storage, content adaptation and adaptive interface modules. The authoring module structures and organizes the instructional content via an authoring tool (creating a structure of concepts, establishing prerequisites criteria and associating files, examples, exercises and complementary ma-

terials for each the concept). The storage module receives this structure and stores data in an XML file, which serves as a basis to generate files for the adaptation modules. The content adaptation and adaptive interface modules work together, adapting the content, presentation and navigation through the XML generated and the user model (De Oliveira et al., 2003). Based on user's educational level, AdaptWeb® applies some techniques such as fragment variants to adapt content and direct guidance and link removal, disabling and annotation to adapt navigation. Based on user's working environment, AdaptWeb® adapts content using conditional text technique and, based on user's knowledge, it adapts the navigation using link disabling and annotation techniques. AdaptWeb® has been constantly updated and improved. Recently, the system incorporated a recommender system, learning analytics and gamification techniques, which are some of the trends and future research directions described in the next section.

FUTURE RESEARCH DIRECTIONS

Although it is not a new area, many trends have been explored in recent years. A study conducted by Somyürek (2015) raised some of these trends, divided into seven categories: standardization, semantic web, modular frameworks, data mining, machine learning techniques, social web and device adaptation. Here we describe some of these trends described by (Somyürek 2015) and we identified another: gamification.

Standardization aims to develop a set of interoperable, reusable and accessible resources, besides facilitating the management of them in the educational context. Standardization in AH systems usually occurs at the metadata level, which are ways of describing learning objects and user model to enable the integration with other AH systems (Fischer, 2001). Some metadata standards currently used are: Dublin Core, IMS Learning Resource Metadata and IEEE Learning Object Metadata (LOM).

Data mining is used to construct an understandable structure by means of analyzing large quantities of data, being divided into clustering and classification. Clustering assigns students with similar characteristics in the same group, while classification determines which these characteristics are (e.g., interests, skills, cognitive and learning styles). These data are usually used to recommend learning objects more effectively to each characteristic or even to implement a recommender system (Somyürek, 2015).

Social web has become very popular in recent years since knowledge has become more personal, social, distributed and dynamic by nature. In addition, mobile devices (e.g., laptops, smartphones and tablets) are becoming increasingly available and, therefore, they can be used for educational purposes. Because of that, device adaptation is necessary to provide a better presentation of the content based on the students' devices and, also, on their skills (Somyürek, 2015). In this sense, the work of (Machado & Oliveira, 2014) proposes the adaptation and resource recommendation for students using mobile devices.

Gamification is a new approach that can be used to motivate and engage students with game elements and design in non-game contexts (Deterding et al., 2011). In the educational area, there are many initiatives involving the use of gamification (e.g., (Grant & Betts, 2013), (Ibanezet al., 2014) and (Hanus & Fox, 2015)). Although the theme is new, there are already initiatives in order to unite the AH systems and gamification, as the work of (Klock et al., 2015). This work is in development and aims to perform the adaptation of gamification based on the characteristics of the students, since not all students will be motivated by the same game elements.

CONCLUSION

Adaptive Hypermedia (AH) systems are able to adjust their status in accordance with the users' profile, behavior and context (Gasparini, 2013),

building a user model with various information about user characteristics, such as: skills, knowledge, needs, preferences, behavior, way of interaction with the system, etc. AH systems bring a new perspective and they can improve the user interaction with computer systems (Gasparini, 2013), being also an alternative to one-size-fits-all approach and, since they allow different students to use the same form of environment adapted to their profile, being useful in educational area.

Through the adaptation of hypermedia systems, it is possible to know the characteristics of students and facilitate a learning process centered on the user. In this chapter, it was exposed some techniques and methods used for adaptation and identified some of the main characteristics of students considered for such adaptation. Thus, it is possible to visualize the importance of AH systems when we analyze the students' learning difficulties. This area is expanding and it presents various trends such as standardization, data mining, social web, device adaptation and gamification.

REFERENCES

- Akscyn, R. M., McCracken, D. L., & Yoder, E. A. (1988). KMS: A Distributed Hypermedia System for Managing Knowledge in Organizations. *Communications of the ACM*, 31(7), 820–835. doi:10.1145/48511.48513
- Brusilovsky, P. (1996). Methods and techniques of adaptive hypermedia. *User Modeling and User-Adapted Interaction*, 6(2-3), 87–129. doi:10.1007/BF00143964
- Brusilovsky, P. (2001). Adaptive Hypermedia. *User Modeling and User-Adapted Interaction*, 11(1-2), 87–110. doi:10.1023/A:1011143116306
- Brusilovsky, P. (2007). Adaptive Navigation Support. In P. Brusilovsky, A. Kobsa, & W. Nejdl (Eds.), *The Adaptive Web: Methods and Strategies of Web Personalization* (pp. 263–290). Berlin: Springer-Verlag. doi:10.1007/978-3-540-72079-9_8
- Brusilovsky, P., Eklund, J., & Schwarz, E. (1998). Web-based education for all: a tool for development adaptive courseware. *Computer Networks and ISDN Systems*, 30(1–7), 291 – 300.
- Brusilovsky, P., Kobsa, A., & Vassileva, J. (Eds.). (1998). *Adaptive Hypertext and Hypermedia*. Dordrecht: Springer Netherlands. doi:10.1007/978-94-017-0617-9
- Brusilovsky, P., & Millán, E. (2007). User Models for Adaptive Hypermedia and Adaptive Educational Systems. In P. Brusilovsky, A. Kobsa, & W. Nejdl (Eds.), *The Adaptive Web: Methods and Strategies of Web Personalization* (pp. 3–53). Berlin: Springer Berlin Heidelberg. doi:10.1007/978-3-540-72079-9_1
- Brusilovsky, P., Schwarz, E., & Weber, G. (1996). ELM-ART: An intelligent tutoring system on World Wide Web. In *Proceedings of the 3rd International Conference on Intelligent Tutoring Systems* (pp. 261–269). Springer. doi:10.1007/3-540-61327-7_123
- Bunt, A., Carenini, G., & Conati, C. (2007). Adaptive Content Presentation for the Web. In P. Brusilovsky, A. Kobsa, & W. Nejdl (Eds.), *The Adaptive Web: Methods and Strategies of Web Personalization* (pp. 409–432). Berlin: Springer Berlin Heidelberg. doi:10.1007/978-3-540-72079-9_13
- De Bra, P., Aerts, A., Berden, B., de Lange, B., Rousseau, B., & Santic, T., ... Stash, N. (2003). AHA! The Adaptive Hypermedia Architecture. In *Proceedings of the Fourteenth ACM Conference on Hypertext and Hypermedia* (pp. 81–84). New York: ACM. doi:10.1145/900051.900068

- De Bra, P., & Calvi, L. (1998). AHA: a generic adaptive hypermedia system. *Proceedings of the 2nd Workshop on Adaptive Hypertext and Hypermedia*, 5–12.
- De Bra, P., Smits, D., & Stash, N. (2006). Creating and Delivering Adaptive Courses with AHA! In *Innovative Approaches for Learning and Knowledge Sharing* (Vol. 4227, pp. 21–33). Springer Berlin Heidelberg. doi:10.1007/11876663_4
- de Koch, N. P. (2001). *Software Engineering for Adaptive Hypermedia Systems-Reference Model, Modeling Techniques and Development Process*. Ludwig-Maximilians-Universität, Munique.
- de Oliveira, J. P. M., Muñoz, L. S., de Freitas, V., Marçal, V. P., Gasparini, I., & Amaral, M. A. (2003). AdaptWeb: an Adaptive Web-based Courseware. *III Annual ARIADNE Conference*.
- Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: defining gamification. In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments* (pp. 9–15). ACM. doi:10.1145/2181037.2181040
- Gasparini, I. (2003). *Interface Adaptativa no ambiente AdaptWeb: navegação e apresentação adaptativa baseada no modelo do usuário (Dissertação)*. Porto Alegre: Universidade Federal do Rio Grande do Sul.
- Gasparini, I. (2013). *Aspectos Culturais no Modelo do Usuário em Sistemas Adaptativos Educacionais: Fundamentos, Proposta e Experimentação (Tese)*. Porto Alegre: Universidade Federal do Rio Grande do Sul.
- Grant, S., & Betts, B. (2013). Encouraging User Behaviour with Achievements: An Empirical Study. In *Proceedings of the 10th Working Conference on Mining Software Repositories* (pp. 65–68). Piscataway, NJ: IEEE Press. doi:10.1109/MSR.2013.6624007
- Hanus, M. D., & Fox, J. (2015). Assessing the effects of gamification in the classroom: A longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance. *Computers & Education*, 80, 152–161. doi:10.1016/j.compedu.2014.08.019
- Ibanez, M.-B., Di-Serio, A., & Delgado-Kloos, C. (2014). Gamification for Engaging Computer Science Students in Learning Activities: A Case Study. *IEEE Transactions on Learning Technologies*, 7(3), 291–301. doi:10.1109/TLT.2014.2329293
- Klock, A. C. T., da Cunha, L. F., de Carvalho, M. F., Rosa, B. E., Anton, A. J., & Gasparini, I. (2015). Gamification in e-Learning Systems: A Conceptual Model to Engage Students and Its Application in an Adaptive e-Learning System. In *Learning and Collaboration Technologies* (pp. 595–607). Springer International Publishing.
- Knutov, E., De Bra, P., & Pechenizkiy, M. (2009). AH 12 Years Later: A Comprehensive Survey of Adaptive Hypermedia Methods and Techniques. *New Review of Hypermedia and Multimedia*, 15(1), 5–38. doi:10.1080/13614560902801608
- Kobsa, A., Koenemann, J., & Pohl, W. (2001). Personalised hypermedia presentation techniques for improving online customer relationships. *The Knowledge Engineering Review*, 16(2), 111–155. doi:10.1017/S0269888901000108
- Machado, G. M., & de Oliveira, J. P. M. (2014). Context-aware adaptive recommendation of resources for mobile users in a university campus. *2014 IEEE 10th International Conference on Wireless and Mobile Computing, Networking and Communications*, 427–433. doi:10.1109/WiMOB.2014.6962206
- Schiaffino, S., & Amandi, A. (2009). Intelligent User Profiling. In M. Bramer (Ed.), *Artificial Intelligence An International Perspective* (Vol. 5640, pp. 193–216). Springer Berlin Heidelberg. doi:10.1007/978-3-642-03226-4_11

Shute, V.J., & Zapata-Rivera, D. (2012). Adaptive educational systems. In *Adaptive technologies for training and education* (pp. 7–27). Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9781139049580.004

Somyürek, S. (2015). The new trends in adaptive educational hypermedia systems. *The International Review of Research in Open and Distributed Learning*, 16(1). doi:10.19173/irrodl.v16i1.1946

Tseng, S.-S., Su, J.-M., Hwang, G.-J., Hwang, G.-H., Tsai, C.-C., & Tsai, C.-J. (2008). An object-oriented course framework for developing adaptive learning systems. *Journal of Educational Technology & Society*, 11(2), 171–191.

Turel, V. (2015). Hypermedia and its Role in Learning. In M. Khosrow-Pour (Ed.), *Encyclopedia of Information Science and Technology* (3rd ed.; pp. 2495–2505). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-5888-2.ch243

Weber, G., & Brusilovsky, P. (2015). ELM-ART – An Interactive and Intelligent Web-Based Electronic Textbook. *International Journal of Artificial Intelligence in Education*.

ADDITIONAL READING

Chatti, M. A., Jarke, M., & Specht, M. (2010). The 3P learning model. *Journal of Educational Technology & Society*, 13(4), 74–85.

Dey, A. K. (2001). Understanding and Using Context. *Personal and Ubiquitous Computing*, 5(1), 4–7. doi:10.1007/s007790170019

Fischer, S. (2001). Course and Exercise Sequencing Using Metadata in Adaptive Hypermedia Learning Systems. *Journal of Educational Resources in Computing*, 1(1), 1–21.

Gasparini, I., Pimenta, M. S., Palazzo, M., de Oliveira, J., & Bouzeghoub, A. (2010). Combining Ontologies and Scenarios for Context-aware e-Learning Environments. In *Proceedings of the 28th ACM International Conference on Design of Communication* (pp. 229–236). New York: ACM. doi:10.1145/1878450.1878489

Pernas, A. M., Gasparini, I., Pires, L. V., Lopes, J. L. B., Yamin, A. C., & de Oliveira, J. P. M. (2015). Integrated Infrastructure for Ubiquitous Learning. *IEEE Revista Iberoamericana de Tecnologías Del Aprendizaje*, 10(3), 92–99. doi:10.1109/RITA.2015.2452533

Weber, G. (1996). Episodic learner modeling. *Cognitive Science*, 20(2), 195–236. doi:10.1207/s15516709cog2002_2

Weber, G., & Brusilovsky, P. (2001). ELM-ART: An adaptive versatile system for Web-based instruction. [IJAIED]. *International Journal of Artificial Intelligence in Education*, 12, 351–384.

KEY TERMS AND DEFINITIONS

Adaptable Hypermedia System: A hypermedia system that allows the user to customize its preferences in a non-automatic way.

Contextual Links or Real Hypertext Links: Links that can be embedded in the context of the page (e.g., using “hotwords” in texts, “hot spots” in images), but they cannot be removed. They are not easy to manipulate as the local non-contextual links.

Links from Index or Content Pages: Appear on pages that contain only links presented in a fixed order (by alphabetical order for index pages and by content for content pages). These links are usually easy to manipulate, unless the page is implemented as an image.

Links on Local or Global Hyperspace

Maps: Appear on maps that represent a local area or even an entire hyperspace as a network of nodes connected by arrows. Using these maps, users can directly navigate to all visible nodes just by clicking on the representation of the desired node.

Local Non-Contextual Links: Are links independent from the content, being easy to manipulate and appearing as a set of buttons, links or a pop-up menu.

Group Synchronization for Multimedia Systems



Dimitris N. Kanellopoulos
University of Patras, Greece

INTRODUCTION

An important feature of multimedia applications is the integration of multiple media streams that have to be presented in a synchronized fashion (Li *et al.*, 1997). Synchronization is mainly the preservation of the temporal constraints within and among multimedia data streams at the time of playout. Temporal relations define the temporal dependencies between media objects (Blakowski & Steinmetz, 1996). An example of a temporal relation is the relation between a video and an audio object that are recorded during a concert. If these objects are presented, the temporal relation during the presentations of the two media objects must correspond to the temporal relation at the time of recording. Discrete media like text, graphics, and images are time-independent media objects, while the semantic of their respective content does not depend upon a presentation to the time domain. A discrete media object is frequently presented using one presentation unit. On the contrary, a time-dependent media object is presented as a media stream. In a continuous media stream (e.g. video), the presentation durations of all units of a time-dependent media object are equal (Li *et al.*, 1997). For example, a video consists of a number of ordered frames. Each of these frames has fixed presentation duration.

Most of the components of a multimedia system support and address temporal synchronization. These components include the operating system, communication subsystem, databases, documents and even applications. In distributed multimedia systems, networks introduce random delays in the delivery of multimedia information, and there

are four sources of asynchrony that can disrupt synchronization (Akyildiz & Yen, 1996):

- Delay Jitter and Local Clock Drift.
- Different Initial Collection Times. Let us consider two media sources, one providing voice and the other video. If these sources start to collect their media units (MUs) at different times, the playback of the MUs of voice and video from the two sources at the receiver loses semantic meaning.
- Different Initial Playback Times. If the initial playback times are different for each user, then asynchrony will arise.

To ensure synchronized delivery of multimedia information, intelligent synchronization protocols/techniques are required. This article illustrates various issues on intra- and inter-media synchronization and presents the basic schemes for inter-destination media synchronization (IDMS). It presents IDMS standardization efforts and novel solutions for new multimedia applications. Finally, it outlines future research directions for IDMS.

BACKGROUND

Intra-Media Synchronization

Intra-media synchronization refers to the time relations between the presentation units of one time-dependent media object. An example is the time relations between the single frames of a video sequence. The spacing between subsequent frames is dictated by the frame production rate. Jitter

may destroy the temporal relationships between periodically transmitted MUs that constitute a real-time media stream, thus hindering the comprehension of the stream. *Playout adaptation algorithms* undertake the labor of the temporal reconstruction of the stream, which is referred to as the *restoration of its intra-stream synchronization quality* (Park & Choi, 1996). *Adaptive media playout* (AMP) improves the media synchronization quality of streaming applications by regulating the playout time interval among MUs at a receiver. To mitigate the effect of the jitter, MUs have to be delayed at the receiver in order a continuous synchronized presentation can be guaranteed. Therefore, MUs have to be stored in a buffer and the size of this buffer will correspond to the amount of jitter in the network. As the synchronization requirements can vary according to the multimedia application on hand, we must control the individual synchronization (delay sensitivity, error tolerance etc.) for each media separately. To this direction, Park and Choi (1996) investigated an efficient multimedia synchronization method that can be applied at intra-media synchronization in a consistent manner. They proposed an adaptive synchronization scheme, based on the delay offset and playout rate adjustment that can match the application's varying synchronization requirements effectively. Park and Kim (2008) introduced an AMP scheme based on a discontinuity model for intra-media synchronization of video applications over the best-effort networks. They analyzed the temporal distortion (i.e., discontinuity) cases, such as playout pause and skip, to define a unified discontinuity model. Finally, Laoutaris and Stavrakakis (2002) surveyed the work in the area of playout adaptation.

Inter-Media Synchronization

Inter-media synchronization refers to the synchronization between media objects of a multimedia object. An example of inter-media synchronization is the *Lip Synchronization* that refers to the temporal relationship between an audio and video stream for the particular case of human speak-

ing (Aggarwal & Jindal, 2008). There are many systematic specification methods to describe synchronization problems. Blakowski and Steinmetz (1996) illustrated the main synchronization specification methods: interval-based specifications, control flow-based specification, axes-based synchronization, event-based synchronization, scripts, and comments. A *Synchronization Specification* of a multimedia object can describe all temporal dependencies of the included objects in a multimedia object. It is comprised of:

- Intra-object synchronization specifications for the media objects of the presentation.
- Quality of Service (QoS) descriptions for intra-object synchronization.
- Inter-object synchronization specifications for media objects of the presentation.
- QoS descriptions for inter-object synchronization.

To achieve inter-media synchronization various algorithms have been applied. There are several types of synchronization control such as *Skipping* (Ishibashi et al., 2002a), *Buffering* (Ishibashi et al., 2002a), *Adaptive Buffer Control* (ABC) (Wongwirat & Ohara, 2006), *Queue Monitoring* (QM) (Hikichi et al., 2002), *Virtual-Time Rendering* (VTR) (Ishibashi et al., 2002b), and *media adaptive buffering* (Isomura et al., 2011). Boronat et al. (2009) have reviewed the most powerful inter-media synchronization algorithms. The building blocks of these algorithms are the synchronization techniques utilized both at the sender and receiver sides. These algorithms can use multiple of synchronization techniques to achieve the synchronization mechanism even from different categories (Din & Bulterman, 2012).

Classification of Inter-Media Techniques

Synchronization techniques can be categorized according to the '*location*', '*purpose*', '*content*', and '*information used*' (Boronat et al., 2009).

- Location of Synchronization Technique:**
 The synchronization control can be performed either by source or receiver. If control is performed by the source, most of the time it will require some feedback information from the receiver. The receiver will tell the source about the degree of asynchrony at the current instance.
- Live vs. Synthetic Synchronization (Type of Media):** In *live* media, the temporal relations are exactly reproduced at a presentation, as they existed during the capture process. Synthetic synchronization techniques are used for *stored* media.
- Information Used for Synchronization Technique:** The information included in the MU can be different like timestamp, sequence number. Some techniques use either sequence number or timestamp, while others may use both.
- Purpose of Synchronization Technique:**
 The techniques can be divided into four subcategories with respect to its purpose:
 - The basic control techniques, which are required in almost all the algorithms. Examples are the attachment of synchronization information in MUs at source and buffering of MUs at receiver.
 - The common control techniques, which can be applied in both ways.
 - The preventive control techniques, which are used to prevent the asynchrony in the streams.
 - The reactive control (or corrective) techniques are designed to recover synchronization in the presence of synchronization errors.

Table 1 shows a classification of inter-media techniques, based on the above criteria.

Table 1. Classification of inter-media techniques

Technique	Location	Description
Basic Control	Source control	Add information useful for synchronization: timestamps, sequence numbers (identifiers), event information and/or source identifiers.
	Receiver control	Buffering techniques
Common Control	Source control	- Skip or pause MUs in the transmission process. - Advance the transmission timing dynamically. - Adjust the input rate. - Media Scaling.
	Receiver control	- Adjust the playout rate. - Data interpolation.
Preventive Control	Source control	- Initial playout instant calculation. - Deadline-based transmission scheduling. - Interleave MUs of different media streams in only one transport stream.
	Receiver control	- Preventive skips of MUs (e.g. discardings) and/or preventive pauses of MUs (repetitions, insertions or stops). - Change the buffering waiting time of the MUs. - Enlarge or shorten the silence periods of the streams.
Reactive Control	Source control	- Adjust the transmission timing. - Decrease the media streams transmitted. - Drop low-priority MUs.
	Receiver control	- Reactive skips (eliminations or discardings) and/or reactive pauses (repetitions, insertions or stops). - Make playout duration extensions or reductions (playout rate adjustments). - Use of a virtual time with contractions or expansions. - Master/slave scheme - Late event discarding (Event-based). - Rollback techniques (Event-based)

MAIN FOCUS OF THE ARTICLE

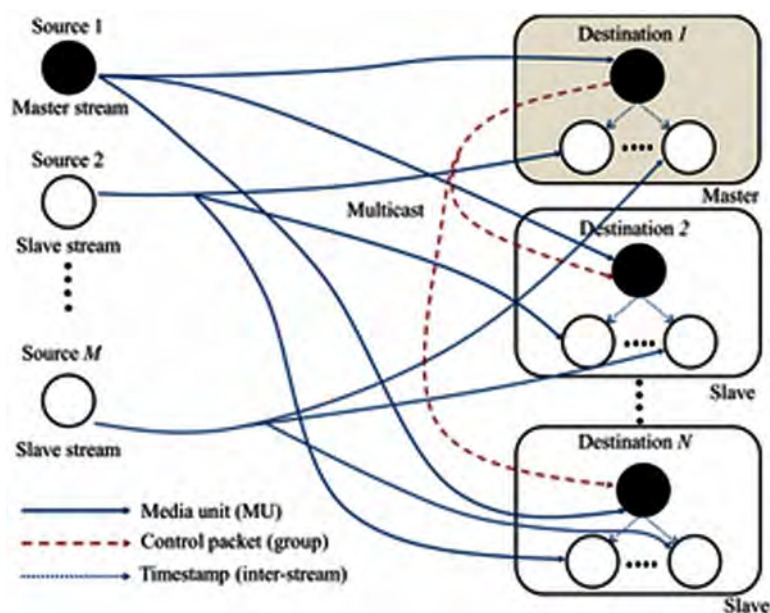
Group or inter-destination media synchronization (IDMS) addresses the presentation of a stream at all the receivers of a group, simultaneously. For example, in real-time distance learning, tutor multicasts a multimedia lesson to a group of students located at different areas. The levels of required synchrony among the receivers depend on the application on hand. Akyildiz and Yen (1996) introduced group synchronization protocols for real-time, multimedia applications, including teleconference and multimedia on demand services. Their protocols achieve synchronization for all configurations (one-to-one, one-to-many, many-to-one, and many-to-many), and do so without prior knowledge of the end-to-end delay distribution, or the distribution of the clock drift. The only a-priori knowledge the protocols require is an upper bound on the end-to-end delay. Recently, Boronat *et al.* (2009) reviewed the most popular multimedia group and inter-stream synchronization approaches. Group synchronization techniques can be classified at three schemes. These schemes are based on the

Virtual-Time Rendering (VTR) media synchronization algorithm to determine the output timing of each MU, so that the timing can be the same at all the destinations. For better understanding of these schemes, consider that M sources and N destinations/receivers are connected through a network. MUs of M different streams have been stored with timestamps in M sources, and they are broadcasted to all the receivers. The timestamp contained in a MU indicates its generation time. The streams fall into a master stream and slave streams. At each receiver, for inter-media synchronization, the slave streams are synchronized with the master stream by using an inter-media synchronization mechanism.

Master/Slave Receiver Scheme (MSRS)

In MSRS (Ishibashi *et al.*, 1997), the receivers are categorized into one Master receiver and Slave receivers. Multiple streams are received at each receiver and one of these streams acts as master stream in order inter-media synchronization to be achieved at each receiver (Figure 1).

Figure 1. Master/slave receiver scheme
Sithu and Ishibashi, 2015.



None of the slave receivers send any feedback information about the timing of the playout processes. It adjusts the playout timing of MUs to that of the master receiver. Only the master receiver sends (multicasts) its playout timing to all the other (slave) receivers. The master receiver controls and computes the presentation time of the MUs according to its own state of the received stream data. Group synchronization is achieved by adjusting the presentation time of the MUs of master stream at the slave receivers to that of the master receiver. Therefore, the slave receivers should present MUs at the same timing as the master receiver. The synchronization of the slave receivers is achieved as follows:

- The master receiver multicasts a control packet to all slave receivers. This control packet includes the presentation time of its first MU of master stream. This process is called “initial presentation adjustment”.
- When the target presentation time of the master receiver changes, the master receiver notifies all the slaves about this modification by multicasting a control packet. This control packet contains the amount of time that is modified and the sequence number of the MU for which the target presentation time has been changed.
- The master receiver periodically multicasts proper control packets to accommodate the newly joined slave receivers.

Boronat *et al.* (2008) presented MSRS by extending the RTP/RTCP (Real-time Transport Protocol/RTP Control Protocol) messages for containing the synchronization information. Figure 1 presents the different type of message exchanges in the basic MSRS. The advantage of the MSRS technique is its simplicity and decreased amount of information exchange (i.e. control packets) to support group synchronization. However, the selection of the master receiver can influence the performance of the scheme. If the faster receiver is selected as the master receiver, it can cause the

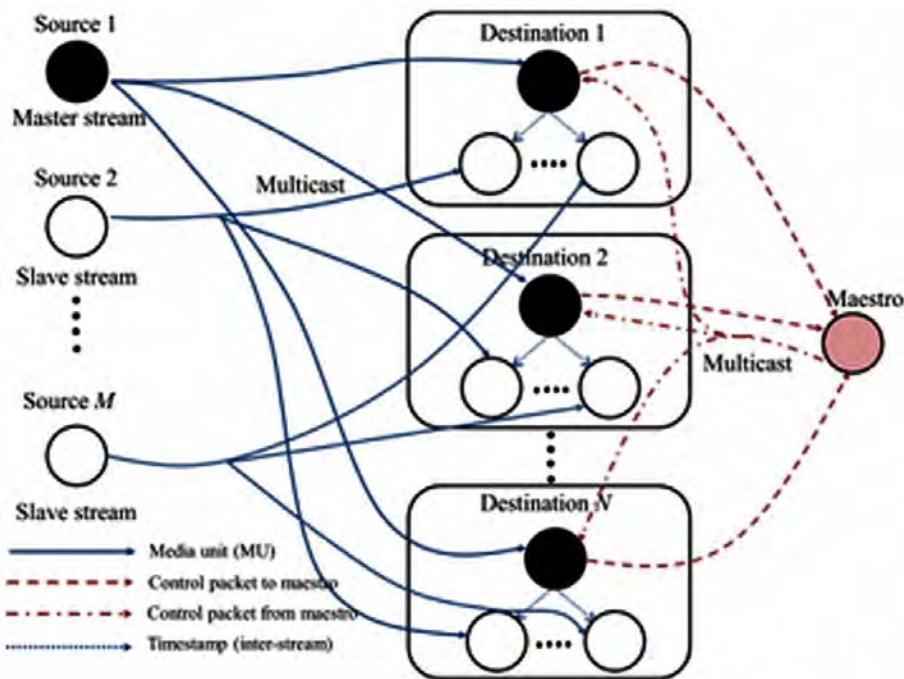
Buffer underflow in the (slower) slave receivers, which can result as the poor presentation quality at slow receivers. On the contrary, if the slowest receiver is selected as master, it can cause *Buffer overflow* on fast slave receiver(s). This will result as high packet drops at faster slave receiver(s). A problem with the MSRS technique is that the master can act as a bottleneck in the system. A second problem deals with the associated degree of unfairness with the slave receivers. Boronat *et al.* (2011) discussed possible options with pros and cons for the master selection in this scheme.

Synchronization Maestro Scheme (SMS)

In SMS scheme, all the receivers are handled fairly as there are not master and slaves (Ishibashi & Tasaka 1997). SMS involves a *Synchronization Manager* (SM), which can be performed by one of the source or receiver. For example, in Figure 2, one receiver/destination performs the SM.

Each receiver estimates the network delay and uses the estimates to determine the local presentation time of the MU. Then, each receiver sends this estimated presentation time of MU to the SM. After that, the SM gathers the estimates from the receivers, and adjusts the presentation timing among the receivers by multicasting control packets to receivers. The SMS scheme assumes that clock speed at the sources and receivers is the same, and that the current local times are also the same (i.e., globally synchronized clocks). Figure 2 depicts the basic principle of the SMS technique. Boronat *et al.* (2009) presented the RTCP based schemes, which follow the same basic principle. The SMS scheme (like the MSRS) is a centralized solution, and thus it can confront the bottleneck problem. The advantage of the SMS scheme over MSRS is its fairness to the receivers, as the feedback information of all the receivers is accounted for determining the presentation time of the MU. However, this fairness costs more communication overhead among the receiver and the synchronization manager.

Figure 2. Synchronization maestro scheme
Sithu and Ishibashi, 2015.



Distributed Control Scheme (DCS)

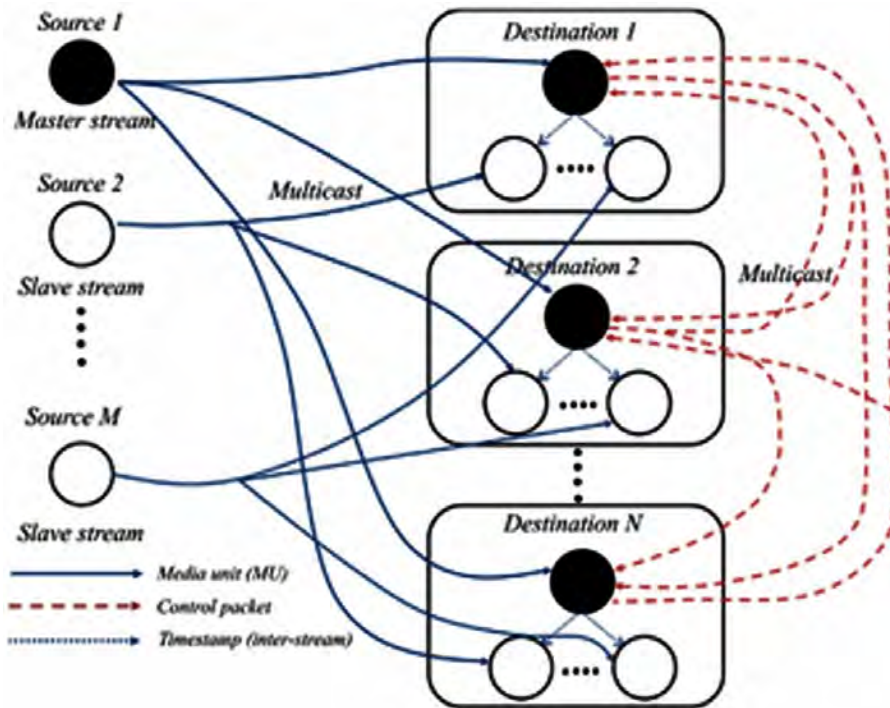
Figure 3 illustrates the DCS scheme (Mauve *et al.*, 2004). Each receiver estimates the network delay, and then determines the presentation time of the MU. Then, it sends (multicasts) this presentation time to all the receivers. After that, every receiver will have the entire view of the estimated time of MU. Each receiver has the flexibility to decide the reference play out time among the timing of all the receivers. The DCS scheme provides higher flexibility to each receiver to decide the presentation time of MU. For example, it is possible that by selecting the presentation time of other receiver, it can achieve higher group synchronization quality, but it may cause the inter-media or intra-media synchronization degradation. In this case, the receiver has the flexibility to choose between the types of synchronization depending upon the nature of application on hand. If the application on hand requires the higher inter-media or intra-media synchronization and can sacrifice on the group synchronization to certain limit, then

receiver can select its own determined presentation time and vice versa. DCS is a distributed scheme by nature and does not suffer from the bottleneck problem. If one or more receivers leave the system, it will not disturb the overall scheme. This greater flexibility and the distributed nature of DCS make it complex in terms of processing. This happens because the receiver does more calculations and comparisons before deciding the presentation time of MU. Finally, DCS has higher message complexity, as each receiver multicasts the estimated presentation time.

SOLUTIONS AND RECOMMENDATIONS

Standardization of IDMS has been carried out within ETSI (European Telecommunications Standards Institute) TISPAN (Telecoms & Internet converged Services & Protocols for Advanced Networking). In addition, such standardization is a highlight for the IETF AVTCORE WG (In-

Figure 3. Distributed control scheme
Sithu and Ishibashi, 2015.



Internet Engineering Task Force - Audio/Video Transport Core Maintenance Working Group). The specification (ETSI, 2009) does pose IDMS and the synchronization of media streams from different sources as a requirement for providing synchronization-sensitive interactive services. These use cases are mostly in the categories of low or medium synchronization, and not very high requirements are posed to delay differences between various user equipments. The protocol specification gives a delay difference of between 150 and 400 ms as a guideline for achieving transparent interactivity, based on ITU guidelines for interactivity in person-to-person communication. ETSI TISPAN has done the first work on standardizing RTCP usage for IDMS. The ETSI proposal is a dedicated solution for use in large-scale IPTV deployments, with low to medium level synchronization requirements. Other services such as Internet-based video services may also benefit from IDMS, and other use cases require higher levels of synchronization, and are not supported by

the ETSI solution (Montagud et al., 2012). Within the Internet Engineering Task Force (IETF), the AVTCORE working group (Brandenburg et al., 2012) carries out standardization of the RTCP-based IDMS protocol. This is the core group that is responsible for the RTP and accompanying RTCP protocol. Most RTCP extensions are developed within the IETF. Boronat et al. (2009) described most IDMS solutions that define new proprietary protocols, with specific control messages, that should increase the network load. Montagud et al. (2012) presented up to 19 use cases for IDMS, each one having its own synchronization requirements.

Presentation adaptation and flow control is a feasible solution to achieve a smooth multiple-stream distributed multimedia presentation. Huang et al. (2001) proposed the *Pause-And-Run* approach for k -stream (PARK) multimedia presentations over Internet to achieve reliable transmission of continuous media. They evaluated the application of the PARK approach over the Internet. The evaluation results revealed a suitable

buffering control policy for the audio and video media, respectively. The characteristics of the PARK approach are:

- PARK adopts TCP to achieve reliable transmission for continuous media.
- A novel flow adaptation scheme is proposed to reduce the overhead of the network and end-hosts as the slow-start scheme is embedded in TCP. The server adapts its transmission rates to the buffer situation of the client and prevents the client's buffers from overflow and underflow as much as possible.
- With the provision of multiple-stream synchronization and the multi-level adaptation control, the client achieves smooth multimedia presentations and graceful presentation degradation.

Realization of *Synchronous shared experiences* requires that users feel that they are *coherently* communicating with each other. Vaishnavi *et al.* (2011) analyzed challenges that need to be tackled to achieve coherence: *Quality of service, mobility and distributed media synchronization*. They presented their solution to distributed media synchronization. Their design uses the local lag mechanism over a distributed control or master-slave signaling architecture. Montagud and Boronat (2012) presented an evolved version of an RTCP-based IDMS approach, including an AMP scheme that adjusts the play-out timing of each one of the geographically distributed consumers in a specific cluster, if an allowable asynchrony threshold between their play-out states is exceeded. Montagud *et al.* (2014) presented an IDMS solution based on extending the capabilities of RTP/RTCP protocols. To enable an adaptive, highly accurate and standard compliant IDMS solution, they specified RTCP extensions, in combination with several control algorithms and adjustment techniques. Montagud *et al.* (2015) reviewed the existing sync reference models by examining the involved features, components and layers in

each one of them. Their study reflects the need for a new modular and extensible theoretical framework to efficiently comprehend the overall media sync research area. Belda *et al.* (2015) presented *Wersync*, a web-based platform that enables distributed media synchronization and social interaction across remote users. Costa and Santos (2014) surveyed the existing media sync solutions focused on the TV area, classifying them in terms of: types of involved devices, types of media content, types of sync techniques, targeted applications or scenarios, and evaluation methodologies.

Ishibashi *et al.* (2014) carried out QoE (Quality of Experience) assessment of fairness between players in a networked game with olfaction. They investigated the influence of the time it takes for a smell to reach a player on the fairness. They illustrated that the fairness is hardly damaged when the constant delays are smaller than about 500 ms. The used media synchronization algorithm considers the human perception of intra-stream and inter-stream synchronization errors. Ghinea and Ademoye (2010) conducted a perceptual measurement of the impact of a synchronization error between smell sensory data and audiovisual content, assuming the audiovisual lip skew is zero. Their results show a synchronization threshold of 30 s, when olfaction is ahead of audiovisual data, and of 20 s when olfaction is behind.

In *joint musical performance*, multiple users play their respective same or different types of musical instruments together, but the media synchronization quality and interactivity may seriously be deteriorated owing to the network delay. Sithu and Ishibashi (2015) proposed a new media synchronization control called the '*dynamic local lag control*'. By QoE assessment, they demonstrated that the new control can achieve high quality of media synchronization and keep the interactivity high.

Huang *et al.* (2013) have presented a historical view of temporal synchronization studies focusing on continuous multimedia. They demonstrated how the development of multimedia systems has

created new challenges for synchronization technologies. They concluded with a new application-dependent, multi-location, multi-requirement synchronization framework to address these new challenges.

FUTURE RESEARCH DIRECTIONS

Group synchronization techniques could benefit from the cross-layer optimization. Cross-layer optimization allows communication between OSI RM layers by permitting one layer to access the data of another layer to exchange information and enable interaction (Bin-Salem & Wan, 2012). It contributes to an improvement of quality of services under various operational conditions. The cross-layer control mechanism provides a feedback on concurrent quality information for the adaptive setting of control parameters of a multimedia system. Thus, it can help to the utilization of synchronization techniques (e.g. preventive control techniques). Soon, researchers will develop more comprehensive multimedia synchronization subsystems that will integrate preventive and reactive methods and will take into account the cross layer optimization and other components (the operating system, communication subsystem, databases, documents and applications). Such systems must be able not only to achieve intra- and inter-media synchronization, but also the IDMS to a diversity of new multimedia applications.

In RTP-based multimedia streaming services, client-driven media synchronization mechanisms must be developed to provide accurate media synchronization such as to reduce: the initial synchronization delay, the processing complexity at the client device, the number of required user datagram protocol ports, and the amount of control traffic injected into the network. Such a synchronization mechanism was recently proposed by Jung and Seo (2016). In their method, the server does not need to send any RTCP SR packets for synchronization. Instead, the client

device derives the precise normal play time for each video and audio stream from the received RTP packets containing an RTP timestamp.

CONCLUSION

This article has illustrated various issues on intra and inter-media synchronization. It has presented basic schemes of IDMS and has mentioned IDMS standardization efforts and novel solutions for new multimedia applications. Also, it has outlined future research directions on IDMS. Certainly, standardization of IDMS will facilitate the uptake of implementations and of the interoperability between different implementations, ensuring a more extensive use of IDMS. Interactive 3D tele-immersive applications are future applications that can provide geographically distributed users with a realistic and immersive multimedia experience. Such applications can lead to new requirements for synchronization, and thus novel multidimensional synchronization models for next-generation multimedia environments must be proposed. There is also a strong demand to describe the impact of future application heterogeneity on human synchronization perception.

REFERENCES

- Akyildiz, I. F., & Yen, W. (1996). Multimedia group synchronization protocols for integrated services networks. *IEEE Journal on Selected Areas in Communications*, 14(1), 162–173. doi:10.1109/49.481702
- Belda, J. (2015). Wersync: A web-based platform for distributed media synchronization and social interaction. *Proceedings ACM TVX 2015*.
- Bin-Salem, A., & Wan, T. (2012). Survey of cross-layer designs for video transmission over wireless networks. *IETE Technical Review*, 29(3), 229–247. doi:10.4103/0256-4602.98865

- Blakowski, G., & Steinmetz, R. (1996). A media synchronization survey: Reference model, specification, and case studies. *IEEE Journal on Selected Areas in Communications*, 14(1), 5–35. doi:10.1109/49.481691
- Boronat, F., Guerri, J. C., & Lloret, J. (2008). An RTP/RTCP based approach for multimedia group and inter-stream synchronization. *Multimedia Tools and Applications*, 40(2), 285–319. doi:10.1007/s11042-008-0208-1
- Boronat, F., Lloret, J., & García, M. (2009). Multimedia group and inter-stream synchronization techniques: A comparative study. *Information Systems*, 34(1), 108–131. doi:10.1016/j.is.2008.05.001
- Boronat, F., Montagud, M., & Vidal, V. (2011). Master selection policies for inter-destination multimedia synchronization in distributed applications. *Proceedings of the IEEE 19th Int. Symposium on MASCOTS*, 269–277. doi:10.1109/MASCOTS.2011.43
- Brandenburg, R. (2012). RTCP for inter-destination media synchronization. In IETF Audio/Video Transport Core Maintenance Working Group. Internet Draft.
- Costa, R., & Santos, C. A. S. (2014). Systematic review of multiple contents synchronization in interactive television scenario. *ISRN Communications and Networking*, 1–17.
- Din, S., & Bulterman, D. (2012). Synchronization techniques in distributed multimedia presentation. *MMEDIA 2012: The Fourth International Conferences on Advances in Multimedia*, 1–9.
- ETSI TS 181 016 V3.3.1. (2009). Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN). *Service Layer Requirements to integrate NGN Services and IPTV*.
- Ghinea, G., & Ademoye, O. A. (2010). Perceived synchronization of olfactory multimedia. *IEEE Transactions on Systems, Man, and Cybernetics*, 40(4), 657–663. doi:10.1109/TSMCA.2010.2041224
- Hikichi, K. (2002). The evaluation of delay jitter for haptic collaboration over the Internet. *Conference Record of IEEE Global Communications Conference*, 1492–1496. doi:10.1109/GLOCOM.2002.1188447
- Huang, C.-M., Kung, H.-Y., & Yang, J.-L. (2001). Synchronization and flow adaptation schemes for reliable multiple-stream transmission in multimedia presentations. *Journal of Systems and Software*, 56(2), 133–151. doi:10.1016/S0164-1212(00)00093-5
- Huang, Z., Nahrstedt, K., & Steinmetz, R. (2013). Evolution of temporal multimedia synchronization principles: A historical viewpoint. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 9(1).
- Ishibashi, Y. et al.. (2014). QoE assessment of fairness in networked game with olfaction: Influence of time it takes for smell to reach player. *Multimedia Systems*, 20(3), 1–11.
- Ishibashi, Y., & Tasaka, S. (1997). A group synchronization mechanism for live media in multicast communications. *Proceedings of the IEEE GLOBECOM*, 97, 746–752.
- Ishibashi, Y., Tasaka, S., & Hasegawa, T. (2002b). The Virtual-Time Rendering algorithm for haptic media synchronization in networked virtual environments. *Proceedings of the 16th International Workshop on Communications Quality and Reliability*, 213–217.
- Ishibashi, Y., Tasaka, S., & Miyamoto, H. (2002a). Joint synchronization between stored media with interactive control and live media in multicast communications. *IEICE Transactions on Communications*, E85-B(4), 812–822.

- Ishibashi, Y., Tsuji, A., & Tasaka, S. (1997). A group synchronization mechanism for stored media in multicast communications. In *Proceedings of the Sixth Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM)*, (vol. 2, pp. 692–700). Kobe, Japan: IEEE Press. doi:10.1109/INFOCOM.1997.644522
- Isomura, E., Tasaka, S., & Nunome, T. (2011). QoE enhancement in audiovisual and haptic interactive IP communications by media adaptive intra-stream synchronization. *Proceedings of IEEE TENCON*, 1085–1089. doi:10.1109/TENCON.2011.6129278
- Jung, T., & Seo, K. (2016). A client-driven media synchronization mechanism for RTP packet-based video streaming. *Journal of Real-Time Image Processing*, 12(2), 455–464. doi:10.1007/s11554-015-0497-3
- Laoutaris, N., & Stavrakakis, I. (2002). Intrastream synchronization for continuous media streams: A survey of playout schedulers. *IEEE Network*, 16(3), 30–40. doi:10.1109/MNET.2002.1002997
- Li, M., Sun, Y., & Sheng, H. (1997). Temporal relations in multimedia systems. *Computers & Graphics*, 21(3), 315–320. doi:10.1016/S0097-8493(97)00008-3
- Mauve, M., Vogel, J., Hilt, V., & Effelsberg, W. (2004). Local-lag and timewarp: Providing consistency for replicated continuous app. *IEEE Transactions on Multimedia*, 6(1), 47–57. doi:10.1109/TMM.2003.819751
- Montagud, M., & Boronat, F. (2012). Enhanced adaptive RTCP-based inter-destination multimedia synchronization approach for distributed applications. *Computer Networks*, 56(12), 2912–2933. doi:10.1016/j.comnet.2012.05.003
- Montagud, M., Boronat, F., Stokking, H., & Cesar, P. (2014). Design, development and assessment of control schemes for IDMS in a standardized RTCP-based solution. *Computer Networks*, 70(9), 240–259. doi:10.1016/j.comnet.2014.06.004
- Montagud, M., Boronat, F., Stokking, H., & van Brandenburg, R. (2012). Inter-destination multimedia synchronization: Schemes, use cases and standardization. *Multimedia Systems*, 18(6), 59–482. doi:10.1007/s00530-012-0278-9
- Montagud, M., Jansen, J., Cesar, P., & Boronat, F. (2015). Review of media sync reference models: Advances and open issues. *MediaSync2015 Workshop*.
- Park, S., & Choi, Y. (1996). Real-time multimedia synchronization based on delay offset and play-out rate adjustment. *Real-Time Imaging*, 2(3), 163–170. doi:10.1006/rtim.1996.0016
- Park, S., & Kim, J. (2008). An adaptive media playout for intra-media synchronization of networked-video applications. *Journal of Visual Communication and Image Representation*, 19(2), 106–120. doi:10.1016/j.jvcir.2007.09.002
- Sithu, M., & Ishibashi, Y. (2015). Media synchronization control in multimedia communication. In D. Kanellopoulos (Ed.), *Emerging research on networked multimedia communication systems* (pp. 25–61). Hershey, PA: Information Science Publishing.
- Vaishnavi, P., Cesar, P., Bulterman, D., Friedrich, O., Gunkel, S., & Geerts, D. (2011). From IPTV to synchronous shared experiences challenges in design: Distributed media synchronization. *Signal Processing Image Communication*, 26(7), 370–377. doi:10.1016/j.image.2011.01.006
- Wongwirat, O., & Ohara, S. (2006). Haptic media synchronization for remote surgery through simulation. *IEEE MultiMedia*, 13(3), 62–69. doi:10.1109/MMUL.2006.54

KEY TERMS AND DEFINITIONS

Delay Jitter: The variation of end-to-end delay; an inherent characteristic of best-effort networks.

Preventive Control Techniques: They used to prevent the asynchrony in the media streams. They minimize latencies and jitters and involve disk-reading scheduling algorithms, network transport protocols, operating systems, and synchronization schedulers.

Quality of Experience (QoE): A measure of a customer's experiences with a service (e.g. IPTV).

Reactive Control Techniques: These techniques are designed to recover synchronization in the presence of synchronization errors.

Scripts: A synchronization specification method that describes synchronization scenarios in multimedia objects.

Synchronous Shared Experiences: New social multimedia applications and services wherein

groups of users in different locations can watch multimedia content, while synchronously communicating with each other.

Synthetic Synchronization: Such synchronization is used in presentation and retrieval-based systems with stored data objects that are arranged to provide new multimedia objects. The temporal relations have been artificially assigned to media that were created independently of each other.

Virtual-Time Rendering (VTR) Algorithm: A popular intra- and inter-stream synchronization algorithm applicable to networks with unknown delay bounds. It makes use of globally synchronized clocks, and consists of the dynamic adjustment of the MUs rendering-time, according to the network condition.

Metadata Standards in Digital Audio

M**Kimmy Szeto***Baruch College, City University of New York, USA*

INTRODUCTION

Audio metadata are an essential tool that supports control and management of systems that create, transmit, describe, manage, and store digital audio. Throughout the lifecycle of digital audio objects—pre-production, acquisition and production, post-production, distribution, storage, transmission, and archiving—metadata “describe the attributes of a resource, characterize resource relationships, and support resource discovery, management, and effective use” (Vellucci, 1999). Technical and structural metadata enable audio in devices and software applications; descriptive metadata provide context. Metadata also relate and incorporate audio in a multimedia environment resulting in profound effects on conception, reception, and consumption.

Issues of audio metadata are intertwined with audio equipment, music production, and information technology. Holmes (2006) gives a broad overview of the technology; Gilmer (2004) provides the background for usage in multimedia production; Lu & Hanjalic (2009) and Kriechbaum (2009) detail the technical aspects in database systems; Casey & Gordon (2007) offer the information professional perspective. International audio metadata standards are chiefly issued by the European Broadcasting Union (EBU), the Society of Motion Picture and Television Engineers (SMPTE), and the Audio Engineering Society (AES). Guidelines for usage and implementation are routinely published by these three organizations, as well as by manufacturers, libraries, and archives.

BACKGROUND

Analog sound recording technologies had been in use for over a century when the Compact Disc Digital Audio (CD-DA) technology emerged in the late 1970s following breakthroughs in signal processing and the optical medium. Audio continued to be produced in a mixture of analog and digital systems consisting of single-purpose devices designed for recording and producing audio to be transmitted in either analog or digital form. The production system was largely determined by the transmission medium—wire, wax cylinder, shellac disc, vinyl disc, magnetic tape, broadcast, optical disc—each requiring an assortment of specialty equipment, typically supplied by a single manufacturer. In such vertically integrated systems, controls and switches that supplied parameters and characteristics (metadata) to the audio were generally integrated in the hardware. Where metadata were specified and shared, the necessity for interoperability was limited to the devices immediately before and after in the production chain.

As digital signal processing technologies evolved, the proliferation of recording formats, compression methods, and media formats made it no longer sensible to build a separate integrated system for each device combination. In such a mixed production environment, devices need to be able to pass along and recognize standardized metadata throughout the production process. For several decades since the 1970s, the broadcast industry attempted to standardize metadata within device categories to allow for interchangeable equipment for each step of the production. But

demands for higher flexibility and broader scope mounted as increasing prevalence of computer software and data networks shifted production workflow and transmission toward a distributed model. The industry was compelled to re-conceive use of metadata in not only audio, but also video production in a digital environment where data and content would be stored and transmitted in the form of computer files. These files, ideally, would encapsulate all necessary metadata, interoperable through the production chain, regardless of media format or equipment.

EBU and SMPTE established the Task Force for Harmonized Standards for the Exchange of Program Material as Bit Streams in 1995 to study the long-term interoperability and stability of production systems. In its 1997 report, the Task Force elevated metadata to be equal in significance to the signals, casting the formulas “Content = Essence + Metadata” and “Asset = Content + Rights.” (Essence includes video, audio, and data; metadata represents information about the essence, including rights.) And, among other technical details, the Task Force, anticipating the need to store and transmit digital multimedia as computer files, specified the generalized file format, or “wrapper.” This format consists of three parts—the preamble, consisting solely of metadata, the body, containing the essence and additional metadata, and the end-of-file marker. The wrapper provided the means to link metadata and essence logically and physically. This structure allowed audio files to contain an ever-increasing variety of metadata. SMPTE published the Metadata Element Dictionary Structure (SMPTE 335:2012) and established the SMPTE Registration Authority (2010), a public metadata registry, to serve as a basis and the administrative body for the development of future metadata standards. Open standards will guide the development of audio and multimedia metadata standards to handle increasingly complex and horizontal system design. Global adoption of well-defined international standards will facilitate data exchange and re-use, as well

as guarantee long-term preservation and combat format obsolescence.

Efforts to standardize audio metadata also dramatically shaped the consumer experience and conception of audio. With analog audio, consumers’ ability to produce and manipulate audio was limited to using audiocassette systems. As the Sony Corporation introduced the Walkman, a portable cassette player, in 1979, listening became personalized. Later, personal portable players were extended to the Compact Disc medium, even though recording devices did not reach consumers until much later. With the advent of storing and transmitting audio as computer files, however, audio consumption was no long tied to specific media or carriers. Personal computers allowed manipulation of audio and the associated metadata, and distribution of audio files over data networks. Widespread integration between personal computers, data networks, and personal portable players spurred consumer demands for describing and organizing digital audio. Descriptive metadata standards evolved and expanded in scope alongside techniques to generate audio semantics and analyze consumer preferences and behaviors, resulting in the ability to create personalized and customizable audio collections.

EMBEDDED METADATA IN DIGITAL AUDIO FILE FORMATS

Pulse-Code Modulation

Pulse-code modulation (PCM) is a method of representing analog signals in digital binaries. It was first developed in 1937 by British engineer Alec H. Reeves for audio transmissions over telegraph. Reeves was granted patents in France in 1938, in Britain in 1939, and in the United States in 1942, but lacked the equipment to produce the digital audio efficiently (Deloraine & Reeves, 1965; Robertson, 2005). The digital coding of analog signals is accomplished in three steps. Sound signals are

sampled at specific time intervals (8,000 times per second for typical telephone speech; 44,100 for Compact Disc Digital Audio; up to 96,000 and beyond for high quality recordings). Then, the sampled signals are assigned values, or quantized. Finally, each of these values is represented in binary code. The uncompressed digital audio is called Linear PCM (LPCM), while compressed variants are variously called μ -law, A-law, differential PCM, adaptive differential PCM, etc., depending on technique and application. Today, with the widespread use of semiconductor equipment, PCM serves as the basis for virtually all forms of digital communications.

Compact Disc Digital Audio

Compact Disc Digital Audio is the standard format for audio on compact discs first released in 1980 by Koninklijke Philips and Sony in a publication commonly known as the Red Book. The Red Book specifies the maximum duration of a disc (79.8 minutes), tracks per disc (99), and index points per track (99), the minimum duration of a track (4 seconds), the use of the International Standard Recording Code (ISRC) (ISO 3901), and the audio format (2 channels of 16-bit LPCM sampled at 44.1 Hz). The Red Book standard was codified in 1987 in the International Electrotechnical Commission publication IEC 60908, and its second edition was published in 1999.

Several extensions to the Red Book specification have been introduced. CD-Text, introduced by Sony in 1996 as part of its Multimedia Command Set, provides the ability to store 4 megabytes of textual information about the audio contents such as album title, track title, and authors in the Interactive Text Transmission System (IEC 61866), which is also used for MiniDiscs and digital radio broadcasting. CD-Text has been widely adopted, and CD-Text metadata have been harvested and made available via online databases such as Gracenote (formerly Compact Disc Database, or CDDb) and freedb.

Compact Disc + Graphics (CD+G) was introduced by Philips and Sony in 1985 for storing low-resolution graphics for on-screen display during audio playback. This is an early specification involving storing images as metadata. CD+G and its later revision, Compact Disc + Extended Graphics (CD-EG) are chiefly used for interactive audio playback such as karaoke, but, while part of the Red Book standard, they are otherwise not widely adopted.

Interchange File Format

The Interchange File Format (IFF) (EA IFF 85), developed by Electronic Arts, was originally introduced in 1985 for graphic files in Commodore-Amiga systems, but it is a generic file format that can be used for a variety of media types, including audio. The IFF file consists of a sequence of chunks. Each chunk includes an area that declares the chunk, followed by an area that specifies size, and then the area for the data (known as Key-Length-Value encoding, or KLV), which may include other nested sub-chunks. The concepts of chunks, as well as the requirement for each chunk to provide metadata to self-identify and self-describe, lay the foundation for the basic audio file formats, including Audio Interchange File Format, developed by Apple Computer, and the Resource Interchange File Format, developed by the Microsoft Corporation and the International Business Machines Corporation (IBM), on which their Waveform Audio File Format is based.

Audio Interchange File Format

The Audio Interchange File Format (AIFF) is an audio application of IFF developed in 1988 by Apple. AIFF originally supported only uncompressed PCM audio; support for compressed audio was added in its 1991 revision, AIFF-C. Metadata elements specified in AIFF are listed in Table 1. AIFF requires technical metadata (Common Chunk) and the audio (Sound Data Chunk), while AIFF-C

requires an additional version number (Format Version Chunk). Other optional chunks include additional technical and descriptive metadata designed to provide machine-actionable capability (such as looping), user- and application-interactive features (such as commenting), and extensibility. The concept of time-stamped comments pegged to specific points in the audio data is widely applied today for versioning, subtitling, user tagging, and provenance tracking. Extensibility provided by the Application Specific Chunk allows applications to create custom metadata and incorporate external metadata for storing state parameters such as date last played, last stop position, and user ratings, additional descriptive metadata such as ID3 tags, as well as other binary files such as album cover art—all popular features of audio file management software today.

Resource Interchange File Format

The Resource Interchange File Format (RIFF) was introduced in 1991 by Microsoft and IBM. The format is identical to IFF except for the labeling of chunk names and endianness, the ordering of bytes within computer memory, which are opposite in the processors used in Apple and IBM computers. Microsoft, which develops software on IBM computers, has implemented several file formats based on RIFF, including the WAV, also introduced in 1991.

Waveform Audio File Format

Metadata elements specified in WAV are listed in Table 2. The functionalities offered by WAV are similar to those offered by AIFF/AIFF-C, and also allow custom-defined chunks to be included in the file. The latest addition to the WAV format, Wave Format Extensible, supports multichannel audio with additional technical data.

Table 1. Metadata elements specified in AIFF

Chunk (All Optional Unless Marked)	Comment
Format Version (AIFF-C only, required)	The version of the AIFF-C format.
Common (required)	Technical metadata: number of channels, number of sample frames, bits per sample, and sample rate. Compression type code and a human-readable compression type name were added to AIFF-C.
Instrument	Defines the basic parameters of the playback sound sampler, i.e. the “instrument” as well as looping of audio data segments.
Saxel (AIFF-C draft only)	Sound accelerator data for eliminating decompression artifacts when playback begins at a point other than the beginning of the audio data. This chunk was not officially adopted.
Comments	Free text that can be time-stamped and linked to a marker.
Marker	A marker points to a position in the audio data. It can be used by the Instrument Chunk, the Comments Chunk, or any other element in the Application Chunk.
Sound Data (required)	The audio data. Uncompressed LPCM in AIFF; in AIFF-C, the audio data can be compressed as defined in the Common Chunk.
Name	The name given to the audio.
Author	The creator of the audio.
Copyright	The copyright statement for the audio.
Annotation(s)	Free text, but using the Comments Chunk is preferred.
Audio Recording	Information that identifies the recording device (according to AES3: AES standard for digital audio - Digital input-output interfacing - Serial transmission format for two-channel linearly-represented digital audio data, section 7.1, channel status data).
MIDI Data (s)	Additional metadata included in the MIDI specification but not in the AIFF / AIFF-C.
Application Specific	Custom chunks created for any purpose by device or software application manufacturers.

Table 2. Metadata elements specified in the WAV format

Chunk (All Optional Unless Marked)	Description
Format (required)	Specifies the format, and technical metadata: number of channels, sampling rate, average sampling rate, and data block size.
Fact	Specifies the length of the audio.
Cue Points	Identifies a series of positions in the audio.
Playlist	Specifies the play order of a series of cue points.
Associated data list	Provides Label (title) and Note sub-chunks to be associated with a cue point; defines a Text with Data Length Information sub-chunk that associates a segment of the audio chunk with free text, purpose, country, language, and dialect.
Wave data	The audio data. Uncompressed LPCM or compressed as defined in the Format Chunk.
File	Contains data or metadata described in other file formats.
Instrument (added in the 1994 update)	Defines the basic parameters of the playback sound sampler, i.e. the “instrument”
Sample (added in the 1994 update)	Metadata for sampling keyboards to recognize the WAV audio as an instrument.

Broadcast Wave Format

The Broadcast Wave Format (BWF) is an extension of the WAV format. The BWF specification was first published by EBU in 1997 (EBU Tech 3285) as Version 0. Its Version 1, published in 2000, added support for the Unique Material Identifier in the SMPTE 330 standard, and was followed by 6 supplemental updates listed in Table 3, all of which were subsequently incorporated into the latest Version 2, published in 2011. BWF gained worldwide acceptance and has been incorporated into other file specifications such as MXF, AES57, and EBUCore/AES60. BWF restricts the audio coding to uncompressed LPCM and

Table 3. Supplements to BWF version 1

Supplement	Description
MPEG Audio	Specifies the additional technical metadata necessary to describe MPEG audio.
Capturing Report	Specifies additional information for the full transmission history of the analog-to-digital conversion and digital transfers, in order to safeguard the original quality of the audio.
Peak Envelope	Specifies additional information about peak audio signal levels in the file, in order to speed up opening, display, and processing.
Link	Provides a mechanism to link multiple BWF files for seamless output. Metadata is expressed in XML.
AXML	Formally defines the use of XML metadata in this BWF chunk.
Dolby	Specifies additional information associated with Dolby technologies, such as Dolby E, Digital and Dolby Digital Plus.

compressed MPEG and specifies an additional Broadcast Audio Extension Chunk (BEXT). The additional metadata facilitate exchange of audio files between computer platforms and applications, documentation of processing activities, as well as synchronization with other audio files. Table 4 lists BEXT metadata fields in BWF Version 2. Supplements of BWF specify additional metadata chunks for various applications. BWF has been extended to incorporate multichannel (MBWF) support of Wave Format Extensible. EBU also introduced RF64, which extended the definition of RIFF to support file size exceeding 4 gigabytes. The latest update in 2009 incorporates MBWF in R64 (EBU Tech 3306).

Other Digital Audio File Formats

Formats that have been developed for more specialized use and narrower functionalities include Sun Microcomputer’s AU format, Computerized Speech Labs’s NSP format, Entropic Research Laboratories’s ESPS and HTK formats, INRS

Table 4. BEXT metadata fields in BWF version 2

BEXT Data Field	Description
Description	A brief description of the audio.
Originator	Name of the producer of the audio.
Originator Reference	An identifier for the Originator, defined in EBU R 99.
Origination Date	Creation date of the audio.
Origination Time	Creation time of the audio.
Time Reference	Time code of the start of the audio.
Version	Version of BWF of the file.
UMID	Unique Material Identifier defined in SMPTE 330M.
Loudness Value	
Loudness Range	
Max True Level	
Max Momentary Loudness	
Max Short Term Loudness	
Reserved	Reserved for future extension.
Coding History	Record of the transmission chain—processing applied to the audio.

Telecom’s file format, IRCAM’s Soundfile format, and AT&T Bell Laboratories’s SPPACK format (Kabal, 2011).

Proliferation of digital audio and video on the Internet prompted rapid development of compressed audio and multimedia technologies. Some digital audio file formats have become known by their compression method, or codec, most notably the popular formats MP3, Advanced Audio Encoding (AAC), and Microsoft’s proprietary Windows Media Audio (WMA) format; some are subsumed under wrapper file formats designed for multimedia and streaming delivery, such as the MXF, Ogg, a series of MPEG standards, and Microsoft’s proprietary Advanced Systems Format. Consequently, the development of audio metadata diverged—technical metadata continue to be embedded within the audio file or specified as part of the multimedia file format, while descriptive metadata evolved independently to coordinate with these formats.

DESCRIPTIVE METADATA

ID3v1 and ID3v2

ID3v1 and ID3v2 are two related versions of metadata containers for descriptive metadata developed without a standardization body, but are standards in the de facto sense for their wide adoption by software applications and hardware manufacturers. The introduction of ID3 in 1996, now called ID3v1, is attributed to programmer Eric Kemp. A chunk of data was appended to the end of the MP3 file container to store descriptive metadata. The format was extended in 1997 to ID3v1.1, attributed to Michael Mutschler, which increased the storage size for the metadata (Nilsson, 2006). The ID3v1 chunk is 128 bytes long with fixed-length metadata tags for title, artist, album, year, comment, indicator for the use of the track number, track number, and genre code (listed in Table 6). The current standard ID3v2 provides an additional 227 bytes for extensions of title, artist, and album, as well as new tags for speed, genre (free text), and fading. In 1998, a group of developers introduced ID3v2 (Nilsson, 2012), which makes use of frames of variable size (up to 16 megabytes each and a total of 256 megabytes). The list of tags is shown in Table 5. In addition to textual metadata, ID3v2 allows for external URL links as well as embedding graphic images through the “APIC” frame.

The ID3v2 frame structure has been incorporated in AIFF, but not in WAV and BWF, although incorporation is possible by way of a custom ID3 Chunk. The set of ID3v2 tag names is also supported in WMA and the MPEG-4 (MP4) file formats, but the metadata are stored in a different data structure.

VorbisComments in Ogg

VorbisComments is a textual metadata specification for the open source Ogg file container format developed by the Xiph.org Foundation. Appeared in 2000, this specification is primarily concerned

Table 5. List of ID3v2 tags

Frame Name	Frame Code			Description
	2.2.0	2.3.0	2.4.0	
GENREID				genre ID
CONTENTGROUP	TT1	TIT1	TIT1	content group description
TITLE	TT2	TIT2	TIT2	title/song title/content description
SUBTITLE	TT3	TIT3	TIT3	subtitle/description refinement
ARTIST	TP1	TPE1	TPE1	lead performer(s)/soloist(s)
BAND	TP2	TPE2	TPE2	band/orchestra/accompaniment
CONDUCTOR	TP3	TPE3	TPE3	conductor/performer refinement
MIXARTIST	TP4	TPE4	TPE4	interpreted, remixed, modified by
COMPOSER	TCM	TCOM	TCOM	composer
LYRICIST	TXT	TEXT	TEXT	lyricist/text writer
LANGUAGE	TLA	TLAN	TLAN	language(s)
CONTENTTYPE	TCO	TCON	TCON	content type
ALBUM	TAL	TALB	TALB	album/movie/show title
TRACKNUM	TRK	TRCK	TRCK	track number/position in set
PARTINSET	TPA	TPOS	TPOS	part of set
ISRC	TRC	TSRC	TSRC	international standard recording code
DATE	TDA	TDAT		date
YEAR	TYE	TYER		year
TIME	TIM	TIME		time
RECORDINGDATES	TRD	TRDA		recording dates
RECORDINGTIME			TDRC	recording time
ORIGYEAR	TOR	TORY		original release year
ORIGRELEASETIME			TDOR	original release time
BPM	TBP	TBPM	TBPM	beats per minute
MEDIATYPE	TMT	TMED	TMED	media type
FILETYPE	TFT	TFLT	TFLT	file type
COPYRIGHT	TCR	TCOP	TCOP	copyright message
PUBLISHER	TPB	TPUB	TPUB	publisher
ENCODEDBY	TEN	TENC	TENC	encoded by
ENCODERSETTINGS	TSS	TSSE	TSSE	software/hardware settings for encoders
SONGLN	TLE	TLEN	TLEN	length (ms)
SIZE	TSI	TSIZ		size (bytes)
PLAYLISTDELAY	TDY	TDLY	TDLY	playlist delay
INITIALKEY	TKE	TKEY	TKEY	initial key
ORIGALBUM	TOT	TOAL	TOAL	original album/movie/show title
ORIGFILENAME	TOF	TOFN	TOFN	original filename
ORIGARTIST	TOA	TOPE	TOPE	original artist(s)/performer(s)

continued on following page

Table 5. Continued

Frame Name	Frame Code			Description
	2.2.0	2.3.0	2.4.0	
ORIGLYRICIST	TOL	TOLY	TOLY	original lyricist(s)/text writer(s)
FILEOWNER		TOWN	TOWN	file owner/licensee
NETRADIOSTATION		TRSN	TRSN	Internet radio station name
NETRADIOOWNER		TRSO	TRSO	Internet radio station owner
SETSUBTITLE			TSST	set subtitle
MOOD			TMOO	mood
PRODUCEDNOTICE			TPRO	production credit
ENCODINGTIME			TDEN	encoding time
RELEASETIME			TDRL	release time
TAGGINGTIME			TDTG	tagging time
ALBUMSORTORDER			TSOA	album sort order
PERFORMERSORTORDER			TSOP	performer sort order
TITLESORTORDER			TSOT	title sort order
USERTEXT	TXX	TXXX	TXXX	user defined text information frame
WWWAUDIOFILE	WAF	WOAF	WOAF	official audio file URL
WWWARTIST	WAR	WOAR	WOAR	official artist/performer URL
WWWAUDIOSOURCE	WAS	WOAS	WOAS	official audio source URL
WWWCOMMERCIALINFO	WCM	WCOM	WCOM	commercial information URL
WWWCOPYRIGHT	WCP	WCOP	WCOP	copyright/legal information URL
WWWPUBLISHER	WPB	WPUB	WPUB	publisher's official URL
WWWRADIOPAGE		WORS	WORS	official Internet radio station homepage URL
WWWPAYMENT		WPAY	WPAY	payment URL
WWWUSER	WXX	WXXX	WXXX	user defined URL link frame
INVOLVEDPEOPLE	IPL	IPLS		involved people list
MUSICIANCREDITLIST			TMCL	musician credits list
INVOLVEDPEOPLE2			TIPL	involved people list
UNSYNCEDLYRICS	ULT	USLT	USLT	unsynchronized lyrics/text transcription
COMMENT	COM	COMM	COMM	comments
TERMSOFUSE		USER	USER	terms of use
UNIQUEFILEID	UFI	UFID	UFID	unique file identifier
CDID	MCI	MCDI	MCDI	music CD identifier
EVENTTIMING	ETC	ETCO	ETCO	event timing codes
MPEGLOOKUP	MLL	MLLT	MLLT	MPEG location lookup table
SYNCEDTEMPO	STC	SYTC	SYTC	synchronized tempo codes
SYNCEDLYRICS	SLT	SYLT	SYLT	synchronized lyrics/text
VOLUMEADJ	RVA	RVAD		relative volume adjustment
VOLUMEADJ2			RVA2	relative volume adjustment (2)
EQUALIZATION	EQU	EQUA		equalization

continued on following page

Table 5. Continued

Frame Name	Frame Code			Description
	2.2.0	2.3.0	2.4.0	
EQUALIZATION2			EQU2	equalization (2)
REVERB	REV	RVRB	RVRB	reverb
PICTURE	PIC	APIC	APIC	attached picture
GENERALOBJECT	GEO	GEOB	GEOB	general encapsulated object
PLAYCOUNTER	CNT	PCNT	PCNT	play counter
POPULARIMETER	POP	POPM	POPM	popularimeter
BUFFERSIZE	BUF	RBUF	RBUF	recommended buffer size
CRYPTEDMETA	CRM			encrypted meta frame
AUDIOCRYPTO	CRA	AENC	AENC	audio encryption
LINKEDINFO	LNK	LINK	LINK	linked information
POSITIONSYNC		POSS	POSS	position synchronization frame
COMMERCIAL		COMR	COMR	commercial frame
CRYPTOREG		ENCR	ENCR	encryption method registration
GROUPINGREG		GRID	GRID	group identification registration
PRIVATE		PRIV	PRIV	private frame
OWNERSHIP		OWNE	OWNE	ownership frame
SIGNATURE			SIGN	signature frame
SEEKFRAME			SEEK	seek frame
AUDIOSEEKPOINT			ASPI	audio seek point index

Compiled from Harvey (2013) and Nilsson (1998, 1999, 2000b), and adapted from Solboda (2011).

with the structure and encoding. There is no mandatory field, and users are free to create any field. The Xiph.org Foundation recommends a “minimal list of standard field names” (listed in Table 7) and later added a mechanism for embedding digital images. Additional VorbisComments field names as well as other mechanisms for embedding descriptive metadata are being proposed, discussed, and tested by the user community.

Material Exchange Format

The Material Exchange Format (MXF) consists of a series of specifications published by SMPTE beginning in 2004. MXF grew out of the Advanced Authoring Format (AAF), and is a comprehensive file format that addresses the interoperability issue

in the production chain. The format is supported by EBU, the Association of Radio Industries and Businesses (ARIB), the Professional-MPEG Forum, the Advanced Authoring Format Association, as well as equipment manufacturers. The specifications focus on technical and structural metadata, and leaves open mechanisms for embedding descriptive metadata. Several recommended practices have been published (SMPTE 377-1, EBU R 121, SMPTE 380M). Although MXF metadata are KLV-encoded, mechanisms for incorporating XML metadata have been developed (EBU R 121, SMPTE 380M, SMPTE ST 434:2014). The *SMPTE Metadata Registries and Related Items* web site serves as a clearinghouse for metadata standards developed for MXF, including EBUCore.

Table 6. ID3v1 and ID3v2 genre codes

Code	Genre	Code	Genre	Code	Genre
0	Blues	65	Cabaret	130	Terror
1	Classic Rock	66	New Wave	131	Indie
2	Country	67	Psychedelic	132	Brit Pop
3	Dance	68	Rave	133	Afro-Punk
4	Disco	69	Showtunes	134	Polish Punk
5	Funk	70	Trailer	135	Beat
6	Grunge	71	Lo-Fi	136	Christian Gangsta Rap
7	Hip-Hop	72	Tribal	137	Heavy Metal
8	Jazz	73	Acid Punk	138	Black Metal
9	Metal	74	Acid Jazz	139	Crossover
10	New Age	75	Polka	140	Contemporary Christian
11	Oldies	76	Retro	141	Christian Rock
12	Other	77	Musical	142	Merengue
13	Pop	78	Rock & Roll	143	Salsa
14	R&B	79	Hard Rock	144	Thrash Metal
15	Rap	80	Folk	145	Anime
16	Reggae	81	Folk-Rock	146	J-Pop
17	Rock	82	National Folk	147	Synthpop
18	Techno	83	Swing	148	Abstract
19	Industrial	84	Fast-Fusion	149	Art Rock
20	Alternative	85	Bebop	150	Baroque
21	Ska	86	Latin	151	Bhangra
22	Death Metal	87	Revival	152	Big Beat
23	Pranks	88	Celtic	153	Breakbeat
24	Soundtrack	89	Bluegrass	154	Chillout
25	Euro-Techno	90	Avant-garde	155	Downtempo
26	Ambient	91	Gothic Rock	156	Dub
27	Trip-Hop	92	Progressive Rock	157	EBM
28	Vocal	93	Psychedelic Rock	158	Eclectic
29	Jazz+Funk	94	Symphonic Rock	159	Electro
30	Fusion	95	Slow Rock	160	Electroclash
31	Trance	96	Big Band	161	Emo
32	Classical	97	Chorus	162	Experimental
33	Instrumental	98	Easy Listening	163	Garage
34	Acid	99	Acoustic	164	Global
35	House	100	Humor	165	IDM
36	Game	101	Speech	166	Illbient
37	Sound Clip	102	Chanson	167	Industro-Goth

continued on following page

Table 6. Continued

Code	Genre	Code	Genre	Code	Genre
38	Gospel	103	Opera	168	Jam Band
39	Noise	104	Chamber Music	169	Krautrock
40	Alt. Rock	105	Sonata	170	Leftfield
41	Bass	106	Symphony	171	Lounge
42	Soul	107	Booty Bass	172	Math Rock
43	Punk	108	Primus	173	New Romantic
44	Space	109	Porn Groove	174	Nu-Breakz
45	Meditative	110	Satire	175	Post-Punk
46	Instrumental Pop	111	Slow Jam	176	Post-Rock
47	Instrumental Rock	112	Club	177	Psytrance
48	Ethnic	113	Tango	178	Shoegaze
49	Gothic	114	Samba	179	Space Rock
50	Darkwave	115	Folklore	180	Trop Rock
51	Techno-Industrial	116	Ballad	181	World Music
52	Electronic	117	Power Ballad	182	Neoclassical
53	Pop-Folk	118	Rhythmic Soul	183	Audiobook
54	Eurodance	119	Freestyle	184	Audio Theatre
55	Dream	120	Duet	185	Neue Deutsche Welle
56	Southern Rock	121	Punk Rock	186	Podcast
57	Comedy	122	Drum Solo	187	IndieRock
58	Cult	123	A Cappella	188	G-Funk
59	Gangsta Rap	124	Euro-House	189	Dubstep
60	Top 40	125	Dance Hall	190	Garage Rock
61	Christian Rap	126	Goa	191	Psybient
62	Pop/Funk	127	Drum & Bass	CR	Cover
63	Jungle	128	Club-House	RX	Remix
64	Native American	129	Hardcore	255	None

Compiled from Harvey (2013), "Liste" (2013), and Strnad (2002).

0-79 were introduced with ID3v1; extended to 80-110 by the Winamp application in 1997; extended to 111-115, 116-141, 142-147 in 1998; extended to 148-191 in 2010; CR and RX values appear in ID3v2 only.

EBU and AES Metadata Specifications

The co-publication of EBUCore (EBU Tech 3293) in 2008 and AES60 (formerly AES-X098A) in 2011 was the result of the metadata harmonization project between EBU and the AES. EBUCore assumes the use of BWF, and, an extension of the Dublin Core metadata schema expressed in XML, was designed as a minimum list of elements for de-

scribing audio and video. The EBUCore metadata set includes 19 elements: Title, Creator, Subject, Description, Publisher, Contributor, Date, Type, Format, Identifier, Source, Language, Relation, Coverage, Rights, Version, Publication History, Metadata Provider, and Entity Type. EBUCore has been widely adopted by production and broadcasting organizations internationally. The EBUCore standard is now in Version 1.6, published in 2015; the corresponding AES standard is due

Table 7. Recommended fields for VorbisComments metadata

Recommended Field	Description
TITLE	Title of the work.
VERSION	May be used to differentiate multiple versions of the same TITLE.
ALBUM	The collection name to which this track belongs.
TRACKNUMBER	The track number of this piece if part of a specific larger collection or album.
ARTIST	The author(s), creator(s) or composer(s) of the work.
PERFORMER	The artist(s) who performed the work. Omitted if it is the same as ARTIST.
COPYRIGHT	Copyright attribution.
LICENSE	License information.
ORGANIZATION	Name of the organization or label producing the track.
DESCRIPTION	A short text description of the contents.
GENRE	A short text indication of music genre.
DATE	Date the track was recorded.
LOCATION	Location where track was recorded.
CONTACT	Contact information for the creators or distributors of the track.
ISRC	The International Standard Recording Code.
METADATA_BLOCK_PICTURE	Embedded digital image.

to be revised in 2016. In 2015, EBU generalized the technical metadata area of EBU Core with the Audio Definition Model (ADM) (EBU Tech 3364), which enables descriptions of audio containing tracks of mixed formats.

AES also published complementary standards for embedding technical metadata in MPEG Layer II compressed audio file format (AES41) and for the archiving of analog audio in digital form (AES57, formerly AES-X098B). Administrative metadata for audio production workflow (AES-X155) and the history of signal processing (AES-X098C) are currently in development.

EBU has registered EBUCore with the SMPTE Registration Authority, and continues its development of metadata specifications towards linked

data and semantic web applications. EBUCore has been simultaneously published as a RDF/OWL ontology, which has formed the basis for the development of the W3C Ontology for Media Resources, an interlingua ontology and an associated API.

PBCore

PBCore, first introduced in 2005 with the support of the Corporation for Public Broadcasting, is a community-developed metadata standard for describing analog and digital audio assets primarily in broadcasting. An extension of Dublin Core, PBCore provides elements that enhance description in the areas of instantiation (analog and digital), intellectual content, and intellectual property. PBCore 2.0, introduced in 2011 (updated to version 2.1 in 2015), enhances the ability to draw relationships and include rights information at the instantiation level, including the handling of segments, multi-part assets, and derivatives. The PBCore schema allows external URI references for all element values.

MPEG-7 Audio

The Multimedia Content Description Interface (MPEG-7), developed by the Motion Picture Expert Group, is a suite of ISO standards (ISO/IEC 15938) in 13 parts published between 2002 and 2008. Following a comprehensive approach to describing multimedia content, MPEG-7 allows fast and efficient searching and retrieval. The standard defines the XML-based Description Definition Language (DDL), which serves as the basis for developing Description Schemes (DSs) and Descriptors (Ds). MPEG-7 Part 4 includes structures for describing technical parameters of audio signals such as silence, timbre, harmonic features, spectral features, and signal features, as well as descriptive metadata such as audio signature, musical instrument sounds, melody description, general sound recognition and indexing, and spoken content. The division between technical

and descriptive is blurred since MPEG-7 audio metadata are in part based on the characteristics of the audio itself rather than of human description. These metadata enable searching and filtering based on sound segments, sound characteristics, and lexical inputs. Since its inception, MPEG-7 has spurred research and development in semantic indexing as well as in automatic and interactive extraction and retrieval of multimedia information.

IEEE 1599

The Institute of Electrical and Electronics Engineers published a standard in 2008 for music encoding (IEEE 1599) that provides an XML language to encapsulate a holistic sound object. The goal of this standard is to integrate symbolic computer notations, printed scores, machine rendering, recorded performances, and related objects to enhance real-time human interaction with these facets of a single musical work, as well as between musical works.

FUTURE RESEARCH DIRECTIONS

Digital audio will continue to evolve in a distributed environment with increasing demand for search and retrieval in a multimedia context. The development of the semantic web will bridge audio essence and semantics through automatic and semi-automatic semantic extraction processes. These uses of digital audio necessitate mapping KLV-style metadata to XML, refining syntactic structures to better describe audio essence, developing multimedia ontologies, and harmonizing relationships among metadata elements and schemes.

Audio information segmentation, classification, and retrieval will remain major topics of research, as will automatic speech recognition and audio-enriched metadata extraction based on knowledge of music theory. These audio metadata will form the basis of semantic audio, which will significantly impact not only the production,

transmission, and consumption of audio, but also open up new areas of applications in scientific, social, informatics, business, and artificial intelligence research.

CONCLUSION

By the turn of the 21st century, metadata standards have evolved together with production and transmission technologies from equipment-dependent formats to the Open System Interconnection model (ISO/IEC 7498-1) where digital audio objects are stored and transmitted in a distributed network. Metadata standards have also developed alongside the transition from carrier- and device-dependent audio formats to web-based, personalized delivery. In this new environment, audio metadata enable the re-contextualization, repackaging, and commodification of audio, integrating it with multimedia contextual elements and transforming it into a new cultural product.

REFERENCES

- Casey, M., & Gordon, B. (Eds.). (2007). *Sound directions: Best practices for audio preservation*. Retrieved from <http://www.dlib.indiana.edu/projects/sounddirections/bestpractices2007/>
- Deloraine, E. M., & Reeves, A. H. (1965). The 25th anniversary of pulse code modulation. *IEEE Spectrum*, 2(5), 56–63. doi:10.1109/MSPEC.1965.5212943
- Gilmer, B. (Ed.). (2004). *File interchange handbook for images, audio and metadata*. Burlington, MA: Focal Press.
- Harvey, P. (2013). ID3 tags. In *ExifTool by Phil Harvey*. Retrieved from <http://www.sno.phy.queensu.ca/~phil/exiftool/TagNames/ID3.html>
- Holmes, T. (2006). *The Routledge guide to music technology*. New York: Routledge.

Kabal, P. (2011). *Audio file format specifications*. Retrieved from <http://www-mmsp.ece.mcgill.ca/Documents/AudioFormats/>

Kriechbaum, W. (2009). Audio metadata. In *Encyclopedia of Database Systems* (pp. 157–160). New York: Springer.

Liste der ID3v1-Genres. (2013). *Wikipedia, die freie Enzyklopädie*. Retrieved on September 10, 2013 from http://de.wikipedia.org/wiki/Liste_der_ID3v1-Genres

Lu, L., & Hanjalic, A. (2009). Audio. In *Encyclopedia of Database Systems* (pp. 146–148). New York: Springer; doi:10.1007/978-0-387-39940-9

Morris, J. W. (2012). Making music behave: Metadata and the digital music commodity. *New Media & Society*, 14(5), 850–866. doi:10.1177/1461444811430645

Nilsson, M. (1998). *ID3 tag information*. Retrieved from <http://www.id3.org/id3v2-00>

Nilsson, M. (1999). *ID3 tag version 2.3.0*. Retrieved from <http://www.id3.org/id3v2.3.0>

Nilsson, M. (2000). *ID3 tag version 2.4.0 - Native Frames*. Retrieved from <http://www.id3.org/id3v2.4.0-frames>

Nilsson, M. (2006). Low tech history. In *ID3v2: The audience is informed*. Retrieved from <http://www.id3.org/History>

Nilsson, M. (2012). Contributors. In *ID3v2: The audience is informed*. Retrieved from <http://www.id3.org/Contributors>

Robertson, D. (2005). Reeves, Alec Harley (1902–1971). In *Oxford Dictionary of National Biography*. Oxford, UK: Oxford University Press. Retrieved on October, 2005, from <http://www.oxforddnb.com/view/article/45670>

SMPTE. (2010). *SMPTE Registration Authority*. White Plains, NY: Society of Motion Picture and Television Engineers Registration Authority. Retrieved from <http://www.smp-te-ra.org/mdd/>

SMPTE. (2012). *SMPTE 335:2012: Metadata Element Dictionary Structure*. White Plains, NY: Society of Motion Picture and Television Engineers Registration Authority. Retrieved from; doi:10.5594/SMPTE.ST335.2012

Solboda, T. (2011). *Side-by-side comparison of ID3 versions, fields/frames*. Retrieved from http://www.unixgods.org/~tilo/ID3/docs/ID3_comparison2.html

Strnad, P. (2002). Tag format specs. In *MP3 Manager*. Retrieved from <http://www.mpx.cz/mp3manager/tags.htm>

Vellucci, S. L. (1999). Metadata for music: Issues and directions. *Fontes Artis Musicae*, 46(3/4), 205–217.

ADDITIONAL READING

Apple Computer. (1989). *Audio interchange file format: “AIFF.”* Retrieved from <http://www-mmsp.ece.mcgill.ca/Documents/AudioFormats/AIFF/Docs/AIFF-1.3.pdf>

Apple Computer. (1991). *Audio interchange file format: “AIFF-C”: A revision to include compressed audio data*. Retrieved from <http://www-mmsp.ece.mcgill.ca/Documents/AudioFormats/AIFF/Docs/AIFF-C.9.26.91.pdf>

Audio Engineering Society. (2011). *AES57: AES standard for audio metadata: Audio object structures for preservation and restoration*. New York: Audio Engineering Society.

Audio Engineering Society. (2011). *AES60: AES standard for audio metadata: Core audio metadata*. New York: Audio Engineering Society.

Audio Engineering Society. (2012). *AES41: AES standard for digital audio: Audio-embedded metadata (Parts 1-5)*. New York: Audio Engineering Society.

- Baggi, D. L., & Haus, G. M. (Eds.). (2009a). The new standard IEEE 1599: Interacting with music contents by XML symbols. *Journal of Multimedia*, 4(1).
- Baggi, D. L., & Haus, G. M. (2009b). IEEE 1599: Music Encoding and Interaction. *IEEE Computer*, 42(3), 84–87. doi:10.1109/MC.2009.85
- Bourke, P. (1996). *Creating AIFF audio formatted files*. Retrieved from <http://paulbourke.net/dataformats/audio/>
- Bradley, K. (Ed.). (2009). *IASA-TC04: Guidelines on the production and preservation of digital audio objects* (2nd ed.). Auckland Park, South Africa: International Association of Sound and Audio Visual Archives.
- Brighton, J. (2011). Introducing PBCore 2.0. In *PBCore: Public broadcasting metadata dictionary project*. Retrieved from <http://www.pbcore.org/news/introducing-pbcore-2-0>
- Documentation. (2011). In *PBCore: Public broadcasting metadata dictionary project*. Retrieved from <http://v2.pbcore.org/documentation/>
- Emmerson, S. (2015). Listening with machines: A shared approach. *Organised Sound*, 20(1), 68–75. doi:10.1017/S1355771814000442
- European Broadcasting Union. (2002). *Tech 3295: P_META Metadata Exchange Scheme*. Geneva: European Broadcasting Union.
- European Broadcasting Union. (2003). *EBU Technical Information I36-2003: Metadata Implementation Considerations for Broadcasters*. Retrieved from https://www.ebu.ch/CMSimages/fr/tec_text_i36-2003_tcm7-11055.pdf
- European Broadcasting Union. (2007). *Material Exchange Format basic user metadata implementation* [EBU Recommendation R 121]. Geneva: European Broadcasting Union.
- European Broadcasting Union. (2009). *MBWF/RF64: An extended file format for audio*. Geneva: European Broadcasting Union. Retrieved from <http://tech.ebu.ch/docs/tech/tech3306-2009.pdf>
- European Broadcasting Union. (2011). *Specification of the Broadcast Wave Format (BWF): A format for audio data files in broadcasting*. Geneva: European Broadcasting Union. Retrieved from <https://tech.ebu.ch/docs/tech/tech3285.pdf>
- European Broadcasting Union. (2014). *Tech 3364: Audio Definition Model: Metadata Specification, Version 1.0*. Geneva: European Broadcasting Union.
- European Broadcasting Union. (2015). *Tech 3293: EBU Core Metadata Set (EBUCore), Version 1.6*. Geneva: European Broadcasting Union.
- Evain, J.-P. (n.d.). *ebucore* [OWL ontology]. Retrieved from <http://www.ebu.ch/metadata/ontologies/ebucore/>
- Federal Agencies Audio-Visual Working Group. (2012). *Guidelines: Embedded metadata in Broadcast WAVE files*. Retrieved from <http://www.digitizationguidelines.gov/guidelines/digitize-embedding.html>
- IBM Corporation, & Microsoft Corporation. (1991). *Multimedia programming interface and data specifications 1.0*. Retrieved from <http://www.mmsp.ece.mcgill.ca/Documents/AudioFormats/WAVE/Docs/riffmci.pdf>
- IEEE. Computer Society. (2008). IEEE recommended practice for defining a commonly acceptable musical application using XML (IEEE Std 1599-2008). New York: The Institute of Electrical and Electronics Engineers.
- Institut für Rundfunktechnik. (2013). *MXF Test Center*. Retrieved from <http://mxf.irt.de/information/eburecommendations/>

- International Electrotechnical Commission. (1987). *Compact Disc Digital Audio system*. Geneva: International Electrotechnical Commission.
- International Electrotechnical Commission. (1999). *Audio recording: Compact Disc Digital Audio system*. Geneva: International Electrotechnical Commission.
- Ive, J. G. S. (2006). MXF. *Broadcast Engineering*, 48(8), 14–17.
- Jayant, N. S., & Noll, P. (1984). *Digital coding of waveforms: Principles and applications to speech and video*. Englewood Cliffs, NJ: Prentice Hall.
- Kim, H.-G., Moreau, N., & Thomas, S. (2005). *MPEG-7 Audio and Beyond: Audio Content Indexing and Retrieval*. Hoboken, NJ: J. Wiley; doi:10.1002/0470093366
- Lacinak, C., & Forsberg, W. (2011). A study of embedded metadata support in audio recording software: Summary of findings and conclusions. Annapolis, MD: Association for Recorded Sound Collections; Retrieved from http://www.arsc-audio.org/pdf/ARSC_TC_MD_Study.pdf
- Lee, W., Bailer, W., Bürger, T., Champin, P.-A., Evain, J.-P., Malaisé, V.,... (Eds.). (2012). *Ontology for Media Resources 1.0*. Retrieved from <http://www.w3.org/TR/mediaont-10/>
- Martinez, J. M. (Ed.). (2004). *MPEG-7 Overview (version 10)*. Palma de Mallorca: International Organisation for Standardisation.
- Microsoft Corporation. (1994). *Microsoft multimedia standards update*. Retrieved from <http://www-mmsp.ece.mcgill.ca/Documents/Audio-Formats/WAVE/Docs/RIFFNEW.pdf>
- Microsoft Corporation. (2007). Multiple Channel AudioData and WAVEFiles. In *Windows DevCenter – Hardware*. Retrieved from <http://msdn.microsoft.com/en-us/windows/hardware/gg463006.aspx>
- Morrison, J. (1985). “EA IFF 85” standard for interchange format files. Retrieved from <http://www.martinreddy.net/gfx/2d/IFF.txt>
- Moving Picture Experts Group. (2005). *MPEG-7 Audio*. Retrieved from <http://mpeg.chiariglione.org/standards/mpeg-7/audio>
- Müller, M. (2015). *Fundamentals of Music Processing: Audio, Analysis, Algorithms, Applications*. Cham, Switzerland: Springer International Publishing. doi:10.1007/978-3-319-21945-5
- National Digital Information Infrastructure and Preservation Program. (2013). Format Descriptions for Sound. In *Sustainability of digital formats: Planning for Library of Congress collections*. Retrieved from <http://www.digitalpreservation.gov/formats/fdd/fdd000016.shtml>
- Newcomb, S. B. (1991). Standard Music Description Language. *Computer*, 24(7), 76–79. doi:10.1109/2.84842
- Nilsson, M. (2000a). *ID3 tag version 2.4.0 - Main Structure*. Retrieved from <http://www.id3.org/id3v2.4.0-structure>
- Nilsson, M. (2000b). *The MPEG audio meta data format ID3*. Retrieved from <http://tools.ietf.org/html/draft-nilsson-id3-00>
- PBCore Talk. (2013). Retrieved from <http://groups.google.com/group/pbcore-talk>
- PBCore_2.1. (2015). Retrieved from http://github.com/WGBH/PBCore_2.1
- Pruett, M. (2013). *Audio file library*. Retrieved from <http://www.68k.org/~michael/audiofile/>
- Schmitt, T., & Bernstein, R. (2012). *CD Text format*. Retrieved from <http://www.gnu.org/software/libcdio/cd-text-format.html>
- Seebach, P. (2006). Standards and specs: the Interchange File Format (IFF). In *developer-Works*. Retrieved on June 13, 2006 from <http://www.ibm.com/developerworks/power/library/pa-spec16/?ca=dgr-lnxw07IFF>
- SMPTE. (2004a). *SMPTE EG 42: Material Exchange Format (MXF) – MXF descriptive metadata Scheme*. White Plains, NY: SMPTE.

SMPTE. (2004b). *SMPTE 380M: Television – Material Exchange Format (MXF) – Descriptive Metadata Scheme-1 (standard, dynamic)*. White Plains, NY: SMPTE.

SMPTE. (2012). *SMPTE Metadata Registries and related items*. White Plains, NY: Society of Motion Picture and Television Engineers Registration Authority. Retrieved from <http://www.smp-te-ra.org/mdd/>

SMPTE. (2014). *SMPTE ST 434: Material Exchange Format (MXF) – XML encoding for metadata and file structure information*. White Plains, NY: SMPTE.

SMPTE. (n.d.) *Online Registry*. Retrieved from <http://www.smp-te-ra.org/diffuser/>

Sony Corporation. (1996). *CD TEXT authoring and mastering operation manual*. Retrieved from <http://web.archive.org/web/20070204035327/http://www.sonydadc.com/file/cdtext.zip>

Stegmaier, F., Bailer, W., Bürger, T., Suarez-Figuera, M. C., Mannens, E., Evain, J. P., & Kosch, H. (2013). Unified access to media metadata on the web. *IEEE MultiMedia*, 20(2), 22–29. doi:10.1109/MMUL.2012.55

Supirovic, P. (1999). *MPEG audio frame header*. Retrieved from http://mpgedit.org/mpgedit/mpeg_format/mpeghdr.htm

Task Force for Harmonized Standards for the Exchange of Programme Material as Bit Streams. (1997). Final Report. *EBU Technical Review 277*.

Task Force for Harmonized Standards for the Exchange of Programme Material as Bit Streams. (1998). Final Report: Analyses and Results [Special supplement]. In *EBU Technical Review*. Retrieved from <http://tech.ebu.ch/docs/techreview/ebu-smp-te-tf-bitstreams.pdf>

The state of recorded sound preservation in the United States: A national legacy at risk in the digital age. (2010). Washington, DC: Council on Library and Information Resources and The Library of Congress. Retrieved from <http://www.clir.org/pubs/reports/pub148/pub148.pdf>

Waggener, B. (1995). *Pulse code modulation techniques: With applications in communications and data recording*. New York: Van Nostrand Reinhold.

Xiph.org Foundation. (2012). *VorbisComment*. Retrieved from <http://wiki.xiph.org/index.php/VorbisComment>

Xiph.org Foundation. (2015). *Ogg Vorbis I format specification: Comment field and header specification*. Retrieved from http://xiph.org/vorbis/doc/Vorbis_I_spec.html#x1-820005

KEY TERMS AND DEFINITIONS

Descriptive Metadata: Describe a resource for the purpose of identification, discovery, and retrieval.

Digital Audio: Digital coding of analog audio signals.

Embedded Metadata: Metadata that are stored in the same file that carries the object described by the metadata.

Extensible Markup Language: A markup language that defines a set of rules for encoding textual data recommended for semantic web publishing.

Metadata Schemas: Structures that establish and define data elements, the relationships between them, and the rules governing their use.

Metadata Standards: Specifications or definitions that establish common meaning, usage, and interpretation of metadata.

Semantic Audio: The extraction of symbols and meaning related to audio, including speech and non-speech content.

Technical Metadata: Describe the state, the creation process, and the digital characteristics of a resource.

Multimedia-Enabled Dot Codes as Communication Technologies

Shigeru Ikuta

Otsuma Women's University, Japan

INTRODUCTION

Augmentative and alternative communication (AAC) technologies are widely used, providing students with severe speech, language, and communication difficulties the opportunity to improve their communication, and by extension, their relationships with others. AAC systems utilize assistive technology (AT) devices that range from no-tech to high-tech. Modifying young children's environments by using AT, defined as any tool, device, or adaptation that allows them more ways to gain access to the people, places, and setting where they can be exposed to typical developmental activities, increases opportunities for learning (Sadao & Robinson, 2010). Dell, Newton, and Petroff (2016) described the practical use of such devices in a classroom. Carpenter, Johnston, and Beard (2014) published a text for both in-service and pre-service educators to introduce AT. Jonge, Scherer, and Rodger (2007) provided an opportunity to try to understand the experiences of AT users in the workplace.

A widely used AAC tool, voice output communication aids (VOCAs) utilize single-level or multi-level outputs to convey sounds. Although there are a variety of VOCAs catering to students with different abilities and needs (Inclusive design research center, 2016; RESEARCH AUTISM, 2016), most devices are severely hampered by their low-output numbers and short lengths of time that they can record.

Approximately ten years ago, to address the above problems the present author started using Scan Talk dot codes developed by Olympus Co. (1999). Such codes transform voices and sounds

into two-dimensional dot codes directly outputted on ordinary paper. Students with severe hand, finger, or mental challenges, however, could not correctly trace Scan Talk codes using the Scan Talk Reader. The present author, therefore, used new dot codes developed independently by Gridmark, Inc. (2009) and Apollo Japan (2005) and conducted school activities with original handmade teaching materials overlaid with these dot codes. In our work, just touching the "invisible" dot codes printed on the paper or symbol icons by using a sound pen clearly reproduces voices and sounds. By using the identical sound pen or a scanner pen connected to a tablet or PC, multimedia sources such as movies, Web pages, and PowerPoint files, in addition to voices and sounds, can be reproduced on its screen.

In this article, state-of-the-art dot code technology is outlined, and basic information regarding the creation of original handmade materials using dot codes and the use at both general and special needs schools is presented.

BACKGROUND

Outline of Dot Code Technology

GridOutput Dot Codes

GridOutput (Gridmark, 2009) is a set of novel two-dimensional codes comprising extremely small dots. Such dot codes can invisibly overlay any graphically printed letters, photos, and illustrations with no impact on the designed visual images, meaning that letters, photos, and illustra-

tions can be changed into information-trigger icons. A maximum of four voices and sounds can be linked to each icon, as well as other media such as movies, Web pages, and PowerPoint files. Simply touching the dot codes printed on ordinary paper with a sound pen (e.g., G-Talk or G-Speak) or a scanner pen (e.g., G-Pen) enables students to directly access the corresponding digital information.

To print document content that includes the “invisible” GridOnput dot codes, industry-standard Cyan-Magenta-Yellow-Black (CMYK) processes are required. More specifically, carbon ink that absorbs infrared rays is used only for dot code printing, while non-carbon ink is used to print graphics. The sound and scanner pens read the invisible dot codes using built-in infrared cameras.

ScreenCode Dot Codes

Developed by Apollo Japan (2005), ScreenCode is a microscopic barcode that allows developers to map data onto printed surfaces, such as printed documents. Unlike other data-coding systems that map data onto expensive metallic-based inks, ScreenCode allows users to map data using regular ink similar to that found in household printers. Apollo Japan has recently published their ScreenCode smartphone and tablet application in which the built-in camera of the smartphone and tablet, using a specially designed lens, can identify ScreenCode and reproduce linked multimedia such as movies, audio files, photos, and Web pages on their screen.

Other Dot Codes

Developed by DENSO WAVE, Inc., the well-known QR code (1994) is an open code that anyone is allowed to use; it has become quite widespread as Micro QR codes, iQR codes, SQRCs, and Frame QRs. These QR codes incorporate high-level features such as illustrations, letters, and logos, and provide high-capacity data encoding, small printout sizes, Japanese Kanji

and Hirakana character support, dirt and damage resistance, 360-degree readability, and structured appending features. DENSO WAVE distributes their QR-code-generation programs, QRdraw Ad (DENSO WAVE, 2016a) and QRmaker Ad (DENSO WAVE, 2016b), free of charge.

Developed by TOPPAN TDK LABEL Co., LTD, the Z code (TOPPAN TDK LABEL, 2016) is composed of tiny dot codes and does not spoil illustrations or texts. The Z code is now used to integrate digital media with analog media and assess whether printed matter is genuine or false.

Developed by ZAK Co., the Simple Microdot Code (SMC) (ZAK, 2009) creates a sophisticated algorithm to print data on a 2.5 mm² area that can be decoded with an optical device. These tiny swatches of encoded dots can also be used to trigger electronic links, applications, databases, etc. The SMC dot code patterns can easily be integrated into a print file by end users using any of today’s desktop publishing software that support CMYK printing processes, thus allowing SMC dot code patterns to be placed anywhere on a printed surface, either as a standalone pattern or embedded in a directional graphic element.

Franklin Electronic Publishers, Inc. developed AnyBook (Franklin Electronic Publishers, 2016) using dot code technology, which is a set of reading enhancement products that allow children to hear their parents or a loved one reading any book for a more intimate learning experience, even when these readers are not physically present. These products consist of a reading pen that enables the words and pictures to talk with any book using vocal recording software. A variety of AnyBooks are available; for example, holding up to 200 h of audio, AnyBook Anywhere (DRP-5100) has 420 reusable stickers with pre-recorded sounds that are created using a special glue that does not harm the pages of the books, a headphone jack for private listening, and backup and sharing capabilities with AnyBook Case.

Afaya Technology Co., Taiwan, has developed a dot code incorporating charge coupled device scanning technology, which identifies the printed

K-layer that hides index data in two-dimensional bar codes of four to five ink-coated sheets of paper or books. This technology also includes an MP3 database and a visual-authoring tool. They sell a variety of Afaya-Pens (2016) in cooperation with their business partners and have produced many publications with pre-printed dot codes.

In Japan, under the new Course of Study (revised in 2008), “Foreign Language Activities” in the 5th and 6th grades are required at all elementary schools. The overall objectives of these new activities are “to deepen the understanding of languages and cultures through the experience of foreign language learning; to encourage efforts towards communication; to familiarize children with foreign language sounds and fundamental expressions; and to develop a foundation of basic communication abilities.” (TOKYO SHOSEKI, 2016a) With this as an opportunity, many English textbooks, side-readers, and picture books overlaid with dot codes have entered the market (EC kids, 2016; ECC, 2016; Kumon, 2016; Pendoku, 2016; Sailor, 2016, TAKAHA, 2016; TOKYO SHOSEKI, 2016b). The hope here is that these new foreign language activities may help students acquire conversational English skills more smoothly.

MAIN FOCUS OF THE ARTICLE

Each and every student has individual hopes, needs, and desires, as well as a history of learning. Therefore, schoolteachers ideally should develop original independent content suitable for each individual. Given the challenges in doing so, easy-to-handle and less-costly software and hardware solutions are crucial to schoolteachers for creating original content for students in their class (Ikuta et al., 2013, 2015, 2016).

We have been collaborating with two Japanese venture business companies, Gridmark, Inc. and Apollo Japan, to help students, especially those with various handicaps, by developing original handmade content and conducting fruitful school activities. Gridmark, Inc. has kindly provided us,

free of charge, with their valuable software for creating handmade content for sound pens, allowing us to distribute it to collaborating schoolteachers.

We recently created an original sheet (called a magical sheet) overlaid with GridOnput dot codes. Each sheet, as shown in Figure 1, has 117 “Post-it” sticker icons, which can be taken off and pasted elsewhere. Up to four audio clips can be linked to each icon by using the specially designed software (*NANA.exe*), which was kindly given to us by Gridmark. Schoolteachers can now create original handmade content and conduct school activities in their class without buying anything, instead of using a specially designed linkage table file, *filelist.csv*, that can be opened and edited with Microsoft Excel. In this file, each mp3 audio file is linked to the corresponding dot code icon’s number. The content folder, produced by double-clicking *NANA.exe*, can then be copied onto the Micro SD card of a sound pen. Touching the symbol icons pasted just beside the illustrations and texts with a sound (G-Talk or G-Speak) or scanner pen (G-Pen) plays back the original voices and sounds, as well as multimedia such as movies on the screen of a tablet or PC.

The executable file that can reproduce multimedia has also been distributed to collaborating schoolteachers free of charge; as such, they can edit a specially designed *URLLIST.csv* and replace old movies with new ones. Here, *NANA.exe*, *filelist.csv*, and *the folder* including an executable to reproduce multimedia are sent on request free of charge.

Gridmark, Inc. has just published their software (i.e., *SoundLinker* and *FileLinker*) and make it available to the market; their software can easily link audios and multimedia to dot code symbol icons. The executable file, created with *FileLinker*, enables the *sound* pen, G-Speak, connected to your tablet or PC to reproduce the multimedia sources like movies in addition to the *scanner* pen. More costly software called the *Gridmark Authoring Tool (GAT)*, which can directly overlay the GridOnput dot codes on illustrations or text of a PDF file, is also available now; unfortunately,

Figure 1. Magical sheet with GridOnput dot code symbol icons, G-Speak, and G-Pen



this *GAT software* requires a costly color printer to print the overlaid file.

Apollo Japan has recently published a new application called *ScreenCode*, which can be used with an iPhone or iPad with a built-in camera; note that the built-in camera must have a specially designed small lens to reproduce multimedia such as movies, audio files, and photos already saved on the iPhone or iPad. Entering a uniform resource locator (URL) with a software keyboard also reproduces Web pages. As shown in Figure

2, the present author has started to distribute sheets with *ScreenCode* dot codes free of charge, as schoolteachers do not need to buy the costly software, *Garyu Tensei for Smartphone*. Clearly, *ScreenCode* is a very attractive application for connecting to digital media from a smartphone to the dot codes on paper. University students are now developing original handmade content for younger students at both general and special needs schools, as detailed further in the next section.

Figure 2. Handmade sheet generated by *ScreenCode* and a specially designed lens



SOLUTIONS AND RECOMMENDATIONS

Original Handmade Content and School Activities

The teachers at Honmachidahigashi elementary school in Machida, Tokyo, conducted an activity using dot code symbol icons, as shown in Figure 3, for a 5th grade male student with a reading disability. At the school, the homeroom teacher gave the student a test with a worksheet after finishing each unit, but the student consistently could not answer anything. The homeroom teacher could not understand this.

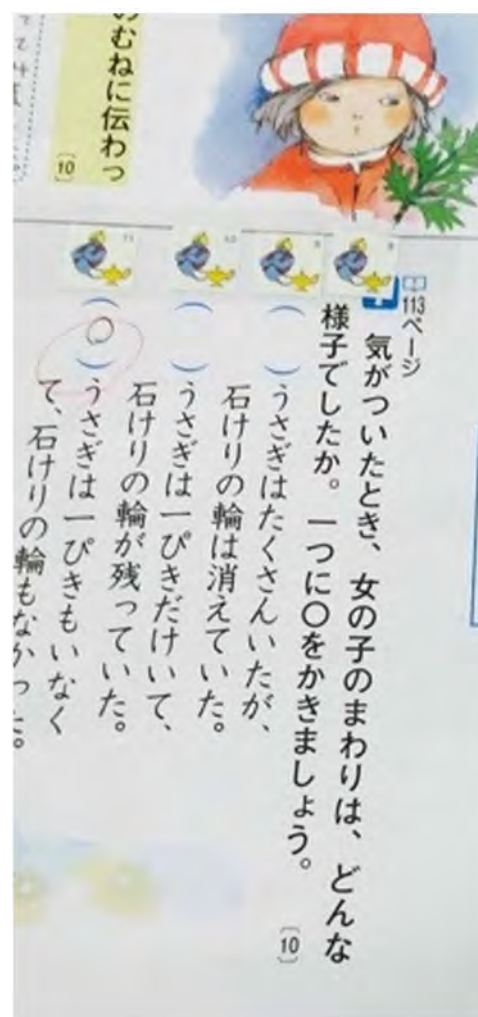
We provided a sound pen to the homeroom teacher, and she recorded every question using her voice on the sound pen, linking it to the corresponding symbol icons beside each question. The student could now listen to the audio with a sound pen instead of reading by himself.

As the first achievement since the given activity started, the student earned a grade of 70; all of the teachers at the school were at first surprised, but soon realized he had a reading disability. They confirmed the importance of knowing and understanding the difficulties each student faces and trying to remove such difficulties with the help of teaching aids and tools. The homeroom teacher now continues to create suitable content for this student.

A student of Otsuma Women's University, Miki Endo, also created content; as shown in Figure 4-a, younger students can learn about our solar system (NASA 2016) using audio clips and movies. The GridOnput dot codes were overlaid on each planet using the *GridLayouter* software (which has unfortunately been withdrawn from the market); Gridmark, Inc. has just published new software with similar functionality (i.e., the *GAT*).

Younger students at several elementary schools near the university enjoyed the content very much; by touching each planet with sound and scanner pens, they were able to listen to explanations for each planet and watch movies of the correspond-

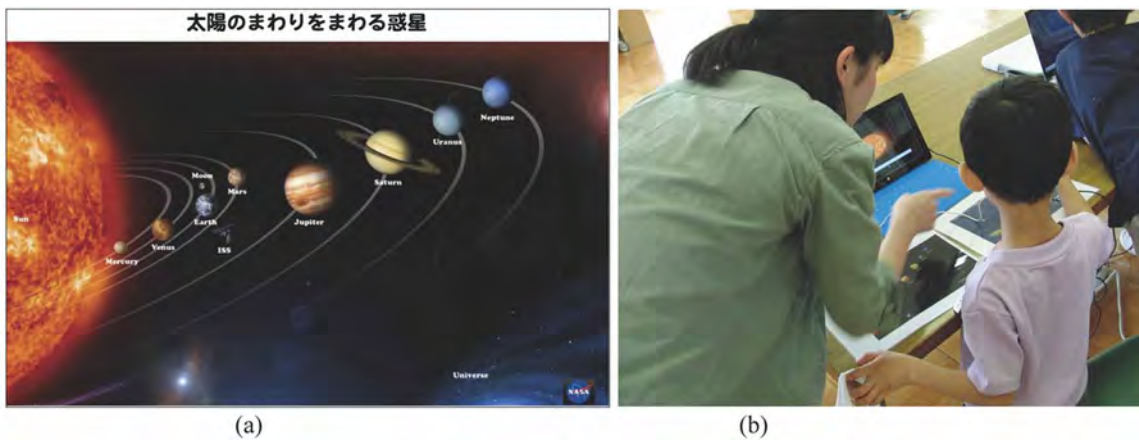
Figure 3. Worksheet with GridOnput dot code symbol icons



ing planets. A snapshot of such activities is shown in Figure 4-b.

Teachers at the Osaka Prefectural Special Needs School for the Visually Impaired also created handmade content for blind students, in which even student with total blindness could find the symbol icons and listen to the voices with a sound pen by touching the symbol icons beside the three-dimensional stickers with numeric numbers. A snapshot of these activities is shown in Figure 5. Here, the teacher intentionally placed dot code symbol icons all over the sheet such that the students could enjoy finding them and listening to a fortune slip with voices. The teacher also

Figure 4. (a) Handmade sheet with GridOnput dot codes to learn about our solar system and (b) activities at elementary school



created a booklet that could be used at a school event focused on music appreciation.

The present system using a sound pen and handmade content was also introduced to special training school students for the visually impaired. A student with weak eyesight said, with deep emotion, “I am always afraid how I can read picture books for my children in the future. When I try to read a picture book for my child, I have to hold the book to my eyes and cannot look at my child’s face at all; however, now, with the help of this

sound pen, I record my voice beforehand using the recording functionality of the sound pen and can really read a picture book to my child.” The sound pen has been able to realize a dream for people with weak eyesight and the totally blind.

The *ScreenCode* application for iPhone and iPad, recently developed by Apollo Japan, is now put on the market, with multimedia such as movies, photos, audio clips, and Web pages being able to be linked to ScreenCode dot codes overlaid on ordinary paper. A specially designed lens is

Figure 5. Activities with handmade sheet for a totally blind student at the Osaka Prefectural Special Needs School for the Visually Impaired



attached to a built-in camera of the iPhone and iPad; this camera focuses on the dot codes. The augmented reality application *Aurasma* can also reproduce multimedia such as movies by focusing on the trigger illustration with the built-in camera of an iPhone or iPad. Focusing on the trigger illustration, however, is not so easy; a smaller illustration on paper cannot be focused. On the other hand, the *ScreenCode* application, with a built-in camera, can focus fairly well on smaller sized dot code symbols.

Apollo Japan is now distributing “a sample notebook with Screen Code” overlaid with ScreenCode on every page of a blank booklet in which you can write your own messages and link multimedia to the pages quite easily. The present author with university students is now creating handmade worksheets (overlaid with ScreenCode) for younger students at general schools and preparing school activities, as depicted in Figure 6. The present author also created a specially designed sheet with ScreenCode dot codes, as shown in Figure 2, created using the *Garyu tensei for Smartphone* software, and then started to share them with schoolteachers free of charge. Each symbol icon in the sheet can be cut-and-pasted

on any item and linked with multimedia such as movies, audio clips, photos, and Web pages. Figure 7 shows snapshots of using the ScreenCode dot codes with an iPhone; note that the built-in camera is attached with a specially designed lens. Apollo Japan is now developing the system to focus the ScreenCode without using the specially designed small lens; this *ScreenCode* application may promote more widespread use of smartphones and tablets in the classroom.

FUTURE RESEARCH DIRECTIONS

The present author creates original handmade content and conducted many activities at both general and special needs schools in collaboration with schoolteachers not only in Japan but also in the United States, the United Arab Emirates, China, and Oman. All the schoolteachers wish to use easy-to-handle and less-costly software to create their original contents for each student in their class; it is difficult for schoolteachers to obtain the funds to buy costly software and hardware to do so. The software presented in this article, for example, *Garyu Tensei for Smartphone* developed

Figure 6. Handmade content with ScreenCode dot codes



Figure 7. (a) Special lens attached to an iPhone and (b) snapshot of iPhone screen (school activity movie), focusing on ScreenCode dot codes



by Apollo Japan, costs more than 1,000 USD; thus, schoolteachers cannot buy it. Even cheaper software, such as *SoundLinker* and *FileLinker* developed by Gridmark, Inc. for linking audio clips and multimedia files to GridOnput dot codes, are still costly. Collaboration with university professors is crucial to start and continue developing easy-to-handle and less-costly software and creating outstanding original handmade contents for each student in a different set of needs and desires, degree of disability, and learning history.

The present author has conducted many long-term activities with schoolteachers at both special needs and general schools, and performed several assessments of such school activities using the original handmade teaching materials with the present multimedia-enabled dot codes (Ikuta et al., 2013, 2015, 2016). Each student has responded to these activities with confidence and joy, providing that such long-term activities could also improve the students' learning. The various effects noticed during the running of the school activities should be carefully analyzed and classified more in details for each kind of disability, so as to be able to focus the activities more precisely; such analysis is crucial to clarify the usefulness of assistive technology in the classroom.

Through many school activities performed at general and special needs schools, it was noticed that the sound pens require some changes. The positions of the power and audio volume buttons on the new sound pen G-Speak are located just at the pen center, and students with disabilities and in lower grades often make errors. The present author therefore suggest that Gridmark moves these buttons to the side of the pen. The sound pens with Wi-Fi and Bluetooth connectivity could be used in the near future classroom. More effective uses of such new sound pens with original contents should be explored by the researchers in collaboration with schoolteachers in many parts of the world.

The help of university students is also crucial to developing original handmade content, since schoolteachers, especially in Japan, always have too much to do in their daily jobs. Collaboration with schoolteachers is very important and attractive for university students, especially those who wish to become teachers; these university students can learn much from current schoolteachers.

Good relationships and collaborations with developers, Gridmark, Inc., and Apollo Japan are also crucial to continuing and developing the present project. Also, funds provided by the

Ministry of Education, Culture, Sports, Science and Technology, Japan, and the Institute of Human Culture Studies, Otsuma Women's University are essential for us to purchase software and hardware, and provide them to schoolteachers free of charge.

CONCLUSION

In collaboration with schoolteachers, a variety of original handmade content has been created and various activities at both general and special needs schools have been conducted. The present author has created specially designed sheets with "Post-it" symbol icons overlaid with dot codes; such sheets have been provided with sound pens and software to collaborating schoolteachers free of charge, so that they can create handmade content for students in their classrooms themselves without purchasing anything.

We asked schoolteachers to create original handmade content for students in their own classes instead of purchasing teaching materials already on the market, since each student has his or her own needs and desires and own history of learning. Perfect teaching materials and tools for one student do not always perform well with other students. Easy-to-handle and cheaper software and hardware are required if the schoolteachers are to proceed with these voluntary activities. Collaboration with software and hardware companies and their kind help, therefore, are an important issue for the development of focused teaching materials and school activities.

Support given by the Grants-in-Aid for Scientific Research and the Institute of Human Culture Studies, Otsuma Women's University, is also important for ensuring the continuation of larger collaborations with worldwide schoolteachers. The help of university students is also a strong motivation for schoolteachers who wish to develop individual content for each of their students. Collaborations with foreign schoolteachers in the United States, United Arab Emirates, and Oman

are also very important to extending and assessing present school activities performed in the classroom with new information communication technologies.

ACKNOWLEDGMENT

The author thanks "Grant-in-Aids for Scientific Research" (C) (#22530992 and 16K04844) and the "Institute of Human Studies" (s2605, s2713, and s2810) at Otsuma Women's University. The author also thanks Fumie Shimada, Chikako Kaneko, Miki Endo, Erika Kaneko, and Noriko Kuroda for their collaborations. The author would like to heartily thank Enago (www.enago.jp) for their English language review.

REFERENCES

- Afaya Japan. (2016). *Afaya*. Retrieved August 25, 2016, from <http://www.afaya.co.jp/afayapen.html>
- Apollo Japan. (2005). *ScreenCode*. Retrieved August 25, 2016, from http://www.apollo-japan.ne.jp/index_en.html
- Carpenter, L. A. B., Johnston, L. B., & Beard, L. A. (2014). *Assistive Technology: Access for All Students* (3rd ed.). Upper Saddle River, NJ: Pearson Education.
- de Jonge, D., Scherer, M. J., & Rodger, S. (2007). *Assistive Technology in the Workplace*. St. Louis, MO: Mosby.
- Dell, A. G., Newton, D. A., & Petroff, J. G. (2016). *Assistive Technology in the Classroom: Enhancing the School Experiences of Students with Disabilities* (3rd ed.). Upper Saddle River, NJ: Pearson Education.
- DENSO WAVE. (2016a). *QRdraw Ad*, Retrieved August 25, 2016, from https://www.denso-wave.com/en/adcd/support/technical/software_intro/qrdraw_ad.html

- DENSO WAVE. (2016b). *QRmaker Ad*. Retrieved August 25, 2016, from <https://www.denso-wave.com/ja/adcd/product/software/building/qrmaker-ad.html>
- EC Kids. (2016). Retrieved August 25, 2016, from <http://www.eckids.jp/pen/>
- ECC Co. (2016). *ECC Funtalk*. Retrieved August 25, 2016, from <http://www.kids.ecc.jp/course/funtalk.html>
- Franklin Electronic Publishers, Inc. (2016). *Anybook*. Retrieved August 25, 2016, from <http://www.anybookreader.com>
- Gridmark Inc. (2009). *GridOnput*. Retrieved August 25, 2016, from <http://www.gridmark.co.jp/english/index.html>
- Ikuta, S., Endo, E., Nemoto, F., Kaiami, S., & Ezoe, T. (2013). School Activities Using Hand-made Teaching Materials With Dot-codes. In D. G. Barres, Z. C., Carrion, & R. L.-C. Delgado (Eds.), *Technologies for Inclusive Education: Beyond Traditional Integration Approaches* (pp. 220-243). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-2530-3.ch011
- Ikuta, S., Ishitobi, R., Nemoto, F., Urushihata, C., Yamaguchi, K., & Nakui, H. (2016). Handmade Content and School Activities for Autistic Children with Expressive Language Disabilities. In Y., Katz (Ed.), *Supporting the Education of Children with Autism Spectrum Disorders*. Hershey, PA: IGI Global.
- Ikuta, S., Morton, D., Kasai, M., Nemoto, F., Ohtaka, M., & Horiuchi, M. (2015). School Activities with New Dot code Handling Multimedia. In L. Lennex, & K. Nettleton (Eds.), *Cases on Instructional Technology in Gifted and Talented Education* (pp. 314-340). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-6489-0.ch015
- Inclusive design research center (OCAD University) (2016). *Voice Output Communication Aids*. Retrieved August 25, 2016, from <http://idrc.ocad.ca/index.php/resources/technical-glossary/49-voice-output-communication-aids>
- Kumon Institute of Education Co. Ltd. (2016). *E-Pencil*. Retrieved August 25, 2016, from http://www.kumon.ne.jp/eigo/index.html?lid=eigo_002
- NASA. (2016). *Image Galleries*. Retrieved August 25, 2016, from <http://www.nasa.gov/multimedia/imagegallery/index.html>
- Olympus Co. (1999). *ScanTalk, Scan Talk Reader, & Sound Print Kobo*. Retrieved August 25, 2016, from <http://cs.olympus-imaging.jp/jp/support/cs/Scantalk/>
- Pendoku.com. (2016). Retrieved August 25, 2016, from <http://www.pendoku.com>
- QR code.com. (1994). *History of QR Code*. Retrieved August 25, 2016, from <http://www.qrcode.com/en/history/>
- Research Autism Improving the Quality of Life. (2016). *Voice Output Communication Aids and Autism Publications*. Retrieved August 25, 2016, from <http://researchautism.net/autism-interventions/types/alternative-and-augmentative-communication/voice-output-communication-aids-and-autism/voice-output-communication-aids-and-autism-publications>
- Sadao, K. C., & Robinson, N. B. (2010). *Assistive Technology for Young Children: Creating Inclusive Learning Environments*. Paul H Brookes Publishing Co.
- Sailor. (2016). *Sound pen*. Retrieved August 25, 2016, from <http://www.sailor.co.jp/electronic-bungu>
- Takaha. (2016). *Voice Rev*. Retrieved August 25, 2016, from <http://takaha-j.jp/voice-rev.html>

Tokyo Shoseki. (2016a). *2008 Course Of Study for Junior High Schools Foreign Languages (English) & Course of Study for Elementary Schools Foreign Language Activities*. Retrieved August 25, 2016, from https://ten.tokyo-shoseki.co.jp/ten_download/dlf41/cey70148.pdf

Tokyo Shoseki. (2016b). *Disney English*. Retrieved August 25, 2016, from <http://www.tokyo-shoseki.co.jp/books/80742/>

Toppan TDK LABEL Co. Ltd. (2016). *Z code*. Retrieved August 25, 2016, from http://www.toppan-tl.co.jp/products/de_eq/Zc.html

ZAK Co. (2009). *Simple Micro Dot Code (SMC)*. Retrieved August 25, 2016, from <http://www.zak-corp.jp/en/>

ADDITIONAL READING

Beukelman, D. R., & Mirenda, P. (2013). *Augmentative and Alternative Communication: Supporting Children and Adults with Complex Communication Needs* (4th ed.). Baltimore, MD: Paul H. Brookes Publishing Co.

Boone, D. R., McFarlane, S. C., von Berg, S. L., & Zraick, R. I. (2014). *Voice and Voice Therapy* (9th ed.). Upper Saddle River, NJ: Pearson Education.

Bouck, E. C. (2016). *Assistive TECHNOLOGY*. Los Angeles, CA: Sage Publications.

Cook, A. M., & Polgar, J. M. (2012). *Essentials of Assistive Technologies*. St. Louis, MO: Mosby.

Cook, A. M., & Polgar, J. M. (2015). *Assistive Technologies: Principles and Practice* (4th ed.). St. Louis, MO: Mosby.

Hall, T. E., Meyer, A., & Rose, D. H. (2012). *Universal Design for Learning in the Classroom: Practical Applications*. New York, NY: The Guilford Press.

Kaneko, S., Ohshima, M., Takei, K., Yamamoto, L., Ezoe, T., Ueyama, S., & Ikuta, S. (2011). School activities with voices and sounds: Handmade teaching materials and sound pens [in Japanese]. *Computers & Education*, 30, 48–51.

Loncke, F. (2014). *Augmentative and Alternative Communication: Models and Applications for Educators, Speech-Language Pathologists, Psychologists, Caregivers, and Users*. San Diego, CA: Plural Publishing Inc.

Nemoto, F., Ishitobi, R., & Ikuta, S. (2016). Current status and issues of practical use of educational aids based on information communication technology at elementary and secondary schools: Through activities at special needs schools [in Japanese]. *Computers & Education*, 40, 38–43.

Theng, L. B. (2011). *Assistive and Augmentative Communication for the Disabled: Intelligent Technologies for Communication, Learning and Teaching*. Hershey, PA: IGI Global. doi:10.4018/978-1-60960-541-4

KEY TERMS AND DEFINITIONS

Assistive Technology: Assistive technology (AT) devices range from no-tech to high-tech and help to properly assess the correct communication systems for students with severe speech-language deficits; as such, professionals work with students, teachers, and individuals who interact with students on a daily basis.

Communication Aid: Communication aids based on augmentative and alternative communication (AAC) technologies have been widely used to provide students with severe speech, language, and communication difficulties the opportunity to achieve successful communication and relationships with others.

Dot Code: Invisible dot codes developed by Gridmark, Inc. are a novel two-dimensional code

technology consisting of extremely small dots practically imperceptible to the human eye. Each symbol icon can be linked to up to four audio clips. Multimedia such as movies can also be linked to the same symbol icon. A simple touch by sound and scanner pens on the symbol icon enables links between paper and digital content.

Handmade Content: *Original handmade* teaching content plays a key role in learning, as each student has different thoughts, feelings, needs, and desires. Independent teaching material should be prepared and used for each individual, especially in preschool and special needs classrooms.

Magical Sheet: This specially designed sheet has the order for each dot code imprinted on each symbol icon. The sheet has additional symbol icons with audio-recording functionality and mode changing. The icons can then be taken off and pasted onto a target object and touched with sound and scanner pens. The “magical sheet” enables the costly GridOnput system to be used at any school at a low cost.

Scanner Pen: Scanner pens such as the G-Pen can read dot codes printed on paper and play back multimedia (such as movies) on the screen of a G-Pen connected to a tablet or personal computer.

School Activity: School activities at special needs schools can be improved through the use of *original* and *individual* handmade teaching materials and aids suitable for each student with disabilities.

Sound Pen: Sound pens, such as G-Talk and G-Speak, reproduce original voices and sounds by simply touching the dot-coded symbol icons or the dot codes directly overlaid on the paper.

Voice Output Communication Aid: Voice Output Communication Aids (VOCAs) are widely used AT devices that utilize single-level or multi-level outputs to convey sounds. While a variety of VOCAs exist that cater to students of different abilities and needs, most devices are severely hampered by their low-output numbers and short recording times.

Semantically Enhanced Authoring of Shared Media

Charalampos A. Dimoulas

Aristotle University of Thessaloniki, Greece

Andreas A. Veglis

Aristotle University of Thessaloniki, Greece

George Kalliris

Aristotle University of Thessaloniki, Greece

INTRODUCTION

The rapid evolution of digital technology, among others, has revolutionized multimodal content production and distribution processes, propelling novel mediated communication services. Interactive media authoring and sharing technologies are currently being launched, bringing forward new ways of audiovisual (AV) content exchange. Web documentaries (web-docs) and hypermedia have appeared as a natural extension of filmed documentaries and digital TV, inheriting also some of their advantages. Narrative documentaries adopt AV mediated communication mechanisms that humans have been accustomed to be informed and communicate with each other, so that they are more informative and vivid compared with other documents (books, web-pages, multimedia, etc.). Thus, AV-documentaries, web-docs and generally interactive videos and hypermedia can be more easily distributed and attended from most ages and social groups (Dimoulas, Veglis, & Kalliris, 2015; Dimoulas, Kalliris, Chatzara, Tsipas, & Papanikolaou, 2014a; Kotsakis, Kalliris & Dimoulas, 2012; Matsiola, Dimoulas, Kalliris, & Veglis, 2015; Veglis, Dimoulas & Kalliris, 2016). Moreover, the continuous evolution of the computing power and the digital storage media favor digital video production and distribution. This is also fueled by the increased network speed, the efficiency of compression algorithms and the continuous

decrease of the corresponding costs (Kotsakis et al., 2012; Dimoulas, Kalliris & Veglis, 2014b). High quality AV capturing equipment is currently available at low cost and size as part of smart phones and other mobile computing terminals with inherent networking capabilities, allowing easy AV-content production, contribution and sharing (Atzori, Delgado & Giusto, 2012; Dimoulas & Symeonidis, 2015; Dimoulas et al., 2014a; 2014b; 2015; Sidiropoulos, Konstantinidis, Kotsakis & Veglis, 2015; Veglis et al. 2016; Vrysis, Tsipas, Dimoulas & Papanikolaou, 2015). In this context, more and more users are involved in the AV production and consumption chain, so that creative experience and AV media culture are cultivated. Nevertheless, AV media related achievements are still far from the progress that has been made in textual information management during the outspread of social media and Web 2.0 services.

Interactive services have also been introduced into the AV production industry, aiming at augmenting human-machine interaction (HMI). AV content is enhanced in functional and informative level, further stimulating users to actively participate in arousing interactive scenarios. While the transition from Web 2.0 to Web 3.0 is ongoing, intelligent AV-content processing and management services are pursued, facilitating users' involvement on media sharing, commenting and multichannel publishing. In this context, semantic annotation, social tagging and meta-

DOI: 10.4018/978-1-5225-2255-3.ch562

processing can be part of collaborative media syncing, editing and multimedia management, thus propelling more sophisticated authoring (and sharing) of semantically-enhanced media. Although these technologies are rapidly evolving, there are still open challenges regarding upcoming semantic web services (De Bra, Freyne, & Berkovsky, 2013; Dimoulas et al., 2014a; 2014b; 2015; Dimoulas & Symeonidis, 2015; Matsiola et al., 2015; Monaghan, Handschuh, & O'Sullivan, 2011; Veglis et al. 2016). This chapter examines current trends and future perspectives of semantically-enhanced media /multimedia (SeMM), considering all forms of non-linear storytelling, sharing and authoring (web-docs, multichannel media publishing, interactive videos, adaptive hypermedia and generally multimedia services). Background is presented providing basic definitions, involved technology, achieved progress and limitations. Recommendations and future research direction are then stated, aiming at serving a two-folded target: firstly, to present new, user-friendly forms of collaborative creativity, multimedia authoring and storytelling that current technology allows to be successfully deployed; secondly, to suggest innovative adaptation mechanisms that can be utilized in both the media production and consumption ends, allowing for intelligent media management and augmented semantic interaction services to be launched. In this context, the importance of SeMM toward the transition to the Web 3.0 era is revealed.

BACKGROUND

Nowadays, hybrid models of mixed Page-Based Media (PBM) and Time-Based Media (TBM) are the most common multimedia web content, combining both spatial and spatiotemporal arrangement of all media entities (i.e. text, images, audio, video). These so-called Multimodal Media Assets (MMA) (Veglis et al., 2016; Dimoulas et al., 2015) may include multichannel audio and video,

non-linear animations and photos, titles and more elongated textual descriptions that offer versatile presentation of information, enabling different views and audio-tracks selection, multilingual narration and subtitling. MMA schemes have many similarities to the early digital versatile disc (DVD-video) and are closer to the content entities and structures that contemporary web-docs and hypermedia utilize. The term hypermedia refers to hyperlinked multimedia, the same way that hypertext is about grouping relevant information in the form of hyperlinked PBM nodes. It is known that, while the term appeared in the early '90s, it was put aside since then, an issue that is associated to the broader meaning that the multimedia definition received. Thus hypermedia can be considered as a subcase of multimedia web services. Today, both terms are used somewhat as synonyms, with the multimedia definition to be wider, incorporating both TBM and MMA compositions with linear and non-linear navigation, but also referring to new, more sophisticated forms of creativity, digital storytelling and interaction. For this reason, the terms media and multimedia are mostly used in this chapter, instead of hypermedia. Obviously, there are still many technical issues and limitations associated with the involved diversities (i.e. different content types, formats and publishing channels, encoding and reproduction compatibilities, HMI and authoring technologies, semantics and meta-data management, etc.), so that SeMM is more than simple combinations of shared MMA entities.

One of the key terms of hypermedia is interactivity, which has advanced with the evolution of hypertext technology. Along with interactivity, new forms of digital storytelling are considered and pursued, aiming at providing rich media experience through highly valued Quality of Experience (QoE) media services having appropriate emotional impact on the users' side, thus stimulating their active engagement (Kalliris, Matsiola, Dimoulas & Veglis, 2014; Kotsakis, Dimoulas, Kalliris & Veglis, 2014; Matsiola et al.,

2015). Hence, content contribution and sharing models along with social tagging and semantic interaction services can be considered as contemporary interaction forms that can be very useful in various ways. Today, every Internet user acts both as a receiver and transmitter of information by generating and disseminating textual content, along with audio, video and generally multimedia components (Dimoulas et al., 2014b; 2015; Veglis & et al., 2016). This is also the case for the user generated content (UGC) in social networks.

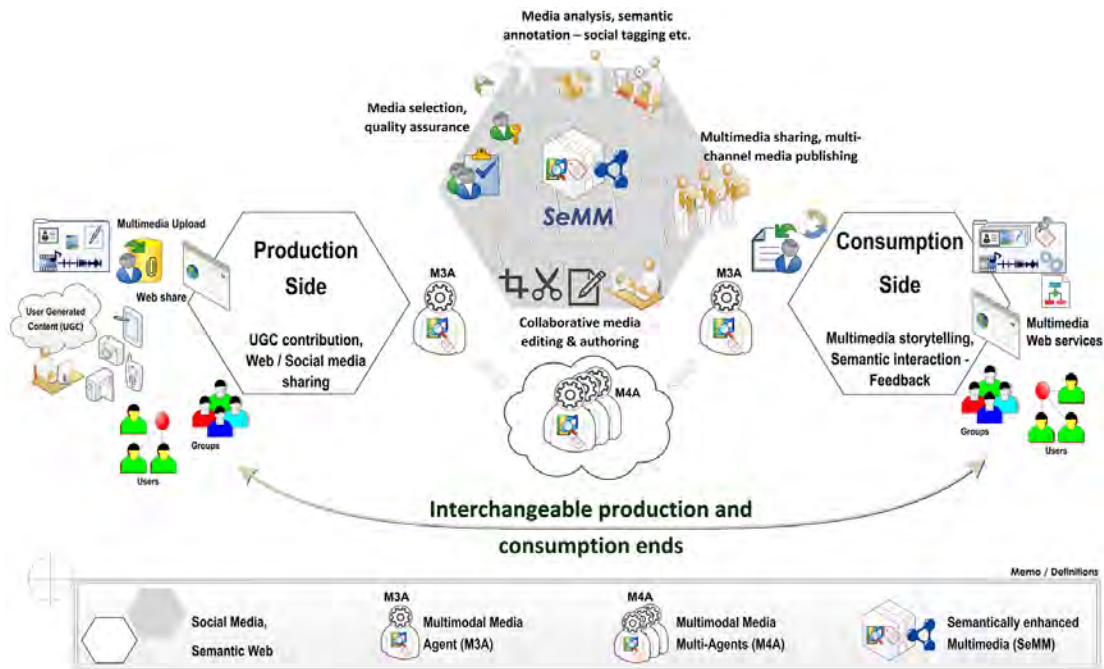
As previously implied, not all of the features that have been already achieved and adopted in textual information and PBM media can be easily applied and accommodated to the cases of TBM and MMA. Indeed, there are still many open issues regarding efficient AV content description and management automation, but also interoperability with transparent multimedia access. Hence, content-based searching and retrieval techniques are currently being implemented, using multimodal content detection, segmentation and summarization-highlighting algorithms along with intelligent processing and hybrid expert systems for pattern recognition and semantic analysis purposes (Cordara, Bober & Reznik, 2013; Dimoulas, Avdelidis, Kalliris & Papanikolaou, 2008; Dimoulas, Papanikolaou & Petridis, 2011; Diou et al., 2010; Kotsakis et al., 2012; Vegiris, Dimoulas & Papanikolaou, 2009). Furthermore, while much progress has been made in textual data mining approaches, semantic interaction and conceptualization efforts are still left behind, not only in AV-related media, but also for text and general PBM information, as well (Dimoulas et al., 2014b; Diou et al., 2010; Symeonidis & Mitkas, 2005; Veglis et al., 2016). This hysteresis in the progress of AV versus textual information processing also explains why hypermedia did not flourish as hypertext. As it is analyzed later on, intelligent MMA processing is also considered to be a crucial prerequisite for the deployment of largescale semantically-enhanced collaborative multimedia authoring and sharing.

SEMANTICALLY ENHANCED MULTIMEDIA

Multimedia Semantics and Authoring Tools: Issues, Controversies, Problems

The term Multimedia (instead of media) in the SeMM model emphasizes the fact that multiple content types and communication means are usually encountered and involved in the associated sharing and authoring processes (without the necessity for each one of them to encompass all kinds of media entities). Many Web-TV platforms and popular AV social media (i.e. Youtube, Vimeo, Ustream, SoundCloud, etc.) have propelled AV-content contribution and dissemination (Dimoulas et al., 2014b; 2015; Tsipas, Zapartas, Vrysis & Dimoulas, 2015). These web environments are currently massively utilized from a variety of simple users, experts and media professionals for both AV content consumption and production, offering on demand upload-access and live streaming services, along with online processing, annotation, commenting, authoring and sharing throughout the entire Internet community (Dimoulas et al., 2014b; 2015; Spyridou, Matsiola, Veglis, Kalliris, & Dimoulas, 2013; Veglis et al., 2016). The availability of multiple broadcasting media led to the implementation of multichannel publishing models, offering alternative navigation routes and new forms of interactivity. In this context, collaborative media authoring and sharing can be considered as a multichannel publishing sub-case, where semantic meta-data and content quality metrics can be exploited toward intelligent cross-media publishing and mediated communication models (Kalliris et al., 2014; Kotsakis et al., 2014; Veglis et al., 2016). In most cases, these processes need to be expedited with the use of intelligent systems and media agents, so that content production, processing and authoring guidelines can be provided as they needed (especially from the average user) along with content recognition and semantic

Figure 1. The processes of semantically-enhanced multimedia production, authoring and sharing



labeling automations (Dimoulas & Symeonidis, 2015; Sidiropoulos et al., 2015; Tsipas et al., 2015; Vrysis et al., 2015). As it is depicted in Figure 1, a recent modeling approach is the utilization of the so-called multimodal media agents (M3A) and multi-agents (M4A), which refer to media services able to sense its surrounding environments (at multiple levels, semiotic modes and/or communication channels) and autonomously adapt on their changes, responding on behalf of other (software) entities and human interactions (Matsiola et al., 2015). Hence, user-related adaptations and social networking intelligence are incorporated into mediating mechanisms that can facilitate multimedia communication, authoring and sharing, thus offering augmented and semantically-enhanced interactions, both at the media production and consumption ends.

In any case, multimedia production involves two complementary phases: content production (including UGC contribution) and authoring. The former can be delivered with a variety of tools and services, namely: discussion groups, weblog, collaborative publishing, Peer-to-Peer, XML

Syndication (Dimoulas et al., 2014; Monaghan et al., 2011). The latter is usually accomplished with the use of software programming, declarative languages, web developing and scripting tools (Dimoulas et al., 2014b; 2015; Dimoulas & Symeonidis, 2015; De Bra et al., 2013; Dos Santos & Muchaluat-Saade, 2012; Matsiola et al., 2015). These tools have been utilized during the last years for the development of user friendly web-doc authoring platforms that offer easy operation, increased creativity choices, with different distribution and pricing models. As a consequence, many worth-mentioning featured web-docs have been appeared during the last decade, which fused and projected new levels of creativity, user interaction and participation, outclassing multichannel media publishing. User engagement is enhanced by means of adaptation and personalization, collaborative social tagging, resources recommendation and semantic interaction. The latter are dominant in the so-called adaptive hypermedia systems (AHS), which are very common in mediated learning and public administration services, narrative tools and serious

gaming approaches, AV heritage and cultural projects (De Bra et al., 2013; Dimoulas et al., 2014a; Dos Santos & Muchaluat-Saade, 2012; Kalliris et al., 2014; Kotsakis et al., 2014). In addition, much effort is put for inter-platform compliance and interoperability, although there are still issues that need to be addressed, especially for the majority of upcoming Web 3.0 services (Dimoulas & Symeonidis, 2016; Dimoulas et al., 2013; 2014; 2015; Veglis et al., 2016). Hence, while multimedia semantics require massive /diverse content labeling and meta-data processing through time-consuming and tiresome users-enabling tasks, media authoring and browsing decisions can offer such kind of ground-truth information in an easier, entertaining and self-extracted way. These potentials of semantically enhanced multimedia (through collaborative authoring and sharing models) have not been fully revealed or exploited yet, and are attempted to be enlightened in the current chapter.

Solutions and Recommendations: Modeling Semantically-Enhanced Authoring of Shared Media

SeMM authoring involves the incorporation of different MMA into various dialogic modes and HMI scenarios, including media sharing, UGC posts and comments, which is usually accomplished through the collaboration of many different users and groups. Semantic interaction, content annotation and social tagging that are usually involved in Web 2.0/3.0 services can be augmented by and/or combined with the multimedia authoring choices. Thus, production organization and scripting can be part of the annotation and visualization meta-information (Ma, Liu, Wang, Teng, & Dai, 2012), while editing and authoring information can be logged into semantic meta-data (Dimoulas et al., 2014a; 2015). Most of the AV-authoring tasks (i.e. AV-production, clip selection-editing, navigation flow programming, etc.) can contribute to content description and meta-processing (i.e. keyword labeling, semantic tagging, topic and context classi-

fication, event detection-segmentation, long-term story summarization-highlighting, etc.). In fact, these tasks can now be easier and more creatively completed as part of the authoring process and not as separate obligatory annotation procedures. Similarly, end-users' interactions can also provide collaborative feedback for hypermedia adaptation and personalization, search-log-based annotations, social tagging and media recommendation (Dos Santos & Muchaluat-Saade, 2012; Monaghan et al., 2011; Tsikrika, Diou, de Vries & Delopoulos, 2011; Tsirellis & Delopoulos, 2011). Hence, storyline extraction, semantic conceptualization, intelligent content-based searching and retrieval can be supported, offering increased media management automation. Along with these, state of the art pervasive context- and location-aware services, serious games, ubiquitous mobile computing and augmented reality further enhance HMI experience, extending users' participation through a plethora of platforms and interface terminals (Atzori et al., 2012; Cordara et al., 2013; Dimoulas et al., 2014a; 2014b; 2015; Dimoulas & Symeonidis, 2015; Kalliris et al., 2014; Kotsakis et al., 2014; Matsiola et al., 2015).

A semantically enhanced media authoring model (SeMAM) is proposed in Figure 2, where the term authoring has a broader meaning, incorporating all the processes of content (and meta-data) production, processing (selection, editing, etc.), programming (multimedia authoring) and sharing, thus integrating the previously mentioned multimodal semantic interaction, context-awareness and adaptation features. Based on this model, MMA entities are assembled along with semantic and storytelling meta-data, forming the Semantically-enhanced Multimedia Assets (SeMMA), which correspond to the SeMM resources involved in an authoring project. SeMMA incorporate basic multimedia communication and interaction functionalities and can be combined as building blocks of semantically-enhanced multimedia storytelling services (SeMMSS), favoring the construction of semantic composition-decomposition models. Thus, semantic meta-data of all the edited-

authored assets are also assembled, generating a composite semantic outcome. In this context, a complex semantic analysis tasks is delivered via semantic decomposition-composition of all the combined hypermedia assets and their editing (/authoring) decision lists (Dimoulas et al., 2014a; 2015). The construction of SeMM/SeMMSS repositories promotes sophisticated management with improved content selection, adaptation and reuse capabilities in the entire media life-cycle. The above are graphically presented in the upper part of Figure 2, where tools and main processes of the entire media lifecycle are depicted. It is obvious that an infinite number of users and user-groups can participate throughout the end-to-end shared-media chain, where the separate roles of producers and consumers no longer exist.

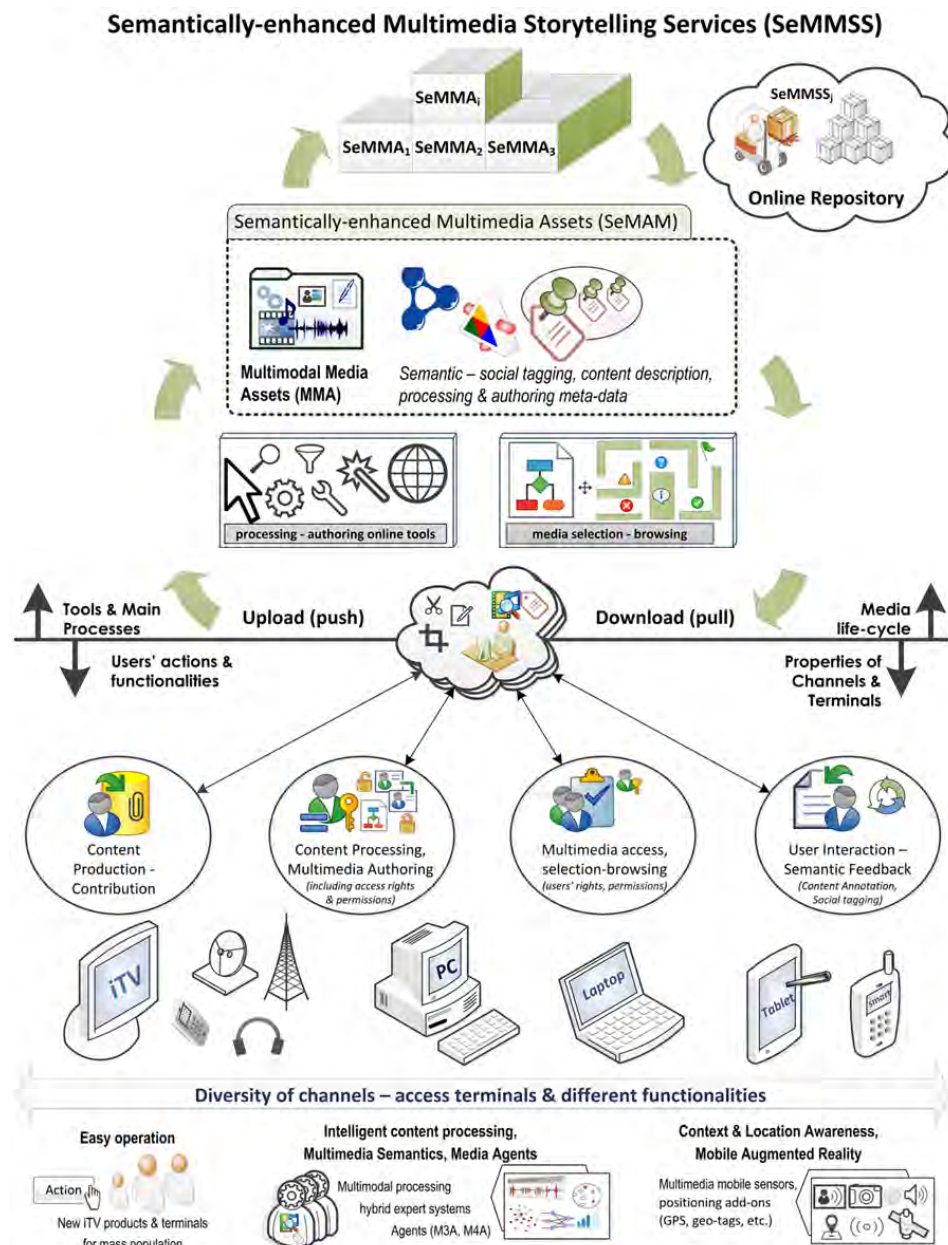
A diversity of devices is also exploited offering different features (depicted in lower part of Figure 2). Specifically, servers and powerful cloud computing terminals can be used for content storage, processing and intelligent semantic analysis services. Mobile terminals and their multimedia sensors are applicable in pervasive and/or immersive applications, while future iTV sets will allow for easy multimedia browsing that can be offered to the mass population without requiring special skills and knowhow. Furthermore, intelligent systems and media agents (M3A, M4A, etc.) are expected to play a very crucial role, sensing and adapting to the users' needs, while diminishing large-scale content and meta-data heterogeneity. Hence, content recognition and semantic conceptualization, voice commanding and multilingual adaptation or even virtual presentation through avatars and digital characters can be deployed, diminishing interfacing limitations (especially in mobile terminals), while facing various issues related to users' privacy. In addition to the versatile and/or multilingual information structuring, SeMMSS offer new presentation capabilities. Following the citizens-journalism paradigm, multiple UGC contributions on the same story can provide different points of view, offering increased user experience (i.e. multi-view selection, augmented

projection, 3D-reproduction, panoramic and time-lapse virtual navigation, etc.). This can be very constructive in heritage projects, promoting multichannel content enhancement-restoration, archiving and documentation. Again, all these processes and their multimedia consumption counterparts can also contribute to semantic meta-data enhancement. On the other hand, security issues arise regarding systems, content rights and users' privacy, creating further difficulties and computational overheads, but also requiring the standardization of applicable mechanisms (Dimoulas & Symeonidis, 2015; Dimoulas et al., 2014a; 2014b; 2015; Kotsakis, Mislow, Kalliris & Matsiola, 2015; Matsiola et al., 2015; Sidiropoulos et al., 2015; Tsipas et al., 2015; Vrysis et al., 2015). The answer to these questions will determine future research directions and progress of hypermedia.

FUTURE RESEARCH DIRECTIONS

Considering the processes that SeMMSS authoring and sharing involve, there are many functional issues and technical challenges associated with the increased processing and bandwidth demands, and the applicable interaction modes that need to be properly addressed. It is quite difficult to encounter examples of multimedia services that incorporate all the above augmented interaction concepts. Neither it is easy, as in the case of text and images, to post, process online and author AV material in multimedia storytelling. One of the reasons can be found on the very interdisciplinary nature of web-docs and hyper-videos. Multimedia web authoring technologies cannot be easily adopted and utilized by the creators, since they usually don't possess software programming skills and knowhow. Thus, a plethora of user friendly web-doc authoring and hypermedia platforms were developed during the last years. Many of them appear in the cyberspace with many ambitious plans but soon disappear. This fact implies that although a high demand exists due to audience

Figure 2. An integrated semantically enhanced media authoring model (SeMAM)



attraction for such kind of expression media, a viable and sustainable business model has not yet been found for these platforms. Furthermore, large-scale interoperability between different media formats, terminals and platforms is difficult to achieve, while security issues remain crucial, especially for augmented hypermedia services (Dimoulas et al., 2014a; 2014b; 2015; Veglis

et al., 2016). Similarly, censorship actions are required and are expected to be more demanding for AV content (Chen & Wang, 2010; Koumartzis & Veglis; 2011).

Another important issue is related to the fact that multimodal intelligent content processing and semantic interaction techniques have many prerequisites and they have not yet reached a

maturity for large-scale real world applications. Nevertheless, research is ongoing, investigating new descriptive AV-features (Cordara et al., 2013; Dimoulas et al., 2011; Kalliris et al., 2014; Kotsakis et al., 2012; 2014; Vegiris et al., 2009), flexible ground truth acquisition and linking mechanisms (Monaghan et al., 2011; Tsikrika et al., 2011; Tsirelis & Delopoulos, 2013), hybrid smart systems and multimodal decision-making approaches (Dimoulas et al., 2008; 2011; Diou et al., 2010). In this context, emotional control, augmented interaction, content filtering, semantic annotation and retrieval are further promoted (Chen & Wang, 2010; Dimoulas & Symeonidis, 2015; Kalliris et al., 2014; Kotsakis et al., 2012; 2014; Matsiola et al., 2015). Perhaps, this is the real opportunity for the proposed SeMAM model to simultaneously serve, integrate and enhance all the augmented interaction and adaptation features of the future semantic-web services.

CONCLUSION

The current chapter focuses on the technologies of semantically-enhanced authoring of shared media. Current trends and future perspectives are analyzed. A semantically enhanced media authoring model (SeMAM) is proposed, integrating most of the expected progress in augmented user interaction and the upcoming Web 2.0/3.0 services. The targeted Semantically-enhanced Multimedia Storytelling Services (SeMMSS) aim at engaging audience members individually, validating their involvement and positively reinforcing personal participation in the narration. The expected result is intense loyalty, long-term engagement and a desire to share the experience. Although there are many issues that still require careful attention, current state of the art and research progress reveal that semantically-enhanced (collaborative) authoring of shared media will dominate. Hence, new perspectives are offered for the preservation of AV documents from the past, the exploitation

of present and upcoming multimedia technologies and their dissemination with added value storytelling services in the future.

REFERENCES

- Atzori, L., Delgado, J., & Giusto, D. D. (2012). Special issue on pervasive mobile multimedia. *Signal Processing Image Communication*, 27(8), 785–934. doi:10.1016/j.image.2012.05.002
- Chen, T. M., & Wang, V. (2010). Web filtering and censoring. *Computer*, 43(3), 94–97. doi:10.1109/MC.2010.84
- Cordara, G., Bober, M., & Reznik, Y. (2013). Special issue on visual search and augmented reality. *Signal Processing Image Communication*, 28(4), 309–402. doi:10.1016/j.image.2012.11.001
- De Bra, P., Freyne, J., & Berkovsky, S. (2013). Special issue on adaptive hypermedia. *New Review of Hypermedia and Multimedia*, 19(2), 81–212. doi:10.1080/13614568.2013.821320
- Dimoulas, C., Avdelidis, A., Kalliris, G., & Papanikolaou, G. (2008). Joint Wavelet video denoising and motion activity detection in multi-modal human activity analysis: Application to video-assisted bioacoustic/psycho-physiological monitoring. *EURASIP Journal on Advances in Signal Processing*, 2008(1), 792028. doi:10.1155/2008/792028
- Dimoulas, C., Kalliris, G., Chatzara, E., Tsipas, N., & Papanikolaou, G. (2014a). Audiovisual production, restoration-archiving and content management methods to preserve local tradition and folkloric heritage. *Journal of Cultural Heritage*, 15(3), 234–241. doi:10.1016/j.culher.2013.05.003
- Dimoulas, C., Papanikolaou, G., Petridis, V. (2011). Pattern classification and audiovisual content management techniques using hybrid expert systems: a video-assisted bioacoustics application in abdominal sounds pattern analysis. *Expert Systems with Application*, 38(10), 13082–13093.

- Dimoulas, C., Veglis, A., & Kalliris, G. (2014b). Application of Mobile Cloud-Based Technologies in News Reporting: Current Trends and Future Perspectives. In J. Rodrigues, K. Lin, & J. Lloret (Eds.), *Mobile Networks and Cloud Computing Convergence for Progressive Services and Applications* (pp. 320–343). Hershey, PA: Information Science Reference; doi:10.4018/978-1-4666-4781-7.ch017
- Dimoulas, C., Veglis, A., & Kalliris, G. (2015). Audiovisual Hypermedia in the Semantic Web. In M. Khosrow-Pour (Ed.), *Encyclopedia of Information Science and Technology* (3rd ed.; pp. 7594–7604). Hershey, PA: Information Science Reference; doi:10.4018/978-1-4666-5888-2.ch748
- Dimoulas, C. A., & Symeonidis, A. L. (2015). Syncing Shared Multimedia through Audiovisual Bimodal Segmentation. *IEEE MultiMedia*, 22(3), 26–42. doi:10.1109/MMUL.2015.33
- Diou, C., Stephanopoulos, G., Panagiotopoulos, P., Papachristou, C., Dimitriou, N., & Delopoulos, A. (2010). Large-scale concept detection in multimedia data using small training sets and cross-domain concept fusion. *IEEE Transactions on Circuits and Systems for Video Technology*, 20(12), 1808–1821. doi:10.1109/TCSVT.2010.2087814
- Dos Santos, J. A. F., & Muchaluat-Saade, D. C. (2012). XTemplate 3.0: Spatio-temporal semantics and structure reuse for hypermedia compositions. *Multimedia Tools and Applications*, 61(3), 645–673. doi:10.1007/s11042-011-0732-2
- Kalliris, G., Matsiola, M., Dimoulas, C., & Veglis, A. (2014). Emotional Aspects and Quality of Experience for Multifactor Evaluation of Audiovisual Content. *International Journal of Monitoring and Surveillance Technologies Research*, 2(4), 40–61. doi:10.4018/IJMSTR.2014100103
- Kotsakis, R., Dimoulas, C., Kalliris, G., & Veglis, A. (2014). Emotional Prediction and Content Profile Estimation in Evaluating Audiovisual Mediated Communication. *International Journal of Monitoring and Surveillance Technologies Research*, 2(4), 62–80. doi:10.4018/IJMSTR.2014100104
- Kotsakis, R., Kalliris, G., & Dimoulas, C. (2012). Investigation of broadcast-audio semantic analysis scenarios employing radio-programme-adaptive pattern classification. *Speech Communication*, 54(6), 743–762. doi:10.1016/j.specom.2012.01.004
- Kotsakis, R., Mislow, A., Kalliris, G., & Matsiola, M. (2015). Feature-Based Language Discrimination in Radio Productions via Artificial Neural Training. In *Proceedings of the Audio Mostly 2015 on Interaction With Sound (AM '15)*. ACM. doi:10.1145/2814895.2814928
- Koumartzis, N., & Veglis, A. (2011). Internet regulation: The need for more transparent Internet filtering systems and improved measurement of public opinion on Internet filtering. *First Monday*, 16(10). doi:10.5210/fm.v16i10.3266
- Ma, C.-X., Liu, Y.-J., Wang, H.-A., Teng, D.-X., & Dai, G.-Z. (2012). Sketch-based annotation and visualization in video authoring. *IEEE Transactions on Multimedia*, 14(4), 1153–1165. doi:10.1109/TMM.2012.2190389
- Matsiola, M., Dimoulas, C., Kalliris, G., & Veglis, A. (2015). Augmenting user interaction experience through embedded multimodal media agents in social networks. In *Social Media and the Transformation of Interaction in Society*. IGI Global. doi:10.4018/978-1-4666-8556-7.ch010
- Monaghan, F., Handschuh, S., & OSullivan, D. (2011). ACRONYM: Context metrics for linking people to user-generated media content. *International Journal on Semantic Web and Information Systems*, 7(4), 1–35. doi:10.4018/jswis.2011100101

Sidiropoulos, E. A., Konstantinidis, E. I., Kotsakis, R. G., & Veglis, A. A. (2015). RecApp: A mobile application for ubiquitous and collaborative audio processing. In *Proceedings of the Audio Mostly 2015 on Interaction With Sound (AM '15)*. ACM. doi:10.1145/2814895.2814934

Spyridou, L. P., Matsiola, M., Veglis, A., Kalliris, G., & Dimoulas, C. (2013). Journalism in a state of flux: Journalists as agents of technology innovation and emerging news practices. *International Communication Gazette*, 75(1), 76–98. doi:10.1177/1748048512461763

Symeonidis, A. L., & Mitkas, P. A. (2006). *Agent intelligence through data mining* (Vol. 14). Springer.

Tsikrika, T., Diou, C., de Vries, A. P., & Delopoulos, A. (2011). Reliability and effectiveness of clickthrough data for automatic image annotation. *Multimedia Tools and Applications*, 55(1), 27–52. doi:10.1007/s11042-010-0584-1

Tsipas, N., Zapartas, P., Vrysis, L., & Dimoulas, C. (2015). Augmenting Social Multimedia Semantic Interaction through Audio-Enhanced Web-TV Services. In *Proceedings of the Audio Mostly 2015 on Interaction With Sound (AM '15)*. ACM. doi:10.1145/2814895.2814907

Tsirelis, T., & Delopoulos, A. (2011). Automatic ground-truth image generation from user tags. *12th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS 2011)*.

Vegiris, C., Dimoulas, C., & Papanikolaou, G. (2009). Audio Content Annotation, Description and Management Using Joint Audio Detection, Segmentation and Classification Techniques. *126th Audio Engineering Society Convention 2009*.

Veglis, A., Dimoulas, C., & Kalliris, G. (2016). Towards intelligent cross-media publishing: media practices and technology convergence perspectives. In A. Lugmayr & C. Dal Zotto (Eds.), *Media convergence Handbook* (Vol. 2). Springer-Verlag. doi:10.1007/978-3-642-54484-2_8

Vrysis, L., Tsipas, N., Dimoulas, C., & Papanikolaou, G. (2015). Mobile Audio Intelligence: From Real Time Segmentation to Crowd Sourced Semantics. In *Proceedings of the Audio Mostly 2015 on Interaction With Sound (AM '15)*. ACM. doi:10.1145/2814895.2814906

ADDITIONAL READING

Aghaei, S., Nematbakhsh, M. A., & Farsani, H. K. (2012). Evolution of the World Wide Web: From Web 1.0 to Web 4.0. *International Journal of Web & Semantic Technology*, 3(1), 1–10. doi:10.5121/ijwest.2012.3101

Bernal, P. A. (2010). Web 2.5: The symbiotic web. *International Review of Law Computers & Technology*, 24(1), 25–37. doi:10.1080/13600860903570145

Bornman, H., & Von Solms, S. H. (1993). Hypermedia, multimedia and hypertext: Definitions and overview. *The Electronic Library*, 11(4-5), 259–268. doi:10.1108/eb045243

Chapman, N., & Chapman, J. (2009). *Digital multimedia* (3rd ed.). Chichester: John Wiley & Sons, Ltd.

Dimoulas, C., & Kalliris, C. (2013). Investigation of wavelet approaches for joint temporal, spectral and cepstral features in audio semantics. In *134th Audio Engineering Society Convention 2013*, 509–518.

Dimoulas, C., Vegiris, C., Avdelidis, K., Kalliris, G., & Papanikolaou, G. (2007). Automated audio detection, segmentation, and indexing with application to postproduction editing. In *122nd Audio Engineering Society Convention 2013*(3), 1462–1471.

Kanellopoulos, D. (2012). Semantic annotation and retrieval of documentary media objects. *The Electronic Library*, 30(5), 721–747. doi:10.1108/02640471211275756

- Karampiperis, P., & Sampson, D. G. (2013). Performance evaluation of decision-based content selection approaches in adaptive educational hypermedia systems. In A. Peña-Ayala (Ed.), *Intelligent and Adaptive ELS, SIST 17* (pp. 161–182). Springer-Verlag Berlin Heidelberg. doi:10.1007/978-3-642-30171-1_7
- Kompatsiaris, Y., & Hobson, P. (2008). *Semantic Multimedia and Ontologies*. London, UK: Springer-Verlag Limited. doi:10.1007/978-1-84800-076-6
- Kompatsiaris, Y., Merialdo, B., & Lian, S. (Eds.). (2012). *TV Content Analysis: Techniques and Applications*. CRC Press.
- Kotsakis, R., Kalliris, G., & Dimoulas, C. (2012a). Investigation of salient audio-features for pattern-based semantic content analysis of radio productions. In *132nd Audio Engineering Society Convention 2012*, 513-520.
- Kotsakis, R. G., Dimoulas, C. A., & Kalliris, G. M. (2012b). Contribution of stereo information to feature-based pattern classification for audio semantic analysis. In *Seventh International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP 2012)*, 68-72. doi:10.1109/SMAP.2012.26
- Koumartzis, N., & Veglis, A. (2011). On the pursue for a fair internet regulation system: A blueprint for a content blocking system encouraging participation by the internet users. In *Proceedings - IEEE Symposium on Computers and Communications*, 789-791. doi:10.1109/ISCC.2011.5983937
- Liao, S., Chu, P., & Hsiao, P. (2012). Data mining techniques and applications - A decade review from 2000 to 2011. *Expert Systems with Applications*, 39(12), 11303–11311. doi:10.1016/j.eswa.2012.02.063
- Mahnane, L., Laskri, M. T., & Trigano, P. (2013). A model of adaptive e-learning hypermedia system based on thinking and learning styles. *International Journal of Multimedia and Ubiquitous Engineering*, 8(3), 339–350.
- Morelli, A. K., Chaves, G. C., & Belchior, T. M. (2009). Digital video and interactivity. *Lecture Notes in Computer Science*, 5915 LNCS, 320-323. doi:10.1007/978-3-642-10643-9_39
- Moustakas, N., Floros, A., & Grigoriou, N. (2011). Interactive audio realities: An augmented / mixed reality audio game prototype. In *130th Audio Engineering Society Convention 2011*, 427-434.
- Psarras, V., Floros, A., Drosos, K., & Strapatsakis, M. (2011). Emotional control and visual representation using advanced audiovisual interaction. *International Journal of Arts and Technology*, 4(4), 480–498. doi:10.1504/IJART.2011.043446
- Quénot, G., Benois-Pineau, J., & André-Obrecht, R. (Eds.). (2012). Content-based Multimedia Indexing, Special issue of MTAP following CBMI 2010. *Multimedia Tools and Applications*, 60(2), 257–482. doi:10.1007/s11042-010-0648-2
- Tsipas, N., Dimoulas, C., Kalliris, G., & Papanikolaou, G. (2013). Collaborative annotation platform for audio semantics. In *134th Audio Engineering Society Convention 2013*, 218-222.
- Veglis, A. (2012a). Journalism and cross media publishing: The case of greece. In A. Veglis, & E. Siaperas (Eds.), *The wiley-blackwell handbook of global online journalism*, 209-230, Blackwell Publishing.
- Veglis, A. (2012b). From Cross Media to Transmedia Reporting in Newspaper Articles. *Publishing Research Quarterly*, 28(4), 313–324. doi:10.1007/s12109-012-9294-z

Vrysis, L., Dimoulas, C., Kalliris, C., & Papanikolaou, G. (2013). Mobile audio measurements platform: bringing audio semantic intelligence into ubiquitous computing environments. In *134th Audio Engineering Society Convention 2013*, 183-189.

KEY TERMS AND DEFINITIONS

Hypermedia: Combination of the linking mechanisms of hypertext and the multimodal presentation of information that is inherent in multimedia.

Multimodal Media Agents (M3A): Media services able to sense its surrounding environments (at multiple levels, semiotic modes and/or communication channels) and autonomously adapt on their changes, responding on behalf of other (software) entities and human interactions.

Multimodal Media Assets (MMA): A hybrid model of mixed time-based and page-based media (TBM, PBM) that are combined in non-linear presentation forms.

Semantic Interaction: Multimodal feature-based HMI utilizing artificial intelligence for concept recognition, semantic conceptualization and communication.

Semantically-Enhanced Media Authoring Model (SeMAM): A multimedia storytelling and authoring model that combines semantically-enhanced multimedia assets (SeMMA) with additional meta-data related to authoring decisions, augmented interaction and semantic feedback.

Semantically-Enhanced Multimedia (SeMM): Multimedia resources enhanced in semantic level through meta-data provided from many different users and groups, during the processes of production, authoring, sharing and consumption.

Semantically-Enhanced Multimedia Assets (SeMMA): Composite assets containing multimodal media (MMA) with authoring and semantic meta-data (SeMM serving as SeMMSS building blocks).

Semantically-Enhanced Multimedia Storytelling Services (SeMMSS): Multimedia services that combine various semantically-enhanced multimedia assets along with semantic meta-data and associated storytelling properties.

Web-Documentary: Documentary delivered through web, usually exploiting non-linear narration with multiple media and enhanced interaction scenarios.

Transmedia and Transliteracy in Nemetical Analysis

Michael Josefowicz

Nemetics Institute Kolkata, USA

Ray Gallon

The Transformation Society, France

Maria Nieves Lorenzo Galés

The Transformation Society, Spain

INTRODUCTION

The increasing diversification of interconnected media platforms, which provide a complex discourse, demands an effective use of the space that is now called “transmedia.” This article provides terms and definitions for transmedia and for the new set of personal skills and abilities required to participate in it: “transliteracy.” It also presents the nemetic system, which facilitates analyzing, tracking, and visualizing communication interactions in virtual transmedia environments.

Since humans started to use both gestural and oral codes to communicate, messages have been elaborated and expressed differently when different communication channels were in use. In recent times, with the dawn of radio and television, that fragmentation of content has become of interest to researchers (Steinberg, 2012), and has been identified as a characteristic of mass media (McLuhan, 1994).

With social media, content is fragmented across multiple virtual and physical platforms, with varying degrees of interaction that add complexity to social communication. Interactivity among multiple authors and multiple audiences generates dynamic “cross-media” seriality, “transmedia narrative” that has been studied from educational, entertainment, and sociological points of view (Dena, 2009).

Learning to use these media requires skills beyond the traditional listening and reading, to be able to integrate multiple messages in multiple codes, as an essential skill both for personal and professional communication. This transliteracy is a complex ability of intertextual navigation, the strategy for coding and decoding the multidiscourse in the digital ecosystem.

These recursive communication experiences are the subject of recent research (Duarte, 2014) that explores cognitive patterns in narrative that can be represented through geometric models, consolidating the use of the term “fractal narrative” in the transmedia context. The aim of this multilevel analysis is to take into account individual discourse (micro level), collective interaction (meso level) and community knowledge building (macro level). Interested readers will find a practical example of this in the documentation of the co-creative process that led to Daniel Durrant’s representation of a NEME (Figure 2 of this article) (Nemetics Institute, 2015).

In December 2010, Mark Frazier had explored the fractal essence of digital discourse, and debated with Spiro Spiliadis, Daniel Durrant, and Michael Josefowicz the possibilities of expressing its complexity using a symbolic language (Frazier, 2010). After this early work with Ebdish (Emergent by Design’ish), the nemetic system has emerged as a more elaborated code to express and visual-

ize interactive communication processes in the transmedia ecosystem (De, 2014).

BACKGROUND: TERMS AND CONCEPTS

Transmedia

The term “transmedia” is attributed to Marsha Kinder, who in 1991 used it to refer to an emerging entertainment supersystem, involving intertextuality and multiple sources with different levels of interaction (Kinder, 1991). It applied to tools, processes, and concepts, and opened the door to media that had not been invented then, such as wearables, implants, or augmented reality devices.

In 2003 Henry Jenkins described a process of “transmedia storytelling” in which “each medium does what it does best, so that a story might be introduced in a film, expanded through television, novels, and comics, and its world might be explored and experienced through game play.” (Jenkins, 2003) Later, he defined transmedia storytelling as a process “where integral elements of a fiction get dispersed systematically across multiple delivery channels for the purpose of creating a unified and coordinated entertainment experience.” (Jenkins, 2007)

Probably the best-known example of transmedia storytelling is the Star Wars franchise. The fictional universe of Luke Skywalker, Yoda, Han Solo, Darth Vader, and Lord Sith is created through the synergy of films, books, role playing games, comics, video games, toys, and animated shows, to create a collective imaginary world. But transmedia storytelling is not limited to the entertainment world.

The notion of multiplatform narrative is expanding now to encompass every type of human communication, including marketing (Tenderich, 2014), political debates (Costanza-Chock, 2014), or personal learning networks (Richardson & Mancabelli, 2011). There is a need to identify the skills required to use transmedia, the processes

for teaching and learning transmedia skills, and the techniques to analyze transmedia production and its metadata.

Transliteracy

The evolution from media to transmedia requires individuals to pass from personal linear communication skills and abilities (reading, writing, speaking, etc.) to intertextual skills. They include not only analogy, correlation, mental association, context awareness, or synthesis, but also empathy, engagement, and other emotional appreciations, essential to integrate and combine fragments of meaning into the holistic comprehension of a story.

The combination of these complex abilities can be understood as a new competency called transliteracy. Transliteracy has been defined as “the ability to read, write and interact across a range of platforms, tools and media from signing and orality through handwriting, print, TV, radio and film, to digital social networks.” (Thomas, 2005)

In transliteracy, a coherent discourse is perceived through a series of transmedia fragments. It requires participants to move from discrete, perceptive skills to compound, intangible projective skills that can facilitate strategic thinking and collective problem solving. Interactions such as debating, negotiating, conciliating, or collaborating on social platforms are high value-added skills and become the energy for emerging collective creativity.

It would be impossible, for example, to understand the phenomenon of global warming and its varied impacts, without some comprehension of how the climate changes; the effect of human activity on the atmosphere; human migratory patterns; changes in food sources and food production due to changes in land, oceanic, and atmospheric conditions, and thus on refugee crises and terrorist activity, as well. No one medium or information source is capable of adequately presenting all these interrelationships, but if we are able, not only to “read” information from a variety of sources, platforms, and media, but to integrate it,

and understand it as a complex holistic tangle of phenomena, we are using transliteracy to gain a higher level of consciousness, or “mindfulness,” about the subject.

Both transmedia and transliteracy are going to change the way we perceive learning and professional training, in universities and professional environments. Language and communication skills, necessary to understand and produce useful knowledge, are more and more related to the abilities to create, join, and maintain communities of interest, and build personal digital networks (Rajagopal, 2012).

Nemetics

Where transmedia is the framework for communication, and transliteracy is the skill for interaction, Nemetics is the analytic tool. Nemetics functions as a fractal meta-language that facilitates communication among researchers in different disciplines to debate about complexity. The multilayer nemetics system provides a methodology for connectivist action-research and action-reflection in transmedia, including several meta-codes for visualizing procedures and results.

The essentials of Nemetics can be summarized in a simple mnemonic acrostic, which describes learning in any context at any level. At its most effective it is:

- Notice without preconceptions (N).
- Engage without judgment (E).
- Mull before communicating or acting (M).
- Exchange in the appropriate way and time (E)

This basic path, (Notice. Engage. Mull. Exchange,) recalling the traditional Bloom taxonomy (Anderson et al. 2001), retrieves four action levels that may or may not be performed during interactions (after each verb, add the option, “or not”).

The whole conversation is then conceptualized as a single identified process, a NEME that can be seen as a coherent unit, represented visually by the

interactions that took place during the debate. The analysis of these nemes shows patterns and waves of exchange that offer extremely rich information (big data) both about the media environment and the participants. The basic initial model is the communicative sociogramme [Figure 1].

At the individual level, a NEME translates as a meta-cognitive routine; at collective levels, the study of NEME patterns can be useful for showing mental models that can serve as universals in further communicative analysis, and in designing debate strategies.

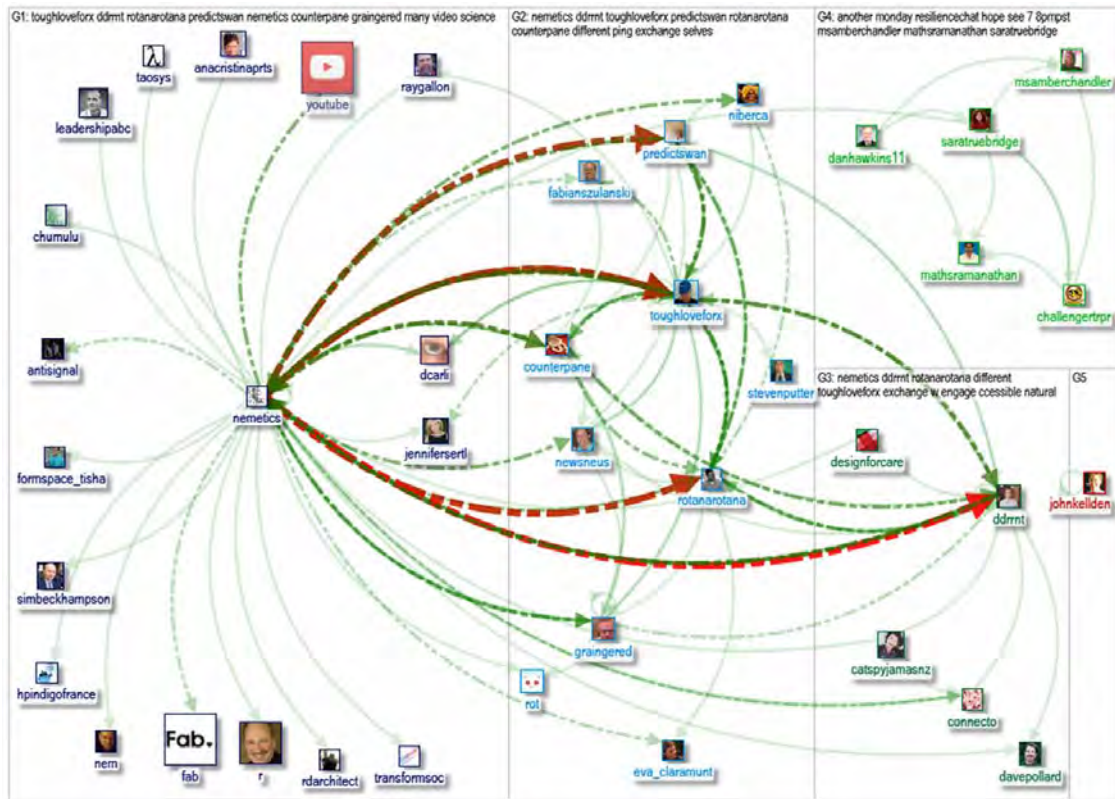
Taken as a unit, a NEME routine - that concludes with consensual knowledge - can be the starting point from which new debates can arise. This recursivity had already been studied in The Knowledge Forum, another international Computer Mediated Communication experience (Bereiter & Scardamalia, 2006), and has been identified as a path into high order thinking (HOT) processes (Rehage, 1994).

From routine building to big data analysis, nemetics includes the dimension of self-reflection, professional development, and organizational transformation. The obtained global vision contributes to leadership and resource management, focused on integrated learning - self and collective - and helps expand consciousness and engagement towards complex problem solving.

Nemetics helps express a new way of thinking about human communication that involves co-creation in complex adaptive/creative environments, and is being developed by the International Nemetics Institute: Care is taken not to define the discipline or its boundaries very rigidly. This is because once it is rigidly defined it also rigidly limits the development of the subject, its evolution and expanse it might cover. In short, any rigid definition would limit the discipline's adaptability to future changes in human conceptual understanding and knowledge or application in a different domain than what is envisaged right now. (De, 2012)

Figure 1. Communicative sociogramme. The graph represents a network of 40 Twitter users whose recent tweets contained “nemetics,” or who were replied to or mentioned in those tweets, taken from a data set limited to a maximum of 18,000 tweets. The network was obtained from Twitter on Monday, 06 July 2015 at 22:51 UTC.

© 2015 Marc Smith, used with permission.



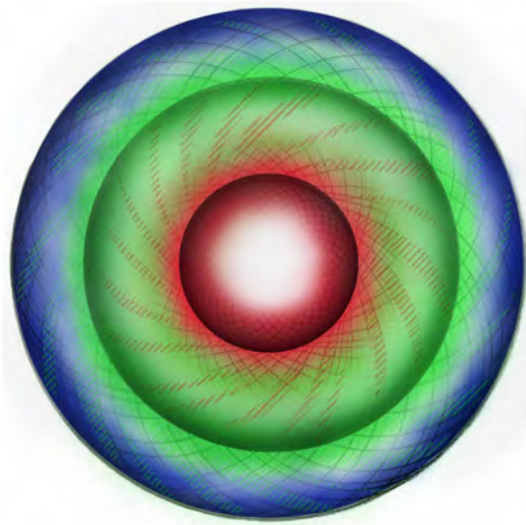
Used as the meta-language for recursive analysis, nemetics shows interactive thinking processes, collaborative debates and a collective cognitive resonance, through debate, discussion, negotiation, and compromise. This nemetical world is hierarchical, in the sense commonly used in complexity discourse, and it is composed of several nested spheres:

- NEMES,
- nemiString,
- nemiTube,
- nemiPlex,
- nemiCell,
- nemiShape.

Avoiding bottom up or top down hierarchies, the nemetic language works as in concentric spheres, from inside to outside, and from outside to inside. Any sphere can be seen as a NEME, where the proximate inner sphere is the source of energy. The proximate outer sphere is the source of constraints [Figure 2]. In addition, there are flows of energy within and between spheres.

This nemetic code can be used as a common language to shape complexity, and the fuzziness of the code is the source of its strength. It has allowed networks to research and exchange about innovation, education, history, political economy, design, or art across the fields of different disciplines, specialities and cultures. Nemetic research

Figure 2. Spheres of nested complexity
© 2015 Daniel Durrant, used with permission.



itself evolves by reproducing experiences of transliteracy in transmedia spaces.

NEMETICS RESEARCH ON TRANSMEDIA AND TRANSLITERACY

Context Analysis

In transmedia, the process is the product. Emerging experiences can be identified on the Internet by studying interaction sequences, exchange paths, and conversational structure, in augmented communication situations.

Nemetics should be understood as an ongoing process that is also a series of products. Context analysis becomes part of the production, and shows the internal process of collective debate and creation:

The power of transmedia is to make collaboration on even serious academic documents simple with almost no strain. No meetings. No phone calls. No schedules. Everyone works in their personal time. Inserts into any number of places. Goes away and waits for a response. (Josefowicz, 2015)

The International Nemetics Institutes provides a collaborative space for debating and exploring complexity from different scientific disciplines, by developing training and learning scenarios for hands-on problem solving in business, education and management fields. The main aim of this initiative is both process and product oriented. The process, begun in 2010, is an emergent intentional learning community that publishes blogs and social media posts, and performs open debates on twitter and other synchronic channels, offering the publication of a print newspaper and presentations in face-to-face training courses. The products include the founding of The International Nemetics Institute (TINI), based in Kolkata, India, and its subsidiary, The Nematic Institute of Art and Science (NIASK).

Case Study

The creation of the International Nemetics Institute is itself the result of collaborative interaction in the transmedia sphere with proactive intentions. The main aim was to provide a professional environment for Nemetics training and Nemetics community building. With a wide presence on the web, the International Nemetics Institute emerged from debates on Twitter, reflections on Google Plus forums, exchange of visual ideas in PowerPoint presentations, and dynamic groups of interest on LinkedIn. The Institute building process, in itself, is an example of collective transmedia emergence.

Overview

The process that led to the emergence of the Institute is best described by the term *synchrodipty* (Robertson, 2011), used to express the combination of synchronicity and serendipity that describes the starting sequence.

Main steps and key players:

- It was generated on the *Emergent by Design Blog*.

- It evolved into the Ebdish wiki (Michael Josefowicz, Daniel Durrant & Mark Frazier).
- Planning conversations emerged on twitter, under the hashtag #nemetics, with references to texts, images, sound, and videos.
- The dynamic conversations attracted international experts in Red Deer, Alberta, Canada (Sean Grainger), and Kolkata, India (Dibyendu De), Barcelona, Catalonia (Ray Gallon, Neus Lorenzo), and in Paris, France (Rotana Ty).
- Conversations expanded to Google Plus forums and LinkedIn groups.
- Mashups and image galleries are complementing textual discourse.
- Face to face meetings have taken place in different countries.
- Simultaneous conversations are flowing through different platforms, expanded into mind mapping, wave tank experiments, and 3D modeling.

Transmedia Environments

Nemetics presence can be tracked across several platforms and co-creation spaces:

- **Blog:** <http://rgbwaves.com/category/nemetics/>
- **YouTube Channel (TINI. The International NemeticsInstitute):** <http://www.youtube.com/channel/UCAes2xe3XVYTiaOyCCkdoTA>
- **Google Plus (+RapidInnovationOrg):** <https://plus.google.com/+RapidinnovationOrg/about>, ([+google.com/+NemeticsinstituteGlobal](https://plus.google.com/+NemeticsinstituteGlobal))
- **Twitter:** @NEMETICS
- **LinkedIn:** <https://www.linkedin.com/pub/international-nemetics-institute/84/14/3b3>
- **Collaborative Organization:** The Transformation Society, <http://www.transformationssociety.net/> publishing on Slideshare (<http://www.slideshare.net/TransformationSociety/presentations>) and on Twitter as @TransformSoc.

Different literacies are being developed in each one of these platforms, and across all of them, resulting in a new collaborative transliteracy.

Results: Analyzing the Behavior of Complex Creative Systems

The process has led to several results:

- From the transmedia debate, a research community has emerged and aggregated.
- The discipline of nemetics has been built across a network of connections, shaping an analytical tool set to engage wicked problems, such as the creation of an International Institute.
- A glossary of nemetic terms has been mapped to the nervous system, using a biomimicry model of communication in the tradition of memetics (Dawkins, 1989). Where a meme is a replicator in the cultural and cognitive context, equivalent to the gene in biology, the NEME is a replicator in the context of Complex Creative System (CCS).
- The same tools used to analyze the complexity of biology, are also used to analyze any other organic complexity. In nemetics, the model is the nervous system. Under this paradigm, a set of fractal analogies can be derived for studying transmedia co-creation and collective resonance:
 - The NEME is like a neuron. (In Transmedia it can be a shared idea, a closed dialogue, a formulated hypothesis...).
 - Bundles of NEMEs, called nemiStrings, are neural networks. (It can be an open debate, teamwork, group action, shared discussion...)
 - Bundles of nemiStrings are nemiTubes, like the nerve bundles found in the spinal column. (Communication exchanged in simultaneous discussions, joined teams from different

- platforms, augmented information management on several channels...)
- A nemiShape is the well-bounded interaction of the elements above, like the brain functioning within the skull. (It could be the community engaged in the same disciplinary debate, a social group sharing an aim, a formulated paradigm wherein debates can evolve).
- This structured landscape of neurons generates a complex creative system that is organically evolving, transforming, and reshaping itself, reproducing the recursive evolution of conversations, debates, and concept creation.
- Further results include mathematical representations of this complexity, diversification of narrative, and integration of meaning: mind maps, 3D animations, and visualized fragmented dialogues.

Current Hypothesis

Nemetics analysis provides a common code to researchers in different fields.

The recursive, fractal nature of NEME means that any level of complexity can be categorized as a NEME to clarify analysis. In a different discipline, for example, history, a singular event is a NEME at the level of events. Zooming in, the event devolves to vibrant nemiTubes of a diversity of social consequences, economic resonances, and cultural impacts. Following the fractal principle of self-similarity, an economic consequence is also a NEME on the level of analysis of consequences, and zooming in still further, consequences would have nemiTubes of market fluctuations and policies. In Geography, the metaphor can include trophic layers, like soil, plants, or animals (Provenza et al, 2013).

The concept of “narrative fractal” is transposed, in nemetics, as a “nemiPlex.” This nemiPlex is a complex of connected NEMEs that work the

way an enzyme works in biology, as a catalyzer of processes. The existence of a nemiPlex in the system catalyzes the process of the emergence of nemiStrings and nemiTubes. Importantly, a nemiPlex is strongest when it results from a diverse collaborative effort.

This is similar to what Lakoff refers to as a “metaphor” that facilitates an actor in “creating meaning.” The metaphor is an at-hand “explanation” for a new event (Lakoff, 2003). In nemetics a metaphor is seen as a constellation of concepts that are triggered by the need to interact (Exchange). It points to the fact that knowledge flows in networks of people while information resides in static web resources. Good transliteracy practice in a well-developed nemiPlex can speed up transmedia integration, for example, in High Performance Teams or Intentional Communities.

FUTURE RESEARCH DIRECTIONS

Nemetic Analysis of transmedia spaces helps understand the emerging communication systems in complex environments, and to design formal and informal transliteracy learning. Several action-reflection itineraries can be developed, in formal and informal training, for improved efficiency in identifying leadership, team dynamics, and coordination applied to transmedia communication. The main objective is social action.

Future benefits will come from applying results to personal and professional fields that are now emerging in the hybrid physical-digital environment:

- Interacting with wicked problems (favoring useful and valuable emergences from inherently chaotic systems)
- Designing and planning organizational action (using big data to adapt social behavior)
- Integrating gamification into professional fields (creating new transmedia spaces)

- Aggregating expanded connections and Internet of Things (new kinds of transliteracies)
- Relating personally and professionally to wearables and bionic implants (facing ethical and moral issues in collective debates)
- Developing lifelong learning environments (dynamizing and moderating MOOCs, eLearning platforms and mLearning communities)

These lines of research tend naturally towards the creation of a transformation society, where collaboration and self-emerging initiatives generate and explore transmedia learning environments.

CONCLUSION

Examples such as the creation of The International Nemetics Institute are useful to tell the story of transliteracy emergence and transmedia evolution and develop an academic corpus for integrating both transmedia and transliteracy into the communicative disciplines.

The increasing proliferation of communications platforms and modalities resulting from emerging technologies means that transmedia and transliteracy are going to become ever more important. Nemetics provides a simple model for understanding and integrating both transmedia contexts and transliteracy skills, and for working in complex adaptive/creative systems.

REFERENCES

- Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., & Wittrock, M. C. et al. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A revision of Bloom's Taxonomy of Educational Objectives*. New York: Pearson, Allyn & Bacon.
- Bereiter, C., & Scardamalia, M. (2006). Knowledge building: Theory, pedagogy, and technology. In K. Sawyer (Ed.), *Cambridge Handbook of the Learning Sciences*. New York: Cambridge University Press.
- Clark, R., & Chopeta, L. (2004). *Graphics for Learning: Proven Guidelines for Planning, Designing, and Evaluating Visuals in Training Materials*. San Francisco: Jossey-Bass/Pfeiffer.
- Costanza-Chock, S. (2014), *Out of the Shadows, Into the Streets! Transmedia Organizing and the Immigrant Rights Movement*. Cambridge, MA: The MIT Press. Retrieved Sept 30, 20015 from: https://mitpress.mit.edu/sites/default/files/titles/free_download/9780262028202_Out_of_the_Shadows.pdf
- Dawkins, R. (1989). *The Selfish Gene* (2nd ed.). Oxford, UK: Oxford University Press.
- De, D. (2012). *Nemetics - Some questions and answers*. Retrieved Sept 30, 20015 from <http://dibyendu.tumblr.com/post/18006947480/nemetics-some-questions-and-answers>
- De, D. (2014). *Winning Anywhere - The Power of 'See': The N-E-M-E Way*. Partridge Publishing India.
- Dena, C. (2009). *Transmedia Practice: Theorising the Practice of Expressing a Fictional World across Distinct Media and Environments*. University of Sydney. Retrieved Sept 30, 20015 from: http://dl.dropbox.com/u/30158/DENA_TransmediaPractice.pdf
- Duarte, G. A. (2014). *Fractal Narrative - About the Relationship Between Geometries and Technology and Its Impact on Narrative Spaces*. Bielefeld: Transcript-Verlag, Media Studies.
- Frazier, M. (2010), The early roots of the Nemes of Ebdish. *Storify documents*. Retrieved Sept 30, 20015 from: <https://storify.com/ebdish/nemes-of-ebdish-p1>

- Jenkins, H. (2003). Transmedia Storytelling. *Technology Review*. Retrieved Sept 30, 2015 from: <http://www.technologyreview.com/news/401760/transmedia-storytelling/>
- Jenkins. (2007). Transmedia Storytelling 101. *Confessions of an Aca-Fan*. Retrieved Sept 30, 2015 from: http://henryjenkins.org/2007/03/transmedia_storytelling_101.html
- Josefowicz, M. (2015). Notes for the #nemetichapter. *Print in the Communication Ecology*. Retrieved Sept 30, 2015 from: <http://sellingprint.blogspot.fr/2015/09/notes-for-nemetichapter.html>
- Kinder, M. (1991). *Playing with Power in Movies, Television, and Video Games: From Muppet Babies to Teenage Mutant Ninja Turtles*. Berkeley, CA: University of California Press.
- Lakoff, G., & Johnson, M. (2003). *Metaphors We Live By*. Chicago: University of Chicago Press. doi:10.7208/chicago/9780226470993.001.0001
- McLuhan, M. (1964). *Understanding Media: The Extensions of Man*. New York: McGraw Hill.
- Nemetics Institute. (2015). Co-creating a new view of the NEME. *Nemetics Is... Learning Ourselves & Our World via Arts & Complexity Science*. Retrieved Jan 2, 2016 from <https://medium.com/nemetics-is/co-creating-a-new-view-of-the-neme-5db123c88abb-.4z6v150>.
- Provenza, F., Pringle, H., Revell, D., Bray, N., Hines, C., Teagure, R.,... Barnes, M. (2013). *Complex Creative Systems: Principles, processes, and practices of transformation*. Retrieved from: <http://whitegumwool.com.au/wp-content/uploads/2013/04/Complex-Creative-Systems.pdf>
- Rajagopal, K. (2012). Understanding personal learning networks: Their structure, content and the networking skills needed to optimally use them. *First Monday*, 17(1). Retrieved Sept 30, 2015 from: <http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/3559/3131>
- Rehage, K. J. (1994). *Bloom's taxonomy: A forty-year retrospective*. Chicago: National Society for the Study of Education.
- Richardson, W., & Mancabelli, R. (2011). *Personal learning networks: Using the power of connections to transform education*. Bloomington, IN: Solution Tree Press. Retrieved Sept 30, 2015 from: http://soltreemrls3.s3-website-us-west-2.amazonaws.com/solution-tree.com/media/pdf/study_guides/Personal_Learning_Networks.pdf
- Robertson, S. (2011). *Synchrodipity, What's the difference between synchronicity and serendipity?*. Retrieved Sept 30, 2015 from: <http://sladerobertson.com/manifesting/synchrodipity.html>
- Steinberg, M. (2012). *Anime's Media Mix: Franchising Toys and Characters in Japan*. Minneapolis, MN: University of Minnesota Press. doi:10.5749/minnesota/9780816675494.001.0001
- Tenderich, B. (2014). *Transmedia Branding*. Retrieved Sept 30, 2015 from: <http://www.eimo.org/bilder/transmedia.pdf>
- Thomas, S. (2005). Transliteracy – Reading in the Digital Age. *Higher Education Academy English Subject Centre Newsletter*. Retrieved Sept 30, 2015 from: <http://www.english.heacademy.ac.uk/explore/publications/newsletters/newsissue9/thomas.htm>

ADDITIONAL READING

- Gallon, R. (2015). Technical communication: a wicked problem. In: *TC World, The Magazine for International Information Management*, December, 2015. Retrieved Jan 2, 2016 from <http://www.tcworld.info/e-magazine/technical-communication/article/technical-communication-a-wicked-problem>
- Kinder, M., & McPherson, T. (Eds.). (2014). *Transmedia Frictions: The Digital, the Arts, and the Humanities*. Oakland: University of California Press.

Lorenzo, N. (2015), *Nemetics visualization: The Nemetics system in conversational environments*. Presentation. Retrieved Jan 2, 2016 from http://www.slideshare.net/nlorenzo/nemetics-visualization-data-and-big-data-analysis?qid=50b5f014-0f9e-49ea-97e2-19b2d4bb89a6&v=default&b=&from_search=7

KEY TERMS AND DEFINITIONS

Intertext: A coherent text that shows a relationship to one or more other texts, where “text” is understood to mean any type of communicative content, typically forming a connected piece of work (includes images, sounds, video, etc.).

NEME: Mnemonic acronym for the fractal learning process of complex creative systems: Notice, Engage, Mull, Exchange.

Nemetics: A fractal code, evolved into a meta-language to facilitate communication among researchers in different disciplines in order to debate about complexity.

Synchrodipity: The compound interaction of synchronicity and serendipity, to produce a sense of discovery, delight, or well-being, and a sense of connectedness between people, ideas and actions,

derived from the flow and the interconnectivity of all things.

Transliteracy: The ability to read, write, and interact across a variety of communication tools, media, and platforms, from text, orality, signing, or drawing, through handwriting, print, TV, radio, and films to electronic networks and social media on digital platforms. It is a necessary complex skill for receiving, interiorizing or producing Transmedia.

Transmedia: Complex communication interaction based on multimedia, multimodal, multiplatform, intertextual human communication, in which each medium or platform has a distinct role to play in communicating the complete content. This interaction acquires meaning with each participating element by rebuilding the fragmented discourse.

Wicked Problem: A problem that is difficult or impossible to solve, because of incomplete, contradictory, and changing requirements that are difficult to define, identify, or recognize. It often involves stakeholders who have radically different worldviews. In addition, complex interdependencies make it so that the effort to solve one aspect of a wicked problem may reveal or create new problems.

Category N

**Networking and
Telecommunications**

Autonomic Cooperative Communications

N**Michał Wódczak***Samsung Electronics, Poland*

INTRODUCTION

Since the number of devices interconnected world-wide is growing at so unprecedented a pace, the orchestration of durable and resilient operation of networked systems becomes critically substantial. Especially when analysed in the context of network resiliency, such durability translates into the provision of key features of reliability, availability, safety, confidentiality, integrity, and maintainability. For this reason, it became highly valid to advocate for the integration of networking with the rationale behind autonomic computing in terms of self-configuration, self-optimisation, self-healing, and self-protection. This way the concept of Autonomic Cooperative Networking was proposed not only to integrate and capitalise on the above-mentioned building blocks, but, in particular, to incorporate the notion of Autonomic Cooperative Behaviour (Wódczak, 2014a). As such, stemming from and being triggered by the said Distributed Cooperative Relaying, the Autonomic Cooperative Behaviour intends to provide a mediation capability, not only aiming to facilitate cooperation among devices, but especially to integrate the relevant routines of the Optimised Link State Routing protocol. Consequently, the said devices are expected to share their computational capabilities and memory to perform joint data processing for the benefit of meeting the global performance indicators through increased resiliency.

As such, the paradigm of autonomic system design assumes that a networked system follow the operating principles of the Human Autonomic Nervous System and, thus, be able to self-manage without any external intervention. Yet, given the fact that autonomic designs are inherently char-

acterised by their own field of applicability, by no means should they be confused with the notion of autonomous or automated operation. In other words, while an autonomous development may, on the one hand, pertain to being stand-alone, and, on the other hand, to being cognitive, the concept of being automated is solely rooted in the operation of scripting. The difference in meaning does not exclude, however, certain dose of synergy so that autonomics could be supported through the inclusion of autonomous and automated routines. Following, to provide more details, the concept of Cooperative Relaying will be described on the basis of Virtual Antenna Arrays to pave the ground for the incorporation of Network Layer routines with special emphasis on the Optimised Link State Routing Protocol and its inherent Multi-Point Relay station selection heuristics. This way, such a routing enabled cooperation will be translated into Autonomic Cooperative Behaviour and integrated with the entities of the Generic Autonomic Network Architecture under the umbrella of Autonomic Cooperative Networked System design (Wódczak, 2014).

BACKGROUND

As the number of globally interconnected devices is becoming substantially large, the resulting networked systems are getting prone to configuration issues and resiliency becomes one of their key characteristics. Following the rationale behind self-management of autonomic computing, the main trend in networking nowadays is to put emphasis on the ability of a networked system to self-configure, self-optimize, self-heal, and self-protect without any explicit need for external

DOI: 10.4018/978-1-5225-2255-3.ch564

human intervention. This is crucial, for complexity reasons, as complete automation appears to be the only reasonable and justified way forward. In particular, devices may improve the related system robustness by sharing their computational resources through the application of cooperative schemes having been elevated to the level of Autonomic Cooperative Behaviour. For this reason the autonomic system design behind the Generic Autonomic Network Architecture was applied to synergise both the concept of Virtual Antenna Arrays and Multi-Point Relay station selection heuristics of the Optimised Link State Routing protocol, so that substantially large set-ups of devices could be considered to imitate the operation of Human Autonomic Nervous System. This was achieved with the aid of Autonomic Cooperative Networked System design allowing for the overall system to be controlled by Decision Making Entities within Autonomic Control Loops.

SYSTEM COMPONENTS

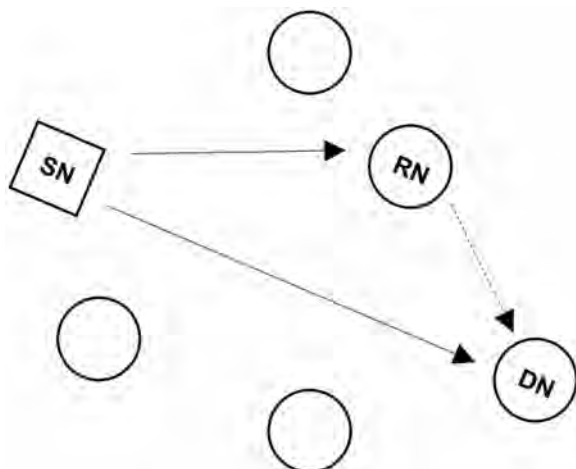
Cooperative Relaying

Cooperative relaying, also known as cooperative transmission, is undoubtedly one of the key advancements in the realm of mobile communications intended to facilitate the mitigation of the impairments induced and imposed by the characteristics of radio propagation (Pabst, Walke, Schultz, Herhold, Yanikomeroglu, Mukherjee, Viswanathan, Lott, Zirwas, Dohler, Aghvami, Falconer & Fettweis, 2004). Based on the assumption that Relay Nodes (RN) are to assist the process of transmission between the Source Node (SN) and the Destination Node (DN), this approach advocates that such a process be carried out in two phases. As outlined in Figure 1, during the first stage, both the DN and RN are assumed to receive the transmitted signal. Only after that may the RN additionally resend its copy towards the DN in order to, potentially, improve the transmission performance through the provision

of the so desirable diversity (Doppler, Redana, Wódczak, Rost & Wichman, 2009). In fact, there are a number of different variations of cooperative relaying known and this idea may also be referred to as cooperation diversity, cooperative diversity, or coded cooperation (Herhold, Zimmermann & Fettweis, 2004). In this work, however, the term of cooperative transmission is preferred as the most generic one (Wódczak, 2012).

Going further, based on the work described by (Laneman & Wornell, 2003) and according to the later classification given in (Herhold, Zimmermann & Fettweis, 2004), cooperative transmission schemes may be categorised either with regard to the forwarding strategy or protocol nature. Consequently, one may distinguish the amplify-and-forward, decode-and-forward, and decode-and-reencode categories. While the amplify-and-forward class is normally referred to as a non-regenerative approach, both its decode-and-forward and decode-and-reencode counterparts are typically named the regenerative ones (Herhold, Zimmermann & Fettweis, 2004). Additionally, the decode-and-reencode group may be subdivided even further to encompass fixed, adaptive, as well as feedback-based variations. One of the most comprehensive solutions of this type is known under the name of Virtual Antenna Arrays

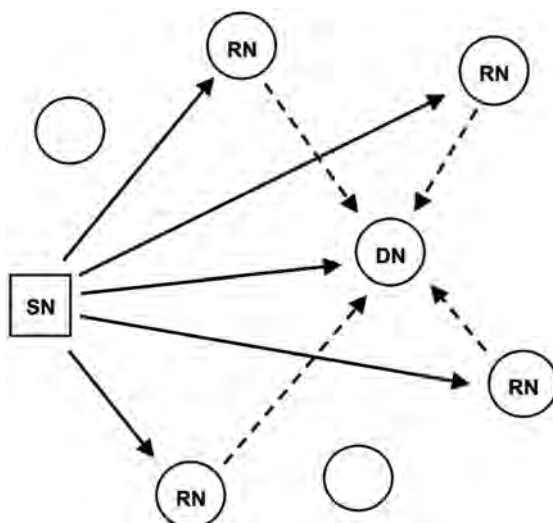
Figure 1. Cooperative relaying



(VAAs) (Dohler, Lefranc & Aghvami, 2002), as such, being, most interestingly, usually perceived as complementary to Distributed Space-Time Block Coding (DSTBC) (Laneman & Wornell, 2003), where both the repetition-based and space-time coded cooperative diversity techniques may be practically applied (Alamouti, 1998), (Tarokh, Jafarkhani & Calderbank, 1999).

In particular, such a DSTBC-enabled and VAA-based system model may be clearly depicted as in Figure 2, where the process of transmission between the SN and DN comprises, once again, two phases. What is more, also during the first phase, the SN broadcasts its signal so that, fairly similarly, it is delivered to the DN, as well as the potential RNs. Following, upon correct reception, all the intermediate RNs process the received data and, eventually, resend it towards the DN, during the second phase. Yet, as it was already signalled, such a retransmission may be performed either in the repetition-based or space-time coded manner, thereby creating a completely new, added value. Moreover, one should also take into consideration that even though both these approaches are capable of achieving full spatial diversity, the space-time coded approach outperforms the repetition-based one, as it can be used more effectively for higher spectral efficiencies (Laneman & Wornell, 2003).

Figure 2. Virtual antenna array



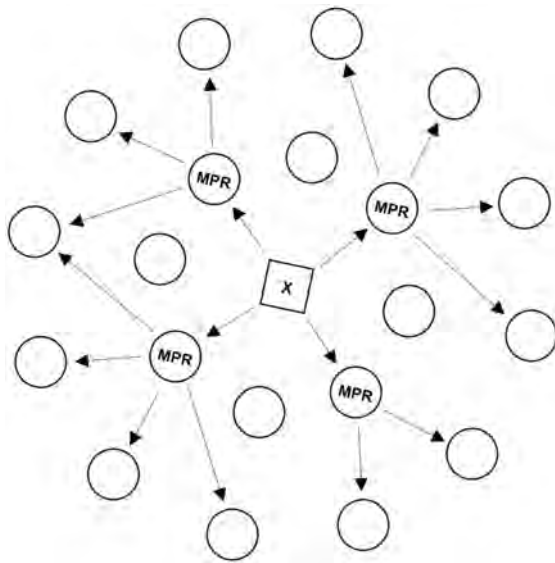
This is why, despite the fact that there are different retransmission schemes readily applicable, as far as the second phase is concerned the described concept mostly capitalises on the idea of VAAs, where, most interestingly, the direct radio link between the SN and DN may not even exist (Wódczak, 2012).

Routing Protocol

The presented concept of Autonomic Cooperative Communications puts special emphasis on the integration of the commonly recognised Optimized Link State Routing (OLSR) protocol, being classified as a proactive solution. As such, not only does the Optimized Link State Routing protocol belong to so desirable a class of routing solutions, but it also employs the said Multi-Point Relay (MPR) station selection heuristics intended to reduce the control overhead understood as the number of messages broadcast for the purposes of accomplishing the task of network topology information dissemination (Clausen & Jacquet, 2003). Most generally, the idea is to transmit the so-called Topology Control (TC) messages exclusively through a set of selected neighbour nodes, chosen in such a manner that they belong to the one-hop neighbourhood of a given SN performing the task of broadcasting and have been chosen with utmost care and due diligence to cover the entire strict two-hop neighbourhood of that node. Such one-hop neighbours are recognised by means of Hello messages, which are naturally received by each of them, but cannot be retransmitted any farther.

As illustrated in Figure 3, the Source Node includes in its MPR set all the symmetric one-hop neighbours being the only ones to provide reachability to a particular two-hop neighbour residing in the strict symmetric two-hop neighbourhood of SN, and, additionally, always denoted as willing to carry and forward traffic (Jacquet, Muhlethaler, Clausen, Laouiti, Qayyum & Viennot, 2001). Following, should there still exist any uncovered two-hop neighbours of SN, the heuristics keeps

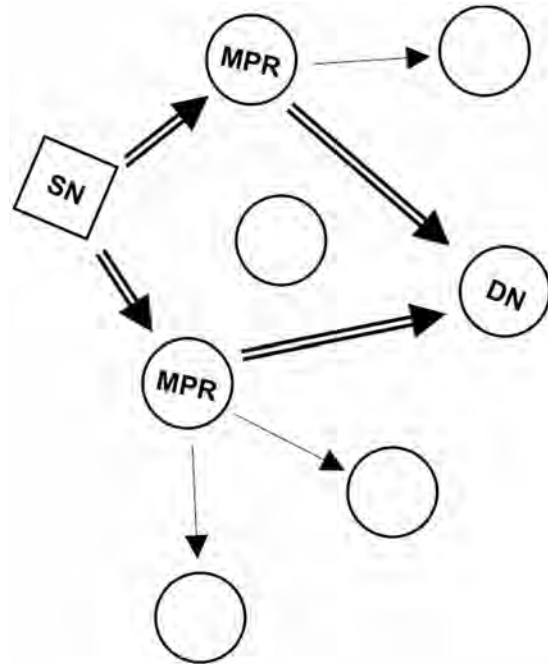
Figure 3. Multi-point relay station selection heuristics



being executed to select such a node in the one-hop neighbourhood which has not been inserted into the MPR set thus far and is characterised by the highest willingness to carry and forward traffic. In the event of the case that a multiple choice would hold, the one is selected which provides the highest reachability, i.e. through which the highest number of still uncovered two-hop neighbours may be reached. Otherwise, should it be impossible to select a single node, the node with the highest degree is chosen, where the degree of a one-hop neighbour denotes the number of its symmetric neighbours, excluding all the members of the one-hop neighbourhood and the SN performing the computation (Clausen & Jacquet, 2003).

In particular, as mentioned before, typical Relay Nodes may cooperate at Link Layer according to a more or less sophisticated scheme. However, the knowledge of such RNs about the network topology and the parameters of the separate radio links is solely limited to their closest one-hop neighbourhood. While it is still possible to imagine a more complex approach able to collect additional details at the said Link Layer, it seems a far better idea to utilise the information which is readily available at Network Layer instead. This

Figure 4. MPR-based VAA pre-selection



way the Routing information Enhanced Algorithm for Cooperative Transmission (REACT) was proposed which first removes each neighbour of the SN, characterised by a zero degree, from the one-hop neighbourhood, and executes the classic MPR heuristics iteratively over the set of one-hop neighbours until all the potential MPR nodes have been subdivided into the primary and additional MPR sets (Wódczak, 2012). In fact, each iteration results in a redundant MPR set, i.e. secondary, ternary and so on. Based on this, all the nodes are assigned to the most relevant VAAs being capable of providing cooperative connectivity between the SN and the DN, as depicted in Figure 4. The additional redundancy, in turn, may be utilised in the case of very sudden topology changes.

Autonomic Entities

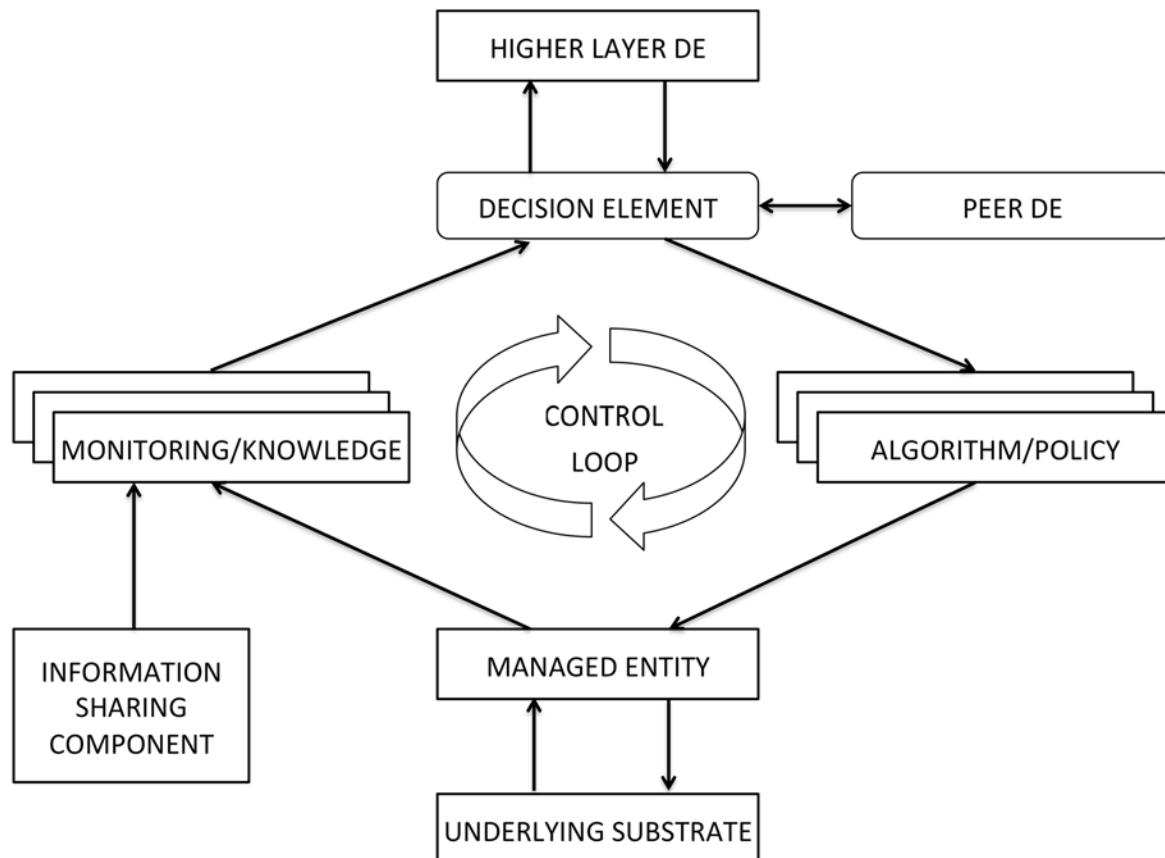
For the needs of further integration with the Generic Autonomic Network Architecture, the concept of REACT, being in fact the key enabler for the fusion between MPR and VAA, was proposed to be translated into a more generic notion

of Autonomic Cooperative Behaviour (ACB) (Wódczak, 2014). Conceptually, such ACB stems directly from the OLSR protocol and it capitalises on the VAA-aided cooperative transmission scheme. However, from the general perspective, the above assumptions might be relaxed so that ACB would be agnostic to both the protocol class and cooperation type. On the other hand, the key element of the GANA architecture is the assumption of the existence of Decision Elements (DEs) at the Network Level, Node Level, Function Level, and Protocol Level (Chaparadza, Papavassiliou, Kastrinogiannis, Vigoureux, Dotaro, Davy, Quinn, Wódczak & Toth, 2009). This way each DE has its exclusive responsibilities and, at the same time, it is able to interact with other DEs. It means that, based on the control and monitoring information available within Autonomic Control Loops (ACL),

as depicted in Figure 5, each DE should be able to take its autonomic decisions which might be to some extent affected by the data it exchanges with its higher or lower level, peering or sibling counterparts.

For this reason, in order to instantiate the concept of Autonomic Cooperative Communications, it was primarily assumed that the definition of the Autonomic Node would need to be enhanced with the aforementioned rationale behind Autonomic Cooperative Behaviour, as outlined in Figure 6. This resulted directly in the introduction of the highly relevant Autonomic Cooperative Node intended to act as an enabler for cooperation among the Managed Entities, as orchestrated by their corresponding Decision Elements. In particular, the Generic Autonomic Network Architecture defines the aforementioned four levels of abstraction at

Figure 5. Autonomic control loop



which DEs may operate (Wódczak, 2012). Starting from the Protocol Level, there resides the Cooperative Processing Decision Element (CP_DE) which is responsible for controlling the aspects of cooperative transmission protocol requiring the cooperative computation of code words (Wódczak, 2011). The operation of CP_DE has to be aligned with the already existing Routing Management Decision Element (RM_DE), located at the Function Level, which needs to interact with its sibling DEs, so that the computation of the routing tables maintained at the cooperating devices could be properly synchronised (Wódczak, 2007).

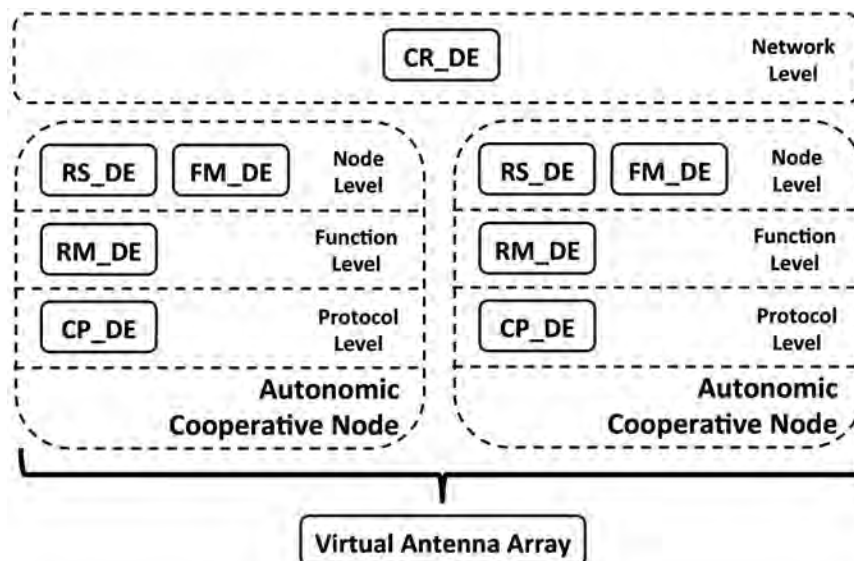
Following, the RM_DE also needs to act pursuant to the directions from the other existing DEs, such as the Resilience and Survivability Decision Element (RS_DE) and the Fault Management Decision Element (FM_DE), both located at the Node Level and responsible for the provision of increased dependability (Wódczak, 2010). More specifically, the RS_DE alone is typically assumed to cover aspects of service resilience and survivability. At the same time, it is supposed to interact with its FM_DE counterpart which, in turn, controls the symptoms suggesting that a failure, for example, in terms of service continuity, may be imminent. Finally, while all these

DEs are located within Autonomic Cooperative Nodes, it is still necessary to provide substantial coordination at the Network Level. This task is accomplished with the aid of the Cooperative Re-Routing Decision Element (CR_DE), which is responsible for overseeing the system situation from a higher-level perspective (Wódczak, 2012). This way a comprehensive approach may be formed encompassing all the aforementioned components so necessary for the instantiation of Autonomic Cooperative Communications.

FUTURE RESEARCH DIRECTIONS

The evolution of autonomic cooperative networking appears to more and more emphasize the importance of the realm of computing itself, thus creating a new notion of autonomic cooperative computing enabled networked systems. In this respect, certain dose of further work is necessary, especially from the architectural perspective so that the fields of autonomic networking and computing may be fused very closely together, especially taking into account the ongoing standardisation efforts.

Figure 6. Architectural perspective



CONCLUSION

In this work the rationale behind the concept of Autonomic Cooperative Communications was detailed. First of all, the workings of Distributed Cooperative Relaying were outlined through the introduction of the idea of Virtual Antenna Arrays as an enabler for further analysis. Following, the incorporation of the relevant routines of the Network Layer was described which put special emphasis on the Optimised Link State Routing Protocol and its Multi-Point Relay selection heuristics. Finally, the Routing information Enhanced Cooperative Transmission was detailed as enabler for Autonomic Cooperative Behaviour, being tightly integrated with the entities of the Generic Autonomic Network Architecture.

REFERENCES

- Alamouti, S. (1998). A Simple Transmit Diversity Technique for Wireless Communications. *IEEE Journal on Selected Areas in Communications*, 16(8), 1451–1458. doi:10.1109/49.730453
- Chaparadza, R., Papavassiliou, S., Kastrinogiannis, T., Vigoureux, M., Dotaro, E., Davy, K. A., & Toth, A. et al. (2009). Creating a viable Evolution Path towards Self-Managing Future Internet via a Standardizable Reference Model for Autonomic Network Engineering. In G. Tselentis, J. Domingue, A. Galis, A. Gavras, D. Hausheer, S. Krco, & T. Zahariadis et al. (Eds.), *Towards the Future Internet – A European Research Perspective*. IOS Press.
- Clausen, T., & Jacquet, P. (2003). Optimised Link State Routing Protocol (OLSR). *RFC 3626*.
- Dohler, M., Lefranc, E., & Aghvami, H. (2002). *Space-time block codes for virtual antenna arrays*. 13th IEEE Proceedings on Personal, Indoor and Mobile Radio Communications. doi:10.1109/PIMRC.2002.1046733
- Doppler, K., Redana, S., Wódczak, M., Rost, P., & Wichman, R. (2009). Dynamic resource assignment and cooperative relaying in cellular networks: Concept and performance assessment. *EURASIP Journal on Wireless Communications and Networking*, 2009(1), 475281. doi:10.1155/2009/475281
- Herhold, P., Zimmermann, E., & Fettweis, G. (2004, June). *Relaying and Cooperation - A System Perspective*. Paper presented at the 13th IST Mobile & Wireless Communications Summit, Lyon, France.
- Jacquet, P., Muhlethaler, P., Clausen, T., Laouiti, A., Qayyum, A., & Viennot, L. (2001). *Optimised link state routing protocol for ad hoc networks*. Paper presented at the IEEE International Multi Topic Conference.
- Laneman, J. N., & Wornell, G. W. (2003). Distributed space-time-coded protocols for exploiting cooperative diversity in wireless networks. *IEEE Transactions on Information Theory*, 49(10), 2415–2425. doi:10.1109/TIT.2003.817829
- Pabst, R., Walke, B., Schultz, D. C., Herhold, P., Yanikomeroglu, H., Mukherjee, S., & Fettweis, G. et al. (2004). Relay-Based Deployment Concepts for Wireless and Mobile Broadband Radio. *IEEE Communications Magazine*, 42(9), 80–89. doi:10.1109/MCOM.2004.1336724
- Tarokh, V., Jafarkhani, H., & Calderbank, A. R. (1999). Space-time block coding for wireless communications: Performance results. *IEEE Journal on Selected Areas in Communications*, 17(3), 451–460. doi:10.1109/49.753730
- Wódczak, M. (2007). *Extended REACT – Routing information Enhanced Algorithm for Cooperative Transmission*. Paper presented at the 16th IST Mobile & Wireless Communications Summit, Budapest, Hungary.
- Wódczak, M. (2010). *Future Autonomic Cooperative Networks*. Second International ICST Conference on Mobile Networks & Management 2010, Santander, Spain.

Wódczak, M. (2011). Autonomic cooperative networking for wireless green sensor systems. *International Journal of Sensor Networks*, 10(1/2), 83–93. doi:10.1504/IJSNET.2011.040906

Wódczak, M. (2012). *Autonomic Cooperative Networking*. New York: Springer-Verlag. doi:10.1007/978-1-4614-3100-8

Wódczak, M. (2014). *Autonomic Computing Enabled Cooperative Networked Design*. New York: Springer-Verlag. doi:10.1007/978-1-4939-0764-9

Wódczak, M. (2014a). *Autonomic Cooperative Networking*. *Encyclopedia of Information Science and Technology* (3rd ed.). Hershey, PA: IGI Global; doi:10.4018/978-1-4666-5888-2

Wódczak, M., Meriem, T. B., Chaparadza, R., Quinn, K., Lee, B., Ciavaglia, L., & Duault, M. et al. (2011). Standardising a Reference Model and Autonomic Network Architectures for the Self-managing Future Internet. *IEEE Network*, 25(6), 50–56. doi:10.1109/MNET.2011.6085642

KEY TERMS AND DEFINITIONS

Autonomic System: A concept advocating for a networked computer system to follow the principle behind the functioning of the Human Autonomic Nervous System, thus enabling it to self-manage without any need for external intervention for the majority of the time of its operation.

Distributed Cooperative Relaying: The mapping of the rationale behind spatio-temporal processing onto networked systems where designated Relay Nodes may form Virtual Antenna Arrays and, given proper synchronization, provide better transmission robustness or throughput.

Multi-Point Relays: A set of network nodes chosen with the aid of the MPR station selection heuristics of the Optimised Link State Routing protocol in such a way that the protocol control overhead becomes substantially reduced.

Spatio-Temporal Processing: A generic term for the space-time block and trellis coded concepts allowing for the exploitation of the radio channel orthogonality for the purposes of increasing either the system robustness or its throughput with the aid of the Multiple Input Multiple Output (MIMO) technology.

Clique Size and Centrality Metrics for Analysis of Real-World Network Graphs

N

Natarajan Meghanathan
Jackson State University, USA

INTRODUCTION

Network Science is a fast-growing discipline in academics and industry. It is the science of analyzing and visualizing complex real-world networks using graph theoretic principles. Several metrics are used to analyze the characteristics of the real-world network graphs; among them “centrality” is a commonly used metric. The centrality of a node is a measure of the topological importance of the node with respect to the other nodes in the network (Newman, 2010). It is purely a link-statistics based measure and not based on any offline information (such as reputation of the node, cost of the node, etc). The commonly used centrality metrics are degree centrality, eigenvector centrality, closeness centrality and betweenness centrality. Degree centrality (DegC) of a node is simply the number of immediate neighbors for the node in the network. The eigenvector centrality (EVC) of a node is a measure of the degree of the node as well as the degree of its neighbor nodes. We refer to DegC and EVC as degree-based centrality metrics. Closeness centrality (CIC) of a node is the inverse of the sum of the shortest path distances of the node to every other node in the network. Betweenness centrality (BWC) of a node is the ratio of the number of shortest paths the node is part of for any source-destination node pair in the network, summed over all possible source-destination pairs that do not involve the particular node. We refer to CIC and BWC as shortest path-based centrality metrics. Computationally efficient polynomial-time algorithms have been proposed in the literature (Brandes, 2001; Strang, 2005; Cormen et. al., 2009; New-

man, 2010) to determine exact values for each of the above centrality metrics; hence we categorize centrality as a computationally lightweight metric.

A “clique” is a complete sub graph of a graph (i.e., all the nodes that are part of the sub graph are directly connected to each other). Cliques are used as the basis to identify closely-knit communities in a network as part of studies on homophily and diffusion. Unfortunately, the problem of finding the maximum-sized clique in a graph is an NP-hard problem (Cormen et. al., 2009), prompting several exact algorithms and heuristics to be proposed in the literature (Pattabiraman et. al., 2013; Fortunato, 2010; Palla et. al., 2005; Sadi et. al., 2010; Tomita & Seki, 2003). In this chapter, we choose a recently proposed exact algorithm (Pattabiraman et. al., 2013) to determine the size of the maximum clique for large-scale complex network graphs and extend it to determine the size of the maximal clique that a particular node is part of. We define the maximal clique size for a node as the size of the largest clique (in terms of the number of constituent nodes) the node is part of. Note that the maximal clique for a node need not be the maximum clique for the entire network graph; but, the maximum clique for the entire graph could be the maximal clique for one or more nodes in the network.

Since the maximal clique size problem is a computationally hard problem and exact algorithms run significantly slower on large network graphs, our goal in this chapter is to explore whether the maximal clique size correlates well to one of the commonly studied computationally lightweight metrics, viz., centrality of the vertices, for complex real-world network graphs: if we observe a

DOI: 10.4018/978-1-5225-2255-3.ch565

high positive correlation between maximal clique size and one or more centrality metrics, we could then infer the ranking of the vertices based on the centrality values as the ranking of the vertices based on the maximal clique size in real-world network graphs. Ours will be the first chapter to conduct a correlation study between centrality and maximal clique size for real-world network graphs. To the best of our knowledge, we did not come across such a work that has done correlation study between these two metrics (and in general, a computationally hard metric vis-a-vis a computationally lightweight metric) for real-world network graphs. Throughout the chapter, we use the terms ‘node’ and ‘vertex’ as well as ‘link’ and ‘edge’ interchangeably. They mean the same.

Background

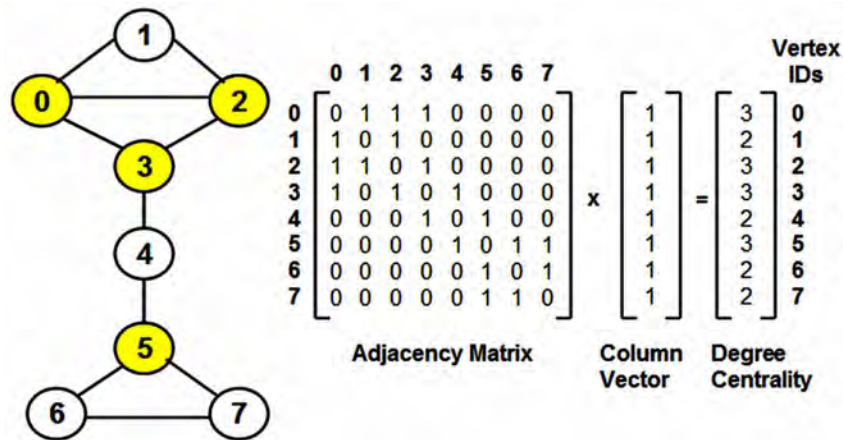
To the best of our knowledge, ours is the first work to focus on a correlation coefficient analysis between a computationally hard metric (maximal clique size for the individual vertices) with that of a computationally lightweight metric (centrality values of individual vertices) for complex real-world network graphs. The work available in the literature so far considers these two metrics separately. Recently, Li et al (2014) conducted a correlation coefficient analysis study among the centrality metrics for real-world network graphs. Centrality metrics have been widely studied for analysis and visualization of complex networks in several domains, ranging from biological networks to social networks (e.g., Koschutski & Schreiber, 2008; Opsahl et. al., 2010). The research focus with regards to cliques in the context of complex networks is to come up with efficient heuristics to reduce the run-time complexity in determining the maximum size clique for the entire network graph. Though branch-and-bound has been the common theme among these works, the variation is in the approach used to arrive at the bounds and enforce them in the search space. Strategies used for pruning the search space are typically based on node degree (e.g., Pattabiraman et. al.,

2013), vertex ordering (e.g., Carraghan & Pardalos, 1990) and vertex coloring (e.g., Ostergard, 2002). Recently, a parallelized approach (Rossi et. al., 2014) for branch and bound has also been proposed for determining cliques in real-world networks ranging from 1000 to 100 million nodes. Nevertheless, none of the research so far has focused on identifying correlation between the maximal clique size for an individual vertex (the size of the largest clique that a particular vertex is part of) with any of the commonly studied metrics (like centrality) for network analysis. Ours is the first step in this direction. With the problem of determining maximum size clique for the entire network graph and maximal size cliques for the individual vertices being NP-hard and computationally time-consuming for complex real-world networks of larger size, it becomes imperative to analyze the correlations of the maximal clique size values of the individual vertices to that of the network metrics that can be easily computed so that meaningful inferences about maximal clique size values can be made.

Centrality Metrics

This section discusses the four centrality metrics that are studied in this chapter. The highest ranked vertex or set of vertices with respect to each of the centrality metrics is shown shaded in the graphs of these figures. The degree centrality (DegC) of a vertex is the number of neighbors adjacent to it (example illustrated in Figure 1). Eigenvector centrality (EVC) of a vertex is a measure of the degree of the vertex as well as the degree of its neighbors. EVC of the vertices in a graph (example illustrated in Figure 2) are the entries in the principal eigenvector of the adjacency matrix for the graph. The larger the value of the entry for an vertex in the principal eigenvector, the higher is its ranking with respect to EVC. The principal eigenvector is determined by running the power-iteration algorithm (Strang, 2005) on the adjacency matrix of the network graph. The closeness centrality (CIC) of a vertex is the inverse

Figure 1. Example to illustrate the computation of degree centrality



of the sum of shortest path distances from the vertex to every other vertex (example illustrated in Figure 3). Betweenness centrality (BWC) is a measure of how significant a node is in facilitating communication between any two nodes in the network. The betweenness centrality for a node is the ratio of the number of shortest paths a node is part of for any source-destination node pair in the network, summed over all possible source-destination pairs that do not involve the particular node (example illustrated in Figure 4).

From Figure 1-4, we notice that the betweenness centrality-based ranking of the vertices is different from the degree centrality and eigenvector centrality-based ranking of the vertices. The degree and EVC-centralities take into consideration only the degree of the vertices and they are positively correlated. However, the BWC centrality takes into consideration the contribution of a vertex in facilitating communication between any two vertices in the network on the shortest path; such vertices are likely to be more central to the

Figure 2. Example to illustrate the calculation of eigenvector centrality using power iteration method

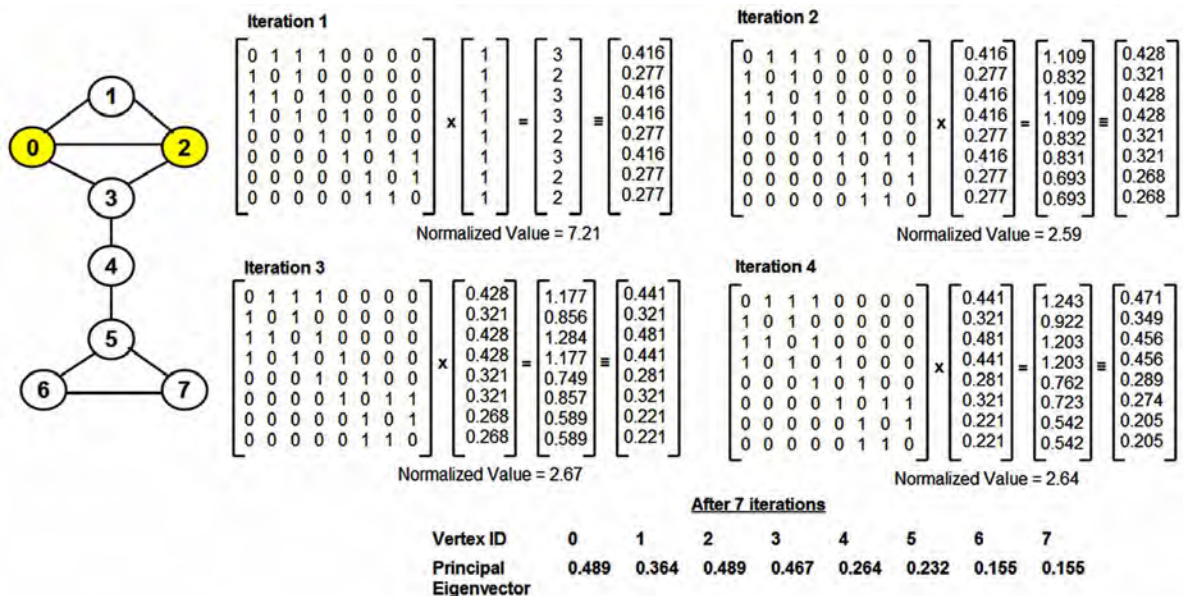
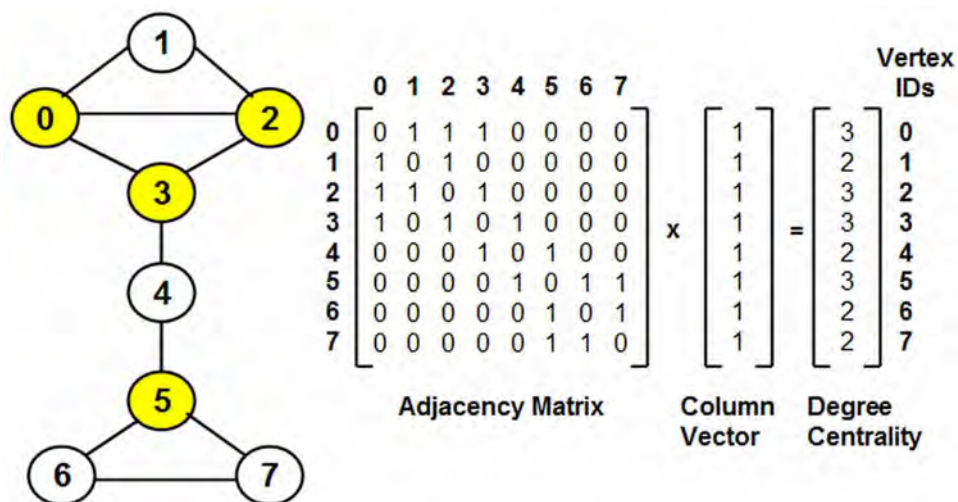


Figure 3. Example to illustrate the calculation of closeness centrality



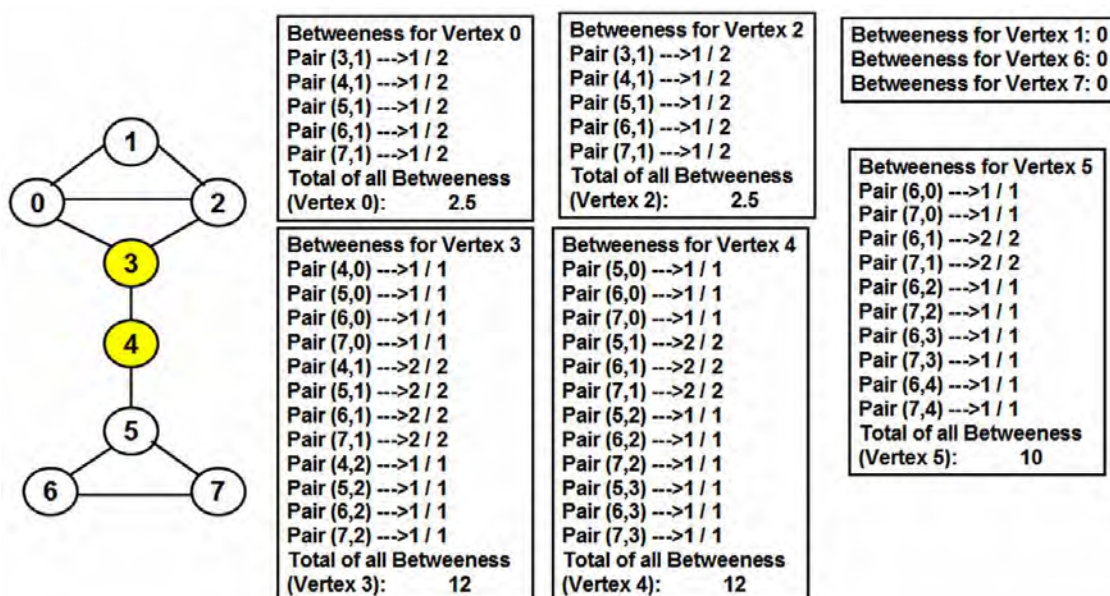
network and form the backbone or the core of the network. As can be seen in the example run of Figure 4, vertex 4 that was ranked lower with respect to degree and Eigenvector centrality is ranked the highest (along with vertex 3) with respect to Betweenness centrality. Also note that in this example graph, no source-destination pair need to go through vertices 1 or 6 or 7 on a short-

est path. As a result, the betweenness centrality for each of these three vertices (1, 6 and 7) is 0.

CLIQUE

A clique is a sub graph of a graph in which all the vertices are adjacent to each other. The problems

Figure 4. Example to illustrate the calculation of betweenness centrality



of finding maximum size clique for the entire graph as well as the maximal size cliques for the individual nodes are NP-hard problems (Cormen et. al., 2009). Several exact algorithms (that at the worst case incur exponential time for a NP-hard problem) have been proposed to determine maximum size cliques for sparse graphs. Recently, with the surge in interest to analyze large real-world networks from a graph theoretic point of view, researchers have proposed efficient exact algorithms (e.g., Tomita & Seki, 2003; Palla et. al., 2005; Fortunato, 2010; Sadi et. al., 2010; Pattabiraman et. al., 2013) to determine maximum size cliques for large/dense graphs. The common theme (Ostergard, 2002) behind these algorithms is a branch and bound approach of searching through all possible candidate cliques and limiting the search to only viable candidate sets of vertices whose agglomeration has scope of being a clique of size larger than the currently known clique found as part of the search; the variation among these exact algorithms is the pruning strategy (the approach taken to compute the bounds and use them) to limit the search. In this section, we will describe one such branch and bound-technique based exact algorithm that has been recently proposed in the literature (Pattabiraman et. al., 2013) to determine maximum size clique in large network graphs and explain our modification to the algorithm so that it can be used to determine the maximal cliques that each vertex in the graph

is part of; the largest among these cliques is the maximum size clique for the entire graph.

Figure 5 outlines the pseudo code of the algorithm, proposed originally by Pattabiraman et al (2013), to determine the maximum size clique for an entire graph. The procedure MAXCLIQUE proceeds in iterations, with each iteration designed to determine the maximum size clique for the entire graph that could also include vertex v_i (considered in the increasing order of the IDs). The sub routine CLIQUE called with vertex v_i as the first constituent vertex of the largest possible clique involving v_i , expands with one vertex at a time through a combination of iterations and recursions. During the sequence of returns from the recursive calls, it is possible that a new sequence of recursions and iterations is triggered due to the presence of a neighbor u of v_i that has scope for being in a clique (involving v_i) of size larger than the clique found so far for the entire graph. The algorithm explores all such possible cliques involving vertex v_i that have scope for exceeding the currently known maximum size clique for the entire graph.

At the end, the algorithm returns the maximum size clique for the entire graph that also happens to be the maximal size clique involving some vertex v_i such that there is no other vertex v_j ($i > j$) that is also part of the clique. Since the algorithm proceeds with vertices in the increasing order of their IDs, if the maximum size clique for the entire

Figure 5. Exact algorithm to determine the maximum size clique for an entire graph

```

Procedure MAXCLIQUE ( $G = (V, E)$ )
   $max \leftarrow 0$ 
  for  $i : 1$  to  $|V|$  do
    if  $\text{degree}(v_i) \geq max$  then
       $U \leftarrow \phi$ 
      for each  $v_j \in \text{Neighbor}(v_i)$  do
        if  $\text{degree}(v_j) \geq max$  then
           $U \leftarrow U \cup \{v_j\}$ 
      CLIQUE( $G, U, 1$ )

Subroutine CLIQUE( $G = (V, E), U, size$ )
  //  $size$  is the size of clique found so far
  if  $U = \phi$  then
    if  $size > max$  then
       $max \leftarrow size$ 
    return
  while  $|U| > 0$  do
    if  $size + |U| \leq max$  then
      return
    select any vertex  $u$  from  $U$ 
     $U \leftarrow U \setminus \{u\}$ 
     $N(u) := \{w \mid w \in \text{Neighbor}(u) \wedge \text{degree}(u) \geq max\}$ 
    Clique( $G, U \cup N(u), size + 1$ )

```

graph involves at least one vertex v_i with a smaller ID, the presence of the maximum size clique is detected much earlier and the subsequent iterations (with vertices whose IDs are greater than v_i , but could be part of only cliques of size smaller or equal to the maximum size clique of the entire graph involving v_i) are merely pruned, contributing to the time-efficiency of the algorithm. Hence, the labeling of the vertices with their IDs plays a significant role in the run-time complexity of the algorithm; the algorithm is capable of quickly determining the maximum size clique if the latter comprises of at least one vertex with a smaller ID.

Figure 6 illustrates our modifications (to determine the size of the maximal clique that each vertex is part of) to the pseudo code of the algorithm presented in Figure 5. The tradeoff is an increase in the run-time of the algorithm: we cannot just prune our search based on the vertex IDs; we have to explore the neighborhood of each of the vertices to determine the maximal size clique that each vertex is part of. Since to start with, the maximal size clique known for vertex v_i is 0, there is no need to filter the neighbors of v_i in procedure MAXIMALCLIQUE based on the degree of the neighbors; all neighbors of v_i are included in the set U and passed onto the

sub routine CLIQUE. However, we could retain all of the pruning steps in sub routine CLIQUE (called to find the maximal size clique for each of the vertices v_i) and recursive calls to the same: there is no need to explore the neighbors of vertex u whose degree is less than that of the currently known maximal clique size for vertex v_i .

REAL-WORLD NETWORK GRAPHS AND CORRELATION STUDIES

The network graphs analyzed are briefly described as follows: (i) *Zachary's Karate Club* (Zachary, 1977): Social network of friendships (78 edges) between 34 members of a karate club at a US university in the 1970s; (ii) *Dolphins' Social Network* (Lusseau et. al., 2003): An undirected social network of frequent associations (159 edges) between 62 dolphins in a community living off Doubtful Sound, New Zealand; (iii) *US Politics Books Network* (Krebs, 2008): Nodes represent a total of 105 books about US politics sold by the online bookseller Amazon.com. A total of 441 edges represent frequent co-purchasing of books by the same buyers, as indicated by the “customers who bought this book also bought these other

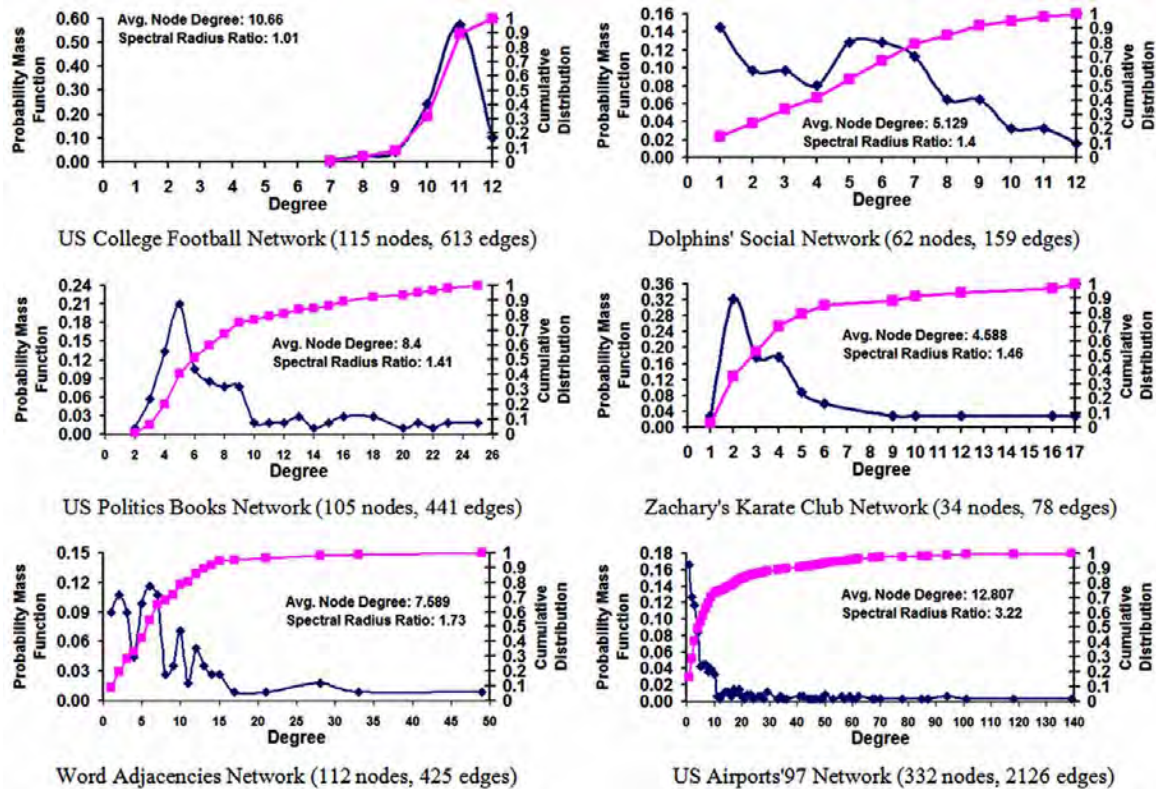
Figure 6. Exact algorithm to determine the maximal clique size for each vertex in a graph

```

Procedure MAXIMALCLIQUE ( $G = (V, E)$ )
  for  $i : 1$  to  $|V|$  do
     $\text{maximalCliqueSize}[v_i] \leftarrow 0$ 
     $U \leftarrow \phi$ 
    for each  $v_j \in \text{Neighbor}(v_i)$  do
       $U \leftarrow U \cup \{v_j\}$ 
    CLIQUE( $G, v_i, U, 1$ )

Subroutine CLIQUE( $G = (V, E), v_i, U, \text{size}$ )  // size is the size of clique found so far for vertex  $v_i$ 
  if  $U = \phi$  then
    if  $\text{size} > \text{maximalCliqueSize}[v_i]$  then
       $\text{maximalCliqueSize}[v_i] \leftarrow \text{size}$ 
    return
  while  $|U| > 0$  do
    if  $\text{size} + |U| \leq \text{maximalCliqueSize}[v_i]$  then
      return
    select any vertex  $u$  from  $U$ 
     $U \leftarrow U \setminus \{u\}$ 
     $N'(u) := \{w \mid w \in \text{Neighbor}(u) \wedge \text{degree}(u) \geq \text{maximalCliqueSize}[v_i]\}$ 
    CLIQUE( $G, v_i, U \cup N'(u), \text{size} + 1$ )
  
```

Figure 7. Distribution of node degrees (probability mass function and cumulative distribution)



books” feature on Amazon; (iv) *WordAdjacencies Network* (Newman, 2006): This is a word co-appearance network representing adjacencies of common adjective and noun in the novel “David Copperfield” by Charles Dickens. A total of 112 nodes represent the most commonly occurring adjectives and nouns in the book. A total of 425 edges connect any pair of words that occur in adjacent position in the text of the book; (v) *US College Football Network* (Girvan & Newman, 2002): Network represents the teams that played in the Fall 2000 season of the American Football games and their previous rivalry - nodes (115 nodes) are college teams and there is an edge (613 edges) between two nodes if and only if the corresponding teams have competed against each other earlier; (vi) *US Airports 1997 Network*: A network of 332 airports in the United States (as of year 1997) wherein the vertices are the airports and two airports are connected with an edge (a total of 2126 edges) if there is at least one direct

flight between them in both the directions. Data for networks (i) through (v) can be obtained from <http://www-personal.umich.edu/~mejn/netdata/>. Data for network (vi) can be obtained from: <http://vlado.fmf.uni-lj.si/pub/networks/pajek/data/gphs.htm>.

The spectral radius degree ratio is a measure of the variation in the node degree with respect to the average node degree; the closer the ratio is to 1, the smaller the variations in the node degree and the degrees of the vertices are closer to the average node degree (characteristic of random graph networks). The farther is the ratio from 1, the larger the variations in the node degree (characteristic of scale-free networks). Figure 7 presents the degree distribution of the network graphs in the increasing order of their spectral radius ratio for node degree (1.01 to 3.23).

Table 1 presents a correlation coefficient analysis of the four centrality metrics and the maximal clique size observed for the vertices in

Table 1. Correlation between centrality metrics and maximal clique size for the nodes

Network Index	Network Name (Increasing Order of Spectral Radius Ratio)	Degree vs. Clique	Eigenvector vs. Clique	Closeness vs. Clique	Betweenness vs. Clique
(v)	US College Football Network	0.315	0.348	-0.028	-0.168
(ii)	Dolphins' Social Network	0.776	0.563	0.418	0.277
(iii)	US Politics Books Network	0.701	0.747	0.321	0.367
(i)	Zachary's Karate Network	0.641	0.767	0.615	0.458
(iv)	Word Adjacencies Network	0.706	0.815	0.835	0.478
(vi)	US Airports 1997 Network	0.868	0.953	0.843	0.404

each of the six real-world network graphs studied in this chapter. Values of correlation coefficient greater than or equal to 0.8 (high correlation) have been highlighted in yellow; values below 0.5 (low correlation) are highlighted in light blue; and values between 0.5 and 0.8 (moderate correlation) are not highlighted in any color. If $\bar{X}$ and $\bar{Y}$ are the average values of the two metrics (say X and Y) observed for the vertices (IDs 1 to n , where n is the number of vertices) in the network, the formula used to compute the Correlation Coefficient between two metrics X and Y is as follows:

$$\begin{aligned}
 & \text{Corr Coeff}(X, Y) \\
 &= \frac{\sum_{ID=1}^n (X[ID] - \bar{X}) * (Y[ID] - \bar{Y})}{\sqrt{\sum_{ID=1}^n (X[ID] - \bar{X})^2} \sqrt{\sum_{ID=1}^n (Y[ID] - \bar{Y})^2}}
 \end{aligned}$$

As we can see in Table 1, in general, the correlation between the centrality metrics and the maximal clique size increases as the spectral radius ratio for node degree increases. This implies, the more scale-free a real-world network is, the higher the correlation between the centrality value and the maximal clique size observed for a node. With several of the real-world networks being mostly scale-free, we expect these networks to exhibit a similar correlation to that observed in this chapter. The visualization figures presented in the chapter were obtained by porting the network data as well as the centrality/maximal clique size results to Gephi (Cherven, 2013) and making appropriate changes to the settings in the latter. The layout algorithm chosen in Gephi for visualization is the Fruchterman Reingold algorithm (Fruchterman & Reingold, 1991) that presents the network in a circular format (like a globe).

Figure 8. US College Football Network: Correlation of maximal clique size with centrality

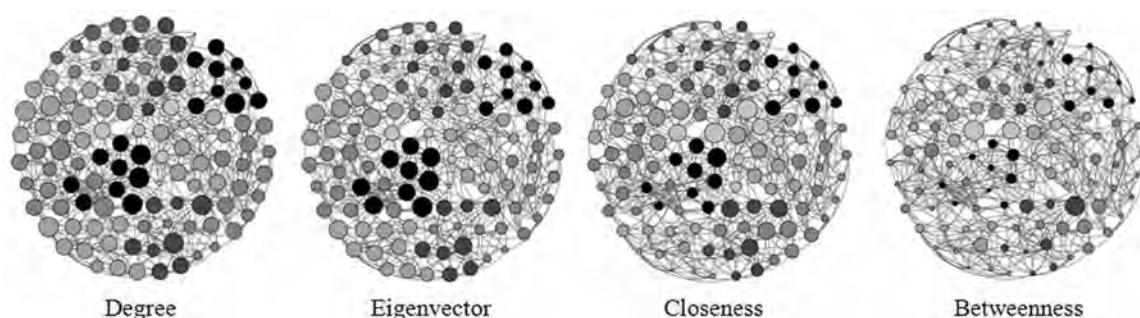


Figure 9. Dolphins' Social Network: Correlation of maximal clique size with centrality

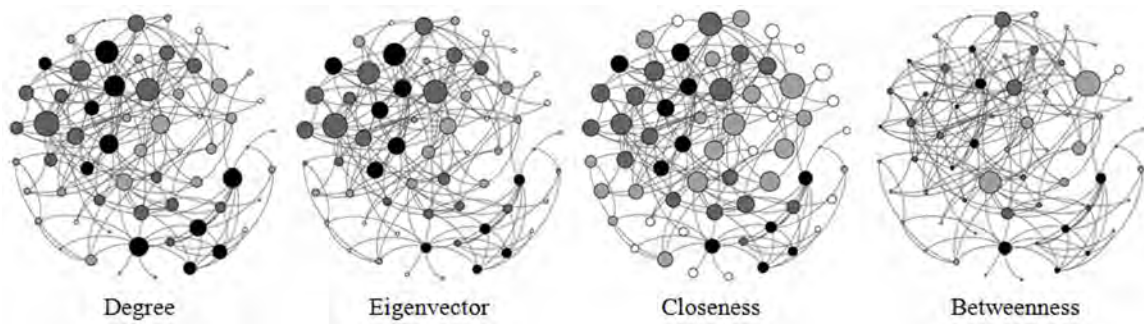


Figure 10. US Politics Books Network: Correlation of maximal clique size with centrality

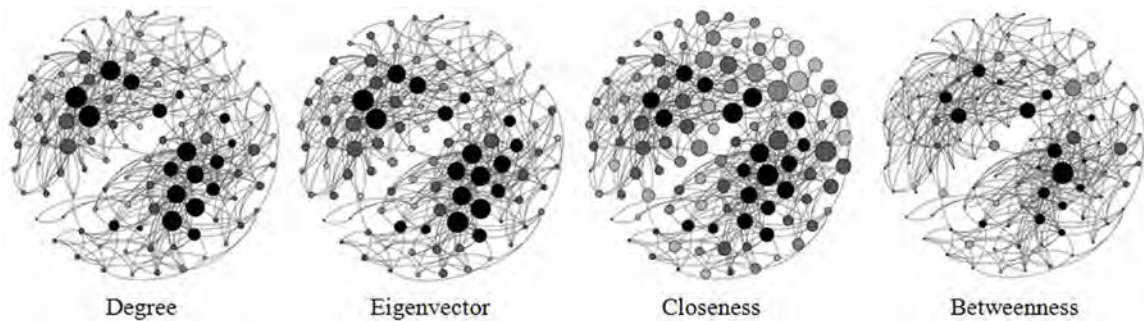


Figure 11. Zachary's Karate Club Network: Correlation of maximal clique size with centrality

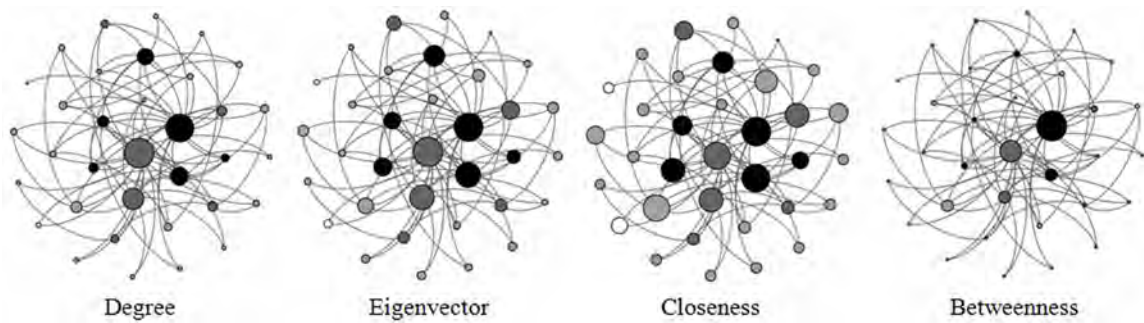


Figure 12. Word Adjacencies Network: Correlation of maximal clique size with centrality

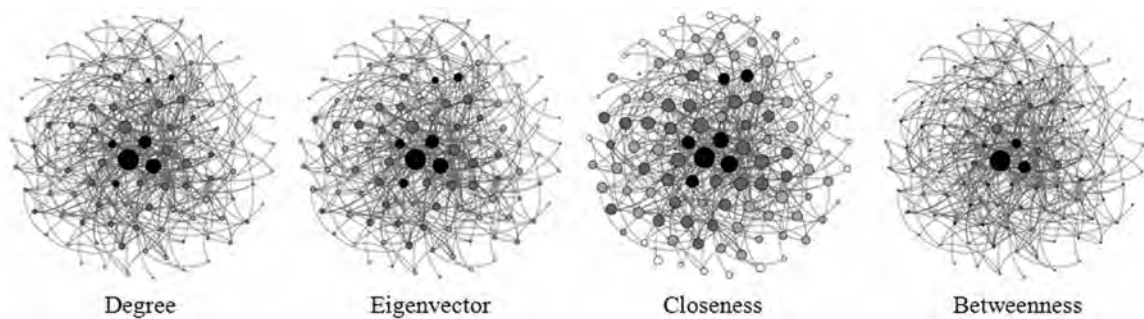
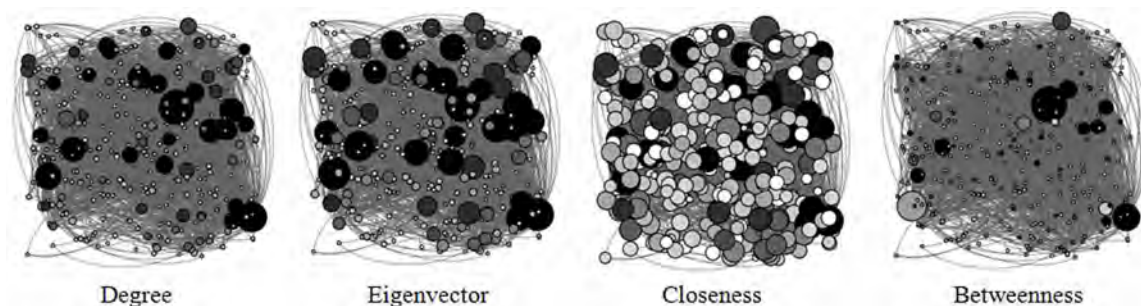


Figure 13. US Airports (1997) Network: Correlation of maximal clique size with centrality



Figures 8-13 depict the correlation observed for the four centrality metrics with that of the maximal clique size for the vertices in the real-world network graphs. In these figures, the node size is a measure of the node centrality (the larger a node is, the larger is its centrality value); the node color is a measure of the maximal size clique the vertex is part of (the darker a node is, the larger is the size of the maximal clique for the node). Among the two classes of centrality metrics, we observe the degree-based centrality metrics (degree centrality and eigenvector centrality) to be very positively and highly correlated with the maximal clique size observed for the nodes. Between the two degree-based centrality metrics, the eigenvector centrality metric shows higher positive correlations to the maximal clique size. This could be attributed to the eigenvector centrality of a node being a measure of both the degree of the node as well as the degrees of its neighbors. That is, a node with high degree as well as located in a neighborhood of high degree vertices is more likely to be part of a maximal clique of larger size. In addition, as the networks get increasingly scale-free, nodes with high degree are more likely connected to other similar nodes with high degree to facilitate an average path length that is almost independent of network size: characteristic of scale-free networks (Newman, 2010), contributing to a positive correlation between degree-based centrality metrics and maximal clique size.

With respect to the two shortest-path based centrality metrics, the betweenness centrality

metric is observed to exhibit a low correlation with maximal clique size for all the six real-world network graphs; the correlation coefficient increases as the network becomes increasingly scale-free. In networks with minimal variation in node degree (like the US College Football network that is more closer to a random network), nodes that facilitate shortest-path communication between several node pairs in the network are not part of a larger size clique; on the other hand, nodes that are part of larger size cliques in such random networks exhibit a relatively lower betweenness centrality. Since the degrees of the vertices in random networks are quite comparable to the average node degree, there is no clear ranking of the vertices based on the degree-based centrality metrics and maximal size cliques that they are part of. Also, if at all a vertex ends up being a larger sized clique in random network graphs, it does not facilitate shortest path communication between the majority of the vertices (contributing to a negative/zero correlation or at best a low correlation with betweenness centrality). As the network becomes increasingly scale-free, the hubs that facilitate shortest-path communication between any two nodes in the network exhibit higher betweenness and closeness centralities as well as form a clique with other high-degree hubs - exhibiting the ultra small-world property; the average path length is $\ln(\ln N)$, where N is the number of nodes in the network (Newman, 2010). The correlation of the closeness centrality values and the maximal clique size values observed for the vertices in

real-world network graphs is significantly higher (i.e., positive correlation) for networks that are increasingly scale-free.

Overall, the degree-based centrality metrics exhibit a relatively better correlation with the maximal clique size compared to that of the shortest-path based centrality metrics (especially in networks with low-moderate variation in node degree). For real-world networks that exhibit moderate-high variation in node degree, the shortest-path based centrality metrics (especially closeness centrality) fast catch up with that of the degree-based centrality metrics and exhibit higher levels of positive correlation with maximal clique size. We anticipate that as the networks become increasingly scale-free, the hubs (that facilitate shortest-path communication between any two nodes) are more likely to form the maximum clique for the entire network graph - contributing to higher levels of positive correlation between any of the four centrality metrics and maximal clique size.

FUTURE RESEARCH DIRECTIONS

With the problem of determining maximal clique sizes for individual vertices being computationally time consuming, our approach taken in this chapter to study the correlation between maximal clique sizes to centrality can be the first step in identifying positive correlation between cliques/ clique size in real-world network graphs to one or more network metrics (like centrality) that can be quickly determined and henceforth appropriate inferences can be made about the maximal size cliques of the individual vertices. We observe the degree-based centrality metrics (especially the eigenvector centrality) to show promising positive correlations to that of maximal clique sizes of the individual vertices, especially as the networks get increasingly scale-free; this observation could form the basis of future research for centrality-clique analysis for complex real-world networks.

CONCLUSION

The correlation coefficient analysis studies between the four centrality metrics (degree, eigenvector, betweenness and closeness centralities) and the maximal clique size for the vertices in the real-world network graphs illustrate several significant findings that have been so far not reported in the literature: (i) the degree-based centrality metrics (especially the eigenvector centrality) exhibit a significantly high positive correlation to the maximal clique size as the networks get increasingly scale-free; (ii) the betweenness centrality of the vertices exhibits a low correlation with that of the maximal size cliques the vertices can be part of; (iii) in real-world networks that are close to random network graphs, the centrality metrics exhibit a low correlation to maximal clique size (especially in the case of shortest-path based closeness and betweenness centrality metrics); (iv) for all the four centrality metrics, the extent of positive correlation with maximal clique size increases as the real-world networks become increasingly scale-free.

REFERENCES

- Brandes, U. (2001). A Faster Algorithm for Betweenness Centrality. *The Journal of Mathematical Sociology*, 25(2), 163–177. doi:10.1080/0022250X.2001.9990249
- Carraghan, R., & Pardalos, P. (1990). An Exact Algorithm for the Maximum Clique Problem. *Operations Research Letters*, 9(6), 375–382. doi:10.1016/0167-6377(90)90057-C
- Cherven, K. (2013). *Network Graph Analysis and Visualization with Gephi* (1st ed.). Birmingham, UK: Packt Publishing.
- Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2009). *Introduction to Algorithms* (3rd ed.). Cambridge, MA: MIT Press.

- Fortunato, S. (2010). Community Detection in Graphs. *Physics Reports*, 486(3-5), 75–174. doi:10.1016/j.physrep.2009.11.002
- Fruchterman, T. M. J., & Reingold, E. M. (1991). Graph Drawing by Force-Directed Placement. *Software, Practice & Experience*, 21(11), 1129–1164. doi:10.1002/spe.4380211102
- Girvan, M., & Newman, M. E. J. (2002). Community Structure in Social and Biological Networks. *Journal of the National Academy of Sciences of the United States of America*, 99(12), 7821–7826. doi:10.1073/pnas.122653799 PMID:12060727
- Koschutzki, D., & Schreiber, F. (2008). Centrality Analysis Methods for Biological Networks and their Application to Gene Regulatory Networks. *Gene Regulation and Systems Biology*, 2, 193–201. PMID:19787083
- Krebs, V. (2008). Retrieved from <http://www.orgnet.com/divided.html>
- Li, C., Li, Q., Van Mieghem, P., Stanley, H. E., & Wang, H. (2014). Correlation between Centrality Metrics and their Application to the Opinion Model. *Physics and Society*. arXiv:1409.6033
- Lusseau, D., Schneider, K., Boisseau, O. J., Haase, P., Slooten, E., & Dawson, S. M. (2003). The Bottlenose Dolphin Community of Doubtful Sound Features a Large Proportion of Long-lasting Associations. *Behavioral Ecology and Sociobiology*, 54(4), 396–405. doi:10.1007/s00265-003-0651-y
- Newman, M. (2010). *Networks: An Introduction* (1st ed.). Oxford, UK: Oxford University Press. doi:10.1093/acprof:oso/9780199206650.001.0001
- Newman, M. E. J. (2006). Finding Community Structure in Networks using the Eigenvectors of Matrices. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, 74(3), 036104. doi:10.1103/PhysRevE.74.036104 PMID:17025705
- Opsahl, T., Agneessens, F., & Skvoretz, J. (2010). Node Centrality in Weighted Networks: Generalizing Degree and Shortest Paths. *Social Networks*, 32(3), 245–251. doi:10.1016/j.socnet.2010.03.006
- Ostergard, P. R. R. (2002). A Fast Algorithm for the Maximum Clique Problem. *Discrete Applied Mathematics*, 120(1-3), 197–207. doi:10.1016/S0166-218X(01)00290-6
- Palla, G., Derenyi, I., Farkas, I., & Vicsek, T. (2005). Uncovering the Overlapping Community Structure of Complex Networks in Nature and Society. *Nature*, 435(7043), 814–818. doi:10.1038/nature03607 PMID:15944704
- Pattabiraman, B., Patwary, M. A., Gebremedhin, A. H., Liao, W.-K., & Choudhary, A. (2013). Fast Problems for the Maximum Clique Problem on Massive sparse Graphs. *Proceedings of the 10th International Workshop on Algorithms and Models for the Web Graph: Lecture Notes in Computer Science* (vol. 8305, pp. 156-169), Cambridge, MA: Springer. doi:10.1007/978-3-319-03536-9_13
- Rossi, R. A., Gleich, D. F., & Patwary, M. A. (2014). Fast Maximum Clique Algorithms for Large Graphs. *Proceedings of the 23rd International Conference on World Wide Web Companion* (pp. 365-366). Seoul, Korea: ACM. doi:10.1145/2567948.2577283
- Sadi, S., Oguducu, S., & Uyar, A. S. (2010). An Efficient Community Detection Method using Parallel Clique-Finding Ants. *Proceedings of IEEE Congress on Evolutionary Computation* (pp. 1-7). Barcelona, Spain: IEEE. doi:10.1109/CEC.2010.5586496
- Strang, G. (2005). *Linear Algebra and its Applications* (4th ed.). Boston, MA: Cengage Learning.
- Tomita, E., & Seki, T. (2003). An Efficient Branch-and-Bound Algorithm for Finding a Maximum Clique. *Proceedings of the 4th International Conference on Discrete Mathematics and Theoretical Computer Science* (pp. 278-289), Dijon, France: Springer. doi:10.1007/3-540-45066-1_22

Zachary, W. W. (1977). An Information Flow Model for Conflict and Fission in Small Groups. *Journal of Anthropological Research*, 33(4), 452–473. doi:10.1086/jar.33.4.3629752

ADDITIONAL READING

Barabasi, A.-L., & Albert, R. (1999). Emergence of Scaling in Random Networks. *Science*, 286(5439), 509–512. doi:10.1126/science.286.5439.509 PMID:10521342

Cardinal, J., Labbe, M., Langerman, S., Levy, E., & Melot, H. (2005). A Tight Analysis of the Maximal Matching Heuristic. *Lecture Notes in Computer Science*, 3595, 701–709. doi:10.1007/11533719_71

Chatterjee, A., Das, D., Naskar, M. K., Pal, N., & Mukherjee, A. (2013). Heuristic for Maximum Matching in Directed Complex Networks. *Proceedings of the International Conference on Advances in Computing, Communications and Informatics* (pp. 1146–1151), Mysore, India: IEEE. doi:10.1109/ICACCI.2013.6637339

Creamer, G. G., Ren, Y., & Nickerson, J. V. (2013). Impact of Dynamic Corporate News Networks on Asset Return and Volatility. *Proceedings of the International Conference on Social Computing* (pp. 809–814), Alexandria, VA, USA: IEEE. doi:10.1109/SocialCom.2013.121

Demange, M., & Ekim, T. (2008). Minimum Maximal Matching is NP-Hard in Regular Bipartite Graphs. *Theory and Applications of Models of Computation. Lecture Notes in Computer Science*, 4978, 364–374. doi:10.1007/978-3-540-79228-4_32

Edmonds, J. (1965). Paths, Trees, and Flowers. *Canadian Journal of Mathematics*, 17(0), 449–467. doi:10.4153/CJM-1965-045-4

Edmonds, J., & Karp, R. M. (1972). Theoretical Improvements in Algorithmic Efficiency for Network Flow Problems. *Journal of the ACM*, 19(2), 248–264. doi:10.1145/321694.321699

Gao, Q. (2010). Network Analysis Applied to a Bibliographic Network. *Proceedings of the Second International Conference on Social Computing* (pp. 615–622), Minneapolis, MN, USA: IEEE. doi:10.1109/SocialCom.2010.96

Guillaume, J.-L., and Latapy, M. (2006). Bipartite Graphs as Models of Complex Networks. *Physica A: Statistical Mechanics and its Applications*, 371(2), 795–813.

Horne, B. G. (1997). Lower Bounds for the Spectral Radius of a Matrix. *Linear Algebra and Its Applications*, 263, 261–273. doi:10.1016/S0024-3795(96)00539-3

Horton, J. D., & Kilakos, K. (1993). Minimum Edge Dominating Sets. *SIAM Journal on Discrete Mathematics*, 6(3), 375–387. doi:10.1137/0406030

Liu, Y.-Y., Slotine, J.-J., & Barabasi, A.-L. (2011). Controllability of Complex Networks. *Nature*, 473(7346), 167–173. doi:10.1038/nature10011 PMID:21562557

Maslov, S., & Sneppen, K. (2002). Specificity and Stability in Topology of Protein Networks. *Science*, 296(5569), 910–913. doi:10.1126/science.1065103 PMID:11988575

Meghanathan, N. (2014). Spectral Radius as a Measure of Variation in Node Degree for Complex Network Graphs. *Proceedings of the 3rd International Conference on Digital Contents and Applications* (pp. 30–33), Hainan, China: IEEE. doi:10.1109/UNESST.2014.8

Muller-Hannemann, M., & Schwartz, A. (1999). Implementing Weighted b-Matching Algorithms: Insights from a Computational Study. *Algorithm Engineering and Computation. Lecture Notes in Computer Science*, 1619, 18–36. doi:10.1007/3-540-48518-X_2

Pena, J., Gutman, I., & Rada, J. (2007). Estimating the Estrada Index. *Linear Algebra and Its Applications*, 427(1), 70–76. doi:10.1016/j.laa.2007.06.020

Peng, T., Zhang, D., Liu, X., Wang, S., & Zuo, W. (2013). Central Author Mining from Co-authorship Network. *Proceedings of the 6th International Symposium on Computational Intelligence and Design* (pp. 228-232), Hangzhou, China: IEEE.

Piraveenan, M., Prokopenko, M., & Zomaya, A. (2008). Local Assortativeness in Scale-Free Networks. *Europhysics Letters*, 84(2), 280–282. doi:10.1209/0295-5075/84/28002

Shaikh, M. A., & Wang, J. (2006). Discovering Hierarchical Structure in Terrorist Networks. *Proceedings of the International Conference on Emerging Technologies* (pp. 238-244), Peshawar, Pakistan: IEEE.

Van Mieghem, P., Wang, H., Ge, X., Tang, S., & Kuipers, F. (2010). Influence of Assortativity and Degree-Preserving Rewiring on the Spectra of Networks. *The European Physical Journal B*, 76(4), 643–652. doi:10.1140/epjb/e2010-00219-x

Wang, H., Winterbach, W., & Van Mieghem, P. (2011). Assortativity of Complementary Graphs. *The European Physical Journal B*, 83(2), 203–214. doi:10.1140/epjb/e2011-20118-x

Winterbach, W., de Ridder, D., Wang, H. J., Reinders, M., & Van Mieghem, P. (2012). Do Greedy Assortativity Optimization Algorithms Produce Good Results? *The European Physical Journal B*, 5(5), 151–160. doi:10.1140/epjb/e2012-20899-2

Yang, C., Poser, B. A., Deng, W., & Stenger, V. A. (2012). Spectral Decomposition of Susceptibility Artifacts for Spectral-Spatial Radiofrequency Pulse Design. *Magnetic Resonance in Medicine*, 68(6), 1905–1910. doi:10.1002/mrm.24208 PMID:22334396

Yannakakis, M., & Gavril, F. (1980). Edge Dominating Sets in Graphs. *SIAM Journal on Applied Mathematics*, 38(3), 364–372. doi:10.1137/0138030

Yarahmadi, Z., Doslic, T., & Ashrafi, A. R. (2010). The Bipartite Edge Frustration of Composite Graphs. *Discrete Applied Mathematics*, 158(14), 1551–1558. doi:10.1016/j.dam.2010.04.010

KEY TERMS AND DEFINITIONS

Algorithm: A sequence of steps to perform a particular task.

Betweenness Centrality: A measure of the fraction of shortest paths between any two vertices that go through a vertex, considered over all pairs of vertices excluding the vertex.

Branch and Bound: An algorithmic strategy that prunes the search space and explores only candidate solutions that have the potential to be better than the currently known solution.

Centrality: A quantitative measure of ranking the vertices of a graph based on its topological structure.

Clique: A sub graph of a graph such that there is an edge between any two vertices.

Clique Size: The size (number of vertices) of the largest clique a vertex is part of.

Closeness Centrality: A measure of the sum of the lengths of the shortest paths from a vertex to the rest of the vertices.

Computationally Lightweight Metric: A network metric whose computation is not time consuming.

Correlation: A statistical dependence between two variables or sets of data.

Degree Centrality: A measure of the number of adjacent neighbors of a vertex.

Eigenvector Centrality: A measure of the degree of a vertex as well as the degree of its neighbors.

Random Network: A network whose degree distribution follows a Poisson pattern wherein the degree of all the vertices lies close to the average degree and the probability of finding a vertex with a degree much farther away from the average degree is zero.

Spectral Radius: The largest Eigenvalue of the adjacency matrix of a graph.

Scale-Free Network: A network whose degree distribution follows a power-law pattern wherein a majority of the vertices have a smaller degree,

but there exists a non-zero probability for finding a vertex with a relatively much larger degree.

Small-World Property: A characteristic property of complex networks wherein the number of hops on the shortest paths between any two vertices is at most the logarithm of the number of vertices in the network.

Topology: An arrangement of the nodes in the network that can be used for communication between any two nodes in the network.

N

Distributed Methods for Multi-Sink Wireless Sensor Networks Formation

Miriam A. Carlos-Mancilla

CINVESTAV Unidad Guadalajara, Mexico

Ernesto Lopez-Mellado

CINVESTAV Unidad Guadalajara, Mexico

Mario Siller

CINVESTAV Unidad Guadalajara, Mexico

INTRODUCTION

A Wireless Sensor Network (WSN) is composed of a set of hundreds of devices deployed in a common area, which are connected among them. Each device has the same capabilities; it is able to collect information from the surrounding environment through one or more sensors, process this information in a local manner. Data collection is usually made by a sink node or base station. A WSN can be used to control multiple tasks in one environment; the sensors require of data sensed from multiple sources by delivering information to multiple sinks at the same time.

There are a wide variety of current research problems in this area, such as tracking, surveillance, military applications, building automation, disaster management, and agriculture, among others. The applications are designed to obtain precise information from the environment regardless of the human presence. Radio communication, processing, transmission/reception generally are the main cause of power consumption. A multi-sink approach is becoming an efficient scheme to reduce the energy consumption in the whole network, either in a centralized or distributed application. In this paper, distributed techniques are presented and analysed.

The analysis is focused on the used strategies in a multiple sink environment, whether the nodes are static or mobile and how applications can be

improved by mixing more than one strategy; the objective of the application such as event detection, following a target, finding the best route, collecting and data aggregation, among others. The survey covers recent and traditional works and presents a discussion of their advantages and drawbacks. A simple and operational distributed algorithm that performs network formation and data collection using a multi-sink environment is presented.

This article is organized as follows: the next section presents the generalities and features of WSNs. The third section presents a classification of distributed approaches based on multi-sink environments; the fourth section describes the classification of the related works and discusses their proposals. The fifth section focuses on the simple and operational proposal of a distributed and multi-sink protocol for a WSN; finally, the last section presents the concluding remarks, which drive further research in the area.

BACKGROUND

Wireless Sensor Networks

Wireless Sensor networks with a single sink node have constraints; one of them is presented during the data collection, where bottleneck (Hochbaum & Shmoys, 1986) and hotspot problems (Sungjin

& Kim, 2006) can arise. Another limiting factor emerges when the purpose of the application is finding a target and the sink is far away from it; in this case, the routing can require great expenses of energy. Another drawbacks are presented when the sink require of responses in a short period of time and the routing is not efficient, when the sink node is not available, or there is a high demand for transmissions and one sink is not enough.

Multi-sink platforms have arisen as response to this kind of problems; a multi-sink platform provides redundancy, avoids problems in the data collections and increases the efficiency of the application. A set of sinks can be strategically or randomly distributed in the coverage area to collect and process the information from the environment. The sink or sinks can be static or mobile according to the available resources of the application.

Once the topology strategy is specified, the information can be collected following a distributed or centralized policy via multi-hop mode. In this article, we focus on the distributed strategies. *Distributed techniques* allow to the nodes take their own decisions in a locally way and dispense the information to a limited set of neighbours nodes within a scope area, which provide energy savings and spreading the information in a uniform manner to avoid the overhearing or hot spot problems. The flexibility of distributed systems

enables targeting random environments without complications during the information collection.

A WSN aims to gather environmental data, the node devices placement may be known or unknown a priori. Network nodes can have a *driven* (IP communication) or *logical* communication (topological communication defined by users or by the application) with all devices; such communication defines the topology according to the application. The logical topology is mainly defined based on the nodes communication. It can be either ad-hoc or strategy based (self-organization, clustering, pheromone tracking, and so on) (Carlos-Mancilla et al. 2016).

Certainly, one of the most important advances in the current proposals has been the *self-organization* (SO) approach in which the nodes are considered autonomous; they interact with the neighbour nodes and are independent of the information of the rest of the nodes. The aim is to achieve collective tasks that exceed its individual capabilities. Examples of these techniques are found in nature (insect colonies, biological cells, the flock of birds, the foraging behaviour of ants, etc.) (Mamei et al., 2006), (Schmeck et al., 2010). Finally, distributed multi-sink approaches support large density scenarios without compromising the network performance. In Table 1 a summary of the main features considered in distributed and centralized multi-sink applications is presented.

Table 1. Metrics in multi-sink applications

Metrics	Centralized	Distributed
Large scale	It is not applicable in big scale networks; the handled information is high.	The handled information is within a range of transmission; small or large ranges do not affect the performance of the network.
Routing	It can be efficient for small networks.	The routing is fast for the amount of information managed.
Energy saving	The number of transmissions and receptions needed to refresh the whole network is expensive.	It is low for transmissions, receptions, and notification of changes in the network.
Redundancy	It is good when there is a failure shown in the system.	According to the application, the redundancy can be faced in an efficient manner.
Reconfiguration	It is expensive spreading the changes to the whole network	The reconfiguration accuracy depends on the application.

CLASSIFICATION OF MULTI-SINK TECHNIQUES

Distributed Multi-Sink Applications (DMSA) have been characterized to be one of the robust and reliable techniques in wireless sensor networks. In such techniques, every node is responsible for providing operation, such as event detection, routing, topology formation, sensing, data collection, and information processing according to the available information in their coverable area. Distributed and multi-sink techniques together allows increasing the efficiency of a network. Multiple sink environments are classified according to the support operation of the nodes in the network, either *mobile* or *static*. Besides, a subdivision of both techniques can also be made *by the application*, and the *technique* used during any procedure. The taxonomy of the classification is shown in Figure 1.

The subclass *By application* in both cases static and mobile can be divided into the main research of any application, such as *Routing*, *Event*

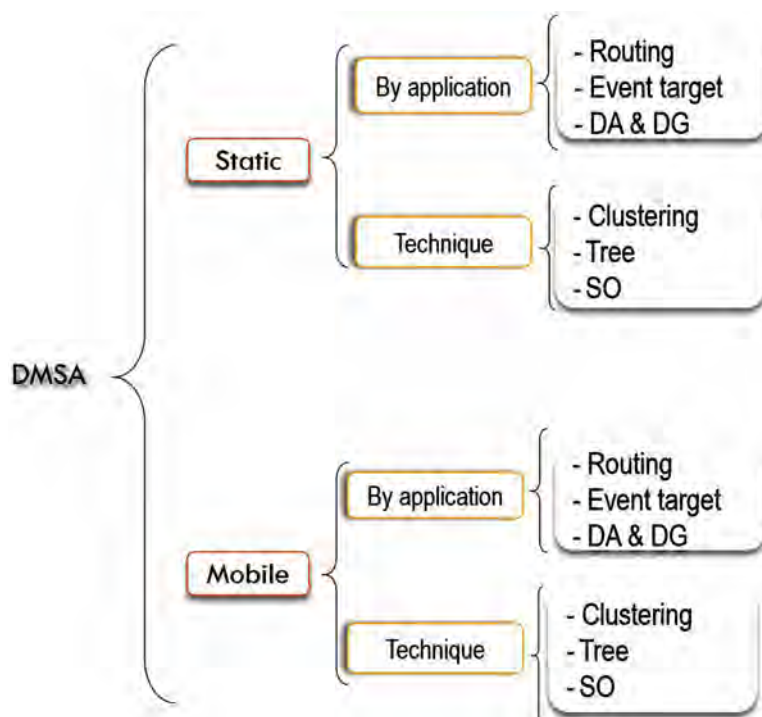
target, and *DA & DC* (Data Aggregation and Data Collection). Whereas the *Technique* subclass is divided into *Clustering*, *Tree formation*, and *SO* approaches. Hybrid approaches can appear to improve an application. An explanation in detail of this classification, protocols, and techniques are reviewed in next section.

DISTRIBUTED APPROACHES USING MULTI-SINK ENVIRONMENTS

Distributed protocols in wireless sensor networks must provide a reduction in the energy consumption, robustness, and reliability algorithms despite the limited energy, noise, loss of messages, and obstacles.

Recently, many works focus on reducing the energy consumption, minimising the number of hops, providing robustness to the network, increasing the throughput, and so on. However, a route with the minimum number of hops may not always be the optimal and reliable route. A

Figure 1. Taxonomy of the multi-sink environments



network with a large number of devices is not always the best option, or even minimising the number of transmissions can have adverse effects. For the above reasons, according to the nature of the application, it is important to define the set of metrics, constraints, and features to evaluate the performance of a network (Muhammad et al., 2015).

In this chapter, we propose a classification and discuss relevant research works included in every one of them. A work can be included in more than one classification and they are focused in distributed multi-sink environment using diverse strategies.

Application Oriented

An application is defined by the available resources and the environment where the devices will be deployed. According to the application, the nodes have an assigned task; every task of a node has a main purpose such as finding the best available route, discovery a specific target, collect data, etc., in order to reduce the energy consumption, prolong the lifetime, generate a robust network or reducing the use of the resources.

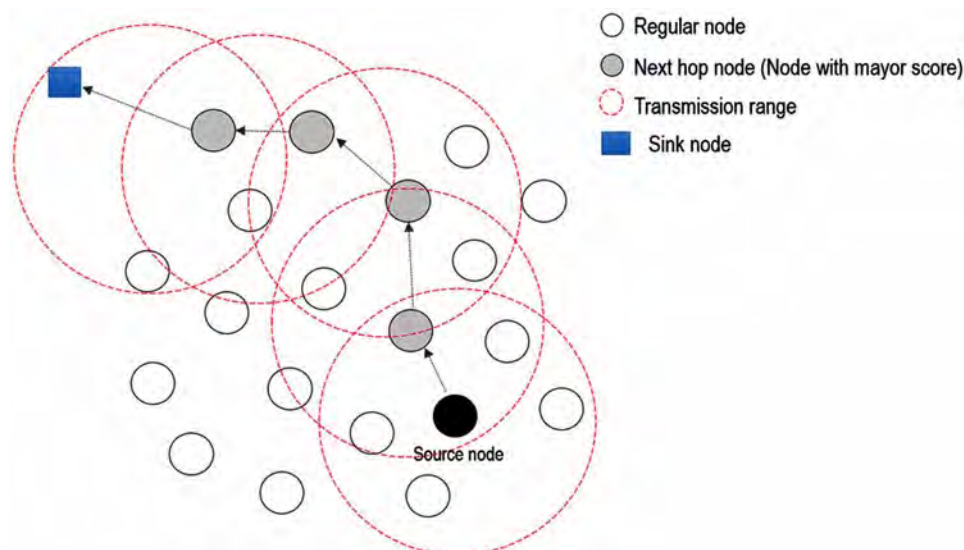
We consider that the available applications can be classified into three categories: *Routing*, *Event target*, and *Data Aggregation & Data Collection* (DA & DC).

Routing

Routing is a process that selects the best route available in a defined environment, in a common area or using a specific number of nodes. The route is determined based on different metrics such as evaluating the number of hops to get to the nearest sink node, the minimum distance between the nodes, the weight of the nodes, the residual energy, the number of visited nodes, and so on. This is shown in Figure 2.

For instance, a tolerance control for enhancing a wireless sensor network using a multi-sink environment is presented in (Huang, 2010). In this work a tolerance area is defined; the nodes out of this area select the shortest path to route their data from source nodes to the nearest sinks. The sensor nodes in this area might switch to different paths to balance the load of sensors in the heavy area. Sensors are deployed uniformly in the network. Sensors outside the area, which are enough close to the area, such that the distance is less than a

Figure 2. Routing process



threshold τ , are considered as located in the area. The properly enlargement of the tolerance area would enhance effectively the data delivery rate.

Drawbacks of this approach appear when the loading of sensors in the network is not balanced; thus, sensors sometimes should select a candidate route to bypass the sensors with too much load or low battery level. The data rate decrease with few sink nodes and the energy depletion is high when looking for alternative routes based on the tolerance area.

An energy conserving dynamic routing with heterogeneous sensor networks is presented in (Lin, 2010); the authors present a fully distributed routing algorithm to support large scale networks based on the Bellman-Ford algorithm (Sedgewick, 2015) for routing. Two kinds of sensors are used; a large number of *L-sensor* which is powered by batteries with limited energy and communication capacity, and several *H-sensor* that serve as sink node with significantly higher computing power and communication capability. It is assumed that *L*-nodes only dissipate energy in data transmission.

The routing exploration is based on the broadcasting method and the nodes build a routing table. Every *L-sensor* receive its neighbours' routing messages, computes a new PBEC value, and broadcasts its updated routing information if the PBEC (Path Bottleneck Energy Cost) value is changed. This paper presents a significant improvement compared with two centralized methods. The energy cost for discovering a neighbourhood is not considered and the cost for data broadcasting method is expensive. The authors do not take into account reconfiguration and data seems to be forwarded always following the same route. This proposal also belongs to the tree formation classification.

Another proposal based on the formation of the clusters and routing is presented in (Xu, 2012); this paper is focused on the reduction of the energy consumption during the cluster formation and submitting information; explained in a posterior section.

A routing protocol to balance the loads in the network with multiple sinks to prolong the lifetime of the system is presented in (Zhang, 2015). After a children node generates or receives a packet, it randomly chooses a parent node using a probability list and takes the chosen node as the next-hop destination of the packet.

The sinks select one of them to be the head sink, which serves as an arbitrator. The arbitrator sinks decide the assignment of nodes to sinks belong based on two criteria: the hop number and the number of nodes of the sinks. In some situations, the tree structure cannot guarantee the balancing of the loads, some collisions can appear, and messages can be lost. Data aggregation is not considered.

An intelligent Agent-based Routing protocol that provides data delivering to mobile sinks is presented in (Kim, 2010). The communication with the agents is performed through a broadcasting method. In this protocol the messages can be lost during the data collection process and does not support channel variability. The energy consumption is also related to the source data rate, which is independent of the protocol.

Event Target

This kind of strategies is used when the nodes in the environment look for a specific node, a target, or a specific area for sensing in the environment. Figure 3 presents an example of an event target network; once the nodes detect an event, they send the information to the nearest sink node.

For instance, a virtual infrastructure based on honeycomb tessellation for data dissemination is presented in (A. T. Erman, 2012). This work can be classified as a routing or event target proposal.

The virtual infrastructures support mobile sinks in a WSNs and it acts as a rendezvous region for storing and retrieving collected event data. When the mobile sink crosses the network, the sensors in the rendezvous region are queried to notify the event data. The virtual infrastructure is based on honeycomb tessellation, the data dissemination is

hexagonal cell-based. A drawback in this proposal is when an event is detected and there is not a sink near the sensed area; the data aggregation and failures are not considered.

A multiple data sink node election in a multi-sink environment is presented in (Pietrabissa, 2016). This algorithm works in a distributed and iterative manner, in which the sink role is periodically reassigned to avoid the hot spot problem and increase the lifetime of the network. The network partition is according to a Centroidal Voronoid Tessellation method which leads to a spatially well-balanced distribution. This proposal does not take into account the energy depletion in the topology formation nor in the overhearing problem.

Algorithms that build a backbone connecting mobile devices in a distributed fashion using an adaptive learning scheme is presented in (Tzong, 2012). The authors consider a load balancing strategy with a zone partition in a multi-sink

WSN environment (QAZP). The movements of a node are predefined; reconfiguration to failures is not addressed.

Data Aggregation and Data Gathering (DA and DG)

In this kind of applications, the formation of the network is not important; the main goal is the way the data is collected. Some applications look for spread the data collection through defined nodes to get the Sink node, while other applications focus on the energy conservation, improving the performance, increasing the throughput, and so on. Figure 4 presents an example of Data collection in a multi-sink environment; the nodes look for the nearest sink in their area to send the information. The election of the routing nodes can be based on different parameters, defined according to the application.

Figure 3. Event detection process in the network

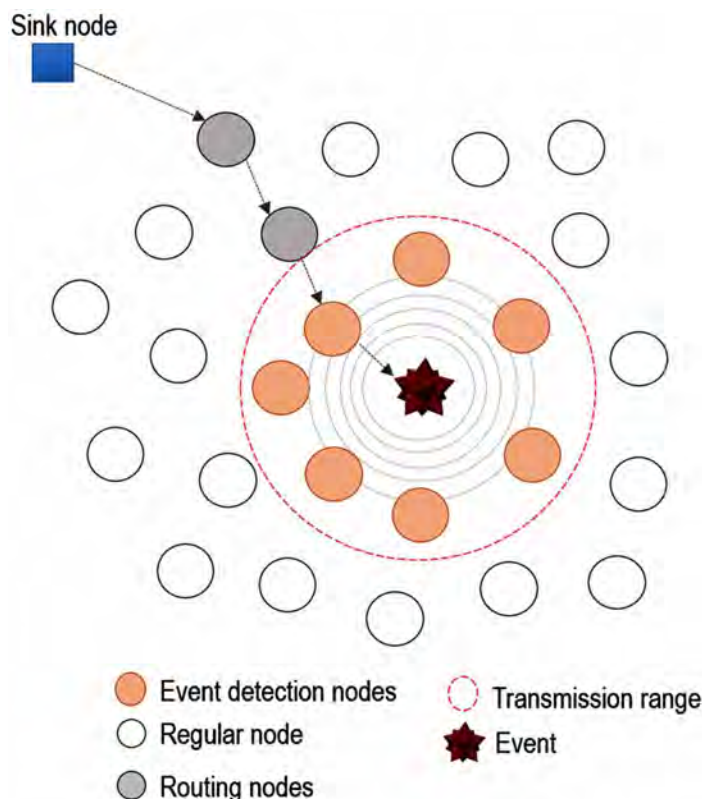
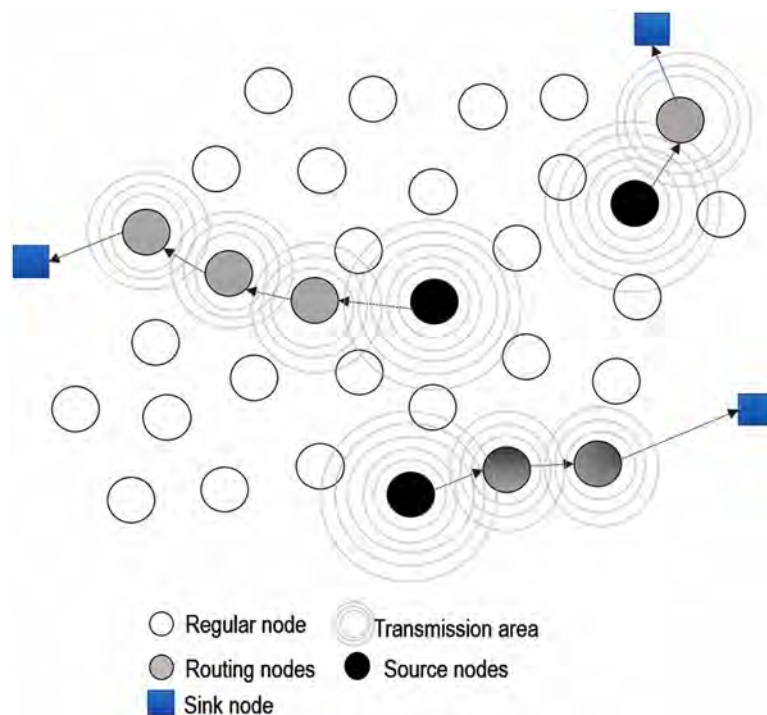


Figure 4. Data collection in a multi-sink environment



An example of data collection through defined nodes is presented in (Winston K. G. Seah, 2006). In this work, a virtual sink architecture for WSN that mitigates the near-sink contention by defining a group of spatially physical sink is proposed. A multi-path routing is adopted to provide alternative paths to increase the probability of successful delivery.

The authors propose a multi-tier topology by introducing local aggregation points and distributing them amongst sensing nodes and multiple local sinks. A node sends a packet simultaneously for diverse routes to multiple sinks. The network increases the probabilities to deliver the data but the energy consumption increase. The data transmission cost is high and the channel conditions are assumed without bad conditions. It is assumed that the local links are connected via high-speed links to a network where the resources are enough to support communications needs.

A data gathering for large scale sensor networks with multiple sinks (M-collectors), which

traverse several shorter sub-tours concurrently to satisfy the distance/time constraints is presented in (Ma, 2013). The objective is to find a set of data-gathering sub-tours in the network such that the number of M-collectors can be minimized.

An M-collector is responsible for gathering data from local sensors in the subarea and the information is forwarded to only one M-collector which has a connection with a final data sink. It is not guaranteed always an available sub-tour, the energy depletion during the sub-tour searching is not considered, and the transmission would be not possible if the M-collector, which has direct communication with the data sink, dies.

Classification by Technique

Clustering

In this strategy the deployed sensor nodes are organized to form clusters; every cluster has a cluster head (CH). The CH is elected according

to the node capabilities with respect to other member nodes in the cluster. CHs are able to collect data from their member nodes, processing, adding data, and forward it to one or multiples sink nodes (Figure 5).

For instance, an Energy-efficient Multi-sink Clustering Algorithm (EMCA) is proposed in (Zhanyang Xu, 2012) to solve the *energy hole* problem. Its inter-cluster and intra-cluster routing algorithm focuses on energy consumption and ensures the optimal path from nodes to its cluster head or sink. The network is divided into n equal clusters; each cluster has a cluster head chosen by considering the residual energy and sends aggregated data to the relevant sink.

The routing is used to calculate the nearest distance between the cluster head nodes and the sink nodes. The data is sent in a multi-hop manner. The energy consumption decreases while the number of the sink increases. No one reconfiguration method is considered.

In (Wang, 2015) a routing protocol algorithm that combines Ant Colony Optimization (ACO) (Dorigo, 2006), clustering, and sinks mobility techniques for home automation networks is presented. The network is divided into several clusters; each one led by a cluster head that com-

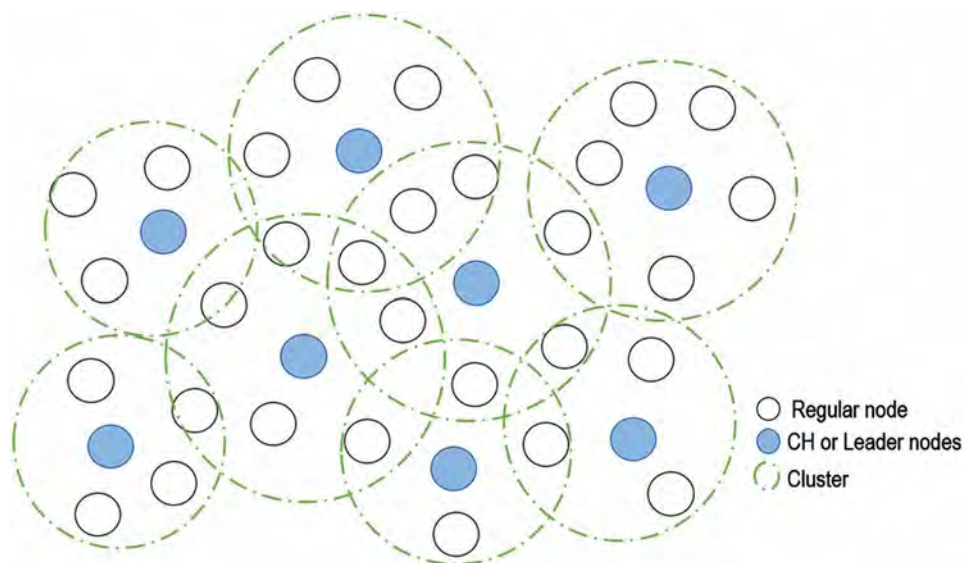
municates with a mobile sink. ACO is applicable to the guidance of the sink mobility; if the ACO fails, the mobility of the sink is being affected. Obstacles and reconfiguration are not considered, the cluster formation is energy expensive.

In (Wang et al, 2013), two algorithms for sink mobility based on clustering strategy and reduction of the energy consumption are proposed. The objective of this work is to study the influence of fixed and mobile sink strategies on home network performance in terms of energy consumption, network lifetime, and mitigation of the hot spot problem. The authors consider the study of diverse parameters including number of mobile sinks, velocity, location, and moving trajectory. Simulations are presented and compared with LEACH (Heinzelman, 2000).

Tree Formation

A tree formation strategy is a known technique to reduce the overhead and the energy consumption; it also helps to create routes to send data through the network (Figure 6). However, a bad implementation of a tree formation can create deep branches, which become an inefficient way to collect information.

Figure 5. Clustering technique



One technique for tree formation is presented in (Mottola, 2015) where distributed heuristics are embodied to minimize the number of nodes involved in routing; furthermore, there is load balancing strategy using information available in a 1-hop neighbourhood. A tree formation strategy is used to build the best available route to the sink nodes; for this purpose, multiple routes are mixed into a single one to send information to different sink nodes and to assure loaded balanced routes. The technique does not consider the density of the network nor transmission of excessive messages, depletion of energy in the routing formation, or reconfiguration.

Self-Organization

The self-organization is an effective method to prolong the lifetime of ad-hoc networks and let the nodes decide the actions to perform according to available information of the environment. The nodes have communication with their neighbour nodes and the decisions are made locally (Figure 7).

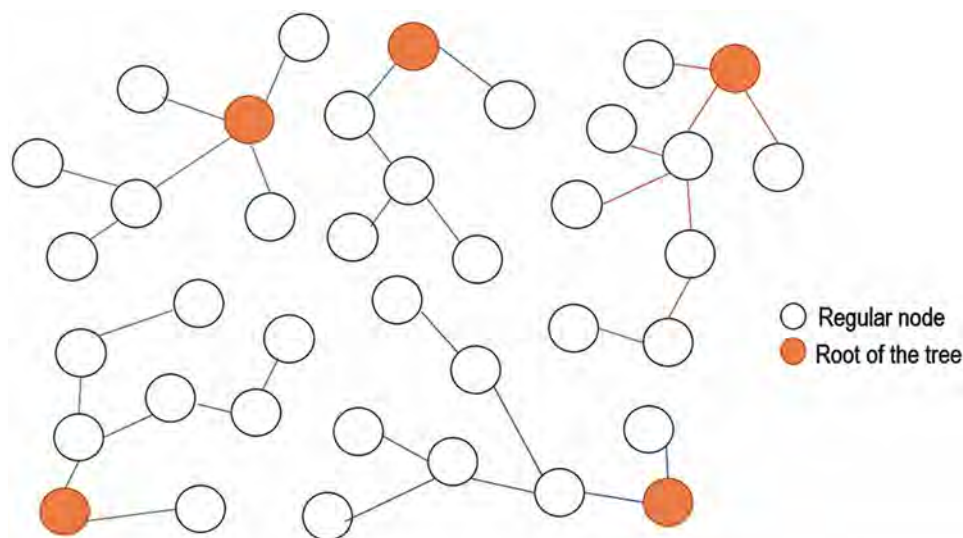
A multi-hop architecture *ICatchYou* is proposed in (Silva, 2009); it provides self-configuration abilities and allows the easy and quick integration of nodes in the network. Self-configuration

procedures are required to support the deployment of a random multi-sink.

ICatchYou performs two processes: *registration* and *updating* of information. The sink nodes should be accessible from conventional IP networks, providing the measured data in real-time for remote destinations, and allowing a simpler management and interoperability with the WSN. Depletion of the energy is not considered, and the efficiency of the nodes depends on the link quality.

In (Kiri, 2007) a system for gathering data from sensor networks with multi-sink configurations is presented; the strategy is inspired by the swarm intelligence of ants, in which each sensor node determines its next action through repeated interaction with its neighbours. Clustering and routing strategies emerge in a self-organized manner. The authors focused the problem on the communication from the sensor to the sink nodes; the number of sensors is equal to the number of clusters in the environment. Data processing and energy consumption are not considered. If a sink fails the cluster is not able to collect data. Data collection over an area may be impossible due to cluster-level power depletion where a number of events occur in a particular area.

Figure 6. Tree formation technique



A DISTRIBUTED METHOD FOR MULTI-SINK ENVIRONMENTS

In this chapter, we present an extension to a previous technique for formation and networking operation in multi-sink environments proposed in (Carlos-Mancilla, et al., 2015).

The main feature of this proposal is that the algorithm is distributed completely; nodes only use local information. The algorithm is divided into three main stages: First, the backbone formation, which is based on self-organization and clustering strategies (Figure 8). This stage allows low energy consumption, information redundancy, and robustness. In the second stage, the sink nodes initiate a tree formation procedure using a load balanced strategy. Finally, in the third stage, the nodes sense and collect the data from the environment according to the role of the node.

The main contributions of this algorithm include:

- A reliable multi-hop multi-sink algorithm, which creates tree routes from the network formation.
- Local load balancing is introduced from the tree formation.

It is assumed that a finite number of static nodes are randomly deployed in a multi-sink environment without any a-priori information. Different roles for the nodes are considered, they are: *Leader*, *Member*, *Gateway* and *Bridge*; a bridge role connects segments of the network. A segment is a set or different cluster which cannot be seen each other by a gateway node explained in more detail in (Olascuaga-Cabrera, 2011) presented in Figure 8.

The management of the nodes, reduction of the number of transmissions, delays, scheduling, and overhead control are issues addressed in the second stage. The sink nodes are the roots of the trees (level 0) and start the procedure in which the levels are assigned and distributed; also the relationship successor/predecessor is established. Redundant links and cycles are deleted, without loss of connectivity in the whole network (Figure 9).

Once the tree formation is over, the nodes change their role to Leaf node (L-node) which used to have Member and Bridge role, and Intermediate nodes (Inter-node) which used to have Leader and Gateway roles. The data collection starts in the leaf nodes until the data reach the sink node which is attached to.

Figure 7. Ad hoc networks

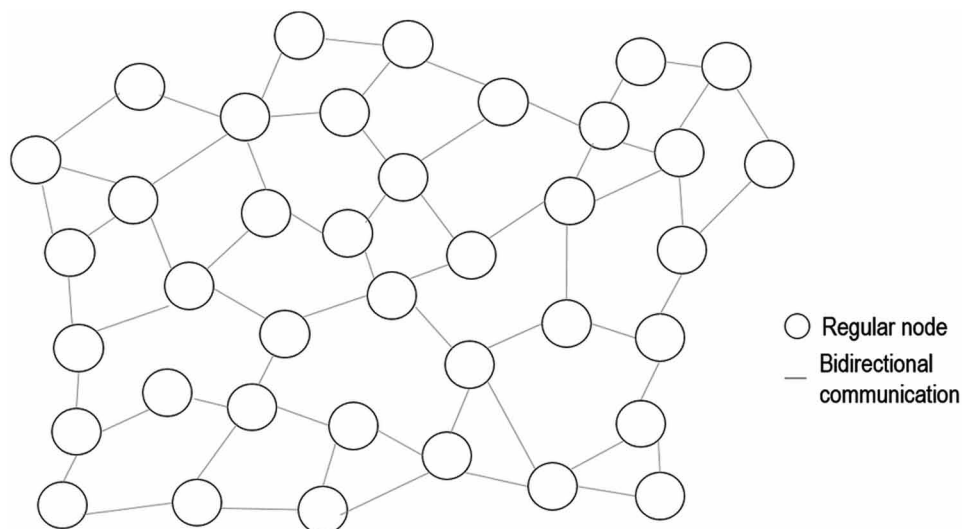


Figure 8. Cluster formation using a self-organized strategy

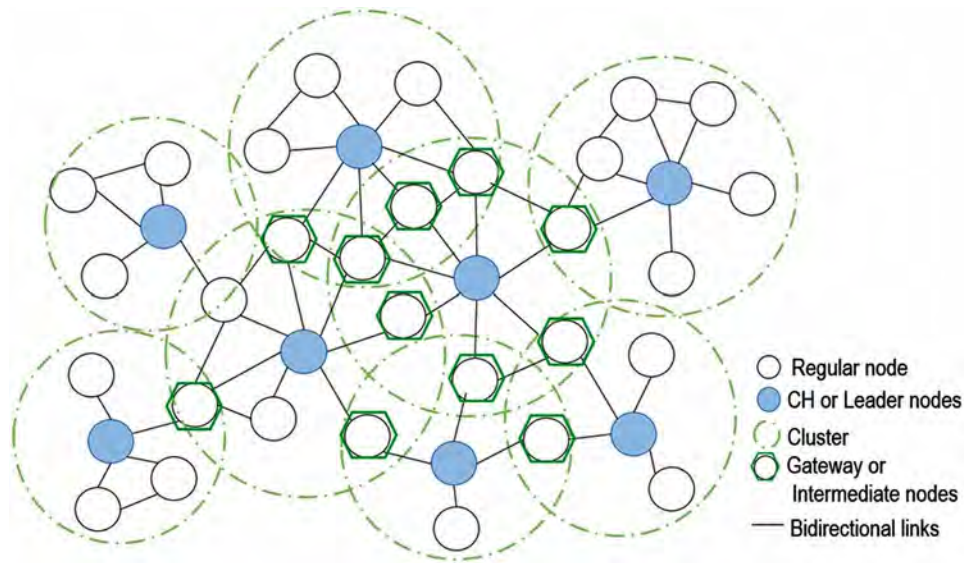
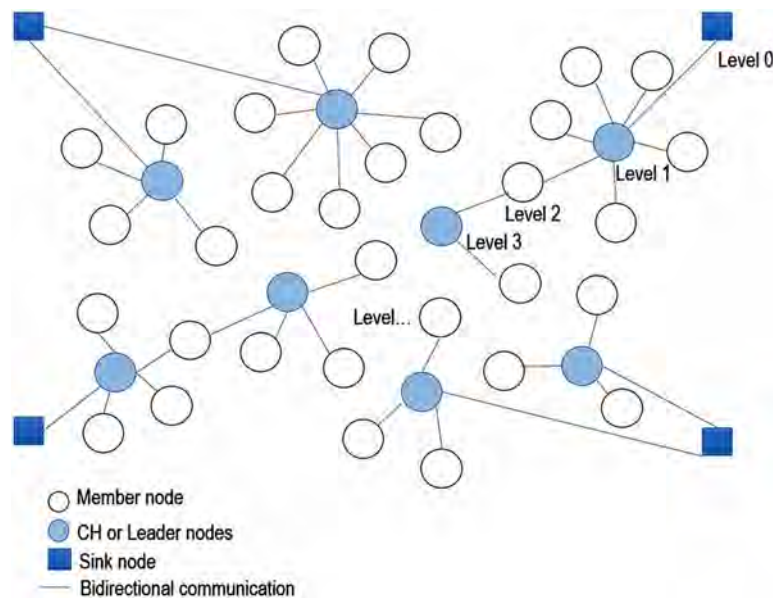


Figure 9. Tree formation algorithm



FUTURE RESEARCH DIRECTIONS

Many applications can be improved using more than one technique in one application. For instance, a lot of research includes routing with tree formation technique, or cluster/tree formation strategies, event target, routing, self-organization, etc. The application of more than one method depends

on the available resources and the environment where will be applied. In (Carlos-Mancilla, 2015) a reconfiguration method for failures has to be considered and the data collection has to contemplate other possible collection strategies for unexpected event detections.

Many strategies have been proposed to improve the performance of one application considering

one or more metrics to the detriment of others. For instance, if the work is focused on routing, most of the cases the number of transmissions is excessive and the energy consumption decrease quickly, when energy consumption is the purpose of the proposal, other metrics will be sacrificed according to the number of devices, distance, channel availability, etc. In Carlos-Mancilla-2015 the proposal considers important to preserve all the mentioned metrics.

Some considerable improvements to this kind of networks are the implementation of solar panels in devices to avoid the energy constraints. Additional works look to send the information to the cloud (Internet of things- IoT) to access easily everywhere, also other researches want to add smart characteristics to the sensor for smart applications, such as home automation, health caring, special body care, and so on.

CONCLUSION

In this chapter, relevant works on distributed multi-sink wireless sensor networks (WSN) have been reviewed. It presents the evolution, design, and implementation of some important WSN techniques in the last years and the most used protocols and standards to improve the sensor applications.

One has remarked that distributed solutions are preferred over centralized ones, since distributed techniques support scalability, autonomous nodes, deployment and elimination of nodes; also, it is possible to use self-organization strategies inspired from nature, in which the interaction is made only with neighbour nodes.

As an example of distributed strategy for multi sink network formation, the last section over-viewed a proposal that involves a self-organization strategy for backbone formation; afterward, the tree formation strategy uses local information and builds the load balanced trees over the formed

network in a multi-sink environment. Finally, sensing and data collection processes are executed at every defined period of time. The presented algorithm copes with challenging problems inherent to wireless sensors, such as energy consumption, memory, latency, loss of messages, routing, and the number of retransmissions that affect the performance of the whole network.

REFERENCES

- Ahn, S., & Daeyoung, K. (2006, February). Pro-active context-aware sensor networks. *Proceedings of European Workshop on Wireless Sensor Networks*. doi:10.1007/11669463_6
- Carlos-Mancilla, M., López-Mellado, E., & Siller, M. (2015). A localized multi-sink multi-hop algorithm for Wireless Sensor Networking. In IEEE (Ed.), *Global Information Infrastructure and Networking Symposium (GIIS)*. IEEE. doi:10.1109/GIIS.2015.7347181
- Carlos-Mancilla, M., López-Mellado, E., & Siller, M. (2016). Wireless Sensor Networks Formation: Approaches and Techniques. *Journal of Sensors*, 1-18.
- Chiming Huang, R.-H. C.-R.-I. (2010, August). Enhancing Network Availability by Tolerance Control in Multi-Sink Wireless Sensor Network. IEEE In (Ed.), *Conference on Information Technology Convergence and Services (ITCS)* (pp. 1-7). Cebu: IEEE.
- Dorigo, M., Birattari, M., & Stützle, T. (2006). Ant Colony Optimization – Artificial Ants as a Computational Intelligence Technique. *IEEE Computational Intelligence Magazine*.
- Dressler, F. (2008, August). A study of self-organization mechanisms in ad hoc and sensor networks. *Published in Computer Communications*, 31(13), 3018–3029. doi:10.1016/j.comcom.2008.02.001

- Erman, A. T. (2012). A virtual infrastructure based on honeycomb tessellation for data dissemination in multi-sink mobile wireless sensor networks. *EURASIP Journal on Wireless Communications and Networking*, 1(17).
- Heinzelman, W. R., Chandrakasan, A., & Balakrishnan, H. (2000, Jan). Energy efficient communication protocol for wireless microsensor networks. *Proc. of the 33rd Annual Hawaii International Conference on System Sciences (HICSS)*, 1-10. doi:10.1109/HICSS.2000.926982
- HochbaumDorit, S., & Shmoys, D. B. (1986, July). A unified approach to approximation algorithms for bottleneck problems. *Journal of the ACM*, 33(3), 533-550.
- Jae-Wan, K., In, J.-S., Hur, K., Kim, J.-W., & Eom, D.-S. (2010, November). An Intelligent Agent-based Routing Structure for Mobile Sinks in WSNs. *Proceedings on IEEE Transactions on Consumer Electronics*, 53(4), 2310–2316.
- Jim, W., Jiayi, C., Bin, L., & Sungyoung, L. (2015). Bio-inspired ant colony optimization based clustering algorithm with mobile sinks for applications in consumer home automation networks. *Published in Consumer Electronics on IEEE Transactions*, 61(4), 438–444. doi:10.1109/TCE.2015.7389797
- Lin Yunyue, Q. W. (2010, April). Energy-Conserving Dynamic Routing in Multi-Sink Heterogeneous Sensor Networks. *IEEE In (Ed.), International Conference on Communications and Mobile Computing (CMC)*. Shenzhen: IEEE.
- Mahmood, M. A., Seah, W. K. G., & Welch, I. (2015). Reliability in wireless sensor networks: A survey and challenges ahead. *Published on Computer Networks*, 29, 166–187. doi:10.1016/j.comnet.2014.12.016
- Mamei, M., Menezes, R., Tolksdorf, R., & Zambonelli, F. (2006, April). Case studies for Self-organization in Computer Science. *Journal of Systems Architecture*, 52(8), 443–460. doi:10.1016/j.sysarc.2006.02.002
- Ming, M., Yang, Y., & Zhao, M. (2013, May). Tour planning for Mobile Data-Gathering Mechanisms in Wireless Sensor Networks. *Published in IEEE Transactions on Vehicular Technology*, 62(4), 1472–1483. doi:10.1109/TVT.2012.2229309
- Mottola Luca, G. P. (2011). MUSTER: Adaptive Energy-Aware Multisink Routing in Wireless Sensor Networks. *Mobile Computing on IEEE Transactions*, 10(12), 1694–1709. doi:10.1109/TMC.2010.250
- Olascuaga-Cabrera, J. G., López-Mellado, E., Mendez-Vazquez, A., & Ramos-Corchado, F. F. (2011). A self-organization algorithm for robust networking of wireless devices. *IEEE Sensors Journal*, 11(3), 771–780. doi:10.1109/JSEN.2010.2068286
- Pietrabissa, A., Liberati, F., & Oddi, G. (2016). A distributed algorithm for Ad-hoc network partitioning based on Voronoi Tessellation. In *Ad-Hoc Networks*. Elsevier Publications.
- Schmeck, H., Muller-Schloer, Cakar, Mnif, & Richter. (2010, Sep). Adaptivity and Self-organization in Organic Computin Systems. *ACM Transactions on Autonomous and Adaptive Systems*, 5(3), 10:1-10:32.
- Sedgewick & Wayne. (2015). Algorithms. In *Shortest Paths*. Pearson.
- Silva Ricardo, J. S. (2009). A New Approach for Multi-Sink Environments in WSNs. *IFIP/IEEE International Symposium on Integrated Network Management (IM 2009)*. IEEE. doi:10.1109/INM.2009.5188794
- Tzong Sheng Cheng, T.-Y. C. (2012, August). An adaptive learning scheme for load balancing with zone partition in multi-sink wireless sensor network. *Expert Systems with Applications*, 39(10), 9427–9434. doi:10.1016/j.eswa.2012.02.119

Wang, J., Yin, Y., Zhang, J., Lee, S., & Sherratt, R. S. (2013, February). Mobility based energy efficient and multi-sink algorithm for consumer home networks. *IEEE Transactions on Consumer Electronics*, 59(1), 77–84. doi:10.1109/TCE.2013.6490244

Winston, K. G., & Seah, H. P. (2006). Multipath virtual sink architecture for wireless sensor networks in harsh environments. *Proceedings of the first international conference on Integrated internet ad hoc and sensor networks*.

Yuichi, K., Sugano, M., & Murata, M. (2007). *Self-Organized Data-Gathering Scheme for Multi-sink Sensor Networks Inspired by Swarm Intelligence*. First International Conference on Self-Adaptive and Self-Organizing Systems (SASO 2007), Cambridge, UK.

Zhang Chongqing, W. Y. (2015). Load-balancing Routing for Wireless Sensor Networks with Multiple Sinks. *12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*. IEEE.

Zhanyang Xu, Y. Y.-U. (2012, December). An Energy-efficient Multi-sink Clustering Algorithm for Wireless Sensor Networks. *International Journal of Control & Automation*, 5(4), 131-142.

KEY TERMS AND DEFINITIONS

Ad-Hoc Networks: The network topology is determined by the sensor locations, which are not known a priori, and do not require a router or a controlled wireless base station. The data collection is dynamically defined with the connectivity of the nodes.

Distributed Strategies: They are schemes in which the task coordination is not performed by a single sensor or sink; instead, the devices interact among them to achieve a common goal through message passing.

Multi-Sink Environments: It is a networking problem in which two or more base nodes or sinks are deployed in an environment to collect data; multiple sinks allow the distribution of traffic with the aim of reducing congestions.

Self-Organisation: The coordination and control needed to form a network emerge through the local interaction of a group of sensors/devices from an initially disordered network y/o system.

Sensor Networks: They are formed by a group of tiny, typically battery-powered devices and wireless infrastructure that sense, observe and record conditions in diverse environments.

Wireless Sensors: They are autonomous devices provided with sensing communication capabilities and limited memory and power resources.

A Graph–Intersection–Based Algorithm to Determine Maximum Lifetime Communication Topologies for Cognitive Radio Ad Hoc Networks

Natarajan Meghanathan
Jackson State University, USA

INTRODUCTION

A cognitive radio is a software-defined radio that can dynamically adapt its transmission parameters to the channels (frequencies) available for use in the operating environment (Akyildiz et. al., 2006). A cognitive radio network (CRN) is thus a network of wireless devices embedded with cognitive radios that can sense the available channels in the neighborhood and switch the communication channel, if needed. Cognitive radios have been considered as a promising solution to alleviate the spectrum scarcity problem (Zhao & Sadler, 2007). Accordingly, CRNs typically comprise of two categories of users: primary users (PUs) who own licensed channels and secondary users (SUs) who do not own any licensed channel (Akyildiz et. al., 2006). The SUs use the licensed channels of the PUs when the latter do not use them. When the PUs of the currently used channels become active, an SU relinquishes the channel and switches to any other available PU channel.

In this chapter, we target a well-known category of CRNs called the cognitive radio ad hoc networks (CRAHNs) - a self-organized ad hoc network of the licensed PU nodes and the unlicensed SU nodes (Akyildiz et. al., 2009). Depending on the activity status of the PU nodes, the set of common PU channels in the neighborhood of the SU nodes changes dynamically with time. Thus, even in a static network of SU nodes and PU nodes, communication topologies (like paths and trees) that connect the SU nodes may have to be frequently

reconfigured depending on the availability of the PU channels in the SU-SU neighborhoods. Not much work has been done on determining stable paths or trees that could exist for a longer time in a CRAHN. Specifically, to the best of our knowledge, there has been no work done to determine stable sequence of a communication topology that spans (i.e., connect all the SU nodes) the entire CRAHN network of SU nodes.

Our focus in this chapter is to develop a generic benchmarking algorithm that can be used to arrive at upper bounds for the lifetime of any communication topology that spans the entire network of SU nodes. Referred to as the Maximum Lifetime Communication Topology (MLCT) algorithm, the algorithm can be used to determine a stable sequence of instances of any communication topology that spans the entire network of SU nodes (say, shortest path tree, minimum spanning tree, connected dominating set, etc) for the lifetime of the network as long as there is a polynomial-time algorithm or heuristic to determine that topology.

BACKGROUND

Most of the work done so far in the CRAHN literature focused on developing routing solutions that are either full spectrum knowledge based or local spectrum knowledge based. The full spectrum knowledge based solutions assume each SU node to be completely aware of all the available PU channels in the network and choose optimal

routes with respect to either minimum number of hops per SU-SU path (Xin et al., 2005), maximum conflict-free assignment (Zhou et al., 2009) of PU channels or minimum number of channel switches per SU-SU path (Xin et al., 2008); there is bound to be switching of channels when none of the common available PU channels for the end nodes of an SU-SU link are the same as the preferred PU channels for one or both the end nodes to which they stay tuned by default for transmission and reception. Such full spectrum knowledge-based solutions take a centralized approach like we took in this chapter; however, the full spectrum knowledge for the current time instant alone cannot be used to arrive at benchmarks for the routing metric, if one intends to stay with a route as long as it exists.

The local spectrum knowledge based routing solutions are distributed in nature and rely only on the spectrum information gathered in the neighborhood through the common control channel (Lo, 2011). The local spectrum knowledge based routing solutions proposed so far could be classified into sub classes that target at optimizing a particular metric in a distributed fashion. The minimum power routing protocol (Pyo & Hasegawa, 2007) is designed to discover SU-SU paths that incur lower energy consumption by taking into consideration the energy loss incurred due to transmission, reception, broadcast route discovery as well as channel switching. The bandwidth footprint (BFP) minimization-based routing protocol (Shi & Hou, 2008) attempts at discovering an s - d (source-destination) path that will minimally impact the on-going s - d sessions with respect to the interference area of the SU nodes (called the bandwidth footprint). Xie et al (2010) evaluated the tradeoffs associated with farthest neighbor routing (FNR) and nearest neighbor routing (NNR) for CRAHNs; results indicate FNR to achieve better end-to-end channel utilization and reliability and NNR to be relatively more energy-efficient. Cheng et al (2007a), Cheng et al (2007b) and Ma et al (2008) attempted to develop delay-sensitive routing protocols for CRAHNs: while

Cheng et al (2007a) and Cheng et al (2007b) focus on minimizing the sum of the channel switching and access delays at the intermediate nodes, Ma et al (2008) focus on minimizing the sum of the queuing delays at the intermediate nodes.

In earlier works, we had proposed separate benchmarking algorithms based on the idea of taking graph intersections (Meghanathan, 2008) to determine stable sequence of unicast paths, multicast Steiner trees and broadcast connected dominating sets for mobile ad hoc networks (MANETs) and to determine stable sequence of data gathering trees (Meghanathan & Mumford, 2013) for wireless mobile sensor networks (WMSNs). The characteristic of both MANETs and WMSNs is that the nodes are mobile and it is the mobility of the nodes that triggers the topology changes. On the other hand, nodes in the CRAHNs considered in this research are static and it is the availability of the PU channels that changes dynamically with time, triggering changes in the communication topology of interest.

NETWORK MODEL AND ASSUMPTIONS

We assume a centralized setup of the CRAHN comprising of the licensed primary users (PU nodes) and unlicensed secondary users (SU nodes) uniform-randomly distributed. Each PU node is assumed to own a licensed channel that has a unique frequency and is identified with the ID of the PU node itself. Both the PU nodes and SU nodes are assumed to be static. Let R be the fixed transmission range for both the categories of nodes. A PU node is said to be a neighbor of an SU node if the Euclidean distance between the two nodes is less than or equal to R . Accordingly, we say that a PU channel is available for use by an SU node at a particular time instant only if the corresponding PU node is in the neighborhood of the SU node (which is always the case, as the nodes are static) and that the PU node is idle (i.e., the PU node is turned OFF) and not using its licensed

channel at that time instant. We assume a PU channel is turned ON and OFF alternatively for a random time period uniform-randomly chosen each time from the range $[0 \dots MAX_{Random_ON}]$ and $[0 \dots MAX_{Random_OFF}]$ respectively.

We model a time-invariant SU graph $G_t(SU)$ of the CRAHN as the graph comprising of only the SU nodes; a link exists between two SU nodes u and v in $G_t(SU)$ if and only if the Euclidean distance between the two SU nodes is within the transmission range, R . We sample the network periodically for every t_{sample} seconds. We model a time-variant static SU graph $G_t(SU)$ of the CRAHN as a snapshot graph of the CRAHN at time instant t : the graph comprises of both the SU nodes and PU nodes; there is no link between any two PU nodes or between a PU node and an SU node; there exists a link between two SU nodes u and v in $G_t(SU)$ if and only if there exists a link between u and v in $G_t(SU)$ as well as there exists at least one common PU node whose licensed channel is available in the neighborhood of both u and v for use (i.e., the PU node is turned OFF). We model a time-variant mobile SU graph $G_{t \dots t+k}(SU) = G_t(SU) \cap G_{t+1}(SU) \cap \dots \cap G_{t+k}(SU)$ such that $G_{t \dots t+k}(SU)$ is connected (i.e., spans all the SU nodes), wherein $G_t(SU)$, $G_{t+1}(SU)$, $G_{t+2}(SU)$, ..., $G_{t+k}(SU)$ are respectively the static SU graphs at time instants sampled at time instants t , $t+1$, $t+2$, ..., and $t+k$.

GENERIC ALGORITHM TO DETERMINE MAXIMUM LIFETIME COMMUNICATION TOPOLOGY

Given the knowledge of the future availability of the PU channels, the MLCT algorithm (pseudo code in Figure 1) works as follows: At a time instant t for which we want to determine a stable communication topology that spans the entire network of SU nodes (say a shortest path tree rooted at source SU node s), we check if the static SU graph $G_t(SU)$ is connected (i.e., there exists a

path between any two SU nodes). We can check the connectivity of an SU graph (static graph or mobile graph) using the Breadth First Search (BFS) algorithm (Cormen et al., 2009). If all the vertices of the graph can be visited by running BFS from any arbitrarily chosen node, then the graph is connected; otherwise, not. If $G_t(SU)$ is connected, the network-wide communication topology of interest exists at time instant t . We then proceed to the next sampling time instant $t+1$ and take the intersection of the graphs $G_t(SU)$ and $G_{t+1}(SU)$ and refer to the intersection graph as the mobile graph $G_{t \dots t+1}(SU) = G_t(SU) \cap G_{t+1}(SU)$. If $G_{t \dots t+1}(SU)$ is connected, we continue to proceed to the next sampling time instant $t+2$ and check whether $G_{t \dots t+2}(SU) = G_t(SU) \cap G_{t+1}(SU) \cap G_{t+2}(SU)$ is connected. We continue like this until we come across sampling time instants $t+k$ and $t+k+1$ such that $G_{t \dots t+k}(SU)$ is connected and $G_{t \dots t+k+1}(SU)$ is not connected. If $G_{t \dots t+k}(SU)$ is connected, it implies the communication topology of interest (say a shortest path tree rooted at source node s) exists in each of the static graph snapshots $G_t(SU)$, $G_{t+1}(SU)$, ..., $G_{t+k}(SU)$. We could use that communication topology across all of these time instants t , $t+1$, ..., $t+k$. We repeat the above procedure starting from time instant $t+k+1$ to find the next stable mobile graph and communication topology of the sequence and continue likewise until the end of the network session.

The MLCT algorithm finds for us a stable sequence of instances of the mobile graph and the communication topology of interest (say the shortest path tree) such that the number of transitions from one instance of the topology to another in the sequence is the global minimum. The average lifetime of the mobile graphs in the stable sequence found by the MLCT algorithm would serve as an upper bound (benchmark) for any network-wide communication topology (be it a shortest path tree, spanning tree, connected dominating set, etc) that spans all the SU nodes found by any centralized or distributed algorithm for CRAHNs.

Figure 1. Pseudo code of the maximum lifetime communication topology algorithm and its application to determine stable sequence of maximum lifetime shortest path trees (MLSPTs)

Input: Static graphs $G_t(\text{SU})$ for all the sampling time instants $t \in 1 \dots T$ of the network session
Source node s
Output: Stable sequence of mobile graphs and maximum lifetime shortest path trees (MLSPTs)
Auxiliary variables: i, j
Initialization: $i = 1, j = 1, \text{MLSPT}_{1 \dots T}(\text{SU}) = \phi; \text{Stable_G}_{1 \dots T}(\text{SU}) = \phi$
Begin MLCT Algorithm
1 **while** ($i \leq T$) **do**
2 Find a mobile graph $G_{i,j}(\text{SU}) = G_i(\text{SU}) \cap G_{i+1}(\text{SU}) \cap \dots \cap G_j(\text{SU})$ such that
 $G_{i,j}(\text{SU})$ is connected and $G_{i,j+1}(\text{SU})$ is not connected.
3 $\text{Stable_G}_{1 \dots T}(\text{SU}) = \text{Stable_G}_{1 \dots T}(\text{SU}) \cup \{G_{i,j}(\text{SU})\}$
4 Find a shortest path tree $\text{SPT}_{i,j}(\text{SU}) = \text{BFS}(G_{i,j}(\text{SU}), s)$
5 $\text{MLSPT}_{1 \dots T}(\text{SU}) = \text{MLSPT}_{1 \dots T}(\text{SU}) \cup \{\text{SPT}_{i,j}(\text{SU})\}$
6 $i = j+1$
7 **end while**
8 **return** $\text{MLSPT}_{1 \dots T}(\text{SU})$ and $\text{Stable_G}_{1 \dots T}(\text{SU})$
End MLCT Algorithm

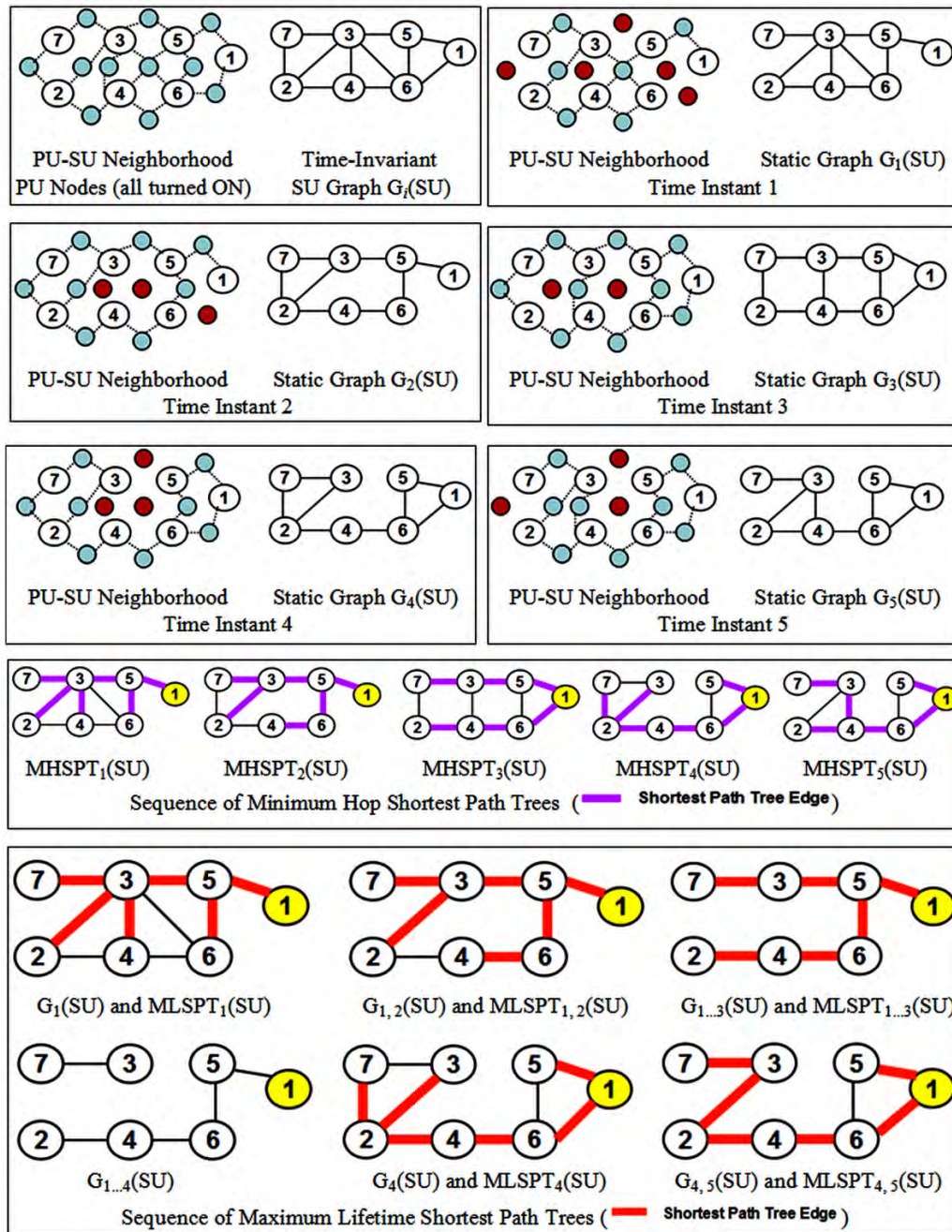
Time Complexity of the MLCT Algorithm

The time complexity of the MLCT algorithm depends on the time complexity to determine the mobile graphs and the number of times we run the BFS algorithm to determine the connectivity of the static graphs and mobile graphs as well as on the time complexity of the algorithm used to determine the underlying communication topology of interest. In general, if $\Theta(X)$ is the run-time complexity to determine the communication topology of interest on a particular SU graph, the overall-time complexity to run the MLCT algorithm to determine a stable sequence of the communication topology of interest would be: $\Theta(T*(V+E+X))$, where $\Theta(V+E)$ is the time complexity of the BFS algorithm and T is the number of instances on which we have to either run the BFS algorithm to check for connectivity of the mobile graphs or run the underlying algorithm of complexity $\Theta(X)$ for the communication topology of interest or both.

Example to Illustrate the Execution of the MLCT Algorithm

Figure 2 illustrates an informative example for the execution of the MLCT algorithm to determine a stable sequence of shortest path trees (node 1 as the source node) across five time instants 1, 2, ..., 5. The SU nodes (with white background) are numbered, while the PU nodes (blue or red-colored) are not numbered. In the static graphs at time instants 1, 2, ..., 5, an active PU node (i.e., is turned ON) is shown as a red-colored node and an idle PU node (i.e., is turned OFF) is shown as a blue-colored node. The dotted links represent the presence of a PU node in the neighborhood of the SU node(s). The time-invariant initial graph of SU nodes represents the network of SU nodes with all possible links: there exists a link between any two SU nodes in this graph if the two SU nodes are within the transmission range of each other. In the time-variant static SU graphs for time instants 1, 2, ..., 5, there exists a link between any two SU nodes if and only if there exists a

Figure 2. Example to illustrate the execution of the maximum lifetime communication topology algorithm for shortest path trees in a cognitive radio ad hoc network of SU and PU nodes



link between the two nodes in the time-invariant SU graph as well as there exists an available/ idle PU node in the neighborhood of the two SU nodes at that time instant. The minimum-hop/ minimum-height shortest path trees (MHSPTs)

incur four transitions (tree changes) with a time-averaged height of $(3+3+3+4+4)/5 = 3.4$. The MLCT-MLSPTs incur only one tree transition (at time instant 4), with a time-averaged height of $(4+4+4+5+5)/5 = 4.4$.

Proof of Correctness of the MLCT Algorithm

We now prove that the MLCT algorithm does determine a stable sequence of the mobile graphs and the corresponding instances of the communication topology of interest such that the number of transitions from one instance of the topology to another in the sequence is the global minimum (i.e., optimum). We use the approach of proof by contradiction. Let m be the number of transitions (change from one instance to another) in the sequence of instances of the communication topology determined by the MLCT algorithm. Let there be a hypothetical algorithm that determines a sequence of instances of the communication topology such that the number of transitions is m' wherein $m' < m$. If such a hypothetical algorithm exists, for m to be greater than m' , (without loss of generality) there should be a mobile graph $G_{t+p...t+s}$ (SU) for which the MLCT algorithm should have incurred at least one transition in the time span from $t+p$ to $t+s$, where t is the beginning time instant of the network session and $p < s$ are the increments to t indicating sampling time instants $t+p$ and $t+s$ respectively, but that the hypothetical algorithm did not incur any transition in the communication topology from $t+p$ to $t+s$. If the hypothetical algorithm were to find a network-wide communication topology that exists from time instants $t+p$ to $t+s$, the mobile graph $G_{t+p...t+s}$ (SU) must be connected. But, the MLCT algorithm underwent at least one transition during the time period $t+p$ to $t+s$ because the mobile graph $G_{t+p...t+s}$ (SU) was not connected. This implies the hypothetical algorithm cannot find an instance of the mobile graph and the communication topology of interest whose duration of existence is larger than that determined by the MLCT algorithm. Hence, the sequence of instances of the communication topology determined by the hypothetical algorithm should have undergone at least the same number of transitions as the sequence of instances of the communication topology determined by the MLCT algorithm, which means $m' \geq m$. This is a

contradiction to the initial hypothesis that $m' < m$ and hence the hypothesis cannot be true.

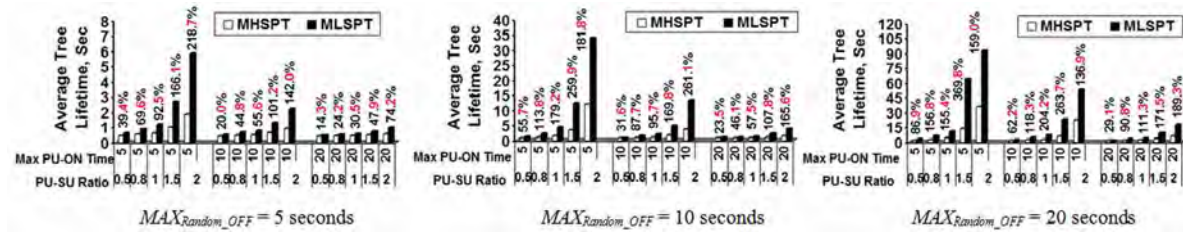
N

SIMULATIONS

We developed a discrete-event simulator in Java for Cognitive Radio Ad hoc Networks (CRAHNs) and implemented the Maximum Lifetime Communication Topology (MLCT) algorithm and adapted it to determine a sequence of maximum lifetime shortest path trees (MLSPTs) among the SU nodes. We also implemented the algorithm to determine a sequence of minimum hop shortest path trees (MHSPTs) that determines a MHSPT on a particular static SU graph and uses it in the static SU graphs for the subsequent time instants as long as the MHSPT exists. We sample the network topology for every 0.25 seconds. The total simulation time is 1000 seconds. The source node for a particular MHSPT or MLSPT is chosen randomly from the set of SU nodes. The network dimensions are 1000m x 1000m. The number of SU nodes is fixed at 50; the number of PU nodes is varied from 25 to 100 nodes. The transmission range of both the SU nodes and PU nodes is fixed to be 250m. The values of the parameters MAX_{Random_ON} and MAX_{Random_OFF} are each: 5, 10 and 20 seconds. The simulation results presented in Figure 3 and Figure 4 are averaged over 20 runs for each combination of values for PU-SU ratio, MAX_{Random_ON} and MAX_{Random_OFF} .

The MLSPTs incur significantly larger values of tree lifetime compared to the MHSPTs for all the scenarios (Figure 3). The percentage values shown in Figure 3 are based on the ratio of the lifetime of the MLSPTs to that of the MHSPTs, indicating the percentage difference in the lifetimes of the two trees. For a fixed value of the MAX_{Random_ON} and MAX_{Random_OFF} parameters, the rate at which the average lifetime of the MLSPTs increases is significantly faster than the rate at which the average lifetime of the MHSPTs increases. The MLCT algorithm makes use of the knowledge about the availability of the PU channels in future and is able

Figure 3. Lifetime of maximum lifetime shortest path trees and minimum hop shortest path trees



to find a long-living shortest path tree that exists in several of the static SU graphs spanning a mobile graph; on the other hand, the MHSPT algorithm could make use of only information about the availability of the PU channels in the particular SU graph at which it is run. The increase in the average lifetime of the MLSPTs (with increase in the PU-SU ratio) gets enhanced with increase in the availability time of the PU channels (i.e., with increase in the value of MAX_{Random_OFF}) and with increase in the number of PU channels. When the PU channels are used by their PU nodes for a longer time and remain unavailable for the SU nodes, the rate of increase for the average lifetime of the MLSPTs reduces.

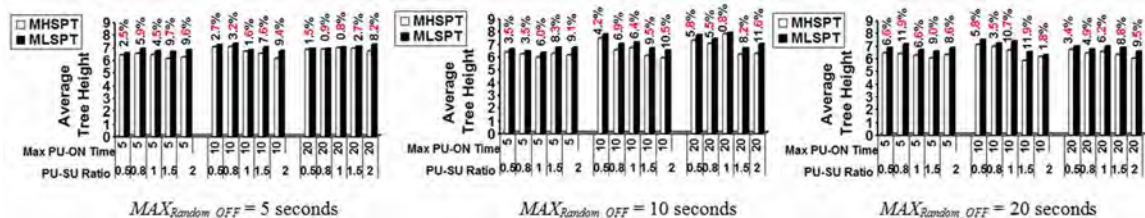
The average tree height (Figure 4) is the time-averaged height of the shortest path trees determined for the sequence of MLSPTs and MHSPTs. The tree height is a measure of the diameter of the network (the maximum of the minimum hop count of the paths between any two nodes). We observe the MLSPTs to incur at most 12% larger height than MHSPTs. There is no definite trend in the difference in the height between the two trees across the simulation conditions. With the significant gains in the lifetime of the MLSPTs

and an inconsequential increase in the height of the trees, we could infer that there is no stability-hop count tradeoff in CRAHNS.

FUTURE RESEARCH DIRECTIONS

In future, we plan to develop a distributed version of the MLCT algorithm to determine a stable sequence of network-wide communication topologies (such as shortest path trees) based on the predicted link-lifetime information (Guan et al., 2010; Li et al., 2013) of a CRAHN. We could then compare the relative performance of the distributed vs. centralized versions of the MLCT algorithms and analyze the stability-hop count tradeoff in constrained real-time settings. Besides, we plan to apply the proposed MLCT algorithm on a CRAHN of mobile SU nodes and study the impact of topology changes due to the fluctuations in the availability of the PU channels as well as due to the mobility of the SU nodes. In addition, we plan to extend the stability-hop count tradeoff studies to simulations involving energy-constrained SU nodes so that any potential tradeoff involving node lifetime (time of first node failure

Figure 4. Height of maximum lifetime shortest path trees and minimum hop shortest path trees



due to energy exhaustion) and/or network lifetime (time of network disconnection due to the failure of one or more nodes due to energy exhaustion) could be analyzed.

CONCLUSION

The high-level contribution of this chapter is the proposal of a centralized benchmarking algorithm to determine stable sequence of mobile graphs and the corresponding instances of a communication topology that spans all the secondary user (SU) nodes in a cognitive radio ad hoc network (CRAHN) environment. Referred to as the Maximum Lifetime Communication Topology (MLCT) algorithm, the algorithm is generic in nature and can be applied to determine a stable sequence of any communication topology of interest as long as there is an algorithm or heuristic to determine the topology. The MLCT algorithm uses a greedy strategy of determining a mobile graph that spans for the largest number of time instants starting from the current time instant and running an appropriate algorithm for the communication topology of interest on the mobile graph. We have shown that such a greedy strategy does indeed gives an optimal solution (benchmark) for the minimum number of transitions (and hence a maximum lifetime) among the instances of the communication topology determined for the duration of the network session.

REFERENCES

- Akyildiz, I. F., Lee, W. Y., & Chowdhury, K. R. (2009). CRAHNs: Cognitive Radio Ad hoc Networks. *Ad Hoc Networks*, 7(5), 810–836. doi:10.1016/j.adhoc.2009.01.001
- Akyildiz, I. F., Lee, W.-Y., Vuran, M. C., & Mohanty, S. (2006). Next Generation/ Dynamic Spectrum Access/Cognitive Radio Wireless Networks: A Survey. *Computer Networks*, 50(13), 2127–2159. doi:10.1016/j.comnet.2006.05.001
- Cheng, G., Liu, W., Li, Y., & Cheng, W. (2007a). Spectrum Aware On-demand Routing in Cognitive Radio Networks. *Proceedings of the 2nd International Symposium on New Frontiers in Dynamic Spectrum Access Networks* (pp. 571–574). Dublin, Ireland: IEEE. doi:10.1109/DYSPAN.2007.80
- Cheng, G., Liu, W., Li, Y., & Cheng, W. (2007b). Joint On-demand Routing and Spectrum Assignment in Cognitive Radio Networks. *Proceedings of the International Conference on Communications* (pp. 6499–6503). Glasgow, UK: IEEE. doi:10.1109/ICC.2007.1075
- Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2009). *Introduction to Algorithms* (3rd ed.). Cambridge, MA: MIT Press.
- Guan, Q., Yu, F. R., Jiang, S., & Wei, G. (2010). Prediction-Based Topology Control and Routing in Cognitive Radio Mobile Ad Hoc Networks. *IEEE Transactions on Vehicular Technology*, 59(9), 4443–4452. doi:10.1109/TVT.2010.2069105
- Li, X., Wang, D., & McNair, J. (2013). On the Accuracy of Maximum Likelihood Estimation of Primary User Behavior in Cognitive Radio Networks. *IEEE Communications Letters*, 17(5), 888–891. doi:10.1109/LCOMM.2013.031913.122829
- Lo, B. F. (2011). A Survey of Common Control Channel Design in Cognitive Radio Networks. *Physical Communication*, 4(1), 26–39. doi:10.1016/j.phycom.2010.12.004
- Ma, H., Zheng, L., Ma, X., & Iuo, Y. (2008). Spectrum Aware Routing for Multi-hop Cognitive Radio Networks with a Single Transceiver. *Proceedings of the 3rd International Conference on Cognitive Radio Oriented Wireless Networks and Communications* (pp. 1–6). Singapore: IEEE. doi:10.1109/CROWNCOM.2008.4562509
- Meghanathan, N. (2009). A Location Prediction Based Reactive Routing Protocol to Minimize the Number of Route Discoveries and Hop Count per Path in Mobile Ad hoc Networks. *The Computer Journal*, 52(4), 461–482. doi:10.1093/comjnl/bxn051

- Meghanathan, N., & Farago, A. (2008). On the Stability of Paths, Steiner Trees and Connected Dominating Sets in Mobile Ad hoc Networks. *Ad Hoc Networks*, 6(5), 744–769. doi:10.1016/j.adhoc.2007.06.005
 - Meghanathan, N., and Mumford, P. (2013). A Benchmarking Algorithm to Determine the Sequence of Stable Data Gathering Trees for Wireless Mobile Sensor Networks. *Informatica - An International Journal of Computing and Informatics*, 37(3), 315-338.
 - Perkins, C. E., & Royer, E. M. (1999). Ad hoc On-demand Distance Vector Routing. *Proceedings of the 2nd Annual International Workshop on Mobile Computing Systems and Applications* (pp. 90-100). New Orleans, LA: IEEE.
 - Pyo, C. W., & Hasegawa, M. (2007). Minimum Weight Routing based on a Common Link Control Radio for Cognitive Wireless Ad hoc Networks. *Proceedings of the International Conference on Wireless Communications and Mobile Computing* (pp. 399-404). Shanghai, China: IEEE. doi:10.1145/1280940.1281026
 - Shi, Y., & Hou, Y. (2008). A Distributed Optimization Algorithm for Multi-hop Cognitive Radio Networks. *Proceedings of the 27th International Conference on Computer Communications* (pp. 1292-1300). Phoenix, AZ: IEEE. doi:10.1109/INFOCOM.2008.186
 - Xie, M., Zhang, W., & Wong, K.-K. (2010). A Geometric Approach to Improve Spectrum Efficiency for Cognitive Relay Networks. *IEEE Transactions on Wireless Communications*, 9(1), 268–281. doi:10.1109/TWC.2010.01.090180
 - Xin, C., Ma, L., & Shen, C.-C. (2008). A Path-centric Channel Assignment Framework for Cognitive Radio Wireless Networks. *Mobile Networks and Applications*, 13(5), 463–476. doi:10.1007/s11036-008-0084-y
 - Xin, C., Xie, B., & Shen, C.-C. (2005). A Novel Layered Graph Model for Topology Formation and Routing in Dynamic Spectrum Access Networks. *Proceedings of the 1st International Symposium on New Frontiers in Dynamic Spectrum Access Networks* (pp. 308-317). Baltimore, MD: IEEE.
 - Zhao, Q., & Sadler, B. (2007). A Survey of Dynamic Spectrum Access. *IEEE Signal Processing Magazine*, 24(3), 79–89. doi:10.1109/MSP.2007.361604
 - Zhou, X., Lin, L., Wang, J., & Zhang, X. (2009). Cross-layer Routing Design in Cognitive Radio Networks by Colored Multi graph Model. *Wireless Personal Communications*, 49(1), 123–131. doi:10.1007/s11277-008-9561-7
- ### ADDITIONAL READING
- Abbagnale, A., & Cuomo, F. (2010). Connectivity-Driven Routing for Cognitive Radio Ad-Hoc Networks. *Proceedings of the 7th Annual International Conference on Sensor Mesh and Ad Hoc Communications and Networks* (pp. 1-9), Boston, MA, USA: IEEE. doi:10.1109/SECON.2010.5508269
 - Chen, T., Zhang, H., Maggio, G. M., & Chlamtac, I. (2007). Topology Management in CogMesh: A Cluster-Based Cognitive Radio Mesh Network. *Proceedings of the International Conference on Communications* (pp. 6516-6521), Glasgow, Scotland, UK: IEEE. doi:10.1109/ICC.2007.1078
 - Chowdhury, K. R., & Akyildiz, I. F. (2011). CRP: A Routing Protocol for Cognitive Radio Ad Hoc Networks. *IEEE Journal on Selected Areas in Communications*, 29(4), 794–804. doi:10.1109/JSAC.2011.110411
 - Liang, H., Lou, T., Tan, H., Wang, Y., & Yu, D. (2015). On the Complexity of Connectivity in Cognitive Radio Networks through Spectrum Assignment. *Journal of Combinatorial Optimization*, 29(2), 472–487. doi:10.1007/s10878-013-9605-0

Liu, S., Wang, Y., & Cui, C. (2014). Connectivity-based Performance Evaluation for Mobile Cognitive Radio Network. *Journal of Computers*, 9(11), 2635–2641. doi:10.4304/jcp.9.11.2635-2641

Lu, D., Huang, X., Li, P., & Fan, J. (2012). Connectivity of Large-Scale Cognitive Radio Ad Hoc Networks. *Proceedings of the International Conference on Computer Communications* (pp. 1260–1268), Orlando, FL, USA: IEEE. doi:10.1109/INFCOM.2012.6195487

Salim, S., & Moh, S. (2013). On-Demand Routing Protocols for Cognitive Radio Ad hoc Networks. *EURASIP Journal on Wireless Communications and Networking*, (102): 1–10.

Shin, D., Kim, J., & Ko, Y.-B. (2014). A Hybrid Topology-based Multicast Routing for Cognitive Radio Ad hoc Networks. *Proceedings of the International Conference on Computing, Communication and Networking Technologies* (pp. 1–5), Hefei, China: IEEE. doi:10.1109/ICCCNT.2014.6963026

Singh, K. D., Rawat, P., & Bonnin, J.-M. (2014). Cognitive Radio for Vehicular Ad hoc Networks (CR-VANETs): Approaches and Challenges. *EURASIP Journal on Wireless Communications and Networking*, (49): 1–22.

Thomas, R. W., Komali, R. S., MacKenzie, A. B., & DaSilva, L. A. (2007). Joint Power and Channel Minimization in Topology Control: A Cognitive Network Approach. *Proceedings of the International Conference on Communications* (pp. 6538–6543), Glasgow, Scotland, UK: IEEE. doi:10.1109/ICC.2007.1082

KEY TERMS AND DEFINITIONS

Ad hoc Network: A decentralized and self-organized network of wireless nodes without any a priori dependence on any infrastructure.

Algorithm: A sequence of steps to perform a particular task.

Cognitive Radio: A software-defined radio that can dynamically adapt its transmission parameters to the channels (frequencies) available for use in the operating environment.

Communication Topology: An arrangement of the nodes in the network that can be used for communication between any two nodes in the network.

Mobile Graph: A sequence of static graphs generated for a time period.

Primary User (PU): Licensed user of a cognitive radio network.

Secondary User (SU): Unlicensed user of a cognitive radio network who uses one or more available channels of a PU.

Shortest Path Tree: A communication topology that is rooted at a specific node and connects every other node in the network on the shortest path (minimum number of hops) to the root node.

Static Graph: Snapshot of a network of PUs and SUs at any particular time instant.

Topology Lifetime: The duration of existence of a communication topology.

Improving Quality of Business in Next Generation Telecom Networks

Vesna Radonjić Đogatović
University of Belgrade, Serbia

INTRODUCTION

Business processes in telecommunication sector have evolved from rigid structures to a highly competitive environment due to the dynamics of the open market. Future telecommunications infrastructure are expected to be built upon the concept of next generation network (NGN) referring to an architecture of telecommunication core and access networks, which assumes transport of all information and services over a common network, typically built around the Internet Protocol (IP).

Many service providers (SPs) are looking to NGN services as a means to attract and/or retain the most gainful users. In NGN users are expected to choose the SP offering the best price and quality of service (QoS) combination. As a result, SPs operating in the same telecommunication market will end up competing for users by adjusting QoS they offer and the price they charge for their services.

Quality of NGN services can be estimated based on the three aspects: QoS, quality of experience (QoE) and quality of business (QoBiz). The main goal for SP is to maximize his revenue while providing users the required QoS at the acceptable price. QoS can be defined as a set of characteristics of a telecommunications service that bear on its ability to satisfy users' requirements (ITU-T E.800, 2008). While QoS is related to the service performances that can be measured and controlled, QoE relates to the experience realized by a user when using the service and it depends upon users' actions and subjective opinions. In addition to SPs' aspirations to ensure the required QoS to their users, profitability is of the most impor-

tance to them. QoBiz in particular covers service providers' profitability. It deals with the financial aspects of service provisioning and refers to all those parameters that are expressed in monetary units. QoS, QoE and QoBiz are integral parts of a service level agreement (SLA) which can be contracted between two SPs or a SP and a user.

This chapter aims to provide new possibilities for SPs to enhance their revenues using the appropriate pricing scheme. Features and applicability of responsive pricing scheme for charging end users in NGN are discussed. Game theory is used as an underlying concept for the implementation of pricing.

BACKGROUND

Pricing with QoS guarantees has gained a strong momentum in telecommunication networks in past decade. It has led to a new interdisciplinary research area of "Telecommunication Economics", which investigates telecommunication networks from an economical rather than from a technical perspective and allows innovative solutions in network management, control and pricing (Courcoubetis & Weber, 2003).

QoS describes the ability of a network to provide a service with an assured service level but it appears that in NGN QoS differentiation will not provide a suitable economic framework for the trade-off between quality delivered by SPs and willingness to pay from users' side. QoE is an alternative framework for pricing service quality according to the user perception (ITU-T G.1011, 2013). It is affected not only by technical (i.e. QoS)

aspects, but on non-technical aspects of service too, such as service set-up, content, price and customer support, which are essential for both QoE and QoBiz evaluation. The distinction between QoS, QoE and QoBiz metrics in case of Internet services are first emphasized by Moorsel (2001). In Wolter and Moorsel (2001) possible relationships between QoS and QoBiz metrics are pointed out, particularly the effects of QoS degradations on the profitability of e-services. This research considers dynamic relationship between QoS and QoBiz metrics with pricing as an important tool for balancing users' behaviour. Further, for the purpose of evaluating the impact of service composition and utility computing on QoBiz aspects of SPs, a distributed architecture of control systems that manages SLAs is proposed in Machiraju et al. (2002). Yu and Bouguettaya (2007) proposed the model for supporting optimized access of web services through service-oriented queries. Besides QoBiz parameters, this model captures functionality and users' behaviour, as the key features of web services. Bjekovic and Kubicki (2011) address the need for an integration of non-functional aspects of service from the business perspective with aim of improving cooperation among business enterprises. In Radonjic Djogatovic et al. (2013) the model performing transparent mapping from QoS to QoBiz parameters is proposed. This QoBiz model is based on several QoS parameters that significantly affect both users' demand and a SP's requirements from the business perspective.

Telecom provider's QoBiz is tightly related to the selection of the appropriate network pricing method. Price based bandwidth allocation has been the focus of many research efforts that aimed to guarantee an appropriate QoS to users. Some methods allow congestion which results in blocking of the lower-priced traffic classes and the acceptance of the higher-priced traffic classes in case of congestion (Marbach, 2004). The concept of responsive pricing was proposed with the aim to incorporate a feedback generated by the network. When the network announces a price based on the cost of using network resources

price-sensitive users adjust their traffic in accordance of their own network service valuation (MacKie-Mason, Murphy, L. & Murphy, J., 1997). Chod and Rudi (2005) considered responsive pricing with resource flexibility as well as effects of demands variability and correlation, assuming normally distributed demand curve. In Ninan and Devetsikiotis (2005) a model for incorporating pricing in NGNs with users sharing bandwidth under a fixed charge per bandwidth amount was presented. Optimal resource allocation of NGN services under a flat pricing scheme and QoS policies were considered by Kallitsis et al. (2007). Congestion pricing with various user demands over time was analyzed by Hande et al. (2010). With the goal of maximization SPs' revenue the same authors applied the optimal combination of flat-rate and usage-based access price components. The importance of selection of the appropriate pricing method for a business model framework in case of providing assured quality services is emphasized in Ghezzi et al. (2014). An overview of usage-based pricing schemes and accounting protocols that can be used in NGN are given in Radonjic Djogatovic and Kostic-Ljubisavljevic (2015).

PRICING AND QUALITY ISSUES IN NGN

The evolution of networks to NGNs must allow the continuity of, and interoperability with, existing networks while enabling the implementation of new capabilities (ITU-T Y.2012, 2010). Suitable pricing and business models need to be designed which is expected to provide the appropriate incentives for both SPs and users.

Generally, the concept of pricing implies the process of determining prices which should be based on the appropriate pricing model and controlled by a pricing policy. It is expected that competition will force SPs to rapidly create and deploy different pricing concepts with aim of fulfilment a trade-off between providing satisfying

user's utility and provider's profit, still preserving implementation efficiency and feasibility. User's utility can be expressed as a function of available network resource offered to a user which indicates a user's sensitivity to changes in QoS. A wide range of different pricing schemes is likely to be applied in NGN (Ninan & Devetsikiotis, 2005; Pandey, Ghosal & Mukherjee, 2007; Radonjic & Acimovic-Raspopovic, 2010). Pricing schemes that should be implemented in NGN have to be defined and evaluated with respect to the heterogeneous technical, economic and social aspects. The main evaluation criteria encompass efficiency in the context of maximizing utilities of users and providers, fairness and feasibility.

In NGN there is a need for shifting from static pricing schemes, in which users are charged independently of the resource consumption and QoS delivered, towards dynamic pricing schemes. In dynamic pricing, price is determined as a cost per unit of consumption and according to the level of QoS guarantees provided for the observed service class. The main problem with dynamic pricing refers to the need of intensive monitoring of network resources in order to dynamically adjust per-class prices to resources usage and the QoS provided for each service class. Any NGN architecture must guarantee fair access to the shared resources in the access network and control load distribution with aim of avoiding focused overload in the core. Also, NGN should support hard guarantees to users and pricing different QoS classes.

Due to the rapidly increasing deployment of interactive and multimedia applications in telecommunication services, QoS becomes an integral part of various protocols and mechanisms for enabling computing and telecommunication systems. QoS is defined by SLA which consists of two parts: the technical part and the administrative part. The technical part encompasses a set of descriptors and associated attributes that describe the particular service class and the traffic profile. The administrative part covers financial and legal aspects: information about pricing, charging, billing and payment, penalties for both user and

service provider in the case of contract violation, etc. (Stojanovic, Kostic-Ljubisavljevic & Radonjic Djogatovic, 2013). As users' needs are constantly increasing, as well as competition between service providers, SLA becomes more complex.

Although QoS is usually represented by delay, loss and jitter, which are difficult to measure precisely, frequently used QoS parameters for determination whether a required service level is being achieved, are network availability and bandwidth. Network availability includes the availability of many items the network consists of, e.g. multiple physical connections, networking device redundancy. It is also important to distinguish available and guaranteed bandwidth. In many dynamic pricing schemes users are allowed to compete for the available bandwidth and their obtained bandwidth depends upon amount of traffic from other users in the network at any observed time (Chod & Rudi, 2005). The term guaranteed bandwidth implies a guaranteed minimum bandwidth SP provides or burst bandwidth in SLA. The service with guaranteed bandwidth has higher priority and is priced higher compared to the available bandwidth service.

By introducing QoS differentiation users are encouraged to choose the service that meets their needs in the most adequate manner, which can be achieved through pricing. Providing service with strong QoS guarantees keeps users satisfied and thereby maintains the confidence in the SP. However, it is necessary to take into account user experience and business indicators too. Therefore, quality of NGN services can be estimated based on three aspects: QoS, QoE and QoBiz.

Obviously, QoS is related to the service performances that can be measured and controlled, while QoE relates to the experience realized by a user when using the service. QoE depends upon users' actions and subjective opinions. It takes into consideration users' satisfaction with the service, subjective evaluation, the degree of their expectations fulfilment and in what context they use it or intend to (Barakovic, S., Barakovic, J. & Bajric, 2010).

In addition to SPs' aspirations to ensure the required QoS to their users, QoBiz is an important aspect in providing quality of NGN services that is being increasingly used in the pursuit of better business. It deals with the financial aspect of service provisioning and refers specifically to measures such as service price, service provisioning costs, revenue from the service provisioning, revenue per transaction, lost transactions etc. In general, QoBiz parameters are all those parameters that are expressed in monetary units. According to the more precise interpretation, it is a monetary value that matches the quality of delivered service, expressed through connection of QoS parameters with monetary value within the SLA (Radonjic Djogatovic, Kostic-Ljubisavljevic & Stojanovic, 2013).

A MODEL FOR QUALITY OF BUSINESS IMPROVEMENT

A model for QoBiz improvement of next generation telecom network should include business aspects that are specific and important for telecom sector. It should reflect involved parties' preferences which allow discovering, negotiating and contracting a service between a user and a provider.

This chapter proposes an approach for eliminating or at least alleviating congestion along with service price optimization regarding QoS and QoE and revenue maximization, which is reflected on QoBiz. It has been done via proper bandwidth assignment considering the capacity constraint, i.e. adapting the network price to optimize network performances. It is achieved by a usage-based scheme with users getting charged for the amount of traffic they consume. For maintaining social optimality these charges would have to be set equal to the marginal cost of usage. Since bandwidth scarcity occurs only during congestion, this marginal cost is essentially the same as the congestion cost.

In order to achieve satisfying network and economic efficiency the responsive pricing scheme is

used (Chod & Rudi, 2005). Price is emphasized as an alternative means for congestion control to ensure proper network operation and in particular to guarantee different service levels. In this pricing scheme, in the case of high network utilization, provider increases the prices for the resources and adaptive users then reduce traffic offered to the network. Similarly, in the case of low network utilization, by decreasing the price adaptive users are encouraged to increase their offered traffic.

In the proposed model bandwidth management server (BMS) is introduced in order to enhance scalability and to allow connection control functions independently of the underlying network, which makes it suitable for NGN. BMS acts as an interface between the connection control functions and the network specific bearer functions. The interface between BMS and the underlying network elements is provided to allow BMS to set up and tear down aggregate bandwidth reservations across the network. The most important features of BMS are the ability to release bandwidth when it is no longer needed and to reserve bandwidth to the particular destination of the particular QoS class. The underlying network elements must be able to inform BMS of any changes affecting current reservations. The interface should allow this information to be passed upwards to BMS immediately such an event takes place and the information should reach BMS with the shortest possible delay.

During SLA negotiation, the SP estimates resource usage according to the required traffic profile. Although it is common that the agreement between the SP and a user for a service consumption has to be settled for a long time period, in this model allocations are performed in short time scales. In practice, it would be impossible for each user to update his bandwidth in a short time scale. Instead of that, with the aim of maximizing SP's revenue, updates are performed by BMS considering network congestion and network capacity utilization. Updates are based on users' parameters of desired bandwidth, maximal price they are willing to pay and their individual utility

functions. During the service provisioning interval, the usage-based traffic monitoring should be applied only with respect to deviations from the contracted traffic profile. Such approach is called hybrid pricing scheme. It needs less storage of monitoring data in comparison with monitoring per unit of consumption (i.e. dynamic pricing) and should be implemented only at the edge routers, in which traffic conditioning is performed. The solution of the problem encompasses the optimal bandwidth allocation and the optimal price for that allocation. The actual price is calculated at the end of service provisioning interval according to SP's specific policy, which should aim to force users to wisely select the most appropriate traffic profiles and to properly adjust to them. Each user has to pay the price that is equal to the sum of products of each optimal price and his bandwidth consumption during all his sessions in the agreed time period.

In the proposed model, problem of determining service prices is divided into optimization with respect to users' utility and the optimization of the provider's revenue. In this fashion, the model incorporates not only QoS but QoE and QoBiz too. Users' preferences may be modelled with utility functions that describe users' sensitivity to changes of QoS. Since it is not easy to predict actual QoS parameters, such as delay or packet losses in most real networks, utility is often expressed as a function of the amount of available network resources (i.e. bandwidth) offered to each user (Ninan & Devetsikiotis, 2005). Thus, the utility indirectly indicates user's sensitivity to changes in QoS.

The optimization procedure is performed for different categories of users, which can be alternatively done for different service classes. According to the elasticity criteria, users are classified into three categories: inelastic, partially elastic and elastic users (Radonjic Djogatovic & Kostic-Ljubisavljevic, 2015). Inelastic users are users who have strict requirements in terms of delay but can tolerate losses to some extent and their utility has been most commonly described

by the sigmoid function. A utility function which best models elastic and partially elastic users' behaviour is a generalization of the logarithmic function. Elastic users do not tolerate losses but can accept a delay to some extent. Partially elastic users are also not tolerant of losses but they have stronger requirements in respect to delay. Users' utility functions vary in accordance to the elasticity criterion of a user. For all users' types, QoS is defined by bandwidth, θ obtained from the SP. Users are assumed to be non-cooperative, meaning that they refuse to reveal their utility functions to one another in the fear of being misused.

The SP employs a usage-based pricing policy by charging M per unit bandwidth consumed. Depending upon the QoS requested, i.e. the class of service he chose, each user would require a minimum bandwidth γ . Fewer bandwidth than γ on average is of no utility to the user (Ninan & Devetsikiotis, 2005). The law of diminishing marginal utility ensures that the user derives the same amount of satisfaction from any bandwidth more than the maximum π . It is considered that the user is willing to pay a maximum price m per unit of bandwidth. When the price M equals the maximal price m , the user will desire only the minimum acceptable bandwidth, γ . For all types of users, any price beyond the maximal price reduces a user's desired bandwidth to zero. Each user will try to choose his bandwidth θ so as to maximize his net benefit (i.e., utility minus cost). Ideally any resource allocation between users should ensure that the total user utility is maximized. On the other hand, the SP's utility T depends on the total revenue generated and hence is a function of the market price and the bandwidth allocated to the various users. In order to maximize his revenue, the SP settles the price per bandwidth unit according to the concept of responsive pricing reflecting the state of resource utilization.

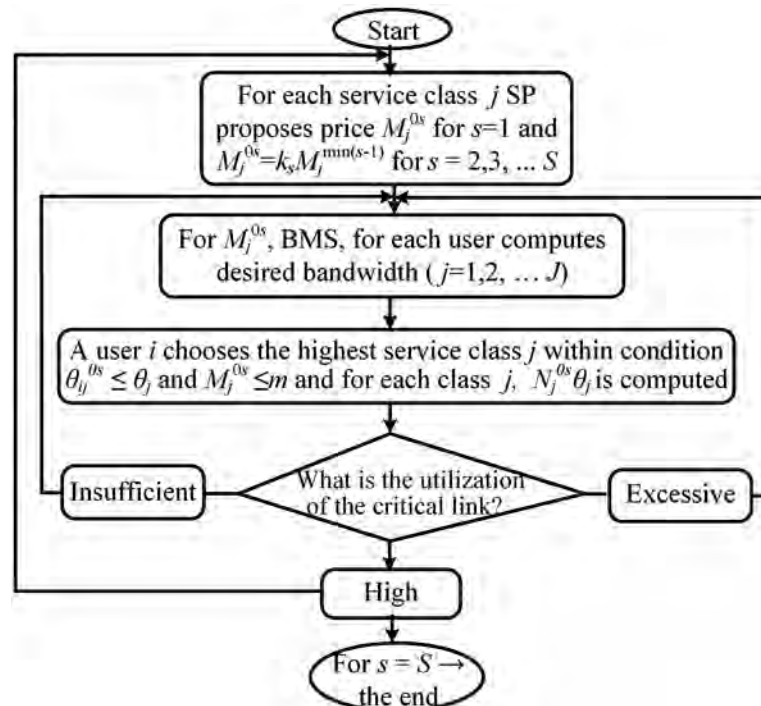
The interaction of the SP and users is defined as a Stackelberg leader-follower game within responsive pricing scheme driving the system towards optimality. Generally, in Stackelberg model at least one player is defined as the leader who

can make the decision and commit the strategy before other players who are defined as followers (Briest, Hoefer & Krysta, 2008). In the proposed model, players are the SP, acting as the leader and users, acting as followers. They act in a definite sequence and interaction between them is dynamic rather than static. The strategy chosen by the SP can be observed by users, so they can adapt their decisions in accordance to their preferences. The SP can choose such a strategy which allows him to maximize revenues, assuming that users will choose their best responses. With backward induction, the best responses of users are obtained first. The best responses of both SP and users in this leader-follower competition can be obtained in the following way. Given the price offered by the SP, based on a demand function, the amount of bandwidth users requested can be determined. For example, if the price is high, the amount of bandwidth requested by users will be small and vice versa. Then, this best response is used to compute the revenue of the SP, and he chooses a strategy that leads to his revenue maximization.

In the proposed usage-based pricing algorithm (Figure 1), prices optimization is performed for a single critical resource link in a SP's network. Such a link usually does not behave badly all the time, but only under certain worst-case conditions. On the observed link, the total number of users is N . For each user, bandwidth and price parameters, representing QoS and QoE respectively, are parts of the SLA. Each user gives his parameters of desired bandwidth and maximal price that can't be changed during the agreement period. The bandwidth allotment to user i is denoted as θ_i^* and that is his best response to the price SP determined. The algorithm can be performed for scenarios with different number of service classes in order to optimize bandwidth usage, price per bandwidth unit and a total capacity of the critical link in the network.

There are S rounds in total and in each round s , l_s iterations are performed, where $s = 1, 2, \dots, S$. Each round s consists of several iterative steps. The algorithm, shown in Figure 1, converges quickly after few tens of rounds. Optimization is

Figure 1. Pricing algorithm



performed for each service class j . The algorithm begins when the provider assigns an initial price, M^0 for each service class j . The initial price can be set either randomly or based on historical data. BMS simulates users' behavior in a manner that for each user requiring a certain service class, calculates the desired bandwidth and forwards the data to the provider. The provider then checks whether all user requirements can be met with the existing capacity and if the answer is yes and if the facilities are not used enough, it proceeds to a new iteration with lower prices. In case that requests exceed the network capacity, next iteration is performed with higher prices. In this fashion, users' responses are adapted to the state of network capacity utilization. A round ends when satisfactory utilization of the network capacity is achieved (over 90%). In each round different values of the network capacity are adopted as well as different values of the initial price for each service class. The price offered by the SP, which maximizes his revenue along with the best response bandwidth of each user constitutes the Stackelberg equilibrium (Radonjic & Acimovic-Raspopovic, 2010). In this manner, revenue, as a significant aspect of QoBiz can be improved.

FUTURE RESEARCH DIRECTIONS

In order to provide the appropriate incentives for providers, suitable business and pricing models need to be designed. QoBiz stands out as a significant concept which deals with the financial aspects of service provisioning. For QoBiz evaluation, it is essential to consider service demand, price structures and revenue trends. The need for new pricing solutions that coordinate the competing requirements of providers and users in NGN environment is evident.

In future research, new possibilities for solving similar problems of pricing and interconnection between multiple providers in NGN should be considered. A general research direction should comprise market modelling, cost and risk consid-

eration as well as network design. The proposed algorithm can be extended to cover contemporary real market situation in which several service providers compete for users.

Different types of providers should be considered along with quantifying their impact on the price setting and overall QoBiz. More specific research direction could include applying different game models (e.g. repeated/cooperative games, Cournot and Bertrand games) under dynamic and hybrid pricing schemes. Furthermore, future research should be directed towards more precise mapping between QoE and QoBiz parameters.

CONCLUSION

The adoption of an appropriate pricing scheme can significantly affect business operation of a NGN provider. NGNs must be flexible enough to enable the use of different pricing and business models. Growth in demand due to the popularity of NGN applications requires significant investment in infrastructure but brings a negligible return.

Service providers, therefore, strive for a new network model giving priority to managed services whose usage is controlled by a particular pricing scheme. They are searching for ways to brand and bundle new services, achieve operational cost reductions and strategically position themselves in relation to their competition. Many SPs are looking to NGN services as a means to attract and/or retain the most gainful users. In NGN users are expected to choose the SP offering the best price and QoS combination. As a result, SPs operating in the same telecommunication market will end up competing for users by adjusting QoS they offer and the price they charge for their services.

With the aim of telecom provider's QoBiz improvement, in this chapter the model for the service price optimization is proposed. It has been shown that applying responsive pricing scheme along with Stackelberg game, leads to improvement of the SP's revenue, as a significant aspect of QoBiz. The target market for the application

of the proposed model are business users, such as small and medium enterprises. The introduction of the proposed model should be done gradually to the coexistence with the current pricing models.

REFERENCES

- Barakovic, S., Barakovic, J., & Bajric, H. (2010). QoE Dimensions and QoE Measurement of NGN Services. *Proceedings of 19th Telecommunications Forum* (pp. 15-18). Belgrade Telecommunications Society.
- Bjekovic, M., & Kubicki, S. (2011). Service quality description – a business perspective. *Proceedings of the Federated Conference on Computer Science and Information Systems* (pp. 513-520). IEEE.
- Briest, P., Hoefer, M., & Krysta, P. (2008). Stackelberg Network Pricing Games. *Proceedings of 25th International Symposium on Theoretical Aspects of Computer Science* (vol. 1, pp. 133-142). Springer-Verlag.
- Chod, J., & Rudi, N. (2005). Resource Flexibility with Responsive Pricing. *Operations Research*, 53(3), 532–548. doi:10.1287/opre.1040.0191
- Courcoubetis, C., & Weber, R. (2003). *Pricing Communication Networks*. Chichester, UK: John Wiley & Sons Ltd. doi:10.1002/0470867175
- Ghezzi, A., Dramitinos, M., Rangone, A., Agiatzidou, E., Johanses, F. T., Balocco, R., & Løsethagen, H. (2014). Internet Interconnection Techno-Economics: A Proposal for Assured Quality Services and Business Models. *Proceedings of 47th Hawaii International Conference on System Science* (pp. 708-717). IEEE. doi:10.1109/HICSS.2014.95
- Hande, P., Chiang, M., Calderbank, R., & Zhang, J. (2010). *Pricing under Constraints in Access Networks: Revenue Maximization and Congestion Management*. *Proceedings of IEEE INFOCOM 2010* (pp. 1–9). San Diego, CA: IEEE.
- ITU-T Recommendation E.800. (2008). *Definitions of terms related to quality of service*. ITU-T.
- ITU-T Recommendation G.1011. (2013). *Reference Guide to Quality of Experience Assessment Methodologies*. ITU-T.
- ITU-T Recommendation Y.2012. (2010). *Functional requirements and architecture of next generation networks*. ITU-T.
- Kallitsis, M., Michailidis, G., & Devetsikiotis, M. (2007). Pricing and Measurement-based Optimal Resource Allocation in Next Generation Network Services. *Proceedings of the First IEEE Workshop on Enabling the Future Service-Oriented Internet* (pp. 1-6). IEEE. doi:10.1109/GLOCOMW.2007.4437788
- Machiraju, V., Rolia, J., & Moorsel, A. (2002). *Quality of Business Driven Service Composition and Utility Computing*. Software Technology Laboratory, HP Laboratories.
- MacKie-Mason, J., Murphy, L., & Murphy, J. (1997). Responsive Pricing in the Internet. In L. W. McKnight & J. P. Bailey (Eds.), *Internet Economics* (pp. 279–303). MIT Press.
- Marbach, P. (2004). Analysis of a static pricing scheme for priority services. *IEEE/ACM Transactions on Networking*, 12(2), 312–325. doi:10.1109/TNET.2004.826275
- Moorsel, A. (2001). Metrics for the Internet Age: Quality of Experience and Quality of Business. *Proceedings of the Fifth International Workshop on Performability Modeling of Computer and Communication Systems* (vol. 34, pp. 26-31). University of Erlangen.
- Ninan, B. M., & Devetsikiotis, M. (2005). Game-Theoretic Resource Pricing For The Next Generation Internet. In A. Girard, B. Sanso, & F. Vazquez Abad (Eds.), *Performance Evaluation and Planning Methods for the Next Generation Internet* (pp. 141–163). Springer. doi:10.1007/0-387-25551-6_6

Pandey, V., Ghosal, D., & Mukherjee, B. (2007). Pricing-based approaches in the design of next-generation wireless networks: A review and a unified proposal. *IEEE Communications Surveys and Tutorials*, 9(2), 88–101. doi:10.1109/COMST.2007.382409

Radonjic, V., & Acimovic-Raspopovic, V. (2010). Responsive Pricing Modeled with Stackelberg Game for Next Generation Networks. *Annales des Télécommunications*, 65(7-8), 461–476. doi:10.1007/s12243-010-0174-2

Radonjic Djogatovic, V., & Kostic-Ljubisavljevic, A. (2015). Telecommunications Pricing Fundamentals. Belgrade, Serbia: University of Belgrade – Faculty of Transport and Traffic Engineering.

Radonjic Djogatovic, V., Kostic-Ljubisavljevic, A., & Stojanovic, M. (2013). Quality of Business Considerations in Telecommunication Networks. *Proceedings of the 7th International Quality Conference* (pp. 201–205). University of Kragujevac.

Stojanovic, M., Kostic-Ljubisavljevic, A., & Radonjic Djogatovic, V. (2013). SLA-controlled Interconnection Charging in Next Generation Networks. *Computer Networks*, 57(11), 2374–2394. doi:10.1016/j.comnet.2013.04.013

Wolter, K. & Moorsel, A. (2001). *The Relationship between Quality of Service and Business Metrics: Monitoring, Notification and Optimization*. HP Labs Technical Report, HPL-2001-96.

Yu, Q. & Bouguettaya, A. (2007). Framework for Web Service Query Algebra and Optimization. *ACM Transactions on the Web*, 1-34.

ADDITIONAL READING

Collins, H. (2009). *Next Generation Networks Creating a Dedicated Cost Model*. Cheltenham, UK: InterConnect Communications Ltd.

ITU-T Recommendation Y.2233 (2010). Requirements and framework allowing accounting and charging capabilities in NGN.

Sharif, S. M., & Ahmed, A. (2015). Next Generation Networks (NGN) Billing and Charging System. *International Journal of Engineering and Computer Science*, 4(1), 9825–9830.

Stankiewicz, R., Cholda, P., & Jajszczyk, A. (2011). QoX: What is It Really? *IEEE Communications Magazine*, 49(4), 148–158. doi:10.1109/MCOM.2011.5741159

Stojanovic, M., Vukasinovic, M., & Radonjic Djogatovic, V. (2015). Approaches to Quality of Experience Management in the Future Internet, *Proceedings of the 12th International Conference TELSIKS*, Nis, Serbia (pp. 281–288). IEEE. doi:10.1109/TELSKS.2015.7357788

Webb, J. N. (2007). *Game Theory - Decisions, Interaction and Evolution*. London, UK: Springer.

KEY TERMS AND DEFINITIONS

Dynamic Pricing: The process of allocating the price as a cost per unit of resource consumption and according to level of QoS guarantees provided for the particular service class.

Hybrid Pricing: The process of applying the static price in regular network operation mode while during congestion the dynamic pricing is enforced allowing deviations from contracted traffic profile.

Next Generation Network (NGN): An architectural concept of future telecommunication core and access networks, which assumes transport of all information and services over a common network, typically built around the IP.

Quality of Business (QoBiz): A monetary value that matches the quality of delivered service, expressed through connection of QoS parameters with monetary value within the SLA.

Quality of Service (QoS): A set of service requirements that a network should meet when transferring traffic flows.

Responsive Pricing: A pricing scheme which incorporates feedback generated by a network assuming that users are adaptive to the service price changes.

Service Level Agreement (SLA): A contract between service providers or a service provider and a user, which defines responsibilities of all

contracting parties, QoS guarantees, performance metrics, measurement methods and pricing principles.

Service Provider (SP): A general reference to an enterprise that provides telecommunication services to users.

Stackelberg Game: A strategic game in which at least one player is defined as the leader who can make a decision and commit a strategy before other players who are defined as followers.

N

Information–Centric Networking

Mohamed Fazil Mohamed Firdhous

University of Moratuwa, Sri Lanka

INTRODUCTION

Over the years, the Internet not only has increased in size in terms of the number of devices, users and applications but also in the type of usage. The original Internet was designed for remote host access along with transferring emails from one host to the other (Almagor, 2011). These applications made the Internet architecture predominantly to be host centric where each and every accessible device on the Internet was assigned a globally unique IP address. The invention of world wide web along with the connected invention of the multimedia browser in the 1990s started pushing the Internet towards an information store. With this transformation, users started searching for the device, where the required information is stored and downloading that information onto his or her computer for viewing. This is commonly known as browsing and still based on the host-centric architecture of the Internet.

Over the years, the information dissemination has shifted from the host centric browsing paradigm to a data (content) centric paradigm, where the information is given the prominence over the device that holds it (Hassan, Elbreiki, Firdhous & Habbal, 2015). This new paradigm of information dissemination has led to the popularity of one-to-many and many-to-many distribution and retrieval of information to increase over the host-to-host communication in the Internet (Katsaros, Chai, Wang, Pavlou, Bontius & Paolone, 2014). Presently users are more interested in retrieving information faster from anywhere on demand rather than knowing where it was originally stored or currently retrieved from. This shift has prompted the storing of information in intermediate nodes with better support for mobility.

DOI: 10.4018/978-1-5225-2255-3.ch569

For supporting this paradigm shift in information retrieval, many incremental solutions in terms of overlay networks have been proposed and implemented over time (Pandey, Garg & Gore, 2012). Though these solutions could meet the requirement of the time, they did not alter the host centric nature of the Internet architecture. Hence they could not overcome the inherent shortcomings of the host centric architecture resulting in issues with scalability, security, mobility and manageability. The continuous minor modifications in terms of patches and incremental solutions has made the Internet architecture more complex and vulnerable leading to a less-flexible architecture with limited manageability. All these has led to the demand for a design for totally new clean slate architecture that can meet the current as well as the future requirements of an information intensive world. In this regard, many independent projects have been initiated by researchers throughout the world. Though the names and finer implementation details of many of these projects differ from each other, many of the important attributes are either common or shared including assumptions, objectives and architectural properties (Hassan et al, 2015; Ahlgren, Dannewitz, Imbrenda, Kutscher & Ohlman, 2012). Paul, Pan and Jain (2011) have discussed many of the projects in detail highlighting the salient features of them along with the experimentations carried out for validating the claims put forward by the researchers.

In this chapter, the author takes a critical look at emerging networking paradigm of Information Centric Networking (ICN) that has been proposed for overcoming the shortcomings of the current Internet architecture. The main focus of the chapter will be features of the proposed ICN architecture with special emphasis on naming,

addressing and routing along with the challenges and requirements that need to be addressed for the development of ICN.

BACKGROUND

The information-centric networking (ICN) is a term coined for identifying the multiple projects carried out by several researchers all over the world for coming up with the future safe Internet architecture while serving the current demand for seamless distribution of information on demand. The architecture of the proposed networking paradigm is to overcome the shortcomings of the present day Internet including security of information and hosts, support for mobility multicast delivery of information, scalability and quality of service guarantees demanded by different users (Ghodsi, Shenker, Koponen, Singla, Raghavan & Wilcox, 2011). The proposed architecture leverages in-network caching, i.e., replication of information in networking devices such as routers, multiparty distribution of information through replication and user interaction models decoupling senders and receivers ultimately achieving a networking architecture that is more resilient to disruptions, failures flash-crowd effects happening in the networks. The some of the prominent projects carried out with the above mentioned objectives include Data-Oriented Network Architecture (DONA), Publish-Subscribe Internet Routing Paradigm (PSIRP) currently known as the Publish-Subscribe Internet Technology (PURSUIT), Network of Information (NetInf) designed as part of the Design for the Future Internet (4WARD) project and currently known as the Scalable and Adaptable Internet Solutions (SAIL), Content-Centric Networking (CCN) currently absorbed into the Named Data Networking (NDN) project, Content Mediator Architecture for Content-aware Networks (COMET) and Delay Tolerant Networking (DTN) (Koponen, Chawla, Chun, Ermolinskiy, Kim, Shenker, Stoica, 2007; Dimitrov & Koptchev, 2010; Dannewitz, Kutscher,

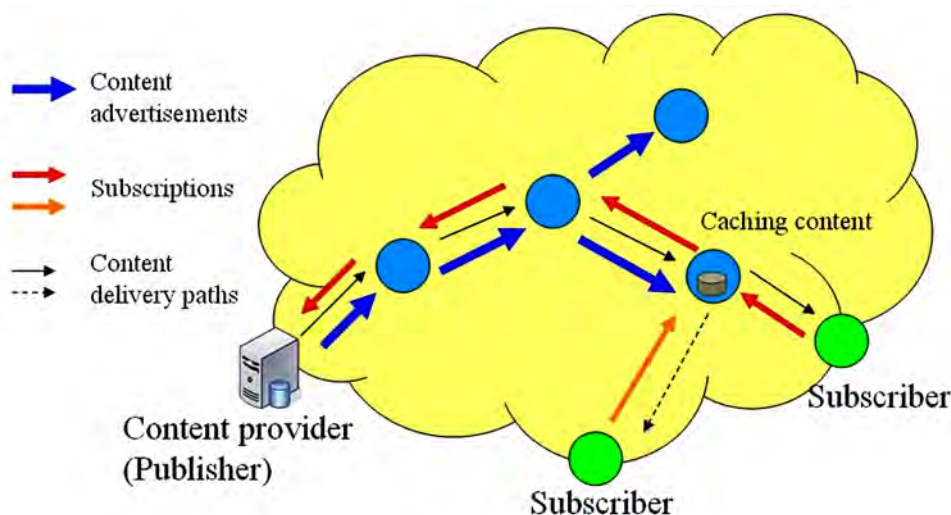
Ohlman, Farrell, Ahlgren, & Karl, 2013; Jacobson, Smetters, Thornton, Plass, Briggs & Braynard, 2009; García, Beben, Ramon, Maeso, Psaras, Pavlou, Wang, Sliwinski, Spirou, Soursos, Hadjioannou, 2011; Fall, 2003). The DTN has been based on a message oriented architecture used along with information-centric addressing and routing concepts.

Concepts of Information-Centric Networking

The ICN paradigm is founded on the idea of production and consumption of information based on user interest. Under ICN, the principle objective of the network is to expose, find and deliver information rather than the reachability of end-hosts and the maintenance of conversations between them (Almeida & Lourenco, 2012). The ICN paradigm mainly consists of two functional parts. They are namely the information dissemination or exposure and information retrieval. Thus the network is assumed to consist of interconnected pieces of information commonly known as content, information or data objects. The content can be of any type including an email, static or user generated data, real time media streams, online videos and music, more complex interactive multimedia communications, or even devices used for network management. Objects can be organized into clusters depending on the requirements to define social relationships or some ontology between them. They can also be mutable, combined or aggregated to form new objects. All these objects are addressed using names for the purpose of identification and routing while managed by applications or services at a upper middleware level. The object names are globally unique and independent of the location where they are stored.

Figure 1 shows the content advertisement and retrieval process in the information centric networks. The creator of any information who is commonly known as the publisher starts the process by advertising the availability of the content in the network without knowing who may be interested

Figure 1. Content advertisement and retrieval process in ICN (Almeida & Lourenco, 2012)



in the said content. Similarly, a receiver known as a subscriber or consumer may also declare its interest on certain quantity without knowing the creator or the current holder of the content in the network. When a receiver's intention matches a published information object, the network initiates a delivery path between a sender and the receiver of the content. Once the delivery path has been established, the content retrieval process would be begun for the receiver. As shown in Figure 1, an information delivery path is generally made up of two end points and several middle nodes. One of the end points is always the receiver who expressed its desire for the contents to the network while the other end point is the sender node that holds a copy of the information object. Hence the sender node need not always be the original creator of the information object but any node that holds (caches) a copy of the information object. Hence in ICN, the match interest rather than the locatability of the end points dictates the establishment of communication. The ICN connects the producers of information to consumers of that information disregarding the underlying hosts. Hence, for increasing the availability of information, dynamic in network caching is enabled in ICN. This in network caching not only improves the availability of information but also reduces the

network latency between the request and response, enables the faster reliable and scalable delivery with improved bandwidth usage reducing congestion. An intermediate router in the delivery path between the sender and receiver may cache the content objects so that the subsequent requests for the same content can be served rapidly by the router itself. This way, routing in ICN consists of finding and delivering copies of information objects from the most efficient location in the network rather than locating the host where it was published originally.

When information is retrieved from a cache rather than from the original source, it is important that the receiver must be provided with guarantees that the cached content indeed came from the original source in the first place. In order to satisfy this requirement, ICN embeds trust directly onto the contents rather than the nodes hosting them. This enables users to verify the quality of the content immediately eliminating the host related threats and vulnerabilities.

Naming of Information Objects

As information rather than hosts plays the central role in information centric networking, naming them with globally unique identifiers is the key

to the success of the entire paradigm. Once an information object is assigned with a globally unique name, it can be addressed and located in the network by users providing the name before the delivery path is created. The naming in ICN not only provides the contents with uniquely identifiable identifiers, but also includes other important properties such as pertinence, usability, scalability and security (Wong & Nikander, 2010). Due to the use of content caching and replication in ICN, name to address mapping creates a one-to-many relation.

Properties of Naming

Everything that is accessible in a network is considered information in ICN (van Adrichem & Kuipers, 2013). The objective of naming in ICN is not only to uniquely identify information objects in the network, but also to include important properties such as pertinence, usability, scalability and security.

- **Uniqueness:** Information objects need to be named in a globally unique fashion in the network. This is required to identify the contents and route them by using their name.
- **Persistence and Location-Independence:** The name needs to be invariant and independent of the location of the host that stores it. Content may generally be replicated or moved from one host to another within the network without service disruption, provided that the provider continues to serve it.
- **Usability and Scalability:** As a network of information, it would be possible to have a large number of data objects in the network where the objects are interconnected with each other and may have multiple interdependencies between them. Furthermore, objects may not be necessarily static and can be mutable. Thus, the naming scheme must not only be scalable but also be us-

able for dynamic objects. The scheme should also allow the deletion of objects automatically through some mechanism such as TTL.

- **Security:** ICN embeds security directly into the content without assuming trust on users or securing the communication channels. Content-based security ensures a cryptographic binding between content and its name to enable information objects to self certify themselves. This binding generally consists of a hash of the information provider's private key in the name of the information object. This allows the authentication of the information provider by creating a signature of the information object using the provider's private key that can be verified with its self-certified public key sent as metadata along with the information object. Confidentiality and integrity of the information are also guaranteed by a public key encryption. An important issue that needs special attention in this self-certification of information is the revocation of information objects when the public key is compromised or information is updated.

The naming schemes used in ICN can be divided into two types based on their structure (Vasilakos, Li, Simon & You, 2015). They are namely, hierarchical and flat naming architectures. Each naming structure has its own advantages as well as disadvantages. The hierarchical naming structure is similar to that of classless inter-domain routing structure of IP addresses. The IP addresses can be aggregated into prefixes performing either longest or shortest match. Similar to domain name systems used in IP networks, hierarchical names used in ICN can also be aggregated and extended (Jacobson et al., 2009). This helps to improve security and authenticity of the provenance of content that is useful in digital rights management. The unique names thus developed are human readable as well as used directly for routing contents

from source to destination. However, the binding between hierarchical names and administrative domains compromises the persistence of information objects. On the other hand, the flat naming structure is not suitable for aggregation but enables easier and faster DHT-like lookups (Stoica, Morris, Karger, Kaashoek, & Balakrishnan, 2001).

The unique names used in ICN include additional metadata for describing contents (Duval, Hodgins, Sutton & Weibel, 2002). The metadata is important in ICN as it describes the content better. Also it helps users to search and find information using keywords rather than the name itself. The metadata can also be used to establish an association between similar information objects. Metadata can also provide cryptographic inputs to perform more complex security checks on information objects. For example, metadata of an information object may contain the provider's public key combined with a digital signature that helps the users to verify the origin of the information. The network can also use the metadata to perform QoS guarantees on information objects, especially for real-time streaming multimedia contents. On the other hand, metadata can be defined in such a manner that it provides an efficient method for monitoring the network health by providing the administrators with useful statistics.

Addressing and Routing in Information Centric Networking

Addressing provides the reachability information of the objects in the network by mapping names to hosting locations (Benmohamed, Cole & Doshi, 2011). Routing in ICN has been implemented using two different strategies (Vasilakos et al., 2015). They are undirected name resolution service or the direct name based routing. The name resolution requires one or several centralized servers such as rendezvous points, register servers, trigger points, etc., in the network. The content publications are collected in these servers and they have a global view of all the published information objects and the network topology. When an ICN router wants

to forward a request message, the routing path is calculated in the centralized server by implementing a link state protocol providing the shortest path between the source and destination. On the other hand, the name based routing is directly performed by the ICN routers. Each router contains a local forwarding information base filled with content publication messages. The request forwarding paths are calculated in the local routers using their own forwarding strategies.

FUTURE RESEARCH DIRECTIONS

Many ICN implementations are still in their early stages and do not address all the requirements of a future internet architecture yet (Wang, 2013). The important aspects such as security, QoS considerations, scalability and network management require to be addressed in more detail and comprehensively. The following is brief description of the issues that demand the immediate attention of the research community to make the information centric network a success.

Security

Security is an essential part of any network architecture that aspires to be future safe (Bellovin, Clark, Perrig & Song, 2005). Hence the ICN researchers have also concentrated on security from the beginning. In ICN, instead of securing communication channels, the ICN model enforces security directly on to the information objects by providing for protection of the authenticity of the objects (Abdallah, Hassanein & Zulkernine, 2015). The use of self-certifying naming based on public-key cryptography leaves certain issues such as key compromise, revocation and management to be resolved by external public key infrastructures or the naming scheme itself. For example, CCN recommends for the distribution of keys as a specific type of information objects using the SDSI/SPKI model. In DONA, key revocations for a principal P can be handled by publication of the

form $P:L$ for some reserved name L . Some of the open security issues in ICN that have attracted the attention of researchers include the problems relating denial of service attacks, content poisoning, cache proliferation attacks and, fuzzy pub/sub message flooding or disturbance (Tourani, Mick, Misra & Panwar, 2016; Abdallah et al., 2015). There can be many different solutions to the above mentioned issues, but it may be interesting to investigate how far probabilistic data streaming algorithms integrated into the ICN for enforcing rate limiting or signature analysis on information objects exchanged between users.

Denial of service attacks generally target either the content provider himself or intermediate routers holding copies of information objects. The general nature of the attack is to target a caching router with invalid requests filling the router's pending interest table (PIT) memory. This results in genuine users being prevented from serving due to lack of resources in the routers. Several researchers have proposed solutions to this kind of DoS attack using different theories, mechanisms and techniques as their base. Some of the proposed solutions include queuing theory based approach, modified token bucket algorithm based vanilla algorithms, per-name prefix rate limiting, Poseidon mechanism, maintaining m-lists and statistical hypothesis testing theory based interest flooding detecting (Tourani et al., 2016). Though these solutions could successfully mitigate the effects of DoS attacks under different conditions, assumptions made during the design of these mechanisms limit the general applicability of these mechanisms. Hence this area is still open for further investigation.

Content poisoning attacks target the memory of the content caches and fill them with invalid contents (Tourani et al., 2016). For injecting the network with invalid contents, an attacker must first get the control of one or more content providers or intermediate routers. Once an attacker brings a content provider under its control, it can flood the network with malicious contents forcing the intermediate routers to store them in their

caches. The counter measures proposed include self-certifying interest/data packet helping the routers to verify and validate the contents before storing or forwarding, ranking mechanism for cached content using exclusion based feedback and signature verification techniques. Though these mechanisms help reduce the amount of content poisoning, they put additional burdens on the network and routers increasing the latency and energy consumption. Hence there is still room for coming up with light-weighted schemes for mitigating the effects of this problem.

Cache pollution attack targets the basic premise on which the entire caching mechanism is built on. Popularity of content in the Internet follows Zipf distribution where a smaller number of popular content are requested frequently. Hence these popular contents can be cached in the edge routers reducing the latency and loading on the Internet links saving bandwidth. The attackers exploits this basic premise and request less popular contents more frequently skewing the distribution. This makes caching less effective polluting the cache with less popular contents. The mitigating solutions proposed include *CacheShield*: a mechanism providing robustness against the locality disruption attack, machine-learning based algorithms and detection scheme based on randomness check. Many of the above proposed solutions incur high computation cost at the intermediate routers making them less scalable. Hence opportunities are still there for coming up with more scalable lightweight mechanisms.

Other areas of ICN security that also require the attention of researchers and developers include, secure naming and routing schemes, application level security, privacy and anonymity in ICN and access control.

Quality of Service Considerations

Quality of Service is concerned with the optimal use of shared resources by allocating them among different classes of users meeting their stipulated requirements (Kaur, 2011; Reddy, Hemalatha

& Mubeen, 2014). The principal objective of QoS management is to apply different priority to different data flows increasing the utility of the network while guaranteeing the required level of performance. With respect to other types of networking schemes such as IP and MPLS, ICN is better placed to manage quality of service based on different traffic flows (Al-Naday, Bontozoglou, Vassilakis & Reed, 2014). This advantage is achieved due to the fact that ICN has a knowledge of the information type before it has been located and transported. But different ICN architectures provide different methods for content identification making it difficult to port mechanisms developed for one type to another.

Scalability

Similar to the existing IP based Internet, the ICN would also grow to become a network with billions of devices, users and objects (Ghodsi et al., 2011). Hence the future ICN network would be required to distribute these billions of information objects to billions of users (devices). This creates two important issues with respect to scalability. They are naming scalability and routing (forwarding) scalability (Xylomenos, Ververidis, Siris, Fotiou, Tsilopoulos, Vasilakos, Katsaros & Polyzos, 2014). With the growth of the namespace, it would be extremely difficult to avoid naming conflicts guaranteeing globally unique names. Both flat and hierarchical naming structures would face this infinitely growing namespace problem at one or the other time. The other associated problem is routing as ICN employs name based routing. Hence this still remains an open research area to be solved.

Network Management

Many of the ICN proposals lack network management as part of the core architecture (Almeida & Lourenco, 2012). Not only the original proposals have failed to include network management as a

core concept within the architectures, even the later researches have also paid scant attention to this important area. Eum et al (2015) have proposed to design an ICN architecture within the framework of software defined network (SDN) in order to achieve the combined advantage of both networking schemes. SDN provides the advantage of decoupling network management from the underlying network infrastructure. This lowers the cost and complexity of network management enabling the continuous evolution of the network architecture in a flexible manner. Hence combining both ICN and SDN design principles will achieve a synergy between them supporting efficient dissemination of information providing flexible management framework. Other than the above mentioned proposal, this area has not seen any other contribution. Hence it remains an open area for further research.

CONCLUSION

In this chapter, the author has presented a comprehensive overview of the emerging networking paradigm of information centric networking. ICN aims to overcome the limitations of the current Internet architecture moving away from the host centric model to a data (information) centric model. The chapter presented the design principles of ICN in detail along with the major topics of naming, addressing and routing. Finally the chapter took a look at the open research challenges in the chosen area that needs the immediate attention of the researchers.

REFERENCES

Abdallah, E. G., Hassanein, H. S., & Zulkernine, M. (2015). A survey of security attacks in information-centric networking. *IEEE Communications Surveys and Tutorials*, 17(3), 1441–1454. doi:10.1109/COMST.2015.2392629

- Ahlgren, B., Dannewitz, C., Imbrenda, C., Kutscher, D., & Ohlman, B. (2012). A Survey of Information-Centric Networking. *IEEE Communications Magazine*, 50(7), 26–36. doi:10.1109/MCOM.2012.6231276
- Al-Naday, M. F., Bontozoglou, A., Vassilakis, V. G., & Reed, M. J. (2014). *Quality of service in an information-centric network*. Paper presented at the IEEE Global Communications Conference. doi:10.1109/GLOCOM.2014.7037079
- Almagor, R. C. (2011). Internet history. *International Journal of Technoethics*, 2(2), 45–64. doi:10.4018/jte.2011040104
- Almeida, F., & Lourenco, J. (2012). Information centric networks – Design issues, principles and approaches. *International Journal on Latest Trends Computing*, 3(3), 58–66.
- Bellovin, S. M., Clark, D. D., Perrig, A., & Song, D. (2005). *A clean-slate design for the next-generation secure internet*. Paper presented at the National Science Foundation Workshop on Next-Generation Secure Internet, Pittsburgh, PA.
- Benmohamed, L., Cole, R. G., & Doshi, B. T. (2011). Naming and addressing for routing scalability. *Johns Hopkins APL Technical Digest*, 30(2), 168–180.
- Dannewitz, C., Kutscher, D., Ohlman, B., Farrell, S., Ahlgren, B., & Karl, H. (2013). Network of Information (NetInf) - An information-centric networking architecture. *Journal of Computer Communications*, 36(7), 721–735. doi:10.1016/j.comcom.2013.01.009
- Dimitrov, V., & Koptchev, V. (2010). *PSIRP project - publish-subscribe internet routing paradigm: New ideas for future Internet*. Paper presented at the 11th International Conference on Computer Systems and Technologies, Sofia, Bulgaria. doi:10.1145/1839379.1839409
- Duval, E., Hodgins, W., Sutton, S., & Weibel, S. L. (2002). Metadata principles and practicalities. *D-Lib Magazine*, 8(4), 1–15. doi:10.1045/april2002-weibel
- Eum, S., Jibiki, M., Murata, M., Asaeda, H., & Nishinaga, N. (2015). *A design of an ICN architecture within the framework of SDN*. Paper presented at the Seventh International Conference on Ubiquitous and Future Networks, Sapporo, Japan. doi:10.1109/ICUFN.2015.7182521
- Fall, K. (2003). *A delay-tolerant network architecture for challenged internets*. Paper presented at the Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications, Karlsruhe, Germany. doi:10.1145/863955.863960
- García, G., Beben, A., Ramón, F. J., Maeso, A., Psaras, I., Pavlou, G.,... Hadjioannou, E. (2011). *COMET: Content mediator architecture for content-aware networks*. Paper presented at the Future Network & Mobile Summit, Warsaw, Poland.
- Ghodsí, A., Shenker, S., Koponen, T., Singla, A., Raghavan, B., & Wilcox, J. (2011). *Information-centric networking: Seeing the forest for the trees*. Paper presented at the 10th ACM Workshop on Hot Topics in Networks, Cambridge, MA. doi:10.1145/2070562.2070563
- Hassan, S., Elbreiki, W., Firdhous, M., & Habbal, A. M. M. (2015). End-to-end networks Vs named data network: A critical evaluation. *Jurnal Teknologi*, 72(5), 71–76. doi:10.11113/jt.v72.3943
- Jacobson, V., Smetters, D. K., Thornton, J. D., Plass, M. F., Briggs, N. H., & Braynard, R. L. (2009). *Networking named content*. Paper presented at the 5th International Conference on Emerging Networking Experiments and Technologies, Rome, Italy. doi:10.1145/1658939.1658941

- Katsaros, K. V., Chai, W. K., Wang, N., Pavlou, G., Bontius, H., & Paolone, M. (2014). Information-centric networking for machine-to-machine data delivery – A case study in smart grid applications. *IEEE Network*, 28(3), 58–64. doi:10.1109/MNET.2014.6843233
- Kaur, A. (2011). An overview of quality of service computer network. *Indian Journal of Computer Science and Engineering*, 2(3), 470–475.
- Koponen, T., Chawla, M., Chun, B. G., Ermolinskiy, A., Kim, K. H., Shenker, S., & Stoica, I. (2007). *A data-oriented (and beyond) network architecture*. Paper presented at the Conference on Applications, Technologies, Architectures, And Protocols for Computer Communications, Kyoto, Japan.
- Pandey, T., Garg, D., & Gore, M. M. (2012). Publish/subscribe based information dissemination over VANET utilizing DHT. *Frontiers of Computer Science*, 6(6), 713–724.
- Paul, S., Pan, J., & Jain, R. (2011). Architectures for the future networks and the next generation Internet: A survey. *Computer Communications*, 34(1), 2–42. doi:10.1016/j.comcom.2010.08.001
- Reddy, D. R. K., Hemalatha, D., & Mubeen, A. (2014). A study on quality of service for computer networks. *IOSR Journal of Computer Engineering*, 16(1), 51–55. doi:10.9790/0661-16155155
- Stoica, I., Morris, R., Karger, D., Kaashoek, M. F., & Balakrishnan, H. (2001). Chord: a scalable peer-to-peer lookup service for internet applications. Paper presented at the Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications, San Diego, CA. doi:10.1145/383059.383071
- Tourani, R., Mick, T., Misra, S., & Panwar, G. (2016). *Security, privacy, and access control in information-centric networking: A survey*. Available at arXiv:1603:03409v1 [cs.NI].
- van Adrichem, N. L. M., & Kuipers, F. A. (2013). Globally accessible names in named data networking. Paper presented at the IEEE Conference on Computer Communications Workshops, Turin, Italy. doi:10.1109/INFCOMW.2013.6970714
- Vasilakos, A. V., Li, Z., Simon, G., & You, W. (2015). Information centric network: Research challenges and opportunities. *Journal of Network and Computer Applications*, 52, 1–10. doi:10.1016/j.jnca.2015.02.001
- Wang, Y. (2013). *Caching, routing and congestion control in a future information-centric internet* (PhD Thesis). North Carolina State University, Raleigh, NC.
- Wong, W., & Nikander, P. (2010). *Secure naming in information-centric networks*. Paper presented at the Workshop on the Re-Architecting the Internet, Philadelphia, PA.
- Xylomenos, G., Ververidis, C. N., Siris, V. A., Fotiou, N., Tsilopoulos, C., Vasilakos, X., & Polyzos, G. C. et al. (2014). A survey of information-centric networking research. *IEEE Communications Surveys and Tutorials*, 16(2), 1024–1049. doi:10.1109/SURV.2013.070813.00063

ADDITIONAL READING

- Bartolomeo, G., & Kovacikova, T. (2013). *Identification and management of distributed data: NGN, content-centric networks and the web*. Boca Raton, FL: CRC Press. doi:10.1201/b14966
- Clark, R. M., & Mitchell, W. L. (2016). *Target-centric network modeling: Case studies in analyzing complex intelligence issues*. London, UK: SAGE Publications.
- de Brito, G. M., Velloso, P. B., Igor, M., & Moraes, I. M. (2013). *Information centric networks: A new paradigm for the Internet*. Hoboken, NJ: Wiley-ISTE. doi:10.1002/9781118579848

Feng, B., Zhou, H., Zhang, H., Jiang, J., & Yu, S. (2016). A Popularity-based cache consistency mechanism for information-centric networking. To appear in IEEE Globecom, Washington DC.

Ohsugi, K., Takemasa, J., Koizumi, Y., Hasegawa, T., & Psaras, I. (2016). Power consumption model of NDN-based multicore software router based on detailed protocol analysis. *IEEE Journal on Selected Areas in Communications*, 34(5), 1631–1644. doi:10.1109/JSAC.2016.2520278

Pathan, M., Sitaraman, R. K., & Robinson, D. (2014). *Advanced content delivery, streaming, and cloud services*. Hoboken, NJ: Wiley-ISTE. doi:10.1002/9781118909690

Tarkoma, S. (2012). *Publish / subscribe systems: Design and principles*. Hoboken, NJ: Wiley-ISTE. doi:10.1002/9781118354261

KEY TERMS AND DEFINITIONS

Attack: An attempt to gain unauthorized access to, steal, expose, alter, disable or destroy computing resources by unauthorized personnel or through unauthorized means.

Content Centric Networking (CCN): A networking paradigm that emphasizes content by making it directly addressable and routable.

Future Internet: A general term coined for describing all the research activities carried out for developing new architectures for the Internet with the view of overcoming its shortcomings.

Information Centric Networking: An approach to evolve the Internet infrastructure to identify named information independently of the distribution channel by introducing uniquely named data as a core Internet principle.

Metadata: A set of data (information) used for describing the actual user data.

Named Data Networking (NDN): A Future Internet architecture designed with the objective of solving problems associated with contemporary internet architectures.

Network Architecture: A framework for the specification of a network's physical components, functional organization, configuration, operational principles and procedures along with the data formats used.

Publish/Subscribe Paradigm: A messaging scheme where senders of messages known as publishers stores the messages in an intermediary location without sending them directly to specific receivers (subscribers). Subscribers can retrieve these messages from these intermediary nodes as and when required.

Security: Denotes the protective mechanisms including policies, guidelines and algorithms applied to computers, computer networks, users and data stored and transmitted over them.

Interoperability Frameworks for Distributed Systems

José Carlos Martins Delgado
Universidade de Lisboa, Portugal

INTRODUCTION

A system is *distributed* with respect to another if their lifecycles are not dependent, i.e., if one can evolve (to a new version) without having to change, to suspend or to stop the behavior or interface of the other.

Different systems usually need to interact, raising the concept of *interoperability* (ISO/IEC/IEEE, 2010), which literally means the *ability* (of two or more systems) to *operate together*. An *interoperability framework* is a set of principles, assumptions, rules and guidelines to analyze, to structure and to classify the concepts and concerns of interoperability.

What these words really mean largely depends on the domain which the systems belong to, although there is a pervasive, underlying notion that these systems are active, reacting upon stimuli sent by others and cooperating to accomplish higher-level goals than those achievable by each single system.

Interoperability has been studied in domains such as enterprise cooperation (Popplewell, 2014; Rezaei, Chiew, & Lee, 2014), e-government services (Sharma & Panigrahi, 2015), military operations (Hussain, Mehmood, Haq, Alnafjan, & Alghamdi, 2014), cloud computing (Zhang, Wu, & Cheung, 2013), healthcare applications (Robkin, Weininger, Preciado, & Goldman, 2015), digital libraries (Agosti, Ferro, & Silvello, 2016) and metadata (Chen, 2015).

In this article, we adopt a more general perspective, exploring interoperability in the generic context of distributed systems, independently of what they are or which domain is the most rel-

evant to them. The following section describes some of the most relevant existing interoperability frameworks. The section after that one introduces the basic concepts that establish a foundation for interoperability in distributed contexts. Next, a multidimensional interoperability framework is proposed and its advantages discussed. Finally, future research directions are hinted and conclusions drawn.

BACKGROUND

The Open Systems Interconnection (OSI) reference model (ISO/IEC, 1994), constitutes one of the first systematizations of distributed interoperability, considering seven layers (Table 1). This standard deals mostly with communication issues, with the objective of sending data and reproducing it at the receiver. How those data are interpreted by the receiver and how it reacts to the data is left unspecified, encompassed by the topmost layer, Application. However, interoperability must ensure not only the exchange of data but also the meaningful use of information (ISO/IEC/IEEE, 2010), which means that this layer must be detailed.

Table 1 depicts the basic structure of several interoperability frameworks (referred to by acronym or first author) that use this layered approach, establishing a rough horizontal correspondence between layers.

The C4IF framework (Peristeras & Tarabanis, 2006), is based on four layers: Connection (basic use of a channel), Communication (data formats), Consolidation (meaning through semantics) and Collaboration (through compatible processes). It

Table 1. Comparison between several layered interoperability frameworks

OSI (1994)	C4IF (2006)	Lewis (2008)	Stamper (2000)	LCIM (2009)	EIF (2010)	Monfelt (2011)
Application	Collaboration	Organizational	Social world	Conceptual	Political	SWOT
						Cultural
						Ethical
					Legal	Legal
	Consolidation	Semantic	Pragmatic	Dynamic	Organizational	Managerial
				Pragmatic		Organizational
				Semantic	Semantic (includes syntactic)	Adaptation
Presentation	Communication	Syntactic	Syntactic	Syntactic		Application
Session						Presentation
Transport						Session
Network	Connection	Machine	Empirics	Technical	Technical	Transport
Link						Network
Physical Medium			Physical world			Link
						Physical Medium

simplifies the lower levels (distinguishing only connectivity and communication) and refines the application layer, distinguishing information semantics from behavior.

Lewis, Morris, Simanta, and Wrage (2008) proposed a similar framework, with slight differences but with basically the same structure.

Stamper, Liu, Hafkamp, and Ades (2000) applied *semiotics* (the study of signs, stemming from linguistics) to the field of information systems and proposed a semiotic ladder, a layered structure in which each layer builds on the previous one (just like a ladder) in an increasing level of abstraction and complexity. Besides the usual syntax and semantics, pragmatics was used to refer to the interaction context and the effect caused by the reception of a message by a receiver. Empirics refer to the lower levels using the physical world, which details are well established and become less relevant to the understanding of interoperability as a whole. The social world tackles the higher levels, in which people become more involved.

Wang, Tolk, and Wang (2009) described the LCIM framework, which follows the semiotic

ladder in essence, with the interesting addition of a dynamic layer that considers evolution along the system's lifecycle.

The European Interoperability Framework (EIF, 2010) was conceived as a broad framework for the interoperability of public services and establishes four main interoperability levels (legal, organizational, semantic and technical), with an upper political context that should ensure compatible visions and aligned priorities.

Monfelt, Pilemalm, Hallberg, and Yngström (2011) further refined the social layer of the semiotic ladder and extended the basic OSI reference model with another seven layers, to take care of higher-level issues, such as risk (SWOT analysis) and dependencies on social and organizational aspects concerning cultural, ethical and legal values, as well as existing administrative and managerial issues. The organizational layer pertains to the pragmatics of the message (its effects in the interaction context) and the adaptation layer pertains to the semantics of the message, adapting the new layers to the technical layers provided by the OSI model.

Other interoperability frameworks, particularly those conceived for complex systems, such as enterprises, try to complete the scenario by considering more than one dimension of interoperability, so that aspects of a different nature can be tackled in an orthogonal way. For example, the lower layers in Table 1 have a clear operational nature, but the higher layers are more conceptual and as such should be dealt with in a previous stage of the systems' lifecycle. Table 2 gives a brief description of these multidimensional frameworks.

Even with multidimensional frameworks to capture the various aspects of interoperability, there is the need to perform an assessment of to which degree systems are interoperable. Doing this on every dimension and aspect is complex and reduces the global perception of the interoper-

ability goals. Therefore, a typical approach is to perform a qualitative assessment in a single scale, by considering the maturity of the systems and of their relationship in terms of interoperability. This involves establishing a set of levels that give a qualitative indication of how far in the abstraction scale has interoperability been considered.

In fact, several frameworks are defined more in terms of the maturity of systems' interoperability than on layer structure or dimensions. Table 3 illustrates the classification of several of these frameworks. Each framework has its own slant and terminology, but overall they possess rather similar structures, ranging from a complete lack of interoperability (level 0) to full integration (level 4).

Table 2. Several multidimensional interoperability frameworks

Framework	Dimensions	Description
SOSI (Morris, Levine, Meyers, Place, & Plakosh, 2004)	Operational	Issues between existing systems that need to interoperate, such as lack of information and non-intuitive user interfaces
	Programmatic	Interoperability between computer applications and solutions
	Constructive	Architectural issues such as data schema interoperability
AIF, ATHENA Interoperability Framework (Berre et al., 2007)	Conceptual	Integration models at the levels of enterprise/business, process, service and information/data
	Application	Methodology with methods to support the development of interoperability projects
	Technical	Implementation basis around technologies such as BPEL, Web Services and XML
FEI, Framework for Enterprise Interoperability (ISO, 2011)	Concerns	At the levels of business, processes, services and data
	Barriers	Difficulties in making systems interoperable (conceptual, technological and organizational)
	Approach	Plan to overcome the barriers. Can be organized in three ways: • Integrated (a common format is used by all systems); • Unified (common goals and semantic mapping); • Federated (common goals and ontology are possible, but each system must adapt to others dynamically, without any imposed model).
ULS Interoperability Framework (Ostadzadeh, Shams, & Badie, 2015)	Abstract	Six classical questions of the Zachman framework (O'Rourke, Fishman, & Selkow, 2003), in the interoperability context: what, how, where, who, why, when
	Perspective	Layered interoperability concerns (contextual, conceptual, logical, physical and out-of-context)
	Barriers	Considers the dimensions of SOSI (Operational, Constructive and Programmatic) as barriers and extends them with cultural barriers
ENSEMBLE European Project (Charalabidis, Lampathaki, & Jardim-Goncalves, 2014)	Domains	Interoperability problem space: social, applied and formal sciences
	Areas	Scientific areas underlying interoperability: data, process, rules, objects, software, cultural, knowledge, services, social networks, electronic identity, cloud and ecosystems
	Elements	Scientific interoperability elements: semantics, models, tools, orchestration

Table 3. Interoperability maturity levels of several frameworks

Level	Typical meaning	LISI (C4ISR, 1998)	OIM (Fewell & Clark, 2003)	OIAM (Kingston, Fewell, & Richer, 2005)	MMEI (Guédria, Chen, & Naudet, 2009)
4	Capable of dynamically adapting to partner's intent	Enterprise	Unified	Dynamic	Adapted
3	Capable of meaningfully exchange semantic content	Domain	Combined	Open	Organized
2	Capable of invoking functionality at the syntax level (common formats)	Functional	Collaborative	Accommodating	Aligned
1	Basic connectivity at the message level	Connected	Ad hoc	Amenable	Defined
0	Unable to interact	Isolated	Independent	Static	Unprepared

THE FOUNDATIONS OF INTEROPERABILITY

Defining the Meaning of Interoperability

Figure 1 illustrates the most basic interoperability scenario, with a system in the role of *consumer* and another in the role of *provider*. The consumer sends a request to the provider, which reacts to it, typically executing some actions and eventually sending back a response to the consumer. System interaction does not have to be message based. A file written by the consumer and later read by the provider will have the same effect. This basic scenario can be extended by composition. The same system can act as a consumer and as a provider, in bidirectional interactions, or several

systems can take part in a multipartite interaction (a *choreography*) towards some common goal.

The 24765 standard (ISO/IEC/IEEE, 2010) provides one of the most cited definitions of interoperability: “The ability of two or more systems or components to exchange information and to use the information that has been exchanged”. This means that interoperability involves more than mere communication. Interacting systems must be able to understand the information they receive, so that they can use it, at various levels of abstraction, as already detailed by Table 1.

Table 4 provides our own proposal for the interoperability layers, which constitutes the first dimension of our interoperability framework. The goal with respect to existing frameworks is to provide an additional systematization of the upper interoperability layers.

Figure 1. A simple request-response interaction between two systems

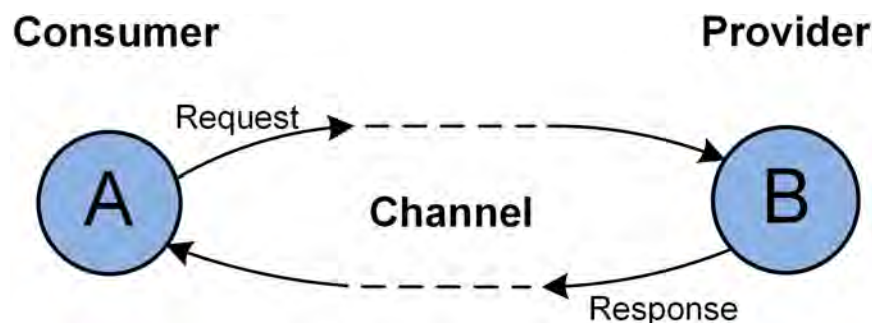


Table 4. Layered perspective of interoperability

Category	Layer	Main artifact	Description
Symbiotic (purpose and intent)	Coordination	Governance	Motivations to have the interaction, with varying levels of mutual knowledge of governance, strategy and goals
	Alignment	Joint-venture	
	Collaboration	Partnership	
	Cooperation	Outsourcing	
Pragmatic (context and reaction effects)	Contract	Choreography	Management of the interaction context and of the reaction effects of the interactions at the levels of choreography, process and service
	Workflow	Process	
	Interface	Service	
Semantic (meaning of content)	Inference	Rule base	Interpretation of a message at the levels of rule, known system components and relations, and definition of concepts
	Knowledge	Knowledge base	
	Ontology	Concept	
Syntactic (notation of representation)	Structure	Schema	Representation of system components, in terms of composition, primitive components and their serialization format in messages
	Predefined type	Primitive resource	
	Serialization	Message format	
Connective (transfer protocol)	Messaging	Message protocol	Lower-level formats and network protocols involved in transferring a message from the context of the sender to that of the receiver
	Routing	Gateway	
	Communication	Network protocol	
	Physics	Media protocol	

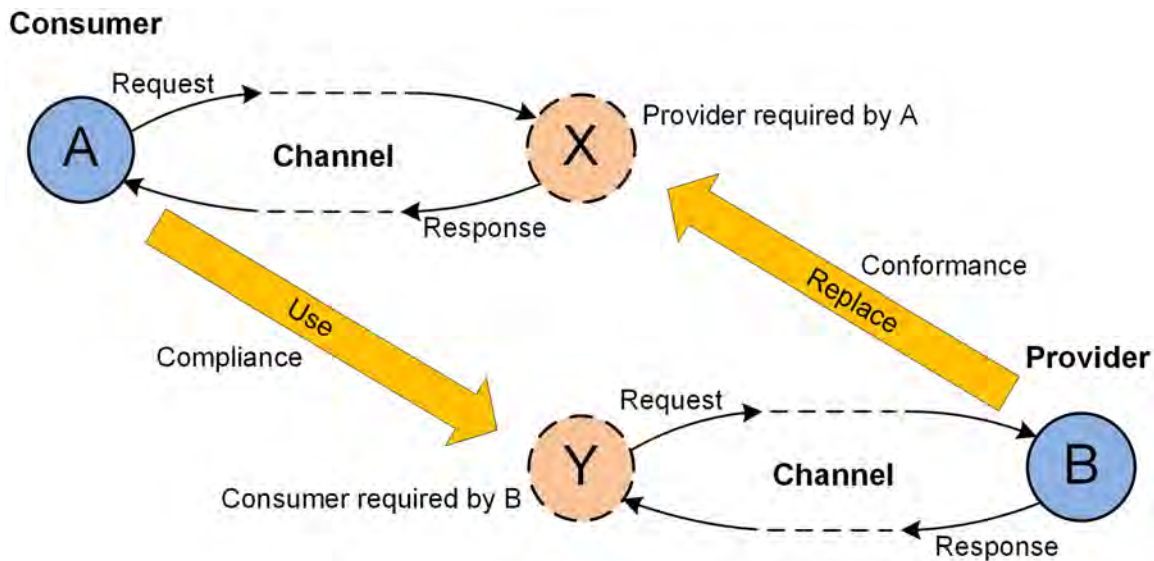
Partial Interoperability: Compliance and Conformance

The 24765 standard (ISO/IEC/IEEE, 2010) actually provides three definitions of interoperability, stemming from previous and now superseded standards. The second definition relevant to this article is: “The capability to communicate, execute programs, and transfer data among various functional units in a manner that requires the user to have little or no knowledge of the unique characteristics of those units”. Here the emphasis is on decoupling, meaning that interacting systems need not be designed for a full match, as long as some characteristics support the interaction, and can evolve independently, as long as these characteristics are maintained. This leads to the concept of *degree of interoperability*, an indicator that expresses how much of the functionality or characteristics of a system can be accessed by another, ranging from none to the full set.

Consider Figure 2, an extension of the scenario of Figure 1, in which system *A* (the consumer) and system *B* (the provider) have been designed independently. *A* was designed to interact with a provider *X* and *B* was designed to expect a consumer *Y*. Having *A* interoperable with *B* involves two conditions (Delgado, 2014):

- Compliance (Tran, Zdun, Oberortner, Mulo, & Dustdar, 2012). *A* must *comply with Y*, which means that *A* must satisfy all the *requirements* of *Y* in order to send requests to *B*. Therefore, *A* can use *Y* to reach *B* (“Use” arrow), without noticing the difference;
- Conformance (Khalfallah, Figay, Barhamgi, & Ghodous, 2014). *B* must *conform to X*, which means that *B* must fulfill all the *expectations* of *A* regarding the effects of requests it sends to *X*. Therefore, *B* can effectively replace (take the form of) *X* without *A* noticing it (“Replace” arrow).

Figure 2. Compliance and conformance. X is how A views provider B . Y is how B views consumer A . X and Y are specifications, not actual systems



Partial interoperability can be achieved by *subsumption*, with the set of provider's features that A uses as a subset of the set of features offered by B , and as long as B (or another system that replaces it) supports the expectations of A regarding the reactions to requests.

In many cases, systems are made interoperable by design, i.e., conceived and implemented to work together (as in Figure 1). When systems are complex (such as enterprises) and evolve in an independent way, ensuring interoperability is not an easy task. A typical and pragmatic solution is to resort to Web Services and XML data, sharing schemas and namespaces, but this entails a high degree of coupling since it involves the entire interface of the provider.

The notion of partial interoperability introduces a different perspective, stronger than similarity but weaker than commonality (resulting from using the same schemas and ontologies). The trick is to consider only the intersection between what the consumer needs and what the provider offers. If the latter subsumes the former, the degree of interoperability required by the consumer can be granted, regardless of whether the provider supports additional features or not.

The Role of System Lifecycle in Interoperability

Compliance and conformance apply not only to message-based interactions, in both content (syntax and semantics, including ontology) and protocol (flow of messages and their effects), but also to non-operational relationships, such as strategy alignment and regulatory compliance. This acknowledges the fact that interoperability is not merely about the operation stage of the system's lifecycle but starts much earlier, in the conception stage. After all, what happens in the operation stage is a consequence of what has been conceived and designed.

Figure 3 expresses this perspective by depicting information on two interacting systems, each schematized in two dimensions, lifecycle stages (horizontally) and interoperability abstraction layers (vertically), which express the detail at which interoperability is considered (see Table 4).

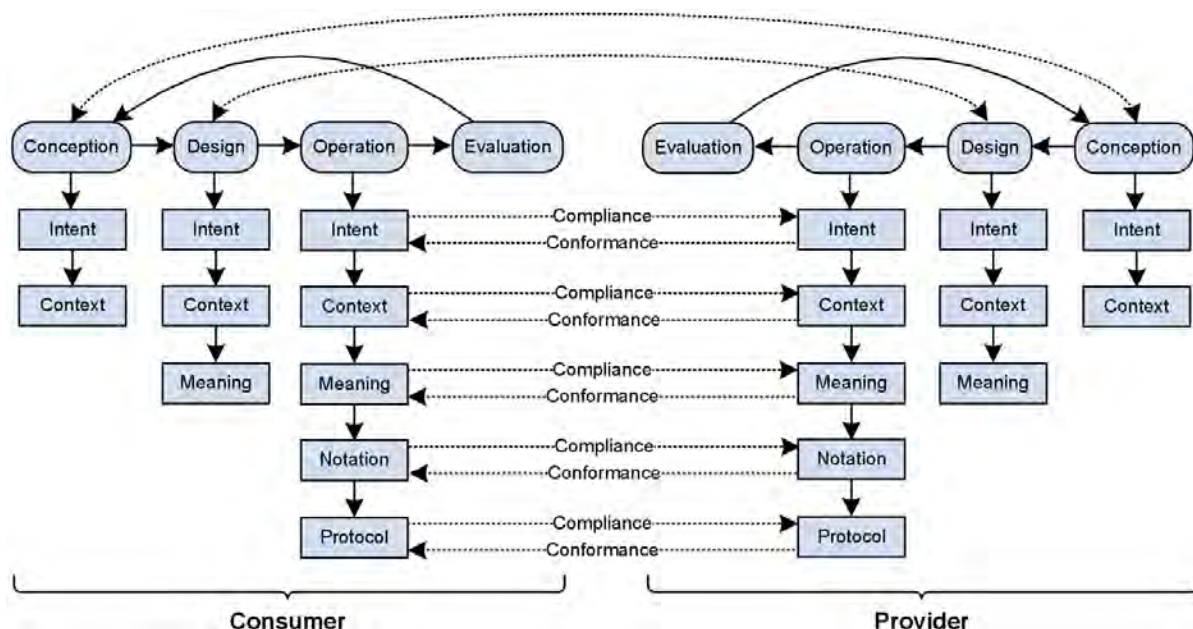
Figure 3 should be interpreted as follows:

- The lifecycles are simple, with just four stages, but they could easily be refined by

- detailing each of these stages while maintaining the whole interoperability rationale;
- All the layers of abstraction of interoperability should be considered in every stage. However, it is only natural that the method used to evolve the system along its lifecycle considers interoperability at an abstract layer in the initial stages and at full detail in the operation stage;
- For simplicity, we consider only the most relevant concept in each category of abstraction layers of interoperability (first column of Table 4: Intent, Context, Meaning, Notation, and Protocol), instead of each layer individually;
- There must be compliance (in the consumer to provider direction) and conformance (in the provider to consumer direction) in each corresponding cell (in a given stage and abstraction layer) of both interacting partners. This is illustrated in detail by Figure 3 in the operation stage only. However, the same should happen in other stages (the dashed arrows at the top of the figure indicate this). The rationale for this is twofold:

 - Partners must be interoperable at each abstraction layer of interoperability, be it at the intent layer (one must understand and accept what the other intends to achieve), at the meaning layer (interoperable ontologies and semantic relationships) or at the basic protocol layer (there must be message connectivity);
 - Each stage is a consequence of the previous stage, in a model-driven fashion. For systems to be interoperable at operation time, their conception and design must also be made interoperable;
- The evaluation stages need not necessarily be made interoperable. Each system can evaluate its design independently, although the specifications for interoperability must not be overlooked and need to be considered in the context of the partners' relationship.

Figure 3. System lifecycle combined with interoperability abstraction layers



A MULTIDIMENSIONAL INTEROPERABILITY FRAMEWORK

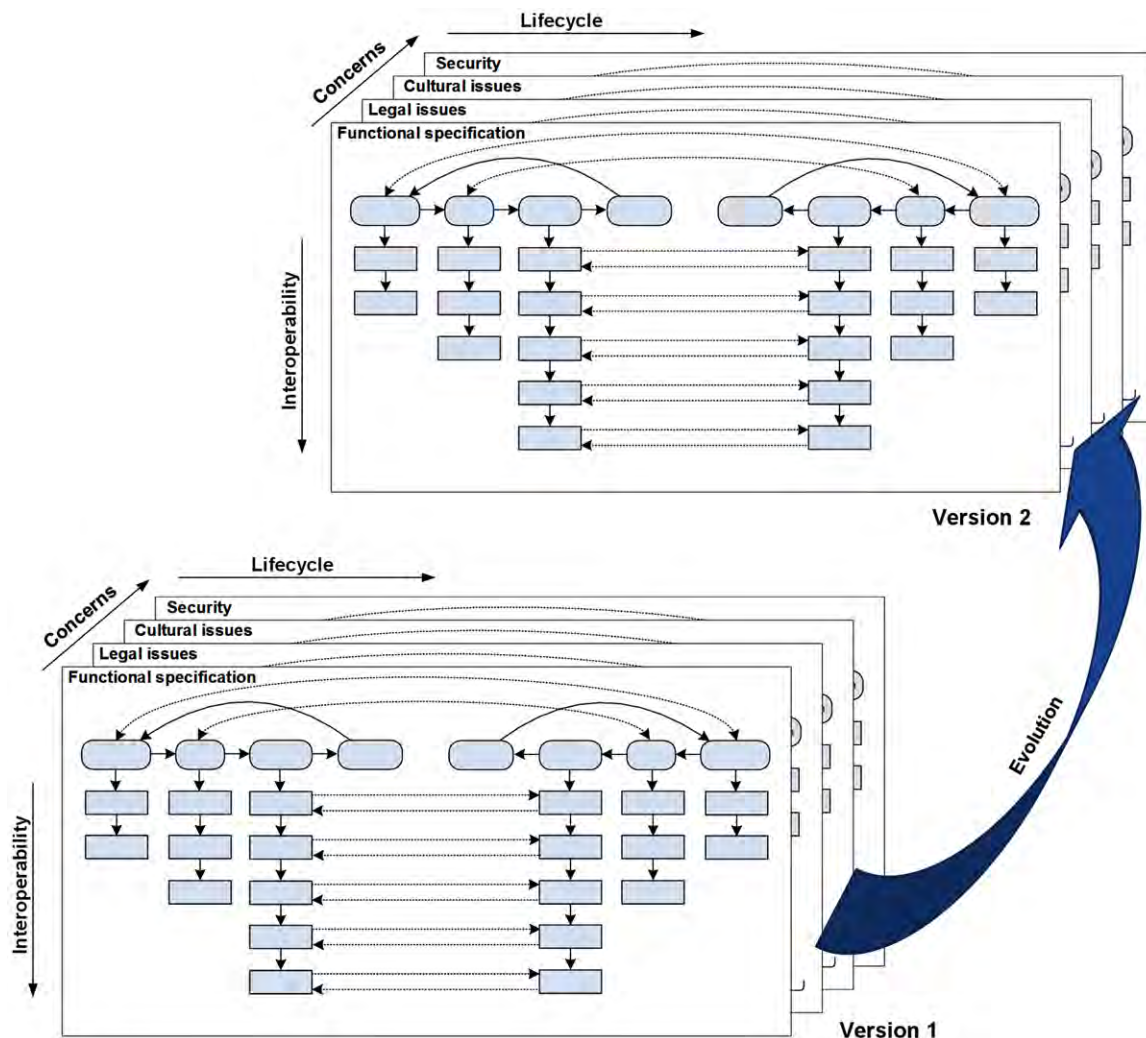
Figure 4 depicts a domain-independent and multidimensional interoperability framework, conceived generically for any type of systems.

The two main structuring dimensions are the Lifecycle (with its stages) and the Interoperability (with its abstraction layers), forming the plane shown in Figure 3. Other aspects, more specific or more domain-oriented, are introduced by a Concerns dimension, expressed in Figure 4 by one plane for each concern, such as the functional specification and legal, cultural and security issues.

The plane formed by the Concerns and Interoperability dimensions (in each stage of the lifecycle) corresponds to considering all aspects of the relationship with a partner in every abstraction layer, while developing the system. The plane formed by the Concerns and Lifecycle dimensions corresponds to considering the entire development of a partner, in all stages and aspects, for each layer of abstraction of the relationship between partners.

This framework is extensible, not only by considering additional concern planes but also by the ability to promote a concern to a full-fledged dimension, if its relevance and applicability breadth justifies that (if it has a variability that

Figure 4. A multidimensional interoperability framework



is relevant across the entire span of existing axes and refers to issues orthogonal to those dealt with by other axes).

A development method is needed to exercise this framework, from the origin of the axes, where everything is conceptual and abstract, to a point in this multidimensional space in which everything is defined and concrete. This method can vary, according to the path chosen along the dimensions (breadth-first, depth-first, and so on) to reach that point, but typically an iterative approach is followed, by progressing along all dimensions, in successive versions of increasing detail (as Figure 3 had already hinted). Figure 4 illustrates two versions of this evolution.

It should be noted that, although the Lifecycle-Interoperability plane shows two partners, their development is coupled only inasmuch as required by compliance and conformance. This illustrates that the development of systems is only partially independent. If interoperability is required, some degree of coupling and co-development is unavoidable.

The main advantages of this multidimensional framework with respect to those mentioned in Table 2 are:

- Generic and domain independent, by considering interacting partners as systems of any kind and complexity. Many other frameworks tend to favor some domain, such as enterprise engineering or military environments;
- Recursive, in the sense that it can be applied to a system or recursively to its components;
- Based on system development lifecycle, considering all stages, instead of just adding interoperability to existing systems;
- System coupling based on partial interoperability, expressed by compliance and conformance;
- Extensible, by promoting a specific concern (e.g., security) to an axis on its own right;

- Clear separation of framework from method. Most frameworks include an approach, which is part of the method, but these are different concepts that should be kept separate (e.g., to allow different methods to exercise a given framework).

FUTURE RESEARCH DIRECTIONS

The CEN EN/ISO 11354-1 standard (ISO, 2011), sets the trend in interoperability consolidation by standardization. However, we cannot expect this standard to apply to all domains and other interoperability standards will appear.

The ENSEMBLE project (Charalabidis, Lampathaki, & Jardim-Goncalves, 2014) has spurred a movement towards the establishment of enterprise interoperability as a science and the development of an interoperability body of knowledge (Jardim-Goncalves, Grilo, Agostinho, Lampathaki, & Charalabidis, 2013).

The interoperability framework described in this article can make a contribution to this effort, although it needs to be improved and made more complete, in several ways:

- The Concerns axis needs to be detailed, to include a better perspective of relevant concerns such as security and common domain-specific aspects and problems, such as those uncovered by the ENSEMBLE project;
- Compliance and conformance are basic concepts, applicable to all domains and layers of abstraction and complexity, but need formalization and definition of their exact meaning in each of the abstraction layers of interoperability of Table 4;
- A method to exercise the framework needs to be devised, with a comparative case study with other interoperability methods, namely those promoted by relevant bodies, such as the European Interoperability Framework (EIF, 2010) or standardized, such as the Framework for Enterprise Interoperability (ISO, 2011).

CONCLUSION

The world is increasingly both distributed and interconnected. Interoperability is crucial and standards are not enough to ensure cooperation (Lewis, Morris, Simanta, & Wrage, 2008). Interoperability frameworks provide a classification structure to organize interoperability concerns, and several frameworks have been proposed, both single and multidimensional.

This article briefly described some of these frameworks and argued that, instead of establishing a framework according to the main issues in a given domain (such as e-government or healthcare), a set of foundational interoperability concepts (such as lifecycle, compliance, conformance and concerns) is enough to define a domain-independent framework, with truly orthogonal dimensions and a clear separation from a method that uses it. This framework and its dimensions were described.

Nevertheless, interoperability seems to constitute, by its very nature of conflicting goals, an unsolvable problem:

- Systems want to interact with others, but not to be dependent upon them;
- Sharing is good (and necessary), but privacy and information hiding is even better;
- Standardization is good but enforced normalization can hamper differentiation between providers.

The compromise is to apply the principle of least coupling, by which systems should interact by using just the minimum possible set of features. Compliance and conformance are the basic concepts that drive the applicability of this principle and an interoperability framework (such as the one described in this article) constitutes the basic tool to achieve it.

REFERENCES

- C4ISR. (1998). *Levels of Information Systems Interoperability (LISI)*. C4ISR Architecture Working Group (AWG), Department of Defense. Retrieved November 13, 2015, from <http://www.eng.auburn.edu/~hamilton/security/DODAF/LISI.pdf>
- Agosti, M., Ferro, N., & Silvello, G. (2016). Digital library interoperability at high level of abstraction. *Future Generation Computer Systems*, 55, 129–146. doi:10.1016/j.future.2015.09.020
- Berre, A., Elvesæter, B., Figay, N., Guglielmina, C., Johnsen, S., & Karlsen, D. (2007). The ATHE-NA interoperability framework. In R. Gonçalves, J. Müller, K. Mertins, & M. Zelm (Eds.), *Enterprise Interoperability II* (pp. 569–580). London, UK: Springer. doi:10.1007/978-1-84628-858-6_62 doi:10.1007/978-1-84628-858-6_62
- Charalabidis, Y., Lampathaki, F., & Jardim-Goncalves, R. (2014). On the scientific foundations of enterprise interoperability: The ENSEMBLE Project and Beyond. In Y. Charalabidis, F. Lampathaki, & R. Jardim-Goncalves (Eds.), *Revolutionizing Enterprise Interoperability through Scientific Foundations* (pp. 336–355). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-5142-5.ch014
- Chen, Y. (2015). A RDF-based approach to metadata crosswalk for semantic interoperability at the data element level. *Library Hi Tech*, 33(2), 175–194. doi:10.1108/LHT-08-2014-0078
- Delgado, J. (2014). The role of compliance and conformance in software engineering. In I. Ghani, W. Kadir, & M. Ahmad (Eds.), *Handbook of Research on Emerging Advancements and Technologies in Software Engineering* (pp. 392–420). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-6026-7.ch017

- EIF. (2010). *European Interoperability Framework (EIF) for European public services Annex 2*. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of Regions 'Towards interoperability for European public services'. Retrieved November 13, 2015, from http://ec.europa.eu/isa/documents/isa_annex_ii_eif_en.pdf
- Fewell, S., & Clark, T. (2003). *Organisational interoperability: Evaluation and further development of the OIM model*. Defence Science And Technology Organisation. Retrieved November 13, 2015, from <http://www.dtic.mil/dtic/tr/fulltext/u2/a466378.pdf>
- Guédria, W., Chen, D., & Naudet, Y. (2009). A maturity model for enterprise interoperability. In R. Meersman, P. Herrero, & T. Dillon (Eds.), *On the Move to Meaningful Internet Systems Workshops* (pp. 216–225). Berlin, Germany: Springer.
- Hussain, T., Mehmood, R., Haq, A., Alnafjan, K., & Alghamdi, A. (2014). Designing framework for the interoperability of C4I systems. In *International Conference on Computational Science and Computational Intelligence* (102–106). Piscataway, NJ: IEEE Computer Society Press. doi:10.1109/CSCI.2014.102
- ISO. (2011). *Advanced Automation Technologies and their Applications, Part 1: Framework for Enterprise Interoperability (CEN EN/ISO 11354-1)*. Geneva, Switzerland: International Standards Office.
- ISO/IEC. (1994). *ISO/IEC 7498-1, Information technology – Open Systems Interconnection – Basic Reference Model: The Basic Model* (2nd ed.). Geneva, Switzerland: International Standards Office. Retrieved November 13, 2015, from <http://standards.iso.org/ittf/PubliclyAvailableStandards/index.html>
- ISO/IEC/IEEE. (2010). *Systems and software engineering – Vocabulary*. International Standard (ISO/IEC/IEEE 24765:2010(E)). Geneva, Switzerland: International Standards Office.
- Jardim-Goncalves, R., Grilo, A., Agostinho, C., Lampathaki, F., & Charalabidis, Y. (2013). Systematisation of interoperability body of knowledge: The foundation for enterprise interoperability as a science. *Enterprise Information Systems*, 7(1), 7–32. doi:10.1080/17517575.2012.684401
- Khalfallah, M., Figay, N., Barhamgi, M., & Ghodous, P. (2014). Model driven conformance testing for standardized services. In *IEEE International Conference on Services Computing* (pp. 400–407). Piscataway, NJ: IEEE Computer Society Press. doi:10.1109/SCC.2014.60
- Kingston, G., Fewell, S., & Richer, W. (2005). *An organisational interoperability agility model*. Defence Science And Technology Organisation. Retrieved November 13, 2015, from <http://www.dtic.mil/dtic/tr/fulltext/u2/a463924.pdf>
- Lewis, G., Morris, E., Simanta, S., & Wrage, L. (2008). Why standards are not enough to guarantee end-to-end interoperability. In C. Ncube & J. Carvallo (Eds.), *Seventh International Conference on Composition-Based Software Systems* (pp. 164–173). Piscataway, NJ: IEEE Computer Society Press.
- Monfelt, Y., Pilemalm, S., Hallberg, J., & Yngström, L. (2011). The 14-layered framework for including social and organizational aspects in security management. *Information Management & Computer Security*, 19(2), 124–133. doi:10.1108/09685221111143060 doi:10.1108/09685221111143060
- Morris, E., Levine, L., Meyers, C., Place, P., & Plakosh, D. (2004). *System of Systems Interoperability (SOSI): final report*. Report No. CMU/SEI-2004-TR-004. Carnegie Mellon Software Engineering Institute. Retrieved November 13, 2015, from <http://www.sei.cmu.edu/reports/04tr004.pdf>

O'Rourke, C., Fishman, N., & Selkow, W. (2003). *Enterprise architecture using the Zachman framework*. Boston, MA: Course Technology.

Ostadzadeh, S., Shams, F., & Badie, K. (2015). An architectural model framework to improve digital ecosystems interoperability. In K. Elleithy & T. Sobh (Eds.), *New Trends in Networking, Computing, E-learning, Systems Sciences, and Engineering* (pp. 513–520). Cham, Switzerland: Springer International Publishing. doi:10.1007/978-3-319-06764-3_65

Peristeras, V., & Tarabanis, K. (2006). The Connection, Communication, Consolidation, Collaboration Interoperability Framework (C4IF) for information systems interoperability. *IBIS - International Journal of Interoperability in Business Information Systems*, 1(1) 61–72.

Popplewell, K. (2014). Enterprise interoperability science base structure. In K. Mertins, F. Bénaben, R. Poler, & J. Bourrières (Eds.), *Enterprise Interoperability VI: Interoperability for Agility, Resilience and Plasticity of Collaborations* (pp. 417–427). Cham, Switzerland: Springer International Publishing. doi:10.1007/978-3-319-04948-9_35

Rezaei, R., Chiew, T., & Lee, S. (2014). A review on E-business interoperability frameworks. *Journal of Systems and Software*, 93, 199–216. doi:10.1016/j.jss.2014.02.004

Robkin, M., Weininger, S., Preciado, B., & Goldman, J. (2015). Levels of conceptual interoperability model for healthcare framework for safe medical device interoperability. In *Symposium on Product Compliance Engineering* (pp. 1–8). Piscataway, NJ: IEEE Computer Society Press. doi:10.1109/ISPCE.2015.7138703

Sharma, R., & Panigrahi, P. (2015). Developing a roadmap for planning and implementation of interoperability capability in e-government. *Transforming Government: People, Process and Policy*, 9(4), 426–447.

Stamper, R., Liu, K., Hafkamp, M., & Ades, Y. (2000). Understanding the roles of signs and norms in organizations—a semiotic approach to information systems design. *Behaviour & Information Technology*, 19(1), 15–27. doi:10.1080/014492900118768

Tran, H., Zdun, U., Oberortner, E., Mulo, E., & Dustdar, S. (2012). Compliance in service-oriented architectures: A model-driven and view-based approach. *Information and Software Technology*, 54(6), 531–552. doi:10.1016/j.infsof.2012.01.001

Wang, W., Tolk, A., & Wang, W. (2009). The levels of conceptual interoperability model: Applying systems engineering principles to M&S. In G. Wainer, C. Shaffer, R. McGraw, & M. Chinni (Eds.), *Spring Simulation Multiconference* (Article no. 168). San Diego, CA: Society for Computer Simulation International.

Zhang, Z., Wu, C., & Cheung, D. (2013). A survey on cloud interoperability: Taxonomies, standards, and practice. *Performance Evaluation Review*, 40(4), 13–22. doi:10.1145/2479942.2479945

ADDITIONAL READING

Adriansyah, A., van Dongen, B., & van der Aalst, W. (2010). Towards robust conformance checking. In M. Muehlen & J. Su (Eds.), *Business Process Management Workshops* (pp. 122–133). Berlin, Germany: Springer.

Jardim-Goncalves, R., Agostinho, C., & Steiger-Garcia, A. (2012). A reference model for sustainable interoperability in networked enterprises: Towards the foundation of EI science base. *International Journal of Computer Integrated Manufacturing*, 25(10), 855–873. doi:10.1080/0951192X.2011.653831

Jardim-Goncalves, R., Grilo, A., Agostinho, C., Lampathaki, F., & Charalabidis, Y. (2013). Systematisation of interoperability body of knowledge: The foundation for enterprise interoperability as a science. *Enterprise Information Systems*, 7(1), 7–32. doi:10.1080/17517575.2012.684401

Kokash, N., & Arbab, F. (2009). Formal behavioral modeling and compliance analysis for service-oriented systems. In F. Boer, M. Bonsangue, & E. Madelaine (Eds.), *Formal Methods for Components and Objects* (pp. 21–41). Berlin, Heidelberg: Springer-Verlag; doi:10.1007/978-3-642-04167-9_2

Loutas, N., Kamateri, E., Bosi, F., & Tarabanis, K. (2011). Cloud computing interoperability: the state of play. In C. Lambrinoudakis, P. Rizomiliotis, & T. Wlodarczyk (Eds.) *International Conference on Cloud Computing Technology and Science* (pp. 752–757). IEEE Computer Society Press. doi:10.1109/CloudCom.2011.116

Naudet, Y., Latour, T., Guedria, W., & Chen, D. (2010). Towards a systemic formalisation of interoperability. *Computers in Industry*, 61(2), 176–185. doi:10.1016/j.compind.2009.10.014

KEY TERMS AND DEFINITIONS

Choreography: Contract between two or more systems, which establishes how they cooperate to achieve some common goal.

Compliance: Asymmetric property between a consumer C and a provider P (C is compliant with P) that indicates that C satisfies all the requirements of P in terms of accepting requests.

Conformance: Asymmetric property between a provider P and a consumer C (P conforms to C) that indicates that P fulfills all the expectations of C in terms of the effects caused by its requests.

Consumer: A role performed by a system A in an interaction with another B , which involves making a request to B and typically waiting for a response.

Dimensions of Interoperability: Organization of an interoperability framework that defines several axes, or dimensions, of orthogonal interoperability concerns. This allows a better specification of an interoperability problem than by using just a single dimension of interoperability layers.

Distributed Interoperability: Interoperability between systems that have independent lifecycles. This means that they can evolve (to a new version) without having to change, to suspend or to stop the behavior or interface of the other. Distribution does not necessarily imply geographical dispersion.

Interoperability: The ability of a consumer C to be partially or fully compatible with a provider P . By composition, it can also refer to multilateral compatibility between several systems, interacting in the context of some choreography.

Interoperability Framework: Set of principles, assumptions, rules and guidelines to analyze, structure and classify the concepts and concerns of interoperability.

Layers of Interoperability: Organization of interoperability concepts and concerns along a single dimension, in layers of monotonically varying degree of complexity and abstraction.

Provider: A role performed by a system B in an interaction with another A , which involves waiting for a request from A , honoring it and typically sending a response to A .

Neural Networks and Their Accelerated Evolution From an Economic Analysis Perspective

N**Stelian Stancu***Bucharest University of Economic Studies, Romania***Constanța-Nicoleta Bodea***Bucharest University of Economic Studies, Romania***Oana Mădălina Popescu(Predescu)***Bucharest University of Economic Studies, Romania***Alina Neamțu(Idorași)***Bucharest University of Economic Studies, Romania*

INTRODUCTION

The expression ‘neural networks’ is a relatively new one in the history of universal science. It represents an automated system, similar to the human brain and works under its corresponding rules, leading to substantial outputs, based on the economic phenomenon analysed. ANN is a separate section of Artificial Intelligence, which in turn designates a research area in Computer Science. Neural networks, as a component of Artificial Intelligence, deal with the processing of specific data structures, learning and classification data algorithms, based on human brain operation.

From the perspective of recognition models, neural networks can be considered an extension of various conventional techniques that have been developed during several decades, namely conventional recognition model like the statistical model that is considered essential for a clear understanding of neural networks. Extensions of this topic can be found in many articles about statistical recognition models that laid the statistical foundations of neural networks: Duda and Hart (1973), Hand (1981), Devijer and Kitter (1982), Fukunaga (1990), Ripley (1994), Cheng and Titterton (1994).

BACKGROUND

The history of neural networks can be divided into five stages: the beginning of neural networks; the golden age; the quiet years, years of renewed enthusiasm showing the interaction between biological experimentation, modelling and computer stimulation, with hardware implementation, finishing with the fifth stage – permanent development.

1940-1950: The Beginning of Neural Networks

In 1943 the neuropsychologist Warren McCulloch and the mathematician Walter Pitts published the paper “A Logical Calculus of the Ideas Immanent in Nervous Activity”, laying the foundations of neural networks. The first precursors of computers were developed as true electronic brains, being supported by Konrad Zuse, who calculated ballistic trajectories using manual procedure. In 1941, in Berlin, at the German Institute for Aviation Research, Z3, Konrad Zuse (1993) designed an electromechanical computer, which was the first programmable computing machine, fully automated, being used to perform statistical analysis for wings vibrations.

DOI: 10.4018/978-1-5225-2255-3.ch571

Warren McCulloch and Walter Pitts (1947) indicated a practical field for the application and recognition of spatial models by neural networks.

Another researchers like Norbert Wiener and von Neumann, showed that research on the human brain design using computers could be a very interesting thing.

In 1949, Donald Hebb wrote *The Organization of Behaviour* and showed that neuronal connection is becoming stronger as it is used, being a fundamental concept for the learning process of a network. Moreover, Hebb developed the rule that bears his name, and which is the basis for almost all of neuronal learning procedures. Hebb could not postulate this rule due to the absence of neurological research, the only able to confirm this result.

Karl Lashley (1950) argued, as a neuropsychologist, that the storing of information by the brain is designed as a distributed system. His thesis was based on experiments on rats.

1950-1960: The Golden Age of Neural Networks

The most remarkable event of that time was the building of the first neuro-computer (The Snark) by Marvin Minsky (1951), which was able to adapt automatically the weights. Snark operated successfully from a technical point of view, but never exercised extremely interesting functions of information processing.

In 1956, the Dartmouth Summer Research Project on Artificial Intelligence has provided a momentum for artificial intelligence and implicitly for neural networks. The next years, John von Neumann proposed the functioning of a simple neuron by using telegraph relays or vacuum tubes.

At the same time, in 1956, Frank Rosenblatt, a neurobiologist at Cornell, started to study about the perceptron. Rosenblatt, as the founder of neural computation, besides inventing the perceptron, was mostly interested in pattern recognition and wrote a book about neural computing, entitled *Principles of Neurodynamics*. This field was developed dur-

ing 1967-1968 through the efforts of Frank Rosenblatt and his co-workers, who designed the first successful neuro-computer, the Perceptron Mark I, which was able to recognize simple average sized numbers of about 20x20 pixels, through an image sensor. For the electromechanical part, they used 512 potentiometer engines, each of them having a variable weight. A bit later, Widrow and Hoff (1960) were distinguished by their studies made by following similar steps.

In 1958, Rosenblatt introduced a simple single-layer artificial neural network, which later will be called perceptron.

In 1959, Rosenblatt described different versions of the perceptron, and verified the perceptron convergence theorem, describing neurons layers that imitate the retina, threshold switches and a learning rule by adjusting the connection weights.

In 1960, Widrow and Hoff have developed the ADALINE models (*ADaptive LInear NEuron*) and the MADALINE models (*Multiple ADaptive LInear Elements*). MADALINE, a quick and accurate adaptive learning system, was the first neural network used to solve a real problem and is still widely used especially in air traffic control. Also, they are the authors of Widrow-Hoff rule or delta rule used in learning process.

1960-1980: The Quiet Years of Neural Networks

Steinbuch (1961) introduced technical achievements for the associative memory, which can be seen as a predecessor for today's neural associative memories. He also described concepts for neural techniques by analysing their possibilities and limitations.

In his book "Learning Machines", Nilsson (1965) has given an overview of the progress and achievements of this period of research in neural networks. He took into consideration the basic principles of self-learning and therefore were discovered the intelligent systems. Later, this assumption proved to be an overstatement, but at that time it provided a very popular area of research.

In 1969, Minsky and Papert published an accurate mathematical analysis of the perceptron, and showed the limitations of the single layer perceptron in the book *Perceptrons*. The analysis demonstrated that the perceptron model is not able to represent several important issues, such as the XOR problem and linear separability. However, research in neural computation area has continued, particularly in fields as adaptive signal processing, pattern recognition and biological modelling.

Kohonen (1972) introduced a linear association model for associative memory. The same year, a similar model was independently introduced by the neurophysiologist Anderson (1972).

Marsburg (1973) used a nonlinear neuron model, with a better performance regarding its conformity to reality.

In 1974, neural networks began to be widely used, with the introduction of the back propagation of signal algorithm (“backpropagation”). The algorithm was initially described by Paul Werbe (1974) and further developed by Rumelhart, Hinton and Williams (1986).

In fact, many of the leaders of neural networks area started to publish their studies during 1970-1980: Amari (1967), Fukushima (1980), Grossberg (1972, 1976). In these articles, numerous neural models were mathematical analysed. They were the ones who put neural networks on a firm position and paved the way for the recovery of this field.

Moreover, Grossberg studied the problem of keeping a neural network capable of learning without destroying already learned associations. Later cooperation with Gali Carpenter (1987) led to development of Adaptive Resonance Theory (ART).

1980-2000: Renewed Enthusiasm

In the early 1980s, many researchers in the field of neural computation have become bold enough to start submitting proposals for developing the neuro-computer area and therefore the applications based on neural networks.

Between 1982-1986, John Hopfield, a physicist with worldwide reputation, became interested in neural networks, and through its numerous conferences in the world, he convinced hundreds of highly qualified scientists, mathematicians and engineers to join the emerging field of neural networks.

Hopfield (1982, 1984, 1985) proposed the bi-directional link in a network, giving birth to recurrent neural networks (ANN Hopfield), which are inspired by the physics laws of magnetism. Also, in 1982, at the conference *Cooperative/Competitive Neural Networks* (US-Japan) was launched the idea of a new generation of computers based on artificial intelligence (first generation – based on electrical switches, second generation – based on transistors, third generation – based on integrated circuits and advanced programming languages, fourth generation – based on codes generators and fifth generation – based on artificial intelligence).

Kohonen (1982, 1989) developed the network theory with unsupervised training (self-organizing), describing characteristics of the self-organizing map, which later will bear his name. Although they were not widely used in technical applications, neural networks slightly regained importance.

Fukushima (1980), Fukushima, Miyake and Ito (1983) introduced the neural model called Neocognitron that can recognize handwritten characters and which was an extension of Cognitron network developed in 1975.

Through the influence of John Hopfield, who convinced many researchers about the importance of the field, and through the publication of back-propagation by Rumelhart, neural networks made a modest progress.

By 1985, the American Institute of Physics began what has become an annual meeting on Neural Networks for computers.

Parker (1985) and Le Cun (1986) rediscovered the learning method for back-propagation of the signal.

Also in 1985, John Hopfield published an article describing a way to find acceptable solu-

tions for Travelling Salesman problem by using Hopfield nets.

In 1986, Rumelhart, Hinton & Williams introduced the back-propagation of error learning procedures, as a generalization of the delta rule. The non-linear-separability problem could be solved by multilayer perceptron. Since then, research development in this field has been almost explosive.

Also in 1986 with the publication of “PDP books” (Parallel Distributed Processing, Volumes I and II), edited by Rumelhart and McClelland, this area has exploded.

In 1987, the Institute of Electrical and Electronics Engineer (IEEE) organized the first International Conference on Neural Networks held in San Diego, and there was created the International Neural Network Society which gathered over 1800 participants.

In 1988 was founded the journal of Neural Networks by INNS, followed by Neural Computation in 1989 and by the IEEE Transaction on Neural Networks in 1990.

In 1989, at the meeting based on neural networks role in the defense strategy Bernard Widrow told his audience that “they were engaged in World War IV, “World War III never happened” where the battlefields are world trade and manufacturing”.

In 1990 the US Defense Department for Small Business Innovation research program named 16 topics specific to neural networks and another 13 topics that include the possible use of neural networks. In the same year, the theory on artificial neural networks based on radial activation functions (ANN-RBF) was developed.

2000-2016: Years of Permanent Development

In 2000, the studies highlighted the overall strength of NN and of the vector support machines. After that, the utility of neural networks has significantly broadened the area. Therefore, artificial

neural networks were used for: economic forecasts (at micro and macro level), shapes detection in biological pictures, chemical problems, solving hydrological problems, bioelectrical impedance analysis for body composition assessment for old people (Hsieh, Chen Lu, Huang Chen, 2012), nuclear physics (Akkoyun Bayram, Kara and Sinan, 2013).

Today, discussions about neural networks take place everywhere, their future being closely related to hardware development. Currently, most part of the neural networks development is improving. The companies are working on three types of chips: neuro-digital, analog and optical. However, some companies are performing to create a network application appropriate to a Neural Integrated Circuit. These Neural Integrated Circuits together with the neurons from digital sphere seem to be the forefront in the near future.

Although optical chips look very promising, it seems that it takes some time to be used in commercial applications.

ARTIFICIAL NEURAL NETWORKS STRUCTURE

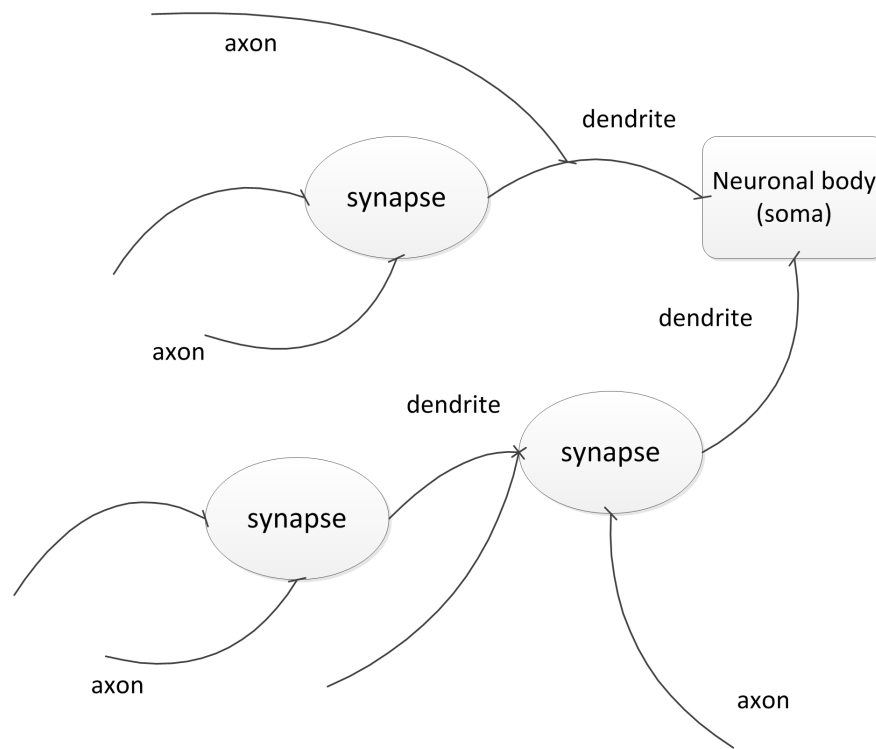
In order to understand the functioning of an artificial neural network, a preliminary study of the biological neuron is necessary.

Biological Neuron

The biological neuron consists of: *neuronal body* (*soma*), consisting of membrane and intercellular substance, its role being to transform input signals into output signals; *dendritic tree*, representing neuronal extensions of the body, designed to collect incoming signals from other neurons; *axon*, which is a prolongation of the neuronal body and has the role to transfer the output signal to other neurons through axonal arbor.

Figure 1. The biological neuron

Source: Adapted from Stancu, S., ASE, 2012.



Neurons are interconnected through links called synapses, the contact points between dendritic tree limbs of a presynaptic neuron and the axonal arbor limbs of a postsynaptic neuron.

The neuron receives information as an electrical nerve impulse through the dendrites. These impulses cause neural depolarization of the membrane, that is, changes in the potential difference between the outside and the inside of the cell.

The local potentials generated on the membrane surface are summarized and if the potential result exceeds a certain value – called threshold – then a potential action is generated and it is transmitted along the axon through the synaptic connection.

Figure 1 shows the basic structure of a biological neuron.

When the impulses reach the synapsis, they cause the release of the chemical substance that carries the information, named neurotransmitters,

which depends on received impulses and other factors. The post-synaptic membrane receives the neurotransmitters and leads to changes in ionic permeability of the membrane called depolarization. Synapses can be excitatory (if they cause positive depolarization) or inhibitory (if they cause negative depolarization).

Artificial Neuron

In a neural network, the artificial neuron is similar to the biological neuron, called simple neuron. Each neuron has m inputs and one output.

The inputs u_i , where $i = 1, m$. $i = 1, m$ are real numbers and represent signals from other neurons or from the outside world, corresponding to electrical signals from the biological model. A weight is assigned to each input u_i , related to

synaptic value of the biological neuron. The output signal is given by:

$$y = f(< w, u >) = f(w'u) = f\left(\sum_{i=1}^m w_i u_i\right) \quad (1)$$

Where:

$u = (u_1, u_2, \dots, u_m)'$ $u = (u_1, u_2, \dots, u_m)' \in R^m$ - inputs vector ;

$w = (w_1, w_2, \dots, w_m)'$ $u = (u_1, u_2, \dots, u_m)' \in R^m$ - weights vector;

$f(w'u)$ - activation function (or transfer, training, adaptation)

Let z be the variable (merge function) represented by the scalar product:

$$z = < w, u > = w'u = \sum_{i=1}^m w_i u_i \quad (2)$$

As a result, in the simplest case, the output (the activation function) is calculated as

$$y = f(z) = \begin{cases} 1 & \text{if } w'u \geq \tau \\ 0 & \text{otherwise} \end{cases}$$

where τ is the activation threshold.

NEURAL NETWORKS ARCHITECTURE

Neural networks are designed as a black box, that is a device that accepts inputs and produce outputs. Figure 3 shows a typical structure of multi-layer neural network.

The number of input neurons corresponds to the number of input variables of the neural network, as the number of output neurons corresponds to the number of output variables of the network. The number of neurons in the hidden layer depends on the particular application of the network.

INTERNAL AND EXTERNAL NEURAL NETWORKS STRUCTURE

There are two different classification levels of the structure of neural networks. The first level is known as the *external structure*. The term '*external*' describes the layout of inputs, outputs, hidden layers that compose the network and the connections between them.

The second level is the *internal structure*. This is related to the actual connections between individual nodes, both within each layer and between layers.

Figure 2. The anatomy of a neuron with single output
Source: Adapted from Stancu, S., ASE, 2012.

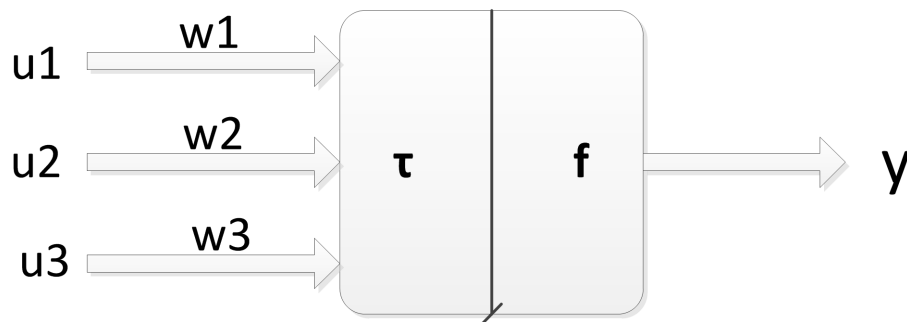
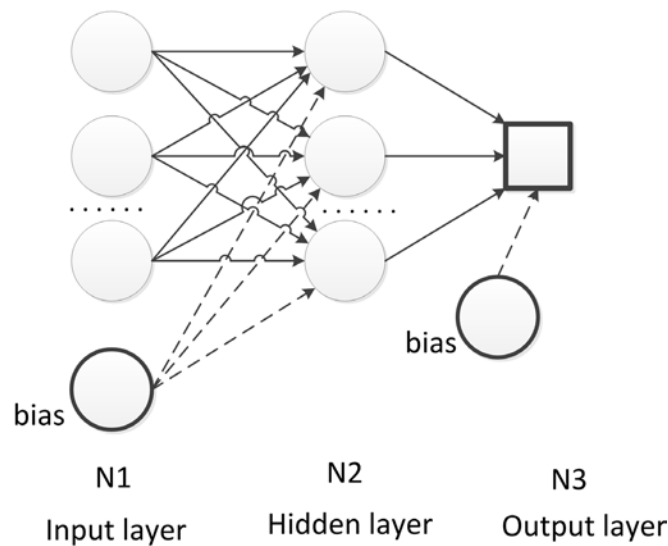


Figure 3. Multilayer ANN architecture

Source: Adapted from Stancu, S., ASE, 2012.



The External Structure of a Neural Network

There are several external structures common to a neural network, namely

- Single input and single output (SISO);
- Multiple input and single output (MISO);
- Multiple input and multiple output (MIMO)

- Single input and multiple output (SIMO).

Figures 4 and 5 show a SISO network with and without threshold.

The next network, classified by complexity is the MISO network. This type of network takes inputs from many variables and uses them to predict the value of a unique output variable.

Figure 6 shows an example of a MISO system.

Figure 4. SISO Network structure with single hidden layer and three nodes, without threshold

Source: Adapted from Stancu, S., ASE, 2012.

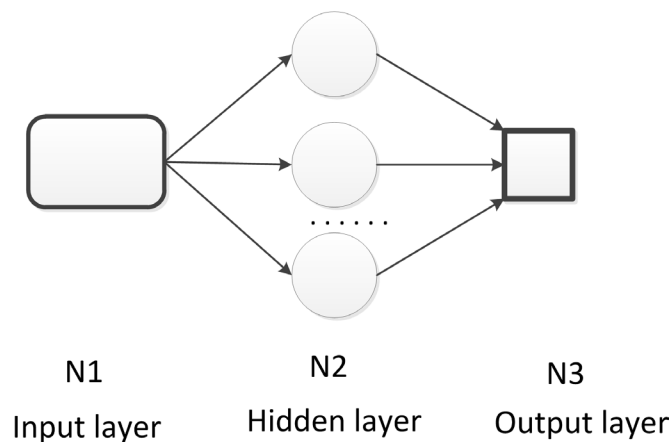


Figure 5. SISO Network structure with single hidden layer and three nodes, with threshold
Source: Adapted from Stancu, S., ASE, 2012.

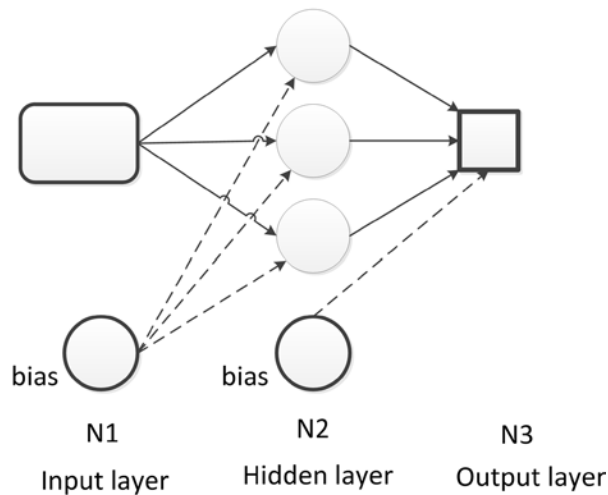
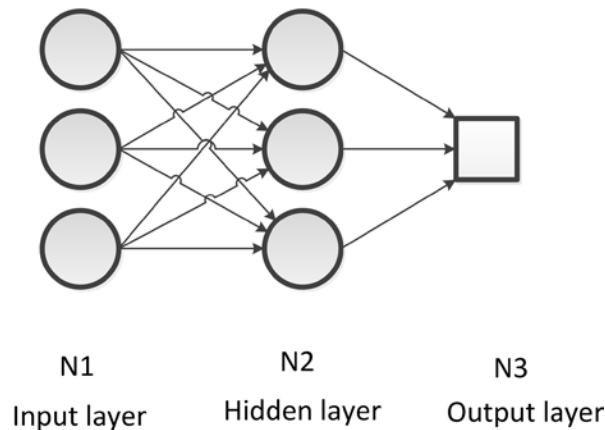


Figure 6. MISO Network structure with single hidden layer and three nodes
Source: Adapted from Stancu, S., ASE, 2012.



The most complex network architecture is the MIMO network. In this network, we will use many variables for input in order to predict values for many output variables. Figure 7 illustrates an example of MIMO network.

MIMO network is very suitable to be used for online applications, as it can predict values for several variables that may be of interest in the process, with only one input pass-through the network.

The Internal Structure of a Neural Network

Individual connections between the nodes show the internal structure of a neural network. We can connect one node to any other node in the networks. Figures 8-10 show three options for connecting nodes to each other. The layers J, K, L may be any layer in the network.

If there is at least one hidden layer, there are two main architectures of an ANN, namely:

Figure 7. MIMO Network structure with 3 input variables, 3 output variables, single hidden layer and 3 nodes

Source: Adapted from Stancu, S., ASE, 2012.

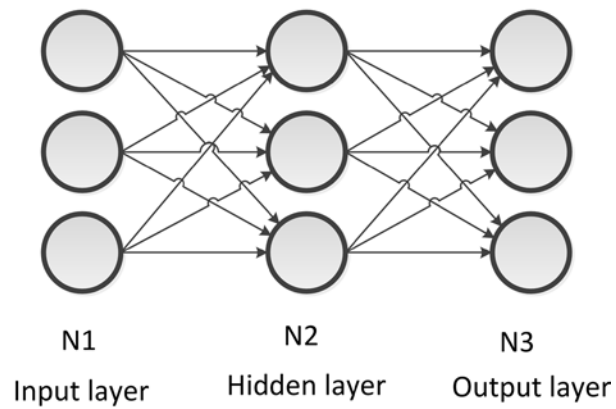


Figure 8. Connection options in a neural network. Inter-layer connection

Source: Adapted from Stancu, S., Constantin, A.M., ASE, 2014.

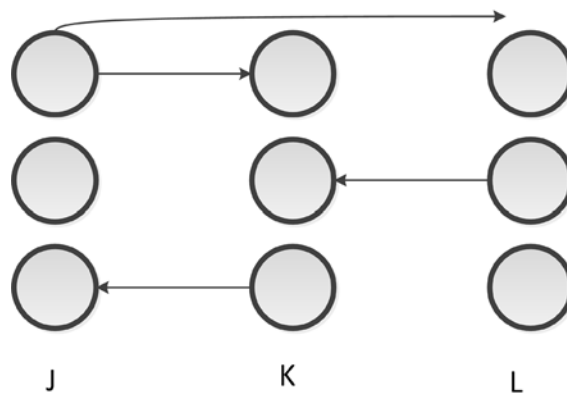


Figure 9. Connection options in a neural network. Intra-layer connection

Source: Adapted from Stancu, S., Constantin, A.M., ASE, 2014.

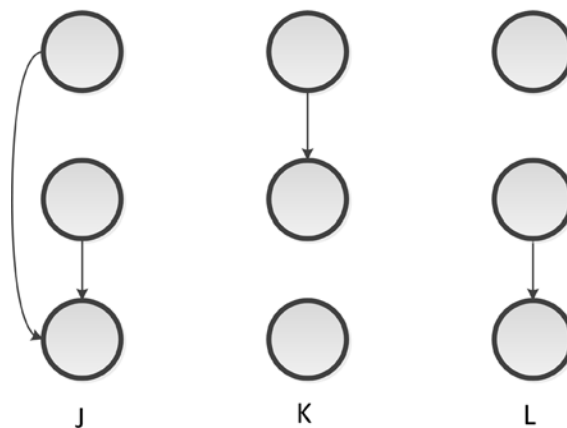
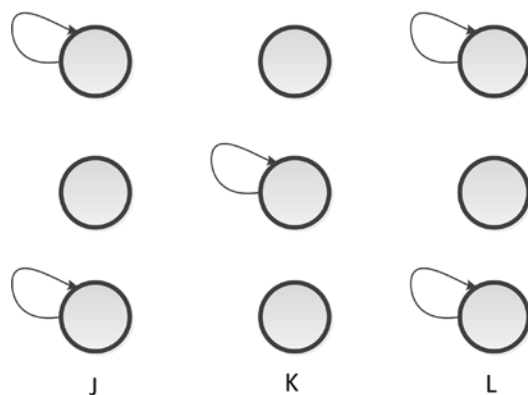


Figure 10. Connection options in a neural network. Recurrent connection
Source: Adapted from Stancu, S., Constantin, A.M., ASE, 2014.



- Feedforward networks (where the propagation arises before the signal)
- Feedback network (where the propagation arises after the signal).

The two types of architecture are shown in Figures 11 and 12.

Multilayer ANN

Multilayer networks are feedforward networks with one or many hidden layers between the input layer and the output layer. Figure 11 shows an example with three hidden layers.

In general, the more complex is the ANN – in terms of layers number and nodes number in a

layer- the higher the computing capabilities are. These multilayered networks have a greater power of representation compared to the single layer ANN, where nonlinearity is introduced through the driving function.

CLASSIFICATION OF THE NEURAL NETWORKS

The main types of neural networks are:

- Neural network with forward signal propagation (e.g. simple perceptron, multilayered perceptron);
- Radial networks;

Figure 11. Feedforward network architecture
Source: Adapted from Stancu, S., ASE, 2012.

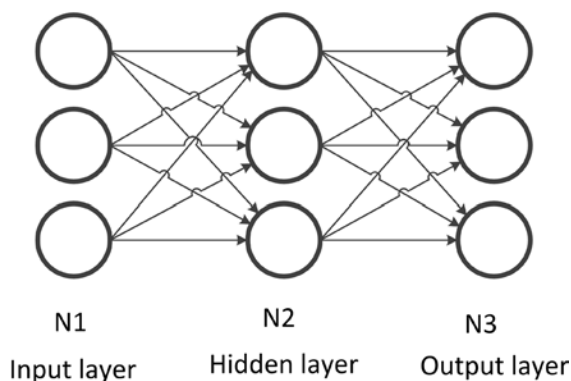


Figure 12. Feedback network architecture

Source: Adapted from Stancu, S., ASE, 2012.

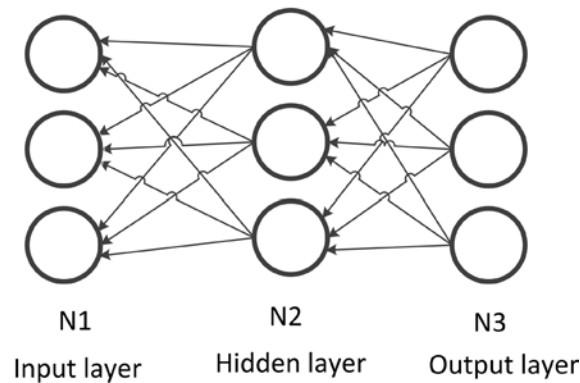
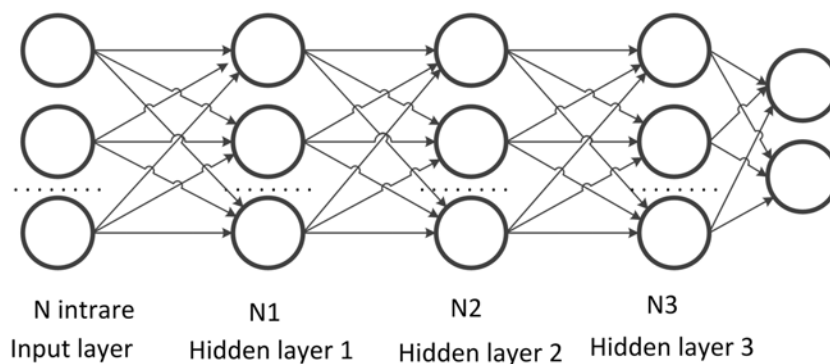


Figure 13. ANN feedforward with 3 hidden layers

Source: Adapted from Stancu, S., ASE, 2012.



- Self-organizing networks (e.g. Kohonen network);
- Recurrent networks;
- Stochastic neural networks (e.g. Boltzmann machines);
- Modular neural networks;
- Associative neural networks.

Given the characteristics of a neural network:

- Linear activation functions vs. non-linear activation functions;
- Directly operating mode (feed-forward) vs recurrently operating mode;
- Supervised learning law vs. unsupervised learning law;
- Competing layers vs. passive layers;

we can offer an overview of the achievements in neuronal computation, namely the taxonomy of neural networks.

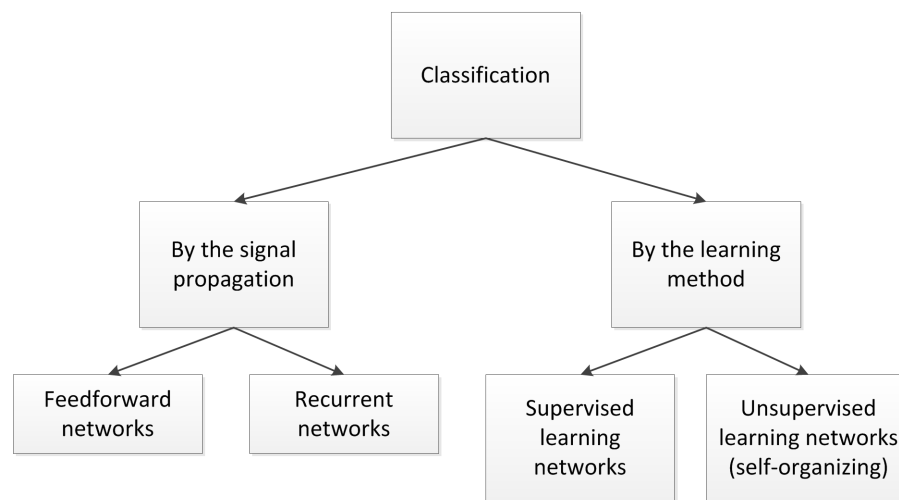
Without claiming to fully include the existing class of neural networks, the above four features allow to frame most of neural networks.

Any neural network is characterized by three elements: the model of the neuron, network architecture and the training algorithm used.

Among the neural network architectures, feed-forward networks (connections are unidirectional, no feedback loops) were the most studied. In this type of network, neurons are arranged in layers, output neurons of the lower layers being applied to the inputs of neurons in the layer immediately below.

Figure 14. Basic classification of artificial neural networks

Source: Adapted from Stancu, S., ASE, 2012.



The well-known multilayer perceptron has this type of structure, the wide range of applications where it was used being a proof of the capabilities if such a system.

The main quality of neural networks is the ability of classification. The most commonly used neural classifiers are presented in the following taxonomy, depending on the type of input (binary or continuous), on the driving mode, (supervised or unsupervised) and on the network architecture:

The blocks located at the bottom of Figure 15 indicate classical algorithms closest to the appropriate neuronal classifier.

THE USE OF NEURAL NETWORKS IN ECONOMICS

In general, neural networks produce more accurate forecasts than a corresponding linear models. However, neural networks have small disadvantages and therefore they should be considered complementary instruments.

Neural networks have been successfully applied to the macroeconomic variables. Applications in macroeconomics are still developing, standing at the border to empirical economic methods.

ANN has a broad spectrum of use, namely:

- In determining the output as an economic result, both in the case of single-output and in the case of multi-output;
- In forecasting macro or micro indicators, such as: forecasting GDP, inflation, unemployment, exchange rate;
- Symbiosis ANN fuzzy, given the role of fuzzy sets in practice
- Symbiosis ANN- genetic algorithms;
- ANN applications in game theory, in finance etc;
- ANN applications in pattern recognition theory;
- The role of ANN in the complex called Big Data from an economic perspective.

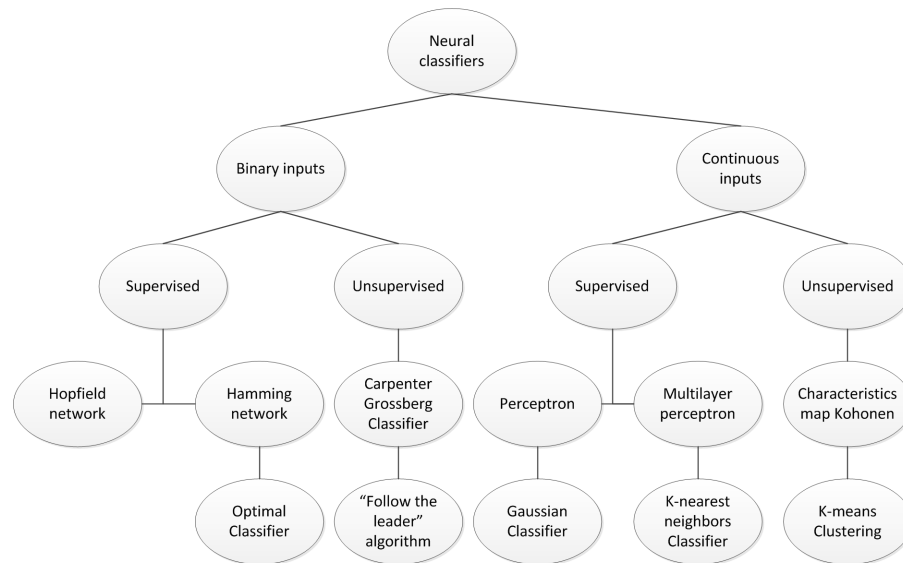
CONCLUSION

During recent years, neural networks began to be studied in all research fields and implemented in many types of expert systems, automatic etc.

The research interests into neural networks has strongly increased as neuronal computation addresses complex issues in an effective and easy

Figure 15. Neural classifiers

Source: Adapted from Stancu, S., ASE, 2012.



method to implement by applying algorithms in automated systems. Moreover, it performs better, with greater accuracy, in extracting and manipulating information in the input data set.

FUTURE RESEARCH DIRECTIONS

Increasing the scope of applying artificial neural networks, their use in all fields will help to streamline decision making, control activity and the activity of micro and macroeconomic modelling.

Currently, the main concern is finding an unique model for the optimization of decisional strategies using neural network in nonlinear modelling. The dependence on a single model for more decision-making strategies can produce significant errors from 10% to 75%, thus it is necessary to build sets of models/strategies in order to form all predictors.

The simplest neural network model is used in the process of validating the transparency of the data which will be entered in predictive-decision process. The other two strategies are designed

to create new disturbed versions of the learning sets, thus all network components will be trained on different data sets. Therefore, an unique set for each component is created by replacing the original data with uniform probability distribution.

Another important issue in forecasting is the accuracy of the neural network ensemble: the three strategies to optimize decisions on applications for scorecards and default prediction.

At the macro-level, the main issue is to find those economic models that aim to formulate and characterize the equilibrium.

REFERENCES

- Akkoyun, S., Bayram, T., Kara, S. O., & Sinan, A. (2013). An artificial neural network application on nuclear charge radii. *Journal of Physics. G, Nuclear and Particle Physics*, 40(5). doi:10.1088/0954-3899/40/5/055106
- Amari, S. (1967). A theory of adaptive pattern classifiers. *IEEE Transactions on Electronic Computers*, 16(3), 299–307.

- Anderson, D., McNeill, G., (1992). *Artificial Neural Networks Technology*. A DACS State-of-the-Art Report, Rome Laboratory RL/C3C, Griffiss AFB.
- Anderson, J. A. (1972). A simple neural network generating an interactive memory. *Mathematical Biosciences*, 14(3-4), 197–220. doi:10.1016/0025-5564(72)90075-2
- Carpenter, G. A., & Grossberg, S. (1987). ART 2: Self-organization of stable category recognition codes for analog input patterns. *Applied Optics*, 26(23), 4919–4930. doi:10.1364/AO.26.004919 PMID:20523470
- Carpenter, G. A., & Grossberg, S. (1987). A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision Graphics and Image Processing*, 37(1), 54–115. doi:10.1016/S0734-189X(87)80014-2
- Cheng, B., & Titterton, D. M. (1994). Neural networks: A review from a statistical perspective. *Statistical Science*, 9(1), 2–54. doi:10.1214/ss/1177010638
- Cun, L. (1986). Learning Processes in an Asymmetryc Treshold Network. Fogelmani-Souli.
- Devijver, P. A., & Kittler, J. (1982). *Pattern Recognition: A Statistical Approach*. Englewood Cliffs, NJ: Prentice-Hall.
- Duda, R. O., & Hart, P. E. (1973). *Pattern Classification and Scene Analysis*. New York: John Wiley.
- Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, 14(2), 179–211. doi:10.1207/s15516709cog1402_1
- Fukunaga, K. (1990). *Introduction to Statistical Pattern Recognition* (2nd ed.). San Diego, CA: Academic Press.
- Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36(4), 193–202. doi:10.1007/BF00344251 PMID:7370364
- Fukushima, K., Miyake, S., & Ito, T. (1983). Neocognitron: A neural network model for a mechanism of visual pattern recognition. *IEEE Transactions on Systems, Man, and Cybernetics*, 13(5), 826–834. doi:10.1109/TSMC.1983.6313076
- Grossberg, S. (1972). *A neural theory of punishment and avoidance*, ii: Quantitative theory. *Mathematical Biosciences*, 15(3-4), 253–285. doi:10.1016/0025-5564(72)90038-7
- Grossberg, S. (1976). *Adaptive pattern classification and universal recoding*, I: Parallel development and coding of neural feature detectors. *Biological Cybernetics*, 23(3), 121–134. doi:10.1007/BF00344744 PMID:974165
- Hand, D. J. (1981). *Discrimination and Classification*. New York: John Wiley.
- Hebb, D. O. (1949). *Organization of behavior*. New York: John Wiley & Sons.
- Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States of America*, 79(8), 2554–2558. doi:10.1073/pnas.79.8.2554 PMID:6953413
- Hopfield, J. J. (1984). Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of the National Academy of Sciences of the United States of America*, 81(10), 3088–3092. doi:10.1073/pnas.81.10.3088 PMID:6587342
- Hopfield, J. J., & Tank, D. W. (1985). Neural computation of decisions in optimization problems. *Biological Cybernetics*, 52(3), 141–152. PMID:4027280
- Hsieh, K-C., Chen, Y-J., Lu H-K., Lee, L-C., Huang, Y-C., & Chen, Y-Y. (2012). The novel application of artificial neural network on bio-electrical impedance analysis to assess the body composition in elderly. *Nutrition Journal*, 12-21. doi:10.1186/1475-2891-12-21

- Jordan, M. I. (1986). Attractor dynamics and parallelism in a connectionist sequential machine. *Proceedings of the Eighth Conference of the Cognitive Science Society*, 531–546.
- Klopf, A. H. (1988). A neuronal model of classical conditioning. *Psychobiology (Austin, Tex.)*, 16, 85–125.
- Kohonen, T. (1972). Correlation matrix memories. *IEEEtC, C-21*, 353–359.
- Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43(1), 59–69. doi:10.1007/BF00337288
- Kohonen, T. (1989). *Self-Organization and Associative Memory* (3rd ed.). Berlin: Springer-Verlag. doi:10.1007/978-3-642-88163-3
- Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE*, 78(9), 1464–1480. doi:10.1109/5.58325
- Kohonen, T. (1998). The self-organizing map. *Neurocomputing*, 21(1-3), 1–6. doi:10.1016/S0925-2312(98)00030-7
- McCulloch, W., & Pitts, W. (1943). A Logical Calculus of Ideas Immanent in Nervous Activity. *The Bulletin of Mathematical Biophysics*, 5(4), 115–133. doi:10.1007/BF02478259
- Minsky, M. L. (1954). *Neural Nets and the Brain Model Problem* (Ph.D. dissertation). Princeton Univ.
- Minsky, M.L. (1956). *Heuristic Aspects of the Artificial Intelligence Problem*. Lincoln Lab., M.I.T.
- Minsky, M.L., & Papert, S. A. (1969). *Perceptrons*. Cambridge, MA: MIT Press.
- Nilsson, N. J. (1965). *The Mathematical Foundations of Learning Machines*. New York: McGraw-Hill.
- Parker, D. (1985). *Learning Logic*. Technical Report TR-87. Cambridge, MA: Center for Computational Research in Economics and Management Science MIT.
- Ripley, B. D. (1994). Neural networks and related methods for classification. *Journal of the Royal Statistical Society. Series B. Methodological*, 56(3), 409–456.
- Rosenblatt, F. (1958). *The Perceptron*. Cornell Aeronautical Lab., Inc.
- Rosenblatt, F. (1962). *Principles of neurodynamics*. New York: Spartan.
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Vol. 1: Foundations. MIT.
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986, October). Learning representations by back-propagating errors. *Nature*, 323(6088), 533–536. doi:10.1038/323533a0
- Stancu, S. (2012). *Calcul neuronal – teorie și aplicații economice*. București: Editura ASE.
- Stancu, S., & Constantin, A. M. (2014). *Rețele neuronale artificiale. Teorie și aplicații*. București: Editura ASE.
- Steinbuch, K. (1961). Die Lernmatrix. *Kybernetik*, 1(1), 36–45. doi:10.1007/BF00293853
- von der Malsburg, C. (1973). Self-organizing of orientation sensitive cells in striate cortex. *Kybernetik*, 14(2), 85–100. doi:10.1007/BF00288907 PMID:4786750
- Werbos, P. (1974). *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences* (PhD thesis). Harvard University.
- Widrow, B., & Hoff, M. E. (1960). Adaptive switching circuits. *Proceedings WESCON*, 96–104.
- Zuse, K. (1993). *Der Computer – Mein Lebenswerk* (3rd ed.). Berlin: Springer-Verlag. doi:10.1007/978-3-662-06514-3

ADDITIONAL READING

Gregor, K., Danihelka, I., Graves, A., Wierstra, D. (2015). *DRAW: A recurrent neural network for image generation*. CoRR, abs/1502.04623, 2015.

Liu, T., Cui, Y., Yin, Q., Wang, S., Zhang, W., & Hu, G. (2016). *Generating and exploiting large-scale pseudo training data for zero pronoun resolution*, arXiv preprint arXiv:1606.01603.

Sordoni, A., Bachman, P., & Bengio, Y. (2016). *Iterative alternating neural attention for machine reading*, arXiv preprint arXiv:1606.02245.

Sukhbaatar, S., Szlam, A., Weston, J., Fergus, R. (2015). *End-to-end memory networks*. CoRR, abs/1503.08895, 2015.

Tischler, A., Ye, Z., Yuan, X., & Suleman, K. (2016). *Natural language comprehension with the epireader*, arXiv preprint arXiv:1606.02270.

KEY TERMS AND DEFINITIONS

Action Potential: The stereotypical voltage spike that constitutes an active output from a neuron. They are propagated along the axon to other neurons.

Arbor: Usually used in the context of a dendritic arbor, the tree-like structure associated with dendritic branching.

Axon: The fibre that emanates from the neuron cell body or soma and that conducts action potentials to other neurons.

Dendrite: One of the branching fibres of a neuron, which convey input information via PSPs.

Feedback Neural Networks: Those networks in which the output neurons can be interconnected with those of the input layer, thereby giving rise to an iterative process.

Feed-Forward Neural Networks (Where the Propagation Arises Before the Signal): Those networks in which the neurons of a layer are interconnected with those of the next layer whereas the neurons of the same later are not connected.

Potential Difference: The voltage difference across the cell membrane.

Presynaptic Membrane: That part of a synapse which is located on the axon terminal.

PSP: Postsynaptic Potential. The change in membrane potential brought about by activity at a synapse.

Receptor Sites: The sites on the postsynaptic membrane to which molecules of neurotransmitter bind. This binding initiates the generation of a PSP.

Recurrent Neural Networks (Feedback): That RNA forward type, where outputs are connected with their inputs.

Soma: The cell body.

Synapse: The site of physical and signal contact between neurons. On receipt of an action potential at the axon terminal of a synapse, neurotransmitter is released into the synaptic cleft and propagates to the postsynaptic membrane. There it undergoes chemical binding with receptors, which, in turn, initiates the production of a postsynaptic potential (PSP).

Optimization of Antenna Arrays and Microwave Filters Using Differential Evolution Algorithms

N

Sotirios K. Goudos

Aristotle University of Thessaloniki, Greece

INTRODUCTION

Several evolutionary algorithms (EAs) have emerged in the past decade that mimic biological entities behavior and evolution. Darwin's theory of evolution is the major inspiration source for EAs. The foundation of Darwin's theory of evolution is natural selection. The study of evolutionary algorithms began in the 1960s. Several researchers independently developed three mainstream evolutionary algorithms, namely, genetic algorithms (Goldberg, 1989), evolutionary programming (Fogel, 1995), and evolution strategies (Beyer & Schwefel, 2002). EAs are widely used for the solution of single and multi-objective optimization problems. Swarm Intelligence (SI) algorithms are also a special type of EAs. SI can be defined as the collective behavior of decentralized and self-organized swarms. SI algorithms among others include Particle Swarm Optimization (PSO) (Kennedy & Eberhart, 1995), Ant Colony Optimization (Dorigo & Stutzle, 2004), and Artificial Bee Colony (ABC) (Karaboga & Basturk, 2007).

An evolutionary algorithm that has recently gained popularity is Differential Evolution (DE) (R. Storn & Price, 1995; R. Storn & Price, 1997). DE is a population-based stochastic global optimization algorithm. DE has been used in several real world engineering problems like fuzzy logic controller design problem (Cheong & Lai, 2007), molecular sequence alignment problem (Kukkonen, Jangam, & Chakraborti, 2007), and automatic image pixel clustering (Das & Konar, 2009). The fact that the DE algorithm can handle efficiently arbitrary optimization problems has made it

DOI: 10.4018/978-1-5225-2255-3.ch572

popular for solving problems in electromagnetics. Therefore, DE has been applied successfully to a variety of constrained or unconstrained design problems in electromagnetics (Goudos, Siakavara, Samaras, Vafiadis, & Sahalos, 2011a; Goudos, Siakavara, Vafiadis, & Sahalos, 2010; Goudos, Zaharis, & Yioultsis, 2010 ; Kurup, Himdi, & Rydberg, 2003).

The purpose of this chapter is to briefly describe the DE algorithm and its variants and present their application to antenna and microwave design problems. This chapter presents results from design cases using self-adaptive DE. The chapter is supported with an adequate number of references. This chapter is subdivided into five sections. The "Background" Section presents the issues, problems and trends with DE for wireless communications. Then we briefly present the different DE algorithms. In the next Section, we describe the design cases and present the numerical results. An outline of future research directions is provided in the following Section while in the "Conclusion" Section we conclude the chapter and discuss the advantages of using a self-adaptive DE-based approach in the design and optimization of microwave systems and antennas. Finally, an "Additional Reading Section" gives a list of readings to provide the interested reader with useful sources in the field.

BACKGROUND

Differential evolution was introduced proposed by Kenneth V. Price and R. Storn in 1995. It uses

real operators for mutation and crossover, instead of the binary operators used in the first GAs. That fact has made DE suitable for solving real-valued problems. DE is a very simple but very powerful stochastic global optimizer. It has been used to solve problems in many scientific and engineering fields and proved to be a very efficient and robust technique for global optimization. In 1997, Storn established a website (Rainer Storn) to where DE source code is publically available for several popular programming languages. Since then there is an explosive growth in differential evolution research.

One of the DE advantages is that very few control parameters have to be adjusted in each algorithm run. However, the control parameters involved in DE are highly dependent on the optimization problem. Therefore, one of the major issues with DE is the correct selection of the control parameters. A basic trend in DE research is the control parameter setting, which has been extensively studied in the literature (Eiben, Hinterding, & Michalewicz, 1999). The effect of the population size was reported in (Feoktistov & Janaqi, 2004).

Another issue is the selection of the appropriate strategy for trial vector generation, which requires additional computational time using a trial-and-error search procedure. Therefore, it is not always an easy task to fine-tune the control parameters and strategy. Since finding the suitable control parameter values and strategy in such a way is often very time-consuming, there has been an increasing trend among researchers in designing new adaptive and self-adaptive DE variants. A DE strategy (jDE) that self-adapts the control parameters has been introduced in (Brest, Greiner, Boskovic, Mernik, & Zumer, 2006). This algorithm has been applied successfully to a microwave absorber design problem (Goudos, 2009) and linear array synthesis (Dib, Goudos, & Muhsen, 2010). SaDE, a DE algorithm that self-adapts both control parameters and strategy based on learning experiences from previous generations is presented in (Qin, Huang, & Suganthan, 2009).

SaDE has been applied to microwave filter design, (Goudos, Zaharis, et al., 2010), and to linear arrays synthesis (Goudos, Siakavara, Samaras, Vafiadis, & Sahalos, 2011b).

Composite DE (CoDE) (Y. Wang, Cai, & Zhang, 2011) is an adaptive DE variant, which combines three different trial vector generation strategies with three preset control parameter settings. The above combination is performed in a random way in order to generate trial vectors. The main advantage of CoDE is that it has a simple structure and thus it is very easy to be implemented in any programming language.

Most of the DE strategies or variants use the binomial crossover operator, which has been found to produce better results than the exponential crossover operator (Mezura-Montes, Velazquez-Reyes, & Coello Coello, 2006). The authors in (Guo & Yang, 2015) propose an alternative crossover operator, namely the eigenvector-based crossover. This operator utilizes the eigenvector information of the covariance matrix of the population to rotate the coordinate system. Additionally, this crossover operator can be applied to any DE variant. In (Goudos, 2015) the eigenvector-based operator is combined with CoDE to introduce a new algorithm the Composite DE with Eigenvector-Based Crossover (CODE-EIG).

The performance comparison of DE among other popular algorithms is another open issue. DE produced better results than PSO on numerical benchmark problems with low and medium dimensionality (30 and 100 dimensions) (Vesterstrom & Thomsen, 2004). However, on noisy test problems, DE was outperformed by PSO. In (Goudos, 2009) a comparative study between DE and PSO variants is presented for the design of radar absorbing materials (RAM). The number of problem dimensions was 10 and DE outperformed the PSO variants in terms of convergence speed and best values found. In (Panduro, Brizuela, Balderas, & Acosta, 2009) a comparison between DE, PSO and Genetic algorithms (GAs) for circular array design is presented. DE and PSO showed similar performances and both of them had better

performance compared to GAs. In (Goudos, et al., 2011a) the jDE outperformed the classical DE and PSO in antenna design problems. However, these results cannot lead to the general conclusion that DE outperforms PSO in all optimization problems in electromagnetics.

We cannot consider the findings about DE-variants obtained from a testbed as being universally applicable. For a specific application, it may be worthwhile to revisit certain variants of DE depending on the properties of the objective functions at hand.

Differential Evolution Algorithm

A population in DE consists of NP vectors $\bar{x}_{G,i}, i = 1, 2, \dots, NP$, where G is the generation number. The population is initialized randomly from a uniform distribution. Each D-dimensional vector represents a possible solution. The initial population evolves in each generation with the use of three operators: mutation, crossover and selection. Depending on the form of these operators several DE variants or strategies exist in the literature (R. Storn, 2008; R. Storn & Price, 1997). The choice of the best DE strategy depends on problem type (Mezura-Montes, et al., 2006). In jDE and SaDE the following four strategies are used for trial vector generation. These include *DE/rand/1bin*, *DE/rand-to-best/2/bin*, *DE/rand/2/bin*, and *DE/current-to-rand/1* (Iorio & Li, 2004). In these strategies a mutant vector $\bar{v}_{G+1,i}$ for each target vector $\bar{x}_{G,i}$ is computed by:

$$\begin{aligned} & \text{DE / rand / 1 / bin} \\ & \bar{v}_{G+1,i} = \bar{x}_{G,r_1} + F(\bar{x}_{G,r_2} - \bar{x}_{G,r_3}), \quad r_1 \neq r_2 \neq r_3 \\ & \text{DE / rand - to - best / 2 / bin} \\ & \bar{v}_{G+1,i} = \bar{x}_{G,i} + F(\bar{x}_{G,best} - \bar{x}_{G,i}) + F(\bar{x}_{G,r_1} - \bar{x}_{G,r_2}) + F(\bar{x}_{G,r_3} - \bar{x}_{G,r_4}), \quad r_1 \neq r_2 \neq r_3 \neq r_4 \\ & \text{DE / rand / 2 / bin} \\ & \bar{v}_{G+1,i} = \bar{x}_{G,r_1} + F(\bar{x}_{G,r_2} - \bar{x}_{G,r_3}) + F(\bar{x}_{G,r_4} - \bar{x}_{G,r_5}), \quad r_1 \neq r_2 \neq r_3 \neq r_4 \neq r_5 \\ & \text{DE / current - to - rand / 1 / bin} \\ & \bar{v}_{G+1,i} = \bar{x}_{G,i} + K(\bar{x}_{G,r_1} - \bar{x}_{G,i}) + F(\bar{x}_{G,r_2} - \bar{x}_{G,r_3}), \quad r_1 \neq r_2 \neq r_3 \end{aligned}$$

where r_1, r_2, r_3, r_4, r_5 are randomly chosen indices from the population, which are different from index i , F is a mutation control parameter, K a

coefficient responsible for the level of recombination that occurs between $\bar{x}_{G,i}$ and $\bar{x}_{G,r_1}$.

After mutation in each generation for every vector of the population, the crossover operator is applied to generate a trial vector $\bar{u}_{G+1,i} = (u_{G+1,1i}, u_{G+1,2i}, \dots, u_{G+1,ji}, \dots, u_{G+1,Di})$ whose coordinates are given by:

$$u_{G+1,ji} = \begin{cases} v_{G+1,ji}, & \text{if } rand_{j[0,1]} \leq CR \text{ or } j=rn(i) \\ x_{G+1,ji}, & \text{if } rand_{j[0,1]} > CR \text{ and } j \neq rn(i) \end{cases}$$

where $j = 1, 2, \dots, D$, $rand_{j[0,1]}$ is a number from a uniform random distribution within the interval $[0,1)$, $rn(i)$ a randomly chosen index from $(1, 2, \dots, D)$, and CR the crossover constant from the interval $[0,1]$. DE uses a greedy selection operator, which for minimization problems is defined by:

$$\bar{x}_{G+1,i} = \begin{cases} \bar{u}_{G+1,i}, & \text{if } f(\bar{u}_{G+1,i}) < f(\bar{x}_{G,i}) \\ \bar{x}_{G,i}, & \text{otherwise} \end{cases}$$

where $f(\bar{u}_{G+1,i})$, $f(\bar{x}_{G,i})$ are the fitness values of the trial and the old vector respectively. Therefore, the newly found trial vector $\bar{u}_{G+1,i}$ replaces the old vector $\bar{x}_{G,i}$ only when it produces a lower objective-function value than the old one. Otherwise, the old vector remains in the next generation. The stopping criterion for the DE is usually the generation number or the number of objective-function evaluations.

jDE

Storn has suggested (R. Storn & Price, 1997) that the differential evolution control parameters are adjusted in the following way: $F \in [0.5, 1]$, $CR \in [0.8, 1]$ and $NP = 10D$. In (Brest, Boskovic, Greiner, Zumer, & Maucec, 2007), a novel approach is proposed for the self-adapting

of DE control parameters. This strategy is based on *DE/rand/1/bin* scheme. Each vector is extended with its own F and CR values. The control parameters are self-adjusted in every generation for each individual. Therefore, by using the self-adaptive algorithm the user does not have to adjust the F and CR parameters while the time complexity does not increase. More details about the jDE algorithm can be found in (Brest, et al., 2007; Goudos, et al., 2011a).

DE with Strategy Adaptation (SaDE)

In the SaDE (Qin, et al., 2009) algorithm both the trial vector generation strategies and the control parameters are self-adapted according to previous learning experiences. SaDE maintains a strategy candidate pool, consisting of four commonly used strategies. Each strategy is assigned a certain probability. The sum of all probabilities is equal to one. These probabilities are initialized with a value of 0.25 and gradually adapted during evolution. An additional parameter called the learning period (LP) is introduced in (Qin, et al., 2009). This corresponds to the number of the previous generations that store the success and fail statistics. After LP generations, the probabilities of selecting different strategies are updated. Therefore, the strategies with high success rates have higher probability to be applied at the current generation. More details about the SaDE algorithm can be found in (Qin, et al., 2009).

APPLICATIONS, RESULTS, AND DISCUSSION

In this Section, we provide representative microwave and antenna design problems that are solved with the above-mentioned algorithms. First, we present an example case linear array synthesis. The design method is JDE. In the second example, we address the problem of sparse array synthesis using the SADE algorithm. The third example uses the jDE algorithm for the design of a bandpass

microwave filter under constraints for wireless communications.

Linear Array Synthesis Using jDE

Several applications of modern wireless communications systems require radiation characteristics such as higher directivity that cannot be achieved by using a single antenna. The use of antenna arrays is the obvious solution for such a case. Among others, antenna arrays are employed for several radar and wireless communications applications in space and on earth. Their advantages include the possibility of fast scanning and precise control of the radiation pattern.

Synthesis of linear antenna arrays has been extensively studied in the last decades using several analytical or stochastic methods. The linear array design problem is multimodal and therefore requires the use of an optimization method that does not easily get trapped in local minima. Common optimization goals in array synthesis are the sidelobe level suppression (while preserving the main lobe gain) and the null control to reduce interference effects. For a uniformly excited linear array, the above goals can be achieved by finding the optimum element positions and phases. We assume a $2N$ -element linear array, which is symmetrically placed along the x -axis, see Figure 1.

We examine the case of position-phase synthesis for a 32-element array. For this case, the total number of unknowns is $2N$. We assume that $d_{\min} = 0.5\lambda$, and $d_{\max} = \lambda$ as in (Kurup, et al., 2003), where $d_{\min}, d_{\max}$ is the minimum and

Figure 1. Geometry of the antenna array

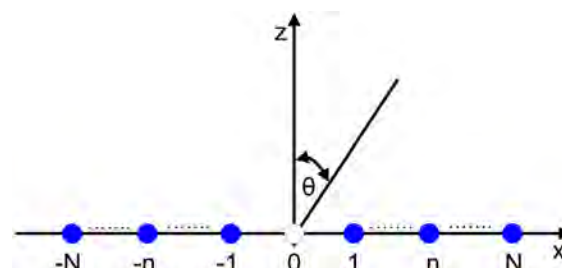


Table 1. Comparative results for 32 elements position-phase synthesis

Algorithm	Best	Worst	Mean	St. Dev.
DE	-23.06	-22.61	-22.73	0.116
jDE	-23.59	-22.78	-23.24	0.278

maximum distance between two adjacent elements respectively.

Table 1 presents the statistical results of the objective-function values for this case. The jDE algorithm clearly outperforms the classical DE. Figure 2 shows the convergence-rate plots. DE and jDE converge at similar speeds.

Figure 3 presents the pattern of the best design obtained. The maximum SLL value is -23.59dB (instead of -23.34dB in (Kurup, et al., 2003)).

Sparse Array Synthesis Using the SADE Algorithm

The second example addresses also the array design problem using a different approach. The use of a uniformly excited array has the advantage of reduced cost and complexity regarding feeding

network design. In such a case, sidelobe level suppression can be achieved by finding the optimum element positions. The constraints that the element positions must satisfy increase the difficulty of the design problem. These could include the minimum and maximum distance between two adjacent elements, the total array length, the sidelobe level suppression in specified angular intervals, and the mainlobe beamwidth set within predefined limits. SaDE and the common DE/rand/1/bin strategy are applied to a 37-element array design cases. The algorithms are run for 100 independent trials. In order to compare results with [15] the number of objective-function evaluations is set to 60,000. Therefore, the population size is set equal to 50 and the number of iterations to 1200. The total array length is set to 21.996λ . The comparative results are presented in Table 2. Although both DE algorithms obtain the same best value, the standard deviation value is smaller for SaDE. The results indicate that the DE algorithms outperform again the real GA. Figure 4 shows the convergence rate graph. It seems that DE and SADE converge at similar speeds. Figure 5 presents the pattern of the best design obtained by the DE algorithms.

Figure 2. Convergence rate for the 32-element array case

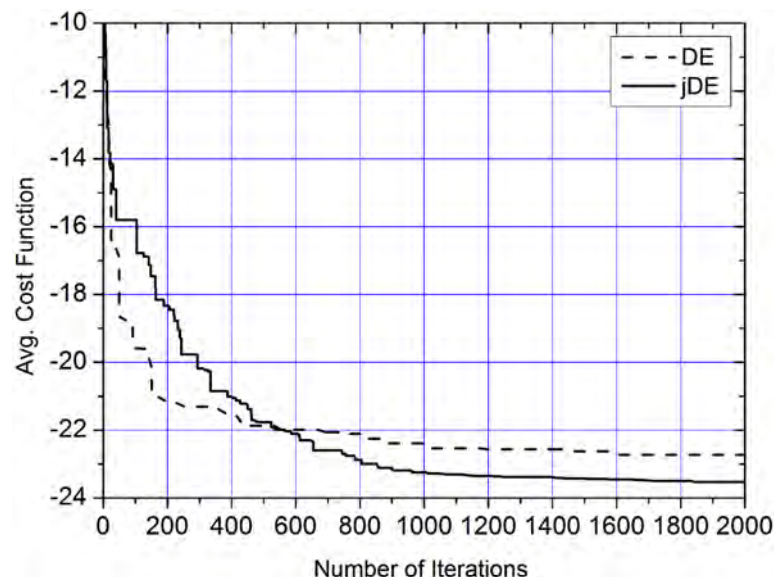


Figure 3. Best Array pattern obtained by jDE for the 32-element array case

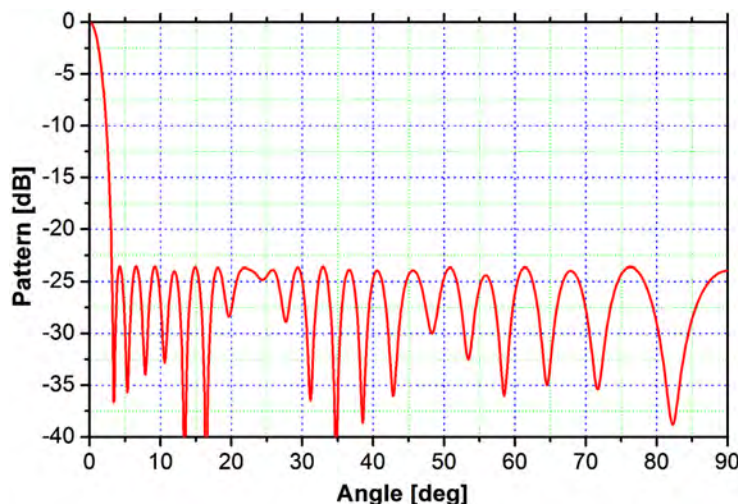


Table 2. Comparative results for the 37-element sparse array design case

Algorithm	Best	Worst	Mean	St. Dev.
GA (Chen, He, & Han, 2006)	-20.562	-20.175	-20.340	N/A
SADE	-20.942	-20.925	-20.941	0.0017
DE	-20.942	-20.916	-20.941	0.0032

Microstrip Bandpass Filter Design Using jDE

Microwave filters are among the important components of a modern wireless communication system. Several papers exist in the literature that address the filter design problem (X. H. Wang, Wang, & Chen, 2008).

The final example presents the design of a bandpass microstrip filter. We consider a substrate with dielectric constant equal to 9 and 0.66mm thickness (Hennings, Semouchkina, Baker, & Semouchkin, 2006). Such a filter design problem can be defined by two objectives subject to two constraint functions. The first objective is to maximize the $|S_{21}|$ in the passband frequency range. The second objective is to minimize the $|S_{21}|$ in the stopband frequency range. Additionally, constraints can be set for $|S_{21}|$ levels in both

the passband and the stopband. These are set to a level below -20dB in the stopbands and to a level above -3dB in the passband. The bandpass filter is designed for operation between 4.8GHz and 5GHz where is the passband. The two stopbands are defined from 4.3GHz to 4.6GHz and from 5.2GHz to 5.5GHz. The geometry of the bandpass filter is shown in Figure 6. The design parameters of the filter are 6.

Using a trial and error search we have found that 15 vectors is an adequate population size for that design problems. The total number of generations is set to 100. The best results after 10 independent trials are selected. For each run, 17 frequency sweeps are taken in the frequency range 4.3-5.5 GHz. The results for both (DE and jDE) algorithms are shown in Table 3. It is evident that the jDE outperforms the DE algorithm. Figure 7 presents the convergence rate. The DE strategies seem to converge at similar speeds.

Figure 8 shows the simulated frequency response of bandpass filter best obtained design.

FUTURE RESEARCH DIRECTIONS

The research domain of evolutionary algorithms is growing rapidly. A current and growing research trend in evolutionary algorithms is their hybridiza-

Figure 4. Convergence rate for the 37-element sparse array

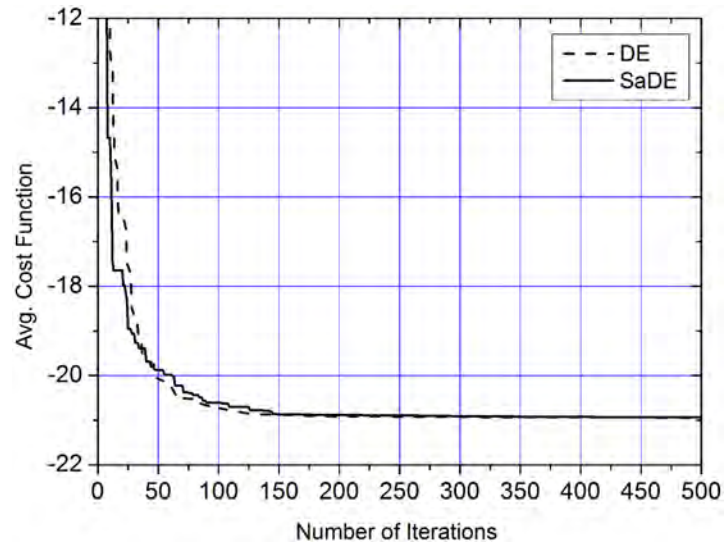


Figure 5. Best Array pattern obtained by SADE and DE for the 37-element sparse array

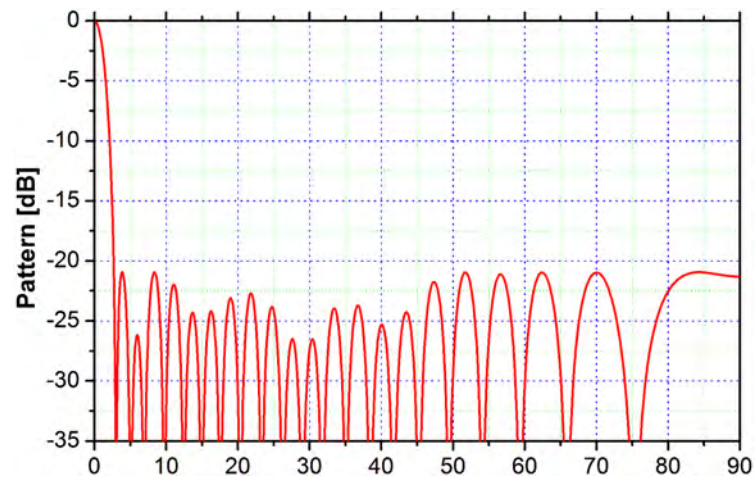


Figure 6. Tri-band filter geometry

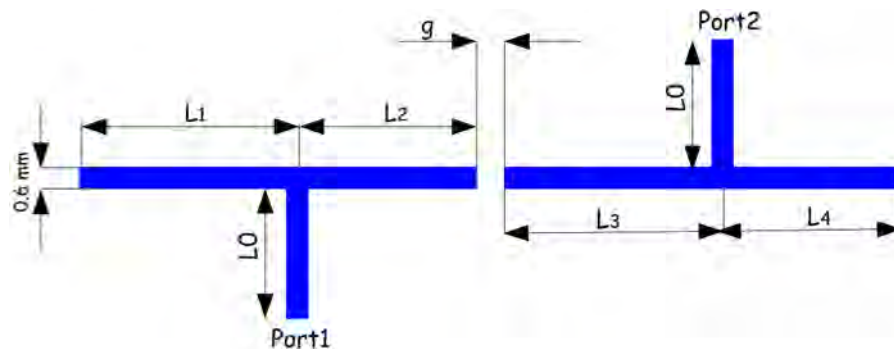


Table 3. Comparative results for the bandpass filter case

Algorithm	Best	Worst	Mean	St. Dev.
DE	-21.52	-21.35	-21.46	0.066
jDE	-21.50	-21.45	-21.48	0.015

tion with local optimizers. These algorithms are called Memetic Algorithms (MAs). MAs can be highly efficient due to this combination of global exploration and local exploitation.

The combination of DE algorithms with the eigenvector-based crossover operator is another research option. One example is the CODE-EIG, which extends the Composite DE with the eigenvector-based crossover operator. The CODE-EIG algorithm is a powerful and robust global optimizer, which selects the appropriate strategy for trial vector generation and control parameters from a predefined candidate pool. An additional control parameter that is included in CODE-EIG is the eigenvector ratio. Thus, CODE-EIG requires only the adjustment of three parameters: the population size, the number of iterations and the eigenvector ratio. Possible modifications of the CODE-EIG algorithm could be another field of study.

The application of multi-objective DE algorithms to antenna and microwave design problems introduces new challenges regarding performance and computational cost. Recently a novel Memetic Multi-objective DE algorithm has been applied to subarray design (Goudos, Gotsis, Siakavara, Vafiadis, & Sahalos, 2013). This new multi-objective DE version uses the Random Walk algorithm for local optimizer in order to search around small regions in the parameter space. Therefore, new research directions have to be explored.

CONCLUSION

In this chapter, we discussed and evaluated the application of self-adaptive DE optimizers to antenna and microwave design problems. We have presented common and state-of-the-art DE strategies found in the literature. These were the jDE and the SaDE algorithms. The DE algorithms are generally robust optimizers. In classical DE algorithms, the selection of the appropriate strategy for trial vector generation and control parameters requires additional computational time using a trial-and-error search procedure. Therefore, it is not always an easy task to fine-tune the control

Figure 7. Convergence rate for the microstrip bandpass filter design case

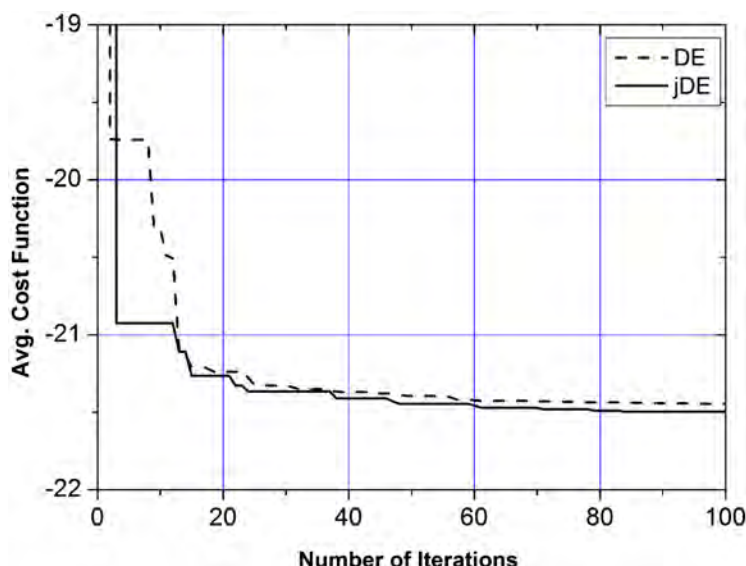
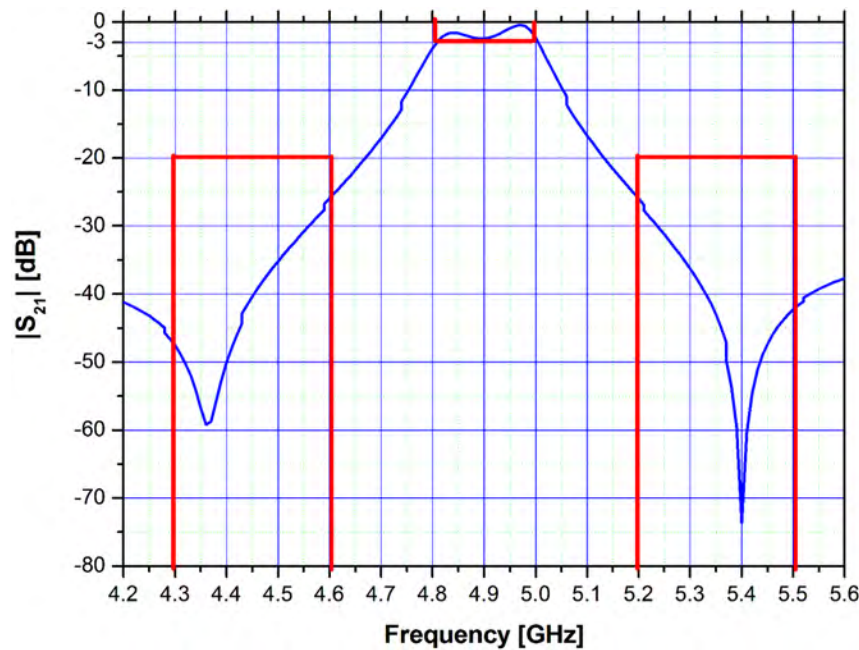


Figure 8. Simulated frequency response of the best filter obtained by jDE



parameters and strategy given also that commonly the appropriate control parameters and strategy selection are problem dependent. The SaDE and jDE advantage though, is the fact that no additional time for solving a given problem is required. SaDE and jDE require only the adjustment of two parameters: the population size and the number of iterations.

The jDE algorithm has been applied to the design of an E-shaped patch antenna and a 60-element array. In both cases the results were compared with the *DE/rand/1/bin* strategy. Compared with the classical DE strategy the jDE strategy converges faster and obtains better results.

We have also presented the design case of a tri-band microwave filter for WiFi and WiMax. The optimizer used was SaDE. The filter found exhibits low loss in the passbands and high isolation between the passbands. The proposed synthesis procedure can be used to any filter design problem that requires optimization of the geometry parameters. The correct selection of the objective function is essential for the efficient application of the SaDE algorithm.

Recent self-adaptive DE algorithms and their application to electromagnetic design will be part of our future work. DE-based optimizers can be used in cases where the computational time plays an important role. The DE algorithms combined with a numerical method can be valuable tools for constrained single or multi-objective optimization problems in wireless communications.

REFERENCES

- Beyer, H. G., & Schwefel, H. P. (2002). Evolution strategies: A comprehensive introduction. *Natural Computing*, 1(1), 3–52. doi:10.1023/A:1015059928466
- Brest, J., Boskovic, B., Greiner, S., Zumer, V., & Maucec, M. S. (2007). Performance comparison of self-adaptive and adaptive differential evolution algorithms. *Soft Computing*, 11(7), 617–629. doi:10.1007/s00500-006-0124-0

- Brest, J., Greiner, S., Boskovic, B., Mernik, M., & Zumer, V. (2006). Self-adapting control parameters in differential evolution: A comparative study on numerical benchmark problems. *IEEE Transactions on Evolutionary Computation*, 10(6), 646–657. doi:10.1109/TEVC.2006.872133
- Chen, K., He, Z., & Han, C. (2006). A modified real GA for the sparse linear array synthesis with multiple constraints. *IEEE Transactions on Antennas and Propagation*, 54(7), 2169–2173. doi:10.1109/TAP.2006.877211
- Cheong, F., & Lai, R. (2007). Designing a hierarchical fuzzy logic controller using the differential evolution approach. *Applied Soft Computing*, 7(2), 481–491. doi:10.1016/j.asoc.2006.12.001
- Das, S., & Konar, A. (2009). Automatic image pixel clustering with an improved differential evolution. *Applied Soft Computing*, 9(1), 226–236. doi:10.1016/j.asoc.2007.12.008
- Dib, N. I., Goudos, S. K., & Muhsen, H. (2010). Application of taguchis optimization method and self-adaptive differential evolution to the synthesis of linear antenna arrays. *Progress In Electromagnetics Research, PIER*, 102, 159–180. doi:10.2528/PIER09122306
- Dorigo, M., & Stutzle, T. (2004). *Ant Colony Optimization*. Cambridge, MA: The MIT Press.
- Eiben, A. E., Hinterding, R., & Michalewicz, Z. (1999). Parameter control in evolutionary algorithms. *IEEE Transactions on Evolutionary Computation*, 3(2), 124–141. doi:10.1109/4235.771166
- Feoktistov, V., & Janaqi, S. (2004). Generalization of the strategies in differential evolution. *Proc of the 18th Int Parallel and Distributed Processing Symposium*, 165–170. doi:10.1109/IPDPS.2004.1303160
- Fogel, D. B. (1995). *Evolutionary Computation: Toward a New Philosophy of Machine Intelligence*. Piscataway, NJ: IEEE Press.
- Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*. New York: Addison Wesley.
- Goudos, S. K. (2009). Design of microwave broadband absorbers using a self-adaptive differential evolution algorithm. *International Journal of RF and Microwave Computer-Aided Engineering*, 19(3), 364–372. doi:10.1002/mmce.20357
- Goudos, S. K. (2015). Shaped beam pattern synthesis of antenna arrays using composite differential evolution with eigenvector-based crossover operator. *International Journal of Antennas and Propagation*.
- Goudos, S. K., Gotsis, K. A., Siakavara, K., Vafiadis, E. E., & Sahalos, J. N. (2013). A multi-objective approach to subarrayed linear antenna arrays design based on memetic differential evolution. *IEEE Transactions on Antennas and Propagation*, 61(6), 3042–3052. doi:10.1109/TAP.2013.2254437
- Goudos, S. K., Siakavara, K., Samaras, T., Vafiadis, E. E., & Sahalos, J. N. (2011a). Self-adaptive differential evolution applied to real-valued antenna and microwave design problems. *IEEE Transactions on Antennas and Propagation*, 59(4), 1286–1298. doi:10.1109/TAP.2011.2109678
- Goudos, S. K., Siakavara, K., Samaras, T., Vafiadis, E. E., & Sahalos, J. N. (2011b). Sparse linear array synthesis with multiple constraints using differential evolution with strategy adaptation. *IEEE Antennas and Wireless Propagation Letters*, 10, 670–673. doi:10.1109/LAWP.2011.2161256

- Goudos, S. K., Siakavara, K., Vafiadis, E. E., & Sahalos, J. N. (2010). Pareto optimal yagi-uda antenna design using multi-objective differential evolution. *Progress in Electromagnetics Research*, 105, 231–251. doi:10.2528/PIER10052302
- Goudos, S. K., Zaharis, Z. D., & Yioultis, T. V. (2010). Application of a differential evolution algorithm with strategy adaptation to the design of multi-band microwave filters for wireless communications. *Progress in Electromagnetics Research*, 109, 123–137. doi:10.2528/PIER10081704
- Guo, S. M., & Yang, C. C. (2015). Enhancing differential evolution utilizing eigenvector-based crossover operator. *IEEE Transactions on Evolutionary Computation*, 19(1), 31–49. doi:10.1109/TEVC.2013.2297160
- Hennings, A., Semouchkina, E., Baker, A., & Semouchkin, G. (2006). Design optimization and implementation of bandpass filters with normally fed microstrip resonators loaded by high-permittivity dielectric. *IEEE Transactions on Microwave Theory and Techniques*, 54(3), 1253–1261. doi:10.1109/TMTT.2006.869709
- Iorio, A. W., & Li, X. (2004). Solving rotated multi-objective optimization problems using differential evolution. *Lecture Notes in Artificial Intelligence*, 3339, 861–872. doi:10.1007/978-3-540-30549-1_74
- Karaboga, D., & Basturk, B. (2007). A powerful and efficient algorithm for numerical function optimization: Artificial bee colony (ABC) algorithm. *Journal of Global Optimization*, 39(3), 459–471. doi:10.1007/s10898-007-9149-x
- Kennedy, J., & Eberhart, R. (1995). *Particle swarm optimization*. Paper presented at the IEEE International Conference on Neural Networks, Piscataway, NJ.
- Kukkonen, S., Jangam, S. R., & Chakraborti, N. (2007). Solving the Molecular Sequence Alignment Problem with Generalized Differential Evolution 3 (GDE3). *Proceedings of IEEE Symposium on Computational Intelligence in Multicriteria Decision Making*. doi:10.1109/MCDM.2007.369105
- Kurup, D. G., Himdi, M., & Rydberg, A. (2003). Synthesis of uniform amplitude unequally spaced antenna arrays using the differential evolution algorithm. *IEEE Transactions on Antennas and Propagation*, 51(9), 2210–2217. doi:10.1109/TAP.2003.816361
- Mezura-Montes, E., Velazquez-Reyes, J., & Coello Coello, C. A. (2006). *A comparative study of differential evolution variants for global optimization*. Paper presented at the GECCO 2006 - Genetic and Evolutionary Computation Conference, Seattle, WA. doi:10.1145/1143997.1144086
- Panduro, M. A., Brizuela, C. A., Balderas, L. I., & Acosta, D. A. (2009). A comparison of genetic algorithms, particle swarm optimization and the differential evolution method for the design of scannable circular antenna arrays. *Progress In Electromagnetics Research B*, (13), 171–186.
- Qin, A. K., Huang, V. L., & Suganthan, P. N. (2009). Differential evolution algorithm with strategy adaptation for global numerical optimization. *IEEE Transactions on Evolutionary Computation*, 13(2), 398–417. doi:10.1109/TEVC.2008.927706
- Storn, R. (2008). Differential evolution research - Trends and open questions. *Studies in Computational Intelligence*, 143, 1–31.
- Storn, R. (n.d.). *Differential Evolution Homepage*. Retrieved from <http://www.icsi.berkeley.edu/~storn/code.html>

Storn, R., & Price, K. (1995). *Differential Evolution—A Simple and efficient adaptive scheme for global optimization over continuous spaces*. Tech. Rep. TR-95-012. Retrieved from <http://citeseer.ist.psu.edu/article/storn95differential.html>

Storn, R., & Price, K. (1997). Differential evolution - A simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, 11(4), 341–359. doi:10.1023/A:1008202821328

Vesterstrom, J., & Thomsen, R. (2004). *A comparative study of differential evolution, particle swarm optimization, and evolutionary algorithms on numerical benchmark problems*. Paper presented at the Congress on Evolutionary Computation, Portland, OR. doi:10.1109/CEC.2004.1331139

Wang, X. H., Wang, B. Z., & Chen, K. J. (2008). Compact broadband dual-band bandpass filters using slotted ground structures. *Progress in Electromagnetics Research*, 82, 151–166. doi:10.2528/PIER08030101

Wang, Y., Cai, Z., & Zhang, Q. (2011). Differential evolution with composite trial vector generation strategies and control parameters. *IEEE Transactions on Evolutionary Computation*, 15(1), 55–66. doi:10.1109/TEVC.2010.2087271

ADDITIONAL READING

Boskovic, B., Greiner, S., Brest, J., Zamuda, A., & Zumer, V. (2008). An adaptive differential evolution algorithm with opposition-based mechanisms, applied to the tuning of a chess program. *Studies in Computational Intelligence*, 143, 287–298.

Fogel, L. J. (1962). Autonomous automata. *Industrial Research*, 4, 14–19.

Fogel, L. J., Owens, A. J., & Walsh, M. J. (1966). *Artificial Intelligence Through Simulated Evolution*. New York: John Wiley & Sons, Inc.

Goudos, S. K., Siakavara, K., & Sahalos, J. N. (2013). Novel spiral antennas design using swarm intelligence for passive UHF RFID tags. Paper presented at the 7th European Conference on Antennas and Propagation, EuCAP 2013, Gothenburg.

Haupt, R. L. (1994). Thinned arrays using genetic algorithms. *IEEE Transactions on Antennas and Propagation*, 42(7), 993–999. doi:10.1109/8.299602

Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems*. Ann Arbor: The University of Michigan Press.

Huang, V. L., Qin, A. K., & Suganthan, P. N. (2006). Self-adaptive Differential Evolution Algorithm for Constrained Real-Parameter Optimization. Paper presented at the IEEE Congress on Evolutionary Computation, 2006. CEC 2006. doi:10.1109/CEC.2006.1688285

Huang, V. L., Qin, A. K., Suganthan, P. N., & Tasgetiren, M. F. (2007). Multi-objective optimization based on self-adaptive differential evolution algorithm. Paper presented at the IEEE Congress on Evolutionary Computation, 2007 doi:10.1109/CEC.2007.4424939

Jin, N., & Rahmat-Samii, Y. (2005). Parallel particle swarm optimization and finite-difference time-domain (PSO/FDTD) algorithm for multiband and wide-band patch antenna designs. *IEEE Transactions on Antennas and Propagation*, 53(11), 3459–3468. doi:10.1109/TAP.2005.858842

Kukkonen, S., & Lampinen, J. (2005). GDE3: the third evolution step of generalized differential evolution. Paper presented at the Proceedings of the 2005 IEEE Congress on Evolutionary Computation, (CEC 2005). doi:10.1109/CEC.2005.1554717

- Lin, C., Qing, A., & Feng, Q. (2010). Synthesis of unequally spaced antenna arrays by using differential evolution. *IEEE Transactions on Antennas and Propagation*, 58(8), 2553–2561. doi:10.1109/TAP.2010.2048864
- Liu, J., & Lampinen, J. (2002). On setting the control parameter of the differential evolution method. *Proceedings of the 8th International Conference on Soft Computing (MENDEL 2002)*, 11–18.
- Mitchell, M. (1996). *An Introduction to Genetic Algorithms* (S. Printing, Ed.). MIT Press.
- Mondal, P., & Mandal, M. K. (2008). Design of dual-band bandpass filters using stub-loaded open-loop resonators. *IEEE Transactions on Microwave Theory and Techniques*, 56(1), 150–155. doi:10.1109/TMTT.2007.912204
- Qin, A. K., & Suganthan, P. N. (2005). Self-adaptive differential evolution algorithm for numerical optimization. Paper presented at the The 2005 IEEE Congress on Evolutionary Computation, 2005. doi:10.1109/CEC.2005.1554904
- Qing, A. (2003). Electromagnetic inverse scattering of multiple two-dimensional perfectly conducting objects by the differential evolution strategy. *IEEE Transactions on Antennas and Propagation*, 51(6), 1251–1262. doi:10.1109/TAP.2003.811492
- Rekanos, I. T. (2007, March 19–23). Conducting Scatterer Reconstruction Using Differential Evolution and Particle Swarm Optimization. Paper presented at the 23rd Annual Review of Progress in Applied Computational Electromagnetics (ACES), Verona, Italy. doi:10.1109/TGRS.2008.916635
- Schwefel, H. P. (1995). *Evolution and optimum seeking*. New York: John Wiley & Sons, Inc.
- Siakavara, K., & Damianidis, C. (2009). Microwave filtering in waveguides loaded with artificial single or double negative materials realized with dielectric spherical particles in resonance. *Progress in Electromagnetics Research*, 95, 103–120. doi:10.2528/PIER09061506
- Soliman, O. S., & Bui, L. T. (2008, 1 June 2008 through 6 June 2008). A self-adaptive strategy for controlling parameters in Differential Evolution, Hong Kong.
- Sotiroudias, S. P., Goudos, S. K., Gotsis, K. A., Siakavara, K., & Sahalos, J. N. (2013). Application of a composite differential evolution algorithm in optimal neural network design for propagation path-loss prediction in mobile communication systems. *IEEE Antennas and Wireless Propagation Letters*, 12, 364–367. doi:10.1109/LAWP.2013.2251994
- Wang, Z., Lai, Q., Xu, R., Yan, B., Lin, W., & Guo, Y. (2009). A millimeter-wave ultra-wideband four-way switch filter module based on novel three-line microstrip structure band-pass filters. *Progress in Electromagnetics Research*, 94, 297–309. doi:10.2528/PIER09060501
- Zhang, J., & Sanderson, A. C. (2009). JADE: Adaptive differential evolution with optional external archive. *IEEE Transactions on Evolutionary Computation*, 13(5), 945–958. doi:10.1109/TEVC.2009.2014613
- Zhang, X. Y., Chen, J. X., Xue, Q., & Li, S. M. (2007). Dual-band bandpass filters using stub-loaded resonators. *IEEE Microwave and Wireless Components Letters*, 17(8), 583–585. doi:10.1109/LMWC.2007.901768

KEY TERMS AND DEFINITIONS

Genetic Algorithms: A stochastic population-based global optimization technique that mimics the process of natural evolution.

Sidelobe Level (SLL): The ratio, usually expressed in decibels (dB), of the amplitude at the peak of the main lobe to the amplitude at the peak of a side lobe.

WiMAX: A wireless communications standard designed to provide 30 to 40 megabit-per-second data rates.

Wireless Local Area Network (WLAN): A network in which a mobile user can connect to a local area network (LAN) through a wireless (radio) connection.

QoS Architectures for the IP Network

N

Harry G. Perros

North Carolina State University, USA

INTRODUCTION

When we call someone over the Internet using a service such as Skype or Google talk, we may experience certain undesirable problems. For instance, we may not be able to hear the other person very well, or even worse, the call may be dropped. This is in contrast to using the regular telephone system where the quality of the voice is always very good. Similarly, during a conversational video call, the picture may freeze, or there may be pixilation, or the call may be dropped. The reason for these problems is that the IP packets that carry the contents of our call are not delivered on time at the destination so that they can be played out at the right time. Also, some of the packets may be lost while they are traversing the Internet. In order to eliminate these problems, the underlying IP network has to be able to provide Quality of Service (QoS) guarantees. Several schemes have been developed that enable the IP network to provide such guarantees. Of these schemes, the Multi-Protocol Label Switching (MPLS) and the Differentiated Services (DiffServ) are the most widely used. In this article, some of the salient features of MPLS and DiffServ are reviewed.

BACKGROUND

QoS is a well-understood and studied topic within the networking community. It is typically expressed in term of the following three metrics: the end-to-end delay, the jitter, and the packet loss rate. The end-to-end delay is the amount of time it takes to transfer a packet from the transmitter to the receiver, and it consists of a) the end-to-end propagation delay, b) delays induced by transmission systems and processing times inside the routers, and c) delays a packet encounters due to queueing in the buffers of the routers. Jitter refers to the variability of the inter-arrival times of the packets at the destination, and the packet loss rate is the percent of packets that are lost.

Different applications have different tolerance to these QoS metrics. Table 1 relates various common networking applications to the end-to-end delay and packet loss rate. For instance, for conversational voice and video it is important that packets should be delivered to the destination in less than 150 msec in order to maintain user satisfaction. (Studies have showed that in fact an end-to-end delay of up to 220 msec can be tolerated.) On the other hand, a packet loss rate of about

Table 1. QoS metrics for common networking services

Tolerance for packet loss	<i>Tolerant</i>	Conversational voice and video	Voicemail	Streaming audio and video	Fax
	<i>Intolerant</i>	Remote app., command and control games	e-commerce web browsing	Texting, file transfer (foreground)	File transfer (background), email
		<i>Interactive delay</i> < 1 s	<i>Responsive delay</i> ~ 1 s	<i>Timely delay</i> ~ 10 s	<i>Background delay</i> >> 10 s
		Tolerance for delay			

DOI: 10.4018/978-1-5225-2255-3.ch573

1 in 100 can be tolerated. That is, conversational voice and video type of applications are packet-loss tolerant but they have a strict end-to-end delay constraint, i.e. they are delay intolerant. On the other hand, a file transfer service is delay tolerant but packet-loss intolerant. This is because we do not expect a file to be delivered immediately, but the integrity of the file is important, and any lost packets have to be re-transmitted.

In view of the above, the question that arises is how can the network provide different QoS to different applications. In order to answer this question, let us first take a look at how the IP network routes packets. Each IP packet consists of a header and a payload, and the header contains different fields one of which is the destination IP address. When a packet arrives at an IP router, the header is examined and the destination address is used in a forwarding routing table in order to find out the next IP router to which the IP packet should be sent. This forwarding operation is carried out at each router along the path followed by the packet, until the packet reaches its destination. The forwarding routing table in each IP router is constructed using a routing protocol, such as the Open Shortest Path First (OSPF). The path that a packet follows is the shortest path in terms of the number of hops, i.e., routers. The advantage of this type of routing is that it is simple. However, since it minimizes the number of hops, it is difficult to guarantee any QoS metrics, such as end-to-end delay, jitter, and packet loss rate. For instance, the fact that the path that a packet follows has the smallest number of hops, does not necessarily mean that it has the shortest end-to-end delay. On the other hand, if all routers have approximately the same packet loss rate, then the shortest path will result to the lowest end-to-end packet loss rate.

Another problem is that a router cannot distinguish packets without an additional mechanism, such as, packet inspection or packet classification. Therefore, it cannot give packets from delay-intolerant applications a higher priority for transmission out of an output port over packets from delay-tolerant applications. (This is necessary, if we want to minimize the end-to-end delay of packets from

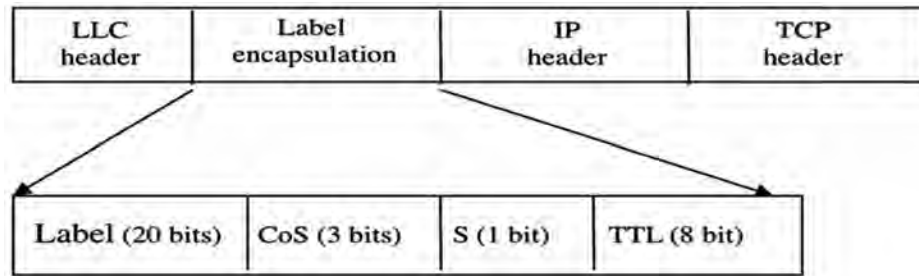
a delay-intolerant application.) In view of this, the only way that delay sensitive applications can be served satisfactorily is to under-utilize the entire network so that the queues of packets waiting for transmission in the output ports of the routers are never too long, and as a result, the delay to transmit a packet out of an output port is negligible. This is known as *over-engineering* the network. This solution is expensive since the links are under-utilized, and also it does not prevent the occurrence of transient traffic congestion. An advantage of over-engineering the network, is that when a link failure occurs, traffic can be redirected over other links without saturating them.

In order to provide QoS guarantees in an IP network without having to operate it at very low utilization, we need a scheme that can carry out *call admission control* and *packet classification*:

- **Call Admission Control:** Before a user or an application starts transmitting packets over the network, we have to make sure that the network has the necessary bandwidth to carry the new flow of packets that will be generated at the expected QoS, without affecting the QoS of the flows of packets that are currently being transmitted over the network. This procedure is known as call admission control;
- **Packet Classification:** Packets should be classified to different classes of service with different priorities, so that they are transmitted out of the output port of a router according to the priority of their class of service.

Several schemes have been developed that enable the IP network to provide QoS guarantees. Of these schemes, the Multi-Protocol Label Switching (MPLS), see IETF RFC 3031 (2001), and the Differentiated Services (DiffServ), see IETF RFC 2474 (1998), are the most widely used. Below, some of the salient features of MPLS and DiffServ are reviewed. For further details the reader is referred to Perros (2014).

Figure 1. The MPLS header



THE MULTI-PROTOCOL LABEL SWITCHING (MPLS) ARCHITECTURE

MPLS was introduced around 2000, and it is currently used extensively in IP networks to provide QoS guarantees for real-time multimedia services, such as voice and video calls, and also to interconnect clouds. MPLS is an architecture that turns the IP network into a *connection-oriented* network. In such a network, the transmission of packets between two users is done over a connection that is established prior to the transmission. This is not a logical connection, such as those established at the TCP level. Rather, a path between the two users is first calculated through the IP network, and then each router along the path reserves bandwidth for this connection on its outgoing link to the next router on the path. It also generates and stores state information regarding the connection. After the connection has been setup end-to-end, the transmission of packets can start. All the packets related to a connection follow the same path through the IP network and they arrive at the destination in the order in which they were transmitted. When the transmission of the packets is completed, the connection is torn down and the bandwidth reserved on each link for the connection is freed. Such a connection is also known as a *virtual circuit*.

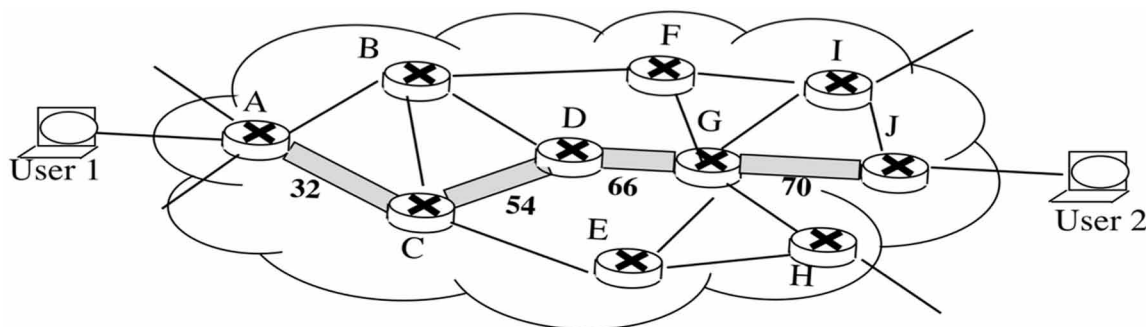
In MPLS, a connection is identified by a number that is referred to as the *label*. In IPv4, the label is encapsulated in a header that is inserted between the LLC and the IP headers, as shown in Figure 1. This header is referred to as the MPLS header.

The first field of the MPLS header is a 20-bit field that carries the label. The second field, still known as the experimental field, is a 3-bit field that is used to carry a Class-of-Service (CoS) indication. Multiple MPLS headers can be stacked in an IP packet, and for this the S field is used to indicate the end of the stack. Finally, the Time-to-Live (TTL) field contains the TTL value copied from the IP TTL field at the time the MPLS header is created after it is decremented by 1.

Each router maintains a table of labels and other relevant information for all active connections. When an IP packet arrives at the router, the label carried in the MPLS header is looked up in the table of labels to determine the next router. The IP packet is then switched to the output port of the router that connects to the next destination. In addition, the IP packets of a connection are associated with a QoS class. In MPLS, there are several different QoS classes, and each class has its own scheduling priority for transmission out of an output port of a router. The QoS class of a packet is stored in the table of labels together with the packet's label at the time the connection is setup or it may be carried in the CoS field of the MPLS header. Such a classification scheme allows packets from a delay-sensitive application to be given higher priority for transmission over packets belonging to a delay-insensitive application.

In MPLS, a connection is known as a Label Switched Path (LSP) and an MPLS-enabled router is referred to as a Label Switching Router (LSR). An example of an MPLS connection is shown in Figure 2. Each connection is associated with an

Figure 2. An example of a connection



ingress LSR and an egress LSR. For instance in Figure 2, LSR A is the ingress LSR and LSR J is the egress LSR.

As mentioned above, a connection is identified by a number which is referred to as a label. Whereas it is possible to identify a connection end-to-end by a single label, the normal practice is to identify a connection by a series of labels, with each label being valid only on a single hop. As we can see in Figure 2, the connection is routed through LSRs C, D, and G, and the labels associated with this connection are: 32, 54, 66, and 70. These labels are allocated at connection setup time. Table 2 gives the labels allocated for the connection. (For presentation purposes all the entries are listed together in a single table, although each entry belongs to a different label table. That is, the first entry belongs to the table of labels of LSR A, and so on.). Packets sent to LSR A from user 1 do not normally contain an MPLS header. The MPLS header with the label 32 is added by LSR A. It is then sent to LSR C, where the incoming label is looked up in the table of labels to find the new label and the output port number. The MPLS header is updated with the new label and the IP packet is switched to that particular output port through the switch fabric of the LSR, and eventually it is transmitted out to LSR D. The same action takes place at LSR D. Finally, when it gets to LSR J, its MPLS header is removed and the packet is sent to user 2. We note that the IP packets in the MPLS network are routed using their incoming label and not the destination IP address.

In MPLS the establishment of a connection is done using the RSVP-TE signaling protocol. Call admission control is carried out during the time the connection is setup. Specifically, a router is aware of all the connections that pass through its switch fabric, and therefore, it can decide whether to accept a new connection or not based on the amount of traffic that will be transmitted and the requested QoS. The router stops accepting new connections when it runs out of bandwidth.

The path through the IP network is computed by means separate from the MPLS architecture and it can be the same path as the one calculated using OSPF, referred to as the hop-by-hop path. It may also be a path that satisfies a QoS criterion, such as, minimization of the end-to-end delay, maximization of throughput, etc. Such a QoS criterion may not necessarily be satisfied by the hop-by-hop route, which typically minimizes the number of hops. A path that satisfies a QoS criterion is known as an *explicit route*. Explicit routing can be also used to provide load-balancing in a network, by forcing some of the traffic to

Table 2. Label entries

LSR	Incoming Label	Outgoing Label	Outgoing Port
A	-	32	to LSR C
C	32	54	to LSR D
D	54	66	to LSR G
G	66	70	to LSR J
J	70	-	to user 2

follow different paths through a network, so that the utilization of the network links is as even as possible.

THE DIFFERENTIATED SERVICE (DIFFSERV) ARCHITECTURE

DiffServ predates MPLS by a couple of years. It is widely used by network providers and also in enterprise networks, i.e., privately owned and operated networks, to provide different service to different classes of traffic. Unlike MPLS, it is not a connection-oriented architecture, and call admission control is not done on a per connection basis.

Customers of a network provider are not individual users, but rather companies and other organizations acting on behalf of their members. A customer enters into a Service Level Agreement (SLA) with a network provider, which is basically a contract that spells out how much traffic per class the customer will submit, the customer's expected QoS, and the price for the service. Call admission control is done at the aggregate level, whereby the operator has to figure out whether it can serve a new customer according to the requested SLA without affecting the SLA agreements of the existing customers.

DiffServ operates on the principle of packet classification, whereby each packet in a DiffServ-enabled IP network is classified into a traffic class. Packet classification (or marking) is done at the ingress node using different parameters, such as, source address, destination address, and port numbers. A DiffServ-enabled network may also accept traffic already marked by a user. A DiffServ class is indicated by a number known as the DiffServ Code Point (DSCP). The DSCP is stored in the 8-bit DS field in the IP header. In IPv4, this field is carried in the type of service field, and in IPv6 it is carried in the traffic class field.

Marked packets are treated differently within a DiffServ-enabled router. Typically, the queueing encountered in a router is at the output ports. In DiffServ, multiple queues are maintained at each

output port, one for each DSCP, and a packet waiting to be transmitted out joins the queue associated with its own DSCP. The queues are served according to a scheduler that guarantees that packets in higher-priority queues are transmitted out first. In addition, packets can get dropped when congestion occurs, and dropping priorities are also associated with some of the DiffServ classes of traffic. The set of controls, i.e., queue selection at an output port, scheduling priority for transmission out of the router, and packet dropping priority, is known in the DiffServ architecture as the Per-Hop Behavior (PHB).

In addition to marking and PHB, DiffServ polices the traffic submitted by the users at the ingress routers, the entry points of a DiffServ-enabled IP network. This done to make sure that the customer does not submit more traffic than agreed upon in the SLA agreement. Excess traffic may be dropped, or it may be marked as eligible to be dropped and let into the network. If congestion occurs in a router then the marked packets are dropped first.

The following four DiffServ classes have been defined:

- Expedited Forwarding (EF);
- Assured Forwarding (AF);
- Default class;
- Class Selector (CS).

The Expedited Forwarding (EF) class is used to support applications that require low delay, low jitter, low packet loss, and assured bandwidth, such as VoIP. The Assured Forwarding (AF) class was designed to support data applications with assured bandwidth requirements. That is, packets will be forwarded with a high probability as long as the class rate submitted by the user does not exceed a predefined contracted rate. Twelve different sub-classes, each with a different DSCP, have been defined within the AF class, designated as AF_xy, where x=1,2,3,4 and y=1,2,3. The scheduling priority is determined by the value of x (with 4 being the highest priority) and the dropping

priority by the value of y (with 1 being the lowest priority). The Class Selector (CS) class, is used to represent the group of 8 classes that existed prior to the introduction of DiffServ. They are designated as CS x , where $x=0,1,\dots,7$. CS0 is used to implement the default class, which is the best effort class. Because of its high use, a fair amount of bandwidth needs to be allocated to it.

An example of how different networking applications can be mapped to DiffServ classes is given in Table 3, see Cisco for further details. The meaning of most of these applications is self-evident. The locally-defined mission-critical data class is used to provide an organization with a premium class of service for a select subset of their applications that have the highest business priority for them. For example, an enterprise may have provisioned Oracle, SAP, BEA, and DLSw+, in the transactional data class. However, the majority of their revenue may come from SAP, and therefore they may want to give this application an even higher level of preference by assigning it to a dedicated class such as the locally-defined mission-critical data class. The transactional data class is a combination of transactional client-server applications and interactive messaging applications. The bulk data class is intended for applications that are relatively non-interactive, insensitive

to packet loss, and they typically span their operation over a long period of time as background occurrences. Such applications include: file transfer, e-mail, backup operations, database synchronizing or replicating operations, content distribution, and any other type of background operations. Finally, the scavenger class has a lower priority than the best effort service. Applications assigned to this class are typically entertainment-oriented. These include: peer-to-peer media-sharing applications, such as, KaZaa, Morpheus, Grokster, Napster, and iMesh, gaming applications, such as, Doom, Quake, and Unreal Tournament, and entertainment video applications.

DIFFSERV WITH MPLS

MPLS provides the capability to set up a connection using either the OSPF next hop routing (hop-by-hop route) or an explicit route that optimizes a QoS metric. Explicit routes are calculated by some means outside MPLS. Bandwidth is allocated on each router along the path of the connection, and different QoS classes of traffic can be defined. It is assumed that a policer is set up at the ingress of the MPLS domain to assure that the traffic is conformant, and a scheduler in each router makes sure that packets are transmitted out according to their priority and requested QoS. MPLS does not define QoS classes, nor does it define the policer and the scheduler. On the other hand, a scheduler along with traffic classification and policing are already available through the DiffServ architecture. Consequently, DiffServ and MPLS have been used together in order to provide a single QoS scheme, see IETF RFC 3270 (2002).

FUTURE RESEARCH DIRECTIONS

The MPLS and DiffServ architecture are mature networking solutions that have been in place for well over a decade. These architectures are used within the domain of an operator, and they do

Table 3. Mapping of applications to DiffServ classes

Application	DiffServ Class
IP routing	CS6
Voice	EF
Interactive video	AF41
Streaming video	CS4
Locally defined mission-critical data	AF31
Call signaling	AF31/CS3
Transactional data	AF21
Network management	CS2
Bulk data	AF11
Best effort	CS0
Scavenger	CS1

not provide a solution to the problem of how to guarantee end-to-end QoS over multiple domains.

The issue of multi-domain routing subject to end-to-end QoS constraints was addressed in IETF RFC 4655 (2006), where an entity referred to as the Path Computational Element (PCE) was introduced. This entity is responsible for route selection, admission control, and resource reservation within a domain. A PCE interacts with the PCEs of other domains in order to setup a multi-domain route with end-to-end QoS measures. In Geleji and Perros (2011) several algorithms were studied that permit a PCE of a domain to setup a path across multiple domains under end-to-end QoS constraints.

Quite independently to the PCE effort, there has been a new trend to centralize the management of routers and Ethernet switches. This trend comes under the name of Software-Defined Networks (SDN), with OpenFlow being a well-known SDN protocol, see OpenFlow (2011). Currently, SDN solutions work only within a single domain, but there is a strong effort to expand SDN to guarantee QoS over multiple domains, in which case the PCE solution can be easily incorporated.

Finally, the recent trend of Network-Function Virtualization (NFV), ETSI GS NFV 002 (2013), is likely to revolutionize the way networks devices and services are implemented. This initiative aims at implementing networking device types, currently running on proprietary dedicated hardware, in software that runs in virtual machines in a data center. How the end-to-end QoS will be guaranteed in a virtualized environment is an open issue.

CONCLUSION

In our daily life we use regularly various multimedia services, such as, voice and video calling (Skype, Facetime, Hangouts), video conferencing (WebEx, Telepresence), TV delivered over the Internet (IPTV), video on demand (YouTube, Netflix, Amazon, Google movies) and instant mes-

saging (texting). We also use other apps that often include some of the above multimedia services, such as, online games and assisted living. The QoS of a voice/video call transported over the IP network is critical to the operators and it can be guaranteed by using appropriate QoS schemes in the IP network, such as MPLS with DiffServ. It is expected that the advent of SDN will facilitate the provision of the necessary QoS for critical multimedia services.

REFERENCES

- Cisco. (n.d.). *Quality of service design overview*. Retrieved from http://www.cisco.com/en/US/docs/solutions/Enterprise/WAN_and_MAN/QoS_SRND/QoSIntro.html
- ETSI GS NFV 002. (2013). *Network Functions Virtualisation (NFV); Architectural Framework*. ETSI.
- Geleji, G., & Perros, H. G. (2011). H., QoS routing across multiple autonomous systems using the path computation elements architecture. *Annales des Télécommunications*, 66(5-6), 293–306. doi:10.1007/s12243-010-0206-y
- IETF RFC 2474. (1998). *Definition of the differentiated services field (DS field) in the IPv4 and IPv6 headers*. IETF.
- IETF RFC 3031. (2001). *Multiprotocol label switching architecture*. IETF.
- IETF RFC 3270. (2002). *MPLS support of differentiated services*. IETF.
- IETF RFC 4655. (2006). *A path computation element (PCE)-based architecture*. IETF.
- OpenFlow. (2011). Retrieved from <http://www.openflow.org/documents/openflow-spec-v1.1.0.pdf>
- Perros, H.G. (2014). *Networking services: QoS, Signaling, Processes*. Amazon.com.

ADDITIONAL READING

ETSI ES 282 003 V3.4.2 (2010-04), *Resource and Admission Control Sub-system (RACS); Functional Architecture*.

3. GPPTS 23.207 V10.0.0 (2011-03), *Technical Specification Group Services and System Aspects; End-to-End Quality of Service (QoS) Concept and Architecture*, (Release 10).

IETF 4761 (2007), *Virtual Private LAN Service (VPLS) Using BGP for Auto-Discovery and Signaling*.

IETF RFC 2205 (1997), *Resource ReSerVation Protocol (RSVP)*.

IETF RFC 2702 (1999) *Requirements for Traffic Engineering Over MPLS*.

IETF RFC 3037, “LDP Applicability

IETF RFC 3209 (2001), *RSVP-TE: Extensions to RSVP for LSP Tunnels*.

IETF RFC 3985 (2005), *Pseudo Wire Emulation Edge-to-Edge (PWE3) Architecture*.

IETF RFC 4142 (2005), *Protocol Extensions for Support of Diffserv-aware MPLS Traffic Engineering*.

IETF RFC 4364 (2006), *BGP/MPLS IP Virtual Private Networks (VPNs)*.

IETF RFC 4447 (2006), *Pseudowire Setup and Maintenance Using the Label Distribution Protocol (LDP)*.

IETF RFC 4448 (2006), *Encapsulation Methods for Transport of Ethernet over MPLS Networks*.

IETF RFC 4762 (2007), *Virtual Private LAN Service (VPLS) Using Label Distribution Protocol (LDP) Signaling*.

IETF RFC 4875 (2007), *Resource Reservation Protocol – Traffic Engineering (RSVP-TE) for Point-to-Multipoint TE Label Switched Paths (LSPs)*.

IETF RFC 5036 (2007), *LDP Specification*.

IETF RFC 5151 (2008), *Inter-Domain MPLS and GMPLS Traffic Engineering – Resource Reservation Protocol-Traffic Engineering (RSVP-TE) Extensions*.

Perros, H. (2005). *Connection-Oriented Networks: SONET/SDH, ATM, MPLS, and Optical Networks*. Wiley. doi:10.1002/0470016361

KEY TERMS AND DEFINITIONS

Call Admission Control: This is a scheme that assures that there is sufficient bandwidth in the network for a new flow of packets.

DiffServ: A mechanism used in the IP network to provide QoS guarantees on aggregates of flows.

Enterprise Network: A privately owned and operated network.

Explicit Route: A route through the IP network that satisfies a QoS criterion, such as, minimizations of the end-to-end delay, maximization of throughput, etc.

MPLS: A mechanism used in the IP network to provide QoS guarantees. IP packets are forwarded from one network node to the next based on labels rather than IP address.

Network-Function Virtualization (NFV): It is a concept that uses the technologies of IT virtualization in order to virtualize networking functions that may be connected together to create communication services.

Open Shortest Path First (OSPF): A protocol used to construct routes through the IP network that typically minimizes the number of hops (routers).

Propagation Delay: This the time it takes for a bit of data to travel across the network from one node or endpoint to another.

Quality of Service (QoS): It describes how well a flow of packets is served by the network. It is typically measured by different metrics, such as, end-to-end delay, jitter, packet loss rate, network availability, and bandwidth.

RSVP-TE: A signaling protocol for setting connections in an MPLS network.

Service Level Agreement (SLA): A contract between a network provider and a customer that spells out that spells out how much traffic per class the customer will submit, the customer's expected QoS, and the price for the service.

Software Defined Networks (SDN): A centralized solution for managing routers and Ethernet switches.

Voice Over IP (VoIP): A telephone service implemented over the Internet. Voice is packetized into IP packets which are transmitted to the destination over the Internet. At the destination, the voice data is recovered from the packets and played back.

N

Throughput Dependence on SNR in IEEE802.11 WLAN Systems

Ikponmwosa Oghogho

Delta State University, Abraka-Oleh Campus, Nigeria

INTRODUCTION

The widespread growth and popularity of the internet and wireless devices have placed greater demand on the need for wireless local area networks (WLANs) (Motchanov, 2010). WLANs are presently being used to enable many users share available network resources in a fast and convenient way (Obaidat et al., 2008).

WLANs popularity began when the United States Federal Communications Commission (FCC) opened up the Industrial Scientific and Medical Band (ISM) to the public (Obaidat et al., 2008). Several companies developed wireless solutions during this period but these solutions did not interoperate among themselves.

The IEEE802.11 WLAN working group addressed this issue by developing different IEEE802.11 WLAN standards which can be followed by different vendors. They defined the physical layer to specify multiple communication data rates to be selected for transmission depending on the quality of the current link (SNR observed) (Metreud, 2006).

The degree of performance enjoyed by WLAN users depend largely on the network design. Throughput is an important network parameters identified to be directly influenced by the SNR received by the end users (Metreud, 2006). Throughput depends on the selected communication data rate which is also directly dependent on the SNR observed.

This chapter presents a background study on 802.11 WLANs, their network performance metrics (throughput, SNR, Received Signal Strength Indication, etc.), interactions between TCP

throughput, losses, link adaptation and SNR in WLANs. Cross layer modelling principles and research that models throughput based on SNR only are also presented. Throughput which is largely dependent on the observed SNR is identified in this chapter as a major metric for characterizing IEEE802.11 WLANs.

BACKGROUND OF STUDY

IEEE 802.11 WLANs

The first IEEE802.11 WLAN standard which supported data rates up to 2 Mbps was released in 1997 (Obaidat et al, 2008). The standard defined both the physical layer and the medium access control (MAC) layer. Other IEEE802.11 standards such as IEEE802.11a, IEEE802.11b, IEEE802.11g, IEEE802.11n, (with higher data rates), IEEE802.11e (for better quality of service support) and IEEE802.11i (for security), IEEE802.11ac, IEEE802.11ax have been developed since then.

At the physical layer IEEE 802.11n can provide data rates up to 54Mbps in a 20MHz band with at least 100Mbps at the MAC data services access point using only mandatory features specified in the standard (Metreud, 2006). Newer standards such as IEEE802.11ac and IEEE802.11ax offer up to 1,300Mbps and 2Gbps respectively (Jamie, 2015). The popularity of the IEEE802.11 “family of standards” has continued to increase due to their convenience and the reduction in prices of their hardware. The discussion in this chapter is however limited to the base standards because

available in literature are throughput models which predict throughput directly from observed SNR only for these base standards.

Radio frequency (RF) based WLANs (which are more common than the IR-based ones) operate in the International Scientific and Medical (ISM) band which does not require any licensing from the US FCC (Obaidat et al., 2008). They use RF signals which can pass through obstacles to transmit and receive data through the air without any connecting cables thus providing freedom and flexibility unmatched by Wired LAN.

OSI Model and the IEEE802.11 Physical Layer

Open System Interconnect (OSI) was created by the International standards organization to provide a design standard and reference model to explain how different networking technologies work together and interact (Rachelle, 2001). Network protocols do not necessarily have to follow the OSI model. Seven layers are specified by the model, namely: physical, data link, network, transport, session, presentation and application layers as shown in Table 1. Table 1 also shows the TCP/IP model layers.

Each layer in the OSI model is responsible for specific functions. All layers in the model interact correctly to move data through the network. Understanding and Implementation of networks would be tedious without the OSI model because

it allows networks to be broken up into small pieces that can be easily managed. It also provides a general language which explains components and their functions.

Specifications of connector and interfaces and the cable requirements or medium are defined by the OSI model physical layer. The physical layer handles all aspects of physically moving data from one node to the other and ensures data conversion from the upper layers into 1s and 0s before they are transmitted over the media.

The IEEE802.11 standard defines two physical layer methods: (i) the frequency hopping spread spectrum and (ii) the direct sequence spread spectrum. The base IEEE802.11 DSSS physical layer supports two rates: 1Mbps and 2Mbps. Subsequent IEEE 802.11 standards support higher rates (5.5Mbps, 11Mbps, 54Mbps, 100Mbps, 1,300Mbps, 2Gbps, etc.). Table 2 shows a summary of the different Physical layer alternatives of some IEEE802.11 standard.

Performance Comparison of some IEEE802.11 WLAN Products

Table 3 shows the performance comparison of Wireless LAN products on the market published by Khanduri and Rattan (2013).

Although IEEE802.11g and IEEE802.11n allow higher data rates, the presence of an IEEE802.11b device in the network forces them to function at lower rates (Sarkar and Sowerby, 2009).

Table 1. Description of layers in the OSI model

Layer	OSI Model Layers	TCP/IP Layer	Information or Message Type	
7	Application	Process layer	Data	Mostly software
6	Presentation		Data	
5	Session	Host to host layer	Data	
4	Transport		Segment	
3	Network	Internet layer	Packet	Mostly hardware
2	Data Link	Network Access layer	Frame	
1	Physical		Bit	

(Paul, 2006; Rachelle, 2001)

Table 2. Different physical layer alternatives of some IEEE802.11 standard

Standard	Spectrum: United States (GHz)	Data Rates (Mbps)	Modulation Scheme
Base IEEE802.11	2.402-2.479	1, 2	GFSK, FHSS
	2.402-2.479	1, 2	B/QPSK, DSSS
	850-950 nm	1, 2	PPM, IR
IEEE802.11a	5.150-5.250, 5250-5350, 5725-5825	6-54	OFDM
IEEE802.11b	2.402-2.479	1, 2, 5.5, 11	CCK
IEEE802.11g	2.402-2.479	1-54	OFDM and CCK
IEEE802.11n	2.4 and 5.0	> 100	MIMO and OFDM

(Poole, 2012)

Table 3. Comparison of some wireless products on the market

Product	Compatible With	Major Disadvantages	Major Advantages
802.11a	None	Smallest range of all 802.11 standards	High bit rate in less crowded spectrum
802.11b	802.11	Bit rate too low for many emerging applications	Widely deployed, higher range, Cheapest hardware
802.11g	802.11 and 802.11b	Limited number of collocated WLANs	High bit rate in 2.4 GHz spectrum
802.11n	802.11a, 802.11b and 802.11g	Difficult to implement	Highest bit rate

IEEE802.11b WLANs system provides the largest range but fails to provide the required data speed for many emerging applications. IEEE802.11g, n, ac, ax are becoming more popular due to the increased data speed they provide.

IEEE802.11 Network Performance Characterization

Throughput is the most important network performance statistics for characterizing IEEE802.11 WLANs. Network statistics parameters related to the topic of this chapter are presented in the following sections.

Throughput

Throughput is a measurement of the average rate that data (in bits) can be sent between one user and another (Henty, 2001). Throughput is the data in bits per second that the user experiences after overheads are removed. Poor throughput

causes dissatisfaction for the users as it will take more time to complete a task such as an upload or download on the network.

Poor throughput is always associated with longer round trip times (RTTs). RTT is the time required for a signal pulse or packet to travel from a specific source to a specific destination and back again (Ali and Khuder; 2012). Poor throughput and long RTTs are very uncomfortable for the users hence network designers make efforts to avoid such conditions.

The throughput and RTT experienced by users on a network are dependent on how well the network was designed (Henty, 2001). When the SNR degrades, the throughput observed reduces appreciably and the delay experienced becomes longer. These conditions are unpleasant for WLAN users.

When calculating throughput the protocol overhead is usually not included. A Transmission control protocol (TCP) throughput rate of 1Mbps, implies that 125kilobytes of actual data

payload were sent between two network nodes (e.g. Client and Server) during one second, not including TCP, Internet Protocol, and Ethernet or IEEE802.11 headers.

Packet loss is more applicable to Universal datagram protocol (UDP) because in UDP the traffic sending node can send as much traffic as the system can handle without knowing how much of it will be lost unlike in TCP where all packets must be acknowledged and no data loss is expected to occur.

For the same network connection, the throughput can change greatly based on the protocol used for transmission, the type of data traffic sent as well as the quality (SNR) and data bandwidth of the network connection (Henty, 2001). Among these parameters that influence throughput, the SNR is the most prominent because a poor SNR can sometimes result in zero throughput.

Throughput is measured at the transport layer to accurately reflect the performance that a user will experience. This is why it is usually computed using the amount of data in the payload area of the highest protocol layer (e.g., UDP or TCP payload size) of the packets being transmitted.

Network traffic is described in terms of upstream or downstream traffic. Upstream traffic refers to data that flows away from the local computer or network toward the remote destination. Downstream traffic refers to data that flows to the user's computer or network. Traffic on most networks flows in both upstream and

downstream directions simultaneously because when data traffic is flowing in one direction, the network protocols usually send control traffic in the opposite direction.

Received Signal Strength Indication (RSSI)

In telecommunications, RSSI is a measurement of the power present in a received radio signal. Due to recent improvements, the RSSI on Windows is the RSSI reported from the Wi-Fi card. RSSI reads directly in dBm in relation to the signal strength and is usually negative. The lower the negative number the better the signal strength as illustrated in Table 4.

Signal to Noise Ratio (SNR)

When performing RF site survey, it's important to define how far an access point (AP) signal can go based on SNR ratio. The SNR at any instant in time varies with: (i) signal losses due to physical channel condition variations, (ii) variations in noise floor level and (iii) changes in received signal strength indication (RSSI) due to changes in distance or direction of the Client from the WLAN radio.

SNR in decibel (dB) is the signal level (in dBm) minus the noise level (in dBm) (Geier, 2008; Serano, 2013). It should be noted that the dB here merely represents a difference in two logarithmic

Table 4. Received signal strength indication (dBm) classification legend of WLANs

Subjective Score	Signal Strength range (dBm)	Signal Quality	Possible Coverage Worth
5	$-60 \leq RSSI \leq -20$	Excellent Signal	Excellent Coverage
4	$-75 \leq RSSI \leq -60$	Good Signal	Strong Coverage
3	$-85 \leq RSSI \leq -75$	Low Signal	Weak Coverage
2	$-90 \leq RSSI \leq -85$	Very Low Signal	Unacceptable coverage
1	$-90 \leq RSSI \leq -108$	No Signal	No Coverage

(Isiagbona and Obahiagbon, 2013; Sarkar and Sowerby, 2009)

values, such as dBm. By comparison, the dB is a dimensionless unit, used for quantifying the ratio between two values.

The measured SNR of an AP at the user device decreases as range increases because signal level is decreased by the corresponding free space loss between the user and the AP. This is also true for signals propagating from the user device to the AP. Increase in RF interference due to the presence of devices such as microwave oven, cordless phones, etc. which increases the noise level, also decreases SNR.

A higher SNR value implies a stronger RSSI in relation to the noise levels, thereby allowing higher data rates to be selected for transmission and fewer retransmissions will occur. This offers better throughput and shorter RTTs. A lower SNR requires WLAN devices to operate at lower data rates, which decreases throughput and increases delay.

Recent studies have shown that frequency diversity can lead to limitations in the accurate prediction of SNR hence effective SNR has been proposed and widely adopted as a metric for measuring wireless channel quality (Mohammed & Lili, 2016). About 20dB is recommended as the minimum SNR for defining the range boundary of each access point. SNR can be categorized into three namely: Strong signals ($SNR > 24dB$), Grey signals ($25dB > SNR > 18dB$) and Weak signals ($SNR < 19dB$) (Henty, 2001; Geier, 2008; Sarkar and Sowerby, 2009; Oghogho, 2014a).

Interactions BETWEEN TCP Throughput, Losses, Link Adaptation, and SNR in WLANs

The SNR is the predominant metric for determining when to change the data link rate (DLR) in WLANs (Metreaud, 2006). In TCP losses trigger congestion control algorithm which reduces the sending rate and ensures the retransmission of lost packets (Oghogho, 2014b).

Overall losses across the different layers in WLANs are due to (i) the varying nature of the wireless channels resulting in incorrect reception of channel symbols at the physical layers, (ii) packet collisions and inefficiency of error correction schemes or mechanisms for the received channel symbols and (iii) queuing process and buffer overflows at the IP layer (Moltchanov, 2010; Detti et al., 2011).

These overall losses have significant impact on the throughput observed for a particular application because link adaptation in WLANs determines the selected data rate used for transmission. In link adaptation, stations choose a data rate depending on underlying channel conditions which is predominantly the SNR sensed by the station. This process changes the throughput behaviour of WLANs significantly (Mahmood, 2010). Link adaptation enables stations to use higher data rates to transmit their frames as SNR increases thus throughput increases close to maximum channel capacity.

The DLR variations which are caused by channel condition variations and losses in different layers have effects on the average throughput attainable for a given average SNR and the variation in throughput observed (Metreaud, 2006).

Cross Layer Modelling Principles

The Physical, data link, network, IP, transport and application layers have different losses associated with them (Moltchanov, 2012). Researchers have suggested that cross-layer modelling principles can provide the practical methodology to evaluate: (i) the effects of wireless channel characteristics, (ii) channel adaptation mechanisms and (iii) buffering process at the IP layer on the throughput of TCP connections sharing a wireless bottleneck (Moltchanov, 2012).

Three basic steps are involved in Cross-layer modelling: (i) wireless channel modelling, (ii) cross-layer extension of the wireless channel of interest to the layer of interest (iii) performance

evaluation at the layer of interest. We can replace a mathematical model that was applied at one layer by a better model at another higher layer in order to obtain better results of the overall performance of the WLAN system (Moltchanov, 2012).

The cross layer modelling of the aggregate performance of the IEEE802.11 WLAN system is difficult to model analytically. The performance that a given application provides is a complex function of the properties and interactions between the traffic, environmental conditions, protocols, etc. which are usually very difficult and complicated to analytically model (Moltchanov, 2010).

This process can be simplified by carrying out extensive measurements of an important network metric such as throughput at a higher layer (e.g. transport layer) while varying the RSSI at a lower layer (e.g. the physical layer) so as to provide data that aggregately represents information of the WLAN system performance over different layers in real environments.

All processes involved between the lower layer and the higher layer where measurement is taken are implicitly taken into account regardless of whether they can or cannot be isolated or separately recognised (Oghogho et al 2014a). An empirical approach simplifies the model and throughput can be directly predicted from SNR calculations.

Research That Modelled Throughput Dependence on SNR Only

In this section, empirical findings of some researchers using cross layer modelling principles to model the dependence of throughput against the received signal (either RSSI or SNR) are discussed.

Henty's Models

Henty (2001) developed a linear throughput model from an empirical study which was compared with results from measured data for different environments. The models predict the throughput as a

function of SNR only. There are three parameters that must be specified:

1. $T_{\max}$, the throughput experienced in ideal circumstances (that is under maximum signal to noise ratio and without interference).
2. A or α , a “slope” or rate of drop-off as the signal to noise ratio decreases.
3. T_0 or SNR_0 , Intercept points for the model. SNR_0 is the SNR when throughput falls to zero Mbps.

Below a critical SNR represented as SNR_C , the throughput observed shows a linear drop. Henty (2000) formula for the linear throughput model is given in equation 1.

$$T = \begin{cases} T_{\max} & \text{if } SNR > SNR_C \\ A * SNR - T_0 & \text{if } SNR \leq SNR_C \end{cases} \quad (1)$$

$T_{\max}$ (a critical throughput value), A and T_0 are all vendor and application specific. They are needed to predict the throughput of the WLAN for non-ideal situations. When the SNR drops below SNR_C , the throughput falls with a slope “ A ” and drops to a throughput “ T_0 ” when the SNR drops to 0dB. The value of SNR_C is the point of intersection of the two lines of the model. From equation 1, SNR_C can be obtained from $T_{\max}$, A and T_0 values as shown in equation 2.

$$SNR_C = \frac{T_{\max} + T_0}{A} \quad (2)$$

Henty (2001) also presented an exponential throughput model. Above a critical SNR value the throughput predicted is close to a maximum ($T_{\max}$). Below the critical SNR, the throughput drops exponentially to zero. The formula used for the exponential model is given in Equation 3.

$$T_1 = T_{\max} \left[1 - \exp(-\alpha (SNR - SNR_0)) \right] \quad (3)$$

The values for $T_{\max}$, α and SNR_0 can be tuned by measurement. They are also application and vendor specific. $T_{\max}$ is the WLAN throughput under ideal conditions. α is the slope of the exponential curve and determines the rate at which the exponential curve falls off. SNR_0 is the SNR value at which the WLAN shows zero throughput.

For two users or two Clients, Henty (2001) introduced T_{loss} factors into the throughput prediction models to adjust the single user throughput to match the actual throughput in the presence of a second user. The loss factor is based on the path loss from Client one (C_1) to the Client two (C_2) and is represented by $(PL(C_1, C_2))$. A major draw-back in this method is the difficulty in estimating the loss factors under other scenarios.

Using Matlab Henty (2001) presented simplified throughput equations in Mbps for two different vendors from their data as shown in equations 4-5 for WaveLAN and 3Com respectively.

$$T = T_1 (SNR_1) - 0.000309 (T_{\max} - T_1 (SNR_2)) - 0.00384 (PL(C_1, C_2)) + 0.0918 \quad (4)$$

$$T = T_1 (SNR_1) - 0.000072 (T_{\max} - T_1 (SNR_2)) - 0.00646 (PL(C_1, C_2)) + 0.0647 \quad (5)$$

Metreud Throughput Models

Metreud (2006) presented models which predict the average throughput as functions of the average received power. The models are single user models but they consider different types of channels for IEEE802.11b and IEEE802.11g WLAN Systems.

Metreud One Tap Constant Channel Model

This model represents an ideal channel used as a baseline for comparing throughput measured with more realistic channels. The channel is a representation of the maximum throughput that can be achieved for wireless devices. The piecewise linear model for the one tap constant channel describing the maximum achievable throughput (Mbps) for IEEE802.11b and IEEE802.11g WLANs are given by equations 6 and 7 respectively.

$$S_{ave(b)}(p) = \begin{cases} 6.15 & p > -84.73 \\ 0.94p + 86.04, & -91.25 < p \leq -84.73 \\ 0 & p \leq -91.25 \end{cases} \quad (6)$$

$$S_{ave(g)}(p) = \begin{cases} 29.15 & p > -72.82 \\ 1.51p + 138.79, & -91.18 < p \leq -72.82 \\ 0 & p \leq -92.18 \end{cases} \quad (7)$$

Where p is the received power (RSSI) in dBm.

Metreud One-Tap Model A Channel

This is a one-tap Raleigh fading channel having only one propagation path hence it is not a realistic channel. The piecewise linear model for the one tap Raleigh fading channel describing the maximum achievable throughput (Mbps) for IEEE802.11b and IEEE802.11g WLANs are given by equations 8 and 9 respectively.

$$S_{ave}(p) = \begin{cases} 6.06 & p > -76.42 \\ 0.38p + 35.37, & -92.21 < p \leq -76.42 \\ 0 & p \leq -92.21 \end{cases} \quad (8)$$

$$S_{ave}(p) = \begin{cases} 26.94 & p > -65.84 \\ 1.3p + 101.30, & -89.69 < p \leq -65.84 \\ 0 & p \leq -89.69 \end{cases} \quad (9)$$

Metreud Multi-Tap Model B Channel

This is a multi-tap, two cluster multipath model used to represent an indoor residential area. The piecewise linear model for the multi-tap Model B channel describing the maximum achievable throughput (Mbps) for IEEE802.11b and IEEE802.11g WLANs are given by equations 10 and 11 respectively.

$$S_{ave}(p) = \begin{cases} 6.12 & p > -78.55 \\ 0.54p + 48.16, & -89.98 < p \leq -78.55 \\ 0 & p \leq -89.98 \end{cases} \quad (10)$$

$$S_{ave}(p) = \begin{cases} 26.61 & p > -65.49 \\ 1.12p + 99.92, & -89.27 < p \leq -65.49 \\ 0 & p \leq -89.27 \end{cases} \quad (11)$$

Metreud Multi-Tap Model C Channel

This is a multi-tap, and a slightly harsher channel than channel B having more taps and a greater RMS delay spread. The piecewise linear model for the multi-tap Model C channel describing the maximum achievable throughput (Mbps) for IEEE802.11b and IEEE802.11g WLANs are given by equations 12 and 13 respectively.

$$S_{ave}(p) = \begin{cases} 6.14 & p > -79.40 \\ 0.62p + 55.21, & -89.33 < p \leq -79.40 \\ 0 & p \leq -89.33 \end{cases} \quad (12)$$

$$S_{ave}(p) = \begin{cases} 26.71 & p > -66.33 \\ 1.16p + 103.79, & -89.32 < p \leq -66.33 \\ 0 & p \leq -89.32 \end{cases} \quad (13)$$

Only UDP traffic was used and interfering sources and the effect of humans were completely

isolated in their work hence they give higher values than what will be expected in real-life measurements.

Oghogho, Edeko and Emagbetere TCP Upstream Throughput Models

Oghogho et al (2014a) and Oghogho et al (2015a) presented IEEE802.11b WLAN throughput models which describe single and multiple users TCP upstream throughput as functions of SNR only for all SNR, strong signals (SNR>24dB), grey signals (18<SNR<25dB) and weak signals (SNR<19dB).

The single user models are shown in equations 14-17. C, C_1 and C_2 are model constants while a_1, a_2 and a_3 are model coefficients.

$$(General) TCP_{up} T_s = f(SNR) = \begin{cases} C_1 & SNR > 38dB \\ a_1 SNR + a_2 SNR^2 + a_3 SNR^3 & 39dB > SNR > 16dB \\ |a_1 SNR + a_2 SNR^2 + a_3 SNR^3 - C_2| & SNR < 17dB \end{cases} \quad (14)$$

$$(Strong) TCP_{up} T_s = f(SNR) = SNR^{a_1} \quad (15)$$

$$(Grey) TCP_{up} T_s = f(SNR) = a_1 SNR \quad (16)$$

$$(Weak) TCP_{up} T_s = f(SNR) = \begin{cases} a_1 SNR^2 + a_2 SNR^3, & 19dB > SNR > 11dB \\ 0 & SNR \leq 11dB \end{cases} \quad (17)$$

Oghogho et al (2015a) multiple users models are shown in equations 18-21.

$$(General) TCP_{up} T_M = f(SNR) = C * SNR^{a_1} \quad (18)$$

$$(Strong) TCP_{up} T_M = f(SNR) = a_1^{SNR} \quad (19)$$

$$(Grey)TCP_{up}T_M = f(SNR) = e^{\left(\frac{a_1}{SNR}\right)} \quad (20)$$

$$(Weak)TCP_{up}T_M = f(SNR) = e^{\left(\frac{a_1}{SNR}\right)} \quad (21)$$

Oghogho et al (2015b) presented empirical probability models for predicting IEEE802.11b WLAN TCP upstream throughput for single and multiple users. These models predict the probability that the TCP upstream throughput falls within a certain range based on the SNR observed.

For a single user on the network, Z values can be obtained using model equations 22-25 for all SNR, Strong signals, Grey signals and Weak signals respectively. X is the TCP upstream throughput value whose probability is to be predicted and C is the correction factor they introduced into the respective equations.

$$Z = \frac{x - 5.1038}{1.81327} + C, (0 \leq x \leq \infty) \quad (22)$$

$$Z = \frac{x - 5.8942}{0.64389} + C, (0 \leq x \leq \infty) \quad (23)$$

$$Z = \frac{x - 2.2078}{1.54209} + C, (0 \leq x \leq \infty) \quad (24)$$

$$Z = \frac{x - 0.736}{0.67621} + C, (0 \leq x \leq \infty) \quad (25)$$

For a multiple users on the network, Z values can be obtained using model equations 26-29 for all SNR, Strong signals, Grey signals and Weak signals respectively.

$$Z = \frac{x - 2.510255}{1.8888} + C, (0 \leq x \leq \infty) \quad (26)$$

$$Z = \frac{x - 2.991923}{1.7747314} + C, (0 \leq x \leq \infty) \quad (27)$$

$$Z = \frac{x - 0.7517}{1.07056} + C, (0 \leq x \leq \infty) \quad (28)$$

$$Z = \frac{x - 0.487813}{0.6485703} + C, (0 \leq x \leq \infty) \quad (29)$$

Oghogho, Edeko and Emagbetere TCP Downstream Throughput Models

Oghogho et al (2015c) presented IEEE802.11b WLAN throughput models which describe single user TCP downstream throughput as functions of SNR only. Equations 30-33 show the different model equations for General model (all SNR), Strong signal model, Grey signal model and Weak signal model respectively.

$$\begin{aligned} (General)TCP_{down}T_s = f(SNR) = \\ SNR^{a_1}, \quad SNR > 15dB \\ SNR^{a_1} + C_1, 16dB > SNR > 8dB \\ 0, \quad SNR < 9dB \end{aligned} \quad (30)$$

$$\begin{aligned} (Strong)TCP_{down}T_s = f(SNR) = \\ C_2, \quad SNR > 64dB \\ C_3, \quad 64dB > SNR > 56dB \\ SNR^{a_1}, 57dB > SNR > 24dB \end{aligned} \quad (31)$$

$$(Grey)TCP_{down}T_s = f(SNR) = SNR^{a_1} \quad (32)$$

$$(Weak)TCP_{down}T_s = f(SNR) = \begin{cases} a_1 SNR, & SNR \geq 7dB \\ 0, & SNR < 7dB \end{cases} \quad (33)$$

Oghogho et al (2016) presented IEEE802.11b WLAN throughput models which describe multiple user TCP downstream throughput as functions of SNR only. Equations 34-37 show the different model equations.

$$\begin{aligned} & (General) TCP_{down} T_M = \\ & f(SNR) = a_1 SNR + a_2 SNR^2, SNR > 9dB \end{aligned} \quad (34)$$

$$\begin{aligned} & (Strong) TCP_{down} T_M = \\ & f(SNR) = a_1 SNR + a_2 SNR^2, SNR > 24dB \end{aligned} \quad (35)$$

$$\begin{aligned} & (Grey) TCP_{down} T_M = \\ & f(SNR) = a_1 \ln(SNR), 18dB < SNR < 25dB \end{aligned} \quad (36)$$

$$\begin{aligned} & (Weak) TCP_{down} T_M = \\ & f(SNR) = a_1 SNR, SNR > 19dB \end{aligned} \quad (37)$$

Oghogho et al (2015d) presented empirical probability models for predicting IEEE802.11b WLAN TCP downstream throughput for single and multiple users. For a single user on the network, Z values can be obtained using model equations 38-41 for all SNR, Strong signals, Grey signals and Weak signals respectively. x is the TCP downstream throughput value and C is the correction factor they introduced into the respective equations.

$$Z = \frac{x - 5.2323}{1.33735} + C, (0 \leq x \leq \infty) \quad (38)$$

$$Z = \frac{x - 5.7489}{0.60438} + C, (0 \leq x \leq \infty) \quad (39)$$

$$Z = \frac{x - 3.3915}{1.37022} + C, (0 \leq x \leq \infty) \quad (40)$$

$$Z = \frac{x - 2.1143}{1.51961} + C, (0 \leq x \leq \infty) \quad (41)$$

For multiple users on the network, Z values can be obtained using model equations 42-45 for

all SNR, Strong signals, Grey signals and Weak signals respectively.

$$Z = \frac{x - 1.59385}{2.5025} + C, (0 \leq x \leq \infty) \quad (42)$$

$$Z = \frac{x - 1.48491}{2.6833} + C, (0 \leq x \leq \infty) \quad (43)$$

$$Z = \frac{x - 1.8242}{1.81778} + C, (0 \leq x \leq \infty) \quad (44)$$

$$Z = \frac{x - 1.4034}{1.55312} + C, (0 \leq x \leq \infty) \quad (45)$$

FUTURE RESEARCH DIRECTIONS

The models presented in this chapter focused on those that will enable researchers and WLAN users to estimate TCP throughput for various observed values of SNR when there is a single or there are multiple users on the network. To fully describe the behaviour of throughput there is the need to:

1. Consider models specifically developed for other IEEE 802.11 WLAN systems based on other standards (e.g. IEEE 802.11g, n, e, ac, ax, etc.) as functions of SNR only.
2. Consider models specifically developed from other protocols (e.g. UDP) as functions of SNR only.

This will help to facilitate the development of a software tool with a graphic User interface to aid accurate prediction of throughput over a wide range of SNR for different IEEE802.11 WLAN standards, protocols, environments, users, vendors, etc. The Network designer will only need to carry out an RSSI and noise floor level survey of their environments and the tool will estimate the throughput after the appropriate network parameters have been fed into the software.

CONCLUSION

In this chapter the author presented the dependence of Throughput on SNR in IEEE 802.11 WLAN systems. The author emphasized that the throughput experienced by the users is a major parameter for characterizing WLAN performance. The multiple communication data rates specified by the physical layer which vary depending on the quality of the current link (basically the SNR observed) was shown to appreciably influence the throughput experienced by the users. Finally, recent research findings which apply cross modelling principles to model the dependence of throughput on SNR was presented along with future directions in this research area.

REFERENCES

- Ali, Q., & Khuder, E. (2012). Analysis of Industrial Networks Using Several WLAN Standards. *Information Engineering*, 1(1), 8–13.
- Detti, A., Blefari-Melazzi, N., Habib, I., & Ordine, A. (2011). Per-station throughput fairness in a WLAN hot-spot with TCP traffic. *Computer Networks*, 55(8), 1820–1833. doi:10.1016/j.comnet.2011.01.004
- Geier. (2008). *Wi-Fi: Define Minimum SNR Values for Signal Coverage*. Enterprise Networking Planet. Retrieved 4/04/13 at <http://www.enterprisenetworkingplanet.com/netsp/article.php/3747656/WiFi-Define-Minimum-SNR-Values-for-Signal-Coverage.htm>
- Henty, B. E. (2001), *Throughput Measurements and Empirical Prediction Models for IEEE 802.11b Wireless LAN (WLAN) Installations* (Masters Thesis). Virginia Polytechnic Institute and State University. Retrieved 21/08/12 at <http://scholar.lib.vt.edu/theses/available/etd-08142001-172523/unrestricted/hentythesis.pdf>
- Isiagbona, J., & Obahiagbon, K. (2013). A Practical-Based Radio Channel Site Survey Approach for Optimal Deployment of a Campus Wireless LAN. *African Journal of Computing and ICT*, 6(4), 133–146.
- Jamie, L. (2015). *What is 802.11ax Wifi, and will it really deliver 10Gbps?* EXTREMETECH. Retrieved 01/09/16 at <http://www.extremetech.com/computing/184685-what-is-802-11ax-wifi-and-do-you-really-need-a-10gbps-connection-to-your-laptop>
- Khanduri, R., & Rattan, S. S. (2013). Performance Comparison Analysis Between IEEE802.11a/b/g/n Standards. *International Journal of Computers and Applications*, 78(1), 13–20. doi:10.5120/13452-1071
- Mahmood, M. H., Chang, C., Jung, D., Mao, Z., Lim, H., & Lee, H. (2010). Throughput behavior of link adaptive 802.11 DCF with MUD capable access node. *Int. J. Electron. Commun.*, 64(11), 1031–1041. doi:10.1016/j.aeeu.2009.10.007
- Metreaud, L. T. (2006). *An RF-Isolated Real-Time Multipath Test bed for Performance Analysis of WLANs* (Masters Thesis). Worcester Polytechnic Institute.
- Moltchanov, D. (2010). Performance models for wireless channels. *Computer Science Review*, 4(3), 153–184. doi:10.1016/j.cosrev.2010.04.001
- Moltchanov, D. (2012). A study of TCP performance in wireless environment using fixed-point approximation. *Computer Networks*, 56(4), 1263–1285. doi:10.1016/j.comnet.2011.11.012
- Muhammad, O. K., & Lili, Q. (2016). Accurate WiFi packet delivery rate estimation and applications. *Computer Communications, IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on*. IEEE. Doi:10.1109/INFOCOM.2016.7524394

Obaidat, M. S., Papadimitriou, G. I., & Obeidat, S. (2008). Wireless LANs. *The Handbook of Computer Networks* (vol. 2). John Wiley and Sons, Inc.

Oghogho, I., Edeko, F. O., & Emagbetere, J. E. (2014b). *Empirical Probability Models for Predicting TCP Downstream throughput in a WLAN System*. Retrieved from <http://www.istp.org.in/Spec.Issue.html>

Oghogho, I., Edeko, F. O., & Emagbetere, J. E. (2015a). Investigation on the Dependence of TCP Upstream Throughput on SNR For Single and Multiple Links in a WLAN System. *Review of Information Engineering and Applications*, 2(1), 15-32. DOI: 10.18488/journal.79/2015.2.1/79.1.15.32

Oghogho, I., Edeko, F. O., & Emagbetere, J. E. (2015b). Probability Models for Predicting TCP Upstream Throughput in an IEEE802.11b WLAN System. *Journal of Electronic and Electrical Engineering*, 12(2), 1–19.

Oghogho, I., Edeko, F. O., & Emagbetere, J. E. (2015c). *Empirical Investigation on the Dependence of TCP Downstream Throughput on SNR in an IEEE802. 11b WLAN System*. JKsUES. doi:10.1016/j.jksues.2015.06.006

Oghogho, I., Edeko, F. O., & Emagbetere, J. E. (2015d). Probability Models for Predicting TCP Upstream Throughput in an IEEE802.11b WLAN System. *Journal of Electronic and Electrical Engineering*, 12(2), 1–19.

Oghogho, I., Edeko, F. O., & Emagbetere, J. E. (2016). *Measurement and Modeling of TCP Downstream Throughput Dependence on SNR in an IEEE802.11b WLAN System*. JKsUES. doi:10.1016/j.jksues.2016.03.003

Oghogho, I., Edeko, F. O., Emagbetere, J. E., & Victor, M. (2014a). Empirical investigation on the dependence of TCP upstream throughput on SNR in an IEEE802. 11b WLAN system. In *Telecommunications (ICT), 2014 21st International Conference on* (pp. 442-446). IEEE. Doi:10.1109/ICT.2014.6845155

Paul, S. (2006). *The OSI Model: Understanding the Seven Layers of Computer Networks*. Global Knowledge Training LLC. Retrieved 27/07/16 at http://faculty.spokanefalls.edu/Rudlock/files/WP_Simoneau_OSIModel.pdf

Poole, I. (2012). *IEEE 802.11n Standard- IEEE 802.11n Wi-Fi/WLAN standard uses technologies including OFDM and MIMO to enable it to provide high speed data transport at 600 Mbps peak*. Retrieved 04/04/12 at <http://www.radio-electronics.com/info/wireless/wi-fi/ieee-802-11n.php>

Rachelle, L. M. (2001). *The OSI Model, an Overview*. SANS Institute. Retrieved 27/07/16 at <https://www.sans.org/reading-room/whitepapers/standards/osi-model-overview-543>

Sarkar, N. I., & Sowerby, K. W. (2009) The Combined Effect of Signal Strength and Traffic Type on WLAN Performance. *IEEE in Wireless Communications and Networking Conference*. Doi:10.1109/WCNC.2009.4917982

KEY TERMS AND DEFINITIONS

Cross Layer Modelling: It is a type of modelling that involves three basic stages namely: (i) wireless channel modelling, (ii) cross-layer extension of the wireless channel of interest to the layer of interest (iii) performance evaluation at the layer of interest.

Signal to Noise Ratio (SNR): It is a network metric measured in decibel (dB) and is equal to the signal level (in dBm) minus the noise level (in dBm).

Throughput: It is a measurement of the average rate that data (in bits) can be sent between one user and another.

Category N

Neural Networks

Applications of Artificial Neural Networks in Economics and Finance

N

Iva Mihaylova

University of St. Gallen, Switzerland

INTRODUCTION

The abilities of the brain, a complex, nonlinear and parallel computer, to learn, memorize and generalize, have motivated research in biological neural systems, referred to as “artificial neural systems” (ANNs) (Engelbrecht, 2007, p. 5). ANNs are a powerful technique for multivariate dependence analysis. The current capabilities of modern computing power and storage space allow for easy solutions of problems with a single objective by means of moderate-sized ANNs (Ibid.). Unlike other traditionally used techniques in economics and finance, artificial neural networks (ANNs) do not necessarily rely on assumptions about the population distribution (Shachmurove, 2003). Furthermore, the universal approximation theorem (Cybenko, 1989; Funahashi, 1989) postulates that an ANN with one hidden layer is capable of approximating any continuous function.¹ Due to these facts, ANNs are becoming an increasingly attractive and flexible analytic tool for applications in the area of economics and finance. The purpose of this article is to present such important applications with an emphasis on recent research trends.

BACKGROUND

Please see the Artificial Neural Networks entry, this volume for further discussion on ANNs basics. A frequently used ANN in economics and finance is the feedforward ANN. In this architecture, information stringently flows from the neurons situated in the input layer to the neurons in the

subsequent layers without any feedback connections. Each node is a neuron except for the input nodes. Rather than moving directly from input to output, new neurons as weighted combinations of input variables located in a hidden layer are created. This is a common ANN feature that ensures the achievement of modelling flexibility. When all neurons of a preceding layer are fully connected with all neurons in the subsequent layer, the ANN is referred to as “a multilayer perceptron” (MLP).

Similarly to MLPs, ANNs of a radial basis function type (RBF) (Broomhead & Lowe, 1988) are universal function approximations (Haykin, 1998) that have been a frequent choice in business applications. The difference between MLP and RBF ANNs consists in the type of approximation to non-linear mapping they deliver. The RBF delivers a local approximation while the MLP – a global one (Haykin, 1998). The detailing of contributions focused on the technical specifics of ANNs concerning optimization and architecture is beyond the scope of this article. Instead, the interested reader is referred to their comprehensive coverage in the books of Haykin (1998), Engelbrecht (2007), Haykin (2009) and Rojas (2013), while McNelis (2005), Zhang (2008), and Alexandridis and Zapanis (2014) contain applications in the area of economics and finance.

MAIN FOCUS OF THE ARTICLE

The purpose of this section is to survey important applications of ANNs in economics and finance. Emphasis has been laid on recent research trends in contributions in which ANNs are the main em-

DOI: 10.4018/978-1-5225-2255-3.ch575

pirical method or are combined with alternative research tools.

Applications of ANNs for Prediction

A major part of the ANN contributions in the area of economics and finance is dedicated to predictions. The implemented input data are either technical (for example, historical data of assets' statistical distribution moments) or fundamental (for example, data on the current economic activity). A tendency to focus on robustness checks is followed by some authors: (1) to compare the performance of ANNs to alternative methods, such as ARCH/GARCH models, which are appropriate for financial and economic data that frequently exhibit volatility clustering; (2) to combine them in hybrid models, or (3) to implement several ANN architectures in order to encourage the choice of the optimal one. In this section, such contributions are given priority over works that implement a single ANN model because of their goal of finding the best solution from several alternatives.

Falat et al. (2015) studied the volatility dynamics of the EUR/GBP currency with the traditionally applied in such research setting statistical approach that relies on ARCH and GARCH models, as well as several RBF ANN architectures. The authors reported higher but not considerable accuracy for the statistical approach. However, Falat et al. (2015) attributed this to factors such as their non-ideal but improvable optimization strategy. Kristjanpoller et al. (2014) combined ANNs and a GARCH model in a hybrid ANN-GARCH model for predicting the volatility of three emerging Latin American stock markets. On the basis of several robustness checks, the authors reached the conclusion that the hybrid model improved the performance of the GARCH model with respect to their sample. Kristjanpoller and Minutolo (2015) obtained analogous results concerning the spot and future volatility of gold. They also proposed an algorithm that incorporated variables in their model only if they improved it, assessed in terms of a measure of the model error.

Abrishami and Varahami (2011) implemented, for the prediction of gas prices, the popular in financial and economic research multilayer feedforward ANN and a group method of data handling (GMDH) ANN. The latter architecture is based on pattern recognition and can be used as a refined alternative to the traditional technical analysis (Abrishami & Varahami, 2011). The authors proposed a hybrid ANN containing GMDH and rule-based expert system modules. They reached the conclusion that it performed better during the recent 2008 financial crisis. Similarly, based on a GJR-GARCH ANN model for predicting the volatility of the NASDAQ Composite Index, Monfared and Enke (2014) reached the conclusion that hybrid models were appropriate for forecasting extreme events because the structure of volatility becomes more complex. ANNs should be avoided in low-volatility periods because they bring unnecessary complexity (Monfared & Enke, 2014).

A class of ANN models with a growing number of financial and economic applications is the wavelet ANNs, proposed by Zhang and Benveniste (1992). They combine a wavelet function as an activation function and a classical ANN. Bernard et al. (1998) explained their superior performance with their lower computational cost if wavelets were used instead of other activation functions. At the same time they preserve their approximation capacity due to their ANN component. Alexandridis and Zapranis (2013 a,b) demonstrated that wavelet ANNs could be efficiently used for weather derivative pricing and modelling as well as for the prediction of the prices of financial weather derivatives. Alexandridis et al. (2008) also reported accurate performance of wavelet ANNs for predicting crude oil returns. The interested reader is referred to the book of Alexandridis and Zapranis (2014) for a comprehensive account of the model identification framework for utilizing wavelet ANNs in various applications.

Sermpinis et al. (2016) proposed two hybrid ANNs: an adaptive evolutionary multilayer per-

ceptron (aDEMLP) and an adaptive evolutionary wavelet neural network (aDEWNN). The authors compared their performance to the traditionally applied MLP benchmark for forecasting and trading of the Dow Jones Industrial Average, iShares NYSE Composite Index Fund and SPDR S&P 500 exchange-traded funds. Sermpinis et al. (2016) demonstrated the superior accuracy of their hybrid models compared to the traditionally applied MLP, as well as wavelet ANN, autoregressive moving average models, a smooth transition autoregressive model and a random walk model. A distinct contribution of their work is that the adaptive hybrid ANNs they propose are deprived of any data-snooping bias.

Higher-order artificial neural networks (HO-ANNs) form part of another new direction in the ANN literature that deals with their shortcoming that, without the application of additional methods,² they are a “black box”, which means that the decision-making behind the results is challenging to explain. Zhang (2008) refers to HOANNs as “open box”, due to the transparent mapping of each neuron and weight to a functional variable and coefficient. This book, edited by Zhang (2008), contains a collection of the few contributions in finance and economics that apply HOANNs, for example for exchange rate, stock return, electricity load demand and price prediction.

Among the other ANNs applications can be mentioned: inflation forecasting (Nakamura, 2005); fundamental analysis and prediction of stock returns based on financial ratios (Yildiz & Yezegel, 2010); financial distress prediction (Chen & Du, 2009; Chen et al., 2009); prediction of public expenditure (Radulesku et al., 2015); risk management (Krollner, 2011; Lai et al., 2007); trading strategies (Zhang et al., 2007); real gross domestic production and industrial production (Aminian et al., 2006); sovereign debt crises (Fioramanti, 2008); and forecasting of the future value of financial assets (Giebel & Rainer, 2013).

Applications of ANNs for Classification



Classification is a supervised learning statistical problem that consists in extracting selected information from finite samples. Similarly to ANN applications for predictions, its usefulness lies in the possibility that its results can be used for subsequent decision-making. ANNs for classification in the area of finance and economics are predominantly used for analysis of the fraudulent behaviour and insolvency of economic agents (for example, customers and banks). The purpose of this section is to present such recent common applications.

Baesens et al. (2005) compared the performance of various ANN models and a proportional hazards model in relation to a data set with 15,000 observations for personal loans. The authors reached the conclusion that with respect to the chosen data set, ANNs perform better at distinguishing good from bad borrowers. Kirkos et al. (2007) used various competing statistical methods for detecting fraudulent financial statements: an ANN, a decision tree (Arminger et al., 1997) and a Bayesian belief network (Pearl, 1986), while Chen (2016) also implemented a support vector machine (Vapnik, 1995) with the same research purpose. In both studies the implemented traditional ANNs are not the best performing method. Since ANNs have a universal approximation property, and, similarly to the conclusion of Falat et al. (2015), their performance may be enhanced by an improved optimization strategy or by the application of a different ANN architecture. For example, Ravisankar et al. (2011) applied the probabilistic neural network (PNN) proposed by Specht (1990), based on the Bayes classifier, in addition to the traditional classification methods (logistic regression, a support vector machine and a MLP ANN). The authors report a superior performance of the PNN for financial statement fraud detection after applying the standard for

classification applications selection of the most significant inputs and cross-validation. The interested reader is referred to the survey of Ngai et al. (2011) for a comprehensive account of the existing data mining techniques in addition to ANNs for detection of financial fraud.

Applications of ANNs for Modelling

ANNs are applied less frequently in economics and finance for modelling. One of their applications is for modelling of bounded rational adaptive artificial agents. This approach emerged in response to the observed inconsistency of the assumption for agents' full rationality (see, for example, Allais, 1953). Introduced by Simon (1955), the concept "bounded rationality" relies on the realistic assumption that agents produce decisions based on the available and not complete information. In this setting, the ANN's neurons are interpreted as agents that receive information from the environment and produce decisions on this basis.

Repetitive games are a well-developed strand of the mathematical modelling literature for studying human behaviour. Chernobelskiy et al. (2013) applied an ANN for simulating agents' behaviour with different levels of sympathy and selfishness. This is a more realistic approach than the traditional assumption in economic models that participants take decisions that maximize their personal utility. Sgroi and Zizzo (2008) studied the question of whether ANNs can learn to play a $n \times n$ game by playing a finite set of other $n \times n$ games. Svalestuen et al. (2015) proposed an application of an ANN and evolutionary game theory (EGT) for predicting social behaviour through multi-agent-based simulation. The applied ANN adapts its weights by using the EGT "learning-by-imitation" method in order to reflect the agent's decision-making rule. This simulation-based modelling strategy can be integrated into a decision support system for fair policies for stakeholders (Svalestuen et al., 2015).

The prospect theory of Kahneman and Tversky (1979), which builds on psychological biases, is

another response to the controversial assumption of agents' full rationality and marks the emergence of behavioural finance and behavioural economics. Furthermore, the advent of non-invasive human brain imaging techniques makes it possible to study biologically economic theories of choice and the emergence of neuroeconomics as new research fields. Farmer and Lo (1999, p. 9991) assess positively the prospects of this trend with respect to finance:

One of the most promising directions is to view financial markets from a biological perspective and, specifically, within an evolutionary framework in which markets, instruments, institutions, and investors interact and evolve dynamically according to the "law" of economic selection. Under this view, financial agents compete and adapt, but they do not necessarily do so in an optimal fashion. Evolutionary and ecological models of financial markets are truly a new frontier whose exploration has just begun.

The works of Wang, (2002), Wong and Wang (2006), and Rustichini and Padoa-Schioppa (2015) can be given as examples of ANN applications in this context.

Padoa-Schioppa (2008, p. 451) reflects on the future of neuroeconomics:

If neuroscience will contribute to economics, it will do so by the way of psychology. In other words, discoveries about the brain will hopefully lead to better theories of thought, which in turn will hopefully lead to better theories of choice behaviour. The syllogism is based on two premises – that neuroscience can contribute to psychology, and that psychology can contribute to economics. [...] It thus follows that neuroscience can contribute to economics. Notably, this proposition is a statement of possibility. Whether any discovery about the brain will in fact lead to better economic theories is ultimately an empirical question. Current examples might not be the most compelling, but the field is in its infancy and time will tell.

FUTURE RESEARCH DIRECTIONS

A future trend in the evolution of ANN research in economics and finance will build on a major difference between ANNs and the human brain: the principle of self-organization of the latter (Kelso, 1995), which ensures its adaptability. Traditional ANNs, for example the neuron-fixed feedforward networks, evolve according to a pre-defined architecture and are not self-organized. The elaboration of new adaptive neural network models is promising for applications in economics and finance because they successfully adapt to non-stationary input data. That is, the ANN system remains optimal over time because its parameters change in line with the input. For example, Patra et al. (2015) proposed an adaptive local linear optimized RBF ANN model for financial time series prediction. Their model has displayed superior performance than the traditional architectures based on an RBF and MLP.

An essential and integral part of any model in economics and finance is its interpretation. A common misconception concerning ANNs is the claim that they are a “black box”, wrongly attributed to their failure to explain the contribution of each independent variable in the prediction process. However, although undertaken by very few authors, studies that develop techniques for tracking the explanatory capacity of ANNs exist and hold great promise for their wider future successful applicability in different business fields. Olden et al. (2004) and Gevrey et al. (2003) surveyed and tested the main existing methods for explaining the capacity of ANNs. Paliwal and Kumar (2011) proposed a new method for ranking the relative importance of each input variable in the ANNs based on the interquartile range of the empirical distribution of the network weights that connect input to hidden nodes. The authors also considered the important question in economics and finance of how their method would perform when the input data exhibit multicollinearity. Paliwal and Kumar

(2011) reached the conclusion that it performs well in the presence of such data provided that the sample size is big enough.

The observed tendency towards increased availability of big data provides opportunities for more refined testing of existing theories and exploration of new research questions in economics and finance that previously were impossible to answer based on macro-level aggregate data. Furthermore, it can be expected that big data statistical methods and topics from related fields will be adapted for economic and financial ANN applications. For example, the traditionally applied in social marketing sentiment analysis or opinion mining has been applied for analysis of financial news by Heston and Sinha (2014). Another application of ANN that is potentially extendable to economic and financial big data was studied by Chen et al. (2010) and Zhu et al. (2010). Both contributions rely on game theory in order to study social networks. The interested reader is referred to the presentation of Roncalli (2014) in section “Additional readings” for further references and a comparison of the performance of ANN and other methods in various big data financial applications (bond scoring, asset picking, trend filtering, index tracking, technical analysis, mean reverting).

As stated by David Hendry based on Mayer-Schönberger and Cukier (2013), all authors cited by Taylor et al. (2014, p. 6), “much work using Big Data is essentially descriptive, dealing with correlation rather than causality. [...] if economics cannot seek causality, it similarly loses one of its mainstays [...]” Consequently, methods that enhance and adapt the explanatory capacity of ANNs to big data business applications have promising potential.

Deep learning applications can also be expected to attract professionals’ and scholars’ attention due to increased availability of high quality data. A selection with examples of such recent applications is provided in section “Additional readings”.

CONCLUSION

There is well-documented evidence that the observed economic and financial phenomena frequently have a complex non-linear behaviour. ANNs are attracting growing research interest due to their capacity to deal with the modelling challenges of this type of data. There is a tendency in the new research contributions to combine ANNs and other methods in more flexible hybrid models. Consequently, emphasis is placed on fine-tuning ANNs in order to add power to their explanatory capacity. A parallel tendency with a similar goal is to compare the performance of ANNs to other Alternative models.

An essential and integral part of any model in economics and finance is its interpretation. Although undertaken by very few authors, studies that develop techniques for tracking the relative contribution of each input variable of ANNs have exhibited a tendency towards a gradual increase in recent years. These particular studies hold great promise for the wider future successful applicability of ANNs in different business fields.

Finally, the tendency towards increased availability of data with much refined granularity opens new research opportunities. Both more refined testing of existing theories and exploration of new economic and financial research questions, as well as adaptation of methods and topics of related to the finance and economics research fields, can be expected to draw scholars' attention.

REFERENCES

- Alexandridis, A., & Zapranis, A. (2013a). Wind derivatives: Modeling and pricing. *Computational Economics*, 41(3), 299–326. doi:10.1007/s10614-012-9350-y
- Alexandridis, A., & Zapranis, A. (2013b). *Weather derivatives: Modeling and pricing weather-related risk*. New York: Springer-Verlag. doi:10.1007/978-1-4614-6071-8
- Alexandridis, A. K., & Zapranis, A. D. (2014). *Wavelet Neural Networks: With applications in financial engineering, chaos, and classification*. Hoboken, NJ: John Wiley & Sons. doi:10.1002/9781118596272
- Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'école américaine. *Econometrica*, 21(4), 503–546. doi:10.2307/1907921
- Aminian, F., Suarez, E. D., Aminian, M., & Walz, D. T. (2006). Forecasting economic data with neural networks. *Computational Economics*, 28(1), 71–88. doi:10.1007/s10614-006-9041-7
- Arminger, G., Enache, D., & Bonne, T. (1997). Analyzing credit risk data: A comparison of logistic discrimination classification tree analysis and feed forward networks. *Computational Statistics*, 12(2), 293–310.
- Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science*, 211(4489), 1390–1396. doi:10.1126/science.7466396 PMID:7466396
- Baesens, B., Van Gestel, T., Stepanova, M., Van den Poel, D., & Vanthienen, J. (2005). Neural network survival analysis for personal loan data. *The Journal of the Operational Research Society*, 56(9), 1089–1098. doi:10.1057/palgrave.jors.2601990

- Bernard, C., Mallat, S., & Slotine, J. J. (1998). Wavelet interpolation networks. *Proceedings of ESANN '98*, 47–52.
- Broomhead, D., & Lowe, D. (1988). Multivariable functional interpolation and adaptive networks. *Complex Systems*, 2, 321–355.
- Chen, S. (2016). Detection of fraudulent financial statements using the hybrid data mining approach. *SpringerPlus*, 5(1), 1. PMID:26848429
- Chen, W., Liu, Z., Sun, X., & Wang, Y. (2010). A game-theoretic framework to identify overlapping communities in social networks. *Data Mining and Knowledge Discovery*, 21(2), 224–240. doi:10.1007/s10618-010-0186-6
- Chen, W. S., & Du, Y. K. (2009). Using neural networks and data mining techniques for the financial distress prediction model. *Expert Systems with Applications*, 36(2), 4075–4086. doi:10.1016/j.eswa.2008.03.020
- Chernobelskiy, A., Dixit, V., Cala, A., Pandya, S., & Rosas, H. J. (2013). Modeling, learning and cooperation in Iterative Games. *SIAM Undergraduate Research Online*, 6, 42–52.
- Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems*, 2(4), 303–314. doi:10.1007/BF02551274
- Engelbrecht, A. P. (2007). *Computational intelligence: an introduction*. Chichester, UK: John Wiley & Sons. doi:10.1002/9780470512517
- Falat, L., Stanikova, Z., Durisova, M., Holkova, B., & Potkanova, T. (2015). Application of Neural Network models in modelling economic time series with non-constant volatility. *Procedia Economics and Finance*, 34, 600–607. doi:10.1016/S2212-5671(15)01674-3
- Farmer, J. D., & Lo, A. W. (1999). Frontiers of finance: Evolution and efficient markets. *Proceedings of the National Academy of Sciences of the United States of America*, 96(18), 9991–9992. doi:10.1073/pnas.96.18.9991 PMID:10468547
- Fioramanti, M. (2008). Predicting sovereign debt crises using artificial neural networks: A comparative approach. *Journal of Financial Stability*, 4(2), 149–164. doi:10.1016/j.jfs.2008.01.001
- Frank, M. J., & Claus, E. D. (2006). Anatomy of a decision: Striato-orbitofrontal interactions in reinforcement learning, decision making, and reversal. *Psychological Review*, 113(2), 300–326. doi:10.1037/0033-295X.113.2.300 PMID:16637763
- Funahashi, K. (1989). On the approximate realization of continuous mappings by neural networks. *Neural Networks*, 2(3), 183–192. doi:10.1016/0893-6080(89)90003-8
- Gevrey, M., Dimopoulos, I., & Lek, S. (2003). Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecological Modelling*, 160(3), 249–264. doi:10.1016/S0304-3800(02)00257-0
- Giebel, S., & Rainer, M. (2013). Neural network calibrated stochastic processes: Forecasting financial assets. *Central European Journal of Operations Research*, 21(2), 277–293. doi:10.1007/s10100-011-0234-3
- Haykin, S. S. (1998). *Neural Networks: A comprehensive foundation*. Patparganj, Delhi: Pearson Education. Singapore: Indian Branch.
- Haykin, S. S. (2009). *Neural networks and learning machines*. Upper Saddle River, NJ: Pearson Education.
- He, K., Xie, C., Chen, S., & Lai, K. K. (2009). Estimating VaR in crude oil market: A novel multi-scale non-linear ensemble approach incorporating wavelet analysis and neural network. *Neurocomputing*, 72(16), 3428–3438. doi:10.1016/j.neucom.2008.09.026
- Heston, S. L., & Sinha, N. R. (2014). *News versus sentiment: Comparing textual processing approaches for predicting stock returns*. Robert H. Smith School Research Paper.

- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2), 263–291. doi:10.2307/1914185
- Kelso, J. A. S. (1995). *The self-organization of brain and behavior*. Cambridge, MA: MIT Press.
- Kirkos, E., Spathis, C., & Manolopoulos, Y. (2007). Data Mining techniques for the detection of fraudulent financial statements. *Expert Systems with Applications*, 32(4), 995–1003. doi:10.1016/j.eswa.2006.02.016
- Kristjanpoller, W., Fadic, A., & Minutolo, M. C. (2014). Volatility forecast using hybrid neural network models. *Expert Systems with Applications*, 41(5), 2437–2442. doi:10.1016/j.eswa.2013.09.043
- Kristjanpoller, W., & Minutolo, M. C. (2015). Gold price volatility: A forecasting approach using the Artificial Neural Network–GARCH model. *Expert Systems with Applications*, 42(20), 7245–7251. doi:10.1016/j.eswa.2015.04.058
- Krollner, B. (2011). *Risk Management in the Australian Stock market using Artificial Neural networks* (Unpublished doctoral dissertation). School of Information Technology, Bond University, Queensland, Australia.
- Lai, K. K., He, K., & Yen, J. (2007). Modeling VaR in crude oil market: A multi scale nonlinear ensemble approach incorporating wavelet analysis and ANN. In Y. Shi, G. D. van Albada, J. Dongarra, & P. M. A. Sloot (Eds.), *Computational Science–ICCS2007* (pp. 554–561). Berlin: Springer Verlag. doi:10.1007/978-3-540-72584-8_73
- Laney, D. (2001) *3D data management: Controlling data volume, variety and velocity* (META Group File 949). Retrieved September 4, 2016, from <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>
- Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. London: John Murray.
- McNelis, P. D. (2005). *Neural networks in finance: gaining predictive edge in the market*. Burlington, MA: Elsevier Academic Press.
- Monfared, S. A., & Enke, D. (2014). Volatility forecasting using a hybrid GJR-GARCH neural network model. *Procedia Computer Science*, 36, 246–253. doi:10.1016/j.procs.2014.09.087
- Nakamura, E. (2005). Inflation forecasting using a neural network. *Economics Letters*, 86(3), 373–378. doi:10.1016/j.econlet.2004.09.003
- Olden, J. D., Joy, M. K., & Death, R. G. (2004). An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. *Ecological Modelling*, 178(3), 389–397. doi:10.1016/j.ecolmodel.2004.03.013
- Padoa-Schioppa, C. (2008). The syllogism of neuro-economics. *Economics and Philosophy*, 24(03), 449–457. doi:10.1017/S0266267108002071 PMID:19050765
- Paliwal, M., & Kumar, U. A. (2011). Assessing the contribution of variables in feed forward neural network. *Applied Soft Computing*, 11(4), 3690–3696. doi:10.1016/j.asoc.2011.01.040
- Patra, A., Das, S., Mishra, S. N., & Senapati, M. R. (2015). An adaptive local linear optimized radial basis functional neural network model for financial time series prediction. *Neural Computing & Applications*, 1–10.
- Pearl, J. (1986). Fusion, propagation, and structuring in belief networks. *Artificial Intelligence*, 29(3), 241–288. doi:10.1016/0004-3702(86)90072-X

- Radulescu, M., Logica, B., & Zamfiroiu, T. (2015). Forecasting public expenditure by using feed-forward neural networks. *Economic Research-Ekonomska Istraživanja*, 28(1), 668–686. doi:10.1080/1331677X.2015.1081828
- Ravisankar, P., Ravi, V., Raghava Rao, G., & Bose, I. (2011). Detection of financial statement fraud and feature selection using data mining techniques. *Decision Support Systems*, 50(2), 491–500. doi:10.1016/j.dss.2010.11.006
- Rojas, R. (2013). *Neural networks: a systematic introduction*. Berlin: Springer Verlag.
- Rustichini, A., & Padoa-Schioppa, C. (2015). A neuro-computational model of economic decisions. *Journal of Neurophysiology*, 114(3), 1382–1398. doi:10.1152/jn.00184.2015 PMID:26063776
- Schroeder, R. (2014). Big Data: Towards a more scientific social science and humanities? In M. Graham & W. H. Dutton (Eds.), *Society and the Internet: How Networks of Information are Changing our Lives*. Oxford, UK: Oxford University Press. doi:10.1093/acprof:oso/9780199661992.003.0011
- Sermpinis, G., Verousis, T., & Theofilatos, K. (2016). Adaptive Evolutionary Neural Networks for Forecasting and Trading without a Data-Snooping Bias. *Journal of Forecasting*, 35(1), 1–12. doi:10.1002/for.2338
- Sgroi, D., & Zizzo, D. J. (2008). Learning to Play 3x3 Games: Neural Networks as Bounded-Rational Players. *Journal of Economic Behavior & Organization*, 69(1), 27–38. doi:10.1016/j.jebo.2008.09.008
- Shachmurove, Y. (2003). *Applying artificial neural networks to Business*. Economics and Finance, Dept. of Economics, University of Pennsylvania. Retrieved September 4, 2016, from economics.sas.upenn.edu/system/files/02-08.pdf
- Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, 69(1), 99–118. doi:10.2307/1884852
- Specht, D. F. (1990). Probabilistic neural networks. *Neural Networks*, 3(1), 110–118. doi:10.1016/0893-6080(90)90049-Q PMID:18282828
- Svalestuen, Y., Öztürk, P., Tidemann, A., & Tiller, R. (2015). Agent-Based Simulation of Stakeholder Behaviour through Evolutionary Game Theory. Artificial Life and Computational Intelligence. In Chalup, S., Blair, A. D., Randall, M. (Eds.). *Artificial Life and Computational Intelligence First Australasian Conference, Proceedings*. Cham: Springer Verlag. doi:10.1007/978-3-319-14803-8_8
- Taylor, L., Schroeder, R., & Meyer, E. (2014). Emerging practices and perspectives on Big Data analysis in economics: Bigger and better or more of the same? *Big Data & Society*, 1(2).
- Vapnik, V. (1995). *The nature of statistical learning theory*. New York: Springer Verlag Inc. doi:10.1007/978-1-4757-2440-0
- Wang, X. J. (2002). Probabilistic decision making by slow reverberation in cortical circuits. *Neuron*, 36(5), 955–968. doi:10.1016/S0896-6273(02)01092-9 PMID:12467598
- Wong, K. F., & Wang, X. J. (2006). A recurrent network mechanism of time integration in perceptual decisions. *The Journal of Neuroscience*, 26(4), 1314–1328. doi:10.1523/JNEUROSCI.3733-05.2006 PMID:16436619
- Xu, S., & Zhang, M. (2007). A new Adaptive Neural Network model for financial data mining. In D. Liu, S. Fei, Z.-G. Hou, H. Zhang, & C. Sun (Eds.), *Advances in Neural Networks-ISNN 2007* (pp. 1265–1273). Berlin: Springer Verlag Inc. doi:10.1007/978-3-540-72383-7_147

Yildiz, B., & Yezegel, A. (2010). Fundamental analysis with artificial neural network. *The International Journal of Business and Finance Research*, 4(1), 149–158.

Zhang, D., Jiang, Q., & Li, X. (2007). Application of Neural Networks in financial data mining. *International Journal of Computer, Electrical, Automation, Control and Information Engineering*, 1(1).

Zhang, M. (2008). *Artificial higher order neural networks for economics and business*. Hershey, PA: IGI Global.

Zhang, Q., & Benveniste, A. (1992). Wavelet networks. *IEEE Transactions on Neural Networks*, 3(6), 889–898. doi:10.1109/72.165591 PMID:18276486

Zhu, X., Wu, J., Wang, C., & Xie, J. (2010). *On the design of utility functions in community formation game*. Retrieved September 4, 2016, from http://www.academia.edu/5626801/on_the_design_of_utility_functions_in_community_formation_games

ADDITIONAL READING

Batres-Estrada, B. (2015). Deep learning for multivariate financial time series. Retrieved September 4, 2016, from <http://www.diva-portal.org/smash/get/diva2:820891/FULLTEXT01.pdf>

Ding, X., Zhang, Y., Liu, T., & Duan, J. (2015, June). Deep learning for event-driven stock prediction. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence (ICJAI'15)* (pp. 2327–2333).

Dixon, M. F., Klabjan, D., & Bang, J. H. (2016). Classification-based Financial Markets Prediction using Deep Neural Networks. Retrieved September 4, 2016, from <http://arxiv.org/pdf/1603.08604.pdf>

Roncalli, T. (2014). Big Data in Asset Management. Retrieved September 4, 2016, from <http://www.thierry-roncalli.com/download/big-data-in-asset-management.pdf>

Xiong, R., Nicholas, E. P., & Shen, Y. (2015). Deep Learning Stock Volatilities with Google Domestic Trends. Retrieved September 4, 2016, from <http://arxiv.org/pdf/1512.04916v3.pdf>

Zhu, C., Yin, J., & Li, Q. (2014). A stock decision support system based on DBNs. *Journal of Computer Information Systems*, 10(2), 883–893.

KEY TERMS AND DEFINITIONS

Artificial Neural Network Architecture (Topology): This concept involves decision-making concerning the size and number of the ANN layers, as well as the design of the connections among the ANN units.

Data Snooping: Denotes the use of statistical inference methods for detecting statistically significant patterns without prior theorizing and formulating hypotheses regarding their causal relationship.

Deep Learning: A branch of machine learning to whose architectures belong deep ANNs. The term “deep” denotes the application of multiple layers with a complex structure.

Explanatory Capacity of Artificial Neural Networks: It consists in the implementation of methods that are able to determine the relative contribution of each ANN input variable to the ANN output. See Gevrey et al. (2003) for further details.

Multicollinearity: This concept explains the situation in regression analysis when two or more of the independent variables are moderately to highly correlated. Consequently, the regression results are biased.

Stationarity: A time series is stationary when its statistical distribution moments do not change over time.

Supervised Learning: It consists in learning from data with a known-in-advance outcome that is predicted based on a set of inputs, referred to as “features”.

ENDNOTES

- ¹ This statement is true if a global minimum can be reached for each concrete ANN application.
- ² See section “Future research directions” for a discussion on key contributions in this area.

Artificial Neural Networks and Their Applications in Business

Trevor J. Bihl

Air Force Institute of Technology, USA

William A. Young

Ohio University, USA

Gary R. Weckman

Ohio University, USA

INTRODUCTION

Perspectives of Chapter

Knowledge discovery in datasets (KDD), the process of finding non-obvious patterns or trends in datasets primarily to assist in understanding complicated systems (Mannila, 1996), leveraging computational pattern recognition methods to predict, describe, classify, group, and categorization data (Jain, Duin, & Mao, 2000). The nonlinearity of ANNs lend to modeling complex data structures; however, this also results in ANNs being complex and opaque to many users (Weckman, et al., 2009). While ANNs see application in business applications, some hesitation and misconceptions appear due to the ‘black-box’ nature of ANNs (Dewdney, 1997); hence traditional statistical-based models are far more used in practice. Although some interpretability aspects exist when applying ANNs (de Marchi, Gelpi, & Grynawski, 2004), it should be noted that the interpretability of ANNs is subjective relative to other commonly used tools.

Objectives of Chapter

ANNs methods are interconnected networks of nodes, which are trained on patterns in data through statistical learning methods (Jain, Duin, & Mao, 2000). Although ANNs are computation-

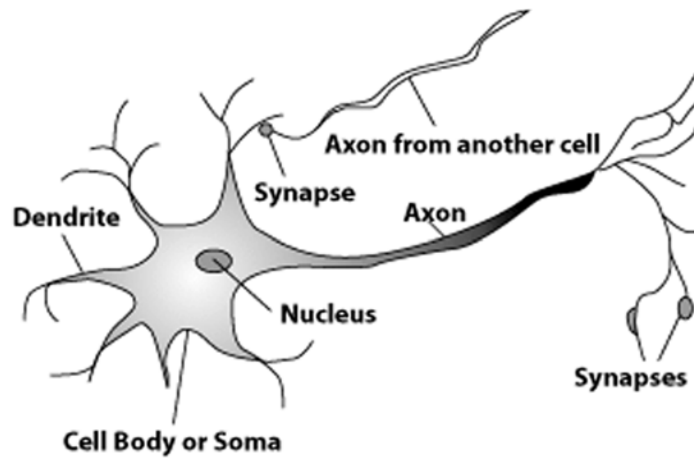
ally complex, inherently ANNs are statistical in nature and epistemologically similar in function to Bayesian (Beck, King, & Zeng, 2004) and likelihood methods (Verikas & Bacauskiene, 2002). Various software packages are now available for practitioners, including NeuroDimensions (2005), Matlab (2010), JMP (Sall, Lehman, Stephens, & Creighton, 2012), and R (2008). The objective of this chapter is to provide readers with a general background of ANNs, their business applications, and developing quality ANN models. The target audience is intended to be readers who may not be familiar with this form of mathematical modeling practice but may want to pursue it for their business need.

In Young, Bihl, and Weckman (2014), the authors provided a brief overview of ANNs and their applications in business. Herein, the authors have revised this discussion and included new material on ANN architectures and an example end-to-end analysis of business data using ANNs with the JMP13 Pro platform.

Biological Inspiration of ANNs

ANNs are computational machine learning models that are neurologically inspired with the intent of representing complex non-linear input and output relationships. The earliest known development of the ANN model was in the 1940s (McCulloch &

Figure 1. Biological Neuron (Young, Holland, & Weckman, 2008)



Pitts, 1943), with the fundamental building block of the ANN model being the neuron. Figure 1 displays a basic sketch of the biological neural network model (Neuralpower, 2007).

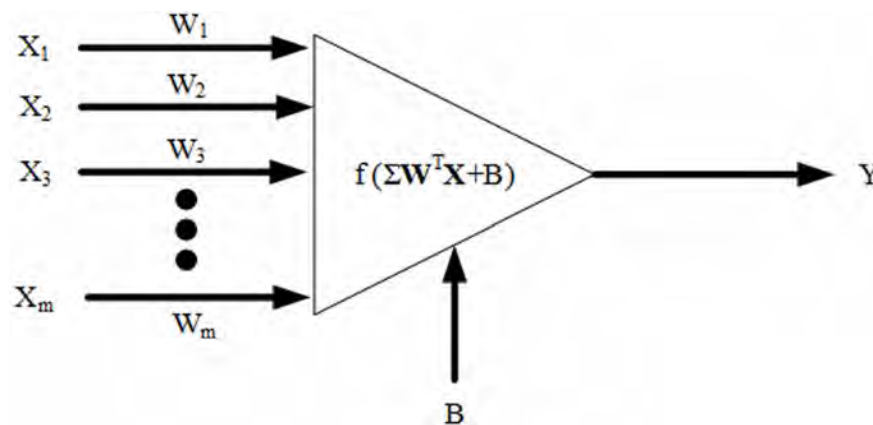
Computational ANNs extend biological neuron models by considering multiple interconnected nodes termed “neurons” which employ statistical methods to learn patterns between inputs and outputs (Jain, Duin, & Mao, 2000). Through organizational and iterative principles, connection weights between neurons, inputs and outputs are computed to learn a nonlinear input-output relationship (Jain, Duin, & Mao, 2000).

BACKGROUND: ANN PROPERTIES

Underlying Artificial Neuron Operations

Perceptrons, aka “nodes”, computationally inspired models of biological neurons, employ mathematical and statistical methods to determine a mapping between independent inputs (X_1 , X_2 , X_3 , and X_m) and a dependent variable (Y). In operation, as represented in Figure 2 and with an example linear activation function, perceptron connection weights (W_1 , W_2 , W_3 , and W_m) and

Figure 2. Artificial Neuron, adapted from (Young, Holland, & Weckman, 2008; Li, Artificial neural networks and their business applications, 1994)



bias (B) are determined through iterative methods such as gradient descents. Weights represent the determined strength of the association between independent and dependent variables. Weights can be positive, negative, or zero; the weighted sum is then processed by the neuron's activation function and then sent through the rest of the network.

Table 1 contains the three most commonly used activation functions for a single perceptron (Duch & Jankowski, 1997) along with the step function. While the step function could be used, ANN training frequency employs gradient descents for learning which involve a derivative of the step function. Therefore, the hyperbolic tangent, linear, and logistic/sigmoid activation functions are more common due to their derivational advantages. Due to interconnections and training, an ANN using linear activation functions will still result in a nonlinear model.

To determine which activation to use, a trial-and-error procedure is required; however, the sigmoid function is the default for many operations. One simple ANN example is logistic regression whereby a single ANN neuron, a *perceptron*, with one input and one output node is essentially considered (Timmerman, et al., 1999).

Architectures

In operation, ANNs generally consist of three types of layers: input, hidden, and output. Layers between the input and output layers are typically termed 'hidden' because they act as 'black-box' transformations where the mechanism for prediction is not transparent to users (Olden & Jackson, 2002). The most commonly utilized ANN architecture in practice is the multilayer perceptron (MLP) model as seen in the neural interpretation diagram (NID) of Figure 3 (Özesmi & Özesmi, 1999). The four layer MLP ANN in Figure 3 uses hyperbolic tangent activation functions and, moving from left to right, one sees that the first layer (or input layer) is connected to the first hidden

Table 1. Common activation functions

Activation Functions	Definitions	Range
Hyperbolic Tangent	$\frac{e^x - e^{-x}}{(e^x + e^{-x})}$	(-1, 1)
Linear	x	$(-\infty, +\infty)$
Logistic/Sigmoid	$\frac{1}{(1 + e^{-x})}$	(0, 1)
Step	$\begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases}$	(0, 1)

layer, with the output of the first hidden layer then feeding into any subsequent hidden layers, with the results passed to the output layer. In this diagram, solid lines represent excitation signals (or positive weights), dashed lines represent inhibitory signals (or negative weights), and line thickness reflect the magnitude of the weight. To determine a network's weights one trains the network, where the values and activation functions determine what relationships are passed through the network.

However, ANNs come in a variety of structures, as seen in Figure 4. Although detailing all methods available is beyond the scope of this overview, some of variants include probabilistic neural networks, evolutionary ANNs (Yao & Liu, 1997; Yao & Liu, 1998; Yao, 1999), radial basis function networks (Looney C., 1997), recurrent ANNs (Looney C., 1997), and convolution networks (Duda, Hart, & Stork, 2001), among many other methods. These methods largely differ by the nature of connections (feedforward or feedback), the presence of an output layer (e.g. learning vector quantization (LVQ) ANNs, a type of Self-Organizing Map (SOM), have only an input and hidden layer), and the nature of the learning rules (Jain, Mao, & Mohiuddin, 1996; Looney C., 1997, p. 368).

Figure 3. Multi-Layered ANN, adapted from (Young, Holland, & Weckman, 2008)

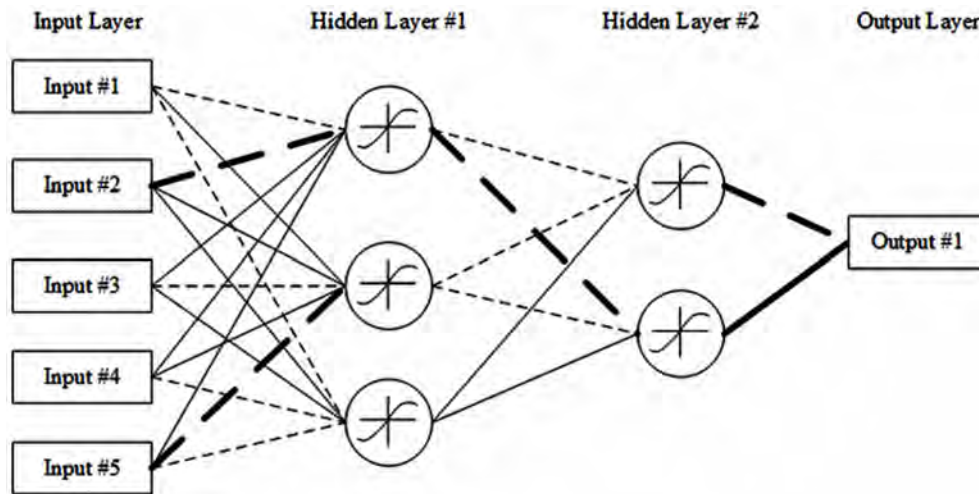
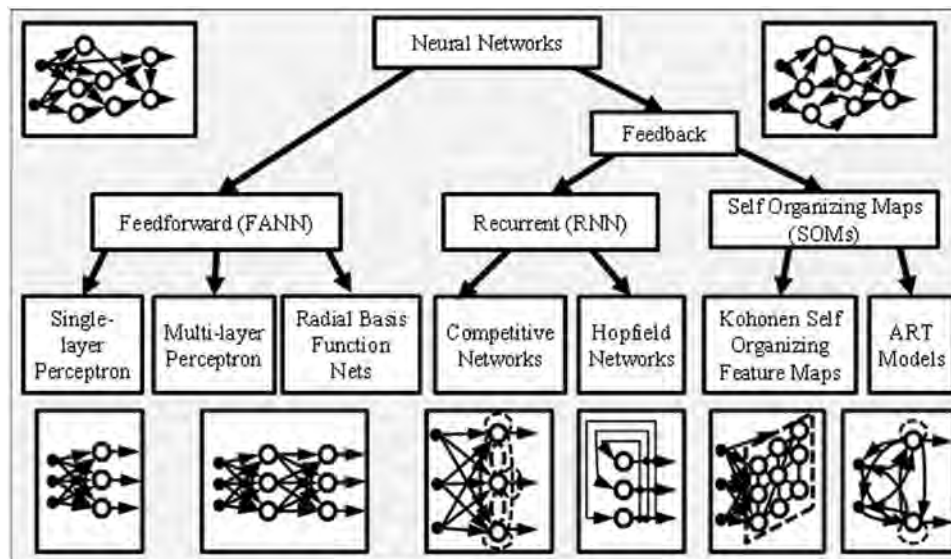


Figure 4. General Taxonomy of ANN Architectures, taken from (Bihl, 2015) and adapted from (Jain, Mao, & Mohiuddin, 1996) using the ANN families of (Looney C. G., 1997, p. 368)



DECISIONS IN DEVELOPING ANNs

Data Preprocessing Issues

Preprocessing steps include any necessary normalization, imputing missing values, and data quality checks. Practically, models should be trained with more data than free parameters, with ideal ratios of 10:1, 5:1 or 3:1 (observations to

free parameters), to prevent over-fitting (Schunn & Wallach, 2005). Conversely, one could penalize the performance index for every free parameter in a model (Myung, 2000; Grünwald, 2001). Additionally, understanding the class balance in the data is critical. If one aims to understand a minority class (e.g. loan defaults), strategies such as under-sampling (Liu, Wu, & Zhou, 2009) can be beneficial.

Data Partitioning Philosophies

As with any pattern recognition problem, one should avoid training and testing on the same observations; to avoid this situation, training, testing, and cross-validation observations should be sequestered (Zhang G. P., 2007). In building ANNs, one uses the training and test sets to build an ANN, with weights fine-tuned for performance, the validation set then evaluates final ANN accuracy (Zhang G. P., 2007).

Two issues exist in creating training/testing/validation data subsets: 1) selecting subset percentages, and 2) the data splitting method to create representative subsets. Consensus does not exist on subset percentages, as seen in Table 2. Multiple strategies also exist for appropriately selecting training/testing/validation observations (May, Maier, & Dandy, 2010); generally, these are either random or deterministic. Random selection is one of the most common methods (May, Maier, & Dandy, 2010); while deterministic means follow different philosophies (Wu, May, Dandy, & Maier, 2012), and include methods such as CADEX (Kennard & Stone, 1969), DUPLEX (Snee, 1977), and SSTATS (Mindrup, Friend, & Bauer, 2012).

SELECTING AND SPECIFYING ANN ARCHITECTURE

Overfitting is a known issue in ANNs, c.f. (Clark, 1999), therefore appropriately specifying the ANN architecture, such as the number of layers and nodes, is essential. An example of the importance

of proper architecture selection is presented in Figure 5, on a small dataset ($N = 27$, $N_F = 3$ data features, $N_c = 3$ classes). Here averaged results from 100 replications are presented. One can see that instantiating too few hidden nodes yields poor, but consistent, accuracy, while instantiating too many hidden nodes provides excellent training accuracy but poor testing accuracy on unseen observations, a condition of overfitting.

Selecting the appropriate number of hidden nodes, N_{Hidden} , involves both science and art; too few will not be accurate, while too many nodes introduce noise (Looney C., 1997). However, a variety of heuristics exist, see (Gao, Chen, & Qin, 2010), to facilitate this process. One basic and reasonable ANN heuristic involves instantiating an amount proportion to the number of classes:

$$N_{Hidden} = aN_c$$

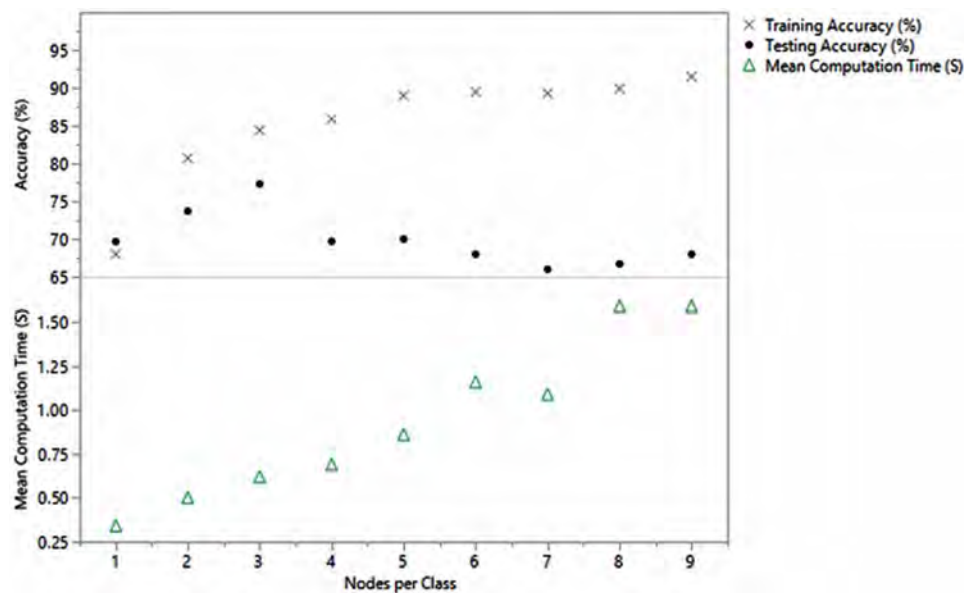
where a is a user specified constant and N_c classes (Looney C. G., 1997, p. 101). For a small number of classes twice the number of nodes as there are classes can be sufficient (Looney C., 1997); however, selecting an ideal number frequently involves testing multiple architectures.

Differing opinions exist on the appropriate number of hidden layers. Although, theoretically, one hidden layer is sufficient for properly posed problems (Kurkova, 1992) complicated problems may best be solved with multiple hidden layers (Zhang G. P., 2007; LeCun, Bengio, & Hinton, 2015). Thus it is advisable to begin with a single layer and only add additional layers (complexity and computation time) if needed.

Table 2. Example philosophies on data splitting for ANNs

Reference	Training	Test	Validation
(Poh, 1991)	65%	10%	25%
(Looney C., 1997)	60%	15%	25%
(Wu, May, Dandy, & Maier, 2012)	60%	20%	20%
(Balbin, Safieva, & Lomakina, 2009)	varies	varies	10%

Figure 5. Example of Overfitting from (Bihl, 2015) using the Three-Class Insect Dataset of (Lindsey, Herzberg, & Watts, 1987) (Hand, Daly, McConway, Lunn, & Ostrowski, 1993) with an LVQ ANN



Training Method

Determining weights for a given network involves training on a dataset; two main paradigms exist in this process: supervised (classification/prediction) and unsupervised (clustering) learning (Haykin, 1999). Both supervised and unsupervised schemes are iterative, with weights updated after investigating all observations in the training set. While the learning algorithm investigates all samples, known as an epoch, statistical information is collected, such as mean squared error (MSE), in order to adjust the weights of a network.

Table 3. Common ANN training methods

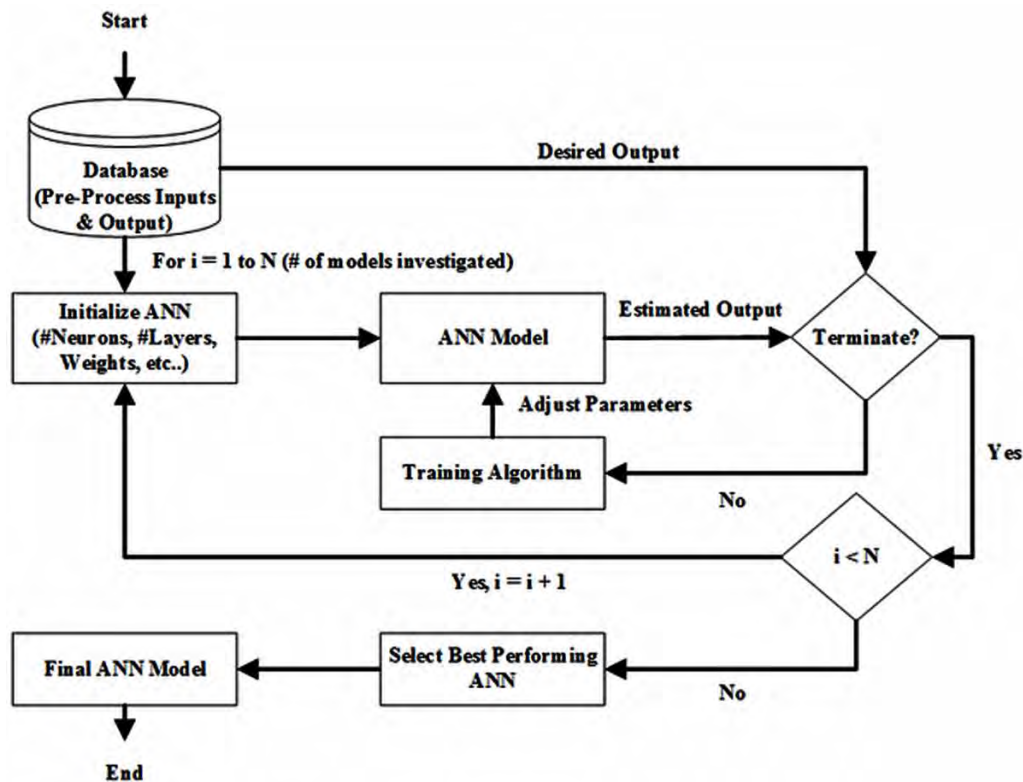
Training Method	Reference for Additional Information
Back-propagation	(Rumelhart & McClelland, 1986)
Batch training	(Duda, Hart, & Stork, 2001)
Conjugate gradient	(Looney C., 1997)
Gradient descent	(Looney C., 1997)
Levenberg-Marquardt	(Levenberg, 1944)

Various strategies exist to train ANNs, Table 3, with back-propagation being one of the most popular training practices (Rumelhart & McClelland, 1986). Back-propagation assigns error through the network to each neuron, and then it adjusts the neurons that account for more error more aggressively than other neurons with the process iterating to minimize a cost function, such as MSE, or until a maximum epoch point is reached.

Recommended ANN Procedure

For best results, experimental procedures are still required to determine the optimal number of neurons, hidden layers, architecture, training settings, and data splitting scheme. Finding the ‘right’ parameters for an ANN is important for robustness. A generalized procedure widely used in practice, is shown in Figure 6 and described further in (Young W., 2010). Additionally, design of experiments approaches have also been proposed (Chiu, Cook, Pignatiello, & Whittaker, 1997) to determine optimal ANN settings. In general, three phases, including data pre-processing, training, and testing, are considered.

Figure 6. Procedural flow for ANN training (Young, Holland, & Weckman, 2008)



Example Business ANN Analysis Using JMP13

For an example analysis, bank marketing buyer analysis data from (Moro, Cortez, & Rita, 2014) will be used. Here, 20 input variables are considered with an output binary variable Y determining if the client subscribed to a long term deposit (“yes”) or not (“no”). As described in (Moro, Cortez, & Rita, 2014), the 20 X variables include numeric and categorical information, including demographics, credit, type and frequency of contact between bank and client, and external market data.

For example, ANN analysis, the Neural Network toolbox of JMP13 Pro (SAS Inc.) will be used for one example iteration of the process in Figure 6. For analysis, one selects the Y variable as the response and all other columns as the *Factors*. Settings were specified as described above with 20% of the data sequestered for validation, $N_{Hidden} = 4$ based on $a = 2$, hyperbolic tangential nodes, and one hidden layer.

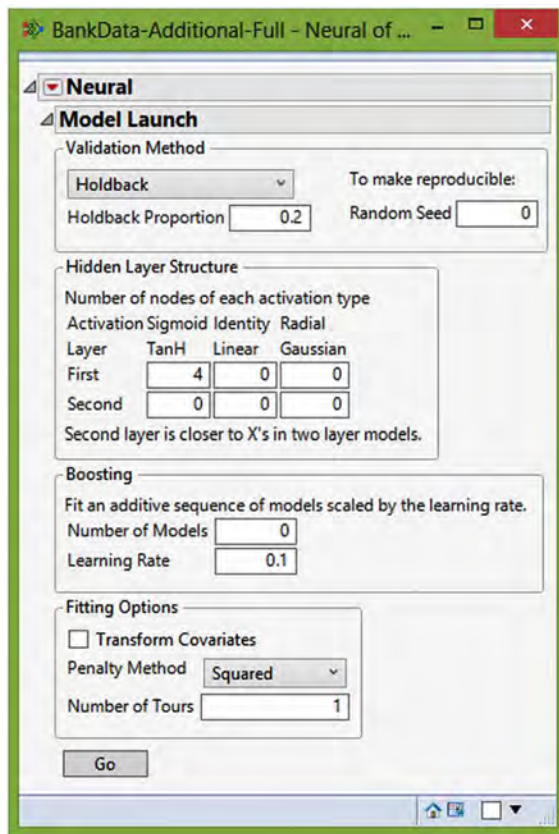
Results, seen in Figure 8, mirror those of (Moro, Cortez, & Rita, 2014) where the training/testing accuracy and validation accuracy are similar with receiver operating characteristic (ROC) curve AUCs being 0.941, thus indicating that the data is effectively modeled through the ANN and accurate when predicting the response of future customers.

APPLICATIONS OF ANNs IN BUSINESS

Domains

A variety of business domains employ ANNs as the authors noted and listed in (Young, Bihl, & Weckman, 2014); the authors present Table 4 as an update on their prior work where only selected references since 2014 are included. Incidentally, ANNs are not universally employed across business domains. Some domains, such as fashion,

Figure 7. JMP13 Neural Network Setup and Initialization



require very fast response times and offer short; therefore, ANNs are largely unexplored in this area (Nenni, Giustiniano, & Pirolo, 2013).

Big Data Implications

Training of an ANN is computational demanding; however, once trained, the underlying ANN model is not computationally costly to use. When analyzing 'Big Data,' many challenges exist as documented by the authors in (Bihl, Young, & Weckman, 2015). As a strategy, developing ANNs on a subset of the data and then retraining the model as needed would be beneficial. Additionally, many large datasets contain more features than necessary and non-salient features can degrade performance; thus, feature selection and dimensionality reduction analysis (DRA) are needed

Table 4. Recent applications of ANN for business purposes

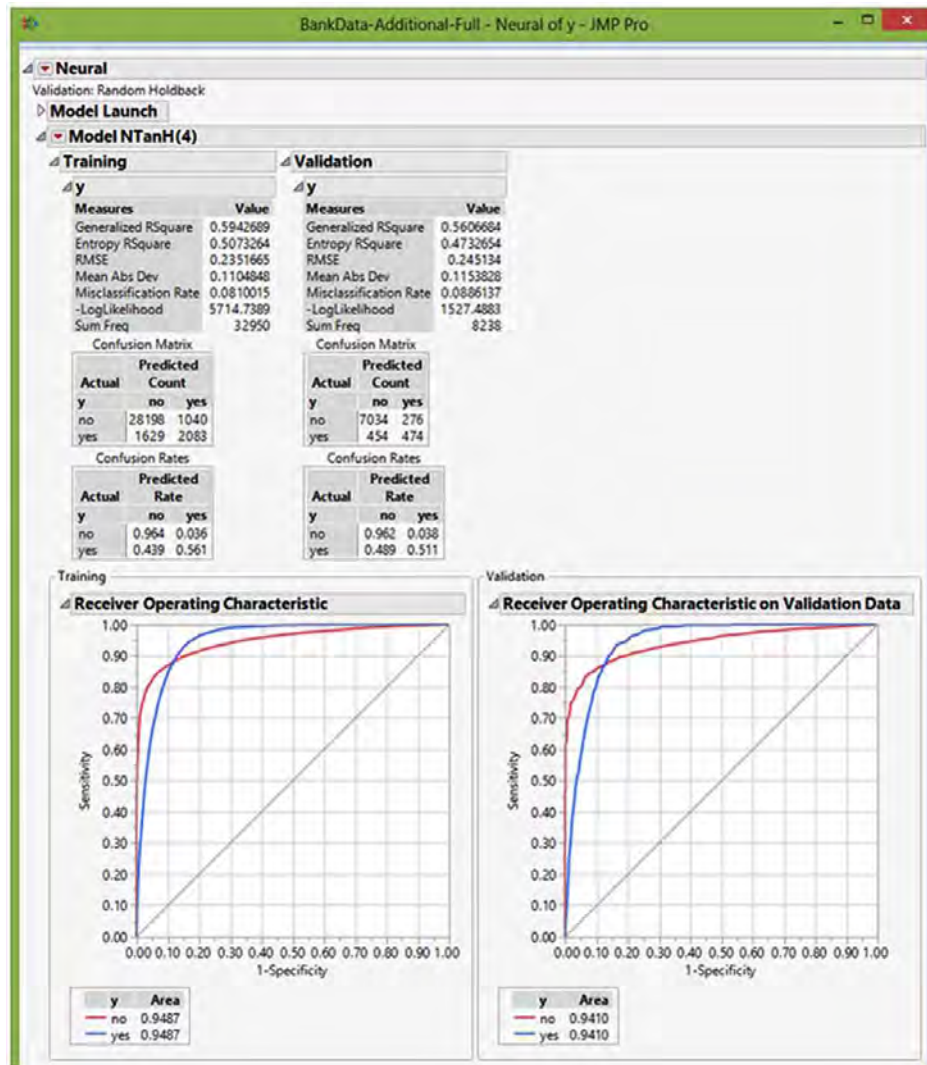
Discipline	References
Auditing	(Chen, Chi, & Wang, 2015)
Buyer Analysis	(Moro, Cortez, & Rita, 2014)
Bankruptcy	(Jiang, 2015) (Sharma, 2012)
Credit	(Grace & Williams, 2016)
Forecasting	(Sharma, 2012) (Maithili, Kumari, & Rajamanickam, 2012)
Human Resources	(Napoli, 2015)
Insurance	(Hassan & Abraham, 2016)
Loans	(Adewusi, Oyedokun, & Bello, 2016)
Marketing and Sales	(Moro, Cortez, & Rita, 2014)
Production and Manufacturing	(Opritescu & Volk, 2015) (Betiku & Taiwo, 2015)
Real Estate	(Zhang, Gao, Seiler, & Zhang, 2015)
Retail	(Machado & al., 2015)
Search Engines	(LeCun, Bengio, & Hinton, 2015)
Stock Markets	(Sen & Das, 2015) (Weckman & al., 2016)

to make quality ANNs. DRA methods include input reduction methods (sensitivity, correlation studies, etc.) (Young W., Weckman, Thompson, & Brown, 2008; Hernandez, Nesic, Weckman, & Ghai, 2006; Bihl, Bauer, & Temple, 2016), and ANN-based feature screening methods (Bauer, Alsing, & Greene, 2000; Verikas & Bacauskiene, 2002), to find a parsimonious set of features.

Small Data Implications

Limited, or scarce, data is also a serious problem that often impedes the development of efficient ANNs due to ANN training procedures requiring the partitioning of data into subsets to prevent over-fitting (Porter, et al., 2001). Since the cross-validation dataset acts as an additional testing set, some researchers cannot afford to hold out a separate independent testing set (Lendassea, Simonb, Wertzc, & Verleysen, 2005).

Figure 8. ANN results using JMP13



Based on these training and data partitioning philosophies utilized within the application of ANN design, larger datasets are at times more idea than smaller ones. However, issues exist when small datasets or imbalanced datasets are analyzed. To overcome these limitation, some methods do exist such as under-sampling, over-sampling and artificial sampling methods. Over-sampling, or boosting, is a simple or naïve strategy that simply reproduces sampled data to improve the predictive accuracy of ANN when sample sizes are small (Seiffert & et al., 2008).

Under-sampling, involves making the training data groups as equal as possible to avoid biasing (Seiffert & et al., 2008). Artificial sampling provides uses real sample information to create additional synthetic samples to increase sample size for training (Bui, 2004). While all strategies have been utilized successfully in practice, 1) more research is needed for robust comparisons and 2) simpler methods, such as under-sampling, perform very well in practice when compared to more complicated methods (Seiffert & et al., 2008; Long & Servedio, 2010).

Knowledge Extraction

Practitioners employ the KDD process for a variety of reasons; one being to understand complex systems better (Browne, et al., 2003). Knowledge extraction (KE) techniques are often applied to mathematical models in order to gain insight. ANNs, though powerful in their prediction and applicability, are still viewed today as black-box model (Lek & Guegan, 1999). To eliminate the black-box mentality, three classification approaches are used to extract knowledge from an ANN. These methods include decomposition, pedagogical, and eclectic methods.

Decomposition methods use the internal structure of the ANN architecture to derive a primitive form of KE. One technique used to inspect a network's activity visually is the NID, as seen in Figure 3 (Özesmi & Özesmi, 1999). Pedagogical methods do not use the internal structure to derive or extract knowledge from a network. Instead, pedagogical methods use the network's map of input to output relationships and reformulate the relationships via another representation, e.g. decision trees (Schmitz, Aldrich, & Gouws, 1999; Young W., Weckman, Hari, Whiting II, & Snow, 2012). Finally, eclectic methods utilize a collection of decomposition and pedagogical methods, as seen in Validity Interval Analysis (VIA) method (Thrun, 1995).

Deep Learning ANNs

Deep learning methods employ multiple layers that aim to model high-level abstractions (LeCun, Bengio, & Hinton, 2015). Deep learning ANNs employ a variety of ANN variants to learn highly complex and abstract patterns, such as search engine queries and ideal responses (LeCun, Bengio, & Hinton, 2015). However, despite the many advantages of deep learning ANNs, issues can exist in their unchecked development. One example is the failure (unwarranted associations) of Google's

Photo App whereby it tagged black people as "gorillas" (Tramer & et al., 2015) (Guynn, 2015). However, it should be noted, that these issues could manifest in any pattern recognition algorithm due to biases and limitations in the data, see (Dreyfus & Dreyfus, 1992).



CONCLUSION

The authors presented a brief primer on ANNs, which summarizes issues related to ANN background, architecture selection, model building, applications, and example implementation. Since publishing an earlier version of this paper (Young, Bihl, & Weckman, 2014), ANNs have increased in business applications, but are still not as well received as simpler methods, e.g. (Nenni, Giustiniano, & Pirolo, 2013).

FUTURE RESEARCH DIRECTIONS

Another emerging direction is the use of deep learning; however, issues with deep learning show that it is not a solution to all problems. It should be noted that additional reading has been provided in Table 4, which provides a brief summary listing of business disciplines. Additionally, readers are directed to summary papers by Li (1994), Wong et al. (1997), Vellido et al. (1999), Smith and Gupta (Smith & Gupta, 2000), Haikmpoor et al. (2011), and Tkáč and Verner (2016), as extended reviews of ANN applications in business.

REFERENCES

Adewusi, A., Oyedokun, T., & Bello, M. (2016). Application of artificial neural network to loan recovery prediction. *International Journal of Housing Markets and Analysis*, 9(2), 222–238. doi:10.1108/IJHMA-01-2015-0003

- Balabin, R. M., Safieva, R. Z., & Lomakina, E. I. (2009). Universal technique for optimization of neural network training parameters: Gasoline near infrared data example. *Neural Computing & Applications*, 18(6), 557–565. doi:10.1007/s00521-008-0213-3
- Bauer, K. W. Jr, Alsing, S. G., & Greene, K. A. (2000). Feature screening using signal-to-noise ratios. *Neurocomputing*, 31(1-4), 29–44. doi:10.1016/S0925-2312(99)00147-2
- Beck, N., King, G., & Zeng, L. (2004, May). Theory and Evidence in International Conflict: A Response to de Marchi, Gelpi, and Grynaviski. *The American Political Science Review*, 98(2), 379–389. doi:10.1017/S0003055404001212
- Betiku, E., & Taiwo, A. E. (2015). Modeling and optimization of bioethanol production from breadfruit starch hydrolyzate vis-à-vis response surface methodology and artificial neural network. *Renewable Energy*, 74, 87–94. doi:10.1016/j.renene.2014.07.054
- Bihl, T. J. (2015). *Feature Selection and Classifier Development for Radio Frequency Device Identification*. Air Force Institute of Technology.
- Bihl, T. J., Bauer, K., & Temple, M. A. (2016). *A Comparison of Feature Selection Methods for RF-DNA Fingerprinting Using ZigBee Device Emissions*. *IEEE Transactions on Information Forensics and Security*.
- Bihl, T. J., Young, W. A. II, & Weckman, G. R. (2015). Defining, Understanding and Addressing Big Data. *International Journal of Business Analytics*, 3(2), 1–32. doi:10.4018/IJBAN.2016040101
- Browne, A., Hudson, B., Whitley, D., Ford, M., Picton, P., & Kazemian, H. (2003). Knowledge extraction from neural networks. *Proceedings of the 29th Annual Conference of the IEEE Industrial Electronics Society*, 1909-1913.
- Bui, T. D. (2004). *Neural network analysis of sparse datasets: An application to the fracture system in folds of the Lisburne formation, North-eastern Alaska* (Ph.D. Dissertation). Texas A&M University, Department of Petroleum Engineering, College Station, TX.
- Chen, F., Chi, D., & Wang, Y. (2015). Detecting biotechnology industry's earnings management using Bayesian network, principal component analysis, back propagation neural network, and decision tree. *Economic Modelling*, 46, 1–10. doi:10.1016/j.econmod.2014.12.035
- Chiu, C.-C., Cook, D. F., Pignatiello, J. P., & Whittaker, A. D. (1997). Design of a radial basis function neural network with a radius-modification algorithm using response surface methodology. *Journal of Intelligent Manufacturing*, 8(2), 117–124. doi:10.1023/A:1018504704266
- Clark, J. W. (1999). Neural networks: New tools for modeling and data analysis. In *Scientific Applications of Neural Nets* (pp. 1–96). Springer Berlin Heidelberg. doi:10.1007/BFb0104277
- de Marchi, S., Gelpi, C., & Grynaviski, J. D. (2004, May). Untangling Neural Nets. *The American Political Science Review*, 92(2), 371–378. doi:10.1017/S0003055404001200
- Dewdney, A. K. (1997). The Apprentice Builds a Brain: Misled by Metaphors. In *Yes, We Have No Neutrons*. John Wiley & Sons.
- Dreyfus, H., & Dreyfus, S. (1992). What artificial experts can and cannot do. *AI & Society*, 6(1), 18–26. doi:10.1007/BF02472766
- Duch, W., & Jankowski, N. (1997). New Neural Transfer Functions. *Neural Computing Surveys*, 7, 639–658.
- Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern Classification* (2nd ed.). John Wiley & Sons.

- Gao, P., Chen, C., & Qin, S. (2010). An optimization method of hidden nodes for neural network. *International Workshop on Education Technology and Computer Science (ETCS)*, 53-56. doi:10.1109/ETCS.2010.300
- Grace, A., & Williams, S. (2016). Comparative Analysis of Neural Network and Fuzzy Logic Techniques in Credit Risk Evaluation. *International Journal of Intelligent Information Technologies*, 12(1), 47-62. doi:10.4018/IJIT.2016010103
- Grünwald, P. (2001). Model selection based on minimum description length. *Journal of Mathematical Psychology*, 44(1), 133-152. doi:10.1006/jmps.1999.1280 PMID:10733861
- Guynn, J. (2015, July). Google photos labeled black people 'gorillas'. *USA Today*.
- Hakimpoor, H., Bin Arshad, K. A., Tat, H. H., Khani, N., & Rahmandoust, M. (2011). Artificial neural networks' applications in management. *World Applied Sciences Journal*, 14(7), 1008-1019.
- Hand, D. J., Daly, F., McConway, K., Lunn, D., & Ostrowski, E. (1993). *A Handbook of Small Data Sets*. CRC Press.
- Hassan, A. K., & Abraham, A. (2016). Modeling Insurance Fraud Detection Using Imbalanced Data Classification. *Advances in Nature and Biologically Inspired Computing*, 117-127.
- Haykin, S. (1999). *Neural Networks: A Comprehensive Foundation*. Prentice Hall.
- Hernandez, S., Nesic, S., Weckman, G., & Ghai, V. (2006). Use of Artificial Neural Networks for Predicting Crude Oil Effect on CO₂ Corrosion of Carbon Steels. *Corrosion Journal*, 62(6).
- Jain, A. K., Duin, R. P., & Mao, J. (2000, January). Statistical Pattern Recognition: A Review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1), 4-37. doi:10.1109/34.824819
- Jain, A. K., Mao, J., & Mohiuddin, K. M. (1996). Artificial neural networks: A tutorial. *Computer*, 29(3), 31-44. doi:10.1109/2.485891
- Jiang, X. (2015). *Corporate Bankruptcy Prediction Using Artificial Neural Network* (PhD Dissertation). Baylor University.
- Kennard, R. W., & Stone, L. A. (1969). Computer aided design of experiments. *Technometrics*, 11(1), 137-149. doi:10.1080/00401706.1969.10490666
- Kurkova, V. (1992). Kolmogorov's Theorem and Multilayer Neural Networks. *Neural Networks*, 5(3), 501-506. doi:10.1016/0893-6080(92)90012-8
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444. doi:10.1038/nature14539 PMID:26017442
- Lek, S., & Guegan, J. (1999). Artificial neural networks as a tool in ecological modeling, an introduction. *Ecological Modelling*, 120(2-3), 65-73. doi:10.1016/S0304-3800(99)00092-7
- Lendassea, A., Simonb, G., Wertz, V., & Verleysen, M. (2005). Fast bootstrap methodology for regression model selection. *Neurocomputing*, 64, 161-181. doi:10.1016/j.neucom.2004.11.017
- Levenberg, K. (1944). A Method for the Solution of Certain Non-Linear Problems in Least Squares. *Quarterly of Applied Mathematics*, 2(2), 164-168. doi:10.1090/qam/10666
- Li, E. Y. (1994). Artificial neural networks and their business applications. *Information & Management*, 27(5), 303-313. doi:10.1016/0378-7206(94)90024-8

- Li, E. Y. (1994). Artificial neural networks and their business applications. *Information & Management*, 27(5), 303–313. doi:10.1016/0378-7206(94)90024-8
- Lindsey, J. C., Herzberg, A. M., & Watts, D. G. (1987). A method for cluster analysis based on projections and quantile-quantile plots. *Biometrics*, 43(2), 327–341. doi:10.2307/2531816
- Liu, X., Wu, J., & Zhou, Z. (2009). Exploratory undersampling for class-imbalance learning. *IEEE Transactions on Systems, Man, and Cybernetics. Part B, Cybernetics*, 39(2), 539–550. doi:10.1109/TSMCB.2008.2007853 PMID:19095540
- Long, P. M., & Servedio, R. A. (2010). Random classification noise defeats all convex potential boosters. *Machine Learning*, 78(3), 287–304. doi:10.1007/s10994-009-5165-z
- Looney, C. (1997). *Pattern Recognition Using Neural Networks*. Oxford University Press.
- Machado, M., Gomes, L. F. A. M., Santos, D. J., & Caldeira, A. M. (2015). Using a Bipolar Choquet Neural Network to Locate a Retail Store. *Procedia Computer Science*, 55, 741–747. doi:10.1016/j.procs.2015.07.163
- Maithili, A., Kumari, R. V., & Rajamanickam, S. (2012). Neural network towards business forecasting. *IOSR Journal of Engineering*, 2(4), 831–836. doi:10.9790/3021-0204831836
- Mannila, H. (1996). Data mining: machine learning, statistics, and databases. *Eight International Conference on Scientific and Statistical Database Management*, 1-8.
- MATLAB. (2010). *MATLAB version 7.10.0 (R2010a)*. Natick, MA: MathWorks Inc.
- May, R. J., Maier, H. R., & Dandy, G. C. (2010). Data splitting for artificial neural networks using SOM-based stratified sampling. *Neural Networks*, 23(2), 283–294. doi:10.1016/j.neu-net.2009.11.009 PMID:19959327
- McCulloch, W., & Pitts, W. (1943). A Logical Calculus of the Ideas Immanent in Nervous Activity. *The Bulletin of Mathematical Biophysics*, 5(4), 115–133. doi:10.1007/BF02478259
- Mindrup, F. M., Friend, M. A., & Bauer, K. W. (2012). Small sample training and test selection method for optimized anomaly detection algorithms in hyperspectral imagery. *Journal of Applied Remote Sensing*, 6(1), 063563-1. doi:10.1117/1.JRS.6.063563
- Moro, S., Cortez, P., & Rita, P. (2014). A Data-Driven Approach to Predict the Success of Bank Telemarketing. *Decision Support Systems*, 62, 22–31. doi:10.1016/j.dss.2014.03.001
- Myung, J. (2000). Journal of Mathematical Psychology. *The Importance of Complexity in Model Selection*, 44(1), 190-204.
- Napoli, C. e. (2015). Toward work groups classification based on probabilistic neural network approach. *Artificial Intelligence and Soft Computing*, 79-89.
- Nenni, M. E., Giustiniano, L., & Pirolo, L. (2013). Demand forecasting in the fashion industry: A review. *International Journal of Engineering Business Management*, 5, 1–6.
- Neuralpower. (2007). Retrieved from Neuralpower Technology: <http://www.neuralpower.com/images/tecnology/figure1.gif>
- NeuroDimensions. (2005, Dec). *Newsletters*. Retrieved 2009, from NeuroDimensions: <http://www.neurosolutions.com/newsletters.html>
- Olden, J. D., & Jackson, D. A. (2002). Illuminating the black box: A randomization approach for understanding variable contributions in artificial neural networks. *Ecological Modelling*, 154(1-2), 135–150. doi:10.1016/S0304-3800(02)00064-9

- Opritescu, D., & Volk, W. (2015). Automated driving for individualized sheet metal part production—A neural network approach. *Robotics and Computer-integrated Manufacturing*, 35, 144–150. doi:10.1016/j.rcim.2015.03.006
- Özesmi, S., & Özesmi, U. (1999). An artificial neural network approach to spatial habitat modeling with interspecific interaction. *Ecological Modelling*, 116(1), 15–31. doi:10.1016/S0304-3800(98)00149-5
- Poh, H. L. (1991). *A Neural Network Approach for Marketing Strategies Research and Decision Support* (PhD Dissertation). Stanford University.
- Porter, C., O'Donnell, C., Crawford, E., Gamito, E., Errejon, A., Genega, E., & Tewari, A. et al. (2001). Artificial Neural Network Model To Predict Biochemical Failure After Radical Prostatectomy. *Molecular Urology*, 5(4), 159–162. doi:10.1089/10915360152745830 PMID:11790277
- Rumelhart, D., & McClelland, J. (1986). *Parallel Distributed Processing*. Cambridge, MA: MIT Press.
- Sall, J., Lehman, A., Stephens, M., & Creighton, L. (2012). *JMP start statistics: a guide to statistics and data analysis using JMP*. SAS Institute.
- Schmitz, G., Aldrich, C., & Gouws, F. (1999). ANN-DT: An algorithm for extraction of decision trees from artificial neural networks. *IEEE Transactions on Neural Networks*, 10(6), 1392–1401. doi:10.1109/72.809084 PMID:18252640
- Schunn, C. D., & Wallach, D. (2005). Evaluating goodness-of-fit in comparison of models to data. In W. Tack (Ed.), *Psychologie der Kognition* (pp. 115–154). Reden and Vorträge anlässlich der Emeritierung.
- Seiffert, C. (2008). A comparative study of data sampling and cost sensitive learning. *IEEE International Conference on Data Mining Workshops*, 46–52. doi:10.1109/ICDMW.2008.119
- Sen, P. K., & Das, D. (2015). Artificial Neural Network Applications for Predicting Volatility in Stock Markets: A Review. *The MA Journal*, 50(10), 108–112.
- Sharma, V. (2012). Artificial neural network applicability in business forecasting. *International Journal of Emerging Research in Management & Technology*, 62–65.
- Smith, K. A., & Gupta, J. N. (2000). Neural networks in business: Techniques and applications for the operations researcher. *Computers & Operations Research*, 27(11), 1023–1044. doi:10.1016/S0305-0548(99)00141-0
- Snee, R. D. (1977). Validation of regression models: Methods and examples. *Technometrics*, 19(4), 415–428. doi:10.1080/00401706.1977.10489581
- Team, R. D. (2008). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.
- Thrun, S. (1995). Advances in Neural Information Processing Systems: Vol. 7. *Extracting rules from artificial neural networks with distributed representations*. Cambridge, MA: MIT Press.
- Timmerman, D., Verrelst, H., Bourne, T. H., De Moor, B., Collins, W. P., Vergote, I., & Vandewalle, J. (1999). Artificial neural network models for the preoperative discrimination between malignant and benign adnexal masses. *Ultrasound in Obstetrics & Gynecology*, 13(1), 17–25. doi:10.1046/j.1469-0705.1999.13010017.x PMID:10201082
- Tkáč, M., & Verner, R. (2016). Artificial neural networks in business: Two decades of research. *Applied Soft Computing*, 38, 788–804. doi:10.1016/j.asoc.2015.09.040
- Tramer, F. (2015). *Discovering Unwarranted Associations in Data-Driven Applications with the FairTest Testing Toolkit*. arXiv preprint.

- Vellido, A., Lisboa, P. J., & Vaughan, J. (1999). Neural networks in business: A survey of applications. *Expert Systems with Applications*, 17(1), 51–70. doi:10.1016/S0957-4174(99)00016-0
- Verikas, A., & Bacauskiene, M. (2002). Feature selection with neural networks. *Pattern Recognition Letters*, 23(11), 1323–1335. doi:10.1016/S0167-8655(02)00081-8
- Weckman, G., Dravenstott, R. W., Young, W. A. II, Ardjmand, E., Millie, D. F., & Snow, A. P. (2016). A Prescriptive Stock Market Investment Strategy for the Restaurant Industry using an Artificial Neural Network Methodology. *International Journal of Business Analytics*, 3(1), 1–21. doi:10.4018/IJBAN.2016010101
- Weckman, G., Millie, D., Ganduri, C., Rangwala, M., Young, W., Rinder, M., & Fahnenstiel, G. (2009). Knowledge Extraction from the Black Box in Ecological Monitoring. *Journal of Industrial and Systems Engineering*, 3(1), 38–55.
- Wong, B. K., Bodnovich, T. A., & Selvi, Y. (1997). Neural network applications in business: A review and analysis of the literature (1988-1995). *Decision Support Systems*, 19(4), 301–320. doi:10.1016/S0167-9236(96)00070-X
- Wu, W., May, R., Dandy, G. C., & Maier, H. R. (2012). A method for comparing data splitting approaches for developing hydrological ANN models. *International Congress on Environmental Modelling and Software Managing Resources of a Limited Planet*. Retrieved from <http://www.iemss.org/society/index.php/iemss-2012-proceedings>
- Yao, X. (1999). Evolving artificial neural networks. *Proceedings of the IEEE*, 87(9), 1423–1447. doi:10.1109/5.784219
- Yao, X., & Liu, Y. (1997). A New Evolutionary System for Evolving Artificial Neural Networks. *IEEE Transactions on Neural Networks*, 8(3), 694–713. doi:10.1109/72.572107 PMID:18255671
- Yao, X., & Liu, Y. (1998). Towards designing artificial neural networks by evolution. *Applied Mathematics and Computation*, 91(1), 83–90. doi:10.1016/S0096-3003(97)10005-4
- Young, W. (2010). *A Team-Compatibility Decision Support System to Model the NFL Knapsack Problem: An Introduction to HEART* (PhD Dissertation). Athens, OH: Ohio University.
- Young, W., Holland, W., & Weckman, G. (2008). Determining Hall of Fame Status for Major league Baseball Using an Artificial Neural Network. *Journal of Quantitative Analysis in Sports*, 4(4). doi:10.2202/1559-0410.1131
- Young, W., Weckman, G., Hari, V., Whiting, I. I. H., & Snow, A. (2012). Using artificial neural networks to enhance CART. *Neural Computing & Applications*, 21(7), 1477–1489. doi:10.1007/s00521-012-0887-4
- Young, W., Weckman, G., Thompson, J., & Brown, M. (2008). Artificial Neural Networks for Knowledge Extraction of Concrete Shear Strength Prediction. *International Journal of Industrial Engineering - Theory, Applications, and Practice*, 15(1).
- Young, W. A., Bihl, T. J., & Weckman, G. R. (2014). Artificial Neural Networks for Business Analytics. In J. Wang (Ed.), *Encyclopedia of Business Analytics and Optimization* (pp. 193–208). doi:10.4018/978-1-4666-5202-6.ch019
- Zhang, G. P. (2007). Avoiding Pitfalls in Neural Network Research. *IEEE Transactions on Systems, Man, and Cybernetics - Part C*, 37(1), 3–16. doi:10.1109/TSMCC.2006.876059
- Zhang, H., Gao, S., Seiler, M., & Zhang, Y. (2015). Identification of Real Estate Cycles in China Based on Artificial Neural Networks. *Journal of Real Estate Literature*, 23(1), 65–84.

KEY TERMS AND DEFINITIONS

Back-Propagation: A supervised learning method used to determine the weights of an ANN, where the difference between the desired and the model's output is minimized.

Epoch: The representation of an entire training set of sample data through the learning algorithm so that an ANN's weights can be determined.

Knowledge Extraction: The process of discovering how input attributes are used within an ANN to formulate the output such that one can validate functional relationships within the model.

Neuron: An individual building block of an ANN in which weighted input values are transformed via a transfer function into an output, which is typically passed to other portions of the network.

Over-Fitting: Occurs when a mathematical model describes random error or noise instead of the real underlying relationships within a dataset, which artificially produces desirable goodness of fit metrics for training data, but produces poor metrics for testing data.

Post-Processing: A process of utilizing a trained mathematical model in order to improve the understanding of the database that has been modeled.

Pre-Processing: A process of preparing a dataset in order to develop a mathematical model.

Supervised Learning: A learning strategy in which the desired output, or dependent attribute, is known.

Unsupervised Learning: A learning strategy in which the desired output, or dependent attribute, is unknown.

Recurrent Neural Networks for Predicting Mobile Device State

Juan Manuel Rodriguez

ISISTAN, UNICEN-CONICET, Argentina

Alejandro Zunino

ISISTAN, UNICEN-CONICET, Argentina

Antonela Tommasel

ISISTAN, UNICEN-CONICET, Argentina

Cristian Mateos

ISISTAN, UNICEN-CONICET, Argentina

INTRODUCTION

Smartphones have become a key part of everyday life as an essential tool for their users. People have fully integrated mobile devices into their lives by using them to communicate with friends, check e-mails, play games, record physical activities and take pictures, among other possible uses. Moreover, smartphones are equipped with several features (WiFi, GPS, and Bluetooth among others) that can record activities and contextual information, such as location, application usage, and even messaging and calling behaviour. Hence, smartphones are interesting options for tracking and mining user behaviour in daily life (Do and Gatica-Perez, 2014). This information offers new opportunities to analyse human behaviour aiming at enhancing the user experience with mobile devices and, at the same time, helping to ease the use of smartphones' services (Rios et al., 2014).

Several domains leverage on the prediction of mobile devices' states (Pejovic and Musolesi, 2015; Niroshinie et al., 2013; Ravi et al., 2008). For example, in the development of context aware applications, the predictions could be useful for determining the context in which applications

are running. Such context aware applications are encompassed in a concept called "Anticipatory Mobile Computing" (AMC) (Pejovic and Musolesi, 2015). The goal of AMC is deciding which actions should be taken based on predicted future states to improve the outcome. AMC concepts are present in personal assistance technology (such as Google Now, Microsoft Cortana or Siri), healthcare applications and smart cities. Personal assistance technology uses state prediction to provide relevant information to the user before such information is requested. For example, predicting that the hour in which the user goes to work allows personal assistance technology to provide information about traffic and weather, which might be relevant to the user.

Other use of predictions can be found in mobile cloud computing (Niroshinie et al., 2013). One of its key proposals is moving computing from mobile devices to the cloud to reduce battery consumption. However, to effectively reduce battery consumption, it is necessary to predict whether the energy requirements for communicating are lower than those of processing. In addition, if the mobile device is not going to be connected to the Internet when the cloud finishes its work,

computation results will be unavailable. This might lead to the repetition of the computation in the mobile device, which would waste more energy than if the mobile device had performed the computation in the first place. These are just a few examples of the developing mobile device state prediction techniques' importance.

The current state of mobile devices can be regarded as the consequence of the previous states. Consequently, future behaviour can be predicted based on how a user has been using his/her device. The generation of predictive models of human behaviour has emerged as a topic of interest in several areas, such as recommendation systems, context-aware services, and personalised and adaptive interfaces. For example, several studies have focused on predicting the probability of users to be at a particular place at a given time. Also, Do and Gatica-Perez (2014) and Liao et al. (2012) aimed at predicting which application the user will use next based on contextual information to reduce the time users spend on searching for a specific application.

Although mobile devices constantly evolve as they provide increasing functionality due to the improvements in processing power, storage capabilities, graphics and connectivity; battery capacities do not experience the same growth (Ravi et al., 2008). Power emerges as a critical resource for battery-powered systems as mobile devices. Hence, battery management becomes a crucial requirement to users. Providing battery management information requires the ability to accurately predict remaining battery life in a dynamically changing system. Interestingly, most of the studies in the literature focus on offering location prediction, or application personalisation, instead of analysing the impact of user behaviour in battery level and life. In this context, this work evaluates the suitability of Recurrent Neural Networks (RNN) for predicting future battery levels of mobile devices based on the users' usage pattern of different features, such as the WiFi connection or screen status, among others.

BACKGROUND

The analysis of human behaviour through smartphone usage has attracted considerable interest in recent years. Works on smartphone mining include physical activity recognition by applying learning techniques to accelerometer data, and personalisation of content and user interface. For example, Do and Gatica-Perez (2014) predicted human behaviour based on smartphone sensors using statistical methods commonly used in signal processing for forecasting time series. Particularly, user location within a ten-minute window was predicted along with the applications users were more likely to use based on the location, time, Bluetooth proximity and communication logs. The approach linearly combined least-squares linear regression, logistic regression and Markov models. Location prediction results showed that the most important predictors were Bluetooth proximity, location and time, and confirmed the dependency between human mobility and social interactions.

Application usage prediction was also studied in (Shin et al., 2012; Liao et al., 2012). Both approaches aimed at presenting users with the most probable applications to be used aiming at reducing the time spent searching for a desired application. Shin et al. (2012) based the prediction on a probabilistic Naïve Bayes model that was personalised to each user according to his/her application usage pattern. Prediction was based on information from GPS, cellular network, accelerometer, calls, WiFi, Bluetooth, screen status, battery status, illumination and running applications. Results showed that the previously used applications, location and hour of day were useful predictors.

Liao et al. (2012) created temporal profiles of application usage and proposed a three step probability-based scoring model to compute the probability of using an application. First, the periodicity of application usage was detected by a Discrete Fourier Transform. Second, captured periodicities were hierarchically clustered to

identify users' behaviour. Clusters represented different usage patterns. For example, one cluster corresponded to weekday behaviour, whereas other comprised weekend behaviour. Finally, the possible application usages were predicted based on the Chebyshev's inequality.

Regarding connectivity prediction, Rios et al. (2014) aimed at predicting Bluetooth and WiFi usage based on mining association rules from information regarding data, WiFi and Bluetooth connections, GPS status changes, ringer status (silent, vibrate, normal) changes and headset connection. The rules were able to accurately capture the relations between WiFi connection habits relative to location and GPS status. The authors highlighted the potential of their approach for assisting users in the activation/deactivation of connectivity features to increase battery life.

Finally, closely related to this work are the studies carried out in (Wen et al., 2003; Ravi et al., 2008), which aimed at predicting battery life. Wen et al. (2003) predicted battery based on information regarding variations in workload, application usage, battery charge rates, the charging-cycle effect and a voltage curve. Then, such information was combined into an online linear curve fitting statistical approach. Ravi et al. (2008) also aimed at predicting the next charging opportunity, call time predictor (i.e. how much call-time will the user require over a determined time period), and application battery consumption. With such information the authors intended to inform users not only when they will run out of battery, but also, which applications should they terminate to increase battery life. The analysis considered location, battery status, list of running processes, id of the cell the smartphone is connected to, past calling behaviour, average number of minutes of call time that users need during each hour of the day separated between weekdays and weekends, and the base curve of battery life.

PREDICTING MOBILE DEVICES STATES USING RNN

This work aims at predicting a mobile device future state based on previous states, i.e. predicting sequences of states. Mobile device state changes as the result of an event in time. For example, a change in the battery level can be defined as an event, which triggers a new state of the mobile device in which the battery level is modified, and all other features remain the same. Particularly, this work defines a state as a combination of the following features:

- **Time Between Events:** Minutes passed in-between the current state and the previous one. Type: Integer.
- **Battery Level:** Mobile device current battery level ranging between 0 and 100. Type: Integer.
- **External Energy Supply:** Whether the mobile device is plugged to an external energy supply, e.g. AC adapter or USB connection. Type: Boolean.
- **Screen On/Off:** Whether the mobile device screen is active. Type: Boolean.
- **WiFi:** Whether the mobile device is connected to a WiFi network. Type: Boolean.

To predict future states, RNN will be used. RNN (Williams and Zipser, 1989) differ from feed-forward Artificial Neural Networks (ANN) in that they can have cycles in the computational networks. These cycles allow the network to have dynamic contextual windows, instead of fixed size windows like the standard feed forward models. This dynamic window contributes to the enormous potential of RNN for successfully predicting sequences of events. In this context, the sequence of events models the sequence of states for a mobile device. Although this work does not evaluate the

feasibility of deploying the proposed predictor in a mobile device, a work (Alsharif et al., 2015) on decoding keyboard gesture patterns into text by means of an RNN shows that it is possible to use RNN within a mobile device.

SOLUTIONS AND RECOMMENDATIONS

The proposed ANN is based on the RNN known as Long Short-Term Memory (LSTM) (Hochreiter and Schmidhuber, 1997). RNN were used because they are very effective for predicting sequences (Alsharif et al., 2015) as new predictions depends on past predictions. This allows to use an arbitrary number of previously known states to make a prediction.

The proposed ANN needs at least two consecutive states for predicting the following state. Figure 1 depicts the architectural structure of the

ANN. In a first step, each state is processed by a multi-layer perceptron (MLP). Each layer linearly combines its input vector using a weighting matrix (W) and an internal bias (B), with the combination being processed via

$$L_n(i) = \tanh\left(\left(i_n^T W_n\right)^T + B_n\right),$$

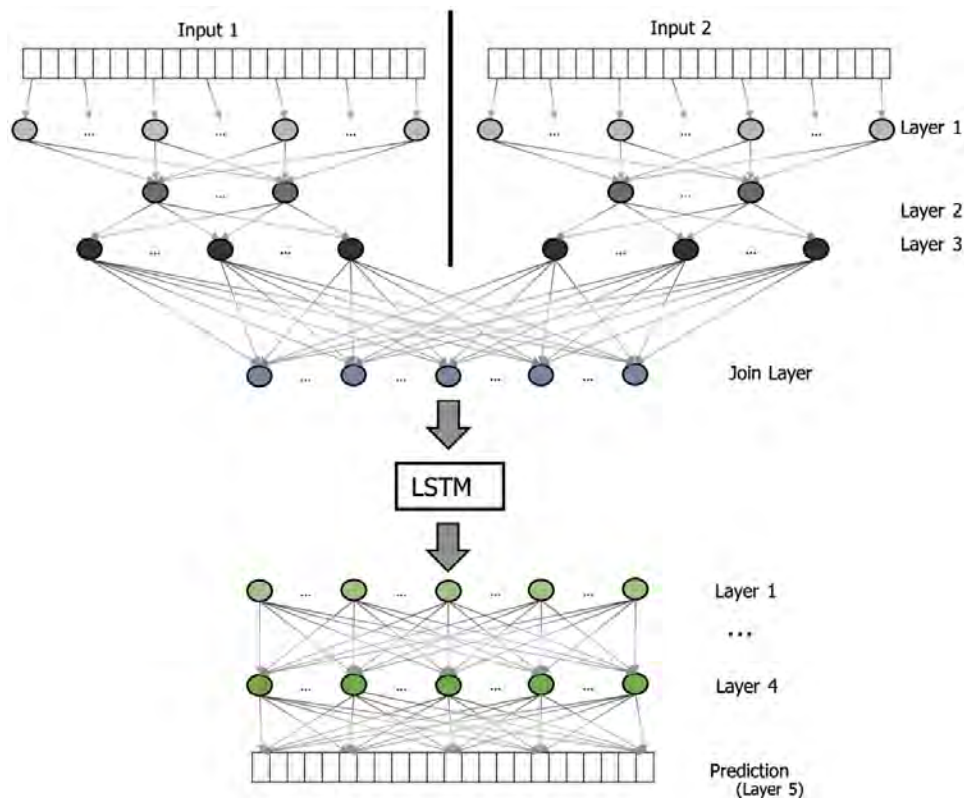
which describes how each layer behaves. Note that functions such as $\tanh$ and sigmoid are applied element wise to the vectors. Then, the different layers are combined in a MLP via

$$L_n\left(L_{n-1}\left(\dots L_1(x)\dots\right)\right).$$

As a result of these operations, the newly computed vector can be used as input for another layer.

Since the proposed ANN uses two different MLP for processing the two states, the results

Figure 1. Neural Network structure



of each MLP need to be combined. For this, the proposed approach uses a layer similar to the layers in MLP, but that accepts several inputs. In the approach this layer is called Join Layer ($JL(i..j)$), which can be defined as

$$\tanh\left(\left(i_n^T W_{JL-i} + \dots + j_n^T W_{JL-j}\right)^T + B_{JL}\right)$$

The Join Layer result is fed to the recurrent layer of the ANN, namely the LSTM layer. This layer is recurrent because it has two variables that take their values from previous outputs: the context and the hidden state. Since the output of this layer depends on previous outputs, the prediction of the proposed ANN depends on previous predictions making it an RNN. The context (c_{t-1}) is the LSTM layer previous output, and the hidden state (s_{t-1}) is the LSTM layer previous internal state. For the first prediction, the context c_0 and the hidden state s_0 are vectors full of zeros. The LSTM layer has three inputs: the current state (x_t), the hidden state and the context. The hidden state s_t depends on s_{t-1} , $g_t(x_t, c_{t-1})$ and $y_t^{in}(x_t, c_{t-1})$, and can be defined as

$$s_t = s_{t-1} + g_t(x_t, c_{t-1}) \circ y_t^{in}(x_t, c_{t-1}),$$

where

$$g_t(x_t, c_{t-1}) = \text{sigmoid}\left(\left(x_t^T W_g^x + c_{t-1}^T W_g^c\right)^T + B_g\right)$$

and

$$y_t^{in}(x_t, c_{t-1}) = \text{sigmoid}\left(\left(x_t^T W_{y^{in}}^x + c_{t-1}^T W_{y^{in}}^c\right)^T + B_{y^{in}}\right).$$

After computing the new hidden state, the LSTM layer output (c_t) can be defined as the dot product $h_t \circ y_t^{out}(x_t, c_{t-1})$, which takes as input

$$h_t = \text{sigmoid}\left(\left(s_t^T W_h\right)^T + B_h\right),$$

and

$$y_t^{out}(x_t, c_{t-1}) = \tanh\left(\left(x_t^T W_{y^{out}}^x + c_{t-1}^T W_{y^{out}}^c\right)^T + B_{y^{out}}\right).$$

Finally, the LSTM layer output is taken as input by other MLP for obtaining the final prediction.

In order to be analysed by the NN, mobile device state features were encoded into a vector of elements, in which each element can adopt either -1 or 1. The selection of such values corresponds to the fact that ANN usually rely on $\text{sigmoid}(i)$ and $\tanh(i)$ functions, which when applied to -1 or 1 result in values close to their asymptotes. Also, applying such functions to 0 results in the mean value of their asymptotes. For example, $\tanh(i)$ has two asymptotes: -1 when $i \rightarrow -\infty$ and 1 when $i \rightarrow \infty$. Applying the function to the selected values results in: $\tanh(-1) \cong -0.7616$, $\tanh(1) \cong 0.7616$, while $\tanh(0) = 0$, i.e. the mean value of the asymptotes. Integer features were encoded using their binary representation. For instance, the “Hour of the day” feature was encoded using the five binary digits needed for representing any number between 0 and $31 = 2^5 - 1$. In the case of the “Time between events” feature, the maximum time was set to 3 hours or 180 minutes. Thus, it was encoded using 8 elements. Finally, for binary features, -1 stands for false, and 1 stands for true, e.g. when the mobile device is connected to a WiFi network the “WiFi” feature is 1, otherwise it is -1.

Figure 2 illustrates the application of the approach for predicting the states from time $n+1$ to m using the previous n states. In a first step, the ANN is fed with the known states x_1 and x_2 resulting in a prediction x_3' . Such step is not performed for obtaining the prediction, as x_3 is known, but for the sake of the ANN internal state. After this step, the ANN is fed with x_2 and x_3 . The ANN is fed repeatedly with the known states until the x_{n+1}' prediction is obtained. At this moment, as no states are known, x_{n+1}' is the first state to be predicted. Unlike the input, in which each element can only adopt two values, the output elements adopt ap-

proximations to the expected values. For instance, if a particular element of x'_{n+1} , is -0.9 it is likely that the real value of x_{n+1} is -1. Hence, x'_{n+1} needs to be pre-processed before it is fed to the ANN. Eq. 1 shows the pre-process function. The sequence prediction algorithm has been proved to be effective in other sequence prediction domains (Sutskever et al., 2014).

$$p(x') = \begin{cases} 1 & x' \leq 0 \\ -1 & \text{otherwise} \end{cases}$$

Experimental Settings

The dataset used in the experimental evaluation was gathered from a mobile device using the application described in (Rios et al., 2014). The mobile device owner was a University employee who volunteer for the study. The mobile device was a Samsung Galaxy SII and the dataset was collected between 5th of June of 2015 and 8th of October of 2015. Since the mobile device owner could turn on and off the logging application at will, the dataset does not have information for every day between those dates. The resulting dataset comprises data belonging to 19,067 records of mobile device states collected during 79 days. Records were grouped into chunks, where each chunk contains states from a single day. However, days in which states were separated by more than 3 hours were split into several chunks. In total, 83 chunks were created, which were split into training and validation splits. The training split comprised 70% of the chunks, which were randomly selected. Table 1 summarises the dataset characteristics.

Table 1. Dataset description

	Full Dataset	Training Split	Validation Split
Days	79	56	23
Chunks	83	59	24
Events	19,067	12,809	6,258
Average	230	217	261
St. Dev.	168	124	246
Min. Size	63	63	75
Max. Size	979	638	979

The proposed ANN has 546,862 trainable parameters, which are denoted by the weights matrices (W), and the biases vectors (B) denote the ANN parameters to be trained. Table 2 shows the size of the matrices and vectors involved in each of the first MLPs. The parameters of the join layer and LSTM are detailed in Table 3 and Table 4, respectively. Finally, the parameters of the MLP that produces the final prediction are presented in Table 5.

To fit the ANN parameters, different approaches were evaluated. Firstly, two error functions were considered, namely Mean Squared Error (MSE) and Cross-Entropy Error (CEE). MSE consists in adding the element wise squared difference between the expected value and the predicted value, as shown in Eq. 2 CEE is used in logistic regression and represents the entropy between the obtained prediction and the expected value. For the expected values being $y \in \{0,1\}$, Eq. 3 presents the cross entropy formula. Since the elements of the expected values are $y_i \in \{0,1\}$, the correction formula $c(x) = (x+1)/2$ was applied element wise

Figure 2. State prediction

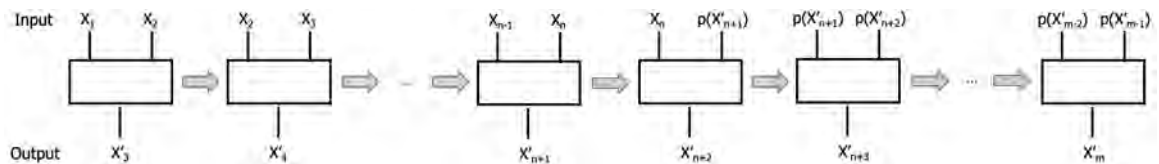


Table 2. First MLP parameters

Parameter	Dimensions	Parameter	Dimensions
W_1	23 x 23	B_1	23
W_2	23 x 40	B_2	40
W_3	40 x 50	B_3	50

Table 3. Joint layer parameters

Parameter	Dimensions
W_{JL-1}	50 x 100
W_{JL-2}	50 x 100
B_{JL}	100

Table 4. LSTM parameters

Parameter	Dimensions	Parameter	Dimensions
W_g^x	100 x 1000	$W_{y^{out}}^x$	100 x 100
W_g^c	100 x 1000	$W_{y^{out}}^c$	100 x 100
B_g	1000	$B_{y^{out}}$	100
$W_{y^{in}}^x$	100 x 1000	W_h	1000 x 100
$W_{y^{in}}^c$	100 x 1000	B_h	100
$B_{y^{in}}$	1000		

Table 5. Final MLP parameters

Parameter	Dimensions	Parameter	Dimensions
W_1	100 x 40	B_1	40
W_2	40 x 23	B_2	23
W_3	23 x 40	B_3	40
W_4	40 x 23	B_4	23
W_5	23 x 23	B_5	23

for both expected values and predictions before computing the Cross Entropy Error.

$$J(y, y') = \frac{\sum (y_i - y'_i)^2}{N}$$

$$J(y, y') = \frac{-\sum (y_i \cdot \log(y'_i) + (1 - y_i) \cdot \log(1 - y'_i))}{N}$$

To avoid overfitting, the impact of adding a regularisation factor to both MSE and CEE was also assessed. The regularisation factor was calculated as the sum of the parameters' L_1 -norm. Finally, two different techniques were evaluated to approximate the ANN parameters. The first one was the well-known Stochastic Gradient Descent. The second one was ADADELTA (Zeiler, 2012), a variation of Stochastic Gradient Descent in which the learning rate is computed based on the historic values of the gradient. The two error functions,

regularised or not, and the techniques for fitting parameters can be freely combined resulting in eight different settings.

Experimental Results

The first experiment aimed at assessing all potential training settings described above. For each setting, the evolution of the ANN up to 50 epochs was evaluated. One-state-in-the-future prediction stands for predicting the state x'_{n+1} using all the states between x_1 and x_n . Since the ANN requires at least two previous states for generating a prediction, the first possible prediction is x'_3 using x_1 and x_2 . In a chunk of m states, the last prediction is x'_m using $(x_1, x_2, \dots, x_{m-1})$.

Firstly, the training dataset was evaluated using MSE, as shown in Figure 3. Within this figure, Figure 3a and Figure 3b present the errors obtained when training the ANN using MSE or CEE, respectively. The settings using Gradient Descent and MSE achieved the lowest errors. Furthermore, the setting without regularisation performed slightly better, which is expected as regularisation is intended to reduce overfitting the training dataset. However, these results alone are not sufficient for evaluating as overfitting is not completely avoided. Hence, the same evaluation was performed using the validation dataset. Figure 4 shows the results for the validation dataset. Particularly, Figure 4a shows the errors for the ANN trained using MSE, and Figure 4b shows the errors for ANN trained using CEE. As expected, the obtained MSE are higher than the errors obtained for the trained dataset. The shape of the curves in Figure 4a and Figure 4b are similar to the ones in Figure 3a and Figure 3b, respectively. This implies that the ANN did not overfit the training data. Moreover, Gradient Descent and MSE without regularisation also achieved the lowest error.

The second experiment was designed for evaluating the sequence prediction capabilities of the ANN. Thus, predictions of up to 10 states

into the future were generated using the validation dataset and the best performing ANN from the previous experiment. Such ANN was the one obtained after the 34th epoch using Gradient Descent and MSE without regularisation. Note that for predicting the n state into the future, the $n-1$ previous predictions were used. For instance, suppose that m states are known and that the state $m+n$ is to be predicted. In this context, the ANN would use as input the following state sequence

$$(x_1, \dots, x_m, p(x'_{m+1}), \dots, p(x'_{m+n-1})).$$

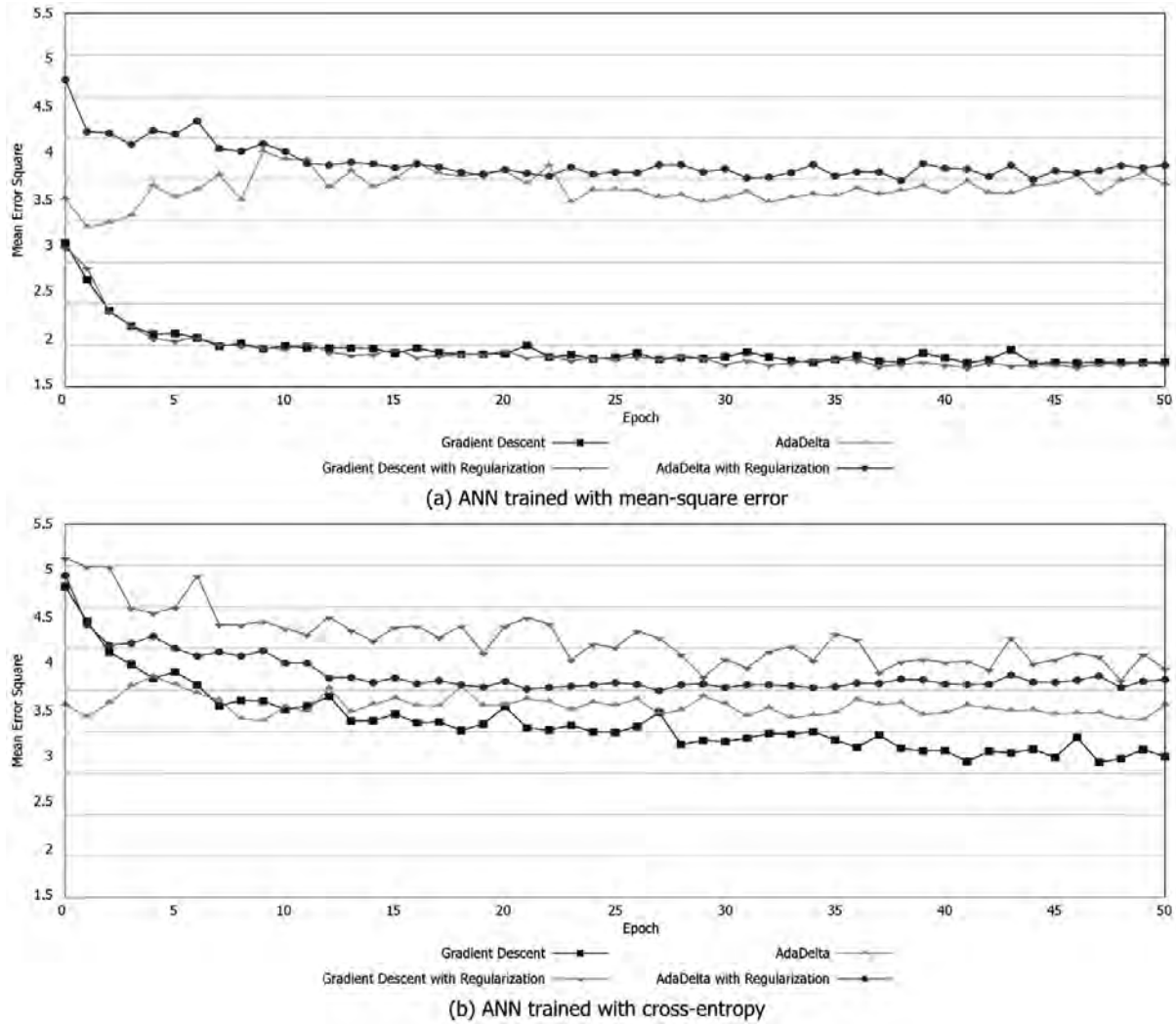
The average time between 10 states is 48'26.82" with a standard deviation of 58'47.27". The minimum registered time between 10 states was 10 minutes and the maximum was 697 minutes.

For the integer type features, namely "Hour", "Time between events", and "Battery", the error was calculated as the Mean Absolute Error (MAE) as defined in Eq. 4. Figure 5 presents the obtained MAE. For "Hour" and "Battery", the MAE increases as the number of states in the future increases, whereas for "Time between events" remains invariant. This is expected due to the fact that "Hour" and "Battery" values are correlative with previous values, while "Time between events" values are not. For instance, if "Battery" has a value of 60% and the battery is not in charging state, it is very unlikely that the next state has a value for "Battery" of 10% or 90%. Hence, an error predicting a "Battery" value would have a heavy impact on further predictions.

$$MAE = \frac{\sum |expected\ value - predicted\ value|}{\# prediction}$$

Since the error in predicting the time between n states is given by the sum of each individual prediction error for "Time between events", the mean accumulated absolute error (MAAE) is defined in Eq. 5. MAAE accounts for the accumulated error in a prediction sequence. In this context, the MAAE for "Time between events"

Figure 3. MSE for training dataset



was 121 minutes. Considering these results, it can be concluded that the ANN did not performed effectively for predicting “Time between events”, but performed with a low error rate for predicting “Hour” and “Battery”.

$$MAAE = \frac{\sum_{i=1}^N |expected\ value_i - predicted\ value_i|}{\# prediction}$$

The relation between the “Hour” and “Battery” features’ Absolute Error (AE) and the “Time between events” feature Accumulated Absolute Error (AAE) for the 10-state-in-the-future prediction was studied to determine whether failing at

predicting the “Time between events” is correlated to errors in the prediction of other features. Table 6 presents the Pearson correlation between “Hour” and “Battery” features’ AE, and the “Time between events” feature AAE. There is a strong correlation between the errors of predicting “Hour” and “Time between events” probably because both are related to the prediction of the time between states. However, there is no strong correlation with the “Battery” prediction, i.e. errors in “Battery” prediction are independent from errors in the time interval prediction.

Finally, a True Positive (TP) and True Negative (TN) analysis for the prediction of the binary fea-

Figure 4. MSE for validation dataset

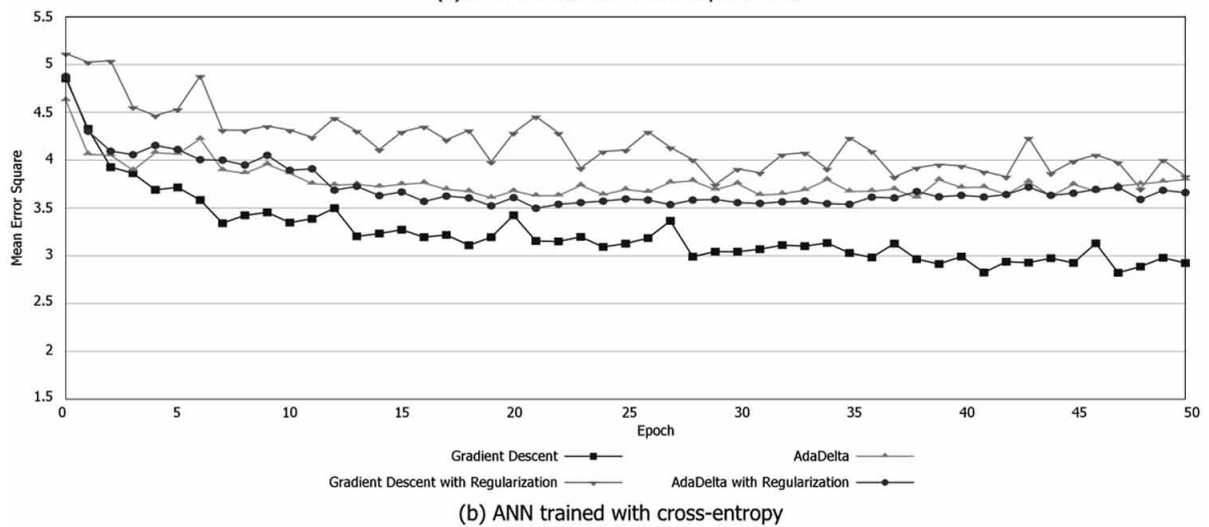
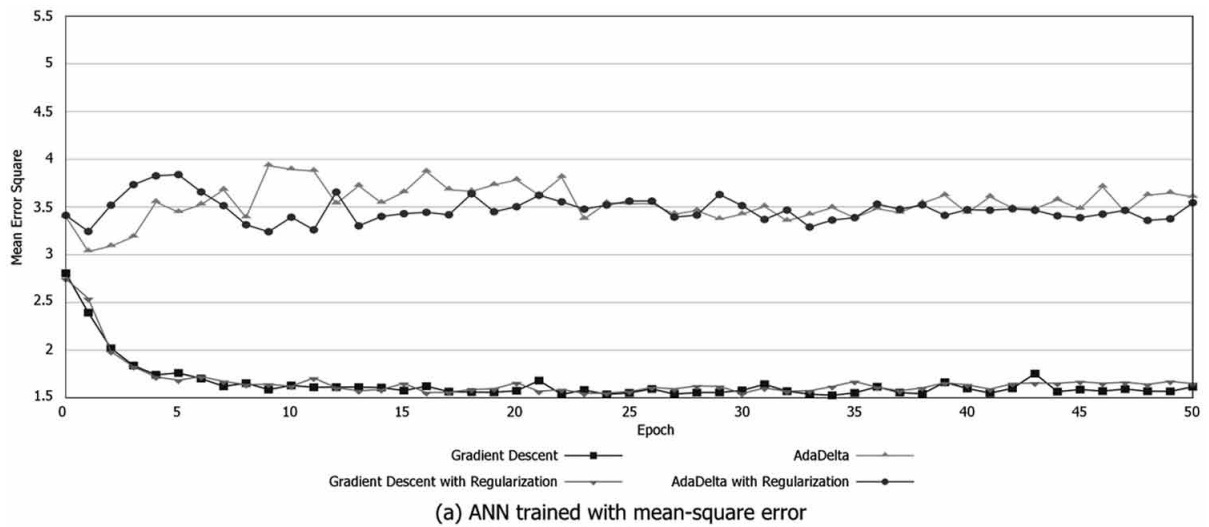


Figure 5. Mean Absolute Error up to 10-state-in-the-future prediction

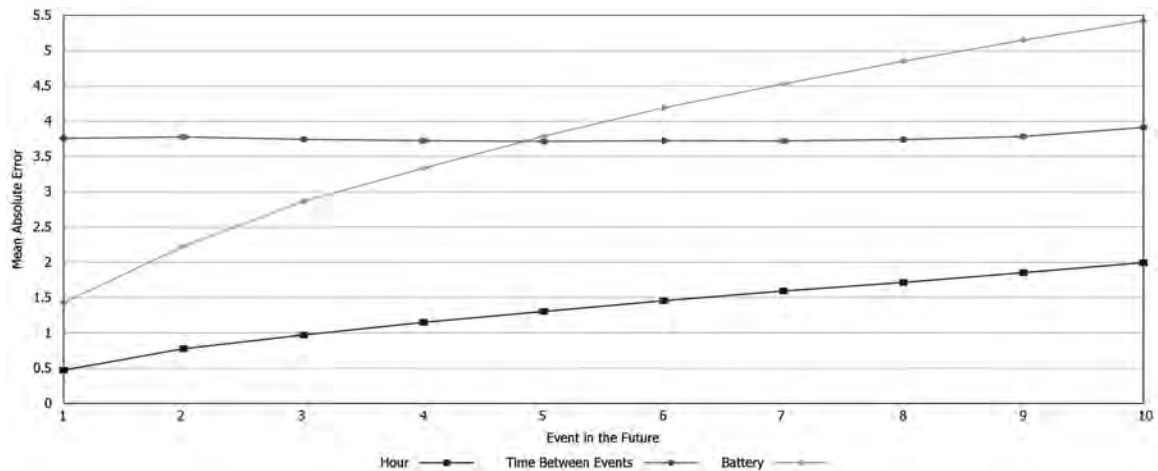


Table 6. Pearson correlations with “Time between events”

Feature	Correlation	p-value
Hour	0.400	>0.001
Battery	0.005	0.719

tures (“External Energy Supply”, “Screen on/off” and “WiFi”) was performed, as Figure 6 reports. The best results were obtained when predicting “External Energy Supply”. Although the TP and TN decrease as more states into the future are predicted, they are never below 80% and 95%, respectively. The ANN also achieved good performance when predicting the “WiFi” feature with a TP rate of 70% and a TN rate of 80%. However, predicting the “Screen on/off” feature resulted in low TP and TN rates. Although the performance when predicting this feature was relatively good when considering a 1-state-in-the-future prediction, it rapidly decreased as more predictions into the future were performed. For example, predicting that the screen would be on in the second future state provided no real information about whether the screen would be on or off, as the TP rate was about 50%. Furthermore, after three states into the future, predicting that the screen would be on

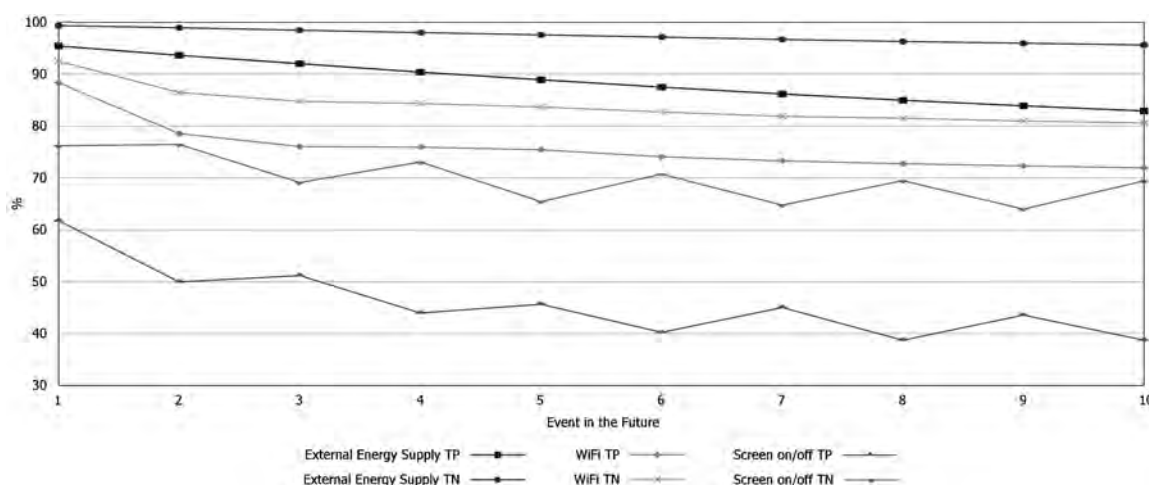
meant that the screen would be most likely off. This was probably a result of a lack of information about user incentive for activating the screen, such as application notifications or user location.

FUTURE RESEARCH DIRECTIONS

Since this paper defines mobile device states using only a small subset of all the possible features, future studies will consider additional features, such as location, Bluetooth connection, incoming/outgoing calls, cellular connection, or 3G signal strength. The main limitation of the current approach is the high number of parameters. Future research should evaluate other ANN configurations with fewer parameters.

Future works will further validate the proposed approach using a larger dataset, such as the Cambridge Device Analyser dataset (Wagner et al., 2014). Additionally, the possibility of using just one ANN for modelling a number of users could be studied. This could be useful in domains in which group behaviour is more important than individual behaviour. Another line of work is comparing ANN based state prediction with other machine learnings techniques. Finally, more complex

Figure 6. True-Positive/True-negative analysis



ANN models should be evaluated in the future. Such models would not only include other type of ANN layers, but also expand the definition of the mobile device state by adding other features, such as location, CPU usage, and day of the week.

CONCLUSION

This study aimed at evaluating the suitability of ANN for predicting mobile device future states. The obtained evidence suggests that RNN have potential for predicting future states of mobile devices. However, the results were mixed, as the ANN was able to effectively predict with low error rates the state of features, such as “Battery” or “External Energy Supply”, while it was ineffective for predicting other features, such as “Screen on/off” or “Time between events”.

REFERENCES

- Alsharif, O., Ouyang, T., Beaufays, F., Zhai, S., Breuel, T., & Schalkwyk, J. (2015). Long short-term memory neural network for keyboard gesture decoding. *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*, 2076–2080. doi:10.1109/ICASSP.2015.7178336
- Do, T. M. T., & Gatica-Perez, D. (2014). Where and what: Using smartphones to predict next locations and applications in daily life. *Pervasive and Mobile Computing*, 12, 79–91. doi:10.1016/j.pmcj.2013.03.006
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780. doi:10.1162/neco.1997.9.8.1735 PMID:9377276
- Liao, Z.-X., Lei, P.-R., Shen, T.-J., Li, S.-C., & Peng, W.-C. (2012). Mining temporal profiles of mobile applications for usage prediction. *Data Mining Workshops (ICDMW), 2012 IEEE 12th International Conference on*, 890–893. doi:10.1109/ICDMW.2012.11
- Musolesi, M., Piraccini, M., Fodor, K., Corradi, A., & Campbell, A. T. (2010). *Pervasive Computing: 8th International Conference, Pervasive 2010, Helsinki, Finland, May 17-20, 2010. Proceedings*. Springer Berlin Heidelberg.
- Niroshinie, F., Seng, W. L., & Wenny, R. (2013). Mobile cloud computing: A survey. *Future Generation Computer Systems*, 29(1), 84–106. doi:10.1016/j.future.2012.05.023
- Pejovic, V. and Musolesi, M. (2015). Anticipatory mobile computing: A survey of the state of the art and research challenges. *ACM Comput. Surv.*, 47(3), 47:1–47:29.
- Ravi, N., Scott, J., Han, L., & Iftode, L. (2008). Context-aware battery management for mobile phones. In *Pervasive Computing and Communications, 2008. PerCom 2008. Sixth Annual IEEE International Conference on*, (pp. 224–233). IEEE. doi:10.1109/PERCOM.2008.108
- Rios, C., Rodriguez, J. M., Godoy, D., Schiaffino, S., & Zunino, A. (2014). Usage pattern mining for smartphone use personalization. In *Proceedings of the 13th Brazilian Symposium on Human Factors in Computing Systems*, (pp. 377–380). Porto Alegre, Brazil: Sociedade Brasileira de Computação.
- Shin, C., Hong, J.-H., & Dey, A. K. (2012). Understanding and prediction of mobile application usage for smart phones. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, (pp. 173–182). New York, NY: ACM. doi:10.1145/2370216.2370243

Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. *Proc. NIPS*.

Wagner, D. T., Rice, A., & Beresford, A. R. (2014). Device analyzer: Large scale mobile data collection. *SIGMETRICS Perform. Eval. Rev.*, 41(4), 53–56. doi:10.1145/2627534.2627553

Wen, Y., Wolski, R., & Krintz, C. (2003). Online prediction of battery lifetime for embedded and mobile devices. *Lecture Notes in Computer Science*, 3164, 57–72.

Williams, R. J., & Zipser, D. (1989). A learning algorithm for continually running fully recurrent neural networks. *Neural Computation*, 1(2), 270–280. doi:10.1162/neco.1989.1.2.270

Zeiler, M. D. (2012). *ADADELTA: An adaptive learning rate method*. CoRR, abs/1212.5701

ADDITIONAL READING

Burbey, I., & Martin, T. L. (2012). A survey on predicting personal mobility. *International Journal of Pervasive Computing and Communications*, 8(1), 5–22. doi:10.1108/17427371211221063

Campbell, A., & Choudhury, T. (2012). From smart to cognitive phones. *IEEE Pervasive Computing/IEEE Computer Society [and] IEEE Communications Society*, 11(3), 7–11. doi:10.1109/MPRV.2012.41

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. doi:10.1038/nature14539 PMID:26017442

Rosa, F. D., Malizia, A., & Mecella, M. (2005). Disconnection prediction in mobile ad hoc networks for supporting cooperative work. *IEEE Pervasive Computing / IEEE Computer Society [and] IEEE Communications Society*, 4(3), 62–70. doi:10.1109/MPRV.2005.55

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, 61, 85–117. doi:10.1016/j.neunet.2014.09.003 PMID:25462637

Zambonelli, F. (2015). Engineering self-organizing urban superorganisms. *Engineering Applications of Artificial Intelligence*, 41, 325–332. doi:10.1016/j.engappai.2014.10.004

KEY TERMS AND DEFINITIONS

Artificial Neural Network: A kind of machine learning algorithms loosely based on how biological neural networks work.

Epoch: A single training pass through the entire training set.

Error Function: A function used for assessing how well a machine learning method performs.

Gradient Descent: Technique for fitting parameters of a function.

Mobile Device: A computational device that people carries, such as smartphones or tables.

Mobile Device State: A set of variables that describe the mobile device's current conditions.

Recurrent Neural Network: ANN that uses previous states for making new predictions.

Category O

Optical Engineering

Visible Light Communication Numerous Applications

Ala' Fathi Khalifeh

German Jordan University, Jordan

Hasan Farahneh

Ryerson University, Canada

Christopher Mekhiel

Ryerson University, Canada

Xavier Fernando

Ryerson University, Canada

INTRODUCTION

The initial theory of Visible Light Communication (VLC) was founded in the 1880s when Alexander Graham Bell invented the photo-phone which was used to transmit a voice signal using the modulated sunlight. Since the time of Graham Bell, optical communication research has attracted the interest of scholars around the world and has evolved into a new IEEE standard namely the P802.15.7 - Standard for Short-Range Wireless Optical Communication (standard, 802.15.7 (2015)). In 2003 at the Nakagawa Laboratory in Keio University, Japan, they have proposed using the Light Emitting Diodes (LEDs) for data transmission.

A major factor that contributes to the evolution of VLC technology is the existing infrastructure. Hence, previously installed facilities, such as LED traffic lights or LED sign boards are readily used. Since the transmitters for VLC are light sources, they function for lighting purposes and illuminate the surrounding environment, hence the radiation power and signal-to-noise ratio (SNR) is high; paving the way for a stable communication link (T. Yamazato, I. H. (2014)).

With respect to the emergence of green communication, VLC is highly energy efficient as it utilizes LEDs. The United States Department of

Energy further corroborated the importance of LED technology, as shown in Table 1. There is superiority in terms of power consumption and operating lifetime in LED technology as compared to traditional lighting technology, such as incandescent and fluorescent lighting. This clearly shows the potential of the LED lighting technology to replace all the conventional illumination tools as well as serve as a reliable transmitter for a VLC link (Chung, Y.-Y. T.-Y. (2014)).

Radio Frequency (RF) wireless connectivity has been used for several decades as it allows for indoor and short distance links to be established without any physical connection. However, these solutions remain relatively expensive and have low to medium data rates. RF wireless links require that spectrum licensing fees are paid to federal regulatory bodies and are required to be contained within strict spectral zones. These frequency allocations are determined by local authorities and may vary from country to country, making a standard interface difficult. Since the visible light spectrum is not in the licensed band (400 to 790 THz), licensing fees can be avoided which effectively reduces system cost. In addition, the broadcast nature of the RF link is beneficial for mobile connectivity but this may result in interference between devices located within close prox-

Table 1. Performance of the conventional and LED lighting technology

Lamp Type	Watts	Lumens	Operating Lifetime
Incandescent	60	900	1000
Compact florescent lamp	15	900	8500
LED (2011)	12.5	800	25000
LED – future(2015)	5.8	800	40000

imity. Due to the RF wavelength, it is difficult to contain within boundaries and can impede system performance (Hranilovic, S. (2005)).

Optical radiation in the infrared or visible range is easily contained by opaque boundaries. As a result, interference between adjacent devices can be minimized easily and economically. Additionally, inexpensive LEDs and photodiodes are able to interchangeably work between baseband and transmission frequencies where as high-frequency RF circuit design techniques are required in the RF domain. Free-space optical (FSO) links with an inherent low probability of intercept and anti-jamming characteristics is among the most secure of all wide-area connectivity solutions (Hranilovic, S. (2005)).

Unlike many RF systems that radiate signals in all directions, thus making the signal available to all within the receiving range, FSO transceivers use a highly directional and cone-shaped beam with a dominant line-of-sight (LOS) propagation path. Therefore, interception is extraordinarily difficult and anyone tapping into the systems can easily be detected as the intercept equipment must be placed within the very narrow optical foot print (Ghassemlooy, Z., P. W. (2012)). Although this contributes to the security of wireless optical links and reduces interference it also greatly impacts the high mobility of such devices.

The aim of this book chapter is to introduce the concept of VLC as an emerging technology from a system and hardware design point of view, and shed the light on the rich features and potential of this technology that make it a viable substitution for other wireless technologies. The importance of

this technology will be demonstrated by covering various applications and scenarios.

BACKGROUND

VLC is the process of transmitting digital data by using the visible light spectrum. This can be achieved by modulating the data using a light source such as LEDs that can be switched fast enough to avoid observable flickering or light dimming, another possible way is to change the light intensity in a way that is not observable to the human eye but can be detected using an appropriate sensor such as a photo diode (PD). Along with its prime function of lighting, LEDs can also serve to transmit data as long as there are suitable receivers.

Additionally, image sensor pixels can be used as an effective VLC receiver. The ability to spatially separate multiple sources of image sensors provide an attractive feature to VLC. Image sensor pixels can sense LED transmission sources and discards other pixels which detects ambient noise. More specifically, outdoor usage of VLC is possible by discarding pixels associated with noise sources such as the sunlight or streetlights. Hence, image sensor based VLC is an attractive solution for outdoor mobile applications (T. Yamazato, I. H. (2014)). With regards to short-range VLC applications, the SNR of a direct detection receiver is proportional to the square of the received optical power. Therefore, VLC links can tolerate only a comparatively limited amount of signal path loss.

Figure 1 shows a general block diagram for a simple VLC transceiver. As shown in the figure,

the upper part represents the transmitter, where binary data is first passed to a coder such as a Reed-Solomon coder (Khalifeh, A. Y. (2010)) where error correction codes are added. After that, binary data is modulated using a digital modulation scheme such as binary On-Off Keying (OOK) where data is modulated by passing it to an amplification circuit that drives an LED(s), which converts the data into light over the wireless communication channel.

One can notice from Figure 2 that either a single or several LEDs can be used to transmit data. The same applies to the receiving circuit, thus four different configurations can be used to increase transmission rates, as well as improve the received SNR; (a) Single-Input Single-Output (SISO), where only one LED transmitter and one photodiode receiver is used, (b) Single-Input Multiple-Output (SIMO), where one LED transmitter and multiple photodiode receivers are used, (c) Multiple-Input Single-Output (MISO) where multiple LED transmitters and one photodiode receiver is used, and finally, (d) Multiple-Input Multiple-Output (MIMO) where multiple LED transmitter and multiple photodiode receivers are used.

The data path is then transmitted to the receiver circuit, where the light pulses are detected by a photodiode(s). The output signal is then passed to an amplifier, pulse shaping and re-generation

blocks, a channel decoding module to correct the introduced errors, and a demodulator circuit that will detect and convert the signal back into binary data. Without any ambient noise or distortion, the transmitted bit stream should reach the destination without any errors. However, due to the internal and external noise sources, some binary data will be falsely detected which causes bit errors. In addition to the typical additive white Gaussian noise that is locally generated by the receiver electronic circuit, VLC receivers are prone to the ambient surrounding and background lights that can cause errors in the detection process.

However, in order to mitigate the effect of these external sources, various techniques are used such as the one proposed in (Ya'nez, V. T. (2009)) where the channel characteristic is estimated and the received signal is equalized based on the channel estimation model. Another technique presented in (Ya'nez, V. T. (2009)) uses the wavelength filtering to reduce the unwanted inference signal.

VLC APPLICATIONS

This section discusses in details the major applications where VLC has a great potential. Applications are divided into two main categories; indoor and outdoor. VLC can be used in various indoor

Figure 1. VLC transceiver block diagram

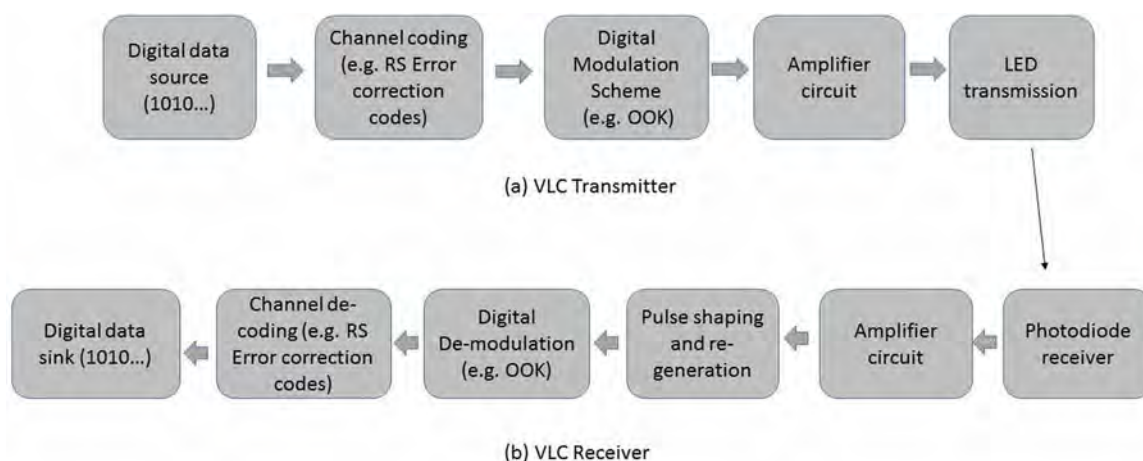
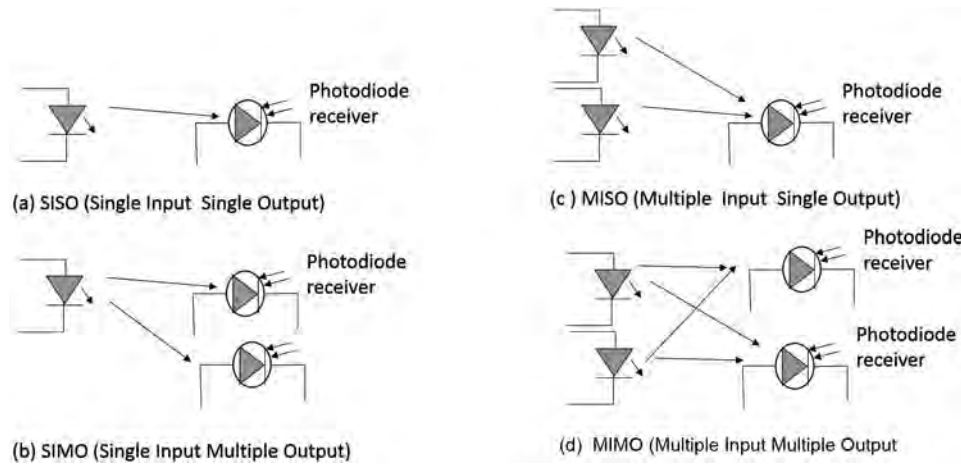


Figure 2. Several transmitter-receiver configurations



applications and scenarios, namely, in high-density indoor areas such as classrooms, conference halls, convention centers and other assembly spaces where light can be used to broadcast relevant information to the audience. Entertainment applications can also utilize VLC for video and audio streaming and broadcasting. Indoor navigation and localization is another important application of VLC, and finally, VLC can play an important role in data transmission in sensitive areas where radio wireless communication is not preferable, some examples are in hospitals and inside the airplanes. A closer look for these potential applications will be described next.

VLC can also be used in various environments where it is not safe or recommended to use the radio spectrum, since it may interfere with other critical electronic equipment. The other set of applications that can efficiently utilize this technology is some of the outdoor-based applications. Research has been done on the field of vehicle-to-vehicle (V2V) communications where cars can communicate with each other to exchange critical and general messages. Additionally, it can be easily used for Intelligent Transportation Systems (ITS) and vehicle-to-infrastructure and infrastructure-to-vehicle (V2I and I2V) communication.

Indoor Applications

VLC has many potential indoor applications. This is due to the fact that an indoor channel changes slowly when compared to outdoor channels, where the channel may vary more rapidly. Three main applications will be discussed: Indoor localization and positioning, medical services, and VLC in-flight communication systems.

Indoor Localization and Positioning Services

One important application of VLC is indoor localization and positioning. This is a vibrant research area that is gaining much attention recently. Positioning applications cover a wide area where the technology can be included into various consumer electronics. Indoor positioning technology can be used to guide users in large areas. Moreover, positioning systems using VLC can also detect products inside large warehouses, automating some record management processes.

As known in literature, radio signals coming from Global Positioning System (GPS) satellites cannot penetrate well through walls of large buildings, consequently, fast and accurate indoor posi-

tioning is difficult to achieve by GPS. To overcome this problem, there are two possible alternatives; RF and VLC-based techniques. Examples of RF based techniques are: wireless local area network, radio-frequency identification (RFID), cellular, ultra-wide band and Bluetooth. These technologies will have positioning accuracies from tens of centimeters to several meters. But unfortunately, this amount of accuracy is not sufficient for many indoor applications. Hence, the techniques based on VLC are gaining greater attention.

LEDs are currently being installed in many large buildings such as museums and shopping malls, because they have the advantage of much longer life and lower operating cost. For these reasons, indoor positioning techniques based on VLC and LEDs are the desirable and preferable options (LVX Minnesota lightning. (2015)).

Medical Services

Another major concern these days are regarding the compatibility of medical devices in healthcare centers with the incorporation of wireless technology (S.S. Muhlen, D. D. (2008)). RF wireless technology has always been associated with the emission of electromagnetic interference (EMI) (H. Hong, Y. R. (2008)). Intrusion of the EMI jeopardize the quality of medical monitoring, as the accuracy and efficiency of data transmission is crucial for the medical staff to provide corresponding measures or treatment, based on the real-time information received (K.S. Tan, I. H. (2001), Chung, Y.-Y. T.-Y. (2014)).

VLC can be applied to indoor medical applications by transmitting patient data as well as management data. Healthcare information such as electrocardiography (ECG), photoplethysmogram (PPG) signals and text information can be transmitted simultaneously, using a single channel VLC. This allows for a more precise and accurate monitoring and diagnosis.

VLC In-Flight Communication Systems

The use of VLC technologies for offering data networking for passengers during flight is another promising indoor application. Internet access, Wi-Fi is traditionally used to connect the passenger devices to an access point that will send their request to the internet either via a satellite connection or via plane-to-earth data link. This Wi-Fi scheme shares the radio spectrum among many passengers, which may cause congestion and reduce the speed of communication. In addition, Wi-Fi may affect the airplane navigation systems especially at take-off and landing causing EMI.

A better solution can be implemented by utilizing a hybrid radio and optical system, where passengers may use the radio spectrum for uploading their requests, which usually require less bandwidth than the downlink. The downlink can be provided to the passengers via the personal LED lights available on top of their seats. This scheme will increase the internet speed due to abundant bandwidth available in the optical domain and will reduce the interference to the airplane navigation system. Brighter overhead lighting will result in higher SNR and greater connectivity.

Outdoor Applications

In this section, potential outdoor applications for VLC will be discussed. Despite the more challenging problems such as fast time varying channels and users' high mobility, VLC, if designed properly, can be used in several outdoor applications such as automotive communication, intelligent transportation systems, and underwater communications.

Visible Light Communication in Automotive Industry

Due to VLC being optimal for short range communication, it has significant potential in the

automotive industry such as car-to-car (C2C) and Intelligent transportation systems (ITS). Additionally, existing infrastructure is in place to take advantage of VLC such as headlights and streetlights. In the next few years, the implementation of LED taillights, brake lights, will all be more than 50% (Cui, K. C. (2012)). All these conditions make it desirable to consider VLC in the automotive industry. Vehicular Visible Light Communications (V2LC) offers a dependable solution to put into operation vehicle-to-vehicle (V2V) communications and requires minimal installation cost.

V2LC is challenged by dynamic ambient noise, high mobility, and moderately low transmitter heights on communicating systems such as vehicles and roadside units. These limitations make the V2LC channel modelling a challenging task. V2LC utilizes either a photodiode detector or a camera installed on the vehicle to be used as a receiving element. It was found that V2LC systems perform better in high volume vehicular traffic; this is because of the fact that visible light can mainly propagate within proximity and LOS to reduce interference and increase link scalability (CALM. (2011)).

C2C Communication System

A traditional C2C -VLC scenario is shown in Fig. 3, where the car on the left side communicates with the car on the right side using its headlamp. As shown, the projected beam pattern is single, and the received light consist of the LOS and reflected or Non-Line of Sight (NLOS) components. It can

be noticed that a multi-path interference will occur due to the reflected rays from the road surface, which depends on many factors such as the pavement material, the angle of incidence as well as the weather condition (rain, snow, fog, etc.).

There are many regulations issued by Economic Commission of Europe and Federal Motor Vehicle Safety Standards in United State of America related to headlamps to insure that vehicles will provide good road illumination. At the same time, the light emitted from headlamps should not visually disturb other road users. Hence, the lamp, its reflective devices, and other associated equipment must achieve specific requirements. Therefore, the Lambertian model which has been used widely in indoor VLC LED modelling, cannot be used to model the vehicle's light pattern (Kumar, N., et al. (2012)).

In C2C VLC links, usually the LOS and the NLOS components form the received optical power. For FSO, there are additional noise sources, such as the background solar radiation, streetlights, vehicles, and secondary reflections which can be treated as artificial light (Ghassemlooy. Z., P. W. (2012)). The background solar radiation is mostly the dominant noise source as it is composed of direct and scattered radiation (Cui, K. C. (2012)). As such, good noise cancellation mechanisms are required to account for the nature of these noise sources.

In addition, appropriate signal processing algorithms are needed to equalize the received signal and improve its condition. Especially, the channel coherence time will be very short in an outdoor vehicular environment due to increased

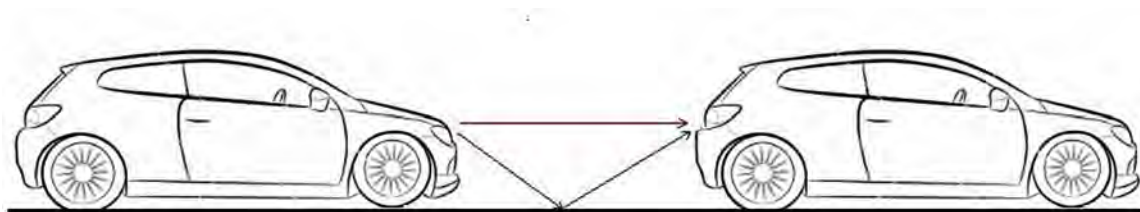


Figure 3. C2C VLC and projected LOS and NLOS beam patterns

mobility of fast moving vehicles. Hence, any channel estimation and equalization has to be achieved seamlessly.

ITS Traffic Light to Car Communication

Improvements to safety, fuel consumption, and transportation time are the key objectives of an intelligent transportation system. ITS refers to the potential of adding information and smart communication technologies to the transportation infrastructure and vehicles in order to enhance transportation safety, mobility, and support productivity through the use of advanced communications technologies. VLC can play an important role in ITS, such as broadcasting important traffic information. Here several configurations, such as V2V using LED-based rear panel lights, and bidirectional infrastructure-to-vehicle are possible. Traffic information system using LED-based traffic lights were investigated by many authors. They touched upon topics such as analyzing the communication performance and defining a service area, in which communications using a particular data rate together with intensity modulation OOK can be maintained.

VLC is suitable for both a broadcast system in I2V communication systems and effective in V2V as well. In a V2V scenario, a vehicle in front of a traffic light can receive the information sent by the traffic light, and then it can relay this information using the brake lights to the vehicle running behind. This can be extended to establish a vehicle ad-hoc network. With respect to I2V broadcasting, there is no need for extra power to broadcast traffic safety related information, which can be continuously broadcasted to support smooth traffic flow as well as reducing accidents and fatalities (CALM. (2011)).

LED-based traffic lights are also considered to be a suitable choice for traffic light Road Side Units (RSUs) in the ITS architecture without the need for additional infrastructure. One of the most

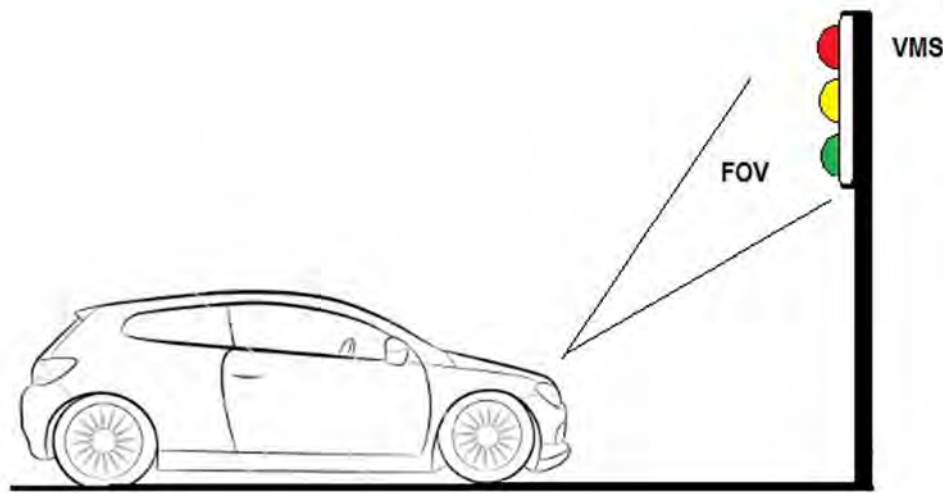
possible approaches of communication between an LED-based RSU and a vehicle is presented in Fig. 4. The first stage is collecting and passing the detected information to a central control unit, which consists of many parts such as data processing, geographical distribution of the data, and information broadcasting. After the data collection stage, traffic control actions are made. Traditional ITS systems are normally based on wireless radio solutions and short range infrared systems.

In order to deliver information from the traffic control infrastructure, which is represented by traffic lights to vehicles, VLC can be effectively integrated into this scenario. This technique offers many advantages, like cost effectiveness, alleviation of additional wireless communication support, and energy efficiency. The most important benefit is that it does not require additional power to support future communication systems. Fig. 4 shows that the ITS use control technology and access networks to achieve communication facilities. In this case, it becomes important for an ITS architecture design to be flexible enough to accommodate integration of new systems (Kumar, N. L. (2012)).

Underwater Communication

Providing an efficient and reliable data transmission for underwater communication is a challenging task, this is due to the nature of water, which absorbs the energy of the transmitted signal and refracts it. Data transmission underwater suffers from high-energy consumption, long propagation delay, and limited link bandwidth. RF-based solutions are not efficient due to the high attenuation and due to the multipath signal propagation (Gabriel, C. K. (2011)). As an alternative, acoustic waves are in use for underwater communications. Table 2 compares acoustic and optical communication methods for underwater use. As shown in the table, both acoustic and optical mediums suffer from power loss. The acoustic waves suffer

Figure 4. VLC integration with ITS architecture depicting the vehicle Field-of-View (FOV) and the Variable Message System (VMS)



from more than 0.1 dB per meter per Hz, so the higher the frequency and distance, the higher the attenuation and power loss.

For optical transmission, the amount of power loss depends on the turbidity, which is a measure of the water purities and its clarity. The higher the turbidity, the higher the power loss. In addition, the signal strength is inversely proportional with the distance (Lanbo, L. S.-H. (2008)). Although acoustic transmission has a greater range, its transmission rate is very limited. While the optical communication link has higher data rates for low distance, it becomes detrimental over a long distance. The propagation delay for optical communication is much lower when compared

to an acoustic link, since light speed inside water is significantly faster than the speed of acoustic transmission. Moreover, optical transceivers are relatively cheap and less complicated when compared with acoustic transceivers. It is important to note that the performance metrics in Table 2 are based on a seawater environment, so it may vary for other underwater environments.

DISCUSSION

Ever since the growing demand and use of data, there has been a need to reliably transmit significant amounts of data in real time. Current

Table 2. Comparison between underwater communication using acoustic and optical communication methods, assuming a seawater environment

Property	Acoustic	Optical
Require alignment	No	Yes
Receiver complexity	Higher	Lower
Recommended communication range	Up to 5 km	Up to 100m
Data rate	Up to 100 Kbps	Up to 1 Gbps
Power loss	> 0.1 dB/m/Hz	proportional to water turbidity
Propagation delay /speed	large/slow (1500-2000 m/s)	low/ high (2.255×10^8)
Frequency band	~ kHz	10^{15} Hz

Sources: Lanbo, L. S.-H. (2008), Yang S., K., (2011), Lurton, X. (2010)

technology, as well as a growing number of users limit the communication capacity. Hence, there is a great need for high bandwidth, energy efficiency, cost, and reliability for a communication link. Hence, VLC offers an excellent alternative to traditional technology.

However, there exists some challenges that were investigated to make this a viable communication technology. These challenges are dealing with the LOS requirement of VLC, inability to communicate through opaque obstacles, high mobility, ambient light interference, limited LED photon emission rate, required DC biasing, and limited modulation bandwidth as well as the non-linearity nature of LEDs (Dong-Fang Zhang, Y.-J. Z.-Y. (2013)).

As previously discussed, there are several approaches which can be taken to alleviate the main challenges with VLC. To name a few, implementing a MIMO or LED array can be used to increase the SNR and improve diversity in impeding scenarios, channel modelling and equalization at the receiver, camera image sensors used to eliminate ambient light-treated as noise, and finally advanced modulation techniques to overcome limited modulation bandwidth.

FUTURE RESEARCH DIRECTIONS

Since VLC has several challenges as previously mentioned, there has been a fair amount of lab research toward improving its performance as well as improving its distance range. In order to mitigate the channel effect on a VLC system, several techniques have been proposed, such as the usage of Orthogonal Frequency Division Multiplexing (OFDM), and the use of MIMO (Mohammed S. A. Mossaad, S. H. (2015)).

With respect to the limited LED modulation bandwidth, polarization multiplexing methods can be combined with other multiplexing methods which can further increase the communication rate. Cvijetic et al. studied Bit Error Rate (BER)

performance of a polarization multiplexed optical wireless transmission (Chiharu Mukai, K. O. (2012)). Additionally, advanced modulation techniques have been carried out to achieve high data rates up to 100Tbit/s by implementing orbital angular momentum (OAM) with other multiplexing domains and present a free-space data link that uniquely combines OAM-, polarization-, and wavelength-division multiplexing by using three-dimensional multiplexing. (Hao Huang, X. (2014)).

In addition, researchers have looked at combining the Power Line Communication (PLC) technology with VLC, such that the power lines become the data source of these lights (Ding W., Y. F. (2015), Hao M., L. L. (2013)).

CONCLUSION

In this chapter, we have discussed several indoor and outdoor potential applications that VLC can utilize in the upcoming years. Despite its performance challenges, VLC showed significant potential in several indoor and outdoor applications, such as: indoor localization and positions, in-mall advertising and user behavior tracking, and in-plane multimedia broadcasting. For outdoor applications, it shows a promising potential in the automotive industry as well as underwater communications.

REFERENCES

CALM. (2015, Nov 1). *Communications in Co-operative Intelligent Transport Systems - CALM for C-ITS*. Retrieved from Introduction to Communications Access for Land Mobiles: <http://calm.itsstandards.eu/>

Chiharu Mukai, K. O. (2012). *BER Performance of Multiplexed CSK using Linear/Circular Polarization in Optical Wireless Communications*. ISITA.

- Chung, Y.-Y. T.-Y. (2014). Mobile health-monitoring system through visible light communication. *Bio-Medical Materials and Engineering*, 608–737.
- Cui, K. C., Chen, G., Xu, Z., & Roberts, R. D. (2012). Traffic light to vehicle visible light communication channel characterization. *Applied Optics*, 51(27), 6594–6605. doi:10.1364/AO.51.006594 PMID:23033030
- Ding, W. Y. F., Yang, F., Yang, H., Wang, J., Wang, X., Zhang, X., & Song, J. (2015). A hybrid power line and visible light communication system for indoor hospital applications. *Computers in Industry*, 68, 170–178. doi:10.1016/j.compind.2015.01.006
- Dong-Fang Zhang, Y.-J. Z.-Y. (2013). Multi-LED Phase-Shifted OOK Modulation Based. *IEEE Photonics Technology Letters*. Retrieved from <https://www.eldoled.com>
- Gabriel, C. K. (2011). *Channel modeling for underwater optical communication. Proceeding in IEEE GLOBECOM Workshops*. GC Wkshps.
- Ghassemlooy, Z. P. W. (2012). *Optical wireless communications: system and channel modeling with Matlab*. CRC Press.
- Grantham, C. (1998). Visible Light Communication for Audio Systems. *IEEE Transactions on Consumer Electronics*, 1–11.
- Hao, M. L. L. (2013). Integration of indoor visible light and power line communication systems. *17th IEEE International Symposium in Power Line Communications and Its Applications (ISPLC)*. doi:10.1109/ISPLC.2013.6525866
- Hao Huang, X. (2014). 100 Tbit/s free-space data link enabled by three-dimensional multiplexing of orbital angular momentum, polarization, and wavelength. *Optics Letters*, 39(2), 197–200. doi:10.1364/OL.39.000197 PMID:24562105
- Hong, H., Y. R. (2008). Information illuminating system for healthcare institution. *Proceedings of International Conference on Bioinformatics and Biomedical Engineering*, 801–804.
- Hranilovic, S. (2005). *Wireless Optical Communication Systems*. Springer.
- Khalifeh, A. Y., & Yousefizadeh, H. (2010). Optimal Audio Transmission Over Error-Prone Wireless Links. *IEEE Transactions on Multimedia*, 12(3), 204–214. doi:10.1109/TMM.2010.2041096
- Kumar, N. L. (2012). Visible Light Communications in Intelligent Transportation Systems. *Intelligent Vehicles Symposium Alcalá de Henares*, 748–753. doi:10.1109/IVS.2012.6232282
- Lanbo, L. S.-H. (2008). *Prospects and problems of wireless communication for underwater sensor networks*. Wirel. Commun. Mob. Comput.
- Lurton, X. (2010). *An Introduction to Underwater Acoustics Principles and Applications*. Springer-Verlag Berlin Heidelberg. doi:10.1007/978-3-642-13835-5
- LVX Minnesota Lightning. (2015, Nov 1). Retrieved from <http://purelifi.com>
- Masayuki Kinoshita, T. Y. (2014). *Motion Modeling of Mobile Transmitter for Image Sensor Based I2V-VLC, V2I-VLC, and V2V-VLC*. Globecom Workshop-Optical Wireless Communications.
- Mohammed, S. A., & Mossaad, S. H. (2015). Visible Light Communications Using OFDM and Multiple LEDs. *IEEE Transactions on Communications*, 4304–4313.
- Muhlen, S. S. D. D. (2008). New challenges in controlling EMI in hospitals. *Proceedings of International Federation for Medical & Biological Engineering*, 834–837.
- Pengfei, L., e. a. (2014). Fundamental Analysis of a Car to Car Visible Light Communication System. *9th International Symposium on Communication Systems, Networks & Digital Sign (CSNDSP)*.

Perez-Jimenez, R. R.-H. (2011). Visible Light Communication Systems for Passenger In-Flight Data Networking. *IEEE International Conference on Consumer Electronics (ICCE)*, 445-446. doi:10.1109/ICCE.2011.5722675

Popoola, W. O., & Ghassemlooy, Z. (2009). BPSK subcarrier intensity modulated free-space. *Journal of Lightwave Technology*, 27(8), 967-973. doi:10.1109/JLT.2008.2004950

Ramirez-Iniguez, R. S. M. (2008). Optical Wireless Communications: IR for Wireless Connectivity. Boca Raton, FL: CRC Press.

Standard 802.15.7. (2015, Nov 1). Retrieved from <https://standards.ieee.org/develop/project/802.15.7.html>

Takaya Yamazato, I. T. (2014). Image-Sensor-Based Visible Light Communication for Automotive Applications. *IEEE Communications Magazine*, 10.

Tan, K. S. (2001). Electromagnetic interference in medical devices: Health Canada's past current perspectives and activities. *IEEE International Symposium on Electromagnetic Compatibility*, 1283-1284.

U.S. Department of Energy. (2013). *Part 1: Review of Life-Cycle Energy Consumption of Incandescent, Compact Fluorescent, and LED Lamps*. Author.

Yamazato, T., Takai, I., Okada, H., Fujii, T., Yendo, T., Arai, S., & Kawahito, S. et al. (2014). Image sensor based visible light communication for automotive. *IEEE Communications Magazine*, 52(7), 88-97. doi:10.1109/MCOM.2014.6852088

Yañez, V. T. (2009). Illumination interference reduction system for VLC Communications. *Proc. WSEAS Int. Conf. Math. Methods, Computer. Tech. Intel.*

Yang, S. K. (2011). Reduction of optical interference by wavelength filtering in RGB-LED based indoor VLC system. *Proceeding of Opto-Electronics and Communications Conference (OECC)*.

Zahmati & Fernando. (2011). Emerging Wireless Applications in Aerospace: Benefits and Challenges. *Exponent, the Honeywell Technical Journal*.

ADDITIONAL READING

Gallico, P. V. (2007). *Optics research trends*. New York: Nova Science.

Haas, H. (2015). Visible Light Communication. *Optical Fiber Communication Conference*.

Haas, H., & Chen, C. (2015). What is LiFi? *2015 European Conference on Optical Communication (ECOC)*. doi:10.1109/ECOC.2015.7341879

Haruyama, S. (2012). Advances in Visible Light Communication Technologies. *European Conference and Exhibition on Optical Communication*. doi:10.1364/ECEOC.2012.We.3.B.5

Koteles, E. S., & Willner, A. E. (1995). *Emerging components and technologies for all-optical networks: Conference: Papers*. SPIE.

Whyte, B. (1999). *Networked futures: Trends for communication systems development*. Chichester, England: J. Wiley. doi:10.1002/0470841850

KEY TERMS AND DEFINITIONS

CDMA: A channel access method, which allows several users to share a band of frequencies.

Intelligent Transportation Systems (ITS): The potential of adding information and smart communications technologies to the transportation infrastructure and vehicles, in order to enhance transportation safety, mobility, and support productivity.

Light-Emitting Diode: A semiconductor device that emits visible light at a single wavelength when an electric current passes through it. The output from an LED can range from red (at

a wavelength of approximately 700 nanometers) to blue-violet (about 400 nanometers).

OFDM: A method of encoding digital data on multiple carrier frequencies.

Signal to Noise Ratio: A ratio of desired signal to undesired signal (noise) in the average power level of a transmission.

Smart Grid: An electrical grid which includes a variety of operational and energy measures in-

cluding smart meters, smart appliances, renewable energy resources, and energy efficiency resources.

Visible Light Communication (VLC): Data communications medium which uses visible light between 400 and 800 THz (780–375 nm).

Wireless Communication: Transfer of information between two or more points that are not connected by an electrical conductor.



Category P

Public Sector Management

Community Outreach



Loriene Roy

The University of Texas at Austin, USA

Antonia Frydman

The University of Texas at Austin, USA

INTRODUCTION

Community outreach is the extension of library services outside of the physical library. In public libraries, this calls on the library to connect with actual and potential library users. Once primarily associated with serving disadvantaged potential library users, community outreach now refers to any activities libraries develop that deliver services to new constituencies. Community outreach covers the following process: identifying targeted communities, assessing a need, considering responses to this need, prioritizing options, targeting a response to the need, developing the initiative, marketing the response, delivering the program, and conducting continuous and end-of-service evaluation. Of those actions, the process of marketing has received the most attention over recent years. The concept of marketing or ‘selling’ the library and its services has been interpreted as a form of advocacy; not only do libraries inform the public about an individual service, but individual librarians are called on to advocate for their work setting and for the professional at large. The American Library Association (ALA) now considers “Advocacy for Libraries and the Profession” as one of its eight key action areas (American Library Association, 2013). ALA has developed tools for library workers, grouped in a section of its website called “AdvocacyU.”

BACKGROUND

As Heim (now, McCook) pointed out, outreach is much more than opening the doors of an in-

DOI: 10.4018/978-1-5225-2255-3.ch579

stitution; it extends across a spectrum of activities and actions in what Heim referred to as the Stimulation Continuum (Heim, 1984). Outreach involves reaching out to the underserved, and even intervening during times of need.

Early outreach services were designed with education in mind. Early public libraries in the mid to late nineteenth century were founded to further the education of an American public that had little access to other forms of schooling. Libraries, including state library agencies, coordinated broad educational efforts under offices of library extension.

These efforts were followed by programs designed to acclimate new citizens. During that time the Immigrant Publication Society issued small pamphlets to assist librarians in responding to the needs of Jews from Eastern Europe, Russian Jews, and Poles (Carr, 1919). These publications were issued with support from the ALA Committee on Work with the Foreign Born. This committee existed for thirty years, from 1918 to 1948. Over time, these library efforts have been scrutinized more closely and considered by some to be expressions of acculturation.

Heim noted that after 1950, outreach services returned to the educational motif, focusing on adults’ learning needs. And since the 1950s, public libraries provided both general educational services and targeted services for specific clientele. ALA reflected this interest by establishing the Office of Library Service to the Disadvantaged (OLSD) in 1970. In the 1980s, targeted services addressed the needs of latchkey children, business people, and the homeless. The OLSD was renamed the Office of Literacy and Outreach Services (OLOS).

OLOS's mission describes the communities often associated with community outreach:

OLOS focuses attention on services that are inclusive of traditionally underserved populations, including new and non-readers, people geographically isolated, people with disabilities, rural and urban poor people, and people generally discriminated against based on ethnicity, sexual orientation, age, language, and social class. (American Library Association, Office of Literacy and Outreach Services, 2013)

OLOS's involvement with outreach currently includes planning and delivering nationally visible professional events. This includes the annual Dr. Martin Luther King Jr. Sunrise Breakfast Celebration held during each ALA Midwinter meeting, the National Bookmobile day which is celebrated during National Library Week, and events scheduled for the ALA Annual Conference, including the Diversity and Outreach Fair and the Jean Coleman Library Outreach Lecture. In addition, the OLOS website maintains 'toolkits' to assist library workers in their outreach missions. These toolkits include practical advice on serving audiences including the homeless, older adults, and tribal libraries.

For Heim, outreach necessitates creating a climate that supports many activities (Heim, 1982). In 2004, Osborne grouped cases of outreach into six classes: services outside library walls, outreach inside the library, outreach using information technology, technical services, advocacy, and staff development (Osborne, 2004). In her 2010 book entitled *Librarians as Community Partners: An Outreach Handbook*, Smallwood organized 66 examples of library outreach into ten broad categories (Smallwood, 2010). Some categories illustrate that outreach can be designed for specific clientele such as seniors, youth, culturally and ethnically diverse patron groups, and individuals residing in correctional care facilities. Approaches may include classroom outreach, book festivals, or special collections. Outreach can also be grouped

with approaches that share specific techniques such as those that employ local media and those whose success results from collaborations with other agencies, institutions, or companies including arts centers, day cares, restaurants, book stores, and historical societies.

Outreach is not limited to the public library setting or within the library's physical facility. Academic libraries increasingly provide services to their wider community of parents, alumni, and potential library supporters. The services that libraries develop might include instruction, exhibits, lectures, events, and other public programs. Across the globe, outreach services have included various conveyance means to deliver resources and circulating materials, as well as services. Bookmobiles travel to remote areas in the United Kingdom and Scandinavia, while donkeys travel with a cart of books to children in Ethiopia, or use side packs to bring books to children in Columbia (Smallwood, 2010). Alternatively, within library facilities, the redesign of space addresses outreach missions, as patrons have access to learning commons, information commons, and/or social commons.

Outreach has additionally become associated with social justice. One example of this is The People's Library, which was founded by members of the Occupy Wall Street movement (Gray, 2012). Those involved in outreach are recognized for their innovative work, and outreach efforts are often tied to support for intellectual freedom. For example, the Robert B. Downs Award, awarded by the Graduate School of Library and Information Science at the University of Illinois at Urbana-Champaign "acknowledges individuals or groups who have furthered the cause of intellectual freedom, particularly as it impacts libraries and information centers and the dissemination of ideas" (The University of Illinois at Urbana-Champaign, 2013).

Controversies and Problems

Extension of library services beyond traditional roles of collection-building and support for the

reader is controversial to some library workers, and surprising to some members of the library's public. Individual ALA members and even members of the ALA Council (the policy setting body of the association) sometimes express disagreement over whether involvement in social justice issues is supportive of the profession's mission. Those in the profession may question whether outreach aligns with the original missions of education, information, and recreation, and whether extending services is in some way disadvantages the traditional library audience of people seeking a good book to read. This tension is more pronounced during economic downturns, when finances may be limited to improving routine services by developing and implementing new physical and virtual services. Those supportive of outreach reply that literacy programming is an active expression of the service continuum. When library workers develop services to support both existing readers and patrons seeking the skills to become readers, they are affirming not only the library itself, but also the library's commitment to outreach (Heim, 1982).

Libraries have long wrestled with measuring their impact. Starting in the 1930s, national standards were released for public libraries (McCook, 2011). Until the 1960s, these standards were revised each decade, and measured a public library's achievement in numbers of recommended books or hours of service. This approach to evaluation changed in the early 1980s with the publication of the first of a series of planning documents for public librarians. These planning documents provided local libraries with assistance in identifying their roles for a limited period of time, usually five years, and instruction on how to measure whether they had served their community well during that time. Achievement was then measured in ratios called output measures, such as turnover rate or circulation divided by number of volumes held. Public libraries continue to use a suite of planning documents to assess their services.

Academic libraries were typically ranked by the number of volumes held. New measures have

been incorporated over time, including those that measure user perceptions of library services such as the LibQUAL+ project (Snyder, 2002). A number of studies have estimated the economic impact of a library. One such study found that for every one dollar invested in public libraries in the state of Texas, a community receives \$4.42 in benefits (The University of Texas at Austin, 2012). These measurements, however, do not specifically assess the impact of community outreach.

Another less-mentioned challenge regarding community outreach is that it requires special interpersonal skills on the part of the library workers. Librarians involved in community outreach efforts should be aware of the skillsets and competencies required to fulfill a specific service, and attitudes required to serve a specific community. (Montiel Overall, 2009) Being a 'well meaning person' is not the sole predictor of success in working with communities of color (Roy, 2011). In developing outreach services, information professionals must examine their potential role as outsider and the advantages and disadvantages associated with their presence (Roy, 2009). Knowledge about a community, and how to work best within that community is a process that takes time. It involves sensitivity to protocols, good etiquette that guides interpersonal communication within a community.

Solutions and Recommendations

What was once considered outreach is now linked to the concept of community engagement. Responses and solutions to the above challenges are discernable in a myriad of cases of community outreach. Successful community engagement activities include the following key features: belonging, commitment, communication, flexibility, genuineness, relevance, sustainability, and accountability (Sung & Hepworth, 2013).

In the first decades of the 21st century, libraries are responding through *crisis informatics*: assisting those facing disasters or challenges such as job loss. Newer audiences are created out of need. Many libraries reacted to the worldwide

economic recession by developing programs for job seekers (Roy, Brzozowski, Arnold-Garza, & Beauchemin, 2011). In America, libraries contributed to the launch of the Affordable Health Care Act (referred to familiarly in the United States as Obamacare) by offering information sessions and direct assistance to patrons who want to explore, understand, and enroll in healthcare coverage options (Goldberg, 2013).

Contemporary work on community outreach includes studies of communities of need, the preparation of information professionals to serve communities of need, crisis informatics, and work on digital equity.

Selected library services have demonstrated flexibility in the area of community outreach. For instance, reference services were once only patron-initiated and took place through face-to-face contact. With the emergence of new communication technologies, reference services can now be conducted on the telephone, and through an array of virtual deliveries. Whereas a librarian might once have had difficulty finding anything related to a patron's request, now the difficulty lies in narrowing down the search to find the most relevant selections within a vast number of results. Reference library staff have responded to shifting patron needs by incorporating efficiency, instruction, point-of-need services, in addition to more traditional reader advisory services.

Some examples of incorporating outreach into elemental library services are tiered reference, roving reference, and embedded librarianship. Library efficiency is demonstrated through *tiered reference*: a reference model that provides a pyramid of service contacts in which library support staff answer uncomplicated questions and subject specialist librarians answer higher level questions. *Roving reference*—in which reference staff seek out patrons in their facilities to ensure that their questions are answered—is one approach for the internal outreach of reference services. This concept has been expanded in the performance and philosophy of embedded librarianship. *Em-*

bedded librarianship is the extension of the triad of public services—reference, instruction, and reader's advisory—into the community (Carlson & Kneale, 2011).

Serving Communities of Need

As demonstrated by Gratia Countryman, Director of the Minneapolis (Minnesota, USA) Public Library in the early twentieth century, community outreach found a home in libraries (Benidt, 1984). In this case, libraries reached out to community members directly where they lived and worked in locations as varied as factories, hospitals, flophouses, and homeless shelters. State library agencies and national libraries focused on library development, bringing traveling libraries to remote and/or small communities and establishing services for often overlooked community members such as the blind and visually impaired. Recognizing that not all potential library patrons frequented the physical library, public libraries moved outside of their buildings. Over time this was seen in bookmobile services, services in storefronts, and in portable buildings.

Librarians have a lengthy history of serving the reader, and traditional services have been defined by offering print resources and e-resources. Many studies attempt to redefine patron needs by first examining how people seek information in everyday life. Subjects of these studies include people searching for information about adoptions, immigrants living in urban settings, families coping with chronic illness, and survivors of intimate partner violence.

While outreach has had its historical foundation in the public library settings, academic libraries are also reinterpreting their missions through community outreach. Leong summarized outreach activities in academic libraries in three countries: Canada, China, and the United States. He found that outreach was expressed through (1) extending library access to the public; (2) creating information literacy services; (3) entering into

exchanges and other cooperative arrangements; and (4) hosting events such as exhibitions and lecture series (Leong, 2013).

Preparing Information Professionals to Serve Communities of Need

Professional organizations have developed a range of documents that illustrate the skills and competencies expected of information professionals. These include competencies for reference librarians, librarians serving children and youth, and those working in the subject specific areas such as music librarianship and art librarianship. Some national competencies exist, such as the Professional Registration program developed by the Library Association of New Zealand Aotearoa (LIANZA) with the National Library of New Zealand and Te Ropu Whakahau, the national association for Maori in libraries and information management (Library Association of New Zealand Aotearoa, 2013).

Whether explicitly or implicitly, graduates of programs preparing entry-level professionals to work in libraries, archives, and museums are expected to demonstrate competency in the area of community outreach. Three sample competencies are identified, as follows:

1. The Association of Library Service to Children, a division of the American Library Association, issued a competency document that outlines what is expected of public librarians serving patrons ages 0 to 14 years. Such information professionals are expected to provide “library outreach programs which meet community needs and library goals and objectives” (American Library Association. Association of Library Services to Children, 2009).
2. Librarians working in special collections are expected to be “skilled in planning and implementing programs and publications that promote and interpret the collections, such as exhibits, conferences, guest lectures,

public speaking, and other active forms of outreach” (American Library Association. Association of College & Research Libraries, 2008).

3. Standard four of the diversity standards for academic librarians asks that “Librarians and library staff shall develop collections and provide programs and services that are inclusive of the needs of all persons in the community the library serves” (American Library Association. Association of College & Research Libraries, 2012).

Specific academic programs have provided opportunities within their curricula for students to participate in community outreach, oftentimes through service learning activities. These can be operated as individual initiatives such as “If I Can Read, I Can Do Anything,” a national reading club for American Indian children and youth that is managed by students and their advising faculty member in the School of Information at the University of Texas at Austin. Students in the School of Library and Information Studies at the University of Alabama (USA) provided computer training with community members with disabilities living in west Alabama. Students can focus on specific civic engagement through independent studies, internships, Capstone or graduation projects, or through opportunities embedded within a formal course.

Several graduate programs accredited by the American Library Association have provided special opportunities for students to work directly with communities of need and/or to prepare librarians in these communities. The School of Library and Information Studies in the University of Alabama USA’s College of Communication and Information Studies offered Project ALFA, Accessible Libraries For All. AFLA focused on gathering data and helping students acquire skills in serving library patrons with disabilities. Simmons College in Boston, USA, collaborated with Harvard by means of a grant from the National Endowment for the Humanities to assist staff at

Iraqi libraries. The School of Information and Library Science at the University of North Carolina USA hosted ELIME-21, Educating Librarians in the Middle East: Building Bridges for the 21st Century. Students at the University of North Texas's Department of Library and Information Sciences collaborated to help address the needs of those living in rural areas in Jamaica. The Graduate School of Library and Information Science recognized the needs of communities and students by forming the Community Informatics Research Laboratory.

Crisis Informatics

Emergency responses may replace traditional library services, especially when a natural disaster strikes. Faculty at the University of Toronto (Canada), Florida State University (USA), the University of South Carolina, and University of Washington (USA) have studied what information is needed during the early warning phases of emergencies, technology used during the 2010 Chilean earthquake, use of information in reducing HIV/AIDS in Uganda, and preparing for and responding to hurricanes.

Digital Equity

Efforts toward reducing the digital disparity involve bringing technology and training to audiences that might not otherwise have access to equipment and learning opportunities. The Bill & Melinda Gates Foundation has demonstrated how issues of technological equity can be addressed nationally through its American libraries initiatives. Over a period of five years, the Foundation brought hardware, software, and training to public libraries in all fifty U. S. states, starting in the states with the highest percentage of children qualifying and receiving free lunches in the public schools. Before this initiative launched, only about one in four of American public libraries offered free public access computing. At the end of the program,

99 percent of the libraries provided these services (Gordon, Gordon, & Moore, 2001).

The Foundation extended the U.S. library program to tribal communities in the states of Utah, Colorado, New Mexico, and Arizona, with the Native American Access to Technology Program. After this program wrapped up, the Foundation supported the development and services of a web portal called WebJunction. WebJunction has built an online community that offers distance delivery of continuing education to library workers, as well as targeted services, including technology training in rural libraries, and among Spanish speaking clientele and job seekers (Mason, 2009). At the international level, the Gates Foundation has annually recognized innovation in providing free public access to the Internet through the Access to Learning Award. Past recipients have provided Internet access at rural markets in Guatemala, to children and their families in rural schools in China, to urban centers in Columbia, and to farmers in Africa. Recipients are recognized at the International Federation of Library Associations and Institutions (IFLA)'s World Library and Information Congress along with an award of \$1 million USD.

FUTURE RESEARCH DIRECTIONS

Those teaching and conducting research in schools of information might be involved in a wide range of investigations that address community outreach. Today, these areas include community informatics, medical informatics, global information justice, service learning, action research, evidence placed practice, information seeking, and information coping.

While an individual or organization can make a great impact on addressing community needs, the wants are unending and vast. Connectivity is still a challenge. Broadband access is practically inconceivable in many areas around the world. In Alaska, USA, for example, telephone confer-

ence calls are still more reliable than web based teleconferencing. In rural Native Alaskan villages, Internet access may be limited to one user and one computer at a time (American Library Association, 2011). Along with connectivity, a major challenge to digital equity is sustainability. Even with the help of major funders, initiatives usually have limited financial support. At the same time, there is a constant demand for retooling the skills of information workers, refreshing equipment, and meeting the needs of wave after wave of new library patrons.

There is a call for libraries and information settings to continually document their impact. Those working in and supporting information settings are encouraged to look beyond input measures such as resources added or outputs such as circulation of materials and attendance, in order to consider new ways to measure not only cost effectiveness but return on investment. In 2007, the Urban Libraries Council identified outreach to caregivers and parents as a key strategy for supporting early literacy. Outreach initiatives such as this one would create a network of partnerships committed to helping prepare the youngest citizens for school readiness through early literacy support (Urban, 2007).

CONCLUSION

Community outreach has existed as an essential reflection of the vision and mission of libraries since the beginnings of the public library movement. The elements of outreach have involved serving all members of the library's geographic boundary and beyond. These expressions have included supporting and advancing literacy as well as assisting patrons in meeting the needs, challenges, opportunities, and tragedies of their everyday lives. Such services are challenged by budgetary constraints, problematic motivations among some service providers and patrons, and varying access to technology. Still, community outreach is the local, national, and international

stage that showcases what libraries can do best: providing an equitable playing field for access to the world of knowledge.

REFERENCES

American Library Association. Association of College & Research Libraries. Rare Books and Manuscripts Section. (2008). *Guidelines: Competencies for special collections professionals*. Retrieved March 3, 2015, from <http://www.ala.org/acrl/standards/comp4specollect>

American Library Association. Association of Library Services to Children. (2009). *Competencies for librarians serving children in public libraries*. Retrieved March 3, 2015, from <http://www.ala.org/alsc/edcareers/alsccorecomps>

American Library Association. Office for Information Technology Policy. (2011). *Libraries vital in FCC's efforts to build out broadband to Native nations*. Retrieved March 3, 2015, from <http://www.ala.org/offices/oitp/publications/officialfilings/officialfilings>

American Library Association. Association of College & Research Libraries. Racial and Ethnic Diversity Committee. (2012). *Diversity standards: Cultural competencies for academic libraries*. Retrieved March 3, 2015, from <http://www.ala.org/acrl/standards/diversity>

American Library Association. (2013). *Key action areas*. Retrieved March 3, 2015, from <http://www.ala.org/aboutala/missionpriorities/keyactionareas/>

American Library Association. Office of Literacy and Outreach Services. (2013). *Mission*. Retrieved March 3, 2015, from <http://www.ala.org/offices/olos>

Benidt, B. W. (1984). *The library book: Centennial history of the Minneapolis Public Library*. Minneapolis, MN: Minneapolis Public Library and Information Center.

- Carlson, J., & Kneale, R. (2011). Embedded librarianship in the research context; Navigating new waters. *College & Research Libraries News*, 72(3), 167–170.
- Carr, J. F. (1919). *Library work with the foreign born. Exploring a neighborhood: Our Jewish people from Eastern Europe and the Orient*. New York: Immigrant Publication Society.
- Goldberg, B. (2013). Libraries stress neutrality as they prep for queries on health care law. *American Libraries*, 44(9/10), 13.
- Gordon, M. T., Gordon, A. C., & Moore, E. (2001). New computers bring new patrons. *Library Journal*, 126(9), 134–138.
- Gray, S. W. (2012, August). *Reference librarianship on the fly: Taking the librarian out of the library*. Paper presented at the World Library and Information Congress, Helsinki, Finland.
- Heim, K. M. (1982). Stimulation. In G. A. Schlachter (Ed.), *The Service imperative for libraries: Essays in honor of Margaret E. Monroe* (pp. 120–154). Littleton, CO: Libraries Unlimited.
- Leong, J. H. T. (2013). Community engagement – building bridges between university and community by academic libraries in the 21st century. *Libri*, 63(3), 220–231.
- Library and Information Association of New Zealand Aotearoa. (2013). *Professional registration*. Retrieved January 7, 2013, from <http://www.lianza.org.nz/registration>
- Mason, M. G. (2009). WebJunction: A community for library staff. *Journal of Library Administration*, 49(7), 701–705.
- McCook, K. P. (2011). *Introduction to public librarianship*. New York: Neal-Schuman.
- Montiel Overall, P. (2009). Cultural competence: A conceptual framework for library and information science professionals. *Library Trends*, 79(2), 175–204. doi:10.1086/597080
- Osborne, R. (2004). *From outreach to equity: Innovative models of library policy and practice*. Chicago: American Library Association.
- Roy, L. (2009). Service learning connecting diverse communities and LIS students and faculty. In L. Roy, K. Jensen, & A. H. Myers (Eds.), *Service learning: Linking library education and practice* (pp. 73–82). Chicago: American Library Association.
- Roy, L. (2011). Weaving partnerships with the American Indian peoples in your community to develop cultural programming. In L. Roy, A. Bhasin, & S. K. Arriaga (Eds.), *Tribal libraries, archives, and museums: Preserving our language, memory, and lifeways* (pp. 141–156). Lanham, MD: Scarecrow Press.
- Roy, L., Brzozowski, B., Arnold-Garza, S., & Beauchemin, K. (n.d.). ‘Are you searching for a new job?’: Texas public libraries provide services for job seekers. *Texas Library Journal*, 87(1), 30–33.
- Smallwood, C. (2010). *Librarians as community partners: An outreach handbook*. Chicago: American Library Association.
- Snyder, C. A. W. (2002). Measuring library service quality with a focus on the LibQUAL+ project: An interview with Fred Heath. *Library Administration & Management*, 16(1), 4–7.
- Sung, H., & Hepworth, M. (2013). Modelling community engagement in public libraries. *Malaysian Journal of Library & Information Science*, 18(1), 1–13.
- The University of Illinois at Urbana-Champaign. Graduate School of Library and Information Science. (2013). *Robert B. Downs Intellectual Freedom Award*. Retrieved March 3, 2015, from <http://www.lis.illinois.edu/about-gslis/awards/downs-award>

The University of Texas at Austin. Bureau of Business Research. (2012). *Texas public libraries: Economic benefits and return on investment*. Retrieved March 3, 2015, from <https://www.tsl.state.tx.us/roi>

Urban Libraries Council. (2007). *Making cities stronger: Public library contributions to local economic development*. Chicago: Urban Libraries Council.

Welburn, W. C., Welburn, J., & McNeil, B. (2010). *Advocacy, outreach and the nation's academic libraries: A call for action*. Chicago: American Library Association.

ADDITIONAL READING

American Library Association. Office for Literacy and Outreach Services. (2008). *Handbook for mobile services staff*. Retrieved March 3, 2015, from <http://www.ala.org/offices/olos/bookmobiles/mobileservices>

American Library Association. Office for Literacy and Outreach Services. (2010). *Keys to engaging older adults @ your library*. Retrieved March 3, 2015, from <http://www.ala.org/offices/olos/toolkits/olderadults>

American Library Association. Office for Literacy and Outreach Services. (2012). *Extending our reach: Reducing homelessness through library engagement*. Retrieved March 3, 2015, from <http://www.ala.org/offices/extending-our-reach-reducing-homelessness-through-library-engagement>

Armour, G. (2010). Communities communicating with formal and informal systems: Being more resilient in times of need. *Bulletin of the American Society for Information Science and Technology*, 36(5), 34–38. doi:10.1002/bult.2010.1720360510

Blake, B., Martin, R. S., & Du, Y. (2011). *Successful community outreach: A how-to-manual for librarians*. Chicago: American Library Association.

Boff, C., Singer, C., & Stearns, B. (2006). Reaching out to the underserved: More than thirty years of outreach job ads. *Journal of Academic Librarianship*, 32(2), 137–147. doi:10.1016/j.acalib.2005.12.007

Courtney, N. (2008). *Academic library outreach: Beyond the campus walls*. Westport, CT: Libraries Unlimited.

Davis, D. (2009). Outreach to non-English speakers in U. S. public libraries: Summary of a 2007 study. *Public Libraries*, 48(4), 13–19.

DePo, A., Colletti, M., Hallock, E., & Lewis, N. (2011). Free to Learn: Best practices for serving former prisoners in public libraries. *Colorado Libraries*, 36(1), 1–5.

Diaz, R. (2005). Developing library outreach programs for migrant farm workers. *Florida Libraries*, 47(1), 12–14.

Featherstone, R. M. (2012). The disaster information specialist: An emerging role for health librarians. *Journal of Library Administration*, 52(8), 731–753. doi:10.1080/01930826.2012.746875

Ferguson, S. (2012). Are public libraries developers of social capital? A review of their contribution and attempts to demonstrate it. *The Australian Library Journal*, 61(1), 22–33. doi:10.1080/00049670.2012.10722299

Fontenot, M. (2013). Five “typical” years as an outreach librarian and five things I have learned. *College & Research Libraries News*, 74(8), 431–432.

Grace, D., & Sen, B. (2013). Community resilience and the role of the public library. *Library Trends*, 61(3), 513–541. doi:10.1353/lib.2013.0008

Hawk, M. (2008). Bookmobile services in rural America: Past, present, and future. *Bookmobile & Outreach Services*, 11(2), 57–78.

Jensen, B. (2002). Service to day laborers: A job libraries have left undone. *Reference and User Services Quarterly*, 41(3), 223–233.

Jones, P. A. Jr. (2004). *Still struggling for equality*. Westport, CT: Libraries Unlimited.

LeRoux, C. J. B. (2010). Social informatics vs. community informatics: A brief overview of their origins and current status. *Mousaion*, 28(1), 34–44.

Lewis, J. (2013). Information equality for individuals with disabilities: Does it exist? *The Library Quarterly*, 83(3), 229–235. doi:10.1086/670697

Marco, G. A. (2012). *The American public library handbook*. Santa Barbara, CA: Libraries Unlimited.

Mars, A. (2012). Library service to the homeless. *Public Libraries*, 51(2), 32–35.

Pointon, S. E. (2009). Library outreach is the future! *Public Libraries*, 48(3), 2–5.

Rudin, P. (2008). No fixed address: The evolution of outreach library services on university campuses. *The Reference Librarian*, 49(1), 55–75. doi:10.1080/02763870802103761

Singer, D., & Agosto, D. (2013). Reaching senior patrons in the digitized library. *Public Libraries*, 52(6), 38–42.

University of North Texas. *PEARL project: Community outreach plans*. Retrieved March 3, 2015, from <http://pearl.unt.edu/community-outreach-plans>

KEY TERMS AND DEFINITIONS

Advocacy: Informing the public of the library offerings, the library as work setting, and the library profession.

Community Informatics: Designing services for targeted members of the library's user community.

Crisis Informatics: Designing services that assist library patrons who are impacted by, or might be impacted by, traumatic events.

Digital Equity: Promoting and designing services that provide all members of the community with access to online information and the skills and tools to access, use, and evaluate online content.

Information Coping: The ability to handle news, alerts, messages, data, music or other information formats in a way that results in learning and understanding with minimal anxiety.

Information Seeking: The habits and patterns that humans undertake in locating answers to questions, satisfying needs, and fulfilling desires for learning and entertainment.

Service Learning: Addressing community needs through course content and personal reflection.

Exploring “Hacking,” Digital Public Art, and Implication for Contemporary Governance

P**Amadu Wurie Khan***University of Edinburgh, UK***Chris Speed***University of Edinburgh, UK*

INTRODUCTION

This article presents the application of the online (Internet) ‘hacking’ concept to community life and processes from two hypothetical contexts: First, it was hypothesised that technology could be ‘hacked’ into by disadvantaged communities to create a digital public art. Second, the community-generated digital art platform could in turn be used to ‘hack’ into images and memories to facilitate the sharing of conversations and identity, social engagement, and digital inclusion among residents. The article therefore presents how these contexts of the characteristics and practicality of online ‘hacking’ inspired the design and functionality of a community digital artwork in a disadvantage urban estate in Edinburgh, UK. In addition, the article considers the implication of the ‘hacking’ practices by and among disadvantaged communities for realizing the social action, social engagement and networked society goals of the UK Government’s ‘Big Society’ policy. This is significant because the ‘Big Society’ agenda promotes an interactive networked culture that has transformative potential to connect citizens, build knowledge and continuous learning and regenerate communities at a time of economic austerity in the UK (Mayo & Steinberg, 2007; Speed, Khan & Phillips 2016).

The article is presented as follows. The following section is the conceptual path-clearing.

It traces the etymology and usage of the concept of ‘hacking’ from the techno-scientific domain. Section three makes an attempt at disambiguating the kinds of ‘hacking’ practices that are relevant to issues of community relations and processes. Section four presents the practical application of the ‘hacking’ concept that culminated in the social design of a physical digital public art, the ‘totem pole’. In section five, the implications of ‘hacking’ for the ‘Big Society’ policy is considered. Section five provides suggestions for a research agenda that could generate future design interventions that are inspired by concepts associated with digital media culture and for realizing forms of contemporary governance such as the ‘Big society’.

CONCEPTUAL BACKGROUND: THE “HACKING” FOLKLORE

In tracing the etymology of ‘hacking’, folklore rather than the history of the concept should be prioritised. This is because there are many narratives to explain the emergence of the concept and its incorporation into contemporary public discourse (Devitt, 2001). History could be subjective, but mainly expected to be a precise and accurate record. Folklore, although rooted in historical narrative and passed down across generations, is not expected to carry the kind of accuracy as history should. This is because its mainly verbal

form of transmission and its inherent performance element makes folklore vulnerable to variability and manipulation (Khan, 2009). A word of caution though! History could be handed down verbally too, and like folklore could be accurate, written down and passed through generations. Nonetheless, folklore is associated with traditional stories, gossip, myths and legends, all of which are traditional art forms that are characterised by dubiety, as might be the case with history. A look at the many romanticised accounts of the origin of ‘hacking’, lends weight to prioritising the folklore around the origin of the concept over historical accounts.

Against this background, folklore has it that ‘hacking’ originated from the realm of technology as student slang at the Massachusetts Institute of Technology (MIT) between 1950 and 1960 (see Levy, 2002; Devitt, 2001). MIT is said to have been amongst the first institution to offer courses in computer programming and computer science, and that it was on such a course that group students taking a class on artificial intelligence came to coining the word ‘hacker’. Students used the term to refer to their ability to manipulate a computer to perform actions not intended for that program. It has also been suggested that the term was used to convey a sense of performing a “practical joke and feeling of excitement because the team member would ‘hack away’ at the keyboard hours at a time.” (Moore, 2006). Examples of ‘hacking’ folklore associated with MIT include: in 1964, MIT students (hacks) placing a convincing replica of a campus police car on top of the Institute’s Great Dome as a form of practical joke, manipulating electric trains to make it perform faster and more efficiently (see Levy, 2002; Moore, 2006). Ward (BBC 27 Oct 2000), argued that ‘hacking’ originally meant “an elegant, witty or inspired way of doing almost anything”. Since then, the practice of ‘hacking’ has been evident in or applied to other fields and scholarly or technical communities, and not just limited to technology. The next section briefly explores these folkloric dimensions of

‘hacking’, namely ‘phreaking’ & ‘cracking’, and ‘hobby-ism’ & ‘prosumerism’.

“Phreaking” and “Cracking”

In the 1970s, ‘phreaking’ or phone ‘hacking’ emerged by which ‘hacks’ manipulated telephones to make free calls. A legendary figure in this respect was, a ‘phreaker’ John Draper, who discovered that a whistle included as a free gift in boxes of Captain Crunch cereal emitted a 2,600 hertz pitch which was the frequency used to indicate operator calls to phone exchanges (Burnham, 2009; Lapsley, 2011). Blowing the whistle into the mouthpiece of the phone meant that the call was seen to come from an operator and hence no charges were levied on the call. Not only did John Draper become known in the hacking world as Captain Crunch, but his work is seen to have inspired Steve Wozniak and Steve Jobs who went on to found Apple Computers (Burnham, 2009). By the 1980s, ‘phreaking’ was evident in computers in the form of Bulletin Board Systems (BBS), which is believed to be the precursor to the yahoo groups of today (Levy, 2002; Lapsley, 2011). In addition to enabling individuals to post messages of any kind of topics, the BBS specialized in disseminating information on how to break into computers, how to use stolen credit card numbers and share stolen computer passwords (see Levy, 2002). Known as ‘cracking’, the practice entails the circumventing computer security and unauthorized remote computer break-ins via a communication networks such as the Internet.

From the above practices, two distinct but interconnected senses of the term ‘hacker’ and ‘hacking’ in the domain of technological science are discernible: the modification of use and the breaking of codes/security (Raymond, 2001). This often came together in particular ways to establish particular but interconnected areas in the domain of technological science: computer programming and computer security. In the former sense of the term ‘hacker’, emphasis is put on modifying computer programmes/technologies

so that they can perform new uses, while in the latter (code-breaking) the emphasis is placed on breaking security features, either as an end in itself or as a way so that the technology can be used in ways other than originally intended. It has been suggested that the terms ‘hacker’/‘hacking’ might be restricted to the first sense, whilst the terms ‘cracker’/‘cracking’ might be used to refer to the second set of practices (Cramer, 2003). According to Raymond (2001), ‘hackers’ from the field of programming usually work openly and use their real name, while computer security hackers (or crackers) prefer secretive groups and identity-concealing aliases. Also, their activities in practice are largely distinct. The former focus on creating new and improving existing infrastructure (especially the software environment they work with), while the latter primarily and strongly emphasize the general act of circumvention of security measures, with the effective use of the knowledge (which can be to report and help fixing the security bugs, or exploitation for criminal purpose) being only rather secondary.

“Hobby-ism” and “Prosumerism”

There is another sense in which ‘hacking’ and ‘hackers’ include ‘hobby-ism’, in which case individuals show commitment and passion in being creative in a given field in ways that are similar to professionals (Burnham, 2009). An example often cited is that of ‘circuit bending’, a practice associated with the techno-music sub-culture. By this way, ‘hobbyists’ customized the circuits within electronic devices such as low voltage, battery-powered guitar effects and small digital synthesizers to create new musical sound (Burnham, 2009; Early, 2007; Levy, 2002). This results in the creation of the strange, dis-harmonic digital tones that became part of the techno-music style. ‘Hobbyists’ are in the main customers of products who rather than just accept the manufacturer’s design of a product, will try to adapt it to their tastes in order to get a desired aesthetic and utility

value. It has been observed that such activities have been tapped into by manufacturers in an attempt to widened customer base and profit margins. This is by the appropriation of ‘hacker’/‘hobbyist’ generated designs or specifications into new products that are sold into the market. The ‘hacker’ or ‘hobbyist’ therefore transforms from a consumer to what Toffler (1980) referred to a ‘prosumer’. In this situation (prosumerism), the role of producer/professional and consumer merges to participate in the design and production requirements of products through mass customisation. Manufacturers adapt or create their products to the specific requirements of consumers in order to keep the latter satisfied, while generating profit.

Others argued that ‘prosumers’ include consumers who rather than just accept or purchase what is on sale would actively search the market and seeks to influence the market (Crakburn, 2003). ‘Prosumers’ therefore do not only influence or dictate design, aesthetic and utility values, but also influence the markets. In this sense, ‘prosumerism’ arguably makes consumers part of the creative and marketing or pricing process, as they impart their individual and or collective preferences.

The above folkloric account illustrates the nebulous origins of ‘hacking’ and its manifestations (phreaking, cracking, hobby-ism, and prosumerism) within science and technology: mechanics, computers (both programming and security), and telephonics, as well as in manufacturing. These manifestations of ‘hacking’ continue to be shrouded by accounts that have been subjected to manipulation and variability, and therefore continues to be the terrain of folklore. This, in turn, makes ‘hacking’ a romanticized construct, and a contested concept (Devitt, 2002; Thomas, 2002). ‘Hacking’ has therefore become a metaphor for the practices and actions that exploit (or explore) weaknesses or deficiencies in a system to behave or function in a certain way (Kulikauskas, 2004). The next section takes a look at some key social characteristics of ‘hacking’.

“HACKING”: AND ITS SOCIAL CONTEXT

Arguably, the actions by ‘phreakers’, ‘crackers’, ‘hobbyists’ and ‘prosumers’ constitute a creative process that is performed in a social environment (see Kulikauskas, 2004). The socially creative process involves risk-taking or the circumvention of limitations, and entails self-reliance, and learning through the sharing of skills and resources (Fitch, 2003). The creativity is motivated by intellectual, functional and aesthetic challenges (see Cramer, 2003). It is also a conscious technical act or practice that individuals undertake towards a preconceived outcome that was not intended by the original creators of the ‘thing’ that is ‘hacked’ (Levy, 2002; Dan, 2011). It therefore has the following social characteristics, namely reciprocity, resilience and moral ambiguity (Fitch, 2003; Lemos, 2002).

“Reciprocity” and “Resilience”

‘Hacking’ depends on reciprocity. This is better exemplified in computer gaming. As Crabtree (2003) explains, when people are stuck in a computer game, they will go to a gaming community online, and ask others for advice. Other gamers will help on the principle of reciprocity; they will sometime in future expect someone else (including those they have helped) to help them out with their gaming problems. Reciprocity is therefore a form of harnessing social capital among society’s members. Through reciprocity people turn to each other when they encounter everyday problems (civic, social, political, economic etc.) to access knowledge, skills, and advice from others who have expressed similar problems in a bid to solve or help overcome them. ‘Hacking’, in this sense, becomes a social, or in Crabtree’s conception a ‘civic’, phenomenon (Crabtree, 2003). It thrives on symmetrical and synchronous relationship. as well as on individual and collective resilience to cope with emerging problems and demands that citizens (non-citizens too) encounter in their everyday lives in the community (see Crabtree, 2003).

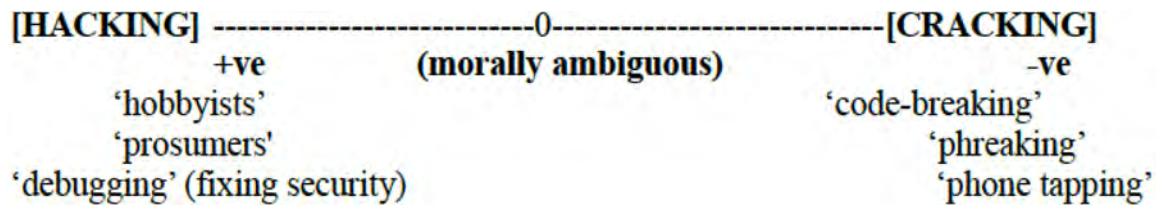
This is not to say that ‘hacking’ is always motivated by reciprocity. It might be that there are altruistic and communitarian reasons behind an individual’s sharing of skills and resources to facilitate ‘hacking’. For example, helping others may be a source of ‘feel-good’ factor as well as perceived as a civic and social responsibility that an individual owes to others. It might also be that some individuals might not give back or reciprocate either in the same way or in kind as other participants. They simply just take away, consume or benefit from what others have contributed without returning such favours; neither would they expect similar favours from others who benefit from their contribution. In this case, the relationship is asymmetrical and asynchronous by which an individual may not reciprocate magnitude and content or information as required. It is also possible that in cases of reciprocal response, this is not spontaneous, but deferred to a later date (Hrastinki, 2008).

The “Moral Ambiguity”

From the foregoing discussion, it could be suggested that hackers’ socially creative practices could be perceived to be ‘legal’ or ‘illegal’ (Fitch, 2003, Lemos, 2002). For want of clarity, we suggest that, the term ‘hacking’ is reserved or applied to ‘legal’ activities, while ‘cracking’ is for the ‘illegal’ or transgressive ones as Figure 1 shows.

Yet, we know that these conceptualisations are fraught with ‘moral ambiguity’. By ‘moral ambiguity’ is meant that individuals’ perceptions and interpretations of the legitimacy of a ‘hacking’ practice is subjective and influenced by their moral values and beliefs. This begs a three-dimensions to conceptualizing ‘hacking’ as Figure 1 shows. At the extreme left and moving towards the right of the continuum [up to point zero (0)] of Figure 1, are forms of positive ‘hacking’ that are normally perceived as good and legitimate or legal. Examples of this kind of ‘legal’ activity encompasses ‘hobbyists’ and ‘prosumers’ as indicated in Figure 1. As ‘hobbyists’, they are not necessarily profes-

Figure 1. The “hacking”/“cracking” continuum



sionals, but amateur enthusiasts seeking to adapt and influence the aesthetic and utility value of a product. ‘Hackers’ are ‘prosumers’ because they openly and ‘legally’ participate in the designing and production chain from the outset as well as influencing market forces. The activities are, in the main, facilitated and supported by manufacturers.

‘Cracking’ is diametrically opposite to this practice, widely perceived as illegitimate or illegal and indicated on the extreme right of the continuum [moving from point zero (0) to the right]. Examples normally perceived as illegitimate and transgressive actions include code-breaking, ‘phreaking’ and phone tapping. However, ‘code-breaking’ can be perceived as ‘legal’ and or ‘transgressive’ depending on an individual’s moral standpoint. There is therefore the possibility of a continuum of ‘hacking’ and ‘cracking’ practices, which could be designated as ‘morally ambiguous’, which is around ‘point zero’ (0) or in the middle of the continuum. It should be envisaged that there will exist a plethora of on-line and off-line practices that could be located in the continuum, albeit it underpinned by the moral persuasion of individuals. More importantly, what constitute ‘legal’ or ‘illegal’ and their boundaries of demarcation are not only based on an individual’s ‘subjectivities’, but also constructed by legislation and policies of states, supra-national institutions and industry. Nonetheless, the illustration is intended to serve a heuristic purpose that is informed by the perception of individuals in ascribing ‘good’/‘bad’ or ‘legal’/‘illegal’ to ‘hacking’. The illustration also demonstrates that both ‘hacking’ and ‘cracking’ practices are underpinned by imaginativeness, intentionality and functionality (see Lemos, 2002; Fitch, 2003).

The above characteristics and application of ‘hacking’ provided the inspiration for a community-generated digital art, known as the digital ‘totem pole’ (see Chris et al. 2016). The key hypothesis was that technology could be ‘hacked’ into to create a digital public art, which have implications for the realization of UK’s ‘Big Society’ policy among disadvantage communities. The rest of the article presents how online ‘hacking’ inspired the social design of the ‘totem pole’ and its utility for community life and processes.

FROM THEORY TO DESIGN APPLICATION

This section is divided into two parts. The first describes the design locale and process of community engagement that culminated in the construction of the public digital artwork. This is followed by a discussion of the practicality and functionality of the digital ‘totem pole’ for the ‘Big Society’ as a form of contemporary governance.

Wester Hailes and the Design Process

Wester Hailes is a huge housing estate constructed in the 1970s on the Western outskirts of Edinburgh, the capital city of Scotland, UK. It has been afflicted with high levels of social and economic deprivation, crime and unemployment. The social and economic challenges of the area have provided the impulse for local residents to organise community development and service delivery initiatives. In addition, Wester Hailes has historically deployed community art towards

achieving community development, regeneration and empowerment. Community art has also been central to the desire to project a positive image of the community to contrast a mainly negative representation of its residents.

Two local service providers, Prospect Housing Association (aka Prospect) and Whale Arts expressed an interest in using social media as a platform to share ideas, photos and memories. Prospect set up a Facebook page and began posting images of the area that were originally published in the community newspaper; the *Wester Hailes Sentinel*; latterly the *West Edinburgh Times*. The page quickly became popular and with photographs attracting many comments about who, when and where they were taken (<http://on.fb.me/mOPPwp>). This ‘write back’ facility began to enable residents to recover memories of the past and drew out many connections beyond the image itself. By the summer of 2010, the research team and local residents agreed on a design method that encapsulated ‘hacking’. This involved the development of platforms that facilitated the public ‘writing back’ on to the photos and images of Wester Hailes. To enable this process, community members chose the construction of

a large wooden digital ‘totem pole’ to be located within the neighbourhood.

WHALE Arts coordinated the production of the ‘totem pole’ (see Chris, Khan & Phillips, 2016). A steering group of community members and project partners was created to ensure that clear targets are set and achieved. The steering group also facilitated networking, engagement and capacity building that underpins the five participatory workshops within the community (Figure 2).

The workshops were held during local civic events including the Tenth anniversary celebration of the Union Canal, at the AGM of service providers including the Prospects Housing Association and WHALE community arts project, and at the annual community Road Show in the Westside Plaza shopping centre. Participants to the workshops include local residents, staff of services and community activists. Project staff used the workshops to explain the rationale, process and outcomes with a view to generate interest and recruit local residents to participate in its development and delivery.

A central element of the workshop was the display of a portable banner with embedded quick

Figure 2. Residents learning to scan



response (QR) codes (as proof of concept) and historical photographs depicting people, events and places of Wester Hailes on the outer wall of a portable shed (see Chris, Khan, & Phillips, 2016). Both the portable banner and photographs were used in combination to introduce local residents with different levels of technological expertise and interest to the technology used in the project. Participants were asked to look at the photographs to trigger their memories or any other associations that they might have with the

people, event or place depicted. They were then encouraged to share their story/memories about these, which were captured through a voice and film recorder. Participants were also encouraged to scan the QR codes embedded on the photographs and to record their stories and upload them to the *Tales of Things* website (and the Wester Hailes social history archive). They were then asked to scan and listen to the replay. This exercise was to give them experience of the opportunities that the technology would offer, which is the ability to read and write into the codes. The workshops were also opportunities for participants to ask questions and provide any suggestions or views about the project and how best to improve it and to get their sustained involvement in its future development (see Chris, Khan & Phillips, 2016). The overall ethos of the workshops was one that promoted engagement with local residents and their exposure to new developments in web-based technology in a way that was empowering, collaborative, non-threatening and meaningful. The workshops were therefore central to community participation in all aspects of the totem pole project including the design, time-scales and location of the 'totem pole' (see Chris, Khan, & Phillips, 2016).

During the carving of the pole and the ensuing deliberations, both residents and service providers realised that the pole had a potential for service delivery in the locality. It was therefore decided that, in addition to sharing stories and memories of the area, the 'totem pole' should include 5 QR codes that give access to a variety of services. The final product was a carved wooden digital 'totem pole', which was situated within Wester Hailes (Figure 3 above). Its QR codes provided a physical platform for 'hacking' images (through the ability of people being able to comment and create new meanings for the images) and sharing conversations about the area. The QR barcodes were also gateways to cloud based material relevant to the location of the pole. Local residents could scan one of the labelled tags to access and contribute to historical photographs, stories, video and audio clips (Figures 4 and 5). By so doing, the pole act

Figure 3. The installed digital "totem pole"



Figure 4. Local residents scanning the QR codes



Figure 5. Use of historical photos to evoke memories of past community activists. Residents could “write” memories on to QR codes and attach them to the photograph.



as a social resource to help build connections between the people and the place, as well as drawing upon online resources (see Marggets, 2011; Chris, Khan & Phillips, 2016).

What then can be concluded from the design and practicality of the ‘totem pole’ that is relevant to developing an understanding of the intersection of the digital media culture of ‘hacking’, public art, and the ‘Big Society’ policy? To answer this question, a brief account of the main features and aims of the policy would be imperative.

“Hacking” and Implications for the “Big Society”

The ‘Big Society’ is a policy mainly championed by the governing Conservative Party in the UK (Szieter & Ishkanian, 2012). It is aimed at encouraging citizens (and non-citizens) to participate collectively and collaboratively in local initiatives. It is intended to make policy-makers transparent and accountable to the recipients of public services. By so doing, the policy is a bottom-up

process that transforms policy-making and service delivery from a centralised bureaucratic control system to one that is grass roots-led. It's proponents argued that this would make service provision decentralised and localised. Proponents also argued that this 'localism' enable services to be tailor-made to the needs, tastes and specifications of local people, and at the same time delivered in a better improved quality (see Lesley2010). In this sense, service users (citizens) are constructed as not only consumers, but 'prosumers' as they become stakeholders in the design of services from the outset of their conception and formulation. As 'prosumers', citizens depend on each other to pull skills and resources together for mutual benefit. Grass-roots involvement in the formulation, delivery and making choices on services, access information and knowledge, and hold public officials to account make them become empowered social actors (see Chadwick, 2009). The 'Big Society's' originator, British Prime Minister David Cameron claimed that the policy will enable people powerful enough to help themselves and is the "biggest, most dramatic redistribution of power from elites to the man in the street" (Green Paper No. 14). The policy could enable a fluid and evolving response by individuals to challenging circumstances, eradicate structural inequalities, and facilitate civic engagement, social connectivity, networking, and innovation (see Szieter & Ishkanian, 2012).

Given these goals of the policy, the 'totem pole' became relevant to the 'Big Society' agenda among the research team, and residents. It serves as a networked museum or repository for audio, written and visual recordings of current and past memories, images, narratives, works and ideas of local residents and others in the Diaspora. While these benefits of community digital art have been observed elsewhere (see for instance, the 'Talking Poles' in Moulder et al., 2011), the totem pole moved beyond its aesthetic and archival value to have a symbolic relevance to Wester Hailes as a community with historic problems of marginalisation. The general feeling among

residents was that the 'totem pole' served a symbolic function – that of community resilience and regeneration. It symbolises the community's resilience to contest negative depictions of their community as afflicted with social delinquencies and deprivation. The process of designing and delivering the totem pole depended on the input of residents. As others have observed, most public artwork that incorporate digital technology and produced by non-professional artists tended to exclude people from the final stages of the creative process (Moulder et al., 2011, p.2). In contrast, the pole's social design approach facilitated community participation at every stage - from the design process of the pole and the QR codes, the content and themes therein contained to the installation of the finished product. The pole's design therefore reflected residents' design preferences, aspirations and expectations of their locality (see Chris, Khan & Phillips, 2016). It made residents to be co-creators of the artwork and for them to have equal ownership of the whole project.

The design process also contributed to the skills of locals for it to be a successful participatory community art project (Moulder et al., 2011, p.8; Ackoff, 1974). The lack of digital technology skills including social media, therefore, did not preclude anyone from participating. As explained earlier, the workshops built such skills among those that lacked them. Overall, the process of design nurtured an engagement between professional and non-professional artists, academics and non-academics, and digital media technologists and non-technologists. The social connectedness and networked communities was only possible by the in-built 'read-write' component of the 'totem pole'. In this way, the project generated significant social capital by engaging with the diverse ages, backgrounds and interests present in the Wester Hailes community to explore and capture their memories of the area. It is also to articulate a collective future ambition for the community.

In summary therefore, as a 'hacking' inspired design intervention, the collaborative production of the 'totem pole' became tangible to the ideals

of the ‘Big Society’ agenda. Firstly, it was an embodiment of social action, civic engagement, and networking. This is because it was co-produced artwork between residents, service providers and the academic/research team to realise the community’s aspirations. Secondly, it generated a sense of community, empowerment, and identity with the estate among residents (Tonnies, 1957). This sense of community moved across geographies of locality, Diaspora, shared interest and generations. By so doing, it generates a feeling of identity with Wester Hailes among these disparate geographies. Thirdly, the sharing of images and resources constitute resilience, resourcefulness and reciprocity. For instance, the renewed sense of identity, participation in the design and use of the pole demonstrates the resilience of residents to regenerate their community at a time of decline in services such as local newspapers and economic austerity. The sharing of information, memories and images largely depend on the community pulling together its resources in a way that is reciprocal (see Chris, Khan & Phillips, 2016). These positive social processes are akin to ‘civic/social hacking’.

However, it was the view of some residents, service providers and the research team that the ‘pole’ might be used by the public in deviant or transgressive ways. For instance, there were concerns that its ‘read-write’ facility might attract unfair, impolite and gratuitous comments against residents, service providers or politicians. If these were to occur, they would constitute ‘cracking’, a phenomenon associated with the ‘hacking’ culture as enunciated earlier. To address this unwanted and transgressive use of this public art technology, an editorial policy was put in place to moderate and monitor public comments before being posted on to the ‘pole’. Although this was done in consultation with residents, it presents a moral and ethical dilemma that would confront any digital project that is community-led as would be with the ‘Big Society’ policy. For instance, how would the powers that be (meaning researchers,

services, policy-makers and politicians) respond to activities of communities that are outside the stipulations or rules of engagement of a public art or service? Would the transgressive behaviour by individuals and communities, albeit it in their own interest and benefit, be tolerated? If tolerated, then it would make citizens to act like ‘prosumers’ who are involved in deciding on the type of service and how it is used or its usage. The extent to which such individual and collective interventions are acceptable or not acceptable, justified or not justified and whether or not they constitute social deviant behaviour are bound to be debatable.

In addition, the participation and reciprocal relationship among stakeholders (meaning researchers, services and residents) in the production and functionality or workings of the pole could be asynchronous, rather than synchronous. For example, in a socially disenfranchised community like Wester Hailes, not every one had access to the Internet or interactive mobile phones (see Chris, Khan & Phillips, 2016). Many individuals could not afford the financial resources and skills required to use this technology, despite the training provided as explained earlier. These individuals are not likely to benefit from this ‘hacking’ activities that the functionality of the ‘pole’ depended on. This could further widened the gap between the empowered and the marginalized as level of participation based on their abilities and resources at their disposal are uneven. It is a potential risk to achieving the ‘Big Society’ agenda: not all individuals and communities will participate either in all services or activities or in equal measure. The project claimed to promote digital inclusion in so far as many residents who had no experienced of interactive media were provided with the skills to use the ‘pole’. Nonetheless, the structural inequalities in society are bound to affect, in different ways, individual and community participation in the technology.

Another moral dilemma was that, Internet-facilitated ‘hacking’ has an uneasy relationship with the goals of the ‘Big Society’. Firstly, this

type of 'hacking' is mainly virtual (internet), and confronts the purpose of face-to-face social contact, social interaction and social engagement among citizens. Secondly, the 'hacking' through everyday social practices and sharing is conducted on a voluntary basis, albeit with the expectation of reciprocity. In the context of the 'Big Society', it is likely that such commitment and voluntarism will vary among society's members. The received wisdom is that volunteering is the terrain of the middle class, and so deploying 'hacking' as a mechanism for pursuing the 'Big Society' agenda might exclude disenfranchised communities (see Chris, Khan & Phillips, 2016).

FUTURE RESEARCH DIRECTIONS

The exploration of the 'hacking' concept and its application in social design and policy interventions is heuristic. Research and design energies should therefore be devoted towards understanding the intersections between online practices (virtual communities) and offline practices, or the daily experiences of individuals in the policy context of contemporary governance. The nature of the investigation should aim at offering radical insights into how disadvantaged communities will develop any means possible to overcome the challenges they face especially at a time of financial cuts that are likely to impact upon them. These design and research interventions could be understood through the use of cross-disciplinary research: social science, arts and humanities and industrial models of co-design with communities that are impacted by policies.

In addition, it has been highlighted that 'hacking' is a practice that is fraught with moral ambiguity: that of whether it is acceptable or non-acceptable to practice 'hacking' (Kulikaskas, 2004). The pejorative conception of the term is compounded by contemporary media representation of 'hacking' as criminal and social delinquency. 'Hacking' in this sense is associated with criminal acts such as identity theft, credit card

fraud and other computer-related crimes (see also Moore, 2006). The 'phone hacking' saga in the UK culminating in the judicial 'Leveson Inquiry' in 2011/2012 is a reminder that, in addition to legislative and policy considerations of the state and industry, moral ambiguity play a role in the public determination of what constitute legitimate or illegitimate 'hacking' (Leveson, 2012). The criminal connotation of the term is believed to have caused a segment of the computer community to ascribe the term 'hacker' to those engaging in legal activities. As already explained, they ascribe 'cracker' to those engaging in nefarious 'hacking' activities including those performing computer break-ins. Nonetheless, the term continues to be used by actors in the techno-scientific field for two reasons: firstly, the intended meaning can be based on the context of usage in order to clarify which meaning is intended (pejorative or complimentary); secondly, the practice describes a set of skills which are used for various reasons including nefarious criminal activity (Moore, 2006). Researchers should therefore engage in empirical investigation to gather evidence to inform this debate about what is morally legal or illegal, and who has the right or moral authority to decide these activities as indiscretions.

Research should also investigate (empirically and or conceptually) the utility in employing the terms 'social or civic hacking' and 'social or civic cracking' for depicting forms of 'hacking'. Questions that could be empirically explored include: Should the term 'social/civic cracking' be applied to hacking activity among communities that are 'illegal', transgressive or contravenes social norms, policy and legislation? Conversely, should the perceived community 'legal' hacking practices be ascribed as 'social/civic hacking'? Exploring these areas would develop our understanding on how communities use digital technology for empowerment, resilience and resourcefulness beyond that which the state is currently willing or able to provide.

Finally, all of the aforementioned research areas could also focus on affluent or middle class com-

munities. This would enable an understanding of the motivations, aspirations and resourcefulness of this segment of society in harnessing technology for transgressive or legal pursuits to cope with the challenges they face. Such research should generate insights that would enable a comparative understanding of the use of technology along socio-economic demographics in the polity. It will also counterbalance stereotypes of 'hacking' as a practice undertaken by a certain section of society.

CONCLUSION

This article has explored how the concept of 'hacking' provided the inspiration for a community-based digital artwork that facilitated social engagement, connections, and shared identity in a disadvantaged neighbourhood in an urban setting. This example highlights that digital media culture have utility for addressing everyday life situations, and the capacity of individuals and communities to develop or organize innovative social solutions, whether transgressive or conformist to established protocols (norms, laws) for the improvement of their lives and neighbourhoods.

The chapter also considered the moral dimension to evaluating 'hacking' and the distinction between 'hacking' and 'cracking'. Nonetheless, it has been explored that, whether or not social processes are 'hacking' or 'cracking', they have the characteristics of a community (virtual or actual). This is because they are a manifestation of empowerment, resilience and resourcefulness by individuals to access or generate benefit beyond that which the state is currently willing or able to provide. The chapter also provides suggestions for a research agenda that could generate future design interventions to develop an understanding of how digital media culture could inform forms of contemporary governance such as the 'Big society'.

REFERENCES

- Ackoff, R. L. (1974). *Redesigning the Future*. New York: Wiley.
- Andreyev, J. (2010). *HIVE*. Retrieved from <http://julieandreyev.wordpress.com/>
- Bowie, F., & Fels, S. (2009). *Flow*. Retrieved from <http://fionabowie.org/>
- Chadwick, A. (2009). Web2.0: New Challenges for the Study of E-Democracy in an Era of Information Exuberance. *I.S: A Journal of Law and Policy for the Information Society*, 10. Retrieved from <http://static1.1.sqspcdn.com/static/f/127762/17971662/1335910108743/>
- Crabtree, J. (2003). Civic hacking: a new agenda for e-democracy. *OpenDemocracy*. Retrieved from <http://www.mafhoum.com/press4/136C35.htm>
- Cramer, F. (2003). *Social Hacking, Revisited*. Retrieved from http://cramer.pleintekst.nl:70/all/social_hacking_revisited_sollfrank/social_hacking_revisited_sollfrank.pdf
- Dan, A. (2011). *Social Hacking*. Retrieved from <http://danariely.com/2011/01/20/social-hacking/>
- Devitt, M. (2001). A brief history of Computer Hacking. *Dynamic Chiropractic*, 17. Retrieved from <http://www.dynamicchiropractic.com/mpacms/dc/article.php?id=18078>
- Devitt, M. (2001). *A Brief History of Computer Hacking*. Retrieved from <http://www.chiroweb.com/archives/19/13/04.html>
- Earley, S. (2007). *Folksonomy Versus Taxonomy*. Retrieved from <http://sethearley.wordpress.com/2007/02/15/folksonomy-versus-taxonomy/>
- Fitch, C. (2003). *Crime and Punishment: The Psychology of Hacking in the New Millennium*. Retrieved from <http://www.giac.org/paper/gsec/3560>

Hudson, A., & Ciesla, M. (2011). *Do you own the gadgets you buy?*. Retrieved from http://news.bbc.co.uk/1/hi/programmes/click_online/9406690.stm

Kulikauskas, A. (2004). *Social Hacking: The Need for an Ethics*. Retrieved from http://www.freeebay.net/site/index.php?option=com_content&task=view&id=289&Itemid=99999999

Lapsley, P. (2011). *Phone Phreakers*. Retrieved from <http://voicegal.wordpress.com/2011/04/10/phone-phreakers/>

Lemos, R. (2002). New Laws Making Hacking a Black and White Choice. *CNET News*. Retrieved from http://news.com.com/2009-1001_3-958129.html

Lesley, C. (2010). *Localism & the Coalition – a neglected aspiration*. Retrieved from <http://idea.gov.uk/Idk/core/page.Id=24346328>

Leveson, B. (2012). *An Inquiry Into the Culture, Practices and Ethics of the Press* (vol. 1). Retrieved from <http://www.gov.uk/government/publications/leveson-inquiry-report-into-the-culture-practices-and-ethics-of-the-press>

Levey, S. (2002). *Hackers: Heroes of the Computer Revolution*. Penguin.

Margetts, H. (2011). *The Internet and Political Science: Re-examining Collective Action, Governance and Citizen-Government Interactions in the Digital Era*. Retrieved from <http://www.oii.ox.ac.uk/research/projects/?id=71>

Mayo, E., & Steinberg, T. (2007). *The Power of Information: An Independent Review*. Retrieved from <http://opsi.gov.uk/advice/poi/power-of-information-reviewpdf>

Moore, R. (2006). *Cybercrime: Investigating High-Technology Computer Crime* (1st ed.). Cincinnati, OH: Anderson Publishing.

Moulder, V., Boschman, L. R., & Wakkary, R. (2011). *The talking Poles – Public Art based in Social Design*. CHI 2011, Vancouver, Canada.

Raymond, E. S. (2001). *How To Become A Hacker*. Retrieved from <http://www.catb.org/~esr/faqs/hacker-howto.html>

Speed, C., Khan, A., & Phillips, M. (2015). Contemporary Governance Discourse and Digital Media: Convergences, Prospects & Problems for the 'Big Society' Agenda. In D. O'Brien & P. Matthews (Eds.), *After urban regeneration: Communities, policy and place* (p. 147). Bristol: Policy Press.

Szieter, S., & Ishkanian, A. (2012). Introduction: What is Big Society? Contemporary social policy in a historical and comparative perspective. In S. Szieter & A. Ishkanian (Eds.), *The Big Society Debate* (pp. 1–26). London: Edward Elgar Publishing Ltd. doi:10.4337/9781781002087.00007

Tacchi, J., Slater, D., & Hearn, G. (2003). *Ethnographic action research: A user's handbook developed to innovate and research ICT applications for poverty eradication*. New Delhi: UNESCO.

Thomas, D. (2002). *Hacker Culture*. Minneapolis, MN: University of Minnesota Press.

Toffler, A. (1980). *The Third Wave*. New York: Morrow.

Tonnies, F. (1957). *Community and Society* (C. P. Looms, Trans.). East Lansing, MI: Michigan State University Press.

Wal, V. T. (2004). 'Folksonomy': Coinage and Definition. Retrieved from <http://www.vanderwal.net/folksonomy.html>

ADDITIONAL READING

Albrow, M. (2012). 'Big Society' as a rhetorical intervention. *The Big Society debate: A new agenda for social welfare*, 105-115.

Borko, F. (2011). *Handbook of Augmented Reality*. Springer.

Bulley, D., & Sokhi-Bulley, B. (2014). Big society as big government: Cameron’s governmentality agenda. *British Journal of Politics and International Relations*, 16(3), 452–470.

Buzzetto-More, N. A. (2013). Social media and prosumerism. *Issues in Informing Science and Information Technology*, 10, 67–80.

Chandler, A. (1996). The changing definition and image of hackers in popular discourse. *International Journal of the Sociology of Law*, 24(2), 229–251. doi:10.1006/ijsl.1996.0015

Gitrell, J. F. (2002). Relationships between Service Providers and Their Impact on Customers Journal of Service Research May 2002, vol. 4, no. 4, 299–311 doi:10.1177/1094670502004004007

Jørgensen, K., Sandqvist, U., & Sotamaa, O. (2015). From hobbyists to entrepreneurs On the formation of the Nordic game industry. *Convergence (London)*.

Limón, J. E., & Young, M. J. (1986). Frontiers, settlements, and development in folklore studies, 1972–1985. *Annual Review of Anthropology*, 15(1), 437–460. doi:10.1146/annurev.an.15.100186.002253

Lin, Y. (2007). Hacker culture and the FLOSS innovation. Handbook of research on open source software: technological, economic, and social perspectives, 34–46.

MacKinnon, D., & Derickson, K. D. (2013). From resilience to resourcefulness A critique of resilience policy and activism. *Progress in Human Geography*, 37(2), 253–270. doi:10.1177/0309132512454775

O’Brien, D., & Matthews, P. (Eds.). (2015). *After Urban Regeneration: Communities, Policy and Place*. Policy Press. doi:10.1332/policy-press/9781447324157.001.0001

Paltrinieri, R., & Esposti, P. D. (2013). Processes of Inclusion and Exclusion in the Sphere of Prosumerism. *Future Internet*, 5(1), 21–33. doi:10.3390/fi5010021

Spafford, E. H. (1990). Is a computer break-in ever ethical? *Information Technology Quarterly*, IX(2), 9–14.

KEY TERMS AND DEFINITIONS

Folklore: Generally refers to forms of cultural expressions, which people use to convey and share their history and identity. It incorporates narratives, jokes, beliefs, proverbs, legends, myths, music, songs, dances, costumes, food, and festivals.

Hacking: Describes the practices and actions by which individuals and communities’ exploit (or explore) weaknesses of a system and product, which could be technological or social, to get by. It encapsulates both transgressive or ‘illegal’ and resourceful or ‘legal’ practices that individuals undertake to circumvent or respond to technological and social challenge or circumstance.

Hobby-ism: This refers to a form of ‘hacking’ process and practice by which an individual customize or adapt a product to one’s preferences or tastes to get a desired utility value. People who indulge in this practice are called ‘hobbyists’, and by extension are those customers of products who rather than just accept the manufacturer’s design of a product, will try to adapt it to their specification.

Prosumerism: This is a form of ‘hacking’ process by which the role of producer or professional or manufacturer and consumer merges to participate in the design and production requirements of products through mass customisation. The end product is a situation where manufacturers adapt or create their products to the specific requirements of consumers to satisfy the customers and generate profit.

QR Code: An abbreviation of Quick Response Code. It is a barcode with a machine-readable optical label of different black squares arranged in a large square or rectangular grid. When scanned, the bar codes respond quickly to allow access to its embedded information about the item to which it is attached.

Service Providers: A term used to refer to a whole range of agencies, organisations and institutions that provider services to the public (individuals and communities).

P

Political Context Elements in Public Policy of Radio Frequency Information Technology and Electromagnetic Fields

Joshua M. Steinfeld

Old Dominion University, USA

INTRODUCTION

A pinnacle in e-governance was reached with the development of information systems that utilize radio frequency technologies. For example, Radio Frequency (RF) towers' capabilities as communication and monitoring devices enable efficiency maximization and real-time solutions. Also, Advanced Imaging Technology (AIT) allows for quick and reliable information processing for purposes of tracking and surveillance. There are several advantages to government's use of Radio Frequency Information Technology (RADFIT) such as the ability to quickly communicate across a wide range of global positioning systems, management of communication portals, and survey of visitors entering secure environments in the case of millimeter wave scanning.

The main issue with the presence of radio frequency electromagnetic fields (RFEMF) is balancing the benefits provided from supporting the RADFIT systems with the environmental effects of electromagnetism. Regulation of technologies is controversial as agencies and stakeholders struggle to weigh benefits and costs. Hood et al. (2001) presents a framework for understanding regulatory policy domains by classifying benefits and costs of Information Technology (IT) sciences according to competing political systems: interest group, entrepreneurial, client, and majoritarian. The interaction of these political context elements influences the corresponding regulatory regimes, which are ideographs that describe the track of reasoning of RADFIT systems consumers, producers, and regulatory bodies. By examining

regulatory regimes of the RADFIT sphere, public policy implications and future research directions emerge that may improve participatory confidence and informational effectiveness while mitigating threats to communities.

The main purpose of the manuscript is to discuss political context elements and their impact on the public policy arena surrounding RFEMF issues, in addition to touch upon organizational initiatives and public management alternatives. Political context elements of regulatory regimes are presented first. According to the *interest group* political system, incrementalism and the status quo are introduced as encumbrances to policy change. Lack of organization in the public policy arena limits viable alternatives and contributes to government lethargy. The *entrepreneurial system*, indicative of rational choice and new public management, is subsequently discussed as the prompt and elicitor of RADFIT solutions. Modernism and progression serve as societal themes that steer entrepreneurialism in IT and public sector activities in general. Next, the *client system*, which involves administrative responsibility, is highlighted as the regime offering the most potential for bureaucratic discretion and inquiry. The opportunities for interaction between regulatory agencies and resident stakeholders, creates inconsistencies and marginalization of particular societal participants. Then, the *majoritarian system*, serving as the basis for democratic forms of governance, is detailed so as to review the unresolved paradoxes involved in representative decision rules such as voting. Subsequently, organizational and community leadership initiatives, despite the obstacles posed

DOI: 10.4018/978-1-5225-2255-3.ch581

by political context elements, are illustrated to show the current state of organized opposition to RADFIT proposals. Recommendations and areas for further research follow in an attempt to consolidate diffuse community efforts.

There are several objectives of this entry. The controversy over RADFIT solutions is examined to explain how political context elements dictate regulatory regimes of the RFEMF sphere. This entry aims to provide an overview of the bureaucratic considerations underlying RADFIT guidelines and public policy as well as the response by communities. Public administration theory and recommendations for future action serve to provide frameworks for additional policy analysis. The goal is not to provide a comprehensive review of the RFEMF regulatory arena but instead to illuminate indicators that create the onset and resonance of various regulatory regimes, or ideographs, that dictate the decision-making of public policy participants of RADFIT applications including RFEMF, along with implications for communities.

BACKGROUND

There are millions of RF tower base sites in the world. The United States alone has more than 301,779 radio frequency tower and transmission base sites (Cellular Telecommunications Industry Association, 2013). In many cases the towers stand from 50 feet to 200 feet tall. RF base sites also exist in the form of small individual devices less than 2 feet by 1 foot in size that may be mounted on building roofs or siding. Broadly, the United States Department of Commerce identifies several objectives for the use of the radio spectrum and RFEMF to carry out national policies and achieve national goals such as national security, safeguarding of life and property, support crime prevention and law, foster conservation of natural resources, provide for dissemination of information and entertainment, promote research and exploration, stimulate social and economic

progress, and generally improve the well-being of man (National Telecommunications and Information Administration, 2014: Chapter 2). More specifically, several areas of strategic interest are identified according to these objectives including agriculture, consumer expenditures and saving, education, health, oceanography, public safety, outer space, social welfare, transportation, and urbanization (National Telecommunications and Information Administration, 2014: Ch. 2).

At the local level, by 2011, automatic license-plate readers were utilized by about three-quarters of police departments surveyed (American Civil Liberties Union, 2013), with civilian uses of license-plate readers emerging (Hardy, 2014). Commercially, retailers employ companies that track shoppers through cell phones in order to identify return shoppers and learn about other shopping habits or patterns while in the store for the purposes of improving store layouts, marketing, and overall profitability (Clifford, 2013a; Clifford, 2013b). In turn, government enjoys larger tax receipts from the increases in purchases that result.

The RADFIT data collected is transmitted, stored, processed, and analyzed at data centers. These large warehouse-type buildings house servers and hard-disk drives that sit on concrete slabs as large as several football fields and are typically located in rural areas where power is less expensive (McLellan, 2013). There are over 6000 data centers in the world with more than half in the U.S.; each data center consumes the amount of power equivalent to a city with 20,000-40,000 residents (President's Council of Advisors on Science and Technology, 2014, p. 30).

The environmental effects of RF tower bases have been of concern to community residents and researchers alike. For example, RF tower installation at public schools has been strongly opposed by various community voluntary initiatives and political activist groups (Steinfeld, 2013). Similar to the RF towers that utilize RFEMF technology, AIT scanners are commonly used to scan subjects entering secure areas. In May 2013, the Transpor-

tation Security Administration (TSA) disposed of approximately 800 Rapiscan full body scanners, called “puffer” machines, due to concerns regarding irradiation of subjects (Jansen, 2013). Millimeter wave imaging systems replaced the puffer machines and function by emitting radio frequency electromagnetic waves (RFEMW) in the millimeter spectrum (30-300 gigahertz) to render images that look like photographic negatives (Elias, 2012). As of September 2012, TSA deployed about 700 AIT scanners, or Whole Body Imagers (WBI), with plans to have a total of 1800 in use by 2014 (Elias, 2012).

One major issue with the systemic use of RADFIT and the immersive presence of EMF in society has to do with health concerns that have been cited in scientific studies, concluding that RFEMF can be dangerous to human health. Golbach et al. (2015) conduct a study of immune cells, in which it was concluded that exposure to LF EMF enhanced neutrophil extracellular formation, which leads to increased antimicrobial properties and damage to surrounding cells, essentially damaging the immune system. Additionally, Pall (2015) reviews 18 epidemiological studies in concluding that exposure to microwave and radio frequency EMF from cell/mobile phone bases trigger widespread neuropsychiatric effects such as depression, sleep disturbance, headache, fatigue, concentration dysfunction, memory changes, irritability, restlessness, and loss of body weight. Manzella et al. (2015) also determine that EMF affects the human biological clock. Pall (2015) reviews 6 studies that conclude radio and television antenna may have the same effects.

Differently, Shi et al. (2014) use in-vitro isolation techniques to test the effects of intermediate wireless magnetic fields and report no genotoxic effects from exposure. Furthermore, Shirai et al. (2014) recorded no significant negative impacts on the brain function of three generations of rats that were exposed to cell phone radiation for 20 hours per day. Yet, Du et al. (2014) report that

obese mice lose weight and fat when treated with direct current electromagnetic fields. Laramee et al. (2014) detected an elevation of heat shock gene expression as a result of static magnetic field exposure, indicating that RFEMF heat-up or “cook” its subjects. Lerchl et al. (2015) conclude that the number of tumors in the lungs and livers of animals were higher, in addition to elevated numbers of lymphomas, than in the sham-exposed group. The tumor producing effects are attributed to metabolic changes due to EMF exposure (Lerchl et al., 2015).

Without the hindsight of the recent scholarly literature and experimental studies on RFEMF, on January 3, 1996, the Federal Communications Commission (FCC) passed the Telecommunications Act “to promote competition and reduce regulation in order to secure lower prices and higher quality services for American consumers and encourage the rapid deployment of new telecommunications technologies” (Telecommunications Act, 1996). According to Section 332(c)(7)(B)(iv), “No State or local government or instrumentality thereof may regulate the placement, construction, and modification of personal wireless service facilities on the basis of the environmental effects of radio frequency emissions to the extent that such facilities comply with the Commission’s regulations concerning such emissions.” In conjunction with this limitation, Section 214(e)(1)(a) grants telecommunications carriers eligibility to provide their own telecommunications facilities in areas designated as “universal service,” even if a competitor is already providing wireless service coverage for the area (O’Neill, 1999).

Section 253(a) reinforces carriers’ rights in citing “No state or local statute or regulation, or other state or local legal requirement, may prohibit or have the effect of prohibiting the ability of any entity to provide any interstate or intrastate telecommunications service.” Incidentally, the Telecommunications Act may be in violation of fifth and tenth amendment Constitutional protections.

The fifth amendment, states that persons cannot “be deprived of life, liberty, or property, without due process of law,” and the tenth amendment declares that “The powers not delegated to the United States by the Constitution, nor prohibited by it to the states, are reserved to the states respectively, or to the people” (U.S. Constitution, 1789).

For AIT, Sections 101 and 109 in the Aviation and Transportation Security Act of 2001 provide for federal government’s authority to utilize RADFIT. Section 101(f)(8) makes it lawful for TSA to “identify and undertake research and development activities necessary to enhance transportation security” and Section 109(a)(5) provides for the lawful use of wireless technologies to aid in passenger screening (Aviation and Transportation Security Act, 2001). However, privacy issues related to AIT surfaced and thus Automated Target Recognition (ATR) technology was developed as an alternative means of security scanning (Elias, 2012). Essentially, WBI were transmitting and recording images of users, creating privacy issues since RFEMF pass right through clothing. Privacy-enhancing morphological edge and gradient detection software algorithmic technology was subsequently developed for millimeter wave scanners to combat the issue of privacy through distortion of facial and other physical features (Cavoukian, 2009), effectively silencing AIT opposition founded upon the privacy issue.

The lack of concrete experimental data on health effects of RADFIT and shortage of innovative remedies to combat health concerns has brought into question the impact of invisible risks posed by RFEMF. Meanwhile, the Telecommunications Act has authority over local and state powers but may in fact be unconstitutional. Consequently, public policies governing RADFIT are being inconsistently applied and the regulatory environment remains ambivalent. With the failure of public management to achieve universal clarity and resident support on issues such as rights-of-way, regulatory stances continue to be conceded by political elements.

POLITICAL CONTEXT ELEMENTS AND VOLUNTARY INITIATIVES

P

Interest Group Politics

The confrontation between advocacy coalitions representing competing interest groups leads to incrementalism and the status quo. As a result, it becomes unlikely that current trends in RF tower base proliferation or use of RADFIT solutions for communicative and security applications will be curtailed. There is a status quo and changing it requires special justification (Baumgartner et al., 2009). The main issues such as whether or not to allow RF towers to be erected near schools and residential neighborhoods, or if AIT systems should be required for entry into secure areas have been at deadlock since the major benefits of RADFIT were discovered and capabilities implemented. Additionally, the interest group narratives have largely remained unchanged. For example, the American Cancer Association (ACA), a federal funding beneficiary, continues to maintain that scientific evidence has not conclusively indicated that electromagnetic fields created by RF towers and other base sites are hazardous to human health due to the longer wavelength of RF and its lower energy level as compared to gamma rays or x-rays (American Cancer Society, 2013). Meanwhile, opponents to RFEMF have advocated a not-in-my-backyard approach to RF towers and employed bounded rationality when dealing with RADFIT security apparatuses.

Bounded rationality emphasizes the cognitive limitations of decision-makers and is the arbitrary underpinning of the advocacy coalition approach (True et al., 2007). Health hazards from RFEMF absorption, an invisible risk, remain disputed. From consumers and producers’ perspectives, the idea of losing mobile connectivity from absence of RF towers or being susceptible to security breaches and inefficiencies by avoiding utilization of AIT systems is tangible. Jones (1994) argued that decision changes do not arise from changes

in preferences or irrationality; they emerge from shifts in attention called “serial shifts.” The senses process information equally but attention is serially committed to one idea at a time (Simon, 1983). The dominant, or tangible concerns, as opposed to the invisible health risk, occupy the psyche. Typically, only one primary aspect of the choice situation serves as the critical decision factor (Simon, 1957, 1985; Zaller, 1992). As a result, both individual and tactical initiatives seeking to oppose RADFIT applications are lulled by even the supporters who share the utmost concern and consciousness over the issue. The status quo is maintained and incrementalism persists through steady growth of RADFIT applications as new technologies that create efficiencies.

Entrepreneurial Politics

New public management and undertones of progressivism continue to guide bureaucratic conduct in lieu of rational choice logic. While rational choice theory has been founded upon the notion that market mechanisms should be used to settle collective choice problems (Hill, 2009), as in the case of new public management (Hood, 1995), a core competency of any sound public policy involves the conception that public value is being created for society (Moore, 1995). In the case of RADFIT applications, the benefits of utilization are real and the potential costs associated with health risks are hypothetical. Meanwhile, remnants of Wilson’s (1887) progressive movement continue to trigger the cogitation of clever solutions to microcosmic issues that could have been left unresolved, such as in the case of AIT as a solution to laborious manual screening.

Rational choice, in the form of economic individualism, has become a decision rule under the new public management accordance in bureaucracy, enticing public leaders and the public at large to look to the private sector for solutions and then to the government and nonmarket institutions upon market failure (Bozeman, 2007). For example, public agencies purchase AIT ap-

paratuses in place of hiring trained workers and corporations are placated by political authorities with permissive jurisdiction that enable RF tower producers to enjoy productive economies of scale with RF tower construction and maintenance. Without major indication of widespread epidemic from RFEMF exposure and documented incurrence of costs, the entrepreneurial policy domain continues to be dictated by technocratic principles of automating bureaucratic processes.

Consumers of RADFIT applications are similarly limited in scope by rational choice decision-making as it relates to social exchange and public choice theory. Under the auspices of social exchange theory, people choose courses of action that yield the least amount of blame and will behave differently based on the number of people present (Latane & Darley, 1970; Heath, 1976). Users of RFEMF are unlikely to dissent if numerous bystanders are present in the case of AIT screening or a large number of people stand to lose through loss of wireless services. Instead, users exercise public choice and abide by laws of collective action. Policy change advocates have a common interest in benefits but are unwilling to incur the costs of providing the collective good (Olson, 1965). Miller (1994) attempts to cast doubt on social exchange and public choice theory through witness of periodic altruistic behavior. Considerately, policy entrepreneurs tend to thrive from outside their formal positions of government or participative roles (Roberts & King, 1991); there is no opportunity, or policy stream (Kingdon, 2008), for advocates of regulatory change to act extemporaneously or institutionalize concerted efforts, respectively.

Client Politics

While dangers to human health and wellbeing may be natural and inevitable, the exacerbation or mitigation of risks, and how costs are disbursed, are not (Minow, 2012). Public administrators are oftentimes in tension with the citizenship role, creating potential for recurring conflict between

the two identities (Cooper, 2012). The extent to which public administrators are responsive to challenges or remain responsible to the bureaucracy is determined by client politics. Bureaucrats with knowledge-in-the-process (Lasswell, 1970), those administrators involved in RADFIT solicitation and implementation, are bound by characteristics of bureaucracy. For example, government or contracted operators and repair personnel of potentially hazardous RF hardware fail to express health and wellbeing concerns because current tasks are thought to be temporary given hopes of ascending in rank or being delegated task variation, even though the position continues to be in existence and held by another incumbent nonetheless (Weber, 1946). In this case, the public administrator or contractor is a client of the bureaucracy, furloughed by self-interest (see Downs, 1967; Dunleavy, 1991).

Oppositely, Weber (1947) and Fayol (1949) posit bureaucracy as client-oriented and administrators as bureaucratic servants that subordinate individual interest to general interest and create a division of labor with clearly defined responsibilities. According to this rigid structure, RF tower and AIT technicians and operators employed by the government are unlikely to ever converge with clients that are external to the bureaucracy, such as concerned subjects of RADFIT applications. In accordance with Weberian traditionalism, public administrators cling to old ways despite individuals' tendency to be progressive in other spheres of life (Manheim, 1936). Public administrators and technical experts become mesmerized with current innovative capacities instead of embracing advancement in thought, thereby neglecting ideological evolution (Burke, 1790). Soon thereafter, the bureaucrat is unable to retrospectively imagine society without a propagation of RF towers or reliance on AIT machines.

Majoritarian Politics

The calculus supporting denunciation of majoritarian politics has been aptly applied over the course of policy domains (Arrow, 1951; Buchanan &

Tullock, 1962; Tullock & Wagner, 1978). From a regulatory standpoint, the economic-type limits to control, *de minimis non curat fiscus*, parallel downfalls of majoritarian decision rules (Hood, 1976). For instance, bureaucrats may simply lack the authority to impact implementation and utilization aspects of RADFIT components and systems, a control that could rest with a few policy makers and loitering advocacy coalitions. In other cases, colossal consumer markets welcome RADFIT solutions that promise to create even more wealth for government and societal members. The economic and ideological grandeur of RFEMF applications circumscribes the eclipsing echoes citing dangers to RF absorption.

Other modes of majoritarian politics take shape through consumer federalism (Wildavsky, 1998). When one AIT model is found to be ineffective or believed to be exceptionally hazardous to users' health, as in the case of "puffer" security scanning machines, then it is possible for consumer federalists to play off one producer against another in order to pursue agency procurement ambitions. Governmental consumers and RFEMF regulatory bodies achieve cohesion as a spectrum emanates for the devices market through not only determining the lawfulness of the technology but also devising a range of acceptable exposure restrictions. Davis, Dempster, & Wildavsky (1966) identify these decision rules as the deviant case and shift point. The deviant case involves the underlying controversy over RADFIT solutions whereas the shift point rests along the spectrum of technological capacity. These decision rules are non-incremental and therefore typically exist in a state of stasis, left to coagulate solely according to further advancements.

Voluntary Initiatives

Despite the unfavorable political context elements that present obstacles to effective regulation of RADFIT, numerous initiatives have surfaced that contribute to regulation of RADFIT and EMF. The European Union, on November 14, 2011, was the first organization to ban use of "backscatter" ion-

ized radiation technology in favor of millimeter wave scanners (DiSalvo, 2011; Jaslow, 2011). Additionally, the International Association of Fire Fighters officially opposes the idea of mounting RFEMF applications on firehouses until further research (International Association of Firefighters, 2014). In Australia, community groups have succeeded in organizing at the national level through a group called No Towers Near Schools that provides support and information on current disputes around the country (No Towers Near Schools, 2014). Elsewhere, community leadership efforts have spanned international borders. On December 1, 2013, more than 100 people in Tsawwassen, Canada and 200 people in Point Roberts, U.S. protested near the border, on 1st avenue and 56th street, vowing to stop the installation of five towers that would affect residents on both sides (Hager, 2013).

In the U.S., there are dozens of communities in California currently organizing to reject RF towers such as in San Diego (see Verdin, 2014), Los Angeles (see City of Huntington Beach, 2013; West San Pedro Neighborhood Alliance, 2014; Northwest San Pedro Neighborhood Council, 2009), Carmel (see Arcega-Dunn, 2014), Hillsboro (see Lee, 2013), and San Rafael (see Upshaw, 2009). The presence of EMF base sites has also garnered substantial attention in New York, as an example, where the Department of State released the publication entitled *Planning and Design Manual for the Review of Applications for Wireless Telecommunications Facilities*. Governor Andrew Cuomo believes that municipal regulation of cellular phone towers and antennas is one of the most controversial and heavily debated current local government land use issues (Legal Memorandum LU01, 2014). Similarly, the North Carolina League of Municipalities released guidance on the current tensions between wireless service providers' desire for an efficient, streamlined, predictable, and expedited regulatory process as opposed to citizens' desire for transparency and interest in maintaining rights inherent to democratic opposition (Hibbard, 2013).

SOLUTIONS AND RECOMMENDATIONS

New approaches to policy change are required if regulatory regimes in the RFEMF sphere are going to be expanded or added to. Currently, there is rapid promulgation and application of RADFIT solutions given the apparent stalemate within political context elements. Unchallenged RFEMF base sites vastly outnumber community leadership initiatives. Baumgartner & Jones (1993) found that policymaking leaps and periods of stasis share a tendency toward punctuated equilibrium, or unforeseen impulse; an unfavorable regulatory outcome whether it results in health epidemic, extreme loss of connectivity, or security breach.

Solutions to static regulatory environments can be found by emulating the postmodern movement. Stivers (1994) emphasized bureaucratic listening as a way to compensate for spatial differences between stakeholders in regulatory settings. Regulatory agencies have neglected the wide body of research conducted that implicates RFEMF technology as hazardous. However, it may not be critical for the policy change initiatives to rest with empirical conclusions where validity and reliability remain controversial. Miller (2013) builds on postmodern approaches to public policy and administration through examination of ideographs. The narratives that produce symbolic images of policy domains have the potential to bypass the empirical struggle plaguing the certainty of RFEMF research. Scientific reference may be replaced by ideographs that portray RF towers as nuisances, eyesores, or institutional threats to privacy. The very notion of empirical uncertainty may emerge as a lasting narrative that paints ideographic imagery similar to that of the questioned invisible risk such as viable slogans that could include "you can't regulate what you can't see" and "go easy with radiofrequency."

According to new public management, a counterpart to postmodernism, narratives of efficiency, cost-effectiveness, capacity building, and innovation serve as underpinnings for policy implementa-

tion, yet creativity remains through mobilization such as outsourcing. The Telecommunications Act of 1996 aimed to expedite telephone carriers' efforts and conserve their vital resources while striving to provide resident, business, and government consumers with affordable and "universal service." In *Reinventing Government* (1998), Osborne and Gaebler argue that government "steers rather than rows," which is exactly what the federal government did by contracting out telecommunications infrastructure, services, and maintenance to private companies while at the same time providing authoritative means with regards to competition, site placement, and procurement (Telecommunications Act, 1996). Instead, government should "serve rather than

steer," whereby public servants help citizens articulate and meet their shared interests rather than steering society to new directions (Denhardt & Denhardt, 2000).

Local and state governments may help communities to consolidate their efforts so that advocacy can be directed to federal government action. An amendment to the Telecommunications Act is required to consider health effects and unsightliness of RADFIT. Under the revision, Title 47 of the Code of Federal Regulations needs to be specifically added to Section 332(c)(7)(B)(iv) of the Act (Table 1). Federal limits for Maximum Permissible Exposure (MPE) to radiofrequency radiation clearly state that exposure by the general public is not to exceed 30 minutes.

Table 1. Title 47: U.S. code of federal rules

Code of Federal Rules				
Title 47: Telecommunication				
Volume: 1				
Date: 2011-10-01				
Title: Section 1.1310 - Radiofrequency radiation exposure limits.				
Context: Title 47-Telecommunication. Chapter 1-Federal Communications Commission.				
Subchapter A- Part 1. Practice And Procedure.				
Subpart I-Procedures Implementing the National Environmental Policy Act of 1969.				
Frequency (MHz)	Electric Field (V/m)	Magnetic Field (A/m)	Power (mW/cm2)	Time (mins)
Limits for Occupational/Controlled Exposures				
0.3-3.0	614	1.63	*100	6
3.0-30	1842/f	4.89/f	*900/f ²	6
30-300	61.4	0.163	1	6
300-1500			f/300	6
1500-100,000			5	6
Limits for General Population/Uncontrolled Exposure				
0.3-1.34	614	1.63	*100	30
1.34-30	824/f	2.19/f	*180/f ²	30
30-300	27.5	0.073	0.2	30
300-1500			f/1500	30
1500-100,000			1	30

Occupational/controlled limits apply in situations that persons are exposed from employment.

General population/uncontrolled exposures apply in situations that the general public is exposed.

f = frequency in MHz

* = Plane-wave equivalent power density

Therefore, RF towers that are mounted on or adjacent to schools, buildings, parks, neighborhoods, playing fields, or other sites may technically be rejected due to excess general population/uncontrolled exposure beyond 30 minutes. However, a strong lobby exists that demands connectivity, and connectivity cannot occur without exposure; it is the presence of RFEMF that enables RADFIT. It is unlikely that community voluntary initiatives will have the impact of freezing RADFIT expansion or destructing current base sites. However, base sites that are opposed by communities should be rejected or removed if already in place and a federal mandate is necessary to achieve equality on this matter across all stakeholder communities.

FUTURE RESEARCH DIRECTIONS

Hood and Margetts (2007) see the future as an extrapolation of the past; the tools of the digital age may create heightened levels of nodality such as cyber detection and group targeting. While rapid installation of RF towers and continued usage of AIT devices will increase the number of systems users and the overall outreach of RADFIT application, it is unlikely that growth alone in the number of RF towers or AIT devices would increase nodality; advancements in RFEMF capabilities would likely be required. Furthermore, RADFIT enhancements may be met with incremental policy challenges as regulatory regimes gradually shift according to new research developments. Unfortunately, consumers and producers are forced to continually exercise a responsive-type of policy influence, as new technologies can typically be more quickly implemented than empirically deemed to be unsafe. Meanwhile, there is immense competition between researchers seeking to conduct progressive research on RFEMF utility and those experimenters aiming to uncover any negative externalities associated with RADFIT systems.

RFEMF research has been conducted on rats and human organelles. Research on rats is largely

conclusive in determining toxic effects of exposure. Yet, rats are much smaller mammals than humans and are not viewed to be similar by the general public. On the other hand, research on humans that could involve immense danger to human subjects is not feasible and there would be significant challenges to securing a control group in any case, given the numerous factors that can affect human health over the course of a study, especially if long-term effects of RF absorption are the foci of study. Nonetheless, the intrusive placement of many existent RF towers provides an opportunity for surveying human subjects, especially if the RF towers are not going to be removed.

Furthermore, expansive environmental impact studies need to be conducted to ascertain the broad impact of RADFIT. On September 14, 2000, the U.S. Department of Interior, on behalf of the Fish and Wildlife Service, issued a cautionary statement of RF towers' impact on migratory birds. A Communication Tower Working Group consisting of government agencies, industry, academic researchers, and non-governmental organizations was formed by the Service to reduce the estimated 4 to 5 million bird deaths per year that occur from colliding with communication towers (Clark, 2000). It may be possible that RFEMF has physiological effects on birds as well. In either case, birds' migration is being obstructed which could influence other animals. Impact studies of communication tower base sites and isolated RFEMF on living organisms within complex ecosystems would advance the field's body of knowledge.

For now, the precautionary principle is suggested for dealing with the regulatory sphere of RADFIT and EMF. Previously, the precautionary principle has been advocated for application to regulatory decision-making regarding sustainable development including the Rio Convention and Kyoto Protocol (Steele, 2006), and it is now being recommended when it comes to government's regulation for the use and power of electromagnetic fields (Portier, 2015). The precautionary principle involves employing alternatives in situations when

irreparable health or environmental risk may result from a particular action while the research is inconclusive or not yet realized in terms of the risks. Considering the set of feasible alternatives determines whether the actions of independent agents need to be restricted given certain levels of risk to the environment and human health (Tickner, 2003).

CONCLUSION

Expansion and application of RADFIT solutions continues to progress at a rapid pace. Empirical research regarding health effects from RFEMF absorption are surrounded by controversy and lag behind innovation. Currently, the four major political context elements and respective bureaucratic limitations of the regulatory environment results in public policy stasis. Interest group politics have contributed to incrementalism and the status quo while at the same time entrepreneurial modes of political will are defeated by deficiencies inherent to economic individualism and social exchange. In addition, client politics posit the public administrator as a participant nullified by dual roles as bureaucrat and resident client, just as majoritarian approaches to decision rules are malignantly constructed along a spectrum that serves to favor strategic public procurement of RADFIT devices.

Modernism and progression have dominated the narrative platform to date. Bureaucratic listening and postmodernism in the form of ideographs serve as communicative devices that can be used by community organizations to minimize the stakeholder gap between users, consumers, and producers of RADFIT systems. Reform of the Telecommunications Act and consolidation of community leadership initiatives are concrete solutions. Given the propagation of RFEMF infrastructure, empirical studies using survey data of residents living near RF towers or frequently exposed to other AIT devices is logical and does not contribute additional threat to subjects already experiencing continual RF absorption. Irrespective

of health concerns related to RFEMF and RADFIT solutions, sociological constraints remain that are detrimental to the contributions of individual stakeholder and regulatory participants.

P

REFERENCES

- American Cancer Society. (2013). *Cellular Phone Towers*. Retrieved from <http://www.cancer.org/cancer/cancercauses/othercarcinogens/athome/cellular-phone-towers>
- American Civil Liberties Union. (2013). *You Are Being Tracked: How License Plate Readers Are Being Used To Record Americans' Movements*. Retrieved from <http://www.aclu.org/files/assets/07613-aclu-alprreport-opt-v05.pdf>
- Arcega-Dunn, M. (2014). Carmel Valley Residents Opposed to AT&T's Proposed Plan to Build Cell Tower on Public Library. *ABC 10 News*. Retrieved from <http://www.10news.com/news/carmel-valley-residents-opposed-to-atts-proposed-plan-to-build-cell-tower-on-public-library-01282014>
- Arrow, K. (1951). *Social Choice and Individual Values*. New York: John Wiley & Sons.
- Aviation and Transportation Security Act. (2001). *Public Law 107-71, 107th Congress*. Retrieved from http://www.tsa.gov/sites/default/files/assets/pdf/ssi/atsa_public_law_107_1771.pdf
- Baumgartner, F., Berry, J., Hojnacki, M., Kimball, D., & Leech, B. (2009). *Lobbying and Policy Change: Who Wins, Who Loses, and Why*. Chicago, IL: The University of Chicago Press. doi:10.7208/chicago/9780226039466.001.0001
- Baumgartner, F., & Jones, B. (1993). *Agendas and Instability in American Politics*. Chicago, IL: University of Chicago Press.
- Bozeman, B. (2007). *Public Values and Public Interest: Counterbalancing Economic Individualism*. Washington, DC: Georgetown University Press.

- Buchanan, J., & Tullock, G. (1962). *The Calculus of Consent*. Indianapolis, IN: Liberty Fund. doi:10.3998/mpub.7687
- Burke, E. (1790). *Reflections on the Revolution in France*. New York: Penguin Books.
- Capri, M., Scarcella, E., Fumelli, C., Bianchi, E., Salvioli, S., Mesirca, P., & Franceschi, C. et al. (2004). In Vitro Exposure of Human Lymphocytes to 900 MHz CW and GSM Modulated Radiofrequency: Studies of Proliferation, Apoptosis, and Mitochondrial Membrane Potential. *Radiation Research*, 162(2), 211–218. doi:10.1667/RR3209 PMID:15387149
- Cellular Telecommunications Industry Association. (2013). *Wireless Quick Facts*. Retrieved from <http://www.ctia.org/advocay/research/index.cfm/aid/10323>
- City of Huntington Beach. (2013). *City Wins Court Case Against T-Mobile – No Cell Tower at Bolsa View Park*. Retrieved from <http://www.huntingtonbeachca.gov/announcements/announcement.cfm?id=627>
- Clark, J. (2000). *Service Guidance on the Siting, Construction, Operation, and Decommissioning of Communications Towers*. United States Department of Interior, Fish and Wildlife Service. Retrieved from <http://www.fws.gov/migratorybirds/currentbirdissues/hazards/towers/comtow.html>
- Cleary, S., Du, Z., Cao, G., Liu, L., & McCrady, C. (1996). Effect of Isothermal Radiofrequency Radiation on Cytolytic T-lymphocytes. *The FASEB Journal*, 10(8), 913–919. PMID:8666169
- Clifford, S. (2013a, July 14). Attention, Shoppers: Store Is Tracking Your Cell. *The New York Times*. Retrieved from <http://www.nytimes.com/2013/07/15/business/attention-shopper-stores-are-tracking-your-cell.html>
- Clifford, S. (2013b, June 19). Using Data to Stage-Manage Paths to the Prescription Center. *The New York Times*. Retrieved from <http://bits.blogs.nytimes.com/2013/06/19/using-data-to-stage-manage-paths-to-the-prescription-counter/>
- Constitution, U. S. (1789). *Fifth and Tenth Amendments of the United States Constitution*. Retrieved from http://www.law.cornell.edu/constitution/tenth_amendment
- Cooper, T. (2012). *The Responsible Administrator: An Approach to Ethics for the Administrative Role* (6th ed.). San Francisco, CA: Jossey-Bass.
- Covoukian, A. (2009). *Whole Body Imaging in Airport Scanners: Building in Privacy by Design*. Information and Privacy Commissioner of Ontario. Retrieved from <http://www.ipc.on.ca/images/resources/wholebodyimaging.pdf>
- Davis, O., Dempster, M., & Wildavsky, A. (1966). A Theory of the Budgetary Process. *The American Political Science Review*, 60(3), 529–547. doi:10.2307/1952969
- Denhardt, R., & Denhardt, J. (2000). The New Public Service: Serving Rather Than Steering. *Public Administration Review*, 60(6), 549–559. doi:10.1111/0033-3352.00117
- Disalvo, D. (2011). Europe Bans Airport Body Scanners For “Health and Safety” Concerns. *Forbes*. Retrieved from <http://www.forbes.com/sites/daviddisalvo/2011/11/15/europe-bans-airport-body-scanners-over-health-and-safety-concerns/>
- Downs, A. (1967). *Inside Bureaucracy*. Boston, MA: Little, Brown, and Company.
- Dunleavy, P. (1991). *Democracy, Bureaucracy, and Public Choice: Economic Explanations in Political Science*. London: Prentice Hall.

- Elias, B. (2012). *Airport Body Scanners: The Role of Advanced Imaging Technology in Airline Passenger Screening*. Congressional Research Service. Retrieved from <http://www.fas.org/sgp/crs/homsec/R42750.pdf>
- Fayol, H. (1949). *General and Industrial Administration*. New York, NY: Pitman.
- Golbach, L. A., Scheer, M. H., Cuppen, J. J. M., Savelkoul, H., & Verburg-van Kemenade, B. M. L. (2015). Low-Frequency Electromagnetic Field Exposure Enhances Extracellular Trap Formation by Human Neutrophils through the NADPH Pathway. *Journal of Innate Immunity*, 7(5), 459–465. doi:10.1159/000380764 PMID:25871408
- Hager, M. (2013, December). Tsawwassen and Point Roberts Residents Continue to Fight Radio Towers. *Vancouver Sun*, 1. Retrieved from <http://www.vancouversun.com/news/Tsawwassen+Point+Roberts+residents+continue+fight+radio+towers/9234438/story.html>
- Hardy, Q. (2014, April 19). How Urban Anonymity Disappears When All Data Is Tracked. *The New York Times*. Retrieved from https://bits.blogs.nytimes.com/2014/04/19/how-urban-anonymity-disappears-when-all-data-is-tracked/?_r=0
- Heath, A. (1976). *Rational Choice and Social Exchange: A Critique of Exchange Theory*. Cambridge, UK: Cambridge University Press.
- Hibbard, K. (2013). *City Authority to Regulate Wireless Telecommunications*. Retrieved from <http://www.nclm.org/SiteCollectionDocuments/Legislative/2013%20--%20Cell%20TowerWireless%20Telecommunications%20Regulatory%20Authority.pdf>
- Hill, M. (2009). *The Public Policy Process* (5th ed.). Harlow, UK: Pearson Longman.
- Hood, C. (1976). *The Limits of Administration*. London: John Wiley & Sons.
- Hood, C. (1995). Contemporary Public Management: A New Global Paradigm? *Public Policy and Administration*, 10(2), 104–117. doi:10.1177/095207679501000208
- Hood, C., & Margetts, H. (2007). *The Tools of Government in the Digital Age*. Hampshire, UK: Palgrave Macmillan. doi:10.1007/978-1-137-06154-6
- Hood, C., Rothstein, H., & Baldwin, R. (2001). *The Government of Risk: Understanding Risk Regulation Regimes*. Oxford, UK: Oxford University Press. doi:10.1093/0199243638.001.0001
- Hook, G., Zhang, P., Lagroye, I., Li, L., Higashikubo, R., Moros, E., & Roti, J. et al. (2004). Measurement of DNA Damage and Apoptosis in Molt-4 Cells after *In Vitro* Exposure to Radiofrequency Radiation. *Radiation Research*, 161(2), 193–200. doi:10.1667/RR3127 PMID:14731070
- Huang, T., Lee, M., Oh, E., Zhang, B., Seo, J., & Park, W. (2008). Molecular Responses of Jurkat T-cells to 1763 MHz Radiofrequency Radiation. *International Journal of Radiation Biology*, 84(9), 734–741. doi:10.1080/09553000802317760 PMID:18821387
- International Association of Fire Fighters. (2014). *Division of Occupational Health, Safety, and Medicine*. Retrieved from <https://www.iaff.org/hs/Facts/CellTowerFinal.asp>
- Jansen, B. (2013). TSA Dumps Near-Naked Rapiscan Body Scanners. *USA Today*. Retrieved from <http://www.usatoday.com/story/travel/flights/2013/01/18/naked-airport-scanners/1845851/>
- Jaslow, R. (2011). Europe Bans Airport Scanners Over Cancer Fears: How About U.S.? *CBS News*. Retrieved from <http://www.cbsnews.com/news/europe-bans-airport-scanners-over-cancer-fears-how-about-us/>

- Jones, B. (1994). *Reconceiving Decision-Making in Democratic Politics: Attention, Choice, and Public Policy*. Chicago, IL: University of Chicago Press.
- Kingdon, J. (2003). *Agendas, Alternatives, and Public Policies* (2nd ed.). New York, NY: Longman.
- Laramée, C., Frisch, P., McLeod, K., & Li, G. (2014). Elevation of Heat Shock Gene Expression From Static Magnetic Field Exposure in Vitro. *Bioelectromagnetics*, 35(6), 406–413. doi:10.1002/bem.21857 PMID:24839179
- Lasswell, H. (1970). The Emerging Conception of the Policy Sciences. *Policy Sciences*, 1(1), 3–14. doi:10.1007/BF00145189
- Latane, B., & Darley, J. (1970). Social Determinants of Bystander Intervention in Emergencies. In J. Macaulay & L. Berkowitz (Eds.), *Altruism and Helping Behavior* (pp. 13–27). New York, NY: Academic Press.
- Lee, V. (2013). Hillsboro Residents Oppose New Cell Towers. *ABC News*. Retrieved from <http://abclocal.go.com/kgo/story?section=news/local/peninsula&id=9029667>
- Legal Memorandum LU01. (2014). *Municipal Regulation of Cellular Telephone Towers and Antennas*. Department of State, Office of General Counsel. Retrieved from <http://www.dos.ny.gov/cnsl/lu01.htm>
- Lerchl, A., Klose, M., Grote, K., Wilhelm, A. F. X., Spathmann, O., Fiedler, T., & Clemens, M. et al. (2015). Tumor Promotion by Exposure to Radiofrequency Electromagnetic Fields Below Exposure Limits for Humans. *Biochemical and Biophysical Research Communications*, 459(4), 585–590. doi:10.1016/j.bbrc.2015.02.151 PMID:25749340
- Manheim, K. (1936). *Conservatism: A Contribution to the Sociology of Knowledge* (Collected Works Volume 11). New York, NY: Routledge.
- Manzella, N., Bracci, M., Ciarapica, V., Staffolani, S., Strafella, E., Rapisarda, V., & Santarelli, L. et al. (2015). Circadian Gene Expression and Extremely Low-Frequency Magnetic Fields: An In Vitro Study. *Journal of Bioelectromagnetics*, 36(4), 294–301. doi:10.1002/bem.21915 PMID:25808738
- McLellan, C. (2013). The 21st Century Data Center: An Overview. *ZDNet*. Retrieved from <http://www.zdnet.com/the-21stcentury-data-center-an-overview-7000012996/>
- Miller, H. (1994). Post-Progressive Public Administration: Lessons from Policy Networks. *Public Administration Review*, 54(4), 378–386. doi:10.2307/977386
- Miller, H. (2013). *Governing Narratives: Symbolic Politics and Policy Change*. Tuscaloosa, AL: The University of Alabama Press.
- Minow, M. (2012). Seeing, Bearing, and Sharing Risk. In J. Hacker & A. O’Leary (Eds.), *Shared Responsibility, Shared Risk: Government, Markets, and Social Policy in the Twenty-First Century* (pp. 253–269). Oxford, UK: Oxford University Press. doi:10.1093/acprof:oso/bl/9780199781911.003.0013
- Moore, M. (1995). *Creating Public Value: Strategic Management in Government*. Cambridge, MA: Harvard University Press.
- No Towers Near Schools. (2014). *Community Battles Over Siting of Mobile Phone Facilities Around Australia*. Retrieved from <http://www.notowersnearschools.com>

- Northwest San Pedro Neighborhood Council. (2009). *Board and Stakeholder Meeting Minutes*. Retrieved from <http://Nwsanpedro.org/wp-content/uploads/2012/10/November-9-2009-minutes.pdf>
- O'Neill, K. (2009). Wireless Facilities Are a Towering Problem: How Can Local Zoning Boards Make the Call Without Violating Section 704 of the Telecommunications Act of 1996? *William and Mary Law Review*, 40(3), 975–1018.
- Olson, M. (1965). *The Logic Of Collective Action: Public Goods and The Theory of Groups*. Cambridge, MA: Harvard University Press.
- Osborne, D., & Gaebler, T. (1992). *Reinventing Government: How The Entrepreneurial Spirit Is Transforming The Public Sector*. Middlesex, UK: Penguin Books.
- Phillips, J., Ivaschuk, O., Ishida-Jones, T., Jones, R., Campbell-Beachler, M., & Haggren, W. (1998). DNA Damage in Molt-4 T-lymphoblastoid Cells Exposed to Cellular Telephone Radiofrequency Fields in Vitro. *Bioelectrochemistry and Bioenergetics*, 45(1), 103–110. doi:10.1016/S0302-4598(98)00074-9
- Portier, C. (2015). *The Precautionary Principle Should Be Invoked for Radio Frequency Electromagnetic Fields*. Keynote Speech. Annual Meeting of the Bioelectromagnetics Society.
- President's Council of Advisors on Science and Technology. (2014). *Big Data and Privacy: A Technological Perspective. Report to the President, May 2014*. Executive Office of the President.
- Roberts, N., & King, P. (1991). Policy Entrepreneurs: Their Activity Structure and Function in the Policy Process. *Journal of Public Administration Research and Theory*, 1(2), 147–175.
- Sannino, A., Di Costanzo, G., Brescia, F., Sarti, M., Zeni, O., Juutilainen, J., & Scarfi, M. (2009). Human Fibroblasts and 900 MHz Radiofrequency Radiation: Evaluation of DNA Damage After Exposure and Co-Exposure to 3-chloro-4-(dichloromethyl)-5-hydroxy-2(5 h)-furanone (MX). *Radiation Research*, 171(6), 743–751. doi:10.1667/RR1642.1 PMID:19580481
- Shi, D., Zhu, C., Lu, R., Mao, S., & Qi, Y. (2014). Intermediate Frequency Magnetic Field Generated by a Wireless Power Transmission Device Does Cause Genotoxicity in Vitro. *Bioelectromagnetics*, 35(7), 512–518. doi:10.1002/bem.21872 PMID:25196478
- Shirai, T., Imai, N., Wang, J., Takahashi, S., Kawabe, M., Wake, K., & Fujiwara, O. et al. (2014). Multigenerational Effects of Whole Body Exposure to 2.14 GHz W-CDMA Cellular Phone Signals on Brain Function in Rats. *Bioelectromagnetics*, 35(7), 497–511. doi:10.1002/bem.21871 PMID:25196377
- Simon, H. (1957). *Models of Man*. New York: Wiley.
- Simon, H. (1983). *Reason in Human Affairs*. Stanford, CA: Stanford University Press.
- Simon, H. (1985). Human Nature in Politics: The Dialogue of Psychology with Political Science. *The American Political Science Review*, 79(02), 293–304. doi:10.2307/1956650
- Steele, K. (2006). The Precautionary Principle: A New Approach to Public Decision-Making? *Law Probability and Risk*, 5(1), 19–31. doi:10.1093/lpr/mgl010
- Steinfeld, J. (2013). Radio Frequency Towers: Public School Placement. In S. Jorgensen (Ed.), *Encyclopedia of Environmental Management* (pp. 2224–2233). New York: Taylor & Francis.

Stivers, C. (1994). The Listening Bureaucrat: Responsiveness in Public Administration. *Public Administration Review*, 54(4), 364–369. doi:10.2307/977384

Telecommunications Act. (1996). *Preamble*. Retrieved from <http://transition.fcc.gov/reports/tcom1996.pdf>

Tickner, J. A. (2003). Precautionary Assessment: A Framework for Integrating Science, Uncertainty, and Preventive Public Policy. In J. A. Tickner (Ed.), *Precaution: Environmental Science and Preventive Public Policy* (pp. 265–278). Washington, DC: Island Press.

True, J., Jones, B., & Baumgartner, F. (2007). Punctuated-Equilibrium Theory: Explaining Stability and Change in Public Policymaking. In P. Sabatier (Ed.), *The Theories of the Policy Process* (pp. 155–187). Boulder, CO: Westview Press.

Tullock, G., & Wagner, R. (1978). *Policy Analysis and Deductive Reasoning*. Lexington, MA: D.C. Heath and Company.

Upshaw, J. (2009). Marinwood Residents Worried About Talk of Cell Tower. *Marin Independent Journal*. Retrieved from http://www.marinij.com/marinnews/ci_12820360

Verdin, A. (2014). Residents Oppose Cell Tower Project; Fallbrook Community Planning Group Denies AT&T Project Amidst Concerns. *Fallbrook Bonsall Village News*. Retrieved from <http://www.thevillagenews.com/story/76500/>

Weber, M. (1946). Bureaucracy. In H. Gerth & C. Mills (Eds.), *From Max Weber: Essays in Sociology* (pp. 196–264). New York, NY: Oxford University Press.

Weber, M. (1947). *The Theory of Social and Economic Organizations* (A. Henderson & T. Parsons, Eds.). New York, NY: Free Press.

West San Pedro Neighborhood Alliance. (2014). *Fight T-Mobile Cell Towers In Residential RI Neighborhoods*. Retrieved from <http://www.westsanpedro.org/index.html>

Wildavsky, A. (1998). *Federalism & Political Culture*. New Brunswick, NJ: Transaction Publishers.

Wilson, W. (1887). The Study of Administration. *Political Science Quarterly*, 2(2), 197–222. doi:10.2307/2139277 PMID:4591257

Zaller, J. (1992). *The Nature and Origins of Mass Opinion*. New York: Cambridge University Press. doi:10.1017/CBO9780511818691

ADDITIONAL READING

Bell, D. (1960). *The End of Ideology: On the Exhaustion of Political Ideas in the Fifties*. Cambridge, MA: Harvard University Press.

Bloor, D. (1976). *Knowledge And Social Imagery* (2nd ed.). Chicago, IL: The University of Chicago Press.

Cooper, P. (2009). *The War Against Regulation: From Jimmy Carter to George W. Bush*. Lawrence, KS: University of Kansas.

Dahl, R., & Stinebrickner, B. (2003). *Modern Political Analysis* (6th ed.). Upper Saddle River, NJ: Prentice Hall.

Dunleavy, D., Margetts, H., Bastow, S., & Tinker, J. (2006). *Digital Era Governance: IT Corporations, The State, And E-Government*. Oxford, NY: Oxford University Press. doi:10.1093/acprof:oso/9780199296194.001.0001

Dunleavy, P., & Carrera, L. (2013). *Growing The Productivity Of Government Services*. Northampton, MA: Edward Elgar Publishing, Inc. doi:10.4337/9780857934994

Emmett, S., & Wright, P. (2011). *Excellence in Public Sector Procurement: How To Control Cost And Add Value*. Cambridge, England: Cambridge Academic.

Fesler, J. (1980). *Public Administration: Theory And Practice*. Englewood Cliffs, NJ: Prentice Hall Inc.

Fiorino, D. (2006). *The New Environmental Regulation*. Cambridge, MA: MIT Press.

Hill, M., & Hupe, P. (2009). *Implementing Public Policy* (2nd ed.). Thousand Oaks, CA: Sage Publications Inc.

Hindmoor, A. (2006). *Rational Choice*. New York, NY: Palgrave Macmillan. doi:10.1007/978-0-230-20997-8

Hood, C., & Lodge, M. (2006). *The Politics of Public Service Bargains: Reward Competence, Loyalty And Blame*. Oxford, NY: Oxford University Press. doi:10.1093/019926967X.001.0001

Kerwin, C., & Furlong, S. (2011). *Rulemaking: How Government Agencies Write Law And Make Policy* (4th ed.). Washington, DC: CQ Press.

Mahoney, J., & Thelen, K. (2010). *Explaining Institutional Change: Ambiguity, Agency, And Power*. New York, NY: Cambridge University Press.

Manin, B. (1997). *The Principles of Representative Government*. New York, NY: Cambridge University Press. doi:10.1017/CBO9780511659935

Miller, H. (2002). *Postmodern Public Policy*. Albany, NY: State University of New York Press.

Moore, M. (2013). *Recognizing Public Value*. Cambridge, MA: Harvard University Press.

Morgan, G. (2006). *Images of Organization*. Thousand Oaks, CA: Sage Publications Inc.

National Telecommunications and Information Administration. (2014). *Manual of Regulations & Procedures for Federal Radio Frequency Management* (Redbook). United States Department of Commerce. 14, May, 2014.

Pall, M. (2015). Microwave Frequency Electromagnetic Fields (EMFs) Produce Widespread Neuropsychiatric Effects Including Depression. *Journal of Chemical Neuroanatomy*. 2015.08.001.10.1016/j.jchemneu

Parsons, T. (1937). *The Structure Of Social Action*. New York, NY: The Free Press.

Perrow, C. (1970). *Organizational Analysis: A Sociological View*. London, England: Tavistock Publications, Ltd.

Shadish, W., Cook, T., & Campbell, D. (2002). *Experimental and Quasi Experimental Designs: For The Generalized Causal Inference*. Belmont, CA: Wadsworth Cengage Learning.

Stone, D. (2002). *Policy Paradox: The Art Of Political Decision Making*. New York, NY: Norton & Company.

Teppati, V., Ferrero, A., & Sayed, M. (2013). *Modern RF And Microwave Measurement Techniques*. New York, NY: Cambridge University Press. doi:10.1017/CBO9781139567626

Ulaby, F., Michielssen, E., & Ravaioli, U. (2010). *Fundamentals of Applied Electromagnetics* (6th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Weidenbaum, M. (1969). *The Modern Public Sector: New Ways of Doing The Government's Business*. New York, NY: Basic Books, Inc.

KEY TERMS AND DEFINITIONS

Advocacy Coalition: A group of like-minded policy participants and stakeholders that seek policy change through lobbying and communication with policy makers.

Bounded Rationality: The cognitive limitations of decision-makers and the arbitrary underpinning of the advocacy coalition approach.

New Public Management: An emphasis on privatization and outsourcing in order to maximize profitability and optimize resource allocation.

Postmodernism: A new consideration in public administration that tends to ignore productivity and profitability constraints in favor of more humanistic indicators of preference.

Public Choice Theory: A theoretical proposition that explains policy participant and stakeholder hesitancy despite shared common interests.

Rational Choice: Policy analysts make decisions that maximize public value while minimizing the cost of such benefits.

Risk Regulation Regime: Ideological and narrative underpinnings of the regulatory framework governing specific public policy issues.

Social Exchange: The notion that stakeholder behaviors change according to the situational context such as the number of bystanders present.

Stasis: Public policy issues reach a stalemate where the political context elements underlying risk regulation regimes are unable to influence effective policy change.

Public Policies for Providing Cloud Computing Services to SMEs of Latin America



Mohd Nayyer Rahman
Aligarh Muslim University, India

Badar Alam Iqbal
Aligarh Muslim University, India

INTRODUCTION

SMEs are an integral part of the Latin American Economic Region as they represent 99% of all the businesses and account for around 67% employment (LACFORUM, 2013). Thus, SMEs are the backbone of the Latin American economic region: a region characterised by high aspirations of the people towards trade and development. Since times immemorial governments are trying to boost the performance of SMEs due to their power of generating domestic employment. In the present world of technologically advanced operations, there is huge potential in SMEs to contribute towards the development of the region. Challenges are posed by the MNCs in the same region or the ones entering the region. MNCs are characterised by the use and employment of technologically sophisticated method of manufacturing and operations. Apart from this MNCs are also interested in R & D activities and widely invest in R & D department for improving efficiency in order to gain economies of scale. They benefit from the use of IT enabled services to increase the efficiency of the business. For managing knowledge, MNCs widely use cloud computing services provided by specialised companies. Cloud computing has been used as a tool by which ITES can be utilised by firms and it can help in better economical and operational decision making. SMEs of the Latin American economic region too aim for

increasing efficiency and optimising the use of ITES. Cloud computing will empower the SMEs of Latin America to compete with the MNCs. Cloud computing as a modern concept under felicitation of business by adding value to the business and operations achieving cost efficiency in the business. If SMEs of the Latin American economic region are provided cloud computing services through policy initiatives, it would not only add value to the SMEs businesses but also will increase the sustainability of the economic region. Cloud Computing is an opportunity to utilise shared resources for optimising business operations in the technologically driven global economic environment. Typically, cloud computing services include access to databases for the businesses, access to software that is important for decision making and contribution to knowledge and information sharing. Cloud Computing aims to reduce the amount of complexity, minimise costs, and enhance organisational agility (Ghaffari et al, 2014). Cloud Computing decreases the obstacles to conducting information process intensive activities. Indeed, people do not need to maintain their own technology infrastructure as they transfer the burden of system management and data protection to the cloud computing service provider (Jeager et al, 2008). Thus, the study will focus on the issues related to the formulation of public policies for providing cloud computing services for SMEs of Latin America.

BACKGROUND/CONCEPTUAL FRAMEWORK

Cloud Computing

It involves the use of appropriate hardware and software along with networks that allow centralised data storage and online access to the same. It also includes free or restricted (depending on the political and economic environment of a country) access to computer services and resources. Its aim is to achieve economies of scale and coherence by sharing knowledge resources. The term cloud is used as a metaphor for a setup both tangible and intangible that is a collection of tools and resources related to IT-enabled services. With cloud computing services, a business can optimise both its operations and decisions as more easy and fast sharing of knowledge is possible. Cloud Computing is a set of services that provide infrastructure resources using internet media and data storage on a third party server.

Cloud Computing comprises of three services:

1. **Software-as-a-Service (SaaS):** Under this, particular service software is provided online for the consumption of the end user. It stands in opposition to the purchase of the software and then regular updates by the client businesses. The prominent software under this category are applications like Word Processing, CRM (Customer Relationship

Management), ERP (Enterprise Resource Planning) etc. This is a matured model and through it, businesses can achieve economies of scale. Commercial vendors include Yahoo mail, Gmail, Hotmail, TurboTax Online, Facebook, Twitter, Microsoft Office Live, Google Apps, Cisco WebEx conferencing etc.

2. **Platform-as-a-Service (PaaS):** Under this service, software development kits and tools are provided on platforms. The tools include Java, NET, Python, Ruby on Rails. Prominent commercial vendors include Microsoft Azure Services, Amazon Web Services (AWS), Salesforce, Google App Engine Platform, IBM Cloudburst, Amazon's relational database services, Rackspace Cloud Sites etc.
3. **Infrastructure-as-a-Service (IaaS):** This refers to devices such as storage devices, servers, virtual computers, network transfers etc. which are physically located in one central place which is known as data centre but they can be accessed and used over the internet from anywhere using the login authentication systems.

Within organisations there are different cloud deployment models such as:

1. **Public Cloud:** Easily and economically available from a third party provider via

Figure 1. Service model of cloud computing

Source: Prepared by author

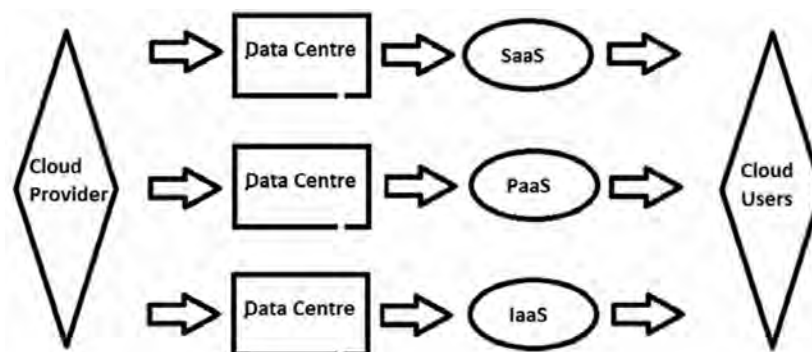


Table 1. SME definitions used by multilateral institutions

Institution	Max. of Employees	Max. Revenues/Turnover (\$)	Max. Assets (\$)
World Bank	300	15000000	15000000
MIF-IADB	100	3000000	None
African Development Bank	50	None	None
Asian Development Bank	No official definition/ Individual national definition used	No official definition/ Individual national definition used	No official definition/ Individual national definition used
UNDP	200	None	none

Source: Gibson & Vaart, 2008, Brookings Global Economy and Development

- internet for deploying IT solutions. The famous example is Google Apps.
2. **Private Cloud:** It is suitable for large organisations and remains within organisations. The US government cloud product (Federal Information Security Management Act) FISMA certifies such clouds and is being handled by a third party provider e.g. Google Private Clouds.
 3. **Community Cloud:** It is used and controlled by a group of enterprises that have shared interests in the cloud computing services e.g. US Federal government is using cloud community such as forms.gov, file.gov, cars.gov, USA.gov, Apps.gov.
 4. **Hybrid Cloud:** It is a combination of the public and private cloud.

Public Policy

Public policy refers to the administrative steps taken by executive branches of states towards a specific or a class of issues in a way that is consistent with the legal framework of the state. The aim of the strong public policy is to solve problems efficiently and effectively, serve justice, support governmental institutions and encourage active citizenship. In other words, the term public policy always refers to the actions of government and the intentions that determine those actions. The policy is a series of actions coordinated to achieve a goal. Three qualifications are necessary for public policy:

1. The idea of the intentional course of action stating not to take a certain action.
2. The requirement that official actions be sanctioned by law.
3. Laws should not be mistaken for the whole realm of the policy.

Small and Medium Enterprise (SMEs)

SMEs are defined differently by different organisations and countries. Their definition may be based on the size of the firm, revenue/sales/turnover, number of employees, access to international markets etc. Different definitions of SMEs as given by multilateral institutions are presented in Table 1.

Among Latin American countries there are at least two different definitions of an SME. One is based on the number of employees in the firm and the other uses sales revenue to determine the economic size of production units. The first definition ignores what was usually major differences among sectors (and among branches within each sector), often resulting in SMEs contribution to the economy being overestimated (LAEO, 2013). In Argentina and Chile, the two criteria have been used to estimate the difference between the official figures and the real figures. The results found that the official figure for the contribution of SMEs to the total number of the job was 7 percentage points too high in Argentina and more than 20 points too high in Chile (Stumpo, 2007). However, this is the criterion used in national

statistics departments, which often provide the data available in these countries, while the policy making bodies define company sizes based on the sales revenue criteria.

REVIEW OF LITERATURE

Several studies highlight the importance and need of cloud computing in general and few others in particular about the SMEs improvement through cloud computing. From a policy perspective very few studies are available on cloud computing but no such study in reference to Latin American SMEs is available. Authors have identified the paradigm shift in businesses due to cloud computing and have equated the transformation as analogous to the replacement of individual generators by the centralised electricity grid (Etro, 2011; Li & Wang, 2011). The development of cloud computing was captured as an opportunity by large enterprises throughout the world by adopting cloud computing bandwagon (Klie, 2011; Li et al, 2011; Walsh et al, 2011). But the decision has remained tough for the SMEs to decide whether to take cloud computing services or not and if yes up to what extent (Gupta, Seetharaman& Raj, 2013). Studies have highlighted that SMEs are a potential market for cloud computing and companies are ready to offer cloud computing services at fewer prices as compared to the same service provided to large corporations. The prime need of SMEs is software and other services that help them to manage emails, licensing to software's, other assets etc (Ferguson, 2008). King, 2008 has identified several points in relation to small businesses moving to cloud computing services such as IT infrastructure availability, minimal upfront investment, disaster recovery, software upgrades etc. Through this SMEs can reduce the cost of knowledge operations, potential avoidance of natural disaster mishaps but the fear is for reliability in using cloud computing services by the SMEs. In the study, King concluded that

SMEs are showing positive inclination towards cloud computing. The important cloud computing parameters that are looked by SMEs includes trust in cloud providers, incremental cost, and reliability. The benefits identified for SMEs using cloud computing services includes security, reliability, trust, cost reduction, online collaboration etc. The success of the cloud computing adoption by SMEs was discussed and it was concluded that SMEs can explore cloud computing with relatively little risk (Clark, 2009). It was observed and empirically searched that security and privacy are top concerns of 51% small and medium businesses. Other factors considered by SMEs includes availability versus sudden downturn and migration across cloud services. The importance of moving to cloud step by step was recommended using a couple of tips for SMEs and thus moving to cloud computing was strongly emphasised (Martin, 2010). The variables for adoption of cloud computing services by SMEs includes cost reduction (in terms of data storage, subscription, low upfront cost, cost control via scalability, elasticity of resources), convenience (easy usage in terms of accessibility and availability from anywhere and anytime), reliability (indicates dependability to use it whenever needed), sharing and collaboration; and security and privacy (Gupta, Seetharaman& Raj, 2013). Large number of authors have classified three services under cloud computing: Software-as-a-Service, Platform-as-a-Service and Infrastructure-as-a-Service (Mahesh et al, 2011; Sultan, 2011; TruongandDustdar, 2011; Ojala&Tyrvaainen, 2011; Creeger, 2009; Liet al, 2011; Durkee, 2010; Marston, Li, Bandyopadhyay, Zhang, and Ghalsasi, 2011; Karadsheh, 2012; Rath, 2012; Neves, Marta, Correia, and de Castro, 2011; McAfee, 2011). There has been no common standard or definition for cloud computing seems to exist (Grossman, 2009; Voas& Zhang, 2009). In the history of IT, the manner in cloud computing is providing elasticity of using resources without having to pay a premium for large scale investment is unprecedented in the history of IT (Armbrust et

Table 2. Cloud revenue (in billion dollars)

Region/ Country	2011	2014 (Estimate)	2016 (Estimate)	CAGR in %
Argentina	0.2	0.4	0.6	28.5
Brazil	1.4	2.7	4.4	25
Mexico	0.6	1.1	1.8	26
Other	0.2	0.5	0.8	26
Latin America	2.4	4.7	7.6	26.4

Source: Cloud Computing in Latin America, ECLAC-Projects Documents Collection, 2014

al, 2009). A research conducted by Easynet Connect showed that UK SMEs are increasingly eager to adopt cloud computing with 47% planning to do so within the next 5 yrs. Of those which indicated their preparedness to move to cloud computing, 35% of them cited cost savings as the key driver (Stening, 2009).

European Network and Information Security (ENSIA) in its survey found that 68% of the SME responses it received indicated that avoiding capital expenditure in hardware, software, IT support and information security is behind their possible engagement in cloud computing (ENISA, 2009). Sultan, 2011 concluded in its study that cloud computing is likely to be an attractive option for many SMEs due to its flexible cost structure and scalability. Weintraub and Cohen (2015). Attempted in their study to minimize the cost of Cloud Computing services provided to consumers. They came up with three models such as hierarchical, simple pricing model and the complete pricing model. In order for the price models to be effective, they also identified three pre-conditions and these will exteriorize in the long run. In another study, same authors reached to a conclusion that consumers of cloud computing services are attempting to maximize overall utility while encountering open tariffs. Bundling of Cloud computing services blocks market competition in cloud computing market (Weintraub & Cohen, 2015).

OPPORTUNITIES FOR PROVIDING CLOUD COMPUTING SERVICES TO SMEs OF LATIN AMERICA



It is of paramount importance to highlight the concept of cloud computing services in Latin America as one unit and then move on to build the linkages with the SMEs of Latin America. While building this discussion it must be kept in mind that no specific study is available on the current status of SMEs of Latin America with respect to cloud computing services. The cloud computing services usage is indicated by the revenue generated by the public cloud. The data for revenue is also scarce and is available only in few reports that too of one or two specific years. Thus, it remains difficult for the researchers to highlight the increasing use of cloud computing services. In the coming lines, that would be attempted with the help of data available.

Above mentioned Table 2 highlights the cloud revenue generated in Latin America as a sign of the increase in the use of cloud computing services. The figures for 2011 are real and for other years its estimated data. Argentina shows an increasing trend in the revenue collected through cloud computing services with a CAGR of 28.5%. On the other hand, Brazil and Mexico show CAGR of 25% and 26%, respectively. Both the countries also indicate an increase in the use of cloud computing services. Other countries (apart from Argentina, Brazil, and Mexico in the Latin American Region) also shows an increasing trend in the cloud computing revenues. Overall the Latin American region shows a CAGR of 26.4% which indicates potential opportunity in the future and is a clear indication that the users of cloud computing will increase. It would be justified to state that SMEs of the Latin American region can take benefit of this opportunity and can use cloud computing services. Figure 2 presents the data of Table 2 to highlight the present opportunities in Latin America.

Figure 2. Cloud revenue

Source: Prepared by author from Table 2 data

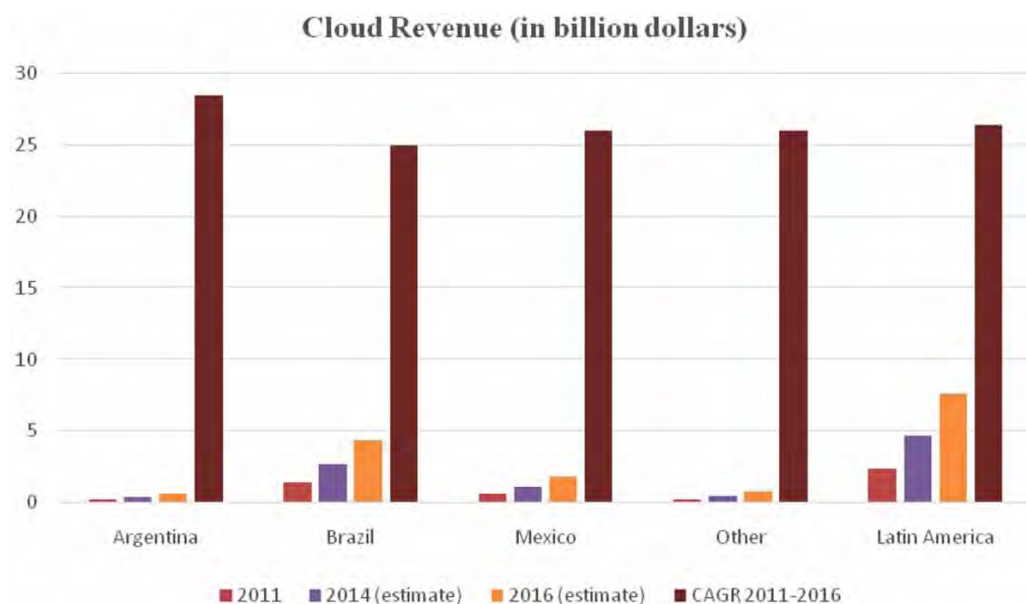


Figure 2 clearly shows that Compounded Annual Growth rate (CAGR in %) is promising for all individual countries as well as for the Latin American region. This may be extended towards the opportunities for SMEs for using the cloud computing services. Further, Table 3 highlights the important variables relevant to the cloud computing services in Latin America.

Elucidation of the details of Table 3 includes variables such as Cloud Traffic Growth, Data Centre Workload, Cloud Workload and Traditional Data Centre Workload. For the period 2013-2018, cloud traffic shows an increase of 35% as Com-

pounded Annual Growth Rate. The rate is high and shows that use of cloud computing services is increasing. The second variable, that is, Data Centre Workload is also increasing with a CAGR of 21%, again highlighting the increase in the use of cloud computing. The third variable, that is, Cloud Workload shows a whopping growth of 34% CAGR and aligns itself with the findings of above data. The final variable given in the table states that traditional data is showing a very slow rate due to the slow transformation in the devices used. New devices are now preferred by the users instead of old ones.

Table 3. Latin America cloud computing data

Variables	2013	2014	2015	2016	2017	2018	CAGR %
Cloud Traffic Growth (exabytes)	89	130	180	240	312	394	35
Data Center Workloads (millions)	2.6	3.2	3.9	4.7	5.7	6.9	21
Cloud Workloads (millions)	1.2	1.7	2.3	3.1	4.1	5.2	34
Traditional Data Center Workloads (millions)	1.4	1.5	1.6	1.6	1.7	1.7	3

Source: CISCO Visual Networking Index: Forecast and Methodology, 2013-2018

Table 4. Regional consumer cloud storage users by 2018 in Latin America

Consumer internet users in millions (% of population)	355 (55%)
Average number of devices per consumer internet user	3.87
Consumer Cloud Storage users in millions (% of internet users)	109 (31%)

Source: CISCO Visual Networking Index: Forecast and Methodology, 2013-2018

Table 4 highlights the estimate for 2018 about key variables that shows the use of cloud computing services. Consumer internet and cloud storage users, both will increase in the years to come. According to the estimate, 55% of the population will be using consumer internet services by 2018 in Latin America and 31% would be using cloud storage services. The average number of the device per consumer will also increase to 3.87. Remember that consumer here does not refer to individuals but the connections (where each connection is taken as a consumer). Respective data related to SMEs would be worth discussing. Table 5 highlights that data.

Table 5 highlights the impact of cloud computing services on SMEs growth and job creation by SMEs in Latin America. As estimated by ECLAC, beginning from 2013, in next 5 years (i.e 2014 – 2019) SMEs in Argentina can create 117300 jobs and number of SMEs will rise by 24700 due to the presence of cloud computing services. With respect to Brazil, the figures are 861000 jobs and 202100 new SMEs in the next 5 years. The estimated figures for the next 10 years are also promising and shows the strength of cloud computing towards SMEs in Latin America. In next 10 years, Argentina will have 30300 more firms

and 128900 new employees under SMEs due to cloud computing services in Latin America. In Brazil, in the next 10 years, the figures would be 402650 new SMEs and 945000 new jobs. This shows that cloud computing has the potential to boost SMEs in Latin America over the next decade. This opportunity must be tapped by the respective countries with the help of public policies.

PUBLIC POLICIES FOR CLOUD COMPUTING SERVICES

This section will be a formulation platform for policies that will boost the SMEs involvement in cloud computing services. Following policies are suggested:

- Improving Access to Cloud Computing Service:** The primary objective of the administrative and executive set up in Latin America should focus on improving access to cloud computing services to SMEs. While corporations are working strategically for this, the administrative initiative is important which can be in the form of another ICT revolution or a new phase of ICT revolution in the region. For this ECLAC must take strong initiatives that should include collaborating with the cloud computing service providers. As it has been stated earlier that corporations are ready to provide cloud computing services to SMEs at low cost as compared to large corporations, they must be taken into confidence for doing so at an accelerated pace. Regulatory bodies of the respective country under Latin American economic

Table 5. Cloud computing impact on no. of SMEs and job creation

	Firms (in 5 yrs)	Employment (in 5 yrs)	Firms (in 10 yrs)	Employment (in 10 yrs)
Argentina	24700	117300	30300	128900
Brazil	202100	861000	402650	945000

Source: Etro & Colciago, Cloud Computing, structural change and job creation in SMEs, ECLAC, 2013

region should assure SMEs of the benefits they may tap in future.

- **Improving Quality of Broadband Internet:** Many countries in Latin America have low fixed or mobile broadband penetration such as Honduras, Nicaragua, Paraguay, Guatemala, Bolivia etc. By improving the quality of broadband, cost efficiency can be achieved by the SMEs. In this regard, policies should be framed that target to increase the quality of broadband. The quality of broadband has a direct relation with devices used by the firms. The cost of installation of devices must be reduced by state sponsoring or making a purchase of cloud computing devices under the category of tax deductible expense. SMEs heavily relying on the use of cloud computing services must be considered as key players and their quality of broadband be accessed regularly.
- **Strong Legal and Regulatory Frameworks:** The reality is that, in many countries within Latin America, there is no strict legal or regulatory framework. At administrative level laws must be more stringent so as to provide safety and security to the cloud computing devices may not be misused. For this fiscal norms must also be adhered to. This will boost the confidence of SMEs towards cloud computing and they will also take benefits of cloud computing. SMEs, in general, are not aware of existing legal and regulatory framework related to cloud computing. The government of the respective country should take the initiative in training SMEs about the existing laws and regulatory framework. Changes in existing framework that may benefit SMEs must be considered by policymakers on a serious note. Step by step strategies should be adopted to universalise the legal framework in the Latin American economic region.
- **Service Level Agreements for the Private Provider:** The main concern for SMEs is about security and portability. Due to this SMEs are reluctant to fully migrate to cloud services. Private providers must be asked under regulations and cloud guidelines to follow service level agreements with SMEs and it should cover service adaptability, system security, and latency, service reliability, data security etc. This will act as a motivation for the SMEs to benefit from the cloud computing services. Strict regulation must be made for penalising the service providers in case of breach of the agreement. Presently, there is no clause of penalising the service providers. This will minimise any possible loss to the operations of SMEs and will not affect their working capital or operations negatively.
- **Ease of Use and Convenience:** SMEs must be motivated through policy to outsource their accounting and finance work to the cloud which will leave more time for SME executives to spend on strategic work and initiatives. Their move from PC-based accounting packages to cloud-based ones would be strategically beneficial as well as a source of cost reduction in operations. This will also allow avoidance of continuous hardware updates by SMEs thereby eliminating maintenance woes for utilising different machines. The cloud approach will thus eliminate administrative overhead and will permit access from any geographical location, any device, and from any organisation.

FUTURE RESEARCH DIRECTIONS

The data for cloud computing services is available only in a haphazard manner. No time series data or panel data is available for cloud computing

services across countries. Due to this reason, the present study has used only descriptive methodology. In future, the researchers should work on primary data and should come up with sharing of the cloud computing data so that statistical analysis is possible. On the availability of the panel or time series data, specific causality between cloud computing services and other macroeconomic variables may be identified.

CONCLUSION

In the end, it would be justified to conclude that SMEs play an important role in the Latin American economic region and cloud computing services will improve the operations of SMEs thus benefiting the whole region. The future of cloud computing services in Latin America, particularly for SMEs is promising and will increase both number of SMEs and employment opportunities. There is a need for policies in favour of SMEs to reap the benefit of cloud computing services and for the same policies must target the issues such as access to cloud computing services, quality of broadband, legal and regulatory frameworks, service level agreements and ease and use of convenience.

REFERENCES

- Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R. H., & Konwinski, A. (2009). *Above the clouds: A Berkeley view of cloud computing*. Electrical Engineering and Computer Sciences, the University of California at Berkeley. Retrieved from <http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-28.pdf>
- Clark, L. (2009). SMEs can benefit most from the cloud. *Computer Weekly*, 19.
- Creeger, M. (2009, August). CTO roundtable: Cloud computing. *Communications of the ACM*, 52(8), 50–56. doi:10.1145/1536616.1536633
- Durkee, D. (2010, May). Why cloud computing will never be free. *Communications of the ACM*, 53(5), 62–69. doi:10.1145/1735223.1735242
- ECLAC. (2014). *Cloud Computing in Latin America Current situation and policy proposals*. ECLAC Project Documents Collection.
- ENISA. (2009). *An SME perspective on cloud computing*. Retrieved from <http://www.enisa.europa.eu/act/rm/files/deliverables/cloud-computingsmesurvey/?searchterm=survey>
- Etro, F. (2011, May). The economics of cloud computing. *IUP Journal of Managerial Economics*, 9(2), 7–22.
- Ferguson, S. (2008). Rising above the din. *Education Week*, 25(17), 29.
- Ghaffari, K., Delgosha, S. M., & Abdolvand, N. (2014). Towards Cloud Computing: A SWOT Analysis on its adoption in SMEs. *International Journal of Information Technology Convergence and Services*, 4(2).
- Grossman, R. (2009). The case for cloud computing. *IT Professional*, 11(2), 23–27. doi:10.1109/MITP.2009.40
- Gupta, P., Seetharaman, A., & Raj, J. R. (2013). The usage and adoption of cloud computing by small and medium businesses. *International Journal of Information Management*, 33(5), 861–874. doi:10.1016/j.ijinfomgt.2013.07.001
- International Economic Forum Latin America and the Caribbean. (2013). Retrieved from http://www.oecd.org/site/lacforum/LACForum_summary_ENG.pdf

- Jaeger, P.T., Lin, J., & Grimes, J.M. (2008). Cloud computing and information policy: Computing in a policy cloud?. *Journal of Information Technology & Politics*.
- Karadsheh, L. (2012, May). Applying security policies and service level agreement to IaaS service model to enhance security and transition. *Computers & Security*, 31(3), 315–326. doi:10.1016/j.cose.2012.01.003
- King, R. (2008). Cloud computing: Small companies take flight. *BusinessWeek Online*, 4.
- Klie, L. (2011, December). SMB hosted CRM market set to triple by 2015. *CRM Magazine*, 15(12), 16.
- Latin American Economic Outlook. (2013). Retrieved from http://www.eclac.org/publicaciones/xml/5/48385/LEO2013_ing.pdf
- Li, Q., Wang, C., Wu, J., Li, J., & Wang, Z.-Y. (2011, November). Towards the business-information technology alignment in cloud computing environment: An approach based on collaboration points and agents. *International Journal of Computer Integrated Manufacturing*, 24(11), 1038–1057. doi:10.1080/0951192X.2011.592994
- Mahesh, S., Landry, B. J. L., Sridhar, T., & Walsh, K. R. (2011, July–September). A decision table for the cloud computing decision in small business. *Information Resources Management Journal*, 24(3), 9–25. doi:10.4018/irmj.2011070102
- Marston, S., Li, Z., Bandyopadhyay, S., Zhang, J., & Ghalsasi, A. (2011, April). Cloud computing—The business perspective. *Decision Support Systems*, 51(1), 176–189. doi:10.1016/j.dss.2010.12.006
- Martin, J. A. (2010, April). Should You move your business to the cloud? *PC World*, 28(4), 29–30.
- Neves, F. T., Marta, F. C., Correia, A. M. R., & de Castro, N. (2011). *The adoption of cloud computing by SMEs: Identifying and coping with external factors*. Paper presented at 11a Conferência da Associação Portuguesa de Sistemas de Informac, ão (CAPSI 2011) – A Gestão de Informac, ão na era da Cloud Computing, Lisboa, Portugal.
- Ojala, A., & Tyrvaïnen, P. (2011, July). Developing cloud business models: A case study on cloud gaming. *IEEE Software*, 28(4), 42–47.
- Rath, A. (2012). *Cloud computing: Facing the reality*. Bhubaneswar, India: Batoi.
- Stening, C. (2009). *Every cloud has a Silver Lining*. Retrieved from <http://www.easynetconnect.net/portals/0/downloadfiles/industryinsight/industrynews/Cloud-computing-website-article-final.pdf>
- Stumpo, G. (2007). *Políticas de apoyo a las pequeñas y medianas empresas en América Latina: situación actual y desafíos*. Buenos Aires: SEPYME.
- Sultan, N. A. (2011, June). Reaching for the cloud: How SMEs can manage. *International Journal of Information Management*, 31(3), 272–278. doi:10.1016/j.ijinfomgt.2010.08.001
- Truong, H.-L., & Dustdar, S. (2011). Cloud computing for small research groups in computational science and engineering: Current status and outlook. *Computing*, 91(1), 75–91. doi:10.1007/s00607-010-0120-1
- Voas, J., & Zhang, J. (2009). Cloud computing: New wine or just a new bottle? *IT Professional*, 11(2), 15–17. doi:10.1109/MITP.2009.23
- Weintraub, E., & Cohen, Y. (2015). Cost Optimization of Cloud Computing Services in a Networked Environment. *International Journal of Advanced Computer Science and Applications*, 6(4), 148–157. doi:10.14569/IJACSA.2015.060420

Weintraub, E., & Cohen, Y. (2015). Optimizing User's Utility from Cloud Computing Services in a Networked Environment. *International Journal of Advanced Computer Science & Applications*, 1(6), 153–163.

ADDITIONAL READING

Azodolmolky, S., Wieder, P., & Yahyapour, R. (2013). Cloud computing networking: Challenges and opportunities for innovations. *IEEE Communications Magazine*, 51(7), 54–62. doi:10.1109/MCOM.2013.6553678

Berry, R., & Reisman, M. (2012). Policy challenges of cross-border cloud computing. *Journal of International Commerce and Economics*, 4(2), 1–38.

Cordeiro, T. D., Damalio, D. B., Pereira, N. C. V. N., Endo, P. T., de Almeida Palhares, A. V., Goncalves, G. E., & Mângs, J. E. et al. (2010, November). Open source cloud computing platforms. In *2010 Ninth International Conference on Grid and Cloud Computing* (pp. 366–371). IEEE. doi:10.1109/GCC.2010.77

Furht, B., & Escalante, A. (2010). *Handbook of cloud computing* (Vol. 3). New York: Springer. doi:10.1007/978-1-4419-6524-0

Gupta, P., Seetharaman, A., & Raj, J. R. (2013). The usage and adoption of cloud computing by small and medium businesses. *International Journal of Information Management*, 33(5), 861–874. doi:10.1016/j.ijinfomgt.2013.07.001

Knorr, E., & Gruman, G. (2008). What cloud computing really means. *InfoWorld*, 7.

Kshetri, N. (2010). Cloud computing in developing economies. *IEEE Computer*, 43(10), 47–55. doi:10.1109/MC.2010.212

Rittinghouse, J. W., & Ransome, J. F. (2016). *Cloud computing: implementation, management, and security*. CRC Press.

KEY TERMS AND DEFINITIONS

Cloud Computing Services: Services which provide the facility of data storage both offline and online. It is data storage on a virtual world platform.

Community Cloud: It is used and controlled by a group of enterprises which have shared interests in the cloud computing services.

Hybrid Cloud: It is a combination of the public and private cloud.

Private Cloud: It is suitable for large organisations and remains within organisations.

Public Cloud: Easily and economically available from a third party provider via internet for deploying IT solutions.

Public Policy: It refers to the administrative steps taken by executive branches of states towards a specific or a class of issues in a way that is consistent with the legal framework of the state.

Small and Medium Enterprises: Enterprises whose scale of production is relatively small in the economy.

Category R

Research Methods and Scholarly Publishing

Advancement and Application of Scientometric Indicators for Evaluation of Research Content

R**Tazeem Zainab***University of Kashmir, India***Zahid Ashraf Wani***University of Kashmir, India*

INTRODUCTION

The 20th century may be designated as the century of the growth and development of metric sciences (Kumar et al., 2009). There has been a tremendous development of metric based fields like Bibliometrics, Scientometrics, Librametrics, Technometrics, Sociometrics, Econometrics, Biometrics, Cybermetrics or Webometrics in this epoch.

Scientometrics can be considered as an analogous conception to bibliometrics. Scientometrics is a novel scientific field joining science and technology with information science and expending numerous mathematical, statistical, data mining techniques and procedures to measure and quantify scientific information. It can be perceived as a discipline of science that encompasses bibliometrics, informetrics, webometrics, librametrics and other metric sciences. As matter of fact, today when the phase of scientific and technological revolution is massive but funding resources are trifling, the measurement of eminence of research productivity is essential and acknowledged by many. The basic requirement is to endorse the progression of science as competently as possible, i.e. to in the best way upsurge the power/prize degree from the financing science (Fiala, 2013). The term scientometrics was introduced by Vassily V. Nalimov & Z. M. Mulchenko in 1969 as Naukometriya in Russian, meaning the study of the evolution of science through the measure-

ment of scientific information (Glanzel, n.d.). As per Tague-Sutcliffe (1992) Scientometrics is concerned with the quantitative facets of science. The focus of scientometrics as a discipline is the literature of science and technology. Price (1961, 1963) defines Scientometrics is a science about science. It offers numerous perceptions, representations, and practices to researchers that when functional in an academic field helps to understand its fundamentals, position, knowledgeable core, and probable forthcoming progress. Wilson (1999) indicated that everything that encompasses quantitative features of science of science, science communiqué and science policy are in the content of scientometrics. While defining the term Van Raan (1997) also accentuated the quantitative learning of science and technology. Vinkler (2010) defined that Scientometrics cannot be circumscribed within the circle of a scientific discipline. He widened the description as quantitative study of people, sets, materials and phenomena in science and their relationships. Further, he adds that scientometrics also covers various other aspects like practices of researchers, socio-organizational arrangements, administration, procedures, national economy. He also specified that Scientometrics could be foundation of statistics and also can indicate the policy in science like performance checking, research precedence assortment, science-society or science-economics relative scholarships. Wilson (1999) considers scientometrics as an organized

DOI: 10.4018/978-1-5225-2255-3.ch583

method to assess the past, present, and future progression of science as he believes that its origin is from the interest of trivial group of scholars in the subtleties of science.

BACKGROUND

The term “Scientometrics” was not noticed in Western scientific circles until it was translated into English. The roots of Scientometrics lie in the survey piloted by Galton in 1874 among 180 prominent scientists in Britain in order to measure, comprehend, and define the eminence of significant scholars and their potentials (Godin, 2006). Then in 1900’s James McKeen Cattell, the Psychologist and Editor of Science, enthused by Galton measured scientific growth by observation and classification. He offered two important facets of scientific output: quality (i.e. worth as adjudicated by peers) and quantity (i.e. production). After that, in 1926 Lotka’s mathematical model to estimate the frequency of author’s publication in a field, in 1934, Bradford’s distribution law for articles through a set of journals.. Further the outline put forth by Price (1961, 1963) regarding the historical evolution of science and Bernal’s (1939) theory of the social function of science, idea of sociology of science by and Merton (1968, 1973) provided a back bone to the development of the field. Finally in 1969 the field was given a comprehensive name “Scientometrics” by the Russian mathematician Vasilii Nalimov. In 1977, this area took a great leap when a maiden issue of Scientometrics journal by T. Braun was available, and the term received an academic acknowledgement. And today, there are numerous journals dedicated to scientometrics and its related fields like, Research Policy, Journal of Informetrics, Social Studies of Science, Journal of the American Society for Information Science and Technology, etc.

SCIENTOMETRIC INDICATORS

Bibliometric indicators aka Scientometric indicators (when referred in the field of Science) provide mathematical measures that are envisioned to quantitatively define the worth of scientific research and the scholarly publication in a particular domain. Scientometrics indicators are valuable assets for determining the capacity in terms of quality and quantity of scientific research of researchers, establishments or nations. El-Maamiry and Gauri (2013) define bibliometric indicator as a maneuver founded on bibliographic facts and figures employed to quantify and evaluate scientific scholarly output of an individual, institution, nation and so on. According to Vinkler (2001) Scientometric indicator is a scientometric measure which is accredited to scientometric organizations. To illustrate he has put forth the examples like Garfield Factor for a journal in a given year; Publication Productivity of an institution. Durieux and Gevenois (2010) emphasize that scientometric indicators are specifically essential for investigators and research institutions as these dimensions are frequently used in research funding, activities, and performance of researchers. They further elaborate that with increasing advances in scientific developments novel research is continuously asserted by new researchers thus making bibliometric indicators more significant. Scientometrics does not exist without the practice of quantitative information and indicators. According to him indicators are the spirit of scientometrics as they illustrate the communication process in science. Scientometric indicators are key parameters for the purpose of assessment and evaluation of scientific research output as they allow us to understand the prospective competitive aptitude of scientific research output and provide arithmetic measures to quantitatively define the worth of scholarly and scientific research, the impact and degree of scientific works

-as an auxiliary for the total scientific research output, as being reliable measures for evaluating and assessing the scientific research output and most importantly policy-making.

TYPES OF SCIENTOMETRIC INDICATORS

El-Maamiry and Gauri (2013) in their study have given three categories of bibliometric indicators. *Quantity indicators*, *Quality indicators* and *Structural Indicators*. *Quantity indicators* measure the total production of a researcher. *Quality indicators* measure the prominence and performance of researchers and *Structural indicators* signify the correlation between publications, researchers and research fields. A similar categorization of scientometric indicators has also been given by Durieux and Gevenois (2010). According to them, there are three types of bibliometric indicators: *Quantity indicators* measure the productivity of a particular researcher or research group. *Performance indicators* measure the quality of a journal, researcher, or research group. *Structural indicators* measure connections between publications, authors, or research fields. In this chapter, the scientometric indicators will be discussed under three classes, Quantity indicators, Quality indicators and Structural indicators. As per Durieux and Gevenois (2010) there are three types of bibliometric indicators: *Quantity indicators* measure the productivity of a particular researcher or research group. *Performance indicators* measure the quality of a journal, researcher, or research group. *Structural indicators* measure connections between publications, authors, or research fields.

PROMINENT SCIENTOMETRIC INDICATORS

Various scientometric indicators have been derived so far to measure the performance of research

production of institutions, researchers, research groups, journals and so on. Most of the indicators used to evaluate the scientific research eminence are citation based, which traditionally were derived from the Citation databases like SCI, SSCI, JCR etc. Later in 21st century, web-based databases like Web of Science, Scopus and citation search engines like Microsoft Academic Search, Google Scholar, CiteSeerX began to be used to derive indicators that can be used to evaluate research. These indices help in analyzing and assessing the quality, quantity and also the co-authorship networks, collaborations between researcher, groups, institutes or countries. Some of the prominent scientometric indicators that have been evolved so far are discussed here:

1. Hirsch-Index

Hirsch-index or h-index was proposed by a Physicist J.E. Hirsch in 2005. According to him a researcher has index h if h of his/her N_p papers have at least h citations each and the other papers have no more than h citations each (Hirsch, 2005). H-index is considered to be the appraisal of the influence of a researchers' aggregate research contribution. Today it is highly applied to enumerate the prominence, rank, worth, and comprehensive impression of researchers, research groups or research institutions. Some of the advantages of h-index have been put forth by Hirsch (2005) himself. The advantages pointed out are:

- It pools a measure of quantity in terms of publications and quality by indicating citations received by a research.
- It permits us to illustrate the research production of a scientist with objectivity helping in decision making regarding upgrades, scholarships and awards.
- It is better executed than various other indicators used to evaluate research output like impact factor, citation per paper, number of papers, etc.

- It is easy to understand and can be accessed simply via Thomson ISI Web of Science, Google Scholar.

Though h-index has received a great accolade in terms of research evaluation worldwide but it has come under the scanner for its various critical points. Bornmann and Marx (2003), BAR-ILAN (2008), Bornmann, Mutz, and Daniel (2008), Ball (2012), and Schreiber (2013) give following disadvantages of h-index:

- More citations in h-core is of no value.
- H-index is field-dependent like most of the pure citation measures.
- Self-citations can easily influence the calculations of h-index.
- There is a delinquency in finding of reference standards.
- Data collection for h-index calculation is quite hectic. A complete list of publications with their citations of a researcher is required for its calculation.
- A meticulous problem of h-index is that a complete name of a researcher is required to distinguish between researchers with same name.
- More specifically the index handicaps newcomers as their publications and citation rates are very low.
- It sanctions researchers to respite on their glories as citations increase over-time even if no new publication is added.

Due to a number of limitations of h-index, a number of variants of h-index like g-index, a-index, hg-index, m-index and so on have been introduced to overcome its deficiencies.

2. Impact Factor

Citations are an important factor in depicting the quality of scientific literature. The impact factor is based upon the count of these citations to determine the value of research. According to Saha,

Saint and Christakis (2003) Impact factor is the citations received by a journal article in scientific literature. It is given by the number of citations in year $n+1$ to a published work in a journal in year n divided by the total number of articles published in year n in that journal. This index as a journal quality index is based on the conception of citation frequency.

Many advantages of Impact factor have been put forth by experts. It is considered to be quite beneficial for knowledge administrators and researchers; information professionals and also the librarians. Some of the advantages of Impact factor given by Honekopp and Kleber, 2008, Khan and Hedge (2009), Garfield and Pudovkin (2015) are:

- It enables experts to draw a comparative relation between different journals and research groups.
- It's easily available to use and understand.
- Since Impact factor is fundamentally based on the citation patterns in a subject area it is considered to be an objective measure of quality of research thus having a wider acceptance worldwide.
- Impact factors reveal the actual status of a journal within a research field, as the number of citations increases or declines in a particular timespan.
- Impact factor is specifically used for the prediction of the forthcoming citations of an article. Further than this, there certainly are some intrinsic snags in journal Impact factor (IF). The disadvantages have been discussed below.

Disadvantages

Researchers Honekopp and Kleber (2008), Khan and Hedge (2009), Nisonger (2004), and Seglen (1997) have stated various problems that are associated with IF. Some of them are:

- Impact factor does not relate with definite citations of specific articles and hence can-

not be considered to be statistically representing individual journal articles.

- Inclusion of self-citations creates a mess.
- Citations to “noncitable” stuff are speciously encompassed.
- High citations to review articles leads to inflation of the impact factor of journals.
- Length of articles matters a lot. Lengthy articles receive more than short length papers.
- Citations within the same journals inflates IF.
- Books are not included as a source of citation in various databases which drops the IF.
- IF varies from year to year as citation number keeps changing.
- Prejudiced towards English language, hence leading American publications.
- IF depends upon expansion of research area and also the relations between various research fields.

3. G-Index

G-index is considered to be a well-known index to calculate the scientific output at a broader level (Woeginger, 2008). It was given by Leo Egghe in 2006 as the improved variant of h-index which defines the performance in terms of citations of an article at the global level. Egghe (2006) defines it as the f number of papers which receive total or more than g^2 citations. He further elaborates that $g \geq h$ and also the g-score is high when citations are high in the g-core.

Advantages

- G-index provides information regarding both the size and the impact of the g-core.
- More comprehensive than h-index.
- Rather than using many indexes like h-index, r-index separately for measuring output and eminence of research, combined g-index can be employed (Schreiber, 2009).

Disadvantages

Schreiber (2008, 2009, 2010) proposes various cons of g-index. According to him, following points should be reconsidered to improve the efficacy of g-index.

- Self-citations are ignored and excluded.
- The publication count are fractionalized.
- The collaboration index or the number of co-authors should be considered while calculating g-index.

4. M-Quotient

The m-quotient proposes to measure the academic term and length of a researcher despite his period of scientific career. It is given by:

$$h\text{-index}/k$$

where k designates the number of years since a researcher's first publication.

Advantages

- Career length is an important factor in calculation of m-quotient.
- Based on citation metrics.

Disadvantages

- The m-quotient of an author is not properly established till initial years of his/her career.
- Minor changes in h-index of an author can alter the m-quotient altogether.
- As it is based on h-index, it nevertheless furnishes it anomalies in terms of quality and quantity measurement.

5. Shanghai Ranking

Shanghai Ranking or Academic Ranking of World Universities (ARWU) is a yearly publication of

university rankings by Shanghai Ranking Consultancy. It is based on four criteria namely quality of education, quality of faculty, research output and Per Capita Performance. Quality of faculty represents highly cited researchers, Nobel Laureates and Field Medalists. The research output is measured on the basis of indicators like papers in NATURE and SCIENCE and also papers indexed by SCI (Zhang et al., 2015).

Advantages

- Its widely accepted Institutional Ranking System because of its transparency, methodology used and purpose.
- It is based on Scientific indicators like citations, research output, etc.

Disadvantages

- It is dependent on award factors decreasing the importance of instruction and humanities.
- It is widely criticized for the criteria it uses for ranking universities (Van Raan, 2005).
- Florian (2007) consider that any kind of ranking and numbering universities cannot be accurate in any kind of ranking system.

CONCLUSION

The evaluation of qualitative research is being prioritized throughout the world for grading of universities and research institutions which in turn affects the research funding of such institutions by different funding agencies. Therefore analyzing research in different areas has always been an uphill task for evaluating bodies that resulted in the development of different evaluation indicators for analyzing the qualitative research. The scientometric indicators particularly are of great use in the measurement and evaluation of the scientific research output but at the same time it

requires careful efforts in use. Despite criticism at various levels, most professionals believe in it and is one of the widely used indicator categories. Though there is abundant literature emphasizing critics' limits and partialities but academicians and researchers have confidence in it consider it the eminent method and agree that it is the only way to assess and evaluate the scientific research.

FUTURE RESEARCH DIRECTIONS

The arena of research in scientometric evaluation is evolving with the development of more specific variants of scientometric indicators. These indicators are widely being employed in evaluation of different parameters of research including the impact of Journal publications, articles and author performance. The field is continuously evolving with the incorporation of nascent indicators each day. However there is a dearth of most comprehensible indicators that can be employed to evaluate and analyse each aspect of the scientific productivity. Therefore further research needs to be carried out to study different dimensions associated with Scientometric indicators so as to be more precise and appropriate in revealing qualitative aspects of research.

REFERENCES

- Bernal, J. D. (1939). *The social function of science*. London: George Routledge & Sons.
- Durieux, V., & Gevenois, P. A. (2010). Bibliometric indicators: Quality measurements of scientific publication. *Radiology*, 255(2), 342–351. doi:10.1148/radiol.09090626 PMID:20413749
- El-Maamiry, A. A., & Ghauri, M. A. (2013). Measuring information quality: Concerns on the use of Bibliometric studies. *International Journal of Information Dissemination and Technology*, 3(4), 274–278.

- Fiala, D. (2013). Suborganizations of institutions in library and information science journals. *Information*, 4(4), 351–366. doi:10.3390/info4040351
- Garfield, E., & Pudovkin, A. (2015). Journal Impact Factor Strongly Correlates with the Citedness of the Median Journal Paper. *Collnet Journal of Scientometrics and Information Management*, 9(1), 5–14. doi:10.1080/09737766.2015.1027099
- Glanzel, W. (n.d.). *A Concise Introduction to Bibliometrics & its History*. Retrieved from <https://www.ecoom.be/en/research/bibliometrics>
- Godin, B. (2006). On the origins of bibliometrics. *Scientometrics*, 68(1), 109–133. doi:10.1007/s11192-006-0086-0
- Hinze, S., & Glänzel, W. (2012). *Scientometric indicators in use: An over-view*. Presentation at the European summer school for scientometrics. Retrieved from http://www.scientometrics-school.eu/images/21_13Hinze.pdf
- Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National academy of Sciences of the United States of America*, 102(46), 16569–16572. doi:10.1073/pnas.0507655102
- Hood, W., & Wilson, C. (2001). The literature of bibliometrics, scientometrics, and informetrics. *Scientometrics*, 52(2), 291–314. doi:10.1023/A:1017919924342
- Khan, K. M., & Hegde, P. (2009). Is impact factor true evaluation for ranking quality measure? *DESIDOC Journal of Library & Information Technology*, 29(3), 55–58. Retrieved from <http://www.publications.drdo.gov.in/ojs/index.php/djlit/article/viewFile/253/163>
- Kumara, A., Prakasan, E. R., Mohan, L., Kademani, B. S., & Kumar, V. (2009). *Bibliometric and Scientometric Studies in Physics and Engineering: Recent Ten Years Analysis*. Putting Knowledge to Work: Best Practices in Librarianship, Mumbai, India. Retrieved from http://eprints.rclis.org/14829/1/BIBLIOMETRIC_AND_SCIENTOMETRIC_BOSLA-CDAC_Conf__2009-2.pdf
- Merton, R. K. (1973). *The sociology of science: Theoretical and empirical investigations*. Chicago: University of Chicago Press.
- Nisonger, T. E. (2004). The benefits and drawbacks of impact factor for journal collection management in libraries. *The Serials Librarian*, 47(1-2), 57–75. doi:10.1300/J123v47n01_04
- Price, D. D. (1961). *Science since Babylon*. Yale Univ. Press.
- Price, D. D. (1963). *Little Science, Big Science*. Columbia Univ. Press.
- Saha, S., Saint, S., & Christakis, D. A. (2003). Impact factor: A valid measure of journal quality? *Journal of the Medical Library Association: JMLA*, 91(1), 42. Retrieved from <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC141186/> PMID:12572533
- Schreiber, M. (2008). The influence of self-citation corrections on Egghe's g index. *Scientometrics*, 76(1), 187–200. doi:10.1007/s11192-007-1886-6
- Schreiber, M. (2009). Revisiting the g-index: The average number of citations in the g-core. *Journal of the American Society for Information Science and Technology*, 61(1), 169–174. doi:10.1002/asi.21218
- Schreiber, M. (2010). How to modify the g-index for multi-authored manuscripts. *Journal of Informetrics*, 4(1), 42–54. doi:10.1016/j.joi.2009.06.003
- Schreiber, M. (2013). How to derive an advantage from the arbitrariness of the g-index. *Journal of Informetrics*, 7(2), 555–561. doi:10.1016/j.joi.2013.02.003
- Seglen, P. O. (1997). Why the impact factor of journals should not be used for evaluating research. *BMJ (Clinical Research Ed.)*, 314(7079), 497. doi:10.1136/bmj.314.7079.497 PMID:9056804
- Tague-Sutcliffe, J. (1994). Quantitative methods in documentation. Fifty Years of Information Progress: a Journal of Documentation Review, 147–188.

- Van Raan, A. F. J. (1997). Scientometrics: State-of-the-Art. *Scientometrics*, 38(1), 205–218. doi:10.1007/BF02461131
- Van Raan, A. F. J. (2005). Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods. *Scientometrics*, 62(1), 133–143.
- Vinkler, P. (2010). Indicators are the essence of scientometrics and bibliometrics: Comments to the book entitled Bibliometrics and Citation Analysis, From the Science Citation Index to Cybermetrics from Nicola De Bellis. *Scientometrics*, 85(3), 861–866. doi:10.1007/s11192-010-0159-y
- Wilson, C.S. (1999). Informetrics. *Annual Review of Information Science and Technology*, 34, 107–247.
- Woeginger, G. J. (2008). An axiomatic analysis of Egghe's g-index. *Journal of Informetrics*, 2(4), 364–368. doi:10.1016/j.joi.2008.05.002
- Zhang, Y., Eckelmann, B., Andrews, T., Cheng, Y., Vught, F., & Morse, R. (2015). *International Ranking & Institutional Research*. Retrieved from https://manoa.hawaii.edu/miro/wp-content/uploads/2015/06/International-rankings-handouts_final_06092015.pdf
- Zitt, M., & Bassecouard, E. (2008). Challenges for scientometric indicators: Data demining, knowledge-flow measurements and diversity issues. *Ethics in Science and Environmental Politics*, 8(1), 49–60. doi:10.3354/esep00092
- Braun, T., Glanzel, W., & Schubert, A. (1985). *Scientometric Indicators: A 32-Century Comparative Evaluation of Publishing Performance and Citation Impact*. Singapore: World Scientific. doi:10.1142/0106
- Foundation for Research Development (South Africa). (1990). *SA science and technology indicators*. Pretoria: Scientometric Advisory Centre.
- Foundation for Research Development (South Africa). (1991). *South African science and technology indicators 1990*. Pretoria: Foundation for Research Development, Scientometric Advisory Centre.
- Frankel, M. S., & Cave, J. (1997). *Evaluating science and scientists: An east-west dialogue on research evaluation in post-communist Europe*. Budapest: Central European University Press.
- Ingwersen, P. (2012). *Scientometric indicators and webometrics -- and the polyrepresentation principle in information retrieval*. New Delhi: Ess Ess Publications.
- Južnič, P., Pečlin, S., Žaucer, M., Mandelj, T., Pušnik, M., & Demšar, F. (2010, January 01). Scientometric indicators: Peer-review, bibliometric methods and conflict of interests. *Scientometrics*, 85(2), 2. doi:10.1007/s11192-010-0230-8
- Karamourzov, R. (2012, January 01). The development trends of science in the CIS countries on the basis of some scientometric indicators. *Scientometrics*, 91(1), 1–14. doi:10.1007/s11192-011-0592-6
- Kumari, G. L. (2009, September 01). Synthetic Organic Chemistry research: Analysis by scientometric indicators. *Scientometrics: an International Journal for All Quantitative Aspects of the Science of Science. Communication in Science and Science Policy*, 80(3), 559–570.

ADDITIONAL READING

Andrés, A. (2009). *Measuring academic research: How to undertake a bibliometric study*. Oxford: Chandos Publishing. doi:10.1533/9781780630182

Mendez, A., & Salvador, P. (1992, January 01). The application of scientometric indicators to the spanish scientific research council. *Scientometrics*, 24(1), 61–78. doi:10.1007/BF02026473

Raan, A. F. J. (1988). *Handbook of quantitative studies of science and technology*. Amsterdam: North-Holland.

Ravichandra, R. I. K. (2010). *Growth of literature and measures of scientific productivity: Scientometric models*. New Delhi: Ess Ess Publications.

Suriya, M. (2011). *Gender studies in informatics: An application of scientometric indicators and mapping techniques*. New Delhi: Ess Ess Publications.

Tijssen, R. J. W. (1992). *Cartography of science: Scientometric mapping with multidimensional scaling methods*. Leiden: DSWO Press, University of Leiden.

KEY TERMS AND DEFINITIONS

R

H-Index: H-index is considered to be the appraisal of the influence of a researchers' aggregate research contribution.

Impact Factor: Impact factor is the citations received by a journal article in scientific literature.

Performance Indicators: These measure the quality of a journal, researcher, or research group.

Scientometric Indicators: Metrics founded on bibliographic facts and figures employed to quantify and evaluate scientific scholarly output of an individual, institution, nation and so on.

Scientometrics: The field within bibliometrics that quantifies and encompasses evaluation and assessment of scholarly content within the field of science.

Electronic Theses and Dissertations (ETDs)

Ralph Hartsock

University of North Texas, USA

Daniel G. Alemneh

University of North Texas, USA

INTRODUCTION

Academic libraries around the world are seeking to take advantage of the powerful forces that transform higher education, including new and rapidly changing technologies, an abundance of digital (mostly open access) resources in a myriad of formats, and changing practices in how scholars communicate and disseminate their research and creative work.

Theses and dissertations, the monograph-length essays required for graduate degrees from institutions of higher education, have evolved with the technology. Electronic Theses and Dissertations (ETDs) constitute the primary contributions to a community of research (Ramirez et al., 2014). The term “Electronic Theses and Dissertations” (ETD) is used primarily to differentiate between analog theses and dissertations (paper, microfilm) and their digital counterparts (digital objects). Since 1998, academic institutions increasingly publish theses and dissertations that are born digital.

BACKGROUND

As forms of scholarship evolve, so do users’ and creators’ expectations. Theses and dissertations represent part of the historical record of graduate education at the institution. Those produced prior to the advent of the photocopier were created by the use of carbon paper. However, often in the domain of music, the need for accompanying material required the student to attach a separate

sheet with the musical notation to each copy. Students attached photographs and illustrations, predominantly black and white, in much the same manner.

The first electronic theses and dissertations (ETD) project was launched in 1987 by a business company and a long-term vendor of theses and dissertations for academic libraries, University Microfilms International (UMI), by converting its large collection of dissertations on microfiches and microfilms going back to 1939 into electronic form. The first non-profit ETD hosted by a university was launched ten years later, in 1997, at Virginia Tech, which made electronic submission of theses and dissertations through its ETD system a requirement for the university’s graduating students (Ramirez et al., 2014). Virginia Tech University, along with representatives from UMI and the American Council of Graduate Schools, was one of the founders of the Coalition for Networked Information’s joint project, with the goal to collaboratively develop collections of ETDs. In 1995 this resulted in creation of the Networked Digital Library of Theses and Dissertations (Fox et al., 1997).

Since the late 1990s, an increasing number of academic institutions have mandated the electronic submission of theses and dissertations. Today, textual dissertations need only be in a word processing file and converted to a more permanent and unchangeable file format to become Electronic Theses and Dissertations (ETDs). The current digital submissions of ETDs experienced significant increased usage of graphics or multimedia contents.

DOI: 10.4018/978-1-5225-2255-3.ch584

During the analog age examples of handwritten music had to be glued into the dissertation, with the typescript below it; this included attaching the original music on the carbon copies. With the introduction of musical software (Finale, Sibelius) or imaging software, writers could place these materials inline inside the dissertation. A move to an all-digital means of providing electronic theses and dissertations is accelerating their discovery and facilitating their use, value, and impact in research.

Accompanying Materials

Rebecca Lubas (2009) and Cedar C. Middleton, Jason W. Dean, and Mary A. Gilbertson (2015) present adequate processes for the cataloging and metadata creation of homogenous textual dissertations. However, dissertations increasingly have accompanying materials, most prevalent in music and the performing arts: these have included audio tapes, compact discs, or video recordings of recitals, concerts, and lectures. Traditionally, these audio tapes, either in reel-to-reel or audio cassette format, or videocassettes, in various configurations, were difficult to preserve. Equipment also went out of date, as certain formats became dominant. Beta and U-Matic declined into more limited use as VHS became the standard for videocassettes. Discs, either CDs or DVDs, became the norm during the 1990s.

Since the introduction of ETDs, illustrations have become predominantly color, particularly in the arts and sciences. In addition to increased usage of graphics, those in biological and chemical fields include video demonstrations of their experiments, or may draw the elements and the design of molecules. Today, these are all submitted as streaming audio files or audio visual files and integrated seamlessly with the original ETDs.

Copyright Issues

While increased availability of interoperable Open Access content helps to integrate and enhance ac-

cess to diverse digital resources, they also bring about great challenges for traditional policies. There have been some concerns, questions, and misconceptions about various issues, ranging from intellectual property to quality issues. Two of the primary concerns about ETDs and their accompanying materials are copyright and fair use. Musical and artistic works created after 1923 are held in the creator's copyright until 75 years after that creator's death, due to the Berne Convention Implementation Act of 1988. Musical scores may be either brief examples that support the author's thesis, or comprise the primary contents of the thesis, as is done in new editions of music. Additionally, this affects open access to performances (audio or video) of these compositions, as it affects public performance rights (Dogan, 2011). Due to this, many institutions limit access to these recordings to those enrolled in classes or patrons on the campus. Open access has also been a concern for emerging authors wishing to publish their research in peer reviewed publications. Although some scholars consider this issue to be contentious, Ramirez et al. (2014) has alleviated this concern, stating that the thesis as a non-peer reviewed paper does not constitute double publishing.

Access and Cataloging

As forms of digital scholarship evolve, so do users' and creators' roles and expectations. The Open Access (OA) movement has become increasingly important in shaping the ways that academic libraries provide services to support the creation, organization, management, and use of digital contents. Making ETDs Open or the removals of barriers (pricing, technical and legal hurdles) facilitates successful management of ETDs across the entire life-cycle to ensure their preservation and continuous availability in a manner that current and future users expect. Even though some ETDs are restricted to their specific institutions' users, consulting the metadata description may convey sufficient information, and can be adequate for

most external users' needs. Making the metadata accessible to providers of search and discovery services via Open Archive Initiative Protocol for Metadata Harvesting (OAI-PMH) facilitates usability.

Successful retrieval of materials that are useful to a user relies on the quality of the information representation. In an ideal world, digital content would also be given the same consideration as other library materials, when conducting collection development, organization and cataloging of works, etc.

In information science, an ontology is a formal naming and definition of the types, properties, and interrelationships of the entities that fundamentally exist for a particular domain of discourse. It is thus a practical application of philosophical ontology, with a taxonomy. Historically, ontologies arise out of the branch of philosophy known as metaphysics, which deals with the nature of reality – of what exists. An ontology compartmentalizes the variables needed for some set of computations and establishes the relationships between them (Pawletta et al., 2014). This necessity of relationships is accentuated in music, due to the relationships between the composer, the work, and those who perform recordings of these works. Ludwig van Beethoven composed nine symphonies, each published by different organizations, and recorded by numerous conductors, from Arturo Toscanini to Leonard Bernstein.

The generation of accurate indexing terms is critical to the discovery and use of specific manifestations. However, in practice it is difficult to comprehensively represent every item with appropriate descriptions and index terms. According to Alemneh and Hartsock (2014a), because ETDs usually constitute original research, each is unique to the bibliographic world; as a result, catalogers need to provide original cataloging (i.e., creation of metadata record from scratch) as opposed to copy cataloging (i.e., use of pre-existing metadata records with or without augmentations) to describe each ETD. When catalogers employ traditional controlled vocabularies, such as the Library of

Congress Subject Headings (LCSH), creating access to these original theoretical researches becomes challenging and labor intensive. (Lubas, 2009) Considering the multi-disciplinary and interdisciplinary characteristics, several subjects and terms need to be supplied frequently to adequately represent ETDs for efficient access. Middleton et al. (2015) notes that it is uncommon to provide full level cataloging to theses and dissertations because of this increased labor. But a research status of a certain number of institutions affects accessibility and expectations of graduate students. Thus, these metadata specialists may provide full level cataloging in MARC (Machine-Readable Cataloging) and more universal descriptive standards like RDA (Resource Description and Access) to their dissertations. MARC was developed during the 1960s at the Library of Congress as a model for data input and retrieval (Seikel et al., 2011). It employed three character numerical codes, which utilized 8 bit memory as opposed to natural language labels. A majority of library catalogs online use MARC data, with public viewable interfaces installed by local system vendors. This translates the numerical code into readable labels. The new RDA standard provides a comprehensive set of guidelines and instructions for the creation of descriptions and access, covering all types of digital contents and media. (American Library Association, 2010).

Furthermore, in some disciplines, the subject can also be an aspect of study. Recently introduced are genres and forms, through seven projects of the Library of Congress Genre/Form Terms (lcgft). Several professional associations provided guidance in the formulation of these terms. Beginning in 2007, the Policy and Standards Division (PSD) of the Library of Congress formulated thesauri of genre/form terms for moving images, cartography, law, literature, religion, music, and non-musical sound recordings (predominantly radio programs) (Young et al., 2013) Many genres and forms are combined into a single term. Genre indicates a category of works that is characterized by a similar plot, theme, setting, situation, or characters (such

as Horror, Thriller, Western, Country). Form denotes a category of works with a particular format and/or purpose (for example, Film, Sonata, Symphony, Opera, etc.). Catalogers may add another Genre/Form for textual dissertations, the lcgt of “Academic theses.” A dissertation author may produce an edition of music, or create an original composition; the cataloger then adds the genre of the music in the ETD, for example, Anthems, Madrigals, Folk songs, or Symphonies.

Within the context of radio, genre/form terms denote categories of programs. They describe a program according to its content (Westerns, Biographies), style (Audience Participation Programs, Call-in Shows), topic (Crime or Mystery Programs), structure (Magazines, Anthologies), intended audience (Children’s Programs), method of transmission (Shortwave Broadcasts) or combinations of these. A genre/form category suggests a common theme, motif, setting, situation or characterization that is easily recognizable.

That means, in addition to denoting “Aboutness,” certain music and other subjects will have “Isness” as well, referred to as content in RDA (McKnight, 2012). To illustrate this, current practice prescribes that Sonata, Symphony, or Opera represent a subject in the LCSH when singular, indicating resources that discuss the form itself. For a genre/form term, the plural indicates that this is a manifestation of the genre: a score or a recorded performance.

In music, the author achieves originality by analyzing a specific composer or composition(s), or applying a unique approach or analysis to a specific composition. This aspect may be applied to several works by various composers. For example, multiple dissertation writers may analyze extended techniques that employ specific instruments (trumpet, vibraphone) in relation to specific compositions (Meredith, 2008; Smith, 2008).

During the analog age, from the invention of the printing press until the 1980s, the carrier determined the content of its material. Thus, books are textual and printed, audiocassettes are sound recordings on tape 1/8 inch wide, while VHS is

the standard for 1/2 inch wide videocassettes. The digital age has not only made texts and images interactive, but releases content types from single carriers. Today, a DVD may be audio or video, contain moving images, or still images, sound, textual files, and even games.

In certain disciplines, institutions require the student to perform a recital (vocal or instrumental), and to present a lecture about a specific aspect of a composition, to be performed at the Lecture-Recital. The dissertation is a textual representation of this lecture. To provide access to all facets of ETDs, metadata departments must integrate the cataloging of all the formats involved, not just the text.

In the transitory period between the late analog and the early digital age, recitals were submitted on audiocassettes, or if visual, on videocassettes. These videos could be in a number of formats, but VHS predominated. Later, students submitted their audio recordings on compact discs. The digital reformation makes possible the submission of these aural and visual data via streaming mechanisms. The electronic manifestation of the ETD becomes the record of choice, and the cataloger may provide full level access, including description, subjects from a controlled vocabulary, classification numbers, and functioning links, not only for the textual dissertation (Middleton et al., 2015), but to all accompanying materials via multiple linking fields, such as the MARC (Machine-Readable Cataloging) 856.

While the subjectivity and objectivity of the process and the need to distinguish functional representation from mere descriptions of a topic still exist, recording these data has evolved from the earlier cataloging methods using *Anglo-American Cataloging Rules* (AACR), and *Anglo-American Cataloging Rules, 2nd edition* (AACR2), to the current practice of *Resource Description and Access* (RDA). Catalogers provide Library of Congress Subject Headings (LCSH) for the topics of each dissertation. In addition, those of musical editions contain a Medium of Performance field, located in the MARC format as the 382 tag. This

presents a granular statement of the personnel necessary for an execution of the work, and uses a controlled vocabulary developed by the Library of Congress, with assistance from the Music Library Association. For Gregory Alan Schneider's *I never saw Another Butterfly: A Composition for SATB Choir and Chamber Orchestra* (Schneider, 1997), the LCSH may be represented by a MARC 650 tag: Choruses, Secular (Mixed voices) with chamber orchestra. The Library of Congress Medium of Performance Thesaurus statement (LCMPT) more granularly exhibits which voices are represented in this chorus.

This would display to the cataloger in MARC as:

382 01 mixed chorus \$v SATB \$a orchestra \$2 lcmpt

The public would see this as:

Performance medium: mixed chorus (SATB), orchestra.

FUTURE RESEARCH DIRECTIONS

Theses and dissertations represent a wealth of scholarly and artistic content created by graduate students in masters and doctoral programs in the degree-seeking process. The digital environment has now introduced new resource types into the current information landscape, accompanied by new user expectations. Advancing knowledge requires not only enhancing our capacity to generate more knowledge, but also cultivating our ability to provide seamless discovery processes to the vast quantities of knowledge we continue to generate. As the starting points for new researches are increasingly digital, catalogers and metadata specialists need to provide adequate descriptions of digital contents sufficient to ensure successful information retrieval.

Arguably, ETDs have experienced a continuous evolution in format and structure that require dif-

ferent lifecycle management practices to facilitate access, use, and reuse. Accompanying materials in the ETD landscape can be diverse, ranging from simple visual forms of maps or diagrams to complex and high resolution images of art objects or videos of lab observations. As demonstrated by Alemneh et al. (2014), by employing appropriate metadata elements that link various characteristics and relationships, ETDs' associated contents can be integrated. In light of such diverse contents and formats, institutions integrate and contextualize all parts of ETDs to add value and enhance access and use.

Poole (2015) argued that harnessing the principles and practices of digital curation (ranging from managing quality metadata, provenance and authenticity to ensuring long-term access and trust) facilitate the sharing, access, and reuse of scientific data. Doing so also allows researchers and other stakeholders to address new imperatives in scholarly research.

CONCLUSION

There has been a shift in the way users search, access, and use ETDs. In principle, cataloging or metadata descriptions still serve the same function that they always have: to link creators with users. In other words, by describing information resources adequately, whenever users express their information need, using the best possible descriptor or search term, the results obtained match their information needs. However, in the increasingly self-structured Web 2.0 environment, it is clear that traditional user experience and access methods of a simple ranked list of search results will be of limited utility. Various emerging Web applications – driven by semantic web technologies such as the resource description framework (RDF), semantic Web rules language (SWRL), and other members of the World Wide Web Consortium (W3C) family of specifications – offer powerful data organization, combination, and query capabilities. All these have serious im-

plications for how ETDs are curated. Consequently catalogers have changed their methodology and procedures significantly over the past few decades. Deployment of new cataloging standards including RDA and the exposure of RDA-based data in the linked data cloud have already transformed the traditional practices. Effective metadata descriptions, taxonomies, and ontologies add value and amplify the mostly interdisciplinary ETDs. Complex retrieval systems allow visualization of the information space, among other things—allowing ETD users to explore and delve deeper in multi-dimensional ways.

REFERENCES

- Alemneh, D. G., et al. (2014). *Guidance Documents for Lifecycle Management of ETDs*. Educopia Institute. Retrieved January 29, 2016 from <https://educopia.org/publications/gdlmetd>
- Alemneh, D. G., et al. (2015). Knowledge Representation and Subject Access in ETDs: Analysis of Creators' and Users' Assumptions and Expectations. In T. Watanabe & K. Seta (Eds.), *Proceedings of the 11th International Conference on Knowledge Management*. Retrieved January 29, 2016 from <http://ickm.kis.osakafu-u.ac.jp/program/accepted-papers/>
- Alemneh, D. G., & Hartsock, R. (2014a). Theses and Dissertations from print to ETD: The Nuances of Preserving and Accessing those in Music. In J. M. Krueger (Ed.), *Cases on Electronic Records and Resource Management Implementation in Diverse Environments* (pp. 41–60). Hershey, PA: IGI Global.
- American Library Association. (2010). *About RDA*. Retrieved February 17, 2016 from <http://www.rdatoolkit.org/about>
- Dougan, K. (2011). Dissertations in the Electronic Age: Tapping into Emerging Musicology Research. *Music Reference Services Quarterly*, 14(3), 109–130. doi:10.1080/10588167.2011.596094
- Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2), 199–220. doi:10.1006/knac.1993.1008
- Joint Steering Committee for Revision of AACR. (1982). *Anglo-American Cataloguing Rules* (2nd ed.). Chicago: American Library Association.
- Krueger, J. M. (2013). *Cases on Electronic Records and Resource Management Implementation in Diverse Environments*. Hershey, PA: IGI Global. Retrieved January 29, 2016 from <http://site.ebrary.com/id/10755871>
- Lubas, R. L. (2009) Defining Best Practices in Electronic Thesis and dissertation Metadata. *Journal of Library Metadata*, 9(3-4), 252-263. DOI:10.1080/19386380903405165
- McKnight, M. (2012). Are We There Yet? Toward a Workable Vocabulary for Music. *Fontes Artis Musicae*, 59(3), 286–292.
- Meredith, S. (2008). *Extended techniques in Stanley Friedman's Solus for unaccompanied trumpet*. Retrieved February 15, 2016, from <http://digital.library.unt.edu/ark%3A/67531/metadc6075/>
- Middleton, C. C., Dean, J. W., & Gilbertson, M. A. (2015). A Process for the Original Cataloging of Theses and Dissertations. *Cataloging & Classification Quarterly*, 53(2), 234–246. doi:10.1080/01639374.2014.971997
- Park, E. G., & Richard, M. (2011). Metadata assessment in e-theses and dissertations of Canadian institutional repositories. *The Electronic Library*, 29(3), 394–407. doi:10.1108/026404711111141124

Pawletta, T., Durak, U. & Breitenecker, F. (2014). Ontologies in Modelling and Simulation. *SNE (Simulation Notes Europe)*, 24(2). DOI:10.11128/sne.24.2.1024

Poole, A. H. (2015). How has your science data grown? Digital curation and the human factor: A critical literature review. *Archival Science*, 15(2), 101–139. doi:10.1007/s10502-014-9236-y

Ramirez, M. L., McMillan, G., Dalton, J. T., Hanlon, A., Smith, H. S., & Kern, C. (2014). Do Open Access Electronic Theses and Dissertations Diminish Publishing Opportunities in the Sciences? *College & Research Libraries*, 75(6), 808–821. doi:10.5860/crl.75.6.808

RDA Toolkit. (2012). *Resource Description and Access*. Retrieved February 4, 2016 from <http://access.rdatoolkit.org/>

Schneider, G. A. (1997). *I never saw Another Butterfly: A Composition for SATB Choir and Chamber Orchestra*. Retrieved February 4, 2016, from <http://digital.library.unt.edu/ark%3A/67531/metadc278223/>

Seikel, M., & Steele, T. (2011). How MARC Has Changed: The History of the Format and Its Forthcoming Relationship to RDA. *Technical Services Quarterly*, 28(3), 322–334. doi:10.1080/07317131.2011.574519

Smith, J. D. (2008). *Extended performance techniques and compositional style in the solo concert vibraphone music of Christopher Deane*. Retrieved February 15, 2016, from <http://digital.library.unt.edu/ark%3A/67531/metadc9030/>

Sowa, J. F. (1995). Top-level ontological categories. *International Journal of Human-Computer Studies*, 43(5-6), 669–685. doi:10.1006/ijhc.1995.1068

Spalding, C. S. (1967). *Anglo-American Cataloging Rules*. Chicago: American Library Association.

Young, J. L., & Mandelstam, Y. (2013). It Takes a Village: Developing Library of Congress Genre/Form Terms. *Cataloging & Classification Quarterly*, 51(1-3), 6–24. doi:10.1080/01639374.2012.715117

ADDITIONAL READING

Alemneh, D. G. (2014). Metadata for ETD Lifecycle Management. In M. Schultz and K. Skinner (Eds.) *Guidance Documents for Lifecycle Management of ETDs*. Educopia Institute. Retrieved November 23, 2016, from <http://digital.library.unt.edu/ark:/67531/metadc279711/>

Chandrasekaran, B., Josephson, J. R., & Benjamins, R. (1999). What are ontologies, and why do we need them? *IEEE Intelligent Systems*, 14(1), 20–26. <https://www.csee.umbc.edu/courses/771/papers/chandrasekaranetal99.pdf> Retrieved November 23, 2016 doi:10.1109/5254.747902

Hartsock, R., & Alemneh, D. G. (2016). *Beyond Text: Best Practices for Cataloging Music-ETDs and Associated Audio-Visual Materials*. Retrieved November 23, 2016, from <http://digital.library.unt.edu/ark:/67531/metadc848622/>

Library of Congress. Cataloging and Acquisitions Home: Home page of the Acquisitions and Bibliographic Access Directorate. November 23, 2016, from <https://www.loc.gov/aba/>

Library of Congress. Genre/Form Headings at the Library of Congress. Retrieved November 23, 2016, from <http://www.loc.gov/catdir/cpsol/genreformgeneral.html>

Potvin, S.,.... (2015). *Texas Digital Library Descriptive Metadata Guidelines for Electronic Theses and Dissertations, Version 2.0*. Retrieved November 23, 2016, from <https://tdl-ir.tdl.org/tdl-ir/handle/2249.1/68437>

KEY TERMS AND DEFINITIONS

Catalog: Textual representation of bibliographic information for materials, either in analog (i.e., card catalog) or digital (i.e., online catalog) forms; online catalogs employ natural language labels, each derived from MARC (Machine Readable Cataloging) numerical data.

Digital Curation: The active management, preservation, and enrichment of digital resources.

ETD: The term “Electronic Theses and Dissertations” (ETD) is used primarily to differentiate between analog theses and dissertations (paper, microfilm) and their digital counterparts (digital objects). However, “ETDs” may also be digitized or born digital.

Intellectual Property Rights: Refers to creations of the mind: inventions; literary and artistic works and symbols, names, images, and designs used in commerce.

LCGFT: Library of Congress Genre Form Thesaurus (LCGFT) is a controlled vocabulary of genres and forms maintained by the United States Library of Congress, for use in bibliographic and authority records.

LCMPT: Library of Congress Medium of Performance Thesaurus (LCMPT) is a controlled vocabulary maintained by the United States Library of Congress, for use in bibliographic

and authority records. Terms provide maximum granularity of the medium of performance.

LCSH: Library of Congress Subject Headings (LCSH) is a controlled vocabulary maintained by the United States Library of Congress, for use in bibliographic records.

MARC: Machine-Readable Cataloging (MARC), a data retrieval system developed during the 1960s at the Library of Congress. It employed three character numerical codes, which utilized 8 bit memory as opposed to natural language labels.

OA: Open Access (OA) refers to online research outputs freely accessible, without restrictions on costs (e.g. access tolls) or use (e.g., copyright, performance rights, and licensing).

OAI-PMH: Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) is a mechanism developed for harvesting metadata descriptions of records in an archive.

RDA: Resource Description and Access (RDA) is a standard for descriptive cataloging initially released in 2010, providing instructions and guidelines on formulating bibliographic data. RDA divides bibliographic entities into works, expressions, manifestations, and items.

RDF: Resource Description Framework (RDF) is a standard model for data interchange on the Web, facilitates data merging across several schemas.

The Nature of Research Methodologies

Ben Tran

Alliant International University, USA

INTRODUCTION

In the research community, according to Tran (2015), the process of research, is generally defined as the procedures used in research that involve introducing a problem, narrowing the research problem into purpose statements, research questions, and hypotheses, collecting and analyzing data to address these questions and hypotheses, and using a writing structure that best fits the problem and the methods. Hence, terms such as *investigator* (often associated with quantitative research) and *inquirer* (often associated with qualitative research) are used interchangeably. Thus, a *methodology* refers to the philosophical framework and the fundamental assumptions of research (van Manen, 1990), *research design* refers to the plan of action that links the philosophical assumptions to specific methods (Creswell, 2003; Crotty, 1998), and *methods* are techniques of data collection and analysis (Creswell, 2003; van Manen, 1990). Some mixed methods writers, like Tashakkori and Teddlie (1998), consider this form of research a methodology and focus on the philosophical assumptions. To call a process a *methodology* introduces a complexity to the process of research. Other mixed methods writers, like Creswell, Plano-Clark, Guttman, and Hanson (2003), Greene, Caraceli, and Graham (1989), and Onwuegbuzie and Teddlie (2003), emphasize the techniques or methods of collecting and analyzing data. To call mixed methods research a *method* is clean and concise and resonates with many researchers (Elliott, 2005). The purpose of this chapter is to cover the three types (trends) of research methodologies: the traditional (quantitative, qualitative), the universal (mixed-methods),

and the trends (blogs, webinars, virtual intercepts, and virtual reality). This chapter will also cover a brief history of research methods and the usage of research methodologies.

BRIEF HISTORY OF RESEARCH METHODOLOGIES

Before the advent of mixed methods, many studies used multiple methods to achieve the benefits of triangulation (Galton & Wilcocks, 1983) without restricting themselves to any paradigmatic membership or methodological category (Tashakkori & Teddlie, 2003). Thus, during the last 50 years, writers have used different names, making it difficult to locate articles that might relate to mixed methods research. Mixed methods has been called *multitrait/multimethod research* (Campbell & Fiske, 1959), *integrated* or *combined* (Johnson & Onwuegbuzie, 2004, p. 17; Steckler, McLeRoy, Goodman, Bird, & McCormick, 1992), and *quantitative and qualitative methods* (Fielding & Fielding, 1986). It has been called *hybrids* (Ragin, Nagel, & White, 2004), *methodological triangulation* (Morse, 1991a), *combined research* (Creswell, 1994), and *mixed methodology* (Tashakkori & Teddlie, 1998). It has also been called the *third methodological movement* (Tashakkori & Teddlie, 2002, p. 5), the *third research paradigm* (Johnson & Onwuegbuzie, 2004, p. 15), and *a new star in social science sky* (Mayring, 2007, p. 1). Nevertheless, the beginning of mixed methods is cited by some (Creswell & Plano-Clark, 2007, p. 5; Johnson, Onwuegbuzie, & Turner, 2007) to Campbell and Fiske (1959) as *multitrait* of *multimethod* research, a concept later formalized by

Webb, Campbell, Schwartz, and Sechrest (1966) as *triangulation* (Greene, Caracelli, & Graham, 1989), and is often cited as having methodological superiority over single methods (Johnson et al., 2007; Tran, 2014a). For the first 60 years or so of the 20th century, mixed research can be seen in the work of cultural anthropologists and, especially, the fieldwork sociologists (Gans, 1963; Hollingshead, 1949; Jahoda, Lazarsfeld, & Zeisel, 1931/2003; Lynd & Lynd, 1929/1959).

In social science methodological literature, Campbell and Fiske's (1959) article introduced the idea of triangulation, referring to *multiple operationalism* (Bouchard, 1976). Today, the most frequently used name is *mixed methods research*, a name associated with the *Handbook of Mixed Methods in Social and Behavioral Research* (Tashakkori & Teddlie, 2003). Furthermore, early researchers' idea of multiple operationalism follows more closely what today is called *multimethod research*, in contrast to what currently is called *mixed methods research*. However, Campbell and Fisk (1959) are rightfully credited as being the first to show explicitly how to use multiple research methods for validation purposes, and were extended further by Webb, Campbell, Schwartz, and Sechrest (1966). Thus, Webb et al. are credited with being the first to coin the term *triangulation*. With that said, Cook (1985) is credited for coining the term critical multiplism (also see Houts, Cook, & Shadish, 1986).

THE USAGE OF RESEARCH METHODOLOGIES

In the usage of research methodologies, it was Denzin (1978, p. 291) who first outlined how to triangulate methods. Denzin outlined the following four types of triangulation: (1) data triangulation, (2) investigator triangulation, (3) theory triangulation, and (4) methodological triangulation (Tran, 2014a). According to Tran (2014a), Denzin also distinguished *within-methods* triangulation, from

between-methods triangulation (Tran, 2014a). According to Denzin, three outcomes arise from triangulation: convergence, inconsistency, and contradiction. Furthermore, Jick (1979) noted the following advantages of triangulation: (1) it allows researchers to be more confident of their results, (2) it stimulates the development of creative ways of collecting data, (3) it can lead to thicker, richer data, (4) it can lead to the synthesis or integration of theories, (5) it can uncover contradictions, and (6) by virtue of its comprehensiveness, it may serve as the litmus test for competing theories (Tran, 2014a). Morse (1991b) on the other hand, outlined two types of methodological triangulation: simultaneous or sequential.

Sieber (1973) provided a list of reasons to combine quantitative and qualitative research data that can play a role in providing baseline information and helping to avoid *elite bias*. Rossman and Wilson (1985) identified three reasons for combining quantitative and qualitative research: (1) used to enable confirmation or corroboration of each other through triangulation, (2) used to enable or develop analysis in order to provide richer data, and (3) used to initiate new modes of thinking by attending to paradoxes that emerge from the two data sources. By examining published research, Greene, Caracelli, and Graham (1989) inductively identified the following five broad purposes of rationales of mixed methodological studies: (1) triangulation, (2) complementarity, (3) development, (4) initiation, and (5) expansion. Sechrest and Sidana (1995) listed four reasons for methodological pluralism: (1) for verification purposes, (2) to provide some basis for estimating possible error in the underlying measures, (3) to facilitate the monitoring of data collected, and (4) to probe a data set to determine its meaning. Also, Dzurec and Abraham (1993) identified the following six *pursuits* that link qualitative and quantitative research: (1) the pursuit of mastery over self and the world, (2) the pursuit of understanding through recomposition, (3) the pursuit of complexity reduction to enhance understanding,

(4) the pursuit of innovation, (5) the pursuit of meaningfulness, and (6) the pursuit of truthfulness (pp. 76-77).

Furthermore, Collins, Onwuegbuzie, and Sutton (2006) identified four rationales for conducting mixed research: participant enrichment, instrument fidelity, treatment integrity, and significance enhancement. In sum, the 20th century, according to Tran (2015), started with some use of what later came to be called mixed research, but social and psychological research quickly became primarily quantitative and coalescing into a qualitative research paradigm in the 1980s and 1990s (Guba, 1990). In reaction to the polarization between quantitative and qualitative research, another intellectual movement occurred, and it has come to be called mixed methods research, all thriving and coexisting. In contrast to Thomas Kuhn's (1962) expectation for single paradigms characterizing *normal science*, the research community suggest that a three-paradigm *methodological* world might be healthy because each approach has its strengths and weaknesses and times and places of need.

Research Methodologies: Traditional (Quantitative and Qualitative)

Quantitative research options have been predetermined and a large number of respondents are involved. By definition, measurement must be objective, quantitative and statistically valid. Simply put, it is about numbers, and objective hard data. The sample size for a survey is calculated by statisticians using formulas to determine how large a sample size will be needed from a given population in order to achieve findings with an acceptable degree of accuracy. Generally, researchers seek sample sizes which yield findings with at least a 95% confidence interval, plus/minus a margin error of 5% points. Hence, quantitative data are divided into two main categories, descriptive and continuous (Tran, 2014a, 2015).

According to Tran (2014a, 2015), qualitative research is collecting, analyzing, and interpreting data by observing what people do and say. Where-

as, quantitative research refers to counts and measures of things, qualitative research refers to the meanings, concepts, definitions, characteristics, metaphors, symbols, and descriptions of things. Qualitative research is much more subjective than quantitative research and uses very different methods of collecting information, mainly individual, in-depth interviews, and focus groups. The nature of this type of research is exploratory and open-ended. Qualitative data consists of open-ended information that the researcher gathers through interviews with participants. Hence, qualitative data are divided into two main categories, nominal and ordinal.

Research Methodologies: Universal (Mixed Methods)

Mixed methods research has been established as a third methodological movement over the past twenty years, complementing the existing traditions of quantitative and qualitative movements (Tashakkori & Teddlie, 2003; Teddlie & Tashakkori, 2009). The term *mixed methods* has come to be used to refer to the use of two or more methods in a research project yielding both qualitative and quantitative data (Creswell & Plano-Clark, 2007; Greene, 2007; Teddlie & Tashakkori, 2009). Multimethods do not have the same paradigmatic problem as do mixed methods since they can adopt the paradigm appropriate to the single type of data being collected. The paradigm problem for mixed methods arises because of the so called paradigm wars of the 1970s and 1980s where the positivist paradigm of quantitative research came under attack from social scientists supporting qualitative research and proposing constructivism, or variants thereof, as an alternative paradigm (Guba & Lincoln, 1994; Reichhardt & Rallis, 1994). To deal with this problem a range of alternative approaches have been developed (Creswell & Plano-Clark, 2007; Tashakkori & Teddlie, 2003). These approaches, according to Hall (2013), can be classified into three basic categories: a-paradigm stance, multiple paradigm

approach, and the single paradigm approach. The first of these simply ignores paradigmatic issues altogether, the second asserts that alternative paradigms are not incompatible and can be used in the one research project, and the third claims that both quantitative and qualitative research can be accommodated under a single paradigm. A paradigm may be viewed as a set of *basic beliefs* that deals with ultimate or first principles. It represents a *worldview* that defines, for its holder, the nature of the *world*, the individual's place in it, and the range of possible relationships to that world and its parts (Guba & Lincoln, 1994).

Mixed Methods Approaches

Research studies are becoming increasingly diverse and inclusive of both quantitative and qualitative methods—that is, they are mixing methods to address specific objectives (Tran, 2014a, 2015). The basic premise behind using a mixed methods research design is that *the combined* of both approaches provides a better understanding of a research problem than either approach could alone. Creswell and Plano-Clark (2011) argue that integrating methodological approaches strengthens the overall research design, as the strengths of one approach offset the weaknesses of the other, and can provide more comprehensive and convincing evidence than mono-method studies. Another more practical benefit is that mixed method research can encourage interdisciplinary collaboration and the use of multiple paradigms. There are more than a dozen mixed methods research typologies in the literature, each emphasizing different angles. For the most part, however, typologies include at the very least two basic dimensions—timing of data integration and purpose of integration (Guest et al., 2012). The two most commonly used terms in this regard are *sequential* and *concurrent* designs (Creswell & Plano-Clark, 2007; Morgan, 1998; Morse, 1991b).

Mixing the Data: Quantitative and Qualitative

R

The mixing of data is a unique aspect of the research community when it comes to definition. By mixing the datasets, the researcher provides a better understanding of the problem than if either dataset had been used alone. There are three ways in which mixing occurs: merging or converging the two datasets by actually bringing them together, connecting the two datasets by having one build on the other, or embedding one dataset within the other so that one type of data provides a supportive role for the other dataset. Thus, it is not enough to simply collect and analyze quantitative and qualitative data, they need to be *mixed* in some way so that together they form a more complete picture of the problem than they so when standing alone (Tran, 2015).

FUTURE RESEARCH DIRECTIONS: RESEARCH METHODOLOGIES: TRENDS (BLOGS, WEBINARS, VIRTUAL INTERCEPTS, AND VIRTUAL REALITY)

Design decisions depend on the purposes of the study, the nature of the problem, and the alternatives appropriate for its investigation. Once the purposes have been specified, the study should have explicit scope and direction, and attention can be focused on a delimited target area. Design alternatives can be organized into nine functional categories based on these problem characteristics (Isaac & Michael, 1995, pp. 45-47): historical, descriptive, developmental, case or field, correlational, causal-comparative, true experimental, quasi-experimental, and action. Traditional qualitative research methods, like study groups, polls and observational studies, are going digital and expanding the ability for researchers and busi-

nesses to target participants and collect information. The following are four types of qualitative trends: virtual intercepts, virtual reality (Bryson, N. Y.), blogs, and webinars.

- **Blogs:** A blog is a personal internal journal, a discussion, or an informational site published on the World Wide Web and consisting of discrete entries typically displayed in reverse chronological order.
- **Webinars:** Web conferencing may be used as an umbrella term for various types of online collaborative services including web seminars (or webinars), webcasts, and peer-level web meetings. It may also be used in a more narrow sense to refer only to the peer-level web meeting context, in an attempt to disambiguate it from the other types of collaborative sessions.
- **Virtual Intercepts:** A web intercept (survey) is a survey which is used by websites to get feedback from users and visitors about the website. Such surveys are used to collect information regarding the website and how helpful it is for the users. A web intercept survey is an easy and quick way to intercept a website's performance consisting of a few main questions which are related to the topics directly and are usually objective in nature. Researchers can intercept respondents online from social media or surveys and route them into a one-on-one interview.
- **Virtual Reality** A technology-based Tavistock Method (Tran, 2014b) for research that harnesses virtual and augmented reality to enhance the research experience—for participant and client alike. Virtual reality captures new data, creates new experiences, and gives new insight by creating amazing reproductions of the physical world. With virtual reality technology, “where” becomes “anywhere”. Researchers can bring more people together—in virtual environments—than ever be-

fore. Researchers can bring people together from all across the country, or the world, and capture fresh data and experience it all “through the eyes” of the participants.

CONCLUSION

According to Tran (2015), debates about singular or universal truths or approaches to viewing the world (Socrates, Plato), versus multiple or relative truths (the Sophists such as Protagoras and Gorgias), versus balances or mixtures of the extremes (Aristotle's “golden mean” or principle of balance, moderate skepticism, Cicero, Sextus Empiricus), go back, at least, to ancient Western philosophy, and the spirit of these debates lives today in the different views of the three major approaches to social research. This debate continues to affect how we view knowledge, what we look for, what we expect to find, and how we believe we are to go about finding and justifying *knowledge*. Mixed research is between the extremes Plato (quantitative research) and the Sophists (qualitative research), with mixed research attempting to achieve a workable middle solution for many (research) problems of interest. Hence, research paradigms address the philosophical dimensions of social sciences (Tran, 2015). A research paradigm is a set of fundamental assumptions and beliefs as to how the world is perceived which then serves as a thinking framework that guides the behavior of the researcher (Jonker & Pennink, 2010).

Some writers (e.g. Berry & Otley, 2004; Creswell, 2009; Neuman, 2011; Saunders, Lewis, & Thornhill, 2009) emphasizes that it is important to initially question the research paradigm to be applied in conducting research because it substantially influences how one undertake a social study from the way of framing and understanding social phenomena. According to Arbnor and Bjerke (2008), Guba (1990), Guba and Lincoln (1994), and Wahyuni (2012), the two main philosophical dimensions to distinguish existing research paradigms are ontology and epistemology (Kalof,

Dan, & Dietz, 2008; Laughlin, 1995; Saunders, Lewis, & Thornhill, 2009), whereas only Arbner and Bjerke (2008), Guba (1990), and Guba and Lincoln (1994) claim that is a *third paradigm*: methodology. As such, quantitative research and quantitative research data are static through time, compared to qualitative and qualitative data, but still have functional uses and are relevant in the digital world. According to Bryson (N. Y.), deliberative and participative research methods are gaining popularity due to their increased ability to discover information on a larger scale. For instance, the largest growth is found in social media and mobile market research, and such technologies ranges from high-definition video conferencing and instant communication around the world to the ability to reach participants on their mobile devices and access to demographics that are traditionally hard to reach, the Internet is providing technology based research methods like blogs, webinars, virtual intercepts, and virtual reality. According to Bryson (N. Y.), three of the five trends gaining traction in digital qualitative research are: (1) market research online communities, (2) social media and qualitative research, and (3) mobile ethnography.

REFERENCES

- Arbner, I., & Bjerke, B. (2008). *Methodology for creating business knowledge*. London: Sage Publications.
- Berry, A. J., & Otley, D. T. (2004). Case-based research in accounting. In C. Humphrey & B. Lee (Eds.), *The real life guide to accounting research: A behind-the-scenes view of using qualitative research methods* (pp. 231–256). Oxford, UK: Elsevier. doi:10.1016/B978-008043972-3/50016-5
- Bryson, J. (n.d.). *Current and emerging trends in qualitative market research*. Retrieved on March 4, 2015. Available at <http://qualblog.com/current-and-emerging-trends-in-qualitative-market-research/>
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81–105. doi:10.1037/h0046016 PMID:13634291
- Cook, T. D. (1985). Postpositivist critical multiplism. In L. Shotland & M. M. Mark (Eds.), *Social science and social policy* (pp. 21–62). Beverly Hills, CA: Sage Publications.
- Creswell, J. W. (1994). *Research design: Qualitative and quantitative approaches*. Thousand Oaks, CA: Sage Publications.
- Creswell, J. W. (2003). *Research design: Qualitative, quantitative, and mixed methods approaches* (2nd ed.). Thousand Oaks, CA: Sage Publications.
- Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches* (3rd ed.). Thousand Oaks, CA: Sage Publications.
- Creswell, J. W., & Plano-Clark, V. L. (2007). *Designing and conducting mixed methods research*. Thousand Oaks, California, London: Sage Publications.
- Creswell, J. W., & Plano-Clark, V. L. (2011). *Designing and conducting mixed methods research*. Thousand Oaks, CA: Sage Publications.
- Creswell, J. W., Plano-Clark, V. L., Guttman, M. L., & Hanson, E. E. (2003). Advanced mixed methods research design. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 209–240). Thousand Oaks, CA: Sage Publications.
- Crotty, M. (1998). *The foundations of social research*. Thousand Oaks, CA: Sage Publications.
- Denzin, N. K. (1978). *The research act: A theoretical introduction to sociological methods*. New York: Praeger.

- Dzurec, L. C., & Abraham, J. L. (1993). The nature of inquiry: Linking quantitative and qualitative research. *ANS. Advances in Nursing Science*, 16(1), 73–79. doi:10.1097/00012272-199309000-00009 PMID:8311427
- Elliott, J. (2005). *Using narrative in social research: Qualitative and quantitative approaches*. Thousand Oaks, CA: Sage Publications. doi:10.4135/9780857020246
- Fielding, N. G., & Fielding, J. L. (1986). *Linking data: Sage university paper series on qualitative research methods*. Beverly Hills, CA: Sage Publications. doi:10.4135/9781412984775
- Galton, M., & Wilcocks, J. (1983). *Moving from the primary school*. London: Routledge and Kegan Paul.
- Gans, H. J. (1963). *Urban villagers: Group life and class in the life of Italian-Americans*. New York: Free Press.
- Green, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis*, 11(3), 255–274. doi:10.3102/01623737011003255
- Greene, J. C. (2007). *Mixed methods in social inquiry*. San Francisco, CA: John Wiley & Sons.
- Guba, E. G. (1990). *The paradigm dialog*. Newbury Park, CA: Sage Publications.
- Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. In N. K. Dezin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 105–117). Thousand Oaks, CA: Sage Publications.
- Guest, G., MacQueen, K., & Namey, E. (2012). *Applied thematic analysis*. Thousand Oaks, CA: Sage Publications. doi:10.4135/9781483384436
- Hall, R. F. (2013). Mixed methods: In search of a paradigm. In T. Le & Q. Le (Eds.), *Conducting research in a changing and challenging world* (pp. 71–78). Nova Science Publishers, Inc.
- Hollingshead, A. B. (1949). *Elmtown's youth*. New York: John Wiley.
- Houts, A. C., Cook, T. D., & Shadish, W. R. (1986). The person-situation debate: A critical multiplist perspective. *Journal of Personality*, 54(1), 52–105. doi:10.1111/j.1467-6494.1986.tb00390.x
- Isaac, S., & Michael, W. B. (1995). *Handbook in research and evaluation: A collection of principles, methods, and strategies useful in the planning, design, and evaluation of studies in education and the behavioral sciences* (3rd ed.). San Diego, CA: Educational and Industrial Testing Services.
- Jahoda, M., Lazarsfeld, P. F., & Zeisel, H. (2003). *Marienthal: The sociography of an unemployed community*. New Brunswick, NJ: Transaction Publishers. (Original work published 1931)
- Jick, T. D. (1979). Mixing qualitative and quantitative methods: Triangulation in action. *Administrative Science Quarterly*, 24(4), 602–611. doi:10.2307/2392366
- Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, 33(7), 14–26. doi:10.3102/0013189X033007014
- Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research*, 1(2), 112–133. doi:10.1177/1558689806298224
- Jonker, J., & Pennink, B. (2010). *The essence of research methodology: A concise guide for master and PhD students in management science*. Heidelberg, Germany: Springer.
- Kalof, L., Dan, A., & Dietz, T. (2008). *Essential of social research*. New York: McGraw-Hill.

- Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago: University of Chicago Press.
- Laughlin, R. (1995). Empirical research in accounting: Alternative approaches and a case for middle-range thinking. *Accounting, Auditing & Accountability Journal*, 8(1), 63–87. doi:10.1108/09513579510146707
- Lynd, R. S., & Lynd, H. M. (1929/1959). *Middletown: A study in modern American culture*. Orlando, FL: Harcourt Brace.
- Mayring, P. (2007). Introduction: Arguments for mixed methodology. In P. Mayring, G. L. Huber, L. Gurtler, & M. Kiegelmann (Eds.), *Mixed methodology in psychological research* (pp. 1–4). Rotterdam: Sense Publishers.
- Morgan, D. L. (1998). Practical strategies for combining qualitative and quantitative methods: Applications to health research. *Qualitative Health Research*, 8(3), 362–376. doi:10.1177/104973239800800307 PMID:10558337
- Morse, J. M. (1991a). Strategies for sampling. In J. M. Morse (Ed.), *Qualitative nursing research: A contemporary dialogue* (pp. 127–145). Newbury Park, CA: Sage Publications. doi:10.4135/9781483349015.n16
- Morse, J. M. (1991b). Approaches to qualitative-quantitative methodological triangulation. *Nursing Research*, 40(2), 120–123. doi:10.1097/00006199-199103000-00014 PMID:2003072
- Neuman, W. L. (2011). *Social research methods: Qualitative and quantitative approaches* (7th ed.). Boston, MA: Pearson/Allyn and Bacon.
- Onwuegbuzie, A. J., & Teddlie, C. (2003). A framework for analyzing data in mixed methods research. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 351–383). Thousand Oaks, CA: Sage Publications.
- Ragin, C., Nagel, J., & White, P. (2004). Sociology program methodology, measurement & statistics program directorate for social, behavioral & economic sciences. *Workshop on Scientific Foundations of Qualitative Research*, Arlington, VA: National Science Foundation.
- Reichhardt, C. S., & Rallis, S. F. (1994). Qualitative and quantitative inquiries are not incompatible: A call for a new partnership. In C. S. Reichhardt & S. F. Rallies (Eds.), *The qualitative-quantitative debate: New perspectives*. San Francisco, CA: Jossey Bass. doi:10.1002/ev.1670
- Rossmann, G. B., & Wilson, B. L. (1985). Numbers and words: Combining quantitative and qualitative methods in a single large-scale evaluation study. *Evaluation Review*, 9(5), 627–643. doi:10.1177/0193841X8500900505
- Saunders, M., Lewis, P., & Thornhill, A. (2009). *Research methods for business students*. London: Pearson Education.
- Sechrest, L., & Sidana, S. (1995). Quantitative and qualitative methods: Is there an alternative? *Evaluation and Program Planning*, 18(1), 77–87. doi:10.1016/0149-7189(94)00051-X
- Sieber, S. D. (1973). The integration of fieldwork and survey methods. *American Journal of Sociology*, 78(6), 1335–1359. doi:10.1086/225467
- Steckler, A., McLeroy, K. R., Goodman, R. M., Bird, S. T., & McCormick, L. (1992). Toward integrating qualitative and quantitative methods: An introduction. *Health Education Quarterly*, 19(1), 1–8. doi:10.1177/109019819201900101 PMID:1568869
- Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches*. Thousand Oaks, CA: Sage Publications.
- Tashakkori, A., & Teddlie, C. (2002). *Handbook of mixed methods in social & behavioral research*. Sage Publications, Inc.

Tashakkori, A., & Teddlie, C. (2003). *Handbook of mixed methods in social and behavioral research*. Thousand Oaks, CA: Sage Publications.

Teddlie, C., & Tashakkori, A. (2009). *Foundations of mixed methods research*. Thousand Oaks, CA: Sage Publications.

Tran, B. (2014a). Triangulation in organizational research: Validating knowledge in human competence at work. In A. Takhar & A. Ghorbani (Eds.), *Market research methodologies: Multi-method and qualitative approaches* (pp. 93–117). Hershey, PA: Premier Reference Source/IGI Global.

Tran, B. (2014b). Rhetoric of play: Utilizing the gamer factor in selecting and training employees. In T. M. Connolly, L. Boyle, T. Hainey, G. Baxter, & P. Moreno-Ger (Eds.), *Psychology, pedagogy and assessment in serious games* (pp. 175–203). Hershey, PA: Premier Reference Source: Information Science Reference/IGI Global.

Tran, B. (2015). The nature of research methodologies: Terms and usage within quantitative, qualitative, and mixed methods. In J. E. Jones & L. B. Mette (Eds.), *Mixed methods research for improved scientific study*. Hershey, PA: IGI Global.

van Manen, M. (1990). *Research lived experience: Human science for an action sensitive pedagogy*. State University of New York Press.

Wahyuni, D. (2012). The research design maze: Understanding paradigms, cases, methods and methodologies. *Journal of Applied Management Accounting Research*, 10(1), 69–80.

Webb, E. J., Campbell, D. T., Schwartz, R. D., & Sechrest, L. (1966). *Unobtrusive measures*. Chicago: Rand McNally.

ADDITIONAL READING

Bloor, M. A., & Wood, F. A. (2006). *Keywords in qualitative methods: A vocabulary of research concepts. Limited*. London: Sage Publications. doi:10.4135/9781849209403

Creswell, J. W. (2008). *Research design: Qualitative, quantitative, and mixed methods approaches*. Thousand Oaks, CA: Sage Publications.

Davidson, J., & di Gregorio, S. (2011). Qualitative research and technology: In the midst of a revolution. In N. K. Denzin and S. Lincoln (4th Ed.), *Handbook of qualitative research* (pp. 627–643). Thousand Oaks, CA: Sage Publications.

Deshpande, R. (1983). Paradigms lost: On theory and method in research in marketing. *Journal of Marketing*, 47(4), 101–110. doi:10.2307/1251403

Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of Management Review*, 14(4), 532–550.

Flick, U. (2007). *Managing quality in qualitative research*. London: Sage Publications. doi:10.4135/9781849209441

Flick, U. (2014). *An introduction to qualitative research* (5th ed.). London: Sage Publications.

Flick, U., Garms-Homolová, V., Herrmann, W. J., Kuck, J., & Röhsch, G. (2012). I cant prescribe something just because someone asks for it ...: Using mixed methods in the framework of triangulation. *Journal of Mixed Methods Research*, 6(2), 97–110. doi:10.1177/1558689812437183

Given, L. M. (2008). *The sage encyclopedia of qualitative research method: Volumes 1 & 2*. Newbury Park, CA: Sage Publications. doi:10.4135/9781412963909

- Gobo, G. (2012). Globalizing methodology?: The encounter between local methodologies. *International Journal of Social Research Methodology*, 14(6), 417–437. doi:10.1080/13645579.2011.611379
- Hanson, D., & Grimmer, M. (2007). The mix of qualitative and quantitative research in major marketing journals, 1993–2002. *European Journal of Marketing*, 41(1/2), 58–70. doi:10.1108/03090560710718111
- Hennink, M. M., Hutter, I., & Bailey, A. (2011). *Qualitative research methods*. London: Sage Publications.
- Hsiung, P. C. (2012). *The globalization of qualitative research: Challenging anglo-american domination and local hegemonic discourse* [27 paragraphs]. Forum Qualitative Sozialforschung/Forum: Qualitative. *Social Research*, 13(1), 21. Available at <http://nbn-resolving.de/urn:nbn:de:0114-fqs1201216> Retrieved on September 20, 2014
- Layder, D. (1994). *Understanding social theory*. London, England: Sage Publications.
- Layder, D. (1998). *Sociological practice: Linking theory and social research*. London, England: Sage Publications. doi:10.4135/9781849209946
- Marsden, D., & Littler, D. (1996). Evaluating alternative research paradigms: A market-oriented framework. *Journal of Marketing Management*, 12(7), 645–655. doi:10.1080/0267257X.1996.9964442
- Maxwell, J. A., & Loomis, D. M. (2003). Mixed methods design: An alternative approach. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 241–272). Thousand Oaks, CA: Sage Publications.
- Maxwell, S. E., & Delaney, H. D. (1994). *Designing experiments and analyzing data*. Mahwah, NJ: Lawrence Erlbaum.
- McMullen, L. M. (2011). A discursive analysis of Teresa's protocol: Enhancing oneself, diminishing others. In F. J. Wertz, K. Charmaz, L. M. McMullen, R. Josselson, R. Anderson, & E. McSpadden (Eds.), *Five ways on doing qualitative analysis: Phenomenological psychology, grounded theory, discourse analysis, narrative research, and intuitive inquiry* (pp. 205–223). New York: Guilford.
- Nuttall, P., Shankar, A., Beverland, M. B., & Hooper, C. S. (2011). Mapping the unarticulated potential of qualitative research. *Journal of Advertising Research*, 51(1), 153–163. doi:10.2501/JAR-51-1-153-166
- Onwuegbuzie, A. J., & Leech, N. L. (2010). Generalization practices in qualitative research: A mixed methods case study. *Quality & Quantity*, 44(5), 881–892. doi:10.1007/s11135-009-9241-z
- Ryan, A., & Gobo, G. (2011). Managing the decline of globalized methodology. *Special Issue on Perspectives on Decolonizing Methodologies. International Journal of Social Research Methodology*, 14(6), 411–415. doi:10.1080/13645579.2011.611378
- Schnettler, B., & Rebstein, B. (2012). International perspective on the future of qualitative research in Europe. *Qualitative Sociology Review*, 8(2), 6–11.

KEY TERMS AND DEFINITIONS

Between-Methods Triangulation: Involves the use of both quantitative and qualitative approaches.

Mixed-Methods Research: Is an approach to knowledge that attempts to consider multiple viewpoints, perspectives, positions, and standpoints.

Mixing Data: There are three ways in which mixing occurs: merging or converging the two datasets, connecting the two datasets by having one build on the other, or embedding one dataset within the other.

Paradigm (Research): Is a common body of beliefs, assumptions, and rules that govern research.

Qualitative Data: Consists of open-ended information that the researcher gathers through interviews with participants.

Qualitative Research: Is much more subjective than quantitative research and uses very different methods of collecting information, mainly individual, in-depth interviews, and focus groups.

Quantitative Data: Quantitative data includes closed-ended information such as that found on attitude, behavior, or performance instruments.

Quantitative Research: Refers to the meanings, concepts, definitions, characteristics, metaphors, symbols, and descriptions of things.

Triangulation: Is the combination of methodologies in the study of the same phenomenon. The four types of triangulation: (1) data triangulation, (2) investigator triangulation, (3) theory triangulation, and (4) methodological triangulation.

Within-Methods Triangulation: Refers to the use of either multiple quantitative or multiple qualitative approaches.

Research Methodology

R**Swati C. Jagdale***MAEER's Maharashtra Institute of Pharmacy, India***Rahul U. Hude***MAEER's Maharashtra Institute of Pharmacy, India***Aniruddha R. Chabukswar***MAEER's Maharashtra Institute of Pharmacy, India*

INTRODUCTION

Research is the process to find solution to a problem through the planned and systematic collection of data, analysis, verification and interpretation of data. Research is the very important process for accessing knowledge for promoting progress and to enable people to relate more effectively to his environment to accomplish his aim and to resolve his conflicts. The research is primarily carried out to discover new facts, to verify and test important facts, to analyze an event or process or phenomenon, to identify the cause and effect relationship, to develop new scientific tools and techniques, concepts and theories, to solve and understand scientific and nonscientific problems (Rajasekar, Philominathan and Chinnathambi, 2006). In daily life new problems, events, phenomena and processes occur every day. Practically, implementable solutions and suggestions are required for tackling new problems that arise. Scientists have to undertake research on them and find their causes, solutions, explanations and applications (Gogoi and Goowalla, 2015).

The term 'Research' consists of two words; Re and Search. 'Re' means again and again and 'Search' means to find out something (Pandey & Pandey, 2015).

According to Clifford Woody of the University of Michigan, "Research is a carefully inquiry or examination in seeking facts or principles; a diligent investigation to ascertain something."

According to C. Francies Rummel, "Research is an endeavour to discover, develop and verify knowledge. It is an intellectual process that has developed over hundreds of years, ever changing in purpose and form and always searching for truth" (Pandey & Pandey, 2015; Kothari, 2004; Singh, 2006).

BACKGROUND

In the 1600s the origin of modern scientific method occurred in Europe.

- **Copernicus:** A scientific model that could be verified and checked by observation.
- **Tycho Brahe:** Accurate instrumental observations to confirm the model.
- **Johannes Kepler:** Theoretical examination of experimental data.
- **Galileo Galilei:** Scientific laws developed from experiment.
- **Rene Descartes:** Mathematics to quantitatively show theoretical ideas.
- **Isaac Newton:** Theoretical derivation of an experimentally confirmable model.
- **Karl Popper:** Scientific theory should make predictions and can be tested and verified (Frederick, 2011; https://en.wikipedia.org/wiki/History_of_scientific_method).

OBJECTIVE OF RESEARCH

1. To gain knowledge with a phenomenon or to achieve new perceptions into it.
2. To draw accurately the characteristics of a particular situation, individual or a group.
3. To determine the time frame with which something occurs or with which it is associated with something else.
4. To test a theory of a causal relationship between variables that is to analyses process or phenomenon.
5. To discover new facts; verify and test important facts.
6. To develop new concepts, theories and scientific tools to solve and understand the problems.
7. To find answers to scientific, nonscientific and social problems and to overcome the problems occurring in everyday life (Gogoi and Goowalla, 2015; Bhawna, and Gobind 2015).
3. It requires expertise.
4. Research involves collecting new data or information from primary or first-hand sources or using existing information for a new motive.
5. It based upon observable experience or empirical proofs.
6. Research demands accurate systematic observation, description and accurate investigation.
7. It is characterized by carefully designed methods or plan that applies careful analysis.
8. It achieves to organize data in quantitative terms.
9. It generally requires inexpensive informational data.
10. It is based on mutually depends upon causes and effect.
11. It sometimes requires courage (Pandey and Pandey, 2015; Kumar, 2015).

CHARACTERISTICS OF RESEARCH

1. It is directed toward the solution of a problem.
2. It gives special importance to the development of generalizations, principles, or theories that will be helpful in predicting future events.

TYPES OF RESEARCH

The research is broadly classified into two main categories as Fundamental or Basic and Action or Applied. Various ways through which research may classify is summarizes in Figure 1 and Table 1.

Figure 1. Classification of research

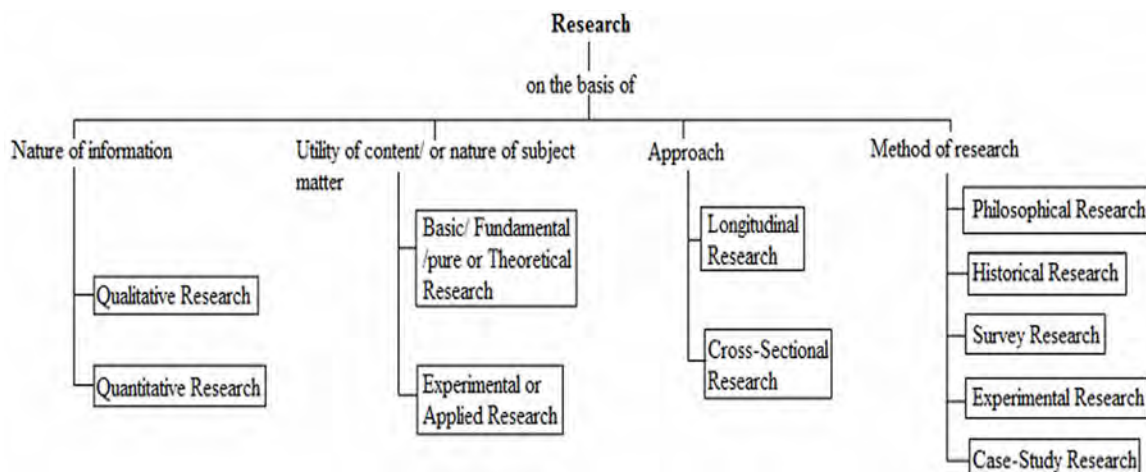


Table 1. Classification of research

On the Basis	Types of Research	Explanations
Levels	Basic level	Basic level as basic research. It does not necessarily produce results of immediate practical value and is designed to add an organized framework of scientific knowledge.
	Applied level	Helps to solve an immediate practical problem and the goal of adding scientific knowledge is secondary.
Objectives of Research	Fundamental	These research are applicable to all cases.
	Action	Applicable for certain conditions so it's utility is limited.
Approach of Research	Longitudinal	Case study, historical research, genetic comes under this type of research.
	Cross sectional	Survey and experimental research are the examples of cross sectional research.
Precision in Research Findings	Experimental	Such research are precise.
	Non experimental	Non-experimental research is not precise.
Nature of Findings	Explanatory	Such researches explain more concerned laws, principles and theories.
	Descriptive	These research are more concerned with the facts.
National Science Foundation	Basic	Those researches which accept origin or unique inspection for the advancement of knowledge.
	Applied	It may be characterized as the usefulness in practice
	Development	It uses scientific knowledge for the production of useful devices, systems, methods, materials for processes excluding design and production engineering.
Another Classification	Adhoc	This research is the class of inquiry used for a purpose alone and special.
	Empirical	Depends upon the observation or experience of phenomena and incident.
	Explained	Explained research is based on a theory.
	Boarder line research	In boarder line research involves main two branches or areas of science. For example study of public school finance.

1. Basic Research

Investigation or research studies of some natural phenomenon or relating to pure science are called as basic research or theoretical research. It is a finding solution on basic principles and reasons for occurrence of a particular process or phenomenon. Such researches some time may not lead to immediate application or use; not concerned with finding solution of any practical problems of immediate interest.

Many applied research uses base that comes from outcomes of basic research. Research concerning some natural phenomenon or process or relating to pure mathematics, research studies deals with human behavior studied with a view to make generalizations about human behavior and elementary particles results in identification

of new particles are the examples of fundamental research.

Basic research is an attempt to find answers to the following questions.

- Why materials are look like that?
- How does sodium melt?
- Why is sound produced when liquid is heated? (Pandey and Pandey, 2015; Patil and Mankar, 2016; http://www.mech.hku.hk/bse/bbse3002/Research_Methodology.pdf).

2. Applied Research

An applied research involves solving certain problems or issues by employing well known and accepted theories and principles. Most of

the case studies, experimental study, and interdisciplinary research are mainly applied research. Aim of Applied research is finding a solution or conclusion for an immediate problem facing a society or organization.

For example, research having immediate potential applications such as increasing efficiency of a production machine, increasing gain factor of production of a material, pollution control and preparing vaccination for a disease etc. (Pandey and Pandey, 2015; Patil and Mankar, 2016; http://www.mech.hku.hk/bse/bbse3002/Research_Methodology.pdf).

3. Experimental Development

Experimental development is systematic approach, knowledge gained from research and empirical experience, which is directed to producing new products, services and devices; to installing new systems, procedures, processes, and services; or to improving significantly those already produced or installed.

For examples, study of a polymerization reactions of given class under various reaction conditions; includes experimental development as scaling up process which has been optimized at the laboratory level and investigating and evaluating possible reaction methods for producing the polymer (Fracati, 2002).

4. Quantitative and Qualitative Methods

The applied and basic researches can be quantitative or qualitative or even both.

- **Quantitative Research/Method:** Is mainly based on the quantification or measurement of quantity or amount. Quantitative research is numerical, non-descriptive and applies statistics or mathematics. The results are conclusive often presented in tables and graphs. It also investigates the what, where and when of decision making.

It finds applications not only in mathematics and physical sciences but also in social sciences, economics and biology.

- **Qualitative Research/Method:** Is deals with qualitative phenomenon involving or relating to quality. Qualitative research are descriptive, non numerical, applies reasoning and uses words; data cannot be graphed. It aims to get the meaning and describe the situation. It is exploratory and investigates the why and how of decision making. Investigation of the reasons for human behavior for example why people think or do certain things? To find why certain data are random then it is a qualitative research. If the aim is to investigate how random the data is, what is the mean, mode, variance and distribution function then it becomes quantitative. Explaining how absorption of drug takes place in our body is a qualitative description. It does not involve any numbers, quantities or data. While at how much or percentage drug is absorbed to blood; involves measurement of quantity hence it becomes quantitative Rajasekar, Philominathan and Chinnathambi, 2006; Kothari, 2004; Patil and Mankar, 2016; Borrego, Douglas and Amelink, 2009; Bernhard and Baillie, 2013; Choy, 2014).

RESEARCH APPROACHES

The generation of data in quantitative form which can be subjected carefully to quantitative analysis in a formal and fixed fashion is a quantitative approach. This approach can be further sub categorized into inferential, experimental and simulation approaches to research. Inferential approach leads to research that forms a data base from which to infer characteristics or relationships of population. This usually survey type research where a sample of population is studied (questioned or observed) to determine its features, and then to conclude that the population has the same features or charac-

teristics. Qualitative approach is concerned with subjective evaluation of opinions, attitudes and behavior. Research in such a situation is a function of researcher's perception and ideas or feeling. This generates results either in non quantitative form or in the form which are not subjected to accurate quantitative analysis.

Experimental approach is characterized by much greater control over the research variables and in this case some variables are manipulated to observe their effect and relation on other variables.

Simulation approach involves the setting up of an artificial environment within which relevant information or facts and data can be generated. This allows an observation of the dynamic behavior of a system under supervise state or controlled situations. (Kothari, 2004; Bhawna and Gobind, 2015).

RESEARCH PROCESS

Research process consists of series of steps as-

1. Formulation of Research Problem

At the very beginning, the researcher must decide the general area or field of interest or aspect of a subject matter that he would like to investigate or inquire into and then research problem should be formulated or prepared.

2. Extensive Literature Survey

For this purpose, the abstracting and indexing journals and published or unpublished bibliographies, academic journals, conference proceedings, government reports, books and internet search engine etc. must be utilized depending on the nature of the problem.

3. Development of Working Hypothesis

Working hypothesis is tentative postulations made in order to draw out and test its logical or empirical results.

4. Preparing the Research Design

After preparing hypothesis, research design or model (conceptual structure within which research would be conducted) is prepared. The framing of such a model helps to research to be as efficient as possible yielding maximal informative data. Research design provides collective relevant proof with ideal effort, time and expenditure.

5. Determining Sample Design

A designed sample plan is decided before any data or information is actually collected for obtaining a sample from a given population. Survey involves a great deal of time, money and energy so it is impossible in practice under many situations.

6. Collecting the Data

Primary data can be collected either through experiment work or through detail survey. In case of survey, data can be collected by any one or more ways like by observation, through personal interview, through telephonic interviews, by mailing of questionnaires or through schedules.

7. Execution of the Project

If the execution of the project proceeds on correct proper lines, the data to be collected would be sufficient and dependable.

8. Analysis of Data

The analysis of data has carried by a number of closely related operations such as establishment of categories coding, tabulation (based on the mathematical calculation of various percentages, standard deviations, coefficients etc.) and then drawing statistical inference.

9. Hypothesis Testing

After analyzing the data, the researcher should go to test the hypothesis, if any, he had prepared

earlier. Do the facts that support the hypothesis or they happen to be contrary? This is to be answered by applying various tests like 't' test, 'F' test, ANOVA etc. Testing will result in either acceptance or in rejection. If the researcher had no hypothesis or theory to start with, generalizations started on the basis of identified data.

10. Generalizations and Interpretation

If a hypothesis is tested and confirm several times, it may be possible for the researcher to reach at generalization that is to prepare a theory. As a matter of fact, the real value of research lies in its ability to reach at certain generalizations. If the researcher had no hypothesis to begin with, he might seek to explain his findings on the basis of some theory or facts. It is known as interpretation.

11. Preparation of the Report or the Thesis

Finally, the researcher has to prepare systematic report of what has been done by him (Rajasekar, Philominathan and Chinnathambi, 2006; Kothari, 2004; Gogoi and Goowalla, 2015; Pandey and Pandey, 2015; Kothari, 2004; Patil and Mankar, 2016).

IMPORTANCE OF RESEARCH

1. Research introduces scientific and rational new thinking, promotes the development of logical habits of thinking also governs the decisions of the policy maker.
2. Research has its special importance in solving various operational problems of business, society and industry.
3. Research helps for nearly all government policies in our economic system.
4. It helps in determination of motivations underlying the consumer (market) behavior.
5. It is important for social scientists in studying social relationships and in finding and solving various problems.

6. Research provides the intellectual satisfaction of knowing a few things just for the sake of knowledge; also practical use for the social peoples to know for the purpose of being able to do something better or in a more efficient manner than present.
7. To professionals research may mean a source of livelihood for philosophers and thinkers, research helps to outlet for new ideas, logics and insights.
8. Research is technique for the development of new styles and creative work makes it delightful and glorious.
9. Research is the base of knowledge to gain knowledge and an important source and guidelines for solving different types of problems in social, business and daily life (Rajasekar, Philominathan and Chinnathambi, 2006; Kothari, 2004).

RESEARCH METHODS AND RESEARCH METHODOLOGY

Research method includes all those methods or techniques that are used for conduction of research shown in Table 2 (Kothari, 2004). Research methods are essentially designed, scientifically formulated and value neutral. They include theoretical procedures, principles, hypothesis, experimental studies, numerical schemes and statistical approaches etc. Research methods help us collect samples, data and investigate a solution to an unsolved problem. Particularly, scientific research methods require explanations which are based on collected facts, information, experiments, measurements, observations and verification and not on reasoning alone. They accept only those explanations which can be confirmed by experiments.

Research methodology is a science of studying how research is to be carried out and is a systematic way to solve a problem. Various steps that are normally followed by a researcher in studying his/her research problem along with the scientific

Table 2. Method and techniques used to perform various researches (Kothari, 2004)

Types	Methods	Techniques
Library Research	1. Analysis of historical records	Recording of notes, Content analysis, Tape and Film listening analysis.
	2. Analysis of documents	Statistical compilations and manipulations, reference and abstract guides.
Field Research	1. Non-participant direct observation	Observational behavioral scales, use of score cards.
	2. Participant observation	Interactional recording, photo graphic techniques.
	3. Mass observation	Recording mass behavior, interview using independent observers.
	4. Mail questionnaire	Identification of social and economic background of respondents.
	5. Opinionnaire	Use of attitude scales, projective techniques, use of sociometric scales.
	6. Personal interview	Interviewer uses a detailed schedule with open and closed questions.
	7. Focused interview	Interviewer focuses attention upon a given experience and its effects.
	8. Group interview	Small groups of respondents are interviewed simultaneously.
	9. Telephone survey	Used as a survey technique for information and for discerning opinion; may also be used as a follow up of questionnaire.
	10. Case study and life history	Cross sectional collection of data for intensive analysis, longitudinal collection of data of intensive character.
Laboratory Research	Small group study of random behavior, play and role analysis	Use of audio-visual recording devices, use of observers, etc.

thought behind them. Thus it necessary for the researcher to know not only about the research methods/techniques but also the methodology. Researchers need to aware how to develop certain tests or indices, how to apply particular research techniques as well as they requires to know which of these methods or techniques are applicable and which are not. They should aware the criteria by which they can decide that certain methods, techniques and procedures will be applicable to solve certain problems and others will not. All this conclude that it is necessary for the researcher to design methodology for his/ her problem as the same may differ from problem to problem. (Rajasekar, Philominathan and Chinnathambi, 2006; Kothari, 2004; Gogoi and Goowalla, 2015; Kothari, 2004; Choy, 2014; <http://www.ais.utm.my/researchportal/files/2015/02/Example3-Res-Design.pdf>; Agbaje and Alarape, 2010; Khairul Baharein Mohd Noor, 2008).

CRITERIA OF GOOD RESEARCH

1. The aim should be clearly defined and common general concepts be used.
2. The procedure used should be described in sufficient detail to allow another researcher to repeat the research for further development, keeping the continuity of what has already been attained.
3. The designed procedure of the research should be carefully planned and thoroughly studied to obtain results that are as goal as possible.
4. The researcher should report with complete frankness, mark in procedural design and estimate their effects upon the investigating.
5. The analysis of data should be sufficiently satisfactory to show its importance and the methods of analysis used should be proper in the circumstances. The validity and reliability of the data should be checked and verified carefully.

6. Conclusions should be confined in area to those justified and explained by the data of the research and limited to those for which the data provide an adequate basis.
7. Greater confidence in research is justify if the researcher is experienced, has a good reputation in research and is a person of integrity.
8. In other words, we can express the good research should have characteristics such as systematic, logical, empirical and replicable (Pandey and Pandey, 2015; Kothari, 2004).

PROBLEMS ENCOUNTERED BY RESEARCHERS

1. The lack of a scientific training in the methodology of research is a great obstacle for researchers. There are small quantities of competent researchers. Research is mostly a scissor and paste job without any awareness shed on the collected materials.
2. There is insufficient or lack of interaction between the university research departments on one side and business establishments, government departments and research institutions on the other side, between researcher having same research interest.
3. Most of the business units do not have the confidence that the material supplied by them to researchers will not be misused and as such they are often resistant in supplying the needed information and data to researchers.
4. Research studies overlapping one another are undertaken completely often for want of sufficient information. This results in duplication and waste away resources.
5. There does not exist a code of conduct for researchers and inter university and inter-departmental competitions are also quite common.
6. Difficulty of sufficient and timely secretarial assistance, including computerial support. This leads to unnecessary delays in the completion and finalization of research studies.

7. Basic library management and functioning is not adequate at many places and much of the time and energy of researchers are spent in tracing out the books, journals, reports, and references etc. (Kothari, 2004).

FUTURE RESEARCH DIRECTIONS

Current Scenario of Research and Global Funding Forecast

The process of creating new products, processes and technologies that can be used and marketed for mankind's benefit in the future is termed as research and development (R&D). Asian countries continue to grow faster than other parts of the world. Because of this, combined Asian R&D investments are growing at a faster rate and their global R&D shares continue to increase at nearly 1% per year, while R&D shares of American and European decrease, even though their absolute R&D investments also continue to increase, but not at as fast rate as they do in Asia. Total global R&D share spending regional wise are shown in Table 3 (Global R&D funding forecast A supplement to R&D magazine, 2016).

The U.S. is largest single country in R&D investments with slightly more than a quarter of all global R&D spending. These U.S. R&D programs are supported and run by means of industrial (66%), federal government (25%) and academic/non-profit (7%) investments as shown in Figure 2 (Global R&D funding forecast A supplement to R&D magazine, 2016).

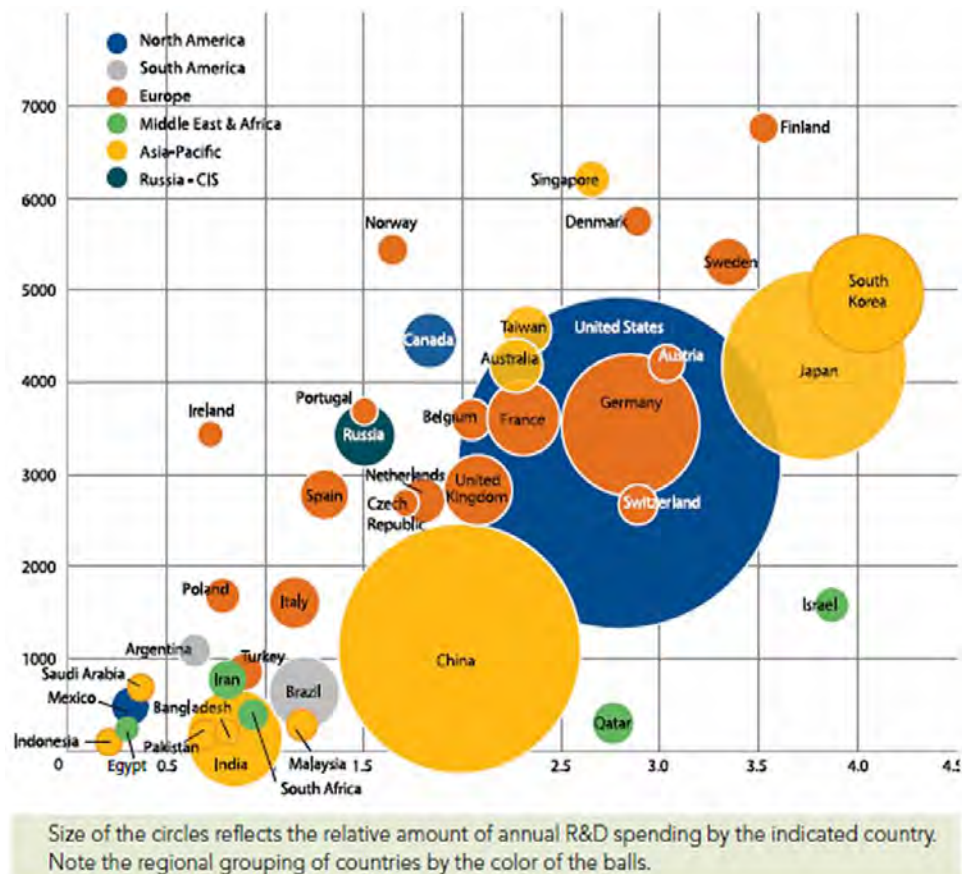
The global Life Science industry is one of the top large global high tech industries. It covers pharmaceuticals, biotechnology, medical devices and instruments, bioscience, veterinary, agricultural and commercial research and testing. However, most of the activities of these industries are operated in the biopharmaceutical sector by R&D, which reports for about 85% of the industries total R&D spending. Prediction shows a modest 1.8% increase in R&D spending for the global Life Science R&D marketplace in 2016 and a weak

Table 3. Share of total global R&D spending (Global R&D Funding Forecast A Supplement to R&D Magazine, 2016)

Region	Year		
	2014	2015	2016
North America	29.1%	28.5%	28.4%
U.S.	26.9%	26.4%	26.4%
Caribbean	0.1%	0.1%	0.1%
All North America	29.2%	28.5%	28.5%
Asia	40.2%	41.2%	41.8%
China	19.1%	19.8%	20.4%
Europe	21.5%	21.3%	21.0%
Russia	3.1%	2.9%	2.8%
South America	2.8%	2.6%	2.6%
Middle East	2.2%	2.3%	2.3%
Africa	1.0%	1.1%	1.1%
Total	100.0%	100.0%	100.0%

Figure 2. Country-wise share of R&D spending

Source: 2016 Global R&D funding forecast A supplement to R&D magazine



0.6% increase in U.S. R&D spending for 2016. (Global R&D funding forecast A supplement to R&D magazine, 2016).

Research on human being is very important for improvement of health and discovering new medicine and technology to treat diseases effectively. Current scenario of clinical trials revealed that there is a continuous increase in number of clinical research studies worldwide (about in 183 countries). Figure 3 shows registered numbers of clinical trials from the year 2007 to march 2015 on website www.clinicaltrials.gov.in. (Global R&D funding forecast A supplement to R&D magazine, 2016; Bangera, 2015).

World health organization (WHO), in U.S. National institute of health (NIH), National cancer institute (NCI) and Department of defense (DOD); in India Department of science & technology (DBT), Department of defense (DOD) and University of grant commission (UGC) etc are the research funding agency.

Information and Communication Technology (ICT) R&D are the backbone of the global digital economy and constitute a key driver of productivity growth in a knowledge-based economy and citizen's quality of life. It is clear fact that the ICT industry and ICT R&D innovation in non ICT industries and services make an important presence

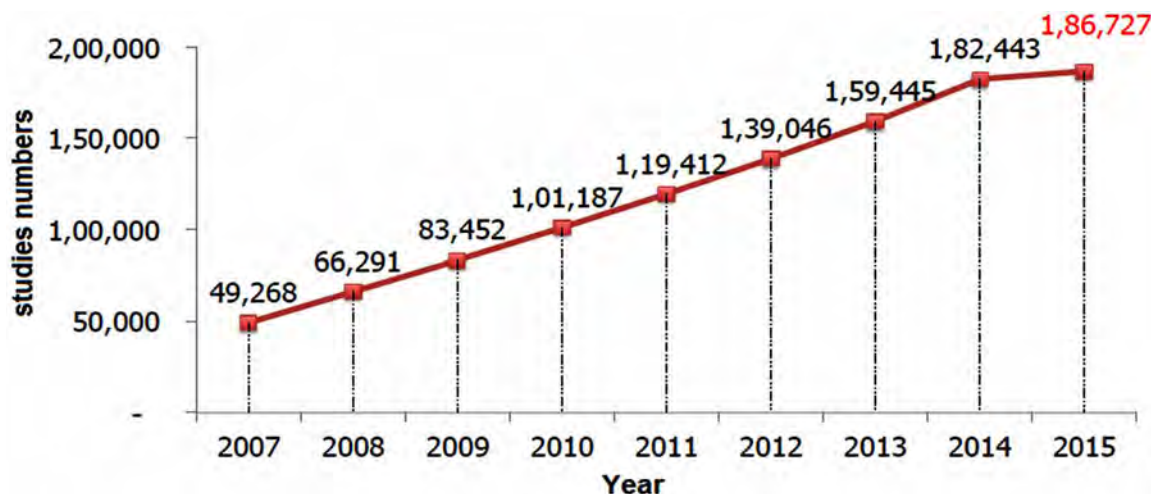
to the economic growth of advanced economies. In the EU, USA and Japan; the ICT field is the largest R&D investing sector of the economy. The EU ICT sector is a significant contributor to achieving the target of investing 3% of GDP in R&D. When comparing ICT spending over GDP, the USA, Japan, Taiwan and Korea are investing significantly more in ICT R&D than the EU. Supporting the speedy pace of revolution in ICT requires high levels of R&D. Actually, in terms of R&D expenditures budget, patents, and venture capital investments; the ICT sector exceeds all other industries by a large margin.

For stronger development in ICT a new strategy was started, primarily to foster the commercialization of research such as policy governance, science base, business R&D and innovation, knowledge flows and commercialization, human resources, green innovation, clusters and regional policies and globalization (Giuditta De Prato, Daniel and Juraj, 2011; Stephen and Scott 2010; OECD, 2014).

CONCLUSION

Research is a systematic process of gaining new knowledge. Main aim of research is to find out the truth which hidden or not yet discovered.

Figure 3. Number of clinical trial studies registered worldwide
Source: 2016 Global R&D funding forecast A supplement to R&D magazine



Research mainly classified into to two groups as basic and applied research. Research process starts from formulating research problem to final report through systematic and accurate investigation. All research Approaches investigates and explore the different assert to knowledge and are designed to label a specific type of research problem.

The methods and procedures used by a researcher during a research study are the research methods. Research methodology is a science and systematic way of studying how research is to be carried out. It mainly provide plan of work for research. Good research should be systematic, logical, empirical and replicable. In some part researcher are facing problems during research because of lack of scientific knowledge and interaction and funding facility. Thus research introduces scientific and rational new ideas and it promotes the development of logical habits of thinking. Research has its special importance in solving various operational and planning problems of business, society and industry.

REFERENCES

- Agbaje, A., & Alarape, A. I. (2010). *Introductory lectures on research methodology*. Retrieved June 10, 2016, from <http://www.ndc.gov.ng/Lectures/Research-Methodology.pdf>
- Bangera, S. (2015). *Current Clinical Research Scenario*. Retrieved June 10, 2016, from http://cdsaindia.in/sites/default/files/00_Current%20Clinical%20Research%20Scenario_Dr.%20Sudhakar.pdf
- Bernhard, J., & Baillie, C. (2013). Standards for quality of research in engineering education. *Proceeding of the Research in Engineering Education Symposium*.
- Bhawna, & Gobind. (2015). Research methodology and approaches. *IOSR Journal of Research & Method in Education*, 5(3), 48-51.
- Borrego, M., Douglas, E., & Amelink, C. (2009). Quantitative, qualitative, and mixed research methods in engineering education. *The Journal of Engineering Education*, 98(1), 53–66. doi:10.1002/j.2168-9830.2009.tb01005.x
- Choy, L. T. (2014). The strengths and weaknesses of research methodology: Comparison and complimentary between qualitative and quantitative approaches. *IOSR Journal of Humanities And Social Science*, 19(4), 99–104. doi:10.9790/0837-194399104
- Frascati Manual Proposed standard practice for surveys on research and experimental development. (2002). OECD Publications Service.
- Frederick, B. (2011). *Managing Science Methodology and Organization of Research, Innovation, Technology, and Knowledge Management*. Springer Science+Business Media, LLC.
- Gogoi, L., & Goowalla, H. (2015). A study on the impact of research methodology in Ph. D course: An overview. *International Journal of Development Research*, 5(11), 6065–6067.
- History of scientific method. (n.d.). Retrieved on August 26, 2016, from https://en.wikipedia.org/wiki/History_of_scientific_method
- Khairul, B. M. N. (2008). Case study: A strategic research methodology. *American Journal of Applied Sciences*, 5(11), 1602–1604. doi:10.3844/ajassp.2008.1602.1604
- Kothari, C. R. (2004). *Research methodology methods and techniques* (2nd ed.). New Age International (p) Limited.
- Kumar, R. (2011). *A step by step guide for beginners* (3rd ed.). London: SAGE Publications Ltd.
- OECD. (2014). *OECD Science, Technology and Industry Outlook, 2014*. OECD Publishing. doi:10.1787/sti_outlook-2014-en
- Pandey, P., & Pandey, M. M. (2015). *Research methodology: Tools and techniques*. Bridge Center.

Patil, S., & Mankar, A. (2016). Research methodology: For beginners. *International Research Journal of Multidisciplinary Studies*, 2(1), 1–6.

Rajasekar, S., Philominathan, P., & Chinnathambi, V. (2006). *Research methodology*. Retrieved June 10, 2016, from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.262.3449&rep=rep1&type=pdf>

Research design and methodology. (n.d.). Retrieved June 10, 2016, from <http://www.ais.utm.my/researchportal/files/2015/02/Example3-Res-Design.pdf>

Research methodology. (n.d.). Retrieved June 10, 2016, from http://www.mech.hku.hk/bse/bbse3002/Research_Methodology.pdf

Singh, Y. K. (2006). *Fundamental of research methodology and statistics*. New Age International (p) Limited.

Stephen, E., & Scott, A. (2010). *ICT R&D policies an international perspective*. IEEE Computer Society.

ADDITIONAL READING

Ashley, P., & Boyd, W. E. (2006). Quantitative and qualitative approaches to research in environmental management. *Australasian Journal of Environmental Management*, 13(2), 70–78. doi:10.1080/14486563.2006.10648674

Neuman, W. L. (2011). *Social research methods – Qualitative and quantitative approaches* (7th ed.). Allyn and Bacon Publication.

Peffer, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, 24(3), 45–77. doi:10.2753/MIS0742-1222240302

Whitesides, G. M. (2004). Whitesides Group: Writing a paper. *Advanced Materials*, 16(15), 1375–1377. doi:10.1002/adma.200400767

KEY TERMS AND DEFINITIONS

Applied Research: It deals with finding a solution or conclusion for an immediate problem by using well known theories or principles.

Basic Research: It deals with the finding the truth behind natural process or pure sciences problems.

Research: Finding the truth or solving the problem on the basis of scientific base is called as research.

Research Approaches: It define as the dealing of research by a particular manner or process.

Research Method: It includes all the procedure used to carry out research.

Research Methodology: It is way or path which shows how research is carried out in systematic manner.

Research Problems: It means that problems encountered during the research.

Scholarly Identity in an Increasingly Open and Digitally Connected World



Olga Belikov

Brigham Young University, USA

Royce Kimmons

Brigham Young University, USA

INTRODUCTION

The perceived role of the scholar has undergone rethinking in recent years as scholarly professionals have expanded conversations about the process of knowledge development and the role of the scholar in society. Boyer (1990) argues that “what we urgently need today is a more inclusive view of what it means to be a scholar – a recognition that knowledge is acquired through research, through synthesis, through practice, and through teaching” (p. 24), and he proposes that scholarship includes the four practices of discovery, integration, application, and teaching. In each of these aspects of scholarship, technology plays a role in defining possibilities, identifying priorities, and shaping practice, and advances in information technology over the past few decades have yielded significant technological artifacts (such as ubiquitous computing devices, data collection and storage systems, the internet, and social media) that influence what it means to be a scholar on an ongoing basis.

This chapter explores the intertwined relationship between technological advances and scholarly practice, and draws attention to emergent forms of scholarship described in the literature. The chapter will then highlight commonalities and differences between these emergent forms and discuss implications of the practices, especially those that affect the identity of the scholar as

they participate in these forms of scholarship. Throughout this conversation, technology will be used as an anchor for connecting scholarly practices to advances and social shifts of our time and will be treated as a co-evolutionary artifact with scholarship rather than as a change agent (cf. Veletsianos & Kimmons, 2012b).

BACKGROUND

As the historical centers of scholarly work for many centuries, universities have gradually developed and evolved in response to a variety of factors and are currently being reshaped in response to “globalization, mass expansion, and economic uncertainty, overlaid by new technologies connecting learners and content” and researchers “in new ways” (Siemens & Matheos, 2010, para. 17). Shifts in social norms and values and advances in technology have always impacted scholarship and the university, or institutionalized scholarship, in ways that reflect the needs and habits of the era (McNeely & Wolverton, 2008). Thus, when we consider emergent forms of scholarship connected to technology innovations, we must recognize that technology, society, and scholarship are all ever-evolving artifacts throughout all eras that influence and impact one another in complex and negotiated ways (Veletsianos & Kimmons, 2012b).

Some specific technologies that have historically impacted the creation and evolution of universities include the printing press, radio, television, microphotography/microfilm, mass publishing, microcomputers, the internet, and social media (Binkley, 1935; Tate, 1947; Siemens & Matheos, 2010; Veletsianos & Kimmons, 2012b). Each of these technologies bring with them different affordances, limitations, assumptions, and challenges that impact how scholars work in each of Boyer's areas of discovery, integration, application, and teaching. *Discovery* or the process of developing new knowledge through research is impacted as technologies improve efficiencies of data collection and analysis and allow for new methods of inquiry (e.g., big data, computational modeling). *Integration* is impacted as data and findings may be shared across distant locations and between experts within disciplines in a timelier manner. *Application* is influenced as scholars can more effectively report, serve, and collaborate with their communities, the public, and diverse colleagues from various disciplines. *Teaching* is impacted as scholars can teach students across geographic distances and employ new pedagogies and media to deliver instruction, assess student learning, and support student knowledge construction. *Identity* is impacted as scholars navigate their use of social medias and their offline and online identities converge while their sense of identities are impacted by their participation in these networks.

Forms of Emergent Technology-Influenced Scholarship

Many of the emerging scholarly practices that respond to recent technological advances associated with the internet and social media have been categorized into at least five general forms: *digital scholarship*, *social scholarship*, *open scholarship*, *public scholarship* and *networked participatory scholarship* (Kimmons, 2015). Each of these identified forms seeks to draw attention to a set of scholarly practices (or in some cases to advocate

for those practices) in contradistinction to previous norms in the following ways:

- **Digital Scholarship:** Emphasizes the power of the internet and digital media for making data sharing and collaboration cheaper, faster, and easier (Andersen, 2003; Borgman, 2007; Pearce, Weller, Scanlon, & Kinsley, 2010; Russell, Weinberger, & Stone, 1999).
- **Social Scholarship:** Highlights the importance of social interaction between scholars to generate quality work and to induct new scholars into academe by utilizing computer-mediated mechanisms like discussion groups (Berge & Collins, 1995), blogs (Chong, 2010), and social networking sites (Greenhow, 2009) to support scholarship that is conversational and less formal than the traditional publication cycle (Oblinger, 2010).
- **Open Scholarship:** Emphasizes the importance of utilizing technologies and practices for teaching and research that espouse openness and sharing for the purpose of broadening access to knowledge, reducing costs, enhancing scholarly impact, and supporting transparent and equitable practices (Furlough, 2010; Wiley & Green, 2012; Eysenbach, 2006; Norris, Oppenheim, & Rowland, 2008; Veletsianos & Kimmons, 2012a, para. 3).
- **Public Scholarship:** Articulates the importance of public participation by scholars as an obligation to the community and a desire to stay relevant in their respective fields through civic engagement (Brown-Dean, 2015).
- **Networked Participatory Scholarship:** Builds off of the three aforementioned forms of scholarship and attempts to bring them together into a unified vision of scholars using digital and social technologies to “share, reflect upon, critique, improve,

validate, and further their scholarship” in ways that fundamentally transform existing institutional structures (Veletsianos & Kimmons, 2012b).

COMMONALITIES AND DIFFERENCES

These emergent forms of scholarship are not mutually exclusive nor clearly delineated but rather represent a collection of connected emerging practices evolving with a variety of technologies, and at least three of the terms have been proposed to only understand or to promote a subset of emergent practices. If we consider all five categorizations and look for common themes across them, we discover some commonalities surrounding the ideas of *digitization*, *sharing*, *democratization*, *social networking*, and *transparency*. These themes are interconnected, and each deserves focused attention for us to recognize the richness of emergent forms of scholarly practice.

- **Digitization:** Refers to the transfer of scholarly artifacts and processes from previous media formats (e.g., print journals, letters, catalogues) to digital formats (e.g., electronic media, online journals, emails, databases). The process of digitization provides many opportunities for improved efficiency and access to scholarly work both for dissemination and collaboration. For example, database technologies allow for the collection, storage, manipulation, analysis, and retrieval of large datasets for research, which has given rise to a wide array of new interdisciplinary fields and profitable research lines. Through the use of shared databases and other digital technologies, the speed and scale at which research may be conducted has increased. All five identified categorizations rely upon digitization and are empowered by it, and without digital media and artifacts that

improve connections, sharing, and efficiencies, we would not be able to explore new forms of scholarship that rely upon information processing, social connections at a distance, and inexpensive publishing and sharing mechanisms.

- **Sharing:** Of data has improved through digitization as well and represents a common theme from all five categorizations. Large datasets can be collaboratively constructed and analyzed by geographically dispersed researchers and shared with scholars from various fields via online portals and other mechanisms. This sharing allows for improved collaboration and interdisciplinary scholarship, as traditional institutional and departmental boundaries are breached or bypassed. Sometimes referred to as primary source sharing, improved access to data sources allows for new types of questions to be asked and research lines to be explored as scholars can combine datasets from and easily bring in experts from a variety of fields to collect and analyze data.

This sharing extends to improved dissemination of findings, research outcomes, reviews, and scholarly opinion, or secondary sources. By publishing findings in online journals and less formal online venues (e.g., discussion fora, blogs), researchers allow their work to be indexed by online catalogs and search engines and to be accessible by researchers throughout the world, thereby circumventing many persistent problems of access to academic material. This improved access gives researchers greater facility to stay abreast of new findings, review peers' work, and to provide responses in a timely manner. Due to the nature of digital media, which can be disseminated and stored at relatively low costs, new sharing possibilities are also available through the innovation of open access journals, and scholars can engage in a variety of practices that are open, social, and networked.

These approaches to collaborating, sharing findings, and reducing or removing traditional barriers to dissemination may have the potential to *democratize* some aspects of the scholarly enterprise by allowing a greater number and diversity of scholars to contribute to research endeavors and to gain access to findings and discussion. Whereas it may be that the establishment of the university at the center of scholarly work occurred in large part due to mass publication and the inability of individuals and smaller institutions to keep pace with large entities' collection and storage of scholarly artifacts (Binkley, 1935), the reduction of these barriers may lead to a rethinking of the role of universities in society and scholars' relationships to them. As physical barriers are removed through the replacement of many traditional libraries with online resources and monetary barriers are removed through the growth of open access journals and other scholarly portals, collaboration and access of scholarly work will likely increasingly occur in virtual spaces rather than geographic locales, which has the impact of increasing access and participation across traditional geographic barriers but also calls into question the usefulness or necessity of many geographically bound sites of scholarly endeavor.

- **Social Networking:** Also influences sharing and dissemination by connecting the networked lives of scholars with their work. Emphasized in social scholarship, public scholarship, and networked participatory scholarship, each scholar has social connections that reflect personal and professional relationships, and scholars' literacies to traverse these connections, develop new relationships, and present meaningful online personas can influence their abilities to conduct meaningful work, share findings, and find and utilize available resources. By using an informal sharing space like a blog, for instance, scholars can share their work with colleagues and the public through all stages of the scholarly

process and elicit feedback or solicit partnerships to improve their work. Similarly, scholars can nurture online relationships through a social networking platform and quickly gain access to a wealth of suggestions regarding research questions, data sources, previous work, funding and job opportunities, and so forth and can quickly gain insight from a variety of people even outside of their field, thereby potentially improving interdisciplinary work. These social networking sites, especially Twitter, have become increasingly prevalent in professional settings as scholars use Twitter as a conference backchannel and a professional networking platform (Kimmons & Veletsianos, under review; Veletsianos, 2012). The wide variety of social networking sites allows scholars to contribute meaningfully to the public sphere both formally and informally.

- **Transparency:** Plays a role in these practices as well, as information is diffused freely and scholars' relationships, thoughts, opinions, and interactions take an increasingly public stage, thereby opening more aspects of their lives to both scrutiny and celebration (Tufekci, 2008). Open scholarship emphasizes the importance of removing barriers between learners and content and between learners and teachers, and for scholars to effectively connect with others socially online requires for them to be transparent in what they share (Dalsgaard & Paulsen, 2009; Mazer, Murphy, & Simonds, 2007). As scholars begin to remove these barriers, they participate in public scholarship and are able to engage in public spheres of their expertise. Such transparency suggests fundamental shifts in the nature of academic institutions and the roles that scholars play in society and may suggest shifts in thinking of scholars less like laboratory-dwelling recluses to micro-celebrities (Marwick & boyd, 2011).

Throughout these discussions, some scholars approach these issues with an interest in utility and focus on technologies' affordances to improve efficiency and scale. However, other scholars, as with open scholarship, take a social advocacy stance and argue that there are moral considerations to supporting forms of technology-enhanced scholarship that are democratized or transparent. Such fundamental shifts in thinking and emphases upon advocacy bring with them certain difficulties and implications that require attention.

Democratic participation in the scholarly process, as with the political process, requires certain literacies, skills, and attitudes and an allotment of time on the part of participants. Thus, though barriers to entry may be reduced in terms of geography and overt monetary constraints, other factors may influence an individual's or institution's ability to effectively participate in these new forms of scholarship, suggesting that barriers may not be altogether removed but rather changed to favor new groups of scholars (Veletsianos & Kimmons, 2012a). As has been pointed out in discussions of the public domain, the openness of any domain (e.g., the internet) favors those with the resources and skills necessary to gain advantage from it (Chander & Sunder, 2004). In terms of resources, those individuals and institutions with the infrastructure necessary to support sustained access and sharing (e.g., networks, data sources) and those who have developed the skills necessary to develop and utilize social networks and to traverse massive amounts of information will gain greater advantage from these technologies. This conversely means that scholars who lack valuable social connections in the real world may have difficulty creating and sustaining new social relationships online and that traditional problems of inequity may be replicated in online spaces (cf. Thelwall, 2009, and the problem of homophily).

Democratic participation also requires participants to be well-informed and to have access to dissenting views.

DIFFICULTIES AND IMPLICATIONS

R

However, two problems have arisen in conjunction with the advent of the social and semantic webs, which pose problems for emerging scholarly practice. The social web, or Web 2.0, allows participants to connect with like-minded individuals via social media and social networks, to create and remix content, and to share content with the world. Because users of social media readily connect only with those that they agree or have some other social connection with, this creates a situation in which online echo chambers emerge and scholars naturally isolate themselves from dissenting opinion (Gilbert, Bergstrom, & Karahalios, 2009). Consider a blog that a scholar uses to share work in progress to solicit feedback. If this blog is followed by like-minded scholars, this can lead to authors only interacting with those who already share their opinions and may have the effect of minimizing or ignoring dissenting views.

Similarly, the semantic web, or Web 3.0, relies upon machine algorithms to present information to users that is relevant to their interests and requirements. As scholars utilize search engines and other mechanisms for navigating web content, filter bubbles can emerge in which a scholar's traversal of online resources is effectively managed to only deliver supporting viewpoints and similarities in thinking (Pariser, 2011). This means that if scholars are looking for evidence to support their work, they can more easily find it and not have to sift through dissenting or irrelevant evidence in the process.

Both of these examples have deep implications for scholarly practice, because though critique and controversy have traditionally been valuable hallmarks of scholarship, emerging practices may lead to scholars operating within new forms of insulation and isolation. Whereas scholars have traditionally been viewed as largely operating within the "ivory tower" of an academic institution that may be disconnected from the problems and

realities of the world at large, social technologies can both supplant these barriers by increasing communicative efficacy with those outside the institution but simultaneously instantiate new barriers based upon the social relationships, beliefs, and values of the scholar. In these new forms of scholarship, then, it behooves scholars to consider the value of dissenting viewpoints and to utilize new technologies in a manner that allows them to meaningfully grow, develop, and work within a diverse global community.

There is also the prevailing issue of identity fragmentation that is affected by all of these forms of scholarly participation. Identity can be understood as a constellation of interconnected fragments of social participation (Kimmons & Veletsianos, 2014). As scholars seek to be participatory and interact with these emergent technologies, they encounter incongruencies between their values and opinions and those of the general public. When it comes to sharing, the scholar may be inclined to share something that is of value to them but may be perceived by institutions, fellow scholars, or students as inappropriate material for a public forum. These blurred lines between the personal and professional sharing done by scholars is largely in part due to the pluralistic use of blogs, social medias and other forms of sharing. With greater potential for interaction, issues of privacy are raised (Lin, Hoffman, & Borengasser, 2013). Social networking sites are used to support educational endeavors, but there is little precedence for scholars to adhere to when it comes to use of these sites. The identity of the scholar participating in social networking sites is rarely an entirely authentic or complete representation of the scholar's identity due to issues of implicit and explicit appropriate uses of these media. The ideal of democratization through networked participatory scholarship may be turned upside down as scholars begin to be wary of participating in the public sphere for fear of scrutiny. Scholars

may then be inclined to turn inwards and remove themselves from the democratic ideals that public and open scholarship promote. Although using these networking tools as scholars can be valuable and can nurture beneficial scholarly relationships, the effects of the convergence between offline and online personas may influence the representational identity of the scholar to the point that it may become fragmented and disjointed from the true identity of the scholar (Kimmons & Veletsianos, 2014). This fragmented and incomplete identity may hinder the scholar from using these social networking sites as a means of truly meaningful social participation.

FUTURE RESEARCH DIRECTIONS

There is a great deal of ambiguity surrounding what appropriate and meaningful participatory scholarship looks like. Without a deeper understanding of how scholars can participate in the public sphere in the most impactful way possible while still preserving their values and those of their institution, they will not be able to contribute meaningfully. It would be valuable to conceive a flexible and plural framework for what networked participatory scholarship should look like. While some of these more specific forms of scholarship such as open scholarship have some prescriptions for optimal participation, the dependence that networked participatory scholarship has upon the value system of the scholar leaves room for discussion regarding guidelines for this type of participation. More specifically there is also the need to analyze what prescriptions institutions are making for the scholars and whether they are held to the standards of explicit or implicit assumptions of acceptable behavior.

The framework of the acceptable identity fragment as it applies in higher education is another area that needs to be looked at more closely.

While the idea of identity fragmentation has been valuable in discussing networked participation across educators of younger ages, it is important to inspect this framework to understand whether this identity fragment truly applies in higher education. The way in which scholars engage in the public sphere is very unique in comparison to other educators and understanding the way in which scholars interact with identity fragmentation needs to be further explored and understood. Along with other educators, there is also a significant amount of teachers using these forms of public participation and social networking in their classroom (Piotrowski 2015). It is crucial that as these medias become widespread as teaching tools, that we recognize the implications that come with classroom use of these tools. A deeper analysis of identity, power relationships, and community perception is vital and as some of these aspects of public participation are better understood, there may be hope for a guide of how scholars approach networked participation in both formal and informal settings.

CONCLUSION

In this chapter, we see that emerging scholarly practices extend existing practice but also stand to continually reshape it, generating new efficiencies, possibilities, and challenges. Clear identification of all emerging practices is difficult, but the literature suggests at least five general emergent forms that include a wide range of technology-influenced scholarly practices, namely: digital scholarship, social scholarship, open scholarship, public scholarship, and networked participatory scholarship. Scholarly practice has historically evolved with technology and other cultural artifacts, and these categorizations of emerging scholarly practice serve to draw attention to a set of interconnected themes permeating the literature that may help guide us to meaningfully direct the ever-morphing forms that scholarship continues to take in our time.

REFERENCES

- Andersen, D. L. (2003). *Digital scholarship in the tenure, promotion, and review process*. M.E. Sharpe.
- Berge, Z., & Collins, M. (1995). Computer-mediated scholarly discussion groups. *Computers & Education*, 24(3), 183–189. doi:10.1016/0360-1315(95)00010-J
- Binkley, R. C. (1935). New tools for men of letters. *The Yale Review*, 24.
- Borgman, C. L. (2007). *Scholarship in the digital age: Information, infrastructure, and the Internet*. Cambridge, MA: MIT Press.
- Boyer, E. (1990). *Scholarship reconsidered: Priorities for the professoriate*. Princeton, NJ: The Carnegie Foundation for the Advancement of Teaching.
- Brown-Dean, K. L. (2015). Emphasizing the scholar in public scholarship. *Political Science and Politics*, 48(1), 55–57. doi:10.1017/S1049096515000426
- Chander, A., & Sunder, M. (2004). The romance of the public domain. *California Law Review*, 92(5), 1331. doi:10.2307/3481419
- Chong, E. K. M. (2010). Using blogging to enhance the initiation of students into academic research. *Computers & Education*, 55(2), 798–807. doi:10.1016/j.compedu.2010.03.012
- Dalsgaard, C., & Paulsen, M. (2009). Transparency in cooperative online education. *International Review of Research in Open and Distance Learning*, 10(3). doi:10.19173/irrodl.v10i3.671
- Eysenbach, G. (2006). Citation advantage of open access articles. *PLoS Biology*, 4(5), e157. doi:10.1371/journal.pbio.0040157 PMID:16683865
- Furlough, M. (2010). Open access, education research, and discovery. *Teachers College Record*, 112(10), 2623–2648.

- Gilbert, E., Bergstrom, T., & Karahalios, K. (2009). Blogs are echo chambers: Blogs are echo chambers. *System Sciences, 2009. HICSS'09. 42nd Hawaii International Conference on System Sciences*, 1–10.
- Greenhow, C. (2009). Social scholarship: Applying social networking technologies to research practices. *Knowledge Quest*, 37(4), 42–47.
- Kimmons, R., & Veletsianos, G. (2014). The fragmented educator 2.0: Social networking sites. Acceptable identity fragments, and the identity constellation. *Computers & Education*, 72, 292–301. doi:10.1016/j.compedu.2013.12.001
- Lin, M. G., Hoffman, E. S., & Borengasser, C. (2013). Is social media too social for class? A case study of Twitter use. *TechTrends*, 57(2), 39–45. doi:10.1007/s11528-013-0644-2
- Marwick, A., & boyd, (2011). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society*, 13(1), 114–133. doi:10.1177/1461444810365313
- Mazer, J., Murphy, R., & Simonds, C. (2007). Ill see you on Facebook: The effects of computer-mediated teacher self-disclosure on student motivation, affective learning, and classroom climate. *Communication Education*, 56(1), 1–17. doi:10.1080/03634520601009710
- McNeely, I. F., & Wolverton, L. (2008). *Reinventing knowledge*. New York, NY: W.W. Norton & Co.
- Norris, M., Oppenheim, C., & Rowland, F. (2008). The citation advantage of open-access articles. *Journal of the American Society for Information Science and Technology*, 59(12), 1963–1972. doi:10.1002/asi.20898
- Oblinger, D. G. (2010). From the campus to the future. *EDUCAUSE Review*, 45(1), 42–52.
- Pariser, E. (2011). The filter bubble: What the Internet is hiding from you. Penguin UK.
- Pearce, N., Weller, M., Scanlon, E., & Ashleigh, M. (2010). Digital scholarship considered: How new technologies could transform academic work. *In Education*, 16(1).
- Piotrowski, C. (2015). Emerging research on social media use in education: A study of dissertations. *Research in Higher Education*, 27.
- Russell, K., Weinberger, E., & Stone, A. (1999). Preserving digital scholarship: The future is now. *Learned Publishing*, 12(4), 271–280. doi:10.1087/09531519950145670
- Siemens, G., & Matheos, K. (2010). Systemic changes in higher education. *Technology and Social Media*, 16(1).
- Tate, V. D. (1947). From Binkley to Bush. *The American Archivist*, 10(3), 249–257. doi:10.17723/aarc.10.3.yr86jr3w0598m881
- Thelwall, M. (2009). Homophily in MySpace. *Journal of the American Society for Information Science and Technology*, 60(2), 219–231. doi:10.1002/asi.20978
- Tufekci, Z. (2008). Grooming, gossip, Facebook, and MySpace. *Information Communication and Society*, 11(4), 544–564. doi:10.1080/13691180801999050
- UNESCO. (2002). *Forum on the impact of open courseware for higher education in developing countries: Final report*. Retrieved from <http://unesdoc.unesco.org/images/0012/001285/128515e.pdf>
- Veletsianos, G. (2012). Higher education scholars participation and practices on Twitter. *Journal of Computer Assisted Learning*, 28(4), 336–349. doi:10.1111/j.1365-2729.2011.00449.x
- Veletsianos, G., & Kimmons, R. (2012a). Assumptions and challenges of open scholarship. *International Review of Research in Open and Distance Learning*, 13(4), 166–189. doi:10.19173/irrodl.v13i4.1313

Veletsianos, G., & Kimmons, R. (2012b). Networked participatory scholarship: Emergent techno-cultural pressures toward open and digital scholarship in online networks. *Computers & Education*, 58(2), 766–774. doi:10.1016/j.compedu.2011.10.001

Wiley, D. (2003). *A modest history of OpenCourseWare*. Autounfocus blog. Retrieved from <http://www.reusability.org/blogs/david/archives/000044.html>

Wiley, D. (2010). Open education. In M. K. Barbour & M. Orey (Eds.), *The foundations of instructional technology*. Retrieved from http://projects.coe.uga.edu/ITFoundations/index.php?title=Open_Education

Wiley, D., & Green, C. (2012). Why openness in education? In D. Oblinger (Ed.), *Game changers: Education and information technologies* (pp. 81–89). Educause.

ADDITIONAL READING

Binkley, R. C. (1935). New tools for men of letters. *The Yale Review*, 24.

Kimmons, R., & Veletsianos, G. (2014). The fragmented educator 2.0: Social networking sites, acceptable identity fragments, and the identity constellation. *Computers & Education*, 72, 292–301. doi:10.1016/j.compedu.2013.12.001

Siemens, G., & Matheos, K. (2010). Systemic changes in higher education. *Technology and Social Media*, 16(1).

Veletsianos, G., & Kimmons, R. (2012b). Networked participatory scholarship: Emergent techno-cultural pressures toward open and digital scholarship in online networks. *Computers & Education*, 58(2), 766–774. doi:10.1016/j.compedu.2011.10.001

KEY TERMS AND DEFINITIONS

Digital Scholarship: An emergent form of scholarship that emphasizes the use of digital technologies to support efficiency.

Networked Participatory Scholarship: An emergent form of scholarship that emphasizes the role that social technologies play in reshaping educational institutions and the role of the scholar.

Open Scholarship: An emergent form of scholarship that emphasizes openness, sharing, and democratization of educational resources.

Public Scholarship: An emergent form of scholarship that emphasizes responsibility of the scholar to engage in the public sphere.

Social Scholarship: An emergent form of scholarship that emphasizes the importance of collaboration and mentoring in the scholarly process.

Category R

Risk Assessment

Fuzzy Logic Approach in Risk Assessment

R**Çetin Karahan***Directorate General of Civil Aviation, Turkey***Esra Ayça Güzeldereli***Afyon Kocatepe University, Turkey***Ashlhan Tüfekci***Gazi University, Turkey*

INTRODUCTION

As an intuitively subjective and ambiguous notion, risk requires a detailed and attentive study, though. Since risk involves the events likely to occur in the future, risk assessment is an area where uncertainty is prevalent. Therefore, making use of experience, previous statistics and prediction ability is crucial in risk studies. The field of risk management is enriched with new techniques and methodologies, which serve the purposes of discovering more data, reducing subjectivity through more quantitative models and building flexible systems conducive to be updated with the obtained data. One of these new tools is fuzzy logic representing the uncertainty and to study with imprecise and uncertain knowledge.

This paper discusses the application of fuzzy logic to risk assessment process as an alternative to the traditional models due to its similarity to human reasoning and its accuracy in interpreting uncertainty. A fuzzy logic-based algorithm is developed for the purposes of enhancing risk assessment accuracy. Impact and likelihood factors, which are fundamental elements of risk, measured by the fuzzy logic-based approach. Beyond the impact and likelihood values, the factors directly effecting impact and likelihood also considered in this study and these factors included in fuzzy operations, in order to reduce subjectivity and increase precision.

In this study, an approach is explained for risk assessment. The aim of this approach is providing insight as a powerful alternative to traditional methods. A comparison between the risk values measured by the new model and those measured by the classical model supports the view that using fuzzy logic in risk assessment helps to produce more effective outcomes.

BACKGROUND

The concept of fuzzy logic was first introduced in 1965 by Prof. Lotfi A. Zadeh who developed Lukasiewicz's multivalued logic to set theory and created what he called fuzzy sets – sets whose elements belong to it in different degrees. At the start, fuzzy logic was a theoretical concept with little practical application. In the 1970's, Prof. Edrahim Mamdani of Queen Mary College, London, built the first fuzzy system, a steam-engine controller, and he later designed the first fuzzy traffic lights. His work led to an extensive development of fuzzy control applications and products (Cirstea, Dinu, McCormick & Khor, 2002, pp. 113-114).

Bellman and Zadeh (1970) developed an initial general theory on decision making in fuzzy environment which include three basic concepts as fuzzy goals, fuzzy constraints and fuzzy decisions. It is concluded that the proposed theory is generally has advantages according to the traditional probability theory.

DOI: 10.4018/978-1-5225-2255-3.ch588

Tah and Carr (1999) claimed that the current risk management techniques mostly based on the operational research techniques developed in 1960s and usually had failed to meet the needs of project managers. They introduced a fuzzy risk analysis model for a construction project to eliminate the past studies' concentration on particular risks and proposed a model which have a generic and generally practicable representation.

The development of fuzzy set theory to fuzzy technology during the first half of the 1990s has been very fast. More than 16,000 publications have appeared since 1965. Most of them have advanced the theory in many areas. Quite a number of these publications describe, however, applications of fuzzy set theory to existing methodology or to real problems. In addition, the transition from fuzzy set theory to fuzzy technology has been achieved by providing numerous software and hardware tools that considerably improve the design of fuzzy systems and make them more applicable in practice (Zimmerman, 2001, p. xxi).

Hajiha, Roodposhti and Askary (2009) provided a risk assessment approach conducted on the basis of fuzzy logic for audit risk, inherent risk and control risk. The results are compared to a real case and the accuracy level of the results is discovered to be relatively higher.

Keropyan and Gil-Lafuente (2011) place the emphasis on the importance of the ability of making right decisions and provide examples of use of fuzzy logic in selection of the decision-making styles within the scope of strategic management.

A. Pesic, D. Pesic and Tepavcevic (2012) proposed fuzzy logic as an innovative strategic management instrument to identify internal risks and to eliminate some of the restrictions imposed by the classical methods.

Nunes and Marques (2012) developed a fuzzy logic-based model for using in risk assessment of work accidents and occupational diseases and discussed the superiority of the fuzzy logic over the classical methods arguing that it allows a more comprehensive evaluation of risks and combination of both subjective and objective criteria.

Dainiene and Dagiliene (2013) calculated the sustainability of the business by using fuzzy logic and the competence of the business was shown experimentally. In the study, the operational value of the business is calculated on the basis of the financial and non-financial data.

Shang and Hossen (2013) provided evidence for the possible use of fuzzy logic as a decision-making system. The study supports the view that fuzzy logic may also be used in complex risk systems along with the other risk models such as decision trees and artificial neural network.

All of the studies that have been investigated support the view that the use of fuzzy logic as a decision-making system in strategic management has a number of advantages particularly in situations where there is a high level of unclarity.

USING FUZZY LOGIC TO IDENTIFY THE LEVEL OF RISK

General Information on Fuzzy Logic

The concept in question is that of a fuzzy set, that is, a "class" with a continuum of grades of membership. As will be seen in the sequel, the notion of a fuzzy set provides a convenient point of departure for the construction of a conceptual framework which parallels in many respects the framework used in the case of ordinary sets, but is more general than the latter and, potentially, may prove to have much wider scope of applicability, particularly in the fields of pattern classification and information processing. Essentially, such a framework provides a natural way of dealing with problems in which the source of imprecision is the absence of sharply defined criteria of class membership rather than the presence of random variables (Zadeh, 1965, p. 339).

In short, fuzzy set is a set where the members of the universal set have appointed values between the closed range of 0-1 and their membership status and membership degrees can be identified by these appointed values. The following should be taken into consideration while forming fuzzy sets:

- Fuzzy sets which are marked in the universal set need to be distributed symmetrically.
- The number of fuzzy sets defined for each variable should be odd number (typically 3, 5 or 7). Thus, some fuzzy sets may remain in the middle. Here, as an example, for temperature value it can be defined 3 fuzzy sets with cold, warm and hot values.
- For providing each value being defined, fuzzy sets should be overlapped to each other with a certain percentage. This will ensure the execution of more than one rule in the process of determining the output.
- Triangle or trapezoidal membership functions had better to be selected, which take less time to compute.

Kulkarni (2001) identifies Fuzzy Inference Systems (FIS) as a tool using if-then rules to state the relation between the input and output fuzzy sets. Fuzzy relations indicate the degree of existence or non-existence of the links and the interactions between the members of one or several sets.

While a fuzzy system is designed, system inputs and outputs are identified and fuzzy sets are built for each input and each output. Relationship between these sets of input and output are built through logical processes. Fuzzy logic system converts these verbal statements into numerical statements and generates a single value at the end.

Risk and Risk Management

Risk is the combination of the likelihood and the consequence of a specified hazard being realized. The type of risk analysis used should be appropriate for the available data and to the exposure, frequency and severity of potential loss. Quantitative risk analysis incorporates numerical estimates of frequency or probability and consequence. In practice a sophisticated analysis of risk requires extensive data which are expensive to acquire or often unavailable (Pokoradi, 2002, p. 64).

Because it facilitates the exploration of the effects of the decisions made, actions of prioritization, mitigation and measurement, risk assessment

is a mechanism which helps a great deal with organizational stability.

Risk management is an iterative and cyclic process whose main aim is to eliminate or at least to reduce the risks according to the ALARP (as low as reasonably practicable) principle. (Nunes & Marques, 2012, p. 22).

Traditionally, probability theory was one of the methodologies used most in risk assessment. Probabilistic risk assessment (PRA) is the general term for risk assessment that uses theory and models to represent the likelihood of different risk levels in a population (i.e. variability) or to characterize the uncertainty in risk estimates. Among various probabilistic techniques to quantify uncertainties, the most widely used approach has been Monte-Carlo analysis (MCA) (Darbra, Eljarrat & Barcelo, 2008, p. 379).

One of the most frequently conducted methods is the matrix method. Risk values are obtained by multiplying likelihood of occurrence value by impact value. Risk values are classified as low, moderate and high and the value range for each classification is determined according to the preferred scale. Below is an example of impact-likelihood matrix.

Fine-Kinney method and fault-tree analysis also popularly used tools for risk assessment. Fuzzy logic approach, has recently come forward as a big contributor to the studies on risk assessment. A superior approach to model situations where there is a substantial amount of uncertainty, fuzzy logic is capable of incorporating human reasoning and judgment into the system.

Designing Fuzzy System

The factors directly effecting the likelihood and impact of an event are selected to use their scores in fuzzy calculations to generate an impact and a likelihood output value. These two output values of impact and likelihood are then become the inputs of the final FIS which generates a risk score. In this sense, fuzzy logic operates on three different systems rather than only one as illustrated in Figure 2.

Figure 1. Example of impact- likelihood matrix used in risk assessment

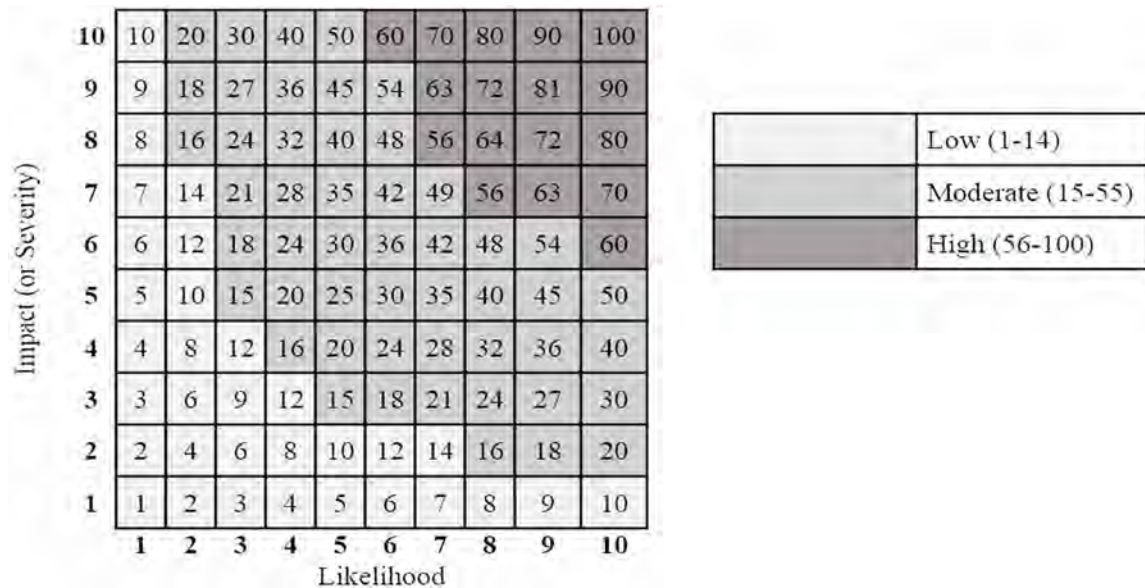
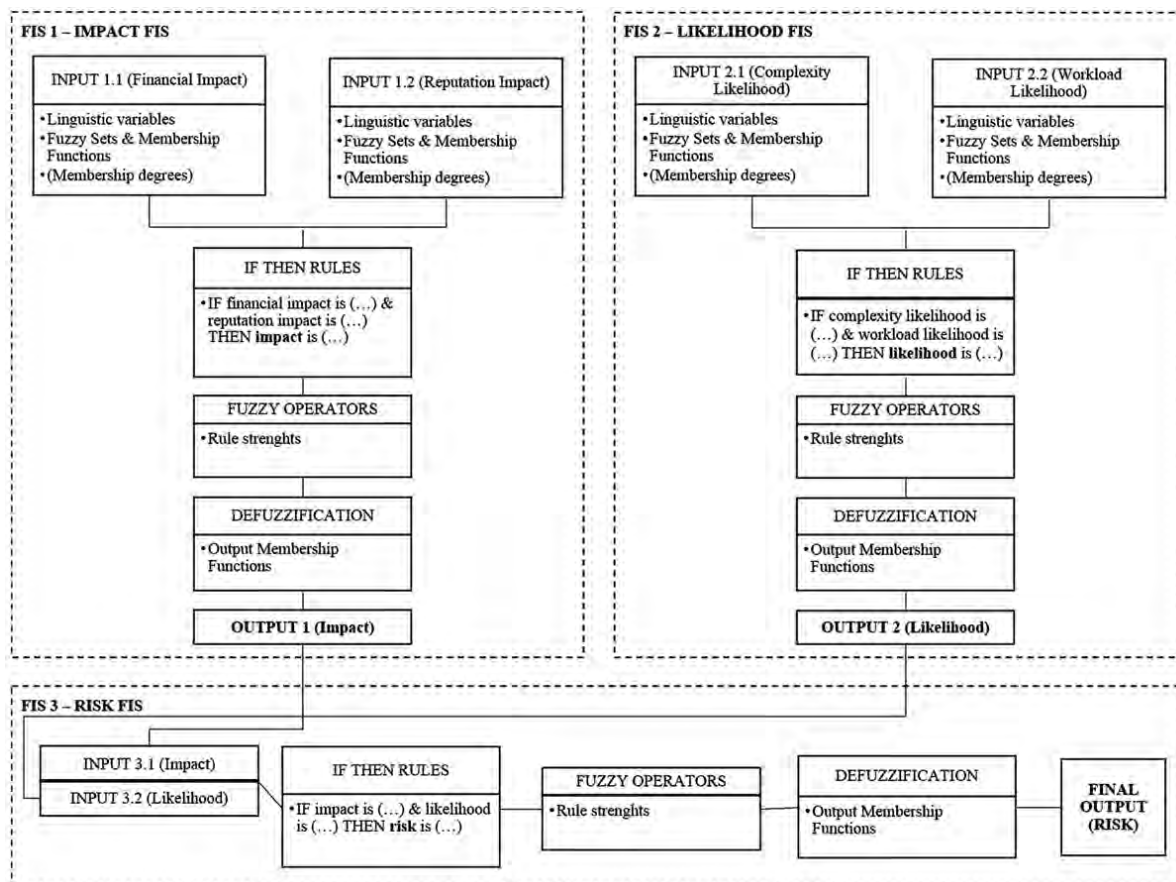


Figure 2. Fuzzy Inference System for risk assessment



The design of fuzzy system involves manually created fuzzy sets, functions and if-then rules as well as Fuzzy Logic component of MATLAB 7.0 program. Hence, the system can be converted to a software with any programming language easily by using these manually created and expressed fuzzy system components. MATLAB 7.0 is used to verify the calculations that are done manually.

Fundamental properties of the designed FIS:

Type : Mamdani

“and” Method : min

Defuzzification Method : centroid

Implication Method : min

Aggregation Method : max

Impact Likelihood Risk

Input : [1x2] [1x2] [1x2]

Output : [1x1] [1x1] [1x1]

Rule : [1x9] [1x9] [1x25]

A numerical example also embedded in the FIS explanations steps. The example includes impact and likelihood factors' scores which supposed to be generated by expert opinions. In order to testify the flexibility and accuracy of the fuzzy logic system, the example uses decimal values, which would otherwise be integer¹. The numerical inputs for the example is as follows:

Impact Factors' Scores Are: Financial = 7,5
and Reputation = 6,4

Likelihood Factors' Scores Are: Complexity = 7
and Workload = 6

Table 1. Scoring criteria for the risk-related impact factor

Impact Factor	Score	Expression	Damage (USD)
Financial loss (likely to occur)	1	Too low	< 3.000
	2	Too low	3.000 – 10.000
	3	Low	10.000 – 17.000
	4	Low	17.000 – 34.000
	5	Moderate	34.000 – 67.000
	6	Moderate	67.000 – 117.000
	7	High	117.000 – 167.000
	8	High	167.000 – 250.000
	9	Too high	250.000 – 333.000
	10	Too high	> 333.000
			Damage expression
Reputational damage (likely to occur)	1	Too low	only at the level of the concerned stakeholders
	2	Too low	only at the level of the stakeholders involved
	3	Low	being in the news in local printed media
	4	Low	being in the news in local printed and visual media
	5	Moderate	being in the news in local and regional printed media
	6	Moderate	being in the news in local and regional printed and visual media
	7	High	being in the news in national printed media
	8	High	being in the news in national printed and visual media
	9	Too high	being in the news in international printed media
	10	Too high	being in the news in international printed and visual media

Design of Impact and Likelihood FISs

Step 1: Identifying Input Data Set and Linguistic Variables

In case a risk occurred, there might be a number of subsequent social, economic, reputation and legal impacts. In this study only two of the factors were considered as inputs (financial and reputation impact). Input values were constructed on the basis of certain criteria, which are shown in Table 1.

The likelihood of a risk occurring depend on several factors such as workload and complexity of the process, the extent to which information technologies are used and the quality of the human resources. This study is concerned with only two factors as input: complexity and workload. The criteria determined for these values are shown in Table 2.

The criteria specified in above tables are configured to allow for the expert opinion to be included in the basis of fuzzy logic processes.

Step 2: Generating Membership Functions

Below are the membership functions built for the impact and likelihood fuzzy systems.

Table 3. Financial Impact membership functions

Condition	Membership Degree	Fuzzy Set
$0 < fI \leq 2$	$fI_{md} = 1$	Low
$2 < fI \leq 4$	$fI_{md} = \frac{4 - fI}{2}$ $fI_{md} = \frac{fI - 2}{2}$	Low Moderate
$4 < fI \leq 6$	$fI_{md} = 1$	Moderate
$6 < fI \leq 8$	$fI_{md} = \frac{fI - 6}{2}$ $fI_{md} = \frac{8 - fI}{2}$	High Moderate
$8 < fI \leq 10$	$fI_{md} = 1$	High

where fI_{md} indicates the membership degree of financial impact.

By referencing Figure 3, the membership functions for “financial impact (fI)” factor are generated as follows (see Table 3).

Since the functions for the financial and reputation impact are the same, the graphs and the formula for the reputation impact are not provided separately. As can be seen from the Figure 4,

Figure 3. Graph showing the membership functions for financial and reputation impact

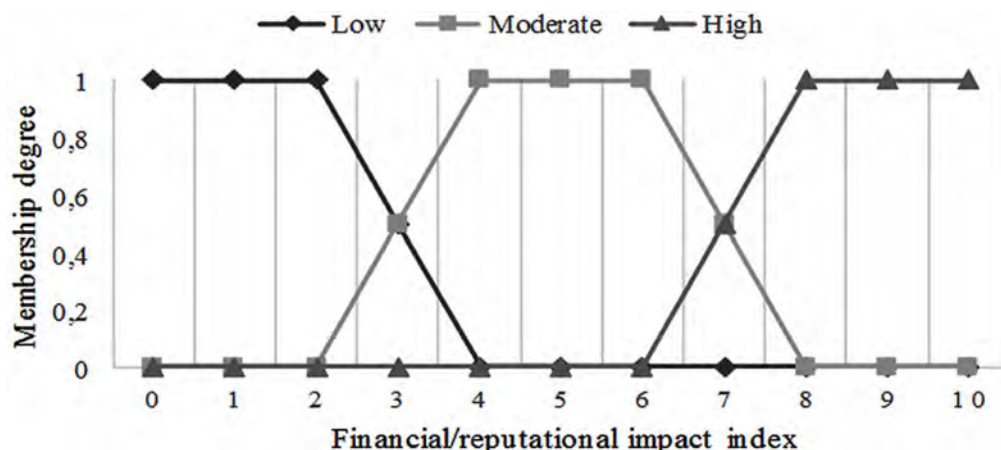


Table 2. Scoring criteria for the risk-related likelihood factor

Complexity (Assesses whether the complexity of the operations or the relevant legislation impedes the exercise of the controls, thereby increasing the likelihood of faults.)	
1	Operation/legislation isn't complex, fault likelihood is TOO LOW.
2	Operation/legislation isn't complex, fault likelihood is TOO LOW.
3	Operation/legislation isn't quite complex, fault likelihood is LOW.
4	Operation/legislation isn't quite complex, fault likelihood is LOW.
5	Operation/legislation is somewhat complex, fault likelihood is MODERATE.
6	Operation/legislation is somewhat complex, fault likelihood is MODERATE.
7	Operation/legislation is complex, fault likelihood is HIGH.
8	Operation/legislation is complex, fault likelihood is HIGH.
9	Operation/legislation is excessively complex, fault likelihood is TOO HIGH.
10	Operation/legislation is excessively complex, fault likelihood is TOO HIGH.
Workload (Assesses whether the understaff in face of the workload increases the likelihood of faults)	
1	The number of staff is adequate to perform the workload.
2	The number of staff is adequate yet the staff is inexperienced to perform the workload.
3	The number of staff is hardly adequate to perform the workload.
4	The number of staff is only adequate to perform the workload with extra effort and time.
5	The number of staff is seldomly inadequate to perform the workload.
6	The number of staff is sometimes not adequate to perform the workload.
7	The workload is high for the number of staff. Work is completed with short delay.
8	The workload is high for the number of staff. Work is completed with substantial delay.
9	The workload is too high for the number of staff. Some work is incomplete.
10	The workload is too high for the number of staff. Substantial amount of work is incomplete.

Figure 4. Graph showing the membership functions for the complexity and workload likelihood factors

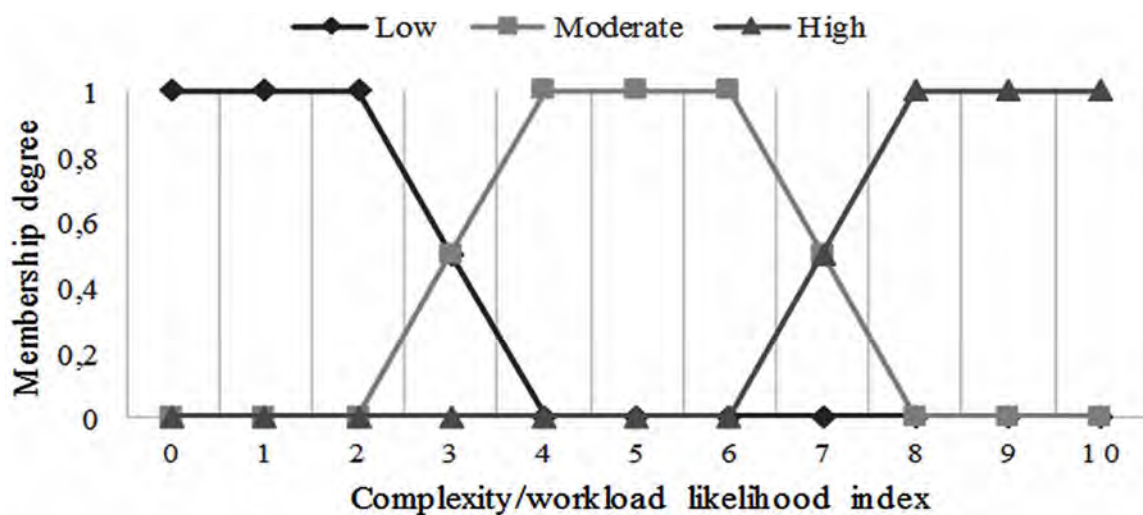


Table 4. If-then rules for “Impact”

		Financial Impact		Reputation Impact		Impact
1	IF	Low	&	Low	THEN	Too low
2		Low		Moderate		Low
3		Low		High		Moderate
4		Moderate		Low		Low
5		Moderate		Moderate		Moderate
6		Moderate		High		High
7		High		Low		Moderate
8		High		Moderate		High
9		High		High		Too high

the membership functions for likelihood factors (complexity and workload) would be the same with Table 3 entries.

In this regard, the financial impact (fI) value of 7,5 coincides with the range of 6-8 in the membership degree functions provided. Therefore;

$fI_{md} = (7,5 - 6)/2 = 0,75$ (high) - 75% “high” set member,

$= (8 - 7,5)/2 = 0,25$ (moderate) - 25% “moderate” set member.

In a similar way, as the reputation impact (rI) value of 6,4 is in 6-8 interval:

$rI_{md} = (6,4 - 6)/2 = 0,2$ (high) - 20% “high” set member,

$= (8 - 6,4)/2 = 0,8$ (moderate) - 80% “moderate” set member.

The complexity likelihood (cL) value of 7 coincides in 6-8 interval in the membership degree functions provided. Therefore;

$cL_{md} = (7 - 6)/2 = 0,5$ (high) -50% “high” set member,

$= (8 - 7)/2 = 0,5$ (moderate) -50% “moderate” set member.

Table 5. If-then rules for “Likelihood”

		Complexity		Workload		Likelihood
1	IF	Low	&	Low	THEN	Too low
2		Moderate		Low		Low
3		High		Low		Moderate
4		Low		Moderate		Low
5		Moderate		Moderate		Moderate
6		High		Moderate		Too high
7		Low		High		Moderate
8		Moderate		High		High
9		High		High		Too high

Table 6. Fuzzy rule strength for the numerical example with the nonzero value

Rule Num.	fI_{md}	rI_{md}	Rule Strength (min)	Impact
5	0,25	0,8	0,25	Moderate
6	0,25	0,2	0,2	High
8	0,75	0,8	0,75	High
9	0,75	0,2	0,2	Too high

As the workload likelihood (wL) value of 6 coincides in 4-6 interval

$wL_{md} = 1$ (Moderate) - 100% “moderate” set member.

Step 3: Rule Base

Once the inputs were constructed, output membership functions (Figure 5) are developed in the

fuzzy system and then if-then rules are determined. If-then rules, which allow association between the inputs and the outputs for the impact and likelihood factors are provided below.

Step 4: Application of Fuzzy Operators

As both impact factors are the members of the 2 fuzzy sets (“high” and “moderate”) there would be $2 \times 2 = 4$ rules with nonzero value (see 5, 6, 8 and 9th lines in Table 4). Following the selection of the rules, strength of each rule is calculated. The strength of a rule implies its lowest or weakest value (minimum). Table 6 indicates the rule strengths for each nonzero valued rules.

As one of the likelihood factors is a member of “1” set and the other is “2” sets, there appear $2 \times 1 = 2$ rules which are nonzero (see 5 and 6th lines in Table 5). The fuzzy rule strength for the example with the nonzero value is shown in Table 7.

Table 7. Fuzzy rule strengths for the numerical example with the nonzero value

Rule Num.	cL_{md}	wL_{md}	Rule Strength (min)	Likelihood
5	0,50	1	0,50	Moderate
6	0,50	1	0,50	Too High

Figure 5. Output membership functions for the impact and likelihood factors

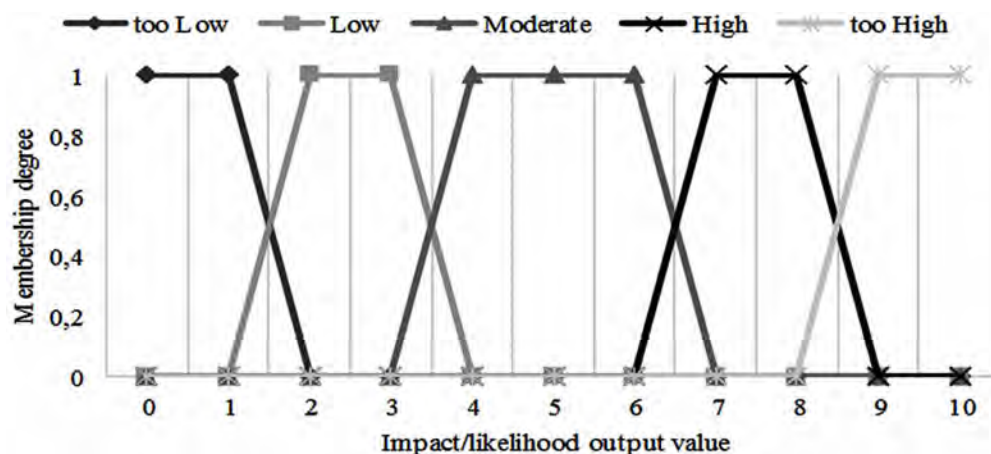


Table 8. Impact/likelihood output fuzzy sets' area formulas and centroid values

Fuzzy set	Area	Centroid (x;y)	Centroid
Too Low	$\frac{(4x - x^2)}{2}$	$\left(\frac{(x_1 + x_2 + \dots + x_n)}{n}, \frac{(y_1 + y_2 + \dots + y_n)}{n} \right)$	0,75
Low	$\frac{(6x - 3x^2)}{2}$		2,5
Moderate	$4x - x^2$		5
High	$\frac{(6x - 3x^2)}{2}$		7,5
Too High	$\frac{(4x - x^2)}{2}$		9,25

where x indicates a single value generated by FIS.

Step 5: Output Membership Functions for Impact and Likelihood Factors (Defuzzification)

Output membership functions, given in Figure 5, are constructed in the same way as the input membership functions. In this study, the outputs for the impact and likelihood factors are expressed as “too Low, Low, Moderate, High and too High”.

Shown below are the area formula for and the centroid² value of the impact and likelihood output membership functions given in the figure above (see Table 8). These formulas and values will eventually be used during defuzzification, which is the final process.

In this stage, using the acceptable rules (those with values other than 0), aggregation process conducted. As in Table 6, more than one rule set may share the same result or system output. Where this is the case, the output which is “the most real” or

“the strongest” (with the maximum value) will get appointed. Among the rules producing the same results, the ones with the maximum strength are selected from Table 6 to form Table 9.

In the same manner, the values received from Table 7 is as follows (see Table 10).

These threshold values will be cited for the area calculation using the system output graph. Below is the table showing the area values calculated for each system output using the values provided above (see Table 11).

As a consequence of the defuzzification process, the impact and likelihood scores are calculated to be 6,86 and 6,38 respectively. In order to verify this manual calculation, the same input values are also executed by the MATLAB program. The impact and likelihood values produced by MATLAB stands at 6,86 and 6,41 respectively as well.

Table 9. Fuzzy set threshold values for the numerical example (impact)

Too Low	Low	Moderate	High	Too High
-	-	0,25	0,75	0,2

Table 10. Fuzzy set threshold values for the numerical example (likelihood)

Too Low	Low	Moderate	High	Too High
-	-	0,50	-	0,50

Table 11. Fuzzy membership functions area values for the numerical example

Fuzzy Set	Area	Impact				Likelihood			
		x	Area	Centroid	(Area x Centroid)	x	Area	Centroid	(Area x Centroid)
Too Low	$\frac{(4x - x^2)}{2}$	0	0	-	0	0	0	-	0
Low	$\frac{(6x - 3x^2)}{2}$	0	0	-	0	0	0	-	0
Moderate	$4x - x^2$	0,25	0,9375	5	4,6875	0,50	1,75	5	8,75
High	$\frac{(6x - 3x^2)}{2}$	0,75	1,40625	7,5	10,54688	0	0	-	0
Too High	$\frac{(4x - x^2)}{2}$	0,2	0,38	9,05	3,439	0,50	0,875	9,125	7,984375
SUM			2,72375		18,67338		2,625		16,73438
$\frac{\sum((centroid)_i x (area)_i)}{\sum (area)_i}$		6,855761				6,375			

Design of Risk FIS

The third fuzzy system is modelled to obtain a risk value after the impact and likelihood factors for an identified risk assessed and the outputs are generated.

By referencing this graphic, the membership functions for “impact (I)” factor are generated (see Table 12).

The impact value of 6,86 and the likelihood value of 6,38 both coincides with the range of 6

Figure 6. Input parameters membership functions for the risk value

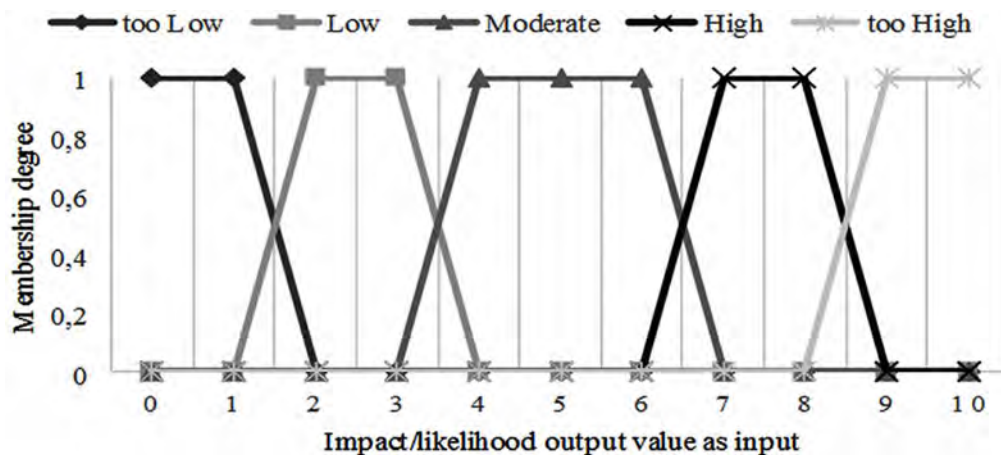


Table 12. Impact membership functions

Condition	Membership Degree	Fuzzy Set
$0 < I \leq 1$	$I_{md} = 1$	Too low
$1 < I < 2$	$I_{md} = 2 - I$ $I_{md} = I - 1$	Too low Low
$2 \leq I \leq 3$	$I_{md} = 1$	Low
$3 < I < 4$	$I_{md} = I - 3$ $I_{md} = 4 - I$	Moderate Low
$4 \leq I \leq 6$	$I_{md} = 1$	Moderate
$6 < I < 7$	$I_{md} = I - 6$ $I_{md} = 7 - I$	High Moderate
$7 \leq I \leq 8$	$I_{md} = 1$	High
$8 < I < 9$	$I_{md} = I - 8$ $I_{md} = 9 - I$	Too high High
$9 \leq I \leq 10$	$I_{md} = 1$	Too high

where “ I ” implies impact, “and “ md ” implies membership degree. The membership functions for likelihood factor is identical with impact factor and was not represented here separately. Below are the if-then rules, which help to make associations between the impact and likelihood inputs and the risk output.

– 7 in the membership degree functions provided. Accordingly:

$$\begin{aligned}
 I_{md} &= (6,86 - 6) = 0,86 \text{ (high) - 86\% “high” set member,} \\
 &= (7 - 6,86) = 0,14 \text{ (moderate) - 14\% “moderate” set member.} \\
 L_{md} &= 6,38 - 6 = 0,38 \text{ (high) - 38\% “high” set member,} \\
 &= 7 - 6,38 = 0,62 \text{ (moderate) - 62\% “moderate” set member.}
 \end{aligned}$$

As both factors (I and L) are members of “2” sets, there appear $2 \times 2 = 4$ rules which are nonzero (see 13, 14, 18 and 19th lines in Table13). The fuzzy rule strength for the example with the nonzero value, calculated through the minimum method, is shown below.

By using the aggregation method of maximum Table 16 is generated.

Table 17 shows the area values calculated for each system output using the values provided above.

As a consequence of the defuzzification process at this stage, the risk score is calculated to be 73,95. The risk score produced by MATLAB stands at 74,5.

Most risk assessment calculations involve the multiplication of the impact and likelihood values. In practice, a risk matrix, which is developed using this calculation, is cited to this end. By classical calculation for the numerical example, we come up with:

- Impact = $(7,5 + 6,4) / 2 = 6,95$
- Likelihood = $(7+6) / 2 = 6,5$
- Risk = 45,175

By this method, the risk value for the relevant process, would be 45. By this point of view, the weight imposed on the risk by the impact and the likelihood is assumed to be the same and the resulting value of 45 would be considered to be a moderate risk. When the risks associated with the all processes and actions are considered, a comparison will reveal sharp distinctions.

Fuzzy logic is considered to be advantageous in particular where the numerical value used to rate the risks as well as the impact and likelihood factors affecting the risks is close to the threshold.

FUTURE RESEARCH DIRECTIONS

Any studies related to a future projection suffers from the vagueness and unforeseen variability. For eliminating these shortcomings, more and more

Table 13. If-then rules for “Risk”

		Impact		Likelihood		Risk
1	IF	Too low	&	Too low	THEN	Low
2		Too low		Low		Low
3		Too low		Moderate		Low
4		Too low		High		Moderate
5		Too low		Too high		Moderate
6		Low		Too low		Low
7		Low		Low		Low
8		Low		Moderate		Low
9		Low		High		Moderate
10		Low		Too high		Moderate
11		Moderate		Too low		Low
12		Moderate		Low		Low
13		Moderate		Moderate		Moderate
14		Moderate		High		High
15		Moderate		Too high		High
16		High		Too low		Low
17		High		Low		Moderate
18		High		Moderate		High
19		High		High		High
20		High		Too high		High
21		Too high		Too low		Moderate
22		Too high		Low		Moderate
23		Too high		Moderate		High
24		Too high		High		High
25		Too high		Too high		High

Figure 7. Output membership function for risk value

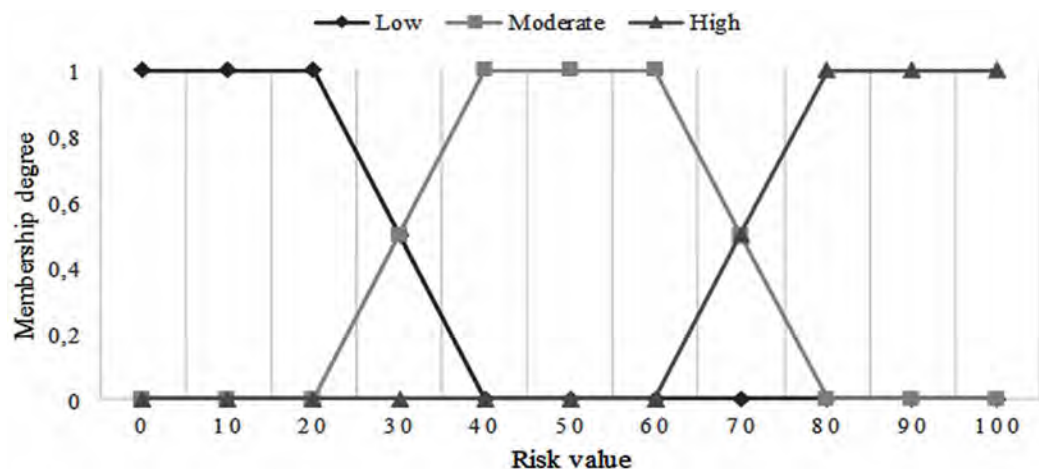


Table 14. Risk output fuzzy sets' area formulas and centroid values

Fuzzy Set	Area	Centroid ($x; y$)	Centroid
Low	$40x - 10x^2$	$\left(\frac{(x_1 + x_2 + \dots + x_n)}{n}, \frac{(y_1 + y_2 + \dots + y_n)}{n} \right)$	15
Moderate	$60x - 20x^2$		50
High	$40x - 10x^2$		85

Table 15. Fuzzy rule strengths for the numerical example with the nonzero value

Rule Num.	Impact	Likelihood	Rule Strength(min)	Risk
13	0,14	0,62	0,14	Moderate
14	0,14	0,38	0,14	High
18	0,86	0,62	0,62	High
19	0,86	0,38	0,38	High

Table 16. Output fuzzy set threshold values for the numerical example

Low	Moderate	High
-	0,14	0,62

data acquisition and processing is very critical. The approaches modelling the vague future have to use expert opinion with real and historical data and have to be flexible to be updated rapidly by

new acquisitions. The fuzzy logic based model which provided here should be enriched by new inputs of likelihood and impact factors for operational usage.

Some risk factors like structural changes, previous audit consequences, information technologies usage, social impacts, irregularities and fraud parameters –or any other specific parameters related to market- to be injected in the impact and likelihood FIS will provide significant rise on output quality and accuracy. Risk management is

Table 17. Fuzzy membership functions area values for the numerical example

Fuzzy Set	Area	x	Area	Centroid	(Area x Centroid)
Low	$40x - 10x^2$	0	0	15	0
Moderate	$60x - 20x^2$	0,14	8,01	50	400,40
High	$40x - 10x^2$	0,62	20,956	83,1	1.741,444
TOTAL			28,96		2.141,84
$\frac{\sum ((centroid)_i (area)_i)}{\sum (area)_i}$			73,95		

a continuous and repetitive process and created risk map have to be monitored and optimized periodically. This dynamic process creates the obligation of updating or revising the fuzzy rules according to these possible changes. The algorithm indicated in this study provides a convenient basic for developing a flexible software for creating a risk map. All the processes that an Organization carries out can be listed with their related risk factors to be processed by fuzzy logic based risk assessment algorithm.

For many Organizations, impact may be more important than likelihood in risk assessment. Considering this hypothesis, new fuzzy logic approaches that uses different weights or coefficients for these two factors can be developed. It is also possible to enhance precision by replacing the Mamdani FIS used in this study by Sugeno FIS.

It is predicted that fuzzy logic approach can also be applied to the other aspects of strategic management in addition to risk assessment.

CONCLUSION

Risk assessment is a continuous and recursive process that includes a cost-benefit analysis in order to maximize the use of the opportunities and minimize the threats. All organizations are risk-prone as they function to achieve their objectives. Achievement of the objectives depends on whether the organization is aware of the risks and able to respond accordingly.

Risk assessment plays a significant role in institutional management. In addition to its status as a prominent actor, risk assessment is also dominated by uncertainty. As the subject matter of risk is any event likely to occur in the future, risk-related studies rely heavily on experience, statistical data and prospecting.

Risk, in classical method, is calculated by multiplication of impact and likelihood factors.

This value is classified as low, moderate and high using a 3x3, 5x5 or 10x10 matrices. One of the restrictions of the classical approach is the separation by precise values. For example, in a scale of 1 and 25, a score of 15 indicates a moderate level whereas a score of 16 is considered high.

Fuzzy logic has advantages in the field of risk assessment thanks to its accuracy in modeling uncertainty and feasibility to be modeled in a similar way to human reasoning. Providing a theoretical and exemplary account of the fuzzy logic process, this study aims to provide the foundation of the concept for those interested in the subject.

The algorithm and the fuzzy sets used in this study are expressed as in plain terms as possible for the purposes of clarity and ease of use. A number of studies on fuzzy logic and strategic management have been investigated for the purposes of this study and it has been observed that fuzzy logic is also applicable to different areas of strategic management in addition to risk assessment.

REFERENCES

- Bellman, R. E., & Zadeh, L. A. (1970). Decision-making in a fuzzy environment. *Management Science*, 17(4), B-141–B-164. doi:10.1287/mnsc.17.4.B141
- Cirstea, M., Dinu, A., McCormick, M., & Khor, J. G. (2002). *Neural and fuzzy logic control of drives and power systems*. Woburn, MA: Elsevier Science.
- Dainiene, R., & Dagiliene, L. (2013). Company's going concern evaluation fuzzy model. *Economics and Management*, 18(3), 374–382. doi:10.5755/j01.em.18.3.3915
- Darbra, R. M., Eljarrat, E., & Barcelo, D. (2008). How to measure uncertainties in environmental risk assessment. *Trends in Analytical Chemistry*, 27(4), 377–385. doi:10.1016/j.trac.2008.02.005

- Hajiha, Z., Roodposhti, F. R., & Askary, S. (2009). *A new approach to evaluate audit risk Model by fuzzy expert Systems - evidences from Iran*. Retrieved from http://baa.group.shef.ac.uk/download_papers/index.php?submit=download_file&username=bafa&token=26684b3867876df7bd3fcac34d8cbddb&paper_id=177
- Keropyan, A., & Gil-Lafuente, A. M. (2011). A fuzzy-based decision model application on strategic management. *African Journal of Business Management*, 5(15), 6586–6590.
- Kulkarni, A. D. (2001). *Fuzzy logic fundamentals, computer vision and fuzzy neural systems*. Prentice Hall.
- Nunes, I. L., & Marques, M. S. (2012). Applications of fuzzy logic in risk assessment – The RA_X case. In M. F. Azeem (Ed.), *Fuzzy inference system - Theory and applications* (pp. 21–40). Rijeka, Croatia: InTech.
- Pesic, A., Pesic, D., & Tepavcevic, A. (2012). A new strategic tool for internal audit of the company based on fuzzy logic. *Computer Science and Information Systems*, 9(2), 653–666. doi:10.2298/CSIS111201013P
- Pokoradi, L. (2002). Fuzzy logic based risk assessment. *Academic and Applied Research in Military Science*, 1(1), 63–73.
- Shang, K., & Hossen, Z. (2013). *Applying fuzzy logic to risk assessment and decision-making, Research report sponsored by Canadian Institute of Actuaries*. Retrieved from <https://www.soa.org/Files/Research/Projects/research-2013-fuzzy-logic.pdf>
- Tah, J. H. M., & Carr, V. (2000). A proposal for construction project risk assessment using fuzzy logic. *Construction Management and Economics*, 18(4), 491–500. doi:10.1080/01446190050024905
- Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, 8(3), 338–353. doi:10.1016/S0019-9958(65)90241-X
- Zadeh, L. A. (1978). Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1(1), 3–28. doi:10.1016/0165-0114(78)90029-5
- Zimmerman, H. J. (2001). *Fuzzy Set Theory and Its Applications*. Boston: Kluwer Academic Publishers. doi:10.1007/978-94-010-0646-0

ADDITIONAL READING

- Chan, H. K., & Wang, X. (2013). *Fuzzy Hierarchical Model for risk assessment: principles, concepts, and practical applications*. London: Springer. doi:10.1007/978-1-4471-5043-5
- Dubois, D., Lang, J., & Prade, H. (1991). *Fuzzy sets and systems, fuzzy sets in approximate reasoning, part 2: logical approaches*. NJ: Prentice Hall.
- Gencer, C., Kizilkaya, E. A., & Karahan, Ç. (2010). An Algorithm Predicting Upper Level Icing Potential by Fuzzy Set Theory and an Application with this Algorithm for Turkey. *The Open Industrial & Manufacturing Engineering Journal*, 3, 7–12.
- Klir, G. J., & Yuan, B. (1995). *Fuzzy sets and fuzzy logic: theory and applications*. New Jersey: Prentice Hall.
- Klir, G. J., & Yuan, B. (1996). *Fuzzy sets, fuzzy logic, and fuzzy systems: selected papers by Lotfi A. Zadeh*. River Edge, NJ: World Scientific Publishing Co.
- Rawabdeh, Í., Al-Refaie, A., & Arabiyat, H. (2013). Developing a fuzzy logic decision support system for strategic planning in industrial organizations. *International Journal of Intelligent Systems and Applications in Engineering*, 1(2), 14–23.

Wang, Y. M., Chin, K. S., Poon, G. K. K., & Yang, J. B. (2009). Risk evaluation in failure mode and effects analysis using fuzzy weighted geometric mean. *Expert Systems with Applications*, 36(2), 1195–1207. doi:10.1016/j.eswa.2007.11.028

Wei, S. H., & Chen, S. M. (2009). Fuzzy risk analysis based on interval-valued fuzzy numbers. *Expert Systems with Applications*, 36(2), 2285–2299. doi:10.1016/j.eswa.2007.12.037

KEY TERMS AND DEFINITIONS

Fuzzy Logic: A field of study which centers on the human reasoning process and operates by converting it into mathematical functions.

Fuzzy Inference System (FIS): The system which is a way of mapping an input space to an output space by using fuzzy set theory. FIS uses a collection of fuzzy membership functions and rules to generate an output.

Risk: The possibility of any event occurring that may stop businesses or companies from

achieving their strategic, financial and operational goals.

Risk Assessment: The process of identifying, evaluating and managing potential events and situations to provide a reasonable level of confidence for the achievement of the business goals.

R

ENDNOTES

- ¹ In general, in risk assessment process, a team consisting related process's responsables and the other related staff rates any certain risks and the average of these points reveals risk score. Thus, risk score is commonly a decimal value. However, especially in traditional matrix method, the decimal values are rounded.
- ² Centroid defuzzification returns the center of area under a curve. This center is calculated by getting the average values of x and y axis of the related shape (triangular and trapezoid in this study). Centroid value is referenced to x axis value of the calculated center's axis.

Predictive Analytics and Intelligent Risk Detection in Healthcare Contexts

Nilmini Wickramasinghe

Epworth HealthCare, Australia & Deakin University, Australia

INTRODUCTION

Healthcare costs globally continue to rise and the US in particular is struggling to contain healthcare costs (Wickramasinghe et al., 2015; 2012). This has led most OECD countries to focus on healthcare reform coupled with greater investment in IS/IT to facilitate superior healthcare delivery (Wickramasinghe et al, 2015). At the same time, the federal government has affected policy to emphasize meaningful use of such technology in healthcare (Wickramasinghe et al., 2015, “Meaningful Use,” 2012). Hence, it is now prudent to develop appropriate technology solutions that not only comply with this requirement but also facilitate superior healthcare delivery to ensue.

An area that can particularly benefit from the application of technology solutions is that of equal experience and trying to provide equality of care and access to all citizens. A key area within healthcare disparities relates to access of language services in healthcare or more specifically supporting limited English proficient patients (LEP patients). In particular, English language proficiency should not impinge on access to- and quality of- service for healthcare. Improving access to language services in healthcare has been an ongoing issue that continues to be at the forefront of various healthcare agendas ((Au et al., 2009; Barrett et al., 2008; Chen et al., 2007).

BACKGROUND

Recent discussions on healthcare disparities (Gibbons, 2011) all note the significant potential benefit technology can make in trying to provide

DOI: 10.4018/978-1-5225-2255-3.ch589

an equal experience to all Americans. Sadly, while there are many good points about the US healthcare system, there also exists a considerable racial and ethnic disparity in the delivery of healthcare across the US (ibid). The underlying root causes for these disparities are all amenable to interventions using IS/IT (information systems/ information technology). The thesis of this paper is that technology is well suited to assist is that of limited English proficiency (LEP).

More than 23 million Americans today have limited English proficiency, which in turn has a negative impact on their ability to receive and comprehend appropriate healthcare delivery (Youdelman, 2008; Flores et al. 2008). Integral to the delivery of care is communication between doctor and patient; however, language barriers typically lead to problems such as delay or denial of services, issues with medication management, and underutilization of preventative services (Green et al, 2005; Jacobs et al, 2004; Ghandi et al, 2000; Karliner et al, 2004). The literature suggests that the quality of communication between Although Title VI of the Civil Rights Act 1964 has always required that entities receiving federal funds provide language services to those with LEP, the law has not often been enforced in healthcare settings (Jacobs et al., 2006). However, awareness of the need to provide language services in healthcare has increased in recent years (Gibbons, 2011).

Current Problem

In 2001 The Institute of Medicine has published two key reports “To err is Human” and “Crossing the Quality Chasm”. Taken together these

reports highlight that patient safety should be one of the essential components of high quality healthcare and that patients should not be harmed by the care that is intended to help them. These profound statements have had far reaching impacts to policy reform and efforts to address patient safety and quality of care delivered today. The role of language barriers and their impact on adverse events is thus now also receiving heightened attention. Especially given that research is consistently highlighting that adverse events affect LEP patients disproportionately more and result in serious consequences to the patient.

Approximately 57 Million people or more than 20 percent of the US population speak a language other than English and this figure is growing while approximately 8.6 percent of the population is defined as LEP. Thus at least 8.6 percent of the US population is at risk for adverse events because of barriers associated with language issues. This adds further cost pressures to an already strained healthcare system

To address this problem a technology mediated solution is proffered to provide multi-lingual support at in-take and registration for LEP patients.

Development of the Conceptual Model

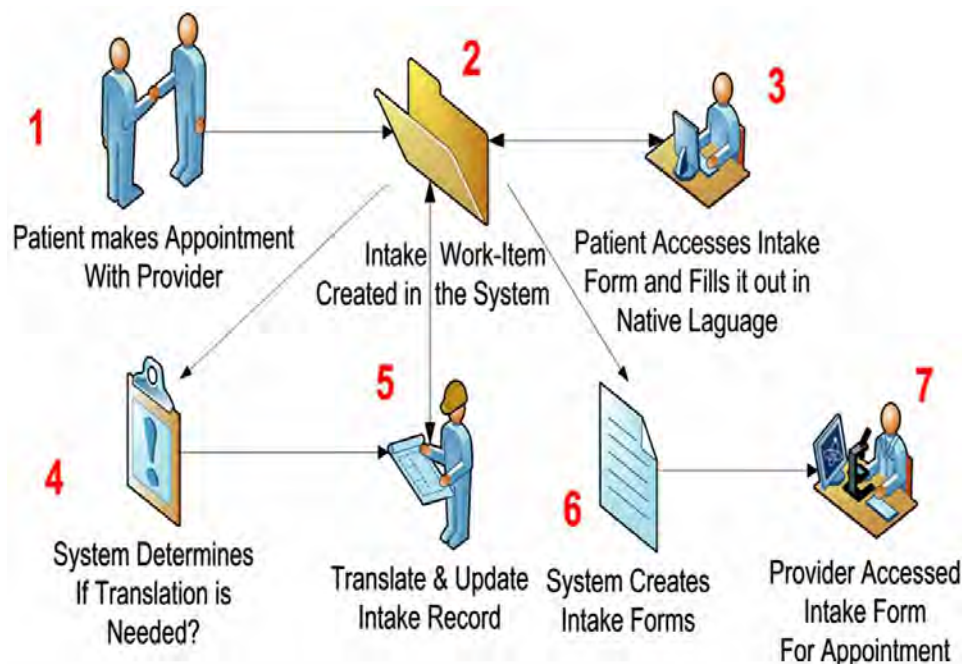
Web 2.0 technology and cloud computing which provides and facilitates anytime anywhere access to (Troshani et al., 2011; Svantesson and Clark, 2010; Mell and Grance, 2010; Gilbert, 2010; Amazon, 2011; Armbrust et al., 2010), affords us the possibility to leverage these technology benefits to design and develop an appropriate solution to address the problem – a portal that support multi language real time translation at intake and registration. Figure 1 provides the conceptual model.

From the above conceptual model in Figure 1, it is possible to model the intake process into critical steps. All these 13 steps of care must be addressed in the technology solution and/or suite of solutions if the solution is to provide the necessary assistance for LEP patients. Taken together these steps traverse the healthcare encounter at intake.

PILOT STUDY

The developed solution was then tested in a small pilot study to assess its reliability and feasibility.

Figure 1. Conceptual model



A 2 arm control study was designed where both English proficient patients and Spanish proficient patients were randomly selected at a leading MidWest Multi-specialty healthcare facility. Before the study commenced appropriate ethical clearances were obtained. Table 1a and b detail the results from the two week pilot trial as well as the activities that were observed in the study.

Results

The study qualitatively measured the interactions between the patients and the electronic kiosk, the value of a clear and concise intake summary for the physician at the beginning of the medical encounter and the ability of the administrative team in the department to take advantage of the completed electronic intake form for other purposes. There was no attempt to integrate the intake form with electronic medical records. Table 2 summarizes the key findings.

FUTURE RESEARCH DIRECTIONS

Two areas are especially critical when considering future developments of the technology mediated solution. Specifically, one area is to assess the health literacy of the patients no matter if they are LEP or proficient English language speakers. The other key direction involves developing and enlarging the languages offered to include Chinese, polish, Farci and German as these are the foreign languages most in prevalent within the US.

CONCLUSION

The US healthcare provider setting is working to deliver the best possible patient care. It must also meet cost of service objectives and provide physicians the best possible tools to meet patient's expectations. Technology in its many applications will continually be employed and adapted to ac-

Table 1a. Study activity

Actor	Before	During Study
Spanish Speaking Patients	Handed an English language intake form and were asked to fill it out to the best of their ability Often had a family member fill out the form, forms were often difficult to read and were incomplete If no family member was present a Spanish speaking staff administrator was taken off their job and asked to serve as interpreter and fill out the form with patient and increased the cost of serving an LEP patient Completed Forms were often difficult to read and of marginal value to physician provider Cultural behaviors limited the ability to gather information, Older Hispanics, men and women are very reluctant to volunteer information Completed form given to administration desk Completed form was placed in patient intake folder and was available to physician, was not available for other purposes Wait times varied as to availability of	Asked to participate in beta study, if agreeable directed to intake kiosk Offered Spanish language form which was translated into English per study description by medically competent interpreter. In some cases family member filled out form in other cases interpreter filled out form. No problems observed in using kiosk technology Form presentation was clear and understandable by patients Completed form given to administration desk Completed form in electronic format was available for other purposes such as Average duration of form completion action was not a factor
English Speaking Patients	Given paper form Filled out form to best of ability Very often completed forms were not easy to read and of minimal value to physician Completed form was placed in patient intake folder and available to physician, was not available for other purposes	Given option of using electronic intake kiosk with drop down menus and input boxes Faster intake form completion Intake form presentation was clear and understandable by patient Completed form was placed in patient intake folder and available to physician and available for other purposes
Physician	Clarity of intake form depended on writing and communication skill of patient	Electronic format provided easy to read and use intake form
Administration	Begins w patient request for appoint and scheduling of appointment, and request for completion of paper forms prior to actual appoint, on appointment date Patients rarely completed intake forms in advance of appointment If patient did not complete intake form prior to visit a blank form was presented to patient at time of visit If patient was a Spanish speaker a family member or staff would assist with intake completion If no family member available administration staff acted as interpreter. Completed form was used presented to administrator	Administration asked English or Spanish speaking patients if they would be interested in completing an intake form electronically and in native language. Once form was complete either handwritten or electronic it was returned to the administration and placed in the intake folder for clinical visit with physician.

Table 1b. Control – English speaking patients

Actor	Event	Outcome
English Speaking Patient	English speaking patients are offered option of traditional paper intake form or completing the form at kiosk with touch screen and drop down menu capabilities.	The Department of Urology serves an aged skewed population. As a general rule the younger the patient the more likely they were to opt for the electronic format. 80% of the patients offered the electronic option chose for to use that. Several patients commented that the electronic version was much faster to use and more consistent with a modern healthcare setting. Other inquired if the electronic form was transferable to other departments. For those patients that chose the electronic version patient satisfaction ranged on the scale of 4 or 5 with 5 being the highest on a 1 to 5 scale.
Spanish Speaking Patient	Spanish speaking patients are offered option of the traditional paper intake form or completing the form at kiosk with touch screen and drop down menu capabilities. The hand written intake form provided by the administration was not translated into Spanish, consequently each Spanish speaking patient opted for the electronic version of the form offered in their native language.	In each case at the electronic kiosk, translator services were incorporated to facilitate the completion of the intake form. Spanish speaking patients were afforded the opportunity to complete the intake in their language while the administration and physician were provided an English language version of the intake form. In each case customer satisfaction was rated a 5 on a 1 to 5 scale. Multiple copies of the completed intake form given to each patient for potential use in other departments.
Physician and Patient	English and Spanish speaking patients were queued to meet with the physician in a business as usual manner.	Physicians who had access to the electronic forms regardless of language spoken by the patient rated the clarity of the information on the electronic form higher than the hand written form. Also, physicians stated that the quantity of the information was greater and offered a more clear picture of the patient medical, family and social status. Physician satisfaction with the electronic form over the handwritten form was ranked higher because the electronic form was easier and faster to read. See appendix xx for sample of handwritten and electronic form.
Administration and Patient	Once the intake forms were completed (English and Spanish), they were handed to the front desk for administration, placed in the patient intake folder and placed in queue for the physician. The form is reviewed by the physician. It is generally not used beyond physician review for other purposes.	After the physician encounter patients were often asked to complete other forms very similar in content to the original intake form. Form redundancy is a major detractor to patient satisfaction and high cost component to administrative efficiency. No other forms have a language assistance component. The impact of this is obvious i.e. when a dietary requirement is needed before a test can be conducted and this requirement is not met on the day of the scheduled procedure because the instructions were not understandable by the non-English speaking patient.

Table 2. Summary of findings

Actor	Event	Outcome
English Speaking Patient	English speaking patients are offered option of traditional paper intake form or completing the form at kiosk with touch screen and drop down menu capabilities.	The Department of Urology serves an aged skewed population. As a general rule the younger the patient the more likely they were to opt for the electronic format. 80% of the patients offered the electronic option chose for to use that. Several patients commented that the electronic version was much faster to use and more consistent with a modern healthcare setting. Other inquired if the electronic form was transferable to other departments. For those patients that chose the electronic version patient satisfaction ranged on the scale of 4 or 5 with 5 being the highest on a 1 to 5 scale.
Spanish Speaking Patient	Spanish speaking patients are offered option of the traditional paper intake form or completing the form at kiosk with touch screen and drop down menu capabilities. The hand written intake form provided by the administration was not translated into Spanish, consequently each Spanish speaking patient opted for the electronic version of the form offered in their native language.	In each case at the electronic kiosk, translator services were incorporated to facilitate the completion of the intake form. Spanish speaking patients were afforded the opportunity to complete the intake in their language while the administration and physician were provided an English language version of the intake form. In each case customer satisfaction was rated a 5 on a 1 to 5 scale. Multiple copies of the completed intake form given to each patient for potential use in other departments.
Physician and Patient	English and Spanish speaking patients were queued to meet with the physician in a business as usual manner.	Physicians who had access to the electronic forms regardless of language spoken by the patient rated the clarity of the information on the electronic form higher than the hand written form. Also, physicians stated that the quantity of the information was greater and offered a more clear picture of the patient medical, family and social status. Physician satisfaction with the electronic form over the handwritten form was ranked higher because the electronic form was easier and faster to read. See appendix xx for sample of handwritten and electronic form.
Administration and Patient	Once the intake forms were completed (English and Spanish), they were handed to the front desk for administration, placed in the patient intake folder and placed in queue for the physician. The form is reviewed by the physician. It is generally not used beyond physician review for other purposes.	After the physician encounter patients were often asked to complete other forms very similar in content to the original intake form. Form redundancy is a major detractor to patient satisfaction and high cost component to administrative efficiency. No other forms have a language assistance component. The impact of this is obvious i.e. when a dietary requirements is needed before a test can be conducted and this requirement is not met on the day of the scheduled procedure because the instructions were not understandable by the non-English speaking patient.

comply with these goals. Technology that facilitates communication between the patient and the provider (physician and administration) will have an important role. All healthcare interactions begin and end with communication between the provider and patient. This is a determining factor in successful health outcomes. With greater clarity and availability of patient information there is a greater probability of meeting patient need and managing the cost of delivering service. For example, it is well documented that language barriers (communication barriers) between patients and providers are a leading cause of healthcare disparities. Communication barriers are not limited to the interactions between the patient and physician but also include the patient's interactions with the physician supporting healthcare infrastructure.

Communication Barriers Impact

Patient: Physician communication, for example,

1. The inability for physicians to understand a patient's need and thus increase the potential for medical errors and the associated higher cost of service.
2. Poor communications between the physician and patient extends the healthcare encounter and creates delays in physician access.

Patient: Administration communication, for example,

1. The ability for the patient to follow with procedural or pharmacy instructions,
2. The ability for patients to provide or understand billing or insurance information and
3. the ability to follow up with next steps in a treatment protocol which leads to higher cost of service.

This study represents a first step to understand the impact of using an electronic intake format for gathering patient information from both English speaking and non-English speaking (LEP) patients. Healthcare disparities plague the US healthcare system and language proficiency is one of these key barriers. Technology does have the potential to reduce this barrier effectively and

efficiently so that superior care delivery might ensue. The proffered solution described in the preceding holds the key to potentially addressing this challenge.

REFERENCES

Amazon. (2011). *Announcing Amazon SimpleDB in the EU Region*. Available: <http://aws.amazon.com/about-aws/whats-new/2009/09/22/announcing-amazon-simplifiedb-in-the-eu-region/>

Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., & Konwinski, A. (2010). A view of cloud computing. *Communications of the ACM*, 53(4), 50–58. doi:10.1145/1721654.1721672

Au, M., Taylor, E., & Gold, M. (2009). *Improving Access to Language Services in Health Care. Policy Brief Mathematica*. Policy Research Inc. White paper.

Barrett, S., Dyer, C., & Westpheling, K. (2008). *Language Access: Understanding the Barriers and Challenges in Primary care Settings*. McLean, VA: The Association of Clinicians for the Underserved.

Chen, A., Youdelman, M., & Brooks, J. (2007). The Legal Framework for Language Access in Healthcare Settings: Title VI and Beyond. *Journal of General Internal Medicine*, 22(S2), 362–367. doi:10.1007/s11606-007-0366-2 PMID:17957427

DrTango Inc. (2011). *A Comparison of DMHC and CDISB853 Requirements*. Available online at <http://www.sb853.com/pdf/regulations-Table.pdf>

Dutta, S., & Mia, I. (2011). *The Global Information Technology Report 2010–2011*. Geneva: World Economic Forum and INSEAD.

Etro, F. (2010). The economic consequences of the diffusion of cloud computing. In S. Dutta & I. Mia (Eds.), *The Global Information Technology Report 2009–2010*. Geneva: World Economic Forum and INSEAD.

- Flores, G. (2005). The Impact of Medical Interpreter Services on the Quality of Health Care: A Systematic Review. *Medical Care Research and Review*, 62(3), 255–299. doi:10.1177/1077558705275416 PMID:15894705
- Flores, G., Torres, S., Holmes, L., Salas-Lopez, D., Youdelman, M., & Tomany-Kprman, S. (2008). access to Hospital Interpreter Services for Limited English Proficient Patients in New Jersey: A Statewide Evaluation. *Journal of Healthcare for the Poor and the Underserved*, 19(2), 391–415. doi:10.1353/hpu.0.0007 PMID:18469412
- Gandhi, T., Burstin, H., Cook, E., Puopolo, A., Haas, J., Brennan, T., & Bates, D. (2000). Drug complications in outpatients. *Journal of General Internal Medicine*, 20(3), 149–154. doi:10.1046/j.1525-1497.2000.04199.x PMID:10718894
- Gibbons, M. (2011). Use of Health Information technology Among Racial and Ethnic Underserved communities. *Perspectives in Health Information Management*, 8(Winter). Available at <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3035830/>
- Gilbert, F. (2010). Cloud service contracts may be fluffy: selected legal issues to consider before taking off. *Journal of Internet Law*, 14(6), 17–30.
- Green, A., Ngo-Metzger, Q., Legedza, A., Mas-sagli, M., Philips, R., & Iezzoni, L. (2005). Interpreter Services Language Concordance and Health Care Quality. *Journal of General Internal Medicine*, 20, 1050–1056.
- Institute of Medicine. (2003). *Unequal treatment: confronting racial and ethnic disparities in health care*. Washington, DC: The National academic Press.
- Interpreter Services Work Group Report to Minnesota Legislature. (2008). *Interpreter Services for Patients in Medical and dental Settings*. Available online at <http://www.health.state.mn.us/dics/ppq/hci/ISWGreport08.pdf>
- Jacobs, E., Shepard, D. S., Suaya, J. A., & Stone, E.-L. (2004). Overcoming Language Barriers in Health care. *American Journal of Public Health*, 94(5), 866–869. doi:10.2105/AJPH.94.5.866 PMID:15117713
- Julisch, K. & Hall, M. (2010). Security and control in the cloud. *Information Security Journal: A Global Perspective*, 19(6), 299–309.
- Kallick, D. (2007). *Working for a better Life*. New York: Fiscal Policy Institute.
- Karliner, L., Perez-Stable, E., & Gildengorin, V. (2004). The Language Divide. *Journal of General Internal Medicine*, 19(2), 175–183. doi:10.1111/j.1525-1497.2004.30268.x PMID:15009797
- Meaningful Use. (2012). Retrieved from <http://www.healthit.gov/policy-researchers-implementers/meaningful-use>
- Mell, P., & Grance, T. (2010). The NIST definition of cloud computing. *Communications of the ACM*, 53(6), 50.
- National Committee for Quality Assurance. (2008). *Standards aimed at Improving Health care for Diverse Groups*. Available online at <http://www.ncqa.org/tabid/916Default.aspx>
- Svantesson, D., & Clarke, R. (2010). Privacy and consumer risks in cloud computing. *Computer Law & Security Report*, 26(4), 391–397. doi:10.1016/j.clsr.2010.05.005
- Troshani, I., Wickramasinghe, N., & Rampersad, G. (2011). *On cloud Nine? An Integrative Risk Management Framework for Cloud Computing*. Bled eConference.
- Wickramasinghe, N., Bali, R., Kirn, S., & Reima, S. (2012). *Creating Sustainable e-health Solutions*. New York: Springer. doi:10.1007/978-1-4614-1536-7

KEY TERMS AND DEFINITIONS

Cloud-Based Computing (or Cloud Computing): When computing resources such as software and hardware are delivered as a service over a network.

Healthcare Delivery: The network, structure and stakeholders involved in administering medical care to individuals.

Human Interpreter: Person who performs language translation.

Language Translation: The act of converting the meaning into a different language.

LEP-Limited English Proficiency: Individuals for whom English is not their native language and thus have difficulty communicating and understanding in English.

Meaningful Use: Federal government stipulation for technology solutions in healthcare.

Real-Time Translation: Translation that happens instantaneously at the point of action.

Stress Testing Corporations and Municipalities and Supply Chains

R

Frank Wolf
CSSTA L3C, USA

INTRODUCTION

Assuring the likelihood of survival is the fundamental reason for stress testing large banks in Europe and the United States under the mandates called Basel I, II and III, and CCAR respectively. Similarly stress testing corporations, not-for-profit organizations, and municipalities, has become a necessary exercise to prevent insolvency in times of economic uncertainty and shortened corporate life expectancy.

The concept of stress testing is not entirely new. For example, computer software gets routinely stress tested in order to assure it can survive volume fluctuations, data storage requirements and hacking attempts. In the field of engineering, stress testing materials under conditions of extreme heat, cold, or loads has been around for ages. In the banking sector of our economy stress testing means having sufficient capital to meet current obligations and even prevent runs on your neighborhood ATM. So the logical extension of this reasonable practice is to stress test mid-sized or large firms, non-profits and municipalities with respect to solvency, with the intent to preserve public services, commercial products, peoples' jobs, invested capital, supplier and customer relationships and quite possibly preserve whole communities!

It is very well established in the academic and business literature that both insolvency and bankruptcy are predictable phenomena for the vast majority of cases.² However, testing for those conditions is not yet customary for businesses and municipalities. During the 1960s managers

presumed that a certain percentage of business failures were to be expected and that these constituted the risk of market entry. More recently the Harvard Business Review⁹, dedicating an entire issue to business failure, postulating that *not* succeeding constitutes a learning experience which one should embrace and not be ashamed of. The clear message is this: learn from your failure and then try again, only do better next time. That is acceptable in theory, just don't tell this to a failing entrepreneur with a great idea who has just mortgaged his house and put his entire future at risk. Failure is not a good thing and many types of failures are both predictable and most importantly preventable. The fiduciary responsibility for preventing business insolvency rests primarily with directors and corporate officers, individually and collectively. States are on the line for municipal failures under Chapter 9 of the U.S. Bankruptcy Code irrespective of whether Chapter 9 applies. States do bear a large reputational risk as future borrowing costs may increase.

BACKGROUND

Legal Considerations for Stress Testing Corporate Entities for Insolvency Risk

Insolvency

Bankruptcy is a legal system set up by the U.S. Congress to administer insolvency of individuals or corporations. That should not be mistaken for

business insolvency which is commonly described as the financial distress caused when either or both:

1. An entity's debts exceed the assets it owns ("balance sheet insolvency"); or
2. When the party does not have the cash, or liquidity, to meet its payments as they become due ("cash flow insolvency").

There are other ways outside the bankruptcy system to attempt to resolve insolvency. Although the code offers many advantages to an entity in financial distress, it could be in the best interest of the company to avoid the formal bankruptcy process, and instead attempt to reorganize outside the federal court system.

The bankruptcy process is extremely costly, kicks in too late, and is time consuming and frustrating to debtors and creditors alike, as the process returns very little to creditors that are unsecured or even to those with priority treatment.

The low plan confirmation and low plan completion rates, and the high conversion and high dismissal ratios, indicates that early detection of insolvency risk would be hugely beneficial. The implementation of effective and efficient preventive measures should be emphasized and promoted, if not required, as part of good corporate governance. In other words, corporate stress testing may soon become a mandate for public corporations!

By implementing a program that helps to identify early-on the risk of insolvency, a company may avoid the many uncontrollable pitfalls that could occur with a late bankruptcy Chapter 11 reorganization filing.

Once a business is within a judicial bankruptcy process, it may lose control of the possibility of dismissing the petition even though it was voluntarily filed, and instead face a forced conversion and a liquidation scenario with a possible hostile trustee and disgruntled creditors.

Statistics show that the advantages of a bankruptcy filing include:

1. The automatic stay of most, but not all, litigations and
2. The possibility of removing or discharging some unsecured debt.

Finally the system can be "gamed" in terms of favoring the secured and some other preferred parties to the detriment of the unsecured and even priority creditors, who receive on the average, a very small distributions from the estate assets.

For example, in chapter 7 liquidations, according to the official court data for calendar year 2008, and based on the 63,000 Chapter 7 cases with assets, closed in that period with over \$2.6 billion in receipts, secured creditors received 28.9% of total amounts distributed; priority creditors received 4.4%; unsecured creditors 25%; and debtors and third parties 8% for a total distribution of \$1.785B or 66.2%. At the same time the total fees paid out to the professionals in these asset cases, mostly legal, trustee fees and administrative charges were 33.8% of the total distributions made. (Haverstock and Crewson, 2012)

However, these figures do not consider that in no-asset cases, which are nearly 97% of all Chapter 7 matters, unsecured creditors receive almost no distribution.

The legal considerations notwithstanding, statistics often quoted (Reeves¹) find that today's public companies have a 1:3 chance of disappearing within five years due to a combination of bankruptcies, liquidation, delisting, mergers and acquisitions. That disappearance rate is 6 times larger than it was 40 years ago and calls for a governance model (Reeves²) of planning for multiple time scales in which stress testing is a good start.

Corporations

The U.S. Bankruptcy Code provides corporations with various reorganization options that must be considered in today's complicated business environment. These options may include corporate

reorganization, recapitalization or sale of substantially all or some assets, and possibly liquidation of an entire business. However, it also entails hidden pitfalls that may outweigh its benefits.

Whether one chooses a straight Chapter 7 liquidation process, or reorganization under Chapters 11 or 13, private businesses that approach the nebulous zone of insolvency, or that are at risk of falling into such an area, must consider the various scenarios available to avoid getting caught in what could be a precarious bankruptcy judicial process.

Also, prepackaged reorganizations, or out of court debtor-creditor arrangements that avoid the time and cost of the judicial code administration, must be understood and considered.

The complexity arising from options the Code provides to individuals, corporations and partnerships facing business failure is enormous, and requires a great deal of expertise and time to sort out. Not only is research and planning involved in the pre-bankruptcy stage, but also complying with the filing requirements and the regulatory provisions under the joint judicial-administrative procedures of Chapter 11. The task is daunting, even for a firm that can afford the time and expense involved in advanced planning and compliance. (UCLA-Lopucki, 2012)¹⁴.

Although “success” in terms of a Chapter 11 bankruptcy case is difficult to define, it has been measured by the time taken to complete a case, to get the plan of reorganization confirmed, whether the company eventually emerges as a stand-alone company, or if the debtor had to re-file after a short period of time.

These measurements will vary according to the type of bankruptcy and chapter selected to be filed. Also, many so called “reorganizations” actually constitute business liquidations and involve a sale of most or substantially all the assets of the company under the Chapter 11 judicial reorganization umbrella.

Such measurements do not necessarily indicate the degree of failure or success of a case that has

been converted to Chapter 7 liquidation, or dismissed by the court for cause, and whose assets, employees and clients get reintroduced somehow into the economy and become productive parts outside the judicial bankruptcy system.

The federal bankruptcy courts received, during the 12 months period ending in March 2012, a total of 1,367, 006 new filings, of which only 46,393 constituted business filings for which corporate stress testing may have been beneficial. That is, cases in which the debtor is a corporation, partnership, or if the debt is primarily related to a business of some kind belonging to this category. This in contrast to consumer-individual cases (non-business) in which debt was incurred primarily for personal family or household purposes. (Administrative Offices of the U.S. Courts, 2012).¹ The number of Chapter 11 cases stood at only 11,339, but of these only 9,616 constituted business filings and 1,723 were non-business debt related.

In terms of comparing different outcomes and success rates between the various bankruptcy code filing options, the numbers show that only a third of the Chapter 11 cases are ever confirmed or approved by the court; that 22% are converted to Chapter 7 liquidation; and that over 35% are dismissed. That is, they never get to the confirmation stage and instead are resolved somehow outside the judicial bankruptcy system.

After the 2005 amendments to the U.S. Code, the average disposition time of confirmation for all Chapter 11 cases filed still exceeds twelve months from date of filing of the petition, and the majority of the cases get converted to Chapter 7 liquidation in nine months and / or dismissed in over seven months. (Flynn and Crewson, 2012)

These official statistics show, that Chapter 11 reorganizations cases filed are rising moderately, the rates are not entirely uniform among the different chapters, they vary among the different states and geographical areas of the nation, and that a very low percentile of cases get to the ideal approval stage of judicial confirmation of a plan

of reorganization. Moreover, about 55% never get confirmed and instead are dismissed or converted to a forced liquidation chapter of the Code.

It is reasonable to infer from these statistics that the financial carnage from business failures is widespread with lasting effects beyond its epicenter. Given an increased fragility in the business environment, preventing the disastrous effects of corporate failures through a well- documented predictive stress testing methodology is called for.

Municipalities

As banner headlines now routinely announce, nations can and do become insolvent, recent examples being Russia, Mexico, Greece, Spain, Ireland and others. Failed US cities include Detroit MI, Hillview KY, San Bernadine CA, Stockton CA, Jefferson City AL, Central Falls RI, and Boise City ID, among others. Since nations cannot be dissolved, the sovereign settlement mechanism that has evolved post WWII is called the Paris Club. U.S. States, with the exception of Vermont, are state- constitutionally bound to balance their budgets. But that usually only applies to the General Funds accounts, into which most tax income flows and from which operational expenses are paid. Many other forms of commitments are not included, and therein is the problem of pending insolvency of municipalities.

Municipalities may qualify under the U.S. Bankruptcy Code Chapter 9, where Vermont and 24 other States are again the exception. For those exceptions, the respective legislature must enable Chapter 9 for a designated town. The Chapter 9 rules are slightly different for a municipality, for example:

1. A town cannot involuntarily be put into Chapter 9 as it only applies to public agencies or instrumentalities of the state.
2. Judges in these matters are assigned by the chief judge of the circuit rather than the clerk.
3. The debtor bears the burden of proving insolvency.

4. Actions against the municipality or its individuals are automatically stayed.
5. Collective bargaining agreements can be modified under the rules, but not unilaterally abrogated, etc.

The net effect in all those Chapter 9 cases sadly is always the same. Once filed, money ceases to flow and can't be raised for new projects, economic activity leaves town, and only few investment can be made. Clearly an undesirable conditions similar to commercial insolvency, which does not happen overnight and which can be predicted, and in many cases steps can be taken to prevent total collapse.

THE CASE FOR STRESS TESTING CORPORATE ENTITIES, INCLUDING NON-PROFITS AND SUPPLY CHAINS

The process of stress testing corporations differs from that of stress testing banks which is governed by Basel I, II, and III rules (EU) and CCAR (US). The difference is that bank stress testing is mandated by governmental regulators worldwide, while in the corporate world stress testing is only advisable. To begin with, the stress testing process, if done well, is organization specific and includes a quantitative analysis based on current conditions, followed by running future scenario models, which are both quantitative and qualitative. A reasonable approach is to first conduct a ratio analysis, as follows:

1. For the target enterprise collecting financial operating data about a company and its major competitors for three to five prior years is essential. Sources are quarterly or year-end balance sheets and operating statements. The next step involves computing of ratios such as the solvency ratio, quick ratio, current ratio, and several more as may be appropriate for a specific case under review. Doing the same for all competitors establishes a basis for comparisons, a benchmark or standard.

2. Insolvency and bankruptcy for most cases are predictable². Producing graphical output of ratios over time often establishes a trend line and some insight into the client, their competitors and the industry averages.
3. Compute the Altman Z-Score², which is a multivariate regression formula of financial ratios that correctly predicts bankruptcy two years out between 70% and 80% of the time. The value Z is a number meaning standard deviations above or below the mean.
4. Risk scenario: This stage in the analyses needs to be reviewed with corporate management in order to chart a path for developing alternative and plausible scenarios.
5. Scenarios are likely future events that could lead to financial distress and eventual business failure. The importance of correctly structuring and modeling these scenarios cannot be overemphasized, as they can provide actionable insights to predicting failure and to its prevention.

Scenario development involves the following steps: (Schoemaker, 1995)¹⁸

1. Scope, dealing with time frame, market and geography
2. Identification of stake holders and their interests, suppliers, clients, employees
3. Identification of trends, economic, scientific, and political as appropriate
4. Identification of uncertainties, management, elections, interest rates, liquidity
5. The initial scenario should focus on positives and all negatives to build clusters and develop themes, to assess the level of preparedness by management for an uncertain future.
6. Plausibility check, look for inconsistencies and lack of a compelling story.
7. Identify research needs to tackle unknowns, such as effects of cloud computing or destructive new technologies
8. Iteratively evolve into defining decision final risk scenarios to be modeled.

Additional risk factors bearing on the solvency also have to be quantified, including:

- Pending law suits as a claimant or as a respondent during the last 3-5 years?
- Discontinued operations during the last 3-5 years, or plans for future discontinuation.
- Write-offs or impairments during the last 3-5 years?
- Debt defaults during the last 3-5 years?
- Having sought financial partners, disposed of operations, last 3-5 years?
- Is a management succession plan in force?
- Internal control issues, accounting, fraud history?
- Technology issues, obsolescence, creative destruction, electronic hacking risk?
- Competition issues, strength-weakness-threats-opportunities,
- Distribution of clients, Pareto problem, 80%/20%?
- Regulatory issues regarding safety, human resources, environment, sustainability
- Supply chain disruptions as a function of world events, partner insolvency, force- majeure?
- Liquidity, credit and market risks
- Globally changing trade compliance due to international trade agreements or WTO decisions
- Demand volatility, increasing or decreasing, social media.
- Talent management and leadership development to accommodate new technologies
- Transportation network design, trip optimization model opportunities for productivity gain, production bottle necks
- Working capital adequacy to meet corporate growth plans in the near future, including that of supply chain partners, orderly retrenchment.
- Interest rates' effect on finance, Forex
- Inflation effect on performance

Given likely future environments and probable internal risk factors, a simulation model is built in order to answer the “what if” questions that can help management prevent insolvency. Corporate officers and especially Board members have fiduciary responsibilities to shareholders when solvent. When insolvent however, Board responsibility extends to all stakeholders and especially creditors. Hence Board members must be the drivers to institute stress testing in companies to meet their fiduciary duties and to protect themselves from possible legal action.

THE CASE FOR STRESS TESTING MUNICIPALITIES

The object of stress testing municipalities is very similar to that of stress testing corporations, namely to prevent insolvency and bankruptcy under possible stressful future conditions. The difference between corporate and municipal stress testing is in the measurements and the nature of the organization. For one thing, municipalities cannot be dissolved in a bankruptcy process because people still live and work in those locations. There are also differences in the data and organization; municipalities are entities for public service and have a fiduciary responsibility towards tax paying citizens vs. corporations having responsibility to dividend expecting shareholders among others. The final obvious difference is that municipal managers are often elected on issues other than managerial and financial skills, while corporate managers are appointed often on the basis of financial performance and business acumen.

Philosophically the process is similar to that of corporate stress testing, namely collecting data, running statistical analyses and simulating future plausible scenarios. However, there are significant differences between corporate and municipal stress testing approaches. One approach is to consider stress testing state/province-wide and all its municipalities within. Demographic data

is readily available from the U.S. Census, such as town population, income- and age distributions. Each town will most likely have a budget and an annual audit for the current year, all of which would under normal circumstances be publicly available. Consistency of definition of variables from one town to another has to be painstakingly established.

Establishment of a state-wide database combining demographic Census data with municipal financial data is the first step. Classifying towns as either urban, urban ring, rural or suburban offers more chances of segmentation later on. The next step is to develop fiscal stress indicators based on (a) town financial position, (b) town financial flexibility and (c) town tax capacity. An overall index will segment each town in a State into 1-5 rating insolvency risk levels, with 5 being the financially weakest and therefore subject to simulated stress scenarios. Suggested actual measurements for categories (a), (b) and (c) are the following: (Rhode Island, 2011)¹⁷

Financial Position:

- General fund balance at end of fiscal year, dollars
- Quick ratio, which is (total cash and investments)/(notes, loans and liabilities)
- Unfunded pension liability, dollars and percent of budget, per capita
- Annual required contribution to pension funds, dollars

Financial Flexibility:

- State aid, dollars and percent of budget
- Education state aid, dollars per student
- Total property tax revenue, dollars and percent of budget
- Growth in services, dollars
- Debt/revenue ratio
- Employment total benefit payments

Tax Capacity:

- Real estate tax rate
- Per capita tax of owner occupied residential real estate
- Per capita tax of owner occupied commercial real estate
- Median family income

Sponsorship of such studies lies with the States themselves, as they are first in line for assistance. Municipalities have a self-interest in conducting stress tests by an independent outside consultant.

The end result is in the prevention of another Detroit or Puerto Rico.

A CORPORATE STRESS TEST STUDY EXAMPLE

The following is an abbreviated case study from 2012, performed by Corporate Solvency Stress Testing L3C.

The client firm under review is a 75 year old family owned enterprise in the business of transporting freight and packages for supply chains, mostly by truck. It operates in the North Eastern States of the US and has annual revenue below \$100 million. Company assets consist of a large fleet of owned trucks that are mostly ten years or older which in some cases have over 4 million miles on their odometers. Other assets are a network of leased terminals, fully equipped repair facilities and the usual office equipment, plus company fuel pumps at each terminal. Triggered by stress in loan repayment and subsequent deteriorating banking relationship, personnel issues, an auditors' going concern statement, rivalry among two generations of family members working at the firm, a consultant was engaged to resolve the short term stress.

Problem

Multiple small problems can blur the vision to see the big picture. Solving a number of short

term crises may restore tranquility temporarily, however, the root causes of the problem must eventually be solved to assure long term growth and stability. The larger problem in this case was to reduce debt and put the enterprise on a more stable operational platform. The restructuring consultant quickly embarked on several employee terminations and replacement hires for critical financial management positions. In addition, a new banking relationship was established resulting in calming of corporate management, who then agreed to undertake a corporate stress testing regimen. The primary goal of the stress testing process was to assess the firms' survivability under plausible adverse conditions. The time frame to complete this assignment was approximately three months and included the following steps:

1. A three prior year financial ratio analysis based on easily available balance sheets and operating statements.
2. A similar financial analysis of the three largest competitors. These data came from publically available balance sheets and operating statements.
3. A side-by-side comparison of those four firms over the last three years with respect to financial ratios.
4. Developing and running a bankruptcy prediction model for client and competitors and a side by side comparison of all.
5. An interim presentation to management was scheduled, and from it evolved a number of plausible scenarios for the next three years. These quantitative scenarios addressed a combination of cost issues like (a) fuel cost, (b) interest rates, (c) maintenance (d) fuel surcharge, plus current year optimistic and pessimistic budget assumption. Qualitative issues were researched and discussed including competitor behavior changes, fleet replacement buy vs. lease, shipping trends and market share, network optimization, and potential benefits from new logistics technology.

6. Three scenarios were modeled, based on the current budget, and cash flow projections. Monte Carlo simulation gave indicators for future outcomes.

Scenarios

The first dealt with the possibility that fuel costs increase 20% with the market only being able to absorb 40% of that. In another scenario fuel prices, interest rates and maintenance costs, were simultaneously changed from current budget value. In the last scenarios, fuel surcharges were eliminated.

In the Monte Carlo simulation, revenues and fuel surcharges were subjected to sampling from normal distributions to produce a total revenue estimate. Fuel, maintenance and fixed expenses were sampled from their respective normal distribution and combined with the revenue and depreciation to produce cash flow estimates for the current fiscal year. The simulation was run for 255 working days and aggregated to show yearly levels.

Results

The client was deemed marginally solvent based on the solvency related ratio and the bankruptcy prediction model Z-score. However, the client's current ratio was below the industry average giving it a relatively low credit rating of BB. For each of the six runs net profit estimations were generated from which the narrow line between profitability, and or loss became painfully apparent. For the Monte Carlo simulation the measurement was EBITA. The results show that efforts of the short term fix is no fix. While short term fixes were necessary, they offer only momentary reprieve and that continuous monitoring is required. The client will need to formulate a longer term strategy to assure long term stability and hopefully achieve moderate growth. Beyond the strategy to get out of debt over a period of time, there will be many projects that need to be professionally managed. These projects involve the evaluation of the cur-

rent network of terminals, fleet replacement and its effect on maintenance and operations cost, etc.

The final step was an Opinion Letter to the Board, in which case the firm was found to be solvent for now, which means that its immediate ability to pay current obligations appeared secure. Long term profitability, however, could only be secured with the Board making a combination of simultaneous expansion and contraction decisions to establish a new growth path. Stress Test related Opinion Letters by third parties are neither accounting nor legal documents but are helpful in showing good faith and governance.

Having gone through this process, the Board was able show good governance practices should the company become insolvent sometime in future. A future bankruptcy court would be less likely to reverse past management decisions made while the company was certifiably solvent.

FUTURE RESEARCH DIRECTION RELATIVE TO INSOLVENCY RISK MANAGEMENT

While stress testing corporations and municipalities for solvency is relatively new, research to establish good measurements and processes will need to be developed in order for this preventive medicine to affect society. Not only is there a burning need to refine the process and make corporate and municipal stress testing as routine as seeing a dentist every six months, there is also a need to understand the forces that resist stress testing efforts, namely the notion that ignorance is bliss. In the corporate market, those forces operate mostly in mature organizations with an ingrained management team. For municipalities those forces are elected managers without management skills.

For corporations and non-profits, there is a pressing need to standardize the process of corporate stress testing and find appropriate measurements. Third party advisors in cooperation with trade associations are suitable entities to undertake such projects. Public accountants

and bankruptcy law firms are too close to their clients to be neutral, and certainly can't write a "Letter of Solvency" likely to stand up in court.

For Board members the responsibility is clear. Their personal liability is such that if knowingly or unknowingly insolvent, their responsibility widens to include creditors, customers and other stakeholders, rather than just shareholders.

With municipal failures on the rise, possibly in Puerto Rico, there is a danger that Chapter 9 bankruptcy will increasingly become a strategy for fiscal relief, just as Chapter 11 became strategy for airlines at one time not too long ago. While a model processes for stress testing municipalities exist for New York State for example, the research goal should be to simplify the process and standardize the measurements so that the municipal stress testing can be done by the municipalities themselves. In the interim, third party stress test advisors are appropriate sources to tailor the process to geographic uniqueness. These models and methods should become part and parcel of the municipal budgeting process with special emphasis on the capture of off-book commitments.

For corporate supply chain partners it is obvious that financial weakness of one member affects all. It is clear that more work needs to be done to assess supply chain solvency risk as well as those risks that fall under the category of force-majeure.

The technology is rapidly approaching the state in which a *software agent* will be able to continuously do the mechanical stress testing activities for a single firm, supply chain, or municipality. In particular to capture the balances on key financial statements, compute financial ratios relative to debt, assets, cash flow etc., to compare those with comparable enterprises from public data bases, and create a management dashboard of solvency risk change. While risk scenarios are unique to each enterprise, even a data bank of scenarios can be defined and stored for semi-automatic Monte Carlo simulations with defined parameters to answer *what-if* questions. The end result and ultimate objective of permanent stress testing can and should be automated and become part of perpetual good governance.

CONCLUSION



The field of corporate and municipal solvency stress testing is an evolving technology with strong implication for a social good. The social good in cases of both corporate and municipal insolvencies is in the preservation of jobs, beneficial use of technology and strengthening services to communities, fostering viable investments and avoiding enormous waste due to preventable insolvencies and bankruptcies. The social good is also in leading to managerial preparedness to face rapid changes in the economic environment head on, in other words, good governance. But, it is hard to convince managers to be prepared for events that may not occur; until they do occur.

In this paper, the authors established the following:

1. A state of insolvency in both corporations and municipalities can be defined, and that insolvencies are a leading indicator for bankruptcy, which is statistically predictable.
2. A meaningful recovery of value from formal bankruptcy is unlikely, as the process kicks in too late and is too costly; leading to the conclusion that preventive mechanisms in the form of corporate and municipal stress testing are needed.
3. Stress testing processes for corporations, whole supply chains, and municipalities differ from each other with respect to applicable measurements and other parameters.
4. The ultimate vision is for stress testing processes to become part of a predictive analytics repertoire for both corporations supply chains and municipalities. Such processes must be easy to use and understand at the citizen level.

In this paper, the authors have presented an interdisciplinary approach to the process and the measurements for corporate and municipal fiscal stress testing. In addition, the authors presented a case for the advisability to undertake such actions sooner rather than later in light of continuing uncertainty.

REFERENCES

- Administrative Offices of the U.S. Courts. (2012). *Bankruptcy data source, Table F, Business and non-business cases commenced by chapter of the bankruptcy Code, 12-months period ending March 31, 2012*. Author.
- Altman, E. (1968). Financial ratio discriminant analysis and the prediction of corporate bankruptcies. *The Journal of Finance*, 23(4), 589–609. doi:10.1111/j.1540-6261.1968.tb00843.x
- Bankrupt24.co.za. (2012). Retrieved from <http://www.bankrupt24.co.za/7/24/2012>
- Chung, C. (2015). Zombieland/The Detroit Bankruptcy. *Fordham Urban Law Journal*, 41(3).
- Cihak, M. (2007). *Introduction to Applied Stress Testing*. International Monetary Fund, IMF Working Paper 07/59.
- Crook, J., & Edelman, D. (2014, March). Special Issue on Credit Risk Modeling. *The Journal of the Operational Research Society*, 65(3), 321–322. doi:10.1057/jors.2014.6
- Flynn, E., & Crewson, P. (2012). Chapter 11 filing trends in history and today. Office of the United States Trustee.
- Grant-Thornton. (2010). *Zone of insolvency: Decision making in times of distress*. Author.
- Haverstock, C., & Crewson, P. (2012). *U.S. trustee program posts bankruptcy data*. Retrieved October 2012, from <http://www.ustrustee.gov>
- Heckman, I. (2015). A Critical Review on Supply Chain Risks – Definition, measures and modeling. *Omega*, 52, 119–132.
- Jones, M. T., Hilbers, P., & Slack, G. (2004). *Stress testing financial systems: What to do when the governor calls*. IMF working paper WP/04/127.
- Lopucki. (2012). *BRD success modeling project*. Retrieved from <http://lopucki.law.ucla.edu>
- Navarro-Galera, A., Lara-Rubino, J., Buendea-Carillo, D., & Rayo-Canton, S. (2015). What can increase the Default Risk in Local Government? *International Review of Administrative Sciences*.
- Reeves, M., & Püschel, L. (2015, July 2). *Die Another Day, bcg Perspectives*. Boston Consulting Group.
- Reeves, M. (2016, January). The Biology of Corporate Survival. *Harvard Business Review*.
- Rhode Island. (2013, January 28). *Report of the Municipal Stress Test Task Force*. Author.
- Schoemaker, P. (1995). Scenario planning: A tool for strategic thinking. *Sloan Management Review*, (Winter), 41.
- Shao, L., Kwok, C., & Zhan, R. (2013, September). National Culture and Corporate Investment. *Journal of International Business Studies*, 44(7), 745–763. doi:10.1057/jibs.2013.26
- Sheaffer, Z., & Brender-ilan, Y. (2014, February). Are Sensation Seekers in Control, a Study in Crisis Preparedness. *Risk Management*, 16(1), 1–24. doi:10.1057/rm.2014.1

KEY TERMS AND DEFINITIONS

Bankruptcy: A successful legal procedure and declaration for seeking protection from creditors. All bankrupt debtors are in fact insolvent. (bankrupt24.co.za, 2012), includes corporations and municipalities. (Jones, Hilbers, Slack, 2004).

Chapter 7: Of the U.S. Bankruptcy code pertains to corporate cases, or individual insolvencies, for which restructuring is unlikely to have success and liquidation the is only option.

Chapter 9: Of the U.S. Bankruptcy code pertains to municipalities, and its applicability varies by State. For example, it does not apply to the Commonwealth of Puerto Rico and is creating challenging legal issues in the 2016 financial crisis there.

Chapter 11: Of the U.S. Bankruptcy code pertains to corporate and personal cases for which reorganization or restructuring is a viable option.

Insolvency: Is defined as the financial state of a person or company or municipality unable to meet its obligations when due. Usually liabilities exceed current assets in those situations.

Scenario: Is a plausible future state of an enterprise or town may find itself in, and it is usually

quite different than what is expected under normal operating conditions.

Stress Testing: Corporations or municipalities for solvency, is a quantitative and qualitative process of showing the likelihood of an organizations' fiscal survival probability under adverse conditions. It addresses the "what-if" question and leads to better decisions, good governance and reduces likelihood of financial stress.

R

Category R

Robotics

Binary Decision Diagram Reliability for Multiple Robot Complex System

R**Hamed Fazlollahtabar***Sharif University of Technology, Iran & National Elites Foundation, Iran***Seyed Taghi Akhavan Niaki***Sharif University of Technology, Iran*

INTRODUCTION

Generally, complex systems in reliability are defined as a system having a combination of series, parallel, R out of N and standby components. Each of these models has the corresponding mathematical formulations for reliability computations leading to decompose the original system (or sub-system) into an equivalent one with a known cumulative distribution function (CDF) or reliability function. Continuing the decomposition procedure enables the decision maker to reduce the whole system to a unique component with a known CDF. For better understanding, an illustrative example for a complex system reduction is given in Figure 1. The system is composed of both series and parallel components which are reduced first to a series system and eventually to a one-component system.

It should be noted that the reduction methods explained before are not effective for all systems. In cases with complicated interrelations of components it is required to develop an efficient methodology. This methodology deals with subjects such as event trees, Boolean representations, coherent structures, cut sets and decompositions.

Network reliability analysis receives considerable attention for the design, validation, and maintenance of many real world systems, such as production, computer, communication, or power networks. The components of a network are subject to random failures, as more and more

enterprises become dependent upon network or networked computing applications. Failure of a single component may directly affect the functioning of a network. So the failure probability of each component is a crucial consideration while considering the reliability of a network. There are so many exact methods for computation of network reliability (Bobbio et al., 2006). The network model is a directed stochastic graph $G=(V, E)$, where V is the vertex set, and E is the set of directed edges. An incidence relation associates with each edge of G a pair of nodes of G , called its end vertices. The edges represent components that can fail with known probability. In real problems, these probabilities are usually computed from statistical data. The problem related with connection function is NP-hard. The same thing is observed for planar graphs (Provan, 1986).

BACKGROUND

The word *reliability* can be traced back to 1816, by poet Coleridge before World War II the name has been linked mostly to repeatability (Saleh and Marais, 2006). A test (in any type of science) was considered reliable if the same results would be obtained repeatedly. In the 1920s product improvement through the use of statistical process control was promoted by Dr. Walter A. Shewart at Bell Labs. Around this time Wallodi Weibull was working on statistical models for fatigue. The

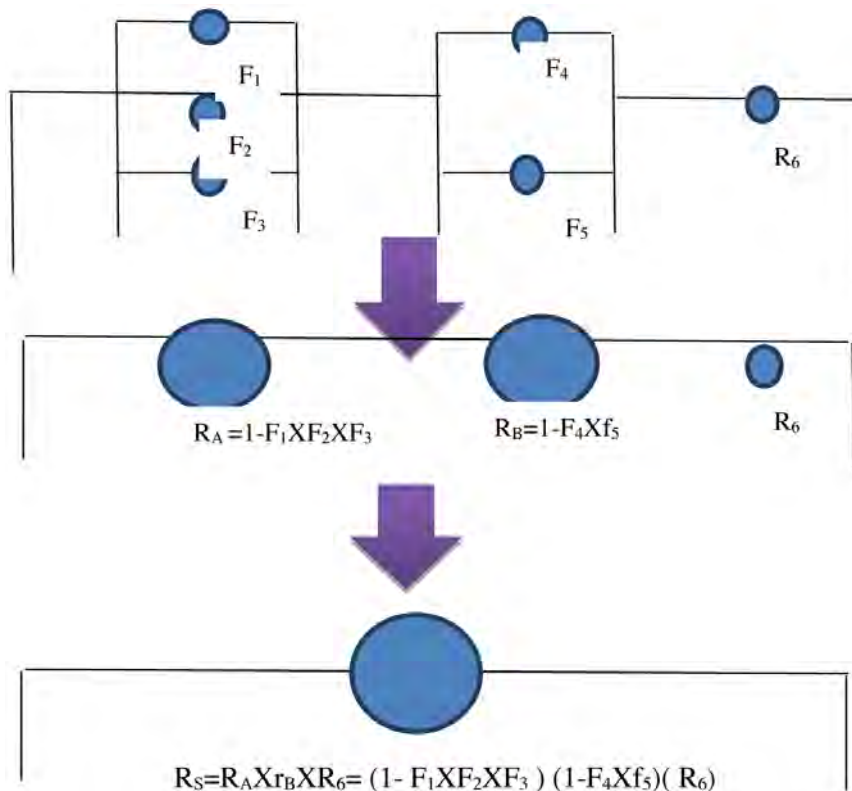
development of reliability engineering was here on a parallel path with quality (Juran and Gryna, 1988). The modern use of the word *reliability* was defined by the U.S. military in the 1940s and evolved to the present. It initially came to mean that a product would operate when expected (nowadays called “mission readiness”) and for a specified period of time. In the time around the WWII and later, many reliability issues were due to inherent unreliability of electronics and to fatigue issues. In 1945, M.A. Miner published the seminal paper titled “Cumulative Damage in Fatigue” in an ASME journal. A main application for reliability engineering in the military was for the vacuum tube as used in radar systems and other electronics, for which reliability has proved to be very problematic and costly. The IEEE formed the Reliability Society in 1948. In 1950, on the military side, a group called the Advisory Group on the Reliability of Electronic Equipment,

AGREE, was born. This group recommended the following 3 main ways of working:

- Improve Component Reliability;
- Establish quality and reliability requirements (also) for suppliers;
- Collect field data and find root causes of failures.

In the 1960s more emphasis was given to reliability testing on component and system level. The famous military standard 781 was created at that time. Around this period also the much-used (and also much-debated) military handbook 217 was published by Radio Corporation of America (RCA) and was used for the prediction of failure rates of components. The emphasis on component reliability and empirical research (e.g. Mil Std 217) alone slowly decreases. More pragmatic approaches, as used in the consumer industries,

Figure 1. A complex system reliability



are being used. The 1980s was a decade of great changes. Televisions had become all semiconductor. Automobiles rapidly increased their use of semiconductors with a variety of microcomputers under the hood and in the dash. Large air conditioning systems developed electronic controllers, as had microwave ovens and a variety of other appliances. Communications systems began to adopt electronics to replace older mechanical switching systems. Bellcore issued the first consumer prediction methodology for telecommunications, and SAE developed a similar document SAE870050 for automotive applications. The nature of predictions evolved during the decade, and it became apparent that die complexity was not the only factor that determined failure rates for Integrated Circuits (ICs). Kam Wong published a paper questioning the bathtub curve (Wong, 1981; see also O'Connor, 2002). During this decade, the failure rate of many components dropped by a factor of 10. Software became important to the reliability of systems. By the 1990s, the pace of IC development was picking up. Wider use of stand-alone microcomputers was common, and the PC market helped keep IC densities following Moore's Law and doubling about every 18 months. Reliability Engineering now was more changing towards understanding the physics of failure. Failure rates for components kept on dropping, but system-level issues became more prominent. Product development time continued to shorten through this decade and what had been done in three years was being done in 18 months. This meant that reliability tools and tasks must be more closely tied to the development process itself. In many ways, reliability became part of everyday life and consumer expectations.

Complex systems are characterized by large numbers of components, cut sets or link sets, or by statistical dependence between the component states. These measures of complexity render the computation of system reliability a challenging

task. Der Kiureghian and Song (2008) developed a decomposition approach which, together with a linear programming formulation, allows determination of bounds on the reliability of complex systems with manageable computational effort. The approach also facilitated multi-scale modeling and analysis of a system, whereby varying degrees of detail can be considered in the decomposed system.

In problems of maintenance optimization, it is convenient to assume that repairs are equivalent to replacements and those systems or objects are, therefore, brought back into an as good as new state after each repair. Standard results in renewal theory may then be applied for determining optimal maintenance policies. In practice, there are many situations in which this assumption cannot be made. The quintessential problem with imperfect maintenance is how to model it. In many cases it is very difficult to assess by how much a partial repair will improve the condition of a system or object and it is equally difficult to assess how such a repair influences the rate of deterioration. Kallen (2011) proposed a superposition of renewal process is used to model the effect of imperfect maintenance. It constitutes a different modelling approach than the more common use of a virtual age process.

Nishijima (2007) addressed the issue of optimization of reliability acceptance criteria for components of complex engineered systems with given criterion to acceptable system risk. To this end, first described how complex engineered systems may be modelled hierarchically by use of Bayesian probabilistic networks. The Bayesian probabilistic network serves as a function relating the reliability acceptance criteria of the individual components of the system to the risk acceptance criteria for the system. Thereafter, a constrained optimization problem is formulated for the optimization of the component reliabilities. In that optimization problem the system risk ac-

ceptance criterion defines the constraint and the expected utility from the system, is considered as the objective function.

During the design phase of a product, reliability engineers are called upon to evaluate the reliability of the system. The question of how to meet a reliability goal for the system arises

when the estimated reliability is inadequate. This then becomes a reliability allocation problem at the component level. Mettas (2000) estimated a general model for the minimum reliability requirement of multiple components within a system that will yield the goal reliability value for the system. The model consisted of two parts. The first part was a nonlinear programming formulation of the allocation problem. The second part was a cost function formulation to be used in the nonlinear programming algorithm. A general behavior of the cost as a function of a component's reliability is assumed for this matter.

In the exact method there are two classes for the computation of the network reliability. The first class deals with the enumeration of all the minimum paths or cuts. A path is a subset of components (edges and/or vertices), that guarantees the source and the sink to be connected if all the components of this subset are functioning. A path is a minimal if a subset of elements in the path does not exist that is also a path. A cut is a subset of components (edges and/or vertices), whose failure disconnect the source and sink. A cut is a minimal if the subset of elements in the cut does not exist that is also a cut (Hariri and Raghavendra, 1987; Ahmad, 1988). The probabilistic evaluation uses the inclusion-exclusion, or sum of disjoint products methods because this enumeration provides non-disjoint events. Numerous works about this kind of methods have been presented in literature (Lucet and Manouvrier, 1999). In the second class, the algorithms are based on graph topology. In the first process we reduce the size of the graph by removing some structures namely, polygon-to-chain (Choi and Jun, 1985) and delta-to-star reductions (Gadani, 1981). By this we will be able to compute the reliability in linear time and the reduction will result in a single edge.

The idea is to decompose the problem in to one failed and another functioning (Carlier and Lucet, 1996). The same was confirmed by Theologou & Carlier (1991) for dense networks. Satyanarayana & Chang (1983) and Wood (1985) have shown that the factoring algorithms with reductions are more efficient at solving this problem than the classical path or cut enumeration methods.

MAIN FOCUS OF THE ARTICLE

Assume a production system having multiple robots and work stations leading to a complex system. To evaluate the reliability of this system a network of components is considered configuring a complex reliability network. A robotic network is considered which has perfect vertices and imperfect links. It means path links may fail with known probability. We obtain the reliability of the given network by using an exact method and with binary decision diagram. Binary decision diagram based reliability evaluation involves three main steps. First, ordering the given path link. Second, generate the reliability function with the help of min-paths from source to sink. At last, apply Shannon's decomposition to compute the reliability of the given network. The reliability evaluation under study is in two states of working and breakdown. Then, multiple robots are considered in two states. To handle such a complex problem in a network structured production system, binary decision diagram (BDD) is proposed. The reasons are: adaptability with network system; Boolean decision systems according to two states of the robots; flexible computational capability with increase of size of the system components. In the proposed BDD working state is shown by 1 and breakdown (stop working) is shown by 0. More details are given in next section.

Binary Decision Diagrams

Akers (1978) first introduced BDD to represent Boolean functions i.e. a BDD is a data structure used to represent a Boolean Function. Bryant

(1992) popularized the use of BDD by introducing a set of algorithms for efficient construction and manipulation of BDD structure. The BDD structure provides compact representations of Boolean expressions. A BDD is a directed acyclic graph (DAG) based on the Shannon decomposition. The Shannon decomposition for a Boolean function is defined as follows:

$$f = l \cdot f_{l=1} + \bar{l} \cdot f_{l=0} \quad (1)$$

where l is one of the decision variables, and f is the Boolean function evaluated at $l = i$. By using Shannon's decomposition, any Boolean expression can be transformed into binary directed acyclic graph. BDD are used to work out the terminal reliability of the links. Coudert and Madre (1992) found BDD usefulness in reliability analysis which was further extended by Rauzy (1993 and 2003). They are specially used to assess fault trees in system analysis. In the network reliability framework, Imai et al. (1999) have shown how to functionally construct the corresponding BDD. An alternate approach was shown by Singhal et al. (2010, 2011, 2012), to compute BDD based network reliability. Table 1 shows the truth table of a Boolean function f and its corresponding Shannon tree is shown in Figure 2. Sink nodes are labelled either with 0, or with 1, representing the two corresponding constant expressions. Each internal node u has two out-edges called 0-edge, and 1-edge. The node linked by the 1-edge represents the Boolean expression when $l_i = 1$, i.e. $f_{li} =$

1; while the node linked by the 0-edge represents the Boolean expression when $l_i = 0$, i.e. $f_{li} = 0$.

Maintenance department records the failures and thus probability distribution is obtained leading to reliability evaluation. Since we have a network of components in the production system being active or inactive (binary state), binary decision tree is configured. The following notations are noted for this purpose:

L_j : robot j th

F_u : robot performance function

K : counter for robots; $j=1,2,\dots,K$.

SOLUTIONS AND RECOMMENDATIONS

Let's consider an example for the implementation of the proposed robotic system reliability evaluation. A production system having 9 work stations and 4 robots to process material handling in considered. The movement paths of the robots are in series and parallel with respect to the sequence of work stations they service. If two robots have a common path except source (S) and sink (T) nodes, then they are series and otherwise they are parallel. A schematic presentation of the system network is shown in Figure 3.

According to the series/parallel structure of the components the corresponding reliability block diagram is drawn in Figure 4.

Table 1. Truth table of a Boolean Function f

F_u	L_K	...	L_3	L_2	L_1
0	0	...	0	0	0
0	1	...	0	0	0
1	0	...	1	0	0
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
0	0	...	0	1	1
1	1	...	1	1	1

Figure 2. Binary decision Tree of the given Boolean Function

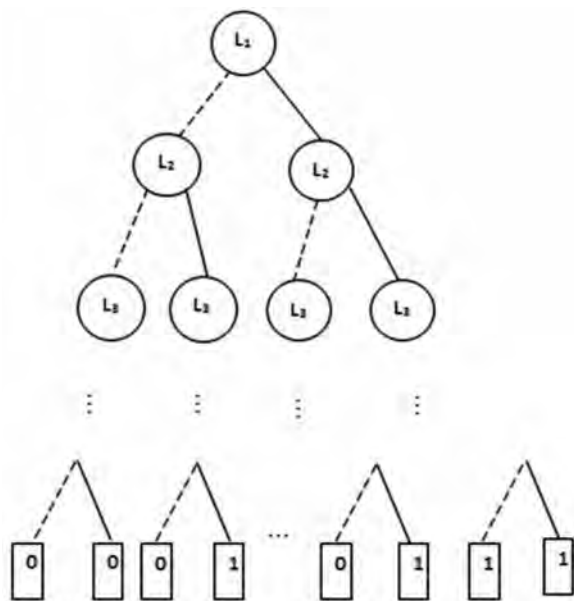
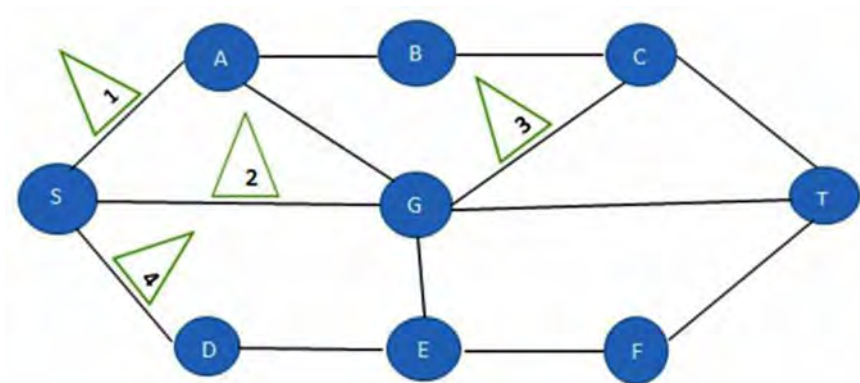


Figure 3. A schematic presentation of the robotic system



Box 1.

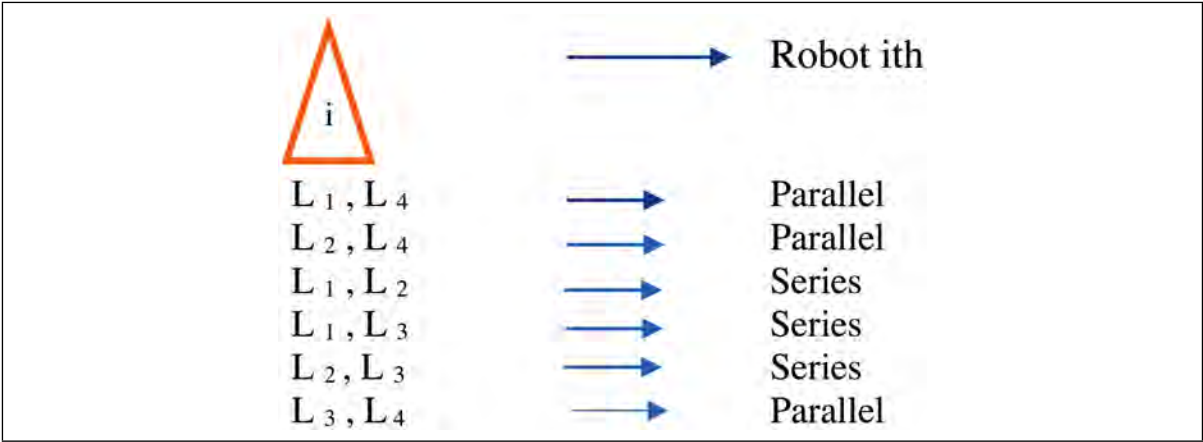
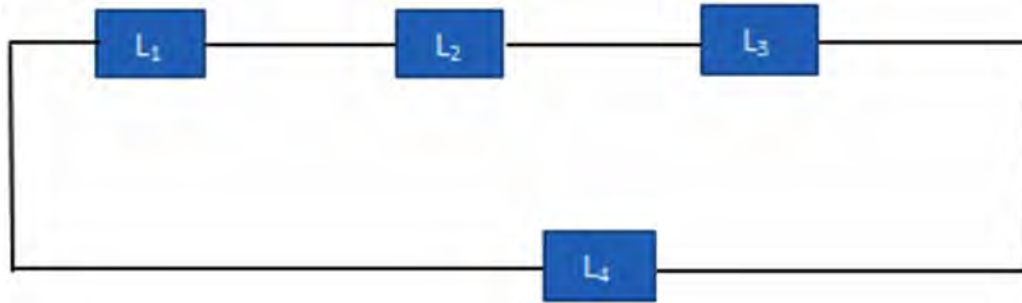


Figure 4. A reliability block diagram



Also, the corresponding table of truth for the 4-robot system is given in Table 2.

According to the truth table, a binary decision tree is formed and shown in Figure 5.

While there are two states of working (S) and failure (F) for each robot, 16 cases are formed as given below. Since robots are similar, the failure rate is considered 0.02 for all of them, i.e., $p(L1)=p(L2)=p(L3)=p(L4)=0.02$. Finally, the reliability of each case considering binary decision tree and Shannon decomposition formula are as follows:

$$\begin{aligned}
 L1_F L2_F L3_F L4_F: & 1-[(1-(0.02*0.02*0.02))*(1-0.02)]=0.020007 \\
 L1_S L2_F L3_F L4_F: & 1-[(1-(0.98*0.02*0.02))*(1-0.02)]=0.02038416 \\
 L1_F L2_S L3_F L4_F: & 1-[(1-(0.02*0.98*0.02))*(1-0.02)]=0.02038416 \\
 L1_F L2_F L3_S L4_F: & 1-[(1-(0.02*0.02*0.98))*(1-0.02)]=0.02038416 \\
 L1_F L2_F L3_F L4_S: & 1-[(1-(0.02*0.02*0.02))*(1-0.98)]=0.98000016 \\
 L1_S L2_S L3_F L4_F: & 1-[(1-(0.98*0.98*0.02))*(1-0.02)]=0.03882384 \\
 L1_S L2_F L3_S L4_F: & 1-[(1-(0.98*0.02*0.98))*(1-0.02)]=0.03882384 \\
 L1_S L2_F L3_F L4_S: & 1-[(1-(0.98*0.02*0.02))*(1-0.98)]=0.98000784 \\
 L1_F L2_S L3_S L4_F: & 1-[(1-(0.02*0.98*0.98))*(1-0.02)]=0.03882384 \\
 L1_F L2_S L3_F L4_S: & 1-[(1-(0.02*0.98*0.02))*(1-0.02)]=0.02038416
 \end{aligned}$$

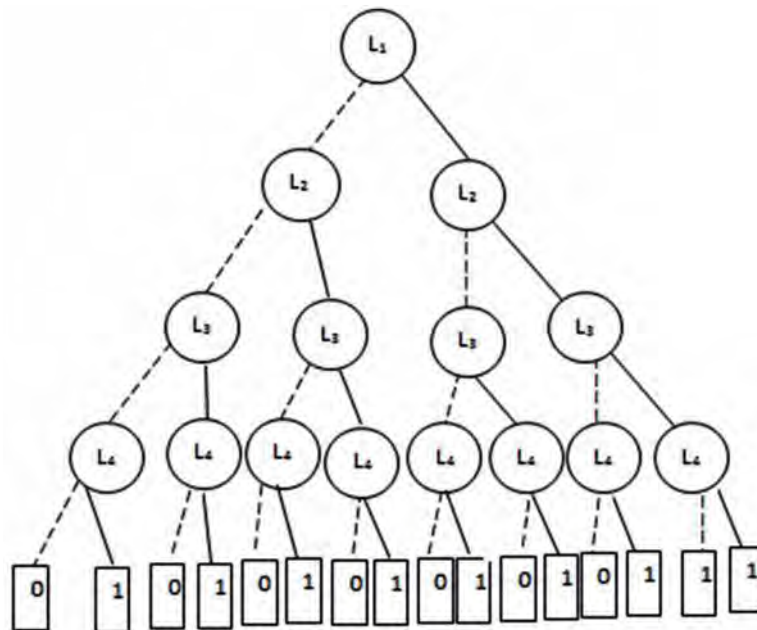
$$\begin{aligned}
 L1_F L2_S L3_S L4_S: & 1-[(1-(0.02*0.98*0.98))*(1-0.98)]=0.98038416 \\
 L1_F L2_F L3_S L4_S: & 1-[(1-(0.02*0.02*0.98))*(1-0.98)]=0.98000784 \\
 L1_S L2_S L3_S L4_F: & 1-[(1-(0.98*0.98*0.98))*(1-0.02)]=0.94236816 \\
 L1_S L2_F L3_S L4_S: & 1-[(1-(0.98*0.02*0.98))*(1-0.98)]=0.98038416 \\
 L1_S L2_S L3_F L4_S: & 1-[(1-(0.98*0.98*0.02))*(1-0.98)]=0.98038416 \\
 L1_S L2_S L3_S L4_S: & 1-[(1-(0.098*0.98*0.98))*(1-0.98)]=0.99882384.
 \end{aligned}$$

This way the reliability of a complex robotic system is computed. According to the obtained solutions in cases having more working components the reliability of the system is more. Thus the maintenance department should conduct a predictive maintenance program to keep components in working states to obtain more system reliability. Also, management can decide on reserve system in which some components are ready to be replaced with the broken down ones when needed. This way, the system is available most of the time. Of course, providing reserve components incur costs to the system but economic analysis associated with reliability evaluations help decision makers in this purpose. Generally, reserve system is justified due to the benefits achieved by a reliable system. Another recommendation could be the configuration of the work stations and robots in the production network so that more parallel sub-systems are formed. In parallel sub-systems the availability and thus the reliability of the

Table 2. Table of truth for 4-robot system

L_1	L_2	L_3	L_4	f_u
0	0	0	0	0
1	0	0	0	0
0	1	0	0	0
0	0	1	0	0
0	0	0	1	1
1	1	0	0	0
1	0	1	0	0
1	0	0	1	1
0	1	1	0	0
0	1	0	1	1
0	1	1	1	1
0	0	1	1	1
1	1	1	0	1
1	0	1	1	1
1	1	0	1	1
1	1	1	1	1

Figure 5. Binary decision tree for the four robot production system



system is more due to more choices for robots to serve in a sequence free manner. While in series system the whole system may be influenced by one component breakdown.

FUTURE RESEARCH DIRECTIONS

Reliability assessment in the current advanced manufacturing systems is significant due to large amount of economic investments. Also, complex system (network) of components motivates developing efficient methods for performance analysis. Binary state systems are the simplest for which is extensively considered in production systems due to maintenance consideration. Combining stochastic process computations such as branching processes with high performance artificial intelligence techniques help the decision makers to obtain more confident results. The research in this field is toward applying neural network methodologies for computing reliability in stochastic and uncertain components specifically in automation applications.

CONCLUSION

An efficient method for generating the BDD of a robotic complex system has been proposed. The reliability has been evaluated via BDD and by applying a truth table and Shannon decomposition formula. A network of robots in a production system was considered. The block diagram was configured and employed for developing different cases of the system performance. Considering the binary decision tree the computations of reliability for each case was presented.

ACKNOWLEDGMENT

This research has been supported by Iran National Elites Foundation.

REFERENCES

- Ahmad, S. H. (1988). Simple enumeration of minimal cut sets of acyclic directed graph. *IEEE Transactions on Reliability*, 27(5), 484–487. doi:10.1109/24.9868
- Akers, B. (1978). Binary decision diagrams. *IEEE Transactions on Computers*, 27(6), 509–516. doi:10.1109/TC.1978.1675141
- Bobbio, A., Ferraris, C. and Terruggia, R. (2006). *New Challenges in Network Reliability Analysis*. Technical Report, TR-INF-UNIPMN, 1-8.
- Bryant, R. E. (1992). Symbolic Boolean manipulation with ordered binary-decision diagrams. *ACM Computing Surveys*, 24(3), 293–318. doi:10.1145/136035.136043
- Carlier, J., & Lucet, C. (1996). A decomposition algorithm for network reliability evaluation. *Discrete Applied Mathematics*, 65(1-3), 141–156. doi:10.1016/0166-218X(95)00032-M
- Choi, M. S., & Jun, C. H. (1985). Some variant of polygon-to-chain reductions in evaluating reliability of undirected network. *Microelectronics and Reliability*, 35(1), 1–11. doi:10.1016/0026-2714(94)P1833-X
- Coudert, O., & Madre, J. C. (1992). Implicit and incremental computation of primes and essential primes of Boolean functions. *Proc. of the 29th ACM/IEEE Design Automation Conference*, 36-39. doi:10.1109/DAC.1992.227866
- Der Kiureghian, A., & Song, J. (2008). Multi-scale reliability analysis and updating of complex systems by use of linear programming. *Reliability Engineering & System Safety*, 93(2), 288–297. doi:10.1016/j.res.2006.10.022
- Gadani, J. P. (1981). System effectiveness evaluation using star and delta transformations. *IEEE Transactions on Reliability*, 30(1), 43–47. doi:10.1109/TR.1981.5220959

- Hariri, S., & Raghavendra, C. S. (1987). SYREL- A symbolic reliability algorithm based on path and cut set methods. *IEEE Transactions on Computers*, 36(10), 1224–1232. doi:10.1109/TC.1987.1676862
- Imai, H., Sekine, K., & Imai, K. (1999). Computational investigations of all terminal network reliability via BDDs. *IEICE Transactions on Fundamentals*, 82(5), 714–721.
- Juran, J., & Gryna, F. (1988). *Quality Control Handbook* (4th ed.). New York: McGraw-Hill.
- Kallen, M. J. (2011). Modelling imperfect maintenance and the reliability of complex systems using superposed renewal processes. *Reliability Engineering & System Safety*, 96(6), 636–641. doi:10.1016/j.ress.2010.12.005
- Lucet, C., & Manouvrier, J. F. (1999). Exact methods to compute network reliability. In D. C. Ionescu & N. Limnios (Eds.), *Statistical and Probabilistic Models in Reliability* (pp. 279–294). Birkhauser Boston. doi:10.1007/978-1-4612-1782-4_20
- Mettas, A. (2000). Reliability allocation and optimization for complex systems. *Proceedings of the Annual Reliability and Maintainability Symposium*. doi:10.1109/RAMS.2000.816310
- Nishijima, K. (2007). *Optimal Reliability of Components of Complex Systems using Hierarchical System Models*. Institute of Structural Engineering, ETH Zurich.
- O'Connor, P. D. T. (2002). *Practical Reliability Engineering* (4th ed.). New York: John Wiley & Sons.
- Provan, J. S. (1986). The complexity of reliability computations on planar and acyclic graphs. *SIAM Journal on Computing*, 15(3), 694–702. doi:10.1137/0215050
- Rauzy, A. (1993). New algorithms for fault tolerant trees analysis. *Reliability Engineering & System Safety*, 5(59), 203–211. doi:10.1016/0951-8320(93)90060-C
- Rauzy, A. (2003). A new methodology to handle Boolean models with loops. *IEEE Transactions on Reliability*, 52(1), 96–105. doi:10.1109/TR.2003.809272
- Saleh, J. H., & Marais, K. (2006). Highlights from the Early (and pre-) History of Reliability Engineering. *Reliability Engineering & System Safety*, 91(2), 249–256. doi:10.1016/j.ress.2005.01.003
- Satyanarayana, A., & Chang, M. K. (1983). Network reliability and the factoring theorem. *Networks*, 13(1), 107–120. doi:10.1002/net.3230130107
- Singhal, M., Chauhan, R. K., & Sharma, G. (2010). An Alternate Approach to Compute the Reliability of a Computer Communication Network Using Binary Decision Diagrams. *Communications in Computer and Information Science*, 94(4), 160–170. doi:10.1007/978-3-642-14834-7_16
- Singhal, M., Chauhan, R. K., & Sharma, G. (2011). A New approach for finding the Various Optimal Variable Ordering to generate the Binary Decision Diagrams (BDD) of a Computer Communication Network. *International Journal of Computers and Applications*, 31(3), 1–8. doi:10.5120/3801-2502
- Singhal, M., Chauhan, R. K., & Sharma, G. (2012). Binary Decision Diagrams and Its Variable Ordering for Disjoint Network. *International Journal of Advanced Networking and Applications*, 3(6), 1430–1437.
- Theologou, O., & Carlier, J. (1991). Factoring and reductions for networks with imperfect vertices. *IEEE Transactions on Reliability*, 40(2), 210–217. doi:10.1109/24.87131
- Wong, K. (1981). Unified Field (Failure) Theory- Demise of the Bathtub Curve. *Proceedings of Annual RAMS*, 402–408.
- Wood, R. K. (1985). A factoring algorithm using polygon-to-chain reductions for computing K-terminal network reliability. *Networks*, 15(2), 173–190. doi:10.1002/net.3230150204

ADDITIONAL READING

Ebeling, C. E. (1997). *An Introduction to Reliability and Maintainability Engineering*. Boston: McGraw-Hill.

Gnedenko, B., & Ushakov, I. (1995). *Probabilistic Reliability Engineering*. Wiley. doi:10.1002/9780470172421

Govil, A. K. (1983). *Reliability Engineering*. New Delhi: McGraw Hill.

Kuo, W., & Zuo, M. (2003). *Optimal Reliability Modeling: Principles and Applications*. Hoboken, NJ: John Wiley & Sons.

Levitin, G., & Lisnianski, A. (2013). *Multi-state system reliability: assessment, optimization and applications*. Singapore: World Scientific Publishing.

Lisnianski, A., & Levitin, G. (2003). *Multi-state System Reliability Assessment, Optimization, Applications*. Singapore: World Scientific. doi:10.1142/5221

Nishijima, K. (2007). *Optimal Reliability of Components of Complex Systems using Hierarchical System Models*. Institute of Structural Engineering, ETH Zurich.

Stirzaker, D. (2005). *Stochastic Processes and Models*. New York: Oxford University Press.

KEY TERMS AND DEFINITIONS

Binary System: A system in which information can be expressed by combinations of the digits 0 and 1 or a system consisting of two parts.

Complex System: A system having multiple components and sub-components making it difficult to evaluate its performance.

Failure: Failure is the state or condition of not meeting a desirable or intended objective, and may be viewed as the opposite of success.

Reliability: Reliability is the ability of a system or component to perform its required functions under stated conditions for a specified period of time.

Repair: To restore to sound condition after damage or injury.

Robot: An automated driverless controllable vehicle which is used as transportation and transferring device.

Work Station: Work station is an industrial unit for processing a function allocated due to the manufacturing plan using input material and delivering semi-produced or final products to the next manufacturing unit.

A Bio-Inspired, Distributed Control Approach to the Design of Autonomous Cooperative Behaviors in Multiple Mobile Robot Systems

Gen'ichi Yasuda

Nagasaki Institute of Applied Science, Japan

INTRODUCTION

This chapter concerns the design and implementation of bio-inspired control architectures for intelligent multiple mobile robot systems. In a multi-robot system, robots work together efficiently to accomplish tasks intrinsically distributed in space, time, or functionally. Due to the cost of increasing system complexity especially in control and communication, it is difficult to implement a centralized controller for multi-robot systems under unknown dynamic environments (Khoshnevis & Bekey, 1998; Klavins, 2003). Focusing on building control systems for intelligent mobile robots, this chapter presents a non-centralized, behavior-based methodology for autonomous cooperative control, inspired by the adaptive and self-organizing capabilities of biological systems, such as turning over behavior of starfishes (Suzuki, et al., 1971), which can generate robust and complex behaviors through limited local interactions in the presence of large amount of uncertainties (Pfeifer et al., 2007; Salazar-Ciudad et al., 2000; Shen et al., 2004).

Traditionally, the fundamental feature of asynchronous concurrent processes in robot control systems was already employed in the control system design as a centralized control, where a single multitask operating system controls different programs, or processes, to be performed concurrently, such as locomotion control, environment recognition, path planning, decision making in uncertain environments, etc. under a set of tim-

ing requirements. From the view of asynchronous concurrent processes, modularity and localization of control are critical issues in control system design. Thus, in this chapter, the distributed control architecture, based on discrete event systems as a formal approach, is configured as a collection of asynchronous, concurrent processes in a hierarchical but non-centralized way, and consequently it can be developed in such a way to reduce the programming requirements. While actuator and sensor controls, such as vehicle control and image processing, are prerequisite at the lowest control level, a unified and systematic methodology has been developed to combine given techniques for controlling the robot motion and for integrating sensory information at some higher control levels, into one complete robot control system for real-world robotic applications.

Asynchronous concurrent processing on parallel distributed architecture can support the coordination between the global intelligence or control center and the set of local centers as well as among local centers (Alon, 2006; Kondacs, 2003; Taylor, 2014), where each local center supports at least one robotic module and each module is governed by only one local center. The channel between two local centers is also open for emergent responsive motions. The object-oriented concept can be employed to support the system requirements for modularization of robot intelligence including the reconfiguration of global-local interaction through parallel distributed processing, utilizing the traditional technique in spatial and kinematic

DOI: 10.4018/978-1-5225-2255-3.ch592

planning, and dynamic control as building blocks (Werfel, 2004). Thus, the distributed autonomous control system design based on asynchronous concurrent processing would provide the necessary flexibility to overcome some of limitations of current intelligent mechatronic system design. Furthermore, the flexibility for different tasks and control environments can be increased, by integrating various robots and components into systems in a heterogeneous robot environment where each robot has its own kinematic structure and programming language.

This chapter describes the hierarchical task specification in terms of event driven state based Petri net modeling (Yasuda, 2014), where the robotic task model is hierarchically decomposed into subtasks in such a manner that every subtask, represented as a subnet of the task specification net model, is assigned to a local module, based on the geographical distribution of the local centers.

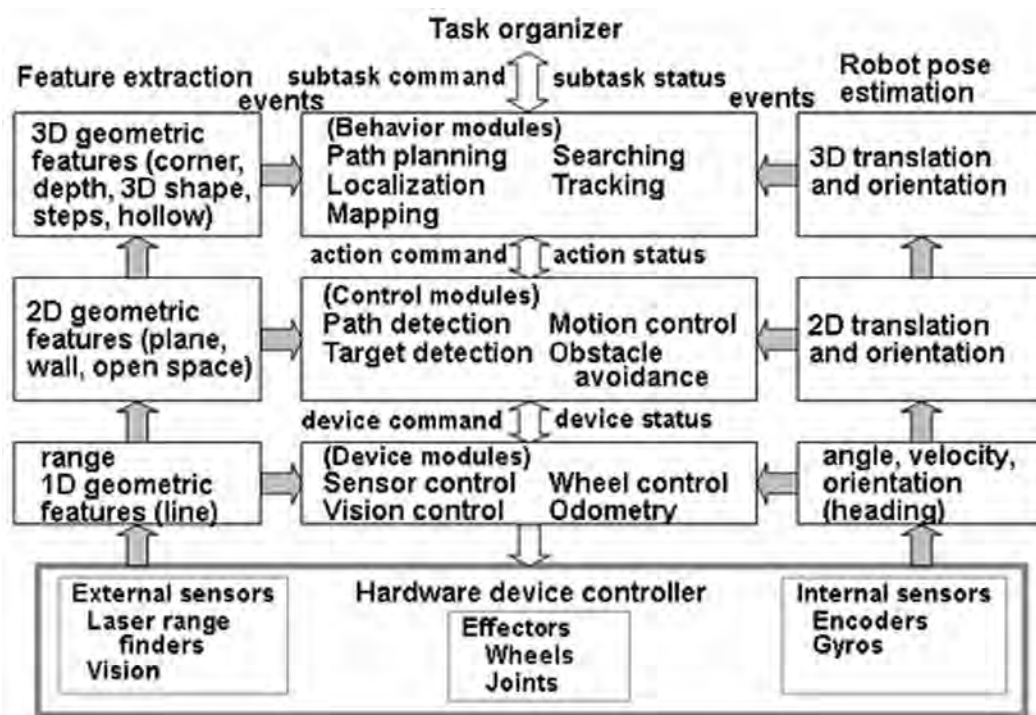
BACKGROUND

The control of autonomous robots evolves from the control of an individual robot to the self-coordination of a group of robots that interact and function together to form a social group, where the interactions among robots can produce behaviors that no single robot can accomplish without the other individual robots. In recent studies focusing on cooperation of distributed autonomous multiple robot systems, the cooperation strategy is mostly based on distributed problem solving and distributed decision (Shi et al., 2012), like a human individual, considering each of these robots as an agent in the social interaction. The behavior of a robot can extend beyond itself to include other individual robots. However, a robotic task is considered as a succession of many primitive reactive behaviors or actions, each activated with some events or external sensing information. For example, in collision avoidance among mobile robots and transporting many discrete litters by multiple robots, each robot gets some information

about other robots by communication, and then uses the information for making a distributed decision. Collective execution of cooperative tasks with dynamic constraints, such as large object manipulation, may have more dynamic factors. Because the robots which are working on manipulation, interfere each other dynamically through the manipulated object, the increase in communication required for achieving harmonious cooperation between moving motion and manipulating motion in a robot makes the control system difficult to realize in a pure distributed system. So, some global information about the object and robots should be collectively reflected from sensors on each robot, so that all the robots can react synchronously to environmental change or upper level mastering event in real time.

The behavior-based approach emphasizes the decomposition from a task to a set of pure reactive behaviors for real-time control, which implies fully parallel control architecture composed of multiple behavior modules. The important problem is that, in cases where control should be shared between multiple modules, the output would result from the arbitration of multiple conflicting module outputs. The activation and inhibition mechanism of conflicting modules composes the intelligence center of the distributed control system. The behavior network is another example where a behavior is chosen by comparing and updating activation levels for each behavior. Each enabled behavior selects the required actions, computes run-time parameters, and generates a bid describing the appropriateness of the behavior. The most appropriate behavior is determined in a distributed manner through inter-behavior bidding without any centralized mechanism. Although the problem of arbitration between behaviors is better solved in a centralized system with some knowledge, the arbitration occurs only for specific cases in a system where tasks are partially ordered with a planner. Because, in dynamic environment, a robot needs to quickly decide what to do and how to do it, the deliberative components must keep pace with changes in the environment to reactively

Figure 1. Hierarchical and distributed control architecture for mobile robot activities in exploring and mapping task



produce intelligent behavior where each behavior is also responsible for planning the required actions, so that fixed computational resources need to be distributed among the behaviors. Thus, the reactions, composed of reflex, local control and replanning, should run in parallel on hierarchical software architecture, in order to insure reactivity at every level. Behavior and other control modules should have their own control autonomously with knowledge grouped at each level in order to have a better management of conflicts.

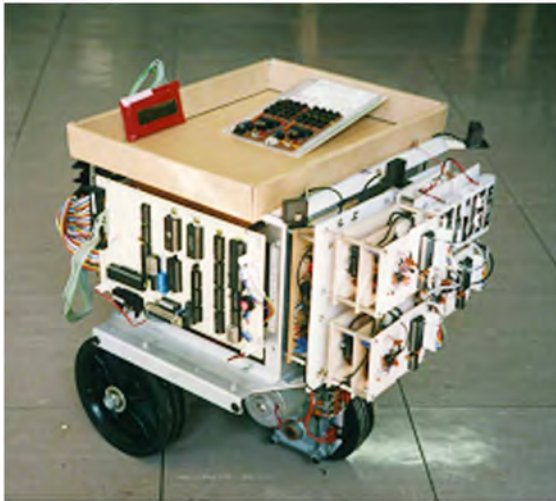
HIERARCHICAL DISTRIBUTED ARCHITECTURE FOR BEHAVIOR BASED MOBILE ROBOT CONTROL

The conceptual framework of hierarchical and distributed control software architecture for integrated deliberative planning and reactive execution of mobile robots is shown in Figure 1. The software architecture is constructed upon

the distributed hardware structure, where each microcomputer-based intelligent device for sensing and actuation is connected via a serial bus. The software control structure is not only functionally parallel/distributed but also hierarchically distinguished for effective development of a robot control system, considering the different abstraction levels that correspond to the different real-time constraints and the complex nature of its signals. The distinction of planning and reactive control is based on the different time scales of interaction. A view of an experimental mobile robot is shown in Figure 2.

The major subsystems of the hierarchical architecture are the coordination and control layers, which are implemented on the system controller. Each control module at the control layer provides for a dynamic link between perception and action to control robot motion using external sensory data through input lines carrying data from sensor modules, where output lines issue commands to effector modules or send data to another module.

Figure 2. View of an experimental mobile robot



Sensing information is organized into several levels corresponding to increasingly complex representations and abstract interpretations. Any representation of sensing information at a level may be used for event detection to achieve the condition test by managing all the conditions at the level, and to send a signal to the upper layer whenever the condition is met. For example, the range-finding module manages a belt of range sensors, to process their echoes, and to verify conditions or events, where three levels of interpretation are considered: point distance threshold level, segment location level, poly-line recognition level. Similar to the levels of interpretation of sensing information, there are different levels of actuation command. The effector module to drive the robot's motorized wheels with internal sensors is organized into trajectory control level and tracking control level. The trajectory control level receives a path (straight or circular), and generates a smoothed sequence of straight lines and clothoid arcs. The tracking control level tracks a point, supplied by the trajectory control level, by servoing of the robot motion using perceptual data. For example, the robot can detect and avoid obstacles while carrying out wall following or object pattern tracking, through cooperative parallel execution of the obstacle avoidance and the motion

execution modules. For primitive goal-directed action control, besides motion execution modules, the control layer embodies landmark and obstacle detection, self-localization, and self-referencing modules. Basically, for the same actuator system, such as arm, leg and hand, one motion execution module can run at a time and their parallel or sequential execution can be decided with efficient multi-agent system computation based on mutual exclusion to a limited computer resource.

The coordination layer is in charge of the overall behavior control as a whole, coordinating the control modules by sending the activity start events and waiting for the incoming events and clock events which have a global influence on its achievement. The coordination layer is built as a set of concurrent layers that embodies the different behaviors. Behavior modules have specific inputs and outputs: sensory and actuation information. They may share some of the input information with other modules. The access to the hierarchical sensory information is managed by the shared information manager. While a behavior module is activated by its specific inputs, it is executed by the attached executor as a software agent. The set of active behaviors and their resulting cooperation is controlled by the incoming events.

The task organization layer generates a task specification with a global planer based on specified task plan. Generally each step of a plan corresponds to a task, while a task is represented as a succession of commands written with some command language. Its event based task description coordinates the system behaviors or subtasks: environment map building, global navigation, and particular manipulating operations. At the task organization layer, global planning is executed as one small part of the main controller's nominal cycle of activities. A geometrical map of the local environment is built in the coordination layer. Given a set of target locations, the coordination layer plans a detailed medium distance local path to the closest target position only and executes this plan. Upon reaching this target location, its local map will change based on the perceptual

information extracted during motion. Only then the local path planner is triggered to generate a path from the new location to the next target location. When a reference object is detected by a vision sensor, a trajectory planner is triggered to generate the short distance motion for docking or manipulation.

MODELING AND DESIGN OF ROBOTIC TASKS

For the Petri net modeling of robotic activities, a place represents a condition or state of processes or resources. A transition corresponds to an event or action. The input places of the transition define the pre-conditions to the execution of the action. The output places define the results or post-conditions of the action (Milutinovic, et al., 2002). For real-time cooperative control of some robotic activities, a transition representing a robotic activity is decomposed into start and end transitions and a place representing the process, or the code associated to the activity. Temporal constraints, such as a delay, duration, time-out, or periodicity can be included in the start transition. If a place is an action place, a token placed in the place indicates that the state of the process corresponding to the place is holding. The execution starts when the place becomes marked. When the execution ends, the end transition is fired and the processing time is determined by external conditions. In our research, for compact representation of interaction with the environment, the extended Petri net adopts the following elements as input and output interfaces which connect the net to its environment: gate arcs and output arcs. A gate arc connects a transition with a status signal source (or places) in the internal or external environment, and depending on the signal, it either permits or inhibits the occurrence of the event. An output arc connects a place with a resource or external machine and sends a command signal to the resource.

An extended Petri net representation of robotic actions is specified using transition firing with

gate arcs based on the command/response concept. When the start transition is enabled, the transition triggering function is activated to check the input permissive gate signals and request the resources needed by the token. After the transition is fired, the initialization command is sent to the resource through the output signal arc. Next, after the initialization is finished, the acknowledge signal is sent from the resource through the inhibitive gate. Then, the token sends the associated parameters and action start signal to the resource. At the end of actions, the resource sends the signal status through the inhibitive gate. After the termination process, the resource sends the status to acknowledge the completion of the process through the inhibitive gate arc, and finally the resource is released, where the termination activity may be interlocked with another subtask in the system through the inhibitive gate arc.

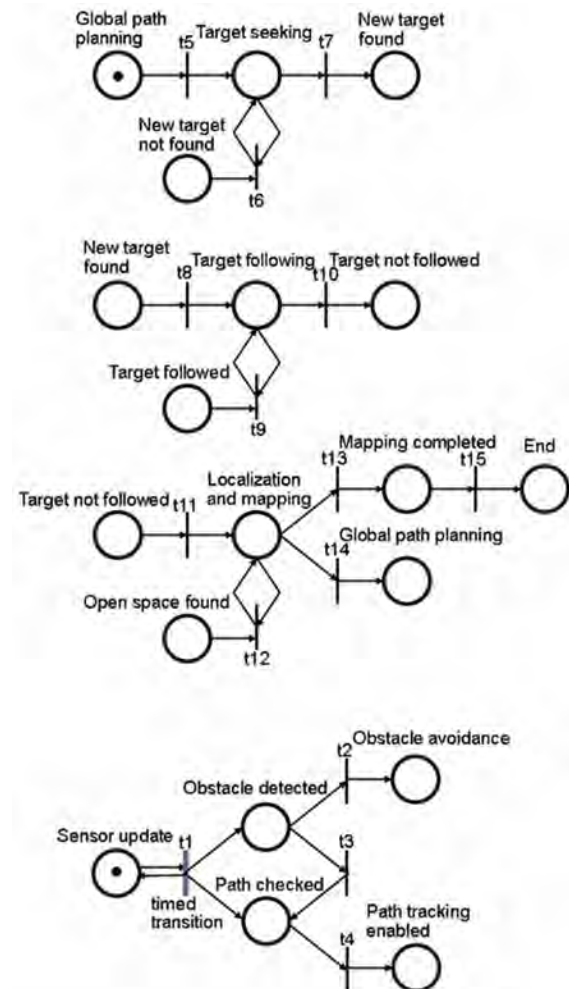
Basically, from a task-oriented view of robotic activities, a robotic task is defined as a collection of state changes of the objects or the robot's environment in a system, and a task can be performed in a sequence of these subtasks and each subtask, in turn, will take place according to the behaviors dependent on the successive sensor data. Primitive behaviors can be enabled if their pre- and post-conditions are fulfilled, and their executions are triggered by their specific signals. When two or more behaviors are enabled, one of them should be selected because they use the same resource or motion execution module. So, the net model should be enabled in parallel with the resource place. By the analogy with instinct behaviors in animal ethology (Tinbergen, 1951), complex behaviors are viewed as a resultant behavior of different kinds of robotic actions combined using control structures such as sequential, conditional, iterative, and concurrent constructs. Given all the possible actions, the resultant, desired behavior is brought about by a set of explicit rules of transitions between actions through sensor information, leading to the unexpected or surprising appearance. Top-down decomposition and bottom-up construction, referred as refinement and abstraction by place

or transition substitution, support the definition of activities and their associated resources, offering a vertical structuring mechanism similar to object composition in object-oriented language. The hierarchical structuring mechanism can be defined in static or dynamic form, corresponding to the concept of macro and procedure invocation, respectively. For a macro-place, where all the arcs are connected to transitions, the boundary of the subnet is composed by places, and a set of interface ports are used to connect the net and its subnets for flat (non-hierarchical) or hierarchical simulation. Bottom-up, horizontal net composition is defined as the merging of nets into a single net one through transition or place fusion to support reusability of net modules already available. High-level net folding based on tokens and node vectors is seen as a kind of fusion.

Petri nets have been applied to a simple mobile robot task of sensory environment exploring and mapping, composed of obstacle avoidance, target seeking and wall following behaviors (Nishiwaki, et al., 2014). Upon these primitive behavior modules, a Petri net based approach was applied to coordination of several concurrent activities of modules for the higher-level behavior or task. As shown in Figure 3, the overall net model is conceptually configured through sequential constraints between behavior modules defined based on the task specification sent from the task organization layer so that the behaviors of the whole system are brought about for achieving the global task specification. The firing of transition t_1 , representing the sensing action, models the updating of data from the range sensors. The places that receive tokens from the sensing module model data information checked and supplied about obstacle and wall distances and directions and target positions. If an obstacle is detected as indicated by the firing of transition t_3 , the obstacle avoidance behavior takes place. Otherwise, if no obstacle is found as indicated by the firing of transition t_4 , after the firing of transition t_5 , the firing of transitions defining which behavior is initiated depending on current sensory data in exploration and mapping

is brought about. The activity of each behavior module is represented by a place in the net. In Figure 3, the robotic task is divided into three behaviors, “Target seeking”, “Target following”, and “Localization and mapping”, where each behavior module is triggered by particular sensing information, plans the corresponding actions, calculates the strength, or bid, of the behavior command using the importance of the sensing information and the specified priorities among the different behaviors. The motor control mode models the motion states of the mobile robot according to the behavior decision, and is stored at the lowest level in the hierarchical shared memory.

Figure 3. Overall Petri net model composed of behaviors involved in exploring and mapping task



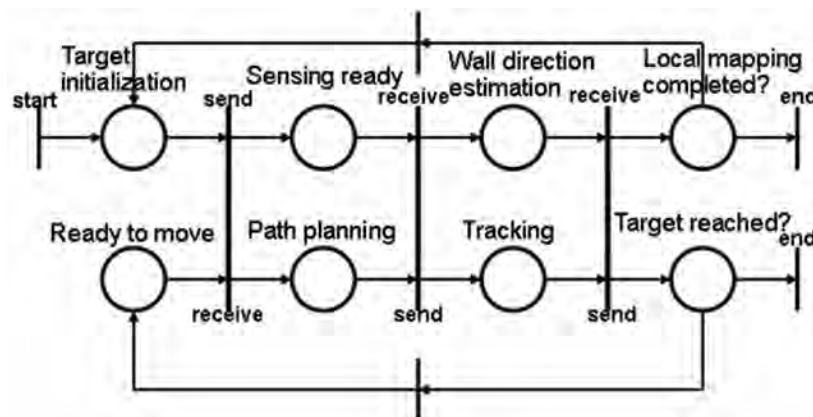
Thus, the overall behavior will take place through the activation of the behavior modules dependent on the successive sensor data, while the obstacle avoidance behavior is always concurrently checking whether an obstacle is present or not.

Actually, behavior “Target seeking” starts when the robot starts environment search, that is, the corresponding transition t_5 fires. Since a wall is detected, as indicated by the firing of transition t_6 , behavior “Target following” begins, that is, the robot keeps following the wall using range sensors and builds a local map. The end of the behavior and the start of behavior “Localization and mapping” take place once the wall is no longer present as indicated by the firing of transition t_{10} , that is, the robot performs the localization and mapping task. Thus, the robot behavior may oscillate between the exploring and mapping behaviors. At the end of “Localization and mapping”, as indicated by the firing of transition, if the whole environment in the global map is searched, then the mobile robot stops as indicated by transition t_{13} and place “End”, if not the next target is determined and the search and map behavior is continued as shown by transition t_{14} .

Figure 4 shows the detailed net model of the wall following behavior in exploring and mapping. When the behavior is active, it carries out the two primitive behaviors, target tracking and local mapping, continuously using scanned distance measurements information. The primitive

behaviors must start and end simultaneously, synchronizing with events for sending and receiving the required distance information. The target tracking module requests wheel control and distance measurement control modules to perform tracking control using scanned measurements. The locations of sensed points are corrected taking into account the motion of the robot to estimate the direction of the wall with respect to the robot coordinate. The measurement module integrates the most recently sensed points to have complete updated data about the wall direction. The local mapping module updates the local map with the current wall direction. If the wall is not found, the two behaviors terminate the tracking control and local map building. Then the localization behavior is activated to localize the robot on its own local map and transfer the positional information to the global map builder. The current goal or subgoal and the current state of the robot are accessed and read as global, shared variables using synchronous sending and receiving actions between associated modules based on mutual exclusion. The synchronous actions are represented as a shared transition for synchronous communication. The target tracking behavior is a periodic process, while the local mapping behavior is executed each time that a new location of wall comes from the tracking module. The wall following behavior is executed not periodically, to react to environment status change as a supervising process, such as

Figure 4. Petri net model of wall following behavior with parallel constructs



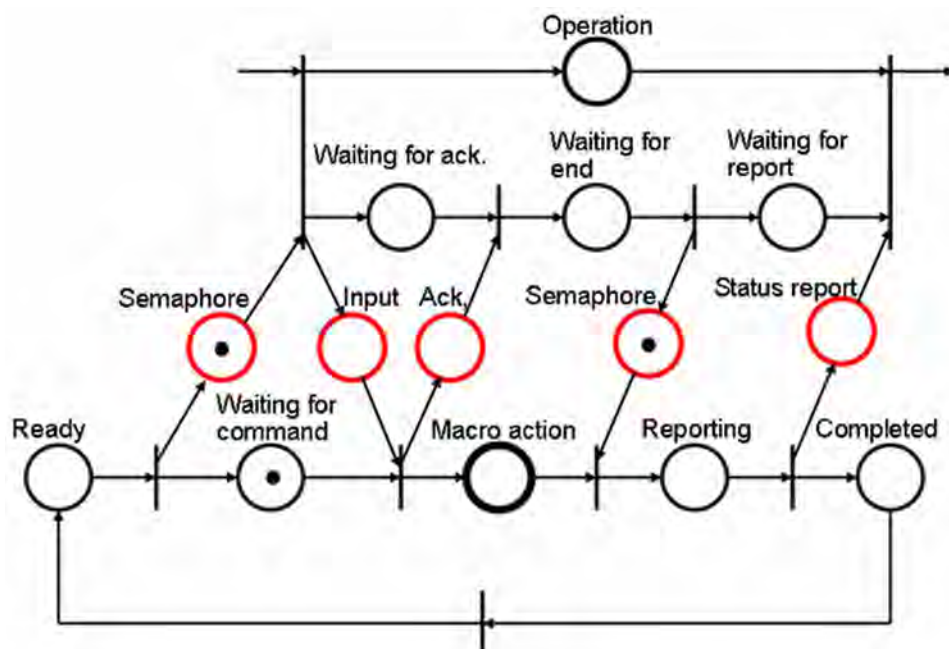
obstacle detection or command reception from the upper level. When the behavior is started, a triggering function is activated to detect the termination condition. A triggering function is also activated to detect obstacles, such that if an obstacle is encountered, the function is fired and the robot stops. For path tracking using ultrasonic range sensor signals, if the straight-line path to the target is free, the target position is redirected to the sensor module and obstacle avoidance is stopped.

SOFTWARE CONTROL OF MOBILE ROBOTS

All the modules in the coordination layer are realized as asynchronous concurrent processes activated under a multitasking operating system. The local planning module receives a task plan, parses or analyses each task into a command language composed of a set of command statements. A command represents a behavior or function provided by specified control modules in the

control layer. Then the behaviors in the coordination layer are activated as a whole, taking into account events that might have a global influence on its achievement, coordinating the activities of control modules such that they are in accordance with transition firing in the hierarchical net model. Thus, the coordination layer carries out tasks and manages robot resources, by detecting and arbitrating conflicts or multiple access to a sensor/effector module. The robot localization may use different resources to provide robot position, such as odometry, laser or vision data, and some external or contextual knowledge. When an event occurs at any control module, the module interrupts the system controller. The behavior module receives the event and updates corresponding net model data, or enabled transitions, in the hierarchical shared memory, where the states of activity of all modules in the system are managed with the associated input and output event identifiers and the parameter list, so that the information provided by sensor modules may be used by dedicated modules at several levels of interpretation. Using the

Figure 5. Petri net model of communication protocol between operation and control modules in hierarchical Petri net



shared memory and an environment map created by a global map manager, the behavior module processes the event and generates global actions for navigation. Then, the behavior module requests commands to the relevant control modules and these modules perform real-time control actions according to the commands, as shown in Figure 5.

A token points to a command statement and contains information related to its current execution status. Triggering functions for transition firing play a central role in the activation and inhibition of the behavior modules. To provide for event based programmable reactivity composed of three stages, that is, reflexive reaction, action determined by the current task and task re-planning, a logical formula of conditions to be tested for transition firing is associated to each behavior or control module. If the conditions are true, a set of reactive actions are carried out and commands specified by the token, including control loop, are sent to and executed by the specified modules or resources. At the same time, the transition triggering functions of conditions for ending of the actions, which are also specified by the token, are activated. Robot resources such as sensor and effector modules are managed to detect and arbitrate conflicts by multiple access. If an error occurs during task execution, the origin of the error is located and a procedure or a sequence of intermediate actions to reach the sub-goal, is inserted in the current task for possible repairs locally handled at the control level or re-planning using global knowledge of the whole plan at the task level.

FUTURE RESEARCH DIRECTIONS

The main disadvantage of using low-level Petri nets has been overcome with the use of hierarchical, modular Petri nets. Practically the decomposition strategies for obtaining subnets may be adopted according to the number of processors, the similarity structure of the subnets or the geographical distribution of the subsystems. To make further

improvement on the real-time performance, parallel and distributed execution of the Petri nets in robot controllers should be augmented, such that the net model is maximally partitioned into subnets in such a manner that every subnet may be assigned to an agent capable to control one active component or effector, involving one subsystem or workplace and avoiding situations where two or more agents handle the same subsystem. Innovative multi-core parallel processing architecture with FPGAs for efficient parallel execution is now being employed for the development of hierarchical and distributed Petri net models in the robot controller, in order to overcome limitations on the status of implementation due to hardware and low-level software, such as inter-process communication.

CONCLUSION

This chapter presents a design methodology of behavior modules for discrete event distributed control of autonomous mobile robot systems. A modular, Petri net based behavioral control software has been implemented in accordance with a hierarchical distributed microcontroller based hardware structure. The behavior modules with respective pre-conditions and post-conditions can be dynamically connected in response to status event information from action control modules at the lower level to achieve the specified overall task. The approach involving planning, control and reactivity can integrate high-level command input with the behavior modules through the distributed autonomous control architecture, to correspond better to the needs of general purpose mobile robots.

REFERENCES

Alon, U. (2006). *An Introduction to Systems Biology: Design Principles of Biological Circuits*. Chapman & Hall/ CRC.

- Arbib, M. A. (2004). Evolving emotions in animal and robot. *International Journal of Computational Intelligence and Applications*, 4(03), 225–236. doi:10.1142/S1469026804001343
- Gadanho, S. C. (2003). Learning behavior-selection by emotions and cognition in a multi-goal robot task. *Journal of Machine Learning Research*, 4, 385–412.
- Khoshnevis, B., & Bekey, G. A. (1998). Centralized sensing and control of multiple mobile robots. *Computers & Industrial Engineering*, 35(3-4), 503–506. doi:10.1016/S0360-8352(98)00144-2
- Klavins, E. (2003). Communication complexity of multi-robot systems. *Advanced Robotics, Springer*, 7, 275–292. doi:10.1007/978-3-540-45058-0_17
- Kondacs, A. (2003). Biologically inspired self-assembly of two-dimensional shapes using global to local compilation. *Proceedings of International Joint Conference on Artificial Intelligence*, 633–638.
- Mamey, M., Vasirani, M., & Zambonelli, F. (2004). Experiments in morphogenesis in swarms of simple robots. *Applied Artificial Intelligence*, 18, 9–10, 903–919.
- Meng, Y., & Gan, J. (2007). Local interaction via virtual stigmergy coordination in distributed search and collective cleanup. *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, 1371–1376. doi:10.1109/IROS.2007.4399254
- Meng, Y., & Gan, J. (2008). A distributed swarm intelligence based algorithm for a cooperative multi-robot construction task. *Proceedings of IEEE Swarm Intelligence Symposium*, 21–23. doi:10.1109/SIS.2008.4668296
- Milutinovic, D., & Lima, P. (2002). Petri net models of robotic tasks. *Proceedings of the IEEE International Conference on Robotics and Automation*, 4, 4059–4064.
- Nishiwaki, K., Chestnutt, J., & Kagami, S. (2014). Planning and control of a humanoid robot for navigation on uneven multiscale terrain. In *Experimental Robotics, Advanced Robotics* (Vol. 79, pp. 401–415). Springer. doi:10.1007/978-3-642-28572-1_28
- Pfeifer, R., Lungarella, M., & Iida, F. (2007). Self-organization, embodiment, and biologically inspired robotics. *Science*, 318(5853), 1088–1093. doi:10.1126/science.1145803 PMID:18006736
- Salazar-Ciudad, I., Garcia-Fernandez, H., & Sole, R. V. (2000). Gene networks capable of pattern formation: From induction to reaction-diffusion. *Journal of Theoretical Biology*, 205(4), 587–603. doi:10.1006/jtbi.2000.2092 PMID:10931754
- Shen, W., Will, P., Galstyan, A., & Chuong, C.-M. (2004). Hormone-inspired self organization and distributed control of robotic swarms. *Autonomous Robots*, 17(1), 93–105. doi:10.1023/B:AURO.0000032940.08116.f1
- Shi, Z., Tu, J., Zhang, Q., Liu, L., & Wei, J. (2012). A survey of swarm robotics system. *Advances in Swarm Intelligence. Lecture Notes in Computer Science*, 7331, 564–572. doi:10.1007/978-3-642-30976-2_68
- Suzuki, R., Katsuno, I., & Matano, K. (1971). Dynamics of neuron ring. *Kybernetik*, 8(1), 39–45. doi:10.1007/BF00270832 PMID:5558431
- Taylor, T. (2014). A genetic regulatory network-inspired real-time controller for a group of underwater robots. *Proceedings of the 8th Conference on Intelligent Autonomous Systems*, 403–412.
- Tinbergen, N. (1951). *The Study of Instinct*. Oxford, UK: Oxford Univ. Press.
- Werfel, J. (2004). Building blocks for multi-robot construction. *Proceedings of 7th International Symposium on Distributed Autonomous Robotic Systems*, 285–294.

Yasuda, G. (2014). Behavior-based autonomous cooperative control of intelligent mobile robot systems with embedded Petri Nets. *Proceedings of the 8th International Conference on Soft Computing and Intelligent Systems, the 15th International Symposium on Advanced Intelligent Systems (SCIS&ISIS 2014)*, 1085-1090. doi:10.1109/SCIS-ISIS.2014.7044830

KEY TERMS AND DEFINITIONS

Behavior: Sequence of interactions between an organism and its environment where the actions affect its own perceptions.

Environment Mapping: Building a natural internal representation of a robot's environment including places in the environment and other information such as target objects and obstacles.

Hierarchical Structure: Complex system structure organized into a hierarchy of levels of

abstraction such that a system at one level becomes a building block for the system defined at the next higher level.

Localization: Processing to maintain an on-going estimate of the location of a robot with respect to the environment map.

Modularity: Capability of reuse of the programs describing the construction of substructures as if they were subroutines that can be invoked multiple times.

Multi-Agent System: A system composed of multiple cooperative agents or robots, where each controls its own local subsystem, communicating with other agents within limited range, to achieve a common goal collectively.

Subsumption Architecture: Control architecture consisting of a number of behavior modules that directly map sensation to action, arranged in a hierarchy similar to motor patterns used in animal ethology, where lower levels have an implicit priority over higher levels.

Improving Dependability of Robotics Systems



Nidhal Mahmud
University of Hull, UK

INTRODUCTION

The use of robotics systems is widespread and spans a variety of application areas. From health-care, to manufacturing, to nuclear power plants, to space missions, these systems are typically conceived to perform difficult, dangerous or critical tasks. The nature of such tasks (e.g., surgery operations, radioactive waste clean-up or space mining) places high demands on the dependability of robotics systems.

The preoccupations in the dependability of robotics systems are not new. Fault Tree Analysis (FTA; Vesely, 1981) and Failure Modes and Effects Analysis (FMEA; IEEE Std.352, 1987) are among the most often used techniques in various domains of robotics. For instance, Visinsky, Walker, and Cavallaro (1993) describe the use of FTA for robots operating in remote and hazardous environments. Other fields of application include industrial robots like in Karbasian, Mehr, and Agharajabi (2012), and modular and swarm robots like in Murray, Liu, Winfield, Timmis, and Tyrrell (2012).

The widespread use of FTA in the dependability assessment of complex systems is mainly due to the flexibility and ease of use of the fault trees. These are static (i.e., 'pure' Boolean) models, and therefore enable the use of efficient Boolean calculus in the elimination of component failures that are irrelevant to the total failure of the system. This logical reduction (known as qualitative analysis) simplifies the process to produce overall probabilities of system hazards (i.e., quantitative analysis). Nevertheless, such convenience comes with the loss of the significance of the sequencing

of failure events—i.e., the dynamic features often exhibited by modern systems cannot be captured by combinatorial models like this type of fault trees.

Robotics systems are certainly not an exception when it comes to sequence-dependent failures. For example, preclusion of the dynamic aspects due to the use of static fault trees in the analysis of modular robotic systems is clearly noted in Murray et al. (2012). To overcome such drawback, an alternative can be the utilization of fault trees that are extended with capabilities to capture the dynamic features. A well-known example is the Dynamic Fault Tree (DFT) approach (Dugan, Bavuso, & Boyd, 1992). This method was primarily conceived for quantitative analysis, which is often state-based (i.e., Markov analysis which is based on state transition diagrams [Markov models] is the DFT most prominent solving technique). That is, the full power of the Boolean methods was sacrificed here, especially when it comes to analyzing the dynamic parts of the system at the level of the fault tree (i.e., reducing the DFT).

Theoretically, some later research efforts have provided workarounds to the question of FTA with dynamic aspects. To deal with it, a technique which is relevant to this article consists of extending the Boolean methods with temporal logic calculus. In this connection, a set of temporal laws that enable qualitative analysis of fault trees extended with dynamic features can be found in Walker and Papadopoulos (2009). In the same vein, the algebraic formalism in Merle, Rousset, Lesage, and Bobbio (2010) proposes formal descriptions of dynamic behaviors and provides proofs of a number of theorems useful for the qualitative analysis of this type of fault trees. The

latter approach also deals with the corresponding probabilistic algebraic analysis.

In practice, automation of such advanced FTA as part of integrated dependability and systems engineering processes requires an automated generation and synthesis of these fault trees from failure behavioral models that are linked to the system specifications. The work in Mahmud, Walker, and Papadopoulos (2012) describes a suitable approach to generating and synthesizing fault trees that preserve the significance of the event-order from hierarchical models. Application areas for this approach include the automotive domain (Chen, Mahmud, Walker, Feng, Lönn, & Papadopoulos, 2013). More details about integration in an extended FTA through a Model-Based development process can be found in Kolagari et al. (2015).

In this article, emphasis is put on the significance of the sequencing of failure events and its implications in dependability analysis of robotics systems using FTA. Accordingly, a suitable technique for automated generation and synthesis of extended fault trees from system models is presented. Furthermore, we outline a novel approach to automated reduction of these fault trees. The article is structured as follows: the background section provides a literature review and highlights the relevant approaches to generating and synthesizing fault trees from systems models. The next section outlines an algorithm for the logical reduction of fault tree algebraic expressions that are extended with temporal semantics. Then, an advanced fault tree synthesis approach is presented in the following section. Finally, we present some future research directions, then we conclude.

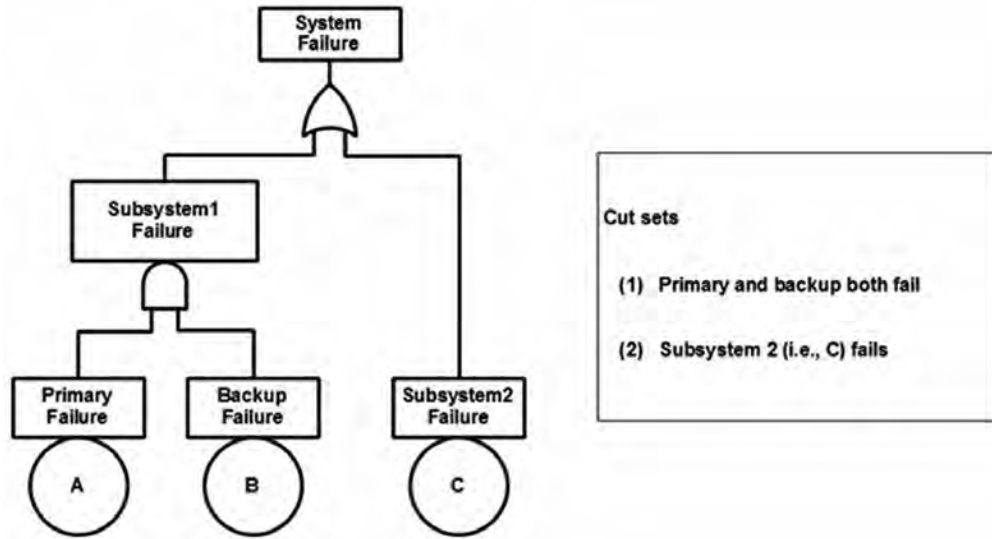
BACKGROUND

A number of approaches to generating fault trees from system models can be found in the literature. However, a special focus is put on automata representations of the dysfunctional behaviors of systems from which the fault trees get generated.

In this regard, the existing approaches fall into two main categories. The first category concerns the generation of static fault trees; it includes approaches that have been used in the context of influential modeling languages like Altarica (Rauzy, 2002) and AADL (Joshi, Vestal, & Binns, 2007). The former has been used in several aerospace projects including Airbus civil aircraft programs. The latter is increasingly being accepted by the aerospace community as a future standard. Moreover, there have been efforts recently to generate static fault trees from modeling languages specific to robotics systems, e.g. RobotML (Yakymets, Dhouib, Jaber, & Lanusse, 2013).

The second category of methods for fault tree generation includes approaches to generating dynamic fault trees (Dehlinger, & Dugan, 2008), and fault trees that are extended with the significant temporal information about the order in which failure events occur (Mahmud, Papadopoulos, & Walker, 2010; Mahmud et al., 2012; Mahmud & Mian, 2013). The DFTs are solved quantitatively, often, by converting them into equivalent Markov models. Although, there have been efforts on the minimization of the obtained Markov models (Crouzen, Hermanns, & Zhang, 2008; Boudali, Crouzen, & Stoelinga, 2010), there is still less focus on qualitative analysis at the level of the fault tree in this method. However, the approach in Mahmud et al. (2010, 2012) and Mahmud and Mian (2013) allows to reduce the extended fault trees into equivalents which can be proven by using the temporal laws and theorems in Walker and Papadopoulos (2009) and Merle et al. (2010). This is a State Automata-based top-down deductive approach to Fault tree synthesis with Order-dependent behaviors, see SAFORA in Mahmud (2012). It has demonstrated its value in the framework of a rich model-based design founded on EAST-ADL (Chen et al., 2013; Kolagari et al., 2015). EAST-ADL is an approach for describing automotive electronic systems through an information model that captures engineering information in a standardized form (EAST-ADL, 2010).

Figure 1. Example of a SM path for the fault tree algebraic representation



The fault tree generation algorithm used in the Safora method is based on the approach outlined in Mahmud et al. (2010). It converts State Machines (SMs) showing transitions from normal to degraded and failed states into fault tree algebraic expressions. The transitions are labeled by failure events (denoted by upper case letters taken from the alphabet, e.g. A, B, C ... like in Figure 1) and the states that are relevant to the system dependability are marked (denoted by the lowercase letter q with the subscript m). An algebraic expression is a logical combination of failure events —using the logical AND (symbol ‘.’) and the logical OR (symbol ‘+’) — which can be extended with sequence information by using the Priority-OR temporal logic operator (symbol ‘|’). The latter, abridged as POR, can be found in Walker and Papadopoulos (2009) and represents a priority situation where one event must occur first and other events may or may not occur subsequently. For example, in the expression ‘A|B’, A occurs and B does not occur; or A and B both occur, but in sequence. Similarly, in A|B|C, A occurs first or alone; should B and/or C occur, this must happen after A, but the order in which B and C occur is unimportant.

A sequence can be represented by using only ‘.’ and ‘|’, which is a central issue in Mahmud et al. (2010). For example, the sequence A occurs first then B occurs second, which is denoted by $A < B$ (‘<’ represents the Priority-AND operator, which is abridged as PAND), can be represented by $A|B.B$, where ‘|’ has priority over ‘.’. Proofs of this equivalence and others that are useful to this article, like the following, can be found in Walker and Papadopoulos (2009) and Merle et al. (2010). We wish to note, though, that POR is referred to as non-inclusive BEFORE in Merle et al. (2010).

$$A < B \Leftrightarrow (A|B).B$$

$$A < B + B < A \Leftrightarrow A.B$$

$$(A|B).(A|C) \Leftrightarrow A|B|C$$

$$A|(B+C) \Leftrightarrow A|B|C$$

In Mahmud et al. (2010), the algebraic expression deduced from a marked state (denoted by q_m) in a SM is essentially the disjunction over the paths π from q_m to the initial state (traversed backwards) of the conjunction of the events that

label π . Moreover, an event E , which is incident from a state q in which paths diverge, will have priority over any other event E' , also incident from q , if and only if there is an event in the path starting from q through E to q_m , which is repeated (i.e., reappearing) in a path starting from q through E' to any state. For example, the (sub) path from q_i to q_m (top right of Figure 1) is essentially represented by $C.\pi_2$, where C is the sub-path's event which is incident from q_i and π_2 is the conjunction of the remaining events to reach q_m . Since C is repeated in a path divergent from q_i (slightly downwards in the figure), also there is an event F in π_2 which is repeated in another divergent path (to the bottom of the figure), C will then have priority over D and E , the other events which are also incident from q_i . Therefore, $C.\pi_2$ is now fine-tuned as $CIDIE.\pi_2$. Similarly, $A.\pi_1$, to the left of Figure 1, is essentially the algebraic representation of the sub path from q_0 (i.e., the initial state) to q_i . This representation is also fine-tuned into $AIB.\pi_1$ since there is an event B that is repeated in a divergent path. Therefore, the full path from q_0 to q_m is algebraically represented by $AIB.\pi_1.CIDIE.\pi_2$, and without any loss of the significance of the event order. There might be other paths reaching q_m , which are denoted as disjunctions (sums) of conjunction (product) terms. Thus, the marked state expression is directly obtained in a standard sum-of-product canonical form like in Equation (1), which is convenient for the logical manipulation thereof.

$$q_m = \sum \left(\prod E_i . \prod (E_j | E_k) \right), j \neq k \quad (1)$$

where ' Σ ' and ' Π ' represent respectively the Boolean 'OR' and 'AND'.

REDUCTION ALGORITHM

In this section, we propose an algorithm for the logical reduction of the generated algebraic expressions. The algorithm is demonstrated on

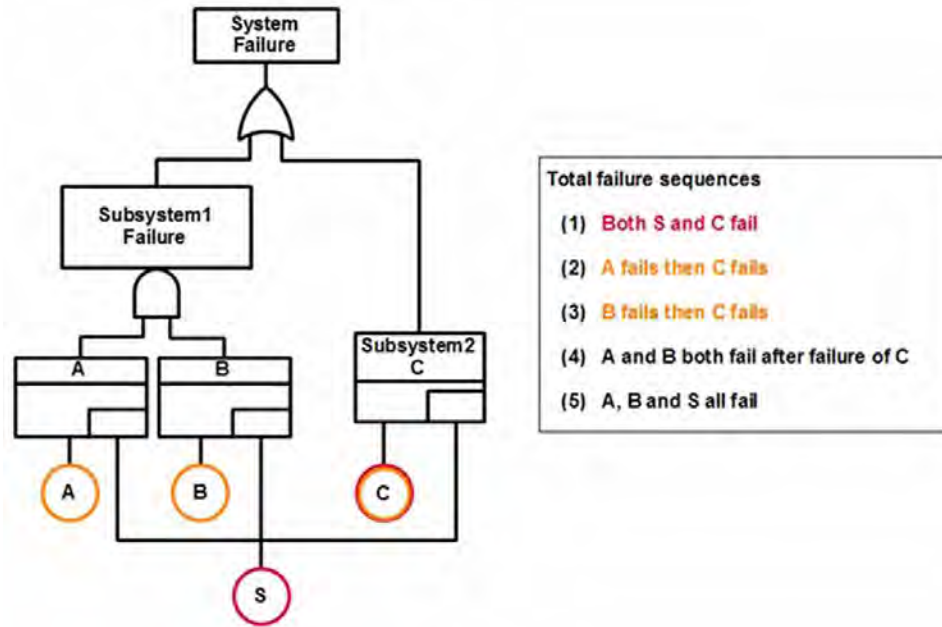
a primary-standby (PS) redundant system, like the example in Figure 2 (a). "A" is the primary component, "B" is the backup, and "S" is a monitoring sensor whose role is to activate B upon detection of a deviation from the output of A (e.g., omission of output). "I" represents the input to each of the two redundant components, and "Out" is the system output. Out must receive input from either A or B for a continuous operation of the system. This pattern is not uncommon in dependability-critical systems as A and B can be any sensing, control or actuating device. A SM example which describes the PS failure behavior is shown on Figure 2 (b), where $X \in \{A, B, S\}$ denotes an internal failure (i.e., basic event) of the corresponding component to the left of the figure. For the sake of clarity, the SM is simplified by assuming that input I is always provided, Out has no failure mode on its own, and total failure states are all merged into one marked state (i.e., state 7). Figure 2 (b) shows that failure of either primary or backup component needs to be followed, whether immediately or not, by failure of the other unit in order to reach state 7. Also, failure of the sensor is relevant to the total failure of the system only if it occurs too soon—i.e., S fails before failure of the monitored (primary) component occurs. Otherwise, the system can still function in standby mode upon failure of A.

Application of the conversion algorithm outlined in the previous section gives the following algebraic expression for state 7, and in the form of a sum of Product Terms (PTs) as in Equation (1):

$$\begin{aligned} \text{state 7} = & B.SIB.AISIB + (\text{PT\#1}) \\ & A.BIA.SIBIA + (\text{PT\#2}) \\ & A.SIA.BISIA + (\text{PT\#3}) \\ & AIB.SIBIA + (\text{PT\#4}) \\ & BIS.AISIB + (\text{PT\#5}) \\ & AIS.BISIA (\text{PT\#6}) \end{aligned}$$

Notice that there are 6 product term expressions, each corresponds to one path reaching state 7.

Figure 2. (a) a primary-standby example system and (b) the corresponding failure behavior



In fact, each product term $\prod E_i \cdot \prod (E_j | E_k), j \neq k$ in Equation (1) is a sequence set. For example, $A.B|C = \{ABC, BAC, BCA, AB, BA\}$. In the 2-event sequences (i.e., AB and BA), C does not occur. In other words, $A.B|C$ is the set of all sequences of A, B, and C should C occur, subject to inclusion of any order in which A and B occur (logical AND) and exclusion of the order in which C occurs before B (due to the term 'B|C')—assuming that simultaneity of occurrences of events is due to common cause failures, e.g., a power failure which causes several components to fail at the same time. Such failures are treated separately and are assigned separate failure rates should they occur.

Concerning the expression of state 7, we have: PT#1 is the singleton $\{ASB\}$. Similarly, PT#2 = $\{SBA\}$ and PT#3 = $\{BSA\}$. However, PT#4 = $\{SA, SAB\}$, PT#5 = $\{AB, ABS\}$, and PT#6 = $\{BA, BAS\}$. This way, each PT can be interpreted as a set of sequences as well as an algebraic expression—i.e., able to be manipulated by unions or by disjunctions (sums). Let's say we have n product terms denoted by $PT_i, i=1,2,\dots,n$. The canonical

form minimization algorithm proposed in (Merle et al., 2010) removes the redundant elements from the set S of all product terms.

$$S = \bigcup_{i=1}^n PT_i \quad (2)$$

According to Merle et al. (2010), a product term PT_i is redundant if:

$$PT_i \cdot \sum_{j \neq i} PT_j = PT_i \quad (3)$$

This algorithm aims at removing the completely redundant PTs. In many cases, we also need to check whether some events are irrelevant, especially when it comes to the order-sensitive part (i.e., the POR input events) of a PT. For instance, while the algorithm doesn't remove any $PT_i, i=1,\dots,6$ that compose the canonical form of state 7 of the PS example system, this doesn't mean that they cannot be reduced. Observing, e.g., $PT\#4 = \{SA, SAB\}$, it shows that B may not occur, or may occur but after A. The other

options to reach state 7 are via $PT\#2 = \{SBA\}$, which means that B occurs between S and A, and via $PT\#3 = \{BSA\}$ which means that B, S, and A, all occur in sequence. In other words, there is a sub-sequence SA common to $PT\#2$, $PT\#3$, and $PT\#4$; whether B occurs before SA, in the middle of SA, after SA, or doesn't occur at all, simply means that B is irrelevant and can be removed, thereby merging these product terms into a reduced (minimal) product term represented by $PT\#2,3,4 = \{SA\}$. But, the intermediate $PT_i, i \in \{2, 3, 4\}$ need not be removed yet, as they can serve to merge and reduce other PTs.

For example, $PT\#6 = \{BA, BAS\}$ can be merged with $PT\#2$ and $PT\#3$, then reduced into $PT\#2,3,6 = \{BA\}$ —since, overall, whether S is a prefix with BA (i.e., $PT\#2$), S occurs in the middle of BA (i.e., $PT\#3$), or S is a suffix with BA or does not occur at all (i.e., $PT\#6$), all show that S is irrelevant. Similarly, the $PT\#i, i \in \{2, 3, 6\}$ need not be removed yet until all iterations are processed. The last iteration concerns $PT\#5 = \{AB, ABS\}$ which can be similarly merged with $PT\#1$ and $PT\#4$, then reduced into $PT\#1,4,5 = \{AB\}$.

At this stage, state 7 (Figure 2b) canonical form is reduced (minimized) to $S = PT\#2,3,4 \cup PT\#2,3,6 \cup PT\#1,4,5$. That is, $S = \{SA, BA, AB\}$, which is algebraically written as:

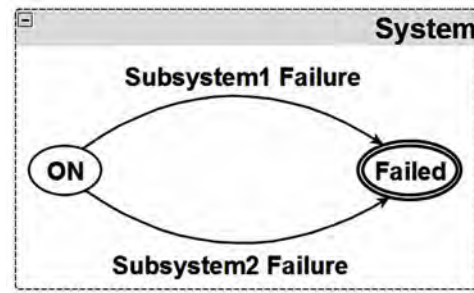
$$Sal_g = A.S|A + A.B|A + B.A|B \quad (4)$$

The intermediate $PT\#i, 1 \leq i \leq 6$ are now each completely redundant with Sal_g (i.e., $PT\#i, Sal_g \supset PT\#i$) and can be removed as they cannot serve to reduce a PT any further.

Equation (4) can be written as $Sal_g = S < A + A.B$ which is graphically represented by the reduced SM in Figure 3a (with 3 full paths only instead of 6 initially), and by the reduced fault tree shown in Figure 3b where the gate to the top of the figure is a logical OR, to the left is a logical AND, and to the right is a Priority-AND.

For large systems, this approach is practical via composition. That is, failure expressions local to the system constituents are generated from their

Figure 3. (a) reduced SM and (b) reduced DFT of the PS example system



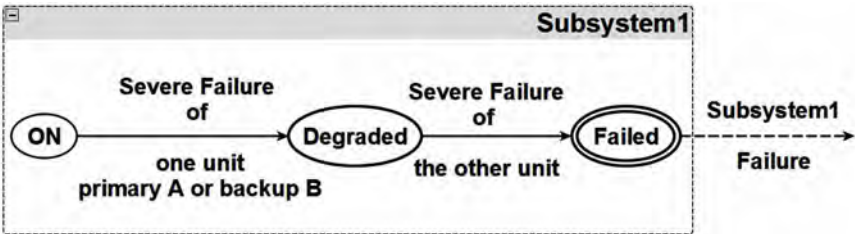
corresponding SMs, reduced as outlined in this section, and then synthesized into expressions of system failure conditions which likewise can be reduced further if necessary (Mahmud et al., 2012; Mahmud, 2012).

FAULT TREE SYNTHESIS

For instance, let us consider the example adapted from Visinsky et al. (1993) of two redundant sensors in a single robot joint. These are the primary and backup sensors which compose Subsystem1 of Figure 4a (i.e., the fault tree representation of the system). The basic events A and B (denoted by circles) correspond to the failures of primary and backup sensors, respectively. Subsystem1 fails if both sensors fail, hence the logical AND gate on the left-hand side of the figure. Moreover, the motion at the joint is coupled with another motion (Subsystem2) such that failure of either motion causes a joint total failure, and hence the logical OR gate to the top of the figure. However, Subsystem2 is not supplied with a fault tolerance mechanism. Thus, its basic event C is a single point of failure.

To increase the reliability of the system, the actively redundant sensors of Subsystem1 and the sensor of Subsystem2 all share the same spare unit S, as in Figure 4b —assuming that they are all functionally equivalent. S is supposed to take over any unit (A, B, or C) whichever fails first. In case a unit fails but cannot be rescued by the

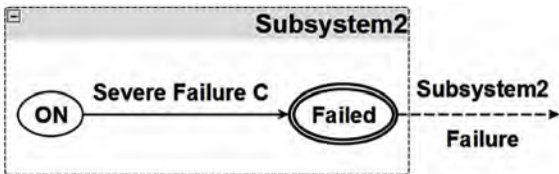
Figure 4. (a) fault tree of a robot joint sensing system with active redundancy and (b) with an extra passive redundancy



spare, the output event of the corresponding spare gate to the middle of the figure becomes true (i.e., a fault event). A spare unit cannot take over a failed unit if it is also failed or is not available (e.g., it took over another unit that failed earlier). Such capability can be useful to implement cost-effective fault tolerance solutions that meet dependability requirements, especially if the sensors are unlikely to fail.

Now the order in which failure events occur can affect the overall outcome (i.e., total failure of the system). For example, if A fails first, it will be rescued by the spare, then if C fails second, in this case the spare is no longer available to take over C, and thereby causing a total failure. However, if C fails first, then it will be rescued by S, but a

Figure 5. Highly abstract state machine depicting the monolithic behavior of the system



subsequent failure of A is not severe since A is in active redundancy with B.

The dysfunctional behavior of this example system can be seen as an abstraction of an automata-hierarchy describing its embedded components. That is, it can be highly abstracted as in Figure 5, and the SMs of the constituents

Figure 6. Abstract state machine describing the dysfunctional behavior of Subsystem1

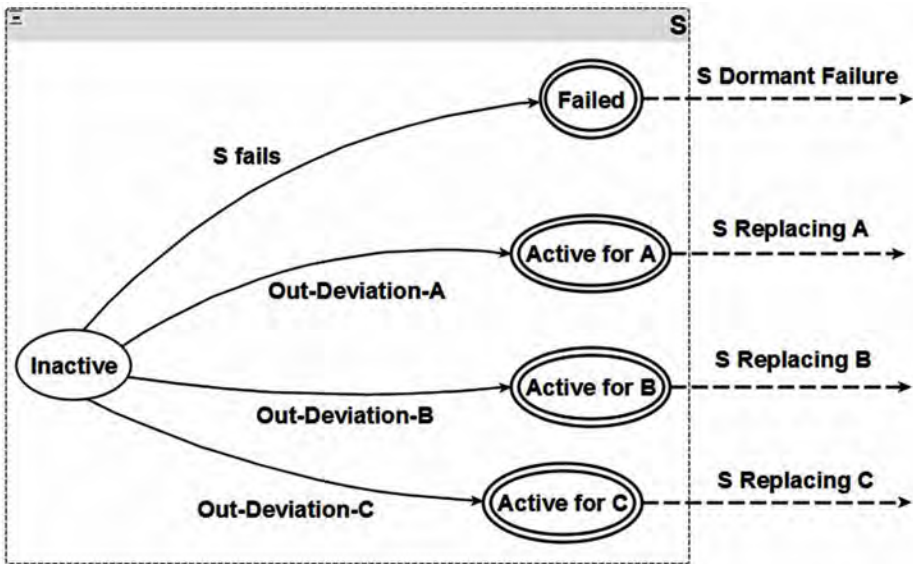
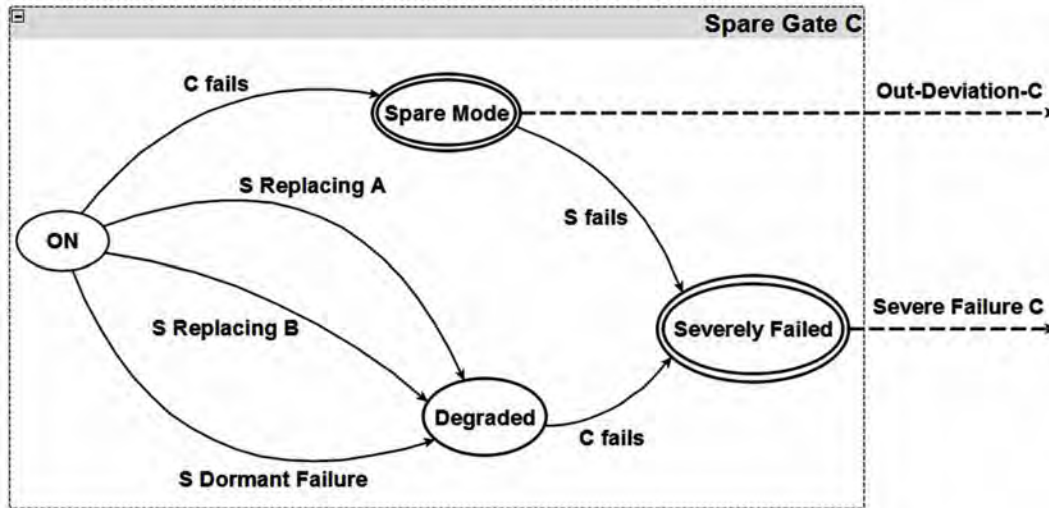


Figure 7. Abstract state machine describing the dysfunctional behavior of Subsystem2



located at lower levels of the hierarchy. An event causing a transition into a SM is typically either a basic event or a non-atomic event. The former means a component internal failure and the latter is due often to a fault caused by failures in some other components.

For example, Figure 5 shows that the system is initially working (state ON), but it totally fails due to propagation of errors caused by failures occurring downwards in the event causal chain,

i.e. from either Subsystem1 (Figure 6) or Subsystem2 (Figure 7).

Figure 6 shows that Subsystem1 is initially ON, but is seriously degraded as soon as one of its actively redundant units fails. A complete failure of Subsystem1 occurs upon the second severe failure (propagated from the other redundant unit).

Similarly, Subsystem2 fails upon a severe failure of its unit C, which is backed by S.

The spare unit S (Figure 8) is in passive redundancy and is, thus, inactive initially. Since

Figure 8. State machine of the spare unit S

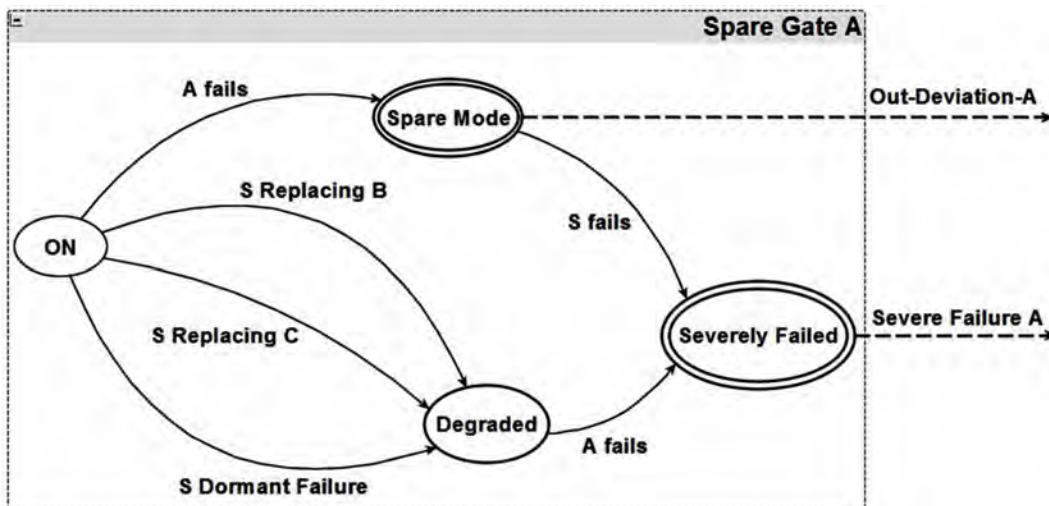
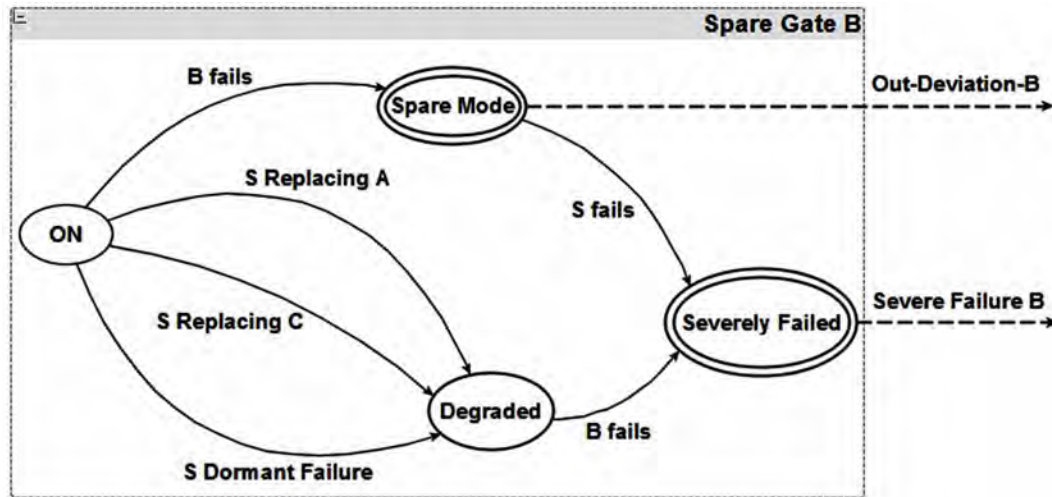


Figure 9. State machine of the spare gate C



it is shared between spare gates A, B and C; it would take over any of these, whichever fails first (i.e., in response to the earliest deviation detected from the output of a unit). Alternatively, it may fail dormant.

Since all sensors of the robot joint sensing system exhibit similar dysfunctional behavior, we put emphasis on one unit, e.g. the spare gate C (Figure 9). Initially operational (state 'ON'), if C fails before any loss of the spare; then the situation is still safe, which means that the spare gate C is functioning in spare mode, in which case only a subsequent failure of S would worsen the situation. However, if C fails after the loss of the spare, the situation becomes severe.

Application of the generation algorithm on the SMs of the constituents and the reduction algorithm on the generated expressions, then synthesizing the obtained reduced expressions, ultimately yield the following result for the global failure condition, and without any loss of the significance of the event-order:

$$\text{Total failure} = S.C + A.C + B.C + S.A.B + C.A.C.B.$$

This result shows that the system fails due to any of the following causes: both S and C fail; A

fails then C fails; B fails then C fails; A, B and S all fail; or A and B both fail after failure of C.

FUTURE RESEARCH DIRECTIONS

Automation of the generation, synthesis, and analysis of advanced fault trees from RobotML models as part of integrated model-based design and analysis processes is an open issue. Moreover, there is still a lack of tools for the automated analysis of fault trees that are extended with event-order features. This is mainly due to the complexity of the manipulation as well as probabilistic assessment of this type of fault trees. Also, repairable events would be worth being considered as many users aim at modeling repairable systems. We wish to note, though, that the approach in this article does not totally preclude the existence of cyclic behaviors (failures and repairs)—i.e., after the transformation of the embedded acyclic behaviors into algebraic expressions, and the reduction thereof, the obtained results represent, therefore, minimal circuitless sequences onto which are grafted the repair circuits. However, this would imply analysis of (coherent) systems with perfect repair policy. Repairable systems with different maintenance policies are likewise worth being considered in future extensions of this work.

CONCLUSION

In this article, we outlined a lack of suitable techniques that generate, synthesize and analyze fault trees with the significant sequencing of events from models specific to robotics systems. Most of the existing work deals with static fault trees; less work is on generation of dynamic fault trees but with a focus on quantitative analysis. In this regard, this article puts emphasis on such important issue. The second contribution is a solution proposed for the generation then reduction of fault trees with more expressive power from automata descriptions of failure behaviors. We demonstrated the proposed algorithm on a primary-standby example system and shown the obtained analysis results. Moreover, a compositional approach to computing fault tree expressions of global failure conditions from the dysfunctional behavior local to the components of the system was presented. An example depicting a robot joint sensing system was used in the study. It has few components, but with very complex dependencies. This helped to illustrate the solution which sets the stage for a high-level of automation and integration of such advanced fault tree analysis with model-based design to effectively improve development of robotics systems.

REFERENCES

- Boudali, H., Crouzen, P., & Stoelinga, M. (2010). A rigorous, compositional, and extensible framework for dynamic fault tree analysis. *Dependable and Secure Computing. IEEE Transactions on*, 7(2), 128–143.
- Chen, D., Mahmud, N., Walker, M., Feng, L., Lönn, H., & Papadopoulos, Y. (2013). Systems Modeling with EAST-ADL for Fault Tree Analysis through HiP-HOPS. In *Proceedings of the 4th International Federation of Automatic Control (IFAC) Workshop on Dependable Control of Discrete Systems (DCDS)*, (pp. 91-96). IFAC doi:10.3182/20130904-3-UK-4041.00043
- Crouzen, P., Hermanns, H., & Zhang, L. (2008). On the minimisation of acyclic models. In *International Conference on Concurrency Theory* (pp. 295-309). Springer Berlin Heidelberg.
- Dehlinger, J., & Dugan, J. B. (2008). Analyzing dynamic fault trees derived from model-based system architectures. *Nuclear Engineering and Technology: An International Journal of the Korean Nuclear Society*, 40(5), 365–374. doi:10.5516/NET.2008.40.5.365
- Dugan, J. B., Bavuso, S. J., & Boyd, M. A. (1992). Dynamic fault-tree models for fault-tolerant computer systems. *Reliability. IEEE Transactions on*, 41(3), 363–377.
- EAST-ADL. (2010). *EAST-ADL Domain Model Specification, D4.1.1*. Retrieved from http://www.atesst.org/home/liblocal/docs/ATESST2_D4.1.1_EAST-ADL2-Specification_2010-06-02.pdf
- Joshi, A., Vestal, S., & Binns, P. (2007). Automatic generation of static fault trees from aadl models. *Workshop on Architecting Dependable Systems of The 37th Annual IEEE/IFIP Int. Conference on Dependable Systems and Networks*.
- Karbasian, M., Mehr, M. R., & Agharajabi, M. (2012). Improvement of the Reliability of Automatic Manufacture Systems by Using FTA Technique. *International Journal of Industrial Engineering*, 23(2), 85–89.

- Kolagari, R. T., Chen, D., Lanusse, A., Librino, R., Lönn, H., Mahmud, N., & Yakymets, N. et al. (2015). Model-Based Analysis and Engineering of Automotive Architectures with EAST-ADL: Revisited. *International Journal of Conceptual Structures and Smart Applications*, 3(2), 25–70. doi:10.4018/IJCSSA.2015070103
- Mahmud, N. (2012). *Dynamic Model-based Safety Analysis: From State Machines to Temporal Fault Trees* (Ph.D. dissertation). Department of Computer Science, University of Hull, Hull, UK.
- Mahmud, N., & Mian, Z. (2013). Automatic generation of temporal fault trees from AADL models. In *Safety, Reliability and Risk Analysis: Beyond the Horizon* (pp. 2741–2749). CRC Press. doi:10.1201/b15938-415
- Mahmud, N., Papadopoulos, Y., & Walker, M. (2010). A Translation of State Machines to Temporal Fault Trees. In *Proceedings of the 40th IEEE/IFIP International Conference on Dependable Systems and Networks*, (pp. 45-51). IEEE. doi:10.1109/DSNW.2010.5542620
- Mahmud, N., Walker, M., & Papadopoulos, Y. (2012). Compositional synthesis of temporal fault trees from state machines. *Performance Evaluation Review*, 39(4), 79–88. doi:10.1145/2185395.2185444
- Merle, G., Roussel, J. M., Lesage, J., & Bobbio, A. (2010). Probabilistic algebraic analysis of fault trees with priority dynamic gates and repeated events. *Reliability. IEEE Transactions on*, 59(1), 250–261.
- Murray, L., Liu, W., Winfield, A., Timmis, J., & Tyrrell, A. (2012). Analysing the Reliability of a Self-reconfigurable Modular Robotic System. In *Bio-Inspired Models of Networks, Information, and Computing Systems* (pp. 44-58). Springer Berlin Heidelberg. doi:10.1007/978-3-642-32711-7_4
- Rauzy, A. (2002). Mode automata and their compilation into fault trees. *Reliability Engineering & System Safety*, 78(1), 1–12. doi:10.1016/S0951-8320(02)00042-X
- IEEE Std.352. (1987). *IEEE Guide for General Principles of Reliability Analysis of Nuclear Power Generating Station Safety Systems*. ANSI/IEEE.
- Vesely, W. E. (1981). *Fault Tree Handbook*. US Nuclear Regulatory Committee Report, NUREG-0492.
- Visinsky, M. L., Walker, I. D., & Cavallaro, J. R. (1993). Robotic fault tolerance: algorithms and architectures. *Robotics and Remote Systems in Hazardous Environments*, 53-73.
- Walker, M., & Papadopoulos, Y. (2009). Qualitative temporal analysis: Towards a full implementation of the Fault Tree Handbook. *Control Engineering Practice*, 17(10), 1115–1125. doi:10.1016/j.conengprac.2008.10.003
- Yakymets, N., Dhouib, S., Jaber, H., & Lanusse, A. (2013). Model-driven safety assessment of robotic systems. In *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on* (pp. 1137-1142). IEEE. doi:10.1109/IROS.2013.6696493

ADDITIONAL READING

- Avizienis, A., Laprie, J.-C., & Randell, B. (2001). Fundamental Concepts of Computer System Dependability. *IARP/IEEE-RAS Workshop on Robot Dependability: Technological Challenge of Dependable Robots in Human Environments*.
- Dal Cin, M., Huszerl, G., & Kosmidis, K. (1999). Quantitative evaluation of dependability critical systems based on guarded statechart models. In *High-Assurance Systems Engineering, 1999. Proceedings. 4th IEEE International Symposium on* (pp. 37-45). IEEE. doi:10.1109/HASE.1999.809473

KEY TERMS AND DEFINITIONS

AADL: AADL is an Architecture Analysis and Design Language standardized in 2004 by the Society of Automotive Engineers. It is used in the specification and analysis of the software and hardware architecture of real time embedded systems. AADL is mainly devoted to performance-critical aerospace and automotive systems.

Dynamic Fault Tree: Dynamic Fault Tree (DFT) was invented in response to a shortage in modelling sequence-dependent failures by using standard combinatorial fault trees. It thus added capabilities for capturing the dynamic aspects exhibited by modern complex systems. The DFT is designed primarily for quantitative analysis, in general by using Markov techniques. However, there has been less focus on qualitative analysis at the level of the DFT, i.e. determination of its (minimal) cut-sequences.

EAST-ADL: EAST-ADL is an Electronic Architecture and Software Technology Architecture Description Language for automotive embedded systems, which was developed by a consortium of universities and automotive companies. The language was further refined within the framework of the Model-based Analysis & Engineering of Novel Architectures for Dependable Electric Vehicles (MAENAD) EU FP7 project. Aspects covered by EAST-ADL include vehicle features, functions, requirements, variability, dependability, software components, hardware components and communication.

Failure Mode and Effects Analysis: Failure Mode and Effects Analysis (FMEA) is commonly known (and often used) as a bottom-up analysis technique. It proceeds by analyzing the system components individually, or sometimes collec-

tively, to inductively derive the consequences of their failures on the system. FMEA aims at addressing those effects and the technique is widely practiced in reliability engineering in high-hazard industries.

Fault-Tolerance Systems: Fault-tolerance enables a system to continue its intended operation in the event of faults in some parts of it. Fault-tolerance systems rely typically on redundant units. The redundant components can be active and operating in parallel, or passive but switched into active use upon failure of the primary units. They can also be designed to be shared by a number of functionally identical systems.

Fault Tree Analysis: Fault Tree Analysis (FTA) postulates a hazard (top event of the fault tree) which must be avoided. It then reasons backward to identify all logical combinations of events which could lead the system to that hazard. FTA can be quantitative by combining figures for component failure rates to calculate overall probabilities of system hazards. It can also be qualitative by eliminating the component failures that are irrelevant to the total failure of the system, and thereby attempting to produce minimal failure scenarios. FTA is widely practiced in reliability engineering and in high-hazard industries.

State Automata to Fault Trees: State Automata to Fault trees with Order-dependent behaviors (SAFORA) is a top-down deductive synthesis approach that computes expressions of global failure conditions from the dysfunctional behavior local to the components of the system (described as hierarchical state machines). The purpose is to accurately analyze the reliability of architectures and the technique is suitable for complex systems featuring dynamic aspects.

Robotics and Programming Integration as Cognitive–Learning Tools

R**Nikleia Eteokleous***Frederick University Cyprus, Cyprus*

INTRODUCTION

The technological improvements within the robotics field and its expansion to various fields such as medicine, industry and education, calls for robotics integration within the educational practice as learning tools. Alimisis (2012) supports that robotics draw the attention and interest of academicians, researchers and teachers in all educational levels as well as other stakeholders (policy makers and society leaders). Educational systems are responsible in preparing students (future citizens) for this ever-changing Hi-Tech, globalized, interconnected world. Numerous 21st century skills are reported in the literature as important to be developed by future citizens as the means to address the needs and demands of the society. Digital literacy is one of them and robotics and programing are becoming important elements within the educational settings. The students need to be provided with the opportunities to experience tinkering, fabrication, design and create technological artifact & interactive objects, construct their own meaningful projects, experience the scientific method of inquiry (Bers, 2008a; Bers, 2008b; Bers, Matas & Libman, 2013; Bernstein, Mutch-Jones, Cassidy, Hamner, & Cross, 2016; Eteokleous, 2016). Consequently, educators need to design learning environments enhanced by new technologies where students have the opportunity to experience them as cognitive-learning tools within their learning processes. In order for robotics to be integrated within the educational practice,

teachers need to be appropriately and adequately prepared by universities (for pre-service teachers at the undergraduate level and in-service teaches at the graduate level) and professional development authorities (for in-service teachers) (Vollsted, Robinson, & Wang, 2007).

The current chapter takes into consideration numerous research studies suggesting that robotics integration for educational purposes is an effective teaching method; arguing that if robotics activities are appropriately designed and implemented have great potential to significantly improve and enhance the teaching and learning process (Bauerle, & Gallagher, 2003; Benitti, 2012; Bers, Flannery, Kazakoff, & Sullivan, 2014; Eteokleous, Demeetriou, & Stylianou, 2013; Papert, 1993). Robotics in the classroom has taken a global momentum especially because of its positive contributions in the teaching of science, technology, engineering and mathematics (STEM) (Benitti, 2012; Bers, et al, 2014; Nugent, et al., 2009; Sullivan, 2008; Williams, Ma, Prejean, Lai, & Ford, 2007). Additionally, research has shown that robotics integration in education promotes the development of various non-cognitive skills, however extremely important life skills. For example, reasoning, problem solving, tinkering, sequencing, computational thinking, decision making, scientific investigation, collaboration, knowledge construction, critical thinking, creativity, communication (Bers et al., 2002; Benitti, 2012; Bers, 2008a; Bers, et al., 2014; Chambers & Carbonaro, 2003; Eteokleous, 2015; Eteokleous, 2016; Miglino, Lund, &

Cardaci, 1999; Resnick, Berg, & Eisenberg, 2000; Sullivan, 2008; Williams, et al., 2007; Williams, Ma, & Prejean, 2010).

Main Aim

Having in mind the above, the current study moves one-step further aiming to examine another aspect of educational robotics. The study focuses on robotics and programming as medium for developing cognitive skills in disciplines non-related to STEM. The purpose of the study is to examine how robotics and computer programming can be employed in the elementary classrooms (2nd to 6th grade) in order to achieve disciplinary learning objectives across disciplines (besides STEM). The following research questions guided the current study:

- In what degree did the disciplinary learning objectives are met? What did students learn when robotics and programming were used as cognitive-learning tools across disciplines?
- In what ways did teachers integrate robotics and programming across disciplines? What pedagogical and instructional approaches were employed by the teachers?
- What factors may influence robotics and programming integration as cognitive-learning tools in meeting disciplinary learning objectives?
- What were teachers' experiences, reflections, problems faced, and lessons learned from robotics and programming integration across disciplines?

BACKGROUND

Educational Robotics as a Cognitive - Learning Tool

It is suggested that there are two approaches related to robotics in education; 1) robotics as a

subject matter and 2) integrating robotics as a tool within the educational practice to achieve specific learning objectives (Eteokleous, et al., 2013). The integration of robotics as a subject matter, as an autonomous entity, and not within a well-designed lesson plan, provide limited educational potential and value. On the other hand, robotics integration as a learning tool, in selected teaching cases exploits its full potential; therefore, it upgrades and enhances the teaching and learning process and promotes school transformation (Eteokleous, et al., 2013). The intention of this approach is not to learn how to use the robotics package, and its programming software, but to use it as a tool within a specific educational context to meet disciplinary learning objectives (Bers, 2010; Bers, et al., 2014; Bernstein, et al., 2016, Eguchi, 2007, Eteokleous, et al., 2013). In other words, robotics is employed as a tool to teach and deliver concepts within various disciplines such as Mathematics, Engineering, Science, Physics, and even in non-technology related fields such as Language and Linguistics, Arts, Biology, Psychology (Bers, Ponte, Juelich, Viera & Schenker, 2002; Eguchi, 2007, Eteokleous, et al., 2013). *Robotics integration* in the teaching and learning practice is defined as the use of robotics by students as a tool that enhances their learning experience and supports the achievement of specific learning objectives (Eteokleous, et al., 2013, Thomaz et al, 2009).

This approach is related to the *learning with computers* or computers as mindtools (Jonassen, 1999a), where computers and overall technology is introduced as students' partners within the teaching and learning process. *Learning with technology* requires integrating computers and overall technology as mindtools in the classrooms to support constructive learning. Educators embed technology capacity in the context of ongoing teaching and learning in different school subjects. Thus, students learn how to use technology not as an end in itself, but as a tool that help them execute their tasks and promote the balanced development of their mental abilities. As a result they do not learn from technology, but technologies support

meaning generated by students. (Becker, 1993; Becker & Ravitz, 2001; Bielaczyc & Collins, 1999; CTGV, 2003; Cuban & Pea, 1998; Earle, 2002; Haugland, 2000; Jonassen, 1999a, 1999b).

A definition provided by Jonassen (1999b), successfully captures the essence of using various technology means as mindtools. He writes “Mindtools are computer-based tools and learning environments that have been adapted or developed to function as intellectual partners with the learner in order to engage and facilitate higher order thinking and learning” (p. 10). Jonassen (2000) was the first to provide the stimulus and motive for developing the appropriate theoretical background for the integration of robotics (technology) as an educational tool. He argued that technology and the various technological means can be considered as cognitive tools or “Mindtools”, which enhance and enrich the educational process (Jonassen, 2000; Chambers & Carbonaro, 2003).

The same approach can be applied for robotics and programming as part of the new technologies. By replacing the words *computer-based tools*, *computers* and *technology* with *robotics and programming*, the same approach is suggested by the current study to be used in the educational settings. Robotics and programming are employed as cognitive-learning tools to meet disciplinary learning objectives. For years, educational technology researchers argued for importance of meaningful technology integration in the educational settings within appropriate learning environments, supported by pedagogical and theoretical frameworks. Bers (2008a, 2010) recently suggested that robotics and programming can be integrated in developmentally appropriate ways in the educational practice.

Development of Knowledge and Skills: Cognitive and Non-Cognitive

Robotics has been connected to constructivism (Bauerle & Gallagher, 2003; Williams, Ma & Prejean, 2010). The teaching and learning theory that supports constructivism; highlights

the educational value of such exercises as the ones the integration of robotics in the educational practice can provide. Such exercises are based on the philosophy of “learning by constructing”, giving students the opportunity to develop interactive “thinking objects” (Bers et al., 2014; Kafai & Resnick, 1996; Papert, 1993 ; Harel, & Papert, 1991; Papert, 1986). Students get into the process of evaluating, changing, differentiating, re-designing, re-constructing and re-programming their objects (Puntambekar & Kolodner, 2005). The construction of such objects helps students to represent reality (Bers et al., 2014; Eteokleous, 2016; Kazakoff, & Bers, 2012, Kazakoff, & Bers, 2014). It is in this context, where students have an active role, they understand the connection between theory and practice, relating the abstract to the concrete, as well as relate what they learn within the classroom to reality and nature (Miglino, Lund, & Cardaci, 1999; Sullivan & Bers, 2015; Sullivan & Moriarty, 2009).

Within a constructivistic learning environment, students have the opportunity to manage their learning; they are free to discover, and decide on how to proceed. It promotes a student-centered environment where students through collaboration and social interaction have active role in the teaching and learning process. Along the same lines, the developmentally appropriate integration of technology is informed by Piaget’s ideas and the theory of constructionism and constructionist (Kazakoff, & Bers, 2014).

Research has shown that robotics integration in education promotes the development of various non-cognitive skills such as thinking skills such as application, synthesis, evaluation, problem solving, decision making, and scientific investigation (Bers et al. 2002; Bers et al., 2014; Chambers & Carbonaro, 2003; Eteokleous, 2016; Kazakoff, & Bers, 2012, Kazakoff, & Bers, 2014; Resnick, Berg, & Eisenberg, 2000). Besides the achievement of non-cognitive skills, when students are provided opportunities to become involved with robotics and programming, deeper and better understanding and engagement with STEM content

can be achieved (Gura, 2011). Along the same lines, the developmentally appropriate ways of robotics and programming integration can promote students cognitive skills development in STEM and beyond (Kazakoff, & Bers, 2012, Kazakoff, & Bers, 2014). Additionally, disciplines integration improve learning, not only among subject matters related to STEM. Finally, it is argued that when students take the role of programmers new skill and concepts are likely to be developed (Kazakoff, & Bers, 2012, Kazakoff, & Bers, 2014).

The Educational Robotics Curriculum

The Robotics Academy at Frederick University (Department of Education) (<http://akrob.frederick.ac.cy>, (<https://www.facebook.com/AkadimiaRompotikis>), a research and educational unit that aims to promote and conduct research in the area of robotics education, was launched in order to establish and stimulate robotics mainly to the Cypriot educational system and society. Specifically, it researches how to best integrate robotics in the educational system as a subject-matter, as well as a cognitive-learning tool within the teaching and learning process (employ robotics as a tool in order to achieve disciplinary learning objectives and the development of non-cognitive skills). The Academy offers courses using various robotics packages. The target audience is educators and students (all educational levels) and anyone interested in robotics (hobbyists and professionals).

The Robotics Academy developed an innovative curriculum for students and educators in all educational levels. The curriculum is constructed based on the philosophy *Robots- Partners in Learning: Examine, Explore and Discover through Construction and Programming*, integrating various theories and approaches. The educational robotics curriculum aims to develop participants' knowledge and skills in robotics (theory), in constructing various robotics models and in programming the robots. Various educational robotics packages and visual programming platforms are

used. The participants are engaged in hands-on, technology-based as well as unplugged activities related to robotics, based on the grounds of gamification, project, problem and inquiry based learning.

MAIN FOCUS OF THE ARTICLE

Research Methodology

A descriptive case study methodology was implemented (Cohen, Manion, & Morrison, 2008; Yin, 2003) collecting mainly qualitative data by using the following data collection methods:

- Student portfolios: Students' programming samples (coding), and robotic projects.
- Teacher artifacts: Lesson plans, exercises, handouts, teacher reflective journals completed after each course
- Classroom observations
- Student assessment exercises/ rubrics
- Teacher technology self-efficacy questionnaire (prior classroom interventions) (Kassotaki & Roussos, 2006)
- Teacher personal interviews (30 minutes each) prior and after robotics integration.

The teachers voluntarily participated in the study following announcement of the opportunity via email to their respective principals. The students and parents completed consent forms. In order to provide equal educational opportunities to schools and educators, emails were sent to rural and urban public schools in order to have students from different areas and socio-economic status.

Three of the schools are located in rural areas and five in urban areas. Overall, 8 teachers (aged 30-45 years old) and 169 students participated in the classroom interventions, 82 boys and 87 girls, aged 7-12 years old (from 2nd to 6th grade). The classroom interventions took place in public elementary schools, in different educational elementary levels, and disciplines, employing three

robotics educational packages: the Bee Bots, Lego WeDo, Lego Mindstorms NXT. The classroom interventions varied in regards to duration (5 to 11 lessons, 40 minutes each) due to the grade, the robotics packages used and the discipline areas. All of the interventions took place across various disciplines (See Table 1) and lasted from September 2013 to June 2015.

The Robotics Academy had the overall responsibility to professionally train the teachers and provide the required robotics packages and software. All teachers participated at the same professional development training provided by the Robotics Academy staff. Specifically, the teachers attended one-week intensive professional development training on educational robotics. The training focused on the following thematic units: robotics and educational robotics, philosophy of integration, learning theories related to robotics integration, added value, pedagogical and instructional strategies and approaches to robotics integration, robotics integration across disciplines. The teachers were also trained on how to use three different robotics packages (building and programming the robots).

The teachers were given the flexibility to choose disciplines and topics that they thought it would better fit to robotics integration. Along the same lines, the teachers decided upon the pedagogical and instructional strategies to employ for robotics integration. They needed though to ensure that robotics were integrated as cognitive-learning tools to support disciplinary knowledge and achieving disciplinary-related learning goals. The same approach adopted for the assessment. Each teacher was responsible to develop his/her own assessment based on the course nature and the learning goals set, to assess students' understanding, knowledge and skills development. The teachers were given templates to complete in regards to their reflective journals and lesson plans. A classroom observation sheet and an interview protocol were also developed. The author provided academic and educational support, guidance and help when requested.

Qualitative data collected through student portfolios, teacher artifacts, classroom observations, assessment exercises/rubrics and teachers' personal interviews. The MonoConc Pro 2.0. software was used and the qualitative data were analyzed implementing the continuous comparison of data approach (Maykut & Morehouse, 1994). The SPSS (v.19) was used for analyzing the quantitative data collected through the assessment exercises/rubrics and the teacher efficacy questionnaire.

Teachers' Profile

Overall, eight teachers participated in the study working in three rural and five urban schools. There were four male and four female teachers participated at the study, aged 30-42 years old). Four teachers had some experience in integrating robotics and programming. For all teachers technology self-efficacy was rated as medium to high. All of the teachers held master's degree in numerous fields (Environmental Education, Curriculum and Instruction, Educational Leadership, Special Education) and only three of them had master's degree in Educational Technology. Gender and teacher's self-efficacy (comfort with technology) did not reveal to influence robotics integration, however their educational backgrounds did. All teachers appeared to be open to challenges, not afraid to take risks and it seems that this is what is needed in order to implement new approaches and new learning tools.

Robotics and Programming Integration Across Disciplines

The classroom interventions took place across various disciplines besides STEM (Life Education, Arts, Environmental Education, Language and Linguistics, Geography, Health Education, English) (See Table 1). At the cases where the Lego Mindstorms NXT package was used, the teachers spent time within the Design and Technology discipline to teach students about the robotics package and programming platform.

The same happened in regards when the Lego WeDo package was used, however less time needed (only one period – 40 minutes). In only two cases there was no clear connection between robotics and programming and the disciplinary content delivered. Two teachers did not manage to meet the disciplinary learning objectives and successfully integrate robotics and programming within the educational practice. The two teachers have some similar characteristics and several factors contributed to this result. Both of them were very optimistic and decided to integrate robotics and programming in several disciplines, their master degree was not in educational technology and did not have prior experiences to robotics and programming integration. Even though their self-efficacy was rated medium, when they faced technical problems they tried to solve them by themselves, refusing to request technical support.

Assessment Exercises / Rubrics

The assessment exercises/ rubrics showed that in six classroom interventions the disciplinary objective were achieved and students managed to master disciplinary learning in a great degree. Teachers were very satisfied with the results and reported that the results were an encouraging factor to repeat this experience for their students and themselves. The five most enthusiastic ones declared their interest in further enhancing robotics and programming integrating experience with more robotic packages and in an increased frequency throughout the school year.

Teachers' Pedagogical and Instructional Strategies Implemented

The teachers implemented various pedagogical and instructional strategies to robotics and programming integration. The variations were due to grade, duration, the disciplines involved, robotics package used, teacher's educational background and prior experience to robotics and programming. Furthermore, the extent to which robotics

and programming were integrated varied as well among the eight teachers. Six of the teachers successfully managed to find ways (adopted the appropriate pedagogical and instructional strategies) in connecting the disciplines to robotics and programming. While examining the data collected the question to be answered was: *What does it mean when a teacher manage to integrate robotics and programming to disciplinary learning?* For the purposes of this study, robotics and programming are integrated as cognitive-learning tools when the students managed to connect the discipline concepts/ terms (theory) to robotics and programming (practice). When robotics and programming gave the opportunity to students to represent reality. For example, when in Safety Education the students programmed the robot to stop when the light turned red and to move when the light turned green. The particular unit aimed in developing students' knowledge and skills in regards to the rules and regulations that vehicles and pedestrians need to follow when driving/ walking in the streets.

Additionally, the four teachers that had prior experience to robotics and programming (minimum experience) designed and delivered more complicated exercises. Additionally, they developed graded difficulty robotics and programming exercises. Two of them hold a master's degree in Educational Technology. These two teachers managed to easily and effectively design the appropriate learning environment and adopted appropriate pedagogical and instructional strategies. The other two *experienced* teachers managed to unlock hidden possibilities of Bee Bots integration with 5th graders.

Teachers' Experiences, Reflections, Problems Faced and Lessons Learned

The teachers overall reported positive experiences to robotics and programming integration. They felt that it was valuable and beneficial to their teaching and learning and to students' learning.

They had the chance through this process to learn many new things. Two of them also reported that it is important to take risks in order to succeed. Three of the teachers faced difficulties in designing and delivering their courses in order to meet the disciplinary objectives. Overall, the teachers did not face any serious technical problems. Teachers managed to solve by themselves the ones that appeared and in few cases they asked for assistance from the school IT department. Teachers needed lots of preparation time to design and organize the lessons. Five of them of them discussed that they found it necessary to make adjustments while the lessons where evolved based on the experiences of the previous courses.

Other Factors

Additionally, various factors revealed to influence robotics and programming integration. Gender and class size are not included in the list. However, students' group synthesis appeared to be a factor that needs special attention. Mixed-gender groups worked much better together in constructing and building the robots, though female single-gendered groups worked better throughout the lessons' activities. Additionally, the choice of the appropriate robotics package in relation to students' age, discipline and learning objectives is an important aspect to consider. The selection of the appropriate robotics package highly influence the activities/exercises to be designed and delivered. Further, the results showed that there is no need to use a more sophisticated robotics package to young age students (i.e. Lego Mindstorms NXT to be used with 2nd graders). On the other hand, robotics packages for lower elementary grades students (i.e. Bee Bots) can be used for older elementary students when the appropriate pedagogical approaches and instructional activities are adopted..

Solutions and Recommendations

Overall, the study reveals robotics' effectiveness as a tool to develop students' knowledge and

skills, and achieve specific learning objectives across various disciplines, beyond STEM. It also highlights the need for teachers to be professionally trained, supported and guided. Difference in teachers' experiences and educational backgrounds are important to consider. Finally, the study suggests the value of adopting appropriate learning pedagogies and teaching approaches when robotics is integrated within classroom activities. The potential of robotics and programming was revealed in disciplines beyond STEM. The focus should be to engage teachers in using robotics and programming in their classrooms across disciplines

FUTURE RESEARCH DIRECTIONS

The paper has important educational and theoretical significance. It adds to the body of literature related to robotics integration within the teaching and learning practice and its impact on disciplinary learning objectives and skills development. The results of the study revealed the great potential of integrating robotics as a cognitive-learning tool to achieve cognitive skills in disciplines non-related to STEM. It sets the foundation to promote research in the field of educational robotics in order to further examine and define the appropriate learning pedagogies and teaching approaches to be employed. Additionally, it underlies the value of integrating robotics as an innovative form of teaching and learning to be applied in schools, in order to promote the development of skills beyond STEM.

CONCLUSION

The current analysis drawn from two years integration of educational robotics and programming across various disciplines. Overall, the study demonstrated the positive impact of robotics integration in mastering discipline learning in elementary educational level. The analysis revealed that in

six out of the eight classroom interventions, the students managed to master discipline learning and meet the learning objectives set by the teachers. The study also revealed the variations and commonalities among the eight teachers as well as the factors that may influence robotics integration. In general, it is examined how teachers integrated robotics, in which degree the integration was achieved and the learning goals met. The study also managed to capture their experiences, reflections, ongoing challenges, anticipated outcomes and achievements. The results and discussion take into consideration teacher digital literacy/ comfort level with technology, educational background, experiences, robotics packages used, class size, teacher age, gender, and experiences.

REFERENCES

- Alimisis, D. (2012). Robotics in Education & Education in Robotics: Shifting Focus from Technology to Pedagogy. In *Proceedings of the 3rd International Conference on Robotics in Education 2012* (pp. 7-14). Charles University in Prague.
- Bauerle, A., & Gallagher, M. (2003). Toying With Technology: Bridging the Gap Between Education and Engineering. In *Proceedings of Society for Information Technology & Teacher Education International Conference 2003* (pp. 3538-3541). Chesapeake, VA: AACE.
- Becker, H. J. (1993). Teaching with and about computers in secondary schools. *Association for Computing Machinery*, 36(5), 69–55. doi:10.1145/155049.155066
- Becker, H. J., & Ravitz, J. L. (2001, March). *Computer Use by Teachers: Are Cuban's Predictions Correct?* Paper presented at the 2001 Annual Meeting of the American Educational Research Association, Seattle, WA.
- Benitti, F. (2012). Exploring the educational potential of robotics in schools: A systematic review. *Computers & Education*, 58(3), 978–988. doi:10.1016/j.compedu.2011.10.006
- Bernstein, D., Mutch-Jones, K., Cassidy, M., Hamner, E., & Cross, J. (2016). *Robots and Romeo and Juliet: Studying Teacher Integration of Robotics into Middle School Curricula*. Paper presented at the 2016 Annual Meeting of the American Educational Research Association, Washington, DC.
- Bers, M. (2008a). *Blocks to Robots: Learning with Technology in the early childhood classroom*. New York: Teachers College Press.
- Bers, M. (2008b). Engineers and Storytellers: Using robotics manipulatives to develop technological fluency in early-childhood. In O. Saracho & B. Spodek (Eds.), *Contemporary perspectives on science and technology in early childhood education* (pp. 105–125). Charlotte, NC: Information Age.
- Bers, M., Matas, J., & Libman, N. (2013). Linvot ULehibanot, To Build and to Be Built: Making Robots in Kindergarten to Explore Jewish Identity. *Diaspora, Indigenous, and Minority Education: Studies of Migration, Integration, Equity, and Cultural Survival*, 7(2), 164–179. doi:10.1080/15595692.2013.787062
- Bers, M. U. (2010). The TangibleK Robotics Program: Applied Computational Thinking for Young Children. *Early Childhood Research and Practice*, 12(2).
- Bers, M. U., Flannery, L. P., Kazakoff, E. R., & Sullivan, A. (2014). Computational thinking and tinkering: Exploration of an early childhood robotics curriculum. *Computers & Education*, 72, 145–157. doi:10.1016/j.compedu.2013.10.020
- Bers, M. U., Ponte, I., Juelich, C., Viera, A., & Schenker, J. (2002). Teachers as Designers: Integrating Robotics in Early Childhood Education. *Information Technology in Childhood Education Annual*, (1): 123–145.

- Bielaczyc, K., & Collins, A. (1999). Learning communities in classrooms: A reconceptualization of educational practice. In C. Reigeluth (Ed.), *Instructional-design theories and models: A new paradigm of instructional theory* (pp. 269–292). Mahwah, NJ: Erlbaum.
- Chambers, J. M., & Carbonaro, M. (2003). Designing, Developing, and Implementing a Course on LEGO Robotics for Technology Teacher Education. *Journal of Technology and Teacher Education*, 11(2), 209–241.
- Cognition and Technology Group at Vanderbilt. (2003). Connecting learning theory and instructional practices: Leveraging some powerful affordances of technology. In H. O'Neil & P. Perez (Eds.), *Technology application in Education: A learning view* (pp. 173–209). Mahwah, NJ: Erlbaum.
- Cohen, L., & Manion, L. (1994). *Research Methods in Education*. London: Routledge and Kegan Paul.
- Cohen, L., Manion, L., & Morrison, K. (2008). *Research methods in education* (5th ed.). London: Routledge.
- Creemers, B. P. M., & Kyriakides, L. (2010). School factors explaining achievement on cognitive and affective outcomes: Establishing a dynamic model of educational effectiveness. *Scandinavian Journal of Educational Research*, 54(1), 263–294. doi:10.1080/00313831003764529
- Cuban, L., & Pea, R. (1998, February). *The Pros and Cons of Technology in the Classroom*. Paper presented at the Funder's Learning Community Meeting, Palo Alto, CA.
- Earle, R. S. (2002, January-February). The Integration of Instructional Technology: Promises and Challenges. *ET Magazine*, 42(1), 5-13. Available from: <http://BooksToRead.com/etp>
- Eguchi, A. (2007). Educational Robotics for Elementary School Classroom. In *Proceedings of Society for Information Technology & Teacher Education International Conference 2007* (pp. 2542-2549). Chesapeake, VA: AACE.
- Eteokleous, N. (2015). *Employing Educational Robotics for the Development of Problem-Based Learning Skills*. Paper presented at the 2015 Annual Meeting of the American Educational Research Association, Chicago, IL.
- Eteokleous, N. (2016). *Developing 21st Century Skills Through an Educational Robotics Curriculum*. Paper presented at the 2016 Annual Meeting of the American Educational Research Association, Washington, DC.
- Eteokleous, N., Demetriou, A. Y., & Stylianou, A. (2013). The Pedagogical Framework for Integrating Robotics as an Interdisciplinary Learning – Cognitive Tool. In J. Roselli & E. Gulick (Eds.), *Information and Communications Technology: New Research* (pp. 141–158). Nova Science Publishers, Inc.
- Gura, M. (2011). *Getting started with Lego Robotics: A guide for K-12 Educators*. ISTE.
- Harel, I., & Papert, S. (1991). *Constructionism*. Ablex.
- Haugland, W. S. (2000). Early Childhood classrooms in the 21st century: Using Computers Maximize Learning. *Young Children*, 12–18.
- Jonassen, D. H. (1999a). Designing Constructivist Learning Environments. In C. Reigeluth (Ed.), *Instructional Design Theories and Models: A new paradigm of Instructional Theory* (pp. 215–239). Mahwah, NJ: Erlbaum.
- Jonassen, D. H. (1999b). *Computer as Mindtools in Schools: Engaging Critical Thinking* (2nd ed.). Columbus, OH: Prentice Hall.
- Jonassen, D. H. (2000). *Computers as mindtools for schools: Engaging critical thinking* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

- Kafai, Y., & Resnick, M. (1996). *Constructionism in Practice: Designing, Thinking, and Learning in a Digital World*. Mahwah, NJ: Lawrence Erlbaum.
- Kassotaki, S., & Roussos, P. (2006). The Greek Computer Self-Efficacy Scale. *Proceedings of the 5th Panhellenic Conference ICT in Education*, 726–733. (in Greek)
- Kazakoff, E., & Bers, M. (2012). Programming in a robotics context in the kindergarten classroom: The impact on sequencing skills. *Journal of Educational Multimedia and Hypermedia*, 21(4), 371–391.
- Kazakoff, E. R., & Bers, M. U. (2014). Put your robot in, Put your robot out: Sequencing through programming robots in early childhood. *Journal of Educational Computing Research*, 50(4), 553–573. doi:10.2190/EC.50.4.f
- Maykut, P., & Morehouse, R. (1994). *Beginning Qualitative Research, A Philosophic and Practical Guide*. London: The Falmer Press.
- Miglino, O., Lund, H. H., & Cardaci, M. (1999). Robotics as an Educational Tool. *Journal of Interactive Learning Research*, 10(1), 25–47.
- Nugent, G., Barker, B., Grandgenett, N., & Adamchuk, V. (2009). The use of digital manipulatives in k-12: Robotics, GPS/GIS and programming. In *Frontiers in education conference, 2009. FIE '09*. 39th IEEE (pp. 1–6, 18–21). IEEE.
- Papert, S. (1980). *Mindstorms: Children, Computers, and Powerful Ideas*. New York, NY: Basic Books.
- Papert, S. (1986). *Constructionism: A New Opportunity for Elementary Science Education*. A MIT Proposal to the National Science Foundation.
- Papert, S. (1993). *Mindstorms: children, computers, and powerful ideas* (2nd ed.). New York, NY: BasicBooks.
- Puntambekar, S., & Kolodner, J. L. (2005). Toward implementing distributed scaffolding: Helping students learn science from design. *Journal of Research in Science Teaching and Teacher Education*, 42(2), 185–217. doi:10.1002/tea.20048
- Resnick, M., Berg, R., & Eisenberg, M. (2000). Beyond black boxes: Bringing transparency and aesthetics back to scientific investigation. *Journal of the Learning Sciences*, 9(1), 7–30. doi:10.1207/s15327809jls0901_3
- Sullivan, A., & Bers, M. U. (2015). Robotics in the early childhood classroom: Learning outcomes from an 8-week robotics curriculum in pre-kindergarten through second grade. *International Journal of Technology*, 26(1), 3–20.
- Sullivan, F. R. (2008). Robotics and science literacy: Thinking skills, science process skills, and systems understanding. *Journal of Research in Science Teaching*, 45(3), 373–394. doi:10.1002/tea.20238
- Sullivan, F. R., & Moriarty, M. A. (2009). Robotics and Discovery Learning: Pedagogical Beliefs, Teacher Practice, and Technology Integration. *Journal of Technology and Teacher Education*, 17(1), 109–142.
- Thomaz, S., Aglae, A., Fernandes, C., Pitta, R., Azevedo, S., & Burlamaqui, A. (2009). RoboEduc: a pedagogical tool to support educational robotics. *Proceedings of ASEE/ IEEE frontiers in education conference*. doi:10.1109/FIE.2009.5350439
- Vollsted, A. M., Robinson, M., & Wang, E. (2007). Using robotics to enhance science, technology, engineering and mathematics curricula. *Proceedings of American Society for Engineering Education Pacific Southwest Annual Conference*.
- Williams, D., Ma, Y., & Prejean, L. (2010). A Preliminary Study Exploring the Use of Fictional Narrative in Robotics Activities. *Journal of Computers in Mathematics and Science Teaching*, 29(1), 51–71.

Williams, D., Ma, Y., Prejean, L., Lai, G., & Ford, M. (2007). Acquisition of physics content knowledge and scientific inquiry skills in a robotics summer camp. *Journal of Research on Technology in Education*, 40(2), 201–216. doi:10.1080/15391523.2007.10782505

Yin, R. (2003). *Case study research: Design and Methods*. Thousand Oaks, CA: Sage Publications.

ADDITIONAL READING

Alimisis, D., Arlegui, J., Fava, N., Frangou, S., Ionita, S., Menegatti, E., & Pina, A. et al. (2010). Introducing robotics to teachers and schools: experiences from the TERECoP project. In J. Clayson & I. Kalas (Eds.), *Proceedings for Constructionism 2010* (pp. 1–13). Paris: American University of Paris. Barker.

Arlegui, J., Menegatti, E., Moro, M., & Pina, A. (2008). Robotics, Computer Science curricula and Interdisciplinary activities, In Proceedings of the TERECoP Workshop “Teaching with robotics, Conference SIMPAR 2008”, Venice.

Barker, B. S., & Ansorge, J. (2007). Robotics as means to increase achievement scores in an informal learning environment. *Journal of Research on Technology in Education*, 39(3), 229–243. doi:10.1080/15391523.2007.10782481

Bishop, A.P., Bertram, B.C., & Lunsford, K.J. & al. (2004). Supporting Community Inquiry with Digital Resources. *Journal of Digital Information*, 5(3).

Blikstein, P. (2013). Digital fabrication and ‘making’ in education: The democratization of invention. In J. WalterHerrmann & C. Bóching (eds.), *FabLabs: Of Machines, Makers and Inventors* (pp. 1–21). Bielefeld: Transcript Publishers.

Detsikas, N., & Alimisis, D. (2011). Status and Trends in Educational Robotics Worldwide with Special Consideration of Educational Experiences from Greek Schools. Proceedings of the International Conference on Informatics in Schools: Situation, Evolution and Perspectives (ISSEP, Comenius University, Bratislava, Slovakia.

diSessa, A. A. (2000). *Changing minds: computers, learning, and literacy*. Cambridge, MA: MIT Press.

Druin, A., & Hendler, J. (2000). *Robots for kids: exploring new technologies for learning experiences*. San Francisco: Morgan Kaufman/Academic Press. Lawhead.

Moro, M., & Alimisis, D. (2009). From the Logo Turtle to the Tiny Robot Turtle: practical and pedagogical issues. Proceedings of the 5th Pan-Hellenic Conference “ICT in Education, Syros (Greece).

KEY TERMS AND DEFINITIONS

BeeBots: A colorful, easy-to-operate, and friendly little robot that can be used by students from the age of four to the age of ten, depending on the level of programming difficulty of each exercise.

Cognitive Skills: The skills developed that are related to specific disciplines. For example, for mathematics discipline, skills to be developed are: sequencing, calculation, multiplication; for physics discipline skills to be developed are: distances, angle. Newton’s law of motion.

Discipline Integration or Interdisciplinarity: It is the integration of various disciplines within the same unit, to examine one concept. The purpose is to relate/combine two or more disciplines into one exercise/ activity in achieving discipline learning objectives.

Educational Robotics: The use of robotics in the teaching and learning process (in the educational practice) as a subject matter and/or as a cognitive-learning tool to achieve disciplinary learning objectives.

Floor Mat: A paper or foam based mat (mainly located on the floor or high-table) which is separated in blocks of 15X15 centimeters. Each floor mat represents a specific theme/ concept and each block depicts different sub-theme/ sub-concept related to the overall theme/ concept. Each floor mat may have from 9 to 24 blocks (or even more!) Floor mats can be developed for various disciplines.

Lego Mindstorms NXT Robotics Package: An educational robotics package appropriate for secondary and higher education level students. The package contains a brick to be connected to the PC either through a wire or through Bluetooth, three motor sensors, touch sensor, ultrasonic sensor, sound sensor and light sensor. Additionally, it contains 431 building blocks, axles, gears, and cams. A visual programming software is also part of the package.

Lego WeDo Robotics Package: An educational robotics package appropriate for elementary

school students (7+ years old). There are two package available: the construction and the resource set. A visual programming software is also part of the package. The construction set includes a hub (to be connected to the PC), 3 different kinds of sensors (tilt, motor and distance sensors), 3 different sizes of axles, gears, cams, 158 colorful building blocks and other essential pieces to build the mini figure: the feet, hair and hut. The resource set includes additional building blocks, axles, gears and cams.

Non-Cognitive Skills: The skills developed that are not related to specific discipline however are considered to be important and necessary for students to develop in survive and succeed in the 21st century. Those are, but not limited to: problem solving, critical thinking, collaboration, communication, global citizenship, digital literacy, knowledge construction, creativity, innovation, self-directed learning.

Robotics Packages: There are various robotics packages that allow users/students to construct and program robots. These packages mainly contain bricks, wires, sensors and a visual programming software.

APPENDIX

R

Table 1.

A/A	Educators	Grade	Disciplines	Duration	Number of students	Educational Robotics Package
1	Teacher 1	6 th	Design & Technology Life Education Environmental Education	8 lessons (40 mins each)	25 students (13 boys & 12 girls)	Lego Mindstorms NXT
2	Teacher 2	6 th	Design and Technology, Safety Education Arts Language and Linguistics	16 lessons (40 mins each)	20 students (10 boys & 10 girls)	Lego Mindstorms NXT
3	Teacher 3	2 nd	Design and Technology Language and Linguistics	11 lessons (40 mins each)	21 students (10 boys & 11 girls)	Lego Mindstorms NXT
4	Teacher 4	4 th	Language and Linguistics Arts	8 lessons (40 mins each)	22 students (10 boys & 12 girls)	Bee Bots
5	Teacher 5	2 nd	Geography Health Education Arts	8 lessons (40 mins each)	16 students, (7 boys & 9 girls)	Bee Bots
6	Teacher 6	5 th grade	Language and Linguistics English	5 lessons (40 mins each)	20 students, (10 boys & 10 girls)	Bee Bots
7	Teacher 7	2 nd	Design & Technology Language and linguistics Arts Health Education	8 lessons (40 mins each)	22 students (9 boys and 13 girls)	Lego WeDo
8	Teacher 8	5 th	Design & Technology Language and linguistics Arts	5 lessons (40 mins each)	23 students (13 boys & 10 girls)	Lego WeDo

State of the Art and Key Design Challenges of Telesurgical Robotics

Sajid Nisar

National University of Science and Technology, Pakistan

Osman Hasan

National University of Science and Technology, Pakistan

INTRODUCTION

Telesurgical robotic systems allow surgeons to perform surgical operations from remote locations. It could be the same room where a surgical operation is being performed or a place somewhere else outside. It could be across two different countries or even cross-continental as demonstrated in “Operation Lindbergh, (Jacques Marescaux, 2002)” where surgeons performed laparoscopic cholecystectomy from New York, USA while the patient was at Strasbourg, France. Similarly, a telesurgery service was established in Ontario, Canada, between a teaching hospital and a rural community hospital located over 400 km away in 2003. It has been nearly twenty years since the emergence of telesurgical robotics in the operating room, but it has been only the last five years or so, that the potential of surgical robotics is being recognized by the surgical community as a whole. Now, it can be said with considerable confidence that robotic surgery has demonstrated numerous advantages over conventional surgery and it has revolutionized the operation theater (OT) as well as the surgical techniques.

Some prominent features of this technology are;

- More degrees-of-freedom for the surgical tool manipulation are available than the conventional one.
- Greater precision (even up to <10 micrometers) than the conventional surgical techniques where it is highly dependent of the

human hand resolution (100 micrometers typically).

- Less fatigue for surgeons, who now sit in an ergonomic console rather than being in an uncomfortable standing posture required by the conventional surgical procedures.
- Enhanced safety and increased patient trust in surgery by making use of comprehensive safety-critical features.

However, these advantages come at the expense of extremely high costs of these complex machines. It is yet to be determined when the benefits will outweigh the cost associated with these (Camarillo, 2004). Staggering capital and operational costs are a great challenge for the designers and engineers to bring down to a minimum possible level.

Moreover, telesurgical robotics is a multidisciplinary field and requires shared understanding among various professionals like medical doctors, engineers and computer scientists. Due to this diversity, there are various design objectives and meeting them altogether is a challenge in itself. A number of telesurgical robots have been developed so far with the objective of optimizing certain metrics and thus each one of these has its own advantages as well as shortcomings. The main scope of this chapter is to identify key design metrics for telesurgical robots and compare some of the notable existing telesurgical robotic systems accordingly. This information is expected to play a vital role while designing the next generation of telesurgical robots.

MAIN FOCUS OF THE ARTICLE

The main focus of this chapter is to provide a comprehensive survey about telesurgical robotics which can be equally beneficent to engineering and as well as medical professionals. After a brief background, a detailed discussion of various state-of-the-art telesurgical systems is provided. Key design approaches and challenges are identified and their solutions are recommended. A set of parameters that can be used to ascertain the usefulness of a telesurgical robot are discussed. These parameters not only allow one to choose the most suitable option among the existing systems but also could serve as guidelines for the development of next generation telesurgical systems. A separate section is dedicated for the future research directions followed by conclusions.

BACKGROUND

The stereotactic brain surgery by using Unimation PUMA 200 robot is considered the first robot assisted surgery (Kwong, 1988). In the beginning, there was a dearth of specialized robots custom-designed for medical surgeries. However, industrial robots had developed fully at that time and they were modified to cater the needs of surgery. Since then, many specialized robots have been designed to match the needs of surgical environments. A successful system requires concerted effort of engineers, medical practitioners, scientists and, to some extent, business professionals as well. To summarize the role and contributions, the medical professionals are the one who generally raise the need or a problem and engineers provide a doable solution. Surgeons act as its evaluators as they happen to be the directed customer-community as well. The business people provide a feasible way to commercialize the technology to make it commercially viable and meaningful. This scheme determines the dynamics of interaction amongst the contributing communities.

STATE-OF-THE-ART

The recent advancements in surgical robotics have led to an enormous usage in practice, making surgical robotics industry worth of more than \$0.5 billion per annum. For example, till 2009, da Vinci surgical system of Intuitive Surgical Inc. for laparoscopy was reportedly installed at more than 1,000 locations worldwide (Alterovitz & Desai, 2009). Moreover, it has been successfully used for thousands of surgical procedures globally. This success has increased the interest of researchers and investors, and therefore, a number of other systems have been built since then. According to one study, the worth of surgical robotic industry has been estimated to reach \$1.5 billion till 2018 (Global Industry Analysts, 2012).

Now, we describe some of the successful and well-researched telesurgical robots while highlighting their key features and capabilities.

Zeus Surgical System

The Automated Endoscopic System for Optimal Positioning (AESOP), a four DOF endoscopic manipulator, was developed by the Computer Motion Inc. AESOP was not a telesurgical system itself but it had been utilized as a building block for many other systems, including Zeus (Holt, 2004), developed by the merger of Computer Motion Inc. and Intuitive Surgical Inc. in 1995. It received the US Food and Drug Administration (FDA) approval in 2001.

Zeus comprised of three robotic arms attached to the OT bed. One robotic arm used the AESOP technology with a 3D high quality digital camera, while the other two were used for the surgical procedures. Zeus was the first telesurgical robot based on master-slave topology. The surgeons used a small master manipulator of 6 DOF to move the robotic arms (Holt, 2004). Zeus system is no more in production (Jyotsna Dwivedi, 2012) and has been superseded by the da Vinci surgical system.

da Vinci Surgical System

The da Vinci surgical system, also developed by Intuitive Surgical Inc., got its Food and Drug Administration (FDA) approval in year 2000 and became the first commercially used surgical robot. It had performed more than 360,000 procedures worldwide in 2011 alone (Blake Hannaford, 2013). With a net worth of approximately \$1.5 million (Yash, 2008), the da Vinci surgical system is one of the most expensive tools used by the surgeons today.

Like other existing telesurgical robotic systems, da Vinci is also based on the master-slave topology. It is mainly composed of a surgeon console, a patient-side cart, EndoWrist surgical instruments and a 3D vision system. Its surgeon console has two master manipulators and an incorporated 3D vision system. It can be used with three or four arms mounted over a trolley offering easy docking. Robotic arms use double-parallellogram mechanism to achieve remote center of motion. Specially designed metallic cables with high fatigue life are used for transmission. Surgical instruments are detachable during operation and need a replacement after a certain number of surgical operations, 10 as mentioned in subsequent portion of this chapter. This makes the running cost of the da Vinci surgical system quite higher and difficult to sustain for small and medium-sized medical organizations. The bigger footprint of system makes it difficult to transport and adjust in OT environment. However, the ergonomic ease and enhanced vision systems of da Vinci has been quite popular among the surgeons. The limited feedback has also been very useful during the medical procedures.

Raven Surgical System

Raven I and II are telesurgical systems developed by the BioRobotics Lab, University of Washington. Raven II is the latest and an improved version of Raven I. It has been presented as an open source system for collaborative research for advances in medical robotics (Blake Hannaford, 2013).

Main components of the Raven II include the patient-side (slave), the surgeon-side (master) and an integrating unit. Raven II manipulators are based on spherical mechanism which is relatively complex to actuate. Raven can be mounted on either side of the surgical site and up to four arms can be used in a single surgical field. It supports 4 DOF and removable surgical instruments during operation. Raven II uses Phantom Omni devices as its master manipulators. However, Phantom Omni are multipurpose devices and are not tailored for medical robots. These have certain limitations such as being not able to retain a particular posture freely, and thus, lacking the ergonomic ease for surgeons.

SOFIE Surgical System

“Surgical robot with force feedback” (SOFIE) is developed by the University of Technology (TU/e) Eindhoven. The system provides surgeons a compact layout with haptic (force) feedback. Based on a master-slave topology, it has specifically designed master manipulators for surgical maneuvers.

The Sofie system is mainly composed of two slave robotic arms mounted on a bed side, two master controllers incorporated with haptic feedback and software control and integration of master and slave. It uses customized master manipulators which provide considerable ergonomic ease. The biggest short coming of this system is bulky tool head. The tools are not detachable during a surgical procedure. The systems makes use of an ordinary vision system as well. Furthermore, to the best of our knowledge, it has not been tested in an OT environment yet.

Mirosurge Surgical System

MiroSurge is developed by the German Aerospace Center DLR for MIS and open surgery. The use of bimanual haptic force and the partial tactile feedback gives operator direct perception of the remote surgical field. It also gives surgeon the

flexibility to control instruments by optically tracked hand-held forceps (Tobergte, et al., 2011). Based on the master-slave topology, MiroSurge has its own master devices and it also has the ability of force reflection. It uses controlled RCM technique along with a robust control system. In case of a failure, robotic manipulators could lose RCM which is highly dangerous. However, it uses force-torque sensors which make it considerably safer during operation.

MiroSurge is mainly composed of two surgical arms with haptic feedback for holding instruments, one surgical arm with two endoscopic High Definition cameras that provide stereoscopic 3D image and a contact free interface as surgeon console. It provides a dedicated left and right handed control comprising of 7 DOF. Currently, MiroSurge uses bed-side mounting, which somewhat hinders its portability. It also requires specially customized OT bed for operation. However, the surgeon console is quite ergonomic and controlled RCM provides the flexibility of easy setup.

Al-Zahrawi Surgical System

Al-Zahrawi surgical robot is intended for MIS and is developed by National University of Sciences and Technology, Pakistan. The system is composed of master manipulator mounted over surgeon console, slave manipulator supported by a ground support trolley, instrument module and a control unit. Based on the planar parallelogram RCM kinematics, the system implements master-slave topology. Uniquely designed modular surgical tool wrist is interchangeable during operation. Furthermore, the distal end of the slave manipulator and tool wrist do not contain any circuitry and, therefore, are fully sterilizable (Hassan, 2014). The overall size and footprint of the system is relatively smaller and makes it suitable for MIS. However, it lacks haptic feedback.

KEY DESIGN APPROACHES AND CHALLENGES

Telesurgical robotics, being a multidisciplinary field, requires a comprehensive approach to design a successful system. Hence, it is imperative to figure out the key design approaches being used so far and their respective challenges. The above discussion about various robotic systems makes it clear that the following key design challenges are of prime importance for a robust telesurgical robotic system.

Manipulator Configuration

Selection of an appropriate configuration of a manipulator mainly depends on its application. For example, for fulfilling the large workspace requirements, as in the case of brain, thorax and abdomen surgery, series structures are preferable (Zhijiang, 2003). Similarly, parallel structures can be used for narrow, long and slender workspaces which may require more power to manipulate the end-effectors, as in the case of minimal invasive surgeries for retina, nose and ear. Configuration also determines the complexity of control and manufacturing. For example, series structures are generally difficult to control as compared to parallel structures. Moreover, actuator mounting and positioning becomes critical for series structures whereas in parallel structure it is relatively easier.

Remote Center of Motion

The remote center of motion (RCM) is a key requirement of minimal invasive techniques. Mechanisms are built such that there is no lateral or side-ways motion at the entry port of surgical tool. This entry port coincides with the RCM of manipulator to reduce the stress over adjacent tissues. A detailed study of 1-DoF remote center of motions is presented by (Guanghua Zong, 2008).

However, RCM in existing systems is primarily achieved using following two basic techniques.

1. Using Control Algorithms

In this technique, the RCM is achieved via control algorithms that govern the manipulator. It gives the flexibility to alter and can be adjusted on more than one points in the manipulator workspace. However, in case of a failure, there is always some risk of manipulator going rogue, which is highly undesirable. DLR Mirosurge makes use of this type of RCM. However, given the flexibility to change RCM position at any time makes system ideal for both open surgery and MIS. Whereas, the mechanism singularities encountered in such type of arrangements often need extensive research to deal with.

2. Using Remote Center of Motion Mechanisms

This technique uses the mechanical design of the manipulator to achieve the RCM. The mechanism is designed in such a way that the end-effector revolves around a single point. This approach is intrinsically safe and has been applied in most of the present telesurgical robotic systems like da Vinci, Sofie, Raven and Al-Zahrawi. But, systems based on such mechanisms usually have bigger motion envelop which has its own downsides. The fixed shape and form of the manipulator makes it impossible to change the position of the remote center of motion without moving the whole mechanism.

Miniaturization

The space and volume constraints always attach compulsion to a surgical system to be as compact as possible. The combined footprint of da Vinci surgical system, comprising of surgeon console, patient-cart and central unit is about 20 square feet as measured by authors at Qatar Hospital, Karachi, Pakistan. A similar observation of 22

square feet has been presented by (Eichel, 2005) as well. This makes the system bulkier and difficult to transport. Bigger size means more time and effort required to setup the system as well.

However, decreasing the overall size, weight and power required to actuate the system usually results in increased design complexity and short life-cycle of various key components. For example, the EndoWrists of da Vinci system are made compact, lighter and smaller which provides small space for cables to turn around. This decreases the fatigue life of cables and limits the number of usage of EndoWrist tool to 10 times at maximum. This, in turn, increases the overall operational cost of the system.

Tool Detachment

In one typical laparoscopic surgery, for example, more than five different surgical tools are used. This makes tool detachment an important requirement during a surgical procedure. The requirement of miniaturization impedes the process of detachment as it increases design complexity manifolds. In currently available systems, da Vinci and Al-Zahrawi surgical system provide tool detachment facility while performing a surgical procedure.

Power Transmission

A suitable power transmission scheme is key for the precision and accuracy of a surgical system. Currently available telesurgical systems mainly imply two types of actuation schemes;

1. Geared Actuation,
2. Cable Actuation.

Sofie and Mirosurge make use of the geared actuation scheme whereas da Vinci, Al-Zahrawi and Raven mainly make use of metallic cables. Gear actuation is considered more efficient and has longer life but the requirement of miniaturization makes it difficult to incorporate. Cable actuation scheme has the advantage of flexibility and ap-

plicability even to the complex designs but it has relatively shorter life. This also requires specially designed medical precision cables for the best performance and longer life. The slip factor is also critical for smooth and longer operation.

Backdrivability

Backdrivability is an important concept which defines the ability of a mechanism to be manually operable in backward direction. In case of a surgical robot, it means the possibility of the robot to be manually extracted out from the surgical site without doing any damage in case of a machine failure. Different kind of kinematic structures have different degree of backdrivability. Generally, manipulators having higher transmission ratios and active prismatic joints are considered non-backdrivable. Few of the existing surgical manipulators have backdrivability available for distal end degrees-of-freedom.

Haptics

Haptics (force and tactile feedback) provides a sense of touch to the remotely located surgeon and has long been a desirable feature in telesurgical robotic systems. Through haptics, the surgeon can feel and respond to a situation whether it's in training mode or operational mode. Early systems like Zeus did not support this feature. Incorporation of haptics makes the master controller design considerably difficult and, hence, requires a great amount of effort to fully evolve out yet. The technology for realistic haptic feedback is being actively researched upon.

Safety and Reliability

Since the birth of surgical robotics, safety and reliability has been the biggest challenge, given the safety-critical nature of applications. The fact that the da Vinci surgical system - only robotic system that made its way to the hospitals - is still under scrutiny by the FDA (Tanner,

2013) reflects the stringent criteria of safety and reliability. The seamless transmission of signals between master and slave robots during a surgical procedure is critical. It has been concluded by M. D. Fabrizio and others that a maximum time delay of 700 milliseconds is acceptable whereas, generally, a time delay of 300 milliseconds is considered completely safe. The reliability of teleoperation network is also crucial for a safe and reliable telesurgical robotic system. Frequent interruptions or failures result in loss of patient trust which needs to be avoided at all costs.

Multidisciplinary Nature

Multidisciplinary nature is a fascinating aspect of telesurgical robotics. However, for a relatively new field, there is a dearth of standard procedures which makes the collaboration among various communities challenging and less productive. Every community including engineering and medical has its own specific jargons, standards and communication protocols. Lack of standardization hinders an accurate assessment of needs and problems belonging to the area of one community by the other. This has been an impediment for many research systems to materialize as a successful commercial technology.

Solutions and Recommendations

Keeping the above discussed challenges in view, we now propose some solutions and recommendations for resolving the critical issues.

For miniaturization, the recent developments in the micro electrical mechanical systems (MEMS) augur a great deal of progress. Micro robots, containing microintegrated devices, are being developed which can work inside specific organs. Microtransmissions have been developed containing microgears as small as of a pollen grain (Camarillo, 2004). Despite of its great potential, use of MEMS in surgical robotics has not drawn full attention yet.

For tool detachment, a modular design approach is preferable. Surgical tools should be designed as a separate integrate-able module rather than a part of the manipulator. Doing so provides the flexibility of tool-change during a procedure. Such kind of technique is used in EndoWrist of the da Vinci surgical system. This technique can further be used to develop submodules comprising of tool's distal end. Such type of tools can be sterilized with a considerable ease and safety. Moreover, backdrivable manipulators can be designed by replacing active prismatic joints by revolute joints and balancing transmission ratios wherever possible.

To improve the haptic feedback, some out of the box solutions are needed as the conventional ways make the system design quite cumbersome. The use of various chemical and fluid sensors can provide some novel solutions. High fidelity force sensors can be used for this purpose as they have more force sensation even than that of the human hand (Camarillo, 2004).

To counter the challenges associated with multidisciplinary nature of the field, there is a serious need of some kind of standardization which can bring all related communities to a single page. This is difficult to achieve on an organizational level. However, efforts can be made to stimulate the surgical robotics community as whole to develop such kind of mechanism. The inclusion of relevant curriculum at school levels of all related communities could be a positive step in this direction.

Finally, the risk factors can be reduced and the robotic surgeries could be made further reliable by using more comprehensive reliability evaluation methods and having safety mechanisms. For example, rigorous analysis techniques like formal methods can be used to analyze the software and hardware components of a telesurgical robot. These techniques provide accurate analysis results at the cost of time and effort but their usage can be easily justified based on the safety-critical nature of telesurgical robots. Moreover, both hardware and software interlocks may be used to

avoid catastrophic complications. Better training mechanisms are also required to equip the surgeons with the skills that are essential to operate these sophisticated machines.

Moreover, the curriculum of most medical schools does not include hands-on experience with robotically assisted procedures these days. Hence, forcing surgeons to acquire these skills during their practice in a cluttered and haphazard manner. The students need to be trained in more innovative and productive ways. For this purpose, the virtual reality simulators can play a pivotal role.

FUTURE RESEARCH DIRECTIONS

About a decade ago, the maximum DOF offered by a telesurgical robotic manipulator were confined to 4-7. Whereas, now research is being carried out to increase this number to a certain level where surgeon can enjoy the liberties of open surgery. A further new trend is about the RCM mechanisms which could offer vital DOFs by virtue of the mechanism itself. These new mechanisms (Gijbels, 2014; Li, 2014, Nisar 2017) provide compact distal end which is highly desirable for MIS applications.

Tactile feedback is important for future surgical robots. The more recently developed da Vinci Si and Sofie system provide surgeons with limited force feedback functionality. Further research is being carried out to make these systems more realistic and user friendly. High speed data transmission and telecommunication protocols are the lynchpin of the future progress in telesurgical robotics. To reduce the time delays and ensure seamless transmission of signals it still requires effort to improve the communication and relevant hardware technologies. Reduced footprint and miniaturization is at the core of the future research efforts. The systems like Raven and Sofie are focused to have smaller motion envelopes.

Low cost surgical systems yet seem a distant dream. This can be established by the fact that, in 2004, the older version of da Vinci surgical system

with 03 arms had a cost of around \$1 million, and additional operational costs were about \$130,000 per annum. Later versions with 04 arms, and the da Vinci S, made available in 2006, approximately costs of \$1.33 million and \$1.53 million, respectively (Justin & Albani, 2007). Being the lone commercial telesurgical system so far, this shows the trend of prices going up with each new version. So, to lower the price and reduce the overall cost of a telesurgical system has great potential to be investigated in future.

Autonomous surgical robotic systems would be the ultimate surgical machines. Autonomous and semiautonomous robots are being envisaged as future robots to replace the surgeon. Autonomous suturing and knot-tying has already been attempted by Endobot (Hyosig. Kang, 2001). Similarly, autonomous knot-tying is tried by another robot called EndoPAR by (Hermann Mayer, 2004) as well. However, these ventures are just confined to suturing and knot-tying. Therefore, development of such surgical robots for conducting full surgeries still has a long way to go.

Use of Artificial Intelligence (AI) and 3D visualization techniques is another hot bed of research in surgical robotics. Neuromorphic engineering, genetic algorithms, and artificial evolution are some of the future directions of research regarding this aspect (Camarillo, 2004). 3D visualization of surgical field increases depth perception capabilities of surgeons while performing operations. 3D mapping, virtual construction of the surgical field can help surgeons to preplan and even rehearse their surgeries. This can greatly reduce chances of any mistakes or errors.

CONCLUSION

Last two decades have seen a phenomenal growth in telesurgical robotics and, therefore, brought a paradigm shift in the field of surgery. Now surgeons using robots have access to enhanced dexterity, control, precision, 3D vision systems,

and image and motion scaling during the surgeries while sitting back in an ergonomic console. Moreover, the robotic hands do not shake and there are less chances of errors. Smaller incisions, less blood loss and faster recovery promise telesurgical robotics a bright future. These robots are also paving paths for many interesting areas of research in medical surgery. For example, surgical robots have been used to do many successful kidney transplants of obese patients, which was classified as a very risky procedure earlier.

Despite enormous advantages, the usage of telesurgical robots in the OT is quite limited. The major reasons include high costs and the tough safety and reliability constraints. In order to develop more cost-effective and reliable systems, cooperation between engineers, computer scientists and surgeons is vital. This chapter provided a brief overview of the existing telesurgical robotic systems and the associated challenges and key design approaches in a non-technical language in order to provide a common ground for healthy discussions among the diverse communities. We identified the safety and reliability issues as the main bottle neck in this area and have proposed some methods for improving the state-of-the-art. We observe that surgeons' training need to be more intuitive. In terms of technological advancements, we see an increased incorporation of haptics and the introduction of more DOF in the next generation of telesurgical systems. Exploration of autonomous and semi-autonomous systems are the prime future research initiatives. The challenges of accuracy, precision and miniaturization are being confronted with great zeal whereas the issues related to reliability, cost and telecommunication still need to be considered.

REFERENCES

Alterovitz, R., & Desai, J. P. (2009, June). Surgical Robotics. *IEEE Robotics & Automation Magazine*, 16(2), 16–17. doi:10.1109/MRA.2009.932616

- Blake Hannaford, J. R. (2013, April). Raven-II: An Open Platform for Surgical Robotics Research. *IEEE Transactions on Bio-Medical Engineering*, 60(4), 954–959. doi:10.1109/TBME.2012.2228858
- Camarillo, D. B., Krummel, T. M., & Salisbury, J. K. (2004, October). Robotic technology in surgery: Past, present, and future. *American Journal of Surgery*, 188(4), 2–15. doi:10.1016/j.amjsurg.2004.08.025
- Eichel, L., Ahlering, T. E., & Clayman, R. V. (2005). Robotics in Urologic Surgery: Risks and Benefits. *AUA Update Series*, 24(13).
- Gijbels, A. N. (2014). Design and Realisation of a Novel Robotic Manipulator for Retinal Surgery. *IEEE/RSJ International Conference on* (pp. 3598–3603). Tokyo: IEEE.
- Global Industry Analysts. (2012). *Medical Robotics: A Global Strategic Business Report*. Global Industry Analysts.
- Guanghua Zong, X. P. (2008). Classification and type synthesis of 1-DOF remote center of motion mechanisms. *Mechanism and Machine Theory*, 43(12), 1585–1595. doi:10.1016/j.mechmach-theory.2007.12.008
- Hassan, T., & Hasan, O. (2014). Al-Zahrawi: A Telesurgical Robotic System for Minimal Invasive Surgery. *IEEE Systems Journal*.
- Hermann Mayer, I. N. (2004). The Endo[PA]R System for Minimally Invasive Robotic Surgery. *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 3637–3642.
- Holt, D. E. (2004). Telesurgery: Advances and Trends. *University of Toronto Medical Journal*, 82, 3.
- Jacques Marescaux, M. J. (2002). Transcontinental Robot-Assisted Remote Telesurgery: Feasibility and Potential Applications. *Annals of Surgery*, 235(4), 487–492. doi:10.1097/00000658-200204000-00005
- Justin, M., & Albani, M. (2007). Mach/April). The Role of Robotics in Surgery: A Review. *Science & Technology Review*, 104(2), 169.
- Jyotsna Dwivedi, I. M. (2012). Robotic Surgery - A Review on Recent advances in Robotic Surgical Systems. *Florida Conference on Recent Advances in Robotics*, 1-7.
- Kang, H., & Wen, J. T. (2001). EndoBot: a Robotic Assistant in Minimally Invasive Surgeries. In *Proceedings of the 2001 IEEE International Conference on Robotics & Automation* (pp. 2031-2036). Seoul: IEEE. doi:10.1109/ROBOT.2001.932906
- Kwong, Y. S., Hou, J., Jonckheere, E. A., & Hayati, S. (1988). A robot with improved absolute positioning accuracy for CT-guided stereotactic brain surgery. *IEEE Transactions on Bio-Medical Engineering*, 35(2), 153–160. doi:10.1109/10.1354
- Li, J. G. (2014). A Class of 2-Degree-of-Freedom Planar Remote Center-of-Motion Mechanisms Based on Virtual Parallelograms. *ASME Journal of Mechanisms and Robotics*.
- Nisar, S., Endo, T., & Matsuno, F. (2017). Design and Kinematic Optimization of a Two Degrees-of-Freedom Planar Remote Center of Motion Mechanism for Minimally Invasive Surgery Manipulators. *ASME. J. Mechanisms Robotics*, 9(3), 031013–031013-9. doi:10.1115/1.4035991
- Satava, R. M. (2002). Surgical Robotics: The Early Chronicles: A Personal Historical Perspective. *Surgical Laparoscopy, Endoscopy & Percutaneous Techniques*, 12(1), 6–16. doi:10.1097/00129689-200202000-00002
- Tanner, L. (2013, April 9). *FDA Investigating Potential Problems with Popular Surgical Robot*. Retrieved April 10, 2013, from <http://healthland.time.com/2013/04/09/fda-investigating-potential-problems-with-popular-surgical-robot/>
- Tobertge, A. P., Hagn, U., Rouiller, P., Thielmann, S., Grange, S., & Albu-Schäffer, A. (2011). The sigma.7 haptic interface for MiroSurge: A new bi-manual surgical console. *Proceedings of IROS*, 3023-3030. doi:10.1109/IROS.2011.6094433

Yash, S. (2008, May). *The DaVinci Robot: Reducing Recovery Time and Medical Costs in the Operating Room*. Retrieved on April 26, 2013, from <http://www.jyi.org/issue/the-davincireg-robot-reducing-recovery-time-and-medical-costs-in-the-operating-room/>

Zhijiang, D., & Lining, S. (2003). Review of surgical robotics and key technologies analysis. *Proceeding of the 2003 IEEE International Conference on Robotics, Intelligent Systems and Signal Processing*, 1041-1046. doi:10.1109/RIISP.2003.1285733

ADDITIONAL READING

Faust, R. A. (2007). *Robotics in Surgery: History, Current and Future Applications*. New York: Nova Science Publishers, Inc.

Manuel Ferre, M. B. (2007). *Advances in Telerobotics*. Springer-Verlag Berlin Heidelberg; doi:10.1007/978-3-540-71364-7

KEY TERMS AND DEFINITIONS

Motion Envelope: A motion envelope is a virtual volume representing all positions which a mechanism, and its linkages, achieve during its complete range of motion.

Remote Center of Motion (RCM): A remote fixed point, with no physical revolute joint over there, around which a mechanism or part of it can rotate is called remote center of motion.

Workspace: Workspace or reachable space represents all those points which can be reached by the end effector of a robot.

Telesurgical Robotics and a Kinematic Perspective

Sajid Nisar

National University of Sciences and Technology, Pakistan

Osman Hasan

National University of Sciences and Technology, Pakistan

INTRODUCTION

The advent of Minimally Invasive Surgery (MIS) in surgical theaters has revolutionized the centuries-old art of surgery for various surgical procedures. After rapidly gaining its momentum in 1990s, MIS offered huge benefits over the traditional open-surgery techniques. Details of which can be found in (Satava, 2004; Lanfranco, 2004; Holt, 2004; Taylor, 2006; Kuo C. H., 2009; Nisar, 2015). Along with its revolutionary advantages, MIS proved to be a tedious and cumbersome technique for surgeons. For safety and better efficacy, it requires well-trained and skillful surgeons to carry out the surgical procedures. When compared with open-surgery, MIS techniques adversely affect surgeons' ability of direct-sight, sense-of-touch, depth perception and kinesthetic feel. These issues, along with numerous others pose serious challenges to medical doctors and surgeons when performing MIS procedures.

Surgical robotics, primarily, emerged as a very effective solution to the above-mentioned problems and many other related issues. With the passage of time, these robots started taking the role of augmenters and led to the emergence of many innovative, safer, better and user-friendly features pertaining to various MIS procedures. This gave huge impetus to the ongoing research in this field and grabbed a lot of attention from researchers of many mutually varying fields of research. While industry has played an important role, it had been the academia that remained at the forefront of

the research in surgical robotics. Being in operation theaters for roughly more than two decades, surgical robots have demonstrated their potential towards the betterment of surgical procedures and general degree of efficacy (Taylor, 2006; Nisar, 2015). Enhanced precision, greater control over maneuvers, scalable movements and tremor-free tool motion with tactile feedback are some of the sublime features of today's surgical robots. Increased trust of surgeons, medical practitioners (Chitwood, 2001) and, even of the patients, on the use of robots for medical and surgical purposes is becoming a hallmark of this technology. With every new development, the overall size and weight of surgical robots is shrinking while the reliability and safety margins are expanding.

Behind every successful surgical robot, a number of technical and non-technical factors play an important role. For example, kinematics, dynamics, control and manufacturability are some of the technical aspects of a design. Kinematics - being the first and foremost manipulator design step - is considered as the lynchpin of performance for any surgical robot. It is a fundamental aspect of any mechanical design (Kuo C. H., 2012) and plays a decisive role in ascertaining its capabilities and viability vis-à-vis applications. A sound kinematic design is better posed to offer greater performance measures in terms of safety, reliability and surgical task-achievement (Kuo C. H., 2009). Kinematic design acts as a foundation block for the rest of surgical robotic system to be built over. A dismal kinematic design could severely limit the capa-

bilities of a robot for advanced features and vice versa. Therefore, it is important to understand the kinematic design approaches in practice so far and discuss their features and potential shortcomings.

A number of well-researched and peer-reviewed surgical robots in academia and industry for various minimally invasive surgical procedures have been proposed and many novel designs are being continuously investigated. Thus a survey of the kinematic aspects of these existing designs, giving an all-inclusive consideration, can play a vital role in understanding their pros and cons and thus improving performance of the new designs.

BACKGROUND

Most of the existing surgical robotic systems are telesurgical in nature, where the surgeon operates robotic tools remotely (Nisar, 2015). This remote location could be the same room as that of the patient or anywhere outside. Given the extensive focus on telesurgical systems, the terms ‘surgical robotics’ and ‘telesurgical robotics’ are sometimes implied in the latter’s sense despite of a certain distinction between the two.

A typical telesurgical robotic system constitutes of various subsystems, like master manipulator, slave manipulator, surgical tool with wrist, control unit and vision system. A system level explanation of a particular surgical robot is described at length in (Hassan, 2014). One key member, and a more interesting one from the robotics perspective, is the slave manipulator. Every manipulator is built upon some mechanism, which embodies the concept and ingenuity of its designer to achieve some specific task, for example, surgical tool movement in the case of surgery. This ability of a mechanism to achieve the design-objective in terms of links lengths and joint angles is mainly defined by the robot kinematics. Kinematics, being a study of the motion geometry of a mechanism and its bodies, provides understanding about the position of its members and change in their positions with respect to time

as the mechanism changes its posture. Kinematic design plays a fundamental role and is a crucial factor in terms of its ability to achieve a given task and the overall design robustness.

A key approach to describe telesurgical robots is to classify them based on their kinematic designs. As described earlier, kinematic design defines the basic shape and function of a mechanism and dictates the terms to achieve the design goal such as surgical movements. Therefore, it is important to consider the kinematic perspective of the existing surgical robot designs in order to get a broader and clear picture of the overall progress in the field.

Historically, industrial manipulators were the first to be modified and used for surgical applications (Satava, 2004). Later on, special designs and dedicated manipulators were proposed and implemented for surgical purposes. Most of the surgical applications and majority of the surgical robots are intended for the MIS procedures (Taylor, 2006; Kuo C. H., 2009; Nisar, 2015). Therefore, before examining the available kinematic designs, it is important to understand the operational concept of MIS and its kinematic requirements.

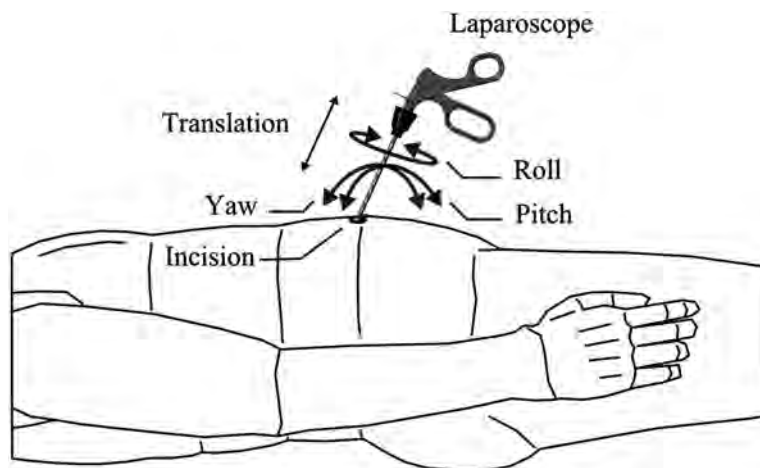
Concept of MIS

Minimally Invasive Surgery (MIS) involves long and slender surgical tools called laparoscopes. Surgeons insert these tools inside the patient body and get visual feedback from one of the tools having an endoscopic camera at the distal end. Generally, three or more such tools are inserted inside the patient body through small holes called as incisions (Howe, 1999). A device, called trocar is used to provide a safer passage to operate the surgical tools at the incision point without causing any damage to the neighboring tissues.

DoFs Requirement

A general MIS procedure requires 4 degrees-of-freedom (DoF) for tool manipulation at the point of incision. The open-close and tool tip movements are generally obtained through dedicated distal end

Figure 1. Typical tool setup in MIS



wrists and, thereby, do not play any significant role in the manipulator (slave) kinematics. The four DoFs required, referred here as primary MIS DoFs, are shown in Figure 1 (labeled as pitch, yaw, roll and translation).

Remote-Center-of-Motion

Figure 1 also demonstrates that the surgical tool is required to be virtually hinged around the incision point. It can translate inwards/outwards through the incision and can rotate along its center. This point is usually called as Remote-Center-of-Motion (RCM). Being a fundamental require-

ment for telesurgical robots, intended for MIS, it is defined as a point around which a mechanism or a part of a mechanism rotates around (Taylor, 2006; Nisar, 2015). The mechanism that achieves RCM is generally called an RCM mechanism (shown in Figure 2). When actuated for the pitch and yaw DoFs, its distal end remains virtually hinged around and rotates along the RCM point.

Workspace Requirements

It has been concluded through experimental observations (Lum M. J., 2006) that during a typical MIS process, 95 per cent motion of the surgical

Figure 2. A general remote center of motion mechanism

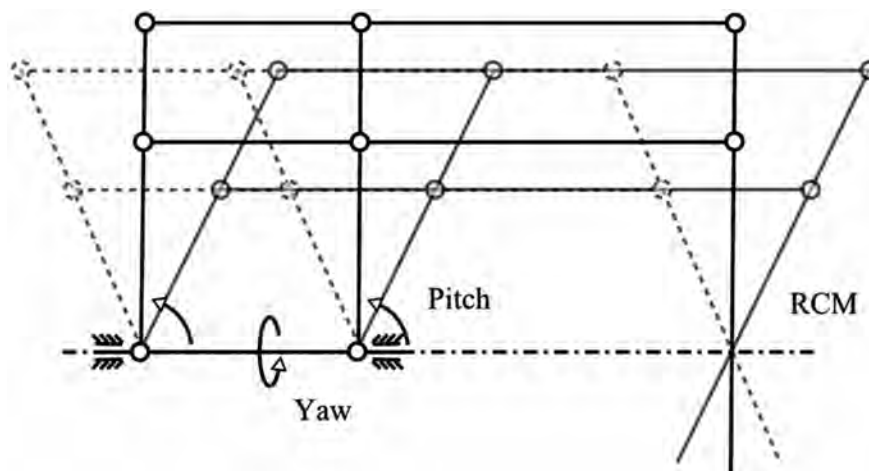
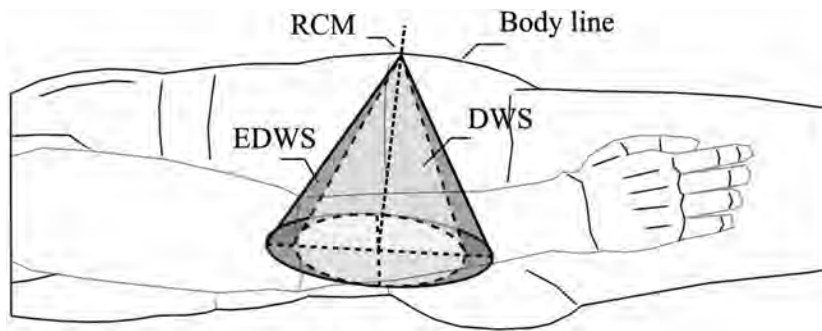


Figure 3. Workspace requirements for MIS



tool is confined in a 60 degree conic region, named as Dexterous Workspace (DWS). The remaining 5 per cent tool movements are contained in another conic region called as Extended Dexterous Workspace (EDWS), which is defined by the intersection of a 90 degrees cone with elliptical base and DWS, as shown in Figure 3. Dexterous Workspace and Extended Dexterous Workspace together define the required workspace for MIS (Lum M. J., 2006).

MAIN FOCUS OF THE ARTICLE

The main focus of this article is to present a general survey of the present day telesurgical robotics technology from a perspective of their kinematics. A number of prevalent kinematic structures in the state-of-the-art telesurgical robots are discussed and the factors shaping their kinematic design are analyzed. Based on that, advantages and potential shortcomings of these kinematic structures are highlighted which could be instrumental for engineers and designers to have a comprehensive understanding of the existing technology. Key design issues and challenges are summarized. To make future surgical robots more effective and useful, possible solutions and recommendations are made in a separate section. The last section gives a discussion about the future trends and possible ways to proceed forward for the telesurgical robotics in near future. The chapter deals with robotics for telesurgical applications, in particular. However,

the discussion and findings in many cases could be applicable to general surgical robotics as well.

PREVALENT KINEMATIC STRUCTURES

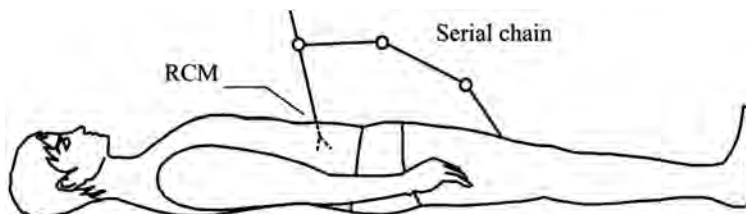
Detailed description of several notable surgical robotic systems, such as Zeus surgical robot, da Vinci surgical system, Raven, Sofie, DLR Miro-Surge and others, can be found in (Kuo C. H., 2009; Nisar, 2015). Here, we describe and discuss the kinematic structures which have been commonly used in existing surgical robotics research including the above systems.

Planar

Planar mechanisms have been one of the most widely used kinematic structures in surgical robotics (Li J. G., 2014). Given their advantages of planar motion, it is easier to place three or more manipulators around the surgical site and operate them simultaneously. Their motion envelop is generally compact (Nisar, 2015) and offer the flexibility of using redundant links to increase the mechanism stiffness. Such mechanisms generally offer higher safety features and better suitability for surgical purposes.

Parallelogram-based RCM mechanisms (Figure 2) are a typical example of planar RCM kinematic structures. Initially proposed by (Taylor et al., 1998) and later on developed in many

Figure 4. Depiction of a serial structure in MIS setup



different forms (U.S Patent No. 5,807,377, 1998; Bedem, 2010; Hassan, 2014), parallelogram mechanism has been the most popular choice in planar RCM mechanisms. The mechanism offers one DoF motion (pitch) and the other DoF (yaw) is obtained by tilting the whole mechanism along its base axis, as shown in Figure 2. The problem with these mechanisms is their limited number of DoFs offered by the mechanism itself. To achieve the translation and roll DoFs, external means are employed, which can have adverse effects on the performance of the manipulator (Gijbels, 2014; Li J. G., 2014, Nisar, 2017). A unique feature of such mechanisms is that they have a number of variants, which have the same function but are geometrically different. Thus, giving the flexibility of implementation in many different surgical application scenarios. Gear drives, cable-pulley schemes and synchronous belt-drives can be used to transmit power while keeping the actuators at the base of the mechanism. da Vinci surgical robot (Intuitive Surgical, 2015), Sofie surgical system (Bedem, 2010) and al-Zahrawi surgical manipulator (Hassan, 2014) make use of planar kinematic structures to achieve RCM.

Serial

Another popular kinematic structure found in surgical robots is of serial type. In a serial chain, tail of one link is connected with the head of the previous link and so on. Such kinematic chains are easier to design and solve for the mechanism singularities. Conventional algebraic and geometric approaches are often sufficient to solve the kinematics of these mechanisms. Serial chains

are usually fully-backdrivable which is a highly desirable feature for a surgical manipulator. The end link in these kinematic structures usually acts as the wrist of the surgical tool while the serial-chain is used to achieve the primary DoFs required for MIS. Lower joint stiffness and higher rate of error propagation are two typical problems encountered with serial chains. Moreover, serial structures need a careful consideration to design against gravitational instabilities. DLR Mirosurge (Hagn, 2010) and Raven (Lum M. J., 2009) are comprised of this kind of structures.

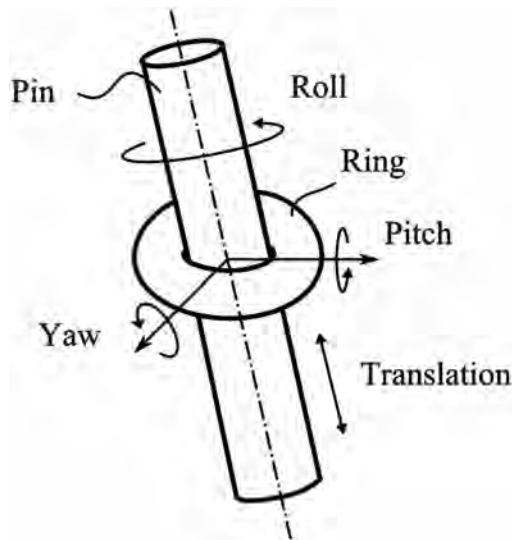
IsoCenters

Isocentric kinematic structures offer a more natural solution to the MIS needs. They consist of a pin-in-ring mechanism; having a ball and cylindrical joint at the center of the ring. The main idea behind such a mechanism is illustrated in Figure 5. The ball and cylindrical joint together offer the four primary DoFs required for MIS. Isocenters have been used in the Zeus Surgical Robot to achieve RCM (Kuo C. H., 2009). However, the mechanism could not gain much interest and is getting less and less attention in the recent years. One reason for this is the physical constraint at the incision point, which makes it difficult to coincide the RCM with the incision point. Also, the implementation of the ball and cylindrical joint at the same point is difficult to realize in a physical system.

Spherical

Spherical kinematic structures make use of spherical/circular arcs to achieve the pivoted motion at

Figure 5. Pin-in-ring arrangement in Isocenters

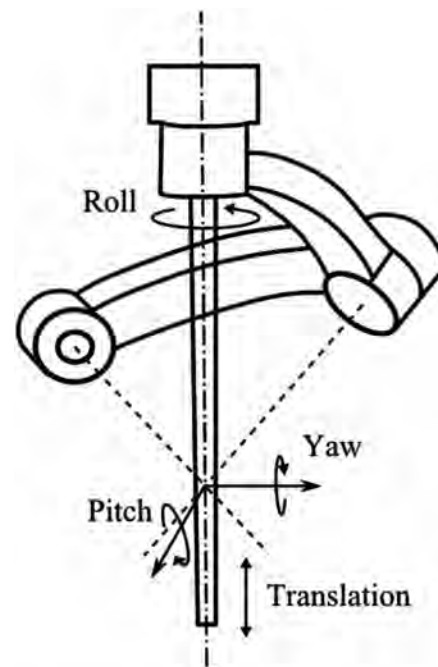


RCM. Like the planar structures, these are also an extension of serial chains but are spatial in nature. The circular members are designed in such a way that together they constitute a virtual sphere and the center of the sphere acts as the RCM point. The mechanism offers pitch and yaw DoFs implicitly while the translation and roll DoFs are achieved through external means. Among the notable systems, the Raven surgical robot (Rosen, 2010) uses this kind of a kinematic structure. Though easier to model and more intuitive for RCM applications, the problem with such mechanisms is their requirement of high-degree machining accuracy during fabrication. This makes them difficult to realize using conventional means, which adds to the overall cost of a surgical system. Figure 6 shows a spherical kinematic structure in an MIS application.

Continuum

Flexible manipulators have recently shown a great potential for surgical applications. In particular, they are being widely advocated for Single Access Port (SAP) surgeries. However, these snake-like structures are difficult to model and control. The computational cost is often higher for such mechanisms in practical situations. The repeatability and

Figure 6. Spherical kinematic structure with MIS DoFs

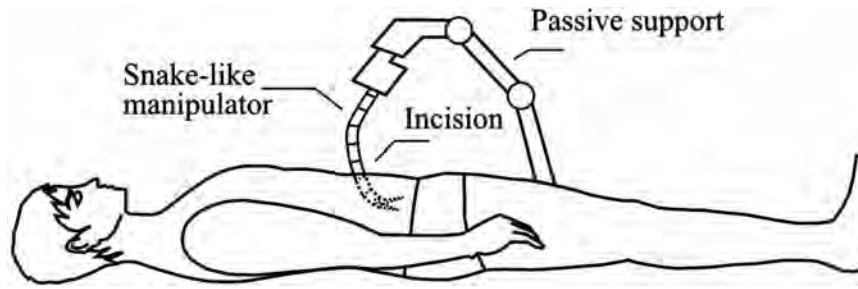


precision of such structures is difficult to maintain which also has been a major bottleneck of their practical utilization in MIS. Various methods and techniques exist to achieve snake-like motions. For example, use of cable-and-pulley and tendon-driven disks have been in practice for some time. More recently, memory-shaped alloys are being used to achieve the surgical DoFs needed for the MIS. Though not fully mature yet, flexible continuum manipulators have the potential to further revolutionize the field of surgery. Figure 7 shows a concept diagram of a flexible snake-like manipulator in an MIS setup.

Parallel/Spatial

Though common in many industrial applications, parallel structures are rarely found in surgical robotics. Generally consisting of spatial platforms, they achieve RCM by using one or more links moving in a controlled fashion. The surgical tool is usually mounted on the platform and the pitch, yaw and translation DoFs are achieved by virtue

Figure 7. A snake-like continuum structure in MIS setup



of the mechanism. Such mechanisms are usually easier to build but are difficult to model and control. Their kinematic analysis is relatively difficult and, in some cases, becomes computationally too expensive. Moreover, due to their spatial nature, they are not ideal for surgical setups where multiple manipulators are required to perform the surgical tasks in a confined space.

FACTORS SHAPING THE KINEMATICS

In this section, we discuss various key factors that affect the kinematic design of a mechanism:

1. Intended applications
 - a. MIS, being the major application, has greatly influenced the existing kinematic designs. The incision-point and conic-shape constraints of the required workspace have given rise to a whole family of remote center of motion (RCM) mechanisms.
2. Mode of operation
 - a. Systems intended for tele-operation mostly employ the master-slave topology. For easier mapping solutions, similar kinematic structures are used for both manipulators. Whereas, in the case of autonomous or semi-autonomous robots, slave manipulator kinematics provide the flexibility to be diverse and thus further extend its functionality.
3. Operational role
 - a. The robots that are designed for assistance (e.g. camera navigation, inspection) usually have simple kinematic design pertaining to the relatively simpler job at hand. For interventional manipulators, the structure is supposed to have extra DoFs due to the complicated nature of intended operations.
4. Operating environment
 - a. Indoor/outdoor: Indoor operations generally require a more compact motion envelop compared to the outdoor operations. On the contrary, outdoor operations require lighter and portable kinematic structures.
 - b. Underwater/space: Though only a few robots (Charles, 2009; Haidegger, 2008) are intended for such extreme environments, but the factors affecting kinematic design greatly differ in such applications. For example, underwater robots need to withstand extra torque and loads, thereby, requiring kinematics to substantiate the manipulator performance on these aspects.
5. Haptics
 - a. Haptics is becoming a necessity these days and many modern kinematic designs are supporting this feature. Incorporation of haptics influences

the overall size and imposes space constraints.

6. Control and Sensing
 - a. Being central to the robot performance, simple kinematics is preferred for easier control. Complex kinematic structures are usually less spacious to accommodate force, torque and many other kinds of sensors.

ISSUES AND DESIGN CHALLENGES

After highlighting the critical factors affecting the overall kinematic design, it is important to identify key issues and design challenges for surgical robots in this regard.

Mechanism Singularities

Mechanism singularities pose a serious bottleneck for many of the mechanism discussed above. Singularities represent configurations of the mechanism where it loses one or more of its DoFs. Mathematically, it is the configuration of a manipulator where its Jacobean matrix, which maps the joint velocities to the end-effector velocities, loses its rank. The mechanism renders itself useless for that particular DOF(s) in such a scenario and it becomes important to avoid any singularities falling within the workspace of the mechanism. A better kinematic design should be less prone to the kinematic singularities and, thus, more suitable for its surgical utility.

Novel RCM Mechanism

In formative years of surgical robotics, a number of new fundamental RCM mechanisms were proposed and many of these proved to be very successful for many surgical applications. However, recently the focus is shifting from innovative mechanical designs to robust control strategies.

Many researchers are trying to introduce tertiary features such as haptics and 3D-vision, and investigations for new RCM mechanisms are becoming very rare nowadays. Most of the kinematic structures used in recently reported surgical robots are based on older concepts and ideas. Surgical robotics have moved much further and we believe that new RCM mechanisms which could augment the shortcomings of the existing ones need to be designed and developed.

Haptics Incorporation

Most of the existing mechanisms and structures do not offer flexibility to incorporate extra features to achieve tactile or kinesthetic feedback. The surgical robots that provide limited forced feedback, like da Vinci and Sofie, are relatively bigger in size and have a larger footprint. One reason for this could be the fact that these kinematic designs were proposed when haptics was not considered as a prime requirement for a surgical system. With surgical technology getting mature and making in-roads towards difficult surgical procedures, haptics is becoming an essential part of a surgical robot now.

Realizability

The ability of a kinematic design to be manufacturable with conventional fabrication resources is a key factor for its success in the surgical market. Many mechanisms, such as spherical and isocentered approaches, are found to be achieving the surgical design objectives but are very difficult to realize as a physical system. Given the need of machining accuracy and fabrication precision, their physical development is costlier, and makes the overall system unaffordable for the health-care organizations. Easier realizability of planar parallelogram-based kinematic structures is the main reason behind their wide application range and high popularity.

Non-Modular Designs

Almost all of the existing kinematic designs are non-modular in nature. Modularity has particular advantages when it comes to surgical applications. It helps to make certain components of a manipulator sterilizable and, thus, reusable. Non-modular designs are usually expensive to maintain and inflexible to make modifications. In the surgical context, a modular design could offer the ease of change of surgical tools or faulty components and could make the robotic system useful and operational for longer periods of time.

Backdrivability

Backdrivability is the ability of a mechanism to be manually operable in backward direction. This is of prime importance when it comes to surgical manipulators. Given their safety-critical application, backdrivable mechanisms can be safely retracted out of the patient body in case of a machine failure without doing any damage to the neighboring tissues. Moreover, most of the revolute joints are usually backdrivable whereas active prismatic joints are considered as non-backdrivable. Many existing kinematic structures make use of active prismatic joints or cable-pulley schemes which have adverse effects on their overall backdrivability.

SOLUTIONS AND RECOMMENDATIONS

The typical planar RCM mechanisms, like the parallelogram-based ones, provide pitch DoF by virtue of the mechanism itself. The yaw DoF is easier to obtain as well. However, the remaining DoFs (tool translation and roll) have to be achieved through external means, such as cable-pulley schemes or actuators mounted on the distal end of the manipulator. This causes many other disadvantages, like increased inertia and vibrations at the distal end of the manipulator. To overcome

these issues, it is important to investigate new planar RCM mechanisms which could offer the translation DoF without using any external means. Recently, some researchers have proposed new mechanisms (Gijbels, 2014; Li J. G., 2014; Li J. Y., 2015, Nisar, 2017) on similar lines. Moreover, singularities can be avoided by altering the kinematics of the mechanism by using efficient and novel design and analysis techniques. Innovative control methods can be implemented to achieve stable and repeatable surgical movements.

Cable-driven systems have been the norm for flexible manipulators so far. Using more fatigue-resistant cables and frictionless joints can increase their utility and life span. Memory-shaped alloys could offer a better solution for the snake-like manipulators where relatively smaller movements are required. Recently, soft robotics has shown some progress in this regard and the results are also applicable to surgical robotics. For haptics, kinesthetic feedback could be given preference over tactile feedback. It has relatively small impact over the size and complexity of a design. Moreover, using compact and miniaturized sensors could further help in this regard.

Modular designs offer great benefits like sterilizable and reusable components. It could help to increase the product life cycle and reduce the operational cost of a surgical manipulator. On the other hand, it could introduce some issues like increased design complexity as well. However, it has the potential to outweigh its shortcomings. Furthermore, to enhance the backdrivability of a kinematic structure it is preferable to replace the active prismatic joints with revolute joints wherever possible. New mechanisms offering pitch, yaw and translation DoFs by virtue of the mechanism itself could be an amicable solution to the backdrivability problem.

In order to test and implement new designs, the use of rapid prototyping technologies and other novel manufacturing techniques could be instrumental. Kinematically robust mechanisms, but otherwise difficult to fabricate, could be realized with significant ease using such manufacturing

techniques. Such manufacturing methods could also be used to fabricate miniaturized components of the complex transmissions and, therefore, reduce the overall size of new robotic systems.

FUTURE RESEARCH DIRECTIONS

Recently, RCM mechanisms with additional degrees-of-freedom are getting an enormous attention. For example, traditional parallelogram mechanism offered the pitch DoF only. The new parallelogram based mechanisms offer pitch and translation DoFs inherently. This in turn helps to counter many adverse effects like higher inertia and energy requirements. Flexible and snake-like continuum kinematic structures are exhibiting great potential for the future surgical applications. Some other interesting directions of research in surgical robots include the investigation of methods for achieving flexible motion and the use of novel materials, like memory-shape alloys. Soft robotics is also a relatively new approach to design new kinematic structures which could bridge the gap between the traditional machine environment and human-machine interaction.

A more concerted effort is needed in order to achieve the primary MIS DoFs through manipulator mechanism without using any external means. Simple and easier to fabricate/assemble structures are highly preferable as well. Given the need of sterilization, future research needs to target kinematic structures with modular design.

CONCLUSION

This chapter presented an overview of the research in surgical robotics from perspective of kinematics. Various prevalent kinematic structures are described and their advantages and disadvantages are highlighted. It is evident from the discussion that simple and robust kinematic structures have shown more success than their complex and larger

counterparts. The factors affecting kinematics of a design have a significant impact on the design choices where most of the surgical robots either go for a spherical-serial kinematic structure or a parallelogram-based planar structure. Among the many available structures, parallelogram-based kinematic structures have been the most widely used ones and offer better prospects for MIS applications. Given the MIS requirements, most of the robots generate enough DoFs but not all of these DoFs are obtained through the slave kinematics itself. Almost every kinematic design needs to employ external means in order to fully achieve the primary MIS DoFs. Moreover, among the existing kinematic structures the ability of a mechanism to incorporate haptics varies largely. Future surgical robotic systems need to be compact, portable and more versatile in their modes of operation. Thus, we believe that there is a great need for new and innovative kinematic RCM structures which could offer all DoFs inclusively and have better support for haptics, modularity and miniaturization.

REFERENCES

- Bedem, L. J. (2010). *Realization of a demonstrator slave for robotic minimally invasive surgery*. Eindhoven: Technische Universiteit Eindhoven.
- Charles, R. D. (2009). Evaluation of Teleoperated Surgical Robots in an Enclosed Undersea Environment. *Telemedicine Journal and e-Health*, 325–335. PMID:19441950
- Chitwood, J. W. Jr, Nifong, L. W., Chapman, W. H. H., Felger, J. E., Bailey, B. M., Ballint, T., & Albrecht, R. A. et al. (2001). Robotic surgical training in an academic institution. *Annals of Surgery*, 234(4), 475–486. doi:10.1097/00000658-200110000-00007 PMID:11573041
- Gijbels, A. N. (2014). Design and Realisation of a Novel Robotic Manipulator for Retinal Surgery. *IEEE/RSJ International Conference on* (pp. 3598–3603). Tokyo: IEEE.

- Hagn, U. K., Konietschke, R., Tobergte, A., Nickl, M., Jörg, S., Kübler, B., & Hirzinger, G. et al. (2010). DLR MiroSurge: A versatile system for research in endoscopic telesurgery. *International Journal of Computer Assisted Radiology and Surgery*, 5(2), 183–193. doi:10.1007/s11548-009-0372-4 PMID:20033517
- Haidegger, T., & Benyo, Z. (2008). Surgical robotic support for long duration space missions. *Acta Astronautica*, 63(7-10), 996–1005. doi:10.1016/j.actaastro.2008.01.005
- Hassan, T. A. (2014). *Al-Zahrawi: A Telesurgical Robotic System for Minimal Invasive Surgery*. *IEEE Systems Journal*. doi:10.1109/JSYST.2014.2331146
- Holt, D. A. (2004). Telesurgery: Advances and Trends. *University of Toronto Medical Journal*, 82(1), 51–54.
- Howe, R. D., & Matsuoka, Y. (1999). Robotics for Surgery. *Annual Review of Biomedical Engineering*, 1(1), 211–240. doi:10.1146/annurev.bioeng.1.1.211 PMID:11701488
- Intuitive Surgical. (2015, October 01). *Intuitive Surgical - da Vinci Surgica System Si*. Retrieved from http://www.intuitivesurgical.com/products/davinci_surgical_system/davinci_surgical_system_si/
- Kuo, C.-H. J. (2009). Robotics for Minimally Invasive Surgery: A Historical Review from the Perspective of Kinematics. In *International Symposium on History of Machines and Mechanisms* (pp. 337-354). Springer Netherlands. doi:10.1007/978-1-4020-9485-9_24
- Kuo, C.-H. J. (2012). Kinematic design considerations for minimally invasive surgical robots: an overview. *Int J Med Robotics Comput Assist Surg*, 127-145.
- Lanfranco, A. R., Castellanos, A. E., Desai, J. P., & Meyers, W. C. (2004, January 1). Robotic Surgery; A Current Perspective. *Annals of Surgery*, 239(1), 14–21. doi:10.1097/01.sla.0000103020.19595.7d PMID:14685095
- Li, J. G. (2014). A Class of 2-Degree-of-Freedom Planar Remote Center-of-Motion Mechanisms Based on Virtual Parallelograms. *ASME Journal of Mechanisms and Robotics*.
- Li, J. Y. (2015). Kinematic Design of a Novel Spatial Remote Center-of-Motion Mechanism for Minimally Invasive Surgical Robot. *Transactions of the ASME; Journal of Medical Devices*, 1-8. doi:10.1115/1.4028651
- Lum, M. J., Friedman, D. C. W., Sankaranarayanan, G., King, H., Fodero, K., Leuschke, R., & Sinanan, M. N. et al. (2009). The RAVEN: Design and Validation of a Telesurgery System. *The International Journal of Robotics Research*, 28(9), 1183–1197. doi:10.1177/0278364909101795
- Lum, M. J., Rosen, J., Sinanan, M. N., & Hannaford, B. (2006, July). Optimization of a Spherical Mechanism for a Minimally Invasive Surgical Robot: Theoretical and Experimental Approaches. *IEEE Transactions on Bio-Medical Engineering*, 53(7), 1440–1445. doi:10.1109/TBME.2006.875716 PMID:16830951
- Madhani, A. J. (1998). *U.S Patent No. 5,807,377*. US Patent Office.
- Nisar, S., Endo, T., & Matsuno, F. (2017). Design and Kinematic Optimization of a Two Degrees-of-Freedom Planar Remote Center of Motion Mechanism for Minimally Invasive Surgery Manipulators. *ASME. J. Mechanisms Robotics.*, 9(3), 031013–031013-9. doi:10.1115/1.4035991
- Nisar, S. O. (2015). Telesurgical Robotics. In M. Khosrow-Pour (Ed.), *Encyclopedia of Information Science and Technology* (3rd ed.). IGI; doi:10.4018/978-1-4666-5888-2.ch541
- Rosen, J. M. (2010). Raven: Developing a Surgical Robot from a Concept to a Transatlantic Teleoperation Experiment. In *Surgical Robotics* (pp. 159-197). Springer US.
- Satava, R. M. (2004). Future Trends in the Design and Application of Surgical Robots. *Seminars in Laparoscopic Surgery*, 11(2), 129–135. PMID:15254652

Stoianovici, D. W. (1998). A modular surgical robotic system for image guided percutaneous procedures. In *Medical Image Computing and Computer-Assisted Intervention* (pp. 404–410). Berlin: Springer. doi:10.1007/BFb0056225

Taylor, R. H. (2006). A Perspective on Medical Robotics. *Proceedings of the IEEE*, 94(9), 1652–1664. doi:10.1109/JPROC.2006.880669

ADDITIONAL READING

Gomes, P. (2011). Surgical Robotics: Reviewing the Past, Analysing the Present, Imagining the Future. *Robotics and Computer-integrated Manufacturing*, 27(2), 261–266. doi:10.1016/j.rcim.2010.06.009

Kuo, C. H., & Dai, J. S. (2009, January). Robotics for minimally invasive surgery: a historical review from the perspective of kinematics. In *International symposium on history of machines and mechanisms*, (pp. 337–354). Springer Netherlands. doi:10.1007/978-1-4020-9485-9_24

KEY TERMS AND DEFINITIONS

Minimally Invasive Surgery (MIS): MIS is a surgical technique in which surgical tools are inserted inside the patient body through small holes and thus allow the surgeon to operate using the video feedback from an endoscopic camera.

Robot Kinematics: Robot kinematics is the study of the position and movement of multi-degree of freedom kinematic chains that form the structure of a robotic manipulator.

Singular Configurations: Singularities are robot configurations where a manipulator loses one or more of its degrees-of-freedom and, therefore, cannot move in the corresponding direction(s).

Spatial Mechanism: Spatial mechanisms are essentially three-dimensional mechanisms such that one or more members of such a mechanism move in a plane different to others.

Workspace: Workspace or reachable space represents all those points that can be reached by the end effector of a robot.

Using Global Appearance Descriptors to Solve Topological Visual SLAM

Lorenzo Fernández Rojo

Miguel Hernandez University, Spain

Luis Paya

Miguel Hernández University, Spain

Francisco Amoros

Miguel Hernandez University, Spain

Oscar Reinoso

Miguel Hernandez University, Spain

INTRODUCTION

Nowadays, the use of mobile robots has extended to many different environments, where they have to move autonomously to fulfill an assigned task. With this aim, it is necessary that the robot builds a model of the environment and estimates its position using this model. These two problems are often faced simultaneously. This process is known as SLAM (Simultaneous Localization and Mapping) and is very common since when a robot begins moving in a previously unknown environment it must start generating a model from the scratch while it estimates its position simultaneously.

To carry out the SLAM process, the robot has to make use of the different sensors it may be equipped with (e.g. odometry, touch, laser, cameras, etc.). During the last years, the use of cameras has extended thanks to the amount of information they can capture from the environment and their relatively low cost. This work is focused on the use of computer vision.

When a vision sensor is used to solve the mapping and localization tasks, we must take into account that scenes contain a great quantity of information. This way, it is necessary to extract the most relevant information from the scenes as it will allow us to work with a lower

number of components. This problem can be approached from two points of view: local and global appearance methods. First, the methods based on local appearance consist in extracting and describing prominent landmarks or regions from the scenes. These methods typically need more computational time to build the map and estimate the position of the robot since it is necessary to extract the distinctive features from each image, describe them and make a complete comparison with the data stored in the map. Second, global-appearance methods describe each scene with a unique descriptor that contains information of the whole appearance. These methods tend to be computationally more efficient.

The SLAM problem can be approached from three different points of view: metric, topological and hybrid metric-topological SLAM. First, the metric approach consists in representing the position of some landmarks in the environment with geometric accuracy, with respect to a reference system. Using this approach, the position of the robot can be estimated with respect to this system. However, these methods often present a high computational cost. Second, the topological approach represents typically the environment as a graph, where the nodes represent prominent localizations (e.g. rooms) and links are the connectivity

relations between areas. These approaches offer a more compact representation that supposes a lower computational cost, but the localization process is inherently less accurate. At last, some researchers have started working on hybrid approaches that try to combine the advantages of the metric and the topological approaches.

The main objective of this work is to develop and test an algorithm to solve the SLAM problem using two sources of information: (a) the global appearance of omnidirectional images captured by a camera mounted on the mobile robot and (b) the robot internal odometry. We have decided to use a hybrid metric-topological approach to solve the SLAM problem.

BACKGROUND

The SLAM problem has been extensively studied. Moravec and Elfes (1985) developed one of the first works in this area. They build a metric map by means of wide-angle sonar range measurements and a probabilistic approach. Subsequently laser sensors were introduced to improve the accuracy and computational efficiency of the algorithms. For example, Thrun (2001) develops a SLAM algorithm in which a team of robots builds a map online using laser sensors and a Monte Carlo approach. Lately, the use of cameras in the field of mobile robotics has become widespread due to the numerous advantages they offer (passive sensors, low cost, large amount of information, low power consumption, etc.). Many authors have studied the SLAM problem both using local features (Gil et al., 2010; Valiente et al., 2015) or global appearance (Paya et al., 2014; Berenguer et al. 2015).

The main problem that local features present is the high computational cost of the necessary processes. As a feasible alternative, the use of topological approaches is a field of great interest in the construction of maps by means of the global appearance of visual information, due to the numerous advantages it presents in terms of

simplicity and computational cost. For example, Menegatti et al. (2004) carried out a study on robot navigation using the omnidirectional visual information captured from the environment, by using a global appearance descriptor.

Cameras have been used both to solve the SLAM problem metrically, topologically and in a hybrid way. For example, Guerra et al. (2014) and Gil et al. (2010) present two metric visual SLAM approaches using respectively a monocular and a stereo-camera and SURF features. On the other hand, Lui et al. (2012) develop a purely vision-based topological approach that builds an incremental model based on Haar Wavelet and Werner et al. (2009) carried out a task of topological SLAM using vision-based techniques and global appearance. They also make use of omnidirectional images, and furthermore they propose a Bayesian approach that combines the odometry information of the robot with the visual information to improve the accuracy. At last, Tully et al. (2012) develop a unified filtering framework for hybrid SLAM.

In this work we present a comparison of three global appearance descriptors in a process of robotic mapping, using a hybrid metric-topological SLAM.

APPEARANCE-BASED SLAM

In this work, we develop a new approach to hybrid SLAM, trying to obtain the advantages of each method. On the one hand, the topological representation permits building a global compact representation of the environment that permits a quick localization. On the other hand, the metric representation makes use of the information provided by the topological method to detect loop closures, so that it is possible to correct the possible errors in the position of the robot. In both cases, the global appearance of the omnidirectional scenes captured by the sensor mounted on the robot is used to create the model and localize

the robot. The next two subsections give details on the topological and the metrical approaches of the algorithm.

Constructing a Topological Map

In the global appearance approach, each scene is described by means of a unique descriptor that contains information on the whole appearance. This way, it is necessary to define first a method to obtain a descriptor per scene. This descriptor should retain the information in a compact and efficient manner and it must be computed quickly. In this work we make use of and compare three different and well known global appearance descriptors: Fourier Signature (FS) (Payá et al., 2010), the gist of the scene (Siagian and Itti, 2007) and the Histogram of Oriented Gradients (HOG) (Amorós et al., 2010).

In this case, the environment is represented through a graph where (a) each node represents an area of the environment and contains one or several images with similar visual appearance and (b) each edge represents the connectivity relationships between the nodes, i.e. the possibility to navigate from one node to another (Romero et al., 2010). Since each node may contain several images, a representative image per node is computed. We define the similarity $S(i, j)$ between two images I_i and I_j that belong to the same node with this expression:

$$S(d_i, d_j) = \frac{1}{\sum_{m=0}^M \sum_{n=0}^N (d_i(m, n) - d_j(m, n))^2} \quad (1)$$

Where $d_i(m, n)$ represents the descriptor of the panoramic image I_i . In general, the descriptor is a matrix of size $M \times N$. The distance criterion we use is the Euclidean distance, which is the term in the denominator. Then, the closer the images are, the smaller the distance and the higher the similarity $S(d_i, d_j)$. Once we have

computed the similarity between every two images in the node, we use the following equations to compute the most representative image:

$$\begin{aligned} N_R &= \max_{i \in P} \left[\min_{j \in P, j \neq i} (S(d_i, d_j)) \right] \\ R &= \arg \max_{i \in P} \left[\min_{j \in P, j \neq i} (S(d_i, d_j)) \right] \end{aligned} \quad (2)$$

Where P represents the set of nodes in the graph and R is the image that best represents the node.

Then, to build the topological map we use the following algorithm. Let's suppose the robot captures a new image I_k and this information has to be added to the topological map.

1. The descriptor d_k of the image is obtained and compared with the descriptor of the image that represents the current node R_c . If the similarity $S(d_k, R_c)$ is over certain threshold Th_{min} , it is added to the current node, and the node representative is re-calculated.
2. If the similarity $S(d_k, R_c)$ does not exceed the threshold Th_{min} , d_k is compared with the node representatives of all the neighbor nodes to the current one. The node with the higher similarity $S(d_k, R_{nei})$ is chosen. Then, if this new similarity exceeds the threshold, the current image is added to that adjacent node, and the node representative is re-calculated including the information of the new descriptor.
3. In the case that the similarity $S(d_k, R_{nei})$ does not exceed the threshold, the current image descriptor is compared with the representative descriptor of the rest of the nodes. If the similarity exceeds the threshold in any case, it is added to that node, and the node representative is re-calculated.
4. If no match is found, a new node is added to the system, so an edge between the new

node and the previous one is added too (i.e. they have become neighbor nodes). The new node includes the current image, which is also the node representative until a new image arrives, when this representative will be updated.

Constructing a Metric Map

The metric map is built using a visual SLAM approach based in Payá et al. (2010). We combine a Monte-Carlo Localization (MCL) algorithm with a landmark estimation process. The robot decides when its current location is included as a metric map landmark. The main improvements over previous work mainly lie in the combination of a mapping process and a localization process simultaneously, combining a topological method with a metric method.

In a MCL problem, the aim is the estimation of the robot's pose $x_{t+1} = (x, y, \theta)$ at time instant $t + 1$ using a set of measurements $z_{1:t+1} = \{z_1, z_2, \dots, z_{t+1}\}$ from the environment and the movements $u_{1:t+1} = \{u_1, u_2, \dots, u_{t+1}\}$ of the robot. The probability density function $p(x_{t+1} | z_{1:t+1}, u_{1:t+1})$ is represented by a set of M random samples $x_{t+1} = x_{t+1}^i$, $i = 1, \dots, M$ extracted from it (particles), where each particle can be understood as a hypothesis of the true state of the robot $x_{t+1}^i = (x^i, y^i, \theta^i)$. Each particle has a weight associated that determines the importance of the particle. The set of particles defines a discrete probability function that approximates the continuous belief. The original MCL algorithm consists of three main phases: (1) *Prediction phase*, in which a set of particles $\bar{\mathcal{X}}_{t+1}$ is generated based on the set of particles $\mathcal{X}_t$ and a control signal u_{t+1} ; (2) *Update phase*, in which the weight ω_{t+1}^i of each particle in the set $\bar{\mathcal{X}}_{t+1}$ is computed using the observation z_{t+1} ; and (3) *Resampling phase*, in which the resulting set $\mathcal{X}_{t+1}$ is computed by

resampling with replacement from the set $\bar{\mathcal{X}}_{t+1}$, where the probability of resampling each particle is proportional to its importance weight ω_{t+1}^i , in accordance with the literature on the SIR algorithm (Sampling Importance Resampling). Finally, the set $\mathcal{X}_{t+1}$ represents the distribution $p(x_{t+1} | z_{1:t+1}, u_{1:t+1})$.

In this work, since the objective is building a map of the environment while the position of the robot is computed (SLAM), the experiment starts without a map of the environment. Taking into account this consideration, the initial set of samples is drawn from a narrow Gaussian centered at the initial point $(x, y) = (0, 0)$. At the start of the experiment, the first map landmark (l_0) corresponds with the first pose of the robot x_0 and when the robot moves and captures a new image, the set of particles of the initial pose also moves adding some error to the movement of each particle, taking into account the movements $u_{1:t+1} = \{u_1, u_2, \dots, u_{t+1}\}$ of the robot. Moreover, the topological algorithm computes the representative pose of the node every time a new image is added, and it will appear on the metric map as the set of particles that represents the metric localization of the node. This way, each representative pose of the topological map has one metric map landmark (l_i) associated. So, although we compute each set of particles that represents each movement of the robot, we only add to the metric map the set of particles that represents the node. On the other hand, our MCL algorithm differs with respect to the traditional MCL algorithm in that we do not carry out an update and resampling process in each movement of the robot but only when the topological algorithm detects a loop closure, even though we perform a prediction phase at each movement of the robot.

When the topological algorithm detects a loop closure, an update and resampling process is activated. To compute the weight of each particle ω_{t+1}^i we use as input both, the metric information

Figure 1. Robot used to develop the SLAM experiments, catadioptric vision sensor and sample omnidirectional and panoramic images



(odometry) and the visual information (global image descriptor) through the following equation:

$$\omega_{t+1}^i = e^{-v_i \mathbf{f}_i^{-1} v_i^T} \cdot e^{-h_j \mathbf{f}_i^{-1} h_j^T} \quad (3)$$

where v_i represents the difference between the position of the landmark l_j and the position (x_i, y_i) of each particle i ($v_i = (l_x^j, l_y^j) - (x_i, y_i)$), $\mathbf{f}_i$ is a diagonal matrix $\mathbf{f}_i = \begin{bmatrix} \sigma_i^2 & 0 \\ 0 & \sigma_i^2 \end{bmatrix}$ where σ_i^2 has been chosen in order to minimize the error in the localization of the robot. On the other hand, $h = |d_j - d_i|$ is the difference between the appearance descriptor associated to the current observed image and the descriptor associated to the landmark l_j . Once the weight of each particle has been computed, the resulting set $\mathcal{X}_{t+1}$ is computed by resampling with replacement from the set $\bar{\mathcal{X}}_{t+1}$, where the probability of resampling each particle is proportional to its importance weight ω_{t+1}^i .

RESULTS AND DISCUSSION

With the objective of evaluating our SLAM algorithm, we carry out a realistic experiment with a

mobile robot in an indoor environment. The robot was manned to travel a specified route into the environment, so that several loop closures occur and it captured a complete set of images. The ground truth is computed from the data collected with a laser. Figure 1 shows the mobile robot used, the catadioptric vision sensor mounted on the robot and two sample images: an omnidirectional scene and its panoramic version. Figure 2 shows a bird eye's view of the environment where the experiments were developed and figure 3 shows some sample panoramic scenes captured in this environment.

Based on previous work (Payá et al., 2010) and to evaluate our method, we have decided to use the Procrustes analysis. It provides us with a measure of how similar is the layout of the landmarks after the SLAM process, comparing to the real layout. As a result of this process we obtain a parameter $\mu \in [0, 1]$ (shape difference), where μ is a measure of the shape correspondence between the sets of points A and B , so that the lower is μ , the more similar are A and B .

Figure 4 shows an example of metric map built after an experiment using the FS, 32 components per row and a number of particles equal to 200. In this figure, the particles that represent the position of the landmarks are plotted as green points and the resulting positions are represented with blue crosses. The internal odometry is plotted as a blue

Figure 2. Bird's eye view of the environment where the experiments have been carried out. The green dots are the positions where the robot captures a new image to be included in the model and the blue point is both the beginning and the end point of the route.

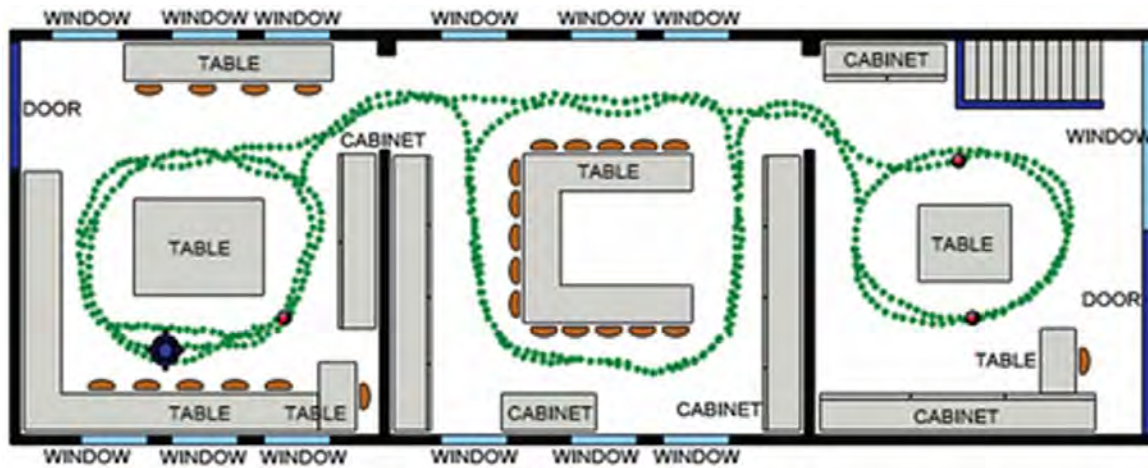


Figure 3. Some examples of panoramic images captured by the robot in the environment



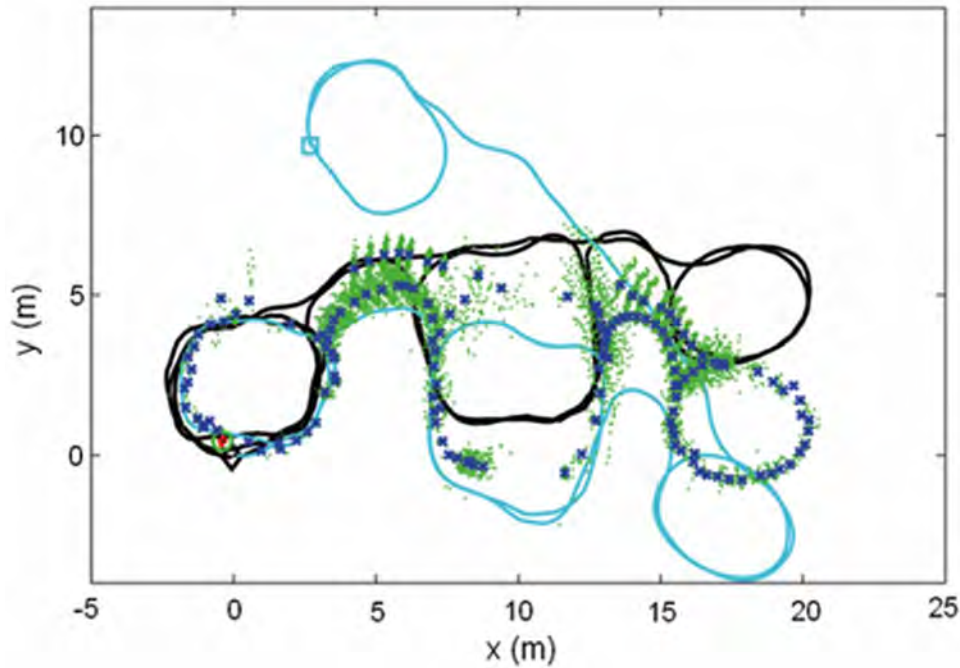
line. Finally, the ground truth is represented with a black line. The ground truth is shown only for comparative effects, but it is not considered in the SLAM algorithm.

According to fig. 4, the particles' dispersion of each landmark grows when the robot moves from the initial position until the robot detects a loop closure. At this moment, the dispersion of the samples decreases due to the fact that the algorithm carries out a resampling of these samples. It can be observed that the map obtained by means of our algorithm is considerably more accurate than the map obtained through the internal odometry.

Once proved that the algorithm works with a specific configuration, a set of experiments has been designed to test the performance of the hybrid algorithm when the three different appearance-based descriptors are considered (FS, GIST and HOG). The influence of the number of components of each descriptor is tested, since this is an important parameter that may have an effect both in the accuracy of the process and in the computational cost.

On the one hand, the μ factor is used to evaluate the accuracy of the resulting metric map and on the other, the step time needed by the al-

Figure 4. Metric map built with the hybrid algorithm (blue crosses) compared to the map build from the odometry data (blue line) and the ground truth (black line). The descriptor is the FS with 32 components per row and 200 particles in the MCL algorithm.



gorithm is used to evaluate the computational feasibility (ability to work in real time). In all cases, the robot odometry is used as the input of the prediction phase in the MCL algorithm.

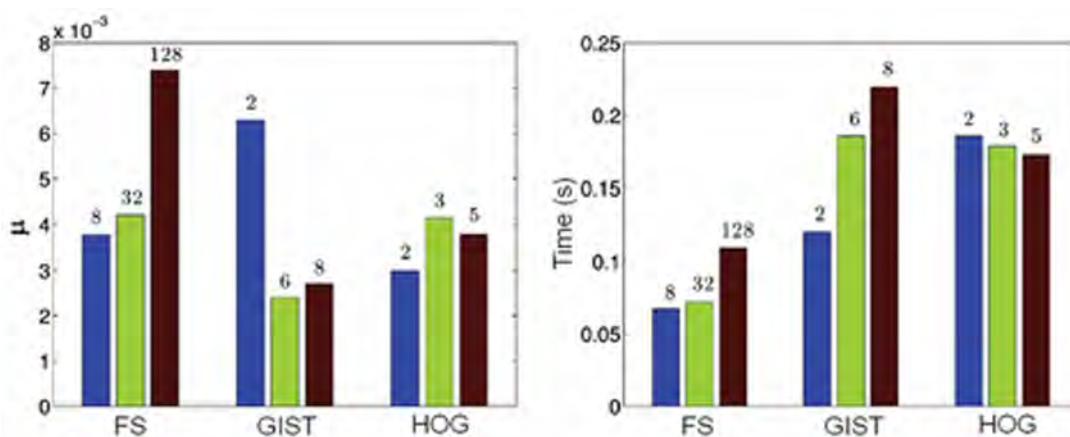
In Figure 5 we can see the μ factor and the step time needed by the algorithm, depending on the size of the descriptor. We have performed several experiments using the three descriptors (separately) and a number of particles equal to 200. We have chosen this number of particles as, in general, it offers a good accuracy with a relatively low computational cost.

As far as the size of each descriptor is concerned (a) in the case of FS we have to decide the number of components per row to store, (b) in the case of *gist*, the size of the mask used to calculate the descriptor can be configured, what has an effect on the size of the descriptor and (c) in the case of HOG, it is possible to modify the size of the descriptor by means of changing the size of the window used to compute the descriptor (Payá et al., 2014).

In general, fig. 5 shows that in the case of FS and *gist*, the accuracy is influenced by the size of the descriptor more than in the case of HOG. However, in general, we can say that the three methods present a good accuracy, even in the worst case. The μ factor is lower than 0.0075 in all the cases. We must take into account that the μ factor we get by using only the internal odometry without our algorithm is around 0.1421. With respect to the time t , FS clearly improves the other two descriptors. Anyway, with all three descriptors it is possible to work in real time since the times obtained are under 0.25 sec per iteration.

We have tested also the influence of the number of particles in the accuracy and computational cost, and the results have shown that the accuracy grows as the number of particles does until this number is equal to 200. From this number, the accuracy remains constant but the computational cost keeps on growing. This way, the optimal number of particles in all cases is around 200 particles.

Figure 5. Shape difference μ (left) and step time of the algorithm (right) versus the number of components of the descriptor. The FS, gist and HOG descriptors are evaluated.



FUTURE RESEARCH DIRECTIONS

The algorithm we have presented constitutes a promising starting point for the implementation of an accurate and efficient hybrid SLAM algorithm using the global appearance of visual information.

Currently, we are working in the improvement of this algorithm from several points of view. First, we are extending the algorithm to larger and dynamic environments including outdoors scenes. Second, we are considering the presence of some typical situations that usually come out when visual information is used: visual aliasing, occlusions (due to the presence of people moving), noise in the scenes and changing lighting conditions in the environment. We are evaluating new global-appearance descriptors and their ability to cope with these situations. At last, we are considering to include a multi-hypothesis framework that keeps some map hypothesis every time a loop closure is detected and that chooses among these hypothesis when it has enough information.

CONCLUSION

In this work, a hybrid metric-topological SLAM algorithm has been proposed and evaluated. The mapping and localization processes are solved

using the panoramic images captured by a catadioptric system mounted on the robot and three global appearance descriptors have been used: Fourier Signature, *gist* and Histogram of Oriented Gradients. This way, the contribution of the work is twofold: a new efficient SLAM algorithm is proposed and tested and several descriptors are evaluated and optimized. Thanks to this framework, compact maps of the environment can be obtained using only the odometry and visual information.

The experiments have been developed in an unstructured indoors environment where a set of real images have been captured. Both the size of the descriptors and the number of particles in the MCL algorithm have been evaluated and optimized to obtain an accurate map with a feasible computational cost. The use of several description methods proves that our algorithm is generalizable to any other image descriptor, including local descriptors.

ACKNOWLEDGMENT

This work has been supported by the Spanish Government through the project DPI2016-78361-R: “Creación de mapas mediante métodos de apariencia visual para la navegación de robots” and by the Generalitat Valenciana (GVa) through

the project GV/2015/031: “Creación de mapas topológicos a partir de la apariencia global de un conjunto de escenas”.

REFERENCES

- Amorós, F., Payá, L., Reinoso, O., Fernández, L., & Marín, J. M. (2010). Visual map building and localization with an appearance-based approach. *Proceeding of the 7th International Conference on Informatics in Control, Automation and Robotics*, 423–426.
- Berenguer, Y., Payá, L., Ballesta, M., & Reinoso, O. (2015). Position estimation and local mapping using omnidirectional images and global appearance descriptors. *Sensors (Basel, Switzerland)*, 15(10), 26368–26395. doi:10.3390/s151026368 PMID:26501289
- Gil, A., Reinoso, O., Ballesta, M., Juliá, M., & Payá, L. (2010). Estimation of visual maps with a robot network equipped with vision sensors. *Sensors (Basel, Switzerland)*, 10(5), 5209–5232. doi:10.3390/s100505209 PMID:22399930
- Guerra, E., Munguía, R., & Grau, A. (2014). Monocular SLAM for autonomous robots with enhanced features initialization. *Sensors (Basel, Switzerland)*, 14(4), 6317–6337. doi:10.3390/s140406317 PMID:24699284
- Lui, W., & Jarvis, R. (2012). A pure vision based topological SLAM system. *The International Journal of Robotics Research*, 31(4), 403–428. doi:10.1177/0278364911435160
- Menegatti, E., Zoccarato, M., Pagello, E., & Ishiguro, H. (2004). Image-based memory for robot navigation using properties of omnidirectional images. *Robotics and Autonomous Systems*, 47(4), 251–267. doi:10.1016/j.robot.2004.03.014
- Moravec, H. P., & Elfes, A. E. (1985). High resolution maps from wide angle sonar. *Proceeding of the IEEE International Conference on Robotics & Automation*, 116–121. doi:10.1109/ROBOT.1985.1087316
- Payá, L., Amorós, F., Fernández, L., & Reinoso, O. (2014). Performance of global-appearance descriptors in map building and localization using omnidirectional vision. *Sensors (Basel, Switzerland)*, 14(2), 3033–3064. doi:10.3390/s140203033 PMID:24534922
- Payá, L., Fernández, L., Gil, A., & Reinoso, O. (2010). Map building and Monte Carlo localization using global appearance of omnidirectional images. *Sensors (Basel, Switzerland)*, 10(12), 11468–11497. doi:10.3390/s101211468 PMID:22163538
- Romero, A., & Cazorla, M. (2010). Topological slam using omnidirectional images: Merging feature detectors and graph-matching. In J. Blanc-Talon, D. Bone, W. Philips, D. Popescu & P. Scheunders (Eds.), *Advanced Concepts for Intelligent Vision Systems: Proceedings of the 12th International Conference (ACIVS 2010)*. Sydney, Australia: Springer.
- Siagian, C., & Itti, L. (2007). Rapid biologically-inspired scene classification using features shared with visual attention. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(2), 300–312. doi:10.1109/TPAMI.2007.40 PMID:17170482
- Thrun, S. (2001). A probabilistic online mapping algorithm for teams of mobile robots. *The International Journal of Robotics Research*, 20(5), 335–363. doi:10.1177/02783640122067435
- Tully, S., Kantor, G., & Choset, H. (2012). A unified Bayesian framework for global localization and SLAM in hybrid metric/topological maps. *The International Journal of Robotics Research*, 31(3), 271–288. doi:10.1177/0278364911433617

Valiente, D., Ghaffari, M., Valls-Miró, J., Gil, A., & Reinoso, O. (2015). Information-based view initialization in visual SLAM with a single omnidirectional camera. *Robotics and Autonomous Systems*, 72(1), 93–104. doi:10.1016/j.robot.2015.05.005

Werner, F., Maire, F., & Sitte, J. (2009). Topological SLAM using fast vision techniques. In *Advances in Robotics: Proceedings of the FIRA RoboWorld Congress 2009*. Incheon, Korea: Springer. doi:10.1007/978-3-642-03983-6_23

ADDITIONAL READING

Amorós, F., Payá, L., Reinoso, O., & Fernández, L. (2012). Map building and localization using global-appearance descriptors applied to panoramic images. *Journal of Computer and Information Technology*, 2(1), 55–71.

Andreasson, H., & Duckett, T. (2004). Topological localization for mobile robots using omnidirectional vision and local features. In *Proceeding of the 5th Symposium of Intelligent Autonomous Vehicles*. Lisbon, Portugal.

Angeli, A., Doncieux, S., Meyer, J. A., & Filliat, D. (2009). Visual topological SLAM and global localization. In *Proceedings of the International Conference on Robotics and Automation*, Kobe (pp. 4300–4305).

Bay, H., Tuytelaars, T., & Van Gool, L. (2006). SURF: Speeded Up Robust Features. In *Proceeding of the ninth European Conference on Computer Vision*, Graz, Austria (pp. 404–417).

Cummins, M., & Newman, P. (2008). Fab-map: Probabilistic localization and mapping in the space of appearance. *Journal of Robotics Research*, 27(6), 647–665. doi:10.1177/0278364908090961

Davison, A. J. (2003). Real-time simultaneous localisation and mapping with a single camera. In *Proceedings of the International Conference on Computer Vision*, Sendai, Japan (pp. 1403–1410). doi:10.1109/ICCV.2003.1238654

Fernández, L., Gil, A., Payá, L., & Reinoso, O. (2010). An evaluation of weighting methods for appearance-based Monte Carlo localization using omnidirectional images. In *Proceedings of the IEEE International Conference on Robotics and Automation (Workshop on Omnidirectional Robot Vision)*, Anchorage, Alaska, USA (pp. 13–18).

Fernández, L., Payá, L., Valiente, D., Gil, A., & Reinoso, O. (2012). Monte Carlo localization using the global appearance of omnidirectional images. Algorithm optimization to large indoor environments. In *Proceedings of the 9th International Conference on Informatics in Control, Automation and Robotics*, Rome, Italy (pp. 439–442).

Fox, D., Burgard, W., & Thrun, S. (1999). Markov localization for mobile robots in dynamic environments. *Journal of Artificial Intelligence Research*, 11, 391–427.

Friedrich, H., Dederscheck, D., Krajsek, K., & Mester, R. (2007). View-based robot localization using spherical harmonics: Concept and first experimental Results. In F. A. Hamprecht, C. Schnörr & B. Jähne (Eds.), *Pattern Recognition: Proceedings of the 29th DAGM Symposium*. Heidelberg, Germany (LNCS 4713, pp. 21–31).

Gil, A., Ballesta, M., Juliá, M., Reinoso, O., & Ubeda, D. (2012). Cooperative simultaneous localisation and mapping using independent Rao-Blackwellised filters. *Journal of Computer Vision, IET*, 6(5), 407–414. doi:10.1049/iet-cvi.2011.0108

Gil, A., Reinoso, O., Ballesta, M., & Juliá, M. (2010). Multirobot visual SLAM using a Rao-Blackwellized particle filter. *Robotics and Autonomous Systems*, 58(1), 68–80. doi:10.1016/j.robot.2009.07.026

- Kröse, B., Bunschoten, R., Hagen, S. T., Terwijn, B., & Vlassis, N. (2004). Household robots look and learn: Environment modelling and localization from an omnidirectional vision system. *Journal of IEEE Robotics and Automation Magazine*, 11(4), 45–52. doi:10.1109/MRA.2004.1371608
- Montemerlo, M., Thrun, S., Koller, D., & Wegbreit, B. (2002). FastSLAM: A factored solution to the simultaneous localization and mapping problem. In *Proceeding of the Eighteenth International conference on Artificial Intelligence*, Menlo Park, CA, USA (pp. 593-598).
- Murillo, A. C., Guerrero, J. J., & Sagüés, C. (2007). SURF features for efficient robot localization with omnidirectional images. In *Proceeding of the IEEE International Conference on Robotics & Automation*, San Diego, CA, USA (pp. 3901-3907). doi:10.1109/ROBOT.2007.364077
- Oliva, A. (2005). Gist of the Scene. In L. Itti, G. Rees, & J. K. Tsotsos (Eds.), *The encyclopedia of neurobiology of attention* (pp. 251–256). San Diego, CA: Elsevier; doi:10.1016/B978-012375731-9/50045-8
- Payá, L., Fernández, L., Reinoso, O., Gil, A., & Úbeda, D. (2009). Appearance-based dense maps creation: Comparison of compression techniques with panoramic images. In *Proceeding of the 6th International Conference on Informatics in Control, Automation and Robotics*, Milan, Italy (pp. 250-255).
- Payá, L., Reinoso, O., Amorós, F., Fernández, L., & Gil, A. (2011). Probabilistic map building, localization and navigation of a team of mobile robots. Application to route following. *Journal of Multi-Robot Systems: Trends and Development*, 1, 191–210.
- Payá, L., Reinoso, O., Gil, A., & Sogorb, J. (2008). Multi-robot route following using omnidirectional vision and appearance-based representation of the environment. *Journal of Lecture Notes in Computer Science*, 5271, 680–687. doi:10.1007/978-3-540-87656-4_84
- Shi, C., Wang, Y., & Yang, J. (2010). Online topological map building and qualitative localization in large-scale environment. *Journal of Robotics and Autonomous Systems*, 58(5), 488–496. doi:10.1016/j.robot.2010.01.009
- Thrun, S., Fox, D., Burgard, W., & Dellaert, F. (2000). Robust Monte Carlo localization for mobile robots. *Journal of Advanced Robotics*, 19(10), 1059–1080.
- Valgren, C., Lilienthal, A. J., & Duckett, T. (2006). Incremental topological mapping using omnidirectional vision. In *Proceeding of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Beijing, China (pp. 3441-3447).
- Valiente, D., Gil, A., Amorós, F., & Reinoso, O. (2013). SLAM of view-based maps using SGD. In *Proceeding of the 10th International Conference on Informatics in Control, Automation and Robotics*, Reykjavik, Iceland (pp. 329-336).
- Wolf, J., Burgard, W., & Burkhardt, H. (2005). Robust vision-based localization by combining an image retrieval system with Monte Carlo localization. *Journal of IEEE Transactions on Robotics*, 21(2), 217–229.
- Zhiwei, L., Xiang, G., Yanyan, C., & Songhao, Z. (2013). A novel loop closure detection method in monocular SLAM. *Journal of Intelligent Service Robotics*, 6(2), 79–87. doi:10.1007/s11370-012-0124-0

KEY TERMS AND DEFINITIONS

Appearance Descriptor: It is a descriptor of an image that represents the global information of the same without extracting landmarks.

Gist: It is the meaning of a scene, or in other words, the spatial envelope of the scene.

Localization: It is the estimation of the position of an autonomous agent in a given map.

Mapping: It is the creation of an internal representation of any given environment.

Metrical Map: It is a representation of the environment through geometrical information with certain accuracy.

Mobile Robot: It is an autonomous vehicle that is capable of movement in any given environment.

Omnidirectional Vision: It is a vision system that is capable of capturing all the information surrounding the system with a single image (360°).

Probabilistic Localization: It is a localization task, where the information of all previous robot locations is used to estimate its current location.

SLAM: It is the process of building a map of an environment while simultaneously the localization of the agent that compute the map is estimated.

Topological Map: It is a representation of the environment by means of a list of locations within a graph with connectivity relationships between them.

R

Category S

Small and Medium Enterprises

Big Data and Simulations for the Solution of Controversies in Small Businesses

S

Milena Janakova

Silesian University in Opava, Czech Republic

INTRODUCTION

General perspective of the chapter is focused on the use of big data and intelligence in the field of information technology. Information technology (IT) supports the majority of the realized activities. The press is focused on time and quality. There are also competitions, customer preferences and downward pressure on prices and costs. In this situation, new products and services have to address both existing and new customers. Such innovations actively use information technology and available information from the Internet. Good and bad advice, experiences or requests are immediately available for everyone via successful story. (Concessao, 2016) The behavior of individuals in a global information society reveals similarities with regard to the collective behavior of animals in nature. Swarm intelligence provides inspiration for various fields such as economics, biology and also computer science.

The global information society creates data in various formats and data is stored in many sources. Data (big data) is everywhere around us. The well-known databases (Barbucha et al., 2015) are default data source, but there are many other sources such as the Internet, e-mails, chats, and also personal notes in PDF or text files. The volume of data is as large as the ocean or the universe. Users of information technology swim in the data without lifebuoys in all companies and organizations. Everyone wishes to get knowledge immediately in the needed format, but knowledge is not data. There are difficulties that cause ambiguous tasks, unrealistic expectations, faulty data or inadequate

procedures and processes. The special problems create a large volume of available data.

The objective of interest is native support of sustainable development based on big data and intelligence in the field of information technology. (Acharjya et al., 2015) In this situation, there is a spectrum of well-known approaches and the fields that are oriented on big data and intelligence such as artificial intelligence, business intelligence, competitive intelligence, customer intelligence, swarm intelligence and also computational intelligence. At first glance, working with data must use all suitable approaches for correct data analysis with respect to sustainable development. Sustainable development is important for supporting society's continuous development without crises. (World Development Report, 2015) The question is "How can be solved existing problems responsibly?" The easy way is to depend on governments or luck. Another way is to merge power of mankind based on cooperation and communication. In this situation, big data brings good results.

BACKGROUND

Working with big data requires optimal IT support, skills of IT users, and more predictive analytics. Sense and simplicity are in the forefront. Interest is focused on true story formation. For these goals, there is computational intelligence bringing IT power to problem solution. The basic step is correct problem formulation in a complex view. CPU speed, memory size, disk volume, and network connection are not critical elements. Information

technology has enough resources to work with big data. (Information Technology Market Reports, 2015) The question is to find the best data for the required analysis and to create a suitable story for the given reality. The main role of computational intelligence must be focused on helping with data analysis and predicting further development to break unexpected conditions in society. (Covington, 2016) Market competition creates other hard requests on realized activities, computers and their software, and also IT users. We need a solution that is better, faster, more user-friendly, more complex, and more predictive.

Regardless on the wide spectrum of hardware, software, available methods and verified methodologies, there are big data projects that end with mistakes. There is contradiction between expectations and reality. IT users rely on information technology as a unique tool supporting realized tasks and processes for solving existing problems, but they are often disappointed over achieved results. IT users often must say “It doesn’t work.” An active solution takes inspiration from nature, specifically swarm intelligence with links to business intelligence. In many cases, foundation work is simulation. The benefit is easy creation and modification of simulations without demands on finance, material, or human resources.

MAIN FOCUS ON EXISTING CONTRAST

Big Data and Controversies

IT product development is perceived with a wide spectrum of conditions and preferences. Optimal IT implementation must be evaluated based on the available IT product, technology issues, data difficulties, company philosophy, employment approach, motivation, the aims and concept, management support, end users, the learning process, security of application, measurement and evaluation. The perception of IT implementation

is a difficult process with a number of relationships. This volume information is beyond human capacity and a suitable arrangement is needed.

For optimal implementation, the verified methods and system architectures are defined (Crawley et al., 2015), but there are doubts about disproportionate demands on IT users, implementation teams, human resources, finances and time, or societal stability. From a long-term perspective of existing controversies, big data will bigger and bigger, and the pressure on reducing costs for resources will continue to be greater and greater. The frequently used keywords show the rapidly expanding sphere of interest in the area of big data in detail: Analytics - 4 380, Big Data – 4 330, Business Intelligence – 2 550, Data Architect - 911, Data Mining - 2 080, Data Science – 4 370, Data Warehouse - 408, Database – 1 460, Excel – 3 800, Hadoop – 2 410, IoT (Internet of Things) – 1 800, Operations Research - 513, Predictive Modeling - 898, Python – 1 130, SQL - 945, Statistician – 1 090, Visualization – 2 620. (Graville, 2013 - 2015; Big Data Science on twitter, 2015)

Controversies also bring absence public of optimal software implementations. Many surveys show an existing disillusionment from results based on processes with data and information. (Gartner Newsroom, 2015) The natural question is “How do data, information, and knowledge help with difficulties in the global society?” Big data is able to derive various relations about the actual reality (useful as well as useless) based on optimal visualization. (Clark, 2014) The global information society and their governments need more. (Simon, 2014) For everyone, there are important changes:

- The exchange of goods and services, data and information.
- The dependence on information technology in all activities.
- The perception of individuals and collective behavior like people themselves.
- The rhythm of life of modern civilization.

But it is not all. Human behavior is strongly affected by the social networks in communities. (Top 15 Most Popular Social Networking Sites, 2015) The default characteristic of decision-making is a requirement on time and quality. Collective experience and thinking with models has a positive influence. (World Development Report, 2015) This variability is important for everyday activities in such a dynamic and rapid society. For the global information society, these changes, as well as information technology, support controversies that are oriented around global challenges. (Rischard, 2002) Available spectrum of IT products with needed designs, simulations and methods create sufficient space for optimal software implementation. (Top 27 Free Data Analysis Software, 2015) With regards to actual conditions, managers, business owners and IT users have selected an optimal approach for use of suitable software. There are difficulties, unfortunately.

Big Data and Intelligence

Computational intelligence often uses verified techniques as neural networks, fuzzy systems, hybrid neuro-fuzzy approaches, or genetic algorithms. (Jain et al., 2014) These approaches are suitable for creating simulation via available tools and applications. The issue is their complexity and variability. Ongoing changes are very dynamic and they occur unexpectedly. Answers to these changes must be realized in very short time. (Zobaa & Vaccaro, 2015) These answers need easy manipulation and instant access to variable support. A little extra time is good to have for an optimal reaction to existing reality. Good prediction may give extra time for adaptation, but the question is about the adopted strategy for predictions and algorithm evolution; therefore, the new challenge is connecting computational intelligence with swarm intelligence, business intelligence, and data science for true story creation based on big data in practice. (Marr, 2016)

Visualization, selection of optimal data, and dimension reduction may help, but it is not easy in today's modern and complex world. (Dzemyda et al., 2014) Visualization brings many graphs in various formats, but which one of them is appropriate and how do we derive the needed knowledge? Selection of optimal data is also problematic with regards to dynamic changes of actual conditions and needs. The challenge is working with data and other sources as with mirrors to human behavior that is similar to animal and swarm intelligence. This mirror of future uses all intelligence systems that may help. Computational intelligence, business intelligence, and swarm intelligence play key roles. The focus is oriented on simulations that cooperate with humans in many areas to find all interesting solutions automatically.

A considerable volume of work is well-manageable in a team, but many IT users prefer instant solutions and independent work in an intuitive interface with examples. (Matei et al., 2015) It is also the reason that IT users rather prefer default MS Excel and available contingency tables and they do not use other software for future prediction. This is a fatal mistake. They voluntarily give up benefits from big data and other intelligences. (Mohanty et al., 2015) Information technology offers processes for the analysis of accessible and interesting links in existing data sources. It is possible to target marketing campaigns to individual customers not only based on their shopping behavior, but behavior on the website or social networking sites.

SOLUTIONS AND RECOMMENDATIONS

This contrast leads to new solutions, attractive innovations, and needed global cooperation. In many cases, nature is, and will be, an inspiration for new products and approaches in the IT field.

Optimal work with data and information is complex. It is not a pile of data, information and

knowledge in a vacuum. Similarly as human brain maps (Lewis, 2015), all relationships must be respected together with reflecting the actual reality. (Mayer-Schönberger & Cukier, 2014) The starting point is defined by a division topic on a few areas of interest based on objects. The default recommendation is to work only with a few objects with regards to understanding the complexities; therefore, the optimal division of “big data” meaningfully respects the following three levels:

- **Basic Level:** Default activities and analyses for data storage in data warehouse (internal matter).
- **Advanced Level:** Searching for links between stored data in a data warehouse and information sources in the global society (external matter).
- **Variable Level:** Searching for unexpected events based on complex statistics and mathematical methods with the support of Artificial Intelligence, Business Intelligence, Customer Intelligence, Competitive Intelligence, Swarm Intelligence (uncertainty matter).

Big data helps solve many difficulties in the global information society based on relations with database systems and intelligence implemented in various systems. There are well-known model cases for predicting and monitoring reserves, prices, or impacting the use of anything such as daily consumer goods, rare commodity, and also customer preferences and expected behavior and success based on specified metrics. This active implementation uses all the benefits of information technology with respect to IT user needs. The practical solution to existing difficulties is to connect information technology not only with well-known principles in this field, but also to respect people’s needs and people’s behavior. It is easy to add the important metrics “benefits and risks” into software evaluation. This metric

indicates the optimal solution in a specific case of IT implementation via all the above-defined levels. This is the multidimensional approach.

The above-mentioned levels are focused on internal, external and unexpected matters via a multidimensional approach. The adopted approach must have adequate reflection in the visualization of the needed data. Visualization is important as “a graphic manual” for quick and immediate work with data. In many cases, visualization has a close relation to simulation. The reason for this is that IT users modify the used data structure in simulation and receive immediate output in the form of various graphs. This basic rule respects almost all applications that work with big data and which provide graphic support for processed data. Big data and suitable analysis need an optimal data collection. Data quality is one of the frequent difficulties in many companies and organizations. Small businesses have similar difficulties. For easy data collection, it is better to divide data into three groups that follow the above defined levels of work with big data.

On the basic level, data are collected in a well-known structure based on verified rules from database systems and a warehouse. All IT users seemingly have information about the need to store data into information systems. There is an essential relational approach, working with objects and creating relations between the designed objects. The available information is focused on pragmatic data storage for customer registration, customer orders, storage records and invoicing. For this purpose, graphs are useful in the form of bar graphs, pie charts or line charts.

On an advanced level, IT users look at processed data from a higher perspective. This is not only a view based on the local conditions in a company or organization, but this view must be extended to global society. The reason is natural, we all live in a global information society. Information has an initial place where it is created and stored, but this information is not hidden and IT users share their own experiences, preferences

and difficulties with the provided services or products. In this situation, it is important to know the skills of the competitions, the opinion of the users on the realized activities, their preferences and requirements. This interest is focused on long time data monitoring via various places, interests, companies and customers. For these cases, good examples of graphical presentation are based on clusters or collections of various graphs (charts). These collections reveal more about outgoing changes, and IT users are interested in predicting future development.

Questions concerning future development bring our interest to the third level of work with big data. This level is characterized by unexpected conditions. These predictions are from one topic which is attractive for all IT users. Who would not want to know what is going to happen? Unfortunately, modern global society has a number of unbalanced conditions and phenomena. Companies and organizations are concerned about losing orders, are interested in the offered services, rapid changes in customer preferences, or development of new technologies. The solution to such difficulties is also a great challenge for big data. Interested IT users may see various clusters of graphs that are interconnected via various numbers of relations (lines). Analysts attempt to increase the transparency of the use of colors, the change of scale, or use statistical methods. The practical benefit is not clear, but this variable level is an inseparable part of the work with big data. The default links are on social media, and many companies and organizations use Facebook, Twitter and LinkedIn as supplementary information source for realized analyses. The problem is how to integrate high variable data into default data structures in static databases and warehouses. Based on these sites, streams of data are created in big volumes.

The difficulties for a number of designers and analysts are caused by indecision as to which work needs to be implemented. A unique solution and method do not exist and implementation of the verified methodology can bring false results.

Rapid developments and new approaches and technologies also bring a new vision for active IT product implementation. It is therefore suitable to use open access for available solutions, to add needed dimensions and phases based on actual reality such as individual pages into a note pad. A book is created which describes IT product implementation. Please see Table 1.

The pages of this book create the needed background for active work with big data. Realized simulations and visualizations may demonstrate an increased respect for current conditions, preferences and available skills, but every IT user asks about the default benefits for everyday activities. Measurable performance is at the centre of interest. For information technology, this performance is evaluated via specified metrics. There are no unique solutions and various metrics may be defined as time required for development, the number of distributions that were sold, representation on the market, the number of customers, the amount of income, the amount of costs. These metrics are evaluated with numbers (hard metrics), but there are metrics without an evaluation with any number (soft metrics). Good examples of soft metrics are the reputation of the company or organization, customer confidence, or also employee motivation. The question is how to divide the responsibility of individual dimensions and phases into measured benefits or weakness via defined metrics. A suitable solution is achievable through a scale that is assigned by individual preferences in an implementation team. Simulation with visualization once again provides an opportunity to display various solutions in relation to a defined scale, although only one is actually correct. Based on inspiration in nature, there are pheromones. For big data, these pheromones will indicate the best combination for dimensions and phases for suitable IT product development.

The practical use of this approach is simple. The designers and analysts select from the available dimensions and phases which are evaluated by their impact on the current reality and known

Table 1. Specification of individual dimensions and phases for IT product development in the dynamic changes of a global information society

Dimensions (With Defined Attributes and Methods)	IT Product Development		Phases
Function and Processes	Actual impact (pheromone) on reality for selected dimensions	Actual impact (pheromone) on selected phases	Communication
Data and Information			Disposition
Documentation			Evaluation
Hardware			Feasibility Study
Software			Coding
Management			Concept
Methodology and Methods			Construction iterations
Organizational issue			Implementation
Social Issue			Inception
Time and Level of Abstraction			Integration
Consultants and User Participation			Operations and Maintenance
Learning			Production
Data quality			Requirement gathering
End User Requirements			Retirement
Employee Approach			Software Design
Company Psychology			Sprint
Designed Concepts			Sprint Planning
Integration and Customization			System Analysis
Measurement and Evaluation			Testing
Simulations and Visualization			Transition

preferences. There is, for example, a project on upgrading software for communication support in a company. Based on the assigned impacts (pheromones), a book is created with the needed dimensions and phases for such an actualization. The suitable dimensions are:

- Consultant and user participation with learning.
- Data and information.
- Data quality.
- Documentation.
- Organizational issue.
- Time and level of abstraction.

And the suitable phases are:

- Coding.
- Communication.
- Evaluation.
- Integration.
- Requirement gathering.
- Sprint.
- Testing.
- Transition (release).

This approach combines various methodologies, methods and advice based on experience from IT product development. The benefit is flexibility with regards to the time and quality

request. The weakness of this approach can be ignorance of one of the important dimensions or phases for an active solution. In this situation, iteration helps add a needed dimension and phases. In other cases, additional dimensions and phases are selected based on current needs and recommendations. There is also the possibility of enabling more swarm intelligence based on current research (evolutionary algorithms as a dynamic complex system, differential evolution for ensemble strategies, decomposition and cooperative co-evolution technique, high exploration particle swarm optimization or a new efficient evolutionary multi-criterion optimization algorithm).

FUTURE RESEARCH DIRECTIONS

The multidimensional approach helps with sustainable development in all countries. The basic recommendation is to work with big data in context based on the multidimensional approach. The big data concept works with many terms and relations:

- Information technology.

(implementation in industry, realized analyses and statistics, transforming running processes, data balance, visualization, databases, applications and tools, simulations, and security)

- Actual conditions in firms, organizations, and individuals.

(to find the greatest benefit, to reduce existing risks, to respect cost reduction, to set optimal scales by preferences, to select suitable software)

- Urgent requests of the global society with preferences of individuals.

(to respond quickly to actual reality, to provide optimum quality, and other global challenges for technological regulation or unemployment)

The multidimensional approach gives rise to various partial views on big data such as:

- Information technology.
- Implemented software in firms, organizations and with individuals.
- Tools for better perception of context.
- Catalyst of people's reactions to ongoing changes.
- Protection against unwanted changes.

For IT development, the actual selection of needed dimensions and phases is based on the current impact (pheromone) for the given software product. This approach also has positive and negative aspects as with other methods and methodology. There is not unfortunately a unique solution for IT product development for all cases. The above-mentioned design of the method for IT product development provides higher flexibility and an opportunity to combine various views of a suitable solution. There is also space of course for mistakes and misunderstanding based on a lack of optimal skills and sources. Prevention of these weaknesses is an issue for further development and closer inspiration in swarm intelligence. It is suggestion for further research in this matter from global view.

CONCLUSION

A global information society brings both a number of benefits and weaknesses. The revolutionary possibilities are caused by information technology entering into everyday activities in various fields of human activities. Positive and negative experiences and requests are therefore well-known and collective experience and thinking have an important place in decision-making. In this reality, software development has major pressure from competitions, costumer requirements, contractors and also the development of new innovations and information technologies.

The suitable recommendation is to implement a multidimensional view on big data with all partial views like information technology, implemented software, tools for better perception of context, catalyst of people's reactions to ongoing changes, and protection against unwanted changes. The multidimensional view is a default part of a data warehouse, and this view also holds an optimal place in big data. Such an approach works with big data along three levels (internal matter as the basic level, external matter as an advanced level, and uncertainty matter as a variable level). These levels support benefits from information technology for the sustainable development of the global information society.

A considerable effort is needed for the definition of the order of this complexity. The basis for an active solution is a multidimensional approach, object access and inspiration from nature with regards to swarm intelligence. Based on pheromones, ants use this way for road marking, referring to the way or warning of danger. These kinds of activities are important for IT product development such as specification of the road for an adopted methodology, definition of a reference for needed dimensions and phases for IT development, and also as a warning against omissions and mistakes. The current impact is consequently defined as a pheromone for IT product development. The same impact is introduced for well-known dimensions and phases.

REFERENCES

- Acharjya, D. P., Dehuri, S., & Sanyal, S. (Eds.). (2015). *Computational Intelligence for Big Data Analysis: Frontier Advances and Applications*. Springer.
- Barbucha, D., Nguyen, N. T., & Batubara, J. (Eds.). (2015). *New Trends in Intelligent Information and Database Systems*. Springer.
- Big Data Science on Twitter. (2015). Retrieved from twitter.com
- Clark, D. (2014). *Data Visualization Is The Future - Here's Why*. Retrieved from <http://www.forbes.com/sites/dorieclark/2014/03/10/data-visualization-is-the-future-heres-why/>
- Concessao, R. (2016). How to tell a Story from Big Data? *CreateSpace Independent Publishing Platform*.
- Covington, D. (2016). Analytics: Data Science, Data Analysis and Predictive Analytics for Business. *CreateSpace Independent Publishing Platform*.
- Crawley, E., Cameron, B., & Selva, D. (2015). *System Architecture: Strategy and Product Development for Complex Systems*. Prentice Hall.
- Dzemyda, G., Kurasova, O., Zilinskas, J. (2014). *Multidimensional Data Visualization: Methods and Applications*. Springer.
- Gartner Newsroom. (2015). Retrieved from <http://www.gartner.com/newsroom/>
- Granville, V. (2013). *Big data set - 3.5 billion web pages - made available for all of us*. Retrieved from <http://www.bigdatanews.com/profiles/blogs/big-data-set-3-5-billion-web-pages-made-available-for-all-of-us>
- Graville, V. (2015). *Most popular data science keywords on DSC*. Retrieved from http://www.datasciencecentral.com/forum/topics/most-popular-data-science-keywords-on-dsc?xg_source=activity
- Information Technology Market Reports. (2015). Retrieved from http://www.reportlinker.com/report/best/keywords/Information%20Technology?utm_source=adwords1&utm_medium=cpc&utm_campaign=High_Tech_And_Media&utm_adgroup=Information_Technology_All&gclid=CNne3s6NksYCFSLmwgo dACYAXg

Jain, L. C., Behera, H. S., Mandal, J. K., & Mohapatra, D. P. (Eds.). (2014). Computational Intelligence in Data Mining. *Proceedings of the International Conference on CIDM*.

Lewis, T. (2015). *Human Brain: Facts, Anatomy & Mapping Project*. Retrieved from <http://www.livescience.com/29365-human-brain.html>

Marr, B. (2016). *Big Data For Small Business For Dummies*. For Dummies.

Matei, S. A., Russell, M. G., & Bertino, E. (Eds.). (2015). Transparency in Social Media: Tools, Methods and Algorithms for Mediating Online Interactions. Springer. doi:10.1007/978-3-319-18552-1

Mayer-Schönberger, V., Cukier, K. (2014). *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Eamon Dolan/Mariner Books.

Mohanty, H., Bhuyan, P., & Chenthati, D. (Eds.). (2015). *Big Data: A Primer*. Springer.

Rischar, J. F. (2002). *High noon: twenty global problems, twenty years to solve them*. New York: Basic Books.

Simon, P. (2014). *The Visual Organization: Data Visualization, Big Data, and the Quest for Better Decisions*. John Wiley Sons Inc.

Top 15 Most Popular Social Networking Sites. (2015). Retrieved from <http://www.ebizmba.com/articles/>

Top 27 Free Data Analysis Software. (2015). Retrieved from <http://www.predictiveanalyticstoday.com/top-data-analysis-software/>

World Development Report 2015: Mind, Society, and Behavior. (2015). Washington, DC: International Bank for Reconstruction and Development / The World Bank.

Zobaa, A. F., & Vaccaro, A. (2015). Computational Intelligence Applications in Smart Grids: Enabling Methodologies for Proactive and Self-Organizing Power Systems. Imperial College Press.

KEY TERMS AND DEFINITIONS

Big Data: The broad spectrum of data that have appeared throughout the development of the global information society.

Competition: A test of skill or ability of offered goods, services, and other realized activities in the global information society.

Global Society: A society where the creation and integration of data via information and knowledge is a significant process in all fields of human activities.

Information: Information is perceived data about the properties of existing objects. Information reduces uncertainty of the system.

Information Technology: Information technology is a scientific discipline that seeks the optimal solution for compiling, creating, linking and improving available technologies.

Sustainable Development: Economic development of the global information society that is conducted with respect to natural resources and global challenges like technological regulation, intellectual property protection, and unemployment.

Swarm Intelligence: Focus is dedicated to intelligent behavior swarms, i.e. large numbers of individuals. Nature is an inspiration for the designed method in many fields such as chemistry, engineering, environmental modeling, financial services and information technology.

Financing Micro, Small, and Medium Enterprises in Indian Industry

Shromona Ganguly

Indian Institute of Management Calcutta, India & Reserve Bank of India, India

INTRODUCTION

During the last five decades, the micro, small and medium enterprises (MSME) in the Indian economy has emerged as a dynamic and vibrant segment having a significant contribution towards employment generation and entrepreneurship formation. In addition, these enterprises play a crucial role in making the growth of the economy more inclusive by reducing regional disparity. During 2012-17, the MSME sector in India accounted for almost 38 per cent of GDP. The sector has grown approximately at a rate of 28.02 per cent and 26.42 per cent in terms of estimated number of enterprises and employment respectively during the period 2001-02 to 2006-07. Within the MSME sector, services constitute 80 per cent and manufacturing constitutes the remaining 20 per cent (Annual Report, Ministry of MSME, 2014-15).

Despite being an important contributor towards the national output and employment of the economy, the issue of timely and adequate availability of credit has been recognised as one of the major impediments to the growth of the sector in recent times. Recognising the problem of credit constraint faced by MSMEs, a slew of initiatives has been undertaken to make the financial system architecture of the sector more conducive for growth. However, the issue of credit availability to small firms became more important after the recent financial crisis which caused further slowdown in credit flow from formal banking system to the small firms.

Across the world, countries have followed different strategies to solve the issue of information

asymmetry and credit rationing in the small value loan market. However, adoption of these strategies primarily depends on the financial and legal system of the country. Typically, countries with advanced capital markets and strong legal and enforcement system use credit scoring, factoring, and leasing as alternative techniques to solve the issue of information asymmetry in small business loan segment (Berger and Udell 2006). However, due to weak enforcement, these strategies have limited effectiveness in developing countries. As a result, small firms in developing countries face larger credit constraints compared with countries which have well-developed financial systems. The existing literature on small firm financing discusses the issues and challenges from a cross-country perspective. There is a lack of detailed research on implications of country specific factors and financial system on small firm financing. The present essay aims to fill this gap by analysing the extent and nature of credit constraint faced by small firms in India.

The existing literature on small firm financing often analyses issues using the cross-country data. There are very few studies to understand the problems or complexities associated with small firm financing in the context of India. Further, deployment of technology assumes a key role in solving some of the structural issues embedded in the financial system of developing countries hindering credit flow to small businesses. At the same time, use of technology in the lending of small businesses raises important issues related to regulatory and policy responses to the online credit markets. So far, there is no comprehensive

study of these issues in the literature. The present chapter aims to fill this gap. The objective of the present chapter is to analyse the financial system of India in the context of lending to small firms and to what extent use of technology can solve the problem of lack of access to credit to small firms in India.

The chapter is organised as follows: The next section (Section 2) provides a brief of theoretical foundation and literature review on the financing of small firms around the world. Section 2 is further divided into two sub-sections focusing on the existing lending technologies in the context of small firm financing around the world and the how infusion of technology in financial market may change the landscape. Section 3 delineates the present status of credit flow to the MSME sector from the formal banking system. The section also discusses the issues and challenges involved in providing finance to the small businesses through the banking system in India. Section 4 describes in detail the present state of technology in the small business credit market in India. Section 5 concludes the study with observations and policy implications.

BACKGROUND AND LITERATURE REVIEW

Worldwide, access to finance has been a major challenge for small enterprises (Beck and Demircuc-Kunt 2006). The information asymmetry in the small business lending market can take the form of i) ex-ante screening of projects (adverse selection) or ii) ex-post monitoring of loan contracts (moral hazard). In the presence of information asymmetry, there exists an interest rate which maximizes expected return to the bank and beyond which, increasing the interest rate as a response to meeting credit demand could attract more risky borrowers resulting in a lower profit for the bank. As a result, banks prefer to ration credit in such a situation.

Lending Technologies in the Context of Small Firm Financing: Relationship-Based vs. Arm's-Length Systems

Worldwide, the financial systems can be classified into two broad categories; i) bank-based system as observed in the case of Germany and Japan ii) capital market based system as the one in UK and US. Typically financial institutions in an arm's length system rely on hard information in the form of financial system statements for loan disbursement as well as loan pricing. On the other hand, in a relationship-based system, the length of the relationship between a financial institution and borrower determines key terms of loan contract (Peterson and Rajan 1995).

Suggesting that the causality runs from finance to growth, Rajan and Zingles (2001) establishes the theoretical underpinning of the interdependence between the financial system and industrial structure. According to the study, the relative growth of industries is largely influenced by financial system structure of the country. Industries which require a higher amount of external financing will thrive in countries with developed financial systems. Similarly, countries with advanced financial system cater to the needs of young and innovative firms well which at a later stage, fuels the wave of creative destruction as alluded by Schumpeter (1911). Also, the study highlights the role of collateral in mitigating information asymmetry problems in developing countries. Accordingly, as the capital market of a country develops, equity financed industries tend to grow faster and use less fixed capital. This theory also establishes the link between equity market development and innovation. As innovation is typically an intangible asset, the growth of innovation-led industries requires the development of equity market to efficiently process soft assets like innovation. According to Wurgler (2000), countries with well-developed financial systems will invest more in growing industries while countries with the underdeveloped financial system will focus on declining industries.

The empirical literature on financial institution size and small firm financing often concludes that large financial institutions use sophisticated lending technologies and more often deals with the small business segment at an arm's length. In contrast, small and niche banks use relationship-based lending for catering the small size loan segment. As a result, these institutions are more efficient in dealing with the issue of information asymmetry prevalent in small business financing (Berger and Udell 2006).

Government-owned financial institutions are typically large and thus, following the argument presented above, these institutions are not best suited to cater the financial needs of small businesses. However, in many countries, large development banks owned by the government are entrusted with the mandate of credit disbursement to small firms. Often these are subsidised loans in the form of directed lending to specific sectors. In literature, there are mixed evidence on the efficacy of development banking model in credit disbursement to small businesses. Often, these institutions lack market discipline and cater the segment of small firms which are not creditworthy. In most cases, empirical evidence suggests that state ownership of banks is associated with more creditworthy firms getting excluded from the formal banking system (Beck and Levine 2004, Berger *et al* 2004). To further illustrate the functioning of development banks, Carvalho (2014) establishes that in Brazil, firms eligible for government subsidised loans expand employment in politically favoured regions. Further, these persistent expansions take place just before competitive elections, and are associated with lower future employment growth by firms in other regions. However, in many developing countries, state-owned development banks have played a crucial role in disbursing credit to industry at cheap cost and timely manner during the early stages of industrialization (Nayyar 2015).

However, Torre *et al* (2010) challenges this conventional wisdom about the relative efficiency of small and niche banks in small firm lending.

According to them, most banks in emerging economies look at the segment of small business as strategically important and deploy multiple techniques to extend loans to small business. Using the survey data of a balanced mix of developed and developing economies, the paper shows that the intensification of bank involvement with SMEs in various emerging markets is neither led by small or niche banks nor is it highly dependent on relationship lending. Moreover, large banks have relative efficiency in offering a wide range of products and services to micro and small businesses, due to increased use of technology and scale economies that these banks enjoy.

The Financial Crisis, “Perfect Storm,” and the Role of Technology in the Small Business Credit Market

The existing scholarly work on small firm financing underscores the adverse impact of the financial as well as debt crisis on the availability of credit to small firms (Jimenez and Ongena 2012, Iyer *et al* 2014, Udell *et al* 2016). In most of the studies, it was found that small firms are hit directly by reduction of loan supply by the financial institutions due to the following reasons (Mills and McCarthy 2014).

Increased Risk Perception

During the financial crisis, small businesses were hit the most in their sale and profit. In the US, a major portion of the small firms started facing increased problem of accessing bank credit as they became less credit worthy with reduced income and sales growth and increased cyclicity of their business operations.

Collateral

In the US, the sources of collateral were squeezed during the financial crisis, which in turn reduced the supply of credit to small businesses.

Rise-Averseness and Regulatory Overhang

During the financial crisis, banking regulators in the US responded by stipulating higher capital requirements for the banks. This, in turn, has reduced the availability of credit to small firms since the banks perceive these firms as high-risk and less creditworthy. Also, the impact of such regulatory overhang has partially neutralised the otherwise positive impact of the low interest rate and unconventional monetary policy pursued by major economies after the crisis.

However, one notable trend observed during the period of crisis as well as recovery is the emergence of a broad range of alternative lending technologies in the small business credit market. Typically, the new technologies have been evolving around receivable purchases and innovations in specialised lending by verticals. In addition, the online lending technologies have sprung up in major economies around the world. The google searches of “term loan” clearly indicate an upward trend during the last few years (Mills and McCarthy 2014). In most cases, the online players are acting as catalyst in infusing new technology in banking industry.

CREDIT FLOW TO SMALL BUSINESS IN INDIA

The Role of Commercial Banks

India’s financial system is known as bank-based. The banking system in India is predominated by public sector banks accounting for more than 70 per cent of total assets of the banking system. These banks played a critical role in mobilising financial savings and fuelling the industrial growth of the country during the early phases of industrialization in India. However, in India, the period before liberalization was characterised by significant command and control by the government in various sectors. In the case of the banking

sector, this took the form of high pre-emptions in the form of high reserve requirements. The policy of high reserve requirements often resulted in misallocation and mispricing of resources. Due to the adverse effects of excessive control of the financial sector in India, the pre-liberalization phase is often termed as the era of “financial repression” (Mohan 2004).

Though steps were taken from the mid-80’s to liberalise the credit delivery mechanism, this gained momentum only during the 90’s. As a part of the overall mandate of liberalization, the statutory liquidity ratio (SLR) was brought down from 38.5 per cent in 1992 to 25 per cent in 1997. Also, selective credit control was phased out in a gradual manner. To impose fiscal discipline as well as provide greater autonomy to monetary policy, automatic monetization of budget deficit was also prohibited.

The bank credit to GDP ratio increased in a slow but steady manner since the initiation of reform. The same trend is observable in the case of the sectoral credit to GDP ratio for the industry (manufacturing). However, it is noteworthy that despite various policy initiatives taken during the early 90’s, the growth in credit to GDP ratio became more pronounced only after 2000-01 (Figure 1). Despite a reduction in the statutory requirement ratio, banks in India continued investing in government securities as these assets were considered a safer haven for investment as compared to extension of credit to other sectors of the economy.

During the last two decades, industry accounts for almost 40 per cent of total bank credit in India. An interesting observation in this regard is that, since 1999, the credit flow from commercial banks to the industry has become a more long term in nature. Most of the long term borrowing is collateralized in nature. Hence, this trend, in some sense, indicates that within the total credit flow to industry, unsecured small account borrowings, typically by the smaller firms, have reduced over the years (Table 1).

Figure 1. Credit to GDP

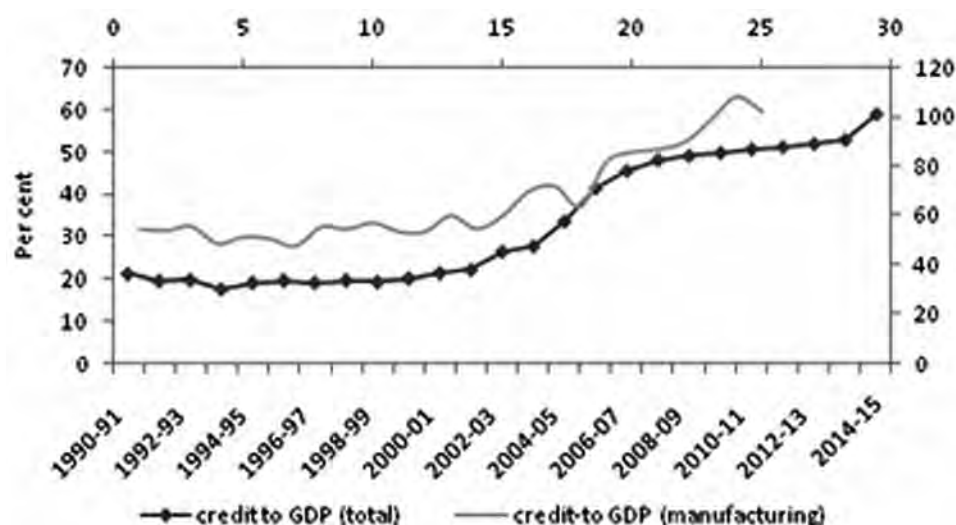


Table 1. Type of account-wise outstanding credit of SCBs to industry (Rs million)

Year	Cash Credit	Overdraft	Demand Loans	Medium-Term Loans	Long-Term Loans	Total Credit to Industry	Total Bank Credit	Credit to Industry as Percentage of Total Credit	Share of Long-Term Loans in Total Credit to Industry
1980*	54,035	6,604	3,126		13,900	1,02,362	2,13,116	48.0	13.6
1985*	1,00,015	14,319	6,498		36,690	2,06,589	4,99,946	41.3	17.8
1990	2,25,169	35,309	17,034	15,858	69,466	5,08,455	10,43,119	48.7	13.7
1993	3,62,067	51,975	35,347	29,756	92,037	7,89,643	16,24,673	48.6	11.7
1996	5,12,787	75,038	97,348	66,200	1,52,731	12,22,595	25,46,921	48.0	12.5
1999	5,98,911	88,069	2,85,328	1,06,603	3,21,509	18,79,469	38,24,250	49.1	17.1
2002	7,01,656	96,878	3,52,934	2,23,138	7,49,100	27,16,255	65,59,931	41.4	27.6
2006	10,36,074	2,23,669	5,48,881	5,80,180	23,02,017	56,62,160	1,51,38,421	37.4	40.7
2008	18,32,043	3,69,985	8,21,815	12,37,281	37,15,787	92,85,358	2,41,70,065	38.4	40.0
2011	34,60,241	5,67,323	16,26,767	23,98,908	62,99,509	1,61,20,468	4,07,56,470	39.6	39.1
2012	46,04,315	6,30,802	15,04,072	25,36,168	78,32,971	1,94,19,646	4,80,32,669	40.4	40.3
2013	56,61,551	7,92,240	21,38,127	27,04,239	93,73,503	231,59,883	552,53,170	41.9	40.5
2014	66,60,380	9,46,887	15,54,544	32,93,926	108,50,402	261,62,579	628,20,824	41.6	41.5

Source: Basic Statistical Return (BSR) of scheduled commercial banks in India, various issues

*: End-June. For all other years, data pertains to end-March

The aforementioned phenomenon becomes more visible from the size-wise composition of total industrial credit from the commercial banks. During the 80's as well as 90's, micro and small

enterprises accounted for, on an average, 25 per cent of total credit to the industry from the banks. This proportion has reduced to 15 per cent substantially after 2004-05. At the same time, large and

medium industries account for almost 85 per cent of total bank credit to industry during 2014-15, much higher than the proportion observed during 2003-04 (78 per cent) (Table 2).

Not only the absolute amount of bank credit flow to the small industries has come down during recent times, but the small industries located in rural areas also witnessed the sharpest fall in bank credit during recent times. Analysing the below table (Table 3), we get the following trends:

1. Out of the total bank credit outstanding in small borrowal accounts, the share of industry has declined from 16 per cent in 1996 to 4.2 per cent in 2006 and further, to 1.6 per cent in 2015.

2. While this decline of share in the industry in small borrowal account is uniform across geographical locations, the same is highest in the case of rural small borrowal accounts.
3. Within the small borrowal accounts (industry), the share of rural accounts has declined over the years (43 per cent in 1996 to 32 per cent in 2015).

In India, the public sector banks account for the largest commercial bank branch network of the country. These banks have a crucial role to play in making the banking system of India more inclusive. Apart from being the lynchpin of directed credit flow, public sector banks also play a pivotal role in furthering financial inclusion in hitherto underdeveloped part of the country. As per

Table 2. Size-wise classification of bank credit to industry (Rs crore)

Year	Micro and SMALL	Medium and Large	Industry	Total Bank Credit	Share of Industry in Total Bank Credit	Share of Micro and Small in Total Bank Credit to Industry	Share of Large and Medium in total Bank Credit to Industry
1981-90	8,451	21,152	29,693	59,056	50.28	28.46	71.24
1994-95	27,638	74,672	1,02,310	2,11,560	48.36	27.01	72.99
1999-2000	52,814	1,47,319	2,00,133	4,35,958	45.91	26.39	73.61
2000-2001	56,002	1,62,837	2,18,839	5,11,434	42.79	25.59	74.41
2001-02	57,199	1,72,324	2,29,523	5,89,723	38.92	24.92	75.08
2002-03	60,394	2,35,168	2,95,562	7,29,215	40.53	20.43	79.57
2003-04	65,855	2,47,210	3,13,065	8,40,785	37.23	21.04	78.96
2004-05	74,588	3,52,304	4,26,892	11,00,428	38.79	17.47	82.53
2005-06	91,212	4,59,232	5,50,444	15,07,077	36.52	16.57	83.43
2006-07	1,17,910	5,79,429	6,97,339	19,31,189	36.11	16.91	83.09
2007-08	1,32,698	7,25,646	8,58,344	23,61,914	36.34	15.46	84.54
2008-09	1,68,997	8,85,393	10,54,390	27,75,549	37.99	16.03	83.97
2009-10	2,06,401	11,05,051	13,11,452	32,44,788	40.42	15.74	84.26
2010-11	2,10,206	13,94,369	16,04,576	36,67,354	43.75	13.10	86.90
2011-12	2,36,657	17,00,669	19,37,325	42,89,745	45.16	12.22	87.78
2012-13	2,84,348	19,45,831	22,30,179	48,69,563	45.80	12.75	87.25
2013-14	3,48,194	21,68,290	25,16,483	55,29,602	45.51	13.84	86.16
2014-15	3,80,028	22,77,599	26,57,627	60,02,952	44.27	14.30	85.70

Source: Bhattacharya (2013), and Sectoral Deployment of Bank Credit, RBI

Table 3. Trend in small borrowal accounts (Rs million)

Year	Small Borrowal Accounts (Industry)				Total Bank Credit to Small Borrowal Accounts			
	Rural	Semi-Urban	Urban/Metropolitan	Total	Rural	Semi-Urban	Urban/Metropolitan	Total
1996	1,098	745	714	2,557	15,674	10,672	9,907	15,674
2000	2,166	2,338	4,202	8,705	34,615	29,711	38,419	1,02,745
2006	2,603	2,483	3,236	8,322	69,064	55,506	75,310	1,99,880
2010	1,744	2,067	9,158	12,968	1,12,960	94,463	1,42,442	3,49,865
2015	23,695	17,941	32,241	73,876	21,90,976	18,32,193	12,43,742	52,66,911

Source: Basic Statistical Returns of Scheduled Commercial Banks, RBI

the bank-group wise data on credit flow to micro and small enterprises (MSE) sector, more than 75 per cent of total credit to MSE comes from the public sector banks, followed by domestic private banks (21 per cent) and foreign banks (4 per cent).

Issues and Challenges of Providing Credit to Small Enterprises Through the Commercial Banking Sector

The above section establishes the fact that there is a growing retrogression within the total credit flow to micro and small enterprises sector in India (Bhattacharya, 2013). Allen et al (2012), analysing the evolving financing channels of firms in India, concluded that alternative financing channel (non-bank, non-financial sources) is

quite significant for Indian firms in general, and small firms in particular. The finding is in sync with our finding in the previous section, which shows that total bank credit to the industry has grown in favour of large firms over the years. As pointed out by Bhattacharya (2013), this is in sharp contrast to what was expected after deregulation of interest rate in the Indian banking sector. Trends in bank credit to small businesses do not indicate any favourable impact resulting from interest rate deregulation, rather it indicates a higher amount of credit rationing in the small value loan segment in India. The preponderance of informal sources in the financing of small, unregistered manufacturing firms in India has been a reason for slower industrial growth of the country (Table 4).

Table 4. Financing pattern of unorganised manufacturing in India

Lending Institutions	2010-11			2005-06			2010-11		
	OAE	MHE	NHE	OAE	MHE	NHE	OAE	MHE	NHE
Term lending institutions, banks and societies	60.4	69.7	70.2	47.7	57.9	58.6	58.9	69.8	72.5
Other institutional agencies	2.3	2.6	4.7	4.3	5.7	7.1	1.6	2.9	1
Money lenders	21	11.3	9.5	16.6	12.9	10	25.3	16.8	13.3
Business partners	0.2	2.3	2.4	2.2	2.1	7.2	0.1	0.3	2.7
Suppliers/contractors	2.6	1.9	2.9	2	2	2.2	2.5	2.8	1
Friends/relatives	12.1	10.8	7.4	21	10.6	8.2	9.6	6.3	7
Others	1.4	1.4	3.1	6.2	8.9	6.7	2.1	1	2.6
Total	100	100	100	100	100	100	100	100	100

Note: OAE: Own Account Enterprises, MHE: Mixed-household enterprises, NHE: Non-household enterprises

Source: Sen and Raj (2015)

As per the priority sector lending target mandated to commercial banks in India, banks are required to lend 40 per cent of their total lending towards specified sectors like agriculture and allied activities and micro and small enterprises. However, a number of public sector banks failed to meet the priority sector lending target as on March 2013. In addition, within the total lending to priority sectors, a trend of credit concentration is discernable towards the large accounts (RBI 2012).

The growing proportion of bank credit to industry in favour of large firms needs to be contrasted with the capital market development in India. Small businesses often lack the required financial track record to tap funds from the capital market. Hence these firms are more dependent on bank credit. As against this, our analysis shows that despite robust overall credit growth, credit to the MSE sector has come down substantially during the last decade. As a result, a growing proportion of small firms is moving to the informal markets as well as inter-firm credit market for financing their operation. As shown below, inter-corporate loan market has grown across all size categories of firms in recent times. During 1990-2010, inter-corporate loan as a proportion of total borrowing is most significant in the case of small firms (Figure 2).

Another seemingly puzzling fact is that despite the high-risk perception of the sector, MSE

sector's contribution remained stable in the non-performing assets (NPAs) of the commercial banks¹. As against this, large firms account for a growing proportion of total NPAs of banks in India. This trend suggests that banks are reluctant to extend credit to this sector due to ex-ante risk perception of the sector. Lack of formal financial intermediation of the sector is structural in nature and requires a re-look at the risk management and project appraisal practices followed in the Indian banking sector. At this juncture, an important policy question remains whether India needs the establishment of a special development bank catering issues related to industrial finance.

ISSUES AND CHALLENGES WITH SMALL FIRM FINANCING IN INDIA AND HOW TECHNOLOGY MAY CHANGE THE GAME

A significant proportion of entrepreneurs is self-financed in India (National Knowledge Commission 2008). Within the self-financed start-ups, a closer examination revealed that nearly two third of entrepreneurs have either invested their own savings or borrowed from an already existing family business. Banks constitute the next important source of finance for entrepreneurs in India and contribute almost 22 per cent of early stage

Figure 2. Inter-corporate borrowing in Indian manufacturing

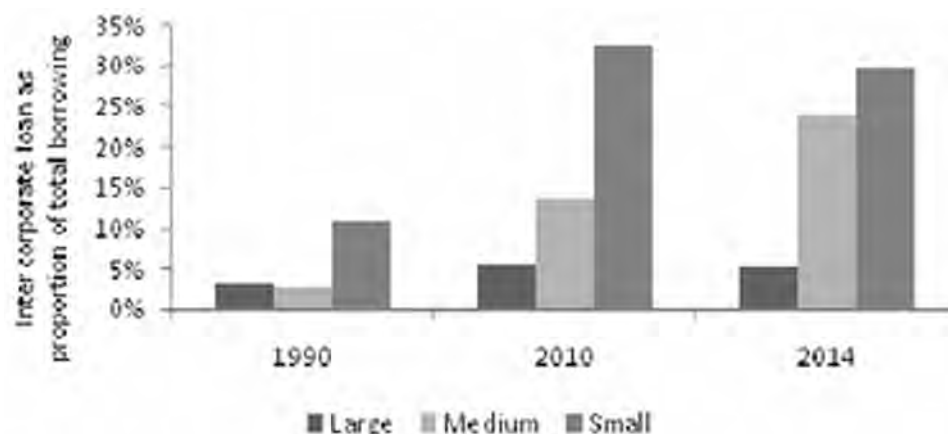


Figure 3. Source of finance for start-up in India



financing of such enterprises. The role of venture capital and angel investors is rather limited in this respect; these together contribute 9 per cent of early stage financing in India (Figure 3).

The problem of financing the small firms has been catching the attention of policymakers in India for long. The need for creating the appropriate set of institutions for financing of small firms in India has been articulated by the committee of financial sector reforms (2008). However, it was also felt that along with creation of new financial institutions, the use of technology may change the financial landscape of small businesses in India. Since the micro and small business in India is largely affected by delayed payments by their corporate buyers, the Reserve Bank of India (RBI) has been contemplating the set up of an electronic trade receivable platform in India, largely in line with the NAFIN model of Mexican Development Bank (RBI, 2014). It is expected that with the infusion of technology, the credit flow to the micro businesses in India without any credit history will improve in the near future.

CONCLUSION

The chapter analyses in detail the role played by formal and informal financial institutions in India

in providing credit to the micro and small firms in Indian industry. The discussions presented in the previous sections suggest three key trends discernible in this respect:

- There is a growing pattern of credit retrogression observable in the case of bank credit flow to industry. Over the years, large firms account for an increasing amount of total credit flow to the industry while the share of small firms has come down.
- The informal financial sector still accounts for a significant amount of total financing of the sector. This trend has not changed much despite various initiatives were undertaken by the Reserve Bank to expand the outreach of the banking system.
- Self-financing predominates the early stage financing pattern for start-ups in India. Banks, though constitute the second largest source of early stage debt finance, have limited outreach for rural entrepreneurs. The alternative financing channels, which include venture capital and private equity funds, constitute less than 10 per cent of total early stage financing in India. This figure is much lower compared with the advanced economies.

FUTURE RESEARCH DIRECTIONS

In recent times, the policy environment for the micro and small enterprises sector in Indian industry has kept its thrust on increasing the credit flow from the formal banking system to the MSEs. This includes the introduction of new innovative techniques like credit scoring, and use of technology to deepen the market for factoring in India. Apart from use of such new techniques, a new bank with active government involvement (Mudra Bank) has been created to monitor and provide credit to rural entrepreneurs. However, use of new lending technology to the small business in India also requires an overhaul of the regulatory approaches towards the new technologies introduced in the credit market. For example, during the last few years the peer to peer lending market has been taking its root in India. The discussion paper released by the Reserve Bank (RBI, 2016) articulates the need for the overall monitoring of the sector from the perspective of financial sector stability, while recognising the benefit it provides to various stakeholders. In this regard, the policy challenge is to achieve the optimal amount of oversight for the new technologies evolving in the small business credit market which will prevent abuse or fraud without preventing innovation.

REFERENCES

- Allen, F., Chakrabarti, R., De, S., Qian, M., & Qian, M. (2012). Financing firms in India. *Journal of Financial Intermediation*, 21(3), 409–445. doi:10.1016/j.jfi.2012.01.003
- Annual Report. (2014-15). Ministry of Ministry of Micro, Small and Medium Enterprises, Government of India. Retrieved from http://msme.gov.in/WriteReadData/DocumentFile/MSME%20ANNUAL%20REPORT%202014-15_English.pdf
- Basic Statistical Returns of Scheduled Commercial Banks in India. (n.d.). Retrieved from <https://rbi.org.in/Scripts/AnnualPublications.aspx?head=Basic%20Statistical%20Returns>
- Beck, T., & Demircuc-Kunt, A. (2006). Small and medium-size enterprises: Access to finance as a growth constraint. *Journal of Banking & Finance*, 30(11), 2931–2943. doi:10.1016/j.jbankfin.2006.05.009
- Beck, T., & Levine, R. (2004). Stock markets, banks, and growth: Panel evidence. *Journal of Banking & Finance*, 28(3), 423–442. doi:10.1016/S0378-4266(02)00408-9
- Berger, A. N., Demircuc-Kunt, A., Levine, R., & Haubrich, J. G. (2004). Bank concentration and competition: An evolution in the making. *Journal of Money, Credit and Banking*, 36(3b), 433–451. doi:10.1353/mcb.2004.0040
- Berger, A. N., & Udell, G. F. (2006). A more complete conceptual framework for SME finance. *Journal of Banking & Finance*, 30(11), 2945–2966. doi:10.1016/j.jbankfin.2006.05.008
- Bhattacharya, A. (2013). Credit Retrogression in Micro and Small Enterprises Sector. *Economic & Political weekly*, 48(35), 105–114.
- Carbó-Valverde, S., Rodríguez-Fernández, F., & Udell, G. F. (2016). Trade Credit, the Financial Crisis, and SME Access to Finance. *Journal of Money, Credit and Banking*, 48(1), 113–143. doi:10.1111/jmcb.12292
- Carvalho, D. (2014). The Real Effects of Government-Owned Banks: Evidence from an Emerging Market. *The Journal of Finance*, 69(2), 577–609. doi:10.1111/jofi.12130
- De la Torre, A., Pería, M. S. M., & Schmukler, S. L. (2010). Bank involvement with SMEs: Beyond relationship lending. *Journal of Banking & Finance*, 34(9), 2280–2293. doi:10.1016/j.jbankfin.2010.02.014
- Iyer, R., Peydró, J. L., da-Rocha-Lopes, S., & Schoar, A. (2014). Interbank liquidity crunch and the firm credit crunch: Evidence from the 2007–2009 crisis. *Review of Financial Studies*, 27(1), 347–372. doi:10.1093/rfs/hht056

- Jiménez, G., Ongena, S., Peydró, J.-L., & Saurina, J. (2012). Credit supply and monetary policy: Identifying the bank balance-sheet channel with loan applications. *The American Economic Review*, 102(5), 2301–2326. doi:10.1257/aer.102.5.2301
- Mills, K., & McCarthy, B. (2014). *The State of Small Business Lending* (No. 15-004, p. 6). Harvard Business School Working Paper.
- Mohan, R. (2005). Financial sector reforms in India: Policies and performance analysis. *Economic and Political Weekly*, 1106–1121.
- National Knowledge Commission. (2008). *Entrepreneurship in India*. Retrieved from http://knowledgecommissionarchive.nic.in/downloads/documents/NKC_Entrepreneurship.pdf
- Nayyar, D. (2015). Birth, Life and Death of Development Finance Institutions in India. *Economic and Political Weekly*, 50(33), 51–60.
- Petersen, M. A., & Rajan, R. G. (1995). The effect of credit market competition on lending relationships. *The Quarterly Journal of Economics*, 110(2), 407–443. doi:10.2307/2118445
- Raj, R. S., & Sen, K. (2015). Finance Constraints and Firm Transition in the Informal Sector: Evidence from Indian Manufacturing. *Oxford Development Studies*, 43(1), 123–143. doi:10.1080/13600818.2014.972352
- Rajan, R. G., & Zingales, L. (2001). Financial systems, industrial structure, and growth. *Oxford Review of Economic Policy*, 17(4), 467–482. doi:10.1093/oxrep/17.4.467
- RBI. (2012). *Report on the Trend and Progress of Banking in India*. Retrieved from <https://rbidocs.rbi.org.in/rdocs/Publications/PDFs/0TPB021112FLS.pdf>
- RBI. (2014). *Concept Paper on Trade Receivables and Credit Exchange for Financing of Micro, Small and Medium Enterprises*. Retrieved from https://rbi.org.in/scripts/bs_viewcontent.aspx?Id=2804
- RBI. (2016). *Consultation Paper on Peer to Peer Lending*. Retrieved from https://rbi.org.in/scripts/Bs_viewcontent.aspx?Id=3164
- Schumpeter, J. A. (1911). *The theory of economic development*. Statistical Table Relating to Banks in India, Reserve Bank of India. Retrieved from <https://rbi.org.in/Scripts/AnnualPublications.aspx?head=Statistical%20Tables%20Relating%20to%20Banks%20in%20India>
- Sectoral Deployment of Bank Credit. (n.d.). Reserve Bank of India. Retrieved from https://rbi.org.in/Scripts/Data_Sectoral_Deployment.aspx
- Wurgler, J. (2000). Financial markets and the allocation of capital. *Journal of Financial Economics*, 58(1), 187–214. doi:10.1016/S0304-405X(00)00070-2

KEY TERMS AND DEFINITIONS

Adverse Selection: The concept of adverse selection applies to a case where the bad customers apply for the service because they have better information about how much benefit can be extracted from the service.

Development Banks: Specialised banks created for subsidised directed lending to specific sectors.

Information Asymmetry: Information asymmetry deals with the decision of transaction in situations where one party has more information compared to the other.

Unconventional Monetary Policy: Unconventional monetary policies are used in a situation where the interest rate has reached near zero. These include quantitative easing, forward guidance and signalling.

ENDNOTE

- ¹ Source: Statistical Table Relating to Banks in India, various issues.

Software Development Process Standards for Very Small Companies

S

Rory V. O'Connor
Dublin City University, Ireland

INTRODUCTION

In recent times quality orientated process approaches and standards have matured and gained acceptance in many software development organizations. Standards emphasize communication and shared understanding more than anything. There are many potential benefits of using standards. In particular for small and very small companies, the benefits that certification can provide include: increased competitiveness, greater customer confidence and satisfaction, greater software product quality, increased sponsorship for process improvement, decreased development risk, facilitation of marketing, and higher potential to export. While good internal software management might help meet the first five claims; the last two can only be the benefits of using a widely recognized standard.

It is commonly agreed that very small software companies, implementing management procedures, and controls to appropriately administer their software development activity is a significant challenge (Laporte et al, 2015). For example, a software company operating in India may have a completely different set of operational problems when compared to a software company in Canada, Mexico or Ireland. Even within a single geographical area such as Ireland, the range of operational issues faced by a small local Irish-owned firm can be radically different to those affecting a multinational subsidiary. The fact that all companies are not the same raises important questions for those who develop software process

and process improvement models. To be widely adopted by the software industry, any process or process improvement model should be capable of handling the differences in the operational contexts of the companies making up that industry. But process improvement models, though highly publicized and marketed, are far from being extensively deployed and their influence in the software industry therefore remains more at a theoretical than practical level.

With this in mind, the standardization body ISO/IEC has recently published the ISO/IEC 29110 standard “Lifecycle profiles for Very Small Entities” with the overall objective being to assist and encourage very small software organization in assessing and improving their software. The purpose of this chapter is provide a primer on the ISO/IEC 29110 standard focusing on two main process areas of Project Management and Software Implementation. This chapter will start with an explanation of the rationale and justification for the development of this new standard, followed by an overview of its structure and explain how to deploy ISO/IEC 29110 in a typical very small software company.

BACKGROUND

This section will introduce the problem with standards and explain the specific case of very small entities, before presenting the ISO/IEC standard as a solution specifically designed to address these problems for very small companies.

Very Small Companies

The definition of “Small” and “Very Small” Entities is challengingly ambiguous, as there is no commonly accepted definition of the terms. The term “Very Small Entity” (VSE) had been defined by the ISO/IEC JTC1/SC7 Working Group 24 and subsequently adopted for use in the new ISO/IEC 29110 software process lifecycle standard as being “an entity (enterprise, organization, department or project) having up to 25 people” (Laporte et al, 2008).

A large majority of enterprises worldwide are VSEs. In Europe, for instance, as illustrated in Table 1, over 92% of enterprises are micro-enterprises. They have fewer than nine employees. Micro enterprises account for 70% to 90% of enterprises in OECD countries and about 57% in USA.

Typically VSEs are economically vulnerable as they are driven by cash flow and depend on project profits, so they need to perform the projects within budget. They tend to have low budgets which have many impacts, such as: lack of funds to perform corrective post delivery maintenance; few resources allocated for training; little or no budget to perform quality assurance activities; no budget for software reuse processes; low budget to respond to risks; and limited budget to perform Process Improvement and / or obtain a certification/assessment. Typically the VSE’s product has a single customer, where the customer is in charge of the management of the system and the software integration, installation and operation. It is normal practice for the customer not to

define quantitative quality requirements and for customer satisfaction to depend on the fulfillment of specific requirements that may change during the project. A close relationship between all involved project members including the customer shows that software development in small and very small companies is strongly human-oriented and communication between them is important.

The internal business process of VSEs is usually focused on developing custom software systems, where the software product is elaborated progressively and which typically does not have strong relationship with other projects. Typically most management processes (such as human resource and infrastructure management) are performed through informal mechanisms, with the majority of communication, decision-making and problem resolution being performed face-to-face.

Problems With Standards

Although commercial SPI models have been highly publicized, they are not being widely adopted and their influence in the software industry therefore remains more at a theoretical than practical level (O’Connor and Coleman, 2009). For example, in the case of Capability Maturity Model Integration (CMMI), an Australian study found that small organizations considered that adopting CMMI would be infeasible (Staples et al, 2007) and an Irish study found significant resistance due to negative perceptions surrounding levels of bureaucracy and required documentation (Coleman and O’Connor, 2006). Further investigation of the CMMI by Staples and Niazi (2006) discovered, after systematically reviewing 600 papers, that there has been little published evidence about those organizations who have decided not to adopt CMMI.

There is evidence that the majority of small and very small software organizations are not adopting existing standards / proven best practice models because they perceive the standards as being developed by large organizations and orientated towards large organizations, thus provoking the

Table 1. Size of enterprises in Europe (Moll, 2013)

Type	Number of Employees	Annual Turnover	No. of Enterprises (% of Overall)
Micro	1-9	≤2M	92.2
Small	10-49	≤10M	6.5
Medium	50-249	≤50M	1.1

debate the in terms of number of employees, size does actually matter (O'Connor and Coleman, 2008a). Studies have shown that small firms' negative perceptions of process model standards are primarily driven by negative views of cost, documentation and bureaucracy. In addition, it has been reported that SMEs find it difficult to relate standards to their business needs and to justify the application of the international standards in their operations. Most SMEs cannot afford the resources for, or see a net benefit in, establishing software processes as defined by current standards and maturity models (O'Connor and Coleman, 2008b).

ISO/IEC 29110 Standard Proposed

Accordingly there is a need to help such organizations understand and use the concepts, processes and practices proposed in the ISO/IEC JTC1/SC7's international software engineering standards. The ISO/IEC 29110 standard "Lifecycle profiles for Very Small Entities" is aimed at addressing the issues identified above and addresses the specific needs of VSEs. The approach (Laporte et al, 2013a) used to develop ISO/IEC 29110 (2001) started with the pre-existing international standard ISO/IEC 12207 (2008) dedicated to software process lifecycles. The overall approach consisted of three steps: (1) Selecting ISO/IEC 12207 process subset applicable to VSEs of up to 25 employees; (2) Tailor the subset to fit VSE needs; and (3) Develop guidelines for VSEs.

Furthermore, in late 2009, the International Council on Systems Engineering (INCOSE) Very Small and Micro Entities Working Group (VSME) was established to evaluate the possibility of developing a standard, using the Generic profile group scheme of the ISO/IEC 29110 series, based on ISO/IEC 15288 (2008), for organizations developing systems. Late 2011 saw the launch of the official development of the systems engineering ISs and TRs for VSEs. In August 2014, ISO published the ISO/IEC 29110 systems engineering and management guide of the Basic profile ISO/IEC TR 29110-5-6-2:2014 (2014). The systems

engineering and management guide of the Entry profile has been published in 2015 as ISO/IEC TR 29110-5-6-1:2015 (2015). Similar to the existing set of software ISO/IEC 29110 TRs, the Management and Engineering Guide for systems engineering should also be made available at no cost by ISO (Laporte et al, 2014).

STRUCTURE OF ISO/IEC 29110

The basic requirements of a software development process are that it should fit the needs of the project and aid project success. And this need should be informed by the situational context where in the project must operate and therefore, the most suitable software development process is contingent on the context. The core situational characteristic (Clarke and O'Connor, 2012) of the entities targeted by ISO/IEC 29110 is size, however there are other aspects and characteristics of VSEs that may affect profile preparation or selection. Creating one profile for each possible combination of values of the various dimensions introduced above would result in an unmanageable set of profiles. Accordingly VSE's profiles are grouped in such a way as to be applicable to more than one category. Table 2 illustrates a Profile Group, which contains three profiles (labeled A, B and C) that are mapped to nine combinations of business models and situational factors.

Profile Groups are a collection of profiles, which are related either by composition of processes (i.e. activities, tasks), or by capability level, or both. The "Generic" profile group is applicable

Table 2. Allocating VSE characteristics to profile groups

Business Models	Profile Situational Factors		
	Critical	User Uncertainty	Environment Change
Contract	Profile A	Profile A	Profile A
In-House	Profile C	Profile B	Profile A
Commercial	Profile B	Profile A	Profile A

to a vast majority of VSEs that do not develop critical software and have typical situational factors. This profile group does not imply any specific application domain, however, it is envisaged that in the future new domain-specific sub-profiles may be developed in the future. Table 3 illustrates this profile group as a collection of four profiles, providing a progressive approach to satisfying the requirements of profile group. To date the Basic Profile has been published, the purpose of which is to define a software development and project management guide for performing one project at a time.

Engineering and Management Guide

At the core of this standard is a Management and Engineering Guide, officially known as ISO/IEC TR 29110-5-1-2 (2011), which focuses on *Project Management* and *Software Implementation* as illustrated in Figure 1. The purpose of the Basic Profile is to define Software Implementation (SI) and Project Management (PM) processes from a subset of ISO/IEC 12207 (2008) and ISO/IEC 15289 (2011) appropriate for VSEs, as illustrated in Figure 1.

Project Management Process

The purpose of the Project Management (PM) process is to establish and carry out the tasks of the

Table 3. Graduated profile of the Generic profile group.

Generic Profile Group			
Entry	Basic	Intermediate	Advanced

software implementation project in a systematic way, which allows compliance with the project's objectives in terms of expected quality, time, and costs (O'Connor and Laporte, 2012). The seven objectives of the PM process are listed in table 4.

Figure 2 illustrates the 4 activities of the project management process as well as their input and output product. The four activities of the Project Management Process are:

- **Project Planning:** The primary objective of this process is to produce and communicate effective and workable project plans. This process determines the scope of the project management and technical activities, identifies process outputs, project tasks and deliverables, establishes schedules for project task conduct, including achievement criteria, and required resources to accomplish project tasks.

Figure 1. ISO/IEC 29110 project management and software implementation relationship

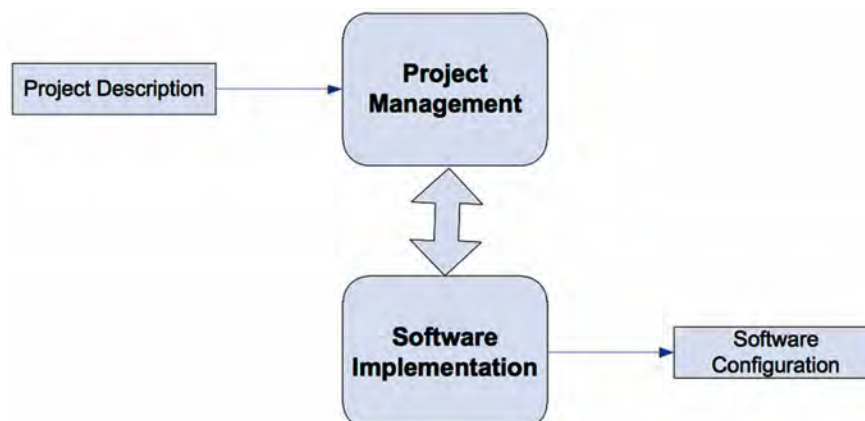


Table 4. Objectives of the project management process of the basic profile (ISO/IEC TR 29110-5-1-2, 2011)

Objective	Description
PM.O1	The Project Plan for the execution of the project is developed according to the Statement of Work and reviewed and accepted by the Customer. The tasks and resources necessary to complete the work are sized and estimated.
PM.O2	Progress of the project is monitored against the Project Plan and recorded in the Progress Status Record.
PM.O3	The Change Requests are addressed through their reception and analysis. Changes to software requirements are evaluated for cost, schedule and technical impact.
PM.O4	Review meetings with the Work Team and the Customer are held. Agreements are registered and tracked.
PM.O5	Risks are identified as they develop and during the conduct of the project.
PM.O6	A software Version Control Strategy is developed. Items of Software Configuration are identified, defined and baselined. Modifications and releases of the items are controlled and made available to the Customer and Work Team including the storage, handling and delivery of the items.
PM.O7	Software Quality Assurance is performed to provide assurance that work products and processes comply with the Project Plan and Requirements Specification.

- **Project Plan Execution:** To implement the actual work tasks of the project in accordance with the project plan. Ideally when the project plan has been agreed and communicated to all teams members, work of the development of the product, which is the subject of the project, should commence.
- **Project Assessment and Control:** Purpose is to determine the status of the project and ensure that the project performs according to plans and schedules, within projected budgets and it satisfies technical objectives.
- **Project Closure:** Typically involves releasing the final deliverables to the customer, handing over project documentation to the business, terminating supplier contracts, releasing project resources and communicating project closure to all stakeholders.

For illustration purposes, two tasks of the Project Planning activity are listed in Table 5. The project manager (PM) and the customer (CUS) are involved in these 2 tasks. The customer is involved, during the execution of the project, when he submits change requests, during project review meetings, for the validation and approval of the requirements specifications and for the acceptance of the deliverables.

Software Implementation Process

The purpose of the Software Implementation (SI) process, illustrated in Figure 3, is to achieve systematic performance of the analysis, design, construction, integration, and test activities for new or modified software products according to the specified requirements. The seven objectives of the SI process are listed in Table 6.

Table 5. Example of 2 tasks of the project planning activity (ISO/IEC TR 29110-5-1-2, 2011)

Role	Task	Input	Output
PM CUS	PM.1.2 Define with the Customer the Delivery Instructions of each one of the Deliverables specified in the Statement of Work.	Statement of Work [reviewed]	Project Plan Delivery Instructions
PM CUS	PM.1.14 Review and accept the Project Plan. Customer reviews and accepts the Project Plan, making sure that the Project Plan elements match with the Statement of Work.	Project Plan [verified]	Meeting Record Project Plan [accepted]

Figure 2. ISO/IEC 29110 project management process

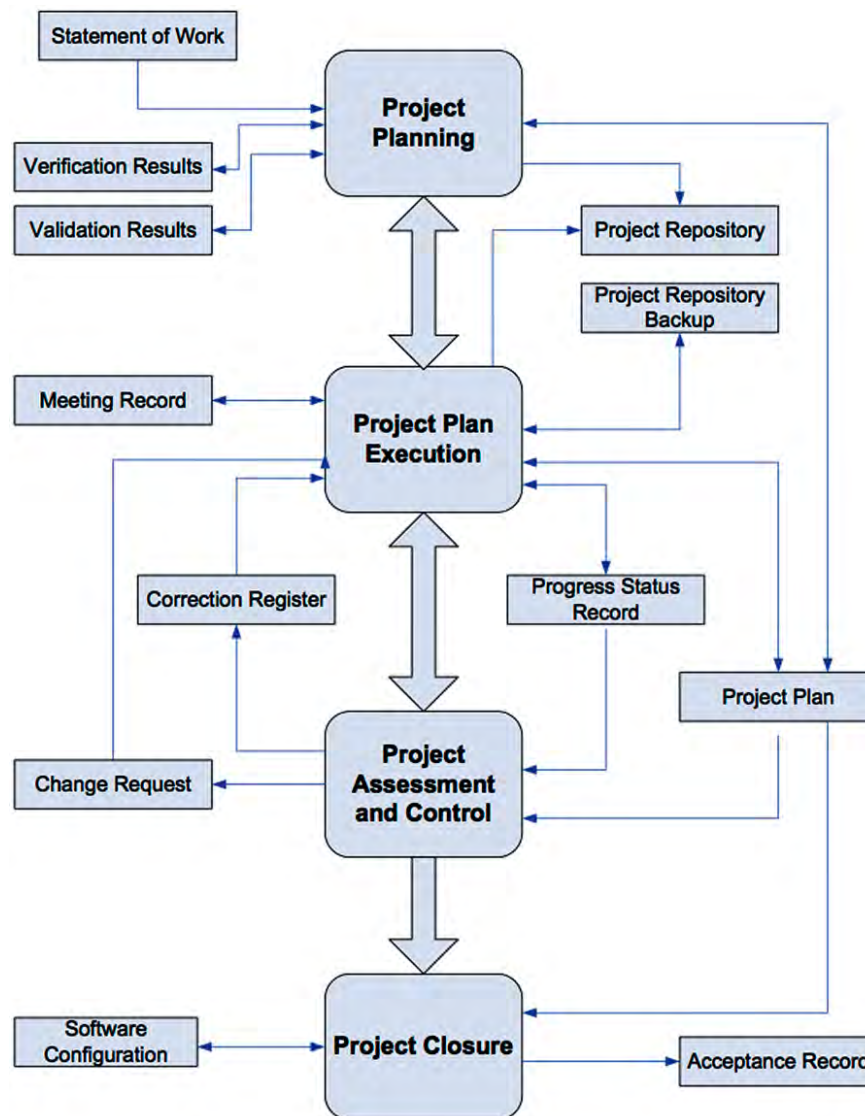


Table 6. Objectives of the software implementation process of the basic profile

Objective	Description
SI.O21	Tasks of the activities are performed through the accomplishment of the current Project Plan.
SI.O2.	Software requirements are defined, analyzed for correctness and testability, approved by the Customer, baselined and communicated.
SI.O3.	Software architectural and detailed design is developed and baselined. It describes the Software Components and internal and external interfaces of them.
SI.O4.	Software Components defined by the design are produced. Unit test are defined and performed to verify the consistency with requirements and the design. T
SI.O5.	Software is produced performing integration of Software Components and verified using Test Cases and Test Procedures. Results are recorded at the Test Report.
SI.O6.	A Software Configuration, that meets the Requirements Specification as agreed to with the Customer, which includes user, operation and maintenance documentations, is integrated, baselined and stored at the Project Repository.
SI.O8.	Verification and Validation Tasks of all required work products are performed using the defined criteria to achieve consistency among output and input products in each activity.

The activities of the Software Implementation Process are:

- **Software Implementation Initiation:** Ensures that the Project Plan established in Project Planning activity is committed to by the Work Team.
- **Software Requirements Analysis:** Analyzes the agreed Customer's requirements and establishes the validated project requirements.

The activity provides:

- **Software Architectural and Detailed Design:** Transforms the software requirements to the system software architecture and software detailed design.
- **Software Construction:** Develops the software code and data from the Software Design.
- **Software Integration and Tests:** Ensures that the integrated Software Components satisfy the software requirements.
- **Product Delivery:** Provides the integrated software product to the Customer.

IMPLEMENTING THE ISO/IEC 29110 STANDARD

In order to facilitate the implementation, by VSEs, of a Profile, a set of Deployment Packages (2013) are available. A deployment package is a set of artifacts developed to facilitate the implementation of a set of practices, of the selected framework, in a VSE. A deployment package is not a process reference model (i.e. it is not prescriptive). The elements of a typical deployment package are: description of processes, activities, tasks, roles and products, template, checklist, example, reference and mapping to standards and models, and a list of tools. Deployment packages are not intended to preclude or discourage the use of additional guidelines that VSEs find useful.

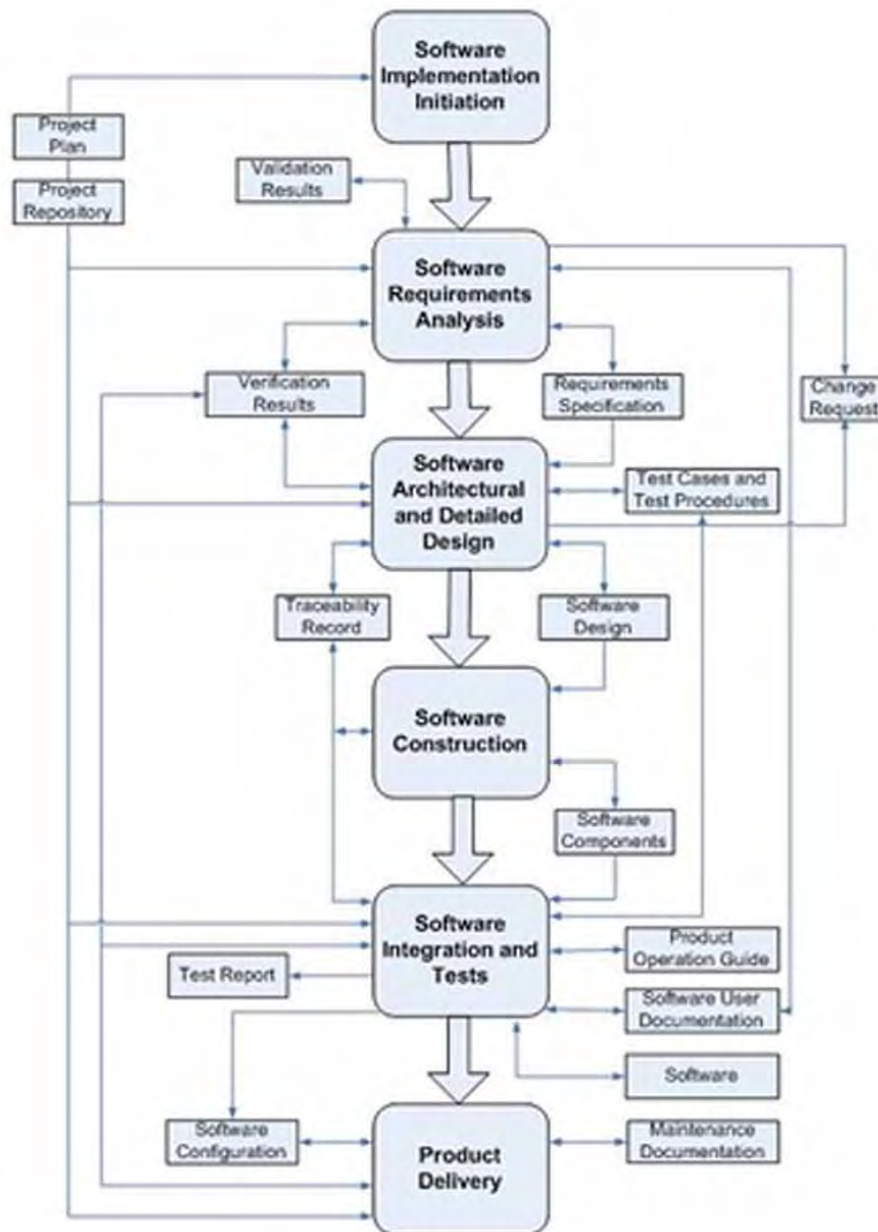
The elements of a typical deployment package are: technical description, relationships with ISO/IEC 29110, key definitions, detailed description of processes, activities, tasks, roles and products, template, checklist, example, references and mapping to standards and models, and a list of tools. The mapping is only given as information to show that a Deployment Package has explicit links to Part 5, ISO standards, such as ISO/IEC 12207, or models such as the CMMI developed by the Software Engineering Institute. Hence by deploying and implementing a package (O'Connor and Sanders, 2013) a VSE can see its concrete step to achieve or demonstrate coverage to Part 5.

The working group (ISO/IEC JTC1/SC7 WG 24) behind the development of this standard is advocating the use of pilot projects as a mean to accelerate the adoption and utilization of ISO/IEC 29110. Pilot projects are an important means of reducing risks and learning more about the organizational and technical issues associated with the deployment of new software engineering practices. Pilot projects are based on the ISO/IEC 29110-5 Management and engineering guide and the deployment package(s).

To date a series of pilot projects have been completed in several countries utilizing some of the deployment packages developed. For example Ribaud et al. (2010) have documented the results of one pilot project that conducted with a 14-person VSE based in France, which successfully implemented ISO/IEC 29110 processes practices utilizing the available Deployment Packages. From which they have identified some potential additional infrastructure and support process activities and suggestions for future evolution of ISO/IEC 29110 Process Profiles. A further series of pilot projects are currently underway in research laboratories and enterprises in Canada, Ireland, Belgium and France, with further pilot projects planned in the near future.

The results from one pilot study in Canada concluded that the tools developed to support the project management processes proved very useful and helped the project managers rapidly

Figure 3. ISO/IEC 29110 software implementation process



integrate the knowledge required to execute the processes (Laporte et al, 2013b). In the case of this trial company, for the first time, the company has documented management processes for small-scale projects. Besides, some project managers have joined forces to promote project management practices within this engineering firm's division. The improvement programme was so successful

that managers of the company's other divisions have shown an interest in learning this approach in order to implement it within their respective divisions.

Laporte et al (2015) report on two successful trials of ISO/IEC 29110, that demonstrate it was possible to properly plan the project and develop the software product using proven software prac-

tices documented in standards as well as not interfering with the creativity during the development of their web site. People who think that standards are a burden, an unnecessary overhead and a treat to creativity should look at this start-up project and revisit their results.

CONCLUSION

For most enterprises, but in particular for VSEs, international certifications can enhance credibility, competitiveness and access to national and international markets. Brazil has led the development of an ISO/IEC 29110 certification process. An ISO/IEC 29110 auditor should be competent in auditing techniques, have expertise in ISO/IEC 29110 and have experience in software development. For VSEs, such a certification should not be too expensive and short. The certification process has been successfully piloted in a few VSEs.

Finally, research studies have been undertaken to understanding the perception of VSEs towards the adoption of process standards (Basri et al, 2010) and also to evaluate management sentiment towards ISO/IEC 29110 (O'Connor, 2012) and management commitment to SPI and ISO/IEC 2910 in particular in Europe (O'Connor et al, 2010) and South America (Sanchez-Gordon et al, 2015). These revealed that the acceptance level of any type or model of software quality or lifecycle standard in VSEs is a very low priority item, but the level of awareness of standards and potential benefits was high. Furthermore these studies showed the main reason for not adopting standards was a lack customer requirement, a lack of resources and the perceived difficulties in defining an organizational process. Furthermore, this analysis reveals a pattern that indicates that the acceptance level of quality standard such as ISO among VSEs are still low even though the staff and management are knowledgeable and aware the benefit of adopting such standards. The main reasons are more related to the lack of the customer requirement and the limited resources

in the company. In addition the perception a heavyweight process especially in terms of documentation, cost and non- alignment with current development process are among the reasons why the companies did not plan to adopt a lifecycle standard in the short to medium term. However from the analysis, VSEs may still be interested in lifecycle standards if certain important criteria are met and such standards are closely related to their needs. Therefore it can be concluded that the market and demand for ISO/IEC 29110 in VSEs has a positive outlook.

FUTURE RESEARCH DIRECTIONS

In terms of future work, as ISO/IEC 29110 is an emerging standard there is much work yet to be completed (O'Connor and Laporte, 2014). The main remaining work item is to finalize the development of the remaining three profiles: (a) Entry – a six person-months effort project or a start-up VSEs; (b) Intermediate - Management of more than one project and (c) Advanced - business management and portfolio management practices. In addition the development of additional Profile Groups for other domains such as critical software, game industry, scientific software developments are being studied.

Whilst work is currently underway on an assessment mechanism for ISO/IEC 29110, a clear niche market need is emerging which may force the process assessment community to change their views on how process assessments are carried out for VSEs. In particular there is a strong need to ensure that VSEs are not required to invest the anything similar in terms of time, money and other resources on process assessments, as may be expected from their larger SMEs (small and medium enterprises), or even MNC (multinational corporations) counterparts. Indeed some form of self-assessment, possibly supported by Internet based tools, along with periodic spot-checks may be suitable alternative to meet the unique needs of VSEs.

REFERENCES

- Basri, S., & O'Connor, R. (2010). Understanding the Perception of Very Small Software Companies towards the Adoption of Process Standards. In *Proceedings of Systems, Software and Services Process Improvement* (pp. 153 – 164). Springer-Verlag doi:10.1007/978-3-642-15666-3_14
- Clarke, P., & O'Connor, R. (2012). The situational factors that affect the software development process: Towards a comprehensive reference framework. *Information and Software Technology*, 54(5), 433–447. doi:10.1016/j.infsof.2011.12.003
- Coleman G. and O'Connor R. (2008b). The Influence of Managerial Experience and Style on Software Development Projects. *International Journal of Technology, Policy and Management*, 8(1).
- Coleman, G., & O'Connor, R. (2008a). Investigating Software Process in Practice: A Grounded Theory Perspective. *Journal of Systems and Software*, 81(5), 772–784. doi:10.1016/j.jss.2007.07.027
- ISO. IEC 12207. (2008). Information technology – Software lifecycle processes. Geneva, Switzerland: International Organization for Standardization/International Electrotechnical Commission.
- ISO. IEC 15288. (2008). Systems and software engineering - System lifecycle processes. International Organization for Standardization/International Electrotechnical Commission.
- ISO. IEC 15289. (2011). Systems and software engineering – Content of lifecycle information products (documentation). International Organization for Standardization/International Electrotechnical Commission.
- ISO/IEC TR 29110-1. (2011). *Software Engineering - Lifecycle Profiles for Very Small Entities (VSEs) - Part 1: Overview*. Geneva: International Organization for Standardization (ISO). Available at http://standards.iso.org/ittf/PubliclyAvailableStandards/c051150_ISO_IEC_TR_29110-1_2011.zip
- ISO/IEC TR 29110-5-1-2. (2011). *Software Engineering - Lifecycle Profiles for Very Small Entities (VSEs) - Part 5-1-2: Management and engineering guide - Generic profile group: Basic profile*. International Organization for Standardization/International Electrotechnical Commission. Available at http://standards.iso.org/ittf/PubliclyAvailableStandards/c051153_ISO_IEC_TR_29110-5-1_2011.zip
- ISO/IEC TR 29110-5-6-1. (2015). *Systems Engineering – Lifecycle Profiles for Very Small Entities (VSEs) - Management and engineering guide: Generic profile group: Entry profile*. International Organization for Standardization/International Electrotechnical Commission Available at <http://standards.iso.org/ittf/PubliclyAvailableStandards/index.html>
- ISO/IEC TR 29110-5-6-2. (2014). *Systems Engineering – Lifecycle Profiles for Very Small Entities (VSEs) - Management and engineering guide: Generic profile group: Basic profile*. International Organization for Standardization/International Electrotechnical Commission. Available at <http://standards.iso.org/ittf/PubliclyAvailableStandards/index.html>
- Laporte, C. Y., Alexandre, S., & O'Connor, R. (2008). A Software Engineering Lifecycle Standard for Very Small Enterprises. In *Proceedings of EuroSPI* (pp. 129-141). Springer-Verlag.
- Laporte, C. Y., Chevalier, F., & Maurice, J.-C. (2013b). *Improving Project Management for Small Projects, ISO Focus*. In *Proceedings of International Organization for Standardization*.
- Laporte, C. Y., O'Connor, R., & Fanmuy, G. (2013a). International Systems and Software Engineering Standards for Very Small Entities. *Proceedings of CrossTalk - The Journal of Defense Software Engineering*, 26(3), 28-33.

- Laporte, C. Y., & O'Connor, R. V. (2014). A Systems Process Lifecycle Standard for Very Small Entities: Development and Pilot Trials. In *Proceedings of EuroSPI*. Springer-Verlag. doi:10.1007/978-3-662-43896-1_2
- Laporte, C. Y., O'Connor, R. V., & Garcia Paucar, L. (2015). Software Engineering Standards and Guides for Very Small Entities: Implementation in two start-ups. *Proceedings of 10th International Conference on Evaluation of Novel Approaches to Software Engineering (ENASE 2015)*. doi:10.5220/0005368500050015
- Moll, R. (2013). *Being prepared – A bird's eye view of SMEs and risk management, ISO Focus+*. Geneva, Switzerland: International Organization for Standardization.
- O'Connor, R., Basri, S., & Coleman, G. (2010). Exploring Managerial Commitment towards SPI in Small and Very Small Enterprises. In *Proceedings of Systems, Software and Services Process Improvement* (pp. 268-278). Springer-Verlag.
- O'Connor, R., & Coleman, G. (2009). Ignoring 'Best Practice': Why Irish Software SMEs are rejecting CMMI and ISO 9000. *Australasian Journal of Information Systems*, 16(1).
- O'Connor, R. V. (2012). Evaluating Management Sentiment towards ISO/IEC 29110 in Very Small Software Development Companies. In A. Mas et al. (Eds.), *SPICE 2012. CCIS* (Vol. 290, pp. 277-281). Heidelberg, Germany: Springer. doi:10.1007/978-3-642-30439-2_31
- O'Connor, R. V., & Laporte, C. Y. (2012). Software Project Management with ISO/IEC 29110. *19th European Conference on Systems, Software and Services Process Improvement (EuroSPI 2012)* (Vol. 301). Springer-Verlag.
- O'Connor, R. V., & Sanders, M. (2013). Lessons from a Pilot Implementation of ISO/IEC 29110 in a Group of Very Small Irish Companies. *Proceedings 13th International Conference on Software Process Improvement and Capability determination (SPICE 2013)*, (Vol. 349). Springer-Verlag.
- O'Connor, R. V., & Laporte, C. Y. (2014). An Innovative Approach to the Development of an International Software Process Lifecycle Standard for Very Small Entities. *International Journal of Information Technologies and Systems Approach*, 7(1), 1-22. doi:10.4018/ijitsa.2014010101
- Package, D. (2013). *ISO/IEC JCT1/SC7 Working Group 24 Deployment Packages repository*. Retrieved November 10, 2013, from <http://profs.logti.etsmtl.ca/claporte/English/VSE/index.html>
- Ribaud, V., Saliou, P., O'Connor, R., & Laporte, C. (2010). Software Engineering Support Activities for Very Small Entities. In *Proceedings of Systems, Software and Services Process Improvement* (pp. 165-176). Springer-Verlag. doi:10.1007/978-3-642-15666-3_15
- Sanchez-Gordon, M. L., O'Connor, R. V., & Colomo-Palacios, R. (2015). Evaluating VSEs Viewpoint and Sentiment Towards the ISO/IEC 29110 Standard: A Two Country Grounded Theory Study. *Proceedings 13th International Conference on Software Process Improvement and Capability determination (SPICE 2015)*, (Vol. 526). Springer-Verlag.
- Staples, M., & Niazi, M. (2006). *Systematic Review of Organizational Motivations for Adopting CMM-based SPI*. Technical Report PA005957, National ICT Australia.
- Staples, M., Niazi, M., Jeffery, R., Abrahams, A., Byatt, P., & Murphy, R. (2007). An exploratory study of why organizations do not adopt CMMI. *Journal of Systems and Software*, 80(6), 883-895. doi:10.1016/j.jss.2006.09.008

ADDITIONAL READING

Basri, S., & O'Connor, R. (2011). Knowledge Management in Software Process Improvement: A case study of very small entities. In M. Ramachandran (Ed.), *Knowledge Engineering for Software Development Life Cycles: Support Technologies and Applications* (pp. 273–288). IGI Global. doi:10.4018/978-1-60960-509-4.ch015

Basri, S., & O'Connor, R. V. (2012). A Study of Knowledge Management Process Practices in Very Small Software Companies. *American Journal of Economics and Business Administration*, 3(4), 636–644.

Clarke, P., & O'Connor, R. V. (2013). An empirical examination of the extent of software process improvement in software SMEs. *Journal of Software: Evolution and Process*, 25(9), 981–998.

Clarke, P., & O'Connor, R. (2012). The situational factors that affect the software development process: Towards a comprehensive reference framework. *Information and Software Technology*, 54(5), 433–447. doi:10.1016/j.infsof.2011.12.003

Clarke, P., & O'Connor, R. V. (2012). The influence of SPI on business success in software SMEs: An empirical study. *Journal of Systems and Software*, 85(10), 2356–2367. doi:10.1016/j.jss.2012.05.024

Laporte, C. Y., Fanmuy, G., Ptack, K., & Marvin, J. (2012). Systems and Software Engineering Standards for Very Small Entities. *Insight (American Society of Ophthalmic Registered Nurses)*, 15(1), 32–33.

Laporte, C. Y., Séguin, N., & Villas Boas, G. (2013). *Seizing the benefits of software and systems engineering standards*, *ISO Focus* (pp. 32–36). International Organization for Standardization.

KEY TERMS AND DEFINITIONS

Process Assessment: The disciplined examination of the processes by an organisation against a set of criteria to determine capability of those processes to perform within quality, cost and schedule goals.

Project Implementation: Is defined as a specified set of activities designed to put into practice an activity or program of known dimensions.

Project Management: This is the process and activity of planning, organizing, motivating, and controlling resources to achieve specific goals.

Software Process: A set of activities, methods, practices and transformations that people use to develop and maintain software and the associated products.

Software Process Improvement (SPI): Aims to understand the software process as it is used within an organisation and thus drive the implementation of changes to that process to achieve specific goals such as increasing development speed, achieving higher product quality or reducing costs.

Very Small Entity: An enterprise, organization, department or project having up to 25 people.

Category S

Social Networking and Computing

Adolescents' Food Communication in Social Media

Christopher Holmberg

University of Gothenburg, Sweden

INTRODUCTION

Social media is ubiquitous in the lives of adolescents. The Swedish Media Council (2015) reports that 91% of Swedish adolescents between 13 and 16 years old use social media sites such as Facebook or Instagram and similar numbers exist in the US (Lenhart, Smith, Anderson, Duggan, & Perrin, 2015). Social media are usually defined as Internet-based services that permit users to construct personal profiles, generate and access searchable online content (e.g., images, videos), form online connections with other users, and view these social connections (CDC, 2011).

Considering the widespread use of social media, it is no surprise that much of the communication in these networks concern food and nutrition. Previous studies show that the dissemination and sharing of food content is prevalent in many of these channels (Holmberg, Chaplin, Hillman, & Berg, 2016). As Neely, Walton, and Stephens (2014) noted, the communication of food mirrors social and cultural values, and the social aspects of young people's food practices are important components of their wellbeing and health. Social media provides a unique opportunity to study adolescents' food communication without interfering with it, and in order to gain a better understanding of what food means to adolescents, it is crucial to study their online food communication.

Adolescents' online food communication is a pressing matter because online food communication affects adolescents in both positive and negative ways with regard to health (Holmberg, 2016). Not only does messages of food in social media platforms and channels serve a symbolic purpose

concerning socio-cultural aspects of communication and discourse. A meta-analysis by van Meer et al. (2015) shows that food communication, such as food images, affects brain areas associated with appetite. Other experimental studies show that food communication in social media can affect dietary behaviors among adolescents (Bevelander, Anschutz, Creemers, Kleinjan, & Engels, 2013).

The goal of this chapter is to introduce a promising research direction regarding adolescents' food-related communication in social media. The objective is to elucidate the complex and interwoven relationship between food and nutrition, social media, and adolescents from a health promotion perspective. As such, this article will draw upon empirical studies and results as well as conceptual and methodological research literature.

BACKGROUND

The social nature of food and food practices - that is, the idea that food is a way to interact and connect with others - is crucial to promoting health. To better understand food and its complex relationship with social media, we need to consider how diets and food habits connect people culturally. While food and nutrients are necessary for human survival, they also function as important objects in culture and as rich sources for metaphor (Korthals, 2008). Some sociologists even aver that food is a total social fact (Mauss, 1967). The cultural dimension of food practices has been termed the "omnivore's dilemma" (Korthals, 2008) or "omnivore's paradox" (Fischler, 1988), based on the assumption that humans can eat a wide variety

of things. Unlike specialized eaters, omnivores such as humans can thrive on a multitude of diets and lack inherent predilections for foods that are healthy. Culture thus becomes a primary factor that dictates human eating behaviors, which suggests that the social meaning and metaphors of food can affect food choices and implicate which types of food confer social acceptance.

Given that young people are heavy users of social media and that their usage is generally difficult for parents and guardians to monitor (McBride, 2011), it becomes critical to examine health and food messages in those channels. Research demonstrates that the adolescent period is marked by changes in body composition, physical activity, diet habits, and psychological issues. These issues put youths at an increased risk of adapting unhealthy lifestyles (Alberga, Sigal, Goldfield, Prud'homme, & Kenny, 2012). As Vuk Pisk et al. (2012) have documented, adolescence is therefore a critical time for identity formation and a very sensitive period where social ties and relationships with others grow in importance. Since adolescents' social networks function as important influences for their dietary habits, researchers in many fields, such as obesity research, emphasize the importance of further exploring these social network effects.

FOOD CONTENTS IN NUMEROUS SOCIAL MEDIA

Communication includes both the sending and receiving of messages. Online, adolescents are exposed to messages relating to food, but they also act as uploaders of food-related content such as images or videos of food items. Along these lines, there are a growing number of studies highlighting ways in which food communication is perpetuated through different types of social media applications.

Food blogs provide information and inspiration regarding recipes, meal ideas, and food items. As such, they have the capacity to shape

adolescents' social norms regarding how, what, and where to eat. Schneider et al. (2013) found that popular American food blogs met energy recommendations but were excessive in saturated fat and sodium. The authors therefore suggested that the public should be aware of the nutritional limitations of common food blogs. Simunaniemi et al. (2011) analyzed how laypeople blogged about fruits and vegetables. The study revealed that the bloggers approached fruit and vegetables through either lived or mediated experiences and that self-expression that aimed to influence others' diets was common in these blogs. There are also examples of how blogs are used to foster controversial dietary practices. Using the spreadable media theory (Jenkins, Ford, & Green, 2013), research has shown that blogs allow for non-conventional experts to circumvent traditional peer-review processes by publishing in the public domain. Holmberg (2015) uses the example of how prominent low-carb high fat diet promoters in Sweden could reach a wide audience and appeal to a like-minded community of dieters by using self-made blogs and websites. Also microblogs such as Twitter have been analyzed in regards to food communication. Hingle et al. (2013) found that Twitter can provide a method for observing real-time food consumption and that this type of data visualization may present a method of identifying relationships between diet and behavior.

Media-sharing sites popular with adolescents, such as YouTube also propagate large amounts of food and diet information. Cerri et al. (2012) showed that a large number of the videos on YouTube labeled with the word "diet" did not contain this information. The researchers also found that a majority of the videos did not contain scientifically sound information and that videos often contained incomplete information. Boyd and Ellison (2007) defined social network sites as:

... web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users

with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system. (p. 211)

Facebook is the world's largest online social network site and it is no surprise that food-and-nutrition related content has been studied on this platform. Lohse (2013) found that Facebook is an effective strategy to recruit participants to nutrition education programs. There are also less healthy occurrences taking place on Facebook. Freeman et al. (2014) found that food companies promoting high calorie low nutrient food items (e.g. Coca-Cola) utilize the interactive aspects of Facebook to market their products. The study showed that these companies capitalize on users' social networks to magnify the reach and personal relevance of their marketing. Young users were particularly prone to this kind of online food marketing. A social network especially popular with younger individuals is Instagram (Pew Research Center, 2015). Studies have investigated both the food-related content on Instagram (Hu, Manikonda, & Kambhampati, 2014) and the overall communication of food images from the perspective of adolescents (Holmberg et al., 2016). Holmberg et al. (2016) found that food images shared on Instagram by adolescents was dominated by foods characterized as high calorie and low nutrient such as sodas and candy.

Why Can It Be Problematic?

As noted in previous examples food marketing transmitted through social media outlets are frequently transmitted by food companies producing less health food items. Montgomery and Chester (2011) outlined five kinds of digital marketing techniques that are used by fast food companies to target adolescents: 1) augmented reality; such as online virtual environments that can decrease conscious attention to marketing techniques, 2) methods that include surveillance of users' behaviors online, 3) data gathering and behavioral tailoring intended to distribute individualized

marketing, 4) location targeting which follow adolescents' activities and are able to link point of influence to point of purchase, 5) neuro-marketing, which draws on neuroscience methods to design digital marketing techniques intended to activate subconscious and emotional arousal. Many of these intrusive practices are intentionally aimed to draw on unconscious processes, circumventing the rational decision making that is central to the system of just marketing.

These integrated food advertisement campaigns reinforce commercial messages across several online platforms to engage individually with users. These campaigns may have greater impact than that of conventional marketing, especially among adolescent consumers (Kelly, Vandevijvere, Freeman, & Jenkin, 2015). Food brands have been noted to be a key factor in the success of food marketing. The perceptions of brands changes as children become older; children learn how to recognize commercial, supermarket and budget brands and attach different meaning to these brands. Roper and La Niece (2009) found that food brands were important for British schoolchildren, and they also noted that peers become more influential to consumption decisions as children move through the tweens. In their study, peer approval substituted family as the major influence behind consumption behavior.

This behavior seems to be perpetuated in social networking sites. Holmberg et al. (2016) found that labeled food products were highly prevalent in food images uploaded by adolescents on Instagram. Like the British study (Roper & La Niece, 2009), the presentation and display of certain food brands were important to present to peers in the adolescents' social networks. The idea of food as a signifier of status is not novel. Bourdieu (1984) critically discussed that our preferences for certain foods are reflected by our social position, and are themselves acts of social positioning. What is new however, is the pace and means by which food is communicated today. Food is no longer seen just as essential for the body; it has become a currency that can be used online for social status.

Adolescents may increase their social capital by sharing desirable meals and food items through social media. In order to understand adolescents' practice of online food communication we need to delve further into the theoretical concepts surrounding adolescents' social food practices. Therefore, this chapter will continue to focus on online social networks which explore social dynamics among adolescents and how they could influence dietary habits and food consumption in different ways.

THEORETICAL FRAMEWORKS THAT CAN ADVISE RESEARCH

It is significant to recognize adolescents' influence on food choices and their active role in conveying family practices and customs. A review of youths' food practices and social relationships showed that food practices play multifaceted roles in adolescents' social lives, and that food communication influence different aspects of their social relationships. Aspects such as caring, sharing, integrating, trusting, interchanging, negotiating, and belonging was some of the features of relationships that were affected through adolescents' different food practices (Neely et al., 2014). Salvy et al. (2012) sorted the influence of peers on adolescents' eating habits into impression management, and social facilitation and modeling. They also refer to the normative framework, put forth by Herman et al. (2003), while attempting to assimilate research concerning these collective mechanisms.

Impression Management and Social Facilitation

boyd (2007) draws on the work of Goffman (1959) when defining impression management among adolescents in online interaction. According to boyd, adolescents seek to define social situations online by using signs from the context around them. Social norms arise out of situations as adolescents learn to interpret cues from the environment and

apprehend what considers appropriate behavior. The management of impressions is a vital social skill that is refined through experience. The adolescents learn how to create meaning out of a situation and others' reactions over time. boyd further argues that adolescents in a sense have more control online as they are able to more deliberately choose what information to put forward, thus reducing impulsive and spontaneous reactions that might have leaked out in everyday interaction. This understanding is important as the adolescents, to varying degrees, have an active choice in what kind of food communication they are transmitting online. Previous research indicates that in the presence of others, individuals tend to eat more than when eating alone. The leading explanation for the social facilitation of eating is the time-extension premise, which claims that the company of others ultimately encourages increased eating by prolonging the duration of the meal (Pliner, Bell, Hirsch, & Kinchla, 2006). However, overweight adolescents in experimental studies eat less snacks in social situations than normal weight adolescents, eat more when alone, and also eat more in the presence of overweight peers than in the presence of leaner peers (Salvy, 2009). For adolescents who are overweight or obese, other processes may overtake social facilitation. For instance, when considering the strong connection between overweight status and social stigma, it is credible to assume that overweight youths are particularly interested in making a successful impression on unacquainted peers. Thus, adolescents might manage their impression online by considering and curating the food messages they transmit to others.

While most studies based in social facilitation theory focuses on the means in which the presence of others influences behavior, researchers have claimed that individuals modify their behavior to that of others; that behavior is learned from the environment in a process of observational learning. This process is often referred to as social modeling (Bandura, 1977; Grusec, 1992). Online social modeling of food consumption has been observed

among children and adolescents. Bevelander et al. (2012) showed by means of computer choice task with food pictures, that the use of fictional peers amplified children's inclination to try unfamiliar foods. A review of social modeling studies regarding food choice found that there was near universal support that food selection are influenced by the behaviors of others (Cruwys, Bevelander, & Hermans, 2015). The authors noted that there is indication that social modeling happens both because a), individuals seek information about suitable behaviors, motivated by reducing uncertainty, and b), because individuals want to identify with others, driven by an affiliation motive.

Whereas research on social modeling focus on modifications in persons' behaviors during the presence of others, research in the field of impression management aims to understand peoples' motivations to behave when in the company of others (Salvy et al., 2012). Impression management is often used interchangeably with self-presentation, where an individual seeks to influence the perception of their own image (boyd, 2007). Kümpel Nørgaard et al. (2013) showed that adolescents purchase and consume snacks that support their self-image when socializing with other peers. Krämer and Winter's (2008) study indicated that self-efficacy with regard to impression management in a social networking site is strongly associated to the number of virtual friends, the level of profile detail, and the style of the profile photo. Eating less seems to be a way to demonstrate a good impression at any age, probably because obesity and overweight is socially stigmatized, and persons are inclined to associate negative characteristics (e.g., physically inactive, overeaters) with persons who eat large quantities of food (Vartanian, Herman, & Polivy, 2007). Nevertheless, significant exceptions do exist: as previously explained; food items have cultural and social implications, and particular types of foods (e.g. desirable food brands or unhealthy snack food) linked with social status may be consumed and presented to others. Researchers have primarily linked this to young individuals' inclination

to express a positive impression amongst peers, which is emphasized during adolescence when peer consent becomes progressively important (Roper & La Niece, 2009).

Normative and Social Network Theories

One way to assimilate the mechanisms of social influences of eating is to adhere to Herman et al.'s (2003) normative structure, accounting for the influences of others on eating behavior. According to Aronson et al. (2010), social norms are defined as the procedures within a group for the tolerable behaviors, values, and beliefs of its members. The normative model suggests that, in the company of appetizing food, and when absent of other constrictions, individuals are motivated to eat as much as possible and that the presence of others is what governs when consumption stops. Consequently, social norms mainly function as inhibitory mechanisms, signifying when people should stop eating if they want to be socially appropriate. There are indications that apparent peer norms (descriptive norms; e.g. what the person thinks important others do) are even stronger connected to adolescents' soft drink consumption than the actual behavior of the peer group (Perkins, Perkins, & Craig, 2010). Lally et al. (2011) also showed that descriptive norms were strongly correlated with consumption of fruit and vegetables, soft drinks, and unhealthy snacks. However, the study did not identify an association between injunctive norms (i.e. beliefs about how a person should behave) and dietary intake.

Adolescents are exposed to the food related norms and influences of peers in their extended online networks. Although adolescents will normally add friends and acquaintances as 'friends', they will also add people since it would be socially inelegant to decline friendship requests, because they make the adolescent seem cool, or merely because it would be fascinating to follow their uploads (boyd, 2007). Because 'friends' are displayed on the adolescents' online profile,

they also provide meaningful information about that person. This reflection is significant for food communication as eating is usually a highly communal experience (Delormier, Frohlich, & Potvin, 2009). When adolescents share photos of foods they experience, the photos also becomes a confirmation of the larger event experience with other individuals; they want people to know who they participated with. A desirable lifestyle therefore includes the opportunity to share exclusive foods via one's online social networks. Facebook users can for example see food-related pages that members of their online social network have endorsed (e.g. 'liked'), and also food content they post. This may establish social norms around certain foods high in social status.

Stok et al. (2013) investigated whether communicating health-promoting descriptive and injunctive norms influences adolescents' intended and actual fruit consumption. The researchers distributed booklets with informational text regarding the consumption of fruit and vegetables to the students. They found that booklets containing descriptive norms influenced consumption but not intentions. The researchers also found that the injunctive norms not only had no positive effect, but that they could potentially decrease fruit intake. The authors discussed that injunctive norms could lead persons to feel like they were being pushed in a certain track by the source distributing the norm, particularly when the norm was not in line with their personal beliefs. Conversely, the researchers thought that descriptive norms could motivate by indicating what would be effective and adaptive behavior in a particular situation.

The normative structure is particularly useful when interpreting recent evidence that obesity and obesogenic behaviors tend to cluster in adolescents' social networks (Salvy et al., 2012). Research conducted in the field of social psychology has showed that adolescents' consumption of snacks and soft drinks was related with their peers' intake (Wouters, Larsen, Kremers, Dagnelie, & Geenen, 2010). The researchers also found that the correlations between individual and peer snack intake was particularly

strong among boys and youths with lower body weights. Food intake is often underreported in adolescents as well as adults with obesity, compared with their normal weight peers. Bevelander et al. (2011) showed that not only are adolescents' influenced by their peers when it comes to consuming foods, teenagers also influence each other when it comes to food purchases; girls' higher calorie food purchases are correlated with their peers' food purchases. Since friends share their snacking behaviors, education and interventions need to address both the individual snacking practice as well as the social component of snacking.

FUTURE RESEARCH DIRECTIONS

Considering the widespread use of social media applications among adolescents, future experimental studies are needed to examine whether social media networks such as Instagram, blogs, and Facebook actually influence adolescents' food habits. Specifically, research is needed to explore how the content and delivery of information on social media appeals and engages adolescents, stimulates behavior change, sustains behavior change, and how social media impact obesity risks in adolescents. This literature review also emphasize that there are theoretical underpinnings that need to be explored and developed further, especially concerning their usability in social media studies involving adolescents.

CONCLUSION

This chapter employs different theoretical and analytic frameworks that suggest social media as an evolving setting for social influence on adolescents' dietary habits, which also have contiguous impacts on children's health risks such as obesity. Social modeling, normative theory, and network theory are all relevant frameworks that can be used to explore how social media influences adolescents' food practices.

REFERENCES

- Alberga, A. S., Sigal, R. J., Goldfield, G., Prudhomme, D., & Kenny, G. P. (2012). Overweight and obese teenagers: Why is adolescence a critical period? *Pediatric Obesity*, 7(4), 261–273. doi:10.1111/j.2047-6310.2011.00046.x PMID:22461384
- Aronson, E., Wilson, T. D., & Akert, R. M. (2010). *Social psychology* (7th ed.). Upper Saddle River, NJ: Pearson Education.
- Bandura, A. (1977). *Social Learning Theory*. Englewood Cliffs, NJ: Prentice Hall.
- Bevelander, A., Anschutz, D. J., Creemers, D. H. M., Kleinjan, M., & Engels, R. C. M. E. (2013). The Role of Explicit and Implicit Self-Esteem in Peer Modeling of Palatable Food Intake: A Study on Social Media Interaction Among Youngsters. *PLoS ONE*, 8(8), e72481. doi:10.1371/journal.pone.0072481 PMID:24015251
- Bevelander, K. E., Anschutz, D. J., & Engels, R. C. (2012). The Effect of a Fictitious Peer on Young Children's Choice of Familiar v. Unfamiliar Low- and High-Energy-Dense Foods. *The British Journal of Nutrition*, 108(6), 1126–1133. doi:10.1017/S0007114511006374 PMID:22313605
- Bevelander, K. E., Anschutz, D. J., & Engels, R. E. (2011). Social Modeling of Food Purchases at Supermarkets in Teenage Girls. *Appetite*, 57(1), 99–104. doi:10.1016/j.appet.2011.04.005 PMID:21527294
- Bourdieu, P. (1984). *Distinction: a social critique of the judgement of taste*. London: Routledge.
- boyd. (2007). Why Youth (Heart) Social Network Sites: The Role of Networked Publics in Teenage Social Life. In D. Buckingham (Ed.), *MacArthur Foundation Series on Digital Learning - Youth, Identity, and Digital Media Volume*. Cambridge, MA: MIT Press.
- boyd, d., & Ellison, N. B. (2007). Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication*, 13(1), 210–230. doi:10.1111/j.1083-6101.2007.00393.x
- CDC. (2011). *The Health Communicator's Social Media Toolkit*. Retrieved from http://www.cdc.gov/healthcommunication/ToolsTemplates/SocialMediaToolkit_BM.pdf
- Cerri, E. M., Fisher, A. R., & Taheri, S. (2012). What is social media feeding you? A study of diet and weight loss information available on YouTube. *Appetite*, 58(3), 1171. doi:10.1016/j.appet.2012.02.024
- Cruwys, T., Bevelander, K. E., & Hermans, R. C. J. (2015). Social Modeling of Eating: A Review of When and Why Social Influence Affects Food Intake and Choice. *Appetite*, 86(0), 3–18. doi:10.1016/j.appet.2014.08.035 PMID:25174571
- Delormier, T., Frohlich, K. L., & Potvin, L. (2009). Food and eating as social practice – understanding eating patterns as social phenomena and implications for public health. *Sociology of Health & Illness*, 31(2), 215–228. doi:10.1111/j.1467-9566.2008.01128.x PMID:19220802
- Fischler, C. (1988). Food, self and identity. *Social Sciences Information. Information Sur les Sciences Sociales*, 27(June), 275–292. doi:10.1177/053901888027002005
- Freeman, B., Kelly, B., Baur, L., Chapman, K., Chapman, S., Gill, T., & King, L. (2014). Digital Junk: Food and Beverage Marketing on Facebook. *American Journal of Public Health*, 104(12), e56–e64. doi:10.2105/AJPH.2014.302167 PMID:25322294
- Goffman, E. (1959). *The Presentation of Self in Everyday Life*. New York: Anchor Books Doubleday.

- Grusec, J. E. (1992). Social learning theory and developmental psychology: The legacies of Robert Sears and Albert Bandura. *Developmental Psychology*, 28(5), 776–786. doi:10.1037/0012-1649.28.5.776
- Herman, C. P., Roth, D. A., & Polivy, J. (2003). Effects of the Presence of Others on Food Intake: A Normative Interpretation. *Psychological Bulletin*, 129(6), 873–886. doi:10.1037/0033-2909.129.6.873 PMID:14599286
- Hingle, M., Yoon, D., Fowler, J., Kobourov, S., Schneider, L. M., Falk, D., & Burd, R. (2013). Collection and Visualization of Dietary Behavior and Reasons for Eating Using Twitter. *Journal of Medical Internet Research*, 15(6), e125. doi:10.2196/jmir.2613 PMID:23796439
- Holmberg, C. (2015). Politicization of the Low-Carb High-Fat Diet in Sweden, Promoted On Social Media by Non-Conventional Experts. *International Journal of E-Politics*, 6(3), 27–42. doi:10.4018/ijep.2015070103
- Holmberg, C. (2016). If You cant beat It-Use It: Why and how clinicians need to consider social media in the treatment of adolescents with obesity. *European Journal of Clinical Nutrition*, 70(9), 977–978. doi:10.1038/ejcn.2016.126 PMID:27601136
- Holmberg, C., Chaplin, E. J., Hillman, T., & Berg, C. (2016). Adolescents presentation of food in social media: An explorative study. *Appetite*, 99, 121–129. doi:10.1016/j.appet.2016.01.009 PMID:26792765
- Hu, Y., Manikonda, L., & Kambhampati, S. (2014). *What We Instagram: A First Analysis of Instagram Photo Content and User Types*. Paper presented at the International AAAI Conference on Weblogs and Social Media.
- Jenkins, H., Ford, S., & Green, G. (2013). *Spreadable Media: Creating Value and Meaning in a Networked Culture*. New York: NYU Press.
- Kelly, B., Vandevijvere, S., Freeman, B., & Jenkin, G. (2015). New Media but Same Old Tricks: Food Marketing to Children in the Digital Age. *Current Obesity Reports*, 4(1), 37–45. doi:10.1007/s13679-014-0128-5 PMID:26627088
- Korthals, M. (2008). Food as a Source and Target of Metaphors: Inclusion and Exclusion of Foodstuffs and Persons through Metaphors. *Configurations*, 16(1), 77–92. doi:10.1353/con.0.0044
- Krämer, N. C., & Winter, S. (2008). Impression management 2.0: The relationship of self-esteem, extraversion, self-efficacy, and self-presentation within social networking sites. *Journal of Media Psychology: Theories, Methods, and Applications*, 20(3), 106–116. doi:10.1027/1864-1105.20.3.106
- Kümpel Nørgaard, M., Nørgaard Hansen, K., & Grunert, K. G. (2013). Peer influence on adolescent snacking. *Journal of Social Marketing*, 3(2), 176–194. doi:10.1108/JSOCM-06-2012-0028
- Lally, P., Bartle, N., & Wardle, J. (2011). Social norms and diet in adolescents. *Appetite*, 57(3), 623–627. doi:10.1016/j.appet.2011.07.015 PMID:21843568
- Lenhart, A., Smith, A., Anderson, M., Duggan, M., & Perrin, A. (2015). *Teens, Technology and Friendships: Video Games, Social Media and Mobile Phones Play an Integral Role in How Teens Meet and Interact With Friends*. Retrieved from <http://www.pewinternet.org/files/2015/08/Teens-and-Friendships-FINAL2.pdf>
- Lohse, B. (2013). Facebook is an effective strategy to recruit low-income women to online nutrition education. *Journal of Nutrition Education and Behavior*, 45(1), 69–76. doi:10.1016/j.jneb.2012.06.006 PMID:23305805
- Mauss, M. (1967). *The gift: forms and functions of exchange in archaic societies*. New York: W. W. Norton & Company.

- McBride, D. L. (2011). Risks and Benefits of Social Media for Children and Adolescents. *Journal of Pediatric Nursing*, 26(5), 498–499. doi:10.1016/j.pedn.2011.05.001 PMID:21930038
- Montgomery, K., & Chester, J. (2011). *Digital Food Marketing to Children and Adolescents: Problematic Practices and Policy Interventions*. Retrieved from http://digitalads.org/sites/default/files/publications/digitalads_digital_marketing_report_final_web_20111017.pdf
- Neely, E., Walton, M., & Stephens, C. (2014). Young Peoples Food Practices and Social Relationships. A Thematic Synthesis. *Appetite*, 82(0), 50–60. doi:10.1016/j.appet.2014.07.005 PMID:25017130
- Perkins, J. M., Perkins, H. W., & Craig, D. W. (2010). Peer weight norm misperception as a risk factor for being over and underweight among UK secondary school students. *European Journal of Clinical Nutrition*, 64(9), 965–971. doi:10.1038/ejcn.2010.106 PMID:20628407
- Pew Research Center. (2015). *Teens, Social Media & Technology Overview 2015*. Retrieved from http://www.pewinternet.org/files/2015/04/PI_TeensandTech_Update2015_0409151.pdf
- Pliner, P., Bell, R., Hirsch, E. S., & Kinchla, M. (2006). Meal duration mediates the effect of social facilitation on eating in humans. *Appetite*, 46(2), 189–198. doi:10.1016/j.appet.2005.12.003 PMID:16500000
- Roper, S., & La Niece, C. (2009). The Importance of Brands in the Lunch-Box Choices of Low-Income British School Children. *Journal of Consumer Behaviour*, 8(2/3), 84–99. doi:10.1002/cb.275
- Salvy, S. J. (2009). Social Influence on Youth Food Intake and Choice of Activities. *Appetite*, 52(3), 856. doi:10.1016/j.appet.2009.04.170
- Salvy, S. J., de la Haye, K., Bowker, J. C., & Hermans, R. C. J. (2012). Influence of peers and friends on childrens and adolescents eating and activity behaviors. *Physiology & Behavior*, 106(3), 369–378. doi:10.1016/j.physbeh.2012.03.022 PMID:22480733
- Schneider, E. P., McGovern, E. E., Lynch, C. L., & Brown, L. S. (2013). Do Food Blogs Serve as a Source of Nutritionally Balanced Recipes? An Analysis of 6 Popular Food Blogs. *Journal of Nutrition Education and Behavior*, 45(6), 696–700. doi:10.1016/j.jneb.2013.07.002 PMID:24206585
- Simunaniemi, A.-M., Sandberg, H., Andersson, A., & Nydahl, M. (2011). Laypeople Blog About Fruit and Vegetables for Self-Expression and Dietary Influence. *Health Communication*, 26(7), 621–630. doi:10.1080/10410236.2011.561520 PMID:21541865
- Stok, F. M., de Ridder, D. T., de Vet, E., & de Wit, J. B. (2013). Don't tell me what I should do, but what others do: The influence of descriptive and injunctive peer norms on fruit consumption in adolescents. *British Journal of Health Psychology*. doi:10.1111/bjhp.12030 PMID:23406475
- Swedish Media Council. (2015). *English summary: Kids & Media 2015*. Retrieved from <http://www.webcitation.org/6jmRir1SJ>
- van Meer, F., van der Laan, L. N., Adan, R. A., Viergever, M. A., & Smeets, P. A. (2015). What you see is what you eat: An ALE meta-analysis of the neural correlates of food viewing in children and adolescents. *NeuroImage*, 104, 35–43. doi:10.1016/j.neuroimage.2014.09.069 PMID:25285373
- Vartanian, L. R., Herman, C. P., & Polivy, J. (2007). Consumption stereotypes and impression management: How you are what you eat. *Appetite*, 48(3), 265–277. doi:10.1016/j.appet.2006.10.008 PMID:17157957

Vuk Pisk, S., Mihanović, M., Silic, A., Bogović, A., & Vidovic, V. (2012). Self-concept in overweight adolescents. *South African Journal of Psychiatry*, 18(1), 27–31.

Wouters, E. J., Larsen, J. K., Kremers, S. P., Dagnelie, P. C., & Geenen, R. (2010). Peer influence on snacking behavior in adolescence. *Appetite*, 55(1), 11–17. doi:10.1016/j.appet.2010.03.002 PMID:20211671

ADDITIONAL READING

Counihan, C., & Van Esterik, P. (2013). *Food and Culture: A Reader* (Vol. 3). New York: Routledge.

de la Haye, K., Robins, G., Mohr, P., & Wilson, C. (2010). Obesity-related behaviors in adolescent friendship networks. *Social Networks*, 32(3), 161–167. doi:10.1016/j.socnet.2009.09.001

Stead, M., McDermott, L., MacKintosh, A. M., & Adamson, A. (2011). Why healthy eating is bad for young peoples health: Identity, belonging and food. *Social Science & Medicine*, 72(7), 1131–1139. doi:10.1016/j.socscimed.2010.12.029 PMID:21429646

Valentine, G. (1999). Eating in: Home, consumption and identity. *The Sociological Review*, 47(3), 491–524. doi:10.1111/1467-954X.00182

KEY TERMS AND DEFINITIONS

Impression Management: Conscious or sub-conscious procedure in which adolescents attempt to influence the perceptions of themselves. This is done by regulating and adjusting information in social interaction.

Online Social Networks: Online services that permits individuals to construct a public or semi-public profile within a network, and specify other users with whom they share a connection.

Social Facilitation: A theory that suggests that there is an improvement in performance produced by the actual, imagined or implied presence of others.

Social Media: Umbrella term for Internet-based services and applications that permit its users to construct individual personal profiles, generate and access searchable online content (e.g., images, videos), form online connections with other users, and view these social connections.

Social Modeling: A theory that stipulates that we learn to imitate others by observing their behavior and that we learn to behave in certain ways by watching others do what they do.

Agent-Based Social Networks

Federico Bergenti

Università degli Studi di Parma, Italy

Agostino Poggi

Università degli Studi di Parma, Italy

Michele Tomaiuolo

Università degli Studi di Parma, Italy

INTRODUCTION

Social networking sites have deeply changed the face of the web in the last decade. Social web sites have attracted users with very weak interest in technology, including people that before the social networking revolution were not even regular users of other popular Internet services and computers in general (Stroud, 2008). The phenomenon is so widespread that many people started using social networking systems to ask questions directly to people instead of querying search engines (Morris et al., 2010) and in place of regular email. Moreover, some of the largest social networking sites constitute a separate and closed network (Sabbag, 2011). After the huge success of the early social networking systems, many other players came in the social networking market and nowadays hundreds of different social networking systems exist. Even if the social networking systems are greatly dissimilar in their user base and functionality, they are almost always centralized systems. The centralized nature allows a simple browser-based user experience and, moreover, many algorithms, e.g., friend suggestion, are far easier and more efficient to implement in this setting. However, such a solution has many drawbacks, e.g., lack of privacy, lack of anonymity, risks of censorship and operating costs. Therefore, a decentralized architecture may offer several advantages, but it may involve some security and design issues that are not present in

traditional social networks. This chapter has the goal of introducing social networks and how they can take advantages of a decentralized architecture and how the use of multi agent technologies help to cope with its security and design issues and, in addition, to support the creation of innovative services for its users.

BACKGROUND

Social networks can be defined as structures consisting of a finite set of actors and the relation or relations defined on them, where an actor is simply a discrete individual or a social unit (Wasserman & Faust, 1994). Social networking systems are implemented as web-sites allowing users to have a profile and managing their online social network, i.e., such systems allow them to: (i) construct a profile which represents them in the system; (ii) create a list of users with whom they share a connection and (iii) navigate their list of connections and that of their friends (Ellison, 2007).

Although we agree that self-presentation and social network management are extremely important and necessary components of a social networking system, we believe that the social networking revolution is far more related to the paradigm shift that transformed most people from mere consumers of information to full-fledged information producers. Most people create information which is essentially personal and, then, it

is mainly of interest for friends and acquaintances. Before the social web revolution such information used to be essentially lost in the web, while, nowadays social networking systems are able to deliver the information to the “right” people.

After the huge success of the early social networking systems, many other players came in the social networking market and nowadays hundreds of different social networking systems exist. Even if the social networking systems are greatly dissimilar in their user base and functionality, they are almost always centralized systems because of the access and implementation advantages.

A minor drawback is that scaling centralized systems to tens or hundreds of million of users is not an easy task. At any rate, we consider this drawback as a minor one, since the problem can be solved providing enough resources. However, the huge operative costs of supporting the infrastructure necessary to provide the service to millions of users can only be justified with robust business plans. While some social networking services have extremely differentiated business models (McGrath 2010), for most of them the primary source of income is advertisement and consequently they have a strong motive for: (i) using user provided data to increase performance for that purpose and (ii) even giving access to authorized commercial third parties to the raw data. This behavior poses serious threats to privacy and data protection issues, especially considering that there is no clear legislation on what uses of the user data are legitimate, and regarding the conditions for disclosing the data to third parties, especially when the subjects involved are from different countries.

Another problem is that social networking systems have terms of service that their users give to the system operators a non-exclusive, transferable, sub-licensable, royalty-free, worldwide license to the submitted content (Facebook, 2016; Twitter, 2016). Such terms are needed for legal reasons: in order to serve webpages containing the users’ data (e.g., their profile page) the service provider needs some rights over that data; it is nonetheless

true that users are essentially allowing the service providers to do with the data whatever they want for free. Moreover, most of the times, the users themselves have not easy and streamlined ways to obtain all the data they inserted in the system in a semantically meaningful or at least in a structured way; this (i) is a serious lock-in problem in its own right and (ii) hinders users’ trust in the platform. Most social networking platforms do not provide their users with easy and standard ways to export user submitted contents in a structured way. According to Fitzpatrick and Lueck (2010) this issue (i) is similar to a serious lock-in problem in its own right, and (ii) it hinders users’ trust in the platform.

The last problem with centralized social networking system is that service providers are in the position to effectively perform a-priori or a-posteriori censorship, or to disclose all the information they have, no matter how private, to other entities. They can perform such actions either motivated by selfish interests or forced under legal terms and other forms of pressure.

DECENTRALIZED SOCIAL NETWORKS

Peer-to-Peer (P2P) technologies are the most suitable mean to build decentralized social networks. In fact, P2P technologies allow the definition of open and decentralized overlay networks on top of the Internet that users can use for directly communicating to find and share resources, often music and movie files (Schollmeier, 2001). Such networks are one of the few largest distributed computing systems ever, and more surprisingly, they can run with great stability and resilient performance in face of possibly the most ferocious dynamics (Qiu & Srikant, 2004).

Thus, the use of P2P technologies for the development of social network is not only viable, but also highly desirable (Wang et al, 2006). First of all, P2P systems essentially achieve automatic resource scalability, in the sense that the avail-

ability of resources is proportional to the number of users. This property is especially desirable for media sharing social networking systems, considering the exceptionally high amount of resources needed. Secondly, the popularity over time of most content on such systems exhibits either a power-law or an exponential behavior and is consequently well suited for P2P distribution (Zink et al., 2009), possibly with fallback strategies for less popular content.

Regarding censorship issues, a P2P system essentially solves them by design. Without a central entity, nobody is in the position of censoring data systematically nor may be held legally responsible for the diffusion of censurable data: the sole owners and responsible of the data are the users themselves. Unfortunately, P2P systems, and especially Distributed Hash Table (DHT) based ones, may be still liable to attacks meant to disrupt the system functionality (Urdaneta et al., 2011), often based on the introduction of a large number of Sybil nodes and the diffusion of bogus information. However, the most popular DHT systems are significantly robust because of the high redundancy they achieve by using data replication and a redundant routing mechanism. Usually, the countermeasures are based on the embedding of the identity notion into the overlay level (Aiello & Ruffo, 2010), and on the use of some notion of “trust”, based on either certification authorities or some reputation mechanism (Aiello & Ruffo, 2012). Common consensus algorithms, including Byzantine agreement, have also been proposed and applied (Balfe et al., 2005; Anceaume et al, 2008).

Although peer-to-peer systems overcome the weakness of a single point of failure, there are some well-known and important vulnerabilities, including (i) Sybil attacks, or node insertion attacks, where multiple nodes are created in the network, each of them representing fictitious identities but all belonging to a single user; and (ii) publish attacks, based on index poisoning, where essentially some bogus content is deliberately spread to the index nodes responsible for other files or keywords. The presence of Sybil

nodes allows other attacks, e.g. routing attacks, eclipse attacks, storage attacks. Also publish attacks may exploit the presence of Sybil nodes, if available. Many of the proposed countermeasures to securing peer-to-peer networks are based on some notion of “trust” among peers. Depending on the approach they use to evaluate and manage trust relationships among peers, those countermeasures can be divided in two main groups: (i) credential and policy based, (ii) reputation based. Urdaneta et al. (2011) provide a detailed analysis of threats to DHTs, together with some proposed countermeasures. While underlining the existing vulnerability to Sybil attacks, authors conclude that “Current DHT deployments are not specifically designed to tolerate the presence of malicious nodes. However, most of them are based on Kademlia, which provides relative security by using data replication and a redundant routing mechanism similar to wide paths”.

Eventually, the lack of a central entity which has, or believes to have, interests in hindering interoperability (Shankland, 2011), creates the opportunity to design the system so that heterogeneous units can interoperate, typically providing a semantic common setting for the data.

Various solutions are being proposed to overcome the centralized architecture of the most widespread social networking platforms. Many of these proposals follow a federated approach, allowing users registered on a certain server to create relationships with users of other servers. Others are full-fledged peer-to-peer systems, usually based on a DHT.

Among the federated social networking systems, two of the best known are Diaspora (Bielenberg et al, 2012) and StatusNet (StatusNet, 2016). Diaspora servers communicate by means of an ad-hoc federation protocol and the standard Salmon protocol for comments. Users can (i) participate in the network by setting up their own server, which is named a “pod”, or (ii) exploit already existing pods. Using a number of existing protocols, StatusNet shows quite strong interoperability with other networks. With regards

to the completely distributed solutions, their origin can be traced back to Freenet (Charke et al., 2001), which is meant as a distributed, cooperative, uncensored and secure file system. It uses a “best-effort” unreliable routing algorithm to find content and namespaces, over both “Open-Net” and “DarkNet” connections. Various quite popular uncensored forums are built on Freenet, but they usually suffer a large amount of spam coming from anonymous sources. Similarly to other more recent systems.

Specifically in the field of social networking, various systems are being developed on the basis of peer-to-peer communications and DHT indexing. Among them, Maze (Chen et al., 2004), TRIBLER (Pouwelse et al., 2008), PeerSoN (Buchegger et al., 2009), Persona (Baden et al., 2009), Safebook (Cutillo et al., 2009) and DECENT (Jahid et al., 2012) are the most interesting.

Maze (Chen et al., 2004) supports a peer-to-peer social network through the use of some centralized services. It uses a ticketing server which issues tickets to all peers to identify them. This ticket is then served as a form of legitimate communication/transaction between peers. The ticket is valid for a single communication. For further communication, all peers need to contact the ticketing server for a ticket. Maze also uses another centralized server which, apart from holding a directory of peers, also checks online status of each of them.

TRIBLER (Pouwelse et al., 2008) is a P2P social-based file sharing network which is built on top of the Bittorrent protocol (Cohen, 2003). Tribler is based on the generation and maintenance of social networks in order to improve content discovery, searching and download performance. In particular, it proposes a decentralized recommendation mechanism based on standard collaborative filtering techniques and that takes advantage of the concepts of friends, friends-of-friends and tastes communities.

PeerSoN (Buchegger et al., 2009) is a system designed to provide encryption, decentralization and direct data exchange in the field of social

networks. The first prototype of PeerSoN is designed around a PKI, though some studies are being conducted for weakening this assumption. Each user has a unique ID, possibly computed as a hash of the user’s email. The DHT is used to trace the user’s network presence. An index file, containing a list of new content generated by the user, is also registered in the DHT.

Persona (Baden et al., 2009) is designed as a set of social networking services. It uses an interesting attribute-based encryption protocol for protecting access to users’ content. It allows each user to create various groups of “friends”, by assigning proper attribute credentials. Content can then be associated with a publication policy and made available only to a restricted audience.

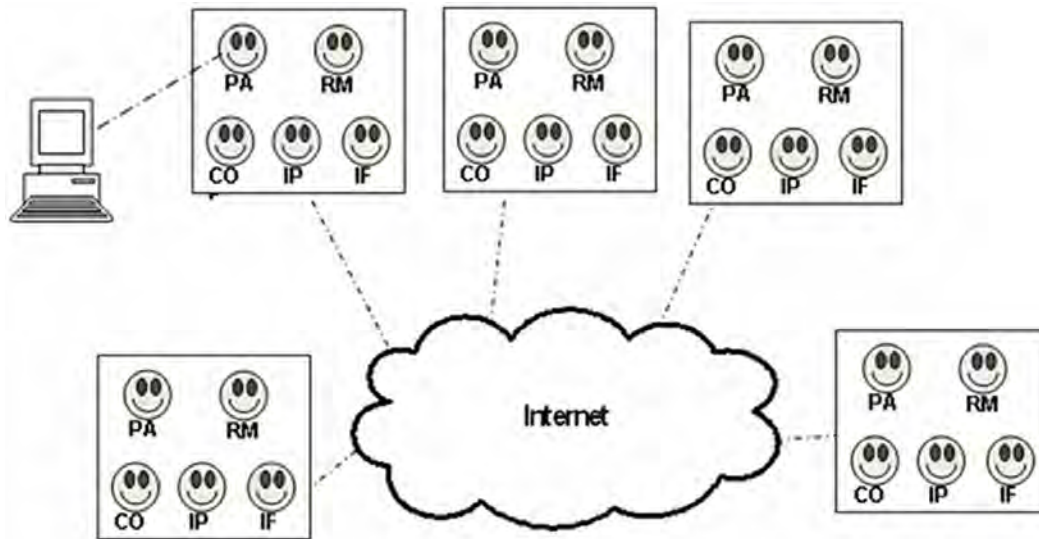
Safebook (Cutillo et al., 2009) is based on a DHT and a network of socially close peers, defined Matryoshka. Peers in a user’s Matryoshka are trusted and support the user by anonymizing communications and replicating content and profile information. Safebook exploits a more traditional certification authority. In fact, a user’s public key cannot be calculated from his identity, and all public/private key pairs are generated locally by the peers.

DECENT (Jahid et al., 2012) is a decentralized architecture for social networks that uses a distributed hash table to store user data, and offers cryptographic protections for confidentiality and integrity, as well as support for flexible attribute policies and fast revocation. DECENT ensures that neither data nor social relationships are visible to unauthorized users and provides availability through replication and authentication of updates.

AGENT-BASED SOCIAL NETWORKS

Agents and multi-agent systems are an interesting areas in software research that contributed to the development of the theory and the practice of complex distributed systems (Jennings et al., 1995). In particular, a multi-agent system offers the right abstraction for representing a decentralized and

Figure 1. An example of architecture of agent-based social network



peer-to-peer system (Moro et al., 2002) and may provide an appropriate framework for developing peer-to-peer applications (Koubarakis, 2003). Moreover, multi-agent systems share with social networks both the structure and the scope, since they are composed of individuals connected with some kinds of relationship and they are realized for accomplishing individual and/or common goals (Franchi & Poggi, 2012). Therefore, it is natural to think about synergies between social network and multi-agent system research and application and about the use of multi-agent coordination algorithms for the development of typical social network services. In fact, multi-agent technologies and techniques offer the necessary functions to the users of a social network such as security and trust supports (Yu & Singh, 2000; Franchi et al., 2013) and collaborative and organizational supports (Hattori et al., 1999; Yoshida et al., 2003). In particular, in an agent-based social network a user can take advantage of a set of agents that can help or even replace the user in her/his actions in the social network. For example, Figure 1 shows an example of architecture of agent-based social network; in particular, in such an architecture, each user can take advantage of a set of different personal agents: a personal assistant (PA) agent

helps the user in the interaction with the other users and with the other personal agents; a resource manager (RM) agent supports the management and sharing of resources with other users; the information filtering (IF) and information pushing (IP) agents help the user to find the relevant information and to exchange messages when she/he is offline; finally, the collaborative and organizational agent supports the coordination with the other users.

Several interesting works demonstrate how multi-agent systems are a suitable means for implementing social networks and their services. ReferralWeb (Kautz et al., 1997), Yenta (Foner, 1997), Community Organizer (Hattori et al., 1999), Shine (Yoshida et al., 2003), MARS (Yu & Singh, 2003), Blogracy (Franchi et al., 2013) and MASCRM (Olszak, 2013) are the most interesting.

ReferralWeb is an agent based interactive system for reconstructing, visualizing, and searching social networks on the World-Wide Web, whose main focus is selecting an expert of a given field in one's (extended) social network (Kautz et al., 1997). In ReferralWeb a social network is modeled by a graph, where the nodes represent individuals, and an edge between nodes indicates that a direct relationship between the individuals has been dis-

covered. For ReferralWeb a direct relationship is implied when the names are in close proximity in any documents publicly available on the Web, e.g., home pages, co-authorship in published papers or organization charts in institutional websites. The constructed network is then used to guide the search for people or documents in response to user queries; a person can: i) ask to find the chain between himself/herself and a named individual; ii) search for an expert in a given topic and providing a maximum social radius (the number of “links” in the chain connecting the person performing the query with the expert); iii) request a list of documents written by people “close” to a given expert.

Yenta is a matchmaking system that helps people with similar interests to get in touch (Foner, 1997). Yenta agents do not query the web; instead, they scan user’s emails, Usenet posts and (possibly) documents in order to discover their users’ interests and hobbies. The idea is that many potentially interesting people do not publicly write and are consequently invisible to tools relying on public data. Collected data are then used to introduce users’ to each other. Considering that in the nineties web communities were built around the idea of common interests rather than personal acquaintance, the system was a truly distributed social networking system for the time.

Community Organizer is a system where agents help the users by gathering and exchanging information, visualizing contexts, and recommending or assisting their users in making a choice (Hattori et al., 1999). Each user has a personal agent and a set of additional community agents have the function of providing shared information, knowledge, or contexts within the community and act as mediators for informal communications between people. In particular, each personal agent acquires the user profile and visualizes potential communities around the user. The community agent collects the user profiles and maintains information on potential communities. Upon a request from a personal agent, the community agent first computes potential communities around the owner of the personal agent, and then sends the necessary

data (users in the potential communities and their relevance) to the personal agent.

Shine is a fully peer-to-peer framework for network community support (Yoshida et al., 2003). The framework also provides design guidelines and enables different applications to share program components and cooperate and features a peer-to-peer architecture through which personal agents can flexibly form communities where users can exchange information with peer agents. Essentially Shine is a middleware for collaborative workspaces especially tailored to implement various collaborative workspaces. Agents in Shine are goal-driven through plans: a plan is description of agent action rules. Multiple plans are executed concurrently in the plan execution module of each agent. Some plans are prepared to perform services of applications while other plans are provided by Shine to do fundamental or common tasks. A plan acts in response to external events, such as receiving a message from another agent, user input or notification about a change in the community.

MARS is a multi-agent referral system that finds experts on the basis of personal agents able to learn the user’s preferences and interests and able to build an expertise model of the other users on the basis of their responses (Yu & Singh, 2003). The expertise model is captured through a classical vector space model (Salton & McGill, 1983) and each personal agent maintains the models of its neighbors. In particular, the model is updated by a personal agent on the basis of the responses of its neighbors. The response can be an answer of its user or a referral: if the agent is reasonably confident about the expertise of its user matches the query, it directly answers; otherwise, it suggests a referral to another personal agent. Since the number of neighbors is bounded, some of them will be discarded to make place for new ones.

Blogracy is a peer-to-peer, anonymous and uncensorable social networking platform (Franchi et al., 2013). The architecture of the platform is modular and is built around a module for basic file sharing and DHT operations, possibly exploiting

an existing implementation, and another module providing a set of social services to the local user through a Web interface. Moreover, the platform provides two additional agent based modules respectively providing a set of pervasive services and a set of information retrieval and pushing services. In particular, the current prototype of Blogracy takes advantage of; i) Vuze (Vuze, 2016), a popular BitTorrent client implemented in Java and available as open source software, for implementing the file sharing and DHT operations, ii) Open Social (OpenSocial and Gadgets Specification Group, 2016), a set of APIs supporting the sharing of social data, for implementing the social services, and iii) JADE (Bellifemine et al., 2008), probably the most known agent development environment enabling the integration of agents and both knowledge and Internet-oriented technologies, for implementing the agent-based services.

MASCRM (Olszak, 2013) is a system based on a group of agents that independently acquire a number of important data necessary to manage the customer relationship by working on the top with a variety of social networks. Agents are able to communicate with each other and activate mutually. In particular there are able to collect a lot of data on the behaviour of the customers, their purchase preferences, etc. A prototype of the system has been experiment on the Facebook network a showed how agents work supports organizations to quickly respond to the opinions and attitudes of their current and future customers, and also to quickly assess the relevance of their marketing actions and decisions.

FURTHER RESEARCH DIRECTIONS

As illustrated above, while decentralized architectures can help in overcoming some of the problems of the most widespread centralized social networking platforms, multi-agent systems can be used for providing advanced social services on such a architectures. In particular,

multi-agent systems might become one of the most important means for the development of intelligent services for social networks and for coordinating the activities of its users. However, a lot of work is necessary for updating and experiment the coordination, knowledge management, and learning capabilities provided by multi-agent systems for the development of distributed services for helping the members in their activities and for, when will be necessary, using agents as delegate of some members for exchanging information and performing tasks.

CONCLUSION

Social networking sites are deeply changing the face of the Web, but their architecture might be improved. In fact, such systems utilize a traditional client-server architecture that relegates all the information in central servers. Although this approach supports highly mobile user access since users can log-in from any web browser, it also presents many drawbacks, e.g., lack of privacy, lack of anonymity, risks of censorship and operating costs. The integration between peer-to-peer technologies and multi-agent systems may be used for developing social networks that do not present the previous drawbacks. Moreover, the use of m multi-agent systems is the right solution to offer strong coordination techniques to the users of social networks and provide them more sophisticated and usable services.

REFERENCES

Aiello, L. M., & Ruffo, G. (2010). Secure and flexible framework for decentralized social network services. In *Proceedings of the IEEE International Conference on Pervasive Computing and Communications* (pp. 594-599). Mannheim, Germany: IEEE. doi:10.1109/PERCOMW.2010.5470506

- Aiello, L. M., & Ruffo, G. (2012). LotusNet: Tunable privacy for distributed online social network services. *Computer Communications*, 35(1), 75–88. doi:10.1016/j.comcom.2010.12.006
- Anceaume, E., Ludinard, R., Ravoaja, A., & Brasileiro, F. (2008). Peercube: A hypercube-based P2P overlay robust against collusion and churn. In *Proceedings of the Second IEEE International Conference on Self-Adaptive and Self-Organizing Systems* (pp. 15-24). Venice, Italy: IEEE.
- Baden, R., Bender, A., Spring, N., Bhattacharjee, B., & Starin, D. (2009). Persona: an online social network with user-defined privacy. In *Proceedings of the ACM conference on Data communication* (pp. 135-146). Barcelona, Spain: ACM. doi:10.1145/1592568.1592585
- Balfe, S., Lakhani, A. D., & Paterson, K. G. (2005). Trusted computing: Providing security for peer-to-peer networks. In *Proceedings of the Fifth IEEE International Conference on Peer-to-Peer Computing* (pp. 117-124). Konstanz, Germany: IEEE. doi:10.1109/P2P.2005.40
- Bellifemine, F., Caire, G., Poggi, A., & Rimassa, G. (2008). JADE: A Software Framework for Developing Multi-Agent Applications. Lessons Learned. *Information and Software Technology Journal*, 50(1-2), 10–21. doi:10.1016/j.infsof.2007.10.008
- Bielenberg, A., Helm, L., Gentilucci, A., Stefanescu, D., & Zhang, H. (2012, March). The growth of diaspora-a decentralized online social network in the wild. In *Computer Communications Workshops* (pp. 13–18). IEEE. doi:10.1109/INFCOMW.2012.6193476
- Buchegger, S., Schiöberg, D., Vu, L., & Datta, A. (2009). PeerSoN: P2P social networking: early experiences and insights. In *Proceedings of the Second ACM EuroSys Workshop on Social Network Systems* (pp. 46–52). Nuremberg, Germany: ACM.
- Chen, H., Yang, M., Han, J., Deng, H., & Li, X. (2004). Maze: A Social Peer-to-Peer Network. In *Proceedings of the IEEE International Conference on E-Commerce Technology for Dynamic E-Business* (pp. 290-293). Beijing, China: IEEE. doi:10.1109/CEC-EAST.2004.44
- Clarke, I., Sandberg, O., Wiley, B., & Hong, T. (2001). Freenet: A distributed anonymous information storage and retrieval system. In H. Federrath (Ed.), *Designing Privacy Enhancing Technologies, Lecture Notes in Computer Science* (Vol. 2009, pp. 46–66). New York, NY, USA: Springer. doi:10.1007/3-540-44702-4_4
- Cohen, B. (2003). Incentives build robustness in BitTorrent. *Proceedings of the First Workshop on Economics of Peer-to-Peer Systems*, 68-72.
- Cuttillo, L. A., Molva, R., & Strufe, T. (2009). Safebook: A Privacy Preserving Online Social Network Leveraging on Real-Life Trust. *IEEE Communications Magazine*, 47(12), 94–101. doi:10.1109/MCOM.2009.5350374
- Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1), 210–230. doi:10.1111/j.1083-6101.2007.00393.x
- Facebook. (2016). *Facebook statement of rights and responsibilities*. Retrieved March 25, 2016, from <http://www.facebook.com/legal/terms>
- Fitzpatrick, B. W., & Lueck, J. J. (2010). The case against data lock-in. *Communications of the ACM*, 53(11), 42–46. doi:10.1145/1839676.1839691
- Foner, L. (1997). Yenta: a multi-agent, referral-based matchmaking system. In *Proceedings of the First International Conference on Autonomous Agents* (pp. 301-307). Marina del Rey, CA: ACM. doi:10.1145/267658.267732

- Franchi, E., & Poggi, A. (2012). Multi-Agent Systems and Social Networks. In M. M. Cruz-Cunha, G. D. Putnik, N. Lopes, P. Gonçalves, & E. Miranda (Eds.), *Handbook of Research on Business Social Networking: Organizational, Managerial, and Technological Dimensions* (pp. 84–97). Hershey, PA: Information Science Reference. doi:10.4018/978-1-61350-168-9.ch005
- Franchi, E., Poggi, A., & Tomaiuolo, M. (2013). Open social networking for online collaboration. *International Journal of e-Collaboration*, 9(3), 50–68. doi:10.4018/jec.2013070104
- Hattori, F., Ohguro, T., Yokoo, M., Matsubara, S., & Yoshida, S. (1999). Socialware: Multiagent systems for supporting network communities. *Communications of the ACM*, 42(3), 55–61. doi:10.1145/295685.295701
- Jahid, S., Nilizadeh, S., Mittal, P., Borisov, N., & Kapadia, A. (2012). DECENT: A decentralized architecture for enforcing privacy in online social networks. In *Proceedings of the IEEE International Conference on Pervasive Computing and Communications* (pp. 326–332). Lugano Switzerland: IEEE. doi:10.1109/PerComW.2012.6197504
- Jennings, N., Corera, J., & Laresgoiti, I. (1995). Developing industrial multi-agent systems. In *Proceedings of the First International Conference on Multi-Agent Systems* (pp. 423–430). San Francisco, CA: Kluwer.
- Kautz, H., Selman, B., & Shah, M. (1997). Combining Social Networks and Collaborative Filtering. *Communications of the ACM*, 40(3), 63–65. doi:10.1145/245108.245123
- Koubarakis, M. (2003). Multi-agent systems and peer-to-peer computing. In M. Klusch, S. Ossowski, Omicini, & H. Laamanen (Eds.), *Methods, systems, and challenges. Cooperative Information Agents VII, Lecture Notes in Computer Science* (Vol. 2782, pp. 46–61). Berlin: Springer.
- McGrath, R. G. (2010). Business Models: A Discovery Driven Approach. *Long Range Planning*, 43(2-3), 247–261. doi:10.1016/j.lrp.2009.07.005
- Moro, G., Ouksel, A. M., & Sartori, C. (2002). Agents and peer-to-peer computing: a promising combination of paradigms. In *Agents and Peer-to-Peer Computing* (pp. 1–14). Berlin, Germany: Springer.
- Morris, M., Teevan, T., & Panovich, K. (2010). What do people ask their social networks, and why?: a survey study of status message Q&A behavior. In *Proceedings of the 28th International Conference on Human Factors in Computing Systems* (pp. 1739–1748). Atlanta, GA: ACM. doi:10.1145/1753326.1753587
- Olszak, C. (2013). Multi-agent framework for social customer relationship management systems. In *Proceedings of the Informing Science and Information Technology Education Conference* (Vol. 2013, No. 1, pp. 367–387). Porto, Portugal: Academic Press.
- OpenSocial and Gadgets Specification Group. (2016). *Open Social Specifications*. Retrieved March 25, 2016, from <https://github.com/OpenSocial/spec>
- Pouwelse, J. A., Garbacki, P., Wang, J., Bakker, A., Yang, J., Iosup, A., & Sips, H. J. et al. (2008). TRIBLER: A social-based peer-to-peer system. *Concurrency and Computation*, 20(2), 127–138. doi:10.1002/cpe.1189
- Qiu, D., & Srikant, R. (2004). Modeling and performance analysis of BitTorrent-like peer-to-peer networks. *Computer Communication Review*, 34(4), 367–378. doi:10.1145/1030194.1015508
- Sabbag, D. (2011). *Google evangelist warns Facebook could be the next AOL or IBM*. Retrieved March 25, 2016, from <http://www.guardian.co.uk/technology/2011/sep/21/google-facebook-vint-cerf>

Salton, G., & McGill, M. (1983). *Introduction to modern information retrieval*. New York, NY: McGraw Hill.

Schollmeier, R. (2001). A definition of peer-to-peer networking for the classification of peer-to-peer architectures and applications. In *Proceedings of the First International Conference on Peer-to-Peer Computing* (pp. 101–102). Linköping, Sweden: Academic Press.

Shankland, S. (2011). *Facebook blocks contact exporting tool*. Retrieved March 25, 2016, from http://news.cnet.com/8301-30685_3-20076774-264/facebook-blocks-contact-exporting-tool/

StatusNet. (2016). *Status Net and GNU Social Network*. Retrieved March 25, 2016, from <http://www.gnu.org/software/social/>

Stroud, D. (2008). Social networking: An age-neutral commodity - Social networking becomes a mature web application. *Journal of Direct Data and Digital Marketing Practice*, 9(3), 278–292. doi:10.1057/palgrave.ddmp.4350099

Twitter. (2013). *Twitter terms of service*. Retrieved March 25, 2016, from <http://twitter.com/tos>

Urdaneta, G., Pierre, G., & Van Steen, M. (2011). A survey of DHT security techniques. *ACM Computing Surveys*, 43(2), 1–49. doi:10.1145/1883612.1883615

Vuze. (2016). *Vuze Software*. Retrieved March 25, 2016, from <http://www.vuze.com/>

Wang, F., Moreno, Y., & Sun, Y. (2006). Structure of peer-to-peer social networks. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, 73(3), 1–8. doi:10.1103/PhysRevE.73.036123 PMID:16605614

Wasserman, S., & Faust, K. (1994). *Social network analysis: methods and applications*. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511815478

Yoshida, S., Kamei, K., Ohguro, T., & Kuwabara, K. (2003). Shine: A peer-to-peer based framework of network community support systems. *Computer Communications*, 26(11), 1199–1209. doi:10.1016/S0140-3664(02)00254-2

Yu, B., & Singh, M. (2003). Searching social networks. In *Proceedings of the Second International Joint Conference on Autonomous Agents and Multi Agent Systems* (pp. 65–72). Melbourne, Australia: ACM. doi:10.1145/860575.860587

Yu, B., & Singh, M. P. (2000). A social mechanism of reputation management in electronic communities. In *Cooperative Information Agents IV-The Future of Information Agents in Cyberspace* (pp. 154–165). Berlin, Germany: Springer. doi:10.1007/978-3-540-45012-2_15

Zink, M., Suh, K., Gu, Y., & Kurose, J. (2009). Characteristics of YouTube network traffic at a campus network - Measurements, models, and implications. *Computer Networks*, 53(4), 501–514. doi:10.1016/j.comnet.2008.09.022

ADDITIONAL READING

Androutsellis-Theotokis, S., & Spinellis, D. (2004). A survey of peer-to-peer content distribution technologies. *ACM Computing Surveys*, 36(4), 335–371. doi:10.1145/1041680.1041681

Bordini, R., Dastani, M., Dix, J., & Fallah-Seghrouchni, A. (Eds.). (2005). *Multi-Agent Programming: Languages, Platforms and Applications. Multiagent Systems, Artificial Societies, and Simulated Organizations*. Berlin, Germany: Springer. doi:10.1007/b137449

Carrington, P.J., Scott, J., & Wasserman, S. (Eds.). (2005). *Models and methods in social network analysis*. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511811395

Ferber, J. (1999). *Multi-agent systems: an introduction to distributed artificial intelligence*. Reading, MA: Addison-Wesley.

Muller, J. (1998). Architectures and applications of intelligent agents: A survey. *The Knowledge Engineering Review*, 13(4), 353–380. doi:10.1017/S0269888998004020

KEY TERMS AND DEFINITIONS

Anonymity: State of being unknown or unacknowledged to the others.

Censorship: Modification or suppression of speech or other public communication which may

be considered objectionable, harmful, sensitive, or inconvenient as determined by a controlling body.

Multi-Agent System: A loosely coupled network of software agents that interact to solve problems that are beyond the individual capacities or knowledge of each software agent.

Peer-to-Peer System: A network based system in which each node can act as both client and server for the other ones of the system.

Privacy: The right to be secluded from the presence or view of others.

Social Networking System: A network based system facilitating the building of social networks.

Software Agent: A computer program that is situated in some environment and capable of autonomous action in order to meet its design objectives.

Aspects of Various Community Detection Algorithms in Social Network Analysis

S

Nicole Belinda Dillen

St. Thomas' College of Engineering and Technology, India

Aruna Chakraborty

St. Thomas' College of Engineering and Technology, India

INTRODUCTION

The 21st century is dominated by social media. Social networking websites like Facebook, Twitter and Myspace connect millions of people from all corners of the world. As a result, the global community is growing at a fast pace with new connections being forged every second. Businesses, government corporations and other organizations are relying increasingly on online networking in order to promote products, conduct surveys and target specific audiences. The immense importance of social media has catalyzed research in social network analysis, a field that had gained prominence in the mid '90s. Social network analysis involves the detection of communities or closely connected groups of individuals, studying the characteristic patterns associated with these communities, identifying important "key" actors in these networks and finally studying the overall behavior associated with people in the network.

BACKGROUND

Community detection in social networks is one of the most important areas of social network analysis. The most widely accepted definition of a community is a closely connected group of individuals with sparse connections to individuals belonging to other communities. With this definition in mind several algorithms have been implemented which serve to identify community structure in social networks.

Algorithms such as those proposed by Kernighan (1970) are traditional graph partitioning algorithms which use clustering techniques like *k*-means (MacQueen, 1967) to form graph clusters and can, therefore, be applied on social networks. The problem with these algorithms is that they can only find either a fixed number of communities or community clusters of a certain size, both of which may not really be the best case for real-world scenarios.

A very popular algorithm was developed by Girvan and Newman (2002) and uses a divisive technique in which edges are constantly removed to split a social network into smaller and smaller component networks. This technique gives rise to the natural community structure inherent in the network which can be represented through a community hierarchy. However, certain issues such as time complexity prompted one of the authors to devise an entirely new approach to detect community structure by optimizing a function known as *modularity* (Newman, 2004). This technique paved the way for a number of subsequent algorithms that collectively constitute the class of modularity-optimization algorithms. These include an improvement on Newman's fast algorithm (Clauset, Newman, & Moore, 2004) and an effort to improve community quality by balancing the communities detected by the CNM method (Wakita, & Tsurumi, 2007). The Louvain method (Blondel, Guillaume, Lambiotte, & Lefebvre, 2008) is, perhaps, one of the most popular greedy modularity-optimization algorithms which

DOI: 10.4018/978-1-5225-2255-3.ch603

is modelled on Newman's original modularity-based algorithm.

Further research developments led to the idea that community structure may not consist of just well-defined “crisp” clusters; rather, most real-world social networks contain communities that naturally overlap to a certain extent. Overlapping community detection algorithms serve to identify this natural community structure and either allocate the overlapping nodes to a single “best-fit” community through modularity optimization (Berti, Sperduti, & Burattin, 2014) or preserve the overlapping structure through the use of fuzzy membership functions (Kundu & Pal, 2015b).

COMMUNITY DETECTION ALGORITHMS: NEW AND OLD

Traditional View of Communities in Social Networks

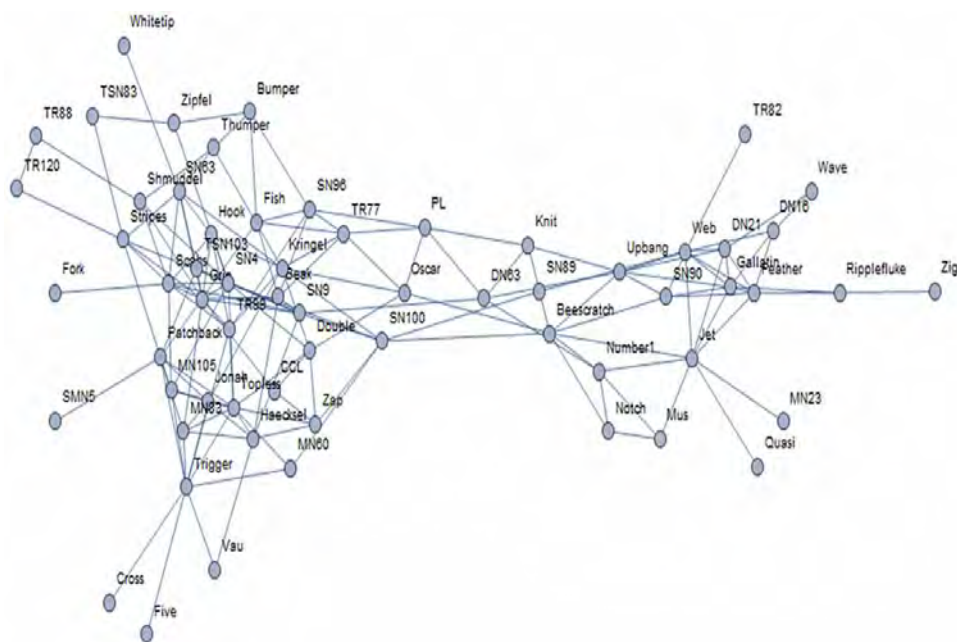
Most community detection algorithms consider a model in which each individual in a network belongs to a single community. The individuals

in this community will have many connections with each other but will have a minimal number of connections to individuals belonging to other communities. In fact, this very phenomenon is exploited by nearly all of the community detection algorithms prevalent in social network analysis.

Overlapping Communities in Social Networks

Contrary to the traditional community model mentioned above, many recent research developments deal with more realistic models in which multiple communities may be assigned to an individual. In real world social networks, it is only natural for an individual to be a part of several communities simultaneously. For example, a student may be part of his/her study group consisting of close friends, and at the same time belong to the school's debate team which is an entirely different community. Many community detection algorithms fail to take this relevant aspect into consideration and are, hence, limited in their approach. The newer models, therefore, describe what are known as “overlapping” communities.

Figure 1. A graph representing the Dolphin Social Network: lines indicate edges and circles indicate vertices



Algorithms pertaining to both the models described above will be discussed shortly, but first some terminology that will be used throughout this article, along with their formal definitions, will be described below.

Terminology

1. **Graph:** A graph is a network of nodes and their interconnections. Each node in the graph is known as a vertex and each connection between two vertices is known as an edge (or link).
2. **Social Network:** A social network is a set of individuals that are interconnected in some manner. The criteria for the interconnections may vary: for example, a link in one network may be formed between two individuals if they are friends, while in another network a link may constitute the presence of a common interest or activity. Formally, a social network may be represented by a graph $G(V, E)$, where V is the set of vertices (individuals) and E is the set of edges. Figure 1 depicts a social network represented by a graph.
3. **Community:** A community is a subset of nodes in a social network that have dense connections between them and which are sparsely connected to nodes belonging to other communities.
4. **Edge Betweenness:** The betweenness of an edge (Girvan & Newman, 2002) is the number of shortest paths between vertices that contain the edge.
5. **Modularity:** Modularity (Newman, 2004) is a measure of the quality of community partitions formed by an algorithm. It is the difference between the actual density of intra-community edges and the corresponding connections in a random network possessing the same degree distribution as that of the actual network. The measure was given by:

$$Q = \sum_i \left(e_{ii} - a_i^2 \right) \quad (1)$$

where, e_{ij} is the number of edges connecting nodes in community i to those in community j , and a_i is the total number of edges that have one end in community i .

Equation 1 was extended to accommodate weighted graphs in (Blondel, Guillaume, Lambiotte, & Lefebvre, 2008):

$$Q = \frac{1}{2m} \sum_{i,j} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j) \quad (2)$$

where, A_{ij} is the weight of the edge between vertices i and j , k_i is the total weight of all edges linked to i , c_i is the community to which i is assigned and $\delta(c_i, c_j)$ is 1 when both i and j belong to the same community and is 0 otherwise. The total weight of all edges in the network is m , where

$$m = \frac{1}{2} \sum_{i,j} A_{ij}.$$

6. **Role Players:** Role players are important nodes in a social network. They perform unique functions such as connecting individuals belonging to different communities or being a central player in an important activity.
7. **Granule:** A granule (Zadeh, 1997) is a collection of similar, indistinguishable objects that can be treated as an independent unit. In a social network, a granule may be denoted by A_c indicating that it is represented by a center vertex c . A node's membership in the granule is given by its relationship with c which is usually a distance function. The set of all granules may be represented as:

$$Gr = \left\{ A_c \mid \forall c \in C, \sum_{v \in V} \widetilde{\mu}_c(v) / v \right\} \quad (3)$$

where, $\widetilde{\mu}_c(v) = \frac{\mu_c(v)}{\sum_{i \in C} \mu_i(v)}$ and $\mu_c(v)$ is the

membership function of node v in the granule represented by c .

8. **Embeddedness:** Embeddedness (Kundu & Pal, 2015a) is the extent to which a pair of granules overlap. If the granules are represented by centers a and b respectively, then the embeddedness is denoted by $\varepsilon(a, b)$ and is nothing but the cardinality of the intersection of the granules:

$$\varepsilon(a, b) = |A_a \cap A_b| = \sum_{v \in V} \min \left(\left(\widetilde{\mu}_a(v), \widetilde{\mu}_b(v) \right) \right) \quad (4)$$

The Girvan-Newman Algorithm

The Girvan-Newman algorithm (Girvan & Newman, 2002) is one of the oldest and most popular community detection algorithms that have been developed. It had, at its core, the concept of betweenness which was used to partition the graph, i.e. the social network, into several communities. The algorithm consists of three simple steps:

1. **Look for the Edge with the Highest Betweenness:** As defined in the previous section, betweenness measures the importance of an edge in terms of the highest volume of ‘traffic’ it conducts. An edge with high betweenness indicates that the edge serves as an important bridge between nodes of different communities. Therefore, removing this edge would most probably separate the communities that it connects or ‘bridges’. With this concept in mind, Girvan and Newman proposed to detect the edge with the highest betweenness and remove it. If multiple edges have the same highest betweenness, all of them are removed. This step results in a possible partition of the graph into a number of initial communities.
2. **Recalculating the Betweenness:** After the first step the betweenness of all the remaining edges in the partitioned graph are recalculated. This is due to the fact that the removal of edges in the previous step will

obviously alter the flow of traffic among the newly formed graph components.

3. **Repeat:** With the new partitioned graph and betweenness values, repeat steps 1 and 2. The repeated application of these steps will break the graph into smaller and smaller components. This process is carried out until no further edges remain in the graph.

The Girvan-Newman algorithm decomposes the social network into a hierarchical structure with each level of the structure containing a set of communities nested within the community set of the previous level. Therefore, with this algorithm, one can observe the structural breakdown of the social network at each point of an iteration of the algorithm. The Girvan-Newman method works well for moderately sized graphs containing up to a few thousand nodes (Newman, 2004) but is less effective for larger networks due to the large computational complexity involved in re-computing the betweenness values during each iteration.

NEWMAN’S FAST ALGORITHM

In 2003, Newman (2004) proposed a more time-efficient method for detecting communities. His algorithm made use of his original concept of modularity which is defined in Equation 1. Newman’s algorithm is an entirely greedy approach that is implemented to address the pitfalls of the Girvan-Newman algorithm. Apart from the computational complexity involved in the Girvan-Newman algorithm, another aspect that the authors had failed to address was the formation of *meaningful* communities. While it is true that the algorithm detects community structure effectively, it does not include any metric for measuring the actual quality of the detected communities. In order to address this issue, Newman proposed to use modularity both as a qualitative measure as well as to find the communities faster. Newman argued that the modularity could be optimized

over all possible community divisions in order to obtain the best fit. Furthermore, the optimization could be done using a greedy approach to reduce the time complexity involved.

Initially, each node in the graph belongs to its own community, i.e. if there are n nodes then the number of communities will also be n . Communities are then joined together in pairs and the pairs that produce the greatest increase or smallest decrease in modularity are finally merged together. This process continues until all the nodes are eventually merged into a single community. The result is a hierarchical structure of communities in which each level contains a larger set of communities than its parent level. An important point to remember is that, when joining communities, only those pairs that have edges between themselves are to be considered as communities that are not connected by edges will never increase the modularity. Again, the change in modularity can always be calculated in constant time making the algorithm much faster than the previous one. Newman acknowledged that the local optimization feature could possibly lead to a decrease in the quality of communities detected as opposed to the non-local nature of the Girvan-Newman algorithm but emphasized the fact that while the latter algorithm is computationally very expensive for large networks, the former would be able to provide a decent community structure in considerably less time.

THE LOUVAIN ALGORITHM

The Louvain community detection algorithm (Blondel, Guillaume, Lambiotte, & Lefebvre, 2008) is a heuristic algorithm that uses modularity (as defined in Equation 2) as a metric for gauging the effectiveness of the detected communities. The algorithm is primarily aimed at detecting communities in large social networks in a relatively short span of time. The Louvain algorithm consists of a greedy “iterative stage” and a “coarse graining” stage.

1. **Iterative Stage:** In this stage, the change in modularity is noted when a particular node i is moved out of its own community and into the community of one of its neighbors. The changes are observed for each neighbor of i . Node i is finally moved into that community for which the change is maximum as well as positive. In the event that the maximum modularity change is negative or 0, the node i will remain confined to its own community. The process is repeated for every other node in the network. After all the nodes have been considered, the entire stage is applied iteratively to the network until no further modularity change is possible. In this manner, the iterative stage seeks to find the local maxima of the modularity obtainable in the network.
2. **Coarse Graining:** In the coarse graining stage, a new network is constructed by taking each community detected in the iterative stage as a single node. For the sake of clarity, the nodes in the old network will be referred to as vertices. An edge between two nodes in this new network is the sum of weights of the interconnections between vertices belonging to the community represented by the first node and those belonging to the community represented by the second node. Of course, by this same logic, there will also exist self-loops whose weights are equal to the sum of the *intra*-connections between vertices belonging to the respective communities represented by the nodes in the new network. The newly constructed network is then passed through the iterative stage which will detect a set of communities pertaining to the new network information.

The iterative and coarse graining stages are repeated until no further modularity change is possible. One can easily see by now that the size of data being processed during the iterative stage drastically reduces with each coarse graining process. Thus, this algorithm is extremely

effective when applied to large network datasets consisting of more than a million nodes. As in the algorithm developed by Newman (2004), one can visualize the community hierarchy detected by the two stages.

OVERLAPPING NODE DETECTION TO IMPROVE COMMUNITY CLUSTERING

Berti, Sperduti, and Burattin (2014) believed that the detection of overlapping nodes that belong to multiple communities is the key to detecting important role players in the network. These nodes bridge communities together and may, therefore, be strong “communicators” that connect individuals belonging to diverse groups. To achieve this end the authors proposed a new measure known as *cuttability* which helps in determining whether a node is an overlapping one or not. If we consider an edge (i, j) , the cuttability of this edge can be defined as follows:

$$cut(i, j) = \min \left\{ \begin{array}{l} \sum_{(i,k) \in E, C(i)=C(k)} w_{i,k} - \sum_{(i,k) \in E, C(j)=C(k)} w_{i,k} \\ \sum_{(j,k) \in E, C(j)=C(k)} w_{j,k} - \sum_{(j,k) \in E, C(i)=C(k)} w_{j,k} \end{array} \right\} \quad (5)$$

From the formula above, it can be easily seen that if both nodes i and j of a given edge belong to the same community, then the cuttability of the edge will be 0. If the nodes belong to distinct communities, the cuttability will increase. As a result, the cuttability will indicate the extent to which a node lies between two communities and is, hence, likely to be an overlapping node. The term “cuttability” arises from the fact that overlapping nodes can be “cut away” or removed from the graph without affecting the overall consistency of the clustering. The cuttability of the entire graph is the sum of the cuttability of all its edges. The

algorithm proposed by the authors consists of the following four steps:

1. **Initial Clustering:** Here, the fast Louvain algorithm (Blondel, Guillaume, Lambiotte, & Lefebvre, 2008) is used to detect an initial set of community clusters in the network.
2. **Detecting the Overlapping Nodes:** Once the initial set of clusters is obtained, the overlapping nodes are detected. Each node is checked by considering its actual community and the set of its adjacent communities. The node is assigned to each of its adjacent communities in turn and the change in cuttability is measured; at any point, if the change is positive then the node is an overlapping node.
3. **Re-clustering on Removal of Overlapping Nodes:** The set of detected overlapping nodes is then removed from the graph and another set of clusters is obtained from the reduced graph. This step once again uses the Louvain algorithm to quickly detect the clusters and the final modularity is measured.
4. **Adding back the Nodes to Maximize Modularity:** Finally, the overlapping nodes are added back to the reduced graph in such a way as to maximize the modularity. Steps 2 – 4 can be iterated until the modularity cannot be increased any further.

Through this approach, it may so happen that nodes which were overlapping nodes in a previous step cease to overlap in a later step due to formation of a better community structure. On the other hand, subsequent iterations may give rise to new overlapping nodes and, even more interestingly, some nodes may be persistent and remain as overlapping nodes throughout the process. These nodes may be regarded as truly overlapping in that they actually do belong to multiple communities even though they are finally assigned to a single community.

FUZZY – ROUGH COMMUNITIES IN FUZZY GRANULAR SOCIAL NETWORKS (FRC – FGSN)

The FRC – FGSN (Kundu & Pal, 2015b) model is a new approach towards detecting communities in overlapping networks. It considers the possibility that nodes may belong to a number of communities with varying degrees of membership. Furthermore, some nodes may not belong to any community at all and will thus be considered as ‘orphan nodes’. The model may be described in two parts:

1. **Fuzzy Granular Social Networks:** The Fuzzy Granular Social Network (FGSN) model (Kundu & Pal, 2015a) represents a social network in terms of ‘granules’. The concept of granularity helps model large-scale social networks in terms of a much smaller number of granules. Each node in the network will have a membership value associated with each granule. Of course, the value is dictated by a membership function that results in a value ranging from 0 (not a member) to 1 (completely a member). Any number in the interval (0, 1) signifies partial membership in a granule. This is, thus, a ‘fuzzy’ membership (Zadeh, 1965) as opposed to the ‘crisp’ membership values of 0 and 1 that are normally encountered. Furthermore, the membership values are normalized so as to gain a better idea of the ‘degree’ of membership of a node to a particular granule.
2. **Fuzzy Rough Communities:** The FGSN model is then analyzed for the detection of communities. There are two concepts that come into play here: embeddedness (defined in Equation 4) and θ – cores. θ – cores are merely those granules whose granular degree is greater than a certain constant θ . This value is generally set equal to the average granular degree in the FGSN. In the Fuzzy Rough Community (FRC) detection

method, all θ – cores are initially computed. Next, all community-reachable θ – cores are grouped together, and separated from the θ – core pool, to form individual communities (community-reachable θ – cores are those θ – cores for which the granule representative of one θ – core lies in the neighborhood (or support set) of the other). Also, all θ – cores of a community must have a mutual normalized embeddedness greater than a certain pre-defined constant $\in$. The communities that have so far been detected are merely in terms of granules. What is required, however, are communities that consists of only vertices and it is here that the concept of Rough Sets (Pawlak, 1982) comes into play. Since the communities are not well-defined, they can be represented as rough sets, each with their own lower and upper approximations. The lower approximation of a community will contain all those nodes which definitely belong only to that community and no other community. The upper approximation contains the lower approximation as well as nodes that belong to both the community in question as well as to other communities:

$$L(C_i) = \{x \mid x \in \text{Support}(A_p) \text{ and } x \notin \text{Support}(A_q)\}; \\ \forall A_p \in C_i, A_q \in C_j \ i \neq j \quad (6)$$

$$U(C_i) = \{x \mid x \in \text{Support}(A_p); \forall A_p \in C_i\} \quad (7)$$

Thus, the membership function for a node in a community C_i is defined as:

$$\delta_{C_i}(x, r) = \begin{cases} 1 & \text{if } x \in C_i \\ \sum_{c \in C_i} \widetilde{\mu}_c(x, r) & \text{if } x \in L(C_i) \setminus U(C_i) \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

GRANLOUV: A PROPOSED COMMUNITY DETECTION ALGORITHM

In a previous work (Dillen & Chakraborty, 2015), the authors had proposed a novel algorithm for community detection. This algorithm, coined as *GranLouv*, was inspired by two of the algorithms mentioned above, namely, the Louvain method of community detection and the Fuzzy Granular Social Network model. There are three basic features associated with this algorithm:

1. **Formation of a Granular Adjacency Matrix:** The network is initially modeled in the FGSN domain by considering a set of granule centers and assigning the node memberships to these granules. The centers considered are those nodes in the original network whose degree is greater than the average degree of the network. The granule radius is taken to be equal to the network diameter. Once the FGSN is constructed, a new granular adjacency matrix is formed. An edge between two granules in this network bears a weight equal to the embeddedness (Equation 4) between the said granules. This would, thus, imply the existence of self-loops in the new network as well. In this case, the self-loop's weight will be the same as the cardinality of the granule with which it is associated.
2. **Applying the Louvain Algorithm on the Granular Matrix:** The granular adjacency matrix thus formed is then fed as input to the Louvain algorithm. The algorithm will detect communities in this 'network' which will be referred to as granular communities. These communities serve as a basis for the final step.
3. **Final Community Detection:** In the final detection stage, fuzzy rough communities are first detected in the network using the properties of lower and upper approximation in FGSN networks described in Equation 8. After this, all vertices are assigned to their

final communities: each vertex will possess a membership of 1 (i.e. complete membership) in the community for which its fuzzy membership is maximum. It will possess a membership of 0 in all the other remaining communities.

Application of *GranLouv*

The *GranLouv* algorithm was tested on three real-world social networks: the Dolphin Social Network (Lusseau, Schneider, Boisseau, Haase, Slooten, & Dawson, 2003), the Les Misérables Social Network (Knuth, 1994) and the American College Football Social Network (Girvan & Newman, 2002). The algorithm achieved a modularity of 0.509, 0.534 and 0.599 respectively in each of the three cases which is quite comparable to other algorithms employing the use of modularity. The results are depicted graphically in Figure 2, Figure 3, and Figure 4 as well as in Table 1, Table 2 and Table 3 respectively.

Table 1. Modularity detected by algorithms on the Dolphin Social Network

Serial No.	Algorithm	Modularity
1.	Louvain	0.518
2.	GranLouv	0.509

Table 2. Modularity detected by algorithms on the Les Misérables Social Network

Serial No.	Algorithm	Modularity
1.	Newman and Girvan (2004)	0.540
2.	Louvain	0.555
3.	GranLouv	0.534

Table 3. Modularity detected by algorithms on the American College Football Social Network

Serial No.	Algorithm	Modularity
1.	Girvan and Newman (2002)	0.601
2.	Louvain	0.604
3.	GranLouv	0.599

Figure 2. Communities detected in the Dolphin Social Network using the GranLouv

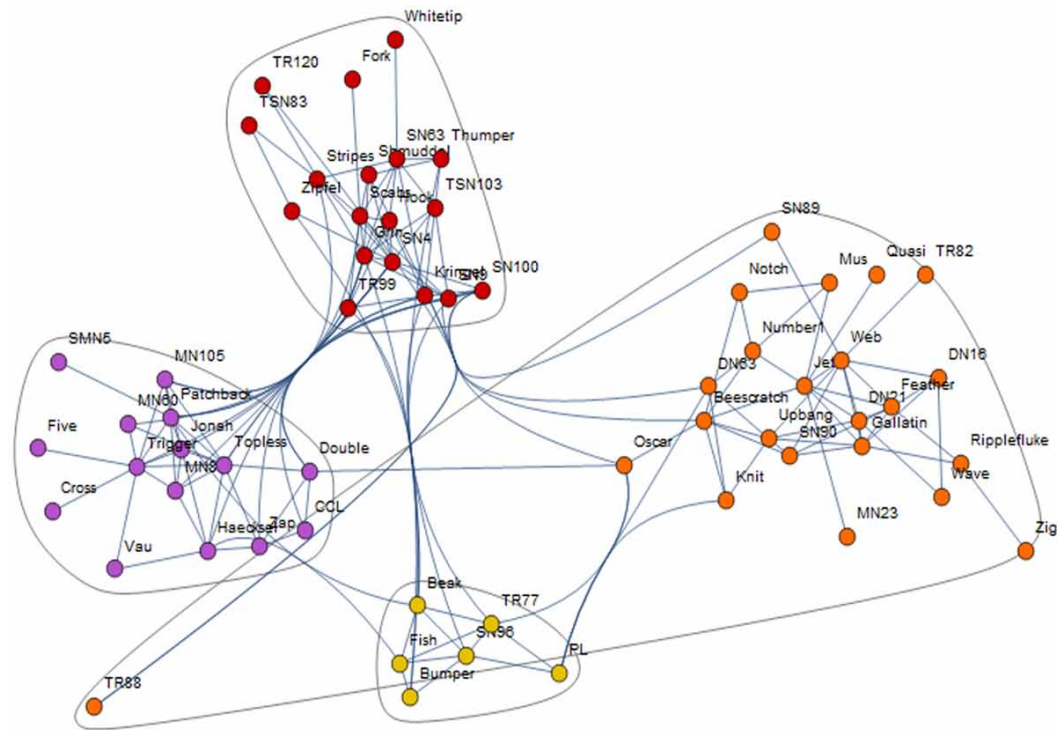


Figure 3. Communities detected in the Les Miserables Social Network using the GranLouv algorithm

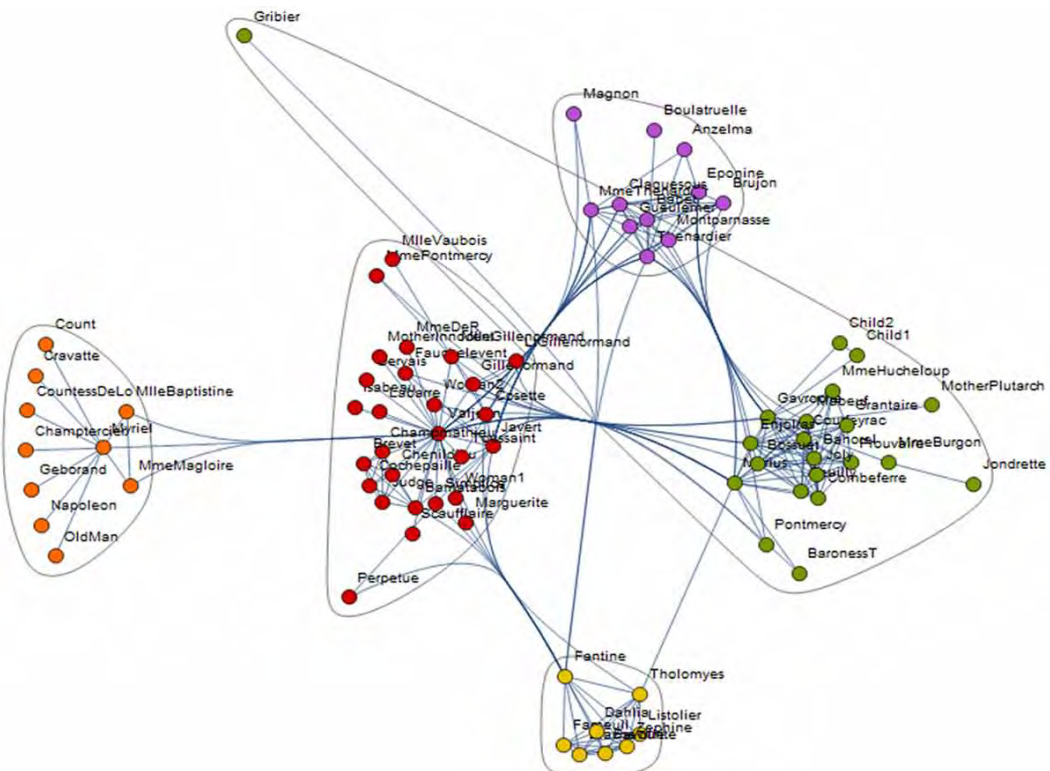
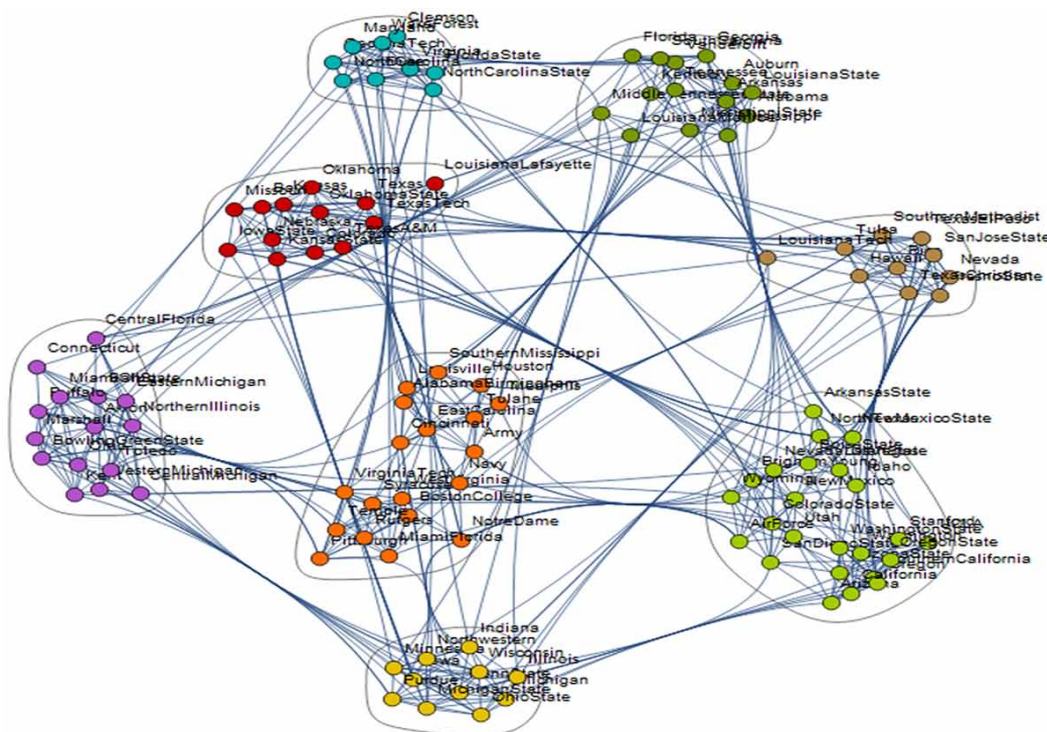


Figure 4. Communities detected in the American College Football Social Network using the GranLouv algorithm



FUTURE RESEARCH DIRECTIONS

Although the granule centers in GranLouv are chosen on the basis of vertex degree, this practice may not necessarily be the best way to do so as it takes only one factor into consideration, i.e. the degree, and so fails to properly and holistically select the nodes which have a higher tendency to form clusters or granules around themselves. This problem must be addressed effectively as granule center selection poses a level of difficulty when the networks consist of a very large number of nodes. It is definitely not feasible to select all nodes in the network as centers as this would go against the actual idea behind using a granular model in the first place, which is to reduce the amount of redundant information in the network and ease the computation process.

Another aspect that must be taken into consideration is the absence of overlapping communities using GranLouv. This problem may be addressed

by changing the modularity function to account for overlapping communities as well instead of only crisp community sets.

Finally, attention must be drawn to the fact that granules are assumed to be circular in nature with a predefined radius set by the programmer. However, a granule should be allowed to assume any arbitrary shape in order to realistically capture information in the network. Therefore, there is much scope for research work which can properly and efficiently consider granules with any shape as opposed to the strict and inflexible one that has been proposed.

CONCLUSION

Social network analysis, and especially community detection, has a multitude of benefits. Through community detection, business organizations can use social network data to determine appropriate

target audiences for their products and selectively advertise to the target audience, thus increasing the probability of attracting potential buyers. Law enforcement agencies may also use social network analysis to identify potential or existing criminal organizations based on the connections of known criminals. Furthermore, “important” players in social networks may also be identified with a higher degree of accuracy through detection factors such as popularity (number of connections), behavior, and other pertinent features. Thus social network analysis plays an important role in economics, business intelligence and law enforcement, as well as other areas ranging from politics to information science. Research in this field could, therefore, significantly address a number of problems in a world that is becoming increasingly globalized and interconnected at a rapid pace.

ACKNOWLEDGMENT

The authors wish to convey their thanks to Prof. Sankar Kumar Pal as well as Mr. S. Kundu from the Center for Soft Computing Research, Indian Statistical Institute, Kolkata, for their continued support and guidance provided during the conception of this paper.

REFERENCES

- Berti, A., Sperduti, A., & Burattin, A. (2014). A novel criterion for overlapping communities detection and clustering improvement. In *IEEE Symposium on Computational Intelligence and Data Mining (CIDM 2014)*. Orlando, FL: IEEE. doi:10.1109/CIDM.2014.7008675
- Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics*, 2008(10), P10008. doi:10.1088/1742-5468/2008/10/P10008
- Clauset, A., Newman, M., & Moore, C. (2004). Finding community structure in very large networks. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, 70(6), 1–6. doi:10.1103/PhysRevE.70.066111 PMID:15697438
- Dillen, N. B., & Chakraborty, A. (2015, October). *Modularity-Based Community Detection in Fuzzy Granular Social Networks*. Paper presented at the International Congress on Information and Communication Technology. Udaipur, India.
- Ganganathy, N., Chenz, G., & Cheng, C. T. (2014). Detecting Hierarchical and Overlapping Community Structures in Networks. *2014 International Symposium on Nonlinear Theory and its Applications, NOLTA2014*.
- Girvan, M., & Newman, M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences USA*, 99(12), 7821–7826. doi:10.1073/pnas.122653799
- Kernighan, B. W., & Lin, S. (1970). An Efficient Heuristic Procedure for Partitioning Graphs. *The Bell System Technical Journal*, 49(2), 291–307. doi:10.1002/j.1538-7305.1970.tb01770.x
- Knuth, D. E. (1994). *The Stanford GraphBase: A platform for combinatorial computing*. New York, NY: Addison-Wesley Publishing Company.
- Kundu, S., & Pal, S. K. (2015). (2015b, in press). Fuzzy-rough community in social networks. *Pattern Recognition Letters*, 67, 145–152. doi:10.1016/j.patrec.2015.02.005
- Kundu, S., & Pal, S. K. (2015a). FGSN: Fuzzy Granular Social Networks – Model and applications. *Information Sciences*, 314, 100–117. doi:10.1016/j.ins.2015.03.065
- Lusseau, D., Schneider, K., Boisseau, O. J., Haase, P., Slooten, E., & Dawson, S. M. (2003). The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations. *Behavioral Ecology and Sociobiology*, 54(4), 396–405. doi:10.1007/s00265-003-0651-y

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley, CA: University of California Press.

Newman, M. E. J. (2004). Fast algorithm for detecting community structure in networks. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, 69(6), 066133. doi:10.1103/PhysRevE.69.066133 PMID:15244693

Newman, M. E. J., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, 69(2), 026113. doi:10.1103/PhysRevE.69.026113 PMID:14995526

Pawlak, Z. (1982). Rough Sets. *International Journal of Computer & Information Sciences*, 11(5), 341–356. doi:10.1007/BF01001956

Pons, P., & Latapy, M. (2006). Computing communities in large networks using random walks. *Journal of Graph Algorithms and Applications*, 10(2), 191–218. doi:10.7155/jgaa.00124

Wakita, K., & Tsurumi, T. (2007). Finding community structure in a mega-scale social networking service. In *Proceedings of IADIS international conference on WWW/Internet 2007*, 153–162.

Zadeh, L. (1965). Fuzzy sets. *Information and Control*, 8(3), 338–353. doi:10.1016/S0019-9958(65)90241-X

Zadeh, L. A. (1997). Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. *Fuzzy Sets and Systems*, 90(2), 111–127. doi:10.1016/S0165-0114(97)00077-8

KEY TERMS AND DEFINITIONS

Community: A community is a subset of nodes in a social network that have dense connections between them and which are sparsely connected to nodes belonging to other communities.

Edge Betweenness: The betweenness of an edge is the number of shortest paths between vertices that contain the edge.

Embeddedness: Embeddedness is the extent to which a pair of granules overlap.

Granule: A granule is a collection of similar, indistinguishable objects that can be treated as an independent unit.

Graph: A graph is a network of nodes, or vertices, and their interconnections, or edges.

Modularity: Modularity measures the quality of community partitions formed by an algorithm. It is the difference between the actual density of intra-community edges and the corresponding connections in a random network possessing the same degree distribution as that of the actual network.

Role Players: Role players are important nodes performing unique functions in a social network.

Social Network: A social network is a set of individuals that are interconnected through some relationship.

Classification of Traffic Events Notified in Social Networks' Texts

S

Ana Maria Magdalena Saldana-Perez

Instituto Politecnico Nacional, Mexico

Marco Antonio Moreno-Ibarra

Instituto Politécnico Nacional, Mexico

Miguel Jesus Torres-Ruiz

Instituto Politécnico Nacional, Mexico

INTRODUCTION

Many researching works implement procedures to use web data and users generated content on critical situations management, or to notify relevant events (Wakefield, 2013). Every day millions of publications, where people describe their environment, and ideas about many interesting topics are done at social networks. From such publications, it is possible to acquire data related to the urban factors that concern most citizens, in order to improve the people's activities and conditions, and with a view to structure smart cities. Web publications that have coordinates or provide geographic information, can be considered as VGI.

The main purpose of this work, is to obtain useful information for a further traffic prediction procedure, by integrating and processing VGI and authoritative data. The VGI data is collected from a social network, where participants are considered as intelligent sensors, which constantly communicate the environment changes they perceive; such reports are influenced by the people context.

Two of the main advantages of VGI, are its continuously updated data, and the easily data consultation. Such kind of source has not been deeply studied, most of the time its social characteristics have been analyzed over its technological usages.

The present approach makes use of VGI and authoritative data sources, to extract, analyze, classify and geocode data related to the traffic in the study area. The case of study is Mexico City, the largest metropolitan area in America, and the largest Spanish-speaking city in the world (Zimmerman, 2015).

The social network consulted is Twitter, also a RSS service is considered. In the analysis stage, a text mining and a recovery information processes are applied over the short texts, extracted from the social network and from the rss service, with a view to identify the traffic events described on the posts; on the classification stage, a Naïve Bayes algorithm is trained to classify the traffic events. The geocoding process is done through the use of the methodology proposed by Salazar et al. (2016). The events and the authoritative data, are related with the purpose to establish links, between the streets and the business located on the area where the traffic is presented; also a probabilistic analysis is applied. Finally, the events and the authoritative data are mapped to visualize the traffic distribution.

The paper is organized as follows: first the background of the work is shown, below the data treatment is described; then, solutions and recommendations are done; after the future research directions and conclusions are presented; finally, the additional readings are mentioned.

BACKGROUND

User generated content (UGC) is any publication on internet, done by users of web services such as blogs, wikis, forums, social networks, podcast and chats. It is used in many applications, including: researches, information and news spread, problems processing, disaster management, and collaborative mapping. UGC has originated some other concepts, volunteered geographic information (VGI) and crowdsourcing are two of them (Chard, 2015).

Crowdsourcing is the process of getting ideas, information, or work done, from a group of interested people; it has been a recurrent data and services source for some businesses and researches (Chard, 2015). Mobile devices as smartphones, mobile GPS, cartographic applications and social networks, make crowdsourcing possible.

According to Wen Lin (2013), VGI is composed of volunteered information generated by users, who have not a geographic specialized knowledge, but are interested on provide data with geographic characteristics; such data are employed on many web services as Open Street Map (OSM), WikiMapia, Google Maps, among others. This association between VGI and web services are the GeoWeb basis.

VGI makes possible the generation of new spatial information usages, such as collaborative mapping, georeferenced content, network vectorization, and collective ranking of places (Gouveia & Fonseca, 2008; Elwood, 2013; Sui, 2012; Goodchild, 2007). There exist web services that invite people to provide personal information, their location, or some other geospatial information, such as landmarks, points of interest (POI), and street segments; with the purpose to increment their data repositories, and make them available for further analysis. OSM, Foursquare, and Twitter are some of those web services (Kunze, 2015).

Kunze (2015), proposes a VGI integration process, to generate estimation models of dwelling occupancy. In his approach, OSM and topographic data from Dresden city are used. When

working with data from many spatial data sources, it is needed to correctly integrate such data on a GIS environment, and to consider some specific features, as the data precision, and the reference system applied over them (Flowerdew, 1991).

Most of the internet users, are part of the social networks; according with previous investigations 70% of them, access to internet by using mobile devices; and 73.4% do it through a computer (Castells, 2010).

Nowadays, internet users are interested on describe their environment, or give information about the events they perceive; on their publications people describe physical aspects of their surroundings, as the traffic or pollution situation, the security degree they consider appropriate for their neighborhood, socio-politic events, and some other topics of interest; also, they provide metadata such as coordinates, URL's, among others (Fleming, 2014).

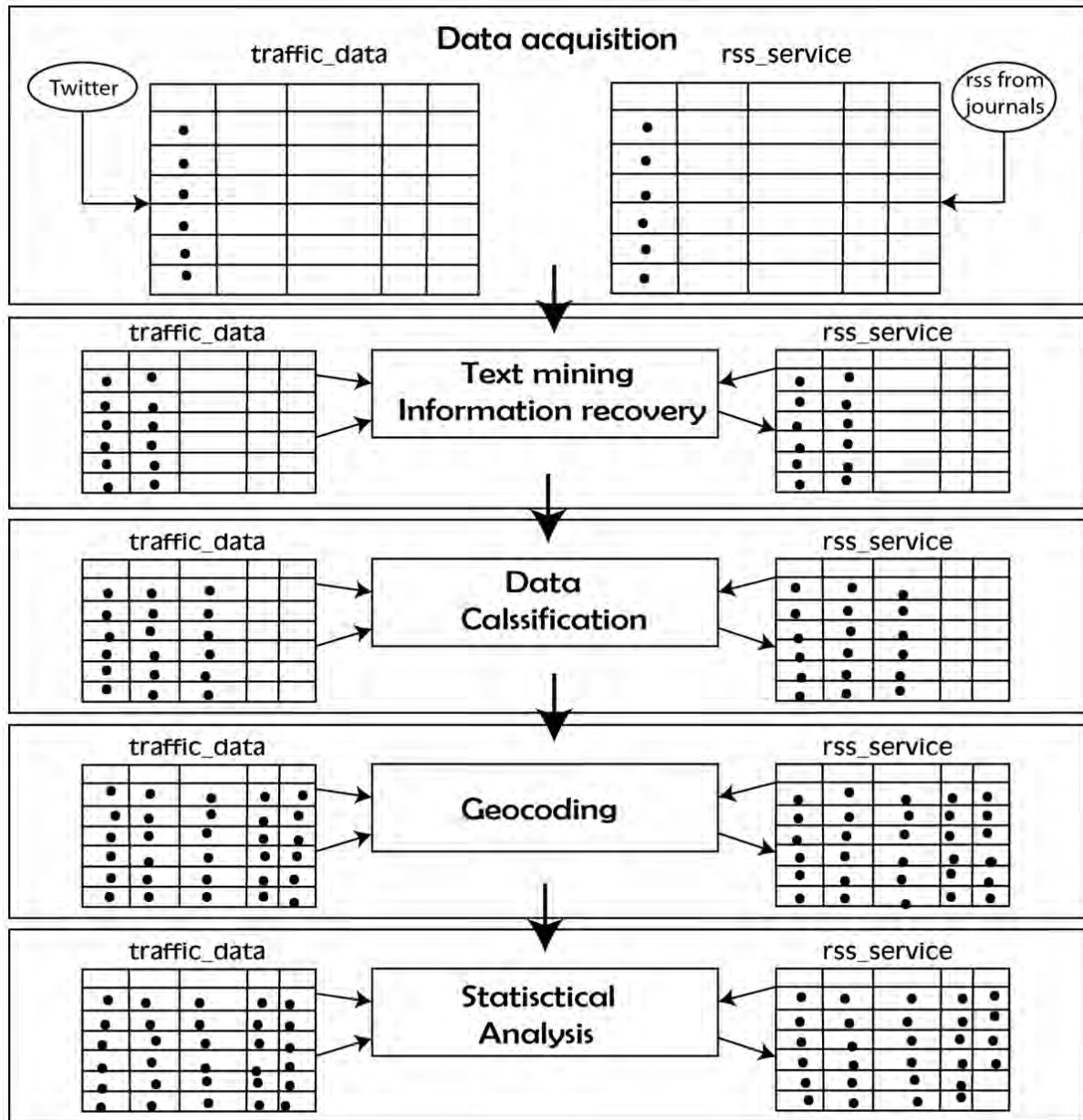
User generated content published on social networks and VGI sources, has an explicit semantic value, since the content has been produced by people interested on provide their data for further analysis, unlike the other web content generated by people without a specific purpose (Resch, 2015).

Considering people as sensors, gives a wide variety of heterogeneous data, to researches that study the human behavior or the city dynamics, through the analysis of VGI and UGC data. The people's point of view is different from one person to another, and it is influenced by the person's living conditions, activities, education, ideas, and the geographic area where he or she lives (Resch, 2015).

DATA TREATMENT TO DETECT TRAFFIC EVENTS ON TEXTS

The present approach has been designed to obtain information about the traffic situation in the study area, by analyzing and processing VGI data. The proposed methodology, consist on 5 stages: the first, is the data acquisition; the second, are

Figure 1. Data acquisition from a social network to describe the urban traffic situation methodology stages



a text mining and a recovery process, to obtain data related to traffic events; the third, is the data classification; the fourth, is the data geocoding; and the fifth, is a statistical analysis. On Figure 1, a diagram shows the methodology stages.

The social network consulted for this research is Twitter; it is one of the most used micro blog services, at present has more than 500 million

users. On its web site, people post their ideas, on short texts called tweets, which have a limit of 180 characters (Bernstein, 2013). The tweets considered as relevant for this approach, are those which describe the traffic situation or traffic events occurred on the study area.

This work has been oriented to detect traffic information published by people; an advantage

of the approach is its flexibility, since it could be used to extract information related to some other urban topics, in order to improve human life.

Data Acquisition

To extract the short texts from Twitter, an API of basic transmission is used on a code script to randomly extract, the 1% of tweets published on the study area. The tweets are continuously extracted, since the API can not to have access to tweets published one week before the consultation date.

The short texts, are obtained from specific Twitter accounts, which main purpose is to notify the traffic on the study area. The Twitter profiles monitored are @*trafico889*, @*dftrafico* and @*MexicoTrafico*.

The collected tweets and their metadata, are stored on a table named *traffic_data*; the features considered are the tweet's publication date and time, its text, and its coordinates (in case they have been published). Some Twitter service studies, have shown that only 3% of the tweets have the coordinates of the the place where they were posted, and only 17% of the short texts have relevant information about the place of publication (Burton, 2012).

To extract the information from the rss service, a PHP script was developed to collect the news related to traffic on the study area, that have been published by some of the most important local journals: *Reforma*, *El Universal* and *La Jornada*. The text, date and time of publication, were extracted from the news collected, and then stored on a table named *rss_service*. Both *rss_service* and *traffic_data* tables belong to the *traffic_situation* database.

Information Recovery

Once the tweets and rss texts have been collected, is needed to extract from the texts, the information related to the traffic situation and to the traffic events. The texts stored on the *rss_service* and *traffic_data* tables, are written in Spanish

language. In order to manipulate such texts it is necessary to apply a text mining process and a recovery information procedure.

The stop words, are all the common words on a language, such as prepositions, connectors, and articles, that do not provide relevant information to the texts (Sidorov, 2014).

On the text mining process, it is necessary to remove the stop words from the texts; also, the special characters used on the Spanish language (':,;.\$%`!¿?/*...') are deleted. After this process, texts are ready to be treated by more complex computing procedures.

The information recovery process, consist on identify the words related to the traffic on texts; this is done by a programmed script, that identifies on each tweet or rss text, the words related to the traffic and the number of times that they appear.

The identified words are grouped, according to the event to which they refer, into five classes: *favorable* and *traffic*, define the situation of the vehicles movement through the streets; *accident* and *closure*, describe traffic events; *rule*, when the text has information about road rules; and *unidentified*, when none of the text's words are related to traffic events.

Data Classification

On this section, the rss texts, and the tweets text are indistinctly called *texts*, in order to simplify the explanation. The classification process described below, is independently applied over the *rss_service* and *traffic_data* table's registries.

To classify the texts, the Naïve Bayes algorithm has been chosen; it is a machine learning algorithm based on the Bayes conditional probability theorem, which main purpose is, to calculate the probability for a document, to belong to a determined class, in case that it contains a specific word (Scharkow, 2013). Naive Bayes, assumes that a feature's presence or absence, is not related to any other feature's presence or absence. The algorithm, has been used as a standard for text classification and

spam filters; its main advantage is, that it is good working with big amounts of data, and needs few data to determine class parameters (Lantz, 2013).

Naïve Bayes requires a training process, where the algorithm is applied over a training dataset, to create probabilistic tables that let it know the differences between the proposed classes, and the kind of element that is being treating.

To create the training corpus, the third part of the texts was classified, by using an algorithm that searches on each registries' text, the words related to each proposed class, and assigns to the registry a class label according to the major number of class' words founded. The class label is stored on the field *event_class* of the *rss_service* and *traffic_data* tables. The third part of the texts with a defined label class are named *training corpus*, meanwhile the two third parts without label are called *test corpus*.

The Naïve Bayes algorithm develops the following steps; first, from the training corpus extracts the *event_class* label of each registry and factorizes it, it means that represents such class label with a number; and then, a data dictionary is created to save all the words that compose the texts. Each word, acts as a key on the dictionary, and its value is the number of times the word was founded on the entirely training corpus, this is done by comparing each registry's text of the *rss_service* and *traffic_data* tables, with the words in the dictionary; if there is a word on a text that does not appear on the data dictionary, it is added as a new key.

Since texts have different lengths, it is needed to homogenize their longitudes, to make them useful for the machine learning's training. For that purpose, a vector is created for each text, each vector's position represents a dictionary's word, catalogued in the same way as in the dictionary; the value of each position, is the number of times the word appears on the treated text. In case that a word does not belongs to the text, its value is zero. Every text (from the tweets or the rss messages), has a representing vector.

Once the vectors have been structured and the classes have been defined, the Naïve Bayes, algorithm calculates the probability for each dictionary's word to belong to any of the defined classes, by counting the times that an specific word appeared on each class, and dividing that number, between the total words on the class $P(B)$. Then, it calculates the probability that the training registries have been correctly classified, based on the words present in them, and the calculated words' probabilities. The process is related to the Bayes theorem shown on Equation 1, where B is the word founded on the registry, A_i is the dependent event (the label of the class we are analyzing), $P(B)$ is the marginal probability of the word, $P(A_i)$ is the prior probability of the class A_i , $P(B|A_i)$ is the likelihood of word B in the class A_i , and $P(A_i|B)$ is the posterior probability of classify the registry on the A_i class, given that the founded word was B .

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)} \quad (1)$$

After the training process, is time to Naïve Bayes learning. When a new registry is being analyzed, the algorithm calculates its posterior probability, in order to determine the appropriate class for it, taking into account the values stored on the text's vector. The *posterior probability* for a vector to belong to each of the classes, is calculated considering the presence probability computed on the training stage for each vector's word (*marginal probability*), and the words absence probability. Then, the *posterior probabilities* are added, the result is called *class probabilities*, as shown in Equation 2; the probability for the vector to belong to a certain class, is calculated by dividing the *conditional probability* between the *class probabilities* calculated previously, as shown in Equation 3. The number of classes is 5, as defined on the information recovery process.

$$class\ probabilities = \sum_{j=1}^5 P \left(\begin{matrix} Class\ j\ |\ -\ probability\ of\ words \\ presence\ or\ absense - \end{matrix} \right) \quad (2)$$

$$P(vector \in Class\ i) = \frac{P \left(\begin{matrix} Class\ i\ |\ -\ probability\ of\ words \\ presence\ or\ absense - \end{matrix} \right)}{class\ probabilities}, \forall i \in [1,5] \quad (3)$$

The class chosen for the vector is the one with the highest numeric value. After the learning process, Naïve Bayes is ready to classify the test data. The classification process follows the next sequence: first, the texts of the test corpus are tokenized to search for new words to be added on the data dictionary, and to identify the number of each word repetitions in the corpus; then the words frequencies are calculated. After that, a vector with the same length as the dictionary is created for each text; the conditional and class probabilities for each vector are calculated; and finally, the highest probability for each vector to be part of a class is selected, in order to assign it a correct class label. The class labels are stored on the *event_class* field, for any of the registries of the tables analyzed.

Data Geocoding

For the data geocoding, a gazetteer based on the methodology proposed by Salazar et al. (2016) has been used; the gazetteer is derived from *GeoNames*, and has 36,236 elements that represent Mexico City's streets with names written in Spanish, and abbreviations that describe the hierarchy of the roads. The gazetteer works with spatial dictionaries that represent the location of transport stations with points, the streets with

lines, and the location of monuments, buildings and some other landmarks with polygons.

On the geocoding methodology the text is analyzed, by identifying words which make reference to monuments, buildings, streets and avenues through the use of n-grams, bi-grams and tri-grams on it, by applying a recovery information process (Sidorov, 2014); once that the elements have been identified, a geospatial representation (point, line or polygon) defined on the spatial dictionaries, is assigned to them. By calculating the possible combinations of the geometric elements, the spatial relations between the detected elements are calculated ((line, line), (line, polygon), (line, point), (point, line), (point, point)), when the resultant relation is possible, the coordinates of the elements intersection, are chosen as the coordinates for the event described.

What the proposed approach does on this stage, is to apply the geocoding methodology of Salazar (2016) over the registries texts stored on the *rss_service* and *traffic_data* tables, in order to obtain its coordinates. The resultant coordinates are stocked on the fields *x_coordinate* and *y_coordinate* of the mentioned tables.

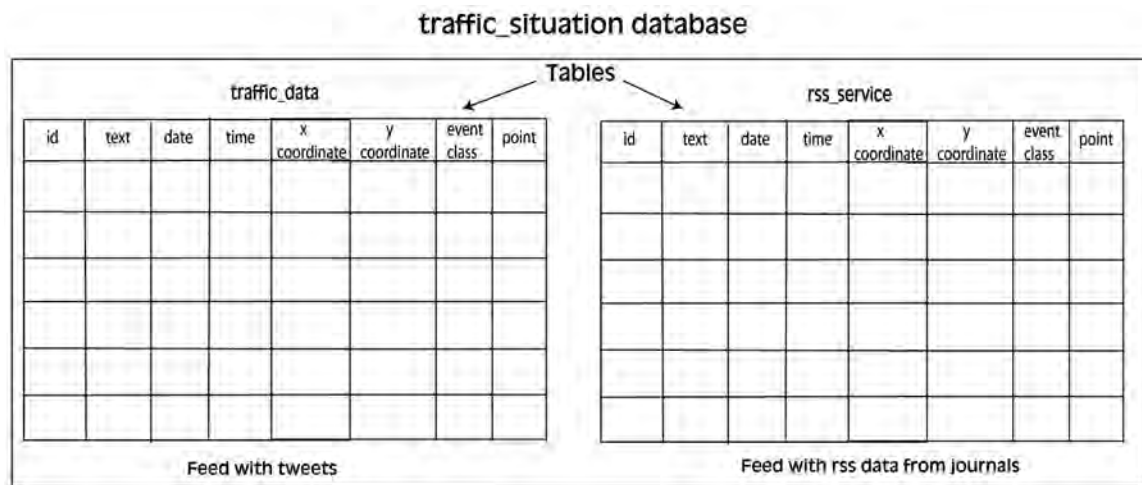
The *traffic_situation* database, has been structured on *Postgresql*, as shown on Figure 2. By using the spatial extension *Postgis*, the texts coordinates, are transformed into geometric points, to represent the events they describe, in the study area's map; such transformation is possible by applying the Postgis function *ST_GeomFromText*.

Statistical Analysis and Results

To make a better analysis of the classification and geocoding results, a statistical analysis was applied over the 5000 tweets and rss texts treated.

It was found that 32% of the processed texts, have information about accidents occurred on the roads; 62.98% describe the traffic situation, and from them, 14.24% describe a good circulation situation, meanwhile the 48.74% show problems for vehicles to move through the streets; 4.0% refer to streets closures; and only 1.02% of the texts, had

Figure 2. Tables from traffic_situation database



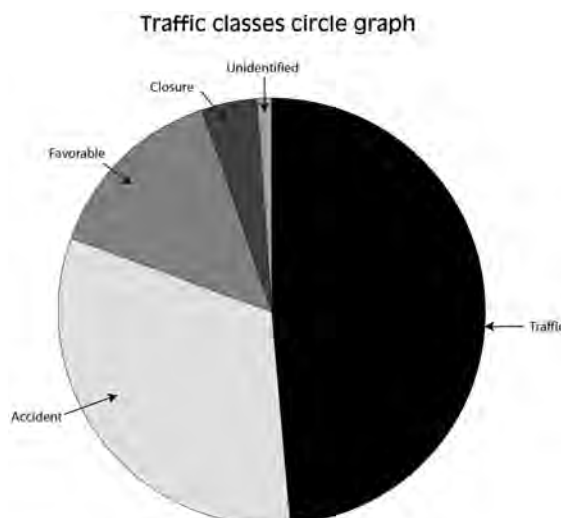
not enough information to be classified. On Figure 3, a circle graph shows the events proportion.

To have a better event's visualization, the points of each class were mapped on the study area. Here the authoritative data is used, to relate the classified and georeferenced traffic events, with the location of commercial businesses of the study area. The authoritative data is a shape file from the geography and statistics department in Mexico, each point in the shapefile, represents a commercial entity in the city, and provides infor-

mation such as the business schedule, commercial activity, and number of workers.

The used cartography belongs to Open Street Map; which lets to this approach make a relation between the VGI data from a mapping platform, and the VGI data from a social network. On Figure 4, a map with the OSM cartography shows the authoritative data (points), and buffers of 100 meters round the businesses, to detect the traffic events reported by citizens near to them. The information from the businesses makes possible to relate its commercial activity, and the number of traffic events detected.

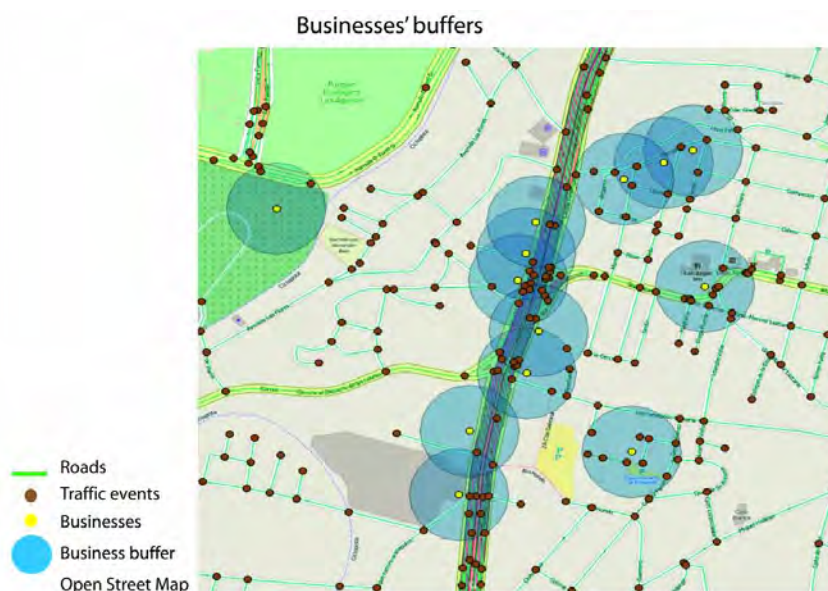
Figure 3. Circle graph that shows the classified events proportions



SOLUTIONS AND RECOMMENDATIONS

The first problem this approach presented, was the texts geocoding. Since a few part of the short texts from the social network have coordinates, and most of the rss texts recovered have not the coordinates of the place where they were published, or of the event they describe, it was necessary to implement a geocoding process. One solution proposed, was using a basic gazetteer to search for the streets names on texts; the issue with this solution is, that it does not exists a complete Mexico's gazetteer, and that there are streets names in Mexico City,

Figure 4. Buffers created round authoritative data, to evaluate the number of traffic events, and its relation to the commercial activity of the represented place



related with names of people and places from other countries, such as Liverpool, Zurich, Benjamin Franklin, among others. When using this proposal, many short texts were geocoded at some other latitudes far from Mexico, as represented on Figure 5, where the street named *Berlin* in Mexico City was geocoded in Germany.

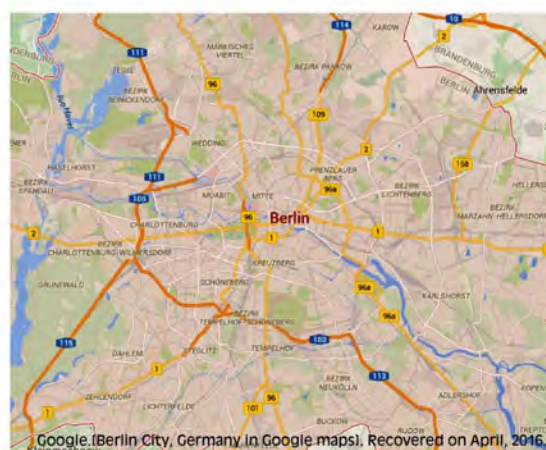
Looking for another solution, it was found that many researches use annotators to geocode tweets manually, this is one of the most accurate ways to geocode, but it is not easily scalable and requires more time and participants. The methodology proposed by Salazar (2016), was the most efficient for this investigation, since by an automatic

Figure 5. Coincidence between Mexico City streets names, and places out of the study area with the same name

Names coincidences that affect the correct data geocoding



Berlin street in Mexico City (study area)



Berlin City in Germany (not the study area)

process, the coordinates for the text were obtained with a good precision level.

Another issue, was the classification model election. Our approach was designed to use a machine learning algorithm, and by studying many tools, it was decided to apply the Naïve Bayes classification algorithm, due to the good results that it has reached by classifying texts such as e-mails and SMS. Naïve Bayes requires a training data set, to learn patterns and to be useful for further data classifications; such data set was classified manually. A recommendation for this approach is to use some other semi-classified machine learning algorithms, to classify the test corpus created by the present research, and make comparisons between the new results and the Naïve Bayes ones.

FUTURE RESEARCH DIRECTIONS

This methodology has been proved to extract information about the traffic on urban areas, with the study case in Mexico City, but it can be used to extract information about some other urban topics, as well as air pollution, floods, robbery, among others; the changes needed on the procedures, to treat other topics, must be done on the relevant words to be considered on the text mining and recovery information processes, and in the classification of the data from the training corpus. On the machine learning stage, the update is not necessary since the algorithm learns the interesting words from the texts, by calculating their probabilities of presence and absence.

The analysis of traffic has been done on this approach, with a view to provide classified traffic events descriptions with their coordinates to a further research, that studies the relation between traffic events in a city and models the behavior of traffic in urban areas by regression models.

CONCLUSION

The present approach, makes use of VGI and authoritative data, to represent and classify the traffic events on an urban area. The authoritative data are geometric points that represent the commercial businesses on the study area; the VGI information is acquired from Twitter messages and from rss services. The social network Twitter has been chosen because of its large number of members and the facility to extract information from it, by using a Twitter API.

The study area is Mexico City, one of the most populated and with more demographic density in America. Since the collected tweets belong to Mexican Twitter accounts, the short texts are written in Spanish language. After the extraction from tweets and rss, texts are stored on a database named *traffic_situation*, into two tables; then a text mining procedure, and an information recovery process are applied over the data to eliminate special Spanish characters, and to extract the words that describe events or situations related to the traffic on the study area.

The data set of tweets and rss texts, are partitioned into three. The identified traffic words in Spanish, such as *lento*, *accidente*, *cierre*, *avance* among others, are used to classify one data set partition, into five classes: *traffic*, *favorable*, *closure*, *accident* and *unidentified*; by using a words comparison method. Such classified partition is called training corpus.

To classify the other two partitions which compound the test corpus, Naïve Bayes a semi-supervised machine learning algorithm is used. The algorithm is trained with the classified texts, and calculates probability tables, in order to learn how to classify new texts based on the words that they contain or not; after the training process, the algorithm classifies the test corpus into the same five classes defined on the information recovery stage.

Despite some of the short texts from such microblogging service have the coordinates of the place where they were published, it is necessary to apply a geocoding process over the data, in order to know the coordinates of the place where the classified traffic event is being described.

When the classification and coordinates of the texts are known, geographic points are created to represent the events on the area's map, with a view to identify the places with more traffic, and its relation with the commercial businesses; a statistical analysis is also done, to know the proportion of each defined class.

The present approach collects, classifies and geocodes traffic events information, nevertheless, it could be useful in the acquisition and data treatment of many topics, or to analyze traffic on some other study areas.

REFERENCES

- Bernstein, M. S., Bakshy, E., Burke, M., & Karer, B. (2013, April). Quantifying the invisible audience in social networks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 21-30). ACM. doi:10.1145/2470654.2470658
- Burton, S. H., Tanner, K. W., Giraud-Carrier, C. G., West, J. H., & Barnes, M. D. (2012). Right time, right place health communication on Twitter: Value and accuracy of location information. *Journal of Medical Internet Research*, 14(6), e156. doi:10.2196/jmir.2121 PMID:23154246
- Castells, M. (2010). The culture of real virtuality: The integration of electronic communication, the end of the mass audience, and the rise of interactive networks. In *The Rise of the Network Society* (2nd ed.). Academic Press.
- Chard, K., & Caton, S. (2015). Social Clouds: Crowdsourcing Cloud Infrastructure. In *Crowdsourcing (pp. 191-217)*. Springer Berlin Heidelberg.
- Elwood, S., Goodchild, M. F., & Sui, D. (2013). *Prospects for VGI research and the emerging fourth paradigm*. Springer Netherlands. doi:10.1007/978-94-007-4587-2_20
- Fleming, L. E., Haines, A., Golding, B., Kessel, A., Cichowska, A., Sabel, C. E., & Cocksedge, N. (2014). Data mashups: Potential contribution to decision support on climate change and health. *International Journal of Environmental Research and Public Health*, 11(2), 1725–1746. doi:10.3390/ijerph110201725 PMID:24499879
- Flowerdew, R. (1991). Spatial data integration. In D. J. Maguire, M. F. Goodchild, & D. W. Rhind (Eds.), *Geographical information systems: Principles and applications* (Vol. 1, pp. 375–387). Harlow: Principles. Longman.
- Goodchild, M. F. (2007). Citizens as sensors: The world of volunteered geography. *GeoJournal*, 69(4), 211–221. doi:10.1007/s10708-007-9111-y
- Gouveia, C., & Fonseca, A. (2008). New approaches to environmental monitoring: The use of ICT to explore volunteered geographic information. *GeoJournal*, 72(3-4), 185–197. doi:10.1007/s10708-008-9183-3
- Kunze, C., & Hecht, R. (2015). Semantic enrichment of building data with volunteered geographic information to improve mappings of dwelling units and population. *Computers, Environment and Urban Systems*, 53, 4–18. doi:10.1016/j.compenvurbsys.2015.04.002
- Lantz, B. (2013). *Machine learning with R*. Packt Publishing Ltd.

Lin, W. (2013). Volunteered geographic information and networked publics? Politics of everyday mapping and spatial narratives. *GeoJournal*, 78(6), 949–965. doi:10.1007/s10708-013-9490-1

Resch, B., Summa, A., Sagl, G., Zeile, P., & Exner, J. P. (2015). Urban emotions—Geo-semantic emotion extraction from technical sensors, human sensors and crowdsourced data. In *Progress in Location-Based Services 2014* (pp. 199–212). Springer International Publishing. doi:10.1007/978-3-319-11879-6_14

Salazar, J. C. (2016). *Monitoreo urbano de entidades y eventos geográficos basado en censado social* (Unpublished master dissertation). Centro de Investigación en Computación, Instituto Politécnico Nacional, Mexico.

Scharkow, M. (2013). Thematic content analysis using supervised machine learning: An empirical evaluation using German online news. *Quality & Quantity*, 47(2), 761–773. doi:10.1007/s11135-011-9545-7

Sidorov, G., Velasquez, F., Stamatatos, E., Gelbukh, A., & Chanona-Hernández, L. (2014). Syntactic n-grams as machine learning features for natural language processing. *Expert Systems with Applications*, 41(3), 853–860. doi:10.1016/j.eswa.2013.08.015

Sui, D., Elwood, S., & Goodchild, M. (Eds.). (2012). *Crowdsourcing geographic knowledge: volunteered geographic information (VGI) in theory and practice*. Springer Science & Business Media.

Wakefield, J. (2013). *Tomorrow's cities: How big data is changing the world*. BBC.

Zimmerman, R. A. (2015). *An economic history of Latin American developmental approaches: the political economy of world trade*. Towson University Institutional Repository.

ADDITIONAL READING

Andrienko, G., Andrienko, N., Bosch, H., Ertl, T., Fuchs, G., Jankowski, P., & Thom, D. (2013). Thematic patterns in georeferenced tweets through spacetime visual analytics. *Computing in Science & Engineering*, 15(3), 72–82. doi:10.1109/MCSE.2013.70

Backstrom, L., Sun, E., & Marlow, C. (2010). Find me if you can: improving geographical prediction with social and spatial proximity. In: *Proceedings of the 19th international conference on World wide web* (pp. 61-70). ACM. doi:10.1145/1772690.1772698

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.

Ribeiro, S. S. Jr, Davis, C. A. Jr, Oliveira, D. R. R., Meira, W. Jr, Gonçalves, T. S., & Pappa, G. L. (2012) Traffic observatory: a system to detect and locate traffic events and conditions using Twitter, In: *Proceedings of the 5th International Workshop on Location-Based Social Networks* (pp. 5-11). ACM. doi:10.1145/2442796.2442800

Saldana-Perez, A. M. M., & Moreno-Ibarra, M. (2016). Traffic Analysis Based on Short Texts from Social Media. *International Journal of Knowledge Society Research*, 7(1), 63–79. doi:10.4018/IJKSR.2016010105

KEY TERMS AND DEFINITIONS

Gazetteer: A geographical dictionary that contains information about socio-economic statistics and physical features of a geographic area.

Geocoding: The process to obtain coordinates from spatial reference data on texts, such as street names or landmarks.

Metadata: The data that contains some other data or labels it, to describe its content, such as coordinates or URL information.

Microblogging Service: A service to distribute content to a group of members by internet, such content could be short texts, small audios or video links.

RSS Service: A document used to publish frequently updated information, which includes text and metadata.

Social Network: A web platform where people shares interests, activities and establish social relations by virtual connections.

Traffic: The movement of people, vehicles or merchandise through the roads, it can be fluent or congested.

Communication Privacy Management and Mediated Communication

S

Debra L. Worthington
Auburn University, USA

Margaret Fitch-Hauser
Auburn University, USA

INTRODUCTION

Sandra Petronio (1991) introduced communication privacy management theory (CPM) to explain how individuals control and reveal private information. While it was originally developed as an organizing principle for understanding disclosure in traditional social interactions, it has since been extended to a number of contexts, most recently to evolving communication technologies and social networking sites, including online blogging (e.g., Child & Agyeman-Budu, 2010; Child, Petronio, Agyeman-Budu, & Westermann, 2011), Facebook usage (e.g., De Wolf, Willaert & Pierson, 2014; Waters & Ackerman, 2011), and Twitter and Short Message Service (SMS) (e.g., Cho & Hung, 2011; Jin, 2013; Patil & Kobsa, 2004). CPM provides a set of theoretical tools to explore the intersection of technology and individual privacy in relationship management. Below privacy is defined, components of communication privacy management theory and their application to mediated communication are outlined, and areas of future research are presented.

BACKGROUND

Both a dynamic and dialectic process, the notion of privacy suggests that individuals regulate boundaries of disclosure, personal identity, and temporality (Palen & Dourish, 2003). More specifically, it refers to our ability to manage when, how, and

the extent to which our personal information is revealed to others (Westin, 1967).

When discussing the intersection of technology and privacy, people often focus on technical issues associated with technology use (see, for example, Boyles, Smith, & Madden, 2012). In reality, individuals focus significant attention on managing privacy in their online digital lives. CPM provides a means to better understand and explain how individuals use and communicate in online and mediated communication contexts (Child & Petronio, 2011).

COMMUNICATION PRIVACY MANAGEMENT THEORY

Originally developed for interpersonal contexts (Petronio, 1991), research associated with CPM initially focused on social and interpersonal interactions in areas such as family and health communication. (e.g., Petronio, 2006; Petronio & Caughlin, 2005; Petronio, Jones, Morr, 2003).

Petronio (2007) describes CPM theory as “an evidenced-based, applied theory construct to be translatable into practices” (p. 219). The CPM system rests on three elements – privacy ownership, privacy control and privacy turbulence. Eight axioms predict privacy practices (Petronio, 2013). The first two axioms are associated with privacy and the ownership of personal information. Axiom 1 proposes that individuals believe in private ownership of their personal information

and in their ability and right to share or protect that information from others. Axiom 2 predicts that when access to private information is granted to others, those gaining access become co-owners of the information, taking on the trust and responsibility that comes with co-ownership.

Axioms 3 – 6 are associated with privacy control. Petronio (2013) described privacy control as the regulating engine for determining the conditions of providing or denying access to private information. Thus, not only do individuals believe they are sole owners of their personal information (i.e., Axiom 1), but they also believe they alone control their personal privacy, even when that information is shared with others (Axiom 3). At the same time, how information is shared is based on the privacy rules individuals develop (Axiom 4). Core and catalyst criteria influence decisions on how and when rules are invoked. Core criteria are the most stable and predictable guidelines for privacy choices, while catalyst criteria result in privacy rule changes based on motivation and risk assessments.

Axiom 5 addresses how, once access to private information is shared with others, the original owner continues to maintain control by continued coordination and negotiation of privacy rules associated with third-party access (Petronio, 2013). However, ownership rights can be challenged when individuals manage multiple, often inter-related, privacy boundaries (e.g., can information revealed by a friend be shared with another mutual friend) (Petronio, 2002). Confidants fall into two categories – deliberate confidants purposely ask for information (e.g., bank employee and customer), while reluctant confidants receive unwanted private information (e.g., a third party present during a mobile phone exchange). Reluctant confidants may experience unwanted feelings of obligation and responsibility (Petronio & Reiersen, 2009). If the parties can reach a consensus about privacy rules, and accept the means by which they became deliberate confidants, then the confidant relationship can be effectively regulated.

The complications of collective co-ownership are seen in Axioms 6 and 7. Co-ownership leads to mutually agreed upon and practiced privacy boundaries where all members of the group can engage in sharing private information (Axiom 6, Petronio, 2013, p. 10). These group held privacy boundaries are regulated by decisions about who may divulge what information to whom and when. (Axiom 7, Petronio, 2013, p. 11). Thicker boundaries suggest that the coordinated rules of those collectively holding private information are relatively closed, while thinner boundaries are more permeable, resulting in information that is more accessible and open to third parties (Petronio & Reiersen, 2009). The original owner of the information and the confidant negotiate the level of access third parties may have, including the scope and extent of private information that can be shared.

The purpose of these boundaries is to govern who has control of and access to information as well as how to protect that information (Petronio, Sargent, Andea, Reganis, & Cichocki, 2004). People manage or coordinate privacy boundaries based on negotiation of privacy rules related to linkages, boundary permeability, and information ownership (Petronio, 2002). Privacy rules are both normative and situational and affected by a number of factors, including cultural expectations, individual motivations, risk-benefit assessments, gender, and the needs of the situation (Petronio, 2009). Importantly, multiple rules may be used during the boundary management process.

The final axiom, Axiom 8, addresses the area of privacy turbulence, and acknowledges that privacy regulation does, at times, fail and rules are broken. Privacy boundary turbulence often results from confidentiality breaches (i.e., privacy expectations of the original owner of information are not met by co-owners) (Petronio & Reiersen, 2009). Violations of confidentiality – discrepancy breaches of privacy, privacy ownership violations, and preemptive privacy control – can negatively affect the relationship of those involved.

APPLYING CPM TO MEDIATED COMMUNICATION CONTEXTS

The principles of CPM have been used to explain how individuals use social networking sites as well as emerging communication technologies. This section illustrates how CPM's principles and axioms can be and have been applied in these areas. The first principle of CPM asserts that individuals own their private information. In many ways, this norm was strengthened with the advent of mobile technology (Bergvik, 2004). For example, mobile phones have shifted from a location-based technology to person-linked devices. Today, they are viewed as a central element to individual identity and public presentation (Arminen, 2007; Campbell, 2008), and subsequently as private and personal property (Rosen, 2004; Häkkinen & Chatfield, 2005). This norm has become so well established that few individuals will answer another's cell phone in recognition that calls and texts are private belonging to the owner of the phone. The result is that incoming calls are viewed as private communication with the phone's owner even when the caller or receiver is in a public setting. Mobile phone users can manage privacy concerns by strategically choosing their method of communication (e.g., instant messaging, SMS, voice, etc.) (Worthington, Valikoski, Fitch-Hauser, Imhof, & Kim, 2012). Each means of communication comes with its own set of privacy concerns with some mediums implying greater ownership and having thicker boundaries than others. For example, Häkkinen and Chatfield (2005) found SMS messages are assigned greater privacy, and are seen as more confidential, than voice calls.

In online communication, individuals will sometimes treat public domain space as if it is private (Child & Agyeman-Budu, 2010; Child & Petronio, 2011). One consequence may be posting, tweeting, blogging or otherwise sharing information that results in follow-up measures to manage privacy (Axioms 3 & 7). Bloggers who remove or "scrub" a message from their blog have reevaluated the risk-benefit ratio of retaining a

message on their blog (Child, et al., 2011). Thus, blog scrubbing results from the recognition of potential privacy turbulence (Axiom 8) that may result from privacy violations. A number of factors may lead bloggers to change their regular privacy rules, among them impression management, safety, and relationship considerations (Child et al., 2011). Blogger perceptions of privacy appear to fall along a continuum ranging from those who are highly aware of privacy issues and proactively weigh the risks (the cautious blogger) to those who never worry about their privacy (high risk-taking bloggers). In these two cases, blog scrubbing will rarely occur. However, the normative blogger is cognizant of triggers and will adjust privacy rules to situational demands.

Axiom 2 predicts that when access to private information is granted to others, these gaining accesses become co-owners of the information, taking on the trust and responsibility that comes with co-ownership. Facebook postings are an example of information that individuals move across their privacy boundary to become collectively co-owned with Facebook friends (Child & Petronio, 2011). In such cases, there is a loss of individual control in favor of being social and interacting with others in their network (Bateman, Pike, & Butler, 2011; Child et al., 2011). These friends become authorized co-owners of the information. However, users who are more protective of their privacy may engage in a variety of strategies, including providing less information, making vague postings, and using coded language (Axiom 4; Child & Starcher, 2016; Child, Duck, Andrews, Butauski, & Petronio, 2015). There are a number of reasons for increased privacy concerns with Facebook and similar social networking sites. Users may be concerned with who they are interacting with or be worried about family members, romantic interests, colleagues, employers and others who may be engaging in mediated lurking (Child & Westermann, 2013; Child et al., 2015; Frampton & Child, 2013). Of course, users can reclaim privacy by deleting postings. Notably, they can also refuse to accept responsibility for

disclosures placed on them by others. This refusal can be seen when individuals delete comments or other information posted on their personal site by others (i.e., reluctant confidants, Axiom 5).

Related to social networking sites, De Wolf and colleagues (2014) suggest that just as individuals create personal boundaries, they also create group boundaries with others with whom they share information. Adequate feelings of control and privacy management reside with the group. Their study supports Petronio (2002) assertion that individuals must establish both individual and group privacy norms in order to have adequate management of privacy issues.

Finally, privacy perceptions differ across cultures and countries. For example, in a Malaysian study, Mohamed and Ahmad (2012) identified several reasons individuals are concerned with information privacy associated with technology, communication, and social networking sites. They note that a user's perception of the gravity of losing any kind of personal information leads to increased privacy behaviors. They also report that protective behavior use will increase when users find it easy to use security settings. In a study of the effect of cultural differences on mobile phone usage, Worthington et al. (2012) reported differences in user privacy concerns across the four countries they studied (Finland, Germany, South Korea, and the U.S.). For example, Finnish and U.S. respondents were more sensitive to the interaction between conversational topic and public location than South Korean and German participants.

FUTURE RESEARCH DIRECTIONS

Arguably, one of the strongest concerns individuals have with communication technology is how to manage and control privacy. Privacy, however, goes beyond how to protect financial information or individual passwords to include how to manage personal disclosures. Applying the theoretical

framework of Communication Privacy Management to current and emerging technology provides insight into communication behaviors of its users as they balance privacy issues against their daily communicative needs. No matter the means of communication, receivers make choices on what to disclose and what will remain private when communicating with others. Decision criteria aid in the development and employment of underlying privacy rules (Petronio, 2002; Petronio & Durham, 2008). A better understanding of the criteria used to make privacy decisions when using emerging communication technology and social networking sites will enable researchers to better predict communication behaviors.

Decision criteria are affected by a number of factors, including cultural and individual differences. Like many areas of communication studies, comparisons across cultures are relatively rare (for an exception, see Worthington et al., 2012). Such studies would allow for the identification of commonalities as well as cultural differences in how individuals manage privacy, assess co-ownership, and develop privacy rules. Along with studies exploring cultural differences, research into individual differences is needed. How might differences in self-consciousness, interaction-involvement, communication apprehension, compulsive communication, and related psycho-social constructs impact privacy management decisions and rule development?

While a number of studies have focused on Facebook users (e.g., De Wolf, Willaert & Pierson, 2014; Waters & Ackerman, 2011), few have addressed CPM by users of other sites (e.g., LinkedIn, MySpace, Meetup, VK, etc.). Such studies may provide a better understanding of underlying factors associated with boundary permeability and turbulence, particularly as related to areas such as the effect of privacy breakdowns on social relationships, self-disclosure, and privacy management practice.

Finally, the intersection of communicative affordance and CPM provide another potential area

of study. An affordance describes the “mutuality of actor intentions and technology capabilities that provide the potential for a particular action” (Majchrzak, Faraj, Kane, & Azad, 2013, p. 39). Schrock (2015) argues that communication has been inherent in many explanations of social affordances, particularly those that highlight the role of perceptual cues on individual perception and social networks. This approach to affordance goes beyond design features (e.g., apps, screens, etc.) to include the impact of a technology on human communication processes, such as communication privacy management. How might perceptions of technology and communication affordance be informed by user concerns of personal privacy?

CONCLUSION

Emerging communication technologies and social networking sites have significantly impacted how individuals and groups communicate with one another. However, while the methods of communication have changed and while they certainly can affect communication processes, at their center are people communicating with other people. Communication Privacy Management theory represents one of many human communication theories that may provide insight into how individuals use technology to communicate with one another. Through its application, researchers may ultimately find that while the means of communication change, individual privacy concerns remain the same.

REFERENCES

- Arminen, I. (2007). Review essay: Mobile communication society? *Acta Sociologica*, 50(4), 431–437. doi:10.1177/0001699307083983
- Bateman, P. J., Pike, J. C., & Butler, B. S. (2011). To disclose or not: Publicness in social networking sites. *Information Technology & People*, 24(1), 78–100. doi:10.1108/09593841111109431
- Bergvik, S. (2004). *Disturbing cell phone behavior – A psychological perspective. Implications for mobile technology in tourism. (Report No. R 46/2004)*. Fornebu, Norway: Telenor Group.
- Boyles, J. L., Smith, A., & Madden, M. (2012). *Privacy and data management on mobile devices*. Pew Research Internet Project. Retrieved July 19, 2016 from <http://pewinternet.org/Reports/2012/Mobile-Privacy.aspx>
- Campbell, S. W. (2008). Mobile technology and the body: Apparatus, fashion, and function. In J. E. Katz (Ed.), *Handbook of mobile communication studies* (pp. 153–164). Cambridge, MA: MIT. doi:10.7551/mitpress/9780262113120.003.0012
- Child, J. T., & Agyeman-Budu, E. A. (2010). Blogging privacy management rule development: The impact of self-monitoring skills, concern for appropriateness, and blogging frequency. *Computers in Human Behavior*, 26(5), 957–963. doi:10.1016/j.chb.2010.02.009
- Child, J. T., Duck, A. R., Andrews, L. A., Butauski, M., & Petronio, S. (2015). Young adults management of privacy on Facebook with multiple generations of family members. *Journal of Family Communication*, 15(4), 349–367. doi:10.1080/15267431.2015.1076425
- Child, J. T., & Petronio, S. (2011). Unpacking the paradoxes of privacy in CMC relationships: The challenges of blogging and relational communication on the internet. In K. B. Wright & L. M. Webb (Eds.), *Computer-mediated communication in personal relationships* (pp. 21–40). New York: Peter Lang.

- Child, J. T., Petronio, S., Agyeman-Budu, E. A., & Westermann, D. A. (2011). Blog scrubbing: Exploring triggers that change privacy rules. *Computers in Human Behavior*, 27(5), 2017–2027. doi:10.1016/j.chb.2011.05.009
- Child, J. T., & Starcher, S. (2016). Fuzzy Facebook privacy boundaries: Exploring mediated lurking, vague-bookings, and Facebook privacy management. *Computers in Human Behavior*, 54, 483–490. doi:10.1016/j.chb.2015.08.035
- Child, J. T., & Westermann, D. A. (2013). Lets be Facebook friends: Exploring parental Facebook friend requests from a communication privacy management (CPM) perspective. *Journal of Family Communication*, 13(1), 46–59. doi:10.1080/15267431.2012.742089
- Cho, V., & Hung, H. (2011). The effectiveness of short message service for communication with concerns of privacy protection and conflict avoidance. *Journal of Computer-Mediated Communication*, 16(2), 250–270. doi:10.1111/j.1083-6101.2011.01538.x
- De Wolf, R., Willaert, K., & Pierson, J. (2014). Managing privacy boundaries together: Exploring individual and group privacy management strategies in Facebook. *Computers in Human Behavior*, 35, 444–454. doi:10.1016/j.chb.2014.03.010
- Frampton, B. D., & Child, J. T. (2013). Friend or not to friend: Coworker Facebook friend requests as an application of communication privacy management theory. *Computers in Human Behavior*, 29(6), 2257–2264. doi:10.1016/j.chb.2013.05.006
- Häkkinä, J., & Chatfield, C. (2005). “It’s like if you opened someone else’s letter”: User perceived privacy and social practices with SMS communication. In M. Tscheligi, R. Bernhaupt, & K. Mihalic (Eds.), *Proceedings of the 7th International Conference on Human Computer Interaction with Mobile Devices & Services* (pp. 219–222). ACM. doi:10.1145/1085777.1085814
- Jin, S.-A. A. (2013). Peeling back the multiple layers of Twitter’s private disclosure onion: The roles of virtual identity discrepancy and personality traits in communication privacy management on Twitter. *New Media & Society*, 15(6), 813–833. doi:10.1177/1461444812471814
- Majchrzak, A., Faraj, S., Kane, G. C., & Azad, B. (2013). The contradictory influence of social media affordances of online communal knowledge sharing. *Journal of Computer-Mediated Communication*, 19(1), 38–55. doi:10.1111/jcc4.12030
- Mohamed, N., & Ahmad, I. (2012). Information privacy concerns, antecedents and privacy measure use in social networking sites: Evidence from Malaysia. *Computers in Human Behavior*, 28(6), 2366–2375. doi:10.1016/j.chb.2012.07.008
- Palen, L., & Dourish, P. (2003). Unpacking “privacy” for a networked world. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 129–136). New York: ACM Press.
- Patil, S., & Kobsa, A. (2004). Instant messaging and privacy. In S. Fincher, P. Markopoulos, D. Moore, & R. Ruddle (Eds.), *People and computers XVIII: Design for life. Proceedings of HCI 2004* (pp. 85–88). London: Springer-Verlag.
- Petronio, S. (1991). Communication boundary management: A theoretical model of managing disclosure of private information between married couples. *Communication Theory*, 1(4), 311–335. doi:10.1111/j.1468-2885.1991.tb00023.x
- Petronio, S. (2002). *Boundaries of privacy: Dialectics of disclosure*. Albany, NY: SUNY Press.
- Petronio, S. (2006). Impact of medical mistakes: Negotiating work-family boundaries for physicians. *Communication Monographs*, 73, 462–467. doi:10.1080/036377750601061174

- Petronio, S. (2007). Translational research endeavors and the practices of communication privacy management. *Journal of Applied Communication Research*, 35(3), 218–222. doi:10.1080/00909880701422443
- Petronio, S. (2009). Privacy management theory. In S. Littlejohn & K. Foss (Eds.), *Encyclopedia of communication theory* (pp. 797–799). Thousand Oaks, CA: Sage. doi:10.4135/9781412959384.n302
- Petronio, S. (2013). Brief status report on communication privacy management theory. *Journal of Family Communication*, 13(1), 6–14. doi:10.1080/15267431.2013.743426
- Petronio, S., & Caughlin, J. P. (2005). Communication privacy management theory: Understanding families. In D. Braithwaite & L. Baxter (Eds.), *Engaging theories in family communication: Multiple perspectives* (pp. 35–49). Thousand Oaks, CA: Sage.
- Petronio, S., & Durham, W. T. (2008). Communication privacy management theory. In L. A. Baxter & D. O. Braithwaite (Eds.), *Engaging theories in interpersonal communication: Multiple perspectives* (pp. 309–322). Thousand Oaks, CA: Sage. doi:10.4135/9781483329529.n23
- Petronio, S., Jones, S. M., & Morr, M. C. (2003). Family privacy dilemmas: Managing communication boundaries within family groups. In L. Frey (Ed.), *Group communication in context: Studies of bona fide groups* (pp. 23–56). Mahwah, NJ: LEA.
- Petronio, S., & Reiersen, J. (2009). Privacy of confidentiality: Grasping the complexities through communication privacy management. In T. Afifi & W. Afifi (Eds.), *Uncertainty, information regulation, and disclosure decisions: Theories and applications* (pp. 365–383). New York: Routledge.
- Petronio, S., Sargent, J., Andea, L., Reganis, P., & Cichocki, D. (2004). Family and friends as healthcare advocates: Dilemmas of confidentiality and privacy. *Journal of Social and Personal Relationships*, 21(1), 33–52. doi:10.1177/0265407504039838
- Rosen, C. (2004, Summer). Our cell phones, ourselves. *New Atlantis (Washington, D.C.)*, 6, 26–45. Retrieved from <http://www.thenewatlantis.com/publications/our-cell-phones-ourselves>
- Schrock, A. (2015). Communicative affordance of mobile media: Portability, availability, locatability and multimodality. *International Journal of Communication*, 9, 1229–1246.
- Waters, S., & Ackerman, J. (2011). Exploring privacy management on Facebook: Motivations and perceived consequences of voluntary disclosure. *Journal of Computer-Mediated Communication*, 17(1), 101–115. doi:10.1111/j.1083-6101.2011.01559.x
- Westin, A. F. (1967). *Privacy and freedom*. New York: Atheneum.
- Worthington, D. L., Valikoski, T., Fitch-Hauser, M., Imhof, M., & Kim, S. (2012). Listening and privacy management in cell phone conversations among young adults: A cross cultural comparison of Finnish, German, Korean, & U.S. American Students *Empedocles. European Journal for the Philosophy of Communication*, 3, 43–60. doi:10.1386/ejpc.3.1.43_1

ADDITIONAL READING

- Bellman, S., Johnson, E., Kobrin, S., & Lohse, G. (2004). International differences in information privacy concerns: A global survey of consumers. *The Information Society*, 20(5), 313–324. doi:10.1080/01972240490507956

Campbell, S. W. (2007). Perceptions of mobile phone use in public settings: A cross-cultural comparison. *International Journal of Communication, 1*, 738–757. <http://ijoc.org/index.php/ijoc/article/download/169/112> Retrieved July 20, 2016

Child, J. T., Haridakis, P. M., & Petronio, S. (2012). Blogging privacy rule orientations, privacy management, and content deletion practices: The variability of online privacy management activity at different stages of social media use. *Computers in Human Behavior, 28*(5), 1859–1872. doi:10.1016/j.chb.2012.05.004

Cho, H., Rivera, M., & Lim, S. S. (2009). A multinational study on online privacy concerns and local response. *New Media & Society, 11*, 409–431. doi:10.1177/1461444808101618

Kim, H.-S. (2016). What drives you to check in on Facebook?: Motivations, privacy concerns, and mobile phone involvement for location-based information sharing. *Computers in Human Behavior, 54*, 397–406. doi:10.1016/j.chb.2015.08.016

Kobsa, A., Patil, S., & Meyer, B. (2012). Privacy in instant messaging: An impression management model. *Behaviour & Information Technology, 31*(4), 355–370. doi:10.1080/01449291003611326

Mullen, C., & Hamilton, N. F. (2016). Adolescents response to parental Facebook friend requests: The comparative influence of privacy management, parent-child relational quality, attitude and peer influence. *Computers in Human Behavior, 60*, 165–172. doi:10.1016/j.chb.2016.02.026

Serewicz, M. C. M., & Petronio, S. (2007). Communication privacy management theory. In B. Whaley & W. Samter (Eds.), *Explaining communication: Contemporary theories and exemplars* (pp. 257–273). Mahwah, NJ: LEA.

Worthington, D. L., & Fitch-Hauser, M. (2015). Communication privacy management and mobile phone use. In Z. Yan (Ed.), *Encyclopedia of mobile phone behavior* (Vol. 2, pp. 937–949). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-8239-9.ch075

KEY TERMS AND DEFINITIONS

Boundary Permeability: The level of access, scope, and extent of private information an individual is willing to share with others.

Boundary Turbulence: Occurs when privacy expectations of the original owner of information are not met and a confidentiality breach occurs.

Communication Privacy Management Theory: A theory addressing the tension people experience when choosing what personal information to reveal and what will remain private when interacting with others.

Confidants: Individuals who are given access to private information.

Confidentiality Breaches: Occur when privacy expectations of the original owner of information are not met by co-owners of information.

Deliberate Confidants: Individuals who are purposely given private information and readily take on obligations and responsibility associated with co-owning the information.

Information Co-Ownership: Occurs when access to private information is granted to others, who then take on the trust and responsibility that comes with co-ownership of previously private information.

Privacy Boundaries: Personal boundaries separating public and private information. They govern who has control of and access to personal information.

Private Information Ownership: The belief that individuals own their personal information and have the ability and right to share or protect that information from others.

Reluctant Confidants: Individuals who receive unwanted private information and may experience unwanted feelings of obligation and responsibility.

The Dual Nature of Participatory Web and How Misinformation Seemingly Travels

S

Sameer Kumar

University of Malaya, Malaysia

INTRODUCTION

Social media is called ‘social’ as it is primarily ‘produced’ by people (user-generated) and there is less or no commercial aspect involved. The platform lets people to communicate in horizontal using multimedia (text, photos, videos, music, etc.), thus, also making them reporters and publishers in the process. Social media has brought in a revolution in the way people express their views and share their likes and dislikes (Asur, Huberman, Szabo, & Wang, 2011; Ellison, 2007; Kumar & Jan, 2011, 2012). This media sits on Web 2.0 technology. In this article the author argues that the web is a neutral medium which could be used from transmitting good and bad information. A hypothetical model of how misinformation travels is also presented.

BACKGROUND

Web 2.0 does not refer to the update in the technical specification, but rather new applications and technology that makes web experience more interactive. While talking of Web 2.0 the applications such as Ajax or Openlaszlo are stated, which make the web as a ‘Participatory Web’. This is in contrast to the earlier version Web 1.0 which was essentially an ‘Information Web’. The main features of Web 2.0 are user-interaction, dynamic content, meta data and scalability (Best, 2006). Social media sites such as Facebook and Flickr make use of Web 2.0 technology in order to give interactivity to their sites. McAfee (2006) refers to the following technological features of Web

2.0, acronymed as SLATES – Search, Links, Authoring, Tags, Extension, Signals. Search enables keyword search, Links helps in linking to other documents within the page or elsewhere on the Net, Authoring refers to the ability to update, iterate and cumulate content, Tags help in searching and helping to avoid category rigidity, Extensions help in leveraging the Web as both document server and application platform, and finally, Signals that help to inform users of changes in the social media content. In more technical terms, Web 2.0 technology is based on two sides – Web browser side (client) and Web server side (server), and they utilize softwares like Ajax, JavaScript, Flash etc. to fetch data to and from the Web server and to carry on instant updates. Flash software technology is also used in music and video sites. YouTube is a good example of Flash technology. The popularity of Web 2.0 has encouraged all web allocations with collaborative and interactive functionality to append the term 2.0 to it – interestingly these name are, for example, Library 2.0, Publishing 2.0, Social Work 2.0, Travel 2.0, and the list goes on. Essentially all these Web applications use Web 2.0 technologies to provide better functionality in their domain of business.

In this article that author argues that although these new technologies are making Web even more popular and accessible, they are also making people spend long hours on seemingly unimportant and irrelevant conversations, dialogues, reviews and opinions. According to Andrew Keen, writer of ‘*Cult of the Amateur*’, Web 2.0 has “created a cult of digital narcissism and amateurism, which undermines the notion of expertise” (“Web 2.0 has created a cult of digital narcissism and

DOI: 10.4018/978-1-5225-2255-3.ch606

amateurism?," 2010). Just as a river, that carries everything – good or bad – as it flows, social media makes no discrimination between what it carries. Some people use it for a good cause, some as a campaign tool, while others use it to spread hatred and false news.

Through surveys and content analysis of some comments of Social Media, which includes blogs, discussion forums, newscasts, market research studies and traditional websites, on three determinants – 1) social service, 2) politics and 3) misinformation – we do a critical examination to see if social media is a double-edged sword – a technology which could be used to transmit both good and bad information.

A Great Vehicle for Political Propaganda

In politics, people elect their representatives. These representatives then govern a county or a country. Political news have always captured the front pages of our newspapers. A country is more secured and safe if its political climate is stable. The way people elect their representatives and the way representatives campaign during the elections, differ from place to place and from country to country. Let us take the example of two biggest democracies in the world – India and USA. In India, a large percentage of people who cast their votes are still farmers, low income group people and the elderly. Political parties in India, thus, go all out in wooing the farmers and it's the farmers who actually bring a candidate to power. Unfortunately, most of the farmers are either illiterate or semi-literate who have almost nothing to do with Computers or latest online technologies. One can forget about how Social media would ever affect their voting decisions. In contrast, USA is a country where most of the voters are literate and have access to internet and other latest electronic gadgetries. The way social networking is bringing people together, its communicating power and speed, its "peer to peer" and "word of mouth" potency have been understood

quite well and quite early in the game of politics in America.

Social media based videos in particular are becoming a popular medium for political propaganda (Noguera & Corretero, 2010). This was clearly evident in the last major US presidential elections (year 2008) which remained in the news for over 2 years – first the elections between Hillary Clinton and Barack Obama within the party and then the fight to the finish between the Democrats and the Republicans. Both McCain and Obama used Social media technologies to the hilt. We particularly know Obama for his Social media savoir-faire. Barack Obama appointed Scott Goodstein as External Online director ("How Obama used social networking tools to win," 2009) for *Obama for America*. According to Scott, there were three main factors instrumental in the effectiveness of the campaign – 1) using deadlines to test new ideas, 2) allowing consumers to engage and, 3) validate and moving with the marketplace. There's lot happening on the Net which one can either ignore or engage in it. Scott believed that they were able to successfully engage and move with the marketplace, which helped them win the elections. His social media strategies helped attract 1.7 million supporters on Twitter, 6.5 million on Facebook and over 2 million supporters on MySpace. Speaking in Singapore, he stated that Social Media put Obama's slogan 'Change you can believe in', front and centre. Obama, himself an exceptional communicator, needed the right messenger and social media proved to be just that right one. It helped Barack Obama's message to clearly reach the audience. People were talking and discussing about him in Facebook, MySpace, Twitter and everywhere.

US has been seeing a shift in media consumption. There are lesser and lesser number of people interested in TV or newspaper. More and more people have been reaching out to Internet for news and entertainment. Understanding this, the online campaign team of Obama meticulously handled social networking campaign platforms, including those that were catering to the ethnic

communities. Obama's willingness to experiment with new communication tools gave the initial fillip. In the present times, the competitiveness in elections is intense and just 5% votes shift can make a candidate lose or win elections. Propaganda through social media coupled with good support from ethnic groups - such as Indian-americans, Afro-americans, etc, probably gave Obama the vital vote shift.

However, if Social Media could give positive results, it could also do the opposite. This is what happened with Sarah Palin and McCain lost the US presidential elections.

One of the bloggers ("The Perils of Modern Communication," 2009) comments and writes about what he read in WaPro:

The difference is that now the masses are technologically enabled, amplified by a twillion tweets. Everybody's got a megaphone, bless democracy's heart.

But when a protest of one (or a few) can instantly morph into a babble of thousands, rabble-rousing becomes a hobby -- and rational debate becomes an oxymoron.

The comment speaks about how a protest can amplify manifold.

In the case of Barack Obama one saw how Twitter updates, Facebook and Myspace communities helped him gather funds, volunteers and publicity. A dedicated social networking site mybarackobama.com, powered by 'six apart' company, was extensively used. However, after the campaign was over, his twitter stream went silent for weeks and there was little to no engagement directly by the Obama camp on other sites as well. Probably, politicians have not yet realized that it is not just enough to make social media as a campaign tool but also to use as an effective tool to connect to its faithful voters post-election results. It may also be true that political parties are scared that the power of social media would soon make them unpopular, if unpopular policies

are enacted (i.e increase in taxes), and hence are not comfortable at engaging with people post elections, as yet.

We, thus, are seeing that successful experiment with Social Media in politics has now made politicians and political parties to stand up and take notice. It also reminds them that it is not just sufficient to win elections but live up to their manifesto. Because the users of Social Media are no bovines – they will write, share and speak-up about any topic or issue. If they do not perform while in power, the same people who voted for them would bring them down, when they stand up for the next term. For political parties and politicians, writings and feedback on social media are a nice way to gauge what people are thinking about them.

A Platform for Social Service

In day-to-day context, the word 'social service' often means 'caring for humanity'. When we hear that certain organization is doing social service, we understand that they are providing humanitarian service. So how can this Social media work for social service or cause? Can it help in social entrepreneurship?

Late C.K. Prahalad, the management Guru, said that it is important to target the BoP (Bottom of Pyramid), in order to eradicate poverty (Prahalad, 2010). Half of the people in the world are living under \$2.5 a day. As commercial companies are more and more targeting people who have money, there is a widening gap between the have and have-nots. This gap could be targeted by social entrepreneurs. It is a huge untapped market that could be harnessed both for profit and eradication of poverty. Melting of borders has also made social entrepreneurs bring their innovations to different parts of the world. The success of Grameen Bank's micro-credit model, for example, is now being implemented in other parts of Asia, Africa and even Europe, with varying degree of success.

Social entrepreneurs in the west are leveraging the power of social media to reach out to the devel-

oping world. Yonus's Grameen bank in Bangladesh is one of the earliest successful examples of social entrepreneurship. Grameen bank's unique lending model went on to prove that people would pay back their loans even when they are given without co-laterals. Grameen bank had a 99% loan return rate which is much higher than commercial banks. The same micro-finance concept is taken up of kiva.com that works as a connector between the people willing to lend money to people in need. People who need money are first verified by the regional field partners and once verified, their profile is posted on the kiva website. By remitting as low as 25\$, one can contribute to one or multiple loan-requesters. People who lend money are satisfied that it is going for the right cause. Once the borrower starts repaying, the money is repatriated to the lenders in a phased manner,. Field partners keep the interest to defray their expenses. The default rate is near zero and the success is huge, so much so, that kiva.com as of 2013 had helped people with about half a billion US\$ of loans. A loan request is fulfilled in just 15 hours, on an average.

The question remains whether globalization has created a new breed to social entrepreneurs with fire in their belly to do something transformational and bring about a social improvement. The internet has brought the awareness of social entrepreneurship. For example, many people have come to know about this kind of entrepreneurship by visiting the asoka.org or kiva.com websites.

Improvement in the level of education is another area where social entrepreneurs have been working on. Negroponte's one-laptop-per-child (<http://laptop.org/en/>) is a wonderful example of this. One laptop per child is MIT's project that has developed a laptop for school children and it costs under US\$200. The project has plans to bring it down to US\$100 eventually. It is a major project that it being implemented in association with the governments of both the developing and the developed worlds. India too is working to come out with an ultra-cheap laptop for school children that some say would well cost below US\$ 100. It

is important here to know that these projects help students to have access to latest digital technologies. Better access to information over social media and learning material undoubtedly improves their knowledge level and grasping power. It the Social media that spreads the word of mouth. You one has something worthwhile to sell or give and if people like it, it will spread like wildfire over the internet. One-laptop-per-child would not have become known across the world had it not been discussed in blogs and other niche networking sites. It is important to understand that Social Media is not only about sites like Facebook and Youtube. It is also about such sites like Kiva.com that are able to provide direct help thousands and thousands of needy individuals.

One of the bloggers ("The Humanitarian Power of Social Media," 2009) a recent impact of social media in motivating people to donate blood. The bloggers decided to donate blood in Austin, Texas, and they had just three days to spread the message. To promote the event, the bloggers used blogs and asked other bloggers to write about the event. The result was satisfying – over 100 people turned up to donate blood at a centre that gets on an average 40 people donating blood on a given day. Among the donors were people who were donating blood for the first time. In another example ("Capital Area Food Bank of Texas," 2009) of how social media could benefit humanity – a company TysonFoods decided to donate 100 pounds of food (limited to a maximum of 35,000 pounds) for every comment they receive on the website. This message was twitted across (through microblogging site Twitter.com) and within 6 hours the limit was reached. The company got over 350 comments and the donation of 35,000 pound of food was made to Capital Area Food Bank of Texas's HAM-up event. Social media tools like weblogs, social networks and video chats are used in both large and small hospitals which are 'reengineering' the way patients and doctors interact (Hawn, 2009)

The consumer to consumer communications have been greatly expedited due to the viral nature of social media (Mangold & Faulds, 2009). This

can send good message very fast. If one have a blood camp in his neighborhood or there is sudden need to help someone in dire need information could be circulated through Facebook. Twitter, the micro blogging network, is making this even faster. Just one twit and all followers of the individual on Twitter are immediately informed.

But Also a Vehicle for Misinformation

Social Media is a popular opinion sharing platform. It is to everybody's benefit if these opinions and views are constructive. Such an endeavour, for example, can help a child in need of blood of a rare blood group. However, what happens if the same platform is used for spreading hate or some other kind of misinformation? Social Media is only a medium - it carries everything and carries it fast. So if someone wants to spread rumour about a UFO or Michael Jackson rising from the dead, just posting a grainy video footage on Youtube can make the rumour spread like wildfire. It will be carried forward in other blogs, copied, Digged, Reddited, Stumbled upon and Twitted in no time. Harmful rumors spread and cause people to panic (Hashimoto, Kuboyama, Shirota, & Ieee, 2011) To prove this point, RTL carried out an experiment to show how fast rumours could travel on the social media. They posted a fake video of Michael Jackson (now no more alive) on Youtube. Within a matter of just one day it was viewed over 7,00,000 times and many believed that the video was true. The video was removed from YouTube the next day.

SOLUTIONS AND RECOMMENDATIONS

In social Media, a person is credible if he has a popular blog or if he active in the forums. For example, Sarah Palin, the vice presidential candidate of the US 2008 elections, was the governor of Alaska, enjoying over 93% popularity, when she was chosen by McCain, in the run-up to the

US Presidential elections. However, in the social media there was grapevine that Sarah Palin's daughter Bristol was infact the mother of her (Palin's) last child ("Governor Palin Babygate Debunked!," 2008). The source that spread this news went blow-by-blow giving reasons and proof of why this was true. Palin camp did not take the initiative to clear the clouds till the time it became a very hot topic in the formal and informal media. Infact, the formal media (newspaper and TV) took cues from what was circulating in social media and made the topic hotter. McCain Camp tried to douse the fire by doing some unusual things like even carrying out *Adwords* advertisement and driving that traffic to their election website. Palin did come up later to tell that her teen-aged daughter was pregnant with her boyfriend Levi Johnston ("Bristol Palin," 2009). However, the damage was so huge that, some say that this issue was one of the contributors to McCain's loss in the elections. Furthermore, once immensely popular Sarah had her popularity hugely diminished. In 2009, she resigned as the Governor of Alaska. Fig. 1 shows a way rumour through video seemingly spreads.

Misinformation is part of Social Media. Not only for popular persons, it could also happen to products or companies. There have been several cases, especially small companies, who have suffered heavily in the hands of this social media con-artists. Those who only have internet business are always at a risk to suffer the maximum damage if false information gets around. They are picked up by search engines, which keep it indexed for ever (blog posts never die). Someone determined to cause brand damage to a company or product could set-up blogs, join forum and post false information. Depending how influential the blogger is, he could make a serious dent in the popularity of the product or company. So the question again arises, how to do the damage control? Here I present a few steps, in Figure 2 that could be taken when misinformation is being spread and the individual/company is made aware of it.

Microblogging with Twitter is increasing the speed of information and misinformation dissemi-

Figure 1. How a rumor or misinformation seemingly travels through video in YouTube

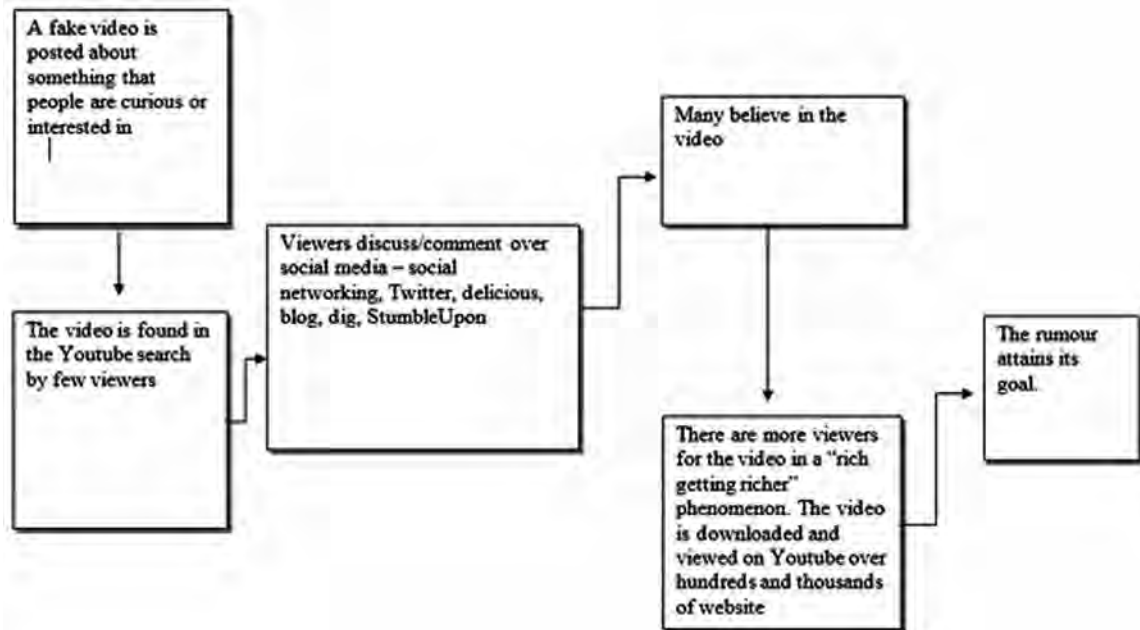
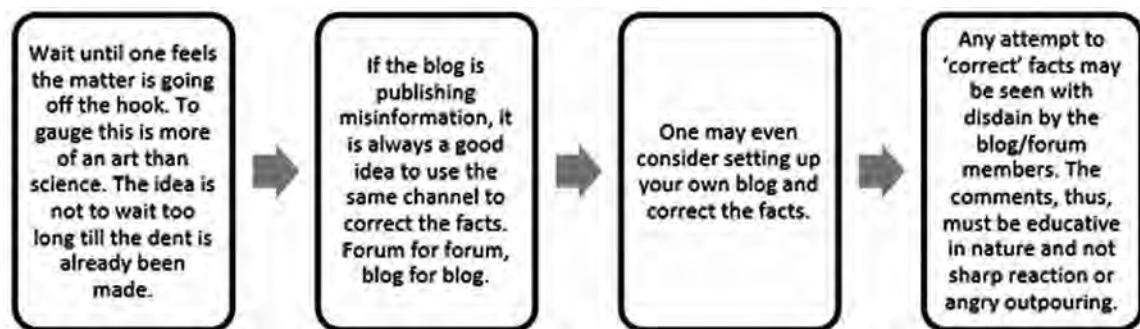


Figure 2. Possible steps to control misinformation damage on social media



nation over the social Media. 80% of Twitter's 140 characters-or-less messages are delivered through the mobile phones or PDA. Suppose some popular Twitter user has 30,000 followers, in one instant he could influence that many followers. Additionally, followers of this popular person may have their own following. If the information is worth transmitting these followers twit to their followers, creating a dizzy chain effect. Millions of people are made aware of information – false or true, in just

a few hours or may be a few minutes. One of the bloggers ("Social media diseases," 2009) writes:

24. Art Says:
August 12th, 2009 at 1:58 pm
@IVAN3MAN:

"FALSE beliefs are everywhere. Eighteen percent of Americans think the Sun revolves around the Earth, one poll has found. Thus it

seems slightly less egregious that, according to another poll, 10 percent of us think that Senator Barack Obama, a Christian, is instead a Muslim. The Obama campaign has created a Web-site to dispel misinformation. But this effort may be more difficult than it seems, thanks to the quirky way in which our brains store memories — and mislead us along the way.”

Who did they survey and why am I not surprised?

In Twitter such misinformation is becoming common. The originator of this misinformation may be doing it for fun or may himself think it is true. However, once it goes few layers down, the information is modified, changed and believed even more. After traveling a few layers, the ‘node’ does not know where it originated from and thus, either believes or disbelieves it.

Information over a network travels fast and when the network is made of people who give credibility to the word of mouth, it travels even faster. It is pertinent to note that misinformation, rumors and negative news travels even faster. As the old adage goes “if you bite a dog it is news”. Because of the nature of humans to give importance to negative news, there is lot of false news circulating and being believed by the people.

Satirical sites such as Onion.com do not deliberately circulate misinformation (visitors know that these news are meant to entertain) hence must not be confused with sites, news or videos that pose themselves as true but in reality may be just hoaxes. One of the possible ways to reduce the harmful repercussions of incorrect and malicious information is to be informed if the information being transmitted is really true nor not. Those news that unbelievable are in the first category to be checked for its truthfulness. There are several sites that debunk misinformation, rumors, myths, etc – snopes.com (Mikkelsen, 2016), being among the most popular. With the US election season (2015-16) now in US, lot of political misinformation is floating around. To get to the bottom of truth, political fact checking could be done by visiting the sites like, politifact.com.

FUTURE RESEARCH DIRECTIONS

Although spread of misinformation is a topic that needs immediate scientific scrutiny, the research in this area has been laggard so far. In the wake of the recent Paris attacks (NBCNews, 2016) at a deluge of misinformation that also flooded the social media, there is an urgent need to conduct studies on how false and fabricated information. A hypothetical model is presented in the article. The spread of misinformation of which social media is the primary carrier nowadays, needs to be understood and checked so that people are aware that not everything that they see or hear on social media would be true. Scientific community too is under its grips where only certain portions of a scientific study for example is used to influence the idea. ‘Echo chambers’ bring about confirmation bias leading to misinformation such as conspiracy theories (Feldman, Myers, Hmielowski, & Leiserowitz, 2014; Harvey, 2016). Hence the future research could look into other models of how misinformation travels and the ways of limiting the spread of fabricated information of social media. Social network theory, for example, is a powerful way to understanding these mechanisms where a “influence limitation” could aid in restraining ‘bad’ information with ‘good’ information (Budak, Agrawal, & El Abbadi, 2011).

CONCLUSION

Unlike traditional media which has an editorial board to oversee what is published and what is not, social media is free for all – there are no editors and one could publish whatever he or she wanted. In Social Media there is no formal filter. The person receiving it has to use his/her judgment to separate misinformation from true information.

Here an attempt was made to understand the impact of Social media on Society. Social media is rapidly evolving, hence it is impossible to gauge its real impact. We learn that politicians in developed countries, like US, are using Social Media

for the purpose of election campaign. However, there is lack of engagement with voters after the campaign. SM is being used for Social Entrepreneurship, esp. Micro-financing. Volatile and viral nature of SM is carrying both information and misinformation at extremely fast pace. It may be rightly concluded that Social media is a medium, which could be a double-edged sword, capable of both good and harm.

REFERENCES

- Asur, S., Huberman, B. A., Szabo, G., & Wang, C. (2011). *Trends in Social Media: Persistence and Decay*. Arxiv preprint arXiv:1102.1402
- Best, D. (2006). *Web 2.0: Next Big Thing or Next Big Internet Bubble?* Lecture Web Information System Technische Universiteit Eindhoven.
- Budak, C., Agrawal, D., & El Abbadi, A. (2011). *Limiting the spread of misinformation in social networks*. Paper presented at the 20th international conference on World wide web. doi:10.1145/1963405.1963499
- Capital Area Food Bank of Texas. (2009). Retrieved from <http://austinfoodbank.wordpress.com/2008/08/12/are-ya-hankerin-for-a-ham-up/>
- Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1), 210–230. doi:10.1111/j.1083-6101.2007.00393.x
- Feldman, L., Myers, T. A., Hmielowski, J. D., & Leiserowitz, A. (2014). The mutual reinforcement of media selectivity and effects: Testing the reinforcing spirals framework in the context of global warming. *Journal of Communication*, 64(4), 590–611. doi:10.1111/jcom.12108
- Harvey, C. (2016). Here's how scientific misinformation, such as climate doubt, spreads through social media. *The Washington Post*. Retrieved from <https://www.washingtonpost.com/news/energy-environment/wp/2016/01/04/heres-how-scientific-misinformation-such-as-climate-doubt-spreads-through-social-media/>
- Hashimoto, T., Kuboyama, T., Shiota, Y., & IEEE. (2011). *Rumor Analysis Framework in Social Media*. New York: IEEE.
- Hawn, C. (2009). Take Two Aspirin And Tweet Me In The Morning: How Twitter, Facebook, And Other Social Media Are Reshaping Health Care. *Health Affairs*, 28(2), 361–368. doi:10.1377/hlthaff.28.2.361 PMID:19275991
- How Obama used social networking tools to win. (2009). *INSEAD Knowledge*. Retrieved from <http://knowledge.insead.edu/contents/HowObamausedsocialnetworkingtowin090709.cfm>
- Kumar, S., & Jan, J. M. (2011). *Social Networks and the Business Value of Social Media*. Paper presented at the International Conference on Humanities, Society and Culture (ICHSC 2011), Kuala Lumpur, Malaysia.
- Kumar, S., & Jan, J. M. (2012). *“I’ll e-mail you a handshake”: Is social media making us virtual?* Paper presented at the National Conference on Social Networking, Hyderabad, India.
- Mangold, W. G., & Faulds, D. J. (2009). Social media: The new hybrid element of the promotion mix. *Business Horizons*, 52(4), 357–365. doi:10.1016/j.bushor.2009.03.002
- McAfee, A. P. (2006). Enterprise 2.0: The dawn of emergent collaboration. *MIT Sloan Management Review*, 47(3), 21–28.
- Mikkelsen, D. (2016). Rumor has it. *Snopes*.

NBCNews. (2016). *Rumors and Misinformation Spread on Social Media Following Paris Attacks*. Author.

Noguera, J. M., & Correyero, B. (2010). The Impact of Politics 2.0 in the Spanish Social Media: Tracking the Conversations around the Audiovisual Political Wars. In J. G. Breslin, T. N. Burg, H. G. Kim, T. Raftery, & J. H. Schmidt (Eds.), *Recent Trends and Developments in Social Software* (Vol. 6045, pp. 152–161). Berlin: Springer-Verlag Berlin. doi:10.1007/978-3-642-16581-8_16

Palin, B. (2009). *Wikipedia*. Retrieved from http://en.wikipedia.org/wiki/Bristol_Palin

Prahalad, C. K. (2010). *The fortune at the bottom of the pyramid: Eradicating poverty through profits*. Wharton.

Social media diseases. (2009). *Bad Astronomy, Discover Magazine*. Retrieved from <http://blogs.discovermagazine.com/badastronomy/2009/08/12/social-media-diseases/>

The Humanitarian Power of Social Media. (2009). Retrieved from <http://www.tradeshowweek.com/blog/1620000362/post/1750034975.html>

The Perils of Modern Communication. (2009). *Social Media Today*. Retrieved from <http://www.socialmediatoday.com/SMC/122872>

Web 2.0 has created a cult of digital narcissism and amateurism? (2010). Retrieved from <http://folksonomy.co/?permalink=1758>

Jin, F., Wang, W., Zhao, L., Dougherty, E., Cao, Y., Lu, C.-T., & Ramakrishnan, N. (2014). Misinformation propagation in the age of twitter. *Computer*, 47(12), 90–94. doi:10.1109/MC.2014.361

KEY TERMS AND DEFINITIONS

BoP: In an acronym of ‘Bottom of Pyramid’ that refers to the poorest, yet largest strata of people. The poorest are usually defined by those group of people having earning of under USD 2.5/day.

Echo Chambers: the area of the virtual media where ideas and beliefs are reinforced through repetitive broadcast.

Influence Limitation: A process where misinformation (or ‘bad’) is diluted or replaced with correct (‘good’) information. Such an endeavor potentially reduces the harmful repercussions of misinformation.

Misinformation: False information, rumors, etc. circulating through the online social media.

Social Media Sites: Websites like Facebook, YouTube, Twitter, Instagram, etc, where people access the Internet to interact (chat, share photos and videos, communicate, etc.) with other people online.

Social Network Theory (SNT): SNT theory looks at associations in terms of ties and nodes, where nodes form a tie if there is some kind of relationship between them.

Web 2.0: Often referred to as the second version of World Wide Web (or simply ‘Web’) that has interactivity. The earlier version of Web had static pages with little or no interactivity.

ADDITIONAL READING

Chen, X., & Sin, S. C. J. (2013). ‘Misinformation? What of it?’ Motivations and individual differences in misinformation sharing on social media. *Proceedings of the American Society for Information Science and Technology*, 50(1), 1–4.

Effective Cultural Communication via Information and Communication Technologies and Social Media Use

Androniki Kavoura

Technological Educational Institute of Athens, Greece

Stella Sylaiou

Aristotle University of Thessaloniki, Greece

INTRODUCTION

Cultural organizations need to adopt contemporary methods and ways of communication for heritage management. Raising people's interest for cultural heritage may take place via the implementation of innovative communication tools that allow for engagement to take place. Research though has shown that cultural organizations are more reluctant than firms are to implement a two-way communication process with new technologies and social media. Do cultural organizations employ Information and Communication Technologies (ICTs) and social media in an interactive approach? This chapter aims to illustrate the significance that should be attributed to innovation through the use of technology for communication purposes by cultural organizations. The way communication trends take place in the field of cultural management and communication is presented, along with an overview of the technological outlook that can be invaluable. The use of innovative communication tools that are employed by cultural organizations for their communication strategies has been little researched as a topic. This study aims to add to the literature on the use of social media, mobile devices and ICTs by cultural organizations.

Technological advances have enabled a dynamic and effective communication of cultural organizations with their audiences and made more attractive. ICTs are used to enrich, transform and enhance the cultural experience. They can become

a powerful communication channel that may be a virtual 'direct' experience. They can disrupt linear sequences and provide usable and accessible in technology, content, navigation solutions for all people, irrespective of their age or abilities. Furthermore, the ICTs may permit exploration of cultural information, learning-by-doing in game-like environments that foster creativity and innovation.

The motivation for this chapter was the examination of the reasons why cultural organizations should use ICTs and social media in the cultural communication process with their public. The aim of this chapter to present the most popular communicative media that can be used by cultural managers to promote their cultural activities and organizations. The implementation of innovative and emerging technologies by cultural organizations has a role to play at their management and communication level of the organization. The chapter describes new applications employed by cultural managers that take into account the needs, preferences and expectations of the public and encourage them to interact with the cultural content and context. This may allow a more holistic approach to take place in regard to the trends and future outlook on the adoption of ICTs.

The adoption and implementation of innovative communication technologies may bring the visitor closer and connect him/her with the cultural organization. Curators and professionals who are in charge of media communication strategies in

cultural organizations may also find this study useful, because it illustrates innovative ways to incorporate web technologies in the promotion of the organizations and in the identification of visitors' needs and preferences.

However, the cultural messages shall use the appropriate medium and communication models to be transmitted via contemporary communication models. The use of innovative communication tools that are employed by cultural organizations for their communication strategies has been little researched as a topic (Padilla-Melendez & del Aguilla-Obra, 2013) where geographical and temporal boundaries are minimized (Bilton, 2015; Jeffres, 2015) and the co-creation of information via social media allows people to create content and post information online to like minded people.

BACKGROUND OF RESEARCH: CULTURAL ORGANIZATIONS, CULTURAL MARKETING, AND ICTS

This chapter presents (a) both the implementation of innovative communication technologies with the use of ICTs via computers, mobile phones, telephone lines to access and manage electronic data on the web from the cultural organizations' point of view in order to illustrate the possibilities on offer for managing and promoting their services and (b) the implementation of social media that are internet-based innovative communication tools that bring people together through the use of user-generated cultural content. In doing so, a two way communication and engagement between cultural organizations, art and audiences can be promoted and cultural managers may more fully adopt both ICTs and social media.

There is a tendency of cultural organizations to overall ignore competition as part of the marketing approach; when they do so, they attend those cultural organizations that they are in their immediate whereabouts (Bennett, 2015). Cultural organizations need to embrace new strategies of marketing and use marketing tools and techniques,

as well as of social media, to disseminate their content and promote their activities to reach and communicate with new audiences (Fletcher & Lee, 2012), to increase their visitors and build stronger relationships and loyalty.

Cultural organizations are interested in developing relationships with different categories of the wide public. Firstly, simple promotional techniques were enough to achieve their goals. Nowadays, due to the economic crisis they needed to adopt a more sophisticated approach to marketing (Kolb, 2013). Furthermore, the explosion of ICTs and the extensive use of social media by cultural organizations suggested a 'transformation in cultural communication' (Russo, 2011) and an adaptation to users' changing usage behavior and expectations related to technological development and cultural changes (Shyam Sundar, Go, Hyang-Sook, & Zhang, 2015). Furthermore, they provided opportunities for user-generated content and co-creation (Holdgaard, & Klasturp, 2014; Sylaiou, Basiouka, Patias, & Stylianidis, 2013), education, enjoyment and entertainment (Charitonos, 2011; Hou, Wu, Lin, Sung, Lin, & Chang, 2014; Katz & Halpern, 2015; Kellner & Kim, 2010; Russo, Watkins, & Groundwater-Smith, 2009; Selwyn, 2009; Vavoula, Sharples, Rudman, Meek, & Lonsdale, 2009), and social utility for online communities (Allen-Greil & MacArthur, 2010).

In a continuously changing environment social media are a major challenge for cultural organizations, since they change the ways cultural meaning is perceived and the cultural products are consumed. They have become a part of everyday life and it is expected that cultural organizations adopt them to communicate with the wide public. Their use and growth have led to significant changes in cultural communication and the marketing strategies need to get updated and become adjusted to the new conditions. Social media can make cultural organizations' aims and scope attractive to a broader audience and include not traditional audiences and provoke their visit not only to their digital extension, but also to the physical site. Furthermore, they permit the interaction with the

wide public and they may collect valuable information about their preferences and opinions. They create new and novel ways for communicating, understanding and experiencing culture. Social media promote innovation, the creation of online communities and participatory culture.

The economic crisis and the budget cuts and the reduction of the state grants changed dramatically the cultural landscape and have led cultural organizations to a shrinkage in their activities, or even to closure. Technological advances have emerged as areas of crucial interest, since they are making it possible to use sophisticated tools to provide solutions to problems created by economic crisis. ICTs may have a multi-faceted contribution to the relationship between culture and the public and can play a decisive role in making cultural content and context available, in enhancing experience through meaningful and constructive interaction between culture and the public.

HOW ARE ICTs AND SOCIAL MEDIA EMPLOYED WITHIN THE CULTURAL COMMUNICATION PROCESS

ICTs and social media are slowly implemented by cultural organizations in order to connect with their public and engage with the users informing them about the activities organized by cultural professionals and interacting with them in a reciprocal relation that allows the co-creation and exchange of experiences. The next sections present innovative ways with which ICTs and social media are implemented by cultural organizations and thus, how they can be further employed by cultural organizations.

Innovative and Emerging Technologies and Challenges of Cultural Organizations

There is an ongoing research that has taken place in regard to the potential of ICTs implementation

to cultural organizations (Gonzalez, 2015; Kéfi, & Pallud, 2011; Royston & Delafond, 2015; Scoffield & Liu, 2014; Sylaiou, Economou, Karoulis, & White, 2008; Sylaiou, Liarokapis, Kotsakis, & Patias, 2009). ICTs have made the content and context of museum collections more accessible and attractive to the wide public and have enriched the museum experience.

The implementation of innovative technologies by cultural organizations allows for the provision of customers with improved services since visitors may have a holistic approach in regard to the exhibits of the organization and be able to interact with them. In addition, the public can be informed at any time about the events and procedures of the organization with the latest technologies and the way cultural heritage is approached and managed. Furthermore, emerging technologies allow for the collection of data as far as cultural preservation and cultural management is concerned of rare collections and archives that may become available to the public with the adoption of new technologies. The following section presents the different purposes of use of emerging technologies and ICTs.

According to Bakhshi & Throsby (2012) have identified four categories of ICT innovation in cultural organizations:

- Innovation in audience reach in terms of engagement with audiences.
- Innovation in artform development in terms of encouragement of new and experimental work and incorporation in their programming.
- Innovation in value creation in terms of new ways to measure the economic and cultural value and to translate these into terms that policymakers, funding agencies, donors and private investors can relate to.
- Innovation in business management and governance in terms of organizations' business models review and of a search for innovative financing strategies.

Web 2.0, the second generation of the Web, or the ‘participative Web’ has revolutionized the way of interaction and participation of the users. Web 2.0 technologies support and advance the information and knowledge dissemination in a social way. More specifically, crowdsourcing technologies make use of the “wisdom of the crowd” and aim at integrating many different audiences, raising public awareness, encourage active participation in the preservation and presentation of cultural heritage, which leads to the creation of a participatory culture and the democratization of knowledge (Sylaiou, Basiouka, Patias, & Stylianidis, 2013).

Interaction of multiple channels from the social media networks can be combined complementarily with real life events that the cultural organization manages that aim to engage with the user in an interactive and expressive way and not with the traditional one way of communication (Lazzeretti, Sartori, & Innocenti, 2015). Social media’s proliferation and extensive use have transformed the cultural experience and provide many opportunities for learning, active participation and creativity. However, the main contribution of ICT is to involve communities and promote public dialogue and collaborative frameworks, create the sense of community and provide more transparency. They have sought to become the vehicle for fulfilling the public’s expectations and change in the relationship between cultural organizations and the public.

ICTs may have a multi-faceted contribution to the relation between cultural organizations and the wide public. More specifically, Internet technologies have the tremendous potential of offering virtual visitors ubiquitous access via the WWW. The Internet combined with:

- Virtual Reality (VR), a simulation of a real or imaginary environment generated in 3D by digital technologies that is experienced visually and provides the illusion of reality.
- Augmented Reality (AR) is considered as an extension of VR. AR tools present a view of the real-world environment whose

elements are augmented in real time by computer-generated sensory input, such as video, graphics, GPS data or sound (Promoter, 2015).

- Augmented reality and holograms systems that ‘augment’ reality with holographic displays (Mavridis & Hanson, 2009, Kim & Wohn, 2014).

Geographic Information Systems (GIS) that store, manage, analyze, visualize and present all types of spatial and geographical data have already transformed the cultural experience through their ability to document, organize, visualize, preserve and disseminate a plethora of types and vast amounts of cultural data. As a result, ICT have been extensively used by cultural organizations and have evolved into a discrete hybrid disciplinary field with its own especially dedicated conferences, such as the Computer Applications and Quantitative Methods in Archaeology (CAA), the Virtual Systems and Multimedia Society (VSMM) and the Virtual Reality, Archaeology and Cultural Heritage (VAST), Digital Strategies for Heritage (DISH) conference, ‘Art and Interaction’ as integral part of the Special Interest Group on Computer–Human Interaction (SIGCHI) Creativity and Cognition series, Digital Arts in Association for Computing Machinery’s Special Interest Group on Computer Graphics and Interactive Techniques (ACM SIGGraph), the Conference on Transdisciplinary Imaging at the Intersections of Art, Science and Culture and many more.

Nowadays, emerging ICTs help cultural organizations to face challenges concerning connections between heterogeneous data from different sources and enhance visitor experience. Some of these emerging technologies are mentioned below:

- Linked Open Data (LOD) linking content among cultural heritage collections and between those collections and collections created for other purposes by Libraries, Archives, and Museums (LAM) (Oard, Levi, Punzalan, & Warren, 2014).

- Internet of Things (IOT), a network environment of things, or everyday physical objects with unique identifiers will be connected to the Internet, able to identify each other as devices and interact in an intelligent way with each other. Its application to arts and culture (Senior, Moreton, & Dovey, 2015) and its interaction with wearable and mobile devices, contents retrieval from the Cloud and services in culture provision to internal and external users (Mighali et al., 2015) are still under exploration. Emerging Radio Frequency Identification Detection (RFID) technologies will connect things between them, objects, places, people, will apply in cultural spaces to help users to get the correct information which is better than traditional guides (Karimi, Nanopoulos, & Schmidt-Thieme, 2012; Sen, Roy, & Sarkar, 2014).
- Mobile apps that can be used for:
 - Learning in the framework of formal (Maher 2015) and informal learning (Katz & Halpern, 2015).
 - Cultural experiences enhanced with Augmented Reality (Damala, Marchal, & Houlier, 2007; De Angeli & O' Neil, 2015).
 - Digital storytelling (Negrini & Di Blas, 2015; Hart & Brownbill, 2015).
 - User-generated content and co-creation (Mason, 2012).

The next section presents the implementation of social media from cultural organizations in their communication activities.

The Implementation of Social Media From Cultural Organizations

This section presents the most important social media that can be employed by cultural organizations in their communication activities. Various online applications facilitate the “sharing of information, experiences, and perspectives throughout

community-oriented websites” (Weinberg, 2009, p. 1). Social media consist of “a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of User Generated Content” (Kaplan & Heinlein, 2010, p. 61) that help cultural organizations to reach new audiences, increase participants’ engagement and the likelihood of a visit (Theocharidis, Nerantzaki, Vrana, & Paschaloudis, 2014).

Among the most commonly used social networks user-driven channels that are employed by cultural organizations, such as museums, are Facebook, Twitter, Blogs, TripAdvisor, Flickr and YouTube.

- **Facebook:** Facebook can be used as a Public Relations’ tool for promotional and communicational purposes (Kelly, 2014; Kotzaivazoglou, Paschaloudis, Sylaiou, Nerantzaki, & Mantzirtzi, 2016; Proctor, 2010; Russo, 2011, 2012; Trembach, & Deng, 2015). It also may be used to promote a cultural organization’s events and announce important but it may also receive participants’ feedback to get their comments regarding their experience allowing the cultural organization to improve its services based on visitors’ comments made. The real case is that cultural organizations do not usually see Facebook as an interactive environment limiting its potential (Lazzeretti et al., 2015).

The social networking site of Facebook can be implemented to create an interactive communication with the target group of the cultural organizations.

- **Twitter:** Micro-blog is a type of blog that lets users publish short text updates while the act of using these services is called micro-blogging. For example, Twitter is a free micro-blogging service where users publish short messages which contain 140

characters or less, the so called tweets on the Twitter website or they may use their mobile phones to send sms text-messages in real time to their connected environment of this network based on the retweets that will take place and based on how big the social network is (Bastos, Galdini Raimundo, & Travitzki, 2013). In that way, information sharing may be accomplished online and many people even if they have never met may participate.

- **Blogs:** Blogs are websites that contain comments, forums for conversations, news and quite often hyperlinks provided by the cultural organizations. They create a friendly space for discussions, they connect people with organizations; blogs provide a valuable engagement tool for interacting the public and forming communities. They can also be considered as educational content platforms. Blogs offer cultural organizations a valuable way to reach a wider audience through search engine optimization and through the blog's inclusion in Google search (Bomboy & Sherman, 2014).
- **TripAdvisor:** TripAdvisor allows for the collection of user generated content in regard to a cultural organization located in a specific area. Available reviews are there for the potential visitor to see and comment on. The possibility of geographically collecting information about the cultural organizations of the region and user generated content under reputation management facilities (Ntalianis, Kavoura, Tomaras, & Drivas, 2015), allow the cultural manager to have feedback for his/her organization. Information that is created by other users about a cultural organization and provides good material and recommendations about it while at the same time, it is free advertisement for the organization.
- **Flickr:** Flickr is a photo sharing site. As such, it is very much associated with photographs taken when on holidays where

cultural tourism and visiting of cultural organizations may be the case. In fact, it has been found that visitors pay attention to Flickr advertisements but also employ it as a communication tool to upload the photos taken from their trips (Kavoura & Stavrianeas, 2015).

User generated content is created on Flickr and Dotan & Zaphiris (2010) found the existence of patterns on language and tagging in regard to specific countries and uploads on Flickr. Lazzeretti et al. (2015) on the other hand, found that the use of Flickr from cultural organizations had little success as a two way communication since there was a lack of interest by users. The use of a specific language in this communication's technological tool may attract visitors and users of cultural organizations that may gather on Flickr based on specific tags that are of interest to them and thus, make it more attractive.

- **YouTube:** YouTube, another technological instrument can be used to broadcast interviews with curators as suggested by Lazzeretti et al. (2015).

Significant exhibits in the room of a museum or the organization of a cultural festival at a well equipped venue, for example, can be briefly presented on You Tube employing narration, sound and image and thus, creating a rich media result.

RECOMMENDATIONS AND FUTURE RESEARCH DIRECTIONS

The implementation of innovative communication tools by cultural managers allows for engagement to take place between the cultural organization and the users. The use of technology for management and communication purposes by cultural organizations allows for a two way communication with their audiences that may bring the visitor closer and connect him/her with the cultural organization.

Cultural managers may use the above mentioned innovative communication technologies presented in this chapter in order to manage and communicate their organizations and services. There is a long way before all cultural organizations adopt such innovative platforms and tools, thus, enhancing their communication strategies.

Cultural organizations need to open their virtual doors to the world and use ICTs in a more two way oriented approach than the way they do it nowadays. New technologies and social media need to be further adopted in the cultural communication process for creating meaningful experiences. In order for this to be better employed, cultural managers need to realize the value of marketing. In fact, Wymer, Boenigk, & Mohlmann (2015) have created a construct for nonprofit organizations as a guide for nonprofit marketing orientation that incorporates services, supporters, the creation of a brand of the cultural organization and commercial aspects for it. It would be thus, interesting to examine from a holistic communication and management perspective the services on offer by the cultural organization, its supporters and the identification of their needs and their characteristics, the elements that can create its brand that will make it well known to the public and commercial aspects. Social media and information technologies have a role to play at the management and communication level of the organization.

Another future research opportunity can be associated with human resources, cultural organizations and the implementation of information and communication technologies. Are cultural organizations willing to correspond with users on the online level? Little research has illustrated so far that on the one hand that there is low involvement from the user in the cultural organizations' accounts of social media, while cultural organizations do not correspond with their users, thus creating a two way communication with them is in need to be further examined (Lazzeretti et al., 2015). In that way, a vicious circle is created. Which is the case when employees of the cultural

organization are devoted to manage this online information directed to and from the cultural organization and interact with the users?

Mobile technologies allow for the interactive participation between the real and virtual worlds and this is where future research and cultural organizations should focus. They should be further researched in regard to their use as communication and promotional tools by cultural organizations but also in regard to their geographic implementation (Lee, Chung, & Jung, 2015; Thompson, Purcell, & Rainie, 2013) since research mainly takes place for the USA while results should be offered for other continents and countries.

CONCLUSION

ICTs that have brought forth the emergence of social media should adopt individually and within a specific framework social networks employing the best elements each one has on offer. Social media is growing rapidly and provide great opportunities for cultural organizations to enhance and reinforce the experiences they offer. However, the simple use of social media is not enough (Effing & Spil, 2016). They should organize and apply social media strategically to benefit (DiStaso & McCorkindale, 2013).

According to social media strategy' definition, it is "a goal-directed planning process for creating user generated content, driven by a group of Internet applications, to create a unique and valuable competitive position" (Effing & Spil, 2016, pp. 2). These new applications are taking into account the needs, preferences and expectations of various categories of public and encourage them to interact with the cultural content and context, as well as with other people, to share their experiences and memories, giving to community the sense of 'ownership' over culture. Fostering innovations with the implementation of the above mentioned technologies described in the chapter may allow for effective and efficient cultural management and cultural communication. Cultural managers

but also users may benefit from the implementation of emerging technologies towards a closer interaction with each other.

REFERENCES

- Allen-Greil, D., & MacArthur, M. (2010). Small Towns and Big Cities: How Museums Foster Community On-line. In J. Trant & D. Bearman (Eds.), *Museums and the Web 2010: Proceedings*. Toronto: Archives & Museum Informatics. Retrieved October 20, 2015, from <http://www.archimuse.com/mw2010/papers/allen-greil/allen-greil.html>
- Bakhshi, H., & Throsby, D. (2012). New technologies in cultural institutions: Theory, evidence and policy implications. *International Journal of Cultural Policy*, 18(2), 205–222. doi:10.1080/10286632.2011.587878
- Basballe, D. A., & Halskov, K. (2010). Projections on museum exhibits: engaging visitors in the museum setting. *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction* (pp. 80-87). ACM. doi:10.1145/1952222.1952240
- Bastos, M. T., Raimundo, R., & Travitzki, R. (2013). Gatekeeping Twitter: Message Diffusion in Political Hashtags. *Media Culture & Society*, 35(2), 260–270. doi:10.1177/0163443712467594
- Bennett, R. (2015). Competitive Myopia in a Not-For-Profit Service Industry: The Case of the Performing Arts Sector in the UK. In H. Spotts (Ed.), *Assessing the Different Roles of Marketing Theory and Practice in the Jaws of Economic Uncertainty* (p. 29). Vienna: Springer. doi:10.1007/978-3-319-11845-1_8
- Bilton, C. (2015). Management in the cultural industries. In K. Oakley & J. O'Connor (Eds.), *The Routledge Companion to the Cultural Industries*. Oxon, NY: Routledge.
- Bomboy, S., & Sherman, R. (2014). Success Strategies for Engaging Audiences with Museum Website Blogs. *MW2014: Museums and the Web 2014*. Retrieved November 1, 2015, from <http://mw2014.museumsandtheweb.com/paper/success-strategies-for-engaging-audiences-with-museum-website-blogs/>
- Charitonos, K. (2011). *Museum Learning via Social Media: (How) Can Interactions on Twitter Enhance the Museum Learning Experience?* In Learning, Media and Technology Doctoral Conference, London, UK.
- Chen, W. (2015). A Moveable Feast: Do Mobile Media Technologies Mobilize or Normalize Cultural Participation? *Human Communication Research*, 41(1), 82–101. doi:10.1111/hcre.12041
- Damala, A., Marchal, I., & Houlier, P. (2007). Merging Augmented Reality Based Features in Mobile Multimedia Museum Guides. *Anticipating the Future of the Cultural Past, CIPA Conference 2007*. Retrieved October 1 2015, from <https://halshs.archives-ouvertes.fr/halshs-00530903>
- De Angeli, D., & O'Neill, E. (2015). A smartphone headset for augmented-reality experiences in museums. *MW2015: Museums and the Web 2015*. Retrieved November 15, 2015, from <http://mw2015.museumsandtheweb.com/paper/a-smartphone-headset-for-augmented-reality-experiences-in-museums/>
- DiStaso, M. W., & McCorkindale, T. (2013). A benchmark analysis of the strategic use of social media for Fortune's most admired U.S. companies on Facebook, Twitter and YouTube. *The Public Relations Journal*, 7(1), 1–33.
- Dotan, A., & Zaphiris, P. (2010). A cross-cultural analysis of Flickr users from Peru, Israel, Iran, Taiwan and the UK. *International Journal of Web Based Communities*, 6(3), 284–302. doi:10.1504/IJWBC.2010.033753

- Effing, R., & Spil, A. A. M. (2016). The Social Strategy Cone: Towards a Framework for Evaluating Social Media Strategies. *International Journal of Information Management*, 36(1), 1–8. doi:10.1016/j.ijinfomgt.2015.07.009
- Fletcher, A., & Lee, M. J. (2012). Current social media uses and evaluations in American museums. *Museum Management and Curatorship*, 27(5), 505–521. doi:10.1080/09647775.2012.738136
- Gonzalez, D. (2015). *Museum making: Creating with emerging technologies in art museums*. MW2015: Museums and the Web 2015. Retrieved October 20, 2015, from <http://mw2015.museum-sandtheweb.com/paper/museum-making-creating-with-emerging-technologies-in-art-museums/>
- Hart, T., & Browbill, J. (2015). *Storyteller-World War One: Love and Sorrow: A hybrid exhibition mobile experiences*. MW 2015: Museums and the Web 2015. Retrieved November 15, 2015, from <http://mw2015.museumsandtheweb.com/paper/storyteller-world-war-one-love-and-sorrow-a-hybrid-exhibition-mobile-experience/>
- Holdgaard, N., & Klasttrup, L. (2014). Between control and creativity: Challenging co-creation and social media use in a museum context. *Digital Creativity*, 25(3), 190–202. doi:10.1080/14626268.2014.904364
- Hou, H.-T., Wu, S.-Y., Lin, P.-C., Sung, Y.-T., Lin, J.-W., & Chang, K.-E. (2014). A Blended Mobile Learning Environment for Museum Learning. *Journal of Educational Technology & Society*, 17(2), 207–218.
- Jeffres, L. (2015). Mass Communication Theories in a Time of Changing Technologies. *Mass Communication & Society*, 18(5), 523–530. doi:10.1080/15205436.2015.1065652
- Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of social media. *Business Horizons*, 53(1), 59–68. doi:10.1016/j.bushor.2009.09.003
- Karimi, R., Nanopoulos, A., & Schmidt-Thieme, L. (2012). RFID-Enhanced Museum for Interactive Experience. *Communications in Computer and Information Science*, 247, 192–205. doi:10.1007/978-3-642-27978-2_17
- Katz, J. E., & Halpern, D. (2015). Can Virtual Museums Motivate Students? Toward a Constructivist Learning Approach. *Journal of Science Education and Technology*, 1–13.
- Kavoura, A., & Stavrianeas, K. (2015). The Importance of Social Media on Holiday Visitors' Choices - the case of Athens, Greece. *EUROMED. The Journal of Business*, 10(3), 360–374.
- Kéfi, H., & Pallud, J. (2011). The role of technologies in cultural mediation in museums: An Actor-Network Theory view applied in France. *Museum Management and Curatorship*, 26(3), 273–289. doi:10.1080/09647775.2011.585803
- Kellner, D., & Kim, G. (2010). YouTube, Critical Pedagogy, and Media Activism. *Review of Education, Pedagogy & Cultural Studies*, 32(1), 3–36. doi:10.1080/10714410903482658
- Kelly, L. (2014). The connected museum in the world of social media. In K. Drotner & K. C. Schroder (Eds.), *Museum communication and social media: The connected museum* (pp. 54–71). London: Routledge.
- Kim, M., & Wohn, K. (2014). Spatial AR Hologram: a new exhibition system combining augmented reality technology with conventional holographic display. In N. Proctor & R. Cherry (Eds.), *Museums and the Web Asia 2014*. Silver Spring, MD: Museums and the Web. Retrieved October 15, 2015, from <http://mwa2014.museumsandtheweb.com/paper/spatial-ar-hologram-a-new-exhibition-system-combining-augmented-reality-technology-with-conventional-holographic-display/>
- Kolb, B. M. (2013). *Marketing for Cultural Organizations New Strategies for Attracting and Engaging Audiences* (3rd ed.). London: Routledge.

- Kotzaivazoglou, I., Paschaloudis, D., Sylaiou, S., Nerantzaki, D.-M., & Mantzirtzi, E. (2016in press). How nonprofit organizations in Greece are using Facebook as a public relations tool. In A. Kavoura, D. Sakas, & P. Tomaras (Eds.), *Springer Proceedings in Business and Economics*. Springer.
- Lazzeretti, L., Sartori, A., & Innocenti, N. (2015). Museums and social media: The case of the Museum of National History of Florence. *International Review on Public and Nonprofit Marketing*, 12(3), 267–283. doi:10.1007/s12208-015-0136-5
- Lee, H., Chung, N., & Jung, T. (2015). Examining cultural differences in acceptance of mobile augmented reality: comparison of South Korea and Ireland. In I. Tussyadiah & A. Inversini (Eds.), *Information and Communication Technologies in Tourism* (pp. 477–491). Springer. doi:10.1007/978-3-319-14343-9_35
- Maher, D. (2015). Connecting Classroom and Museum Learning with Mobile Devices. *Journal of Museum Education*, 40(3), 257–267. doi:10.1179/1059865015Z.000000000102
- Mason, M. (2012). The Dimensions of the Mobile Visitor Experience: Thinking beyond the Technology Design. *International Journal of the Inclusive Museum*, 5(3), 51–72. doi:10.18848/1835-2014/CGP/v05i03/44404
- Mavridis, N., & Hanson, D. (2009). The IbnSina Center: An augmented reality theater with intelligent robotic and virtual characters. *The 18th IEEE International Symposium on Robot and Human Interactive Communication*, (pp. 681–686). RO-MAN 2009.
- Mighali, M., Del Fiore, G., Patrono, L., Mainetti, L., Alletto, S., Serra, G., & Cucchiara, R. (2015). Innovative IoT-aware Services for a Smart Museum. *Proceedings of the 24th International Conference on World Wide Web* (pp. 547–550). doi:10.1145/2740908.2744711
- Negrini, M., & Di Blas, N. (2015). *Digital Storytelling for Cultural Heritage: A Modular, Multi-channel, Multi-scenario Approach, Built Heritage, Monitoring Conservation Management*. Part of the series. *Research for Development*, 367–375. doi:10.1007/978-3-319-08533-3_31
- Ntalianis, K., Kavoura, A., Tomaras, P., & Drigas, A. (2015). Non-Gatekeeping on Social Networks: A Reputation Monitoring Approach. *Journal of Tourism and Services*, 6(10), 19–44.
- Oard, D. W., Levi, A., Punzalan, R., & Warren, R. (2014). Bridging Communities of Practice: Emerging Technologies for Content-Centered Linking. *MW2014: Museums and the Web 2014*. Retrieved October 20, 2015, from <http://mw2014.museumsandtheweb.com/paper/bridging-communities-of-practice-emerging-technologies-for-content-centered-linking/>
- Padilla-Melendez, A., & del Aguilla-Obra, A. R. (2013). Web and social media usage by museums: Online value creation. *International Journal of Information Management*, 33(5), 892–898. doi:10.1016/j.ijinfomgt.2013.07.004
- Proctor, N. (2010). Museum as platform, curator as champion, in the age of social media. *Curator*, 53(1), 35–43. doi:10.1111/j.2151-6952.2009.00006.x
- Promoter. (2015). *Augmented reality: enriching culture. An overview about Augmented Reality experimentations*. Retrieved November 15, 2015, from <http://www.digitalmeetsculture.net/article/augmented-reality-enriching-culture/>
- Royston, C., & Delafond, S. (2015). How To Introduce Digital Transformation To A Museum. In N. Proctor & R. Cherry (Eds.), *Museums and the Web*. Retrieved October 20, 2015 from <http://mw2014.museumsandtheweb.com/paper/how-to-introduce-digital-transformation-to-a-museum/>

- Russo, A. (2011). Transformations in Cultural Communication: Social Media, Cultural Exchange, and Creative Connections. *Curator: The Museum Journal*, 54(3), 327–346. doi:10.1111/j.2151-6952.2011.00095.x
- Russo, A. (2012). The rise of the ‘media museum. In E. Giaccardi (Ed.), *Heritage and social media: Understanding heritage in a participatory culture* (pp. 145–157). London: Routledge.
- Russo, A., Watkins, J., & Groundwater-Smith, S. (2009). The impact of social media on informal learning in museums. *Educational Media International*, 46(2), 153–166. doi:10.1080/09523980902933532
- Scofield, S., & Liu, J. (2014). Online Marketing Communications and the Postmodern Consumer in the Museum context. *Cambridge Conference on Business & Economics*. Murray Edwards College, Cambridge University. Retrieved 20 October 2015, from http://www.gcbe.us/2014_cbec/data/Susan%20Scofield,%20Jie%20Liu.pdf
- Selwyn, N. (2009). Faceworking: Exploring students education-related use of Facebook. *Learning, Media and Technology*, 34(2), 157–174. doi:10.1080/17439880902923622
- Sen, S., Roy, S., & Sarkar, S. K. (2014). A Proposal for Enhancing Museum Visiting Experience Implementing Active RFID Technology. In *Fourth International Conference on Advances in Computing and Communications (ICACC)* (pp. 295–298). doi:10.1109/ICACC.2014.76
- Senior, T. J., Moreton, S., & Dovey, J. (2015). *The Arts and Humanities in the Internet of Things*. Retrieved October 10, 2015, from <http://react.hoffi.com/publications/publications/arts-and-humanities-internet-things>
- Shyam Sundar, S., Go, E., Hyang-Sook, K., & Zhang, B. (2015). Communicating Art, Virtually! Psychological Effects of Technological Affordances in a Virtual Museum. *International Journal of Human-Computer Interaction*, 31(6), 385–401. doi:10.1080/10447318.2015.1033912
- Sylaiou, S., Basiouka, S., Patias, P., & Stylianidis, E. (2013). *The Volunteered Geographic Information in Archaeology*. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XXIV International CIPA Symposium, 2–6 September 2013, Strasbourg, France. Retrieved November 1, 2015, from <http://www.isprs-ann-photogramm-remote-sens-spatial-inf-sci.net/II-5-W1/301/2013/isprsannals-II-5-W1-301-2013.pdf>
- Sylaiou, S., Economou, M., Karoulis, A., & White, M. (2008). *The evaluation of ARCO: a lesson in curatorial competence and intuition with new technology*. New York: ACM Press.
- Sylaiou, S., Liarokapis, F., Kotsakis, K., & Patias, P. (2009). Virtual museums, a survey and some issues for consideration. *Journal of Cultural Heritage*, 10(4), 520–528. doi:10.1016/j.culher.2009.03.003
- Theocharidis, A. I., Nerantzaki, D. M., Vrana, V., & Paschaloudis, D. (2014). Use of the web and social media by Greek museums. *International Journal of Cultural and Digital Tourism*, 1(2), 8–22.
- Thomson, K., Purcell, K., & Rainie, L. (2013). *Arts organizations and digital technologies*. Washington, DC: Pew Research Center’s Internet & American Life Project.
- Trembach, S. A., & Deng, L. (2015). Social media and participatory culture: Opportunities and challenges for reforming the contemporary museum. In A. K. Hastings (Ed.), *Annual Review of Cultural Heritage Informatics* (pp. 99–116). The University of South Carolina’s School of Library and Information Science.
- Tylor, E. B. (1871). *Primitive culture: researches into the development of mythology, philosophy, religion, language, art, and custom*. Retrieved October 20, 2015, from <https://archive.org/details/primitiveculture01tylouroft>

UNESCO. (2001). *Mexico City Declaration on Cultural Policies: UNESCO Universal Declaration on cultural diversity*. Paris: UNESCO.

Vavoula, G., Sharples, M., Rudman, P., Meek, J., & Lonsdale, P. (2009). My Art Space: Design and evaluation of support for learning with multimedia phones between classrooms and museums. *Computers & Education*, 53(2), 286–299. doi:10.1016/j.compedu.2009.02.007

Weinberg, T. (2009). *The New Community Rules: Marketing on the Social Web*. O'Reilly Media.

Wymer, W., Boenigk, S., & Mohlmann, M. (2015). The Conceptualization of Nonprofit Marketing Orientation: A Critical Reflection and Contributions Toward Closing the Practice-Theory Gap. *Journal of Nonprofit & Public Sector Marketing*, 27(2), 117–134. doi:10.1080/10495142.2014.965078

ADDITIONAL READING

Agapito, D., & Lacerda, A. (2014). Marketing and Brand Design of Destination Experiences: The Role of ICT. *Journal of Spatial and Organizational Dynamics*, 2(3), 201–216.

Katz, J. E., & Halpern, D. (2015). Can Virtual Museums Motivate Students? Toward a Constructivist Learning Approach. *Journal of Science Education and Technology*, 24(6), 776–788. doi:10.1007/s10956-015-9563-7

Moro Sundjaja, A. (2015). An Adoption of Social Media for Marketing and Education Tools at Museum Industry. *Advanced Science Letters*, 21(4), 1028–1033. doi:10.1166/asl.2015.5975

Pulh, M., & Mencarelli, R. (2015). Web 2.0: Is the Museum-Visitor Relationship Being Redefined? *International Journal of Arts Management*, 18(1), 43–51.

KEY TERMS AND DEFINITIONS

Cultural Communication: The application/adaptation of communication theories, mechanisms, methods and tools in cultural organizations.

Cultural Organization: A not for profit network that aims to promote culture and the arts.

Culture: The capacity of human beings to classify, codify, and communicate their experiences symbolically so that they can be available to future generations.

Mobile Media Technologies: Interactive use of computing devices such as smartphones or tablets that allow the user to communicate, interact, share and download material with the use of applications.

Social Media: Applications that engage people in sharing user-generated content employing online communication.

From the Psychoanalyst's Couch to Social Networks

Annamaria Silvana de Rosa

Sapienza University of Rome, Italy

Emanuele Fino

Psychologist, Psychometrician, Italy

Elena Bocci

Sapienza University of Rome, Italy

INTRODUCTION

The rise of information technologies and the Internet have dramatically changed the ways in which people interact, communicate and represent themselves and the world that they live in (Edwards et Al., 2013). In particular, the recent appearance of social network sites (SNS) is challenging mainstream social psychological research and it requires social psychologists to question their earmarking theoretical tools.

To date, social psychological research has mainly focused on the application of existing theories to explain the use and impact of SNS (de Rosa, 2012). Nevertheless, the convergence of social evolution and information technologies is laying a new basis for the performance of activities throughout the social structure (Castells, 1996:470-471), so that social psychologists must revise their theoretical traditions in its investigation.

Although research in the past decade has focused on the study of new phenomena generated by online social interactions, the underlying symbolic universe of meanings, products and contents of such interactions is still partially unknown.

Exploration of this semantic production can draw fruitfully on Social Representations Theory (SRT) (Moscovici, 1961/1976; 1995; Farr & Moscovici, 1984). In his classic study on the image of psychoanalysis Moscovici investigated

how different forms of knowledge encounter and cross-fertilize. His specific interest was in how scientific theories circulate within common sense and what happens to those theories when they are elaborated upon the laboratory of society. Moscovici identified the product of this encounter in social representations (SR), defined as common sense theories on key aspects of the world that allow individuals and groups to represent it and master it (Farr & Moscovici, 1984).

In this paper, we present and discuss the results of research on spontaneous conversations in SNS in two cultural contexts, France and Italy, which followed up on Moscovici's classic study on psychoanalysis. This research is part of a broader research project (de Rosa, 2011; 2012; 2013) comprising: a field study, content analysis of the online press and content analysis of spontaneous conversations on SN.

BACKGROUND

It is important to recall here that the conventional approach of social sciences to the study of SN has been rooted in the field of sociometry since the introduction and diffusion in the late 1960s of such well-known theories as the *six degrees of separation* and the *small world phenomenon* (Travers & Milgram, 1969). These theories have influenced generations of social scientists intent

DOI: 10.4018/978-1-5225-2255-3.ch608

on extending mathematical models to the study of social relationships in terms of network theory and relying on concepts such as nodes (individual actors within the network) and ties (relationships between those actors). Social network analysis (SNA) is the modern sociological evolution of this trend (Carrington, Scott & Wasserman, 2005; Scott, 2004). Thanks to the recent affirmation and visibility of the Web, this approach has become hegemonic in the study of SN (Catanese et Al., 2012; De Meo et Al., 2012).

On the one hand, this trend can be seen as the consequence of the increasing interest of social sciences in computationally intense methods with which to analyse and model social phenomena (Williford & Henry, 2012) leading to the reproduction of “‘habitual practices’ employed by quantitative researchers using the procedures which they are comfortable and familiar with” (Stoneman, Sturgis & Allum, 2012:854). On the other hand, it is possible to identify a significant lack of theory in the study of “what lies beneath” the massive and multiform production of social interaction in new online communication channels, especially from a semantic perspective (de Rosa, 2012).

Research on SR investigated via interpersonal exchanges on SNS may be a valid response to this theoretical challenge. In fact, in the 1990s, Moscovici (1995; 1997:7) provided an anticipatory proposal; he emphasised the importance of investigating new communication phenomena by studying “how common sense, the language exchanged, groups themselves are shaped in this cyber-communication”.

We accordingly assumed that exploration of the representational fields underlying the ‘social discourse’ would provide track of the new type of common sense emerging from SN, as well as the social positioning of different actors and groups.

Although there is no single definition of SNS (Boyd and Ellison, 2007; Nadkarni & Hofmann, 2011) we highlight a growing interest in their role in the social construction of knowledge

(de Rosa, 2013; Mazzara, 2008). Briefly, recent decades have seen a structural transformation of the traditional communication channels whereby tele-communication is used to connect people in a new pattern of “connected-presence” (Licoppe & Smoreda, 2005; Castells, Fernandez-Ardevol, Qiu & Sey, 2007), providing interactive commitment of the actors involved, be they scientists, trainees, or simply laypersons in the sharing of knowledge. In fact, ‘Science 2.0’ (Jankowski, 2002), along the lines of the recent web revolution (O’Reilly, 2005) implies a “coupling between science and ICT”, providing new ways “for accessing scientific knowledge and [...] participation” (Ponte & Simon, 2011:150).

The development in the knowledge building through technological innovation enlarges the discourse to relevant issues already developed in the literature like: the digital divide (Castells, 1996/2000) the ethical implications (Drude et al., 2008) the effects of online psychotherapies (Andersson et al., 2009; Shingleton et al., 2013). Regarding these issues, through the SN, the experts have the opportunity to open the doors of their rooms and meet laypersons, establishing social relationships and, most of all, a “familiarity” effect, generating “trust” (Markovà, 2009).

In the SN, the traditional view of media as a ‘dirty mirror’ of scientific theories is completely reversed. Not only can the public elaboration and understanding of science no longer be compared to traditional top-down communication flows (Castells, 1996/2000; Moscovici, 1961/1976), but it also gives way to a new form of coexistence, “whose transgression launches a process of creation of profane knowledge from what was originally a concept elaborated in the midst of a small group of researchers. Thus uprooted from its original context, the concept is appropriated by the different media and thrown around, discussed, amplified, acclaimed, distorted, disproved, discredited, forgotten and rediscovered, circulating in the process among all conceivable factions of society, serving each and every one in turn” (Bangerter, 1995:8).

In this sense, the choice of psychoanalysis is strategic. As suggested by Jesuino (2008:393) the investigation “would be much more problematic if the study focused on a hard science such as physics, which is more distant to daily concerns of common citizens. Psychoanalysis had, and still has, a rather ambiguous status within the scientific community.” In fact, although in 1961 Moscovici described psychoanalysis as “a cultural event that was no longer confined to the limited circle of the sciences, affecting the whole of society” (Moscovici, 1976/2008:xxix), its scientific status and therapeutic validity are still subject to discussion worldwide (Mecacci, 2000; Meyer, 2005; Onfray, 2010; Rowland Smith, 2010).

SPONTANEOUS CONVERSATIONS

Objectives and Hypotheses

In accordance with our theoretical perspective we present and discuss results from the analysis of spontaneous conversations collected from three SNS: *Facebook*, *Twitter* and *Yahoo! Answers*, in the French and Italian contexts. The SNS were selected on the basis of their distinctive aggregative characteristics and communication constraints reflected in the members' modalities of affiliation and segmentation, as documented in the literature (Greitemeyer, & Kunz, 2013; Ryan & Xenos, 2011; Nadkarni & Hofmann, 2011).

Free conversations among members of SN have been analysed from an ecological perspective of non-intrusive research in natural contexts. It is our conviction that a psycho-social approach, oriented to the investigation of semantic aspects (and not only the formal structure of relations among members), aiming to capture and analyse significant conversational excerpts from conversational co-production among SN' users, allow to go over the present sociological supremacy (influenced by mathematical and computer-science-based contributes) to the study of the structure and content of social networking (vs. the paradigm of SNA).

In particular, we explored:

1. The SR of psychoanalysis, psychiatry and mental-health emerging as interrelated systems of representations from online conversations on SNS;
2. The positioning of social actors, groups and communities on the basis of the SR shared;
3. The communication systems emerging from the user-generated content analysed according to different constraints characterizing each of the SNS considered.

In light of the main assumption – that the dynamic relation between SR and public issues in the new ‘social arena’ of SNS is likely to determine the symbolic positioning of groups with similar or conflicting interests and visions of the world (de Rosa, 2013) – the following hypotheses may be summarized in the capacity to identify:

1. Evidence of the co-existence of expert and everyday knowledge, science and common sense, ‘reified universes’ and ‘consensual universes’ with their specific processes and functions, that positions different user groups and styles of communication in specific SNS. These spaces give voice to and negotiate competing definitions of roles, forms of knowledge, practices and/or personal experience related to the domain of mental-health, among actors characterized by different levels of access and proximity to the objects of representations investigated, namely lay users, experts and experts in training;
2. The occurrence of different SR of the target objects (psychoanalysis vs psychiatry) according to the experiential/informative character of the conversational exchanges on *the SNS* as a function of their communicative constraints;
3. Cross-national differences in the SNS users' SR depending on their anchoring in the two cultural contexts (France and Italy) due to the media echo effect of the recent controversy

on the scientific validity of psychoanalysis and to the different legal and institutional frameworks;

4. Different communicational styles defined by the context of *Social Media* and specific aggregative and communicative attributes which differ considerably in regard to their members' patterns of affiliation and segmentation, constraints on the length of messages and their semantic context and emotional expressiveness, like the use of the SN as a *conversational space* for sharing and negotiating personal emotional/psychological problems and providing social support (the members of the SN as therapists).

Method

Data Collection and Sample

This exploration required a systematic and robust methodology for the retrieval and analysis of the semantic structures underlying the online user-generated content.

Consequently, we developed a search engine-assisted strategy for the collection of spontaneous public conversations among SN users, through Google Advanced Search (de Rosa, 2012; 2013).

This approach enabled us to access dialogues, conversational exchanges and discussions on the objects of analysis spontaneously produced in SNS rather than being affected by the researcher's mediation.

Conversations were collected over a one-year period (1/3/2010-1/3/2011). A set of keywords ('psychoanalysis', 'psychiatry', 'psychoanalyst', 'psychiatrist', 'mental-health', 'mental-illness') semantically related to the object of analysis, was employed in the search. All the posts appeared in Google Search, on the basis of the following criteria, are included in the statistical analysis: keywords, temporal window, SN and language (Italian and French).

In order to collect data we applied different steps (de Rosa, 2012):

Step 1: Open the web page www.google.it (or www.google.fr) and access to Google advanced search functions using the link located on the right of the search bar.

Step 2: Define the web search parameters in the specific language of interest.

Specify keywords in Google's "all these words" search bar. Specify the search language in the "Language" bar.

Limit the search to the domains of interest; i.e. for Facebook posts, specify: www.facebook.com in the "Domain" bar

Finally, leave the values of the other parameters as default, turn off - if activated - the SafeSearch filter, and click on the "Search" button to display the results.

Step 3: Limit the search to the pre-specified time frame, by using the function in the left column on the page displaying the results. Note that in our case we opted for a search for each month within the time frame of interest (i.e. "1/1/2010 - 1/1/2011").

Step 4: Opening of all the links obtained (one by one).

Step 5: Save pages in.html format, renaming them by title, page name, group name, discussion topic.

Step 6: Repeat the procedure using keywords composed of two or more words, phrases. The only difference with the previously indicated procedure, resides in the Google search bar in which to place these words, in this case corresponding to the following: "this exact word or phrase", and proceed with the instructions already explained above. However so, it is important to specify that in this case we had to proceed by utilising one phrase at a time, given the inherent characteristics to the logarithm of the search engine.

Twitter required to implement the following specific procedure through the Advanced Search feature:

in the first place, the search parameters remain unchanged, proceeding analogously to Step 1 as previously presented, taking care, however, not to specify in this case the web domain of interest; once obtained the results, to search in the Twitter domain, we used Google's "real-time" function (such function went off-line on July 2, 2011 after the deal with Twitter expired - reference: <http://searchengineland.com/as-deal-with-twitter-expires-google-realtime-search-goes-offline-84175>), located in the left column of Google's results page; Google in fact, automatically returned all Tweets related to the previously specified key-words.

In order to specify the time range, we used the time diagram in the right column of the real time results page, displaying results by month.

By clicking on each individual month, the search engine returned online tweets posted in that specific month. At this point, it was possible to save the Google search page in.html format, such a page including all tweets of interest.

Currently, via Google Trends you can collect interesting descriptive statistics on a theme starting from the following information: keywords, country and time period.

Social media sites have gradually spread over time; we recall in this regard some among those that are free: Socialmention.com; Tagboard.com; Hashtagify.me; Keyhole.co.

Socialmention favours some particular search sources ("wordpress", "ask" etc.) and by inserting the # symbol in front of the keyword it also allows to identify tweets; It also allows to select the time frame and language, with the added benefit of being able to investigate "sentiment" (the ratio of positive mentions against those that are generally negative) for each comment. Tagboard allows research on classical SN like Facebook and Twitter, viewing sentiment globally on the keywords selected, but not comment by comment.

In research, however, it is important to remind that most of the popular social SNS let crawlers extract data only through Their Own API (Application Programming Interface) so as to control

the amount of information about users and their activities.

Data Analysis

We employed an inductive and non-probabilistic approach to automated text content analysis (Lahlou, 2011). Specifically, co-occurrence analysis and descending hierarchical classification -four classes from the French corpus and three classes from the Italian one- were applied to the text corpora by means of the Alceste algorithm (Reinert, 1986) implemented in the *Iramuteq* software. *In particular, we took in account* the most relevant cited "words" in the descending hierarchical classification, selected on the bases of the chi2 value.

Results

Regarding the French context, Class 1 was the biggest (34.1%), followed by Class 2 (24.5%), Class 4 (21.2%) and Class 3 (20.1%). With regard to the Italian context, Class 1 was the biggest (51.2%), followed by Class 2 (37.2%) and Class 3 (11.6%).

The French Class 1 contained conversations mainly collected from *Facebook Pages* ($\chi^2 1802,4$) and related to psychoanalysis ($\chi^2 155,9$), mental-health ($\chi^2 46,2$) and psychoanalyst ($\chi^2 28$). In this class, the general topic of mental-health was prevalent, as testified by the occurrence of terms such as *disorder* ($\chi^2 274,2$), *depression* ($\chi^2 262,4$), *symptoms* ($\chi^2 203,2$) and *anxiety* ($\chi^2 126,6$). As expected, the language and the style emerging from the analysis of the sentences highlighted the presence of users characterized by high levels of expertise in the field, particularly groups and organizations, rather than private individuals, displaying specialized lexicons and information, in agreement with hypothesis a) -focused on position of different social groups in the SN-.

Moreover, we found a specific effort to delimit the boundaries between psychoanalysis and psychiatry in the field of psychopathology, particularly depression. Similarly, the Italian

Class 2 included conversations from *Facebook Pages* ($\chi^2 2276,3$) and to a lesser extent from *Groups* ($\chi^2 27,6$). Psychoanalysis ($\chi^2 260,7$) and psychiatry ($\chi^2 289,9$) were the main objects of discussion. *Psychiatry* ($\chi^2 321,3$) also represented the most relevant cited word in the class, followed by the terms *social* ($\chi^2 221,7$), *psychiatric* ($\chi^2 209$), *psychology* ($\chi^2 199,9$) and *patient* ($\chi^2 197,5$). Differently from the French context and in agreement with hypothesis c) (focused on cross-national differences) the representation of psychiatry and the domain of mental-health lay on a communitarian dimension, whereas intervention by professionals (professional educator) was centred on a rehabilitation process targeted on interpersonal and family relationships.

Again, specification of the boundaries of the psychiatric domain was subject to discussion. Formal and instructional language was used, although it focused on describing the methods and tools characterizing the relevant profession rather than providing diagnostic information on specific mental diseases.

As regards the representations identified -in line with hypothesis b- there emerged a difference between the SR of psychoanalysis and psychiatry in their approaches to mental-illness. In regard to the former, mental-illness was described in terms of a “painful psychic process” (French Class 1). By contrast, the latter conceived mental-illness as a “disease” requiring suitable treatment relying on diagnostic techniques and pharmaceutical methods (Italian Class 2).

Both French Class 2 ($\chi^2 168$) and 3 ($\chi^2 4599$), as well as the Italian Class 1 ($\chi^2 3181,5$), included conversations retrieved from *Yahoo! Answers*. The French Class 2 was focused on mental-illness ($\chi^2 183,5$) and mental-health ($\chi^2 14,5$).

Class 3 instead referred to the psychiatrist ($\chi^2 355$) and psychiatry ($\chi^2 31$) and also in the Italian Class 1 focused on the psychiatrist ($\chi^2 995,7$). Interestingly, the latter was also composed of sentences from *Facebook Profiles*, although to a lesser extent ($\chi^2 6$). Contrary to what was previously observed, the semantic field

of these classes seemed to reproduce a naive view of mental-health, once again in agreement with hypothesis a). The lexicon utilised included words pertaining to everyday language and especially linked to the emotions, actors and actions of ordinary social life and aspects regarding socialization: e.g. *love* ($\chi^2 269,8$), *man* ($\chi^2 193,5$), *eat* ($\chi^2 121,7$) and *wife* ($\chi^2 118$) in the French context and *life* ($\chi^2 167,4$), *speak* ($\chi^2 164,8$), *child* ($\chi^2 158,2$) and *friend* ($\chi^2 104,1$) in the Italian Class 1. Accordingly, the topic of mental-health was approached from a non-professional, heuristic perspective, in which users engaged in dynamics of experiential support rather than pointing out specific guidelines.

In the same vein, in the Italian Class 1, we found an emphasis on the role of peers and the family in situations of psychological distress, as conveyed by words such as *support* ($\chi^2 132,7$), *friend* ($\chi^2 104,1$), *person* ($\chi^2 86,3$) and *boy-friend* ($\chi^2 86$). In this case, users seemed to act as ‘infomediaries’ of expert knowledge, providing informal help and suggestions on choice of a therapist.

It was also possible to distinguish a representation of the psychiatrist in terms of ‘he who prescribes’ therapies, but ‘not enough’ to accomplish an adequate healing process, as opposed to a more functional psychotherapeutic intervention.

This medicalized representation of psychiatry was also present in the French Class 3. The word ‘psy’ ($\chi^2 773,9$) was the most important, followed by *doctor* ($\chi^2 647,5$), *drug* ($\chi^2 318,5$) and *psychiatrist* ($\chi^2 300,3$). To be noted is that this French expression is an effective and succinct way to express a single profile including many and differentiated professional jobs within (psychologist, psychoanalyst, psychiatrist...).

Interestingly, in this class the elements ‘doctor-ill-hospital’ could be identified with the axes of an epistemic triangle (Moscovici, 1961).

As expected (hypothesis c.) a basic difference in the conception of mental-health emerged in the two cultural contexts investigated. Whilst in France this domain was associated with the

practice of hospitalization and psychiatrists were institutionally given responsibility to control such state of individual 'deviance' in Italy professionals were expected to work within a dimension of socialization through schemes to rehabilitate the mentally ill within the community. These cultural differences – based on our hypothesis concerning the anchoring into the French and Italian legal and societal frameworks of the SR of mental-health and the relative social practices – highlight the interest of cross-cultural differences for studies on SN. They do so both from the perspective of content-knowledge building and meaning-sharing among the SNS users (including their representational systems) and from the perspective of shaping the use of technology. This assumption has been adopted by authors who have empirically demonstrated “that even more so than experience of use, culture is a key behavioural determinant” and on revisiting the behaviour chain model, have “found that even though users had to their disposal the same set of features, the uses for creating content and value and involving others and the time investment for staying active and loyal in the site differed across countries.” (Vasalou, Joinson & Courvoisier, 2010:727).

The French Class 4 comprised conversations from *Facebook Pages* ($\chi^2 787.9$) and *Groups* ($\chi^2 64.7$) and from *Twitter* ($\chi^2 52.3$). Psychoanalysis ($\chi^2 31.6$) and the psychoanalyst ($\chi^2 85.4$) were the main objects of discussion within this class. The semantic fields and the underlying representation of psychoanalysis were linked to the psychological and psychotherapeutic domains expressed through words like *psychology* ($\chi^2 371.3$), *health* ($\chi^2 370.7$), *psychologist* ($\chi^2 318.7$) and *psychotherapist* ($\chi^2 232.1$). Notably, in this case users mainly exchanged information on the training and career paths required to achieve the professional competence and status of psychoanalyst, thus introducing a discussion which negotiated the scientific boundaries of the discipline itself.

Similarly, the Italian Class 3 was based on posts retrieved from *Twitter* ($\chi^2 6963.9$), whereas mental-health ($\chi^2 2915.2$) and mental-

illness ($\chi^2 5318$) constituted the main objects of representations. In this case, the discourse focused on promotion of services and activities provided by local aid agencies through the words *mental* ($\chi^2 7493.2$), *health* ($\chi^2 6362.1$), *centre* ($\chi^2 200.6$) and *department* ($\chi^2 155$). Furthermore, the recurrence of Italian city names marked a common reference to specific local and community aid centres. Twitter was evidently employed as a marketing tool to exchange information and updates on services and initiatives, rather than being a 'room' for conversation.

Study of the communication systems underlying these online conversations yielded even more interesting findings. Because changes in communication systems are significant for studying SR, we assumed that the SNS investigated would implement specific aggregative and communicative attributes, differing considerably in regard to their members' patterns of affiliation and segmentation, constraints on the length of messages and their semantic context and emotional expressiveness (hypotheses a. and d.).

In particular, we found in Facebook Pages experts and institutions belonging to the field of mental-health that stimulated and contributed to discussions on the topic of mental disorders in order primarily to promote their expertise and 'advertise' their work and activities so as to provide primary support for persons who might resort to the Internet to find such information and support. The same objectives seemed to characterize communication on Twitter, although in this case it is not possible to speak of conversations, but rather of information exchanges and updates on initiatives and activities characterizing the agenda of professionals and agencies in the field of mental-health. Differently and interestingly, in *Yahoo! Answers* as well as in *Facebook Profiles*, psychological problems and disorders were contextualized in a sort of *mutual aid context* in which users, according to different levels of familiarity with expert knowledge, tried to share and negotiate information to support each other in identifying adequate remedies or professional

aid for precarious psychological conditions (in agreement with our hypothesis d.).

CONCLUSION

Finally, we sought to verify Moscovici's original hypothesis that the dynamic relation between SR and public issues in the new 'social arena' of SNS is likely to determine the symbolic positioning of groups with similar or conflicting interests and visions of the world (de Rosa, 2013). In our research we found a specific struggle for the definition of roles, forms of knowledge and practices related to the domain of mental-health negotiated among actors characterized by different levels of access and proximity to the objects of the representations investigated. On the one hand, we found lay users asking for advice and information from the community, specifically about specialist and professional services in the field of mental-health. They were frequently re-directed to instances of support and aid by the online community, so that scientific mediation was replaced with mutual

Figure 1. The psychologist and the SN



understanding. In this framework, professionals are even averted through denouncing their pointlessness competence, in view of a relative distance from 'real' life and everyday 'common' experiences.

On the other hand, posts produced by professionals in the mental-health field emphasised their

Table 1. The main feature of the research design

THEORETICAL FRAMEWORK	THE RESEARCH IN SHORT		
	OBJECTIVES	HYPOTHESES FOCUSED ON	MAIN RESULTS
We focus the interest of promoting researches on social representations, which take account of new arenas for social transmission and elaboration of knowledge through social exchanges among the members of the thinking society. (Moscovici, 1961/1976). Co-production among social networks' users, allow to go over the present sociological supremacy (influenced by mathematical and computer-science-based contributes) to the study of the structure and content of social networking (vs. the paradigm of Social Network Analysis, SNA) (Carrington, Scott & Wasserman, 2005).	To explore: - the social representations of psychoanalysis, psychiatry and mental health emerging as interrelated systems of representations from online conversations on SNS; - the positioning of social actors, groups and communities on the basis of the social representations shared; - the communication systems emerging from the user-generated content analysed according to different constraints characterizing each of the SNS considered.	a) position of different social groups in the SN.	a) The French Class 1 contained conversations mainly collected from Facebook Pages (chi2 1802,4) and related to psychoanalysis (chi2 155,9), mental health (chi2 46,2) and psychoanalyst (chi2 28). In this class, the general topic of mental health was prevalent, as testified by the occurrence of terms such as disorder (chi2 274,2), depression (chi2 262,4), symptoms (chi2 203,2) and anxiety (chi2 126,6). As expected, the language and the style emerging from the analysis of ECUs highlighted the presence of users characterized by high levels of expertise in the field, particularly groups and organizations, rather than private individuals, displaying specialized lexicons and information. The topic of mental health was also approached from a non-professional, heuristic perspective, reproducing a naive view of mental health. The lexicon utilised included words pertaining to everyday language and especially linked to the emotions, actors and actions of ordinary social life and aspects regarding socialization: e.g. love (chi2 269,8), man (chi2 193,5), eat (chi2 121,7) and wife (chi2 118,0) in the French context and life (chi2 167,4), speak (chi2 164,8), child (chi2 158,2) and friend (chi2 104,1) in the Italian Class 1.
		b) differences between the social representations of psychoanalysis and psychiatry in their approaches to mental illness.	b) As regards the representations identified, there emerged a difference between the social representations of psychoanalysis and psychiatry in their approaches to mental illness. In regard to the former, mental illness was described in terms of a "painful psychic process" (French Class 1). By contrast, the latter conceived mental illness as a "disease" requiring suitable treatment relying on diagnostic techniques and pharmaceutical methods (Italian Class 2).
		c) cross-national differences.	c) the Italian Class 2 included conversations from Facebook Pages (chi2 2276,3) and to a lesser extent from Groups (chi2 27,6). Psychoanalysis (chi2 260,7) and psychiatry (chi2 289,9) were the main objects of discussion. Psychiatry (chi2 321,3) also represented the most relevant cited word in the class, followed by the terms social (chi2 221,7), psychiatric (chi2 209), psychology (chi2 199,9) and patient (chi2 197,5). Differently from the French context and in agreement with hypothesis c) the representation of psychiatry and the domain of mental health lay on a communitarian dimension, whereas intervention by professionals was centred on a rehabilitation process targeted on interpersonal and family relationships: These cultural differences – based on our hypothesis concerning the anchoring into the French and Italian legal and societal frameworks of the social representations of mental health and the relative social practices – highlight the interest of cross-cultural differences for studies on SN.
		d) different communicational styles defined by the context of Social Media.	d) In particular, we found in Facebook Pages experts and institutions belonging to the field of mental health that stimulated and contributed to discussions on the topic of mental disorders in order primarily to promote their expertise and "advertise" their work and activities so as to provide primary support for persons who might resort to the Internet to find such information and support. The same objectives seemed to characterize communication on Twitter, although in this case it is not possible to speak of conversations, but rather of information exchanges and updates on initiatives and activities characterizing the agenda of professionals and agencies in the field of mental health. Differently and interestingly, in Yahoo! Answers as well as in Facebook Profiles, psychological problems and disorders were contextualized in a sort of mutual aid context in which users, according to different levels of familiarity with expert knowledge, tried to share and negotiate information to support each other in identifying adequate remedies or professional aid for precarious psychological conditions.

engagement in re-definition of their identities and the borders between 'expert' knowledge and common sense. They symbolically 'opened the doors of their rooms' through participation in the online discussion, harmonizing presences, styles and lexicons on the basis of the new communication patterns required by these web environments.

Briefly, in this new communicative scenario the results of our study show how different target groups act new practices, showing their positioning:

- Users act as 'infomediaries' of expert knowledge, providing informal help and suggestions online;
- Experts open the doors of their "physical-rooms" to "cyber-rooms".

Echoing the famous words of Woody Allen: "Psychoanalysis is a myth kept alive by the sofas", meaningful results from our research show how the "sofa" has moved in the recent decade "from the psychoanalyst's couch to SN". These are the new "living rooms" – in a cyber world (de Rosa, 2004) – where the familiarity effect induced among the SN members, who perceive themselves as "clubs of friends", contribute also to redefine the borders between the expert's knowledge, the professional's role aimed at the intervention and the lay people's social discourse around psychoanalysis and psychiatry.

FUTURE RESEARCH DIRECTIONS

Even if taken as a specific inquiry on SNS, we consider our research to be a preliminary step in the study of the semantic production of social-interactive phenomena generated by the ongoing global redefinition of patterns of communication and it is our conviction that further research in the field is warranted. We have shown how studies of social media can inform and be informed by, the concept of SR and its inseparable and constitutive role of communication.

REFERENCES

- Andersson, G., Carlbring, P., & Cuijpers, P. (2009). Internet interventions: Moving from efficacy to effectiveness. *Electronic Journal of Applied Psychology*, 5(2), 18–24.
- Bangerter, A. (1995). Rethinking the relation between science and common sense: A comment on the current state of social representation theory. *Papers on Social Representations*, 4, 61–78.
- Boyd, D. M., & Ellison, N. B. (2007). Social Network Sites: Definition, History, and scholarship. *Journal of Computer-Mediated Communication*, 13(1), 210–230. doi:10.1111/j.1083-6101.2007.00393.x
- Carrington, P.J., Scott, J., & Wasserman, S. (Eds.). (2005). *Models and Methods in Social Network Analysis*. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511811395
- Castells, M. (1996/2000). *The Rise of Network Society*. Oxford, UK: Blackwell.
- Castells, M., Fernandez-Ardevol, M., Qiu, J. L., & Sey, A. (2007). *Mobile Communication and Society: A Global Perspective*. Cambridge, MA: The MIT Press.
- Catanese, S., De Meo, P., Ferrara, E., Fiumara, G., & Provetti, A. (2012). Extraction and analysis of Facebook friendship relations. In A. Abraham (Ed.), *Computational Social Networks: Mining and Visualization* (pp. 291–324). London: Springer Verlag. doi:10.1007/978-1-4471-4054-2_12
- De Meo, P., Ferrara, E., Fiumara, G., & Ricciardello, A. (2012). A Novel Measure of Edge Centrality in Social Networks. *Knowledge-Based Systems*, 30, 136–150. doi:10.1016/j.knosys.2012.01.007
- de Rosa, A. S. (2004). Giocare ad auto-battezzarsi in internet: una ricerca sulla scelta dei nicknames tra i membri di due chat lines italiane. *Rassegna di Psicologia*, 1, 179–224.

- de Rosa, A. S. (2011). 1961-1976: a meta-theoretical analysis of the two editions of the "Psychanalyse, son image et son public". *Papers on Social Representations*, 20(2). Retrieved from <http://www.psych.lse.ac.uk/psr/>
- de Rosa, A. S. (2012). La psicoanalisi, la sua immagine, il suo pubblico: 1961, 1976, 2011. Compire 50 anni nell'era dei social networks. In I. Galli (Ed.), *Cinquant'anni di Rappresentazioni sociali. Bilanci e prospettive di una Teoria in continuo divenire* (pp. 59–101). Milano: Unicopli.
- de Rosa, A. S. (Ed.). (2013). *Social Representations in the "Social Arena"*. London: Routledge.
- Edwards, A., Housley, W., Williams, M., Sloan, L., & Williams, M. (2013). Digital social research, social media and the sociological imagination: Surrogacy, augmentation and re-orientation. *International Journal of Social Research Methodology*, 16(3), 245–260. doi:10.1080/13645579.2013.774185
- Farr, R. M., & Moscovici, S. (Eds.). (1984). *Social Representations*. Cambridge, UK: Cambridge University Press.
- Greitemeyer, T., & Kunz, I. (2013). Name-Valence and Physical Attractiveness in Facebook: Their Compensatory Effects on Friendship Acceptance. *The Journal of Psychology*, 153, 257–260. PMID:23724699
- Jankowski, N. W. (2002). Creating Community with Media: History, Theories and Scientific Investigations. In L. Lievrouw & S. Livingstone (Eds.), *The Handbook of New Media* (pp. 34–49). London: Sage. doi:10.4135/9781446206904.n4
- Jesuino, J. C. (2008). Linking Science to Common Sense. *Journal for the Theory of Social Behaviour*, 38(4), 393–409. doi:10.1111/j.1468-5914.2008.00377.x
- Lahlou, S. (2011). A comment on Chartier and Meunier. *Papers on Social Representations*, 20, 38.1–38.7.
- Licoppe, C., & Smoreda, Z. (2005). Are Social Networks Technologically Embedded?: How Networks Are Changing Today with Changes in Communication Technology. *Social Networks*, 27(4), 317–335. doi:10.1016/j.socnet.2004.11.001
- Markova, I. (2009). Dialogicità e conoscenza. In *Paradigmi delle Rappresentazioni Sociali*. Il Mulino.
- Mazzara, B. (Ed.). (2008). *I discorsi dei media e la psicologia sociale*. Roma: Carocci.
- Mecacci, L. (2000). *Il caso Marilyn M. e altri disastri della psicoanalisi*. Bari: Laterza.
- Meyer, C. (2005). *Le livre noir de la psychoanalyse*. Paris: Les Arènes.
- Moscovici, S. (1995). Vygotsky, le Grand Robert et la cyber-représentation. *Cahiers Internationaux de Psychologie Sociale*, 28(4), 15–21.
- Moscovici, S. (1997, May 27). *Social Representations Theory and Social Constructionism*. Retrieved from <http://psyberlink.flogiston.ru/internet/bits/mosc1.htm#back2>
- Moscovici, S. (1961). La Psychanalyse, son Image et son Public. Étude sur la représentation sociale de la psychanalyse. Paris: Presses Universitaires de France.
- Nadkarni, A., & Hofmann, S. G. (2011). Why Do People Use Facebook? *Personality and Individual Differences*, 52(3), 243–249. doi:10.1016/j.paid.2011.11.007 PMID:22544987
- O'Reilly, T. (2005). What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. *O'Reilly Network*. Retrieved from <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>

Onfray, M. (2010). *Le crépuscule d'une idole, l'affabulation freudienne*. Paris: Grasset.

Ponte, D., & Simon, J. (2011). Scholarly communication 2.0: Exploring researchers opinions on Web 2.0 for scientific knowledge creation, evaluation and dissemination. *Serials Review*, 37(3), 149–156. doi:10.1080/00987913.2011.10765376

Reinert, M. (1986). Un logiciel d'analyse textuelle: ALCESTE. *Les Cahiers de l'Analyse des Données*, 3, 187–198.

Rowland Smith, R. (2010). *Death-drive: Freudian Hauntings in Literature and Art (The Frontiers of Theory)*. Edinburgh, UK: Edinburgh University Press. doi:10.3366/edinburgh/9780748640393.001.0001

Ryan, T., & Xenos, S. (2011). Who uses Facebook? An investigation into the relationship between the Big Five, shyness, narcissism, loneliness and Facebook usage. *Computers in Human Behavior*, 27(5), 1658–1664. doi:10.1016/j.chb.2011.02.004

Scott, L. (2004). *Social Network Analysis: A Handbook*. New Haven, CT: Sage Publication.

Shingleton, R. M., Richards, L. K., & Thompson-Brenner, H. (2013). Using technology within the treatment of eating disorders: A clinical practice review. *Psychotherapy (Chicago, Ill.)*, 50(4), 576–582. doi:10.1037/a0031815 PMID:23527906

Stoneman, P., Sturgis, P., & Allum, N. (2012). Exploring public discourses about emerging technologies through statistical clustering of open-ended survey questions. *Public Understanding of Science (Bristol, England)*, 22(7), 850–868. doi:10.1177/0963662512441569 PMID:23825238

Travers, J., & Milgram, S. (1969). An Experimental Study of the Small World Problem. *Sociometry*, 32(4), 425–443. doi:10.2307/2786545

Vasalou, A., Joinson, A. N., & Courvoisier, D. (2010). Cultural differences, experiences with social networks and the nature of the “true commitment” in Facebook. *International Journal of Human-Computer Studies*, 68, 719–728. Retrieved from http://luminainteractive.com/pdfs/facebook_ijhcs10.pdf

Williford, C., & Henry, C. (2012). *Computationally Intensive Research in the Humanities and Social Sciences*. Washington, DC: Council on Library and Information Resources.

ADDITIONAL READING

Abric, J. C. (1994). *Pratiques sociale et représentation*. Paris: P.U.F.

Bonardi, C., & Roussiau, N. (1999). *Les représentations sociales*. Paris: Dunod.

de Rosa, A. S. (2001). The king is naked. Critical Advertisement and Fashion: the Benetton Phenomenon. In K. Deaux & G. Philogène (Eds.), *Representations of the Social* (pp. 48–82). Oxford: Blackwell.

Jodelet, D. (1989). *Les représentations sociales*. Paris: P.U.F.

KEY TERMS AND DEFINITIONS

Facebook: Free social networking service. Activated in 2004, it registered about 100 million users in less than 9 months.

Google Advanced Search: The most popular and currently the most used web search engine.

Social Networks: In the definition of social networks, inter-relationality among diverse social actors, such as individual private users and organizations, is a core element.

Social Representations: Common sense theories on key aspects of the world that allow individuals and groups to represent it and master it.

Twitter: Free social network and micro-blogging service that gives its users a personal page that can be updated by means of text messages with a maximum length of 140 characters.

User-Generated Content: Any form of content such as blogs, forums, posts, chats, tweets... created by users of an online system or service, often made available via social media websites.

Yahoo! Answers: Community-driven question-and-answer (Q&A) site launched by *Yahoo!* in July 2005.

S

The Internet Behavior of Older Adults

Elizabeth Mazur

Pennsylvania State University, USA

Margaret L. Signorella

Pennsylvania State University, USA

Michelle Hough

Pennsylvania State University, USA

INTRODUCTION

Internet activities of persons over age 65 include communication (i.e., email, instant messaging); social media (i.e., Facebook, LinkedIn, Twitter), search engines (e.g., Google), online newspapers (e.g., nytimes.com), magazines (e.g., salon.com), discussion forums, classifieds (e.g., craigslist.org), photo sharing (e.g., Flickr), online banking, and bill paying. Most studies on this topic (i.e., Choi & DiNitto, 2013; Madden, 2010) have concentrated on email, texting, web surfing, banking, shopping, and more recently, social media, and are notable for being conducted by researchers in fields as diverse as psychology, business, communications, social work, gerontology, computer science, engineering, and robotics.

Although this entry emphasizes research with persons over age 65, some studies exclude age information, preferring instead the phrases retirees (Hahm & Bikson, 1989), senior citizens, or, as in one early publication (Edwards & Engelhardt, 1989), members of AARP (age 50 and over). Publications that do specify participant age often group together persons differing in age by as much as 30 to 40 years (e.g. Trocchia & Janda, 2000; Zhang & Kaufman, 2015), inconsistently define *older adult* as beginning at ages 55 (i.e., Wong, Yeung, Ho, Tse, & Lam, 2014; Zhang & Kaufman, 2015), 59 (e.g., Berner et al., 2015;), or 60 (i.e., Choi & DiNitto, 2013), or do not specify age distribution past the defined minimum age

(i.e., Reisenwitz, Iyer, Kuhlmeier, & Eastman, 2007; Zhang & Kaufman, 2015).

BACKGROUND

The earliest publications concerning older persons' Internet behaviors – before the word *Internet* was ever used – speculated that technological solutions, connected computers included, could improve the daily lives of older persons in their homes and communities. Theorists initially centered their attention on two-way communication, especially to mitigate perceived loneliness. For example, Ramm and Gianturco (1973) envisaged a “picture communication system with the aid of computers” (p. 325) that could personalize entertainment, education and home employment in order to fight isolation and feelings of uselessness.

As editor of a special issue of *American Behavioral Scientist* on technology and aging, Monk (1988a, 1988b) emphasized the significance of networked communication to the isolated elderly and argued that the elderly are often active users of technology rather than passive recipients. In 1989, Edwards and Engelhardt (1989) explicitly introduced the concept that older persons are not necessarily technophobic. The self-selected college-educated survey respondents over age 50 were generally positive about computers, notably in terms of a short-term change in attitudes as a result of their brief supervised technology experience.

rience and interest in Internet-related computer use for health insurance benefits, personal and family medical history, genealogy, and stock market information. In a comparison of retired and employed computer inexperienced public utility workers in California, Hahm and Bikson (1989) found that email increased interactions among retirees, though, according to the authors, “older adults must be given multiple sources of instruction” (p. 127).

OLDER ADULTS AS ACTIVE INTERNET USERS

Fortunately, a growing body of theoretically informed, empirically based literature, still mostly focused on electronic mail, began appearing in the 1990s. Beyond their specialized foci, research in this area underscored that the elderly are active users of technology, more capable of understanding and enjoying the Internet than was originally assumed.

The Pew Research Center Describes the “Gray Gap”

Beginning with its initial survey of the “gray gap” between young and older Americans (Lenhart, 2000), Pew Internet & American Life Project consistently has recounted that Americans age 65 and older are among the least likely groups to go online; once online, however, they are typically enthusiastic emailers, information searchers, and social networkers.

Why have older persons been less likely to go online? Initially, few persons over age 65 had Internet access (Lenhart, 2000), partly due to many being retired and some living on fixed incomes. However, economic factors never by itself explained the age differential. A second major factor was lack of contact with computers, and this was true even for older Americans with college or graduate degrees. As detailed by Fox (2004), most older persons lived lives far removed from

the Internet, knew few people who used email or surfed the Web, and could not imagine reasons for spending money and time learning how to use a computer. In addition, Fox (2004) reported that persons over 65 were more likely than any other age group to be living with disability, which could hinder their ability to obtain computer training or read the small type on many Web sites.

Since the initial 2000 report, each subsequent Pew paper has described an increase in older persons’ Internet use, especially between 2000 and 2004, when Fox (2004) reported a 47% jump. In 2001, Fox enumerated the top Internet interests among seniors who go online to include mainly email, but also hobby information, news, health information, browsing “just for fun,” and weather updates. By 2010, although users age 65 and older were still primarily using email to maintain personal contacts, 13% - more than three times as many as in 2009 - were relying on social media to manage their daily communications – sharing links, photos, videos, news and status updates (Madden, 2010). By 2013, the percentage of social media users had jumped to 27%. Yet, older computer users are still significantly less likely than all American adults (63%) to use social networking sites (Smith, 2014). According to Duggan (2015), of the five major platforms of current social media, Facebook is the most popular among older adult Internet users (48%), though they are less likely to go on Facebook than all U.S. adult Internet users (72%). Connected older adults also use – though less frequently – the social media platforms Pinterest (16%), LinkedIn (12%), Twitter (6%), and Instagram (4%). Relatively few older wired users participate in other online platforms such as discussion forums (8%) and Tumblr; only 2% of Internet users ages 65 and older report doing so (Duggan, 2015).

Since 2000, wired seniors consistently have been more likely than their offline peers to be younger, married, white, highly educated, and enjoying relatively high incomes; recent data confirm this trend (Smith, 2014). Lenhart and Duggan (2014) noted some interesting behaviors

among older Internet users in committed relationships; almost half (47%) reported sharing email accounts with their partner, as compared to 27% of all Internet users who were married or in a committed relationship. Older users' sharing of social networking profiles was less frequent (11%), and unlike email account sharing, there was little age variation in this behavior (Lenhart & Duggan, 2014).

Gender differences in Internet usage have been inconsistent over the years. Although in 2000 about 40% of older Internet users were women, in 2004 the gender ratio among U.S. wired seniors appeared to have reached parity, similar to the general Internet population. However, recent data indicates that almost two-thirds (65%) of men ages 65 years and older use the Internet or email as compared to 55% of women (Smith, 2014). Interestingly, whereas disability initially may decrease computer use, more recently older adults living with chronic disease have been increasingly likely to state that they work on a blog or contribute to an online discussion, a listserv, or another Internet forum that helps people with personal issues or health problems (Madden, 2010). According to Smith (2014), older adults with physical challenges often enjoy tablets and e-readers, but at a rate that is somewhat lower (22%) than those without health or disability issues (30%).

There is a burgeoning group of Americans who are slightly younger than retirees and who are more attached to the online world. This "silver tsunami" has gained momentum through the past decade. Internet users age 65 or older are often cited as the fastest-growing demographic group online, but that description can be misleading. Much, though not all, of the growth in this age group over the past decade has come from long-time Internet users aging into senior status rather than from waves of elderly nonusers in their seventies and eighties suddenly going online. Current data indicate that after ages 65-69, Internet and broadband use decreases with age and falls off most dramatically starting at about age 75 (Smith, 2014). At the same time, data indicate that most

older adults who become first-time Internet users in their senior years become enthusiastic users, going online daily or almost every day (Smith, 2014).

Does Use of the Internet Improve Older Adults' Social Involvement and Quality of Life?

The first longitudinal study to examine causal relationships among use of the Internet, social involvement, and psychological consequences suggested potential drawbacks to Internet use, although older persons do not appear to have been included. Kraut and his colleagues (1998) tracked participants longitudinally, and over 12 to 24 months, disturbing trends were identified. Increases in Internet use predicted less communication within the family and more loneliness and depression. The initial reactions in the press were not surprising ("Sad, Lonely World Discovered in Cyberspace"; Harmon, 1998); however, Kraut's research team later found that the negative effects did not persist over longer periods of time (Kraut et al., 2002). Nonetheless, research continues to suggest that there could be social and emotional drawbacks to Internet use. For example, Stepanikova, Nie, and He's (2010) large cross-sectional study of U.S. adults ages 18 to 70 showed negative impacts on interpersonal interaction, communication, loneliness, and life satisfaction. These hypothesized consequences are sometimes called *displacement*; the Internet replaces presumably more meaningful face-to-face interactions (Huang, 2010).

The implication that Internet use might lead to social isolation is problematic for older adults, at least in the US, given their high risk for depression, especially for those who experience loneliness and lack social support (e.g., Seritan, McCloud, & Hinton, 2009). Yet, it is possible that Internet communication actually might facilitate social networks among those who are geographically dispersed (Stepanikova et al., 2010). Early research in the US did suggest significant benefits of computer use for older adults. For example,

White and colleagues (1999) taught retirement community members (average age 77 years) to operate computers for word processing, accessing the Internet, and email. There was an immediate post-training improvement in feelings of loneliness among the 15 computer users compared to the eight members of the control group, but this difference was not significant at the five-month follow-up.

In a review of the literature provocatively titled “Computer use has no demonstrated impact on the well-being of older adults,” Dickinson and Gregor (2006) outlined several reasons supporting their thesis. One problematic issue identified was that many of the intervention studies did not control for variables that likely accompanied computer instruction, such as face-to-face interactions. The authors also noted a tendency for small or even nonsignificant effects to be considered *improvements* when summarized in subsequent research. Finally, Dickinson and Gregor (2006) recognized that individuals whose well-being is higher initially are more likely to successfully learn to use computers and are less likely to drop out.

Recent research using large samples, longitudinal data, and mediational models, as well as reviews and meta-analyses, suggests that there are likely no major negative consequences of Internet use for older adults, and there may be some benefits. The least positive recent report was in a meta-analysis by Huang (2010), who examined the relation between measures of psychological well-being and Internet use. Although Huang found a very small association between greater Internet use and reduced well-being, a major limitation Huang cites is that only 7.5% of the participants were adults over age 55. On the positive side, Chang and colleagues (2015) studied adult Facebook users in the U.S. Younger adults had more Facebook friends, but increasing age was associated with a higher proportion of Facebook friends that were actual friends beyond the Internet. Across all ages, having a higher proportion of actual Facebook friends was associated with lower reports of loneliness.

Cotton and colleagues have conducted several recent studies suggesting positive impacts of Internet engagement on older adults. Initially, the researchers studied residents of communities for aging persons in one U.S. state (Cotton et al., 2013). Although the sample was not diverse and predominantly female, it is notable for having an average participant age of 83. The major positive result was that even after controlling for possible confounding factors, greater Internet use was correlated with lower depression. A longitudinal study with a larger sample size by Cotton and co-authors (2014) similarly showed a reduction in depression in older adults associated with higher Internet use, with a stronger effect for individuals living alone. Thus, taken together, the research studying Internet engagement and psychological well-being suggests that Internet use is unlikely to cause great harm, and shows promise for conditions such as depression.

OBSTACLES TO OLDER ADULTS' INTERNET USE

The three most common barriers to older adult computer use cited in the literature are physical disability, cognitive limitations, and negative attitudes towards computers.

Physical Limitations

Common recommendations to ameliorate physical issues that may impact older adults as well as younger persons with disabilities include using large fonts, providing visual contrast, and minimizing actions such as scrolling that might be impacted by physical difficulties (Charness & Boot, 2009; Kim, 2008). Novice users perform better with input devices that allow direct interaction with the screen; persons with severe mobility problems typically do best with voice input (Fisk et al., 2004). Computer programs and systems now often include such accessibility features that users can activate when needed (Kim, 2008),

and websites have improved in ergonomic design (Nielsen, 2013). Recent tests by Nielsen (2013) showed that computer users over 65 performed better than previous cohorts, but still had more challenges than those under 55. On all measures (success, speed, errors, satisfaction), older users performed more poorly. Visual issues were one problem, with tiny text or poor contrast causing difficulties; another was what Nielsen calls “unforgiving” design, such as forms that cannot handle minor typos or hyphens in phone numbers. Larger mobile devices may help overcome some of these issues. Smith (2014) reports that 27% of older users own a tablet, an e-reader, or both, more than the 18% with smartphones.

Cognitive Limitations

Research has suggested that older computer users are likely to be slower to learn how to operate computers and their programs and that they will commit more errors in doing so (Charness & Boot, 2009; Kim 2008; Nielsen, 2013). Some of these problems may be due to declines in working memory, problems in regulating attentional processes, lesser abilities to form new procedural knowledge, difficulties in multitasking, and movement difficulties that impact the ability to respond to commands or use computer program menus and other interface elements (Fiske et al., 2004). Rogers and Fisk (2010) concluded that the challenge is translating knowledge of people’s potential limitations into universal design.

Much attention has been given in recent years to the question of whether computer gaming can ameliorate cognitive decline. At present there are no clear answers. The online “brain training” site, Luminosity, recently had to refund its customers because the company made unsubstantiated claims that the games could improve cognitive functioning and reduce dementia risk (Span, 2016). There have been numerous failures in recent empirical work to find positive effects of online games on cognitive functioning. A meta-analysis by Powers and colleagues (2013) did indicate some

experimental evidence of positive impacts on older adults, but the domain affected most was motor skills.

Attitudes

Research examining the attitudes of older computer users has been difficult to interpret. Although some studies have shown improved attitudes towards computers after training (Kim, 2008), it cannot therefore be assumed that older persons’ negative attitudes towards computers cause computer or Internet avoidance. Nonetheless, older users remain “slower to adapt to new technologies and more likely to report anxiety...or frustrations” (Rogers & Fisk, 2010, p. 647). Researchers also report older adults’ difficulties in learning as well as the lack of availability of in-person help and access to computers and technology (Chen et al., 2013). However, because these studies lack employed older persons, who are more likely to have computer experience, the challenge for researchers lies in disentangling computer knowledge (and, relatedly, computer experience) from factors such as income, education, and health status.

Little attention has been given to the issue of older users’ knowledge of computer security. Grimes, Hough, Mazur, and Signorella (2010) examined the relationship of age to knowledge of various Internet hazards among 167 US adults ages 19 to 91 years. Older persons (average age 72) showed less knowledge of Internet hazards, such as viruses and spam, than did younger individuals (average age 24). Among the older adults, however, education and prior computer experience were more important in predicting Internet security awareness than were age and gender.

Thus, much of the research suggests that older persons may be more vulnerable to the difficulties that typically plague all computer users, such as horizontal scrolling, inconsistent controls, complex security rules, and other basic design flaws. Yet bad design and complexity may be particularly problematic for the oldest, especially when combined with their higher likelihood of being

computer novices; these challenges may impact confidence and persistence.

CONSUMER IMPLICATIONS OF OLDER ADULTS' INTERNET BEHAVIORS

The consumer implications of older adults' Internet usage have been considerable, at least in the US. According to Roberts and Manolis (2000), age is the most important demographic characteristic for consumer marketers and researchers. Kamal and Patil (2003) note that many older adults in the US are rich in both time and money (see also Hough & Kobylanski, 2009). Yet the buying patterns and attitudes of this segment differ from those of younger consumers. Persons age 65 and older have been less likely than those ages 45 to 64 to use the Internet for online purchasing; in particular, those older persons who lack confidence in their Internet skills use it significantly less frequently (AARP, 2000).

From the results of interviews with 12 American men and women ages 57 to 87 years old, Trocchia and Janda (2000), identified differences between older Internet users and non-users. Expanding on that study, Hough and Kobylanski (2009) theorized that marketing interventions focused on placing technology in consumers' hands can be particularly useful in influencing older persons' reference group affiliation, nature of social relations, and perception of reality. They also noted that marketing focused on price or on special promotions have little to moderate impact on increasing older persons' engagement with computers or the Internet. According to results from surveys of 171 Americans between ages 65 and 85, Iyer and Eastman (2006) found that older persons' attitudes about the Internet were also affected by whether they felt that the Internet was fun, convenient, and efficient. Lee, Han, and Chung (2014) found perceived usefulness, enjoyment, and ease of use to be crucial factors in Internet adoption by mature users.

Reisenwitz and colleagues (2007) posited that other factors that impact the effectiveness of marketing to older consumers include emphasizing the ability to create nostalgic experiences online, promoting aspects of innovation, and addressing perceptions of risk. Bruine de Bruin, Parker, and Strough (2016) note multiple studies that indicate that older adults tend to search for fewer options and prefer limited selection sets as compared with their younger counterparts. Older adults are significantly less inclined to engage in intensive alternate searching with the goal of selecting the optional choice. These results suggest that older consumers are less likely to engage in prolonged on-line research sessions when making purchase decisions. Taken together, these studies indicate that the commercial implications of older adult engagement with the Internet continue to be challenging and complex.

FUTURE RESEARCH DIRECTIONS

The issue of whether the Internet will improve older adults' quality of life and social relationships remains unresolved, due partly to the lack of generalizability of many empirical findings. With the exception of studies from Pew Research Center, much of the research has ignored older persons living in private households, independently surfing the Internet. Researchers also often omit information concerning participants' gender, income, education, health, and employment status.

Despite the commercial advertisements, there are few experimental studies of computer gaming's possible link to reduction of cognitive decline in older adults. There is also little research on telehealth, which promises improved access to medical and psychological services, especially for persons with disabilities and older persons living isolated lives (Khubchandani & Thew, 2016). There is a paucity of cross-cultural research, especially in the global south, where promoting Internet access to resources and fostering social inclusion may be more important than in the west, where most of the research has been completed.

CONCLUSION

As Ogozalek (1991) first predicted in an early review of the impact of computer technology on aging, cohort effects complicate prediction from one age group of older adults to another. Charness and Boot (2009) have questioned whether the “gray gap” in both computer and Internet use described over a decade ago (e.g., Lenhart, 2000) will ever be completely eliminated. Current research still has not answered whether the ongoing pace of change in technology, combined with changes associated with aging, may mean that there will always be fewer older than younger adults using the newest technologies. However, as current Internet users in their 50s and early 60s age into the senior population, they are likely to maintain positive attitudes towards technology and continue frequent use, which the research suggests may benefit mental health and social relationships.

REFERENCES

- AARP. (2000, March). *AARP national survey on consumer preparedness and e-commerce: A survey of computer users age 45 and older*. Washington, DC: AARP Research Group. Retrieved from <http://assets.aarp.org/rgcenter/consume/ecommerce.pdf>
- Berner, J., Rennemark, M., Jogr  us, C., Anderberg, P., Sk  ldunger, A., Wahlberg, M., & Berglund, J. et al. (2015). Factors influencing Internet usage in older adults (65 years and above) living in rural and urban Sweden. *Health Informatics Journal*, 21(3), 237–249. doi:10.1177/1460458214521226 PMID:24567416
- Bruine de Bruin, W., Parker, A., & Strough, J. (2016). Choosing to be happy? Age differences in maximizing decision strategies and experienced emotional well-being. *Psychology and Aging*, 31(3), 295–300. doi:10.1037/pag0000073 PMID:26913927
- Chang, P. F., Choi, Y. H., Bazarova, N. N., & Lockenhoff, C. E. (2015). Age differences in on-line social networking: Extending Socioemotional Selectivity Theory to social network sites. *Journal of Broadcasting & Electronic Media*, 59(2), 221–239. doi:10.1080/08838151.2015.1029126
- Charness, N., & Boot, W. R. (2009). Aging and Information Technology use: Potential and barriers. *Current Directions in Psychological Science*, 18(5), 253–258. doi:10.1111/j.1467-8721.2009.01647.x
- Chen, Y., Lee, B. D., & Kirk, R. M. (2013). Internet use among older adults: Constraints and opportunities. In R. Z. Zheng, R. D. Hill, & M. K. Gardner (Eds.), *Engaging older adults with modern technology: Internet use and information access needs* (pp. 124–141). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-1966-1.ch007
- Choi, N. G., & DiNitto, D. (2013). The digital divide among low-income homebound older adults: Internet use patterns, eHealth literacy, and attitudes toward computer/Internet use. *Journal of Medical Internet Research*, 15(5), e93. doi:10.2196/jmir.2645 PMID:23639979
- Cotten, S. R., Anderson, A. W., & McCullough, M. B. (2013). Impact of Internet use on loneliness and contact with others among older adults: Cross-sectional analysis. *J of Medical Internet Resesearch*, 15(2), e39. doi:10.2196/jmir.2306 PMID:23448864
- Cotten, S. R., Ford, G., Ford, S., & Hale, T. M. (2014). Internet use and depression among retired older adults in the united states: A longitudinal analysis. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, 69(5), 763–771. doi:10.1093/geronb/gbu018 PMID:24671896
- Dickinson, A., & Gregor, P. (2006). Computer use has no demonstrated impact on the well-being of older adults. *International Journal of Human-Computer Studies*, 64(8), 744–753. doi:10.1016/j.ijhcs.2006.03.001

- Duggan, M. (2015, August 19). *Mobile messaging and social media 2015*. Pew Research Center. Retrieved from <http://www.pewInternet.org/2015/08/19/mobile-messaging-and-social-media-2015/>
- Edwards, R., & Engelhardt, K. G. (1989). Micro-processor-based innovations and older individuals: AARP survey results and their implications for service robotics. *International Journal of Technology & Aging*, 2(1), 42–55.
- Fisk, A. D., Rogers, W. A., Charness, N., Czaja, S. J., & Sharit, J. (2004). *Designing for older adults: Principles and creative human factors approaches*. New York, NY: Taylor & Francis. doi:10.1201/9781420023862
- Fox, S. (2001, September 9). *Wired seniors*. Pew Internet & American Life Project. Retrieved from <http://www.pewInternet.org>
- Fox, S. (2004, March 28). *Older Americans and the Internet*. Pew Internet & American Life Project. Retrieved from <http://www.pewInternet.org>
- Grimes, G. A., Hough, M. G., Mazur, E., & Signorella, M. L. (2010). Older adults knowledge of Internet hazards. *Educational Gerontology*, 36(3), 173–192. doi:10.1080/03601270903183065
- Hahm, W., & Bikson, T. (1989). Retirees using social media and networked computers. *International Journal of Technology & Aging*, 2(2), 113–123.
- Harmon, A. (1998, August 30). Sad, lonely world discovered in cyberspace. *The New York Times*. Retrieved from <http://www.nytimes.com>
- Hough, M., & Kobylanski, A. (2009). Increasing elder consumer interactions with information technology. *Journal of Consumer Marketing*, 26(1), 39–48. doi:10.1108/07363760910927037
- Huang, C. (2010). Internet use and psychological well-being: A meta-analysis. *Cyberpsychology, Behavior, and Social Networking*, 13(3), 241–249. doi:10.1089/cyber.2009.0217 PMID:20557242
- Iyer, R., & Eastman, J. K. (2006). The elderly and their attitudes toward the Internet: The impact on Internet use, purchase, and comparison shopping. *Journal of Marketing Theory and Practice*, 14(1), 57–67. doi:10.2753/MTP1069-6679140104
- Kamal, M., & Patil, G. D. (2003). Computer access and utilization patterns among older people. *Journal of American Academy of Business, Cambridge*, 3(1/2), 319–322.
- Khubchandani, A., & Thew, D. (2006). Achieving equity in the health of women with disabilities through telehealth: Challenges and benefits. In S. Miles-Cohen & C. Signore (Eds.), *Eliminating inequities for women with disabilities* (pp. 83–91). Washington, DC: American Psychological Association.
- Kim, Y. S. (2008). Reviewing and critiquing computer learning and usage among older adults. *Educational Gerontology*, 34(8), 709–735. doi:10.1080/03601270802000576
- Kraut, R., Kiesler, S., Boneva, B., Cummings, J. N., Helgeson, V., & Crawford, A. M. (2002). Internet paradox revisited. *The Journal of Social Issues*, 58(1), 49–74. doi:10.1111/1540-4560.00248
- Kraut, R., Patterson, M., Lundmark, V., Kiesler, S., Mukhopadhyay, T., & Scherlis, W. (1998). Internet paradox: A social technology that reduces social involvement and psychological well-being? *American Psychologist*, 53(9), 1017–1031.
- Lee, E., Han, S., & Chung, Y. (2014). Internet use of consumers aged 40 and over: Factors that influence full adoption. *Social Behavior and Personality*, 42(9), 1563–1574. doi:10.2224/sbp.2014.42.9.1563
- Lenhart, A. (2000, September 21). *Who's not online*. Pew Internet & American Life Project. Retrieved from <http://www.pewInternet.org>
- Lenhart, A., & Duggan, M. (2014, February 11). *Couples, the Internet, and social media*. Pew Research Center. Retrieved from <http://www.pewInternet.org/2014/02/11/couples-the-Internet-and-social-media/>

- Madden, M. (2010, August 27). *Older adults and social media*. Pew Internet & American Life Project. Retrieved from <http://www.pewInternet.org>
- Monk, A. (1988a). Aging, loneliness, and communications. *The American Behavioral Scientist*, 31(5), 532–560. doi:10.1177/000276488031005004
- Monk, A. (1988b). Introduction. *The American Behavioral Scientist*, 31(5), 517–523. doi:10.1177/000276488031005002
- Nielsen, J. (2013, May 28). Seniors as web users. Nielson Norman Group. Retrieved from <https://www.nngroup.com/articles/usability-for-senior-citizens/>
- Ogozalek, V. Z. (1991). The social impacts of computing: Computer technology and the graying of America. *Social Science Computer Review*, 9(4), 655–666. doi:10.1177/089443939100900409
- Powers, K. L., Brooks, P. J., Aldrich, N. J., Palladino, M. A., & Alfieri, L. (2013). Effects of video-game play on information processing: A meta-analytic investigation. *Psychonomic Bulletin & Review*, 20(6), 1055–1079. doi:10.3758/s13423-013-0418-z PMID:23519430
- Ramm, D., & Gianturco, D. T. (1973). Computers and technology: Aiding tomorrows aged. *The Gerontologist*, 13(3), 322–326. doi:10.1093/geront/13.3_Part_1.322 PMID:4796064
- Reisenwitz, T., Iyer, R., Kuhlmeier, D. B., & Eastman, J. K. (2007). The elderly's Internet usage: An updated look. *Journal of Consumer Marketing*, 24(7), 406–418. doi:10.1108/07363760710834825
- Roberts, J. A., & Manolis, C. (2000). Baby boomers and busters: An exploratory investigation of attitudes toward marketing, advertising, and consumerism. *Journal of Consumer Marketing*, 17(6), 481–497. doi:10.1108/07363760010349911
- Rogers, W. A., & Fisk, A. D. (2010). Toward a psychological science of advanced technology design for older adults. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, 65B(6), 645–653. doi:10.1093/geronb/gbq065 PMID:20833690
- Seritan, A. L., McCloud, M. K., & Hinton, L. (2009). Geriatric depression - Review for primary care. *Current Psychiatry Reviews*, 5(2), 137–142. doi:10.2174/157340009788167329
- Smith, A. (2014, April 3). Older adults and technology use. Pew Research Center. Retrieved from <http://www.pewinternet.org/2014/04/03/older-adults-and-technology-use/>
- Span, P. (2016). F.T.C.'s Lumosity penalty doesn't end brain training debate. *New York Times*. Retrieved from <http://www.nytimes.com/2016/01/19/health/ftcs-lumosity-penalty-doesnt-end-brain-training-debate.html>
- Stepanikova, I., Nie, N. H., & He, X. (2010). Time on the Internet at home, loneliness, and life satisfaction: Evidence from panel time-diary data. *Computers in Human Behavior*, 26(3), 329–329-338.
- Trocchia, P. J., & Janda, S. (2000). A phenomenological investigation of Internet usage among older individuals. *Journal of Consumer Marketing*, 17(7), 605–612. doi:10.1108/07363760010357804
- White, H., McConnell, E., Clipp, E., Bynum, L., Teague, C., Navas, L., & Halbrecht, H. et al. (1999). Surfing the net in later life: A review of the literature and pilot study of computer use and quality of life. *Journal of Applied Gerontology*, 18(3), 358–378. doi:10.1177/073346489901800306
- Wong, C. K., Yeung, D. Y., Ho, H. C., Tse, K., & Lam, C. (2014). Chinese older adults Internet use for health information. *Journal of Applied Gerontology*, 33(3), 316–335. doi:10.1177/0733464812463430 PMID:24717738
- Zhang, F., & Kaufman, D. (2015). Social and emotional impacts of Internet use on older adults. *European Scientific Journal*, 11(17), 1–15.

ADDITIONAL READING

Burdick, D. C., & Kwon, S. (Eds.). (2004). *Gero-technology: Research and practice in technology and aging*. New York, NY: Springer.

Charness, N., & Schaie, K. W. (Eds.). (2003). *Impact of technology on successful aging*. New York, NY: Springer.

Mann, W. C. (2005). *Smart technology for aging, disability, and independence: The state of the science*. Hoboken, NJ: Wiley-Interscience. doi:10.1002/0471743941

Morell, R. W. (2008). *Older adults, health information, and the world wide web*. Mahwah, NJ: Erlbaum.

Wolfe, D. B., & Snyder, R. (2003). *Ageless marketing: Strategies for reaching the hearts and minds of the new customer majority*. New York, NY: Kaplan.

Zheng, R. Z., Hill, R. D., & Gardner, M. K. (2013). *Engaging older adults with modern technology: Internet use and information access needs*. Hershey, PA: IGI Global. doi:10.4018/978-1-4666-1966-1

KEY TERMS AND DEFINITIONS

S

Accessibility: The goal of making any place or process usable by individuals with disabilities or challenges; in this context, making the Internet and computer technology usable for older adults with physical or cognitive limitations.

Brain Games: Video games, usually online, hypothesized to keep brain functioning at levels of younger individuals.

Gray Gap: Refers to data that indicates that older adults are less likely than those under 55 to be online.

Marketing: Techniques used to encourage consumers to make purchases.

Psychological Well-Being: Usually operationalized in terms of mood or interpersonal functioning; lower wellness would be reflected in indicators such as depression, loneliness, or social isolation.

Silver Tsunami: Phrase referring to the large number of rising seniors in many world economies.

Telehealth: The use of electronic information and telecommunication technologies to support long distance clinical health care and health education.

Issues and Challenges in Enterprise Social Media

Sarabjot Kaur
IIT Kanpur, India

Subhas Chandra Misra
IIT Kanpur, India

INTRODUCTION

Social media is a ubiquitous phenomenon enabling social connections worldwide. This communication platform has given rise to a new social order wherein users can create and disseminate information, opinions, self-created content and knowledge to widespread audience. Business organizations are also catching up with the phenomenon in terms of utilizing the functionality of social media tools for intra firm communication among employees for enterprise wide social networking and collaboration (Fulk & Yuan, 2013). Many organizations are using Enterprise Social Media (ESM) tools for information sharing, crowdsourcing ideas and solving problems, and relationship building (Riemer & Scifleet, 2012). The present article attempts to understand the enterprise social media tools from the perspectives of its functional and interactional aspects along with highlighting the key issues associated with its usage.

BACKGROUND

Social media provides a platform to create and share user generated content along with the functionality of connecting with people. Facebook, Twitter and LinkedIn, wikis, blogs etc. are some examples of popular social media tools. Users can look for people with similar interests, connect with them, create groups and share varied types of content. Web tools similar to popular social media

are being utilized in the enterprise context to foster collaboration and knowledge sharing within the firm. These tools are known by varied names such as Enterprise Social Media (ESM) or Enterprise Social Software Platforms (ESSP).

Defining Enterprise Social Media

Leonardi et al. (2013) define Enterprise Social media (ESM) as “Web-based tools that allow workers to communicate messages with specific coworkers or broadcast messages to everyone in the organization; explicitly indicate or implicitly reveal particular coworkers as communication partners; post, edit, and sort text and files linked to themselves or others; view the messages, connections, text, and files communicated, posted, edited and sorted by anyone else in the organization at any time of their choosing” (p.2). According to Boyd and Ellison (2007), these are “web-based services that allow individuals to construct a public or semi-public profile within a bounded system; articulate a list of other users with whom they share a connection; view and traverse their list of connections and those made by others within the system”(p. 211).

ESM includes tools like as blogs, social network sites (SNSs) and wikis. A lot of different packages are available for use by enterprises. SharePoint by Microsoft, is an online application platform where members can collaborate for team projects, upload and modify files, initiate discussions and work together in a virtual space mak-

ing the collaborative work easier to execute and monitor. Jive is another name in the same list that provides functionality of enterprise wide connections facilitating online sharing and collaboration among employees across the world. Other tools in use are Yammer and Chatter.

Benefit of ESM Tools

ESM provides an ability to maintain social connections along with accessibility to digital content (Kane, 2015). It gives a way of connecting with employees across the organization facilitating interactions and information sharing. Online connections, social interaction and communication patterns can reveal a lot about the various knowledge communities in the organization. Along with providing a functionality similar to traditional media of communication exchange such as email, it provides a way to connect and display the shared information to a preferred set of people who can further benefit from the exchange. The mutual communication partners, shared information and membership of different online groups can help employees understand the prevalent social dynamics along with locating the relevant support for knowledge (Ellison et al., 2015).

Some of the major benefits are listed below.

Acts as an Interactive Platform

Members of an ESM platform can interact in a variety of ways. Subramaniam and Nandhakumar (2013) identified the following ways in which members interact over an ESM platform

1. **Impromptu or Informal:** Ad-hoc problem solving is enabled by collective work of individuals in an Informal way where formal structures do not inhibit communication and interaction. Members come together to work on some common problem that needs immediate attention.
2. **Planned:** This happens in the duration of a planned project where the interaction is

sustained for specific information needs from different members of the community.

3. **Ongoing:** Focused and repetitive interactions help in developing roles as members share knowledge over ESM platform.

Helps to Locate Knowledge Source within Organization

Awareness about the communications of others can help the knowledge seekers understand the potential knowledge sources. An understanding of who knows what and who knows whom helps the employees become aware of the communications among the coworkers and it becomes imperative in eliciting knowledge from them whenever required. This kind of awareness has been referred to as 'ambient awareness (Leonardi & Meyer, 2015) which is facilitated by the use of enterprise wide social media tools. Message transparency and translucent network connections offered by social media (Leonardi, 2015) give employees an edge over the opaque communications handled by conventional technologies.

- Message transparency is the ability of users to see the different conversation threads and the contents generated therein.
- Network translucence or the knowledge of people's communication partners helps in identifying real communication patterns in any organization which may be different from the traditional or expected ones.

Employees can gain an understanding about the nature and kind of projects done by their colleagues along with their individual expertise. The communication threads and patterns can be observed to make sense of individual's tacit knowledge which can be accessed whenever required.

Facilitates Knowledge Sharing Among Employees

Social tools enable knowledge sharing and collaboration which is important for the new and

creative ideas to emerge and a sufficient flexibility of sharing information across the organization provides a foundation for such an activity (Patroni et al., 2015; Subramaniam & Nandhakumar, 2013). Knowledge sharing is an equivocal concept involving a communicative process affected by social dynamics and interpersonal processes. It can be a difficult process in case of geographically dispersed organizations as employees may not be aware of the right source of getting advice and expertise. In such a case, ESM allows employees to connect with their counterparts working in dispersed locations across organization and hence help them to access knowledge from the right source and person. (Ellison et al., 2015).

Facilitates Collaborative Work

ESM tools enable collaborative content generation. It fosters collective intelligence which is a source of new ideas and creative solutions. Capability of sharing knowledge and expertise allows creativity to flourish as it enables the combination of varied perspectives and enables the completion of collective work within shared timeframes (Subramaniam & Nandhakumar, 2013). Virtually co-present actors are bound by focused interaction with other actors in real time to create an interaction order which facilitates collective work (Subramaniam & Nandhakumar, 2013). ESM helps in crowdsourcing solutions for problems by facilitating the provision of diverse information from various sources in the organization (Mäntymäki & Riemer, 2014).

ISSUES IN ESM USAGE AND KNOWLEDGE SHARING

ESM tools act as enabling technology providing a means for the desired ends like collaboration, networking and knowledge sharing across enterprises. However, its actual usage in different contexts may be varied. There are a number of factors that have an impact on knowledge sharing via ESM tools

and subsequent employee performance, some of which are listed below.

Inter vs. Intra Team Usage

Teams have their own peculiar demands w.r.t. task requirements. The kind of information required by the members of a team varies with the context of interaction. Hence, ESM usage in inter-team and intra-team context has different effects on knowledge sharing. The former is more suited for innovative performance associated with finding new knowledge and the latter is more suitable for task performance using knowledge for short term use without any consideration for discovering novel information (Kuegler et al., 2015). Thus, task context is an important aspect to be taken into perspective while assessing the role of ESM in employee performance.

Cultural Norms

Cultural norms of the organization play a significant role in defining the mindset of the employees to collaborate. Competition for better performance and sharing can also motivate employees to read and explore more about a given area. Both collaboration and competition facilitate knowledge sharing leading to faster learning among employees (Patroni et al., 2015).

Tie Strength

Social ties are connections among people which are used to share varied information with tie strength being strong, weak or absent in any given social circle. Granovetter (1973) defined tie strength as “a combination of the amount of time, the emotional intensity, the intimacy (mutual confiding), and the reciprocal services which characterize the tie” (p. 1361). In context of knowledge sharing, social ties and their strength affect the behavior of knowledge seeker and provider. In case where initial ties are strong and knowledge is less complex, there exists a greater possibility that knowledge will be sought

right away (Leonardi & Meyer, 2015). However, the kind of knowledge sought also varies with different social ties. Strong ties have individuals with whom there is information overlap. Hence, weak ties may generate information that is novel (Granovetter, 2005).

Cognitive Overload

Employees all across the organization may exchange different kinds of information interacting over ESM tools. This leads to increased load or volume of communication. Hence, it may be difficult to keep a track of all relevant communication and conversation threads. Also, there is a disadvantage if there are too many conversations going on as people may fight for attention. In such a case, the aim of reducing an employee's knowledge search efforts by developing a contextual understanding of the knowledge sources may not be fully achieved (Leonardi, 2015).

CHALLENGES IN ESM DESIGN AND ADOPTION

Apart from the issues that arise in the usage of ESM tools in particular reference to knowledge sharing and collaborative work, there are certain challenges in its design and adoption which need to be addressed.

ESM Platform Design

Designing a platform requires the consideration of organizational conditions, culture and employee needs. Kane (2015) classified various aspects of platform design for enterprise social media. Different functionalities and design aspects present in the popular social media tools can be utilized in the context of designing a platform for enterprises taking into view their requirements and peculiarities. Some of the interaction and content related aspects identified by Kane (2015) are:

1. Persistence of relationship depicting the network boundaries w.r.t inclusion of various stakeholders in the organization wherein a decision has to be made regarding the inclusion of internal and/or external stakeholders like clients and vendors as well. Including external stakeholders may help in solving shared problems (Kane, 2015).
2. Types of connections which can be either interactions (messaging and chats), proximities (Digital proximities via forums and common online discussion) or flows (movement of information among non-connected users such as in case of Twitter hashtags).
3. Connection capabilities which may offer an option to understand the strength of social ties generated by binary connections and their interaction intensity.
4. Supported content may vary from simple text to audio, video and hyperlinks etc. along with the option of expressing one's opinion on information and posts shared with others.
5. Digital trace enabling a log wherein activities of members on the social media tools can be recorded and summarized for use by other members.
6. Profile authenticity aspect covers the extent of accurate representation of real identity of users who generate and share content thereby increasing its trustworthiness.

ESM tools can be designed in view of existing popular social media tools tailoring their basic functionalities as per organizational needs. Individual usage defines the way technology is understood and utilized in the long run. Users make sense of technology according to their own peculiar needs and may have differentiated usage. Thus, understanding user perspective at various phases in the ESM adoption and implementation can aid better design of the system with consideration of issues like intra-team and inter-team usage. Features like creating filters for eliminating the excessive information, social tie metrics based

on the conversational and exchange frequency etc. may be considered while designing different ESM tools.

ESM Platform Adoption and Usage

Any enterprise system is adopted keeping into view certain key organizational requirements along with assessing the system's compatibility with the idiosyncrasies of organization. In case of ESM tools, factors such as purpose of ESM, participants involved, legal and security risks along with governance and policy mechanisms (Turban et al., 2011) must be considered in adoption decision. Adoption of any enterprise system makes the system accessible and available but there are certain factors that influence its continued usage. On the individual level, self-construal and interdependent tendencies of employees should be taken into view (Liu & Rau, 2014) in respect of ESM tools for intra organizational usage. Employees engage in social sense making of ESM creating their own peculiar communicative practices over such media (Riemer et al., 2012). Utility of ESM can be enhanced by positive organizational factors in terms of conducive knowledge sharing culture and leadership support. Thus, various factors at technological, organizational and individual level are interrelated and their constructive interplay is necessary for ensuring successful assimilation and continued usage of ESM tools.

FUTURE RESEARCH DIRECTIONS

Enterprise social media is a relatively new area for scholarly research in information systems domain. It has previously been explored in context of its knowledge sharing, interactional and communicative consequences mediated by different organizational, technological and individual user related variables. However, a more holistic understanding

can be achieved by adopting a multi-perspective view of the phenomenon. For instance, the effect of organizational support in terms of leadership and supportive team culture may be explored in terms of its effect on knowledge sharing and organizational communication. In addition, user perspective may be understood in greater detail with the help of more qualitative studies that may aid in uncovering design related constraints which can be subsequently worked upon to create efficient tools suitable for organizational requirements. Constant connectivity on ESM tools may lead to a wide variety of information generating attentional conflict (Brzozowski, 2009). Apart from understanding the beneficial consequences of the ESM, future research can explore the effect of this attentional conflict generated due to excessive usage of ESM tools on knowledge sharing behavior of employees and subsequent work performance related consequences.

CONCLUSION

ESM tools have potential to foster collaboration and make information exchange easier. However, the successful assimilation of this technology requires support both at the organizational and individual levels. Technology in itself is a means. The ends to which it is eventually utilized rests on the user. Hence, organization needs to understand various dimensions of adopting and implementing ESM tools in phased manner allowing users to understand and make sense of it. Considering the interplay of various factors at organizational, technological and individual level can help in devising suitable plans for adoption and implementation. Effective usage of ESM can benefit organizations in terms of shared learning and cross-functional expertise in solving problems which can lead to enhanced innovative performance of the organization as a whole.

REFERENCES

- Beck, R., Pahlke, I., & Seebach, C. (2014). Knowledge exchange and symbolic action in social media-enabled electronic networks of practice: A multilevel perspective on knowledge seekers and contributors. *Management Information Systems Quarterly*, 38(4), 1245–1270.
- Boyd, D. M., & Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1), 210–230. doi:10.1111/j.1083-6101.2007.00393.x
- Brzozowski, M. J. (2009, May). WaterCooler: exploring an organization through enterprise social media. In *Proceedings of the ACM 2009 international conference on Supporting group work* (pp. 219–228). ACM. doi:10.1145/1531674.1531706
- Ellison, N. B., Gibbs, J. L., & Weber, M. S. (2015). The use of enterprise social network sites for knowledge sharing in distributed organizations: The role of organizational affordances. *The American Behavioral Scientist*, 59(1), 103–123. doi:10.1177/0002764214540510
- Fulk, J., & Yuan, Y. C. (2013). Location, motivation, and social capitalization via enterprise social networking. *Journal of Computer-Mediated Communication*, 19(1), 20–37. doi:10.1111/jcc4.12033
- Granovetter, M. (2005). The impact of social structure on economic outcomes. *The Journal of Economic Perspectives*, 19(1), 33–50. doi:10.1257/0895330053147958
- Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology*, 78(6), 1360–1380. doi:10.1086/225469
- Huang, Y., Singh, P. V., & Ghose, A. (2015). A structural model of employee behavioral dynamics in enterprise social media. *Management Science*, 61(12), 2825–2844. doi:10.1287/mnsc.2014.2125
- Kane, G. C. (2015). Enterprise social media: Current capabilities and future possibilities. *MIS Quarterly Executive*, 14(1), 1–15.
- Kuegler, M., Smolnik, S., & Kane, G. (2015). Whats in IT for employees? Understanding the relationship between use and performance in enterprise social software. *The Journal of Strategic Information Systems*, 24(2), 90–112. doi:10.1016/j.jsis.2015.04.001
- Leonardi, P. M. (2015). Ambient awareness and knowledge acquisition: Using social media to learn” who knows what” and” who knows whom. *Management Information Systems Quarterly*, 39(4), 747–762.
- Leonardi, P. M., Huysman, M., & Steinfield, C. (2013). Enterprise social media: Definition, history, and prospects for the study of social technologies in organizations. *Journal of Computer-Mediated Communication*, 19(1), 1–19. doi:10.1111/jcc4.12029
- Leonardi, P. M., & Meyer, S. R. (2015). Social Media as Social Lubricant How Ambient Awareness Eases Knowledge Transfer. *The American Behavioral Scientist*, 59(1), 10–34. doi:10.1177/0002764214540509
- Liu, J., & Rau, P. L. P. (2014). Impact of self-construal on choice of enterprise social media for knowledge sharing. *Social Behavior and Personality: An International Journal*, 42(7), 1077–1089.
- Mäntymäki, M., & Riemer, K. (2014). *Information, ideas and input: The value of enterprise social networks*. Paper presented at 25th Australasian Conference on Information Systems, Auckland, New Zealand.
- Patroni, J., Recker, J. C., & Von Briel, F. (2015). How does enterprise social media help retail employees innovate? In *Proceedings of the 23rd European Conference on Information Systems*. Association for Information Systems.

Riemer, K., Overfeld, P., Scifleet, P., & Richter, A. (2012). Eliciting the Anatomy of Technology Appropriation Processes: a Case Study in Enterprise Social Media. In ECIS (p. 134).

Riemer, K., & Scifleet, P. (2012, January). Enterprise social networking in knowledge-intensive work practices: A case study in a professional service firm. In *Proceedings of the 23rd Australasian Conference on Information Systems 2012* (pp. 1-12). ACIS.

Subramaniam, N., Nandhakumar, J., & Baptista John, J. (2013). Exploring social network interactions in enterprise systems: The role of virtual co-presence. *Information Systems Journal*, 23(6), 475–499. doi:10.1111/isj.12019

Turban, E., Bolloju, N., & Liang, T. P. (2011). Enterprise social networking: Opportunities, adoption, and risk mitigation. *Journal of Organizational Computing and Electronic Commerce*, 21(3), 202–220. doi:10.1080/10919392.2011.590109

ADDITIONAL READING

Bradley, A. J., & McDonald, M. P. (2011). *The social organization: How to use social media to tap the collective genius of your customers and employees*. Harvard Business Press.

Butler, M., Butler, D. L., & Chester, J. (2010). *Enterprise social networking and collaboration*. Martin Butler Research Limited.

Keyes, J. (2012). *Enterprise 2.0: social networking tools to transform your organization*. CRC Press. doi:10.1201/b12532

Kilduff, M., & Tsai, W. (2003). Social networks and organizations. *Sage (Atlanta, Ga.)*.

Klososky, S. (2010). *Enterprise social technology: Helping organizations harness the power of social media, social networking, social relevance*. Greenleaf Book Group.

Leistner, F. (2012). *Connecting organizational silos: Taking knowledge flow management to the next level with social media*. John Wiley & Sons. doi:10.1002/9781119205258

Majchrzak, A., Fife, E., Min, Q., & Pereira, F. (2013). *Activating the Tools of Social Media for Innovative Collaboration in the Enterprise*. Springer Science & Business Media.

McAfee, A. (2009). *Enterprise 2.0: New collaborative tools for your organization's toughest challenges*. Harvard Business Press.

White, B., King, I., & Tsang, P. (Eds.). (2011). *Social media tools and platforms in learning environments*. Springer Science & Business Media. doi:10.1007/978-3-642-20392-3

KEY TERMS AND DEFINITIONS

Collective Intelligence: Shared intelligence resulting from group efforts where group members collaborate and compete to reach consensus in respect of mutually relevant problem.

Geographically Dispersed Team: A team whose members are located at geographically distant places and employ collaborative web based tools to accomplish group tasks.

Knowledge Communities: Group of people sharing common interest who come together for generating and exchanging information and ideas on specific issues in an organizational context.

Knowledge Management: The practice of creating, recording, sharing and using knowledge generated within organizations for fulfilling organizational goals.

Sense Making: Process by which people understand and attribute meaning to experiences.

Social Ties: Individual connections maintained by people in their social circle within which they interact and exchange varied kinds of information.

Tacit Knowledge: Knowledge that results from internalized information and experiences and is difficult to explicate in a formal way.

User Generated Content: Content such as text, video, audio, images appearing on blogs, wikis, discussion forums and other social media websites created by users of these online services.

S

Mapping the Dissemination of the Theory of Social Representations via Academic Social Networks

Annamaria Silvana de Rosa

Sapienza University of Rome, Italy

Laura Dryjanska

Sapienza University of Rome, Italy

Elena Bocci

Sapienza University of Rome, Italy

INTRODUCTION

Academic social networks are forms of Internet service, which facilitate the management of relations among scientists, sharing resources for publications, and in some case data, research results and multimedia sources. This chapter concentrates on what are the benefits of academic social networks, how to analyze their impact in spreading knowledge and why they are important. In particular, it aims at mapping the presence of publications using the case study of the theory of social representations in three academic social networks: Academia.edu, ResearchGate and Mendeley.

Academia.edu was founded in September 2008 by Richard Price, who did a PhD at Oxford in philosophy. After finishing his PhD, he founded Academia.edu, which is a platform for academics around the world to connect and share research, which in October 2016 had more than 43 million members. He spotted the need for the platform when doing his PhD. Once freely registered, a user can set his or her profile and fill in their publication list, upload papers and enlist field(s) of interest, finding at the same time researchers with a matching profile. Then, it is possible to follow what academics in the field are working on,

i.e. the latest papers they are publishing, the talks they are giving or the blog posts and status updates they are writing. An important tool that Academia.edu offers is the statistic of one's downloads and page views; it also allows the researcher to know what keywords people use to search for them on Google (Giglia, 2011).

Research Gate, founded in 2008 by physicians Dr. Ijad Madisch (Boston) and Dr. Sören Hofmayer (Berlin), and computer scientist Horst Fickenscher (Berlin), is aimed at creating a working and discovering network among scientists, "Discover", "Communicate" and "Collaborate" are its main purposes (Giglia, 2011). In October 2016 it had more than 11 million members.

London-based *Mendeley*, founded in 2009 by three German PhD students (Victor Henning, Jan Reichelt and Paul Föckler), in October 2016 was used by around 2.5 million researchers worldwide to discover, share and annotate research papers (as a reference manager), and to network and collaborate with other academics (Giglia, 2011). Mendeley has two components: a desktop program and a web-based storage space, which can be used independently or synchronized (MacMillan, 2012).

The main differences among these three academic social networks can be summarized as below:

- *Academia.edu* and *ResearchGate.net* focus more on the producers of research and their networking (main function: “to be contacted”);
- While reference-sharing *Mendeley.com* sites focus on readers, helping users to share and find relevant references for their work (main function: “discover recommended papers”),

One difference still existing in October 2016 is that *Academia.edu* users can post their own papers, but *Mendeley* users can also share others' papers in their *My Library* section (Thelwall & Kousha, 2014).

Overall, it has been found that different disciplines favor different academic social networks and some authors argue that at some point there will be a “winner in the race” (van Norden, 2014). At the moment awareness among scientists of the Academic Social Networks varies, but the most well-known site tends to be Google Scholar, both among natural and social scientists, as stated by van Norden (2014).

If the Personal Social Networks have become exponentially popular among lay people by sharing personal information, snapshots on private life, CV, or even for institutions and companies aimed at their web-marketing; turning to a scholars and researchers target, the academic social networks, born in 2008, have quickly become a fundamental tool to manage, read, share, annotate and cite research papers, among tens of millions of connected users. In the era of bibliometric culture, the academic social networks – moving from the first collaborative aim of global knowledge sharing and co-producing - have also become a tool for the author's popularity. Therefore they have contributed to originate a new disciplinary field called *Altermetrics* (De Bellis, 2009, 2014), aimed at identifying new indicators for measuring their scientific impact.

This chapter first presents the literature review on the topic of Academic Social Networks, which constitutes a fairly new field of study, given their

emergence less than ten years ago. Subsequently, it focuses on the case study of the publications inspired by the theory of social representations and their presence in Academic Social Networks, exploring their characteristics (such as publication year and language) and mapping the geo-cultural contexts of the location of institutions of first authors. Follows the discussion of open networked science and bibliometric culture and possible future research directions, including further statistical analyses of existing material and switching from publications to authors as units of analysis. Finally, the conclusions concern the diffusion of the theory of social representations outside of Europe and the role of academic social networks in this process.

BACKGROUND

Academic Social Networks have become a significant part of informal scholarly communication (Thelwall & Kousha, 2014). According to Hoffman, Lutz and Meckel (2015), they provide channels for quick dissemination of research results and interaction with both peers and lay audiences, while the open access philosophy increases their appeal (Nielsen, 2012). Academic social networks address the researchers' need to ensure that their publications are accessible and visible to a wide audience (Thelwall & Kousha, 2014). They also form a part of academic identity, akin to a business card, or serve as a personal repository (Jordan, 2016).

The established services are constantly changing, hoping to improve user experience, including the design of the user interface, which has to be attractive but also simple, providing a low barrier for newcomers (Goodwin, Jeng & He, 2014). Moreover, academic social networks allow a responsiveness and informality, unlike the formal publishing process (Ovadia, 2014).

Current trends of the research on Academic Social Networks have often concentrated on users, for example Rosenzweig, Grinstein and Ofek

(2016) suggest women and authors from less economically advanced countries are more likely to utilize them. According to Williams and Woodacre (2016), the overarching outcomes of the review of research on academic social networks suggest that it falls into two primary areas: promises and perils. There has also been some research (He & Jeng, 2016), dedicated to the development of scholarly collaboration facilitated by academic social networks. On the other hand, some authors have explored the relationship between social and usage metrics, and traditional bibliometric indicators at author level (Ortega, 2015). This type of research is possible thanks to the fact that academic social networks encompass both levels, suggesting that such metrics could be used as proxies or predictors of research impact (Priem & Hemminger, 2010). Since bibliometrics (traditional techniques for measuring scholarly impact) have become well known for generating conflict and concern (Roemer & Borchardt, 2012), alternatives proposed by academic social networks appeal more attractive. Readership data also constitutes a useful supplementary measure to remedy some limitations of citation analysis across the social sciences and humanities (Mohammadi & Thelwall, 2014).

In the emerging field of research on academic social networks there have been no studies so far dedicated to a specific case of an interdisciplinary theory, comparing publications disseminated via these innovative tools. The method proposed below could be employed in other fields, taking under scrutiny a distinct school of thought.

Serge Moscovici developed the theory of social representations when assessing the process assimilation and transformation of expert knowledge on psychoanalysis among the lay people (1961/1976). According to de Rosa (1994), it can be operationalized on three different levels: a) social representations as *phenomenon* – “ways of knowing” characteristic of social reality that emerge in daily life during interpersonal communication and are directed toward comprehension

and control of the physical-social environment; b) *theory* of social representations – the collection of conceptual definitions, methodological operations and formulation of constructs that have social representations as their object; c) *meta-theory* of social representations – the collection of critical comments, ripostes and comparisons with other theoretical models which emerges from the critical debate on the theory of social representations. For a stock of the wide scientific field developed in more than 50 years since 1961, see de Rosa, (2011, 2013a) de Rosa & d’Ambrosio (2008).

ACADEMIC SOCIAL NETWORKS AND SOCIAL REPRESENTATIONS

This research is part of a wider research program led by de Rosa (2013b) dedicated to the investigation of the spread of the Theory of Social Representations (Moscovici, 1961/1976, 2000; de Rosa, 2013a), in online contexts, specifically in Academic Social Networks, taking into consideration the most important ones: Academia.edu, Research Gate and Mendeley.

This paper use data and meta-data collected by a research tool - the Grid for theoretical Meta-Analysis of Social Representation Literature - created by de Rosa in 1994 and and progressively updated depending on the development of the scientific field and the needs linked to the new forms and practices of sociology of knowledge in the *era of networked science* (Nielsen, 2012). The grid is a fundamental tool of the *SoReCom* “A.S. de Rosa” @-Library (de Rosa, 2016b). For the purpose of the research line presented in this chapter, a specific section has been added to the grid, recording each publication’s presence/absence in each of the three Academic Social Networks.

The research presented in this chapter was conducted in November 2015. Of 9414 publications analyzed, filed in the bibliographic repertories of the online *SoReCom* “A.S. de Rosa” @-Library, 6458 (69% of the total) were not found in any of

the three Academic Social Networks examined, while it has been found the presence of the remaining 2956 articles in at least one or more of them.

Concerning the distribution of the sources in the three Academic Social Networks, the most numerous contributions were found in Research Gate, followed by Mendeley and Academia. Different combinations are presented in Figure 1.

The analysis of the years of publication of literature in Social Representations is organized in decades: from 1952, the date of the article by Moscovici (1952) which can be considered a sign of the embryo-genesis of the theory of social representations (de Rosa, 2011), to the date of this empirical investigation, in November 2015. It consists of 6 decades, through which there is a progressive and increasingly widespread use of the theory of social representations:

- 1952-1961
- 1962-1971
- 1982-1981
- 1992-2001
- 2002-2011
- 2011-2015

In the graphical representation in Figure 2, there seems to be a sharp decline of production in the timeframe 2012-2015, due quite surely to the fact that the amount of time taken into account is limited to four years, and not ten as it is instead for other time periods.

Another criterion that has been adopted to analyze the sample was the language, framing the theory of Social Representations as a multi-cultural, multi-generational and multi-linguistic scientific field. As it was plausible to imagine, English ($f=1614$, 54.60%) is always the predominant language, maintaining this role in all samples analyzed - the set of 2956 items and specific items in each Academic Social Network. It seems to be the hegemonic language of the web. Moreover, English has become over the years the main vehicle of scientific communication, means of shared communication, beyond the country of publication of the native language of the scientists. Follows the order in French ($f=497$, 16.81%), the native language of the theory, born and developed in France, and Spanish ($f=406$, 13.73%) and Portuguese ($f=354$, 11.98%), incontrovertible sign that sees Latin America as the most fertilized scenario.

Figure 1. The frequencies distribution of 2956 items related to social representations in different academic social networks

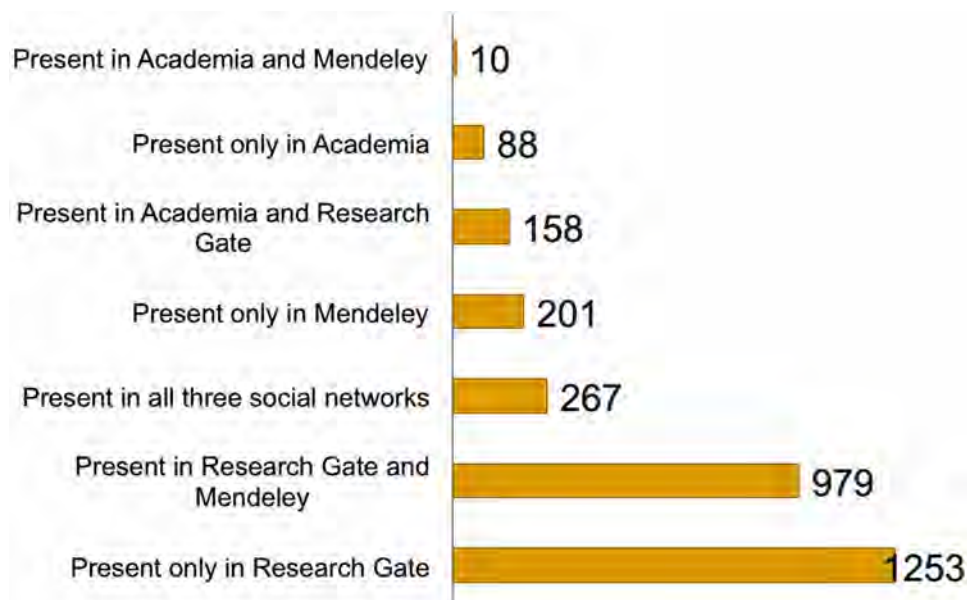
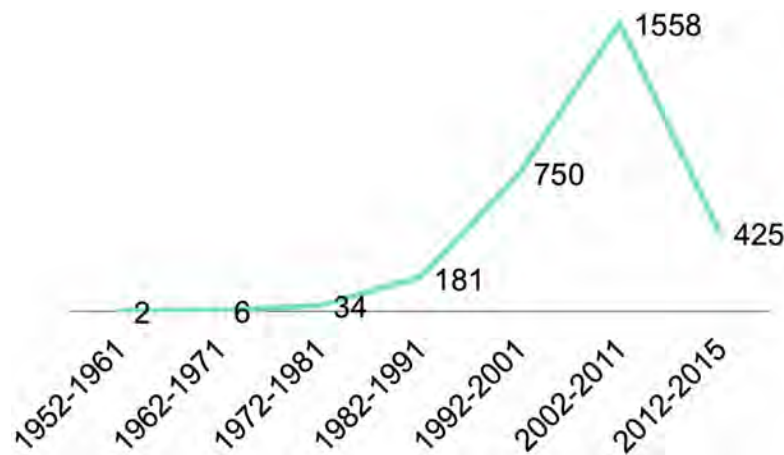


Figure 2. The frequencies distribution of 2956 items related to social representations in academic social networks by decade



The classification according to the Resource Type encompasses multiple sources from which the extracted contributions that draw on the paradigm of social representations, are distinguished according to the type of publication. Various types of publications have been found in the Academic Social Networks, with the overwhelming majority of articles in scientific journals ($f=2307$, 78.04%), followed by book chapters ($f=282$, 9.54%), conference presentations ($f=224$, 7.58%) and books ($f=68$, 2.30%).

Mapping the expansion of the theory and its different spread within Academic Social Networks has always been one of the main goals of the “A.S. de Rosa” @-Library, as well as evaluating its development through various paradigmatic approaches and its dissemination in different thematic domains of applications, identifying precise geographical boundaries. Europe ($f=1856$, 62.79%) tends to be always the most dynamic area in the production and dissemination of scientific literature on Social Representations, followed by Latin America ($f=706$, 23.88%) and North America ($f=232$, 7.85%), when considering the institutions of the authors of publications on social representations that can be found in the Academic Social Networks. Latin America’s importance is in line with the overall trends of literature production

in this field, where in particular Brazilian authors are starting to replace authors from European institutions as leaders (Wachelke, Matos, Ferreira, & Costa, 2015).

In a more specific manner, this contribution examines the country of origin from which the major contributions are present in Academic Social Networks. Through the use of *Tableau software* (<http://www.tableau.com>), it was possible to create a graphical representation that would take into account the diffusion of the Theory of Social Representations in Academic Social Networks from a geographical point of view, for nations and continents. Figure 3 shows how France ($f=517$, 17.49%) occupies the undisputed first place of the greatest visibility in the Academic Social Networks of the literature inspired by the theory of social representations, followed by the United Kingdom ($f=456$, 15.43%), Brazil ($f=390$, 13.19%) and Italy ($f=198$, 6.70%).

The following continent, Latin America, presented in Figure 4 sees as the most prominent countries Brazil, Mexico and Argentina (among others with less frequencies) in this fertilised scenario.

In the remaining continents, the numbers of publications present in Academic Social Networks are progressively lower, which confirms earlier

Figure 3. The frequencies distribution of 1856 items in Europe



results of geo-mapping (de Rosa, in press, forthcoming). Figure 5 groups these “*new emerging scenarios*”.

SOLUTIONS AND RECOMMENDATIONS

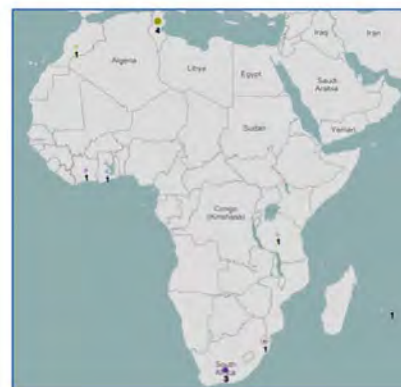
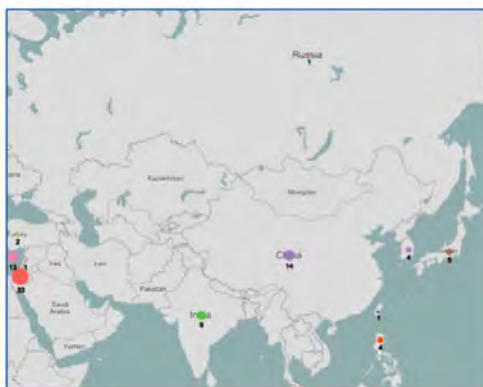
Due to space limitations, it is not possible to go into detail for each of the three Academic Social Networks, where the numerical values differ for each country, but in proportion the trends tend to be the same, positioning Europe as the main continent where the authors who post their papers that use the theory of social representations, followed by Latin America and other scenarios. Certainly, international conferences play a crucial role in the dissemination of the theory (de Rosa, & d'Ambrosio, 2008), although conference presentations are not that frequently posted in the Academic Social Networks.

It is worth noting that the Academic Social Networks also play a crucial role in the debate on the evolution of the bibliometric culture in the recent decades from the scope of *information retrieval* and *science citation* (Garfield, 1955) to Webometrics (De Bellis, 2009) and beyond to *Altmetrics*, based on the transactions of users in the new scenario of the Web 2.0 and the on-growing scenario of *social networking* of especial interest to scholars now undertaking large-scale migration to online publishing and moving toward a universe of web-native communication (de Rosa, 2015a). This will not be irrelevant also for the evaluation of the science impact; in fact, some reasonable doubts about the use of social media in the research evaluation and the need to distinguish between authors' *social popularity* (based on opinion) and *scientific impact* (based on peer reviewed quality filter of scientific facts and results) have been expressed by Moed (2014).

Figure 4. The frequencies distribution of 706 items in Latin America



Figure 5. The frequencies distribution of the remaining items in North America, Oceania, Asia and Africa



The evolutionary scenario of the new bibliometric culture from *Science Citations* to *Sciento-metrics* to *Altmetrics* is widely documented in a multi-disciplinary research field which has moved from information science, informatics, statistics, mathematics, technology, communication and new media studies, but which, due to asymmetric applications in the domain of social sciences and humanities compared natural and applied sciences, has crossed epistemological issues in the history of sciences and their disciplinary policies (de Rosa, 2015a, 2016b).

FUTURE RESEARCH DIRECTIONS

While this chapter concentrated on mapping the geo-cultural trends, the limitation of presented results is the focus on descriptive statistics (e.g. frequencies). Further analyses should be carried out, including correlation of papers featured in Academic Social Networks with their bibliometrics and analysis of variance, taking into account multiple variables, such as whether the papers are empirical or theoretical, what constructs and theories they employ, etc.

A possible future research direction consists of the assessment of the “*personal profiles*” on the academic social networks of authors who publish using the theory of social representations. While the presented research is based on the presence/absence of specific publications, it is possible that not all of their authors actually have a personal profile. On the other hand, it may be interesting to trace who extensive are the personal profiles and how much information the authors are willing to include on them.

Another possible development concerns the willingness of authors to send privately or post full-text publications when prompted by other users, based on the functionalities of different Academic Social Networks that allow such exchange.

Consistently with the main goals of the overall research program aimed at the meta-theoretical analysis (de Rosa, 2013b, 2015a, 2016a), the reconstruction also in the academic social networks of the kind and evolution of inter-individual and inter-institutional co-operations is another goal of the wider meta-theoretical analysis research program. Therefore further efforts will be dedicated to the reconstruct also “within” and “by” the different academic social networks the dynamics of the knowledge epidemiology via the inter-institutional collaborations between authors belonging to institutions in different countries and continents: *who* works with *whom*, (on what) and *where*? Some tools like the analytics of the “followers” (*who is following of followed by whom*) may help to reconstruct the map of the relations among the scientists.

Moreover, a qualitative research line will be integrated with the empirical research line based on the systematic meta-theoretical analysis of the overall scientific production. It will enrich the multi-year collection of interviews already conducted by de Rosa for the main purpose of reconstructing a “*biography of the theory*” based on personal narratives of the protagonists of the theory of social representations (starting from his founder Serge Moscovici) and scholars of different generations about the relevance that this theory had in their intellectual life and the evaluation of the prospective scenario about the theory development. This qualitative extension of the study based on individual interviews will be pursued also with scientists belonging to this scientific field present in the academic social networks, orienting specific questions also aimed at revealing their attitudes towards these new environments and their motivations for using them.

Additionally, this research could be extended to other types of academic social networks aside from the three considered up to date in the So.Re.Com. “A.S. de Rosa” @-Library, which is a flex-

ible instrument, where additional variables can be included in the future.

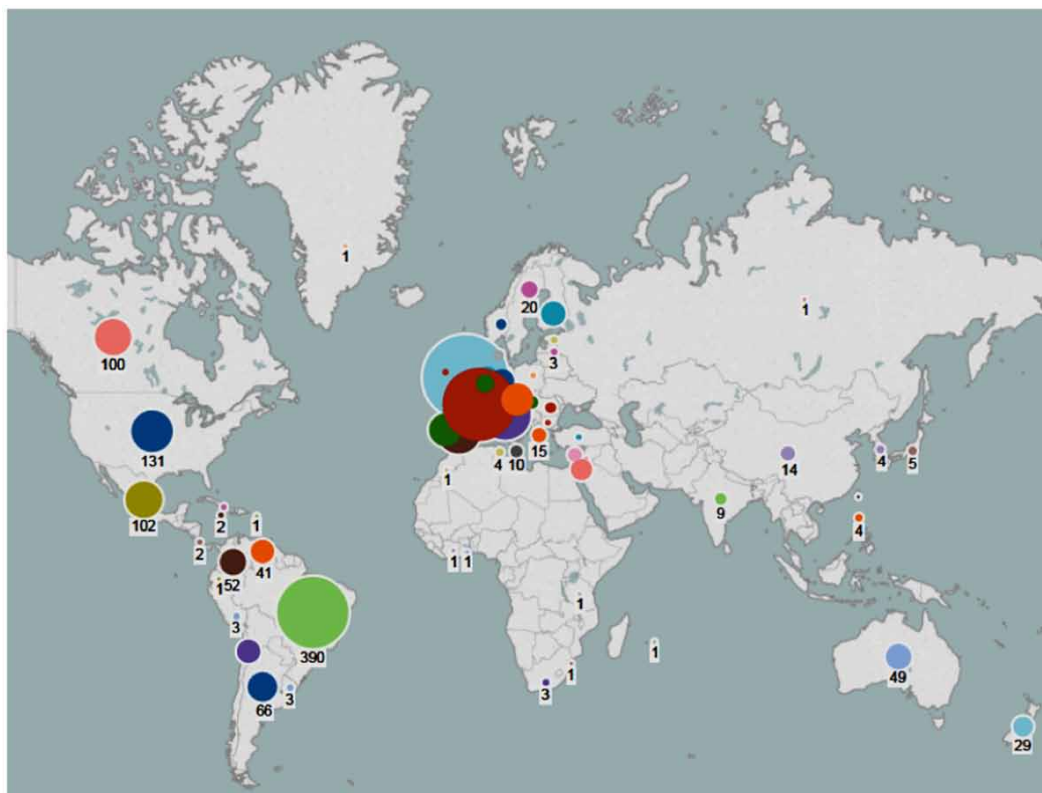
CONCLUSION

In conclusion, it may be stated that this research has shed some more light on the diffusion of the Theory of Social Representations, the origin of a number of works and debates in social psychology, which tends to occupy a central position in the social sciences and that, as now established, meets an interest growing in different countries, in Europe and across the Atlantic. The Academic Social Networks can constitute excellent allies in spreading knowledge and - though they still relatively modestly refer to the field of Social Representations - in time it is likely to expect a progressive, comprehensive and very useful dis-

semination of scientific production using these channels, as demonstrated by literature review. The hard work done by the team of the European/International Joint Ph.D. on Social Representations and Communication Research Centre and Multimedia Lab, founded and directed by Anamaria Silvana de Rosa (2015b) has allowed generations of students and early stage researchers to work on a large sample of items relating to social representations.

The impressive number of contributions from French, English, Brazilian and Italian institutions (among many others worldwide) demonstrates that both European and non-European researchers engaged in the dissemination of the theory are succeeding in efforts to spread it using the Academic Social Networks, identifying a microcosm that mirrors a much larger universe. However, “emerging scenarios” also deserve the attention,

Figure 6. The frequencies distribution of 2956 items related to social representations in academic social networks



because it is very interesting if and how the data described in this paper (combined in Figure 6) will undergoes continuous evolution compared with the set of data gathered in different times through follow-up investigations.

REFERENCES

De Bellis, N. (2009). *Bibliometrics and Citation Analysis: From the Science Citation Index to Cybermetrics*. Lanham, MD: Scarecrow Press.

de Rosa, A. S. (1994). From theory to meta-theory in Social Representations: The lines of argument of a theoretical-methodological debate. *Social Sciences Information. Information Sur les Sciences Sociales*, 33(2), 273–304. doi:10.1177/053901894033002008

de Rosa, A. S. (2011). 50 anos depois: a 'Psicanálise, sua imagem e seu público' na era do Facebook. In A. Almeida, Z. Trindade, F. Santos (Eds.). *Teoria das Representações sociais - 50 anos*, (pp. 491-561). Technopolitik e Centro Moscovi, Brasília: MEMORES/UERJ, REPSO/UFES, LABINT/UFPE.

de Rosa, A. S. (Ed.). (2013a). *Social Representations in the "social arena"*. New York: Routledge.

de Rosa, A. S. (2013b). Taking stock: a theory with more than half a century of history. In A.S. de Rosa (Ed.), *Social Representations in the "social arena"*. New York: Routledge.

de Rosa, A. S. (2015a). The Use of Big-Data and Meta-Data from the So.Re.Com A.S. de Rosa @-Library for Geo-Mapping the Social Representation Theory's Diffusion over the World and its Bibliometric Impact. *9th International Technology, Education and Development Conference, INTED2015*. Madrid: INTED Publications.

de Rosa, A. S. (2015b). The European/International Joint PhD In Social Representations and Communication: a Pioneer Case of Triple "I" (International, Interdisciplinary, Intersectoral) Networked Joint Doctorate. *9th International Technology, Education and Development Conference*. Madrid: INTED Publications.

de Rosa, A. S. (2016a). Mise en réseau scientifique et cartographie de la dissémination de la théorie des représentations sociales et son impact à l'ère de la culture bibliométrique. In G. Lo Monaco, S. Delouvée, & P. Rateaux (Eds.), *Les représentations sociales* (pp. 51-68). Brussels, Belgium: Editions de Boeck.

de Rosa, A. S. (2016b). The So.Re.Com. "A.S. de Rosa" @-library: Mission, Tools and Ongoing Developments. In M. Khosrow-Pour (Ed.), *Encyclopedia of Information Science and Technology* (4th ed.). Hershey, PA: IGI Global.

de Rosa, A. S., & d'Ambrosio, M. L. (2008). International conferences as interactive scientific media channels: the history of the Social Representations theory through the eight editions of ICSR from Ravello (1992) to Rome (2006). *Rassegna di Psicologia*, 2, 153-207.

de Rosa, A. S., & Dryjanska, L. (2017 forthcoming). Social Representations, Health and Community. Visualizing selected results from the meta-theoretical analysis through the "Geo-mapping" technique. In A. Oliveira & B. Camargo Vizieu (Eds.), *Representações sociais do envelhecimento e da saúde*. João Pessoa, Brasil: Editora Universitária da UFPB.

de Rosa, A. S., Dryjanska, L., & Bocci, E. (2017). The impact of the impact: Meta-Data Mining from the SoReCom "A.S. de Rosa" @-Library. In M. Khosrow-Pour (Ed.), *Encyclopedia of Information Science and Technology* (4th ed.). Hershey, PA: IGI Global.

- Garfield, E. (1955). Citation indexes for science. A new dimension in documentation through association of ideas. *Science*, 122(3159), 108–111. doi:10.1126/science.122.3159.108 PMID:14385826
- Giglia, E. (2011). Academic social networks: It's time to change the way we do research. *European Journal of Physical and Rehabilitation Medicine*, 47(2), 345–349.
- Goodwin, S., Jeng, W., & He, D. (2014). Changing communication on ResearchGate through interface updates. *Proceedings of the American Society for Information Science and Technology*, 51(1), 1–4. doi:10.1002/meet.2014.14505101129
- He, D., & Jeng, W. (2016). Scholarly Collaboration on the Academic Social Web. *Synthesis Lectures on Information Concepts, Retrieval, and Services*, 8(1), 1–106. doi:10.2200/S00698ED1V01Y-201601ICR047
- Hoffmann, C. P., Lutz, C., & Meckel, M. (2015). A relational altmetric? Network centrality on ResearchGate as an indicator of scientific impact. *Journal of the Association for Information Science and Technology*, 65(4), 765–775.
- Jordan, K. (2016). Digital scholarship and the social networking site: How academics conceptualise their networks on academic social networking sites and Twitter. *Selected Papers in Internet Research, proceedings of the 17th Annual Conference of the Association of Internet Researchers*. Academic Press.
- MacMillan, D. (2012). Mendeley: Teaching scholarly communication and collaboration through social networking. *Library Management*, 33(8/9), 561–569. doi:10.1108/01435121211279902
- Moed, H. F. (2014). *New development in the evaluation of HSS research, Workshop Internazionale “La valutazione della ricerca nelle Humanities and Social Science”*. ANVUR (National Agency for the Evaluation of Universities and Research Institute). Retrieved from http://www.anvur.org/index.php?option=com_content&view=article&id=693:workshop-internazionale-la-valutazione-della-ricerca-nelle-humanities-and-social-sciences-it&catid=48&Itemid=363&lang=it
- Mohammadi, E., & Thelwall, M. (2014). Mendeley readership altmetrics for the social sciences and humanities: Research evaluation and knowledge flows. *Journal of the Association for Information Science and Technology*, 65(8), 1627–1638. doi:10.1002/asi.23071
- Moscovici, S. (1952). Premiers résultats d'une enquête concernant la psychanalyse. *Revue Française de Psychanalyse*, 3(16), 386–415. PMID:13027870
- Moscovici, S. (1961/1976). *La psychanalyse son image et son public. Etude sur la représentation sociale de la psychanalyse*. Paris, France: Presses Universitaires de France.
- Moscovici, S. (2000). *Social Representations: Explorations in social psychology*. Cambridge, UK: Polity Press.
- Nielsen, M. (2012). *Reinventing discovery: The new era of networked science*. Princeton, NJ: Princeton University Press.
- Ortega, J. L. (2015). Relationship between altmetric and bibliometric indicators across academic social sites: The case of CSICs members. *Journal of Informetrics*, 9(1), 39–49. doi:10.1016/j.joi.2014.11.004

Ovadia, S. (2014). ResearchGate and Academia.edu: Academic social networks. *Behavioral & Social Sciences Librarian*, 33(3), 165–169. doi:10.1080/01639269.2014.934093

Priem, J., & Hemminger, B. H. (2010). Scientometrics 2.0: New metrics of scholarly impact on the social Web. *First Monday*, 15(7). doi:10.5210/fm.v15i7.2874

Roemer, R. C., & Borchardt, R. (2012). From bibliometrics to altmetrics. A changing scholarly landscape. *College & Research Libraries News*, 73(10), 596–600.

Rosenzweig, S., Grinstein, A., & Ofek, E. (2016). Social network utilization and the impact of academic research in marketing. *International Journal of Research in Marketing*, 33(4), 818–839. doi:10.1016/j.ijresmar.2016.02.002

Thelwall, M., & Kousha, K. (2014). Academia.edu: Social network or Academic Network? *Journal of the Association for Information Science and Technology*, 65(4), 721–731. doi:10.1002/asi.23038

Thelwall, M., & Kousha, K. (2015). ResearchGate: Disseminating, communicating, and measuring Scholarship? *Journal of the Association for Information Science and Technology*, 66(5), 876–889. doi:10.1002/asi.23236

Van Noorden, R. (2014). Online collaboration: Scientists and the social network. *Nature*, 512(7513), 126–129. doi:10.1038/512126a PMID:25119221

Wachelke, J., Matos, F. R., Ferreira, G. C. S., & Costa, R. R. L. (2015). Um panorama da literatura relacionada às representações sociais publicada em periódicos científicos. *Temas em Psicologia*, 23(2), 293–308. doi:10.9788/TP2015.2-05-Eng

Williams, A. E., & Woodacre, M. A. (2016). The possibilities and perils of academic social networking sites. *Online Information Review*, 40(2), 282–294. doi:10.1108/OIR-10-2015-0327

ADDITIONAL READING

Boston, I. (2009). Racing towards academic social networks. *On the Horizon*, 17(3), 218–225. doi:10.1108/10748120910993240

De Bellis, N. (2014). *Introduzione alla bibliometria: dalla teoria alla pratica*. Milano: Associazione Italiana Biblioteche.

de Oliveira, J. P. M., Lopes, G. R., & Moro, M. M. (2011). Academic social networks. In *Advances in Conceptual Modeling. Recent Developments and New Directions* (pp. 2-3). Springer Berlin Heidelberg. doi:10.1007/978-3-642-24574-9_2

Duong, T. H., Nguyen, N. T., & Jo, G. S. (2010). Constructing and mining a semantic-based academic social network. *Journal of Intelligent & Fuzzy Systems*, 21(3), 197–207.

Jeng, W., He, D., & Jiang, J. (2015). User participation in an academic social networking service: A survey of open group users on Mendeley. *Journal of the Association for Information Science and Technology*, 66(5), 890–904. doi:10.1002/asi.23225

Jordan, K. (2014). Academics and their online networks: Exploring the role of academic social networking sites. *First Monday*, 19(11). doi:10.5210/fm.v19i11.4937

Kirkwood, K. (2010). The SNAP Platform: Social networking for academic purposes. *Campus-Wide Information Systems*, 27(3), 118–126. doi:10.1108/10650741011054429

Kroski, E. (2007). The social tools of Web 2.0: Opportunities for academic libraries. *Choice (Chicago, Ill.)*, 44(12), 2011–2021. doi:10.5860/CHOICE.44.12.2011

Taranto, G., & Abbondanza, M. (2009). Powering Students Up. *Principal Leadership*, 10(4), 38–42.

Taranto, G., Dalbon, M., & Gaetano, J. (2011). Academic social networking brings Web 2.0 technologies to the middle grades. *Middle School Journal*, 42(5), 12–19. doi:10.1080/00940771.2011.11461778

Zaugg, H., West, R. E., Tateishi, I., & Randall, D. L. (2011). Mendeley: Creating communities of scholarly inquiry through research collaboration. *TechTrends*, 55(1), 32–36. doi:10.1007/s11528-011-0467-y

KEY TERMS AND DEFINITIONS

Academia.edu: Founded in September 2008, it is an academic social network for researchers around the world to connect and share research, which currently has more than 30 million members.

Academic Social Networks: Online platforms aimed at free exchange of publications, resources and information in the realm of academia, open to students, researchers and professors. Once freely registered, a user can set his or her profile and fill in their publication list, upload papers and enlist field(s) of interest, finding at the same time researchers with a matching profile.

Altmetrics: Metrics based on social media (like blogs, Twitter, and Mendeley) which inform broader and faster measures of impact of a scientific publication, complementing traditional citation metrics.

European/International Joint PhD on Social Representations and Communication: An international joint doctorate led by the Sapienza University of Rome, dedicated to the research training in the field of social representations (<http://www.europhd.eu>).

Geo-Mapping: Technique to visualize data from different geo-cultural contexts or specific geographic locations that takes into account the cultural characteristics of inhabitants, which demonstrates the diffusion of the theory.

Mendeley: Founded in 2009 and used by around 2 million researchers worldwide to discover, share and annotate research papers (as a reference manager), and to network and collaborate with other academics.

ResearchGate: Founded in 2008, aimed at creating a working and discovering network among scientists, “Discover”, “Communicate” and “Collaborate” are its main purposes, in January 2016 it had more than 8 million members.

So.Re.Com. “A.S. de Rosa” @-Library: A multi-purpose web-platform for integrating scientific documentation, networking and training in the field of Social Representations and Communication (So.Re.Com.) (de Rosa, 2016b).

Social Representations: A construct developed by Serge Moscovici (1961/1976), which – according to de Rosa (1994) - can be operationalized as “ways of knowing” characteristic of social reality that emerge in daily life during interpersonal communication and are directed toward comprehension and control of the physical-social environment.

The NetLab Network

S

Dimitrina Dimitrova*York University, Canada***Barry Wellman***NetLab Network, Canada*

INTRODUCTION

The NetLab Network (“NetLab”) is an interdisciplinary scholarly network studying the intersection of social networks, communication networks, information networks, and computer networks. NetLab gets its identity and zeitgeist from its distinctive subject matter, multi-disciplinary nature, and the way in which it functions as a social network.

As a network in its own right, NetLab comprises shifting teams, spatially dispersed relationships, and permeable boundaries. Its members have come from many disciplines: Communication Science, Computer Science, Geography, Information Science, Management Science, and Sociology. NetLab has been inclusive in academic status, including faculty, graduate, undergraduate, and high school students. Although centered at Toronto, Canada, NetLabbers are elsewhere in Canada, as well as Australia, Chile, China, England, Israel, Italy, Japan, Norway, Singapore, and the United States. Connecting them is a shared sensibility of interpreting behavior from a social network perspective rather than seeing the world as composed of bounded groups, tree-like hierarchies, or aggregates of disconnected individuals.

Not only ideas connect the network. NetLab is an informal network of collaborators - faculty and students - that function as a community of practice: a self-selected, self-organizing, informal group of collaborators who solve problems together and learn from each other. In addition to congruent intellectual perspectives, frequent communication and a culture of inclusiveness and

mutual supportiveness also connect NetLabbers. With its paramount interest in social networks, as well as its collaborative focus, interdisciplinary nature, remote team members, and partnerships with government and industry, NetLab exemplifies key trends in research (Wellman, et al, 2016). It is important not only for what it does but for also how it achieves it.

This chapter summarizes and updates an earlier description published in the *Encyclopedia of Cyber Behavior*. As the space allotted for this chapter is less than half of the original chapter, we emphasize recent research and refer readers to the earlier version for fuller discussions, citations and references (Dimitrova & Wellman, 2012).

GUIDING PRINCIPLES

NetLab research has been informed by a set of guiding principles:

1. *The world is composed of networks, not groups.* People function more as individuals connected via partial memberships in multiple networks and less as people embedded in tightly-bounded, densely-knit, settled groups.
2. Many people meet their social, emotional, and economic needs by tapping into multiple, loosely knit networks of diverse associates rather than relying on tight connections to a relatively small number of core associates.
3. The social structures people are in largely determine the operation of two-person re-

DOI: 10.4018/978-1-5225-2255-3.ch612

- lationships: it is sociology, not psychology. Ties are usually asymmetrically reciprocal, differing in content and intensity.
4. Ties link network members indirectly as well as directly.
 5. Asymmetric ties and complex networks differentially distribute scarce resources.
 6. Information and communication technologies (ICTs) are usually extensions and enhancers of ongoing relationships. Few people have most of their ties in segregated virtual worlds.
 7. Households have become more networked, with ICTs keeping mobile spouses and their children in contact.
 8. At work, less-formal, fluctuating and specialized peer relationships are common, and the benefits of boss/subordinate hierarchical relationships are less obvious. The organization of work has become more spatially distributed, with ICTs connecting people, and appreciable numbers working at home full or part-time.
 9. As the dividing line between work and home has weakened, so has the more general boundary between the private and public spheres of life. In the less hierarchical and less bounded networked environment where expertise is more in dispute than in the past and where relationships are more tenuous, there is more uncertainty about whom and what information sources to trust.
 10. Social movements arise out of both existing social networks and more organized groups; they rarely are disconnected bunches of alienated individuals.

BACKGROUND

Although social networking services such as Facebook, LinkedIn, and Twitter have highlighted the connection between social networks and technology, NetLabbers have looked at social

networks – spatially distributed and sparsely knit – for decades. The International Network for Social Network Analysis was founded at the University of Toronto in 1977, and many local scholars have used a common network analytic approach in a variety of substantive areas and topics. Some focused on large-scale structures and their implications while others focused on interpersonal relations.

In the early 2000s, a cohesive set of scholars started examining how the internet fits into community and family life, friendships, civic involvement, and health practices. The network jelled, linked not only by overlapping research interests but also by a propensity to collaborate and help each other. They had an understanding of their overlapping intellectual pursuits and the connections among them, but they lacked a name and a distinct identity. Creating NetLab formalized existing practices of collaboration and gave an identity to an already well-functioning network of faculty and students.

The creation of NetLab – as a brand independent of any scholarly discipline – also facilitated the involvement of scholars from a variety of disciplines. Then and now, the boundaries of the network have been flexible: Faculty members are engaged to a different degree depending on the mix of projects they have at a particular moment in time. Students come in and out, graduate from university, or finish internships. Members move away or change interests. Yet the network remains: Former students come back for a visit or engage as collaborators; network members collaborate and ask for help with relevant references, a survey protocol, or an ethics submission.

Branding the network as “NetLab” provided a useful identity. It gave visibility to the University of Toronto among the global network of social network scholars; it provided an easy way for other scholars to connect with network analysts. It provided a quasi-formal identity to scholars, the general public, the media, and government bodies, NGOs, and corporations interested in the

social network approach. Former students at NetLab have graduated to launch their own research programs that build on and expand on NetLab's research concerns. Some continue their interests in social capital or the way media is involved in social networks.

All this happened with a loosely bounded and fuzzily defined informal network. So NetLab became the name and the brand. It reflects the combination of serious scholarship and joyous collaboration. It reflects NetLab's openness to studying all sorts of networks: social, communication, information, and computer. No longer based at the University of Toronto, it is today "The NetLab Network," but informally still "NetLab".

COMPUTER NETWORKS MEET SOCIAL NETWORKS

The essence of NetLab is *social* networks, not *computer* networks. But, starting in 1990, a series of collaborative projects focused on the ways people used ICTs for work and community. Toronto computer scientists in the Cavecat and Telepresence projects were attracted to NetLab's focus on work and leisure communities that transcended the traditional local proximities of work groups and neighborhoods. Several projects on the remote collaboration at work highlighted the importance of social context. For instance, a study of home-based teleworkers found that work tasks and supervision shaped their patterns of mediated and face-to-face communication. Similarly, the ways in which employees of a distributed organization used a pioneering desktop video conferencing system reflected their need of autonomy.

Significant efforts concentrated on how the emerging internet affected scholarly networks and community. Digital divide studies addressed an early concern: which kinds of people were actively using the internet? Scholarly network studies discovered that friendship was as important for the connectivity of white-collar employees as their collaborative work, and that weak ties

among scholars were especially important for increasing the pool of available advice givers. The "Netville" community study showed that rather than destroying neighboring, the internet increased both local and long distance connectivity. These studies also had the effect of shifting research away from experimental groups and phenomena such as telework or video conferencing to looking at how ordinary people incorporated ICTs into their lives.

NetLab's projects in the early 2000s solidified its theoretical orientation, substantive focus, methodologies, and interdisciplinary orientation. Among the key ideas was the understanding that computer networks link people as well as machines. In other words, computer networks are social networks. At a time when the analysis of online relations tended to focus on ties between two people, NetLabbers emphasized the embeddedness of such ties in networks connecting people in-person as well as online. This interpretation had major implications for the study of ICTs: it meant that NetLabbers could fruitfully use the same intellectual apparatus for studying the internet that they had used to study work, community and households. This background enabled NetLab researchers to wade into major debates about the significance and implications of internet and mobile technologies on interpersonal relations, communities and organizations. Instead of armchair theorizing, they were armed with a broad theoretical approach, systematic methods, and a knowledge base of research lore. A continuing research thread in findings is: "the more, the more, the more." That is, the more social ties people have, the more they use digital media, and the more social support they exchange.

The central presence of social scientists in NetLab provided an emphasis on social context and linked its research to broad societal trends and the history of science and technology. This allowed the group to avoid the common pitfalls of early research on ICTs. Instead of implicit technological determinism – where computerization directly determined behavior -- NetLab researchers saw technology as providing social

affordances, i.e. opportunities and constraints for social relations. The broader social context of NetLab's studies allows scholars to question ahistorical assumptions about ICTs. Rather than assuming that ICTs have no precedent in the past, NetLabbers have shown how debates about whether ICTs enhance or diminish community are continuations of centuries-old debates about the impact of the industrial-bureaucratic revolution on community. Instead of pedantically arguing about the theoretical effects of ICTs, NetLabbers have collected systematic evidence from surveys, in-depth interviews, and ethnographies. Instead of adopting parochial views that ICTs could be understood in isolation from face-to-face and phone communication interaction, and that only online phenomena were relevant for understanding the internet, NetLab researchers looked for the interplay between communication online and offline.

Such an approach has coupled with an appreciation of the social affordances of a series of successively developed technologies. NetLab research started with email and early collaborative technologies. Later studies turned to digital media, both interpersonal and social (e.g., Takhteyev, et al., 2012; Gruzdz, et al., 2012). Two key phenomena discovered in the research were *local virtualities* – where physically proximate people communicate online and *virtual localities* – where physically dispersed networks form communities online.

Researchers at NetLab have situated analyses in the context of historical trends and broader social phenomena. This leads them away from simplified utopian/dystopian interpretations of the affordances of ICTs. NetLabbers have shown that digital media have become incorporated in everyday life. They not only give people speed in data-heavy exchanges but also enable them to stay always connected and to make their communication personalized and portable. For example, NetLab researchers demonstrated that the use of ICTs could support both strong and weak ties, and that ICTs rarely replaced face-to-face interactions but were used instead to communicate in-between

face-to-face contacts (Wang & Wellman, 2010; Rainie & Wellman, 2012). Nor have ICTs decreased the importance of proximity or precluded travel: Rather, they have been incorporated in the variety of tools people use to maintain their networks (Dimitrova, et al., 2015). Online connectivity may even increase face-to-face interactions and community engagement. People stay both locally embedded and globally connected, and have more close relationships, support, trust in others, and civic engagement.

Anabel Quan-Haase's book, *Technology and Society: Social Networks, Work, and Inequality* (2015), presents these and other findings in depth, examining the places where technology and society intersect; connecting technology to issues of social networks, communication, work, power, and inequality. Focusing on Canadian society, she provides an overview of how ICTs are changing peoples' perceptions of themselves and their relationships.

COMMUNITY AND HOUSEHOLD STUDIES

“Connected Lives” and “Networked Individuals”

To understand the networked nature of community, NetLabbers have studied a residential area of Toronto – East York – via surveys and in-depth interviews in four waves: 1969, 1979, 2004-2005, with the latest study starting in 2012. The most recent wave focuses on how the internet fits in with friendship, community, social capital, domestic relationships and civic involvement. This research has found that people happily integrate the internet into their personal networks: at home, at work, and with their families. There is strong evidence that the internet is promoting community, and that it is not a separate, alienating world. Communities are no longer bounded groups but are partially connected networks whose ties reach

out into different spheres, linking people near and far. A culture of frequent connections has permeated families, keeping in touch during the day via email, texting and mobile phones. Many East Yorkers use digital media to work at home, either part-time or full-time (Kennedy, et al. 2011; Rainie & Wellman, 2012; Mathias, 2013).

This research is congruent with a national U.S. survey showing that the number of close friendships had increased in the United States between 2002 and 2007. Belying fears that internet use had caused the atrophying of relationships, the analysis revealed that not only do heavy internet users have more friends than light or non-users, but also that heavy users have experienced the greatest growth in the number of friends during that time period (Wang & Wellman, 2010).

Elsewhere in the world, NetLab's Japanese research in the Yamanashi prefecture depicted how young adults used web-enabled phones instead of personal computers almost a decade earlier than their North American counterparts. The findings indicate a division of ICT use in Japan, with smartphone messages used to contact socially-close supportive relationships and email used to expand and diversify messages (Miyata, et al., 2008).

NetLab's East Asian research suggests that "networked individualism" is more than a Western world phenomenon. The researchers find similar situations in Japan, China, Korea, and Singapore: multiple partial communities, using digital media, trains and planes to connect over long distances. Key differences between North American and East Asian networked individuals are the heightened importance of kinship ties and age seniority (Chua, 2011; Chua & Wellman, 2015, 2016).

Another longstanding divide has been the elderly who are less likely to use ICTs. Neves, et al. (2013) in Lisbon and Quan-Haase, et al. (2016, 2017) in East York suggest this gap is lessening. The widespread need for ICTs means the young are teaching the old, and the formerly young are growing older and bringing digital literacy with them.

SOCIAL CAPITAL

Many NetLab members are interested in interpersonal social capital: network ties that may be mobilized as resources. Keith Hampton, et al. (2016) show how ICTs provide heightened awareness of life events in the lives of both close and more distant acquaintances. Awareness of undesirable events can lead to stress. Women tend to report greater stress than men and from a wider range of events. Hampton (2016) argues that ICTs affordances of persistent contact and pervasive awareness are ushering in changes to the structure of community that may revive the constraints and opportunities of premodern communities. Through the ambient, lean, asynchronous nature of ICTs, awareness supplements surveillance with the informal watchfulness typified in preindustrial community.

Such results fit with experimental work focusing on how individuals use ICTs to maintain and build their personal networks in their daily lives. Thus, smartphones both enable or hinder the transfer of information and support within social networks. For example, smartphones can mitigate tele-cocooning by stimulating communication between weak ties (Boase & Ling, 2013; Kobayashi and Boase, 2015). At the same time, ICTs are associated with an increased use of public space, with people being more connected than a generation ago (Hampton, et al., 2015).

Other NetLab research looks at networked individuals using their ties for advancement. Indigenous entrepreneurs use various forms of social and cultural capital in their networks to mobilize resources (Côté, 2013). Marin (2012) reveals how information holders choose to share or withhold job information from network members who they believe to be suitable for known job openings. They are more likely to share job information with strong ties and when job openings are located in closed labor markets.

Networked information pervades health awareness and care. Gruzdt & Haythornthwaite (2013) described a community in which both formal

health providers and informal advisors used ICTs to discuss health topics. By contrast, those who seek alternative health care, such as naturopathy, rely on ties with friends and relatives for advice and connections and often do not tell their physicians (Wellman & Kelner, 2015).

NETWORKED STRUCTURES

Networked Scholars

The NetLab's Network Assessment and Validation for Effective Leadership (NAVEL) team spent more than five years participating in and simultaneously studying a large multidisciplinary and multi-institutional Canadian research network, GRAND. The project examined how scholars from computer science, social science, health sciences, and the humanities collaborated over five years. The study sheds light on research collaboration and demonstrates how networked organizations actually operate (Dimitrova, et al., 2015a, 2015b). While networked organizations are especially common in research, most discussions are more cheerleading and anecdote than analytic research.

NAVEL's research focused on how new digital technologies and network structures affect team performance and innovation. Researchers found a network whose uneven connectivity was associated with formal organizational position, academic seniority, and disciplinary background. Leading researchers--usually older and more senior--had twice the ties of their junior collaborators; computer scientists were more connected than academics in humanities and social sciences, and cross-disciplinary collaboration linked functionally close discipline (Dimitrova, et al. 2013; 2015c). To connect with each other, GRAND researchers relied mostly on email and in-person chats. Despite ICTs' facilitating of long-distance ties, researchers tended to connect and work with nearby scholars (Hayat & Mo, 2015; Wellman, et al., 2016).

As a longitudinal study, NAVEL offers insights on the evolution of research networks (Dimitrova, et al. 2015c; Hayat & Mo, 2015). Cross-disciplinarity, geographic dispersal, and institutional diversity fostered the creation of more volatile and easily dissolved ties. While networks did not substantially change, ties were frequently added and dropped—especially ties across disciplines, institutions, and locations. The researchers liked the intellectual challenges of cross-disciplinary collaboration and actively created new ties, but did not always sustain them. Both the scholars' social status in GRAND and the structural holes in their networks shaped their exchanges of advice and their collaborations. We caution that because such multidisciplinary, multilevel networks are more complex, they can be slower to produce traditional academic outputs and they cannot be evaluated only by traditional measures.

NAVEL's results echo the findings of two earlier NetLab studies of collaborative research networks of academic, government, and private sector participants that highlighted the continuing relevance of disciplinary boundaries (Dimitrova & Koku, 2009). Nor are such networking dynamics confined to the scientists. Another NetLab study of digital humanities suggested that in the humanities, large-scale collaborative networks had greater network density and integration, without necessarily increasing the level of in-depth collaboration typically found in the sciences (Quan-Haase, et al. 2015).

Networked Learning

Just as communities used to be (mostly) bound up in neighborhoods, formal education used to be bound up in school classrooms. Yet digital media and networked individualism are enabling school-agers and lifelong learners to obtain knowledge from a variety of sources, from Facebook and Google to massively open online courses. Haythornthwaite, et al. (2016) are leading the explorations of this area.

Networked Influence

In a networked society, social influence has become networked influence: it occurs in social networks and propagates through online communication networks (Gruzd & Wellman, forthcoming 2016). Working from this premise, Gruzd's coordinate Social Media Lab finds that while there are some pockets of political polarization on the Twitter social media platform, it also facilitates open, cross-party, and cross-ideological discourse (Gruzd, et al., 2016; Gruzd & Wellman 2016; see also Gruzd, et al., 2012; Miyata, et al., 2015).

Transnational Entrepreneurs

The Information Technologies and Transnational Entrepreneurship Project studied how ICTs and planes allow global connectivity. The "net and jet" helped Chinese immigrants to stay connected in North America and in their homeland, relying on glocalized networks and the internet to engage in transnational entrepreneurship (Chen & Wellman, 2009). Chen currently collaborates with University of Texas colleagues analyzing the relation of social media, social networks, and global media flows (Chen & Reese, 2015).

DEVELOPING NETWORK METHODS

Advanced methods have been a continuing thread in studying networked relationships. Bonnie Erickson pioneered the "position generator" to measure the resources people are people linked to and how differences in network contacts with different jobs have life consequences (Lin & Erickson, 2008).

Anatoliy Gruzd is developing automated text mining and visualization tools for uncovering and representing online social networks (Gruzd & Wellman, 2014). He and the associates at his Social Media Lab have put these tools to good use studying how academics, professionals, and ordinary people use social media and how social media are implicated in knowledge exchange,

collaboration, and social influence processes. The continuing development of tools and methods enables and, in turn, is reinforced by the expanding research agenda of the Social Media Lab.

Concomitantly, Boase and Ling (2013) have incorporated digital trace data into their project designs, merging it with more traditional survey and interview data. Marin and Hampton (2007) have shown ways of simplify data collection using name generators while ensuring reliability. Bernie Hogan is developing novel data collection techniques that blend qualitative and quantitative approaches to understand how people perceive and act on social networks. He developed participant-aided sociograms with the Connected Lives team and later turned to online audit studies (Hogan & Berry, 2011). He has also blended this empirical work with a theoretical expansion of the social affordances concept (Hogan & Wellman, 2014). Hogan (2015) argues that instead of alphabetical or chronological order, information providers should use the logic of machine learning with data-as-graphs to train multidimensional systems.

The complex reality of the scholarly network studied by NAVEL enabled the development of multilevel, multimember modeling and sequencing analyses. These are useful tools for longitudinal studies of networked organizations and networked workers who are partial members of multiple teams (Mo, 2012; Mo & Wellman, 2016).

CONCLUSION

As a community of practice, the NetLab Network has had two main achievements. First, its researchers have made substantive contributions to the issues at the intersection of social networks, communication and computer networks. The accumulation of tacit and explicit interdisciplinary knowledge at the NetLab Network not only lends depth and sophistication to research results; it has reached a stage which enables the development of theory.

Second, NetLab has demonstrated that its model of scholarly collaboration works. Its members practice, analyze, and preach being networked. Their collaborative relationships span both disciplines and academic status and have connected a diverse network of scholars. Their informal, but serious collaboration enables bouncing off ideas and has proven especially conducive to novelty and creativity.

A recent book – *Networked: The New Social Operating System* (Rainie & Wellman, 2012) -- is an exposition of much of NetLab's intellectual capital and research lore. It brings together different strands of the work done by NetLabbers for the past four decades and adds other relevant contributions about social networks, interpersonal relations, the family, work, information, and creativity. The key thread running through the book is the "Triple Revolution": the turn to social networks, the internet and mobile connectivity and the way these revolutions affect everyday life. The authors contend that the Triple Revolution is building social relationships based on "networked individualism", a paradigm shift in the way people are connected: from relatively homogenous, broadly-embracing, densely-knit, and tightly-bounded groups to more heterogeneous, specialized, sparsely-knit, and loosely-bounded social networks. These developments serve as both a summary of research knowledge and a guide for future research. Above all, they demonstrate that the shared perspective binding the NetLab network together is not frozen in time but reflects the intellectual growth of its members. NetLab is well-positioned to make its next steps.

REFERENCES

- Boase, J., & Ling, R. (2013). Measuring mobile phone use. *Journal of Computer-Mediated Communication*, 18(4), 508–519. doi:10.1111/jcc4.12021
- Chen, W., & Reese, S. D. (Eds.). (2015). *Networked China*. Routledge.
- Chen, W., & Wellman, B. (2009). Net and Jet. *Information Communication and Society*, 12(4), 525–547. doi:10.1080/13691180902858080
- Chua, V. (2011). Social networks and labour market outcomes in a meritocracy. *Social Networks*, 33(1), 1–11. doi:10.1016/j.socnet.2010.08.001
- Chua, V., & Wellman, B. (2016). Networked Individualism—East Asia Style. In *Oxford Research Encyclopedia of Communication*. Oxford, UK: Oxford University Press.
- Côté, R. (2013). From disadvantage to success? Indigenous entrepreneurship in the urban marketplace. In *International Sociological Association: Research Committee on Social Stratification Conference*.
- Dimitrova, D., & Koku, E. (2010). Managing Collaborative Research Networks: The Dual Life of a Virtual Community of Practice. *International Journal of Virtual Communities and Social Networking*, 2(4), 1–23. doi:10.4018/jvcsn.2010100101
- Dimitrova, D., Mok, D., & Wellman, B. (2015c). Changing Ties in a Far-Flung, Multidisciplinary Research Network. *The American Behavioral Scientist*, 59(5), 599–628. doi:10.1177/0002764214556803
- Dimitrova, D., & Wellman, B. (2012). NetLab. In Z. Yan (Ed.), *Encyclopedia of Cyber Behavior* (pp. 42–63). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-0315-8.ch004
- Dimitrova, D., Wellman, B., Gruz, A., Hayat, Z., Mo, G. Y., Mok, D., ... Zhuo, X. (2013). NAVEL Gazing: Studying a Networked Scholarly Organization. *Mathematics in Industry*, 18, 287–313.
- Gruz, A., & Haythornthwaite, C. (2013). Enabling community through social media. *Journal of Medical Internet Research*, 15(10), e248. doi:10.2196/jmir.2796 PMID:24176835

- Gruzd, A., Jacobson, J., Mai, P., & Wellman, B. (Eds.). (2016). Understanding communities in an age of social media. *Information, Communication & Society*, 1187-1320.
- Gruzd, A., & Roy, J. (2014). Investigating Political Polarization on Twitter. *Policy & Internet*, 6(1), 20–21. doi:10.1002/1944-2866.POI354
- Gruzd, A., & Tsyganova, K. (2015). Information Wars and Online Activism during the 2013/2014 Crisis in Ukraine. *Policy & Internet*, 7(2), 121–158. doi:10.1002/poi3.91
- Gruzd, A., & Wellman, B. (Eds.). (2014). American Behavioral Scientist, 58, 10. Special issue on Social Influence Online. 58(10):1251-1392
- Gruzd, A., & Wellman, B. (2016). Networked Influence in Social Media. *The American Behavioral Scientist*, 58(10), 1251–1260.
- Gruzd, A., Wellman, B., & Takhteyev, Y. (2012). Imagining Twitter as an Imagined Community. *The American Behavioral Scientist*, 55(10), 1294–1318. doi:10.1177/0002764211409378
- Hampton, K. N. (2016). Persistent and Pervasive Community. *The American Behavioral Scientist*, 60(1), 101–124. doi:10.1177/0002764215601714
- Hampton, K. N., Goulet, L. S., & Albanesius, G. (2015). Change in the social life of urban public spaces. *Urban Studies (Edinburgh, Scotland)*, 52(8), 1489–1504. doi:10.1177/0042098014534905
- Hampton, K. N., Lu, W., & Shin, I. (2016). Digital media and stress. *Information Communication and Society*, 19(6), 1267–1286. doi:10.1080/1369118X.2016.1186714
- Hayat, Z., & Mo, G. Y. (2015). Advice Giving and Receiving Within A Research Network. *The American Behavioral Scientist*, 59(5), 582–598. doi:10.1177/0002764214556802
- Haythornthwaite, C., Andrews, R., Fransman, J., & Meyers, E. M. (Eds.). (2016). *SAGE Handbook of E-learning Research* (2nd ed.). Thousand Oaks, CA: Sage. doi:10.4135/9781473955011
- Haythornthwaite, C., & Wellman, B. (1998). Work, Friendship and Media Use for Information Exchange in a Networked Organization. *Journal of the American Society for Information Science*, 49(12), 1101–1114. doi:10.1002/(SICI)1097-4571(1998)49:12<1101::AID-ASI6>3.0.CO;2-Z
- Hogan, B. (2015). From Invisible Algorithms to Interactive Affordances. In Roles, Trust, and Reputation in Social Media Knowledge Markets (pp. 103-117). Springer.
- Hogan, B., & Berry, B. (2011). Racial and Ethnic Biases in Rental Housing. *City & Community*, 10(4), 1–28. doi:10.1111/j.1540-6040.2011.01376.x
- Hogan & Wellman, B. (2014). The Relational Self-Portrait. In M. Graham & W. Dutton (Eds.), *Society and the Internet*, (pp. 53-66). Oxford, UK: Oxford University Press.
- Kennedy, B., Wellman, B., & Amoroso, J. (2011). Can You Take It With You? Mobility, ICTs and Work-Life Balance. In J. Katz (Ed.), *Mobile Communication* (pp. 310–344). Piscataway, NJ: Transaction.
- Kobayashi, T., Boase, J., Suzuki, T., & Suzuki, T. (2015). Emerging From the Cocoon? *Journal of Computer-Mediated Communication*, 20(3), 330–345. doi:10.1111/jcc4.12116
- Lin, N., & Erickson, B. (Eds.). (2008). *Social Capital*. New York: Oxford University Press.
- Marin, A. (2012). Dont mention it: Why people dont share job information, when they do, and why it matters. *Social Networks*, 34(2), 181–192. doi:10.1016/j.socnet.2011.11.002
- Marin, A., & Hampton, K. (2007). Simplifying the personal network name generator alternatives to traditional multiple and single name generators. *Field Methods*, 19(2), 163–193. doi:10.1177/1525822X06298588

- Matthias, K., van den Berg, P. E. W., Carrasco, J.-A., Arentze, T. A., Axhausen, K., Mok, D., & Wellman, B. et al. (2013). Distance Patterns of Personal Networks in Four Countries. *Journal of Transport Geography*, 31(3), 236–248.
- Miyata, K., Yamamoto, H., & Ogawa, Y. (2015). What affects the spiral of silence and the hard core on Twitter? *The American Behavioral Scientist*, 59(9), 1129–1141. doi:10.1177/0002764215580618
- Mo, G. Y., & Wellman, B. (2012). Understanding sequencing in social network communication. *Bulletin de Methodologie Sociologique*, 113(1), 76–87. doi:10.1177/0759106311426997
- Mo, G. Y., & Wellman, B. (2016). The Effects of Multiple Team Membership on Networking Online and Offline. *Information Communication and Society*, 19(6), 1250–1266. doi:10.1080/1369118X.2016.1187194
- Mok, D., Wellman, B., & Carrasco, J. (2010). Does distance matter in the age of the Internet? *Urban Studies (Edinburgh, Scotland)*, 47(13), 2747–2783. doi:10.1177/0042098010377363
- Neves, B. B., Amaro, F., & Fonseca, J. R. (2013). Coming of (old) age in the digital age. *Sociological Research Online*, 18(2), 6. doi:10.5153/sro.2998
- Quan-Haase, A. (2015). *Technology and Society* (2nd ed.). Toronto: Oxford University Press.
- Quan-Haase, A., Martin, K., & Schreurs, K. (2016). Interviews with digital seniors. *Information Communication and Society*, 19(5), 691–707. doi:10.1080/1369118X.2016.1140217
- Quan-Haase, A., Suarez, J. L., & Brown, D. M. (2015). Collaborating, Connecting, and Clustering in the Humanities. *The American Behavioral Scientist*, 59(5), 565–581. doi:10.1177/0002764214556806
- Rainie, L., & Wellman, B. (2012). *Networked*. Cambridge, MA: MIT Press.
- Takhteyev, Y., Gruz, A., & Wellman, B. (2012). Geography of Twitter Networks. *Social Networks*, 34(1), 73–81. doi:10.1016/j.socnet.2011.05.006
- Wang, H., & Wellman, B. (2010). Social Connectivity in America. *The American Behavioral Scientist*, 53(8), 1148–1169. doi:10.1177/0002764209356247
- Wellman, B., Dimitrova, D., Hayat, T., & Mo, G. Y. (2017). Fifteen Implications of Networked Scholar Research for Networked Work. *International Journal of Communication in Press*.
- Wellman, B. & Kelner, M. (2015). *Why Do Patients Seek Out Integrative Health Care?* National University of Singapore.

ADDITIONAL READING

- Boase, J. (2008). Personal networks and the personal communication system: Using multiple communication media to connect. *Information Communication and Society*, 11(4), 490–508. doi:10.1080/13691180801999001
- Boase, J., & Kobayashi, T. (2008). Kei-tying teens: Using mobile phone email to bond, bridge, and break with social ties – A study of Japanese adolescents. *International Journal of Human-Computer Studies*, 66(12), 930–943. doi:10.1016/j.ijhcs.2008.07.004
- Carrasco, J., Hogan, B., Wellman, B., & Miller, E. (2008). Agency in social activity and ICT interactions: The role of social networks in time and space. *Tijdschrift voor Economische en Sociale Geografie*, 99(5), 562–583. doi:10.1111/j.1467-9663.2008.00492.x

- Carrasco, J., Wellman, B., & Miller, E. (2008). How far – and with whom – Do people socialize? *Transportation Research Record*, 2076, 114–122. doi:10.3141/2076-13
- Côté, R., Plickert, G., & Wellman, B. (2009). Does the golden rule, rule? In R. Hsung, N. Lin, & R. Breiger (Eds.), *Contexts of social capital: Social networks in markets, communities and families* (pp. 49–71). London, UK: Routledge.
- Hampton, K. (2007). Neighborhoods in the network society: The e-neighbors study. *Information Communication and Society*, 10(5), 714–748. doi:10.1080/13691180701658061
- Hampton, K. (2010). Internet use and the concentration of disadvantage: Globalization and the urban underclass. *The American Behavioral Scientist*, 53(8), 1111–1132. doi:10.1177/0002764209356244
- Hampton, K., Sessions, L., Her, E., & Rainie, L. (2009, November 4). *Social isolation and new technology*. Washington, DC: Pew Internet and American Life project.
- Haythornthwaite, C., & Andrews, R. (2011). E-inclusion and exclusion. In C. Haythornthwaite & R. Andrews (Eds.), *E-learning theory and practice* (pp. 179–190). London, UK: Sage. doi:10.4135/9781446288566.n11
- Hogan, B., Carrasco, J., & Wellman, B. (2007). Visualizing personal networks: Working with participant-aided sociograms. *Field Methods*, 19(2), 116–144. doi:10.1177/1525822X06298589
- Kelner, M. J., Wellman, B., Pescosolido, B., & Saks, M. (2000). *Complementary and alternative medicine: Challenge and change*. Amsterdam, The Netherlands: Gordon and Breach/hap.
- Kennedy, T., Amoroso, J., & Wellman, B. (2010). Can you take it with you? Mobility, ICTs and work-life balance. In J. Katz (Ed.), *Mobile communication* (pp. 310–344). Piscataway, NJ: Transaction Publishers.
- Koku, E., & Wellman, B. (2004). Scholarly networks as learning communities: The case of TechNet. In R. Kling, J. Gray, & S. Barab (Eds.), *Designing for virtual communities in the service of learning* (pp. 299–337). New York, NY: Cambridge University Press. doi:10.1017/CBO9780511805080.015
- Marin, A., & Hampton, K. (2007). Simplifying the personal network name generator: Alternatives to traditional multiple and single name generators. *Field Methods*, 19(2), 163–193. doi:10.1177/1525822X06298588
- Quan-Haase, A. (2007). University students local and distant social ties. *Information Communication and Society*, 10(5), 671–693. doi:10.1080/13691180701658020
- Quan-Haase, A. (2010). Self-regulation in instant messaging (IM): Failures, strategies, and negative consequences. *International Journal of e-Collaboration*, 6(3), 22–42. doi:10.4018/jec.2010070102
- Veenhof, B., Wellman, B., Quell, C., & Hogan, B. (2008). *Isolation, cohesion or transformation? How Canadians' use of the Internet is shaping society*. *Connectedness Series*, 16. Ottawa: Statistics Canada.
- Wellman, B. (2001). Physical place and cyber place: The rise of networked individualism. *International Journal of Urban and Regional Research*, 25(2), 227–252. doi:10.1111/1468-2427.00309

KEY TERMS AND DEFINITIONS

Community: Tightly bound groups of people with shared interests and values, common identity, and a sense of belonging. Current approaches treat communities as delineated by interaction and commitment (e.g. virtual communities) rather than by location (e.g. neighbourhoods).

Networked Individuals: Connected individuals embedded in multiple partial networks instead of than being members of dense tightly bound groups.

Networked Organizations: Organizations where extensive use of computer networks and mobile technology is coupled with changes in group dynamics, communication, and authority.

Networked Workers: Rather than working in a single work group or independently from others, networked workers work in multiple, often distant, teams and projects.

Scholarly Networks: The networks of academics and researchers linked by one or more relations; such networks are increasingly becoming formalized, multi-disciplinary, geographically distributed, and reliant on technology.

Social Network Analysis (SNA): An interdisciplinary perspective which focuses on the patterns of relations among social actors and interprets these patterns as social structure.

Social Networks: A set of social actors - be they individuals, groups, organizations, or countries - and the relations among them.

Triple Revolution: Social transformation comprised of: (a) the change in social networks from people being embedded in tightly bound groups to people being embedded in multiple partial network (aka *networked individualism*); (b) proliferation of internet technology; and (c) spread of mobile technologies.

Online Dating/Dating Apps

S

Vladimir Santiago Arias
Texas Tech University, USA

Narissra Maria Punyanunt-Carter
Texas Tech University, USA

INTRODUCTION

Over the past several years, online dating services are increasingly becoming popular venues for finding romantic relationships. In 2012, Match.com reported that one in six marriages started online (Ramirez, Sumner, Fleuriet & Cole, 2015). In 2013, the online mating services brought \$2.1 billion (Ginsberg, 2015) whereas compared to ten years ago, in 2004, the dating industry revenue was only \$473 million. Nowadays, there are many online dating sites such as Match.com, eHarmony, and PerfectMatch.com, with over 50 million users combined (Consumer Rankings., 2012), and the online dating business keeps growing (Visual Economics Credit Loan blog, 2015). Online dating refers to web sites and apps that facilitate romantic relationships' initiation by offering users (1) access to the profiles of potential romantic candidates, (2) a communication channel to initiate contact, and (3) a romantic compatibility matching-algorithm to be paired for potential romantic initiation (see Finkel, Eastwick, Karney, Reis & Sprecher, 2012). Indeed, most online dating platforms are similarly structured (Rosen, Cheever, Cummings & Felt, 2008), in general: users post a photograph and answer questions in regards to personal information and other relevant demographics; however, there is considerable variance among online dating forums with regard to users' level of involvement, interaction, and self-disclosure.

Despite the array of online dating sites and apps, a new online dating app entered to the online dating market, and it is taking over the entire online love business: Tinder. The new app

just entered to the market in 2012, and, two years later, it reached approximately 30 million users, almost a third of the total online dating population (e.g., 96 million users) (Forbes, November 2014). Thus, the popularity of the app has rapidly grown. Tinder app innovates the usual online dating service explained above, by providing users a seemingly endless selection of photos of potential mates without the need to answer questionnaires or forms (Bertoni, 2014a); then, the algorithm of the app links users' contacts from Facebook profiles to provide photographs of potential romantic candidates. After solely looking at photos of potential mates, users swipe right if they like a person and, by the contrary, swipe left if not (Bertoni, 2014a); finally, if both parties like each other, the platform provides a parallel interface to send messages to each other to decide whether or not to meet in person and exchange personal contact information.

Besides the successfulness of online dating market, the online dating service has always been severely criticized for its 'overemphasis' on physical appearance. However, disregarding the communication context (i.e., Face-to-Face and Online), physical appearance is the initiator for communication behaviors in most of the cases. The online dating success trend has been widely explained by the new media pervasiveness argument or the idea that this service is prosperous 'only' or 'mostly' because the access to personal computers and smartphone is wide spread, then focusing only on related phenomena such as self-presentation, self-disclosure, and/or social anxiety. If new media pervasiveness explains this new

social trend, why did commercial video-dating not become so popular during the 90s when the access to video cameras was also pervasive in the U.S.? Little attention has been directed to how online dating mirrors human perception of first impression while forming interpersonal relationships.

Therefore, the present chapter understand the role of human perception of physical appearance during first impression formation which may better tune with Tinder' interface; in other words, Tinder feels more natural to users compared to other dating apps, swipe to left or to the right feels better than 'browsing' profiles; in addition to the pervasiveness argument of new media and apps, this chapter explains Tinder' increasing popularity seems to match to interface architecture as well. Consequently, the concept of technological affordance tuning will be discussed to explain the success on online dating in addition to new media pervasiveness argument; however, the idea of affordance tuning is not widely discussed in the CMC research on online dating, and this concept is the main contribution from this chapter.

BACKGROUND

The online dating technology has been around since 1970, but the rapid rate of development of cheap, fast, reliable, and user personal computers with Internet made online dating technology to evolve from being just an online interface for personal romantic advertisement (see Byrne, Ervin & Lamberth, 1970), then to become an algorithm-based matching system (i.e., e-Harmony.com or Match.com) to finally a combination of both; with the inclusion of smartphone-based dating applications and GPS technology (i.e. Blendr or Skout), the new version of this CMC technology also became into satellite dating (Quiroz 2013). Nowadays, online dating users cannot only browse romantic candidate profiles, but also know where they are given information to decide whether or not meet them in person. But, if online dating success is explained by the argument of new media

pervasiveness solely, then why did commercial video-dating not enjoy similar popularity during the 1990s when the access to video cameras was also pervasive in the U.S.?

Then a new app entered to the market in 2012: Tinder. In only two years, in 2014, Tinder reached approximately 30 million users who have used this app to find a partner making, more than 15 million matches daily; then, users are checking out about a 1.2 billion profiles -14.000 per second- (Bertoni, 2014). While the entire U.S. dating business is worth somewhere between 5 and 6 billion dollars (Forbes, 2014), just Tinder is worth somewhere between \$.1 billion and \$.1.5 billion, and some big bank analysts said that Tinder could even top \$. 5.5 billion in few years, which is almost the entire online market soon (Forbes, 2014). The app is very popular and is one of the top producing online dating apps.

This chapter argues online dating popularity responds more to how technology closely mirrors actual human interaction in the early stages of forming a romantic relationship than other (and earlier) dating services venues; furthermore: why did Tinder get that level of popularity in only 4 years, whereas other online dating sites have been around for more than a decade without obtaining the same users' preference? There exists a constant tension between tasks and technologies in interface design; Gaver (1991) accurately explains that a design based on only technology innovations is functionally awkward, and, by the contrary, a design just based on users' needs may lead to overlook technological innovations; indeed, the main purpose for artifact interface design should be to create a CMC architecture which reflects an interaction between human sensory systems and CMC affordances. Consequently, in order to answer the main queries of this chapter, non-verbal research on physical attractiveness and first impression formation with the concept of affordance tuning along the way will be discussed to explain online dating increasing engagement across the world, an idea which has been ruled out by scholarly research.

MAIN FOCUS OF THE ARTICLE

Issues, Controversies, Problems

Many individuals have assumed new, and even counter, identities in their pursuit for online romantic relationships (Alapack, Blichfeldt, & Elden, 2005). The Internet and CMC (Computer-Mediated Communication) have become suitable forms of communication between individuals looking for controlled forms of relational engagement (Hardey, 2004); the online dating phenomenon reflects the same motivation that seeks romantic outcomes (Whitty, 2008; Heino, et al., 2010; Kang & Hoffman, 2011; Finkel, Eastwick, Karney, Reis, & Sprecher, 2012); Ramirez, Summer, Fleuriet, & Cole, 2015) such as pursuing to attain a long-term relationship commitment (Mahfouz, Philaretou & Theocharaous, 2008).

However, the observable human behaviors associated with online dating have traditionally been isolated to textual inferences. For example, online dating has been research on the basis of Walther's (1996) 'Hyperpersonal' model which aims to understand how users take and advantage from CMC affordances to overcome the absence of nonverbal cues (Walther, 1996, 2007). Thus, in CMC, users create and exchange messages in physical isolation from receiver, masking involuntary cues which make users perform overattributions based on stereotypical impressions of their partners without qualifying the spontaneity of such impressions. In other words, as senders, users do not show their "*their natural physical features and non-deliberate actions into the receiver's realm of perception*" (Walther, 2007, p. 2541, Italics added); then, it can be inferred that online dating fits better for users' dating goals in terms of having more control on first impression formation.

For many individuals, the thought of going online to pursue a romantic relationship seems different and unorthodox (Anderson, 2005). In a study attempting to measure attitudes related to Internet romance, numerous results indicated var-

ied levels of romantic satisfaction. For individuals with high levels of *Internet affinity*, or the desire to interact via the Internet, their perceptions of romantic relationships were seen as positive and enduring (Anderson, 2005). Research suggests that there is a relationship between levels of perceived realism and individual perceptions of online romantic relationships. Surprisingly, the results did not support a positive relationship between levels of perceived realism and online romantic relationships. Anderson (2004) discovered that there would be a positive relationship between amount of time (hours per week) on the Internet and perceptions of online romantic relationships. Anderson (2004) found that a high degree of Internet use was instrumental in establishing positive perceptions of online romantic relationships. Anderson (2004) noted that individuals must adjust their behaviors to the new environment created by online dating. Further, individuals engaging with online relationships must adopt new forms of nonverbal behaviors in order to offset the absence of traditionally implemented offline cues.

Just as with face to face relationships, online relationships are governed by certain codes and standards (Hardy, 2004). *Netiquette* asserts itself as the ruling body for all forms of romantic online exchange (Hardy, 2004). Research dedicated to online romances has reported that online dating services allow for high volumes of information to be transferred between sender and receiver (Hardy, 2004). Unlike print forms of dating services such as the use of newspaper classifieds, Internet services allow individuals to disseminate a plethora of biological facts at a faster rate (Hardy, 2004).

One of the greatest forms of romantic information dissemination is based on nonverbal elements. According to Hardy (2004), the posting of one's photograph onto the web is a very personal decision. For some the posting of a photograph takes away from the *otherworld* experience they are trying to manifest through online interaction. Individuals that choose to eliminate their photo seek to maintain a sense of relationship based on emotional and intellectual criterion and not physi-

cal characteristics (Hardy, 2004). For others, the exchanging of photographs act as a mechanism for the reduction of feelings related to ambiguity and uncertainty (Hardy, 2004). The practice of sending a picture electronically allows both parties to construct a more vivid, while still nonverbal representation of the other person. Further, by sending a picture, both parties are better prepared for face to face interaction (Hardy, 2004).

Social relationships start from first impression formation (Richmond, McCroskey & Hickson, 2007), and, simultaneously, first impression is formed through physical attractiveness assessment. Consequently, dating is not the exception for physical attractiveness as the basis for first impression formation (see Richmond, et al., 2007; Tidwell, Eastwick & Finkel, 2012); Eastwick & Hunt, 2014). Indeed, nonverbal research in interpersonal relationships suggests that physical attractiveness is what determines initial communication approach (Richmond, et. al., 2007). But, attractiveness assessment is individually performed and visually driven, it is in the eye of the beholder (Richmond, et. al., 2007); for example, in Eastwick and Hunt (2014) measured romantic acquaintanceship in three different times in three separated studies with a total of 309 undergraduate participants, the results suggested that the romantic evaluation of potential mates is more unique to a person that consensual over time.

Physical attractiveness primarily sets off first dates in online settings as well (see Walther, 2007; Kang & Hoffman, 2011; Finkel et. al., 2012; Ramirez, et al., 2015). The problem arises when first impression formation is mediated by computer technology because it extrapolates and enhances attractiveness in a hyperbolic way: during the receiving stage in the communication system, users perform an overreliance (idealize) on others' traits because they only have access to the available pictures and textual information and messages which prompts minimal similarity; on the other hand, in the sender communication stage, individuals selectively self-present in a controlled and social desirable fashion; while as a

communication channel, CMC facilitates editing, discretion, and convenience to form impressions (see Walther, 1996, 2007). As a matter of fact, physical attractiveness is found to be the most valued aspect in online dating (see for a review Whitty, 2008; Rosen, Cheever, Cummings, & Felt, 2008; Turner & Hunt, 2014); for example, Ramirez, et. al., (2015) found that perceptions of higher levels of intimacy, informality, composure, and social orientation were positively predicted by the number of photographs that users had on their profiles. Therefore, attractiveness is the strongest predictor for online relationship initiation (Whitty, 2008; Walther, 2008; Finkel, et al., 2012; Ramirez, et al, 2015). Consequently, the important role of physical attractiveness for dating explains the central role of personal pictures in online dating interface architecture, but this is an idea not widely discussed in the scholarly literature.

CMC offers a virtual environment in which the physical isolation prompts anonymity and makes users experience more disinhibition; for example Lapidot-Lefter and Barak (2012) examined the effect of anonymity/non-anonymity, visibility/invisibility, and eye contact/absence of eye contact in an experimental design with 142 participants (71 men and 71 women) from ages 18 to 34 years during shorts debates through CMC; thus, a factorial experimental design was conducted, and the results suggest that lack of eye contact was the primary contributor to online disinhibition effects. As a consequence from acquiring control of the new language system through CMC, individuals are able to openly express their feelings and emotions with one another. According to Alapack et al. (2005), the courtship behaviors created by the evolving language system tend to limit romantic feelings such as awkwardness and conflict. Moreover, the use of tactical punctuations during romantic interludes causes the online environment to become strikingly similar to that of the real world (Alapack et al., 2005).

Besides the currently broad access to personal computer and smartphones that is also a factor to the increasing popularity of online dating, it

is worth to consider that any technology device become popular when its use is more close to naturalness (Xie & Newhagen, 2012); in other words, individuals tend to use more the devices that are feel natural to perform a given task. The concept of affordances refer to the process in which individuals perceive the objects in their environment in terms of its potential for action, without significant intermediate stages that require memory or inferences. For example, a chair is perceived in terms of its 'seatability'; hence, people perceive the affordance of 'seatability' (Gaver, 1991). Accordingly to Gaver's example, an affordance is determined in terms of the attributes of both actor and object; subsequently, there should be perceptual information for an existent affordance. Ergo, affordances are basically determined by the physical properties, shape and scale of artefacts that are more or less easy to be perceived by specific human perception processes, and first impression formation process seems to fit better with online dating.

Online dating affordances give users a broad range of choices to 'control' self-presentation in terms of physical attractiveness and personal information that permits impression management behavior. Hence, online dating interface architecture seems to line up in parallel to Gaver's (1991) requisites of affordance because for interface design the following postulates should be considered: 1) the artifact physical attributes should be compatible with those of the actor, and, in online dating, the visual perception and assessment of physical attractiveness is paired with the central role of personal pictures in this platform; 2) the information of the artifact' attributes are available in a manner compatible with a perceptual system of the actor, and this CMC technology is a visually driven device to perform romantic assessment; and 3) the attributes and the action should be culturally and personally relevant, and, in online dating, the profiles are individually accessed, but culturally valued, as physical appearance evaluation is performed. The affordance tuning for technology design seems to be more appropriate

in Tinder in which users do not need to fill in long questionnaires, and simply go visually during attractiveness assessment to then asking questions to candidates, as it happens during FtF settings.

SOLUTIONS AND RECOMMENDATIONS

As online romantic conversations begin to evolve and take shape, scholars must continue to observe and examine online relational phenomena through the use of online dating apps which have stronger effects on individuals due to its perceived 'naturalness' for its 'proximity'. Through CMC affordances such as the use of emoticons and other tactical punctuations, nonverbal communication plays a crucial role in the development of Internet-based courtship behaviors and romantic online relationships. Online dating and online apps have changed the way that people pursue romantic relationships; it is not only important for scholarly research to understand how users exploit CMC affordances to overcome its limitations as Walther's model proposes, but also it is paramount to include the analysis of how those affordances specifically tune to human perception process making in it more or less 'popular', or more or less 'interactive' which is also part of the market successfulness of any technology, as it can be inferred while trying to answer the question about why video dating did not become as successful as online dating did.

FUTURE RESEARCH DIRECTIONS

Future research should look at how closely online and offline courtship behaviors overlap each other by introducing the affordance tuning conceptualization and nonverbal research. In addition, future research should investigate the communication behaviors that individuals use on online apps compared to face-to-face interactions; for example, Hunt, Eastwick & Finkel (2015) compared

the length of acquaintanceship [first impression formation] of 167 couples taken from previous longitudinal studies, and they found a negative interaction between physical attractiveness and the time couples got to know each other (i.e. up to 9 months). In other words, the longer individuals had known each other, the less probability for choosing a mate by the attractiveness criterion, but most likely by psychological similarity such as personality traits. By the contrary, in the online dating context, Ramirez, et al. (2015) investigated what is the association between the amount of time spent online before meeting the potential candidate FtF, a hierarchical regression model was conducted in five dimensions (i.e., intimacy, composure, formality, task social orientation, dominance) where a curvilinear association among the first four was found, and it indicated a significant association to suggest that online daters benefit from the interface if the time period of online interaction is brief. Basically, the longer time spent on online platform decreases the motivation to meet face-to-face because it dampen the perceptions of closeness because candidates start to seek more online information about a candidate. However, there is no longitudinal study that simultaneously includes both dating settings (i.e. FtF and Online) to compare how the manner in which a romantic relationship starts may impact further romantic relationship outcomes in the long term; since the current state of research lacks of this type of study, it is difficult to establish strong claims in favor or against online dating for the long run.

CONCLUSION

As online romantic conversations begin to evolve and take shape, scholars must continue to observe and examine online relational phenomena through the use of online dating apps, which will be the next ubiquitous trend in terms of technology use in the upcoming years due to the wide spread

smartphone presence (Gerlich, Drmheller, Babb & D'Armond, 2015). Through the use of personal pictures, emoticons and other tactical punctuations, nonverbal communication plays a crucial role in the development of Internet-based courtship behaviors. The popularity of online Tinder is better calibrated to interpersonal attraction than the values-infused and text-heavy quality of the typical online dating site or app, which places users in a critical shopping mindset rather than an orientation based solely on attraction. Moreover, research on online dating should include interface design understanding to bring in the level of 'affordance tuning' to human sensory system to perform a specific task (i.e. dating) to assess CMC popularity and impact on human behavior.

REFERENCES

- Alapack, R., Blichfeldt, M. F., & Elden, A. (2005). Flirting on the Internet and the hickey: A hermeneutic. *Cyberpsychology & Behavior*, 8(1), 52–61. doi:10.1089/cpb.2005.8.52 PMID:15738693
- Anderson, T. L. (2005). Relationships among Internet attitudes, Internet use, romantic beliefs, and perceptions of online romantic relationships. *Cyberpsychology & Behavior*, 8(6), 521–531. doi:10.1089/cpb.2005.8.521 PMID:16332163
- Finkel, E. J., Eastwick, P. W., Karney, B. R., Reis, H. T., & Sprecher, S. (2012). Online dating: A critical analysis from the perspective of psychological science. *Psychological Science in the Public Interest*, 13(1), 3–66. doi:10.1177/1529100612436522 PMID:26173279
- Gaver, W. W. (1991). Technology Affordances. *Proceedings of the CHI 1991*. ACM Press.
- Hardy, M. (2004). Mediated relationships: Authenticity and the possibility of romance. *Information Communication and Society*, 7(2), 207–222. doi:10.1080/1369118042000232657

Kang, T., & Hoffman, L. (2011). Why would you decide to use an online dating site? Factors that lead to online dating. *Communication Research Reports*, 28(3), 205–213. doi:10.1080/08824096.2011.566109

Kindred, J., & Roper, S. (2004). Making connections via instant messenger (IM): Student use of IM to maintain personal relationships. *Qualitative Research Reports in Communication*, 5, 48–54.

Krohn, F. B. (2004). A generational approach to using emoticons as nonverbal communication. *Journal of Technical Writing & Communication*, 34(4), 321–328. doi:10.2190/9EQH-DE81-CWG1-QLL9

Mahfouz, A. Y., Philaretou, A. G., & Theocharous, A. (2008). Virtual social interactions: Evolutionary, social psychological and technological perspectives. *Computers in Human Behavior*, 24(6), 3014–3026. doi:10.1016/j.chb.2008.05.008

Ramirez, A., Summer, E. M., Fleuriot, C., & Cole, M. (2015). When online dating partners meet offline: The effect of modality switching on relational communication between online daters. *Journal of Computer-Mediated Communication*, 20(1), 99–114. doi:10.1111/jcc4.12101

Rosen, L. D., Cheever, N. A., Cummings, C., & Felt, J. (2008). The impact of emotionality and self-disclosure on online dating versus traditional dating. *Computers in Human Behavior*, 24(5), 2124–2157. doi:10.1016/j.chb.2007.10.003

Walther, J. B. (2007). Selective self-presentation in computer-mediated communication: Hyperpersonal dimensions of technology, language, and cognition. *Computers in Human Behavior*, 23(5), 2358–2557. doi:10.1016/j.chb.2006.05.002

Whitty, M. T. (2008). Revealing the real me, searching for the actual you: Presentations of self on an internet dating site. *Computers in Human Behavior*, 24(4), 1707–1723. doi:10.1016/j.chb.2007.07.002

Xie, W., & Newhagen, J. E. (2014). The effects of communication interface proximity on user anxiety for crime alerts received on desktop, laptop, and hand-held devices. *Communication Research*, 41(3), 375–403. doi:10.1177/0093650212448670

ADDITIONAL READING

Hunt, L. L., Eastwick, P. W., & Finkel, E. J. (2015). Leveling the playing field: Longer acquaintance predicts reduced assortative mating on attractiveness. *Psychological Science*, 1–8. doi:10.1177/0956797615579273 PMID:26068893

Rözer, J. J., Mollenhorst, G., & Volker, B. (2015, March). Romantic relationship formation, maintenance and changes in personal networks. *Advances in Life Course Research*, 23, 86–67. doi:10.1016/j.alcr.2014.12.001 PMID:26047843

Tidwell, N. D., Eastwick, P. W., & Finkel, E. J. (2012). Perceived, not actual, similarity predicts initial attraction in a live romantic context: Evidence from the speed-dating paradigm. *Journal of the International Association for Relationship Research*, 20(2), 199–215. doi:10.1111/j.1475-6811.2012.01405.x

Turner, M., & Hunt, N. (2014). What does your profile picture say about you? The accuracy of thin-slice personality judgements from social networking sites made at zero-acquaintance. In G. Meisewitz (Ed.), *Social Computing and Social Media* (pp. 506–516). Zurich, Switzerland: Springer International publications.

Wendorf, J. E., & Yang, F. (2015, June). Benefits of a negative post: Effects of computer-mediated venting on relationship maintenance. *Computers in Human Behavior*, 52, 271–277. doi:10.1016/j.chb.2015.05.040

KEY TERMS AND DEFINITIONS

Affordances: “The perceived properties of the things that determine just how they could possibly be used” (Norman 1998).

Away Messages: The messages that one uses to indicate that they are away from the computer or unavailable to communicate online.

Computer Mediated Communication: The use of computer or computer technology to communicate with other individuals.

Emoticons: Nonverbal expressions that are expressed via text.

Internet Affinity: The desire to initiate a romantic relationship via the Internet.

Netiquette: The etiquette and manners that individuals use while on the Internet.

Online Dating: CMC technology designed to facilitate romantic relationships initiation by anticipating face-to-face interaction through the access to personal profiles without owners’ awareness.

Other World: The notion that something is not real.

Online Prosocial Behaviors

S**Michelle F. Wright***Pennsylvania State University, USA***William Stanley Pendergrass***American Public University System, USA*

INTRODUCTION

Over three billion people use electronic technologies (e.g., cell phones, the internet) everyday (Internet Live Stats, 2016). Although there are many investigations and news stories about negative online behaviors, less attention has been given to positive online behaviors. There are many opportunities to receive help or to perform prosocial acts through electronic technologies. This chapter focuses on online prosocial behaviors. The chapter includes eight sections:

- Section one provides the definition of offline and online prosocial behaviors.
- Section two examines the unique characteristics of the cyber context and how such characteristics are conducive to prosocial behaviors.
- Section three focuses on various online prosocial behavior, including helping through electronic groups, online mentoring, online donations to charities, virtual voluntarism, and helping in other electronic contexts (e.g., social networking sites).
- Section four investigates the value of online prosocial behaviors to the giver and receiver.
- Section five provides theoretical explanations for why people engage in online prosocial behavior.
- Section six describes solutions and recommendations for organizations wanting to harness electronic technologies for various helping opportunities.

- Section seven presents suggestions for future research on online prosocial behavior.
- The last section provides concluding remarks regarding the chapter.

BACKGROUND

Prosocial behaviors are defined as voluntary acts directed toward people or society (Eisenberg & Miller, 1987). Such behaviors may include helping, sharing, donating, and volunteering. Online prosocial behaviors can take various forms, including donating time and attention to electronic discussion boards and Wikis (e.g., Antin, 2011; Butler, Sproull, Kiesler, & Kraut, 2007), helping among corporate employees (e.g., Duranova & Ohly, 2016), helping players in computer games (e.g., Molyneux, Vasudevan, & de Zuniga, 2015), online mentoring (e.g., Cheng, Hanuscin, & Volkmann, 2016), sharing and contributing to open source software (e.g., Lakhani & Hippel, 2003), virtual voluntarism (e.g., Kim & Lee, 2014), and making charitable donations to organizations online (e.g., Bennett, 2006). There are various characteristics of the online environment that are favorable for helping online.

CHARACTERISTICS OF ONLINE PROSOCIAL BEHAVIORS

Online prosocial behaviors have some characteristics that set them apart from the same behaviors offline. Search engines make it easier to

DOI: 10.4018/978-1-5225-2255-3.ch614

find opportunities to help or receive help online (Sproull, Conley, & Moon, 2013). It is easier to give or receive help online because one's physical appearance or personal attributes do not influence other's opinions (Brennan, Moore, & Smyth, 1992). Individuals can use fake names or screen names and hide their identities online, which reduces stigmas associated with seeking help (Wright & Li, 2012). The online environment offers flexibility to individuals wanting to help or give help, allowing them to give help or receive help even with restricted schedules. There is high controllability over online prosocial behaviors. The online environment allows givers to choose when they want to help and if they want to help again without feeling pressured (Sproull et al., 2013; Wright & Li, 2012).

Although there are noticeable differences between online and offline prosocial behaviors, there are similarities. The relationship between the giver and receiver of prosocial behaviors in either environment can include strangers (e.g., Sproull et al., 2013), friends (e.g., Cornejo, Tentori, & Favela, 2013), and business colleagues (e.g., Duranova & Ohly, 2016). Prosocial behaviors are rewarding for givers in either social context (Butler et al., 2007; Eichhorn, 2008). Furthermore, prosocial behaviors can occur through formal and informal organizational institutions (Wright & Li, 2011). There is typically no expectation of direct reciprocity of prosocial behaviors in offline and online contexts (Sproull et al., 2013).

PROSOCIAL BEHAVIORS IN THE CYBER CONTEXT

This section presents a review of the literature on opportunities for prosocial behaviors via open source software and Wikis, electronic support groups, online mentoring, electronic fundraising and crowdfunding, virtual voluntarism, and other

technologies, such as social networking sites (SNS) and online gaming.

Open Source Software and Wikis

Online prosocial behavior began with IBM's sharing of their open source software code and the SHARE user group (i.e., an online association designed to provide technology professionals with continuing education). People can use the internet to volunteer and contribute code, documentation, and technical support to open source projects (Sproull & Kiesler, 2005). In 1991, a Finnish student posted a program on the internet and invited others to contribute their own code. This program was the beginning of Linux. Its development still continues today and is largely voluntary. Other source code information is available for Mozilla, StarOffice, Apache webserver, Python, and the free BSD operating system (Barcellini, Detienne, & Burkhardt, 2009; Raymond, 1999). Some investigations have focused on people's motivations for providing help to open source code projects, with findings revealing that people were more likely to contribute to these projects if they valued the goals of the program and believed their time would benefit themselves and others (Hertel, Niedner, & Herrman, 2003; Lakhani et al., 2003). The sharing of open source software continues into the 2000s with improved technology, such as smartphones and Web 2.0 (Barcellini et al., 2009).

Similar to the sharing of open source software is the contributions people make to Wikis. Wikis or Wikipedias are websites that allow people to collaboratively edit its content (Antin, 2011). Antin (2011) examined the characteristics associated with contributing to Wikis. He found that assumptions about the type of person who contributes to Wiki content, either hacker or geek stereotypes, affect whether people participate. People were more likely to contribute to Wikis when they feel accomplished, felt like they were part of a

community, and believed the work allowed them freedom (Kuznetsov, 2006).

Electronic Support Groups

Clinicians recognized that the internet removes the boundaries and stigmas associated with help-seeking often present in the offline environment. Clients can receive support and advice about their illnesses over the internet without the fear of being judged. Although one of the first studies (i.e., Schneider & Tooley, 1986) on online help seeking did not specifically investigate online prosocial behaviors, it was one of the first to examine electronic technologies, specifically computer-based support groups, to help adults quit smoking. This study set the precedence for future investigations on electronic support groups.

Researchers have continued to investigate electronic support groups. In one study, Finholt and Sproull (1990) investigated prosocial acts through electronic groups among cooperate employees. Employees engaged in a variety of prosocial behaviors related to both work and outside of work activities. Research continues to focus on how electronic support groups contribute to employee satisfaction and organizational commitment (Duranova & Ohly, 2016; Ragsdale & Hoover, 2016). Other research has focused on providing support to special populations. In particular, Brennan and colleagues (1992) found that an electronic network helped caregivers of someone with Alzheimer's disease feel supported by individuals "who really understood what they were going through" (p. 668). Similarly, Hassett and colleagues (1992) found that an electronic support group increased disabled individuals' feelings of social support from other community members.

The benefit of social support and help received through electronic support groups has been recognized in other populations as well, including sexual abuse survivors (e.g., Finn & Lavitt, 1994), people living with AIDs/HIV (e.g., Mo & Coulson,

2013), individuals with disordered eating (e.g., Stommel & Meijman, 2011), individuals with epilepsy (e.g., Wicks et al., 2012), caregivers of premature infants (e.g., Thoren, Metze, Buhner, & Garten, 2013), parents of children with cancer (e.g., Coulson & Greenwood, 2012), hearing impaired individuals (e.g., Thoren et al., 2011), breast cancer patients (e.g., Chen, 2012), and couples dealing with infertility (e.g., Malik & Coulson, 2010). Findings from these studies revealed that electronic support groups are successful at connecting and supporting individuals within each of these populations. Research has continued to support the benefits of online support groups for various populations.

Online Mentoring

There are advantages for the mentoring of students and underrepresented populations online, such as providing access to professional expertise and assistance that may not be available in the mentees' community (Knapczyk, Hew, Frey, & Wall-Marencik, 2005). Mentor-mentee interactions can occur more frequently and at convenient times online (Ensher, Thomas, & Murphy, 2001). The online mentoring process also offers greater privacy and anonymity.

Given that women are underrepresented in technology related fields, like science and engineering, Bennett and colleagues (1998) hypothesized that online mentoring may be beneficial to female adolescents interested in these fields. Their findings revealed that female adolescents felt supported by their mentors and that their confidence had increased. Studies examining online mentoring provide support for improving the professional development of mentees through online mentoring programs (Cheng et al., 2016; Hooley, Hutchison, & Neary, 2016; McAleer & Bangert, 2011). Online mentoring has been used for various populations, such as college students (e.g., Barczyk, Buckenmeyer, Feldman, & Hixon,

2011), new teachers (e.g., Hutchison & Colwell, 2011), women entering management positions (e.g., Loureiro-Koechlin & Allan, 2010), adjunct faculty (e.g., Rogers, McIntyre, & Jazzar, 2010), and mental health professionals (e.g., Lee & del Carmen Montiel, 2010). These studies support the use of online mentoring.

Electronic Fundraising/Donating

With the increasing availability of the internet, it has become a convenient method to make and collect charitable donations. The American Red Cross raised \$1.3 million for the Kosovo crisis through their website and \$110 million was raised through online donations to help the victims of the September 11 attacks (Waters, 2007). Donations raised online for the 2004 Asian Tusanmi exceeded those raised through traditional methods (i.e., calling to pledge money). After the 2010 Haiti earthquake, the donations raised online or through text messages exceeded \$30 million within the first month (Heath, 2010). Other researchers have examined the motivations behind donating to electronic charities. Olsen and colleagues (2001) found that successful nonprofit organizations raised funds by expressing to the donors how the gifts will help specific people. Furthermore, Bennett (2006; 2009) and Eller (2008) found that holding a positive attitude toward a population in need and having prior knowledge of the population were related to donating more money to that population.

Another mechanism for online donations is crowdfunding. Crowdfunding involves raising money from the general public to fund a project (Chen, Thomas, & Kohli, 2016). It works by having someone initiate the idea or project to be funded (Ordanini, Miceli, Pizzetti, & Parasuraman, 2011). Next, individuals support the idea and then the system for collecting donations brings these people together to execute the idea or project.

There were \$5.1 billion raised in 2013 through crowdfunding (Broderick, 2014). Regardless of where money is donated, people are making monetary contributions online to man-made and natural disasters, projects, and ideas.

Virtual Voluntarism

Many websites advertise various online voluntarism opportunities (Spencer, 2002). One of the first investigations to examine virtual voluntarism, the Virtual Volunteering Project, assessed the experiences of almost 200 agencies (Cravens, 2000). Findings from the project revealed that clearly written task descriptions and good communication are essential for keeping volunteers engaged with the organization. Haase and colleagues (2002) further investigated the characteristics of individuals who participated in virtual voluntarism. Volunteers were typically well educated, watched less television, and engaged in an active lifestyle. Other researchers (e.g., Butler et al., 2007) have examined ways to keep volunteers active through online organizations, with findings revealing that websites should be easily accessible to volunteers and that organizations should make their opportunities more fulfilling for volunteers.

Some researchers have considered how organizations can harness the power of social media for virtual voluntarism. In one study, Raja-Yusof and colleagues (2016) examined the activities that volunteers engaged in via social media. The major activity was knowledge sharing, with less time spent on training, fundraising, and problem-solving. Other research has focused on volunteer characteristics. In particular, Kim and Lee (2014) found that social capital and subjective norms predicted volunteering via SNS. The popularity of SNS has created many opportunities for volunteering. Furthermore, organizations should understand how they can attract and maintain volunteers committed to their organizational mission.

Other Electronic Technologies

The popularity of SNS, such as Facebook and Twitter, and online games has generated researchers' interest in prosocial acts through these technologies. Some investigations have provided evidence that gamers are more likely to help other players when they are altruistic and engage in offline prosocial behaviors (Ferguson & Garza, 2011; Molyneux et al., 2015; Wang & Wang, 2008). Sudzina, Razmerita, and Kirchner (2011) found that playing online games helped alleviate daily stress when these individuals received help through Facebook games. Other investigators have examined broader forms of online prosocial behaviors. For instance, face-to-face prosocial behaviors was associated positively with prosocial behaviors via SNS, chat programs, email, and text messages (Wright & Li, 2011).

VALUE OF ONLINE PROSOCIAL BEHAVIORS

There are benefits of online prosocial behaviors to the receiver. For example, receivers report health benefits from their participation in online support groups (e.g., Brennan et al., 1992), feel stress relief after receiving online gifts through Facebook games (e.g., Sudzina et al., 2011), and receive support and gain confidence from online mentors (e.g., Loureiro-Koechlin & Allan, 2010). Receivers benefit from the creation of relationships with users in online communities (Cummings et al., 2002). These online relationships offer receivers social support and advice. There is also evidence that online prosocial behaviors benefit the helper. Helpers gain personal satisfaction and support within their online communities and health benefits through virtual voluntarism (e.g., Butler et al., 2007; McAleer & Bangert, 2008; Mukherjee, 2010). Furthermore, contributors to open source software support groups reported learning and reputational benefits after helping others with questions (Hertel et al., 2003; Lakhani & Hippel, 2003).

THEORETICAL FRAMEWORKS

Social Cognitive Theory

The social cognitive theory suggests that online prosocial behaviors are learned by observing other people (Bandura, 1977). Newcomers may visit a discussion group for a while before posting, allowing them to get an idea about what messages are viewed as helpful. This information is then utilized when posting messages. If the post is rewarded with praise, newcomers might further contribute to the community. Supporting this idea, McKenna and Bargh (1998) found that a positively evaluated response contributed to newcomers' active involvement in the electronic discussion group. Positive reinforcement encourages contributions to the electronic community.

Co-Construction Theory

The co-construction theory was proposed to explain the construction of offline and online identities. Adolescents construct their online identities similar to their offline identities (e.g., Boneva, Quinn, Kraut, Kiesler, & Shlovski, 2006; Huffaker & Calvert, 2005; Whitlock, Powers, & Eckenrode, 2006). The co-construction theory has been applied to online prosocial behaviors. Wright and Li (2011) explained that online prosocial behaviors occur because people generalize their prosocial disposition to the digital environment.

Both the social cognitive and co-construction theories explain why individuals engage in and experience online prosocial behaviors. Both theories provide a valuable framework for understanding the continued involvement in prosocial behaviors and under which conditions people ask for help online.

SOLUTIONS AND RECOMMENDATIONS

While online prosocial behaviors may not receive as much attention as harmful online behaviors,

these behaviors are important to society. Online prosocial behaviors are important and need to be recognized for their benefits. The quality of help received online is an important factor for organizations to consider. In their review of prosocial behaviors through electronic discussion groups, Sproull and colleagues (2013) argued that the quality of help received online should be examined as it might help determine the value to the helper and receiver. Bad advice does not benefit the receiver and it can ruin the helper's reputation. Receiving good advice may relate to the desire to seek help again and determine whether individuals will give advice at another time. Online prosocial behaviors can promote social justice, assist individuals in need, heal and help the wounded, and elevate the level of good in the world. As such, online prosocial behaviors should be recognized for its benefit and promoted.

FUTURE RESEARCH DIRECTIONS

Most research on online prosocial behaviors is cross-sectional. Consequently, changes overtime in online prosocial behaviors are not well understood. Understanding attributional patterns of the helper and receiver is another important direction as attributional patterns relate to future behaviors. Such a research direction might allow researchers to predict when individuals will ask for help and when individuals will give help online.

Individual differences are also important to consider in regard to online prosocial behaviors. Certain characteristics, such as those with high social and emotional competence, empathetic concerns, self-esteem, and the desire to be part of the group may help to explain why individuals act prosocially online. Cultural values may also be important to consider. Collectivism may encourage online prosocial behaviors to maintain group cohesiveness, whereas individualism may encourage the use of online prosocial behaviors to achieve personal goals. Cross-cultural and intra-

cultural investigations linking cultural values to online prosocial behaviors may shed light on the characteristics associated with such behaviors.

CONCLUSION

There are a variety of opportunities to help and receive help in the multi-faceted online environment, including contributing to open source software and Wikis, online mentoring, virtual voluntarism, online donating/fundraising and crowdfunding, and providing support or receiving support through online support groups. People act prosocially toward others or ask for help online for a variety of reasons. Although research on online prosocial behaviors is developing, many areas await investigation. Online prosocial behaviors balance individuals' online experiences.

REFERENCES

- Antin, J. (2011). My kind of people? Perceptions about Wikipedia contributors and their motivations. *CHI 2011 Conference on Human Factors in Computing Systems*, (pp. 3411-3420). doi:10.1145/1978942.1979451
- Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Barcellini, F., Detienne, F., & Burkhardt, J. (2009). Participation in online interaction spaces: Design-use medication in an Open Source Software community. *International Journal of Industrial Ergonomics*, 39(3), 533-540. doi:10.1016/j.ergon.2008.10.013
- Barczyk, C., Bukenmeyer, J., Feldman, L., & Hixon, E. (2011). Assessment of a university-based distance education mentoring program from a quality management perspective. *Mentoring & Tutoring: Partnership in Learning*, 19(1), 5-24. doi:10.1080/13611267.2011.543567

- Bennett, D., Tsikalas, K., Hupert, N., Meade, T., & Honey, M. (1998). *The benefits of online mentoring for high school girls: Telementoring young women in science, engineering, and computing project – Year 3 evaluation*. Retrieved August 1, 2011, from <http://cct.edc.org/report.asp?id=77>
- Bennett, R. (2006). Innovation generation in charity promotional web sites: A suggested model and empirical test. *European Journal of Innovation Management*, 9(4), 347–369. doi:10.1108/14601060610707821
- Boneva, B., Quinn, A., Kraut, R., Kiesler, S., & Sklovski, I. (2006). Teenage communication in the instant messaging era. In R. Brynin & S. Kiesler (Eds.), *Information technology at home* (pp. 612–672). New York: Oxford University Press.
- Brennan, P., Moore, S., & Smyth, K. (1992). Alzheimers disease caregivers uses of a computer network. *Western Journal of Nursing Research*, 14(5), 662–673. doi:10.1177/019394599201400508 PMID:1529609
- Broderick, D. (2014). *Crowdfunding's untapped potential in emerging markets*. Retrieved from: <http://www.forbes.com/sites/hsbc/2014/08/05/crowdfundings-untapped-potential-in-emerging-markets/>
- Butler, B., Sproull, L., Kiesler, S., & Kraut, R. (2007). *Community effort in online groups: Who does the work and why?* Retrieved August 1, 2011, from <http://repository.cmu.edu/hcii/90>
- Chen, A. T. (2012). Exploring online support spaces: Using cluster analysis to examine breast cancer, diabetes and fibromyalgia support groups. *Patient Education and Counseling*, 87(2), 250–257. doi:10.1016/j.pec.2011.08.017 PMID:21930359
- Chen, S., Thomas, S., & Kohli, C. (2016). What really makes a promotional campaign succeed on a crowdfunding platform? Guilt, utilitarian products, emotional messaging, and fewer but meaningful rewards, drive donations. *Journal of Advertising Research*, 56(1), 81–94. doi:10.2501/JAR-2016-002
- Cheng, Y., Hanuscin, D., & Volkmann, M. J. (2016). Face-to-face, online, and hybrid mentoring in a professional development program. In L. J. Searby & S. K. Brondyk (Eds.), *In Best practices in mentoring for teacher and leader development* (pp. 73–92). Charlotte, NC: Information Age Publishing.
- Cornejo, R., Tentori, M., & Favela, J. (2013). Ambient awareness to strengthen the family social network of older adults. *Computer Supported Cooperative Work*, 22(2-3), 309–344. doi:10.1007/s10606-012-9166-2
- Coulson, N. S., & Greenwood, N. (2012). Families affected by childhood cancer: An analysis of the provision of social support within online support groups. *Child: Care, Health and Development*, 38(6), 870–877. doi:10.1111/j.1365-2214.2011.01316.x PMID:21916931
- Cravens, J. (2000). *About the virtual volunteering project*. Retrieved August 1, 2011, from <http://www.coyotecomunications.com/vv/vvabout.shtml>
- Cummings, J., Sproull, L., & Kiesler, S. (2002). Beyond hearing: Where real-world and online support meet. *Group Dynamics*, 6(1), 78–88. doi:10.1037/1089-2699.6.1.78
- Duranova, L., & Ohly, S. (2016). *Persistent work-related technology use, recovery and well-being processes*. New York, NY: Springer. doi:10.1007/978-3-319-24759-5

- Eisenberg, N., & Miller, P. (1987). The relation of empathy to prosocial and related behavior. *Psychological Bulletin*, 101(1), 91–119. doi:10.1037/0033-2909.101.1.91 PMID:3562705
- Eller, A. (2008). Solidarity.com: Is there a link between offline behavior and online donations? *Cyberpsychology & Behavior*, 11(5), 611–613. doi:10.1089/cpb.2007.0218 PMID:18785833
- Ensher, E., Heun, C., & Blanchard, A. (2003). Online mentoring and computer-mediated communication: New directions in research. *Journal of Vocational Behavior*, 63(2), 264–288. doi:10.1016/S0001-8791(03)00044-7
- Ferguson, C., & Garza, A. (2011). Call of (civic) duty: Action games and civic behavior in a large sample of youth. *Computers in Human Behavior*, 27(2), 770–775. doi:10.1016/j.chb.2010.10.026
- Finholt, T., & Sproull, L. (1990). Electronic groups at work. *Organization Science*, 1(1), 41–64. doi:10.1287/orsc.1.1.41
- Finn, J., & Lavitt, M. (1994). Electronic support/mutual aid groups for sexual abuse survivors. *Social Work with Groups*, 17(1/2), 21–46. doi:10.1300/J009v17n01_03
- Haase, A., Wellman, B., Witte, J., & Hampton, K. (2002). Capitalizing on the internet: Social contact, civic engagement, and sense of community. In B. Wellman & C. Haythornthwaite (Eds.), *The Internet and Everyday Life*. Oxford, UK: Blackwell. doi:10.1002/9780470774298.ch10
- Hassett, M., Lowder, C., & Rutan, D. (1992). Use of computer network bulletin board systems by disabled persons. *Proceedings of the Annual Symposium on Computer Application in Medical Care*, 151–155.
- Hertel, G., Niedner, S., & Herrmann, S. (2003). Motivation of software developers in Open Source projects: An Internet-based survey of contributors to the Linux kernel. *Research Policy*, 32(7), 1159–1177. doi:10.1016/S0048-7333(03)00047-7
- Hooley, T., Hutchison, J., & Neary, S. (2016). Ensuring quality in online career mentoring. *British Journal of Guidance & Counselling*, 44(1), 26–41. doi:10.1080/03069885.2014.1002385
- Huffaker, D., & Calvert, S. (2005). Gender, identity, and language use in teenage blogs. *Journal of Computer-Mediated Communication*, 10(2).
- Hutchison, A., & Colwell, J. (2011). Using a wiki to facilitate an online professional learning community for induction and mentoring teachers. *Education and Information Technologies*, 17(3), 273–289. doi:10.1007/s10639-011-9159-7
- Internet Live Stats. (2016). *Internet live stats*. Retrieved from: <http://www.internetlivestats.com>
- Kiesler, S., Zdaniuk, B., Lundmark, V., & Kraut, R. (2000). Troubles with the internet: The dynamics of help at home. *Human-Computer Interaction*, 15(4), 323–351. doi:10.1207/S15327051HCI1504_2
- Kim, Y., & Lee, W. (2014). Networking for philanthropy: Increasing volunteer behavior via social networking sites. *Cyberpsychology, Behavior, and Social Networking*, 17(3), 160–165. doi:10.1089/cyber.2012.0415 PMID:24102569
- Knapczyk, D., Hew, K., Frey, T., & Wall-Marencik, W. (2005). Evaluation of online mentoring for limited license teachers. *Teacher Education and Special Education*, 28(3/4), 207–220. doi:10.1177/088840640502800407
- Kuznetsov, S. (2006). Motivations of contributors to Wikipedia. *Association for Computing Machinery Special Interest Group on Computers and Society*, 36(2), 1–7.
- Lakhani, K., & von Hippel, E. (2003). How open source software works: free user-to-user assistance. *Research Policy*, 32(6), 923–943. doi:10.1016/S0048-7333(02)00095-1

- Lee, C., & del Carmen Montiel, E. (2010). Does mentoring matter in the mental health field? *Social Work in Mental Health*, 8(5), 438–454. doi:10.1080/15332981003744503
- Lieberman, M., & Goldstein, B. (2005). Self-help on-line: An outcome evaluation of breast cancer bulletin boards. *Journal of Health Psychology*, 10(6), 855–862. doi:10.1177/1359105305057319 PMID:16176962
- Loureiro-Koechlin, C., & Allan, B. (2010). Time, space and structures in an e-learning and e-mentoring project. *British Journal of Educational Technology*, 41(5), 721–735. doi:10.1111/j.1467-8535.2009.00975.x
- Malik, S., & Coulson, N. (2010). Coping with infertility online: An examination of self-help mechanisms in an online infertility support group. *Patient Education and Counseling*, 81(2), 315–318. doi:10.1016/j.pec.2010.01.007 PMID:20144521
- McAleer, D., & Bangert, A. (2011). Professional growth through online mentoring: A study of mathematics mentor teachers. *Journal of Educational Computing Research*, 44(1), 83–115. doi:10.2190/EC.44.1.e
- McKenna, K., & Bargh, J. (1998). Coming out in the age of the internet: Identity Demarginalization through virtual group participation. *Journal of Personality and Social Psychology*, 75(3), 681–694. doi:10.1037/0022-3514.75.3.681
- Mo, P. K. H., & Coulson, N. S. (2013). Online support group use and psychological health for individuals living with HIV/AIDS. *Patient Education and Counseling*, 93(3), 426–432. doi:10.1016/j.pec.2013.04.004 PMID:23669152
- Molyneux, L., Vasudevan, K., & de Zuniga, H. G. (2015). Gaming social capital: Exploring civic value in multiplayer video games. *Journal of Computer-Mediated Communication*, 20(4), 381–399. doi:10.1111/jcc4.12123
- Mukherjee, D. (2010). An exploratory study of older adults engagement with virtual volunteerism. *Journal of Technology in Human Services*, 28(3), 188–196. doi:10.1080/15228835.2010.508368
- Olsen, M., Keevers, M., Paul, J., & Covington, S. (2001, November). E-relationship development strategy for the nonprofit fundraising professional. *International Journal of Nonprofit and Voluntary Sector Marketing*, 6(4), 364–373. doi:10.1002/nvsm.161
- Ordanini, A., Miceli, L., Pizzetti, M., & Parasuraman, A. (2011). Crowd-funding: Transforming customers into investors through innovative service platforms. *Journal of Service Management*, 22(4), 443–470. doi:10.1108/09564231111155079
- Ragsdale, J. M., & Hoover, C. S. (2016). Cell phones during nonwork time: A source of job demands and resources. *Computers in Human Behavior*, 57, 54–60. doi:10.1016/j.chb.2015.12.017
- Raja-Yusof, R., Norman, A., Abdul-Rahman, S., Narzri, N., & Mohd-Yusoff, Z. (2016). Cyber-volunteering: Social media affordances in fulfilling NGO social missions. *Computers in Human Behavior*, 57, 388–397. doi:10.1016/j.chb.2015.12.029
- Raymond, E. (1999). *The cathedral & the bazaar: Musings on Linux and open source by an accidental revolutionary*. Cambridge, UK: O'Reilly Media.
- Rogers, C., McIntyre, M., & Jazzar, M. (2010). Mentoring adjunct faculty using the cornerstones of effective communication and practice. *Mentoring & Tutoring: Partnership in Learning*, 18(1), 53–59. doi:10.1080/13611260903448375
- Schneider, S., & Tooley, J. (2003). Self-help computer conferencing. *Computers and Bio-medical Research, an International Journal*, 19(3), 274–281. doi:10.1016/0010-4809(86)90022-4 PMID:3519069

- Sproull, L., Conley, C. A., & Moon, J. Y. (2013). The kindness of strangers: Prosocial behavior on the Internet. In Y. Amichai-Hamburger (Ed.), *The social net: Understanding our on-line behavior* (pp. 143–164). New York, NY: Oxford University Press. doi:10.1093/acprof:oso/9780199639540.003.0008
- Sproull, L., & Kiesler, S. (2005). *Public volunteer work on the Internet*. In *Transforming enterprise: the economic and social implications of information technology* (pp. 361–374). Cambridge, MA: The MIT Press.
- Stommel, W., & Meijman, F. J. (2011). The use of conversation analysis to study social accessibility of an online support group on eating disorders. *Global Health Promotion and Education*, 18(2), 18–26. doi:10.1177/1757975911404764 PMID:21596936
- Subrahmanyam, K., Smahel, D., & Greenfield, P. (2006). Connecting developmental processes to the Internet: Identity presentation and sexual exploration in online teen chatrooms. *Developmental Psychology*, 42(3), 395–406. doi:10.1037/0012-1649.42.3.395 PMID:16756432
- Sudzina, F., Razmerita, L., & Kirchner, K. (2011). *Strengthening weak ties through on-line gaming*. ACM Webscience. Retrieved August 1, 2011, http://www.websci11.org/fileadmin/websci/Posters/192_paper.pdf
- Thoren, E., Svensson, M., Tornqvist, A., Andersson, G., Carlbring, P., & Lunner, T. (2011). Rehabilitative online education versus internet discussion group for hearing aid users: A randomized controlled trial. *Journal of the American Academy of Audiology*, 22(5), 274–285. doi:10.3766/jaaa.22.5.4 PMID:21756843
- Thoren, E. M., Metze, B., Buhner, C., & Garten, L. (2013). Online support for parents of preterm infants: A qualitative and content analysis of Facebook premie groups. *Archives of Disease in Childhood. Fetal and Neonatal Edition*, 98(6), F534–F538. doi:10.1136/archdischild-2012-303572 PMID:23759520
- Wang, C., & Wang, C. (2008). Helping others in online games: Prosocial behavior in cyberspace. *Cyberpsychology & Behavior*, 11(3), 344–346. doi:10.1089/cpb.2007.0045 PMID:18537505
- Waters, R. (2007). Nonprofit organizations use of the internet: A content analysis of communication trends on the internet sites of the Philanthropy 400. *Nonprofit Management & Leadership*, 18(1), 59–76. doi:10.1002/nml.171
- Whitlock, J., Powers, J., & Eckenrode, J. (2006). The virtual cutting edge: The internet and adolescent self-injury. *Developmental Psychology*, 42(3), 407–417. doi:10.1037/0012-1649.42.3.407 PMID:16756433
- Wicks, P., Keininger, D. L., Massagli, M. P., de la Loge, C., Brownstein, C., Isojarvi, J., & Heywood, J. (2012). Perceived benefits of sharing health data between people with epilepsy on an online platform. *Epilepsy & Behavior*, 23(1), 16–23. doi:10.1016/j.yebeh.2011.09.026 PMID:22099528
- Wright, M. F., & Li, Y. (2011). The associations between young adults face-to-face prosocial behaviors and their online prosocial behaviors. *Computers in Human Behavior*, 27(5), 1959–1962. doi:10.1016/j.chb.2011.04.019

ADDITIONAL READING

- Amichai-Hamburger, Y. (2008). Potential and promise of online volunteer. *Computers in Human Behavior*, 24(2), 544–562. doi:10.1016/j.chb.2007.02.004
- Eller, A. (2008). Solidarity.com: Is there a link between offline behavior and online donations? *Cyberpsychology & Behavior*, 11(5), 611–613. doi:10.1089/cpb.2007.0218 PMID:18785833
- Ensher, E., Heun, C., & Blanchard, A. (2003). Online mentoring and computer-mediated communication: New directions in research. *Journal of Vocational Behavior*, 63(2), 264–288. doi:10.1016/S0001-8791(03)00044-7

Ferguson, C., & Garza, A. (2011). Call of (civic) duty: Action games and civic behavior in a large sample of youth. *Computers in Human Behavior*, 27(2), 770–775. doi:10.1016/j.chb.2010.10.026

Hertel, G., Niedner, S., & Herrmann, S. (2003). Motivation of software developers in Open Source projects: An Internet-based survey of contributors to the Linux kernel. *Research Policy*, 32(7), 1159–1177. doi:10.1016/S0048-7333(03)00047-7

Wang, C., & Wang, C. (2008). Helping others in online games: Prosocial behavior in cyberspace. *Cyberpsychology & Behavior*, 11(3), 344–346. doi:10.1089/cpb.2007.0045 PMID:18537505

Wright, M. F., & Li, Y. (2012). Prosocial behaviors in the cyber context. In Z. Yan (Ed.), *Encyclopedia of cyber behaviors*. Hershey, PA: IGI Global. doi:10.4018/978-1-4666-0315-8.ch028

KEY TERMS AND DEFINITIONS

Co-Construction Theory: This theory proposes that people's identities are constructed in similar ways online and offline.

Online Donating/Fundraising: Online donating and fundraising involves raising money or

accepting donations via electronic technologies and the internet.

Online Mentoring: Online mentoring involves someone (the mentee) receiving knowledge on a topic or career via electronic technologies and/or the internet by an expert (the mentor).

Online Prosocial Behaviors: Online prosocial behaviors involve helping, sharing, donating, and volunteering that is carried out via electronic technologies and the internet.

Online Support Groups: Online support groups allow people with the same challenges to connect via electronic technologies and/or the internet.

Open Source Software: Open source software is software in which the source code is available to others for modification or enhancement.

Prosocial Behaviors: Prosocial behaviors are acts that involve helping, sharing, donating, and volunteering.

Social Cognitive Theory: This theory states that people's knowledge and behaviors are related to their observations of others.

Virtual Voluntarism: Virtual voluntarism involves using electronic technologies and/or the internet to assist an organization.

Online Social Networking Behavior and Its Influence Towards Students' Academic Performance

Maslin Masrom

Universiti Teknologi Malaysia, Malaysia

Selisa Usat

Universiti Teknologi Malaysia, Malaysia

INTRODUCTION

Online social networking (OSN) is an online application that has grown rapidly in prevalence and popularity in recent years. Millions of Malaysian youngsters, teenagers, students (either at school, college or university's level) use OSN websites such as Facebook, Twitter, Myspace, Friendster and LinkedIn every day. Besides, Malaysian higher education also has adopted OSN on a wider scale and young Malaysians are very active users of OSN (Zakaria, Watson, & Edwards, 2010). In other words, the use of online social network has become necessary among younger generation.

The emergence of online social network has become a major trend over the years with growing popularity amongst the younger generation (Onuoha & Saheed, 2011). When online social networking becomes more popular, literature, however, suggests that there is an ongoing debate on the role of online social networks in the academic performance of users.

This article synthesizes the literature review concerning use of OSN in school, college and university, and academic performance of students. It offers definitions of OSN, followed by the status of current knowledge about the use of OSN and academic performance of students. Based on the secondary sources, it concludes the entry with recommendations for future research direction on the relationship between use of OSN and academic performance.

BACKGROUND

Online social networking sites (OSNs) are websites that give users a range of services based on web technologies that allow individuals to build a public or semi public profile with relationships system, have a list of other users with whom they share a connection, and finally, view and navigate through the list of users' connections with those who share a connection in the system (Boyd & Ellison, 2007).

OSNs are also defined as a range of activities enabled by social technologies or social media tools include blog, microblog, wiki, social networking site, video sharing site and online discussion board or forum, and operationalized by a group of people (Hamid et al., 2009). It enables users to socialize and create networks online. Examples of OSNs that are used on a regular basis by millions of people nowadays are Facebook, Twitters, MySpace, Friendster, Youtube and Skype. Ahmed and Qazi (2011) meanwhile cited that, the most successful and largest social networking site is Facebook with more than 500 million members had been found as active users of Facebook in July 2010. In Malaysia, social interaction in cyberspace by using social networking has been adapted by many people and has changed their communication (Mustafa & Hamzah, 2011).

We define online social networking as the latest online communication tool that allows users to create a public or private profile to interact

with people in their networks, share their profile information, communicate with others, and share data and information within that system.

The social technologies can support interaction among students by allowing them to actively participate in a discussion. The students can work collaboratively in an online social environment to solve problems with their peers, or to organize social events. The collaborative production's principle embedded in social technologies enable learners and teachers to share and publish artifacts produced as a result of the learning activity (for example, course materials such as course syllabus, course notes and assignments). In this regard, the use of social technologies has changed the demand of education.

Online social networking can be classified into five categories (Fraser & Dutta, 2008), namely, (i) egocentric networks - act as a platform to build a network of friends; (ii) web communities - collecting members with identity ties based on interest, gender, race, nation, religion and others; (iii) opportunistic web – the members gathered for business purpose or professional relationship using OSN site such as LinkedIn; (iv) passion-centric network – gathered people who share interests or hobbies (communities of interest); and (v) media-sharing site – this site is defined based on its contents (such as Youtube for those who want to share videos).

Students are increasingly using these social networks for friends' news feeds, personal updates, events and activities, notes, and messages. According to Michael, Robyn and Kate (2013) the widespread use of media among college students from texting to chatting on cell phones to posting status updates on Facebook may be taking an academic toll. They found media use, in general, was associated with lower grade point averages (GPAs) and other negative academic outcomes. Thus, this has led to a rise of below question:

Can the extensive use of OSNs in the younger generation affect students' academic performance?

The issue of whether OSN give positive or negative impact on students' academic performance is often dependent on the larger issues identified with the overall use of OSN sites such as individual self-discipline and self-regulation (behavior in terms of time spent on- and off-line), psychological effects and user adaptability (Egedegbe, 2013).

STUDENTS' PRACTICES OF OSN AND BEHAVIOR

The most used features by OSNs users are: uploading and sharing photos and videos, comments on other profiles, friends and private messages between users. In this context, users of OSN sites also share a number of documents, and interact and communicate with each other. What makes the OSNs unique, not because they allow students or users to meet others in the network, but because they make possible for students to manage and make visible their own social network (Fardoun et al., 2012).

The teenagers and youth especially students have embraced online social networking as one method to connect with their friends, share information and showcase their social lives (egedegbe, 2013). In general, the students spend a lot of time on OSN sites creating their profile, doing research concerning their academic assignments or works, chatting with friends and posting pictures of event they attended. Young et al. (2009) found that students keen on interacting with others in order to exchange information about their interests, to discuss about interesting or new topics and follow news about certain topics on OSN sites.

USE OF ONLINE SOCIAL NETWORKING

Online social networking can be regarded as a platform allowing teachers and students to com-

municate and collaborate on school subjects and projects outside the classroom (Khedo et al., 2012). According to Khedo et al. (2012), teachers can post school related works on these online communities and students can enrich their learning experiences by teaming up with their class mates to work on assignments and projects. These networks can also go beyond the classroom walls by uniting multiple classrooms from different schools, thus creating a richer environment for collaboration and knowledge sharing.

Hamid et al. (2011) have investigated the impacts of the use of OSN on enhancing student engagement and interaction from the students' perspectives. Their findings revealed that students showed positive inclination towards the use of OSN in facilitating their learning. The OSN benefited the students in enhancing their engagement and interaction, in promoting critical thinking, discovering new knowledge, tracking their own learning progress and being a platform to be more vocal.

In general, benefits of OSN use for students' education purpose among others are as follows:

1. The use of OSN has significant potential to support and enhance in-class teaching and learning (Arnold & Paulus, 2010; Techehaimanot & Hickman; 2010; Kabilan et al., 2010);
2. Increase student interaction with other students and their lecturers via electronic medium;
3. Increase student-to-student and student-to-lecturer interaction outside of traditional class time (Arnold & Paulus, 2010; Gray, Chang, & Kennedy, 2010);
4. Increase students' satisfaction with the course;
5. Improve students' learning and writing ability;
6. Provide greater access to information and information sources.

Other benefits associated with the use of OSN for other purposes are (Mozee, 2012; Connolly, 2011; Zwart et.al, 2011; Rosen, 2011):

1. Encouraging creativity among and between individuals and groups;
2. Creating a sense of belonging among users of common social media tools;
3. Providing more choices to promote engagement among different individuals and groups;
4. Reducing barriers to group interaction and communications such as distance and social or economic status;
5. Increasing the technological competency levels of frequent users of social media.

According to Rosen (2011), the daily use of all online social media technologies for those born between 1965-1979 (Generation X) consumed approximately 13 hours of social media per day; those born between 1980-1989 (Net Generation or Generation Y) consumed approximately 19 hours of social media per day; and those born between 1990-1999 ('I Generation') consumed approximately 20 hours of social media per day.

ACADEMIC PERFORMANCE

There has been considerable discussion regarding the frequent use of OSN sites by high school and college students, and the possible effect of those tools on students' academic performance (Connolly, 2011; Hargittai & Hsieh, 2010; Karpinski, & Duberstein, 2009). With the increased number of young users, it is presumed that OSNs may have some impacts on college students' academic performance, lifestyle, and personal development (Tham & Ahmed, 2011). The frequent involvement of students (at school, college and university level) who constitute a huge percentage of young people in online social networking has, however, led to deliberations on how their academic per-

formance could be affected due to their taking part in this activity.

Based on their findings, Onuoha and Saheed (2011) recommends that: (1) for students who find the use of online social networks distracting to academics, there is need to reduce the time spent on these networks; and (2) time spent on online social networks should be devoted more to academic than social matters in order to make maximal use of its education potentials.

Academic performance usually appears in research into education and educational psychology (Rouis, Limayem, & Salehi-Sangari, 2011). According to Rouis et al., there are two main approaches that offer different visions of academic performance, that is, (i) approaching a specific goal but avoiding adverse outcome offers an alternative, and (ii) goal achievement focused on the task or final results (Valle et al., 2009). Since OSNs use is considering as a leisure activity that interrupts students' academic performance, an effect is presumed on students' overall academic results or grades of the students. In this context, the conceptualization of academic performance is results focused, and therefore focuses only the final results or grades of the students.

THE RELATIONSHIP BETWEEN THE USE OF OSN AND ACADEMIC PERFORMANCE

There is a growing concern whether OSN use is contributing to a decline in academic performance by school, college or university students. Many students stated sparing time with media or online social networking sites take away time from completing homework. This issue can be viewed as a form of displacement. Students saw OSN use as being a hindrance to academic performance when it displaced the quantity of time spent studying. Karpinski (2009) showed that college students who use Facebook often spend less time studying and have lower grade point averages (GPAs) compared to students who have

not signed up for the social networking website. Thus, the relationship between the use of OSN and academic performance of student is one issue that is essential to be measured based on various degrees the level of OSN use and the academic performance level of the user.

Regard to the relationship between the use of OSN and academic performance, the findings of some past studies are listed as following:

- Stollak, Vandenberg, Burklund, and Weiss (2011), Rouis, Limayem, and Salehi-Sangari (2011), Karpinski and Duberstein (2009), and Canales, Wilbanks, and Yeoman (2009) found a negative effect; that is, higher use of OSN typically leading to lower academic performance as measured by grades;
- Ahmed & Qazi (2011) and Hargittai and Hsieh (2010) found either no-to-little relationship between the use of OSN and student academic performance;
- Pasek and Hargittai (2009), Junco, Heiberger, and Loken (2011), and Rizzuto, LeDoux, and Hatala (2009) found a positive effect; that is, the use of OSN leading to an increase in student academic performance;
- In his article, Social Networks in Nigeria, Oche (2010), stated that the recently released results for the National Examination Council (NECO) showed that 87% of the candidates failed English Language and Mathematics which the author attributed to use of Facebook;
- Vanden Boogart (2006), in a similar study also found out that lower GPAs is associated with heavy use of Facebook (i.e., more time spent on Facebook);
- Kolek and Saunders (2008) in a study of students from a public Northeast research university found out that there was no correlation between Facebook use and GPA;
- Hahhad M. (2013) study using sample which consisted of undergraduate students

enrolled in fall 2012 courses at California State University of Monterey Bay concludes that there was a weak correlation between GPA and time in regards to the three variables age, gender and major;

- Junco (2015) in his recent studies on undergraduate students showed that seniors spent significantly less time on Facebook and spent significantly less time multitasking with Facebook than students at other class ranks. Time spent on Facebook was significantly negatively predictive of GPA for freshmen but not for other students. Multitasking with Facebook was significantly negatively predictive of GPA for freshmen, sophomores, and juniors but not for seniors.

Other factors that are also should be considered on measuring the relationship of OSN and students' academic performance include the frequency and intensity of social media usage; the personality type of the social media user (for example, extroverted, introverted); the socioeconomic and cultural background of the user; the ability of the user to self-regulate their behavior in terms of time spent on- and off-line; the differences among the social networking sites in terms of their content, purpose, and structure; and the academic ability of the social media user (Mozee, 2012).

CONCLUSION

Online social networking sites are not only emerging as important tools in today's schools, but they also provide very popular out-of-school computer activities among students and/or young adults. It represents a potential technology that can be exploited to enhance learning in order to help students in their education. The academic performance can have long-term consequences for society and the individual in terms of overall quality of life

(Phusavat, Ketsarpong, Ooi, & Shyu, 2012). The research literature indicated that for some students the use of social media can be beneficial and/or harmful to their academic performance, and for other students it appears to have no effect. The school teachers and college and university's lecturer need to look at the relationship between students' practices on social network sites and their academic learning, and to keep creating a vision of continual technology integration in their classrooms (Greenhow & Burton, 2011). Online networks can be used for academic purposes such as peer-to-peer knowledge sharing and collaboration. In this respect, the school teachers and college or university's lecturer need to recognize the power of OSNs in college students' academic learning. In future, using mixed methods educators could understand how online social networking changes their educational attainment and students' academic performance.

FUTURE RESEARCH DIRECTIONS

Future research should (1) identify the level use of OSN among students; (2) examine the effect of OSN' use to academic performance as measured by grades; and (3) investigate the relationship between the use of OSN and student academic performance.

REFERENCES

- Ahmed, I., & Qazi, F. T. (2011, June). A look out for academic impacts of Social Networking Sites (SNSs): A student based perspective. *African Journal of Business Management*, 5(12), 5022–5031.
- Arnold, N., & Paulus, T. (2010). Using a social networking site for experiential Learning: Appropriating, lurking, modeling and community building. *The Internet and Higher Education*, 13(4), 188–196. doi:10.1016/j.iheduc.2010.04.002

- Boyd, D., & Ellison, N. (2007, October). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1), 210–230. doi:10.1111/j.1083-6101.2007.00393.x
- Canales, C., Wilbanks, B., & Yeoman, A. (2009, Spring). Facebook usage in relation to personality and academic performance. *Modern Psychological Studies*.
- Connolly, M. (2011, October). *Benefits and drawbacks of social media in education*. Wisconsin Center for Education Research. Retrieved from http://www.wcer.wisc.edu/news/coverstories/2011/benefits_and_drawbacks.php
- Egedegbe, N. (2013). *The effect of social networking sites on students' academic performance in Girne American University, North Cyprus*. Retrieved from <http://ydemokrat.blogspot.com/2013/05/the-effect-of-social-networking-sites.html>
- Fardoun, H. M., Alghazzawi, D., López, S., Penichet, V., & Gallud, J. (2012). Online social networks impact in secondary education. In *Proceedings of International Workshop on Evidence-Based Technology Enhanced Learning* (pp. 37–45). Springer Berlin/Heidelberg. doi:10.1007/978-3-642-28801-2_5
- Fraser, M., & Dutta, S. (2008). *Throwing sheep in the boardroom: How online social networking will transform your life*. In *Work and World*. London: Wiley.
- Gray, K., Chang, S., & Kennedy, G. (2010). Use of social web technologies by international and domestic undergraduate students: Implications for internationalizing learning and teaching in Australian universities. *Technology, Pedagogy and Education*, 19(1), 31–46. doi:10.1080/14759390903579208
- Greenhow, C., & Burton, L. (2011). Help from my friends: Social capital in the social network sites of low-income students. *Educational Computing Research*, 45(2), 223–245. doi:10.2190/EC.45.2.f
- Haddad, M. (2013). *Correlation between Minutes Spent on Facebook and Students' Academic Performance*. Retrieved from <https://csumb.edu/sites/default/files/images/st-block-21-1426021047146-raw-699marwanhaddadresearch.pdf>
- Hamid, S., Kurnia, S., Waycott, J., & Chang, S. (2011). *Exploring Malaysian students' perspectives of Online Social Networking (OSN) use for higher education*. Retrieved from www.proceedings.com.au/isana/2011/paper_Hamid.pdf
- Hamid, S., Waycott, J., Chang, S., & Kurnia, S. (2011). Appropriating Online Social Networking (OSN) activities for higher education: Two Malaysian cases. In G. Williams, P. Statham, N. Brown, & B. Cleland (Eds.), *Changing Demands, Changing Directions*. Proceedings Ascilite Hobart 2011 (pp. 526–538). Ascilite.
- Hargittai, E., & Hsieh, Y.-P. (2010, June). Predictors and consequences of differentiated practices on social network sites. *Information Communication and Society*, 13(4), 515–536. doi:10.1080/13691181003639866
- Junco, R. (2015). Student class standing, Facebook use, and academic performance. *Journal of Applied Developmental Psychology*, 36, 18–29. doi:10.1016/j.appdev.2014.11.001
- Junco, R., Heiberger, G., & Loken, E. (2011). The effect of Twitter on college student engagement and grades. *Journal of Computer Assisted Learning*, 27(2), 119–132. doi:10.1111/j.1365-2729.2010.00387.x
- Kabilan, M. K., Ahmad, N., & Zainol Abidin, M. J. (2010). Facebook: An online environment for learning of English in institutions of higher education? *The Internet and Higher Education*, 13(4), 179–187. doi:10.1016/j.iheduc.2010.07.003
- Karpinski, A. C., & Duberstein, A. (2009, April). Facebook® and academic performance. *Journal Computers in Human Behavior*, 26(6), 1237–1245.

- Khedo, K. K., Ally Elaheebocus, S. M. R., Mock-toolah, A., & Suntoo, R. (2012). Online social networking as a tool to enhance learning in the Mauritius education system. *Journal of Emerging Trends in Computing and Information Sciences*, 3(2), 907–912.
- Kolek, E. A., & Saunders, D. (2008). Online disclosure: An empirical examination of undergraduate Facebook profiles. *NASPA Journal*, 45(1), 1–25. doi:10.2202/0027-6014.1905
- Michael, P. C., Robyn, L. F., & Kate, B. C. (2013, March 26). Female college students' media use and academic outcomes: Result from a longitudinal cohort study. *Emerging Adulthood*.
- Mozee, S. (2012). The impact of social media use on academic performance at One Urban University: A pilot study. *Online Journal of Rural and Urban Research*, 2(1), 1–10.
- Mustafa, S. E., & Hamzah, A. (2011). Online social networking: A new form of social interaction. *International Journal of Social Science and Humanities*, 1(2), 96–104. doi:10.7763/IJSSH.2011.V1.17
- Oche, M. (2010). *Social networking and the future of Nigerian students*. Retrieved April 20, 2011 from http://www.leadershipeditors.com/ns/index.php?option=com_content&view=article&id=16800:michaeloche&catid=34:features&Itemid=152
- Onuoha, U. D., & Saheed, F. O. (2011). Perceived influence of Online Social Networks on academic performance: A study of undergraduates in selected universities in Ogun State, Nigeria. African Journal Online. *The Information Manager*, 11(1&2), 6–13.
- Pasek, J., & Hargittai, E. (2009, May). Facebook and academic performance: Reconciling a media sensation with data. *First Monday*, 14(5). doi:10.5210/fm.v14i5.2498
- Phusavat, K., Ketsarpong, S., Ooi, K.-B., & Shyu, S. (2012). Sustaining higher education reforms: Knowledge and policy implications learned from Thailand. *International Journal of Educational Management*, 26(3), 284–301. doi:10.1108/09513541211213354
- Rizzuto, T. E., LeDoux, J., & Hatala, J. P. (2009). Its not just what you know, Its who you know: Testing a model of the relative importance of social networks to academic performance. *Social Psychology of Education*, 12(2), 175–189. doi:10.1007/s11218-008-9080-0
- Rosen, L. (2011, August). *Poke me: How Social Networks can both help and harm our kids*. Washington, DC: American Psychological Association.
- Rouis, S., Limayem, M., & Salehi-Sangari, E. (2011). Impact of Facebook usage on students' academic achievement: Role of self-regulation and trust. *Electronic Journal of Research in Educational Psychology*, 9(3), 961–994.
- Stollak, M. J., Vandenberg, A., Burklund, A., & Weiss, S. (2011, February). *Getting social: The impact of Social Networking usage on grades among college students*. Paper presented at AS-BBS Annual Conference, Las Vegas, NV.
- Tham, J., & Ahmed, N. (2011). The Usage and implications of social networking sites: A survey of college students. *Journal of Interpersonal, Intercultural and Mass Communication*, 2(1), 1–11.
- Valle, A., Rodriguez, S., Cabanch, R. G., Nunez, J. C., Gonzales-Pianda, J. A., & Rosario, P. (2009). Academic goals: Historical and conceptual perspectives and educational implications. *Electronic Journal of Research in Educational Psychology*, 7(3), 1073–1106.
- Vanden Boogart, M. R. (2006). *Uncovering the social impacts of Facebook on a college campus* (Unpublished Masters Thesis). Kansas State University.

Young, S., Dutta, D., & Dommety, G. (2009). Extrapolating psychological insight from Facebook profiles: A study of religion and relationship status. *Cyberpsychology & Behavior*, 12(3), 347–350. doi:10.1089/cpb.2008.0165

Zakaria, M. H., Watson, J., & Edwards, S. L. (2010). Investigating the use of Web 2.0 technology by Malaysian Students. *Multicultural Education & Technology Journal*, 4(1), 17–29. doi:10.1108/17504971011034700

Zwart, M. d., Lindsay, D., Henderson, M., & Phillips, M. (2011). *Teenageers, legal risks and social networking sites*. Victoria, Australia: Victoria Law Foundation.

ADDITIONAL READING

Anonymous. (2010). *Student grades not affected by social networking*. Retrieved from University of New Hampshire: <http://www.newswise.com/articles/student-grades-not-affected-by-social-networking-new-researchfinds>

Halverson, E. R. (2011). Do social networking technologies have a place in formal learning environments? *Horizon*, 19(1), 62–67. doi:10.1108/10748121111107717

Hamid, S., Waycott, J., Chang, S., & Kurnia, S. (2011). Appropriating Online Social Networking (OSN) activities for higher education: Two Malaysian cases. In G. Williams, P. Statham, N. Brown, & B. Cleland (Eds.), *Changing Demands, Changing Directions. Proceedings ascilite Hobart 2011* (pp. 526–538).

Kirschner, P. A., & Karpinski, A. C. (2010). Facebook and academic performance. *Computers in Human Behavior*, 26(6), 1237–1245. doi:10.1016/j.chb.2010.03.024

Kite, S., Gable, R., & Filippelli, L. (2010). Assessing middle school students knowledge of conduct and consequences and their behaviors regarding the Use of Social Networking sites. *Clearing House (Menasha, Wis.)*, 83(5), 158–163. doi:10.1080/00098650903505365

Notley, T. M. (2008). Online network use in schools: Social and educational opportunities. *Youth Studies Australia*, 27(3), 20–29.

Santos, I., Hammond, M., Durli, Z., & Chou, S.-Y. (2009). Is there a role for social networking sites in education? In A. Tatnall & A. Jones (Eds.), *Education and Technology for a Better World: Proceedings of the 9th IFIP TC 3 World Conference on Computers in Education* (pp. 321–330). Berlin, Germany: Springer. doi:10.1007/978-3-642-03115-1_34

Selwyn, N. (2009). Faceworking: Exploring students education-related use of Facebook. *Learning, Media and Technology*, 34(2), 157–174. doi:10.1080/17439880902923622

Zwart, M. d., Lindsay, D., Henderson, M., & Phillips, M. (2011). *Teenagers, legal risks and social networking sites*. Victoria, Australia: Victoria Law Foundation.

KEY TERMS AND DEFINITIONS

Academic Performance: The academic performance is defined by students' reporting of past semester CGPA/GPA and their expected GPA for the current semester. The grade point average or GPA is now used by most of the tertiary institutions as a convenient summary measure of the academic performance of their students. The GPA is a better measurement because it provides a greater insight into the relative level of performance of individuals and different group of students.

Facebook: Facebook is a popular online social networking site or service that offers an online platform on which users create profiles, generate and share contents and information, and interact with other contacts.

Friendster: Friendster is an online social network site where we can meet people online, find new and old school friends and play games.

MySpace: MySpace is online social networking sites that allow users to create webpages to interact with other users. Then, users are able to create blogs, upload videos and photos, and design profiles to showcase their interests and talents.

Online Social Networking Sites: Online social networking sites (OSNs) generally refer to Internet-based locations that allow individuals and groups to interact. Specifically, it refers to those Internet-based services that: promote online social interaction between two or more persons within a bounded system for the purposes of friendship, meeting other persons, and/or exchanging informa-

tion; contains a functionality that lets users create public or semi-public personal profile pages that contain information of their own choosing; serves as a mechanism to communicate with other users; and contains mechanisms that allow users to search for other users according to some specific criteria. Examples of the most visited OSNS are Facebook, Twitter, Youtube, Friendster, Myspace, and LinkedIn.

Online Social Networking: Online social networking is defined as the latest online communication tool that allows these users to create a public or private profile to interact with people in their networks.

Social Media: Technologies that facilitate social interaction, make possible collaboration, and enable deliberations across stakeholders.

Twitter: Twitter is an online social networking and microblogging site or service that enable users to send and read tweets (text messages).

Parental Mediation of Adolescent Technology Use

S

J. Mitchell Vaterlaus

Montana State University, USA

INTRODUCTION

Adolescence is a developmental time period marked by physical (e.g., puberty), psychological (e.g., identity formation), and psychosocial (e.g., negotiation of parent-child relationships) changes (Erikson, 1950, Grotevant & Cooper, 1986). Puberty is typically seen as the beginning of the adolescent developmental time period, although there is some disagreement regarding when adolescence ends (Bynner, 2007). Hall (1904), an early developmental scholar, posited that adolescence continued into the early twenties, and some current scholars have supported this proposition because the developmental tasks of adolescence are continuing into the mid-twenties (Shwartz, Côté, & Arnett, 2005). Consistent with these historical and contemporary conceptualizations (Bynner, 2007; Hall, 1904; Shwartz et al., 2005), in this chapter adolescence is defined as the time period between the onset of puberty and until the individual reaches their mid-twenties.

Contemporary adolescents have grown up with access to a variety of technologies. The majority of adolescents (13-17 years old) in the United States have access to cell phones (88%) and computers (87%), and 92% of adolescents report going online daily (Lenhart, 2015). Adolescents are the most frequent users of social media and 71% report using more than one social media site (Lenhart, 2015). Parents, too, are technology consumers with 91% reporting that they use the internet and 83% using social media. Technology appears to have become a normative aspect of family life, but parents and adolescents use technology differently (Vaterlaus & Tulane, 2015). Parents primarily use technol-

ogy for instrumental purposes (e.g., cell phones to track location of child), while adolescents use technology for social reasons (e.g., cell phones to build social relationships).

As interactive technologies (e.g., cell phones, social media) have become a common feature during adolescence, parents have reported concerns about these technologies. Concerns about adolescent technology use have included worry about the potential psychological outcomes from viewing content (e.g., violent or pornographic), the risk of victimization, and access to illegal activities (e.g., illegal file sharing; boyd & Hargittai, 2013). Because of these concerns for children and adolescents, some limited government policy has emerged regulating website use among minors (Costello, McNiel, & Binder, 2016). For example, the United States' Children's Online and Privacy Protection Act (COPPA; Federal Trade Commission, 2016) requires parental permission for websites/online services to collect or use personal information from children under the age of 13. This particular policy has been criticized because it does not account for the privacy risks for adolescents older than 13 (Costello et al., 2016) and it is difficult to regulate with minors commonly falsifying their ages to access websites (O'Keeffe & Clarke-Pearson, 2011). "As legal regulations in this [technological] sphere are difficult to formulate and enforce, policy makers rely substantially on increasing risk awareness among parents and delegating to them the responsibility for protecting children from online risks." (Kirwil, 2009, p. 394). To mitigate the potential negative effects and facilitate the potential positive effects of adolescent interactive technology use, some parents

DOI: 10.4018/978-1-5225-2255-3.ch616

have made attempts to be involved in their adolescents' technology use. These parental attempts have been researched under the term *parental mediation*, which refers to parental interventions and interactions with their adolescents regarding technology use (Livingstone & Helsper, 2008; Vaterlaus, Beckert, Tulane, & Bird, 2014).

BACKGROUND

The early research on parental mediation focused on parent's attempts to mediate children's television viewing (Austin, 1990; Nathanson, 1999). Dr. Amy Nathanson (Nathanson, 1999), of The Ohio State University, has been a leader in identifying the methods parents have used to mediate children's television viewing. Dr. Sonia Livingstone (Livingstone & Bober, 2006), of the London School of Economics and Political Science, is a pioneer in researching the role of parental mediation with interactive technologies—publishing on parental mediation of children and adolescent's internet use. Additionally, Dr. Laura Padilla Walker and Dr. Sarah Coyne (Padilla-Walker & Coyne, 2011), of Brigham Young University, have made recent contributions to the understanding of the implementation of parental mediation with adolescent interactive technology use.

The evolution of technology has led to technological convergence, which allows a single media source to be accessed from several devices. Brooks-Gunn & Donahue (2008) explained:

Thanks to convergence, a teen can watch a television show on a computer long after the show has aired on television and can use a cell phone to surf the internet. Children, particularly adolescents, thus have almost constant access to media—often at times and in places where adult supervision is absent. (p. 3)

Because interactive technologies facilitate private access to a variety media and digital social opportunities, parents have voiced their concerns

about adolescent interactive technology use (Boyd & Hargittai, 2013) and some have sought ways to reduce the potentially negative effects of adolescents use through parental mediation (Livingstone & Helsper, 2008; Vaterlaus et al., 2014). The term parental mediation represents several different strategies for parental involvement in adolescent technology use. It is important to note that not all researchers in this area of study have adopted the term parental mediation. For example, some prefer “proactive media monitoring” because the implemented parental strategies “may not mediate media effects rather, they may prevent them from occurring in the first place or may protect [adolescents] against them” (Padilla-Walker, Coyne, Fraser, Dyer, & Yorgason, 2012, p. 1154). Regardless of the broader term used to describe parental involvement in their children's technology use, researchers have identified congruent strategies parents use which include: active mediation, restrictive mediation, and co-viewing (Livingstone & Helsper, 2008; Nathanson, 2001). This chapter summarizes different parental mediation strategies, details parent and adolescent perceptions related to parental mediation, and explores the challenges in implementing parental mediation.

CURRENT SCIENTIFIC KNOWLEDGE IN PARENTAL MEDIATION

Parental Mediation Strategies

Parental mediation of adolescent interactive technology use has been investigated with both quantitative (Livingstone & Helsper, 2008) and qualitative (Vaterlaus et al., 2014) research methods. The term parental mediation is representative of a range of strategies parents use to influence their adolescent's technology use. Nathanson's (1999, 2001) early work with parental mediation and children's television viewing provided some broad strategies that parents implement, which include: (a) active mediation, (b) restrictive mediation, and (c) co-viewing. Parents implement

variations of these television parental mediation strategies for adolescent interactive technology use (Livingstone & Helsper, 2008; Vaterlaus et al., 2014). The adapted parental mediation strategies for adolescent interactive technology use can be discussed in terms of active mediation, restrictive mediation, and monitoring strategies.

Active Mediation

The process of active mediation has been described as parents engaging adolescents in “discussions regarding questionable content in the media, and offer[ing] strategies and ways in which children might be more aware of this content, or might avoid it” (Padilla-Walker et al., 2012, p. 1154). At the heart of this strategy is parent-child discussion. The parental mediation strategy of co-viewing (e.g., watching television with a child; Nathanson, 2001) or co-use may not intuitively fit within broad category of active mediation because co-use could include little discussion. Livingstone and Helsper (2008) provide support for inclusion of co-use within the category of active mediation:

Parents and children may watch television together with little conversation (i.e., co-use), perhaps with the parent also reading the paper while the child does their homework in front of the screen. However, to sit together in front of the computer while the child goes online, or even to be in the same room, makes co-use more active, for conversation about the online activity, including interpretive or evaluative comments or guidance, is more likely. (p. 589)

Livingstone and Helsper (2008) reported that active mediation was the most widespread technique parents ($n = 905$) implemented for mediating adolescent internet use. Active mediation was evidenced by parents talking to their adolescent about their internet use or by watching the screen and staying in the vicinity of the computer while the adolescent was on the internet. Vaterlaus and colleagues (2014) asked parents ($n = 80$) and

adolescents ($n = 113$) to describe how parents mediated adolescent interactive technology use (i.e., internet and cell phone use). Parents and adolescents reported that the second most frequent mediation technique was active mediation—parents mediated by asking their adolescent who they were communicating with or what they were accessing via technology. Participants also indicated that parents were friends with their children on social media sites and in some instances used the internet together. Padilla-Walker and colleagues (2012) reported in their longitudinal study that active mediation was the most common approach among a sample of 276 of mother-adolescent dyads, but the mother’s use of the strategy decreased as the adolescent aged. Parenting qualities were found to be predictive of the use of active mediation in a study with 478 families (Padilla-Walker & Coyne, 2012). Parents who demonstrated connection (e.g., responsiveness to adolescents needs) and regulation (e.g., emphasize the importance of rules) were more likely to implement increased levels active mediation when compared to other parents.

Restrictive Mediation

Parents who use restrictive mediation block access to interactive technology or implement limits on adolescents interactive technology use (Vaterlaus et al., 2014). In a qualitative study, parents and adolescents reported that restrictive mediation included preventing adolescent ownership of cellphones or blocking specific features (e.g., text messaging, apps) on technology (Vaterlaus et al., 2014). Others reported implementing content (e.g., internet filters), time (e.g., internet access blocked after a certain time), and location restrictions (e.g., computer located in a public place). Padilla-Walker and Coyne (2011) identified that parents who had the parenting quality of regulation (e.g. emphasize the importance of the rules) implemented higher levels of restrictive mediation. Also, in a longitudinal study with mother-adolescent dyads, Padilla-Walker and colleagues (2012) found that

restrictive mediation strategies decreased as the adolescent aged.

Livingstone and Helsper (2008) discussed restriction in terms of interaction restriction (i.e., preventing peer-to-peer social facilitation through technology by banning email, instant messaging, or like features) and technical restriction (i.e., blocking activities through the use of filters or parental controls). The practice of restricting interactions was the only parental mediation technique Livingstone and Helsper (2008) identified that was associated with decreasing adolescents' online risks. The researchers cautioned that the sociability facilitated by the internet is a major appeal among adolescents, so blocking interactions may be keeping "teenagers safe at a cost" (p. 597). Lee (2013) identified similar benefits for the implementation of restrictive mediation with Korean mother-adolescent dyads ($n = 566$). Maternal implementation of restrictive mediation was associated with decreased time spent online and lower online risks among adolescents.

Monitoring

The practice of monitoring is defined as "checking up on the child's [internet] activity, covertly or overtly, after use" (Livingstone & Helsper, 2008, p. 589). This was the most frequently reported parental mediation technique in the Vaterlaus et al. (2014) study—participants reported that monitoring involved parents checking adolescents' digital footprint, which included checking things like internet history, text messages, and pictures. In terms of cell phone use parents and adolescents specifically talked about parents monitoring the number of minutes and amount of data used on the cell phone bill. In their comparison between online risks and various parental mediation strategies, Livingstone and Helsper (2008) found no significant relationship between monitoring internet use and decreased online risks for adolescents. Further, there is some evidence that the more parents monitor adolescent internet use the more adolescents do to prevent parent's attempts

to uncover their online activities (Livingstone & Bober, 2006).

Perceptions of Parental Mediation

Parent-Adolescent Discrepancies

Differences have been documented in reports of parental mediation of adolescent internet use between parents and adolescents in the same household (Livingstone & Bober, 2004; Wang, Bianchi, & Raley, 2008). Vaterlaus and colleagues (2014) asked parents and adolescents to report the implementation of parental mediation of adolescent internet use within the same household. Mothers and fathers reported significantly more parental mediation of adolescent internet use when compared to their adolescents' reports. Similarly, Fletcher and Blair (2014) reported that some families experience conflicted parental authority when technology rules are not clearly communicated to the adolescent—resulting in parents and adolescents having different perspectives of parental mediation of adolescent technology use. Livingstone and Bober (2004) postulated that these differences between parents and adolescents may result from generationally different perceptions of the technology rules in the household. Parents may report the official rules without addressing the potential exceptions, while adolescents interpret the rules differently because of the exceptions granted by parents. Adolescents seem to follow their own interpretations of parental technology rules.

Privacy, Authority, and Trust

Parents and adolescents have different perspectives regarding the purpose and use of interactive technologies (Vaterlaus & Tulane, 2015). Oksman and Turtiainen (2004) reported discrepancies between Finnish parents' and adolescents' motivations for cell phone ownership. Parents view adolescent cell phone ownership as a way to increase security (e.g., monitoring teen's location) and adolescents are motivated by wanting to use the device to stay

in contact with their friends. New technologies are often preferred by adolescents because of the privacy they afford. For instance, Oksman and Turtiainen (2004) concluded, “For teenagers, text messaging is a quiet and simple way to maintain their social network without their parents’ knowledge” (p. 336). Adolescents may oppose parental mediation strategies because the practice could be perceived as a violation of privacy:

Children relish the opportunities the internet affords them—for identity play, relationships, exploration and communication—and may not wish to share this experience with their parents. (Livingstone & Bober, 2004, p. 43)

Granting adolescents access to interactive technology and then implementing parental mediation could also be perceived as a violation of parental trust. Vaterlaus and colleagues (2014) qualitatively explored adolescents’ perceptions of parental mediation practices. Approximately half of the adolescents (47%) indicated that parents had the authority and responsibility to mediate their adolescents’ interactive technology use, but then explained how parental trust needed to be part of the equation. Adolescents clarified that parental mediation techniques should be implemented when an adolescent gained access to a new technology and then mediation practices should decrease as the adolescent gained parental trust. Adolescents also recommended that parents communicate with them openly about parental mediation and include adolescents in establishing the parental mediation practices in their home.

Challenges in Parental Mediation

Digital Generation Gap

Parents do experience challenges in their attempts to regulate their youth’s media use because of the privacy afforded by new technologies and the technological expertise required with the evolving technologies (Livingstone & Helsper,

2008). Vaterlaus and Tulane (2015) concluded that there are generational differences between parents and adolescents in their conceptualizations of the purpose and use of technology (e.g., instrumental vs social purposes), where they learn about technology (e.g., parents often rely on children to teach them), and in knowledge of how to use technology. A difference between parents and adolescents in knowledge about how to use technology is referred to as a digital generation gap (Livingstone, 2003, Vaterlaus & Tulane, 2015). One study documented that late adolescents ($n = 555$) perceived that they knew significantly more about how to use interactive technology (i.e., cell phones, social networking, email, and video chat) than their parents (Vaterlaus, Jones, & Tulane, 2015). Parents ($n = 604$) in the study also rated their own knowledge about interactive technologies significantly lower than their late adolescents’ knowledge.

The presence of a digital generation gap could challenge the traditional top-down (parent to child) socialization patterns in families, which may make it challenging to implement parental mediation strategies (Correa, 2014). Parents often report that they do not know what their children are doing online and feel at a loss in how to help them use the internet in positive ways (Livingstone & Bober, 2004). Having more technological knowledge may allow adolescents to evade parental attempts at mediation. For example, a study conducted with teens in the United Kingdom, reported that the teens had implemented efforts to maintain their privacy from their parents through deleting emails, clearing internet history, or minimizing screens when someone came into the room (Livingstone & Bober, 2004). Parents’ lower levels of technological expertise may limit the types of mediation that they implement or prevent the implementation of parental mediation at all. Parents implement a variety of attempts to mitigate the potentially harmful impacts of adolescents’ internet use, but it appears that parents have a preference of using co-use over approaches that may require more technological expertise (i.e.,

restriction or monitoring; Livingstone & Bober, 2004). Fletcher and Blair (2014) interviewed 20 adolescents (seventh grade students) and their mothers regarding mothers' authority over their adolescents' interactive technology use. Mothers who reported limited expertise with a technology were more likely to report that they did nothing to mediate their adolescent's use of that technology.

Parental Mediation in Later Adolescence

One of the major tasks during adolescent development is developing autonomy from parents (i.e., becoming independent while still maintaining connection with parents; Zimmer-Gembeck & Collins, 2003). Following completion of high school many adolescents pursue post-secondary education. Approximately 20.2 million students were expected to attend American colleges and universities in the fall of 2015 (National Center for Education Statistics, 2015). This often involves moving away from home, but does not mean adolescents stop relying on their parents for support in various areas. Adolescents over the age of 18 continue to rely on their parents for economic and emotional support (Aquilino, 2006) and adolescents may continue to benefit from parental involvement in their interactive technology use (Vaterlaus & Tulane, 2015). In their review of technology use among late adolescents (18-25 years old), Coyne, Padilla-Walker, and Howard (2013) suggest that it would be challenging to monitor adolescents' technology use when they leave for college because of proximity limitations.

Vaterlaus, Beckert, and Bird (2015) queried 82 college students (18-25 years old) and 72 parents regarding parental mediation practices in late adolescence. College students (and their parents) who lived with their parents reported low levels of parental mediation of late adolescent cell phone (7%) and internet (11%) use. Similarly, college students (and their parents) who lived away from home also reported low levels of parental mediation of cell phone (28%) and internet use

(13%). Participants who reported parental mediation of college students' technology use indicated that the same strategies (i.e., monitoring, active mediation, and restriction) that have been identified in research with younger adolescents were employed. Furthermore, Vaterlaus et al. (2015) asked participants who reported no parental mediation to explain why parental mediation was not implemented. Some parents and college students reported that parental mediation did not happen because parents lacked the technological capabilities to mediate internet or cell phone use. Parent-child trust was thought to be earned while an adolescent was in high school and the earned trust eliminated the need to implement parental mediation when the adolescent entered college. Coyne and colleagues (2013) posited that parental mediation might be conceptualized as parental control, which could be interpreted by college students as parental intrusion. The Vaterlaus et al. (2015) study focused on adolescents who pursued post-secondary education and presently it is unclear if adolescents who seek different opportunities following high school graduation have similar experiences with parental mediation.

FUTURE RESEARCH DIRECTIONS

The majority of the research on parental mediation has centered on descriptions of the different techniques parents use and fewer studies have investigated the actual influence of parental mediation on adolescents' outcomes. Future research should prioritize studies focused on evaluating the effectiveness of different parental mediation techniques in reducing negative behavioral outcomes for adolescents. If specific parental mediation techniques are found to reduce risks associated with adolescent interactive technology use, researchers should then identify ways to aid parents in overcoming the barriers to implementing parental mediation techniques. It appears that parents' limitations in technological ability serves as a barrier to implementing some of the

parental mediation techniques. Research focused on methods for increasing parents' technological abilities may be a way of decreasing the digital generation gap and increasing parental mediation. Finally, researchers and practitioners should then create, implement, and evaluate interventions and educational programs for parents and adolescents to increase the use of effective parental mediation techniques.

CONCLUSION

Interactive technology use has become an integral aspect of adolescents' lives. Parents do have concerns about adolescent interactive technology use and, because legal regulations on interactive technology are difficult to create and regulate, policy makers rely on parents to take responsibility for keeping their adolescents safe from these risks. Some parents have implemented various parental mediation strategies (i.e., active mediation, restrictive mediation, and monitoring) in an attempt to prevent negative consequences for adolescents. There remains much to be learned regarding the effectiveness of parental mediation at decreasing the potentially negative consequences and increasing the potentially positive consequences associated with adolescent interactive technology use.

REFERENCES

- Aquilino, W. S. (2006). Family relationships and support systems in emerging adulthood. In J. J. Arnett & J. L. Tanner (Eds.), *Emerging adults in America* (pp. 193–218). Washington, DC: American Psychological Association. doi:10.1037/11381-008
- Boyd, D., & Hargittai, E. (2013). Connected and concerned: Variation in parents' online safety concerns. *Policy & Internet*, 5, 245–269.
- Brooks-Gunn, J., & Donahue, E. H. (2008). Introducing the issue. *The Future of Children*, 18(1), 3–10. doi:10.1353/foc.0.0008
- Correa, T. (2014). Bottom-up technology transmission within families: Exploring how youths influence their parents digital media use with dyadic data. *Journal of Communication*, 64(1), 103–124. doi:10.1111/jcom.12067
- Costello, C. R., McNiel, D. E., & Binder, R. L. (2016). Adolescents and social media: Privacy, brain development, and the law. *Journal of the American Academy of Psychiatry and the Law Online*, 44(3), 313–321. PMID:27644864
- Coyne, S. M., Padilla-Walker, L. M., & Howard, E. (2013). Emerging in a digital world a decade review of media use, effects, and gratifications in emerging adulthood. *Emerging Adulthood*, 1(2), 125–137. doi:10.1177/2167696813479782
- Duggan, M., Lenhart, A., Lampe, C., & Ellison, N. B. (2015). *Parents and social media*. Retrieved from <http://www.pewinternet.org/2015/07/16/main-findings-14/>
- Federal Trade Commission. (2016). *Children's online privacy protection rule (COPPA)*. Retrieved from: <https://www.ftc.gov/enforcement/rules/rulemaking-regulatory-reform-proceedings/childrens-online-privacy-protection-rule>
- Fletcher, A. C., & Blair, B. L. (2014). Maternal authority regarding early adolescents social technology use. *Journal of Family Issues*, 35(1), 54–74. doi:10.1177/0192513X12467753
- Kirwil, L. (2009). Parental mediation of childrens internet use in different European countries. *Journal of Children and Media*, 3(4), 394–409. doi:10.1080/17482790903233440
- Lee, S. J. (2013). Parental restrictive mediation of childrens internet use: Effective for what and for whom? *New Media & Society*, 15(4), 466–481. doi:10.1177/1461444812452412

- Lenhart, A. (2015). *Teens, social media, & technology*. Pew Internet and American Life Project. Retrieved from <http://www.pewinternet.org/2015/04/09/teens-social-media-technology-2015/>
- Livingstone, S. (2003). Childrens use of the Internet: Reflections on the emerging research agenda. *New Media & Society*, 5(2), 147–166. doi:10.1177/1461444803005002001
- Livingstone, S., & Bober, M. (2004). *UK children go online: Surveying the experiences of young people and their parents*. Retrieved from http://eprints.lse.ac.uk/395/1/UKCGOsurveyreport.pdf?origin=publication_detail
- Livingstone, S., & Bober, M. (2006). Regulating the Internet at home: Contrasting the perspectives of children and parents. In D. Buckingham & R. Willett (Eds.), *Digital generations* (pp. 93–113). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Livingstone, S., & Helsper, E. J. (2008). Parental mediation of childrens internet use. *Journal of Broadcasting & Electronic Media*, 52(4), 581–599. doi:10.1080/08838150802437396
- Nathanson, A. I. (1999). Identifying and explaining the relationship between parental mediation and childrens aggression. *Communication Research*, 26(2), 124–143. doi:10.1177/009365099026002002
- Nathanson, A. I. (2001). Parent and child perspectives on the presence and meaning of parental television mediation. *Journal of Broadcasting & Electronic Media*, 45(2), 201–220. doi:10.1207/s15506878jobem4502_1
- OKeeffe, G. S., & Clarke-Pearson, K. (2011). The impact of social media on children, adolescents, and families. *Pediatrics*, 127(4), 800–804. doi:10.1542/peds.2011-0054 PMID:21444588
- Oksman, V., & Turtiainen, J. (2004). Mobile communication as a social stage meanings of mobile communication in everyday life among teenagers in Finland. *New Media & Society*, 6(3), 319–339. doi:10.1177/1461444804042518
- Padilla-Walker, L. M., & Coyne, S. M. (2011). Turn that thing off! parent and adolescent predictors of proactive media monitoring. *Journal of Adolescence*, 34(4), 705–715. doi:10.1016/j.adolescence.2010.09.002 PMID:20943262
- Padilla-Walker, L. M., Coyne, S. M., Fraser, A. M., Dyer, W. J., & Yorgason, J. B. (2012). Parents and adolescents growing up in the digital age: Latent growth curve analysis of proactive media monitoring. *Journal of Adolescence*, 35(5), 1153–1165. doi:10.1016/j.adolescence.2012.03.005 PMID:22507383
- Vaterlaus, J. M., Beckert, T. E., & Bird, C. V. (2015). At a certain age its not appropriate to monitor ones child: Perceptions of parental mediation of emerging adult interactive technology use. *Emerging Adulthood*, 3(5), 353–358. doi:10.1177/2167696815581277
- Vaterlaus, J. M., Beckert, T. E., Tulane, S., & Bird, C. V. (2014). They always ask what Im doing and who Im talking to: Parental mediation of adolescent interactive technology use. *Marriage & Family Review*, 50(8), 691–713. doi:10.1080/01494929.2014.938795
- Vaterlaus, J. M., Jones, R. M., & Tulane, S. (2015). Perceived differences in knowledge about interactive technology between young adults and their parents. *Cyberpsychology (Brno)*, 9(4). doi:10.5817/CP2015-4-3
- Vaterlaus, J. M., & Tulane, S. (2015). Digital generation gaps in parent-adolescent relationships. In C. J. Bruess (Ed.), *Family Communication in the Digital Age* (pp. 426–446). New York, NY: Peter Lang Publishing.

Wang, R., Bianchi, S. M., & Raley, S. B. (2005). Teenagers Internet use and family rules: A research note. *Journal of Marriage and the Family*, 67(5), 1249–1258. doi:10.1111/j.1741-3737.2005.00214.x

Zimmer-Gembeck, M. J., & Collins, W. A. (2003). Autonomy development during adolescence. In G. R. Adams & M. Berzonsky (Eds.), *Blackwell Handbook of Adolescence* (pp. 175–204). Oxford: Blackwell Publishers.

ADDITIONAL READING

Lee, S. J. (2013). Parental restrictive mediation of childrens internet use: Effective for what and for whom? *New Media & Society*, 15(4), 466–481. doi:10.1177/1461444812452412

Livingstone, S., & Helsper, E. J. (2008). Parental mediation of childrens internet use. *Journal of Broadcasting & Electronic Media*, 52(4), 581–599. doi:10.1080/08838150802437396

Padilla-Walker, L. M., Coyne, S. M., Fraser, A. M., Dyer, W. J., & Yorgason, J. B. (2012). Parents and adolescents growing up in the digital age: Latent growth curve analysis of proactive media monitoring. *Journal of Adolescence*, 35(5), 1153–1165. doi:10.1016/j.adolescence.2012.03.005 PMID:22507383

Vaterlaus, J. M., Beckert, T. E., & Bird, C. V. (2015). At a certain age its not appropriate to monitor ones child: Perceptions of parental mediation of emerging adult interactive technology use. *Emerging Adulthood*, 3(5), 353–358. doi:10.1177/2167696815581277

Vaterlaus, J. M., Beckert, T. E., Tulane, S., & Bird, C. V. (2014). They always ask what Im doing and who Im talking to: Parental mediation of adolescent interactive technology use. *Marriage & Family Review*, 50(8), 691–713. doi:10.1080/01494929.2014.938795

KEY TERMS AND DEFINITIONS

Adolescence: The human developmental time period between childhood and adulthood.

Active Mediation: A parental mediation technique where parents use parent-child discussion and participation in their adolescents' technology use to alert adolescents to risks associated with interactive technology.

Digital Generation Gap: A difference between parents and adolescents in knowledge about how to use interactive technology.

Interactive Technology: Technology that digitally facilitates interaction between people or allows for user content creation or manipulation.

Monitoring: A parental mediation technique where parents view adolescents' activity (e.g., internet history, text messages) on interactive technology with or without the adolescents' knowledge.

Parental Mediation: Parental interventions and interactions with their adolescents regarding technology use implemented to promote positive technology use.

Restriction Mediation: A parental mediation technique where parents block access to interactive technology or implement limits (e.g., rules) on adolescents' interactive technology use.

The Qualities and Potential of Social Media

Udo Richard Averweg

eThekweni Municipality, South Africa

Marcus Leaning

University of Winchester, UK

INTRODUCTION

Technology has long been thought to bring about change in human behavior. This varies from minor changes in individual behavior to larger societal level transformations. Twitter and a variety of other social media have been considered as playing a significant part in facilitating, if not causing, recent political revolutions in certain North Africa and Middle Eastern countries in which long standing political regimes established in the post-colonial period were overthrown often by internally driven anti-systemic movements. In a number of instances the overthrow has resulted in less stable states (Khamis & Vaughn, 2011). Similarly, social media has been credited heavily in the spread and organization of the various occupations and civic unrests that occurred in Western societies under the auspices of the Occupy movement (Gleason, 2013). Occupy is a globally distributed network of anti-systemic groups broadly opposed to what they perceive to be systems of economic inequality and the pan governmental legal instruments that facilitate such inequality. A significant volume of commentary on blogs and in newspaper columns has clearly identified social media (often accessed by a mobile phone) as one of the driving forces and key tools of these uprisings (Marzouki, 2011).

Similarly, in commentary on less headline-grabbing events, we also find social media as credited with a change causing potential and social media has also been considered as a way to improve organizational performance (Curran, O'Hara & O'Brien, 2011). In business and management so-

cial media is seen as a way to empower customers and users, improve service levels through feedback loops and generally offer an improved service users' experience of bureaucracy. In this article we examine how the qualities and potential of social media can be understood to cause social change.

BACKGROUND

Social Change

Collective human life is structured by social practices; they define accepted behavior and how we act in different circumstances. Sociologists often understand such social practices as changing and shifting in response to social dynamics, logics and developments as well as external drivers (Giddens, 1990). We may understand technology as a major agent in driving such social change though there is disagreement in the manner in which such change occurs (Leaning, 2009) and whether such change is socially beneficial. One argument is that social media are contributory to the general decline of collective action in social life - social media contribute to the atomization of society and a decline in social capital - the binding connections between individuals. An opposing position contends that social media and similar technology actually contributes to social capital since there "is clear evidence that social capital has been on the ascent in the past decade: in the form of networks in cyberspace" (Lin, 2001). That is, the interpersonal connectedness facilitated by

social media is beneficial and indeed social media afford communities opportunities to internally 'bond' or change in a positive, affirming manner.

Social Media

Social media refers to a range of World Wide Web (hereinafter referred to as 'web') applications that facilitate communication between individuals. Numerous technical sources identify the ancestry of social media in web 2.0 technology. Many of the features considered unique to social media (such as interest-driven communities, peer commentary and horizontal rather than hierarchical information flows) were present in early internet communication systems. Indeed much contemporaneous commentary on social media echoes statements made about the internet in the 1990s and early 2000s.

New Technologies

In analyzing technology we often seek to find characteristics, features or qualities that were not present in whatever technology went before – social media must do something that previous media simply did not do. The cause of this perception lies in the marketing and financing of new technologies; there has to be a unique selling proposition of a technology, it has to offer a feature that solves a problem that previous technologies did not. However, this runs counter to the way in which technology is developed. Technology does not progress in terms of major changes, it moves forwards cautiously rather than in leaps and bounds. We might argue that technological innovations occur on a continuum; new technologies slightly improve a particular aspect of what went before but they rarely take things in a new direction entirely. A new technology may do the task we want performed better but only very rarely does it perform a new task.

Moreover, we should recognize that technological history is often not a single, linear pathway (Nicholson, 2007). While from our current per-

spective a technology may seem to be the result of a continuous path of development we find that there are often many developments that leading to technological cul-de-sacs. We may see the development of technology as occurring in a similar fashion to natural evolution - many different approaches to problems are tried out, some succeed but many fail. However, the important aspect is that each instance of technology is designed as an end in itself, to solve a problem or meet a need. Technologies are not (normally) developed as a stepping-stone to a clearly defined future technology. However, when we examine technology in hindsight that is exactly what they appear to be. Technology is not consciously moving towards a specific goal, rather it is changing and transforming and fighting for survival in an environment of limited resources (Ziman & Ziman, 2003). Thus if we look at social media we find that many of the features are actually improvements upon already existing patterns of communication rather than 'true' innovations. Different forms of social media may share certain features in common and seem to be advances of earlier technologies but whether they are radically different from previous forms of technology and communication is contestable.

THE CHARACTERISTICS OF SOCIAL MEDIA

If the argument that social media do possess new characteristics is correct then we can start to identify the particular areas in which social media is an improvement on what went before. In this article we focus on four key areas:

- The contemporaneity of communications;
- Productive audiences;
- Its dialogical and network nature; and
- Its searchable and 'taggable' nature.

While these characteristics were present in previous internet enabled media (and even earlier in some cases) they are modified, adjusted and

combined in new ways in social media that allow us to see social media as a new technology.

The Contemporaneity of Communications

The contemporaneity of communications refers to the way in which social media are constant in people's lives and how they afford the user continuous routes of communication. Initial social research on internet communications articulated a divide between those technologies that afforded synchronous communications (live, real time discussions such as the various chat systems) and those which afforded asynchronous communications (turn taking systems such as new forums and email lists).

Social media confound this simple division. This occurs for two main reasons: first, many social media platforms incorporate both synchronous and asynchronous communication systems allowing users to both discuss issues in a serial manner asynchronously while also engage in synchronous communication – thus Facebook allows serial communication through status updates and comments but also to directly message a fellow user and engage in synchronous communication if they are logged in at the same time. Second, social media platforms are accessed through a range of hardware platforms including desk-bound personal computers, tablet devices and mobile devices. Accordingly social media is far more accessible and invasive into contemporary personal life than previous communications systems. While social media do not really offer anything new in terms of sheer communicative practice – they still offer an interface through which we can communicate – the means of accessing and communicating through the media is significantly advanced. Because of the penetration of computing devices into work, home and personal spaces and the presence of 'always on' 3G (and more recently 4G) mobile device networks, communication using social media is constant and 'network presence' is increased.

Using social media through 'always on' devices means that we are almost continually in communication. Communications are not stored and delivered to us on a particular occasion but relayed directly to us via social media platforms on mobile devices. Communication becomes cotemporary and we are deemed ever-present in the network.

Productive Audiences

While the amateur production of media content has long been possible through low cost media technology, social media add the dimension of being able to distribute this content easily. Many social media platforms offer the user the ability to upload media content. While there has been much concern (with considerable justification) about users disseminating commercial content without full recompense being made to the copyright owners, what is most striking is that there are significant volumes of media content produced and uploaded without any financial reward for the user/producer. This represents a marked shift from the previous conception of the relationship of media producer to audience.

In the previous older model there was a clear line between those producing media content: professionals who had undergone a period of training and development and part of a media industry – and those consuming it, the audience. Though there has been a reconsideration of the passivity of the audience of the audiences and audiences are no longer considered as passive as once thought even for old media, social media are thought to offer a new dimension of activity. In some instances these productive activities are aligned to their consumption and engagement with professionally produced content (phenomena such as fan media, parody and 'mash up' of professional media content into new content) but many people also produce their own entirely new content (Jenkins, 2006).

Dialogical and Network Nature

Social media are inherently social; they permit users to communicate with one another easily. While many forms of media allow users to communicate with other users and a significant number of pre-web 2.0 internet communication platforms are explicitly based on this, in social media this is elevated to a new level. The manner in which this occurs is through the integration of intra-user communication into other forms of association. This occurs in three main ways. First, social media allow communication between users who have engaged in content dissemination or consumption of content uploaded by others. Thus, many of the social media platforms dedicated to content dissemination such as YouTube or Flickr permit users to post comments tied to specific media texts and to 'rate' the posted texts. A consequence of this is that sustained discussions emerge around texts and often veer into new areas with links to other texts being posted. Second, social media permit communication by prior existing communities that have a presence online. Innumerable geographical, interest-based, friendship and other forms of associational groups that exist off-line also have presence in social media. Moreover these groups are extended in new ways by their being online – users are made aware of prior existing groups and are able to join them and learn of their activities easily. Third, users are in multiple networks and users function as 'nodes' between separate networks. Social media allow users to facilitate new associations and forms of sociality. A consequence of this is the emergence of new networks and groups that only exist through social media.

Searchable and 'Taggable' Nature

The 'taggable' nature of media content refers to the way in social media allows users to add descriptors or 'tags' to media content so that it becomes identifiable and findable by themselves and other users. For example, wireWAX (a taggable video startup) allows interactive media content to be

overlayed on YouTube streaming media. The benefit of the 'tagging' aspect is that the manner of description of an object stored online becomes far more personal and familiar to the individual. In this manner the user does not have to become skilled in the arcane and technical means of description. Accordingly, the labelling of content becomes far more democratic, personal, useable and epistemologically 'local' what Vander Wal refers to as 'Folksonomy' (Vander Wal, 2005). This is the opposite of technical taxonomy where a single authority determines classification: folksonomy is a bottom up approach to defining information and content. As such the practice affords strong resonance with grassroots and anti-hierarchical interpretations of the impact of new and particularly social media.

SOCIAL MEDIA CAUSING A CHANGE

We consider the above noted differences as resulting in four practices that can facilitate social change. These practices are: folksonomy, virality, collective intelligence and the emergence of 'long tail' patterns of production.

Folksonomy

As we noted above, social media allows users to add tags to digital content. Such tagging permits new realms of information to build up around digital content. Of particular interest is the impact this has upon classifications of digital content and data. As the amount of tagged data rises, it begins to pose a threat to the existing definitions and descriptions given to the data when it was first uploaded. If people add their own descriptions and do so in numbers, then the importance of the original description gradually declines.

This affords new ways in which information can be identified and found. Media theorist Dennis McQuail noted that electronic media caused a change in how information was stored, disseminated and accessed. McQuail (McQuail, 1992)

was in turn developing ideas proposed in 1986 by Bordewijk and van Kaam (Bordewijk & van Kaam, 1986). They proposed that interpersonal information transfer works in one of four ways:

- Allocation (from a central point to individuals - as in broadcasting);
- Conversation (exchange of information between peers);
- Consultation (where information is stored in a location and individuals retrieve it - as in a library system); and
- Registration (where those in the centre record information from different end points).

McQuail contends that in an age of new media the traditional model of allocation is challenged by the model of consultation. We now longer passively receive information but can go in search of it. However, as many have pointed out how data is defined, labelled and stored has a significant impact upon whether it can be retrieved. Historically data has been classified by experts in the field, such as librarians, often using complicated systems. What tagging does is allow individuals to add their own labels. The consequence of this is that expert systems of information description are challenged by one driven by non-expert users – folk – to use the colloquialism. Folksonomy is the description of information by the people who use it and the consequent challenge to the power structure of those who have historically labelled it.

Virality

If folksonomy is a challenge to those who label information, virality is often considered a challenge to those who have traditionally controlled the pathways of information flow. The idea of virality emerges from network society theory - a broadly sociological approach to understanding the density and multi-directional forms of communication present in the contemporary world. It

was developed in the work of (amongst others) Manuel Castells (Castells, 2011a, 2011b). In addition to the significant amount of academic commentary, network theory has also attracted much popular commentary. This commentary has constructed the network society in both positive and negative ways.

One ‘vision’ has been the interpretation of contemporary society as being highly integrated with individuals being part of many different communities of networks. We are part of work communities (our colleagues), family and friendship communities and geographic communities (our neighbors) and many historical communities (people in places we have lived in before or met) which social media permits us to remain in even if we leave the physical site of its original creation. Of course such links were possible in the pre-social media age and the pre-internet age. But the links are made far easier to maintain in the social media age and certain qualities of social media applications and sites such as Facebook make locating and contacting old acquaintances far easier. On top of this technologically facilitated ‘social network’, new changes such as the proliferation of platforms and platform independent content means that content can now easily flow across the network.

In previous eras, information tended to follow quite strict pathways through society. Old media forms such as newspapers and television afford a unidirectional flow of information from a central point outwards. The internet and other forms of new media challenged this in the 1990s and 2000s and social media has greatly accelerated the challenge. Information now easily crosses between social networks and it does so that patterns of flow now appear chaotic and not managed or planned. Indeed they seem to spread not with a rational, intelligent plan behind them but with a biological or chaotic pattern. This resemblance to the biological spread led to the coining of the term viral to explain the spread of a piece of film, image or other media file. The link to biological

phenomena is strengthened when the idea of virality is linked to the concept of a 'meme'. In the work of evolutionary biologist Richard Dawkins (Dawkins, 1989), a meme is a concept or idea that 'floats free' of its originating discourse and spreads with a society or culture in a similar way to how successful genes spread across populations. If specific discreet units of media are considered memes then the idea of a viral pattern of distribution becomes stronger.

Virality as a form of media distribution has proven to be a significant upset to those historically in control of media. While entertainment clips that spread virally are not considered of much consequence, virality achieves more significance when the spread information conflicts with dominant systems of representation and challenges the existing political order. In such instances virality becomes a subversive social practice that adds a political dimension to social media.

Collective Intelligence

A wiki is a type of website that allows the collaborative production of information. Possibly the best known wiki is Wikipedia. The Wikipedia website contains information that has been gathered and systematically organized in a collaborative manner by multiple users. The collective of collaboration and information sharing (which is sufficiently dynamic and also co-ordinated) over physical distance, results in properties which may be seen as 'smart' or 'intelligent'. This spawns emerging capacities beyond the individual components. Collective intelligence involves the ability to receive, transmit, utilize and share flows of information. Sociologist, Kingsley L. Dennis, argues that successful wikis (such as Wikipedia) make 'smart' decisions since they embody forms of an emerging hybridized collective intelligence – where the weaknesses of the individual are compensated by the contribution of the many.

James Surowiecki (Surowiecki, 2005) identified that in many instances collective decisions

by many non-experts produce better results than those by individual experts. For Surowiecki, crowds can be 'wise' but they also need to be diverse and beyond physical homogeneity in order to be so. Dennis suggests that there are therefore degrees of 'intelligence' to be found in collectives that are diverse, distributed and heterogeneous yet sufficiently connected to share affiliations similar to information flows. The 'new agora' that wikis offer is informed through a synchronization of material in tandem with communication and digital technologies. As Wagner (2009) notes, by using wikis and related technologies, significant insights may be drawn from many small chunks of knowledge that carry relatively little meaning individually but become exceedingly meaningful when combined.

Long Tail

The 'long tail' is a description of the way in which retailing in the contemporary world allows small businesses that produce a niche product can exist in a world of mass production. The term describes the statistical phenomenon of a chart showing frequency distribution of a low number of instances of many cases. Such a chart will detail a dataset which has a number of cases with a high number of incidences and a gradual decline in the occurrence across the rest of the cases. The chart will have a small number of cases with a high number which forms a peak on the left of the chart and a gradual decline in the occurrences to the right forming a 'long tail'.

The 'long tail' was popularized in Chris Anderson's text (Anderson, 2007) when he noted that it described a number of new internet business such as Amazon which sold a high volume of a very small number of products but a small volume of a large number of products. Anderson proposed that companies could make money by selling very little of many different products; a challenge to the traditional wisdom of retailers of selling a significant amount of a very limited stock range.

The ‘long tail’ describes how media content in the social media environment is consumed and produced. There are a number of texts that achieve high audience numbers – successful internet memes as described in the virality section above; yet there are many media texts that never achieve a large audience. What is interesting is that there are a significantly high number of these.

There are innumerable videos on YouTube with few viewers, many blogs go unread by more than a handful of readers and the vast majority of Twitter streams have only a very small number of followers. This does not lessen the importance of these media; rather it simply indicates that very different consumption patterns of social media. The ‘long tail’ is a description of contemporary consumption patterns which owe much to the impact of the characteristics of social media.

FUTURE RESEARCH DIRECTIONS

While future gazing is always a problematic and a questionable activity, we identify a number of possible trends in social media development and extrapolate from these to the future.

The Self-Leveraging Network

The rich data in networks offers marketing professionals many opportunities for targeting potential customers more intelligently. While this has been possible for a number of years through the use of algorithms that detect patterns of interaction between users and data assets and then target the user with new assets, a number of companies are now developing this approach in a more intelligent or strategic manner. For example, the start-up *Insightfully* has developed an application that takes information from a person’s Twitter and LinkedIn accounts and identifies potential relationships based upon data and previous encounters. The application (or simply ‘app’) explores the interaction between a person and their networks by calculating how much the network assists the person and

how much the person assists the network. In this way it exploits extant data in new ways delivering nuanced information that can be used to assist the person in leveraging their networks.

Deeper Personal Profiles and the Universal Identity (ID)

The social media market is currently very heterogeneous - numerous platforms abound and users maintain many different accounts on these platforms. An increasing problem for platform providers is users not completing profiles. Users often only partially complete the various steps and this has an impact upon both the value of the data (partial data is far less valuable to advertisers) to the social media company and value to the user. In addition to inaccurate information being targeted at them, partial profiles often result in far less traffic and awareness among peer communities.

During 2010, Facebook initiated *Facebook Connect* – this allowed users to use their Facebook login on third party sites. This was followed by *Open Graph API* which has now been adopted and developed by numerous other organizations. Such practices presage the advent of what we may term the Universal ID or login – a system in which we use one login to access all our personal data stored across a range of platforms and organizations. This would facilitate being able to access and straddle multiple platforms and systems without multiple login processes.

Hyper Locality

When electronic mass media emerged during the early 20th century one of the main fears was that the world would become homogenized - a mass media, largely produced in the United States of America and western Europe would dominate all aspects of culture and would eventually drive out local, traditional values. Such fears are still felt today and significant effort is placed on preserving local cultures.

However, there is a gradual trend towards the production and dissemination of localized content. It is this aspect that is predicted to grow. Indeed given the productive qualities of social media content will become increasingly local to the individual. This serves to bring the technology 'closer' to the individual; social media will serve to assist in the narration and engagement not with a distant other (as corporate and state news seeks to do) but with our everyday lives.

CONCLUSION

We have examined how the qualities and potential of social media can be understood to cause social change. It is evident that during the last few years, social media has caught the public interest in a way that few technologies have achieved. While social media can be seen as the media used to be social, it has also become a powerful force since it has shaped government elections, impacted business decisions and altered the landscape of what news is and how it is reported. Social media has indeed become part of our everyday life.

ACKNOWLEDGMENT

Text for this article has been extracted and adapted from the authors' four-series articles which appeared in the *Journal of the Southern Africa Institute of Management Services (SAIMAS)* during the period December 2012-5. Permission to reuse the text was granted by the Editor of the SAIMAS journal.

REFERENCES

- Anderson, C. (2007). *The Long Tail: How Endless Choice is Creating Unlimited Demand*. Random House Business Books.
- Bordewijk, J. L., & van Kaam, B. (1986). Towards a New Classification of Tele Information Services. *InterMedia*, 14(1), 16–21.
- Castells, M. (2011a). *The Power of Identity: The Information Age: Economy, Society, and Culture* (2nd ed.). Hoboken, NJ: Wiley-Blackwell.
- Castells, M. (2011b). *The Rise of the Network Society: The Information Age: Economy, Society, and Culture* (2nd ed.). Hoboken, NJ: Wiley-Blackwell.
- Curran, K., OHara, K., & OBrien, S. (2011, July-September). The Role of Twitter in the World of Business. *International Journal of Business Data Communications and Networking*, 7(3), 1–15. doi:10.4018/jbdcn.2011070101
- Dawkins, R. (1989). *The Selfish Gene*. Oxford University Press.
- Giddens, A. (1990). *The Consequences of Modernity*. Stanford University Press.
- Gleason, B. (2013, March 15). #Occupy Wall Street: Exploring Informal Learning About a Social Movement on Twitter. *The American Behavioral Scientist*, 57(7), 966–982. doi:10.1177/0002764213479372
- Jenkins, H. (2006, September). *Fans, Bloggers, and Gamers: Exploring Participatory Culture*. New York: New York University Press.
- Khamis, S., & Vaughn, K. (2011, July). 'We Are All Khaled Said': The potentials and limitations of cyberactivism in triggering public mobilization and promoting political change. *Journal of Arab & Muslim Media Research*, 4(2-3), 145–163.
- Leaning, M. (2009). *The Internet, Power and Society: Rethinking the Power of the Internet to Change lives*. Oxford, UK: Chandos Publishing. doi:10.1533/9781780631684
- Lin, N. (2001). *Social capital: A theory of social structure and action*. London: Cambridge University Press. doi:10.1017/CBO9780511815447

Marzouki, N. (2011, January 19). *Tunisia's Wall Has Fallen*. Middle East Research and Information Project. Retrieved 18 November 2015 from <http://www.merip.org/mero/mero011911>

McQuail, D. (1992). *Media performance: Mass communication and the Public Interest*. London: Sage.

Nicholson, P. (2007). A History of E-Learning: Echoes of the Pioneers. In B. Fernández-Manjón, J. M. Sánchez-Pérez, J. A. Gómez-Pulido, & M. A. Vega-Rodríguez (Eds.), *Computers and education: e-learning from theory to practice*. New York: Springer. doi:10.1007/978-1-4020-4914-9_1

Surowiecki, J. (2005). *The Wisdom of Crowds*. New York: Anchor Books.

Vander Wal, T. (2005). *Folksonomy Definition and Wikipedia*. Retrieved 18 November, 2012 from <http://www.vanderwal.net/random/entrysel.php?blog=1750>

Wagner, C. (2009). Wikis and Beyond: Harnessing Collective Intelligence through Web 2.0 Technologies. *AMCIS 2009 Proceedings*.

Ziman, J. M., & Ziman, J. (2003). *Technological innovation as an evolutionary process*. Cambridge, UK: Cambridge University Press.

ADDITIONAL READING

Basalla, G. (1988). *The evolution of technology*. Cambridge: Cambridge University Press.

Castells, M. (2002). *The Internet Galaxy: Reflections on the Internet, Business, and Society*. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199255771.001.0001

Castells, M. (2010). *End of Millennium: The Information Age: Economy, Society, and Culture* (2nd ed., Vol. III). Hoboken, NJ: Wiley-Blackwell.

Jansen, B. J., Zhang, M., Sobel, K., & Chowdury, A. (2009). Twitter power: Tweets as electronic word of mouth. *Journal of the American Society for Information Science and Technology*, 60(11), 2169–2188. doi:10.1002/asi.21149

Jenkins, H. (2006). *Convergence Culture: Where Old and New Media Collide*. New York: NYU Press.

Juris, J. S. (2012, May). Reflections on #Occupy Everywhere: Social media, public space, and emerging logics of aggregation. *American Ethnologist*, 39(2), 259–279. doi:10.1111/j.1548-1425.2012.01362.x

Tufekci, Z., & Wilson, C. (2012). Social media and the decision to participate in political protest: Observations from Tahrir Square. *Journal of Communication*, 62(2), 363–279. doi:10.1111/j.1460-2466.2012.01629.x

Wilson, C., & Dunn, A. (2011). The Arab Spring! Digital Media in the Egyptian Revolution: Descriptive Analysis from the Tahrir Data Set. *International Journal of Communication*, 5, 1248–1272. <http://ijoc.org/ojs/index.php/ijoc/article/view/1180> Retrieved 18 November 2015

KEY TERMS AND DEFINITIONS

Collective Intelligence: The weaknesses of the individual are compensated by the contribution of the many.

Contemporaneity of Communications: Refers to the way in which social media are constantly in people's lives and how they afford the user continuous routes of communication.

Folksonomy: The description of information by the people who use it and the consequent challenge to the power structure of those who have historically labelled it.

Long Tail: A description of the way in which retailing in the contemporary world allows small

businesses that produce a niche product can exist in a world of mass production.

Social Media: Refers to a range of World Wide Web applications that facilitate communication between individuals.

‘Taggable’ Nature of Media: Refers to the way in which social media allows users to add descriptors or ‘tags’ to media content.

Virality: The tendency of information to be circulated rapidly and widely from one Internet user to another; the quality or fact of being viral.

Wiki: A unique type of website that eliminates physical distance between the reader and the producers of the information.

S

Short History of Social Networking and Its Far-Reaching Impact

Liguo Yu

Indiana University – South Bend, USA

INTRODUCTION

Social networking is a platform built with information technology that supports virtual socialization (Lenhart & Madden, 2007). Although social networking has a short history compared with the long history of human socialization, it is growing fast and its impact on our life has not yet been fully realized and understood.

For thousands of years, face-to-face socialization is the major form to build social relations (Grusec & Lytton, 2012). Communities are formed locally, economy is developed locally, education, entertainment and religious activities are all performed locally. People interact with each other through physical contact and physical communications. Virtual communication appeared with the birth of mail service (Scheele, 1970). We have been using this kind of virtual communication for hundreds of years. However, virtual communities are rarely formed with surface mail, because mail correspondence usually happens between only two people, which are not enough to form a community. Recall the saying “One person is single, two persons are couple, and three persons form a society”, we can say that a regular mail service between two people forms no virtual community.

Virtual communication, virtual social group, and virtual community become a reality with the advancement of information technology, especially internet-based technology. Social networking was first introduced about 40 years ago. In less than half a century, it has become an

important part of our lifestyle, and moreover, it is still evolving and penetrating into every aspect of our life and our society (Papacharissi, 2010). Its future impact will be profound.

This chapter describes social networking’s different perspectives. The main objective is to help readers understand the technology potential and limits of social networking, its impact and its possible future. The remaining of this chapter is organized as follows. We first describe the background knowledge, including the evolution of social networking together with its supporting technology and the latest research development in this area. Then, we analyze the major social networking services and their features. Next section describes the new business and new socialization built on top of social networking, and their impact on social relations and social dimensions. Future research directions and conclusions are presented finally.

BACKGROUND

Social networking began with the introduction of computer network (Warschauer, 2004). The early form of virtual group is the email mailing list. Email is a communication mechanism established originally on mainframe computers in 1972 (Merritt, 2012). Later, this mechanism is introduced into personal computers on the network. A simple email communication between two users might not be too much different from the traditional surface

mail communication, because it only involves two users and no social groups or social relations are formed here.

However, an email mailing list service provides an unprecedented mechanism to group users together and form virtual communities. Users do not need to meet face to face and could share their feelings, information, and ideas, remotely. With the support of email attachment, pictures, documents, and links to information or data on the Internet could be shared within the group. Although mailing list is an asynchronous communication mechanism, it is better enough to shadow the traditional mail correspondence. Despite the fact that many modern social networking services have emerged in the past ten years, email mailing list is still widely used in business, education, and some organizations. One of the most famous mailing lists is Linux Kernel Mailing List that connects Linux developers and users and forms the Linux community. This virtual community was established in 1995 and had over 14 thousand members in 2008 (Chu, 2009), and it is still actively used as the main platform for Linux community members to post announcements, ask questions, and address issues.

Besides email mailing list, there are also some other asynchronous social networking services, which include newsgroup, discussion forum, and bulletin board. These mechanisms are similar to mailing list: one user's post could be published openly online or broadcasted to the group members. Newsgroup, discussion forum, and bulletin board are usually built into a web portal or a specific web service. For example, discussion forum is an important component of distance education software (Branon & Essex, 2001). Another example is Baidu Tieba, China's largest communication platform that allows users to set up, join in, and search different discussion forums. The service is provided and supported by Baidu, China's search engine.

With the advancement of hardware technology, software applications are also improved. Due to the increase of computer network speed and personal

computer processing power, synchronous communications became possible in 1990s. Accordingly, synchronous social networking services emerged. Because synchronous communications could provide more convenient mechanisms to connect people together, more and more users switched from asynchronous communication platforms to synchronous communication platforms.

Instant messaging is one of the most popular synchronous communication methods used in late 1990s (Lewis & Fabos, 2005). It allows real time text transmission and information sharing. The most successful example of instant messaging service is Microsoft's Messenger, which was released in 1999 and widely used in 2000s. It was basically integrated with Microsoft's Windows Operating Systems and anyone with a Microsoft account can use it. In 2013, Microsoft's Messenger was merged into Skype's instant messaging system after Microsoft acquired Skype. Once, the users of Microsoft's Messenger had reached over 300 million.

Original social networking services only support text transmission. With the evolution of information technology, audio and video transmissions also became possible in the 2010s. For example, Microsoft's Sky Group Call is one of the most popular video conference services. Latest social networking services include more than text, photo, audio and video sharing. Other business and entertainment features are also incorporated. These features will be detailed in the following sections.

Research in social networking has been actively conducted in the past years. The work includes the study of communications through emails (Yu et al., 2008), mobile phones (Matic et al., 2012), instant messaging (Leskovec & Horvitz, 2008), and online social media (Kwak et al., 2010). For example, mobile communication data is used to study social network structures and social relations (Yu et al., 2013); emails are used to understand communication patterns (Malmgren et al., 2009); and social media is used to study recommendation systems (Carullo et al., 2015).

These studies revealed interesting facts about virtual relations and virtual communities (Chianese et al., 2015; Rheingold, 2000). Social networking in professional communities are also extensively studied. For example, Yu, Guan, and Ramaswamy (2016) studied the communication patterns of Linux virtual community; Russell (2013) studied the social relations of LinkedIn, Google+, and GitHub professional groups.

THE COMPETITIVE SOCIAL NETWORKING SERVICES

As with any other business domain, social networking services are evolving. Currently, a few internet giants are dominating the social networking service industry, while many others are fighting aggressively to obtain their market share. It is worth of studying the business strategy, service feature, user base, and financial status of different social networking services in order to better understand this industry.

Facebook

It is fair to say that Facebook is currently the most popular social networking platform. Facebook was launched on February 4, 2004. Its founder is Mark Zuckerberg (Kirkpatrick, 2011). The basic features of Facebook allow a user to build a profile,

connect with friends, share photos, videos or other information, and join virtual groups. According to Alexa.com, Facebook is the second most visited website worldwide, only next to Google. Although Facebook is banned in a few countries, including China, Iran, North Korea, Syria, and Cuba, there are over a billion monthly active users on Facebook platform. The impact of Facebook goes beyond our daily life. From domestic issues to international affairs, virtual communities on Facebook are playing more and more important roles. Table 1 shows the growth of Facebook in terms size, MAU (monthly active users), revenue, and net income.

Facebook is one of the most valued tech companies based on market capitalization. Other highly valued tech companies include Apple in computer hardware industry; Goggle in search engine; Amazon in e-commerce; and Microsoft in software industry. We can see that Facebook is the only high valued tech company of USA that is specialized in social networking services. Moreover, Facebook is still growing, with a 3.7% user growth rate in 2015.

Tencent

Tencent is China's internet company. Its founder is Pony Ma (Tse, E. 2015). Tencent provides many business services, such as gaming, news media, finance, sports, and entertainment. However, the

Table 1. The growth of Facebook

Year	Number of Employees	MAU (billion)	Revenue (\$ million)	Net Income (\$ million)
2008	850	0.1	271	-194
2009	1,218	0.2	777	35
2010	2,127	0.5	1,974	641
2011	3,200	0.8	3,711	1000
2012	4,619	1.0	5,089	1694
2013	6,337	1.2	7,872	1,491
2014	9,199	1.4	12,466	2,925
2015	12,691	1.6	17,928	3,669

Data source: Various web sites on the Internet

most important service provided by Tencent is social networking. Tencent's first social networking service is Qzone (QQ). Qzone was originally a PC-based instant messaging service. Later it was made available for mobile devices. Qzone is mostly used in China with over 800 million active accounts. It has become a platform for Chinese remote socialization.

The latest social networking service developed by Tencent is called WeChat, which is a pure mobile-based social media platform. WeChat is not only used in China, it is globally notable. WeChat was launched in 2011. In less than five years, it has got over 650 million users by 2015. Different from Qzone, the number of WeChat users grows fast internationally. Currently, most international users are in Southeast Asia, India, and Africa. But more and more users from America and Europe are joining in this social networking platform.

WeChat provides text messaging, voice messaging, broadcast messaging, video calls, and data sharing services. Other features include walkie talkie, group chat, shake, and friend radar. Because of its popularity, more and more users have switched from Qzone to WeChat. It is worth to note that both Qzone and WeChat have developed across the social networking domain. They have become platforms for ecommerce, gaming, and financial transactions.

Tencent is the most valued tech company in China. Domestically, it is competing with Alibaba on various internet-based services, such as online payment, online taxi-dispatch, music, sport, and entertainment. Globally, it is competing with Facebook on social networking services.

Twitter, Sina Weibo, Tumblr, and Others

Twitter is a social networking service that enables users to broadcast short messages to their followers. It was launched in 2006 and went public in 2014. Twitter is currently the most popular micro blogging service worldwide. Twitter has a limit on the size of the text message and usually only text could be sent. Image, audio, video, or other data formats are not supported.

Sina Weibo is mainly China's microblogging service. Like Twitter, users of Sina Weibo could choose to post messages publicly or broadcast messages only to their followers. Tumblr is Yahoo's social networking service. Similar to Sina Weibo, Tumblr users could post images, audio or video files.

Table 2 compares Twitter, Sina Weibo, and Tumblr. It should be noted here that Tumblr is still a private company. No revenue data and net income data are available to the public.

Besides Facebook, Tencent, Twitter, Sina Weibo, and Tumblr, there are also many other social networking services. The commonly used ones include LinkedIn for business and professional networking, Renren (China) for campus networking, Habbo for teenagers socialization, and VK (Vkontakte) used mainly in Europe.

From Desktop to Mobile

The latest trend in social networking service is that the platform is shifting from desktop computers to mobile devices. Some pure mobile-based

Table 2. Comparisons of Twitter, Sina Weibo, and Tumblr

	Twitter	Sina Weibo	Tumblr
Number of Employee	3,639	2,915	368
Monthly Active Users	305 million	236 million	226 million
Revenue	1,403 million (2014)	\$477.9 million	na
Net income	-\$578 million (2014)	\$34.7 million	na

Data source: Various websites on the Internet

apps were developed and launched recently. In a short period of time, these apps have attracted the majority of social networking users.

Instagram is an online mobile photo-sharing, video-sharing, and social networking service that enables users to take pictures and videos, and share them either publicly or privately on the app (Holmes, 2015). Instagram was launched in 2010 and acquired by Facebook in 2012. It has over 400 million users in 2015.

Momo is China's another widely used mobile social networking app. It is location-based and allows users to chat with nearby friends and strangers. Momo's main feature different from WeChat is stranger socialization. For example, users could use Momo to look for lost pets, find new friends in a new area, and form a travel group on site. Momo currently has over 80 million monthly active users. It is worth to note that Alibaba, China's internet giant, is one of Momo's major shareholders. With the support of Alibaba's e-commerce, Momo has successfully commercialized its social networking services.

Snapchat is a mobile video messaging app. Its users could set a time limit for how long recipients can view their snaps (photos or videos). After time out, snaps will be deleted from the server. Snapchat is mainly used by teenagers. By 2015, Snapchat has over 200 million active users. Currently, Snapchat is still a private company and its major investors include Yahoo, Alibaba, and Tencent.

Alibaba's first mobile social networking service, Laiwang, is proved to be unsuccessful, because it lost the battle with Tencent's WeChat in China. However, Alibaba did not give up its dream of social networking services. Besides investing in emerging social networking services,

such as Momo and Snapchat, lately, Alibaba has upgraded its mobile payment system, Alipay app with a social networking feature. Although it is hard to evaluate this business decision now, it is worth to note that due to the large user base of Alipay (400 registered users, 270 active monthly users), Alibaba's new effort of providing social networking service could be proved successful.

Besides aforementioned mobile social networking apps, there are also many others, including Vines, Foursquare, Pinterest, Spotify, Hootsuite, and more. For social networking services that are supported by both desktop computers and mobile devices, Table 3 summarizes the distribution of their users on these two platforms. For other social networking services not listed in Table 3, they have close to 100% mobile users. It is clear to see that mobile devices have almost completely replaced desktop computers for socialization. Because of the shift of user preference, companies are also adjusting their business strategies. For example, it is reported that Tencent has dropped Qzone's Windows 10 development. Instead, Qzone's new development will be focused on mobile devices.

ECONOMIC AND SOCIAL IMPACT

Social networking is a business. Making profit is the goal of most social networking services. Currently, both Facebook and Tencent have successfully commercialized their services. Online advertisement is the major source of revenue. Collecting membership fees is another way to monetize virtual social relations. In addition, social networking reduces the cost of some business operations, such as marketing and payment

Table 3. Distribution of desktop users and mobile users of the popular social networking services

	Facebook	QZone	Weibo	Tumblr	LinkedIn
Desktop	32%	26%	15%	46%	26%
Mobile	68%	74%	85%	54%	74%

Data source: Various websites on the Internet

transactions. New e-commerce models have emerged on social networking platforms. Two of these e-commerce models are micro business and ceWebrity business.

Micro business could be built on top of Facebook or WeChat platforms, where a user could sell products or services to his/her friends on the contact list. Products/services information could be posted online or could be broadcasted to friends. This kind of business model certainly reduces the cost of marketing. In addition, micro business could grow fast if the user's friends are willing to share the advertisement with their friends and their social groups, and so on and so forth. Micro business is rather common on WeChat platform, partly because Tencent's payment system is integrated into WeChat, which simplifies the payment transactions.

CeWebrity business could be built on top of Twitter or Sina Weibo platforms. This business is mainly operated or marketed by a celebrity on the Web, who could have thousands or millions of followers. The products or services are usually sold to the followers of the celebrity through broadcasting advertisement on social networking platforms. Because Twitter has a text size limit and does not support photo or video sharing, currently, it is not ideal for ceWebrity business. However, ceWebrity business grows fast on Sina Weibo's platform. Another common platform for ceWebrity business is YouTube. Celebrities on the Internet could post their products or services information on YouTube to make them available to the public, and especially to their subscribers. In this sense, YouTube is also a social networking platform similar to Instagram.

The effect of social networking goes beyond business. It affects every aspect of our life, including education, healthcare, entertainment, and more. Various new social groups and social relations are built on social networking platforms. Members in the virtual community can share their feelings, ideas, and experiences. They could also seek helps and advices from other community members. Users could play together, work to-

gether, and shop together. Using entertainment as an example, social networking has been used as a platform for live music shows. Moreover, other apps could be integrated into social networking apps, making it more convenient for social networking users to do different things, such as booking a ticket or paying utility bills.

Social networking not only broadens the horizon of our social relations, but also provides us better opportunities to develop mutual understandings in various social dimensions. Cross border and cross cultural relations could be easily developed with social networking. Through interacting with people of different cultures and different belief, we could better value other's living system and lifestyle. Social networking provides a cost-effective way to foster the respect of diversities. The traditional social dimension development is achieved through reading books, watching TVs/movies, browsing the internet, traveling, joining clubs, and participating in various events. Social networking provides additional mechanisms for social dimension development. It enables us to join remote cultural groups and form remote cross-cultural relations easily, allowing us to live in a virtually diverse environment. Therefore, in general social networking has a positive influence on individual's development of universal ethical principles, which are essential for us to resolve social and cultural conflicts and reduce social and cultural inequalities.

With the expansion of social networking services, people are joining in more and more virtual groups and spend more time on virtual socialization. For example, there are over 600 million groups on Facebook. Both Facebook, Qzone, and WeChat allow users to setup their own groups. Latest data shows that Facebook users spend twenty more minutes on it a day and China's WeChat users spend more than forty minutes on it a day. These numbers are expected to grow as people depend more and more on social networking. The latest development in this area shows that different social networking services are trying to connect their services together in order to

build more communities and larger communities. For example, WeChat has been integrated with LinkedIn, allowing WeChat users to invite their LinkedIn contacts to join WeChat.

FUTURE RESEARCH DIRECTIONS

It is reasonable to say that we are living in two worlds now, a real world through which we interact and communicate with others physically, and a virtual world which is built on top of social networking. In our virtual world, we interact and communicate with others remotely. Technology has helped us build this virtual world, and it will continue to improve our experience of virtual socialization. The technologies that are under development and might have great potential to enhance social networking services include VR (Virtual Reality), IoT (Internet of Things), and SE (Smart Environment).

VR (Virtual Reality) is the emerging technology that can simulate the physical environment and bring close-to-real experience for the users. For example, with the help of physical devices, such as special screens, head phones, and sensory devices, users could feel existing in a different environment like real. IoT (Internet of Things) is the network to connect all physical objects online, which also means these objects can be controlled remotely by users. With this technology, physically separated users are connected more closely, which allows physical interactions through remote objects. SE (Smart Environment) is the idea to build an environment with embedded sensors, displays, and computing devices. The main purpose of SE (Smart Environment) is to seamlessly merge human being into the environment, making us easy to feel, understand, and control the environment.

VR, IoT, and SE are closely related when they are going to be utilized in social networking. Although these technologies are still on their early stages of development, they are believed to be part of the driving forces for next economic boom. Once these technologies are matured and incorporated

into social networking services, the impact could be revolutionary. Potential applications include gaming, virtual sport, entertainment, and more. Technology and internet giants, like Microsoft, Google, Facebook, Tencent, and Alibaba, are all actively investing in these technologies.

CONCLUSION

In this chapter, we discussed the evolution of social networking together with its supporting technology. We also described the most popular social networking services, their features, and their business status. The new business and new social relations that are built on social networking platforms are described. The economic and social impact and future potentials of social networking development are also illustrated.

Now days, technology is an integral part of our daily life. We are utilizing technology, and at the same time technology is affecting our lifestyle. Social networking is a perfect example showing how human activities and human society are affected by technology.

We are living in a changing world and changing society: the borders between countries are disappearing and the distance between people is shortening. The world is getting smaller. As for the future, we really do not know how it will look like. But, we can imagine. With the advancements of technologies, our imagination could become real.

REFERENCES

- Branon, R. F., & Essex, C. (2001). Synchronous and asynchronous communication tools in distance education. *TechTrends*, 45(1), 36–36. doi:10.1007/BF02763377
- Carullo, G., Castiglione, A., De Santis, A., & Palmieri, F. (2015). A triadic closure and homophily-based recommendation system for online social networks. *World Wide Web (Bussum)*, 18(6), 1579–1601. doi:10.1007/s11280-015-0333-5

- Chianese, A., Piccialli, F., & Valente, I. (2015). Smart environments and cultural heritage: A novel approach to create intelligent cultural spaces. *Journal of Location Based Services*, 9(3), 209–234. doi:10.1080/17489725.2015.1099752
- Chu, S. (2009). *Managing Knowledge for Global and Collaborative Innovations* (Vol. 8). World Scientific.
- Grusec, J. E., & Lytton, H. (2012). *Social Development: History, Theory, and Research*. Springer Science & Business Media.
- Holmes, J. (2015). *Instagram: Instagram Blackbook: Everything You Need To Know About Instagram For Business and Personal - Ultimate Instagram Marketing Book Paperback*. CreateSpace Independent Publishing Platform.
- Kirkpatrick, D. (2011). *The Facebook Effect: the Inside Story of the Company that Is Connecting the World*. Simon and Schuster.
- Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is Twitter, a social network or a news media? In *Proceedings of the 19th International Conference on World Wide Web* (pp. 591–600). ACM. doi:10.1145/1772690.1772751
- Lenhart, A., & Madden, M. (2007). *Social Networking Websites and Teens: An Overview*. Pew/Internet.
- Leskovec, J., & Horvitz, E. (2008). Planetary-scale views on a large instant-messaging network. In *Proceedings of the 17th International Conference on World Wide Web* (pp. 915–924). ACM. doi:10.1145/1367497.1367620
- Lewis, C., & Fabos, B. (2005). Instant messaging, literacies, and social identities. *Reading Research Quarterly*, 40(4), 470–501. doi:10.1598/RRQ.40.4.5
- Malmgren, R. D., Hofman, J. M., Amaral, L. A., & Watts, D. J. (2009). Characterizing individual communication patterns. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 607–616). ACM. doi:10.1145/1557019.1557088
- Matic, A., Osmani, V., & Mayora-Ibarra, O. (2012). Analysis of social interactions through mobile phones. *Mobile Networks and Applications*, 17(6), 808–819. doi:10.1007/s11036-012-0400-4
- Merritt, T. (2012). *Chronology of Tech History*. Lulu.com.
- Papacharissi, Z. (2010). *A Networked Self: Identity, Community, and Culture on Social Network Sites*. Routledge.
- Rheingold, H. (2000). *The Virtual Community: Homesteading on the Electronic Frontier*. MIT press.
- Russell, M. A. (2013). *Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub, and More*. O'Reilly Media, Inc.
- Scheele, C. H. (1970). *A Short History of the Mail Service*. Smithsonian Books.
- Tse, E. (2015). *China's Disruptors: How Alibaba, Xiaomi, Tencent, and Other Companies Are Changing the Rules of Business*. Portfolio.
- Warschauer, M. (2004). *Technology and Social Inclusion: Rethinking the Digital Divide*. MIT press.
- Yu, L., Ramaswamy, S., & Zhang, C. (2008). Mining email archives and simulating the dynamics of open-source project developer networks. *The 4th International Workshop on Enterprise and Organizational Modeling and Simulation*, 17–31.
- Yu, L., Zhong, G., & Ramaswamy, S. (2016). The effect of time zone difference on asynchronous communications in global software development. *International Journal of Computer Applications in Technology*, 53(3), 213–225. doi:10.1504/IJCAT.2016.075523

Yu, Z., Zhou, X., Zhang, D., Schiele, G., & Becker, C. (2013). Understanding social relationship evolution by using real-world sensing data. *World Wide Web (Bussum)*, 16(5/6), 749–762. doi:10.1007/s11280-012-0189-x

ADDITIONAL READING

Atwan, G., & Lushing, E. (2011). *The Facebook Book*. Abrams.

Bandura, A. (1986). *Social Foundations of Thought and Action: A Social Cognitive Theory*. Prentice-Hall, Inc.

Carrington, P. J., Scott, J., & Wasserman, S. (Eds.). (2005). *Models and Methods in Social Network Analysis* (Vol. 28). Cambridge university press. doi:10.1017/CBO9780511811395

Diani, M., & McAdam, D. (Eds.). (2003). *Social Movements and Networks: Relational Approaches to Collective Action*. OUP Oxford. doi:10.1093/0199251789.001.0001

Flew, T. (2005). *New Media: An Introduction*. Oxford University Press.

Harasim, L. M. (1993). *Global Networks: Computers and International Communication*. MIT Press.

Horst, H. A., & Miller, D. (Eds.). (2013). *Digital Anthropology*. A&C Black.

Jackson, M. O. (2008). *Social and Economic Networks* (Vol. 3). Princeton university press.

Shih, C. (2009). *The Facebook Era: Tapping Online Social Networks to Build Better Products, Reach New Audiences, and Sell More Stuff*. Prentice Hall.

KEY TERMS AND DEFINITIONS

Asynchronous Communications: A communication mechanism that does not require the sender and the receiver to be online at the same time. A delay of response is allowed and expected for this kind of communication.

Celebrity on the Web: Someone who is famous on the Internet.

CeWebrrity Business: An e-commerce business operated and/or marketed by a celebrity on the Web. The products or services are usually sold to the followers of the celebrity through social networking.

Desktop Computers: A personal computer that is mainly used in a single location. Two representative desktop computers are IBM PC and Apple Mac.

Internet of Things: A technology that connects physical objects online so that users can control these objects remotely.

MAU (Monthly Active Users): A social networking service metric that counts the number of users who perform at least one activity (sending a message or reading a message) on the platform in one month period of time.

Micro Business: An e-commerce business built on top of social networking services, in most cases, mobile social networking services, where a store owner sells products/services to his/her friends in the app's contact list.

Mobile Devices: Smartphone and tablet computers that could be carried and used anywhere. Usually wireless network is needed for mobile devices to be connected to the Internet.

Smart Environment: The idea to build an environment with embedded sensors, displays, and computing devices so that users can better understand and control the environment.

Social Networking: A service based on internet technology that allows users to communicate and socialize remotely.

Synchronous Communications: A communication mechanism that requires the sender and the receiver to be online at the same time. An immediate response is expected for this kind of communication.

Virtual Community: A group formed on top of a social networking service. In most cases, group members have the same interest but are remotely located.

Virtual Reality: A technology that can bring close-to-real physical experience for the user.

Virtual Socialization: People socialize through using computer-based technology. No face-to-face interactions and communications are needed.

Social Media and Business Practices

Ashish Kumar Rathore

Indian Institute of Technology Delhi, India

P. Vigneswara Ilavarasan

Indian Institute of Technology Delhi, India

INTRODUCTION

In today's competitive environment, businesses are continuously exploring ways by which customers can be better engaged to sustain long-term growth (Aral, Dellarocas, & Godes, 2013). Because of the unpredictable and wide dispersion of social media (SM) platforms, businesses today face many challenges that did not exist a few years ago (Dong & Wu, 2015). SM platforms are economical and user-friendly and facilitate the consumption, generation, and sharing of user-generated content by the consumers (Sigala & Marinidis, 2009). As most of their targeted customers are present on various SM platforms, businesses are aligning their strategies and tactics to incorporate these platforms at all levels. This shift toward SM platforms can be determined by several factors including declining response rates vis-a-vis traditional customer engagement methods, technology development through customer participation, open communication of customer preferences, low cost of information dissemination, and the demographic shifts toward use of new technologies (Gillin, 2007).

SM platforms offer an easier and more cost-effective way for businesses to reach customers, and consequently strengthen brand awareness through numerous applications and tools (Rodriguez-Donaire, 2012). Using SM platforms, businesses can define or re-define relationships with both new and old users, respectively, and develop communities that interactively collaborate to identify issues and solutions for both themselves and businesses (Meredith & O'Donnell, 2011). Businesses appear to believe that such SM initiatives are justified

because of their potential to generate profits, for instance, through advertising (Aral et al., 2013). Businesses can improve branding and direct more traffic to its web site utilizing SM advertising.

Furthermore, customers can add value by generating their own content to influence the purchase decisions of others through peer-to-peer communications. SM platforms enhance the communication power of individuals by providing different avenues without demanding much effort of the users (Curran & Lennon, 2011). Apart from creating and sharing knowledge, users can create or join various communities with other like-minded individuals based on their similarities of interest and purpose. Also referred to as virtual communities, they help users create personal relationships in an enabling environment through emotion-laden discussions (Rheingold, 2000). These communities have resulted in new opportunities for businesses (Brodie, Hollebeek, Juri, & Ili, 2011). Some of the activities undertaken by businesses include sharing of content, interaction with customers, gathering customer feedback, provision of customer services, and effective collaboration with employees or business partners, and so on. (Bowden, 2009). Furthermore, SM has not only strengthened the existing relationship between businesses and users but also resulted in innovative changes in traditional communication methods, thereby enhancing the capability of businesses to better interact and dialog with users.

In light of above, this chapter attempts to explain SM in greater detail and as delineated by various business practices. This understanding of the use of SM by businesses is preceded by

a discussion on the definition of SM platforms, their different types, and the associated business models.

BACKGROUND

Social Media

The definition of SM has been evolving over a period of time. According to Terry (2009), it refers to “digital technologies emphasizing user-generated content or interaction” (p. 508). User-generated content supported through SM is “a mixture of fact and opinion, impression and sentiment, founded and unfounded titbits, experiences, and even rumor” (Blackshaw & Nazzaro, 2006: 4). The content available on these SM platforms consists of various pieces of online information which are generated and shared by users about brands, products, and services. Often SM is referred to by its channel characteristics, either identifying directionality of messages (Kent, 2013) or using particular tools such as Facebook or Twitter for engagement and communication (Howard & Parks, 2012).

Although there is a lack of a formal and concise definition, SM is often defined as Internet-based applications that transmit user-generated content. Some definitions are simply based on the nature of communication in SM. For instance, Russo, Watkins, Kelly, and Chan (2008: 22) explain SM as “those that facilitate online communication, networking, and/or collaboration.” Kaplan and Haenlein (2010: 61) offer a similarly definition of SM as “a group of Internet based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of User Generated Content.” In the same manner, Lewis (2009: 2) noted that SM simply serves as a “label for digital technologies that allow people to connect, interact, produce, and share content.”

In addition, Aula (2010: 43) defines SM as a place where customers can communicate directly

with their favorite organizations and gather more information about the organizations’ products instead of simply providing a platform for individuals to stay in touch with their family and friends.

These definitions could easily be applied to other communication technologies, such as e-mail, thereby missing the unique technological character that distinguishes SM.

Furthermore, a more extended and complex definition of SM is as follows:

(a) the information infrastructure and tools used to produce and distribute content; (b) the content that takes the digital form of personal messages, news, ideas, and cultural products; and (c) the people, organizations, and industries that produce and consume digital content (Howard and Parks, 2012: 362).

Kent (2010: 645) broadly defined SM as

any interactive communication channel that allows for two-way interaction and feedback, further by their potential for real-time interaction, reduced anonymity, a sense of propinquity, short response times, and the ability to “time shift,” or engage the social network whenever it suits each particular member.

The definitions become broader by combining SM and social network sites (SNSs). Boyd and Ellison (2007: 211) seminally defined SNSs as

web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system.

The social network is “one of the most typical features of social media in addition to the content aspect” (Enders, Hungenberg, Denker, & Mauch, 2008: 200).

SM allows individuals to meet strangers, but rather they enable users to articulate and make visible their social networks. It also allows users to “enhance their profiles by adding multimedia content or modifying the look and feel of their profiles” (Hanna, Rohm, & Crittenden, 2011: 266).

As discussed, definitions in earlier literature, SM have been defined by specific features, minimizing their unique communicative properties. To sum up all definitions, a new definition of SM given by Carr and Hayes (2015: 50) is as follows:

SM are Internet-based channels that allow users to opportunistically interact and selectively self-present, either in real-time or asynchronously, with both broad and narrow audiences who derive value from user-generated content and the perception of interaction with others.

As observed above, the definitions are diverse and continuously evolving. This chapter will use SM platform as an inclusive term and in an operational sense.

SM Platforms

SM platforms can be differentiated on the basis of technology, content, and functions. They represent different online sources: review website, blogs, discussion boards, chat rooms, and social networking sites, for creating and sharing of information and knowledge about businesses and their products/services (Blackshaw & Nazzaro, 2006). Kietzmann, Hermkens, McCarthy, and Silvestre (2011) identify seven fundamental building blocks in understanding SM in what has come to be known as the honeycomb model: *identity*—the extent to which customers choose to reveal their identities; *presence*—the extent to which customers know whether others are available; *relationships*—the extent to which customers relate to each other; *conversations*—the extent to which customers communicate with each other; *groups*—the extent to which customers are ordered or form communi-

ties; *reputations*—the extent to which customers know the social standing and content of others; and *sharing*—the extent to which customers exchange, distribute, and receive content. Irrespective of the name of the SM platforms, one can differentiate them on the basis of these functional blocks.

SM platforms can be categorized into the following types (Lietsala & Sirkkunen, 2008): SNSs (LinkedIn, Facebook, MySpace), collaborative productions (Wikipedia), content creation and publishing (blogs, v-blogs, podcasts), content sharing (Flickr, YouTube), virtual worlds (FarmVille, Second Life), add-ons (Slide, Friends for Sale), and social commerce (Tripadvisor, Groupon, and Facebook Connect).

The number of SM platforms’ users, worldwide, is continuously increasing at a phenomenal rate. It is predicted that in 2016 there will be around 2.13 billion SM users around the globe, up from 1.79 billion in 2014. Almost two-thirds (64%) of these users use or will use such SM platforms at least once a day through their computers. However, users are now increasingly using their smart phones to access SM platforms (Nielsen, 2014). Almost 47% of smart phone owners visit SM platforms every day. Furthermore, they are logging their presence with their profiles and connecting on to multiple SM platforms. While Facebook remains the largest SM platform, users are now taking to other social platforms such as LinkedIn, Pinterest, and Instagram.

REVENUE MODELS OF SM

SM platforms were thriving primarily on user-generated content and did not focus much on revenue generation in the early phases of their growth. With the increase in the number of SM platforms with hyper-specialized focus, the modes of revenue generation are also evolving. In literature, the most discussed revenue models are advertising models and subscription models (Enders et al., 2008; Vukanovic, 2011; Wang, Ye,

Zhang, & Nguyen, 2005). The other models are: licensing of content and technology to third parties, selling goods and services to the community, and voluntary donations (Canzer, 2006; Lietsala & Sirkkunen, 2008; Zeng & Reinartz, 2003).

Advertising-Based Models

Advertising in SM platforms is now an approach that is as commonly adopted by businesses as advertising in traditional media such as newspapers and magazines. SM platforms have begun to incorporate brand pages, ad videos, and additional user-generated content for revenue. In addition, users and businesses can create their own profiles or brand sites using SM platform to advertise their particular products or services. On SM platforms, two types of advertising can be observed: affiliate models and banner advertising. In banner advertising, fees are charged by the platforms in exchange for the display of advertisements on their websites (Canzer, 2006). Facebook and YouTube are well-known examples of banner advertising. In affiliate models, SM platforms direct traffic to an “affiliate” website and, in turn, might get a percentage of revenues from the resulting sales. The popular Japanese social networking site Mixi allows users to rate and review books, games, and other items and connects users directly to Amazon Japan with one click to facilitate the purchase of those items or to listen to music which can later be bought from the iTunes store. Facebook, for example, offers businesses the option to construct individual groups in return for a sponsorship. For SM platforms relying on an advertising model, it is essential that the intended audience be large and highly differentiated to increase revenues. Thus, the key revenue driver for this model is the number of users.

Subscription-Based Model

Some SM platforms charge users for accessing content which is generated on their site. This

model requires users to subscribe by paying a specified amount for accessing content (Enders et al., 2008) on the platform. In some practices, this is referred to as the Freemium model, in which users have two choices, either to subscribe to a basic free account package which guarantees a limited amount of content/services or a pro account, at an additional cost, that provides enhanced features. For example, LinkedIn initially offers free subscription only for limited features, but for advanced features a subscription fee is charged. Sometimes users are not willing to pay a subscription fee for a service just because of a lack of trust. A certain degree of user trust is needed because users need to be convinced for them to be willing to pay for a service (Wang et al., 2005). For that, SM platforms provide free trials for users to comprehend the service offering. A good example of a subscription-based model is The Auteurs, an SM platform built around classic films. Users need to subscribe for accessing movies. The crucial driver in subscription-based models is the creation of high levels of unique customer value, which then determines their willingness to pay for a service. The attractiveness of content becomes a significant factor when companies want to charge their customers for what is essentially free content.

Another model similar to subscription model is Pay-per-item model where users need to pay SM platforms for accessing particular content of interest to them. For example, iStockphoto offers audio, video, and photos from its user stock and charges a fixed amount for each. Such SM platforms could recommend or sell user-generated content as part of their own stock on a pay-per-item terms. YouTube, another example, sells videos as per a pay-per-item model. Users need to pay a particular amount mentioned in the video itself.

The platforms either serve as a host wherein users upload their content as in YouTube and Flickr or provide additional features for the additional fees as in LinkedIn. In both the cases, user-generated content is key.

Licensing of Content and Technology to Third Parties

Licensing content to third parties may also be a source of revenue because of the tremendous growth in user-generated content on SM. An SM platform can be licensed to other businesses who want to employ similar features or services for their individual use. Twitter, for example, gains more than 15% of its sales, that is, around \$47.5 million through such arrangements. It is generated by the licensing of its massive quantity of data to analytics businesses. Although this model is not as popular and profitable as advertising and subscription, it has the potential to grow into a huge revenue stream for Twitter. Sometimes it contains provision for the licensing of content on a revenue sharing basis between itself and content generators. YouTube, for example, primarily compensates the content creators through a program called Content License Agreement, and through revenue sharing arrangement a certain amount of money is made from some videos by both YouTube and the content creator.

Voluntary Donations

In this type of model, SM platforms can use the freely available user-generated content, like that of small girl singing on the street or musician performing, for donation from others. For this, a significant number of SM platforms such as blogs, wikis, online video, and online music users ask for donations using a donate-button available on these platforms for various purposes including content access, platform maintenance, and web hosting (Goh, Heng, & Lin, 2013). This facilitates web development such as open source and user-driven innovations. For example, Wikipedia gets moneys from donations from users and other individuals. Every year, there are banners asking users to donate some money to the Wikimedia Foundation, the non-profit organization which runs Wikipedia. A common feature of such SM platforms is that they run their operations with

limited funding (often thanks to pro-bono time invested by volunteers and users).

Selling Goods and Services to Community

SM platforms have large audiences because of their networking and community mobilization capabilities and this creates an opportunity for businesses to promote and sell products or services directly to these users. SM platforms incorporate gaming techniques on their networks, driven by the monetary opportunities that they present. The sale of virtual goods will remain the primary source of revenue for these platforms. Some SM platforms, for example, allow the sale of online virtual games and virtual accessories to their users. For instance, the Mypurchase service of MySpace offers a platform to creators to sell their music, taking a portion of sales revenues in exchange.

To summarize, these revenue/business models used by SM are very new approaches, and more empirical research is required to optimize these model for better returns. Many SM platforms are now trying to integrate various revenue models to maximize revenues. For example, SoundCloud, an audio sharing platform, uses a combination of subscriptions, selling goods, and revenue sharing models. Furthermore, businesses are trying to integrate SM platforms into their activities. In the next section, use of SM platforms for various business functions is described.

SM PLATFORMS IN BUSINESSES

Many businesses use SM platforms as an advertising channel only because they are accessed by the public at large. Hence, the most evident application of SM platforms in business is to use them for different kinds of online promotional activities (Mangold & Faulds, 2009). Mangold and Faulds (2009) stated that SM platforms are an opportunity to construct brands, exhibit leadership behaviors, enlarge resources, achieve new audiences, and

find new sources of ideas. Direction (2011) proposed a collective benchmark for businesses by classifying whether they can gain benefits from SM platforms or not after incorporating them into their business functions. An essential set of related questions is how businesses should engage and communicate with customers through SM platforms (Jussila, Karkkainen, & Leino, 2011). Andriole et al. (2010) enumerated six latent factors that are influenced using SM platforms, and that ultimately affect business performance: (1) collaboration and communication: capacity to coordinate discussions, to audit communication streams, and to reach more people faster; (2) rapid application development: capacity to engage experts, customers, suppliers, and company employees to modify and to develop applications easier and faster; (3) customer relationship management: solving customer service issues, using forums, wikis, and others, mine customer data effectively, ask for customer feedback, reach more customers, and communicate effectively with customers; (4) innovation: exchange of ideas between experts, fuelled by user-generated content and mass co-creation, capability to improve success rates, to produce efficiently; and (v) knowledge management: capabilities to share, retrieve, organize, and leverage knowledge. Some researches (Bruhn, Schoenmueller, & Scheafer, 2012; Hoffman & Fodor, 2010) show that businesses have already employed SM platforms for value proposition by incorporating it into their various business activities including marketing, recruiting, product innovation, and customer service. Needless to say, SM platforms will not become effective until they are used creatively, with a systematic approach and day-to-day effort toward building confidence in the brand and a loyal user community.

Customer Engagement

Brodie et al. (2011) defined *customer engagement* as “a psychological state that occurs by virtue of interactive, co-creative customer experiences with a business.” Because of these inherent interactive

characteristics, SM platforms may connect customers and businesses by facilitating conversation and providing user-generated content, thus fostering customer engagement. In addition to this, customers who actively participate in online activities on SM platforms are regarded as highly valuable for a business. This provides benefits in terms of customer commitment, trust, customers’ emotional connect, and loyalty (Brodie, Ilic, Juric, & Hollebeek, 2013). SM enables customers as well as non-customers to contribute in marketing mix decisions and value adding (Sashi, 2012). User-generated content from SM platforms can significantly improve customer loyalty and satisfaction, particularly as customer requirements transform over a period of time (Lorenzo-Romero, Constantinides, & Brünink, 2014).

Max Bupa¹, India’s premier health insurance company, uses its engagement platform, Get Help, on Facebook to provide services to customers daily. This innovative platform enables customers to buy a health policy using their Facebook account, interact and share their experience with Max Bupa, and also get instant customer service at their convenience, anywhere, anytime. This has resulted in an increase in positive conversations with complaints decreasing by more than 60%. The turnaround time has consequently decreased by 50%.

Brand Awareness

The brand is a symbol, feature, or any other observable mark of a product or company that differentiates it from others. They are markers of recognition and composite social phenomena (Schau, Muñiz, & Arnould, 2009). Nowadays, branding is considered as an ongoing social process (Fueller, Schroll, Dennhardt, & Hutter, 2012). The social character of brands and the significance of engagement in co-creating brand value (Fueller et al., 2012) reinforce the role of SM platform as a brand awareness channel. The participatory potential of SM enables the users to affect brands in multiple ways. The recent

researches might not focus on brand reputation but on related concepts such as customer trust, emotional appeal, and brand attitude and they are showing positive effects overall (Van Noort & Willemsen, 2011). Brand reputation is a measure of how users feel about, talk about, and act toward an organization's brand (Tsimonis & Dimitriadis, 2014). Users are more likely to choose businesses with a positive brand reputation and are willing to pay more for their products. Applying SM strategies for branding can definitely help businesses. To attract more consumers, businesses are constructing their presence on SM platforms after receiving in-depth information on where and how they or their brands are positioned on SM platforms (Hutton & Fosdick, 2011). Furthermore, it is found that users' online purchasing decision is dependent on the brand perception while using the SM (Chen & Xie, 2008). The online reviews serve as critical approach in measuring brand sentiment (Stelzner, 2012).

For example, India International Jewellery Week (IIJW), an initiative to showcase India's finest jewellery, has built awareness through SM platforms as the biggest international jewellery festival in India. Various communication strategies are adopted to create brand awareness for IIJW on SM platforms. IIJW-related images and coverage of live events are uploaded on Facebook. The use of hashtags like #IIJW and tagging the celebrities are also used for promoting the event. As a result, 1166 people saw IIJW live online and its Twitter followers increased by 50% in a span of 5 days. In addition to this, #IIJW was used by 581 users and about 6,98,920 people were reached through Twitter which accounts for 1,454,957 impressions.

Marketing

Marketing is defined as an effort to find the requirements of specific communities and delivering the desired service/product to them. Earlier businesses used various conventional marketing approaches to identify such communities and assess their requirements. Today, exclusive dependency on

traditional media for marketing is inadequate for businesses to survive. In addition to creating online presence, businesses need to know how to interact with their customers online. SM platforms provide opportunities to expand relationships between users and businesses (Hlavinka & Sullivan, 2011; Lipsman, Mudd, Rich, & Bruich, 2012; Mangold & Faulds, 2009). Chi (2011: 46) defines SM marketing as a "connection between brands and consumers, offering a personal channel and currency for user centred networking and social interaction." Many studies define SM platforms as a new marketing tool (Berinato & Clark, 2010; Paquette, 2013) and appropriate for awareness building, influence, and the attainment of marketing objectives that help improve marketing communication effectiveness and make for better marketing impact (Lipsman et al., 2012). Kumar, Bhaskaran, Mirchandani, and Shah (2013) hailed SM marketing as a customer-loyalty-building and promotional tool. As supported by available literature, if SM platforms are used effectively, brand advocates can be controlled and exploited which would result in both strategic and operational benefits (Chan & Guillet, 2011).

For example, Fork You Too, a casual dining cafe in Delhi NCR, decided to go to SM platforms for restaurant marketing campaigns. They created a campaign around Bollywood movie posters with a Fork You Too-style twist on Facebook, Twitter, and Instagram. The content was strictly aligned toward food, drinks, cuisine, and events with a quirky and fun factor. As a result, Fork You Too has a fan base of more than 25,000 which increased their fan base 12 times on Facebook. They have 200 plus followers on Twitter and 250 plus followers on Instagram.

Product Design and Development

Product design and development is defined as a process which starts with identifying customer requirements and ends with converting these into a product. Customers often share their thoughts and feeling on product or event through SM platforms.

Hence, SM platforms have empowered customers to circulate their creations and opinions and thus add new content. It has also become a new method for businesses to collect people's opinions and understand their preference. User-generated content may embed customers' experiences of a product, with information on product engagement (e.g., product features) and other related information (e.g., shopping experiences; Chua & Banerjee, 2013). The strategic use of SM platforms for developing product features and form user-generated content is a comparatively new area that connects marketing and product design and is a fine example of how two formerly distinct business functions are brought together by the advent of SM (Ng, 2013).

Starbucks, a coffee house, has demonstrated its versatility in engaging customers through various SM platforms, such as MyStarbucksIdea, Facebook, Twitter, and Foursquare (Gallaughier & Ransbotham, 2010). Starbucks closely followed the principle of "design with customers" in defining the role of customers, allowing them to play the role of creators and evaluators of ideas. As a result, Starbucks received more than 120,000 product ideas, 41,000 experience ideas, and 24,000 involvement ideas all posted on MyStarbucksIdea. Starbucks has more than 36 million likes on the Starbucks corporate page and has more than 180,000 users talking about it. Seventeen million users have visited this page on Facebook. It has 18,7000 tweets on its corporate blog with 5.95 million followers following and focusing on the Starbucks' brand on Twitter, and more than 40 million followers on Foursquare.

Sales

From the sales/business function perspective, previous studies have shown that direct communication between customers and suppliers is essential in capturing the expectations with regard to a product or service and the desired requirements (Moncrief, Marshall, & Rudd, 2015). Marshall, Moncrief, Rudd, and Lee (2012) argue that the

SM platforms provide the required information to customers and share strong evidence to support the positive linkages between use of SM platforms and business success. Furthermore, it is evident that online reviews on SM platforms positively influence product sales. The prevailing positive and negative sentiments emerging out of online conversations have a significant impact on sales (Sonnier, McAlister, & Rutz, 2011; Tirunillai & Tellis, 2012). Businesses need to embrace SM platforms and related technology strategies in sales for gaining a competitive advantage (Stockdale, Ahmed, & Scheepers, 2012).

LuLu Mall, one of the most happening malls in the state of Kerala, India, launched an NRI Shopping Festival to attract NRIs, tourists, and the general public to the mall during its promotion period. The main tool used for this purpose was SM platforms to create a buzz around the events and to reach out to potential shoppers. The Facebook page and Twitter handle were updated with these promotional posts on a daily basis. The promotional content was posted on Facebook page of LuLu hypermarket, Dubai. This page had more than 40,247 fans from the Indian subcontinent including Kerala. As a result, LuLu mall increased its sales by 29%. There was also a growth in the number of Facebook likes and fans: 1,51,164 likes to 1,77,268, an increase of 26,104, with reach among fans of 53,99,573 and an impression of 86,83,222.

Promotion

A promotional approach often refers to direct marketing, advertising, and interactive marketing to encourage their products/services (Thackeray, Neiger, Hanson, & McKenzie, 2008). With the arrival of advanced technologies on SM platforms, there is an enhanced potential for promotion by businesses (O'Reilly, 2005). In the case of social promotion, this information is about product-related behaviors, ideas, and services (Thackeray et al., 2008). Mayzlin (2006) examined user-generated content which is a combination of

business promotions and user recommendations using analytical models and found that it can still be influential despite the obvious promotional aim by businesses in such online settings.

MindShift Interactive, a leading digital marketing and research firm, completed a comprehensive SM campaign for HomeShop18's new "Shopping Makes Me Happy" jingle. The promotional campaign was built on Twitter revolving around happiness and excitement among key opinion leaders with influencers exchanging tweets with reasons/people that made them happy. As a result of this, HomeShop18 surprised people with vouchers to increase the happiness quotient. This garnered an overall outreach of 23 million users across SM, trending for more than 24 hours across India.

FUTURE RESEARCH DIRECTIONS

Aral et al (2013) highlighted four major areas of research: design and features, strategy and tactics, management and organization, and measurement and value from the perspective of three levels of analysis—users and society, platforms and intermediaries, and firms and industries. Although some of the above-identified gaps are being explored, the area of research-related SM platforms is still largely unexplored. There is also a need for research in terms of differences between information-rich and information-poor countries. Many businesses are located in the developing world and are targeting customers from both domestic and foreign markets and whose use of SM platforms is different. There is a need for research specifically focusing on businesses' adoption of SM platforms in the information-poor contexts of the developing world.

SM platforms are also breeding ground for negative sentiments vis-a-vis products. An impulsive customer can damage the brand despite it being his or her own mistake. Given the dynamic nature of SM platforms, the risk mitigation strategies need to be robust and implementable. The

other challenges involved include the lack of control over the content on the platforms owned by others and the control over users in terms of access and nature of content to be viewed. There is a need for research in the creation of risk mitigation frameworks based on empirical data available.

Further research is also required for developing methodological solutions to deal with the representativeness of the SM data. As the user-generated content is limited only to the users on the SM platforms, this might not be identical to the overall population. Because not all the SM platforms share the background information about the users, generalized validation of SM data for the populations needs to be strengthened.

CONCLUSION

SM is increasingly becoming an important component that is impacting different business functions. Despite the larger number of users, most SM platforms are primarily generating revenue from advertisements and through affiliate relationships. Businesses can use SM platforms for customer engagement, brand awareness building, marketing, product design and development, sales, and promotion. For each of these areas of business, extant literature was presented along with the examples. Future research in this domain should look at at least four areas design, strategy, management, and measurement from the perspective of users, platforms, and firms. A focus on the developing world as the context will enrich the field by showcasing the differential adoption strategies adopted by or available to businesses. The development of risk mitigation frameworks will be helpful to businesses in dealing with the negative sentiments and other risks associated with the emotional outburst of users. The representativeness of SM data is still an open question and further work to strengthen the methodological rigor will benefit the field greatly.

ACKNOWLEDGMENT

The authors are thankful to Ms. Nalini Srinivasan for wonderful copyediting.

REFERENCES

- Andriole, G. L., Bostwick, D. G., Brawley, O. W., Gomella, L. G., Marberger, M., Montorsi, F., & Rittmaster, R. S. et al. (2010). Effect of dutasteride on the risk of prostate cancer. *The New England Journal of Medicine*, 362(13), 1192–1202. doi:10.1056/NEJMoa0908127 PMID:20357281
- Aral, S., Dellarocas, C., & Godes, D. (2013). Introduction to the special issue-social media and business transformation: A framework for research. *Information Systems Research*, 24(1), 3–13. doi:10.1287/isre.1120.0470
- Aral, S., & Walker, D. (2011). Creating social contagion through viral product design: A randomized trial of peer influence in networks. *Management Science*, 57(9), 1623–1639. doi:10.1287/mnsc.1110.1421
- Aula, P. (2010). Social media, reputation risk and ambient publicity management. *Strategy and Leadership*, 38(6), 43–49. doi:10.1108/10878571011088069
- Berinato, S., & Clark, J. (2010). Six ways to find value in Twitter's noise. *Harvard Business Review*, 88(6), 34–35.
- Blackshaw, P., & Nazzaro, M. (2006). *Consumer-generated media (CGM) 101: Word-of-mouth in the age of the web-fortified consumer*. New York: Nielsen BuzzMetrics.
- Bowden, J. L.-H. (2009). The process of customer engagement: A conceptual framework. *Journal of Marketing Theory and Practice*, 17(1), 63–74. doi:10.2753/MTP1069-6679170105
- Boyd, D., & Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1), 210–230. doi:10.1111/j.1083-6101.2007.00393.x
- Brodie, R. J., Hollebeek, L. D., Juri, C. B., & Ili, C. A. (2011). Customer engagement: Conceptual domain, fundamental propositions, and implications for research. *Journal of Service Research*, 14(3), 252–271. doi:10.1177/1094670511411703
- Brodie, R. J., Ilic, A., Juric, B., & Hollebeek, L. (2013). Consumer engagement in a virtual brand community: An exploratory analysis. *Journal of Business Research*, 66(1), 105–114. doi:10.1016/j.jbusres.2011.07.029
- Bruhn, M., Schoenmueller, V., & Scheafer, D. B. (2012). Are social media replacing traditional media in terms of brand equity creation? *Management Research Review*, 35(9), 770–790. doi:10.1108/01409171211255948
- Canzer, B. (2006). *E-business: Strategic thinking and practice*. Boston, MA: Cengage Learning.
- Carr, C. T., & Hayes, R. A. (2015). Social Media: Defining, Developing, and Divining. *Atlantic Journal of Communication*, 23(1), 46–65. doi:10.1080/15456870.2015.972282
- Chan, N. L., & Guillet, B. D. (2011). Investigation of social media marketing: How does the hotel industry in Hong Kong perform in marketing on social media websites? *Journal of Travel & Tourism Marketing*, 28(4), 345–368. doi:10.1080/10548408.2011.571571
- Chen, Y., & Xie, J. (2008). Online consumer review: Word-of-mouth as a new element of marketing communication mix. *Management Science*, 54(3), 477–491. doi:10.1287/mnsc.1070.0810
- Chi, H. H. (2011). Interactive digital advertising vs. virtual brand community: Exploratory study of user motivation and social media marketing responses in Taiwan. *Journal of Interactive Advertising*, 12(1), 44–61. doi:10.1080/15252019.2011.10722190

- Chua, A. Y., & Banerjee, S. (2013). Customer knowledge management via social media: The case of Starbucks. *Journal of Knowledge Management*, 17(2), 237–249. doi:10.1108/13673271311315196
- Curran, J., & Lennon, R. (2011). Participating in the conversation: Exploring adoption of online social media. *Academy of Marketing Studies Journal*, 15(1), 21–38.
- Direction, S. (2011). Friends in online places: Business and social media. *Strategic Direction*, 27(10), 19–22. doi:10.1108/02580541111171229
- Dong, J. Q., & Wu, W. (2015). Business value of social media technologies: Evidence from online user innovation communities. *The Journal of Strategic Information Systems*, 24(2), 113–127. doi:10.1016/j.jsis.2015.04.003
- Enders, A., Hungenberg, H., Denker, H. P., & Mauch, S. (2008). The long tail of social networking. Revenue models of social networking sites. *European Management Journal*, 26(3), 199–211. doi:10.1016/j.emj.2008.02.002
- Fueller, J., Schroll, R., Dennhardt, S., & Hutter, K. (2012, January). Social brand value and the value enhancing role of social media relationships for brands. In *System Science (HICSS), 2012 45th Hawaii International Conference on* (pp. 3218–3227). IEEE. doi:10.1109/HICSS.2012.533
- Gallaugh, J., & Ransbotham, S. (2010). Social media and customer dialog management at Starbucks. *MIS Quarterly Executive*, 9(4), 197–212.
- Gillin, P. (2007). *The new influencers: A marketer's guide to the new social media*. Sanger, CA: Quill Driver Books.
- Goh, K. Y., Heng, C. S., & Lin, Z. (2013). Social media brand community and consumer behavior: Quantifying the relative impact of user-and marketer-generated content. *Information Systems Research*, 24(1), 88–107. doi:10.1287/isre.1120.0469
- Hanna, R., Rohm, A., & Crittenden, V. L. (2011). Were all connected: The power of the social media ecosystem. *Business Horizons*, 54(3), 265–273. doi:10.1016/j.bushor.2011.01.007
- Hlavinka, K., & Sullivan, J. (2011). *Urban Legends: Word-of-Mouth Myths, Madvocates and Champions Debunking Myths About Brand Conversations and Word-of-Mouth in Social Media*. Colloquy.
- Hoffman, D. L., & Fodor, M. (2010). Can You Measure the ROI of Your Social Media Marketing. *MIT Sloan Management Review*, 52(1), 41–49.
- Howard, P. N., & Parks, M. R. (2012). Social media and political change: Capacity, constraint, and consequence. *Journal of Communication*, 62(2), 359–362. doi:10.1111/j.1460-2466.2012.01626.x
- Hutton, G., & Fosdick, M. (2011). The Globalization of Social Media: Consumer Relationships with Brands Evolve in the Digital Space. *Journal of Advertising Research*, 51(4), 564–570. doi:10.2501/JAR-51-4-564-570
- Jussila, J. J., Karkkainen, H., & Leino, M. (2011, September). Benefits of social media in business-to-business customer interface in innovation. In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments* (pp. 167–174). ACM. doi:10.1145/2181037.2181065
- Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of social media. *Business Horizons*, 53(1), 59–68. doi:10.1016/j.bushor.2009.09.003
- Kent, M. L. (2010). Directions in social media for professionals and scholars. In R. L. Heath (Ed.), *Handbook of Public Relations* (pp. 643–656). Thousand Oaks, CA: Sage.
- Kent, M. L. (2013). Using social media dialogically: Public relations role in reviving democracy. *Public Relations Review*, 39(4), 337–345. doi:10.1016/j.pubrev.2013.07.024

- Kietzmann, J. H., Hermkens, K., McCarthy, I. P., & Silvestre, B. S. (2011). Social media? get serious! Understanding the functional building blocks of social media. *Business Horizons*, 54(3), 241–251. doi:10.1016/j.bushor.2011.01.005
- Kumar, V., Bhaskaran, V., Mirchandani, R., & Shah, M. (2013). Practice prize winner-creating a measurable social media marketing strategy: Increasing the value and ROI of intangibles and tangibles for hokey pokey. *Marketing Science*, 32(2), 194–212. doi:10.1287/mksc.1120.0768
- Lewis, B. K. (2009). *Social media and strategic communications: Attitudes and perceptions among college students* (Doctoral dissertation). Oklahoma State University, OK, USA.
- Lietsala, K., & Sirkkunen, E. (2008). *Social Media: Introduction to the tools and processes of participatory economy*. Tampere University Press.
- Lipsman, A., Mudd, G., Rich, M., & Bruich, S. (2012). The power of "like": How brands reach (and influence) fans through social-media marketing. *Journal of Advertising Research*, 52(1), 40–52. doi:10.2501/JAR-52-1-040-052
- Lorenzo-Romero, C., Constantinides, E., & Brünink, L. A. (2014). Co-creation: Customer Integration in Social Media Based Product and Service Development. *Procedia: Social and Behavioral Sciences*, 148, 383–396. doi:10.1016/j.sbspro.2014.07.057
- Mangold, G., & Faulds, D. (2009). Social media: The new hybrid element of the promotion mix. *Business Horizons*, 52(4), 357–365. doi:10.1016/j.bushor.2009.03.002
- Marshall, G. W., Moncrief, W. C., Rudd, J. M., & Lee, N. (2012). Revolution in sales: The impact of social media and related technology on the selling environment. *Journal of Personal Selling & Sales Management*, 32(3), 349–363. doi:10.2753/PSS0885-3134320305
- Mayzlin, D. (2006). Promotional chat on the Internet. *Marketing Science*, 25(2), 155–163. doi:10.1287/mksc.1050.0137
- Meredith, R., & O'Donnell, P. (2011). A framework for understanding the role of social media in business intelligence systems. *Journal of Decision Systems*, 20(3), 263–282. doi:10.3166/jds.20.263-282
- Moncrief, W. C., Marshall, G. W., & Rudd, J. M. (2015). Social media and related technology: Drivers of change in managing the contemporary sales force. *Business Horizons*, 58(1), 45–55. doi:10.1016/j.bushor.2014.09.009
- Ng, C. S. P. (2013, June). Factors affecting the business performance of firms utilizing social media. Paper presented at the 17th Pacific Asia Conference on Information Systems (PACIS), Jeju Islands, Korea..
- Nielsen. (2014). *The Digital Consumer Report*. Retrieved March 24, 2015, from <http://www.nielsen.com/us/en/insights/reports/2014/the-us-digital-consumer-report.html>
- O'Reilly, T. (2005). *Web 2.0: compact definition*. Message posted to http://radar.oreilly.com/archives/2005/10/web_20_compact_definition.html
- Paquette, H. (2013). *Social Media as a Marketing Tool: A Literature Review*. Available at http://digitalcommons.uri.edu/cgi/viewcontent.cgi?article=1001&context=tmd_major_papers
- Rheingold, H. (2000). *The virtual community* (Revised edition). Reading: Addison-Wesley Publishing Company.
- Rodriguez-Donaire, S. (2012, June). Changing business model dynamics by means of social media. In *Management of Innovation and Technology (ICMIT), 2012 IEEE International Conference on* (pp. 370–377). IEEE. doi:10.1109/ICMIT.2012.6225834

- Russo, A., Watkins, J., Kelly, L., & Chan, S. (2008). Participatory communication with social media. *Curator: The Museum Journal*, 51(1), 21–31. doi:10.1111/j.2151-6952.2008.tb00292.x
- Sashi, C. M. (2012). Customer engagement, buyer-seller relationships, and social media. *Management Decision*, 50(2), 253–272. doi:10.1108/00251741211203551
- Schau, H. J., Muñiz, A. M. Jr, & Arnould, E. J. (2009). How brand community practices create value. *Journal of Marketing*, 73(5), 30–51. doi:10.1509/jmkg.73.5.30
- Sigala, M., & Marinidis, D. (2009). Exploring the transformation of tourism firms' operations and business models through the use of web map services. In *European and Mediterranean Conference on Information Systems* (pp. 1–13). London, UK: Academic Press.
- Sonnier, G. P., McAlister, L., & Rutz, O. J. (2011). A dynamic model of the effect of online communications on firm sales. *Marketing Science*, 30(4), 702–716. doi:10.1287/mksc.1110.0642
- Stelzner, M. (2012). *2012 Social media marketing industry report*. Retrieved April 13, 2015 from <http://www.socialmediaexaminer.com/SocialMediaMarketingIndustryReport2012.pdf>
- Stockdale, R., Ahmed, A., & Scheepers, H. (2012, September). Identifying business value from the use of social media: An SME perspective. In Pacific Asia Conference on Information Systems (PACIS) proceedings (p. 169).
- Terry, M. (2009). Twittering healthcare: Social media and medicine. *Telemedicine Journal and e-Health*, 15(6), 507–510. doi:10.1089/tmj.2009.9955 PMID:19659410
- Thackeray, R., Neiger, B. L., Hanson, C. L., & McKenzie, J. F. (2008). Enhancing promotional strategies within social marketing programs: Use of Web 2.0 social media. *Health Promotion Practice*, 9(4), 338–343. doi:10.1177/1524839908325335 PMID:18936268
- Tirunillai, S., & Tellis, G. J. (2012). Does chatter really matter? Dynamics of user-generated content and stock performance. *Marketing Science*, 31(2), 198–215. doi:10.1287/mksc.1110.0682
- Tsimonis, G., & Dimitriadis, S. (2014). Brand strategies in social media. *Marketing Intelligence & Planning*, 32(3), 328–344. doi:10.1108/MIP-04-2013-0056
- Van Noort, G., & Willemsen, L. M. (2011). Online damage control: The effects of proactive versus reactive webcare interventions in consumer-generated and brand-generated platforms. *Journal of Interactive Marketing*, 26(3), 131–140. doi:10.1016/j.intmar.2011.07.001
- Vukanovic, Z. (2011). New media business models in social and web media. *Journal of Media Business Studies*, 8(3), 51–67. doi:10.1080/16522354.2011.11073526
- Wang, C. L., Ye, L. R., Zhang, Y., & Nguyen, D. D. (2005). Subscription to fee-based online services: What makes consumer pay for online content. *Journal of Electronic Commerce Research*, 6(4), 304–311.
- Zeng, M., & Reinartz, W. (2003). Beyond online search: The road to profitability. *California Management Review*, 45(2), 107–130. doi:10.2307/41166168

ADDITIONAL READING

- Blanchard, O. (2011). *Social media ROI: Managing and measuring social media efforts in your organization*. New Delhi: Pearson Education.
- Casalo, L. V., Flavian, C., & Guinaliu, M. (2008). Promoting consumers participation in virtual brand communities: A new paradigm in branding strategy. *Journal of Marketing Communications*, 14(1), 19–36. doi:10.1080/13527260701535236

Culnan, M. J., McHugh, P. J., & Zubillaga, J. I. (2010). How large US companies can use Twitter and other social media to gain business value. *MIS Quarterly Executive*, 9(4), 243–259.

Jussila, J. J., Karkkainen, H., & Aramo-Immonen, H. (2014). Social media utilization in business-to-business relationships of technology industry firms. *Computers in Human Behavior*, 30, 606–613. doi:10.1016/j.chb.2013.07.047

Lovett, J. (2011). *Social media metrics secrets*. Indianapolis, IN: Wiley Publishing, Inc.

Ngai, E. W., Moon, K. L. K., Lam, S. S., Chin, E. S., & Tao, S. S. (2015). Social media models, technologies, and applications: An academic review and case study. *Industrial Management & Data Systems*, 115(5), 769–802. doi:10.1108/IMDS-03-2015-0075

Rodriguez, M., Peterson, R. M., & Krishnan, V. (2012). Social medias influence on business-to-business sales performance. *Journal of Personal Selling & Sales Management*, 32(3), 365–378. doi:10.2753/PSS0885-3134320306

Siricharoen, W. V. (2012). Social Media for Supporting Businesses Innovatively in Digital World. *International Journal of Engineering Research and Development*, 3(8), 01-08.

Sterne, J. (2010). *Social media metrics: How to measure and optimize your marketing investment*. Hoboken, NJ: John Wiley.

Tuten, T. L., & Solomon, M. R. (2013). *Social Media Marketing*. Boston, MA: Pearson.

KEY TERMS AND DEFINITIONS

Advertising Model: A revenue model for SM platforms in which brand pages and ad videos

containing product/service information are created on the platform.

Brand Awareness: It is a measure of how identifiable the brand is for target customers.

Brand Community: It is a structured composite of social relationships among the admirers of a brand.

Brand Reputation: It is defined as a measure of how customers react to brand (what they think about the brand, how they talk about it and how their inclination towards the brand).

Customer Engagement: Defines the behaviours evoked in a customer by a specific brand that are over and above a basic connect.

Online Product Review: It is an account on SM of a customer's personal experience vis-a-vis a product.

Revenue Sharing: It means that the revenue made from user generated content, SM platform share some part of it to content creator.

Social Media Marketing: It is a term which describes the actual activities involved in using SM platforms for marketing purposes.

Social Media: Online platforms in which users can create and share the content.

Social Promotion: It can be defined as direct marketing, advertising, and interactive marketing using SM platforms to encourage the sale of products/services.

Subscription Model: In this revenue model, the subscription fees charged for accessing content or services available on SM platform constitute the revenue source.

ENDNOTE

¹ All examples for this chapter are sourced from www.socialsamosa.com.

Social Media Credit Scoring

Billie Anderson

Ferris State University, USA

J. Michael Hardin

Samford University, USA

INTRODUCTION

Credit scoring is a method of modeling potential risk of credit applicants. It involves using different statistical techniques and past historical data to create a credit score that financial institutions use to assess credit applicants in terms of risk. Credit scoring is essentially a type of classification problem: which credit applicants should be considered good risks and which applicants should be considered bad risks.

A scorecard model is built from a number of characteristic inputs. Each characteristic is comprised of a number of attributes. In the example scorecard shown in Figure 1, age is a characteristic and “25–33” is an attribute. Each attribute is associated with a number of scorecard points. These scorecard points are statistically assigned to differentiate risk, based on the predictive power of the variables, correlation between the variables, and business considerations.

For example, in Figure 1, the credit application of a 32 year old person, who owns his own home and makes \$30,000, would be accepted for credit by this institution. The total score of an applicant is the sum of the scores for each attribute present in the scorecard. Lower scores imply a higher risk of default, and higher scores imply lower risk.

Credit, as it has evolved since the 1950’s, it is cold and impersonal, completely based on numbers--what you owe, what you’ve paid, how much money you have. The numbers all get hacked down to one number: a credit score. The bigger

and more complex the global financial system gets, the less it cares about anything other than the applicant’s credit score. Banking institutions are becoming less and less personal with online and mobile banking applications. Very few people know the local bank manager at their hometown bank anymore. A credit applicant’s personal story is not taken into account when applying for a financial loan. The credit applicant’s “character” fits into the bank’s algorithms about as well as peanut butter fits into the workings of a Swiss watch.

One consequence of this system is that it is biased against poor people, people with no bank accounts or very little credit history (“thin files”) and young people. These types of applicants cannot obtain credit because they cannot generate the financial metrics that would provide evidence to determine if these applicants should be granted or rejected for a financial loan. It is these several billion people who would be the most enthusiastic first customers of a new kind of credit based less on numbers and more on character. The need to market to and provide credit to these types of applicants marks the emergence of a social media credit score.

This chapter will describe how banks and financial organizations are starting to incorporate big data sources, such as data from social media websites, into the credit lending process. A discussion of how more established organizations, such as FICO and SAS, are incorporating big data in their scorecard methodology will be given. A description of two start-up companies

Figure 1. Example scorecard

Example Scorecard

Let Cutoff=500

A new customer applies for credit.

AGE 32 120 points

HOUSE OWN 225 points

INCOME \$45K 200 points

Total 545 points

ACCEPT FOR CREDIT

Characteristic Name	Attribute	Scorecard Points
AGE	Up to 25	100
AGE	26-33	120
AGE	34-45	185
AGE	45+	225
HOUSE	OWN	225
HOUSE	RENT	110
INCOME	Up to \$10K	120
INCOME	10K-25K	140
INCOME	26K-35K	160
INCOME	36K-50K	200
INCOME	50K+	240

which are using electronic and big data sources such as social media to provide banking services and grant loans will be discussed.

BACKGROUND

The statistical methods used to categorize objects into groups can be traced to 1936 in Fisher's publication (Fisher, 1936). Durand (1941) was the first to use Fisher's methodology to distinguish between good and bad loans. Using this research, the founders of credit scoring, Bill Fair and Earl Isaac, built the first credit scoring system for the United States in 1958. Although credit scoring has been in use since that time, it is only recently that credit scoring has become widespread.

There are some older, yet still very relevant, credit scoring resources that discuss the statistical issues in developing a credit scorecard (Siddiqi, 2006; Thomas, Oliver, & Hand, 2005; Hand & Henley, 1997). These references provide a traditional framework that credit analysts have used in the past to develop credit scores.

As with any other business sector that is highly data dependent, the credit scoring world

has felt the impact of the big data phenomenon that is sweeping through modern businesses (Lohr, 2015). Banks are moving away from using traditional statistical techniques to build a credit scorecard. Financial organizations are combining the traditional techniques with new big data technology and big data analytics to build their scorecards. With the rise of companies using social media data as a tool to build better predictive models, financial institutions are integrating big data sources in calculating credit risk (Fei et al., 2015). Recently, even Facebook is hinting at entering the social media credit arena. Facebook has recently made its Messenger application the ability to make financial payments, similar to Apple Pay. What is interesting about Facebook's decision is that it will allow Facebook to collect data on their customer's financial payments and coupled with the already-abundant social media data it has, this decision allows Facebook to become a stand-alone financial lending institution that can incorporate character into the credit score (Maney, 2016).

Wei et al. (2015) provide the most contemporary and comprehensive review of incorporating social media data into an applicant's credit score.

The article examines credit scoring models that are built with and without social media data in order to study whether the incorporation of social media data into a credit score provides an advantage to financial lending companies. The article also examines whether potential credit applicants change their social media profiles and perhaps start making more strategic social media connections once a credit applicant knows they are being assessed based on their social media activity and persona. And, lastly the paper examines how incorporating social media provides a more even-level playing field for some of those disenfranchised groups that have been traditionally overlooked.

MAIN FOCUS OF THE ARTICLE

In the last several years, financial institutions have begun exploring ways to incorporate an applicant's social media activity such as Facebook and Twitter into the risk associated with a loan. The main advantage for a financial institution using such expanded data sources is the shorter time frame to deploy a credit scoring model. In a traditional credit scoring scenario, it can take 12 to 18 months to update a scorecard model with new data. This time lag results from lenders having to verify the new data, determining if the new data impacts other departments of the institution, and taking time for employees to learn how to use the new information to make decisions on credit worthiness, pricing and cross-selling (Adams, 2012).

Another advantage of using big data sources such as social media data is that social media data allows financial institutions with a broader view of the applicant's background other than traditional data sources, such as the bank's own customer data. The type of data collected from social media sources is wide and varied and provides insight into an applicant's lifestyle and behaviors. Social media data can provide genuine demonstrations of behavior that point to an applicant's capability and likelihood of meeting repayment terms. For those applicants

with "thin files" or young people, utilizing social media data to establish a credit footprint may be a sensible idea (Thomas, 2015).

By using big data sources and the associated big data analytics, a bank can increase:

- **Efficiency:** Is essential to banks that collect and analyze vast volumes of data. Accessing and analyzing various types of structured and unstructured data as soon as it is captured is essential in staying relevant to electronic banking customers.
- **Profitability:** Big data sources allow a bank to reach customers in ways that were never possible. By using online banking and social media applications, a bank can correspond with a customer at precisely the right time through online campaigns and messaging.
- **Customer Interactions:** Many banks are using big data and social media sources to shift their focus from product-centric to customer-centric. The bank's goal is to understand what the customer needs from each banking interaction so that each point of contact with the bank will be ideal. Big data analytics mine millions and billions of rows of customer and transaction data in order to develop specialized campaigns and offers for the customer.

A specific facet of big data is information collected from social media. Using social media data is still an evolving avenue for many financial organizations. Social media is one of the most effective ways to find out what customers like and do not like, as well as the type of banking relationship the customer expects from the financial institution. Social media can be used as a marketing tool for a bank; in a recent article, Ally Bank utilized social media exclusively to promote retirement products (DePaula, 2015). The aim of the social media campaign was to encourage younger people to start thinking about and saving for retirement but also to engage a retiree crowd over the age

of 65. Some of the social media campaigns Ally developed was to give away \$150 gift cards to generate interest in savings accounts on Pinterest, created Google hangouts and Twitter chats so that people could talk about retirement and savings, and created YouTube videos. The benefit of creating a social media marketing campaign for Ally is that the interest does not immediately dwindle like a traditional television or radio ad. Bank customers have the ability to keep the conversation going.

Another aspect of social media that banks are leveraging is customer activity on social media sites; this would be used to calculate part of their credit score. Banks hope to use social media activity to develop a sort of social media credit score. This concept is based on the traditional credit score, but it is very difficult to define exactly how to calculate a social media credit score. Currently, a social media credit score concerns how others in your social network view you. For example, what would be the impressions of a bank official if he were to research you? Are you an influential person -- that is, do you have the power to persuade others in your social network to buy a product you recommend, or sign up for a particular service? Are you positive or negative regarding the social media decisions you make? Financial institutions are currently turning the answers into these questions to develop a financial measurement and use that measurement either as a component of traditional credit scores or as the sole credit worthiness measurement of a customer.

What exactly constitutes social media activity? Below is a list of common social media outlets that banks access to develop a social media credit score.

- **Facebook:** The world's largest social media site is a wealth of information for anyone permitted access using the 'friend' status. Anyone who is a friend can see your profile and information unless the appropriate privacy settings are activated.
- **Twitter:** 140 characters can say a lot about someone, as this messaging system proves. Many Twitter fans tweet about themselves

daily. Searching Twitter can reveal any comment a person has made about themselves or someone else.

- **LinkedIn:** LinkedIn is an online resume network. You can post current and past jobs, active organizations, and awards.

One way that financial institutions are measuring a customer's social media influence is using a Klout score. Klout is a social credit score (Wagstaff, 2012). Klout measures an individual's influence online, with influence defined as the ability to drive others to action. When online content is produced, Klout looks at how the individual's network responds to that content. Klout takes in data from social networks and gives an overall score based on online influence. The data is from public networks such as Facebook, Twitter, Google+, LinkedIn, and Foursquare. Klout process about 3 billion pieces of content and connections daily and produces a new, accurate score for each Klout member every morning (Fernandez, 2011).

In the sections that follow, we will examine how organizations are incorporating big data sources such as social media into their banking and credit scoring applications.

FICO

FICO is a leading analytics software company, helping businesses in 90+ countries make better decisions that drive higher levels of growth, profitability and customer satisfaction. The company's groundbreaking use of big data and statistical algorithms to predict consumer behavior has transformed entire industries. FICO provides analytics software and tools used across multiple industries to manage risk, fight fraud, build more profitable customer relationships, optimize operations and meet strict government regulations. Probably the most well-known analytical tool is the FICO score, the standard measure of consumer credit risk in the United States.

FICO has recently started experimenting with incorporating social media data into their

credit scores. FICO has determined, using big data algorithms to analyze social media, that if a potential credit applicant has used the word ‘wasted’ in their profile that it has predictive power of whether the applicant is going to repay a loan or not (Hardekop, 2015). However, just because FICO has figured out this ‘nugget of gold’ in the big data landscape, it does not mean that FICO is using social media data to construct a credit score. In fact, FICO does not use social media data to calculate a credit score due to regulatory restrictions. However, FICO is piloting a new score called FICO Score XD that does incorporate big data sources. This credit score is aimed at the cohort of individuals that the Introduction section discusses; people who have no credit history or are invisible to the credit scoring system. FICO Score XD incorporates mobile phone, utility, and rent payment histories, property and tax data are also incorporated into the credit score. According to Christina Goeth, a FICO spokeswoman,

FICO Score XD is currently designed to only score consumers that are not scorable with traditional credit data. The algorithm checks to determine if a traditional FICO Score can be generated first, and if it can, the traditional score is returned to the lender. If it cannot, FICO Score XD provides a second chance to get approved. The goal of FICO Score XD is to expand access to credit. (Selyukh, 2016)

SAS

SAS is well known for its analytical software. SAS offers a series of products aimed at credit risk managers, such as SAS Credit Scoring for Banking and SAS Real Time Decision Manager. Credit Scoring for Banking mines and organizes data, then uses that information to develop consumer scorecards and perform compliance with Basel II requirements for management of credit risks for pools of loans. Decision Manager produces real-time recommendations via data channels such as websites, call centers, and ATMs to aid

a bank in making credit management decisions such as loan approvals and credit card increases. To further speed updates and expansion of data used in scoring, SAS’s High Performance Analytics searches data sets that include billions of information items to locate correlations between different data sources that can be combined to accelerate the development of new predictive scorecards.

Lenddo

One of the larger players in the social media credit scoring business is Lenddo. Lenddo is one of the more established companies, with its inception in 2011. The co-founders of Lenddo realized that more than one billion people in developing countries were starting to move into the middle class and would need access to credit (King, 2014). The initial intent of Lenddo was to offer small, targeted loans to individuals in developing countries. The purpose of the loans was to give the rising middle class in developing countries access to credit, as many individuals in these countries lack credit histories because there is no financial infrastructure for an applicant to establish a traditional credit history. The purpose of Lenddo loans must be to enrich the customer’s life through education, healthcare, or home improvement. These loans could work well for an individual who wants to take a career-development class, but lacks cash and cannot qualify for a traditional loan due to a lack of credit history. The applicant can leverage what he or she does have -- a connected network of online friends and colleagues.

Lenddo uses non-traditional data sources to compute applicant’s credit scores. Lenddo uses only social media data to determine the creditworthiness of applicants (Anderson & Hardin, 2014). To apply for a loan, the first step in Lenddo is to create a profile and upload a photo to the website. Then, the individual connects to social networks and invites friends and family to be part of what Lenddo calls the “trusted network.” Lenddo uses a social network algorithm to evaluate the “trusted

network” and develop a Lenddo score. The higher the Lenddo score, the more access to loans an applicant will have. If the applicant is granted the loan, the applicant’s entire social network is kept informed of the applicant’s payment history (Anderson & Hardin, 2014).

Moven

Moven, formerly known as Movenbank is currently generating excitement with their use of social media data in a banking environment. Moven is a completely online and mobile company; its motto is “no paper, no plastic, no hidden fees” (King, 2012). The bank has no physical locations and will not issue any physical debit or credit cards. The premise behind Moven is that it will control the digital front end of the customer relationship and partner with traditional banks to handle the deposits. Moven provides financial services through a mix of mobile and web modes such as banking applications enabled using smart phones.

The founding of a company like Moven is in response to banking customers living in a digital age. A recent study showed that it is not only millennials that are wanting personalized, anytime, anywhere services from their banking relationships. Cisco conducted a survey of over 7,000 individuals in 12 different countries and found that there was a gap between the consumer behavior and their current bank’s ability to offer personalized and convenient services and products (Cisco, 2014). To understand and respond to digital customer’s better, Moven is providing applications that offer real-time banking advice and incorporate social media data into the development of a credit score.

Moven recently developed a smartwatch application that gives a customer real-time banking information and budget advice. The new timepiece version of the app provides a stripped-down analysis of the wearer’s spending habits. For example, as a customer walks out of Starbucks, the application provides an alert telling how much was just spent and whether the customer is in good shape budget-wise or should start using the office Keurig. The

application’s spending meter turns red, green or yellow depending on how much the customer has spent overall in the current month versus the previous month (Crosman, 2015).

Moven is using social media profiles including Facebook, LinkedIn, and Twitter to determine personality traits about a consumer which might help to predict their ability and willingness to payback credit. Movenbank uses the personal information stored in social media sites such as Facebook to construct a one-on-one personal financial relationship with the customer. Online engagement methods such as messaging will also encourage customers to increase money in their savings accounts.

Moven has developed a product called CRED™ score, which combines traditional scoring elements and social media data. Moven seeks to determine a sort of digital credibility standing of a customer using their social media information. Taking into account an individual’s traditional credit score in addition to a customer’s use of digital payment channels, social connectivity and money management beliefs, CRED™ score assesses risk as well as a customer’s financial potential. Unlike a credit score, CRED™ score is designed to be a sort of financial health or wellness score, like a calorie counter on your phone – a score that goes up when a customer gets better at saving or managing their money (Moven, 2016). CRED™ score is trade-off for individuals who use it, where individuals have to be willing to share their social data to participate in building better financial health.

FUTURE DIRECTIONS

As we become more and more of a digital society through online banking and shopping, there are several areas of technology that will see drastic changes in the financial field.

- **Storage:** Hard drives, jump drives, and servers will soon be items of the past. New

emerging technology will be needed to collect, store, preserve, manage, and analyze unstructured streaming data that social media outlets produce.

- **Commerce:** In time, all credit agencies and financial organizations may use social media information as a component of a consumer's credit score. As more shopping moves online and physical money such as \$10 and \$20 bills are vastly becoming digital 1's and 0's, it will more important to calculate the financial risk of individuals.
- **Software Developments:** In the next few years, we will see the development of distributed computing technology, such as Hadoop, to perform serious scalable computing. New advances in big data analytics will be developed to deal with the flood of web data produced from clickstreams from a website with millions of users.

More credit data is available than ever before through unstructured data sources and social media outlets. The challenge for banks and other financial institutions is to quickly capture that data and make decisions. Similarly, the credit industry has never experienced such a growth of available data. The challenge will be to invent innovative credit scorecard models that can rapidly adapt to changing data in real-time.

CONCLUSION

Modern data processing systems have ensured that virtually all banks and financial institutions have rich data sources to mine. However, merely having volumes of data is no longer a source of competitive advantage. Data has become a commodity and therefore needs to be honed and enriched further to derive that extra advantage needed to differentiate results. Therefore, to form and sustain a differentiator strategy, the existing data has to be supplemented with unstructured

and social media data sources to give the overall information a whole new dimension.

The rapidly expanding social media arena provides a new source of information flow. Tapping into the collective intelligence and sentiments of the crowd through aggregated mining of their social posts provides clues to consumer thought. At a bank, the concept of the 360-degree view of the customers ascends to an entirely new level in the context of accessing their social media interactions.

REFERENCES

- Adams, J. (2012, May 22). Credit scoring models become more agile. *American Banker*, 177(79).
- Anderson, B., & Hardin, J. M. (2014). Credit scoring in the age of big data. In J. Wang (Ed.), *Encyclopedia of Business Analytics and Optimization* (pp. 549–557). Hershey, PA: Information Science Publishing. doi:10.4018/978-1-4666-5202-6.ch049
- Cisco Study: Digital Lifestyles Drive a Widening Gap Between Consumers' Expectations and Bank Delivery. (2014, November 14). M2 Communications. Retrieved from http://go.galegroup.com.ezproxy.ferris.edu/ps/i.do?id=GALE%7CA390044775&v=2.1&u=lom_ferrissu&it=r&p=ITOF&sw=w&asid=6ba91710790d792a30167022f9d1479d
- Crosman, P. (2015). Moven smartwatch app to test demand for real-time PFM advice. *American Banker*, 1(2), 1–2.
- DePaula, M. (2015). How to Reach a Retiree? Try YouTube. An Ally Bank marketing campaign eschews the traditional approach and finds an older audience in social media. *American Banker*, 124(5), 30.
- Durand, D. (1941). *Risk Elements in Consumer Installment Financing*. New York: National Bureau of Economic Research.

- Fei, W., Gu, J., Yang, Y., & Zhou, Z. (2015). Credit risk evaluation based on social media. *Procedia Computer Science*, 55, 725–731. doi:10.1016/j.procs.2015.07.165
- Fernandez, J. (2011, December 2). *The Official Klout Blog*. Retrieved July 16, 2012, from What Does Klout Measure?: <http://corp.klout.com/blog/2011/12/what-does-klout-measure/>
- Fisher, R. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7(2), 179–188. doi:10.1111/j.1469-1809.1936.tb02137.x
- Hand, D., & Henley, W. (1997). Statistical classification methods in consumer credit scoring: A review. *Journal of the Royal Statistical Society. Series A, (Statistics in Society)*, 160(3), 523–541. doi:10.1111/j.1467-985X.1997.00078.x
- Hardekopf, B. (2015, October 23). *Your Social Media Posts May Soon Affect Your Credit Score*. Retrieved June 27, 2016, from Forbes: <http://www.forbes.com/sites/moneybuilder/2015/10/23/your-social-media-posts-may-soon-affect-your-credit-score-2/#2f48f3a93207>
- King, B. (2012, May 6). *A Vision for 'Frictionless' Banking*. Retrieved July 6, 2012, from American Banker: http://www.americanbanker.com/btn/25_5/innovators-of-the-year-movenbank-brett-king-1048912-1.html
- King, J. (2014). IMF world bank: Credit scoring - friends, followers and settling scores. *The Banker*. Retrieved from <http://search.proquest.com/docview/1609313368?accountid=10825>
- Lohr, S. (2015, January 8). *Banking Start-Ups Adopt New Tools for Lending*. Retrieved June 26, 2016, from The New York Times: http://www.nytimes.com/2015/01/19/technology/banking-start-ups-adopt-new-tools-for-lending.html?_r=0
- Maney, K. (2016, April 22). Facebook Could Blow Up Credit Cards and Make Loans to Billions; The lending industry should be very afraid. Your friends are a great predictor of whether you'll pay back a loan. *Newsweek*, 166(15). Retrieved June 26, 2016, from http://go.galegroup.com.ezproxy.ferris.edu/ps/i.do?id=GALE%7CA449524469&v=2.1&u=lom_ferrissu&it=r&p=ITOF&sw=w&asid=ca849a5f6bd915c14da61ee23ea5df3f
- Moven: From mobile banking to mobile money. (2013, February 17). Retrieved from <http://thefinancialbrand.com/37169/moven-mobile-banking-mobile-money/>
- Selyukh, A. (2015, November 4). *Could your social media footprint step on your credit history* [Interview]. Retrieved from <http://www.npr.org/sections/thetwo-way/2015/11/04/454237651/could-your-social-media-footprint-step-on-your-credit-history>
- Siddiqi, N. (2006). *Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring*. Hoboken, NJ: John Wiley & Sons, Inc.
- Thomas, L., Oliver, R., & Hand, D. (2005). A survey of the issues in consumer credit modeling research. *The Journal of the Operational Research Society*, 56(9), 1006–1015. doi:10.1057/palgrave.jors.2602018
- Thomas, P. (2015). Data from social media has a place in credit decisions. *Credit Control*, 36(3), 64–68.
- Wagstaff, K. (2012, June 20). *Time Techland*. Retrieved July 16, 2012, from Klout: A good idea done wrong: <http://techland.time.com/2012/06/20/klout-a-good-idea-done-wrong/>
- Wei, Y., Yildirim, P., Van den Bulte, C., & Delarocas, C. (2015). Credit scoring with social network data. *Marketing Science*, 35(2), 234–258. doi:10.1287/mksc.2015.0949

ADDITIONAL READING

- Bigonha, C., Cardoso, T., Moro, M., Goncalves, M., & Almeida, V. (2011). Sentiment-based influence detection on Twitter. *Journal of the Brazilian Computer Society*, 1–15.
- Campo-Avila, J., Moreno-Vergara, N., & Trella-Lopez, M. (2011). Analyzing factors to increase the influence of a Twitter user. In J. Perez, J. Corchado, M. Moreno, V. Julian, P. Mathieu, & J. Canda-Bago et al. (Eds.), *Highlights in practical applications of agents and multiagent systems* (pp. 69–76). Berlin: Springer-Verlag. doi:10.1007/978-3-642-19917-2_9
- Karppi, T., & Crawford, K. (2016). Social media, financial algorithms and hack crash. *Theory, Culture & Society*, 133(1), 73–92. doi:10.1177/0263276415583139
- Kolaczyk, E. (2009). *Statistical analysis of network data*. New York: Springer. doi:10.1007/978-0-387-88146-1
- Laleh, N., & Azgomi, M. A. (2010). A hybrid scoring and spike detection technique in streaming data. *Intelligent Data Analysis*, 14(6), 773–800.
- Lin, W.-Y., Hu, Y.-H., & Tsai, C.-F. (2012). Machine learning in financial crisis prediction: A survey. *IEEE Transactions on Systems, Man and Cybernetics. Part C, Applications and Reviews*, 42(4), 421–436. doi:10.1109/TSMCC.2011.2170420
- Mantelero, A. (2016). Personal data for decisional purposes in the age of analytics: From an individual to a collective dimension of data protection. *Computer Law & Security Report*, 32(2), 238–255. doi:10.1016/j.clsr.2016.01.014
- Poston, L. (2012). *Social media metrics for dummies*. Hoboken: John Wiley & Sons, Inc.
- Russell, M. (2011). *Mining the social web: analyzing data from Facebook, Twitter, LinkedIn, and other social media sites*. Sebastopol: O'Reilly Media, Inc.
- Thomas, D., & Barlow, M. (2011). *The executive's guide to enterprise social media strategy: how social networks are radically transforming your business*. Hoboken: John Wiley & Sons, Inc.
- Ting, I.-H., & Wu, H.-J. (2009). *Web mining applications in e-commerce and e-services*. Berlin: Springer-Verlag. doi:10.1007/978-3-540-88081-3
- Ting, I.-H., Wu, H.-J., & Ho, T.-H. (2010). *Mining and analyzing social networks*. Berlin: Springer-Verlag. doi:10.1007/978-3-642-13422-7
- Wang, J. (2008). *Data warehousing and mining: concepts, tools, methodologies, and applications*. Hershey: Information Science Reference. doi:10.4018/978-1-59904-951-9
- Yap, B. W., Ong, S. H., & Husain, N. H. (2011). Using data mining to improve assessment of credit scoring worthiness via credit scoring models. *Expert Systems with Applications*, 38(10), 13274–13283. doi:10.1016/j.eswa.2011.04.147
- Zhang, M., Jansen, B., & Chowdhury, A. (2011). Business engagement on twitter: A path analysis. *Electronic Markets*, 21(3), 161–175. doi:10.1007/s12525-011-0065-z
- Zhou, X., Xu, Y., Li, Y., Josang, A., & Cox, C. (2012). The state-of-the-art in personalized recommender systems for social networking. *Artificial Intelligence Review*, 37(2), 119–132. doi:10.1007/s10462-011-9222-1
- Zhu, T., Wu, B., & Wang, B. (2009). Social influence and role analysis based on community structure in social network. In R. Huang, Q. Yang, J. Pei, J. Gama, X. Meng, & X. Li (Eds.), *Data mining and applications* (pp. 788–795). Berlin: Springer-Verlag. doi:10.1007/978-3-642-03348-3_84

KEY TERMS AND DEFINITIONS

Credit Score: A measure of risk associated with how likely an individual is to re-pay a financial loan.

Hadoop: Open source software that stores and analyzes massive unstructured data sets.

Klout Score: A Klout score is a numerical measure on the scale of 1 to 100 that measures how influential an individual is in an online social network setting. The higher the Klout score the more influential the individual is considered. For example, is a person influential in getting others to buy a product or sign up for a service.

Social Media: A web-based social outlet. Social media may be in the form of online web-sites that allow individuals to communicate and

establish social networks such as Facebook and web-based videos like those viewed on YouTube.

Social Media Credit Score: A measure of social risk that is created using information from social media outlets.

Structured Data Source: Data that contains an identifiable structure. Structured data is typically stored in traditional databases with identifiable rows and columns.

Tweet: A posting on the social media site Twitter that consists of 140 characters or less.

Unstructured Data Source: Data that has no defined identification such as e-mails, images, text and web logs, and videos. Unstructured data is data that is not contained in a database.

S

Social Network Analysis and the Study of University Industry Relations

Fernando Cabrita Romero
University of Minho, Portugal

INTRODUCTION

The methodological approach of Social Network Analysis (from now on, abbreviated SNA) can bring many benefits for the study of university and industry relations (from now on, abbreviated UIR). Relations between university and industry are important in innovation studies because they can be considered as an innovation network, in the sense that the interactions established by their participants have more or less defined innovation goals (Mansfield & Lee, 1996). SNA is the study of social structure (Wellman & Berkowitz, 1988). It describes a group of quantitative methods for analysing the ties among social entities and their implications (Wasserman & Faust, 2007). With the use of SNA it is possible to explore and to better understand different innovation outcomes involving UIR, by analysing the different measures and the structure of the social network. In this work the SNA methodological approach is described and its fundamental concepts are presented. The paper then reviews the applications of this approach on the study of the relationships between university and industry. The paper aims at systematizing the available information and knowledge, highlighting the main research pathways, the main conclusions and pointing possible future research questions.

BACKGROUND

UIR is an increasingly important and researched phenomena. Theoretically and structurally, the theme is linked to the study of the innovation process and to the need of understanding it and in-

fluencing it. Linear perspectives of the innovation process placed the university at the beginning of a linear sequence of innovation and the firm at the end of it, largely ignoring the interaction between the two types of institutions. New, interactive, dynamic, complex perspectives see the innovation process as a system involving many institutional actors, as a network of relations and as a complicated web of knowledge exchange and utilization. Science and technology are increasingly complex and costly and no single actor commands the necessary resources, relying on multiple sources of information and knowledge exchanges that are crucial for a successful innovation process. In this context, and considering the many types of relations and actors that may be possible in the innovation process, the relationships between academia and industry stand out as particularly relevant, because of the type of institutions that participate and the nature of the information and knowledge that is exchanged. The literature on UIR has raised many issues on the theme, and debates are ongoing, which can be found in some review articles (Baldwin & Green, 1984; D'Este & Patel, 2007; Perkmann et al., 2013). Only some of the issues and debates will be explored here, namely those researched by SNA concepts. They will be referred in more detail in the following section, along with the presentation of the results of the literature review, which is the main focus of this work.

SNA is the study of social structure using a group of quantitative methods. It analyses ties among social entities and looks for key players and group patterns. SNA uses concepts that are related to the structural properties of the network

and indicators that are related to relational properties of the network. The most used concepts related to structural properties of the network in UIR studies are the concepts of density, component, and subgroups. The most used social network analysis concepts related to relational properties of the network are the concepts of degree centrality, betweenness centrality and geodesic distance.

The concepts related to structural properties of the network are basic and important concepts that characterize the overall structure of the network. Through the concept of density (which indicates the level of connection between all the nodes of the network) it characterizes its global cohesion. Through the concept of component (which indicates the existence of large groups inside the network) and through the concepts of subgroups or cliques (which indicates the existence of smaller, cohesive, specifically defined groups) it characterizes its internal structure. The combination of these indicators and an adequate interpretation of their meaning provides useful descriptions and characterizations of the network, in terms of the position of their nodes and constituents. The characterization is frequently complemented with visual aids, namely through sociograms. Sociograms are diagrams composed of nodes and lines. The nodes are the actors of the network and the lines are the connections between the actors. The concepts related to relational properties of the network are often at the centre of the analytic procedure, and are used in several ways according to specific research objectives. The concept of degree centrality (which indicates the extent of a node connection) identifies the most connected actors. The concept of betweenness centrality (which captures the intermediary nature of a particular node's connection) is used to identify and characterize the intermediary positions of actors in the network. Besides the main concepts referred above, other concepts related to these ones are also used, but less often. All these concepts are mathematically defined, but it is beyond the scope of this paper to provide detailed algebraic specifications, which are available in several books (Scott, 2000; Wasserman & Faust, 2007).

THE LITERATURE REVIEW

General Aspects and a Synthesis of the Main Results

There is not a great number of articles that addresses specifically the problem of UIR using SNA techniques. There is a variety of perspectives that reflect specific and idiosyncratic concerns of the authors. Few papers follow the same guidelines or share identical perspectives. However, there are small groups of authors that build on past works or use identical databases, such as patent databases.

In terms of the main study object or main research preoccupation the studies can be classified in the following categories: 1) the study of the characteristics of personnel/institutional networks that are prominent in university-industry relations; these studies generally rely on the use of patents that are co-produced jointly by university and non-university members, and the patterns of collaboration are analysed; 2) the study of university-industry relations in the context of specific industrial settings or in the context of specific institutional conditions; these studies may rely also on patent databases but other types of data may be used, either primary data, obtained through questionnaires, or secondary data, obtained through diverse documental sources; 3) the contribution of the study of university-industry relations to the validation of theories; these studies also rely on a mix of patent, primary and secondary data.

In addition to these themes there are other themes that are addressed in these studies, either in a parallel way or as themes that frame the former or the research approach. The themes that could be common to the papers are, in broad terms, the three main themes above indicated but, within each one, the approach and main research concerns and targets are quite different. The literature will be analysed not through the lens of the broad themes, but through the details of the specific papers. This methodology will permit to extract from the papers the main academic debates and to highlight the respective contributions to

knowledge. The next sections will perform that task. Table 1 synthesises the results.

The table indicates the main identified concepts on UIR that have been researched using SNA. It is possible to divide those concepts in four groups. The first one includes the two first concepts, which are related to intrinsic structural properties of the network. The second group includes the two following concepts, which are related to institutional and knowledge characteristics of the network. The third group includes the next three concepts, which are related to specific patterns found in networks. The fourth group includes the last two concepts of the table, which are related to the validation of existing theories. The last concept

is related to the search for new methodological approaches, mainly related to the use of new data sets. The main results and the literature for each concept will be described and explored in the following sections.

Strong and Weak Ties, and Structural Holes

The influence of the nature of the relation on the performance of the network is a debated issue. The concepts of strong and weak ties were introduced by Granovetter (1973). Strong ties represents strong and regular interactions between the actors

Table 1. University-industry relations and social network analysis: main debates and conceptual propositions arising from the literature review

Main Concepts	References (Authors, Year)	Conceptual Propositions Proposed by the Literature
Strong and weak ties, structural holes	(Gilsing & Duysters, 2008; Rost, 2011; van der Valk, Chappin, & Gijssbers, 2011; Villanueva-Felez, Molas-Gallart, & Escriba-Esteve, 2013)	• Balanced social structures (strong ties with some weak ties) seem to be more innovative. • Differential outcomes on the nature of knowledge contingent on the specific balance of the structure of social capital.
“Small worlds” networks	(Balconi, Breschi, & Lissoni, 2004; Guan & Zhao, 2013; Protogerou, Caloghirou, & Siokas, 2013; van der Valk et al., 2011)	Networks with high clustering and short average geodesic paths are more conducive to inventive or innovative activity.
Open-science and proprietary technology	(Balconi et al., 2004; Jason Owen-Smith & Powell, 2004)	• The institutional attributes of open science and proprietary technology influences network structure. • Open science networks are more connected and dense than proprietary networks that are more fragmented and disperse.
Knowledge base and environment as relational factors	(Gilsing & Duysters, 2008; Krätke & Brandt, 2008; Leydesdorff, 2004; J. Owen-Smith, Riccaboni, Pammolli, & Powell, 2002; Plum & Hassink, 2011)	Different knowledge bases affect network structural properties, the position of individual entities in the network and their capacity to access knowledge
Patterns of university-industry relations	(Gilsing & Duysters, 2008; Krätke & Brandt, 2008; Leydesdorff, 2004)	• Patterns are influenced by regional industrial structures. • Biotechnology has a specific pattern of university-industry interaction.
Public research organizations as central actors in innovation networks	(Breschi & Catalini, 2010; De Stefano & Zaccarin, 2013; Lissoni, 2010; Minguillo & Thelwall, 2012; Protogerou et al., 2013)	• Academic authors-inventors assume more brokerage positions. • Public research organization are at the centre of innovation programmes.
Influence of commercial orientation on fundamental science production	(Balconi & Laboranti, 2006)	Academics more connected to industry are more productive in scientific terms.
Industrial districts	(Capo-Vicedo, Molina-Morales, & Capo, 2013; Morrison, 2008)	• Public research organization as main intermediaries of knowledge flows to the district. • Weak knowledge exchanges but strong information exchanges inside the district actors.
Triple-helix theory	(Heimeriks, Hörlesberger, & Van Den Besselaar, 2003; Khan & Park, 2013)	• Triple helix assumptions on institutional role intersections are supported. • Multiple communication channels with differential roles in the Triple Helix relation.
Methodological contributions	(Heimeriks et al., 2003; Kim, 2012; Minguillo & Thelwall, 2012)	Asides from patents indicators, other indicators and data unmask fundamental structural or relational properties.

(nodes) of the network and weak ties represent sporadic and temporary interactions.

Coleman (1988) claimed that cohesive groups and strong ties were effective ways to coordinate an exchange of knowledge flows, while Burt (1992) argued that strong ties resulted in redundant information and that innovation required new knowledge inflows and perspectives coming from weak ties. Unfulfilled connections between groups with internal strong ties are said to be structural holes, which can be filled by specific actors, with strong intermediary positions. Villanueva-Felez et al. (2013) apply these concepts to assess in which way the structure of researchers' social capital affects academic performance. The authors distinguish between academics that are completely embedded in a network that has no weak ties (establishing links with members of his or her own department, without ties with government, industrial, or other societal actors), academics which are in a network that is formed predominantly by weak ties, and academics that are in an integrated network that contains both strong and weak ties. The results show that the academics in the network with no weak ties are the less productive. On a study of a network of inventors and on the assessment of the impact of patents (based on forward citations) and integration of knowledge (based on backward citations), Rost (2011) concludes that inventors with balanced social capital (strong ties but also some weak ties) come up with the most innovative solutions, or integrate the most knowledge or have the highest impact on future knowledge. He concludes that Coleman's and Burt's perspectives are complementary and that in the presence of strong ties, weak network structures (structural holes or peripheral position) leverage the strength of strong ties in the creation of innovation. Similar arguments are advanced in a visual network analysis of two government sponsored programmes that aimed to foster innovation through public-private partnerships (van der Valk et al., 2011) and also by other studies of university-industry relations or industry networks (Ahuja, 2000; Gilsing & Duysters, 2008).

"Small Worlds" Networks

The open science characteristics of scientific communities translate, in social network terms, into the so-called networks with "small worlds" characteristics (Albert & Barabási, 2002). The small world properties, in the context of scientific networks in a specific discipline, are defined by the existence of a large component connecting almost all nodes and within the large component, all nodes (scientists) are close to each other (Albert & Barabási, 2002; Newman, 2001). These characteristics of academic networks are not found in networks of non-academic inventors, except in since-based fields. The influence of small world properties on innovativeness is addressed in studies of industry networks (Verspagen & Duysters, 2004) and in university-industry networks (Guan & Zhao, 2013), and generally considered to be positive, although there are disagreements concerning this positive influence (Fleming, King Iii, & Juda, 2007).

Open Science and Proprietary Technology

The analyses of patent databases provides the basis for the exploration of another important concept, which is debated in multiple forms and in its multiple consequences in studies of university-industry relations, which is the distinction between the characteristics of open science and proprietary technology (Cowan & Jonard, 2003; Merton, 1957). The debate can be inserted in a larger debate concerning the influence of diverse institutional conditions on processes of relations between organizational entities. Balconi et al. (2004) conduct a study of Italian academic and industrial inventors whereby, departing from assumptions on the behaviour or characteristics of "open science" networks and "proprietary networks", expect to find differences between the networks of academic and non-academic inventor. In fact, the study found that networks of industrial inventors are much more fragmented

than networks of academic inventors, except in the chemistry field (defined in a broad sense, i.e. including biotechnology). The chemistry sector, science-based field, was different because it was influenced by the institutional weight of scientific inputs in commercial technology.

Knowledge Based and Environmental Conditions as Relational Factors

Other articles support the importance of environmental factors in shaping specific properties of networks. A study of the Boston biotechnology sector (Jason Owen-Smith & Powell, 2004) found that the information flows between the actors of the network, which included firms and public research organizations, depended not only on network participation and geographic proximity, but also on the institutional characteristics of the network. In public-dominated networks firm performance depended only on net participation, unlike in networks dominated by private entities, where innovative performance depended on position factors, i.e., their closeness to central actors (although this characteristic was weak in terms of statistical significance).

An important determinant of cooperation between university and industry, and an important factor in terms of innovative performance, seems to be related to the position of the firm in the network. That position may be related to geography (Balconi & Laboranti, 2006) or that position may be related to the knowledge base that the firm possesses and that may confer the firm the possibility to connect with more or less central actors of the network. A study of an industrial network in Germany (Cantner & Graf, 2006) argues that a prerequisite for future cooperation is not based on past cooperation but rather on a shared knowledge base. It puts in question ideas that argue that persistent cooperation, based on trust, is necessarily the basis for collaboration. In

a study of two industrial networks (biotechnology and multimedia) in a period that was characterised by the transition of an existing dominant design and a shift away from rules, norms, routines or activities, Gilsing and Duysters (2008) argue that environmental conditions related to different knowledge bases, and the validation and selection mechanisms inherent to each of the two fields, explain the relational and structural properties of the two networks. For instance, the connection of public research organizations is centrally present in the biotechnology field but absent in the multimedia field (Gilsing & Duysters, 2008).

Differences in the knowledge base show up as an important factor in the determination of collaboration structures in another study involving biotechnology firms in a regional context (Plum & Hassink, 2011). It points to differences related to internal competencies of the firms regarding differential capabilities in terms of the nature of the knowledge required to develop the differential products of each firm, in which the knowledge of the market also has a role.

Although in a quite different perspective, a study of the differences between the structures of two networks emphasises the importance of environment in shaping the properties of the network (Capellari & De Stefano, 2014). Patents that are owned by the university (which is the assignee) or invented by the university (the university is not the assignee but at least one of the inventors is a tenured academic), are analysed separately, showing differences in terms of size of components, number and size of subgroups and the brokerage position of inventors. The institutional factors are mediated by two universities that have different policies related to patenting ownership.

Patterns of University-Industry Relations

Databases on scientific literature have been extensively used to analyse the patterns of collaboration

between scientists. Patent databases are also being explored to analyse the patterns of collaboration between academia and industry.

An exploratory analysis of the simultaneous embeddedness of researchers in scientific and technological networks (Breschi & Catalini, 2010), which compares networks of authors, inventors and authors-inventors, and the overlap between them, argues that author-inventors play a crucial role in connecting the two other networks (only authors and only inventors) and occupy important positions in each community. The role of academics as fundamental intermediaries between public and private research is explored in a study (Lissoni, 2010) that finds that academic inventors tend to be more central actors in broker and gatekeeping positions, although strong brokerage positions are very few and held by scientist with many patents and publications. De Stefano and Zaccarin (2013) reach similar conclusions regarding the larger relational activity of academic authors-inventors vis-a-vis industrial authors-inventors.

Two important differences were also apparent in Balconi et al. (2004): academic inventors were more connected than non-academic inventors, and had a more central position. The central position of academics, or of the university, is a characteristic that often shows up in analysis of networks where public research organisations are involved (Balconi & Laboranti, 2006; Breschi & Catalini, 2010; Jason Owen-Smith & Powell, 2004; Protogerou et al., 2013).

The knowledge base of patents is researched to see how much innovation is really based on science (Leydesdorff, 2004). This question is important because theories about university-industry relations are historically influenced by the biotechnology sector. The biotechnology is a science-based sector whose inventive activities tend to be performed in close collaboration with public research organizations and whose output is patented through co-authorships or co-assignments between academic and industrial inventors. The access and the analysis of patents

databases have become easier and many studies have thus relied on these data to infer general conclusions to other fields of science, that are not so formalized in terms of literature relations. The study analysis two sets of patents, extracted from the USPTO, one based on patents that have a university as a co-assignee, and another that has a Dutch address as an assignee. The structure of the co-words networks linking patents and their citations to other patents and scientific literature is analysed. The analysis is entirely based on the visualization of sociograms, while nodes are (co) words). The two networks are quite different. In the set of university patents (which represents university-industry relations) the fields of biotechnology and molecular biology dominate the set and the knowledge base of the patents, and the visualising shows a neat organization around the intellectual organization of the disciplines. In the set of Dutch patents (representing the knowledge base of the international economy) the visualization shows a recognizable representation of the Dutch industrial structure with a dominance of electro-technical and chemical applications and large multinational corporations. Although biomedical application integrates the patents they are not central to the whole set. These results strongly suggest that inferences of university-industry relations based on literature and patent analyses are heavily conditioned by the specificity of the biotechnology sector.

Public Research Organizations as Central Actors in Innovation Networks

The central position of public research organizations shows up in descriptive analyses of networks that involve heterogeneous actors. Both a study of the network structure of science parks (Minguillo & Thelwall, 2012), using web links as indicators of connections, and a study of the collaborative networks established during the seven Framework Programme on Research and Technological De-

velopment of European Commission, show the central position of public research organizations. In the study of science parks, governmental agencies also play an important role, and in the case of the Framework Programmes, although firms are present in larger numbers, they are not the central actors.

Influence of Commercial Orientation on Fundamental Science

The impact on fundamental research of an orientation to patenting and commercialization has been researched through the relationship between patenting activity and publication record of university researchers, and in general the results point to a positive correlation between patenting and publication activity (Czarnitzki, Glänzel, & Hussinger, 2009). This theme is revived with a social network approach (Balconi & Laboranti, 2006) and the results support the positive relationship between publication record and patenting activity. The author argues, in line with other similar arguments (Rosenberg & Nelson, 1994), that industry feeds on academic research but that academic research also needs inputs from high technology industries in order to find direction to its research. So, academics that are close and collaborate with industry producing patents are also the ones that are more productive in purely scientific terms.

Industrial Districts

There is a strand of research of university-industry relations using social network analysis methods that adopt a deductive approach and try to validate some relatively entrenched conceptual implications of some theories.

One of the researched theories looks at the implications of the industrial district approach. Morrison (2008), in her study of the furniture sector in Italy, argues that the community of informal ties appears to be rather small and that

'know how' sharing is also rather limited, contrary to assumptions from industrial district theorists that based their ideas on the development of these concentrated regions on intense knowledge exchange between the actors. It, however, supports the argument that public research organizations, more than large firms, play a central role and as intermediaries in the knowledge flows for innovation that occur in the industrial district, and that knowledge for innovation does not arise only from the close interactions of the firms of the district, an idea that is also supported by a study of a Spanish textile industrial district (Capo-Vicedo et al., 2013).

Triple-Helix Theory

The implications of the triple-helix approach are also examined. Using webometric indicators and semantic analysis of the contents of the webpages Kim (2012) found that university and industry websites were similar, thus suggesting there is an intersection or interchangeability of the roles and function of the two types of organizations, as suggested by the triple-helix theory (Etzkowitz & Leydesdorff, 1998). Diverse channels of communication and relations between the diverse institutional actors (co-authorship, participation in projects, information diffusion) is also explored in Heimeriks et al. (2003) which argues that each communication channel or media has different functional purposes in maintaining in the maintenance of the links of the triple-helix relation.

Methodological Contributions

Finally, there is a search for alternative methodological approaches and indicators in the studies of networks of university-industry relations. Some authors propose the use of webometric approaches (Kim, 2012; Minguillo & Thelwall, 2012) and other authors propose the use of simultaneous indicators of relational characteristics, such as citations, project participation, questionnaires or

other data (Almodovar & Teixeira, 2014; Furu-kawa, Shirakawa, & Okuwada, 2011; Heimeriks et al., 2003), arguing that analysis based on a single indicator underestimate the level and may not capture all of the complexities of the col-laboration patterns.

FUTURE RESEARCH DIRECTIONS

Some possible research paths are open. Eventually, the use of more complex and elaborated concepts of network analysis could improve the analysis of data, it may have the potential to reach different or stronger evidence and conclusions, and it may be an aspect that must be improved. The diversity and plurality of university-industry relations has not been properly addressed in the literature, which tends to use patents as indicators of collaboration. New sources of data must be explored. Environmental and institutional influences of diverse sorts are clearly very important factors that condition and determine university-industry relations, and research is open to greater exploratory efforts. There is a considerable potential to test theoretical and conceptual propositions which are assumed but have scarce empirical support.

CONCLUSION

The use of social network analysis in the study of university-industry relations was reviewed in this study. There are not many studies that combine the two perspectives and the ones that exist fol-low different research objectives and concerns and different methodological proposals. It seems evident that this particular knowledge quest is in a highly exploratory phase. Nevertheless, the contributions to knowledge have been varied and important, ranging from purely descriptive stud-ies and methodological explorations to deductive testing of established theories.

REFERENCES

- Ahuja, G. (2000). Collaboration networks, struc-tural holes, and innovation: A longitudinal study. *Administrative Science Quarterly*, 45(3), 425–455. doi:10.2307/2667105
- Albert, R., & Barabási, A.-L. (2002). Statisti-cal mechanics of complex networks. *Reviews of Modern Physics*, 74(1), 47–97. doi:10.1103/RevModPhys.74.47
- Almodovar, J., & Teixeira, A. A. C. (2014). As-sessing the Importance of Local Supporting Or-ganizations in the Automotive Industry: A Hybrid Dynamic Framework of Innovation Networks. *European Planning Studies*, 22(4), 841–865. doi:10.1080/09654313.2013.771621
- Balconi, M., Breschi, S., & Lissoni, F. (2004). Networks of inventors and the role of academia: an exploration of Italian patent data. *Research Policy*, 33(1), 127–145.
- Balconi, M., & Laboranti, A. (2006). University–industry interactions in applied research: The case of microelectronics. *Research Policy*, 35(10), 1616–1630. doi:10.1016/j.respol.2006.09.018
- Baldwin, D. R., & Green, J. W. (1984). University-Industry Relations - a Review of the Literature. *Sra-Journal of the Society of Research Adminis-trators*, 15(4), 5–17.
- Breschi, S., & Catalini, C. (2010). Tracing the links between science and technology: An exploratory analysis of scientists and inventors networks. *Research Policy*, 39(1), 14–26. doi:10.1016/j.respol.2009.11.004
- Burt, R. S. (1992). *Structural Holes*. Cambridge, MA: Harvard University Press.
- Cantner, U., & Graf, H. (2006). The network of innovators in Jena: An application of social net-work analysis. *Research Policy*, 35(4), 463–480. doi:10.1016/j.respol.2006.01.002

- Capellari, S., & De Stefano, D. (2014). University-owned and university-invented patents: A network analysis on two Italian universities. *Scientometrics*, 99(2), 313–329. doi:10.1007/s11192-013-1211-5
- Capo-Vicedo, J., Molina-Morales, F. X., & Capo, J. (2013). The role of universities in making industrial districts more dynamic. A case study in Spain. *Higher Education*, 65(4), 417–435. doi:10.1007/s10734-012-9553-0
- Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology*, 94(Supplement), S95–S120. doi:10.1086/228943
- Cowan, R., & Jonard, N. (2003). The dynamics of collective invention. *Journal of Economic Behavior & Organization*, 52(4), 513–532. doi:10.1016/S0167-2681(03)00091-X
- Czarnitzki, D., Glänzel, W., & Hussinger, K. (2009). Heterogeneity of patenting activity and its implications for scientific research. *Research Policy*, 38(1), 26–34. doi:10.1016/j.respol.2008.10.001
- De Stefano, D., & Zaccarin, S. (2013). Modelling Multiple Interactions in Science and Technology Networks. *Industry and Innovation*, 20(3), 221–240. doi:10.1080/13662716.2013.791130
- DEste, P., & Patel, P. (2007). University-industry linkages in the UK: What are the factors underlying the variety of interactions with industry? *Research Policy*, 36(9), 1295–1313. doi:10.1016/j.respol.2007.05.002
- Etzkowitz, H., & Leydesdorff, L. (1998). The endless transition: A triple helix of university-industry-government relations. *Minerva*, 36(3), 203–208. doi:10.1023/A:1004348123030
- Fleming, L., King Iii, C., & Juda, A. I. (2007). *Small Worlds and Regional Innovation* (Vol. 18). Academic Press.
- Furukawa, T., Shirakawa, N., & Okuwada, K. (2011). Quantitative analysis of collaborative and mobility networks. *Scientometrics*, 87(3), 451–466. doi:10.1007/s11192-011-0360-7
- Gilsing, V. A., & Duysters, G. M. (2008). Understanding novelty creation in exploration networks—Structural and relational embeddedness jointly considered. *Technovation*, 28(10), 693–708. doi:10.1016/j.technovation.2008.03.004
- Granovetter, M. (1973). The Strenght of Weak Ties. *American Journal of Sociology*, 78(6), 1360–1380. doi:10.1086/225469
- Guan, J. C., & Zhao, Q. J. (2013). The impact of university-industry collaboration networks on innovation in nanobiopharmaceuticals. *Technological Forecasting and Social Change*, 80(7), 1271–1286. doi:10.1016/j.techfore.2012.11.013
- Heimeriks, G., Hörlesberger, M., & Van Den Besselaar, P. (2003). Mapping communication and collaboration in heterogeneous research networks. *Scientometrics*, 58(2), 391–413. doi:10.1023/A:1026296812830
- Khan, G. F., & Park, H. W. (2013). The e-government research domain: A triple helix network analysis of collaboration at the regional, country, and institutional levels. *Government Information Quarterly*, 30(2), 182–193. doi:10.1016/j.giq.2012.09.003
- Kim, J. H. (2012). A Hyperlink and Semantic Network Analysis of the Triple Helix (University-Government-Industry): The Interorganizational Communication Structure of Nanotechnology. *Journal of Computer-Mediated Communication*, 17(2), 152–170. doi:10.1111/j.1083-6101.2011.01564.x
- Krätke, S., & Brandt, A. (2008). Knowledge Networks as a Regional Development Resource: A Network Analysis of the Interlinks between Scientific Institutions and Regional Firms in the Metropolitan region of Hanover, Germany. *European Planning Studies*, 17(1), 43–63. doi:10.1080/09654310802513930

- Leydesdorff, L. (2004). The University-Industry Knowledge Relationship: Analyzing Patents and the Science Base of Technologies. *Journal of the American Society for Information Science and Technology*, 55(11), 991–1001. doi:10.1002/asi.20045
- Lissoni, F. (2010). Academic inventors as brokers. *Research Policy*, 39(7), 843–857. doi:10.1016/j.respol.2010.04.005
- Mansfield, E., & Lee, J. (1996). The modern university: Contributor to industrial innovation and recipient of industrial R&D support. *Research Policy*, 25(7), 1057–1058. doi:10.1016/S0048-7333(96)00893-1
- Merton, R. K. (1957). Priorities in Scientific Discovery: A Chapter in the Sociology of Science. *American Sociological Review*, 22(6), 635–659. doi:10.2307/2089193
- Minguillo, D., & Thelwall, M. (2012). Mapping the network structure of science parks An exploratory study of cross-sectoral interactions reflected on the web. *Aslib Proceedings*, 64(4), 332–357. doi:10.1108/00012531211244716
- Morrison, A. (2008). Gatekeepers of Knowledge within Industrial Districts: Who They Are, How They Interact. *Regional Studies*, 42(6), 817–835. doi:10.1080/00343400701654178
- Newman, M. E. J. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences of the United States of America*, 98(2), 404–409. doi:10.1073/pnas.98.2.404 PMID:11149952
- Owen-Smith, J., & Powell, W. W. (2004). Knowledge Networks as Channels and Conduits: The Effects of Spillovers in the Boston Biotechnology Community. *Organization Science*, 15(1), 5–21. doi:10.1287/orsc.1030.0054
- Owen-Smith, J., Riccaboni, M., Pammolli, F., & Powell, W. W. (2002). A comparison of US and European university-industry relations in the life sciences. *Management Science*, 48(1), 24–43. doi:10.1287/mnsc.48.1.24.14275
- Perkmann, M., Tartari, V., McKelvey, M., Autio, E., Broström, A., DEste, P., & Sobrero, M. et al. (2013). Academic engagement and commercialisation: A review of the literature on university–industry relations. *Research Policy*, 42(2), 423–442. doi:10.1016/j.respol.2012.09.007
- Plum, O., & Hassink, R. (2011). On the Nature and Geography of Innovation and Interactive Learning: A Case Study of the Biotechnology Industry in the Aachen Technology Region, Germany. *European Planning Studies*, 19(7), 1141–1163. doi:10.1080/09654313.2011.573128
- Protogerou, A., Caloghirou, Y., & Siokas, E. (2013). Twenty-five years of science-industry collaboration: The emergence and evolution of policy-driven research networks across Europe. *The Journal of Technology Transfer*, 38(6), 873–895. doi:10.1007/s10961-012-9278-3
- Rosenberg, N., & Nelson, R. R. (1994). American universities and technical advance in industry. *Research Policy*, 23(3), 323–348. doi:10.1016/0048-7333(94)90042-6
- Rost, K. (2011). The strength of strong ties in the creation of innovation. *Research Policy*, 40(4), 588–604. doi:10.1016/j.respol.2010.12.001
- Scott, J. (2000). *Social network analysis: a handbook*. London: Sage Publications.
- van der Valk, T., Chappin, M. M. H., & Gijsbers, G. W. (2011). Evaluating innovation networks in emerging technologies. *Technological Forecasting and Social Change*, 78(1), 25–39. doi:10.1016/j.techfore.2010.07.001
- Verspagen, B., & Duysters, G. (2004). The small worlds of strategic technology alliances. *Technovation*, 24(7), 563–571. doi:10.1016/S0166-4972(02)00123-2

Villanueva-Felez, A., Molas-Gallart, J., & Escibar-Esteve, A. (2013). Measuring Personal Networks and Their Relationship with Scientific Production. *Minerva*, 51(4), 465–483. doi:10.1007/s11024-013-9239-5

Wasserman, S., & Faust, K. (2007). *Social network analysis: Methods and applications*. Cambridge, UK: Cambridge University Press.

Wellman, B., & Berkowitz, S. (1988). *Social structures: A network approach*. Cambridge, UK: Cambridge University Press.

KEY TERMS AND DEFINITIONS

Betweenness Centrality: A social network analysis measure that indicates how much a node is in the middle of the connections between other nodes.

Degree Centrality: A social network analysis measure that indicates the number of other nodes to which the node is connected.

Density: A social network analysis measure that describes the level of linkages between nodes in a network. The more nodes are connected to each other, the denser the network is.

Innovation Process: Is a complex social and technical process that transforms ideas and

technologies into new or improved products or services.

Open Science: Open science is generally, but not exclusively, performed in university settings and is characterized by the wide non-commercial dissemination of research results and scientific knowledge.

Proprietary Technology: Is characterized by the appropriation by private entities of specific claims on technology, generally, but not exclusively, through the legal mechanism of patenting.

Social Network Analysis: A methodological approach that employs quantitative techniques to analyse social structures.

Strong and Weak Ties: A strong tie represents a person with whom there is a regular interaction, and a weak tie represents a person with whom there are sporadic or punctual contacts.

Structural Holes: The connection potential between elements or groups of elements that are not connected.

University-Industry Relations: A set of connections between people in university and people in industry. There are many forms of relations, including informal ones (the flow of graduates to industry, mobility of researchers, public meetings, professional networks) and formal ones (research contracts, licensing, joint labs).

Social Networking and Knowledge Sharing in Organizations

S

Sarabjot Kaur*Indian Institute of Technology Kanpur, India***Subhas Chandra Misra***Indian Institute of Technology Kanpur, India*

INTRODUCTION

Employees working in an organization have both explicit and tacit knowledge. Explicit knowledge is easier to document and share than tacit knowledge which stems from practical and experiential understanding. Solutions to contemporary business problems can be achieved by utilizing the available expertise spanning across different functional areas of an organization (Wenger & Snyder, 2000). Staying connected helps in advancing the understanding about various functional domains and helps in generating a multi-perspective outlook for tackling any issue. One such way of staying connected can be achieved by using social networking tools within the organizational boundaries. Yammer, Salesforce Chatter, Microsoft Sharepoint online are a few examples of collaborative and social tools used in enterprises. Enterprise wide social networking can help in connecting people from different backgrounds and domains making it easier to transfer knowledge to wider audiences (Riemer & Scifleet, 2012). The present chapter draws from extant literature to highlight the usefulness of social networking for knowledge sharing purposes along with discussing the relevant factors and adoption issues.

BACKGROUND

In its simplest form, social networking can be described as the use of internet technology to

create and share user generated content by a web of people connected with each other online. Sophia van Zyl (2009) defines social networking as “applications or websites that support the maintenance of personal relationships, the discovery of potential relationships and should aid in the conversion of potential ties into weak and strong ties, by utilizing emergent Web 2.0 technologies.” Social networking allows people to make connections, join groups of their interest, contribute information and discuss issues of relevance. All this is accomplished through various modes such as online communities, blogs, discussion forums and other online collaborative tools.

Enterprise Social Media

An emerging trend in the use of online collaborative tools in organizational context is Enterprise Social Media (ESM). Leonardi et al. (2013) define enterprise social media (ESM) as “web-based platforms that allow workers to communicate messages with specific coworkers or broadcast messages to everyone in the organization; explicitly indicate or implicitly reveal particular coworkers as communication partners; post, edit, sort text and files linked to themselves or others; view the messages, connections, text, and files communicated, posted, edited and sorted by anyone else in the organization at any time of their choosing.” ESM is an effective platform for connecting people across functional domains and making visible the information shared with a wider

audience. Similar to popular social platforms like Facebook and LinkedIn, ESM platforms may vary in the functionality offered and features supported but primarily they serve similar inherent purpose of bringing people together facilitating information flow and knowledge conversations, crowd source solutions across multiple domains, share relevant expertise and foster collaboration (Riemer & Scifleet, 2012).

SOCIAL NETWORKING AND KNOWLEDGE SHARING

Nature of Knowledge

Every organization wants its employees to share their knowledge with their co-workers as it will eventually benefit the organization in terms of improved performance. Knowledge exists in explicit and tacit forms. While explicit knowledge is easier to communicate and acquire, tacit knowledge is acquired by way of practice and encompasses skills and experiences held by people. It is difficult to document and explicate tacit knowledge as it is embedded in action and involves commitment (Nonaka, 1994). Such knowledge is helpful when there is a specific problem to be solved. The knowledge seeker approaches the potential knowledge provider with the problem and a solution can be reached by consensus after considering the circumstantial issues and making appropriate judgments. Creative solutions can be achieved by mutual knowledge sharing leading to enhanced creative performance when employees search information, share and co-create more information by way of interactions and discussions thereby adding to existing knowledge base of discussants (Sigala & Chalkati, 2015).

Interactions Over Social Media Platforms

Social media platforms can be instrumental in facilitating the above said interactions where it

can occur in the form of immediate unplanned instances or in events of specific information needs (Subramaniam & Nandhakumar, 2013). Tracking the activity of individual users can give a fair idea about their interest areas and relevant domain expertise (Kane, 2015). It leads to an overall understanding about the tacit knowledge resources possessed by various employees, a phenomenon termed as “ambient awareness” (Leonardi & Meyer, 2015).

Social networking tools enable collective completion of task with virtually co-present actors (Subramaniam & Nandhakumar, 2013). Problems can be viewed from different perspectives and social networking tools can be instrumental in crowdsourcing solutions by integrating cross functional expertise (Mäntymäki & Riemer, 2014). Thus, collective intelligence enabled by inputs from various sources across organization has significant potential of improving work performance leading to survival and growth in contemporary competitive environment. Although technology is adopted with intentions of functional improvement and advancement of organizational performance, its actual usage is imperative to understand its effects on organizational variables. Some of the factors that have an influence over knowledge sharing are listed below.

Social Ties

Social networking tools may help in fostering and maintaining connections that can be used to facilitate such discussions subsequently leading to consensual viewpoint on the issue. Strong social ties enhance the knowledge sharing behavior among employees (Chow & Chan, 2008). Dyadic knowledge exchange is enhanced in case of the virtual communities supported by social networking as it increases the possibility of frequent communication and increased flow of information among the dyads (Pan et al., 2015). Hence, better mechanism can be in place to increase the social interaction in both online and offline modes (Chen & Hung, 2010) along with shared goals (Chow &

Chan, 2008) to foster better understanding among the group.

Shared Goals

Shared goals have an indirect effect on knowledge sharing intention among coworkers. Also, a sense of being a part of the group instills a responsibility towards knowledge contribution in the group (Chang & Chuang, 2011). An environment of equality among individuals can also help in facilitating knowledge sharing (Yu et al., 2010). Working towards a collective goal can help to generate knowledge contributions from all members of a particular community.

Social Trust

Social trust influences knowledge sharing intentions (Chow & Chan, 2008; Din & Haron, 2012). Intensive interactions and mutual trust leads to more reliable knowledge sharing (Chang & Chuang, 2011). Strong interpersonal trust helps in reposing faith among knowledge seeker and provider thereby easing the transfer. However, fostering trust online is not always possible and the same can be accomplished by way of bringing people together in offline situations too (Chen & Hung, 2010).

Reciprocity

People want to share their knowledge as they feel that they need to reciprocate the help they received earlier (Yu et al, 2010; Chen & Hung, 2010). Thus, a knowledge contribution initiated by one member can lead to a chain of reciprocal responses adding to the cumulative knowledge base of its members.

Self-Efficacy

Individual self-efficacy will lead to higher contribution towards knowledge sharing activities (Hsu

et al., 2007). Relative advantage of the knowledge sharing exercise, the belief in one's own abilities (Hsu et. al., 2007) and mutual understanding among the knowledge partners helps in facilitating the sharing process (Chen & Hung, 2010).

Relational Context

Mechanisms to support and foster knowledge sharing vary according to relational context (Boer et al., 2011). The congruency between the relational assumptions of knowledge contributors has an effect on the sharing behavior. Similarity in assumptions about knowledge and the respective knowledge communities facilitates knowledge transfer.

Sharing Culture

The overall culture of the organization plays an important role in knowledge sharing activity. Highly open members view such platforms as the opportunity to discuss ideas, have opinions and debate issues along with coming up with potential solutions to problems with the help of mutual knowledge contributions (Yu et al, 2010). Social networking can be even more effective in an environment conducive for utilizing it to its best potential. Cultural norms affect the employee mindset and subsequently their need to seek collective expertise (Patroni et al., 2015). Employees may explore an area in greater depth due to both competitive and collaborative intent which eventually expedites their learning (Patroni et al., 2015).

Motivation

People are motivated by extrinsic and intrinsic rewards in respect of sharing their expertise or domain knowledge. Sharing knowledge over social platforms leads to an increase in the reputation of the knowledge provider which serves as an incentive and further increases both the quality and the quantity of the content shared (Hung et al., 2011).

Apart from this, intrinsic factors like altruism also affect the knowledge sharing behavior positively (Ma & Chan, 2014).

Similarity

Lingual similarity can enable better exchange and understanding which leads to subsequent effective exchange of information (Chang & Chuang, 2011). Similarity of status and location of an individual in the hierarchy may also have an effect on the knowledge sharing behavior. As the individual experience increases, members tend to share more with the people of similar expertise than with similar status (Hwang et al., 2015).

Context

Social networking acting as a medium of knowledge exchange eventually effects employee performance. However, the effect on performance differs according to task context (Kuegler et al., 2015). Performance varies in inter-team and intra-team conditions. Inter-team task context requires the contribution of novel ideas and new information suitable for innovative performance in contrast to intra-team usage where routine task specific information is required (Kuegler et al., 2015).

Excess Information

Too much information on social networking platform can even create clutter and make things worse as it may become difficult to find the relevant information. Different conversation threads may be difficult to follow. There can be attention conflict which may undermine the purpose of generating an overall awareness about knowledge conversations in the organization (Leonardi, 2015).

ADOPTION ISSUES

Any new technology should be viewed from multiple perspectives before introducing it in

the organization. Thus, actors at organizational, individual and technological level are considered for successful adoption and post-adoption usage of social networking tools. Organizational support in terms of user awareness and training is required for increasing the self-efficacy of users. Moreover, motivational support in terms of recognition and reputation feedback (Hung et al., 2011) can lead to more positive contribution by users. Purpose of social networking tools in organization, participants, and adequate risk analysis along with policy mechanisms should be considered before taking any decision to adopt such a system (Turban et al., 2011). Individual acceptability can be increased by ensuring the ease of their functional operation. Individual tendencies like self construal and interdependence (Liu & Rau, 2014) should be considered in adoption decisions. Also, individual usage is increased to as much as 25% when peers are using a given social media technology (McKinsey, 2015). Relative ease of use will render the tools accessible thereby increasing the possibilities of their continued usage. Existing usage of social networking tools in popular context can be examined to design tools that allow for added functionality as per organizational requirements. 35% of companies from a data set of 1500 were reported to adopt ESM platforms following the adoption by competitors (McKinsey, 2015). Irrespective of the reasons of adoption, it is the prerogative of company to decide about the extent of identity to be revealed. More identity leads to accountability while anonymity may help to raise genuine concerns and provide way for unique solutions (Kane, 2015).

FUTURE RESEARCH DIRECTIONS

Although social networking is a widely researched area, there are many aspects that can be taken up for future research in the context of its usage in organizations. Social networking is having an impact on the way people connect and share with each other. However, there can be instances

of misunderstandings due to poorly represented information which can be explored in respect of collaborative work. In addition, there is a need to explore these tools in light of design constraints. Often these tools are third party applications and may not be fully in sync with organizational or group requirements. Having a user oriented perspective in this regard can advance the design research in finding better ways to adapt them to organizational peculiarities which may be quite distinct from general and popular designs. Working in groups fosters collaborative thinking but excessive interaction over social networks may be a source of proliferating groupthink among employee which needs to be researched further. People tend to avoid any controversial and political issues in organizational communication. Hence, sharing in case of sensitive or political information is also another potential area worth exploration which may lead to interesting results.

CONCLUSION

Contemporary organizations are primarily knowledge driven. Knowledge is a co-created phenomenon. Organizations are employing various tools to provide employees with the right knowledge at the right time. Social networking tools have been used in variety of ways to make this knowledge exchange easier and accessible to employees. The article was an attempt to highlight some factors that have an influence over the knowledge sharing behavior of employees in an organization. However, there is still a need to consider various structural and design issues so as to fully harness the functionality of networking platforms. Along with structural improvements, successful implementation and usage of these platforms requires a conducive culture with organizational support. These are just means that can be utilized to the best of their potential by employing them in the right conditions.

REFERENCES

- Boer, N. I., Berends, H., & van Baalen, P. (2011). Relational models for knowledge sharing behavior. *European Management Journal*, 29(2), 85–97. doi:10.1016/j.emj.2010.10.009
- Chang, H. H., & Chuang, S. S. (2011). Social capital and individual motivations on knowledge sharing: Participant involvement as a moderator. *Information & Management*, 48(1), 9–18. doi:10.1016/j.im.2010.11.001
- Chen, C. J., & Hung, S. W. (2010). To give or to receive? Factors influencing members knowledge sharing and community promotion in professional virtual communities. *Information & Management*, 47(4), 226–236. doi:10.1016/j.im.2010.03.001
- Chow, W. S., & Chan, L. S. (2008). Social network, social trust and shared goals in organizational knowledge sharing. *Information & Management*, 45(7), 458–465. doi:10.1016/j.im.2008.06.007
- Din, N., & Haron, S. (2012). Knowledge sharing as a culture among Malaysian online social networking users. *Procedia: Social and Behavioral Sciences*, 50, 1043–1050. doi:10.1016/j.sbspro.2012.08.104
- Hsu, M. H., Ju, T. L., Yen, C. H., & Chang, C. M. (2007). Knowledge sharing behavior in virtual communities: The relationship between trust, self-efficacy, and outcome expectations. *International Journal of Human-Computer Studies*, 65(2), 153–169. doi:10.1016/j.ijhcs.2006.09.003
- Hung, S. Y., Durcikova, A., Lai, H. M., & Lin, W. M. (2011). The influence of intrinsic and extrinsic motivation on individuals knowledge sharing behavior. *International Journal of Human-Computer Studies*, 69(6), 415–427. doi:10.1016/j.ijhcs.2011.02.004
- Hwang, E. H., Singh, P. V., & Argote, L. (2015). Knowledge sharing in online communities: Learning to cross geographic and hierarchical boundaries. *Organization Science*, 26(6), 3. doi:10.1287/orsc.2015.1009

- Kane, G. C. (2015). Enterprise social media: Current capabilities and future possibilities. *MIS Quarterly Executive*, 14(1), 1–15.
- Kuegler, M., Smolnik, S., & Kane, G. (2015). Whats in IT for employees? Understanding the relationship between use and performance in enterprise social software. *The Journal of Strategic Information Systems*, 24(2), 90–112. doi:10.1016/j.jsis.2015.04.001
- Leonardi, P. M. (2015). Ambient awareness and knowledge acquisition: Using social media to learn” who knows what” and” who knows whom. *Management Information Systems Quarterly*, 39(4), 747–762.
- Leonardi, P. M., Huysman, M., & Steinfield, C. (2013). Enterprise social media: Definition, history, and prospects for the study of social technologies in organizations. *Journal of Computer-Mediated Communication*, 19(1), 1–19. doi:10.1111/jcc4.12029
- Leonardi, P. M., & Meyer, S. R. (2015). Social Media as Social Lubricant How Ambient Awareness Eases Knowledge Transfer. *The American Behavioral Scientist*, 59(1), 10–34. doi:10.1177/0002764214540509
- Liu, J., & Rau, P. L. P. (2014). Impact of self-construal on choice of enterprise social media for knowledge sharing. *Social Behavior and Personality: An International Journal*, 42(7), 1077–1089.
- Ma, W. W., & Chan, A. (2014). Knowledge sharing and social media: Altruism, perceived online attachment motivation, and perceived online relationship commitment. *Computers in Human Behavior*, 39, 51–58. doi:10.1016/j.chb.2014.06.015
- Mäntymäki, M., & Riemer, K. (2014). *Information, ideas and input: The value of enterprise social networks*. Paper presented at 25th Australasian Conference on Information Systems, Auckland, New Zealand.
- McKinsey, Q. (2015, October). *Taking the measure of the networked enterprise*. Retrieved from <http://www.mckinsey.com/business-functions/business-technology/our-insights/taking-the-measure-of-the-networked-enterprise>
- Nonaka, I. (1994). A dynamic theory of organizational knowledge creation. *Organization Science*, 5(1), 14–37. doi:10.1287/orsc.5.1.14
- Pan, Y., Xu, Y. C., Wang, X., Zhang, C., Ling, H., & Lin, J. (2015). Integrating social networking support for dyadic knowledge exchange: A study in a virtual community of practice. *Information & Management*, 52(1), 61–70. doi:10.1016/j.im.2014.10.001
- Patroni, J., Recker, J. C., & Von Briel, F. (2015). How does enterprise social media help retail employees innovate? In *Proceedings of the 23rd European Conference on Information Systems*. Association for Information Systems.
- Sigala, M., & Chalkiti, K. (2015). Knowledge management, social media and employee creativity. *International Journal of Hospitality Management*, 45, 44–58. doi:10.1016/j.ijhm.2014.11.003
- Sophia van Zyl, A. (2009). The impact of Social Networking 2.0 on organisations. *The Electronic Library*, 27(6), 906–918. doi:10.1108/02640470911004020
- Subramaniam, N., Nandhakumar, J., & Baptista John, J. (2013). Exploring social network interactions in enterprise systems: The role of virtual co-presence. *Information Systems Journal*, 23(6), 475–499. doi:10.1111/isj.12019
- Turban, E., Bolloju, N., & Liang, T. P. (2011). Enterprise social networking: Opportunities, adoption, and risk mitigation. *Journal of Organizational Computing and Electronic Commerce*, 21(3), 202–220. doi:10.1080/10919392.2011.590109

Wenger, E. C., & Snyder, W. M. (2000). Communities of practice: The organizational frontier. *Harvard Business Review*, 78(1), 139–146. PMID:11184968

Yu, T. K., Lu, L. C., & Liu, T. F. (2010). Exploring factors that influence knowledge sharing behavior via weblogs. *Computers in Human Behavior*, 26(1), 32–41. doi:10.1016/j.chb.2009.08.002

ADDITIONAL READING

Bradley, A. J., & McDonald, M. P. (2011). *The social organization: How to use social media to tap the collective genius of your customers and employees*. Harvard Business Press.

Bratianu, C. (Ed.). (2015). *Organizational Knowledge Dynamics: Managing Knowledge Creation, Acquisition, Sharing, and Transformation: Managing Knowledge Creation, Acquisition, Sharing, and Transformation*. IGI Global.

Dasgupta, S. (Ed.). (2009). *Social Computing: Concepts, Methodologies, Tools, and Applications: Concepts, Methodologies, Tools, and Applications*. IGI Global.

Huysman, M. H., & De Wit, D. (2002). *Knowledge sharing in practice* (Vol. 4). Springer Science & Business Media. doi:10.1007/978-94-015-9992-4

KEY TERMS AND DEFINITIONS

Altruism: Intention or action showing selfless concern for others potentially benefitting the receiver.

Self-Efficacy: The belief in one's ability to accomplish a task successfully. It affects the way an individual approaches the task.

Social Ties: The connections with fellow individuals in one's communication network. Information moves in different ways across various social ties in a social network.

Tacit Knowledge: The knowledge which is acquired by experience and practice, often difficult to explicate.

Virtual Communities: A group of individuals communicating over a social network to accomplish a shared goal.

Web 2.0: Web characterized by interactive and collaborative user generated content. The information generated is more dynamic in nature.

Understanding the Potentials of Social Media in Collaborative Learning

Adem Karahoca

Bahcesehir University, Turkey

İlker Yengin

*A*STAR, Institute of High Performance Computing, Singapore*

INTRODUCTION

Social media is a potential technology to support collaborative learning practices. This chapter illustrates the potential benefits of using social media in collaborative learning. To explain the potentials of social media, the first part of the chapter discusses the nature of the social media technologies and web 2.0. The second part illustrates the idea and importance of collaborative learning to address the changing demands of the 21st century. After introducing the fundamental concepts and ideas related to social media in first two parts, final section discusses the potentials of social media in collaborative learning following with a discussion section and a summary in the conclusion section.

BACKGROUND

Web 2.0 and related developments enabled the growth of social media concept. Web 2.0 provides both technical and philosophical bases that empower the underlying properties of social media. Web 2.0 created a new opportunity for the end users to feel related to the other users and/or institutions who are the creator, co-creator or shareholder of the content. The latest development in the Web 2.0 technology offered a platform that allows internet users to create and share many audio-visual knowledge artifacts (Yengin, 2014 a). Especially, the user owned content creation

and sharing qualities makes Web 2.0 empowered social media services as a potential platform for collaborative learning, in which the working together and co-creation of knowledge is essential.

The idea of using social media for learning is widely discussed in the literature (see following sections). Since there is a literature already discussing the social media in learning in general, this chapter aims for turning our attention to a more specific discussion on the potentials social media brings for collaborative learning activities. Hence, to analyze the issue with a comprehensive understanding, this chapter discuss the underlying mechanisms of Web 2.0 that empower the social media concept, definitions of social media with examples and list of social media services, dynamics and importance of collaborative learning in modern education and connections between collaborative learning and social media.

SOCIAL MEDIA

Before starting to talk about social media, it should be defined with a clear distinction from the interchangeably used terms such as “social networking”. Social networking is defined in Meriem-Webster dictionary as “forms of electronic communication (as websites for social networking and microblogging) through which users creates online communities to share information, ideas, personal messages, and other content (as videos)”. The same dictionary defines the social media as “a

creation and maintenance of personal and business relationships especially online”.

Several authors defined “social networking” as a venue for users to share their activities and interests with others in a particular community (Fenton, 2012). Social networking allows users to have an online profile in a bounded system to connect with other users to communicate and collaborate (Boyd & Ellison, 2007).

Different authors defined “social media” as a way of creative expression (Gauntlett, & Thomsen, 2013; Zagalo & Branco, 2015), exchanging user-generated content (Kaplan & Haenlein, 2010), spreading and sharing meaningful and valuable content (Jenkins, Ford, & Green, 2013) in a culture of collaboration and connectivity (Dijk, 2013) using web technologies such as web 2.0 (Power 2007 ;Tuten 2008 & Brown 2009).

Although dictionaries and different authors make an attempt to define the “social networking” and “social media” to show that these are not exactly same terms, people often use the term of “social networking” and “social media” interchangeably (Cohn, n.d.). This confusion could be a result of the lack of clear cut between social media and social networking concepts. To approach this issue of providing a clear and definite deception of “social networking” and “social media”, this part will provide literature about the common and different properties of these terms and provide a short and practical definition.

The basic commonalities of “social networking” and “social media” are that they both rely on the internet as a communication channel. Another common property between two is allowing people to interact in social ways. The online social interaction or online sociality is defined as “collective action, communication, building communities, connecting and networking, creative content making, collaborative knowledge building, sharing, playing etc.” by Fuchs (2013).

Comparing social media and social networking terms, the main difference is that social media allows publication of knowledge. The two main distinctive characteristics of social media are en-

abling the participation and power of publishing (Hanna, Rohm & Crittenden, 2011). While social media is related to user-generated publishing and distribution of the content knowledge, social networking creates the necessary basis and a technical framework for the social media to live on.

Interestingly, a social networking service may change its characteristics within time and transform to a social media platform. For example, a study (Kwak et al., 2010) showed that hat users’ way of using social networking tool may re-define it as a social media platform. In their study, researchers crawled 41.7 million user profiles, 1.47 billion social relations, 4, 262 trending topics, and 106 million tweets to analyze the topological characteristics of Twitter, which is an online social networking service allowing users to follow and being followed on published content (Kwak et al., 2010). According to the results of this study, Twitter had become a social media platform even though it had started as a social networking service.

As seen with the examples, sometimes the precise categorization of the technologies and services as social media or social networking platforms may be impractical since the terms are used changeable by many people and the services may transform to a different nature in time. For practical reasons of avoiding confusions, the definitions would be that the “social media” as the services that allow users to generate and share content online and the “social networking” as the services that allow to managing the relationships in online social communities.

In addition to the definitions, seeing the examples of social media services may be more helpful for readers to better understand the social media services. Using website traffic data from different sources, eBizMBA Rank website (2016) published the following list in Table 1, showing the most popular 15 social media and social networking services as of Feb 2016.

These social media services can be categorized in two formats: “users’ profiles based” and “content-based services” (Kamila & Bhattacharjee, 2014). User profile based services focus on

Table 1. The most popular social media and social networking services as of Feb 2016

Service	Estimated Unique Monthly Visitors
Facebook	1,100,000,000
Twitter	310,000,000
LinkedIn	255,000,000
Pinterest	250,000,000
Google Plus+	120,000,000
Tumblr	110,000,000
Instagram	100,000,000
VK	80,000,000
Flickr	65,000,000
Vine	42,000,000
Meetup	40,000,000
Tagged	38,000,000
Ask.fm	37,000,000
MeetMe	15,500,000
ClassMates	15,000,000

content and connections around the profile of a member of the service (e.g. personal info, hobbies, likes-dislikes, interest etc.) and the content-based services emphasis on the content development and publishing, while still providing pages of public profiles and connections.

For example, the first 3 in Table 1 are the users' profiles based social networking services emphasizing connecting with users (both in professional and non-professional manner) who are usually like-minded in order to share content and ideas for creating new knowledge using the power of the crowds in a networked community. Pinterest and Instagram may be categorized as the content-based networking services in which the produced user' contents are stored in an archive or a repository like services and stored in a social network (Andersen, 2007).

SOCIAL MEDIA AND WEB 2.0

Currently, social media technologies depend on Web 2.0 technology. In order to understand the social media, it is important to take a look at the

term of Web 2.0 technologies and its distinctive properties. Web 2.0 allows making connections through individuals, creating collaborative communities, publishing of user-generated content, integration of multiple source of knowledge and communication, using mash-up strategies to integrate many contents and harnessing collective intelligence in a decentralized manner using web services and cloud computing (o'Reilly 2007 ; Rouse & Haughn, n.d.).

Before Web 2.0, Web 1.0 offered a limited way of connection between individual users. In order to create a community, the website author, who is the only sole owner of the content, had to create a circle of connections (Fuchs et.al. 2010). Also, the publishing in Web 1.0 required a set of skills in website building and programming. Web 2.0 removed such barriers for the end users and empowered them to connect each other and become the owner of the content (Gylfason, 2010). For instance, users can co-create Web 2.0 content as a story by performing distributed comments and discussions in addition to editing and manipulating the text and pictures in the story (Alexander & Levine, 2008). In such a scenario, a reader or

a follower of the story can contribute to the story development and become a co-author. Participating in the content creation in collective interactions rather than just the consuming the information, users build collective intelligence in communities (O'reilly, T., 2007).

Shortly, the changing role and interactions between web users in production and distribution of information brings new possibilities in the intellectual growth of the societies. These possibilities provided by Web 2.0 and employed in social media platforms open new ways for activities in collaborative learning.

COLLABORATIVE LEARNING

The first idea of collaborative learning term was developed in the 1950s and 60s by a group of secondary school teachers (Bruffee, 1984). It is a type of learning experience in which learners work and be in a learning process together (Whipple, 1987). In the research literature, collaborative learning is described as cooperative activities in the educational context (e.g. reviewing course book together or sharing assignments etc.), as joint problem solving in learning processes and/or as a development of a lifelong learning process evolving in many years in communities such as in professional work field (Dillenbourg, 1999). Collaborative learning improves the relationships intensity between learners, create dynamic communicational flow, reduce duplication, lower the risks and increase reward satisfaction once the students achieve their learning goal (Keas & Mandell, 2009).

Supporting collaborative learning is important to prepare learners for the future business world. Today's economy demands the creation of new ideas and products in collaborative ways (Bonnet, 2016). The latest changing global environment forces us to move on knowledge economies (Quinn, 1992; Acs, de Groot, & Nijkamp 2013; Jorgenson & Vu, 2016), which are economic models that growth is dependent on the knowledge to generate values.

Knowledge economies need the establishment of decentralized business structures and frameworks (Madanipour, 2013) to allow the creation of economic value by employing knowledge in goods and services (Crevoisier, 2016).

This new way of doing business requires us to adopt the idea that "the creation and broadcasting of the knowledge is shifting from centralized points – where the knowledge is accumulated in entities that are controlled by certain stakeholders (e.g. teachers, books, agencies) to more distributed points – where the knowledge is created and distributed by the well-connected users in the networks" (Yengin, 2014). In order to successfully respond the changing dynamics of the new economy, institutions controlling the policies of the education systems should create new learning strategies that allow addressing the requirements and demands of this new economy. To adopt the requirements of the new economic systems in which social media plays an important role (Choi, Huang, Palmer, & Horowitz 2014), education may need to make a shift from the factory model of learning environments toward a collaborative learning environment in a social networked setting (Quintana et al., 2016; Baumeister 2005; Gibson 2006).

POTENTIALS OF SOCIAL MEDIA IN COLLABORATIVE LEARNING

Recent studies already discussed the rise (Zawacki-Richter et al., 2015) and importance (Ford, Bowden, & Beard 2011; Popescu 2014) of social media in facilitating the collaborative learning activities. Rather than focusing again on the increased used and the importance, this part focuses on functional potentials of social media collaborative learning to give the reader wider perspectives on the possibilities with social media in collaborative learning.

Before starting presenting the potentials of social media, it should be kept in mind that the social media is not an ultimate solution to all col-

laborative learning activities. First of all, social media cannot be considered as the exact replacement of face to face communication which has many benefits and superiorities over the digital communication channels (Hiltz & Turoff, 1993; Daly, 1993). Rather, social media should be considered as a supplement or an expanded way of computer-mediated communication (Watson & Hill, 2015) that allows remote communication, documentation and recording of communication, creating channels to allow generation of more ideas and have greater equality of participation (Bordia, 1997). Similarly, social media is not an exact replacement of face to face social and collaborative learning activities in the classroom, especially for friendship building for social well-being. Several authors discussed that there is no relationship between friendship quality based on social media use and well-being of students (Wang et al., 2014); but negative social well-being was positively associated with levels of uses of media (Pea et al. 2012). These authors and studies add weight to the argument that teachers and students should not use the social media as an entire replacement of traditional collaborative learning activities in the classroom, but use the social media as a supplementary communication method.

Although social media is not an exact replacement of face to face teaching, it brings new opportunities and changes for education in institution level and pedagogic level. In institution level, it is a fact that using social media for collaborative learning activities require several transformations in policies. Social media is highly influencing the educational institutions to change their traditional way of doing business in order to keep their competitive power and to strengthen their services providing cost effective and personalized learning activities, compared to the traditional learning environments (Dabbagh & Reo, 2010). To implement collaborative learning activities with social media, schools should change their processes to address issues such as “meeting specific the learning objectives, time constraints, complexity of classroom management, constraint on Internet

access, limited social media tools in schools, teachers’ limited understanding of using social media and limited guidance” (Bull et al., 2008).

In the pedagogic level, social media also introduces new approaches to the collaborative learning. Several studies showed that social media services provide new appealing and engaging learning activities since it allows applications for self-motivated, autonomous and personalized activities. In other words, the allowance of self-motivated, autonomous and personalized learning is the why the social media would be a good supplementary method for collaborative learning. For example, authors (Junco, Heiberger & Loken, 2011) investigating the impact of social media use on 125 students’ learning and engagement showed that using social media tool (Twitter) for the different academic activities and co-curricular discussions may highly engage both students and faculty to collaborate actively to the learning process. In addition to engaging activities, authors also reported that social media has effects on students’ positive feelings to collaborative activities (Hung & Yuen 2010) and social media is engaging with self-motivated and autonomous collaborative activities in and out of the classroom (McGloughlin & Lee, 2010). Also, several authors discussed that the basic elements of social media support creation of personal learning settings which lead to self-regulated learning experiences which are essential in collaborative learning (Dabbagh, & Kitsantas, 2012; Dabbagh & Kitsantas, 2013).

In addition to potentials described above, another answer for why social media is beneficial would be that the social media’s role in positive community building for collaborative learning. For example, in social level, social media creates a cycle in which students collaborate on learning task to build knowledge and create new ideas while socializing and communicating. Community engagement creates opportunities for students to participate in events and collective activities. In such social media enabled collaborative learning communities, students may construct a digital identity (Dalton & Crosby, 2013) and develop

social skills (Bauman and Rivers, 2015), form online friendship (Koutamanis et al., 2013), build rapport (Cain et al., 2013) and show responsible communication patterns to satisfy their psychosocial needs (Manago, Taylor, & Greenfield, 2012).

Ultimately, establishing good relationships and friendship are important factors in collaborative learning activities. Thus, using social media for altering the communication for friendship and the roles in a learning community becomes an important effect of social media. For example, social media addresses this by providing communications platforms for understanding and being aware of own and others' emotions, which also influence collaborative learning performances (Chanel et al., 2016).

Having argued why social media engage students' positively in learning, this part focuses on answering of for what kind of activities and when social media should be used. Actually, research showed that students use social media as guided collaborative learning activities include socializing purposes (meet other students), asking questions during self-studies, exchanging documents and making literature reviews and preparing for exams (Zawacki-Richter et al., 2015). In another earlier study, authors reported that students use social media for doing daily multi-tasking, developing friendships in a networked way, building knowledge and learn collaboratively in a decentralized way (Ulbrich, Jahnke & Martensson 2011).

Although social media provides positive potentials and opportunities as discussed above, introducing social media to the class without the teacher facilitations will not bring out these potentials. The role of the teachers' becomes a critical success factor. Students shouldn't let to use social media for learning purposes without any guidance and pedagogical strategy employed by teachers. A study (Selwyn, 2009) conducted to investigate 909 students' education-related use of social media, illustrated that students interactions are not related to academic tasks when there was no guidance and pedagogic strategy. To put it another way, using social media in learning should be guided by the

teachers. Once it is a guided activity, it opens many positive ways to collaborative learning as explained above paragraphs.

As teachers' roles become critical, teachers' pedagogic styles and use of social media should be also discussed to benefit from the social media. For example, similar to the students' usage, a survey conducted with thousands of participants (Stansbury, 2011) showed that teachers use social media mainly for networking with colleagues, using real world examples in learning, collaborative learning activities, communication, creating assessments, accessing the parents, announcements, community outreach and professional development. Another study (Moran, Seaman & Tinti-Kane, 2011) investigating the faculties attitudes toward the using social media reported that although faculty sees the social media with some value in teaching activates, they complain about the time required to engage in social media activities, and they have concerns about privacy and integrity issues.

To conclude the investigation on the potential of social media in collaborative learning and to comprehend the full potentials, finally, the risk and challenges should be presented. First of all, it should be noted that not every student are readily capable of operating on social media platforms safely. According to recent studies, such students can be exposed to some risk of peer pressure (Milner, 2015), cyberbullying, clique-forming sex abuses (O'Keeffe & Clarke-Pearson, 2011) and privacy issues (Barnes, 2006) while they use social media services. Similarly, a recent report indicated that the privacy of faculty and students is the one of the biggest concern and blockages in adopting social media in the classroom, especially widely used commercial social media tools (Seaman, & Tinti-Kane, 2013).

Several authors also discussed the difficulties that teachers face when identifying the authenticity of the students' thoughts and ideas when they use social media. For example, plagiarism is one of the concerns (Hayes & Introna, 2005; Evering & Moorman, 2012). Keen (2007) argues that using

basic cutting and pasting makes young generation believe that the idea and thought belong to them, which also allows the creation of “younger generation of intellectual kleptomaniacs”. On the other hand, the digital culture brings a new way of learning in which borrowing, sharing or promoting others’ content may be welcomed in some learning fields, if there is an awareness of the risks (Iliescu 2008;). Moreover, guidelines to use social media in learning highly encourages sharing of knowledge with resource citing (Rennie & Morrison, 2013).

Finally, the digital culture difference between students and the adults also may be an issue for using social media in collaborative learning. Teachers and parents who are not savvy in social media culture may not be able to understand the students’ lives which are heavily embedded in the social media culture (Boyd, 2009). Parents and teachers’ lack of knowledge and unfamiliarity of social media use may create a huge disconnection between generations in order to construct social communities to tackle the life and society in general (Yardi, & Bruckman, 2011).

DISCUSSIONS

To adopt social media practices and make transitions, institutions should understand and conceptualize social media as a platform for the community (Hanna, Rohm & Crittenden, 2011). In the education sector, this platform should support the students’ access as well as integration with the current systems. Furthermore, educational institutions should provide infrastructure to implement social media technologies and synthesize them into the existing school learning management system. For instance, social media and other online services of educational institutions should support students’ mobile devices and various operating systems (Zawacki-Richter et al., 2015). Also, web 2.0 based social media services that enable collaborative learning should be embedded into the web-based learning systems.

In order to avoid privacy and reputation issues, educators may be role models of responsible social media usage. They need to avoid conflicts of interest, harming public image of their institutions and creating false impressions that their personal comments are official. Educational institutions need to address the challenges of privacy, plagiarism and other ethical issues by creating policies. There may be also a need for presenting teacher training programs to instruct on the meaningful use of social media technologies in collaborative learning. Also, these institutions should provide guidance for students in using social media and self-regulated learning activities both in formal and informal learning environments.

FUTURE RESEARCH DIRECTIONS

Social media provides different ways of engaging students in collaborative activities. The research on the effects of the risks and benefits of social media in collaborative learning should be extended in order to be able to safely employ the concept with its full potentials while avoiding the risks. In this chapter, a general review of the social media benefits and risks has been provided. However, with a carefully designed study and/or series of studies, it would be possible to have a more specific analysis of social media effects on learning. For example, considering the range of the social media applications with many attributes that may cause multiple effects on different elements of the learning, it may be hard to identify the exact effects of each attribute of social media without controlling isolated effects in experimental studies. As a future research, several controlled studies investigating the effects of social media by isolating the number of variables should be carried. Once such controlled studies are conducted, it would be also easier to compare the social media’s specific effects with other technologies’ effects.

In addition, future research on social media may be extended with the studies that investigate social media effects on the development of skills

that are important for matching the 21-century learning requirements such as collaboration, communication, creativity, critical thinking, problem-solving etc. The future research on the effects of social media also should not be limited only to the positive effects; it should also cover the negative effects such as the likely effects of social media on possibly downgraded writing, face to face communication and attention-focusing skills.

Although there is a need for further research on pedagogic aspects, the social media phenomenon can't be ignored as the previous research literature signals the potentials. Thus, the role of social media in addressing the global changes in the 21 century shouldn't be ignored and there should be more investigations to understand the different forms of social media as well as potential risks and benefits for education to be able to adapt these changes by conducting future research studies.

CONCLUSION

In this chapter, the potential benefits of using social media in collaborative learning have been documented and discussed with examples from the recent research studies. The first part of the chapter investigated the essentials of the social media technologies by reviewing the terminology of social media and social networking with examples of recent social media services, connections to the Web 2.0 and the concept and importance of collaborative learning. The second part of the chapter provided a review of benefits, risks and challenges of using social media which illustrated the potentials of these technologies in a collaborative learning. Finally, the discussion part provided a set of suggestions on using social media in collaborative learning.

In conclusion, social media provides a potential platform for collaborative learning. Social media tools open a way of addressing the changing demands of 21 century. Benefiting from the enabling functions of social media that are

running on the frameworks of web 2.0, students can actively engage in the lesson content building, they become self-regulated learners owning the content rights and the responsibility of their learning as well as becoming socially connected to a learning community while forming a relationship with their learning partners. Teachers, establishing new pedagogic methods to integrate the social media tools into the classroom and curriculum may also be a part and the facilitator of such learning communities. While providing positive potential, teachers and students should be also aware of the risks and challenges of the social media such as peer pressure, cyberbullying, clique-forming, sex abuses, privacy, plagiarism and other security issues.

Finally, adaptation of social media in daily school practices may bring some extra challenges and adaptation obstacles such as meeting specific the learning objectives, time constraints, complexity of classroom management, constraint on Internet access, limited social media tools in schools, teachers' limited understanding of using social media, limited guidance and finding standard measurement and assessment of collaborative learning with social media tools.

REFERENCES

- Acs, Z. J., de Groot, H. L., & Nijkamp, P. (Eds.). (2013). *The emergence of the knowledge economy: A regional perspective*. Springer Science & Business Media.
- Alexander, B., & Levine, A. (2008). Web 2.0 storytelling: Emergence of a new genre. *EDUCAUSE Review*, 43(6), 40–56.
- Andersen, P. (2007). What is Web 2.0? Ideas, technologies and implications for education. Bristol: JISC.
- Barnes, S. B. (2006). A privacy paradox: Social networking in the United States. *First Monday*, 11(9). doi:10.5210/fm.v11i9.1394

- Bauman, S., & Rivers, I. (2015). Being Connected: Friendships and Social Interactions. In *Mental Health in the Digital Age* (pp. 103-116). Palgrave Macmillan UK.
- Baumeister, H. P. (2005). Networked learning in the knowledge economy-A systemic challenge for Universities. *European Journal of Open, Distance and E-Learning*, 8(1).
- Bonnet, J. (2016). *From Knowledge to Innovation Economy: Developing Education and Creating Entrepreneurial Ecosystems* (No. 2016-02). University of Caen and CNRS.
- Bordia, P. (1997). Face-to-face versus computer-mediated communication: A synthesis of the experimental literature. *Journal of Business Communication*, 34(1), 99–118. doi:10.1177/002194369703400106
- Boyd, D. M. (2008). *Taken out of context: American teen sociality in networked publics*. ProQuest.
- Boyd, D. M., & Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1), 210–230. doi:10.1111/j.1083-6101.2007.00393.x
- Brown, R. (2009). *Public Relations and the Social Web: How to use social media and web 2.0 in communications*. Kogan Page Publishers.
- Bruffee, K. A. (1984). Collaborative learning and the conversation of mankind. *College English*, 46(7), 635–652. doi:10.2307/376924
- Bull, G., Thompson, A., Searson, M., Garofalo, J., Park, J., Young, C., & Lee, J. (2008). Connecting informal and formal learning: Experiences in the age of participatory media. *Contemporary Issues in Technology & Teacher Education*, 8(2), 100–107.
- Cain, J., Scott, D. R., Tiemeier, A. M., Akers, P., & Metzger, A. H. (2013). Social media use by pharmacy faculty: Student friending, e-professionalism, and professional use. *Currents in Pharmacy Teaching and Learning*, 5(1), 2–8. doi:10.1016/j.cptl.2012.09.002
- Chanel, G., Lalanne, D., Lavoué, E., Lund, K., Molinari, G., Ringeval, F., & Weinberger, A. (2016). Grand Challenge Problem 2: Adaptive awareness for social regulation of emotions in on-line collaborative learning environments. In *Grand Challenge Problems in Technology-Enhanced Learning II: MOOCs and Beyond* (pp. 13-16). Springer International Publishing.
- Choi, N., Huang, K. Y., Palmer, A., & Horowitz, L. (2014). Web 2.0 Use and Knowledge Transfer: How Social Media Technologies Can Lead to Organizational Innovation. *Electronic Journal of Knowledge Management*, 12(3), 174–184.
- Cohn. (n.d.). *Social Media vs Social Networking*. Retrieved 13 June, 2016, from <http://www.com-pukol.com/social-media-vs-social-networking/>
- Crevoisier, O. (2016). The economic value of knowledge: Embodied in goods or embedded in cultures? *Regional Studies*, 50(2), 189–201. doi:10.1080/00343404.2015.1070234
- Dabbagh, N., & Kitsantas, A. (2012). Personal Learning Environments, social media, and self-regulated learning: A natural formula for connecting formal and informal learning. *The Internet and Higher Education*, 15(1), 3–8. doi:10.1016/j.iheduc.2011.06.002
- Dabbagh, N., & Kitsantas, A. (2013). The role of social media in self-regulated learning. *International Journal of Web Based Communities*, 9(2), 256–273. doi:10.1504/IJWBC.2013.053248
- Dabbagh, N., & Reo, R. (2010). Impact of Web 2.0 on higher education. *Technology integration in higher education: Social and organizational aspects*, 174-187.

- Dalton, J. C., & Crosby, P. C. (2013). Digital identity: How social media are influencing student learning and development in college. *Journal of College & Character*, 14(1), 1–4. doi:10.1515/jcc-2013-0001
- Daly, B. L. (1993). The influence of face-to-face versus computer-mediated communication channels on collective induction. *Accounting, Management and Information Technologies*, 3(1), 1–22. doi:10.1016/0959-8022(93)90006-R
- Dillenbourg, P. (1999). What do you mean by collaborative learning? *Collaborative-Learning: Cognitive and Computational Approaches*, 1, 1–15.
- eBizMBA Rank. (2016, Feb). *Top 15 Most Popular Social Networking Sites*. Retrieved 13 June, 2016, from <http://www.ebizmba.com/articles/social-networking-websites>
- Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1), 210–230. doi:10.1111/j.1083-6101.2007.00393.x
- Evering, L. C., & Moorman, G. (2012). Rethinking plagiarism in the digital age. *Journal of Adolescent & Adult Literacy*, 56(1), 35–44. doi:10.1002/JAAL.00100
- Fenton, N. (2012). The internet and social networking. In *Misunderstanding the Internet*. Abingdon, UK: Routledge.
- Ford, N., Bowden, M., & Beard, J. (2011). Learning together: using social media to foster collaboration in higher education. *Higher education administration with social media*, 105–126.
- Fuchs, C. (2013). Social media: A critical introduction. *Sage (Atlanta, Ga.)*.
- Fuchs, C., Hofkirchner, W., Schafranek, M., Raffl, C., Sandoval, M., & Bichler, R. (2010). Theoretical foundations of the web: Cognition, communication, and co-operation. Towards an understanding of Web 1.0, 2.0, 3.0. *Future Internet*, 2(1), 41–59. doi:10.3390/fi2010041
- Gauntlett, D., & Thomsen, B. S. (2013). *Cultures of creativity*. Academic Press.
- Gibson, I. W. (2006). Enhanced learning and leading in a technology rich, 21st century global learning environment. *Enhancing Learning Through Technology*, 75–102.
- Gylfason, B. P. (2010). *The Future of the Web*. Reykjavic University paper. Retrieved from <http://www.olafurandri.com/nyti/papers2010/Semantic20Web.pdf>
- Hanna, R., Rohm, A., & Crittenden, V. L. (2011). Were all connected: The power of the social media ecosystem. *Business Horizons*, 54(3), 265–273. doi:10.1016/j.bushor.2011.01.007
- Hayes, N., & Introna, L. D. (2005). Cultural Values, Plagiarism, and Fairness: When Plagiarism Gets in the Way of Learning. *Ethics & Behavior*, 15(3), 213–2. doi:10.1207/s15327019eb1503_2
- Hiltz, S. R., & Turoff, M. (1993). *The network nation: Human communication via computer*. MIT Press.
- Hung, H. T., & Yuen, S. C. Y. (2010). Educational use of social networking technology in higher education. *Teaching in Higher Education*, 15(6), 703–714. doi:10.1080/13562517.2010.507307
- Iliescu, S. (2008). Research and Debate--Beyond Cut-and-Paste: The Promise of Collage in Contemporary Design. *Places*, 20(1).
- Jenkins, H., Ford, S., & Green, J. (2013). *Spreadable media: Creating value and meaning in a networked culture*. NYU Press.
- Jorgenson, D. W., & Vu, K. M. (2016). The ICT revolution, world economic growth, and policy issues. *Telecommunications Policy*.
- Junco, R., Heiberger, G., & Loken, E. (2011). The effect of Twitter on college student engagement and grades. *Journal of Computer Assisted Learning*, 27(2), 119–132. doi:10.1111/j.1365-2729.2010.00387.x

- Kamila, K., & Bhattacharjee, S. (2014). Social Networking: An Overview. *International Journal of Research in Management & Social Science*, 56.
- Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons*, 53(1), 59–68. doi:10.1016/j.bushor.2009.09.003
- Keas, R., & Mandell, M. P. (2009). What is Collaboration?. *Australian Research Alliance for Children and Youth*, 2009, 1–3.
- Keen, A. (2007). *The cult of the amateur: How today's Internet is killing our culture*. London: Nicholas Brealey.
- Koutamanis, M., Vossen, H. G., Peter, J., & Valkenburg, P. M. (2013). Practice makes perfect: The longitudinal effect of adolescents instant messaging on their ability to initiate offline friendships. *Computers in Human Behavior*, 29(6), 2265–2272. doi:10.1016/j.chb.2013.04.033
- Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is Twitter, a social network or a news media? In *Proceedings of the 19th international conference on World wide web* (pp. 591-600). ACM. doi:10.1145/1772690.1772751
- Madanipour, A. (2013). *Knowledge economy and the city: Spaces of knowledge* (Vol. 47). Routledge.
- Manago, A. M., Taylor, T., & Greenfield, P. M. (2012). Me and my 400 friends: The anatomy of college students Facebook networks, their communication patterns, and well-being. *Developmental Psychology*, 48(2), 369–380. doi:10.1037/a0026338 PMID:22288367
- McGloughlin, C., & Lee, M. J. W. (2010). *Personalised and self regulated learning in the Web 2.0 era: International exemplars of innovative pedagogy using social software*. Academic Press.
- Milner, M. (2015). *Freaks, Geeks, and Cool Kids: Teenagers in an Era of Consumerism, Standardized Tests, and Social Media*. Routledge.
- Moran, M., Seaman, J., & Tinti-Kane, H. (2011). *Teaching, Learning, and Sharing: How Today's Higher Education Faculty Use Social Media*. Babson Survey Research Group.
- O'reilly, T. (2007). What is Web 2.0: Design patterns and business models for the next generation of software. *Communications & Stratégies*, (1): 17.
- OKeeffe, G. S., & Clarke-Pearson, K. (2011). The impact of social media on children, adolescents, and families. *Pediatrics*, 127(4), 800–804. doi:10.1542/peds.2011-0054 PMID:21444588
- Pea, R., Nass, C., Meheula, L., Rance, M., Kumar, A., Bamford, H., & Zhou, M. (2012). Media use, face-to-face communication, media multitasking, and social well-being among 8-to 12-year-old girls. *Developmental Psychology*, 48(2), 327–336. doi:10.1037/a0027030 PMID:22268607
- Popescu, E. (2014). Providing collaborative learning support with social media in an integrated environment. *World Wide Web (Bussum)*, 17(2), 199–212. doi:10.1007/s11280-012-0172-6
- Power, D.J. (2010). How will Web 2.0 impact design and development of decision support systems?. *DSS News*, 8(8).
- Quinn, J. B. (1992). *Intelligent Enterprise: A Knowledge and Service Based Paradigm for Industry*. Simon and Schuster.
- Quintana, C. D. D., Mora, J. G., Pérez, P. J., & Vila, L. E. (2016). Enhancing The Development Of Competencies: The role of UBC. *European Journal of Education*, 51(1), 10–24. doi:10.1111/ejed.12162
- Rennie, F., & Morrison, T. (2013). *E-learning and social networking handbook: Resources for higher education*. Routledge.
- Rouse, M., & Haughn, M. (n.d.). *Web 2.0*. Retrieved June 20, 2016, from <http://whatis.techtarget.com/definition/Web-20-or-Web-2>

Seaman, J., & Tinti-Kane, H. (2013). *Social media for teaching and learning*. Pearson Learning Systems.

Selwyn, N. (2009). Faceworking: Exploring students education-related use of Facebook. *Learning, Media and Technology*, 34(2), 157–174. doi:10.1080/17439880902923622

Stansbury, M. (2011). Ten ways schools are using social media effectively. *eSchool News*. Retrieved June, 6, 2016, from <http://www.eschoolnews.com/2011/10/21/ten-ways-schools-are-using-social-media-effectively/3/>

Tuten, T. L. (2008). *Advertising 2.0: Social Media Marketing in a Web 2.0 World: Social Media Marketing in a Web 2.0 World*. ABC-CLIO.

Ulbrich, F., Jahnke, I., & Martensson, P. (2011). Special Issue on knowledge development and the net generation. *International Journal of Sociotechnology and Knowledge Development*.

Van Dijck, J. (2013). *The culture of connectivity: A critical history of social media*. Oxford University Press. doi:10.1093/acprof:oso/9780199970773.001.0001

Wang, J. L., Jackson, L. A., Gaskin, J., & Wang, H. Z. (2014). The effects of Social Networking Site (SNS) use on college students friendship and well-being. *Computers in Human Behavior*, 37, 229–236. doi:10.1016/j.chb.2014.04.051

Watson, J., & Hill, A. (2015). *Dictionary of media and communication studies*. Bloomsbury Publishing USA.

Whipple, W. R. (1987). Collaborative learning: Recognizing it when we see it. *AAHE Bulletin*, 4, 6.

Yardi, S., & Bruckman, A. (2011, May). Social and technical challenges in parenting teens' social media use. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 3237-3246). ACM. doi:10.1145/1978942.1979422

Yengin, I. (2014 a). Knowledge negotiation using technology. *Educating in Dialog: Constructing Meaning and Building Knowledge With Dialogic Technology*, 24, 203.

Yengin, I. (2014 b). Using educational technology to create effective learning societies in 21st century. In *Information Technology Based Higher Education and Training (ITHET)*, 2014 (pp. 1-7). IEEE.

Zagalo, N., & Branco, P. (2015). The Creative Revolution That Is Changing the World. In *Creativity in the Digital Age* (pp. 3–15). Springer London. doi:10.1007/978-1-4471-6681-8_1

Zawacki-Richter, O., Müskens, W., Krause, U., Alturki, U., & Aldraiweesh, A. (2015). Student media usage patterns and non-traditional learning in higher education. *The International Review of Research in Open and Distributed Learning*, 16(2). doi:10.19173/irrodl.v16i2.1979

ADDITIONAL READING

Davis III, Charles HF, Regina Deil-Amen, Cecilia Rios-Aguilar, and Manuel Sacramento Gonzalez Canche. "Social Media in Higher Education: A literature review and research directions." (2012).

Feller, S., & Yengin, I. (Eds.). (2014). *Educating in Dialog: Constructing meaning and building knowledge with dialogic technology* (Vol. 24). John Benjamins Publishing Company. doi:10.1075/ds.24

Kietzmann, J. H., Hermkens, K., McCarthy, I. P., & Silvestre, B. S. (2011). Social media? Get serious! Understanding the functional building blocks of social media. *Business Horizons*, 54(3), 241–251. doi:10.1016/j.bushor.2011.01.005

Kozma, R. B. (1994). Will media influence learning? Reframing the debate. *Educational Technology Research and Development*, 42(2), 7–19. doi:10.1007/BF02299087

Lenhart, A., Purcell, K., Smith, A., & Zickuhr, K. (2010). Social Media & Mobile Internet Use among Teens and Young Adults. Millennials. *Pew internet & American life project*.

Lewis, S., Pea, R., & Rosen, J. (2010). Beyond participation to co-creation of meaning: Mobile social media in generative learning communities. *Social Sciences Information. Information Sur les Sciences Sociales*, 49(3), 351–369. doi:10.1177/0539018410370726

Madge, C., Meek, J., Wellens, J., & Hooley, T. (2009). Facebook, social integration and informal learning at university: It is more for socialising and talking to friends about work than for actually doing work. *Learning, Media and Technology*, 34(2), 141–155. doi:10.1080/17439880902923606

O'Reilly, T., & Battelle, J. (2009). *Web squared: Web 2.0 five years on*. O'Reilly.

Stahl, G. (2006). *Group Cognition: Computer Support for Building Collaborative Knowledge. Acting with Technology*.

Yengin, I., Karahoca, A., Karahoca, D., & Uzunboyly, H. (2011). Deciding which technology is the best for distance education: Issues in media/technology comparisons studies. *Procedia Computer Science*, 3, 1388–1395. doi:10.1016/j.procs.2011.01.020

KEY TERMS AND DEFINITIONS

21st Century Learning: Cognitive skills required to address the demands and challenges of the 21st century (e.g. creativity and Innovation -inventive thinking, cross-cultural understanding -global awareness and cross-cultural skills, critical thinking, effective communication, collaborating, the flexible & self-directed learning, skills related to being information and communication technology literate).

Collaborative Learning: The type of learning experiences or learning strategies that allow learners to create knowledge and meaning in a form of social construct.

Knowledge Economy: Economic models that based on the growth dependent on the knowledge to generate values.

Social Media: The services that allow users to generate and share content online.

Social Networking: The services that allow to managing the relationships in online social communities.

Web 1.0: This is the first stage of the internet that only allows the static content publication with no option for direct modification and/or creation of content by other users.

Web 2.0: The platform technology that allows making connections through users and user communities to enable users to publish and share user generated.

Using Social Media to Increase the Recruitment of Clinical Research Participants

S

Saliha Akhtar

Seton Hall University, USA

INTRODUCTION

Clinical trials are research designed to evaluate ways to prevent, detect, or treat diseases and conditions (National Institutes of Health, 2015). Despite their importance, it has become increasingly challenging to enroll participants in clinical trials. According to Mahon, Roberts, Furlong, Uhlenbrauck, and Bull (2015), a high number of clinical trials fail to recruit the necessary number of participants. Multiple studies have been conducted to estimate how many fail, and although they vary in the exact number, researchers recognize and agree that insufficient recruitment is a significant problem that continues to exist in clinical trials. As the number of clinical trials increases (Inventiv Health, 2013), as there is an increased focus on research that involves more complex diseases and specialized medicines, and as researchers work to develop treatments that are better targeted to patients (Akhtar, Israel, & D'Abundo, 2015), the problem of insufficient recruitment will continue to grow unless changes are made to the recruitment process.

In general, researchers rely on traditional recruitment methods to enroll participants. Traditional recruitment methods revolve around the physicians and their support staff having the primary role to locate and recruit these participants. This could include them personally recruiting patients, searching their patient databases, or through printed materials such as flyers and newspapers (Tanner, Kim, Friedman, Foster, & Bergeron, 2015). Although these ways can lead to

the recruitment of some participants, they prove to be limited to a small patient population and are not aligned with changing times. In recent years, medical information is becoming more available to patients on the internet. Similarly, according to the Pew Research Center (2014), 87% of adults use the internet. In fact, many patients are now using the internet as the first source to retrieve health-related information such as on physicians, diagnoses, and therapies (Omurtag, Jimenez, Ratts, Odem, & Cooper, 2012). Furthermore, individuals are becoming more informed when it comes to health-related issues and are taking a bigger role in their healthcare decisions. In fact, when experiencing symptoms, many individuals are first searching the internet before deciding to see a physician (Shere, Zhao, & Koren, 2014). Despite this, recruitment strategies have not adapted with these changes. “What most in the industry have not yet adjusted to is the fact that patients and caregivers have changed how and where they seek and find information – especially healthcare information” (Inventiv Health, 2013, p. 6). With individuals using online platforms to retrieve information, this creates an opportunity for research site personnel to use it as a way to relay information on clinical trial opportunities.

More specifically, social media is an online platform where discussions and engagement can take place between individuals, online communities, and research staff (Shere et al., 2014). It is a platform that can be used to discuss or disseminate information on clinical trials. Grajales III, Sheps, Ho, Novak-Lauscher, and Eysenbach

(2014) found that social media is being used to increase and maintain communications across the different stakeholders, which could include communications between research site personnel and potential participants. Therefore, using social media could be an avenue for investigative site personnel to use to disseminate information on clinical trials to potential participants.

This chapter will review clinical research studies that have successfully utilized social media strategies in recruitment, along with the conclusions and recommendations that future researchers should consider when deciding whether to implement this type of strategy in their clinical trial recruitment.

BACKGROUND

Clinical trials are designed with the objective to enroll a specific number of participants who are to receive a specified treatment such as a drug, procedure, or behavioral program (Akhtar et al., 2015). However, trials frequently discontinue, even in late stage Phase III trials, due to failure to accrue (Schroen et al., 2010). The failure to recruit the necessary number of participants can have an effect on the overall success of the study (Embi et al., 2005). Insufficient recruitment could lead to delays in the approval of necessary medications and higher costs due to the extended recruitment period. In addition, the inability for a study to enroll the target sample size can lead to an insufficient statistical power (Carlisle, Kimmelman, Ramsay, & MacKinnon, 2015). This can prohibit the ability to answer the study's research questions or lead to conclusions which are not accurate. Further, the inability to enroll the necessary number of participants can have an effect on all stakeholders including the investigative site personnel led by principal investigators who carry out the clinical trials, and the individuals with diseases or conditions.

Social Media

Social media use is expected to continue to increase. As of 2015, the Pew Research Center found that approximately 65% of American adults use social networking sites, an increase from 7% as of 2005 (Perrin, 2015). Social media are “web-based tools that are used for computer-mediated communication” (Grajales III et al., 2014, para. 5). There are different types of social media platforms including blogs, microblogs such as Twitter, social media sites such as Facebook, and thematic networking sites such as forums that center on a particular disease or condition. The applications are diverse in their traditional primary functions – for example, LinkedIn for professional networking, Facebook for social networking, YouTube for media sharing, and blogs and Twitter for content production (Dizon et al., 2012).

The traditional uses of many of these platforms are evolving into more of knowledge-based platforms. These types of social media applications are now platforms to share new information and encourage the exchange of ideas (Dizon et al., 2012). Therefore, they can be used to communicate in a variety of ways including the sharing and discussion of clinical trial recruitment opportunities. In fact, strategies that inform and educate on the diseases and conditions being studied have shown to lead to the improvement in recruitment (Dizon et al., 2012). Therefore, the benefits for providing this type of information are two-fold – educating the population of interest about the disease or condition and increasing their interest in the research opportunity. Patel, Doku, and Tennakoon (2003) described it as a process of providing information to the potential participants and then creating an interest for the study.

Social media applications and tools are widely accessible and are usually of no to low cost (Dizon et al., 2012). In consequence, they can be advantageous to use versus traditional methods, which are limited to a small patient pool and can be costly

at times. Another advantage is the ability to target specific populations. Social media networks such as Facebook can be used to target specific individuals based on their user profiles and have the flexibility to be changed in real-time if necessary (Frandsen, Walters, & Ferguson, 2014). Therefore, there is the opportunity to monitor and adapt to responses in real-time and maximize the ability to recruit potential participants.

Changing Recruitment Practices

In order to keep up with current trends, it is important to change or enhance clinical trial recruitment practices. Traditional recruitment methods, which entail the physician along with their investigative site personnel as the primary roles to locate potential participants can no longer be the primary method as they are limited to a small group of participants. For example, the literature has shown that traditional recruitment methods exclude minorities populations and women, and thus, they are significantly under-represented in clinical trials. However, utilizing social media as an avenue to recruit participants for clinical research also presents an opportunity to increase the enrollment of these populations. According to Jens Michael Krogstad from Pew Research Center (2015), although there are some differences in the preference of social media sites, Latinos, Blacks, and Whites use social media platforms about the same. Further, research has also shown that women are more likely than men to use social networking sites (Perrin, 2015) and studies using social media in recruitment have been able to successfully them (Fenner et al., 2012; Tweet, Gulati, Aase, & Hayes, 2011; Shere et al., 2014). Finally, traditional recruitment methods also exclude other groups such as those who do not have a primary physician and those in rural areas (Tanner et al., 2015), which is not a problem with social media. Therefore, recruitment through social media can resolve some of the issues that are raised with traditional recruitment methods.

There are advantages to using social media to recruit participants to clinical research and trials. However, the number of clinical trials using social media platforms for recruitment is limited. The next section of this chapter will review some studies that have utilized social media to recruit participants.

Integrating Social Media in Recruitment

One study that implemented social media in the recruitment of a clinical trial focused on women who were either early in their pregnancy or trying to conceive. In this study, the researchers used traditional recruitment methods in the first phase of the study and added in social media as a supplementary recruitment source in the second phase of the study (Shere et al., 2014). The traditional recruitment methods consisted of targeting those women who called a telephone counselling program, using notice board postings, and sharing the brochures during clinic appointments. Potential participants were also identified through patient chart review. After 4 years of using these recruitment methods, the recruitment strategy was re-evaluated and it was decided to not only expand advertising to other healthcare establishments, but to also use social media. In the six months when social media recruitment methods were used, the researchers were able to recruit 12 times more women. The researchers used a variety of social media methods including Craigslist, Facebook, Twitter, and pregnancy discussion forums and message boards such as *Baby and the bump* and *Baby on the way* to post ads on the study. Individuals who were interested to participate in the study were asked to contact the study coordinator. After discussion of the study details with a healthcare professional involved in the study, the potential participants were invited to the hospital to go over the informed consent and to continue with the study. The researchers found that despite expanding the use of traditional methods in the

second phase, social media contributed to a large portion of the recruitment. In fact, 78% of the recruitment was achieved through social media methods. The authors concluded that although traditional recruitment methods were used in both phases of the study, the introduction of the social media strategies led to a surge in the recruitment. Online social media platforms have the ability to access individuals who are actively looking for information about a specific topic, which would lead them to research opportunities, whereas traditional recruitment methods are generally passive recruitment strategies. This goes back to the point made earlier in the chapter that individuals are using the internet as a primary platform to seek information. It is also important to note that the researchers strategically not only used social media platforms which have a wide following and ability to target a large number of people, but they also specifically targeted sections of the social media where a large number of women and families are typically found. The researchers also thought that the social media created awareness to the study that may have led to some recruitment from the traditional methods.

Another clinical trial that used social media for recruitment was a health intervention trial involving young adult smokers. The researchers used a Facebook ad campaign to target and recruit young adults to the smoking cessation study (Ramo, Rodriguez, Chavez, Sommer, & Prochaska, 2014). Thirty-six ads, which included a combination of different ads including sponsored stories and picture/text ads, ran over a 7-week period. The campaign was successful and showed that Facebook is an efficient and affordable method to enroll young adults, especially as the researchers were able to target specific study populations via this method. Due to the success in the targeted approach, the researchers advise that it is possible to target a specific population by location, demographics, and keywords in Facebook. In addition, simplicity seemed to be important in relaying the main study message. Given the limited space available, complex images and too

detailed messaging is advised against as it can lead to confusion and the possibility of the main message being missed. Also, newsfeed ads were found to be more successful than the other ad types in regards to the number of people it reached, the number of times it was clicked on, and its' cost. Moreover, the researchers found that Facebook is an effective way to reach a very large sample of young adults across the United States in a short amount of time. However, it is also important to point out that using Facebook may limit the generalizability of the sample population. In this study, the sample population was predominantly non-Hispanic Caucasian urban-residing men compared to what is reported in the literature. Therefore, it is suggested that future researchers use other social media platforms, which although might be used less overall, will have a stronger appeal to specific populations. For example, research has shown that that Twitter has a higher appeal among younger adults, urban dwellers, and non-whites (Duggan & Smith, 2013).

Another smoking cessation clinical trial also attempted to explore the advantages of online social media for recruitment. The researchers explored the feasibility of social media recruitment in a clinical trial evaluating the effectiveness of a behavioral support program on smoking cessation (Frandsen et al., 2014). This study used a combination of both traditional and social media methods for recruitment. The traditional recruitment methods consisted of flyers around two university campuses and ads in newspapers. The social media recruitment method consisted of multiple Facebook advertisements that included different wording and logo combinations. When an individual would click on the advertisement, it would direct them to another website where they would receive brief information on the study and a link to enter in their contact details so they could be contacted to be screened for the study. The researchers closely monitored the advertisements, activating and removing them based on the number of clicks and the capacity at the investigative sites. Similar to the previous

studies, the researchers targeted a specific population – adults living within a certain distance from the sites. The researchers were able to enroll 51.9% of the participants through the Facebook advertisements. Although the participants who were recruited via social media were seven years younger than those who were recruited through the traditional methods, there were no other differences in the characteristics between them. It is also important to note that in this study, the cost of online advertising was found to be more than the traditional method of advertising the study in the newspaper. In fact, the online advertising cost two times more per participant versus the traditional recruitment methods. However, Frandsen et al. (2014) note that this is an estimate and various factors can affect the cost and determining the exact amount. For example, a higher percentage of contacts from the social media avenue ended up enrolling in the study, perhaps because they had more detailed information on the study prior to screening, not only lowering the cost of sharing information but also expediting the process up to enrollment. Based on the results of the study, the researchers recommend using online social media to advertise as a supplement to traditional recruitment strategies.

Another case study that will be presented is a research study conducted by researchers at Mayo Clinic who utilized social media for the recruitment of their study. A member of a social networking site approached the researchers to perform research on her condition, spontaneous coronary artery dissection (Tweet et al., 2011). The researchers who recognized the value of increasing their knowledge of this condition designed a pilot study. The researchers used Inspire, a social networking site for women with heart disease, to recruit women who self-identified with spontaneous coronary artery dissection. The study is different from the others presented as it both recruited and collected data from the participants purely online. The researchers were able to recruit the target sample size of 12 after just one week of receiving IRB approval. Given the already established use of

Inspire for sharing information on heart disease, resources, and opportunities, the researchers were able to utilize this strong aspect of the platform to recruit the participants and collect their medical records and imaging data. The pilot study illustrated the advantage of using a disease-specific social networking community online to recruit for their study. Their study also demonstrated that social networking can enable research participant engagement and lead to recruitment. Furthermore, although recruiting potential participants with rare diseases and conditions can be challenging, the study demonstrated that it might not be difficult when using a social networking platform.

The last study that will be discussed is an online preventative intervention study. The Mood Memos study was a randomized controlled trial with the objective to evaluate if promotional messages could improve self-help behaviors for depression (Morgan, Jorm, & Mackinnon, 2013). Participants were to receive emails over six weeks, which encouraged the use of effective self-help strategies. A variety of online-based sources were used to recruit participants over 14 months, which included Google advertising, Facebook, forums, links from websites, online community noticeboards, and email groups or lists. The researchers tested Facebook but found it to be less cost-effective than advertising on Google so did not continue with it. Also, other researchers have found, Facebook is a useful platform for targeting populations with specific demographics. However, as this study was not targeting specific demographic populations, the researchers did not find Facebook to be a viable resource. Surprisingly, this study also did not find online depression forums to be good recruitment source. The researchers did a search through Google and Bing and followed links from other sites to identify appropriate forums. Permission was received from less than half of the 58 forums that were approached. While smaller forums did give permission to post, it was found that it was a time-consuming process to post and moderate responses. Moreover, recruitment via this method

was found to be limited as many of the individuals did not meet the eligibility criteria. Although the study was able to demonstrate that it is possible to recruit a large sample size purely through online sources, there was difficulty recruiting through them; however, the most effective method was found to be Google.

The next section will discuss some of the key pros and cons from the studies discussed and provide some guidance on what future researchers and investigative site personnel should consider when deciding whether to use social media as a recruitment method.

FUTURE RESEARCH DIRECTIONS

As the studies presented in this chapter have found, there are both pros and cons to using social media in the recruitment of clinical trials. One of most commonly reported benefits of using social media in recruitment is the ability to have a wider reach and to be able to use a targeted approach to find the right participants. Traditional recruitment methods are considered to be mostly a mass advertising approach while social media has the ability to target specific demographics and populations. As clinical research becomes more focused, it will become even more important to find the right patients. Another positive aspect of social media, specifically when focusing on established online community networks and forums, is the ability to share information and research opportunities with individuals who are already well-connected and involved with the sharing of information around a specific disease or condition. These individuals are more likely to be interested to learn and potentially participate in research opportunities.

On the other hand, as Morgan et al. (2013) pointed out, there are also some difficulties in using social media. Unlike traditional recruitment methods, in order to be the most effective, social media recruitment methods require continuous monitoring and effort, especially since they are in “real-time.” Shere et al. (2014) recommended

closely monitoring response rates and revising methods based on responses. In fact, the dynamic nature of social media compels researchers to continuously monitor ads and postings to determine if changes are needed based on the responses received. This will help ensure that the recruitment strategies that are implemented are the most appropriate at any given time. In addition, there have been conflicting findings in the literature about the cost of using social media for recruitment compared to traditional methods. Frandsen et al. (2014) estimated the cost of using social media to be more compared to traditional recruitment. However, they also note that this can vary study by study depending on various factors. Finally, it is also important for researchers to be cautious about the potential security and confidentiality issues that could occur with social media. Just like with most information that is online, there are some concerns that are introduced when social media is used for participant recruitment. Bull et al. (2011) discussed a study that used Facebook in trial on preventative HIV education. They found that confidentiality and data security concerns may be raised when using social media in research.

Future research studies should consider utilizing social media for recruitment in order to have the most efficient process. While no formal guidelines have been set forth by health authority agencies on how future research studies should use social media for recruitment, there have been some recommendations set forth by researchers on how to successfully use this avenue as a recruitment strategy. Shere et al. (2014) first recommended understanding the target population in order to be able to determine which platforms would be the most appropriate for them. In doing so, the researchers are not just using an approach that is geared to the general population, but instead are taking into account the characteristics of the targeted population and utilizing an approach that would be best geared towards them. In turn, this will help ensure that the recruitment strategies are best tailored for the targeted population. Another way to maximize the recruitment process would

be to use a variety of recruitment methods in order to maximize the reach of potential participants and to be the most time- and cost-effective. Shere et al. (2014) recommended using a combination of passive, broad-spectrum and targeted, active recruitment techniques while Inventiv Health (2013) specifically proposed using a combination of social media methods and traditional methods. Both of these ways could help increase the efficiency and probability of successful recruitment. Further, similar to the suggestion by Ramo et al. (2014), Inventiv Health (2013) also noted that information and advertisements on social media should contain terms that are understood by the population and to ensure simplicity through empathy and clarity. Through simplicity, there is a greater probability that the social media user will understand the message that is being conveyed. Moreover, through all of these approaches, future research can maximize on the benefits of social media in recruitment.

CONCLUSION

Traditional recruitment methods are no longer viewed as the most effective way to recruit participants to clinical trials. The ability to access healthcare information on the internet has paved the way for patients to become more involved in their healthcare decision-making. In turn this means that investigative site personnel should consider providing information on diseases and conditions as these patients can then become potential participants. Further, by using social media platforms to share health-related information and clinical research opportunities, there is an ability to reach potential participants who might not have been reached through traditional recruitment methods. Social media allows investigative site personnel to reach a larger patient population and a more diverse patient pool, including those populations who have continuously been under-represented in clinical trials such as minorities and women. Although there is still a need for

investigative site personnel to have a formal set of guidelines to follow when using social media recruitment practices, there are some studies and recommendations that shed light on how to achieve success through this type of recruitment strategy.

Recruitment strategies need to shift to an approach that is not only more targeted, but also has a larger reach. By evaluating the success of studies that have used social recruitment strategies so far, it is evident that future researchers can also achieve recruitment success through social media. Social media could be a promising new avenue for clinical trial recruitment that allows for a more positive experience for both investigative site personnel and potential participants.

REFERENCES

- Akhtar, S., Israel, C., & D'Abundo, M. L. (2015). Diversifying clinical research participants: The potential role of health information technologies and online strategies. In V. C. X. Wang (Ed.), *Handbook of Research on Advancing Health Education Through Technology* (pp. 234–258). Hershey, PA: IGI Global.
- Bull, S. S., Breslin, L. T., Wright, E. E., Black, S. R., Levine, D., & Santelli, J. S. (2011). Case study: An ethics case study of HIV prevention research on Facebook: The Just/Us Study. *Journal of Pediatric Psychology*, 1-11.
- Carlisle, B., Kimmelman, J., Ramsay, T., & MacKinnon, N. (2015). Unsuccessful trial accrual and human subjects protections: An empirical analysis of recently closed trials. *Clinical Trials*, 12(1), 77–83. doi:10.1177/1740774514558307 PMID:25475878
- Dizon, D. S., Graham, D., Thompson, M. A., Johnson, L. J., Johnston, C., Fisch, M. J., & Miller, R. (2012). Practical guidance: The use of social media in oncology practice. *Journal of Oncology Practice*, 8(5), e114–e124. doi:10.1200/JOP.2012.000610 PMID:23277774

- Duggan, M., & Smith, A. (2013). *Social media update*. Retrieved from <http://www.pewinternet.org/2013/12/30/social-media-update-2013/>
- Embi, P. J., Jain, A., Clark, J., Bizjack, S., Hornung, R., & Harris, C. M. (2005). Effect of a clinical trial alert system on physician participation in trial recruitment. *Archives of Internal Medicine*, 165(19), 2272–2277. doi:10.1001/archinte.165.19.2272 PMID:16246994
- Fenner, Y., Garland, S. M., Moore, E. E., Jayasinghe, Y., Fletcher, A., Tabrizi, S. N., & Wark, J. D. et al. (2012). Web-based recruiting for health research using a social networking site: An exploratory study. *Journal of Medical Internet Research*, 14(1), 1–14. doi:10.2196/jmir.1978 PMID:22297093
- Frandsen, M., Walters, J., & Ferguson, S. G. (2014). Exploring the viability of using online social media advertising as a recruitment method for smoking cessation clinical trials. *Nicotine & Tobacco Research*, 16(2), 247–251. doi:10.1093/ntr/ntt157 PMID:24127266
- Grajales, F. J. III, Sheps, S., Ho, K., Novak-Lauscher, H., & Eysenbach, G. (2014). Social media: A review and tutorial of applications in medicine and health care. *Journal of Medical Internet Research*, 16(2), 1–23. PMID:24518354
- Inventiv Health. (2013). *E-recruiting: Using digital platforms, social media, and mobile technologies to improve clinical trial enrollment*. Retrieved from http://www.inventivhealth.com/docs/e-Recruiting_Using_Digital_Platforms_Social_Media_and_Mobile_Technologies_to_Improve_Clinical_Trial_Enrollment.pdf
- Krogstad, J. M. (2015). *Social media preferences vary by race and ethnicity*. Retrieved from <http://www.pewresearch.org/fact-tank/2015/02/03/social-media-preferences-vary-by-race-and-ethnicity/>
- Mahon, E., Roberts, J., Furlong, P., Uhlenbrauck, G., & Bull, J. (2015). *Barriers to clinical trial recruitment and possible solutions: A stakeholder survey*. Applied Clinical Trials.
- Morgan, A. J., Jorm, A. F., & Mackinnon, A. J. (2013). Internet-based recruitment to a depression prevention intervention: Lessons from the Mood Memos study. *Journal of Medical Internet Research*, 15(2), e31. doi:10.2196/jmir.2262 PMID:23403043
- Omurtag, K., Jimenez, P. T., Ratts, V., Odem, R., & Cooper, A. R. (2012). The ART of social networking: How SART member clinics are connecting with patients online. *Fertility and Sterility*, 97(1), 88–94. doi:10.1016/j.fertnstert.2011.10.001 PMID:22088209
- Patel, M. X., Doku, V., & Tennakoon, L. (2003). Challenges in recruitment of research participants. *Advances in Psychiatric Treatment*, 9(3), 229–238. doi:10.1192/apt.9.3.229
- Perrin, A. (2015). *Social media usage: 2005-2015*. Retrieved from <http://www.pewinternet.org/2015/10/08/social-networking-usage-2005-2015/>
- Pew Research Center. (2014). *Internet users in 2014*. Retrieved from <http://www.pewinternet.org/files/2014/02/12-internet-users-in-2014.jpg>
- Ramo, D. E., Rodriguez, T. M., Chavez, K., Sommer, M. J., & Prochaska, J. J. (2014). Facebook recruitment of young adult smokers for a cessation trial: Methods, metrics, and lessons learned. *Internet Interventions*, 1(2), 58–64. doi:10.1016/j.invent.2014.05.001 PMID:25045624
- Schroen, A. T., Petroni, G. R., Wang, H., Gray, R., Wang, X. F., Cronin, W., & Slingluff, C. L. et al. (2010). Preliminary evaluation of factors associated with premature trial closure and feasibility of accrual benchmarks in phase III oncology trials. *Clinical Trials*, 7(4), 312–321. doi:10.1177/1740774510374973 PMID:20595245
- Shere, M., Zhao, X. Y., & Koren, G. (2014). The role of social media in recruiting for clinical trials in pregnancy. *PLoS ONE*, 9(3), 1–9. doi:10.1371/journal.pone.0092744 PMID:24671210

Tanner, A., Kim, S. H., Friedman, D. B., Foster, C., & Bergeron, C. D. (2015). Promoting clinical research to medically underserved communities: Current practices and perceptions about clinical trial recruiting strategies. *Contemporary Clinical Trials*, 41, 39–44. doi:10.1016/j.cct.2014.12.010 PMID:25542611

Tweet, M. S., Gulati, R., Aase, L. A., & Hayes, S. N. (2011). Spontaneous coronary artery dissection: A disease-specific, social networking community-initiated study. *Mayo Clinic Proceedings*, 86(9), 845–850. doi:10.4065/mcp.2011.0312 PMID:21878595

ADDITIONAL READING

Batterham, P. J. (2014). Recruitment of mental health survey participants using Internet advertising: Content, characteristics and cost effectiveness. *International Journal of Methods in Psychiatric Research*, 23(2), 184–191. doi:10.1002/mpr.1421 PMID:24615785

Brief, D. J., Rubin, A., Keane, T. M., Enggasser, J. L., Roy, M., Helmuth, E., & Rosenbloom, D. et al. (2013). Web intervention for OEF/OIF veterans with problem drinking and PTSD symptoms: A randomized clinical trial. *Journal of Consulting and Clinical Psychology*, 81(5), 890–900. doi:10.1037/a0033697 PMID:23875821

Gupta, A., Calfas, K. J., Marshall, S. J., Robinson, T. N., Rock, C. L., Huang, J. S., & Raab, F. et al. (2015). Clinical trial management of participant recruitment, enrollment, engagement, and retention in the SMART study using a Marketing and Information Technology (MARKIT) model. *Contemporary Clinical Trials*, 42, 185–195. doi:10.1016/j.cct.2015.04.002 PMID:25866383

Reuter, K., Ukpolo, F., Ward, E., Wilson, M. L., & Angyan, P. (2016). Trial Promoter: A Web-Based Tool for Boosting the Promotion of Clinical Research Through Social Media. *Journal of Medical Internet Research*, 18(6), e144. doi:10.2196/jmir.4726 PMID:27357424

Yuan, P., Bare, M. G., Johnson, M. O., & Saberi, P. (2014). Using online social media for recruitment of human immunodeficiency virus-positive participants: A cross-sectional survey. *Journal of Medical Internet Research*, 16(5), e117. doi:10.2196/jmir.3229 PMID:24784982

KEY TERMS AND DEFINITIONS

Clinical Research: Research that directly involves humans or that utilizes materials from humans (such as blood or tissue).

Clinical Trials: A type of clinical research where individuals are assigned to intervention(s) to evaluate the outcomes.

Investigative Site Personnel: Teams of healthcare professionals who are led by principal investigators and are responsible for carrying out the protocol.

Participant Recruitment: Finding and enrolling the best-qualified patients in a clinical trial in order to answer research question.

Principal Investigator: A primary researcher who is responsible for conducting a clinical research study or trial as per the protocol.

Social Media: Web-based tools that allow for various types of communication or information sharing.

Social Networking: Websites or applications allowing individuals to interact with other users such as those with similar interests.

Why It Is Difficult to Disengage From Facebook

Sonda Bouattour Fakhfakh

University of Tunis El-Manar, Tunisia

INTRODUCTION

In many countries, social life has become heavily reliant on the use of Social Network Sites (SNS). The leader of SNS is Facebook (FB) with more than one billion monthly active users (Patterson, 2015). Founded in 2004 mainly by Mark Zuckerberg, FB was exclusively destined to the Harvard University. Later, FB began to expand to reach the entire world. The biggest demographic using the site is the 18-29 age range. The graduated adults from college are the leaders and women seem to use it more than men (Patterson, 2015).

The deeper intrusion of FB in the society and in the professional context has removed concern not to enhance connection but to break it. Many connected people begin expressing their willingness to be disengaged. They suffer from being invaded by technologies and steered to adapt their FB experience to an economic object (Van Dijck, 2013). Even though, they could not detach and continue to integrate FB into their daily lives they feel bored or busy. So, the power is already handed on those who could disconnect (Jauréguiberry, 2014). Thus, the main question that we try to give some responses to is: why it is so difficult to disengage from FB and alter its use?

We should note here that disconnection/disengagement is not the opposite of connection. It is obviously coupled with it and could not be understood without its existence. It involves the

choose-nots category by distinction with the have-nots and the want-nots, who consciously choose to not engage with FB partially or totally (Jauréguiberry, 2014). We consider disengagement as an approach of re-appropriation of FB use; a way of changing one's practices and a situation where the relation to and with FB is at risk (Karppe, 2014). When users will be faced by the choice to be separated from FB or individuals, attachment will be highlighted. Based on the attachment theory and the Actor Network Theory (ANT), we try to underline the nature and the quality of the attachment within FB network social structure. We then attempt to study the human/human attachment and the human/non-human attachment in order to border the disengagement phenomenon. Hence, the aim here is not to reject (or not) some formulated hypothesis but to develop a theoretical frame from the existing theories.

This analysis could help users as well managers and technology designers to understand how attachments to FB could impact people emotion and behavior. It answers to the serious preoccupation of the professional context about the permanent connection of their employees to FB and information technologies in general. This analysis could also help companies willing to implement disengagement measures to apprehend this choice from a relational perspective in order to get the implication of all users.

BACKGROUND

Preliminary Debate

While many studies emphasize the good effects of FB on users and its propensity to satisfy the need to belong and the need for positive self-presentation (Nadkarni & Hofman, 2012), some others draw its negative side such as the social media fatigue (Laumer et al., 2013) and the risk of dependence and addictions (Xu & Tan, 2012).

The question of whether the consequences of FB use are positive or negative is still debatable. However, it seems to influence / be influenced by the degree of user's engagement involved in/by/with FB (Karppi, 2014). If the disengagement phenomenon is related to the increase of connection (Sheldon et al., 2011), we should, at the beginning ask why people are so attached to the use of FB?

Amichai-Hamburger (2002) and Hart et al. (2015) assert that personality is a major factor in Internet use and in determining people's engagement causes and consequences with social media (here FB). Nadkarni and Hofman (2012) sustain that a high level of extraversion, neuroticism, narcissism, and low levels of self-esteem and self-worth are associated with high FB use. In two surveys of nearly 600 people aged 18 to 83, Hart et al. (2015) conclude that people who are more engaged with FB are higher in attachment anxiety and extraversion. However, they find that extraversion, neuroticism, and self-esteem were not as consistently related to user FB engagement.

Then, as FB is a virtual social network through which users create and entertain relations with each other, the attachment theory seems to be one of the leading theories for studying interpersonal relationships (Rom & Alfasi, 2014). Using FB seems to be influenced by how people successfully build relationships with their primary caregivers during their childhood period (Andangsari et al., 2013) which is successfully transmitted to adulthood (Hazan & Shaver, 1987).

Furthermore, we cannot treat the attachment of FB user only from the user perspective and so, from the study of human/ human relation. User practices should also be examined through the network conditions that empower them and through all things which can possess the user as much as he/she possesses them (Fenwick & Edwards, 2010). In fact, all the components offered by FB (messages, photos...) allow users to create/ destruct links, intensify/decrease ties and choose to interact with near/distant friends.

These complex associations, formed by human and non-humans entities, make the ANT (Latour, 1999) useful when studying FB user attachment. The identification of actors and their associations enables the understanding of their expectancies and objects and the appreciation of their role in the technology construction, adjustment and appropriation.

The Human/Human Attachment via FB Platform

Being connected with others is considered by the self-determination theory as one of the basic psychological human need. Nowadays, users compete to register the high score of FB friends/ likes/sharing regardless of their type of use to show their great socialization and famous. As the use of Internet, FB might be of particular utility for who are physically distant or who avoid face-to-face contact as they find difficulties to interact with others (Nadkarni & Hofman, 2012). It can so connect the absence (spatially) with the presence (Giddens, 2013), lower barriers to communication, enhance the control of time and space of interaction (Mckenna & Bargh, 2000), which provide a sense of security and belonging, the main elements of attachment.

The theory of attachment describes and explains the attachment styles of adult to close personal relationship (Hazan & Shaver, 1987). Within FB network, users are not simple individuals related to each other. They are relational individuals having, at the same time, different facets

and identities, which shape their representation (Dalsgaard, 2008) not only about themselves and the others but also about interpersonal loss, rejection, or abandonment. Founded on Bowlby's work on children's attachment to their primary caregiver, Hazan and Shaver (1987) distinguish between two dimensions (based on the internal working model) of attachment: secure and insecure. The latter is formed by three patterns (anxious, avoidant and disorganized).

Several studies have examined the relation between the attachment styles and SNS behaviors.

Despite the fact that some researchers have reported the absence of difference in SNS use (FB and Twitter) by attachment styles (Hansen, 2008), others have stated a significant dependence between use of FB and insecure attachment style (Hart et al., 2015; Oldmeadow et al., 2013; Rom & Alfasi, 2014). In fact, while the secure attachment people tend to be more comfortable with intimacy and independence when interacting with others (both on/off line), the insecure attachment people (the disorganized and anxious attachment styles) are significantly predictive of using SNS to avoid a more personal face-to-face discussion (Nitzburg & Farber, 2013). However, the disorganized attached people are likely to not feel intimate on these platforms and so they may express less demand for care on the SNS sites than the anxiously attached people. On the other hand, the avoidant attachment users show the opposite behavior of anxiously and disorganized attached people in using FB. They are reluctant to use FB. This could be due to their self-sufficiency and their lower involvement in building personal relationships over SNS (Hart et al., 2015; Oldmeadow et al., 2013).

So, anxiously attached people seem to be the most engaged on FB (Andangsari et al., 2013). The anxiety aspect of attachment refers here to the degree of concern above others' evaluation and approval. This type of individuals appears to be sensitive to positive feedback (likes, comments...)

which affects their feeling about themselves and intensifies their frequency of posting and updating status (Hart et al., 2015). Since they perceive others as available for their requests and problems, they are prone to greater importance to manage their self-presentation in order to get their attention because they anticipate and fear rejection (Hart et al., 2015; Oldmeadow et al., 2013).

However, and as noted by Rom and Alfasi (2014), by being active on FB use, anxiously attached people seem to perpetuate the same characteristics of social inhibition shown in the real world. If it is so, it may be possible then to understand why they cannot disengage. Disengagement would be assimilated to separation. When the anxiously attached persons are faced even to short separations, they could be profoundly disturbed and sometimes, this situation can lead them to panic states (Karen, 1994). So, they resist to this separation and could be so clingy that they may suffocate other by their claim of attention.

Indeed, if becoming more detached (but not less connected) during a period caused increased use of FB (Sheldon et al., 2011), it is likely to suggest, as argued by Bauman (2010), that users are becoming governed by the strong desire to be attached with others. If it is so, could we advance that insecurity, dependence on the emotions and acts of others and the obsessive behaviors to call on attention from the network and to not miss any information will be the cost supported by society?

"Numerous studies indicate that almost half of the human species can be classified as insecurely attached or insecure with respect to attachment" (Ein-Dor et al., 2010, p.123). Insecure attachment is not a pathology in itself (although it can be a major risk factor of disorder) and it is often the result of insecure parental figure (O'Brien et al., 2012). This figure is unable to give to her child the comfort and the security that she has not. Thus, a vicious circle will be created and the insecure attachment style would be perpetuated.

Yet some children spend eight hours per day with technology which is more than the quadruple of the number of hours recommended by the American Academy of Pediatrics. How will this digital generation grow, live and what attachment style will it have? If anxiously attached people perceive FB as the community to which they belong (Rom & Alfasi, 2014) and which provides them with security, so, these online relations will be the external sources that regulate their affection. This attention provided will thus be a reinforcement and reward to continue using FB (learning theory of attachment). Overall, the disengagement practices will be then more difficult.

The Human/Non-Human Attachment: FB Platform in Loop

Users wrongly believe that FB has only some specific known and pre conceived functions to realize (Latour, 1999). In fact, FB has a clear and simple functioning. Its features enable users to form, sustain or interrupt relations easier and without cost. That is why the degree of ease of use perceived by users is found to be deterministic for using FB (Nasri & Charfeddine, 2012). This conceptualization of technology (here FB) as an exterior thing to humans, draws from the acceptance models such as the TAM (Technological Acceptance Model) or the UTAUT (Unified Theory of Acceptance and Use of Technology). Regarding these models, humans decide to use the technology through their perceptions, evaluations and attitudes. Other external factors such as gender, culture or experience are also supposed to influence the relationship between humans and technology. However, technology is not just a material apparatus constructed by humans' actor to satisfy their interest in a particular context and on which they have power. It is also a product of their action. From this point of view, there is a dynamic interaction between actor, content and context which emphasizes the definition of technology by the ANT as a social construct.

By using FB, the actors are connected to their platforms... and this latter connects them (Latour, 2000) in such a way that they become united with FB (Karppi, 2014): they cohabit, co-evolve, co-act and have a mutual dependence (Brangier & Hammes-Adel , 2011). So, the inherent dichotomy between object/subject and the technological determinism are thus rejected and replaced by an equal ontological footing with one another (Graham, 2005) and a high level of technosymbiosis (Brangier & Hammes-Adel , 2011). Thus, when studying the relationship between the user and FB, the question is not the ability of the user as independent subject to manage or to accept/reject the technology as object, but the nature and the quality of the link that exists between them. Hence, the aim is to determine whether their attachment is bad or good, how it can condition the user's action (Latour, 2000). Overall, how user "enacts" with FB.

Based on the cognitive dissonance, Festinger (1957) argues that, when doing an action, the actor tends to avoid any feeling of discomfort by changing his/her way of actions. For example, the disliked object when associated with the liked person/content becomes less disliked. Applying this principle of congruity, even if users dislike their use of smartphone or computer screen and aim to disengage, their attachment to what is shown, heard or read through this referential interface prohibits them to do so. The face-to-face is increasingly replaced by this active interface which becomes users' eyes, hearing and even by which they smell and touch things.

This role is also played by the FB software. In fact, FB communicates with users and replaces their physical absence by sending a heard and seen signal to inform about interaction. It also acts like an extension of the human memory when reminding them of the birthday of their friends or of their participation in events. Moreover, when users logs on to FB for the first time, an affective bond seems to be constructed in such a way that when they attempt to quit FB, an emotional request will be appeared showing the potential

loss of not only the online relations but the offline also (Karppi, 2014). Users' response will be then distracted and they will be worried about the implications of this request. These links, described by Licklider (1960) as a symbiosis, show a direct relationship and a mutual dependence between FB and users and underlines that FB is not just an object but a partner.

Nevertheless, this is the visible category of the iceberg. There are hidden facets that mask the users' "interactive loneliness" (Lipovetsky, 2006) and their "accepted isolation" (Guillaud, 2011). The molecular forces, which work behind the visible side, invite users, unconsciously, to interact (Sampson, 2012), to navigate only where friends or friends of friends had navigated (Guillaud, 2011) and to remain into a personalized "filter bubbles" (Pariser, 2011) in which pertinent information predicted by the EdgeRank algorithm from user's profile will be selected (Taffel, 2015). This algorithm masks the messages that are incongruent with the background or the political orientation of a given user, hides the status less liked by him/her and even deletes FB's accounts of revolutionary activists (as it was done in Tunisia when Ben Ali was President) (Lindenberg, 2011). So, it allows a preferential attachment in such a way that "the rich get richer and the good get better". Consequently, the information received by users is not objective nor it is completely randomized; and the social link that they, mistakenly, believe to maintain is selected and not independent. This way of FB commodification and algorithmic agencies could be described as being "anti-social" (Taffel, 2015), unethical and lack transparency (Pariser, 2011). Furthermore, FB transforms the users' profile, shares, like (even the past forgotten ones), friends and their amount of time spent connected on the platform into an economic opportunities by delivering information to the relevant commercial sites which personalize ads and shape user relation and behavior. Indeed, FB forces sometimes the users to accept the violation of certain laws which were conceived to protect their private sphere before allowing starting a service. In some, FB structure

orients the action to a preferential way and tries to guard this "merchandise" by manipulating emotion, behavior and past memories.

Overall, through the FB technical platforms, affective relations and strong ties with users are created. By the selection of who they should contact, view and share, FB gives the unaware users what they want to read, reinforces their self-presentation and meets their needs. This relation will make the disengagement difficult and a painful choice.

When FB is inaccessible (voluntary or mandatory), users could express distress/anxiety/awkward as if they will be separated or anticipated separation from an attachment figure. This negative feeling seems to be also expressed with the mobile phone, an interface that can allow FB connection (Lee and Katz, 2014).

Hence, if the hardware has contaminated all physical space, the software has passively affected the users' participation by analyzing their habits and their desire (Karppi, 2014). So, Software influences individual or collective behavior as do subjective norms and social influence. Based on this analogy, users seem to conform to the influence of FB platform as to the pressure of "normal" society to continue to be connected. To disengage will be thus regarded as abnormal behavior and will show an uncared feeling of others loss.

FUTURE RESEARCH DIRECTIONS

This investigation has FB as unit of focus. So, generalizing this analysis into other SNS is limited although some facets could be shared (Hargittai, 2008). Additionally, other variables could influence relations regarding FB use. While some studies indicated the relation between gender differences and parental education in FB use (Hargittai, 2008) and the attachment styles, others, have stated the influence of age on frequency of SNS use (Nitzburg & Farber, 2013) as well as differences across cultures (Nadkharni & Hofmann, 2012). Moreover, Stieger and al. (2013) show that people who disengage from FB are especially those who

are more worried about privacy, have an average age of 31 (those who still using Facebook has 24 years old), are men, have fewer friends and they are (slightly) more hooked on the web. Future studies could take into account these variables and test their relations with the human/human and human/non-human attachment. They could also initiate some FB disengagement practices in a voluntary or mandatory context and evaluate their relation with user attachment styles as well as the “translation” process with regard to the technology mediator. Besides, considering the impact of culture, it is possible to test the influence of people who are orally oriented (i.e: Arab and African countries) on the FB disengagement.

CONCLUSION

The aim of this analysis is not to determine whether users are independents from or slaves of FB, but rather, how attachment shapes the FB and the users, could enable or disable users’ actions and could enhance/constraint their rational/irrational behavior. Surely FB platform allows people to keep in touch with others and makes new relationships more easily accessible. However, the nature and quality of attachment that it maintains seem to transform the relation human/human and human/non-human, which requires more investigations and raises more awareness.

Our analysis shows that the attachment could be a major predictor of FB use. In fact, anxiously attached users are more inclined to find difficulty to disengage since they receive the attention and reassurance required from the online community. Besides, FB platform keeps users in touch thanks to its features but also its affective interaction. However, these relations are not equal. There is an exchange asymmetry about information at the expanse of users and a manipulation of their cognition, emotion and behavior. Since 1964, Ellul argued that moral considerations between technique and its use cannot exist in the new

social order. Even though, the users, unworried, continue to be attached.

Through disengagement, it would be possible to show the hidden facets of FB platforms and to select what attachment should be preserved and what should be removed or remodeled. Based on the rational choice theories, human actors have the ability to distinguish between the costs and the benefits of their action guided by their personal interest. Then, more awareness about the misuse of personal data by FB for commercial goals should enhance the degree of care and limit the degree of openness of single users.

It is delusional to believe that it will be a full disengagement (Jauréguiberry, 2014). But, some emergent disengagement practices could be applied (use of technical filters, rational click on like function...) which can be helpful in optimizing users/platform’ relationships and changing the rules of the FB game.

By reference to the learning theory, resistance to disengage results mainly from the user’s operant learning experience. If so, the nature of attachment could be relearned and the FB use re-appropriated.

REFERENCES

- Amichai-Hamburger, Y. (2002). Internet and Personality. *Computers in Human Behavior*, 18(1), 1–10. doi:10.1016/S0747-5632(01)00034-6
- Andangsari, E. W., Gumilar, I., & Godwin, R. (2013). Social networking sites use and psychological attachment need among Indonesian young adults population. *International Journal of Social Science Studies*, 1(2), 133–138. doi:10.11114/ijsss.v1i2.66
- Bauman, Z. (2010). *Hermeneutics and social science: Approaches to understanding*. Routledge.
- Bowlby, J. (1969). Attachment and loss: Vol. 1. *Attachment*. Basic Books.

- Brangier, E., & Hammes-Adel , S. (2011). Beyond the Technology Acceptance Model: Elements to Validate the Human-Technology Symbiosis Model. In M. M. Robertson (Ed.), *Ergonomics and Health Aspects* (pp. 13–21). Springer-Verlag. doi:10.1007/978-3-642-21716-6_2
- Dalsgaard, S. (2008). Facework on Facebook: The Presentation of Self in Virtual Life and Its Role in the US Elections. *Anthropology Today*, 24(6), 8–12. doi:10.1111/j.1467-8322.2008.00626.x
- Ein-Dor, T., Mikulincer, M., Doron, G., & Shaver, P. R. (2010). The attachment paradox: How can so many of us (the insecure ones) have no adaptive advantages? *Perspectives on Psychological Science*, 5(2), 123–141. doi:10.1177/1745691610362349 PMID:26162120
- Ellul, J. (1964). *The technological society*. Translated from the French by J. Wilkinson. Vintage books.
- Fenwick, T., & Edwards, R. (2010). *Actor-Network Theory in Education* (1st ed.). Tonbridge, UK: Routledge.
- Festinger, L. (1957). *A theory of cognitive dissonance*. Palo Alto, CA: Stanford University Press.
- Giddens, A. (2013). *The consequences of modernity*. John Wiley & Sons.
- Graham, H. (2005). *Guerrilla Metaphysics: Phenomenology and the Carpentry of Things*. Illinois, Peru: Open Court.
- Guillaud, H. (2011). Comprendre Facebook. *Futura-Science*. Retrieved January 15, 2016, from <http://www.futura-sciences.com/magazines/high-tech/infos/dossiers/d/informatiquecomprendre-facebook-1360/>
- Hansen, J. K. (2008). *Virtually connected: The role of attachment style in college students' use of online social networking* (Unpublished doctoral dissertation). The Wright Institute, Berkeley, CA.
- Hargittai, E. (2008). Whose space? Differences among users and non-users of social network sites. *Journal of Computer-Mediated Communication*, 13(1), 267–297.
- Hart, J., Nailling, E., Bizer, G. Y., & Collins, C. K. (2015). Attachment theory as a framework for explaining engagement with Facebook. *Personality and Individual Differences*, 77, 33–40. doi:10.1016/j.paid.2014.12.016
- Hazan, C., & Shaver, P. (1987). Romantic love conceptualized as an attachment process. *Journal of Personality and Social Psychology*, 52(3), 511–524. doi:10.1037/0022-3514.52.3.511 PMID:3572722
- Jaur guiberry, F. (2014). La d connexion aux technologies de communication. *Reseaux*, 186(4), 15–49.
- Karen, R. (1994). *Becoming Attached: First Relationships and how They Shape Our Capacity to Love*. New York: Oxford University Press.
- Karppi, T. (2014). *Disconnect.Me – User Engagement and Facebook* (Doctoral Dissertation). Department of Media Studies, University of Turku. Retrieved December 18, 2015, from <http://doria32-kk.lib.helsinki.fi/bitstream/handle/10024/95616/AnnalesB376Karppi.pdf?sequence=2>
- Latour, B. (1999). On recalling ANT. In J. Law & J. Hassard (Eds.), *Actor Network Theory and After*. Malden, MA: Blackwell Publishers. doi:10.1111/j.1467-954X.1999.tb03480.x
- Latour, B. (2000). Factures/fractures. De la notion de r seau   celle d'attachement. In A. Micoud & M. Peroni (Eds.), *Ce qui nous relie* (pp. 189–208). La Tour d'Aigues:  ditions de l'Aube.
- Laumer, S., Maier, C., & Weinert, C. (2013). The Negative Side of ICT-Enabled Communication: The Case Of Social Interaction Overload In Online Social Networks. *Proceedings of the ECIS*.

- Lee, S. K., & Katz, J. E. (2014). Disconnect: A case study of short-term voluntary mobile phone non-use. *First Monday*, 9(12).
- Licklider, J. C. R. (1960). Man-Computer Symbiosis. *IRE Transactions on Human Factors in Electronics*, HFE-1(1), 4–11. doi:10.1109/THFE2.1960.4503259
- Lindenberg, H. (2011). *Le réseau social, fer de lance de la révolution*. Retrieved February 26, 2016 from <http://www.lesinrocks.com/2011/02/14/web/le-reseau-social-fer-de-lance-de-la-revolution-1119967/>
- Lipovetsky, G. (2006). *Le Bonheur Paradoxal, essai sur la société d'hyperconsommation*. Paris: Gallimard.
- McKenna, K. Y. A., & Bargh, J. A. (2000). Plan 9 from cyberspace. The implications of the internet for personality and social psychology. *Personality and Social Psychology Review*, 5(1), 57–75. doi:10.1207/S15327957PSPR0401_6
- Nadkarni, A., & Hofmann, S. G. (2012). Why do people use Facebook? *Personality and Individual Differences*, 52(3), 243–249. doi:10.1016/j.paid.2011.11.007 PMID:22544987
- Nasri, W., & Charfeddine, L. (2012). An exploration of facebook.com adoption in Tunisia using Technology Acceptance Model (TAM) and Theory of Reasoned Action (TRA). *Interdisciplinary Journal of Contemporary Research in Business*, 4(5), 948–968.
- Nitzburg, G. C., Barry, A., & Farber, B. A. (2013). Putting Up Emotional (Facebook) Walls? Attachment Status and Emerging Adults Experiences of Social Networking Sites. *Journal of Clinical Psychology*, 69(11), 1183–1190. doi:10.1002/jclp.22045 PMID:23996299
- O'Brien, P. G., Kennedy, W. Z., & Ballard, K. A. (2012). *Psychiatric Mental Health Nursing: An Introduction to Theory and Practice* (2nd ed.). Jones & Bartlett Publishers.
- Oldmeadow, J. A., Quinn, S., & Kowert, R. (2013). Attachment style, social skills, and facebook use amongst adults. *Computers in Human Behavior*, 29(3), 1142–1149. doi:10.1016/j.chb.2012.10.006
- Pariser, E. (2011). *The filter bubble: What the internet is hiding from you*. Penguin.
- Patterson, M. (2015). *Social Media Demographics to Inform a Better Segmentation Strategy*. Retrieved January 20, 2016, from <http://sproutsocial.com/insights/new-social-media-demographics/>
- Rom, E., & Alfasi, Y. (2014). The Role of Adult Attachment Style in Online Social Network Affect, Cognition, and Behavior. *Journal of Psychology and Psychotherapy Research*, 1, 24–34.
- Sampson, T. (2012). *Virality. Contagion Theory in the Age of Networks*. London: University of Minnesota Press. doi:10.5749/minnesota/9780816670048.001.0001
- Sheldon, K. M., Abad, N., & Hirsch, C. (2011). A two-process view of Facebook use and relatedness need-satisfaction: Disconnection drives use, and connection rewards it. *Journal of Personality and Social Psychology*, 100(4), 766–775. doi:10.1037/a0022407 PMID:21280967
- Stieger, S., Burger, C., Bohn, M., & Martin Voracek, M. (2013). Who commits virtual identity suicide? Differences in privacy concerns, Internet addiction, and personality between Facebook users and quitters. *Cyberpsychology, Behavior, and Social Networking*, 16(9), 629–634. doi:10.1089/cyber.2012.0323 PMID:23374170
- Taffel, S. (2015). Anti-Social | Asocial | Associated: Mapping the Social in Social Media. *Global Media Journal*, 9(1).
- Van Dijck, J. (2013). *The Culture of Connectivity. A Critical History of Social Media*. Oxford, UK: Oxford University Press. doi:10.1093/acprof:oso/9780199970773.001.0001

Xu, H., & Tan, B. C. Y. (2012). Why do I keep checking Facebook: Effects of message characteristics on the formation of social network services addiction. *International Conference on Information Systems*.

ADDITIONAL READING

Alam, M. M. & Wagner, C. (2013). "Facebook Distress": A Model to investigate Discontinuation of Social Networking Site Use. *PACIS 2013 Proceedings*.

Anders, S. L., & Tucker, J. S. (2000). Adult attachment style, interpersonal communication competence, and social support. *Personal Relationships*, 7(4), 379–389. doi:10.1111/j.1475-6811.2000.tb00023.x

Boyd, D. M., & Ellison, N. B. (2008). Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication*, 13(1), 210–230. doi:10.1111/j.1083-6101.2007.00393.x

Breton, P., & Proulx, S. (1989). *L'explosion de la communication*. Paris: La Découverte.

Buote, V., Wood, E., & Pratt, M. (2009). Exploring similarities and differences between online and offline friendships: The role of attachment style. *Computers in Human Behavior*, 25(2), 560–567. doi:10.1016/j.chb.2008.12.022

Burt, R. (2000). Decay functions. *Social Networks*, 22(1), 1–28. doi:10.1016/S0378-8733(99)00015-5

Callon, M. & Ferrary, M. (2006). Les réseaux sociaux à l'aune de la théorie de l'acteur-réseau. *Sociologies pratiques*, 2(13), 37-44.

Deci, E. L., & Ryan, R. M. (2000). The what and why of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11(4), 227–268. doi:10.1207/S15327965PLI1104_01

Deuze, M. (2011). Media Life. *Media Culture & Society*, 33(1), 137–148. doi:10.1177/0163443710386518

Duck, S. (1988). *Relating to others*. Dorsey Press.

Hennion, A. (2013). *D'une sociologie de la médiation à une pragmatique des attachements*. Sociologies.

Jauréguiberry, F. (2003). *Les branchés du portable*. *Sociologie des usages*. Paris: PUF.

Jones, C. R., Ferreday, D., & Hodgson, V. (2008). Networked learning a relational approach: Weak and strong ties. *Journal of Computer Assisted Learning*, 24(2), 90–102. doi:10.1111/j.1365-2729.2007.00271.x

Katz, J. E. (2003). Do machines become us? In J. E. Katz (Ed.), *Machines that become us: The social context of personal communication technology* (pp. 15–25). New Brunswick, N.J.: Transaction.

Kraut, R., Kiesler, S., Boneva, B., Cummings, J., Helgeson, V., & Crawford, A. (2002). Internet paradox revisited. *The Journal of Social Issues*, 58(1), 49–74. doi:10.1111/1540-4560.00248

Langlois, G. (2013). Participatory Culture and the New Governance of Communication: The Paradox of Participatory Media. *Television & New Media*, 14(2), 91–105. doi:10.1177/1527476411433519

Latour, B. (2005). Third Source of Uncertainty: Objects too have agency. In *Reassembling the Social: An Introduction to Actor-Network-Theory* (pp. 63–86). New York: Oxford University Press.

Law, J. (1992). Notes on the Theory of the Actor Network: Ordering, Strategy and Heterogeneity. *Systems Practice*, 5(4), 379–393. doi:10.1007/BF01059830

Le Coadic, Y. F. (2004). *Usages et usagers de l'information*. Paris: ADBS.

Light, B. (2014). *Disconnecting with Social Networking Sites*. UK: Palgrave Macmillan. doi:10.1057/9781137022479

O'Keeffe, G. S., & Clarke-Pearson, K. (2011). The impact of social media on children, adolescents, and families. *Pediatrics*, 127(4), 800–804. doi:10.1542/peds.2011-0054 PMID:21444588

Portwood-Stacers, L. (2012). Media refusal and conspicuous non-consumption: The performative and political dimensions of Facebook abstention. *New Media & Society*, ■■■, 1–17.

Rouis, S. (2012). Impact of cognitive absorption on Facebook on students achievement. *Cyberpsychology, Behavior, and Social Networking*, 15(6), 296–303. doi:10.1089/cyber.2011.0390 PMID:22703035

Schäfer, M. T. (2011). *Bastard Culture! How User Participation Transforms Cultural Production*. Amsterdam: AUP. doi:10.5117/9789089642561

Selwyn, N. (2003). Apart from technology: Understanding peoples non-use of information and communication technologies in everyday life. *Technology in Society*, 25(1), 99–116. doi:10.1016/S0160-791X(02)00062-3

KEY TERMS AND DEFINITIONS

Actor Network Theory (ANT): It is an approach that treats an object as a social construct and assumes that it can act as the human in a network.

Anxiously Attached Person: It refers to the degree of concern exhibited by a person over others' evaluation and approval. It is manifested by an excessive need for attention and reassurance.

Attachment Theory: It presents different styles of attachment that have been built in the childhood period and transmitted to the adulthood. These styles can explain some aspects of the attachment to Facebook and the consequences of disengagement perceived as a separation from caregiver's figure.

Disengagement: It is a way of changing one's practices of connection and an approach of re-appropriation of technology/FB use.

Facebook: It is a SNS which allows a variety of free ways of communication for multiple users and use but with an asymmetry exchange of information and a manipulation of users' cognition, emotion and behavior.

Human and Non-Human Attachment: It is a relationship between human and material generated from the ANT. These two actors are considered to have an equal status until they enter into association within the network.

Social Network Sites (SNS): They are sites that provide possibility to interact with others but which may impact the interpersonal relationships, the human well-being and behavior.

Category S

Socio–Economic Development

Community Science and Technology and Its Meaning to Potential Requirement

S

P. K. Paul

Raiganj University, India

A. Bhumali

Raiganj University, India

INTRODUCTION

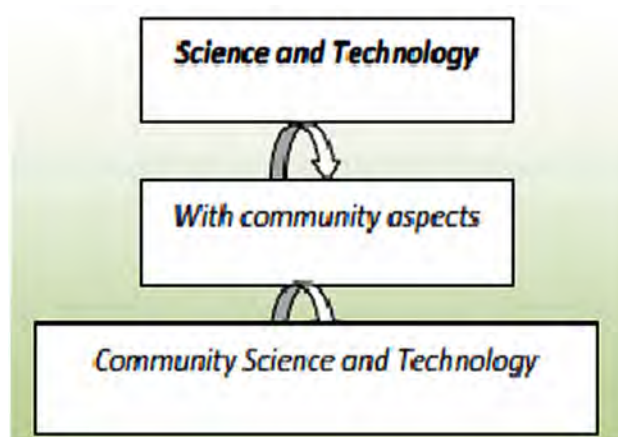
Community Science and Technology (CST) is an emerging concept as well as a broader domain that deals with various aspects of community and societal development. Community Science and Technology may be deemed as subject responsible and dedicated for community and society development. This domain talks about Science and Technology and its development for society and community. On other hand, Community Science and Technology (CST) may be expressed as a domain of Science, Technology, and Engineering responsible for social and community development. CST may come as a full-fledged education programme which should talk about community. On the other hand, it is a domain with focus of community and side by side other domains. Chapters mentioned are community science its meaning, role, values with special reference to emerging and possible domain. Community Science and Technology is actually an emerging concept and it's a domain which is directly and indirectly related to the community or society or development studies (or related subjects) (Avgerou, 2003; Paul et al, 2012a). Information Science is an interdisciplinary domain and responsible for information vis-à-vis technological solution with social science perspective. And thus the domain may be treated as a domain of Community or Social touch or more clearly, a domain of Community Science and Technology.

BACKGROUND

CST is a conceptual domain dedicated to societal development through proper science and technological tools (see Figure 1). Community Science and Technology may be brought as both, individual course of study or may be an individual subject with CST flavour. If it is a full-fledged CST programme then it may deal with several aspects of Community including its nature, role, diversity and side by side valuable for expressing the tools and techniques responsible for community or society and ultimately for its development today. Many aspects like virtual community, social network, and community development may also fall under the CST. Science and Technology are the main sources for the development whether it is industry, communication, education, governance and others. Such development and modernization of society is also possible with better and healthy Community Science and Technology related applications. The main background of this chapter includes but not limited:

- To find out about and learn on Community Science and Technology including its basic nature and characteristics.
- To learn and find out about Community Science and Technology as a stakeholder of development and modernization.
- To know about future potentiality of CST related programs as an academic specialization courses.

Figure 1. The basics of community science and technology



- To draw an exact picture and possibilities of CST program related to Community Science and Technology and its development.
- To find out Service Science and Community related aspects in Information Sciences and Technology.
- To design and development of Information Science related curriculum focused with Community Science and Technology gradients or specialization.
- To find out main problems and challenges of MSc- Information Science/ Information Sciences and Technology course with CST flavor.

This conceptual vis-à-vis policy study has been undertaken by the several tools and methods. Among others, review of literature is the important source of information. The web review, study of existing Community Science and Technology related programme has also played an important role to do this research work. It is a fact that material in the form of paper or chapter on Community Science and Technology is very limited. Hence web review has been undertaken to learn about the term and emerging concept. Role of specific subject to the community like computer in community, communication in community, energy

informatics and management in the community and others critically analyzed and thus huge literature has been consulted and indirectly used to prepare this study.

COMMUNITY SCIENCE AND TECHNOLOGY (CST): BASICS WITH POSSIBLE NOMENCLATURES AND FUTURE POTENTIALITIES

Community Science and Technology or CST is a growing concept and may be deemed as a broad discipline with interdisciplinary nature. It is actually a combination of many academic domains and fields. These fields may be from community, society, science, technology, engineering and some related facets. CST is dedicated or more clearly responsible for healthy application and utilization of science, technology and engineering fundamentals to the service sector, industries, commercial hub, governance, and education as directly or indirectly to the community and its development. In other sense Community Science and Technology may be defined as Applied Science domain having societal or humanities touch. CST in other way may be viewed as a domain for the community, of the community and by the community (Corneo G., Jeanne O, 1998 and P

K Paul & Sridevi, 2012b). Today many domains are indirectly connected with several aspects of Community or Society (see Figure 2). Thus CST is also there, indirectly.

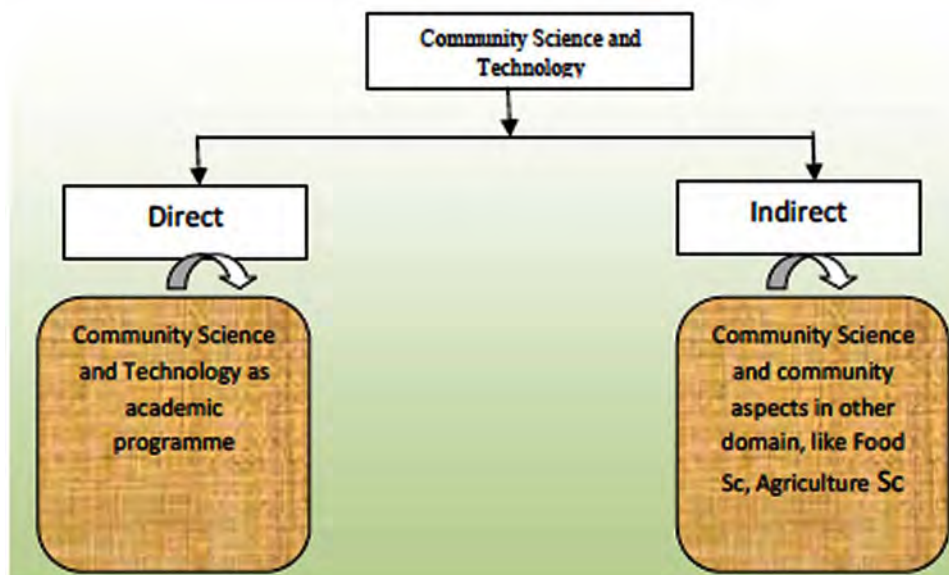
Importantly, Community Science and Technology (CST) is responsible for community and overall societal development as it deals with community aspects and direct utilization of science and technologies in the community. The frequent community development is no doubt responsible for sophisticated social development. Community Science and Technology (CST) may be treated as the domain which has direct or indirect relationship with community or development or comes as service science nature (Lucas R., 1988 and Paul, PK, 2012a). Agriculture, environmental science, fisherman, food and nutrition, horticulture— are indirectly acted as community development domain and thus these subjects may be treated as a Community Science (or Community Technology in the some cases). For example the domain of food and nutrition is a domain responsible for development of healthy nutritional infrastructure building and as it helps in better product building etc. Thus it builds nutritional utilization properly among the community or society (La Porta R,

1997) and in this sense it may be deemed as a Community Science and Technology.

Still Community Science and Technology is as full-fledged domain or an academic domain not much popular or available in the universities around the world. Though it is possible to design and develop Community Science and Technology domain with BSc/MSc or in other related program. The domain may be treated as Social Technologies or Social Engineering. The basic and possible nature and gradients of Community Science and Technology may be related on the following aspects:

- Community and its nature.
- Stakeholders of community development.
- Key Issues on Community.
- Development of community development.
- Community structure and formation.
- Diversity, Development and technologies.
- Stages of community and rural development.
- Social Psychology and Promotion.
- Business, Economics and Society.
- Rural Development and Allied Technologies.

Figure 2. Some community science existences



- Social Inclusion and Standardization.
- Stakeholders of community.
- Science for community and its allied branch.
- Technology for Community and Social development.
- Social Engineering and Community Management.
- Computing for society and some other domain.

However, Community Science and Technology (CST) may also bring with some other domain specialization. For example, food and nutrition science is very much related with community and society as people are the largest stakeholders of food domain. Thus here it is possible to offer Community Science and Technology (CST) touch or flavour in Food and Nutrition science and the program may be offered either as BSc/MSc-Community Science and Technology (Food and Nutrition science) or there may be potentially of offering BSc/MSc- Food and Nutrition science (Community Science and Technology). Here in this structure 'Food Science' as well as the 'Community Science' (or CST) may be included for wider benefit and advantage of the community. Ultimately, this type of program is helpful to use the Food as well as Nutrition to the door step of the community. Here academicians may be aware about various topics as well as some aspects of food and nutrition for the complete development. Their integration in CST is also an important issue and helps in relationship with community or society. Hence indirectly, for specialization of BSc/MSc- Food & Nutrition Science (Community Science) the above mentioned (bolded) aspects of community science may be included.,

These subjects may be (but not limited to the) Family welfare, Community Medicine, Community Nursing, Rural Development, Mass Communication, Micro Biology, Environmental Science, Sustainable Development, Communication Science, Ecology, Population Studies, Defense Studies, Himalayan Studies, Geo Information

Science, Strategic Studies, Food Science and Nutrition, Dietetics Studies, Agriculture, Horticulture, Public Administration, Information Science, Information Studies. Thus the nomenclature may be offered as follows:

- BSc/MSc— CST (Home Science).
- BSc/MSc— CST (Nutrition Science).
- BSc/MSc— CST (Food Science).
- BSc/MSc— CST (Textile Science).
- BSc/MSc— CST (Renewal Energy).
- BSc/MSc— CST (Agricultural).
- BSc/MSc— CST (Environmental Science).
- BSc/MSc— CST (Community Computing).
- BSc/MSc— CST (Social Computing).
- BSc/MSc— CST (Community Informatics).
- BSc/MSc— CST (Communication Science).
- BSc/MSc— CST (Information Science) and similar one.

In the field of Humanities the Community Science aspects may be included and may offer as an area of study or specialization like:

- BA/MA—CST (Public Administration).
- BA/MA—CST (Defense Studies).
- BA/MA—CST (Himalayan Studies).
- BA/MA—CST (Strategic Studies).
- BA/MA—CST (Sustainable Development).

Though apart from Science stream these programs are also possible to offer with Engineering and Technology domain but, as community is an interdisciplinary domain and this is related with some other aspects of community and humanities thus it is better to offer as Science stream (BSc/MSc). It is important to note that depending upon academic structure of the countries, option need to open-up for other degrees like- BS/MS/BE / ME/MTech/BTech. Some science characteristics is actually in built in Community Science and

Technology and this some other nomenclature are also possible to use like Social Computing, Community Informatics, Social Informatics, Communication Science, Information Science, Informatics and so on (Furuholt, B., & Kristiansen, S. 2007 and Thompson, M. P. A. 2005).

Though research program and project work funded by several agencies and in the degree programs may be an interesting tool for the development of Community Science and Technology and allied domain. It is valuable to note that though the nomenclature CST is comparatively new but the concept is old and already exists and thus research activities leading to Project Work/ PhD or higher are possible in this domain with existing setup.

COMMUNITY SCIENCE AND TECHNOLOGY (CST) AND ACADEMICS

Community Science and Technology (CST) may get the wider possibilities and benefit by universities and particularly from the related departments. Here the community aspects related departments may choose community knowledge gradients in their course curriculum for direct utilization of that particular program like if computing/ Information Technology program includes some community related aspects then social and community interaction is directly possible.

In health sciences, some domains have been arrived like community medicine, community nursing. Thus universities may establish separate departments / schools on Community Science and Technology (CST) or may include Community Science and Technology (CST) in some allied or related subjects/ domains. Research and development activities are possible in the areas of Community Science and Technology (CST). Even separate university may be established on Community Science and Technology (CST) with wide range of subjects having connection with

Community Science and Technology (CST). Thus full department may be established on Community Science and Technology (CST) and other domain as listed in the figure for creating better community based applied science domain (Paul, P.K., 2013a, c).

INFORMATION SCIENCES AND TECHNOLOGY (IST) VIS-À-VIS COMMUNITY SCIENCES & TECHNOLOGY (CST)

Service Sciences are the domain which are directly and indirectly helps in service offering and its promotion. Due to this hospitality nature, Information Science is also treated as a service science and in some cases service technologies. Information Science programs are available with several levels and out of these levels MSc-Information Science is considered as most valuable and popular around the world. Ultimately better information transformation leads economic development (Chattopadhyay, P., 2014 and Paul, P.K., 2012b).

Though the basic differences of Information Science with other domains like Information Technology, Computer Science, Computing Engineering is that Information Science is far more related with nature of Science and aspects on behavioral Information including its proper management and uses for the people or communities. Other hand, IT/Computing is mainly responsible and dedicated to the technological applications.

Community Science and Technology (CST) is a domain and a field which has relationship with the community and society related aspects. As an academic program the nomenclature of Community Science and Technology (CST) not gains popularity and thus very limited numbers of universities are offered Community Science and Technology (CST) as Bachelor/ Masters Degree (Paul, P.K., 2013b). In academic community some subjects gaining with Community Science atmosphere (or for the community or society) like:

- Home Science.
- Nutrition Science.
- Horticulture.
- Agriculture.
- Food Science and Technology.
- Textile Management.
- Rural Development.
- Renewal Energy.
- Agro Business Management.
- Urban Planning and Development and some other subjects.

These subjects are responsible for the community development as well as building and obviously societal progress in many ways.

Community Science mainly talks about development, modernization and up gradation of community and society through proper use and utilization of science and technological subjects. Community Science and Technology (CST) may also be treated as subject *of* the, *by* the and *for* the community.

CST as a course and department name available throughout the world. In India also, IEST, Shibpur has the unit –School of Community Science and Technology (SOCSAT) and some other university running programs on Community Science and Technology (CST) like- Community Medicine, Community Nursing, Community Informatics and so on. The MS University, Baroda offers a full-fledged Degree on Community Science (Corneo G., Jeanne O, 1998 and Paul, P.K., & Dangwal, K.L, 2012c).

CURRICULLUM OF INFORMATION SCIENCE: CST POTENTIALITY

Information Science is an advance field of study and practice responsible for several information activities such as information identification, collection and selection and their proper dissemination or delivery. The similar name of Information Science *with* Information Technology/ Computer

Science denotes that both are same domain but practically there are many differences between two. First of all Information Science is an Information domain whose main aim is restricted on how and whom information is needed to deliver for the common people or sector or industries by information systems designing. On other hand, Information Technology or Computer Science is responsible for design and development of computing systems. IS and its increasing interaction with the Technologies lead the newest nomenclature Information Science and Technology (IST).

The advancement of Information field changes Information Science many ways and several technologies as well as tools have been added which includes the Database Technology, Multimedia Technology, Communication and Networking Technology, Information Retrieval Systems, Web Technology and HCI in the IS knowledge stack. Thus the traditional information collection and dissemination systems has been changed and come as field with IT and Technology cluster. Still Information Science talks about information and its various natures like collection, selection, organization, nature and behaviours of information, information ethics and other facet and simultaneously on uses and applications of IT and Computing for better information practice.

Thus, Information Science is actually a science about information and its various facets. A basic M. Sc curriculum of Information Science in European and Indian style leads two year duration and in some cases it comes with specialization of Information Literacy, Information Management, Communication and Networking.

This chapter plays an attention on ‘Information Science and Technology’ and on ‘Community Science and Technology’ atmosphere. Now, let’s see about the curriculum and gradients of Information Science. A basic IST curriculum is actually integrates with computing and Information Studies, Management Studies, Information Technologies, Cognitive Sciences and some gradients /papers like:

- Information and Knowledge/ Information Society and Culture/ Information and Knowledge Management/ Knowledge Organization (Cataloguing)/ / Knowledge Organization (Classification)—as Information Studies related gradients.
- Database Management Systems/ Networking Technology/ Communication Technology/ Multimedia and Graphics Technology/ Information Retrieval Systems/ Search Engine—as computing and Technology related knowledge gradients.
- Management Basics/ Management Information Systems/ Organization and Business Informatics/ Information and Organization/ Knowledge Management—as Management related knowledge gradients;
- Cognitive Science/ Psychology and Information/ Human Computer Interaction/ Usability Engineering and related areas as Social Science related domains.

Hence, whether it has included Technological paper or Information, the Societal and Business related touch has been lying with IS/IST. Importantly, Information Science and Community Science based curriculum is proposed here as 2 years of duration with 4 semesters. Here each semester holds 6 papers of 4 credits each credit is actually education for 10 hours of duration (*see Table 1*).

Here all the semesters have been dealing with either technology or information centric paper but closely related with the society or community.

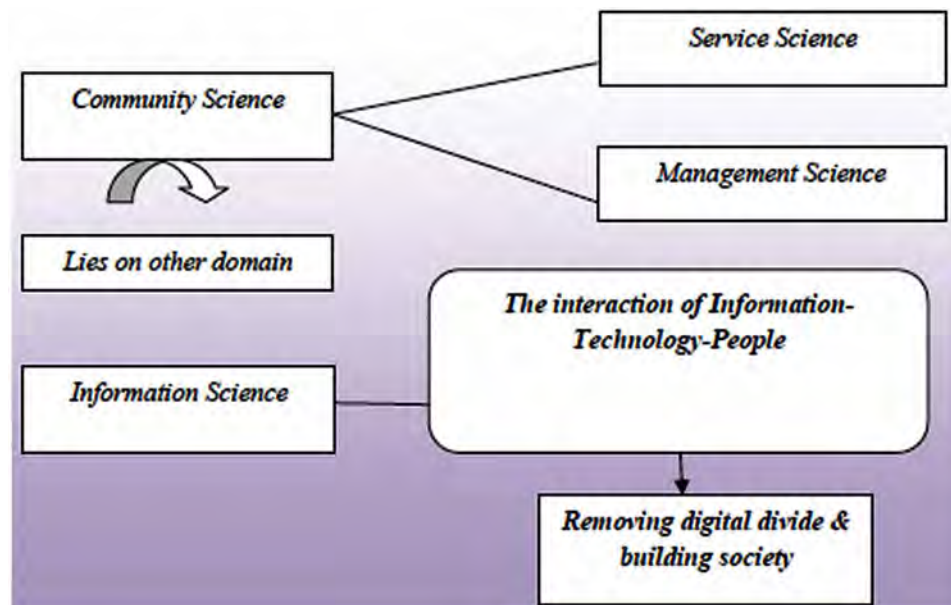
The proposed syllabus of MSc-Information Science (Community Science) presented here as Information Science and Communication Technological integration. Here gradients have presented based on aspects of information, computing and side by side community and society aspects. Community and Information Science integration may be treated as a new domain of Information Science or as a new flavour of Community Information Science.

This is mainly responsible for the design and development of Information Systems and Information infrastructure for the community need. Here directly information and technological uses on community related matter and field are provided with importance. It may be a healthy tool for societal and community development. In the third semester one internship program has been proposed due to its impact. Internship may be carried out in the technological settings, information dealing entities or social sector (such as NGO, Community health point, Telemedicine centre and so on). The fourth semester is proposed to undertake dissertation of around 18-20 credit. The dissertation may undertake as a project to solve a problem or may also undertake an Internship (Industrial/ Research) of around 180-200 hours at any approved and allied settings of the program interest.

Table 1. Showing the proposed structure of MSc-information science with community science flavor

Semester-I Information and Knowledge Information Society and Development Information Organization-I Computer Basics Community and Community Science Societal Development	Semester-II Knowledge Organization-II IT Application Networking, Information and Communication HCI and Usability Engineering Knowledge and Business Information Applications –Community and Industries
Semester-III Community Informatics- Development and Design Advance Computing and Community Informatics Societal Development, NGO and IT Project or Dissertation Internship Networking and Community	Semester-IV Dissertation (18-20 credit) Or Internship (Industrial/Research) Of 180-200 hours

Figure 3. Depicted information science and role of community science for its up-liftmen



CHALLENGES AND FINDINGS

Practically each and every activity has some challenges and as far as this topic and research is concerned the following points are important:

- CST still not recognized as an academic field or with degree program but got recognition as a concept and constitutes some interdisciplinary aspects like Nutrition, Community Medicine and others. And thus in academic world long destination is waiting for.
- CST is not a subject like physics or computer science it is a big interdisciplinary cluster and aspect. And thus building a real curriculum and healthy implementation and availability of HR is an important challenge.
- CST is an applied science and interdisciplinary in nature and it has wider connection in social and humanities domain and thus proper collaboration of Science department (such as for CST with Information Science need to collaborate

with Management Science) with humanities need to undertake.

- Information Science/IST has higher possibilities as community information nomenclature building domain and removing digital divide.
- The proposed curriculum is based on near about 65% of information and technological gradients and remaining are deals with community, society and development studies including business studies. Thus the seeker and providers both need interdisciplinary approach.

RESEARCH DIRECTION AND SUGGESTION

This chapter is the outcome of possible integration of science, technologies with humanities and management and thus following fact and step may be undertaken—

- Community Science is required to start as an academic program and there are possi-

bilities exist to offer Information Science as specialization. Though the vice versa or offering CST specialization is also possible in the IS/IST program (Paul, P.K., 2013c).

- CST may include in the course catalogue of universities for sophisticated development and utilization of education and academic program. Here, individual level of programs and courses may be designed in future for the betterment of technology and humanities or more clearly improving digital humanities.
- CST gradients need to include in other domains of interest and thus, there is a potentialities to find out and possible inclusion of the domain in future research or study.
- A center/ schools need to establish in each university and other institute of academics for interacting various departments.
- Workshop, seminars need to conduct on various aspects on social science and technology and for its further development.
- It is possible to offer Community Information Systems merging or collaborating some related departments like IT, Information Science, IST, Management with social science related department like social work, societal and rural development.

CONCLUSION

All about, 'community and we' and CST is responsible for better community and social development— directly and indirectly. The indirect relationship is useful with the electronics, agriculture or environmental sciences for the community or society. There is an urgent need for development of CST focused academic program. Or in other words utilization of these programs may bring educational, commercial, business, governmental development and development of these aspects ultimately can brings all round development.

Community and rural development play an important role for overall societal development. Information is treated as power and important gradients for overall information infrastructure and technological infrastructure development. The society many ways can grow with the integration of information and community aspects. Thus here several things are possible from academic and professional side for the development and modernization of CST as a discipline and Information Science as Computer Science field. In the developing countries still Digital Divide is an important and challenging issue and thus such problem may be minimized as much as possible by the possible integration of CST as a domain and CST flavour in other domain.

REFERENCES

- Avgerou, C. (2003). In M. Korpela, R. Montealegre, & A. Poulymenakou (Eds.), *The Link Between ICT and Economic Progress in the Discourse of Development* (pp. 373–386). Organizational Information Systems in the Context of Globalization.
- B.Sc Family & Community Science - University of Calicut. (n.d.). Retrieved on 23-10-2015 from the URL: www.universityofcalicut.info/syl/familyco.pdf
- Chattopadhyay, P. (2014). *Transforming Indian Economy into Knowledge Economy: The role of Human Resources with reference to India* (Unpublished thesis). NBU.
- Corneo, G., & Jeanne, O. (1998). Social Organization, Status, and Savings Behavior. *Journal of Public Economics*, 70(1), 37–51. doi:10.1016/S0047-2727(98)00059-0
- Economic Growth and Social Development. (n.d.). Retrieved on 11-04-2015 from the http://www2.dse.unibo.it/vanin/tesi_Pavia.pdf

Furuholt, B., & Kristiansen, S. (2007). A Rural-Urban Digital Divide? - Regional Aspects of Internet Use in Tanzania. *The Electronic Journal on Information Systems in Developing Countries*, 31, 1–15.

La Porta, R., Lopez-de-Silanes, F., Shleifer, A., & Vishny, R. (1997). Trust in Large Organizations. *The American Economic Review*, 87(2), 333–338.

Lucas, R. Jr. (1988). On the Mechanics of Economic Development. *Journal of Monetary Economics*, 22(1), 3–42. doi:10.1016/0304-3932(88)90168-7

Paul, P. K., Chatterjee, D., & Karn. (2012). Information Science Education and Research: emphasizing contemporary Indian scenario- an overview. *IEM/IEEE Sponsored International Conference Proceedings (IEMCON-12)*, 2(1), 349-353.

Paul, P.K. (2013a). Community Science and Technology (CST) From Meaning to possible Areas and Facet at a glance. *Abhinav National Journal of Arts and Education*, 2(3), 13-18.

Paul, P K. (2013b). Community Informatics and its Contemporary Conceptual, Methodological and Technical Issues with Special Reference to Social Issue. *Abhinav National Journal of Commerce and Management*, 4(2), 149-156.

Paul, P K (2013c). Social Computing and Social Informatics: The stakeholders of Knowledge Society emphasizing similarities and dissimilarities at a glance. *Abhinav National Journal of Science and Technology*, 2(4), 25-32.

Paul, P. K., & Sridevi, K. V. (2012). Information Science (IS) Education Challenges, Issues and Opportunities in Indian Context. *International Journal of Mathematics and Engineering with Computers*, 03(02), 87–93.

Paul, P.K., & Dangwal, K.L. (2012). Green Computing: Opportunities and Problems in the perspective of Developing Countries. *Asian Journal of Applied Science and Technology*, 1(4), 15-19.

School of Community Science and Technology (SOCSAT) - An Overview. (n.d.). Retrieved on 23-10-2015 from the URL: <http://www.iests.ac.in/academics/schools/socsat>

Thompson, M. P. A. (2005). ICT, Power, and Developmental Discourse: A Critical Analysis. *Electronic Journal on Information Systems in Developing Countries*, 20(4), 1–25.

ADDITIONAL READING

Chinman, M., Hannah, G., Wandersman, A., Ebener, P., Hunter, S. B., Imm, P., & Sheldon, J. (2005). Developing a community science research agenda for building community capacity for effective preventive interventions. *American Journal of Community Psychology*, 35(3-4), 143–157. doi:10.1007/s10464-005-3390-6 PMID:15909791

Lu, M. T. (2001). Digital divide in developing countries. *Journal of Global Information Technology Management*, 4(3), 1–4. doi:10.1080/1097198X.2001.10856304

Luke, D. A. (2005). Getting the big picture in community science: Methods that capture context. *American Journal of Community Psychology*, 35(3-4), 185–200. doi:10.1007/s10464-005-3397-z PMID:15909794

Paul, P. K., & Rajesh, R. (2015). MTech-information science: A new age proposed programme and its requirement with SWOT analysis in Indian educational context. *International Journal of Enterprise Network Management*, 6(4), 286–298. doi:10.1504/IJENM.2015.073871

Van Dijk, J. A. (2006). Digital divide research, achievements and shortcomings. *Poetics*, 34(4), 221–235. doi:10.1016/j.poetic.2006.05.004

Wandersman, A. (2003). Community Science: Bridging the Gap between Science and Practice with Community-Centered Models. *American Journal of Community Psychology*, 31(3-4), 227–242. doi:10.1023/A:1023954503247 PMID:12866681

Wersig, G., & Neveling, U. (1975). The phenomena of interest to information science. *The Information Scientist*, 9(4), 127-140.

KEY TERMS AND DEFINITIONS

BSc: It is a University Degree or Higher Degree. Normally is treated as graduate program in the European countries and other followers. Normally the program offers with 2-4 year of duration.

Community: It is actually a social unit of any size that shares common values; however he embodied or face-to-face communities are usually treated as small, larger or more extended communities such as a national community, international community and virtual community are also studied.

Community Science: It is a conceptual domain dedicated to societal development through proper science and technological tools. Community Science may bring as both, individual course of study or may be an individual subject with CST flavor.

Developing Countries: It is the countries which are economically and socially trying towards betterment by economic and social maintenances and proper policy implementation.

Digital Divide: It is the concept and aspect of disparities of Information Technology product uses among the common mass. More clearly, still in the society some are having access and condition of using several IT product and some are not able to use due to economical and social status.

India: The Republic of India is located in South Asia. Virtually it is the seventh largest country by area and most importantly second most popular

with 1.2 billion people. Many religions originated from here and some of them are Hinduism, Buddhism, Jainism and Sikhism.

Information: In general processed data is called Information. Though, clearly Information as a concept has a diversity of meanings, from everyday usage to technical settings. Generally speaking, the concept of information is closely related to notions of constraint, communication, control, data, form, instruction, knowledge, meaning, mental stimulus, pattern, perception, and representation.

Information Science: Information Science is actually a discipline of disciplines. It has come from the subjects and fields which are information, information services, information activities and digital information.

Interdisciplinary: Connection between some subjects irrespective of domain and direct and indirect connection with main facet or sub field. Such as Bio Chemistry, Information Sciences, Community Sciences.

Internet Community: It is a Group of people are complex, in ways that make those groups hard to form and hard to sustain; much of the shape of traditional institutions is a response to those difficulties. In the Internet Community, an online community builds weaker bonds and allows users to be anonymous.

MSc: It is a University Degree or Higher Degree. Normally is treated as graduate program of advance level in the European countries and other followers. Normally the program offers with 1-2 year of duration.

Science: It refers to this body of knowledge itself, of the type that can be rationally explained and reliably applied. In the West during the early modern period the term “Science” and “Natural Philosophy” were sometimes used interchangeably to refer to the study of natural phenomena.

Service Science: It is a domain of interdisciplinary nature dedicated to serve or more clearly services to the wide scale such as organization, institution, community and so on.

Technology: It is the collection of techniques, skills, methods as well as the processes which are used in the production of goods or services or in the accomplishment of objectives, such as scientific investigation. Technology term and concept can also be the knowledge of techniques, processes.

UGC: University Grants Commission is the main regulator of Universities of India, Including Central, State, Private and Deemed and other institutes which are listed and granted funds etc.

APPENDIX

S

Abbreviations

BE: Bachelor of Engineering.

BS: Bachelor of Science.

BSc: Bachelor of Science.

BTech: Bachelor of Technology.

CST: Community Science and Technology.

ME: Master of Engineering.

MS: Master of Science.

MSc: Master of Science.

MTech: Master of Technology.

SOCSAT: School of Community Science and Technology.

Financial Inclusion, Content, and Information Technologies in Latin America

Alberto Chong

Georgia State University, USA & Universidad del Pacifico, Peru

Cecilia de Mendoza

Ministry of Production of Argentina, Argentina

INTRODUCTION

Information and communication technologies are becoming widespread all over the world, changing the way people interact, consume, produce and learn. Personal computer, Internet, fixed, and mobile broadband penetration are all trending upward in Latin America, although the overall penetration still levels remains low (ITU, 2013a, 2013b). ICT have brought truly new and innovative possibilities to developing countries and Latin America has not been the exception. Here are just a few far-flung examples. In Honduras, farmers receive market price information via SMS (Pineda Burgos, Agüero Rodríguez, & Espinoza, 2010), in Colombia, coffee workers in rural areas can receive and make electronic payments (Chong, Galindo, & Pinzón, 2011); in Bolivia and Peru, individuals receive text messages to remind them to save (Karlan, McConnell, Mullainathan & Zinman, 2010); in Chile, financial literacy campaigns can be delivered with the help of information technology (Hastings, Mitchell, & Chyn, 2010); in Peru, researchers take advantage of the role of new media in order to increase savings among the poor (Chong & Valdivia, 2016); in Bolivia, public officials use computers in order to reduce discrimination (Chong & Yanez-Pagans, 2016).

Policymakers and academics agree that cyber-instruments can be beneficial for economic and social development, yet virtually no systematic or solid assessment exists on the mechanism and impact of ICT on the welfare of people. The

lack of adequate data to better understand basic problems in development economics has been a recurring problem for decades. A number of the interventions summarized in this chapter avoid this problem by relying on a set of field experiments—actual projects that were tested in the field—in several countries in Latin America and the Caribbean. They rely on randomized experiments, or randomized controlled trials (RCTs), which are increasingly used in formal empirical research in development economics. In a RCT, individuals or communities are randomly assigned to different “treatments” – different programs or different versions of the same program. Since the individuals assigned to treatments are exactly comparable (because they were chosen at random), any difference between them is the effect of the treatment. One of the reasons RCTs are becoming a widespread method of testing is that they help identify the specific variable that may have caused the particular result under investigation. RCTs help to disentangle a particular outcome from the various factors that may have caused it. Thus, the method can have many uses in policy-relevant applications.

With adequate policies in place, governments can take great advantage of cyber-instruments for development. But it’s important to keep in mind that greater access to these tools won’t bring about development on its own. Countries must also focus on how these tools are used. The trials presented in this chapter show that in order for cyber instruments to contribute to development it

is necessary to take into account equally crucial elements, such as human capital, the institutional context, and the policy goals of governments. In many cases, adopting the latest technology will not necessarily generate the best development outcome if, for example, the population lacks the basic literacy skills to use such technology properly. This chapter calls for countries to invest in infrastructure, regulation and human capital to improve their capacity to benefit from greater access to technology. And it recommends that governments systematically evaluate the impact of ICT on projects.

BACKGROUND

The earliest scholars studying the link between Information Technologies and development are Professors L. Roller and L. Waverman, who studied the overall impact of cyber-technologies on economic growth. Their most influential publication investigates how telecommunications infrastructure affects economic growth, finding a significant positive causal link, especially when a critical mass of telecommunications infrastructure is present (Roller & Waverman, 2001).

Current leading scholars have dealt with several interventions and cyber-behaviors in South Asia and Africa. To cite some of the most relevant work, Professor Robert Jensen (2007) showed that the adoption of mobile phones by fishermen and wholesalers in Kerala, a state in India with a large fishing industry, was associated with a dramatic reduction in price dispersion, the complete elimination of waste, and near-perfect adherence to the Law of One Price. Professors Banerjee, Esther Duflo and others (2007) evaluated a program in India in which fourth graders shared the use of software for two hours a week to improve learning in mathematics. The authors revealed that the program produced impressive results in learning, although the effect was temporary and most gains faded away a year after the program ended.

For the case of Latin America, the available evidence has been based mostly on anecdotal cases that describe success stories, but provide little solid empirical evidence on the link between cyber-behaviors and purported related gains in productivity and welfare. In evaluating the impact of information and communication technologies—in Latin America and the Caribbean and elsewhere—a critical problem has been the lack of reliable data that may allow the specific role of a particular ICT tool to be isolated. While some advances in data collection have been made in recent years, this central issue persists for the most part. More recently, Chong (2011) made one of the first efforts to provide systematic evidence on the role of ICTs on development outcomes in the region. In more recent years, several researchers have further studied the role of ICT on development outcomes and in particular, have paid attention to financial issues and in the fact that ICT as tools, need primarily good content to be effective.

CYBER INSTRUMENTS AND FINANCIAL INCLUSION

As financial sectors develop, financial inclusion increases and income inequality diminishes. Despite the benefits of establishing links with the financial system, very few households in the developing world use such links. On average, only 35 percent of Latin American and Caribbean households have a bank account—a very low percentage compared to advanced economies, where no less than 90 percent of the population has this type of link with the financial system (Demirgüç-Kunt, Beck, & Hohohan, 2008).

Households can enjoy access to financial services through several types of technological improvements. The development of the Internet has allowed many households across the world to exploit the benefits of online banking, such as paying utilities or transferring money. The possibility of banking through cellular phones is a

very popular and expanding avenue in the financial world. It can help diminish financial exclusion by offering services to low-income groups that have access to mobile telephones but not to financial services. ICT can also be a very useful tool to provide information to help people make better financial decisions.

Mobile finance has been boosted by the strong growth of mobile phones throughout developed and developing countries. Worldwide, people have more cell phones than bank accounts, and Latin America and the Caribbean are no exception (World Bank 2010). There are many reasons why people do not have bank accounts. Having a telephone will not change important underlying factors determining that decision—but having a cellular phone can solve one issue that limits financial inclusion: limitations in the brick-and-mortar infrastructure of a financial institution. In many places, reaching the poorest is a geographical challenge. Imagine a coffee farmer who lives high in the Andean mountains in Colombia. Despite producing one of the most valued commodities in the world, this farmer would find it very difficult to be linked to the financial system. There is little incentive for the coffee grower to have a bank account, and there is even less incentive for a bank to open an office in a remote area for a handful of coffee farmers. Technology can break this constraint and align the interests of both parties.

Using mobile banking services could vastly improve this farmer's lifestyle. He probably has a cell phone, since Colombia's coffee industry has joined the growing trend of merging cell phones with financial services. In 2009 a pilot to provide financial services to the coffee-growing community was promoted by the National Federation of Coffee Growers of Colombia, the Inter-American Development Bank, cell-phone provider Telefónica, and the Banco de Bogotá, a large Colombian bank. The mechanism used is a very straight forward one in the mobile banking community.

The coffee growers sell their coffee at the nearest cooperative. The proceeds of the sale

are deposited in a bank account that each coffee grower has opened at the bank—an account that they can fully manage through their cell phones. Coffee growers can purchase supplies and production inputs at local facilities by accessing their bank account with their cell phones. This not only eliminates the risks of managing cash, but significantly reduces transaction costs by saving coffee growers a trip to the nearest town to purchase goods and services.

REMINDERS INCREASE SAVINGS

Cell phones can be used not only to provide financial services, but also to encourage certain types of behavior, particularly to help people exercise greater financial discipline and self-control. An example is saving. People usually promise themselves that they will set aside money until they achieve a certain goal. However, these promises are often broken because people postpone savings to the future, turn a blind eye to future needs, or choose current over future consumption.

To date, Latin America and the Caribbean has been moving relatively slowly to exploit technology to foster financial inclusion. Nonetheless, many pilot projects have taken off and researchers in the region are beginning to take stock of the lessons learned (Chong, Galindo, & Pinzón, 2011). A recent study by Dean Karlan, Margaret McConnell, Sendhil Mullainathan and Jonathan Zinman explores the impact of reminding deposit holders about commitments on how much money they want to save and how often they should make deposits. An experiment was carried out in Bolivia, Peru, and the Philippines in which specific banks sent messages to their clients to remind them about their savings commitments. The reminders have a positive and significant impact on savings, compared to clients who also made saving commitments but did not receive the reminders. On average, the people that received the reminders achieved their deposit goals 3 percent more often and were able to save 6 percent more.

Karlan et al show that reminders change intertemporal allocations and improve consumer welfare by linking future expenditure opportunities to today's choices. In the authors' words, they mitigate "attentional failure:" that is, inattention to future consumption opportunities.

The experiment conducted by Karlan and his team varied slightly from country to country. In each experiment, individuals opened a bank savings account that included varying degrees of incentives or commitment features designed to encourage them to reach a savings goal. Some individuals were randomly assigned to receive a monthly reminder via text message or letter, while a control group received no reminder.

In Peru, the researchers worked with a state-owned bank, Caja de Ica, and offered a new product called Plan Ahorro to clients. The clients selected a time frame to achieve a certain savings goal (6 or 12 months) and a minimum deposit to make each month. If the client reached a goal, he or she was rewarded with an 8 percent interest rate, twice the regular interest rate of the savings product offered. A fraction of clients were selected to receive reminders. Some reminders included the goal and others did not. The wording changed in some reminders to emphasize a gain from savings or a loss from failing to save. Late reminders were also sent. Cell phone coverage in the area was low, so the bank sent the reminders by mail. Two additional treatments were tested. One randomly sent a letter to some people in the reminder group that focused on their particular goal (in addition to containing the boilerplate reminder). Another treatment independently and randomly offered clients a gift upon opening the account: a jigsaw puzzle of their goal, a photo of their goal, or a pen. Those in the jigsaw puzzle group received a piece of the puzzle after each deposit.

In Bolivia, the researchers worked with a private bank, Ecofuturo, and sent text messages to clients of the Ecoaguinaldo program. This savings strategy aims to encourage people to deposit a certain amount of money every month; by the end of the year, they would have saved

the equivalent of one month's wage, the amount formal workers receive at the end of the year by law. When signing up for the program, clients determined how much they would save every month. If clients made all their deposits within the first ten months, they would be rewarded with a 6 percent interest rate (twice as high as they would receive otherwise) and a free life and accident insurance policy. Clients were randomly assigned to receive a text message. Some messages included reminders about the benefits, while others did not.

Using reminders proved useful not only in increasing savings, but also in increasing the likelihood that clients in Bolivia would meet their target amount (3 percent). Not only did the reminders have an effect, but the way in which the message was phrased mattered, as well. Reminders that highlighted the client's particular goal, such as school fees, were twice as effective as reminders that did not mention the goal. These findings are novel empirical evidence about the importance of reminders in household finance. This innovative use of modern technology to influence savings decisions proved effective, particularly given the low cost of cellular phone texting. The advantage of using text message reminders over mailings lies mainly in its cost advantage in terms of both money and labor. Comparing the results of the experiments in Bolivia and Peru, the researchers found that the bank receives a profit of about US\$0.20 per client with text messages, compared to a loss of US\$2.32 per client with mailings.

EXPANDING FINANCIAL LITERACY

Another tool to provide useful information to help clients make better financial decisions is information technologies. Hastings, Mitchell, and Chyn (2010) conducted a randomized experiment in Chile to determine whether providing workers with information on the variations in fees charged for pension funds made a difference in which fund workers finally chose.

In the experiment, the researchers used an on-site spreadsheet that computed the expected balance of each respondent's pension fund based on past returns and fund commissions, and each individual's wage, balance and age. The balances were computed for each pension fund, and shown to the interviewees in different ways. For one group, the returns were shown in comparison to the highest cost fund (highlighting the relative loss). For the second group, the returns were compared to the cheapest fund (highlighting the relative benefit).

The experiment found that people with lower levels of education, income, and financial literacy rely more on the advice of employers, friends, and coworkers than on fundamentals in choosing a fund. Further, the researchers found that such individuals are more responsive to the way the information is presented (information framing) when interpreting the relative benefits of different investment choices.

The Chilean experiment shows that financial literacy campaigns can be delivered with the help of information technology. However, the possibility of extrapolating these results to other contexts is limited. The use of information technology as a tool of financial education must take into account local market conditions and the likelihood that the ICT can be adequately adopted.

THE IMPORTANCE OF CONTENT

Researchers recognize that whereas ICT may provide a valuable tool to achieve financial inclusion and improve development outcomes in Latin America, they have become increasingly aware on the role of content that can be delivered using new and traditional ICT tools. In this context, recent research shows the role of soap operas as a potentially very useful mechanism that if paired with ICT tools can provide significant development results.

The key reasons of why soap operas work are summarized by La Ferrara (2016) and La Ferrara,

et al (2014). These and related work provide compelling evidence that creating characters, who are fully identifiable with viewers in terms of race, mannerisms, language, household characteristics, and the like can effectively help provide 'teaching moments' to viewers in a dramatized context. This tool has proven very effective in the case of financial inclusion (Chong & Valdivia, 2016) as well as in other relevant development outcomes frequently related with gender differences. In fact, a crucial objective of soap operas is to empower women through direct actions (e.g., savings on her name) and indirect actions (e.g., conversations with other women, and her partner) in order to reshape the current social construct. This approach is sometimes accompanied with the use of knowledgeable facilitators that conduct discussions with the viewers after each soap episode is shown as this further helps internalize the teaching moments as well as give them a forum to discuss.

The reason for the success of soap operas can be traced to three aspects, which summarize the theory of change. First, soaps are set up in easily recognizable locations and deal with the daily life of people, so that viewers can relate to the story. Second, soaps use colloquial language, identifiable setups, and often include an element of independence and social mobility that is highly appealing to viewers. Third, TV companies spare no expense in order to produce soaps of the highest technical and artistic quality, which makes them more attractive. The theory of change poses that soap operas play a central role for the functioning and process of articulating new behavioral patterns as it shows to give a sense of belonging and identification. It can have effects on values and attitudes, among other things, a reorientation of beliefs on the role of women in society toward greater autonomy and empowerment.

Soap operas in Latin America have shown promise in order to deal with difficult issues related to behavioral economics. There is recent academic evidence that shows that television and soap operas can have important social effects. Relevant work are Chong and La Ferrara (2009)

and La Ferrara, Chong, and Duryea (2012) who show that soap operas have an impact on divorce and fertility rates in Brazil. In addition, Chong and Yanez-Pagans (2015) provide evidence that television may impact negatively homicides in Brazil. Also, Dahl and DellaVigna (2009), focus on the short-run effects of movie violence on crime. In other disciplines, such as psychology Paluck (2009) estimates the effects on beliefs and norms of a radio soap opera featuring messages of intergroup tolerance in Rwanda. Furthermore, in Political Science Gentzkow and Shapiro (2008) have studied the effects of television on social as opposed to political behavior, in particular children's school performance and Olken (2009) has focused on adult participation in social activities and trust. Finally, in Public Health, Jensen and Oster (2009) estimate the impact that the entry of cable TV had on subjective measures of female autonomy, school enrollment and fertility.

The potential for scale of this type of innovations in the context of ICT is extremely high. Once the soap opera is produced, the marginal cost of showing it elsewhere is essentially zero, which means that the costs of massively broadcasting a soap opera as would be close zero. Furthermore with some caveats, it might be possible to scale it up to different Latin American countries, in particular, those that share similar cultural, racial, and historical traits. An example of this area commonality would be the South American Andes.

A LONG WAY TO GO

In spite of the enormous benefits associated with these new tools up to now, Latin America and the Caribbean have lagged well behind the rest of the developing world. There still is a long way to go before the poorest households in Latin America will be able to afford a computer and Internet service of their own. To date, Latin America and the Caribbean has been moving relatively slowly to exploit technology to foster financial inclusion.

Nonetheless, many pilot projects have taken off with the aim of achieving what Asian and African countries accomplished years ago—with the advantage of being able to take stock of the lessons learned.

ESCAPING POVERTY THROUGH TECHNOLOGY

Living in poverty is not only about not having enough money to purchase valuable goods and services; it is also about lacking access to the resources to break out of poverty and guarantee well-being. Access to information—through something as simple as a phone call—can enable the poor to use their own knowledge and strengths to escape poverty traps.

Information and communication technologies are a potentially effective weapon for fighting poverty as they can help people improve their incomes. The role of ICT in improving social and economic conditions in the world's poorest countries has been strongly advanced by the international community and many examples worldwide can be found, including in Latin America. Chong and de Mendoza (2011) documented a number of interventions in Latin American countries where the use of cyber-instruments contributed to earning opportunities for the population.

Furthermore, ICT have shown that they can be very effective in reducing discrimination in particular, towards the poor. Taking advantage of two randomized natural field experiments occurring in the context of the renewal of national identification cards in Bolivia, Chong and Yanez-Pagans (2016) provide evidence that computer technologies can be a very useful tool to significantly reduce discrimination by reducing discretion and by increasing implicit accountability. In fact, computers and related technologies have the potential to transform the way in which governments interact with citizens as well as on the manner in which services are delivered to the

public. They may be of important help in countless situations where the power and discretion of economic agents is unbalanced.

These researchers show that the introduction of computer technologies leads to impressive improvements in efficiency and productivity, and provide strong evidence that the introduction of these tools can lead to a drastic reduction in discrimination, something that is impressive given the significant differential service provision to individuals in their field experiment. Another important consequence of the introduction of computer technologies is that they also help reduce corruption in the public sector. In fact, given that discrimination and corruption appear to go hand in hand, this finding is of relevance as public service delivery in many countries is rife with corruption and as it is reasonable to believe that such link is not exclusive to their study. As these technologies continue to evolve, policy makers and regulators may increasingly adopt them not only as a tool to enhance government efficiency and transparency, but also as a tool to achieve more equitable societal outcomes

In general, there are compelling reasons to expect significant economic development from the adoption of ICT. One way in which these technologies can help bring about economic improvement is by reducing disparities and gaps in information (asymmetric and imperfect information) in markets. For example, individuals and firms can use ICT to research prices of products, look for jobs, or locate potential buyers of their products. This may be especially relevant for certain groups that are at a disadvantage with respect to others, either because they lack information or because their access to it is delayed or in a lower quality form. Low-cost access to useful information can help low-income families find economic and social opportunities that could benefit them. ICT can become a powerful tool to alleviate poverty, reducing uncertainty, promoting inclusion, and increasing income.

Asymmetric information and search costs are particularly detrimental in two specific markets:

agriculture and labor. In the agricultural sector, small producers usually sell their products knowing only prices in their local area. Cell phones give producers access to prices in other markets and open their eyes to the relative worth of their produce. Expanded mobile phone service, together with needed complementary public services (gathering, processing, and disseminating information in real time by agriculture and production ministries) give poor farmers the fodder they need to negotiate higher prices with intermediaries and wholesalers.

ICT have also proven their worth in the performance of labor markets. Job search activities require time and resources, and may lead workers to many dead ends. ICT can help reduce job search time and costs and therefore impact positively on unemployment rates.

FUTURE RESEARCH DIRECTIONS

Information and communication technologies are a potentially effective weapon for fighting poverty. They can help close information gaps, spread the knowledge available to a broader sector of the population and expand the opportunity frontier of the most vulnerable sectors. Armed with more and better information, the poor can be better equipped to act on their own behalf, boost their earning potential and use their own knowledge and abilities to lift themselves out of poverty. Furthermore, taking advantage on the anonymity of cyber-transactions, in the future these tools should focus on how they can help in reducing discrimination and corruption, something that the poor suffer disproportionately.

CONCLUSION

This chapter summarizes a number of specific cyber-interventions in the areas of finance, content, and poverty-reduction that have taken place in Latin America in recent years, which may lead to a

better understanding of the impact of interventions to know what works, what does not, and why. The common thread of all the existing cyber-behavior research in Latin America is that tools that take advantage of new information and communication technologies are not the panacea that policymakers want them to be, but with adequate content such tools may be very useful, as early evidence shows in the case of financial inclusion.

REFERENCES

- Banerjee, A., & Duflo, E. (2011). *Pooreconomics: A Radical Rethinking of the Way to Fight Global Poverty*. New York: Public Affairs.
- Chong, A. (Ed.). (2011). *Development Connections: Unveiling the Impact of New Information Technologies*. New York: Macmillan. doi:10.1057/9780230118379
- Chong, A., & de Mendoza, C. (2011). Development.com: Using ICT to Escape Poverty. In A. Chong (Ed.), *Development Connections: Unveiling the Impact of New Information Technologies* (pp. 69–104). New York: Macmillan. doi:10.1057/9780230118379_3
- Chong, A., Galindo, M., & Pinzon, M. (2011). Banking on Technology for Financial Inclusion. In A. Chong (Ed.), *Development Connections: Unveiling the Impact of New Information Technologies* (pp. 69–104). New York: Macmillan. doi:10.1057/9780230118379_3
- Chong, A., & La Ferrara, E. (2009). Television and Divorce: Evidence from Brazilian Novelas. *Journal of the European Economic Association*, 7(2-3), 458–468. doi:10.1162/JEEA.2009.7.2-3.458
- Chong, A., & Valdivia, M. (2016). *Soap Operas and Financial Inclusion in Rural Peru*, Manuscript. Georgia State University.
- Chong, A., & Yanez-Pagans, M. (2015). Television and Homicides in Brazil. *The Journal of Development Studies*.
- Chong, A., & Yanez-Pagans, M. (2016). *Do Computers help Reduce Discrimination?* Manuscript. Georgia State University.
- Dahl, G., & DellaVigna, S. (2009). Does Movie Violence Increase Violent Crime? *The Quarterly Journal of Economics*, 124(2), 677–734. doi:10.1162/qjec.2009.124.2.677
- Demirguc-Kunt, A., Beck, T., & Honohan, P. (2008). *Finance for all: Policies and Pitfalls in Expanding Access*. Washington, DC: The World Bank.
- Gentzkow, M., & Shapiro, J. (2008). Preschool television viewing and adolescent test scores: Historical evidence from the Coleman study. *The Quarterly Journal of Economics*, 123(1), 279–323. doi:10.1162/qjec.2008.123.1.279
- Hastings, J., Mitchell, O., & Chyn, E. (2010). *Fees, Framing, and Financial Literacy in the Choice of Pension Manager*. Pension Research Council Working Paper WP2010-09. The Wharton School, University of Pennsylvania, Philadelphia, PA.
- ITU. (2013a). *World Telecommunication/ICT Indicators Database*. Geneva: International Telecommunication Union.
- ITU. (2013b). *Information and Communication Statistics Online*. Available at: www.itu.int/ITU-D/ICTEYE/Indicators/Indicators.aspx#
- Jensen, R. (2007). The Digital Divide: Information (Technology), Market Performance, and Welfare in the South Indian Fisheries Sector. *The Quarterly Journal of Economics*, 112(3), 879–924. doi:10.1162/qjec.122.3.879
- Jensen, R., & Oster, E. (2009). The Power of TV: Cable Television and Womens Status in India. *The Quarterly Journal of Economics*, 124(3), 1057–1094. doi:10.1162/qjec.2009.124.3.1057

Karlan, D., McConnell, M., Mullainathan, S., & Zinman, J. (2010). *Getting to the Top of the Mind: How Reminders Increase Savings*. Unpublished document. New Haven, CT: Yale University. doi:10.3386/w16205

La Ferrara, E. A. C., & Duryea, S. (2012). Soap Operas and Fertility: Evidence From Brazil. *American Economic Journal. Applied Economics*, 4(4), 1–31. doi:10.1257/app.4.4.1

Olken, B. (2009). Do Television and Radio Destroy Social Capital?: Evidence from Indonesian Villages. *American Economic Journal. Applied Economics*, 1(4), 1–33. doi:10.1257/app.1.4.1

Paluck, E. L. (2009). Reducing intergroup prejudice and conflict using the media: A field experiment in Rwanda. *Journal of Personality and Social Psychology*, 96(3), 574–587. doi:10.1037/a0011989 PMID:19254104

Röller, H. L., & Waverman, L. (2001). Telecommunications infrastructure and economic development: A simultaneous approach. *The American Economic Review*, 91(4), 909–923. doi:10.1257/aer.91.4.909

World Bank. (2010). *World Development Indicators*. Available at <http://databank.worldbank.org/ddp/home.do?Step=12&id=4&CNO=2>

ADDITIONAL READING

Banerjee, A., & Duflo, E. (2011). *Pooreconomics: A Radical Rethinking of the Way to Fight Global Poverty*. New York: Public Affairs.

Chong, A. (Ed.). (2011). *Development Connections: Unveiling the Impact of New Information Technologies*. New York: Macmillan. doi:10.1057/9780230118379

Jensen, R. (2007). The Digital Divide: Information (Technology), Market Performance, and Welfare in the South Indian Fisheries Sector. *The Quarterly Journal of Economics*, 112(3), 879–924. doi:10.1162/qjec.122.3.879

KEY TERMS AND DEFINITIONS

Development Economics: Branch of Economics that deals with aspects related to the development process in countries.

Financial Inclusion: Financial operations and tools that help increase availability and affordability to low-income segments of society.

Information and Communication Technologies: Tools that help manage information electronically. They may be catalogued as “traditional”, as in the case of television and radio and “modern” as in the case of smartphones.

Soap Operas: Serial drama that deals with the lives of characters in a dramatic and emotional manner.

The Growing Impact of ICT on Development in Africa

S

Sherif H. Kamel

The American University in Cairo, Egypt

INTRODUCTION

Information and communication technology for development (ICT4D) has been evolving for almost four decades realizing business and socioeconomic development and growth and capitalizing on channels for information acquisition and knowledge dissemination allowing the creation of an emerging global knowledge-based society with innovative means of communication that can help increase competitiveness for individuals, organizations and societies. ICT tools and applications could be used as an effective platform to minimize the digital divide and promote social inclusion. The transformation process enabled through ICT is affecting different aspects of the economy and is forcing governments to rethink the way they handle their socioeconomic challenges as they move forward in their development paths. This article addresses some of the initiatives in Africa to minimize the intra and inter digital divide, and join actively the global knowledge society by capitalizing on emerging technologies but more importantly by focusing on human capital as the most invaluable element in the technology ecosystem. Many African nations have invested heavily in building their ICT infrastructure with its different building blocks to help improve the readiness of their communities and smoothly integrate into the global digital world¹.

BACKGROUND

During the recent few decades, ICT tools and applications have transformed the way people live,

work, study, travel, and get entertained amongst other daily routines. The global ICT sector is currently a significant sector in its own right accounting for up to 7.5% of worldwide gross domestic product. ICT has not only changed the world, but it has also increased its potential (Figueres, 2003). In addition, with more challenges facing the world today, more opportunities are being created for ICT to have an effective role. On a different note, emerging and innovative ICT coupled with globalization and the role of societal norms, values and cultures are constantly affecting societies around the world. It is forcing organizations both private and public to rethink and reengineer the way they manage their operations and resources and face competition both locally and globally. Moreover, it is fair to claim that the processes of globalization are increasingly depending on ICT (Musa, 2006). This situation has generated new forms and structures of economic, business and social organizations that are no longer affected by geographic or time constraints but depend mainly on teleworking, telecommuting and overcoming barriers of time and distance which is emerging as the platform for business and socioeconomic development in the 21st century. The digital platform is increasingly becoming at the center stage of most businesses and industries. ICT has the potential, given proper infrastructure in place, trained skilled human resources and timely infrastructure, to improve the balance in economic and social progress, increase growth of the economy, boost the capacity to face societal challenges, enhance progress of democratic values and augment cultural creativity, traditions and identities (Kamel, 2009).

DOI: 10.4018/978-1-5225-2255-3.ch628

During the 1990s, there was an unprecedented link between the technological innovation process and the economic and social organizations. Moreover, as the links between economic development, productivity and the availability of information resources became invaluable, governments around the world started to invest heavily in building their national information infrastructure (Petrazzini & Harindranath, 1997). This led to major changes and transformations in the activities and relationships of individuals and organizations within the society, leading to the evolution of the information society, where the services provided by ICT represent a set of challenges and opportunities for the global society. However, it is important to note that, although access to ICT is a prerequisite to its use, individual differences in time and space as well as capabilities and choice may play a role on the use, value and application of ICT (Alampay et al, 2003).

In today's global environment, it is becoming easier within the context of business and socio-economic development and growth to identify and evaluate as well as compete across all sectors and industries using the wealth of information and knowledge that are disseminated through global information networks. The ICT infrastructure makes information more accessible with more benefits to the society (Shapiro & Varian, 1999), which puts more pressure on firms around the world to exploit all possible opportunities to leverage productivity and efficiency. Businesses are increasingly becoming more aware of the vitality of ICT to stay competitive, with other global implications to productivity, employment and profits to the extent that organizational operations are becoming unthinkable without the effective and efficient use of ICT, especially in a global society, where information travels across national boundaries (Branscomb, 1994). Moreover, societies, especially in the emerging world, are increasingly encouraging high-value innovative entrepreneurs in the ICT space who can seize the opportunities in the marketplace and offer solutions to various economic sectors (Strategy&, 2011). Therefore,

many developing nations including African nations have taken concrete measures in restructuring their ICT policies as part of their overall national strategies for development. This has resulted in having the number of mobile phones subscriptions growing twice as fast as that of the world during the period 2003-2008. Today, mobility, Internet dissemination and the growing diffusion of digital applications and tools are taking the African continent by storm, a leading example is M-Pesa Kenya's renowned mobile payment service, channeling 25% of the country's gross domestic product (GDP) showing how important mobile payment is becoming integral to economic transactions (Hossain, 2015).

The policies introduced included deregulation, encouraging private investment and foreign direct investment (FDI), and the use of tools such as public private partnerships. ICT is an opportunity for the developing world such as the African continent because it is a powerful tool for economic growth, social inclusion and poverty eradication, which can facilitate the integration of African nations into the emerging, digital global marketplace (Annan, 2003). Africa stands to gain a great deal from participating in the globally connected economy; however, it must first establish the necessary ICT infrastructure, and restructure government processed and avail suitable economic conditions to attract and maintain an effective position in the global economy (Ajayi, 2004; Akinola et al, 2012). One decade into the 21st century, the world is becoming smaller and the public is rapidly gaining access to emerging ICTs (Shapiro & Varian, 1999). The information society is becoming a global force and a fundamental element of change in the global society (Garito, 1996). In that respect, Africa is taking clear and robust steps to build its information, technology and communication infrastructure.

Investment in the ICT sector is perceived to have the ability to improve the lives of people with low income who have limited access to services such as healthcare and education (Qureshi, 2007). Moreover, ICT holds the promise of development

and growth by connecting people across the world irrespective of their background, location or cultural setting to more accurate and timely sources of information and knowledge (Ahmed, 2007a) and Africa should be at the center stage of that development given that still in the majority of African countries, less than 5% of the population use the Internet resulting in the fact that most African countries have an Internet penetration equal to or below 11% as compared to the global average of 23%. In recent years, that has resulted in having the percentage of growth of Internet and mobility usage among the highest in the world as the African nations attempt to close the digital divide with the rest of the world.

The Role of ICT in Development in Africa

Since the early years of the 21st century, there has been a high-level commitment from African leaders to bring about change in the way ICT is perceived as part of the Africa's infrastructure boom. The commitments were declared following the meetings of the World Summit on the Information Society (WSIS) in Geneva in 2003 and Tunis in 2005. Today, African leaders are more into the information age and they are more prepared than ever to initiate new ideas, formulate a vision, set the strategy and support it throughout the implementation phases. It is perceived that such a trend represents a unique opportunity for Africa's younger and growing generations to adapt and adopt new tools and techniques using state-of-the-art ICT.

The indicators demonstrate that the African ICT market is gradually rising with huge potentials to have positive implications on different economies. Given the size of the African continent, there is a huge demand for ICT that needs to be satisfied effectively in order to ensure effective participation of African nations in the global information society. In today's growing and fast changing marketplace, it is broadly believed that universal access to ICT through broadband and

access can contribute to economic development supported by governments, the private sector and the civil society. However, despite the developments in the ICT sector in Africa over the last few years, there are still a number of challenges to overcome before the continent is fully connected. The focus in the coming period should be mainly on the underprivileged communities and the rural areas. Moreover, ICT prices remain high and are unaffordable by the majority of the population who live in one of the poorest regions in the world.

While teledensity in Africa and especially in Sub-Saharan Africa is rather low, teledensity in rural areas is very limited. The gap between urban and rural areas in terms of telecommunication access is significant because the majority of the population is living in rural areas and underprivileged settings. It is estimated that in Sub-Saharan Africa nearly 66% of the population resided in rural areas. Within the context of Sub-Saharan Africa, there are around 400,000 localities, 99% are villages. Less than 3% of these have a fixed telephone line connection and less than 4% of the villages have access to Internet in public access centers. Accordingly, it is estimated that 55% of village population are not connected without access to fixed, mobile and/or data services, all resulting in negative competitive edge. Moreover, education, awareness and training are vital platforms for ICT to play an effective role in development. It is a two-way process. While education can help diffuse ICT, emerging technologies could also help spread awareness, education and knowledge. ICTs has the potential to increase access to education, improve teaching and learning and improve the efficiency of educational management and dissemination with the growing distance and blended learning environments. ICT could be the ideal platform for Africa to reach out to all remote locations and help develop a large and growing human capital expected to reach 2 billion people by 2050.

In the context of the developing world, the number of cellular M2M connections known as Internet of things (IoT) will reach a total of 128 million or 52 percent of the global connections

Table 1. M2M connections by region

Region	M2M % Total Connections (2013)	M2M CAGR (2010- 2013)	Connections CAGR (2010-2013)
Africa	1.0%	41.3%	15.0%
Asia	2.1%	55.0%	10.4%
Europe	5.1%	28.6%	2.4%
Latin America	2.1%	43.7%	7.8%
North America	9.3%	22.5%	3.6%
Oceania	5.1%	25.8%	5.5%
Global	2.8%	37.6%	8.8%

Source: GSMA, 2014

during 2015 and will rise to 575 million connections or 60 percent by 2020 (Reuters, 2014). Table 1 demonstrates the M2M connections by region.

African Information Society Initiative (AISI)

The information infrastructure is a strategic resource to be used by governments, corporations, and individuals to rationalize their decision-making process. The conference of ministers of the Economic Commission for Africa (ECA) in May 1995 adopted the Africa Information Society Initiative (AISI). The idea of AISI was to focus on policies, strategies, initiatives and projects that can help establish nation-based information societies that can help create the African information society. AISI is a common vision for Africa's quest to bridge the digital divide (UNECA, 2003a). One of the key components of AISI is the formulation of national eStrategies, also called National Information and Communication Infrastructure (NICI) plans with a view to achieve their development goals through the collaboration of key stakeholders from the government, private sector and the civil society (UNECA, 2004a; UNECA, 2004b). Three major development goals have been articulated by African leaders who represent the driving force behind the embarkation on ICT evolution across

the continent. They include improving the quality of life for every African, working on the integration of the economies of the different African nations and leveraging trade linkages with other regions outside Africa based on mutual development purposes and growth targets. The development of AISI enables African leaders, decision makers and planners to position Africa in the world's rapidly expanding global economic system and accelerate the pursuit of the continent's development goals. The emerging ICTs offer the potential to create jobs at much lower levels of capital investment and exploit Africa information resources without the need for corresponding financial wealth. However, the institutionalization of the process is an integral part of its success, sustainability and scalability.

ICT IN AFRICA

Emerging ICTs are perceived to be the possible driver for socioeconomic development in Africa. There is no doubt that African lives have greatly improved over the last decade with more promises for development and improvement in the coming decade. However, despite the success of the ICT penetration including mobiles and Internet over the last decade, there is limited proof of ICT contribution to development and growth. One of the reasons could be attributed to the lack of ICT market reforms that is essential to integrate ICT within the economic structure of nations, let alone the uneven balance and divide across different African regions. Both areas have seen concrete improvement over the last few years but more needs to be done moving forward.

Competition, although available and growing over the last few years, yet it is still limited and not as effective to introduce major changes and transformations. Therefore, there is additional work that needs to be done in terms of market regulations to ensure competitiveness, quality and service delivery. This is not to undermine the role of other stakeholders such as the private sector or the civil society but they all compliment and

are highly needed for ICT to have an effective role in development and growth. The regulatory environment will also address the issues of equality and inclusion within the information society. Additionally, a regulated and open environment is conducive to investments both local and through foreign channels such as foreign direct investment (FDI). There are different models in place whereas in the case of South Africa, they began using government investments for ICT infrastructure development, however, in the case of Uganda, the government entered into long-term agreements with the private sector to expand networks that are operating more efficiently within a competitive environment (Gillwald and Stork, 2008). There are a number of aspects that represents the ICT ecosystem and that need to be addressed to assess the evolution of ICT in Africa and that includes; (a) liberalization of telecommunication, (b) regulatory environment, (c) market segments, (d) revenues and investments, and (e) employment. These elements represent the building blocks based upon which ICT can effectively contribute to the continent's economic development.

In terms of ICT infrastructure, different means of communication are now widely available. In terms of statistics, Africa has three mobiles for every four people which matches the same penetration level as in India. Moreover, the bulk of mobile machine-to-machine (M2M) connections are in developing countries, including Africa. In Africa, there were 7 million mobile M2M connections in 2014 and the number can quadruple by 2020 whereas most Africans access the Internet from mobile services (Reuters, 2014). The proliferation of M2M started in the developed world but it is an area where the developing world has overtaken the developed world, and the African continent is taking the lead. The Internet of things initiatives and programs are mainly focusing on vehicle tracking, mobile payments (mPayments) and smart cities (Reuters, 2014). Today, across the African continent, capacities are being developed to connect everything to the Internet; moving forward is a world where every rural village,

urban city and community in Africa has wireless access (Presterio, 2014). Examples include South Africa, the continent's most advanced economy is the leader in rolling out M2M devices, one of the projects being smart meters for a power utility and Rwanda using SIM cards to connect point of sales (POS) terminals in remote locations enabling merchants to accept debit or credit card payments (Presterio, 2014).

Youth as the Opportunity

With a young population spread across the African continent, opportunities are being created around the clock. Youth including students are all over the continent increasingly using computers and tablets from home, schools, universities and Internet Cafés. Consequently, African statistics suggest that human development in sub-Saharan Africa has made huge leaps and there's an overall improvement in their skills and capacities (The Economist, 2013). For example, secondary school enrolment grew by 48% during the period 2000 and 2008. According to UNCTAD (2012); there has been an overall economic boom that in many ways made all the difference. Africa is the world's fastest growing continent. While over the last decade per capita income increased by around 30%, the previous 20 years had seen a low down in the continent's economy by 10% (The Economist, 2013). The opportunity is there for Africa to leap-frog and introduce, implement and institutionalize innovative technology applications.

Liberalization of Telecommunications

Historically, the role for government with respect to ICT in Africa has been that of an owner and operator of the incumbent public telecommunication operator (Touré, 2007). However, there has been transformation in the policy direction in recent years whereas governments in Africa are currently promoting competitiveness by establishing a sound policy framework and stable institutions.

Liberalization of the telecommunications sector has been a growing trend over the last couple of decades with a number of African governments committing to open up their domestic telecommunication market and introduce competition with over 50% of all Africa's fixed line markets are now subject to competition and nearly 50% have private-sector participation in their ownership structure. Consequently, the role of the private sector is increasing which effectively is healthier for the industry and can help create an environment that is more competitive, efficient and customer-oriented.

Regulatory Environment

UNECA has placed priority on building an information-rich continent through AISI with its framework that is geared towards developing NICI plans to complement liberalization and deregulation measures whereas nations are judged whether they are information rich or information poor (Ahmed, 2007b). Respectively, over 35 nations have been engaged in defining national policies and electronic strategies (eStrategies) as part of their socioeconomic development process. The impact of the liberalization process and the effectiveness of regulatory policies constitute a key indicator for measuring the effects of the information society in Africa. The rationale for the introduction of deregulatory measures in African telecommunication sectors relates to the economic principle of competition that could increase access and lead to the expansion of the ICT infrastructure. The measures were also geared towards stimulating incumbent national telecommunication entities to provide services in a more efficient and effective manner. It is important to note that in 1993, there was no country with a competitive market environment for telecommunication. Today, example nations that established organizations to regulate the sector include Morocco, which established the National Telecommunication Regulatory Agency in 1998 and Egypt, which established the National Telecommunication Regulatory Authority in 2003.

Market Segments

The market segments of ICT with an emphasis on telecommunications are divided into fixed lines, mobile lines, Internet penetration and broadband access. Most of the fixed telephone lines are concentrated in 6 of the 54 African economies. Algeria, Egypt, Morocco, Nigeria, South Africa and Tunisia account for almost 80% of all fixed lines in Africa. However, most these lines are located in major cities with some random diffusion in rural communities. Africa's fixed telephone line penetration was 3.1 per 100 inhabitants compared to a world average of 19.4 per 100 inhabitants, more than 6 times higher than the penetration rate of Africa. The continent has taken the lead in the shift from fixed to mobile telephony. Therefore, the mobile market in Africa is characterized for its impressive growth rates. North and Sub-Saharan Africa were able to add the largest number of new subscribers. Based on the above that indicates that mobile penetration and growth was more than fixed lines, Africa is the only continent where more revenues were generated from mobile services than from fixed line services showing the prospects of a full-fledged ICT infrastructure. It is important to note that low subscription prices, prepaid services and low recharging cards have contributed to the African mobile boom. Some African countries are approaching full mobile population coverage; Egypt reached that status in 2012 which was coupled with reduction in prices due to increasing competition.

In 2006, less than 3% of the world's Internet subscribers (10.7 million) were located in Africa, which means given the African population, this means that 1.3% of the inhabitants were subscribers to an Internet Service Provider (ISP). The estimated number of effective Internet users' stands at 4.8 users per 100 inhabitants. Affordability, lack of fixed line infrastructure and a low level of ICT literacy are the most striking reasons for the low levels of Internet use in Africa. However, it is important to note that African nations vary greatly in their Internet penetration rate ranging

from the penetration of 35.7 per 100 inhabitants in the Seychelles to 11 in South Africa and 8 in Egypt. The broadband access in Africa serves a limited number of inhabitants. However, with the decrease in prices of both voice and data the volume of bandwidth users is on the rise. The diffusion of bandwidth access is on the agenda of different African governments to be able to capitalize on the infrastructure of different electronic business applications. Moreover, bandwidth could be a determining factor in boosting investments and availing opportunities for offshore outsourcing (Touré, 2007). The deployment of universal broadband access could help create cost-effective and flexible standards for wireless technologies, extend coverage to geographically challenging locations and provide public service electronic applications. Africa's best performers when it comes to the state of the web are Tunisia, Egypt, Kenya and South Africa.

Employment and Job Creation

The telecommunications sector is a platform of income generation at the individual, organizational and governmental levels. The liberalization process and increasing competition helped expand the sector and made available more employment opportunities in the sector. The challenge of the sector is to avail regularly qualified human resources acquainted with the current and timely developments in telecommunications services and technologies. The drive of African leaders is based on a number of elements including the announcement of the G8 summit in 2000 that nations who will succeed in harnessing ICT potentials can look forward to leapfrogging conventional obstacles of infrastructure development whereas everyone should be able to access information and communication networks. The deployment of emerging ICT provides a unique opportunity to developing nations to capitalize on previous experiences and lessons learnt by various developed nations.

Challenges Create Opportunities

S

The formulation of the information society requires the collaboration of multiple stakeholders in the community including the roles of (a) government, (b) private sector, (c) civil society and (d) media. This can help transforming existing challenges into opportunities. The role of government should reflect an enabling environment to develop the national ICT infrastructure and to ensure that all sectors of the economy can benefit from it. Thus, each African government should assign a national entity to be responsible for the coordination between different stakeholders involved in the buildup of its NII. Moreover, African governments should promote the use of ICT to improve decision-making effectiveness. They should design, develop and implement national policies and plans adopting ICT within various sectors. The role of the private sector should be to stimulate growth and assume market leadership in developing national ICT infrastructure through investing in relevant areas and seizing new business opportunities that arise from the implementation of AISI. The role of the civil society should represent the catalyst coordinating with other stakeholders in establishing AII through identifying the needs of rural communities, contributing in human capacities development, reflecting public concerns and needs in terms of ICT and promoting the concerns of the workforce to employers and government. The role of the media is spreading awareness in Africa of the importance and benefits of the information revolution. Newspapers, radio, and television provide an easy, accessible, and cheap means of diffusing information to the end user.

CONCLUSION

The role of ICT for development is growing worldwide and associated policies and strategies are regularly amended to pave the way for an effective

ecosystem that can help ICT realize its targeted objectives. ICT is helping to avail the information society through ICT penetration across the community both urban and rural, which indicates the necessity to minimize the digital divide, work to socially, and economically include the underprivileged communities to become electronically ready. The information society will only be realized when there is a mature ICT infrastructure at a reasonable cost and with ICT penetration at both individual and organizational levels. The African Information Society is the least developed globally reflecting the most primitive information infrastructure indicating one of the least socioeconomically developed regions, despite the waves of change introduced over the last couple of decades in capitalizing on ICT and enabling ecosystem. The transition to an information-based society is so massive and comprehensive that, from a policy perspective, no nation can proceed on all fronts at the same time and it will all depend on the national priorities. African nations must develop their own infostructure with a view to a future connection to the global information infrastructure to bridge the gap with the developed world and to establish links for cooperation and exchange of experiences, skills, capacities and knowledge.

In that respect, ICT is causing an industrial and societal evolution based on information acquisition and knowledge dissemination and allowing the creation of an emerging information society with innovative means of communication that could help increase competitiveness for individuals, organizations and societies. To conclude, there is huge potential through the deployment of the Internet of things in developing economies during the next decade where the estimation reflects higher potential value in advanced economies because of higher value per use yet the developing world can benefit from around 40 percent increase of value generated. Another proof that emerging ICT platforms could effectively help African economies to grow. It will all depend on the business, industry and application and the readiness of both the technology

infrastructure and the timeliness, accuracy and access of the infostructure. Accordingly, the level of impact will vary dramatically. Examples for that include the health sector where healthcare spending in advanced economies is twice that in developing economies and in the retail sector, while higher adoption and values exist in advanced economies, there are large numbers of retail settings in developing markets. The potential is there and the opportunity could be created through the proper deployment of ICT, the key element is to have the clear set of objective and capitalize on the continent's most precious resource, people.

REFERENCES

- Ahmed, A. (2007a). Open access towards bridging the digital divide – policies and strategies for developing countries. *Journal of Information Technology for Development*, 13(4), 337–361. doi:10.1002/itdj.20067
- Ahmed, A. (2007b). Technological transformation and sustainability in the MENA region. In A. Ahmed (Ed.), *Science, Technology and Sustainability on the Middle East and North Africa* (pp. 3–18). Brighton, UK: Inderscience Enterprises Ltd.
- Ajayi, S. (2004). What's in it for us, how Africa can make good on globalisation? *Convergence*, 3(3), 66–68.
- Akinola, O. S., Akinkunmi, B. O., & Alo, T. S. (2012). A data mining model for predicting computer programming proficiency of computer science undergraduate students. *African Journal of Computing and ICT*, 5(1), 43–52.
- Alampay, E., Soliva, P., Justimbaste, L., & Tenedero, C. (2003). *Evaluating the impact of universal access models, strategies and policies in ICTs on poor communities in the Philippines: final report*. National College of Public Administration and Governance Forum, Manila.

- Annan, K. A. (2003). Information and communications technologies: a priority for Africa's development. In J. O. Okpaku (Ed.), *Information and Communications Technologies for African Development: An Assessment of Progress and the Challenges Ahead* (pp. xv–xvii). New York: United Nations ICT Task Force Publications.
- Branscomb, A. (1994). *Who owns information?* New York: Basic Books.
- Figueres, J. M. (2003). Information and communications technologies for African development. In J. O. Okpaku (Ed.), *Information and Communications Technologies for African Development: An Assessment of Progress and the Challenges Ahead*. New York: United Nations ICT Task Force Publications.
- Garito, M. (1996). The creation of the Euro-Mediterranean information society. *Proceedings of the European Union Meeting on the Creation of the Information Society*.
- Gillwald, A., & Stork, C. (2008). *Towards evidence-based ICT policy and regulation*. Retrieved 10 January 2009, from <http://www.researchictafrica.net>
- GSMA. (2014). *Understanding the Internet of things*. GSM Association.
- Hossain, M. (2015). M2M in developing countries. *Proceedings of the 7th CMI conference on mobile communications developing countries*.
- Kamel, S. (2009). Building the African information society. *International Journal of Technology Management*, 45(1/2), 62–81. doi:10.1504/IJTM.2009.021520
- Musa, P. F. (2006). Making a case for modifying the technology acceptance model to account for limited accessibility in developing countries. *Information Technology for Development*, 12(3), 213–224. doi:10.1002/itdj.20043
- Petrizzini, B., & Harindranath, G. (1997). Information infrastructure initiatives in emerging economies: the case of India. In B. Kahin & E. Wilson (Eds.), *The National Information Infrastructure Initiatives*. Cambridge, MA: Massachusetts Institute of Technology Press.
- Prester, T. (2014). Emerging markets and the Internet of things: why devices designed for high-income countries don't work and what to do about it. *Proceedings of the Solid Conference*.
- Qureshi, S. (2007). Information technology innovations for development. *Information Technology for Development*, 13(4), 311–313. doi:10.1002/itdj.20080
- Reuters. (2014). *Africa fast off blocks in adopting Internet of things – industry group*. Reuters Industries. Retrieved 30 August 2015, <http://www.reuters.com/article/2014/11/06/africa-tech-idUSL6N0SW5O920141106>
- Shapiro, C., & Varian, H. (1999). *Information rules*. Boston: Harvard Business School Press.
- Strategy. (2011). *Accelerating entrepreneurship in the Middle East and North Africa*. Retrieved 20 November 2015, from http://www.strategyand.pwc.com/me/home/press_media/management_consulting_press_releases/article/49943176
- The Economist. (2013). *Africa rising a hopeful continent*. Retrieved 10 November 2015, from <http://www.economist.com/node/21572377/>
- Touré, H. (2007). *Competitiveness and information and communication technologies (ICTs) in Africa, African Competitiveness Report*. World Bank Publications.
- United Nations Conference on Trade and Development. (2012). *Foreign direct investment into Africa*. Retrieved 15 November 2015, <http://www.unctad.org>

United Nations Economic Commission for Africa. (2003a). *E-Strategies national, sectoral and regional ICT policies, plans and strategies*. The 3rd Meeting of the Committee on Development Information, Addis Ababa.

United Nations Economic Commission for Africa. (2004a). *Report to the 10th SRO-SA Meeting of ICE*. National Information and Communication Infrastructure (NICI) Plans.

United Nations Economic Commission for Africa. (2004b). *Report to the 7th SRO-WA ICE Meeting*. National Information and Communication Infrastructure (NICI) Plans.

ADDITIONAL READING

Bahta, A. (2005). Emerging features of the African information economy: ICTs in the national budgets of African Countries. *4th Meeting of the Committee on Development Information*, Addis Ababa.

Cogburn, D. L. (1995). Background to Africa's information society initiative (AIS). *African Regional Symposium on Telematics for Development*, April.

Dutta, S., & Mia, I. (2009). The global information technology report – mobility in a networked world. Geneva, World Economic Forum.

Hudson, H. E. (2001). The potential of ICTs for development: opportunities and obstacles. *Background Paper for the World Employment Report*.

Kamel, S. (1998). Building an information highway. *Proceedings of the 31st Hawaii International Conference on System Sciences (HICSS)*, Hawaii, United States, January 6-9, pp. 31-41.

Kamel, S. (1999). *Building the African information infrastructure*. *Journal of Scientific and Industrial Research, Special Issue on Management of Information Technology Organizations and Beyond* (pp. 118–144). New Delhi: National Institute of Science Communication, March-April.

Lance, K., & Bassole, A. (2006). SDI and national information and communication infrastructure (NICI) integration in Africa. *Information Technology for Development*, 12(4), 333–338. doi:10.1002/itdj.20051

Mbarika, V. W. A. (2002). Re-thinking information and communications technology policy focus on Internet versus teledensity diffusion for Africa's least developed countries. *Electronic Journal of Information Systems in Developing Countries*, 9(11), 1–13.

McNamara, K. (2008). Building local capacity for ICT policy and regulation: a needs assessment and gap analysis for Africa, the Caribbean, and the Pacific, *Summary Report*, InfoDev, Working Paper No. 16.

Okpaku, J. (2003a). Background on information and communications technologies for development in Africa. In J. O. Okpaku (Ed.), *Information and Communication Technologies for African Development – An Assessment of Progress and Challenges Ahead* (pp. 23–46). New York: United Nations ICT Task Force Publications.

Okpaku, J. (2003b). Information and communications technologies as a tool for African self-development: towards a re-definition of development. In J. O. Okpaku (Ed.), *Information and Communication Technologies for African Development – An Assessment of Progress and Challenges Ahead* (pp. 1–22). New York: United Nations ICT Task Force Publications.

United Nations Economic Commission for Africa (2003b). Policies and plans on the information society: status and impact. October.

KEY TERMS AND DEFINITIONS

Africa: Referring to the African continent.

Developing Economy: Describing the economies of developing or emerging nations.

Digital Divide: Describing the gap between the haves and have nots in different countries when it comes to information and information technology access and usage.

Information Highway: Describing the digital platform enriched by repositories of information and knowledge that is universally accessible by the different stakeholders in society.

Information Society: Describing the type of society that is information-based that is supported by full-fledged information infrastructure (infostructure).

Internet of Things: The growing development of machine interconnectivity where every day more objects and machines have network connectivity, allowing them to send and receive data between each other.

IT for Socioeconomic Development: Describing the role of information and communication technology in the social and economic development process.

IT Transfer to Developing Nations: Describing the challenges and opportunities that emerge when transferring information and communication technology to developing and emerging nations.

Social Inclusion: Describing the efforts that are exerted to close the gap between the different social segments in the society with respect to social and economic aspects.

ENDNOTE

- ¹ This article is an update of previous publications by Sherif Kamel, “Building the African Information Society”, *International Journal of Technology Management*, Volume 45, Number 1/2, 2009, ISSN 1741-5276, pp. 62-81; and “The Evolution of the ICT Industry in Egypt”, in A. Ahmed (Ed.), *Science, Technology and Sustainability in the Middle East and North Africa*, Brighton: Inderscience Enterprises Limited, 2007, ISBN 0-907776-30-2, pp. 65-79.

Manufacturing vs. Services and the Role of Information Technology

Arnab Adhikari

Indian Institute of Management Ranchi, India

Shromona Ganguly

Indian Institute of Management Calcutta, India

INTRODUCTION

The role of technology in the context of structural changes of economies is often debated, mainly due to the mixed empirical evidence on the relationship between the technology adoption and productivity enhancement across countries. This has led to a policy debate regarding technology adoption by countries. In the context of developing countries, the debate is more on the appropriate set of technologies to be adopted rather than whether increased use of technology should be embraced. The endogenous growth theories (Romer, 1986; Rebelo, 1990) explain the role of technology as set of externalities which offsets the diminishing returns to capital. However, it is widely recognized that modern information technology (IT) is distinctly different from some of the past technological breakthroughs like industrial revolution. This difference comes from the unique feature of the use of IT which results in increased separation of the economics of information from the economics of things (Negroponte, 1995). In recent times, a number of studies have explored the role of IT in explaining the convergence of growth rate among countries (Rodrick, 2012; Greiner et al., 2016). However, there is a dearth of adequate studies on the impact of information technology revolution on the broader socio-economic context or structural transformation of a country. The present chapter aims to understand the same by analyzing the rather unique pattern of structural transformation of the Indian economy, which is characterized

by predominance of services sector mainly led by IT services.

The services sector, which now accounts for close to 60 percent of total GDP of India, has been playing an instrumental role in stimulating the growth of the country. A closer look at the GDP composition of India since independence reveals that while the share of agriculture declined and that of industry stagnated, services sector's share has exhibited an increasing trend to reach at 58 percent in 2009-10 from around 30 percent in 1950-51. The stagnating share of manufacturing and in contrast to that, rising share of services in GDP, is something noteworthy if one compares it with the structural change of countries comparable to India in terms GDP per capita. In most of these countries, the structural change has taken place in a common manner where an initially agrarian economy goes through the phase of industrialization, and then at a more advanced stage, services sector predominates. Many researchers concluded that India has leapfrogged the phase of industrialization and instead directly moved to a service-led growth phase. In the case of India, the stagnating share of manufacturing in national output as well as its reduced share in organized sector employment during the last two decades presages a phenomenon of de-industrialization (Chaudhuri, 2015). Does technology play a role in explaining the strange structural change of India? What is its impact in terms of employment generation and export performance? The chapter analyses some of these questions in detail. The analysis presented

in the chapter adds country-specific insights to the literature on the role of technology in economic development and structural change.

The remaining part of the chapter is organized as follows: the next section of this chapter presents a brief discussion of the existing literature on structural change and the role technology plays in it, followed by a description of the Indian experience of structural change in section 3. The later part of the section 3 also delineates the services sector in its present form in India. Section 4 and 5 provide an assessment of the consequences of services-led growth that India has experienced regarding employment and export performance, respectively. Finally section 6 concludes the chapter with policy implications that follows from the analysis and aims at making the technology services driven growth of India sustainable in the longer run.

BACKGROUND

In this section, first, we present a summarized description of existing scholarly works on structural changes of the economy. The second half of this section traces out the role of technology in the context of structural transformation.

Literature on Structural Change: Concepts and Theoretical Underpinning

The traditional literature of structural changes, describes industrialization as a main catalyst of technological innovation and productivity enhancement (Verdoorn, 1949; Kaldor, 1967). However, the same literature defines services as the outcome of “unproductive labour” (Smith, 1976), *i.e.*, products of labour that perishes the moment the labour is performed. However, a greater integration of national economies since the era liberalization has resulted in increasing importance of the services sector through international trade. In a similar fashion, we observe that

the literature on structural change and economic development during the 80’s and 90’s have started recognizing the potential role of services in the development process (Francois, 1990; Eswaran & Kotwal, 2002). There is no consensus in the literature regarding the relative efficiency of industry or services-led growth. However, when the sectoral composition of output of a country increasingly moves towards services at the expense of manufacturing before the country reaches a high level of per capita income, a phenomenon often termed as de-industrialisation (Rowthorn & Ramaswamy, 1999; Palma, 2005), it affects the employment generation and income distribution of the country.

A parallel stream of empirical literature probes into the relative importance of services versus manufacturing in the growth process across countries. (Eichengreen & Gupta, 2013) Many countries have experienced the wave of service-led growth twice in their growth process; once when they move from lower income to lower-middle income and again when they move from middle-income of upper-middle income. The first phase of services that had predominated the growth process while countries moved from lower-middle income to middle income were mainly low-end services. In the latter phase, services were mainly high-skilled and technology embedded. The second wave of services is more predominant in countries with strong democracy and locations closer to the financial hubs. One of the most important reasons underlying the services-sector boom in many countries is the “splintering effect”-the increased use of services by industry of these countries (Francois, 1990). This happens due to the change in production technology and subcontracting of services by many manufacturing firms, as they grow larger in size. According to some studies (Gordon & Gupta, 2005; Shingal, 2014) the boom of the services sector in India is attributed to demand-side factors, such as increased use of services in the manufacturing sector, and increase in domestic and foreign demand for services. From supply side factors, the reform and increased

flow of FDI boosted the services growth in India, which is most visibly prominent in the case of the telecom sector.

Role of Technology in the Context of Structural Change and Economic Growth

The role of technology in economic development is well recognized since the era of Schumpeter (1942), who, through his concept of creative destruction, first illustrated the role of invention in a capitalist economy. The literature on economic growth emphasizes the importance of technical progress in a more explicit manner since the early 90s (Aghion & Howitt, 1990). A parallel stream of literature focuses on the market failure that happens in the case of production of knowledge due to the infinite expansibility and non-rivalry (David, 1985). Quah (1999) explained how the IT, which is at the centre stage of the recent wave of technical innovation, differs from earlier technical breakthrough. To understand this, it is important to draw a distinction between those economic ideas which have economic value in their physical manifestation and those ideas which retain their value independent of the physical medium containing them. The unique feature of IT is that it reduces the distance between consumers and knowledge producers. Since IT products can be easily disseminated, even with the presence of copyright laws, it is not the supply but the demand side factors that determine the spread of such technologies in an economy. This is more so as the cost to consumers of using IT involves the associated cost of learning the technologies. An important policy conclusion that emerges from the nature of IT is that countries need to develop a sufficiently strong demand base for effective diffusion of IT in the economy. This can happen either through creating necessary skill and expertise among the working population or through specific government policies aimed at enhancing private use of new technologies.

In the context of developing countries, adoption of IT as a policy is often debated and remains a crucial public policy decision. The genesis of this debate can be traced back to the fact that even in some developed countries like the USA; the penetration of IT has not necessarily been productivity enhancing (Strassman 1997). In the context of developing countries as well, the empirical evidences on the relations between IT investment and productivity has thrown up mixed evidences (Kraemer & Dedrick 2001; Dedrick et al., 2013; Jin & Cho, 2015).

The conceptual model developed by Kraemer and Dedrick (2001) provides an excellent framework to understand the relation between IT investment and economic factors through a virtuous circle (Figure 1). The first half of the circle shows how economic factors determine the level of investment in IT. The other half of the circle shows how IT influences the national wealth, wage level and growth rate. The model concludes that IT enhances the national productivity through increasing the productivity of capital by complementing other investments.

MANUFACTURING VS. SERVICES- THE INDIAN EXPERIENCE

Following the theoretical foundation on the role of technology in economic development, this section deals with the structural change of the Indian economy since independence. The later part of this section describes the definitional concepts of the services sector in India.

Structural Change in Indian Services Sector

To understand fully the uniqueness of the structural change that India has experienced (Table 1), it is pertinent to compare the sectoral composition of India in Figure 2. In Figure 2, it is evident that services predominate India's composition of GDP.

Figure 1. IT investment and growth

Source: Kraemer & Dedrick, 2001

S

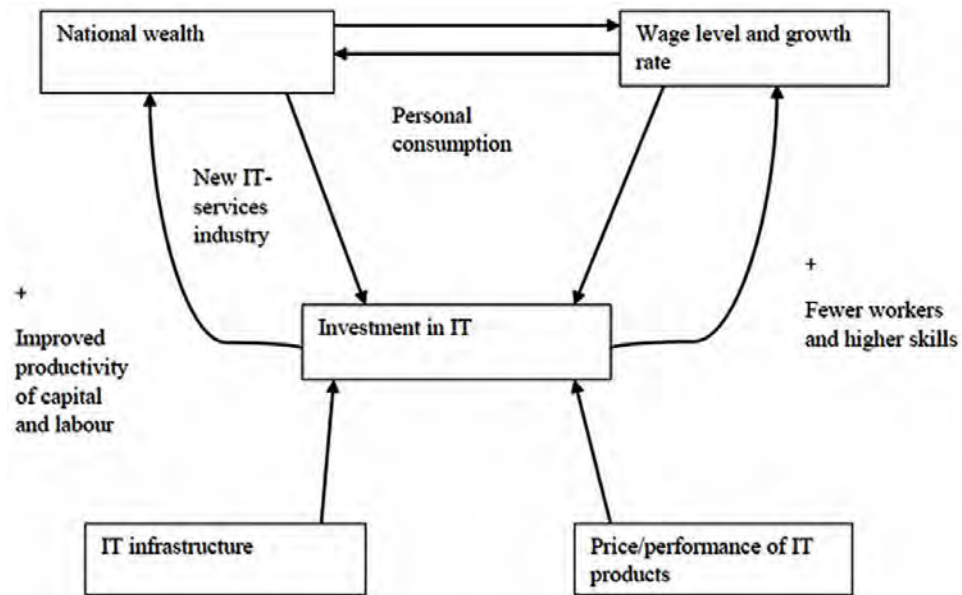
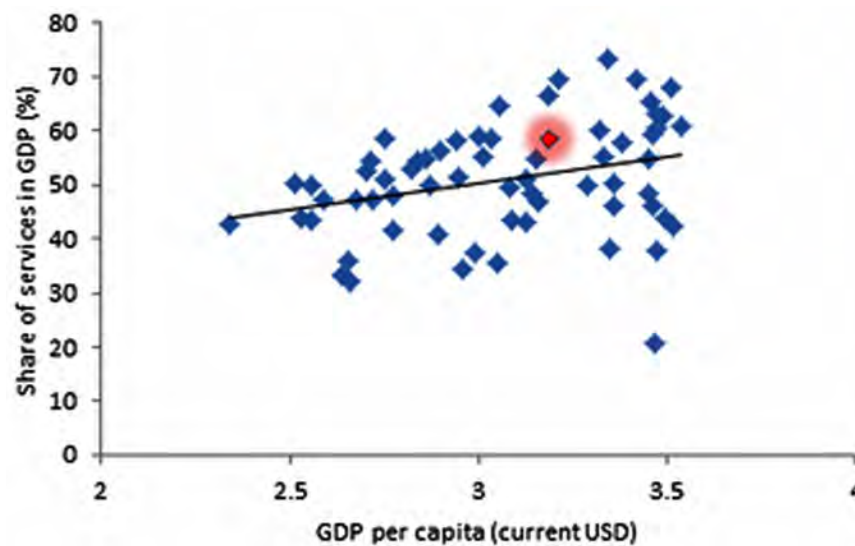


Table 1. Sectorial share in India's GDP

	1950-51	1960-61	1970-71	1980-81	1990-91	2000-01	2004-05	2012-13	Changes (1950-2010)
Agriculture	55.9	46.7	46.0	38.9	31.3	24.6	21.1	17.39	-41.3
Industry	14.3	19.0	20.3	24.5	27.6	26.6	27.2	17.64	13.8
Services	29.8	34.3	33.7	36.6	41.1	48.8	51.5	64.97	27.5

Figure 2. Share of services in GDP

Note: India's position is indicated by the red dot in the scatter plot. Data pertains to 2010



In India, the national accounts data classifies services sector into the following group of activities: wholesale and retail trade, hotel and restaurants, transport services, storage services, communication services, financial services, real estate, renting, business and other social, community and personal services (Ministry of Statistics and Programme Implementation (MoSPI), Government of India). A major limitation of carrying out in-depth empirical analysis in the area of services is the lack of detailed data on services in India. The services sector, in its present form in India, differs distinctly than the way it is defined in the traditional economics. Within the services sector, there exists wide heterogeneity regarding the nature of the industry (public versus private), capital intensity, employment generation capability, economies of scale, technological advancement and so on. The sectors such as telecommunication, banking, and insurance, IT are highly organized by nature with the proportion of organized sector being in the range of 80-90 percent in total output. On the other hand, there are sectors like wholesale and retail trade, real estate and renting and business services where this proportion is less than 20 percent.

The IT Industry in India

India has emerged as one of the major source for providing low cost technology solutions in the global market. The source of such comparative advantage is mainly attributed to the English speaking talent pool available in India. The software Industry in India began to take shape in its present form since the entry of Tata Consultancy Services (TCS) in 1970s in the industry. Following the success of TCS, many software companies were set up in subsequent years. The Indian software industry has a pyramid structure, with few large domestic firms accounting for more than 30 percent of total software export of the country and the smaller firms mainly providing low cost technology solutions to the domestic small and medium sized firms. While the initial trend in the

Indian software industry was “body shopping” *i.e.*, working onsite for US firms, the recent trend suggests a reversal of the same, with off-shoring activities increasing at a rapid pace. This recent trend also indicates that overtime:

- India’s IT industry has matured in terms of its competence
- The necessary infrastructure to work off-shore has been gradually building up in India (Bhatnagar, 2006).

At present, the software sector accounts for around 40 percent of India’s total services sector export and 9 percent of India’s GDP. The IT service industry in India grew at a rate of 10.3 percent in recent years in terms of export. The sector plays a strategic important role not only in shaping India’s growth story in recent years by growing at around 8.5 percent per annum, but also by employing 3.5 million people in the domestic market (National Association of Software and Services Companies (NASSCOM), 2016).

MANUFACTURING VS. SERVICES-FROM THE EMPLOYMENT PERSPECTIVE

India’s workforce is still predominantly agrarian in nature (Table 2). Moreover, the proportion of employment absorbed by both manufacturing and services are quite at a low level if compared with the other low-income countries. As pointed out by Kochhar *et al.* (2006), the lower employment generation capability of Indian manufacturing can be attributed to three important reasons:

1. The state-led heavy industrial policy adopted by India thriving capital intensive sectors in manufacturing
2. The rigid labour laws
3. The education policy of India favoring tertiary education

Figure 3 clearly indicates that the state-led heavy industrialisation policy adopted by India after independence has made the manufacturing sector of India more capital intensive and skill-based as compared with any other comparable countries. The rigid labour laws of India further add to the limited employment generation capability of Indian industry. For an example, an amendment to the Industrial Disputes Act (1947) in 1976 made it mandatory to the firms with more than or equals to 300 workers to seek the permission of the relevant government for any downsizing activity. Apart from the factors mentioned above, the fact that India's investment in education is highly skewed towards tertiary education also explains why the endowment that India possesses today is different from any other low-income countries (Kochhar et al. 2006). This relatively skill-intensive workforce also explains the rise in the service sector of India.

Banga (2005) pointed out three reasons behind the slower employment generation of services sector in India:

- Sectors that have high employment generation potential are growing more slowly than skill-intensive sectors.
- The faster growing services sectors are also sectors with higher productivity.
- Trade in services is accelerating the growth of sectors which are highly skill-intensive in nature.

Before examining the relative importance of these three points in explaining the slower growth of employment in India's services sector, a closer look at the break up of services sector according to their employment capacity would be pertinent. From Figure 4, it emerges that wholesale and retail trade is the largest employment generator within the services industry, followed by transport, and hotels and restaurants, all these together accounting for more than 60 percent of total employment in services. The five-yearly moving average growth rate of different sub-sectors of services (Figure 5) reveal that trade, hotel, transport, and communica-

Figure 3. Value-added share by skill-intensity

Source: Kochhar et al., 2006

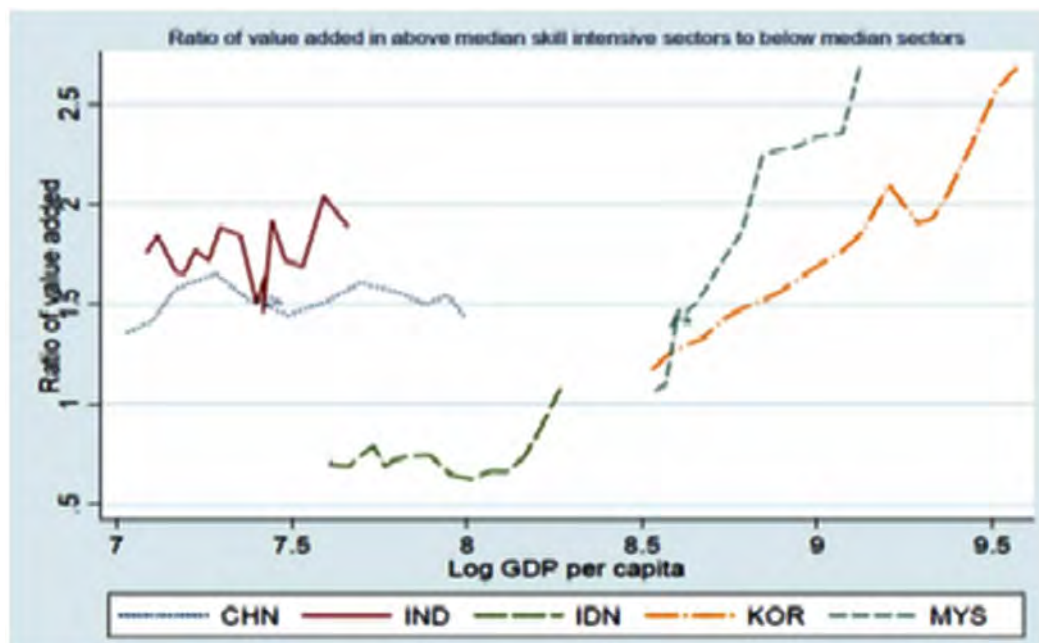


Figure 4. Employment within services-sub sectors

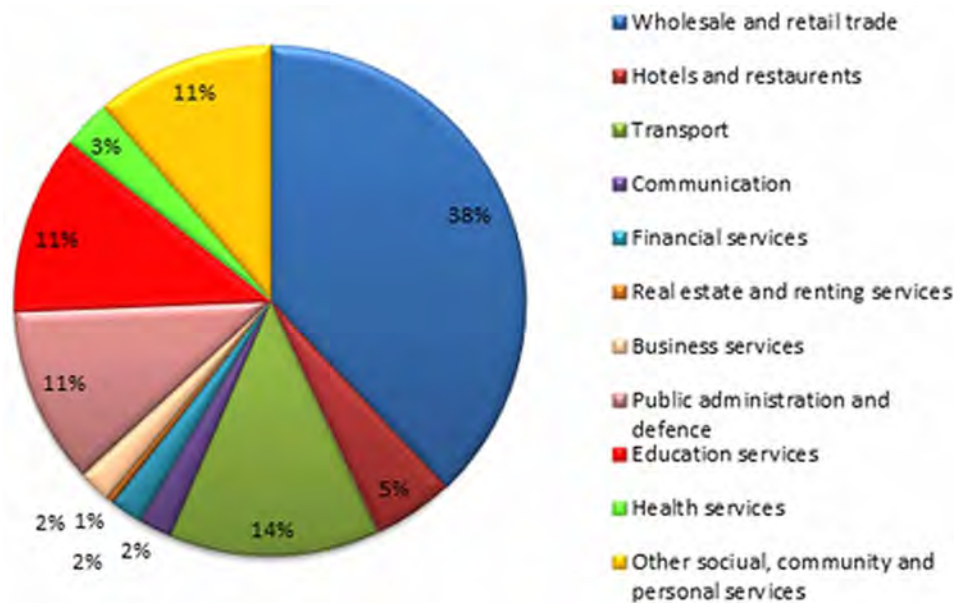
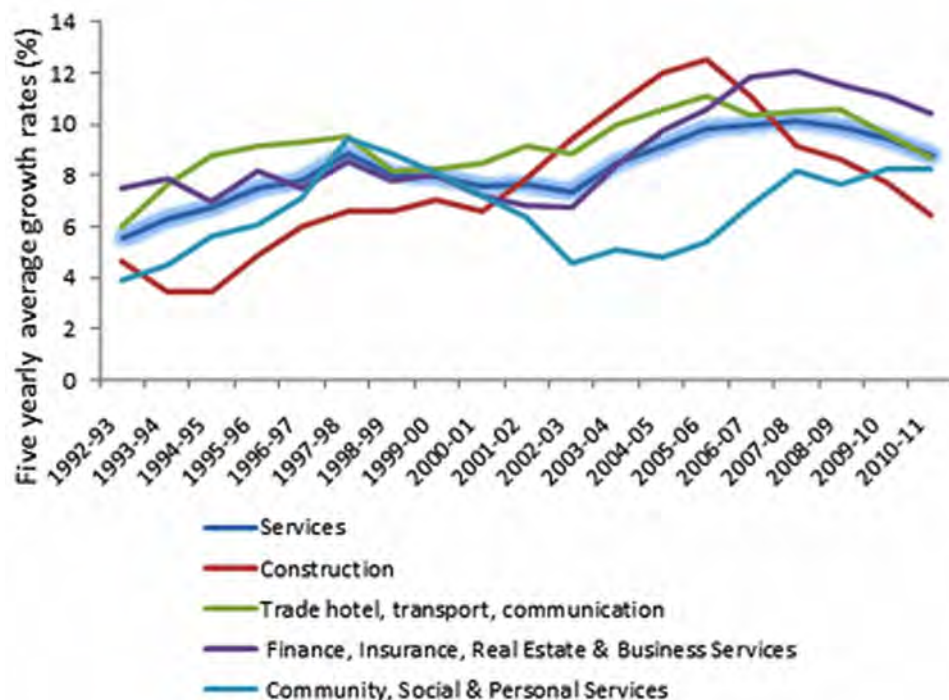


Figure 5. Sectoral growth rates



tion has exhibited higher growth rate as compared with the average growth rate of services shown by the shaded line. Hence, it is difficult to argue that sub-sectors with greater employment generation

capability within services have slowed down in recent years. In the context of productivity growth in Indian services industry, there are very few studies that empirically examine the productivity

growth of sub-sectors with services. One such study by McKinsey and Co (2001) concludes that sectors like telecom, banking, construction, and retail distribution exhibit high productivity level. This could explain the limited role played by these sectors in employment generation.

Lower employment capacity of services sector is another reason behind falling employment elasticity of the sector. This is mostly visible in case of wholesale and retail trade, the segment accounting for the largest share of employment within services (Table 2).

Another explanation of the slower growth of employment in services is the skill-intensive nature of these sectors. However, as discussed earlier, services are a highly heterogeneous set of activities and there are large variations regarding skill requirement within sub-sectors. Also, the urban informal services sector in India is low-skill based and supposedly to have higher employment generation capability compared with the high skill intensive sectors like IT and financial services. The extent to which education acts as barriers to entry in services sectors could be partly captured by the proportion of professional, managerial and technical workers in the total workforce.

MANUFACTURING VS. SERVICES-EXPORT OPPORTUNITIES OF INDIA

During the last two decades, India outperformed all other countries in terms of services exports.

India runs a sizable trade deficit in merchandise trade balance as opposed to a trade surplus in services. While India's share in the world merchandise trade export stood at around 1.52 percent in 2009, India's penetration in the world market in case of service is more than double, standing at 3.6 percent during the same time. The revealed comparative advantage (RCA) index of India for merchandise *vis-à-vis* services clearly indicates the fact that India's comparative advantage in services exports is significantly higher than that of manufacturing. Figure 6, 7, 8 and 9 clearly depict the facts mentioned above.

An in-depth analysis of export data for different sub-sectors of manufacturing indicates that the deteriorating trade balance in merchandise exports in case of India is primarily attributable to chemicals other than pharmaceuticals, aircraft, specialized industrial machinery, telecommunication equipment and parts, computers and general industrial machinery. According to Lall (1999) and Chaudhuri (2013), the decelerating merchandise exports of India is attributable to the structural weakness embedded in the composition of manufacturing goods in India's export basket, which is predominated by technologically primitive products, and thus not able to reap the benefit of sustained growth in world merchandise trade. The same becomes visible on comparing the proportion of high-technology products in total manufacturing exports of India with that of some newly industrialized countries (Table 3).

Table 2. Employment elasticity in different sub-sectors in services

	1989-90 to 1993-94	1993-94 to 1999-2000	1999-2000 to 2004-05	2004-05 to 2009-10
Construction	1.00	1.00	1.09	0.48
Trade	0.63	0.55	0.35	-0.07
Transport, storage and communication	0.49	0.69	0.48	0.08
Banking and insurance	0.92	0.73	1.24	0.27
Community, social and personal services	0.50	0.07	-0.10	-0.14
Services	0.41	0.15	0.45	-0.01

Figure 6. Average growth rate of services exports: 2000-2010

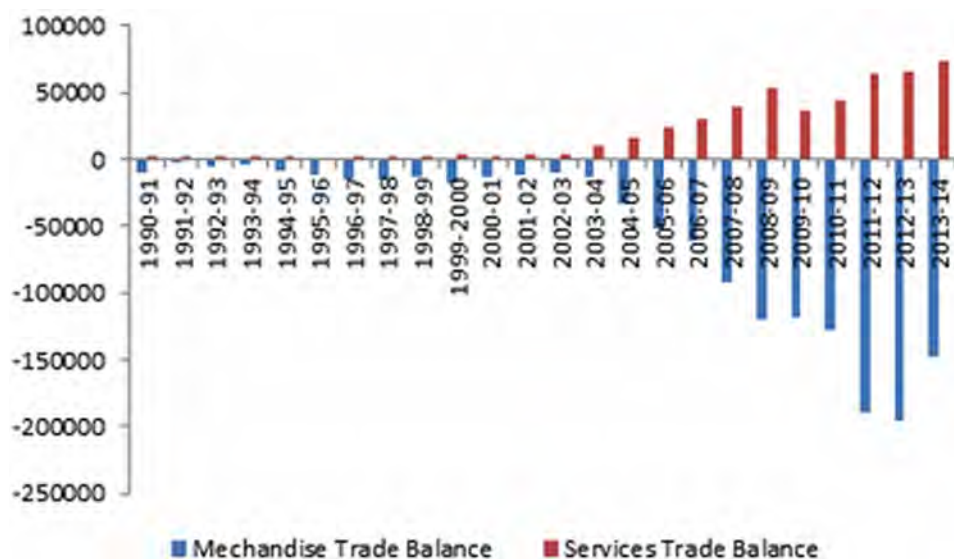
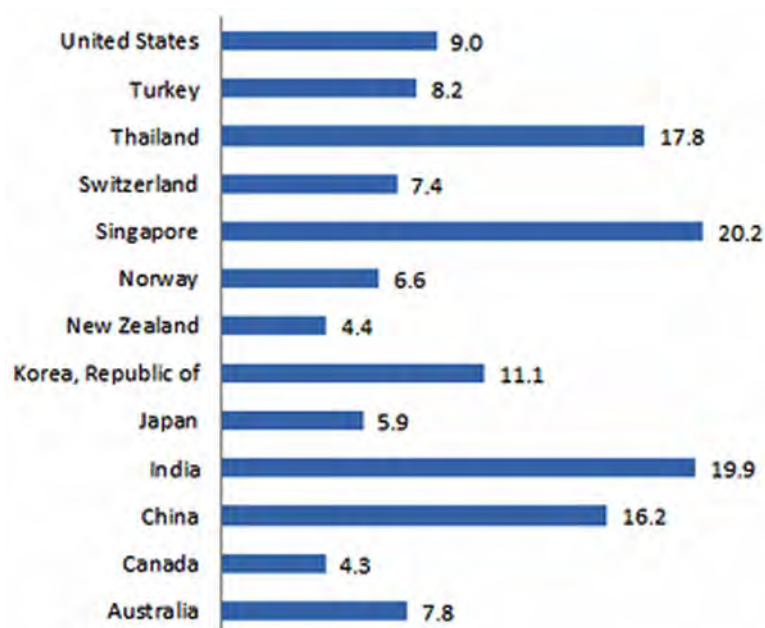


Figure 7. Trade balance (USD Million)



According to the estimates provided by Chaudhuri (2013), services such as telecommunication, IT, air transportation has grown phenomenally, thus creating huge demand for computer hardware, telecom equipment, and aircrafts. Due to the inadequate expansion of these sectors, most of the

need are met by way of importing, thus worsening the merchandise trade balance of India during last one decade (Chaudhuri, 2013).

India's RCA in services exports is mainly explained by the large pool of labour force, with varied skill-sets and English-speaking ability. Ac-

Figure 8. India's share in world market

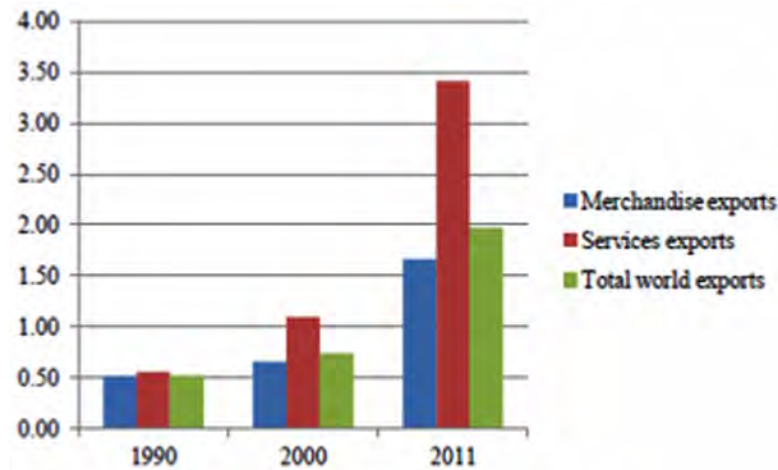


Table 3. High technology exports as percentage of total manufacturing exports

Countries	1990	2000	2010	2014
India	3.94	6.26	7.18	8.59
Korea, Rep.	18.04	35.07	29.47	26.88
Thailand	20.87	33.36	24.02	20.43
China	-	18.98	27.51	25.37
Malaysia	38.21	59.57	44.52	43.87
Singapore	39.89	62.79	49.91	47.18
Vietnam	-	11.07	8.61	26.93
Turkey	1.21	4.83	1.93	1.94
Indonesia	1.59	16.37	9.78	6.97
Philippines	-	72.63	55.26	49.00
East Asia & Pacific	22.10	33.25	26.08	24.19

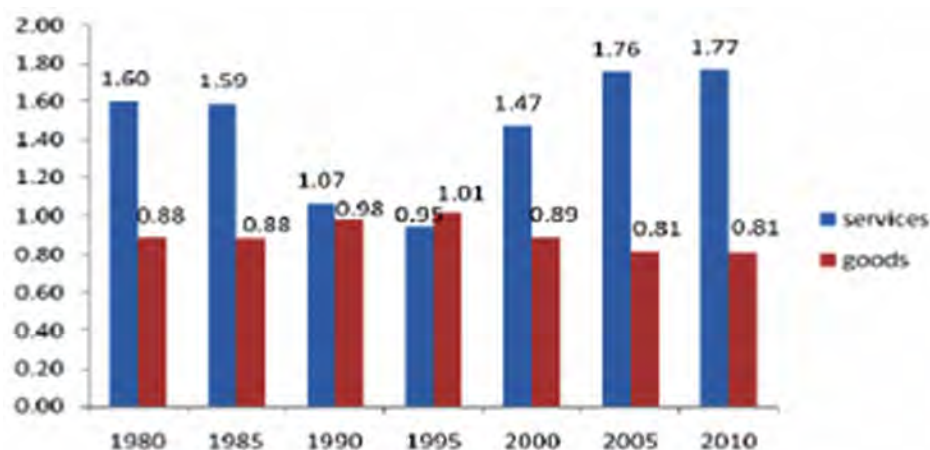
cording to Chanda (2010), there has been a shift in the export composition of services of India from traditional services such as transport and travel towards emerging services like software, accountancy, engineering, and health in recent years. The analysis presented in the same paper indicates that the RCA index has risen sharply for services like software, consulting management, financial and communication. Other regulatory barriers to trade in services have significantly been reduced as a part of the economic reforms

introduced in 90's, accelerating FDI flow and growth of services in India.

FUTURE RESEARCH DIRECTIONS

Based on the analysis of earlier sections, it becomes evident that technology played an important role in shaping India's structural transformation. The slower growth of manufacturing in India is due to lower technology adoption of the sector.

Figure 9. Revealed comparative advantage of India



Also, emergence of IT-BPO industry in India is mainly driven by demographic factors rather than advancement of indigenous technology. In this context, the following policy perspectives aim at spreading the use of technology to make the growth story of India more broad-based and sustainable in long run.

Need to Invent Technology to Boost the Industrial Growth of India

Despite the presence of a thriving services sector, there are few instances where India has been able to invent indigenous technology to boost its manufacturing sector. One of the crucial reasons for the poor performance of Indian manufacturing is identified as low technology adoption (Planning Commission, 2011). At this backdrop, there is a need to create the necessary institutional set up required for invention, diffusion and adoption of technology in the India's industrial sector, a major part of which remains informal in nature even decades after the reform took place (Raj & Sen, 2016). This also implies that India needs technology not only to improve its export performance, but to enhance the inter-sectoral linkage as well. Hence, technology which works at grass roots level is the need for the hour in the case of India

to facilitate its growth story. A slew of recent initiatives taken by the Government in recent times including Digital India, Start Up India and Udyog Aadhar are steps to the right direction.

Need to Spread Use of Technology in Education

The sustainability of the services-led growth of India crucially depends on availability of English-speaking efficient workforce. Based on the survey conducted by Kuruvilla and Ranganathan (2008), it is found that there are a number of human resource related issues challenging India's sustainability in this field in the longer term. Also, there exists a vast urban informal services sector which presently plays a critical role in absorbing low-skilled migrants from rural parts of the country. In the setup, India needs to improve the spread and quality of its vocational education which is crucial in improving productivity in urban informal services segment. The importance of Tele-learning and Tele-training remains significant in the context of India. The National Education Policy of India 2016 specifically highlights the use of information and communication technology (ICT) to spread education at grassroots level in India.

CONCLUSION

Due to the heterogeneity in terms of skill-intensiveness within the service sector, the nature of employment also varies a lot. Though many organizations employ highly skilled salaried personnel, there exists no stable employment or social safety network for people working in the urban informal services sector. In this context, the setting up Unique Identification Authority of India (UIDAI) as a separate entity by the Government of India is an important and significant step which uses information technology to enhance inclusion of marginalized population in the growth process. In order to improve the employment condition in the urban informal sector, specific government policy measures are required for the workers' health and education perspective. One such example is setting up of the community service centers first conceived by the National E-governance Plan (NeGP) of Government of India in 2006. This, along with a broad-based policy of technology adoption would help India sustain its growth story in medium to long term.

REFERENCES

- Aghion, P., & Howitt, P. (1990). *A model of growth through creative destruction* (No. w3223). National Bureau of Economic Research. doi:10.3386/w3223
- Banga, R. (2005). *Trade and foreign direct investment in services: a review*. Indian Council for Research on International Economic Relations (ICRIER), Working Paper, (154).
- Bhagwati, J. N. (1984). Why are services cheaper in the poor countries? *The Economic Journal*, 94(374), 279–286. doi:10.2307/2232350
- Bhatnagar, S. (2006). India's software industry. In *Technology, adaptation, and exports: How some developing countries got it right*. World Bank Publications.
- Chanda, R. (2010). *Services-led Growth. The Concise Oxford Companion to Economics in India*. OUP.
- Chaudhuri, S. (2013). Manufacturing Trade Deficit and Industrial Policy in India. *Economic and Political Weekly*, 48(8), 41–50.
- Chaudhuri, S. (2015). Import Liberalisation and Premature Deindustrialization in India. *Economic and Political Weekly*, 50(43), 61.
- David, P. A. (1985). Clio and the Economics of QWERTY. *The American Economic Review*, 75(2), 332–337.
- Derrick, J., Kraemer, K. L., & Shih, E. (2013). Information technology and productivity in developed and developing countries. *Journal of Management Information Systems*, 30(1), 97–122. doi:10.2753/MIS0742-1222300103
- Eichengreen, B., & Gupta, P. (2013). The two waves of service-sector growth. *Oxford Economic Papers*, 65(1), 96–123. doi:10.1093/oep/gpr059
- Eswaran, M., & Kotwal, A. (2002). The role of the service sector in the process of industrialisation. *Journal of Development Economics*, 68(2), 401–420. doi:10.1016/S0304-3878(02)00019-6
- Francois, J. F. (1990). Producer services, scale, and the division of labor. *Oxford Economic Papers*, 42(4), 715–729.
- Gordon, J., & Gupta, P. (2005). *Understanding India's services revolution*. Palgrave Macmillan, UK.
- Greiner, A., Semmler, W., & Gong, G. (2016). *The forces of economic growth: a time series perspective*. Princeton University Press.
- Jin, S., & Cho, C. M. (2015). Is ICT a new essential for national economic growth in an information society? *Government Information Quarterly*, 32(3), 253–260. doi:10.1016/j.giq.2015.04.007

- Kaldor, N. (1967). *Strategic factors in economic development, Strategic factors in economic development*. Ithaca, NY: New York State School of Industrial and Labor Relations, Cornell U.
- Kochhar, K., Kumar, U., Rajan, R., Subramanian, A., & Tokatlidis, I. (2006). Indias pattern of development: What happened, what follows? *Journal of Monetary Economics*, 53(5), 981–1019. doi:10.1016/j.jmoneco.2006.05.007
- Kraemer, K. L., & Dedrick, J. (2001). *Information technology and productivity: results and policy implications of cross-country studies*. In *Information technology, productivity, and economic growth: International evidence and implications for economic development*. Oxford University Press.
- Kuruvilla, S., & Ranganathan, A. (2008). Economic development strategies and macro- and micro-level human resource policies: The case of Indias outsourcing industry. *Industrial & Labor Relations Review*, 62(1), 39–72. doi:10.1177/001979390806200103
- McKinsey & Co. (2001). *India- the Growth Imperative: Understanding the Barriers to Rapid Growth and Employment Creation*. Accessed at 20th March, Retrieved from <http://www.mckinsey.com/global-themes/india/growth-imperative-for-india>
- National Policy on Education. (2016). Ministry of Human Resource Development, Government of India. Accessed at 20th March, Retrieved from: <http://www.nuepa.org/New/download/NEP2016/ReportNEP.pdf>
- Negroponte, N. (1995). *Being digital*. Alfred A. Knopf. Retrieved from: http://mospi.nic.in/mospi_new/upload/nas_13.html
- Palma, G. (2005). Four sources of de-industrialisation and a new concept of the Dutch Disease. In *Beyond reforms: Structural dynamics and macroeconomic vulnerability*. Stanford University Press and World bank.
- Quah, D. (1999). *The weightless economy in economic development*. In *Information technology, productivity, and economic growth: International evidence and implications for economic development*. Oxford University Press.
- Raj, R. S., & Sen, K. (2016). *Out of the shadows? The informal manufacturing in post-reform India*. Academic Press.
- Rebelo, S. T. (1990). *Long run policy analysis and long run growth (No. w3325)*. National Bureau of Economic Research. doi:10.3386/w3325
- Romer, P. M. (1986). Increasing returns and long-run growth. *Journal of Political Economy*, 94(5), 1002–1037. doi:10.1086/261420
- Rowthorn, R., & Ramaswamy, R. (1999). Growth, trade, and deindustrialisation. *IMF Staff Papers*, 18–41.
- Schumpeter, J. (1942). *Creative destruction. Capitalism, socialism and democracy*. Routledge.
- Shingal, A. (2014). The Services Sector in Indias States: A Tale of Growth, Convergence and Trade. *World Economy*, 37(12), 1773–1793. doi:10.1111/twec.12163
- Strassman, P. (1997). Computers are yet to make companies more productive. *Computerworld*, 15, 28–29.
- The Industrial Dispute Act. (1947). *India Code*. Ministry of Law and Justice, Government of India.
- Verdoorn, P. J. (1949). On the factors determining the growth of labor productivity. *Italian Economic Papers*, 2, 59–68.

ADDITIONAL READING

- Chenery, H. B. (1960). Patterns of industrial growth. *The American Economic Review*, 50(4), 624–654.
- Fisher, A. G. B. (1935). *The clash of progress and security*. London: Macmillan.

KEY TERMS AND DEFINITIONS

De-Industrialization: The term is used to describe declining industrial profit and labour displacement due to increased innovation and use of technology in industry.

Revealed Comparative Advantage Index: It is based on the notion of comparative advantage pioneered by Ricardo in the theory of international trade. The index is calculated as: $RCA = (E_{ij} / E_{it}) / (E_{nj} / E_{nt})$ where: E: Exports, I: Country index N: Set of countries J: Commodity index t: Set of commodities.

Services: The tertiary sector of the economy (also known as the service sector or the service industry) is one of the three economic sectors, the others being the secondary sector (approximately

the same as manufacturing) and the primary sector (agriculture, fishing, and extraction such as mining). It mainly comprises of the “intangible goods” like tourism, healthcare, hospital and other public services.

Splintering Effect: Splintering occurs when a part of the activities in manufacturing sector is outsourced to the services sector. For example, painting a car is often outsourced from the automobile industry to a services firm. This is also known as “fragmentation” of activities in the economy.

Verdoorn's Law: Named After Petrus Johannes Verdoorn (1949), the law states that in the long run productivity generally grows proportionally to the square root of output. As a result, in the presence of increasing returns, faster growth in output results in faster growth in productivity.

New Faces of Digital Divide and How to Bridge It

Viktor Freiman

Université de Moncton, Canada

Dragana Martinovic

University of Windsor, Canada

Xavier Robichaud

Univesité de Moncton, Canada

INTRODUCTION

The objective of our chapter is to explore, through the lenses of digital divide, what are challenges to alleviating socio-economic and intellectual limitations for prosperity of each individual. Will new technologies and access to them really help to develop citizens who are able to contribute in creative and democratic ways to society? In order to answer this question, we extend our work presented in Martinovic and Freiman (2013) to:

- Identify factors to consider in designing flexible, innovative, and inclusive programs for all citizens to enable them to successfully function in the era of the Internet, new media, and computer technologies.
- Analyze how digital divide problematizes one's chances to be involved in the knowledge economy.
- Investigate ways in which digital divide may be circumvented.

In the past 15 years or so, Information and Communication Technologies (ICT; e.g., personal computers, cell phones, Internet) have become pervasive in developed countries, such as Canada and the USA. These tools can be used for both in-school and out-of-school activities, and are particularly suitable for connecting individuals and communities globally (Beetham, McGill,

& Littlejohn, 2009). However, many countries could not provide to their citizens the same level of access to digital technology, which, in its turns risks to deprive them of opportunities to develop abilities necessary for a meaningful use of digital technology and computer networks, including the Internet to gather, manage, and evaluate information, to create documents in multiple media forms, and to communicate at distance, are all aspects of digital literacy, as delineated in various policy documents (e.g., California Emerging Technology Fund, 2008). In fact, digital literacy has become almost a prerequisite for creativity, innovation, and entrepreneurship, all the irreplaceable attributes of the 21st century citizen (Beetham, McGill, & Littlejohn, 2009). However, although it is recognized that technology may positively affect both social and cognitive development of every citizen, it may also create or maintain economic divide across students, teachers, and schools (AERA, 2013), as well as between social groups or societies, even in the most developed countries..

Our effort to examine these tensions that in our opinion go beyond the usual issues of “have’s” and “have not’s” arises from the literature review study (Freiman, Martinovic, & Karadag 2011) we conducted upon request of the Ontario Ministry of Children and Youth Services (MCYS, 2012). Authors of numerous sources we reviewed, including position papers, government-ordered reports, as well as research data, relate digital divide to

DOI: 10.4018/978-1-5225-2255-3.ch630

several aspects of ICT, such as machine vintage, connectivity, online skills, autonomy and freedom of access, technical support, and interest in using the technology (Hawkins & Oblinger, 2006; Oblinger, 2008).

BACKGROUND

By introducing the concept in early 2000s, an OECD report (OECD, 2001) identifies two types of digital divide, the first one, based on the opportunity to access digital tools, including the Internet, and the second one, related to the ability to use those tools. For example, Norris (2001) argues that digital divide appears as result of uneven growth of the Internet resources, their multi-dimensional character, as well as transitory nature of the process associated to this growth. Also, in early 2000s, Attewell (2001) and Swain and Pearson (2001) identify the discrepancies in access to technology as the first digital divide, while differences in the effective use of ICT belong to the second digital divide; thus users with only basic access to technology can be disadvantaged, being technologically illiterate because of their limited opportunity to use digital resources. This view is shared by Dance (2003), who compares the cyberspace netizenry to citizenship in the ancient Athenian democracy that excluded women, slaves, and those with foreign origins. In like fashion, today's cyber netizenry ends up forming cyber-elite (i.e., an information-rich, digerati, and virtual class). Dance (2003) further points to some older reports that demonstrated that the digital divide in the USA was growing and had a deeper and more far-reaching impact than before. In this way, the digital divide resembles the economic divide: The (information) rich get richer, while the (information) poor get poorer.

In order to better grasp a potential impact of digital divide on individual, groups and societies, we should start with a brief clarification of terminology. First, as pointed out by Chen and Wellman (2003), the digital divide has multiple

faces and should be termed “digital divides” (p.3). In order to be addressed, various aspects need to be considered: physical access, financial access, cognitive access, content access, and political access (Wilson & Wilson, 2000). Second, the term “digital divide” has both technological and social resonances. Next, DiMaggio and Hargittai (2001) identify five dimensions of digital inequality: equipment, autonomy of use (location of access), skill, social support, and purposes for using the Internet, while Cuneo (2002) lists twelve perspectives on the digital divide. These include: demographics (e.g., computer per person ratio), age, gender, geography (i.e., where one lives-infrastructure), disposition to ICT (e.g., fear, lack of confidence), learning (e.g., traditional vs. on-line), disabilities, and economic, social, labor-related, cultural, and political factors. The author further declares that “in some ways education is at the heart of the Internet and the Digital Divide because of the importance to society of transmitting information and knowledge” (p. 25). This view is shared by Negroponte whose quotation “... The digital divide is a learning divide - digital is the means through which children learn learning” can be found in many reports of the launch of the “\$100 laptop” project (see, for example, Twist, 2005, para. 36-37). Moreover, van Dijk (2006) resumes his review of studies on digital divide conducted in 2000-2005 by arguing that “in terms of physical access the divide seems to be closing in the most developed countries; concerning digital skills and the use of applications the divide persists or widens” (p. 221).

In this optics, an analysis of digital divide conducted by Dewan and Riggins (2005) pointed at the importance of taking into consideration three levels of digital divide, namely individual, organizational and global ones. According to them, at the individual level, variations in access and/or the ability to use technology can be analysed among different segments of a social system, as well as policies that address these divides. At the organizational level, several factors (size, geographical location, industry, and ownership

status) need to be considered when looking at adoption and the ability to exploit technology in organizations. Finally, at the global level, differences between countries can exist in terms of the access and the use of technology related to their wealth, education levels, infrastructure, and other socio-economic factors.

MAIN FOCUS OF THE ARTICLE

Issues, Controversies, Problems

A cross-country analysis using data on computer use and Internet use determinants in 161 countries in the world, conducted by Chinn and Fairlie (2004) found that economic variables (income per capita, years of schooling, illiteracy), demographic variables (youth and aged dependency ratios, urbanization rate), infrastructure indicators (telephone density, electricity consumption), telecommunication regulatory quality are largely responsible for disparities between countries. The authors put income differentials, telephone density and regulatory quality on the top of their list of factors explaining the global digital divide (*idem.*)

There seem also to be an overall agreement that some groups may face additional vulnerabilities and opportunities in “the digital Era.” Several studies provide us with examples of facts that digital technologies do not necessarily promote fairness and equality between individuals, groups, and societies. According to Willms and Corbett (2003), having access to both a computer and the Internet at home is strongly positively correlated with socio-economic status (i.e., the higher the parents’ socio-economic status, the more likely that students have access) and also with gender (i.e., girls are less likely to have access than boys). Boyd (2007) argued that Facebook members, the so-called “hegemonic teens,” represented primarily the middle and upper middle class, whereas MySpace members, the so-called “subaltern teens,” came from mixed-class or lower class backgrounds (p. 41). The researchers also found

that children from immigrant families appeared to have a better chance of having a home computer and Internet access than children from mainstream families with higher socio-economic status. Immigrant families tended to invest more in their children’s education to ensure that they possessed highly sought computer skills.

Also, it is notable that the Internet is still used by minority groups and indigenous peoples primarily for community building and for access to alternative viewpoints that may not have found a place in the mainstream media (Budka, Bell, & Fiser, 2009). Indigenous Internet practices, for example, appeared to distinguish between inward and outward activities (Landzelius, 2006), that is, between how indigenous people present themselves to the outside world and the activities that serve for identity-finding and for strengthening intercommunity relations. Landzelius concluded that the Internet, as a connecting technology, helps some boundaries to dissolve, but intensifies others. While, according to Landzelius (2002), the Internet may help indigenous youth to find their cultural identity through connections with similar youth on the Internet, this may exacerbate their exclusion from the mainstream societies in which they reside.

At the global level, saying that the debate about the first digital divide related to the access issue is closed seem to be too optimistic, as mentioned by Steyaert and Gould (2009) who cite www.internetworldstats.com (December 2008) that only approximately 1.5 billion people, or 22 per cent of the world’s population of 6.6 billion have access to the Internet. Moreover, the UNDP (2007) report, to which refer Steyaert and Gould (2009), reveals that industrialized countries, with only 15 per cent of the world’s population, are home to 88 per cent of all Internet users. Other facts given by the authors extend a quest for investigation and discussion, namely that London, UK, appears to still have more Internet domains than the whole of Africa. Already in early 2000s, Finland alone had more Internet users than the whole of Latin America without mentioning that ‘the telephone

directory for Manhattan is bigger than that of the African continent, and 50 per cent of the world's population has never made or received a telephone call' (UNDP, 2007, p.743).

The results of a study conducted by Mardikyan et al. (2015) show that, global digital divide has a significant difference in almost every categorical variable which is studied. There are significant differences between continents, and income levels in terms of ICT Development Index. Similarly, there is a significant difference between developed and developing countries and between OECD member and not member countries in terms of this index (p. 9). At the same time Pick and Sarkar (2015) draw attention to the fact that besides differences between nations, generally well documented in the literature, that can be significant discrepancies within one country, on the state, province, prefecture levels which are less known.

In terms of progresses made by different countries in different parts of the worlds, a recent study conducted by Skaletsky (2013) reveals different patterns in terms of improving the access to technology and its meaningful use. It shows, for instance, important progress made by the Eastern European countries whereas in African and Latin American countries issues seem to remain unsolved (idem.). Even in developed countries, like the U.S., while some gaps (e.g., gender-related) are diminished, others (e.g., ethnicity, education level, and age) seem to remain and even widen, which puts a question mark on 'the promise of information technology to benefit traditionally disadvantaged demographic groups and provide a more level playing field for academic and economic marketplace achievement' (Losh, 2010, p. 218).

Moreover, if one recognizes the existence of digital divide, and this, at different levels, its impact on people's life and society, as whole must be questioned, according to Bourdeau de Fontenay and Beltran (2008) who argue, at the same time that 'where digital inequality amplifies inequality, the digital divide needs to be evaluated on its own' (p. 4). While research evidence

points at the access to technology as one of major indicators of digital inclusion, simply providing an access is not helpful without the commitment to a concomitant development of higher levels of critical thinking. Even if we recognize a well-established fact that countries with increased ICT penetration rates can provide more access to ICTs than before, and as a result, could decrease adolescents' reliance on school ICT, according to Zhong (2011), increased ICT penetration rate does not enhance adolescents' self-evaluation of their digital skills. In other words, as mentioned by van Deursen and van Dijk (2011), the "original digital divide of physical Internet access has evolved into a divide that includes differences in skills to use the Internet" (p.893).

At the same time, Schrock and Boyd (2008) have raised the concern that digital literacy phenomena themselves are in a constant state of flux and thus are moving targets. Understanding what online/mobile environments are available for learning and skill development, as well as how these environments can enhance purposeful collaboration among the young, may affect, for example, programs dealing with such matters as anti-cyberbullying, reduction of poverty, social inclusion of recent immigrants or Aboriginal populations, and promotion of digital literacy among girls. Significance of their study is in opening up a discussion beyond an understanding that the first digital divide relating to access to technological infrastructure and hardware has been almost resolved, at least in Canada. Worldwide, example of the second divide can be easily extended. For example, Minghetti and Buhalis (2010) found that a number of causes, such as lack of knowledge, trust, literacy, language skills, and content availability of credit card and low bandwidth can explain digital divide in the tourism industry.

Further, Chen and Liu (2013) point at the rural-urban gap, as well as imbalanced allocation of educational resources as one of the major factors of digital divide. Also, the authors show limited access to digital teaching materials and online learning in disadvantaged areas due to license

restrictions, copyright issues and the high cost of information sources as example of how this factor affects small (and often in rural and remote areas) communities. For example, Rekhari (2009) shows that in Australia, twice as many as non-Indigenous people those from Indigenous communities do not have yet Internet connection.

In some way, wireless networks and small affordable computers based on the open source software (such as a “\$100 laptop” initiative, MIT project) promise to make computers not only potentially but actually ubiquitous, can help in reducing costs. However, it is incumbent upon educators not only to know about these developments but also how to adapt to new societal needs for citizens who must become capable problem solvers and lateral thinkers; persistent and determined learners; open-minded and progressive strategists; and, confident users of technology (Martinovic & Magliaro, 2007).

Therefore, we emphasize that addressing the second digital divide must happen through improving digital literacy competencies among all citizens. This is why educational issues should be brought at the upfront of societal debates that would extend meaningful use of technologies above formal ones found in academic routines of educational institutions but very important in everyday life of learners (Karsenti & Collin, 2013). In the same way, Zhong (2011) reports findings showing no significant interaction between what is done with ICT access at home and ICT access in school. Also, there are studies expressing concerns about a ‘failure of traditional academic literacies to engage marginalized populations’ (Pirbhai-Illich, Turner, & Austin, 2009).

On the other side, Wiseman and Anderson (2012) found that “ICT in instructional practice has not gone much past the entry level of use in the Gulf Cooperation Council and around the world...This indicates that even where schools and classrooms are well-equipped for computer resources, educational systems cannot ‘buy’ learning and knowledge with resources” (p.614).

Thus, different groups may become disadvantaged in terms of usage of digital technologies, especially regarding the digital literacy. For example, recent study of Kuo, Tseng, Lin and Tang (2013) reports that of the 771 million adults who are still digitally illiterate, the majority are women, which is, according to them the third most important issue that women face globally. With a growth of social networks worldwide, some authors point at reluctance to some aspects of their use by elderly citizens in situations of learning, self-expression, and building, maintenance and sharing of identity through digital technology, as they may perceive those situations as posing a threat to their identity (Martin, 2009). This can be an indicator of another group that may need special attention and help to become fully integrated in a digital world.

SOLUTIONS AND RECOMMENDATIONS

Based on the literature, it is plausible to conclude that while some divides disappear or get minimized, the others appear. Based on Canadian context Cuneo (2003) lists twelve perspectives on digital divide, arguing that it is necessary to deal with issues of disposition and skill that teachers have with ICT, as well as labor conditions under which teachers work. Professional development opportunities that range across all education layers need to be provided, including reasonable evaluation of increased workload in presence of ICT. Karsenti & Collin (2013) express the same view: “Il revient aux institutions éducatives d’assurer la formation des apprenants aux usages éducatifs des technologies (...)” (p.205). For Pirbhai-Illich, Turner and Austin (2009), to build future selves for students include succeeding in school and overcoming their current life circumstances, which means that the researchers “acknowledge the need for instruction that develops both academic and new literacies, exemplified in the multimodal media production where they can envision new

possibilities (p.159). Zhong (2011) adds that “both the school and the family can construct supportive social environment for adolescents” because “when an individual child has trouble using ICTs teachers, family members and peer students can offer help” (p.745).

In fact, different countries appear to handle digital divide with different levels of priority. For instance, as mentioned by Ferro et al. (2008) the theme of digital divide has moved higher on the list of priorities in Europe, as a key pillar of the Strategic Plan i2010, while the latest initiative of the US administration for expanding broadband access in schools nationwide (The White House, 2013), opens more opportunities for inclusive education of new generations of students. At the same time, many authors stress an urgent need to elaborate government’s technology policy in narrowing different types of digital divide, such as mentioned above Rural–Urban Knowledge Divide in higher education (Chen & Liu, 2013) for instance, to authorize teachers to provide online learning courses for pedagogical purposes, and to offer the right for rural student and residents to access online learning resources and knowledge with the financial help of the government. In a similar way, the ICT use can be effective in empowering young Aboriginal people (Singleton, Rola-Rubzen, Muir, Muir & McGregor, 2009). For Rekhari (2009) for instance, “indigenous people recognise the role that ICTs can play as a tool in self-determination” (p.179), and an online space can provide opportunities for “sparked an exchange of culture, creativity and experience between non-Indigenous and Indigenous young people” (p.178)

Kuo, Tseng, Lin, and Tang (2013), on their own account, take interest in female learners and show that with “sufficient social support, female learners build up their computer self-efficacy, persevere with ICT learning, and promote their subjective well-being” (p.214). As for Martin (2009), he focuses on seniors and asserts that “the perception of seniors’ relevant and effective digital action is important in challenging the “gen-

eration gap” (p.12). At their turn, Wiseman and Anderson (2012) consider that addressing a “need to provide training and professional development in ICT-based instruction” can help developing educational systems, such as Gulf Cooperation Council’s ones’ (p.615).

In case where disparities exist between different social groups within one country, like in the South Africa, where an English dominated ICT field gives advantages to the English-speaking learners in relation to using ICT, it is suggested by Gudmundsdottir (2010) to place increased focus on adapting ICT to the local curriculum, local needs and the linguistic environment of the users as a recommendation meaningful for other countries with similar issues.

Besides education, other factors should be addressed, once we talk about a digital divide convergence. For example, a model developed by Kyriakidou, Michalakelis, and Spicopoulos (2011) allows for predicting that by 2018, European countries will come closer, in terms of broadband diffusion.

Overall, despite a widespread optimistic assessment of potential of education to bridge digital divide, concerns expressed by Sciadas (2001) more than a decade ago about lack of understanding barriers to access and use of new technologies still remain actual field of research and practice. In fact, as pointed by Sinclair and Bramley (2011), “digital inclusion will only be achieved if the social rather than technological barriers inhibiting digital engagement are addressed” (p.1).

Through synergistic partnership between teachers, administrators, researchers, and students, the school system will be able to move closer to achieving the goal of purposeful and pedagogically meaningful use of ICT within daily classroom pedagogy, alleviating digital divide and inequity.

FUTURE RESEARCH DIRECTIONS

As mentioned by Billon, Lera-Lopez and Marco (2010), investigating ICT adoption and access

is necessary, but studies must now focus on the acquisition of skills needed for a more productive use of technologies, such as the Internet. It could be expressed by a nation level study of ICT use in education systems (Wiseman & Anderson, 2012) joined with a rigorous conception of digital practices and expert knowledge of learning design. For instance, it will be important that the researches centre on the development of teachers' pedagogy and ICT use (Wiseman & Anderson, 2012).

In addition, Kuo, Tseng, Lin, and Tang (2013) consider that "It is important to identify the reasons for (...) illiteracy and the methods required to foster and encourage female ICT learning" (p.217). Also as pointed out by Salman and Rahim (2012) the literacy, knowledge and technology motivation of the Internet of the users need to be further study in order to achieve a digital inclusion. Asiedu (2012) suggests that one direction could be finding innovative ways of incorporating attractive features of dominant communication mediums (the radio in his study) into ICTs. Rekhar (2009) proposes research on a more inclusive future for Indigenous and non-Indigenous people.

CONCLUSION

Consequently, we have come to the conclusion that new technologies and access to them really help to develop citizens who are able to contribute in creative and democratic ways to society. Yet, a number of conditions must be carried out. First, we have to include all the aspects of digital divide: physical, financial cognitive, content and political accesses; second we have to consider the technological and social resonances of digital technologies because, these latter do not necessarily promote fairness and equality. At last and over all, we have to take in consideration that providing an access to digital divide is not helpful without a concomitant development of critical thinking.

REFERENCES

- American Educational Research Association. (2013). *Annual Meeting Call for Submissions*. Retrieved from <http://www.aera.net/Portals/38/docs/New%20Date%20Call%20for%20Submissions%20AM13%20FINAL%20REVISED.pdf>
- Asiedu, C. (2012). Information communication technologies for gender and development in africa: The case for radio and technological blending. *International Communication Gazette*, 74(3), 240–257. doi:10.1177/1748048511432606
- Attewell, P. (2001). The First and second digital divides. *Sociology of Education*, 74(3), 252–259. doi:10.2307/2673277
- Beetham, H., McGill, L., & Littlejohn, A. (2009). *Thriving in the 21st century: Learning literacies for the digital age* (LLiDA project). Retrieved from <http://www.jisc.ac.uk/media/documents/projects/llidareportjune2009.pdf>
- Billon, M., Lera-Lopez, F., & Marco, R. (2010). Differences in digitalization levels: A multivariate analysis studying the global digital divide. *Review of World Economics. Weltwirtschaftliches Archiv*, 146(1), 39–73. doi:10.1007/s10290-009-0045-y
- Bourdeau de Fontenay, A. and Beltran, F. (2008). Inequality and economic growth: Should we be concerned by the digital divide? *ITS Montreal*, 1-37.
- Boyd, D. (2007). Viewing American class divisions through Facebook and MySpace. *Apophenia Blog Essay*. Retrieved from <http://www.danah.org/papers/essays/ClassDivisions.html>
- Budka, P., Bell, B., & Fiser, A. (2009). MyKnet.org: How Northern Ontario's First Nation communities made themselves at home on the World Wide Web. *The Journal of Community Informatics*, 5(2). Retrieved from <http://ci-journal.net/index.php/ciej/article/view/568/450>

- California Emerging Technology Fund. (2008). *California ICT Digital Literacy Assessments and Curriculum Framework*. Retrieved from <http://www.ictliteracy.info/>
- Chen, R.-S., & Liu, I.-F. (2013). Research on the effectiveness of information technology in reducing the Rural–Urban Knowledge Divide. *Computers & Education*, 63, 437–445. doi:10.1016/j.compedu.2013.01.002
- Chen, W., & Wellman, B. (2003). *Charting Digital Divides: Comparing Socioeconomic, Gender, Life Stage, and Rural-Urban Internet Access and Use in Eight Countries*. Retrieved from http://homes.chass.utoronto.ca/~wellman/publications/amd_ses/charting-divides_long.pdf
- Chinn, M. D., & Fairlie, R. W. (2004). *The Determinants of the Global Digital Divide: A Cross-Country Analysis of Computer and Internet Penetration*. Retrieved from <http://ideas.repec.org/p/egc/wpaper/881.html>
- Cuneo, C. (2002). *Globalized and Localized Digital Divides Along the Information Highway: A Fragile Synthesis Across Bridges, Ramps, Cloverleaves, and Ladders. 33rd Annual Sorokin Lecture*. Saskatoon, Canada: University of Saskatchewan.
- Dance, F. E. X. (2003). The digital divide. In L. Strate, R. J. Jacobson, & S. Gibson (Eds.), *Communication and cyberspace: Social interaction in an electronic environment* (2nd ed.). Cresskill, NJ: Hampton Press, Inc.
- Dewan, S., & Riggins, F. J. (2005). The digital divide: Current and future research directions. *Journal of the Association for Information Systems*, 6(2).
- DiMaggio, P., & Hargittai, E. (2001). *From the 'Digital Divide' to 'Digital Inequality': Studying Internet Use as Penetration Increases*. Princeton University Center for Arts and Cultural Policy Studies, Working Paper Series Number 15, Princeton NJ.
- Ferro, E., Dwivedi, Y. K., Ramon Gil-Garcia, J., & Williams, M. D. (2008). *Overcoming Digital Divides: Constructing an Equitable and Competitive Information Society*. Retrieved from http://www.enricoferro.com/html/cfc_book.html
- Freiman, V., Martinovic, D., & Karadag, Z. (2011). *Transition from childhood to adulthood from digital literacy perspective: What does recent research say?* IADIS International Conference, e-Learning 2011, Rome, Italy.
- Gudmundsdottir, G. (2010). From digital divide to digital equity: Learners' ICT competence in four primary schools in Cape Town, South Africa. *International Journal of Education and Development using Information and Communication Technology*, 6(2), 84–105.
- Hawkins, B. L., & Oblinger, D. G. (2006). The myth about the digital divide: "We have overcome the digital divide. *EDUCAUSE Review*, 41(4), 12–13.
- Karsenti, T., & Collin, S. (2013). Usages des technologies en éducation: Analyse des enjeux socioculturels. *Éducation et francophonie*, 41(1), 192–210.
- Kuo, F.-Y., Tseng, F.-C., Lin, C., & Tang, W.-H. (2013). Critical success factors for motivating and sustaining womens ICT learning. *Computers & Education*, 67, 208–218. doi:10.1016/j.compedu.2013.03.006
- Kyriakidou, V., Michalakelis, C., & Sphicopoulos, T. (2011). Digital divide gap convergence in Europe. *Technology in Society*, 33(3-4), 265–270. doi:10.1016/j.techsoc.2011.09.001
- Landzelius, K. (2006). Introduction: Native on the Net. In K. Landzelius (Ed.), *Native on the Net: Indigenous and diasporic peoples in the virtual age* (pp. 1–42). London: Routledge.

- Losh, S. C. (2010). Generation, education, gender and ethnicity in American digital divides. In E. Ferro, Y. K. Dwivedi, J. R. Gil-Garcia, & M. D. Williams (Eds.), *Overcoming Digital Divides: Constructing an Equitable and Competitive Information Society*. Hershey, PA: IGI Global. doi:10.4018/978-1-60566-699-0.ch011
- Mardikyan, S., Yıldız, E. A., Ordu, M. D., & Şimşek, B. (2015). Examining the Global Digital Divide: A Cross-Country Analysis. *Communications of the IBIMA*.
- Martin, A. (2009). Digital Literacy for the Third Age. *eLearning Papers*, 12, 1-15.
- Martinovic, D., & Freiman, V. (2013). Digital skills development for future needs of the Canadian labour market. Final report to the Social Sciences and Humanities Research Council.
- Martinovic, D., & Magliaro, J. (2007). Openness in the context of computer networks: Educators' Perspective. *The International Journal of Technology. Knowledge in Society*, 3(3), 43-52.
- MCYS. (2012). *Stepping Stones: A Resource on Youth Development*. Retrieved from http://www.children.gov.on.ca/htdocs/English/topics/youthopportunities/steppingstones/youth_policy.aspx
- Minghetti, V., & Buhalis, D. (2010). Digital Divide in Tourism. *Journal of Travel Research*, 49(3), 267-281. doi:10.1177/0047287509346843
- Norris, P. (2001). *Digital Divide. Civic Engagement, Information Poverty, and the Internet Worldwide*. New York: Cambridge University Press. doi:10.1017/CBO9781139164887
- Oblinger, D. (2008). Growing up with Google: What it means to education. *Emerging Technologies for Learning*, 3(1), 11-29.
- Oblinger, D., & Oblinger, J. L. (2005). Educating the net generation. *EDUCAUSE*. Retrieved from <http://net.educause.edu/ir/library/pdf/pub7101.pdf>
- OECD. (2001). *Understanding digital divide*. Retrieved from <http://www.oecd.org/dataoecd/38/57/1888451.pdf>
- Pick, J. B., & Sarkar, A. (2015). *The Global Digital Divides: Explaining Change*. Springer.
- Pirbhai-Illich, F., Turner, K. C. N., & Austin, T. Y. (2009). Using digital technologies to address aboriginal adolescents education: An alternative school intervention. *Multicultural Education & Technology Journal*, 3(2), 144-162. doi:10.1108/17504970910967573
- Rekhari, S. (2009). Indigenous communities and new media: Questions on the global digital age. *Journal of Information. Communication & Ethics in Society*, 7(2/3), 175-181. doi:10.1108/14779960910955882
- Salman, A., & Rahim, S. A. (2012). From access to gratification: Towards an inclusive digital society. *Asian Social Science*, 8(5), 5-15. doi:10.5539/ass.v8n5p5
- Schrock, A., & Boyd, D. (2008). *Online Threat to Youth: Solicitation, Harassment, and Problematic Content. An Internet Safety Technical Task Force report. Berkman Center for the Internet and Society*. Harvard University.
- Sciadas, G. (2001). *The Digital Divide in Canada*. Ottawa: Statistics Canada. Retrieved from <http://www.statcan.gc.ca/pub/56f0009x/56f0009x2002001-eng.pdf>
- Sinclair, S., & Bramley, G. (2011). Beyond virtual inclusion – communications inclusion and digital divisions. *Social Policy and Society*, 10(1), 1-11. doi:10.1017/S1474746410000345
- Singleton, G., Rola-Rubzen, M. F., Muir, D., Muir, K., & McGregor, M. (2009). Youth empowerment and information and communication technologies: A case study of a remote Australian aboriginal community. *GeoJournal*, 74(5), 403-413. doi:10.1007/s10708-009-9277-6

Skaletsky, M. (2013). *Essays on the digital divide - explorations through global, national and individual lenses* (Order No. 3587811, Bentley University). ProQuest Dissertations and Theses, 217. Retrieved from <http://search.proquest.com/docview/1427340691?accountid=14592>

Steyaert, J., & Gould, N. (2009). Social Work and the Changing Face of the Digital Divide. *British Journal of Social Work*, 39(4), 740–753. doi:10.1093/bjsw/bcp022

Swain, C., & Pearson, T. (2001). Bridging the digital divide: A building block for teachers. *Learning and Leading with Technology*, 28(8). Retrieved from http://people.virginia.edu/~har4y/edlf345/elementary/files/digital_divide.pdf

The White House. (2013). *Implementing Executive Order 13616: Progress on Accelerating Broadband Infrastructure Deployment*. Retrieved from http://www.whitehouse.gov/sites/default/files/microsites/ostp/broadband_eo_implementation.pdf

Twist, J. (2005). *BBC News*. Retrieved from <http://news.bbc.co.uk/go/pr/fr/-/2/hi/technology/4445060.stm>

UNDP. (2007). *Human Development Report 2007/2008*. New York: UNDP.

van Deursen, A., & van Dijk, J. (2011). Internet skills and the digital divide. *New Media & Society*, 13(6), 893–911. doi:10.1177/1461444810386774

van Dijk, J. (2006). Digital divide research, achievements and shortcomings. *Poetics*, 34(4-5), 221–235. doi:10.1016/j.poetic.2006.05.004

Willms, J. D., & Corbett, B. A. (2003). *Tech and teens: Access and use*. Canadian Social Trends. Retrieved from <http://www.statcan.gc.ca/pub/11-008-x/2003001/article/6530-eng.pdf>

Wilson, E., & Wilson, E. (2000). *Closing the Digital Divide: an Initial Review. Briefing the President*. Washington, DC: Internet Policy Institute.

Wiseman, A., & Anderson, E. (2012). ICT-integrated education and national innovation systems in the Gulf Cooperation Council (GCC) countries. *Computers & Education*, 59(2), 607–618. doi:10.1016/j.compedu.2012.02.006

Zhong, Z. (2011). From access to usage: The divide of self-reported digital skills among adolescents. *Computers & Education*, 56(3), 736–746. doi:10.1016/j.compedu.2010.10.016

ADDITIONAL READING

Bonfadelli, H. (2002). The Internet and Knowledge Gaps – A Theoretical and Empirical Investigation. *European Journal of Communication*, 17(1), 65–84. doi:10.1177/0267323102017001607

Cleary, P. F., Pierce, G., & Trauth, E. M. (2006). Closing the Digital Divide: Understanding Racial, Ethnic, Social Class, Gender and Geographic Disparities in Internet Use among School Age Children in the United States. *The Information Society*, 4, 354–373.

Dibello, L. C. (2005). Issues in Education: Are We Addressing the Digital Divide? Issues, Access, and Real Commitment. *Childhood Education*, 81(4).

Dolničar, V. (2011). Measuring the Dynamics of Cross-national Digital Divides. In L. Haddon (Ed.), *The Contemporary Internet: National and Cross-National European Studies* (pp. 190–205). Frankfurt: Peter Lang.

Dolničar, V., Christou, C., Gannon, R., Haddon, L., Louca, S., Puga, P., & Vieira, J. (2011). Cross-national Broadband Digital Divides. In L. Haddon (Ed.), *The Contemporary Internet: National and Cross-National European Studies* (pp. 121–138). Frankfurt: Peter Lang.

Epstein, D., Nisbet, E., & Gillespie, T. (2011). Whos Responsible for the Digital Divide? Public Perceptions and Policy Implications. *The Information Society*, 27(2), 92–104. doi:10.1080/01972243.2011.548695

Godoy, S., & Gálvez, M. (2011). The Corresponding Digital Divide: Barriers and Facilitators of the Use of ICTs in Middle and Lower-middle Class Parents in Chile. *Revista CTS*, 18(6), 199–219.

Selwyn, N. (2004). Reconsidering political and popular understandings of the digital divide. *New Media & Society*, 6(3), 341–362. doi:10.1177/1461444804042519

Van Deursen, A., & van Dijk, J. (2014). The Digital Divide Shifts to Differences in Usage. *New Media & Society*, 16(3), 507–526. doi:10.1177/1461444813487959

KEY TERMS AND DEFINITIONS

“\$100 Laptop” Initiative: MIT project related to developing a cheap, yet powerful enough laptop suitable to outlive the harshest environmental conditions and provide connectivity and learning to children most affected by the digital divide and poor social conditions.

Digital Divide: Difference in opportunities that individuals, groups, and societies have in adequately accessing and using digital technologies.

Digital Inclusion: Providing equal opportunities to all to live successfully in the Digital Era.

Digital Literacy: Ability to use digital technologies appropriately for learning, working, and functioning in a modern society.

Global Digital Divide: Global disparities related to access to computing and information resources, mainly between developed and developing countries which results in less opportunities derived from such access.

Information and Communication Technologies: ICTs include personal computers, laptops, PDAs, cell phones, computer networks and the Internet.

Rural–Urban Gap: Usually as a consequence of a lacking interface and poor broadband conditions, some rural and remote areas may be digitally disadvantaged.

The Potential Role of the Software Industry in Supporting Economic Development:

S

Sherif H. Kamel

The American University in Cairo, Egypt

INTRODUCTION

Emerging information technologies provide a solid and promising platform to support economic development. In that respect, the information and communication technology (ICT) sector has been one of Egypt's strongest economic sectors with increasing prospects for job creation, productivity, scalability, impact and growth (Mahmoud, 2011). One of the subsets of ICT, the software industry, is gradually becoming more visible, stronger and mature in terms of output and impact both locally and beyond. Software is virtually embedded in every single aspect of our daily life with growing impact on individuals, organizations and society (Schroeder, 2013). In the information age, everything needs a software, one way or another. The software development industry is increasingly becoming global and consequently, there is a growing number of software companies that are providing their services overseas through outsourcing and offshoring. Software is arguably the best entry platform for developing nations, like Egypt, into the global ICT production complex (Heeks, 1998). The exposure to products and services from different global markets and the accompanying growing competition have effectively led to improved software products delivered coupled with an increasing pool of ICT professionals and developers who are well-educated, multilingual, and technology-savvy and at the same time interested to take the local software industry to the next level.

During the last two decades, pro-business government policies and regulations have helped grow the software industry and have encouraged a number of multinationals to invest in Egypt including, but not limited to, the likes of IBM, Intel, Cisco, Google, Microsoft, EMC, Valeo and Oracle. They have all invested substantially in setting-up facilities and growing their operations, and more importantly deciding to offer their services out of their Egypt-based operations. Moreover, more local companies and startups in the software development space have started their operations over the last decade trying to capture a segment of the regional and global market opportunities. This article assesses the developments in the software industry in Egypt and the potential role it can play as a contributor to economic development through the establishment of an export-oriented outsourcing software industry capitalizing on one of Egypt's primary resources, human capital.

BACKGROUND

The ICT sector in Egypt has been steadily growing since the late 1980s. During the period 2011-2012 given the political developments that took place in the country there was an apparent slowdown in all economic aspects, including ICT. However, in 2013, the minister of ICT expected the sector to go back to its double-digits growth by the end of 2014 to match the growth that was taking place during the period leading to the fiscal year 2009-

2010. Moreover, it would come at an ideal time as a solid platform and a strong and effective support mechanism to the overall economic revival of Egypt. For the long term, projections indicate that by 2017-18 annual growth rate will reach 20% and will account for 6% of the gross domestic product (GDP) and the sector will help create 500,000 jobs (Helmy, 2013). These projections naturally change on a regular basis. However, judging from the potential and trends in the marketplace, there is enough evidence that the software industry is probably on course to meet these set objectives.

The ICT industry includes different aspects such as people, hardware, and telecommunications. However, software represents an enabling platform linking all different aspects (Tayia & Wahba, 2001) and a possible area of growth and comparative advantage for Egypt. The key determining factor remains is human capital. Within the ICT industry, the innovation ratio in software development has been lower than that of hardware and telecommunications systems (Bozzetti, 1999) which led a number of developing nations such as India to position the development of a software industry as a strategic option (Tayia & Wahba, 2001). This was supported by the fact that the cost of establishing the software industry in developing nations is relatively low (El-Deeb, 2012). Therefore, since the late 1990s, Egypt has decided to follow that path and different stakeholders have collaborated to help establish a software industry capable to gradually become one of the primary support vehicles to economic development while emphasizing the export-oriented elements in the mix to help improve the nation's current massive and growing trade deficit. The importance of software is that it emerged as the digital brainpower of ICT becoming a platform for economic growth and as the driving force among different economic sectors (Nordhaus, 2000). In 2015, Egypt ranked 100 by the global innovation index among 141 countries worldwide indicating improvement compared to a few years back and promising to offer new developments and advancements in a

variety of sectors including the software industry (WIPO, 2015).

The developing world started to use and produce software in the 1950s and 1960s (Heeks, 1998). During the 1960s computing was introduced to Egypt. At that time, its primary use and applications were limited to the government and the public sector with some modest use within the private sector. During the 1980s, the diffusion of computing gained momentum and was widespread following the global personal computer (PC) evolution. Although computing started in Egypt over 50 years ago, it was only in 1985 that the active role played by the government caused a change in the way ICT was perceived as a vehicle for socioeconomic development and as a tool to improve the decision making process (Rizk & Kamel, 2013; Kamel, 1999). With the establishment of a ministry for communications and information technology (MCIT) in 1999, Egypt's information society initiative (EISI) was launched in 2001 to provide a broad perspective on the strategic plan for ICT diffusion in Egypt (Kamel, 2005).

The general perception indicates that the way developing nations will manage the computer driven processes of change will undoubtedly influence whether its socioeconomic development goals will be promptly, effectively and efficiently achieved. This has also been explicitly mentioned in the Millennium Development Goals (MDGs) and articulated in both meetings of the World Summit on the Information Society (WSIS) in 2003 and 2005. According to Kamel (2010) this could be achieved through "focusing on stimulating the growth of export-oriented activities as opposed to local infrastructure development where attention will be directed towards innovation as a primary driver for future growth." It is all about developing a vision, providing the right enabling environment and be driven by clear objectives supported by the required resources financially and more importantly human capital.

Since 2005, there has been various efforts and steps taken to position Egypt as an alternative

location for investment in intellectual capital as well as offering software development, business process outsourcing, business process offshoring and call center services. This has led Egypt to invest in accelerating its high-tech infrastructure by developing technology parks with over 130 multinational company in the areas of ICT, software development, outsourcing and call centers. This included the establishment of the Smart Village in Giza in 2001, followed by a similar smaller version a few years later in the suburb of Maadi in Cairo and most recently the Tahrir Alley Technology Park (TATP) was established at the Greek Campus of the American University in Cairo as a hub for tech startups. This is an area that can grow fast and that already saw the establishment of multiple startups in the space of IT. Moreover, the growing interest and operations of some of the global players in the call centers and outsourcing space in Egypt has been very encouraging. Sample of these companies includes Sykes, Stream, Convergence and Sutherland and more (Mahmoud, 2011). The number of companies, their capacities and services offered have been gradually increasing over the last decade, creating employment opportunities both directly and indirectly and supporting a variety of economic industries and sectors.

As one of the world's leading outsourcing destinations, Egypt is expanding its operation aiming to increase earnings and contributing to the economic growth as well as emphasizing higher-value services and niche segments (Oxford Business Group, 2012). Through the ICT sector, Egypt operates virtually everything from inbound call centers to multiple offshore functions and services. Sample multinationals operating in Egypt include the Oracle global product support center established in 2005, Stream Global Services, Sutherland Global Services and a finance and accounting base for IBM (Oxford Business Group, 2012). The call center industry is growing fast with companies like Xceed pioneering in that field in Egypt with branches in Morocco offering their services in nine different languages (www.xceedcc.com). Given the demographics of

the population in Egypt where the overwhelming majority (58%) are under the age of 25, the outsourcing industry provides youth with thousands of job opportunities on an annual basis, contributing effectively to the 800 thousand jobs that need to be created annually (www.idsc.gov.eg).

Table 1 demonstrates the current demographics of the ICT sector in Egypt. It is clear that there was an exponential increase in all indicators that was realized post Egypt's uprising in January 2011. This was caused by an overwhelming engagement of the nation's youth in learning, communicating, and exchanging ideas and starting-up IT companies and the creation of an environment that aims to promote freedom of expression, inclusion, entrepreneurial activities and more engagement of different constituencies in policy making and development at large. It is important to note that despite the economic challenges that Egypt is facing, one of the few sectors that managed to float and maintain its steady operation and in some ways grow is the ICT sector which reflects many positive signs of its potentials and how it can effectively contribute to the future of Egypt. The overall theme that has been gradually but forcefully disseminating across the community is three-folds including entrepreneurship, innovation and responsible business.

In 2007, MCIT formulated a strategic three-year plan for the ICT sector with a primary mandate to surpass 1 billion US dollars in terms of IT export revenues by 2010, that target has actually been realized with an additional 10% reaching 1.1 billion US dollars by 2011 largely due to the growth rate in the sector which had reached 14.9% in 2009 and the growing number of multinationals in the ICT sector that had expanded their businesses and/or outsourced their operations to companies in Egypt (Abdelazim, 2010). According to the Information Technology Industry Development Agency (ITIDA), there are plans that revenues should hit 10 billion US dollars by 2025. Undoubtedly, this could have been realized based on the growth rates that were realized in the years 2009 and 2010. Both the government and the

Table 1. Electronic readiness in Egypt

Indicators	Dec 2006	Dec 2008	Dec 2011	Dec 2012	Dec 2013	Dec 2014	Dec 2015	July 2016
Internet Subscribers	6 million	11.4 million	29.8 million	22.6 million	25.7 million	28.3 million	29.8 million	30.8 million
ADSL Subscribers	206,150	593,042	1.65 million	2.24 million	2.49 million	3.12 million	3.03 million	3.79 million
ADSL Penetration per 100 Inhabitants	0.35%	1.02%	2.31%	2.8%	3.21%	3.66%	4.42%	4.68%
Internet Penetration per 100 Inhabitants	8.25%	15.59%	34.83%	26.4%	29.4%	31.7%	32.6%	33%
Mobile Phones	18 million	38 Million	79 million	88.5 million	92.1 million	93.6 million	94.02 million	95.45 million
Mobile Penetration per 100 Inhabitants	23.07%	50.7%	97.93%	116.94%	118.19%	110.90%%	107.41%	108.23%
ICT Companies	2,211	2,621	4,250	3,750	3,860	4,005	4,110	4,250
Number of Employees in the ICT Sector	148K	174K	212K	216K	221K	223K	225K	226K

Source: Egypt ICT Indicators, 2015

business sector relentlessly continue to position Egypt as one of the key outsourcing destinations enticing the likes of Motorola, Orange Business Services, Videotron, HSBC and many others to either establish or expand their outsourcing and offshoring operations in Egypt.

Egypt's current plan given the relative saturation in the telecommunications infrastructure, is to focus on digital Arabic-language content, value-added services, digital services and software development (Kamel, 2010). In addition, with more than 800,000 university graduates annually, MCIT works with local companies and multinationals in the ICT sector to ensure that the required balance of skills and capacities are embedded in the curricula or through specialized educational programs to bridge the gap between the skills of Egyptian youth and the required competencies by the job market.

Since the establishment of Egypt's information society initiative in 2001 and based on models of public-private partnerships (PPP), there has been

a growing and effective role being played by the private sector through a win-win formula that is applied on a variety of ICT projects that related to a number of aspects included but not limited to software incubation and development (Kamel & Ahmed, 2006). In terms of the software industry, knowledge acquisition and transfer are key where 95% of the software development business is intangible capital (Grant, 2000; Hoch et al., 2000). Since the late 1980s, Egypt has been playing a leading role in software publishing in the Middle East and 80% of its software exports are regularly going to countries in the Gulf region and mainly Saudi Arabia (Fergany, 2002).

An Agile Software Industry with an Economic Impact

With a population of about 90 million people, Egypt is the most populous country in the Middle East North Africa (MENA) region with over 25% (over 23 million) of its population enrolled in

education programs, over 20 million represent its private sector workforce and around 7.1 million work for the government (www.idsc.gov.eg). The government of Egypt is more determined than ever to complete its national infrastructure and keep pace with the IT evolution worldwide. With the growing size of the global software market, Egypt has promising potentials to compete in the middle market segment of companies because it can not compete with the likes of India on the basis of prices. It can only compete based on the relatively low price of labor compared with the higher level of value added, business vision, and innovation (Rizk, 2002). Recently, focus was mainly on mobile applications for different sectors including but not limited to banking, health, education, trade, and industry as well as an increasing variety of services. The platform is a cutting-edge startup environment driven by passionate and technology-savvy youth opportunity. In addition, a growing number of software companies owned by Egyptian diasporas located in Europe and the U.S. are relying mainly on their software development activities on a pool of talented software developers located in Egypt.

The software industry in Egypt is growing and is attracting considerable investment from the private sector, in addition to the government incentives supporting the sector. Nevertheless, it is still in an early development stage covering four main market segments (Kamel, 2004) including software tools, packaged applications, customizations of existing applications and Arabization. The need for Arabic-language software remains highly needed especially with the size of the population in the region and its steady annual growth, which could represent a niche software companies (El-Kassas, 2005). The industry is labor intensive and faces opportunities to expand in global markets. There are various local, regional and global motives for Egypt to heavily invest in building an export-oriented software industry. Today, there are over 750 companies working in the software industry with over 18,000 professionals employed.

In 1999, Egypt supplied 70% of the demand for software from the Gulf region, primarily represented by the United Arab Emirates and Saudi Arabia (Kamel and Hussein, 1999). Multimedia-related products such as cultural, educational, entertainment and religious CDs are the leading products exported to the region. The majority of Egyptian software companies have established branches in United Arab Emirates, Saudi Arabia, Kuwait, Bahrain and Oman. Some are primarily working on outsourcing their services to the UK and the United States as well as other European and Far East countries. The government has a comprehensive program to support the software industry that includes a series of incentives such as formulating legislations offering five-year tax holiday for all software companies and 10-year tax holiday for the software companies establishing their operations in the established industrial zones. More is being done in that space especially supporting tech-startups and promoting entrepreneurship and innovation in the software development ecosystem.

The software industry is divided into four categories including (a) software tools, (b) packaged applications, (c) tailored applications and multimedia applications and (d) Arabization of applications. Software development companies range in size from 1 to 5 staff members' startups to relatively mature firms with around 50 to 150 employees. The majority of firms are located in and around the major cities including Cairo, Giza and Alexandria; where more than 60% of the population resides. However, recently, attracting talent has been growing across Egypt's different provinces with some of the promising ideas coming from Upper Egypt and attracting venture capital and angel investors to finance them. Diversity and pooling talents from across Egypt has been one of the elements of the recent developments in the software industry. Opportunities are also created in the various technology parks established to provide support in positioning Egypt as a global destination for outsourcing and offshoring, a

Table 2. Projections of IT exports by 2010

ICT Industry Element	Target Revenue (Million US\$)	CAGR	% of Total
Software and IT Services	280	27%	34%
Arab Content Management	200	46%	12%
Technical Support Centers	200	27%	24%
Contact Center Markets	115	50%	6%
R&D Engineering Design	100	38%	8%
Business Process Outsourcing Services	75	72%	2%
IT Product Exports (Emerging Markets)	45	25%	6%
IT Localization	40	22%	6%
Knowledge Process Outsourcing Services	30	43%	2%
Total	1,085		100%

Source: A. T. Kearney, 2005

position Egypt achieved in 2011 by being ranked 4th according to A. T. Kearney (2011). Table 2 demonstrates the projections of IT exports for 2010 as projected by A. T. Kearney.

Since 2005, the ICT sector, through public-private partnership programs have been focusing more on encouraging innovation and developments in ICT applications and in the development of the software industry with an emphasis on exports. Therefore, the government established four centers of excellence in data mining and computer modeling; wireless technologies, mobile and electronic services and electronic design. Egypt's dream is to become a leading software exporter in the MENA region with a global outreach which is relatively realistic considering the boom in the software industry in Egypt and the potentials in the regional marketplace with an increasing need for Arabization in a market with a population of 350 people and growing. This is also boosted by a growing trends in electronic commerce and the digital economy at large with all the associated opportunities created.

The competitive advantages of Egypt's domestic software production environment have attracted numerous international producers to subcontract programming of tailored applications. The industry was further boosted in 2004 with the establishment of ITIDA, a governmental entity established through Law 15, and aiming at paving the way for the diffusion of electronic business services and supporting an export-oriented IT sector. This was coupled with a number of strong attributes that placed Egypt as a potential center for offshore IT services such as a) favorable technical staff and infrastructure costs, b) language capabilities, c) improving infrastructure, d) pro-business governmental reforms, e) strong government support for the IT sector and f) a strategic geographic location. However, among the multiple challenges facing software development is the issue of intellectual property rights as the most critical barrier for the future growth of the information technology infrastructure in general (El-Kassas, 2005). Enforcement of existing laws should take place given the multimillion dollar losses incurred in the regional markets in general, Egypt included, on an annual basis. Further efforts need to be deployed at the institutional and implementation levels. The provision of laws without the proper follow-up and addressing the evolving problems has not met the expected levels yet (Corea, 2003).

An Industry-Based SWOT Analysis

The software industry in Egypt is better demonstrated through a SWOT analysis in order to understand where it stands and where it is heading with an overview of its strengths and weaknesses and an identification of the opportunities available and the threats faced as shown in Table 3.

FUTURE RESEARCH QUESTIONS

Based on the assessment of the software industry in Egypt, an action plan needs to be formulated in

Table 3. SWOT analysis of the software industry in Egypt

Strengths	Weaknesses
■ Young population that is tech savvy. ■ Educated graduates interested in ICT, entrepreneurship and startups. ■ Computer science majors in all public and private universities. ■ Training programs in IT applications. ■ Growing technical skills in ICT. ■ Low and competitive labor costs. ■ Good command of English for dealing with overseas customers (other languages include French, German and Spanish). ■ Same time zone advantage with Europe and provides a second-shift for the United States (outsourcing and offshoring services). ■ Government facilitating procedures and logistics related to the software industry.	■ Lack of sufficient expertise in any technology. ■ Need to improve expertise to compete in global markets. ■ Software companies spend 6+ months to prepare graduates for the marketplace (skills gap). ■ Lack of project management, communication marketing and managing startup skills. ■ Limited local demand for software. ■ Effective role of government is still limited. ■ Infrastructure level and cost is high compared to the capacities of different stakeholders. ■ Lack of government enforcement of regulations regarding piracy and intellectual property rights violations. ■ Perception that software has little intrinsic value among commercial and government customers.
Opportunities	Threats
■ Creation of more software business incubators-technology parks such as the smart village model. ■ Provision of internships and scholarships from software vendors both local and multinationals. ■ Promotion role played by software associations to activate the role of software development companies. ■ Changes in taxation and the introduction of new intellectual property laws. ■ Growing local, regional and global markets for diversified software development.	■ Intellectual copyright violations. ■ Piracy rates are relatively high. ■ Lack of market research in domestic and overseas markets that Egyptian companies could target. ■ Lack of financial support to the industry. ■ Limited outreach programs to penetrate the global software marketplace. ■ Other competing countries are rapidly developing their stature in the software industry space.

an attempt to capitalize on the opportunities available and overcome the challenges in the market. Such plan needs to be dynamic and iterative in a way to allow introducing changes based on market transformations and constant assessment of needs. The plan can assist Egypt in improving its software industry performance and cope with the global trends and drivers especially in the domain of outsourcing and offshoring. The plan could include issues such as identifying and penetrating global markets to realize growth; investing in people to leverage the technical skills and capacities in the industry; introducing business-incubator programs to link industry to educational institutions and the government; improving the infrastructure to ensure that firms have the vehicle to operate their business; and, increasing the government support role to demonstrate the value of the software development industry to Egypt's economic development (Kamel, 2003).

The plan should attempt to answer some of the research questions that could be addressed

as the software industry develops and grows and that includes whether developing nations can earn billions from software exports? Whether developing nations are relatively less expensive locations in terms of infrastructure, infostructure and humanware for software development? Whether software piracy is negatively affecting the production of software in developing nations? Whether the research and development efforts support the organization of a growing and an agile software industry? If the human capital in emerging economies have the ability to compete on a global scale? In addition to many more questions that address the different elements that relate to the proper development, sustainability and scalability of the software industry.

CONCLUSION

Egypt has an excellent opportunity to establish a relatively small but effective software industry.

The efforts and support enabled by the government, industry, financial institutions and the educational sector will ultimately determine the level of development and success of the industry hopefully making it an effective contributor to Egypt's economic development. The initial required resources to become a leading player in the regional market and possibly establish a global reputation as a software development hub are available in terms of technical capacities, skills and educational infrastructure. However, there are a number of issues that need to be addressed including the unbalanced distribution of qualified human resources, the lack of business non-technical experiences and skills needed to compete in global markets and the need for more support from different financial institutions primarily for startups. In addition, there needs to be a long-term strategy aimed at reconfiguring the educational ecosystem to cater for the growing needs of information and knowledge-based societies and that should be coupled with programs and incentives to minimize the possible industry-related brain drain.

Based on the development of the software industry globally and the potential for the software industry in Egypt; there is a number of issues that need to be restructured to benefit from opportunities available both in Egypt and other countries in the region. These issues relate to enabling better linkages and coordination between industry and academia; formulating an overall assessment mechanism; improving the education and learning ecosystem to deliver the capacities required by local and global markets; improving the industry salary scale to relatively match other regional and global markets in order to minimize the implications of brain-drain situations; setting-up a regulatory authority for the software industry, and excelling globally through offshore and outsource business development and contact centers.

REFERENCES

- Abdelazim, H. (2010). In A. Jeffreys (Ed.), *A growing presence, Interview the CEO of the Information Technology Industry Development Agency*. London: Oxford Business Group.
- Bozzetti, M. (1999). The technological evolution of ICT and standards. In EITO (Ed.), *European information technology observatory* (pp. 80-161). Mainz: Eggebrecht-Press.
- Corea, M. (2003). Managing the provision of knowledge: the design of intellectual property laws. In I. Kaul, P. Conceicao, K. Le Goulven, & R. Mendoza (Eds.), *Providing global public goods, managing globalization* (pp. 410-415). New York: Oxford University Press. doi:10.1093/0195157400.003.0017
- Egypt I. C. T. Indicators. (2015). *Egypt ICT Stats*. Retrieved October 28, 2015, from <http://www.egyptictindicators.gov.eg>
- El-Deeb, A. (2012). A view on the state of software engineering practice: experience from Egypt. *ACM SIGSOFT Software Engineering Notes*, 37(6), 1-4.
- El-Kassas, S. (2005). *Challenges and opportunities for a competitive software industry in Egypt*. WIPO/LAS national seminar on intellectual property for journalists and members of the media, Cairo, Egypt.
- Fergany, N. (Ed.). (2002). *Arab human development report*. New York, NY: United Nations Development Program, Regional Bureau for Arab States.
- Grant, R. M. (2000). Shifts in the world economy: the drivers of knowledge management. In C. Despres & D. Chauval (Eds.), *Knowledge horizons: the present and promise of knowledge management* (pp. 27-54). Philadelphia, PA: Butterworth-Heinemann. doi:10.1016/B978-0-7506-7247-4.50005-7

- Heeks, R. (1998). *Myths of software development in developing countries*. Manchester, UK: University of Manchester.
- Helmy, A. (2013). Information technology in Egypt set to blossom, speech at the American Chamber of Commerce in Egypt. *Business Monthly*, 30(4), 54.
- Hoch, D. J., Roeding, C. R., Purkert, G., Linder, S. K., & Muller, W. R. (2000). *Secrets of software success: management insights from 100 software firms around the world*. Boston: Harvard Business School Press.
- Information and Decision Support Center. (2015). *Egypt Profile*. Retrieved November 10, 2015, from <http://www.idsc.gov.eg>
- Kamel, S. (1999). Information technology transfer to Egypt. In D. F. Kocaoglu, & T. Anderson (Eds.), *Proceedings of the Portland International Conference on Management of Engineering and Technology (PICMET) Technology and Innovation Management: Setting the Pace for the Third Millennium* (pp. 567-571). Portland, OR: Portland State University.
- Kamel, S. (2003). The implications of the digital economy on a growing digital divide in developing nations. In M. Farag (Ed.), *Proceedings of the 8th The American University in Cairo Research Conference on Globalization Revisited: Challenges and Opportunities* (pp 60-70). Cairo: The American University in Cairo Press.
- Kamel, S. (2004). The software industry in Egypt. In M. Khosrowpour (Ed.), *Encyclopedia of Information Science and Technology* (pp. 2836-2840). New York: IGI Global.
- Kamel, S. (2005). Assessing the impacts of establishing an Internet Cafe in the context of a developing nation. In M. Khosrowpour (Ed.), *Proceedings of the 16th Information Resources Management Association International Conference on Managing Modern Organizations with Information Technology* (pp. 176-181). New York: IGI Global.
- Kamel, S., & Ahmed, H. (2006). The impact of eReadiness on eGovernment in developing nations - case study of Egypt. In M. Khosrowpour (Ed.), *Proceedings the 17th Information Resources Management Association (IRMA) International Conference on Emerging Trends and Challenges in Information Technology Management* (pp. 728-731). New York: IGI Global.
- Kamel, S., & Hussein, M. (1999). Internet in Egypt: a tool for development, *Proceedings of the 2nd BITWorld Conference on Business Information Technology Management: the Global Imperative*. Cape Town: BITWorld.
- Kearney, A. T. (2005). *Assessment of Egypt's potential as a competitive location for offshore services, Market Study Report*. Chicago, IL: A. T. Kearney.
- Kearney, A. T. (2011). *Global Services Location Index, Global Index Report*. Chicago, IL: A. T. Kearney.
- Mahmoud, R. (2011). A strong double-digit performance in 2010. *Business Monthly*, 30(1), 43-44.
- Nordhaus, W. D. (2000). *Policy rules in the new economy*. Presentation for the discussion on the new economy, sponsored by the congressional budget committee and the senate budget committee.
- Rizk, N. (2002). Information technology and growth: will the software industry lead Egypt into a new economy? *Proceedings of the Middle East Economic Association, Topics in Middle Eastern and North African Economies Journal*, 4.
- Rizk, N., & Kamel, S. (2013). ICT and building a knowledge society in Egypt. *International Journal of Knowledge Management*, 9(1), 1-20. doi:10.4018/jkm.2013010101
- Schroeder, C. (2013). *The Entrepreneurial Revolution Remaking the Middle East: Startup Rising*. New York: Palgrave MacMillan.

Tayia, A., & Wahba, K. (2001). A system dynamics model for developing the software industry: the case of Egypt. *Proceedings of the international conference of system dynamics society*.

World Intellectual Property Organization. (2015). *Global Innovation Index Report*. Author.

Xceed. (2015). *Company profile*. Retrieved November 15, 2015, from <http://www.xceedcc.com>

ADDITIONAL READING

Fergany, N., Ali, A. A. G., Al-Imam, M. M., Corm, G., El-Kholy, O., Kanaan, T., & Reda, M. G. et al. (2002). *Arab Human Development Report* (p. 180). New York: United Nations Development Programme, Regional Bureau for Arab States.

Gramos, M. (2006). *Market analysis, Egypt IT services 2006-2010 forecast and 2005 vendor shares*. Framingham, MA: IDC.

Harvard Consulting Group. (1999). *Sector Assessment of the Egyptian Software Industry*. Cairo: Egyptian Software Association.

Heeks, R. (1996). Promoting software production and export in developing countries. In E. Roche & M. Blaine (Eds.), *Information Technology, Development and Policy* (pp. 77–94). Aldershot: Avebury.

International Telecommunication Union. (2006). The next India? Egypt's software dream. Website, retrieved June 10, 2011, from <http://www.itu.int>

Kamel, S., & Hussein, M. (2008). Xceed—pioneering the contact center industry in Egypt. *Journal of Cases on Information Technology*, 10(1), 67–91. doi:10.4018/jcit.2008010105

Loch, K., Straub, D., & Kamel, S. (2003). Diffusing the internet in the Arab world: The role of social norms and technological culturation. *IEEE Transactions on Engineering Management*, 50(1), 45–63. doi:10.1109/TEM.2002.808257

Pyramids Research. (2000). Communications market in Egypt. Middle East Communications Market Report, March, pp. 30–34.

Rizk, N., & Kamel, S. (2012). ICT strategy development: Public-private partnerships – the case of Egypt. *International Journal of Strategic Information Technology and Applications*, 3(2), 70–90. doi:10.4018/jsita.2012040105

Seleim, A., Ashour, A., & Khalil, O. (2005). Knowledge acquisition and transfer in Egyptian software firms. *International Journal of Knowledge Management*, 1(4), 43–72. doi:10.4018/jkm.2005100103

KEY TERMS AND DEFINITIONS

Arabization: The transformation of software applications into the Arabic language in terms of usage as well as interface to be able to cater for a community that stands in 2013 at around 370 million people.

Building Blocks: Reflects all the critical success factors of the information and communication technology industry and that include: hardware, software, human resources “humanware”, networking and information “infostructure”.

Business Process Outsourcing: Means the use of outsourcing mechanisms locally or offshore for the design and/or development of software applications.

Information Technology Outsourcing: Reflects the outsourcing of computer or Internet related work, such as programming to other individuals or companies.

Information Technology Startups: They are the small and medium-sized enterprises and startups that are focused on information technology applications and services.

Diffusion of Information Technology: It reflects the spreading of information technology concepts among the society of implementation whether that could be within an organization or within the community at large.

Government-Private Sector Partnership:

The teaming of different entities in the government and the private sector to realize a change and a transformation in the development of information technology at large and in the software industry in specific.

Incubator Programs: A form of collaboration between industry, corporations, the business community and the educational sector through the availability of “space” that is used to formulate solutions that meet market needs and consequently help create employment opportunities especially among the youth.

Informatics Projects: The projects that involve in any way possible the use, design, delivery, implementation and management of information technology irrespective of the elements involved including software, hardware, telecommunications and others.

Information Technology Industry: The accumulation of all elements of information technology design, delivery and management coupled with support services and other human capital elements.

Smart Village: The concept reflects a model technology park represented by a landscape comprising office space designed for different development and service-oriented offerings of information and communication technology companies; thereby giving each the benefit of economies of scale; usually located outside the city.

Software Industry: Focuses on the needs of the software development industry in terms of infrastructure, know-how, capacities, research and development and lifelong learning.

Tailored-Applications: Applications based on industry or organizational needs to complement the off-shelf software applications available in the marketplace.

S

Socio–Economic Processes, User Generated Content, and Media Pluralism

Androniki Kavoura

Technological Educational Institute of Athens, Greece

INTRODUCTION

Media pluralism that is associated with the presentation of different points of view and ideas in media content, consists of a recurrent issue of discussion because it is in close relation with the socio-economic, technological and political environment within which media operate. Major developments take place with the worldwide economic instability that exists but also with the advancement of new technologies and social media that enhance online communication. International organizations such as the European Union have recently created a monitoring tool about media pluralism for European member states to utilize. It is at members' states own discretion the extent to which they will enforce media pluralism that at times relates to socio-economic and technological processes within their boundaries. In other geographic frameworks, such as Latin America, political issues influence media diversity and limit the possibility of different voices to be heard.

This paper aims (a) to present media pluralism's components and (b) to describe the reciprocal role between media pluralism's dimensions and communication technologies, socio-economic and political processes that take place. These processes are based on the technological advancements of new media platforms such as online video platforms and streaming services as media types for online media content. In doing so, this chapter adds to the literature by illustrating the role of the abovementioned processes on media pluralism. Media pluralism is under continuous negotiation. Current real-world scenarios are discussed that

aim to illustrate the dimensions of media pluralism and are of value to professionals in the media.

Which is the case when information and communication technologies allow for new online content material to be uploaded and become viral? Does this safeguard democracy in media and ensure that all voices are heard? The cases that are presented in this chapter as far as media content is concerned, are mainly taken from electronic media, television and press. Alternative approaches are presented as recommendations in need to take place for media pluralism and how online communication and new technologies can guarantee such expression of multiple points of view. These are renegotiated issues that the agents involved with the production and distribution of media content, -policymakers, media managers, marketers, firms-, can contribute so that media pluralism prevails. The creation and distribution of media content that presents different points of view, -minorities, underrepresented groups-, is affected by the rapid development and substantial impact of technological and communication changes. The chapter argues that diversity in media content may be influenced in periods of economic recession that a country may face. In addition, political and social reasons may influence the advertising budgets spent that in turn have an effect on media pluralism.

The next sections briefly describe media pluralism's dimensions and the concept of concentration of media ownership in relation to the technological advancements, user generated content created online as alternative media content, socio-cultural and economic processes in order to provide the framework and current trends. The use of exam-

ples aims to illustrate the way advertising media planning may be also influenced due to limited diversity in media content. Presentation of cases from countries where socio-economic and political processes and technological advancements may influence the media content diversity allows for recommendations to take place with the provision of alternative ways that enforce media pluralism. Recommendations to agents involved in the implementation of media pluralism are provided and suggestions for further research are given.

The researcher searched well-known databases which were Science Direct, Scopus, Emerald, Eric, E-List for available literature and search engines such as Google Scholar and Science Direct. The specific keywords under examination were the ones which were associated with media pluralism, media diversity, social media, user generated content and pluralism in the current socio economic processes.

BACKGROUND OVERVIEW: MEDIA PLURALISM

The definition of the concepts involved in the creation of the argumentation of this chapter entails media pluralism that consists of the umbrella that embraces aspects of media diversity. Media pluralism refers to more than one, in fact, numerous points of view and a variety in media content so that many people and groups can be represented (Valcke, Sukosd, & Picard, 2015).

Media pluralism is associated with and influenced by the presence of a wide framework of ownership and merger control requirements, a range of content in programming via different media types and genres, the existence of a variety of sources that contribute to the media content, viewpoint diversity and freedom of expression, independence of public service broadcasting in the content and program services provided, representation of local and regional interests, cultural expressions, limited or non-existence of clientelism relations

between media and political actors (K.U. Leuven et al., 2009a). The abovementioned dimensions when followed and implemented by agents, aim to provide access to information sources to all people. Pluralism though, continues to be the object of political contestation because it is influenced by the environment within which it operates, where continuous processes take place and the countries' specificities need to be taken into consideration from time to time (Leandros, 2010; Brogi & Dobrev, 2014). Towards this end, media pluralism needs to be holistically approached.

Valcke et al. (2015) categorize media pluralism as 'external' and 'internal' where the former has to do with the multiplicity of media outlets, the existence of many media owners so that media concentration does not take place and multiplicity in the operation of media in regard to the way expression of opinions and different points of view takes place; the latter, 'internal' pluralism is associated with the activities implemented by a broadcaster or a publisher to make every possible effort to produce a variety of and different media content material.

Within the European Union, a study was commissioned titled Media Pluralism Monitor (MPM) to develop a monitoring tool for providing member states with the possibility of assessing media pluralism and justifying whether or not pluralism exists in countries according to specific dimensions: media ownership, cultural, political and geographic diversity, regulatory framework, media types and genres (K.U. Leuven et al., 2009a). In 2013, the European Parliament awarded another grant to continue the implementation of MPM so that member states can be included (Brogi & Dobrev, 2014). At the time of writing this chapter, an ongoing research takes place on MPM. Although measures and a regulatory framework exists, each member state may act independently on the enforcement of specific legislation for plurality at national level. Thus, advocates of plurality and diversity argue that there is freedom of will that may safeguard media pluralism. Habermas

(1989) argued for the public sphere indicating the principled pluralism that guarantees freedom of public opinion, freedom of speech, freedom of press and the absence of constraints in expression. This takes place in environments where democratic institutions prevail. Nonetheless, in reality, different agents, -advertisers, media producers, politicians and others, may make use of the creation and distribution of media content based on the connections they have with media (McNair, 2011), thus, influencing the end result.

At the other side of the coin, competition and the free market may prevail. This, though, may be harmful for pluralism because it does not allow for differentiation in media content and thus, leads to reduction in diversity content (K.U.Leuven, 2009b). Competition is associated with the free will of all to participate in a free market provided they have the means to participate so that everyone can play with the rules of the market. Issues of agents' ownership, market, competition, exclusion, marginalization may take place and we may speak of an uncontrolled free market systems (Gantzias, 2014; Vamvakas, 2006). In this way, unfair competition is created, the media with the most powerful force and greater financial standing dominate while there are television stations that have the power and can influence public opinion even if web TV broadcasts. In press, a similar situation may take place even if the advancements of new technologies and digital communication allow for an easier dissemination of communication messages.

The element of commercialization generates communication associated with the dominance of advertising and the disparity in communicative resources (Vamvakas, 2006; Masouras, 2015). Media content uses the results of surveys of audience measurement and other marketing and communication tools that influence the viability of these programs based on the number of people that follow the program. Whatever cannot be sold, is not broadcasted and everything depends on such results.

CONCENTRATION OF MEDIA OWNERSHIP

Concentration of different media ownership under one owner is a dimension that may influence media pluralism. It usually takes place after mergers and acquisitions take place and brings forth the issue of representativeness and freedom of expression in media content. Lack of different points of view may be the result of such concentration because production and distribution from one or different media is gathered in one's hands and control over information may take place. Decrease in independent media owners' diversity, brings forth decrease in representation of multiculturalism, in the form of limited local community news; advertising in the related themes is also another consequence (Hiller, Savage, & Waldman, 2015).

Thus, high result of media ownership concentration is a dimension that may influence content diversity and advertising expenditure. When there is high concentration, program and/ or content media homogeneity may take place. This is further translated in decrease in sales of advertising time and space to advertisers. Whatever is saleable, survives and representation of different points of view may not be heard since the small group that may be interested in these points do not add up financial numbers for the big organizations. Media concentration may thus, influence content diversity with regard to program content and can be a threat to the freedom of expression (Leandros, 2010). It may also lead to a decrease in audiences' viewership due to their disinterest in homogeneity and further, a vicious circle is created because the advertising budgets are minimized. High ownership levels influence source diversity and existence of different opinions. Ownership consolidation reduces diversity and research on 1600 editorial positions in 25 top newspapers in USA illustrated stability in divergences of opinions after major mergers and acquisitions took place in the media sector (Ho & Quinn, 2009).

USER GENERATED CONTENT, COMMUNICATION TECHNOLOGIES AND MEDIA PLURALISM

Traditional media such as television and press may address a large audience in a short period of time on a daily basis (Belch & Belch, 2014). The advent of new communication technologies and the presence of electronic media, those operating namely online such as social networking sites that create and deliver content online, adds more material to the already existing media sources. Social media are associated with mobile and web applications that promote online user generated content that is created and is uploaded by users (Kaplan, 2015). Social media contain a wide range of online, word of mouth created content that is presented on blogs, Internet discussions, social networking sites, boards and forums (Mangold & Faulds, 2009). This allows for diversity of content to exist since content is presented online by different media.

Social media, allow for the interactive user feedback from digital devices and traditional media seem to lose ground in regard to the websites and online platforms. In order to have a more holistic approach, communication strategies that connect both traditional and online media take place. There is connection of the print era with the digital one for example and this is implemented taking into consideration tools that measure the medium's effectiveness and data analytics (Schlesinger & Doyle, 2015) to see whether combining traditional and new media is an effort worthy to be followed. The combination of print and digital content consists of a new business model that is implemented in different countries and is of particular significance especially when issues about countries' definition as independent states exist (Scotland, Quebec, Catalonia) (Schlesinger & Benchimol, 2015). Minorities and all voices need to be heard when the print press shrinks and digital press provides the forum to do so with the user generated content that is created.

Technological advancements in information and communication technologies have enabled the creation of online material that is widely spread in the digital world and is easy to access. They allow the presence of new content creators who emerge and distribute content in different media platforms and social networking sites that are 24 hours available (Leuven et al., 2009b). Media type information plays a significant role in media consumption that should be taken into account in combination with the fact that people consume differently on social networking sites (Choi, 2016).

New media platforms and streaming video services via video platforms may have an effect on the TV and video firms (Gimpel, 2015). Such implementation via information and communication technologies, allows for the consumption of media content at a different pace than that with the traditional media (Choi, 2016). A new e-culture is created where viewers may download videos from Apple iTunes via a search engine and a file transfer system, news content and TV programs from their mobile devices. 'TV everywhere' is being introduced (Waterman, Sherman, & Ji, 2013). Advertising agencies and media content producers need to take these challenges into account.

At the same time, technological developments can be a threat to pluralism (K.U. Leuven, 2009a). While more ideas and points of view exist in regard to media content, people tend to focus on ideas that are similar to those they share, creating groups of like minded people (Ntalianis, Kavoura, Tomaras & Drivas, 2015). Thus, pluralism does not really operate even if information and communication technologies offer ground to so in the sense that different points of view are not really heard. Online users can be engaged within a process of myopic dialogue and connect with those ideas and points of views that initially shared.

Deregulation, caused by the freedom of markets and the advent of new technologies that contributes so that every communication message can be on air or rather online, is not without problems. Freedom of expression allows for

many different voices to be heard, nonetheless, more players and new content come to the table. This, though, means, that the advertising pie will be divided in different market shares, a pie that diminishes because the new groups that will consume the programs are small and thus, not worth the advertising budget according to the financial estimates of the organizations. Having said that, diversity of sources and resources may influence the allocation of the advertising budget leading to its shrinkage. Papathanasopoulos argued for the role of advertising as a factor that influences what is broadcasted since new content is created on the basis that it can be sold, thus, influencing content diversity (Papathanasopoulos, 2005). New media have multiplied the creation of user generated content and is broken into small pieces.

User generated content can be uploaded in online communities, online websites, blog posts, social media profiles. Apart from the combination of print and digital as a new business communication model presented above, another business communication model that has been recently adopted since user generated content has made a powerful presence in media, is the integration and creation of a dialogue around the online user generated content material that is created by a competitive media channel or platform for a specific media program. This may act as a response to the existence of content diversity in media where viewership of user generated content of competing cable news shows in USA was found to be significant (Sabnis & Grewal, 2015).

SOCIO CULTURAL PROCESSES AND MEDIA PLURALISM

Content diversity in media based on socio-cultural and language issues of a geographic area consists of another important factor that may influence media plurality. Media programmes and their content may reflect the special language needs of different groups. Culture may act as a mechanism of social integration and community cohesion

in modern European cities (Kavoura & Bitsani, 2014). If we take into consideration that diversity may be in need due to the migrations that take place in Europe from Asian countries, then, we may realize the significance for the creation of media content that caters for those groups' needs. Migration is not a new phenomenon, yet, the tendency not to represent migrants in media exists in some countries. News associated with migration is depicted as something negative. Taking Poland as an example, 1% of its population is Polish, yet, stories and news coverage associated with different ethnicities are not covered (Triandafyllidou & Ulasiuk, 2015).

Best cases of cities like Barcelona and Copenhagen that aim to incorporate matters in a bilingual way so that everyone is represented (Guidikova, 2015; Schlesinger & Benchimol, 2015) are worth adopting. Issues of legislation need to be acknowledged, taking into account European institutions such as the European Broadcasting Union (Triandafyllidou & Ulasiuk, 2015).

In regard to Greece, the only recognized minority by the Greek state is a religion one (the Muslim population of Thrace) although in the Greek national elections in 2015, all three Parliament positions were won by Muslim candidates. Nonetheless, there are still not TV channels of national coverage recognized as minority media in Greece (Brogi & Dobрева, 2014). The above-mentioned minority is not represented, there is space for heterogeneity in the media content to be created, this though, has not taken place in the Greek media environment, yet.

The political system also influences media pluralism. Taking into consideration the action of banning of 2 channels from broadcasting one month before the parliament elections as an effort to silence opposition media in another continent, Asia in the country of Turkey (Toksabay & Solaker, 2015) is typical to note that based on political reasons, diversity on such TV content, news programs for example, is limited. Political diversity and media policies go hand in hand with the political stability and freedom of media that

in turn influences diversity in media content. Cultural interference in media pluralism may be also found in the People's Republic of China, where Facebook, one of the most popular social network websites, is currently banned (Riley, 2014).

The technological advancements and user generated content can be used as a platform of expression of free opinions nonetheless, democratic procedures are not always safeguarded. Costa Rica, Latin America, consists of another case, that illustrates how politicians may intervene in the media content since it did not have media pluralism for a period of 10 years -from 2002 till 2011 (Guerrero, 2015).

Geographical and geopolitical issues associated with a country may also influence the way media diversity is enforced. Taking the Cypriot Democracy as an example, - that is united and divided at the same time, half Cypriot, half Turkish due to the Turkish invasion in 1974-, Cypriot media life has been influenced and thus, programming as well. The official languages of the Cypriot Democracy are the Greek and Turkish (article 3, paragraph 1) and programs are transmitted in both Greek and Turkish (article 171, paragraph 1) based on the 1960 Cypriot Institutional legislation. Another typical example is the rapid expansion of Hispanics in the United States that illustrates how multiculturalism is a necessity to be taken into account since there are 45 million Hispanophones who speak Spanish as a first or second language in the United States (Shumow & Vigon, 2015). Examining the issue of language in regard to whether there is media content that represents these people's points of view and the need to be informed in in their own language, allows for media pluralism to exist and flourish.

Activities that deserve to be promoted and implemented by media are those similar to the ones initiated by Google. In late 2015 and following the migration wave of people from Asia to Europe, Google has asked people to contribute to a multilingual site with useful information for refugees and migrants (Fuller, 2015). Applica-

tions exist and can be implemented with the use of mobile devices for the defense of freedom of expression and of rights of all people.

ECONOMIC CONSIDERATIONS AND MEDIA PLURALISM

Media pluralism is also in close connection with the economic environment within which media content is created and distributed. Different kinds of programs need to represent the choices of different groups of people and this is an asset of the pluralistic approach of a society. This may not be the case in periods of economic recession due to high media production costs that in turn limit the range of content programming but may also control the independence of public service broadcasting.

In Greece, that is under deep economic crisis, local private channels broadcast the same content with that of national private channels. Thus, representation of local and regional interests is far from being catered. The New Greek TV, Radio and Internet Broadcast (NERIT), the Greek national public service media station that aims to promote different genres and types in its program, does not always succeed in acting independent due to its limited financial budget and funding that mainly comes from the Greek state. In a period of socio-economic crisis, allocation of resources is limited and thus, quality of content is underrepresented. Lack of financial resources to allocate to public service media is associated with lack of variety content programming. Thus, phenomena like the one described below may take place: the teaser for Mundial in NERIT's athletic program titled "our national team on the left (foot) in Mundial", - that was on air when the Greek team lost-, was negatively criticized by the president of the Greek Parliament from the leftist party Ms Konstantopoulou who mentioned that the use of the word 'left' implied negative connotations for the leftist party she belongs (T.X., 2015). This consists of an obvious attempt to censor freedom of speech and

media content as a time when NERIT's funding comes from the annual Greek budget.

Economic factors may thus, influence new productions where copy costs in producing media content require huge investments that do not exist (K.U.Leuven, 2009a).

RECOMMENDATIONS TO AGENTS INVOLVED IN THE IMPLEMENTATION OF MEDIA PLURALISM

Media pluralism needs the contribution of all to succeed its implementation. Interactions between market agents, policymakers, media are necessary.

- Policymakers at national and international level need to enforce regulatory frameworks to safeguard media pluralism for the media content that is created.
- Media channels and producers need to enforce freedom of expression. Productions that are created need to respect minorities and all points of view while media synergies and alliances can be further explored to deliver media content based on the latest and emerging technologies.
- Media professors may enforce the code of ethics in communication among potential media managers.
- Firms that operate in environments where socio-economic and political instability exists that may influence media pluralism may make the necessary arrangements to ensure that all points of view are heard. Sponsoring activities by means of financing productions and groups for the creation of multilingual TV programmes or the digital representation of different ideas regarding minorities and other underrepresented groups is an activity towards this end.
- Users of online media need to be active and create user generated content that represents many different points of view.

MEDIA PLURALISM: WHERE NEXT? FUTURE RESEARCH DIRECTIONS

Even if democracy exists in a country, it is not definite that the regulatory framework will prevail towards safeguarding media plurality. Inability and unwillingness by politicians and populations to institute reliable regulative mechanisms in public and private sectors (Gantzias, 2014). Research needs to take place at national level in countries that undergo socio-economic and political crisis at to the way plurality is understood within such frameworks.

Further research should take place on the initiation of television content diversity to represent migrants and/or minorities. Issues of language and diversity but also of multiculturalism need to be taken into consideration. Research is welcome for different ethnicities such as the anticipated creation of the federal state consisted of Cypriots and Turks to add to the literature in regard to issues of media content, language and association of ideas of those populations involved. At the time of writing this chapter, the peripheral parliament in Catalonia, Spain approved a 'solemn resolution' declaring the start of a formal process leading to independence (Buck, 2015). Thus, multiple cases offer ground form further research to take place.

Furthermore, future research is also encouraged based on the populations of migrants and refugees who move from one country to live in other countries, have money to spend, yet, the existing multicultural issues need to be acknowledged. Marketers should know the environment and special needs in order to be able to effectively manage their advertising campaigns.

Media regulatory policies are necessary to be implemented. The specificities of media markets of relatively young media markets within the EU members (K.U.Leuven, 2009a) or the socio-economic and cultural crisis that countries may undergo, brings forth the significant issue of the role of the environment and the effect it has on media pluralism. Therefore, research on these

markets may also add to the literature on media plurality.

Video platforms can be an alternative way for generation of media content and media types. Pay TV services offer the possibility to the viewer to stream premium video, an issue that should be further examined.

The incorporation of the competitors' comments as user generated content to programs on television was found to increase the viewership (Sabnis & Grewal, 2015). Programs could take into consideration such a contemporary communication strategy that creates a dialogue and defends media diversity by adding competitors' points of view to their own material.

Combining media content, print and digital, as a new business communication model (Schlesinger & Doyle, 2015; Schlesinger & Benchimol, 2015), with the contribution of information and communication technologies, can be another alternative communication strategy for preserving media pluralism and this needs to be further researched. Research on advertising media planning that incorporates the combination of traditional and online media content to its strategic decisions is also in need rather than examining them separately, in contrast one to the other.

CONCLUSION

Technological advancements in regard to the creation of media content and user generated content that is available online, offer new opportunities so that web technologies can guarantee freedom of expression of many different voices.

This chapter adds to the existing literature on media communication by highlighting emerging trends that bring media pluralism's dimensions under continuous social negotiation. The socio-economic and political environment needs to be considered to safeguard that the constituent elements and dimensions of media plurality are implemented under the continuous renegotiation that takes places.

REFERENCES

- Belch, G., & Belch, M. (2014). The role of new and traditional media in the rapidly changing marketing communications environment. *International Journal of Strategic Innovative Marketing*, 1(3), 131–137.
- Brogi, E., & Dobрева, A. (2014). *Monitoring media pluralism in Europe-Testing and implementation of the media pluralism monitor 2014*. European University Institute: CMPF. Retrieved October 13, 2015, from <http://cmpf.eui.eu/Documents/MPM2014-PolicyReport.pdf>
- Buck, T. (2015). How the Catalonia vote will ramp up tensions with Madrid. *Financial Times*. Retrieved November 11, 2015, from <http://www.ft.com/intl/cms/s/0/ca99506c-86d5-11e5-90def44762bf9896.html#axzz3rCYX2BAF>
- Choi, J. (2016). Why do people use news differently on SNSs? An investigation of the role of motivations, media repertoires, and technology cluster on citizens news-related activities. *Computers in Human Behavior*, 54, 249–256. doi:10.1016/j.chb.2015.08.006
- Euromonitor International. (2015). Retrieved October 15, 2015, from <http://www.euromonitor.com/greece>
- Fuller, J. (2015). *An update of our response to the refugee and migrants crisis*. Retrieved October 23, 2015, from <https://googleblog.blogspot.in/2015/10/update-response-refugee-migrants.html>
- Gantzias, G. (2014). The info-communication globalisation and digital tax payments culture: the Greek crisis and the global info-cash. In N. Petropoulos & G. Tsobanoglou (Eds.), *The debt crisis in the Eurozone. Social Impacts* (pp. 481–509). Cambridge Scholars.
- Gimpel, G. (2015). The future of video platforms: Key questions shaping the TV and video industry. *International Journal on Media Management*, 17(1), 25–46. doi:10.1080/14241277.2015.1014039

- Guerrero, M. A. (2015). Latin America Media: The Challenges to Pluralism. In P. Valcke, M. Sukosd, & R. Picard (Eds.), *Media Pluralism and Diversity: Concepts, Risks and Global Trends* (pp. 211–225). Basingstoke, UK: Palgrave. doi:10.1057/9781137304308_12
- Guidikova, I. (2015). Intercultural integration: A new paradigm for managing diversity as an advantage. In R. Zapata-Barrero (Ed.), *Interculturalism in Cities: Concept, Policy and Implementation* (pp. 136–151). Edward Elgar. doi:10.4337/9781784715328.00015
- Habermas, J. (1989). *The structural transformation of the public sphere. An inquiry into a category of bourgeois society* (T. Burger & F. Lawrence, Trans.). Cambridge, MA: MIT Press.
- Hiller, R., Savage, S., & Waldman, D. (2015). Market Structure and Media Diversity. *Economic Inquiry*, 53(2), 872–888. doi:10.1111/ecin.12153
- Ho, D., & Quinn, K. (2009). Viewpoint diversity and media consolidation. An empirical study. *Stanford Law Review*, 61(4), 781–868.
- Kaplan, A. M. (2015). Social Media. In D.T. Cook & J.M. Ryan (Eds.), *The Wiley Blackwell Encyclopedia of Consumption and Consumer Studies* (pp. 1-5). Wiley. doi:10.1002/9781118989463.wbeccs214
- Kavoura, A., & Bitsani, E. (2014). Investing in culture and intercultural relations for advertising and sustainable development of the contemporary European city within the framework of international city branding and marketing: The case study of the Intercultural Festival of Trieste, Italy. In L. Barreau (Ed.), *Advertising: Types of Methods, Perceptions and Impact on Consumer Behavior* (pp. 35–66). Nova Publishers.
- Leandros, N. (2010). Media concentration and systemic failures in Greece. *International Journal of Communication*, 4, 886–905.
- Leuven, K. U. (2009a). *Independent Study on Indicators for Media Pluralism in the Member States-Towards a Risk Based Approach. Final Report*. European Communities, Katholieke Universiteit Leuven, Leuven. Retrieved October 15, 2015 from https://ec.europa.eu/digital-agenda/sites/digital-agenda/files/final_report_09.pdf
- Leuven, K. U. (2009b). *Independent Study on Indicators for Media Pluralism in the Member States-Towards a Risk Based Approach. Final Report-Annex 1, User Guide*. European Communities, Katholieke Universiteit Leuven, Leuven. Retrieved October 15, 2015 from https://ec.europa.eu/digital-agenda/sites/digital-agenda/files/user_guide_09.pdf
- Mangold, W., & Faulds, D. (2009). Social Media. The new hybrid element of the promotion mix. *Business Horizons*, 52(4), 357–365. doi:10.1016/j.bushor.2009.03.002
- Masouras, A. (2015). *Understanding Competition and Diversity in Television Programming*. Hamburg, Germany: Anchor Academic Publishing.
- McNair, B. (2011). *Introduction to Political Communication* (5th ed.). London: Routledge.
- Ntalianis, K., Kavoura, A., Tomaras, P., & Drigas, A. (2015). Non-Gatekeeping on Social Networks: A Reputation Monitoring Approach. *Journal of Tourism and Services*, 6(10), 19–44.
- Papathanasopoulos, S. (2005). *Politics and media: the case of Southern Europe*. Athens: Kastaniotis. (in Greek)
- Riley, C. (2014). *Banned! 8 things you won't find in China*. Retrieved October 1, 2015, from <http://money.cnn.com/gallery/technology/2014/12/30/banned-china-8/3.html>
- Sabnis, G., & Grewal, R. (2015). Cable news wars on the Internet: Competition and user generated content. *Information Systems Research*, 26(2), 301–319. doi:10.1287/isre.2015.0579

Schlesinger, P., & Benchimol, A. (2015). Small nations, the press and the digital challenge. *Media Culture & Society*, 37(1), 101–106. doi:10.1177/0163443714553561

Schlesinger, P., & Doyle, G. (2015). From organizational crisis to multi-platform salvation: Creative destruction and the recomposition of news media. *Journalism*, 16(3), 305–323. doi:10.1177/1464884914530223

Shumow, M., & Vigon, M. (2015). *News Diversity and Minority Audiences: Using Real Simple Syndication (RSS) to assess the democratic functions of Spanish-language media in the digital age*. Journalism Practice. doi:10.1080/17512786.2015.1014668

Toksabay, E., & Solaker, G. (2015). *Turkey bans TV channels close to Erdogan foe ahead of polls*. Retrieved October 8, 2015, from <http://www.reuters.com/article/2015/10/08/us-turkey-politics-media-idUSKCN0S217A20151008#GmlFHgCOx1xEZjeS.97>

Triandafyllidou, A., & Ulasiuk, I. (2015). Assessing the capacity of the media to reflect diversity and promote migrant integration. In R. Zapata-Barrero (Ed.), *Interculturalism in Cities: Concept, Policy and Implementation* (pp. 152–162). Edward Elgar. doi:10.4337/9781784715328.00016

T.X. (2015). *Zoi's Konstantopoulou negative comment on the title of NERIT for Mundial*. Retrieved October 1, 2015, from <http://www.contra.gr/Columns/contra-epithesi/greece/epekrine-titlths-nerit-gia-to-moyntial-h-zwh-kwnstantopouloy.3492313.html>

Valcke, P., Sukosd, M., & Picard, P. (2015). *Media Pluralism and Diversity: Concepts, Risks and Global Trends*. Basingstoke, UK: Palgrave. doi:10.1057/9781137304308

Vamvakas, V. (2006). *Elections and Communication in the after the political regime era. Politics and Sights*. Athens: Savvalas. (in Greek)

Waterman, D., Sherman, R., & Ji, S. W. (2013). The economics of online television: Industry development, aggregation, and TV everywhere. *Telecommunications Policy*, 37(9), 725–736. doi:10.1016/j.telpol.2013.07.005

ADDITIONAL READING

Evens, T., & Donders, K. (2016). Television distribution: Economic dimensions, emerging policies. *Telematics and Informatics*, 33(2), 661–664. doi:10.1016/j.tele.2015.06.012

Lin, T. C., & Oranop, C. (2016). Responding to media convergence. Regulating multi-screen television services In Thailand. *Telematics and Informatics*, 33(2), 722–732. doi:10.1016/j.tele.2015.07.005

Trenz, H. J., Ruzza, C., & Guiraudon, V. (2015). *Europe's Prolonged Crisis. The making of the unmaking of a Political Union*. London: Palgrave MacMillan.

KEY TERMS AND DEFINITIONS

Concentration of Media Ownership: Ownership of different media usually but not necessarily after mergers and acquisitions that brings forth the issue of representativeness and freedom of expression and lack of presence of different points of view because production, distribution from one or different media is gathered in one's hands and control over information may take place.

Content Diversity: Production of a variety of material for media consumption that is consisted of a variety of content that is different one from the other.

Media Pluralism: Safeguarding representation of different points of view so that numerous people and opinions have an equal chance to be heard, read, broadcasted on media.

Online Video Platform: A media type service that allows users to search, file, upload video content on the Internet that requires subscription.

Social Media: Ways of interacting among people using web online information and communication technologies.

Socio Economic Processes: Related to social and economic issues, identifying social and economic needs within a community and the creation of strategies to address those needs.

User Generated Content: Any form of online content that is created and distributed by consumers and users.

Category S

Sociology

Bipolar Model in Collective Choice

Ayeley P. Tchangani

Université Fédérale Toulouse Midi-Pyrénées, France

INTRODUCTION

Collective choice is a sub-domain of decision analysis (a discipline comprising the philosophy, theory, methodology, and professional practice necessary to address important decisions in formal manner according to Wikipedia) that addresses decision making problems where a certain number of decision makers must select a subset (possibly reduced to a singleton) of alternatives from a large set of potential alternatives in order to achieve some collective as well as individual objectives, preferences, or desires. Such decision making problems are encountered in many practical situations such as management, engineering, economics, social, politics etc., see for instance (Bouyssou *et al.*, 2000), Steuer (1986), and references therein for some real world applications even though in these references the problems are most of the time treated as a single decision maker problems. The existence of many decision makers necessitates to have a *coordination mechanism* (how to aggregate the view points of all decision makers) to address collective choice problems. The purpose of this chapter is to address such coordination mechanism. Using bipolar analysis that consists in evaluating alternatives by two opposite measures (a measure gathering positive aspects of the alternative and that resuming its negative aspects) with regard to pursued objectives at individual level as well as at community level permits in some extent to embed human attitude into the decision process.

The remainder of this chapter is organized as follows. In the second section a background on (collective) choice problems resolution methods will be recalled; then the main focus of this chapter will be presented in third section; section four will be dedicated to the main contribution of

this chapter: bipolar framework for modeling and solving collective choice problem; section five presents a sketch of future directions researches; a conclusion is presented in section six and finally references, additional reading, and some keys terms and definitions end the chapter.

BACKGROUND

In political science, methods for realizing a collective choice (mapping individual preferences onto collective preferences) are dominated since the advent of democracy by simple majority voting process (Picavet, 1996). But many theoretical results such as that of Borda, see (Borda, 1781), Arrow impossibility theorem (Arrow, 1951) show that this way of aggregating individuals preferences can lead to inconsistency. In decision analysis, that actually does have many steps such as formulating decision goal or objectives, identifying attributes that characterize potential alternatives that can respond to the decision goal and making recommendation regarding these alternatives given the decision goal, choice is the final step. But to choose, one must evaluate first; the construction of an evaluation procedure, often carried up by an expert known in the literature as the analyst (Bouyssou *et al.*, 2000) is an important step in the decision process; this step is the main purpose of this chapter. This construction consists in aggregating individual preferences, understood in a broad sense to obtain a way that permits to rank, at least partially, different potential alternatives. Classically, two main approaches have dominated evaluation process in modern decision analysis: value or utility type approach (a value function or an utility measure is derived for each

DOI: 10.4018/978-1-5225-2255-3.ch633

alternative to represent its adequacy with decision goal), see for instance Steuer (1986) and Saaty (1980); outranking methods (a pair comparison of alternatives is carried up under each attribute or criteria to derive a pre-order over the alternatives set), see (Bouyssou *et al.*, 2000), (Brans *et al.*, 1986, 1986a). The approach that will be described in this chapter can be considered as an intermediate one compared to those two approaches evoked previously; indeed by using numerical values to evaluate alternatives look like utility type approach, but as two “opposite” measures are used, it permits incomparability as it is the case in outranking approaches.

MAIN FOCUS OF THE ARTICLE

In many situations, collective decision making is made through a majority voting process where each decision maker casts a ballot for only one alternative and the alternative that obtains the maximum of voices is considered as the community choice. But voting process does not capture in our opinion all attitudes of human beings such as ambiguity, indecision, social values consideration, etc. Indeed, decision makers often face uncertainties (impossibility of decision makers to clearly express their objectives, to elicit and assess attributes, etc.) and interactions (a decision maker may be influenced by other decision makers when expressing his or her judgment). Furthermore, French mathematician Jean-Charles de Borda and other have noticed since 18th century that in an election where the winner is the candidate who got the majority of votes and where there are more than 3 candidates, candidate who obtains the majority of voices is not necessarily the preferred one by the majority of voters. In this chapter we adopt an approach that highlights *bipolarity notion* between all components of collective decision analysis problem. We are motivated by the fact that cognitive psychologists have observed for long time that human evaluate alternatives by considering separately their positive aspects and

their negative aspects, see for instance (Caciopo & Berntson, 1994) and (Osgood *et al.*, 1957). To this end, we introduce *supporting* and *rejecting* notions (Tchangani, 2010) that relate attributes to objectives leading to an evaluation model in terms of two measures or indices (selectability and rejectability) for each alternative in the framework of satisficing game theory (Stirling, 2003) so that a decision maker can be in position of not being able to discriminate between two alternatives. These notions permit to partition attributes set into three subsets given an objective: attributes that support this objective, attributes that reject this objective and attributes that are neutral with regard to this objective; of course only supporting and rejecting attributes are interesting for evaluation process. Selecting and rejecting degrees of an attribute with regard to an objective may be assessed using known techniques such as analytic hierarchy process (AHP), see (Saaty, 1980) or any method that could assign a measure to an attribute with regard to a pair of objective and alternative. This model allows alternatives to be characterized by heterogeneous attributes (at individual level) and possibly different attributes from an individual to another; it allows also incomparability between alternatives in terms of Pareto-equilibrium, see (Pareto, 1896). Collective decision making situations with such issues are pervasive in real world applications; for instance a government evaluating projects that belong to different domains such as health, infrastructures, social, economics, etc. with the main objective to enhance a country developing process (Tchangani, 2015). In such situations, though attributes characterizing projects may be completely different, the important thing is their adequacy with regards to the pursued objectives, so that alternative projects can be ultimately compared on the same basis (decision maker’ desires). The social influence between decision makers and decision makers’ attitude will be taken into account through different degrees such as concordance/discordance degrees within the group, selfishness degree, risk averse degree, see (Tchangani, 2014). When making decision in complex situation, it

is rather rare that a human reach a final decision without hesitation; most of the time there exist some dubitative attitude, some indecision between several possible alternative decisions; this attitude is carried up in this chapter through satisficing game (Stirling, 2003).

BIPOLAR APPROACH FOR SOLVING COLLECTIVE CHOICE PROBLEM

Modeling Process

Formerly, collective choice problem is described by following elements: a set U of potential alternatives from which an alternative or a subset of alternatives must be selected by a group of decision makers; a generic decision maker is designated by i and a generic alternative by u ($u \in U$). In the bipolar analysis framework, evaluation process consists in determining for each alternative $u \in U$, two measures known as *satisfiability* measures: namely a *selectability* measure denoted $\mu_{S_i}(u)$ derived by decision maker i or $\mu_S(u)$ derived by the community of decision makers or stakeholders and a *rejectability* measure denoted by $\mu_{R_i}(u)$ derived by decision maker i or $\mu_R(u)$ derived by the community. The modeling process adopted in this chapter goes from individual to global, global satisfiability measures $\mu_S(u)$ and $\mu_R(u)$ will be obtained by aggregating individual satisfiability measures $\mu_{S_i}(u)$ and $\mu_{R_i}(u)$ over all decision makers. Locally each decision maker i may wish to satisfy own objectives or desires; to this end he or she will use a set of features, criteria or attributes to evaluate the adequacy of an alternative with regard to his or her objectives; these sets may be common to all actors for some particular choice problems but it is not an obligation for the choice procedure being developed in this chapter. We propose to elicit satisfiability measures in two steps: firstly, each decision maker i will derive its categorical

(that is without taking into account the existence of other members) satisfiability measures $\mu_{S_i}^0(u)$ and $\mu_{R_i}^0(u)$ for each alternative $u \in U$ using BOCR analysis for instance. The convergence of *supporting/rejecting* notions and uncertainty leads to an analysis framework known as BOCR analysis, see (Tchangani, 2010; Tchangani, 2015) that may constitute an interesting alternative method for elicitation and evaluation. In this framework the set of attributes characterizing an alternative with regards to an objective is divided into four subsets: benefit subset (B) (certain attributes of alternative that support the objective); opportunity subset (O) (uncertain attributes of alternative that support the objective); cost subset (C) (certain attributes of alternative that reject the objective); and risk subset (R) (uncertain attributes of alternative that reject the objective). Final evaluation of alternative u with regards to objective o will be obtained by aggregating separately positive aspects in terms of benefit and opportunity and negative aspects in terms of cost and risk to measure how well is this alternative with regard to that objective and opposite forces in the realization of that objective by this alternative respectively. Let us denote by $b_k^o(u)$, $o_k^o(u)$, $c_k^o(u)$ and $r_k^o(u)$ the normalized values of benefit, opportunity, cost and risk attribute k of alternative u with regard to objective o ; and by $b^o(u)$, $o^o(u)$, $c^o(u)$, and $r^o(u)$ vectors gathering these normalized benefit, opportunity, cost and risk measures respectively. Let suppose that decision maker is able to supply for these components normalized relative degrees vectors in terms of ω_b , ω_o , ω_c , and ω_r ; then the overall benefit, opportunity, cost, and risk measures of alternative u for objective o are obtained by aggregating over the corresponding attributes set using Choquet integral associated to a weighted cardinal fuzzy measure for instance, see (Tchangani, 2014), as shown by the expressions of Table 1.

In Table 1, $C_\omega^{wefm}(x)$ stands for Choquet integral of numerical vector x associated to a

Table 1. BOCR components for objective o

$B^o(u) = C_{\omega_b}^{wcfm}(b^o(u))$
$O^o(u) = C_{\omega_o}^{wcfm}(o^o(u))$
$C^o(u) = C_{\omega_c}^{wcfm}(c^o(u))$
$R^o(u) = C_{\omega_r}^{wcfm}(r^o(u))$

weighted cardinal fuzzy measure with relative importance weights vector ω of elements to aggregate, see Table 2 and (Tchangani, 2013).

Indeed, given the synergy created by grouping attributes by category, Choquet integral with a weighted cardinal fuzzy measure is well suited as an aggregation operator. Let us denote by $b(u)$, $c(u)$, $c(u)$ and $r(u)$ vectors gathering benefit, opportunity, cost and risk measures with regard to all objectives and by ω^o , the relative importance vector of objectives then the overall opinion regarding the alternative u is captured by the benefit measure $B(u)$, opportunity measure $O(u)$, cost measure $C(u)$, and risk measure $R(u)$ will be obtained similarly by aggregation as shown in Table 3.

Categorical selectability measure $\mu_{S_i}^0(u)$ and categorical rejectability measure $\mu_{R_i}^0(u)$ of alter-

native u in the point of view of decision maker i are finally obtained by normalizing over the alternatives set U , the measures

$$\phi_i B(u) + (1 - \phi_i) O(u)$$

and

$$(1 - \phi_i) C(u) + \phi_i R(u),$$

as shown by expressions of Table 4, see (Tchangani, 2014).

Parameter $0 \leq \phi_i \leq 1$ in Table 4 can be interpreted as the risk averse index of actor i ; indeed, this index permits to adjust the attitude of a decision maker toward uncertainty; for instance a risk averse decision maker, for who $\phi_i \rightarrow 1$, will balance immediate benefit (B) with potential harm (R) regardless of potential benefit (O) and immediate cost to pay (C).

In collective decision analysis situation, there is always some *conditionality* in the sense that the preferences of a given decision maker may be conditioned to that of other decision makers because when making decision, a decision maker may be influenced by its social vicinity positively or negatively. Indeed, influence does not mean positive perception (altruism, deference) of

Table 2. Choquet integral associated to a weighted cardinal fuzzy measure

Given a n dimension numerical valued vector x and a relative importance vector ω of the elements of x, the Choquet integral $C_{\omega}^{wcfm}(x)$ of x associated to a weighted cardinal fuzzy measure with relative importance vector ω is given by $C_{\omega}^{wcfm}(x) = \sum_{k=1}^n \left\{ \left(\frac{n - (k-1)}{n} \right) \left(\sum_{j \in A_k} \omega_j \right) (x_{\sigma(k)} - x_{\sigma(k-1)}) \right\}$ where σ is a permutation over the elements of x given by: $x_{\sigma(1)} \leq x_{\sigma(2)} \leq \dots \leq x_{\sigma(n)} \text{ and by convention } x_{\sigma(0)} = 0$ and A_k is defined by: $A_k = \{\sigma(k), \sigma(k+1), \dots, \sigma(n)\}$
--

Table 3. BOCR components for alternative u

$$\begin{aligned} B(u) &= C_{\omega_o^{wcfm}}^{wcfm}(b(u)) ; O(u) = C_{\omega_o^{wcfm}}^{wcfm}(o(u)) \\ C(u) &= C_{\omega_o^{wcfm}}^{wcfm}(c(u)) ; R(u) = C_{\omega_o^{wcfm}}^{wcfm}(r(u)) \end{aligned}$$

the influenced decision maker toward influencing one; it may represent repulsion or aggressiveness. One can admit that if two decision makers are in discordance then the conditioned selectability of one should be proportional to the categorical rejectability of the other and in the case they are in concordance it must be proportional to the categorical selectability of the other decision maker; a similar observation holds for the rejectability, see (Tchangani, 2014). So let denote by $\mu_{S_i/V(i)}$ and $\mu_{R_i/V(i)}$ the conditional selectability and rejectability of decision maker i given his position with regard to his social vicinity ($V(i)$) and by $\mu_{S_{V(i)}}^0(u)$ and $\mu_{R_{V(i)}}^0(u)$ the vector of categorical selectability measures and rejectability measures of all member of social vicinity of this decision maker. The measures $\mu_{S_i/V(i)}(u)$ and $\mu_{R_i/V(i)}(u)$ are therefore obtained by aggregation of categorical satisfiability measures and normalization as given in Table 5.

In expressions given in Table 5, the attitude of decision makers is taken into account with regard to his/her vicinity through relative concordance weights vector ω_i^c and discordance weights

Table 4. Categorical satisfiability measures

$$\begin{aligned} \mu_{S_i}^0(u) &= \frac{\phi_i B(u) + (1 - \phi_i) O(u)}{\sum_{v \in U} \{ \phi_i B(v) + (1 - \phi_i) O(v) \}} \\ \mu_{R_i}^0(u) &= \frac{(1 - \phi_i) C(u) + \phi_i R(u)}{\sum_{v \in U} \{ (1 - \phi_i) C(v) + \phi_i R(v) \}} \end{aligned}$$

vector ω_i^d ; see (Tchangani, 2014). Weighting parameter $0 \leq \gamma_i \leq 1$ can be interpreted as the altruist degree of decision maker i ; indeed, $\gamma_i \rightarrow 1$ means the total altruist behavior of decision maker i as he does not consider the contribution of his discordance attitude in the formation of his conditional satisfiability measures whereas $\gamma_i \rightarrow 0$ corresponds to the case where decision maker i behaves aggressively as he considers only his discordance attitude about the opinions of the members of his vicinity.

The ultimate selectability measure $\mu_{S_i}(u)$ and rejectability measure $\mu_{R_i}(u)$ of the alternative u in the opinion of decision maker i are corrected as shown by expressions of Table 6 where $0 \leq \delta_i \leq 1$ is the selfishness degree of decision maker i .

Final choice procedures will be derived from the ultimate and individual satisfiability measures; basically two main procedures may guide how to reach a collective choice: reasoning from an “ag-

Table 5. Conditional satisfiability measures

$$\begin{aligned} \mu_{S_i/V(i)}(u) &= \frac{\gamma_i C_{\omega_i^c}^{wcfm}(\mu_{S_{V(i)}}^0(u)) + (1 - \gamma_i) C_{\omega_i^d}^{wcfm}(\mu_{R_{V(i)}}^0(u))}{\sum_{v \in U} \{ \gamma_i C_{\omega_i^c}^{wcfm}(\mu_{S_{V(i)}}^0(v)) + (1 - \gamma_i) C_{\omega_i^d}^{wcfm}(\mu_{R_{V(i)}}^0(v)) \}} \\ \mu_{R_i/V(i)}(u) &= \frac{\gamma_i C_{\omega_i^c}^{wcfm}(\mu_{R_{V(i)}}^0(u)) + (1 - \gamma_i) C_{\omega_i^d}^{wcfm}(\mu_{S_{V(i)}}^0(u))}{\sum_{v \in U} \{ \gamma_i C_{\omega_i^c}^{wcfm}(\mu_{R_{V(i)}}^0(v)) + (1 - \gamma_i) C_{\omega_i^d}^{wcfm}(\mu_{S_{V(i)}}^0(v)) \}} \end{aligned}$$

Table 6. Individual satisfiability measures

$$\begin{aligned}\mu_{S_i}(u) &= \delta_i \mu_{S_i}^0(u) + (1 - \delta_i) \mu_{S_i/V(i)}(u) \\ \mu_{R_i}(u) &= \delta_i \mu_{R_i}^0(u) + (1 - \delta_i) \mu_{R_i/V(i)}(u)\end{aligned}$$

gregation” of individual satisfiability measures or from individual short lists obtained locally using individual procedures.

Solution Procedures

Given the idea of going from individual preferences to community preferences considered here, to reach a community level solution one can consider two possibilities: obtaining an aggregated satisfiability measures and then analyzing the selection problem as a single decision maker problem, or obtaining short lists of local satisficing equilibrium alternatives and then trying to converge to a community satisficing alternative(s).

Aggregated Satisfiability Measures Approach

Let denote by $\mu_S(u)$ and $\mu_R(u)$ the aggregated selectability and aggregated rejectability measures of alternative u obtained from individual measures of all decision makers. Arguing that selected or rejected alternative by two decision makers is sounder than selected or rejected alternative by one decision maker appeals for a synergetic aggregation scheme such as Choquet integral considered in previous sections. Once these measures are obtained, the analysis process in order to reach a final decision can rely on indicators built upon a value function

$$\pi(u) = v(\mu_R(u), \mu_S(u))$$

such as: maximum discriminant,

$$\pi(u) = \mu_S(u) - q\mu_R(u);$$

maximum boldness, $\pi(u) = \frac{\mu_S(u)}{\mu_R(u)}$; maximum selectability index or minimum rejectability index, $\pi(u) = \mu_S(u)$ or $\pi(u) = \frac{1}{\mu_R(u)}$ respectively.

Convergence From Local Analysis

Instead of aggregating individual satisfiability measures to reason on a single decision maker basis, one may consider that each decision maker i comes with his own selected subset of satisficing equilibrium alternatives $\mathcal{E}_{q_i}$ obtained by balancing his selectability and rejectability measures up to a caution or boldness index q_i (Tchangani, 2014). If it happens that a common alternative belong to all subsets, it constitutes the consensus alternative and the choice problem is solved. But this will constitute an exceptional situation and in most cases, local best alternatives will differ from one decision maker to another so that an adequate mechanism is needed to reach a final decision. In this case, one possibility is to define a score $v(u)$ over these individual selected subsets; some possible such scores are given in Table 7 where 1_Ω is the indicator of the set Ω given by

$$1_\Omega(x) = \begin{cases} 1 & \text{if } x \in \Omega \\ 0 & \text{if } \text{not} \end{cases}$$

Table 7. Global score for alternatives

$$\begin{aligned}v(u) &= \prod_{i=1}^n \left\{ \frac{\mu_{S_i}(u)}{\mu_{R_i}(u)} \right\} \\ \text{or} \\ v(u) &= \sum_{u \in \Sigma_{q_i}} \left\{ 1_{\Sigma_{q_i}}(u) \right\}\end{aligned}$$

It may happen (for sake of simplicity for instance) that decision makers be disposed to choosing by a voting process; this case is considered in the following subsection.

Choice by Voting: BORDA Revisited

Choice by voting generally consists in a simple procedure where each decision maker gives his voice to only one alternative and at the end the alternative that obtains the maximum of voices will be declared the winner. But this procedure does have some limitations; indeed, in his paper entitled « *Sur les élections au scrutin* » published in 1781 in « *Mémoires de l'Académie Royale* », see (Borda, 1781), the French mathematician Jean-Charles de Borda has shown that, in a majority election where there are more than 3 candidates, candidate who obtains the majority of voices is not necessarily the preferred one by the majority of voters. To overcome this situation, he proposed to choose the elected candidate using a merit order: each voter will rank candidates according to his preference by assigning, if there are m candidates, the score $m-1$ to the most preferred candidate, the score $m-2$ to next most preferred and so on until the score 1 to the last preferred candidate. So if there are n voters and denoting by $\omega_i(j)$ the score assigned to candidate j by voter i , then the global score or merit $\omega(j)$ of candidate j is obtained by adding all of its scores as given by the expression $\omega(j) = \sum_{i=1}^n \omega_i(j)$ and the elected candidate will be that who got the most important merit score.

But in practice, for a given candidate, a particular voter may have positive opinion for some aspects of that candidate and negative opinion for other aspects. So Borda's voting procedure means that each voter must do a balance of his opinions (positive and negative) in order to obtain a sort of global merit score for each candidate before voting. But it is well known that it is easier to aggregate homogeneous aspects (positive for instance) than heterogeneous aspects (combining

positive and negative aspects simultaneously). So we propose, building on the bipolar notion, a voting process according to the following scheme: voter i will rank all candidates according to their positive aspects in his opinion to obtain a global positive merit score $\omega_i^P(j)$ for candidate j and then rank them according to their negative aspects to obtain a global negative merit score $\omega_i^N(j)$. The final choice process can be done using two different procedures: selection from individual choice (a global merit score $\omega(j)$ for candidate j is obtained similar to expressions of Table 7) and selection from aggregated scores (voters opinions are aggregated separately to obtain for a given candidate j , a global positive merit score $\omega^P(j)$ and a global negative merit score $\omega^N(j)$ and then obtain the global score by balancing $\omega^P(j)$ and $\omega^N(j)$).

Decision regarding which procedure must be used in a particular case, how to aggregate positive and negative opinions separately as well as how to make trade-off between such aggregated opinions, are upper level and law oriented issues that regulate life in community such as a constitution in democratic nations.

FUTURE RESEARCH DIRECTIONS

Decision making approach described in this chapter falls into soft computing paradigm. If bipolar approach presented in this chapter permits to reach possibly robust solution, there remains, nevertheless, possible improvements regarding methodologies and modeling tools in order to reach usability of this framework in solving real world complex problems. Here are some possible improvements directions.

- **Context:** How decision makers and/or experts view positive and negative aspects of an alternative on one hand and their attitude towards risk on other hand may depend on their personal situation as well as

the behavior of decision environment that we refer to as the context. Indeed, preferences depend upon psychological attributes of the person who judges; therefore to dispose of a framework that is as close as possible to human way of deciding, this context component should be considered in modeling stage.

- **Dynamics:** As the context sensitive consideration mentioned above, attitude of decision makers and/or experts may vary from one instant to another so that a same problem may be viewed differently by the same decision maker from an instant to another. This dynamicity should be considered in the modeling process to allow parameters variation at each instant; mainly if a computer assisted tool to support the decision process must be developed.
- **Sensitivity Analysis:** Considering context and actors psychological attributes and interactions appeal for developing a way to do a sensitivity analysis with regards to these parameters. This sensitivity analysis may result in realizing scenarios analysis and answering questions like “what if?” and/or “is it possible?”.

CONCLUSION

Collective choice problem as a decision making problem where possibly antagonistic preferences of many actors or stakeholders have to be taken into account is considered in this chapter. Indeed, the main purpose of this chapter has been to propose a modeling process of collective choice that cope as much as possible with individual behavior in early stages of the decision process, collectivity issues being postponed as much as possible in later stages. The proposed approach of this chapter relies on the cardinal concept or notion of bipolarity at the individual as well as community level. This concept consists in evaluating (individually or collectively) each potential alternative or can-

didate by two measures: a measure that reflects positive aspects of the considered alternative and a measure that gathers its negative impacts with regard to the pursued goal(s). It has also been shown that even classical choice using voting (by merit as proposed in 18th century by Borda) process can be formulated in this framework to integrate individual preferences as much as possibly into the community choice process.

REFERENCES

- Arrow, K. (1951). *Social Choice and Individual Values*. John Wiley & Sons.
- Borda, J.-C. (1781). Sur les élections au scrutiny. *Mémoires de l'Académie Royale*, 657-664.
- Bouyssou, D., Marchant, T. P., Pirlot, P. M., Tsoukiàs, A., & Vincke, P. (2000). Evaluation and Decision Models: A critical Perspective. Kluwer Academic.
- Brans, J. P., Mareschal, B., & Vincke, Ph. (1986). PROMETHEE: A new family of outranking methods in multicriteria analysis. *Operations Research*, 84, 477-490.
- Brans, J. P., Mareschal, B., & Vincke, Ph. (1986a). How to select and how to rank projects: The PROMETHEE method. *European Journal of Operational Research*, 24(2), 228-238. doi:10.1016/0377-2217(86)90044-5
- Caciopo, J. T., & Berntson, G. G. (1994). Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates. *Psychological Bulletin*, 115(3), 401-423. doi:10.1037/0033-2909.115.3.401
- Osgood, C. E., Suci, G., & Tannenbaum, P. H. (1957). *The Measurement of Meaning*. Chicago: Univ. of Illinois Press.
- Pareto, V. (1896). *Essai sur la répartition de la richesse*. Université de Lausanne.

Picavet, E. (1996). *Choix rationnel et vie publique*. Paris: PUF.

Saaty, T. (1980). *The Analytic Hierarchical Process: Planning, Priority, Resource Allocation*. New York: McGraw Hill.

Steuer, R. E. (1986). *Multicriteria Optimization: Theory, Computation, and Application*. Wiley.

Stirling, W. C. (2003). *Satisficing Games and Decision Making: With Applications to Engineering and Computer Science*. Cambridge University Press. doi:10.1017/CBO9780511543456

Tchangani, A.P. (2010). Considering Bipolarity of Attributes with regards to Objectives in Decisions Evaluation. *Inzinerine Ekonomika - Engineering Economics*, 21(5), 475-484.

Tchangani, A. P. (2013). Bipolar Aggregation Method for Fuzzy Nominal Classification using Weighted Cardinal Fuzzy Measure (WCFM). *Journal of Uncertain Systems*, 7(2), 138–151.

Tchangani, A. P. (2014). Bipolarity in Decision Analysis: a way to cope with human judgement. In *Exploring Innovative and Successful Applications of Softcomputing*. IGI Global. doi:10.4018/978-1-4666-4785-5.ch012

Tchangani, A. P. (2015). BOCR Analysis: a Framework for Forming Portfolio of Developing Projects. In M. Al-Shammari & H. Masri (Eds.), *Multiple Criteria Decision Making in Finance, Insurance, and Investment* (pp. 189–204). Springer. doi:10.1007/978-3-319-21158-9_9

ADDITIONAL READING

Brams, S. J., & Fishburn, P. C. (2002). Voting Procedures. In K. J. Arrow, A. K. Sen, & K. Suzumura (Eds.), *Handbook of Social Choice and Welfare* (Vol. 1, pp. 173–236). Elsevier Science. doi:10.1016/S1574-0110(02)80008-X

Hogarth, R. M., & Kunreuther, H. (1992). Decision-Making under Uncertainty: The Effects of Role and Ambiguity. In F. Heller (Ed.), *Decision-Making and Leadership*. Cambridge U. Press.

Kahneman, D., & Tversky, A. (2000). *Choices, Values, and Frames*. Cambridge University Press.

Saaty, T. (2005). *Theory and Applications of the Analytic Network Process: Decision Making with Benefits, Opportunities, Costs, and Risks*. RWS Publications.

Sen, A. K. (1984). *Collective Choice and Social Welfare*. North Holland.

Stirling, W. C. (2012). *Theory of Conditional Games*. Cambridge University Press.

Tchangani, A. P., Bouzarour-Amokrane, Y., & Pérès, F. (2012). Evaluation Model in Decision Analysis: Bipolar Approach. *INFORMATICA: an International Journal*, 23(3), 461–485.

Tchangani, A. P., & Pérès, F. (2010). BOCR Framework for Decision Analysis. In *Proceedings of 12th IFAC Symposium on Large Scale Systems: Theory and Applications, LSS'2010*, 11-14 July, Villeneuve d'Ascq, France.

Vincke, P. (1992). *Multicriteria decision-aid*. Chichester, UK: Wiley.

KEY TERMS AND DEFINITIONS

Aggregation: A process that permit to go from data represented in a high dimension set to a representation in a set of low and compact dimension.

Bipolarity: A notion that consists, for an actor, in viewing or evaluating anything in two directions: a direction positively seen by the actor and a direction considered to impede his aspirations.

BOCR Analysis: A framework of decision analysis where potential actions, options, deci-

sions, or alternatives are evaluated through four indicators: benefit (B), their certain or immediate positive contribution to decision goal; their opportunity (O), the positive uncertain or long term contribution; cost (C) that summarizes their immediate or certain aspects that work against the achievement of decision goal; and risk (R), an indicator that aggregates uncertain or long term potential threats.

Collective Choice: A decision making problem where a certain numbers of actors, decision makers, stakeholders or players must choose a subset (possibly reduced to a singleton) of alternatives or actions among a large number of potential actions or alternatives in order to achieve some collective as well as individual objectives.

Coordination Mechanism: Process by which the views points of many decision makers are aggregated to address collective choice problems.

Majority Voting: A coordination mechanism in collective choice where each decision maker must pronounce himself over a single alternative and the alternative that obtains the maximum voices win.

Rejecting: A notion to express the fact that behavior of something (attribute for instance) is negatively correlated with that of another thing (objective for instance).

Satisficing Game: A decision analysis framework where each action or alternatives is evaluated over two measures or degrees: a *selectability* measure that works towards the achievement of decision maker goal and the *rejectability* measure that constitutes the cost, in a broad sense, to pay to achieve the goal.

Supporting: A notion to express the fact that behavior of something (attribute for instance) is positively correlated with that of another thing (objective for instance).

S

Censorship in the Digital Age the World Over

Kari D. Weaver

University of South Carolina – Aiken, USA

INTRODUCTION

Censorship is an issue that has impacted information access for as long as humans have shared information. As communication technologies have grown and changed, from oral traditions, to the printing press, to the rapid rise of the Internet, many have sought to limit the ability of others to create, find, use, and share information with their fellow citizens. This has had important implications for the power and freedom of individuals within these societies, and has helped to shape the face of the modern world.

Censorship is also of critical importance to managers and entrepreneurs worldwide. Although international business has become increasingly “flattened” by globalization, information policy varies widely from nation to nation and has a significant impact upon both business climate and personnel safety. Employees in some countries may not be able to receive or share certain information, and government censorship may require additional functionality to be included in products or services. Cultural norms, practices, and ethical guidelines for capturing and sharing information are governed by the laws of the host nation. All of these factors must be taken into account when considering expanding operation into new locations around the globe.

BACKGROUND

Although censorship is a familiar concept, a precise definition is challenging. In this article, “censorship” is defined as “the action or the use

of authority to limit access to information that would otherwise be available in the public sphere.” Common justifications for censorship include morality, obscenity, blasphemy, or national security. It can be and is performed by individuals, groups, corporations, or governments and their agents.

Censorship has existed throughout civilized history. The word “censor” is derived from Latin, when censors within the Roman Empire conducted the census and other state functions while regulating morality under Roman rule (Pina Polo, 2012). Censorship has been practiced worldwide, primarily through religious institutions or by governments acting as agents of the church. Book banning and burning has played an important role in European history, such as in France during the Enlightenment, in the USSR during the Bolshevik revolution, and in Germany during the Weimar Republic (Lyons, 2011). There were few laws regulating censorship of information until 1789, when the Constitution of the United States of America was adopted, and individual freedoms, including the right to free speech, became protected. Since then, similar laws have been adopted by other countries around the world (Passavant, 2002).

Today, information has become a digital commodity, and individuals are now affected both by the laws of their home countries and by international standards. This is further complicated as individual jurisdictions seek to apply laws from the offline world to the online context (Worstell, 2012). However, no overarching international laws directly address censorship, and it was only in 2011 that the United Nations Human Rights Council identified unfettered Internet access as a human right (La Rue, 2011). International law

regarding information theft lags behind even the UN report, as prosecutions are based on the laws of the countries in which the crimes were perpetrated. Depending upon the country, these laws can be extremely restrictive, very broad, or potentially even in violation of internationally-recognized human rights agreements (La Rue, 2011). This inconsistent treatment of information across countries, in concert with increasing technological complexity, has led to challenges with the implementation of new internet standards (DeNardis, 2009).

Educational institutions have historically had a special role within society, with libraries and universities acting as both repositories of knowledge and points of information access for their communities. Libraries are common targets for censorship since removing a book from a personal collection blocks access for one person, whereas removing it from the library blocks access for the whole community. As Byrne (2003) notes, librarians cultivate a “professional narrative of non-judgmental, disinterested provision of access to information” which “confers a legitimacy on their professional choices to make [information] available or not make available” (Byrne, 2003, p. 7). Libraries have historically taken all sides on censorship debates - sometimes removing materials, keeping them in place, or relocating them to special sections. Broadly, though, libraries tend to focus on preserving, protecting, and providing access to information, as can be seen in their central role in preserving culture in locations as varied as Afghanistan and Colombia (Knuth, 2003, 2006).

CENSORSHIP IN THE MODERN WORLD

Throughout the modern world, agents of censorship fit broadly into three categories: individuals, governments, and corporations. In countries with laws protecting freedoms of speech, religion, and the press, acts of censorship are typically initiated by individuals or corporations, while, in others,

authoritarian governments are the primary agents of censorship. Although China is most commonly associated with censorship, there is widespread agreement that governments in Eritrea, North Korea, Syria, and Iran are even more restrictive (Reporters Without Borders, 2013). Other countries, such as Ethiopia, Cuba, Saudi Arabia, and Belarus, are also often listed as “most censored” countries, but rankings depend on the perspective of the organization reviewing censorship activities in each nation.

Regardless of the censoring agent, there are three main strategies for censoring materials. The first is removal or blocking of offensive material, which prevents individuals from viewing or experiencing the material at all. As the most common form of censorship, this strategy is typically employed where information is contained within physical artifacts, such as books or recordings. Second, instead of removing the material directly, organizations may instead remove any references to the material, such as a catalog entry or its visibility to a search engine (Zittrain & Palfrey, 2008; Chen & Wang, 2010). This strategy may extend to keyword filtering at an internet service provider (ISP) or router level, depending on the level of control of the government organization (Xu, Mao, & Halderman, 2011). With the growth of the Internet and users’ dependence on search engines for finding information, this is often as effective as removing the information entirely. Another variant of this method is blocking certain search terms that may lead to the offensive material, such as is used with the Chinese microblogging platform Weibo (Ng, 2013). The final method is intimidation. Authoritarian regimes make it known through a variety of means that exploration of certain materials or ideas is forbidden and will carry consequences for the user. While the ultimate decision remains with the user, the potential for punishment may prevent the search from ever taking place.

Due to its rapid growth and power for sharing information, the Internet has proven an important battleground for censorship activity. While most

Western democracies impose only very limited restrictions on Internet content (such as on child pornography), some governments limit public access to certain websites or databases (Morozov, 2011). This is accomplished in multiple ways. First, nationally-owned network hardware at the main network connections for the country may be configured to filter out certain websites or information from specific IP address ranges (Joyce, 2010; Yang, 2009). Second, private companies that operate their own servers or networking equipment within the nation may be required to themselves filter or block certain information, websites, or address ranges (Piety, 2012). Finally the nation may actively employ censors who review Internet content, removing or blocking content that is deemed unacceptable (Joyce, 2010; Yang, 2009). Beyond governments, individuals in countries in the European Union also may self-censor personal information, asking search engines to remove content under the recent Google v. Spain “Right to be Forgotten” ruling in the European courts (Arthur, 2014).

In the corporate sphere, another important consideration is “net neutrality,” or the willingness to pass information to other systems in the global network independent of content, point of origin/destination, or interfacing equipment. While discussions in the United States have primarily been framed in commercial terms, requiring all ISP’s to provide all information equally to their subscribers, there are also significant implications to the free exchange of information worldwide (MacKinnon, 2012). Currently the United States Federal Communications Commission (FCC) is drafting a ruling to protect net neutrality and limit corporate censorship, becoming only the second country in the world to do so despite legal challenges from corporate interests (Selyukh, 2015). The FCC originally planned to recommend against adopting net neutrality, but reversed their decision after receiving significant public comment from U.S. citizens, largely following a viral video explaining net neutrality by comedian John Oliver (Brody, 2015). However,

the limited adoption of net neutrality legislation worldwide likely means this form of censorship is still possible, as the Internet’s infrastructure does not specify the path information must take, and it may pass through the infrastructure of a nation that does not support net neutrality (Subramanian & Katz, 2011).

CURRENT PERSPECTIVES ON CENSORSHIP

As with the difficulty in defining censorship, there are multiple interests and perspectives regarding censorship, both for and against it. Here is a review of the major perspectives, their motivations, and their impact.

For those opposed to censorship, a common perspective advocates the right of people to intellectual freedom and to access publicly-available information in an unrestricted manner. As stated by the American Library Association (ALA):

Intellectual freedom is the right of every individual to both seek and receive information from all points of view without restriction. It provides for free access to all expressions of ideas through which any and all sides of a question, cause or movement may be explored. (American Library Association, Office for Intellectual Freedom, 2013a)

ALA further explains their reasoning:

Intellectual freedom is the basis for our democratic system. We expect our people to be self-governors. But to do so responsibly, our citizenry must be well-informed. Libraries provide the ideas and information, in a variety of formats, to allow people to inform themselves (American Library Association, Office for Intellectual Freedom, 2013a).

In contrast to this perspective, many seek to limit the availability of information based on ideological perspective. For instance, according to Hilgers (1913):

In general, censorship of books is a supervision of the press in order to prevent any abuse of it. In this sense, every lawful authority, whose duty it is to protect its subjects from the ravages of a pernicious press, has the right of exercising censorship of books (p. 519).

In this light, censorship is viewed as a positive activity, counteracting a potential corruption of a message. According to Baker (1989), “suppression of dissident political views is likely to be the most common and most vigorously pursued form of censorship” (p. 34). This perspective also commonly arises in a religious context. Based on data collected by the American Library Association Office of Intellectual Freedom (2013b) since 1990, the fifth, sixth, and seventh most common reasons for book challenges in the United States are due to “the occult or Satanism,” “homosexuality,” and “religious viewpoint,” respectively. This has also been seen in circumstances where religion is prominent in the public sphere. In Afghanistan, the Taliban have repeatedly destroyed books or other artifacts that did not fit their religious beliefs (Knuth, 2006).

In other cases, censorship is sought because some believe the material does not meet a common standard of decency (Bean, Dunkerly-Bean, & Harper, 2014). A “community standard” is frequently invoked with topics involving sexuality, profanity, and violence. While generally limited, this approach has even extended to consumer agreements with mobile technology providers in countries from Peru to Kenya (MacKinnon, 2012). This standard is also commonly invoked in the context of child welfare, since publicly available information may also be accessible by children, for whom, some believe, this information may be confusing or corruptive (Robinson, 2013). A central challenge with the “community standard”, however, arises from its vague definition. While religious censorship is delineated by the teachings of the religion, community standards can be fluid and amorphous. This challenge of definition was famously framed

by Justice Potter Stewart in his 1964 opinion on pornography as “I know it when I see it” (Jacobellis, 1964).

National security is also frequently invoked as a justification for censorship. From this perspective, it is appropriate to limit the availability of certain information if this information may weaken national security protocols or provide adversaries with damaging information (Bohlen, 2013). The central difficulty of this perspective arises from the assessment of which information could pose a risk. Assessments can vary widely and are usually determined solely by the authorities who already possess the information. While restrictions due to national security concerns may be legitimate, authorities have also historically censored information because it was embarrassing or personally damaging.

Perspectives on censorship also differ based on culture and government. Numerous organizations assess freedom of speech around the world, including Freedom House and Reporters Without Borders, scoring countries based on the extent to which freedom of the press and free speech is enjoyed there. In discussion of the 2013 World Press Freedom Index, Reporters Without Borders secretary-general Christophe Deloire noted:

The Press Freedom Index published by Reporters Without Borders does not take direct account of the kind of political system but it is clear that democracies provide better protection for the freedom to produce and circulate accurate news and information than countries where human rights are flouted. (Reporters Without Borders, 2013)

Authoritarian regimes are routinely assessed as the most pervasive censors of information in the world, with China and Iran the most studied (Reporters Without Borders, 2013). Citizens often assume their own countries’ standards for censorship are common throughout the world, which can lead to misunderstanding, confusion, or anger if inflammatory material is shared in some countries but not others.

ACTIVISTS AND ORGANIZATIONS WORKING AGAINST CENSORSHIP

Numerous organizations around the world work to fight censorship, including Amnesty International, Campaign Against Censorship, Freedom House, Index on Censorship, and Reporters Without Borders. In the United States, key organizations fighting censorship include the American Library Association, Electronic Freedom Foundation, and the American Civil Liberties Union. The activities of these organizations vary widely, from advocacy and reporting to legal support and protest actions.

WikiLeaks, founded by Julian Assange, is one of the most influential organizations fighting censorship worldwide (Krotoski, 2011). The WikiLeaks Internet platform is designed to give whistleblowers a safe and anonymous platform through which to share censored information. Although there is some disagreement on the ethics of the WikiLeaks sharing process, the volunteer organization has continued to be influential over time.

Numerous artists are also prominent critics of censorship. Ai Weiwei has openly criticized the Chinese government for corruption and cover-ups related to human rights abuses and censorship of the press. He is involved in a campaign to release the names of all children killed in the May 12, 2008 Sichuan earthquake, despite the Chinese government's efforts to thwart volunteer investigators (Pickowicz, 2009). In Russia, Nadezhda Tolokonnikova and Maria Alekhina, members of the band Pussy Riot, were sent to prison (though subsequently released) for playing a song critical of the Russian state (Walsh & Watkins, 2014). Amnesty International, among other organizations, was heavily involved in a successful campaign to free both musicians (Amnesty International, 2013).

SOLUTIONS AND RECOMMENDATIONS

Censorship is a complex societal issue that will likely never be completely resolved. While cen-

sorship has a largely negative impact on society, legitimate concerns exist that necessitate some forms of censorship. Over time, the challenge for the global community will be to determine the types and extent of censorship that can be tolerated, along with those that may be moderated by the nature, context, and impact of the censorship itself.

As technology progresses, censorship by individuals will become increasingly difficult. Censorship by individuals has historically occurred through appeals to authorities over information rather than direct intimidation. As technology has advanced, many modes of information dissemination have developed, with information often shared through multiple modes simultaneously. The number of authorities controlling information, such as libraries, publishers, retailers, and webmasters, has expanded, while the scope of their control is shrinking. This means that individuals now must appeal to multiple authorities to prevent all access to certain information, and the disagreement of one point of authority subverts the entire effort. Thus, decentralization provides a natural limit to the ability of individuals to censor information. In addition, the growing availability of education for all people will increase resistance to censorship and reduce the motivation to censor information available to others.

Corporate censorship may be legitimately necessary for reasons such as limiting employee distractions, managing company communications, and protecting intellectual property. However, there is also a strong profit motive to limiting the sharing of information that may negatively impact the company, regardless of the ethical implications. For instance, the recent Sony Corporation hacking scandal uncovered significant gender-based pay discrepancies in the company, which ultimately cost the company millions of dollars (Shanahan & Goldstein, 2014). Other companies censor financial disclosures that may materially impact their business, quashing research indicating harm caused by their product, or not disclosing the theft of sensitive customer information. Although corporate censorship is primarily limited to information produced by the company itself, companies have more avenues to censor

information than individuals, including filtering content, limiting search results, and intimidating potential whistleblowers. Most societies have laws that protect some rights of workers and whistleblowers who expose unethical business practices, and sites such as WikiLeaks also support whistleblower protection. In addition, corporate integrity has become an increasingly important issue to consumers in recent years. Damage to corporate reputation arises not just from the leak itself, but also through company actions since the information was discovered, the steps being taken to resolve the issue, and through the company's actions toward whistleblowers. These trends toward openness and accountability should be encouraged by society and codified into law where appropriate. The absence of worldwide net neutrality is also a form of corporate censorship, as telecommunications companies stand to reap enormous benefits if they can control the information consumers access.

Censorship by governments is potentially the most far-reaching, restrictive, and dangerous. It is widely understood that some government censorship is necessary to protect security interests. However, censorship outside of national security interests can hamper research and development, damage international credibility, conceal human rights' abuses, and undermine the citizens' right to know about reasonable government activities. In addition, censorship has the potential to subvert the functioning of government, as those within the censored nation cannot assess programs and policy beyond the point at which the law is implemented. Because governments generate vast quantities of information, there is a tendency to broadly over-censor rather than invest the resources to precisely determine which information can and cannot be shared. Governments are also in the most powerful position to censor information through blocking, filters, and intimidation, and they typically possess the means and organization to engage in all such practices. Where possible, citizens should engage their lawmakers to ensure that government censorship is restricted as much as possible, and that

appropriate oversight is provided to those areas where censorship is necessary. Internationally, organizations such as the UN and its member states should continue taking steps to encourage other nations to increase open access to information and expand the freedom of their people.

FUTURE RESEARCH DIRECTIONS

In the future, research in censorship will be strongly guided by the growth of new communication technologies. For instance, recent research regarding Weibo and censorship or social media use in the Arab Spring uprisings suggest that a greater insight into the role of censorship on social media platforms is under way (Ng, 2013; Robertson, 2013). The growing use of information communication technologies by people worldwide will further enhance this research trend.

While much of the research will be centered on censorship of expression in the information age, larger international discussions will stimulate research regarding information access as a basic human right. Research investigating the agents of censorship and the necessary control of information offer unique possibilities for scholars to simultaneously look at both the history and future of censorship. Much of this discussion will be shaped by the American perspective on the freedom of speech, freedom of the press, and limitations on government information control as similar laws are frequently adopted worldwide.

Finally, historical censorship has centered heavily on the removal of cultural materials of developed societies during times of conflict. Research may be conducted that involves interviewing individuals leaving such areas and documenting knowledge of destroyed materials to prevent further loss and maintain a more consistent cultural record. For instance, the current conflict in Syria has led to the destruction of at least one UNESCO World Heritage site by the Islamic State (Stack, 2015).

CONCLUSION

Censorship has existed for as long as humans have shared information, and it is expected to continue for the foreseeable future. Development of information technologies, such as the printing press, telephone, and Internet, has added new capabilities for both freedom and censorship in society. Both the diversity of perspectives on censorship and further technological developments will surely make this topic even more complex in the future.

The importance of information freedom to the overall freedom of society is clear. Going forward, it is crucial that citizens of all nations continue to debate and address complex problems such as Internet access, government control of information, and international law as societies determine the acceptable limits and uses of censorship within, outside, and across national boundaries.

REFERENCES

- American Library Association, Office for Intellectual Freedom. (2013a). *Intellectual freedom and censorship Q&A*. Retrieved from <http://www.ala.org/Template.cfm?Section=basics&Template=/ContentManagement/ContentDisplay.cfm&ContentID=60610>
- American Library Association, Office for Intellectual Freedom. (2013b). *Challenges by reason, initiator & institution for 1990-99 and 2000-09*. American Library Association. Retrieved from <http://www.ala.org/bbooks/frequentlychallenged-books/statistics>
- Amnesty International. (2013). *Pussy Riot Write-a-thon*. Retrieved from http://takeaction.amnesty-usa.org/c.6oJCLQPAJiJUG/b.8465995/k.6D0D/Pussy_Riot__Writeathon/siteapps/advocacy/ActionItem.aspx
- Arthur, C. (2014, May 14). Explaining the ‘right to be forgotten’ – The newest cultural shibboleth. *The Guardian Technology Blog*. Retrieved from <http://www.theguardian.com/technology/2014/may/14/explainer-right-to-be-forgotten-the-newest-cultural-shibboleth>
- Baker, C. E. (1989). *Human liberty and freedom of speech*. Oxford, UK: Oxford University Press.
- Bean, T. W., Dunkerly-Bean, J., & Harper, H. J. (2014). *Teaching young adult literature: Developing students as world citizens*. Los Angeles, CA: Sage. doi:10.4135/9781506374611
- Bohlen, C. (2013, August 31). Saying ‘no’ to our right to know. *New York Times*.
- Brody, B. (2015, Feb. 26). How John Oliver transformed the net neutrality debate once and for all: The story behind the clip that launched 45,000 comments on the FCC’s web site. *BloombergPolitics*. Retrieved from <http://www.bloomberg.com/politics/articles/2015-02-26/how-john-oliver-transformed-the-net-neutrality-debate-once-and-for-all>
- Byrne, A. (2003). The end of history: Censorship and libraries. In *Proceedings of the beacon for freedom of expression conference*. Alexandria, Egypt: Bibliotheca Alexandrina.
- Chen, T. M., & Wang, V. (2010). Web filtering and censoring. *Computer*, 43(3), 94–97. doi:10.1109/MC.2010.84
- Hilgers, J. (1913). Censorship. In C. G. Herbermann, E. A. Pace, C. B. Pallen, J. J. Wynne, & T. J. Shahan (Eds.), *The Catholic encyclopedia: An international work of reference on the constitution, doctrine, discipline, and history of the Catholic Church*. Academic Press.
- Jacobellis v. Ohio. 378 U.S. 184. (1964). Retrieved from http://www.oyez.org/cases/1960-1969/1962/1962_11_2

- Joyce, M. (2010). *Digital activism decoded: The new mechanics of change*. New York, NY: International Debate Education Association.
- Knuth, R. (2003). *Libricide: The regime-sponsored destruction of books and libraries in the twentieth century*. Westport, CT: Praeger.
- Knuth, R. (2006). *Burning books and leveling libraries: Extremist violence and cultural destruction*. Westport, CT: Praeger.
- Krotoski, A. (2011). Wikileaks and the new, transparent world order. *The Political Quarterly*, 82(4), 526–530. doi:10.1111/j.1467-923X.2011.02250.x
- La Rue, F. (2011). *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression* (UN Report Number A/HRC/17/27). Geneva: United Nations.
- Lyons, M. (2011). *Books: A living history*. Los Angeles, CA: The J. Paul Getty Museum.
- MacKinnon, R. (2012). *Consent of the networked: The worldwide struggle for Internet freedom*. New York, NY: Basic.
- Morozov, E. (2011). *The net delusion: The dark side of Internet freedom*. New York, NY: Public Affairs.
- Ng, J. Q. (2013). *Blocked on Weibo: What gets suppressed on China's version of Twitter (and why)*. New York, NY: The New Press.
- Passavant, P. A. (2002). *No escape: Freedom of speech and the paradox of rights*. New York, NY: New York University Press.
- Pickowicz, P. G. (2009). Chinas soft power: The case for a critical and multidimensional approach. *China Review International*, 16(4), 439–455. doi:10.1353/cri.2009.0100
- Piety, T. R. (2012). *Brandishing the first amendment: Commercial expression in America*. Ann Arbor, MI: The University of Michigan Press. doi:10.3998/mpub.291038
- Pina Polo, F. (2012). *The consul at Rome: The civil functions of the consuls of the Roman Republic*. Cambridge, UK: Cambridge University Press.
- Reporters Without Borders. (2013). *2013 world press freedom index: Dashed hopes after spring* [Press release]. Retrieved from <http://en.rsf.org/press-freedom-index-2013,1054.html>
- Robertson, A. (2013). Connecting in crisis: Old and new media and the Arab spring. *The International Journal of Press/Politics*, 18(3), 325–341. doi:10.1177/1940161213484971
- Robinson, K. H. (2013). *Innocence, knowledge and the construction of childhood: The contradictory nature of sexuality and censorship in children's contemporary lives*. London, UK: Routledge.
- Selyukh, A. (2015, June 11). *Court declines to suspend U.S. net neutrality rules*. Retrieved from <http://www.reuters.com/article/2015/06/11/us-usa-internet-neutrality-idUSKBN0OR2QJ20150611>
- Shanahan, M., & Goldstein, M. (2014, Dec. 15). Hacking reveals gender pay gap at Sony. *The Boston Globe*. Retrieved from <https://www.bostonglobe.com/lifestyle/names/2014/12/15/name-shuttle/0Pca1iPikwJEgobmZgWofM/story.html>
- Stack, L. (2015, Aug. 24). Islamic State blows up temple at Palmyra ruins in Syria. *The New York Times*, p. A6.
- Subramanian, R., & Katz, E. (2011). *The global flow of information: Legal, social, and cultural perspectives*. New York, NY: New York University Press. doi:10.18574/nyu/9780814748114.001.0001
- Thierer, A., & Crews, C. W. (2003). *Who rules the net? Internet governance and jurisdiction*. Washington, D. C.: The Cato Institute.
- Walsh, N. P., & Watkins, T. (2014, Feb. 19). *Members of Pussy Riot released in Sochi*. Retrieved from <http://www.cnn.com/2014/02/18/world/europe/russia-pussy-riot-arrests/>

Worstell, T. (2012, Jan. 27). *Twitter's new censorship system and international internet law*. Retrieved from <http://www.forbes.com/sites/timworstell/2012/01/27/twitters-new-censorship-system-and-international-internet-law/>

Xu, X., Mao, Z. M., & Halderman, J. A. (2011). Internet censorship in China: Where does filtering occur? *Passive and Active Measurement*. doi:10.1007/978-3-642-19260-9_14

Yang, G. (2009). *The power of the Internet in China: Citizen activism online*. New York, NY: Columbia University Press.

Zittrain, J., & Palfrey, J. (2008). Corporate ethics on a filtered Internet. In R. Diebert, J. Palfrey, R. Rohozinski, & J. Zittrain (Eds.), *Access denied: The practice and policy of global Internet filtering* (pp. 69–78). Cambridge, MA: The MIT Press.

ADDITIONAL READING

Andrews, L. (2012). *I know who you are and I saw what you did: Social networks and the death of privacy*. New York, NY: Free Press.

Avery, K., Carvell, T., Gurewitch, D., Haggerty, G., Maurer, J., Oliver, J.,.... (Writers), & Perota, J. (Director). (2014) Episode #1.5 [Television series episode]. In T. Carvill & J. Oliver (Producers). Last Week Tonight with John Oliver [Television Broadcast]. New York, NY: Home Box Office.

Barbour, S. (2010). *Censorship*. Detroit, MI: Greenhaven.

Bollier, D. (2008). *Viral spiral: How the commoners built a digital republic of their own*. New York, NY: The New Press.

Censorship dateline. (2015). Newsletter on Intellectual Freedom, 64(4), 93-122.

Coetzee, J. M. (1997). *Giving offense: Essays on censorship*. Chicago, IL: University of Chicago Press.

Cohen, N. (2012). *You can't read this book: Censorship in an age of freedom*. London, UK: Fourth Estate.

Cohen-Almagor, R. (2005). *Speech, media, and ethics: The limits of free expression*. New York, NY: Palgrave.

Dainotti, A., Squarcella, C., Aben, E., Claffy, K. C., Chiesa, M., Russo, M., & Pescapé, A. (2011). Analysis of country-wide Internet outages caused by censorship. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement* (pp. 1-18). New York, NY: Association for Computing Machinery. doi:10.1145/2068816.2068818

DeNardis, L. (2009). *Protocol politics: The globalization of Internet governance*. Cambridge, MA: The MIT Press; doi:10.7551/mitpress/9780262042574.001.0001

Freedman, L. (2009). *The offensive art: political satire and its censorship around the world from Beerbohm to Borat*. Westport, CT: Praeger.

Howard, P. N. (2011). *The digital origins of dictatorship and democracy: Information technology and political Islam*. Oxford, UK: Oxford University Press.

Howard, P. N., & Jones, S. (2004). *Society online: The Internet in context*. Thousand Oaks, CA: SAGE.

Hutchenson, R. (2013). *Negotiating censorship in modern Japan*. New York, NY: Routledge.

Karolides, N. J., Bald, M., Sova, D. B., & Karolides, N. J. (2005). *120 banned books: Censorship histories of world literature*. New York, NY: Checkmark Books/Facts On File.

Laachir, K., & Talajooy, S. (2013). *Resistance in contemporary Middle Eastern cultures*. New York, NY: Routledge.

Meinrath, S. D., Losey, J., & Lennett, B. (2011). A growing digital divide: Internet freedom and the negative impact of command-and-control networking. *IEEE Internet Computing*, 15(4), 75–79. doi:10.1109/MIC.2011.85

Müller, B. (2004). *Censorship & cultural regulation in the modern age*. Amsterdam, NL: Rodopi.

Nunziato, D. C. (2009). *Virtual freedom: Net neutrality and free speech in the Internet age*. Stanford, CA: Stanford Law Books.

Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. New York, NY: Penguin.

Ryan, J. (2010). *A history of the Internet and the digital future*. London, UK: Reaktion.

Shmuger, M. (Producer), Gibney, A. (Producer & Director), & Bloom, A. (Producer). (2013). *We steal secrets: The story of Wikileaks*. [motion picture]. United States: Universal Studios.

Simon, J. (2014). *The new censorship: Inside the global battle for media freedom*. New York, NY: Columbia University Press.

Sunstein, C. R. (2007). *Republic.com 2.0*. Princeton, NJ: Princeton University Press.

Wu, T. (2011). *The master switch: The rise and fall of information empires*. New York, NY: Vintage.

KEY TERMS AND DEFINITIONS

Censorship: The action or the use of authority to limit access to information that would otherwise be available in the public sphere. Censorship is often undertaken for concerns over morality, obscenity, blasphemy, or national security. It can be performed by individuals, groups, corporations, or governments and their agents.

Information Communication Technologies (ICT): Hardware, software, or network technologies that allow a user to create, edit, find, store, or share information. Examples include mobile phones, computers, intranets, social media platforms, telecommunications services, and the Internet.

IP Address: A set of numbers which specify a unique network address for a computer on a network.

Microblog: A blogging service focused on short entries, usually limited by a number of words or characters.

Net Neutrality: A concept of Internet governance stating that all network traffic should be handled without regard to content, source/destination, or interfacing equipment.

Wikileaks: A website designed to allow whistleblowers to anonymously submit information for distribution to the public.

Information-Based Revolution in Military Affairs

Rafal Kopec

Pedagogical University of Krakow, Poland

INTRODUCTION

Revolution in Military Affairs (RMA) is a general term describing a fundamental change in the character and conduct of armed conflicts. The majority of RMAs rely on the development of new technologies, however, they also include different aspects. RMAs typically consist of three elements: technological change, innovative operational concept, and organizational adaptation (Krepinevich, 1994). The RMA term was introduced by Michael Roberts in the context of revolutionizing the methods of waging war based on changes in tactics and technology between 1560 and 1660 (Roberts, 1956).

The subject of this article is the last RMA. The crux of it is to apply information technology (IT) in warfare. Being precise, the last RMA should be called Information-based Revolution in Military Affairs (Davis, 1996). However, the RMA term has been commonly attributed to Information-based Revolution in Military Affairs since the 90s. In 1993, Andy Marshall, the Pentagon's Director of Net Assessment, replaced the previous term "military-technical revolution" with "Revolution in Military Affairs" to underline a comprehensive change in military. It not only represents a technological change, but also a change in doctrine and organization (Watts, 2011).

BACKGROUND

Looking for the roots of RMA, it is necessary to recall a number of technological breakthroughs, whose synthesis provides the foundation of the

revolution. RMA – similarly to the majority of military revolutions in the history – it is not based on one groundbreaking innovation, but it takes advantage of a number of technological changes. In this case, the miniaturization of computer hardware and creation of highly efficient, decentralized data communication networks are the crucial technologies.

Applying these technologies in military was for the United States and others NATO countries the way to overcome the Warsaw Pact's quantitative dominance in conventional forces. A number of new types of weaponry were developed in the 80s, and they are generally divided into two groups: reconnaissance systems and striking systems (U.S. Congress, Office of Technology Assessment, 1986). The JSTARS (Joint Surveillance Target Attack Radar System) airborne reconnaissance system was the example of the first group, whereas the MGM-140 ATACMS (Army Tactical Missile System) surface-to-surface missile system represented the second one. These two elements were based on new technologies and cooperated closely thanks to highly-efficient data communication networks. They built together the innovative reconnaissance-attack system, and created background for further transformation.

The roots of RMA go to a given strategic problem, that is, anticipated confrontation with outnumbered Soviet forces. It is typical for military revolutions, which occur at a certain time and place just because they offer solutions to actual, not hypothetical, problem (Cohen, 2002; Bjerregaard, 2012).

The assumptions of the new concept were successfully tested during the Desert Storm operation

in 1991. Hence, RMA indicated the direction of transforming the American army, and afterward other modern forces. In simple terms, the transformation was aimed at converting armed forces into one coherent reconnaissance and striking system. It should be based on all available sensors and effectors connected with decentralized data networks (US Department of Defense, 2005).

REVOLUTION IN MILITARY AFFAIRS: ASSUMPTION AND PRACTICE

The Key Elements of RMA

The core of RMA consists of the following interpenetrated elements:

1. Computerization and Networking

Battlespace awareness would replace traditional determinants of dominance like firepower, survivability and mobility. Constantly updated knowledge about location of friendly and foe forces should become a critical factor which determines the military success. The quest for the information superiority has been the leading motive of military transformation. Implementing the new types of weapon is not considered as the essence of the revolution. The integration of all elements of battle formation with information distribution network is regarded here as the novelty. The C4IRS (Command, Control, Communications, Computers, Intelligence, Surveillance, Reconnaissance) system has become a critical component of modern armed forces. It is designated to the real-time coordination of the stimulus-response (from sensors to effectors) process. This *modus operandi* should overcome the eternal problem of warriors, which is uncertainty and opacity inherent in any military campaign, and should “lift” the Clausewitz’s “fog of war” (Owens, 2001).

RMAs are often linked to a broader change (Vickers, Martinage, 2004). Consequently, the

transformation driven by computerization and networking harmonizes with a wide-ranging social change toward an information society. The “Internet metaphor” is often recalled in each sphere (Shapiro, 1999). The background of RMA consists of elements typical for Internet: lack of central control over a system (creative anarchy), easy access to information from almost every place all over the world, and interconnectivity (peer-to-peer architecture) (Shapiro, 1999).

2. Smart Weapon (Precision-Guided Munitions)

It allows hitting the target with accuracy which was not accomplishable earlier. This capability implies the decrease in material requirements. Destroying an enemy object needs considerably less munitions compared to “dumb” munitions, but it requires precise information provided by “battle networks” (Watts, 2007). This trend corresponds with contemporary business, where information is the substitute of material and transport resources (Cebrowski & Garstka, 1998). Information dominance makes it possible to manage resources optimally.

3. Stand-Off

Increasing combat distance means the use of weapons can be launched far from the target. It makes a launcher invulnerable to enemy counter-measures. This kind of weapons is represented, for example, by cruise missiles used among others to carry out “punished” operation – the so-called Tomahawk diplomacy, which is a modern incarnation of gunboat diplomacy (Pretsch, 1999). A particular part of this trend is a widespread tendency towards robotisation (Doaré, Danet, 2014). The use of robots is more and more common not only in aviation (drones) and navy (first of all for minesweeping), but also on land. However, land is more demanding environment and projects of fully autonomous armed vehicles are still under development.

All these things may lead to make warfare resemble a computer game. Direct combat will be the exception rather than the rule as it increases the security level of own forces. Thus a stand-off operational model provides a means to carry out a military campaign despite the casualty aversion, typical for post-heroic Western societies.

RMA's Influence on Military

Deep transformation of leading (first of all American) armed forces is the effect of implementing the elements mentioned above. The crux of the transformation could be summarized as the list of the following oppositions:

1. Quantity vs. Quality

Since technological advantage is much more important than numerical superiority, the tendency toward decreasing the amount of armed forces is very clear. Modern technology replaces the potential expressed in simple numbers, that is, in the number of soldiers, tanks, guns or planes. It means that the tendency observed since the French Revolutionary Wars (the beginning of total wars) has been reversed. The quantity was the evidence of power in the past – the quantity of soldiers at first, then also the quantity of the essential types of weapon. Military power was determined by population power and economical power (e.g. heavy industry). Technological innovation was not a decisive factor, because it could be relatively simple assimilated by the enemy. The quantity was of crucial significance, which caused mass production of weapon and “mass production of men” (Howard, 1976, p. 121).

RMA has changed the situation significantly. The Gulf War in 1991 was a breakthrough moment. It positively examined the thesis about the significance of technology in modern warfare. Not only did the technology become a profoundly influential element of military campaigns, but it started to play a major role and to overshadow further elements.

2. Mass vs. Information

Mass – both in terms of quantity and “massiveness” of firepower and armor – is no longer the condition of military advantage. It has been superseded by fire precision and battleship awareness resulting from informational advantage (Blaker, 2006). According to Paul Wolfowitz, as warfare reflects the characteristics of the times, “we move from the Industrial Age with its emphasis on mass to Information Age warfare, which highlights the power of networked distributed forces and shared situational awareness” (US Department of Defense, 2005, p. 7).

The baseline of the advantage is to create armed forces based on the concept of network-centric warfare. Heavy armored units were declared an anachronism on the grounds of the belief that information replaces armor (Joint Vision 2010, 1996). A light military unit could avoid hits due to possession of complete information about an enemy. Multidimensional awareness and the ability to identify all forces in the battle space could build full-dimensional protection (Joint Vision 2010, 1996). When it comes to a single vehicle it could be possible because of so-called active protection systems (detecting an incoming bullet and using softkill or hardkill countermeasures). The eternal conflict of sword versus shield (or bullet versus armor as its modern interpretation) should be a thing of the past. The third element – information – has been announced the winner of this confrontation. As a result, the possibility to decisively defeat superior forces (when it comes both to the quantity and capability in traditional measures) by less numerous and “less capable,” but digitized and networked platforms will be achieved (US Department of Defense, 2005).

The huge project of the US Army's transformation – Future Combat Systems – was the appearance of that concept. Formally launched in 2003, it led to a new type of a brigade equipped with a manned and unmanned vehicle. Communication network was supposed to be a core of a brigade providing the connectivity and situational aware-

ness of every element of the unit. Heavy armor seemed to be redundant. Extensive information about an enemy's activity was thought to make it possible to build lighter combat vehicles (Future Combat Systems program was supposed to create "medium-weight" combat unit with 19-ton platforms transportable on the C-130 aircraft) (RAND Corporation, 2012). As a result, strategic mobility and capability to deploy troops in the theatre of operation should dramatically increase (Shinseki, 1999).

3. Riflemen vs. Sustainment Soldiers

In the information-based armed forces tending to avoid close combat due to stand-off capabilities, traditional soldier skills have ceased to be crucial. They have been reduced to the background and replaced by technological skills. The ability to operate technologically advanced equipment has become essential even for front-line soldiers, not to mention specialists like unmanned vehicle operators. As a result, the percentage of soldiers in the front line has been diminishing. This in turn has increased the role of broadly defined sustainment (soldiers, whose mission does not require direct contact with enemy). This tendency is derived from the assumption that network-centric units will be able to dramatically reduce or even eliminate the necessity for getting entangled in risky close-combat operations (Bingham, 2001).

4. Hierarchical Organizational Structure vs. Flat Organizational Structure

A flat interrelated organizational structure corresponds more closely to network-centric warfare requirements. Knowledge and consequently autonomy in decision making is going down the hierarchical ladder. Access to information on lower levels is less controlled, which entails more latitude in decision taking. A horizontal structure should involve such functions as fire, maneuver, transportation, and logistics (Blaker, 2006). This is the same process as in the business sphere – a

rigid organizational structure has been replaced by a more flexible network structure with a wide range of self-direction.

RMA in Contemporary Armed Conflicts

RMA-based armed forces were successful in the 1991 Gulf War and in the 2003 invasion of Iraq, however, experience with RMA is ambiguous. Insurgency phases of Afghan and Iraqi conflicts have verified the usefulness of the technology-driven military transformation in the conditions of the so-called asymmetric warfare. These conflicts highlighted the gaps between the claims to full-spectrum dominance and reality (Der Derian, 2009). The expectations that high-tech machines dominate the ground just like in the air and on the sea, where technological advantage almost directly results in strategic superiority, were not fulfilled (Kitfield, 2009). The outcomes of practical application of transformation have brought about a radical shift in the thinking of RMA. It has come a long way from enthusiasm to skepticism symbolized by entering the opposite concept of counter-revolution in military affairs.

The flagship project of RMA – Future Combat Systems – was limited and finally canceled (Krepinevich, Montgomery, 2009). The operations in Iraq and Afghanistan have undermined the RMA framework. Why such a radical change in attitude? Does it mean that the principles of RMA are invalid? I answer these questions by creating a few, not mutually exclusive, possible explanations.

1. The assumptions of RMA were unrealistic. They created a vision of changing the nature of war (not only the character of war) (Deptula, 2001). It involved the rejection of the Clausewitz's vision of war with its inherent unpredictability, thanks to the crucial importance of technology. Superior Information Position should become the key, if not the only factor determining the possibility to bring operation to a successful conclusion

(Alberts, Garstka, 1999). In effect, there are two fallacies from this approach: firstly, the absolute information superiority is possible; secondly, information could play a prominent role in conflict, in some ways substituting brute force.

The symbol of this approach was the concept of Effect-Based Operations. This approach was dependent on mathematical methods for predicting and measuring effects (Vego, 2006). The vision of future conflict as a one-sided undertaking was extremely chancy. This perception of conflict led to replacing strategy with technology as a crucial factor (victory as a simple consequence of technological advantage).

The vision of how to carry out military action was quite unrealistic and excessively technology-saturated. Contemporary armed conflicts show that Clausewitz's fog and friction cannot be ruled out (Seah, 2006) and the nature of war is still the same, despite the changes in equipment, organization, doctrine and military culture (Gray, 2005). Finally, the US Army abandoned the Effect-Based Operation concept, concluding that it assumes a level of unachievable predictability, is overengineered and discounts the human dimension of war (for example passion, imagination, willpower, and unpredictability) (Mattis, 2008).

Fantasies of high technology-driven military dominance repeat every once in a while. The Donald Rumsfeld's Department of Defense was dominated by technocrats similarly to the Robert McNamara's DoD at the time of Vietnam War (Guha, 2011). They both managed technologically sophisticated forces which were unable to decisively defeat far less advanced foes.

2. RMA as a legacy of the Cold War is irrelevant in the environment of contemporary armed conflicts including first of all counterinsurgency operations (COIN). Technologies developed in order to gain superiority in classic, high-intensity, state-to-state conflicts have appeared to be an "overinnovation." Such

technologies as aircrafts' stealth capability or counter-battery capabilities of modern (self-propelled and equipped with integral reconnaissance systems) artillery seem to be superfluous. They are expressions of the conventional orientation of military that engages the enemy directly (Tucker, 2006) while unconventional conflicts require rather an indirect approach, which is focused more on political and social solutions than on military tools (Freedman, 2006).

Light armor and protection based on information superiority have turned out to be a mistake in anti-rebel operations. Iraqi and Afghani operations have re-established traditional, non-information-based solutions. Heavy armored vehicles, including tanks, and Mine-Resistant Ambush Protected (MRAP) vehicles are still very usable. Paradoxically, MRAP vehicles are based on a simple but effective design with V-shaped hulls developed during the Rhodesian Bush War in the 70s. What is more, the sophisticated technology has not become the most desired resource, but the number of soldiers. The prospect of quick, "surgical" operations has not come to fruition. The direct physical presence of troops ("boots on the ground") in a conflict area is still valid. According to McGrath, the starting point of counterinsurgency operation should be to estimate a size of the forces in terms of troops to inhabitants ratio (McGrath historical analysis suggests a figure of about 13,26) (McGrath, 2006, p. 103). RMA has led to a paradoxical situation – in the past the army able to defeat regular forces was naturally numerous enough to occupy the terrain, but RMA-based forces are too small due to the pressure to reduce units.

RMA can be also perceived as a revolution distorted by politics. The demands to build new, lighter platforms were raised in order to adjust RMA principles to post-Cold War determinants (requirement of quick military interventions on a global scale related to the capability of global power projection). Platforms' "slimming" was not a part of core principles of RMA, but it

was justified in the spirit of RMA (information replacing heavy armor). Technological superiority should assert the American dominance round the globe allowing US forces to intervene almost immediately anywhere (Dalby, 2008). Its effect was a challenging program of Future Combat Systems that was based on controversial assumptions (combat platforms transportable on C-130).

3. RMA assumptions have not been fully realized because of the lack of technology. RMA-based transformation can be divided into two stages. The first stage, computerization, based on the implementation of miniaturized and highly-effective computer units in advanced types of weapons, has led to a gain in strike precision and stand-off launch capability. This stage might be recognized as relatively advanced one (Watts, 2011). However, the second stage, networking, responsible for making the battle space “transparent” and reaching indisputable information superiority is still immature.

RMA could be defined as the synergy between three elements: information acquisition, information processing and transfer, and the use of information in order to enhance firepower (in simple words: between sensors, connectors, and effectors). The first and the third elements are the strait extension of the reconnaissance-attack system concept, but the second element is crucial when it comes to real capabilities of RMA-based forces and the potential to cause a true revolutionary change. Sensors and effectors have been developed over decades, whereas the connection sphere relies on innovative technologies. However, their development is not as fast as it has been expected (one can compare ambitious objectives and a relatively slow progress of the Land Warrior program, that is, an integrated fighting system for infantry soldiers).

FUTURE RESEARCH DIRECTIONS

RMA has highly influenced the way technology advanced armed forces conduct military operations. This article signalizes major trends and significant controversies. However, there is a need to intensify research on particular issues and to focus on the following remarkable future directions:

1. The conflict of post-heroic Western mentality (symbolized by technology-driven trends allowing to avoid face-to-face confrontation) with heroic non-Western mentality (symbolized by suicide bombers called ultimate precision weapon) (Peters, 2009).
2. Do we deal with the limits of technological development in some areas, which poses a significant inhibitor of RMA-based transformation?
3. The issue that can be called “Dreadnought dilemma,” that is, building the Dreadnought battleship in 1907 (a faster, better armored battleship with more firepower than previous ones) re-established arms race between Great Britain and Germany and, consequently, undermined previous British advantage (Poundstone, 1993). So, implementing technologically advanced weapon can be perceived as a way to maintain advantage of a current leader, however, it also carries a risk of undermining this advantage.
4. Who is a leader in implementing a flat organizational structure: technology advanced armed forces or, paradoxically, decentralized, networked terrorism (Bousquet, 2009)?

CONCLUSION

RMA is an important step forward in the development of methods and means of warfare. Battle network systems like blue force tracking, which provides a common operation picture, offer sig-

nificant advantage over the enemy (Arkin, 2006). RMA should be perceived in that way: as a useful tool of waging war. However, the technology factor has been considered as something more. It has been claimed to be a decisive factor determining the outcome of every armed conflict (Maj, 2010).

Properly applied RMA can be the source of great advantage, in state-to-state and in counterinsurgency conflict as well. In fact, RMA is the continuation of an eternal pursuit, that is, to acquire more knowledge about the enemy than it could acquire about us. And so, the fundamental principles of war remain immutable, and the idea of changing the very nature of war has turned out to be an illusion.

REFERENCES

- Alberts, D. S., Garstka, J. J. F., & Stein, P. (1999). *Network Centric Warfare: Developing and Leveraging Information Superiority*. DoD Command and Control Research Program. Retrieved November 5, 2013, from http://www.iwar.org.uk/rma/resources/ncw/ncw_2nd.pdf
- Arkin, W.A. (2006, February). Spiralling ahead. *Armed Forces Journal*.
- Bingham, P. T. (2001). Transforming Warfare with Effects-Based Joint Operations. *Aerospace Power Journal*, 58-66. Retrieved November 1, 2013, from <http://www.airpower.maxwell.af.mil/airchronicles/apj/apj01/spr01/ingham.pdf>
- Bjerregaard, T. (2012). *Hybrid Warfare: A Military Revolution or Revolution in Military Affairs?* Fort Leavenworth. U.S. Army Command and General Staff College.
- Blaker, J. (2006, Spring). Arthur A. Cebrowski: A Retrospective. *Naval War College Review*, 59(2), 129-145. Retrieved November 20, 2013, from <http://www.usnwc.edu/getattachment/d1c5384c-f7fa-41b7-a6e6-490450835809/Arthur-K--Cebrowski--A-Retrospective---Blaker,-Jam.aspx>
- Bousquet, A. (2009). *The Scientific Way of Warfare: Order and Chaos on the Battlefields of Modernity*. London: Hurst and Company.
- Cebrowski, A. K., & Garstka, J. H. (1998). Network-Centric Warfare – Its Origin and Future. *Proceedings Magazine*, 139. Retrieved November 13, 2013, from <http://www.usni.org/magazines/proceedings/1998-01/network-centric-warfare-its-origin-and-future>
- Cohen, E. (2002). Technology and Warfare. In J. Baylis, J. Wirtz, E. Cohen, & C. S. Gray (Eds.), *Strategy in the Contemporary World* (pp. 235–252). New York: Oxford University Press.
- Dalby, S. (2008). *Geopolitics, the Revolution in Military Affairs, and the Bush Doctrine*. YCISS Working Paper Number 49. Ottawa: Carleton University.
- Davis, N. (1996). An Information-Based Revolution in Military Affairs. *Strategic Review*, 24(1), 43–53.
- Deptula, D. A. (2001). *Effects-Based Operations: Change in the Nature of Warfare*. Arlington, VA: Aerospace Education Foundation.
- Der Derian, J. (2009). *Virtuous War: Mapping the Military-Industrial-Media-Entertainment Network*. London: Routledge.
- Doaré, R., Danet, D., Hanon, J.-P., & de Boisboissel, G. (Eds.). (2014). *Robots on the Battlefield. Contemporary Perspectives and Implications for the Future*. Combat Studies Institute Press.
- Freedman, L. (2006). *The Transformation in Strategic Affairs*. New York: Routledge.
- Gray, C. (2005). *Transformation and Strategic Surprise*. Carlisle, PA: Strategic Studies Institute.
- Guha, M. (2011). *Reimagining War in 21st Century: from Clausewitz to Network-Centric Warfare*. London: Routledge.
- Howard, M. (1976). *War in European History*. Oxford, UK: Oxford University Press.

- Joint Vision 2010. (1996). Washington, DC: Chairman of the Joint Chiefs of Staff. Retrieved November 13, 2013, from <http://www.dtic.mil/jv2010/jv2010.pdf>
- Kitfield, J. (2009). The Counter-Revolution in Military Affairs. *National Journal*, 14.
- Krepinevich, A. F. (1994). Cavalry To Computer; The Pattern Of Military Revolutions. *The National Interest*. Retrieved November 3, 2013, from <http://nationalinterest.org/article/cavalry-to-computer-the-pattern-of-military-revolutions-848?page=1>
- Krepinevich, A. F., & Montgomery, E. B. (2009, July). *Correcting Course: The Cancellation of the Future Combat Systems Program*. Washington, DC: Center for Strategic and Budgetary Assessments. Retrieved November 16, 2013, from <http://www.csbaonline.org/search/?q=correcting&x=0&y=0>
- Maj, T. E. H. F. (2010). Technological Competence and Military Professionalism—Decisive Factors in A Modern War? *Journal of the Singapore Armed Forces*, 36(2), 30–40.
- Mattis, J. N. (2008). USJFCOM Commander's Guidance for Effects-based Operations. *Parameters*, 38(3), 18–25.
- McGrath, J. J. (2006). *Boots on the Ground: Troop Density in Contingency Operations*. Global War on Terrorism. Occasional Paper 16. Fort Leavenworth, KS: Combat Studies Institute Press.
- Owens, B. (2001). *Lifting the Fog of War*. Baltimore, MD: Johns Hopkins University Press.
- Peters, R. (2009). *Wars of Blood and Faith: The Conflicts That Will Shape the Twenty-First Century*. Mechanicsburg, PA: Stackpole Books.
- Poundstone, W. (1993). *Prisoner's dilemma*. New York: Anchor Books.
- Pretsch, R. A. (1999). *Tomahawk Diplomacy and U.S. National Security*. United States Army Command and General Staff College.
- RAND Corporation. (2012). *Lesson from the Army Future Combat Systems Program*. Santa Monica, CA: RAND Corporation. Retrieved November 12, 2013, from http://www.rand.org/content/dam/rand/pubs/monographs/2012/RAND_MG1206.pdf
- Roberts, M. (1956). *The Military Revolution, 1560-1660: An Inaugural Lecture Delivered before the Queen's University of Belfast*. Belfast, Ireland: Boyd.
- Seah, E. (2006). *Temporal Dominance: Military Transformation and the Time Dimension of Strategy*. Singapore: Institute of Defence and Strategic Studies.
- Shapiro, J. (1999). Information and War: Is It a Revolution? In Z. Khalilzad, J. P. White, & A. W. Marshall (Eds.), *The Changing Role of Information in Warfare* (pp. 113–153). Santa Monica, CA: RAND Corporation.
- Shinseki, E. K. (1999, October 12). Address to the Eisenhower Luncheon. *45th Annual Meeting of the Association of the United States Army*.
- Tucker, D. (2006). *Confronting the Unconventional: Innovation and Transformation in Military Affairs*. Carlisle, PA: Strategic Studies Institute.
- U.S. Congress, Office of Technology Assessment. (1986, July). *Technologies for NATO's Follow-On Forces Attack Concept – Special Report*. A Special Report of OTA's Assessment on Improving NATO's Defense Response OTA-ISC-312. Washington, DC: U.S. Government Printing Office.
- US Department of Defense. (2005). *The Implementation of Network-Centric Warfare*. Washington, DC: Department of Defense, Office of Force Transformation. Retrieved November 2, 2013, from http://www.au.af.mil/au/awc/awcgate/transformation/oft_implementation_ncw.pdf
- Vego, M. N. (2006). *Effect-Based Operations: A Critique*. Washington, DC: National Defense University, Institute for National Strategic Studies.

Vickers, M. G., & Martinage, R. C. (2004). *The Revolution in War*. Washington, DC: Center for Strategic and Budgetary Assessments.

Watts, B. D. (2007). *Six Decades of Guided Munitions and Battle Networks: Progress and Prospects*. Washington, DC: Center for Strategic and Budgetary Assessments.

Watts, B. D. (2011). *The Maturing Revolution in Military Affairs*. Center for Strategic and Budgetary Assessments. Retrieved November 5, 2013, from <http://www.csbaonline.org/publications/2011/06/the-maturing-revolution-in-military-affairs/>

ADDITIONAL READING

Alach, Z. J. (2008). *Slowing Military Change*. Carlisle, PA: Strategic Studies Institute.

Alvis, M. W. (1999). *Understanding the role of Casualties in U.S. Peace Operations, Landpower Essay Series, No. 99-1, January 1999*. Arlington: Association of the United States Army, Institute of Land Warfare.

Alwin-Foster, N. (2005). Changing the Army for Counterinsurgency Operation. *Military Review*, 85(6), 2–15.

Biddle, S. (1996, Fall). Victory Misunderstood, What the Gulf War Tells Us about the Future of Conflict. *International Security*, 21(2), 139–179. doi:10.1162/isec.21.2.139

Biddle, S. (1998, Autumn). The Past As Prologue: Assessing Theories of Future Warfare. *Security Studies*, 8(1), 1–74. doi:10.1080/09636419808429365

Black, J. (2004). *Wars since 1990*. London: Reaction Books.

Coker, Ch. (2002). *Waging war without warriors? The changing culture of military conflict*. Boulder: Lynne Rienner.

Coker, Ch. (2012). The collision of modern and post-modern war. In J. Lindley-French & Y. Boyer (Eds.), *The Oxford Handbook of War. Oxford Handbooks in Politics & International Relations* (pp. 57–68). Oxford: Oxford University Press; doi:10.1093/oxfordhb/9780199562930.013.0005

Coker, Ch. (2013). *Warriorgeeks: how 21st century technology is changing the way we fight and think about war*. London: Hurst Publishing; doi:10.1093/acprof:oso/9780199327898.001.0001

Der Derian, J. (2003, Summer/Fall). War as Game. *The Brown Journal of World Affairs*, X(1), 37–48.

Donnelly, W. M. (2007). *Transforming an Army at War: Designing the Modular Force 1991-2005*. Washington, DC: Center of Military History.

Echevarria, A. J. II. (2005). *Fourth-Generation of Warfare and other Myths*. Carlisle: Strategic Studies Institute.

Freedman, F. (1998). *The Revolution in Strategic Affairs. Adelphi Papers 318*. London: International Institute for Strategic Studies.

Hundley, R. O. (1999). *Past Revolutions, Future Transformation. What can the history of revolution in military affairs tell us about transforming the U.S. military?* Santa Monica: RAND Corporation.

Knights, M. (2005). *Cradle of Conflict: Iraq and the Birth of the Modern US Military*. Annapolis: Naval Institute Press.

Knox, M., & Murray, W. (Eds.). (2001). *The Dynamics of Military Revolution, 1300-2050*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511817335

Libicki, M. (1996, Spring). The emerging primacy of information. *Orbis*, 40(2), 261–274. doi:10.1016/S0030-4387(96)90064-2

Linn, B. M. (2007). *The Echo of Battle: The Army's War of War*. Cambridge: Harvard University Press. doi:10.4159/9780674033528

Lonsdale, D. J. (2004). *The Nature of War in the Information Age: Clausewitzian Future*. New York: Frank Cass.

Parker, G. (1976, June). The Military Revolution, 1560-1660—a Myth? *The Journal of Modern History*, 48(2), 195–214. doi:10.1086/241429

Reardon, R. J., & Charlston, J. A. (2007). *From Transformation to Combat: The First Stryker Brigade at War*. Washington, DC: Center of Military History.

Richter, A. (2004). Lessons From the Revolution: What Recent US Military Operations Reveal About the Revolution in Military Affairs and Future Combat. *Journal of Military and Strategic Studies*, 7(3), 1–27.

Schörnig, N. (2010). *Robot Warriors. Why the Western investment into military robots might backfire*. Frankfurt am Main: Peace Research Institute Frankfurt.

Smith, R. (2005). *The Utility of Force: The Art of War in the Modern World*. London: Penguin Books.

Taw, J. M. (2012). *Mission Revolution: The U.S. Military and Stability Operations*. New York: Columbia University Press. doi:10.7312/taw-15324

van Crevald, M. (1991). *Technology and War: From 2000 B.C. to the Present*. New York: The Free Press.

Zucchini, D. (2004). *Thunder Run: The Armored Strike to Capture Baghdad*. New York: Grove Press.

KEY TERMS AND DEFINITIONS

Effect-Based Operations: The United States military concept for planning, executing and assessing operations to bring about particular enemy's behavioral effects by using the combination of military and non-military methods.

Future Combat Systems: The United States Army's modernization program that integrates manned and unmanned vehicles with decentralized data network. The program assumed creating a "medium-weight" brigade with increased strategic mobility thanks to using platforms transportable on the C-130 aircraft.

Information-Based Revolution in Military Affairs: The revolution in military affairs based on the application of information technology. It relies on the synergy between three elements: information acquisition, information processing and transfer, and the use of information in order to enhance firepower.

Network-Centric Warfare: The United States military doctrine assuming the possibility of transforming informative advantage (achieved by applying advanced information technology; first of all, connecting every element of a battle unit with decentralized data networks) into a full-spectrum dominance in the battle space.

Revolution in Military Affairs: A fundamental change in the character and conduct of armed conflicts. It typically consists of three elements: technological change, innovative operational concept, and organizational adaptation.

Smart Weapon: Munitions which allow hitting the target with high accuracy using a guidance system.

Stand-Off Weapon: Weapons launched at a long distance from the target to avoid defensive fire from the target area.

The Networked Effect of Children and Online Digital Technologies

Teresa Sofia Pereira Dias de Castro

University of Minho, Portugal

António Osório

University of Minho, Portugal

Emma Bond

University Campus Suffolk, UK

INTRODUCTION

Given the ubiquity of technology and the changes it has undergone in terms of its multiple uses, meanings and effects in the later decades of the twentieth century, a growing attention started to be paid to social studies of science and social studies of technology (Goggin, 2006); and to technology (world of production, or what technology does to people) and the social (world of consumption, or what people do with technology), as part of a sociotechnical order (Matthewman, 2011).

As the society and technology context became more complex, thinking about the meaning and use of technology, especially when referring to children, entails reasoning about its effects, which sometimes are good and sometimes are bad. When unforeseen consequences occur, more questions are raised about technology. In this respect, the debates around politics cannot be ignored (e.g. Berg & Lie, 1995; Joerges, 1999; Latour, 2004; Winner, 1980) behind the design of technology, paradoxically encapsulating architectures of social control and domination (e.g. school, prison, hospital as institutions of modernity grounded in the principle of panopticism, see Foucault, 1979) that determine “human experience, behaviour and action” (Matthewman, 2011, p. 50). While at the same time it enables rich experiences, new freedoms, heightened pleasures of consumption, and new forms of public privacy (Matthewman,

2011, referring to Benjamin, 2004 dissertation about the Arcades of Paris project, a structure made of iron and glass he designated by the ‘human aquarium’).

Reactions against technological determinism¹, strongly suggested in the literature on perceptions of risk (see Anderson, 2006; Beck, 1992; Buckingham, 2009; Douglas & Wildavsky, 1982; Furedi, 1997; Giddens, 1990; MacKenzie & Wajcman, 1999), invited theorists to rethink technology in light of its social effects and on “explaining how social processes, actions and structures relate to technology” (Mackay & Gillespie, 1992, p. 685). Such theories are the social shaping of technology² (SST) and social construction of technology (SCOT). However, the interwoven relationship between online digital technologies and children’s everyday lives is far more complex than can be accounted for or reduced by technological or social determinisms (Bijker & Law, 1992; MacKenzie & Wajcman, 1999). Since both positions are equally flawed (Van Loon, 2002), a more cautioned position between contingency and control is proposed. Actor-network theory (ANT) appears as an alternative to the previous deterministic models. This model assumes a high degree of contingency foregrounding the role of technology in the construction of society (Matthewman, 2011), where “the power of things depends on how they are (as Latour says) ‘syntagmatically’ networked with other things, in competition with

paradigmatic counter-programmes of differently coupled actants. The power of things does not lie in themselves. It lies in their associations” (Joerges, 1999, p. 5). When referring to children, as Prout (2005) proposes, ANT solves the technology/society dualisms unresolved by the previous proposals and is a valuable possibility to understand the hybrid and interconnected phenomenon of childhood growing up in the context of the globalised, mobile and wireless late modern society. Ultimately, he finds in ANT a flexible epistemological possibility that can reach the ambiguity of contemporary life and the remarkable transformations brought by progress that have changed drastically childhood and children’s contemporary lives.

Corroborating Prout stand, the text develops from empirical examples that illustrate the embeddedness of online digital technologies within children’s everyday lives, in which one is transformed and transformative of the other (Holloway & Valentine, 2003), though neither “cannot be understood as either human or technical, as neither human nor technology controls the resulting patterns of relationships” (Bond, 2014, p. 62). Children’s accounts are presented here as part of a doctoral thesis guided by two objectives: to understand the personal values and meanings children might use to interpret their technologized lives; and to explore the various aspects enclosed in children’s heterogeneous and complex networked lives with the purpose of uncovering short and long term unintended harmful outcomes that may be (more or less) hidden in their digital experiences. This is a qualitative study that recognises children as competent ‘agents in their own right’ and is informed by a multi-lens approach drawing perspectives from the sociology of risk, childhood studies, socio-technical studies and zemiology in order to get a better understanding of children’s active choices and decisions in how they use the online digital technologies in their everyday lives.

Put simply, online and digital technologies frame children’s worlds and how children shape online digital technologies in their everyday lives.

In this scope, technologies and children are both agents in a relationship that implies more than an object/subject correlation. Technologies stress interactivity (between human and non-human materials), convergence (by implying the combination of different technologies – the object, the activity, and the knowledge), and agency (by stimulating and allowing the usage), while ‘enframing’³ and revealing their non-neutrality (Heidegger, 1977). Children, on the other hand, give subjectivity to technology in the way they use them to act in their own worlds and by how they relate to, with and through technologies (Heidegger, 1977; Matthewman, 2011), while affirming and confirming a modern mode of thought and being, as proclaimed by Heidegger (Van Loon, 2002).

BACKGROUND

The Actor-Network Theory: A Brief Contextualisation

In cultural and media studies one theory attracted attention, the Actor-Network Theory (ANT) (Goggin, 2006), a “conceptual frame for exploring collective sociotechnical processes” (Crawford, 2004, p. 1) and a conducive approach to benefit from a deeper understanding of the mutually constituting interaction, between technological affordances (Parchoma, 2014) and modern childhood (Prout, 2005) in the intersection of late modernity. Deriving from Bruno Latour’s, Michel Callon’s, and John Law’s work, the ANT method “draws on semiotics, social constructionism and symbolic interactionism, while being quite distinct from each of them in certain crucial respects” (Prout, 2005, p. 70). Despite its debilities and limitations – ANT is criticized for reducing the importance of human materials (Matthewman, 2011), it is argued that ANT offers a radical analytical method capable of opening the black-box of technology. Tracing its complex and heterogeneous relationships, reframing, engaging and intervening in familiar

issues using an unfamiliar perspective (Fenwick & Edwards, 2010), “while controversies are still ranging” (Matthewman, 2011, p.107).

This approach is particularly helpful in understanding shifting landscapes, like children and online digital technologies, since it avoids the distinction between ‘technical’ and ‘social’, replacing them by ‘actants’⁴ and ‘networks’⁵ (an actant is a network and vice-versa) enrolled in relationships of negotiation and translation (Neyland, 2006). And “[s]ociety, technology, and even agency, are network-effects” (Matthewman, 2011, p. 110) combining and exchanging competencies and exerting agency within the actor-network structure (Matthewman, 2011). The heterogeneous structure of ANT reclaims to technology the ontological materiality and agency previously replaced by stabilization and meaning in the SCOT model (Matthewman, 2011).

ANT examines social life through the relational associations between human and non-human entities (Prout, 1996), enrolled in a network and performing micro-interactions (e.g. social, economic, natural, educational) (Fenwick & Edwards, 2010). Human and non-human entities share the same ontology and agency within symmetric relationships, i.e. none has a privileged status over the other, and it is through these networked associations that all entities change and are changed by each other, since all (things, persons, knowledge, processes) are relational effects (Crawford, 2004; Fenwick & Edwards, 2010; Matthewman, 2011). In line with this, all entities confer significance, action, subjectivity, intention and essence to each other (designated as relational materiality) (Crawford, 2004). In short, ANT scrutinizes how the network is created, its associations and what holds the network together temporarily (e.g. resistances, negotiations, exclusions) (Fenwick & Edwards, 2010). Drawing from Foucault’s intellectual influence, ANT also envisages the materiality of politics and power affecting the position of actants in the network; and shares his definition of power as a strategy to affect action in a chain of human

and non-human entities (where, simultaneously, objects can act on subjects; and subjects can become objects) (Matthewman, 2011).

To deal with the ambiguity brought by the dissolving boundaries (e.g. public/private; agency/structure; inclusion/exclusion; adult/children) ANT discards the social dualisms that mark contemporary networked children’s experiences. The ANT model is a “more rigorous monistic strand” (Prout, 2005, p. 71) – a framework to understand the complexities that surround late modern childhood and the technological texts⁶ (Fenwick & Edwards, 2010), because things only have significance through their relationships within a collective (Matthewman, 2011).

Nevertheless, Star (1991) offering a criticism to ANT, contends that also those outside the network must be brought in to the analysis:

[...] we should consider the marginalized and the oppressed, those that do not get to design technologies, those that cannot access them, and those that, nonetheless, are compelled to feel their effects. Being outside the privileged network does not mean that you will be free of the effects of the technologies. (Star, 1991, cited in Matthewman, 2011, p. 121).

The network metaphor is a theory that describes – in Latour’s work description coincides with explanation – the social context as it arises from the connections between the actants through their activities⁷ (Mutzel, 2009). As Latour asserts, “[t]he meanings of artifacts are to be found in their use” (Matthewman, 2011, p. 71), which means that the social context does not arise a priori out of the networks; the social context is built from the networks and from the links that modify the actants, which means that nothing really exists outside the networked relations (Bajde, 2013; Latour, 1996). In line with this, ANT suggests that the relations between actants denote movement, displacement, transformation, enrolment and any contact and action is mediated:

[...] actants create the social by association and by translating one meaning into another – and therefore are always much more than ‘mere informants’. Analysts follow the actors and their constitution of categories instead of defining categories a priori. (Mutzel, 2009, p. 877).

Regarding the human-technology relation, the ANT model introduces the concept of affordances, proposed by Akrich and Latour. By affordances one understands “[w]hat a device allows or forbids from the actors – humans and nonhuman – that it anticipates; it is the morality of a setting both negative (what it prescribes) and positive (what it permits)” (Cited in Bloomfield et al., 2010, p. 416). For Latour (2002) the affordance is at the same time permission and promise in which the tool becomes an embodied experience, where technology fulfils more than a purpose, their mediation creates new possibilities, what the author calls being-as-another:

[...] thanks to the hammer, I become literally another man, a man who has become ‘other’, since from that point in time I pass through alterity, [...] This is why the theme of the tool as an ‘extension of the organ’ makes such little sense. Those who believe that tools are simple utensils have never held a hammer in their hand, have never allowed themselves to recognize the flux of possibilities that they are suddenly able to envisage. (Latour, 2002, p. 250).

Technologies also delegate and exhort morality, which means humans do not need to be present, as they can act on their behalf, displacing public morality to public interest to achieve changes in form and substance (Matthewman, 2011). An interesting example (see Christie, 2014) recently highlighted in the news, illustrates how through the panopticon principle technology may exert morality. The following example unfolds how human, non-human, knowledge and symbolic relations (Matthewman, 2011) work together to explain the public morality-interest movement.

A mother, tired of her son’s inability to answer her calls or texts (public morality – e.g. ‘answer the phone, it is important’), decided in association with an application developer, to create an application (named ‘Ignore no more’)⁸ that enables the deactivation of children’s mobile phone in case they don’t answer parents calls (morality role of technology), making it impossible to call or text anyone but the parents (public interest – e.g. ‘answer the phone, it is going to be deactivated’).

The affordance concept is compatible with the analytical tool for the social study of technology that later would be proposed by Ian Hutchby – inspired in Gibson’s work about ‘action possibilities’ – as a ‘third way’ to overcome the technological determinism/social constructionism impasse (Bloomfield et al., 2010; Matthewman, 2011). Hutchby (2001b) explains that different technologies have different affordances and to ignore this is to constrain technologies possibilities – in terms of meanings and uses. Affordances emerge during the interaction process between user and digital device (Still & Dark, 2013). To study the affordances of the tool, it may be a helpful methodology to develop a better insight of the properties (positive and negative affordances (Scarantino, 2003)) of different technologies and take advantage of this knowledge, in terms of avoiding unintended consequences and abuses (Conole & Dyke, 2004), that may arise from misinterpreting the technology.

Affordances can be designed⁹ and anticipated in order to enhance their detection (Xenakis & Arnellos, 2013) as pre-existent properties of the artefact (independent of users perception) or dependent of user’s experience or goals (Kannengiesser and Gero 2012, cited in Still & Dark, 2013). The designer invites users to behave (Withagen et al., 2012) and communicate meaning through the artefact (Xenakis & Arnellos, 2013). By the end of the process, the final consumer perceives (or not) the affordances and may also decide how he/she wants to give meaning to them (Hutchby, 2001b). In agreement with the users context, knowledge, personal needs and intentions, he/she redefines

affordances that were not found previously (e.g. as in the case of the telephone, PC or the Walkman). Hartson (2003, cited in Tsai & Ho, 2013) predicts four types of affordance in the design and evaluation process: cognitive, physical, sensory and functional:

[...] cognitive affordance is design feature that helps users in knowing something; physical affordance is design feature that helps users in doing a physical action in the interface; sensory affordance is design feature that helps users sense something such as seeing or hearing, and functional affordance is design feature that helps users accomplish work. (Tsai & Ho, 2013, pp. 1250-1251)

Tsai and Ho (2013) explored how design affordance may affect smartphone usage and explain how the four types of affordance considered in the design process apply in the design-device and inferential-user interaction:

The label font size in the icon of icon [sic] on smartphone screen is large enough to read easily is an example of sensory affordances. In comparison with other three types of affordances, the functional affordance is a higher-level affordance and it could be seen as the usefulness of a system function. For instance, the process of accomplishing screen unlocking (functional affordance) on iPhone 4 might be decomposed into three stages – seeing the arrowhead displayed on the screen (sensory affordance), understanding on the meaning of arrowhead (cognitive affordance) and swiping the arrowhead touch panel (physical affordance). (Tsai & Ho, 2013, p. 1251).

A BRIEF EXPLANATION ABOUT METHODOLOGY AND ETHICS

This text builds from selected data collected using a qualitative approach informed by the constructivist tradition (Guba & Lincoln, 1998). Participatory

strategies were privileged to generate in-depth narrative data with school-aged children from diverse socio-economic backgrounds and from younger ages in line with gaps identified in the literature (Livingstone & Bulger, 2013). A total of 41 participants aged mainly between 10 and 12 years old were enrolled in this participatory research.

The ethical framework that guided this study endorsed the consequentialist model elaborated from the feminist ethic of care (Denzin & Lincoln, 1994) and ethical symmetry (Christensen & Prout, 2002). To protect children's rights to privacy and confidentiality, identities and personal information about the participants were concealed. Respondents' identities were replaced by a pseudonym.

Dialogical, flexible, reflexive techniques supporting participation and power balance were privileged with the aim of grasping the subjectivities, complexities and contradictions inherent to children's digital lives and giving children the opportunity to express their knowledge, share experiences and reflect on everyday situations. In line with this, group activities (e.g. games, drawing and role playing games), focus groups, group and individual interviews and participant observation were the techniques applied to stimulate discussion about a topic, to observe the participants' behaviour, routines and events occurring in their everyday contexts, offering interesting insights about their digital habits and interactions. During the research meetings children were invited to express their opinions, share their experiences and reflect on situations that emerge within the context of their digital and online lives. With the purpose of gaining a rich interpretation and reflexivity across the true purport of children's understandings, feelings and experiences, two methods, including thematic and narrative analysis, were applied to scrutinise and organise the data.

To solve the quality issue from a social-constructivist research standpoint and ensure the quality of the analytical generalizations (Yin, 2010), triangulation was applied by combining

different techniques for data collection: inter-methods triangulation, triangulation of data, and theoretical triangulation.

RESULTS AND DISCUSSION

Children's relationship *to, with and through* technology (Heidegger, 1977; Matthewman, 2011), *how they use and what they use it for* (Buckingham, 2006), is changing and challenging both adults' expectations about childhood, as well as children's everyday experiences (Bond, 2014). New forms of sociality (Matthewman, 2011), the dynamic construction of the self (Giddens, 1991), and new meanings and cultures of use (du Gay et al., 1997) are outcomes that nourish and are nourished by post-modernity corollaries – anxiety, insecurity, uncertainty and ambivalence (Ekberg, 2007) and are heightened by complex and complicated relationships and interdependencies where children, family, society, and digital online technologies are network-effects exerting agency within a complex actor-network structure (Latour, 1996; Matthewman, 2011; Prout, 2005).

Due to the changes and challenges that ubiquitous communication technologies brought to social dynamics, the boundaries between positive and negative experiences are often blurred, so one has to rely more on ambiguity than on certainty. In accordance with this matrix, data analysis demonstrates how a complex network of human and technical agents is challenging children's meaningful use of their participation rights online, reconfiguring the landscape of childhood, contributing to the magnification of old problems and the creation of new forms of vulnerability (Yar, 2012).

Upholding boyd's argument (2014) the way social media are designed challenges private/public and security/insecurity boundaries and encourages certain practices. On the one hand, challenge children's control when "balancing privacy and safety" (boyd, 2014, p. 47) and, on the other hand, afford children's online participation to

reach a broader, (un)known and (in)visible audience while, simultaneously, by default make them more easily searchable by such audience – often inaccurate with the one imagined by the children. During the fieldwork some participants reported concerning about their privacy and safety online. However, field impressions suggest that does not mean they are actually taking advantage of private settings, because they simply "don't know how to do that" (Grace, aged 12) and, to complicate matters, at least in the case of the Facebook, privacy settings are complex and constantly changing (boyd, 2014). When asked if they would publish a picture of themselves in a public space, all participants responded with a peremptory "no" as an interesting excerpt, depicted from one of the groups, reveals.

All participants in unison: No.

Grace [aged 12]: Everyone would see us. [...] 'Hey how embarrassing.' [She laughs] (PG4_251013)

Nevertheless, some children took real interest in changing their social profiles setting, making them more private, protected and secure. Participants seem to also be aware that a wider public can observe what they publish on the social networks. As Figaro clarifies Facebook is "always public" and "everyone can see" (PG1_270114) and "all the world can know", Tinker Bell adds (PG2_280114). Nonetheless, and despite their parents' anxieties, some children reported accepting friend requests from strangers, which according to them is not 'a big deal' as long as they do not talk to them.

Alana [aged 12]: I add [strangers], but I don't... [Interrupted]

Figaro [aged 12]: [Finishes Alana's sentence] But she never chats with them¹⁰ [...] (PG1_111113)

As Figaro afterwards explains, "It's just to have friends on Facebook" (PG1_211013). This seems to be reason enough to overlook some

precautions, such as for instance privacy settings regarding their personal pictures, making it possible for the wider public to lurk in, disregarding that their photographs can be downloaded and distributed. Children seem to adopt a ‘nothing happens to me’ attitude that confers them an illusory sense of control.

Moreover, other digital devices and online platforms, such as Ask.fm, Facebook and Stardoll were often part of children’s narratives, the scenario where gossip, rumours, rudeness, pranks, abusive requests and abuse of trust may occur in the fluidity of their on/offline interactions. Taking advantage of its main feature – anonymity, Ask.fm enables users to sometimes interact in a harmful/abusive fashion as the following participants’ accounts depict. Tinker Bell talks about a girl friend who published a comment saying that she cut herself (not on purpose) and how “everybody sent messages saying: ‘kill yourself’, ‘cut yourself’ ‘die’” (PG2_040214). And Grace, Rose and Maggie address the case of a young and well-known Portuguese actress who was targeted by rude comments on Ask.fm. In this situation, it is noteworthy how Maggie censures people’s character while Grace blames social software affordances, exemplifying the double-sword of modernity (Giddens, 1990). Maggie and Grace empathise with the young actress episode, because when this happens to them they feel sad and “everyone sees it” (PG4_251013).

Rose [aged 11]: *She took a picture. And a boy said: ‘If I were you, I would take forty pills and die’.*

Grace [aged 12]: *Ah, you know why? She has little breasts and people tease with her: ‘Hey, when will your boobs grow?’ She is 16 years old. ‘And your boobs? Are they hidden in your ass?’ People always say that, on Ask. That thing you ask questions.*

[...]

Researcher: *Do you think that people would have the courage to say that in people’s faces?*

Maggie [aged 12]: *They are cowards.*

Grace [aged 12]: *[...] do you know why? Because they can say it in private. (PG4_251013)*

Ask.fm and Facebook are platforms where rumours and gossip start and spread very quickly as the girls’ accounts below reveal. Giselle sent emoticons with little hearts to Kiara using a boy’s Facebook profile to which Kiara responded. The plot was set for what was to follow. Despite not having an account on Ask.fm, Kiara suffered from gossip and questions on the social network through Tinker Bell’s account, validating Star’s (1991) argument that even the ones outside the network are constrained by its effects.

Tinker Bell [aged 11]: *Two people from the school and two people on Ask came to me to tell me that I gave a kiss to [name of the boy]. And I gave him no kiss. Then, they said that [...] [name of the boy] had S-E-X with Kiara.*

[...]

Tinker Bell [aged 11]: *And the questions appear on my Ask. [Name of a girl] also received it.*

[...]

Tinker Bell [aged 11]: *They told they were a couple because Giselle sent little hearts to Kiara from [name of the boy]’s Facebook. And she replied: ‘You’re always in my heart.’ And someone must have read it in the Library and came to ask questions, to Ask.fm, ‘How long are they are sweethearts?’ And so. (PG2_180214)*

Taking advantage of SNS audiences, rumours published and republished on the Facebook can circulate largely to the point where no one can trace the source of the story (Best & Bogle, 2014). Nevertheless, the stories circulate as real, and to authenticate the evidence they tend to be localized in terms of space (nearby) and time (recently) while they also carry the message that the world is a dangerous place (Best & Bogle, 2014). During a meeting, participants were talking about a

paedophile that was supposedly acting ‘nearby’. According to Tinker Bell, a friend of hers published the alert on the Facebook and shared it on Tinker Bell’s wall.

Tinker Bell [aged 11]: *I don’t know if I told you that this girl posted twice on Facebook saying that a paedophile was in the neighbourhood.*

[...]

Tinker Bell [aged 11]: *She asked me for permission to put it on my Facebook too. I told her ‘Yes’. There’s nothing wrong about that.*

Giselle [aged 10]: *Have they already caught the paedophile of [locality]?*

Tinker Bell [aged 11]: *I don’t know. I never heard of him again. (PG2_261113)*

Also the source of the alert makes the story believable as several others corroborate the story. Weeks later, during a meeting, Tinker Bell announces that the paedophile was caught by the police. When asked if she saw it in the news, once again the story develops untraceable.

Tinker Bell [aged 11]: *They [police] already caught the paedophiles of [locality].*

Researcher: *Yes?*

Tinker Bell [aged 11]: *They caught one. It was in the newspaper.*

Researcher: *But have you seen the newspaper?*

Tinker Bell [aged 11]: *No, a friend of mine told me. Her godmother lives there too. (PG2_140114)*

Children regard digital and online worlds as just another setting where they manage their daily lives and relationships surpassing the dichotomy off/online, because they are not different or separate, but constitutive (Holloway & Valentine, 2003). In this sense, children’s everyday social relationships and activities are embedded within an off/online continuum.

Online social communities are popular among the participants, and many of their daily interactions happen within these platforms. However,

they are sometimes used to play pranks or being uncivil.

Excerpt A

Tinker Bell [aged 11]: *Ah in Stardoll I’m always blocking people, but that’s also because I want to.*

Researcher: *Because you are playful?*

Tinker Bell [aged 11]: *But not with my doll, I do it with another one I only created to annoy people.*

Researcher: *But on Stardoll, you can block and unblock [people]?*

Tinker Bell [aged 11]: *Yes.*

[...]

Kiara [aged 10]: *That is to make fun of people, you know? Because she... [interrupted]*

Tinker Bell [aged 11]: *Because she [talking about Giselle] creates dolls to tease Kiara.*

Researcher: *Who?*

Tinker Bell [aged 11]: *Giselle. She creates guys who are always nagging Kiara like this: ‘Do you want to date me? Do you want to date me? Do you want to date me?’*

Kiara [aged 10]: *‘Do you want to date me, you crooked teeth?’*

Tinker Bell [aged 11]: *‘You are ugly with crooked teeth, but I want to date you.’ It’s so cool to irritate people. It is better not to say this but [pause]. I like to call Kiara bad names. Silly, stupid, fool, sucker. Kiara talks to me: ‘Hello stupid.’ and I say: ‘Hello.’ And she says: ‘Shut up stupid.’ And I say, ‘I have not even talked’, and then she begins: ‘Shut up stupid.’*

Researcher: *Do you get sad when she says that? Never ever? In any day, any time?*

[She nods to me negatively]

Tinker Bell [aged 11]: *Ah. When she calls me bad names very [pause]. As once she called to [name of a girl], unintentionally. She called her a bitch. And [name of a girl] got very upset with her. (PG2_121113)*

Excerpt B

Tinker Bell tells me she has several dolls (Stardoll). One of them has her name, and when people ask for her name she doesn't lie, because, ultimately, they 'don't know if it's my true name or not'. And then she adds, 'I have three dolls. One of them is 'Evilllllll', with eight L's'. She uses this doll with 'ugly hair' when she wants to mistreat someone, because in her own words, it's 'fun'. She says: 'I call someone ugly, but I don't know if the person is really ugly, it's just for fun'. After explaining that 'when someone mistreats you, you can block and report' (or when people ask 'where are you from'), she also explains that she has 'fake' dolls, because if someone reports her, she is banned from the game. Having more than one doll is a strategy to do bad things without running the risk of being banned with the dolls they use to effectively play the game. And she adds, 'when I see someone mistreating a person, I stay quiet in a corner'. I got curious and asked her why she did that. She answered it was to avoid being banned. Tinker Bell and Kiara often play the game and mingle together in the game, where they sometimes simulate to mistreat each other just for fun. (Researcher's Field report PG2_221013)

The conversations above uphold Vandebosch and van Cleemput's (2008) argument that these practices are not intended to cause negative feelings or hurt someone. As Tinker Bell reinforces at the end of the first transcript it is *unintentional*. In their own words it is meant to be *fun*, "a sign of common understanding, or a kind of playful interaction between friends" (Vandebosch & van Cleemput, 2008, p. 501), nevertheless, despite common such practices may be interpreted differently by adults moral values, in particular, parents. In one exchange of text messages between the two girls, Kiara by accident sent the text messages to her mother and got punished because of the content (GP2_040214), which illustrates how children and adults have different perceptions of what is acceptable or not acceptable.

Taking advantage of mobile phone tariffs affording free voice and text communications, as well as anonymity, participants sometimes use this medium to insult each other. In the following situation Giselle is harassed by a girl via text message and, afterwards, by voice call. In the episode is noteworthy how the group got involved and how the situation loses significance when Giselle realizes that the stranger is in fact a girl with special educational needs from her class. This episode, despite being an isolated case, may indicate a new direction in research on bullying involving children with special educational needs and the use of online digital technologies that needs to be addressed and further investigated. The following event portrayed below reveals how online digital technologies empowered the girl with special needs to take an active and hostile role that challenges the taken for granted image of this group depicted as victims.

Kiara [aged 10]: Call her.

[...]

Giselle [aged 10]: Ya. I'll call her to see if I know the voice.

Tinker Bell [aged 11]: It's free to call her.

[...]

[The phone rings]

Giselle [aged 10]: Hello.

Girl X: Hello.

Giselle [aged 10]: Who is this?

Girl X: Clown.

Giselle [aged 10]: Who is this?

Girl X: [laughs] Stupid.

Giselle [aged 10]: Tell me your first name.

Girl X: [laughs] Geek.

Tinker Bell [aged 11]: It's [name of a girl] [whispering].

Giselle [aged 10]: I have her phone number [whispering].

Jenny [aged 10]: Call ended [whispering].

Giselle [aged 10]: Call terminated? I'll insist again.

[...]

[Girl X answers the phone again]

Girl X: What do you want, clown?
Giselle [aged 10]: Is that you Girl X?
Girl X: Clown.
Giselle [aged 10]: Piiiiiii.
Girl X: [laughs] Geek.
Giselle [aged 10]: Do you know me?
Giselle [aged 10] and Mary [aged 12] in unison:
Call ended. [Both girls Laugh]
Mary [aged 12]: Try again.
 [...]
[Girl X rejects the call]
 [...]
Tinker Bell [aged 11]: I also do this with Kiara.
Kiara [aged 10]: That's right. And I do it with her.
Giselle [aged 10]: Dumb.
Researcher: This?
Tinker Bell [aged 11] and Kiara [aged 10] in unison: Yes.
Researcher: Seriously?
Kiara [aged 10]: Hum hum.
Researcher: Why?
Tinker Bell [aged 11]: Unknown number. I call her and pretend I'm someone else.
Kiara [aged 10]: Ya.
Researcher: Why?
 [...]
Tinker Bell [aged 11]: It's funny.
Researcher: It's funny?
Tinker Bell [aged 11]: Yes.
Kiara [aged 10]: Hum hum.
Tinker Bell [aged 11]: But when we send messages, it has her name and has my name.
[Both girls laugh]
 [...]
Researcher: What are you going to do about this?
Giselle [aged 10]: I don't know. If it's a friend of mine, I'll do nothing.
 [...]

(GP2_140214)

CONCLUSION

The combination of childhood and technologies serves as a powerful focus for debates (not new in

the media field) that have been struggling between renewed hopes and fears, and risks and opportunities, where digital and online technologies are responsible for liberating and empowering versus destructing and betraying the essence of childhood. Clearly, as unsupervised personal, online and mobile technology becomes more entrenched in children's lives more complex are the challenges to face. And though most of children's digital and online experiences are positive, evidence suggests that children can easily engage in ambivalent interactions and decisions.

Nevertheless, this embeddedness of technologies in children's quotidian has to be understood in light of other macro, micro and meso societal changes that need further attention as these affect the social reality of families and consequently children's everyday lives. Conforming with this, the narratives included in this text provide the unsettling context adjoining the social processes arising within the scope of late modernity, enhancing tensions (e.g. agency-structure, public-private) that impact on social representations of childhood and through which children become visible or invisible; and the rationales that may shed further light on the extent to which the digital realm is embedded in children's everyday lives through heterogeneous and complex networks that link together human and non-human entities in which "children's social worlds are formed, not just through their associations with other children, but also through their association with their material surroundings" (Valentine & Holloway 2002, p. 317).

Drawing from everyday situations, children's voices were privileged as the main source of understanding to inform a multi-lens approach drawing perspectives from the sociology of risk, childhood studies, socio-technical studies and zemiology. In particular, this study benefit from a conducive approach, the Actor-Network Theory, as a conceptual framework to gain a enriching understanding of the mutually constituting interaction between technological affordances and modern childhood, arising from the intersection of late modernity. Considering the rapid pace of

technological innovation and how children (re) adapt to its fast pace, creatively interpret technological affordances and (re)construct media trends, the fieldwork follows a research with children approach in which children were encouraged to reflect and share their understanding vis-à-vis their technologized quotidian. Considering participants' accounts, interpersonal relationships can indeed become sour when the technology is misinterpreted, as it is suggested, demonstrating how heterogeneous and complex networks influence human action and relationships.

In summary, ANT avoids technological and social reductionisms/determinisms, since what entities do when they meet is unpredictable (see Latour key example, "[t]he key does not act by itself, but is also acted upon by other entities linked in the network" (Fenwick & Edwards, 2010, p. 10)) meaning that,

Social life cannot, therefore, be reduced either to the 'purely' human or to the 'purely' technological (or animal, vegetable, mineral, abstract...). Neither the human nor the technological determines the overall patterning or ordering that results from their combination. (Prout, 2005, p. 70-71)

According to Prout (2005), with ANT new forms of childhood arise as a result of complex and heterogeneous network connections established between children and technologies. Artefacts influence human actions, intentions, meanings, relationships, routines, memories, perceptions of the self; and ideas, desires, meanings and actions influence the artefacts. Drawing from a school playground as an example, Fenwick and Edwards (2010) note that non-human materials (playground) when combined with human entities (children) "produce particular activities, speech, social groupings and exclusions, injuries, even gender identities" [...] [that] exclude, invite and regulate particular forms of participations. And such social practices can determine meaning by reinforcing the political effects of technology. As Moinian states, "we can neither simply blame nor

celebrate these new media and their impact on our conceptions of childhood without understanding their complexity and their potential contradictions" (2006, p. 51, paraphrasing Buckingham, 2000).

FUTURE RESEARCH DIRECTIONS

The examples included in this text offer a small fragment of knowledge, centred on specific problems and situations. However, these are extremely helpful to reach the subjectivity, the delicacy and complexity of matters and meanings children bring to their online experiences and to understand the networked complexities enclosed in the embedded relationship between children and online digital technologies in the context of late modernity. Children's accounts reflect the heterogeneity, the complexities and the context of late modernity and life itself with "new capacities to act and hence new fields of power" (Prout, 2005, p. 115) that create new possibilities and problems in need of ongoing research.

REFERENCES

- Anderson, A. (2006). Media and Risk. In G. Mythen & S. Walklate (Eds.), *Beyond the Risk Society: critical reflections on risk and human security* (pp. 114–131). Open University Press.
- Bajde, D. (2013). Consumer culture theory (re) visits actor-network theory: Flattening consumption studies. *Marketing Theory*, 13(2), 227–242. doi:10.1177/1470593113477887
- Beck, U. (1992). *Risk society: towards a new modernity*. SAGE Publications Ltd.
- Benjamin, W. (2004). *The Arcades project* (H. Eiland & K. McLaughlin, Trans.). Cambridge, MA: Belknap Press.
- Berg, A.-J., & Lie, M. (1995). Feminism and Constructivism: Do artifacts have gender? *Science, Technology & Human Values*, 20(3), 332–351.

- Best, J., & Bogle, K. A. (2014). *Kids gone wild. From rainbows parties to sexting, understanding the hype over teen sex*. New York: New York University Press.
- Bjiker, W. E., & Law, J. (1992). General introduction. In W. E. Bjiker & J. Law (Eds.), *Shaping Technology/Building Society*. Cambridge, MA: MIT Press.
- Bloomfield, B. P., Latham, Y., & Vurdubakis, T. (2010). Bodies, Technologies and Action Possibilities: When is an Affordance? *Sociology*, 44(3), 415–433. doi:10.1177/0038038510362469
- Bond, E. (2014). *Childhood, mobile technologies and everyday experiences - changing technologies = changing childhoods?* Palgrave Macmillan.
- boyd, d. (2014). *It's complicated: the social lives of networked teens*. New Haven, CT: Yale University Press.
- Buckingham, D. (2000). *After the death of childhood: growing up in the age of electronic media*. Polity Press.
- Buckingham, D. (2006). Is there a digital generation? In D. Buckingham & R. Willett (Eds.), *Digital generations - Children, young people, and new media* (pp. 1-13): Lawrence Erlbaum Associates, Taylor Francis Group.
- Buckingham, D. (2009). New media, new childhoods? In M. J. Kehily (Ed.), *An introduction to Childhood Studies* (2nd ed.; pp. 124–138). Open University Press.
- Christensen, P., & Prout, A. (2002). Working with Ethical Symmetry in Social Research with Children. *Childhood*, 9(4), 477–497. doi:10.1177/0907568202009004007
- Christie, J. (2014). Mom who served in the Gulf War creates 'genius' app that will shut down her son's phone when he ignores her calls... and now all parents can buy it. *Mail Online*. Retrieved from <http://www.dailymail.co.uk/news/article-2727267/Mom-served-Gulf-War-creates-genius-app-shut-sons-phone-ignores-calls.html>
- Conole, G., & Dyke, M. (2004). What are the affordances of information and communication technologies? *Alt-J*, 12(2), 113–124. doi:10.1080/0968776042000216183
- Crawford, C. (Ed.). (2004). *The Encyclopedia of Social Theory* (Vol. I). SAGE Publications, Inc.
- Denzin, N., & Lincoln, Y. (Eds.). (1994). *Handbook of qualitative research*. SAGE.
- Douglas, M., & Wildavsky, A. B. (1982). *Risk and Culture: An essay on the selection of technical and environmental dangers*. Berkeley, CA: University of California Press.
- du Gay, P., Hall, S., Janes, L., Mackay, H., & Negus, K. (1997). *Doing cultural studies - The story of the Sony walkman*. SAGE Publications in association with The Open University.
- Ekberg, M. (2007). The Parameters of the Risk Society: A Review and Exploration. *Current Sociology*, 55(3), 343–366. doi:10.1177/0011392107076080
- Fenwick, T., & Edwards, R. (2010). *Actor-Network theory in education*. Routledge.
- Foucault, M. (1979). *Discipline and punish: The birth of the prison* (A. Sheridan, Trans.). New York: Vintage Books.
- Furedi, F. (1997). Culture of fear revisited: Risk-taking and the morality of low expectation. *Continuum*.
- Giddens, A. (1990). *The consequences of modernity*. Polity Press.
- Giddens, A. (1991). *Modernity and self-identity: Self and society in late modern age*. Polity Press.
- Goggin, G. (2006). *Cell Phone Culture: Mobile technology in everyday life*. Routledge.
- Guba, E., & Lincoln, Y. (1998). Competing paradigms in qualitative research. In N. Denzin & Y. Lincoln (Eds.), *The landscape of qualitative research: theories and issues* (pp. 195-220). SAGE Publications.

- Hartson, H. R. (2003). Cognitive, physical, sensory, and functional affordances in interaction design. *Behaviour & Information Technology*, 22(5), 315–338. doi:10.1080/01449290310001592587
- Heidegger, M. (1977). *The question concerning technology and other essays* (W. Lovitt, Trans.). New York: Harper & Row.
- Holloway, S. L., & Valentine, G. (2003). *Ciberkids: Children in the information age*. Routledge Falmer.
- Hutchby, I. (2001). Technologies, Texts and Affordances. *Sociology*, 35(2), 441–456. doi:10.1177/S0038038501000219
- Joerges, B. (1999). Do politics have artefacts? *Social Studies of Science*, 29(3), 411–431. doi:10.1177/030631299029003004
- Kannengiesser, U., & Gero, J. S. (2012). A process framework of affordances in design. *Journal of Design Issues*, 28(1), 50–62. doi:10.1162/DESI_a_00123
- Latour, B. (1991). Technology is society made durable. In J. Law (Ed.), *A sociology of monsters: essays on power, technology and domination*, *Sociological Review Monograph N°38* (pp. 103–131). London: Routledge.
- Latour, B. (1996). On actor-network theory. A few clarifications plus more than a few complications. *Soziale Welt*, 47, 369–381.
- Latour, B. (2002). Morality and Technology: The end of the means. *Theory, Culture & Society*, 19(5/6), 247–260.
- Latour, B. (2004). *Which politics for which artifacts?* Retrieved from http://www.sts-biu.org/images/file/COURSE_READINGS/27-815
- Latour, B., & Venn, C. (2002). Morality and Technology: The end of the means. *Theory, Culture & Society*, 19(5/6), 247–260. doi:10.1177/026327602761899246
- Livingstone, S., & Bulger, M. (2013). A global agenda for children's rights in the digital age - Recommendations for developing UNICEF's Research Strategy. UNICEF.
- Mackay, H., & Gillespie, G. (1992). Extending the social shaping of technology approach: Ideology and appropriation. *Social Studies of Science*, 22(4), 685–716. doi:10.1177/030631292022004006
- MacKenzie, D., & Wajcman, J. (Eds.). (1999). *The social shaping of technology* (2nd ed.). Buckingham, UK: Open University Press.
- Matthewman, S. (2011). *Technology and social theory*. Palgrave Macmillan.
- Moinian, F. (2006). The Construction of Identity on the Internet: Oops! Ive left my diary open to the whole world! *Childhood*, 13(1), 49–68. doi:10.1177/0907568206058610
- Mutzel, S. (2009). Networks as Culturally Constituted Processes: A Comparison of Relational Sociology and Actor-network Theory. *Current Sociology*, 57(6), 871–887. doi:10.1177/0011392109342223
- Neyland, D. (2006). Dismissed Content and Discontent: An analysis of the Strategic aspects of Actor-Network Theory. *Science, Technology & Human Values*, 31(1), 29–51. doi:10.1177/0162243905280022
- Parchoma, G. (2014). The contested ontology of affordances: Implications for researching technological affordances for collaborative knowledge production. *Computers in Human Behavior*, 37, 360–368. doi:10.1016/j.chb.2012.05.028
- Prout, A. (1996). Actor-Network theory, technology and medical sociology: An illustrative analysis of the metered dose inhaler. *Sociology of Health & Illness*, 18(2), 198–219. doi:10.1111/1467-9566.ep10934726

Prout, A. (2005). *The future of childhood: Towards the interdisciplinary study of children*. London: Routledge Falmer, Taylor and Francis Group.

Scarantino, A. (2003). Affordances explained. *Philosophy of Science*, 70(5), 949-961.

Star, S. L. (1991). Power, technology and the phenomenology of conventions: on being allergic to onions. In J. Law (Ed.), *A Sociology of monsters: essays on power, technology and domination* (pp. 26-56). London: Routledge.

Still, J. D., & Dark, V. J. (2013). Cognitively describing and designing affordances. *Design Studies*, 34(3), 285-301. doi:10.1016/j.destud.2012.11.005

Tsai, J.-P., & Ho, C.-F. (2013). Does design matter? Affordance perspective on smartphone usage. *Industrial Management & Data Systems*, 113(9), 1248-1269. doi:10.1108/IMDS-04-2013-0168

Van Loon, J. (2002). *Risk and technological culture: towards a sociology of virulence*. London: Routledge. doi:10.4324/9780203466384

Vandebosch, H., & van Cleemput, K. (2008). Defining Cyberbullying: A Qualitative Research into the Perceptions of Youngsters. *CyberPsychology & Behaviour*, 11(4), 499-503. doi:10.1089/cpb.2007.0042 PMID:18721100

Williams, R., & Edge, D. (1996). The social shaping of technology. *Research Policy*, 25(6), 865-899. doi:10.1016/0048-7333(96)00885-2

Winner, L. (1980). Do artifacts have politics? *Daedalus*, 109(1), 121-136.

Withagen, R., de Poel, H. J., Araújo, D., & Pepping, G.-J. (2012). Affordances can invite behavior: Reconsidering the relationship between affordances and agency. *New Ideas in Psychology*, 30(2), 250-258. doi:10.1016/j.newideapsych.2011.12.003

Xenakis, I., & Arnellos, A. (2013). The relation between interaction aesthetics and affordances. *Design Studies*, 34(1), 57-73. doi:10.1016/j.destud.2012.05.004

Yar, M. (2012). E-Crime 2.0: The criminological landscape of new social media. *Information & Communications Technology Law*, 21(3), 207-219. doi:10.1080/13600834.2012.744224

Yin, R. (2010). Analytic Generalization. In A. J. Mills, G. Durepos, & E. Wiebe (Eds.), *Encyclopedia of case study research*. SAGE Publications.

KEY TERMS AND DEFINITIONS

Actor-Network Theory: Social life understood through a network where human and technological entities have agency to interact and transform each other within symmetric relationships.

Children: The term 'children' follows the United Nations Convention on the Rights of the Child (UNCRC) definition as those under 18 years.

Network: Network is a concept employed to suggest flow and connection among the heterogeneous human and technical entities.

Online Digital Technologies: Refers to devices and services connected to the internet that afford flexible communication and connectivity for the user. Nevertheless, sometimes the conjunction 'and' is used between 'online' and 'digital' to include the reference to devices or experiences that do not afford or imply a connection to the internet.

Social Construction of Technology: Technology understood as a social phenomenon embodied in complex social processes.

Social Determinism: The notion that society shapes technology.

Social Shaping of Technology: Technology understood as a social product.

Technological Affordances: The possibilities and permissions that a technological artefact encloses. These possibilities and permissions emerge during the interaction process between user and the technological artefact.

Technological Determinism: The notion that technology shapes society.

ENDNOTES

- ¹ “Technological determinism is the notion that technological development is autonomous with respect to society; it shapes society, but is not reciprocally influenced. Rather, it exists outside society, but at the same time influences social change.” (Mackay & Gillespie, 1992, p. 686).
- ² The SST “is seen as playing a positive role in integrating natural and social science concerns; in offering a greater understanding of the relationship between scientific excellence, technological innovation and economic and social well-being; and in broadening the policy agenda, for example, in the promotion and management of technological change” (Williams & Edge, 1996, p. 545).
- ³ The word ‘enframing’ is used by Heidegger (1977) “to describe modern technology’s way of revealing the world as standing-reserve” (Matthewman, 2011, p. 14).
- ⁴ An actant is an actor with agency after being translated and becoming part of the network. However, it is not helpful to distinguish between them once ‘actant’ is not much used in ANT analysis (Latour, 1991). For the sake of clarity, the concept ‘actant’ will be used instead of ‘actor’, to avoid being misinterpreted with ‘actor’ suggesting human agency (Matthewman, 2011).
- ⁵ Referring Latour’s and Woolgar’s work, Fenwick and Edwards state that “[i]n ANT’s early years, the term network was employed to suggest both flow and clear points of connection among the heterogeneous entities that became assembled to perform particular practices and processes” (2010, p. 15). However, with the advent of technological networked systems, the network metaphor became problematic.
- ⁶ Hutchby refers to Grint and Woolgar work to explain what ‘texts’ are. “[T]echnologies should be treated as ‘texts’ that are ‘written’ (i.e. configured) in certain ways by their developers, producers and marketers, and have to be ‘read’ (i.e. interpreted) by their users or consumers” (Hutchby, 2001, p. 445).
- ⁷ See Latour (1991) example of the hotel-room key.
- ⁸ Information retrieved from <http://www.theguardian.com/lifeandstyle/shortcuts/2014/aug/18/app-forces-teenagers-to-call-parents>
- ⁹ About the interaction of aesthetics and affordances in the design process of a product see Xenakis & Arnellos, 2013.
- ¹⁰ In Portuguese: “nunca dá à língua.”

Suggestions for Communication of Information for Multicultural Co-Existence

S

Noriko Kurata

Tokyo University of Science, Suwa, Japan

INTRODUCTION

Given the advancement of globalization today, an increasing number of people are emigrating from the country they were born in or from the country where people of their same ethnicity live, to other countries seeking economic affluence and a safe and secure life. Even in Japan, a country that is surrounded by sea, the number of resident foreigners is increasing every year. In any country and in any region, it is difficult to stop globalization. Local governments all over the world are facing the common challenge of fostering mutual understanding among existing local residents and people with multicultural backgrounds or multi-culture residents (MCR), rather than excluding them, and encouraging mutual acceptance.

In Japan, Japanese residents (JR) and MCR are frequently experiencing each other's culture through local events such as cooking classes and gymnastics exercises. Local governments have established many places for such experience-based cultural exchange. However, culture includes things that can only be expressed through language. These are ordinances. Ordinances are the norms of the region that regulate the rights and duties of the residents. If one does not understand the language accurately one cannot interpret the ordinances. Even for those people who cannot interpret ordinances correctly, if they violate an ordinance it is punishable by fine. Persons who do not know about an ordinance suffer disadvantages due to their ignorance. Ordinances are a part of the local culture that should be known to those who intend to live in the region.

As a means of knowing the local culture, there are ways of finding information related to local governments by searching on the website. Most information about local governments in Japan is distributed in Japanese, but through the introduction of machine translation, it is possible to convert it into a language that users themselves find easy to understand.

In this chapter, the examinations on the trend towards an increase in MCRs in Japan were made first, and examples of introduction of machine translation in Japan were also presented. Further, the results of surveys were described as status of communication of information in Japan's basic local governments, some issues were extracted, and available countermeasures were proposed.

BACKGROUND

Increase in Resident Foreigners in Japan

As shown in Figure 1, resident foreigners in Japan are increasing every year. According to the population census (Ministry of Internal Affairs and Communications, 1995; 2000; 2005; 2010; 2015), resident foreigners numbered about 1,140,000 in 1995, but have increased to more than 3,130,000 in 2015, increasing to more than double in 20 years. The composition ratio of resident foreigners in the total population of Japan in 1995 was about 0.9% but it was about 2.5% in 2015, increasing to nearly three times.

Figure 1. Trend of Resident Foreigners in Japan (Tabulated by the author based on “Population census”, Ministry of Internal Affairs and Communications)

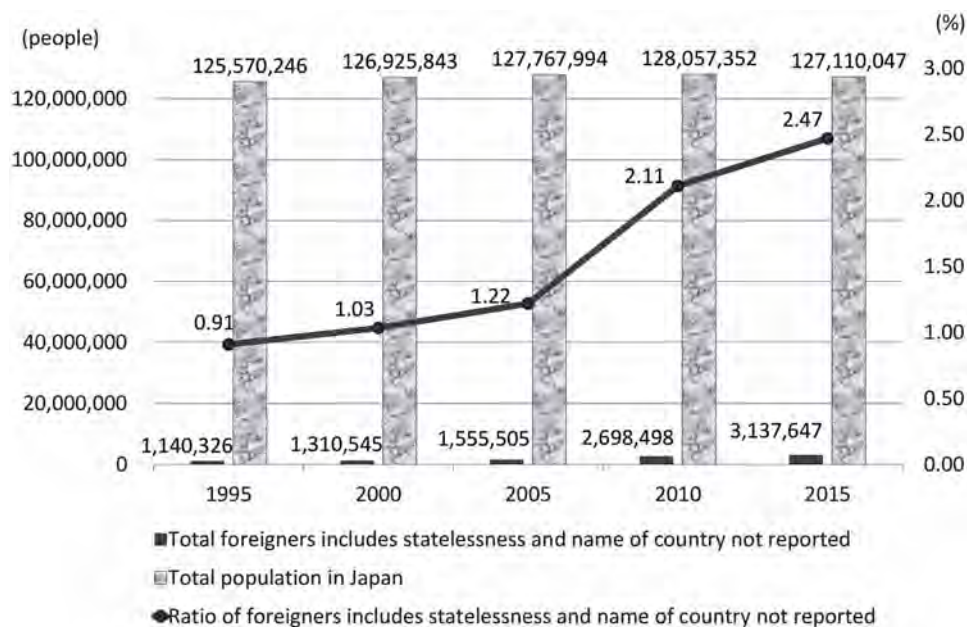
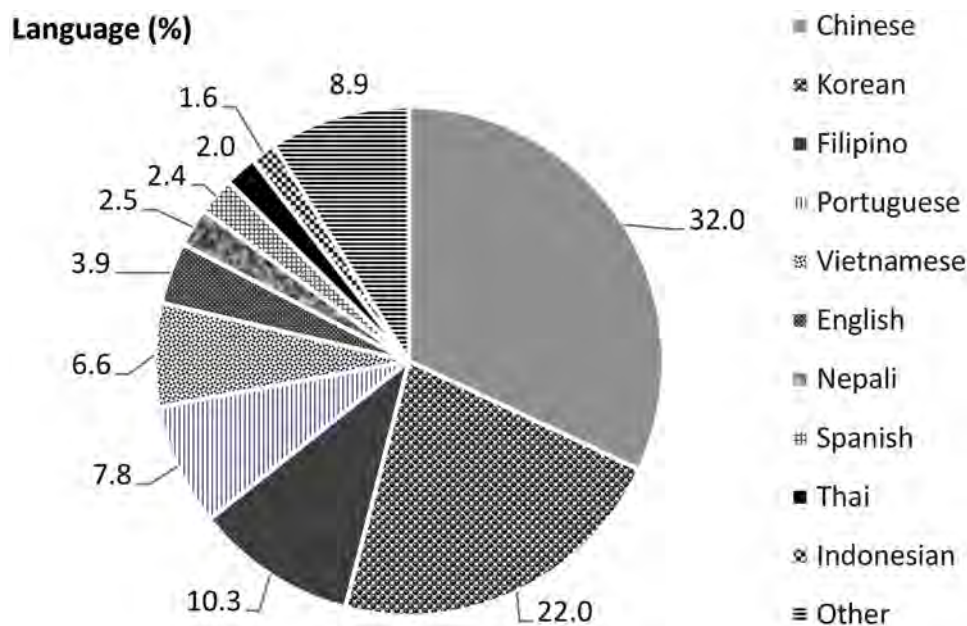


Figure 2 is based on December 2015 statistics on resident foreigners (Ministry of Justice, 2015). The language used by the resident foreigners is shown. Thirty countries from where there are a

large number of resident foreigners were extracted from this data and the calculation presents a total of the national languages and official languages used in these countries. Although Chinese-language

Figure 2. Languages Used by Resident Foreigners in Japan (Based on “Statistics on Resident Foreigners”, Ministry of Justice; tabulated by the author based on the languages of the top 30 countries from where there are a large number of resident foreigners)



users occupy more than 30% of the total, one can observe that users of a variety of languages reside in Japan.

This not only shows that there are diverse languages but also shows that there are people with various cultural backgrounds residing in Japan.

In the “Field Study Related to the Resident Foreigners’ Awareness of the Japanese Language” conducted by the Agency for Cultural Affairs (2001) targeting 600 foreigners residing in Japan, 10 questions were asked related to problematic situations and unpleasant situations arising due to lack of proficiency in Japanese-language conversation, reading and writing (multiple answers allowed). As a result, the maximum percentage was for “hospitals” at 21.3%, followed by “relations with neighbors” coming second at 17.0%, “workplace” being third at 15.1%, and “City Hall Office” being fourth at 11.5%. Difficulties were experienced in daily life and at the City Hall where explanations about the public system are received.

The above survey also examined Japanese-language proficiency. Regarding listening abilities, 21.0% answered “can understand well”, and regarding speaking skills, 17.4% answered “can speak well”. About reading abilities, 84.3% could read hiragana, 75.2% could read katakana, 51.5% could read romaji, 48.5% could read kanji a little, and 19.6% could read kanji and also understand the meaning.

Further, in the above survey, when asked whether they consciously learn Japanese in some way, about 90% (89.7%) answered “learning”. Those who were “learning” were asked eight questions about the method of learning, the most number

of responses was “at a Japanese-language class in the municipality” at 59.9%. People who attended Japanese-language classes were asked 12 questions about how they wanted to utilize the results of their learning at the Japanese-language classes (multiple answers allowed). Table 1 shows the top answers with high selectivity. These results show that resident foreigners have a high consciousness of wanting to know the culture of Japan through the Japanese language.

MACHINE TRANSLATION

Technological Progress

Automatic translation by machines is being researched as an aid to enable humans to understand multiple languages. It is a technology called machine translation (MT) or automatic translation. MT refers to the translation of text in one language (the source language) to another language (the target language), using software.

The methods of MT include the rule-based method in which experts write the translation rules and translation based on the paginal translation method. Translation based on paginal translation can be broadly classified into example-based machine translation (EBMT) and statistical machine translation (SMT). (Sumida, 2009)

The rule-based method can lead to a highly precise translation when there is compliance, but as making grammar rules for minor languages is difficult, it takes a huge amount of time for multi-lingualization.

Table 1. How would you like to utilize the results of learning at Japanese-language classes?

Order	Utilization of Content	Ratio (%)
1	Speaking Japanese	71.7
2	Learning about daily life and culture	54.2
3	Increasing the number of friends	47.7
4	Talking to people (finding a conversation partner)	43.7
5	Deepening understanding about different values (different cultures) from one's own	30.8

EBMT searches for similar parallel translation data from previously stored bilingual translation data, and conducts translation by revising this. (Kurohashi, 2009)

SMT extracts the model required for translation from the bilingual data and translates by maximizing the probability based on this. (Tsumamoto, 2009).

Research on SMT is increasing in accuracy in recent years, due to techniques of automatic evaluation of translation results.

Case Studies of Introduction of Machine Translation in Japan

In private translation systems, translation sites such as @nifty translation or Yahoo! Translation use the rule-based method and translation sites such as Google Translate and Bing Translator use the SMT method.

The Japan Patent Office has adopted SMT and in order to enhance the accuracy of MT of patent literature, they have provided a Japanese-English Machine Translation dictionary from 1999, and a Chinese-Japanese Machine Translation dictionary from 2012, on the web. (Japan Patent Office, 2015)

At The Ministry of Justice the “Japanese Law Translation Database System” was developed from 2009, offering the translation guide, “The Standard Legal Terms Dictionary”. In this database system, the following disclaimers are displayed:

This system is operated by the Ministry of Justice and is a website that provides translations of Japanese laws and regulations. All translations contained on the Japanese Law Translation Database System are unofficial. Only the original Japanese texts of the laws and regulations have legal effect, and the translations are to be used solely as reference materials to aid in the understanding of Japanese laws and regulations. The government of Japan is not responsible for the accuracy, reliability or currency of the legislative material provided in this website, or for any

consequence resulting from use of the information in this website. For all purposes of interpreting and applying law to any legal issue or dispute, users should consult the original Japanese texts published in the Official Gazette. (Ministry of Justice, 2009)

Evidence based on laws in the Japanese language is required for legal problems such as lawsuits. Rather than litigations, the information published here is sufficient if one wants to gain a rough understanding of rights and obligations in daily life.

CHALLENGES AND COMMUNICATION OF INFORMATION IN JAPAN’S LOCAL GOVERNMENTS

Websites Operated by Local Governments

Japan’s local governments have each developed websites and operate them (hereafter, local government webs). The local government webs disseminate information about garbage disposal, Japanese-language lessons, and information about daily life and events.

Usually all the pages of the local government webs are created in the Japanese language. Only a few local governments have established web pages in English or Portuguese and added them to local government webs in the Japanese language. These special web pages have accurate expressions as they have been translated by translators (human) and have high reliability. However, the amount of information displayed on these special pages is much less than the amount of information on the entire local government web in the Japanese language.

Recently, some local governments have introduced MT and the initiative to display the Japanese-language local government web itself in multiple languages is starting. Although, MT

is not accurate enough, a lot of information gets automatically translated in one click. Majority of the local governments that have introduced it support not only English but about five languages. If MCRs are living in an unknown culture, communication of information about the local culture through MT should be very useful. Nevertheless, not all the web pages of the Japanese language local government webs are suitable for MT.

The following survey was conducted to shed light on such “case studies that are incompatible” with MT.

Survey

In this survey the following three points were confirmed based on the observation of local government webs.

- Status of provisions for multicultural co-existence in the cultural and comprehensive plans
- Status of provisions related to communication of information in multiple languages in the cultural and comprehensive plans
- Status of dissemination of ordinances in multiple languages

The target of the survey were cities and special wards among local governments where the composition ratio of MCRs was more than 4% taking as basis the “Basic Count of Population” in the population census conducted by the Ministry of Internal Affairs and Communications in 2010.

In the 2010 population census (Ministry of Internal Affairs and Communications, 2010), the municipality with the highest composition ratio of MCRs in the population was Kawakami Village at 15.7%. The next highest was Oizumi Town at 13.0%, followed by Nanmoku Village at 11.4%, Shinjuku Ward at 7.9%, and Minokamo City at 7.7%. Towns and villages are local governments but as the scale of the population and finance is small and the number of employees are also few, the administrative services are not so substantial.

Therefore, in this survey, towns and villages are excluded and cities and special wards are targeted for study.

This survey was conducted in May 2016.

Survey Results and Discussion

Table 2 shows local governments that fulfil the abovementioned requirements. There were 15 cities and special wards targeted for the survey. Among these, Shinjuku Ward in Tokyo had the highest number of MCRs at 7.9%.

Table 3 shows the presence of publishing of “cultural plans” focusing on culture in the survey targets. Six cities or 40% of them had cultural plans published.

Further, in Table 3 shows the status of provisions for multicultural and multilingual communication of information in the plans of the local governments.

In the survey, in case of local governments that have not published cultural plans or who have published cultural plans but do not have provisions, the highest level plans, that is, the comprehensive plans (also called basic plans) of local governments were confirmed.

As a result, there were four local governments who had made provisions regarding multicultural co-existence, and eight local governments who had provisions in the comprehensive plans. Therefore, a total of 12 local governments had made provisions in any one of the plans. This constituted about 80% of the 15 local governments.

On the other hand, three local governments had made provisions for multilingual communication of information and five had made provisions in the comprehensive plans. A total of eight local governments had made provisions in either one of the plans.

If these are reviewed, provisions for multicultural co-existence were incorporated about 50% in the comprehensive plans and about 60% in the cultural plans, and provisions for multilingual communication of information were incorporated about 33% in comprehensive plans and about 60%

Table 2. Cities and special wards with more than 4% in the population composition ratio of the foreigners

Rank	City/Special Ward(-ku)	Prefecture	Composition Ratio of Foreigners (%)	Nationality Except Japan (Total Number) (People)	Total Population (People)
1	Shinjyuku-ku	Tokyo	7.9	25, 742	326, 309
2	Minokamo City	Gifu	7.7	4, 189	54, 729
3	Minato-ku	Tokyo	6.3	12, 999	205, 131
4	Toshima-ku	Tokyo	6.3	17, 888	284, 678
5	Arakawa-ku	Tokyo	5.7	11, 625	203, 296
6	Kikugawa City	Shizuoka	5.1	2, 413	47, 041
7	Chiryu City	Aichi	5.1	3, 466	68, 398
8	Chuo City	Yamanashi	4.7	1, 484	31, 322
9	Taito-ku	Tokyo	4.5	7, 886	175, 928
10	Joso City	Ibaraki	4.4	2, 890	65, 320
11	Kani City	Gifu	4.2	4, 094	97, 436
12	Moka City	Tochigi	4.1	3, 410	82, 289
13	Iga City	Mie	4.1	3, 978	97, 207
14	Isesaki City	Gunma	4.1	8, 419	207, 221
15	Kosai City	Shizuoka	4.0	2, 375	60, 107

Source: Adapted from reference Kurata, N., Kurata, Y., and Ohashi., 2016

Data from “Population census 2010”.

Table 3. Current status of prescription on multicultural coexistence and multilingual information dissemination of 15 local governments shown in Table 2 (Unit: City/Special ward)

	Found on the Web	Prescribed Multicultural Coexistence	Prescribed Multilingual Information Dissemination
General plan	15	8	5
Culture plan	6	4	3
Total	--	12	8

in cultural plans. It was estimated that local governments that had developed and published cultural plans had a high awareness of culture compared to local governments that had not developed or published them.

Table 4 shows the status of multilingual communication of information on websites operated by local governments and the status of multilingual communication of information of ordinances. Almost all local governments were conducting multilingual communication. On the other hand,

there were no local governments found that were conducting multilingual communication of ordinances.

PROPOSALS

From this survey, in local governments with a high composition ratio of MCRs, even though there were local governments that stated the necessity of multicultural co-existence and multilingual

Table 4. Current status on multilingual Web and multilingual ordinances on the Web of 15 local governments shown in Table 2 (Unit: City/Special ward)

	Having Multilingual Web	Details of Multilingual Web	Having Multilingual Ordinance on Web
With multilingual information dissemination prescription	8	By MT* 7 By Translator 3 Both 2	0
Without multilingual information dissemination prescription	7	By MT * 6 By Translator 3 Both 2	0
Total	15	--	0

*MT: Machine translation.

communication of information, there were no local governments that disseminated multilingual information about ordinances on the web.

For co-existence with MCRs, it is important to have a bi-directional compromise, that is, both an understanding of the culture of JRs and an understanding of the culture of the concerned region of MCRs. Ordinances are norms of daily life determined by the JRs and are simply the local culture. Measures that enable MCRs to understand local culture that is not easily visible in daily life such as ordinances are inadequate.

There are four points that can be considered factors for this situation.

- The style of the ordinances is strictly prescribed. As the ordinances are created using word processing documentation software such as Word, each local government uses a consistent style and the number of rows, characters, and indentation are implemented. When published on the web, these styles collapse.
- Ordinances prescribe rights and obligations and if accurate translation is not done, it can be misleading for the reader and there is a risk of hindering rights and obligations in some way.
- As for ordinances that are revised a number of times each year, translations by a translator (not machine) cannot be performed on time.

- As special applications are used for ordinances, it is possible to view not only the latest ordinance but also the history of past revisions. In ordinances with several revisions, this function of being able to quickly view the ordinance before the revision is very useful. In order to do machine translation without disturbing this function, for all ordinances, it is essential to recreate, in the web base, the current as well as the past ordinances prior to revision.

Hence, in this section we propose the following four points.

- Regardless of existing styles, it is important to disseminate ordinances focusing on content. So as not to interfere with the exercising of rights and obligations, it is more critical to make the content known than to adhere to the format.
- As machine translation is not suitable for ordinances, Japanese residents who are skilled in the specific language and MCRs can collaborate and come up with a proper human-powered translation. In this process, it is important to invite language experts such as translators and listen to their opinions as advisors.
- It is extremely difficult to conduct human-powered translation of all ordinances accurately and in a timely manner. Local gov-

ernments have established all ordinances systematically. In the system, basic ordinances are top priority and other ordinances are placed lower than other ordinances (Isozaki, 2012). Local governments should conduct multilingual information dissemination on a priority basis from basic ordinances that constitute the basis of the community. In recent years, 20% of local governments in Japan have prescribed the “fundamental ordinance on local autonomy” (Kurata, N, Kurata, Y, and Ohashi, 2016). These ordinances tend to be positioned at the highest level as the “constitution” of the concerned local government (Tsujiyama, 2002; Kisa, & Osaka, 2003). Therefore, it is expected that local governments that have prescribed the fundamental ordinance on local autonomy translate this ordinance on a top priority basis.

- It is mostly researchers, or government officials who create amendments for the ordinances that need the prior-to-revision version of the ordinance. It is fine to limit the display of revision history to the internal administration or to its use in the intranet. On the web, information dissemination for MCRs should be prioritized and only the latest ordinances that can be machine translated should be displayed in the web base. In order to exercise rights and obligations, there is no problem if latest ordinances are displayed as long as they are not litigation measures.

FUTURE RESEARCH DIRECTIONS

Among the abovementioned proposals, the process of creating translation in collaboration with MCRs can also be considered to contribute to the development of artificial intelligence.

In other words, if one considers the definition of artificial intelligence in detail, the following two perspectives exist. (Sugawara, 2007)

- Scientific viewpoint: Artificial intelligence aims to reveal the principles of intelligence and thought processes and explain them. Computer technologies are used as the means of manifestation of this theory. It is based on philosophical and psychological research methodology.
- Engineering viewpoint: Artificial intelligence refers to technologies that enable machines to operate things in ways similar to the use of intelligence by humans to operate things. For this, the human information processing functions are the model for computer-processable formats which are then put to use.

The process of translating in collaboration with MCRs, from the point of view of scientific definition, contributes to “revealing the process of understanding of an MCR who understand Japanese to a certain extent”. Also, the completed ordinances become an example and can be useful for example-based machine translation. If modeled, it can also be useful for statistical machine translation and can be considered as contributing to the engineering definition.

Further, as the native languages targeted in this survey are a few in number, the author would like to conduct further research in local governments with high composition ratio of MCRs.

CONCLUSION

Co-existence with MCRs is a challenge all over the world. In Japan where immigration policies have not progressed, multilingual communication of information is lacking. Specifically, since informa-

tion related to rights and obligations specified in ordinances differs depending on the local government, this information is not disseminated even though it is local culture that MCRs must know.

For the realization of a mature society that recognizes diversity, measures that facilitate the understanding of the local culture are first important. In order to protect the dignity of the MCR, it is essential to disseminate information to MCRs in an easy-to-understand language.

REFERENCES

- Agency for Cultural Affairs. (2001). *Field Study Related to the Resident Foreigners' Awareness of the Japanese Language In 2001*. Retrieved from http://www.bunka.go.jp/tokei_hakusho_shuppan/tokeichosa/nihongokyoiku_jittai/zaiju_gaikoku-jin.html
- Isozaki, H. (2012). Lecture on policy legal affairs in local government. Tokyo: Daiichihoki. (in Japanese)
- Japan Patent Office. (2015, May 11). *Preparation for machine translation dictionary of patent literature*. Retrieved from https://www.jpo.go.jp/shiryoutoushin/chousa/tokkyo_dictionary.htm
- Kisa, S., & Osaka, S. (2003). The constitution of our city –Challenges in Niseko town. Tokyo: Nihon Keizai Hyouronsha. (in Japanese)
- Kurata, N., Kurata, Y., & Ohashi, M. (2016). Semi-Quantitative Analysis of how the Preambles in Ordinances are designed: Observing the Change of People's Motivation towards "Inheritance" after the Great East Japan Earthquake. *International Journal of Learning. Teaching and Educational Research.*, 15(3), 29–41.
- Kurohashi, S. (2009). Example-based machine translation: EBMT. In the Association for Natural Language Processing (Ed.), *Encyclopedia of natural language processing* (pp.264-265). Tokyo: Kyoritsu Shuppan.
- Ministry of Internal Affairs and Communications. (1995). *1995 population census of Japan-The third basic tabulation*. Retrieved from <http://www.e-stat.go.jp/SG1/toukeidb/GH07010101Forward.do>
- Ministry of Internal Affairs and Communications. (2000). *2000 population census of Japan-Final report of the population*. Retrieved from http://www.e-stat.go.jp/SG1/estat/GL08020103.do?_toGL08020103_&tclassID=000000030587&cycle
- Ministry of Internal Affairs and Communications. (2005). *2005 population census of Japan-Final report of the population*. Author.
- Ministry of Internal Affairs and Communications. (2010). *2010 population census of Japan-Final report of the population*. Author.
- Ministry of Internal Affairs and Communications. (2015). *2015 population census of Japan-Results of preliminary sample*. Retrieved from http://www.e-stat.go.jp/SG1/estat/GL08020103.do?_toGL08020103_&tclassID=000001072818&cycleCode=0&requestSender=search
- Ministry of Justice. (2009, April). *Japanese law translation database system -Explanatory notes*. Retrieved from http://www.japaneselawtranslation.go.jp/index/about_this_site/?re=01
- Ministry of Justice. (2015). *2015 statistics on resident foreigners*. Retrieved from http://www.e-stat.go.jp/SG1/estat/GL08020103.do?_toGL08020103_&listID=000001150236&requestSender=dsearch
- Sugawara, K. (2007). *Artificial intelligence*. Tokyo: Morikita Publishing. (in Japanese)
- Sumita, E. (2009). Machine translation. In the Association for Natural Language Processing (Ed.), *Encyclopedia of natural language processing* (pp.262-263). Tokyo: Kyoritsu Shuppan.
- Tsujiyama, Y. (2002). Matsushita, K., Nishio, M., & Shindou, M. (Eds.). *Iwanami course: Conception of local government-4-structure*. Tokyo: Iwanami Shoten. Japanese.

Tsukamoto, H. (2009). Example-based machine translation: EBMT. In the Association for Natural Language Processing (Ed.), *Encyclopedia of natural language processing* (pp.266-269). Tokyo: Kyoritsu Shuppan.

ADDITIONAL READING

Mizukami, M., Neubig, G., Sakti, S., Toda, T., & Nakamura, S. (2014). Building a Free, General-Domain Paraphrase Database for Japanese. In *Proceeding of COCOSDA*. doi:10.1109/ICSDA.2014.7051433

Onishi, T., Utiyama, M., & Sumita, E. (2010). Paraphrase Lattice for Statistical Machine Translation. In *Proceeding of ACL*, pp. 1–5.

Sumita, E. (2014). Present state of machine translation with a focus on statistical approach. *Journal of Information Processing and Management*, 57(1), 12–21. doi:10.1241/johokanri.57.12

KEY TERMS AND DEFINITIONS

Artificial Intelligence: This refers to technologies in which intellectual activity such as understanding or reasoning of languages, and problem-solving skills conducted by humans are performed by computers instead. It is also called AI. Application fields that have incorporated artificial intelligence include expert systems that organize a particular area of human knowledge, store it as data, infer and judge using data, while understanding the meaning of inquiries. Further, artificial intelligence technologies are also utilized in pattern recognition that understands languages spoken by human beings and handwriting and in machine translation systems.

Fiscal Year: In Japan, the majority of companies, city halls, and schools operate according to the fiscal year. One fiscal year is from April to March next year.

Fundamental Ordinance on Local Autonomy:

The ordinances are regulations which each municipality enacts at its own discretion, the enactment of which is not obligated by the law of the national government. In the ordinances, the participation of citizens in community planning, along with provision of information relating to the community from the government to the citizens, are provided for, and provisions on administering the local area with citizens, local governments, area businesses, and organizations such as NPOs working together are spelled out. In many municipalities which have enacted them, the ordinances are positioned as supreme law which takes precedence over other regulations.

Local Government: This refers to municipalities and special wards established by the stipulations of the local government of a region. It refers to a local public organization whose boundary is of a smaller scale than that of a prefecture.

Machine Translation: This refers to the use of computers to translate into other languages. As translation of natural languages can contain several mistranslations due to omissions of words or mode of expression, it is used most often in the translation of technical documents such as manuals. Currently, it is sold as an application using the technologies of AI (artificial intelligence), and services are provided on the web pages.

Ordinance: Ordinances are government legislations applied in the prefecture or the local government, and are enacted within the range of the country's laws and regulations. A vote is required in the assembly of concerned local public organizations. In the prefecture or the local government, rules come under ordinances and guidelines come under rules. Among these, only ordinances require the decision of the council. As items related to the rights and obligations of the citizens are prescribed in ordinances, it is essential that they are enacted through a democratic process.

Population Census (National Population Census): This refers to the population survey conducted by the Statistics Bureau, Ministry of

Internal Affairs and Communications every five years on October 1 as of date. Questionnaires are mailed to each household and the members of the household answer providing the family composition. From 2015, it has become possible to provide answers via the internet as well. For

households that do not respond, survey officials appointed by the municipality personally visit the site and confirm the family composition among other things. According to the 2015 survey results, answers through the internet survey constituted 36.9% of the total number of households.

S

Vitalizing Ancient Cultures Mythological Storytelling in Metal Music

Uğur Kiliç

Ondokuz Mayıs University, Turkey

INTRODUCTION

The myth, which has been emerged at first because of the need of the humans “to understand”, then “to let the others to understand” by narrating, has a language structure full of symbols. This narrative characteristic of the myths leads this narrative genre to have a dynamic structure impeding it to be stuck in a certain period of time. Thereby, the myths have gotten a foothold existing in different narrative genres, and have even become a culture all by itself.

For instance, as the most known example, Classical-era Greek Mythology has been a part of the Western culture for centuries. It is known that Greek mythology has reached to this day nourishing various narrative genres. Aforementioned mythology is still in action in those genres thanks to its symbolic structure. Mythology enriches its own universe by interacting with different narrative genres. In fact, writers, artists and intellectuals have enhanced this enrichment including mythology in other fields such as literature, painting, sculpture, philosophy, theatre and music for centuries. The use of mythological motif in a work of art not only fosters the culture that it has emerged into, but also leads to a positive result for the recurrent mythology.

Today, the myths are used being included in different narrative genres. Thus, myths still continue to be active in cultural production. For instance, one can find traces of mythology in academic studies, literature, fine arts, theatre, computer games, TV series, movies and music. As it is seen, myths are positioned as a discipline

that protects its dynamics in expressional fields, e.g., literary, visual and audio.

Today, the other factor that ensures that myths being active in the cultural field is the myths of different geographies can be included in the production field thanks to the increasing number of various narrative mediums in modern day. While in the past the motifs of the dominant narratives, e.g., Greek and Roman Mythology, were mainly used in the works that were created, in today’s world almost every mythology that has been found and put together by studies may find itself a place in the cultural production. The narratives ranked among mythologies like Egyptian, Celtic, Norse and Hindu Mythology, used in fields like literature, films, painting and music is presented as an example for this statement.

The relation between music and mythology gets ahead of aforesaid narrative genres. The most important reason of this pioneering position of music is that music, just like mythology, is one of the ancient forms of narration. Music, whose origins go back to the sounds of nature, has a history of thousand of years. Cavidan Selanik (2010) defines music as “...with the aim of thinking with sounds, feeling and enhancing life through sounds, the art of researching and narrating the reality of man in all of its contexts”. The desire of man to narrate the facts and the events around him with the aim of feeling and enhancing life is similar to the starting point of mythology, inasmuch as myths also emerged thanks to the effort of man to interpret and narrate ‘the things’ happening around him.

Among the music genres of this day, number one genre that includes mythology and the myths in its area of usage is metal music which has begun to evolve around the UK and the USA in the late '60s and in the early '70s. Many bands in the history of metal music have used mythology in their lyrics, the video images, on the album covers, for their costumes and stage shows. Hence, mythology is renewed and enriched also in today's cultural production process. The relation between two of the most ancient narrative genres, mythology and music, which maintains on a contemporary level, is a subject that has been manifested to be studied about.

Based on the manifested research subject, it is intended that the subjects of metal music and mythology are of the top priority to be analyzed. In the chapter of praxis, in the aim of showing the connections of mythology, the lyrics, album covers, videos and concert performances of contemporary metal bands such as Amon Amarth, Eluveitie, Týr, Therion will be analyzed briefly. In addition to the praxis, Swedish metal band Amon Amarth's album "Twilight of the Thunder God" (2008) will be studied using semiotic method of analysis.

BACKGROUND

In this study, principally the conceptual framework will be specified by surveying literature reviews on 'metal music' and 'mythology'. In the praxis, semiotic analysis will be practiced upon. The reason of utilizing this method is that the band Amon Amarth enables us to read and study the mythological images on their album cover under the light of semiotic analysis, inasmuch as the reading method which semiotic analysis offers, assists us to reveal the mythological meanings in the visual product that the band released.

As it is known, the fundamental principles of semiotics depend on the studies of Swiss linguist Ferdinand de Saussure and American philosopher Charles Saunders Peirce. Saussure approached se-

miotics within the limits of linguistics. According to Saussure, a sign is composed of the signifier and the signified. The signifier is the image of sound that the message is ensured to be heard by the receptor. The signified is the meaning that occurred in the receptor's mind (Atabek & Atabek, 2007). The connection between the signifier and the signified is created by the rules and the codes that the society has determined. Dyer (2010) explains the arbitrariness between the signifier and the signified as: "The signifier isn't similar to the signified or the signifier doesn't cause the signified, but they are related by convention or 'contract' ".

On the other hand, Peirce made a connection between semiotics and logic, and saw every subject he had researched as a phenomenon of semiotics, and established the theory of signs in this context (Atabek ve Atabek, 2007). Peirce divided semiotic, which he argued that logic is formal semiotic, into speculative grammar, logic proper and rhetoric. In addition to that, Peirce divides the 'sign' itself into three. These are icon, index and symbol (Rifat, 2009). Icon is the sign that appears to have a similarity to the object (e. g., picture or photograph); index is the sign that is directly related to the signified (e. g., smoke, footprint); symbol is the sign that has a relation of arbitrariness with the object (e. g., flag, traffic lights)" (Atabek & Atabek, 2007). The arguments of Saussure and Peirce pioneered the development of semiotics. Thus, many semiotical applications that improved the conceptions of Saussure and Peirce have come to light since '20s. For instance, the semioticians in Europe have tried to contribute to semiotics based on the ideas of Saussure.

One of these semioticians, L. Hjelmslev, re-considered the bilateral of 'signifier/signified' and 'form/substance'. In this context, Hjelmslev named them as the expression plane, which includes the substance of sound, and content plane, which is the conceptual manifestation of the sign, and he also distinguished between form and substance. According to Mehmet Rifat (2009) in terms of the semiotic studies, one of the crucial conceptual

dualities of Hjelmslev is ‘denotation and connotation’. “Simply it means that any linguistic unit contains other and novel meanings (i. e., connotation) beyond its literal meaning (i.e., denotation).” Another semiotician who argued about the content plane is Roland Barthes (Fiske, 2003). “According to Barthes, who reorganized the distinction of Hjelmslev’s denotative semiotic, connotative semiotic and metalinguistic semiotic with a few small alterations, signs may refer to connotations that have other meanings apart from their denotations” (Atabek & Atabek, 2007).

To sum up, semiotics can be defined as ‘the key to any type of semantic coding system that allows communication to occur’. “Code is the semantic system that a culture or subculture members share” (Fiske, 2003). Thereby, semiotics sets a semantic structure manifesting the sign, the signifier and the signified in decoding.

Based on this possibility that semiotic studies allow, the praxis chapter of this study begins with the iconic explanation of the album cover. Iconic explanation is the part in which the images on the album cover are described and the locations of the images are pointed out. After the iconic explanation, signs are analyzed as signs at primary level, signs at secondary level and color code. In research of ‘signs at a primary level category’ the images on the album cover that draw the attention at first sight are determined. It is important to determine these images in terms of the explanation of the composition that has been put together on the album cover. In the category of signs at a secondary level, the rest of the images that is within the frame are explained. In the category of color code, the dominating color scale is analyzed and the meaning that is represented by this color is interpreted. Following this analysis, the visuals of the album are viewed in the terms of Denotation and Connotation, and they are analyzed in the context of ‘signifier of the denotation-signified of the denotation’ and ‘signifier of the connotation-signified of the connotation’.

On the other hand, in the praxis chapter it is aimed to support the analyzation with the ex-

amples of the album titles, the lyrics and stage performances of Amon Amarth because there are references to mythology in the song titles and lyrics stated in the albums of Amon Amarth. Furthermore, the band designs their stages in parallel to mythological narratives to support and recreate the content of their songs. Therefore, the analysis of these aforementioned elements is important in terms of revealing the connection between the band and mythology.

METAL MUSIC AND MYTHOLOGY: ROOTS AND CULTURAL CONNECTION

Since music is a cultural and artistic production, and it cannot be considered outside of the actions of the society that it has been emerged into, it is a product of thousands of years old adventure. Music has started to be formed by man having the sound in a thinking process and dominating it and has had various functions depending upon the societies’ transformations through historical process. Music, which is a product of thought, is related to the society and geography the person who engages with, and the mentality of the era it is produced in (Selanik, 2010).

By the 20th century, it has become harder to identify the characteristics of the history of music compared to previous periods. The main reason is that the developments in technology has enhanced various music genres in the same period of time. By the modern era, in consideration of technological resources with all the background it accumulated since the primitive age, music has continued to evolve zestfully, vigorously, and normatively at times. “Briefly, 20th century has opened up horizons for musical narratives, provided new instruments and in this context our era is the richest one in history of music” (Mimaroglu, 1999). Metal music, which this study has been bounded by, is a genre emerged under the light of these technological and socio-cultural changes, and in the pursuit of renewal.

Heavy metal or more widely known as metal music is a music genre which evolved in the late '60s and early '70s in the UK and the USA. In fact, the two mainstream genres are known as "New Wave of British Heavy Metal" and "New Wave of American Heavy Metal" surpassing all the subgenres. In this context, the majority of the bands that formed a basis for metal music and led the way in its early period is British and American (e. g., Black Sabbath, Led Zeppelin, Judas Priest, Iron Maiden, Alice Cooper). Aforementioned bands established the characteristics of metal music.

In technical terms, in the early years metal bands included a lead vocal, guitarist, bassist and drummer, later they have included another guitarist that played in sync with the ongoing guitarist creating a stronger and richer sound. In regard to the content, it is seen that various themes and subjects are used in metal music. For instance, in trash metal along with feelings like love, hate, fear and aggression, the variety can be seen in themes like fear of death, destruction, evil, mythological stories, creatures and themes belonging to horror stories (Charlton, 1990).

The technological means that constitutes the reason for the variety of musical movements in 20th century has had its share in also metal music, and these innovations that have been gone through since '70s has led the division of many subgenres, differentiation and disintegration in musical forms. The most well-known subgenres of metal are glam metal, thrash metal, death metal, black metal, folk metal and viking metal. Common trait of these subgenres is that they have created a universe and culture for metal music with the output they have generated. Any of this subgenre has benefited from the culture of the society it has emerged into, directly or indirectly. Folk metal and viking metal, which fulfill the aims of this study, are the proper examples on this topic.

Folk metal, emerged in Europe in '90s, is a genre of metal. Folk metal combines both traditional folk music and heavy metal and it includes

folk instruments in addition to heavy metal instruments. Themes of folk metal are generally nature, history, paganism, fantasy and mythology. In today's world, folk metal is popular especially in Europe. Finntroll, Ensiferum, Korpiklaani and Turisas can be seen as the prominent bands of folk metal. On the other hand, Swedish band Bathory that was formed in 1983 is the pioneer of viking metal. It can be said that viking metal utilized the elements of black metal and folk metal. This music genre adopted the pagan belief of Scandinavia, Norse mythology, the narratives of Viking Age, wars, epic tales and legends. Some of the viking metal bands, like typical folk metal bands, has been using Norse instruments (e.g., kantele) in addition to traditional metal music instruments. Although the lead vocal can be subject to change, brutal vocal is mainly used. Bands such as Wolfchant, Hel, Månegarm ve Týr are shown as the examples of today's viking metal. Considering the lyrics, album visuals, videos and stage performances of these two subgenres of metal music, mythological orientation cannot be overlooked. In order to reveal mythological traces of aforementioned genres, it is necessary to analyze mythology and its structure.

In history, human who is conscious and the most active being towards outside world has been trying to understand "the things" and find correlations between infinite number of phenomena. Knowledge that has been accumulated through this endless continuum, is a natural result of this mental and dynamic process that has been tried by the societies. It is possible to see this vast knowledge manifested in many fields as a production that meets the need of understanding, narrating and sharing of the humankind. Storytelling whose origins goes back to ancient times is the most efficient product of this need of narrating. According to Dell (2012), point of origin of mythology is based on the tendency of aforementioned "understanding and narrating of the origins of 'things'". The values of the societies, their views on concepts and phenomena they contemplated about were effective on the occurrence of the myths

of the societies. Yet, myths aren't just viewed as a part of storytelling. Instead, they are redefined from different perspectives.

It is likely to consider myths as “bo tree” that gathers their sacred values alongside with the lifestyle and wisdom of primeval communities, shaped by symbolic narratives. The myths that were shaped by the belief, lifestyle and wisdom that they were born in, although they seem like a narration medium that serves for the mental production of the individual, become important with the contributions to the culture thanks to the various elements they've been formed of. In other words, although the myths are man-made, they have the essence that characterise and guide the society in lots of ways because the myths are viewed as sacred, unlike the other narrative genres (e.g., tales, epic stories), by the society and therefore they become more efficient in people's lives.

Refusing the effects of the myths on the cultural production means refusing the society itself because when considering both modern and the archaic societies, one can find the traces of mythology in their language, daily lives, literature, art, sculpture and music. When discussing the connection of mythology with music, the first prominent study field is Greek mythology because of the etymology of “music”. Indeed, the origin of the word “music” goes back to nine goddesses known as the Muses or the Mousai in Greek mythology. On the other hand, not as many as in Greek mythology, but it is possible to track the traces of music in other mythologies. Eventually, there is a god or goddess that is related to music in all mythologies.

In today's world, the most obvious genre mythology is represented in music is metal music. Amon Amarth, Ensiferum, Turisas, Blind Guardian, Manowar, Hel, Heidevolk, Týr, Therion are some of the bands that use mythology in their music and distinguish themselves with this characteristic of them. Amon Amarth, who is analyzed in the praxis chapter, includes Norse mythology in every field of their music (e.g., lyrics, album images, stage shows and set). With this aspect of theirs,

the band has become important in cultural terms for producing music in aforementioned tendencies.

MYTHOLOGICAL STORYTELLING IN METAL MUSIC

As it was mentioned in the previous chapter, there is orientation toward mythology in certain subgenres of metal music. This orientation can be observed in lyrics, album titles and visuals, videos and stage shows. Amon Amarth, that is included in the praxis of this study, is a Swedish metal band from Tumba, Stockholm, founded in 1992. They released nine studio albums which are: *Once Sent From The Golden Hall* (1998), *The Avenger* (1999), *The Crusher* (2001), *Versus The World* (2002), *Fate Of Norns* (2004), *With Oden On Our Side* (2006), *Twilight Of The Thunder God* (2008), *Surtur Rising* (2012), *Deceiver Of The Gods* (2013). As it is clear in the album titles, the focus point of the band is Norse mythology. In fact, the title “Fate of Norns” represents The Norns in mythology which are female beings who rule the destiny of gods and men. The title “With Oden On Our Side” has a reference to god Odin. The title “Twilight of the Thunder God” connotes god Thor with the use of ‘thunder god’. The title “Surtur Rising” represents the the giant of fire, Surtr, in mythology. The title “Deceiver of Gods” connotes Loki who is characterised as a trickster god in mythology. As Imke Von Helden (2010) stated “Considering the content of lyrics, references to Norse mythology are seen to have given by viking metal bands.” The songs in Amon Amarth's albums are strong proof of this inference. Furthermore, when the band's albums are studied, it is seen that in all of their nine albums there are songs that include the motifs of Norse mythology. The references involve gods like Odin, Thor, Loki, Balder, Frej, Mimir, Hel, the creatures from different races like Surtr, Jörmungandr (*Midgard Serpent*), Fenrir, as well as toponyms like Yggdrasil, Midgard, Nifelheim, Muspelheim, Hel and narratives in mythology.

The symbols used in these narratives can be a transformed version of the mythological narrative or the exact version of certain parts of those narratives. As it is seen, the album titles have direct and clear references to Norse mythology. For instance, in the song “Twilight of Thunder God” Thor’s battle against the monstrous serpent Jörmungandr in Ragnarök has been depicted.

*Thor, Odin’s son, protector of mankind,
Ride to meet your fate, your destiny awaits.
Thor, Hlôdyn’s son, protector of mankind,
Ride to meet your fate, Ragnarök awaits!*

In the song “War of Gods”, as it is understood from its title, the war of gods in the Norse pantheon is described. This narrative, as it is studied in mythology before, is about Æsir-Vanir war, which are two groups of deities. In the lyrics below, it’s been told that after the Æsir-Vanir war, Mimir who has been sent to Vanir as a captive and his head cut off and sent to Odin.

*The Vanir failed to see Hænir was a fool
Without Mimir at his side Hænir could not rule
In a field of violent rage Mimir’s blood was shed
And to Odin’s court they sent Mimir’s severed head*

As it is seen, content of the lyrics of Amon Amarth as well as their song titles has references to various narratives and gods in Norse mythology. When one reads Norse mythology in depth, these references are identified one by one and one can see the correlations between the song and the mythology clearly. The other way Amon Amarth uses the visuals is stage designs. The band uses the decor and settings that are associated with mythical narratives in their concerts. The motifs associated with Norse mythology, the runes on columns in Old Norse language, the flags, and the hammer of Thor, Mjölnir are some of these decors. It is possible to watch the band’s decor use in 2013 in Summer Breeze Open Air heavy metal festival that takes place annually in Germany (<https://www.youtube.com/watch?v=IGeoJLIuow>). The band

used the decor of Jörmungandr in their stage show, in Wacken Open Air festival in 2014, which is an important festival for the metal music community that takes place annually again in Germany (<https://www.youtube.com/watch?v=mHCNfGQZpG8>).

The most apparent traces related to mythology in metal music are to be found in the album visuals of the band. Hence, one can see an example of this in the album visuals of “Twilight of the Thunder God (2008)” (<http://amonamarth.com/deceiver-of-the-gods/>) Considering the content of the album, the songs and the mythological references in those songs are: Twilight Of The Thunder God (*Jörmungandr, Thor, Odin, Hlôdyn, Ragnarök, Mjölnir*), Free Will Sacrifice, Guardians Of Asgard (*Asgard*), Where Is Your God?, Varyags Of Miklagard (*Norns*), Tattered Banners And Bloody Flags (*Loki, Æsir, Yggdrasil, Muspel, Midgard, Nifelheim, Fenris – Fenrir, Heimdall, Odin, The Norns, Valkyrie*), No Fear For The Setting Sun, The Hero, Live For The Kill, Embrace Of The Endless Ocean.

The album cover of Amon Amarth’s “Twilight of the Thunder God (2008)” contains a digital image. A dragon-like serpent is seen on the right side and the subsection of the frame. The mouth of the serpent is open threateningly and is faced to the man that holds its tongue. The man is standing on the curves of the serpent in the middle of the frame. The man has long and blonde hair and beard. The man has warrior clothing and the belt that hold it together on him. The man holds a short and shiny hammer, which has runes-like engravings on, in his right hand. He is holding the serpent’s tongue with his left hand. His posture gives the impression that he’s about to hit the serpent with the hammer in his hand. The ocean-like imagery takes place up to the middle of the frame and around the serpent and the man. The flames and thunders on the rough water can be seen. An effect like rain and thunder draws attention and stormy sky colors are dominant all across the frame. On upper left side of the frame, the name of the band (AMON AMARTH) is written with the color of flame and rune-like fonts. Under the

name of the band, the album title with smaller fonts (TWILIGHT OF THE THUNDER GOD) is seen.

- **Signs at Primary Level:** The serpent and the man with the hammer holding its tongue. The size of the serpent and the dragon-like appearance are the signs of a mythological creature. The man represents a warrior with his clothing and the hammer. In addition to that, the imagery of the man's battle against a mythological character and his engraved hammer in his hand gives the impression that he's a mythological sign/character as well.
- **Signs at Secondary Level:** The ocean, the flames and the thunder. The rough and stormy ocean is the sign of danger. The flames beyond the ocean are the images that support the notion of danger and threat. The lightnings are the signs of storm, thereby they altogether support the chaotic atmosphere created by the flames and the stormy weather.
- **Color Code:** The dominant colors in the image are the mixture of black and grey, tones of twilight and red. The colors that constitute twilight represent pessimism and death. The redness of the flames is the sign of terror and danger.

The dragon-like serpent and the man with a hammer holding the serpent's tongue appear in terms of denotation on the album cover of *Twilight Of The Thunder God* (2008). The man represents Thor in the Norse mythology and the serpent he attacks represents Midgard Serpent (*Jörmungandr*) existing in the same myth. The hammer in his hand and the belt indicate that he is Thor because the belt stands for Thor's power-belt and the hammer stands for his sacred weapon, Mjölnir. The dragon-like serpent that we see in various parts of the frame stands for Midgard Serpent that will make the world end, alias *Jörmungandr*, in Norse mythology. Thor and the flames, storm and the lightnings behind Midgard

Serpent give the impression of chaos. According to Norse mythology, Thor and Midgard Serpent battle against each other on judgement day, i.e., Ragnarök. Therefore, the images of the lightning, the flames and the storm in the frame represent the judgement day, i.e., Ragnarök. The lightnings are the signs of god of lightning, Thor, as well.

Looking at the image, tones of black and grey are dominant in general. This toning on the twilight is another supporting sign of Ragnarök in the frame. The writings on the upper left side are the information about the band's name and album title.

As it is seen, *Twilight of the Thunder God* (2008) is abundant in mythical symbols. It must be known that analyzing just one album cannot let us generalize the mythological effect on metal music. However, analyzing the other eight albums of this band, one cannot avoid discovering mythical symbols. Furthermore, Amon Amarth is not the only band that includes mythology in its music production. For instance, Týr, the metal band from Faroe Islands, is another band that produces music using mythology. One can see clearly the first step of using mythology is the band's name which is a god in Norse mythology. In fact, the albums of the band such as *How Far To Asgaard* (2002), *Ragnarok* (2006), *The Lay of Thrym* (2011) and *Valkyrja* (2013) incorporate mythical symbols directly in terms of both the names of the albums and the lyrics (<http://www.tyr.fo/discography>). And also, Swedish symphonic metal band Therion's album *Secret of the Runes* (2001) is directly related to Norse mythology in terms of album cover, the names of the songs and the lyrics (<http://therion.se/discography/secret-of-the-runes-2001/>). And another contemporary example of the use of mythology in metal music is Swedish folk metal band Eluveitie (<http://eluveitie.ch/discography/>). Eluveitie includes symbols of Celtic mythology in their albums they have released until today.

Consequently, analyzing the album names, the lyrics, album images and stage performances of the bands that have been included in the praxis of

this study, mythical symbols are included in music production by those bands. Indeed, the elaborate analysis of the band Amon Amarth supports this argument.

FUTURE RESEARCH DIRECTIONS

Technological developments in music increase the control on “sound”. Thus, metal music bands strengthen mythological stories in both lyrical and sound dimension. In today’s world metal music can be considered as an uncharted territory for communication, semiological, mythological and sociological studies. Many metal music bands produce their music adopting various symbols, languages and philosophical approaches. On that sense, every subgenre of metal music, in fact every metal music band, represents a whole universe of subculture. Semiological analysis in this study shows that mythological images can be examined in many form in metal music bands. Every interdisciplinary study in the academic field will act as a key for the secret door of metal music that has been waiting to be opened. In recent years the fact that mythology has been used mostly in metal music is another subject that must be studied in depth. Inasmuch as these tendencies of music bands, the attitudes of the audiences towards those tendencies and the desires of them must be analyzed. The academic studies in mythology and music will be useful for explaining today’s cultural changes.

CONCLUSION

Every society has its own myths and every myth has its own messages in its limited capacity. Those myths that narrate nature, creation, death, belief, war, chivalry, morals; to be brief, life, includes various experiences and wisdom that existed all around the world. For instance, Greek, Celtic, Turkic, Egyptian, Norse and lots of other mythologies have different classifications, functions and

also symbolic narratives. The myths, with their rich contents, have been a source of inspiration full of mental and spiritual elements for languages of today’s and the past societies, daily life, literature, art, sculpture and music. It is possible to observe the reflections of mythology in every field of literature, cinema, television and music. Music is another cultural production for mental and spiritual satisfaction. Thus it’s recognized as a very crucial point in this study.

This study has been limited to metal music which is a music genre that emerged in the late ‘60s and early ‘70s led by the United States and the United Kingdom with the technological, socio-cultural changes and innovative researches. As in lots of music genres, in time various subgenres began to appear in metal music in terms of content and technique. It is possible to say that certain subcultures have come forward as a result of adopting the subgenres of metal music in society. Because of their differences of content, style and technique, the subgenres of metal music such as glam metal, death metal, black metal, folk metal and viking metal are accepted as the subgenres that shape their own culture of music and listeners. Metal music which has been diversified with different equipments in terms of content and technique, has become important for this study owing to the inclusion of mythology.

As it’s been regarded in the praxis study, the birth of music and mythology go back to ancient times, and today they are two narrative systems that keep on building up, renewing and reviving. Myths are the first examples of storytelling and music, which continues to help its survival in its own possibilities, and they still share the same platform in the productions of metal music bands. Metal music bands play the part of today’s ‘mythological storytellers’ using mythology. Thus, mythology and music carry on playing an active role in cultural production. This deduction approaches mythology to metal music more and manifests the need for further study to determine the reflections of mythology on various subgenres of metal music.

REFERENCES

Amon Amarth Discography. (n.d.). Retrieved September 6, 2015 from <http://amonamarth.com/deceiver-of-the-gods/>

Amon Amarth Live At Summer Breeze. (n.d.). Retrieved September 6, 2015 from <https://www.youtube.com/watch?v=IGeoJLIuow>

Amon Amarth Live At Wacken Open Air. (n.d.). Retrieved September 6, 2015 from <https://www.youtube.com/watch?v=mHCNfGQZpG8>

Atabek, G. Ş. (2007). İletişim çalışmalarında göstergebilimsel yöntem. In Medya metinlerini çözümlemek (pp. 65-85). Ankara: Siyasal.

Charlton, K. (1990). *Rock music styles a history*. Dubuque, IA: Wm. C. Brown Publishers.

Dell, C. (2014). *Mitoloji hayali dünyalara eksiksiz rehber* (N. Elhüseyni, Trans.). İstanbul: Yapı Kredi. (Original work published 2012)

Dyer, G. (2010). *İletişim olarak reklamcılık* (N. Ö. Taşkiran, Trans.). İstanbul: Beta. (Original work published 1982)

Fiske, J. (2003). *İletişim çalışmalarına giriş* (S. İrvan, Trans.). Ankara: Bilim ve Sanat. (Original work published 1982). Retrieved September 6, 2015 from <http://eluveitie.ch/discography/>

Helden, I. V. (2010). Barbarians and literature: Viking metal and its links to old Norse mythology. In W. R. S. Niall & I. V. Helden (Eds.), *The metal void* (pp. 71–80). Oxfordshire, UK: Inter-Diciplinary Press.

Mimaroglu, İ. (1999). *Müzik tarihi*. İstanbul: Varlık.

Rifat, M. (2009). *Göstergebilimin ABC'si*. İstanbul: Say.

Selanik, C. (2010). *Müzik sanatının tarihsel serüveni*. İstanbul: Doruk. Therion Discography. Retrieved September 6, 2015 from <http://therion.se/discography/>

Týr Discography. (n.d.). Retrieved September 6, 2015 from <http://www.tyr.fo/discography>

KEY TERMS AND DEFINITIONS

Connotation: It represents the secondary, suppressive meaning of a sign, that is related to the personal knowledge of the one interprets the sign with.

Denotation: The first and direct meaning of a sign.

Jörmungandr: In Norse mythology, the serpent that lives in Midgard, the world inhabited by humans, and that Thor will battle against on judgement day.

Loki: The god that represents evil in Norse mythology.

Mjölknir: The magical hammer that belongs to god Thor in Norse mythology.

Myth: A narrative genre that explains the values of a society (its belief, philosophy, lifestyle, guidance, art etc.) it emerges into through supernatural events and people.

Norns, The: The female beings who rule the destiny are named Urd, Verdandi and Skuld in Norse mythology.

Odin: The main god that is “the father of all things” in Norse mythology.

Ragnarök: Judgement day in Norse mythology.

Thor: The most powerful god in Norse mythology.

Category S

Sports and Entertainment

Mining Sport Activities

Iztok Fister Jr.

University of Maribor, Slovenia

Iztok Fister

University of Maribor, Slovenia

INTRODUCTION

For many people today, sport is one of the stress relieving activities. People usually get involved in sport to achieve attractive shape, healthy lifestyle and better feeling. However, some people get involved in sport to compete professionally. Competitions were recently reserved for the professionals only. Few years ago, the professional community became more open to amateur athletes whose main objective is not competing. Nowadays, these can join to professional sports communities (Rauter, & Doupona Topič, 2011). Especially in massive sports events (running marathons, triathlons, cycling competitions), organizers give the opportunity to amateurs for competing with professionals in the same competition.

Amateurs need to train properly in order to fulfill their competition goals. Performing hard training is difficult not only for amateurs, but even for professionals. Athletes usually train at different stages and places to improve the effectiveness of training – gym, fitness, swimming pool, road, macadam, etc. Another obstacle that usually makes planning sports training sessions difficult for amateurs are properly balancing the two typical training phases, i.e., resting and eating that accompany each end of the training session. Consequently, athletes need to hire personal sports trainers in order to continuously improve their performance. Unfortunately, this hiring becomes very expensive after some time for amateur athletes.

Recent expansion of smart sports watches and smart phones allow athletes to carry out the sports training sessions smarter. Usage of mobile devices

(Baca, Dabnichki, Heller, & Kornfeind, 2009, Baca, Kornfeind, Preuschl, Bichler, Tampier, & Novatchkov, 2010) typically causes production of the dozens of activity datasets containing data about athlete's position, speed, heart rate and power rating over the sports training sessions. Analyzing these activity datasets enable an athlete to gain the deeper insight into every training session. Furthermore, a computer program using the latest data mining methods is even able to extract athlete's habits during sports training. Moreover, it may notify the athlete for possible over-training syndrome to prevent injury, advice nutrition using prediction and in the end, generate various sports training plans by clustering similar activities together.

Consequently, there is a lot of possibilities how to use these data for a further analysis and apply them for improving the athlete's performances. In a nutshell, these activity datasets offer an infrastructure for developing the artificial sports trainer (AST) as proposed by (Fister, Ljubič, Suganthan, Perc, & Fister 2015) that will be capable to help athletes during all sports training phases. Unfortunately, the most of athletes (especially amateurs) are not familiar with sports training theory and thus every assistance in any phase of sports training is welcome. As a matter of fact, the proposed AST computer framework based on Computational Intelligence (CI) algorithms that is open and therefore suitable for incorporating the new features easily.

The primary objective of the AST is to generate the sports training sessions of similar quality as a real trainer. At the moment, the AST serves as a

Figure 1. Illustration of training plan generation in the past and future



framework in which a various features covering different phases of sports training are incorporated especially in individual sports disciplines, where a lot of sports activities have been tracked using the wearable mobile devices. It definitely will present a bright star in planning, performing, controlling and evaluating phases of the sports training also in team sports disciplines in the future.

In this chapter, we focus on opportunities of data mining in sports. In line with this, the association rule mining method is proposed that was applied up to the present for solving two problems, i.e., discovering hidden habits of athletes in sports training and detecting the over-training syndrome. The results of both applications showed a true potential of these methods for data mining the sports activities. However, both applications can be incorporated in the AST framework as the features in the future.

Organization of this chapter is as follows. The next section is devoted to highlight background and fundamental information. In line with this, sports trackers and data mining paradigm are discussed. Then, the principles of data mining the sports activities are clarified. After that, the current applications in mining the sports activities are discussed shortly. The article concludes with a discussion and future research directions.

BACKGROUND

Sports Trackers

In order to track athlete's sports activities in the past, running athletes had worn very simple

watches (later renamed to simple sports watches). These watches allowed to monitor only time and were therefore used for analyzing the duration of workouts and interval trainings. One of the pioneers of sports watches, the finish manufacturer Polar, released first sports watch enabled measuring and monitoring the heart rate in 80's. Heart rate monitors allowed heart rate measurements in a real-time, thus allowing athletes to adjust performance in training/race. Heart rate was also stored into sports watch's memory suitable for the later analysis. Consequently, all measured data helped athletes to perform better training and competition by achieving much better results. At the beginning of the last century, the huge expansion of cycling sensors for measuring a speed and cadence in real-time emerged. The manifestation of those sensors improved also cyclist's performance. Today, modern sports bike is equipped with so called power meter. Power meter offers real-time power monitoring, and nowadays presents the state-of-the-art technology in cycling.

Another revolution was born with an emergence of mobile and pervasive computing. The mobile devices have been renamed to sports trackers when a massive production of smart phones and appropriate software applications has started. Sports tracker is basically an application running on the smart phone in order to track a sports training session (Fister, Fister, Fister, & Fong, 2013). These mobile devices obtain information from a global positioning system (GPS). There is a lot of sports watches on the market equipped with GPS that behave almost the same as the sports trackers (e.g. Garmin Forerunner, etc.). Consequently,

Code snippet 1. An example of GPX file

```
<trkpt lon="14.05636821873486" lat="46.03463463485241">
  <ele>714.4000244140625</ele>
  <time>2015-05-16T14:11:34.000Z</time>
  <extensions>
    <gpstpx:TrackPointExtension>
      <gpstpx:atemp>22.0</gpstpx:atemp>
      <gpstpx:hr>136</gpstpx:hr>
    </gpstpx:TrackPointExtension>
  </extensions>
</trkpt>
<trkpt lon="14.056426053866744" lat="46.03455643169582">
  <ele>714.7999877929688</ele>
  <time>2015-05-16T14:11:35.000Z</time>
  <extensions>
    <gpstpx:TrackPointExtension>
      <gpstpx:atemp>22.0</gpstpx:atemp>
      <gpstpx:hr>136</gpstpx:hr>
    </gpstpx:TrackPointExtension>
  </extensions>
</trkpt>
```

today's almost every type of sports training sessions can be monitored using the sports trackers (even swimming). Data about sports activities can be uploaded to the internet after finishing the training sessions. Powerful software applications allow the online analysis of these training data, sharing sports activities with friends in a "social network" and rearranging virtual competitions between these. Much more functions are accessible to users, e.g. weight monitoring, commenting on social networks, pace prediction and personal records storing. Data tracked during the workout is actually written into the XML datasets. There are two special formats of sports training datasets:

- GPX (GPS Exchange Format) (Code snippet 1),
- TCX (Training Center XML) (Code snippet 2).

The first code snippet presents an example of the GPX dataset, while the second an example of the TCX dataset. Both formats are easily for parsing, i.e., extraction of parameters. In fact, there are many tools which allow researchers to extract features from datasets. How to deal with these data is outlined in the paper (Rauter, Fister, & Fister, 2015).

Data Mining

Data mining (Han, Kamber, & Pei, 2011) is a popular discipline of computer science dealing with information extracted from large databases. This means, data mining methods help us to discover knowledge hidden in data. These methods have become more popular in the current era of big data. In line with this, various data mining applications were developed for application in

business, finance, medicine, engineering and even sport domain. At the moment, the most frequently used methods in data mining encompasses:

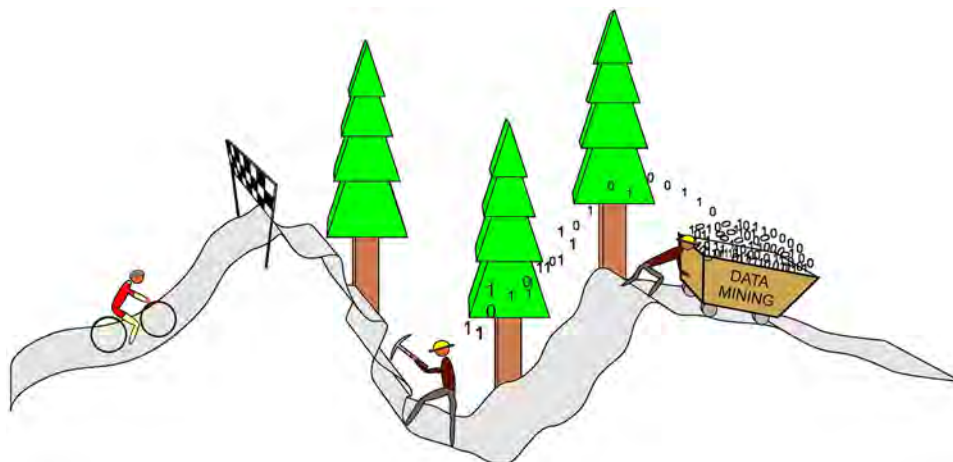
- Classification,
- Clustering,
- Association rule mining and
- Regression.

Classification is a process of searching a model for describing data classes or concepts. The model is built using an analysis of learning dataset, where the results of classification are known. This learned model can be used for prediction of test dataset, where the classification results are unknown. Clustering enables an integration of similar data into the so called clusters

Code snippet 2. An example of TCX file

```
<Activity Sport="Biking">
  <Id>2014-09-13T12:53:15.000Z</Id>
  <Lap StartTime="2014-09-13T12:53:15.000Z">
    <TotalTimeSeconds>8156.731</TotalTimeSeconds>
    <DistanceMeters>24078.68</DistanceMeters>
    <MaximumSpeed>13.526000022888184</MaximumSpeed>
    <Calories>683</Calories>
    <AverageHeartRateBpm>
      <Value>132</Value>
    </AverageHeartRateBpm>
    <MaximumHeartRateBpm>
      <Value>152</Value>
    </MaximumHeartRateBpm>
    <Intensity>Active</Intensity>
    <TriggerMethod>Manual</TriggerMethod>
    <Track>
      <Trackpoint>
        <Time>2014-09-13T12:53:15.000Z</Time>
        <Position>
          <LatitudeDegrees>46.042062006890774</LatitudeDegrees>
          <LongitudeDegrees>14.055101796984673</LongitudeDegrees>
        </Position>
        <AltitudeMeters>769.0</AltitudeMeters>
        <DistanceMeters>0.0</DistanceMeters>
        <HeartRateBpm>
          <Value>80</Value>
        </HeartRateBpm>
        <Extensions>
          <TPX xmlns="http://www.garmin.com/xmlschemas/ActivityExtension/
v2">
            <Speed>0.0</Speed>
          </TPX>
        </Extensions>
      </Trackpoint>
```

Figure 2. Mining sports activities



and does not use any output variables (Jain & Dubes, 1988). Association rule mining searches for relations between attributes that stay normally hidden in databases. Regression differs from the classification in the number of output variables, because there is only one output variable. The most popular classification methods are: support vector machines (SVM) (Vapnik, 1998), decision trees (DT) (Quinlan, 1986) and artificial neural networks (ANN) (Gershenson, 2013). Interestingly, a high impact of nature-inspired algorithms tailored for data mining tasks (Fong, 2013; Fister, 2015) is increased.

MINING THE SPORTS ACTIVITIES

Mining the sports activities (Figure 2) offers athletes to get a deeper insight into their own characteristics, like a body capacity, physical conditions, mental abilities, etc. Usually, casual sports trainers measure a training load of an athlete indirectly by using some indicators of sports training intensity, like heart rate, lactate, VO2 max etc. An estimation of the athlete's progress and increasing the performance over the years bases directly on tracking this information.

Typically, athletes do not realize any insight in relations between indicators of sports training intensity that are saved into dataset as attributes.

Knowing these relations (also associations) could have a great impact on an improvement of the athlete's performance. For example, the athlete performing the sports training session of medium duration, but high intensity will normally sleep badly that day. He will probably not regenerate till the next day. On the other hand, the same athlete will feel motivated and will have good eating habits after performing the long training session of low intensity.

Shall we find any special relations between attributes in large databases, and thus the athlete is guided to be motivated for sports training and consequently influences on his/her faster recovering? Of course. Using one of the data mining methods, named association rule mining, these hidden relations can be discovered. However, some of the association rules presents the positive relations that need to be promoted by the trainer, while another determining the negative relations should be prevented by trainer. For example, if an athlete has problems with eating and sleeping (consequently bad recovering) after the interval training session, he/she can still decrease the load or shorten the duration of the training session.

Mathematically, the association rules are expressed as an implication $X \rightarrow Y$, where X and Y are two subsets of items $I = \{i_1, \dots, i_n\}$, and $X \cap Y = \emptyset$. In addition, a set of transaction

named dataset $D = \{t_1, \dots, t_m\}$ is defined, where each transaction has unique identification and contains a subset of items. Typically, each mined rule is estimated according to the set of transactions by two measures, as follows:

- Support, which is defined as proportion of transactions containing X and Y , and the total number of transactions in database, i.e.:

$$\text{supp}(X \Rightarrow Y) = \frac{n(X \cup Y)}{N}$$

- Confidence, which is defined as proportion of transaction that contains X also contains Y , i.e.:

$$\text{conf}(X \Rightarrow Y) = \frac{n(X \cup Y)}{n(X)}.$$

Thus, the support must be higher than the prescribed minimum support $S_{\min}$ and the confidence higher than the prescribed minimum confidence $C_{\min}$.

Let us notice that there is a lot of association rule mining methods. Anyway, the procedure of the proposed mining sports activities consists of the following four steps:

- Data collection, preparation and preprocessing,
- Selecting the proper association rule mining method,
- Applying the selected data mining method to extracted data and
- Presenting the results to athletes or trainers using visualization.

In the first step, data needs to be prepared. Obviously, sports training datasets are collected by athletes over many years. All these tracked data

about past sports activities should be imported into databases. It is a good practice by athletes to add some additional information about present training session in a form of short questionnaire. Questionnaire describes some characteristics about athlete's feeling (e.g., eating and sleeping after training).

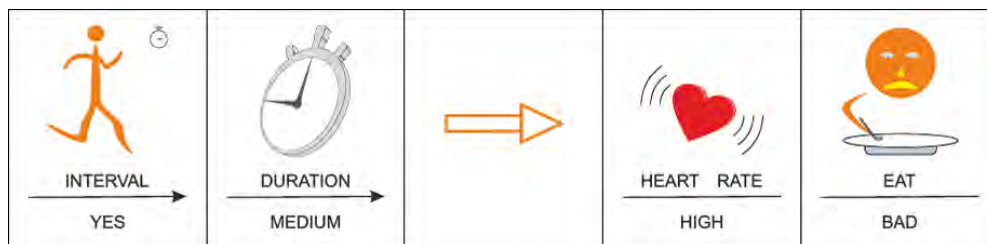
After the importing data about past training session, the proposed method for mining the sports activities is ready for preprocessing. The aim of the preprocessing is to extract attributes from data. These attributes are values extracted from the corresponding activity datasets (Rauter, Fister, & Fister, 2015) and questionnaires, as for example: total duration time, total distance, average speed, average and maximum heart rate and others. The structure of preprocessed features depends on the selected data mining method later. Extraction of these features are presented by various parsers that are described in paper (Rauter, Fister, & Fister, 2015).

In the second step, a decision needs to be received, where the expected results of the data mining should be determined. Consequently, a proper association rule mining method must be selected in the step. For example, if the athlete's behavior during or after training session is necessary, then the methods, like Apriori (Agrawal, & Srikant, 1994) or BatMiner (Fister, 2015), should be used. While the former is the traditional method, the later bases on the known swarm intelligence algorithm. According to the selected data mining method, the preprocessed data should be prepared.

The purpose of third step is to apply the selected data mining method to dataset, in other words, the selected data mining algorithm is being to run.

The final step is devoted to visualization of results to athletes or sports trainers. This step is very important due to successfully presentation of results obtained by the data mining method. Human are visual beings to whom the visual information say more than equivalent alphanumerical information. Therefore, the visualized, graphical information are more suitable for presentation of

Figure 3. An example of association rule



the results. One of the more popular possibility for graphical presentation are glyph-based diagrams (Ward, M. O. 2002).

CURRENT APPLICATIONS

Recent efforts in mining sports activities domain revealed two interesting applications for mining the sports activities:

- Finding hidden habits of athletes in training (Fister, 2015), and
- Detecting the over-training syndrome (Fister, Hrovat, Rauter, & Fister, 2014).

The purpose of the first application is to find some association rules in the sports training datasets obtained during athlete's training session. Usually, athletes are not aware about capacity of their body, influence of interval training on their sleeping, eating and other factors that influence on their performance. In order to help especially amateur athletes, the proposed mining the sports activities were developed. Results are presented athletes visually using glyphs (Figure 3, Figure

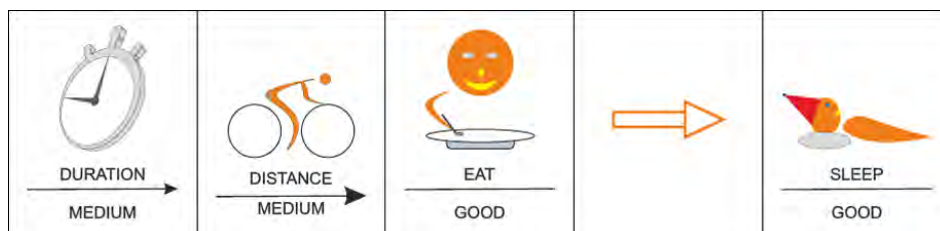
4). In the application, the association rule mining was implemented using the BatMiner.

The second application was devoted for detecting the over-training syndrome using association rule mining, where the Apriori was applied to real datasets created by mountain biker with mobile device. Experiments showed that this method is useful for detecting such kind of problems in sport. However, interpretation of results was still a bit weak in the application.

FUTURE RESEARCH DIRECTIONS

There is many opportunities and challenges for further research in this sports domain. Sports trackers are relatively young devices and therefore only a piece of their potentials were yet explored. On the other hand, athletes must be every year more and more fit in order to be prepared for competing on the higher level. Thus, many more parameters obtained from mobile devices during the sports sessions must be analyzed nowadays. Extracting athlete's advantages and weaknesses in dealing with various training types (intervals, long runs/rides) seems to be a promising way for further

Figure 4. An example of association rule



research besides other challenges that should be summarized as follows:

- **Finding the Proper Training Course:** The task of finding the proper training course is even more complex for sports trainers. Especially, setting up cycling course is very hard and demands many parameters by creation into account. These parameters are mostly connected with elevation, degrees of hills and even traffic conditions. In order to break this wall, data mining seems to be a promising method to find the best combination of training course.
- **Visualization:** Human beings prefer a graphic presentation against the alphanumeric presentation. In process of sports training and racing, athletes produce a lot of data that is normally presented in the alphanumeric form and usually means no special to athletes. However, a visualization of these strings enables also non-scientists a helpful view on results. One of the better solutions for visualization states on the glyph-based diagrams (Ward, 2002). Obviously, the visualization still offers many opportunities for research, e.g., visualization of segments of sports training, visualization of group of trainings, presentation of wall in performance and even virtual competitions among athletes.
- **Food Assistance:** What to eat at the proper moment is still a very peculiar question. Sports trainers and scientists had made many studies and simulations, and discovered that many factors influent on athlete's body and performance. Imagine the double ultra-triathlon (7.6 km swim, 360 km bike, 84.4km run), where an athlete's final success strongly depends on the proper food. Moreover, the proper eating is one of the most important factors in this kind of sports competition. In line with this, food prediction using data mining methods

might help athletes in choosing the proper food for long-term competitions.

- **Analysis of Rivals:** Nowadays, many data about various competitions in different disciplines are available on the Internet. Therefore, many athletes dream about analysis the advantages and weaknesses of their rivals or competitors not only in individual sports disciplines but also in team sports. On the basis of the result history, there is a possibility to achieve opponents' weaknesses using the data mining methods.

CONCLUSION

In this chapter, we presented a short overview of interesting research topic that proposes data mining in sports. Features for improving the performances of athletes, get special insight of their habits and body capacities are described. Steps of mining the sports activities are systematically showed, as well as challenges and opportunities for the future research. As a matter of fact, all these features could be integrated into the AST framework and therefore more and more specific phases of the sports training could gradually be realized by AST. Thus, the potential of the AST would be growth until abilities of the real sports trainer could finally be achieved.

REFERENCES

- Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules. In *Proc. 20th int. conf. very large data bases, VLDB* (Vol. 1215, pp. 487-499).
- Baca, A., Dabnichki, P., Heller, M., & Kornfeind, P. (2009). Ubiquitous computing in sports: A review and analysis. *Journal of Sports Sciences*, 27(12), 1335–1346. doi:10.1080/02640410903277427 PMID:19764000

- Baca, A., Kornfeind, P., Preuschl, E., Bichler, S., Tampier, M., & Novatchkov, H. (2010). A server-based mobile coaching system. *Sensors (Basel, Switzerland)*, 10(12), 10640–10662. doi:10.3390/s101210640 PMID:22163490
- Fister Jr, I. (2015). *Podatkovno rudarjenje športnih aktivnosti*. Poster session presented at IFAM 2015.
- Fister, I., Jr., Fister, D., Fister, I., & Fong, S. (2013). Data mining in sporting activities created by sports trackers. In *Computational and Business Intelligence (ISCBI), 2013 International Symposium on* (pp. 88–91). IEEE. doi:10.1109/ISCBI.2013.25
- Fister, I., Jr., Hrovat, G., Rauter, S., & Fister, I. (2014). Am I overtraining? A novel data mining approach for avoiding overtraining. In *StuCoSReC: proceedings of the 2014 1st Student Computer Science Research Conference* (pp. 47–52).
- Fister, I. Jr, Ljubič, K., Suganthan, P. N., Perc, M., & Fister, I. (2015). Computational intelligence in sports: Challenges and opportunities within a new research domain. *Applied Mathematics and Computation*, 262, 178–186. doi:10.1016/j.amc.2015.04.004
- Fister, I., Rauter, S., Yang, X. S., Ljubič, K., & Fister, I. Jr. (2015). Planning the sports training sessions with the bat algorithm. *Neurocomputing*, 149, 993–1002. doi:10.1016/j.neucom.2014.07.034
- Fong, S. (2013). Opportunities and challenges of integrating bioinspired optimization and data mining algorithms. *Swarm Intelligence and Bio-inspired Computation*, 385–401.
- Gershenson, C. (2003). *Artificial neural networks for beginners*. arXiv preprint cs/0308031
- Han, J., Kamber, M., & Pei, J. (2011). *Data mining: concepts and techniques*. Elsevier.
- Jain, A. K., & Dubes, R. C. (1988). *Algorithms for clustering data*. Prentice-Hall, Inc.
- Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1(1), 81–106. doi:10.1007/BF00116251
- Rauter, S., & Doupona Topič, M. (2011). Perspectives of the sport-oriented public in Slovenia on extreme sports. *Kineziologija*, 43(1), 82–90.
- Rauter, S., Fister, I., & Fister, I. Jr. (2015). *A collection of sport activity files for data analysis and data mining*. University of Ljubljana and University of Maribor.
- Rauter, S., Fister, I., & Fister, I. Jr. (2015). How to Deal with Sports Activity Datasets for Data Mining and Analysis: Some Tips and Future Challenges. *International Journal of Advanced Pervasive and Ubiquitous Computing*, 7(2), 27–37. doi:10.4018/IJAPUC.2015040103
- Vapnik, V. N. (1998). *Statistical learning theory* (Vol. 1). New York: Wiley.
- Ward, M. O. (2002). A taxonomy of glyph placement strategies for multidimensional data visualization. *Information Visualization*, 1(3–4), 194–210. doi:10.1057/palgrave.ivs.9500025

ADDITIONAL READING

- Baca, A. (2001). Computer science and biomechanics in sports-interdisciplinary aspects in research and education. *International Journal of Performance Analysis in Sport*, 1(1), 62–73.
- Baca, A. (2014). *Computer Science in Sport: Research and Practice*. Routledge.
- Begg, R., & Palaniswami, M. (Eds.). (2006). *Computational intelligence for movement sciences: neural networks and other emerging techniques*. IGI Global. doi:10.4018/978-1-59140-836-9
- Brzostowski, K., Drapała, J., Grzech, A., & Świątek, P. (2013). Adaptive decision support system for automatic physical effort plan generation—data-driven approach. *Cybernetics and Systems*, 44(2–3), 204–221. doi:10.1080/01969722.2013.762260

Corrado, L., Tessitore, A., Capranica, L., & Rauter, S. (2012). Motivation for a dual-career: Italian and Slovenian student-athletes/Motivacija Za Dvojno Kariero: Italijanski in Slovenski Studenti Sportniki. *Kinesiologia Slovenica*, 18(3), 47.

Fister, I., Fister, D., Fong, S., & Fister Jr, I. (2013). Widespread mobile devices in applications for real-time drafting detection in triathlons. *Journal of emerging technologies in web intelligence*, 5(3), 310-321.

Fister, I., Strnad, D., Yang, X. S., & Fister, I. Jr. (2015). Adaptation and hybridization in nature-inspired algorithms. In *Adaptation and Hybridization in Computational Intelligence* (pp. 3–50). Springer International Publishing.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1), 10-18.

Tan, P. N., Steinbach, M., & Kumar, V. (2006). *Introduction to data mining* (Vol. 1). Boston: Pearson Addison Wesley.

S

KEY TERMS AND DEFINITIONS

Artificial Sport Trainer: Software that mimics behavior of human sport trainer in order to create a plan of sport training sessions.

Bat Miner: Algorithm for association rule mining based on bat algorithm.

Sport Tracker: Software for tracking sport activity in real-time.

Sport Training: A process built on scientific and pedagogical principles, which affects the performance of an athlete using planned and systematic training sessions.

Sport Exergames for Physical Education

Pooya Soltani

University of Porto, Portugal

João Paulo Vilas-Boas

University of Porto, Portugal

INTRODUCTION

Insufficient physical activity is one the main parameters for mortality, and obesity is a growing concern in post-industrial countries. A combination of physical exercise and healthy nutrition is essential for decreasing obesity. Active video games (exergames) are becoming popular as ways of motivating people to exercise more. However, it is not clear whether these serious games could also be used in physical education (PE) and serve as more than mere entertainment. In this chapter, we will examine existing academic literature based on the characteristics of sport exergames that are important in the domain of PE. We also provide a practical example of psycho-biophysical evaluation of a sport exergame to see how close and encouraging these games are, compared to the real sports.

BACKGROUND

PE is considered to be a crucial part of primary school curriculum around the world which include both physical and educational contents (Lindberg, Seo, & Laine, 2016). Thanks to digital technologies, new learning environments provide opportunities for skill acquisition and socializing, and researchers are now examining the role, efficacy, and opportunities that these environments provide. On the other hand, several reasons including lack of time, lack of skills of PE instructors, and lack of support might reduce the quality and quantity of PE education (Lindberg, Seo, & Laine, 2016).

Moreover, pedagogy has always tried to innovate teaching, and PE is also emerging regarding integrating technology into regular classes. One exciting area in which technology and education could merge is by using video games that include visual (and/or audio) stimulus. Video games can be applied to improve attention, executive functions, and reasoning (Neugnot-Ceroli, Gagner, & Beauchamp, 2015). They are also shown to increase several types of intelligence (e.g. visual-spatial and bodily-kinesthetic) while providing a playful-formative experience (del Moral-Pérez, Fernández-García, & Guzmán-Duque, 2015).

On the other hand, previous psychological research has linked aggression with video gaming (Anderson et al., 2010) and content analysis of video games was mainly concerned about violence and role of gender (Lee & Peng, 2006). As excessive use of technology, which also includes video gaming, is suggested to be a contributing factor in obesity, a new approach has been proposed to include active video games (exergames) that incorporate motion sensor technology and could be played using whole body movements. While gamification of regular exercise activities (e.g. GPS-based virtual reality zombie run) has been previously reported, active video games are not frequently used in the context of PE (Lindberg, Seo, & Laine, 2016), and insufficient evidence about efficacy of exergames exist within schools (Norris, Hamer, & Stamatakis, 2016). Previously, Ennis (2013) considered exergames in three categories of recreation (light to moderate intensity), public health (moderate to intensive physical activity - PA), and educational (to facilitate skills).

DOI: 10.4018/978-1-5225-2255-3.ch640

MAIN FOCUS OF THE ARTICLE

Motivation/Literacy to Play and Exercise

Several methods are used to increase players' motivation in sporting activities (Keegan, Harwood, Spray, & Lavalley, 2009) and game-based learning and storytelling are the primary ways to provide intrinsic motivation for learning (Laine, Nygren, Dirin, & Suk, 2016). Exergames provide a non-scary environment to develop components of mastering of fundamental movement skills (George, Rohr, & Byrne, 2016). For example, children might improve aiming and catching skills during virtual tennis without the fear of getting hit by the physical ball. In children with sensory dysfunction, exergames are used to increase their learning motivation and to make them more confident in facing various learning challenges (Chuang & Kuo, 2016). By embedding elements of nature, exergames can also provide a sense of connectedness and environmental concern, which might be important for exercising outdoor (Öhman, Öhman, & Sandell, 2016). Wittland, Brauner, and Ziefle (2015) also suggested that accepting serious games for physical fitness, is not dependent on gender, expertise, and gaming habits. However, when used with older adults, "guided hands-on" and "1-on-1" teaching methods might be used to increase their engagement when facing technology (Seides & Mitzner, 2015), because older adults need more time to master the skills necessary to play active video games (Santamaría-Guzmán, Salicetti-Fonseca, & Moncada-Jiménez, 2015). Sun (2013) evaluated the effects of exergame in primary school students and showed that PA situational interest decreases over time, but exergame intensity increases. Therefore, strategies to balance the activities should be considered during the game design phase. Many models of have been created to explore game characteristics from designers and consumers' perspectives (cf. Mildner, Stamer, & Effelsberg, 2015). Gender, age, game type, players' characteristics and personalities are motivators

for gameplay (cf. Jabbar & Felicia, 2015). For example, Shaw, Tourrel, Wunsche, & Lutteroth (2016) showed that considering personality and motivation of players in a virtual training exergame with two modes of competitive and cooperative gameplay might increase exercise, especially in competitive individuals.

Learning and Skill Acquisition

Digital games have become great tools in knowledge transfer due to fostering intrinsic motivation in players to acquire more knowledge (Mildner, Stamer, & Effelsberg, 2015). Some sports games may be used to simulate the real sports skills; for example, shooting exergames might have a positive skill transfer for increasing hitting scores (Eliöz, Vedat, Küçük, & Karakaş, 2016), and Vernadakis, Papastergiou, Zetou, & Antoniou (2015) showed that exergame-based interventions could improve object control skills in children. Moreover, players' interactions with the game and other players, affect their learning while sports exergaming (Meckbach, Gibbs, Almqvist, & Quennerstedt, 2014). Exergames might also increase PA while developing motor skills among overweight children and adolescents (do Carmo, Goncalves, Batalau, & Palmeira, 2013). Body tracking technologies have also been used as a live correcting tool for free weight exercises (Conner & Poor, 2016), and improved motor skills (balance) after exergaming was observed with higher scores in female players (Norris et al., 2016).

On the other hand, these games may not be as effective as traditional PE instructions for psychomotor development (Pedersen, Cooley, & Cruickshank, 2016). For example, virtual swimming in the air does not replicate the physical fidelity connected with moving water. A previous systematic review also showed that virtual reality applications have the ability to change behavior but have little gain in knowledge (Omaki et al., 2016). While cognitive functions are crucial for the functional autonomy, Monteiro-Junior et al. (2016) showed that a single bout of virtual real-

ity training does not have any effect on cognitive function in older adults. Johnson, Ridgers, Hul-teen, Mellecker, & Barnett (2015) also showed that short bouts of exergame playing may not influence perceived and actual ball control skill competence. Previous research also argues that when mastery is achieved, a gradual decrease in performance is expected (Adi-Japha, Karni, Parnes, Loewenschuss, & Vakil, 2016). Therefore, player balancing techniques should be used to provide an equilibrium between enjoyment and skill acquisition for players with different levels of expertise (Vicencio-Moreira, Mandryk, & Gutwin, 2015).

Competition vs. Cooperation

To gamify traditional sports, researchers and game designer are often challenged to provide a socially-enriching experience without bombarding players with game information (Choi, Oh, Edge, Kim, & Lee, 2016). Characteristics of different players should also be taken into account; for example, Staiano et al. (2012) concluded that in a group of adolescents, competitive players outperform cooperative participants on executive function test. Girls also play exergames mostly for social interaction and rather than competition, and boys play games for intrinsic reasons (O'Loughlin et al., 2012). Online cooperative exergaming can also increase social relatedness in players compared to the time when they play against a computer character (Kooiman & Sheehan, 2015).

Opinion

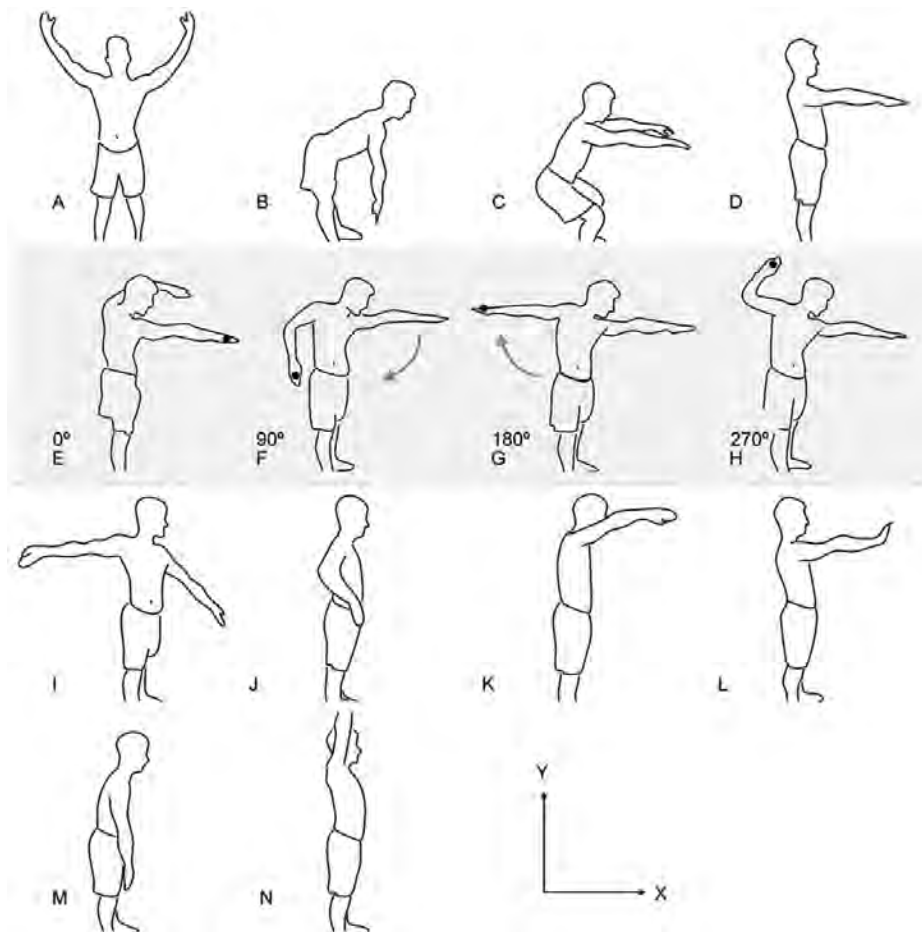
Previous research suggests that if PE instructors decide to use exergames in their teaching, they should justify why and how they want to apply them, and design strategies for the students to interact with exergames (Gibbs, Quennerstedt, & Larsson, 2016). For example, a meta-analytic review suggests that playing exergames can improve older adults' physical balance, balance confidence, functional mobility, executive function,

and processing speed (Zhang & Kaufman, 2015), and coaches may use these ideas to design more effective exercise programs. Exergame inclusion in PE curriculum may also provide additional tools to meet PA standards and to motivate and engage the student in practice (Rudella & Butz, 2015). On the other hand, Hulteen, Johnson, Ridgers, Mellecker, & Barnett (2015) showed that while players had correct skill performance (during table tennis, tennis, and basketball), they also had proper movement during skill assessment, which shows that presence of evaluator may affect the way participants play. According to Kooiman, Sheehan, Wesolek, & Reategui (2016), exergames should not be seen as the primary tool in PE because there are several parameters that expand the activity beyond virtual environments. Using technology in education is mostly affected by teachers' attitudes towards using technology (Albirini, 2006). While Bransford et al. (1999) talked about the four different learning environment as learner-centered, knowledge-centered, assessment centered, and community centered, it is important to decide which area benefit learning new skills in exergames more.

Characterization

In this part of the book chapter, the authors mention studies that were included in the Ph.D. thesis of the first author about characterizing a sport exergame. In the first part, the authors compared movement patterns of forty-six college students, with different performance (novice vs. experienced, and real-swimmers vs. non-swimmers) and gender. Reflective markers were placed over the skin, and the 3D position of each marker was recorded using a motion capture system. Subject played different techniques (100 m each) of a swimming exergame designed for Microsoft Xbox and Kinect (Michael Phelps: Push the Limit, 505 Games, Milan, Italy). Subjects had to stand in front of Kinect and start their gameplays by "hying" for the virtual crowd (Figure 1, Panel A). Then they had to slightly bend forward (mimicking start

Figure 1. Body position during each phase of the game



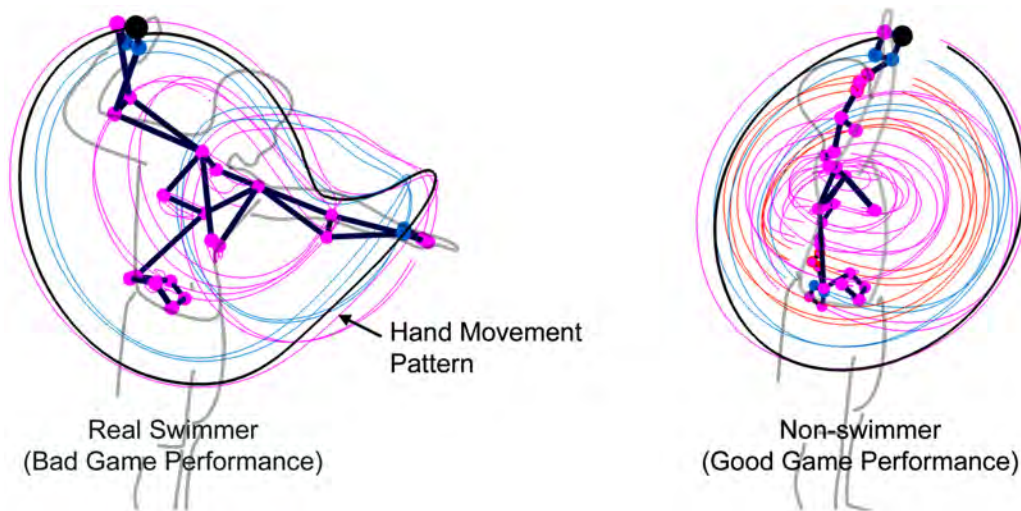
for the crawl, breaststroke, and butterfly; Figure 1, Panel B). For the backstroke technique, the start is shown in Figure 1, Panel C. Following the “Go!” command, they had to extend their back (Figure 1, Panel D) and swim according to the techniques (Figure 1, Panel E to H for crawl; Panel I for backstroke; Panel J for breaststroke, and panel K for butterfly). To start a new lap, they had to extend their arm forward (Figure 1, Panel L) and continue swimming. For terminating the race, they had to drop both arms (Figure 1, Panel M) and raise one arm immediately (Figure 1, Panel N).

The gameplay was divided into two phases of normal (with on-screen visual feedback to prevent players from playing too fast or too slow) and fast (playing without any feedback). Players’ performances were ranked from 1st to 8th. The

results showed that players who ranked better in the game (Figure 2), completed the game faster and with fewer arm cycles. By visually inspecting Figure 2, we can see that movement patterns of good performers were different that bad performers, who were mostly real-swimmers, meaning that the game does not encourage players to swim correctly. Moreover, those who had prior experience with the game had fewer arm cycles and real-swimming experience and gender did not affect biomechanical parameters.

In the second study, the authors measured and compared aerobic and anaerobic energy systems contributions and the activity profiles of participants while sport exergaming. Players were equipped with a gas analyzer to measure oxygen consumption, and blood lactate was measured us-

Figure 2. Movement patterns during swimming exergame



ing a small amount of blood from their ear lobe after each technique and following the gameplay. From these two parameters, we were able to measure the total energy expenditure (EE) during the gameplay. Players' gameplays were also filmed to measure total playing time (TPT), effective playing time (EPT), resting time (RT) and effort to rest ratio (E:R) using video analysis. Our results showed that anaerobic lactic pathway accounted for around 9% of total EE and EE was not different between performing groups. This shows that although the level of EE is lower compared to real-swimming, both energy systems should be considered when analyzing sports exergames. Heart rate (HR) was also measured during the gameplay and differences were observed between real-swimmers and non-swimmers in the first technique. This confirms that real-swimmers tend to swim correctly and exert more at the beginning of the game, but as soon as they realized the mechanisms of the game, they tend to exert less. Players were active around 57% of total time, and E:R was approximately 1.3. Our results also show that although players dedicated more time playing than resting, the changes were not different between players. Experienced players had lower TPT, EPT, and E:R compared to their novice counterparts, suggesting that experienced players learn the strategies to play the game with

lower exertion. This confirms the necessity of designing games with lower resting times (loading of the game, navigating between the menus, etc.), dedicating more time to the actual gameplay.

In the third study, the authors measured muscle activation during sport gaming. Surface electromyography electrodes were placed over Biceps Brachii, Triceps Brachii, Latissimus Dorsi, Upper Trapezius, and Erector Spinae muscles. These muscles were chosen as they are frequently activated during swimming or were relevant because of the game itself. Maximum voluntary isometric contraction (MVIC) of each of the muscle was also measured to normalize the activation in percentages. Our results showed that muscle activation ranged from 5 to 95% MVIC, and differed between normal and fast swimming for all techniques. Muscular coordination was also investigated using biomechanics and observed that when participant plays faster, they complete different stages of the game faster, which results in higher stress some of the muscles more. Measuring muscle activation is important because it can show if players with greater experience, exert and engage less or not. While challenges in sedentary games are usually controlled by adjusting the complexity of mental tasks, employing in-game physical challenges can be unique characteristics of exergame design

and muscle activation evaluation can provide guidelines to make sports exergames closer to real activities. Such evaluations can also address the Concerns regarding long-time computer/video game use, repetitive movements, and musculoskeletal disorders, such as neck and shoulder pain. Moreover, although sports exergames may not produce as much muscle activation as the real activity, such activities might still benefit participants to develop muscular endurance, especially when participation in real sport is not possible or practical due to disability, fear, or injury.

In the last study, the authors evaluated the game experience, usability, and enjoyment during sport exergaming to see if they affect future participation if PA and/or real sport. Enjoyment can both predict and be an outcome of PA participation and is one of the reasons why people play video games. System usability scale (SUS) is a measurement of learning, control, and understanding a game, and higher SUS grade means easier interaction between human, different games and menus, and various controllers and gaming platforms. User experience (UX) deals with consumers' dynamic perceptions and responses of products and systems and is an important factor for game design and positioning of the problems. For measuring enjoyment, a short version of physical activity enjoyment scale (PACES) questionnaire with two additional questions was used to measure future intentions of participation in physical activity and real swimming. Game usability was measured by System Usability Scale and playability aspects using Game Experience Questionnaire. Our results showed that Female players with real swimming and exergame experience enjoyed the game more. Usability score was around 75 which is considered as good with high acceptability. PA intentions did not change within performing groups, but swimming intentions were increased for all players. A possible explanation might be that future PA intentions of those who frequently exercise, may

not be affected by playing exergames. Another explanation might be that those who do not exercise regularly, may think that the health benefits attained through exergaming are enough, and there is no need for further PA participation. It might be possible that novelty and entertainment elements of swimming exergame, played a role in influencing players' attitudes towards real swimming intention.

FUTURE RESEARCH DIRECTIONS

In our research, most of the players had sports science background and were physically active. This might have influenced the way they were interacting with the game. The presence of the researchers themselves might have affected players' performance to play differently. It might be possible that novelty and entertainment elements of swimming exergame, played a role in influencing players' attitudes towards real swimming intention and therefore, longitudinal studies should be conducted to check the changes in intentions in the long run.

CONCLUSION

In general, our data suggest that better performance inside the game does not necessarily mean better performance in real swimming. The motion capture sensor is also not capable of capturing delicate movements of swimming, and it does not encourage players to swim correctly. Therefore, the game may not be used as a training tool for real swimming and may only be used as a motivational tool for teaching basic movements of swimming (e.g. coordination of the upper limbs). While we do not deny the possible role of video games in PE settings, we believe that there is still a long way to use sport exergames in the PE.

REFERENCES

- Adi-Japha, E., Karni, A., Parnes, A., Loewenschuss, I., & Vakil, E. (2008). A shift in task routines during the learning of a motor skill: Group-averaged data may mask critical phases in the individuals acquisition of skilled performance. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 34(6), 1544–1551. doi:10.1037/a0013217 PMID:18980413
- Albirini, A. (2006). Teachers attitudes toward information and communication technologies: The case of Syrian EFL teachers. *Computers & Education*, 47(4), 373–398. doi:10.1016/j.compedu.2004.10.013
- Anderson, C. A., Ihori, N., Bushman, B. J., Rothsetin, H. R., Shibuya, A., & Swing, E. L. et al. (2010). Video Game effects on aggression, empathy, and prosocial behavior and eastern and western countries: A meta-analytic review. *Psychological Bulletin*, 136(2), 151–173. doi:10.1037/a0018251 PMID:20192553
- Bransford, J. D., Brown, A., & Cocking, R. (2000). How people learn: Mind, brain, experience and school, expanded edition. Washington, DC: National Academy Press. Retrieved from www.nap.edu/openbook.php?isbn=0309070368
- Choi, W., Oh, J., Edge, D., Kim, J., & Lee, U. (2016, May). SwimTrain: exploring exergame design for group fitness swimming. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 1692-1704). ACM. doi:10.1145/2858036.2858579
- Chuang, T. Y., & Kuo, M. S. (2016). A motion-sensing game-based therapy to foster the learning of children with sensory integration dysfunction. *Journal of Educational Technology & Society*, 19(1), 4–16.
- Conner, C., & Poor, G. M. (2016, May). Correcting exercise form using body tracking. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 3028-3034). ACM. doi:10.1145/2851581.2892519
- del Moral-Pérez, M. E., Fernández-García, L. C., & Guzmán-Duque, A. P. (2015). Videogames: Multisensory incentives boosting multiple intelligences in primary education. *Electronic Journal of Research in Educational Psychology*, 13(2), 243–270. doi:10.14204/ejrep.36.14091
- do Carmo, J., Goncalves, R., Batalau, R., & Palmeira, A. (2013). *Active video games and physical activity in overweight children and adolescents*. Paper presented at the 2013 IEEE 2nd International Conference on Serious Games and Applications for Health, Vilamoura, Portugal. doi:10.1109/SeGAH.2013.6665323
- Eliöz, M., Vedat, E., Küçük, H., & Karakaş, F. (2016). The effect of motion detecting computer games on the skills training. *Beden Eğitimi ve Spor Bilimleri Dergisi*, 10(1), 13–18.
- Ennis, C. D. (2013). Implications of exergaming for the physical education curriculum in the 21st century. *Journal of Sport and Health Science*, 2(3), 152–157. doi:10.1016/j.jshs.2013.02.004
- George, A. M., Rohr, L. E., & Byrne, J. (2016). Impact of Nintendo Wii games on physical literacy in children: Motor skills, physical fitness, activity behaviors, and knowledge. *Sports*, 4(1), 3. doi:10.3390/sports4010003
- Gibbs, B., Quennerstedt, M., & Larsson, H. (2016). Teaching dance in physical education using exergames. *European Physical Education Review*. doi:10.1177/1356336X16645611

- Hulteen, R. M., Johnson, T. M., Ridgers, N. D., Mellecker, R. R., & Barnett, L. M. (2015). Childrens movement skills when playing active video games. *Perceptual and Motor Skills*, 121(3), 767–790. doi:10.2466/25.10.PMS.121c24x5 PMID:26654991
- Jabbar, A. I. A., & Felicia, P. (2015). Gameplay engagement and learning in game-based learning a systematic review. *Review of Educational Research*, 85(4), 740–779. doi:10.3102/0034654315577210
- Johnson, T. M., Ridgers, N. D., Hulteen, R. M., Mellecker, R. R., & Barnett, L. M. (2016). Does playing a sports active video game improve young childrens ball skill competence? *Journal of Science and Medicine in Sport*, 19(5), 432–436. doi:10.1016/j.jsams.2015.05.002 PMID:26050626
- Keegan, R. J., Harwood, C. G., Spray, C. M., & Lavalley, D. E. (2009). A qualitative investigation exploring the motivational climate in early career sports participants: Coach, parent and peer influences on sport motivation. *Psychology of Sport and Exercise*, 10(3), 361–372. doi:10.1016/j.psychsport.2008.12.003
- Kooiman, B. J., & Sheehan, D. P. (2015). The efficacy of exergames for social relatedness in online physical education. *Cogent Education*, 2(1), 1045808. doi:10.1080/2331186X.2015.1045808
- Kooiman, B. J., Sheehan, D. P., Wesolek, M., & Reategui, E. (2016). Exergaming for Physical Activity in Online Physical Education. *International Journal of Distance Education Technologies*, 14(2), 1–16. doi:10.4018/IJDET.2016040101
- Laine, T. H., Nygren, E., Dirin, A., & Suk, H.-J. (2016). Science Spots AR: A platform for science learning games with augmented reality. *Educational Technology Research and Development*, 64(3), 507–531. doi:10.1007/s11423-015-9419-0
- Lee, K. M., & Peng, W. (2006). What Do We Know About Social and Psychological Effects of Computer Games? A Comprehensive Review of the Current Literature. In P. V. J. Bryant (Ed.), *Playing video games: Motives, responses, and consequences* (pp. 327–345). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Lindberg, R., Seo, J., & Laine, T. H. (2016). Enhancing physical education with exergames and wearable technology. *IEEE Transactions on Learning Technologies*. doi:10.1109/TLT.2016.2556671
- Mildner, P., Stamer, N., & Effelsberg, W. (2015). *From Game Characteristics to Effective Learning Games Serious Games*. Springer.
- Monteiro-Junior, R. S., da Silva Figueiredo, L. F., de Tarso Maciel-Pinheiro, P., Abud, E. L. R., Braga, A. E. M. M., Barca, M. L., & Laks, J. et al. (2016). Acute effects of exergames on cognitive function of institutionalized older persons: A single-blinded, randomized and controlled pilot study. *Aging Clinical and Experimental Research*. doi:10.1007/s40520-016-0595-5 PMID:27256080
- Neugnot-Cerlioli, M., Gagner, C., & Beauchamp, M. H. (2015). The use of games in paediatric cognitive intervention: A systematic review. *International Journal of Physical Medicine & Rehabilitation*, 3(4). doi:10.4172/2329-9096.1000286
- Norris, E., Hamer, M., & Stamatakis, E. (2016). Active video games in schools and effects on physical activity and health: A systematic review. *The Journal of Pediatrics*, 172, 40–46. doi:10.1016/j.jpeds.2016.02.001 PMID:26947570
- Öhman, J., Öhman, M., & Sandell, K. (2016). Outdoor recreation in exergames: A new step in the detachment from nature? *Journal of Adventure Education and Outdoor Learning*. doi:10.1080/14729679.2016.1147965

- OLoughlin, E. K., Dugas, E. N., Sabiston, C. M., & OLoughlin, J. L. (2012). Prevalence and correlates of exergaming in youth. *Pediatrics*, 130(5), 806–814. doi:10.1542/peds.2012-0391 PMID:23027171
- Omaki, E., Rizzutti, N., Shields, W., Zhu, J., McDonald, E., Stevens, M. W., & Gielen, A. (2016). A systematic review of technology-based interventions for unintentional injury prevention education and behaviour change. *Injury Prevention*, injuryprev-2015-041740. doi:10.1136/injuryprev-2015-041740 PMID:26787740
- Pedersen, S. J., Cooley, P. D., & Cruickshank, V. J. (2016). Caution regarding exergames: A skill acquisition perspective. *Physical Education and Sport Pedagogy*, 1–11. doi:10.1080/17408989.2016.1176131
- Rudella, J. L., & Butz, J. V. (2015). Exergames: Increasing physical activity through effective instruction. *Journal of Physical Education, Recreation & Dance*, 86(6), 8–15. doi:10.1080/07303084.2015.1022672
- Santamaría-Guzmán, K., Salicetti-Fonseca, A., & Moncada-Jiménez, J. (2015). Learning curve and motor retention of a video game in young and older adults. *Retos*, 27, 218–221.
- Seides, J., & Mitzner, T. (2015, May). The effect of delivery method, speaker demographics, and physical environment on the engagement level of older adults. *Journal of Emerging Investigators*, 1-10.
- Shaw, L. A., Turrel, R., Wunsche, B. C., Lutteroth, C., Marks, S., & Buckley, J. (2016, February). Design of a virtual trainer for exergaming. In *Proceedings of the Australasian Computer Science Week Multiconference* (p. 63). ACM. doi:10.1145/2843043.2843384
- Staiano, A. E., Abraham, A. A., & Calvert, S. L. (2012). Competitive versus cooperative exergame play for African American adolescents executive function skills: Short-term effects in a long-term training intervention. *Developmental Psychology*, 48(2), 337–342. doi:10.1037/a0026938 PMID:22369339
- Sun, H. (2013). Impact of exergames on physical activity and motivation in elementary school students: A follow-up study. *Journal of Sport and Health Science*, 2(3), 138–145. doi:10.1016/j.jshs.2013.02.003
- Vernadakis, N., Papastergiou, M., Zetou, E., & Antoniou, P. (2015). The impact of an exergame-based intervention on childrens fundamental motor skills. *Computers & Education*, 83, 90–102. doi:10.1016/j.compedu.2015.01.001
- Vicencio-Moreira, R., Mandryk, R. L., & Gutwin, C. (2015, April). Now you can compete with anyone: Balancing players of different skill levels in a first-person shooter game. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 2255-2264). ACM. doi:10.1145/2702123.2702242
- Wittland, J., Brauner, P., & Ziefle, M. (2015). Serious Games for Cognitive Training in Ambient Assisted Living Environments – A Technology Acceptance Perspective. In J. Abascal, S. Barbosa, M. Fetter, T. Gross, P. Palanque, & M. Winckler (Eds.), *Human-Computer Interaction – INTERACT 2015* (pp. 453-471). Springer.

ADDITIONAL READING

- Blumberg, F. C., & Blumberg, F. (Eds.). (2014). *Learning by playing: Video gaming in education*. USA: Oxford University Press. doi:10.1093/acprof:osobl/9780199896646.001.0001

Bredl, K., & Bösche, W. (Eds.). (2013). *Serious games and virtual worlds in education, professional development, and healthcare*. IGI Global. doi:10.4018/978-1-4666-3673-6

Fisher, C. (2014). *Designing Games for Children: Developmental, Usability, and Design Considerations for Making Games for Kids*. CRC Press.

Information Resources Management Association. (2015). *Gamification: Concepts, Methodologies, Tools, and Applications*. IGI Global.

Niemi, H., Multisilta, J., Lipponen, L., & Vivitsou, M. (Eds.). (2014). *Finnish Innovations and Technologies in Schools: A Guide Towards New Ecosystems of Learning*. Springer. doi:10.1007/978-94-6209-749-0

Reinecke, L., & Oliver, M. B. (Eds.). (2016). *The Routledge Handbook of Media Use and Well-Being: International Perspectives on Theory and Research on Positive Media Effects*. Routledge.

Reiners, T., & Wood, L. (Eds.). (2015). *Gamification in Education and Business* (1st ed.). Springer. doi:10.1007/978-3-319-10208-5

Schaaf, R., & Mohan, N. (2014). *Making School a Game Worth Playing: Digital Games in the Classroom*. Corwin Press. doi:10.4135/9781483378534

KEY TERMS AND DEFINITIONS

Exergame: A combination of “exercise” and “game” is a term used for video games that are also a form of exercise.

Game User Research: It focuses on players’ behavior via techniques such as playtesting, analytics, expert analysis, and others.

Game-Based Learning: A type of game play that has defined learning outcomes.

Skill Acquisition: A specific form of learning as the representation of information in memory concerning some environmental or cognitive event.

Category S

Systems and Software Engineering

Adopting Open Source Software in Smartphone Manufacturers' Open Innovation Strategy

S

Mohammad Nabil Almunawar

Universiti Brunei Darussalam, Brunei

Muhammad Anshari

Universiti Brunei Darussalam, Brunei

Heru Susanto

Indonesian Institute of Sciences, Indonesia & Tunghai University, Taiwan

INTRODUCTION

Recent advancement of information and communication technology on mobile communication has changed the way people communicate, interact, and carry out their daily activities. The new era of wireless multimedia communication with smartphones has replaced old cell phones. Major players in the cell phone industry such as Nokia has lost their market shares and new players such as Apple (iPhones) and Samsung (Android phones) dominate the market.

Smartphones are very handy. Although they cannot replace all desktop or laptop's functionalities, they can be carried around conveniently as multipurpose devices. There are a myriad of applications (apps) that can run on smartphones and thousands of new apps are created everyday. The main advantage of smartphones is its ability to connect to the Internet from anywhere, enabling their users to have a complete Internet experience, stay in touch with their families, friends and colleagues, checking emails, making reservations, checking the traffic condition and so forth. Travelling with a smartphone is very helpful. With a digital map app, finding a place is much simpler in comparison to conventional methods. The map can also enlighten us to the location of ATM machines, restaurants, gas stations and others.

While traveling, a person with a smartphone can easily navigate his/her way through unfamiliar routes and keep in touch with family and friends through social networking such as sharing photos with Instagram or video chatting with Skype.

The heart of a cell phone or smartphone is an operating system (OS), a system that controls and manages all resources. As companies that produce cell phones rely on conventional innovation, where they rely on vertically integrated research to foster new technologies (innovations) for competitive advantage, most OSs for cell phones are proprietaries. Consequently an attempt to introduce open source OS in cell phones such as mobile Linux was not very successful since companies producing cell phones need to adopt Open Innovation for their core technology. According to Chesbrough, Vanhaverbeke & West (2006) "Open Innovation is the use of purposive inflows and outflows of knowledge to accelerate internal innovation, and expand the markets for external use of innovation, respectively." This means companies can adopt external technology such as OSS in advancing their technologies and markets.

Open Source Software (OSS) is free software that provides its users with freedom to use, replicate, modify, and distribute for any purpose. Unlike proprietary software where the executable code is commercially distributed under a copyright

DOI: 10.4018/978-1-5225-2255-3.ch641

law, the source code of OSS is available and a user has freedom to modify the source code, creating another version or an extended version of the software.

OSS proponents often state that it offers significant benefits when compared to typical commercial or proprietary software. Commercial software typically favours visible features (marketing advantage) over harder-to measure qualities such as stability, security and similar less glamorous attributes. In short, although OSS visible features may not be as good as the commercial one, OSS quality is normally high. OSS developers are evidently motivated to focus more on quality rather than visible features. For many developers, peer review and acclaim are important. They definitely prefer to build software admired by their peers where clean design, reliability and maintainability, with adherence to standards is highly regarded. "The Open Source community attracts very bright, very motivated developers, who although frequently unpaid, are often very disciplined" (Peeling & Satchell, 2001).

Nowadays most smartphone vendors have adopted OSS in their Open Innovation strategy to expand their business. Open source OS for smartphones such as Android have been widely accepted and in fact has been a dominant OS for smartphones since a few years ago. Interestingly, some smartphone vendors that used proprietary OS, developed and intend to use open source OS such as Maemo and MeeGo (Nokia), and webOS (Palm). Tizen, an alternative open source OS, has been developed by Samsung, Intel and Linux Foundation. Some old OSS players have released open source OS for smartphones such as Firefox OS from Mozilla and Ubuntu Touch from Canonical.

This chapter discusses the Open Innovation and adoption of OSS in smartphone industry. The development of OSS and its use in smartphones will be presented. The competition between proprietary and OSS operating system for smartphones will be discussed as platforms or operation system that shape the smartphone industry. Future direction will be presented in the last part of this chapter.

The next section will present the background followed by a discussion on the development of OSS and smartphone platforms in Section 3. Section 4 will discuss the impact of OSS on smartphone industry. Additionally, Section 5 will be on the future direction and the last section concludes this topic.

BACKGROUND

Open Innovation is a concept introduced by Chesbrough (2003) in which a company use both internal and external knowledge to create values. Chesbrough (2003) defines Open Innovation as "A paradigm that assumes that firms can and should use external ideas as well as internal ideas, and internal and external paths to market, as the firms look to advance their technology." Open Innovation can be contrasted with closed innovation where a company relies on vertically integrated research to foster new technologies or innovations and keeps them secret or strictly protected using intellectual property (IP), meaning that no other companies can use them.

Open Innovation and OSS are two different concepts sharing some similarities as said by Chesbrough (2006): "Open Innovation is sometimes conflated with open source methodologies for software development. There are some concepts that are shared between the two, such as the idea of greater external sources to create values." Technically OSS can be considered as a kind of Open Innovation where a company outsource software needed to the community (open source community) or alternatively a company can contribute its software to the community after turning it to open source software.

Interestingly, though the benefits of OSS are obvious, OSS has not yet been widely used and accepted in the PC world. Despite some of the hindrance factors mentioned above, some OSS gained wider acceptance, such as Linux, Apache, Sendmail, PHP, MySQL, and Firefox. Linux for example is a very reliable and powerful operat-

ing system that can be run on less powerful PCs. Remarkably a well-known website, Google, has used Linux servers since its inception. According to Sergey Brin, who was one of Google founders, said that they chose to use Linux because it is the most cost effective solution (Schumaker, 2000).

OSS is always strong in the platform level. This strength prevails in the smartphone world. Smartphones integrate cell phones and personal digital assistant (PDA) technologies. Cell phones with some features of smartphones were introduced by the end of 90s or early 2000 such as the Nokia 9000 Communicator and BlackBerry. In fact, Ericson used the word smartphone for its GS88 in 1997 (Yousuf, 2013). However, the major breakthrough was the introduction of iPhone by Apple in 2007 with its proprietary OS, iOS. The mobile phone market was shaken by the introduction of iPhone. By 2008, iPhone with iOS dominated the smartphone market with market share close to 80%. The success of iPhone was followed by the introduction of Android by Google in 2008. Domination of iOS was challenged by Android after one year of its introduction. Within a few years, smartphone OS is dominated by Android, an open source OS, with market share of 79.3% in the second quarter of 2013, while iOS is losing its market share, 13.2% in the second quarter of 2013 (Clover, 2013).

LITERATURE REVIEWS

According to Chesbrough (2011), there are two facets to Open Innovation, outside in and inside out. Outside in commonly occurs in Open Innovation. It is when a company imports knowledge, ideas or technologies from external sources and integrate them into the company innovation process. Meanwhile, the less commonly occurring inside out Open Innovation is when a company exports underutilized knowledge, ideas or technologies to be incorporated into other companies' innovation processes.

OSS is a manifestation of Open Innovation in software development (Liu, 2009). Both outside in and inside out approaches are utilized by companies that adopted OSS as part of their innovation strategy. Android is a notable example of outside in Open Innovation adopted by smartphone manufacturers.

The history of OSS can be traced back to early development of large-scale commercial computer in the 1950s and 1960s where software was free and came with source code. The source code can be modified and recompiled to improve the software. At this time the software was not considered as a revenue generator, instead, software was a necessary ingredient to make hardware work and provide useful functionalities or solutions so that people are encouraged to buy expensive hardware or mainframes.

The term Free Software (FS) has been put in place by Richard Stallman (Stallman, 2010). He is the one who established GNU Project and Free Software Foundation and introduced GNU General Public License (GNU GPL). This license guarantees that software under it is free. The term open source software is used instead of free software to embrace the business world (Perens, 1999).

The open source movement shares the basic principles of Stallman's free software movement. As Stallman (2010) said, "free software" and "open source" describe the same category of software, more or less, but say different things about the software, and about values. He further said, "The Free Software Movement and Open Source Movement are two political parties in the same community."

The open source movement has taken a different path from the free software movement to promote the idea of free software. Notable persons in Open Source Movement such as Eric Raymond, Bruce Preens, Tim O'Reilly and others are concerned about the negative connotations of the previously used name Free Software Foundation (DiBona, Ockman, & Stone, 1999), a term considered to be non-business friendly. Thus, they

agreed to promote the idea with a new term, open source software. The new term attracted a lot of support from hacker culture (Raymond, 1998) as well as software companies such as Netscape, Sun Microsystems, and IBM.

The rapid development of OSS has taken place since the 1990s. In the early 1990s, for instance, Net/2, or 386BSD, a UNIX based operating system that was free from AT & T copyrighted code was introduced. Hence, 386BSD was a completely free operating system. The 386BSD was the root of other BSD-based operating systems such as OpenBSD, FreeBSD and NetBSD.

Another milestone in OSS is Linux or to be precise, GNU/Linux. Linus Torvalds originally developed Linux when he created the Linux kernel based on Minix, a UNIX-like small operating system that could run on a PC, designed by Tanenbaum (2006) for teaching purposes. Torvalds and GNU project team collaborated and integrated high-level functions produced by GNU with Torvalds's kernel to form a new operating system called Linux. Linux was released under the GNU GPL license (Towle, Keppler, & McFarland, 2004).

Currently, there are thousands of high-quality OSS which addresses many solutions for different areas. The OSS eventually enjoyed wide acceptance with the support from giant IT companies such as IBM, Hewlett-Packard, Sun Microsystems, and Google. Many government organizations have committed to adopt OSS (van Wendel de Joode, Lin, & David, 2006).

OSS development is different from the traditional software development, which relies on central control that follows a strict schedule. Many developers, based on trust and cooperation or collaboration, develop OSS. Raymond (2001) in his famous work differentiate traditional software and OSS development as "The Cathedral and The Bazaar."

Linus Torvalds introduced the OSS development model when he released the source code of the Linux kernel for Intel 386 machines on the Minix community in the Internet (comp.os.minix)

and was requesting for collaboration in order to further develop it. He received many responses and many programmers contributed code to the project. Later on, in January 1991 the Linux community was formed by launching a news group called alt.os.linux, which became comp.os.linux in March 1992 (Torvalds, 1999).

Collaboration among programmers in the Linux community with the leadership of Torvalds has made Linux a strong and stable operating system in a flash. Raymond (2001) called Torvalds's development model as the bazaar development model.

Prior to smartphones, all cell phones operating systems were proprietary. Until 2009, Symbian (OS used by Nokia cell phones) was a dominant OS followed by RIM BlackBerry OS (Gartner, 2007). Perhaps the only OSS component on cell phones was Java ME, a platform to run applications.

Research in Motion (RIM) introduced BlackBerry in 1999 with an innovative feature, which was not supported by other phones. BlackBerry phones have the capability to send and receive emails or messages through BlackBerry Messenger (BBM) supported by a mini QWERTY keyboard that simplifies typing messages (Hall & Anderson, 2009). A significant breakthrough in mobile phone industry is the introduction of iPhone in 2007. The iPhone OS, iOS, is a completely proprietary system. After its introduction, iPhone enjoyed wide acceptance from end users and quickly held a significant market share in 2008 and iOS became a very popular OS.

Google entered the smartphone market by introducing Android, an OSS of Linux-based operating system, in 2008. Many major smartphones vendors such as Samsung, HTC, and LG adopted Android as the OSS in their products. In addition, the growth of apps that can work on Android staggered, consequently in a very short time Android challenged iOS, took over the market, and became dominant in 2011. Nokia, the dominant player in cell phones keep losing its market shares since the introduction of iOS and then lose further after the wide acceptance of Android.

The popularity of open source OS for smartphones has attracted several vendors such as Nokia, Samsung, and Palm/HP to develop open source OSs. Nokia bought Symbian and turned it to open source OS and subsequently developed Maemo (Maemo.org) and MeeGo (Kenney & Pon, 2011; Morrison, 2013) but lost faith to both of them. Samsung has developed Bada (Bada.com) which has been absorbed by the new open source OS project, Tizen (Tizen.org). Palm/HP developed webOS, which was acquired by LG in February 2013 (Ion, 2013).

There are some new OSS projects on operating systems for smartphones such as Firefox OS (www.mozilla.org/en-US/firefox/os/), Sailfish OS (sailfishos.org) and Ubuntu Touch (www.ubuntu.com/phone), and Jolla (Jolla.com).

WHY OSS PREVAIL IN SMARTPHONES

OSS in the Computer World

OSS has been around for more than two decades. The strength of OSS has been recognized, especially in the platform level. Several researches have been conducted to compare the reliability of OSS as compared to proprietary counterparts and data indicated that OSS is more reliable (Forrester & Miller, 2000; Miller et al., 1995; Miller, Cooksey, & Moore, 2006).

So far, majority of OSS influence is on the platform level, unable to shake the domination of Microsoft on the consumer level (PCs). According to Cosovanu (2006), there are four main reasons as to why OSS is still not widely used. Firstly, cheap pirated copies of proprietary software are widely available. Secondly, proprietary software, especially Microsoft operating system and its main application suite have achieved overwhelming market share, which is over 90% among the PC users. Thirdly, most new PCs are pre-installed with Microsoft operating system and its office suite. Fourthly, there is a conservative bias of many busi-

nesses and state bureaucracies toward proprietary software, especially Microsoft's products.

However, a report from Forrester (2007) indicated a growing adoption of OSS across all geographies and industries. The roles of major vendors (investment, contributions, participation) in this expansion are quite significant. Based on the survey conducted by Forrester, many companies in North America and Europe has trusted OSS in their mission-critical applications although the majority of the companies surveyed used OSS in non-mission-critical applications, application infrastructure (application server, portal server, web server), server operating system and development tools.

The Forrester report further reveals that it is not the cost of acquiring (no license cost for OSS) that appeals to companies, however, the three attributes namely, support for open standards, the ability to use without restriction, and not being locked in to a single vendor. Figure 1 shows the degree of importance of OSS attributes.

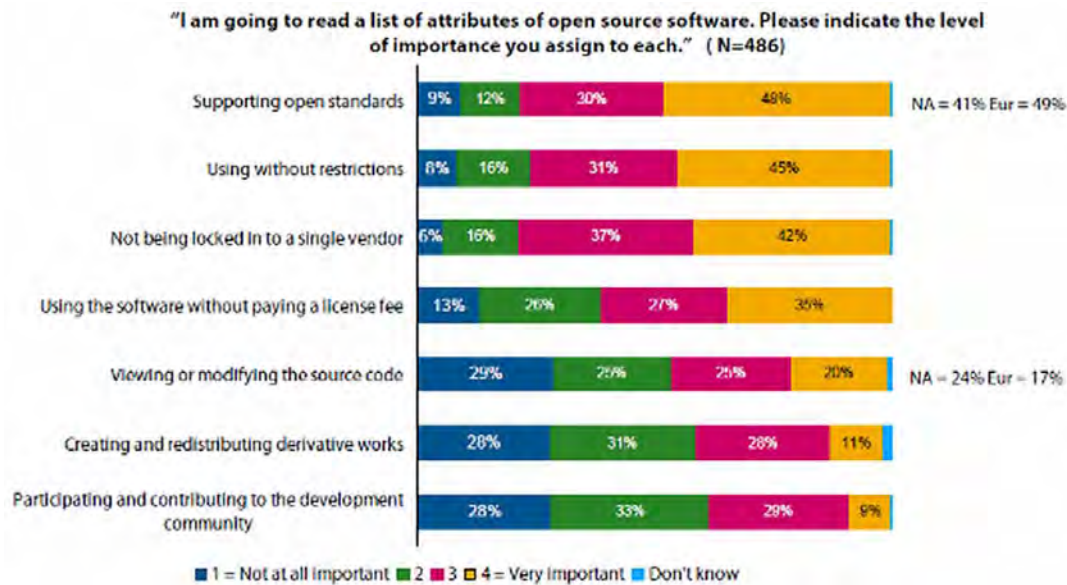
It is interesting to see the change of attitudes of the business world, especially large corporations, toward OSS. At the early stage of OSS development, large companies were suspicious. After sometime, they accept the phenomena of OSS and finally indicate their favorability by adopting OSS or Open Innovation in general as part of their business strategy (Liu, 2009).

OSS in the Smartphone World

Prior to the introduction of smartphones, few players dominated the cell phone industry. The top five vendors in 2006 were Nokia, Motorola, Samsung, Sony Ericson, and LG with market shares of 36.2 percent, 21.5 percent, 11.3 percent, 9 percent, and 6.3 percent respectively and the figures did not change much from the previous year (Gartner, 2007). None of these vendors has OSS in their core products other than the Java platform (Java ME) to run Java-based applications (applets).

In 2009, two years after the introduction of iOS and one year after the introduction of Android,

Figure 1. Preference of Attributes of Open Source Software (courtesy of Forrester, 2007)



the smartphone industry, especially Samsung and HTC, responded positively by embracing Android as the operating system of their smartphones. Similarly, application developers responded with a myriad of apps for Android within a short time. To see the effects of OSS (in this case Android), we have to look at the market shares of major operation systems for smartphones in Table 1 and Figure 2 shows the trend graphically. Note that Android is the only open source OS in Table 1. By 2011 the smartphones world is dominated by two OS, Android and iOS, an increasing trend for Android and a decreasing trend for iOS except for quarter 2 (Q2) in 2015.

According to Yu (2013) a smartphone OS market share has a correlation with its mobile apps developer community as there is a relationship between smartphones, OS and apps. This indicates that Open Innovation prevails in the smartphone industry. Android enjoys strong support from its developer community; consequently, the number of apps is growing fast. The role of apps in attracting users of smartphones is important as they expand the functionality of the smartphones.

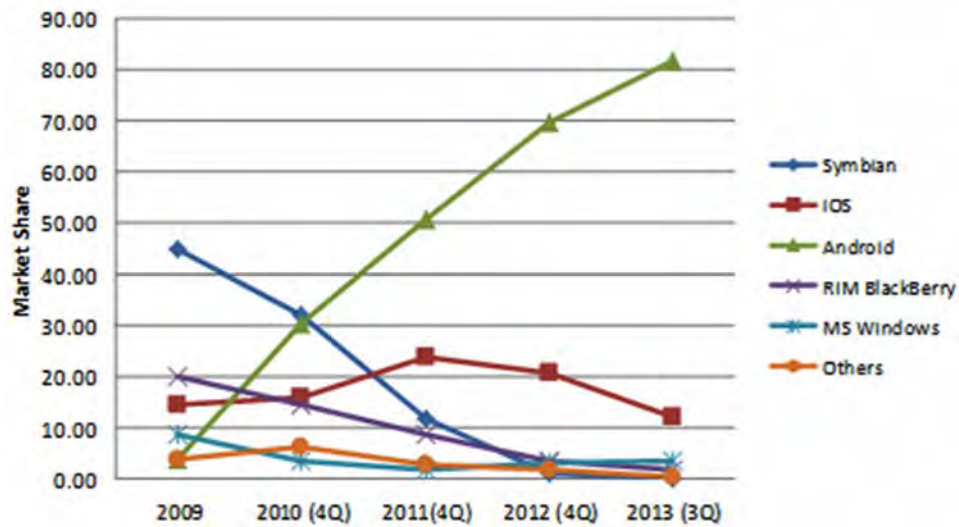
Figure 3. shows the growth of apps in Sourceforge (sourceforge.net), a cloud for developing applications, including mobile apps. If we explore the growth of apps in Sourceforge, the most popu-

Table 1. Smartphone operating systems market share (in percent)

Operating System	2009	2010 (Q4)	2011(Q4)	2012(Q4)	2013(Q2)	2014(Q2)	2015(Q2)
Symbian	44.9	32.3	11.7	1.2			
iOS	14.4	15.8	23.8	20.9	12.9	11.6	13.9
Android	3.9	30.5	50.9	69.7	79.8	84.8	82.8
RIM BlackBerry	19.9	14.6	8.8	3.5	2.8	0.5	0.3
MS Windows	8.7	3.4	1.9	3.0	3.4	2.5	2.6
Others	3.8	6.1	2.8	1.6	1.2	0.7	0.4

Source: various reports from Gartner and IDC

Figure 2. The growth of market share of major mobile OS

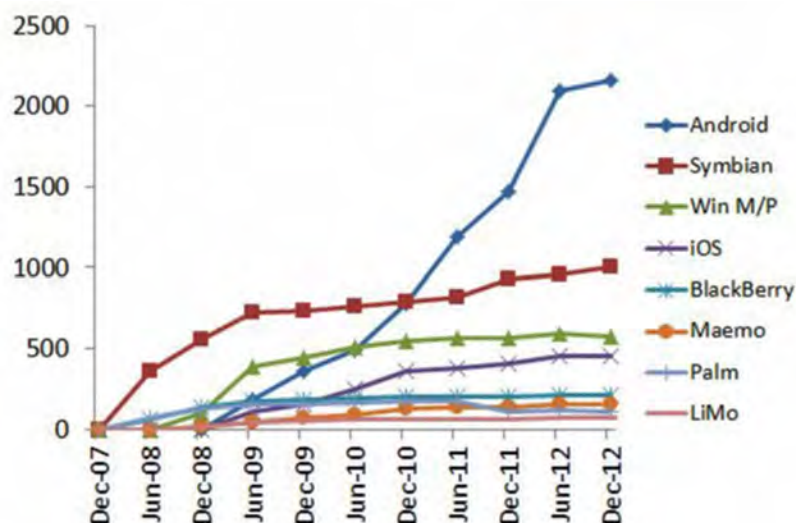


lar and the largest OSS community developer's repository, we can infer that Android has been enjoying a strong support from OSS community starting from its conception.

It cannot be denied that Apple, through iPhone and iOS, created a breakthrough in mobile device technology. iPhone integrates telephony, the Internet and a myriad of applications from third parties, making iPhone outperform other mobile phones. The seamless Internet connection with

full browser capability has let end users experience Internet value network ubiquitously (Kenney & Pon, 2011). In addition, AppStore and iTunes have created an innovative ecosystem containing end users, third-party application developers, and Apple Inc. The openness of iOS to third-party developers further increase the value of iPhone (iPod and iPad), and iOS as innovations from the third-party applications are embraced, creating symbiotic mutualism between Apple and its third-

Figure 3. The growth of apps of several mobile OS in Sourceforge (Source: Yu, 2013)



party developers. Although iOS is a close source software, proprietary and exclusively designed for Apple's products, Apple also adopts Open Innovation strategy to multiply the value of its products in a controlled manner.

The exclusiveness of Apple for its products (Apple will not license iOS to any smartphone manufacturer) has made the introduction of Android, an open source OS, warmly welcomed by smartphone manufacturers such as HTC and Samsung, subsequently producing smartphones with comparable capabilities to iPhone.

Google, the big player in the Internet, decision to enter smartphone market is interesting. Google did not have experience in any mobile platform. Therefore the best way is to pick a good open source OS, adopting fully the Open Innovation approach to secure its position as the top Internet player since smartphones are ubiquitous client machines to access the Internet. Google bought Android Inc, a company co-founded by Andy Rubin that develops a Linux-based operating system for touch screen mobile devices and appointed Andy Rubin as the director of mobile platforms at Google (Beavis, 2008). With the cooperation of Open Handset Alliance (OHA), which includes Samsung, HTC, Sony, and LG, Google unveiled Android on November 5, 2007 and within one year (October 22, 2008), the first Android smartphone, HTC Dream, released (Perez, 2008).

Google does not use Android as a moneymaker directly as Android was released as an OSS and smartphones manufacturers do not need to pay a license fee to use Android as the platform for their smartphones. Google expects Android to attract people to use its Internet services, creating more traffic to get more revenue from the traffic through advertisements and data usage. Consequently, Google spent considerable effort to promote Android, especially to application developers, to create a similar ecosystem to iPhone (Kenney & Pon, 2011). Figure 2 shows the response from application developers that use Sourceforge as their development repository. In addition, the expansion number of apps developed

for Android in several well-known repositories such as Sourceforge, Google Code, and Git Hub is surely highly related to the expansion of available apps in Google Play, which in turn help Google strengthening its Android ecosystem.

FUTURE RESEARCH DIRECTIONS

Android is now a dominant OS for smartphones. However, Google does not acquire direct financial benefit from the popularity of Android. Perhaps the acquisition of Motorola Mobility in May 2012 is an attempt to take direct benefit from Android, but failed. It is interesting to study the reason as to why Motorola Mobility was acquired. One of the reason revealed by Google that the intention in acquiring Motorola Mobility is to gain control on patents owned by Motorola Mobility to protect Android's ecosystem. However, the ownership was not long, Google sold Motorola Mobility to Lenovo in January 2014. On the other side, Google has Chrome OS a product of the Chromium OS Project, an OSS project aimed to produce a web-based operating system. It is not clear how Google treat Android if Chrome OS successfully gain market, when Google has full control on the latter. Does Google want to take advantage of the advancement of mobile computing with a lightweight OS (Chrome OS)?

As a first mover in touch screen mobile devices (smartphones and tablets) Apple with its strong brand and innovative products has always secured its market, although its market share has decreased. In the tablet market, iPad is still leading the market with a market share of 55% (October 2013). Mobile device products with proprietary operating systems keep losing the market such as Windows OS and BlackBerry. The survival strategy of Nokia by taking a proprietary OS (Windows OS) is not clear. Microsoft fully absorbed Nokia in April 2014. The big question is: can Nokia, under Microsoft, gain significant market share in the future?

There are several open source projects working on operating systems for smart mobile devices.

Among others are Firefox OS, Sailfish OS, Tizen and Ubuntu Touch. We do not know the fate of these OSs in the future. However, they are alternative OSs that may challenge the dominant duo, Android and iOS. Perhaps the new open source OS can tap the emerging markets for lower-end mobile devices. As discussed previously, the key for a mobile OS to success in the market is strong support from application developers and creating an ecosystem to attract application developers to contribute their innovation.

Nowadays people live with their smartphones, but unfortunately, information stored in their smartphones is well protected. A growing concern from end users on security and privacy issues need to be addressed. Mobile network is far from secure and the same goes to the Internet. Therefore, research are needed to advance security and privacy issues in smart mobile devices.

CONCLUSION

Open Innovation prevails in smartphone industry. OSS is a kind of Open Innovation in software development majorly utilized by smartphone manufactures as part of their innovation strategy.

OSS is a free software that provide users the freedom to use, replicate, modify, and distribute for any purpose. In the early stages of OSS, the business world, especially large corporations were not interested and some were suspicious of OSS. Strong support from programmers and quality software produced through OSS development methodology have changed the attitude of the business world to favor, support and embrace OSS including many big computer corporations such as IBM, Oracle and Google.

Apple made a breakthrough in consumer mobile devices by introducing touch screen mobile phone (iPhone) and tablet (iPad) equipped with powerful but easy to use operating systems in 2007 with its proprietary OS, iOS. The mobile phone market was shaken by the introduction of iPhone, and within one year, the smartphone

market was dominated by iPhone. Google, a giant Internet player entered the smartphone market by introducing Android, an OSS of Linux-based operating system in 2008. Supported by many smartphone manufacturers, application developers and open source communities, Android quickly grabbed a significant market share after one year of its introduction and continues to grow until it became dominant after three years of its introduction.

One of the important key success factors of any smartphone platform is innovation created from the platform and applications developed for the platform that greatly attracts end users. Both Android and iOS have successfully created good ecosystems to help them grow. However, the power of Open Innovation in open source software to create such ecosystem is stronger than proprietary software, possibly one of the reason for the widening gap between Android and iOS market shares.

The popularity of open source OS for smartphones has attracted several vendors such as Nokia, Samsung, and Palm/HP to develop open source OSs. New players such as Mozilla (Firefox OS) and Canonical (Ubuntu Touch) are trying to penetrate the market. However, their success depends on their ability to create a good ecosystem to grow and challenge the market leader.

REFERENCES

- Bada.com. (n.d.). *Bada - Samsung's smartphone platform*. Retrieved on November 18, 2013, from <http://www.bada.com/>
- Beavis, G. (2008). *A complete history of Android*. Retrieved on November 19, 2013, from <http://www.techradar.com/news/phone-and-communications/mobile-phones/a-complete-history-of-android-470327>
- Chesbrough, H. (2003). *Open Innovation: The new imperative for creating and profiting from technology*. Harvard Business Press.

- Chesbrough, H. (2015). *Everything You Need to Know About Open Innovation - Forbes*. Retrieved November 6, 2015, from <http://www.forbes.com/sites/henrychesbrough/2011/03/21/everything-you-need-to-know-about-open-innovation/>
- Chesbrough, H., Vanhaverbeke, W., & West, J. (2006). *Open Innovation: Researching a New Paradigm*. Oxford, UK: Oxford University Press.
- Clover, J. (2013, August 7). *Android dominates nearly 80% of smartphone market, ios drops to 13% share*. Retrieved November 16, 2013, from <http://www.macrumors.com/2013/08/07/android-dominates-nearly-80-of-smartphone-market-ios-drops-to-13-share/>
- Cosovanu, C. (2006). Open source software in eastern europe and other emerging markets: The moral alternative to piracy? *Journal of Internet Law*, 9, 1–14.
- DiBona, C., Ockman, S., & Stone, M. (1999). Introduction. In *Open sources: voices from the open source revolution*. Sebastopol, CA: O'Reilly Media, Inc. Retrieved from <http://oreilly.com/openbook/opensources/book/>
- Forrester, J. E., & Miller, B. P. (2000). An empirical study of the robustness of windows nt applications using random testing. In *Proceedings of the 4th Conference on USENIX Windows Systems Symposium* (Vol. 4, pp. 6–6). Berkeley, CA: USENIX Association. Retrieved from <http://dl.acm.org/citation.cfm?id=1267102.1267108>
- Forrester. (2007). *Open source software's expanding role in the enterprise*. Retrieved from http://www.anpop.org.eg/Data/Sites/1/media/forrester_research-open_source_buying_behaviors.pdf
- Gartner. (2007, March). *Gartner says worldwide mobile phone sales grew 21 percent in 2006*. Retrieved on November 16, 2013, from <http://www.gartner.com/newsroom/id/501734>
- Google completes purchase of Motorola Mobility. (2012). BBC. Retrieved on from <http://www.bbc.co.uk/news/business-18164190>
- Hall, S. P., & Anderson, E. (2009). Operating systems for mobile computing. *Journal of Computing Sciences in Colleges*, 25(2), 64–71.
- Ion, F. (2013, February). *What LG's acquisition of webOS means for its open source projects*. Ars Technica. Retrieved on November 18, 2013, from <http://arstechnica.com/gadgets/2013/02/what-lgs-acquisition-of-webos-means-for-its-open-source-projects/>
- Kenney, M., & Pon, B. (2011). Structuring the smartphone industry: Is the mobile Internet os platform the key? *Journal of Industry, Competition and Trade*, 11(3), 239–261. doi:10.1007/s10842-011-0105-6
- Liu, Y. W. (2009). The implication of Open Innovation and open source to mobile device manufacturer. Massachusetts Institute of Technology. Retrieved from <http://dspace.mit.edu/bitstream/handle/1721.1/55224/611948447.pdf>
- Maemo.org. (n.d.). *maemo.org - maemo.org: Home of the Maemo community*. Retrieved on November 18, 2013, from <http://maemo.org/>
- Miller, B. P., Cooksey, G., & Moore, F. (2006). An empirical study of the robustness of MacOS applications using random testing. In *Proceedings of the 1st International Workshop on Random Testing* (pp. 46–54). New York: ACM. doi:10.1145/1145735.1145743
- Miller, B. P., Koski, D., Pheow, C., Maganty, L. V., Murthy, R., Natarajan, A., & Steidl, J. (1995). *Fuzz revisited: A re-examination of the reliability of UNIX utilities and services*. Academic Press.

Morrison, G. (2013). *What went wrong with MeeGo? "Nokia lost faith in the project"*. TechRadar. Retrieved on November 18, 2013, from <http://www.techradar.com/news/software/what-went-wrong-with-meego-nokia-lost-faith-in-the-project--1147770>

Netcraft. (2012). *November 2012 web server survey*. Retrieved on November 16, 2013, from <http://news.netcraft.com/archives/2012/11/01/november-2012-web-server-survey.html>

Peeling, N., & Satchell, J. (2001). *Analysis of the impact of open source software*. QINETIQ QINETIQ/KI/SEB/CR010223. Retrieved on November 7, 2015, from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.115.8510&rep=rep1&type=pdf>

Perens, B. (1999). The open source definition. In *Open sources voices from the open source revolution* (pp. 171–188). Sebastopol, CA: O'Reilly Media, Inc.

Perez, S. (2008). *First Android phone approved by FCC*. ReadWrite. Retrieved on November 20, 2013, from http://readwrite.com/2008/08/18/first_android_phone_just_approved

Raymond, E. S. (1998). *Goodbye, "free software"; hello, "open source"*. Retrieved on November 19, 2013, from <http://www.catb.org/~esr/open-source.html>

Raymond, E. S. (2001). *The cathedral & the bazaar: musings on Linux and open source by an accidental revolutionary*. Cambridge, MA: O'Reilly.

Schumaker, J. (2000). Interview with Google's Sergey Brin. *Linux Journal*. Retrieved from <http://www.linuxjournal.com/article/4196>

Stallman, R. (2010). *Free software, free society: selected essays of Richard M. Stallman*. Boston, MA: SoHo Books.

Tanenbaum, A. S., & Woodhull. (2006). *Operating systems: design and implementation*. Upper Saddle River, NJ: Pearson/Prentice Hall.

Tizen.org. (n.d.). *Tizen: An open source, standards-based software platform for multiple device categories*. Retrieved on November 18, 2013, from <https://www.tizen.org/>

Torvalds, L. (1999). The Linux Edge. In *Open sources voices from the open source revolution*. Sebastopol, CA: O'Reilly. doi:10.1145/299157.299165

Towle, H., Keppler, E., & McFarland, C. (2004). Open source issues in business. *Journal of Internet Law*, 3(11).

Van Wendel de Joode, R., Lin, Y., & David, S. (2006). Rethinking free, libre and open source software. *Knowledge, Technology & Policy*, 18(4), 5–16. doi:10.1007/s12130-006-1001-y

Yousuf. (2013). *History Review of the Smartphones Over 20 Years*. Retrieved on November 16, 2013, from <http://www.techinfo2.com/smartphones-history-reveiew-over-20-years.html>

Yu, L. (2013). The coevolution of mobile os user market and mobile application developer community. *COMPUSOFT. An International Journal of Advanced Computer Technology*, 2(2), 44–48.

ADDITIONAL READING

Bonaccorsi, A., & Rossi, C. (2003). Why Open Source software can succeed. *Research Policy*, 32(7), 1243–1258. doi:10.1016/S0048-7333(03)00051-9

Butler, M. (2011). Android: Changing the Mobile Landscape. *IEEE Pervasive Computing / IEEE Computer Society [and]. IEEE Communications Society*, 10(1), 4–7. doi:10.1109/MPRV.2011.1

- Crowston, K., & Howison, J. (2005). The social structure of free and open source software development. *First Monday*, 10(2). doi:10.5210/fm.v10i2.1207
- Goadrich, M. H., & Rogers, M. P. (2011). Smart Smartphone Development: IOS Versus Android. In Proceedings of the 42Nd ACM Technical Symposium on Computer Science Education (pp. 607–612). New York: ACM. doi:10.1145/1953163.1953330
- Heitkötter, H., Hanschke, S., & Majchrzak, T. A. (2013). Evaluating Cross-Platform Development Approaches for Mobile Applications. In J. Cordeiro & K.-H. Krempels (Eds.), *Web Information Systems and Technologies* (pp. 120–138). Springer Berlin Heidelberg; Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-36608-6_8 doi:10.1007/978-3-642-36608-6_8
- Lakhani, K., & Wolf, R. G. (2003). Why Hackers Do What They Do: Understanding Motivation and Effort in Free/Open Source Software Projects (SSRN Scholarly Paper No. ID 443040). Rochester, NY: Social Science Research Network; Retrieved from <http://papers.ssrn.com/abstract=443040>
- Lakhani, K. R., & von Hippel, E. (2003). How open source software works: free user-to-user assistance. *Research Policy*, 32(6), 923–943. doi:10.1016/S0048-7333(02)00095-1
- Linghao, Z., & Ying, L. (2010). On methods of designing smartphone interface. In *2010 IEEE International Conference on Software Engineering and Service Sciences (ICSESS)* (pp. 584–587). doi:10.1109/ICSESS.2010.5552288
- Luo, T., Jin, X., Ananthanarayanan, A., & Du, W. (2013). Touchjacking Attacks on Web in Android, iOS, and Windows Phone. In J. Garcia-Alfaro, F. Cuppens, N. Cuppens-Boulahia, A. Miri, & N. Tawbi (Eds.), *Foundations and Practice of Security* (pp. 227–243). Springer Berlin Heidelberg. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-37119-6_15
- Roberts, J. A., Hann, I.-H., & Slaughter, S. A. (2006). Understanding the Motivations, Participation, and Performance of Open Source Software Developers: A Longitudinal Study of the Apache Projects. *Management Science*, 52(7), 984–999. doi:10.1287/mnsc.1060.0554
- Shih, G., Lakhani, P., & Nagy, P. (2010). Is Android or iPhone the Platform for Innovation in Imaging Informatics. [PubMed]. *Journal of Digital Imaging*, 23(1), 2–7. doi:10.1007/s10278-009-9242-4 PMID:19789925
- Tracy, K. W. (2012). Mobile Application Development Experiences on Apple's iOS and Android OS. *IEEE Potentials*, 31(4), 30–34. doi:10.1109/MPOT.2011.2182571
- Vaughan-Nichols, S. J. (2003). OSs battle in the smart-phone market. *Computer*, 36(6), 10–12. doi:10.1109/MC.2003.1204314
- West, J., & Gallagher, S. (2006). Challenges of Open Innovation: The paradox of firm investment in open-source software. *R & D Management*, 36(3), 319–331. doi:10.1111/j.1467-9310.2006.00436.x

KEY TERMS AND DEFINITIONS

Android: Android is an open source Linux-based operating system designed for touchscreen smart mobile devices (smartphones and tablet computers). Google and Open Handset Alliance (a consortium of hardware, software, and telecommunication companies such as Samsung, HTC, Qualcomm, and T-Mobile) use Android.

Free Software (FS): FS has been introduced by Richard Stallman to denote users the freedom to execute it for any purpose, replicate, modify and redistribute copies of the original or modified software. “Free” in the free software refers to freedom, not refer to price. Free software is under copyleft, a distribution concept to make sure that software under this term is always free.

iOS: iOS is proprietary operating system designed for dedicated touchscreen mobile devices developed by Apple such iPhone, iPad and iPod touch.

Open Innovation: Open Innovation is the use of purposive inflows and outflows of knowledge to accelerate internal innovation, and expand the markets for external use of innovation.

Open Source Software (OSS): Open source software share the same principle of FS. According to Richard Stallman 'free software' and 'open source' describe the same category of software, more or less, but say different things about the software, and about values. He further said, "The Free Software Movement and Open

Source Movement are two political parties in the same community".

Proprietary Software: Proprietary software is a commercial software where the core is kept secret by the developer, users of the software have to pay license fee and do not have the right to lend the software to their colleagues, let alone modify.

Smartphone: A smartphone is a mobile phone that let one not only make telephone calls, send short messages, but also surf the Internet, play games, listen to music, watch movies and other capabilities normally provided by computers such as writing and editing documents, sending and receiving emails and run a myriad of available apps.

S

Assessing Computer–Aided Design Skills

Yi Lin Wong

The Hong Kong Polytechnic University, Hong Kong

Kin Wai Michael Siu

The Hong Kong Polytechnic University, Hong Kong

INTRODUCTION

Computer-aided design (CAD) is the use of computer technology to aid design. It is often referred to the use of computer software to draw two- or three-dimensional drawings or models. It improves the quality of drawings and models, makes the communication in design processes more effective, and increase the productivity of the entire process. Most schools and universities would offer CAD courses for students in design or engineering disciplines and subjects.

Assessing students' knowledge and skills of using CAD computer software is thus inevitable. However, the assessment of CAD is often problematic. Computers are not always available because of the facilities arrangement and administrative issues in schools or universities. It is ineffective and difficult to arrange computers for a large group of students in an examination. The number of CAD software licenses may also be limited, and students are unable to use the software at the same time. Grouping students in different time slots to take the examination may be an eclectic way to solve the issue; however, the tutor or the lecturer would have to develop several sets of questions for the students. The examination in this case also loses its authenticity, as students are able to guess the questions based on the comments of students who took the exam in an earlier time slot. In addition, using particular software in CAD examination can only assess students' familiarity towards the software but not the general knowledge of CAD.

The issues of computer facilities and software license and the limitations of computer-based

examination made CAD assessment difficult. Taking references from some computer programming subjects, it is believed that using paper-and-pencil method is a choice to solve this problem. However, the scope of paper-and-pencil questions is limited, as the questions appeared in the examination can only comprise the most common functions in various CAD software.

This chapter discusses the issue through examining and reviewing existing CAD examination questions in Hong Kong. Based on the review, limitations of the current assessment are highlighted. It is argued that the existing method is inconsistent with the problem solving nature of the software and is unable to assess students with high ability and advanced drawing skills. Recommendation and improvement directions are suggested in the chapter to optimize the quality of assessment. Other subjects which also involve assessment of computer skills and design may also benefit from the recommendations.

BACKGROUND

The term computer-aided design appeared in the late 1960s (Abbas, 1976). Since then, many different disciplines of engineering applied CAD, and the industries also needed staff possessing this kind of skills to make their designs and products to catch up with the trend. Educational and academic institutes had also started to introduce CAD courses to their students. Some institutes had offered CAD course as an elective subject, or combined the knowledge of CAD in other

existing subjects related to engineering design. Specialized and short courses of CAD were very popular in various engineering disciplines and increasingly in design disciplines. As students in design disciplines had not got any technical training, they encountered difficulties in using the computer software, and sometimes they could not get what they wanted through CAD. Apparently, this discrepancy is getting smaller over the years, and now students without technical background are able to use computer software to assist design effortlessly. This is primary due to the popularization of computer technology in the last few decades.

Nowadays, CAD has been extending its application in other disciplines such as mathematics, music and geography. Various kinds of CAD software have been developed to attract the attention of the industries and the academia. The software may possess similar function but different layouts and orientations to provide choices for different needs of the customers. However, due to limitation in resource and time in universities, tutors and lecturers can only teach the selected software based on the actual needs. Students have to explore other software by themselves. Some secondary school teachers have also noticed the needs and started to contact the software companies to buy school license for the computers in the computer laboratories. They have started to rearrange the curriculum and put CAD knowledge and abilities as one of the learning outcomes.

Teaching CAD Software

One of the most effective ways of teaching CAD is to demonstrate the drawing procedures of a particular drawing or model step by step. Teacher can first show the finished drawings and ask students to follow the teacher while the teacher draws on the computer. Teacher will be able to help students and solve students' problems while teaching the steps. At the end of the lesson, all students will be able to finish the same drawings. Through the drawing process, students can learn

different functions of the CAD software. Teacher can also allow students to decide how they use a particular function in the drawings, e.g., whether the function should be applied on this line or the other line. In this way, different students can have different drawings in the end of the lessons. However, this is only applicable to students with higher learning ability. Sometimes the CAD skills learnt are transferable, and they can be applied in using other CAD software. There are many other methods to teach CAD skills. For example, teachers can use problem-based projects to facilitate students learning CAD software.

Learning CAD Software

Prior to learning CAD software, students need to possess basic computer knowledge. They should be very familiar with the operation system of their computer, e.g., Windows and Macintosh, before they can create and generate drawings and models in their mind through the CAD software. They also need to learn the basic shapes that comprise a drawing and analyze the drawings they wanted to create. They need to possess the cognitive ability to manipulate the drawings in the mind before working on the computer. The prior knowledge is essential to facilitate students to learn the functions of the CAD software. Once students are familiar to the functions and the working procedures of the software, they will be able to draw any drawings or make any models they need for their designs. Students will be able to produce neat and clear drawings and present their design ideas more effectively. Some students can make use of the features of the CAD software to design while drawing. The convenience and the effectiveness of CAD software also facilitate the idea generation in design processes. Chester (2007) had made a comprehensive discussion about CAD expertise. He identified two kinds of knowledge: command and strategic knowledge. Command knowledge is the knowledge of the commands and the drawing procedures. Strategic knowledge is the alternate method of finishing a specific drawing or task.

It can be argued that the strategic knowledge is essential in transferring the command knowledge in other CAD software.

Assessment of CAD Learning

Currently, the assessment of the CAD courses often focuses on the application. In general, there are two kinds of formal and formative assessment: computer approach and paper-and-pencil approach.

In computer approach, the typical assessment method is to request students to perform a series of tasks by using designated computer software in a given period of time. The computer software is often the one they used in lessons. When this approach is adopted, resources, time and question setting-up techniques become essential. In order to have a fair examination, the best way is to allow all students in a cohort to sit in the same examination at the same time doing the same questions. However, due to the limitation of computer and software resources, it is hardly possible. Another possible option is to assign students into different groups and take the examination at different time slots doing different set of questions; yet, this requires advanced question setting-up techniques to guarantee the different sets are of the same level of difficulties. In order to solve these issues in formal and formative assessment, paper-and-pencil approach may be one of the most feasible solutions.

In paper-and-pencil approach, students have to use verbal and graphical methods to describe the procedures needed to draw a particular drawing assigned in the examination question. This is the most convenient method, as the arrangement is the same as in the examination of other common subjects. Students are assigned to sit in a room, and what they need to do is to write their answers on the answer sheets.

Research About CAD

Unsurprisingly, educational researchers and researchers concerning the development of engi-

neering and design also start to pay attention to the CAD education. For instance, Gül (2014) studied the instructional methods used in CAD courses of interior architecture education. The four common instructional methods used in CAD courses identified by Dönmez (2013), namely theory based, invention based, searches and analyze based, and collaborative learning based instruction methods, were used to analyze a CAD course in a university of Turkey. Ye et al. (2004) investigated the contents taught in a CAD curriculum and found that the contents taught were not inadequate. Field (2004) discussed what kinds of knowledge are essential to support CAD training and education. Spatial reasoning, solid geometry, mathematics, and exercises in modular design within a team environment should be included. He argued that the CAD education should progress in line with the development of CAD technologies. Rossignac (2004) found that the typical way of teaching CAD was unable to improve students' ability of using CAD and advocated "education-driven research" in CAD.

Rossignac (2004) was not the only researchers who had found difficulties in teaching CAD. According to Piegl (2005), the CAD education still encounters several issues in its development. It lacks comprehensive textbooks for specialized CAD such as rapid prototyping and teachings aids for general purposes. There is no officially recognized CAD curriculum, and this has hindered CAD from globalization. There is also no unique international CAD society to gather all the experts. Up to day, there is no specific journal to support further research and conference about CAD education. The current inattention to CAD education has neglected the learning and teaching aspect of CAD knowledge, and soon universities would become "research university" with secondary focus on learning and teaching (Piegl, 2005).

All the issues and the researches about CAD education mentioned above are related to learning and teaching. None of them concerns the assessment of CAD education. In fact, assessment is important to evaluate the achievement of

students and the success of a curriculum (Morris, 1996). Through assessing students' learning outcome and performance, lecturers and tutors can revise the teaching methods applied and the contents selected for the curriculum. However, currently researches and discussions about assessment of CAD education are lacking. Data about how CAD beginners learn the computer software are also inadequate (Hamade, Artail, & Jaber, 2007).

Hamade et al.'s (2007) research study used time to assess students' performance. They measured the time that students used to draw an assigned model and correlated it with the number of the features selected by the students. This may provide a sounding empirical data about students' performance in CAD; however, this is not applicable and practical in normal classrooms or examinations where in-depth data collection and analysis are not considered.

DIFFICULTIES OF ASSESSING CAD KNOWLEDGE

Apparently, assessing CAD knowledge effectively is not as easy as it seems. Perhaps CAD knowledge teaching at the tertiary level is more flexible and independent that formal assessment is not required. However, when it comes to the context of secondary school level, the issue is more restricted, and it is necessary to develop a formal and formative assessment method to evaluate students' learning outcome in a one-off manner. The following sections of the chapter study the issues and difficulties of CAD assessment through an example of the paper-and-pencil approach that assesses students' 3D CAD drawing skills. Through the investigation and discussion, the difficulties of assessing CAD knowledge are identified. Also, the discussion is able to provide insight to modify and prove the current way of CAD assessment.

An Example of the CAD Assessment

S

Public examination about CAD is considered as a reliable example of paper-and-pencil approach for investigation. The subject of Design and Applied Technology (DAT) in the Hong Kong Diploma of Secondary Education (DSE) includes CAD as one of the elective modules in its curriculum. The elective module is named 'Visualisation and CAD Modelling'. In this module, students have to learn prototyping and drawing (both 2D and 3D) by using computers. However, most of the public examinations focus heavily on 3D drawings, and 3D drawing technique is often assessed. This type of questions is also very popular in the recent years, and it is welcomed by most students. Taking the 2014 DSE examination as an example, one of the questions asked students to 'draw' a fork end. The question is divided into two parts:

1. Sketch the primitive solid models required for constructing the 3D CAD model of the 'fork end' shown in Figure 15. (5 marks)
2. Using 'Boolean operations' and the primitive solid models from (a), draw a series of annotated sketches to show the major steps required in constructing the 3D CAD model of the 'fork end'. (20 marks) (extracted from Hong Kong Examinations and Assessment Authority, 2014)

This is a typical way of assessing computer skills by paper-and-pencil approach. Students have to draw two primitive solid models on paper, and use Boolean operations (which are included in the curriculum) to illustrate how they could work out the form they need. For instance, students have to draw a cube and a tube on the answer sheet provided. They then need to use the cube and the tube to perform union, difference or intersection function so that the desired object, i.e., the fork end, can be formed. Students have to use correct symbols and diagrams to explain the formation process.

At a glance, the setting of the question is logical, and it is able to assess students' ability in drawing 3D objects. However, this kind of questions neglects the flexibility of applying software to draw and design. The question hinders and limits students using primitive solid models and Boolean operations, and surely there are other methods to draw the same object without knowing what primitive solid models and Boolean operations exactly are. Students who are able to draw proficiently and apply the CAD skills in real settings may be unable to score higher marks in this question. On the contrary, students who are not proficient in CAD may be able to score higher marks. It is questionable whether the question is assessing computer knowledge or pure geometry. Besides, the question and the model answers suggested do not allow other ways of drawings, and restrict the students' thinking into one absolute answer only.

This example of assessment is not preferable, as it fails to assess (1) students' practical application of using CAD computer software, (2) alternative methods of drawing the objects, and (3) other knowledge associated with CAD. Seemingly, the inflexibility and the limitation of this kind of questions had turned CAD assessment into something unrelated to CAD.

Possible Issues and Difficulties

Although the DSE DAT example above focuses on 3D drawings and the knowledge concerned is not very comprehensive, it provides insights into the discussion of the issues and difficulties in CAD assessment. The following discusses the possible issues and difficulties which are also applicable to a general CAD assessment. It is noted that the discussions are also applicable in computer approach, as paper-and-pencil approach is just a form of assessment and the content in this assessment form should be the same as that in the other form of assessment. The issues and difficulties revealed from paper-and-pencil approach should also be arguable in other approach of CAD assessment.

Too Straightforward and Simple

The contents concerned in most CAD assessment is too focused that the presentation of the question is too straightforward and the answer to the question is too simple. The only task that students have to do is to draw or design from a blank sheet or virtual environment. However, the practical application of CAD is much more than just developing objects or programs. It also includes modifying and optimizing any existing designs. Trouble shootings are inevitable in CAD while design is complicated and detailed. It is possible that these elements are lacking in CAD assessment, and questions are unable to assess students' ability to trouble shoot. In order to expand the scope of the questions and the CAD knowledge assessed, it is necessary to reconsider the question proposition.

Obscure Assessment Content

As illustrated in the example in the previous section, sometimes what is being assessed in CAD assessment is not exactly the CAD knowledge but knowledge related to mathematics or geometry. Although CAD needs to be supported by different knowledge, it is necessary to be reminded that CAD, as its name, is about design. The ultimate goal of CAD is to use computer technology and other knowledge to support design. Subsequently, design elements are crucial in CAD assessment. CAD assessment without design elements may not be trustworthy. However, it is easy to fall into the trap of just assessing technical or operational aspect of CAD and ignoring its original intention.

Lack of Flexibility

CAD assessment lacks flexibility of considering different software. This happens often when 3D drawing software is concerned. For instance, some 3D drawing software is able to merge or joint objects together even when they overlap

with other, while some other software can only merge objects when they are perfectly matched. It is difficult for a CAD assessment to address both software, especially when the software is not specified in the paper-and-pencil approach. Although it is possible to allow the flexibility and request students to suggest which software should be used, marking tasks would be very complicated, and it would be difficult to judge how many scores a student can get if the student cannot finish the drawing.

Too Output-Oriented

The previous discussions voice a concern that CAD assessment is often too output-oriented. In other words, whether a student is able to complete the requested task is more important than the process of thinking and designing. The computer approach suffers from obvious inadequacy in this regard. It often devaluates the design and development process and emphasize the objects or program. The issue in paper-and-pencil approach is less serious, as the steps and the development process may be clearly illustrated on paper in the examination. It is easily to assess the process by using the paper-and-pencil approach.

SOLUTIONS AND RECOMMENDATIONS

In view of the above issues and difficulties of CAD assessment, recommendations are suggested as below. Each recommendation corresponds to an issue in the previous section. It is noted that these are not the only recommendations that CAD assessment can be improved. There are other feasible methods, and the followings are some directions which help developing and designing CAD assessment.

Problem Solving

Problem solving plays an important role in design (Liddament, 1996; Siu, 2002, 2008), and provided

that the ultimate goal of CAD is to design, problem solving elements should be included in CAD assessment. Modifying and optimizing designs or programs can be one of the tasks in the assessment. The assessment should not be limited in producing new designs but also rectifying old ones. This is able to assess students' ability of applying CAD in different situations, and this is also a way to identify students who have advanced skills of CAD. Only those with advanced skills are able to solve complicated or problematic situation in the CAD context.

Design-Oriented

As mentioned earlier, one of the major goals in CAD is to design. Therefore, CAD assessment, and also CAD curriculum, should be design-oriented. In other words, the role of computer technology in assisting design should be clearly demonstrated. The tasks in CAD assessment should be situated in a design scenario that the computer technology is a tool to achieve the desired results. It is surely absurd in some occasions that the assessment focuses on design but not the knowledge of computer operation; however, it is also the only way to prevent CAD education from turning into a pure technical and operational curriculum.

Concept-Oriented

CAD assessment should also consider the possibility of different computer software. CAD assessment should be flexible enough to allow students answering differently based on the functions of different software. For instance, in the example of the DSE DAT examination, rather than Boolean operations, the question should allow students to use other ways to represent the process of 3D drawing. The marking scheme should also allow flexibility in model development. This means the assessment should focus on the general concept of using CAD instead of limiting to specific software or a particular way of designing.

Process-Oriented

In order to less focus on the design output of CAD, in CAD assessment it is suggested that more emphasis can be placed on the process of designing. The process of how the problem is solved or how the students approach a problem should also be considered in CAD assessment. It is important to be process-oriented because the designing process itself is also the CAD learning process, and this can motivate students to do better in the learning process. However, this does not mean that the output is unimportant. The quality of the output is still important in CAD assessment, as it still governs whether the output is applicable. The suggestion here only highlights the necessity of considering the process together with the design output.

FUTURE RESEARCH DIRECTIONS

The recommendations above are yet put into large-scale practice. Therefore, one of the possible future research directions is to implement the CAD assessment which fulfills the recommendations and receive feedback from students and tutors or lecturers. For instance, tutors or lecturers can be asked to design assessment which focuses on designing and problem solving questions, and check or investigate in (1) the feasibility for the tutors or lecturers to design a proper assessment based on the requirement, (2) the time they needed for designing such assessment, (3) students' performance in this assessment, and (4) verbal feedback from students and tutors. It is also possible to compare students' scores in previous assessment with the new assessment to see the differences. The findings should be able to provide insight on how CAD assessment should be improved in the future.

The future research directions of CAD education and assessment should focus on problem solving skills instead of the practical skills. Focusing on soft skill is crucial to develop a more

mature education, and the skills learnt should be transferable so that students cannot only develop their skills but also expand their cognitive abilities (Siu & Chen, 2011; Siu & Wong, 2014). Correspondingly, future research can emphasize the development of cognitive abilities after learning CAD. The cognitive abilities can be assessed through some well-developed tests such as the Cognitive Abilities Test (CogAT). Pre-test or pro-test can be conducted to analyze whether a specific kind of CAD teaching methods or arrangement could develop students' cognitive abilities. Besides, the cognitive performance of a group of students who have learnt CAD and a control group can be compared to explicate the benefits of learning CAD. Although there is no similar empirical research about the relationship between CAD and cognitive ability, PiegI (2005) pinpointed the issues related human cognition and CAD. There is also a number of researches about the relationship between computer and cognition. For instance, several decades ago Clements and Gullo (1984) had already started to study the effects of computer programming on children's cognitive ability.

Another possibility is to investigate the transferability of CAD knowledge. As CAD is supported by knowledge of mathematics, spatial reasoning and solid geometry (Field, 2004), it is possible to correlate students' performance in CAD with the performance in other related field. The results are able to suggest which kind of students the CAD industry needs and provide information how students will perform in CAD in the future. It is exceptionally important while admission or recruitment is concerned, and this helps allocating resources more wisely.

CONCLUSION

The chapter discusses CAD education and the inadequate focus on CAD assessment. It also highlights the issues and difficulties of CAD assessment through reviewing questions related to

CAD in a public examination in Hong Kong. The four issues and difficulties, namely too straightforward and simple, obscure assessment content, lack of flexibility and too output-oriented results in four corresponding recommendations: problem solving, design-oriented, concept-oriented and process oriented. These recommendations are meant to improve the current practices of CAD assessment and encourage CAD tutors and lecturers aiming for a higher standard of CAD education and assessment. These are just some of many possible resolutions and approaches to improve CAD education and its assessment, and there are still many more waiting for discussion and exploration.

Apparently, there is a dilemma in balancing the breadth and depth of the examination content through the typical assessment method. In addition, the typical CAD assessment is unable to assess students with high ability and advanced CAD skills. The issues identified in this chapter are not new, and they have been existing in various CAD examinations – not only those in Hong Kong. Especially in institutes or schools where formal and formative assessments are imperative, the issues are not yet improved. It can be argued that in a certain extent the development of CAD education and its assessment are stagnating. More efforts of investigating the current situation and improving the current practice is crucial for the future development.

ACKNOWLEDGMENT

The authors would like to thank The Hong Kong Polytechnic University for the research support. Wuhan University of Technology provided partial support for the final preparation of the chapter.

REFERENCES

- Abbas, S. A. (1976). C.a.d. in industry. *Computer Aided Design*, 8(4), 274–278.
- Chester, I. (2007). Teaching for CAD expertise. *International Journal of Technology and Design Education*, 17(1), 23–35. doi:10.1007/s10798-006-9015-z
- Clements, D. H., & Gullo, D. F. (1984). Effects of computer programming on young childrens cognition. *Journal of Educational Psychology*, 76(6), 1051–1058. doi:10.1037/0022-0663.76.6.1051
- Dönmez, S. (2013). Computer aided industrial design software selection in industrial product design education at Turkey using expert choice program. *Procedia: Social and Behavioral Sciences*, 106, 682–689. doi:10.1016/j.sbspro.2013.12.078
- Field, D. A. (2004). Education and training for CAD in the auto industry. *Computer Aided Design*, 36(14), 1431–1437. doi:10.1016/j.cad.2003.10.007
- Gül, Ö. (2014). A study on instructional methods used in CAD courses in interior architecture education. *Procedia: Social and Behavioral Sciences*, 174, 1758–1763. doi:10.1016/j.sbspro.2015.01.834
- Hamade, R. F., Artail, H. A., & Jaber, M. Y. (2007). Evaluating the learning process of mechanical CAD students. *Computers & Education*, 49(3), 640–661. doi:10.1016/j.compedu.2005.11.009
- Hong Kong Examinations and Assessment Authority. (2014). *Hong Kong diploma of secondary education examination design and applied technology: Examination report and question papers (with marking schemes)*. Hong Kong: Hong Kong Examinations and Assessment Authority.
- Liddament, T. (1996). Design and problem solving. In *IDATER 1996 Conference*. Retrieved from <https://dspace.lboro.ac.uk/dspace-jspui/bitstream/2134/1485/3/liddament96.pdf>
- Morris, P. (1996). *The Hong Kong School Curriculum: Development, Issues and Policies*. Hong Kong: Hong Kong University Press.

Piegl, L. A. (2005). Ten challenges in computer-aided design. *Computer Aided Design*, 37(4), 461–470. doi:10.1016/j.cad.2004.08.012

Rossignac, J. (2004). Education-driven research in CAD. *Computer Aided Design*, 36(14), 1461–1469. doi:10.1016/j.cad.2003.10.008

Siu, K. W. M. (2002). *Meeting the new needs: Curriculum development and assessment of technology subjects*. In *25th anniversary commemorative album of the Hong Kong Examinations and Assessment Authority* (pp. 48–54). Hong Kong: Hong Kong Examinations and Assessment Authority.

Siu, K. W. M. (2008). Review on the development of design education in Hong Kong: The need to nurture the problem finding capability of design students. *Educational Research Journal*, 23(2), 179–202.

Ye, X., Peng, W., Chen, Z., & Cai, Y. (2004). Today's students, tomorrow's engineers: An industrial perspective on CAD education. *Computer Aided Design*, 36(14), 1451–1460. doi:10.1016/j.cad.2003.11.006

ADDITIONAL READING

Jonassen, D., Strobel, J., & Lee, C. B. (2006). Everyday problem solving in engineering: Lessons for Engineering Educators. *The Journal of Engineering Education*, 95(2), 139–151. doi:10.1002/j.2168-9830.2006.tb00885.x

Liang, J. (2009). Generation of a virtual reality-based automotive driving training system for CAD education. *Computer Applications in Engineering Education*, 17(2), 148–166. doi:10.1002/cae.20178

McMartin, F., McKenna, A., & Youssefi, K. (2000). Scenario assignments as assessment tools for undergraduate engineering education. *IEEE Transactions on Education*, 43(2), 111–119. doi:10.1109/13.848061

Mourtos, N. J. (2012). Defining, teaching, and assessing engineering design skills. *International Journal of Quality Assurance in Engineering and Technology Education*, 2(1), 14–30. doi:10.4018/ijqaete.2012010102

Robertson, B. F., & Radcliffe, D. F. (2009). Impact of CAD tools on creative problem solving in engineering design. *Computer Aided Design*, 41(3), 136–146. doi:10.1016/j.cad.2008.06.007

Russell, D. L., & Haghi, A. K. (2010). *Web-based engineering education: critical design and effective tools*. Hershey, PA: Engineering Science Reference. doi:10.4018/978-1-61520-659-9

Siu, K. W. M. (2009). Quality assurance in engineering education: An all-round perspective. In A. S. Patil & P. J. Gray (Eds.), *Engineering education quality assurance: A global perspective* (pp. 181–189). New York: Springer. doi:10.1007/978-1-4419-0555-0_14

Siu, K. W. M. (2011). The dilemma and experience of international technology education curriculum assessment and its inspiration. *Studies in Foreign Education*, 38(10), 58–62.

Siu, K. W. M., & Chen, X. Y. (2011). The paradigm reconstruction of technology education from the international perspective: Integrating hard-technology and soft-technology in technology education. *Studies in Foreign Education*, 38(7), 83–88.

Siu, K. W. M., & Wong, Y. L. (2014). Soft and hard technologies in technology education. In V. C. X. Wang (Ed.), *Handbook of research on education and technology in a changing society* (pp. 625–638). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-6046-5.ch046

Yang, L. (1987). *An integrated approach in the education and training of mechanical & production engineers in CAD/CAM design and manufacture*. Bangkok: Association of Southeastern Asian Institutions of Higher Learning.

KEY TERMS AND DEFINITIONS

CAD Assessment: The strategies or methods in assessing the knowledge or skills acquired in CAD. In this chapter, it refers to assessment at any levels.

CAD Education: The learning and teaching activities of CAD. This is a collective term including all kinds of education in various learning institutes and schools at different levels.

Computer Approach: The approach of using computer to assess CAD ability of students.

Computer-Aided Design (CAD): The design activities which are aided by computer technology.

These design activities include creation, analysis, modification and evaluation.

Design and Applied Technology: A secondary school subject in the Hong Kong Diploma of Secondary Education. It focuses on problem solving through hands-on design activities.

Paper-and-Pencil Approach: The approach of using paper and pencil to assess CAD ability of students.

Problem Solving: The physical and cognitive activity of solving a problem. In this chapter, it refers to one of the essential roles of design and also CAD.

S

The Challenges of Teaching and Learning Software Programming to Novice Students

Seyed Reza Shahamiri

Manukau Institute of Technology, New Zealand

INTRODUCTION

Software Engineering is an engineering discipline that involves with all aspect of software development that applies engineering approaches in order to deliver high quality software products (Pressman, 2005). One of the important stages of software production is coding in which software blueprints are realized via a programming language; a programming language is a language that is understandable for computers (Somerville, 2007). Table 1 shows the general phases of software development. Coding and programming language skills are required from phase four but knowledge and understanding of programming language is very helpful in previous phases in order to successfully complete a software project.

There has been a dramatic demand increase for software applications that promises a rewarding carrier for those who poses the required skills. With the fast advances in technology and emerging ones like driverless cars, Internet of Things (IoT) (Pandya & Champaneria, 2015), Big Data

(Sharma & Mangat, 2015), Software Defined Networks (Bizanis & Kuipers, 2016) etc. it is expected that more software programmers will be required in the future. As the result of such demands, various online and offline courses to introductory programming have been provided.

While knowing software coding is a very useful skill, it is difficult to learn programming especially at the beginning level since acquisition of complex new knowledge, associated strategies, and practical skills are required (Robins, Rountree, & Rountree, 2003). Software development courses are generally among difficult subjects and have low pass rates; according to (Dehnadi & Bornat, 2006) the fail rate of first programming papers in university computer science programmes can be up to 60 percent.

However, what are the reasons that make learning and teaching programming difficult? Why do students find it so challenging? And, why the success rates of programming classes are amongst the lowest in computer science papers? The next section tries to identify the reasons and issues

Table 1. Generic software development life cycle phases

No	Name	Description
1	<i>Planning</i>	The planning phase is the fundamental process of understanding why an information system should be built and determining how the project team will go about building it.
2	<i>Analysis</i>	The analysis phase answers the questions of who will use the system, what the system will do, and where and when it will be used.
3	<i>Design</i>	Based on the user requirements, planning and the detailed analysis, the new system must be designed i.e. a blueprint of the system is created by designing the technical architecture.
4	<i>Implementation</i>	Actually implementing the designed system; writing software programs using software languages.
5	<i>Testing</i>	Checking whether the implemented software works according to specified requirements; fixing bugs/errors.
6	<i>Maintenance</i>	To ensure that the implemented system is properly functioning as per the requirements.

DOI: 10.4018/978-1-5225-2255-3.ch643

that make learning introductory programming challenging. Next, some teaching and learning guidelines are provided to facilitate some of the identified challenges. Finally, recommendations for future studies are provided. The guidelines provided in this chapter are based on the literature and the author's extensive experience in teaching software programming.

BACKGROUND

This section provides a background of the problem and explains the issues in software programming teaching and learning.

Since software is intangible and cannot be seen or touched, creating software products can be a very complex task in comparison to most of the other engineering products; a software project can fail easily and lead to poor quality and unreliable products (Ammann & Offutt, 2008). On the other hand, software applications play very important and critical roles in modern life because they control vital operations that require attributes such as security, reliability, performance, etc., qualities that are hard to achieve (Spillner, Linz, & Schaefer, 2007).

Because of the intangible nature of software, and since students cannot directly sense what they have created, it can become very complex to them to successfully implement, debug and verify the product. In addition, it is more challenging to teach and learn the introduction of coding since this is where the students understand the very basics of software nature and expose to entities and concepts that are generally new to them. They also need to design algorithms and implement them via programming language tools and commands in order to perform the required software functionalities.

Although teaching and learning a programming language may be seen as teaching human languages such as English, it is much more complicated. Teaching a human language is mainly concerned about the language syntax, learning the vocabularies, and how to put them in to use.

This part is similar to teaching programming language as there are syntax and vocabularies (i.e. commands) with specific meanings in that particular language. The main difference is, in teaching human languages we show the students the tools to express their thought and feelings in order to communicate, but we do not teach them what to say. To put it differently, we all know what we want to say, typically, but we may not know how to do it in a particular language. On the other hand, with programming languages we need to teach our students what to ask a computer together with how to ask it, and we need to be very specific while following the language syntax and steps of the algorithms thoroughly as computers have zero tolerance in regards to syntax errors. In other words, computer programs do not run even with a single syntax error.

Furthermore, there is a difference between coding knowledge and strategies. The former provides a declarative nature of programming, such as being able to state how an algorithm works, but the last deals with the practical applications of the knowledge, such as how and where to apply the algorithm (Davies, 1993). Programmers need to think outside the box and be creative in applying and creating knowledge in order to formulate a software solution.

Essentially learning programming requires the learners to develop their problem solving skills. As mentioned earlier, the first step of software development is to understand the problem domain, analyze it, and then to propose a solution including design concepts and algorithms to perform the required functionalities. The next step is to relaying commands to computer systems using a programming language in order to implement the designs and algorithms. The challenge is, regardless of how experience the programmer is, there will be various errors either due to syntax mismatch or incorrect semantic, and they need to deal with unexpected program behaviors frequently. Although modern Integrated Development Environments (IDEs) perform analysis on the code to highlight most of the syntax errors and some obvious compile

time errors, in-depth code inspection and analysis are still required to identify and solve the error in most situations specially run-time errors. To put it differently, learning programming involves different types of cognitive activities and mental representations. They are related to designing programs, understanding them, modifying the code, and debugging them. Construction of conceptual knowledge and structuring of basic and advance programming commands are required even at the lower levels.

During the development the students may need to use a variety of other software applications for graphic and multimedia design, note taking, database development, software design, etc., in addition to the IDE being used. The range of these kind of tools can be from familiar applications like Microsoft Word and PowerPoint to more specialized software like Adobe Photoshop. Teaching students how to use these applications is another issue.

The above challenges are mostly connected causing a shock to novice programmers. The fact that they need to deal with all these difficulties at once may discourage them; they may ultimately give up.

According to (Hubert Lederer Dreyfus, 1986) there are five stages to be an expert programmer namely novice, advanced beginner, competence, proficiency, and finally expert. Winslow (Winslow, 1996) explains a novice programmer as:

- Their abilities to organize knowledge is limited and superficial,
- Their mental model is abstract,
- They are incapable of applying relevant knowledge,
- Their approach to programming is line-by-line instead of understanding the structure.

Experience software engineers already develop the skills to act independently or use other resources, such as programming online communities, to find a fix for the errors they face during software development. Nonetheless, novice programmers

highly depend on an expert to advise them- an experience lecturer who can constantly monitor and observe the students may significantly help them to accomplish the first steps until they are skilled enough to be independent.

SOLUTIONS AND RECOMMENDATIONS

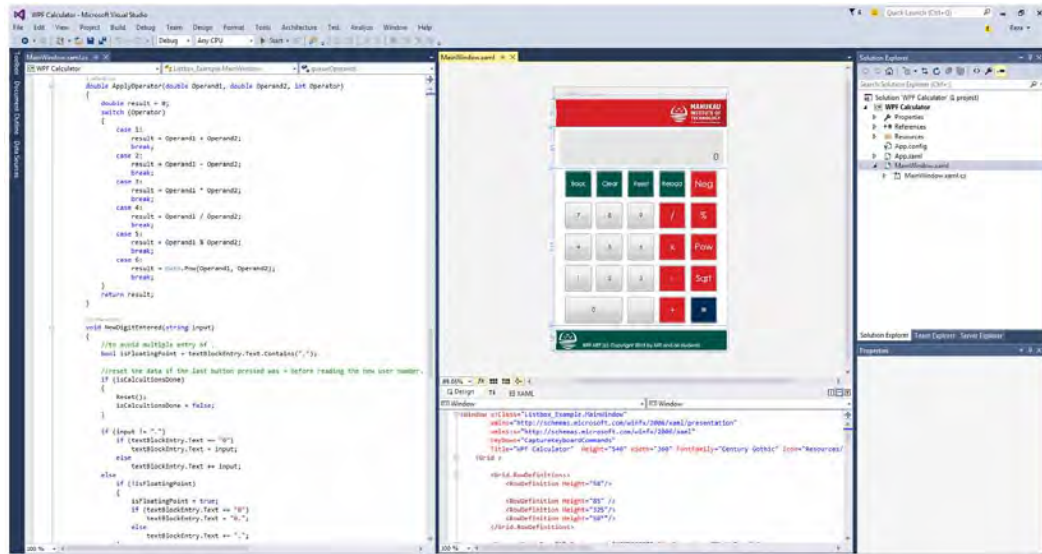
This section tries to answer the question of how to improve the programming classes success rate.

Most students give up early and loose interest when they face the complexities involve in learning software coding; hence, the teaching methodology and the method of delivery is critical. In order to avoid disengaging students, lecturers need to adopt techniques and tools that help students to sense and understand what goes inside the computers and to engage more senses than teaching traditional subjects like Science, Arts, etc. Lecturers need to understand the cognitive requirements of teaching and learning programing and study the successful and unsuccessful strategies.

As an educational program, blended delivery deals with the use of digital and online media in teaching and learning that not only helps learners to involve more senses during the learning process but also give them tools to help them with their studies (Sankey, Birch, & Gardiner, 2010). Furthermore, blended delivery gives more freedom to students in having some control over where and when to learn. It is shown that blended delivery models may ease the learning procedure when face-to-face delivery is combined with computer-mediated activities. In addition, multimodal learning can specifically be helpful for lower-achieving students (Rabiner & Sambur, 1975).

Approaches to teach software programming in which blended delivery combined with activity based learning can be very helpful to maximize the effectiveness of teaching programming. Such approaches help students to visualize the software solutions and understand them better. They may also help to make software somehow tangible. The

Figure 1. Microsoft Visual Studio 2015 IDE



delivery medium is also very important. There are a variety of techniques that can, and must be, considered so that students benefit most from the classes. For example, combining PowerPoints and slides (or other ways of visualizations) with lecturer explanations can be effective to show the main parts of the algorithm being implemented and to provide visualizations of how the algorithm works. Using different colors to highlight key points in the slides and different graphics to explain processes and summarizing content are helpful because they can help students to both hear and see the content and capture information and key points so that more senses are engaged. They can also be used as a reference when students perform self-study, and when face to face lecture is not possible.

Traditionally lecturers use a mixture of techniques that include the lecturer explaining the concepts, and then provide exercises in order to emphasis on learning by doing. Blended delivery can be well integrated with this method of teaching.

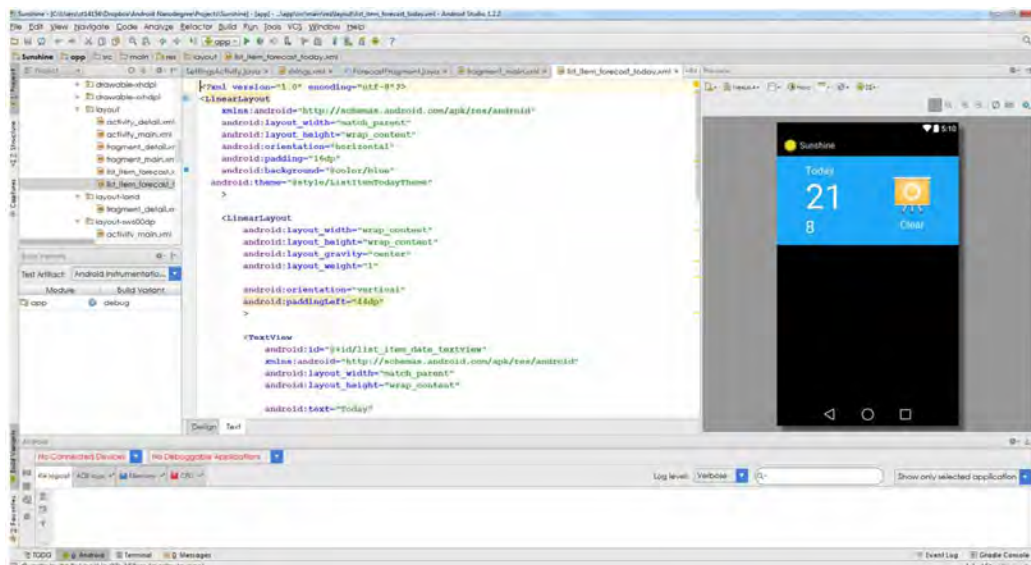
Students should be encouraged to use an IDE to formulate a software solution. IDE itself is a software solution that provides a platform for development by providing tools to help the programmers plan, organize, implement, verify,

and refactor coding projects. Some IDEs are very user friendly and convenient as they provide all the tools required for software development and documentations of how to use them. Nonetheless, there are only few IDEs for some of the programming languages, such as C#.Net, that provide all the required tools; hence, there are limitations in regards to the programming language being though. Current examples of well-known IDEs are Microsoft Visual Studio, Android Studio, NetBeans, and Eclipse.

Figure 1 shows a snapshot of Microsoft Visual Studio that is the primary IDE to develop software applications for Windows based platforms and figure 2 is a snapshot of Android Studio to develop applications for Android platforms. Teaching how to effectively employ tools provided by an IDE must be a part of teaching any programming paper, which can be complicated as they are complex software solutions themselves.

It is important that software programmers be independent and capable of taking initiative. Lecturers need to coach students and help them to search for solutions for their programming problems. Skills to find answers from various sources like textbooks, online communities, online video providers, etc. are crucial.

Figure 2. Android Studio IDE



Programming for different platforms and different types of software, like embedded software, data analytics software, mobile platforms, web services, etc. requires additional understandings of the platforms in addition to knowing the programming and problem solving basics. These basics of programming are generally the same regardless of the software purpose and platform specifically for introduction to programming courses. Once students understand the fundamentals of programming and the way to instruct computers and formulate the instructions, higher level programming courses may concentrate more on programming for specific platforms.

Recently there are new methods of teaching programming widely being used on the Internet in which one-way communication techniques are used to teach software programming. One of the most popular methods is using video lectures that capture the lecturer teaching coding while showing and explaining the code. Students are expected to watch the videos and follow the instructions to learn the subject; lessons are usually followed by related exercises and projects. There are hundreds of hours of video on teaching different program-

ming topics for various programming languages and platforms on YouTube. Some online education providers, like Udacity.com and Lynda.com, employ such techniques as the main method of delivering the lessons. Using such methods are helpful as supplementary materials in conjunction to face-to-face delivery.

FUTURE RESEARCH DIRECTIONS

The best approach to learn programming for novice students is practice. Future works need to study approaches in which lecturers provide many programming exercises for students and monitor the students' performance by completing the exercises in order to identify the gaps in their programming skills. They can then put more emphasis on the identified gaps by designing specific lessons and exercises. Providing resources to point to additional exercises is also helpful.

An important skill that every programmer must well develop is "learn how to learn". Programmers often face coding challenges that are new to them. This is where their research and problem

solving skills play a key role as they need to use available resources to extract information and train themselves to learn the necessary skills in order to solve the coding challenge. Hence, lecturers must consider exercises to develop such skills.

In addition, techniques like pair programming can be effective in which two students work on a tasks together while explaining and debating their understanding of the task, and helping each other to design algorithms and implement them.

CONCLUSION

As software applications play a vital role in our daily life, there is a need to have more skilled programmers to create such software products. There are various emerging disciplines like Internet of Things, driverless cars, Data Science, Software Defined Networks, etc. that demand more programmers in the near future hence a more promising carrier for software developers is expected. Nevertheless, we have seen a low success rate in programming classes where some students lose interest to learn the required skills as they find programming a software a very challenging task; it has been reported that the fail rate of first programming papers in university computer science programmes can be up to 60 percent. This chapter looked at some of the issues in regards to teaching and learning software programming and the nature of programming that may negatively influence the students' attention. Remedies to tackle the issues were also provided emphasizing on blended delivery using the technologies to facilitate the learning, and techniques that encourage more practice by coding exercises. Special attention on developing "learn how to learn" skills is also discussed.

REFERENCES

- Ammann, P., & Offutt, J. (2008). *Introduction To Software Testing* (1st ed.). New York: Cambridge University Press. doi:10.1017/CBO9780511809163
- Bizanis, N., & Kuipers, F. (2016). SDN and virtualization solutions for the Internet of Things: A survey. *IEEE Access*. doi:10.1109/ACCESS.2016.2607786
- Davies, S. P. (1993). Models and theories of programming strategy. *International Journal of Man-Machine Studies*, 39(2), 237–267. doi:10.1006/imms.1993.1061
- Dehnadi, S., & Bornat, R. (2006). *The camel has two humps*. Middlesex University.
- Pandya, H. B., & Champaneria, T. A. (2015). *Internet of things: Survey and case studies*. Paper presented at the 2015 International Conference on Electrical, Electronics, Signals, Communication and Optimization (EESCO). doi:10.1109/EESCO.2015.7253713
- Pressman, R. J. (2005). *Software Engineering: A Practitioners Approach* (6th ed.). McGraw-Hill.
- Rabiner, L. R., & Sambur, M. R. (1975). An Algorithm for Determining the Endpoints of Isolated Utterances. *The Bell System Technical Journal*, 54(2), 297–315. doi:10.1002/j.1538-7305.1975.tb02840.x
- Robins, A., Rountree, J., & Rountree, N. (2003). Learning and Teaching Programming: A Review and Discussion. *Computer Science Education*, 13(2), 137–172. doi:10.1076/csed.13.2.137.14200
- Sankey, M., Birch, D., & Gardiner, M. (2010). *Engaging students through multimodal learning environments: The journey continues*. Paper presented at the Proceedings ascilite, Sydney, Australia.

Sharma, S., & Mangat, V. (2015). *Technology and Trends to Handle Big Data: Survey*. Paper presented at the 2015 Fifth International Conference on Advanced Computing & Communication Technologies. doi:10.1109/ACCT.2015.121

Sommerville, I. (2007). *Software Engineering* (8th ed.). Addison-Wesley.

Spillner, A., Linz, T., & Schaefer, H. (2007). *Software Testing Foundations* (2nd ed.). Santa Barbara, CA: Rocky Nook Inc.

Winslow, L. E. (1996). Programming pedagogy; a psychological overview. *SIGCSE Bulletin*, 28(3), 17–22. doi:10.1145/234867.234872

ADDITIONAL READING

Beck, L. L. (1996). *System Software: An Introduction to Systems Programming*. Pearson.

Farrell, J. (2014). *Programming Logic and Design, Comprehensive* (8th ed.). Cengage Learning.

McConnell, S. (2004). *Code Complete: A Practical Handbook of Software Construction* (2nd ed.). Microsoft Press.

Pressman, R. J. (2014). *Software Engineering: A Practitioners Approach* (8th ed.). McGraw-Hill.

Sankey, M., Birch, D., & Gardiner, M. (2010). *Engaging students through multimodal learning environments: The journey continues*. Paper presented at the Proceedings ascilite, Sydney, Australia.

Sonmez, J. (2014). *Soft Skills: The software developer's life manual*. Manning Publications.

Weinberg, G. M. (1972). *Psychology of Computer Programming*. Van Nostrand Reinhold.

Winslow, L. E. (1996). Programming pedagogy; a psychological overview. *SIGCSE Bulletin*, 28(3), 17–22. doi:10.1145/234867.234872

KEY TERMS AND DEFINITIONS

Blended Delivery: Deals with the use of digital and online media in teaching and learning that helps learners to involve more senses during the learning process and also give them tools to help them with their studies.

Integrated Development Environment: A software solution that provides a platform for software development by providing tools to help the programmers plan, organize, implement, verify, and refactor coding projects.

Introductory to Programming: The initial phase of teaching and learning software programming in which the concentration is on coding basics, problem solving basics, programming language syntax, and primary computer language tools and concept such as variables, loops, decision making, logical and arithmetic operators, etc.

Programming Language: A language that is understandable for computers.

Software: A part of a computer-based system that provides instructions for hardware.

Software Development Life Cycle: Generic steps of constructing software products and the life cycle of a typical software.

Software Engineering: An engineering discipline that involves with all aspect of software development that applies engineering approaches in order to deliver high quality software products.

Developing a Glossary for Software Projects

S

Tamer Abdou*Suez Canal University, Egypt***Pankaj Kamthan***Concordia University, Canada***Nazlie Shahmir***WestJet Airlines Limited, Canada*

INTRODUCTION

There is increasingly significant role played by software in society. This has led to attention on the practices of developing and maintaining software that aim to be successful for all the stakeholders involved.

The discipline of software engineering advocates a systematic and disciplined approach towards the development and evolution of software systems. There is a domain underlying every software project. For a software project to be successful, it is imperative that the domain knowledge be understood and communicated properly to all the stakeholders of that project (Schneider, 2009). Indeed, lack of adequate understanding of the domain has been cited as one of the reasons for software project failures (Kliem, 2007).

The purpose of this chapter is to draw attention to one of the initial steps towards building shareable domain knowledge for a software project, namely a glossary. In doing so, it aims to be relevant to professional as well as pedagogical settings.

The rest of the chapter is organized as follows. First, a motivation and background on glossary are presented, and discussion of relevant previous work is outlined. This is followed by details of a process for developing a glossary. Next, directions for future research are highlighted. Finally, concluding remarks are given.

BACKGROUND

In this chapter, the terms ‘software project’ and ‘project’ are considered synonymous, unless otherwise stated. The term ‘project’ is used to emphasize the fact that the notion of glossary is applicable to a variety of projects, including, but not limited to, software projects. A software project may be about development or about maintenance of a software product.

The following definitions are essential for the rest of the chapter. A *domain* is an area of interest (or the universe of discourse). A *glossary* is a list of terms in a particular domain of knowledge with the definitions for those terms. A *stakeholder* is an individual, group, and/or organization, having an interest in a project.

Glossary in Context

The history of use of glossary in software projects goes back to mid-to-late-1960s, and is therefore is almost as old as the discipline of software engineering itself.

A glossary is similar to, but different from, a dictionary, lexicon, and thesaurus. A comparison can be made using the criteria of goal and scope.

Goal

A glossary, like a dictionary, presents its terms (and corresponding definitions) in a lexicographical (alphabetical) order. Also, a glossary, like a thesaurus, may include synonyms of its terms, but does not include antonyms of any terms. For example, in a *Glossary of Requirements Engineering Terminology* (Glinz, 2014), bug, defect, and fault are considered synonymous. However, unlike a lexicon, a glossary usually does not point to etymology of a term.

Scope

A glossary is specific to the scope of a project, while dictionary, lexicon, and thesaurus are more general in scope as implied by the type of information they include.

Motivation for a Glossary

There are a number of (not necessarily mutually exclusive) reasons for having a glossary for any project.

Support for Knowledge Engineering

It has been acknowledged for some time that software engineering is a knowledge-intensive discipline (Robillard, 1999). For example, there is acquisition of knowledge during all stages of software development. The development of a glossary is an initial step towards knowledge engineering (Schneider, 2009).

The purpose of a glossary is to help make implicit domain expert knowledge *explicit* (Dalkir, 2011), help avoid incurring *analysis debt* due to inconsistent terminology, and to ensure that the knowledge of the domain underlying a software project be communicated properly to all the stakeholders of that project.

Communication Among Stakeholders

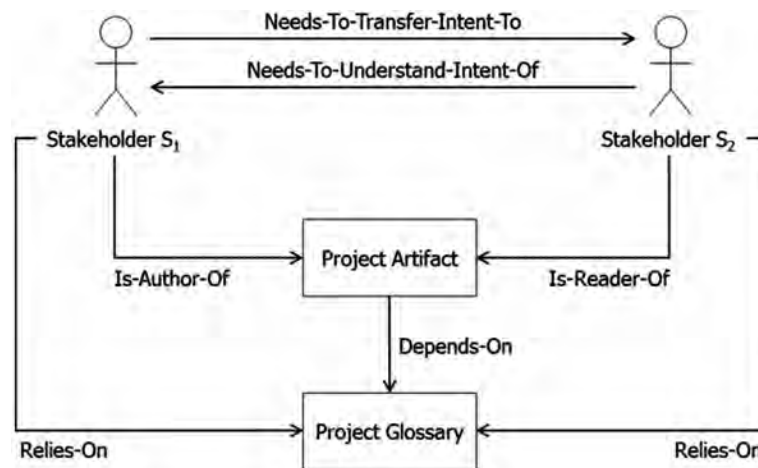
It has been pointed out in a number of contexts that the quality of communication among stakeholders is crucial for the success of a project (Kliem, 2007). There are several dimensions of the quality of communication, one of which is common understanding of the domain underlying the project. It is therefore crucial to create a placeholder of commonly-agreed upon domain knowledge (Kovitz, 1999).

A project-related discourse, whether expressed verbally or in writing, consists of sentences in some language (usually, natural language). A sentence can include terms that may or may not be understood universally. A glossary makes a discourse meaningful by ensuring that all the stakeholders involved in the discourse are on the “same page”. A glossary is a placeholder for commonly-agreed terms (Gottesdiener, 2005).

Figure 1 illustrates the pairwise relationships that can arise between stakeholders and a project artifact, between stakeholders, and the role of a glossary in enabling these relationships.

The need for a glossary is especially acute in organizations where, historically, there has been compartmentalization of departments with more conflict than cooperation between them, but effort is being made to change the trend. For example, *DevOps* is a recent initiative towards a better alignment and working relationship between *development* and *operations* (Hüttermann, 2012). To accomplish that, DevOps extends lessons learned from agile software development and aims to foster close collaboration, culture of shared goals and values, and commitment to measurement and monitoring (Lwakatare, Kuvaja, & Oivo, 2015). For these to manifest successfully, the presence of a common language of communication is a necessary requisite, which a glossary can help provide.

Figure 1. A glossary helps communicate the intent of an author of a project artifact to the reader of that artifact



Reduction of Semantic Heterogeneity in Natural Language

In coming across a term, a stakeholder either (1) is aware of the meaning of the term in one context, but not necessarily other contexts, or (2) is unaware of the meaning of the term. For example, (1) can occur for old terms (that is, terms in existence) and (2) can occur for new terms.

There are certain linguistic structures, for instance, homonyms, metonyms, neologisms, polysemes, portmanteaus, and synonyms, in natural languages such as English. These structures enrich a language, but are also create semantic heterogeneity, which can lead to misunderstandings by stakeholders in different ways.

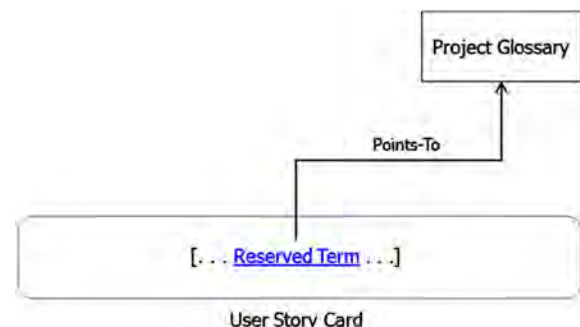
For example, the meanings of the term ‘task’ in *Interaction Design* and in *Scrum*, an agile software development methodology, are different. It has been the experience of one of the authors that even those who are working in industry and specializing in an area (such as, quality assurance) use certain terms with closely-related meanings (such as, ‘smoke test’ and ‘sanity test’) synonymously. The meanings of the same term across different standards can also vary. It is difficult to make meaningful comparisons between entities if the meanings of the two entities are different.

In many agile software development methodologies (Highsmith, 2009), software requirements are expressed as user stories. Figure 2 shows that a reserved term in a user story could point to the project glossary to avoid potential ambiguity and/or to provide clarity to a reader.

Inherent Human Limitations

The software engineers working on the same project may come from different educational backgrounds. These engineers may also be unaware of variations in each other’s backgrounds. This is especially the case for members of teams that are geographically distributed.

Figure 2. A reserved term in a user story card points to the glossary for disambiguation and/or clarification



It is not uncommon for software engineers to work on multiple projects at the same time. In some cases, software engineers are hired on a contract in the middle of a project, and need to familiarize themselves with the domain knowledge. If a software system is aimed for maintenance, then the maintenance team may not be identical to the development team, and may need to consult appropriate documentation for certain activities, such as reengineering, deemed necessary.

For many projects, especially large-scale projects with non-trivial domains, it is unrealistic to expect from every project stakeholder to know, or be able to (correctly) recall, the purpose and meaning of all necessary domain-specific terms, all the time.

In such cases, a glossary can be used as a reference.

Stakeholders of a Glossary

There are a variety of stakeholders for a given project. A glossary, as a *boundary object*, may be used by any stakeholder of a project, and for different reasons.

For example, a project manager may rely on a glossary to support an executive summary; a requirements engineer may use a glossary for domain-specific terms in requirements or otherwise; a designer may use a glossary for defining terms from object-oriented design, user interface design, or other terms in the solution domain; a programmer may use a glossary before deciding upon the style of constant and variable names in a program; a tester may use a glossary to get clarity on a known concept before writing test cases; a non-technical stakeholder (say, a representative user in participatory design) may use a glossary to understand a concept before a meeting; and so on.

Related Work

For the past couple of decades, there have been concerns raised about a lack of a consistent terminology in software engineering (Fenton &

Pfleeger, 1997). This issue, unfortunately, has not been resolved years later, and there are no apparent signs that it will be resolved any time soon.

For example, it has been pointed out (Grimstad, Jørgensen, & Moløkken-Østvold, 2006) that the term ‘effort estimate’ has not been properly used in a number of textbooks and research articles, and it has been pointed out in (Maedche, Botzenhardt, & Neer, 2012) that “a major potential terminological pitfall needs to be addressed [...] user experience (UX) design, user interface (UI) design, graphical user interface (GUI) design, user-centered design (UCD), interaction design (IxD) and additional terms and acronyms are used interchangeably [...] leading to additional confusion”.

The significance of glossary has been acknowledged broadly and explicitly in different contexts in software engineering literature. The significant role that a glossary can play in a software project has been highlighted in form of the *Tip 54 Use a Project Glossary* (Hunt & Thomas, 2000). *Develop a Glossary to Define a Common Language* is one of the recommended best practices in use case modeling (Gottesdiener, 2003). *Define Your Terms* has been proposed as one of the basic principles of writing (Miller, 2004). Finally, *Define a Glossary* is one of the recommended software engineering quality practices (Kandt, 2006). There is native support for glossary in the *Rational Unified Process* (RUP) and the *Open Unified Process* (OpenUP).

The need for a glossary is especially acute in software requirements engineering (Leffingwell, 2011; Wiegers & Beatty, 2013) as it is among the earliest stages in a process where the stakeholders converge. The role of a glossary in reducing lexical ambiguity in software requirements has been emphasized (Kamsties, Berry, & Krieger, 2003). It has been said that “[a requirements engineer] should not make assumptions about the experience or expertise of the reader”, and in stating so, development of a glossary is recommended (Sommerville, 2011). Glossary is part of the *Software Product Certification Model* (SPCM) (Heck, Klabbers, & Eekelen, 2010) and of the *JIT*

Requirements Quality Framework (Heck & Zaidman, 2015). The use of a glossary in contributing towards the traceability of software requirements has been highlighted (Cleland-Huang, Gotel, & Zisman, 2012). In particular, a glossary can serve as a resource for terms in a trace source artifact or in a trace target artifact.

An Exploration of Glossary Engineering

The significance of a glossary for a project calls for a comprehensive approach towards its development.

Requisites for Developing a Glossary

There are certain necessary requisites for developing a glossary properly.

Capabilities of an Author of a Glossary

There can be one or more authors of a glossary. The development of a glossary for a software project requires proper understanding of at least the problem and solutions domains corresponding to the project in question. This, however, can be Catch-22.

It also requires ability in technical writing, as, for example, outlined in the *Software Engineering Professional Practice* (SEPP) Knowledge Area (KA) of the *Guide to the Software Engineering Body of Knowledge* (SWEBOK) or the *Software Engineering Competency Model* (SWECOM). This includes at least a rudimentary background in linguistics.

Experiential Knowledge for Developing a Glossary

The presence of experiential knowledge, that is, knowledge acquired and agreed upon by a collective over time, can be useful for developing a glossary. Indeed, there are guidelines for developing

a glossary (Kovitz, 1999), and there is a pattern named GLOSSARY (Rüping, 2003).

A Process for Developing a Glossary

To be systematic, a process for developing a glossary is proposed and termed GDP for reference. GDP has a number of characteristics:

- **Applicability:** It is independent of any particular application domain, software methodology, or information technology, and is therefore broadly applicable.
- **Correctability:** It is iterative (as suggested by the nonlinearity in the sequence of steps).
- **Feasibility:** It is manual (although it could be partially automated if suitable tools, such as those with support for natural language processing, are available).

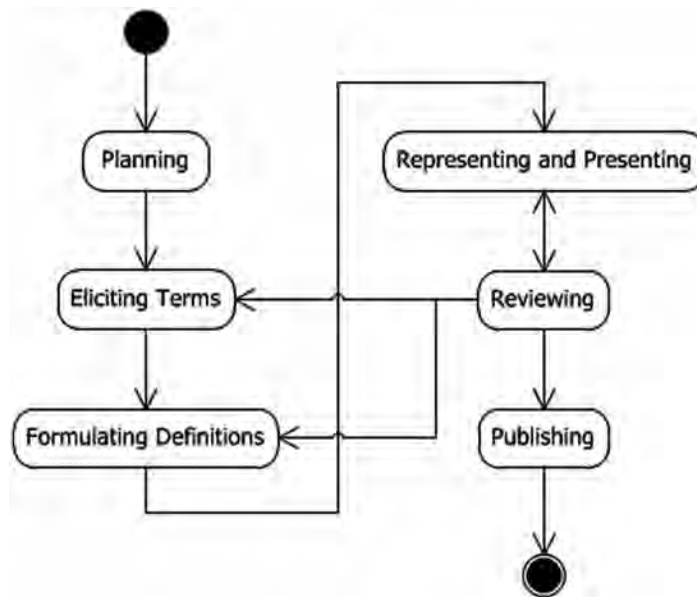
The steps underlying GDP and the interrelationships among the steps are shown in Figure 3, and are described in detail subsequently.

Step 1: Planning

The personnel involved, schedule, and the use of experiential knowledge need to be decided. The candidate personnel may include those who are external to a software project team, as well as those who are internal to a software project team. For example, candidate personnel can be domain experts (also known as subject matter experts) as well as requirements engineers.

The potential for reuse should also be assessed. There could be a number of candidate sources for reuse. For example, there may already exist glossaries in the organization, from past, similar projects, or otherwise. For the *same* software product, the glossary for its development project is a prime candidate for its maintenance project. These other sources could be statement of project vision and documentation on organizational business processes.

Figure 3. A glossary development process



Step 2: Eliciting Terms

The term could be elicited using a number of complementary approaches, including ethnography (interviews or surveys), brainstorming and mind mapping, and past, similar projects.

The terms in the glossary may include those that do not have a universally common interpretation, and those that are newly introduced or at least uncommon. These terms could take the form of abbreviations (including acronyms and initialisms), numeronyms, single word, or a combination of multiple words (phrases).

There needs to be a consensus across the members of a project team on the terms to be included in a glossary. To do that, glossary development should be a collaborative effort.

Step 3: Formulating Definitions

There are different kinds of definitions, of which lexical definitions, operational definitions, and stipulative definitions, are most suited for a glossary. The definition of a term should not be circular.

(For example, this can happen if the definition is tautologous to the term.) The definition should, as far as possible, be self-contained. If possible, multiple definitions of a single term should be avoided. If not, then the context to which each definition applies must be mentioned.

The definitions could be (1) original, (2) variations (adaptations) of existing definitions, or (3) replications of existing definitions. In cases of (2) and (3), there must be a suitable citation.

There needs to be, once again, a consensus across the members of a project team on the definitions to be included in a glossary.

Step 4: Representing and Presenting

The terms and their definitions can be represented (for machine consumption) and presented (for human consumption), in different formats, in different modalities, targeted for different mediums. The order of terms in a glossary is usually lexicographical (alphabetical). The issue of representation and presentation is discussed in more detail later.

Step 5: Reviewing

The necessity and sufficiency of terms, and the definitions of the terms, needs to be inspected, preferably by others not involved in the development of the glossary. This can lead to revisiting one or more previous steps.

Step 6: Publishing

There are a number of means for publishing a glossary. A glossary must be published in a manner that it is available to all relevant stakeholders. Indeed, *Display Models Publicly* is one of the practices of *Agile Modeling* (Ambler, 2002) applicable to the context of a glossary.

Managing the Process for Developing a Glossary

It could be noted that a single-pass, upfront development of a complete glossary can be difficult, as, at inception, required knowledge may not be known, the required expertise may not be present, or new terms may appear during the execution of the project.

Therefore, once an initial version of glossary is developed, it should evolve iteratively and incrementally throughout the project, based on the needs of that project. This means that GDP should start before a software development process, and, subsequently, run in parallel.

In general, a glossary is a ‘living’ document that evolves throughout the life of a project. Therefore, each evolution (increment/iteration) of the glossary should be associated with unique version information.

Indeed, past versions can be useful, especially if they provide information on a term that is deleted in a later version, if they provide information on a certain reference that is removed in a later version, and so on.

Challenges in Developing a Glossary

S

There are a number of challenges to GDP, including the following:

- **Recursion.** The definition of one term may include one or more other terms. For example, it can be expected that the definition of *risk management* will include a number of terms, including *risk*. These other terms themselves may need to be defined. Furthermore, to avoid terminological explosion, there needs to be a stopping criteria.
- **Inconsistency.** The definition of one term should not contradict the definition of another term, at least within the same glossary. It is also important that the definition of a term does not vary drastically across glossaries for different projects, at least across those starting and ending in the same time period. A consistent use of terms is contingent upon long-term stability of those terms. The definition of a term can change over time if new knowledge about the term is discovered.
- **Similarity.** The definitions of two (or more) terms should not be similar to the point that their meanings are difficult to distinguish. For example, the difference between *acceptance test* and *user acceptance test* may not be obvious, unless there is an explicit effort to do so.

Representing a Glossary

There are certain established approaches for representing a glossary, although currently there is no ‘standard’. These approaches rely on the *Extensible Markup Language* (XML).

Figure 4. A partial glossary instance in DocBook 5

```
<glossary>
<title>[Name of Project] Glossary</title>
<!-- ... -->
<glossdiv>
  <title>P</title>
  <glossentry xml:id="pml">
    <glossterm>Pattern Markup Language</glossterm>
    <acronym>PML</acronym>
    <glossdef>
      <para>...</para>
      <glossseealso otherterm="xml"/>
    </glossdef>
  </glossentry>
</glossdiv>
<!-- ... -->
</glossary>
```

Approach 1: DocBook

It is possible to use DocBook (Walsh & Hamilton, 2010). Figure 4 shows a conforming document, in DocBook 5, for a glossary.

The glossseealso element provides relationships among terms in the glossary.

Approach 2: GlossML

It is possible to use the Glossary Markup Language (GlossML). Figure 5 shows a conforming document, in GlossML, for a glossary.

It can be seen in Figure 5 that, by the use of certain attributes, GlossML supports localization and attribution of any resources that are used in the definition of a term.

Presenting a Glossary on the Web

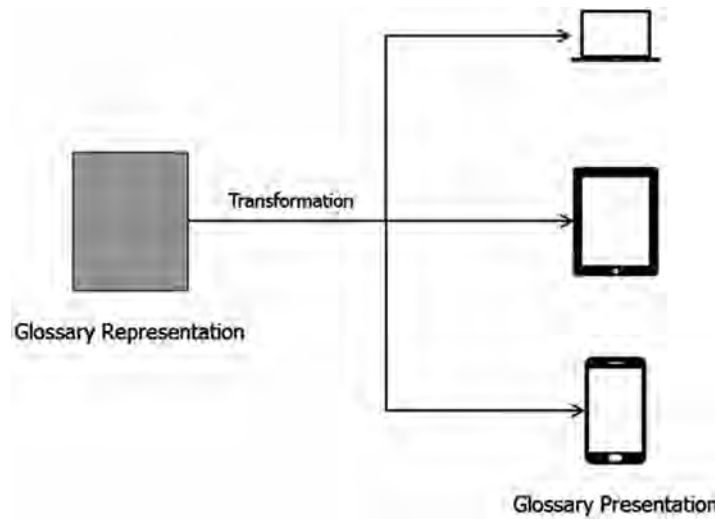
A glossary representation based on one of the aforementioned approaches can be transformed to one or more suitable glossary presentations, as shown in Figure 6.

The set of candidate choices for glossary presentations can vary over time (for example, can be

Figure 5. A partial glossary instance in GlossML

```
<?xml version="1.0" encoding="UTF-8"?>
<glossary version="1.0" srclang="en-US"
  xmlns="http://www.maxprograms.com/gml"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.maxprograms.com/gml GlossML.xsd">
<comment>[Name of Project] Glossary</comment>
<!-- ... -->
<glossentry>
  <langentry xml:lang="en-US">
    <term>...</term>
    <definition source="Merriam Webster">...</definition>
  </langentry>
</glossentry>
<!-- ... -->
</glossary>
```

Figure 6. From a single glossary representation to multiple glossary presentations



influenced by the needs of the stakeholders and the devices deemed relevant for a particular time).

FUTURE RESEARCH DIRECTIONS

There are a few directions for future research that emanate from the work presented in this chapter.

The use of not necessarily textual or not entirely textual modalities for communication in software engineering may necessitate revisitation of the notion of glossary. For example, the use of *semiotic signs*, such as icons or symbols, in requirements engineering to support non-technical stakeholders (Marcus, 2003; Khanom, Heimbürger, & Kärkäinen, 2015) motivates extending the notion of glossary, and modifying the means for representing and presenting it.

The (a) emphasis on, and the use of, a glossary can depend on the (b) culture of an organization involved in software development, in general, and the nature of software development, in particular (Wiegiers, 1996; Hofstede, Hofstede, & Minkov, 2010). For example, communities involved in industrial software development and communities in open source software (OSS) development differ from each other in their style of acquiring

domain knowledge. The (c) impact of incorrect and/or inconsistent domain knowledge on the (d) outcome of a software project can vary considerably. Thus, investigating relationships, such as between (a) and (b), or between (c) and (d), empirically, using a *design science* approach (Johannesson & Perjons, 2014; Wieringa, 2014), for the sake of seeking useful correlations, is of research interest.

The ecosystem of the Social Web includes a variety of applications, including *Wiki*, that enable distributed software engineering. A Wiki has a number of characteristics (Leuf & Cunningham, 2001) that make it a candidate suitable for developing a glossary (and aiding Step 6 of GDP): it is cost-effective, which is especially a consideration for organizations with small budgets; it is possible for people to participate and contribute to the development of a glossary without any particular space or time constraints; there is support for versioning, including reverting back to a previous version if necessary; there is provision for feedback; there are facilities for rich typesetting (such as formatting and hyperlinking); and it is possible for people to access and read the glossary as it develops without any particular constraints on space or time. Therefore, exploring the use of

Wiki for a glossary, in both academia as well as industry, in the realm of agile software projects and otherwise, is another area of research interest.

CONCLUSION

This chapter accentuates the potential uses of a glossary, and contextualizes the challenges in developing a glossary. It presents a systematic approach for developing a glossary, and provides the necessary guidance for doing so.

A glossary does not have any intrinsic value, that is, the value of a glossary does not lie in itself, but that for others (people or products). This has foundational implications towards software engineering education as well as towards software engineering profession. For example, it is crucial that students learning a concept for the first time, or professionals using a concept for the first time, have a clear understanding of that concept and its relation to other concepts in a given context. Indeed, GDP, the glossary development process described in this chapter has implications for *primary learning* (the students learn the significance of a glossary, and learn how to develop and use one properly) as well as *secondary learning* (the students learn how to interact with domain experts, learn how to work collaboratively, and learn how to negotiate and reach consensus in a timely manner), both of which are crucial towards their path of becoming a professional software engineer.

The success of a software project depends intrinsically on effective communication among stakeholders. It has been the experience of the authors that a lack of a proper common ground for terminology continues to be an obstacle towards a transparent communication among software engineers in professional settings, such as those working in industry. There are a number of avenues for creating an environment for an effective communication among stakeholders, and a glossary is one such avenue. A glossary contributes to individual as well as communal understanding of a domain, both of which are necessary for creat-

ing an environment of effective communication among stakeholders.

In conclusion, a glossary is integral to every software project, and therefore developing a glossary is worthwhile.

REFERENCES

- Ambler, S. (2002). *Agile Modeling: Effective Practices for eXtreme Programming and the Unified Process*. Wiley.
- Arabi, A. A. (2008). *Comprehensive Glossary of Telecom Abbreviations and Acronyms*. Auerbach Publications.
- Cleland-Huang, J., Gotel, O., & Zisman, A. (2012). *Software and Systems Traceability*. Springer-Verlag. doi:10.1007/978-1-4471-2239-5
- Dalkir, K. (2011). *Knowledge Management in Theory and Practice* (2nd ed.). The MIT Press.
- Fenton, N. E., & Pfleeger, S. L. (1997). *Software Metrics: A Rigorous and Practical Approach*. International Thomson Computer Press.
- Glinz, M. (2014). *A Glossary of Requirements Engineering Terminology*. International Requirements Engineering Board. IREB.
- Gottesdiener, E. (2003). *Use Cases: Best Practices*. Rational Software White Paper.
- Gottesdiener, E. (2005). *The Software Requirements Memory Jogger: A Pocket Guide to Help Software and Business Teams Develop and Manage Requirements*. GOAL/QPC.
- Grimstad, S., Jørgensen, M., & Moløkken-Østfold, K. (2006). Software Effort Estimation Terminology: The Tower of Babel. *Information and Software Technology*, 48(4), 302–310. doi:10.1016/j.infsof.2005.04.004
- Heck, P., Klabbers, M., & van Eekelen, M. C. J. D. (2010). A Software Product Certification Model. *Software Quality Journal*, 18(1), 37–55. doi:10.1007/s11219-009-9080-0

- Heck, P., & Zaidman, A. (2015). Quality Criteria for Just-in-Time Requirements: Just Enough. *Just-in-Time. The First Workshop on Just-In-Time RE (JITRE 2015)*, Ottawa, Canada.
- Highsmith, J. (2009). *Agile Project Management: Creating Innovative Products*. Addison-Wesley.
- Hofstede, G., Hofstede, G. J., & Minkov, M. (2010). *Cultures and Organizations: Software for the Mind* (3rd ed.). McGraw-Hill.
- Hunt, A., & Thomas, D. (2000). *The Pragmatic Programmer: From Journeyman to Master*. Addison-Wesley.
- Hüttermann, M. (2012). *DevOps for Developers*. Apress. doi:10.1007/978-1-4302-4570-4
- Johannesson, P., & Perjons, E. (2014). *An Introduction to Design Science*. Springer International Publishing. doi:10.1007/978-3-319-10632-8
- Kamsties, E., Berry, D. M., & Krieger, M. M. (2003). *From Contract Drafting to Software Specification: Linguistic Sources of Ambiguity, A Handbook*. Waterloo, Canada: The University of Waterloo.
- Kandt, R. K. (2006). *Software Engineering Quality Practices*. Auerbach Publications.
- Khanom, S., Heimbürger, A., & Kärkkäinen, T. (2015). Can Icons Enhance Requirements Engineering Work? *Journal of Visual Languages and Computing*, 28(3), 147–162. doi:10.1016/j.jvlc.2014.12.011
- Kliem, R. L. (2007). *Effective Communications for Project Management*. Auerbach Publications. doi:10.1201/9781420062489
- Kovitz, B. L. (1999). *Practical Software Requirements: A Manual of Content and Style*. Manning.
- Laplante, P. A. (2012). *Technical Writing: A Practical Guide for Engineers and Scientists*. CRC Press.
- Leffingwell, D. (2011). *Agile Software Requirements: Lean Requirements Practices for Teams, Programs, and the Enterprise*. Addison-Wesley.
- Leuf, B., & Cunningham, W. (2001). *The Wiki Way: Quick Collaboration on the Web*. Addison-Wesley.
- Lwakatare, L. E., Kuvaja, P., & Oivo, M. (2015). *Dimensions of DevOps. The Sixteenth International Conference on Agile Processes, in Software Engineering, and Extreme Programming (XP 2015)*, Helsinki, Finland.
- Maedche, A., Botzenhardt, A., & Neer, L. (2012). *Software for People: Fundamentals, Trends and Best Practices*. Springer-Verlag. doi:10.1007/978-3-642-31371-4
- Marcus, A. (2003). Icons, Symbols, and Signs: Visible Languages to Facilitate Communication. *Interactions*, 10(3), 37–43.
- Miller, J. E. (2004). *The Chicago Guide to Writing about Numbers*. The University of Chicago Press. doi:10.7208/chicago/9780226526324.001.0001
- Robillard, P. N. (1999). The Role of Knowledge in Software Development. *Communications of the ACM*, 42(1), 87–92. doi:10.1145/291469.291476
- Rüping, A. (2003). *Agile Documentation: A Pattern Guide to Producing Lightweight Documents for Software Projects*. Wiley.
- Schneider, K. (2009). *Experience and Knowledge Management in Software Engineering*. Springer-Verlag. doi:10.1007/978-3-540-95880-2
- Sommerville, I. (2011). *Software Engineering*. Pearson Education.
- Walsh, N., & Hamilton, R. L. (2010). *DocBook 5: The Definitive Guide*. O'Reilly Media.
- Wieggers, K. (1996). *Creating a Software Engineering Culture*. Dorset House.

Wiegers, K., & Beatty, J. (2013). *Software Requirements*. Microsoft Press.

Wieringa, R. J. (2014). *Design Science Methodology for Information Systems and Software Engineering*. Springer-Verlag. doi:10.1007/978-3-662-43839-8

ADDITIONAL READING

Anthropy, A., & Clark, N. (2014). *A Game Design Vocabulary: Exploring the Foundational Principles Behind Good Game Design*. Addison-Wesley. Beer, D., & McMurrey, D. (2009). *A Guide to Writing as an Engineer* (3rd ed.). Wiley.

Deissenboeck, F., & Pizka, M. (2006). Concise and Consistent Naming. *Software Quality Journal*, 14(3), 261–282. doi:10.1007/s11219-006-9219-1

DeRespinis, F., Hayward, P., Jenkins, J., Laird, A., McDonald, L., & Radzinski, E. (2012). *The IBM Style Guide: Conventions for Writers and Editors*. IBM Press.

Fenton, N., & Bieman, J. (2015). *Software Metrics: A Rigorous and Practical Approach* (3rd ed.). CRC Press.

Ibrahim, A. M. (1989). Acronyms Observed. *IEEE Transactions on Professional Communication*, 32(1), 27–28. doi:10.1109/47.21857

IEEE. (2009). *IEEE Standard 1044-2009. IEEE Standard Classification for Software Anomalies. The Institute of Electrical and Electronics Engineers (IEEE)*. Computer Society.

Kamsties, E., Hörmann, K., & Schlich, M. (1998). Requirements Engineering in Small and Medium Enterprises. *Requirements Engineering*, 3(2), 84–90. doi:10.1007/BF02919967

Laplane, P. A. (2012). *Technical Writing: A Practical Guide for Engineers and Scientists*. CRC Press.

Rozanski, N., & Woods, E. (2005). *Software Systems Architecture: Working With Stakeholders Using Viewpoints and Perspectives*. Addison-Wesley.

KEY TERMS AND DEFINITIONS

Agile Software Development Methodology:

A software development methodology based on the Agile Manifesto.

Artifact: A document or a model produced during software development.

Process: A set of interrelated or interacting activities that transforms inputs into outputs for some purpose.

Requirement: A statement which translates or expresses a need and its associated constraints and conditions.

Social Web: The perceived evolution of the Web in a direction that is driven by ‘collective intelligence,’ realized by information technology, and characterized by user participation, openness, and network effects.

Software Engineering: The application of a systematic, disciplined, quantifiable approach to the development, operation, and maintenance of software; that is, the application of engineering to software.

User Story: A high-level requirement statement that contains minimally sufficient information to produce a reasonable estimate of the effort to implement it.

Wiki: A Web application developed cooperatively by a community of users, allowing any user to add, delete, or modify information.

Displaying Hidden Information in Glossaries



Marcela Ridao

Universidad Nacional del Centro de la Provincia de Buenos Aires, Argentina

Jorge Horacio Doorn

Universidad Nacional del Oeste, Argentina & Universidad Nacional de La Matanza, Argentina

INTRODUCTION

All models created during the software development process and specifically during the Requirements Engineering Process, are created with structures and purposes clearly defined. These structures have been conceived to maximize the expressiveness of the model for its objective.

Despite this, and very possibly because of this, it may occur that this models hold information no perceptible during its routine use. From an epistemological point of view, this situation is very similar to what happens in data mining (Artz, 2009). For particular case of Requirements Engineering models, which are written in natural language, it has been observed that at least part of this hidden information can be discovered.

In this chapter, a strategy to visualize the grouping of terms in a glossary called Language Extended Lexicon (LEL) is proposed. Experiments performed on real world cases have shown that the clusters obtained using syntactic resources of the LEL model, coincide with semantic nuclei of the application domain. The proposed strategy is based on graphs constructed using hypertext links embedded in the LEL model.

Disciplines dedicated to the comprehension of phenomena whose dominant aspect is the structural complexity instead of the essential complexity of their components have been acquiring more importance (Barabasi, 2002; Dorogovtsev

& Mendes, 2003). There are many examples, in several disciplines, where the detection of groupings provides a notorious contribution to the comprehension of the phenomenon being studied (Sarmah, Kalita, & Bhattacharyya, 2011; Mo, Cao, & Wang, 2012; Zimmermann, Ntoutsis, Siddiqui, Spiliopoulou, & Kriegel, 2012). Most evident examples of this sort of problems are the organizational networks, social networks, bibliographic references networks and interest groups among many others. The classic visual representation of these networks is done by means of graphs. Actually, it happens that when the number of nodes surpasses a moderated limit, those graphs become not useful to observe the relevant characteristics of the network. Configurations of nodes and connections also occur in a great diversity of other applications. They may represent physical networks, such as electrical circuits, roadways, or organic molecules. They are also used when representing less tangible interactions as might occur in ecosystems, sociological relationships, databases, or in the flow of control in a computer program (Gross & Yellen, 2003, 2006; Kamada & Kawai, 1989).

The requirements prioritization method proposed by Duan, Laurent, Cleland-Huang, & Kwiatkowski (2009), analyzes documents of the Requirements Engineering process by means of detecting clusters of requirements. In Duan proposal, the requirements are included in clusters by

means of an iterative algorithm. Those clusters are object of negotiations among the stakeholders to assign priorities. Requirements inherit the priority assigned to the cluster to which they belong.

Some of the models of the Requirements Engineering may be studied from the structural point of view. Particularly, one of the most promising is the LEL (Leite & Franco, 1990). This model records the vocabulary used in the application domain. It describes the words and phrases used by clients and users with a meaning specific to the application domain.

Observing the LEL from a structural point of view, it may be immediately seen that it can be depicted using a graph where the vertices symbolize the LEL entries and the edges represent the hypertext links among entries.

From this point of view, the LEL may be visualized as a sort of linguistic network with a complex structure; in such a way that besides the explicit information stored in each symbol, there is implicit information stored in the relationships among symbols. To build and to analyze the graph of the symbols of the LEL is a sort of knowledge mining.

The most basic knowledge acquirable from the LEL graph is the existence of clusters of entries, which become roots for any taxonomy of the business process. Intuitively it may be said that if the application domain is divided into areas of interest or organization fragments, then it may be expected some coupling among the terms used in such area.

This chapter describes a mechanism to build the graph of any LEL, offering a good visualization of the existing symbol clusters.

The following section analyses the Language Extended Lexicon (LEL), a glossary built during Requirements Engineering process. Then, a section describing force directed algorithms applied to graph visualization is included. After that, the use of these methods in LEL graph construction is studied. And finally, some results, future work and conclusions are presented.

BACKGROUND

LEL: Language Extended Lexicon

The creation of a glossary defining the jargon used in the application domain has been highlighted since many years ago (Arango, 1994; Leite & Franco, 1990). Moreover, several experiences along decades have shown that these glossaries are, by themselves, an important source of information to acquire knowledge about the application domain. (Ben Achour, Rolland, Maiden, & Souveyet, 1999; Oberg, Probasco, & Ericsson, 1998; Prakash, Aurum, & Kox, 2004; Regnell, 1999; Rolland & Ben Achour, 1998). This fact is very well known by linguistics, since they have established that language reflects environment and technology. Just a few examples of that are i) Arabic which has 80 words for camel, ii) Japanese having more than 20 words for rice and iii) Inuit with more than 20 words for snow and ice. It is not needed to explain the importance of camel, rice, snow and ice for those peoples. (Nettle & Romaine, 2000).

The successful use of glossaries by actual practitioners has been reported for more than two decades (Weidenhaupt, Pohl, Jarke, & Haumer, 1998). The authors of this article have a library of more than 200 real cases glossaries developed by undergraduate students, graduate students, practitioners and themselves. The oldest cases were developed as early as 1990 (Leite & Franco, 1990; Rivero, Doorn, del Fresno, Mauco, Ridao, M., & Leonardi, 1998). Some of the most relevant are:

- A reflexive application of the requirement engineering process to the requirement engineering process itself (García & Gentile, 1999).
- Acquisition, administration and maintenance of ATM equipments, 2003.
- Quality Control of Medicaments Supplies, 2007.

- Integrated Administration System of a Private University, 2010.
- A new version of García and Gentile (1999) actually under development.

The Language Extended Lexicon is a meta-model designed to help the elicitation and representation of the language used in the application context (Leite & Franco 1993; Leite, Doorn, Kaplan, Hadad, & Ridao, 2004; Leite et al., 1997). LEL is a glossary proposed by Leite (1997). This model is centered on the idea that a closed description of language terms improves the comprehension of the Universe of Discourse. It is a natural language representation that aims to capture the vocabulary of an application.

The Lexicon main goal is to register symbols (words or phrases) which are peculiar to the application domain. Each entry in the lexicon is identified by a name or names (case of synonyms) and has two types of descriptions, as opposed to the usual dictionary, which provides only one.

The first, called Notion, is the usual one and its goal is to describe the denotation (defines “what the symbol is”) of the word or the phrase. The second, called Behavioral Response, is intended to describe the connotation (describes “how the symbol acts in the system”) of the word or the phrase, providing extra information about the context at hand. Entries are classified in four types according to its general use in the UoD. The types are: Subject, Object, Verb and State. Figure 1 presents the Language Extended Lexicon Model.

While describing the symbols, two principles have to be followed: the *principle of circularity* that intends to maximize the use of symbols in the description of other symbols and the *principle of minimal vocabulary* that intends to minimize the use of symbols that are external to the lexicon. These external symbols must belong to a small subset of a natural language dictionary. These rules stress the description of the vocabulary as a self-contained and highly connected hypertext (Leite & Franco, 1993).

Figure 1. The Language Extended Lexicon Model

```

LEL: representation of the symbols in the application domain language.
Syntax:
    {Symbol}1N

Symbol: entry of the lexicon that has a special meaning in the application domain.
Syntax:
    {Name}1N + {Notion}1N + {Behavioral Response}1N

Name: identification of the symbol. More than one represents synonyms.
Syntax:
    Word | Phrase

Notion: denotation of the symbol. It must be expressed using references to other symbols and using a minimal vocabulary.
Syntax:
    Sentence

Behavioral Response: connotation of the symbol. It must be expressed using references to other symbols and using a minimal vocabulary.
Syntax:
    Sentence

where Sentence is composed only by Symbols and Non-Symbols, the latter belonging to the minimal vocabulary.
+ means composition, {x} means zero or more occurrences of x, () is used for grouping, | stands for or and [x] denotes that x is optional.

```


Automatic Graph Construction

A graph $G = (V, E)$ is a set V of vertices and a set E of edges, in which an edge joins a pair of vertices. Normally, graphs are depicted with their vertices as points in a plane and their edges as line or curve segments connecting those points. There are different styles of representation, suited to different types of graphs or different purposes of presentation. When drawing graphs, some generally accepted aesthetic criteria are used (Fruchterman & Reingold, 1991):

- Distribute the vertices evenly in the frame.
- Minimize edge crossings.
- Make edge lengths uniform.
- Reflect inherent symmetry.
- Conform to the frame.

However, these types of criteria do not contribute to the objective of visualizing the underlying structure of the analyzed problem.

Notoriously effective graphs may be obtained using automatic Drawing Graph methods. Drawing Graphs is an area of mathematics and computer science that combines methods from geometric graph theory and information visualization to derive two-dimensional depictions (Di Battista, Eades, Tamassia, & Tollis, 1999; Kaufmann & Wagner, 2001).

There are many techniques to draw graph. Most of them are devoted to a specific type of graph such as trees, directed graphs and others. A family of techniques to visualize unspecific graphs is known as *force-directed*. Also known as *spring embedders* (see Figure 2), such algorithms calculate the layout of a graph using only information contained within the structure of the graph itself, rather than

relying on domain-specific knowledge. Graphs drawn with these algorithms tend to be aesthetically pleasing, exhibit symmetries, and usually produce crossing-free layouts for planar graphs (Kobourov, 2013). These methods are flexible, easy to implement and provide excellent results (Walshaw, 2003). This sort of methods is based on physics analogies. Their main characteristics are:

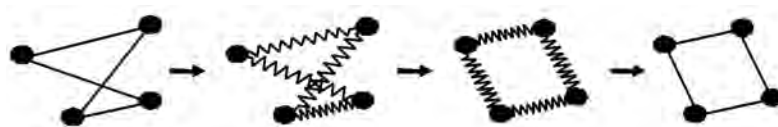
- They model the graph as a mechanical system.
- The graph is obtained equilibrating the mechanical system.

These characteristics define the two main components of the force-directed methods:

- A mechanical model of the graph representing the elements of the graph (vertices and edges), perhaps adding some aesthetic restriction.
- An algorithm to find the equilibrium state of the mechanical system.

Nodes are placed in an arbitrary initial disposition, and the forces move them to its final equilibrium position. The spring layout method of Eades (1984) and the algorithm of Fruchterman and Reingold (1991) both rely on spring forces, similar to those in mechanical laws. In these methods, there are repulsive forces between all nodes, but also attractive forces between nodes which are adjacent. Actually, these repulsive forces come from atomic models where nucleons have strong repulsive forces which prevent the collapse of the nucleus. In the Fruchterman y Reingold algorithm, only adjacent nodes attract themselves, while all nodes have repulsive forces.

Figure 2. Spring Embedder (Adapted from Eades, 1984)



LEL GRAPH CONSTRUCTION USING DIRECTED FORCES

The Fruchterman and Reingold (1991) algorithm was applied to graphs where nodes represent LEL symbols and edges represent the hypertext links among them. Minor changes to the algorithm were done; specifically the forces involved were modified.

The Forces

Used forces are:

$$fa(d) = c1 * \log\left(\frac{d}{c2}\right) \quad (1)$$

$$fr(d) = c3 * \sqrt{d} \quad (2)$$

Figure 3. Force-directed algorithm (Adapted from Fruchterman & Reingold, 1991)

```

area := W * L; { W and L are the width and length of the frame }
G := (V, E); { the vertices are assigned random initial positions }
function fa(delta) := begin return (c1 * log(delta/c2)) end ;
function fr(delta) := begin return (c3 * sqrt(delta)) end ;
for i := 1 to iterations do begin
{ calculate repulsive forces }
for v in V do begin
{ each vertex has two vectors: .pos and .disp }
v.disp := 0;
for u in V do
if (u ≠ v) then begin
{ Δ is short hand for the difference }
{ vector between the positions of the two vertices }
Δ := v.pos - u.pos;
v.disp := v.disp + (Δ/|Δ|) * fr(|Δ|)
end
end
{ calculate attractive forces }
for e in E do begin
{ each edge is an ordered pair of vertices .v and .u }
Δ := e.v.pos - e.u.pos
e.v.disp := e.v.disp - (Δ/|Δ|) * fa(|Δ|);
e.u.disp := e.u.disp + (Δ/|Δ|) * fa(|Δ|)
end
{ prevent from being displaced outside frame }
for v in V do begin
v.pos := v.pos + ( v. disp/ |v.disp|) * v.disp;
v.pos.x := min(W/2, max(-W/2, v.pos.x));
v.pos.y := min(L/2, max(-L/2, v.pos.y))
end
end

```

Where f_a denotes the attraction force and f_r the repulsive force; d represents the distance among nodes.

The constants c_1 , c_2 and c_3 were experimentally adjusted to get a good visualization of the graph (Ridao & Doorn, 2015).

The Frame

The graph must be contained within a predefined frame. The Fruchterman & Reingold algorithm puts many unmovable artificial nodes on the perimeter of the frame with no attractive forces. In this way the frame is modeled as four walls that enclose the graph within them.

Choosing appropriated values for the constants c_1 , c_2 and c_3 , has shown that such wall is completely unnecessary. Furthermore, the wall introduces unrealistic deformation on the parts of the graphs near the wall. This deformation disappears when such wall is removed.

SOLUTIONS AND RECOMMENDATIONS

Some Paradigmatic Results

Graphs were created from LELs of actual organizations for several study cases. Every one of these cases was taken from a repository of LELs ensuring that no one of the authors of this work have been involved in its creations. All of them allowed to clearly see existing clusters. Due to their relevance, three of these cases are briefly summarized in this section.

An additional precaution was taken. Each graph was created many times departing from different initial configurations. The observed result was that for each case studied the final graphs were always notoriously similar.

Case 1: No Previously Known Clusters

In this case, the graph was created with no previous review of the LEL, to ensure that there was no previous knowledge about the existence or not of clusters in the graph.

Figure 4, depicts the obtained graph. The edges were removed from the figure to improve the visualization of nodes location. The graph shows only one cluster. The subsequent semantic analysis confirmed this conclusion.

Symbols with less importance: 24, 62, 65, 32 and 44 appear in the graph in peripheral locations.

Case 2: Two Clusters Previously Known

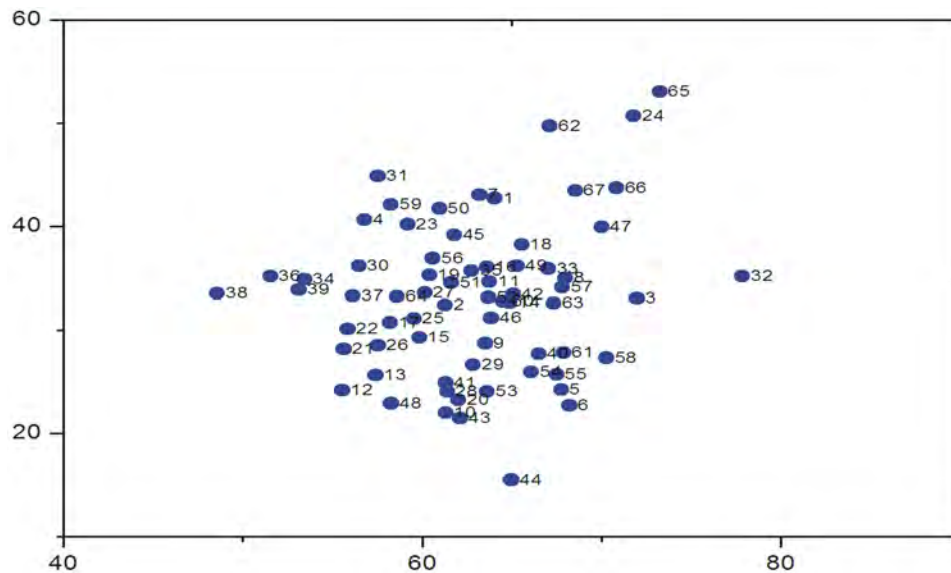
In this case, the graph was created using a LEL which was previously studied to verify if the algorithm will confirm or not what was already known. Figure 5 presents the node distribution for this case. This Figure clearly shows the existence of two clusters. Furthermore, all the nodes are placed within the cluster to which it was supposed they belong. A few nodes interconnecting both clusters are located in the central area of the graph. This is the case of nodes 11, 50, 75, 76, 77 and 78.

Case 3: Two Clusters Previously Known

This case had the same treatment given to case 2; the graph was created using a LEL which was previously studied. This study allowed detecting the existence of two clusters. Surprisingly, the graph clearly depicted three clusters, as seen in Figure 6. To confirm this result, a semantic review was done, and it became obvious that one of the previously known clusters was made up by two different semantic nucleuses. This is the most important result since here the algorithm helped to detect invisible information.

A closer look to the graph shows that connecting nodes are located in intermediate areas close to the clusters they are connecting. Also some less important symbols are placed away from the center of the cluster to which they belong.

Figure 4. Resulting graph for case 1



Use of Indicators to Determine the Number of Clusters

At a first glance, it can be seen that there are geometric differences among the three examples above described. For the case 1 and 2 there are some symmetry of the node distribution over both

axis X and Y. On the other hand, in case 3 this symmetry does not exist.

Symmetry coefficients are:

- Case 1:

$$AS_X = 0.02 \quad AS_Y = 0.14$$

Figure 5. Resulting graph for case 2

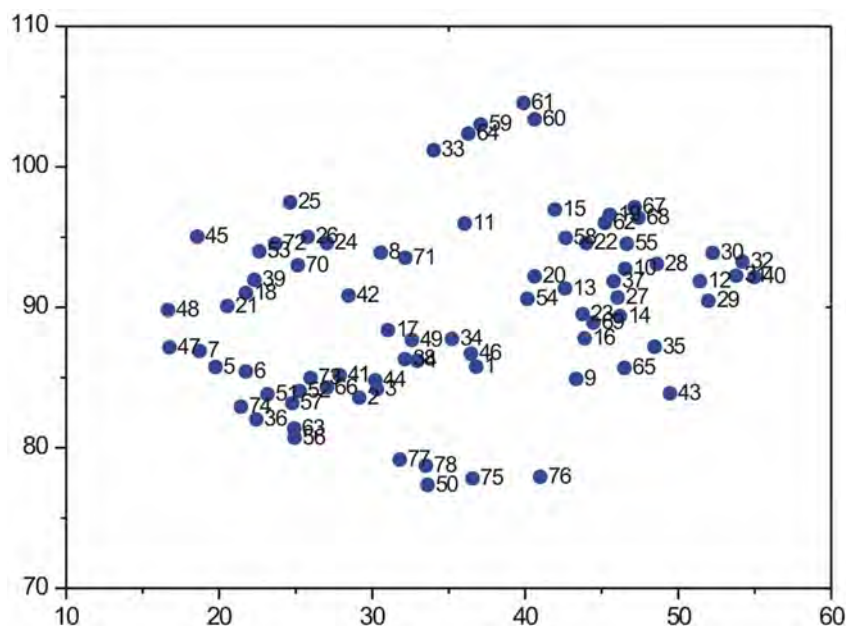
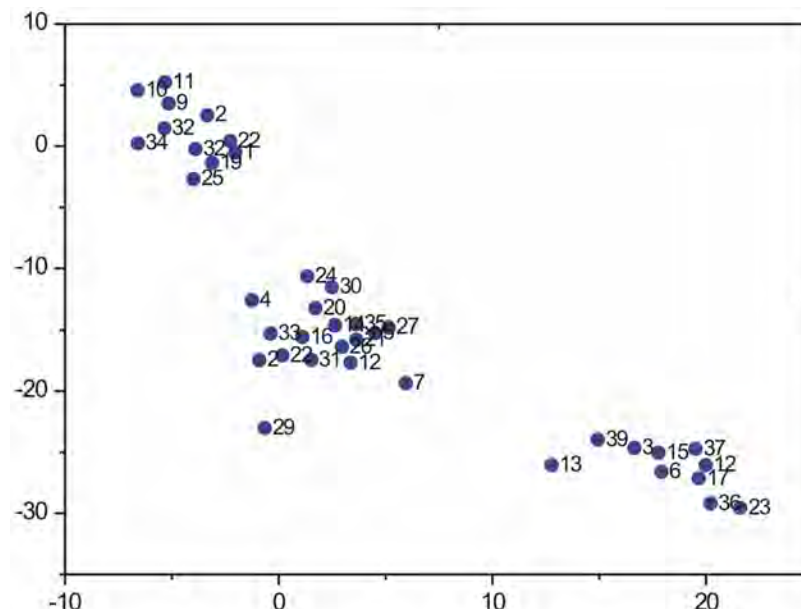


Figure 6. Resulting graph for case 3



- Case 2:

$$AS_X = 0.05 \quad AS_Y = 0.14$$

- Case 3:

$$AS_X = 0.79 \quad AS_Y = -0.48$$

When symmetry coefficients have values rather different from 0, for at least one axis, it is very likely that the LEL under study would have 3 or more clusters. More experiments are needed to infer the limit that may make reliable this distinction.

In the case 2, the nodes are placed in an ellipsoidal pattern, suggesting the existence of two clusters. Vertices belonging to each grouping are closer to one of the focal points of the ellipse. The quotient of the variance of the coordinates distribution for each axis in this case is:

$$CV = \frac{\sigma_X^2}{\sigma_Y^2} = 3,18$$

Case 1 has a variance quotient near to 1. Again, not enough experiments have been done to infer a limit that may allow distinguishing case 1 and case 2.

FUTURE RESEARCH DIRECTIONS

In the near future, several improvements and extensions will be included in the current algorithm used for graph construction. They may be sorted as:

- **Increment in the Number of Dimensions:**
The graphs already constructed have been restricted to two dimensions due to visualization restrictions and to the cultural influence of the traditional use of graph. It has been observed, in just a few cases, that the presence of more than three clusters leads to confusing graphs. Graph used for visualization of real world problems have no reasons to be tied to two dimensions. However, three or more dimensions will bring the necessity of improved visualization techniques.

- **Change in the Visualization Techniques:**
At a first glance, there are two possible ways to improve visualization:
 - Projecting dynamically the n dimension graph over arbitrary planes.
 - Changing automatically the size, color and shape of the nodes and edges using attributes extracted from the actual origin document.
- **Applying the Technique to Other Models:**
Such as Current Scenarios, Future Scenarios and Software Requirements Specifications documents. These models are capable for being visualized using graphs.

It is also necessary a systematic comparison of this cluster construction technique with other proposals, such as Duan et al. (2009).

CONCLUSION

A strategy that helps to detect symbol groupings in the LEL has been proposed. Every one of the detected cluster in all studied cases coincide with existing functional groups in the application domain. This grouping motivates a second reading of the LEL, making easier to acquire more knowledge about the application domain.

One of the most important and unexpected features of this tool is that sometimes the discovered cluster were unknown by the LEL authors. This becomes an effective knowledge discovering feature.

The automatic construction of graph seems to be a very useful tool to visualize part of the implicit information embedded in Natural Language Requirements Engineering documents. This has been proven on glossaries and it is expected to be also true for other Requirement Engineering Process documents.

The obtained results are easily combined with several generic cluster detection strategies, which require a previous knowledge of the number of clusters; such number is one of the results of the automatic graph construction.

REFERENCES

- Arango, G. (1994). Domain analysis methods. In W. Schafer, R. Prieto-Díaz, & M. Matsumoto (Eds.), *Software reusability* (pp. 152–159). Hemel, UK: Ellis Horwood Ltd.
- Artz, J. M. (2009). Data driven vs. metric driven warehouse design. In J. Wang (Ed.), *Encyclopedia of Data Warehousing and Mining* (pp. 382–387). Hershey, PA: IGI Global. doi:10.4018/978-1-60566-010-3.ch060
- Barabasi, A. (2002). *Linked, the new science of network*. New York, NY: Perseus publishing.
- Ben Achour, C., Rolland, C., Maiden, N., & Souveyet, C. (1999). Guiding use case authoring: results of an empirical study. In *Proceedings of International Symposium on Requirements Engineering (RE'99)* (pp. 36–43). Limerick, Ireland: IEEE Computer Society Press. doi:10.1109/ISRE.1999.777983
- Di Battista, G., Eades, P., Tamassia, R., & Tollis, I. (1999). *Graph drawing: Algorithms for the visualization of graphs*. Harlow, UK: Prentice Hall.
- Dorogovtsev, S., & Mendes, J. (2003). *Evolution of networks: From biological nets to the Internet and WWW*. Oxford, UK: Oxford University Press. doi:10.1093/acprof:oso/9780198515906.001.0001
- Duan, C., Laurent, P., Cleland-Huang, J., & Kwiatkowski, C. (2009). Towards automated requirements prioritization and triage. *Requirements Engineering Journal*, 14(2), 73–89. doi:10.1007/s00766-009-0079-7
- Eades, P. (1984). A heuristic for graph drawing. In *Proceedings of Congressus Numerantium 42*. Winnipeg, Canada: Utilitas Mathematica Pub. Inc.
- Fruchterman, T., & Reingold, E. (1991). Graph drawing by force-directed placement. *Software, Practice & Experience*, 21(11), 1129–1164. doi:10.1002/spe.4380211102

- Garcia, O., & Gentile, C. (1999). *Diseño de una herramienta para construcción de LEL y Escenarios* (Graduation dissertation). Universidad Nacional del Centro de la Provincia de Buenos Aires, Argentina.
- Gross, J., & Yellen, J. (Eds.). (2003). *Handbook of graph theory*. Boca Raton, FL: CRC Press. doi:10.1201/9780203490204
- Gross, J., & Yellen, J. (Eds.). (2006). *Graph theory and its applications, (Discrete mathematics and its applications)* (2nd ed.). London, UK: Chapman & Hall/CRC.
- Kamada, T., & Kawai, S. (1989). An algorithm for drawing general undirected graphs. *Information Processing Letters*, 31(1), 7–15. doi:10.1016/0020-0190(89)90102-6
- Kaufmann, M., & Wagner, D. (Eds.). (2001). *Drawing graphs: Methods and models*, LNCS (Vol. 2025). Springer-Verlag. doi:10.1007/3-540-44969-8
- Kobourov, S. (2013). Force-directed drawing algorithms. In R. Tamassia (Ed.), *Handbook of Graph Drawing and Visualization* (pp. 383–408). Boca Ratón, FL: CRC Press.
- Leite, J., Doorn, J., Kaplan, K., Hadad, G., & Ridao, M. (2004). Defining system context using scenarios. In J. Leite & J. Doorn (Eds.), *Perspectives on Software Requirements* (pp. 169–199). Dordrecht, the Netherlands: Kluwer Academic Press. doi:10.1007/978-1-4615-0465-8_8
- Leite, J., & Franco, A. (1990). O uso de hipertexto na elicitação de linguagens da aplicação. In *Proceedings of IV Simpósio Brasileiro de Engenharia de Software (SBES)*.
- Leite, J., & Franco, A. (1993). A strategy for conceptual model acquisition. In *Proceedings of the IEEE International Symposium on Requirements Engineering (RE'93)* (pp. 243–246). San Diego, CA: IEEE Computer Society Press.
- Leite, J., Rossi, G., Balaguer, F., Maiorana, V., Kaplan, G., Hadad, G., & Oliveros, A. (1997). Enhancing a requirements baseline with scenarios. *Requirements Engineering Journal*, 2(4), 184–198. doi:10.1007/BF02745371
- Mo, Y., Cao, Z., & Wang, B. (2012). Occurrence-based fingerprint clustering for fast pattern-matching location determination. *IEEE Communications Letters*, 16(12), 2012–2015. doi:10.1109/LCOMM.2012.111412.121909
- Nettle, D., & Romaine, S. (2000). *Vanishing voices: The extinction of the world's languages*. Oxford, UK: Oxford University Press.
- Oberg, R., Probasco, L., & Ericsson, M. (1998). *Applying requirements management with use cases*. Rational Software White Paper. Cupertino, CA: Rational Software Corporation.
- Prakash, S., Aurum, A., & Kox, K. (2004). Requirements engineering practice in pharmaceutical and healthcare manufacturing. In *Proceedings of the 11th Asia-Pacific Software Engineering Conference* (pp. 402–409). Washington, DC: IEEE Computer Society. doi:10.1109/APSEC.2004.86
- Regnell, B. (1999). *Requirements engineering with use cases – a basis for software development* (Doctoral dissertation). Lund University, Sweden.
- Ridao, M., & Doorn, J. (2015). Agrupamientos en glosarios del universo de discurso. *Journal Tecnología y Ciencia de la Universidad Tecnológica Nacional*, 13(27), 5–16.
- Rivero, L., Doorn, J., del Fresno, M., Mauco, V., Ridao, M., & Leonardi, C. (1998). Una estrategia de análisis orientada a objetos basada en escenarios: Aplicación en un caso real. In *Proceedings of WER'98 - Workshop en Engenharia de Requisitos* (pp. 79–88). Maringá-PR, Brasil.
- Rolland, C., & Ben Achour, C. (1998). Guiding the construction of textual use case specifications. *Data & Knowledge Engineering*, 25(1–2), 125–160. doi:10.1016/S0169-023X(97)86223-4

Sarmah, R., Kalita, J., & Bhattacharyya, D. K. (2011). A pattern matching approach for clustering gene expression data. *International Journal on Data Mining, Modelling and Management*, 3(2), 130–149.

Walshaw, C. (2003). A multilevel algorithm for force-directed graph-drawing. *Journal of Graph Algorithms and Applications*, 7(3), 253–285. doi:10.7155/jgaa.00070

Weidenhaupt, K., Pohl, K., Jarke, M., & Haumer, P. (1998). Scenarios in system development: Current practice. *IEEE Software*, 15(2), 34–45. doi:10.1109/52.663783

Zimmermann, M., Ntoutsis, I., Siddiqui, Z., Spiliopoulou, M., & Kriegel, H. P. (2012). Discovering global and local bursts in a stream of news. In *Proceedings of the 27th Annual ACM Symposium on Applied Computing (SAC '12)*, (pp. 807–812). ACM Press. doi:10.1145/2245276.2245433

KEY TERMS AND DEFINITIONS

Application Domain: The context in which the software system will be used.

Drawing Graphs Methods: An area of mathematics and computer science that combines methods from geometric graph theory and information visualization to derive two-dimensional graph depictions.

Force-Directed Algorithm: A method for calculating a graph layout. It calculates the layout using only information contained within the structure of the graph itself, rather than relying on domain-specific knowledge.

Language Extended Lexicon: A glossary composed by a set of symbols, words or phrases which are peculiar and frequently used in a given application domain. The symbols included in a LEL have a different meaning than the regular use. It contains hypertext links that interconnect symbols.

LEL Graph: A graph where the vertices are the LEL entries and the edges represent the hypertext links among entries.

Requirements Engineering: An area of the Software Engineering which is responsible for acquiring and defining the needs of the software system.

Semantic Cluster: A group of terms in the Language Extended Lexicon, representing an area of interest in the business process.

Dynamic Situational Adaptation of a Requirements Engineering Process

Graciela Dora Susana Hadad

Universidad Nacional del Oeste, Argentina & Universidad de Belgrano, Argentina

Jorge Horacio Doorn

Universidad Nacional del Oeste, Argentina & Universidad Nacional de La Matanza, Argentina

Viviana Alejandra Ledesma

Universidad Nacional del Oeste, Argentina & Universidad Nacional de La Matanza, Argentina

INTRODUCTION

The adaptation of any process to particular situations is considered a good practice in most fields. Literature shows that this practice is quite common in Software Engineering processes, such as the methodologies Rational Method Composer (Haumer, 2005) and OPEN Process Framework (Firesmith & Henderson-Sellers, 2002). However, Requirements Engineering (RE) approaches are seldom tailored to context or project situations (Potts, 1995; Leite, Hadad, Doorn, & Kaplan, 2000; Leffingwell & Widrig, 2003; Seyff et al., 2009). Nevertheless, sometimes the elicitation activity, as part of an RE process, is performed taking into account some environmental characteristics, such as number of information sources, user geographical distribution, user time availability, user experience, among others (Maiden & Rugg, 1996; Hickey & Davis, 2003; Coulin, 2007; Carrizo, Dieste, & Juristo, 2008). Recently, some proposals have appeared to design an RE process for a specific project by selecting existent RE techniques (Lauesen, 2002; Lobo & Arthur, 2005; Alexander & Beus-Dukic, 2009).

There are activities of the requirements process that are invariant regardless of situational factors, while others should be modified, removed or replaced. Not only activities may be adapted, models produced or used in the process may be also suited for the situation (Galster, Weyns, To-

fan, Michalik & Avgeriou, 2014). This means that the RE process may be assembled like a flexible puzzle using interchangeable pieces depending on the situational factors identified.

Situational Method Engineering (SME) is advocated to build methods tailored to specific situations for the development of systems (Kumar & Welke, 1992). Following its principles, the adaptation of any software development process is based on indicators describing the situation (Khan, bin Mahrin & Chuprat, 2014). Part of the task is to compose such indicators based on observable factors, like degree of business processes reengineering, context complexity, developer expertise in the application domain, and project size, among others. Ideally, these situational factors should be taken into account before beginning the software process. However, there are factors not accurately known when initiating a software project, while other factors may change during the course of the project. Hence, a dynamic view of the adaptation of a software development process achieves a better effectiveness of the process itself. Considering that the production of requirements is the starting point of a software development, it should be necessary to pay more attention to factors impacting on the RE process.

A frequent question of practitioners is related with the obligation of performing all the process steps to reach requirements. *Is it possible to shorten the road or to follow a different one?* Under some

circumstances, there is an opportunity to reduce the RE process by deleting or simplifying activities; and sometimes different paths may be followed by choosing other techniques or even extending some activities. Project managers should make decisions depending mainly on the particular case.

Since problem domain knowledge is mostly expressed in natural language (NL), the use of an RE approach based on NL representations improves the commitment of customers and users to the project, increasing the probability of project success (Macaulay, 1993).

Therefore, in this chapter it is presented the tailoring of an RE process based on NL models, according to a particular set of situational factors. Additionally, a process for constructing this RE process, including the evaluation of such factors along software development life cycle, is proposed as an enhanced solution.

BACKGROUND

When working on the creation of an engineering product or system, it is important to have a process. This means having a predictable set of activities, techniques, inputs and outputs helps get a high quality outcome. Hence, the way the work is performed does not depend on individual criteria, allowing repeatability of costs, times and quality, and promoting the accumulation of knowledge about the process. As a consequence, the first activity of a process to develop a product consists in defining precisely the expected outcome. When the product is a software system, this initial activity is an RE process, whose outcome is a Software Requirements Specification (SRS). The RE process is particularly different from other activities of the software development process since it is the one that most interacts with people and their environment, while other activities are mainly carried out within the development team context (Carrizo, 2009). Besides, project decisions impose constraints, tools and methods to carry out those activities. Therefore, if the requirements

process takes into account the particularities surrounding the application context and the project itself, then it will probably result both in a better SRS and in a more effective process.

Additionally, a requirements process needs appropriate and continuous communication to gain customers and users compromise. Good communication is achieved when all stakeholders use the same language. In RE, a proven way to accomplish this is by using the vocabulary of the application context (Leite, Doorn, Kaplan, Hadad, & Ridao, 2004). Communication occurs when stakeholders orally interact, and also when reports, documents and models are exhibited to customers. In this sense, NL models, such as glossaries, use cases and scenarios, stimulate stakeholders communication (Leite et al., 2004), and are the most frequently used in RE (Kaindl, 2000; Leffingwell & Widrig, 2003; Seyff et al., 2009; Antonelli, Rossi, Leite, & Oliveros, 2012).

Software development processes put into practice in real projects are often forced to be adjusted due to contingent circumstances, sometimes in a poorly controlled fashion while the project is ongoing. Hence, the evolving situation should be observed as the process goes forward in order to achieve a better tailoring. Possible adjustments to the process can be known in advance on the basis of certain characteristics, though they may change dynamically, i.e., settings are pre-planned but only implemented when an aspect of the situation changes (Rolland, 2008). In this regard, the process is defined as a set of blocks, having process blocks common to all situations and variant process blocks according to situational factors. Thus, the process is made up by assembling blocks for the particular situation (Henderson-Sellers & Ralyté, 2010). Method Engineering was specifically created to tackle this case, promoting the design, construction and adaptation of methods, techniques and tools in order to develop information systems (Brinkemper, 1996). This discipline considers not only the creation of process blocks but also product blocks, and even blocks that assemble both process and product (Rolland, 2008; Henderson-Sellers &

Ralyté, 2010; Ralyté, 2013). Rolland (2008) not only has pointed out that the software development should begin with a definition phase of the method to be used, but also she proposes that the process must be re-adjusted throughout the life cycle process, an aspect not treated by Method Engineering. SME, as a sub-area of Method Engineering, focuses on building methods for software development, tailored to specific situations (Kumar & Welke, 1992). Thus, SME studies the factors affecting the software project and the application context (Bucher, Klesse, Kurpjuweit, & Winter, 2007). Its principles have been applied in RE to define requirements processes that are adaptable to particular circumstances by using existent modular components (Firesmith, 2004; Jafarinezhad & Ramsin, 2012; Bakhat, Sarwar, Motla, & Akhtar, 2015).

Some of the adaptation factors are related to development time constraints. Naturally, they lead to faster ways of developing software; these are steps toward agile methods (Pinheiro, 2002). A quality factor may depend on a time factor, or the other way around. However, a balance equation for both factors may be frequently achieved (Kohler & Paech, 2002; Pinheiro, 2002). This trade-off is more likely to be better faced by an RE process with the necessary tailoring (Kohler & Paech, 2002). Extremely, when time factors are a major concern, agile methods could be the best way. Though, agile practices are frequently assumed as opposite to RE practices (Cockburn, 2002; Beck, 2004), Kovitz (2002) remarks that both, a phased development and an agile one, do RE but in different styles, since requirements are always present. This vision is also shared by (Sillitti & Succi, 2005; Kohler & Paech, 2002). In this sense, an agile approach dealing with requirements may be also seen as a sort of dynamical adaptation of an RE process. As Cockburn (2002) said, making the decision on the right level of agility is the best way to succeed on a project. RE practices do not deny agility, considering that this property must be combined with quality and attending volatility without constant software fix (Pinheiro, 2002).

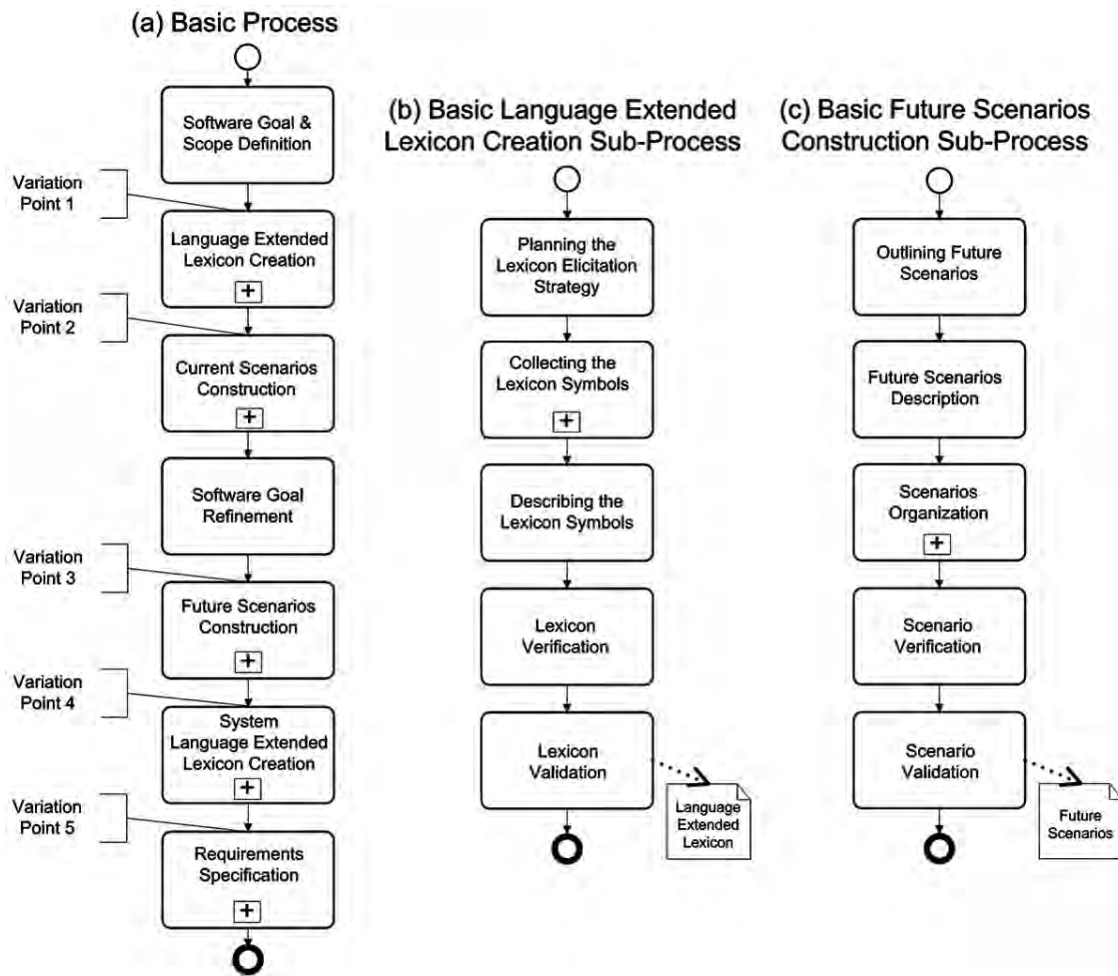
A REQUIREMENTS ENGINEERING PROCESS ADAPTABLE TO SITUATIONAL FACTORS

RE process based on NL models strongly contributes to stakeholders' commitment to the project. It involves the following stages:

- *Defining the general software goal and scope.*
- *Understanding the vocabulary used in the application context*, supported by the Language Extended Lexicon model (Hadad, Doorn, & Kaplan, 2009).
- *Understanding the application context*, supported by a set of current scenarios that represent the situations observed in the application context (Leite et al., 2000).
- *Refining the general goal*, by decomposing it into sub-goals.
- *Defining the software context*, by producing a set of future scenarios that represent situations envisioned in a future application context where the software system will operate (Leite et al., 2004).
- *Defining the vocabulary used in software descriptions*, by producing a Language Extended Lexicon of the System created from the previous one, adding terms used in the descriptions of future scenarios.
- *Making explicit the requirements*, by producing an SRS, after extracting requirements from future scenarios.

This RE process has been initially developed and used in a monolithic fashion, as depicted in Figure 1(a), being independent of any special circumstance. Most main sub-processes are composed of another level of processes, some of which are atomic processes while others are decomposed, as it is shown in Figures 1(b) and 1(c). Recycles in the three diagrams were removed to emphasize the variation points. These recycles imply correcting or updating models as a consequence of a better understanding of

Figure 1. Basic Requirements Engineering Process with two main sub-processes



the problem and of the evolution of the application context.

Though Figure 1(a) shows a monolithic process as part of a cascade model, the strategy can be adapted to be used within different software process models, being this a project factor by itself (see Table 1). For instance, when following an iterative process model, the RE process is performed sequentially for the first four main sub-processes and the last three sub-processes are usually done in iterations along with a requirements change management activity throughout the software development life cycle. It should be noticed that in some cases the fourth sub-process may be included in some of the iterations. Iterations obey to a prioritization activity

within *Software Goal Refinement* sub-process. In an incremental process model, requirements are initially produced, and iterations are based on those defined requirements, obviously allowing requirements evolution. Under this process model, the entire RE process is performed as the starting point, though requirements change management occurs along development life cycle. Regarding agile practices, additionally to time constraints, another leading factor is volatility (Cockburn, 2002), which may be well coped with a change management activity. Though, it may not be the best manner when volatility is a core characteristic of the context. Therefore, an agile method might be a better proposal to satisfy customers in the meanwhile.

This basic RE process may be adapted at the variation points shown in Figure 1(a), depending on a combination of situational factors. Project managers should do it to improve the RE process and, thereby, the requirements product. Situational factors impacting on RE (see Table 1) have been selected based on experience in RE practice and literature proposals (Maiden & Rugg, 1996; Hickey & Davis, 2003; Carrizo et al., 2008; Jafarinezhad & Ramsin, 2012). These factors can be characterized by the following attributes:

- **Value:** Admissible values depend on each factor.
- **Confidence on the Value:** High, Medium, Low.

- **Origin Type:** Application Context or Project.
- **Evolution Type:** Invariant or Contingent.
- **Variation Points:** Where it impacts on.

In order to adapt the RE process, project managers assign values to factors, qualifying them with their degree of confidence on such values. It usually happens that at the very beginning some factors may be not precisely known, since right values are known after gaining a better understanding of the situation. Factors are classified according to their origin into: those related to the specific application context, and those related to the specific software project. From an evolution dimension, factors are classified into: *Invariant*, being those

Table 1. Situational Factors impacting on RE process. Range Very High to Very Low includes five possible values.

Context Factors	Acceptable Values	Evolution	Variation Points
Context Complexity	Very High to Very Low	Invariant	2,3
Target Customer	Tailor-made, Market-driven	Invariant	1,2,3,4
New Business	Yes, No	Invariant	1,2,3,4
Business Process Reengineering	Very High to Very Low	Contingent	3,4
Context Volatility	Very High to Very Low	Invariant	2,3
Volatility of Customer Demands	Very High to Very Low	Contingent	2,5
Users Rotation	Very High to Very Low	Invariant	2,3
Context Inconsistencies	Very High to Very Low	Invariant	2,3
Conflict of Users Interests	Very High to Very Low	Contingent	2,3
Project Factors	Acceptable Values	Evolution	Variation Points
Familiarity with the Domain	Very High to Very Low	Contingent	1,2,3,4
Size of the Project	Very High to Very Low	Contingent	2,3,5
Developers Rotation	Very High to Very Low	Invariant	2,3
Software Quality Required	Very High to Very Low	Invariant	1,2,3,4,5
Reuse existing Requirements Artifacts	Yes, No	Invariant	1,2,3,4,5
Creating Requirements Artifacts for Reuse	Yes, No	Invariant	1,2,3,4,5
Requirements Pre-Traceability	Yes, No	Invariant	1,2,3,4
Requirements Post-Traceability	Yes, No	Invariant	5
Granularity of Requirements Traceability	Individual, Group	Invariant	5
Demand to produce an SRS	Yes, No	Invariant	5
Project Time and Resource Constraints	Very High to Very Low	Contingent	1,2,3,4,5
Software Process Model	Cascade, Incremental, Iterative, Agile	Invariant	1,2,3,4,5

that do not change during the RE process; and *Contingent*, being those that may naturally evolve due to changes in the application context and/or in the project context.

Interaction among factors should be taken into account when tailoring the RE process. *Overriding interaction* indicates that a factor is discarded when another factor takes a specific value. *Limiting interaction* points out that a factor may reduce the range of acceptable values of another factor. *Incompatible interaction* implies that the values of two or more factors cannot be accepted simultaneously at the same project; thereby, the value of at least one factor has to be changed through stakeholders' negotiation to solve the inconsistency. Examples of these interactions are:

- **Overriding Interaction:** The factor Business Process Reengineering is not taken into account if the factor New Business takes the value *Yes*.

- **Limiting Interaction:** The factor Software Process Model cannot take the value *Cascade* if factor Context Volatility is *Very High*.
- **Incompatible Interaction:** The factor Software Quality Required is *Very High* while the factor Project Time and Resource Constraints is *Very Low*.

Each sub-process is defined by a sequence of process blocks, some of which include product models. At each variation point a basic process block may: i) stay as it is; ii) be deleted; iii) be replaced by a particular process block; or iv) be replaced by a process block with a partial internal variation based on parameters. Figure 2, exploiting sub-process of Figure 1(c), shows how the instance of a process should be constructed according to specific situational factors, using a repository containing all possible atomic blocks. This Figure exemplifies two process blocks that may be deleted (*Scenarios Organization* and *Sce-*

Figure 2. Defining the Future Scenarios Construction Sub-Process

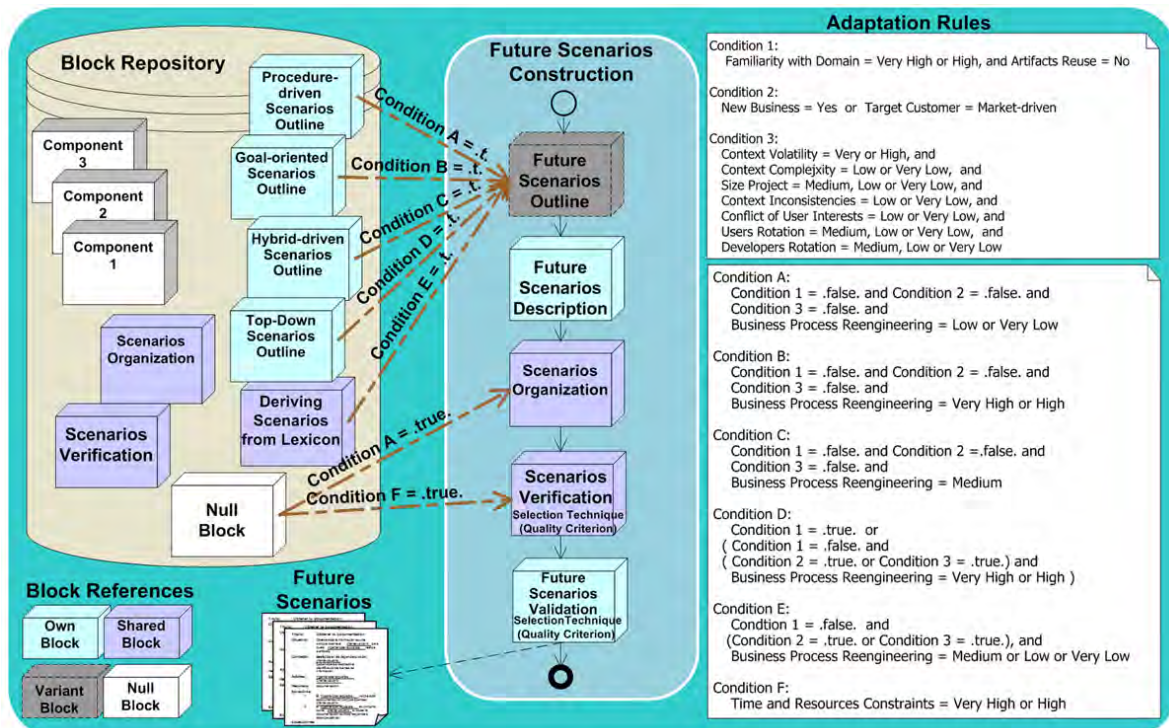


Table 2. Valid operations on atomic process blocks

BLOCK TYPE	OPERATIONS			
	Stay	Delete	Replace	Parameterize
Own	✓	✓		✓
Shared	✓	✓		✓
Variant		✓	✓	

narios Verification), a process block that should be replaced by one of five different atomic processes (*Future Scenarios Outline*), two process blocks that are parameterized (*Scenarios Verification* and *Scenarios Validation*), and a process block that stays independently of any factor (*Future Scenarios Description*).

There are common elements among the various parts of the RE process. For example, Figure 2 shows two process blocks *Scenarios Organization* and *Scenarios Verification* which both are shared with another main sub-process *Current Scenarios Construction*. There are four types of atomic process blocks. An *own block* is that not shared

Figure 3. Instance of the RE Process for a specific situation

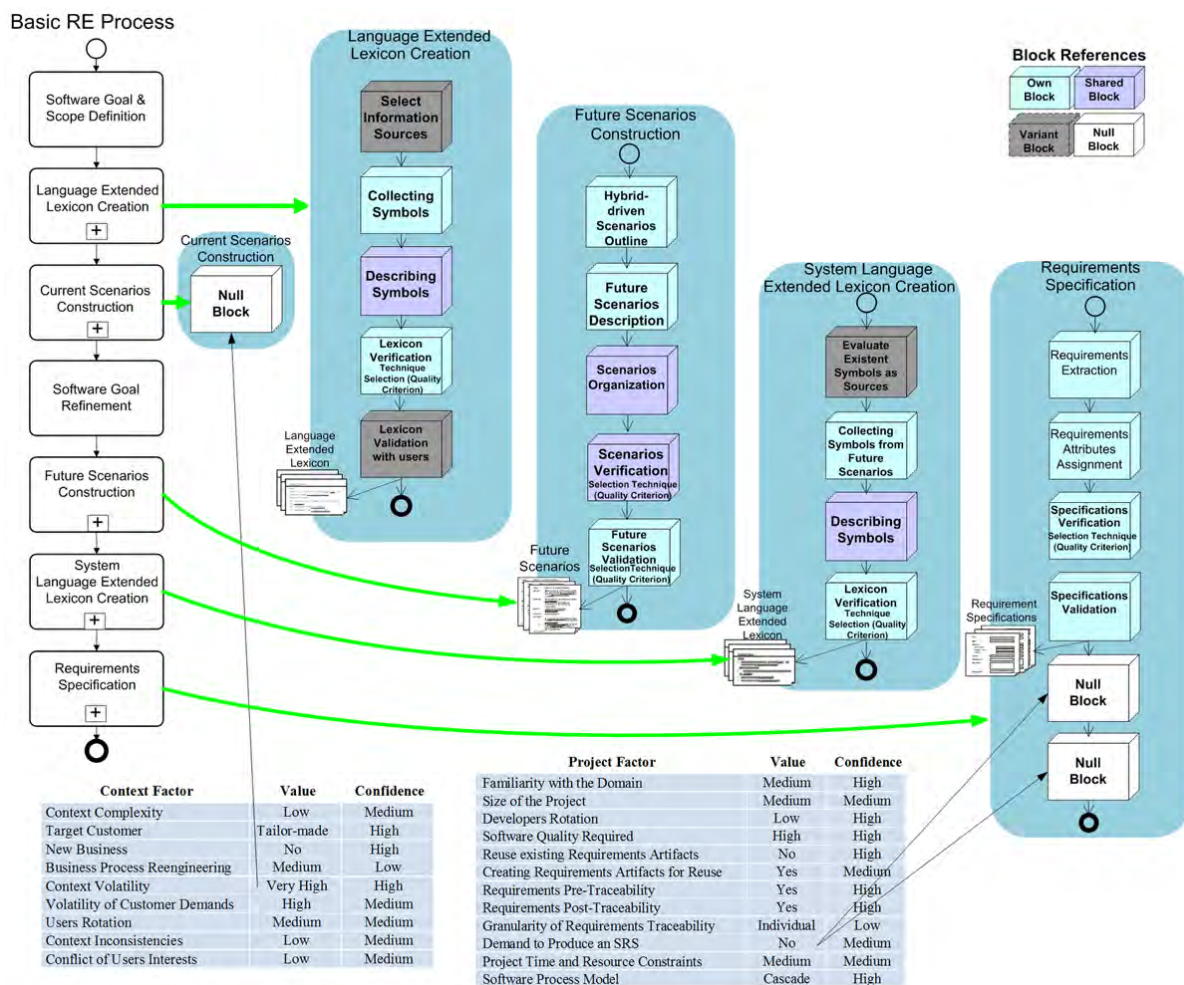
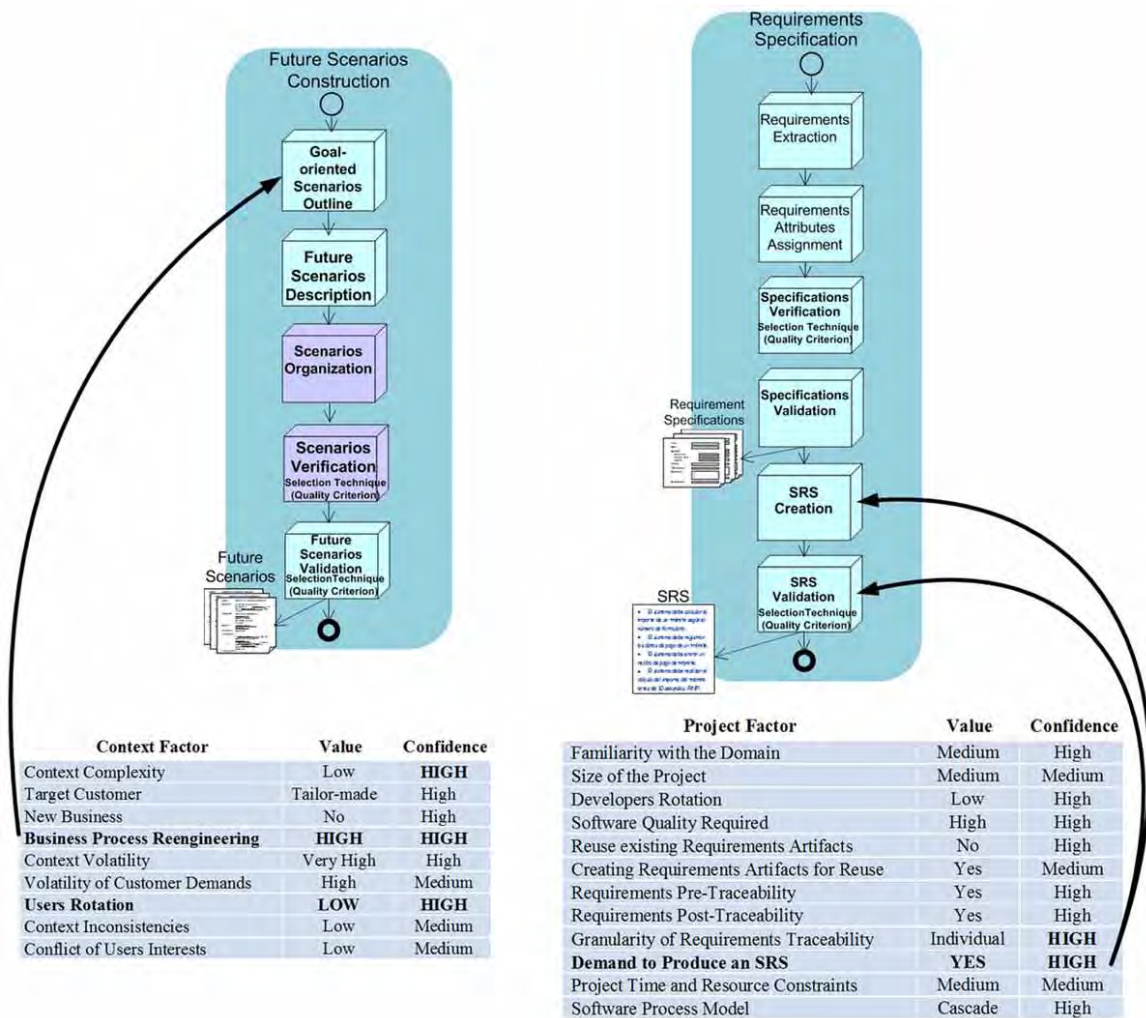


Figure 4. Evolution of the Instance of the RE Process



by any other main sub-process and is independent of situational factors. A *shared block* is common to different sub-processes of the RE process, and may depend on factors. A *variant block* belongs to a specific sub-process and depends on situational factors. A *null block* is an auxiliary block used when an entire sub-process or an atomic process must be deleted or skipped from the basic RE process. Table 2 shows operations that may be applied to each type of process block.

Some process blocks include product blocks, which contain minor variations of the NL models produced by the RE process. For example, if the lexicon model is not created, then the scenario

model does not include hyperlinks to lexicon symbols.

Figure 3 shows an instance of the RE process constructed for a specific situation, which is represented by a set of values given to factors at the beginning of the project. Those factors are sketched at the bottom of this Figure. While ongoing the project, some factors may evolve or change; this is exemplified in Figure 4 where at variation point 3 (see Figure 1), the following factors were updated: Business Process Reengineering, Users Rotation and Demand to Produce an SRS. These three factors, along with some others, gained a better confidence. The first and the third factor

have impacted on two different sub-processes, thus Figure 4 shows only sub-processes requiring a re-design.

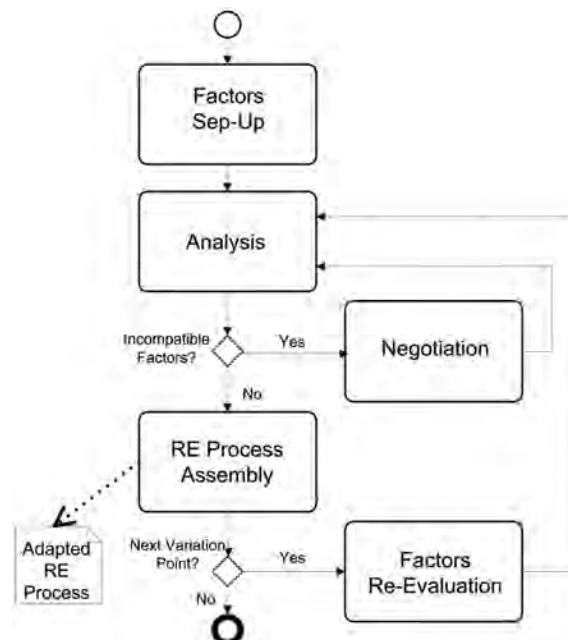
SOLUTIONS AND RECOMMENDATIONS

The adaptation of the basic RE process involves the application of a mechanism of operations on process blocks at every variation point, using a specific combination of situational factors, following SME principles. The variation points are milestones to review the RE process and to indicate possible re-design of remaining RE activities. The process for constructing the RE process for a specific situation includes the following activities (see Figure 5):

1. **Factors Set-Up:** The situational factors are evaluated, assigning a degree of confidence to each value, and taking into account the interaction among factors. A simple form is used to collect the factors value. This form eases its filling with valid values by means of rules for factors interaction.
2. **Analysis:** The factors setting is analyzed considering mainly interactions and factors with low confidence. Project managers should take a position on every factor with low confidence: optimistic, conservative or balanced position. Under an optimistic position, factors are assigned an extreme value of its valid range depending on the meaning of the factor, while under a conservative position factors take the opposite extreme value. A balanced position involves choosing intermediate values. This re-allocation of values is done to allow assembling the process, since the original values are preserved. When incompatible interactions are observed, a negotiation must take place.
3. **Negotiation:** The purpose of this activity is to reach an agreement among stakeholders to re-assign values to incompatible factors.

4. **RE Process Assembly:** The RE process is assembled departing from the basic process, considering a subset of situational factors at each variation point (see Table 1, last column). A set of rules based on a specific combination of those involved factors defines which operation is applied over each basic process block (see an example at Figure 2, right part). Thus, this assembly activity consists in maintaining, deleting, replacing, or parameterizing process blocks.
5. **Factors Re-Evaluation:** At every variation point, those factors involved in remaining variation points must be re-evaluated. Low-confidence factors must be watched, since its level of confidence is supposed to increase due to a better perception of the situation. Contingent factors must be also watched to establish if a change in the situation has occurred. After this activity, a new factors analysis is needed in order to ensure the absence of inconsistencies among dependent factors.

Figure 5. Process for Constructing an RE Process, based on situational factors



Since requirements are the foundations for a good software product, an RE process is crucial to the project success. Project managers should take decisions about selecting the appropriate RE process for a specific situation. A way to define the situation is by its characterization through context and project factors. Therefore, project managers should define as precisely as possible those situational factors in order to construct the suitable RE process. However, they should be aware of the potential evolution of the situation and, thereby, review factors and adjust the RE process at pre-defined milestones.

FUTURE RESEARCH DIRECTIONS

Feasible variants of the process were identified due to the experience achieved after putting into practice the basic RE process in many real cases along two decades.

Situational factors were exhaustively studied in literature, although this set is likely to be extended. Factors impacting on subsequent phases of the software development process will be studied to determine if they may influence backward on the RE process.

The decomposition of the RE process in basic process blocks, along with the set of situational factors, has allowed identifying process commonality and process variability, and hence defining the different types of process blocks needed. It is planned to evaluate the convenience of using mixed blocks, or separated process blocks from model blocks. It is also considered that more cases should be done to confirm the different branches of the adaptable RE process.

CONCLUSION

The proposal presented in this chapter deals with variability in an RE process, driven by a dynamic adaptation based on prevailing situational factors.

Most problems have distinctive features that must be taken into account to carry out a successful requirements process. Project managers should tailor the requirements process by choosing the techniques that are most suitable for specific situations. When they are provided with guidelines on how to adapt the requirements process, they are more willingly to do it.

The process for constructing an RE process involves a set of factors that typifies situations considering both the application context and the software project context. This adaptation process establishes when the RE process should be reviewed and re-tailored based on rules and operations at each variation point. Since it is frequently observed the natural and contingent evolution of situations, it is important to have the possibility of redefining the remaining activities of the RE process once the project is in course.

The adaptation of the requirements process for several real cases has contributed to its acceptance in host organizations. It is considered that the dynamic adaptation process could be implemented with minor adjustments in other RE strategies.

REFERENCES

- Alexander, I. F., & Beus-Dukic, L. (2009). *Discovering Requirements: How to Specify Products and Services*. Chichester, UK: John Wiley & Sons.
- Antonelli, L., Rossi, G., Leite, J. C. S. P., & Oliveros, A. (2012). Deriving requirements specifications from the application domain language captured by Language Extended Lexicon. In *Proceedings of XV Workshop on Requirements Engineering*. Buenos Aires, Argentina: Universidad Nacional de La Matanza.
- Bakhat, K. A., Sarwar, A. A., Motla, Y. H. B., & Akhtar, M. C. (2015). A Situational Requirement Engineering Model for an Agile Process. *Bahria University Journal of Information & Communication Technology*, 8(1), 21–26.

- Beck, K. (2004). *Extreme Programming Explained: Embrace Change*. Boston: Addison-Wesley Professional.
- Brinkkemper, S. (1996). Method Engineering: Engineering of Information Systems Development Methods and Tools. *Information and Software Technology*, 38(4), 275–280. doi:10.1016/0950-5849(95)01059-9
- Bucher, T., Klesse, M., Kurpjuweit, S., & Winter, R. (2007). Situational Method Engineering. In *Situational method engineering: fundamentals and experiences* (pp. 33-48). Springer US. doi:10.1007/978-0-387-73947-2_5
- Carrizo, D. (2009). *Marco para la selección de técnicas para educación de requisitos* (Doctoral Thesis). Universidad Politécnica de Madrid, Madrid, Spain.
- Carrizo, D., Dieste, O., & Juristo, N. (2008). Study of elicitation techniques adequacy. In *Proceedings XI Workshop on Requirements Engineering* (pp. 104-114). Barcelona, Spain: Universitat Politècnica de Catalunya.
- Cockburn, A. (2002). *Agile Software Development*. Boston: Addison-Wesley.
- Coulin, C. R. (2007). *A Situational Approach and Intelligent Tool for Collaborative Requirements Elicitation* (Doctoral Thesis). University of Technology, Sydney, Australia.
- Firesmith, D. (2004). Creating a project-specific requirements engineering process. *Journal of Object Technology*, 3(5), 31-44.
- Firesmith, D.G., & Henderson, B. (2002). *The OPEN Process Framework: An Introduction*. Harlow, UK: Addison-Wesley.
- Galster, M., Weyns, D., Tofan, D., Michalik, B., & Avgeriou, P. (2014). Variability in Software Systems - A Systematic Literature Review. *IEEE Transactions on Software Engineering*, 40(3), 282–306. doi:10.1109/TSE.2013.56
- Hadad, G. D., Doorn, J. H., & Kaplan, G. N. (2009). Creating Software System Context Glossaries. In M. Khosrow-Pour (Ed.), *Encyclopedia of Information Science and Technology* (2nd ed.; pp. 789–794). Hershey, PA: IGI Global. doi:10.4018/978-1-60566-026-4.ch128
- Haumer, P. (2005). *IBM Rational Method Composer: Part 1: Key concepts*. Retrieved August 27, 2015, from <http://www.ibm.com/developerworks/rational/library/dec05/haumer/index.htm>
- Henderson-Sellers, B., & Ralyté, J. (2010). Situational Method Engineering: State-of-the-Art Review. *J. UCS*, 16(3), 424–478.
- Hickey, A. M., & Davis, A. M. (2003, September). Elicitation technique selection: how do experts do it? In *Proceedings of 11th IEEE International Requirements Engineering Conference* (pp. 169-178). IEEE. doi:10.1109/ICRE.2003.1232748
- Jafarinezhad, O., & Ramsin, R. (2012, March). Towards a process factory for developing situational requirements engineering processes. In *Proceedings of the 27th Annual ACM Symposium on Applied Computing* (pp. 1089-1090). ACM. doi:10.1145/2245276.2231946
- Kaindl, H. (2000). A design process based on a model combining scenarios with goals and functions. *IEEE Transactions on Systems, Man, and Cybernetics. Part A, Systems and Humans*, 30(5), 537–551. doi:10.1109/3468.867861
- Khan, H.H., bin Mahrin, M.N., & bt Chuprat, S. (2014). Factors for Tailoring Requirements Engineering Process: A Review. *International Journal of Software Engineering and Technology*, 1(1), 7–18.
- Kohler, K., & Paech, B. (2002). Requirement Documents that Win the Race: Not Overweight or Emaciated but Powerful and in Shape. In *Proceedings of International Workshop on Time-Constrained Requirements Engineering (TCRE'02)*.

Kovitz, B. L. (2002). Hidden Skills that Support Phased and Agile Requirements Engineering. In *Proceedings of International Workshop on Time-Constrained Requirements Engineering (TCRE'02)*.

Kumar, K., & Welke, R. J. (1992). Methodology Engineering: a proposal for situation-specific methodology construction. In W. W. Cotterman & J. A. Senn (Eds.), *Challenges and strategies for research in systems development* (pp. 257–269). John Wiley & Sons.

Lauesen, S. (2002). *Software Requirements: Styles and Techniques*. Addison-Wesley.

Leffingwell, D., & Widrig, D. (2003). *Managing Software Requirements: a unified approach*. Addison-Wesley Professional.

Leite, J. C. S. P., Doorn, J. H., Kaplan, G. N., Hadad, G. D., & Ridao, M. N. (2004). Defining System Context using Scenarios. In J. C. S. P. Leite & J. H. Doorn (Eds.), *Perspectives on Software Requirements* (pp. 169–199). Springer, US. doi:10.1007/978-1-4615-0465-8_8

Leite, J. C. S. P., Hadad, G. D., Doorn, J. H., & Kaplan, G. N. (2000). A scenario construction process. *Requirements Engineering*, 5(1), 38–61. doi:10.1007/PL00010342

Lobo, L. O., & Arthur, J. D. (2005, April). An objectives-driven process for selecting methods to support requirements engineering activities. In *Proceedings of 29th Annual IEEE/NASA Software Engineering Workshop* (pp. 118–130). IEEE. doi:10.1109/SEW.2005.18

Macaulay, L. (1993, January). Requirements capture as a cooperative activity. In *Proceedings of IEEE International Symposium on Requirements Engineering* (pp. 174–181). IEEE.

Maiden, N. A. M., & Rugg, G. (1996). ACRE: Selecting methods for requirements acquisition. *Software Engineering Journal*, 11(3), 183–192. doi:10.1049/sej.1996.0024

Pinheiro, F. (2002). Requirements Honesty. In *Proceedings of International Workshop on Time-Constrained Requirements Engineering (TCRE'02)*.

Potts, C. (1995, August). Using schematic scenarios to understand user needs. In *Proceedings of the 1st Conference on Designing interactive systems: processes, practices, methods, & techniques* (pp. 247–256). ACM. doi:10.1145/225434.225462

Ralyté, J. (2013). Situational Method Engineering in Practice: A Case Study in a Small Enterprise. In *Proceedings of CAiSE'13 Forum at the 25th International Conference on Advanced Information Systems Engineering*, pp. 17–24.

Rolland, C. (2008). Method engineering: towards methods as services. In *Making Globally Distributed Software Development a Success Story* (pp. 10–11). Springer Berlin Heidelberg. doi:10.1007/978-3-540-79588-9_2

Seyff, N., Maiden, N., Karlsen, K., Lockerbie, J., Grünbacher, P., Graf, F., & Ncube, C. (2009). Exploring how to use scenarios to discover requirements. *Requirements Engineering*, 14(2), 91–111. doi:10.1007/s00766-009-0077-9

Sillitti, A., & Succi, G. (2005). Requirements Engineering for Agile Methods. In A. Aurum & C. Wohlin (Eds.), *Engineering and Managing Software Requirements* (pp. 309–326). Springer-Verlag. doi:10.1007/3-540-28244-0_14

ADDITIONAL READING

Browne, G. J., & Ramesh, V. (2002). Improving information requirements determination: A cognitive perspective. *Information & Management*, 39(8), 625–645. doi:10.1016/S0378-7206(02)00014-9

Henderson-Sellers, B., Ralyté, J., Ågerfalk, P., & Rossi, M. (2014). *Situational Method Engineering*. Springer-Verlag Berlin Heidelberg. doi:10.1007/978-3-642-41467-1

Kotonya, G., & Sommerville, I. (1998). *Requirements Engineering: Process and Techniques*. John Wiley & Sons.

Lloyd, W., Rosson, M., & Arthur, J. (2002). Effectiveness of elicitation techniques in distributed requirements engineering. In *Proceedings of 10th IEEE Joint International Conference on Requirements Engineering* (pp. 311-318). doi:10.1109/ICRE.2002.1048544

Ralyté, J., Deneckère, R., & Rolland, C. (2003, January). Towards a generic model for situational method engineering. In J. Eder & M. Missikoff (Eds.), *Advanced Information Systems Engineering* (pp. 95-110). Springer-Verlag Berlin Heidelberg. doi:10.1007/3-540-45017-3_9

Schnieders, A., & Puhlmann, F. (2007). Variability modeling and product derivation in e-business process families. In *Technologies for business information systems* (pp. 63-74). Springer Netherlands. doi:10.1007/1-4020-5634-6_6

Virtanen, P., Pekkola, S., & Paivarinta, T. (2013, January). Why SPI Initiative Failed: Contextual Factors and Changing Software Development Environment. In *Proceedings of 46th Hawaii International Conference on System Sciences (HICSS)*, (pp. 4606-4615). IEEE. doi:10.1109/HICSS.2013.609

Vlaanderen, K., Valverde, F., & Pastor, O. (2008). Improvement of a web engineering method applying situational method engineering. In *Proceedings of 10th International Conference on Enterprise Information Systems*, (3-1), (pp.147-154). Barcelona, Spain.

KEY TERMS AND DEFINITIONS

Process Block: A well-defined unit of process that can be reused as part of any process.

Process Commonality: The common elements of a process that facilitate the definition of a family of processes through reuse.

Process Variability: The variant elements of a process, identified at variation points, which produce deviation from the standard process.

Requirements Engineering Process: A process to produce software requirements by means of methods, techniques and tools during elicitation, modeling, analyzing and evolution of requirements.

Scenario: A representation of an observed or envisioned situation in the application context.

Situational Factor: A characteristic of the project or the application context that may be taken into account when implementing an RE process.

Situational Method Engineering: A discipline that promotes the construction of methods for developing systems according to a pre-defined set of situational factors.

A Formal Approach to the Distributed Software Control for Automated Multi-Axis Manufacturing Machines

S

Gen'ichi Yasuda

Nagasaki Institute of Applied Science, Japan

INTRODUCTION

Nowadays complex automated machinery is composed of multiple independently-driven subsystems, where each subsystem is actuated by at least one effector. The evolution of an overall system is not represented by time but by events, called discrete event based state change systems as a new type of dynamic systems. Typical discrete event systems, which are essentially manmade, include computer operating systems, communication networks, traffic control systems, service systems, and manufacturing systems, especially Flexible Manufacturing Systems (FMS) and Computer Integrated Manufacturing Systems (CIMS). Modern automated intelligent machinery also has the same features as such innovative discrete event systems. In the most part of such systems, the nature of the subsystems is asynchronous and their behaviors are concurrent, independent of other subsystems. During execution of an activity of a system, each subsystem assumes its own unique state without knowledge about the state of other subsystems.

In such a naturally distributed hardware structure, independent axes should be synchronously activated according to events required for cooperative task execution. When activities that depend on some resources are required, the subsystems have to communicate in order to update their status and to synchronize their actions. When two activities needing a shared resource are waiting for the end of the other activity, it may be resolved according to the operational constraints with priority. Different negotiation techniques between subsystems which should cooperate can be adopted based on

distributed communication structures, such as a variety of modified client/server protocols where the client and the server are not fixed, such that they are dynamically chosen according to the status of each in the system. Otherwise, a symptom of deadlock is detected and any supervisory function should be triggered based on centralized communication structures, so that, through any centralized decision making, the resource is assigned to one activity.

This chapter presents a unified and systematic design methodology of discrete event distributed control architecture by embedding intelligent agents with Petri nets. A machine-oriented, agent-based modular distributed software system is configured on a hierarchical distributed microcontroller based hardware structure for real-time intelligent machine control (Yasuda, 2010; 2011). The design of functionally distributed control software for real-time cooperative task execution is described based on the hierarchical modeling of machine tasks. Based on Petri net models of hardware elements of sensor devices such as switches, a set of program modules to execute primitive actions of effectors, is integrated with hierarchical and distributed configuration of behavior and control agents, especially to attain the synchronization of multi-axis motions involved in the complex manufacturing task.

BACKGROUND

Petri nets and automata are the most used to describe discrete event system control. Modeling and

DOI: 10.4018/978-1-5225-2255-3.ch647

analysis of discrete event systems with controllable and uncontrollable events, turned on and off by the supervisor, were proposed based on automata theory (Ramadge & Wonham, 1987). However, finite automata present some drawbacks such as the difficulty to model parallelism, synchronization and resource sharing. Although control commands generation was related to state transitions, since the connection of several finite automata is no longer a finite automaton, the communication specifications between modules can not be achieved using the finite automata framework (Lima & Saridis, 1996; Stadter, 1999; Silva et al., 2014). In spite of a great number of researches concerning advanced methods and tools to analyze the distributed models, these models are mostly constructed by empirical methods based on the knowledge of experts and customized for a particular application. Most approaches are limited by the combinatorial explosion that occurs when attempting to model complex systems.

On the other hand, Petri nets incorporate the notion of a distributed state of a system and a rule of state change (Murata, 1989; David & Alla, 1992), providing a mathematical formalism and a graphic tool for the formal representation of a system whose dynamics is characterized by concurrency, synchronization, nondeterministic decision, mutual exclusion and conflict. Computerized automation systems have the same, typical and most important features as industrial distributed systems.

As a mathematical formalism, a matrix equation can be set up to perform structural and behavioral analysis (Wisniewski et al., 2014) using a formal state-space representation. A Petri net simulates the behaviors through the flow of tokens like a flow chart, providing a visualization of the dynamic system. Petri nets allow a modular and hierarchical synthesis approach (Gomes & Barros, 2005) which can build up Petri nets from specification languages or formal definitions to control code. The ultimate goal of the development of Petri nets is to provide a methodology for control system design that is able to cope with the most dif-

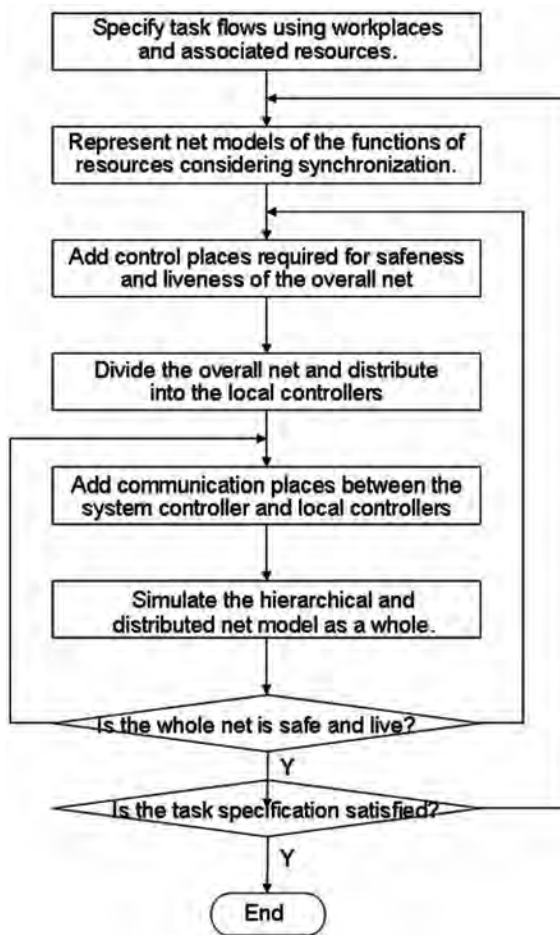
ficult and important aspects of non-deterministic cooperative control (Yasuda & Tachibana, 1991). Thus, the Petri net approach is expected to improve the performance of software control of multi-axis machines in automation systems.

FORMAL APPROACH TO CONTROL SYSTEM DESIGN BASED ON PETRI NETS

In control system design, model based and non-model based are two common approaches. Model based approaches use a model of the process under control for synthesis and verification to derive a controller based on formal specifications, whereas non-model based approaches have minimal or no assumptions about the process under control. The proposed methodology for embedded control system design using hardware-software co-design techniques starts with the description of the system's functionalities through a common system specification language such as Unified Modeling Language (UML). Then the description is translated into a set of formal models using hierarchical Petri nets as the reference model formalism.

At the highest level, the models represent system requirements, independent of any specific implementation platform. At the intermediate level, the models reflect the specific characteristics of system construct. These models are amenable to be translated into code, by being decomposed into sub-models and mapped into specific local controllers. The model based approach assures that a correct model at the highest level maintains its correctness, independently of selecting different mapping and implementation techniques. At the lowest level, the code reflects the concrete syntax of a specific implementation platform. For the model transformations between different abstraction levels, hierarchical Petri nets can be used, although any other behavior formalism with similar characteristics may be used without loss of generality of the proposed methodology.

Figure 1. Generalized flow of Petri net based control system design



The flow of control system design is shown in Figure 1. First the global system model is implemented as a centralized controller. Next, using model decomposition, several sub-models for independent controllers are obtained and implemented in the same platform. Finally the sub-models are implemented in different platforms or local controllers, such as microcomputer based systems, Programmable Logic Controllers (PLCs), etc. Behavioral properties associated with the execution of the distributed models are compared with the execution of the initial centralized model.

An ordinary Petri net is a weighted bipartite defined by a quadruple: a finite set of places represented by circles, a finite set of transitions represented by bars or rectangles, the flow rela-

tion of arcs connecting places and transitions, and the weight function which associates a nonzero value to each arc. Tokens reside in the places, represented by solid circles or dots, and a marking assigns to each place a positive integer, indicating the system state as a marking vector.

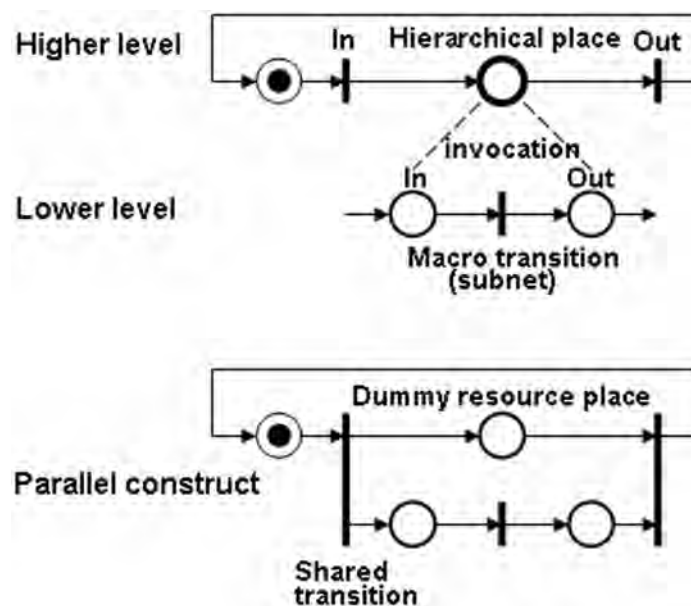
In condition-event Petri nets, where the weight value associated with an arc is 1, a condition that can be either true or false is represented by a place and a token. Events that can occur are represented by transitions. An event is enabled, or can fire or occur, if and only if all its pre-conditions are true and all its post-conditions are false.

A Petri net can be refined hierarchically by putting a subnet in a place and/or a transition. The subnet is subordinated to the higher-level place or transition. Unfolding is done by replacing the hierarchical net with a structure putting the corresponding subnet in parallel to the higher-level element as shown in Figure 2. A subnet definition should be self-contained without taking the other elements of the hierarchical net into account, while the subnet should not influence the outputs of the hierarchical net, so that the behavior of the flat net is identical with the hierarchical net.

SYNTHESIS OF HIERARCHICAL DISTRIBUTED NET MODELS FOR MACHINE TASK EXECUTION

Since Petri nets are a generalized and uninterpreted technique to represent discrete event systems, for modeling a specific system, an interpreted form of the Petri net is necessary to clearly model the functional features of the system including the required control conditions. From the view of discrete event systems, the dynamic behavior of a manufacturing machine is considered as the flow of discrete activities with the control of their start/end events between two successive activities. A machine activity means a manufacturing process performed in a workplace for objects (workpieces, parts or products), such as moving, mechanical processing, assembling, disassembling, buffer-

Figure 2. Hierarchical net modeling for real-time task control: (a) hierarchical control by dynamic subnet invocation and (b) its equivalent form with parallel construct



ing, etc. and can be seen as a task, operation or action according to the level of abstraction in the manufacturing system. A task is decomposed into a successive set of operations and actions, which also correspond to detailed processes for objects.

Typically, a workplace indicates a place where objects can be temporarily and stably held or stored for a specified manufacturing process. The condition for the execution of a process activity to start is the arrival of all specified objects and resources in the workplace, and the result of the process is one or more objects in the next workplace and the release of associated resources. A task flow graph is composed of workplaces and directed connection arcs between workplaces, representing the flow of manufacturing processes.

For explicit description of input/output facilities, the Petri net is extended to control the net model execution through transition firing using special input/output arcs in response to its environmental situations based on signal instead of events. A set of input signals, a set of output signals, a mapping associating every transition with a firing condition, or a Boolean function in the set of input signals, and a mapping associating

every place with an output function in the set of output signals, are specified.

A signal source or external device is represented by a macro place in the environment, which represents a subnet model of the corresponding hardware device. An enabled transition fires immediately when its firing condition associated with a combinatorial logical function is fulfilled. By addition of timing conditions, periodic, delayed or spontaneous firing of transitions can be adopted for autonomous machine control.

In automated manufacturing systems, subnets can be easily identified which realize certain subtasks performed by one machine or effector. A higher-level automation can be achieved through the specification transformations. The expected benefits of using a hierarchical model based approach for generating the models and respective code is to lower the cost of deploying a component and its code on multiple platforms or controllers through reuse. The hierarchical structure of machine control in a manufacturing system is conceptually composed of three levels from lower to higher: action level, operation level and task level. The lower levels are provided as

one of the sublevels determined by the upper level. In manufacturing systems, a machine or system is composed of units, and a unit is composed of devices, where each level is controlled based on sensing and actuating information with a corresponding level of abstraction. At each level, control specifications are represented using Petri nets and associated to corresponding resources required for real-time execution. For multi-axis machines, the action level is concerned with the actions executed by individual effectors. Because an action is executed by an associated effector in its workplace, if two actions are executed by a same effector, access to the workplace is in mutually conflict as a shared resource. The operation level is concerned with the operations executed by several effectors, where the objective is explicitly described by operation programs. At the task level, the discrete event systems representing machine tasks are concretely evaluated as task specifications.

Based on the proposed formal approach, the Petri net based control specifications are designed according to the levels of machine control. First, a task flow graph is defined as a conceptual net model for control specification at the task level, which describes the flow of objects through the machine process and availability of resources (machine units) using high-level Petri nets, where places represent the status of objects and resources. A task involves an object and at least one resource. Each object or resource can be either idle or involved in a task. Identification of the necessary tasks and related resources should be done using a database and/or based on online environment monitoring.

Due to the physical capacity of the machine units, such as assembling, disassembling and transfer devices, or operational limits, the maximum number of objects that can be accepted in the workflow section is prescribed. The restrictive condition is represented through a dummy place with limited capacity connected between the initial and final transitions of the section. Alternatively, inhibitive signal arcs are connected

in the initial transition of the section. The above control conditions are realized as a coordinator at a higher level in the hierarchical net model. A message sending/receiving mechanism based on asynchronous communication allows levels to communicate by taking control and monitoring requirements into account.

At the task level, the whole net model is split up into several sub-models which represent all possibilities of state change of a resource and all relations with other resources. Since the operation level comprises the detailed net models of activities that perform a task at the task level, the task model is detailed into the subnet models which describe causal relations of operations by units in the machine. Thus, a clear view of machine controller's functionalities is provided as rules of reactive input/output conditions relationships seen from the control system. The resultant model should have a structure where the state returns to the initial condition when the specified operations are finished; a Petri net graph with this property is necessarily live or deadlock free. Due to simplified functional descriptions, the whole discrete event manufacturing process performed by a machine can be naturally considered in terms of operations and actions.

Besides a single operation, operations can be combined as a sequential or concurrent construct. In a concurrent construct, operations can coexist in the execution state with a capacity, which start and end their execution through common initial and final events respectively. Several operations (actions) are joined, such that they start separately and finish their execution simultaneously through one final event. Operations (actions) are dispersed such that they start simultaneously through one initial event and finish their execution separately. The above most common types of operations are identified using appropriate firing rules with input signals as basic types. Other types of operations can be represented as combinations of these operations.

Since machine tasks generally involve several operations mutually related, typical problems are

synchronization of operations executed in parallel and/or in conflict (Luo & Zhou, 2014). Because the net model of the whole system is composed by live sub-graphs which describe the contents of the operations, the nature of the system is determined in terms of the structural interrelations between the operations, or due to some appropriate initial marking. When operations are executed in parallel, frequently it involves the synchronization of the start and/or end of some actions. Synchronous action is modeled through the fusion of the transitions that represent the start/end of the corresponding actions. In case of shared resource, since there exist convergence or ramification in the workflow, the starting and/or ending of several operations in conflict should be arbitrated.

For real-time control at the operation level, an operation is represented by one macro place with one input transition and one output transition which correspond to the start and end of its execution, respectively. The macro place represents the progressive status of the operation by the presence or absence of the token in the place. The function of each operation needs to be represented more explicitly and precisely by adding an interpretation with practical control information.

Finally the action level describes all the details regarding input/output signal changes and variables in lowest level elements, or devices such as sensors and actuators, to execute primitive actions. Because a simple action is a single combination of outputs, each action is activated by one or more triggers. For the execution of a primitive action by a PLC, the transition, input places which trigger transition firing and output places which represent firing effects are denoted as logic variables and directly arranged in a respective ladder diagram segment. The Petri net approach significantly simplifies the construction and validation of the control system, by algorithmically building the net models and automatically translating into a program in a common PLC language, such as ladder diagrams (Milik & Hryniewicz, 2012; Peng & Zhou, 2004; Wisniewski et al., 2011).

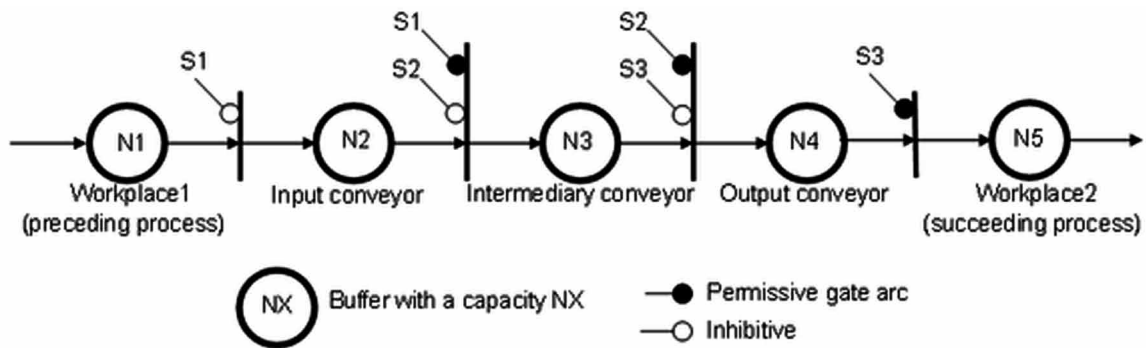
Actions with different control laws may require the same resources (sensors and actuators) or functional procedures, simultaneously. The start of an action engages resources necessary for its execution which are restored at its completion, as indicated by arcs between the action and the resources. In these cases, some external logic circuit, called an arbiter, is introduced to solve the conflict through the connections of input signal arcs between the conflicting transitions and the sub-graph that realizes the arbitration rules.

DESIGN OF NET BASED CONTROL SPECIFICATIONS FOR AUTOMATED MACHINES

An implementation example is a mechanized hand-stretched fine noodle producing machine, which automatically performs noodle stretching with holding and expanding units and transfers the bundles to the drying process with picking and placing devices for automatic handing over. Up/down movements of the units are controlled by position feedback using rotary encoders with a speed-controlled induction motor, and start/stop motions of the devices are controlled by specified infrared sensors. The whole task starts by carrying in the machine a set of bundles with a separate induction motor. After the completion of stretching process, the set of bundles are delivered to the drying process through the transfer machine.

The transfer machine performs the transfer of a set of noodle bundles from the preceding stretching and expanding process to the succeeding drying process. Because the number of bundles from the preceding process is different from that delivered to the drying process, three conveyors to regulate the transfer speed are configured, composed of an input, an intermediary, and an output conveyor. A remote infrared sensor to detect the presence of a bundle is attached to each outlet position of the preceding process and the three conveyors, S0, and S1, S2, S3 from input to output as indicated in Figure 3. Each conveyor is enabled when there

Figure 3. Task flow graph of sequential transfer composed of three conveyors with buffers of different capacities

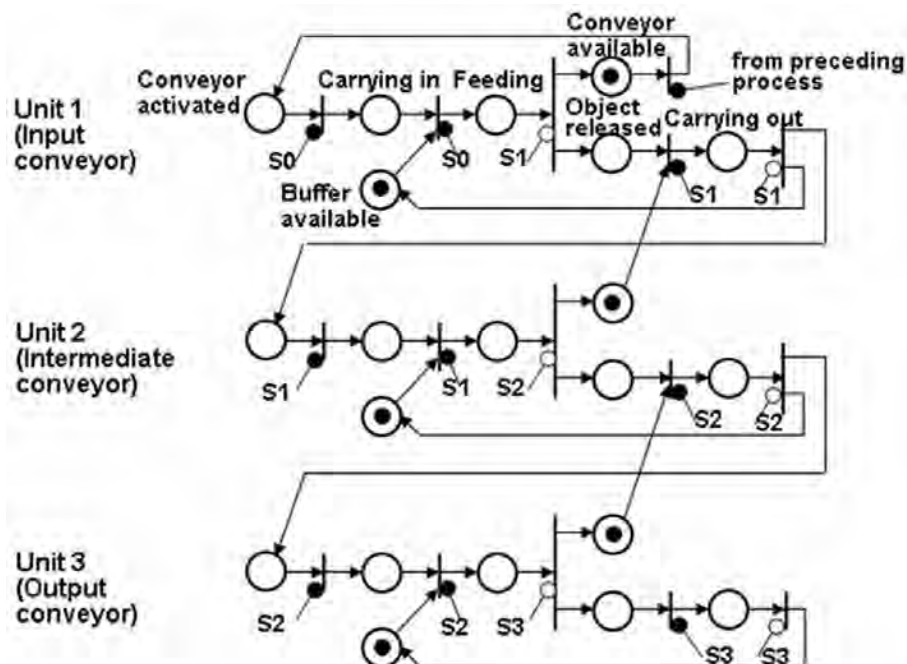


is a bundle in the conveyor and the succeeding conveyor is not full. The remote sensors are used to activate or stop motors to drive conveyors, in order to prevent collisions of bundles during the transfer process. The task flow graph is shown in Figure 3.

For example, when the ready signal from the preceding process is received, if S1 is OFF then the input conveyor is activated and intermediary conveyor is stopped until the buffer of the input

conveyor is full. When the buffer is full or S1 is ON, the intermediary is activated to transfer cooperatively with the input conveyor. When S1 is OFF, the intermediary conveyor is stopped, and the input conveyors may be continued according to the capacity of the input buffer. Then when the S2 is ON, the output conveyor is activated to transfer cooperatively with the intermediary conveyor, and when S2 is OFF, the output conveyor is stopped. The above control conditions and rules are

Figure 4. Detailed net model of sequential transfer machine composed of three unit controllers



represented by an extended Petri net model with different capacities in places representing conveyor actions and trigger functions for transition firing.

For the net model represented as the task flow graph at the task level, the detailed net models of the three transfer units at the operation level are shown in Figure 4, which represent sequentially connected transfer units. Because two neighboring units should be synchronously activated in order to transfer a bundle of noodle, mutual synchronization is realized by asynchronous communication, as shown with a shared place between two units. Distributed execution of the Petri net model is achieved through model decomposition into partitioned subnet models, each controlled by unit controllers.

The whole net model is decomposed into subnet models at the operation level, which are assigned to each unit controller on a microcomputer-based control system with expanded digital input/output interfaces. The task level net model is executed as the main background process, and the operation net models are executed as interrupt handling routines. The global event information necessary for synchronization and conflict resolution is processed as shared variables, which can be accessed by interrupt routines. Based on interrupts

from specified sensors, such as limit switches and encoder counts, transition enabling and firing tests at the lowest action level are executed quickly. After the transition firing, corresponding digital signals are outputted immediately through the interface.

The views of the combined noodle producing and transfer machine and the microcomputer based controller is shown in Figure 5. Although the machine cycle time depends on the number of specified events, under the conditions of ordinary slow motion speeds, the control software of the unit controllers have been satisfactorily executed with fewer states and request/response transmissions. The man-power of software development as well as coded lines could be drastically reduced compared to those written in a common programming language on a common PLC (Globelna et al., 2014). For comparison non-hierarchical Petri net model of a traditional control program for the transfer machine is shown in Figure 6. Because the hierarchical net model can reduce the number of net elements at each level, it greatly simplifies analysis and validation of the controller.

The limitations of the present work are more due to the status of the implementation than the fundamental ideas, for example due to the hard-

Figure 5. Overview of the automated noodle producing machine: (a) overall experimental system, (b) microcontroller based multi-axis sequential controllers

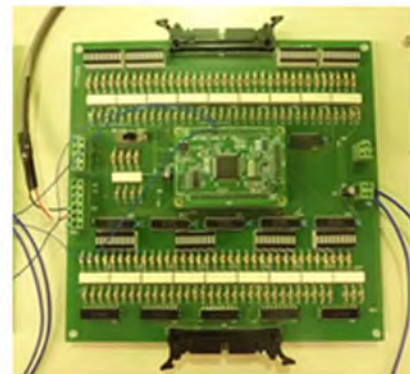
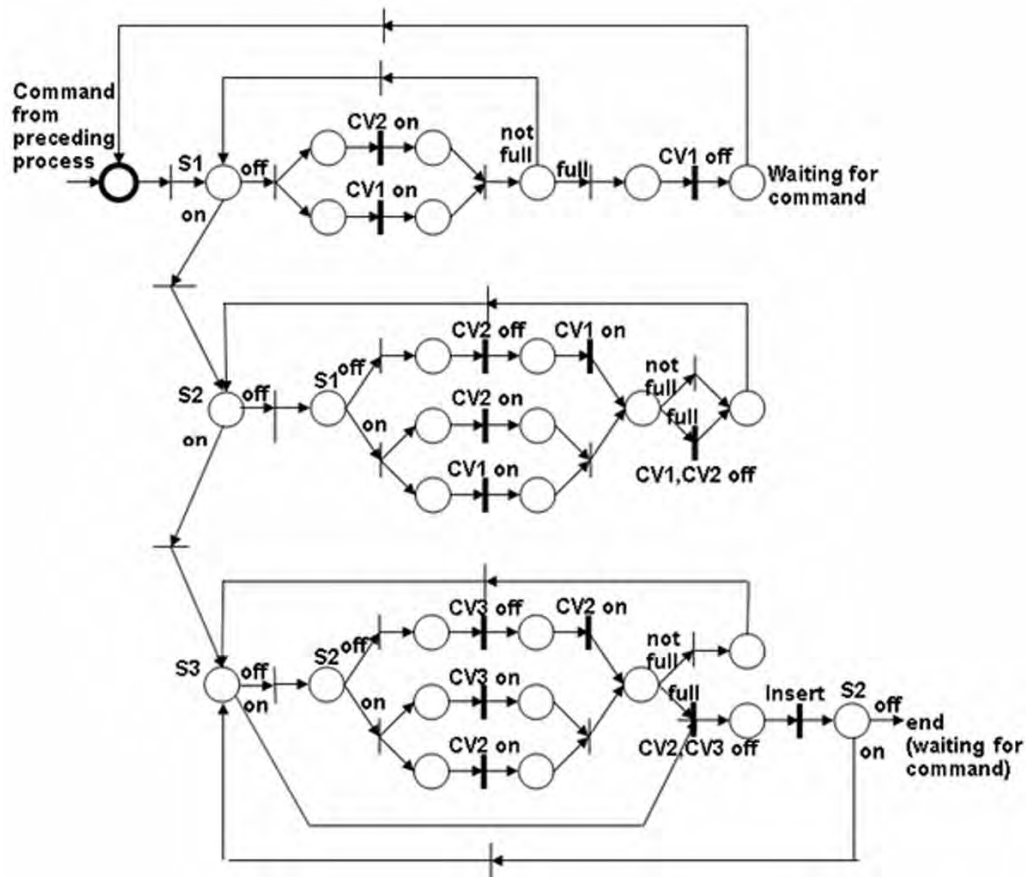


Figure 6. Non-hierarchical net model of control program flow of sequential transfer machine



ware and low-level software such as inter-process communication. Because the three activities (task, operation, action) should run in parallel in order to insure reactivity at every level, communication speed between levels using shared memory is also another important limitation of multithreaded implementation on a single processor. The knowledge bases required for improved robustness and better management of conflicts at every level should be optimally organized using customized retrieval software.

In contrast to the multithreaded implementation, the synchrony paradigm by synchronous communication does not apply in implementation platform of the distributed model on different time regions without a global clock. Based on an asynchronous communication scheme, a communication channel between controllers is

directed, such that an output event is generated by the master transition firing, which is read as an input event by the slave transition, where the delay is represented by a place between the master transition and the slave transition.

FUTURE RESEARCH DIRECTIONS

Some advanced agent technologies can be adopted for further development of the software control of sophisticated manufacturing machines. Because a Petri net based formal approach decentralizes the control activities for complex motion tasks, it can provide basic autonomy to agents according to a hierarchical design methodology. The task model is decomposed into subtasks in such a manner that every subtask may be assigned to an agent.

This decomposition is achieved by a partition of the centralized, functional net model, where each sub-model involves one agent controlling at least one effector. An agent acts as a decision unit when negotiating with higher-level agents, at the same time considered as a system when controlling and coordinating its components. The concepts of object-oriented programming may be applied to intelligent agents, as autonomous entities that have choices and control on their behavior, for integration of modeling, simulation and control software. The agent-based Petri net approach can be integrated with some intelligent control approach such as neural nets, for the activities related to diagnosis and error recovery or exception handling to react to unexpected situations.

CONCLUSION

Based on the formal approach using Petri nets, main control problems for automated multi-axis machines have been identified, such as synchronization, resource allocation and conflict arbitration, to assure that the resultant overall distributed control software is well-designed, coping with the prominent features of asynchronism and concurrency of complex manufacturing processes. The purpose of the formal approach is to facilitate the specification, validation, and generation of discrete event control code by smoothing the transition from specification to implementation. Three levels of detail used to describe the automated manufacturing process, offers a structural technique with well defined steps. Owing to the separation of detailed control, coordination and supervision, the task-level controller deals only with the group of command-level operations for coordination, while the implementation of primitive operations can be grouped into a single control-level macro-action performed by the local controllers. Because each level is concerned with small portions of information at a time without the need to look at all system states, the approach produces notable reduction in the number of con-

sidered states, therefore reducing the complexity of control system design.

REFERENCES

- David, R., & Alla, H. (1992). *Petri Nets and Grafcet - Tools for Modeling Discrete Event Systems*. New York: Prentice Hall.
- Globerna, I., Globery, M., & Adamski, M. (2014). Model checking of UML activity diagrams in logic controllers design. *Proceedings of 9th International Conference on Dependability and Complex Systems*, 233-242.
- Gomes, L., & Barros, J. P. (2005). Structuring and composability issues in Petri nets modeling. *IEEE Transactions on Industrial Informatics*, 1(2), 112–123. doi:10.1109/TII.2005.844433
- Li, Z., Wu, N., & Zhou, M. C. (2012). Deadlock control of automated manufacturing systems based on Petri nets – A literature review. *IEEE Transactions on Systems, Man and Cybernetics. Part C, Applications and Reviews*, 42(4), 437–462. doi:10.1109/TSMCC.2011.2160626
- Lima, P. U., & Saridis, G. N. (1996). *Design of Intelligent Control Systems Based on Hierarchical Stochastic Automata, Series in Intelligent Control and Intelligent Automation* (Vol. 2). World Scientific Publishing.
- Luo, J., Ni, H., & Zhou, M. (2014). Control program design for automated guided vehicle systems via Petri nets. *IEEE Transactions on Systems, Man, and Cybernetics*, 45, 44–55.
- Milik, A., & Hryniewicz, E. (2012). Synthesis and implementation of reconfigurable PLC on FPGA platform. *International Journal of Electronics and Telecommunication*, 58(1), 85–94. doi:10.2478/v10177-012-0012-8
- Murata, T. (1989). Petri Nets: Properties, analysis and applications. *Proceedings of the IEEE*, 77(4), 541–580. doi:10.1109/5.24143

Peng, S. S., & Zhou, M. C. (2004). Ladder diagram and Petri-net-based discrete-event control design methods. *IEEE Transactions on Systems, Man and Cybernetics. Part C, Applications and Reviews*, 34(4), 523–531. doi:10.1109/TSMCC.2004.829286

Ramadge, P. J., & Wonham, W. M. (1987). Supervisory control of a class of discrete event processes. *SIAM Journal on Control and Optimization*, 25(1), 206–230. doi:10.1137/0325013

Silva, E., Campos-Rebelo, R., Hirashima, T., Moutinho, F., Malo, P., Costa, A., & Gomes, L. (2014). Communication support for Petri nets based distributed controllers. In *Proceedings of 2014 IEEE International Symposium on Industrial Electronics*, (pp. 1111–1116). doi:10.1109/ISIE.2014.6864769

Stadter, P. A. (1999). Discrete event command and control for formation flying of distributed small spacecraft systems. *Proceeding of 13th Annual AIAA/USU Conference on Small Satellites, SSC99-VI-4*.

Wisniewski, R., Barkalov, A., Titarenko, L., & Halang, W. (2011). Design of microprogrammed controllers to be implemented in FPGAs. *International Journal of Applied Mathematics and Computer Science*, 21(2), 401–412. doi:10.2478/v10006-011-0030-1

Wisniewski, R., Stefanowicz, L., Bukowiec, A., & Lipinski, J. (2014). Theoretical aspects of Petri nets decomposition based on invariants and hypergraphs. *Lecture Notes in Electrical Engineering*, 308, 371–376. doi:10.1007/978-3-642-54900-7_52

Yasuda, G. (2010). Specification and real-time control of robotic manufacturing systems based on concurrent process modeling. *Proceedings of the 15th International Symposium on Artificial Life and Robotics 2010 (AROB15th'10)*, 782–785.

Yasuda, G. (2011). Implementation of real-time distributed control for discrete event robotic manufacturing systems using Petri nets. *Proceedings of the 16th International Symposium on Artificial Life and Robotics 2011 (AROB16th'11)*, 679–682.

Yasuda, G., & Tachibana, K. (1991). A distributed information-control structure for robotic systems. In *Production Research: Approaching the 21st Century*. Taylor & Francis.

ADDITIONAL READING

Wu, N. Q., & Zhou, M. C. (2009). *System Modeling and Control with Resource-oriented Petri Nets*. CRC Press.

KEY TERMS AND DEFINITIONS

Deadlock: In parallel control systems, a situation where two or more actions are each waiting for the other to finish and thus neither ever does. The liveness of the system implies deadlock-freeness.

Discrete Event Systems: A dynamic system where the state evolution depends entirely on the occurrence of asynchronous discrete events over time, consisting of discrete state spaces and event-driven state transition mechanisms.

Formal Methods: Mathematical techniques for the development of functional specifications and designs of real-time software.

Manufacturing Machines: Complex, high-speed machines which process and combine materials as they move through the machine to produce a finished product.

Multi-Axis Mechanisms: Mechanisms which have complex motion requirements by multiple independent drives, in place of the traditional centralized cam-actuated mechanisms.

Software Control: Control of electro-mechanical actuator system, where the actuator function is defined by software and can be readily modified or reprogrammed.

Synchronization: In multiprocessing systems, joining multiple independent processes in order to reach an agreement or commit to a certain sequence of actions.

Model-Driven Software Modernization



Liliana Maria Favre

Universidad Nacional Del Centro De La Provincia De Buenos Aires, Argentina

Liliana Martinez

Universidad Nacional del Centro de la Provincia de Buenos Aires, Argentina

Claudia Teresa Pereira

Universidad Nacional del Centro de la Provincia de Buenos Aires, Argentina

INTRODUCTION

Most vital software systems in companies have been developed several years ago with technology that today can be considered obsolete and unaligned with their current strategic objectives. These legacy systems play a central role in the company's information system. They are often large business-critical applications that have involved the investment of money, time and other resources through the years. Therefore, companies are facing the problematic of having to modernize or replace their legacy software systems.

Unlike the maintenance, software modernization involves significant changes in the structure and functionality of a legacy system with the aim of increasing its strategic value. Software modernization, understood as technological and functional evolution of legacy systems, provides principles, methods, techniques and tools to support the transformation from an existing software system to a new one that satisfies new requirements. It is related to reverse engineering, restructuring, forward engineering and reengineering processes. The definitions given by Chikofsky and Cross (1990) for these processes are still in force beyond the technological evolution:

- *Reverse engineering* is the process of analyzing a subject system to identify the system's components and their interrelationships and, create representations of the

system in another form or at a higher level of abstraction.

- *Restructuring* is the transformation from one representation form to another at the same relative abstraction level while preserving the subject system's external behavior (functionality and semantic).
- *Forward engineering* is the traditional process of moving from high-level abstractions and logical, implementation-independent designs to the physical implementation of a system.
- *Reengineering* is the examination and alteration of a subject system to reconstitute it in a new form and the subsequent implementation of the new form.

The term *Refactoring*, introduced by Martin Fowler, defines a special kind of restructuring in object-oriented code (Fowler, 1999). One of the advantages of software modernization versus traditional software reengineering is that the former allows discovering and refactoring object-oriented models at a high level of abstraction, such as business process models.

Software modernization faces many challenges due to the proliferation of new technologies such as mobile computing and cloud computing. Thus, novel technical frameworks for information integration and tool interoperability are needed. The *Model Driven Development* (MDD) appears to be an interesting approach to address these chal-

lenges since it provides principles and techniques to represent software systems through models at different levels of abstraction. MDD refers to a range of development approaches based on the use of software models as first class entities. The most well-known realization of MDD is the OMG standard Model Driven Architecture (MDA) (MDA, 2016). The outstanding ideas behind MDA are separating the specification of the system functionality from its implementation on specific platforms, managing the software evolution from abstract models to implementations increasing the degree of automation of model transformations and achieving interoperability with multiple platforms, programming languages and formal languages. The essence of MDA is the Meta Object Facility Metamodel (MOF) that allows different kinds of software artifacts to be used together in a single project (MOF, 2015). Models play a major role in MDA which distinguishes Computation Independent Model (CIM), Platform Independent Model (PIM) and Platform Specific Model (PSM). Some authors also distinguish Implementation Specific Model (ISM) as a description (specification) of the system in source code. To express transformations, OMG defined the MOF 2.0 Query, View, Transformation (QVT) metamodel (QVT, 2015).

OMG Architecture-Driven Modernization Task Force (ADMTF) is developing a set of specifications and promoting industry consensus on modernization of existing applications (ADM, 2016). In the ADM context, software modernization can be summarized as follows. First, information is extracted out of the system artifacts. Second, this information is analyzed in order to take adequate modernization decisions and finally, the information must be transformed to new artifacts. These steps are supported by metamodels to describe existing systems, discoverers to automatically create models of these systems and, tools to understand and transform complex models created out of existing systems. The *Knowledge Discovery Metamodel (KDM)* and the *Abstract Syntax Tree Metamodel (ASTM)* are

two complementary and relevant ADM standards (KDM, 2011) (ASTM, 2011).

This chapter analyzes ADM-based software modernization. It provides an overview of the-state-of-the-art in software modernization techniques. Taxonomy of different techniques is described. A description of how traditional techniques such as static and dynamic analysis can be integrated with ADM standards is presented. We propose a framework for ADM software modernization that integrates different paradigms. Finally, challenges and strategic directions in software modernization are included.

BACKGROUND

25 years ago, modernization focused mainly on reverse engineering for recovering high-level architectures or diagrams from procedural code to face up with problems such as comprehending data structures or databases or the Y2K problem. At that time, many different kinds of static analysis techniques had been developed and several studies had been carried out to compare them.

Later years, new approaches were developed to identify objects into legacy code and translate this code into an object-oriented language. Object-oriented programs are essentially dynamic and present particular problems linked to polymorphism and late binding, abstract classes and dynamically typed languages. Then, the focus of object-oriented software analysis moved from static analysis to dynamic one, more precisely static analysis was complemented with dynamic one. Many works had contributed to reverse engineering object-oriented code. Muller, Jahnke, Smith, Storey, Tilley, and Wong (2000) present a roadmap for reverse engineering research for the first decade of the 2000s. Fanta and Rajlich (1998) describe the reengineering of a deteriorated object-oriented industrial program written in C++. Systa (2000) describes an experimental environment to reverse engineer JAVA software

integrating dynamic and static information. Demeyer, Ducasse, and Nierstrasz (2002) distinguish a variety of techniques for object-oriented reengineering based on patterns.

When the Unified Modeling Language (UML) (UML, 2015) emerged, a new problem was how to extract higher-level views of the system expressed by different kind of UML diagrams. The diagrams that could be reverse-engineered in this way were partial. A new challenge was how to identify different relationships (e.g. dependency, association, aggregation and composition). Although there is relevant work for extracting UML diagrams from source code, many challenges still need to be done. Tonella and Potrich (2005) provide a relevant overview of techniques that have been investigated and applied in the field of reverse engineering of object-oriented code. Authors describe the algorithms involved in the recovery of UML diagrams, such as class diagrams, state diagrams, and sequence diagrams from code. This proposal is based on static and dynamic analysis.

Other works are linked to MDD-based reverse engineering. MacDonald, Russell, and Atchison (2005) report on a project that assessed the feasibility of applying MDD to the evolution of a legacy system. Deissenboeck and Ratiu (2006) show the first steps towards the definition of a metamodel that unifies a conceptual view on programs with the classical structure-based reverse engineering metamodels. Reus, Geers and van Deursen (2006) describe a feasibility study in reengineering legacy systems based on grammars and metamodels. At that moment, Canfora, Di Penta and Cerulo (2011) present a relevant survey that compares existing work in the reverse engineering area, discuss success stories and main achievements, and provide a road map for possible future developments in this area. Fleurey, Breton, Baudry, Nicolas, and Jézéquel (2007) report on the use of MDE as an efficient, flexible and reliable approach for a software migration process. The described process, developed at Sodifrance, includes automatic analysis of the existing code, reverse engineering of abstract high-level models,

model transformation to target platform models and code generation. Sodifrance has developed a tool suite for model manipulation called Model-In-Action (MIA) used as a basis for automating the migration.

New works are related to ADM. Barbier, Deltombe, Parisy, and Youbi (2011) describe a model driven reverse engineering method based on metamodeling and model transformation. The method was implemented in the BLU AGE® Reverse module, an Eclipse IDE plugin. In this approach, a textual DSL is constructed to later describe source code as formal KDM models. Next, these KDM models are transformed to UML PIMs (BLUE AGE, 2016). Authors generalize this method by extending KDM along with an implementation of the ASTM standard. Authors state that the link between KDM and ASTM is not clear even confusing.

Many CASE tools support reverse engineering, however, they only use more basic notational features with a direct code representation and produce very large diagrams (CASE MDA, 2016). Few MDA-based CASE tools support any of the QVT languages. As an example, IBM Rational Software Architect and Spark System Enterprise Architect do not implement QVT. Other tools partially support QVT, for instance Together allows defining and modifying transformations Model-to-Model (M2M) and Model-to-Text (M2T) that are QVT-Operational compliant. Medini QVT partially implements QVT (Medini, 2015). It is integrated with Eclipse and allows the execution of transformations expressed in the QVT-Relation language.

Nowadays, the most complete technology that supports ADM is MoDisco, a Generative Modeling Technology (GMT) component for Model-Driven Reverse Engineering (MoDisco, 2016). As an Eclipse component, MoDisco can be integrated with plugins or technologies available in the Eclipse environment such as Eclipse Modeling Project (EMP), Eclipse Modeling Framework (EMF), and model-to-model transformation. The Model-to-Model Transformation (MMT) Eclipse

project is a subproject of the top-level EMP that provides a framework for model-to-model transformation languages. Transformations are executed by transformation engines that are plugged into the Eclipse Modeling Infrastructure. The main transformation engines developed in the scope of that project are ATL and QVT (ATL, 2016). To date, the QVT declarative component is in its “incubation” phase and provides only editing capabilities to support the QVT language.

SOFTWARE MODERNIZATION TECHNIQUES

This section describes modernization techniques based on traditional techniques and ADM. Besides, it analyzes the influence of formal methods in this area.

Static and Dynamic Analysis

Static analysis extracts static information that describes the software structure reflected in the software documentation (e.g., the source code text) whereas dynamic analysis information describes the structure of the run-behavior and can be extracted by using debuggers, event recorders and general tracer tools. Static analysis is based on classical compiler techniques and abstract interpretation (Tonella, & Potrich, 2005).

Data-flow graph and the generic flow propagation algorithms are specializations of classical flow analysis techniques. Because there are many possible executions, it is usually not reasonable to consider all state of the program. Thus, static analysis is based on abstract interpretation of program model state that is easier to manipulate, although it loses some information. These ideas, applied in compiler optimizations, require information about program semantics and are semantics-preserving program transformations.

Ernst (2003) provides a comparison of static and dynamic analysis from the point of view of their synergy and duality. He argues that static

analysis is conservative and sound. Conservatism means reporting weak properties that are guaranteed to be true, preserving soundness, but not strong enough to be useful. Soundness guarantees that static analysis provides an accurate description of the behavior, no matter on what input or in what execution environment. Dynamic analysis is precise given that it examines the actual run-time behavior of the program, however the results of executions may not generalize to other executions.

The combination of static and dynamic analysis can enrich reverse engineering process. There are different ways of combination, for instance performing first static analysis and then dynamic one or perhaps, iterating static and dynamic analysis.

ADM Techniques

ADM refers to the process of understanding and evolving existing software assets for software improvement, modifications, interoperability, refactoring, restructuring, reuse, porting and migration. Next, we describe the main ADM standards related to modernization.

Standards for Modernization: KDM and ASTM

The key motivation for the establishment of standards for the interchange of software models within the ADMTF context is to facilitate interoperability between the tools and services of the adherents of the standard (ADM, 2015).

KDM and ASTM are two complementary and relevant ADM standards. ASTM acts as the lowest level foundation for modeling of software within the OMG ecosystem of standards, whereas KDM serves as a gateway to the higher-level OMG models.

ASTM complements KDM by providing a continuous framework for mapping between low-level software models that are represented in ASTM and higher-level conceptual views of software that are represented by KDM and other OMG modeling standards such as UML (UML, 2015).

Knowledge Discovery Meta-Model (KDM)

The main goal of KDM is to provide a common interchange format that allows interoperability between tools for maintenance, evolution, assessment and modernization.

KDM is defined via MOF and determines the interchange format via the XMI (XMI, 2011). KDM is a metamodel for knowledge discovery in software that allows representing information related to existing software assets, their associations, and operational environments regardless of the implementation programming language and runtime platform. KDM represents entire enterprise software systems, not just code. KDM separates knowledge about existing systems into several orthogonal facets known as Architecture Views. The KDM specification is organized into the following 4 layers, each layer is further organized into packages which define a set of metamodel elements whose purpose is to represent a certain independent facet of knowledge related to existing software systems (KDM, 2011): Infrastructure Layer, Program Elements, Layer, Runtime Resource Layer and Abstractions Layer.

Abstract Syntax Tree Metamodel (ASTM)

In the MDD context, ASTs are used as a model of the source code. ADMTF has defined ASTM, a specification for modeling elements to express AST in a representation that facilitates the interchange of software models among tools (ASTM, 2011).

ASTM describes the elements used for composing AST models. An ASTM model is a model of how the statements of a software asset are structured reflecting the grammar of the particular programming language.

To provide uniformity as well as a universal framework for extension, the ASTM specification is composed of the Generic Abstract Syntax Tree Metamodel (GASTM), the Specialized Abstract

Syntax Tree Metamodels (SASTMs) and the Proprietary Abstract Syntax Tree Metamodels (PASTM).

Metamodel Formalization

Metamodeling is a central aspect to many modeling frameworks. Then, it is important to reason with formal specification about them. However, progress in this direction is still emerging.

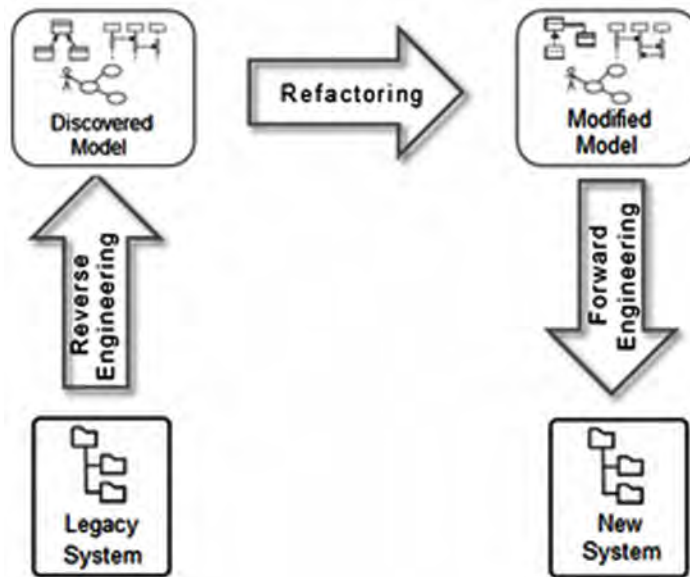
Jackson, Levendovszky, and Balasubramanian (2011) state that there has been a general consensus that MOF is underformalized and attempts at full formalization of MOF or like-MOF metamodels have not yet enabled extensive automated reasoning on metamodeling frameworks. One reason may be the strong association of MOF with OCL. OCL has a denotational semantics that has been implemented in tools allowing dynamic validation of snapshots. However, it cannot be considered strictly a formal specification language due to a formal language must at least provide syntax, some semantics and an inference system. Willink (2011) states that OCL contains many inconsistencies and evolves with MDA standards.

Many approaches, based on different formal specifications, support limited automated reason (Jouault & Bezivin, 2006) (Boronat & Meseguer, 2010) (Anastasakis, Bordbar, Georg & Ray, 2007).

Jackson, Levendovszky, and Balasubramanian (2011) present a relevant approach to specify metamodeling frameworks modularly using algebraic data types and constraint logic programming (CLP). Proofs and test-case generation are encoded as CLP satisfiability problems and automatically solved by using their FORMULA framework. Hassam, Sadou, Le Gloahec & Fleuquin (2011) propose a tool to help the designer to make a decision on the constraints attached to a metamodel during its evolution. This approach consists in identifying the impact of a change made at metamodel level on its associated constraints.

Favre (2009) presents a special-purpose algebraic language called NEREUS to provide extra support for metamodeling. NEREUS takes

Figure 1. Software modernization



advantage of existing theoretical background on formal methods, for instance, the notions of refinement, implementation correctness, observable equivalences and behavioral equivalences that play an essential role in model-to-model transformations. Most of the MOF metamodel concepts can be mapped directly to NEREUS. NEREUS allows specifying metamodels such as the Ecore metamodel, the specific metamodel for defining models in EMF (EMF, 2016). The NEREUS semantics was given by translating it to CASL language (Bidoit & Mosses, 2004). An appropriate set of tools to make formal metamodeling feasible in practice is provided. It includes a parser for NEREUS language, a translator from NEREUS to CASL and application that provides the ability to write specification in NEREUS, integrating the analyzer and the translator. Favre and Duarte (2016) describe formal metamodeling in NEREUS and tool support.

A MODEL-DRIVEN MODERNIZATION FRAMEWORK

We describe a framework for software modernization that integrates traditional reengineering

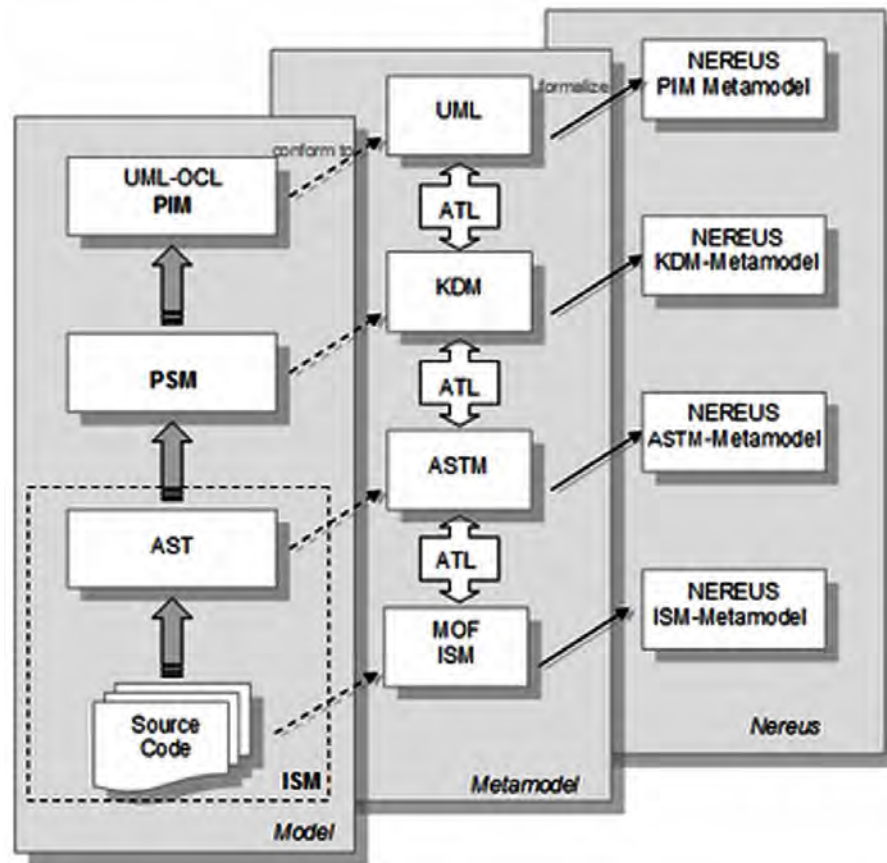
techniques with MDA, ADM and formal specification. The framework identifies three layers linked to reverse engineering, refactoring and forward engineering (Figure 1).

The basic steps are organized as a series of model transformations. Before reverse engineering, it is necessary a previous step for injecting models from the software to be modernized. This transformation is a Text-to-Model (T2M) transformation. After this injection step, Model-to-Model (M2M) transformations (linked to reverse engineering and restructuring stages) are applied until obtain high-level views of the artifacts involved in the modernization process. Finally, in the forward engineering stage, Model-to-Text (M2T) transformations are applied to generate new applications.

The reverse engineering distinguishes three different abstraction levels linked to models, metamodels and formal specifications. These models provide a uniform representation of the system in the ADM standard context (Figure 2).

In the MDD context, the reverse engineering stage extracts elements from existing systems and represents them into a Platform Specific Model (PSM), and subsequently Platform Independent Models (PIMs) are obtained from the PSMs. In the ADM context, KDM is the support for rep-

Figure 2. Model driven reverse engineering



representing PSMs by using ASTM as intermediate representation of a software system. The models at PIM level are built from PSMs by means of transformation (formal mappings) between KDM and UML.

The model level includes code, PIMs, PSMs and ISMs (Implementation Specific Models). PIMs are expressed in UML and OCL and include use case diagrams, activity diagrams, interaction diagrams to model system processes and state diagrams to model lifecycle of the system entities. PSMs are expressed as KDM models. An ISM is a specification of the existing system that is expressed by means of the code and its abstract representation (AST).

At model level, transformations are based on classical compiler construction techniques. They involve processes with different degrees of auto-

mation which range from totally automatic static analysis to processes that require human intervention to dynamically analyze the resultant models. All the algorithms that deal with reverse engineering share an analysis framework. The basic idea is to describe source code or models by an abstract language and perform a propagation analysis in a data-flow graph called in this context Object-Data Flow Graph (ODFG) (Tonella & Potrich, 2005). The static analysis is complemented with dynamic analysis supported by tracer tools.

The metamodel level includes metamodels defined via Meta-Object Facility (MOF), such as UML, KDM and ASTM, which describe the transformations at model level. A metamodel is an explicit model of the constructs and rules needed to construct specific models. MOF metamodels use an object modeling framework that is essentially

a subset of UML. The modeling concepts are classes that model MOF metaobjects, associations that model binary relations between metaobjects, data types that model other data, and packages that modularize the models. The metamodel level includes the UML metamodel that describes families of PIMs, the KDM metamodel that describes families of PSMs and the ASTM metamodel that describes ASTM models. Metamodel transformations are specified as contracts between a source metamodel and a target metamodel. MOF metamodels “control” the consistency of these transformations.

The level of formal specification includes specifications of MOF metamodels and metamodel-based transformations in the metamodeling language NEREUS.

Refactoring is a change made to the internal structure of software to make it easier to understand and cheaper to modify without changing its observable behavior. The Refactoring stage is a sequence of particular M2M transformations in which the source and target models conform to the same metamodel.

The forward engineering stage is the process of moving from high-level abstractions expressed as MDA models into implementations through a sequence of M2M transformations according to MDA principles.

Different realizations of this framework reflect different scenarios of modernization. As a realization of this framework, we proposed a process to migrate Java or C++ applications to mobile platforms (Diaz Bilotto & Favre, 2016).

CHALLENGES ON ADM-BASED MODERNIZATION

The success of legacy system modernization depends on the existence of technical frameworks such as MDA and ADM to cope with the diversity of languages and technologies used to develop a single software system.

Today, the most complete technology that supports ADM is MoDisco. However, MoDisco only supports reverse engineering of class diagrams. Important research issues to be addressed within the context of MoDisco include reverse engineering of different diagrams and advanced composition of heterogeneous models, scalability of model manipulation techniques and traceability during the whole life cycle of a MDD project.

We can summarize benefits of MDD approach as follows. It increases productivity due to the automation introduced in the generation of artifacts. It is more profitable, if the number of migrations is large and there are many applications of the same type to be migrated.

The models play the role of decoupling element to ensure the independence of source and target models with respect to their underlying technologies. Independence is achieved with injectors, M2M and M2T transformations that generate models independent of the source model, and by means of M2M and M2T transformations that construct a target model from independent models. Besides, a modernization process can be extended adding or replacing existing stages. In a chain of transformations, the models provide a valid extension point to add new stages.

The limitations of MDE approach are related to its maturity degree. KDM offers a reduced level of detail and does not distinguish concepts common to many technologies. The debugging process for M2M transformations is limited. Another limitation is the lack of open-source injectors and/or metamodels for programming languages.

The lack of predefined solutions for reverse engineering is another challenge. Sometimes a series of problems such as refactoring, or pre-static analysis of code are relatively general and can be applied to a large number of migration projects. Therefore, it would be interesting to have predefined solutions that express them in terms of standard metamodels

Bruneliere, Cabot, Dupé and Madiot (2014) state that the MoDisco approach is used as a

solid base for real applications, however, some important improvements on model driven reverse engineering techniques are still needed. They affirm that important research issues to be addressed including scalability of model manipulation techniques, advanced composition of heterogeneous models and traceability during the whole life cycle of a project.

To use formal methods that place change and evolution of models and metamodels in the center of the software development process is another challenge. In particular, with the existing verification tools, simple changes in a system require to verify its complete specification again making the cost of the verification proportional to its size.

CONCLUSION

This chapter presents state of the art in the Model Driven Modernization area. Taxonomy of different modernization techniques is described. Particularly, we analyze traditional reverse engineering techniques based on lstatic analysis and dynamic analysis and ADM-based techniques.

Modernization of legacy systems is a new research area in the software industry intended to provide support for transforming an existing software system to a new one that satisfies new demands.

We show that technical frameworks for information integration and tool interoperability that allow managing new platform technologies, such as MDA and ADM, can help to achieve the goals of software modernization. Particularly, we propose a framework ADM compliant to reverse engineering models from object-oriented code based on the integration of static and dynamic analysis, metamodeling and formal specification. The framework distinguishes three different abstraction levels linked to models, metamodels and formal specifications.

We described challenges and open problems in Model Driven Software Modernization. We believe that the adoption of this approach can au-

tomatize crucial steps of a modernization process, reducing costs and development time.

REFERENCES

ADM. (2015). *Architecture-Driven Modernization Task Force*. Retrieved May, 2016, <http://www.omgwiki.org/admtf/doku.php>

Anastasakis, K., Bordbar, B., Georg, G., & Ray, I. (2007). *UML2Alloy: A Challenging Model Transformation. Models 2007. LNCS 4735* (pp. 436–450). Heidelberg, Germany: Springer-Verlag.

ASTM. (2011). *Abstract Syntax Tree Metamodel, version 1.0*. OMG Document Number: formal/2011-01-05. Retrieved May, 2016, <http://www.omg.org/spec/ASTM>

ATL. (2016). *AtlasTransformation Language (ATL)*. Retrieved May, 2016, <http://www.eclipse.org/atl/documentation/>

Barbier, F., Deltombe, G., Parisy, O., & Youbi, K. (2011). Model Driven Engineering: Increasing Legacy Technology Independence. *Second India Workshop on Reverse Engineering (IWRE, 2011) in The 4th India Software Engineering Conference* (pp. 5-10). Thiruvananthapuram, India: CSI ed.

Bidoit, M., & Mosses, P. (2004). *CASL User Manual- Introduction to Using the Common Algebraic Specification Language (LNCS 2900)*. Heidelberg, Germany: Springer-Verlag.

Blu Age Reverse Modeling. (2016). Retrieved on May 2016, www.bluage.com

Boronat, A., & Meseguer, J. (2010). An Algebraic Semantics for MOF. *Formal Aspects of Computing*, 22(3-4), 269–296. doi:10.1007/s00165-009-0140-9

Bruneliere, H., Cabot, J., Dupé, G., & Madiot, F. (2014). MoDisco: A Model Driven Reverse Engineering Framework. *Information and Software Technology*, 56(8), 1012–1032. doi:10.1016/j.infsof.2014.04.007

- Canfora, G., Di Penta, M., & Cerulo, L. (2011). Achievements and Challenges in Software Reverse Engineering. *Communications of the ACM*, 54(4), 142–151. doi:10.1145/1924421.1924451
- CASE MDA. (2016). *Committed Companies and Their Products*. Retrieved May, 2016, www.omg.org/mda/committed-products.htm
- Chikofsky, E., & Cross, J. (1990). Reverse engineering and design recovery: A taxonomy. *IEEE Software*, 7(1), 13–17. doi:10.1109/52.43044
- Deissenboeck, F., & Ratiu, D. (2006). A Unified Meta Model for Concept-Based Reverse Engineering. *3rd International Workshop on Metamodels, Schemes, Grammars, and Ontologies for Reverse Engineering*. Retrieved May, 2016, <http://planet-mde.org/atem2006/atem06Proceedings.pdf>
- Demeyer, S., Ducasse, S., & Nierstrasz, O. (2002). *Object-Oriented Reengineering Patterns*. Amsterdam: Morgan Kaufmann.
- Díaz Bilotto, P., & Favre, L. (2016). Migrating JAVA to Mobile Platforms through HAXE: An MDD Approach. In *Modern Software Engineering Methodologies for Mobile and Cloud Environments* (pp. 240-268). Hershey, PA: IGI Global.
- EMF. (2016). *Eclipse Modeling Framework (EMF)*. Retrieved May, 2016 <http://www.eclipse.org/modeling/emf/>
- Ernst, M. (2003). Static and Dynamic Analysis: Synergy and duality. *ICSE Workshop on Dynamic Analysis (WODA 2003)* (pp. 24-27).
- Fanta, R., & Rajlich, V. (1998). Reengineering object-oriented code. *International Conference on Software Maintenance* (pp. 238-246). Los Alamitos, CA: IEEE Computer Society.
- Favre, L. (2009). *A Formal Foundation for Metamodeling. Reliable Software Technologies, ADA Europe 2009. LNCS* (Vol. 5570, pp. 177–191). Heidelberg, Germany: Springer-Verlag. doi:10.1007/978-3-642-01924-1_13
- Favre, L., & Duarte, D. (2016) Formal MOF Metamodeling and Tool Support. *Proc. 4th International Conference on Model-Driven Engineering and Software Development. MODELSWARD 2016* (pp. 99-110). Roma, Italia: SCITEPRESS (Science and Technology Publications).
- Fleurey, F., Breton, E., Baudry, B., Nicolas, A., & Jézéquel, J. (2007). *Model-Driven Engineering for Software Migration in a Large Industrial Context. MoDELS 2007, LNCS* (Vol. 4735, pp. 482–497). Heidelberg, Germany: Springer-Verlag.
- Fowler, M. (1999). *Refactoring improving the Design of Existing Programs*. Reading: Addison-Wesley.
- Hassam, K., Sadou, S., Le Gloahec, V., & Fleuquin, R. (2011). Assistance System for OCL Constraints Adaptation During Metamodel Evolution. *2011 15th European Conference on Software Maintenance and Reengineering* (pp. 151-160). IEEE Press.
- Jackson, E., Levendosky, T., & Balasubramanian, D. (2011). *Reasoning about Metamodeling with Formal Specifications and Automatic Proofs. Models 2011. LNCS 6981* (pp. 653–667). Heidelberg, Germany: Springer-Verlag.
- Jouault, F., & Bezivin, J. (2006). KM3: A DSL for Metamodeling Specification. In *Proc. 8th IFIP International Conference on Formal Methods for Open Object-Based Distributed Systems (FMOODS 2006). LNCS 4037* (pp. 171-185). Heidelberg, Germany: Springer-Verlag. doi:10.1007/11768869_14
- KDM. (2011). *Knowledge Discovery Metamodel, version 1.3*. OMG Document Number: formal/2011-08-04. Retrieved May, 2016, <http://www.omg.org/spec/KDM/1.3>
- MacDonald, A., Russell, D., & Atchison, B. (2005). Model driven Development within a Legacy System: An industry experience report. *2005 Australian Software Engineering Conference. ASWEC 05* (pp. 14-22). IEEE Press. doi:10.1109/ASWEC.2005.32

MDA. (2016). *The Model-Driven Architecture*. Retrieved May, 2016, <http://www.omg.org/mda/>

Medini. (2012). *Medini QVT*. Retrieved November, 2015, from <http://projects.ikv.de/qvt>

MoDisco. (2016). *Model Discovery*. Retrieved May, 2016, <http://www.eclipse.org/MoDisco>

MOF. (2015). *Meta Object Facility (MOF) Core Specification Version 2.5*. OMG Document Number: formal/2015-08-07. Retrieved May, 2016, <http://www.omg.org/spec/MOF/2.5>

Muller, H., Jahnke, J., Smith, D., Storey, M., Tilley, S., & Wong, K. (2000). *Reverse Engineering: A Roadmap*. 22nd International Conference on Software Engineering (ICSE 2000). Limerick, Ireland: ACM Press. doi:10.1145/336512.336526

QVT. (2015). *MOF 2.0 Query, View, Transformation. Version 1.1*. OMG Document Number: formal/15-02-01. Retrieved May, 2016, <http://www.omg.org/spec/QVT/1.2>

Reus, T., Geers, H., & van Deursen, A. (2006). *Harvesting Software System for MDA-based Reengineering*. LNCS (Vol. 4066, pp. 220–236). Heidelberg, Germany: Springer-Verlag.

Systa, T. (2000). *Static and Dynamic Reverse Engineering Techniques for Java Software Systems* (Ph.D Thesis). University of Tampere, Report A-2000-4.

Tonella, P., & Potrich, A. (2005). *Reverse Engineering of Object Oriented Code*. Monographs in Computer Science. Heidelberg, Germany: Springer-Verlag.

UML. (2015). *OMG Unified Modeling Language. Version 2.5*. OMG Specification formal/2015-03-01. Retrieved May, 2016, <http://www.omg.org/spec/UML/2.5>

Willink, E. (2011). *Modeling the OCL Standard Library*. Electronic Communications of the EASST. Vol. 44. Retrieved May, 2016, <http://journal.ub.tu-berlin.de/eceasst/>

XMI. (2011). *OMG MOF 2 XMI Mapping Specification*OMG. Document Number: formal/2011-08-09. Retrieved May, 2016, <http://www.omg.org/spec/XMI/2.4.1>

ADDITIONAL READING

Brambilla, M., & Cabot, J. M. (2012). *Model-Driven Software Engineering in Practice*. Morgan & Claypool Publishers Synthesis Lectures on Software Engineering.

Favre, L. (2010). *Model Driven Architecture for Reverse Engineering Technologies: Strategic Directions and System Evolution*. Premier Reference Source. IGI Global. doi:10.4018/978-1-61520-649-0

Ulrich, W., & Newcomb, P. (2010). *Information Systems Transformation, Architecture Driven Modernization Case Studies*. Morgan Kaufmann.

KEY TERMS AND DEFINITIONS

ADM (Architecture Driven Modernization):

The process of understanding and evolving existing software assets of a system of interest in the context of MDA.

ATL (Atlas Transformation Language):

A model transformation language and toolkit developed on top of the Eclipse platform that provides ways to produce target models from source models.

MDA (Model Driven Architecture):

An initiative of OMG for the development of software systems based on the separation of business and application logic from underlying platform technologies. It is an evolving conceptual architecture to achieve cohesive model-driven technology specifications.

MoDisco:

A generic, extensible and global Model Driven Reverse Engineering approach and a framework implemented as an official Eclipse project on top of the Eclipse platform.

MOF (Meta-Object Facility): A meta-model that defines a common way for capturing the diversity of modeling standards and interchange constructs involved in MDA.

QVT (Query, View, Transformation): A meta-model for expressing transformation in MDA-based processes.

Reverse Engineering: The process of analyzing available software artifacts such as requirements, design, architectures and code, in order to extract information and provide high-level views of the system.

Mutation Testing Applied to Object-Oriented Languages

S

Pedro Delgado-Pérez
University of Cádiz, Spain

Inmaculada Medina-Bulo
University of Cádiz, Spain

Juan José Domínguez-Jiménez
University of Cádiz, Spain

INTRODUCTION

Mutation testing is a suitable technique to determine the quality of test suites designed for a certain program. This testing technique is based on the creation of *mutants*, that is, versions of the original program with an intentionally introduced fault. Mutations are inserted within the code through some defined rules called *mutation operators*. The underlying idea is that a good set of test cases for the system under test (SUT) should be able to detect any changes generated affecting the behavior of the application.

Test cases are supposed to produce the correct output when they are run on the original program. When the output of a mutant is different from the output of the original program for a test case, the mutation has been revealed and the mutant is classified as *dead*. Otherwise, the mutant is still *alive* and needs to be executed against the rest of the test cases to detect its modification. Hence, if some mutants remain alive after the whole test suite execution, new test cases can be added in order to kill these surviving mutants. However, we classify a surviving mutant as equivalent when the meaning of the program has not actually been modified despite the injected mutation.

Mutation operators represent typical mistakes made when programming and they produce a simple syntactic change in the SUT. Mutation testing is a *white-box* testing technique, i.e., it tests

a program at the source code level. Therefore, the set of mutation operators and the overall technique should be developed around each programming language in particular; the correct choice of the set is one of the keys to successful mutation testing. Thus, we can find an assortment of research studies devoted to the definition of mutation operators for specific languages and tools automating the generation of mutants.

In the same sense, a set of mutation operators can be defined at different levels in each language. Mutation operators mainly dealing with variables, operators or constants were designed for some procedural programs in the early years of the technique. However, other mainstream languages as Java, C# or C++ also include object orientation and completely different mutation operators are needed to test the new structures in these languages. As an example, the operator *IHD* (Hiding Variable Deletion) deletes a variable member in a subclass which is hiding a variable in a parent class:

Original code:

```
class Base{      class Child: public
Base{
public:          public:
    ...          ...
    int v;       int v;
};              };
```

Mutated code:

DOI: 10.4018/978-1-5225-2255-3.ch649

```

class Base{      class Child: public
Base{
public:          public:
    ...          ...
    int v;       /*IHD*/
};              };

```

The purpose of the chapter is to look in depth at the development and the current state of mutation testing, and more specifically, with regard to object-oriented programming languages, in order to widely make known this technique in the computer science research field. Next sections deal with the related work, the steps to accomplish in the mutation testing process, the approaches to evaluate mutation operators and the existing techniques to improve the problems of this technique: equivalent mutant detection, test data generation and the expensive computational cost. Finally, the definition and evaluation of mutation operators for object-oriented languages will be focused.

BACKGROUND

Mutation testing was originally proposed by Hamlet (1977) and DeMillo, Lipton and Sayward (1978) and its development has taken place in parallel with the appearance of the different programming languages (Offutt & Untch, 2001). As a result, in the early years, most of the works centered on procedural programming languages: Agrawal et al. (1989) defined a set of 77 mutation operators for C, the tool *Mothra* was developed including 22 operators to apply mutation testing to Fortran (King & Offutt, 1991) and Offutt and Pan (1996) composed a set of 65 operators for the Ada language. The mutation operators for these procedural languages are known as *traditional* operators.

However, recently, new languages and paradigms have drawn the attention as well as the research has expanded towards other domains (Jia & Harman, 2011). As an illustration, we can find testing tools for rather different languages

like *SQLMutation* for SQL (Tuya, Suárez-Cabal & de la Riva, 2007), *Gamera* for WS-BPEL (Domínguez-Jiménez, Estero-Botaro, García-Domínguez & Medina-Bulo, 2009) or *AjMutator* for AspectJ (Delamare, Baudry & LeTraon, 2009). The existing mutation tools have been enumerated by Jia and Harman (2011). Finally, new mutation frameworks have been also developed lately: *Mutpy* (Derezińska & Halas, 2014) for Phyton 3.x, *Mutant* (n.d.) for Ruby or *PIT* (Van Laeden, 2012) for Java and other JVM languages.

The attention to the object-oriented (OO) paradigm has also risen and several papers and tools have appeared, mainly around Java (Ahmed, Zahoor & Younas, 2010). The first definition of class operators for Java was accomplished by Kim, Clark and McDermid (2000). As exposed in that paper, the aforementioned traditional operators can be applied to test OO programs, but those operators that were developed in programming environments away from this paradigm, do not take into account some types of faults related to features of this kind of programs, so operators at the class level are definitely necessary. Mutation tools including class mutation operators are *MuJava* (Ma, Offutt & Kwon, 2005) for Java, *CREAM* for C# (Derezińska & Szustek, 2009) and *MuCPP* for C++ (Delgado-Pérez, Medina-Bulo, Domínguez-Jiménez, García-Domínguez & Palomo-Lozano, 2015).

All these languages, even though sharing part of the syntax, need a particularized study to define their set of mutation operators and tools to generate the mutants. Mutation testing, usually performed on programs at the unit level, has also been applied at other levels in addition to the class level. Hence, Delamaro, Maldonado and Mathur (2001) studied the technique to be used for integration testing and Mateo, Usaola and Offutt (2012) even to test a complete system. Mutation testing has also been performed on technologies relating the SOA architecture (Bozkurt, Harman & Hassoun, 2013). Furthermore, apart from the code, mutation testing has been used in other domains like the specification of models, such as Finite

State Machines (Fabbri, Delamaro, Maldonado & Masiero, 1994).

As the technique was evolving and it was applied to real-world and bigger applications, it became clearer the barriers that this technique involves, which are discussed by Offutt and Untch (2001): the high computational cost and the time that the user needs to spend, for example, to determine the equivalent mutants. Around these problems have emerged new fields of study so that the technique gets a higher degree of maturation (Usaola & Mateo, 2010; Grun, Schuler & Zeller, 2009). Besides, apart from the mutation score, new calculations are being used to enhance the effectiveness of the technique (Estero-Botaro, Palomo-Lozano, Medina-Bulo, Domínguez-Jiménez & García-Domínguez, 2014).

MUTATION TESTING OVERVIEW

Mutation Testing Steps

The mutation testing process consists basically in three main steps: the analysis of the program, the generation of mutants and the execution of the test suite. Before performing these phases, the correctness of the SUT is checked by executing the test cases on the program. If one of the outputs is incorrect, the tester needs to fix the program until all test cases are successfully executed. Then, the tester can start the process described below:

- **Analysis:** In this first step, the aim is to determine where in the code each mutation

operator can be applied. In other words, the analysis allows us to know the mutants that can be created from the code.

- **Generation of Mutants:** Once the tester finds out that a fault can be introduced, a new program is produced containing that syntactic change, i.e., the mutant is in fact generated.
- **Execution of the Test Suite:** The created mutant can be run against the test suite now. After the execution, the mutant is classified as dead or alive depending on the output.

Afterwards, the mutation score is computed (previously analyzing the equivalence), indicating if some of the mutants are still alive. Then, the tester has to decide whether to create new test cases to try to increase the mutation score or the result passes an established threshold. If new test cases are included, the process has to be performed again and again until the tester is satisfied.

The commented overall process is shown in Figure 1. Moreover, the test suite can be reduced when analyzing the results. In this reduction, the test suite should be non-redundant (a test case is not useful if other test cases detect the same mutants) and adequate (every mutant is killed by at least one test case) (Mresa & Botacci, 1999).

Quality of Mutation Operators

Evaluating the used mutation operators to obtain a test suite with quality is an important issue in this process and several authors have dealt with this aspect. The *mutation score* is the most common

Equation 1. General calculation of mutation score

$$MS(S,C) = \frac{KM}{TM-EM} \times 100$$

MS: Mutation score

S: SUT

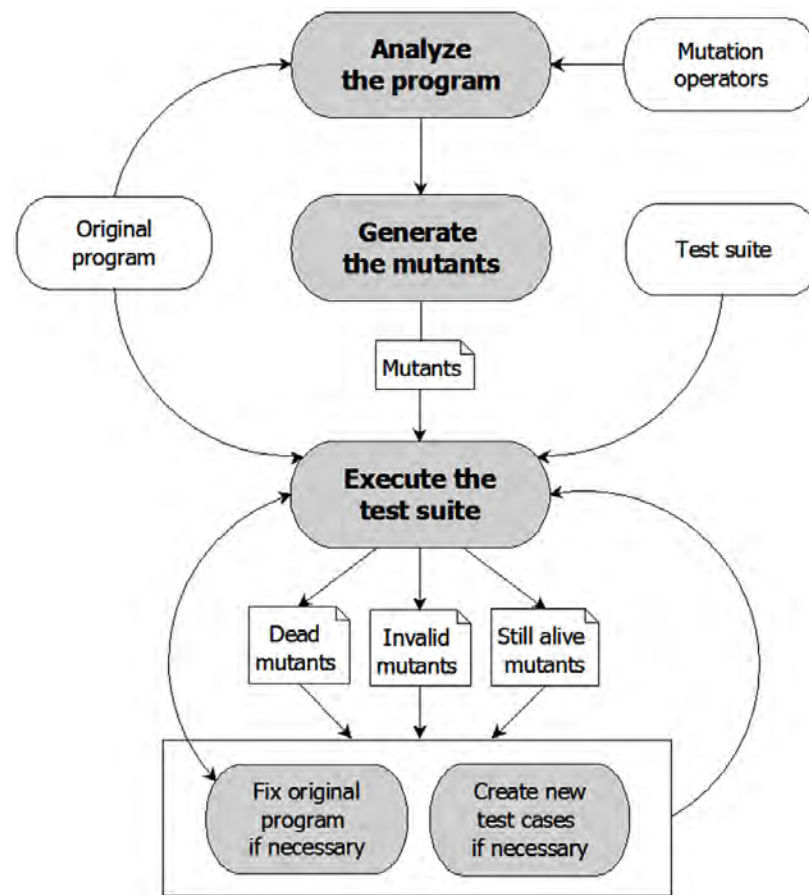
C: Test cases

KM: Killed mutants

TM: Total mutants

EM: Equivalent mutants

Figure 1. General mutation testing process



criterion as this score is calculated to determine the test suite effectiveness distinguishing the mutants (see Equation 1); when all the mutants are killed ($MS=100$), we obtain an adequate test suite for the applied operators.

Mresa and Botacci (1999) used the mutation score in each particular mutation operator, while some studies have considered other dimensions like the number and kind of mutants generated. As a matter of fact, it is preferable not to use operators that generate a large number of equivalent mutants or we have to take care of operators that produce a great amount of mutants as they can entail much computation.

Smith and Williams (2009) proposed to determine the quality of operators by relating the possible mutant states after the test suite execution:

killed by the initial test suite, killed by a new test case, killed by a new test case specifically defined to kill another mutant or not killed. Likewise, Derezińska (2006) exposed an idea of the effectiveness regarding the class mutation operators (specifically for C#) as well as of the test cases for an operator. Derezińska and Rudnik (2012) analyzed both traditional and class operators depending on their influence in the mutation score, and Moghadam and Babamir (2014) measured the mutation score in terms of object-oriented metrics reflecting the structural complexity of a program.

Finally, Estero-Botaro et al. (2014) presented a mathematical formula to measure the operator quality. In short, the quality metric Q_m favors not only those mutants killed by few tests (therefore, the mutation may be difficult to detect), but also

when those few tests kill few mutants (therefore, the tests killing the mutant may be difficult to design analyzing other mutants).

Problems of Mutation Testing

Two are the chief remaining barriers for the definitive practical usage of the technique in the industry: the great amount of computation and the manual time spent by the user. Hence, many studies have been undertaken providing different solutions for these problems in the last years. In the case of the computational expense, several techniques have appeared reducing both the execution cost and the number of mutants generated (Usaola & Mateo, 2010). Within the former, we can stand out the following choices:

- **Weak Mutation:** Comparing the internal state of the mutant and the SUT right after the mutated fragment (Girgis & Woodward, 1985).
- **Firm Mutation:** Less strict than weak mutation, comparing the intermediate state after the execution and the final output (Woodward & Halewood, 1988).
- **Runtime Optimizations:** Such as the *meta-mutant* technique encoding all the mutants in a single program (Untch, Offutt & Harrold, 1993), or specific compilers for faster executions.

Attending the reduction of mutants, we can find:

- **Mutant Sampling:** Only executing some of the available mutants randomly (Budd, 1980).
- **Selective Mutation:** Applying a subset of the defined mutation operators (Mresa & Bottaci, 1999).
- **Mutant Clustering:** Grouping the mutants according to the set of test cases that kill them and then selecting a small number of mutants for each cluster (Hussain, 2008).

- **Higher Order Mutation:** Where the mutant contains more than a single fault (Jia & Harman, 2009).
- **Evolutionary Mutation Testing:** Generating a selected subset of mutants through an evolutionary algorithm (Domínguez-Jiménez, Estero-Botaro, García-Domínguez & Medina-Bulo, 2011).

While the problem of the cost is practically overcome nowadays, the lack of automation is not resolved in spite of the advances. Firstly, the automatic generation of test cases needs to be enhanced so that their creation is not manual. *Constrained-based test data generation* (CBT) is one of the techniques developed, consisting in the creation of values that satisfies certain constraints (DeMillo & Offutt, 1991). Some of their problems were alleviated in the *dynamic domain reduction procedure* (DDR) (Offutt, Jin & Pan, 1999). More recently, new approaches have been exposed, as the promising study to create new test cases with the goal of killing mutants in object-oriented systems (Fraser & Zeller, 2012) or the application of *search based testing* to mutation testing (Papadakis & Malevris, 2013).

Secondly, the analysis of the still alive mutants after the test suite execution to determine those that are equivalent supposes another time consuming labor. This task remains undecided, but it has been tried to be reduced in several works. Baldwin and Sayward (1979) first suggested compiler optimization techniques. Other techniques tackling with this problem are the usage of a *co-evolutionary* algorithm to cut out equivalent mutants in the process through a fitness function (Adamopoulos, Harman & Hierons, 2004), or the analysis of changes in the code coverage (Schuler & Zeller, 2010). Recently, the study by Papadakis, Jia, Harman and Le Traon (2015) succeeded in reducing equivalent mutants by comparing the executables of the original programs and the mutants.

MUTATION TESTING AT THE CLASS LEVEL

Class Mutation Operators for Object-Oriented Programming Languages

As far as the authors know, class mutation operators have been defined for three object-oriented programming languages; Java, C# and C++. The first operators at this level were defined for Java by Kim et al. (2000) and completed by Ma, Kwon and Offutt (2002). Derezińska (2006) and later Delgado-Pérez et al. (2015) defined class operators for C# and C++ respectively based on Java operators, but also presented specific operators for those languages.

Mutation operators are usually gathered in groups because of their structural similarities. Regarding operators at the class level, these are the main categories:

- **Encapsulation or Information Hiding:** Checking the correct accessibility.
- **Inheritance:** Between classes, mainly in presence of overriding members.
- **Polymorphism and Dynamic Binding:** Testing the proper functioning of these properties.
- **Method Overloading:** Confirming the right method is being invoked.

A general list of operators for these categories is shown in Table 1. Most of these operators can be applied to these programming languages, with exceptions such as *AMC* or *IOR* for C# and *JSD*, *JSI* or *EOC* for C++. Moreover, many of these operators have been appropriately tailored to the specifics of each language, changing the scope of application or even the name and the meaning. The concrete differences in the operators among these languages have been commented by Derezińska (2006) in the case of C#, and by Delgado-Pérez et al. (2015) regarding C++.

Table 1. Common class mutation operators

Category	Operator	Description
Encapsulation	<i>AMC</i>	Access modifier change
Inheritance	<i>IHD</i>	Hiding variable deletion
	<i>IHI</i>	Hiding variable insertion
	<i>ISD</i>	<i>super</i> keyword deletion
	<i>ISI</i>	<i>super</i> keyword insertion
	<i>IOD</i>	Overriding method deletion
	<i>IOP</i>	Overriding method calling position change
	<i>IOR</i>	Overriding method rename
	<i>IPC</i>	Explicit call of a parent's constructor deletion
Polymorphism and dynamic binding	<i>PCD</i>	Type cast operator deletion
	<i>PCI</i>	Type cast operator insertion
	<i>PCC</i>	Type cast operator change
	<i>PMD</i>	Member variable declaration with parent class type
	<i>PPD</i>	Parameter variable declaration with child class type
	<i>PNC</i>	<i>new</i> method call with child class type
	<i>PRV</i>	Reference assignment with other compatible type
Method overloading	<i>OMD</i>	Overloading method deletion
	<i>OMR</i>	Overloading method contents replace
	<i>OAN</i>	Argument number change
	<i>OAO</i>	Argument order change
Miscellany	<i>JTD</i>	<i>this</i> keyword deletion
	<i>JTI</i>	<i>this</i> keyword insertion
	<i>JSD</i>	<i>static</i> modifier deletion
	<i>JSI</i>	<i>static</i> modifier insertion
	<i>JID</i>	Member variable initialization deletion
	<i>JDC</i>	Default constructor creation
	<i>EOA</i>	Reference assignment and content assignment replacement
	<i>EOC</i>	Reference comparison and content comparison replacement

Table 2. Mutation operators within the exception handling category

Category	Operator	Description
Exception handling	<i>EHC</i>	Exception handling change
	<i>EHR</i>	Exception handler removal
	<i>EXS</i>	Exception swallowing

Table 3. Specific class mutation operators

Category	Operator	Description
Java-specific	<i>EAM</i>	Accessor method change
	<i>EMM</i>	Modifier method change
C++-specific	<i>IMR</i>	Multiple inheritance replacement
	<i>PVI</i>	<i>virtual</i> modifier insertion
	<i>CDD</i>	Destructor method deletion
	<i>CCA</i>	Copy constructor and assignment operator overloading deletion
C#-specific	<i>PRM</i>	Property replacement with member field
	<i>OID</i>	Overriding indexer deletion
	<i>OPD</i>	Overriding property deletion
	<i>IOK</i>	Override keyword substitution
	<i>DEH</i>	Method delegated for event handling change
	<i>DMO</i>	Delegated method order change
	<i>DMC</i>	Delegated method change

A further group, exception handling, has also been considered as a class-level category (Kim et al., 2000) because exceptions are related with the object-oriented paradigm (see Table 2). As aforementioned, several mutation operators have also been specified for each of these languages, which are presented in Table 3. The Java-specific operators shown in that table are also applicable to C#; still, because of the *property* feature, *EAM* and *EMM* have a limited importance in that language.

Different evaluations about these operators have yielded two main conclusions. First, the mutants created with operators at the class-level are fewer than with traditional operators. Second, the equivalence problem is a more important issue using class operators than traditional ones.

The study by Ma, Kwon & Kim (2009) showed that there were almost twice as many traditional mutants as class mutants for the same classes. In those experiments class operators produced an equivalence percentage over 70%, unlike the 5-15% usually found with traditional operators. Nevertheless, this high equivalence percentage was not confirmed by later studies; in the experiments conducted by Segura, Hierons, Benavides & Ruiz-Cortés (2011), about 45% of the mutants were equivalent and Delgado-Pérez et al. (2015) obtained 37% of equivalence percentage in their study for C++.

FUTURE RESEARCH DIRECTIONS

In brief, mutation testing needs more research to resolve the main objections to the usage of the technique. Regarding the equivalence detection, the future research may address the redefinition or reimplementation of the mutation operators to try to avoid the creation of equivalent mutants. Moreover, in general terms, the currently most used programming languages need more support as well as the most promising languages for the future. In this way, mutation testing can be a real choice in the testing phase and start being widely accepted and having a practical approach.

In the case of class mutation operators, they have not been analyzed at the same extent as traditional operators in the literature, so new studies on the significance of these operators are needed. In addition, an individual evaluation of the operators could be accomplished to obtain a sufficient set of operators at this level.

CONCLUSION

Along this chapter, we have tackled the main topics related to mutation testing, giving details about studies focusing object orientation. The structures related to this paradigm require a tailored analysis and a set of mutation operators at the class level

were defined to that end. These class operators have been defined for three languages, where we can find both common and specific operators for each language.

Although the amount of generated mutants is lower using class operators than traditional ones, equivalence is even a greater problem. Still, important recent advances like the automatic generation of test cases to kill mutants (Fraser & Zeller, 2012) and the reduction of equivalent mutants (Papadakis et al., 2015) are bridging the gap between research and uptake in the industry.

REFERENCES

- Adamopoulos, K., Harman, M., & Hierons, R. M. (2004). How to Overcome the Equivalent Mutant Problem and Achieve Tailored Selective Mutation Using Co-evolution. In *GECCO (2)'04* (pp. 1338-1349). doi:10.1007/978-3-540-24855-2_155
- Agrawal, H., DeMillo, R., Hathaway, B., Hsu, W., Hsu, W., Krauser, E., & Spafford, E. (1989, March). *Design of mutant operators for the C programming language*. Technical Report SERC-TR-41-P. Software Engineering Research Center, Purdue University, West Lafayette, IN.
- Ahmed, Z., Zahoor, M., & Younas, I. (2010, February). Mutation operators for object-oriented systems: A survey. In *Computer and Automation Engineering (ICCAE), 2010 the 2nd International Conference on* (Vol. 2, p. 614-618). doi:10.1109/ICCAE.2010.5451692
- Baldwin, D., & Sayward, F. (1979). *Heuristics for Determining Equivalence of Program Mutations* (Research Report No. 276). Department of Computer Science, Yale University.
- Bozkurt, M., Harman, M., & Hassoun, Y. (2013). Testing and verification in service-oriented architecture: A survey. *Software Testing, Verification and Reliability*, 23(4), 261–313. doi:10.1002/stvr.1470
- Budd, T. A. (1980). *Mutation analysis of program test data* (PhD thesis). Yale University, New Haven, CT.
- Delamare, R., Baudry, B., & Le Traon, Y. (2009, April). AjMutator: a tool for the mutation analysis of AspectJ pointcut descriptors. In *International Conference on Software Testing, Verification and Validation Workshops, 2009. ICSTW '09* (pp. 200–204). doi:10.1109/ICSTW.2009.41
- Delamaro, M. E., Maldonado, J. C., & Mathur, A. P. (2001, March). Interface Mutation: An Approach for Integration Testing. *IEEE Transactions on Software Engineering*, 27(3), 228–247. doi:10.1109/32.910859
- Delgado-Pérez, P., Medina-Bulo, I., Domínguez-Jiménez, J. J., García-Domínguez, A., & Palomo-Lozano, F. (2014). Class mutation operators for C++ object-oriented systems. *Annales des Télécommunications*, 70(3-4), 137–148. doi:10.1007/s12243-014-0445-4
- DeMillo, R., Lipton, R., & Sayward, F. (1978, April). Hints on Test Data Selection: Help for the Practicing Programmer. *Computer*, 11(4), 34–41. doi:10.1109/C-M.1978.218136
- DeMillo, R., & Offutt, A. (1991). Constraint-based automatic test data generation. *Software Engineering. IEEE Transactions on*, 17(9), 900–910. doi:10.1109/32.92910
- Derezińska, A. (2006, October). Quality Assessment of Mutation Operators Dedicated for C# Programs. In P. Kellenberger (Ed.), *Proceedings of VI International Conference on Quality Software* (pp. 227–234). IEEE Computer Societydoi:10.1109/QSIC.2006.51
- Derezińska, A., & Halas, K. (2014). Analysis of Mutation Operators for the Python Language. *Advances in Intelligent Systems and Computing: Vol 286. Proceedings of the Ninth International Conference on Dependability and Complex Systems DepCoS-RELCOMEX* (pp. 155-164). doi:10.1007/978-3-319-07013-1_15

- Derezińska, A., & Rudnik, M. (2012). Quality evaluation of object-oriented and standard mutation operators applied to C# programs. In *50th International Conference on on Objects, Models, Components, Patterns TOOLS 2012* (pp. 42–57). doi:10.1007/978-3-642-30561-0_5
- Derezińska, A., & Szustek, A. (2009). Object-oriented testing capabilities and performance evaluation of the C# mutation system. In *4th IFIP TC2 Central and East European Conference on Software Engineering Techniques, CEE-SET 2009* (pp. 229–242). doi:10.1007/978-3-642-28038-2_18
- Domínguez-Jiménez, J., Estero-Botaro, A., García-Domínguez, A., & Medina-Bulo, I. (2009, November). GAmara: An Automatic Mutant Generation System for WS-BPEL compositions. In R. Eshuis, P. Grefen, & G. A. Papadopoulos (Eds.), *Proceedings of the 7th IEEE European Conference on Web Services* (pp. 97–106). Eindhoven, The Netherlands: IEEE Computer Society Press. doi:10.1109/ECOWS.2009.18
- Domínguez-Jiménez, J., Estero-Botaro, A., García-Domínguez, A., & Medina-Bulo, I. (2011, October). Evolutionary mutation testing. *Information and Software Technology*, 53(10), 1108–1123. doi:10.1016/j.infsof.2011.03.008
- Estero-Botaro, A., Palomo-Lozano, F., Medina-Bulo, I., Domínguez-Jiménez, J. J., & García-Domínguez, A. (2015). Quality metrics for mutation testing with applications to WS-BPEL compositions. *Software Testing. Verification and Reliability*, 25(5-7), 536–571. doi:10.1002/stvr.1528
- Fabbri, S. C. P. F., Delamaro, M., Maldonado, J., & Masiero, P. (1994). Mutation Analysis Testing for Finite State Machines. In *Software Reliability Engineering, 1994. Proceedings 5th International Symposium on* (p. 220-229). doi:10.1109/ISSRE.1994.341378
- Fraser, G., & Zeller, A. (2012). Mutation-Driven Generation of Unit Tests and Oracles. *IEEE Transactions on Software Engineering*, 38(2), 278–292. doi:10.1109/TSE.2011.93
- Girgis, M. R., & Woodward, M. R. (1985). An integrated system for program testing using weak mutation and data flow analysis. In *Proceedings of the 8th International Conference on Software Engineering* (pp. 313–319). Los Alamitos, CA: IEEE Computer Society Press. Retrieved from <http://dl.acm.org/citation.cfm?id=319568.319662>
- Grun, B., Schuler, D., & Zeller, A. (2009, April). The Impact of Equivalent Mutants. In *Software Testing, Verification and Validation Workshops, 2009. ICSTW '09. International Conference on* (pp. 192 -199). doi:10.1109/ICSTW.2009.37
- Hamlet, R. G. (1977, July). Testing Programs with the Aid of a Compiler. *IEEE Transactions on Software Engineering*, 3(4), 279–290. doi:10.1109/TSE.1977.231145
- Hussain, S. (2008). *Mutation clustering*. King's College London. Retrieved 2012-10-25, from <http://www0.cs.ucl.ac.uk/staff/m.harman/Past-MScProjects2007/ShamailaHussain.pdf>
- Jia, Y., & Harman, M. (2009). Higher Order Mutation Testing. *Information and Software Technology*, 51(10), 1379–1393. doi:10.1016/j.infsof.2009.04.016
- Jia, Y., & Harman, M. (2011, October). An Analysis and Survey of the Development of Mutation Testing. *Software Engineering. IEEE Transactions on*, 37(5), 649–678. doi:10.1109/TSE.2010.62
- Kim, S., Clark, J., & McDermid, J. (2000). Class mutation: Mutation testing for object-oriented programs. In *Proceedings of the Net.ObjectDays* (pp. 9–12).
- King, K. N., & Offutt, A. J. (1991). A FORTRAN Language System for Mutation-based Software Testing. *Software, Practice & Experience*, 21(7), 685–718. doi:10.1002/spe.4380210704

- Ma, Y.-S., Kwon, Y. R., & Kim, S.-W. (2009, April). Statistical Investigation on Class Mutation Operators. *ETRI Journal*, 31(2), 140–150. doi:10.4218/etrij.09.0108.0356
- Ma, Y. S., Kwon, Y. R., & Offutt, J. (2002, November). Inter-Class Mutation Operators for Java. In S. Kawada (Ed.), *Proceedings of XIII International Symposium on Software Reliability Engineering* (pp. 352–363). Annapolis, MD: IEEE Computer Society.
- Ma, Y. S., Offutt, J., & Kwon, Y. R. (2005). MuJava: An automated class mutation system. *Software Testing. Verification and Reliability*, 15(2), 97–133. doi:10.1002/stvr.308
- Mateo, P. R., Usaola, M. P., & Offutt, J. (2012). Mutation at the multi-class and system levels. *Science of Computer Programming*, (0). doi:10.1016/j.scico.2012.02.005
- Moghadam, M. H., & Babamir, S. M. (2014). Mutation score evaluation in terms of object-oriented metrics. In *4th International eConference on Computer and Knowledge Engineering (ICCKE)* (pp. 775–780). Mashhad: IEEE. doi:10.1109/ICCKE.2014.6993419
- Mresa, E. S., & Bottaci, L. (1999). Efficiency of mutation operators and selective mutation strategies: An empirical study. *Software Testing. Verification and Reliability*, 9(4), 205–232. doi:10.1002/(SICI)1099-1689(199912)9:4<205::AID-STVR186>3.0.CO;2-X
- Mutant: Mutation testing tool for Ruby. (n.d.). Retrieved from <http://rubymdoc.info/gems/mutant/>
- Offutt, A., & Pan, J. (1996, June). Detecting equivalent mutants and the feasible path problem. In *Computer Assurance, 1996. COMPASS '96, Systems Integrity. Software Safety. Process Security. Proceedings of the Eleventh Annual Conference on* (p. 224-236). doi:10.1109/COMPASS.1996.507890
- Offutt, A. J., Jin, Z., & Pan, J. (1999, February). The dynamic domain reduction procedure for test data generation. *Software, Practice & Experience*, 29(2), 167–193. doi:10.1002/(SICI)1097-024X(199902)29:2<167::AID-SPE225>3.0.CO;2-V
- Papadakis, M., Jia, Y., Harman, M., & Le Traon, Y. (2015). Trivial Compiler Equivalence : A Large Scale Empirical Study of a Simple, Fast and Effective Equivalent Mutant Detection Technique. *Proceeding of the 37th International Conference on Software Engineering (ICSE 2015)*, 936–946. doi:10.1109/ICSE.2015.103
- Papadakis, M., & Malevris, N. (2013). Searching and generating test inputs for mutation testing. *SpringerPlus*, 2(1), 121. doi:10.1186/2193-1801-2-121 PMID:23607084
- Schuler, D., & Zeller, A. (2010). (Un-)Covering Equivalent Mutants. In *2010 Third International Conference on Software Testing, Verification and Validation* (pp. 45–54). IEEE. doi:10.1109/ICST.2010.30
- Segura, S., Hierons, R. M., Benavides, D., & Ruiz-Cortés, A. (2011). Mutation testing on an object-oriented framework: An experience report. *Information and Software Technology*, 53(10), 1124–1136. doi:10.1016/j.infsof.2011.03.006
- Smith, B., & Williams, L. (2009). On guiding the augmentation of an automated test suite via mutation analysis. *Empirical Software Engineering*, 14(3), 341–369. doi:10.1007/s10664-008-9083-7
- Tuya, J., Suárez-Cabal, M. J., & de la Riva, C. (2007). Mutating database queries. *Information and Software Technology*, 49(4), 398–417. doi:10.1016/j.infsof.2006.06.009
- Untch, R. H., Offutt, A. J., & Harrold, M. J. (1993). Mutation analysis using mutant schemata. In *Proceedings of the 1993 ACM SIGSOFT International Symposium on Software Testing and Analysis* (pp. 139–148). New York, NY: ACM. doi:10.1145/154183.154265

Usaola, M., & Mateo, P. (2010). Mutation Testing Cost Reduction Techniques: A Survey. *Software, IEEE*, 27(3), 80–86. doi:10.1109/MS.2010.79

Van Laenen, F. (2012). Mutation Testing. *Overload*, 108, 16–22. Retrieved from <http://accu.org/var/uploads/journals/overload108.pdf>

Woodward, M., & Halewood, K. (1988). From weak to strong, dead or alive? An analysis of some mutation testing issues. In *Proceedings of the Second Workshop on Software Testing, Verification, and Analysis* (p. 152–158). doi:10.1109/WST.1988.5370

Estero-Botaro, A., Palomo-Lozano, F., & Medina-Bulo, I. (2010). Quantitative Evaluation of Mutation Operators for WS-BPEL Compositions. In *Software Testing, Verification, and Validation Workshops (ICSTW), 2010 third international conference on* (p. 142–150). doi:10.1109/ICSTW.2010.36

Horstmann, C., & Budd, T. (2009). *Big C* (2nd ed.). Wiley.

Woodward, M. R. (1993, March). Mutation testing - its origin and evolution. *Information and Software Technology*, 35(3), 163–169. doi:10.1016/0950-5849(93)90053-6

ADDITIONAL READING

Boubeta-Puig, J., Medina-Bulo, I., & García-Domínguez, A. (2011). Analogies and differences between mutation operators for WS-BPEL 2.0 and other languages. In *Software Testing, Verification and Validation Workshops (ICSTW), IEEE Fourth International Conference on* (p. 398–407). doi:10.1109/ICSTW.2011.52

Chevalley, P. (2001, December). Applying mutation analysis for object-oriented programs using a reflective approach. In *Software Engineering Conference, 2001. Eighth Asia-Pacific* (p. 267–270). doi:10.1109/APSEC.2001.991487

DeMillo, R. (1989). Test Adequacy and Program Mutation. In *Software Engineering, 1989. 11th International Conference on* (p. 355–356). doi:10.1109/ICSE.1989.714449

Derezinska, A., & Kowalski, K. (2011). Object-oriented mutation applied in common intermediate language programs originated from C#. In *Software Testing, Verification and Validation Workshops (ICSTW), 2011 IEEE Fourth International Conference in* (pp. 342–350). doi:10.1109/ICSTW.2011.54

KEY TERMS AND DEFINITIONS

Alive Mutant: The mutant that produces the same output as the original program for a certain test case.

Class Mutation Operator: Mutation operator that injects the mutation regarding structures related to the object-oriented paradigm.

Dead Mutant: The mutant that produces a different output from the original program for a certain test case.

Equivalent Mutant: The mutant whose introduced change does not modify the meaning of the original program.

Mutation Operator: Represents a typical mistake produced by a programmer and it is used to introduce a simple syntactic change in the code to generate a mutant.

Mutation Score: The calculation to measure the quality of a test suite detecting the introduced faults in the mutants. It is defined as the quotient between the number of dead mutants and the number of non-equivalent mutants.

Mutation Testing (or Mutation Analysis or Program Mutation): A technique used to evaluate the quality of a test suite in the detection of faults and create new test cases.

Object–Oriented Programming in Computer Science

Rahime Yilmaz

Istanbul University, Turkey

Anil Sezgin

Yildiz Technical University, Turkey

Sefer Kurnaz

Istanbul Esenyurt University, Turkey

Yunus Ziya Arslan

Istanbul University, Turkey

INTRODUCTION

In computer science, a program is composed of a series of commands, which runs within a computer or an electronic circuit, producing information for users. Programming is a that can help programmers while writing a program. Computer programming is the process of writing an algorithm and, it is also the encoding of the algorithm into a notation that can produce and provide information to the users. It can be classified into two groups, that is, system programming and application programming. System programming is a sub branch of the general programming that is composed of low level instructions which are used to operate and handle computer hardware. Application programming is considered as the improved version of the computer programs which can perform specific tasks for the users. One of the application programming types is the object oriented programming (OOP) which is about how information is represented in human mind.

As a computer programming approach, OOP is useful such that it provides easy modeling in designing and developing real entities. This approach is intended to model the entities and, also, the relationships existing between them. OOP allows programmers to define the required classes

to create the objects and to apply modifications (manipulations) on them. It can also supply inheritance, polymorphism and encapsulation features to the developers. With these capabilities, the processed data can be isolated from other redundant applications. Because of its abilities that are readily available to the users, OOP is preferred much more than other available programming languages. The inherent properties of OOP, which do not exist in other application programming, can be stated as modularity, extensibility and reusability. This chapter provides a substantial survey of OOP in computer science.

In this chapter, we have highlighted a number of explanations and reviews that are generally accepted and are in common use in OOP. We explain the heart of this chapter OOP concept in section 1, Object Oriented Programming Features, making up the largest section. Main topic of OOP which are included Inheritance, Polymorphism, Abstraction and Encapsulation titles are explained with details in the subtitles of section 1. In section 2, you can find an example of OOP implementation in Java. There are many kinds of OOP languages in use but in this study, Java was given as a strong example to OOP language (Harel, Marron, & Weiss, 2010). The following section is Future Research Directions which include future works related OOP. The

chapter ends with Conclusion section that gives a brief information about this study.

BACKGROUND

Programming paradigm is a fundamental style of computer programming which classifies programming languages. Different programming paradigms were developed by considering the concepts and abstraction which are used to represent the elements of a program, and steps that compose a computation. Some of the programming languages are designed to support one paradigm and some of them support multiple paradigms. Before OOP languages, there are also some other paradigms such as classic programming, modular programming and structural programming (Bista, Bajracharya, & Dongol, 2015). These programming techniques were helped the programmers while solving their problems. Depending on improving technology, new structure of OOP has emerged so one of these paradigms is OOP which changed radically the programming paradigm continued until the day it appeared. Software methodology used before OOP referred to by the name of the procedural programming. This methodology was based on advancing codes in a particular direction and calling the common function that is used to reduce the workload. This methodology, used in the software world for a long time, has some difficulties. First of all, application of the procedural programming developed as a whole cannot be divided. So, each developer working on the application has to know almost every building of application. Due to its building as a whole, it is hard to make changes on the application. The reason of these difficulties is that procedural programming is an abstract that is not able to model the real world. The real world can be simulated by programmers thanks to OOP by using objects and classes. Object-oriented approach enables to divide a complex system into smaller parts and manageable modules which makes development process easier to grasp and share among members

of a developer team and easier to communicate to users who are needed to provide requirements and confirm how well the system meets the requirements throughout the process (Dennis, Wixom, & Tegarden, 2015). Thus, OOP enables modularity and abstraction with increased code understanding, maintenance and expansion. OOP is formed by the collection of objects which communicate with each other in order to perform tasks. This communication is based on messages in OOP, and objects are created from classes.

Class is a blueprint which is used to define objects describing the contents of the objects itself. It is a user-defined prototype for an object that defines a set of attributes and methods which characterize any object of the class. These attributes are data members or variables (static attributes), and methods are dynamic behaviors, also called member functions. Data members and member functions are called class members. Attributes, which are attached to the classes, store information about the object (Robson, 1981). Data member has a name and a type; it holds a value of that type. Member function receives parameters from the caller (if it's required), performs the tasks which are defined in the function body and returns result or void to the caller.

In most of the programming languages, class keyword is used to define a class. Class declaration must contain the name of the class which programmer declares. Basic class declaration looks like this:

```
Class NameOfClass {
...
}
```

The other term, object, helps users to understand the object oriented notion. Objects, also called instances of a class, are modeled on real world entities. All the instances of a class have similar properties. Basically, objects have 2 characteristics, state and behavior. State is a well-defined condition of an item which captures the relevant aspects of an object. Behavior is the observable

Figure 1. A cat object and its methods



effects of an operation or event. For example, a software object that modeled real world cat would have variables that indicated the cat's current state:

- Its breed is Cymric
- Its color is Yellow
- Its age is 3.

These variables are formally known as instance variables because they contain the state for a particular cat object, and in the object oriented terminology, a particular object is called an instance. In addition to its variables, the software cat would also have methods from real cat's behaviors: to meow, to wag the tail, play the ball etc. Figure 1 shows three different method of cat object.

A class may contain a variable that is shared by all instances of class. These variables are defined within a class. Programmer can create new objects which can inherit the number of characteristics from one of the existing objects. Objects can communicate with each other by passing message. A

message is a method call from sender object to a receiver object. Object responds to a message by executing one of its methods (Legdarg, 1996). Additional information, along with the called arguments, may accompany a method call. A message includes three components. These components are object identifier which indicates the message receiver, method name, and arguments which can be required for the execution of the method. In the example of cat.play (cp), cat is an object and the receiver of the message, play is the method, and cp is the argument of play.

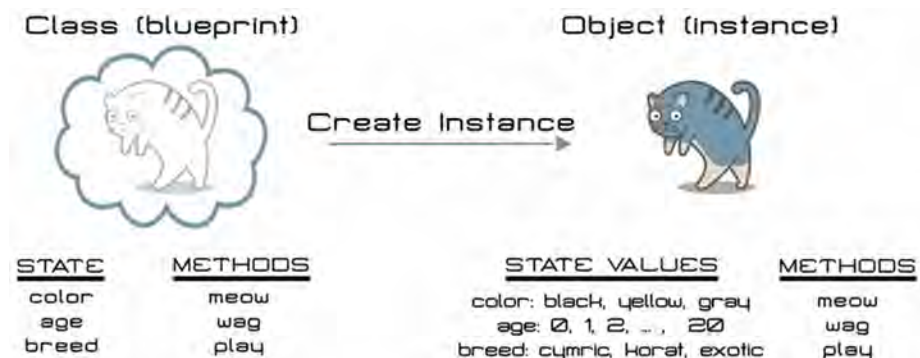
The basic implementation of 'cat' example is given in program 2 of Implementation of OOP Design in Java Programming section and this example is illustrated in Figure 2 with detail which include a diagram of cat class, cat class to address issues regarding the states, methods, attributes, etc. used in OOP.

1. OBJECT ORIENTED PROGRAMMING FEATURES

OOP has many features that make it usable including the followings:

- Inheritance
- Polymorphism
- Abstraction
- Encapsulation

Figure 2. Illustration of cat object



1.1. Inheritance

This concept points code reuse without repeating or rewriting the code. It includes the creation of new classes from existing classes, and these new classes are called derived class or sub class. The existing class is called the base class or parent class or super class. The derived class reuses the base class members, moreover it can add and alter them. So programmers can avoid rewriting and testing of code that already exist. The goal of inheritance is to support refinement of classes into derived classes or subclasses. There are many advantages of using inheritance concept in programming language. Direct modeling of such hierarchies makes the conceptual structure of programs easier to comprehend, inheritance supports that common properties of classes factorized that is described once and reused when needed (Wirfs-Brock & Johnson, 1990). This results in modularity and makes complicated programs easier to comprehend and maintain, since description is avoided. Inheritance hierarchies support a technique where the most general classes containing common properties of different classes. These classes are designed and verified first, and then, specialized classes are developed by adding more details to existing classes (Thomsen, 1986).

Inheritance implements the “Is-A” relationship. In OOP, Is-A relationship means that one object

is type of another one. For example, student is a person, car is a vehicle etc.

A class inherits instance variable declarations as well as methods from its base class. By adding new instance variables and new methods, and by overriding base class methods, the derived class’s attributes may be redefined. A derived class cannot access the private members of its base class; otherwise this situation would be against the concept of encapsulation. A derived class has access to the public and protected members of its base class. They are both inheritable and visible to users.

There are different types of inheritance that depends on programming language. *i)* Single Inheritance: A derived class, which is inherited properties and behaviors from a single base class, is called single inheritance. *ii)* Multi-Level Inheritance: If a class is derived from another derived class, it is called multi-level inheritance. *iii)* Hierarchical Inheritance: If more than one class is derived from a base class, it is called hierarchical inheritance. *iv)* Hybrid Inheritance: It is a combine form of single and multiple inheritance. *v)* Multiple Inheritance: If a class is derived from more than one class, it is called multiple inheritance.

1.2. Polymorphism

Dictionary definition of polymorphism is the property of something having many forms. In com-

Figure 3. A real world example of inheritance



puter science, polymorphism refers to the ability of a programming language to describe objects in different ways based on their class or data type (Gamma, Helm, Johnson, & Vlissides, 1994). It can be allowed to talk to an object even if it is not known exactly what the object is. Thanks to the polymorphism, size of programming applications can be made smaller. In addition, understanding these applications is easier than those of others. If polymorphism does not exist, programmers have to check the objects one by one to determine which type and method are called according the object type. Polymorphism is activated in such situations that frees the designer from this inconvenience and allows flexibility. To better understand the concept of polymorphism, it is needed to comprehend the inheritance well. Polymorphism is tied closely to the concept of inheritance in OOP languages. There is a simple rule that is called as “is-a”, to know whether or not inheritance is the right design for user’s data. This rule states that every object of the subclass is an object of the superclass. To explain the relationship between inheritance and polymorphism, an example is given in Figure 3. The figure represents a transportation class with subclasses including corresponding modes (sea, road, air, rail) and vehicles. For example, all the road vehicles are a vehicle type. Thus, it makes sense for the road vehicle class to be a subclass of the vehicle class. Naturally, the opposite is not true since not every vehicle is a road vehicle. All cars and bikes are road vehicles, so they can be grouped as a subclass of road vehicle class. Race car and normal car that are not shown in the figure are indicated subclasses of car class as seen below example. The most general concept of inheritance can be explained in this way:

Common features of race car and normal car:

- They are road vehicles
- They have an engine.
- They are designed to carry persons.
- They consume the fuel.
- They need a driver.

There are common features of car and bike:

- They are road vehicle
- They are designed to carry persons.
- They need a driver.

There are common features of road vehicle and sea vehicle:

- They are designed to carry persons.
- They need a driver.

As seen in the example, information just like “to carry persons” and “to need a driver” are repeated more than one. Each category of information does not need to be saved in these classes, because inheritance provides it directly. If common features are defined in vehicle class, all subclasses can take that information from only one class. By this way, if it is necessary to update the system, only one change related to the concept in vehicle class can be enough. And this is called polymorphism.

Polymorphism is also related to “overloading” and “overriding”. Overloading is a compile time polymorphism method that has the same name with different parameters. Overloading is a feature that enables a class to possess two or more methods having same name. Unless return type of method is same, error will occur.

Overriding is a run time polymorphism method that the implementation given in the base class is replaced with that in subclass. Override method can be added by rewrite the method that inherited from the base class. In this way, it provides the use of inherited class method. In this case, the software allows the flexibility that can make a different job by using the same method.

1.3. Abstraction

The word abstract means dissociated from any specific instance. Abstraction is to develop models in terms of interface and functionality instead of

implementation details. So it is related to encapsulation and data hiding (Yourdon, Whitehead, Thomann, & Oppel, 1995). Abstraction is applied to the model by considering the process of identifying object. It is used to reduce complexity of the design and implementation that focuses on the meaning of behaviors to avoid specification. Thanks to the abstraction, class internals are protected from user-level errors which breaks state of the objects (Wegner, 1987).

Abstract class is a parent class, which allows inheritance, containing abstract members. These members are only declared, not implemented. Implementation of abstract members is done within the derived class. Another type of a member is virtual member. Unlike abstract member, virtual members are implemented in parent class. To declare an abstract class, abstract keyword is used. Abstract member functions and properties are also declared with this keyword.

1.4. Encapsulation

From the user's point of view, a number of features are packaged in a capsule to form an entity. This entity offers a number of services in the form of interfaces by hiding the implementation details (Canning, Cook, Hill, & Olthoff, 1989). The encapsulation term is used to describe the hiding of the implementation details. The advantages of encapsulation are information hiding and implementation independence. Local variables are hidden in functions and private members are hidden in classes. Therefore, external direct access is prevented. If user does not know the implementation details, it is called information hiding. If user's interface is not affected by changing the implementation mechanics, it is called implementation independence (Booch, Maksimchuk, Young, Conallen & Houston, 2007). Class encapsulates the static attributes and the dynamic behaviors into the limited area to isolate and reuse when necessary. These operations cannot be done in the traditional programming languages. Private access control modifier hides data member of a

class from outside world. Access to these private data members is provided via public accessor functions. Objects do not have permission to know the implementation details of others. The implementation details are hidden within the class.

2. OOP DESIGN

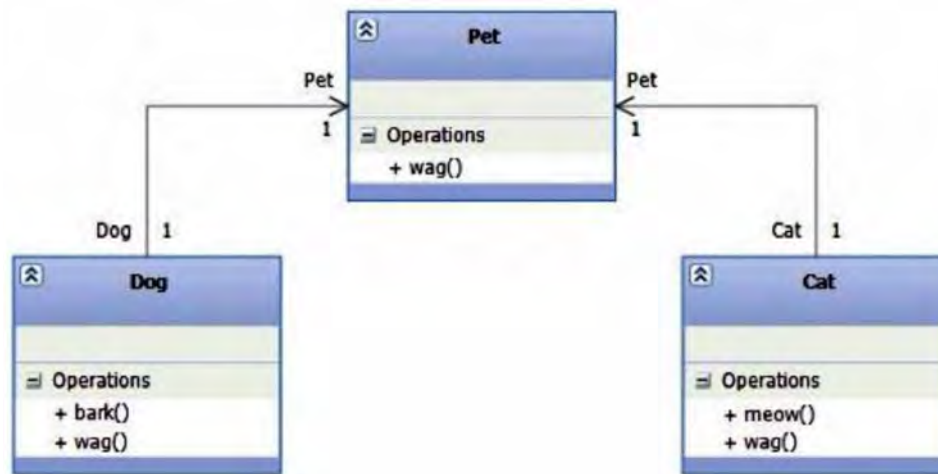
People tell the daily concept with using spoken language. Programmers try to express the concept and the entity that is related to the problem, to computer with using programming languages. To do this, during the design phase, models providing an expression are created. Object-oriented method provides the creation of these models and (if necessary) updating the system.

In section 2.1 we have developed an object-oriented design for our cat system and in section 2.2 we implemented our object-oriented design in Java and showed how to convert class diagrams to Java code.

2.1. OOP Design with the UML Diagrams

Unified Modeling Language (UML) diagrams are used to describe the structure of systems. It is a standard for modeling object-oriented systems which defines sets of rules and vocabulary for conceptual and physical representation of system (Mallick & Das, 2013). It includes graphical notation to create visual model of the system (Booch, Rumbaugh, & Jacobson, 2005) which has well-defined semantics. UML in development process is used for object-oriented analysis and design. There are different diagramming techniques which is used to model a system. These diagrams can be grouped into 2 which are called structure and behavior diagrams. Structure diagram shows the static relationships and represents the data in a system which includes class, object, package, deployment, component, composite structure and profile diagrams. Behavior diagram shows the dynamic relationships between the instances or ob-

Figure 4. Class diagram modeling generalization of superclass Pet and subclasses Cat and Dog



jects. It includes use-case, sequence, statechart and activity diagrams. First of these diagrams called use case diagram is used to capture the requirements of the system and shows the relationships between the system and environment. Sequence diagram is a form of interaction diagram which is used to model the behavior of objects by focusing on time-based ordering of an activity. Statechart diagram shows the behavior of the classes. Activity diagram illustrates the flow of activities in a use case (Booch, Rumbaugh, & Jacobson, 2005).

It has developed an object-oriented design with UML class diagram for pets as shown in Figure 4.

Figure 4 is the class diagram that models generalization of superclass Pet and subclasses Cat and Dog. The following section 2.2, it is implemented our object-oriented design in Java and showed how to convert class diagrams to Java code (Deitel & Deitel, 2012).

2.2. Implementation of OOP Design in Java Programming

A university system can be given as an example, and this system is implemented by java programming language. The Java source code for the Person class is shown in Listing 1.

In this example, there is a base class which is called Person. This class includes basic informa-

tion, which people have, and there are 3 derived classes which inherit data and functions from the base class. These derived classes have their own methods and data members, and also they have methods and data members of Person class.

Keyword public is an accessibility modifier which refers the calling members or methods from external locations like other classes (Lea, 1999). AcademicPersonnel, administrativePersonnel and student classes call methods from Person class. Derived class methods are declared public because, they should be accessed from Main() method which Test class has. AcademicPersonnel class has a method called checkAcademicPersonnel(), this method is declared private. Private access specifier refers hiding data members and methods from other classes. Every class can access its own private data members and methods. If we try to reach this private checkAcademicPersonnel() method from Test class, it is going to fail. deliverCourse() method, which is declared public, can reach this private method because, they are in the same class.

Last class is called Test, which is developed to test other classes, it only has one method called Main. We create 3 different objects from 3 different derived class. Object academicPersonnel1 is created from academicPersonnel class which has deliverCourse() method on its own scope but, this

Listing 1.

```

1. public class Person
2. {
3.     public void printPersonalInformation(String id, String fName, String lName)
4.     {
5.         System.out.println(id+" "+fName+" "+lName);
6.     }
7. }
8. class academicPersonnel extends Person
9. {
10.     private boolean checkAcademicPersonnel()
11.     {
12.         return true;
13.     }
14.     public void deliverCourse(String courseCode, String courseName)
15.     {
16.         if(checkAcademicPersonnel())
17.         {
18.             System.out.println("Academic Personnel");
19.             System.out.println("Course: "+courseCode+ " "+ courseName);
20.         }
21.     }
22. }
23. class student extends Person
24. {
25.     void takeCourse(String courseCode, String courseName)
26.     {
27.         System.out.println("Course: "+courseCode+" "+courseName
28.     }
29. }
30.
31. class administrativePersonnel extends Person
32. {
33.     boolean checkStudentInformation()
34.     {
35.         return true;
36.     }
37. }
38. class Test
39. {
40.     public static void main(String[] args)
41.     {
42.         academicPersonnel academicPersonnel1 = new academicPersonnel();

```

continued on following page

Listing 1. Continued

```

43. student student1 = new student();
44. administrativePersonnel administrativePersonnell = new administrativeP-
ersonnel();
45. academicPersonnell.printPersonalInformation(("100305044", "Rahime", "Yilm
az"));
46. academicPersonnell.deliverCourse("CENG101", "Algorithms and Programming
I");
47. student1.printPersonalInformation(("214700560", "Anil", "Sezgin"));
48. student1.takeCourse("CENG102", "Algorithms and Programming II");
49. administrativePersonnell.checkStudentInformation();
50. }
51. }

```

object can also reach printPersonalInformation() method on Person class. Object student1 is created from student class, which has takeCourse() method on its own scope, and same as academicPersonnel class, student class is derived from Person class, so this object can reach base class methods. Third object is created from administrativePersonnel class which is derived from same base class.

For another example of the Java source code for the pet class is shown in Listing 2.

This example shows using abstract class. Abstract class is used to create a base template for derived classes. Pet class is the abstract base class which includes an abstract method called wag(). This method has no implementation, therefore it is declared abstract. There are two classes called dog and cat which are derived from pet class. These two derived classes provide an override method by using override keyword. Cat and Dog classes override wag() method, dog1.wag() invokes the wag() method declared in dog class, cat1.wag() invokes the wag() method declared in cat class.

FUTURE RESEARCH DIRECTIONS

To create (write) a program in any languages needs detailed training. That means a programmer have to be trained highly in the programming language that programmers wants to use. In the real life,

even if any person do not know any languages, they could write a computer program. Namely, untrained people can also create the program without using any programming languages. We can ask how it can be done. Instead of teaching a programming languages to a person, we can teach how to draw flow charts. When a correct flow chart is drawn, a case at the background of the flow chart can create a program source code by using the flow chart. If these kinds of cases can be created, an untrained person can write his program by using flow chart and produce his programming code. First of this, the basic flow chart symbols are taught and combination of this basic elements are told to the training people who wants to implement their problem by using flow chart. And then, it is expected to solve their problem. This solves the dependency of a programming language problem to implement a computer software.

CONCLUSION

The programming languages before OOP concept were not easy and friendly. Large and complex problems were solved by dividing into small systems. The real world modeling was not implemented by procedural programming. Creation of the real simulation of problems can be implemented

Listing 2.

```

1.  public abstract class pet
2.  {
3.  public abstract void wag();
4.  }
5.  class cat extends pet
6.  {
7.  @Override
8.  public void wag()
9.  {
10.   System.out.println("Cat wags tail.");
11.  }
12.  }
13. class dog extends pet
14. {
15. @Override
16. public void wag()
17. {
18.   System.out.println("Dog wags tail.");
19. }
20. }
21. class Test1
22. {
23. public static void main(String[] args)
24. {
25.   dog dog1 = new dog();
26.   cat cat1 = new cat();
27.   dog1.wag();
28.   cat1.wag();
29. }
30. }

```

by the OOP. The functional programming shows more tendency for losing data while running according to OOP. The system requirements of data structure and data are more flexible in OOP than those in other models. The modification of all systems is more complicated than the object oriented system, since OOP system has a modular structure. All processes are done by using functions in former languages but in OOP, everything is processed by using objects methods.

The data is transferred by using messages among objects. By this way, objects can communicate among themselves. Data hiding properties of OOP attempt to protect the data from the outside modification request.

As a conclusion, the features of OOP, such as encapsulation, abstraction, polymorphism and inheritance help us to model the real life entities on computers. A more qualified and effective software can be created by using OOP.

REFERENCES

- Bista, R., Bajracharya, L., & Dongol, D. (2015). A New Approach To Enhance Efficiency of Object Oriented Programming. *Technia*, 8(1), 1058.
- Booch, G., Maksimchuk, R. A., Engle, M. W., Young, B. J., Conallen, J., & Houston, K. A. (2007). *Object-Oriented Analysis and Design with Applications* (3rd ed.). Boston, MA: Addison-Wesley Professional.
- Booch, G., Rumbaugh, J., & Jacobson, I. (2005, May). *The Unified Modeling Language User Guide*. Addison-Wesley.
- Canning, P. S., Cook, W. R., Hill, W. L., & Olthoff, W. G. (1989). *Interfaces for Strongly-Typed Object-Oriented Programming*. Paper presented at OOPSLA '89, New York, NY. doi:10.1145/74877.74924
- Deitel, P., & Deitel, H. (2012). *Java How to Program* (9th ed.). Pearson Education Limited.
- Dennis, A., Wixom, B. H., & Tegarden, D. (2015, April). *System Analysis & Design: An Object-Oriented Approach with UML* (5th ed.). Academic Press.
- Gamma, E., Helm, R., Johnson, R., & Vlissides, J. (1994). *Design Patterns: Elements of Reusable Object-Oriented Software*. Boston, MA: Addison-Wesley Professional.
- Harel, D., Marron, A., & Weiss, G. (2010, June). Programming coordinated behavior in java. In *European Conference on Object-Oriented Programming* (pp. 250-274). Springer Berlin Heidelberg.
- Lea, D. (1999). *Concurrent Programming in Java: Design Principles and Patterns*. Boston, MA: Addison-Wesley Professional.
- Ledgard, H. F. (1996). *The Little Book of Object-Oriented Programming*. Upper Saddle River, NJ: Prentice Hall.
- Mallick, B., & Das, N. (2013, November). An Approach to Extended Class Diagram Model of UML for Object Oriented Software Design. *International Journal of Innovative Technology & Adaptive Management*, 1.
- Robson, D. (1981). Object-Oriented Software Systems. *Byte.*, 6(8), 74–86.
- Thomsen K., S. (1986). *Multiple Inheritance, a Structuring Mechanism for Data, Processes and Procedures*. Aarhus Universitet, Matematisk Institut, Datalogisk Afdeling.
- Wegner, P. (1987). *Dimensions of Object-Based Language Design*. Paper presented at OOPSLA '87, New York, NY. doi:10.1145/38765.38823
- Wirfs-Brock, R. J., & Johnson, R. E. (1990). Surveying Current Research in Object-Oriented Design. *Communications of the ACM*, 33.
- Yourdon, E., Whitehead, K., Thomann, J., & Opel, K. (1995). *Mainstream Objects: An Analysis and Design Approach for Business* (1st ed.). Upper Saddle River, NJ: Prentice Hall.

KEY TERMS AND DEFINITIONS

Class: Used to define the objects and to describe the contents of the objects.

Java: A high level programming language that is based on object oriented design which was first released by Sun Microsystems in 1995.

Message: Communication type among methods.

Method: Called a function or a procedure that is used to operate on the data.

Object: Instance occurrence of a same class that models real-world items.

Object Oriented: A technique for programming that is based on the objects and on the relationship between those objects.

Programming: A mathematical methodology that helps programmers while writing a program.

The Past, Present, and Future of UML

S**Rebecca Platt***Murdoch University, Australia***Nik Thompson***Curtin University, Australia*

INTRODUCTION

Since its inception, the Unified Modeling Language (UML) has risen to relative ubiquity in the IT community. However, despite its status as an ISO industry standard (International Organization for Standardization, 2005), the UML is still evolving to accommodate the changing needs of industry. This development aims to ensure that UML remains effective and relevant to the most current developments in software engineering techniques. This article charts the progress of this arguably indispensable standard and discusses the ongoing evolution in three sections: The Past, The Present, and The Future. The Past section will detail the reasons for which standardization was needed, the history behind its inception and development, initial reception from the user community and also its initial effectiveness. The Present section then describes the various changes between UML 1.0 and UML 2.5. The reasons behind these changes and the effectiveness of them are then discussed. Finally in The Future section, the article will describe the current state of UML, predictions for the next specification of UML based on the Object Management Group documentation, and also common problems and suggestions from the wider community which may be addressed in future iterations of the specification.

BACKGROUND

The Unified Modeling Language is a form of notation that was developed with the core goal of

creating a standardized representation of general-purpose models, with the focus of functionality of these primarily being for software engineering and systems development. Despite this main focus of approach in the specification design, the language is meant to attain some level of applicability regardless of the subject matter. The reason a modeling language was needed in order to achieve this was to manage the complexity of the subject at hand - whether it was system or software design or another subject entirely. As a model is by nature an abstraction of reality, it allows the user to characterize the design of the subject in an effective manner. This abstract model then enables the user to better evaluate the subject and communicate that in an efficient and meaningful way rather than attempting to demonstrate their intentions using the actual software or system in question. In order to achieve this intended core goal the language has been modified and refined over time, resulting in evolutions of varying effectiveness and popularity.

THE EVOLUTION OF UML

The Past

In the late 1950s, the first object orientated programming language, Simula was introduced, and with it came “a powerful new combination of ideas into structuring computer programs, including instantiation of abstract data types, inheritance, and polymorphism” (Cook, 2012, p. 471). To accompany this new idea of object orientated

DOI: 10.4018/978-1-5225-2255-3.ch651

languages, methods for designing software in this object orientated way also started to emerge, and in time they were referred to as modeling languages. By the late 1980s there were more than fifty separate modeling languages - each with their own syntax, structure and notation. There were many issues with this overwhelming variety of languages and it has been noted that “such open-ended approaches [could] affect and constrain the system in unexpected ways or even result in failure. For example, system development and implementation failure rates remained stubbornly high. Cost overruns and time overruns were still the norm, rather than the exception” (Erickson & Siau, 2013, p. 296). As it was humanly impossible in this kind of environment for all system analysts and other relevant personnel to be trained in all methods, the lack of communication and technical understanding coupled with the fact that the majority of the languages available were unable to meet the demands required of them, led to alarmingly high project failure rates.

This lack of standardization and communication was not only negatively affecting development projects but also limiting the potential of object-orientated technology in general. In response to this very significant concern, The Object Management Group (OMG) was founded in 1989. The initial and presiding goal of OMG was to “create a standard for communication amongst distributed objects” (Cook, 2012, p. 472). This goal was intended to foster progress toward a common object model that would work on all platforms on all kinds of development projects. In order to further this goal specifically in the domain of modeling languages, OMG launched the Object Analysis and Design Special Interest Group to study design methods. This is also the origin point from which any Request For Proposals were issued.

Around the time that OMG was founded, a separate company called Rational was also attempting to implement a solution to the over saturation of modeling languages in use. To this end they recruited Grady Booch and James Rumbaugh in 1996. These men were the creators of

two of the dominant modeling languages of the time. Booch’s method was called Object-Oriented Design (OOD) (Booch, 1991) and Rumbaugh’s method was known as the Object-Modeling Technique (OMT) (Rumbaugh, Blaha, Lorensen, Eddy, & Premerlani, 1990). They were soon joined by Ivar Jacobson, whose Object-Oriented Software Engineering (OOSE) method (Jacobson, 1992) was also a prominent modeling language at the time. “The Three Amigos” as they later came to be known then set to work on the development of the Unified Modeling Language. A potentially universal standard form of notation with the intent to create ease of communication and reduce the risk of failure for projects, with human factors considered above all as this had been identified as a main failure point of previous projects (Erickson & Siau, 2013).

The UML 0.91 specification was the initial result of the unification of OOD, OMT, and OOSE, a somewhat successful endeavor as each base modeling language had unique strengths; Booch’s OOD was good for low level design, Rumbaugh’s OMT was effective for OO analysis, and Jacobson’s OOSE was good for high level design, as well as allowing for the implementation of use cases. Working with “The Three Amigos” were the UML Partners; a software development team who represented a range of different vendors and system integrators, who would collaborate to propose UML as the standard modeling language for the OMG (Kobryn, 1999). Representatives from other companies (such as IBM, Microsoft and Oracle) were consulted during the Object-Oriented Programming, Systems, Languages and Applications (OOPSLA) conference held that year, with the outcome of these consultations resulting in the UML 1.0 draft which was then submitted to OMG in response to the Request For Proposal. UML 1.0 was accepted by OMG in November, 1997.

The initial response after the release of the specification indicated that the Unified Modeling Language was very effective, once the personnel involved had made it past the difficult learning

curve of training in a new modeling language. In fact there is speculation that the response towards UML was actually too great - for while it was proven to be much more effective than its predecessors, it still had issues. The rapid uptake and positive response meant that the uptake of UML ended up growing at an alarming rate before it had finished standardizing properly.

The Present

When initially accepted as a standard, UML 1.0 appeared to meet all stated requirements and to be an effective modeling language. Since then, however, a number of revisions have taken place to alter the notation in order to fix various shortcomings and to become more effective. For example, some of the issues that were resolved between UML 1.1 and UML 1.3 included the lack of integration between certain model types, the absence of certain modelers and that some of the standard elements were named and organized inconsistently. There was also trouble with the architectural alignment – According to OMG “The submitters fell short of their goal of implementing a 4-layer metamodel architecture using a strict metamodeling approach. Instead they settled for the pragmatic, but less rigorous, loose (non-strict) metamodeling approach. This “adversely affected the integration of UML with other OMG modeling standards, such as the Meta Object Facility (MOF)” (Kobryn, 1999, p. 31).

As it is, The Object Management Group oversees standardization and it is through their processes that revisions of the UML are implemented. There are two mechanisms for standard revisions; RFPs and RTFs. The Request For Proposal (RFP) is the primary mechanism for updating specifications, while Revision Task Forces (RTF) are secondary. When a proposal is received, it is the RTF that examines and votes on the validity of it. The RTF is also able to recommend changes to the proposal in order to clarify areas that may be ambiguous. If the proposal is approved, then it becomes OMG adopted technology. If the proposal

is not approved, then the RFP is reissued, with changes made to it to reflect the reasons for the last proposal failing.

Through the OMG system, a number of significant changes have been made in response to some of the shortcomings identified in UML. Between UML 1.1 and UML 1.2, the specification was reformatted in order to better align with other OMG specifications. Typographical and grammatical errors were also targeted in this revision. Between UML 1.2 and UML 1.3, problems that had occurred during the alterations of the last revision were fixed, the activity graph notation was completed, and the standard elements were more formally organized. The revisions of the specification were all rather minor after that, up until UML 2.0 was released in 2005. The following UML Specification to be released was UML 2.4.1, and unlike UML 1.x this specification is organized into four sections. These sections are called the Superstructure, the Infrastructure, the Object Constraint Language, and the UML Diagram Interchange. The Infrastructure “defines the foundational language constructs required” (OMG, 2011, p. 1). This is then balanced by the UML Superstructure, which “defines the user level constructs required” (OMG, 2011, p.1). The current UML Specification in use is UML 2.5, which was released in June 2015. This version has been re-written to simplify the contents, increasing readability of the document. Other major changes include the removal of Infrastructure as a separate section of the Specification, and also “the compliance levels L0, L1, L2, and L3 have been eliminated, because they were not found to be useful in practise” (OMG, 2015a, p.11).

Studies have shown that the ongoing revision implementation has been successfully achieving the goal of standardization. As UML becomes more refined, it also becomes more universally accessible and accepted. A study found that 21% of Australian Computing Society members used UML frequently (Davies, Green, Rosemann, Indulska, & Gallo, 2006), further evidence of this growing standardization was demonstrated by

Dobing and Parsons (2006) who noted that class diagrams were the most frequently utilized aspect of UML as reported by 73% of participants. Since these studies were conducted it has been demonstrated that practitioners have been successfully implementing the Unified Modeling Language more effectively and frequently, to such a point that it is now a part of many undergraduate university curricula in Information Technology fields. More recent studies have shown that the growth and uptake of UML has persisted over time (e.g. Dobing & Parsons, 2010; Budgen, Burn, Brereton, Kitchenham, & Pretorius, 2011).

The Future

The next specification for UML, UML 2.6, currently has no known set or speculated release date. As it is too early in the development process for even the unofficial release specifications to be revealed to the general public, very little is known about this future specification update.

Based on comments regarding UML 2.6, it appears that the next version will cover minor revisions to the 2.5 specification, including fixing two sets of syntactical errors currently shown in the UML 2.5 metamodel. This expectation of only minor revisions is further supported by the list of issues shown to be resolved by the UML 2.6 Revision Task Force, which includes such changes as “Clarification of use case semantics” and “Parameterization of lifelines”, “such that Interactions can be used in slightly different context” (OMG, 2015b).

Issues, Controversies, Problems

Despite years of revision that have successfully yielded incremental improvements to the specification, problems remain that need to be addressed. The standard elements are still rather “bloated” and they contain a level of inconsistency in both naming and organization. This level of complexity and the inconsistencies introduced during revisions have been detrimental to the overall

readability of the specification. There is a concern that the design of the notation is not sufficiently user-friendly, which would discourage potential users from adopting and using UML in favour of other simpler alternatives such as DOT graph description language, as described by Erickson and Siau (2013).

Another potential issue that needs to be addressed concerns the cyclic nature of specification revisions. In the process of updating UML to attempt to deal with the above issues and problems, excessive addition, removal and alteration of major concepts could affect the core structure of UML. The current method of revision leaves the core structure vulnerable. As previously stated, the UML focus of functionality was primarily for software engineering. As a result, software tools offer extensive support for UML when used with this focus in mind. However there is very little support in software tools for any other application, despite the language’s goal of being a standardized representation of general-purpose models.

Solutions and Recommendations

The issue of “bloating” regarding the elements of UML is due to the inconsistent naming and organization within the standards. By phasing in more consistency to the various aspects of future specifications of the language, the volume of elements would be reduced and bloating would cease to be an issue. Reduction of elements and increase in consistency would also aid the uptake of the language among new users. Simplifying the language (and the specification documentation in relation to this) would increase the readability of it, and encourage more users to utilize UML instead of another modeling language. Also, core structure vulnerability can be corrected by the introduction of protocols within the specification revision procedure, to ensure that this remains unaltered. Support for non-software engineering projects is difficult to implement currently due to the fact that UML implementation seems to be “tool-based.” If all of the various tools used for

UML that are widely recognized started creating more support for the language in terms other than software engineering, then the specification would broaden to include this more as a result.

FUTURE RESEARCH DIRECTIONS

It has been shown that UML has been implemented within the field of software engineering increasingly over the years, moving from relatively low industry usage (e.g. Davies et al, 2006) to the present state where the growth in UML usage has led to an abundance of tools and software to better support the language. However, research into the Unified Modeling Language has been limited in recent years. There have been a few surveys conducted based on the use of UML in terms of software engineering and development, but very little in terms of its other applications as a general purpose modeling language. Current study seems to focus on the compliance of tools to UML, rather than the compliance of UML to its intended purpose. Future surveys of the adoption of UML (both within and outside of the field of software engineering) would be well served to include elements concerning the perceived effectiveness of the modeling language by users in real world situations. Another possible research direction may examine how the modeling language has affected development practices and utilization of techniques, and whether the overall project success rate has increased as a result of this.

CONCLUSION

The Unified Modeling Language may be the current industry standard, but it is still evolving and transitioning through constant revisions of the specification. These stages of revision are implemented to ensure that the UML remain effective and viable in the demanding and rapidly changing landscape of software engineering. This article

examined this evolution in terms of three main periods described as The Past, The Present, and The Future. The Past section detailed the reason behind which standardization was needed, the history leading up to and including the development of UML. The initial reception from the user community and initial effectiveness were also discussed. The Present section then described the various changes between UML 1.0 and UML 2.5 and the reasons behind these changes and their ongoing effects. Finally, The Future section described the current state of UML, the expectations for the next specification of UML and also some open issues from the wider community which are yet to be addressed. Some possible solutions and future research directions were also presented in light of these issues. In conclusion the Unified Modeling Language has proven itself to be an effective standard for communication and it will maintain its significant foothold in software engineering for the foreseeable future. However, the requirement for continual revisions to the specifications will also remain as the expected functionality and needs of UML practitioners will continue to change over time.

REFERENCES

- Avison, D. E., & Fitzgerald, G. (2003). Where now for development methodologies? *Communications of the ACM*, 46(1), 78–82. doi:10.1145/602421.602423
- Booch, G. (1991). *Object-oriented design with application*. Benjamin-Cummings.
- Budgen, D., Burn, A. J., Brereton, O. P., Kitchenham, B. A., & Pretorius, R. (2011). Empirical evidence about the UML: A systematic literature review. *Software, Practice & Experience*, 41(4), 363–392. doi:10.1002/spe.1009
- Cook, S. (2012). Looking Back at UML. *Software & Systems Modeling*, 11(4), 471–480. doi:10.1007/s10270-012-0256-x

- Davies, I., Green, P., Rosemann, M., Indulska, M., & Gallo, S. (2006). How do practitioners use conceptual modeling in practice? *Data & Knowledge Engineering*, 58(3), 358–380. doi:10.1016/j.datak.2005.07.007
- Dobing, B., & Parsons, J. (2006). How UML is used. *Communications of the ACM*, 49(5), 109–113. doi:10.1145/1125944.1125949
- Dobing, B., & Parsons, J. (2010). Dimensions of UML Diagram Use: Practitioner Survey and Research Agenda. In K. Siau & J. Erickson (Eds.), *Principle Advancements in Database Management Technologies: New Applications and Frameworks* (pp. 271–290). Hershey, PA: IGI. doi:10.4018/978-1-60566-904-5.ch013
- Erickson, J., & Siau, K. (2013). Unified Modeling Language: The teen years and growing pains. *Human Interface and the Management of Information. Information and Interaction Design* (pp. 295–304). Berlin: Springer.
- Fowler, M., & Scott, K. (2003). *UML Distilled*. Boston: Addison-Wesley.
- International Organization for Standardization. (2005). *ISO/IEC 19501:2005*. Retrieved 2nd Nov 2013, from http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=32620
- Jacobson, I. (1992). *Object-oriented software engineering: a use case driven approach*. New York: Pearson Education.
- Kobryn, C. (1999). UML 2001: A standardization odyssey. *Communications of the ACM*, 42(10), 29–37. doi:10.1145/317665.317673
- Lucas, F. J., Molina, F., & Taval, A. (2009). A systematic review of UML model consistency management. *Information and Software Technology*, 51(12), 1631–1645. doi:10.1016/j.infsof.2009.04.009
- Management Group. (n.d.). Retrieved 10th Nov 2013 from <http://www.omg.org/spec/UML/2.4.1/Infrastructure>
- Miles, R., & Hamilton, K. (2006). *Learning UML 2.0*. O'Reilly.
- MOF Revision Task Force (1999). *Meta Object Facility Specification v. 1.3. Document ad/99-06-05*. Object Management Group.
- OMG. (1999). Policy and procedures of the OMG technical process. Document pp/99-05-01. OMG.
- OMG. (2011). *OMG Unified Modeling Language (OMG UML), Infrastructure, v2.4.1*. OMG.
- OMG. (2013). *OMG Unified Modeling Language (OMG UML), v2.5 FTF Beta1*. Object Management Group. Retrieved 10th Nov 2013 from <http://www.omg.org/spec/UML/2.5>
- OMG. (2015a). *OMG Unified Modeling Language (OMG UML), v2.5*. Object Management Group. Retrieved 4th Oct 2015 from <http://www.omg.org/spec/UML/2.5>
- OMG. (2015b). *Issues for UML 2.6 Revision Task Force Mailing List (OMG UML)*. Object Management Group. Retrieved 4th Oct 2015 from <http://www.omg.org/issues/uml2-rtf.open.html>
- Pilone, D., & Pitman, N. (2009). *UML 2.0 in a Nutshell*. O'Reilly Media.
- Rumbaugh, J. R., Blaha, M. R., Lorensen, W., Eddy, F., & Premerlani, W. (1990). *Object-oriented modeling and design*. Prentice-Hall.
- ADDITIONAL READING**
- Alhir, S. S. (2003). *Learning UML*. O'Reilly.
- Álvarez, A. T., & Alemán, J. L. F. (2000). Formally modeling UML and its evolution: A holistic approach. In S. F. Smith (Ed.), *Formal methods for open object-based distributed systems IV* (pp. 183–206). US: Springer. doi:10.1007/978-0-387-35520-7_9

Bennett, S., McRobb, S., & Farmer, R. (2006). *Object-oriented systems analysis and design using UML* (Vol. 2). UK: McGraw-Hill Berkshire.

Eriksson, H.-E., & Penker, M. (1997). *UML toolkit*. New York: John Wiley & Sons, Inc.

Eriksson, H.-E., & Penker, M. (2000). *Business modeling with UML*. Chichester: Wiley.

Fowler, M. (2004). *UML Distilled: A Brief Guide to the Standard Object Modeling Language*. Boston: Addison-Wesley Professional.

Kobryn, C. (1999). UML 2001: A standardization odyssey. *Communications of the ACM*, 42(10), 29–37. doi:10.1145/317665.317673

Miles, R., & Hamilton, K. (2006). *Learning UML 2.0*. California: O'Reilly.

Object Management Group. (2007). *Unified Modeling Language (OMG UML): Superstructure*. Massachusetts: Object Management Group.

Pooley, R., & Stevens, P. (1999). *Using UML: Software Engineering with Objects and Rules*. Boston: Addison-Wesley Longman.

Rumbaugh, J., Jacobson, I., & Booch, G. (1999). *The Unified Modeling Language reference manual*. Boston: Addison-Wesley Longman.

Satzinger, J. W., Jackson, R. B., & Burd, S. D. (2005). *Object-oriented Analysis and Design: With the Unified Process*. Boston: Thomson Course Technology.

Satzinger, J. W., Jackson, R. B., & Burd, S. D. (2011). *Systems analysis and design in a changing world*. Boston: Cengage Learning.

KEY TERMS AND DEFINITIONS

Model: A conceptual diagram used to represent a system.

Object Management Group: An organization created with the goal to determine a standard method of communication between distributed objects.

OOPSLA: “Object-Oriented Programming, Systems, Languages and Applications” – an annual research conference run by the Association for Computing Machinery.

Software Engineering: The application of systematic methods and approaches for the development and maintenance of software artifacts.

Specification: The set of requirements that must be satisfied in order for any model to comply with the current standards of UML.

Unified Modeling Language: A form of notation developed with the core goal of creating a standardized representation of general-purpose models, with the focus of functionality primarily being for software engineering.

Petri Nets Identification Techniques for Automated Modelling of Discrete Event Processes

Edelma Rodriguez-Perez

CINVESTAV Unidad Guadalajara, Mexico

Ernesto Lopez-Mellado

CINVESTAV Unidad Guadalajara, Mexico

INTRODUCTION

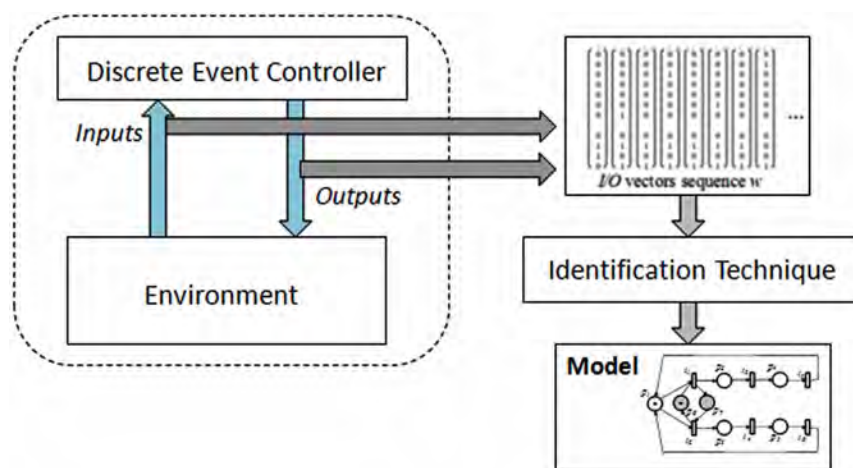
Building Petri net models from system behaviour observations is a hard task when the system is large and complex; then, the use of computer-aided modelling tools is useful. Identification techniques have been useful for building systematically models involving events and states. Finite automata and Petri nets have been used as a formalism to describe the functioning of discrete event processes in operation.

Reactive systems are embedded within an environment interacting with other systems. We focus on systems that interact through binary sig-

nals, which is the case of discrete event processes. The behaviour of the system is then captured as sequences of vectors whose entries are the values of input-output signals; afterwards, the sequences are processed by an identification method to obtain the discrete event model. This is shown in Figure 1.

This chapter surveys relevant identification methods and overviews two approaches that generate models of different levels of abstraction; one that describes in detail the relationship between input events and outputs, and other that yields compact descriptions. Finally, current research problems and trends on discrete event process identification are discussed.

Figure 1.



BACKGROUND

1. Identification Methods

1.1. Language Learning Methods

Pioneer works on identification methods can be found in computer science theory, where the problem of obtaining a language representation from sets of accepted words has been dealt since a long time. Such methods are generally referred as *language learning techniques*.

Gold's method (Gold, 1967) for identification in the limit processes positive samples: an infinite sequence of examples such that the sequences contain all and only all the strings in the language to learn.

The Probably Approximately Correct (PAC) learning technique in (Valiant, 1984) learns from random examples and studies the effect of noise on learning from queries.

The query learning model proposed in (Angluin, 1988) considers a learning protocol based on a "minimally adequate teacher"; this teacher can answer two types of queries: membership query and equivalence query.

Several works adopted state machines as representation model, allowing describing the observed behaviour.

In (Booth, 1967) a method to model a language as Moore or Mealy machines is presented. The method proposed in (Kella, 1971) allows obtaining models representing Mealy machines from a single observed input-output sequence. In (Biermann & Feldman, 1972) a method to identify non deterministic Moore machines based on a set of input-output sequences is presented. The method presented in (Veelenturf, 1978) processes simultaneously a sample of sequences to produce stepwise convergent series of Mealy machines, such that the behaviour of every new machine includes the behaviour of the previous one. Later, in (Veelenturf, 1981) an algorithm to identify a unique Moore machine generating the behaviour observed during m sequences starting at the same

initial state is proposed. This method is improved in (Richetin, Naranjo & Runeau, 1984); it proposes two algorithms to identify multiple systems.

Other works use as description formalism Petri net models. In (Hiraishi, 1992) an algorithm for synthesising Petri net models is presented. The proposed algorithm has two phases. In the first phase, the language of the target system is identified in the form of DFA. In the second phase, the algorithm guesses from the DFA the structure of a Petri net that accepts the obtained language.

1.2. Methods for Process Identification

In recent years, identification approaches (based on Petri net or automata) have been proposed for obtaining approximated models of Discrete event process whose behaviour is unknown or ill-known. In the context of manufacturing processes, identification methods can be complementary to established modelling techniques.

The incremental synthesis approach, proposed in (Meda, Ramirez & Lopez, 2000), (Meda & Lopez, 2001), (Meda & Lopez, 2003), deals with unknown partially measurable concurrent discrete event systems (DES) exhibiting cyclic behaviour. Several algorithms have been proposed allowing the on-line building of interpreted Petri net (IPN) models from the DES outputs. Although the techniques are efficient, the obtained models may represent more sequences than those observed.

In (Giua & Seatzu, 2005) a method to build a free labelled PN from a finite set of transitions strings is presented. This method is based on the resolution of an Integer Linear Programming (ILP) problem; the obtained PN generates exactly the observed language. Both the ILP statement and its solution are computationally demanding. This approach has been extended to other PN classes (Cabasino, Giua, & Seatzu, 2007; Dotoli, Fanti, & Mangini, 2008; Fanti & Seatzu, 2008; Dotoli et al., 2011); however, issues regarding applications to actual industrial DES have not yet been addressed in these works.

Another recent off-line method (Klein, Litz, & Lesage, 2005) allows building a non-deterministic finite automaton (FA) from a set of input/output (I/O) sequences, experimentally measured from the DES to be identified. Under several hypotheses, the constructed FA generates exactly the same I/O sequences of given length than observed ones. The method was conceived for fault detection in a model-based approach (Roth, Lesage, & Litz, 2009). Extensions to this work propose an identification method performing optimal partitioning of concurrent subsystems for distributed fault detection purposes (Roth, Lesage, & Litz, 2010).

Regarding Input-Output identification, in (Estrada, Lesage, & Lopez, 2014) the system behaviour, captured as an I/O sequence is processed to obtain Interpreted Petri nets including cyclic behaviour. Although the proposed methodology is scalable due to the algorithms efficiency, the obtained models are close to finite automata and can be huge, due to the explicit representation of observed input changes that could not be relevant to define the output evolution. Later in (Estrada, Lopez, & Lesage, 2015) a new method for building IPN models from I/O sequences is presented. It consists of two complementary stages; the first one obtains, from the I/O sequence, the reactive part of the model composed by observable places and transitions. Afterwards, The I/O sequence is mapped into a sequence of the created transitions, from which the second stage builds the non-observable part of the model including places that ensure the reproduction of the observed input output sequence.

Recently, in (Rodriguez, Tapia, & Lopez, 2016) a novel method allows building a reduced representation of the observable part of the model which yields consequently, a reduced complete IPN.

Other works on the matter, based on different approaches have been proposed. The techniques for workflow mining, published by van der Aalst and co-workers (Cook et al., 2004; van der Aalst, Weijters, & Maruster, 2004), allow building Petri net models of workflow processes in which all the activities are observable. Other works pursue

the construction of a stochastic PN from recorded event sequences (Ould El Medhi, Leclercq, & Lefebvre, 2006; Ould El Medhi et al., 2012).

2. Petri Nets and Automated Modelling

This section presents the basic concepts and notation of Petri Nets (PN) and Interpreted Petri nets (IPN).

Definition 1: An ordinary Petri Net structure G is a bipartite digraph represented by the 4-tuple $G = (P, T, Pre, Post)$ where: $P = \{p_1, p_2, \dots, p_{|P|}\}$ and $T = \{t_1, t_2, \dots, t_{|T|}\}$ are finite sets of vertices named places and transitions respectively; $Pre(Post): P \times T \rightarrow \{0, 1\}$ is a function representing the arcs going from places to transitions (from transitions to places).

The incidence matrix of G is $W = W^+ - W^-$, where $W^- = (w_{ij}^-)$; $w_{ij}^- = Pre(p_i, t_j)$; and $W^+ = (w_{ij}^+)$; $w_{ij}^+ = Post(p_i, t_j)$ are the pre-incidence and post-incidence matrices respectively.

A marking function $M: P \rightarrow Z^+$ represents the number of tokens residing inside each place; it is usually expressed as a $|P|$ -entry vector. Z^+ is the set of nonnegative integers. In particular, in this paper $M: P \rightarrow \{0, 1\}$; the PN is referred as 1-bounded or *safe*.

Definition 2: A Petri Net system or Petri Net (PN) is the pair $N = (G, M_0)$, where G is a PN structure and M_0 is an initial marking.

In a PN system, a transition t_j is *enabled* at marking M_k if $\forall p_i \in P, M_k(p_i) \geq Pre(p_i, t_j)$; an enabled transition t_j can be fired reaching a new marking M_{k+1} . This behaviour is represented as $M_k \xrightarrow{t_j} M_{k+1}$. The new marking can be computed as $M_{k+1} = M_k + Wu_k$, where $u_k(i) = 0, i \neq j$, $u_k(j) = 1$; this equation is called the PN state equation. The reachability set of a PN is the set of all possible reachable markings from M_0 firing only enabled transitions; this set is denoted by $R(G, M_0)$.

Now it is defined IPN, an extension to PN that allows associating input and output signals to PN models. This definition is adapted from (David & Alla, 2010).

Definition 3: An interpreted Petri net (IPN) (Q, M_0) is a labelled net structure $Q = (G, \Sigma, \Phi, \lambda, \varphi)$ with an initial marking M_0 where:

- G is a PN structure,
- $\Sigma = \{\alpha_1, \alpha_2, \dots, \alpha_r\}$ is the inputs alphabet,
- $\Phi = \{\beta_1, \beta_2, \dots, \beta_q\}$ is the outputs alphabet,
- $\lambda: T \rightarrow Ev \times C$ is a labelling function of transitions, where

$C = \{C_1, C_2, \dots, C_{|T|}\}$ is the set of input conditions in which every C_i is a Boolean function on Σ ; when a C_i is always true it is denoted as “=1”, and

$Ev = \{Ev_1, Ev_2, \dots\}$ is the set of input events conditions; every Ev_i is a Boolean function of input events, built on Σ ; events are denoted as α_i_0 and α_i_1 for representing that the input value changes from 1 to 0, or from 0 to 1 respectively. A condition Ev_i may not exist; this is denoted as “ ϵ ”.

In an IPN, a transition t_j will be fired if a) t_j is enabled, **and** b) condition C_j is true, **and** c) the event in $E(t_j)$ occurs.

- $\varphi: R(Q, M_0) \rightarrow (Z^+)^q$ is an output function, that associates with each marking in $R(Q, M_0)$ a q -entry output vector, where $q = |\Phi|$ is the number of outputs. φ is represented by a $q \times |P|$ matrix, such that if the output symbol β_i is present (turned on) every time that $M(p_j) \geq 1$, then $\varphi(i, j) = 1$, otherwise $\varphi(i, j) = 0$.

The state equation of PN is completed with the marking projection $Y_k = \varphi M_k$, where $Y_k \in (Z^+)^q$ is the k -th output vector of the IPN.

Definition 4: A place $p_i \in P$ is said to be *observable* if the i -th column vector of φ (denoted as $\varphi(\bullet, i)$) is not null. Otherwise it is *non-ob-*

servable. $P = P^{obs} \cup \square P^{nobs}$, and $P^{obs} \cap \square P^{nobs} = \emptyset$; where P^{obs} is the set of observable places and P^{nobs} the set of non-observable places.

3. Input-Output Identification Techniques

The Problem statement considers a DES composed of a Plant and a Programmable Logic Controller (PLC) operating in a closed-loop. It is assumed that the data exchanged between Plant and the PLC consists of binary signals. The input signals of the PLC (outputs of the Plant) are generated by the sensors of the Plant. The output signals of the PLC (inputs of the Plant) control the actuators of the Plant. The external behaviour of such a DES can be observed by the change of values of all input/output (I/O) signals exchanged between the controller and the plant.

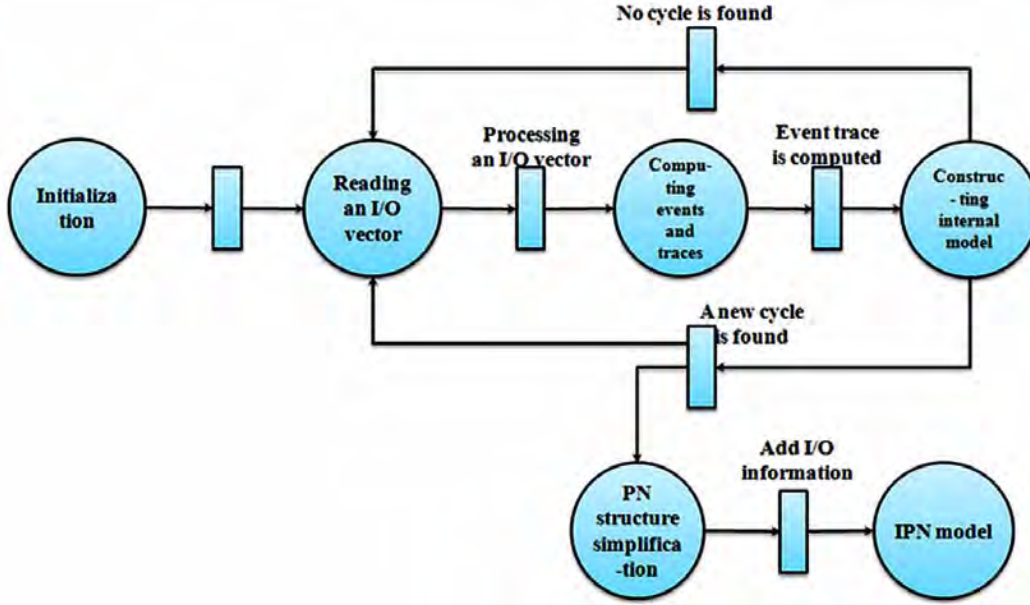
The signal values of all inputs and outputs at a given instant j can be arranged to form an I/O vector $I/O(j) = (\alpha_1(j) \alpha_2(j) \dots \alpha_r(j) | \phi_1(j) \phi_2(j) \dots \phi_q(j))^T$. At the end of every PLC cycle, the current value of all Inputs and Outputs can be easily captured and recorded in a data base. Each new observed I/O vector (when at least one I/O changes its value) belongs to an I/O vector alphabet (Estrada et al., 2015).

3.1. Stepwise Method

The identification problem can be stated as follows: given a DES whose only available information is a single observed I/O sequence w arbitrarily large and an accuracy parameter κ , the aim of the identification process is to obtain a safe IPN model (Q, M_0) such that $\mathcal{L}_{out}^\kappa(Q, M_0) = \mathcal{L}^\kappa(S)$. Where $\mathcal{L}_{out}^\kappa$ is the observed language and $\mathcal{L}^\kappa(S)$ is the system language of length κ . The parameter κ is used to adjust the accuracy of the identified model, similarly as proposed in (Klein, et al., 2005).

The I/O vector sequence w is progressively built by adding new observed I/O vectors. Variations

Figure 2. Stages of the identification algorithm



in any of input or output between two consecutive I/O vectors in w are considered as events $\tau(j)$, which are relevant to the identification method in which a string of observed events are computed (Estrada, Lopez, & Lesage, 2014).

The events can be depicted in a more compact way. A symbolic input event $\lambda'(\tau(j))$ is a string representation of the input event vector $\beta(\tau(j))$. I_i-1 denotes the change from 0 to 1 of the input I_i ; similarly, I_i-0 denotes the change from 1 to 0 of the input I_i . $\lambda'(\tau(j))$.

Then for a sequence w , a sequence τ of observed events $\tau(j)$, $\tau = \tau(1) \tau(2) \dots \tau(j) \dots$ is processed. During the process, if the difference between two consecutive I/O vectors has not been observed before, a new event e_j is created ($\tau(j) = e_j$). The main steps of the method are given in the PN diagram of Figure 2.

Example 1: Consider a DES that has three output signals: $\Phi = \{A, B, C\}$ and three input signals $\Sigma = \{a, b, c\}$. The entries of the binary I/O vectors have the following correspondence:

$$\begin{bmatrix} a & b & c & | & A & B & C \end{bmatrix}^T.$$

An I/O sequence is progressively observed. The first measured I/O vector of the DES is:

$$w(1) = \begin{bmatrix} 0 & 0 & 0 & | & 0 & 0 & 0 \end{bmatrix}^T.$$

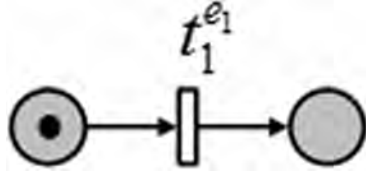
When a second I/O vector

$$w(2) = \begin{bmatrix} 1 & 0 & 0 & | & 1 & 0 & 0 \end{bmatrix}^T$$

is processed, the event vector

$$\tau(1) = e_1 = \begin{bmatrix} 1 & 0 & 0 & | & 1 & 0 & 0 \end{bmatrix}^T$$

is computed; the input event vector is $\lambda(1) = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^T$ and its corresponding symbolic input event is $\lambda'(1) = a-1$, i.e. the rising edge of a .

Figure 3. PN representing e_1


Considering a value $\kappa = 1$, we can compute the first event trace $\tau^1(1) = e_1$. This event trace is related to a transition of the IPN. The IPN constructed after observing two I/O vectors is shown in Figure 3.

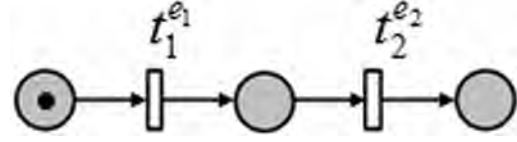
When a third I/O vector

$$w(3) = \begin{bmatrix} 1 & 0 & 0 & | & 0 & 0 & 0 \end{bmatrix}^T$$

is processed,

$$\tau(2) = e_2 = \begin{bmatrix} 0 & 0 & 0 & | & -1 & 0 & 0 \end{bmatrix}^T,$$

$\lambda(2) = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}^T$ and $\lambda'(2) = \varepsilon$ are computed and the model is updated, as shown in Figure 4.

 Figure 4. PN representing the sequence $e_1 e_2$


Until 8th I/O vector, the situation is quite similar: every new event is processed and the model is updated (see Box 1).

When the 9th vector

$$w(9) = \begin{bmatrix} 1 & 0 & 0 & | & 1 & 0 & 0 \end{bmatrix}^T$$

is processed, the event

$$\tau(8) = e_1 = \begin{bmatrix} 1 & 0 & 0 & | & 1 & 0 & 0 \end{bmatrix}^T$$

is computed and the trace $\tau^1(8)$ is identified through Step 3 as an already observed trace e_1 . Since it leads to the same marking than the input place of $t_1^{e_1}$, such a place and the output place of $t_7^{e_7}$ can be merged as shown in Figure 5.

Box. 1.

$$w = \left(\begin{array}{c} \xrightarrow{e_1} \\ \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \end{array} \begin{array}{c} \xrightarrow{e_2} \\ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \end{array} \begin{array}{c} \xrightarrow{e_3} \\ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \end{array} \begin{array}{c} \xrightarrow{e_4} \\ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ -1 \end{bmatrix} \end{array} \begin{array}{c} \xrightarrow{e_5} \\ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 1 \end{bmatrix} \end{array} \begin{array}{c} \xrightarrow{e_6} \\ \begin{bmatrix} 1 \\ 0 \\ 0 \\ -1 \\ 0 \\ -1 \end{bmatrix} \end{array} \begin{array}{c} \xrightarrow{e_7} \\ \begin{bmatrix} 1 \\ 0 \\ 0 \\ -1 \\ 0 \\ 0 \end{bmatrix} \end{array} \begin{array}{c} \xrightarrow{e_8} \\ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \end{array} \right)$$

Figure 5. Internal model for the first detected cycle

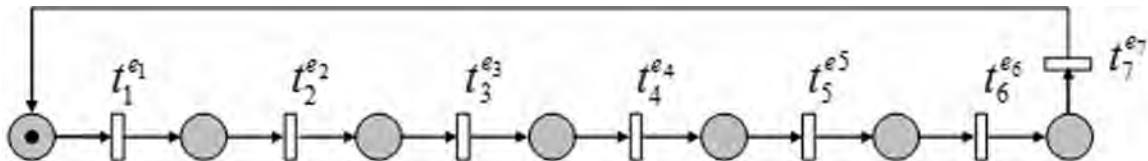
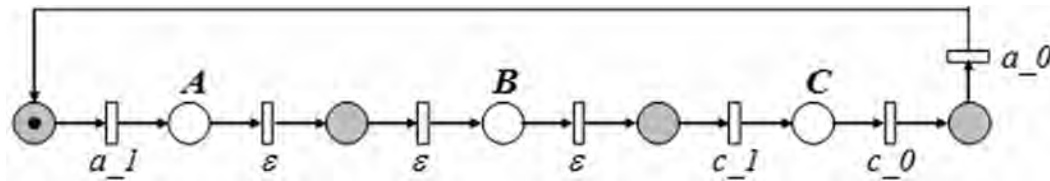


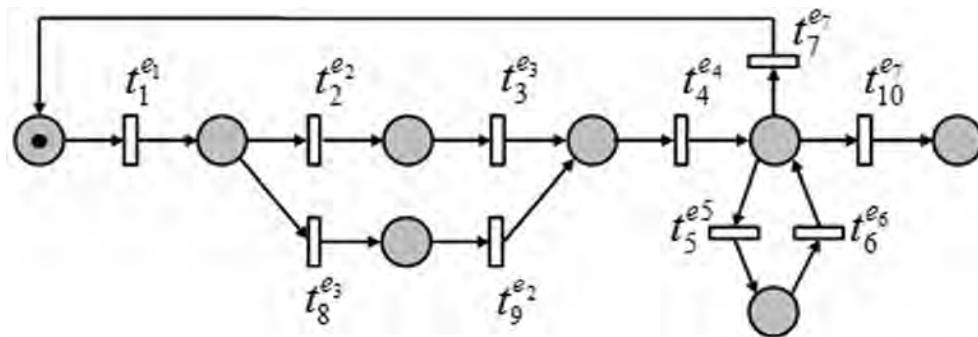
Figure 6. IPN for the first detected cycle



Box 2.

$\begin{matrix} \xrightarrow{e_3} \\ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \end{matrix}$	$\begin{matrix} \xrightarrow{e_2} \\ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \end{matrix}$	$\begin{matrix} \xrightarrow{e_4} \\ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \end{matrix}$	$\begin{matrix} \xrightarrow{e_7} \\ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \end{matrix}$	$\begin{matrix} \xrightarrow{e_1} \\ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \end{matrix}$	$\begin{matrix} \xrightarrow{e_3} \\ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \end{matrix}$	$\begin{matrix} \xrightarrow{e_2} \\ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \end{matrix}$	$\begin{matrix} \xrightarrow{e_4} \\ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \end{matrix}$	$\begin{matrix} \xrightarrow{e_8} \\ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \end{matrix}$
$\begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$
$\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$
$\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$
$\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$
$\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$

Figure 7. PN corresponding to the whole I/O sequence



Since a cycle is found, leading to an intermediate IPN model shown in Figure 6.

Simultaneously to the creation of the intermediate IPN, more I/O vectors are added to the observed sequence and the PN is updated (see Box 2).

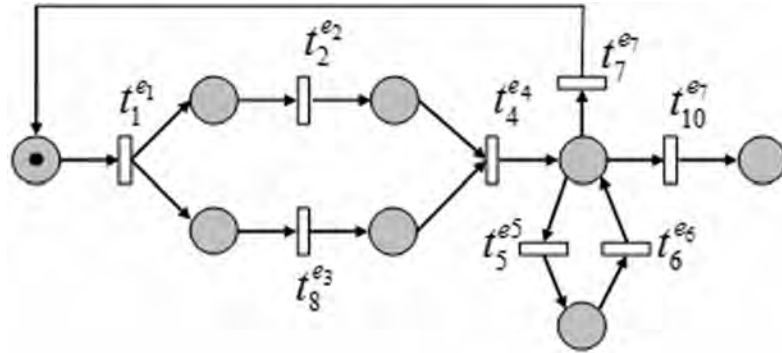
Two more cycles are found in this sequence and intermediate IPN models are created. We show only the PN obtained after finding the second cycle (Figure 7) and its equivalent model transformed by analysing concurrency (Figure 8).

3.2. Two-Phase Method

The purpose in this method is not only to compute an IPN model in which the observed sequence is reproducible, but also to achieve expressivity and compactness in the identified model allowing representing causal relationship and concurrency of the involved operations.

The method consists of two main steps which are outlined below.

Figure 8. Equivalent internal model representing concurrency



Step 1: Discovering the reactive input/output behavior. In this step is determined the observable part of the IPN consisting of subnets, named fragments, composed by observable places labeled with output symbols, and transitions labeled with algebraic expressions of input symbols. From the sequence w , a corresponding sequence of transitions S is obtained.

Step 2: Determining the non-observable part of the IPN and the initial marking M_0 . The sequence S is processed for obtaining causal and concurrency relationships useful for determining the non-observable places that relate the fragments such that S (thus w) can be executed from M_0 .

Observable Behavior

Procedure 1 summarizes the steps of the method to identify the {PLC + Plant} observable behavior. (Estrada, Lopez & Lesage, 2015)

Example 2: The purpose of this system (Figure 9) is to sort parcels according to their size. It has 9 signal sensors from the system: $a_0, a_1, a_2, b_0, b_1, c_0, c_1, k_1, k_2$, and 4 signals to the actuators: A+, A-, B, C. This example has been used in other publications (Estrada, 2014), (Estrada, 2014).

The influence of some input signals over the outputs setting is observed at the same PLC

Procedure 1. Computing the observable IPN components

Input: I/O sequence w

Output: Observable incidence matrix φW and labeling transition function λ

1) Analyze sequence w in order to

- Compute events vector sequence
- Compute elementary events
- Compute Direct and Indirect Causality Matrices
- Construct Output Event Firing Functions
- Find Input events with differed influence

2) Use computed data in the previous step to

- Compute transitions of the IPN and their labeling λ
- Compute observable incidence matrix φW

cycle. In order to discover such an input/output direct relationship, we analyze the relative frequency of the occurrence of both input events IE_i and output events OE_k , with respect to the occurrence of OE_k along the whole sequence of events.

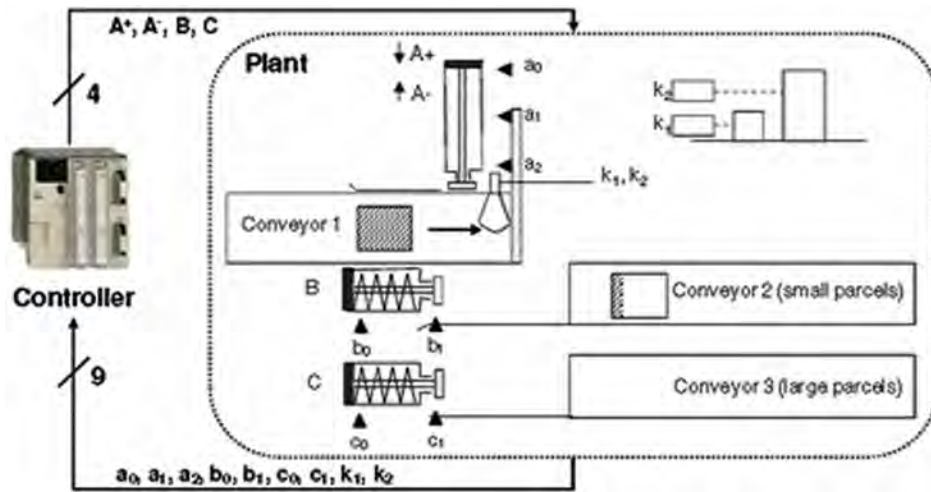
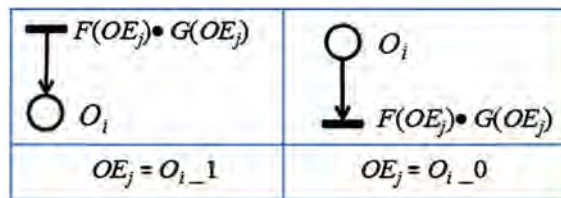
Using computing firing functions of output events can be easily translated into IPN fragments, as shown in Figure 10.

For the *Example 2*, $D = \emptyset$; the computed PN fragments are shown in Figure 11.

The behavior is reproduced by merging such conditions into a unique transition, which is labeled by a firing function computed from individual firing functions of each output event. This is captured in the model as a fusion of IPN fragments as shown in Figure 12.

The corresponding partial model is shown in Figure 13. The inferring procedure which al-

Figure 9. Layout of the system case study


 Figure 10. Rising and falling edges of output O_i


allows discovering the non-observable behavior is described in next section.

Non-Observable Behavior

Once the events and transitions of the net are completely defined, we need to add non-observable

places to translate an aggregation of the non-observable dynamics of the process in such a way that the global PN will reproduce the whole behavior of the system.

Given an observable IPN model whose structure is $G^{obs}=(P^{obs}, T, Pre^{obs}, Post^{obs})$ and a transitions sequence $S = t_1 t_2 \dots t_j \dots \in T^*$ reproducing the I/O sequence w , an ordinary PN structure $G^{nobs}=(P^{nobs}, T, Pre^{nobs}, Post^{nobs})$ that reproduces S and an initial marking M_0 enabling S must be found; (G^{obs}, M_0) must be safe.

Thus, the PN structure of the complete identified model is $G=(P, T, Pre, Post)$ with $P=P^{obs} \cup P^{nobs}$, $Pre=Pre^{obs} \cup Pre^{nobs}$, $Post=Post^{obs} \cup Post^{nobs}$.

Figure 11. IPN fragments for Example 1

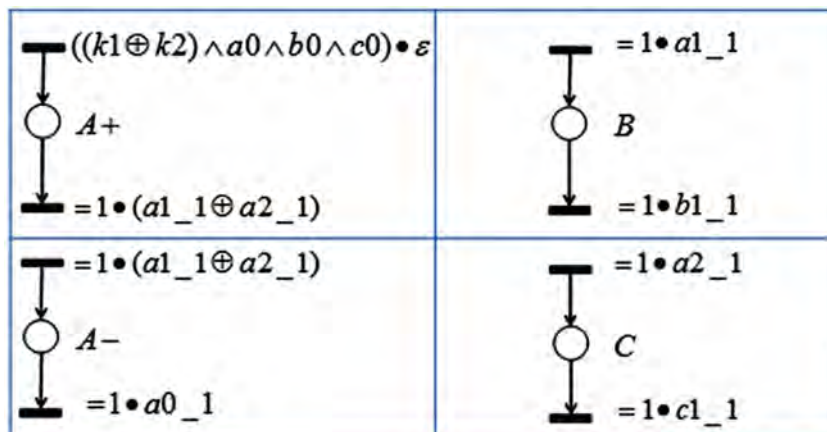
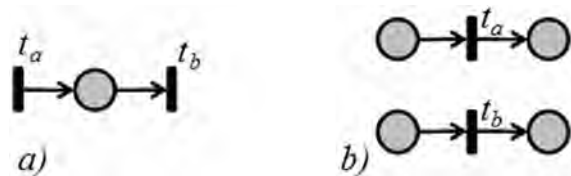


Figure 14. Structures that represent $t_a < t_b$ a) shows a causal relationship from t_a to t_b , whereas b) shows a concurrent relationship between t_a and t_b



Example 3 (Cont.): By considering the couples of consecutive non-concurrent transitions in Seq' which in this example is only (t_3, t_5) —, the places: (t_1, t_2)

(t_2, t_3) $(t_3, t_1 t_5)$ $(t_4, t_1 t_5)$ (t_5, t_6) (t_6, t_7) and $(t_2 t_7, t_4)$ are computed. The PN structure and the computed initial marking are shown in Figure 16.

Finally, the identified IPN of the sorting system described in Example 2 is obtained by merging the observable model in Figure 13 and the non-observable model from Figure 16 after applying the places correction. We can also delete non-observable implicit places. Then, the IPN shown in Figure 17 is the final model after the merging; it reproduces w .

Figure 15. Concurrent threads synchronized by a transition

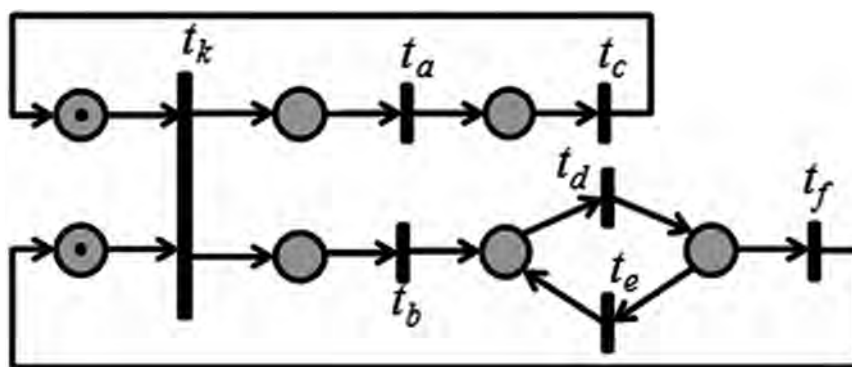


Figure 16. (G^{nobs}, M_0) the non-observable IPN of Example 1

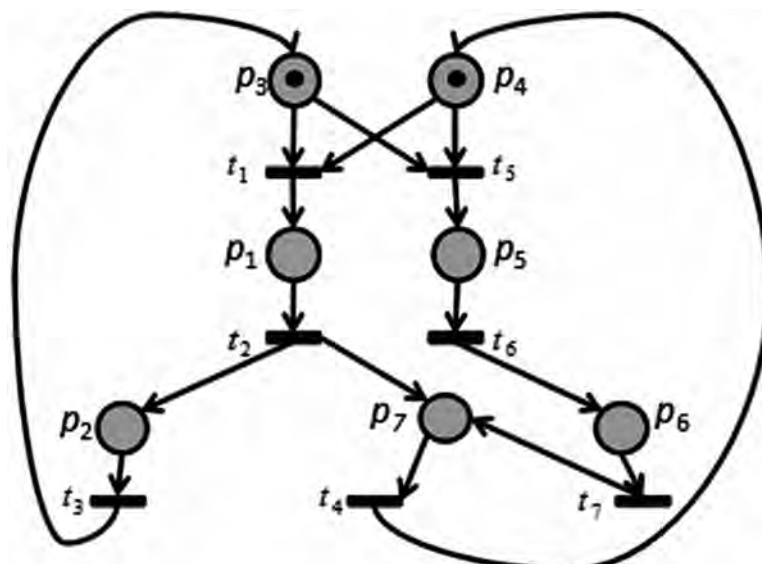


Figure 17. Final IPN model for the process in Example 1

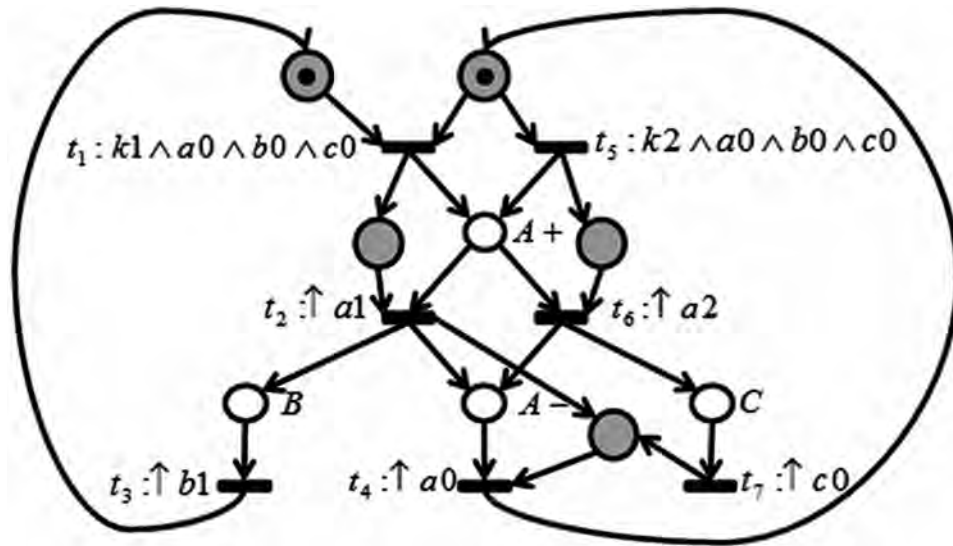
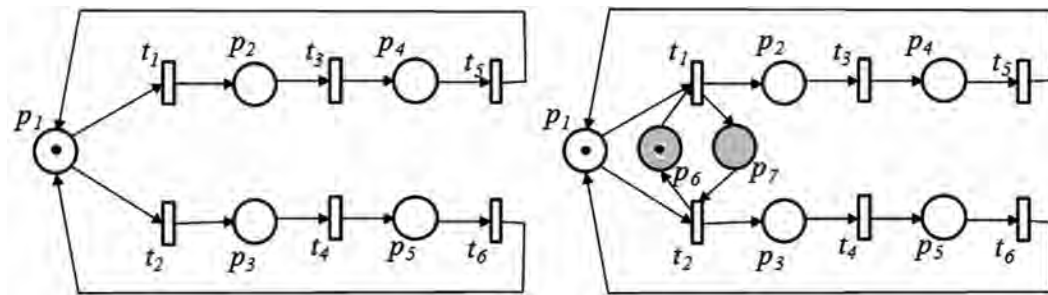


Figure 18. Sequencing choice structures



FUTURE RESEARCH DIRECTIONS

Several proposals address the duration of activities and waiting times in a discrete event process. Timed PN models are obtained from event sequences whose sampling dates of every event are provided. Relevant works on the matter are (Meda, and Medina, 2011) (Muñoz et al., 2014) (Basile, Chiacchio, and Coppola 2013). Recently, in (Rodríguez, Tapia, & López, 2016) a simple and fast algorithm to identify controlled discrete event processes from timed input-output vector sequences has been presented. The method is based on a three-stage strategy: 1) the observable components of the TTIPN are first obtained, 2) the

non-observable part is inferred, and 3) the time parameters associated to transitions are obtained.

Additionally, other important issues regarding the PN structure have to be dealt for discovering the relative frequency of occurrence among the sub-processes represented in a PN model. It is interesting to know the sequencing of such sub-processes or at least the branching factor in alternative production sequences. Figure 18a shows a simple PN model involving two alternative sub-processes. In Figure 18b the model is complemented with a loop formed by p6 and p7, which states a policy that alternates the execution of the internal processes.

CONCLUSION

Identification methods that yield Petri net models have been overviewed. They can be used as automated modelling technique for reverse engineering of unknown or ill known processes. Two input-output black-box methods for PLC-based controlled plants have been presented. First, a stepwise method builds detailed PN models is outlined. The identified model represents the effect of every event in the IPN; this causes huge models close to finite automata. Second, a two-phase method yielding more compact models is described. The separation between the observable and the internal behaviour is clearer allowing a more comprehensible description of causal and concurrent relations among the events. Both approaches require long observations to obtain approximations closer to the actual system behaviour.

REFERENCES

- Angluin, D. (1988). Queries and Concept Learning. *Machine Learning*, 2(4), 319–342. doi:10.1007/BF00116828
- Biermann, A. W., & Feldman, J. A. (1972). On the synthesis of finite-state machines from samples of their behaviour. *IEEE Transactions on Computers*, 21(6), 592–597. doi:10.1109/TC.1972.5009015
- Booth, T. L. (1967). *Sequential machines and automata theory*. New York: John Wiley and Sons, Inc.
- Cabasino, M. P., Giua, A., & Seatzu, C. (2007). Identification of Petri Nets from Knowledge of Their Language. *Discrete Event Dynamic Systems*, 17(4), 447–474. doi:10.1007/s10626-007-0025-0
- Cook, J. E., Du, Z., Liu, C., & Wolf, A. (2004). Discovering models of behaviour for concurrent workflows. *Computers in Industry*, 53(3), 297–319. doi:10.1016/j.compind.2003.10.005
- David, R., & Alla, H. (2010). *Discrete, Continuous, and Hybrid Petri Nets*. Berlin: Springer-Verlag. doi:10.1007/978-3-642-10669-9
- Dotoli, M., Fanti, M. P., & Mangini, A. M. (2008). Real time identification of discrete event systems using Petri nets. *Automatica*, 44(5), 1209–1219. doi:10.1016/j.automatica.2007.10.014
- Dotoli, M., Fanti, M. P., Mangini, A. M., & Ukovich, W. (2011). Identification of the unobservable behaviour of industrial automation systems by Petri nets. *Control Engineering Practice*, 19(9), 958–966. doi:10.1016/j.conengprac.2010.09.004
- Estrada-Vargas, A. P., Lesage, J.-J., & López-Mellado, E. (2014). A Stepwise Method for Identification of Controlled Discrete Manufacturing Systems. *International Journal of Computer Integrated Manufacturing*, 28(2), 187–199. doi:10.1080/0951192X.2013.874591
- Estrada-Vargas, A. P., López-Mellado, E., & Lesage, J.-J. (2014). Input-Output Identification of Controlled Discrete Manufacturing Systems. *International Journal of Systems Science*, 45(3), 456–471. doi:10.1080/00207721.2012.724098
- Estrada-Vargas, A. P., López-Mellado, E., & Lesage, J.-J. (2015). A Black-Box Identification Method for Automated Discrete-Event Systems. *IEEE Transactions on Automation Science and Engineering*, (90): 1–16. doi:10.1109/TASE.2015.2445332
- Fanti, M. P., & Seatzu, C. (2008). Fault diagnosis and identification of discrete event systems using Petri nets. *Proc. of the 9th Int. Workshop on Discrete Event Systems*, 432–435. doi:10.1109/WODES.2008.4605985
- Giua, A., & Seatzu, C. (2005). Identification of free-labeled Petri nets via integer programming. *Proc. of the 44th IEEE Conference on Decision and Control, and the European Control Conference*, 7639–7644. doi:10.1109/CDC.2005.1583395

- Gold, M. E. (1967). Language identification in the limit. *Information and Control*, 10(5), 447–474. doi:10.1016/S0019-9958(67)91165-5
- Hiraishi, K. (1992). Construction of Safe Petri Nets by Presenting Firing Sequences. *Lecture Notes in Computer Science*, 616, 244–262. doi:10.1007/3-540-55676-1_14
- Kella, J. (1971). Sequential Machine Identification. *IEEE Transactions on Computers*, 20(3), 332–338. doi:10.1109/T-C.1971.223241
- Klein, S., Litz, L., & Lesage, J.-J. (2005). Fault detection of Discrete Event Systems using an identification approach. *Proc. 16th IFAC World Congress*, 1–6. doi:10.3182/20050703-6-CZ-1902.01440
- Meda, M., & López, E. (2001). A passive method for on-line identification of discrete event systems. *Proc. of the IEEE Int. Conference on Decision and Control*, 4990–4995.
- Meda, M., & López, E. (2003). Required Transition Sequences for DES identification. *Proc. of the IEEE Conference on Decision and Control*, 3778–3782.
- Meda, M., Ramírez, A., & López, E. (2000). Asymptotic Identification for DES. *Proc. IEEE Conference on Decision and Control*, 2266–2271.
- Ould El Medhi, S., Leclercq, E., & Lefebvre, D. (2006). Petri nets design and identification for the diagnosis of discrete event systems. *IAR Annual Meeting*, 1–6.
- Ould El Mehdi, S., Bekrar, R., Messai, N., Leclercq, E., Lefebvre, D., & Riera, B. (2012). Design and identification of stochastic and deterministic stochastic Petri nets. *IEEE Trans. on Systems, Man and Cybernetics. Part A*, 42(4), 931–946.
- Richetin, M., Naranjo, M., & Luneau, P. (1984). Identification of Automata by Sequential Learning. *Pattern Recognition Letters*, 2(6), 379–385. doi:10.1016/0167-8655(84)90004-7
- Roth, M., Lesage, J.-J., & Litz, L. (2009). An FDI Method for Manufacturing Systems Based on an Identified Model. *Proc. of IFAC Symposium on Information Control Problems in Manufacturing (INCOM 2009)*, 1389–1394. doi:10.3182/20090603-3-RU-2001.0058
- Roth, M., Lesage, J.-J., & Litz, L. (2010). Black-box identification of discrete event systems with optimal partitioning of concurrent subsystems. *Proc. of the American Control Conference*, 2601–2606. doi:10.1109/ACC.2010.5530540
- Valiant, L. G. (1984). A theory of the learnable. *Communications of the ACM*, 27(11), 1134–1142. doi:10.1145/1968.1972
- Van der Aalst, W., Weijters, T., & Maruster, L. (2004). Workflow Mining: Discovering Process Models from Event Logs. *IEEE Transactions on Knowledge and Data Engineering*, 16(9), 1128–1142. doi:10.1109/TKDE.2004.47
- Veelenturf, L. P. J. (1978). Inference of Sequential Machines from Sample Computations. *IEEE Transactions on Computers*, 27(2), 167–170. doi:10.1109/TC.1978.1675053
- Veelenturf, L. P. J. (1981). An Automata theoretical approach to developing learning neural networks. *Cybernetics and Systems*, 12(1-2), 179–202. doi:10.1080/01969728108927670

ADDITIONAL READING

- Basile, F., Chiacchio, P., & Coppola, J. (2013). An approach for the identification of time Petri net systems. In *Proceedings of IEEE 18th Conference on Emerging Technologies & Factory Automation (ETFA)*, Cagliari, Italy, pp. 1–8.
- Cabasino, M. P., Darondeau, P., Fanti, M. P., & Seatzu, C. (2015). Model identification and synthesis of discrete-event systems. In M. Zhou, H.-X. Lim, & M. Weijnen (Eds.), *Contemporary Issues in System Science and Engineering*. IEEE/Wiley.

Estrada-Vargas, A. P., Lopez-Mellado, E., & Lesage, J.-J. (2010). A comparative analysis of recent identification approaches for discrete event systems. *Mathematical Problems in Engineering*.

Meda-Campaña, M. E., & Medina-Vazquez, S. (2011). Synthesis of timed Petri net models for on-line identification of Discrete Event Systems. in *Proc. 9th IEEE International Conference on Control and Automation (ICCA)*, Santiago, Chile, pp. 1201-1206.

Muñoz, D. M., Correcher, A., García, E., & Morant, F. (2014). Identification of Stochastic Timed Discrete Event Systems with st-IPN. *Mathematical Problems in Engineering*, 2014, 1–21. doi:10.1155/2014/835312

Rodríguez-Pérez, E., Tapia-Flores, T., & López-Mellado, E. (2016). Identification of Timed Discrete Event Processes. Building Input-Output Petri Net Models. *Workshop on Algorithms & Theories for the Analysis of Event Data, Torun, Poland, ATAED@Petri Nets/ACSD 2016*, pp. 153-167.

KEY TERMS AND DEFINITIONS

Discrete Event Model: Accurate representation of the system behaviour expressed in a fine formalism, usually a finite automaton or a Petri net.

Discrete Event Systems: A class of dynamical systems in which the behaviour is characterised by successions of steady states delimited by events in general asynchronous.

Identification Methods: Techniques that build systematically formal models from the observation of external behaviour.

Input-Output Identification: Identification approach based on input-output sequences observed from a discrete event process.

Petri Nets: Formalism for specifying discrete event systems behaviours, allowing describing states, events, causal and concurrent relations, information exchange, resource allocation, and other complex behaviours.

Research and Development on Software Testing Techniques and Tools

S**Tamilarasi T***VIT University, India***M. Prasanna***VIT University, India*

INTRODUCTION

Software testing is the process of executing the program with the intent of finding the error. It is done to check if the system meets the requirements specified and be executed successfully in the intended environment. Testing checks the system whether it is “Fit for Purpose”, also verifies whether the system does what it is expected to do. Software testing techniques are the different approaches and ways of ensuring that a software application is fully tested. Now days we can get lots of Software Testing Tools in the market. Selection of tools is totally based on the project requirements & commercial or free tools (Open Source Tools) you are interested. Off Course, free Testing Tools may have some limitation in the features list of the product, so it's totally based on what are you looking for & is that your requirement fulfil in free version or go for paid Software Testing Tools.

BACKGROUND

As software applications get more ever complex and intertwined and with the large number of different platforms and devices that need to be tested. Software testing techniques encompass everything from unit testing individual modules, integration testing an entire system to specialized forms of testing such as security and performance

[1]. Testing is a process used to help identify the correctness, completeness and quality of developed computer software. With that in mind, testing can never completely establish the correctness of computer software. One definition of testing is “the process of questioning a product in order to evaluate it”, where the “questions” are things the tester tries to do with the product, and the product answers with its behavior in reaction to the probing of the tester.

SOFTWARE TESTING TECHNIQUES

Testing Outlook

Testing outlook has an impact on the software testing process. Consider the following definition:

Testing is the process to prove that the software works correctly.

This definition sounds good, but the person who developed the software will only try to show that the software works correctly. This is the typical psychology of testing. The software will work correctly for the inputs that are given by that person which will obtain correct results. If some other input was given, the software will obtain wrong results which is not acceptable in a commercial environment.

Now consider this definition:

Testing is a process to prove that the software does not work.

If the aim of the test engineer is to prove that the software does not work, then the process can be considered as good. If the software performs well, then you can say that the software is very reliable. If the software works efficiently, after some days of testing, it does not mean that the software has no bugs at all.

So the definition would be:

Testing is the process to detect the defects and minimize the risk associated with the residual defects.

Verification and Validation

While going for testing, the two terms verification and validation have to be differentiated. Barry Boehm defines these terms as follows:

- **Verification:** “Are we Building the product right?”
- **Validation:** “Are we building the right product?”

Verification is done by the development team to ensure that the software is as per the specifications in the SRS document. It is to check whether the software conforms to specifications. Validation is to check whether the software meets the customer expectations. It is carried out with the involvement of the client.

Levels of Testing

Unit Testing

Unit testing aims at testing each of the components that a system is built upon. As long as each of them works as they are defined to, then the system as a whole has a better chance of working together. Where possible, all units that could

possibly fail are tested at least one. Judgment is needed to decide what parts need testing. Some things such as accessors/mutators generally don't need to be tested. Components that are erroneous can be detected earlier. The scope of unit tests is even smaller than your traditional tests. If errors are detected, they are generally easier to fix.

Integration Testing

The purpose of integration testing is to verify the functional, performance, and reliability between the modules that are integrated. It is a software development process which program units are combined and tested as groups in multiple ways. In this context, a unit is defined as the smallest testable part of an application. Integration testing can expose problems with the interfaces among program components before trouble occurs in real-world program execution. Integration testing is component of Extreme Programming (XP), a pragmatic method of software development.

System Testing

- In system testing the behavior of whole system/product is tested as defined by the scope of the development project or product.
- It may include tests based on risks and/or requirement specifications, business process, use cases, or other high level descriptions of system behavior, interactions with the operating systems, and system resources.
- System testing is most often the final test to verify that the system to be delivered meets the specification and its purpose.
- System testing is carried out by specialists testers or independent testers.
- System testing should investigate both functional and non-functional requirements of the testing.

Acceptance Testing

Acceptance testing, a testing technique performed to determine whether or not the software system has met the requirement specifications. The main purpose of this test is to evaluate the system's compliance with the business requirements and verify if it has met the required criteria for delivery to end users. The acceptance test cases are executed against the test data or using an acceptance test script and then the results are compared with the expected ones.

Test Plan

Test planning, the most important activity to ensure that there is initially a list of tasks and milestones in a baseline plan to track the progress of the project. It also defines the size of the test effort. It is the main document often called as master test plan or a project test plan and usually developed during the early phase of the project.

Test Oracles

Oracle is a repository or knowledge. Test oracles are the people used to check the correctness of the program for the given test cases. Human test oracles are used extensively if the program does not work. A common way is to give the program and the input test data to a friend and ask them to derive the output from the program. If the output is same as expected, then it is said to be correct, otherwise there is a defect in the software.

Test Cases

A test case is a document, which has a set of test data, preconditions, expected results and postconditions, developed for a particular test scenario in order to verify compliance against a specific requirement.

Typical Test Case Parameters

- Test Case ID
- Test Scenario
- Test Case Description
- Test Steps
- Prerequisite
- Test Data
- Expected Result
- Test Parameters
- Actual Result
- Environment Information
- Comments

TEST DESIGN TOOLS

Test design tools help to create test cases, or at least test inputs (which is part of a test case). If an automated oracle is available, then the tool can also make the expected result, so in point of fact it can generate test cases.

Agile Designer

Agile Designer is a one stop shop for requirements definition and test cases. Agile Designer lets you design the perfect set of *test cases*, find and make the 'right' data, link expected results to tests and data, cut defect creation by up to 95% and build clearly-defined outsourcing work packages from visual flows:

- Maximize Test Coverage
- Manage Changing Requirements
- Find and Make Test Data
- Continuous Integration
- Service Virtualization Testing
- Link Expected Results
- Agile Processes
- Optimize Software Tests
- Reduce Software Testing Costs

- Manage Software Outsourcing
- Estimate Development Costs
- Reverse Engineer Requirements

Bender RBT Test Case Design Tool

Bender RBT has two test design engines: Cause-Effect Graphing and Pair Wise. The Graphing engine is the only test tool that designs tests that ensure that defects are propagated to an observable point. Its tests maximize coverage while minimizing the number of tests. It does a logical consistency check on the requirements and can generate an as built spec from the test model - critical for agile projects. The Pair Wise engine supports both orthogonal pairs and optimized pairs. Both engines do strong negative testing.

Coronys Test Automation Framework

Framework comprises real-time test execution engine, script development environment, test flow drag-and-drop creation, test management. Programmable asynchronous events on communication channels are supported for protocol emulation, data capturing and CLI automation. Serial, Telnet, TCP, UDP, HTTP and more protocols are instrumented. DLL, ActiveX, COM-objects are easily integrated into scripts. Test cases aggregated into suites are launched as jobs targeted to specific setups. Incorporate job launch directives into the software build to enhance your Continuous Integration Process. Built-in MySQL database enables querying for in-depth results analysis.

LOAD AND PERFORMANCE TESTING TOOLS

In client/server applications, many users access the server application simultaneously. In such a case, the software has to be tested for its performance.

A speedy and responsive site is directly correlated with higher levels of visitors and better user experience. *Performance testing* plays a vital role. It is the only way to determine an application's behaviour at different levels of load and users.

1. WebLOAD
2. LoadComplete
3. Apache JMeter
4. LoadRunner

WebLOAD

Load and performance testing tool for web applications. WebLOAD lets you perform load and stress testing on any internet application using Ajax, Adobe Flex, .NET, Oracle Forms, HTML5 and many more technologies. You can generate load from the cloud and on-premises machines. WebLOAD's strengths are its ease of use with features like DOM-based recording/playback, automatic correlation and JavaScript scripting language. The tool supports large-scale performance testing with heavy user load and complex scenarios, and provides clear analysis on the functionality and performance of the web application.

LoadComplete

Easy and affordable performance testing tool. LoadComplete enables you to create and execute realistic load tests for websites and web apps. It automates creating realistic load tests by recording user interactions and simulating these actions with hundreds of virtual users either from your local computers or from the cloud. LoadComplete helps you check your web server's performance under a massive load, determine its robustness and estimate its scalability. It also provides detailed metrics and reports that help you get in-depth insights into infrastructure performance, application behavior, and end user experience.

Apache JMeter

Open source load testing tool. It is a Java platform application. It is mainly considered as a performance testing tool and it can also be integrated with the test plan. In addition to the load test plan, you can also create a *functional test plan*. This tool has the capacity to be loaded into a server or network so as to check on its performance and analyze its working under different conditions. Initially, it was introduced for testing the web applications, but later its scope had widened. It is of great use in testing the functional performance of the resources such as Servlets, Perl Scripts and JAVA objects.

HP LoadRunner

This is a HP product which can be used as a performance testing tool. This can be bought as a HP product from its HP software division. Also, it is very much useful in understanding and determining the performance and outcome of the system when there is actual load. One of the key attractive features of this testing tool is that, it can create and handle thousands of users at the same time. This tool enables you to gather all the required information with respect to the performance and also based on the infrastructure.

TEST MANAGEMENT TOOLS

Features or characteristics of test management tools are:

- To manage the tests (like, keeping track of the same kind of data for a given set of tests, knowing which tests need to run in a common environment, number of tests planned, written, run, passed or failed);
- Scheduling of tests to be executed (manually or by a test execution tool);

- Managing the testing activities (time spent in test design, test execution, whether we are on schedule or on budget);
- Traceability of tests, test results and defects to requirements or other sources;

qTest

Developed by QASymphony, qTest is the only test management solution that allows you to manage your manual, exploratory, and automated testing in a single location, providing the most complete view of your testing coverage and progress on the market. With the help of qTest Connector, it has seamless integrations with JIRA for an entire end-to-end QA solution – but that is not all, it also integrates with other tools like Bugzilla, FogBugz, Rally, VersionOne etc.

PractiTest

An entirely SaaS end-to-end QA and Agile friendly Test management tool. Using their unique and customizable filters you can efficiently organize your requirements, create and run tests, track bugs and generate reports using this tool. It integrates seamlessly with leading bug tracking tools like JIRA, Pivotal tracker, Bugzilla, and Redmine as well as various automation tools such as Selenium, Jenkins etc. However, their API can ensure further customizing for your process' needs. It is not open sourced but is quite affordable. You can benefit from their human methodological support throughout your usage.

Zephyr

Zephyr products are the fastest growing agile test management products in the world with more than 7000 customers in 100 countries. Their real-time solutions offer seamless integrations with tools project teams are already using today like JIRA, Selenium, Jenkins, Bamboo as well as the tools

they will use tomorrow as Agile and DevOps processes evolve overtime.

Test Collab

Test Collab is a modern test management tool which offers complete platform for your application's testing. It offers state-of-the-art integration with all popular bug trackers and test automation tools. Apart from that, it offers time tracking, agile methodology, requirements management, test plans and scheduling.

SOURCE CODE TESTING TOOLS

Source code testing tools are specific to the programming language used for developing the software. As UNIX/ Linux is nowadays used for software development extensively. Lint, a utility used to test the portability of the code. Line profilers used to do the timing analysis. These profilers find out the execution time for the entire program as well as for individual function calls.

EMMA Tool

Emma distinguishes itself from other tools by going after a unique feature combination: development while keeping individual developer's work fast and iterative. Such a tool is essential for detecting dead code and verifying which parts of an application are actually exercised by the test suite and interactive use.

CodeCover Tool

CodeCover is an extensible open source code coverage tool. It provides several ways to increase test quality. It shows the quality of test suite and helps to develop new test cases and rearrange test cases to save some of them. So we get a higher quality and a better test productivity. The main features of CodeCover are: Supports statement coverage,

branch coverage, loop coverage and strict condition coverage; Performs source instrumentation for the most accurate coverage measurement; CLKI interface, for easy use from the command line; Ant interface, for easy integration into an existing build process; Correlation Matrix to find redundant test cases and optimize your test suite.

EvoSuite

To find defects in software, one needs test cases that execute the software systematically, and oracles that assess the correctness of the observed behavior when running these test EvoSuite is a tool that automatically generates test cases with assertions for classes written in Java code. To achieve this, EvoSuite applies a novel hybrid approach that generates and optimizes whole test suites towards satisfying a coverage criterion

JavaCodeCoverage (JaCoCo)

JavaCodeCoverage is a byte-code analyzer tool for test coverage analysis for Java software which neither requires neither the language grammar nor the source code. An important aspect of JavaCodeCoverage is that it stores the coverage information for individual test case thereby facilitating detailed coverage analysis. Another important aspect of JavaCodeCoverage is that it records all vital code elements and test coverage information in open source database software MySQL.

DATA FLOW TESTING

Data flow testing can be performed at two conceptual levels: Static data flow testing, Dynamic data flow testing.

Data Flow Graphs

The data flow graph is a graph consisting of nodes and directed links.

Data Object State and Usage

Data Objects can be created, killed and used. They can be used in two distinct ways: (1) In a Calculation (2) As a part of a Control Flow Predicate. The following symbols denote these possibilities:

1. **Defined:** d - defined, created, initialized etc.
2. **Killed or Undefined:** k - killed, undefined, released etc.
3. **Usage:** u - used for something (c - used in Calculations, p - used in a predicate).

1. Defined (d)

- An object is defined explicitly when it appears in a data declaration.
- Or implicitly when it appears on the left hand side of the assignment.
- It is also to be used to mean that a file has been opened.

2. Killed or Undefined (k)

- An object is killed or undefined when it is released or otherwise made unavailable.
- When its contents are no longer known with certitude (with absolute certainty / perfectness).

3. Usage (u)

- A variable is used for computation (c) when it appears on the right hand side of an assignment statement.

Data Flow Anomalies

An anomaly is denoted by a two-character sequence of actions. For example, ku means that the object is killed and then used, whereas dd means that the object is defined twice without an intervening usage. What is an anomaly depends on the application. There are nine possible two-letter combinations for d, k and u. Some are bugs, some are suspicious, and some are okay.

1. **dd:** Probably harmless but suspicious. Why define the object twice without an intervening usage?
2. **dk:** Probably a bug. Why define the object without using it?
3. **du:** The normal case. The object is defined and then used.
4. **kd:** Normal situation. An object is killed and then redefined.
5. **kk:** Harmless but probably buggy. Did you want to be sure it was really killed?
6. **ku:** A bug. The object does not exist.
7. **ud:** Usually not a bug because the language permits reassignment at almost any time.
8. **uk:** Normal situation.
9. **uu:** Normal situation.

In addition to the two letter situations, there are six single letter situations.

1. **-k:** Possibly anomalous because from the entrance to this point on the path, the variable had not been defined. We are killing a variable that does not exist.
2. **-d:** Okay. This is just the first definition along this path.
3. **-u:** Possibly anomalous. Not anomalous if the variable is global and has been previously defined.
4. **k-:** Not anomalous. The last thing done on this path was to kill the variable.
5. **d-:** Possibly anomalous. The variable was defined and not used on this path. But this could be a global definition.
6. **u-:** Not anomalous. The variable was used but not killed on this path. Although this sequence is not anomalous, it signals a frequent kind of bug. If d and k mean dynamic storage allocation and return respectively, this could be an instance in which a dynamically allocated object was not returned to the pool after use.

Data Flow Anomaly State Graph

Data flow anomaly model prescribes that an object can be in one of four distinct states:

1. **K:** Undefined, previously killed, does not exist.
2. **D:** Defined but not yet used for anything.
3. **U:** Has been used for computation or in predicate.
4. **A:** Anomalous.

These capital letters (K,D,U,A) denote the state of the variable and should not be confused with the program action, denoted by lower case letters.

Forgiving Data: Flow Anomaly Flow Graph

Forgiving model is an alternate model where redemption (recover) from the anomalous state is possible. The point of showing you this alternative anomaly state graph is to demonstrate that the specifics of an anomaly depends on such things as language, application, context, or even your frame of mind. In principle, you must create a new definition of data flow anomaly (e.g., a new state graph) in each situation.

Data Flow Strategies

Various types of data flow testing strategies in decreasing order of their effectiveness are:

1. **All - du Paths (ADUP):** The all-du-paths (ADUP) strategy is the strongest data-flow testing strategy discussed here. It requires that every du path from every definition of every variable to every use of that definition be exercised under some test. The all-du-paths strategy is a strong criterion, but it does not take as many tests as it might seem at first because any one test simultaneously satisfies the criterion

for several definitions and uses of several different variables.

2. **All Uses Strategy (AU):** The all uses strategy is that at least one definition clear path from every definition of every variable to every use of that definition be exercised under some test. Just as we reduced our ambitions by stepping down from all paths (P) to branch coverage (C2), say, can reduce the number of test cases by asking that the test set should include at least one path segment from every definition to every use that can be reached by that definition.
3. **All P-Uses/Some C-Uses Strategy (APU+C):** For every variable and every definition of that variable, include at least one definition free path from the definition to every predicate use; if there are definitions of the variables that are not covered by the above prescription, then add computational use test cases as required to cover every definition.
4. **All C-Uses/Some P-Uses Strategy (ACU+P):** The all c-uses/some p-uses strategy (ACU+P) is to first ensure coverage by computational use cases and if any definition is not covered by the previously selected paths, add such predicate use cases as are needed to assure that every definition is included in some test.
5. **All Definitions Strategy (AD):** The all definitions strategy asks only every definition of every variable be covered by at least one use of that variable, be that use a computational use or a predicate use.
6. **All Predicate Uses (APU), All Computational Uses (ACU) Strategies:** The all predicate uses strategy is derived from APU+C strategy by dropping the requirement that we include a c-use for the variable if there are no p-uses for the variable. The all computational uses strategy is derived from ACU+P strategy by dropping the requirement that include a p-use for the variable if there are no c-uses for the variable.

LOGIC BASED TESTING

“Logic” is one of the most often used words in programmers’ vocabularies but one of their least used techniques. Boolean algebra is to logic as arithmetic is to mathematics. Without it, the tester or programmer is cut off from many test and design techniques and tools that incorporate those techniques.

Knowledge Based System

- The *Knowledge-Based System* (also expert system or “artificial intelligence” system has become the programming construct of choice for many applications that were once considered very difficult.
- Knowledge-based systems incorporate knowledge from a knowledge domain such as medicine, law, or civil engineering into a database. The data can then be queried and interacted with to provide solutions to problems in that domain.
- One implementation of knowledge-based systems is to incorporate the expert’s knowledge into a set of rules. The user can then provide data and ask questions based on that data.
- The user’s data is processed through the rule base to yield conclusions (tentative or definite) and requests for more data. The processing is done by a program called the *inference engine*.
- Understanding knowledge-based systems and their validation problems requires an understanding of formal logic.
- Decision tables are extensively used in business data processing; Decision-table preprocessors as extensions to COBOL are in common use; boolean algebra is embedded in the implementation of these processors.

Decision Tables

- It consists of four areas called the condition stub, the condition entry, the action stub, and the action entry.
- Each column of the table is a rule that specifies the conditions under which the actions named in the action stub will take place.
- The condition stub is a list of names of conditions.

The rule specifies whether a condition should or should not be met for the rule to be satisfied. “YES” means that the condition must be met, “NO” means that the condition must not be met, and “I” means that the condition plays no part in the rule, or it is immaterial to that rule. The action stub names the actions the routine will take or initiate if the rule is satisfied. If the action entry is “YES”, the action will take place; if “NO”, the action will not take place. “Condition” is another word for predicate. Decision-table uses “condition” and “satisfied” or “met”. Let us use “predicate” and TRUE/FALSE. Now the above translations become:

- Action 1 will be taken if predicates 1 and 2 are true and if predicates 3 and 4 are false (rule 1), or if predicates 1, 3, and 4 are true (rule 2).
- Action 2 will be taken if the predicates are all false, (rule 3).
- Action 3 will take place if predicate 1 is false and predicate 4 is true (rule 4).

In addition to the stated rules, we also need a *Default Rule* that specifies the default action to be taken when all other rules fail.

Decision Table Processors

Decision tables can be automatically translated into code and, as such, are a higher-order language.

If the rule is satisfied, the corresponding action takes place. Otherwise, rule 2 is tried. This process continues until either a satisfied rule results in an action or no rule is satisfied and the default action is taken. Decision tables have become a useful tool in the programmers kit, in business data processing.

CONCLUSION

Testing has been widely used as a way to help engineers develop high-quality systems by means of small group of programmers to an organized discipline in software engineering. However, the maturation of testing techniques has been fruitful, but not adequate. Pressure to produce higher-quality software at lower cost is increasing and existing techniques used in practice are not sufficient for this purpose. Fundamental research that addresses the challenging 15 problems, development of methods and tools, and empirical studies should be carried out so that we can expect significant improvement in the way we test software.

FUTURE RESEARCH DIRECTIONS

Researchers should demonstrate the effectiveness of many existing techniques for large industrial software, thus facilitating transfer of these techniques to practice. The successful use of these techniques in industrial software development will validate the results of the research and drive future research. Development of efficient testing techniques and tools that will assist in the creation of high-quality software will become one of the most important research areas in the near future.

REFERENCES

- Aichernig, B. K., Brandl, H., Jöbstl, E., Krenn, W., Schlick, R., & Tiran, S. (2015). Killing strategies for model-based mutation testing. *Software Testing. Verification and Reliability*, 25(8), 716–748. doi:10.1002/stvr.1522
- Akhunzada, A., Ahmed, E., Gani, A., Khan, M. K., Imran, M., & Guizani, S. (2015). Securing software defined networks: Taxonomy, requirements, and open issues. *IEEE Communications Magazine*, 53(4), 36–44. doi:10.1109/MCOM.2015.7081073
- Akram, M., & Ashraf, W. (2015). *Analytical Study of Black Box and White Box Testing for Database Applications*. Academic Press.
- Andreasen, M. S., Nielsen, H. V., Schrøder, S. O., & Stage, J. (2015). Usability in open source software development: Opinions and practice. *Information Technology and Control*, 35(3).
- Beizer, B. (2003). *Software testing techniques*. Dreamtech Press.
- Dalal, S. (2015). *Analysis and design of software testing techniques based on root cause analysis of software system failures*. Academic Press.
- Drasgow, F. (Ed.). (2015). *Technology and Testing: Improving Educational and Psychological Measurement*. Routledge.
- Dreossi, T., Dang, T., Donzé, A., Kapinski, J., Jin, X., & Deshmukh, J. V. (2015, April). Efficient guiding strategies for testing of temporal properties of hybrid systems. In *NASA Formal Methods Symposium* (pp. 127-142). Springer International Publishing. doi:10.1007/978-3-319-17524-9_10
- Garg, P. (2015). Role of Testing Strategies in Improving Quality of Software. *International Journal of Research*, 2(12), 800–805.

Ghani, A., Azim, A., Baharom, S., & Zulzalil, H. (2015). A preliminary study of adaptive random testing techniques. *International Journal of Information Technology & Computer Science*, 19(1), 116–127.

Jaikumar, M., & Ramani, A. V. (2016). Software Defect Prediction–Taxonomy of Literatures and Research Dimensions. *International Journal of Engineering Science*.

Jan, S. R., Shah, S. T. U., Johar, Z. U., Shah, Y., & Khan, F. (2016). An Innovative Approach to Investigate Various Software Testing Techniques and Strategies. *International Journal of Scientific Research in Science, Engineering and Technology*, 2395-1990.

Jorgensen, P. C. (2016). *Software testing: a craftsman's approach*. CRC Press.

Lewis, W. E. (2016). *Software testing and continuous quality improvement*. CRC Press.

Mao, K., Harman, M., & Jia, Y. (2016, July). Sapienz: multi-objective automated testing for Android applications. In *Proceedings of the 25th International Symposium on Software Testing and Analysis* (pp. 94-105). ACM. doi:10.1145/2931037.2931054

Prasad, D. K. (2004). *Software testing tools*. Dreamtech Press.

Rigby, D., & Bilodeau, B. (2015). *Management tools & trends 2015*. London: Bain & Company.

Romano, B. L., de Souza, R. B., & da Cunha, A. M. (2015, April). *Deploying Integrated Environment for Software Testing Tools*. In *Information Technology-New Generations (ITNG), 2015 12th International Conference on* (pp. 777-779). IEEE. doi:10.1109/ITNG.2015.140

Rovida, C., Alépée, N., Api, A. M., Basketter, D. A., Bois, F. Y., Caloni, F., & Fuchs, H. et al. (2015). Integrated Testing Strategies (ITS) for safety assessment. *ALTEX-Alternatives to Animal Experimentations*, 32(1), 25–40. PMID:25413849

Sabharwal, S., Sibbal, R., & Sharma, C. (2015). *A Survey of Testing Techniques for Testing Web based Applications*. Academic Press.

KEY TERMS AND DEFINITIONS

Agile Processes: Introduce Agile processes into your Software Development life cycle.

Continuous Integration: Manage continuous integration frameworks easily.

Estimate Development Costs: Prevent scope by accurately estimate the cost of new software development and changes to existing systems.

Find and Make Test Data: Reduce the time needed to find and make test data by up to 95%. Link the right data to test cases and expected results.

Link Expected Results: Link expected results to both test cases and data.

Manage Changing Requirements: Significantly reduce the amount of time software testers spend manually repairing existing test cases when a change is made.

Manage Software Outsourcing: Provide a systematic way of enforcing SLAs with out-sourcers.

Maximize Test Coverage: Go beyond combinatorial and pairwise testing and cover 100% of requirements in the smallest number of tests.

Optimize Software Tests: Optimize software tests and clean up your existing test cases, reducing your current test cycles by more than 30%.

Reduce Software Testing Costs: Reduce the cost of testing by de-duplicating your existing test cases and automatically creating missing test cases.

Reverse Engineer Requirements: Define your software requirements using Agile Designer and present them in a visual and active flow chart.

Service Virtualization Testing: Maximise the use of virtualization toolkits by designing the correct request/response sets.

The Role of Feedback in Software Process Assessment

Zeljko Stojanov

University of Novi Sad, Serbia

Dalibor Dobrilovic

University of Novi Sad, Serbia

INTRODUCTION

Processes play a pivotal role in business success of enterprises because effective processes increase their adaptability, efficiency and competitiveness in a global market (Lam, 2014). Lepmets et al. (2012) argued that effective software organizations should align their processes to overall business goals and goals on the operational level, which assumes continuous process assessment and improvement. However, in many cases processes are not well defined, which lead to poor and unsystematic implementation. According to Persse (2006), an organization that uses weak or misdirected processes will have weak products and services, and unpredictable outcomes of business activities. For solving problems with poorly defined and implemented processes, a number of approaches have been proposed and shaped in software process improvement (SPI) frameworks. SPI is a set of activities aimed at achieving better process performance, leading to increased quality of software products and services. However, analysis of reported SPI studies revealed a great variability in success of SPI initiatives (Rainer & Hall, 2003; Montoni & da Rocha, 2014). In addition, Bannerman (2008) argued that implementation of SPI initiatives does not always lead to business benefits.

Process assessment helps software organizations to improve themselves through identification of critical problems and appropriate improvement opportunities. Process assessment is the first phase in SPI cycle, which means that assessment

outcomes serve as input for process improvement action plans. Effective assessment requires commitment and involvement of management at all levels in an organization, as well as involvement of staff who implement processes (Mathiassen et al., 2005; Herranz et al., 2013). According to Baddoo et al. (2000), a variety of experiences and attitudes of different groups of people within an organization positively contribute to realization of process assessment and improvement activities. O'Connor & Basri (2012) stated that people involvement in assessment and improvement activities is necessary because employees are the best source of information for these activities and should implement improvements in practice.

Feedback is one of the most valuable tools for achieving continuous organizational development and improvement (Roebuck, 1996). Feedback is essential in assessing and improving organization performance since it is in the core of problem solving activities and it influences decision making while searching for solutions for identified problems (Greve, 2010). Understanding feedback and its effects requires looking at intrinsic sources of information relevant for increasing knowledge and improving performance of individuals and organizations (Greller & Herold, 1975). Considering people as the main sources of feedback information in an organization positively influences the success of improvement initiatives. However, it is necessary to consider and weight potential sources of information, which is usually the task of people within an organization.

Software process assessment includes feedback as a core activity aimed at feeding data back to relevant individuals or groups within an organization. Feedback is usually a part of a typical sequence of activities, which includes collecting, analyzing and interpreting data. Dyba et al. (2004) proposed a measurement and feedback process in the context of software process assessment with the following steps: planning to use data, collecting data, analyzing data, feeding back data, and follow up. Feedback is also used as a method for learning in software organizations based on previous experiences. Heikkilä (2009) suggested that SPI initiatives should be more concerned with organizational learning and change management, which helps in improving processes gradually and based on lessons learned during SPI implementation. In addition, information included in feedback can be used as a basis for decision making. According to Halloran (1999), software process assessment and improvement facilitate organizational learning if all relevant information and knowledge is communicated to organisation's members.

BACKGROUND

The concept of feedback has been researched and used in many different fields, including education, management, marketing, professional training, human resource development, medicine and engineering. Feedback is a complex concept originated in systems thinking and cybernetics. Due to the specificity of different fields, the term feedback has been used and interpreted in many ways, which causes that there is no universally accepted definition of feedback in theory and practice. Ramaprasad (1983) provided the most general definition of feedback: *Feedback is information about the gap between the actual level and the reference level of a system parameter which is used to alter the gap in some way*. Examinations of this definition in recent literature revealed that it cues directions for learning and performance improvement in organizations (Hattie & Tim-

perley, 2007). The most common usage of the term feedback in the research on humans relates to presentation of information to individuals regarding different aspects of performance, such as behavior or outcomes (Atkins et al., 2002). Assessment of functional mechanisms of feedback provides more useful and practical information for managers and practitioners than assessment of feedback characteristics (Alvero et al., 2001).

Feedback is strategically important and valuable tool for assessing and improving organizations in highly dynamic, uncertain and competitive business environment. Based on the literature review on various types of feedback in organizations, De-Nisi & Kluger (2000) found that studies reported inconsistent results concerning the effectiveness of feedback. For achieving the best positive effects on performance, and preventing possible negative effects, feedback should be carefully planned and designed. Feedback is generally used for evaluating the level of performance and for indicating improvements. In this sense, feedback is used for identifying a gap between what is observed in the practice and what is expected in improved practice. According to Grafton et al. (2010), the main uses of feedback in organizational context are for: (1) promoting organizational learning, (2) analyzing the impact of past decisions, (3) prompt examination of adopted strategies and targets, and (4) identifying needs for corrective actions. Feedback involves people that provide information and people that interpret provided information. Based on the interpretation of feedback information people make decisions in their organizations. Feedback forms loops between actual performance and established goals. According to Hattie & Timperley (2007), the main purpose of feedback is to reduce divergence between current understandings and performances and proposed objectives.

Types of Feedback

Several approaches for defining and observing feedback can be found in literature. Based on

extensive literature review, Gabelica et al. (2012) differentiate two main types of feedback: (1) *performance feedback*, which is concerned with the outputs of observed tasks, and (2) *process feedback*, which is concerned with information how individuals or the teams perform tasks. Performance feedback can be used for performance assessment and improvement. Govaerts et al (2013) described a cyclic process of performance improvement, which includes feedback. In this process, performance is observed and interpreted within assessment process. This assessment is cognitive process during which an assessor provides a judgment about observed performance. These information is available as feedback that should be correctly interpreted in order to enable performance improvement. Performance assessment activities result with feedback in both quantitative and qualitative formats. According to Gabelica et al. (2012), the following three types of process feedback can be distinguished: (1) *task related feedback*, which is related to information how individuals or groups deal in with tasks within a process, (2) *interpersonal feedback*, which is concerned with social conditions under which a group create and share understanding, and (3) *cognitive feedback*, which refers to cognitive interactions between people in a group.

Sengupta & Te'eni (1993) described advantages of using cognitive feedback compared with outcome or performance feedback. Cognitive feedback relates to information that enhances understanding of observed phenomena and quality of decisions in a given context. Cognitive feedback helps in clarifying past decisions and assists in selecting appropriate decision strategies, providing better control of change implementation within a system. This is very important in projects focused on assessing the state of the practice within organizations and planning and implementing proposed improvements. Outcome feedback, in contrast to cognitive feedback, provides only information about the outcomes of observed process, which is generally insufficient for building comprehensive view and understanding, and therefore leads to

less effective decisions and further improvement actions (Sengupta & Abdel-Hamid, 1993).

Feedback Influence on Motivation and Commitment

According to Locke (1996), people need to check their progress while effectively pursuing to proposed goals. Feedback is essential for achieving proposed goals, which means that based on the observed situation it affects goal persistence, disengagement, and goal change (Fishbach & Finkelstein, 2012). In practice, feedback can be positive and negative, but effective assessment and improvement approaches assume proper use of both positive and negative feedback for achieving proposed improvement goals (Fishbach et al., 2010). Positive feedback encourages people to pursue in proposed direction towards proposed goals, increasing their motivation and commitment. At the other hand, negative feedback should be used for identifying weaknesses, the lack of accomplishments and inconsistencies in improvement. In addition, Vallerand & Reid (1988) suggested that positive feedback increases intrinsic motivation, which is largely mediated by the feelings of competence. Negative feedback can also be motivating through self-regulation process, which can lead to making progress towards reducing discovered discrepancy and pursuing toward stated objectives. According to Van-Dijk & Kluger (2004), positive feedback connected with wishes, and negative feedback connected with obligations, increase motivation and commitment.

Feedback and Learning in Organizations

Feedback, as an important and powerful tool for shaping learning in organizations, should be observed from the perspective of individuals, teams and whole organizations (Gabelica et al., 2012). According to Mausolff (2004), organizational processes are the primary source of learning in an organization, while feedback helps in initiat-

ing and sustaining organizational learning. The purpose of feedback in organizational learning is to provide information on actual implementation and performance of processes and activities (identification of strengths and weaknesses), encouraging critical reflections of involved people and guiding further activities related to practice improvement. According to Argote (2013), context based learning in organizations is usually based on direct experiences and depends on several factors, among which feedback takes significant role. London & Smither (2002) presented person–environment interaction view of feedback processes over time, and stated that individual positive attitude towards feedback, feedback orientation and feedback culture in an organization enhance performance improvement programs and learning within an organization. In addition, in the context of performance improvement and organizational learning, feedback is not isolated event but an intrinsic part of comprehensive management process over long period of time. Performance feedback that promotes and facilitates organizational learning should be based on meaningful and sensitive communication between individuals at all organizational levels (Mayfield & Mayfield, 2012). Furthermore, feedback understanding should be enhanced by using consistent feedback terminology and appropriate transmission media.

FEEDBACK IN SOFTWARE PROCESS ASSESSMENT

Performance assessment evolved from a purely measurement based focus to one that recognizes the importance of social and cognitive processes in organizations (Levy & Williams, 2004). Assessment activities are usually implemented within an organization and highly depend on the context that should be completely understood in order to maximize positive outputs. Effective performances of processes require access to relevant information, which comes from many sources like process performers, management, customers and other

stakeholders (Heidrich et al., 2006). Collecting information from various sources ensures comprehensive view of processes and more reliable assessment aligned with organization's business objectives. Better process performances lead to better organizational performances and increases maturity of an organization (McBride, 2010).

Due to the longevity and complexity of assessment and improvement projects, the following questions deserve special attention (Aaen et al., 2001): how participants in these projects know whether they as individuals and as the whole organization achieve proposed goals, how they know if perceived benefits are realized, and how they know what are the next steps based on the observed state of the project? According to Aaen et al. (2001), feedback is essential for effective management in software process assessment and improvement initiatives because it helps in: obtaining visible results in several important points during project implementation, keeping the effort in the focus, and maintaining motivation, commitments, and legitimacy. Further, Arent et al. (2000) suggested that project (with included processes) assessment, with feedback provided to relevant people in an organization, supports commitment, participation and learning in an organization conducting SPI project.

Presenting Results

Information included in feedback provides the basis for decision making related to revealing the next activities in an assessment and improvement project. During process assessment, feedback can be seen as information provided by stakeholders regarding understanding of the current state and results. According to Atkins et al. (2002), feedback format should be carefully designed in order to enable integration of data that have emergent characteristics and to provide a relevant basis for judgments. Majority of assessment processes use information in both quantitative and qualitative formats, enabling more comprehensive understanding and evaluation of

processes' performances (Brutus, 2010). Feedback results should be presented to all relevant people in an organization, ensuring active participation of employees that are the main source of knowledge. Therefore, their involvement in discussion and interpretation of results is invaluable for the success of assessment project.

Dyba et al. (2004) presented an overview of several techniques for presenting and interpreting assessment results, such as mind maps, brainstorming, tables, diagrams and charts. Stojanov et al. (2013) outlined an assessment approach in which feedback includes information presented to company personnel in the form of trends that are based on statistical analysis of data extracted from the maintenance repository in a small software company. Trends were presented with different types of diagrams (lines, charts, pies) and tables. Zarour et al. (2015) identified feedback sessions as one of the best practical techniques in process assessment aimed at presenting assessment results to organization being assessed. Presenting results in an appropriate way enables their validation, as well as continuous improvement of assessment methods.

Directing and Maintaining Motivation and Commitment

Commitment and motivation play the key roles in process assessment projects, because it is recognized that motivated and supported people are willing to reward greater effort in their work. Arent et al. (2000) pointed out the positive effect of active participation and commitment to assessment project activities and overall SPI process, while Mishra & Mishra (2008) underlined positive effects of periodic feedback sessions on the personnel motivation in assessment projects. Active involvement of organization's staff creates opportunities for maintaining motivation and overcoming many difficulties that may occur during assessment. Dyba & Moe (1999) stressed the importance of active cooperation between employees in an organization and researchers,

which ensures maintaining the focus in assessment project, especially in feedback sessions that serve as an effective tool for analyzing the state and directing further activities. Baddoo & Hall (2002) identified feedback as an important motivating factor for developers and senior managers, which can positively affect implementation of SPI projects. In addition, Baddoo & Hall (2003) discovered the lack of management commitment and the lack of feedback to developers as important de-motivating factors in SPI projects. Better understanding of actual processes through staff involvement in assessment activities helps in sharing experiences and defining more realistic improvements goals, which increases their motivation toward improvement project (von Wangenheim et al., 2006).

Learning in Software Process Assessment

Feedback can be used as a method for learning in an organization based on the previous experiences and identified issues in assessment process. One of the most effective ways to provide information to people in an organization is by using feedback meetings, where employees can interpret collected data and results from analysis, as well as pack and store knowledge in a way suitable for reusing in future software projects (Mishra & Mishra, 2008). According to Dyba et al. (2004), conducting several feedback meetings helps in transforming individual experiences to collective knowledge in a company. Cater-Steel et al. (2006) pointed out that feedback can be used for collecting experiences from organizations participating in assessment projects, which can be used for providing recommendations for these organizations in future projects, as well as for improving assessment methods.

Dyba & Moe (1999) presented two industrial case studies in which participants from software organizations were provided opportunities to analyze, interpret and learn from the assessment results. Feedback to participants was provided

during half-day workshops by using several techniques such as gap analysis, scatter plots, bar charts and histograms. Dyba & Moe (1999) stressed the importance of active participation of company employees, which contributes to better implementation of identified improvements and to learning from the experience. Arent et al. (2000) underlined the significance of learning effect in assessment projects that are based on active involvement of people and on-time feedback provided to relevant stakeholders. In such software organizations, learning occurs at different levels - people learn from each other, positive experiences are transferred between different projects, and managers gain a higher level of professionalism and deeper understanding of practice. Kim et al. (2005) presented an experience factory model for accumulating and utilizing process assessment experiences, which is based on database designed for saving all analyzed data in forms of root words. Collected experiences help in decision making and establishing optimal improvement strategies.

SOLUTIONS AND RECOMMENDATIONS

Success of assessment and SPI projects requires a cross-disciplinary group of assessors that includes researchers and key experts available in software organizations (Pettersson et al., 2008; Feliz, 2012), as well as proper use of feedback during all phases in projects (Takeuchi et al., 2014; Stojanov et al., 2015; Zarour et al., 2015). Due to the complexity of software processes that should be comprehensively investigated, assessment should be based on both quantitative and qualitative data collected in the organizational context (Feliz, 2012, Stojanov et al., 2015). Appropriate and timely analysis, combined with feedback, contribute to the efficiency of assessment, facilitate decision making and identification of the most relevant improvements.

FUTURE RESEARCH DIRECTIONS

S

Software process assessment plays important role in investigating and improving everyday practice in software organizations. However, new trends in organizing work of geographically distributed teams, and recognized importance of human factors, introduce new challenges regarding practical issues that should be investigated and understood. Based on these observations, development of new methods and techniques that will facilitate work of distributed teams of assessors can be expected. These new methods and techniques will be based on the contemporary Internet and mobile technologies. In addition, the complexity of human factors' effects on practice will require creation of multidisciplinary teams that will include experts specialized in disciplines such as sociology, psychology and human resource management.

CONCLUSION

It has been recognized by both industry and research communities that assessment of software processes is necessary for maintaining and increasing the quality of products and services. This requires effective plans and guidelines for conducting assessments, resulting with identification and implementation of the most valuable improvements. Effective communication between people involved in assessment projects is a premise for success. This communication includes preparation and implementation of assessment plans, but also requires the use of efficient feedback techniques in all phases of assessment in order to achieve the best outcomes.

REFERENCES

Aaen, I., Arent, J., Mathiassen, L., & Ngwenyama, O. (2001). A Conceptual MAP of Software Process Improvement. *Scandinavian Journal of Information Systems*, 13(1), 79–99.

- Alvero, A. M., Bucklin, B. R., & Austin, J. (2001). An Objective Review of the Effectiveness and Essential Characteristics of Performance Feedback in Organizational Settings (19851998). *Journal of Organizational Behavior Management*, 21(1), 3–29. doi:10.1300/J075v21n01_02
- Arent, J., Iversen, J. H., Andersen, C. V., & Bang, S. (2000). Project Assessments: Supporting Commitment, Participation, and Learning in Software Process Improvement. In *Proceedings of the 33rd Hawaii International Conference on System Sciences (HICSS)*. Maui, Hawaii, USA. doi:10.1109/HICSS.2000.926777
- Argote, L. (2013). *Organizational Learning: Creating, Retaining and Transferring Knowledge* (2nd ed.). New York: Springer US. doi:10.1007/978-1-4614-5251-5
- Atkins, P. W. B., Wood, R. E., & Rutgers, P. J. (2002). The effects of feedback format on dynamic decision making. *Organizational Behavior and Human Decision Processes*, 88(2), 587–604. doi:10.1016/S0749-5978(02)00002-X
- Baddoo, N., & Hall, T. (2002). Motivators of Software Process Improvement: An analysis of practitioners views. *Journal of Systems and Software*, 62(2), 85–96. doi:10.1016/S0164-1212(01)00125-X
- Baddoo, N., & Hall, T. (2003). De-motivators for software process improvement: An analysis of practitioners views. *Journal of Systems and Software*, 66(1), 23–33. doi:10.1016/S0164-1212(02)00060-2
- Baddoo, N., Hall, T., & Wilson, D. (2000). Implementing a people focused SPI programme. *Proceedings of the 11th European Software Control and Metrics Conference and The Third SCOPE Conference on Software Product Quality*, 373–381.
- Bannerman, P. L. (2008). Capturing business benefits from process improvement: Four fallacies and what to do about them. *Proceedings of the 1st International Workshop on Business Impact of Process Improvements, BiPi '08*, pp. 1–8. doi:10.1145/1370837.1370839
- Brutus, S. (2010). Words versus numbers: A theoretical exploration of giving and receiving narrative comments in performance appraisal. *Human Resource Management Review*, 20(2), 144–157. doi:10.1016/j.hrmr.2009.06.003
- Cater-Steel, A., Toleman, M., & Rout, T. (2006). Process improvement for small firms: An evaluation of the RAPID assessment-based method. *Information and Software Technology*, 48(5), 323–334. doi:10.1016/j.infsof.2005.09.012
- DeNisi, A. S., & Kluger, A. N. (2000). Feedback Effectiveness: Can 360-Degree Appraisals Be Improved? *The Academy of Management Executive*, 14(1), 129–139.
- Dyba, T., Dingsoyr, T., & Moe, N. B. (2004). *Process Improvement in Practice - A Handbook for IT Companies*. Norwell, MA: Kluwer Academic Publishers.
- Dyba, T., & Moe, N. B. (1999). *Rethinking the concept of software process assessment. Proceedings of the 6th European Software Process Improvement Conference, EuroSPI'99*.
- Feliz, T. (2012). Lightweight Software Process Assessment and Improvement. In *Proceedings of Thirtieth Annual Pacific Northwest Software Quality Conference* (pp. 405–424). Portland, OR: PNSQC.
- Fishbach, A., Eyal, T., & Finkelstein, S. R. (2010). How Positive and Negative Feedback Motivate Goal Pursuit. *Social and Personality Psychology Compass*, 4(8), 517–530. doi:10.1111/j.1751-9004.2010.00285.x

- Fishbach, A., & Finkelstein, S. R. (2012). How feedback influences persistence, disengagement, and change in goal pursuit. In H. Aarts & A. J. Elliot (Eds.), *Goal-Directed Behavior*. New York, NY: Psychology Press.
- Gabelica, C., Van den Bossche, P., Segers, M., & Gijssels, W. (2012). Feedback, a powerful lever in teams: A review. *Educational Research Review*, 7(2), 123–144. doi:10.1016/j.edurev.2011.11.003
- Govaerts, M. J. B., van de Wiel, M. W. J., & van der Vleuten, C. P. M. (2013). Quality of feedback following performance assessments: Does assessor expertise matter? *European Journal of Training and Development*, 37(1), 105–125. doi:10.1108/03090591311293310
- Grafton, J., Lillis, A. M., & Widener, S. K. (2010). The role of performance measurement and evaluation in building organizational capabilities and performance. *Accounting, Organizations and Society*, 35(7), 689–706. doi:10.1016/j.aos.2010.07.004
- Greller, M. M., & Herold, D. M. (1975). Sources of feedback: A preliminary investigation. *Organizational Behavior and Human Performance*, 13(2), 244–256. doi:10.1016/0030-5073(75)90048-3
- Greve, H. R. (2010). Designing Performance Feedback Systems to Guide Learning and Manage Risk. *Organizational Dynamics*, 39(2), 104–114. doi:10.1016/j.orgdyn.2010.01.004
- Halloran, P. (1999). Organisational learning from the perspective of a software process assessment and improvement program. *Proceedings of the 32nd Hawaii International Conference on System Sciences*. doi:10.1109/HICSS.1999.772898
- Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, 77(1), 81–112. doi:10.3102/003465430298487
- Heidrich, J., Münch, J., Riddle, W., & Rombach, D. (2006). People-oriented Capture, Display, and Use of Process Information. In S. T. Acuña & M. I. Sánchez-Segura (Eds.), *New Trends in Software Process Modeling* (pp. 121-179). Singapore: World Scientific Publishing Company.
- Heikkilä, M. (2009). Learning and Organizational Change in SPI Initiatives. *Proceedings of the 10th International Conference on Product-Focused Software Process Improvement, PROFES 2009*, 216-230. doi:10.1007/978-3-642-02152-7_17
- Herranz, E., Colomo-Palacios, R., & Amescua-Seco, A. (2013). Towards a New Approach to Supporting Top Managers in SPI Organizational Change Management. *Procedia Technology*, 9, 129–138. doi:10.1016/j.protcy.2013.12.014
- Kim, G., Lee, M., Lee, J., & Lee, K. (2005). Design of SPICE experience factory model for accumulation and utilization of process assessment experience. *Proceedings of the third ACIS International Conference on Software Engineering Research, Management and Applications*, 368–374.
- Lam, V. S. (2015). The Basis and Core Knowledge of Business Process Management. In M. Khosrow-Pour (Ed.), *Encyclopedia of Information Science and Technology* (3rd ed.; pp. 5303–5314). Hershey, PA: Information Science Reference. doi:10.4018/978-1-4666-5888-2.ch524
- Lepmets, M., McBride, T., & Ras, E. (2012). Goal alignment in process improvement. *Journal of Systems and Software*, 85(6), 1440–1452. doi:10.1016/j.jss.2012.01.038
- Levy, P. E., & Williams, J. R. (2004). The Social Context of Performance Appraisal: A Review and Framework for the Future. *Journal of Management*, 30(6), 881–905. doi:10.1016/j.jm.2004.06.005

- Locke, E. A. (1996). Motivation through conscious goal setting. *Applied & Preventive Psychology*, 5(2), 117–124. doi:10.1016/S0962-1849(96)80005-9
- London, M., & Smither, J. W. (2002). Feedback orientation, feedback culture, and the longitudinal performance management process. *Human Resource Management Review*, 12(1), 81–100. doi:10.1016/S1053-4822(01)00043-2
- Mathiassen, L., Ngwenyama, O. K., & Aaen, I. (2005). Managing Change in Software Process Improvement. *IEEE Software*, 22(6), 84–91. doi:10.1109/MS.2005.159
- Mausolff, C. (2004). Learning from Feedback in Performance Measurement Systems. *Public Performance & Management Review*, 28(1), 9–29.
- Mayfield, M., & Mayfield, J. (2012). Effective performance feedback for learning in organizations and organizational learning. *Development and Learning in Organizations: An International Journal*, 26(1), 15–18.
- McBride, T. (2010). Organisational theory perspective on process capability measurement scales. *Journal of Software Maintenance and Evolution: Research and Practice*, 22(4), 243–254.
- Mishra, D., & Mishra, A. (2008). A Study of Software Process Improvement in Small and Medium Organizations. In H. Oktaba & M. Piattini (Eds.), *Software Process Improvement for Small and Medium Enterprises: Techniques and Case Studies* (pp. 140–157). Hershey, PA: Information Science Publishing. doi:10.4018/978-1-59904-906-9.ch006
- Montoni, M. A., & da Rocha, A. R. C. (2014). Applying grounded theory to understand software process improvement implementation: A study of Brazilian software organizations. *Innovations in Systems and Software Engineering*, 10(1), 33–40. doi:10.1007/s11334-013-0209-8
- O'Connor, R., & Basri, S. (2012). The effect of team dynamics on software development process improvement. *International Journal of Human Capital and Information Technology Professionals*, 3(3), 13–26. doi:10.4018/jhcitp.2012070102
- Persse, J. R. (2006). *Process Improvement Essentials: CMMI, Six SIGMA, and ISO 9001*. Sebastopol, CA: O'Reilly Media, Inc.
- Pettersson, F., Ivarsson, M., Gorschek, T., & Öhman, P. (2008). A practitioners guide to light weight software process assessment and improvement planning. *Journal of Systems and Software*, 81(6), 972–995. doi:10.1016/j.jss.2007.08.032
- Rainer, A., & Hall, T. (2003). A quantitative and qualitative analysis of factors affecting software processes. *Journal of Systems and Software*, 66(1), 7–21. doi:10.1016/S0164-1212(02)00059-6
- Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science*, 28(1), 4–13. doi:10.1002/bs.3830280103
- Roebuck, C. (1996). Constructive feedback: Key to higher performance and commitment. *Long Range Planning*, 29(3), 328–336. doi:10.1016/0024-6301(96)00028-3
- Sengupta, K., & Abdel-Hamid, T. K. (1993). Alternative Conceptions of Feedback in Dynamic Decision Environments: An Experimental Investigation. *Management Science*, 39(4), 411–428. doi:10.1287/mnsc.39.4.411
- Sengupta, K., & Teeni, D. (1993). Cognitive feedback in GDSS: Improving control and convergence. *Management Information Systems Quarterly*, 17(1), 87–113. doi:10.2307/249511
- Stojanov, Z., Dobrilovic, D., & Stojanov, J. (2013). Analyzing Trends for Maintenance Request Process Assessment: Empirical Investigation in a Very Small Software Company. *Theory and Applications of Mathematics & Computer Science*, 3(2), 59–74.

Stojanov, Z., Stojanov, J., & Dobrilovic, D. (2015). Knowledge Discovery and Systematization through Thematic Analysis in Software Process Assessment Project. *Proceedings of IEEE 13th International Symposium on Intelligent Systems and Informatics*, 25-30. doi:10.1109/SISY.2015.7325405

Takeuchi, M., Kohtake, N., Shirasaka, S., Koishi, Y., & Shioya, K. (2014). Report on an assessment experience based on ISO/IEC 29110. *Journal of Software: Evolution and Process*, 26(3), 306–312.

Vallerand, R. J., & Reid, G. (1988). On the relative effects of positive and negative verbal feedback on males and females intrinsic motivation. *Canadian Journal of Behavioural Science*, 20(3), 239–250. doi:10.1037/h0079930

Van-Dijk, D., & Kluger, A. N. (2004). Feedback Sign Effect on Motivation: Is it Moderated by Regulatory Focus? *Applied Psychology*, 53(1), 113–135. doi:10.1111/j.1464-0597.2004.00163.x

von Wangenheim, C. G., Anacleto, A., & Salviano, C. F. (2006). Helping Small Companies Assess Software Processes. *IEEE Software*, 23(1), 91–98. doi:10.1109/MS.2006.13

Zarour, M., Abran, A., Desharnais, J.-M., & Alarifi, A. (2015, March). An investigation into the best practices for the successful design and implementation of lightweight software process assessment methods: A systematic literature review. *Journal of Systems and Software*, 101, 180–192. doi:10.1016/j.jss.2014.11.041

ADDITIONAL READING

Bourque, P., & Fairley, R. E. (Eds.). (2014). *Guide to the Software Engineering Body of Knowledge, Version 3.0, SWEBOOK*. IEEE Computer Society.

Christensen, M., & Thayer, R. (2002). *The Project Manager's Guide to Software Engineering's Best Practices*. Piscataway, NJ, USA: Wiley-IEEE Computer Society Press.

Florac, W. A., Park, R. E., & Carleton, A. D. (1997). *Practical Software Measurement: Measuring for Process Management and Improvement, Guidebook CMU/SEI-97-HB-003*. Pittsburgh, PA, USA: Software Engineering Institute, Carnegie Mellon University.

Greve, H. (2003). *Organizational learning from performance feedback*. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511615139

ISO. IEC 12207. (2008). Systems and software engineering - Software life cycle processes. Geneva, Switzerland: International Organization for Standardization (ISO).

Jones, C. (2010). *Software Engineering Best Practices: Lessons from Successful Projects in the Top Companies*. New York, NY, USA: McGraw-Hill.

Münch, J., Armbrust, O., Kowalczyk, M., & Soto, M. (2012). *Software Process Definition and Management*. The Fraunhofer IESE Series on Software and Systems Engineering. Springer-Verlag Berlin Heidelberg. doi:10.1007/978-3-642-24291-5

Oktaba, H., & Piattini, M. (2008). *Software Process Improvement for Small and Medium Enterprises: Techniques and Case Studies* (1st ed.). Hershey, PA: Information Science Publishing. doi:10.4018/978-1-59904-906-9

Sommerville, I. (2011). *Software Engineering* (9th ed.). Boston, MA, USA: Addison-Wesley.

West, M. (2013). *Return On Process (ROP): Getting Real Performance Results from Process Improvement*. Boca Raton, FL, US: CRC Press. doi:10.1201/b14053

KEY TERMS AND DEFINITIONS

Cognitive Feedback: Information that enhances understanding of observed phenomena and quality of decisions in a given context.

Commitment: The state of being dedicated and focused to do something.

Feedback: Information or individual's responses on difference between actual level and proposed level of a system parameter.

Learning: Activity or process of acquiring knowledge or skills through studying or practical experience.

Motivation: The state when someone desires or is willing to do something.

Performance Feedback: Information about different aspects of performances of a task, a system or individuals, which can be used as a basis for improvement.

Software Process Assessment: A set of activities planned and conducted with the aim to get insight into the current state of software processes.

Software Process Improvement: A set of activities planned and conducted with the aim to achieve better performances of software processes, which leads to higher quality of delivered software products and services, as well as to better organizational performance.

Software Process: A set of activities related to development, maintenance and management of software systems.

Secure Software Development of Cyber–Physical and IoT Systems

S

Muthu Ramachandran

Leeds Metropolitan University, UK

INTRODUCTION

Cyber-Physical Systems (CPS) and Internet of Things (IoT) is on the rapid increase as the demand for such applications is growing exponentially. There is a very strong reason for connecting three technologies such as CPS, IoT and Cloud as the first two are connected to a cloud for receiving and analysing data. Cloud computing has emerged to provide a more cost effective solution to businesses and services while making use of inexpensive computing solutions which combines pervasive, Internet, and virtualisation technologies. Cloud computing has spread to catch up with another technological evolution as we have witnessed Internet technology, which has revolutionised communication and information superhighway. Cloud computing is emerging rapidly and software as a service paradigm is increasing its demand for more services. However, this new trend needs to be more systematic with respect to developing secure software engineering and its related processes such as requirements, design, development, and test. For example, current challenges that are faced with cyber security are: application security flaws and lessons learned which can all be applied when developing applications for CPS and IoT systems. Similarly, as the demand for cloud services increases and so increased importance sought for security and privacy. Cloud service providers such as Microsoft, Google, Sales force.com, Amazon, GoGrid are able to leverage cloud technology with pay-per-use business model with on-demand elasticity by which resources can be expanded or shortened based on service requirements.

Alur (2015) defines CPS as:

“A CPS system is defined as a system consists of computing devices communicating with one another and interacting with the physical world via sensors and actuators.” Examples of such systems include from smart buildings to medical devices to automobiles.

McEwen and Cassimally (2014) defines IoT as:

“An IoT system consists of any physical objects contains controllers, sensors, and actuators are connected with Internet.” Examples of such system include any devices capable of sending and receiving data through the internet such as internet enabled washing machine, dishwasher, etc.

In other words, IoT can also be defined as the network of physical objects or things that are built or embedded with sensors, actuators, software, and connect via the internet which enables these objects to collect and exchange data. The difference between the CPS and IoT needs to be clarified as the applications being deployed over the years. First of all, let us look at a precursor is known as Embedded systems which have been successfully deployed in wider areas such as aerospace, manufacturing, chemical processes, civil infrastructures, etc. The key difference between the CPS and Embedded system is the inter-connectivity of these networked physical objects, whereas it often not embedded but interact with physical world objects. A wireless sensor networks can be mounted around a river to receive and exchange data amongst them to calculate any abnormal level

DOI: 10.4018/978-1-5225-2255-3.ch655

of river overflow to avoid any natural disasters in the region. Therefore, security of CPS and IoT systems are paramount to our research as well as their data has been secured.

Currently, security related flaws are being found on a daily basis that are fixed by adding security patches. This is simply an unacceptable paradigm for sustainability of cloud computing. Therefore, we need to develop and build cloud services with build-in security of services (SaaS, PaaS, IaaS), data centres, and cloud servers. This article aims to provide a number techniques and methods for developing cloud services systematically with build-in security. It will also cover a range of system security engineering techniques have been adopted as part of a cloud development process. A number of examples of scenarios have chosen from Amazon EC2, to illustrate with, emerging cloud system security engineering principles and paradigm (Ramachandran, 2013 & 2014). This real-world case study have been used to demonstrate the best practices on business process modelling and component based design for developing cloud services with Build Security In (BSI). BSI techniques, strategies, and process presented in this article are general systems security principle and are applicable for both in a cloud environment and traditional environment (non-cloud environment). The significant contribution of this research is to illustrate the application of the extended system security method known as SysSQUARE to elicit security requirements, to identify security threats of data as well as integrating build-in security techniques by modelling and simulating business processes upfront in the systems development life cycle.

BACKGROUND

Legacy applications have complex interconnections and are connectionless. For example, a sales manager needed to access a real-time stock on the mainframe applications when travelling requires migrating to SOA. IoT (Internet of Things) has

emerged to address the need for connectivity and seamless integration with other devices. However, there are potential challenges ahead meeting the growing need for IoT based applications. This includes design and implementation challenges, various applications and connectivities such as smart objects and wireless sensor networks, data gathering, storing and analyzing in a cloud based solution, and IoT Security and Privacy issues. Piayre and Seong (2013) discuss an IoT application for a wireless sensor network which is useful in emergency response systems. In addition, CPS systems have a much bigger impact on connected devices, therefore, it is important to understand the clear distinction between these two systems. Table 1 provides features against CPS and IoT systems which considers computational capacity, processing speed, storage capacity, multiple sensor capacity, multiple communication capabilities, mobility, distribution capability, programming, and architectural model that is suitable and secured.

Table 1. Features of CPS and IoT systems

Features	Cyber Physical Systems	IoT
Significant computational capacity	high	low
Processing speed	high	low
Storage capacity	medium	low
Multiple sensory input/output devices, such as touch screens, cameras, GPS chips, speakers, microphone, light sensors, proximity sensors	√	√
Multi-communication connectedness using Wifi, GPS, 3-5G, Bluetooth, etc	√	√
Mobility	Mobile CPS	Mobile
High Level Programming	Java	Java
Distribution Mechanism (Apps Store, Play Store, etc)	√	√
Architectural Design Model	Layer Model, Event-Driven, Web Services	Event-Driven, Web Services

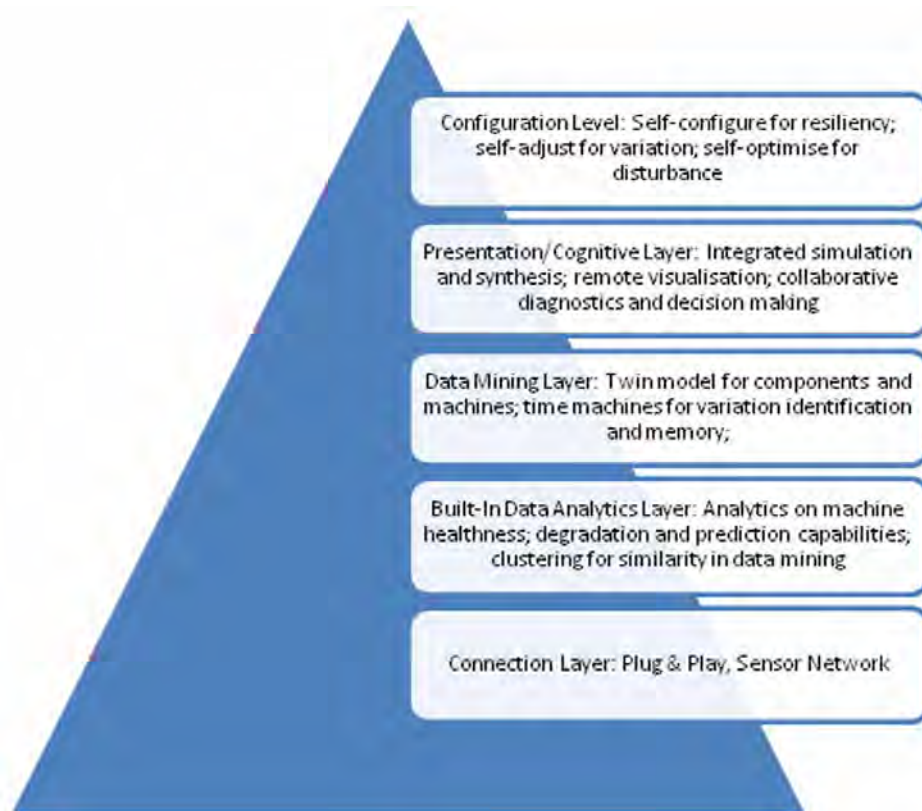
CPS AND IoT ARCHITECTURAL DESIGN CHARACTERISTICS: FUNCTIONS VS. ATTRIBUTES

Design of software architecture for CPS and IoT systems is the key to achieving long term goal of building a sustainable service which is secure and available. There are numerous characteristics that are expected from such devices and their services. Figure 1 shows a five layer model for CPS and IoT architectural layers and their properties. The layers are connection layer 1 (providing plug & play capability as shown at the bottom of the triangle), data analytics layer 2, data mining layer 3, presentation/Cognitive layer 4, and finally a configuration layer 5 (as shown at the top of the triangle providing self configurability and composability of services and devices).

IoT has emerged to address connecting everything possible around us and get real data on behavior otherwise would have not been possible. It all started with RFID technology in the early part of 2000 introduced in the retail market to as product id-tags, in 2010 the technology has been applied to surveillance, healthcare, security, transport, food safety, by 2020 the technology will be used in locating people, products, collecting data on every objects, further on it will be applied in tele-operations and tele-presence, virtual world, touch and feeling, and ability to touch, monitor, and control remote objects, etc.

The following section introduces a systematic approach to developing security specific system requirements for building BSI right from requirements phase of the system engineering life-cycle for IoT and CPS based applications.

Figure 1. CPS and IoT architectural design characteristics



MAIN FOCUS OF THE ARTICLE

Security Requirements Engineering for CPS and IoT Systems

Current examples of UK cyber security attacks on businesses are devastating from Carphone warehouse (a mobile phone sales business), Talktalk (a telephone service provider), and Ashley Madison, a dating website where personal data. Ashford (2009) reports UK business spends 75% of the software development budget on fixing security flaws after delivering the product. This is a huge expenditure and it also creates untrustworthiness amongst customers. Andress (2011) provides an excellent literature survey on the basics of information security techniques. The cyclic security principles known as IAA is not limited pattern of solution for developing secure systems. There are other security concepts that form a pattern of solution known as CIA (Confidentiality, Integrity, and Availability). The CIA considers more towards how well we should design supporting those three characteristics of systems including software and services. In addition, Andress (2011) stated the concept of ParkerianHexad, which consists of six principles CIA (3) + PAU (3) (Possession or control, Authenticity, and Utility).

Traditionally, security has been added and fixed by releasing security patches on a daily basis by major software vendors. This practice needs to change by systematically identifying and incorporating system security right from requirements. This process is known as Building In Security (BSI). Readers are urged to follow the work by McGraw (2004 & 2006) and Ramachandran (2011). This article contributes towards providing a system engineering process for developing and deploying cloud services systematically. It also provides a classification system for cloud security and cloud data security which are useful for developing and maintaining large scale systems with build in security. Finally, data security has been modelled and simulated using the business process methodology. The

results show effectiveness when we develop systems systematically with good systems engineering principles and tools. Therefore, our main recommendation towards building security in (BSI) strategy is to follow one of our guidelines/recommendations:

The aforementioned processes and classification, security principles, and security attributes can be used as a framework for capturing security specific requirements supporting BSI focus by Systems and Software Engineers. In other words, Security requirements = principles of CIA + PAU.

Allen et al. (2008) state that the one of the main goals of Software Security Engineering is to address software security best practices, process, techniques, and tools in every phases and activities of any standard software development life cycle (SDLC). The main goal of building secured software which is defect free and better built with:

- Continue to operate normally in any event of attacks and to tolerate any failure.
- Limiting damages emerging as an outcome of any attacks triggered.
- Build Trust & Resiliency In (BTRI).
- Data and asset protection.

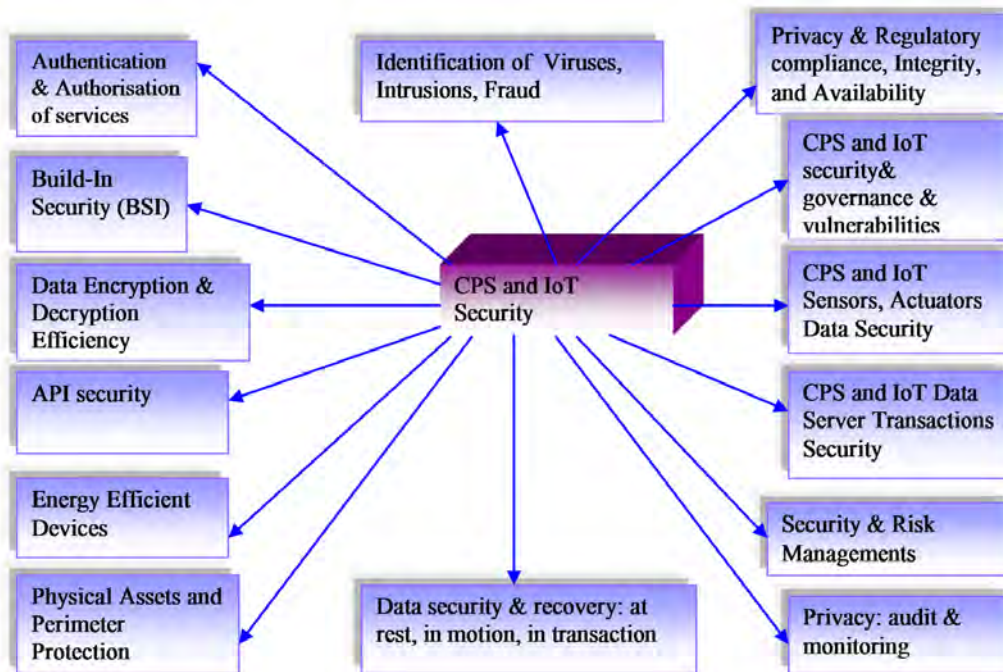
In other words, secured software should operate normally in the event of any attacks. In addition, it involves the process of extracting security requirements from overall system requirements (includes hardware, software, business, marketing, and environmental requirements) and then also further refined and extracted security and software security requirements from software and business requirements. Then the refined software security requirements can be embedded and traced across the software development life cycle (SDLC) phases such as requirements, design, development, and testing. This has not explained well in security related literatures so far. This provides a clear definition of eliciting software security requirements.

Capturing and identifying requirements for security explicitly is one of challenges in software engineering. Often security is considered as one the non-functional requirements which have been considered as constraints identified during and after software has been developed and deployed. However, it has an impact on the functionality of the system. Therefore, we need to be able specify security requirements explicitly throughout the security-specific life-cycle phases as part of achieving BSI (security requirements, design for security, security testing & securability testing). Mead (2005) for the SEI's (software Engineering Institute) has identified a method known as SQUARE (Secure Quality Requirements Engineering) which has been extended SysSQUARE (Systems Engineering SQUARE) towards systems security engineering method. Our extended method consists of nine steps as follow:

- **Agree on Definition:** Which means to define a set of acronyms, definitions, and domain-specific knowledge needs to be agreed by stakeholders. This will help identify and validate security-specific requirements clearly by stakeholders.
- **Identify Security Goals:** Which means to clearly define what is expected of the system with respect to security by business drivers, policies, and procedures.
- **Develop Artefacts:** Which means to develop scenarios, examples, misuse cases, templates for specifications, and forms.
- **Perform Risk Assessments:** Which means to conduct risk analysis for all security goals identified, conduct threat analysis.
- **Select an Elicitation Technique:** Which includes systematic identification and analysis of security requirements from stakeholders in the forms of *interviews, business process modelling and simulations, prototypes, discussion and focus groups*. As part of this phase, one has also to identify level of security, cost-benefits analysis, and organisational culture, structure, and style.
- **Elicit Security Requirements:** Which includes activities such as producing security requirements document based security specific principle structure as part of our goal of developing BSI earlier, risk assessment results, and techniques identifies for analysis such as *business process modelling and simulations, threat modelling, and misuse cases*, etc.
- **Categorise Security Requirements:** Which includes activities such as classifying and categorising security requirements based on company-specific requirements specification templates and to use our recommended security principles as this will help Systems Engineers to apply BSI and track security-specific requirements for validation & verification at all stages of the systems engineering life-cycle.
- **Identify Systems Data Security Requirements:** Which include activities on extracting and carefully identifying data security and relevant sub-systems such as data centres, servers, cloud VM, and software security, SQL security, and other types of security that are relevant to data. This separation of concern allows systems engineers to integrate, track, design, and develop data security as part of enterprise wide systems development.
- **Prioritise Security Requirements:** Which include activities of selecting and prioritising security requirements based on business goals as well as cost-benefit analysis.
- **Inspect Security Requirements:** Which means to conduct requirements validation process using requirements inspection and review meetings.

According to the SysSQUARE model, the first phase starts with identifying security requirements that are achievable and agreed by all stakeholders who are involved in the process. The second step focuses mainly on developing a list all possible security goals as part of the business and functional

Figure 2. CPS and IoT security attributes



goals. Thirdly, to develop a list of artifacts that are needed to achieve those security goals. Fourthly, to conduct a detailed risk assessment for each security goal identified and assessed. Clear identification of requirements of the whole application system and extract security requirements. Interact with stakeholders to clarify security requirements and the technology they want to use, and cost implications. Categorisation and prioritisation of security requirements will help achieve realistic goals against business targets. For example, a network system we need to separate further two categories of security requirements such wired and wireless security systems. The SysSQUARE method elicitation of security requirements has been applied to study the behavior of threat modelling for cloud data security which has been presented in the last section of this article.

Cloud computing has emerged to address the needs of the IT cost-benefit analysis and also a revolution in technology in terms of reduced cost for Internet data and speed. Therefore, the demand for securing our data in the cloud has also increased as a way of building trust for cloud migration

and to benefit business confidence in the cloud technology by cloud providers such as Amazon, Microsoft, Google, etc. Therefore, we also want to make sure our BSI model and strategies are applicable to cloud services as well as traditional systems. Figure 2 shows a model to structure CPS and IoT security attributes to develop and integrate BSI across the system development life-cycle.

The CPS and IoT security attributes shown in Figure 2, are essential and useful to understand non-functional aspects of services development and service provision. These attributes are also useful for building BSI and maintaining security. As shown in the figure, protecting and securing CPS and IoT systems requires energy efficient algorithms, efficient data allocation and retrieval algorithms, and high level of data security using encryption and decryption efficient algorithms. The service availability of these systems is a priority requirement, and often these systems can be developed using readily available APIs such as Google map API, weather forecast APIs, Facebook APIs, Twitter APIs. For example, one could use an IoT to monitor physical premises and send every

data to a Twitter account using those APIs so that relevant people will be alerted quickly.

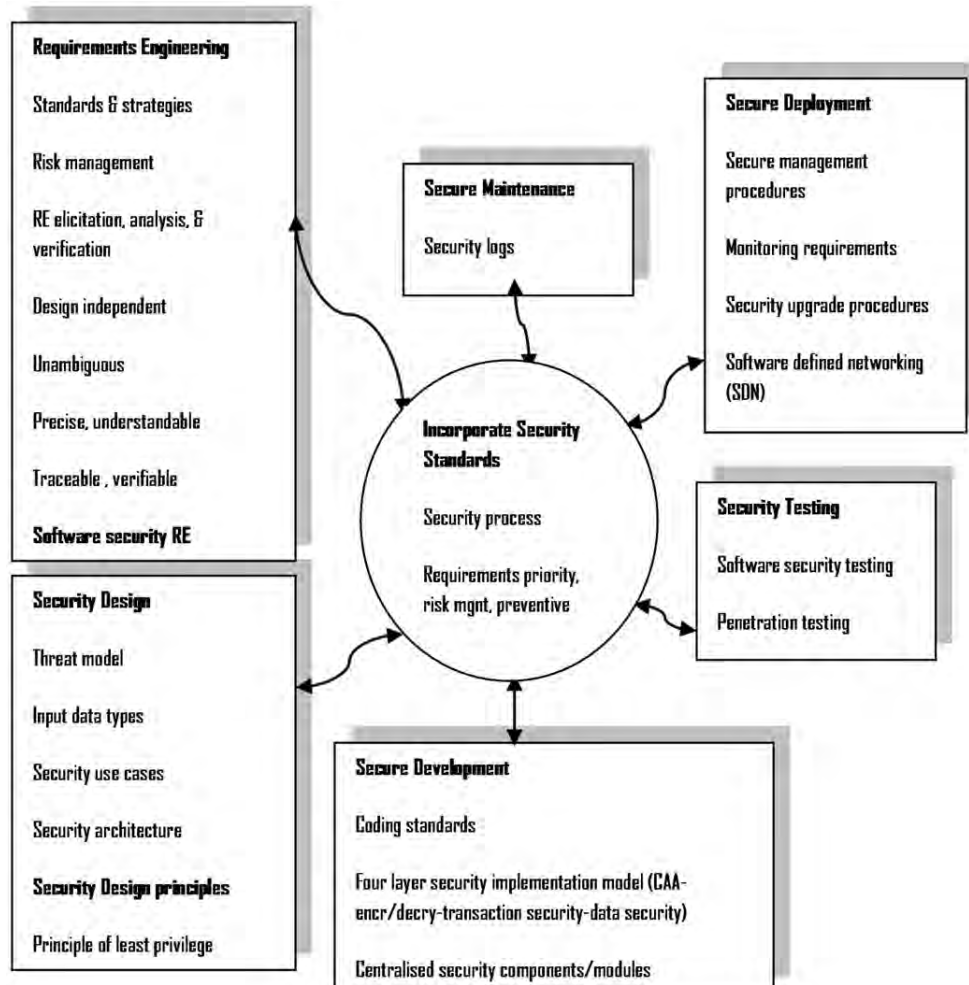
SOLUTIONS AND RECOMMENDATIONS

Integrated Secure Software Development Paradigm

The above discussed drawbacks and requirements for a concise method, lead us to develop a model that integrates various activities of identifying and analysing software security engineering into the software development process, and this new

process and its activities is shown in Figure 3. However, this paper focuses on only software security requirements specific activities. According to this model, SSRE (software security requirements engineering) consists of identifying standards and strategies of the organization with regards to requirements elicitation (including analysis, validation, verification), conducting risk management and mitigation, and identifying software security requirements consists of a further sub-processes of defining security, identifying security strategies, conducting areas and domain scope analysis, business process modeling and simulation, identifying security issues, applying use cases and misuse cases, attack patterns.

Figure 3. Integrated secure software development engineering life cycle (IS-SDLC)



Likewise, this model also provides security-specific processes for identifying security threats during design, development, testing, deployment, and maintenance. There are a numerous number of good design principles that can be found in a vast majority of software design literatures. However, the following is a list of some of the key design principles that are highly relevant to software security design and are part of our IS-SDLC model:

- Principles of least privilege states to allow only a minimal set of rights (privileges) to a subject that requests access to a resource. This helps to avoid intentional or intentional damage that can be caused to a resource in case of an attack.
- Principles of separation of privilege states that a system should not allow access to resources based on a single condition rather it should be based on multiple conditions which has to be abstracted into independent components.
- Design by incorporating known Common Vulnerability Exposures (CVE, <https://cve.mitre.org/>).
- Design for resilience to develop a resilience model which supports system sustainability along side with Building Trust and Security In (BTSI).
- Select software security requirements after performance simulation using BPMN (Business Process Modeling Notation) and is described in detail by Ramachandran.

SSRE activities in our IS-SDLC supports security in a software defined networking (SDN), Cloud computing services (Software as a service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS), Enterprise security includes cloud service providers and service consumers, and design for security principles and techniques. This the unique contribution of this model and for the body of knowledge in software security research.

Service Component Model for CPS and IoT Systems

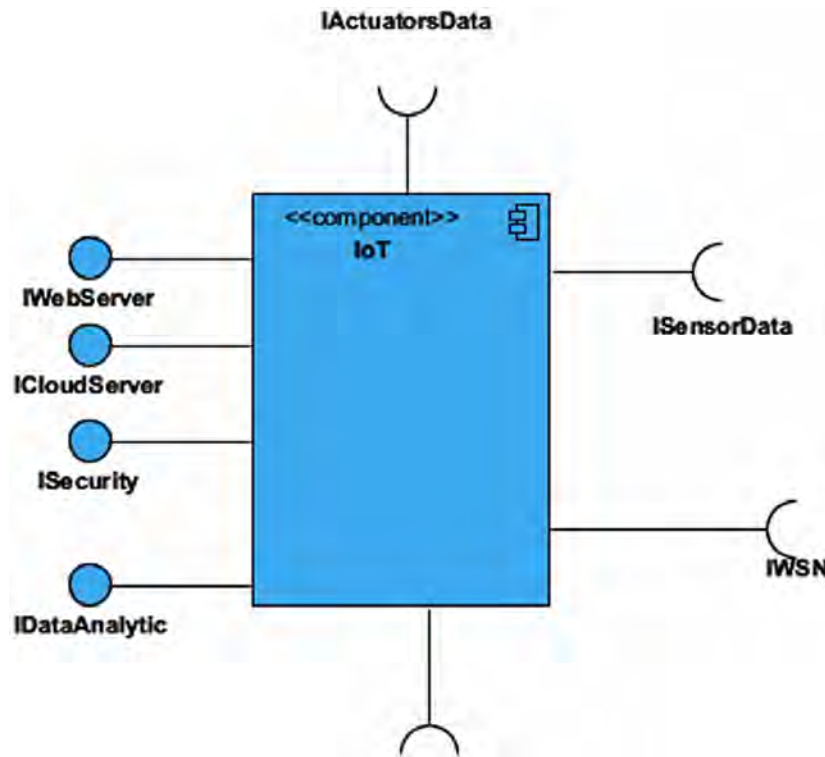
The emergence of IoT main purpose is to be highly interoperable and being able to connect to smart objects, virtual objects, non-deterministic network environment, etc. This can only be achieved with such a high degree of interoperability is by design ToT systems on web services and SOA. Therefore, this paper has developed a service component model based on the IoT requirements now and in the future which is presented in Figure 4.

As shown in the Figure 4, service component model provides two types of interfaces requires services shown as semi-arc for accessing input from wireless sensor (IWSN), sensor data (ISensorData), actuator data (IActuator), and environmental data such as location services (IEnvLocData). There are a number of provider services which this component model offers to connect to other services for composing very complex applications. These are IDataAnalytics, ISecurity (a set of attributes for handing secured services in the event of any intrusion), IWebServer and ICloudServer (connecting to web services and cloud services).

FUTURE RESEARCH DIRECTIONS

This article has presented our approach to developing cloud services systematically with build in security. We have developed a number of security-specific components that can be reused and customised because they are components with message interfaces. We have also developed a number of business processes with simulation to pre-inform us about their performances and security measures that can be taken before service implementation and deployment. As we discussed in this article, to make cloud computing as a new technological business model that is highly successful, profitable, and sustainable, we need to ensure cloud security and privacy can be maintained and trusted. Therefore, most

Figure 4. Service component model for CPS and IoT systems



of the future research will focus mainly on cloud security related issues, in particular, some of more specifically:

- CPS and IoT with Cloud Computing.
- CPS and IoT Development, Tools and Techniques.
- CPS and IoT with Big Data Analytics.
- CPS and IoT with Security issues.
- Control of cloud resources where it is being used and shared and their physical security if this is a hardware resource. In other words, *security concerned with sharing resources and services*.
- Seizure of a company because it has violated the local legislative requirement. Concerns of client's data when it has also been violated. Therefore, forensic investigation of cloud services and cloud data recovery and protection issues will dominate much of the future research.
- Consumer switching to price competition. Storage services provided by one cloud vendor may be incompatible with another vendor's service if user decides to move from one to the other (for example, Microsoft cloud is currently incompatible with Google cloud).
- Security key encryption/decryption keys and related issues. Which is a suitable technique for a specific service request and for a specific customer data? Who should control? Consumers or providers?
- Cloud service development paradigm. What is the suitable development paradigm for this type of business driven delivery model?
- CPS and IoT Service security vs. cloud security vs. data security will dominate most of the future research.
- CPS and IoT Privacy related issues. Who controls, personal and transactional information?

- Audit and monitoring: How do we monitor and audit service provider organisations and how do we provide assurance to relevant stakeholders that privacy requirements are met when their Personally Identifiable Information (PII) is in the cloud?
- Engineering CPS and IoT cloud services. How do we develop, test, and deploy cloud services? Can we continue to follow traditional methods and process?
- Business process modelling integrated with cloud service development will emerge and can address business related issues.
- Integrating data security as part of the system, software, and service engineering processes.

CONCLUSION

CPS, IoT, and Cloud computing have established its businesses and providing services for connected devices. However, this new trend needs to be more systematic with respect to software engineering and its related processes. For example, current challenges that are witnessed today with cyber security and application security flaws are important lessons to be learned. It also has provided best practices that can be adapted. Similarly, as the demand for CPS, IoT, and cloud services increases and so increased importance sought for security and privacy. We can build CPS, IoT, and cloud application security from the start of the cloud service development. CPS, IoT, and Cloud computing are multi-disciplinary that includes social engineering, software engineering, software security engineering, distributed computing, and service engineering. Therefore, a holistic approach is needed to build services. We need to use established architectural and service component model that has been proven over the years in many applications.

REFERENCES

- Allen, J., Barnum, S., Ellison, R. J., McGraw, G., & Mead, N. R. (2008). *Software security engineering: A guide for project managers*. Boston, MA: Addison Wesley.
- Alur, R. (2015). *Principles of cyber-physical systems*. London: MIT Press.
- Andress, J. (2011). *The Basics of Information Security: Understanding the Fundamentals of InfoSec in Theory and Practice*. Syngress (Elsevier).
- Aoyama, M. (2002). *Web Services Engineering: Promises and Challenges*. Presented at International Conference on Software Engineering, ICSE'02, Orlando, FL. doi:10.1145/581339.581425
- Ashford, W. (2009). Retrieved from <http://www.computerweekly.com/Articles/2009/07/14/236875/on-demand-service-aims-to-cut-cost-of-fixing-software-security.htm>
- Bertolino, A. (2006). Audition of web services for testing conformance to open specified protocols. In *Architecting systems with trustworthy components*. Springer. doi:10.1007/11786160_1
- Bias, R. (2009). Cloud Expo Article, Cloud Computing: Understanding Infrastructure as a Service. *Cloud Computing Journal*. Retrieved from <http://cloudcomputing.sys-con.com/node/807481>
- BonitaSoft. (2012). *BOS 5.8. Open source BPMN simulation software*. Retrieved from <http://www.bonitasoft.com/resources/documentation/top-tutorials>
- BPMN2. (2012). *BPMN 2.0 Handbook* (2nd ed.). Future Strategies Inc.
- Caminao Project. (2013). *Caminao's Way: Do systems know how symbolic they are?* Modeling Systems Engineering Project. Retrieved June 21 from http://caminao.wordpress.com/overview/?goback=%2Egde_3731775_member_251475288

- Cause, G. (2012). *Delivering Real Business Value using FDD*. Retrieved April 26 from <http://www.methodsandtools.com/archive/archive.php?id=19>
- Chesbrough, H., & Spohrer, J. (2006). A research manifesto for services science. Special Issue on Services Science. *Communications of the ACM*, 49(7), 35. doi:10.1145/1139922.1139945
- Clarke, R. (2010). *User Requirements for Cloud Computing Architecture*. Presented at the 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing. doi:10.1109/CCGRID.2010.20
- Cobweb. (2009). Retrieved from <http://www.cobweb.com/>
- CSA. (2010). *Cloud Security Alliance. Domain 12: Guidance for Identity & Access Management V2.1*. Retrieved from <https://cloudsecurityalliance.org/guidance/csaguide-dom12-v2.10.pdf>
- Curbera, F. (2007, November). component contracts in service-oriented architectures. *IEEE Computer*, 40(11).
- EC2. (2012). Retrieved from <http://aws.amazon.com/ec2/>
- Erl, T. (2005). *Service-oriented architecture: concepts, technology, and design*. Prentice Hall.
- Farrell, J., & Ferris, C. (2003, June). What are web services?. *Communications of the ACM*, 46(6).
- Gandhi, B. (2011). *Business Process as a Service (BPaaS) delivered from the cloud*. Retrieved from <http://thoughtsoncloud.com/index.php/2011/12/business-process-as-a-service-bpaas-delivered-from-the-cloud/>
- Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. *Future Generation Computer Systems*, 29(7), 1645–1660. doi:10.1016/j.future.2013.01.010
- Helbig, J. (2007, November). Creating business value through flexible IT architecture. *IEEE Computer*, 40(11).
- Iaa, S. (2010). *Cloud computing world forum*. Retrieved April 2nd from <http://www.cloudwf.com/iaas.html>
- IBM. (2010). *Eleven habits for highly successful BPM programs*. IBM Thought Leadership White Paper.
- IThound. (2010). *Video whitepaper*. Retrieved February 3rd from http://images.vnunet.com/video_WP/V4.htm
- Khaled, L. (2010). Deriving architectural design through business goals. *International Journal of Computer Science and Information Security*, 7(3).
- Lakshminarayanan, S. (2010, December). *Interoperable security service standards for web services. IT pro*. IEEE CS Press.
- Linthicum, D. (2009). *Application design guidelines for cloud computing*. InfoWorld. Retrieved from <http://www.infoworld.com/d/cloud-computing/application-design-guidelines-cloud-computing-784?page=0,0>
- Mather, T., Kumaraswamy, S., & Latif, S. (2009). Cloud Security and Privacy: An Enterprise Perspective on Risks and Compliance. Academic Press.
- McEwen, A., & Cassimally, H. (2014). *Designing the Internet of Things*. Wiley.
- McGraw, G. (2004, March-April). Software Security: Building Security. *IEEE Security & Privacy*.
- McGraw, G. (2006). *Software security: building security in*. Addison Wesley.
- Mead, N.R. (2005). *Security Quality Requirements Engineering (SQUARE) Methodology. Technical Report*. CMU/SEI-2005-TR-009. Retrieved from <http://www.sei.cmu.edu/library/abstracts/reports/05tr009.cfm>

- Nano, O., & Zisman, A. (2007, November-December). *Realizing service-centric software systems*. *IEEE Software*.
- Naone, E. (2007). *Computer in the cloud, technology review*. Retrieved from <http://www.technologyreview.com/Infotech/19397/?a=f>
- NIST. (2009). Retrieved March 22nd from <http://csrc.nist.gov/groups/SNS/cloud-computing/index.html>
- Oh, S. H. (2011). A Reusability Evaluation Suite for Cloud Services. In *Proceedings of the Eighth IEEE International Conference on e-Business Engineering*. IEEE CS Press. doi:10.1109/ICEBE.2011.27
- Oracle. (2012). *Data Security Challenges. Oracle9i security overview release number 2(9.2)*. Retrieved November 4th from http://docs.oracle.com/cd/B10501_01/network.920/a96582/overview.htm
- OVF. (2010). *Open Virtualization Format (OVF)*. Distributed Management Task Force. Retrieved from http://dmtf.org/sites/default/files/standards/documents/DSP0243_1.1.0.pdf
- Paa, S. (2010). *Types of PaaS solutions*. Retrieved April 22nd from <http://www.salesforce.com/uk/paas/paas-solutions/>
- Papazoglou, P. M. (2007). Service-oriented computing: state of the art and research challenges. *IEEE Computer*, 40(11).
- Piyare, R., & Seong, R. L. (2013). Towards internet of things (iots): Integration of wireless sensor network to cloud services for data collection and sharing. *International Journal of Computer Networks & Communications*, 5(5).
- Popović, K., & Hocenski, Z. (2010). Cloud computing security issues and challenges. Presented at MIPRO 2010. Croatia: Opatija.
- Ramachandran, M. (2008). *Software components: guidelines and applications*. Nova Publishers.
- Ramachandran, M. (2011). Software components for cloud computing architectures and applications. In Z. Mahmood & R. Hill (Eds.), *Cloud Computing for Enterprise Architectures*. Springer. doi:10.1007/978-1-4471-2236-4_5
- Ramachandran, M. (2011). *Software Security Engineering: Design and Applications*. New York: Nova Science Publishers.
- Ramachandran, M. (2012). Service Component Architecture for Building Cloud Services. *Service Technology Magazine*. Retrieved from <http://www.servicetechmag.com/I65/0812-4>
- Ramachandran, M. (2013). Systems Engineering Processes for the Development and Deployment of Secure Cloud Applications. In *Encyclopedia of Information Science and Technology* (3rd ed.). Academic Press.
- SaaS. (2009). Retrieved from <http://www.saas.co.uk/>
- Science Group. (2006). *2020 Science Group: Toward 2020 science*. Tech. report. Microsoft. Retrieved from http://research.microsoft.com/towards2020science/downloads/T2020S_Report.pdf
- Serugendo, G. (2004). *Self-organisation: paradigms and applications, engineering self-organising systems: Nature-Inspired approaches to software engineering*. *Lecture Notes in CS* (Vol. 2977). Springer.
- Sindre, G., & Opdahl, A. L. (2005). Eliciting security requirements with misuse cases. *Requirements Engineering*, 10(1), 34–44. doi:10.1007/s00766-004-0194-4
- Srinivasan, K. M. (2012). *State-of-the-art Cloud Computing Security Taxonomies: A classification of security challenges in the present cloud computing environment*. Presented at International Conference on Advances in Computing, Communications and Informatics, ICACCI '12, Chennai, India. doi:10.1145/2345396.2345474

Taiyuan, S. (2009). *A Flexible Business Process Customization Framework for SaaS*. Presented at WASE International Conference on Information Engineering.

Tondel, I. A. (2008, January-February). *Security requirements for rest of us: a survey*. *IEEE Software*.

Tyagi, S. (2006). *RESTful web services*. Retrieved from <http://www.oracle.com/technetwork/articles/javase/index-137171.html>

Venkataraman, T. (2010). *A Model of Cloud Based Application Environment*. Cloud Forum.

Verizon. (2010). Retrieved October 20th from <http://www.zdnet.co.uk/news/cloud/2010/10/08/the-cloud-lessons-from-history-40090471/>

Vouk, M. A. (2008). Cloud computing – issues, research and implementations. *Journal of Computing and Information Technology*, 16.

Wang, L., & Laszewski, G. (2008). *Scientific Cloud Computing: Early Definition and Experience*. Retrieved from <http://cyberaide.googlecode.com/svn/trunk/papers/08-cloud/vonLaszewski-08-cloud.pdf>

Weiss, A. (2007, December). *Computing in the clouds*. Academic Press.

Wilson, C., & Josephson, A. (2007). Microsoft Office as a Platform for Software + Services. *The Architecture Journal*, 13. Retrieved April 20th from www.architecturejournal.net

Yang, J. (2003). Web service componentisation. *Communications of the ACM*, 46(10), 35. doi:10.1145/944217.944235

Zhang, L.-J., & Zhou, Q. (2009). *CCOA: Cloud Computing Open Architecture*. Presented at IEEE International Conference on Web Services.

ADDITIONAL READING

ACMQ. (2009). Cloud Computing: An Overview, distributed computing. ACM Queue; Tomorrow's Computing Today.

Buyya, R., Broberg, J., & Goscinski, A. (2011). *Cloud Computing Principles and Paradigms*. NJ, USA: Wiley; doi:10.1002/9780470940105

Mather, T., Kumaraswamy, S., & Latif, S. (2009). *Cloud Security and Privacy: An Enterprise Perspective on Risks and Compliance*. CA, USA: O'Reilly.

OVF. (2010) Open Virtualization Format (OVF), Distributed Management Task Force, http://dmf.org/sites/default/files/standards/documents/DSP0243_1.1.0.pdf, version 1.1.0

Paa, S. 2010. Types of PaaS solutions <http://www.salesforce.com/uk/paas/paas-solutions/>

Papazoglou, P. M. et.al (2007) Service-oriented computing: state of the art and research challenges, Special Issue on Service-oriented Computing, IEEE Computer, V.40, No.11, November 2007.

Popović, K., & Hocenski, Z. (2010) Cloud computing security issues and challenges, MIPRO 2010, May 24-28, Opatija, Croatia

Ramachandran, M. (2011) Software components for cloud computing architectures and applications, Springer, Mahmood, Z and Hill, R (eds.). www.springer.com/computer/communication+networks/book/978-1-4471-2235-7

Ramachandran, M. (2011). *Software Security Engineering: Design and Applications*. New York, USA: Nova Science Publishers; https://www.novapublishers.com/catalog/product_info.php?products_id=26331

Sosinsky, B. (2011). *Cloud Computing Bible*. Wiley.

KEY TERMS AND DEFINITIONS

Build Security In (BSI): The process of identifying service security requirements right from beginning of the service identification to the complete life cycle.

Business Process as a Service (BPaaS): The set of process related to managing process related activities of a service business.

Cloud Data Security (CDS): Security of maintaining and preserving client's data that are kept in the cloud.

Service Reuse: The process of reusing services when composing new services.

Software Security Engineering (SSE): The new discipline of applying engineering principles to develop security requirements to engineer software applications including cloud services such as SaaS is essentially a software application which is delivered as a service.

Software Literacy

S**Elaine Khoo***University of Waikato, New Zealand***Craig Hight***University of Newcastle, Australia*

INTRODUCTION

Software mediates almost every aspect of everyday life. Nearly all of human professional and personal activities are embedded and shaped within systems and interactions that involve software at some level from the seductive ecosystems on devices such as iPhones or iPads through to functional interfaces of automatic teller machines (ATMs), to the tools underlying everyday practices such as word processing and email, through to the sophisticated professional editors for multimedia design. None of these tools are 'neutral'. They derive from social and cultural assumptions about their use and all have particular values embedded in their interfaces and hierarchy of affordances. These embody conceptual frameworks in which particular ways of acting or thinking are more possible and imaginable than others. Despite their ubiquity within contemporary society, software itself is only now emerging as a field of study in its own right. As championed by the proponents of Software Studies this is a vital but neglected area of cultural production intersecting with and potentially shaping all other spheres of cultural, economic and political activity (Fuller, 2008; Livingstone et al. 2014; Manovich, 2008). This article proposes the notion of 'software literacy' as one way to understand the skills and understandings needed for people to be critical and creative users of software (as application, platform and infrastructure) in today's software saturated culture. This contribution argues for the relevancy of software literacy as deeply intertwined with people's engagement with software and how it influences

the way people come to understand, represent, generate and critique knowledge. The discussion below begins by overviewing the background and need for a focus on software studies before proceeding to define, introduce and elaborate on a framework for software literacy. An outline of the framework is then grounded in and exemplified through a case study located within a university teaching and learning context with implications for further thinking and research in the field.

BACKGROUND

The proliferation of digital and networked technologies is an expanding and accelerating feature of modern societies and can be predicted to continue to rise and impact on almost every sphere of human living. Most people develop some level of proficiency with everyday software packages informally through their daily use and incremental engagement over time (Bulfin & Koutsogiannis, 2012; Hague & Logan, 2009). Informal learning practices have been shown to increase students' sense of agency and consequently to have the potential to make learning a richer and more fulfilling experience (Furlong & Davies, 2012). Commenting on the trend of digital penetration, numerous authors have further argued that ubiquitous access to digital technologies has shaped a new internet-centred generation of 'digital natives' (Oblinger, 2003) with the corresponding assumption that access to digital tools has, on its own, facilitated the development of new learning skill sets (Tapscott, 2009). Terms such as the

DOI: 10.4018/978-1-5225-2255-3.ch656

‘digital generation’, ‘millennials’, ‘Net Generation’ (Tapscott, 1999), ‘digital natives’ (Prensky, 2001), ‘Google generation’, ‘Generation Y’ and so forth have derived from a host of assumption about the distinctive skill set of generations immersed within digital technologies. Such labels aim to characterise an emerging class of learners accustomed to engaging with software and technologies such that they can effortlessly adopt new technologies, operate at ‘twitch speed’, are able to multitask, imagine, and visualize while communicating in multiple modalities and consequently possess higher technical skills compared to previous generations (Prensky, 2001). The term ‘digital natives’ itself assumes a generational change in digital literacies fed particularly by informal learning, opinions which are closely informed by emancipatory rhetoric surrounding the digital. Consequently in part the term articulates anxieties amongst educational institutions and practitioners that they are falling behind the literacies students will bring with them to learning contexts.

Educators thus are often encouraged to operate under the assumptions that students already possess the necessary computing skills and conceptual frameworks to learn with and through generic software packages, and tend to neglect the role which the affordances of software themselves play in shaping how students ‘perform’ the software (Adams, 2006). Recent research indicates that such assumptions about students’ digital proficiencies are unfounded and that digital inequalities and marginalization persist around students’ access to, and use of, information and knowledge (Bennett, Maton, & Kervin, 2008; Jones & Czerniewicz, 2010; Kennedy et al., 2008). There is evidence students may not be aware of how to apply software embedded technologies effectively to enhance their learning (Khoo, Johnson, & Zahra, 2012; Valtonen, et al., 2011). In other words, digital inequality may not be defined specifically to the issue of *physical* access to digital tools (software and hardware) but may be intimately shaped by each learner’s ability to learn and apply (software-based) tools in ways

which similarly constrain learning and access to knowledge. A student cohort where everyone has access to digital technologies does not necessarily mean that they have all naturally acquired the literacies to usefully apply and critically engage with those technologies. Inequalities also arise due to differences in the social and cultural support for digital use; the expectation of teaching and learning institutions should perhaps be that individuals may be more or less able to exploit and critique the affordances of software and hardware (Selwyn & Facer, 2007).

For example, common affordances such as copy, cut, and paste are assumed to be naturalized and embedded across different software applications and able to be picked up with ease by this current millennial generation for work and leisure pursuits. However these affordances are also poorly understood as tools contextualized within larger hierarchies of affordances and interfaces that shape both their use and more broadly people’s engagement with knowledge, culture and society in the 21st century (Livingstone, Wijnen, Papaioannou, Costa, & Grandio, 2014).

There is a need to parse the distinctions between distinct media, information and specifically digital-centered literacies relevant to current digital and networked technologies and to examine the nature of student understanding and decision making around which tools might best serve their learning purposes. Labels such as ‘digital natives’ tend to conflate a basic skill in operating new technologies with broader forms of understanding and the ability to critique aspects of technology-based cultures. It is important to challenge these stereotypes and assumed competencies in order to examine the range of skills and other literacies that today’s students bring (and do not bring) to their tertiary learning. Students may not be aware of the full implications of the affordances and constraints offered through particular software; students may be as likely to be ‘captured’ by specific applications as ‘empowered’ by them. The design of software necessarily enables some practices even as it constrains and marginalises

others. An awareness of the nature, rationale and implications of different software becomes more vital in situations where platforms and applications are assumed by their users to be ‘neutral’ tools. Many studies have been conducted on information literacy and on ways of mastering software (Underwood, 2009), but the role of software itself tends to be taken for granted and is not questioned. No studies to date have raised the role of student understanding of how software and its affordances influence knowledge representation, generation, and critique.

Educators know very little about how students develop the skills and expertise needed to attend to the features and use of software (as application, platform, and infrastructure) in the service of completing basic and complex learning tasks. There is evidence that students adopt a range of informal approaches to meet their learning needs due to the proliferation of digital and networked tools (Peeters et al., 2014). Research also indicates that students’ formal software learning backgrounds are diverse (Khoo, Johnson, Torrens, & Fulton, 2011). As with other tools employed during learning, student knowledge and use of software and hardware is highly specific to their formal and informal educational, social, and cultural contexts for learning and use (Jones, Ramanau, Cross, & Healing, 2010; Valtonen, Dillon, Hacklin, & Väisänen, 2010). There is a need for further studies to raise lecturer awareness of the role that software plays in shaping and framing what and how they teach and its effects on student learning. In particular, educators can benefit from ideas and guides that help them to interrogate and reflect on their practice to facilitate more effective pedagogies for teaching and learning with, through and about software (including raising their students’ awareness to similar issues).

SOFTWARE LITERACY

Software studies, a comparatively new field of enquiry, proposes that ‘software’, which encom-

passes many forms of computer programming, is the dominant cultural technology of our time (Fuller, 2008; Johnson, 1997; Manovich, 2013). Manovich argues that we need to ‘think of software as *a layer that permeates all areas of contemporary societies* (emphasis in original, 2013, p. 15), a feature of all cotemporary social and cultural practices which has been comparatively neglected as an object of investigation. In an era in which our daily work routines may revolve around the use of a set of creative and communicative applications, social, cultural and political communication is embedded within software platforms such as Facebook, YouTube and Twitter, and a plethora of software applications run in the background of all major institutions, ‘software’ as a collective needs to become the target of greater investigation. A core premise of software studies is the need to recognise the role of software in shaping the nature of our institutions and our everyday lives; for software users (i.e. the vast majority in modern societies) to develop a more critical awareness of how software operates to both ‘empower and discipline’ people (Kitchin & Dodge, 2011, p. 10-11). For instance, Adams (2007) argues for the development of more informed epistemologies of practice and advocates for software design principles that are more sensitive to pedagogical practices. Others such as van Leeuwen and O’Halloran (2011) underscore the need for software literacy modelling and use by lecturers and students. Being more critically aware of these aspects would allow educators and students more informed use of software (Vallance & Towndrow, 2007). This awareness can lead to opportunities for more equitable exploration and dialogue concerning the social, cultural and technological forces that shape the way in which knowledge within a discipline is developed, shared and communicated, and the implications of these aspects for teaching and learning practices.

In order to extend this line of research the term ‘software literacy’ is proposed and defined as the expertise in understanding, applying, problem solving and critiquing software in pursuit of particular learning and professional goals (Hight,

Khoo, Cowie, & Torrens, 2014). This notion is grounded in a practice-based schema which aims to identify what distinguishes a novice user from an expert user through proposing a progressive transition from novice to expert-like understanding, capabilities and qualities (Jones, 2008; Livingstone, et al. 2014). Three progressive tiers of development towards software literacy exist. They are:

1. A basic skill level where a learner can use a particular software (typically using default affordances),
2. An ability to independently troubleshoot and problem solve issues faced when using the full range of affordances constructed by the software, and finally,
3. The ability to critique the software, including being able to apply such critique to a range of software designed for a similar purpose and to further use these understandings for new software learning.

A key threshold in tier 3 is the point where students are able to transfer a critical understanding of software affordances and use to new software and to familiar software in new contexts and situations. They can be assumed to have the sophisticated understanding needed to then consider how software (at various levels of application, platform and infrastructure) shapes the particular kinds of knowledge and actions that are readily available while at the same time constraining other forms of knowledge and action. At this level, students (and by extension, citizens) are assumed to be capable of recognizing more broadly the role which software plays in various levels within contemporary life, and to more deliberately choose platforms and applications based on a thorough understanding of the conceptual and practical implications of their choices. This third tier accentuates an ability to critique as essential; an attribute also recognised by other researchers as a key characterisation of 21st century learners and working professionals to participate productively in the political, leisure

and social spheres of everyday life (Gilbert 2005; Jenkins et al. 2009).

Developing one's software literacy relies on a combination of general competency with software and technologies, together with the ability to undertake more independent (especially informal) learning of discipline-specific knowledge as and when required. This conceptual framework is a response to current digital literacy and information literacy frameworks and related terms which do not go far enough to identify the implications of the choice of software applications and platforms on what can be achieved (Hegarty et al., 2010; Livingstone et al., 2013). The software literacy framework outlined here is intended to complement existing research on digital and information literacy; to more properly acknowledge the emergence of software as a cultural artefact possessing agency within social, cultural, economic and political practices (Fuller, 2008; Johnson, 1997; Manovich, 2013).

SOLUTIONS AND RECOMMENDATIONS

The potential of the three-tier software literacy framework was evidenced through a case study of university context teaching and learning environment. The aim was to explore, examine, and theorize on how the notion of software literacy is understood, developed, and applied in university teaching and learning contexts, and the extent to which this understanding is useful when translated into new contexts of learning with and through software (Khoo, Hight, Torrens, & Cowie, 2013). To address this broader aim, the study, confined to the disciplines of media studies and engineering, sought to explore the nature of informal and formal learning of software applications as an entry point in investigating the nature of software literacies fostered within distinct tertiary student cohorts. The study initially focused on student learning of PowerPoint as a generic form of software application, typically learned partially or wholly through

informal learning, before progressing to examining student learning of disciplinary-specific software learning in media studies (e.g. Adobe Photoshop, Final Cut Pro), and engineering (SolidWorks). Within both these disciplines, PowerPoint was a common feature of lecture sessions, but not specifically examined as software. While engineering students were required to learn SolidWorks (a 3-D design application) as part of disciplinary knowledge, the media studies students encountered software as part of production electives within a broader degree programme and were encouraged to use a range of applications as needed for creative projects. Each of these disciplines required students to apply professional-level creative and editing applications toward learning tasks, and provided detailed instruction in software which was both comparatively complex and relatively unknown to incoming student cohorts. Each disciplinary context also provided distinct sets of knowledge about software provided to students, contrasting experience with a range of alternative applications to those taught in workshops, but nevertheless shared expectations of the need for students to employ informal learning to complete their understanding of specific pieces of software. The overall objective was to understand the impact of more direct instruction (of sophisticated software applications) on student software literacy development and disciplinary understanding.

The findings of this study support the existence of the three-tier software literacy framework but indicate there is some flexibility in student development and movement between the tiers. It was demonstrated that student ability to achieve the higher tiers of software literacy did not preclude them from needing to revisit earlier levels in contexts where they encounter new but similar software. This suggests lecturers should not assume student competency with software across contexts (e.g., informal to formal, from campus to workplace settings). The framework provides a key conceptual tool for practitioners in terms of understanding the role of troubleshooting, for

example, as an important development stage in learning with and through software.

There was evidence of students having arrived at tier 3 (critical literacy) as well as those not moving beyond tier 2 (troubleshooting). For example, student (mis)assumption that the course lecture notes presented in PowerPoint's bullet point format in and of themselves constitute an adequate elaboration of pertinent disciplinary knowledge. The affordances and constraints that PowerPoint and discipline-based software offer in shaping disciplinary knowledge needs further consideration by students and by lecturers. Lecturers therefore need to be aware of the implications of their choice and modelling of a software application or platform. Findings suggest they would be wise to introduce and help students to develop critical awareness of how such software can inform and shape their own understanding of disciplinary knowledge and practice.

These findings were manifested in distinct but closely aligned patterns within each disciplinary context. However, there was also a difference in the way each discipline prioritized software literacy development as part of their student learning pathway and graduate profiles. For example, in the media studies case, critical thinking as emphasized in the third tier is seen as a core aspect of creative disciplinary knowledge. Being able to judge the creative capacities of competing software has implications for the nature of the practice which is being developed, the form and eventual conceptual complexity of media products themselves, as well as more practical considerations such as budget. In contrast, in the engineering case, students are expected to develop digital and software literacy skills as part of their professional learning. For graduates, the quality of engineering design produced and a design's efficiency and effectiveness are prioritized over the kinds of creative reflexivity observed in the media studies case study. Such disciplinary differences reveal a need for each discipline to continually reflect upon how discipline-specific software teaching and learn-

ing is positioned in relation to graduate profiles and the breadth and depth of literacies required to engage successfully with software.

Another clear finding reiterates that multiple learning pathways exist for exploring the affordances of any particular software, both formal and informal. Students tend to prefer informal strategies as a supplementary to and at times above formal strategies to learn discipline-based software (these included YouTube video tutorials, along with peer networks). Lecturers could usefully be informed by and take advantage of students' informal repertoire of learning strategies and networks including their accessing of (web-based) resources and pooling of peer expertise. Lecturers drawing from students' already established informal learning strategies would recognize the relevance of the social and cultural context in shaping effective technology and software engagement.

The findings importantly challenge current notions and assumptions of the digitally literate generation. When learning to use PowerPoint most students perceived themselves to be competent and confident early adopters of technologies and could identify the general affordances of generic software (PowerPoint) to the extent of critiquing the design and performance of PowerPoint practice in their lectures. They were able to recognize discipline-specific affordances among the applications formally taught and acknowledged these to be central in their engagement with disciplinary knowledge. However, very few were able to critique how the design of the *software* itself might shape their disciplinary knowledge. Very few students demonstrated critical thinking about the nature and role of the software they were using and most were not able to describe or discuss applications beyond those used in their learning. That is, there was very little evidence of tier 3 software literacy. Students' superficial critique of software challenges current assumptions that today's students as 'digital natives' have developed critical awareness simply through familiarity with and regular use of a variety of everyday applications and platforms. To reiterate, lecturers need

to explicitly teach and model software critique if they wish to foster this capacity and/or make this possibility known to students.

The diversity of student cohorts and the range of understandings and the variation in familiarity, skills and experience students bring with them to the formal software learning context constitute a further challenge for teaching of, through and about software. Some students may already have a critical orientation towards software (or, more frequently, toward specific applications). And there is demonstrably an advantage in terms of more advanced software learning for those who have prior experience with other software with a similar conceptual framework. In response to this diversity lecturers could usefully direct time and attention to formatively assessing students' initial software literacy and adapting teaching activities in light of this. Most crucially, however, the overall findings indicate there is no single best approach (one size fits all) to teaching discipline-specific software. Lecturers adopting a range of teaching approaches (formal and informal) and being flexible to address diverse learning needs represents a crucial part of supporting student learning.

FUTURE RESEARCH DIRECTIONS

Software is a site of cultural production and an actor in social, economic, political practices; it needs to be analyzed in terms of its embedded values and effects on such practices. The brief summary of case study findings outlined above is limited due to the situated nature of the investigation into discipline-specific software teaching and learning. The participants in the study represent a convenience purposive sample of lecturers and students from one educational institutional setting. They were from two distinctly different disciplinary contexts, with distinct and different disciplinary foci and expectations of software teaching and learning. Lecturers were also careful to point out the study only focused on some courses within a program and all universities have different inter-

pretations of how software teaching and learning can be enacted. Many educational practices are now intimately associated with software-based tools and applications, suggesting an urgent need for research which explores the implications for such practices and the manner in which they operate and are embedded within specific institutional, disciplinary and specific learning contexts. The software literacy framework outlined above offers a basis for addressing the social and cultural dimensions of software-centered learning, including the interactions between formal tertiary environments and informal learning networks.

CONCLUSION

Software literacy is an essential part of learning and living in the 21st century; something which transcends the use of any particular tool and any particular educational, social and cultural context. Software is an increasingly central part of the palette of understandings and skills which comprise the broadening umbrella of digital literacy. As a cultural artefact, software plays a role in reproducing, reinforcing, and augmenting existing cultural practices, as well as generating new practices. Software such as Google, the iOS software in iPhones and iPads, and Microsoft Office software packages are common examples reflecting the extent to which software has become embedded in everyday personal and professional pursuits. The reliance locally and internationally on algorithmically-centered big data for policy and funding decisions is another example of how software tools can influence the way we live. It is therefore desirable and advantageous that graduates (and other learners) have a critical understanding of software to make more informed choices about their use, can transfer this critical understanding to software they have yet to encounter, and understand that all software has nuanced affordances and limitations. This has implications for education providers, especially universities, in their role in fostering critical thinking and serving

as critic and conscience of society. It is crucial to ensure all students and lecturers are supported in teaching and learning processes when these are mediated through and focused on software. As society moves to exploit the potential of e-learning platforms, and make use of social media and cloud-based and mobile applications, there is a need for further detailed empirical investigation of software literacy. In tertiary settings, this is needed to ensure equitable and critical learning with, through and about software.

ACKNOWLEDGMENT

The authors gratefully acknowledge funding support from the Teaching and Learning Research Initiative, New Zealand Council for Educational Research, Wellington, New Zealand.

REFERENCES

- Adams, C. (2006). PowerPoint, habits of mind, and classroom culture. *Journal of Curriculum Studies*, 38(4), 389–411. doi:10.1080/00220270600579141
- Adams, C. (2007). On the informed use of PowerPoint: Rejoining Vallance and Towns. *Journal of Curriculum Studies*, 39(2), 229–233. doi:10.1080/00220270601175246
- Bennett, S., Maton, K., & Kervin, L. (2008). The digital natives debate: A critical review of the evidence. *British Journal of Educational Technology*, 39(5), 775–786. doi:10.1111/j.1467-8535.2007.00793.x
- Bulfin, S., & Koutsogiannis, D. (2012). New literacies as multiply placed practices: Expanding perspectives on young peoples literacies across home and school. *Language and Education*, 26(4), 331–346. doi:10.1080/09500782.2012.691515
- Fuller, M. (2008). *Software studies: A lexicon*. Cambridge, MA: The MIT Press. doi:10.7551/mitpress/9780262062749.001.0001

- Furlong, J., & Davies, C. (2012). Young people, new technologies and learning at home: Taking context seriously. *Oxford Review of Education*, 38(1), 45–62. doi:10.1080/03054985.2011.577944
- Gilbert, J. (2005). *Catching the knowledge wave? The knowledge society and the future of education*. Wellington, NZ: NZCER Press.
- Hague, C., & Logan, A. (2009). *A review of the current landscape of adult informal learning using digital technologies*. Retrieved from http://preview.futurelab.org.uk/resources/documents/project_reports/becta/Adult_Informal_Learning_educators_report.pdf
- Hegarty, B., Penman, M., Kelly, O., Jeffrey, L., Coburn, D., & McDonald, J. (2010). *Digital Information Literacy: Supported Development of Capability in Tertiary Environments*. Wellington, New Zealand: Ministry of Education. Retrieved March 10, 2012 from http://www.educationcounts.govt.nz/publications/tertiary_education/80624
- Hight, C., Khoo, E., Cowie, B., & Torrens, R. (2014). Software literacies in the tertiary environment. In B. Hegarty, J. McDonald, & S.-K. Loke (Eds.), *Rhetoric and reality: Critical perspectives on educational technology*. Proceedings ascilite 2014 (pp. 410–415). Dunedin, New Zealand. Retrieved from <http://ascilite2014.otago.ac.nz/proceedings/>
- Jenkins, H., Purushotma, R., Weige, M., Clinton, K., & Robison, A. J. (2009). *Confronting the Challenges of Participatory Culture - Media Education for the 21st Century*. Cambridge, MA: MIT Press. Retrieved June 20, 2012 from http://mitpress.mit.edu/sites/default/files/titles/free_download/9780262513623_Confronting_the_Challenges.pdf
- Johnson, S. (1997). *Interface Culture: How New Technology Transforms the Way We Create and Communicate*. HarperCollins Publishers.
- Jones, C., & Czerniewicz, L. (2010). Describing or debunking? The net generation and digital natives. *Journal of Computer Assisted Learning*, 26(5), 317–320. doi:10.1111/j.1365-2729.2010.00379.x
- Jones, C. M. (2008). From novice to expert: Issues of concern in the training of psychologists. *Australian Psychologist*, 43(1), 38–54. doi:10.1080/00050060601089470
- Kennedy, G., Judd, T. S., Churchward, A., Gray, K., & Krause, K.-L. (2008). First year students experiences with technology: Are they really digital natives? *Australasian Journal of Educational Technology*, 24(1), 108–122. doi:10.14742/ajet.1233
- Khoo, E., Hight, C., Torrens, R., & Cowie, B. (2013). *Copy, cut and paste: How does this shape what we know?* TLRI research in progress. Retrieved February 1, 2014 from <http://www.tlri.org.nz/tlri-research/research-progress/post-school-sector/copy-cut-and-paste-how-does-shape-what-we-know>
- Khoo, E., Johnson, M., & Zahra, A. (2012). I learnt a whole lot more than churning out an essay: Using online tools to support critical collaborative inquiry in a blended learning environment. *Journal of Open, Flexible and Distance Learning*, 16(1), 127-140. Retrieved from <http://journals.akoatearoa.ac.nz/index.php/JOFDL/article/view/94>
- Kitchin, R., & Dodge, M. (2011). *Code/Space: Software and Everyday Life*. Cambridge, MA: The MIT Press. doi:10.7551/mitpress/9780262042482.001.0001
- Kvavik, R. B. (2005). Convenience, communications, and control: How students use technology. In D. Oblinger & J. Oblinger (Eds.), *Educating the Net generation* (pp. 7.1–7.20). Retrieved from <http://www.educause.edu/research-and-publications/books/educating-net-generation/convenience-communications-and-control-how-students-use-technology>

Livingstone, S., Wijnen, C., Papaioannou, T., Costa, C., & Grandio, M. (2014). Situating media literacy in the changing media environment: critical insights from European research on audiences. In N. Carpentier, K. Christian Schröder, & L. Hallet (Eds.), *Audience transformations: Shifting audience positions in late modernity* (pp. 210–227). Abingdon, UK: Routledge.

Manovich, L. (2013). *Software takes command. International Texts in Critical Media Aesthetics*. London: Bloomsbury Academic Plc.

Oblinger, D. (2003). Boomers, gen-Xers, and millennials: Understanding the “new students.” *EDUCAUSE Review*, 38(4), 36–45.

Peeters, J., Backer, F. D., Buffel, T., Kindekens, A., Struyven, K., Zhu, C., & Lombaerts, K. (2014). Adult learners informal learning experiences in formal education setting. *Journal of Adult Development*, 21(3), 181–192. doi:10.1007/s10804-014-9190-1

Prensky, M. (2001). Digital natives, digital immigrants. *On the Horizon*, 9(5), 1–6. doi:10.1108/10748120110424816

Selwyn, N., & Facer, K. (2007). *Beyond the digital divide: Rethinking digital inclusion for the 21st century*. Futurelab. Retrieved March 15, 2013 from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.101.3384&rep=rep1&type=pdf>

Tapscott, D. (1999). *Growing up digital: The rise of the net generation*. New York, NY: McGraw-Hill.

Tapscott, D. (2009). *Grown up digital*. New York, NY: McGraw-Hill.

Underwood, J. (2009). *The impact of digital technology*. BECTA. Retrieved from <http://publications.becta.org.uk/display.cfm?resID=41343>

Vallance, M., & Towndrow, P. A. (2007). Towards the informed use of information and communication technology in education: A response to Adams PowerPoint, habits of mind, and classroom culture. *Journal of Curriculum Studies*, 39(2), 219–227. doi:10.1080/00220270601105631

Valtonen, T., Pontinen, S., Kukkonen, J., Dillon, P., Väisänen, P., & Hacklin, S. (2011). Confronting the technological pedagogical knowledge of Finnish net generation student teachers. *Technology, Pedagogy and Education*, 20(1), 1–16. doi:10.1080/1475939X.2010.534867

van Leeuwen, T., & O’Halloran, K. L. (2011). *Towards a social theory of semiotic technology: Exploring PowerPoint’s design and its use in higher education and corporate settings*. Retrieved May 2, 2012, from <http://datasearch2.uts.edu.au/cpce/projects/detail.cfm?ProjectId=2008000350>

ADDITIONAL READING

Adams, C. (2006). PowerPoint, habits of mind, and classroom culture. *Journal of Curriculum Studies*, 38(4), 389–411. doi:10.1080/00220270600579141

Bennett, S., Maton, K., & Kervin, L. (2008). The digital natives debate: A critical review of the evidence. *British Journal of Educational Technology*, 39(5), 775–786. doi:10.1111/j.1467-8535.2007.00793.x

Edwards, E. (2015). Software and the hidden curriculum in digital education. *Pedagogy, Culture & Society*, 23(2), 265–279. doi:10.1080/14681366.2014.977809

Kitchin, R., & Dodge, M. (2011). *Code/Space: Software and Everyday Life*. Cambridge: The MIT Press. doi:10.7551/mitpress/9780262042482.001.0001

Livingstone, S., Wijnen, C., Papaioannou, T., Costa, C., & Grandio, M. (2014). Situating media literacy in the changing media environment: critical insights from European research on audiences. In N. Carpentier, K. Christian Schröder, & L. Hallet (Eds.), *Audience transformations: Shifting audience positions in late modernity* (pp. 210–227). Abingdon, England: Routledge.

Lynch, T. R. (2015). *The Hidden Role of Software in Educational Research: Policy to Practice*. Routledge Research in Education. Abingdon, England: Routledge.

Manovich, L. (2013). *Software takes command*. *International Texts in Critical Media Aesthetics*. London: Bloomsbury Academic Plc.

Prensky, M. (2001). Digital natives, digital immigrants. *On the Horizon*, 9(5), 1–6. doi:10.1108/10748120110424816

van Dijck, J. (2013). *The culture of connectivity: A critical history of social media*. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199970773.001.0001

KEY TERMS AND DEFINITIONS

Affordance: A person's perceived opportunity to utilise a particular tool for action, for example, a doorknob is for turning. Within software, these are typically organised through user interfaces.

Media and Information Literacy: An umbrella term for a number of competencies, skills

and understandings associated with contemporary communication and information practices. This term encompasses the more specific term digital literacy, which is associated with digital media practices.

Pedagogy: Study and practice of teaching.

Software: Machine readable Instructions which directs a computer's processor to perform specific operations. For everyday software applications, these instructions are put into operation as computer users run an application.

Software Application: A computer program designed to perform a set of tasks or functions. Common everyday examples range from an 'app' on a smartphone, to desktop computer programs such as a web browser.

Software Literacy: A person's expertise in understanding, applying, problem solving and critiquing software in pursuit of particular learning and professional goals.

Software Platform: A coherent programming environment or system which supports applications. Examples include operating systems such as iOS or Android, or social media platforms such as Facebook.

Software Process Improvement for Web-Based Projects Comparative View

S**Thamer Al-Rousan***Isra University, Pakistan*

INTRODUCTION

In the last two decades, there have been a significant increase on the expectations and demand of Web-based application due to the influence of the World Wide Web on our modern economy (Pressman, 2004). At the same time, the development, deployment and maintenance processes of the web-based systems which have become more and more complex and difficult to manage, have not progressed at a sufficient rate to meet these demand and expectations (Cardoso, 2007).

Many developers of web-based projects do not take into concern the unique requirements and characteristics of Web applications. They fail to realize that characteristics and requirements of web-based systems considerably different from that of traditional software, and so does their development (Alrifai, 2008). Hence, many developers conduct web applications in an ad hoc manner, and fail to adopt sound design methodologies, resulting in poor quality web systems and applications. A survey by the Cutter Consortium (2006) shows that failure to meet business needs (84%), project schedule delays (79%), and budget overruns (63%), lack of functionality (53%) and poor quality of deliverables (52%) are the main problems cited by the stakeholders of such applications.

In finding solutions to the problems of web-based application development, there has been an increasing pressure on the Web-based application industry. The search for solutions to improve Web-based software development has continued for many years and Web-based applications organizations are now realizing that one of their

fundamental problems is the inability to effectively manage the Web-based applications development process (Sulayman & Mendes, 2011). Software Process Improvement (SPI) has been recognized as an efficient and effective way for organizations to improve their quality of the software they develop and the productivity with which they work with. For this, SPI is an essential tool for improving Web-based applications development process (Sulayman & Mendes, 2011).

Process improvement in Web-based software projects is of growing concern for many reasons such as reducing cost and time, producing high quality software, and improving productivity (Smite et al., 2011). Despite the importance of SPI implementation, there is evidence that the vast majority of Web-based software development organizations are reluctant to adapt SPI models and standard because of their complexity (Abdel-Hamid & Abdel-Kader, 2011). It has been detected that improvement efforts in Web-based software projects based on process improvement frameworks which are designed for traditional software projects fails most of the time (Sulayman & Mendes, 2011).

In the absence of theoretical or empirical work examining the suitability of the traditional SPI standards and models for Web-based projects development, we believe that it is important to examine the suitability of the existing SPI standards and models for Web-based projects development process. The motivation for this study was to fill the gaps in the field of software process improvement. The main contributions of this study are:

1. Examine the suitability of the existing software process improvement models and standards for Web-based projects.
2. Increasing awareness on the importance of software process improvement in Web-based projects.

We believe that achieving these goals will lead to enrich the SPI in Web-based projects with new properties that leads to enhance the SPI projects implementation.

BACKGROUND ON WEB-BASED SOFTWARE DEVELOPMENT

There is a relatively small but a rising research on the differences between the Web-based applications and traditional software development. In general, this literature shows that the Web-based applications have certain unique inherent characteristics that make Web-based development considerably different and possibly more difficult comparing with software development (Rousan et al., 2014). The Web-based application characteristics which are usually built in shorter time-frames, serving as direct interface to various stakeholders, should meet a broad set of requirements and more than often serve a specific group of users. They are usually built from template solutions, by using coarse-grained authoring tools, which were developed by a multidisciplinary team (Ginige, 2002). These characteristics constitute the reasons why many concepts, methods, techniques, and tools of traditional software development are either insufficient to meet the needs of Web-based applications or have to be modified in order to do so (Pressman, 2004). Thus, Web-based project developers must have a reasonable understanding of the characteristics of Web-based project development and how these may affect the outcome of the project.

Web-based project development process is a complex and a demanding activity. Hence, the development of high-quality Web applications

does not occur naturally, there must be a systematic plan, complete with managing and improving the process, testing, and evaluation, to develop and implement the Web-based software project accordingly (Smite et al., 2011). Under such demanding circumstances, SPI is becoming increasingly important for Web-based software project.

RELATED WORKS AND RATIONALE OF SPI

The motivation for execution process improvement activities is to gather information as to what needs to be changed and to set up how to follow the improvements in order to reduce development cost and increase the quality of products created (Pino et al, 2008). SPI is about making things better. Unfortunately, the majority of Web-based software development organizations don't concentrate on process improvement activities. Some cases about the implementation of SPI founded in literature. A case study about the performance of SPI models in Web-based software development was proposed in (Sulayman & Mendes, 2011). Software process improvement in Web-based software development based on agile methods was proposed in (Abdel-Hamid & Abdel-Kader, 2011). The continuous improvement in the basis of the existing method was been provided by Nawazish et al. (2010). It gives some comments and suggestions that must be followed during the Web-based software development process. It concentrates on the gathering information that need to be changed, reducing the cost of development, and increasing the quality of the product produced by setting up how to follow the improvements. Salinas et al. (2012) analyze whether the practices proposed in Scrum meeting the CMMI generic and specific goals or not. They found that the CMMI is not appropriate for Web-based projects that based on scrum.

Software process Improvement is a continuing practice and must always be followed up with the analysis of actual areas of improvement. When SPI implemented effectively, the results can be

measured in the enhancement of product quality, and customer satisfaction. The study by Nawazish et al (2010) gave a description of the strengths and weakness of SPI in terms of Quality Functional Deployment (QFD). It based on the customer satisfaction. It considers the customer's needs according to the CMMI model. Several SPI models are designed particularly to improve product quality and meet customer satisfaction. One of these models is OWPL model (Alexandre et al, 2010). OWPL model is employed to enhance the product quality and to evaluate process improvement on organizational development.

Many SPI models are technically challenging (Sulayman & Mendes, 2011). A lot of organizations undergoing SPI projects have reported encountering problems and difficulties. The study of Amescua et al. (2006) shows a list of the resistance factors that affects organizations that implement the SPI projects. Hall et al. (2002) stated some major problems encountered during the implementation of SPI project. However, these problems include: hardly transferred technology, poorly managed information, changeable goals, unclear picture about the current status of the SPI project, and unclear distribution of roles. In addition, Baddoo et al. (2007) stated that the deciding factor in minimizing project failure is effective communication among developers, users, and project managers.

Despite the importance of SPI, it regularly is seen as unnecessary bureaucracy between software engineers and computer programmers. Some of the SPI models are small and cheap, but they have amazing results in spite of their cost. Some take years and millions of dollars to apply in spite of their popularity. However, SPI models share some common characteristics regardless of their scope, size, and complexity (Falessi et al, 2014). In next section, we attempted to check the suitability of the some of the software process improvement models and standards to meet the special characteristics and requirements for the Web-based projects.

ANALYSIS OF THE RESISTANCE FACTORS IN WEB-BASED PROJECTS

Based on software process improvement (SPI) literature, there are various factors which are influencing the implementation of SPI project. This study has categorized all those factors according to Beecham's research as presented by Clarke & Connor (2012). All those factors have been broadly divided into 2 main categories which are:

1. Organizational factors and
2. Project factors.

The organizational factors related to problems within the scope of the organization and are usually under senior manager's responsibility as reported in (Ali & kalmsteiner, 2014), meanwhile project factors related to problems regarding the software project management such as planning, activities, resources among others.

- **Organizational Factors:** Software processes play an important role in coordinating different teams in large or small. These are the factors which are related within the scope of the organization and are usually under senior managers' responsibility as presented by Clarke & Connor (2012) and others as stated in (Johansen, et al., 2016; Laporte, et al., 2016; Aydan, et al., 2015; Laporte, & Chevalier, 2015; Laporte, & Connor, 2015; Galvan, et al., 2015; Garcia, et al., 2015; ISO/IEC, 2015; Clarke & Connor, 2015; Exter, 2014; Laporte, et al., 2014; Jeners, et al., 2013). There are 5 factors which are categorized under organizational factors namely: Human, Political, Cultural, Goals and Change Management. Table 1 provides a description states the key resistance factor for each one identified.

Table 1. Description of each organizational factor

Factors	Description	Key Resistance Factors
Human	Without commitment from all organizational levels (human) to support SPI, the initiative will most likely fail or the results are not far reaching (Clarke & Connor, 2015). The experience of senior management with an SPI project will give positive impacts to the improvement process. Consultation support such as advice and training of SPI action teams and staffs is one critical aspect in ensuring the success of SPI project (Laporte, et al., 2014).	• Lack of commitment in all levels of the organizations. • Lack of adhesion and participation of all the individual involved in SPI projects. • Lack of professionals experience and skill. • Lack of leadership and backup by top management level. • Lack of adequate training
Political	The political factor is important to the construction for a department to SQA (Garcia, et al., 2015). The establishment of quality policy which is one of the political issues comes after the commitment of senior management. Quality policies and standards for SPI efforts describe the organizational goals and objectives related to the quality.	• Lack of the establishment of organizational policies. • Lack of the establishment of Quality Policy.
Cultural	In order to ensure success in a cultural change program, a champion who can build, deploy, drive, and own each initiative going forward must be properly decided (Larrucea, et al., 2016). However, every cultural change program requires good cooperation from both management and tactical technical staff, improvement programs will fail if either group is left out or underemphasized	• Lack of expertise in implementing cultural changes.
Goals	According to Exter (2014) research, if the goals, deadlines, and expected results by the managers are impractical, the effort towards SPI may be unsuccessful Laporte, & Connor (2015) found a reason why software process needs to have indefinitely ongoing processes. It is essential that clear goals need to be specified earlier, so that progress towards those goals can be continually monitored, and so that revisions to either goals, of processes, or both can be made persistently.	• Lack of consistency between software processes improvement project and the organization's strategic objectives. • Absence of focus on the organization's most urgent needs. • Unrealistic expectation towards the SPI project.
Change Management	Initial analysis needs to be conducted to determine whether the SPI initiative apt with the organization' objectives and interests. This is also discussed and supported by Laporte, et al. (2016). Exter (2014) in conceived that the SPI project's team be used to actively facilitate the efforts toward changes on the part of the project teams rather than simply check the situation of the ongoing process in order to report a long and depressing list of findings. Clarke & Connor (2012) highlighted that in order to have a successful software process improvements, risk from configuration and change management which is not explicitly defined will lead to unsuccessful business process.	• Insufficient and ineffective assessment of the current software process

- **Project Factor:** These are the factor which reflects the resistance factors on ongoing project which contributes and gives impact during the software process and it involves contribution from all level of personnel management as described by Clarke & Connor (2012) and others (Exter, 2014; Laporte, et al., 2014; Jeners, et al., 2013). There are 4 factors which are categorized under project factors namely: Budget, Estimates, Documentation, Quality, Tools

and Technologies. Table 2 provides a description states the key resistance factor for each one identified.

SOFTWARE PROCESS IMPROVEMENT MODELS AND STANDARDS

Different software process improvement models and standards such as Trillium, BOOTSTRAP,

Table 2. Description of each Project factor

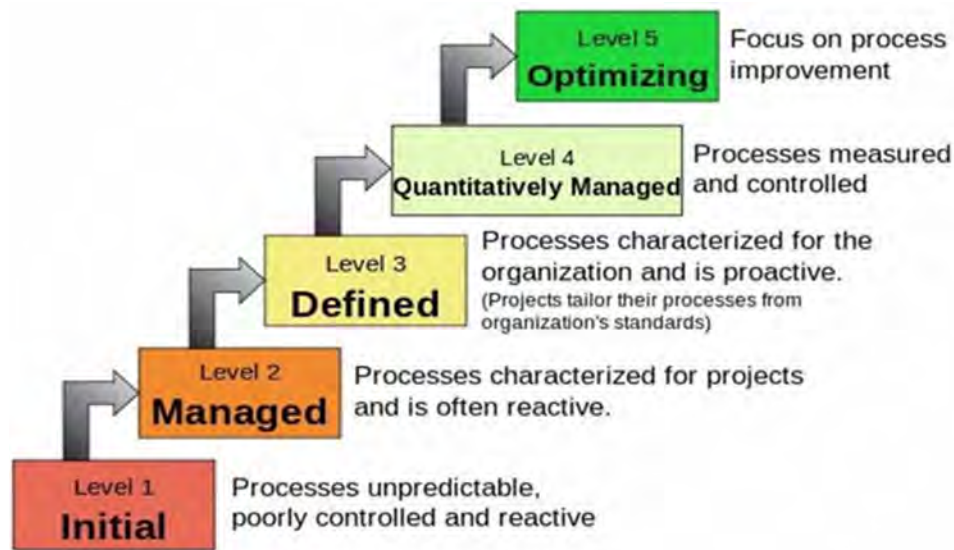
Factors	Description	Key Resistance Factors
Budget and Estimates	According to Laporte, et al. (2016) lack of progress in improvement plans is frustrating to those who really want to achieve progress and this put down the importance of time and costs in the process evaluation. Further research support by Aydan, et al. (2015) stated that the senior management often overlook on the manageability of SPI project size issue and sufficient budget size for SPI project issue. The first year is the most difficult period for a SPI program. Resistance to change will be at its peak, the costs are likely to be higher than in the 17 following years and due to the steep learning curve, the first year goals and targets can easily be missed.	• Current budget and estimates exceeds planning. • Lack of understanding by top management level that the software processes improvement project is a long-term return on investment process. • Lack of visibility about the ongoing software processes improvement project activities.
Documentation	In SPI project, documentation is a must in order to provide proof and dissemination throughout the organization in a formal way. Therefore, it is helpful to have an infrastructure for documentation, since it is a mandatory practice throughout the organization. According to Clarke & Connor (2012) the documentation is also gaining importance in the list of problems associated to SPI. It includes data measurement, proceedings register, coordination and management of the documentation, data collecting the operational framework forms the relationships and dependencies between what is to be done, by whom, and how to do it.	• Excessive documentation and formality. • Lack of infrastructure and of a documentation management. • Lack flexibility in the use of the documentation in projects of different types and sizes.
Quality	Quality refers to the state of the software as it was released or delivered to customers. In order to achieve higher level of quality assurance is by creating consent about how all the requirements definition processes has to be performed and which information should be provided to ensure the successfulness of the SPI project.	• Lack of involvement of top management in the relationship between the project teams and the person or group of quality assurance. • Lack of treatment to guarantee process conformity in instances of hiring and/or dismissal of skilled professionals.
Tools and Technology	The problem associated refers to the implementation of new tools and technologies, amount of work and pressures that hinder the use of new tools. According to Umarji, complexity of SPI tools and technologies need to ease with which developers can adapt to changes in work practices caused by SPI (Exter, 2014). However it has several acceptance issues because it often involves learning new technology, changes in work practices and an additional workload. Also, SPI involves collecting data about projects, resources and deliverables and often practitioners are not keen on sharing this type of data.	• Automation of not well-defined processes. • Lack of training on the support tools and technologies defined as support. • Pressure and absence of planning concerning the adaptation period.

SPICE, CMM, and CMMI have been proposed to help organizations to achieve more expectable results by improving their software process (Nawazish et al, 2010). This study discusses and compares some SPI models for Web-based project from a comparative view. The SPI models examined in this subsections all share common characteristics. They are well defined, measurable, repeatable, helpful, and widespread in use.

CAPABILITY MATURITY MODEL

The Capability Maturity Model (CMM) was developed by the US Department of Defense at Software Engineering Institute (SEI) in the mid-1980s (Salinas et al, 2012). CMM is used for the improvement of the existing software development processes, and can also be used to other processes. The main role of this model is

Figure 1. Capability maturity model



to develop a process maturity framework and to manage the process (Falessi et al, 2014). This framework supports the organization to enhance their software process by using the five maturity levels (Initial, Managed, Defined, Quantitatively Managed, and Optimized Levels) (Dangle et al, 2005). Figure 1 shows the CMM model. Each level in CMM needs to accomplish several Key Process Areas (KPA's) which composed of key practices that contributes to satisfying their aims and objectives.

Maturity levels consist of a predefined set of process areas. The maturity levels are measured by the achievement of the specific and generic goals that apply to each predefined set of process areas. Every Key Area must be implement the predefined objectives, which are, the essential infrastructure of the objective satisfaction or the activities to be developed. Every maturity level creates a different component in the software process, resulting in the capability increase of the organization process (Falessi et al, 2014).

For Web-based projects, this model is having certain disadvantages: First, When organizations use CMM, the developers look to the next level as

a target, this can be a risky for Web-based projects because if they become highly concentrated their attentions on reaching the next level, they may forget the actual target, namely to improve the processes. Second, CMM has sometimes been problematic. The application of multiple models which has not been integrated within an organization could make assessment, training, and improvement activities to be costly and take long time. The fact found in several empirical studies is that Web-based project is very short, usually not more than six months, and the average is less than three months (Al-Rousan et al, 2013). Finally, CMM approach would be very helpful when it is incorporated early enough in software development which mean it cannot be used to recover from a crisis in an emergency situation (Dangle et al, 2005). For Web-based projects the requirements of the target groups are unknown at the beginning of the project and project requirements change continuously, causing the vision of the Web-based application to evolve continuously which make the CMM approach unsuitable for Web-based projects development process.

CAPABILITY MATURITY MODEL INTEGRATION

Capability Maturity Model Integration (CMMI) was created by Software Engineering Institute by combining the CMM models (SW-CMM V2.0, Integrated Product Development (IPD), and System Engineering CMM (SE-CMM) (Salinas et al, 2012). CMMI focuses on a set of process areas that are related to software process improvement and capability. CMMI is the same as CMM, it shows the processes areas in five maturity levels. The difference can be seen from the fact that CMMI is an integrated approach that establishes a framework for the purpose of building an initial set of integrated models, or are used to integrate future and current models. CMMI has two representations (Jaramillo et al, 2015):

1. **Stages Model:** Designed to provide a standard sequence of improvements. It provides a way to assess the company's process capability at 1-5 levels (Initial, Aware, Defined, Quantitatively, and optimizing).
2. **Continuous Model:** Focus on the specific processes areas that are considered important for the company's direct business goals. Each process area is rated in terms of maturity level (0-5).

The main difference between the two representations is that the staged model is used to evaluate the capability of the company as a whole, while the continuous model is used to assess the maturity of certain process areas within the company.

CMMI includes a set of goals, which are brief descriptions of a wanted state that must be reached by a company. The CMMI has also specific goals, which are related to each process area and describe the wanted state for that area. In addition, CMMI includes a set of generic and specific practices that cover a set of process areas. These practices describe the ways of achieving a goals. Many generic and specific practices be related with every goal within a process area (Kerzazi, 2015).

Regardless how good CMMI approach can be; for Web-based projects, this approach is having certain disadvantages: First, CMMI has a lot of technical definition and documentation because it covers practically the whole thing from initial level process to optimized level, which make it ambiguity (Dangle et al, 2005). As the short development cycle is so common for Web-based projects, it is assumed to be the first requirement in the Web-based project improvement process (Pressman, 2004). Second, documentation is a must in order to provide proof and dissemination throughout the organization in a formal way. Therefore, it is helpful to have an infrastructure for documentation, since it is a mandatory practice throughout the organization. Thus, in order to initiate CMMI-based process improvement in Web-based projects, it requires a major shift in organizational culture and attitude, which make it hard and expensive.

Finally, CMMI approach just provides developers practices that are required to improve the process but it does not state how to implement them (Jaramillo et al, 2015). In addition, CMMI has a lot of technical definition and documentation because it covers practically the whole thing from initial level process to optimized level, which make it ambiguity (Dangle et al, 2005). With the high degree of individuality and multidisciplinary team in Web-based projects, the implementation of this approach in Web-based projects become fuzzy. As a result, this led to the argument that CMMI approach is not suitable for the development process of Web-based projects.

TRILLIUM

Trillium model was developed by Bell Canada, Nortel and Research Bell Northern. It combines requirements from the ISO 9000 series, CMM, internal Bellcore standards, and software quality standards from the IEEE (Harrington, 2011).

Trillium model and its application have been developed based on the feedback and experiences

of the suppliers (Model & Eizenman, 2012). It has been used for self-evaluation and for the evaluation of existing and possible suppliers. Trillium model was primarily designed for use with embedded software systems (for instance, telecommunications). So, Trillium model includes process info that is exclusive to the telecommunications field (Harrington, 2011). Nevertheless, much of the model can be useful to other parts of the software industry for example Management Information Systems (MIS). Although the Trillium model is based on the SEI model (CMM version 1.1), its structural design differs from the CMM in the following ways (Harrington, 2011):

1. Trillium model structural design is based on roadmaps, instead of key process areas present in CMM.
2. Trillium model has a wider product perspectives instead of only perspective that based on SPI.
3. Trillium has more awareness about issues that impacting on capability.
4. Trillium model has awareness towards customer concerns, telecommunication industry and technological maturity.

Trillium is comprised of five levels. These are (Gonzales et al, 2006):

1. **Unstructured:** The software development process is ad-hoc. Projects normally cannot meet schedule targets or quality. Success is based on individuals instead of organizational infrastructure.
2. **Repeatable and Project Oriented:** Individual project success is accomplished via stable project management planning and control, with emphasis on estimation techniques, requirements management, and configuration management.
3. **Defined and Process Oriented:** Processes are identified and used at the organizational level, though project customization is still

allowed. Processes are improved and controlled. ISO 9001 requirements such as internal process reviewing and training are integrated.

4. **Managed and Integrated:** Process analysis is used as a major mechanism for process improvement. Process change management and fault restriction programs are integrated into processes. CASE tools are integrated also into processes.
5. **Fully Integrated:** Formal methodologies are widely used. Development history and organizational repositories are used are integrated into processes. Every capability area span several levels.

Every level (except first level) includes numerous capability areas, which in turn cover roadmaps that included activities. These is very similar to CMM. Every capability area combines one or more roadmaps, and each roadmap includes one or more practices that span numerous Trillium levels (Harrington, 2011).

For Web-based project, this approach is having certain disadvantage: The model includes process information that is unique to the telecommunications field, and a significant percentage of the practices described in the model related directly to hardware development. For this, Web project cannot be adopted as-is. In addition, Formal methodologies, stable project management planning and control are critical for this model. In Web-based project, developers must deal with unknown field of businesses and the business requirements may change dramatically, so the formal methodologies are hard to be adopted. Furthermore, Trillium model need adequate training, professional experiences and skills, while Web application developers are significantly young, and they have less experience compared to the traditional software developers. Moreover, the time to learn and apply these kind of models is frequently not available which make the Trillium approach unsuitable for Web project development process

HEALTHCHECK

In 1982, Healthcheck was developed within the British Telecom (Mackie & Rigby, 1993). The goal was to find an approach of evaluating software process versus existing world best practice to allow direct improvement initiatives for software process, product and staff working in software development. The purpose was to improve the quality of the software and so to understand where to allocate investment in changes. The idea was to apply a series of face to face interviews with team members that work on a particular project or unit in order to assess them (Patricia, 2004). All interviews are based on the Healthcheck questionnaire, which consists of nine key process areas, each process area containing from 10 to 16 questions (Norris et al, 1994). Software development team members should be involved from the very start of the interview process, because their opinion counts when making decision on what needs improvement or adjustment. The key process area are scored on a percentage scale. Based on experts' view, the key process area with 100% score indicates that a unit is working according to existing world best practice. The results are showed as radar diagrams using a percentage scale (Patricia, 2004).

In Web-based, the major problems in starting SPI process is with managing the resources such as methods, practices, people, tools and technology they use. Besides, Web-based organizations are expanding their businesses with other client organizations. This is done in order to decrease cost and improve operations. Web-based organizations are employing more qualified and experienced people who are not part of their organizations or organizational process. So it's hard to take their opinion when making decision on what needs improvement or adjustment.

Healthcheck model idea was to apply an improvement cycle that included steps identified as assess, analyses, modify and execute. For Web-based projects, applying this cycle is a complex

job due to relatively short development time and the complexity of Web-based software development environment and the process. In addition, Healthcheck model needs participation of all the individual involved in SPI process. It also need commitment in all levels of the organizations. Boundaries between countries in Web-based project development are disappearing, and software development process is becoming globally oriented, multidiscipline and cross-cultural which make the practices adopted in Healthcheck unsuitable for Web-based projects development process.

SOFTWARE PROCESS IMPROVEMENT AND CAPABILITY DETERMINATION

Software Process Improvement and Capability Determination (SPICE) is an international framework for assessment of software processes. SPICE developed cooperatively by the IEC (International Electro technical Commission) and the ISO (International Organization for Standardization). SPICE is also known as ISO/IEC 15504 (Laporte et al, 2006).

SPICE was build based on existing software estimation models such Trillium, CMM, and Software Technology Diagnostic (STD) (Rout et al, 2007). SPICE provides a group of documents, which are worked as a framework for the measurement of software process. Organizations can employ these documents in numerous stages of production, for example in planning, managing, monitoring, controlling and improving acquisition, supply, development, operation, evolution and support of software (Rout et al, 2007).

The SPICE documents was accordingly structured into seven working groups trying to provide information to create a standard for software process evaluation and improvement. The working groups had specific goals for creating the seven products that were to include the SPICE product group. The products were (Rout, 2010):

1. **Introductory Guide (IG):** Shows how different parts of the standard fit together and offers guidance for their selection.
2. **Baseline Practices Guide (BPG):** Describes the goals and important activities that are essential for good software engineering, and more precisely defines what activities are needed but not how they are implemented.
3. **Assessment Instrument (AI):** A guide for designing tools that elicit data related to the processes undergoing estimation.
4. **Process Assessment Guide (PAG):** A document that identifies the estimation method, which in turn describes how to perform the estimation using the estimation tool and the basic practices.
5. **Process Improvement Guide (PIG):** Offers guidance on how to use the estimation outcomes for process improvement.
6. **Process Capability Determination Guide (PCDG):** Offers generic guidance on how to employ the estimation outcomes for capability evaluation.
7. **Assessor Training and Qualification Guide (ATQG):** Offers generic guidance for the development of programs intended to train staff as evaluators utilizing this standard, and procedures.

This approach has certain disadvantages for Web-based projects. First, the development of ISO/IEC 15504 was based on the weaknesses of previous standards and models. It is considered as to complement other international standards and models for assessing the capability, effectiveness and quality of processes and organizations, for this is not easy to be understood. Second, ISO/IEC 15504 was created in a development context, making it difficult to apply in a service management context. Third, the staff who will take part in SPICE initiative most have relevant education and training and experience, personal qualities such as communication skills, specific skills for particular groups such as management skills for

the management group, and ISO/IEC 15504 related training and experience in process capability assessments. As result, SPICE approach unsuitable for Web-based projects development process.

CONCLUSION

This study reviewed different software process models and techniques that are helpful in software process improvement. All SPI models and standards mentioned above are different in characteristics. There is no single SPI approach studied in this study, which is able to meet all the requirements and characteristics of Web-based project development. Considering the importance of Software Process Improvement, and as the Web-based development projects is still lacks of the SPI models and standards that could serve as a guideline to the development of Web-based applications, it is extremely important that Web-based application developers understand the strengths and weaknesses employed on existing SPI models and standards, and they should adopt and adapt the best SPI models and standards or coming with a new models and standards to meet the requirements and challenges of the Web-based application development.

FUTURE RESEARCH DIRECTIONS

Main objective of this study was to investigate specific SPI models or techniques for Web-based projects development. The review as mentioned earlier helped us in identifying the possible research gaps and directions. One of the research gaps lies in proposing a specific SPI model for Web-based projects, which keeps in view their characteristics and aims to help them measure their success and improve continuously. This can be achieved by either enhancing some existing SPI model or by proposing one from scratch. This is the line of our future work.

REFERENCES

- Abdel-Hamid, A., & Abdel-Kader, M. (2011). Process Increments: An Agile Approach to Software Process Improvement. In *Proceedings on AGILE Conference* (pp. 195-200). Salt Lake City, UT: IEEE.
- Alexandre, S., Renault, A., & Habra, N. (2006). OWPL: A Gradual Approach for Software Process Improvement in SMEs. *Proceedings of the 32th International Conference on Software Engineering and Advanced Applications*, 328-335. doi:10.1109/EUROMICRO.2006.48
- Ali, N., & Kalmsteiner, M. (2014). Use and evaluation of simulation for software process education: A case study. In *Proceedings of the European Conference Software Engineering Education* (pp. 59-73). Seon Monastery, Germany: Shaker Verlag.
- Alrifai, R. (2008). A Project Approach for Teaching Software Architecture and Web Services in a Software Engineering Course. *Journal of Computing Sciences in Colleges*, 23(4), 237-240.
- Al-Rousan, T., Al-Haj Hassan, O., & Aljawarneh, S. (2013). A Comparative Study of Development Methodologies for Web Project: Towards a Viable Alternative. *Journal of Social Sciences.*, 40(1).
- Amescua, A., García, J., Sánchez-Segura, M., & Medina-Domínguez, F. (2006). Software Process Improvement for Practitioners Based on Knowledge Management Tools. *Proceedings of the 5th International Conference on Software Engineering, Parallel and Distributed Systems*, 24-29.
- Aydan, U., Yilmaz, M., & Connor, R. V. (2015). Towards a serious game to teach ISO/IEC 12207 software lifecycle process: An interactive learning approach. In *Proceedings of the 15th International Conference on Software Process Improvement and Capability determination* (pp. 217-229). Berlin: Springer-Verlag. doi:10.1007/978-3-319-19860-6_17
- Baddoo, N., Hall, T., & Wilson, D. (2000). Implementing a people focused SPI program. *proceedings of the 11th European Software Control and Metrics Conference and the Third SCOPE Conference on Software Product Quality*, 373-281.
- Cardoso, J. (2007). The Semantic Web Vision: Where are we? *IEEE Intelligent Systems*, 22(5), 84-88. doi:10.1109/MIS.2007.4338499
- Clarke, P., & Connor, R. V. (2012). The situational factors that affect the software development process: Towards a comprehensive reference framework. *Information and Software Technology*, 54(5), 433-447.
- Clarke, P., & Connor, R. V. (2015). Exploring the relationship between software process adaptive capability and organizational performance. *IEEE Transactions on Software Engineering*, 41(12), 1169-1183.
- Cutter Consortium. (2006). *Poor Project Management Number-one Problem of Outsourced E-projects*. Cutter Research Briefs. Retrieved October 24, 2015, from <http://www.cutter.com/research/2000/crb001107.html>
- Dangle, K., Larsen, P., Shaw, M., & Zelkowitz, M. V. (2005). Software Process Improvement in Small Organization: A Case Study. *IEEE Software*, 22(6), 68-75. doi:10.1109/MS.2005.162
- Exter, M. (2014). Comparing educational experiences and on-the-job needs of educational software designers. In *Proceedings of the 45th ACM Technical Symposium on Computer Science Education* (pp. 355-360). New York, NY: ACM. doi:10.1145/2538862.2538970
- Falessi, D., Shaw, M., & Mullen, K. (2014). Achieving and Maintaining CMMI Maturity Level 5 in a Small Organization. *IEEE Software*, 31(5), 80-86. doi:10.1109/MS.2014.17

- Galvan, S., Mora, M. R., Connor, R. V., Acosta, F., & Alvarez, F. (2015). A compliance analysis of agile methodologies with the ISO/IEC 29110 project management process. *Procedia Computer Science*, 64(1), 188-195.
- Garcia, L., Laporte, C. Y., Arteaga, J., & Bruggmann, M. (2015). Implementation and certification of ISO/IEC 29110 in an IT startup in Peru. *Software Quality Professional*, 17(2), 16-29.
- Ginige, A. (2002). Web Engineering: Managing the Complexity of Web Systems Development. In *Proceedings of the 14th International Conference on Software Engineering and Knowledge Engineering* (pp.721–729). ACM. doi:10.1145/568760.568885
- Gonzales, E., Hamrick, P., Smouse, E., & Dyer, E. (2006). Pollen-Mediated Gene Dispersal within Continuous and Fragmented Populations of a Forest Understory Species. *Molecular Ecology*, 15(8), 2047–2058. doi:10.1111/j.1365-294X.2006.02913.x PMID:16780423
- Hall, T., Rainer, A., & Baddoo, N. (2002). Implementing Software Process Improvement: An Empirical study. *Software Process Improvement and Practice*, 7(1), 3–15. doi:10.1002/spip.150
- Harrington, M. (2011). Empirical Evidence of Priming, Transfer, Reinforcement, and Learning in the Real and Virtual Trillium Trails. *IEEE Transactions on Learning Technology*, 31(2). Retrieved from <http://standards.iso.org/ittf/PubliclyAvailableStandards/index.html>
- ISO/IEC. (2015). ISO/IEC 29110-5-6-1:2015 – Systems and software engineering – Lifecycle profiles for very small entities (VSES) – Part 5-6-1: Systems engineering – Management and engineering guide: Generic profile group: Entry profile. Geneva, Switzerland: International Organization for Standardization.
- Jeners, S., Clarke, P., Connor, R. V., Buglione, L., & Lepmets, M. (2013). Harmonizing software development processes with software development settings: A systematic approach. In *Systems, Software and Services Process Improvement* (pp. 167–178). Berlin: Springer. doi:10.1007/978-3-642-39179-8_15
- Johansen, J., Colomo-Palacios, R., & Connor, R. V. (2016). Towards a manifesto for software process education, training, and professionalism. In *Software Process Improvement and Capability Determination*. doi:10.1007/978-3-319-38980-6_8
- Kerzazi, N. (2015). Conceptual alignment between SPEM-based processes and CMMI. In *Proceedings of the 10th International Conference on Intelligent Systems: Theories and Applications* (1-9). Morocco: IEEE. doi:10.1109/SITA.2015.7358391
- Laporte, C. Y., & Connor, R. V. (2015). Software process improvement in graduate software engineering programs. *Proceedings of the 1st International Workshop Software Process Education, Training and Professionalism*, 18-24.
- Laporte, C. Y., & Chevalier, F. (2015). An innovative approach to the development of project management processes for small-scale projects in a large engineering company. *INCOSE International Symposium*, 25(1), 819-839. doi:10.1002/j.2334-5837.2015.00099.x
- Laporte, C. Y., Connor, R. V., & Garcia Paucar, L. (2016). The implementation of ISO/IEC 29110 software engineering standards and guides in very small entities. *Proceedings of 10th International Conference on Evolution of Novel Approaches to Software Engineering*, 29-30.
- Laporte, C. Y., Hebert, C., & Mineau, C. (2014). Development of a social network website using the new ISO/IEC 29110 standard developed specifically for very small entities. *Software Quality Professional*, 16(1), 4–25.

- Laporte, Y., April, A., & Renault, A. (2006). Applying ISO/IEC Software Engineering Standards in Small Settings: Historical Perspectives and Initial Achievements. *Proceedings of the SPICE 2006 Conference*.
- Larrucea, X., Connor, R. V., Colomo-Palacios, R., & Laporte, C. Y. (2016). Software process improvement in very small organizations. *IEEE Software*, 33(2), 85-89.
- Mackie, C., & Rigby, P. (1993). Practical experience in assessing the health of the software process. *Software Quality Journal*, 2(4), 265-275. doi:10.1007/BF00403769
- Model, D., & Eizenman, M. (2012). A probabilistic approach for the estimation of angle kappa in infants. In *Proceedings of the 12th International Conference on Eye Tracking Research and Applications* (pp. 53-58). New York: ACM. doi:10.1145/2168556.2168564
- Nawazish, K., Mansoor, A., Khokhar, M., Rehman, S., & Rauf, A. (2010). MECA: Software Process Improvement for Small Organizations. In *Proceedings of the International Conference on Information and Emerging Technologies* (pp. 1-6). Karachi: IEEE. doi:10.1109/ICIET.2010.5625678
- Norris, M., Rigby, P., & Stockman, S. (1994). Life after ISO 9001: British Telecom's Approach to Software Quality, *Communications Magazine*. *IEEE Software*, 32(10), 58-63.
- Patricia, A. (2004). *An Evaluation of Software Development Practice and Assessment Based Process Improvement in Small Software Development Firms* (PhD dissertation). Sch.Com Info Tech, Griffith University, Australia.
- Pino, F., García, J., Pino, M., García, F., & Piatini, M. (2008). Software Process Improvement in Small and Medium Software Enterprises: A Systematic Review. *Software Quality Control*, 16(2), 237-261. doi:10.1007/s11219-007-9038-z
- Pressman, R. S. (2004). *Applying Web Engineering in Software Engineering: A Practitioner's Perspective*. New York: McGraw-Hill.
- Rousan, T., Al-Haj, O., & Al-Zobaydi, A. (2014). The Roles of Decision Making and Empowerment in Jordanian Web-based Development Organizations. *Web Engineering Journal*, 13(5&6), 21-30.
- Rout, P., El Emam, K., Fusani, M., Goldenson, D., & Jung, W. (2007). SPICE in Retrospect: Enveloping a Standard for Process Assessment. *Systems and Software*, 80(9), 1483-1493. doi:10.1016/j.jss.2007.01.045
- Rout, P. (2010). Guest Editorial Software Process Improvement and Ad Ability Determination: Selected Articles from SPICE. *Software Maintenance Evaluation*, 22(1), 239-241.
- Salinas, C., Escalona, M., & Mejías, M. (2012). A Scrum-Based Approach to CMMI Maturity Level 2 in Web Development Environments. In *Proceedings of the 14th International Conference on Information Integration and Web-based Applications & Services* (pp. 282-285). New York: ACM. doi:10.1145/2428736.2428782
- Smite, D., Wohlin, C., Gorschek, A., & Robert, F. (2011). Empirical Evidence in Global Software Engineering: A Systematic Review. *Empirical Software Engineering*, 15(1), 91-118.
- Sulayman, M., & Mendes, E. (2011). An Extended Systematic Review of Software Process Improvement in small and Medium Web companies. In *Proceedings of the 15th Annual Conference on Evaluation & Assessment in Software Engineering* (pp. 134-143). Durham, NC: IEEE. doi:10.1049/ic.2011.0017

KEY TERMS AND DEFINITIONS

CMMI: A model that supports evaluating and improving an organization's processes using a best-practice model.

ISO/IEC: An international standard for evaluating and improving an organization's software processes.

Model: An abstract and simplifying representation of an object or phenomenon of the real world.

Model-Based on SPI: Model-based on SPI approaches compare the current processes and practices of a development organization against a reference model or a benchmark.

Process Improvement (SPI): SPI means understanding existing processes and changing

these processes to increase product quality and/or reduce costs and development time.

Process Model: Representation of a real-world process that focuses only on specific features of the real-world process and does not capture all details.

QIP: The Quality Improvement Paradigm (QIP) is a six-step procedure for structuring software development and improvement activities.

Software Process Model: A software process model is a model of a software process.

A Study of Contemporary System Performance Testing Framework

S**Alex Ng***Federation University, Australia***Shiping Chen***CSIRO Data61, Australia*

INTRODUCTION

System Performance Testing is a vital activity spanning the Software Development Life Cycle (Siva & Sharma, 2006). There are different types of test: Functionality, Graphics, Collision, Performance, Gaming, User Interface, Progression, Correlation, Numerical, Conformance, Interoperability and many more. In this chapter, we find special interest in System Performance Testing. The purpose of system performance testing is to identify bottlenecks by measuring a system's responsiveness and scalability under a certain load (Sarojadevi, 2011). Load testing provides a means for assessing the behavior of a system under varying traffic burdens. Performance testing is able to provide simulation and modeling help enterprises experiment not just with load but also with specific situations, such as large numbers of users performing the same activity conducting a search or completing an online transaction simultaneously (Menascé, 2002). Performance tests conducted over time provide a good landscape trend which helps companies plan and test for future needs. Such testing can help determine what hardware resources need to be purchased to achieve desired future capabilities (Jain, 1991).

In this chapter, we discuss some of the contemporary system performance testing frameworks and present a general-purpose testing framework for both simple and complicated performance testing. We observed that few of the system

performance testing frameworks available have been widely used in industry or the research community (Yigitbasi, Iosup, Epema, & Ostermann, 2009). Our framework proposes an abstraction to facilitate performance testing by separating the application logic from the common performance testing functionalities.

The rest of this chapter is organized as follows. We first provide a definition of system performance and the requirements for system performance testing framework. Afterward, we present a discussion of some of the contemporary system performance testing frameworks. Lastly, we propose a system performance framework with a sample implementation to demonstrate the applicability of this framework.

BACKGROUND

According to (Meier, Farre, Bansode, Barber, & Rea, 2007), Performance Testing is defined as the technical investigation done to determine or validate the speed, scalability, and/or stability characteristics of the product under test. Performance-related activities, such as testing and tuning, are concerned with achieving response times, throughput, and resource-utilization levels that meet the performance objectives for the application under test.

(Meier et al., 2007) classify performance metrics into the following categories:

- **Network-Specific Metrics:** A set of metrics about the overall behavior of the network used to support the system.
- **System-Related Metrics:** A set of metrics helps identify the resource utilization of the system.
- **Platform-Specific Metrics:** A set of metrics related to software that is used to host the application system, such as the Microsoft.NET Framework common language runtime (CLR) and ASP.NET-related metrics.
- **Application-Specific Metrics:** These include custom performance counters inserted into the application code to monitor application health and identify performance issues.
- **Service-Level Metrics:** A set of metrics help measure overall application throughput and latency, or they might be tied to specific business scenarios.
- **Business Metrics:** These metrics are indicators of business-related information, such as the number of orders placed in a given timeframe for a particular department.

There are some common system performance metrics for enterprise systems, such as: *Response Time*, *Latency*, and *Throughput*. In some contexts it's customary to call these things by different names: *Throughput and Response Time*, or *Capacity and Delay*, or *Bandwidth and Latency*. We provide the following definitions to avoid ambiguity:

- **Response Time (RT):** It is the total time taken by a client to wait in invoking a server function and coming back with a result. Oxford Dictionary defined RT as *the length of time taken for a person or system to react to a given stimulus or event*.
- **Latency:** It is the total time spent by a system generated message to travel from its source to the destination. According to Oxford Dictionary, "latent" means "exist-

ing but not yet developed or manifest". Together, latency and bandwidth define the speed and capacity of a network. For an online system, latency is the total time spent by a message from its sender (source) to its receiver (destination).

In a synchronous client/server-based environment, RT is equivalent to Latency, where the client is both the message sender (requester) and receiver (respondent). This is why these terms are sometimes used alternatively. However, they are different in the asynchronous mode because the sender and the receiver are two different parties. The total time spent by a client (RT) in invoking an asynchronous request and waiting for the result to come back will be different from the total time for a message (Latency) to get from the source to its destination because the message sender may be different from the message receiver.

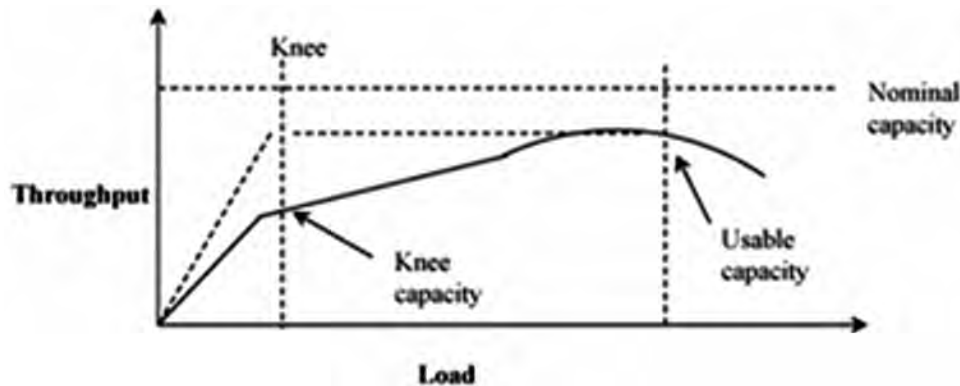
Furthermore, there exists a term called *Processing Time* to describe the amount of time a system takes to process a given request, not including the time it takes the message to get from the user to the system or the time it takes to get from the system back to the user.

With this view, Latency + Processing Time = Response Time.

- **Throughput:** It represents the server's capability in handling a certain number of client requests (such as messages or transactions) within a unit of time (such as second or minute). This shows how well a company's facility is able to cope with the demand of its client.

From Figure 1, we can see the throughput of a typical system demonstrates the following conditions: (1) *Nominal Capacity*: maximum achievable throughput under ideal workload conditions. (2) *Usable capacity*: maximum throughput achievable without exceeding the pre-specified response-time limit, and (3) *Knee Capacity*: A condition that starts to deviate from the linear

Figure 1. Typical throughput of a system



relationship between throughput and load. The *knee* is arrived when the response time is low and with high throughput.

REQUIREMENTS FOR SYSTEM PERFORMANCE TESTING FRAMEWORK

(Jain, 2015) specified seven criteria for selecting and evaluation technique for system performance:

- **Stage:** Select the appropriate technique suitable at different development life-cycle stage.
- **Time Required:** Consider the time effort required to implement the particular technique.
- **Tools:** Select the tools that are readily available that suits the environment.
- **Accuracy:** Specify and design the level of accuracy required by the system.
- **Trade-Off Evaluation:** Make smart consideration on the trade-off of different criteria.
- **Cost:** Select a technique that is most cost effective.
- **Saleability:** Ensure the selected technique should be reusable in the future.

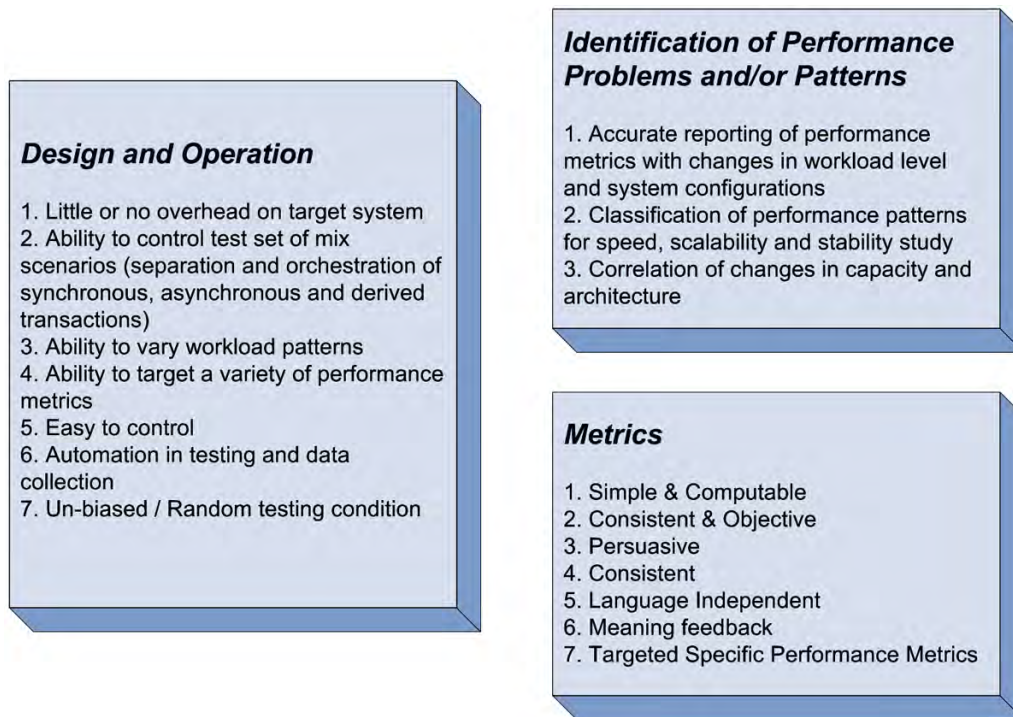
(Pressman, 1992) asserts that there are several common characteristics emerge when devising

metrics to measure system performance: *Simple and computable, Persuasive, Consistent and objective, Consistent in units or dimensions, Programming language independent and Gives feedback.*

According to (Mansharamani, 2011), the requirements for performance test consist of two groups: *understanding specifics of the workload*, and *gathering performance targets*. On the other hand, (Meier et al., 2007) stress that performance testing is conducted to address one or more risks related to *expense, opportunity costs, continuity, and/or corporate reputation*. Furthermore, (Illes, Herrmann, Paech, & Rückert, 2005) specify quality criteria for testing tools basing on the ISO/IEC 9126 standard (ISO/IEC, 2001) for software product quality: *Functionality, Reliability, Usability, Efficiency, Maintainability, Portability, General vendor qualifications, Vendor support, and Licensing and pricing*. In addition, (SOASTA, 2014) stresses that the top ten performance problems are related to issues with: *Architecture, Bandwidth, Load Balancer, Third Party Services, Connectivity, Shared Environment, Database Performance, Configuration Settings, Application, and Culture & Mindset*.

Figure 2 is a summary of the requirements for a system performance testing framework in three different categories: Design and Operation requirements; Identification of performance problems and/or patterns requirements; and Metrics requirements.

Figure 2. Requirements for system performance testing framework



AN OVERVIEW OF CONTEMPORARY SYSTEM PERFORMANCE TESTING FRAMEWORKS

System testing is a costly and time-consuming activity in software development (Rankin, 2002). As a result, a large number of automatic performance testing products and open source tools have been developed to address the above requirements, such as IBM's Rational Performance Tester and Rational Test Workbench, HP's LoadRunner ("HP loadRunner," 2015), Microsoft's Web Application Stress Tool ("Microsoft visual studio testing tools," 2015), Apache Jakarta's JMeter ("Apache jmeter," 2015), and Grinder ("The grinder, a Java load testing framework," 2015). Moreover, there are other products such as CA Technologies Service Virtualisation, HP Performance Centre, and Neotys Neoload ("Performance testing products," 2015).

Recent advances have enabled analysis tools more dynamic interaction with the application under study. The new trends of development in social computing, mobile services, information

explosion, and cloud computing has enabled many visionary elements being evolved rapidly and are fragmented. While these performance products and tools have been successfully used in a range of enterprise applications, few of them have been widely used in industry or the research community, partially due to the fact that they are either too expensive and complicated for small projects and companies, or too specific and simple for testing a wide variety of applications.

The Garner Magic Quadrant for Integrated Software Quality Suites report (Driver et al., 2014) named IBM, Microsoft, HP, Oracle, SOASTA, and Borland are among the Leaders Quadrant in the Integrated Quality Assurance (IQA) tool suite market. This development indicates not only the importance of performance testing, but also the attraction and popularity of these kinds of product. The leader products are generally supporting a broad set of technologies encompassing aspects such as integration, connectivity, synchronization between the suite components and other tools within the software development life cycle.

Table 1. Overview of contemporary system performance frameworks

Vendor: Product Name	Complete Test Management	Load Test Automaton	Support Mobile, Web and Legacy Applications	Versatile Reporting	Persuasive Metric	Ease of Installation and Setup	Cost / License
Borland: Silk Performer	Yes	Yes	Yes	Yes	Yes	Yes	Medium
CA Technologies: Service Virtualization	Yes	Yes	No	Yes	Yes	Medium	Medium
HP: LoadRunner and Performance Centre	Yes	Yes	Yes – Strong for Windows	Yes	Yes	Difficult	High
IBM: Rational Performance Tester & Rational Test Workbench	Yes	Yes	Yes	Yes	Yes	Medium	High
Microsoft: Application Centre Test	Yes	Yes	Yes	Yes	Yes	Yes	Medium
Oracle: Application Test Suite	Yes	Yes	Yes – Strong for database	Yes	Yes	Medium	Medium
SOASTA: SOASTA Platform	No	Yes	No – Focus strongly at cloud & mobile load testing for Web applications	Yes	Yes	Medium	Medium
Open Source: JMeter / Grinder	No	Yes	No – Java based Web Application only	No	Medium	Medium	Nil

Each framework has unique things that it does and conversely, each of the tools have unique deficiencies that will impede the effectiveness of individual test plan.

We assert that there is no one particular system performance testing framework that fits all – any option selected will be something of a trade-off. All the leaders provide friendly user interfaces, powerful automated script generators, and comprehensive performance analysis and reporting features. However, most of these functionalities are for web applications.

In particular, these products are considered very expensive and the prices are usually binding to the number of concurrent loads. For example, one of the leader products can cost around US\$65K

for a license with 100 concurrent clients. This makes it difficult to use the products for small companies and budget-hardship projects.

The Table 1 provides an overview of the features demonstrated by some of the contemporary system performance frameworks.

Apache Jakarta's JMeter ("Apache jmeter," 2015), and Grinder ("The grinder, a Java load testing framework," 2015) are two open source performance testing tools for the Java environment. The Grinder load testing framework can run a distributed test using many load injector machines. It can load test any Java APIs including HTTP web servers, SOAP and REST web services, and application servers (CORBA, RMI, JMS, EJBs), as well as custom protocols.

The Apache JMeter offers test functions including: Web - HTTP, HTTPS, SOAP / REST, FTP, Database via JDBC, LDAP, Message-oriented middleware (MOM) via JMS, Mail - SMTP(S), POP3(S) and IMAP(S), MongoDB (NoSQL), Native commands or shell scripts, and TCP. JMeter has no monitoring built in. Therefore, the user is required to program wrapper scripts to synchronize test data with external performance monitoring data.

A GENERAL-PURPOSE SYSTEM TESTING FRAMEWORK

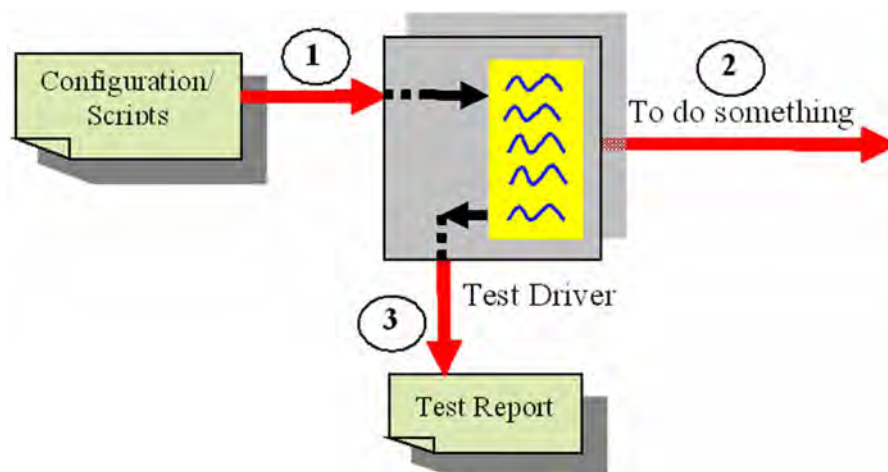
(Kelly, 2003) has highlighted five types of test automation frameworks which were commonly employed:

- **Test Script Modularity Framework:** Small, independent scripts are run in a hierarchical fashion to construct larger tests, realizing a particular test case.
- **Test Library Architecture Framework:** This technique is to divide the application-under-test into procedures and functions library files (e.g APIs, DLLs) that represents modules, sections, and functions of

the application-under-test. These library files are then called directly from the test case script.

- **Keyword-Driven or Table-Driven Testing Framework:** The type of framework develops data tables and keywords, which are independent of the test automation tool, to execute them and the test script code that “drives” the application-under-test and the data. In a keyword-driven test, the functionality of the application-under-test is documented in a table as well as in step-by-step instructions for each test.
- **Data-Driven Testing Framework:** In this type of framework, test input and output values are read from data files (such as, datapools or cvs files) and are loaded into variables in captured or manually coded scripts. Variables are used for both input values and output verification values. Navigation through the program, reading of the data files, and logging of test status and information are all coded in the test script.
- **Hybrid Test Automation Framework:** This is the most commonly implemented framework utilizing a combination of all of the above techniques.

Figure 3. A general purpose performance testing framework



We are now going to introduce a general-purpose system testing framework developed by us. An abstraction of our framework is shown in Figure 3.

This framework consists of three entities, (1) configuration and scripts, (2) test driver and (3) test results. They are interacted in the following manner:

- The test driver requires one or more configurations and scripts as inputs. The configuration file is used to control the setup of the test environment, specifying parameters such as the number of client threads to be used, the duration of the test run, and the payload types. The script is used to specify how to run a test in either product-specific languages (Rational RobotJ) or general-purpose languages (C/C++, Java, Perl, Python, Ruby etc.);
- According to the inputs, the test driver creates a specified amount of loads (threads and processes) to simulate concurrent clients or requests to stress the tested application, such as a web server. The target application under test will receive the test

request messages, performs the request and returns the result to the test driver.

- After all the threads and process finish their testing, the test driver will collect individual test results and combine these results into a final (usually summarized) test report.

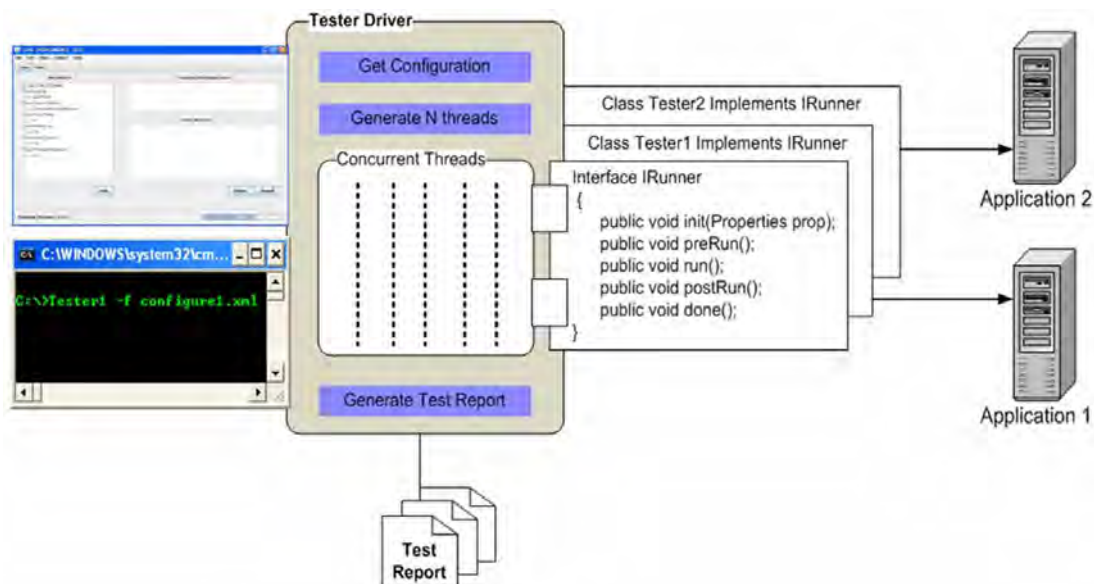
ARCHITECTURE FOR SINGLE PROCESS/MACHINE PERFORMANCE TEST

Based on the above descriptions, we developed a general-purpose performance testing framework for single process performance tests. Figure 4 shows the overall architecture and its basic components and interfaces.

This framework uses a common interface, called *IRunner*, to separate the application logics from the other components. The interface captures the common requirements for testing a variety of applications/technologies, described as follows:

- ***init(app)***: Each thread may need to make preparations by calling *init()* before a test,

Figure 4. Overall architecture of a general-purpose system performance testing framework



to perform actions such as setting up a connection to the server, initialize test data, and clean and populate databases. The argument (app) is a set of application specific properties input by the tester. *init()* is run only once for each test.

- **run():** This method is repeatedly called by each thread during testing to complete a specific business operation, such as invoking a web service function, updating database, etc. This interface gives full control and flexibility to the tester to define and program what to test and how to test.
- **preRun()/postRun():** These methods are called, respectively, to make preparations and/or clean-up before/after each test iteration, i.e. each invocation of the above *run()*. They can also be used to simulate application behaviors, such as thinking and typing time by sleeping for a certain time.
- **done():** This method is called after the completion of all test runs to release some resources (connection to DB/web server) and to export the local summary results to the global results. This method is exclusively executed to ensure the consistency of the global results. Both Java and .NET have native support of the exclusive execution by marking the method as ‘synchronous’.

To test a specific application, the test engineer simply implements all or some of the above interfaces by programming the specific application logic in the corresponding language. This makes the framework capable of testing a wide variety of technologies and applications. Figure 4 illustrates two implementations of *IRunner* (*Tester1* and *Tester2*) to test two different applications, i.e. Application 1 and Application 2. The C# code shown in Box 1 is incomplete which shows the skeleton of each client thread’s run method only.

Another architectural decision of our testing framework is to support two user interfaces, i.e. Graphics User Interface (GUI) and Command

Box 1.

```
public void run()
{
    // instance the common interface IRunner
    IRunner runner = ...
    // before test: prepare test
    runner.init(this.config);
    // during test
    for(alarm.start(); !alarm.isTime(); numTx++) {
        runner.preRun();
        timer.start();
        runner.run(); // To run one iteration of your test
        latency = timer.stop();
        if(alarm.isTesting()) localLog.add(latency);
        runner.postRun();
    }
    // after test: add the local log to the global log
    runner.done();
}
```

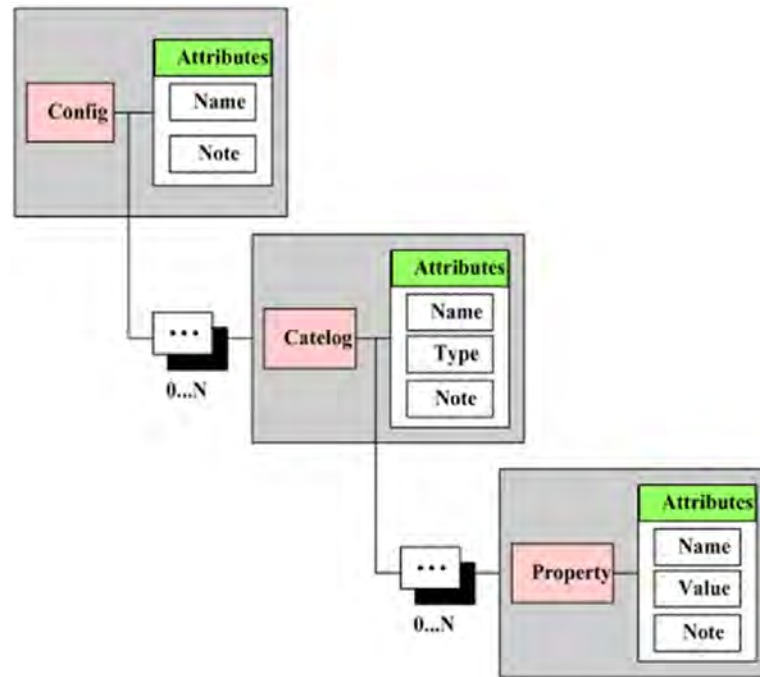
Line (CMD), as shown in Figure 4. As a result, a uniform configuration is designed to be shared by the two user interfaces.

Single Uniform Configuration

Our system uses a single XML file for all configuration parameters of each test, including the general testing parameters (such as number of concurrent clients, running time and result logging file names) and application-specific data. This will simplify the testing configuration management effort. We carefully defined a flexible configuration schema in order to cover most test cases in terms of different applications and technologies. Figure 5 presents an overview of the schema (data model). Our configuration data model has the following features:

- Support for multiple catalogs that can be used to group a set of parameters (properties) for a certain purpose.
- Support any number of properties of each catalog for extension.
- Instead of defining strong schema tag names, we only define a structure of the configuration data model and then pre-define a set of properties ‘keywords’ used for

Figure 5. Schema of the 'universal' configuration



a specific version of the framework implementations. This makes the configuration always extendable without the need for frequently changing the XML schema code.

Box 2 shows an example of testing configuration based on the above schema.

In Box 2, all the property names in the catalog named 'Control' (type='test') are predefined as reserved keywords for all performance tests. These

property values are used by the common test driver to conduct general-purpose performance test tasks, such as how many concurrent thread required to create, how long each thread to run, etc. The C# code in Box 3 shows how the test driver use the parameter 'numClients' to generate the corresponding amount of concurrent threads.

On the other hand, the catalog named 'TPCC' (type='app') can contain any properties used by a specific application, such as TPCC in this case.

Box 2.

```
<Config name="Test1" note="TPCC testing">
  <Catalog name="control" type="test">
    <Property name="numClients" value="100" />
    <Property name="testTime" value="80" note="sec" />
    <Property name="warmup" value="0.1"
      note="10% of testTime" />
    <Property name="cooldown" value="0.1"
      note="10% of testTime" />
  </Catalog>
  <Catalog name="TPCC" type="app">
    <Property name="dbString" value="....." />
    <Property name="numWare" value="10" />
  </Catalog>
</Config>
```

Box 3.

```
public void test()
{
  // create and start multiple threads/clients
  Thread[] client = new Thread[numClients];
  for(int i=0; i<numClients; i++) {
    client[i] = new Thread(new ThreadStart(this.run));
    client[i].Start();
  }
  // To wait all threads to finish
  for(int i=0; i<numClients; i++) client[i].Join();
}
```


Architecture for Distributed Testing

To support large-scale load testing for enterprise systems, multiple testing machines must be used to generate a large amount of concurrent requests to stress enterprise applications. Our framework provides a reference architecture for distributed performance testing by leveraging the above single process testing architecture via loosely coupling messages as shown in Figure 6.

An explanation of the architecture for distributed testing environment is provided as follows:

- The above described single test driver is deployed and run on each testing machine, i.e. Tester 1, Tester 2, ... Tester N.
- The configuration used for running the test instructs the driver to receive further test control information from the Test Manager by subscribing the same or different specific topics (or messaging queue) depending on the Tester's role.
- The Test Manager will control these test drivers by sending the corresponding topic

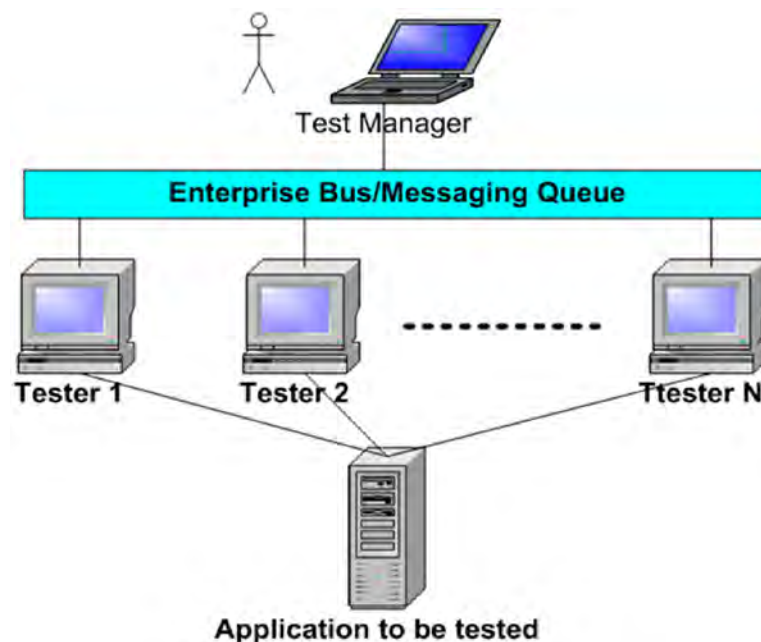
or queue configuration XML via the enterprise bus.

- The Testers who subscribes to the topic will receive the configuration XML and start a test run according to the configuration in exactly the same way as single process/machine testing described in the previous section.
- After the test run, each test driver will send its test report to a topic (message queue) as specified in the configuration file.

Our framework provides a number of test harness designs to support the common functionalities of performance testing:

- **Configuration:** This component hosts all configuration data;
- **Alarm:** This component controls the whole test, such as when the test completes and if to log the test results;
- **High-Resolution Timer:** This component measures the latency and response time with higher resolution (3.27E-07 msec on Windows);

Figure 6. The architecture of a distributed general-purpose system performance testing framework



Box 4.

```

public class ResultLog
{
    // private members
    string name;
    int counter; // Num. of samples
    double sumValue; // Sum of sample values
    double maxValue = 0.0;
    double minValue = 999999999.9;
    int[] valueDistribution;
    .....

    // public methods
    public void add (double sample); // add a sample
    public void add (ResultLog log); // add another log
    public double getAverage (); // return the mean
    public double getMax (); // return maxValue
    public double getMin (); // return minValue
    ...
}

```

- **Random Generator:** It will generate random testing data (number and string) at runtime;
- **Result Log:** This component hosts all test results for one measurement.

For illustration purpose, we provide the design for the **ResultLog** as follows in C# in Box 4.

The C# code in Box 4 shows the result log includes all the performance metrics defined in previous section. Note that instead of hosting the average directly, the log contains only the sum of each measurement and the total number of sampling so that it can derive the mean (average) of the measurement using one after test. To log the distribution of each measurement, we use an array of integers (*valueDistribution*) to host the numbers of samplings fallen into individual elements of *valueDistribution* with the code shown in Box 5.

This framework has been implemented on both Java and .NET platforms. These implementations have been validated by a number of academic research projects; able to support a variety of large-scale performance tests via the extendable configuration (Chen, Ng, & Greenfield, 2013; Chen, Yan, Zic, Liu, & Ng, 2006; Ng, 2007; Ng, Chen, & Greenfield, 2004; Yao et al., 2012).

Box 5.

```

public void add(double value)
{
    sumValue += value;
    counter++;
    if(value > maxValue) maxValue = value;
    if(value < minValue) minValue = value;
    int index = (int)(value/disUnit);
    valueDistribution[index]++;
}

```

disUnit is the unit value (scale) for the element (bucket) of *valueDistribution*.

FUTURE RESEARCH DIRECTIONS

Recently, we have seen the shift from the traditional focus in the application quality management (i.e. load/stress and functional/regression testing) towards greater emphasis on productivity (i.e. agile development such as the Scrum framework (www.scrumalliance.org)) to improve time to market. Apart from that, packages, SaaS and business process management are driving greater business analyst involvement in overall quality efforts, along with a shift in focus from finding defects in validation to meeting business objectives.

In summary, we consider the following four areas are the major driving forces in shaping the development of future system performance testing frameworks:

- **Cloud:** The cloud has already reshaped the load/stress testing (e.g., SOASTA, Neotys, Neustar and Keynote) and for developing solutions for simplifying cross-browser testing (e.g., CrossBrowserTesting.com and BrowserStack). However, there is still plenty of work required to improve the overall solutions for cohesive testing heterogeneous environments in the cloud.
- **Mobile:** The Internet-of-Things market will reshape the landscape for testing. We stress that there is a need for either products or technology to form partnerships due to the fact that mobile devices have added a number of new complexities and challenges to the automation of tests.

- **Agile Development:** The growth of agile development practices continues pushing traditional software vendors to enhance their framework to enable smarter, more efficient tests and a strong partnership between developers and testers.
- **Open Source:** The number of open source testing tools will continue to grow (including those from Appvance and Sauce Labs). The Selenium WebDriver in particular will have significant impact on the functional testing of Web applications.

CONCLUSION

We have presented an overview of some of the contemporary system performance testing frameworks. We observed that most of the commercially available system performance testing frameworks have been designed with a strong view of tightly linking business, testing, and development; that integrates the test tool with developers IDEs. Most of the market leaders have broad portfolios covering test management, and functional and performance test automation. All the frameworks are capable to create, manage and execute test automation on the Windows platform, to test Web applications, and with some frameworks focusing on the cloud and mobile applications.

Furthermore, we have demonstrated our general-purpose performance testing framework is simple to implement yet versatile to test the performance of complicated and large-scale enterprise systems.

REFERENCES

- Apache jmeter. (2015). Retrieved 10 March, 2015, from <http://jmeter.apache.org/>
- Chen, S., Ng, A., & Greenfield, P. (2013). A performance evaluation of distributed database architectures. *Concurrency and Computation*, 25(11), 1524–1546. doi:10.1002/cpe.2891
- Chen, S., Yan, B., Zic, J., Liu, R., & Ng, A. (2006). *Evaluation and modelling of web services performance*. Paper presented at the 2006 IEEE International Conference on Web Services (ICWS2006), Chicago, IL. doi:10.1109/ICWS.2006.59
- Driver, M., Murphy, T. E., Wilson, N., Valdes, R., Norton, D., & Sobejana, M. (2014). *Magic quadrant for integrated software quality suites*. Gartner, Inc.
- HP loadRunner. (2015). Retrieved 10 March, 2015, from <http://www8.hp.com/au/en/software-solutions/loadrunner-load-testing/index.html>
- Illes, T., Herrmann, A., Paech, B., & Rückert, J. (2005). *Criteria for software testing tool evaluation. a task oriented view*. Paper presented at the 3rd World Congress for Software Quality.
- ISO/IEC. (2001). ISO/IEC 9126 standard. *Software engineering -- product quality -- part 1: quality model*. ISO.
- Jain, R. (1991). *The art of computer systems performance analysis: techniques for experimental design, measurement, simulation, and modeling*. Wiley.
- Jain, R. (2015). *Selection of techniques and metrics*. Retrieved from Lecture note of CSE567M: computer systems analysis (spring 2015) website: http://www.cse.wustl.edu/~jain/cse567-15/k_03tam.htm
- Kelly, M. (2003). Choosing a test automation framework. *IBM DeveloperWorks*, 18.
- Mansharamani, R. (2011). *Performance requirements gathering & analysis*. Retrieved from <http://www.softwareperformanceengineering.com/uploads/1/2/6/6/12667295/performancerequirementsgatheringandanalysis.pdf>
- Meier, J. D., Farre, C., Bansode, P., Barber, S., & Rea, D. (2007). *Performance testing guidance for web applications*. Microsoft Corporation.

Menascé, D. A. (2002). Load testing of web sites. *IEEE Internet Computing*, 6(4), 70–74. doi:10.1109/MIC.2002.1020328

Microsoft visual studio testing tools. (2015). Retrieved 10 March, 2015, from <https://www.visualstudio.com/explore/testing-tools-vs>

Ng, A. (2007). An evaluation of the performance of web services using table-driven XML. *International Journal of Business Process Integration and Management*, 2(3), 15. doi:10.1504/IJBPM.2007.015493

Ng, A., Chen, S., & Greenfield, P. (2004). *An evaluation of contemporary commercial soap implementations*. Paper presented at the Fifth Australian Workshop on Software and System Architectures (AWSA2004), Melbourne, Australia.

Performance testing products. (2015). Retrieved 19 March, 2015, from <http://www.revolutionit.com.au/products/products-by-software-type/performance-testing>

Pressman, R. S. (1992). *Software engineering: a practitioner's approach* (3rd ed.). New York: McGraw-Hill.

Rankin, C. (2002). The software testing automation framework. *IBM Systems Journal*, 41(1), 126–139. Retrieved from <http://www.research.ibm.com/journal/sj/411/rankin.html> doi:10.1147/sj.411.0126

Sarojadevi, H. (2011). Performance testing: Methodologies and tools. *Journal of Information Engineering and Applications*, 1(5), 5–13.

Siva, V., & Sharma, S. (2006). Should performance be an integral part of the SDLC? *SETLabs briefings*. Retrieved from <http://www.infosys.com/technology/performance-an-integral-part-of-SDLC.pdf>

SOASTA. (2014). The top 10 tips to improve web application performance. Author.

The grinder, a Java load testing framework. (2015). Retrieved 10 March, 2015, from <http://grinder.sourceforge.net/>

Yao, J., Ng, A., Chen, S., Liu, D., Friedrich, C., & Nepal, S. (2012). *A Performance Evaluation of Public Cloud Using TPC-C*. Paper presented at the 2012 International Workshop on Analytics Services on the Cloud, Shanghai, China.

Yigitbasi, N., Iosup, A., Epema, D., & Ostermann, S. (2009). *C-meter: A framework for performance analysis of computing clouds*. Paper presented at the 2009 9th IEEE/ACM International Symposium on Cluster Computing and the Grid. doi:10.1109/CCGRID.2009.40

KEY TERMS AND DEFINITIONS

Bandwidth: The amount of information that can be transmitted in a fixed amount of time over a particular media channel.

CPU Utilization: Determines the percentage of time the processor is busy by measuring the percentage of time the thread of the idle process is running and then subtracting that from 100 percent.

Disk I/O Utilization: The amount of input and output operations on the secondary storage at a certain unit of time.

Latency: The total time spent by a system generated message to travel from its sender (source) to its receiver (destination).

Main Memory Utilization: The amount of RAM used by a particular system at a certain unit of time.

Response Time: The total time taken by a user to wait in invoking a request to a particular system and coming back with a result.

Scalability: A measure of the capability of a system to increase its total output under an increased load when resources (typically hardware) are added.

Speedup: A measure of the relative performance improvement when executing a task. Speedup can be defined by studying the different types of performance metrics such as: throughput and latency.

System Performance Testing: A special type of technical investigation done to determine or

validate the speed, scalability, and/or stability characteristics of the system under test.

Throughput: The capability of a system in handling a certain number of client requests (such as messages or transactions) within a unit of time (such as second or minute).

A Tale of Two Agile Requirements Engineering Practices

S**Pankaj Kamthan***Concordia University, Canada***Terrill Fancott***Concordia University, Canada*

INTRODUCTION

The discipline of software engineering advocates a systematic and disciplined approach towards the development and evolution of software systems. The ecosystem of software engineering is changing constantly, and influences the practice of software engineering. In the past decade, there have been a number of notable strategic changes in industrial software engineering, including the adoption of agile methodologies (Highsmith, 2009). If industrial surveys are any sign, this trend is likely to continue for the foreseeable future.

There are a number of concerns in agile software development, including requirements engineering (Sillitti & Succi, 2005; Zhu, 2009; Leffingwell, 2011). There are different types of agile requirement, of which currently the most common forms are use cases (Jacobson et al., 1992) and user stories (Cohn, 2004). The use cases and user stories have different origins, both in space and in time, but, by being among the practices of scenario-oriented requirements engineering (SORE) (Alexander, 2011), they are not entirely unrelated. The purpose of this article is to situate use cases and user stories in context of each other.

The need for a comparison among concepts arises naturally in many disciplines, and agile requirements engineering is no exception. Indeed, it is suggested by theories of learning, such as constructivism (Piaget, 1952), that upon initial exposure to a new concept C, a comparison between C and other, closely related and relatively

more established, concepts C' is inevitable. This comparison is a necessary prerequisite for creating an understanding of C through assimilation and accommodation, selection, and application. In this article, C is user story and C' is use case.

The rest of the article is organized as follows. First, background and related work are presented. This is followed by introduction of a conceptual framework for systematically comparing use cases and user stories. Next, directions for future research are highlighted. Finally, concluding remarks are given.

BACKGROUND

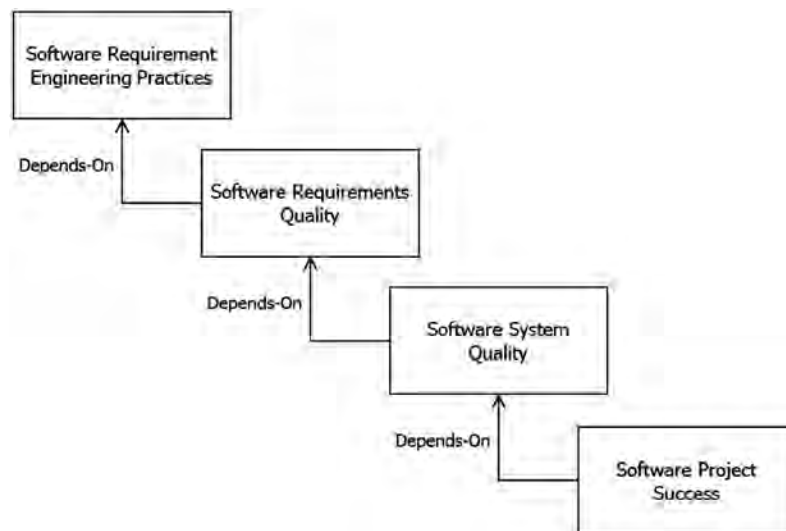
In this section, a brief background on agile methodologies and requirements engineering is presented, and previous work on comparing use cases and user stories is analyzed.

Agile Methodologies

In the 1990s, a number of limitations of rigidity in approaches for development of certain types of software systems were realized. These limitations made it difficult for organizations that aimed for relatively short release cycles of their products to remain competitive in the market, had relatively small budgets to allocate for their projects, and required richer communication between technical and non-technical stakeholders (including customers and users) to address their needs better.

DOI: 10.4018/978-1-5225-2255-3.ch659

Figure 1. A causal relationship between the success of a software project and the practices of software requirements engineering



The drive to cope with these limitations led to the inception of agility.

The *Agile Manifesto* constitutes the basis for a number of agile methodologies, including Agile Experience Design (AXD), Crystal Clear, Extreme Programming (XP), OpenUP, Scrum, and User-Centered Agile Process (UCAP). The agile methodologies have evolved over time and in certain cases, such as the Discipline Agile Delivery (DAD) process framework (Ambler & Lines, 2012), elements of multiple agile methodologies have been included.

Requirements Engineering

In software engineering, significant emphasis is placed on the requirements engineering, in general, and practices of requirements engineering, in particular, in the light of their impact on the later phases of software development and eventually their influence on the success of the software project (Wiegiers, 2003). This causality is depicted in Figure 1. In requirements engineering, the attention is on the problem so as to devise a desirable and viable solution, namely the software system, which can satisfy the stakeholders.

The agile methodology underlying an agile project determines the type of agile requirement to be adopted. For example, in Crystal Clear and OpenUP, an agile requirement can take the form of a use case (Jacobson et al., 1992) and in AXD, XP, Scrum, and UCAP, an agile requirement can take the form of a user story (Cohn, 2004).

Related Work

The need for understanding the differences between use cases and user stories has been, directly or indirectly, expressed in several places in literature. In the following, previous efforts that compare use cases and user stories are analyzed briefly and chronologically.

The notion of simplicity has been a criterion for comparison between use cases and user stories. In one of the earliest work on user stories (Beck, 2000), it has been pointed out that user stories are “simplified use cases.” However, the meaning of ‘simplification’ and exactly what is “simplified” has not been given.

The structure of the means used for a description has often been a criterion for comparison between use cases and user stories. Indeed, it has

been stated that use cases are “structured” while user stories (in XP) are “unstructured” (Cohn & Paul, 2001), that “use cases are a structured representation of a user story” (Decker et al., 2006), and that “[use cases] are expressed using a constrained (semi-formal) syntax” and that “[user stories] are expressed using natural language prose” (Alexander & Maiden, 2004). However, there are a number of ways of expressing a use case (Cockburn, 2001), and not all of them need to be “structured” or follow a “constrained (semi-formal)” syntax.

The issue of a software requirement treading into software design has been a criterion for comparison between use cases and user stories. For example, it has been asserted that “use cases are more prone to including details of the user interface” (Cohn, 2004) and that “it is usual for use cases to include user interface details” (Monochristou & Vlachopoulou, 2007). However, there are guidelines (Cockburn, 2001; Wiegers, 2003) that explicitly suggest against the inclusion of any user interface specifics in use cases.

It has been concluded, based on a controlled experiment, that use cases can be useful as a complement to user stories (Gallardo-Valencia, Olivera, & Sim, 2007).

Finally, it has been pointed out that there are trade-offs in committing to both use cases and user stories, and that each approaches testing differently (Wiegers & Beatty, 2013).

Table 1. A framework for comparing use Cases and user stories

Discipline	Viewpoint	Criterion
Software Engineering	Project	• Estimate • Schedule
	Process	• Methodology • Development
	People	• Users • Value
	Product	• Scope • Description
	Resource	• Purpose • Maturity

The issues underlying related work can be summarized as follows. The comparison between use cases and user stories is treated as a secondary, not a primary, concern; the comparison is not systematic; and the instrument for comparison has a singular perspective. This article is motivated, in part, by the need to address these issues. To do that, a systematic approach for comparison that rests on the foundations of software engineering is required.

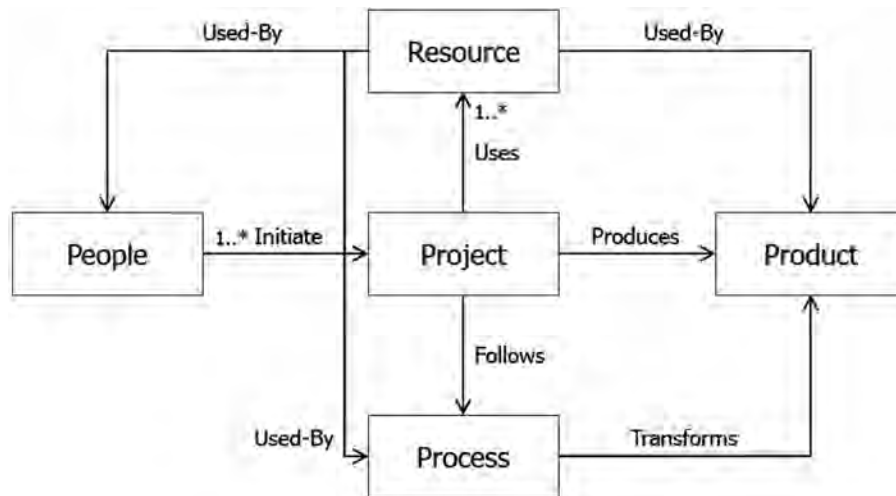
A CONCEPTUAL FRAMEWORK FOR COMPARING USE CASES AND USER STORIES

In this section, use cases and user stories are compared systematically. The purpose of comparison is to facilitate understanding and appropriate use of each.

The comparison is based on a conceptual framework, as given in *Table 1*, consisting of a set of meta-criteria that are decomposed into a set of criteria. The criteria are selected in a manner that they are relevant to software engineering, as well as both to use cases and to user stories. The criteria are considered equally significant.

The conceptual framework is based on the software engineering principle of *separation of concerns* (Ghezzi, Jazayeri, & Mandrioli, 2003). The meta-criteria consist of certain viewpoints considered relevant to software engineering by the *IEEE Software and Systems Engineering Standards Committee*. Figure 2 shows these viewpoints and the interrelationships among them, the dynamics of which can be explained as follows. *People* initiate a *Project* that uses one or more *Resources* to follow a *Process* that produces a *Product* by transforming it (iteratively and incrementally) over space and time. This *Product* could be an intermediate artifact of the software system in question, or the system itself, which evolves from inception to conclusion. For example, a low-fidelity design prototype could be such a *Product*.

Figure 2. The viewpoints and their interrelationships in the conceptual framework



The rest of the section provides details of comparison between use cases and user stories, organized by viewpoints and criteria, highlighting the similarities and differences between them in the process.

Use Cases vs. User Stories: Project-Viewpoint

A comparison of use cases and user stories from a project viewpoint can be based on the following criteria: estimate and schedule.

Criterion: Estimate

Use Cases: In light of their availability early in a software development process, use cases have been recommended as a basis for estimation (Karner, 1993; Mohagheghi, Anda, & Conradi, 2005). The purpose of estimation is to schedule the entire software project. This approach requires the presence of all use cases upfront, and the fact that they all are expressed in a certain manner, for calculating the estimate. However, a use case is not intrinsically related to estima-

tion. Furthermore, software development methodologies that currently support use cases, do not use the estimates based on use cases for planning.

User Stories: A user story must be associated with an estimate. The estimation must be part of a user story development process. In other words, a user story description without an estimate is incomplete. A user story is estimated individually. In both XP and Scrum, the purpose of estimating user stories is short-term planning. In particular, the user story estimate is used for iteration planning and release planning.

Criterion: Schedule

Use Cases: The completion (design and implementation) of a use case is not explicitly time-bound. A single use case could be completed over a number of iterations.

User Stories: The completion of a user story is time-bound. A single user story must be completed in a single iteration. Indeed, if a user story is too large to fit in a single iteration, it must be split into multiple user stories.

Use Cases vs. User Stories: Process-Viewpoint

A comparison of use cases and user stories from a process viewpoint can be based on the following criteria: methodology and development.

Criterion: Methodology

Use Cases: The use cases are originally associated with Objectory. There is explicit support for use cases in Unified Process and its customizations, in ICONIX Process, and in Crystal Clear. In other words, use cases were a part of rigid methodologies initially, and over time became a part of certain agile methodologies. The use case modeling approach itself is iterative, incremental, and interspersed.

User Stories: The user stories are originally associated with XP. There is explicit support for user stories in other agile methodologies such as Scrum, and its extensions such as UCAP and U-SCRUM. The user story creation approach itself is iterative, incremental, and interspersed.

Criterion: Development

Use Cases: The development of a use case model can be based solely on introspection (say, via speculation or self-reference).

User Stories: The development of user stories relies explicitly on extrospection. The social aspect of the development is exhibited in form of proximity (such as, during meetings and interviews), communication, and collaboration, and is considered necessary for elaboration and clarification of user stories. This situation, however, is changing as *crowdsourcing* becomes an accepted means of eliciting users' needs.

Use Cases vs. User Stories: People-Viewpoint

S

A comparison of use cases and user stories from a people viewpoint can be based on the following criteria: users and value.

Criterion: Users

Use Cases: In use case modeling, the users form a subset of human actors. The users elicited during use case modeling can be fictitious or surrogate users (Alexander & Beus-Dukic, 2009). These users may be based on introspection, not necessarily on actual communication with the real users.

User Stories: There must be an actual person behind a (user) role in a given user story. Indeed, a user story is initially expressed for a concrete person, such as a candidate for a *persona*, and is subsequently generalized to a user story for an abstract (user) role.

Criterion: Value

Use Cases: The value provided by a use case to an actor is implicit. This value may or may not be highlighted in a use case description.

User Stories: The value provided by a user story to its (user) role must be explicit. This value serves as a rationale for the user story. The type of value, however, is not specified in the user story definition, and, indeed, may vary across projects and its stakeholders.

Use Cases vs. User Stories: Product-Viewpoint

A comparison of use cases and user stories from a product viewpoint can be based on the following criteria: scope and description.

Criterion: Scope

Use Cases: In general, a use case is larger in scope compared to a user story. An *epic* is a ‘candidate’ user story so large that it can not be completed in a single iteration, and therefore must be split (Cohn, 2004). The scope of a use case is comparable to that of an epic (Collier, 2012).

User Stories: In general, the scope of a user story is comparable to that of a use case ‘normal’ scenario.

Criterion: Description

Use Cases: A use case can be described at multiple levels of details (Cockburn, 2001). The description can be expressed in different modalities, namely text and graphics. A comprehensive textual description of a use case can have several elements, some of which are about the context of use (such as, business rules governing the use case, trigger that initiates the use case, and frequency, pre- and post-conditions of interaction). It is expected that technical terminology is avoided in the description to make it amenable to non-technical stakeholders. However, doing so is unavoidable in a use case model expressed in any of the diagram types of the Unified Modeling Language (UML). The exceptional cases are dealt with in recovery and failure scenarios. The possible types of relationships among use cases are known (Adolph et al., 2003). The description of a use case is not intrinsically related to the existence of test cases. Indeed, the test cases are separated from use cases and, for example, may appear in a test plan, or otherwise in some form of test documentation. A use case description is a relatively permanent artifact that persists throughout the development and maintenance of the software system.

User Stories: A user story has single level of description. The description has a single modality, namely text. A user story description is usually not comprehensive, and contextual information, if any, is minimal and external to the description. It is expected that technical terminology is avoided in the description to make it amenable to non-technical stakeholders. The exceptional cases are dealt with in acceptance tests, not in the user story statement. There is currently no clearly established understanding of the types of relationships among user stories, although this situation is improving (Gomez, Rueda, & Alarcón, 2010). A user story must be associated with acceptance criteria. The acceptance criteria must be part of a user story development process. In other words, a user story description without acceptance criteria is incomplete. The tests are part of the description of a user story. Traditionally, index cards have been used as a medium for disseminating user stories. The tests, by being on the same index card as the user story, are intimately related to that user story. This is in agreement with the ‘T’ of INVEST (Wake, 2002) and INSERT (Patel & Ramachandran, 2009), where ‘T’ expands to Testability, one of the desired quality attributes of a user story. A user story description, at least in the classical sense (Beck, 2000), is intended to be transient: its life is limited to the iteration for which it is formulated. This is in accord with Travel Light, one of the principles of XP. This principle has also been adopted by Agile Modeling (Ambler, 2002). Therefore, user stories need not be archived for long-term use. Indeed, index cards used as a medium for user stories are usually discarded after use (Langr & Ottinger, 2011), although this practice is changing with the increasing use of agile methodologies for geographically dispersed teams, rise of agile project management systems, and needs of maintainability.

Use Cases vs. User Stories: Resource-Viewpoint

A comparison of use cases and user stories from a resource viewpoint can be based on the following criteria: purpose and maturity.

Criterion: Purpose

Use Cases: A use case is a model of the external behavior resulting from the interaction between an actor and the system.

User Stories: A user story is a statement of the desired functionality, from the perspective of a user. This desired functionality is aimed for user's benefit, but it may not be related to user's interaction with the system.

Criterion: Maturity

Use Cases: The notion of a use case was introduced in the middle of 1980s. There is established, experiential body of knowledge, for authoring 'high-quality' use cases, in form of guidelines (Cockburn, 2001), patterns (Adolph et al., 2003; Issa, Odeh, & Coward, 2006; Övergaard & Palmkvist, 2005), and anti-patterns (El-Attar, 2009; El-Attar & Miller, 2012). The relationship of use cases to concepts pertaining to other software artifacts, such as software architecture views, is established. It is common for use cases to be included in requirements engineering pedagogy, as suggested by recent textbooks on software engineering as well as on requirements engineering.

User Stories: The notion of a user story was introduced in the late 1990s. There is growing, yet undeveloped, experiential body of knowledge, for authoring 'high-quality' user stories, in form of 'smells' (Cohn, 2004), guidelines (Wallace, Raggett, & Aufgang, 2002; Cohn, 2004; Patel & Ramachandran, 2009), patterns (Leffingwell, 2011), and standards (ISO/IEC/IEEE, 2012). The

notion of a user story continues to evolve (Moreno & Yagüe, 2012). The possible types of stakeholders of a user story also continues to become granular and grow according to the needs of agile projects in different domains (Hochmüller, 2011; Read, Arreola, & Briggs, 2011). The inclusion of user stories in requirements engineering pedagogy is not common, although there are initial efforts in that direction (Fancott, Kamthan, & Shahmir, 2011).

The Framework in Context

The framework for comparison of use cases and user stories presented earlier uses a particular set of viewpoints. It could be noted that using a different set of viewpoints, or changing the framework itself, can lead to a comparison that has a possibly different outcome. For example, a business model, a cost model, or an organizational capability maturity model could serve as a basis for constructing alternative frameworks.

FUTURE RESEARCH DIRECTIONS

There are a few directions that emanate from the work presented in this article. These avenues are discussed briefly in the following.

The framework for comparison of use cases and user stories presented in this article has industrial implications. For example, the framework could be used as a basis for designing and conducting empirical studies (say, through surveys and/or interviews) of requirements engineering practices in organizations, including those organizations that are preparing for a transition from rigid to agile methodologies.

The framework also has pedagogical implications. For example, the framework could be used in a course on requirements engineering. This could be followed by feedback from students of their preference, if at all, of one type of agile requirement over the other (Fancott, Kamthan,

& Shahmir, 2011). The results obtained from an analysis of the feedback may help improve requirements engineering education, and is therefore of research interest.

The comparison of use cases and user stories carried out in this article is from the perspective of user stories. There are other entities in the requirements engineering body of knowledge that can also serve as possible candidates for a comparison with user stories. For example, such entities include 'conventional' software requirements defined by international standards, such as ISO/IEC/IEEE 29148 (ISO/IEC/IEEE, 2011). Indeed, such a comparison is important as agile methodologies continue to seek acceptance in 'conventional' software engineering contexts, as indicated by the latest version of the *Guide to the Business Analysis Body of Knowledge* (BABOK), the *Guide to the Project Management Body of Knowledge* (PMBOK), the *Guide to the Software Engineering Body of Knowledge* (SWE-BOK), and the *Kernel and Language for Software Engineering Methods* (Essence). Therefore, placing user stories in context of 'conventional' software requirements in a systematic manner, such as by using a framework for comparison as an instrument, is of research interest. It is evident that such framework for comparison will be different from that presented earlier, although certain elements of *Table 1* are likely to remain relevant.

CONCLUSION

The significance of requirements engineering in agile software development, and increasing to commitment to agile methodologies in the industry, necessitates a comparison of use cases and user stories. It is evident that an investment in software project artifacts, regardless of their type, entails costs. For example, such cost may incur from training the agile team, or from purchasing a management system. Therefore, or even otherwise, say, for the sake of requirements engineering education, it is important that such a comparison remain objective and accurate.

The framework for comparison presented in this article highlights both similarities and differences between use cases and user stories. From the comparison between use cases and user stories based on this framework, the following conclusion can be drawn. The use cases and user stories have different, albeit overlapping, goals, and different means of achieving those goals. Furthermore, they can co-exist and support each other. Indeed, it can be expected that the body of knowledge on use cases will continue to serve as a reference point for the development of the body of knowledge on user stories, thereby contributing to the advancement of agile requirements engineering.

REFERENCES

- Adolph, S., Bramble, P., Cockburn, A., & Pols, A. (2003). *Patterns for Effective Use Cases*. Addison-Wesley.
- Alexander, I. (2011). GORE, SORE, or What. *IEEE Software*, 28(1), 8–10. doi:10.1109/MS.2011.7
- Alexander, I., & Beus-Dukic, L. (2009). *Discovering Requirements: How to Specify Products and Services*. John Wiley and Sons.
- Alexander, I., & Maiden, N. (2004). *Scenarios, Stories, Use Cases through the Systems Development Life-Cycle*. John Wiley and Sons.
- Ambler, S. W. (2002). *Agile Modeling: Effective Practices for Extreme Programming and the Unified Process*. John Wiley and Sons.
- Ambler, S. W., & Lines, M. (2012). *Disciplined Agile Delivery: A Practitioner's Guide to Agile Software Delivery in the Enterprise*. IBM Press.
- Beck, K. (2000). *Extreme Programming Explained: Embrace Change*. Addison-Wesley.
- Cockburn, A. (2001). *Writing Effective Use Cases*. Addison-Wesley.

- Cohn, M. (2004). *User Stories Applied: For Agile Software Development*. Addison-Wesley.
- Cohn, T., & Paul, R. (2001). A Comparison of Requirements Engineering in Extreme Programming (XP) and Conventional Software Development Methodologies. *Proceedings of The Seventh Americas Conference on Information Systems (AMCIS 2001)*.
- Collier, K. W. (2012). *Agile Analytics: A Value-Driven Approach to Business Intelligence and Data Warehousing*. Addison-Wesley.
- Decker, B., Ras, E., Rech, J., Klein, B., & Hoecht, C. (2006). Using Wikis to Manage Use Cases: Experience and Outlook. *Proceedings of the Workshop on Learning Software Organizations and Requirements Engineering*.
- El-Attar, M. (2009). *Improving the Quality of Use Case Models and their Utilization in Software Development* (Ph.D. Thesis). The University of Alberta, Edmonton, Canada.
- El-Attar, M., & Miller, J. (2012). Constructing High Quality Use Case Models: A Systematic Review of Current Practices. *Requirements Engineering*, 17(3), 187–201. doi:10.1007/s00766-011-0135-y
- Fancott, T., Kamthan, P., & Shahmir, N. (2011). Using the Social Web for Teaching and Learning User Stories. *Proceedings of The Sixth International Conference on e-Learning (ICEL 2011)*.
- Gallardo-Valencia, R. E., Olivera, V., & Sim, S. E. (2007). Are Use Cases Beneficial for Developers Using Agile Requirements? *Proceedings of The Fifth International Workshop on Comparative Evaluation in Requirements Engineering (CERE 2007)*.
- Ghezzi, C., Jazayeri, M., & Mandrioli, D. (2003). *Fundamentals of Software Engineering* (2nd ed.). Prentice-Hall.
- Gomez, A., Rueda, G., & Alarcón, P. P. (2010). A Systematic and Lightweight Method to Identify Dependencies between User Stories. *Proceedings of The Eleventh International Conference on Agile Processes in Software Engineering and Extreme Programming (XP 2010)*. doi:10.1007/978-3-642-13054-0_17
- Highsmith, J. (2009). *Agile Project Management: Creating Innovative Products*. Addison-Wesley.
- Hochmüller, E. (2011). The Requirements Engineer as a Liaison Officer in Agile Software Development. *Proceedings of The First Workshop on Agile Requirements Engineering (AREW 2011)*. doi:10.1145/2068783.2068785
- ISO/IEC/IEEE. (2011). *ISO/IEC/IEEE 29148:2011. Systems and Software Engineering -- Life Cycle Processes -- Requirements Engineering. The International Organization for Standardization (ISO)/The International Electrotechnical Commission (IEC)/The Institute of Electrical and Electronics Engineers. IEEE*.
- ISO/IEC/IEEE. (2012). *ISO/IEC/IEEE 26515:2012. Systems and Software Engineering -- Developing User Documentation in an Agile Environment. The International Organization for Standardization (ISO)/The International Electrotechnical Commission (IEC)/The Institute of Electrical and Electronics Engineers. IEEE*.
- Issa, A., Odeh, M., & Coward, D. (2006). Using Use Case Patterns to Estimate Reusability in Software Systems. *Information and Software Technology*, 48(9), 836–845. doi:10.1016/j.infsof.2005.10.005
- Jacobson, I., Christerson, M., Jonsson, P., & Övergaard, G. (1992). *Object-Oriented Software Engineering: A Use Case Driven Approach*. Addison-Wesley.

- Karner, G. (1993). *Use Case Points – Resource Estimation for Objectory Projects*. Objective Systems SFAB. Copyright Owner Rational Software.
- Langr, J., & Ottinger, T. (2011). *Agile in a Flash: Speed-Learning Agile Software Development*. Pragmatic Bookshelf.
- Leffingwell, D. (2011). *Agile Software Requirements: Lean Requirements Practices for Teams, Programs, and the Enterprise*. Addison-Wesley.
- Mohagheghi, P., Anda, B., & Conradi, R. (2005). Effort Estimation of Use Cases for Incremental Large-Scale Software Development. *Proceedings of The 2005 International Conference on Software Engineering (ICSE 2005)*. doi:10.1109/ICSE.2005.1553573
- Monochristou, V., & Vlachopoulou, M. (2007). Requirements Specification using User Stories. In I. G. Stamelos & P. Sfetsos (Eds.), *Agile Software Development Quality Assurance* (pp. 71–89). Idea Group. doi:10.4018/978-1-59904-216-9.ch004
- Moreno, A. M., & Yagi e, A. (2012). Agile User Stories Enriched with Usability. *Proceedings of The Thirteenth International Conference on Agile Processes in Software Engineering and Extreme Programming (XP 2012)*.
-   vergaard, G., & Palmkvist, K. (2005). *Use Cases: Patterns and Blueprints*. Addison-Wesley.
- Patel, C., & Ramachandran, M. (2009). Story Card Based Agile Software Development. *International Journal of Hybrid Information Technology*, 2(2), 125–140.
- Piaget, J. (1952). *The Origins of Intelligence in Children*. International University Press. doi:10.1037/11494-000
- Read, A., Arreola, N., & Briggs, R. O. (2011). The Role of the Story Master: A Case Study of the Cognitive Load of Story Management Tasks. *Proceedings of The Forty Fourth Hawaii International Conference on Systems Science (HICSS-44 2011)*. doi:10.1109/HICSS.2011.432
- Sillitti, A., & Succi, G. (2005). Requirements Engineering for Agile Methods. In A. Aurum & C. Wohlin (Eds.), *Engineering and Managing Software Requirements* (pp. 309–326). Springer-Verlag; doi:10.1007/3-540-28244-0_14
- Wake, W. C. (2002). *Extreme Programming Explored*. Addison-Wesley.
- Wallace, D., Raggett, I., & Aufgang, J. (2002). *Extreme Programming for Web Projects*. Addison Wesley.
- Wieggers, K. E. (2003). *Software Requirements* (2nd ed.). Microsoft Press.
- Wieggers, K. E., & Beatty, J. (2013). *Software Requirements* (3rd ed.). Microsoft Press.
- Zhu, Y. (2009). *Requirements Engineering in an Agile Environment* (M.Sc. Thesis). Uppsala University, Uppsala, Sweden.

ADDITIONAL READING

- Boehm, B. W., & Turner, R. (2004). *Balancing Agility and Discipline: A Guide for the Perplexed*. Addison-Wesley.
- Brown, A. W. (2012). *Enterprise Software Delivery: Bringing Agility and Efficiency to the Global Software Supply Chain*. Addison-Wesley.
- Bunse, C., Feldmann, R. L., & D  rr, J. (2004). Agile Methods in Software Engineering Education. In *Proceedings of The Fifth International Conference on Extreme Programming and Agile Processes in Software Engineering (XP 2004)*, Garmisch-Partenkirchen, Germany. doi:10.1007/978-3-540-24853-8_43
- Coplien, J. O., & Harrison, N. B. (2005). *Organizational Patterns of Agile Software Development*. Prentice-Hall.
- Davis, B. (2009). *97 Things Every Project Manager Should Know: Collective Wisdom from the Experts*. O'Reilly Media.

Hazzan, O., Lapidot, T., & Ragonis, N. (2011). *Guide to Teaching Computer Science: An Activity-Based Approach*. Springer-Verlag; doi:10.1007/978-0-85729-443-2

Larman, C. (2003). *Agile and Iterative Development: A Manager's Guide*. Addison-Wesley.

Mahnic, V. (2012). A Capstone Course on Agile Software Development Using Scrum. *IEEE Transactions on Education*, 55(1), 99–106. doi:10.1109/TE.2011.2142311

Rüping, A. (2003). *Agile Documentation: A Pattern Guide to Producing Lightweight Documents for Software Projects*. John Wiley and Sons.

Séguin, N., Tremblay, G., & Bagane, H. (2012). Agile Principles as Software Engineering Principles: An Analysis. In *Proceedings of The Thirteenth International Conference on Agile Processes in Software Engineering and Extreme Programming (XP 2012)*, Malmö, Sweden. doi:10.1007/978-3-642-30350-0_1

KEY TERMS AND DEFINITIONS

Acceptance Criteria: The criteria that a system must satisfy in order to be accepted by a user, customer, or other authorized entity.

Agile Methodology: A software development methodology based on the Agile Manifesto. **Artifact:** A document or a model produced during software development.

Process: A set of interrelated or interacting activities that transforms inputs into outputs for some purpose.

Requirement: A statement which translates or expresses a need and its associated constraints and conditions.

Software Engineering: The application of a systematic, disciplined, quantifiable approach to the development, operation, and maintenance of software; that is, the application of engineering to software.

Stakeholder: An individual, team, organization, or classes thereof, having an interest in a system.

Use Case: A sequence of actions performed by a system, which yields an observable result of value to an actor of the system.

User: An individual who interacts with the software system usually, but not always, with a specific goal.

User Story: A high-level requirement statement that contains minimally sufficient information to produce a reasonable estimate of the effort to implement it.

Towards an Understanding of Performance, Reliability, and Security

Ye Wang

Zhejiang Gongshang University, China

Bo Jiang

Zhejiang Gongshang University, China

Weifeng Pan

Zhejiang Gongshang University, China

INTRODUCTION

According to J. Spohrer, service-oriented systems are “value co-creation configurations of people, technology, internal and external service systems connected by value propositions, and shared information (such as language, laws, measures, models, and so on)” (Spohrer, 2007). A service-oriented system (Espadas et al., 2013) delivers a set of services that are cooperating with each other over the Internet, which are therefore called “web services”. Web services are delivered in a new computing paradigm called “service-oriented computing”. Service-oriented computing paradigm plays an increasingly crucial role in modern world industries. Even in cloud computing, services are irreplaceable elements (Wang & Sun, 2015). In service-oriented systems, software developers create software applications through composing web services. Web service composition helps developers to solve complex problems by combining published basic services and ordering them to meet specific service requirements from user requests (Song & Lee, 2013).

Service-oriented systems offer many benefits to software developers and software users. For example, they allow developers to support many customers with a single version of software and provide services to their customers 24/7. Over the past decade, a large number of service-oriented systems have been developed to aid a wide range

of real-world applications (Wang et al., 2011). For example, E-tourism service systems (see for example: www.visitcalifornia.com) have been used by tourists to plan their trip and book tickets as well as hotels. PayPal (www.paypal.co.uk), an online payment service, has been used by many e-commerce applications worldwide. Many investment banks or brokerages nowadays offer their customers trading services such as order management services, stock trading services, bond trading services and so on, and have been playing an increasingly pivotal role in the financial sector (Yang et al., 2011).

There are a lot of key research issues with regard to service-oriented systems, involving service modeling and specification (Cardoso et al., 2010; De Virgilio, 2010), service selection (Sweeney et al., 2016), web service composition algorithms (Hiel & Weigand, 2010; Leitner et al., 2010), verification of composition correctness (Rai et al., 2015), and handling composition faults (Lemos et al., 2016), etc. Rather than only concentrating on functionality, recent work puts more emphasis on Quality of Services (QoS) properties (Calheiros et al., 2015). With many businesses providing same or similar services, QoS properties have become an important indicator for good services (Zhao et al., 2007). Yet, satisfying QoS requirements in service-oriented systems has proved to be more challenging than satisfying functional requirements, due to the following characteristics (Yang

et al., 2011): First, QoS properties are system-level requirements which cannot be assigned directly to individual system components; instead, they need to be planned at the infrastructure-level as a whole, with their design aspects then entrusted to system components. Second, QoS properties are often interdependent and their realization in a system requires a collective and coordinated behavior of the system components and a system-level design strategy. Third, QoS properties are application-specific and their fulfillment necessarily requires a specific design approach germane to their applications. Finally, QoS properties are not only a design time concern, but also most crucially a runtime concern. Satisfying QoS properties means designing runtime mechanisms that can maintain the system's QoS properties throughout the execution time.

Through the investigation of several service-oriented systems in different sectors (Wang & Sun, 2015; Wang et al., 2011; Yang et al., 2011), this chapter has identified a core set of common QoS properties necessary for most service systems, which are performance, reliability and security. It is believed that these QoS properties are also universal to any service-oriented system. The aim of this chapter is therefore to provide an overview of these common QoS properties, give a precise definition of these QoS properties and offer an in-depth analysis of the issues, challenges and research opportunities of QoS properties in designing and developing service-oriented systems.

BACKGROUND

This section discusses existing research efforts related to QoS definitions and modeling, inter-dependencies among QoS properties and QoS technical support.

The Definitions and Modeling of QoS Properties

According to Zheng et al. (2014), QoS is a broad concept that encompasses a number of non-

functional properties such as price, availability, reliability, and reputation, so on and so forth. These properties apply both to stand-alone services and to composite services.

QoS modeling is a way to precisely define various QoS properties. There are also some QoS modeling approaches. For example, Bocciarelli and D'Ambrogio (2011) proposed an approach that extends Business Process Modeling Notations for modeling non-functional properties of business processes. This approach first constructed several meta-models to enable the modeling of multiple non-functional properties, such as performance, reliability and so on.

Non-functional Requirements (NFR) Framework, proposed by Chung et al. (2012), can be also considered as a way for QoS properties modeling. NFR framework treated non-functional requirements as “soft-goals”, and refined them to a set of fine-grained soft-goals and operationalization finally. i* (Horkoff et al., 2014) was a big improvement that integrates resources and tasks with goals and soft-goals, and defines four major types of relationships among them, which are resource dependency, task dependency, goal dependency and soft-goal dependency. With these four types of dependencies, the relationships between functional requirements and QoS properties were represented explicitly and much clearer. Besides, a QoS properties description model (Desai et al., 2005) was developed for QoS properties definition.

Inter-Dependencies Between QoS properties

There are numerous complex and nontrivial dependencies among quality properties, including *Conflicts* and *Cooperation* with different degrees, as described by Eyged and Grünbacher (2005). It is the same to QoS properties. A conflict occurs between two QoS properties if they make contradicting statements about common quality attributes or quality factors, whereas QoS properties cooperate if they mutually enforce common

quality attributes or quality factors. Quality Attribute Risk and Conflict Consultant (QARCC), designed by Boehm and In (1996), was an early attempt to investigate the conflicts between different QoS properties. Egyed and Grünbacher (2005) addressed the issue on false inter-dependencies, and proposed to use Requirement Tracing (RT) technique to filter out false conflicts and cooperation automatically. They determine whether two QoS properties have true dependencies according to the trace dependency between their related functional requirements. The dependency link is established depending on whether these functional requirements are overlapped. Take the e-commerce system for example, if the performance property “The system should give a response to users within 1 second when a user logs in the system” is constrained on the functional requirement “Login”, and another performance property “The system should give a response to users within 3 second when a user search the goods from the system” is constrained on the functional requirement “Search Goods”, then we can say these two QoS properties have false dependency since there is no overlap between their functional requirements “Login” and “Search Goods”.

Technical Support of Performance, Reliability and Security

Concurrency is mainly adopted to improve a service-oriented system’s performance. Tretola & Zimeo (2007) proposed a new technique, workflow fine-grained concurrency, to enable partially concurrent execution of some sequential activities in a workflow. Garcia et al. (2009) divided the whole application into a set of same application logics that are deployed onto different web servers, and used a QoS properties control mechanism to provide differentiation between distinct categories of service consumers.

Many approaches have been developed to support reliability, divided into two major categories: 1) sequential strategies, such as Retry (Chan et al., 2007) and Recovery Block (Scott et al., 2006),

which invoke backup services to take over primary ones when primary services fails; 2) parallel strategies, such as Active strategy (Salatge & Fabre, 2007), which employ multiple instances of services at the same time. Compared to sequential strategies, parallel strategies are more effective to enhance reliability through reducing failure rates, but parallel strategies consume more resources than sequential strategies and need to maintain the consistency among service instances all the time, which thereafter degrades the performance. Zheng & Lyu (2010) designed an adaptive QoS-aware middleware to provide reliable web services. This middleware dynamically selects sequential or parallel strategies according to user requirements and runtime QoS information.

Protecting messages from attacks is an effective way to support the confidentiality aspect of security. Rahaman & Schaad (2007) proposed an approach to avoid third-party rewriting by extending the message with an account section. OASIS also establishes a series of security specifications to improve the confidentiality of messages, such as Web Services Security (OASIS, 2012), Web Services Trust (OASIS, 2009), etc. In addition, service composition approaches also play a crucial role in supporting security for service-oriented systems (Koshutanski & Massacci, 2003). Secure service composition (Carminati et al., 2006) is to select services that satisfy client requirements by involving security concerns (i.e. confidentiality, authorization and authentication) into the composition process.

PERFORMANCE, RELIABILITY AND SECURITY

Three common QoS properties are discussed in this chapter, i.e. performance, reliability and security. Precise definitions to three common QoS properties and their detailed metrics will be first given as well as the real cases. Then the inter-dependencies between these QoS properties will be presented.

The Definitions and Metrics

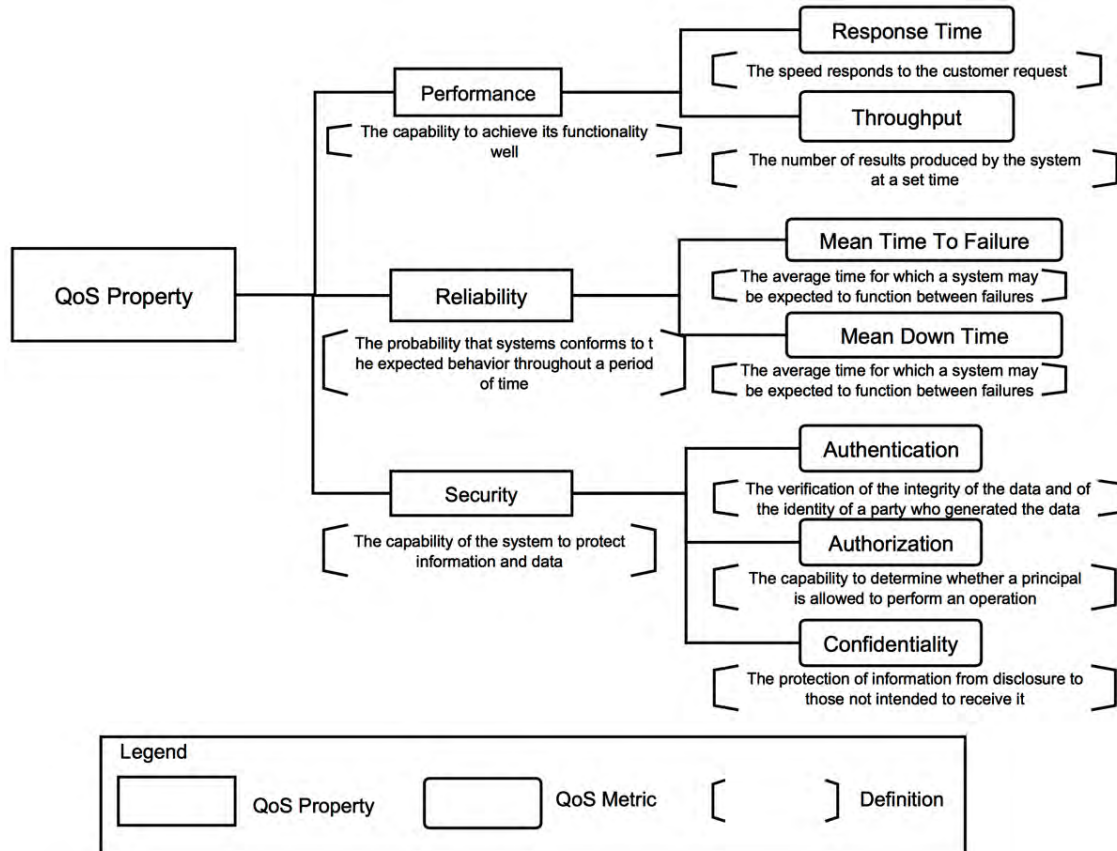
Performance refers to the capability of service-oriented systems to achieve its functionality well. Performance can be measured by the system response time (Jayathilaka et al., 2015) and the system throughput (Yang et al., 2012). The response time of a service-oriented system is defined by the speed with which the service-oriented system responds to the customer request, whereas the throughput is defined by the number of results produced by the system at a set time.

Reliability usually refers to the probability that the system will conform to its expected behavior throughout a period of time. There exist different reliability models for different classes of systems. Some reliability models are based on the failure rate whereas others are based on the recovery rate. The most well known is the mean time to

failure (MTTF), which specifies the average time for which a system may be expected to function between failures. For example, if a service-oriented system is expected to have a failure rate of 10^{-3} failures/hour, its MTTF is once every 1000 hours. Reliability models that build on the recovery rate usually use the mean down time (MDT) to predict the reliability of a system.

Security refers to the capability of the system to protect information and data so that unauthorized persons or systems cannot read or modify them and authorized persons or systems are not denied to access them (ISO, 2001). Usually, security is not measured by a quantitative metric; instead, it is normally transformed into three types of specific requirements: authentication, authorization, and confidentiality. Authentication is the verification of the integrity of the data and of the identity of a party who generated the data. Authorization is

Figure 1. A summary of performance, reliability and security



the process by which one determines whether a principal is allowed to perform an operation, where as confidentiality is the protection of information from disclosure to those not intended to receive it (Reddy et al., 2014). Figure 1 shows a summary of the above QoS properties.

Examples in Industry Practices

A real-world service-oriented system – a Global Equity Trading System (GETS) is used to explain the above QoS properties. As a stock trading service system, GETS's main function is to facilitate its users to sell or buy shares online. GETS's users are traders and operators. Traders are investors who would buy or sell shares (also called orders in GETS) for their institutions or themselves, whereas operators are employees in stock exchanges who would use GETS to process shares. GETS is required to provide three major trading services for its users, which are Order Management, Trading, and Trade Management. Due to space limitation, only a few of QoS properties in GETS are presented as below. `

- **Performance**
 - Response time
 - The response time should be less than 5s during trading.
 - The response time should be less than 2s when placing a price to an order.
 - Throughput
 - More than 300 orders are matched per second.
- **Reliability**
 - Mean time to failure
 - The mean time to failure should be less than once every 10000 hours.
- **Security**
 - Authentication
 - Traders are required to login the system before submit an order.
 - Authorization

- Only operators can do trade management.
- Only traders can do order management.
- Confidentiality
 - Trade information should be sent to printing agents in privacy.

The Inter-Dependencies Between QoS Properties

QoS properties are not independent. Understanding the inter-dependencies between these QoS properties will facilitate a deeper understanding of each QoS property. A potential conflict and cooperation model is used to analyze the inter-dependencies between performance, reliability and security. This model is built on the basis of Egyed and Grünbacher's model (2005) as well as the experience of domain experts. The potential conflict and cooperation model is a two-dimensional matrix (*Table 1*), where rows represent affected QoS properties and columns represent affecting QoS properties. That means the value in each blank represents the potential effects that columns play on rows. To exactly specify the inter-dependencies between two QoS properties, the original potential conflict and cooperation model is adapted by extending the two-level effects to four-level effects, i.e. Make, Help, Hurt and Break. Make means two QoS properties strongly cooperate with each other, i.e. improving one QoS property will strongly increase another QoS property. Help means a low degree of cooperation, i.e. improving one QoS property will increase another QoS property. Break means there are strong conflicts between two QoS properties, i.e. improving one QoS property will strongly degrade another QoS property. Hurt means a low degree of conflict. For example, improving Security will strongly degrade Performance, thus "-" is filled in row 3, column 2. A typical example is checking user password. Checking user password is the design strategy for the security assurance, but this strategy will consume more time and accordingly the system

Table 1. The inter-dependencies between performance, reliability and security.

QoS Properties	Effect		
	Performance	Reliability	Security
Performance	++	-	--
Reliability	0	++	0
Security	-	+	++

Note: ++ represents Make; + represents Help; - represents Hurt; -- represents Break; 0 represents no effect.

will respond slowly to user's request, which makes contradicting statements with the performance. Improving Reliability plays no effect on Security, thus "0" is filled in row 2, column 4.

It is worth noting that the types of QoS properties can be changed according to the actual requirements of different systems. For example, the performance can be divided into two QoS metrics, i.e. response time and throughput, and re-evaluate the conflict and cooperation between these two QoS metrics and other QoS properties.

FUTURE RESEARCH DIRECTIONS

The development of service-oriented systems usually put emphasis on technologies at design and implementation phase, such as service modeling, service discovery, service composition, service evaluation, service monitoring and management. Yet, few considerations have been taken into the requirements analysis phase. It is widely recognized that the requirements of service-oriented systems are fundamental to execution of the subsequent development tasks. Likewise, ensuring the quality of service requirements is also a crucial but challenging task for service-oriented systems development. A service requirement specifies both the functional requirements and QoS properties for an atomic service or a composite service. The composition of services should follow contracts between different service providers, and contracts are derived from service requirements. It is believed that the modeling and analysis of

QoS properties will become more and more important in service-oriented systems development and the approaches or techniques to implement QoS properties modeling and analysis will be used more and more widely.

With regards to QoS properties modeling, there is still a lack of unified recognized classifications and definitions for various QoS properties, although a lot of classifications have been proposed. It is recommended that the classification and the definition of QoS properties should be towards a specific domain or a specific business sector, such as e-tourism systems, e-collaboration systems, and financial systems. In doing so, the modeling approaches of QoS properties will be more accepted and easier to applied in different areas.

With regards to QoS properties analysis, the techniques still remain at the qualitative analysis of inter-dependencies among QoS properties, whereas the quantitative analysis is rarely discussed. However, the quantitative analysis of inter-dependencies among QoS properties is crucial for the improvement of the whole system's quality. For example, to exactly which extent the security can affect the performance. Therefore, it is believed that one of the future research directions of QoS properties analysis is to put more attentions on the qualitative analysis of inter-dependencies among QoS properties.

CONCLUSION

Service-oriented computing is a promising computing paradigm for software applications. However, there are a lot of key research issues in the service-oriented computing paradigm, involving service modeling and specification, service selection, composition algorithms, handling composition faults, and so on. QoS properties are key factors to resolve these issues as well as a crucial aspect in the design of service-oriented systems because it directly touches the concerns of the service users. Therefore it is necessary and important to achieve a

deeper understanding of different types of QoS properties. This chapter introduces three QoS properties namely performance, reliability and security, and provides an overview of these QoS properties and offer an in-depth analysis of the issues, challenges and research opportunities of QoS properties in designing and developing service-oriented systems. It is hoped that through the demonstration and explanation, this chapter offers the readers an interesting snapshot of performance, reliability and security in service-oriented systems.

ACKNOWLEDGMENT

This work is jointly supported by National Natural Science Foundation of China under Grant No. 61402406, Zhejiang Provincial Natural Science Foundation of China under Grant No. LY13F02010 and No. LY15F02004 and Public Welfare Technology Application Research Project of Zhejiang Province under Grant No. 2014C23008.

REFERENCES

Bocciarelli, P., & D'Ambrogio, A. (2011). A BPMN extension for modeling non functional properties of business processes. In *Proceedings of the 2011 Symposium on Theory of Modeling & Simulation: DEVS Integrative M&S Symposium* (pp. 160-168). Society for Computer Simulation International.

Boehm, B., & In, H. (1996). Identifying quality-requirement conflicts. *IEEE Software*, 13(2), 25–35. doi:10.1109/52.506460

Calheiros, R. N., Masoumi, E., Ranjan, R., & Buyya, R. (2015). Workload prediction using arima model and its impact on cloud applications qos. *IEEE Transactions on Cloud Computing*, 3(4), 449–458. doi:10.1109/TCC.2014.2350475

Cardoso, J., Barros, A., May, N., & Kylau, U. (2010). Towards a unified service description language for the internet of services: Requirements and first developments. In *Proceedings of IEEE International Conference on Services Computing* (pp. 602-609). IEEE. doi:10.1109/SCC.2010.93

Carminati, B., Ferrari, E., & Hung, P. C. (2006). Security conscious web service composition. In *Proceedings of 2006 IEEE International Conference on Web Services* (pp. 489-496). IEEE. doi:10.1109/ICWS.2006.115

Chan, K. M., Bishop, J., Steyn, J., Baresi, L., & Guinea, S. (2007). A fault taxonomy for web service composition. In *Proceedings of International Conference on Service-Oriented Computing* (pp. 363-375). Berlin: Springer.

Chung, L., Nixon, B. A., Yu, E., & Mylopoulos, J. (2012). *Non-functional requirements in software engineering*. New York, NY: Springer Science & Business Media.

De Virgilio, R. (2010). Meta-modeling of semantic web services. In *Proceedings of IEEE International Conference on Services Computing* (pp. 162-169). IEEE.

Desai, N., Mallya, A. U., Chopra, A. K., & Singh, M. P. (2005). Interaction protocols as design abstractions for business processes. *IEEE Transactions on Software Engineering*, 31(12), 1015–1027. doi:10.1109/TSE.2005.140

Egyed, A., & Grünbacher, P. (2005). Supporting software understanding with automated requirements traceability. *International Journal of Software Engineering and Knowledge Engineering*, 15(05), 783–810. doi:10.1142/S0218194005002464

Espadas, J., Molina, A., Jiménez, G., Molina, M., Ramírez, R., & Concha, D. (2013). A tenant-based resource allocation model for scaling Software-as-a-Service applications over cloud computing infrastructures. *Future Generation Computer Systems*, 29(1), 273–286. doi:10.1016/j.future.2011.10.013

- García, D. F., Garcia, J., Entrialgo, J., Garcia, M., Valledor, P., Garcia, R., & Campos, A. M. (2009). A QoS control mechanism to provide service differentiation and overload protection to internet scalable servers. *IEEE Transactions on Services Computing*, 2(1), 3–16. doi:10.1109/TSC.2009.3
- Hiel, M., & Weigand, H. (2009). Interoperability changes in an adaptive service orchestration. In *Proceedings of IEEE International Conference on Web Services* (pp. 351-358). IEEE. doi:10.1109/ICWS.2009.23
- Horkoff, J., Barone, D., Jiang, L., Yu, E., Amyot, D., Borgida, A., & Mylopoulos, J. (2014). Strategic business modeling: Representation and reasoning. *Software & Systems Modeling*, 13(3), 1015–1041. doi:10.1007/s10270-012-0290-8
- ISO. (2001). *ISO/IEC 9126-1: Software Engineering - Product Quality - Part 1: Quality Model*. Retrieved May 15, 2016, from http://www.iso.org/iso/catalogue_detail.htm?csnumber=22749
- Jayathilaka, H., Krintz, C., & Wolski, R. (2015). Response time service level agreements for cloud-hosted web applications. In *Proceedings of the Sixth ACM Symposium on Cloud Computing* (pp. 315-328). ACM. doi:10.1145/2806777.2806842
- Koshutanski, H., & Massacci, F. (2003). An access control framework for business processes for web services. In *Proceedings of the 2003 ACM workshop on XML security* (pp. 15-24). ACM. doi:10.1145/968559.968562
- Leitner, P., Michlmayr, A., Rosenberg, F., & Dustdar, S. (2010). Monitoring, prediction and prevention of sla violations in composite services. In *Proceedings of IEEE International Conference on Web Services* (pp. 369-376). IEEE. doi:10.1109/ICWS.2010.21
- Lemos, A. L., Daniel, F., & Benatallah, B. (2016). Web service composition: A survey of techniques and tools. *ACM Computing Surveys*, 48(3), 33.
- Milanovic, N., & Malek, M. (2004). Current solutions for web service composition. *IEEE Internet Computing*, 8(6), 51–59. doi:10.1109/MIC.2004.58
- OASIS. (2009). *WS-Trust 1.4*. Retrieved May 13, 2016, from <https://www.oasis-open.org/standards#wstrustv1.4>
- OASIS. (2012). *Web Services Security v1.1.1*. Retrieved May 13, 2016, from <https://www.oasis-open.org/standards#wssv1.1.1>
- Rahaman, M. A., & Schaad, A. (2007). Soap-based secure conversation and collaboration. In *Proceedings of IEEE International Conference on Web Services* (pp. 471-480). IEEE. doi:10.1109/ICWS.2007.167
- Rai, G. N., Gangadharan, G. R., & Padmanabhan, V. (2015). Algebraic modeling and verification of Web service composition. *Procedia Computer Science*, 52, 675-679.
- Reddy, M. R., Yalla, P., & Chandra, J. V. (2014). Design and Implementation of Integrated Testing Tool based on Metrics and Quality Assurance. *International Journal of Applied Engineering Research*, 9(21), 10463–10472.
- Salatge, N., & Fabre, J. C. (2007). Fault tolerance connectors for unreliable web services. In *Proceedings of 37th Annual IEEE/IFIP International Conference on Dependable Systems and Networks* (pp. 51-60). IEEE. doi:10.1109/DSN.2007.48
- Scott, R. K., Gault, J. W., & McAllister, D. F. (2006). Fault-tolerant software reliability modeling. *IEEE Transactions on Software Engineering*, 13(5), 582–592.
- Spohrer, J., Maglio, P. P., Bailey, J., & Gruhl, D. (2007). Steps toward a science of service systems. *Computer*, 40(1), 71–77. doi:10.1109/MC.2007.33

Tretola, G., & Zimeo, E. (2006). Workflow fine-grained concurrency with automatic continuation. In *Proceedings 20th IEEE International Parallel & Distributed Processing Symposium* (pp. 8). IEEE.

Wang, Y., & Sun, J. (2015). *Quality Enhancement in Designing E-Tourism Service Systems*. In *New Business Opportunities in the Growing E-Tourism Industry* (pp. 43–63). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-8577-2.ch003

Wang, Y., Sun, J., Zhao, L., Wang, X., Yang, X., & Kavs, A. J. (2011). Understanding quality requirements in knowledge intensive service systems. *International Journal of Services Technology and Management*, 16(2), 208–221. doi:10.1504/IJSTM.2011.042597

Yang, X., Zhao, L., Wang, X., Wang, Y., Sun, J., & Cristoforo, A. (2012). Satisfying quality requirements in the design of a partition-based, distributed stock trading system. *Software, Practice & Experience*, 42(2), 131–157. doi:10.1002/spe.1054

Yau, S. S., & Liu, J. (2008). Service functionality indexing and matching for service-based systems. In *Proceedings of IEEE International Conference on Services Computing* (vol. 1, pp. 461–468). IEEE. doi:10.1109/SCC.2008.125

Zhao, J. L., Tanniru, M., & Zhang, L. (2007). Services computing as the foundation of enterprise agility: Overview of recent advances and introduction to the special issue. *Information Systems Frontiers*, 9(1), 1–8. doi:10.1007/s10796-007-9023-x

Zheng, Z., & Lyu, M. R. (2010). An adaptive QoS-aware fault tolerance strategy for web services. *Empirical Software Engineering*, 15(4), 323–345. doi:10.1007/s10664-009-9126-8

Zheng, Z., Zhang, Y., & Lyu, M. R. (2014). Investigating QoS of real-world web services. *IEEE Transactions on Services Computing*, 7(1), 32–39. doi:10.1109/TSC.2012.34

ADDITIONAL READING

Alrifai, M., & Risse, T. (2009). Combining global optimization with local selection for efficient QoS-aware service composition. In *Proceedings of the 18th international conference on World Wide Web* (pp. 881–890). ACM. doi:10.1145/1526709.1526828

Alrifai, M., Risse, T., & Nejdl, W. (2012). A hybrid approach for efficient Web service composition with end-to-end QoS constraints. [TWEB]. *ACM Transactions on the Web*, 6(2), 7. doi:10.1145/2180861.2180864

Calheiros, R. N., Vecchiola, C., Karunamoorthy, D., & Buyya, R. (2012). The Aneka platform and QoS-driven resource provisioning for elastic applications on hybrid Clouds. *Future Generation Computer Systems*, 28(6), 861–870. doi:10.1016/j.future.2011.07.005

Calinescu, R., Grunske, L., Kwiatkowska, M., Mirandola, R., & Tamburrelli, G. (2011). Dynamic QoS management and optimization in service-based systems. *IEEE Transactions on Software Engineering*, 37(3), 387–409. doi:10.1109/TSE.2010.92

Canfora, G., Di Penta, M., Esposito, R., & Villani, M. L. (2005). QoS-aware replanning of composite web services. In *Proceedings of IEEE International Conference on Web Services* (pp. 121–129). IEEE. doi:10.1109/ICWS.2005.96

Cao, B. Q., Li, B., & Xia, Q. M. (2009). A service-oriented qos-assured and multi-agent cloud computing architecture. In *Proceedings of the IEEE International Conference on Cloud Computing* (pp. 644–649). Berlin, Heidelberg: Springer. doi:10.1007/978-3-642-10665-1_66

Cardellini, V., Casalicchio, E., Grassi, V., Iannucci, S., Presti, F. L., & Mirandola, R. (2012). Moses: A framework for qos driven runtime adaptation of service-oriented systems. *IEEE Transactions on Software Engineering*, 38(5), 1138–1159. doi:10.1109/TSE.2011.68

- Cardellini, V., D'Angelo, M., Grassi, V., Marzolla, M., & Mirandola, R. (2015). A Decentralized Approach to Network-Aware Service Composition. In *European Conference on Service-Oriented and Cloud Computing* (pp. 34-48). Springer International Publishing. doi:10.1007/978-3-319-24072-5_3
- Chen, X., Zheng, Z., Liu, X., Huang, Z., & Sun, H. (2013). Personalized qos-aware web service recommendation and visualization. *IEEE Transactions on Services Computing*, 6(1), 35–47. doi:10.1109/TSC.2011.35
- Delimitrou, C., & Kozyrakis, C. (2013). Paragon: QoS-aware scheduling for heterogeneous data-centers. *ACM SIGPLAN Notices*, 48(4), 77–88. doi:10.1145/2499368.2451125
- Delimitrou, C., & Kozyrakis, C. (2014). Quasar: Resource-efficient and QoS-aware cluster management. *ACM SIGPLAN Notices*, 49(4), 127–144.
- Dragone, L., & Rosati, R. (2007). Checking e-service consistency using description logics. In *Proceedings of IEEE International Conference on Services Computing* (pp. 162-169). IEEE. doi:10.1109/SCC.2007.40
- Gao, T., Ma, H., Yen, I. L., Bastani, F., & Tsai, W. T. (2005). Toward QoS analysis of adaptive service-oriented architecture. In *Proceedings of IEEE International Workshop on Service-Oriented System Engineering* (pp. 1-8). IEEE.
- Garg, S., Modi, K., & Chaudhary, S. (2016). A QoS-aware Approach for Runtime Discovery, Selection and Composition of Semantic Web Services. *International Journal of Web Information Systems*, 12(2), 177–200. doi:10.1108/IJWIS-12-2015-0040
- Hammoudeh, M., & Newman, R. (2015). Adaptive routing in wireless sensor networks: QoS optimisation for enhanced application performance. *Information Fusion*, 22, 3–15. doi:10.1016/j.inffus.2013.02.005
- Luo, N., Yan, J., & Liu, M. (2007). Towards efficient verification for process composition of semantic web services. In *Proceedings of IEEE International Conference on Services Computing* (pp. 220-227). IEEE. doi:10.1109/SCC.2007.120
- Mabrouk, N. B., Beauche, S., Kuznetsova, E., Georgantas, N., & Issarny, V. (2009). QoS-aware service composition in dynamic service oriented environments. In *Proceedings of ACM/IFIP/USENIX International Conference on Distributed Systems Platforms and Open Distributed Processing* (pp. 123-142). Berlin, Heidelberg: Springer. doi:10.1007/978-3-642-10445-9_7
- Martinez, E., Díaz, G., Cambroner, M. E., & Schneider, G. (2010). A model for visual specification of e-contracts. In *Proceedings of 2010 IEEE International Conference on Services Computing* (pp. 1-8). IEEE. doi:10.1109/SCC.2010.32
- Menascé, D. A., Ruan, H., & Gomaa, H. (2007). QoS management in service-oriented architectures. *Performance Evaluation*, 64(7), 646–663. doi:10.1016/j.peva.2006.10.001
- Papazoglou, M. P. (2003). Service-oriented computing: Concepts, characteristics and directions. In *Proceedings of the Fourth International Conference on Web Information Systems Engineering* (pp. 3-12). IEEE. doi:10.1109/WISE.2003.1254461
- Papazoglou, M. P., Traverso, P., Dustdar, S., & Leymann, F. (2008). Service-oriented computing: A research roadmap. *International Journal of Cooperative Information Systems*, 17(02), 223–255. doi:10.1142/S0218843008001816
- Qi, L., Dou, W., & Chen, J. (2016). Weighted principal component analysis-based service selection method for multimedia services in cloud. *Computing*, 98(1-2), 195–214. doi:10.1007/s00607-014-0413-x

- Sweeney, J., Sweeney, J., Armstrong, R. W., Armstrong, R. W., Johnson, L. W., & Johnson, L. W. (2016). The effect of cues on service quality expectations and service selection in a restaurant setting: A retrospective and prospective commentary. *Journal of Services Marketing*, 30(2), 136–140. doi:10.1108/JSM-01-2016-0010
- Wan, Z., & Wang, P. (2016). A Novel Pattern Based Dynamic Service Composition Framework. In *Proceedings of IEEE International Conference on Services Computing* (pp. 235-242). IEEE. doi:10.1109/SCC.2016.38
- Wang, G. et al.. (2005). Service level management using QoS monitoring, diagnostics, and adaptation for networked enterprise systems. In *Proceedings of the Ninth IEEE International EDOC Enterprise Computing Conference* (pp. 239-248). IEEE. doi:10.1109/EDOC.2005.30
- Wang, G., Chen, A., Wang, C., Fung, C., & Uczekaj, S. (2004). Integrated quality of service (QoS) management in service-oriented enterprise architectures. In *Proceedings of the Eighth IEEE International Enterprise Distributed Object Computing Conference* (pp. 21-32). IEEE.
- Yuan, M., Huang, Z., Hu, J., Li, X., & Zhu, Y. (2009). Ensuring coordination of multi-business interactions. In *Proceedings of IEEE International Conference on Services Computing* (pp. 356-363). IEEE.
- Zhao, X., Shen, L., Peng, X., & Zhao, W. (2015). Toward SLA-constrained service composition: An approach based on a fuzzy linguistic preference model and an evolutionary algorithm. *Information Sciences*, 316, 370–396. doi:10.1016/j.ins.2014.11.016
- Zheng, Z., & Lyu, M. R. (2015). Selecting an optimal fault tolerance strategy for reliable service-oriented systems with local and global constraints. *IEEE Transactions on Computers*, 64(1), 219–232. doi:10.1109/TC.2013.189
- Zheng, Z., Ma, H., Lyu, M. R., & King, I. (2013). Collaborative web service qos prediction via neighborhood integrated matrix factorization. *IEEE Transactions on Services Computing*, 6(3), 289–299. doi:10.1109/TSC.2011.59
- Zheng, Z., Zhang, Y., & Lyu, M. R. (2010). Distributed qos evaluation for real-world web services. In *Proceedings of 2010 IEEE International Conference on Web Services* (pp. 83-90). IEEE. doi:10.1109/ICWS.2010.10
- Zou, J., Wang, Y., & Lin, K. J. (2010). A formal service contract model for accountable saas and cloud services. In *Proceedings of 2010 IEEE International Conference on Services Computing* (pp. 73-80). IEEE. doi:10.1109/SCC.2010.85

KEY TERMS AND DEFINITIONS

Performance: The capability of the service-oriented systems to achieve its functionality well, which can be measured by the throughput and the response time.

Quality of Services: The capability of the service-oriented systems to achieve its non-functional goals.

Reliability: The probability that the service-oriented system will conform to its expected behavior throughout a period of time.

Security: The capability of the system to protect information and data.

Service-Oriented Computing: A new computing paradigm that delivers services that are cooperating with each other over the Internet.

Service-Oriented System: A distributed software system based on the software delivery SaaS (Software as a Service) model.

Software as a Service: A software delivery model according to which a software vendor develops a web-based software application and then operates that application over the Internet for use by its customers.

Understanding User Experience

S**Camille Dickson-Deane***University of Melbourne, Australia***Hsin-Liang (Oliver) Chen***University of Massachusetts Boston, USA*

INTRODUCTION

As the need for information continues to drive the design and development process in society, understanding how one uses solutions is key. Interactions and the information they produce, drive how often and how successful an experience can be, knowing that each interaction may differ from another. User experience determines the quality of an interaction being used by an actor in order to achieve a specific outcome. The actor can have varying roles and evolving needs thus reviewing and predicting experiences are important. As an actor uses and gains feedback, the feedback guides individual and group behavior thus becoming pertinent to how interactions occur. This then questions how artefacts are designed to promote such interactions and what processes should be incorporated to ensure successful interpretation, use, physical reaction and conation. This chapter discusses the effects of user experiences today based on societal needs and expectations. It shows how the field is delineated into numerous sub-topics all of which can stand on their own yet, still draw from each other. The discussions will include fields such as cognitive science, human computer interaction, learning sciences and even ergonomics to show how design and subsequently interactions can assist in having successful user experiences.

BACKGROUND

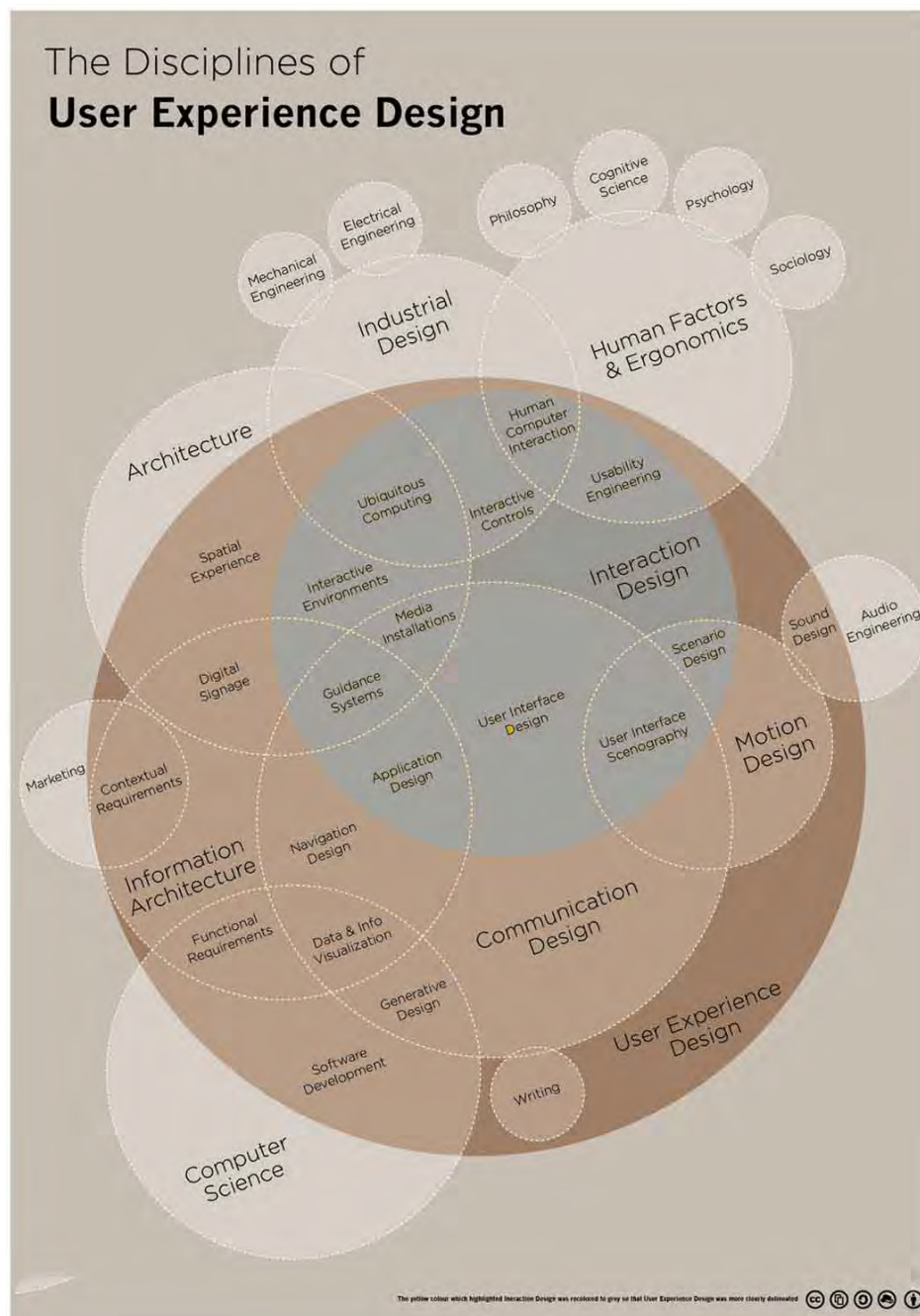
User experience (UX) is defined as a generalised approach to acknowledging the quality of interac-

tion an actor [a user] has with an organisation, its service or products (Nielsen, 1992; Norman & Draper, 1986). This acknowledgement is defined as, but not limited to the actors' perceptual views, cognitive and affective feelings, interpretation of meaning, and attitudes. Today, UX is frequently known to review the quality of interaction an actor has with computer-related technology and is known to include an organisation or product's brand experience (Park, Han, & Kim, 2013). UX is identified as an action as well as a field with relational ties to several different knowledge brackets (i.e., Human Computer-Interaction (HCI), Ergonomics, Cognitive Science, Industrial Science, Human Factors) and has elements which include "usability, affect and user value" (McNamara & Kirakowski, 2006; Norman & Draper, 1986; Notess, 2001; Park et al., 2013 pg. 286; Tullis & Albert, 2008). As such, its history has varying starting points and thus varying influences (see Figure 1).

Norman (1986) known for his work in interaction design, first coined the term. The need to ensure that any experience an actor has with an object is classified not only by the expressed needs of the actor, but also by anticipating the future needs and thus uses of the said object [by the actor]. This means that an object can draw from many disciplines, in order to fulfil this expectation towards a satisfying experience for the actor. The UX is an outcome presented through the interaction of the actor with a user interface (Seow, 2005). The user interface is somewhat physical in its definition and typically describes the layer of an object where an action has an interaction to access

DOI: 10.4018/978-1-5225-2255-3.ch661

Figure 1. The Disciplines of User Experience Design. This figure illustrates the many disciplines which are included in the field of User Experience.



the said object. An example of the user interface, interaction and user experience is as follows

- **Actor:** A user (Man with no inherent challenges to using a door).
- **Object:** A door knob that is circular/round and smooth in design.
- **Context:** The user (the man) is aware of the time and is rushing to enter a law office before it closes. The entry to the law office

is a door with a door knob that is circular/round and smooth in design.

- **User Interface:** The surface of the door knob.
- **Interaction:** The user using their hand to grasp the door-knob with the intent of opening the door. The actions that may be included in the interaction are, turning, pushing, and pulling.
- **(Intended/Expected/Prior) User Experience:** The user based on previous experiences with other doors, turned and pulled the door knob to enter the law office.
- **[Actual] User Experience:** The user used the door knob, turned and pulled the door towards his/her person in an effort to enter the law office. The door provided feedback that it can only be pushed and as a result the user had to change their expectations and push the door.

The above example was elaborated to show how some of the complexity of the experience of opening a door can be for an actor (Dewey, 1925; Pucillo & Cascini, 2014). In this example, the mental model the actor had of the proposed experience of opening the door may have changed after the experience occurred. If there was a change between what was expected and what occurred, this means that the experience had by the actor differed from his/her own expectations and as a result, there are a number of feelings and misinterpretations occurred during the steps required for that one action (Pucillo & Cascini, 2014). This means that the user's experience with this system (i.e., man opening the law office's door), has to accommodate the actor's state; any previous experience including influences from professional practice; any additional properties that the system may have and finally the context in which the interaction will occur (Hassenzahl & Tractinsky, 2006; Lallemand, Gronier, & Koenig, 2015). When looking at what can be used to determine the steps in a user's experience, Pucillo and Cascini (2014) defined a framework

of affordances where a basic psychological action needs to be fulfilled (i.e., experience affordance), using physical and cognitive skills of the actor (i.e., use affordance), to manipulate an artefact (i.e., effect affordance) thus having some compatible interaction between the actor and the artefact (i.e., manipulation affordance). This experience that the actor has, is the result of his or her internal state (Dewey, 1925; Hassenzahl & Tractinsky, 2006 p. 95) and this state combined with the features of the designed system, the context in which the system was design and then the subsequent interaction that occurred creates for a spectrum of responses from the actor (Lallemand et al., 2015). A great example would be Shin, Jung and Chang's (2012) use of the Technology Acceptance Model (TAM) to test the intent of use for QR codes in the marketing of products. This research, showed that the quality of the QR codes influenced the user's perception of use through their own interaction with the artefacts (i.e., the QR codes). This further elaborates on Dewey's (1925) supposition that experiences are contextual to an individual and includes numerous factors such as emotions, expectations, interests and the like and as such [it] is a dynamic and temporary event.

The [user] experiences that currently occur are now referenced to computer screens and websites (Park et al., 2013). The ubiquitous use of portable technology in society means that the field that was once focused on the general use of objects and systems now is focused on technology-related objects and systems. Gross and Sheridan's (2011) study presented research on how students' used a university's library website. Their research focussed on the library's tool, *Summon* and found that most students showed positive experience with this new tool, and identified needed changes to assist students in evaluating retrieved information. In Kohler's (2011) research, he identified five research areas of pragmatism, sociability, usability, hedonism and collaboration as a virtual co-creation system, *Ideation Quest*, was researched to understand a users' co-creation experience. This research highlights not only the participants

experience in virtual worlds, but also illustrates where the experience seamlessly is modified into learning and performance behavior. These studies, reflect today's societal norms and now expands the discussion into other areas such as usability, learnability and accessibility (Park et al., 2013). Both of these terms, have great historical relevance to the field, even as we continue to exist in the Information Age. The meaning of usability and accessibility has changed because of the increase in societal use of technology and thus it means that UX design will accommodate many more pertinent discussions which will enhance the field.

MAIN FOCUS OF THE ARTICLE

Issues, Controversies, Problems

The importance of the user and their ability to access the information as provided, is key. This discussion needs to also include the conversion where the information retrieval experiences affect the learning process - thus allowing for *the user* to now become *the learner*. This broaches the boundaries towards ensuring that *the user experience* is also acknowledging *the learner experience* as well (Moore, Dickson-Deane, & Liu, 2014; Notess, 2001; Seow, 2005; Squires & Preece, 1996). The boundaries between the user and the learner experience is very much more prevalent in environments that are clearly demarcated for learning (i.e., online learning environments, environments associated with educational institutions, etc.). As technology is becoming more and more pervasive, the field of information science has blurred these lines and now encourages the sharing and collaboration of information not only with the intent to inform, but also for users to learn/perform (Moore et al., 2014; Notess, 2001; Squires & Preece, 1996). The field of user experience is exceedingly volatile due to the complex nature of the overarching fields (i.e., Human Computer Interaction) as well as the difficulty in use and interpretation of the topics needed to translate

conceptual concepts (Lallemand et al., 2015). Some of the concerns include:

- The importance of acknowledging that a user's experience, although being subjective to that individual and can also contribute to group behaviors.
- The knowledge that user's can learn bad behaviors from poorly designed solutions and that these bad designs should not diminish the power of a good design.
- That user testing should use the true "users" of the product and not short-change the value of the experience by testing others who can act like a user (i.e., "pseudo-users").
- A lack of awareness in understanding how to reduce the information-retrieval lag time yet enhance the experiences (i.e. information-access) of users.
- That accessibility of information and thus the experience does not mean that users need to be accommodated but is a cultural awareness that all users are to be included.

Thus, the experience is an encompassing realm, which affects the user through all of their senses and is demonstrated in design or usage through tangible and intangible objects, experiences and environments.

User Experience is Subjective but Affects Groups

UX is very subjective which when extrapolated affects a large number of individuals from any user-base. There are two views to a users' subjectivity which are beneficial for designers to understand

1. The act of finding objects in an interface and experiencing an environment or service is a very subjective activity (Glanz, 2012).
2. Accepting subjectivity as an ambiguous action but knowing that it is part of a group experience thus being an essential concept of user experience is important (Hassenzahl, 2010; Wright & McCarthy, 2010).

Law, Roto, Hassenzahl, Vermeeren, and Kort (2009, p. 726) provide a good example from their survey respondents, “only an individual can have an experience but I believe it can be externalised (albeit poorly) and recognised and related to by others,” to illustrate the relationship between “individual subjectivity” and “group”. Topics such as Open Educational Resources (OERs) and the design and access to such materials can sway entire groups away from the purpose due to poorly designed materials (Ives & Pringle, 2013; Rodriguez-Solano, 2015). This relationship is an important issue as researchers study user experience at the individual level, the group level, and the dynamics among the group especially as they are access information as well as learn using the materials being accessed.

Learning from Bad Designs

Another challenge with UX, is managing poorly designed interfaces. When interfaces are poorly designed, and the redesigned solutions provide more poor designs, there is a great chance that the user will learn and adapt their behaviors to accept the poor designs. Looking at a user’s behaviour over time can show that based on their own experiences, their behaviours adjust and sometimes accommodate and accept experiences which are not normally favorable (Hassenzahl et al., 2013; Karapanos, 2013). This creates a two-fold challenge for designers in that they

1. Should be cognizant of their design choices and really look for research to support their design choices.
2. Should research user behaviours over periods of time to illustrate continued usage and potential adoption and adaptability.

Using End-Users for Testing as Opposed to “Others”

One key challenge that occurs in user experience, is not using the end-users to obtain the true “user”

experience (Gruen, Rauch, & Redpath, 2002; Pruitt & Grudin, 2003). The key to getting the best feedback and seeing the value in the feedback is the reviewing of the experience using actual/potential users. Not using true end-users to gain the information needed on the experience can distort the *true* user experience thus hinder the design and development process (Shneiderman, 1998). Geelen, Reinders, and Keyson, (2013) emphasized that the involvement of end-users was critical to their study- design. In their study, they were able to identify needs, behavioral changes, and community initiatives of end-users to foster the design process instead of by technical and financial considerations. The value of end-user is well recognized by software engineers as the input from end-users improve software quality (Barker & Fiedler, 2012; Ko, Abraham, Beckwith, & Blackwell, 2011; Panko & Port, 2012). This is especially important where many UX projects are based on online environments and soliciting feedback from end-users can be easier to obtain (Hedegaard & Simonsen, 2013).

Challenges in Reducing Information-Retrieval Lag

Most design cycles rarely include the user experience and furthermore are not aware of the potential harmful lag has on the information retrieval process (Card, Moran, & Newell, 1983; Holden, Volda, Savoy, Jones, & Kulanthaivel, 2016). Liaw and Huang (2003) found direct relationship between the response time of the internet and a user’s perceived satisfaction. This relationship informed the user’s belief that the solution was easy to use, useful and addressed the core intent of the experience/use. Having an increased time in the retrieval of information is perceived to somewhat affect the user process negatively. The increased time can be internet access (i.e., insufficient bandwidth) or designs that force users to search for information that should be easily accessible (MacKenzie, 1992; Martin & Corl, 1986). When designing, considerations to Fitts’

Law (1954) and Hick-Hyman's Law is pertinent when looking at the user accessing information through computer interfaces. Fitt's (1954) Law discusses when a user attempts to select part of an interface, how quickly they can select that part of the interface is determined by the distance to the target and the size of the target. Hick-Hyman's Law describes the time it takes a user to make a decision as a result of the number of possible choices offered - increased choices is equal to an increased time (Seow, 2005). Both of these laws may not be readily used when determined how to manage user experience and be even more harmful during a learning activity (Moore et al., 2014). When discussing the design of a webpage, discussion on the information retrieval process is a given, but when designing online learning, this discussion tends to be coupled with other elements of the design process. Designing using online learning environments (OLEs) requires a more in depth approach – the designs are not only base on aesthetics but also include the meaning and thus correct interpretation of the content. This process of pedagogical usability, when elaborated also discusses [pedagogical] accessibility. Thus ignoring these key issues can negatively affect interactions (Bakaev & Avdeenko, 2015; Moore et al., 2014).

Accessibility and Usability

When usability was first defined, it focused on the effectiveness, efficiency and satisfaction of a product in a specific context (ISO 9241-11, 1998). Now as it is being redefined to include a user's emotional experience, the types of ways in which a product is used and access to the infrastructure and technology that provide the Internet (Bevan, Carter, & Harker, 2015) is highlighted. As usability continues to be addressed, accessibility is becoming just or even more important (Youngblood & Youngblood, 2013). Accessibility addresses a user's ability to have the same access, interact the same, and enjoy the benefit of the same services no matter their disposition (Section 508,

1998). Coupling both terms (i.e., accessibility and usability) is ideal for the user experience as they both are needed to sufficiently fulfil a user's experience. Being inclusive requires the specific attention to awareness of user characteristics and usage patterns. User's may have cognitive impairments that are not sufficiently disclosed before the user-testing process and this can create results which are not applicable to the tool but also results which are not inclusive to the potential user base.

SOLUTIONS AND RECOMMENDATIONS

UX design is somewhat secretly added to design and development cycles and even though it is a keen discussion with designers, organizations are yet to seriously include the concerns of the user into the general design process. Usability evaluations, if designed well, have many opportunities to capture user experiences and being aware of what an "experience" may really encompass is the most important part of the design and development process. Considering the individual characteristics, the true habits of a user, the differences in time-based experiences and the difference between access and usage will assist in solving some of the issues discussed. Using research to inform design decisions as well as recognising that the user is the initiator of the experience is the key to characterizing experiences and thus providing the impetus for solutions as needed.

FUTURE RESEARCH DIRECTIONS

Due to the numerous associated fields reflected in this chapter, future research directions may vary thus creating a great research field for *user experiences*. When dissecting *the user* and the characteristics associated with, current research includes the review of physical/behavioural attributes and now includes the cognitive attributes. Reviewing affective and now conative and social

abilities will be key going forward (Hilgard, 1980; Reeves, 2006). Some additional thoughts on future research include:

- **Human Factors and Ergonomics:** Technology is becoming smaller and more portable so looking at human well-being within the systems performance may raise issues with keeping up with technology usage. The age gap between those who know technology to be pertinent for their everyday existence and those who remembered when technology was not affordable, is growing. This means that there may be instances for researching physical injury associated with the newer technology designs, thus prompting more focus on the physical designs for everyday use; thus creating ties to accessibility and usability. This can also extend to include patient and doctors at hospitals – hospital records, admittance areas/computer-related forms, etc. All of these can address the support of functional and social health issues.
- **Cognitive Science:** In reviewing the virtual worlds and the immersive technologies, research can look at the field of gaming. The gaming industry has had a head start on what works for these environments but also extrapolating these experiences to simulated environments that can be used to train soldiers, doctors, emergency crew, new automobile drivers or simply a child learning to complete a messy activity in a safe environment, is an option. This also has other references to eye tracking studies and brain/cognitive mapping thus including design issues such as information retrieval and good design methods can assist.
- **Industrial Science:** Improving the quality and the productivity of any product is becoming easier from a user experience perspective. Technology is assisting industrial engineers in the observance of user behaviours and as smart cars and smart homes

are on the horizon, getting immediate feedback from users with varying characteristics to inform design and the use of such, will be one of the main foci.

- **Human Computer Interaction:** This field is heavily focused on the personalization and visualization of interactions. Future research may want to combine these two and push the observance review process to users (similar to the industrial science field) and allow users to illustrate their own personalized interactions thus informing a strategic push for informed personalized and personal designs.

Looking at the entire field of user experience and the future of the research, it seems like a good futuristic research will include, users as researchers/designers thus placing the control of the access, usability, quality and productivity in the hands of the product owner.

CONCLUSION

As the field of user experience continues to grow, so will the expectations of its use. Users are becoming more and more aware of the power they yield as a consumer and this means that they will more and more begin driving the design and development process. Being able to use social media to complain about or promote a product, provide feedback on experiences through recommender systems and even just be quietly observed through behaviour intelligence software is becoming the norm. Drawing more attention to the value of user experiences will in effect create additional sectors, which seek to use such influences to gain improvements on numerous offerings. Understanding not only the behaviours, thoughts and feelings that guide a users' experience, but also specifically looking at the decisions that are made as a result of all of these factors will influence the field of information science and definitely impact the learning.

REFERENCES

- Bakaev, M., & Avdeenko, T. (2015, May). A quantitative measure for information transfer in human-machine control systems. In *Control and Communications (SIBCON), 2015 International Siberian Conference on* (pp. 1-4). IEEE. doi:10.1109/SIBCON.2015.7146989
- Barker, S. K., & Fiedler, B. (2012). *Developers, Decision Makers, Strategists or Just End-users? Redefining End-User Computing for the 21st Century: A Case Study* (Doctoral dissertation). IGI Global.
- Bevan, N., Carter, J., & Harker, S. (2015). ISO 9241-11 revised: What have we learnt about usability since 1998? In *Human-Computer Interaction: Design and Evaluation* (pp. 143-151). Springer International Publishing.
- Card, S. K., Moran, T. P., & Newell, A. (1983). *The psychology of human-computer interaction*. L. Erlbaum Associates.
- Dewey, J. (1958). *Experience and Nature* (1925). New York: Dover.
- Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 47(6), 381–391. doi:10.1037/h0055392 PMID:13174710
- Geelen, D., Reinders, A., & Keyson, D. (2013). Empowering the end-user in smart grids: Recommendations for the design of products and services. *Energy Policy*, 61, 151–161. doi:10.1016/j.enpol.2013.05.107
- Glanz, M. (2012). User experience research: Modelling and describing the subjective. *Interdisciplinary Description of Complex Systems*, 10(3), 235-247.
- Gross, J., & Sheridan, L. (2011). Webscalediscovery: The user experience. *New Library World*, 112(5/6), 236–247. doi:10.1108/03074801111136275
- Gruen, D., Rauch, T., Redpath, S., & Ruettinger, S. (2002). The use of stories in user experience design. *International Journal of Human-Computer Interaction*, 14(3-4), 503–534. doi:10.1080/10447318.2002.9669132
- Hassenzahl, M. (2010). Experience design: Technology for all the right reasons. *Synthesis Lectures on Human-Centered Informatics*, 3(1), 1–95. doi:10.2200/S00261ED1V01Y201003HCI008
- Hassenzahl, M., Eckoldt, K., Diefenbach, S., Laschke, M., Len, E., & Kim, J. (2013). Designing moments of meaning and pleasure. Experience design and happiness. *International Journal of Design*, 7(3).
- Hassenzahl, M., & Tractinsky, N. (2006). User experience-a research agenda. *Behaviour & Information Technology*, 25(2), 91–97. doi:10.1080/01449290500330331
- Hedegaard, S., & Simonsen, J. G. (2013, April). Extracting usability and user experience information from online user reviews. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2089-2098). ACM. doi:10.1145/2470654.2481286
- Hilgard, E. R. (1980). The trilogy of mind: Cognition, affection, and conation. *Journal of the History of the Behavioral Sciences*, 16(2), 107–117. doi:10.1002/1520-6696(198004)16:2<107::AID-JHBS2300160202>3.0.CO;2-Y PMID:11608381
- Holden, R. J., Volda, S., Savoy, A., Jones, J. F., & Kulanthaivel, A. (2016). Human Factors Engineering and Human-Computer Interaction: Supporting User Performance and Experience. In *Clinical Informatics Study Guide* (pp. 287-307). Springer International Publishing.
- ISO 9241-11, I. S. O. Usability Definition (1998).
- Ives, C., & Pringle, M. M. (2013). Moving to open educational resources at Athabasca University: A case study. *The International Review of Research in Open and Distributed Learning*, 14(2), 1–13. doi:10.19173/irrodl.v14i2.1534

- Karapanos, E. (2013). User experience over time. In *Modeling Users' Experiences with Interactive Systems* (pp. 57-83). Springer Berlin Heidelberg. doi:10.1007/978-3-642-31000-3_4
- Ko, A. J., Abraham, R., Beckwith, L., Blackwell, A., Burnett, M., Erwig, M., & Rosson, M. B. et al. (2011). The state of the art in end-user software engineering. *ACM Computing Surveys*, 43(3), 21. doi:10.1145/1922649.1922658
- Kohler, T., Fueller, J., Matzler, K., & Stieger, D. (2011). Co-creation in virtual worlds: The design of the user experience. *Management Information Systems Quarterly*, 35(3), 773-788.
- Lallemand, C., Gronier, G., & Koenig, V. (2015). User experience: A concept without consensus? Exploring practitioners perspectives through an international survey. *Computers in Human Behavior*, 43, 35-48. doi:10.1016/j.chb.2014.10.048
- Law, E. L. C., Roto, V., Hassenzahl, M., Vermeeren, A. P., & Kort, J. (2009, April). Understanding, scoping and defining user experience: a survey approach. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 719-728). ACM. doi:10.1145/1518701.1518813
- Liaw, S. S., & Huang, H. M. (2003). An investigation of user attitudes toward search engines as an information retrieval tool. *Computers in Human Behavior*, 19(6), 751-765. doi:10.1016/S0747-5632(03)00009-8
- MacKenzie, I. S. (1992). Fitts law as a research and design tool in human-computer interaction. *Human-Computer Interaction*, 7(1), 91-139. doi:10.1207/s15327051hci0701_3
- Martin, G. L., & Corl, K. G. (1986). System response time effects on user productivity. *Behaviour & Information Technology*, 5(1), 3-13. doi:10.1080/01449298608914494
- McNamara, N., & Kirakowski, J. (2006). Functionality, usability, and user experience: three areas of concern. *Interactions*, 13(6), 26-28.
- Moore, J. L., Dickson-Deane, C., & Liu, M. (2014). Designing CMS courses from a pedagogical usability perspective. In *Perspectives in Instructional Technology and Distance Education : Research on Course Management Systems in Higher Education* (p. 169). Information Age Publishing.
- Nielsen, J. (1992, June). Finding usability problems through heuristic evaluation. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 373-380). ACM.
- Norman, D. A., & Draper, S. W. (1986). *User centered system design*. Hillsdale, NJ: Academic Press.
- Notess, M. (2001). Usability, user experience, and learner experience. *eLearn*, 2001(8), 3.
- Panko, R. R., & Port, D. N. (2012, January). End user computing: the dark matter (and dark energy) of corporate IT. In *System Science (HICSS), 2012 45th Hawaii International Conference on* (pp. 4603-4612). IEEE.
- Park, J., Han, S. H., Kim, H. K., Cho, Y., & Park, W. (2013). Developing elements of user experience for mobile phones and services: Survey, interview, and observation approaches. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 23(4), 279-293. doi:10.1002/hfm.20316
- Pruitt, J., & Grudin, J. (2003, June). Personas: practice and theory. In *Proceedings of the 2003 conference on Designing for user experiences* (pp. 1-15). ACM. doi:10.1145/997078.997089
- Pucillo, F., & Cascini, G. (2014). A framework for user experience, needs and affordances. *Design Studies*, 35(2), 160-179. doi:10.1016/j.destud.2013.10.001
- Reeves, T. C. (2006). How do you know they are learning? The importance of alignment in higher education. *International Journal of Learning Technology*, 2(4), 294-309. doi:10.1504/IJLT.2006.011336

Rodriguez-Solano, C., Sánchez-Alonso, S., & Sicilia, M. A. (2015). Creation of reusable open textbooks: Insights from the Connexions repository. *British Journal of Educational Technology*, 46(6), 1223–1235. doi:10.1111/bjet.12196

Section 508. Section 508 Law and Related Laws and Policies. (1998). Retrieved from <http://www.section508.gov/content/learn/laws-and-policies>

Seow, S. C. (2005). Information theoretic models of HCI: A comparison of the Hick-Hyman law and Fitts law. *Human-Computer Interaction*, 20(3), 315–352. doi:10.1207/s15327051hci2003_3

Shin, D. H., Jung, J., & Chang, B. H. (2012). The psychology behind QR codes: User experience perspective. *Computers in Human Behavior*, 28(4), 1417–1426. doi:10.1016/j.chb.2012.03.004

Shneiderman, B. (1998). *Designing the User Interface: Strategies for Effective Human-Computer Interaction* (3rd ed.). Reading, MA: Addison-Wesley.

Squires, D., & Preece, J. (1996). Usability and learning: Evaluating the potential of educational software. *Computers & Education*, 27(1), 15–22. doi:10.1016/0360-1315(96)00010-3

Tullis, T., & Albert, B. (2008). *Measuring the user experience: collecting, analyzing, and presenting usability metrics*. Morgan Kaufmann.

Wright, P., & McCarthy, J. (2010). Experience-centered design: Designers, users, and communities in dialogue. *Synthesis Lectures on Human-Centered Informatics*, 3(1), 1–123. doi:10.2200/S00229ED1V01Y201003HCI009

Youngblood, N. E., & Youngblood, S. A. (2013). User experience and accessibility: An analysis of county web portals. *Journal of Usability Studies*, 9(1), 25–41.

KEY TERMS AND DEFINITIONS

Accessibility: Enables a person with a disability to be afforded the opportunity to acquire the same information, engage in the same interactions, and enjoy the same services as a person without a disability in an equally effective and equally integrated manner, with substantially equivalent ease of use (Section 508, 1998).

Affordance: The perception and actual properties which are used to determine the possibilities for actionable use.

Cognitive Science: A discipline that combines the study of the mind and its processes by building off of psychology, linguistics, philosophy, and computer modeling.

Interaction Design: A process where interfaces are created to promote behaviours that are influenced by users as well as influencing users at the same time. It allows for an observation of how users interact with user interfaces in an effort to learn user behaviour.

Usability: According to the ISO 9241(1998) standard usability is: “The ease with which a user can learn to operate, prepare inputs for, and interpret outputs of a system or component.”

User Experience: Sometimes referred to as UX is the subjective yet reactive occurrence an actor has when interacting with a company, its services and its products.

User Interface: A physical area which is used for communicating a user to a system. Due to the pervasive use of the technology, the definition is mostly used to define an information device which uses computer technology [system] as a medium to interact with a user.

The What, How, and When of Formal Methods

S

Aristides Dasso*Universidad Nacional de San Luis, Argentina***Ana Funes***Universidad Nacional de San Luis, Argentina*

INTRODUCTION

Formal Methods (FM) is an area of Software Engineering. They comprise a collection of methodologies and related tools, employing a mathematical basis –as do most engineering disciplines–, to construct its products. Although FM are part of Software Engineering, they extend their scope to the development not only of software products but also of hardware systems.

The goal of this article is to give answers to questions such as *what are Formal Methods*, *how are Formal Methods implemented*, *how are they used in Software Engineering*, *when should they be used*, among other related questions.

The chapter starts answering the question of *what are FM*; also the aims of FM are stated at the same time that their main characteristics are presented.

An example that shows *how* FM can be used to help specifying software requirements as well as the rest of the stages in a software development process is given as answers to the questions *how are FM implemented* and *how FM are used in Software Engineering*.

A discussion about *when they should be used*, explaining why they should be employed when the software system is required to be as secure and reliable as possible and how they can also be used as a complement to traditional development methods, is also provided.

A section on the state of the art in FM, providing an analysis of Lightweight FM and the grow-

ing impact that Model Checking is having in the software and hardware industry as an automatic FM for system verification, is presented. Finally, a discussion on the use of automatic analyzers like Alloy, which replace conventional analysis based in theorem proof by a “non-complete” analysis based in the examination of cases, is also given.

BACKGROUND

The ‘What’ of FM

What Are Formal Methods

FM contain a wide range of methods and related tools oriented to the production of secure and reliable software and hardware systems by employing a logic-mathematical basis.

As traditional development methods, FM consist of a set of techniques and supporting tools to assist developers during the whole software development process. The fundamental difference with traditional methods is that FM are based on mathematics and formal logic leading to unambiguous specifications, where desired properties of a system under development can be formally expressed and verified.

The adoption of FM makes possible the specification and the verification of software and hardware systems. They provide the mathematical tools to develop new concrete formal specifications and, eventually, executable code from abstract

formal specifications, where all the development steps can be formally verified.

Therefore, in contrast to traditional development methods, FM use mathematical proofs as a complement to software testing in order to verify the correctness of the system under development.

There are a number of different Formal Methods, each having its own notation, methodology and supporting tools; a comprehensive list of FM can be found in Formal Methods Wiki (Bowen). Formal notations or formal specification languages are used to produce formal specifications of software and hardware systems. There are different styles in formal notations, and there also are different degrees of rigor in development. Formal specifications can be written either using *abstract* or *concrete* style. Some formal specification languages adopt a *property-oriented* style, allowing the creation of *algebraic specifications*, where the desired properties of the system under development are given by axioms, in a purely declarative way. This kind of specifications is called *algebraic* because specifications are seen like heterogeneous algebras. A different style for specifications is the use of *model-based* notations, which make use of concrete data types (integers, reals, sets, list, etc.). They are more concrete than property-oriented specifications. However, in general, formal specification languages favors abstraction, being oriented to answer the question *what are the software requirements* of a system more than *how the requirements are going to be implemented*, i.e. they are oriented to describe what a system should do more than how it should be done.

Formality in the use of FM can go from the maximum degree of rigor, i.e. the use of formal specification and formal verification of the whole system to the use of formalizations for requirements specification only.

Some confusion can arise with the use of terms *formal language*, *formal notation*, *formal system* and *formal method*. In (Alagar, V. S. & Periyasami, K., 2011) the differences are clearly

stated. In Table 1 a synoptic view of the differences between those terms is shown.

A *formal notation* is a language that has both its *syntax* and *semantics* formally defined. Specification languages used in FM and programming languages are formally defined by their syntax and semantics. Syntax drives the validation of the language constructs. Strictly speaking the syntax of a language is given by the formal grammar of the language. There are a number of methods to do this, among them BNF (Backus – Naur Form). These methods are widely used to represent the syntax of programming languages, specification languages, etc.

Semantics, on the other hand, is a way of associating meanings to the language's valid constructs. Roughly speaking it can be said that given a language $\mathcal{L}$, the semantics of $\mathcal{L}$ can be represented by a pair $(\mathcal{U}, \mathcal{I})$, where $\mathcal{U}$ is the universe of values, i.e. the set of all possible values (numbers, characters, Boolean values, etc.), and $\mathcal{I} : \mathcal{L} \rightarrow \mathcal{U}$ is an interpretation function that assigns to each construct of the language $\mathcal{L}$ a value in the universe of values $\mathcal{U}$.

There are different styles for giving language semantics. *Operational semantics* is concrete; it may define an abstract machine, and is not well suited for proofs. In *Denotational semantics* a denotation (usually a mathematical object) is given to each phrase or construct of the language; since it is abstract, it is well suited for proofs; *Axiomatic semantics* is also abstract; here each phrase or

Table 1. A view of formal methods, formal systems and formal languages

Formal Method	Formal System	Formal Notation	Formal Language: formal syntax + formal semantics.
		Proof system; axioms + inference rules	
	Automatic tool support for specification, proof assistance, code generation, etc.		

construct of the language is described by axioms, and normally are limited to conditional equations.

As Table 1 shows, besides a formal notation normally implemented in a specification language defined by its formal syntax and semantics, a FM has also associated a *proof system* with its own *axioms* and *inference rules* that allows to produce formal deductions of system properties (theorems). These three elements make what is called a *formal system*.

Axioms are formulas of a formal system language that are considered valid. They are premises or ‘starting points’ for reasoning, using the deductive mechanism, in the formal system. *Inference rules* are syntactic rules or functions that given some premises return a conclusion. Together with the axioms of the formal system they can be used to infer new properties that are coherent with the axioms. An example of an inference rule is given in Equation 1. It states that, under the assumption that a logical expression E_1 has been proved true, then as a conclusion we have that a disjunction of E_1 with any other logical expression E_2 is also true.

Formal systems, which not only permit writing specifications but also proving formal properties of a system via their deductive mechanism, can be enriched with (a) techniques that guide the user on the steps to be followed during the development of a software system, going from the creation of abstract specifications that capture user’s requirements to more concrete specifications and eventually to an executable implementation, and (b) automatic tools giving support to these activities, namely edition, type checking, verification and translation to executable code. When all these conditions are met we can say that the formal system is a Formal Method.

Equation 1. Introduction of disjunction

$$\frac{E_1}{E_1 \vee E_2}$$

What Is To Be Done

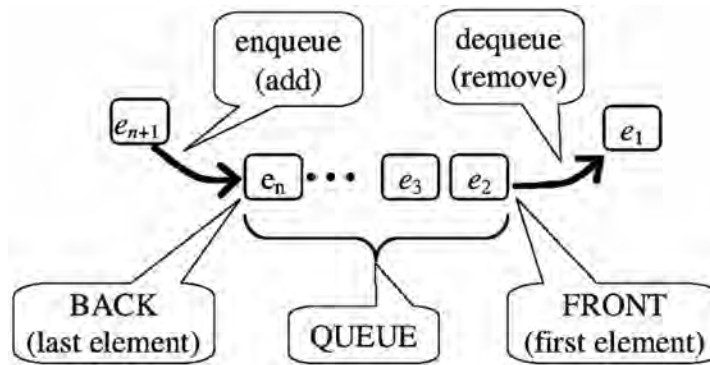
The use of formal notations favors the writing of abstract specifications, focusing on the answer to the question *what is to be done* more than *how is to be done*.

This point is illustrated by a simple example of a queue specification. A queue is a very common data structure both in software and hardware, as well as in other areas of science. It is also known as one of the abstract data type (Black, 2014). It is a very common structure in everyday life as well, e.g. a queue of people at a teller’s window. A queue has a *front* and a *back* and it can be abstractly defined by the operations that can be applied on it: enqueue (add) when incorporating a new element in the queue; dequeue (remove) when taking an element of the queue out of it and first that returns the element at the front of the queue. Figure 1 shows a schematic of a queue.

Some rules must be observed when using a queue: No element can be removed from an empty queue, only the element in front can be removed, and a new element can only be added after the element at the back of a queue. This translates by saying that a queue is a FIFO structure: *First In First Out*.

In an abstract specification, it does not matter what a queue is holding, i.e. what kind of elements the queue has (e.g. people, cars, numbers, etc.), neither how the queue is implemented (by software, hardware, etc.), the important point is that a queue must respect the FIFO principle. Example 1 shows an abstract specification of a queue in RSL (The RAISE Language Group, 1992), the specification language of the RAISE formal method (The RAISE Method Group, 1995). RSL module QUEUE_0 defines two sorts (‘Queue’ and ‘Elem’); ‘Queue’ corresponds to the interest type of the specification and ‘Elem’ to the type of the queue’s elements. Both types have been specified as sorts, i.e. sets of abstract values with only one operator defined on them (the equality). **value** keyword introduces the definition of three functions and a constant, namely the functions

Figure 1. A schematic representation of a queue



‘add’ (enqueue a new element), ‘rem’ (remove or dequeue the first element), and ‘first’ (return the first element in the queue), and the constant ‘empty’, representing the empty queue. Both ‘rem’ and ‘first’ have been specified as partial functions ($\rightharpoonup$) since they are not defined for the empty queue.

Under the **axiom** keyword, the properties that a queue must hold have been stated; they specify the FIFO principle.

This specification does not say *how* the queue or its elements are implemented; it only specifies by means of algebraic definitions of its operations *what* the acceptable behavior of a queue should be. This is done by the axioms that combine the queue operations in such a way that the properties of a queue are clearly established. In the example, the first axiom establishes that the first element of a queue having only one element ‘e’ is element ‘e’; the second axiom establishes that the first element in a queue, obtained after inserting a new element, is equal to the first element in the queue before insertion, i.e. the insertion of a new element into the queue does not change the first element into the queue. The third axiom says that remove an element from the queue after adding a new one produces the same result that add the new element and then remove an element. Finally, the fourth establishes that a remove operation applied to a queue with one element empties the queue. This is an example of a *property-oriented* or *algebraic specification*, a kind of specifications that use sorts and algebraic definitions of functions.

The ‘How’ of FM

How to develop using FM? Traditional FM adhere to two different development approaches: *transformational* and *invent and verify*.

Transformational methods use a calculus; the developer starts from an initial abstract specification and following predefined and verified transformation rules obtain a new more concrete specification that preserves the properties of the initial specification. Successive calculations lead to new specifications and eventually to implementation.

In invent and verify approach, the developer starts with an initial abstract specification and in a successive step, ‘invent’ a more concrete specification that needs to be formally verified as an implementation of the previous one. This process can involve one or more steps until a concrete specification is obtained, which is suitable for translation to executable code. In this category is the RAISE method, which provides a “*refinement relation*” that can drive the development from an abstract specification to a concrete one. The formal definition of the refinement relation can be found in (The RAISE Method Group, 1995).

To illustrate the use of refinements, Example 2 gives a more concrete specification of a queue in the RSL module QUEUE_1, which can be formally proved as a possible implementation of the abstract specification given in QUEUE_0, i.e. one development step towards a more concrete specification of a queue has been ‘invented’.

Example 1. Property-oriented RSL specification of a queue

```

scheme QUEUE_0 =
class
  type
    Queue
    Elem
  value
    empty: Queue,
    add: Elem × Queue → Queue,
    rem: Queue  $\rightarrow$  Queue,
    first: Queue  $\rightarrow$  Elem
  axiom
     $\forall e: \text{Elem}, q: \text{Queue} \bullet$ 
      first(add (e, empty)  $\equiv$  e,
     $\forall e_1, e_2: \text{Elem}, q: \text{Queue} \bullet$ 
      first(add (e1, add(e2, q)))  $\equiv$ 
        first(add(e2, q)),
     $\forall e_1, e_2: \text{Elem}, q: \text{Queue} \bullet$ 
      rem(add(e1, add(e2, q)))  $\equiv$ 
        add(e1, rem(add(e2, q))),
     $\forall e: \text{Elem}, q: \text{Queue} \bullet$ 
      rem(add (e, empty)  $\equiv$  empty
end

```

Using the built-in types list (*) and **Nat** (natural numbers), type ‘Queue’ has been defined as a list of elements (Queue = Elem*) and the type ‘Elem’ as **Nat**. The signatures of the operations are the same. Now the axioms reflect the presence of built-in types and the use of some of its operations: **hd** (that returns the head of a list), and **tl** (that returns the tail of a list). Also the axioms now define explicitly the functions ‘add’, ‘rem’ and ‘first’, and the constant ‘empty’ (as the empty list) employing the operations of the RSL built-in types that were used for the implementation of the queue. Example 2 is also an example of a *model-based specification*, a kind of specification that makes use of concrete data types and explicit function definitions.

Examples 1 and 2 have clearly shown how different levels of abstraction can be achieved when writing formal specifications of a same problem. In QUEUE_0, an abstract, property-oriented RSL specification has been produced, where the proper-

ties that a queue must hold are given. In QUEUE_1, a more concrete, model-based specification has been created, as a possible implementation of the abstract specification.

To illustrate how a new development step (QUEUE_1, in the example) can be proved as an implementation of the previous one (QUEUE_0), the RAISE refinement relationship between these two specifications is informally described below:

QUEUE_1 is an implementation of QUEUE_0 if it satisfies the following two conditions:

1. The *static implementation relationship*, i.e. the signature of QUEUE_1 must include the signature of QUEUE_0. This means that all the entities (types, values, objects, variables and channels) in QUEUE_0 must be in QUEUE_1 with the same names and maximal types. This property can be checked statically by a tool.

Example 2. Model-based RSL specification of a queue

```

scheme QUEUE_1 =
class
  type
    Queue = Elem*
    Elem = Nat
  value
    empty: Queue,
    add: Elem × Queue → Queue,
    rem: Queue  $\tilde{\rightarrow}$  Queue,
    first: Queue  $\tilde{\rightarrow}$  Elem
  axiom
    empty  $\equiv$  <>,
     $\forall e: \text{Elem}, q: \text{Queue} \cdot$ 
      add(e, q)  $\equiv$  q ^ <e>,
     $\forall q: \text{Queue} \cdot \text{rem}(q) \equiv \mathbf{tl} \ q$ 
      pre q  $\neq$  empty,
     $\forall q: \text{Queue} \cdot \text{first}(q) \equiv \mathbf{hd} \ q$ 
      pre q  $\neq$  empty
end

```

2. The *preservation property*, i.e. all the properties in QUEUE_0 must hold in the refinement (QUEUE_1). Properties are expressed by means of axioms. They also include subtypes restrictions, constant and function's definitions and initial values of variables. The preservation property cannot be checked statically, it needs a formal proof.

In this example, the signatures of QUEUE_1 and QUEUE_0 are the same, namely same names and maximal types; therefore the static implementation relationship is satisfied. To prove the preservation property, we must prove that all the queue properties given by the axioms in QUEUE_0 hold also in QUEUE_1, i.e. the following *proof obligations* arise:

```

[ $\forall e: \text{Elem}, q: \text{Queue} \cdot \text{first}(\text{add}(e, \text{empty})) \equiv e$ ]
[ $\forall e1, e2: \text{Elem}, q: \text{Queue} \cdot \text{first}(\text{add}(e1, \text{add}(e2, q))) \equiv \text{first}(\text{add}(e2, q))$ ]
]

```

```

[ $\forall e1, e2: \text{Elem}, q: \text{Queue} \cdot$ 
  rem(add(e1, add(e2, q)))  $\equiv$  add(e1,
  rem(add(e2, q))) ]
[ $\forall e1, e2: \text{Elem}, q: \text{Queue} \cdot$ 
  rem(add(e1, empty))  $\equiv$  empty ]

```

There is a tool in the RAISE method that helps with the proof obligations. It provides help in proving whether going from one specification to another more concrete is a valid development step or, in other words, whether the more concrete one is a correct implementation of the previous one.

Another interesting *how-question* is *how a FM is constructed*? A FM is constructed from a number of elements, such as a logic, type theory and set theory. Additionally, a semantics, which is used in the definition of its specification language and that is closely associated with the underlying theory. The semantics and the syntax of the specification language conform the core of a FM. A theorem prover, with its methods and rules, and with different scope according to the theory chosen, is sometimes provided. Let us note that

not all FM count with theorem provers. Some of them have just a confidence checker based on type theory. Others FM do not rely on semi-automatic theorem provers for verification but on the use of fully automated tools like model-checkers or analyzers.

When to Use FM

Since the full application of FM requires highly trained and expensive experts, its use has not been widely spread. However, the application of FM is highly recommended mainly in the development of safety critical systems where minimal or zero tolerance to errors is mandatory. So, several organizations, among them the USA's Food and Drug Administration and NASA, all involved in medical devices, avionics, and other areas that require a high level of reliability in their systems, recommend employing FM in some form along the development of software and hardware systems.

Something important to point out in regard FM adoption is that its application does not release developers from testing. Testing and formal verification are complementary approaches. Not everything can be proved, as no testing can find all errors in a system. In addition, nevertheless FM is a useful tool for verification, some kind of validation of user requirements must precede a formal development as well as the delivered software. Therefore, an important recommendation when using formal development is testing all the produced software even if this has been subjected to formal proofs. Conversely, projects developed using traditional methods that have critical components can apply FM for the specification and verification of their critical components; hence, an integration of FM with traditional methods can be applied in these situations.

Finally, it is well-known that the use of mathematics is important in the development of error free products in engineering. It is true that a bridge that has been built using mathematics and physics can fall, but that is no excuse to avoid the use of engineering methods that are widely recognized

as providing assurance building. This idea and the problems associated to the costs of the application of traditional FM have motivated the searching for new ways of applying formality to various areas of system development. Two of the more popular ones are *Lightweight FM* and *Model Checking*, which are treated in the two following sections.

Lightweight Formal Methods

A crucial benefit, when using FM, is not only the possibility of writing unambiguous abstract specifications but more concrete specifications can also be given, and then proven by formal verification that a concrete specification is a correct implementation of an initial abstract specification, to finally obtain an executable piece of code from them; i.e. to prove that the desired system properties, given in an initial specification, still hold in the more concrete specification. Nonetheless, this activity is time consuming and expensive and the number of experts in the field is very limited.

However, although producing a full proof of a system from its initial formal specification is complicated and expensive, it is very useful in cases of critical system. To avoid FM drawbacks, other ways of using FM have been proposed.

One alternative is the adoption of *Lightweight FM*. Lightweight FM is the name given to the partial use of FM, for example when they are applied not to a whole system but only to its critical components (since they are particularly sensitive for reason of special requirements of reliability or security in the specification or in their verification). The partial application of FM for specification only, without formal verification is also referred as Lightweight FM.

Other ways of addressing this use of FM has been supported by new specific tools, analyzers that can generate instances of model invariants, simulate the executions of operations defined as part of the model and be able to check properties of the system.

One of these FM is Alloy (Jackson, 2012). Alloy proposes a new approach to FM, adopting the

use of a formal notation that is useful for describing structures, but that replace the conventional analysis based in theorem proving with the use of completely automatic tools.

An important difference of Alloy analyzer with formal theorem proof is that Alloy's analysis is not *complete* since the tool exams only a finite space of cases. However, the space of cases analyzed is usually counted in the billions or more cases, reaching coverages unreachable by testing. The analyzer, unlike testing, does not make use of testing cases; on the contrary, the user provides the properties that need to be proved.

Alloy has been used in a wide range of applications from finding holes in security mechanisms to designing telephone switching networks.

Using techniques that integrate FM with other more traditional methodologies or semiformal notations is also another way of Lightweight FM. There are a number of theoretical works that combine graphical notations with mathematically precise formalisms, such as providing formalism to the UML, either through the OCL language or by giving to UML graphic language a formal semantics – e.g. see (The precise UML group). Also there is much work done on formalizing Java and Java Machine –see (Hartel, P. and Moreau, L. A. V., Dec. 2001); (Bertelsen, 2003). Other good examples can be found in; (France, 1999); (Lano, 1991); (Weber, 1996); (William E. McUmbur and Betty H.C. Cheng, 2001); (DeLoach, S. and Hartrum, T., 2000); (Meyer, E. and Souquieres, J., 1999); (Funes, A. and George, C., 2003); (Hamann, 2015); (Reggio, G. and Larosa, M., 1997); (Gold-sack, S. and Kent, S., 1996); (Soon-Kyeong Kim and Carrington, D., 2000); (Amalio, N. et al., 2003), among others.

Model Checking

The correctness of the end result of a software development or hardware production can be verified in several ways. They all have to contrast the end product against some initial specification. The methods used to conduct the verification

differ in how the verification process is related to the specification. In general, the degree of user involvement and the completeness and soundness of the assertions that they allow to deduce is critical in determining the success of the verification.

In software as in other areas, a common verification technique is testing. In the case of software is impossible to achieve exhaustive testing. As previously stated, another important method that can be used in conjunction with testing is formal verification. A kind of formal verification is deductive verification where a correctness proof of a system specification is produced from the system axioms and inference rules provided by the underlying logic. This kind of verification is expensive and requires the presence of experts. An alternative to deductive verification is *Model Checking*.

Model Checking is a technique that is based on algorithmically verifying properties of a system. It permits verifying that a discrete model of a hardware or software system, which provides an abstract representation of his behavior usually given by a Kripke structure or a Label Transition System, satisfies some desired properties specified as formulas of a logic. The model of the system is specified in a formal language associated with the Model Checking tool, and the properties are usually given in a temporal or modal logic (Emerson, 1990), (Stirling, 1992). Temporal logics are formalisms that allow specifying properties like *reachability*, e.g. *Can we get to a system state where the formula ψ holds, by starting from the initial state?*

The birth of Model Checking can be attributed to Clarke et al (1986). In 1990, Model Checking received a significant enhancement by the publication of Burch et al (1990) of the symbolic model checking method.

There are a large number of tools available for Model Checking, free and commercial, which are proof of its wide use in software and hardware verification processes. A model-checker provides a completely automatic approach for proving that the system satisfies the desired properties. This

powerful assistance together with the existence of efficient model-checkers and expressive logics makes of model checking an attractive formal approach to software verification.

CONCLUSION

Whether FM are useful, whether FM should be used at all has been discussed for a long time. The importance that FM have in the developing of systems where security, reliability and safety are of the utmost importance, and the concern that these issues have in a number of standards and regulatory agencies and how FM plays a role in fulfilling these needs has been extensively discussed since several years ago, e.g. in (Bowen, J. and Hinchey, M., 1995a) (Bowen, J. and Hinchey, M., 1995b), (Bowen, J. and Hinchey, M., 1995a), (Bowen, J. and Hinchey, M., 2006), (Luqi and Goguen, J., 1997).

Their application in several fields of the industry and sensitive areas such as medical devices, space, avionics, hardware design and verification, and the use of other variants of FM, like Lightweight FM and Model Checking, is proof of its importance and usefulness in the development of critical and quality software.

In favor of this argument, the European Community (EC) has a project dedicated to FM, the DEPLOY Project (ICT- FP7-Deploy Project). In their site, they state that *their overall aim is to make major advances in engineering methods for dependable systems through the deployment of formal engineering methods*.

Kaivola et al. (2009) had replaced testing with formal verification in Intel i7 processor. In (M. Achutha KiranKumar V.), there is also a report on Intel using FM. Several other applications can be mentioned arguing in favor of FM, e.g. the Deploy project site holds pages dedicated to Success Stories of the application of FM (ICT-FP7-DEPLOY Success Stories) and on the use of FM in industry (ICT-FP7-FM).

The main benefits of FM must be found in creating unambiguous specifications and the associated facility of formally proving system properties. All this is facilitated by an underlying logic with a set of axioms, rules and tools, which are part of FM. This constitutes an assurance in the quality of the developed system. These advantages are not present in traditional development systems, where there is no formal semantics for their notations and, consequently, their specifications can suffer from ambiguities and contradictions.

However, we cannot elude the drawbacks of FM. FM rely on the fact that their successful application is based on the involvement of algebra and mathematical logic experts able to conduct correctness proofs, making FM adoption hard and expensive. We cannot also elude the fact that formal specifications can complicate the communication with end users, making the requirement validation difficult. In this regard, a graphical notation, although not formal, is perceived as more intuitive and facilitates communication with non-experts. In fact, there is a great deal of effort put into creating tools and methods that combine FM with more traditional development methods, e.g. by integrating graphical notations with formal semantics.

FUTURE RESEARCH DIRECTIONS

Judging by the important concepts and theoretical results that FM research activities have provided to Software Engineering and that are nowadays in use, such as abstract data-types, pre- and post-conditions, symbolic evaluation, invariants, runtime assertions, state-space analysis, and many others, it is to be expected that in the future FM will continue with its significant contributions.

An analysis of the main currently ongoing research conferences dealing with FM, the scope of the subjects and the number of researches in the area can be used as example of the strength and vitality of the FM area and also serve as an

indicator of the research directions that the field is undertaking. The main interests are directed to:

- *Research into hybrid approaches to formal modeling and analysis; i.e., the combination of (formal and semi-formal) methods for system development, regarding both modeling and analysis.* iFM 2016, URL: <http://en.ru.is/ifm/> (Retrieved march 2016).
- *Addressing critical problems faced by the modern railway – how to deliver reliable service to passengers and to freight operators, while maintaining very high levels of safety. So what can be done? Techniques and tools are needed for modelling, analysis, verification and validation that can cope with the new more complex systems; these techniques must support rather than impede the development process and must address and ensure: Required functionality, Safety and integrity, System security, Adherence to standards.* RSSR 2016, URL: <http://conferences.ncl.ac.uk/rssrail/> (Retrieved march 2016).
- *Developing and evaluating systems that interact with physical processes, and systems that use artificial intelligence technology. Examples include autonomous systems, robots, and cyber-physical systems in general. Applying formal methods to these systems of growing interest and importance is challenging because they exhibit much greater non-determinism than traditional systems, making them challenging to assure.* FM 2016, <http://www.fmeurope.org/?p=567> (Retrieved march 2016).
- *Temporal representation and reasoning in (1) Artificial Intelligence, (2) Databases and (3) Logic and Verification.* TIME 2016, URL: <http://time2016.compute.dtu.dk/index.html> (Retrieved march 2016).
- *Theories of computation and programming, foundations of software engineering and formal techniques in software design and verification, as well as tools that sup-*

port formal techniques for system modeling, design and verification. ICTAC 2016, URL: <http://cc.ee.ntu.edu.tw/~ictac2016> (Retrieved march 2016).

- Symbolic and state space-based techniques for the validation and analysis of software systems. Techniques and empirical evaluations based on explicit representations of state spaces, as implemented in the SPIN model checker or other tools, or techniques based on the combination of explicit representations with other representations, are the focus of this symposium. SPIN 2016, URL: <http://www.spin2016.info/#> (Retrieved march 2016).
- Formal methods and advanced software technologies, especially for engineering complex, large-scale artifacts like cyber-physical systems, networks of things, enterprise systems, or cloud-based services. Contributions relating to formal methods or integrating them with software engineering, as well as papers advancing scalability or widening the scope of rigorous methods to new design goals are especially welcome. SETTA 2016, URL: <http://lcs.ios.ac.cn/setta/about/cfp.shtml> (Retrieved march 2016).
- *Model-based design and analysis, Risk reduction and mitigation techniques, Formal methods, verification and validation, In-the-loop and model-based testing.* Safecomp 2016, URL: <http://www.ntnu.edu/safecomp2016> (Retrieved march 2016).

REFERENCES

Achutha & Kumar, V. (n.d.). *Embracing Formal Verification: An Intel® Graphics Experience*. Retrieved March 2016, from http://www.testandverification.com/DVClub/03_Jul_2014/DVClub_2014_Kiran_formal.pdf

- Alagar, V. S., & Periyasami, K. (2011). *Specification of Software System. Texts in Computer Science*. Springer-Verlag. doi:10.1007/978-0-85729-277-3
- Amalio, N. (2003). Modular UML semantics: Interpretations in Z based on Templates and Generics. *Proceedings of FM 2003*.
- Bertelsen, P. (2003). *Semantics of Java Byte Code*. Retrieved March 2016, from <ftp://ftp.dina.kvl.dk/pub/Staff/Peter.Bertelsen/jvm-semantics.ps.gz>
- Black, P. E. (2014, July 7). *Queue*. Retrieved March 2016, from <https://xlinux.nist.gov/dads/HTML/queue.html>
- Bowen, J. (n.d.). *Formal Methods Wiki*. Retrieved April 2017, from Formal Methods Wiki: http://formalmethods.wikia.com/wiki/Formal_Methods_Wiki
- Bowen, J., & Hinchey, M. (1995a). *Ten commandments of Formal Methods*. IEEE Computer.
- Bowen, J., & Hinchey, M. (1995b, July). Seven more Myths of Formal Methods. *IEEE Software*, 12(3).
- Bowen, J. and Hinchey, M. (2006, January). Ten Commandments of Formal Methods ... Ten Years Later. *IEEE Computer*.
- Burch, J. R., Clarke, E. M., McMillan, K. L., Dill, D. L., & Hwang, L. J. (1990). *Symbolic model checking: 10_20 states and beyond*. In *Logic in Computer Science, 1990. LICS '90* (pp. 428-439). Philadelphia, PA: IEEE.
- Clarke, E. M., Emerson, E. A., & Sistla, A. P. (1986, April). Automatic verification of finite-state concurrent systems using temporal logic specifications. *ACM Transactions on Programming Languages and Systems*, 8(2), 244–263. doi:10.1145/5397.5399
- DeLoach, S., & Hartrum, T. (2000). A Theory-Based Representation for Object-Oriented Domain Models. *IEEE Transactions on Software Engineering*, 6(6), 500–517. doi:10.1109/32.852740
- Emerson, E. A. (1990). *Temporal and modal logic*. In *Handbook of Theoretical Computer Science* (Vol. B, pp. 995–1072). Elsevier Science Publishers B. V.
- France, R. (1999). A Problem-Oriented Analysis of Basic UML Static Requirements Modeling Concepts. *Proceedings of OOPSLA '99*. doi:10.1145/320384.320390
- Funes, A., & George, C. (2003). *Formalizing UML class diagrams*. In *UML and the Unified Process*. Idea Group Publishing.
- Goldsack, S., & Kent, S. (1996). LOTOS in the Object-oriented analysis process. In *Formal Methods and Object Technology*. Springer. doi:10.1007/978-1-4471-3071-0_3
- Hamann, L. (2015). *On Formalizing UML and OCL Features and Their Employment to Runtime Verification* (PhD thesis). Universität Bremen. Retrieved March 2016 from elib.suub.uni-bremen.de/edocs/00104250-1.pdf
- Hartel, P., & Moreau, L. A. V. (2001, December). Formalizing the Safety of Java, the Java Virtual Machine and Java Card. *ACM Computing Surveys*, 33(4), 517–558. doi:10.1145/503112.503115
- ICT- FP7-Deploy Project. (n.d.). *Deploy Project*. Retrieved March 2016, from <http://www.deploy-project.eu/>
- ICT- FP7-DEPLOY Success Stories. (n.d.). *DEPLOY Success Stories*. Retrieved March 2016, from http://www.fm4industry.org/index.php/DEPLOY_Success_Stories
- ICT-FP7-FM. (n.d.). *FAQ on Formal Methods in Industry*. Retrieved March 2016, from http://www.fm4industry.org/index.php/FAQ_on_Formals_Methods_in_Industry
- Jackson, D. (2012). *Software Abstractions*. The MIT Press.

Kaivola, R., Ghughal, R., Narasimhan, N., Telfer, A., Whittemore, J., Pandav, S.,... Naik, A. (2009). Replacing Testing with Formal Verification in Intel® Core™ i7 Processor Execution Engine Validation. In *Computer Aided Verification. 21st International Conference, CAV 2009, Proceedings*. (pp. 414-429). Berlin: Springer.

Kim, S.-K., & Carrington, D. (2000). A Formal Specification Mapping between UML Models and Object-Z Specifications. *LNCS*, 1878, 2–21.

Lano, K. (1991). Z++, An object-oriented extension to Z. In *Z user workshop, Workshops in Computing*. Oxford, UK: Springer.

Luqi,., & Goguen, J. A. (1997, January-February). Formal methods: Promises and problems. *IEEE Software*, 14(1), 73–85. doi:10.1109/52.566430

McUmbler & Cheng. (2001). General Framework for Formalizing UML with Formal Languages. In *Proceedings of the 23rd International Conference on Software Engineering* (pp. 433-442). Washington, DC: IEEE Computer Society.

Meyer, E., & Souquieres, J. (1999). A Systematic Approach to Transform OMT Diagrams to a B Specification. *Proceedings of FM '99* (pp. 875–895). Springer. doi:10.1007/3-540-48119-2_48

Reggio, G., & Larosa, M. (1997). A graphic notation for formal specification of dynamic systems. In *Proceedings of FME'97, number 1313 in LNCS*. Springer. doi:10.1007/3-540-63533-5_3

Stirling, C. (1992). Modal and temporal logics. In D. M. S. Abramsky (Ed.), *Handbook of Logic in Computer Science* (pp. 477-563). Clarendon Press.

The precise UML group. (n.d.). Retrieved March 2016, from <http://www.cs.york.ac.uk/puml/>

The RAISE Language Group. (1992). *The RAISE Specification Language*. Prentice Hall.

The RAISE Method Group. (1995). *The RAISE Development Method*. Prentice Hall.

Weber, M. (1996). Combining Statecharts and Z for the Design of Safety-Critical Control Systems. *Proceedings of the 3rd International Symposium of FME'96*. doi:10.1007/3-540-60973-3_94

ADDITIONAL READING

Abrial, J. R. (1996). *The B-Book*. Cambridge University Press. doi:10.1017/CBO9780511624162

Berard, B., Bidoit, M., Finkel, A., Laroussinie, F., Petit, A., Petrucci, L., & Schnoebelen, Ph. (2001). *Systems and Software Verification. Model-Checking Techniques and Tools*. Springer. doi:10.1007/978-3-662-04558-9

Casey, C. (2000). *A Programming Approach to Formal Methods*. McGraw-Hill.

Clarke, E., Grumberg, O., & Peled, D. (2000). *Model Checking*. The MIT Press.

Clarke, E. M., Emerson, E. A., & Sistla, A. P. (1986). Automatic verification of finite-state concurrent systems using temporal logic specifications. *ACM Transactions on Programming Languages and Systems*, 8(2), 244–263. doi:10.1145/5397.5399

Hall, J. A. (1990, September). Seven Myths of Formal Methods. *IEEE Software*, 7(5), 11–19. doi:10.1109/52.57887

Hinchey, M. G., Bowen, J. P., & Vassev, E. (2010). Formal Methods. In P. A. Laplante (Ed.), *Encyclopedia of Software Engineering* (pp. 308–320). Taylor & Francis.

Larsen, P. G., Fitzgerald, J., & Brookes, T. (1996, May). Applying Formal Specification in Industry. *IEEE Software*, 13(7), 48–56. doi:10.1109/52.493020

Pons, C., & Baum, G. (2000). Formal Foundations of Object-Oriented Modeling Notations. *Proceedings of ICFEM, 00*, 2000.

Ross, P. E. (2005). The Exterminators. *IEEE Spectrum*, 42(September), 36–41. doi:10.1109/MSPEC.2005.1502527

Wing, J. M. (1990, September). A Specifiers Introduction to Formal Methods. *IEEE Computer*, 23(9), 8–24. doi:10.1109/2.58215

KEY TERMS AND DEFINITIONS

Algebraic Specification: Formal specification given in an abstract style based in the use of sorts and algebraic function definitions, like in heterogeneous algebras (multiple sets and their operations).

Formal Proof: Finite sequence of well formed sentences of a formal language, where each sentence either is an axiom or it follows from the previous sentences derived using a rule of inference; the last sentence is the assertion to be proved (theorem).

Model-Based Specification: Formal specification using concrete types and explicit function definitions.

Proof System: Provides a way of deducing formulae, beginning with the axioms of the logic and using its rules of inference.

Property-Oriented Specification: Formal specification given in an abstract style, where the system behavior is defined by a set of properties. These properties are given by axioms relating functions and sorts. (See Algebraic Specification).

Semantics: In a language, it is the meaning of a valid syntactical element, as opposed to syntax, which describes how the symbols of the language are combined to form valid constructions. Most programming languages have their syntax defined formally (traditionally in BNF), while formal specification languages have also their semantics defined formally.

Specification: A document describing what a system should do, what a problem is or what a domain is all about. In formal methods this document is written using a formal language.

Verification: The process of determining whether a product fulfills a set of previously established requirements. In formal verification, this is the process of proving that a more concrete specification preserves the properties of a more abstract specification.

Category T

Teacher Education

Constructing Preservice Teachers' Knowledge of Technology Integration

T

Kathleen A. Paciga
Columbia College Chicago, USA

Angela Fowler
Erikson Institute, USA

Mary Quest
Erikson Institute, USA

INTRODUCTION

Citizens today are connecting through social media and consuming more digital content than ever before (Sasseen, et al., 2013). There are nearly 300 new applications being created every day and, according to the Massachusetts Institute of Technology, more than 10 new groundbreaking technologies created every year (2015). These new apps are placed into the app store (Vaala, Ly & Levine, 2015) with little regulation on “educational value” to help educators choose apps that are, in fact, educational as determined by the science of learning (Hirsch-Pasek, et al., 2015).

Though most know our technologically-rich society requires a more technologically-literate educator, teacher preparation has not yet caught up with this reality. There are many reasons for this gap: dwindling higher education budgets that make continually updating technology difficult; preservice students coming into teacher preparation programs with varying competencies and understandings of appropriate technology use (Carroll & Morrell, 2006); and teacher preparation/education faculty members who may have little competence with emerging technologies (Polly et al., 2010).

In order to support preservice teachers (PTs) in developing competency in technology integration we make four suggestions—PTs must:

- Become knowledgeable about what developmentally appropriate practice looks, feels, and sounds like (NAEYC, 2009);
- Explore and develop knowledge about ways experts authentically utilize technologies to consume, create, and connect in their areas of expertise;
- Engage in hands-on experiences with a range of technologies across content areas to build their level of comfort; and,
- Demonstrate their competency in integrating technology with content in developmentally appropriate ways.

This chapter provides a framework for how we construct this foundational understanding with PTs. It shares exemplars-in-practice, and suggests needed changes in order to support this practice across institutions of higher education.

BACKGROUND

We utilize the Technological Pedagogical Content Knowledge (TPACK) framework (Figure 1, Mishra & Koehler, 2006) as a theoretical lens for our argument. TPACK represents one (of many) useful models for thinking about specific knowledge teachers must have to effectively and meaningfully integrate technologies into their

instruction. Every lens and model has its shortcomings and/or pitfalls (Schmidt, et al., 2009), but we selected TPACK because its compartmentalization and overlapping of three main pieces—Content Knowledge, Technical Knowledge, and Pedagogical Knowledge—are relevant given the ways in which American sets of standards for teacher preparation and student achievement have parsed out content knowledge as a distinct part of teaching and learning and have integrated pedagogy and technology into their standards. The framework holds the most effective way to integrate technology is for teachers to simultaneously draw on their technological, pedagogical, and content knowledge.

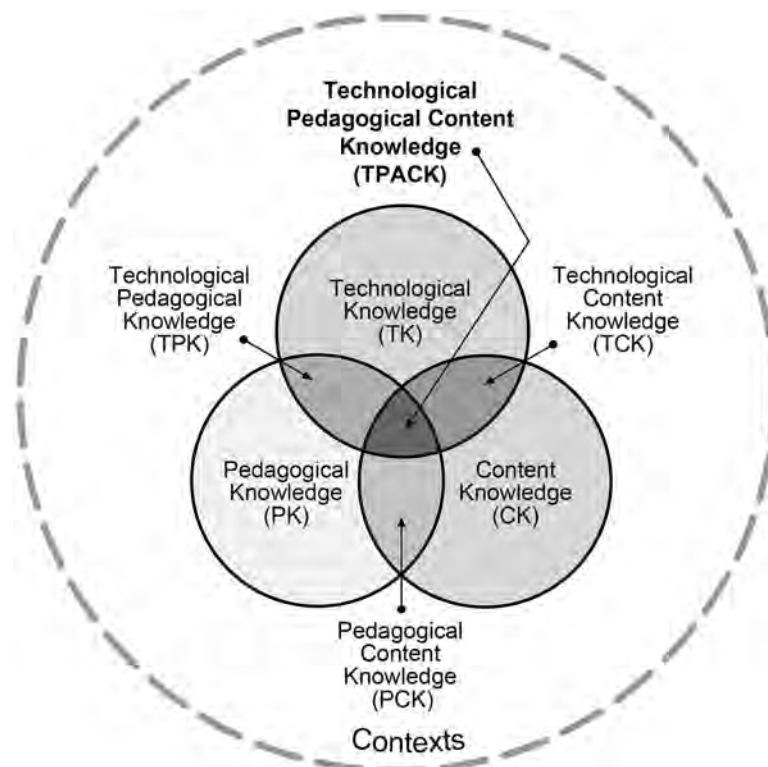
The TPACK model has been successfully applied to inservice teacher training (Hutchison & Woodward, 2014; Niess, 2011) and has been used to explore the growth and development of PT's application of technology in P-12 classrooms as well

(Koh & Divaharan, 2011; Pamuk, 2011). Studies of TPACK have suggested (1) it is difficult to affect change in PTs' technological content knowledge (Hofer & Grandgenett, 2012); (2) higher TPK scores were apparent in lesson plans associated with models of teaching with which they had the most familiarity as learners themselves (Lee, et al., 2014); and (3) technology-infused approaches to teacher preparation yielded more rapid increases in CK and PK, thus impacting the depth of the PT's understanding of the content as well as the ways in which the methods/pedagogy are effective to support content area learning.

The ways in which the school-aged students will be required to utilize the technology impact the requirements for teacher preparation. In America, the Common Core State Standards (National Governors Association Center & Council of Chief State School Officers, 2010) drive the integration of technology/digital components in reading, writ-

Figure 1. The Technological Pedagogical Content Knowledge (TPACK) framework

Note: Image reproduced by permission of the publisher, © 2012 by tpack.org.



ing, language, and speaking & listening (Hoffman, et al., 2014) in many states. Here, the idea is that *all* professions utilize reading, writing, language, and speaking & listening to accomplish their work. Some young children come to school with informal and play-based experiences (Marsh, et al., 2015; Plowman, 2012) in operating a range of devices, coding, digital composition, gaming, viewing, digital reading, online researching so it is imperative that PTs understand these funds of knowledge young children may bring to formal learning contexts.

In order to become a licensed teacher in America, PTs are to demonstrate competency of a range of technology standards (see Council of Chief State School Officers, 2011). In these standards, “technology” is broadly mentioned four times: as a tool for assessment and differentiation; as a support for content and skill development; as a tool for accessing, interpreting, evaluating, and applying information; and as a subject of study (i.e. digital citizenship).

In university-based teacher preparation and certification, coursework often includes a stand-alone course (carrying 1-4 credit hours) in which PTs evaluate and apply technology to content-area lessons. Some institutions integrate technology into their methods blocks as well (and with somewhat encouraging results [Wetzel, et al., 2014]). So, while the TPACK framework (Figure 1) is useful for thinking about what knowledge teachers must have to effectively and meaningfully integrate technologies into their instruction, we suggest teacher preparation programs might take it a step further through applying constructivist theories (Dewey, 1938; Malaguzzi, 1993) of teaching and learning.

TECHNOLOGY INTEGRATION AS A PRESERVICE COMPETENCY

This section presents three components of technology integration in PT coursework that work to unpack the developmental appropriateness of

technology and pull broadly on elements of socio-constructivist pedagogies. These components allow PTs room for exploration, tinkering, and talking about the technologies while learning how they can be utilized as tools to consume, create, or connect people across space and time.

The inspiration for these components is rooted in the Reggio Emilia Approach (REA). The Reggio Emilia schools (in Italy) have a well-known and respected approach that values learners as strong and capable individuals who are co-learners alongside their teachers (Edwards, Gandini, & Forman, 2012). The REA suggests learning takes place through the exploration of many “languages of learning,” referring to the different ways learners construct knowledge of the world around them. Technologies are simply another language/tool for learning.

We encourage and support teacher preparation faculty in becoming familiar with and utilizing technology resources to support their own practice. If faculty are not proficient in at least some technology it is very difficult for them to support appropriate use of technology with PTs and they may be reluctant to use technology or to expose their lack of knowledge to their students. To counter this, Polly and colleagues (2010) present a framework in which faculty are given time to work with the technology themselves and learn from knowledgeable faculty or even students in order to support them in taking risks by incorporating technology into their own teaching.

Third, we require PTs to actually practice technology both as a tool for carrying out their professional work and for teaching content (Puerling & Fowler, 2014; Wetzel, et al., 2014). In this way, the PT is engaged in real-life, culturally important, hands-on experiences in which he or she seeks to construct understanding about the world around him/her--two features of developmentally appropriate practice (NAEYC, 2009). Our PTs are tasked with utilizing technology for personal communication and productivity, entertainment, socialization, research, assessment, and teaching (among other things).

SOLUTIONS AND RECOMMENDATIONS

The following exemplars-in-practice are used to illustrate how our PT programs start to intentionally and systematically support PTs' construction of TPACK knowledge (presented in Figure 1) by utilizing the REA. Our PTs: (1) explore how professionals authentically use technology; (2) build their own background knowledge about a wide range of hardware, devices, and software; and (3) tinker with technology across diverse content and with myriad pedagogical approaches in preservice preparation. In these ways, a socio-constructivist lens overlays onto the TPACK framework.

Exploring Authentic Use of Technology Across Professions

In the foundational courses of teacher preparation, faculty utilize concept-mapping as a strategy to document what PTs know about a range of professions. They view a video clip that was intended to teach children about a profession (the pediatrician field trip on *Mr. Rogers' Neighborhood*) and consult online reading material ("Doctors" on PebbleGo.com; e-how articles). After watching the clip, faculty models for and guides PTs through mapping out who, what, where, why, and how each profession is carried out.

An essential component of the 'how' part of the concept-maps involves looking at the technology utilized in each profession and its relationship to the professional's end goal. Traditional paper and markers are utilized for thinking through one of the explored professions, but other mapping software are presented at each station/profession (Poplet, Mindmup [see example, Figure 2], Explain Everything, etc.) and then the software are replicated on multiple devices (computer with wired/wireless mouse, iPad, iPod Touch; laptop). For example, the PTs might explore how:

- A scientist uses a microscope to zoom in, captures the image with a screenshot, and

uses overlay/annotations to document their findings;

- A doctor utilizes some low-tech technologies (otoscope) to determine health and documents findings on an electronic medical record (see computer-based mapping example, Figure 2);
- A journalist uses multiple outlets to communicate their messages and connect to others (oral or written news interviews, podcasts, twitter or other social media, documentaries);
- A musician uses sound mixing software to record and edit a track.

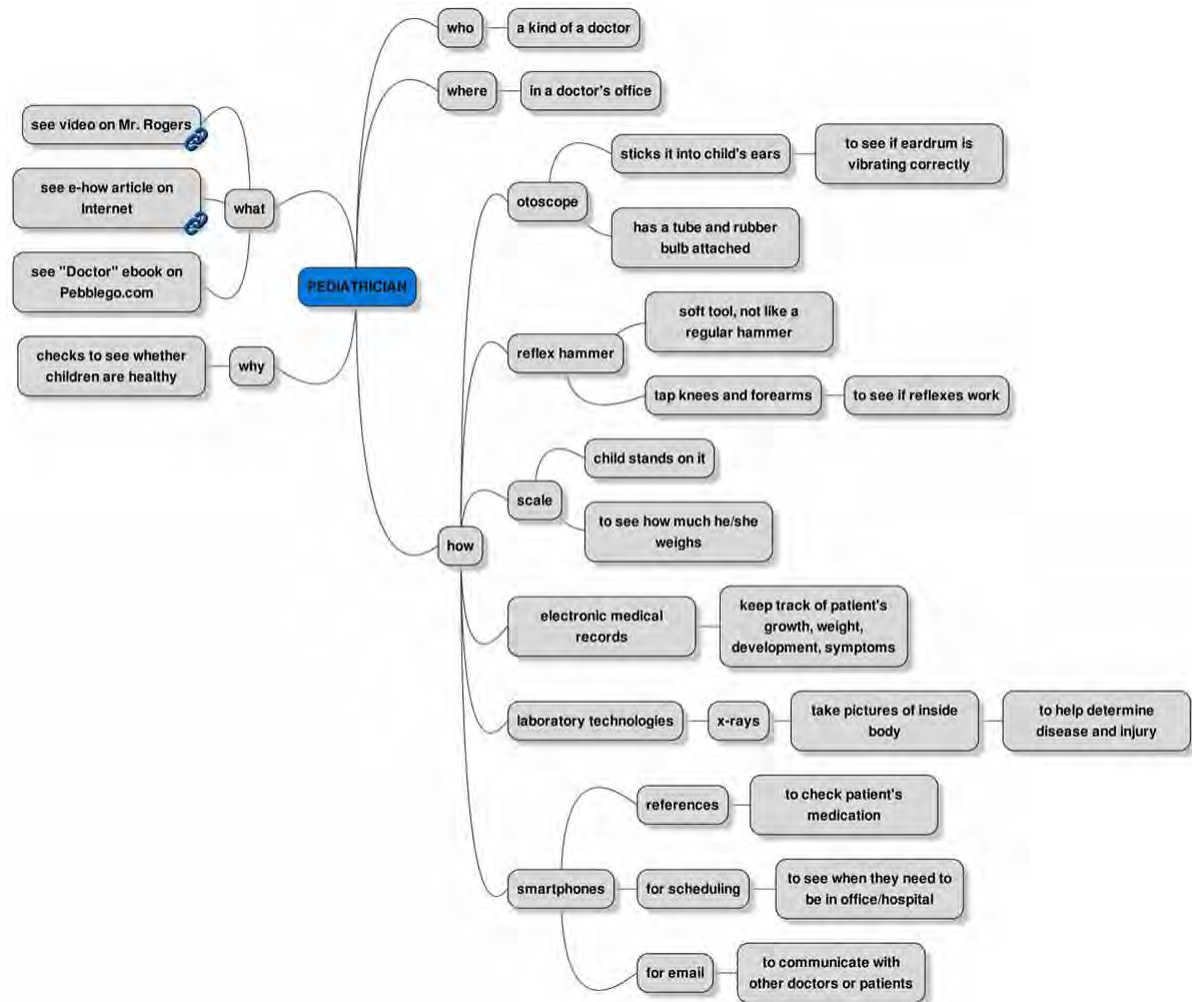
In this example, the technology/media (video-based media clips and a range of software designed for mapping), content (social studies related to community workers/jobs), and the faculty's modeling of pedagogy (modeled and shared experiences with the technology, guided practice to try it out, and independent practice) are integral components that support the PT's development in multiple areas of the TPACK framework, applying a constructivist approach.

Building Background Knowledge About Technology

As indicated in the above example, faculty seamlessly weaves in the range of hardware, devices, and software into the guided practice/exploration component of the lessons. Access, guided exploration, and intentional use of a range of technologies and media ensures that PTs are able to build and construct their background knowledge about the affordances and limitations of each piece of hardware or software.

PTs engage in backchannel discussions (through Twitter, Padlet, and/or a Today's Meet applet embedded into our course learning management system) to document their connections, questions, and thoughts related to course content. PTs utilize Twitter to access readings/news articles related to education and young children and then

Figure 2. Map of “pediatrician” using Mindmup application on a computer



they engage in a backchannel conversation using a hash tag to easily communicate with the group. As an extension and a meaningful way to connect to practical uses of these tools in early education contexts, PTs view and discuss a video of a classroom of first graders engaged in this process as their teacher reads aloud (Harvey et al., 2013).

Trying it out in multiple ways contributes to the PT's understandings of the affordances and constraints of each hardware and software. The PTs then relate those to theories of child development, pedagogical knowledge, and content knowledge in order to determine whether the technology might be useful to augment, modify, or redefine the child's learning process (Puentedura, 2015).

Tinkering While Learning How to Teach

A socio-constructivist view of learning involves constructing knowledge through language-facilitated social experiences (Vygotsky, 1978). For PTs to be prepared to incorporate technology into their classrooms it is less important to 'train' them how to use a specific piece of hardware or software than it is to practice engaging with unfamiliar or newly emerging hardware and software with the mindset to determine the affordances and drawbacks as related to a particular student/class of children. PTs who become comfortable with the process of tinkering to construct knowledge

about these tools are much better able to use them with intention and also to approach unfamiliar technology in the future with more confidence. Some examples of how PTs have tinkered with the technology for various purposes through their preservice programs follow (see also Wilkinson & Petrich, 2013).

Consuming New Information and Supporting Exploration of the World Around

Identifying and consuming new information represents one viable method of acquiring the answers to the inquiries that occur during learning. People from nearly all professions now consume new information from electronic resources—news-papers/magazines; online encyclopedias, wikis, and journals; e-books; online tutorials; and PDF files. These are all used to research and acquire new information about the world around (Pew Research Center, 2014). In the real world, people harness technology to acquire and consume large amounts of new information. When their search strategies are effective (Henry, 2006), they can end up with visual, printed, and audio forms of information on just about any topic.

And, while the reality of “the real world” exists, a large disconnect remains between how people in “the real world” find and consume their information, and the “school world” parallel experience. For example, most PT preparation programs do not require teachers to purchase digital textbooks, yet classrooms as young as Kindergarten and first grade are utilizing digital textbooks for students in multiple content areas (see <http://www.discoveryeducation.com/>, for an example). Moreover, it is far less common to see early childhood teachers reading aloud from an e-book than it is to see them reading from a printed text (Roskos, 2013). To help bridge this gap and to learn through a more constructivist lens, PTs engage in several hands-on experiences.

First, PTs read a PDF version of a children’s book and use the comment and annotate tools

in the program to document their connections, questions, predictions, and vocabulary learning. Next, PTs utilize a web-based electronic version of a children’s chapter book (<http://www.wenovel.com/tag/Charlotte-s-Web.html>) and a web annotation applet (<https://www.diigo.com/index>) to extend the digital reading experience to another device. After, PTs explore a third children’s title using an e-reader application and the built-in text annotation tools. PTs are required to share their text on a projection screen (link from their own device to the teacher podium), and conduct a read aloud of a section of one of the books to the class, thinking aloud about their navigation through the digital text in addition to their comprehension or vocabulary focus. After the three e-reading experiences, we tinker with sharing out our annotations with others in the class via social media outlets.

As a second way to build knowledge about consuming information through electronic sources, PTs tinker with searching for various kinds of information through hash tags on Twitter, Pinterest boards for science experiment ideas, YouTube videos to build playlists for “virtual field trips” (Blachowicz & Obrochta, 2005), conventional library-based searches, and even exploring how different search engines yield different kinds of information on the same key search terms.

As a part of learning about being a good consumer of digitally packaged texts, PTs tinker with at least 5-10 different applications in their methods courses and discuss the quality of the software and identify risks to children (apps with advertising, links to social networks, requesting private information). PTs arrive at the conclusion that the software ranges from inaccurate and distracting to thoughtfully designed and engaging (Hoffman & Paciga, 2014; Paciga & Hoffman, 2014; Schuler, 2012). They are introduced to the position statement on technology use in early education, (NAEYC & Fred Rogers Center, 2012), Common Sense Media (n.d.), and *Children’s Technology Review* (n.d.) as ways to help determine the level of developmental appropriateness of software, apps, and media for their students; our claim is

not that “all PTs need to know THESE resources,” but, rather, that PTs should become familiar with these *kinds* of resources in order to best support their teaching.

Connecting and Communicating Through Technology

In addition to consuming new information with technology, PTs can benefit from tinkering with technology to (1) connect to others and (2) share their messages to those in their networks. In their professional lives, teachers might connect and communicate for a couple of purposes exemplified in the following paragraphs—to communicate daily happenings with parents and other stakeholders, and to communicate a child’s achievements and interests to parents or guardians.

The practical uses of technology can enhance the teachers’ ability to connect (Paciga, 2016). When considering teaching and learning from an REA perspective, one area of pedagogical practice where this connection is very clear is in working with parents and families, as parents and families are considered co-learners alongside their children. For this reason, we consider it especially prudent for PTs to be prepared to help parents understand some of the challenges and barriers they might face in relation to parents’ technology-mediated work with their children. To help prepare them for conversations around parental challenges and barriers, PTs research and pull together a Public Service Announcement pamphlet, video tutorial, or infographic related to technology access, use/over-use, or troubleshooting program specifics.

In many of the required courses, PTs work to develop their knowledge of how to work with families, focusing on the importance of having open lines of communication, building positive relationships. PTs are asked to practice some strategies (writing letters to parents, creating newsletters) to achieve these goals once licensed practitioners. PTs also engage in applying these strategies for communicating and building relationships through technology-applied means--

communicating via email, blogs, text messages, and/or shared calendars. PTs also tinker with photo and video sharing sites and social media pages (Puerling, & Fowler, 2014).

PTs view the Explore More Learning Center (<http://exploremorelearning.bravesites.com>) and The Blue School (<http://www.blueschool.org/>) websites and social media outlets to construct knowledge about connecting and communicating with parents in a technology-rich society—to explore “school appropriate” uses. To help PTs develop competency in these forms of communication and to understand and work with policies related to privacy and children’s identities being put onto social media, PTs practice blog writing—embedding shared calendars, and images of the children’s work/explorations. They also explore strategies for safeguarding children such as taking pictures from behind to show children’s hands and their work, cropping out faces, or blurring out the children’s faces as alternative ways to still utilize the technologies, but remain within the parameters of the school’s policies.

Sharing a child’s learning and interests with their parents is a more specific/narrow form of connecting and communicating, but arguably as important as communicating the general happenings/announcements of a school day. There has been a huge shift in the assessment of young children (see, Paciga, et al., 2015)—away from more authentic forms of assessment such as observation, anecdotal notes, and work samples and toward more data driven systems of checklists and testing. Innovative technologies have the potential to shift assessment back to more developmentally appropriate practices by using authentic assessment applications that can just as easily quantify authentic assessments into hard data; we’ve worked in Evernote to manage our authentic assessments in our college-level coursework (adding tags for goals, students, and inserting audio/video/visual evidence), but there are several electronic portfolio applications currently available and more on the horizon.

T

To support PTs in this regard, faculty use some of the same tools we expect PTs to use in documenting PT progress in preservice coursework—photos, videos, and audio recordings—to provide a richer and more authentic assessment of student progress. The earlier these authentic documentation tools appear in teacher preparation programs, the more confident PTs will become with using the technology; this level of confidence is very important in implementing it into the classroom later (Ertmer & Ottenbreit-Leftwich, 2010).

Creating to Demonstrate Understanding and Problem Solving

The third example of tinkering relates to using technology to create. PTs build an incredible amount of knowledge in a technology integration course by (1) exploring something low-tech, such as an overhead projector, tape measure, flashlight; (2) thinking about possible uses for the object/tool beyond the typical and conventional; (3) using equipment that is more high-tech to record a short “how-to” video for their low-tech object; (4) using movie editing tools to package the video (faculty offer verbal scaffolds as necessary); and (5) publicizing/sharing their video (posted on the class learning management system or YouTube). This creates a situation in which the technology can be used with purpose, but with low stakes attached, fostering more risk-taking and exploring. In addition, PTs are simultaneously developing their understandings of recording, editing, and publishing video—something the PTs engage in as part of the assessment process during student teaching.

Another way PTs have constructed knowledge in this domain is through exploration of a piece of software (Book Creator or Explain Everything) to document their own learning. PTs were instructed to use iPads to document their observations about patterns and numbers in nature while on a hike (i.e., photos, videos, or voice memos with illustrations). Afterward, they took time to explore three

applications that could be used for communicating their learning to a wider audience. Each application had options to import images, add audio tracks, and/or insert text/overlay. PTs chose two of the applications to explore and then presented their observations about patterns and numbers to the larger group by using one application. PTs identified and discussed what they found challenging or intuitive about the applications and how they might overcome or harness those when using the software applications with children. The spin-off of these experiences was that several students began utilizing these applications during their student teaching experiences as a way to document the class' learning through shared writing experiences.

FUTURE RESEARCH DIRECTIONS

There is room for additional research studying the design of coaching/professional development infrastructures as related to technology integration. For example, are there ways in which teacher education programs can use different role modeling strategies and experiences to support understanding? How can mentor teachers and professional learning networks circumvent the under-prepared teacher-preparation faculty? In what ways might teacher preparation coursework facilitate a shift in the societal perceptions of teacher as “knower” toward teacher as “facilitator”? As some research already indicated, this is an institutional problem and relates to the university's ability to develop consistent strategies for ensuring technology integration by all faculty members (Lavonen, et al., 2006). More comprehensive research on how successful institutions have developed strategies to overcome these obstacles would be incredibly helpful.

Also, it is prudent to explore the impact of reflection and analysis of technology integration on PTs' utilization of technology in the content areas. Though the research shows reflection is an

important part of developing an understanding of technology integration (Goktas et al., 2009), more research in this area will help teacher education programs develop strategies for more effective reflective practice.

CONCLUSION

Technology can be a great tool for teaching and learning, but only when teachers understand it, know how to use it and can integrate it in developmentally appropriate ways. This chapter suggested several areas of development that need to be addressed by teacher preparation programs in order to support better technology integration into the early childhood classroom. We provided examples of ways that our teacher preparation programs work to construct this knowledge with preservice teachers—supporting preservice teachers’ understanding of developmentally appropriate technology practices, knowledge of authentic technology uses, and participation in hands-on technology experiences that connect knowledge or theory to practice.

ACKNOWLEDGMENT

A portion of this work was supported by an Early Career Research Fellowship awarded to K. Paciga by the Fred Rogers Center for Early Learning and Children’s Media at Saint Vincent College and the TEC Center at Erikson Institute, with funding provided by The Grable Foundation of Pittsburgh.

REFERENCES

Blachowicz, C., & Obrochta, C. (2005). Vocabulary visits: Virtual field trips for content vocabulary development. *The Reading Teacher*, 59(3), 262–268. doi:10.1598/RT.59.3.6

Buss, R. R., Wetzell, K., Foulger, T. S., & Lindsey, L. (2015). Preparing teachers to integrate technology into K–12 instruction: Comparing a stand-alone technology course with a technology-infused approach. *Journal of Digital Learning in Teacher Education*, 31(4), 160–172. doi:10.1080/21532974.2015.1055012

Carroll, J., & Morrell, P. (2006). A comparison of teacher education faculty and preservice teacher technology competence. *Journal of Computing in Teacher Education*, 23(1), 5–10.

Children’s Technology Review. (n.d.). *CTRX*. Available at <http://reviews.childrenstech.com/ctr/home.php>

Common Sense Media. (n.d.) *Reviews & age rating*. Available at <https://www.commonsense-media.org/>

Council of Chief State School Officers. (2011). *inTASC Model Core Teaching Standards 2011*. Available at http://www.ccsso.org/resources/programs/interstate_teacher_assessment_consortium_%28intasc%29.html

Dewey, J. (1938). *Experience and education*. New York: Collier.

Edwards, C., Gandini, L., & Forman, G. (Eds.). (2012). *The hundred languages of children: The Reggio Emilia experience in transformation*. Santa Barbara, CA: Praeger.

Ertmer, P., & Ottenbreit-Leftwich, A. (2010). Teacher technology change: How knowledge, confidence, beliefs, and culture intersect. *Journal of Research on Technology in Education*, 42(3), 255–284. doi:10.1080/15391523.2010.10782551

Goktas, Y., Yildirim, Z., & Yildirim, S. (2009). Main barriers and possible enablers of ICTs integration into pre-service teacher education programs. *Education, Technology and Society*, 12(1), 193–204.

- Harvey, A., Goudvis, S., Muhtaris, K., & Ziemke, K. (2013). *Connecting comprehension and technology: Adapt and extend toolkit practices*. New York: Heinemann.
- Henry, L. A. (2006). SEARCHing for an answer: The critical role of new literacies while reading on the internet. *The Reading Teacher*, 59(7), 614–626. doi:10.1598/RT.59.7.1
- Hirsh-Pasek, K., Zosh, J. M., Golinkoff, R. M., Gray, J. H., Robb, M. B., & Kaufman, J. (2015). Putting Education in Educational Apps Lessons From the Science of Learning. *Psychological Science in the Public Interest*, 16(1), 3–34. doi:10.1177/1529100615569721 PMID:25985468
- Hofer, M., & Grandgenett, N. (2012). TPACK development in teacher education: A longitudinal study of preservice teachers in a secondary MA Ed. program. *Journal of Research on Technology in Education*, 45(1), 83–106. doi:10.1080/15391523.2012.10782598
- Hoffman, J. L., & Paciga, K. A. (2014). Click, swipe, and learn: Sharing e-Books with toddlers and preschoolers. *Early Childhood Education Journal*, 42(6), 379–388. doi:10.1007/s10643-013-0622-5
- Hoffman, J. L., Paciga, K. A., & Teale, W. H. (2014). Common Core State Standards and early literacy instruction: Confusions and conclusions. Chicago, IL: University of Illinois at Chicago Center for Literacy. Available at <http://cfl.uic.edu/publications/research-and-policy-reports/>
- Hutchison, A., & Woodward, L. (2014). A Planning Cycle for Integrating Digital Technology Into Literacy Instruction. *The Reading Teacher*, 67(6), 455–464. doi:10.1002/trtr.1225
- Koh, J. H. L., & Divaharan, S. (2011). Developing preservice teachers technology integration expertise through the TPACK-developing instructional model. *Journal of Educational Computing Research*, 44(1), 35–58. doi:10.2190/EC.44.1.c
- Lavonen, J., Lattu, M., Juuti, K., & Meisalo, V. (2006). Strategy-based development of teacher educators ICT competence through a cooperative staff development project. *European Journal of Teacher Education*, 29(2), 241–265. doi:10.1080/02619760600617433
- Lee, K. S., Smith, S., & Bos, B. (2014). Pre-service teachers' technological pedagogical knowledge: A continuum of views on effective technology integration. *International Journal of E-Learning & Distance Education*, 29(2).
- Malaguzzi, L. (1993). For an education based on relationships. *Young Children*, 49(1), 9–12.
- Marsh, J., Plowman, L., Bishop, J. C., Lahmar, J., Scott, F., Davenport, A., ... Robinson, P. (2015). *Exploring play and creativity in pre-schoolers' use of apps: Final project report*. Retrieved from http://www.techandplay.org/reports/TAP_Final_Report.pdf
- Massachusetts Institute of Technology. (2015). *Groundbreaking technologies 2015*. Retrieved on April 2015 from <http://www.technologyreview.com/lists/technologies/2015/>
- Mishra, P., & Koehler, M. (2006). Technological Pedagogical Content Knowledge: A new framework for teacher knowledge. *Teachers College Record*, 108(6), 1017–1054. doi:10.1111/j.1467-9620.2006.00684.x
- NAEYC. (2009). *Developmentally appropriate practice in early childhood programs serving children from birth through age 8*. Washington, DC: Author.
- National Association for the Education of Young Children (NAEYC) & Fred Rogers Center for Early Learning and Children's Media at Saint Vincent College. (2012). *Technology and interactive media as tools in early childhood programs serving children birth through age 8*. Washington, DC: Author. Available at <http://www.naeyc.org/content/technology-and-young-children>

- National Governors Association Center & Council of Chief State School Officers. (2010). *Common Core State Standards for English Language Arts, & Literacy in History/Social Studies, Science, and Technical Subjects*. Retrieved from <http://www.corestandards.org/the-standards/english-language-arts-standards>
- Niess, M. L. (2011). Investigating TPACK: Knowledge growth in teaching with technology. *Journal of Educational Computing Research*, 44(3), 299–317. doi:10.2190/EC.44.3.c
- Paciga, K., & Hoffman, J. (2014). Innovations in literacy: Realizing the potential of e-books in early education. In M. Khosrow-Pour (Ed.), *Encyclopedia of Information Science and Technology* (3rd ed., Vol. 7, pp. 4787–4796). Hershey, PA: IGI Global.
- Paciga, K., Hoffman, J., & Teale, W. (2015, March 30). *The CCSS don't say we should exclude the play*. Available at <http://preschoolmatters.org/?s=paciga>
- Paciga, K. A. (2016, January 5). *The power of social networks*. Available at <http://www.fredrogerscenter.org/2016/01/05/the-power-of-social-networks/>
- Pamuk, S. (2011). Understanding preservice teachers technology use through TPACK framework. *Journal of Computer Assisted Learning*, 28(5), 425–439. doi:10.1111/j.1365-2729.2011.00447.x
- Pew Research Center. (2014, December). *Americans feel better informed thanks to the Internet*. Available at <http://www.pewinternet.org/2014/12/08/better-informed/>
- Plowman, L. (2012). *Press for play: Using technology to enhance learning in the early years*. London: Technology Enhanced Learning Research Programme. Retrieved from http://tel.ioe.ac.uk/wp-content/uploads/2011/11/PressForPlay_WebTP.pdf
- Polly, D., Mims, C., Shepherd, C., & Inan, F. (2010). Evidence of impact: Transforming teacher education with Preparing Tomorrows Teachers to Teach with Technology (PT3) Grants. *Teaching and Teacher Education: An International Journal of Research and Studies*, 26(4), 863–870. doi:10.1016/j.tate.2009.10.024
- PuenteDura, R. (2010). *SAMR and TPCK: Intro to advanced practice*. Retrieved January 16, 2015 from http://hippasus.com/resources/sweden2010/SAMR_TPCK_IntroToAdvancedPractice.pdf
- Puerling, B., & Fowler, A. (2014). Technology tools for teachers and teaching: Innovative practices and emerging technologies. In C. Donohue (ed.), *Technology and digital media in the early years: Tools for teaching and learning*. New York: Routledge & NAEYC.
- Sasseen, J., Olmstead, K., & Mitchell, A. (2013). *The state of news media 2103*. The Pew Research Report. Available at <http://www.stateofthemedias.org/2013/digital-as-mobile-grows-rapidly-the-p pressures-on-news-intensify/>
- Schmidt, D. A., Baran, E., Thompson, A. D., Mishra, P., Koehler, M. J., & Shin, T. S. (2009). Technological pedagogical content knowledge (TPACK): The development and validation of an assessment instrument for preservice teachers. *Journal of Research on Technology in Education*, 42(2), 123–149. doi:10.1080/15391523.2009.10782544
- Shuler, C. (2012). *iLearn II: An analysis of the education category of the iTunes app store*. Available at <http://www.joanganzcooneycenter.org/publication/ilearn-ii-an-analysis-of-the-education-category-on-apples-app-store/>
- Vaala, S. E., Ly, A., & Levine, M. (2015). *Getting a read on the app stores: A market scan and analysis of children's literacy apps*. New York, NY: The Joan Ganz Cooney Center at Sesame Workshop. Retrieved from http://www.joanganzcooneycenter.org/wp-content/uploads/2015/12/jgcc_gettingaread.pdf

Vygotsky, L. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.

Wetzel, K., Buss, R., Foulger, T. S., & Lindsey, L. (2014). Infusing educational technology in teaching methods courses: Successes and dilemmas. *Journal of Digital Learning in Teacher Education*, 30(3), 89–103. doi:10.1080/21532974.2014.891877

Wilkinson, K., & Petrich, M. (2013). *The art of tinkering*. San Francisco, CA: Weldon.

ADDITIONAL READING

Donohue, C. (Ed.). (2014). *Technology and digital media in the early years: Tools for teaching and learning*. New York: Routledge.

Edwards, S., & Bird, J. (2015). Observing and assessing young children's digital play in the early years: Using the Digital Play Framework. *Journal of Early Childhood Research*, Online First Edition. doi:1476718X15579746.

Hong, S. B., & Trepanier-Street, M. (2004). Technology: A tool for knowledge construction in a Reggio Emilia inspired teacher education program. *Early Childhood Education Journal*, 32(4), 87–94. doi:10.1007/s10643-004-7971-z

Puerling, B. (2012). *Teaching in the digital age: Smart tools for age 3 to grade 3*. Redleaf Press.

Wohlwend, K. (2013). *Literacy playshop: New literacies, popular media, and play in the early childhood classroom*. New York: Teachers College Press.

KEY TERMS AND DEFINITIONS

Preservice Teacher: Student pursuing a degree for initial certification/licensure in teaching field (early childhood education, elementary education).

Reggio Emilia Approach [REA]: An educational philosophy focused on early education that was founded by Loris Malaguzzi in Reggio Emilia, Italy; its fundamental principles include that (1) children are capable of their own learning; (2) children form their understandings through their interactions with others; (3) children are communicators; (4) the child's environment teaches; (5) the adult is a mentor and guide; (6) documenting children's expression of knowledge helps guide the child's learning.

Tinkering: Process by which students explore, create, and construct knowledge about the world around them.

Effectiveness of Teacher Training in Using Latest Technologies

T

Revathi Viswanathan

B. S. Abdur Rahman University, India

INTRODUCTION

Teachers who teach English as a second language (ESL) in a developing nation often feel the need for updating their knowledge of new teaching trends by attending in-service training programs or professional development courses. It is believed that such learning paves the way for providing a curriculum that integrates the use of technology within and beyond the language classroom. In this context, it must be stated that teacher training programs prove to be effective when they offer practical experience to teachers in new teaching methods. OECD (2009)'s comparative report defined effective professional development as, "on-going, includes training, practice and feedback, and provides adequate time and follow-up support. Successful programmes involve teachers in learning activities that are similar to ones they will use with their students, and encourage the development of teachers' learning communities" (p.3). Mathew (2014) defines the terms, teacher training, teacher education and teacher development. She quotes Widdowson (1983) and Richards and Nunan (1990), according to whom training, "deals with familiarising student teachers with techniques and skills to apply in the classroom" (p.29). Mathew explains teacher development as "a voluntary process, ongoing, bottom-up, since the starting point is the teachers' own experience where new information is sought, shared, reflected on, tried out, processed in terms of personal experience and finally 'owned' by the teachers" (p.29). Similarly, Evans (2002) quotes Grossman (1994, p. 58) according to whom professional development of experienced teachers could be in

the form of "workshops, study groups, Ž reside chats, a district-wide colloquium for middle school teachers, action research projects, and conversations with the professor-in-residence ...'(p.3). A professional development experience could be presented as a workshop or other formally related meetings. (Quattlebaum, 2012). In this chapter, the author would elaborately discuss the workshop, which she conducted for teachers (who represented various Indian states) and trained them in using the hand-held devices for teaching language skills. Further, she would throw light on the responses given by the participants, which reflected the effectiveness of the program.

BACKGROUND

Need for Teacher Development

Leadership and Teacher Development Branch (2005) believe that the high quality of professional learning helps teachers to improve their effectiveness in teaching. Carlson and Gadlo (n.d) consider the importance of teacher training in technical skills and say that teachers "need professional development in the pedagogical application of those skills to improve teaching and learning" (p.119). They advocate the need for providing lifelong professional preparedness to teachers through a three dimensional approach of giving pre-service training, in-service training through workshops, seminars and short courses that provide pedagogical and technical support to teachers. Teachers have to plan their teaching based on learners' needs and learning styles. They

DOI: 10.4018/978-1-5225-2255-3.ch664

need to expand their knowledge by learning the changing trends in teaching and learning processes. With the emphasis on using digital technology, it is advisable for teachers to get trained in using the latest technologies in order to provide individualized learning and instant feedback to students' performance. Sanders and McCutcheon (1986) classify teaching practices in three ways and one among them is 'structuring practices' which discusses the need for creating conditions for teaching-learning processes to take place. Training programs are nowadays conducted both online and face-to-face for educating teachers in using advanced technology and it is believed that such hands-on experience would help them to create proper learning conditions to students.

Teacher development programs have relevance to Mezirow (1997)'s theory of transformational learning. According to the theory, faculty development programs "aim at changing individuals' thoughts and actions" (Saroyan, Amundsen, & Li, 1997, p.96). Mezirow believed in the aspect that adult learners would change their practices when they change their basic assumptions about learners, role of teachers and the educational goal. Applying the transformational theory for conducting teacher training programs in using technology, it could be said that it educates teachers about the need for using the required tools for promoting students' language skills, students' level of comfort in using the devices, the way to integrate the use of tools with regular teaching and the extent to which technology would enhance their teaching style.

Saroyan, Amundsen and Li (1997) discuss the faculty development program that they designed based on Ramsden (1992)'s theories and Mezirow's transformational theory. It was a product-oriented program and focused on enabling participants to design a course and demonstrate skill and exhibit self-confidence in teaching it. In the micro-teaching session, they were asked to select and use the strategies which would foster learning among students. They were encouraged to give a self critique besides getting the feedback from peers. They were asked to give their inputs

in the beginning and after the program so that their change of perception towards teaching could be evaluated. The data collected was related to the instructional formats that were developed on the basis of Ramsden's and Mezirow's theories. It was evident from the analysis of the feedback given by the participants and teachers that they had changed their view of teaching as knowledge transmission to more focused activity of attaining learning outcomes.

Considering the need for educating teachers in using mobile technology, it is worthwhile to mention that Montrieux, Vanderlinde, Schellens and De Marez (2015) conducted a qualitative focus group study for reviewing the impact of tablet devices towards teaching and learning practices. They used Vanderlinde and Braak (2010)'s e-capacity model for framing a few research questions while conducting the semi-structured interview with a select few teachers and students. According to the e-capacity model, teachers need to have the competence for handling technology and showing teacher development. The interview was conducted after the teachers and students made use of tablet devices for six months. The analysis of the responses given by teachers and students revealed that the introduction of innovative technology helps teachers to integrate technology with traditional teaching. It was found that the "innovative teachers" had adopted the role of coach and had changed their teaching style in order to mobilise students' learning processes.

Similarly, Lynch (2015) presents the outcome of the two studies which revealed that school students (35% of them according to the study) were more interested in doing teachers' lessons or activities by using their tablet.

Maduakolam & Bell (2003) discuss the product-based faculty professional development model that was designed in order to train teachers to use and integrate advanced technologies into instruction and infuse technology into all teacher education courses. In the workshops that were conducted, the participants were given hands-on experience of creating instructional resources and

develop web sites using Microsoft Front Page and Dreamweaver. It was believed that such experience would create opportunity for the participants to provide the required resources to students. The participants were asked to use Microsoft Outlook to collaborate and communicate with other faculty members. The outcome of two years of training through workshops and implementation proved that the teaching and learning became very exciting and the faculty efficacy in technology usage and integration was very encouraging besides the networking done among faculty members.

It must be stated that the current workshop too could be related to the above-mentioned theory as it focused on changing tertiary level teachers' misconception of using mobile devices for teaching and educating them about new teaching technique to enhance students' language skills.

METHODOLOGY

Participants

The two-day workshop was conducted for teachers at the collegiate level and about 26 teachers representing various South-western states of India participated in it. These teachers were enrolled based on their primary interest in using ICT in the classroom and their willingness to use mobile phones in the teaching and learning process. They were asked to bring either a Smartphone or a Tablet or an iPad for getting hands-on experience in using them during the workshop.

The technical support was provided by the department of Computer Applications in B.S. Abdur Rahman University, where the workshop was organised. To explain, the wireless network facility (with two additional access points) was given in the seminar hall, which was used for conducting sessions on using mobile phones for accessing resources and applications. Similarly, the Internet connectivity in the language lab was also enhanced for the participants to access the free container

applications in the desktop for creating individual mobile apps for teaching and learning process.

T

Teacher Training Design

The aim of conducting a two-day workshop on Mobile Enhanced Language Learning (MELL) was to promote the concept of mobile learning, which is otherwise called as, M-Learning, among college teachers for teaching the language skills. It was believed that the training offered would develop teachers' teaching skills for facilitating learning even beyond the classroom. Carlson & Gadio (n.d) mention that, "teacher professional development in the use of technology should embody and model the forms of pedagogy that teachers can use in their classrooms" (p.121). They go onto specify the need for empowering teachers and including activities that develop their knowledge and skills, encourage them to use a variety of learning strategies, develop higher order thinking skills and demonstrate how technology facilitates teaching skills, to list a few. Crompton (2015) refers to the International Society for Technology in Education (ISTE), which specifies standards for coaches. It states that one of the ways by which technology coach could determine the professional development need of staff is, "design, develop and implement technology-rich professional learning programs that model principles of adult learning and promote digital age best practices in teaching, learning and assessment" (n.d). As stated in transformative theory, professional development programs should bring a few changes in teachers' teaching techniques. Villegas-Reimers (2003) believes that professional development programs that deal with, "activities that include both technical and conceptual aspects of instruction" (p.19) would prove to be most effective in sustaining change in teachers (Baker and Smith, 1999). To achieve this objective in this workshop, the tertiary level teachers were given hands-on experience in using the mobile phones in the three aspects such as 1. Using the mobile phone for browsing online

resources and activities, 2. Accessing mobile applications relating to language skills, and 3. Creating an app for handling language tasks. The purpose of focusing on this ubiquitous technology was to facilitate learning anywhere and everywhere. It is believed as Bjerede, Atkins & Dede (n.d) mention that “Student engagement in school will increase, as will educational outcomes, because the wireless devices build on strengths and preferences youth have developed by using these tools outside of classroom settings” (p.5).

The first aspect focused on encouraging teachers to search for useful language games and activities in their mobile phones and relate them to the learning modules. The resources included text, audio and video files that could be shared with students and encouraged to learn from those files. The second aspect focused on creating awareness of various language learning apps readily available in Google play store that could be installed in the mobile phones. For example, Google play store offers free language learning applications such as ‘Learn English’, ‘Speak English Daily’, ‘Real Business English’ to name a few, which provide learning content to students. Similarly, the play store also presents a few language teaching tools such as, ‘VoiceThread’, ‘ipadio’ and a few more that offer scope for training students in a particular language skill. ‘VoiceThread’ is a cloud application, which requires an updated version of Adobe Flash. A new user has to create a VoiceThread account with a username and password by accessing the web site www.voicethread.com. The user who wants to use the mobile app could register by accessing the app in the mobile and navigate to the account tab. He has to click on ‘join VoiceThread’ button and fill in the short registration form. The user can complete his registration by accessing the email and retrieving the verification message to confirm his account. This tool helps a user to create presentations, upload, share and discuss different types of media, which includes presentations, images, audio files and videos, and also comment on the slides created by a user. Similarly, ipadio is another application (available in android and

iOS versions) that allows its users to make phone calls and also share through social audio and video platforms. The user has to create an account for using the app. The recorded audio (in MP3, WAV and many more formats) and videos could also be broadcasted with the help of this application. It was felt that the exposure to such applications would help teachers to work on two things like prepare worksheets relating to the content present in the app and integrate mobile learning as part of teaching technique. The final aspect of the workshop that focused on creating an app was to enhance the ICT skills in language teachers and prepare them for sharing teaching content through the ubiquitous tool, mobile phone. As mentioned earlier, the current workshop was conducted with the consideration of above-mentioned aspects and the hands-on experience was provided by two teacher trainers, the author herself and a freelance consultant, Dr.P.N.Ramani.

Training

On the first day of the workshop, the participants were given a questionnaire and were asked to identify the scope of using technology (Appendix) based on their practical usage. The purpose was to make them self-assess their level of comfort in using a smart phone over a desktop. The questionnaire had five questions and the first question dealt with the extent to which the participants use a Smartphone. It had more than 30 options and the participants had to mark unlimited number of options based on their utility. The second question, which was close ended type, reviewed their preference in using a mobile phone over a desktop computer. The third question (an open-ended question) expected participants to list out the similarities between a desktop computer and a mobile phone. For the fourth question, the participants had to specify the multi-media features in a mobile phone and for the fifth question, they were asked to specify the infrastructural facilities required for using a mobile phone in a teaching and learning process. The answers given were

Figure 1. Screenshot of participants' involvement in a workshop session



discussed during the session and the participants were asked to retain the questionnaire for reassessing their perception of using Smartphones, after attending all the sessions. As a continuation to this discussion, an initial orientation on Mobile teaching and learning was given with the following objectives:

1. To create awareness of the advantages of using mobile technology.
2. To show the possibility of offering blended learning through the use of mobile devices.
3. To make participants realize the possibility of providing flipped learning through mobile devices.

4. To educate the teacher participants in technical know-how of providing an infrastructure for offering mobile learning to students.

The interaction along with the presentation provided lot of insights into the challenges faced by teachers in their institutions. While a few pointed out about the facilities available within the language classrooms, a few others shared their lack of knowledge of installing apps and the rest of the participants were apprehensive in using the mobile devices within the classroom. It was decided to address the challenges during the course of the workshop. After the introductory session, the participants were introduced to

Figure 2. Screenshot of participants' involvement in a workshop session



the resources that could be accessed online in the mobile devices. The participants worked in pairs and accessed those resources in their devices. The content was shared with others. They were also asked to focus on how the resources could be made use of while training students in language skills. Teachers shared a few language tasks that integrate the use of the resources accessed in the mobile phones.

In the next session, the resource persons provided input on the language application tools that are readily available in Google play store and iOS store. The features of those tools were discussed and the participants were asked to download and install in their mobile devices. They had to work in groups, explore the method of using the tools, and design one or two tasks pertaining to LSRW skills. Each group was assigned a skill and they had to work on it. It must be admitted that the participants worked with lot of enthusiasm and came out with two to three skill-based tasks. The groups shared their tasks with other group members too. The groups had to again work on the tasks to carry out suggestions given by other members. The participants were asked to keep the tasks safe so that they could be uploaded in the app, which would be created on the next day.

On the second day of the workshop, the participants were exposed to the designing of a free app using two different container apps, YAPP and Appypie. YAPP (which was available as a free container app until the early months of this year) was founded by Seidman and Melia in 2012. It could be downloaded in iOS and Android devices. The latter app, Appypie, is a paid container app and could be accessed in Windows phones or iPhones. It is worthwhile to mention that these container applications offer the facility to share the teaching and learning content through various pages created by a teacher. The objectives of using the applications were

- To enhance teachers' knowledge of ICT skills and progress to the next step of creating a teaching and learning environment in a mobile device.

- To show them the possibility of designing a mobile application without using html language.
- To demonstrate and provide hands-on experience to upload learning material in app.
- To show them the way to broadcast to the users about the material or task uploaded.

The session was conducted in the language lab that has high-speed Internet connectivity and the author demonstrated the way to create an app, create pages like Newsfeed, Simple text, upload Videos, provide links to Audio files and broadcast about the tasks uploaded in the pages. The participants created an app by logging into the websites, www.yapp.us and www.appiepie.com. They posted their images in the home page of the app, gave a name to their app, created a page for schedule, an invitation page for enrolling users and a simple text page.

As mentioned earlier, they uploaded the tasks prepared on the previous day in the simple text page created by them.

When the session continued in the seminar hall, the participants could broadcast the tasks and pages that they uploaded to the trainers and other participants through their app. This proved to be a very interesting and exciting experience to the participants.

Discussion

Harwell (2003) mentions that the content of a professional development program should “generate and contribute new knowledge to the profession and increase the ability to monitor students' work, in order to provide constructive feedback to students and appropriately redirect teaching”(p.4). She goes on to stress the effectiveness of a professional development program, which gives opportunity to teachers to learn new classroom practices. Considering the outcome of the workshop, it must be stated that the current workshop had achieved the objective of exposing

Figure 3. Screenshot of mobile applications created by participants

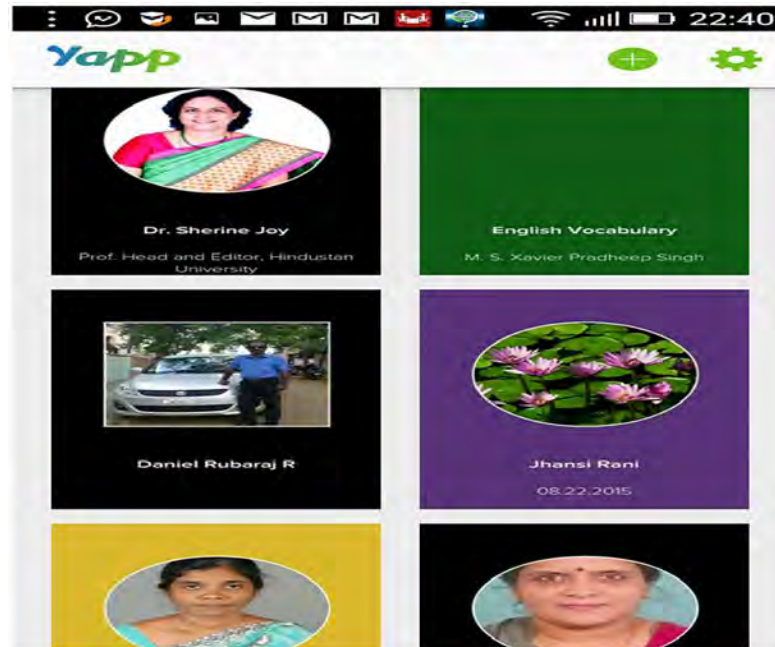


Figure 4. Screenshot of the worksheet uploaded in an app

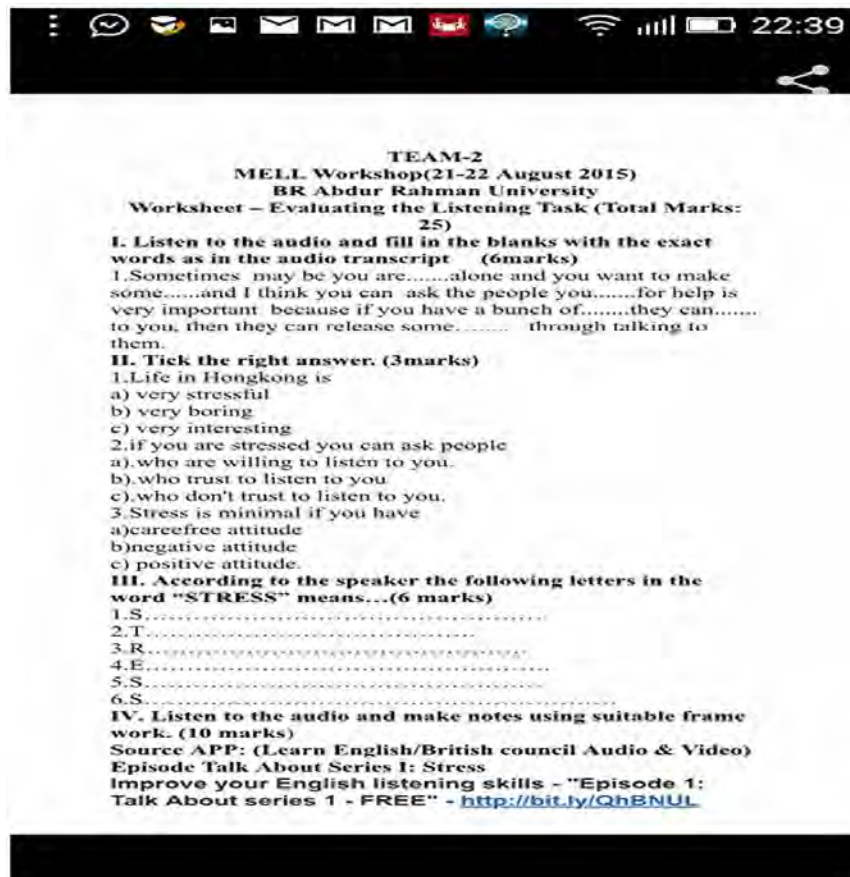


Figure 5. Screenshot of the video task uploaded by a participant



teachers to a new technology and that was evident in their enthusiasm to integrate multimedia with the teaching tasks and upload them in the simple text page. A few even downloaded YouTube videos pertaining to a particular task created by them for facilitating listening skill and uploaded them in the simple text page.

In addition to the various tasks uploaded by the participants, the feedback given by the teacher participants about the workshop showed the extent to which the workshop proved to be an effective teacher development program. The responses could be categorized under knowledge enhancement and application. With regard to knowledge enhancement, the participants felt that the workshop served as an eye opener to the world of mobile technology. They found the sessions to be comprehensive and the content gave them exposure to M-learning. The participants specified that the third session (that dealt with creating an app) was very useful as they could successfully create an app for their students and even attempt uploading tasks in the simple text page. With regard to application of the knowledge gained, they felt confident that they would be able to implement the new technique of using mobile for teaching and learning in their classrooms. Apart from these responses, the participants had generally stated that

the sessions were very interactive, resourceful, motivating, and informative and created curiosity to know more about m-learning.

SOLUTIONS AND RECOMMENDATIONS

It is worthwhile to mention that the participants showed lot of commitment to participate in every session of the workshop. After the practical sessions, the final session was devoted for interacting with the participants relating to the challenges they faced in every session. One of the challenges was the low connectivity to a few mobile phones of participants. Moreover, the battery in their mobile phones could not hold the charge for long due to access of application and had to be charged now and then. Another apprehension that the participants stated was the possibility of keeping track of students' access to the application, which would be created by the concerned teacher, periodically. It was suggested that teachers could include an evaluation component for the activities uploaded in the mobile application to monitor completion of tasks given and that would ensure students' periodic access to application. Considering the need for using technology for teaching and learning process, teachers of English can augment the use of mobile devices by utilizing the existing applications and offer blended learning experience to students along with classroom learning. It would prepare them for lifelong learning and train them in the aspect of self-direction. By preparing skill-based worksheets teachers would enhance their skill of material production and that would pave way for professional development. As Montrieux et al (2015) mention, "Introducing new technology into education generates a simultaneous need for professionalization" (p.15). Further, as Vanderlinde and van Braak (2010) suggest in their e-capacity framework, educators have to follow an educational policy that provides techni-

cal and pedagogical support in order to facilitate the development of new teaching and learning practices (Montrieux et al, 2015). In other words, the concerned institutions have to offer adequate technical support (i.e enhancing wireless connectivity within the campus) for teachers to pursue the above-mentioned endeavour.

FUTURE RESEARCH DIRECTIONS

Considering the growing interest of teachers in using mobile phones for augmenting learning, it is suggested that teacher researchers can do research on using the devices for implementing the concept of Flipped learning to teach any particular language skill or language skills in general. Similarly, teachers can work on interdisciplinary research for creating a common Learning Management System (LMS) for providing teaching components to students. In other words, teachers of English and a core subject teacher (either from Information Technology or Computer Science department) can work together to create a mobile app for the utility of teachers and students. It is worthwhile to mention that this kind of research projects would draw even alumni (who belong to one of the above-mentioned departments) of an institution, who is in the field of application development, to play an active role to design an application for his teachers. It further paves the way for an institute and industry collaboration through a project work.

CONCLUSION

Thus, it must be stated that the workshop had largely covered the method of using mobile devices for teaching and learning purposes. Needless to say, these teacher training programs create more opportunities for teachers to explore new teaching techniques.

REFERENCES

- Bjerede, M., Atkins, K., & Dede, C. (n.d.). *Ubiquitous Mobile Technologies and the Transformation of Schooling*. Retrieved from <https://www.qualcomm.com/media/documents/files/ubiquitous-mobile-technologies-and-the-transformation-of-schooling.pdf>
- Carlson, S., & Gadio, C. T. (n.d.). Teacher professional development in the use of technology. *Technologies for Education*. Retrieved from http://www.ictinedtoolkit.org/user/library/tech_for_ed_chapters/08.pdf
- Crompton, H. (2015). Know the ISTE standards for coaches: PD and program evaluation. *ISTE*. Retrieved from <https://www.iste.org/explore/articleDetail?articleid=504>
- Evans, L. (2002). What is teacher development? *Oxford Review of Education*, 28(1), 123–137. doi:10.1080/03054980120113670
- Harwell, S. H. (2003). Teacher professional development: It's not an event, it's a process. *CORD*. Retrieved from <http://www.cord.org/uploadedfiles/HarwellPaper.pdf>
- Leadership and Teacher Development Branch. (2005). *Professional Learning in Effective Schools: The Seven Principles of Highly Effective Professional Learning*. Department of Education & Training. Retrieved from <http://www.education.vic.gov.au/Documents/school/teachers/profdev/proflearningeffectivesch.pdf>
- Lynch, M. (2015). Do mobile devices in the classroom improve learning outcomes? *The Conversation*. Retrieved from <http://theconversation.com/do-mobile-devices-in-the-classroom-really-improve-learning-outcomes-38740>
- Maduakolam, I., & Bell, E. (2003). A product-based faculty professional development model for infusing technology into teacher education. *Contemporary Issues in Technology & Teacher Education*, 3(3), 340–352. Retrieved from <http://www.citejournal.org/articles/v3i3currentpractice1.pdf>

Mathew, R. (2015). Teacher development as the future of teacher education. *Innovation in English Language Teacher Education*, (pp. 29-37). New Delhi, India: British Council.

Mezirow, J. (1997). Transformational learning. *Theories of Learning in Educational Psychology*. Retrieved from <http://www.lifecircles-inc.com/Learningtheories/humanist/mezirow.html>

Montrieux, H., Vanderlinde, R., Schellens, T., & De Marez, L. (2015). Teaching and learning with mobile technology: A Qualitative explorative study about the introduction of tablet devices in secondary education. *PLoS ONE*. Retrieved from <http://journals.plos.org/plosone/article/asset?id=10.1371/journal.pone.0144008.PDF>

OECD. (2009). *Creating Effective Teaching and Learning Environments: First Result from TALIS*. Retrieved from <http://www.oecd.org/edu/school/43023606.pdf>

Quattlebaum, S. (2012). Why professional development for teachers is critical. *The Evolution- A Destiny Solutions Illumination*. Retrieved from <http://evollution.com/opinions/why-professional-development-for-teachers-is-critical/>

Sander, D. P., & McCutcheon, G. (1986). The development of practical theories of teaching. *Journal of Curriculum and Supervision*, 2(1), 50–67. Retrieved from http://ascd.asia/ASCD/pdf/journals/jcs/jcs_1986fall_sanders.pdf

Saroyan, A., Amundsen, C., & Li, C. (1997). *Incorporating Theories of Teacher Growth and Adult Education in a Faculty Development Program*. To Improve the Academy. Paper 387. Retrieved from <http://digitalcommons.unl.edu/podimproveacad/387>

Villegas-Reimers, E. (2003). Teacher professional development: an international review of the literature. *UNESCO: International Institute for Educational Planning*. Retrieved from <http://unesdoc.unesco.org/images/0013/001330/133010e.pdf>

ADDITIONAL READING

Ally, M., Grimus, M., & Ebner, M. (2014). Preparing teachers for a mobile world, to improve access to education. *Prospects*, 44(1), 1–17. Available at https://www.researchgate.net/publication/260247016_Preparing_teachers_for_a_mobile_world_to_improve_access_to_education Accessed 24 Sep 2016 doi:10.1007/s11125-014-9293-2

Ayres, D., Tyrrell, C., & Poon, K. (2013). Mobile technology: a study on the impact on the role of the initial teacher training (ITT) tutor. *Research in Teacher Education*, 3(1), 27-32. Available at: <https://www.uel.ac.uk/Schools/Cass/Research/Research-in-Teacher-Education/Volume-3-No-1-April-2013> (Accessed 24 Sep 2016)

Baran, E. (2014). A Review of Research on Mobile Learning in Teacher Education. *Journal of Educational Technology & Society*, 17(4), 17–32. Available at http://www.ifets.info/journals/17_4/2.pdf Accessed 23 Sep 2016

Grant, M. M., & Barbour, M. K. (2013). Mobile teaching and learning in the classroom and online: Case studies in K-12. In Z. Berge & L. Muilenburg (Eds), *Handbook of Mobile Learning* (PP.285-292). New York, NY: Routledge. Available at: https://www.academia.edu/3530778/Mobile_teaching_and_learning_in_the_class-room_and_online_Case_studies_in_K-12 (Accessed 13 Sep 2016).

Herrington, J., Herrington, A., Mantei, J., Olney, I., & Ferry, B. (2009). Using mobile technologies to develop new ways of teaching and learning. In Herrington, J, Herrington, A, Mantei, J, Olney, I and Ferry, B (Eds), *New Technologies, New Pedagogies: Mobile learning in Higher Education*, Faculty of Education, University of Wollongong, (pp.138). Available at: <http://ro.uow.edu.au/cgi/viewcontent.cgi?article=1077&context=edupapers> (Accessed 23 Sep 2016)

Ismail, I. F., Bokhare, S. N., Azizan, S., & Azman, N. (2013). Teaching via mobile phone: A case study on Malaysian teachers technology acceptance and readiness. *Journal of Educators Online*, 10(1), 91–129. Available at <https://doaj.org/article/965bf8bb312a4354a97ce347c551d21a> Accessed-16Sep2016 doi:10.9743/JEO.2013.1.3

Keengwe, J., & Bhargava, M. (2013). Mobile learning and integration of mobile technologies in education. *Education and Information Technologies*, 19(4), 737–746. Available at <http://link.springer.com/article/10.1007/s10639-013-9250-3> Accessed14Sep2016 doi:10.1007/s10639-013-9250-3

Liu, M., Navarrete, C. C., Maradiegue, E., & Wivagg, J. (2014). Mobile Learning and English Language Learners: A Case Study of Using iPod Touch As a Teaching and Learning Tool. *Journal of Interactive Learning Research*, 25(3). Available at https://www.researchgate.net/publication/287587248_Mobile_learning_and_English_Language_Learners_A_case_study_of_using_iPod_touch_as_a_teaching_and_learning_tool Accessed 16 Sep 2016

KEY TERMS AND DEFINITIONS

Advanced Technology: Technology that reflects the latest trends.

Flipped Learning: A new pedagogical approach that promotes collaborative learning among students. According to the approach students are expected to gather information relating to a topic, from the resources provided by the teacher and work as a team to discuss the content in the classroom.

Free Container Application: An application that provides lot of features, which could be used by the users for free to suit their requirements.

In-Service Training: Training that is offered to teachers while they are in service to enhance their teaching skills.

Learning Management System (LMS): It refers to the platform created either virtually or physically for promoting teaching and learning process. Teachers use it for sharing information to teach a concept and also for training students in a particular skill.

Mobile Enhanced Language Learning: The support provided by handheld devices for promoting language learning.

Needs Assessment: A questionnaire which is used for finding out the needs of the teachers or learners or educators in order to offer training accordingly.

Transformational Learning: A learning experience that has a greater impact on an individual and brings a change in one's behaviour and teaching technique.

T

APPENDIX: SCOPE FOR USING TECHNOLOGY

Answer the following questions and tick the appropriate options given below:

1. What do you do with your Smart phone?
 - a. Check facts
 - b. Take photos
 - c. Make videos
 - d. Carry out tests
 - e. Read the news
 - f. Dictionary
 - g. Calendar
 - h. Write down ideas
 - i. Listen to music
 - j. Images
 - k. Stopwatch/timer
 - l. Read eBooks
 - m. Voice Recorder
 - n. Discover related subject material
 - o. Document scanner
 - p. Calculator
 - q. Edit videos
 - r. Publish in the class blog
 - s. Track blog visits
 - t. Make presentations
 - u. Remote control
 - v. Communicate
 - w. Updates
 - x. Locate points on the map
 - y. Tweet
 - z. Assess Students
 - i. Share Notes
 - ii. Digital Whiteboard
 - iii. Weather
 - iv. Measure Productivity
 - v. Play
2. Why do you prefer a mobile phone to a desk top computer?
 - a. Easy to carry
 - b. Easy to type out information
 - c. Communicate without delay
 - d. Receive user-friendly commands
 - e. Have more options for sharing information
3. What similarities can you find between a desktop computer and a mobile phone?
4. Can you specify the multi-media features offered by a mobile phone?
5. What infrastructural facilities would you prefer for using mobile phones in the teaching and learning process?

Role of Educational Leaders in Supporting Beginning Teachers in Al Ain Schools in the UAE

T

Salam Omar Ali

Brighton Collage Al Ain, UAE

INTRODUCTION

Principal support of new teachers has been cited as one of the primary factors that influences both general and special education teachers' retention in the profession. However, principal leadership is a critical component of creating environments that support new teachers to meet the complex and diverse needs of their students (Correa & Wagner, 2011). Principals should work on having an effective school-based induction activities to provide all novice teachers with the induction they need.

Novice teachers may face some obstacles which might cause stress, anxiety and lack of self-assurance (Saenz-Lopez, Almagro, & Ibanez, 2011). Therefore, schools should implement an individual structured program to support the beginning.

A principal's role in the induction of novice teachers is multifaceted. Findings from several research studies, such as Andrews, Gilbert, & Martin, 2006; Correa & Wagner, 2011) suggest that building administrators who are effective in supporting novice teachers promotes a positive school climate, serves as an instructional leaders, and actively supports induction and mentoring programs.

BACKGROUND OF THE STUDY

A beginning teacher is a person who has been hired to work in a new environment with people he/she does not know. The new appointed teacher goes through a period of transition from being a

student to into becoming a teacher. In such a case, beginning teachers need supervision and support to adjust to their new roles. Moreover, being exposed to new policies, and untried materials, gives the new teachers the feeling of being lost. It is no wonder that some of the new teachers leave their jobs simply because they are left to "sink or swim" and of course many sink and leave to feel better. However, this paper highlights the importance of the principal's role in supporting novice teachers in Al-Ain in the UAE.

Research Questions

This paper answers the following questions:

1. What are the present practices of principals in Al-Ain schools in supporting their beginning teachers?
2. How can the educational leaders develop these practices to have the best induction program for their beginning teachers?
3. What needs would the new appointed teachers in Al-Ain district, have and what would they expect from an effective school-based induction program?

LITERATURE REVIEW

Around 50% of teachers who remain in education field leave their profession after five years (Evans, 2008). On the other hand, different reports indicate that 25%-50% of beginning teachers resign during their first three years of teaching

DOI: 10.4018/978-1-5225-2255-3.ch665

(Voke, 2002; NCES, 1999). Other studies showed that it is within the first five years of teaching; approximately half of the teachers will leave the occupation (Ingersoll & Smith, 2003; Murnane et. al, 1991). Novice teachers usually face many difficulties in the profession. In a study for Veenman (1984) the perceived problems of beginning teachers were classified into eight kinds of problems; they are 1) classroom discipline, 2) student motivation, 3) dealing with individual differences, 4) assessment of student work, 5) relationships with parents, 6) organization of class work, 7) inadequate instructional resources, and 8) dealing with problems of individual students. Moreover, Corbell (2008), stated that there are ten factors affect the beginning teachers' perceptions of success, such as, Mentor Support, Colleague Support, Administration Support, Classroom Management, Student Success, Instructional Resources, Assignment and Workload, Parental Contacts, Satisfaction, and Commitment. However, teachers' success can be ensured by providing them with a comprehensive, coherent professional development program (Wong, 2004).

Nothing is more important for student learning than the quality of the classroom teacher. Effective teachers manage to produce better achievement regardless of which curriculum materials, pedagogical approach, or reading program is selected (Allington, 2003). Teacher turnover usually leads to lower student achievement, and of course low student achievement leads to teacher turnover. Indeed, teaching has been characterized as an occupation with high levels of attrition or high early career turnover, especially among beginners (Darling-Hammond, 2003). Furthermore, some studies showed that most of the teachers who decided not to continue working in this career were much affected by the lack of knowledge about the school system, classroom management, dealing with the parents and many other different issues that they could not handle. As a result, it is of great importance to know that induction programs of the beginning teachers have an effective impact on their

decision whether to continue teaching or to stop it. Wong stated that, what keep a good teacher are structured, sustained, intensive professional development programs. However, these programs allow new teachers to observe others, to be observed by others, and to be part of networks or study groups where all teachers share together, grow together, and learn to respect each other's work (2003). Any induction program should be evaluated to check its effectiveness in meeting the needs of the new teachers. The increased focus on induction programs as an agent of reform in education has resulted in a great deal of research. Most studies, however, are outcome-based, generally used to ensure the financial support needed to continue induction programs in particular areas (Abell, Dillon, Hopkins, McInerney, & O'Brien, 1995). Induction is seen as the crucial phase of teacher development, which links Initial Teacher Training and in-service education for established teachers (Commonwealth Department of Education, Science and Training, 2002, p 11). Prior to any program establishment, principals should bear in mind teachers' needs before shaping or applying any induction program.

To sum up, referring to the fact that education sector plays an important role in shaping the whole society, it is very crucial to pay attention to the main element in the educational process; the teachers who are the most important factor in maintaining high quality education (Barber & Mourshed, 2007). Gless and Moil (2005) argue that nothing is more important for student learning than the quality of the teacher. Supporting teachers and providing appropriate induction program improves their teaching practices (Breux & Wong, 2003). Induction programs save the time of teachers in discovering the right instructional practices on their own. Literature suggests that effective school-based induction programs should include the following:

- Individual structured induction program (Huling-Austin, 1992). That means to find out the real needs of each teacher.

- Reduced workloads for the new qualified teachers (NQTs) and mentors for observation and assistance (Bleach, 1999 and Commonwealth Department of Education, Science and Training, 2002); opportunities to observe other experienced colleagues in order to explore good teaching and learning practice and opportunities of being observed by other trusted colleagues followed by discussion and feedback (Bleach, 1999; Huling-Austin, 1992 and Ingersoll and Kralik, 2004).
- Ongoing professional development relevant to the needs of NQTs (Commonwealth Department of Education, Science and Training, 2002; and Ingersoll and Kralik, 2004).

METHODOLOGY

A mixed method of quantitative and qualitative methods is adopted to conduct this study.

Sample and Instrumentation

A total of 20 principals and 30 of their new appointed teachers were chosen from public and private schools in Al Ain school district to administer a questionnaire and participate in interviews.

Opportunity sampling was used through a process of randomization. A convenience sample is made up by people who are easy to be part of the study, such as this one.

The study instruments include (1) a 12-item Likert scale, questionnaire with 4 response choices, strongly agree, agree, do not agree, and strongly do not agree, and (2) a semi structured interview

to gather information on principals' and teachers' perceptions and interactions between principals and newly appointed teachers regarding first year experiences at school and induction practices.

Teacher Questionnaire (See Appendix)

Demographic information was collected, including gender, years of experience and whether a mentor has been provided within the school.

The questionnaire then included 25 items; each assessed using a 4 point scale, running from Strongly Agree to Strongly Disagree.

Principal components factor analysis using oblimin rotation was used to look for common factors within the 25 items. Three factors emerged accounting for 52% of the variance ($p < 0.005$). Scales were constructed on the basis of this.

Scale 1: Mentor Support

This scale consists of items 2-6 on the questionnaire. Reliability analysis was used to check for internal reliability. A Chronbach's alpha of $r = 0.75$ was obtained. As such, the items were combined generating a "Mentor Support" scale.

Scale 2: Instructional Resources

This scale consists of items 16-21 on the questionnaire. Reliability analysis was used to check for internal reliability. A Chronbach's alpha of $r = 0.76$ was obtained. As such, the items were combined, generating an "Instructional Resources" scale.

Scale 3: Assignment and Workload

This scale consists of items 22-25 on the questionnaire. Reliability analysis was used to check for internal reliability. A Chronbach's alpha of $r = 0.66$ was obtained. Item 22 was deleted, resulting in a Chronbach's alpha of $r = 0.79$. As such, items 23-25 were combined, generating an "Assignment and Workload" scale. Items 7 to 15 did not load onto

Table 1. Breakdown of sample according to gender and group

Group	Male	Female
Teacher	15	16
Principal	11	9

any of the 3 main factors. As such, these items were not included in any further analysis.

Principal Questionnaire (See Appendix)

Demographic information was requested, including gender, and years of experience. There were then 12 items on the questionnaire, consisting of statements to be answered on a 4 point Likert scales running from strongly disagree to strongly agree.

Principal components factor analysis with oblimin rotation revealed 2 factors in the 12 questions, accounting for 55.30% of the total variance ($p < 0.001$).

Scale 1: Support

According to the factor analysis, items 2, 4, 5, 7, 8, 9, 10, 11, 12, which accounted for 40.25% of the total variance were combined. Reliability analysis for this scale revealed a Chronbach's alpha of $r = 0.86$. These items were used to construct the scale "Support".

Scale 2: Direction

Items 2, 3 and 6 were combined to construct the scale "Direction". The Chronbach's alpha for this scale was $r = 0.61$.

Questionnaire items and interview questions are designed to collect information to answer main questions of the study with use of previous literature review and surveys. They were designed to the UAE context to better inform the case study of Al Ain school district.

DATA ANALYSIS AND FINDINGS

This section presented the results of analysis of the questionnaires, for both principals and teachers, and the interviews, as follows.

Teacher Questionnaire Results

Demographics

Only 7 from the 32 respondents indicated the number of years they had been teaching. Of those 7, the number of years teaching ranged from 3 to 20, with a mean of 10.86 ($SD = 7.54$)

From the 32 respondents, 23 did not give an indication as to whether they had a mentor. Of the 9 who did complete this item, 6 did not have a mentor.

Due to the small number of respondents giving demographic information, demographics, with the exception of gender, were excluded from any further analysis.

Mentor Support, Assignment and Workload, and Instructional Resources Scales

Table 2 shows the means and standard deviations relating to these scales. No differences according to gender were obtained on any of the scales. Table 1 shows the means of the three scales.

T-tests were used to look for differences between the three scales. Significant differences were found between all 3 scales: Mentor Support, Assignment & Workload ($t = 2.28$, (31), $p < 0.05$), Mentor Support and Instructional Resources

Table 2. Mean scores and standard deviations of the three teacher scales

	Mentor Support	Assignment and Workload	Instructional Resources
Mean	13.63	14.87	6.81
SD	2.94	2.96	1.73

($t=12.09$, (31), $p<0.001$), and Assignment & Workload and Instructional Resources ($t=15.04$, (31), $p<0.001$).

The t-tests indicate that respondents felt least positive about Instructional Resources, more positive about Mentor Support and mostly positive about Assignment & Workload. These results seemed similar to what the literature review indicates.

Principal Questionnaire Results

Of the 20 participants, only 9 gave information as to the number of years teaching. Of those, the number of years taught ranged from 11 to 30, with a mean of 19.78 ($SD=7.82$). This indicated that most of the participating principals had many years of teaching before assuming the management positions in their schools.

Due to the small number of respondents giving demographic information, demographics, with the exception of gender, were excluded from any further analysis.

Support and Direction Scales

Table 3 shows the means and standard deviations relating to the Support and Direction scales. No differences according to gender were obtained on any of the scales. Table 2 shows the means of the two scales. A higher mean for support (28.15 vs. 9.25) than direction is shown to indicate that teachers in this study viewed the support needed highly than given directions. Newly appointed teachers have the tendency to feel that they are ready and well prepared graduating from their programs to teach, hence, need of less direction, but greater support.

Table 3. Mean scores and standard deviations of the two principal scales

	Direction	Support
Mean	9.25	28.15
SD	1.74	4.20

T-tests revealed a significant difference between the two scales ($t=21.49$, (19), $p<0.001$). This indicates that principals felt more positive about the level of support they give rather than the amount of direction they give.

Inter-Relationships Between Principal and Teacher Questionnaire

Pearson's correlational analysis found only one relationship between the five scales. The findings showed a positive correlation between 'Mentor' and 'Instructional Resources' ($r=0.45$, $p<0.01$). This indicated that the more positive the teacher felt about their mentoring, the more positive they felt about instructional resources. No relationships were found between how well principals felt they were directing and supporting and the teacher's opinions.

The Interview Results

It is obvious from the findings of the interviews questions (see pages 22&23) that the principals showed that they provide a lot of care to the novice teachers in terms of supporting them, giving them a reasonable workload and providing them with adequate time to plan and meet the other teachers in the same field. On the other hand, teachers kept on complaining about their current situation as their real needs have been never met. According to the interviews, most of the principals said that they are a great help to their novice teachers, but they are restricted by the time; they do not have enough time to guide and assist the new appointed teachers. Some of the principals stated that: there is a big need to appoint a mentor for all new novice teachers to enable them to produce high quality of teaching and to give them the self-confidence they need to continue their career. Others said that they would like to help and support their novice teachers personally, but the principals' duties are too complicated as they should make sure that everything at the school is just on the right track, simply they find their jobs very exhausting and

stressful. Regarding the relationship between the principals and the novice teachers, most of them commented that they are very keen to have a very good relationship with their novice teachers, through which novice teachers can express their ideas and talk about their needs freely without being afraid of the possibility of being fired. This idea is supported by the literature (Brock and Grady, 1998)

FUTURE RESEARCH DIRECTIONS

Although this study explored the role of the principals in supporting novice teachers, at the same time, it gave us ideas for future research, such as to investigate the impact of school climate on novice teachers' retention and the effect of having mentors in supporting novice teachers' motivation.

SOLUTIONS AND RECOMMENDATIONS

There is a big gap between what the novice teachers expect to get from their principals and what they are getting. Therefore, principals have to integrate an induction program into the school system and to become as part of the school's culture.

Depending on the results, in order for the beginning teachers to accept their profession, they need to be provided them with support and guidance. An induction should be taken into the stake holders' consideration to help teachers develop their abilities and skills through an ongoing professional development. The new appointed teachers who will be under that program should have less workload than any other teacher to make sure that they will not be overwhelmed with work. Furthermore, Schools should make it possible for the novice teachers to have enough time to meet the other professional experienced teachers who could provide them with help and assistance regarding

planning and other educational issues. Principals have to be very keen to let their novice teachers meet the experts coming from outside the school. Special courses should be attended according to the teachers' experiences. Furthermore, the following should be taken into consideration:

Mentoring

It is basic need for novice teachers. Principals should make sure that new appointed teachers get mentors as mentoring helps them feel that there is someone cares about them and helps them deal with day-to-day worries.

Peer Coaching

This is one of the most important needs of novice teachers. Many educators across the globe use various coaching models to improve the performance of teachers and raise the attainment of their students. Coaching is about helping people to learn.

Principals and the school management team should be the models of sharing good practice through the distribution of leadership, giving the head of faculties (department) the anatomy to shape their departments. Additionally, they should have their input into the school strategic planning and professional development of all staff.

CONCLUSION

Novice teachers are obviously in need of support to overcome the obstacles they may face. Consequently, it is the role of the principals to understand the stress that novice teachers have and to help them reduce their sufferings. Moreover, the quality of the relationship between the novice teachers and their principals is so important.

Principals have to offer support for their novice teachers in the proper atmosphere through different ways, such as providing them with opportunities

of orientation sessions to understand the school policy. Experienced teachers can also provide basic teaching tips to inexperienced teachers; ideas that can be immediately implemented into the classroom with new ideas in teaching methodologies.

Induction programs that are designed for novice teachers should meet their diverse needs; these needs can be classified into the following:

- **Psychological Needs:** To prepare the novice teacher to endure any psychological issue and to manage stress is an important need.
- **Professional Needs:** Informing teachers about the importance of professional development for their career, their duties, different ways of dealing with students and with the other teachers, legal issues and any other important professional needs that should be met before and during the service.
- **Instructional Needs:** To prepare the new appointed teachers for their classrooms instruction is very important. Principals should make sure that novice teachers are ready to use different instructional methods, techniques and resources.
- **Social Needs:** Providing teachers with opportunities to start having good relations with parents and colleagues using different ways such as, open days for teachers and staff, social activities for teachers, and above all, participating in extracurricular activities.

Finally one should realize the importance of supporting novice teachers, providing them with the right training they need and connecting them with knowledgeable, caring and experienced teachers to help them build up their confidence, efficacy and success in serving diverse students with diverse needs.

REFERENCES

Abell, S. K., Dillon, D. R., Hopkins, C. J., McInerney, W. D., & OBrien, D. G. (1995). Somebody to count on: Mentor/intern relationships in a beginning teacher internship program. *Teaching and Teacher Education*, 11(2), 173–188. doi:10.1016/0742-051X(94)00025-2

Allington, R. (2003). *The six ts of effective elementary literacy instruction*. Retrieved Oct 11, 2005, from www.readingrockets.org/article.php

Andrews, S., Gilbert, L., & Martin, E. (2006). The first years of teaching: Disparities in perceptions of support. *Action in Teacher Education*, 28(4), 4–13. doi:10.1080/01626620.2007.10463424

Barber, M., & Mourshed, M. (2007). *How the world's best-performing schools come out on top*. London McKinsey.

Bleach, K. (1999). New deal for new teachers. *Professional Development Today*, 2(2), 7–18.

Breaux, A., & Wong, H. (2003). *New teacher induction: How to train, support, and retain new teachers*. Mountain View, CA: Harry K. Wong.

Brock, B. L., & Grady, M. L. (1998). Beginning teacher r induction programs: *The role of the principal. The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 7(1), 179–183. doi:10.1080/00098659809599355

Commonwealth Department of Education, Science and Training, (2002). *An Ethic of Care. Effective Programmes for Beginning Teachers*, Retrieved February, 5, 2009 from [http://www.dest.gov.au/archive/schools/publications/2003/Beginning Teachers.pdf](http://www.dest.gov.au/archive/schools/publications/2003/Beginning%20Teachers.pdf)

Corbell, K. A. (2008). *Evaluating the perceptions of success inventory for beginning teachers and its connection to teacher retention* (Doctoral dissertation). North Carolina State University.

Correa, V., & Wagner, J. (2011). Principals' roles in supporting the induction of special education teachers. *Journal of Special Education Leadership*, 24, 17–25.

Evans, A. (2008). *Reasons why new teachers quit*. Retrieved May, 7, 2009 from www.associatedcontent.com/article/670891/reasons

Gless, J., & Moir, E. (2005). *Supporting Beginning Teachers with Heart and Mind: a Decade of Lessons learned from the Santa Cruz New Teacher Project*. Santa Cruz, CA: The New Teacher Center at the University of California.

Huling-Austin, L. (1992). Research on learning to teach: Implications for teacher induction and mentoring programmes. *Journal of Teacher Education*, 5(43), 173–181. doi:10.1177/0022487192043003003

Ingersoll, R. M. (2001a). Teacher turnover and teacher shortages: An organizational analysis. *American Educational Research Journal*, 38(3), 499–534. doi:10.3102/00028312038003499

Ingersoll, R. M., & Kralik, J. M. (2004). *The impact of mentoring on teacher retention: What the research says*. Denver, CO: Education Commission of the States. Retrieved April, 9, 2009, from: <http://www.ecs.org/clearinghouse/50/36/5036.htm>

Inman, D., & Marlow, L. (2004). Teacher Retention: Why Do Beginning Teachers Remain in the Profession? *Education*, 124, 2004.

Murnane, R., Singer, J. D., & Willet, J. B. (1991). *Who will teach? Policies that matter*. Cambridge, MA: Harvard University Press.

Nakpodia, E. D. (2010). Relationship Between Principals And Teachers' Perceptions of Principals' Performance Of Staff Development. *The Online Journal of Academic Leadership*, 8(4), 10–24.

Saenz-Lopez, P., Almagro, B. J., & Ibanez, S. J. (2011). Describing problems experienced by Spanish novice physical education. *The Open Sports Sciences Journal*, 4(1), 1–9. doi:10.2174/1875399X01104010001

Veenman, S. (1984). Perceived Problems of Beginning Teachers. *Review of Educational Research*, 54(2), 143–178. doi:10.3102/00346543054002143

Voke, H. (2002). Understanding the teacher shortage. *ASCD InfroBrief*, 29, 1–17.

Vroom, V. H. (1964). *Work and motivation*. New York: Wiley.

Wong, H. (2003). *Induction: How to train, support, and retain new teachers*. Paper presented at the National Staff Development Council. Retrieved December, 10, 2003, from www.NewTeacher.com

ADDITIONAL READING

Bennetts, C. (2001). Lifelong learners: In their own words. *International Journal of Lifelong Education*, 20(4), 272–288. doi:10.1080/02601370110048782

Bickmore, S. T., & Bickmore, S. T. (2010). A Multifaceted Approach to Teacher Induction. *Teaching and Teacher Education: An International Journal of Research and Studies*, 26(4), 1006–1014. doi:10.1016/j.tate.2009.10.043

Caprara, G. V., Barbaranelli, C., Borgogni, L., & Steca, P. (2003). Efficacy beliefs as determinants of teachers job satisfaction. *Journal of Educational Psychology*, 95(4), 821–832. doi:10.1037/0022-0663.95.4.821

Cross, C., & Rigden, D. (2002). Improving teacher quality. *The American School Board Journal*, 189(4), 24–27.

Hawk, P. (1986–1987). Beginning teacher programs: Benefits for the experienced educator. *Action in Teacher Education*, 8(4), 59–63. doi:10.1080/01626620.1986.10519321

Hiebert, J., Gallimore, R., & Stigler, J. (2002). A knowledge base for the teaching profession: What would it look like and how can we get one? *Educational Researcher*, 31(5), 3–15. doi:10.3102/0013189X031005003

Johnson, S. M., & Birkeland, S. (2003b, Fall). Pursuing a sense of success: New teachers explain their career decisions. *American Educational Research Journal*, 4(40), 608.

Kauffman, D., Johnson, S. M., Kardos, S., Liu, E., & Peske, H. G. (2002). Lost at sea: New teachers experiences with curriculum and assessment. *Teachers College Record*, 104(2), 273–300. doi:10.1111/1467-9620.00163

KEY TERMS AND DEFINITIONS

T

Mentor: A facilitator who usually helps new appointed teachers find their ways in the profession.

Novice Teachers: Teachers who are new to the profession of teaching.

Peer Coaching: When two or more colleagues (teachers) work together to share good practices and help each other build new teaching skills.

APPENDIX

Teacher's Questionnaire

The Role of Principals in Supporting the New Teachers (Teacher Version)

Full Name (optional): -----

Your school's name (optional): -----

Years of experience as principal: -----

Years of experience as teacher (if any): -----

Gender: _____

Do you have a mentor assigned to you by the school or another experienced teacher to provide you with assistance?

_____ Yes

_____ No

Use the following scale to answer each of the questions.

1: strongly disagree

2: disagree

3: agree

4: strongly agree.

Table 4. Mentor support (teacher questionnaire)

#		1	2	3	4
1	The mentoring relationship is or would be important to me.				
2	My mentor or an exemplary teacher has provided assistance with classroom management.				
3	My mentor or an exemplary teacher has provided assistance with instructional concerns.				
4	My mentor or an exemplary teacher encourages me to reflect about my teaching.				
5	Working with my mentor has been a positive experience.				
Colleague Support					
6	I have opportunities for meaningful conversations with other novice teachers in a setting free of evaluation.				
7	I have common planning times with other teachers at my same grade level or subject area.				
8	I have opportunities to visit and observe exemplary teachers.				
9	I have a colleague in my same subject area (secondary) or grade level (elementary) who will answer my questions.				

continued on following page

Table 4. Continued

T

#		1	2	3	4
Administration Support					
10	The administration at my school provides appropriate feedback for my discipline's decisions				
11	The administration at my school encourages me to be an effective teacher.				
12	The administration has oriented me to the school and staff.				
13	I am satisfied with the contact I have with my administration.				
14	The administration provides me with effective instructional leadership.				
Instructional Resources					
15	I have been provided with curriculum that aligns with the Ministry's objectives for my grade level or subject area.				
16	I have the curriculum materials I need to teach effectively.				
17	I feel confident in my ability to use the instructional technology available to me.				
18	The school provides professional development that is valuable to my instruction in the classroom.				
19	I feel confident in my ability to grade student work.				
20	I feel comfortable with reporting the assessment of my students' work.				
Assignment and Workload					
21	I think the number of preparations I have for my classes is appropriate for a beginning teacher.				
22	I have at least one period per day that I can devote without interruption to planning for my classes.				
23	My overall teaching workload is reasonable.				
24	Beginning teachers are allowed to choose whether to take on extra duties or not.				

The Principal Questionnaire

The Role of Principals in Supporting New Teachers

Full Name (optional): -----

Your school's name (optional): -----

Years of experience: -----

Gender: _____

Use the following scale to respond to each questionnaire item by checking the appropriate response box as follows

1: strongly disagree

2: disagree

3: agree

4: strongly agree.

Table 5. Mentor support (principal questionnaire)

#		1	2	3	4
1	I direct new teachers to perform at acceptable level at school				
2	I work with new teachers so they can blend in school culture				
3	I make sure to that new teaches are oriented and made aware of Ministry policies				
4	I am fully satisfied and pleased from the help and support given to the newly appointed teachers				
5	I think about my professional conduct in light of moral and ethical standards.				
6	I allow new teachers to meaningfully converse with other novice teachers in a setting free of evaluation.				
7	We provide new teachers with induction on all aspects of school				
8	I assist new teachers to transform skills to meet teaching and learning needs				
9	I evaluate new teachers more often to give feedback to them				
10	I attend to new teachers' school problems and issues				
11	I believe that new teachers can benefit from assign them mentors				
12	I assign mentors or veteran teacher to guide new teachers				

Do you have any additional comments?

Thank you for your cooperation.

Principal's Interview Questions

- To which extent are you satisfied with the help and support you provide the novice teachers with?
- What kind of obstacles you face in helping and supporting novice teachers?
- Do you think novice teachers need a mentoring program? Justify your answer.
- If you feel that your school is fair with novice teachers regarding the workload and the support they get explain how?

Teacher's Interview Questions

- To which extent are you satisfied with the help and support you get from the principal?
- What kind of obstacles you face in your career as a novice teacher?
- Regarding training and support, do you think novice teachers are getting what they need?
- If you feel that your school is fair with novice teachers regarding the workload and the support they get explain how?

The Technological Pedagogical Content Knowledge of EFL Teachers (EFL TPACK)

T

Mehrak Rahimi

Shahid Rajaei Teacher Training University, Iran

Shakiba Pourshahbaz

Shahid Rajaei Teacher Training University, Iran

INTRODUCTION

ICT-integrated teaching is not as easy as it might seem and it requires compound skills and knowledge base for both teachers and learners. The theoretical framework of the knowledge teachers need to teach with technology has been referred to in the literature by the term TPACK or Technological and Pedagogical Content Knowledge. TPACK is a complex framework that explains teacher knowledge for technology integration and its intertwining concepts.

This heavy demand of teaching professional career is rooted in a string of developments and advancements in technological arena that started in mid-20th century and has been flourishing in this century swiftly. It is not far-fetched to say that every aspect of peoples' lives is changing along with advances of technology in the 21st century. Computers, cell phones, televisions, and other technological devices are no longer considered new inventions, as they are now being used daily by everyone. Education like other fields of science must not fall behind this trend. The curriculum, methodologies of teaching, resources and materials, teachers, students and the school environment need to adapt and change to match the requirements of the modern world in which technology plays an undeniably significant role.

For some people, staying current with technology in the field of education is a more important issue as they look at the progress of ICT (Information and Communications Technology) to be a

solution to many pedagogical problems (Pedersen, 2001). This implies a prevalent force in the modern world to embrace technology more than ever. In the field of education, this is translated into smart schools that are equipped with innovative technological facilities such as interactive whiteboards, developed computer laboratories and a variety of software programs.

Many researchers have investigated the effects of ICT on enhancing teaching and learning in different ways. It is believed that ICT can have positive effects on education such as learning efficiency, learning effectiveness, access, convenience, motivation, and institutional efficiency (Hubbard, 2009). With that being said, ICT integration in education is expected to lead to better learning which automatically demands more qualified teaching conditions.

In this era, educational administrators push educators to empower themselves professionally in line with the trend of technological normalization. Hence, many teachers face the dilemma of 'changing' themselves with the new teaching condition. With this pushed change in the educational system, of course come some unintended consequences for the teachers, some of which are actually not so desirable (Pedersen, 2001). Some of these consequences are already evident: ICT integration can take a huge load off teachers if the required resources and skill are present; however, it is also likely that it reduces the teaching quality in environments that lack the needed resources and/or teachers do not possess enough technology

knowledge. As a matter of fact, research suggests that one of the most important personal factors that hinder technology normalization is the lack of ICT knowledge (Mahdi, 2013). Studies in this regard reveal that in order to benefit from all aspects of technology in education, ICT should be integrated in the educational system in a way that it is used by teachers and students every day, as an integral part of the lesson, just like pen and pencil (Bax, 2002). Naturally, this requires much knowledge of technological affordances from both sides.

According to what was presented above, it is obvious that the lack of knowledge to use technology creates a huge burden for teachers to integrate technology in the process of teaching (Yurdakul, et al., 2012). Teaching is known to be one of the most stressful occupations ever since teachers are believed to suffer from different social and political discriminations leading many of them to feel frustrated (Warrad, 2013). Now in the 21st century, the force of technological empowerment and ICT integration in schools has multiplied the stressful nature of the profession for teachers. TPACK or ICT-literacy can be the answer to teachers' prayers regarding successful ICT integration in schools. To make the whole process of education more successful and to protect the well-being of teachers, empowering teachers to handle their job appropriately is a very crucial issue. TPACK is the knowledge of most value in today's world and understanding this complex knowledge is the very first step in the path of successful ICT integration into the process of teaching. With that being said, this chapter aims at investigating the concept of TPACK in terms of education in general and language teaching in particular.

BACKGROUND

Considering the complexity of teaching profession in general, it is not a surprising thing to claim that the knowledge required by teachers to integrate ICT into their teaching is also complex in nature. The idea of creating a construct for explaining this

knowledge goes back to 1986 in the Shulman's description of pedagogical knowledge (PK). In his seminal work, Shulman suggested three categories of content knowledge required to handle the teaching of a specific subject matter: subject matter content knowledge, pedagogical content knowledge, and curricular content knowledge. By the advancement of technology, however, the attention to teachers' knowledge base was naturally drawn to technological arena and the way it is interwoven with other types of content knowledge. Research on TPACK can generally be divided into three main categories (Yurdakul et al. 2012):

1. Definition and measurement of TPACK,
2. Effects of professional development on TPACK, and
3. Evolution of the TPACK.

In the first phase of TPACK research, the studies tried to systematically define TPACK (e.g., Koehler, Mishra, Yahya, and Yadav, 2004) and create a deep understanding of the complex interrelationships between content, pedagogy and technology knowledge and the contexts in which they occur (Yurdakul et al. 2012).

The second phase of TPACK mostly focused on investigating the role of TPACK in the professional development of pre-service and in-service teachers (Koehler & Mishra, 2005). Most studies in this phase were experimental and tried to explore whether teachers' professional development programs can actually influence their TPACK level. Not surprisingly, it was found by most of these studies that teachers' professional development has a positive influence on developing teachers' level of TPACK (e.g., Doering, Veletsianos, Scharber, & Miller, 2009; Graham et al., 2009).

Following this phase, research on TPACK construct shifted to focus on developing an integrated model of TPACK considering its various dimensions separately and in relation to one another. However, the number of studies with a focus on TPACK as a unified and consistent body

is very limited (Archambault & Crippen, 2009). It might be safe to say that the study that has been carried out in 2012 by Yurdakul and associates has been successful in developing a consistent and comprehensive framework. This study includes technology integration in classrooms based on different competencies and components in teaching-learning process. As a result, this framework has been chosen to define TPACK and extend its components into teaching English as foreign language (EFL) teacher knowledge base in the current study.

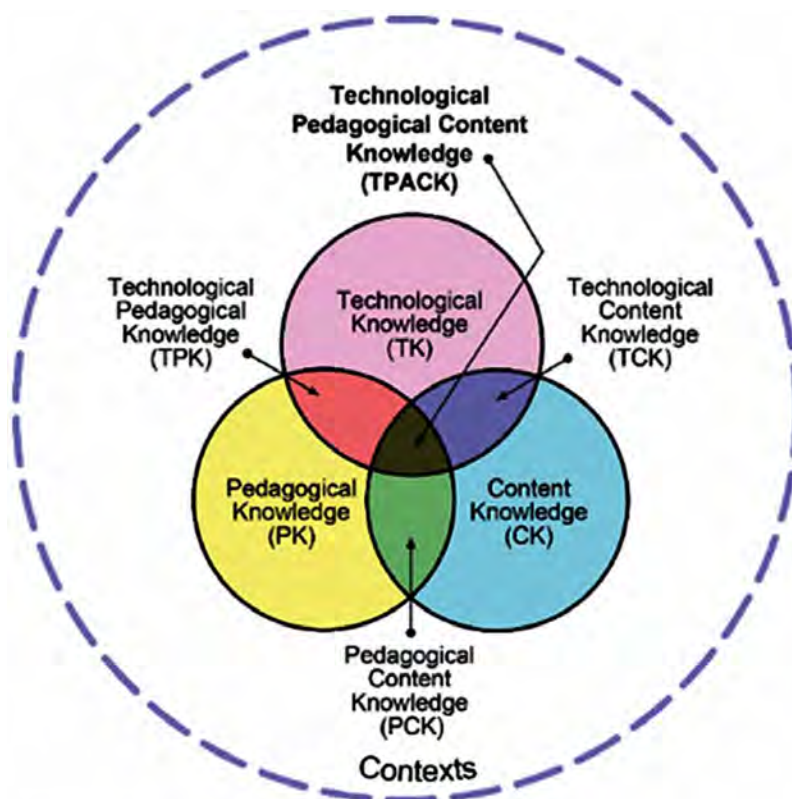
The Definition of TPACK

The fact that teaching is a highly complicated and demanding profession is accepted by most educators. Teachers are constantly expected to apply many domains of complex knowledge structures across different contexts (Mishra, Spiro

& Feltovich, 1996; Spiro & Jehng, 1990). With that being said, integrating technology into the teaching process is not a simple task. The skills needed to use ICT effectively should consider various aspects including subject matter, knowledge of student thinking and learning, and most importantly, knowledge of technology (Putnam & Borko, 2004).

TPACK is defined in literature as a complex framework to explain teacher knowledge for technology integration. Teacher knowledge in this definition is not simply the ability to use technology; it is a complex interaction of three layers of knowledge within the TPACK framework: technology, pedagogy and content (Yurdakul et al, 2012). In a more concise definition, TPACK is defined as “an understanding that emerges from interactions among content, pedagogy, and technology knowledge” (Koehler & Mishra, 2009, p.66). In order to understand and define TPACK

Figure 1. the TPACK framework and its three components
Source: Koehler & Mishra, 2008, p. 12.



in a more comprehensive way, however, it is required to understand the underlying themes and core concepts of TPACK framework.

Figure 1 is one of the early conceptualizations of TPACK construct displaying its core intertwining concepts: Technological Knowledge (TK), Pedagogical Knowledge (PK) and Content Knowledge (CK). Technological Knowledge (TK) refers to one's ability to use and integrate technology into everyday activities and also the ability to recognize the situations which technology can help to enhance or impede reaching a goal. Pedagogical Knowledge (PK) is defined as teachers' deep knowledge about different teaching and learning methodologies, approaches and practices such as planning lessons and classroom management techniques. Content Knowledge (CK) is teachers' knowledge of the specific subject matter that is going to be taught or learned including theories, approaches, ideas, concepts and frameworks (Koehler & Mishra, 2009).

According to this definition, there are other components in TPACK framework which are created as a result of the interaction among the three

domains of teacher knowledge. These components include: TCK (Technological Content Knowledge), PCK (Pedagogical Content Knowledge), TPK (Technological Pedagogical Knowledge), and TPACK (Technological Pedagogical Content Knowledge) (Koehler & Mishra, 2009; Yurdakul et al, 2012). Table 1 contains a short summary of the definitions and examples for each of these seven dimensions synthesized from literature review (Cox & Graham, 2009; Koehler & Mishra, 2009; Mishra & Koehler, 2006 as cited in Chai, Koh & Tsai, 2013, p.33).

Despite the fact that TPACK is the basis of technological knowledge framework, early conceptualizations of TPACK framework focused on seven distinct components of technology knowledge (TK, PK, CK, TPK, TCK, PCK and TPACK) (Shahin, 2011; Shmidt et al, 2009). The lack of attention to the main component TPACK as a whole led to the development of a more comprehensive framework of TPACK considering all these separate parts along with a main core.

Table 1. Definitions and examples of TPACK dimensions

TPACK Constructs	Definition	Example
TK	Knowledge about how to use ICT hardware and software and associated peripherals	Knowledge about how to use Web 2.0 tools (e.g., Wiki, Blogs, Facebook)
PK	Knowledge about the students' learning, instructional methods, different educational theories, and learning assessment to teach a subject matter without references towards content	Knowledge about how to use problem- based learning (PBL) in teaching
CK	Knowledge of the subject matter without consideration about teaching the subject matter	Knowledge about Science or Mathematics subjects
PCK	Knowledge of representing content knowledge and adopting pedagogical strategies to make the specific content/topic more understandable for the learners	Knowledge of using analogies to teach electricity (Shulman, 1986)
TPK	Knowledge of the existence and specifications of various technologies to enable teaching approaches without reference towards subject matter	The notion of Web quest, KBC, using ICT as cognitive tools, computer-supported collaborative learning
TCK	Knowledge about how to use technology to represent/ research and create the content in different ways without consideration about teaching	Knowledge about online dictionary, SPSS, subject specific ICT tools e.g. Geometer's Sketchpad, topic specific simulation
TPACK	Knowledge of using various technologies to teach and/ represent and /facilitate knowledge creation of specific subject content	Knowledge about how to use Wiki as an communication tool to enhance collaborative learning in social science

Source: Chai, Koh & Tsai, 2013, p.33.

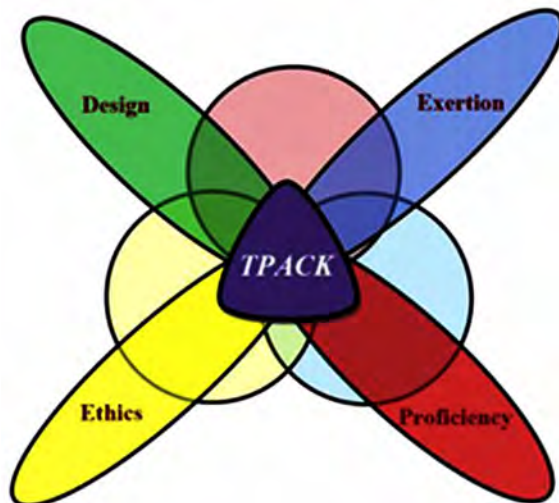
The Conceptual Model of TPACK

TPACK is not a simple framework and this is due to the complicated nature of teaching profession. Knowing about technology and how to use it in everyday activities is just one part of the story; knowing how to integrate this knowledge into instruction by using proper pedagogical skills and classroom management abilities is a quite different paradigm. This matter has been seen within a new framework of TPACK where the original concepts are preserved and the concepts and components are interwoven more comprehensively. The framework (Figure 2) consists of four main components: design, exertion, ethics and proficiency (Yurdakul et al, 2012).

- **Design:** It refers to teacher's ability to create and develop curriculum plans, and combining proper technological tools and resources to enhance the teaching and learning situation.
- **Exertion:** It refers to the ability to implement the design plans in the teaching and learning situation and also using appropriate technologies to assess and evaluate students learning.

Figure 2. The framework and factors of TPACK-deep scale

Source: Yurdakul et al, 2012, p. 973.



- **Ethics:** It focuses on the ability to consider copy right issues and show ethical behavior as a professional in online educational environments and also when using technological tools and resources.
- **Proficiency:** It refers to teachers' leadership ability to integrate technology into the teaching process in the most effective way.

THE EFL TPACK

Teaching English as a foreign language takes place when language learners learn English in a nonnative English speaking country. In this situation, English is usually taught by nonnative teachers who are not expected to be as fluent and accurate as native English speakers as they themselves have been EFL learners. Due to the lack of exposure to real and authentic language input in the EFL environments, EFL teaching strongly depends on technology especially in teaching listening and speaking (Liu et al., 2014). Technological tools have made EFL teaching effectively possible as early as 1960s and thus knowing how to use technology has always been an undeniably important aspect of EFL teachers' professional development (Yang, 2000).

In order to help EFL teachers become successful in their profession, there is a need to define the conceptual model of EFL TPACK. Earlier, the knowledge/competency base of EFL teacher education programs has been proposed to be based on a tripartite model including (Rahimi, 2008, p. 4):

- **Content Knowledge (CK):** Knowledge of the subject matter, English language.
- **Pedagogical Knowledge (PK):** Knowledge of science of teaching and pedagogy, knowledge of generic teaching strategies, beliefs, and practices; along with support knowledge, the knowledge of the various disciplines that would enrich teachers' approach to the teaching and learning of English.

- Pedagogical Content Knowledge (PCK):** Knowledge/competency of teaching in reality, the specialized knowledge of how to represent content knowledge in the classroom and how students come to understand the subject matter in the context of real teaching; the students' problems and ways to overcome those problems by considering all variables related to their learning (teaching materials, assessment procedures, parents, etc.).

Combining this model with the TPACK model presented earlier, the EFL TPACK can be specified more meticulously for EFL profession. The components of EFL TPACK are presented in Table 2.

By going through the EFL TPACK model presented above, it is clear that when ICT use is accompanied by appropriate knowledge base on teachers' side, it can help EFL teachers to enhance all aspects of their teaching; from materials development to classroom management to appropriately diagnose language learners' learning problems and

Table 2. The EFL TPACK components

EFL TPACK Constructs	Definition	Example
CK	Knowledge of the subject matter without consideration about teaching the subject matter	English language proficiency
PK	Knowledge about the students' learning, instructional methods, different educational theories, and learning assessment to teach a subject matter without references towards content	knowledge of generic teaching strategies, beliefs, and practices along with support knowledge, the knowledge of the various disciplines that would enrich teachers' approach to the teaching and learning of English (e.g. educational psychology, second language acquisition), such as knowledge of using metacognitive strategies to enhance learning
PCK	Knowledge of representing content knowledge and adopting pedagogical strategies to make the specific content/topic more understandable for the learners	the specialized knowledge of language teaching and learning; how to represent English as a foreign language in the classroom and how language learners come to understand English in the context of real teaching; discovering the students' problems and ways to overcome those problems by considering all variables related to their language learning (teaching materials, assessment procedures, parents, etc.), such as knowledge of conducting group activities to improve students' learning
TK	Knowledge about how to use ICT hardware and software and associated peripherals	IT literacy, knowledge of technology in general, knowing about basic computer software (web browsers, media players) and devices (printers, scanners, digital cameras)
TPK	Knowledge of the existence and specifications of various technologies to enable teaching approaches without reference towards subject matter	IT integration literacy, the ability to use technologies to teach and interact with students
TCK	Knowledge about how to use technology to represent/research and create the content in different ways without consideration about teaching	Knowledge of CALL teaching/learning: CALL materials development and lesson planning
TPACK	Knowledge of using various technologies to teach, represent, and facilitate knowledge creation of specific subject content	Knowledge of CALL teaching/learning: using multimedia software/games as a tool to enrich teaching language macro skills (listening, speaking, reading, writing) and components (grammar, vocabulary, pronunciation); class management and assessing students' learning; presenting content via appropriate language teaching strategies by using proper technological tools intermingled with appropriate language teaching methodology/instructional materials

solving those problems professionally. In this way, ICT integration would be a truly wise choice for every EFL classroom paving the way of creating a more authentic and native-like language teaching and learning environment for both EFL teachers and learners.

Assessing EFL-TPACK

Research suggests that integrating ICT into education has become significantly important for European and Western countries in the last decade (Elmaifi, 2014). In the same vein, the expansion of technological infrastructures has made ICT one of the most important issues addressed in educational systems of developing countries. One reason is to prepare the labor force of a world in which technology is being normalized. As a result, the mainstream education needs technologically literate teachers to handle classes of digital natives and direct them in the right path of using technology for their academic and professional goals. Having said this fact, teacher professional development is now considered a highly important resource in the field of language teaching to help EFL teachers develop their knowledge base to catch up the needs of today's modern world (Avalos, 2011; Kleinsasser, 2013). Professional development programs in the area of CALL (Computer Assisted Language Learning) are one example (Hong, 2010).

All around the world, short term pre-service and in-service courses have been prevalent ways enabling educators to develop their professional capacities. Needless to say, these programs cannot be very fruitful without a comprehensive needs analysis. A wise step to do so would be investigating the needs of EFL teachers regarding ICT literacy. It would be useful to gain comprehensive insights about EFL teachers' current level of TPACK, and find out accurate facts about their strong and weak points considering different dimensions of TPACK knowledge base. In this way, the program would be certainly more effective and lots of time, money and resources would be saved.

A number of instruments have been developed to assess TPACK. The differences of these instruments lie in different TPACK frameworks that have been developed to this date. The TPACK survey developed and validated by Schmidt et al. (2009) is a widely used instrument to assess TPACK. This scale assesses TPACK in terms of the seven distinct components of TPACK: TK, CK, PK, TCK, PCK, TPK and TPACK. The 24 item TPACK survey developed by Archambault and Crippen (2009) is another example. The survey, like the previous one, examines TPACK by the seven components of TPACK based on the prototypical model. What these two instruments have in common is that they have been built on the commonly used TPACK framework (Koehler & Mishra, 2008; Mishra & Koehler, 2006). The TPACK scale developed and validated by Yurdakul et al., (2012) however is a little different from these two. The scale has been built on the TPACK framework developed by Yurdakul et al., (2012) and consists of four sub-scales including Design (10 items), Exertion (12 items), Ethics (6 items), and Proficiency (5 items).

Two instruments have been specifically developed to measure language teachers' TPACK. The first one is an EFL-TPACK questionnaire that investigates EFL students' perceptions of their teachers' TPACK (Tseng, 2014). In the process of developing the scale, the TPACK prototypical model with seven components was put into test (TK, CK, PK, PCK, TCK, TPK, and TPACK). The result showed that EFL teachers' TPACK consists of five factors, three of which were found to be TK, TPK, and PCK and two of which were the combination of PK and CK; and TCK and TPACK respectively.

In a very recent work, Baserab, Kopchac, and Ozdend (2016) developed and validated a self-assessment survey that examines TPACK of EFL teachers. The TPACK-EFL survey, as it is called by the authors, aims to provide an assessment tool for pre-service foreign language teachers and addresses subject-specific pedagogies and technologies. The result of data analysis revealed

a seven-factor structure for TPACK including TK, CK, PK, PCK, TCK, TPK, and TPACK.

What is vitally important to mention here is that both scales showed that EFL TPACK has the same factor structure as the general scale, verifying the seven-component model of TPACK (Shahin, 2011; Shmidt et al, 2009). However, what distinguishes these scales from the original version is the way items are loaded on the extracted factors. This means that the general model as a whole is valid to be used across subject matters; however, context specific studies are required to explain the details of the construct TPACK.

SOLUTIONS AND RECOMMENDATIONS

In today's digital world, EFL teachers face an immediate need to 'technologize' their pedagogical knowledge, meaning that they need to integrate ICT literacy with their professional knowledge (Liu et al., 2014). The arrival of the Internet and wireless technologies can be considered a huge driving force that guides EFL teachers in the path of ICT integration. With learners being exposed to unbounded Internet resources wherever they are at any time, EFL teachers have no choice but to embrace technology as an integral part of their teaching. Failure to do so will certainly create a situation in which teachers are not able to function effectively in their job (Mishra, Koehler & Kereluik, 2009).

Considering the modern world's need which is global communication, one of the most effective EFL teaching methods is considered to be communicative language teaching (Bygate, 2001). ICT integration can certainly provide opportunities to bring various authentic materials and activities that can provide a suitable near-native language learning atmosphere. Using technology can also be an appropriate solution to big class size dilemma that most EFL teachers deal with (Liu

et al, 2014). With the presence of high TPACK, EFL teachers can assess their students' learning more accurately by using technology to make more room for learners' individual differences. Keeping these issues in mind, it is obvious that EFL teaching cannot effectively move towards the communication goal without using technology alongside the required TPACK. As a result, the role of EFL TPACK in successful technology integration is obviously significant.

FUTURE RESEARCH DIRECTIONS

Living in the era of communication, TPACK has now become a global phenomenon that interests many researchers around the world. Different researchers investigated TPACK in various ways, from developing a sound theoretical framework for TPACK to assessing teachers' TPACK level in different countries.

However, research on TPACK is far from over. Further research is urged to investigate EFL teachers' TPACK considering their individual differences, attitudes, age and working experience. Experimental research can also be done on manipulating certain variables to measure their effects on improving teachers' TPACK level. It is also recommended to do qualitative studies to triangulate the data and corroborate the findings with the literature.

CONCLUSION

In this era, technology has a huge impact on every aspect of people's lives, from their careers and socializing to education and entertainment. In an attempt to keep up with other fields, educational administrators and teachers need to welcome technology as an inseparable part of the curriculum. Teachers are the key agents of education with the professional responsibility of preparing the new

generation and thus they should be able to use ICT in their classrooms effectively. According to a longstanding body of research, this task is simply not possible without the required knowledge and skills to use technology. Understanding the TPACK framework is the primitive step for all educators in the path of developing their TPACK level. This demands that the educators of different disciplines and specialties become familiar with their field-specific TPACK. To enrich EFL teacher education paradigm, this chapter reviewed and presented the available models/frameworks of TPACK and then proposed the EFL TPACK and how to assess it based on the available literature.

REFERENCES

- Archambault, L., & Crippen, K. (2009). Examining TPACK among K-12 online distance educators in the United States. *Contemporary Issues in Technology & Teacher Education*, 9(1), 71–88.
- Avalos, B. (2011). Teacher professional development in Teaching and Teacher Education over ten years. *Teaching and Teacher Education*, 27(1), 10–20. doi:10.1016/j.tate.2010.08.007
- Baserab, D., Kopchac, T., & Ozdend, Y. (2016). Developing a technological pedagogical content knowledge (TPACK) assessment for preservice teachers learning to teach English as a foreign language. *Computer Assisted Language Learning*. doi:10.1080/09588221.2015.1047456
- Bax, S. (2002). CALL - past, present and future. *System*, 31, 13–28.
- Blurton, C. (2002). *New Directions of ICT Use in Education*. Available: <http://www.unesco.org/education/educprog/Iwfdl/edict.pdf>
- Bygate, M. (2001). Speaking. In R. Carter & D. Nunan (Eds.), *The Cambridge guide to teaching English to speakers of other languages*. Cambridge, UK: Cambridge University. doi:10.1017/CBO9780511667206.003
- Chai, C. S., Koh, J. H. L., & Tsai, C. C. (2010). Facilitating pre-service teachers' development of technological, pedagogical, and content knowledge (TPACK). *Journal of Educational Technology & Society*, 13(4), 63–73.
- Cox, S., & Graham, C. R. (2009). Diagramming TPACK in Practice: Using an elaborated model of the TPACK framework to analyze and depict teacher knowledge. *TechTrends*, 53(5), 60–69. doi:10.1007/s11528-009-0327-1
- Doering, A., Veletsianos, G., Scharber, C., & Miller, C. (2009). Using the technological, pedagogical, and content knowledge framework to design online learning environments and professional development. *Journal of Educational Computing Research*, 41(3), 3. doi:10.2190/EC.41.3.d
- Elmaifi. (2014). Advantages of using ICT in learning and teaching process. *Ed Tech Review*. Retrieved May 2015 from <http://edtechreview.in/trends-insights/insights/959advantages-of-using-ict-in-learningteaching processes>.
- Graham, C. R., Burgoyne, N., Cantrell, P., Smith, L., St. Clair, L., & Harris, R. (2009). TPACK development in science teaching: Measuring the TPACK confidence of inservice science teachers. *TechTrends*, 53(5), 70–79. doi:10.1007/s11528-009-0328-0
- Hong, K. H. (2010). CALL teacher education as an impetus for L2 teachers in integrating technology. *ReCALL*, 22(1), 53–69. doi:10.1017/S095834400999019X
- Hubbard, P. (Ed.). (2009). *Computer Assisted Language Learning: Vol. 1. Foundations of CALL. Critical Concepts in Linguistics Series*. New York: Routledge.
- Hughes, J. E., & Scharber, C. (2008). Leveraging the development of English-technology pedagogical content knowledge within the deictic nature of literacy. In AACTE's Committee on Innovation and Technology (Eds.), *Handbook of technological pedagogical content knowledge for educators* (pp. 87–106). Mahwah, NJ: Routledge.

- Kleinsasser, R. C. (2013). Language teachers: Research and studies in language(s) education, teaching, and learning in Teaching and Teacher Education, 19852012. *Teaching and Teacher Education*, 29, 86–96. doi:10.1016/j.tate.2012.08.011
- Koehler, M. J., & Mishra, P. (2005). What happens when teachers design educational technology? The development of technological pedagogical content knowledge. *Journal of Educational Computing Research*, 32(2), 131–152. doi:10.2190/OEW7-01WB-BKHL-QDYV
- Koehler, M. J., & Mishra, P. (2008). Introducing TPACK. AACTE Committee on Innovation and Technology (Ed.), *The handbook of technological pedagogical content Knowledge (tpck) for educators*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Koehler, M. J., & Mishra, P. (2009). What is technological pedagogical content knowledge? *Contemporary Issues in Technology & Teacher Education*, 9(1), 60–70.
- Koehler, M. J., Mishra, R., Yahya, K., & Yadav, A. (2004). Successful teaching with technology: the complex interplay of content, pedagogy, and technology. In *Proceedings from the Annual meeting of the society for information technology & teacher education*.
- Liu, S., Liu, H., Yu, Y., Li, Y., & Wen, T. (2014). TPACK: A new dimension to EFL teachers' PCK. *Journal of Education and Human Development*, 3(2), 681–693.
- Mahdi, H. S. (2013). Issues of computer assisted language learning normalization in EFL contexts. *International Journal of Linguistics*, 5(1), 191–203. doi:10.5296/ijl.v5i1.3305
- Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for integrating technology in teacher knowledge. *Teachers College Record*, 108(6), 1017–1054. doi:10.1111/j.1467-9620.2006.00684.x
- Mishra, P., Spiro, R. J., & Feltovich, P. J. (1996). Technology, representation, and cognition: The prefiguring of knowledge in cognitive flexibility hypertexts. In H. van Oostendorp & A. de Mul (Eds.), *Cognitive aspects of electronic text processing* (p. 28). Academic Press.
- Pedersen, J. (2001). Technological determinism and the school. *Journal of Educational Enquiry*, 2(1), 61–65.
- Putnam, R. T., & Borko, H. (2000). What do new views of knowledge and thinking have to say about research on teacher learning? *Educational Researcher*, 29(1), 4–15. doi:10.3102/0013189X029001004
- Rahimi, M. (2008). What do we want teaching-materials for in EFL teacher training programs? *Asian EFL Journal*, 31, 1–17.
- Schmidt, D. A., Baran, E., Thompson, A. D., Koehler, M. J., Mishra, P., & Shin, T. (2009). Technological pedagogical content knowledge (TPACK): The development and validation of an assessment instrument for pre-service teachers. *Journal of Research on Tech.*
- Shahin, I. (2011). Development of survey of technological pedagogical and content knowledge (TPACK). *TOJET-The Turkish Online Journal of Educational Technology*, 10(1), 97–105.
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14. doi:10.3102/0013189X015002004
- Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57(1), 1–22. doi:10.17763/haer.57.1.j463w79r56455411
- Spiro, R. J., & Jehng, J.-Ch. (1990). Cognitive flexibility and hypertext: Theory and technology for the nonlinear and mutlidimensional traversal of complex subject matter. In D. Nix & R. Spiro (Eds.), *Cognition, education, and multimedia: Exploring ideas*. Academic Press.

Tseng, J. (2014). Developing an instrument for assessing technological pedagogical content knowledge as perceived by EFL students. *Computer Assisted Language Learning*, 29(2), 302–315. doi:10.1080/09588221.2014.941369

Warrad, M. I. (2013). *Teacher burnout: Causes and projected preventative and curative interventions*. Drexel University.

Wilson, E., & Wright, V. (2010). Images over time: The intersection of social studies through technology, content, and pedagogy. *Contemporary Issues in Technology & Teacher Education*, 10(2), 220–233.

Yang, Y. (2000). *History of English Education in China (1919-1998)*. Retrieved from ERIC database.

Yurdakul, I. K., Odabasi, H. F., Kilicer, F., Coklar, A. N., Birinci, G., & Kurt, A. A. (2012). The development, validity and reliability of TPACK-deep: A technological pedagogical content knowledge scale. *Computers & Education*, 58(3), 964–977. doi:10.1016/j.compedu.2011.10.012

ADDITIONAL READING

Ansyari, M. F. (2015). Designing and evaluating a professional development programme for basic technology integration in English as a foreign language (EFL) classrooms. *Australasian Journal of Educational Technology*, 31(6), 699–712. doi:10.14742/ajet.1675

Dinh, H., Jordan, K., & Elsdon-Clifton, J. (2015). Measuring TPACK in Vietnam: Issues still remain. In *EdMedia 2015* (pp. 851-856). Association for the Advancement of Computing in Education.

Ekrem, S., & Recep, Ç. (2014). Examining Pre-service EFL Teachers TPACK Competencies in Turkey. *Journal of Educators Online*, 11(2). doi:10.9743/JEO.2014.2.2

Ersanli, C. Y. (2016). Improving Technological Pedagogical Content Knowledge (TPACK) of Pre-Service English Language Teachers. *International Education Studies*, 9(5), 18. doi:10.5539/ies.v9n5p18

Koehler, M. J., Mishra, P., Wolf, L. G., Zellner, A., & Kereluik, K. (2012). Thematic considerations in integrating TPACK in a graduate program. In D. Polly, C. Mims & K. Persichitte (Eds.), *Creating technology-rich teacher education programs: Key issues*. 1–12. Hershey, PA: IGI Global.

Koehler, M. J., Shin, T. S., & Mishra, P. (2012). How do we measure TPACK? Let me count the ways. In R. N. Ronau, C. R. Rakes, & M. L. Niess (Eds.), *Educational technology, teacher knowledge, and classroom impact: A research handbook on frameworks and approaches* 16–31. Hersey, PA: IGI Global. doi:10.4018/978-1-60960-750-0.ch002

Kwangsawad, T. (2016). Examining EFL Pre-service Teachers' TPACK through Self-report, Lesson Plans and Actual Practice. [EduLearn]. *Journal of Education and Learning*, 10(2).

Mozafari, P. (2016). *The perspectives of EFL teachers on the integration of ICT into their teaching practices in the context of Iranian schools: an activity theory perspective* (Doctoral dissertation, University of Warwick).

Tai, S. J. D. (2014, March). Teachers' TPACK Competency Developed and Observed in EFL Classrooms: The Impact of TPACK-in-Action Workshops. In *Society for Information Technology & Teacher Education International Conference* (1), 49-56.

KEY TERMS AND DEFINITIONS

CALL: Computer-assisted language learning, the search for and study of applications of the computer in language teaching and learning.

CK: Content Knowledge, knowledge of concepts, theories, conceptual frameworks.

Digital Natives: A person born or brought up during the age of digital technology and so familiar with computers and the Internet from an early age.

EFL: English as a foreign language: the study of English by nonnative speakers living in a non-English-speaking environment.

ICT Literacy: The ability to use digital technology, communication tools, and/or networks.

PK: Pedagogical Knowledge, generic knowledge about how students learn, teaching approaches, methods of assessment and knowledge of different theories about learning.

TK: Technological Knowledge, an understanding of the way that technologies are used in a specific content domain.

TPACK: A framework to understand and describe the kinds of knowledge needed by a teacher for effective pedagogical practice in a technology enhanced learning environment.

Ubiquitous Teachers' Training and Lessons Learned with the *uProf!* Model

T

Sabrina Leone

Università Politecnica delle Marche, Italy

Giovanni Biancofiore

giovannibiancofiore.com, Italy

INTRODUCTION

The current landscape of open and smart learning challenges calls for disruptive educational policies and systems (Li, Chang, Kravcik, Popescu, Huang, Kinshuk & Chen, 2015; Middleton, 2015). All stakeholders (i.e., policy makers, educational institutions, teachers, learners, parents) have renewed roles and degrees of participation in the learning process. Undeniably, though, teachers are on the frontline. On the one hand they are required to absorb their shifting function into facilitators in technology-enhanced learning environments, and on the other hand they are pressured to become more aware lifelong learners.

This technological uptake and consistent change and innovation have been urged by the decisively faster pace that economic systems, social relations and individuals have had since the diffused embracing of Web 2.0, that is the web of social networking tools (Leone & Guazzaroni, 2010), towards Web 3.0, that is a more connected, open and intelligent network thanks to semantic applications (Spivack, 2013).

Definitely, living, working and learning are smarter and smarter, that is they seamlessly accommodate next generation technology (i.e., smartphones, tablets, tablet PCs, sensor network nodes, contact-less smart cards, RFID and QR codes), socialised (Leone & Biancofiore, 2015), and ubiquitous, that is wireless and characterised by high mobility and embeddedness (El-Bishouty, Ogata, & Yano, 2007; Leone & Leo, 2011a). Most

European countries have made significant investments over the last years with a view to ensuring universal access to information and communication technologies (ICT), with considerable success (Education, Audiovisual and Culture Executive Agency, 2011). Specifically, embedding ICT in education and training systems has required relevant changes across the technological, organisational, teaching and learning environments of classrooms, workplaces, and informal learning settings; further efforts, though, are required in this direction (European Commission, 2008; Education, Audiovisual and Culture Executive Agency, 2011). A precondition for using computers in educational contexts is that they are widely available and users are familiar with them. With reference to formal education, data (Education, Audiovisual and Culture Executive Agency, 2011) show that currently no great disparity between schools in availability of ICT equipment exist, but a lack of educational software and support staff still affect instruction. Thus, the solution to an effective use of ICT in education and training is not technology itself, but an advancement in understanding how smart technologies are and can be used to support learning, and what are the barriers in the way of success.

Being an ICT-integrating teacher means going beyond ICT skills, and developing an understanding of the complex relationships between pedagogy, content and ICT (Alayyar, Fisser & Voogt, 2012). Moreover, while in most European countries centrally promoted online resources and

general pedagogical support are available to guide teachers' practical implementation of innovative technology-enhanced learning in the classroom (according to TIMSS - *Trends in International Mathematics and Science Study 2007*) (Education, Audiovisual and Culture Executive Agency, 2011), no research literature seems to document the activation of tailored in-class techno-pedagogical tutoring to teachers during the implementation. A later study carried out on behalf of the European Commission (Pelgrum, 2010) shows that teachers often have difficulties in implementing ICT in the teaching-learning process and that they need support to accomplish this task. Finally, the outcomes of more recent research experiences (Leone, 2013) highlight that teachers perceive in-class techno-pedagogical tutoring as a basic need for the enhancement of their professional skills and, as a result, of their students' learning.

This article illustrates features and outcomes of the experiences with *uProf!*, a model for teachers' professional training in ubiquitous learning by tablets and Quick Response (QR) codes. The model was designed, implemented and successfully validated in 2012 within a full-immersion, learner-centred and metacognitive course for 80 Italian high school teachers who are part of a school network for the enhancement of curriculum continuity from middle into high school. Later on, *uProf!* has been used to provide continuous professional training and tailored in-class tutoring during these teachers' first implementations of technology-enhanced learning experiences.

BACKGROUND

Ubiquitous devices allow to create a Computer Supported Ubiquitous Learning (CSUL) environment. A CSUL could be embedded in everyday life (Ogata & Yano, 2004) and is characterised by permanency, accessibility, immediacy, interactivity, situatedness and adaptability (Curtis, Luchini, Bobrowsky, Quintana, & Soloway, 2002; Leone & Leo, 2011a). Learning theories for CSUL are authentic learning (Brown, Collins & Duguid,

1989), situated learning (Lave, & Wenger, 1991) and learning by doing (Schank, 1995).

These premises highlight the strong potential that ubiquitous learning (uLearning) provides to a lifelong learner, that is to the activator of economical, cultural and social growth of our society and of the systems within. In a lifelong learning vision, the individual is a primary resource of knowledge: the individual's empowerment extends progressively to his/her context and to society as in a network. Empowerment has to be considered as the success key of the relation between a community of practice and the organization to which it is linked. The "double link" described by Wenger & McDermott & Snyder (2002) is crucial for the survival and the development of both, community of practice and organization: the first conducts the improvement of its members' expertise through informal interaction and learning; the second relies on experts grown in the community of practice to identify and implement strategies, to focus on objectives and outcomes of its units.

Through this process old and new borders of knowledge, competence and skills merge and widen. Coherently, traditional educational systems need to evolve into more open and flexible systems, systems that could allow users to choose a learning path according to their learning styles, needs and interests, at any time along their life. In a lifelong learning vision, change and learning are strictly linked: change requires learning and, vice versa, no learning takes place without change (Leone, 2010).

As a result of the current evolution of education systems, the aims of vocationally-based education are beyond training and instruction, to encompass the development of standard features essential to professional competence and lifelong learning (Savery & Duffy, 1994). These include the acquisition of domain knowledge and competencies, and of generally applicable competencies (i.e., ability to adapt to change and participate in change, communicate, collaborate in groups, and be self-directed lifelong learners) (Minasian-Batmanian, Koppi & Pearson, 2000).

THE *uProf!* MODEL

In the label of this model, *uProf!*, “u” stands for “ubiquitous” and “Prof” for “a professional”, that is, “a person engaged in an activity and having an impressive competence in it”, for “professional”, that is characterized by or conforming to the technical or ethical standards of a profession”, and for “professor”, that is either a teacher and “a person who affirms a faith in something” (in the definitions from the Merriam-Webster dictionary), the belief in ubiquitous learning.

The *uProf!* model has been devised for teachers’ professional training in ubiquitous learning, in order to support them in acquiring the mind-set (beyond the skill-set) to successfully facilitate technology-enhanced learning. The *uProf!* model has been developed in the vision of lifelong learning and thus of the individuals’ empowerment, that is the teachers’ and subsequently their students’ empowerment, by learner-centred environments characterised by a bottom-up approach, personalisation and adoption of new technologies as means of re-mediation of networked knowledge (Leone & Guazzaroni, 2010).

Needs, Objectives, and Competences

The *uProf!* model aims to meet teachers’ need for knowledge and skills to implement effectively inclusive ubiquitous learning environments, and for in-classroom tailored tutoring during initial implementations.

General learning objectives of the *uProf!* model are: supporting the acquisition and improvement of techno-pedagogical knowledge, competences and skills that teachers could spend in their different domains; providing teachers with tools to personalise and making flexible their students’ learning.

Specific learning objectives are: increasing teachers’ awareness of the relevance of digital skills for the development of active citizenship; enriching the teachers’ professional and personal

background by knowledge exchange practices, on the basis of innovative teaching-learning experiences with ICT; allowing the acquisition of innovative teaching methodologies for the promotion of key competences, with a focus on the *QRcode* format as a framework that supports uLearning (Leone & Leo, 2011b; Leone, 2012; Leone, 2014).

The *uProf!* model can develop higher cognitive, digital and professional competences as follows: by increasing learners’ awareness of the relevance of digital skills for the development of active citizenship; by selecting the most suitable innovative teaching-learning methodologies, on the background of lifelong learning, in relation to the context, aiming at empowerment, inclusion and personalised learning; by using the innovative teaching-learning methodologies and materials acquired; by implementing the *QRcode* format.

Paradigm and Learning Strategies

The *uProf!* model draws on socio-constructivism (Vygotsky, 1986; Varisco, 2002) and andragogy (Knowles, 1970). It foresees the adoption of brainstorming, problem solving, learning by doing, collaborative and cooperative learning, game based learning, simulations in presence and in immersive virtual environments, and metacognition. The model can be an effective inclusive learning tool since it facilitates personalised and flexible learning if a learner-centred approach is adopted, an approach attentive to each learner’s diverse needs (UNESCO, 2009).

Organization: Learning Environments, Tools, and Human Resources

The physical places to implement the *uProf!* model are the classroom with a wireless connection or a hot spot, where learners operate, individually or in a group, on paper-based and digital learning material, by the various ubiquitous devices available, and any other place by ubiquitous devices with an Internet connection. The learn-

T

ing environment is, therefore, both physical (i.e., the classroom and the learners) and virtual (i.e., web-based, with interactive activities in and out of a distributed community of practice - Wenger, McDermott & Snyder, 2002). Various systems (e.g., Learning Management Systems) and tools can be used to implement the *uProf!* model, according to the learners' personal and professional needs, priorities and learning styles. Specifically, personalised learning is provided by the use of ubiquitous devices (anytime, anywhere learning), by the adoption of adaptive mechanisms (e.g., *conditional activities* in Moodle) which operate upon the assessment of learners' learning styles and prior and progressively acquired knowledge, and by the use of multimodal learning materials (Leone, 2013).

The length of the implementation can vary in relation to the number and the extension of the uLearning modules of the designed path.

Necessary human resources are: one or more (teachers') trainers, that define strategies, objectives, contents and activities; a learning technologist, who has extensive knowledge, management and troubleshooting expertise of all the educational technologies that could support learning and teaching within the path; an ICT technician, that installs/checks the QR code decoder software in the mobile devices in use, checks the Internet connection and deals with general hardware and software troubleshooting.

The implementation of the *uProf!* model in a distributed learning environment requires an Internet connection and laptops, smartphones, tablets or tablet PCs, together with the use of sensor network nodes, contact-less smart cards, RFID (Radio Frequency Identification) and QR codes.

Evaluation and Assessment

The *uProf!* model is monitored at the beginning and at the end of its implementation, in order to evaluate its effectiveness and rating from the participants' point of view (evaluation), and to assess the achievement of learning objectives

(assessment). For the evaluation, the data are collected (1) by entry and exit anonymous surveys, to record teachers' digital skills, previous technology-enhanced teaching-learning experiences, motivation and expectations on the model (entry survey), and feedback (exit survey). Assessment is entirely based on self- and peer- assessment; it is carried out by close and semi-structured tests (entry test on prior knowledge and competences), direct observation (ongoing assessment) and project works and role-plays (summative test at the end of the learning path). Assessment parameters are: effectiveness of the teaching/learning methodology adopted in the project works and cohesion with the learning scenario; competence in the use of the teaching/learning methodology adopted in the project works.

Recommendations for an Optimal Implementation

The *uProf!* model is very flexible for both the trainer-facilitator and the learners (i.e., teachers), and it can be adapted to all contexts and used for all knowledge domains. Anyhow, essential elements for the success of the experience are: (1) thorough organisation and management of the necessary hardware and software; (2) an adequate familiarization of the learners with the learning environment (technology, tools and learning approach); (3) a light e-moderation by the facilitator, in order to provide a modulation of self-regulated and shared learning on the basis of the learners' silent and/or expressed requests, with the aim of supporting participants' high motivation.

uPROF! IN COMUNITA' DI PRATICA PROGETTO EUREKA

The *uProf!* model was validated within a full-immersion, learner-centred and metacognitive course for 80 Italian high school teachers, within the *Eureka* project (2012-2014), a network of 11 schools in Apulia, Italy, for the enhancement

of curriculum continuity from middle into high school.

The *QRcode* format (Leone & Leo, 2011b; Leone, 2012; Leone, 2014) was selected as element of techno-pedagogical innovation through the *Eureka* project, and for this reason played a central role in the development of the model. One of the authors of this article was the coordinator of the techno-pedagogical innovation of the project and both were the trainers.

The initial teachers' training for the *Eureka* project was delivered in October-November 2012 in two in-presence 4-hour workshops; participants worked in 16 groups of 5 and in few plenary sessions, and they used tablets and laptops. In addition, a distributed community of practice (Wenger, McDermott & Snyder, 2002) was activated, the *Comunità di Pratica Progetto Eureka*, in the eLearning space www.elearnigplace.it/corsi (Moodle) that the trainers own and manage. A large number of the tools available in Moodle were used: its adaptive mechanism of *conditional activities* to provide trainees with personalised and scaffolded learning; a learners' forum for brainstorming and problem solving, plenary discussions, knowledge exchange and sharing of best practices; group forums for collaborative and cooperative group works; tags for knowledge co-construction and sharing; assignments for individual or group works to be handed in for feedback; questionnaires and grade book for self-assessment; logs for self-monitoring learners' involvement; instant messaging for personal communications. In addition, Google Docs, for the co-production of final project works, and SurveyMonkey, for the completion of entry and exit surveys, were integrated in Moodle.

Besides being the trainees' environment for the metacognitive experience of their collaborative and cooperative learning in presence and in eLearning, the *Comunità di Pratica Progetto Eureka* was proposed as virtual environment for the realisation of the aims of the *Eureka* project itself, that is to create a network among the 11

school partners and enhance curriculum continuity from middle into high school. Hence, the *Comunità di Pratica Progetto Eureka* offered enormous potential to teachers (and, subsequently, to their students and to the institutions of which they are part), provided that the members of this community engaged in a constructive interaction for the construction and sharing of best practices.

A one-week phase of scaffolded familiarisation with the learning environment and materials was held in *Comunità di Pratica Progetto Eureka* before the workshops.

As a whole, in order to better meet institutional time constraints and trainees' learning needs, these first training sessions were made as much flexible and personalised as possible, relying on participants' and partner schools' active participation. The entire training aimed to (1) support the acquisition and improvement of techno-pedagogical knowledge, competences and skills that trainees could spend in their different disciplines; (2) provide trainees with tools to personalise and make flexible their students' learning.

Contents were focused on digital skills as necessary crosscurricular competences to be citizens of the 21st century and on the *QRcode* format (what it is and how to use it – research experiences and case studies carried out, and group *role-plays* of its implementation in an interdisciplinary scenario). The learning path included the development of innovative technology-enhanced teaching-learning methodologies (brainstorming, problem solving, learning by doing, metacognition, collaborative and cooperative learning, peer tutoring, project work). In-depth contents on inherent definitions, concepts and approaches were provided in this first phase, to subsequently facilitate the implementation of possible (interdisciplinary and not) scenarios (single learning modules or entire projects) directly in the classroom by tailored tutoring for each trainee.

Indeed, in the following months 4 of the 11 schools of the *Eureka* project network started interdisciplinary projects with the *QRcode* for-

mat. In all, 23 teachers were engaged, and they were supported by in-class tutoring on the *uProf!* model. In detail:

- In April-May 2013 (4 weeks, 9 hours a week for each class) an interdisciplinary path was developed on Italian, music, history, geography, French and maths with 4 3rd-year classes of a middle school together with 4 1st-year classes of a high school, 193 students and 14 teachers in all.
- In January-February 2014 (4 weeks, 9 hours a week for each class) an interdisciplinary path on citizenship education, geography, Italian, the arts and music with 4 3rd-year classes of a middle school together with 4 1st-year classes of high school, 187 students and 9 teachers in all.

Outcomes

The final feedback surveys pointed out that time constraints were one of the critical elements of the training sessions, but all the causes (limited budget, delays in institutional communication within the project network, missed deadlines in carrying out familiarisation and group activities) were independent from the *uProf!* model, that, instead, was successfully validated by the trainees' satisfaction and learning outcomes.

In detail, participants perceived both experiences of initial training and of in-class tutoring during the implementations of the interdisciplinary paths with the *QRcode* format as new, interesting and useful (100%) (especially for the hands-on and reusable activities and contents), flexible (94.8%), but tiring as well (57.9%) because of poor Internet connection (36.8%), time constraints (48.2%) and consequent difficulty in getting oriented with learning materials and environments (26.3%).

The respondents pointed out that learning with ICT showed the following disadvantages: it required familiarisation with technology-enhanced learning (mindset) (84.2%) and a high degree of technological know-how (57.9%); viceversa,

learning with ICT offered the following advantages: independence from time and space (94.8%), personalised learning (100%) and flexible contents (94.8%).

At the end of this experience, 52.6% of the participants affirmed that they were gratified for having learned to do new things, 21.1% were gratified for having learned new things, and the rest were frustrated because they had not been able to exploit it at the best (21.1%), deceived because it had not met their learning needs (10.5%) and happy because it was over, at last (5.3%). Finally, 78.9% would have been happy to repeat this experience in order to keep learning, to improve as a teacher, to enhance their collaborative and cooperative work skills, and to be part of a learning network.

Finally, the global perception of the experiences was positive because they represent a first opening to what learning has already become in everyday life thanks to the affordances of ICT. Nevertheless, evidence shows the need for more flexible education systems by the development of the culture of knowledge exchange, cooperative work, sense of belonging to a community as signal of growth of the system as a whole and the individual as its activator. This implies necessarily endeavours to bring oneself into question and, thus, to accept the relativism of one's professional eco system; a re-allocation of school funds is vital as well to tutor teachers directly in the classroom during the implementation of various technology-enhanced learning formats.

FUTURE RESEARCH DIRECTIONS

In the near future, the *uProf!* model will be implemented further and extensively in different grades of instruction to provide continuous professional training to teachers.

Accordingly, future research issues will be the analysis of the impact of the model on the teachers' empowerment and on their institutions, the validation of teachers' informal learning during the application of the *uProf!* model, and the ef-

fects of the use of the *uProf!* model on students' achievement of crosscurricular skills.

Further, the model could be included in a research experience in the field of business management to observe its effects and to identify margins of improvement.

CONCLUSION

Technological uptake and consistent change and innovation in education have been urged by the decisively faster pace of the knowledge society since the emergence of Web 2.0, towards Web 3.0.

Undeniably, living, working and learning are more and more ubiquitous. Embedding ICT in education and training systems has required relevant changes across the technological, organisational, teaching and learning environments of classrooms, workplaces, and informal learning settings; further efforts, though, are required in this direction. Teachers today are expected to be able to apply a wide range of digital technologies in the classroom, but they need practical support. In-class tailored techno-pedagogical tutoring is a basic need for the enhancement of their professional skills and, as a result, of their students' learning. In this regard, the *uProf!* model of teachers' professional training in ubiquitous learning by tablets and QR codes was designed, implemented and successfully validated within a full-immersion, learner-centred and metacognitive course for 80 Italian high school teachers who were part of a school network for the enhancement of curriculum continuity from middle into high school. Later on, the model was adopted to provide in-class tutoring to the teachers engaged in technology enhanced interdisciplinary implementations.

The global perception of the experience was positive. Nevertheless, margins of improvement have emerged in the participants' digital skills, in their approach to teaching and learning, in their approach to the adoption of ICT in the curriculum as a chance of holistic change and growth towards smart inclusive education.

REFERENCES

- Alayyar, G. M., Fisser, P., & Voogt, J. (2012). Developing technological pedagogical content knowledge in pre-service science teachers: Support from blended learning. *Australasian Journal of Educational Technology*, 28(8), 1298–1316. doi:10.14742/ajet.773
- Barr, R., & Tagg, J. (1995). From Teaching to Learning: A New Paradigm for Undergraduate Education. *Change Magazine*, 2(12), 8-12. Retrieved January 11, 2008, from www.cic.uiuc.edu/resources/deo/paradigm.html
- Brown, J. S., Collins, A., & Duguid, P. (1989). *Situated Cognition and the Culture of Learning. In Educational Researcher* (pp. 32–42). Jan.-Feb.
- Curtis, M., Luchini, K., Bobrowsky, W., Quintana, C., & Soloway, E. (2002). Handheld Use in K-12: A Descriptive Account. *Proceeding of the International Workshop on Wireless and Mobile Technologies in Education*, Los Alamitos, CA: IEEE Computer Society. doi:10.1109/WMTE.2002.1039217
- Education, Audiovisual and Culture Executive Agency. (2011). *Key data on learning and innovation through ICT at school in Europe 2011*. Brussels: EACEA.
- El-Bishouty, M. M., Ogata, H., & Yano, Y. (2007). PERKAM: Personalized Knowledge Awareness Map for Computer Supported Ubiquitous Learning. *Journal of Educational Technology & Society*, 10(3), 122–134.
- European Commission. (2008). *Commission Staff Working Document on The use of ICT to support innovation and lifelong learning for all – A report on progress. SEC(2008) 2629 final*. Brussels: European Commission.
- Knowles, M. S. (1970). *The Modern Practice of Adult Education: Andragogy versus Pedagogy*. New York: Associated Press.

- Leone, S. (2010). F2F learning vs eLearning: the lifelong learner's point of view. *Proceedings of INTED 2010*.
- Leone, S. (2012). Il format *QRcode* in un'esperienza di apprendimento cooperativo e interdisciplinare. *Proceedings di Didamatica 2012*.
- Leone, S. (2013). *Characterisation of a Personal Learning Environment as a lifelong learning tool*. New York, USA: Springer Science & Business Media. doi:10.1007/978-1-4614-6274-3
- Leone, S. (2014). The *QRcode* format: a tool for inclusive, personalised and interdisciplinary learning experiences. In M. Khosrow-Pour (Ed.), *Encyclopaedia of Information Science and Technology* (3rd ed.). Hershey, PA: IGI Global.
- Leone, S., & Biancofiore, G. (2015). Leveraging teamwork by Google+ in a lifelong learning perspective. *Knowledge Management & E-Learning: An International Journal (KM&EL)*, 7(2).
- Leone, S., & Guazzaroni, G. (2010). Pedagogical sustainability of interoperable formal and informal learning environments. In F. Lazarinis, S. Green, & E. Pearson (Eds.), *Developing and utilizing e-learning applications*. IGI-Publisher. doi:10.4018/978-1-61692-791-2.ch001
- Leone, S., & Leo, T. (2011a). The synergy of paper-based and digital material for ubiquitous foreign language learners. *Knowledge Management & E-Learning: An International Journal (KM&EL)*, 3(2).
- Leone, S., & Leo, T. (2011b). Il format *QRcode* in Learning for All. *Proceedings di Didamatica 2011*.
- Li, Y., Chang, M., Kravcik, M., Popescu, E., Huang, R. Kinshuk, & Chen, N.-S. (Eds.). (2015). *State-of-the-Art and Future Directions of Smart Learning*. Singapore: Springer.
- Middleton, A. (Ed.). (2015). *Smart learning: Teaching and learning with smartphones and tablets in post compulsory education*. Sheffield, UK: Media-Enhanced Learning Special Interest Group and Sheffield Hallam University.
- Minasian-Batmanian, L. C., Koppi, A. J., & Pearson, E. J. (2000). Developing lifelong learners: a novel online problem-based ultrasonography subject. In D. Squires, G. Conole, & G. Jacobs (Eds.), *The Changing Face of Learning Technology* (pp. 137–148). Cardiff, UK: University of Wales Press.
- Ogata, H., & Yano, Y. (2004). *Context-Aware Support for Computer-Supported Ubiquitous Learning*. Paper presented at the 2nd IEEE International Workshop on Wireless and Mobile Technologies in Education, JhongLi, Taiwan.
- Pelgrum, W. J. (2010). *Study on Indicators of ICT in Primary and Secondary Education (IIPSE)*. Brussels: European Commission, Directorate General Education and Culture.
- Savery, J. R., & Duffy, T. M. (1994). Problem based learning: An instructional model and its constructivist framework. *Educational Technology*, 8, 31–38.
- Schank, C. (1995). *What We Learn When We Learn by Doing*. Technical Report No. 60. Northwestern University, Institute for Learning Sciences.
- Spivack, N. (2013). Web 3.0: The third generation of the web is coming. *Lifeboat Foundation*. Retrieved July 29, 2013, from <http://lifeboat.com/ex/web.3.0>
- UNESCO. (2009). *Policy Guidelines on Inclusion in Education*. Paris: UNESCO.
- Varisco, B. M. (2002). *Costruttivismo socio-culturale. Genesi filosofiche, sviluppi psico-pedagogici, applicazioni didattiche*. Roma: Carocci.
- von Glasersfeld, E. (1998). *Il costruttivismo radicale. Una via per conoscere ed apprendere*. Roma: Società Stampa Sportiva.
- Vygotsky, L. S. (1986). *Thought and Language*. Cambridge, MA: MIT Press.
- Wenger, E., McDermott, R., & Snyder, W. M. (2002). *Cultivating Communities of Practice*. Boston: Harvard Business School.

ADDITIONAL READING

Alberici, A., Catarsi, C., Colapietro, V., & Liodice, I. (2007). *Adulti e Università. Sfide ed innovazioni nella formazione universitaria e continua*. Milano, Italia: Franco Angeli.

Bereiter, C. (2002). *Education and Mind in the Knowledge Age*. Mahwah, NJ: Lawrence Erlbaum Associates.

Billett, S. (2008). Learning throughout working life: A relational interdependence between personal and social agency. *British Journal of Educational Studies*, 56(1), 39–58. doi:10.1111/j.1467-8527.2007.00394.x

Broek, S. D., & Buiskool, B. J. (2010). *Key Competences for Adult Learning Professionals*. Zoetermeer: Research voor Beleid.

Broek, S. D., & Buiskool, B. J. (2011). Identifying a common set of key competences for Adult Learning staff: An inventory of European practices. *Journal of Adult and Continuing Education*, 17(1), 40–62. doi:10.7227/JACE.17.1.5

Calvani, A. (2006). *Reti, comunità e conoscenza. Costruire e gestire dinamiche collaborative*. Trento: Erikson.

Chen, N. S., Chia-En Teng, D., & Lee, C. H. (2010). Augmenting Paper-based Reading Activities with Mobile Technology to Enhance Reading Comprehension. *Proceedings of WMUTE2010*, April 12–16 2010, Kaohsiung, Taiwan. doi:10.1109/WMUTE.2010.39

Chen, W., Lim, C., & Tan, A. (2010). Pre-Service teachers' ICT experiences and competencies: New generation of teachers in digital age. *Proceedings of the 18th International Conference on Computers in Education*. Putrajaya, Malaysia: Asia-Pacific Society for Computers in Education.

EAEA (2006). *The Common European Adult Learning Framework*.

Edmunds, R., Thorpe, M., & Conole, G. (2012). Student attitudes towards and use of ICT in course study, work and social activity: A technology acceptance model approach. *British Journal of Educational Technology*, 43(1), 71–84. Retrieved July 29, 2013 doi:10.1111/j.1467-8535.2010.01142.x

European Commission. (2001). *Making a European area of lifelong learning a reality*. Brussels: European Commission.

European Commission. (2002). *European Report on Quality Indicators of Lifelong Learning*. Brussels: European Commission.

Jonassen, D. H., & Land, S. M. (2000). *Theoretical Foundations of Learning Environment*. New Jersey: Lawrence Erlbaum Associates.

Keegan, D. (2004). *The incorporation of mobile learning into mainstream education and training*. Paper presented at The 18th Asian Association of Open Universities Annual Conference.

Kukulska Hulme, A., Sharples, M., Milrad, M., Arnedillo-Sánchez, I., & Vavoula, G. (2009). Innovation in Mobile Learning: A European Perspective. *International Journal of Mobile and Blended Learning*, 1(1), 13–35. doi:10.4018/jmbl.2009010102

Lave, J., & Wenger, E. (1991). *Situated Learning. Legitimate Peripheral Participation*. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511815355

Leo, T., Manganello, F., & Chen, N.-S. (2010). From The Learning Work To The Learning Adventure. *Proceedings of EDEN 2010 Annual Conference*, 9–11 June, Valencia (Spain).

Leone, S. (2008). Lifelong learning and tertiary education in Italy. *The International Journal of Interdisciplinary Social Sciences*, 3(2), 7–14. doi:10.18848/1833-1882/CGP/v03i02/52529

Leone, S. (Ed.). (2014). *Synergic integration of formal and informal eLearning environments*. Hershey, PA: IGI Global.

Nalder, J. (2008). The dawn of uLearning. Master thesis. Retrieved March 10, 2010 from <http://www.scribd.com/doc/12398804/The-dawn-of-uLearning-Jonathan-Nalder-Masters-thesis>

Ramsden, A. (2009.) *The potential of QR Codes in Education. JISC Emerge Users & Innovation Programme Meeting*. Retrieved August 12, 2009, from <http://www.slideshare.net/andyransden/potential-qr-codes-education-emerge-2009-presentation>

Saravani, S.A., & Clayton, J.F. (2009). A conceptual model for the educational deployment of QR codes. *Same places, different spaces. Proceedings Ascilite Auckland 2009*.

Thorpe, M., & Edmunds, R. (2011). Practices with technology: Learning at the boundary between study and work. *Journal of Computer Assisted Learning*, 27(5), 385–398. Retrieved July 29, 2013 doi:10.1111/j.1365-2729.2010.00405.x

UNESCO Institute for Education. (1999). *Glossary of Adult Learning in Europe*. Paris, France: UNESCO.

Wasson, B. (2007). Design and Use of Collaborative Network Learning Scenarios: The DoCTA Experience. *Journal of Educational Technology & Society*, 10(4), 3–16.

Wenger, E. (1998). *Communities of practice: Learning, meaning and identity*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511803932

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into Practice*, 41(2), 64–72. doi:10.1207/s15430421tip4102_2

KEY TERMS AND DEFINITIONS

Adult Education: All forms of non-vocational adult learning, whether of a formal, non-formal or informal nature.

CSUL (Computer Supported Ubiquitous Learning) Environment: Everyday ubiquitous learning informed by the theories of authentic learning (Brown, Collins & Duguid, 1989), situated learning (Lave, & Wenger, 1991) and learning by doing (Schank, 1995).

Formal Learning: Hierarchically structured, chronologically graded educational system running from primary through to tertiary institutions.

Informal Learning: Unstructured learning that allows persons to acquire attitudes, values, skills and knowledge from daily experience, within the individual's environment (i.e., family, friends, peer groups, etc.).

Lifelong Learners: Self-regulated learners, characterized as demonstrating perseverance, initiative, and adaptive abilities. Self-regulation relates to an ability to recognize a need for further learning as well as to be proactive in gaining access to and accomplishing learning.

Lifelong Learning: A holistic vision of learning in different contexts (formal, non-formal and informal) and throughout life, based on the evolution of provider-driven education toward personalised learning and aiming at improving knowledge, skills and competencies within a personal, civic, social and/or employment-related outlook.

Non-Formal Learning: Learning that takes place through education organized for specific learners with specific learning objectives, outside the formal established system.

QR (Quick Response) Code: A bidimensional code (it displays information in both vertical and horizontal directions) that can hold larger amounts and different kinds of contents (e.g., website addresses, texts, numerical information, contact details) than a normal bar code (monodimensional). The information stored in a QR code can be readily decoded and accessed by a mobile device with an embedded camera and free code reading software installed.

QRcode Format: A technology-enhanced learning format to be implemented in a learner-

centred learning environment to offer inclusive, personalised (different learning styles and goals) and flexible (anytime, anywhere) learning by the integration of paper-based and digital learning materials through QR code.

Ubiquitous Learning: Wireless learning supported by a large number of cooperative small nodes with computing and/or communication capabilities (e.g., handheld devices, sensor network nodes, contact-less smart cards, RFID and QR codes) and characterised by high mobility and embeddedness.

T

Video Considerations for the World Language edTPA

Elizabeth Goulette

Georgia State University, USA

Pete Swanson

Georgia State University, USA

INTRODUCTION

For decades in the United States, teacher education has been both a political and social focus. As part of President Lyndon B. Johnson's War on Poverty, Congress passed the *Elementary and Secondary Education Act of 1965* (ESEA) in 1965 (P.L. 89-10). This six-part groundbreaking federal legislation placed an emphasis on equal access to education while setting high standards for academic performance and demanding accountability from schools and districts within a framework of six titles. Additionally, the ESEA funded primary and secondary school education, with the goal of decreasing the achievement gap. A year later, two amendments were proposed and passed: Title VI – Aid to Handicapped Children and Title VII – Bilingual Education Programs. At the turn of the new millennium, the landmark legislation was reauthorized by the George W. Bush administration and retitled as *No Child Left Behind* (NCLB, U.S. Department of Education, 2002).

The unfunded mandate demanded even more rigorous testing and accountability of k-12 student learning and included foreign language as part of the core curriculum. While the legislation's philosophical merits (e.g., a highly effective teacher in every classroom) were difficult to dispute, researchers criticized NCLB because it narrowed the k-12 curriculum and prioritized reading, mathematics, and science instruction over

non-tested content areas, such as foreign languages (Rosenbusch, 2005; Rosenbusch & Jensen, 2004). Additionally, teachers had to be considered *highly qualified*. That is, educators in public schools had to have: 1) a bachelor's degree, 2) full state certification or licensure, and 3) prove that they know each subject they teach (U. S. Department of Education, 2004).

When the legislation was reauthorized by the Obama Administration, the law demanded further scrutiny of schools and practitioners. *Race to the Top* required states to measure beginning and veteran teacher effectiveness in order to receive full federal funding (U.S. Department of Education, 2009). Eager to secure federal funds, states began to pass legislation that centered on pre-service teacher preparation and certification standards, emphasizing teacher performance and effectiveness at the state level (e.g., Georgia Professional Standards Commission, 2014; Illinois State Board of Education, 2012). As funding was awarded, states required to use student learning as evidence in teacher evaluation practices (Darling-Hammond, 2012), and in many states throughout the country, pre-service teacher candidates have to demonstrate "the results of classroom processes, such as impact on student learning" (Goe, Bell, & Little, 2008, p. 4), often through teacher performance assessments.

In December 2015, Congress reauthorized the law, now known as the *Every Student Succeeds*

Act (ESSA). Maintaining the rigor of its predecessors, the new legislation leaves the majority of the details regarding teacher education, qualifications, and certification procedures up to the states (U.S. Department of Education, 2015). Aligned with the federal legislation, a new national teacher portfolio, edTPA, has been developed and pilot tested in a variety of states. However, while novel, this new externally-reviewed portfolio is problematic (Hildebrandt & Swanson, 2014), especially in the area of assessment of learning. In this chapter, the authors present edTPA in the context of world languages and then discuss how the Integrated Performance Assessment can be used as teacher candidates develop their portfolios in order to successfully pass edTPA.

BACKGROUND

Over the past 60 years, educational legislation can be categorized into four distinct topics (Cochran-Smith, 2000): teacher attributes (early 1950s to late 1960s), teacher effectiveness (late 1960s through the mid-1980s), teacher knowledge (early 1980s and continuing through the 1990s), and teacher outcomes (2000 to present). Currently, as promoted by legislation, the goal for education in the US is improving student learning, and the measurement of teacher effectiveness is a critical component, which as Muijs (2006) noted, remains challenging.

Teacher Portfolio Assessment

For decades, pre-service teacher candidates' performance as novice educators has been measured via portfolio assessments. *The Teacher Work Sample*, developed by Renaissance Partnership for Improving Teacher Quality (2002), required teacher candidates to create an assessment plan, provide evidence of instructional decision making, use student learning to adjust their teaching, interpret student data, and communicate with

others about students' progress. A few years later, the *Performance Assessment for California Teachers* (PACT) was designed to measure the candidate's knowledge, skills and ability to teach and assess k-12 students (Sato, 2014). Grounded in the National Board for Professional Teaching Standards processes for veteran teachers, the PACT served as the basis for the development of edTPA (Sato, 2014.)

edTPA

Developed by the Stanford Center for Assessment, Learning, and Equity (SCALE) at Stanford University, edTPA is a nationally-available performance assessment of novice educators' readiness to teach in 27 different content areas (SCALE, 2013). Working in conjunction with the American Association of Colleges for Teacher Education and administered by Pearson, edTPA can be administered across institutions and reliably scored by experts in teaching (Sato, 2014). Presently, edTPA is in various stages of implementation in 666 Educator Preparation Programs in 36 states and the District of Columbia participating in edTPA (American Association for Colleges of Teacher Education, 2016) in order to inform initial teacher licensure and certification decisions. For example, Hawaii is going to require a performance-based assessment such as edTPA for teaching candidates beginning in 2017, while Georgia and Illinois began requiring edTPA for certification purposes in September 2015.

Typically, portfolios such as edTPA are developed during a teacher candidate's final field placement, also known as student teaching. The edTPA focuses on three areas: (1) Planning for Instruction and Assessment, (2) Instructing and Engaging Students in Learning, and (3) Assessing Student Learning. World Language teacher candidate effectiveness is measured by thirteen 5-point Likert-scale rubrics. Thus, total scores range between 13 to 65 points. The 5-point scale describes teacher candidates' knowledge and skills

T

ranging from individuals who are not ready to teach to very well-qualified individuals who are ready to teach (SCALE, 2013). It is important to note that beginning teachers are not expected to be in the advanced range of the rubrics.

States are able to determine the levels to which teacher candidates must perform on edTPA and most teacher assessments. For example, a qualifying score of 29 for the World Language edTPA was set in Georgia for Fall 2015 enactment (Georgia Professional Standards Commission, GAPSC, 2015). In 2017, the required score in Georgia will increase to 32 for World Language teacher candidates (GAPSC, 2015). Other states have set cut scores as well: Washington (30), Illinois (31), and New York (35).

Research on the World Language edTPA shows that of the three tasks, teacher candidates “were most successful in the planning tasks and least successful in the assessment tasks” (Hildebrandt & Swanson, 2014, p. 584). Such findings may be explained by what takes place in traditional world language teacher education programs. That is, teacher candidates have ample opportunities to plan for instruction and even teach lessons with students in k-12 schools when participating in field placements like practica and student teaching. With respect to practica placements, teacher candidates observe teachers of record for pedagogical, classroom management, and even assessment purposes. Nevertheless, the teacher candidates are not given access to student data. During their field placement, student teaching, teacher candidates have access to k-12 student data because they are in charge of the classroom typically, but it is during student teaching when teacher candidates develop their edTPA dossiers. Thus, prior to student teaching, teacher candidates do not have sufficient opportunities to collect and analyze data, which is required by edTPA. By the time they have access to student data, it is too late for them to learn how to analyze and write up their findings.

BEST PRACTICES IN WORLD LANGUAGE ASSESSMENT OF STUDENT LEARNING

The teaching and learning of world languages has seen a paradigm shift in recent years. Once known as teaching the four skills (i.e., reading, writing, listening, and speaking) in the target language, world language instruction and assessment is now conceptualized in the communicative language teaching approach that focuses on the three modes of communication: interpretive, interpersonal, and presentational (World-readiness Standards for Foreign Language Learning, 2015). Best practices in language teaching for proficiency in the target language places an emphasis on developing and demonstrating proficiency through performance-based assessments. Through such assessments, students can work individually or collaboratively, use their collection of skills and knowledge to create a response to a prompt (complex questions or situations) or a product that can have more than one correct response (Liskin-Gasparro, 1996, 1997; Wiggins, 1998). Performance-based assessments reflect the tasks and challenges language learners will face in real world scenarios.

Student Proficiency Assessment in the Target Language

The Integrated Performance Assessment (IPA) serves as an evaluation of student ability in the target language that is used as a cluster assessment featuring three tasks with one task in each of the three modes of communication (Adair-Hauck, Glisan, Koda, Swender, & Sandrock, 2006). The IPA is a multi-task assessment conceptualized within a single thematic context. A graphic representation of an IPA is presented in Figure 1. First, language learners complete an interpretive task (e.g., reading). Afterward they use the information from an interpersonal task (e.g., conversation) before they summarize their learning with a pre-

sentational task. In other words, language learners view, listen to, and/or read authentic texts in the target language, interact with learners in the target language in oral and written form, and then present in oral and written form to audiences of listeners and readers.

USING VIDEO AND VIDEO CONSIDERATIONS

Teacher Candidate Skills Using Video

Teacher candidates working on the edTPA possess a wide range of experiences using technology in the classroom. Some have never video recorded themselves teaching before whereas others have completed class assignments which required them to video record their teaching and reflect upon its effectiveness. While it may be an awkward process for students to watch themselves on camera, research supports using video recording as

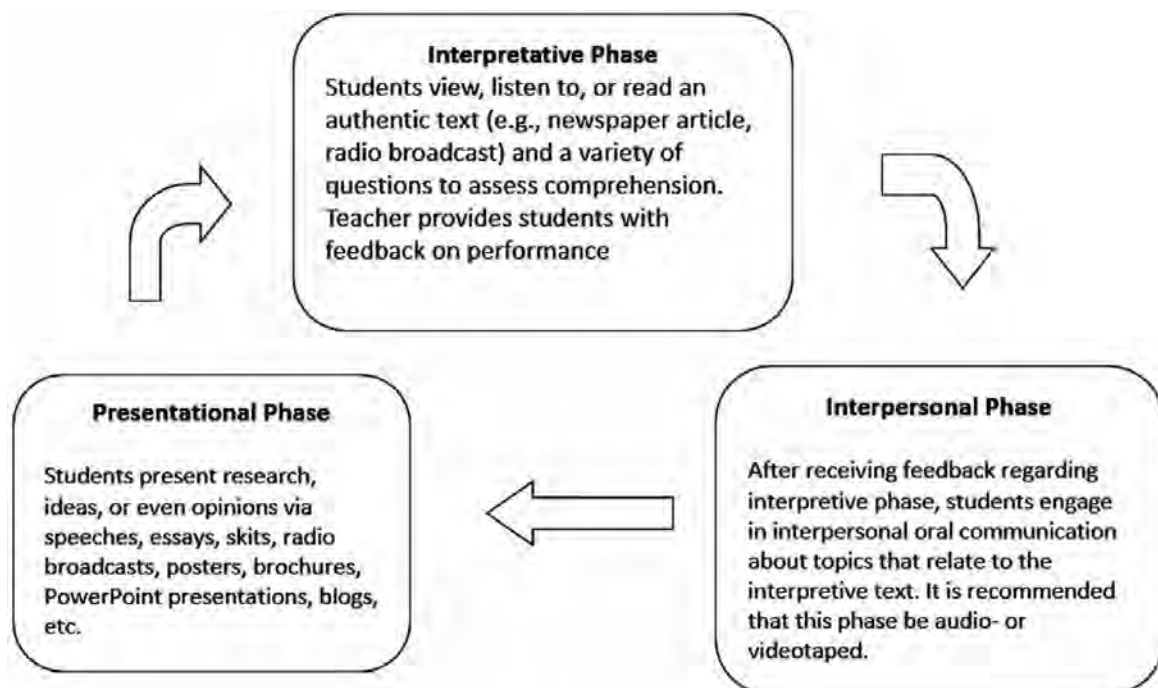
a powerful pedagogical tool to facilitate teacher candidates' self-reflection and growth (Baecher, McCormack, & Kung, 2014; Calandra, Brantley-Dias, Lee, & Fox, 2009; Danielowich & McCarthy, 2013; Tripp & Rich, 2012).

Teacher preparation programs also vary widely when it comes to the degree to which technology is incorporated into required coursework for pre-service teachers. Some teacher preparation programs mandate that their teacher candidates pass a dedicated technology course and other programs embed technology throughout the curriculum. Despite the wide continuum upon which these pre-service teachers fall regarding technology skills in the classroom, successful recording is essential for passing edTPA and cannot be avoided.

Securing Parental/Guardian Consent

An important prerequisite to the recording phase is garnering approval from parents/guardians of the students in the focal class that has been selected. Once teacher candidates have identified the focal

Figure 1.



class that they will record for edTPA, it is imperative that they immediately notify school administration, their k-12 cooperating teacher, students in the focal class, and their parents/guardians about the recording process and requirements. If school administrators and/or cooperating teachers refuse to allow a teacher candidate to record in their schools and/or classrooms, university supervisors will be required to assign new placements. A natural consequence of this type of logistical problem is that teacher preparation programs may begin to narrow the scope of schools that they place their student teachers in for their field experiences, based on prior consent for, and experience with video recording.

Many teacher preparation programs have developed their own video recording permission forms for teacher candidates' use on the edTPA. It is important to note, though, that while some schools require parents/guardians to sign a school specific consent form for video recording at the beginning of each school year, this form does not include recording for edTPA purposes. In the event that some parents/guardians do not consent to have their child video recorded for edTPA, the teacher candidate must avoid capturing these non-consenting students on camera. However, it is essential that all students, regardless of whether their parents have consented to the recording or not, receive the same quality instruction during the learning segment.

In order to provide all students with the highest quality instruction, while respecting non-consenting parents' wishes, teacher preparation programs often recommend that teacher candidates temporarily reconfigure the classroom into two distinct sections. In one section of the room the teacher candidate would place all of the students whose parents have provided recording consent. Those students for whom the teacher candidate has not been given recording permission should be placed in a different section of the classroom, which is outside of the camera's view. Tripods can be useful tools when student teachers face a situation where not all of their students have gotten parental consent to be recorded. At Georgia

State University, the foreign language education university supervisor recommends that teacher candidates use a tripod to position the camera at an angle that focuses on the consenting students. This way, the teacher candidate can walk about the room freely, and engage with all students, without fear of inadvertently capturing a non-consenting student on camera. In the unlikely case that a non-consenting student accidentally appears on the video, teacher candidates may blur out the image of the student's face.

Minimizing Teacher Candidate Anxiety

As one might imagine, teacher candidates are typically most anxious during the recording phase of the edTPA. There are myriad reasons why the video recording process and submission can be anxiety-provoking for these novice teachers. Students in the Georgia State University foreign language education program who recently completed the edTPA reported the following concerns about the recording phase: technological issues, such as ensuring that the camera's batteries are sufficient to record the entire class period; students' sensitivity to the camera, for example, students misbehaving to *show off* for the camera; the pressure to produce the best quality video of excellent teaching which is highly engaging to students; and factors outside of their control which could come to bear on the video recording process, such as fire drills or other interruptions that could negatively affect the instruction. While these potential issues cannot be completely mitigated, university supervisors and cooperating teachers in the k-12 schools can collaboratively assist teacher candidates by offering suggestions based on past experiences working with video cameras in the classroom for edTPA.

Vital Considerations for Video Recording

There are several common technical issues that teacher candidates deal with when working on the edTPA. Although the edTPA website offers

some resources to teacher candidates vis-à-vis the recording process, teacher candidates may not examine them carefully enough. At Georgia State University, foreign language teacher candidates are provided copies of important edTPA documents, such as the edTPA handbook, evidence chart, and video formatting guidelines. To underscore the importance of avoiding technical pitfalls, the university supervisor reiterates the technical elements in seminar sessions as well as reminder emails and individual conferences with teacher candidates during the entire edTPA process.

One of the most important pieces of advice for world language teacher candidates regarding edTPA is to practice with the recording equipment in advance of the official recording. Recording a lesson that is not part of the learning segment for edTPA can be valuable for many reasons. First, the teacher candidate will have an opportunity to face any discomfort or fears he/she may have with recording their teaching. The high stakes nature of the edTPA assessment and its consequences on teacher certification can be debilitating to some teacher candidates who only want to do their best. If teacher candidates are able to desensitize themselves (and their students) to the camera during the practice recording, everyone should be more comfortable during the official recording. Likewise, pre-service teachers may feel more at ease after having the opportunity to work with the camera equipment to ensure that the recording quality is sufficient. Teacher candidates can breathe a sigh of relief once they rehearse with the equipment and the audio is clear and the video shows student engagement and growth.

Teacher candidates should review both the mock recording and the official one to ensure that they meet all edTPA video requirements. Teacher candidates should select two short clips which highlight the best moments of their teaching during the learning segment (which is comprised of three to five connected days of instruction united by a central focus). These clips must be no more than fifteen minutes total and the two files combined must be less than 300 MB total in size (Pearson, 2014). The relatively small file size requirements

mean that recording in standard definition instead of high definition is recommended, if possible. If the recording device does not offer the option to record in standard definition, teacher candidates can reduce the resolution to either “320 x 240” or “640 x 480” as another avenue to meet the file size requirement (Pearson, 2014). As a last resort, large video files can be compressed to condense the file in order to satisfy edTPA’s constraints but this is not preferable due to technical issues that could potentially occur during the compression process.

In addition to file size specifications, edTPA will only accept video files in the following formats: *flv*, *asf*, *qt*, *mov*, *mpg*, *mpeg*, *avi*, *wmv*, *mp4*, and *m4v* (Pearson, 2014). This stipulation is made to ensure that Pearson’s portfolio reviewers will be able to access the teacher candidates’ video files on their own computers. Consequently, at Georgia State University, the foreign language education university supervisor recommends that teacher candidates open their video files on at least three different computers to avoid a potential pitfall where the Pearson reviewer is unable to open a file on his/her computer. In the event that a Pearson reviewer is unable to open a file, the teacher candidate would receive a condition code for that task, and the entire portfolio would not be scored, requiring a resubmission.

Other technical considerations for video recording include various filming techniques. First, edTPA mandates that the clips be “continuous and unedited, with no interruption in events” (Stanford Center for Assessment, Learning, & Equity, 2015, p. 19). That is to say, the submitted video files cannot be spliced or edited in any way. No sophisticated features such as the title or even opening and closing credits are permitted (Stanford Center for Assessment, Learning, & Equity, 2015). For this reason, teacher candidates are advised to record each day of instruction during the learning segment so that they will have ample options to select from for the final video clips.

Secondly, the two video files must include interactions between the teacher candidate and students as well as student-to-student exchanges.

T

As a result, it can be advantageous for teacher candidates to use a tripod to record whole class interactions but enlist the assistance of a video operator (e.g. the cooperating teacher) to record small group work during the lesson. However, it is paramount that the video operator be well equipped to use the video recording device properly. Since the video clips must be continuous and unedited, the video operator must be cognizant of keeping the video running continuously. While it may be tempting for the video operator to start and stop the recording while moving throughout the classroom, this will cause problems when it comes to selecting a continuous clip.

Likewise, video operators must be mindful of placing a video camera's microphone close enough to capture small group interaction clearly without also recording background noise that could interfere with audio quality. On the world language edTPA, teacher candidates must submit video clips that showcase their students using the language interpersonally, often in small group or paired activities. Strategic placement of the video recording device during these small group exercises is critical to ensure that the focal students can be heard clearly using the target language. If the video clips capture too much background noise, or if the videorecorder is placed too far away from the focal students, it will be nearly impossible to distinguish their speech. When teacher candidates have the opportunity to practice with the video recording device in advance, they then have a better idea of how to strategically place the device and move it around to capture optimal sound quality during the recording phase of the official submission.

Peer Review to Avoid Condition Codes

It is important for university supervisors and k-12 cooperating teachers to assist teacher candidates, particularly with the video recording process, to avoid a costly mistake such as a condition code

which would render the entire portfolio unable to be scored, requiring resubmission. Therefore, at Georgia State University, a peer review assignment is a mandatory component before final submission. During the peer review phase, teacher candidates open their assigned classmate's files and check that all files open properly and comply with edTPA's requirements. While the peer reviewers are not able to offer suggestions on the content of their classmate's portfolio, they are allowed to check for technical issues, such as a corrupt file or a file that is formatted incorrectly. Moreover, reviewing their classmate's portfolio may cause them to notice that a document is missing or formatted incorrectly in their own portfolio. The peer review process happens during the last weekend before official submission, so that each teacher candidate has a few days to review their peer's feedback and make any last-minute changes to improve their own portfolio.

All of these video recording strategies are crucial components for teacher candidates' success on the world language edTPA. The stakes are so high for edTPA that receiving a condition code for even one video means the teacher candidate will be required to resubmit one or more of the tasks at an additional cost. Beyond the financial toll, the resubmission process itself can be time consuming, depending upon how many tasks need to be redone. Official scoring reports often take three to four weeks to receive, which could potentially create timing issues for program completion and state certification. As a result, Georgia State University world language teacher candidates are required to submit their official edTPA portfolio by week ten to account for the extra time that could be needed to resubmit in order to receive a passing score by the end of the semester. Without achieving the state determined qualifying score on the edTPA, teacher candidates cannot be recommended for teacher certification, and therefore, may be held back from becoming employed by a school district.

FUTURE RESEARCH DIRECTIONS

edTPA is a relatively new teacher performance assessment on the US educational landscape and there is much to learn about it. At this moment, there is a dearth of research on the World Language edTPA (Hildebrandt & Swanson, 2014, 2016). It would be informative to learn more about how world language teacher education programs have been restructured since edTPA was approved for licensure decisions in so many states. Additionally, more empirical evidence is needed for multiple reasons ranging from programmatic improvement decisions to helping teacher candidates prepare quality dossiers. It would also be helpful to know more about how teacher education programs provide remediation to teacher candidates who are initially unable to pass one or more sections of the edTPA. At the moment, only quantitative research methods have been used to examine teacher candidate performance on the World Language edTPA. Qualitative and mixed methods research designs would provide important insight into this new high-stakes and costly assessment at a time when testing costs to become a teacher are approximately \$1000 in Georgia (Hildebrandt & Swanson, 2014).

CONCLUSION

Since the mid-1960s, the ESEA legislation has been an integral part of US curriculum and instruction. As noted earlier, the law focused on not only equal access to education but also setting high standards for academic performance while demanding accountability from schools and school districts. Every president since then has worked to reauthorize the legislation and arguably the most controversial reauthorization is known as *No Child Left Behind* (NCLB, U.S. Department of Education, 2002). This mandate sought more rigorous testing and accountability of k-12 student learning and included foreign language as part of

the core curriculum. Years later the Obama administration's initiative, *Race to the Top*, further scrutinized the educational process and mandated that states begin measure beginning and veteran teacher effectiveness to receive full federal funding (U.S. Department of Education, 2009).

Now, edTPA has been rolled out in more than half of the states and is a cause for concern, especially given that pre-service world language teacher candidates in many states have to successfully pass edTPA in order to receive teacher licensure (Hildebrandt & Swanson, 2016). Thus, we strongly recommend that teacher preparation faculty members take the time to work with teacher candidates specifically on the aspects of video recording in the classroom. It is important to first gain parental permission to film adolescents in schools and it is equally important for teacher candidates to rehearse such aspects as camera and tripod placement, framing angles to show students, lighting, audio quality so that students can be heard well, all of which will help reduce teacher candidate anxiety and improve overall video quality.

In sum, edTPA is a high-stakes teacher candidate assessment and care must be taken when developing portfolios for external review. Research shows that teacher candidates have been found to score lower on the assessment task of the portfolio (Hildebrandt & Swanson, 2014), and the integrated performance assessment appears to have merit when teacher candidates develop their edTPA dossiers.

REFERENCES

- American Association for Colleges of Teacher Education. (2016). *Participation map*. Available from, <http://edtpa.aacte.org/state-policy>
- Baeher, L. H., McCormack, B., & Kung, S. (2014). Supervisor use of video as a tool in teacher reflection. *The Electronic Journal for English as a Second Language*, 18(3), 1–17.

- Calandra, B., Brantley-Dias, L., Lee, J. K., & Fox, D. L. (2009). Using video editing to cultivate novice teachers practice. *Journal of Research on Technology in Education*, 42(1), 73–94. doi:10.1080/15391523.2009.10782542
- Cochran-Smith, M. (2000). The questions that drive reform. *Journal of Teacher Education*, 51(5), 331–333. doi:10.1177/0022487100051005001
- Danielowich, R. M., & McCarthy, M. J. (2013). Teacher educators as learners: How supervisors shape their pedagogy by creating and using classroom videos with their student teachers. *Action in Teacher Education*, 35(3), 147–164. doi:10.1080/01626620.2013.806231
- Darling-Hammond, L. (2012). The right start: Creating a strong foundation for the teaching career. *Phi Delta Kappan*, 94(3), 8–13. doi:10.1177/003172171209400303
- Elementary and Secondary Education Act of 1965, Pub. L. No. 89-10, 79 Stat. 27.
- Georgia Professional Standards Commission. (2014). *Understanding the 2014 educator certification rule*. Available from, http://www.gapsc.com/Commission/policies_guidelines/Downloads/2014EducatorCertificationRuleChanges.pdf
- Georgia Professional Standards Commission. (2015). *edTPA standard setting decision*. Available from, http://www.gapsc.com/Downloads/edTPA_Standard_Setting_Decision_071015.pdf
- Goe, L., Bell, C., & Little, O. (2008). *Approaches to evaluating teacher effectiveness: A research synthesis*. Washington, DC: National Comprehensive Center for Teacher Quality.
- Hildebrandt, S., & Swanson, P. (2014). World language teacher candidate performance on edTPA: An exploratory study. *Foreign Language Annals*, 47(4), 576–591. doi:10.1111/flan.12117
- Hildebrandt, S., & Swanson, P. (2016). *Understanding the world language edTPA: Research-based policy and practice*. Charlotte, NC: Information Age Publishers.
- Illinois State Board of Education. (2012). Memo. Available from, <http://www.isbe.state.il.us/licensure/pdf/higher-ed/edTPA/edTPA-summary1212.pdf>
- Muijs, D. (2006). Measuring teacher effectiveness: Some methodological reflections. *Educational Research and Evaluation: An International Journal on Theory and Practice*, 12(1), 53–74. doi:10.1080/13803610500392236
- Pearson. (2014). *Recommended video format and settings*. Available from, <https://www.edtpa.com/.../RecommendedVideoFormatsandSettings.pdf>
- Renaissance Partnership for Improving Teacher Quality. (2002). *Teacher work sample: Performance prompt, teaching process standards, scoring rubrics*. Bowling Green, KY: Author.
- Rosenbusch, M. H. (2005). The no child left behind act and teaching and learning languages in the U.S. Schools. *Modern Language Journal*, 89, 250–261.
- Rosenbusch, M. H., & Jensen, J. (2004). Status of foreign language programs in NECTFL states. *NECTFL Review*, 56, 26–37.
- Sato, M. (2014). What is the underlying conception of teacher of the edTPA? *Journal of Teacher Education*, 1–14. Available from <http://jte.sagepub.com/content/early/2014/07/09/0022487114542518>
- Stanford Center for Assessment, Learning, and Equity. (2013). *2013 edTPA field test: Summary report*. Stanford, CA: Stanford University. Available from, <http://edtpa.aacte.org/news-area/announcements/edtpa-summary-report-is-now-available.html>
- Stanford Center for Assessment, Learning, and Equity. (2015). *edTPA world language assessment handbook*. Stanford, CA: Stanford University.
- Tripp, T. R., & Rich, P. J. (2012). The influence of video analysis on the process of teacher change. *Teaching and Teacher Education*, 28(5), 728–739. doi:10.1016/j.tate.2012.01.011

U.S. Department of Education. (2002). No Child Left Behind Act of 2001, Pub. L. No. 107-110, § 115. *Stat, 1425*. Available from <http://www2.ed.gov/policy/elsec/leg/esea02/107-110.pdf>

U.S. Department of Education. (2004, March). *New no child left behind flexibility: Highly qualified teachers*. Washington, DC: U.S. Department of Education. Available from, <http://www.ed.gov/nclb/methods/teachers/hqtflexibility.html>

U.S. Department of Education. (2009). *Race to the top executive summary*. Available from, <http://www2.ed.gov/programs/racetothetop/executive-summary.pdf>

U.S. Department of Education. (2014). *Improving teacher preparation: Building on innovation*. Available from, <http://www2.ed.gov/documents/teaching/teacher-prep.pdf>

ADDITIONAL READING

Adair-Hauck, B., Glisan, E. W., Koda, K., Swender, E., & Sandroock, P. (2006). The integrated performance assessment (IPA): Connecting assessment to instruction and learning. *Foreign Language Annals*, 39(3), 359–382. doi:10.1111/j.1944-9720.2006.tb02894.x

Center for Advanced Research on Language Acquisition. (2013). *Creating an assessment unit*. Available from, http://www.carla.umn.edu/assessment/VAC/CreateUnit/p_1.html

Hildebrandt, S., & Swanson, P. (2014). World language teacher candidate performance on edTPA: An exploratory study. *Foreign Language Annals*, 47(4), 576–591. doi:10.1111/flan.12117

Hildebrandt, S., & Swanson, P. (2016). *Understanding the world language edTPA: Research-based policy and practice*. Charlotte, NC: Information Age Publishers.

The National Standards Collaborative Board. (2015). *World-readiness standards for learning languages* (4th ed.). Alexandria, VA: Author.

T

KEY TERMS AND DEFINITIONS

Central Focus: As defined by SCALE, the important understandings and core concepts that students will develop during the teaching of the learning segment.

Communicative Language Teaching (CLT): An approach to language instruction that prioritizes target language interaction as both the overarching goal as well as an avenue for accomplishing that goal. This approach focuses primarily on the development of students' communicative competence in a meaningful cultural context.

edTPA: The first nationally available subject-specific performance-based assessment adopted by many U.S. states as a means to measure teacher candidates' developing skills and knowledge.

Integrated Performance Assessment (IPA): A standards-based and performance-based assessment which incorporates three tasks, aligned with a common central focus, that focuses on each of the following modes of communication: interpersonal, interpretive, and presentational.

Learning Segment: As defined by SCALE, a series of three to five days of consecutive instruction centered on a common central focus, marked by identifiable beginning and ending points.

Modes of Communication: Based on ACTFL's stance that communication occurs in three modes: interpersonal, interpretive, and presentational.

Pre-Service Teachers: Also known as teacher candidates, this term is used to describe student teachers who are enrolled in a teacher preparation program and working toward teacher certification. They complete supervised field-based teaching experiences with the support and mentorship of university faculty and K-12 cooperating teachers.

Category T

Theoretical Computer Science

Efficient Optimization Using Metaheuristics

T

Sergio Nesmachnow

Universidad de la República, Uruguay

INTRODUCTION

Nowadays, many real-world problems are intrinsically complex. Optimization is a process that deals with hard problems arising in many scientific, academic, and commercial applications. Many optimization problems are difficult to solve, because of having large solution spaces, handling complex functions and hard constraints, or managing large volumes of data. Traditional computing methods are not useful in practice for solving these hard problems, because they demand very large execution times when solving realistic instances.

Metaheuristics are soft computing methods that allow computing accurate solutions for hard-to-solve problems in reasonable execution times. Metaheuristics are conceived under simple paradigms to take advantage of tolerating imprecision, uncertainty, and approximation in the problem data. They can handle partial information and non-exact solutions in order to compute accurate solutions for optimization problems. These features make metaheuristics highly valuable techniques, as they allow meeting realistic resolution times in many application areas, from informatics to industrial and commercial.

Metaheuristics were proposed almost thirty years ago, and nowadays this research area is consolidated. Many metaheuristic proposals have been formulated, regarding both theoretical models and applications. This chapter provides an insight into the main concepts, advances, and results in the field of metaheuristics for efficiently solving optimization problems. A general view of most well-known metaheuristics is presented and the main applications of metaheuristics in nowadays real-world problems from several domains is described. Finally, the main current and future research lines in the field are also summarized.

DOI: 10.4018/978-1-5225-2255-3.ch669

BACKGROUND

Metaheuristics are high-level soft computing strategies that define algorithmic frameworks and techniques to find approximate solutions for optimization problems (Blum & Roli, 2003; Talbi, 2009). Many efficient and accurate metaheuristics have been proposed, which can be applied to solve a variety of optimization problems underlying many applications in science/technology, industry, and commerce.

Metaheuristics Concepts: Definition and Classification

Metaheuristics were originally proposed as high-level problem-independent strategies to coordinate several heuristic search methods, which can be instantiated to solve hard problems. This definition has broadened to include a wide range of search and learning processes (including shaking, construction/deconstruction, adaptation, swarming and collective behavior, hybridization, etc.) applied to improve the search.

The heuristic components in a metaheuristic are conceived to be applied in an intelligent way, providing accurate and balanced methods for diversification and intensification. These two concepts are very important to guarantee the efficacy of the search: *diversification* refers to achieving a good exploration pattern for the search space, providing a reasonable coverage and avoiding stagnation in local optima; *intensification* means exploiting or improving already found accurate solutions to increase their quality.

Regarding the number of solutions handled, metaheuristics are classified in two classes: *trajectory-based* and *population-based*. Trajec-

tory-based metaheuristics work with a single solution, which is replaced by another (often the best) solution found in its neighborhood. The search is characterized by a trajectory in the space of solutions. Trajectory metaheuristics allow quickly exploiting solutions, thus they are referred as intensification-oriented methods. On the other hand, population-based metaheuristics work with a set of candidate solutions, which are modified and/or combined following some common guidelines. Some solutions in the population are replaced by newly generated solutions (often the best). These methods are characterized as diversification-oriented, because having multiple solutions allows significantly increasing the exploration capabilities.

Other classification criteria for metaheuristics are related to specific features of the search strategy, for example: memory vs. memory-less, or dynamic vs. static objective function.

Metaheuristic Techniques

Trajectory-based metaheuristics include many well-known techniques from the early years of this area:

- **Simulated Annealing (SA):** Probabilistically allows moving to a solution with worst objective function value, trying to escape from local optima (Kirkpatrick et al., 1983). SA is inspired on the annealing process of metals and it was the first metaheuristic proposed, although the term “metaheuristic” was not used in those years.
- **Tabu Search (TS):** Enhances a local search strategy by using memory to store information about visited solutions. To promote diversification, returning to recently visited solutions is not allowed. TS was introduced by Glover (1986), who first used the term “metaheuristic”.
- **Greedy Randomized Adaptive Search Procedure (Feo and Resende, 1995):** A metaheuristic which greedily selects com-

ponents to construct a solution, and then applies a local search to improve it.

- **Variable Neighborhood Search (Mladenovic & Hansen, 1997):** Local search techniques based on using different neighborhood structures during the search.
- **Iterated Local Search (Lourenço et al., 2002):** Uses a hill-climbing to find local optima and a stochastic perturbation/restart strategy to prevent the search getting stuck in local optima.

Population-based metaheuristics include techniques that use cooperation in the search:

- **Evolutionary Computation (EC):** Emulates the evolution of species in nature by applying stochastic operators (recombination and random changes) on a population. A selection-of-the-best strategy guides the search to high-quality solutions. EC includes Genetic Programming (Koza, 1992), Evolution Strategies, and Evolutionary Algorithms (EAs) (Goldberg, 1989).
- **Swarm Intelligence (SI):** Uses agents to perform explorations while they interact with neighbors and the environment, based on the collective behavior of self-organized systems. Agents have limited search capabilities, but an “intelligent” pattern emerges in large swarms. SI include Ant Colony Optimization (ACO), emulating the behavior of ants foraging for food (Dorigo, 1992), Particle Swarm Optimization (PSO), simulating flocks of birds (Kennedy & Eberhart, 1995), Artificial Immune System (DasGupta, 1998), and the recently proposed Bacterial Foraging, Fish/Glowworm Swarm, Firefly Algorithm, Cuckoo Search, and Bee Algorithms.
- **Evolutionary-Inspired Metaheuristics:** Apply an evolutionary search, with different features. Estimation of Distribution Algorithms learn by building/sampling probabilistic models of promising solu-

tions (Larrañaga & Lozano, 2002). Scatter Search/Path Relinking applies a specific recombination and ideas from TS (Glover, 1999). Differential Evolution is specialized for optimizing real functions (Storn & Price, 1997). Harmony Search (HS) emulates the improvisation of musicians (Geem et al., 2001). Memetic Algorithms includes a local search phase after the evolutionary operators (Moscato, 1999).

Hybrid models are used to improve the search capabilities of metaheuristics. Strong hybrid metaheuristics are constructed by including problem-oriented representations and operators. Weak hybrids are built by combining several methods to improve the search. An efficient cooperation is achieved by exchanging a small set of solutions or statistical values (Talbi, 2002).

Parallel metaheuristics allow enhancing and speeding up the search by using several computing elements. They reach high quality results in short execution times, even for hard-to-solve optimization problems.

The main models for parallel trajectory metaheuristics are:

- **Parallel Multi-Start:** Launches several instances, which may be independent/co-operative, start from the same/different solution(s), use the same/different parameters, etc.
- **Parallel Moves:** Distributes the current solution to several processes, each one explores by applying a search based on its candidate solution and the results are returned to the main process.
- **Move Acceleration:** Evaluates several moves in a parallel centralized way, by parallelizing the objective function using an aggregation approach.

The main models for parallel population metaheuristics considers the population organization:

- **Master-Slave:** Applies a hierarchical functional decomposition: a master process performs the evolutionary search, and controls a group of slave processes that evaluate the objective function, which often requires larger computing time than the search operators.
- **Distributed Subpopulations:** Splits the population in subpopulations, each one running a sequential metaheuristic. The solutions only can interact within their subpopulation. A migration operator exchanges some selected solutions among subpopulations, introducing a new source of diversity.
- **Cellular:** Uses an underlying spatial structure for the population. Interactions are restricted to neighbor solutions, and characteristics gradually propagate through the grid. This feature is useful to provide diversity in the population, often improving the search.

The research on parallel metaheuristics is very important nowadays, when large instances of real-world problems are solved (Alba et al., 2013).

STATE OF THE ART: CURRENT RESEARCH LINES

Issues

The main research issues are related to methodology, algorithms, and applications.

Methodology

Methodology is well regarded and needed for a comprehensive view of the field. Providing mathematical foundations and theoretical analysis for metaheuristics, such as the convergence for SA (Laarhoven & Aarts, 1987), EC (Goldberg, 1989), PSO (Clerc & Kennedy, 2002), and ACO (Gutjahr, 2010) is important. Theoretical convergence is

desired, but because it is often only guaranteed in inefficient times, analysis based on computing hitting-time or takeover-time are relevant to determine the applicability of metaheuristics as problem solvers. Studying the dynamic behavior of metaheuristics and landscape analysis for specific problems are also relevant.

Parameter tuning of metaheuristics is also important. It significantly impacts on both efficiency and accuracy, and even can prevent metaheuristics to be accepted into real-world software. Specific methodologies have been proposed for fine-tuning metaheuristics, trying to move this task from art (i.e., usually following rules-of-thumb) to science (i.e., following a sound approach) (Birattari, 2009). This issue is related to the more general one about robustness and sensitivity of metaheuristics.

Developing methodologies for the statistical analysis of results and designing benchmark problems for evaluating and comparing both existing algorithms and new proposals is also very important (Bartz-Beielstein, 2010).

Algorithms: Design and Implementation

Traditional metaheuristics, proposed more than twenty years ago, are firmly established as problem solvers. Their main algorithmic features have been studied, but several design alternatives have led to not-so-well-explored variants, for which their actual contribution is yet to be demonstrated. Most classic metaheuristic are implemented in the software frameworks available today.

Many recently proposed metaheuristics are not consolidated yet. Among them, we can mention Harmony Search, Artificial Bee, Cuckoo Search, Firefly Algorithm, etc., which take inspiration in different natural/physical processes for problem solving. The application of these novel methods to different hard-to-solve optimization problems is needed to evaluate the contribution of each proposal. The emergence of many “new” metaheuristics is not free from controversy.

A whole line of research is open concerning the design and implementation of metaheuristic

techniques to solve specific problem classes, including multiobjective optimization, stochastic/uncertain optimization, constraint handling, very large scale optimization, and others.

Applications

Although originally proposed to solve combinatorial optimization problems, nowadays metaheuristics are useful tools to solve real-life problems arising in diverse application domains.

Many works have focused on solving academic optimization problems related to resource assignment, routing, set covering/partitioning, etc. in several variants: discrete combinatorial problems, continuous, multi-objective, and dynamic problems.

Real-world problems from science, industry and commerce, have also been solved using metaheuristics. In *automation*, metaheuristics are applied as accurate/efficient learning techniques helping to reduce the need for human presence. In *bioinformatics*, metaheuristics are efficient techniques to deal with computationally-expensive problems that manage very large amount of data in this emergent scientific field (sequence alignment, drug design, genome assembly, protein structure). In *engineering*, metaheuristics are useful for designing systems with many components, complex functions, and large design spaces (aerodynamic optimization, signal and image processing, machinery design, electronics, hydraulic networks). In the area of *information processing*, the fast exploration of metaheuristics significantly help with the challenge of dealing with huge volumes of data (feature selection, classification/clustering, data mining, natural language processing). In *manufacturing and industry*, metaheuristics allow increasing competitiveness in nowadays globalized economy (warehouse location, packing, assembly line problems). In *planning and scheduling*, metaheuristics have a major impact on the productivity and quality of service by accurately deciding how to commit resources to a group of tasks in very reduced execution times. Metaheuristic

tics are also useful optimization methods in *routing and logistics*, to manage the flow of resources for large complex scenarios. In *telecommunications*, metaheuristics have great impact on addressing the main challenges from classic wired to modern wireless cellular/vehicular/sensor network infrastructures, including dealing with large-sized scenarios, need for real-time results, etc.

Metaheuristics have also been successfully applied in energy optimization, medicine, military applications, economy and finance, and many other areas, frequently using multiobjective, hybrid, and parallel metaheuristics. These facts show the growing maturity of the research in metaheuristics, and allow concluding that many real-life problems will be solved using metaheuristics in the near future (Nesmachnow, 2014).

Controversies

The metaheuristics field has developed fast, but some issues are controversial, including unconvincing methodologies used for the experimental evaluation and the emergence of many “new” methods.

Regarding the experimental evaluation, it has been noticed that many published articles do not provide a proper experimental analysis, e.g., the authors only confront the results against simple ad-hoc heuristics, or they even do not provide a comparison at all. Often the comparison of metaheuristic is not conducted by applying standard metrics and a proper statistical methodology. In some cases, researchers apply metaheuristics to very small problem instances or even to not hard-to-solve optimization problems. Maybe the lack of an accepted generic framework for performance assessment is behind these shortcomings when studying metaheuristics on real-world problems. However, general guidelines for statistical evaluation and analysis exist, thus researchers have no excuses to present poorly studied results.

Several “new” metaheuristics have been proposed, inspired on many natural/physical phenomena (e.g., Glowworm Swarm Optimization (GSO),

Gravitational Search, Intelligent Water Drops (IWD), Bat Algorithm, etc.). Some researchers have a critical view of these new techniques, mostly based on the improper use of metaphors or phenomena emulation to justify the proposals (Sörensen, 2015). Critics argue that many new metaheuristics are only reformulating previous methods, introducing small variants on classic metaheuristics. In some cases, the new variants are not properly evaluated, making difficult to determine the real contribution to the field. For example, while SI methods are based on emergent intelligence, the new GSO shares several features with ACO and PSO, and it is also strongly related to Firefly Algorithm, Bat Algorithm, and Migrating Birds Optimization. Moreover, it has been argued that techniques like IWD are based on doubtful premises, with few links with reality.

The debate about the HS metaheuristic is a clear example of how some researchers are skeptical about new metaheuristics proposals: Weiland (2010) suggested that HS is a special case of EA; Geem (2010) attempted to refute the critic, and Padbreg (2011) proved that HS is equivalent to an EA with a special crossover operator. A plethora of “new” methods have been recently proposed, including bio-inspired (Generalized Response Surface, Open Source Development Model, Biogeography Optimization, Shuffled Frog Leaping, Invasive Weed Colony, etc.) (Yang, 2008); and based on physics/mathematics (Spiral Optimization, Big Bang–Big Crunch, Gravitational Search, Hysteretic Optimization, Immune Gravitation, Artificial Physics Optimization, Charged System Search, Galaxy Search, etc.) (Biswas et al., 2013). Quantum-computing implementations of metaheuristics are popular too. Some of these “new” algorithms are too similar to each other, and they were published in conferences/journals outside the area. This phenomenon is not new: in the 1990s, several metaheuristics were proposed (Search Space Smoothing, Threshold Accepting, Great Deluge, Demons Algorithm), but they were black-box techniques based on local search/SA, so they did not gain popularity. Anyway, as some

new proposals do not have solid foundations and have not been properly validated, researchers and practitioners should take care when proposing or applying a “new” method to solve specific problems.

Problems

The lack of a general methodological framework for evaluating metaheuristics is a key problem in the field. Although statistical analysis is recognized as a must for metaheuristics evaluation, a well-conceived scientific approach is needed to avoid “craft” procedures for parameter setting and experimental analysis.

Regarding implementations, an important challenge nowadays is designing generic frameworks and software libraries for developing and using metaheuristic algorithms. Here, several crucial aspects should be taken into account, including programming decisions, modularity, reusability, flexibility, and the trade-off between usability and computational efficiency of each implementation. Nowadays, many libraries and frameworks exist, including some comprehensive ones, such as ECJ, EO/ParadiseEO, HeuristicLab, JDeal, MALLBA, and Opt4j (see a full list on Luke (2013)). Nevertheless, more effort is required to conceive efficient and easy-to-use implementation guidelines for generic metaheuristic frameworks.

SOLUTIONS AND RECOMMENDATIONS

When searching for a useful optimization method to solve a problem, a wise idea is to apply a systematic bottom-up approach: starting from a simple solver (i.e., using trajectory-based strategies) and then move to more complex metaheuristics. This approach usually leads to a better understanding of the problem and the capabilities of each resolution technique. When a specific method includes many search operators or paradigms, the contribution

of each component must be studied carefully and systematically.

Another important recommendation to determine the contribution of metaheuristics as optimization techniques is to develop a formal background for studying theoretical properties (i.e. convergence) and metrics. The usefulness of specific ideas, components and/or search operators should be carefully evaluated, for all (especially new) metaheuristics. A formal, component-wise experimentation is crucial to provide a clear understanding of the algorithms.

Recently proposed metaheuristics, based on unconventional search paradigms, should be applied when providing real arguments about how useful are their searching strategies for the class of problem targeted.

Applying a careful methodology for obtaining and analyzing the results in the experimental evaluation of metaheuristics is mandatory (Rardin & Uzsoy, 2001). Important aspects related to metaheuristics evaluation are experiments design, finding appropriate test instances, evaluating numerical efficacy and time efficiency, and clearly presenting and analyzing the results. In any publication, a comprehensive number of problem instances should be solved. Instances must be representative of realistic properties and dimension of real-world problems: analysis performed over few and/or small instances are not useful at all, since they do not provide evidence about the applicability and efficiency of the metaheuristic method. Small toy problems, for which even a random search could compute good solutions, must be avoided. When specific benchmarks exist, they must be used, and the metaheuristics results must be compared against those from the best known methods for solving the problem.

Parameter evaluation tests must be performed on a separated set of problem instances, in order to avoid bias in the experiments (Coy et al., 2001). At least 30 (recommended 100) independent executions must be performed for parameter setting and also for validation experiments. Both numerical

results and execution time should be reported and statistically assessed (McGeoch, 2008). Scalability analyses regarding the problem dimension are recommended.

Statistical methods must be employed to analyze the results. A standard methodology for the statistical analysis of metaheuristics is to apply a normality test (e.g., Kolmogorov-Smirnov/Shapiro-Wilk) to check if the results follow a normal distribution; if so, a Student/ANOVA test must be applied next; otherwise, a non-parametric test (e.g., Kruskal-Wallis/Wilcoxon/Friedman) must be performed. This two-step procedure allows controlling the error of incorrectly rejecting the null hypothesis when it is true, since two independent phases (testing for different null hypotheses), are used. For metaheuristics, usually the result distributions are not normal, so non-parametric tests and median analysis are recommended. A comparative analysis of metaheuristics must be performed using well-known statistical tests and metrics, and considering a set of realistic problem instances.

Regarding implementation, some recommendations oriented to develop efficient and easy-to-use libraries include making the most of modularity and reusability. This strategy allows the user to develop easy implementations, not starting from scratch, and focusing on the specific contributions of his work. C/C++ and Java are the most used languages to develop metaheuristics (Nesmachnow et al., 2015). Object-oriented features of high-level languages are useful, but some of them (inheritance, polymorphism, overloading) usually conspire against efficiency and they should be used carefully. At the end, researchers should not forget that the ultimate goal of any computing technique is to provide a real solution for the problem that should be user-friendly, robust, and able to cope with day-to-day real-world operation.

Nowadays, practitioners and researchers have many resources to address the main concepts about metaheuristics. Excellent books published about the subject are: “Essays and Surveys on Metaheuristics” (Hansen and Ribeiro, 2001),

“Handbook of Metaheuristics” (Glover and Kochenberger, 2003), “Metaheuristics: Progress as Real Problem Solver” (Ibaraki et al., 2005), “Nature-Inspired Metaheuristic Algorithms” (Yang, 2008), “Metaheuristics: From Design to Implementation” (Talbi, 2009), “Handbook of Metaheuristics” (Gendreau and Potvin, 2010), “Essentials of Metaheuristics” (Luke, 2013).

A good practice to evaluate the research on metaheuristics is submitting articles to conferences and journals in the field. Nowadays, several conferences focus in metaheuristics. The most respected ones are Metaheuristics International Conference (11 editions, 1996–2015), Parallel Problem Solving from Nature (12 editions, 1990–2012), and EU/ME Workshop (15 editions, 2000–2015). Younger conferences on metaheuristics are International Conference on Hybrid Metaheuristics (8 editions, 2004–2013), and International Conference on Metaheuristics and Nature Inspired Computing (5 editions, 2006–2014). Specific streams about metaheuristics exist on Operation Research societies like ALIO, EURO, INFORMS, and OR Society. Conferences focused on EC include Genetic and Evolutionary Computation Conference (24 editions, 1985–2015); IEEE Congress on Evolutionary Computation (17 editions, 1999–2015); EvoStar (18 editions, 1998–2015); and Evolutionary Multi-Criterion Optimization (8 editions, 2001–2015). Other relevant conferences are International Conference on Unconventional Computation and Natural Computation (14 editions, 1998–2015); Asia-Pacific Symposium of Intelligent and Evolutionary Systems (19 editions, 1997–2015); International Workshop on Nature Inspired Cooperative Strategies for Optimization (7 editions, 2006–2015); IEEE International Conference on Intelligent Systems (7 editions, 2002–2015); EVOLVE (5 editions, 2011–2015), and International Conference on Theory and Practice of Natural Computing (4 editions, 2012–2015). Conferences on SI and other metaheuristic methods include ANTS International Conference on Swarm Intelligence (9 editions, 1998–2014); International Confer-

ence on Artificial Immune Systems (12 editions, 2002–2013); Learning and Intelligent Optimization (9 editions, 2005–2015); and International Conference on Swarm Intelligence (6 editions, 2010–2015).

Two novel journals focusing on metaheuristics started publishing in 2010: *International Journal of Metaheuristics* and *International Journal of Applied Metaheuristic Computing*. Both journals are still consolidating, thus most articles on metaheuristics are often published on soft computing and related journals: *Soft Computing*, *Applied Soft Computing*, or *IEEE Transactions on Evolutionary Computation*. Other related journals are *Swarm Intelligence Journal*, *International Journal of Bio-Inspired Computation*, *Swarm and Evolutionary Computation*, and *International Journal of Swarm Intelligence*.

FUTURE RESEARCH DIRECTIONS

Several relevant issues about metaheuristics are still in a consolidation stage, while many others emerge as promising lines for actual and future work.

Regarding algorithms, there is still room to contribute in establishing a sound foundation for the search patterns proposed by each metaheuristic, especially for the novel methods proposed in the last years. The importance of the intensification and diversification strategies to efficiently compute accurate results was demonstrated in earlier theoretical works, and now those concepts provide a hint for developing a unified view for metaheuristics. In fact, some researchers are working on this topic. Using theoretical results to design more powerful search strategies is also a promising line of work too.

The design and formalization of frameworks for metaheuristics, including algorithms, metrics, methodologies for the statistical analysis, and

comprehensive benchmarks for different problem classes, is also an important line of research. Formal methodologies should be adopted to design and implement generic software libraries for metaheuristics.

The hybridization of well-known metaheuristics with other search techniques, including heuristics and exact methods, is also a promising field of study. Several important issues emerge in this line of work, including if (and when) it is worth to sacrifice the general capabilities of metaheuristics in order to obtain efficiency or accuracy for solving a given class of problems. Determining how useful is introducing problem knowledge and how to incorporate it in a systematic way is very important too.

Another promising line of work is related to parallel models for metaheuristics. Researchers in this field are trying to conceptualize a unified view of parallel metaheuristics, taking into account the points of view of theory and technology. They are also working on including relevant issues such as using new mobile devices and large computing infrastructures (accelerators, many-core devices, graphic processors, grid and cloud computing).

All the previous features are also valuable regarding the application of metaheuristics on different application domains. In real life, concepts such as hybridization, adaptation, comprehensible implementations, and features such as flexibility, robustness, easy-to-use, and efficiency for interactive/online problems can contribute to make metaheuristics the preferred choice as useful optimization methods.

Finally, a wide field of opportunity is still open concerning the application of metaheuristics to different hard-to-solve optimization problems arising in new technological, industrial, and commercial areas. Multiobjective optimization, dynamic problems, stochastic and uncertainty optimization, interactive or online problems and applications, are among the main lines of research in the field.

CONCLUSION

This chapter provided an insight of the field of metaheuristic techniques and their characterization as methods to solve optimization, search, and learning problems.

Nowadays, metaheuristics have established as useful tools in practice to cope with the complexity of many real-world problems, and this chapter summarized the main ideas, concepts, problems and solutions, and future research direction in the field, in a descriptive way which should be useful to researchers, practitioners, and also newcomers to the area.

REFERENCES

- Alba, E., Luque, G., & Nesmachnow, S. (2013). Parallel Metaheuristics: Recent Advances and New Trends. *International Transactions in Operational Research*, 20(1), 1–48. doi:10.1111/j.1475-3995.2012.00862.x
- Bartz-Beielstein, T., Chiarandini, M., Paquete, L., & Preuss, M. (2010). *Experimental Methods for the Analysis of Optimization Algorithms*. Springer-Verlag. doi:10.1007/978-3-642-02538-9
- Birattari, M. (2009). *Tuning Metaheuristics: A Machine Learning Perspective*. Springer. doi:10.1007/978-3-642-00483-4
- Biswas, A., Mishra, K., Tiwari, S. & Misra, A. (2013). Physics-Inspired Optimization Algorithms: A Survey. *Journal of Optimization*.
- Blum, C., & Roli, A. (2003). Metaheuristics in combinatorial optimization: Overview and conceptual comparison. *ACM Computing Surveys*, 35(3), 268–308. doi:10.1145/937503.937505
- Clerc, M., & Kennedy, J. (2002). The particle swarm—explosion, stability, and convergence in a multidimensional complex space. *IEEE Transactions on Evolutionary Computation*, 6(1), 58–73. doi:10.1109/4235.985692
- Coy, S., Golden, B., Runger, G., & Wasil, E. (2001). Using experimental design to find effective parameter settings for heuristics. *Journal of Heuristics*, 7(1), 77–97. doi:10.1023/A:1026569813391
- DasGupta, D. (1998). *Artificial Immune Systems and their Applications*. Springer-Verlag.
- Dorigo, M. (1992). *Optimization, Learning and Natural Algorithms* (Unpublished doctoral dissertation). Politecnico di Milano, Italy.
- Feo, T., & Resende, M. (1995). Greedy randomized adaptive search procedures. *Journal of Global Optimization*, 6(2), 109–133. doi:10.1007/BF01096763
- Geem, Z. (2010). Research Commentary: Survival of the Fittest Algorithm or the Novelest Algorithm? *International Journal of Applied Metaheuristic Computing*, 1(4), 75–79. doi:10.4018/jamc.2010100105
- Geem, Z., Kim, J., & Loganathan, G. (2001). A New Heuristic Optimization Algorithm: Harmony Search. *Simulation*, 76(2), 60–68. doi:10.1177/003754970107600201
- Glover, F. (1986). Future Paths for Integer Programming and Links to Artificial Intelligence. *Computers & Operations Research*, 13(5), 533–549. doi:10.1016/0305-0548(86)90048-1
- Glover, F. (1999). Scatter search and path relinking. In Glover, Corne & Dorigo (Ed.), *New Ideas in Optimization*. McGraw-Hill.
- Glover, F., Corne, D., & Dorigo, M. (Eds.). (1999). *New Ideas in Optimization*. McGraw-Hill.
- Goldberg, D. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley.
- Gutjahr, W. (2010). Convergence Analysis of Metaheuristics. *Annals of Information Systems*, 10, 159–187. doi:10.1007/978-1-4419-1306-7_6

- Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. *IEEE International Conference on Neural Networks*, 1942–1948.
- Kirkpatrick, S., Gelatt, C., & Vecchi, M. (1983). Optimization by Simulated Annealing. *Science*, 220(4598), 671–680. doi:10.1126/science.220.4598.671 PMID:17813860
- Koza, J. (1992). *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. MIT Press.
- Laarhoven, P., & Aarts, E. (1987). *Simulated Annealing: Theory and Applications*. Berlin: Springer. doi:10.1007/978-94-015-7744-1
- Larrañaga, P., & Lozano, J. (Eds.). (2002). *Estimation of Distribution Algorithms: A New Tool for Evolutionary Computation*. Boston: Kluwer. doi:10.1007/978-1-4615-1539-5
- Lourenço, R., Martin, O., & Stutzle, T. (2002). Iterated local search. In Glover & Kochenberger (Eds.), *Handbook of Metaheuristics* (pp. 321–353). Kluwer.
- Luke, S. (2013). *Essentials of Metaheuristics*. Lulu.
- McGeoch, C. (2008). Experimental methods for algorithm analysis. In M.-Y. Kao (Ed.), *Encyclopedia of Algorithms*. Springer. doi:10.1007/978-0-387-30162-4_135
- Mladenovic, N., & Hansen, P. (1997). Variable neighborhood search. *Computers & Operations Research*, 24(11), 1097–1100. doi:10.1016/S0305-0548(97)00031-2
- Moscato, P. (1999). Memetic algorithms: A short introduction. In *New Ideas in Optimization*. McGraw-Hill.
- Nesmachnow, S. (2014). An overview of metaheuristics: Accurate and efficient methods for optimisation. *International Journal of Metaheuristics*, 3(4), 320–347. doi:10.1504/IJMHEUR.2014.068914
- Nesmachnow, S., Luna, F., & Alba, E. (2015). An empirical time analysis of evolutionary algorithms as C programs. *Software, Practice & Experience*, 45(1), 111–142. doi:10.1002/spe.2217
- Padberg, M. (2012). Harmony Search Algorithms for binary optimization problems. In *Operations Research Proceedings* (pp. 343–348). doi:10.1007/978-3-642-29210-1_55
- Rardin, R., & Uzsoy, R. (2001). Experimental Evaluation of Heuristic Optimization Algorithms: A Tutorial. *Journal of Heuristics*, 7(3), 261–304. doi:10.1023/A:1011319115230
- Sörensen, K. (2015). Metaheuristics—the metaphor exposed. *International Transactions in Operational Research*, 22(1), 3–18. doi:10.1111/itor.12001
- Storn, R., & Price, K. (1997). Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, 11(4), 341–359. doi:10.1023/A:1008202821328
- Talbi, E.-G. (2002). A Taxonomy of Hybrid Metaheuristics. *Journal of Heuristics*, 8(5), 541–564. doi:10.1023/A:1016540724870
- Talbi, E.-G. (2009). *Metaheuristics: From Design to Implementation*. Wiley Publishing. doi:10.1002/9780470496916
- Weyland, D. (2010). A Rigorous Analysis of the Harmony Search Algorithm: How the Research Community can be Misled by a Novel Methodology. *International Journal of Applied Metaheuristic Computing*, 1(2), 50–60. doi:10.4018/jamc.2010040104

Yang, X.-S. (2008). *Nature-Inspired Metaheuristic Algorithms*. Luniver Press.

ADDITIONAL READING

Bianchi, L., Dorigo, M., Gambardella, L., & Gutjahr, W. (2009). A survey on metaheuristics for stochastic combinatorial optimization. *Natural Computing*, 8(2), 239–287. doi:10.1007/s11047-008-9098-4

Boussaid, I., Lepagnot, J., & Siarry, P. (2013). A survey on optimization metaheuristics. *Information Sciences*, 237, 82–117. doi:10.1016/j.ins.2013.02.041

Dorigo, M., & Stützle, T. (2004). *Ant Colony Optimization*. MIT Press.

Gendreau, M., & Potvin, J.-Y. (2005). Metaheuristics in Combinatorial Optimization. *Annals of Operations Research*, 140(1), 189–213. doi:10.1007/s10479-005-3971-7

Gendreau, M., & Potvin, J.-Y. (2010). *Handbook of Metaheuristics* (2nd ed.). Springer. doi:10.1007/978-1-4419-1665-5

Glover, F., & Laguna, M. (1997). *Tabu Search*. Kluwer. doi:10.1007/978-1-4615-6089-0

Hansen, P., & Ribeiro, C. (2001). *Essays and Surveys on Metaheuristics*. Kluwer.

Ibaraki, T., Nonobe, K., & Yagiura, M. (2005). *Metaheuristics: Progress as Real Problem Solvers*. Kluwer. doi:10.1007/b107306

Laporte, G., & Osman, I. (1996). Metaheuristics in Combinatorial Optimization. *Annals of Operations Research*, 63.

Michalewicz, Z. (1996). *Genetic algorithms + Data Structures = Evolution programs*. Springer. doi:10.1007/978-3-662-03315-9

Michalewicz, Z., & Fogel, D. (2004). *How to Solve It: Modern Heuristics*. Springer. doi:10.1007/978-3-662-07807-5

Mitchell, M. (1996). *An introduction to genetic algorithms (complex adaptive systems)*. MIT Press.

KEY TERMS AND DEFINITIONS

Diversification: Exploration of the search space providing a reasonable coverage and avoiding stagnation in local optima.

Intensification: Exploitation of already found accurate solutions to increase their quality.

Metaheuristics: High-level strategies that define algorithms to find approximate solutions for optimization problems.

Parallel Metaheuristics: Models that use several computing elements for enhancing and speeding up the search.

Population Metaheuristics: Methods that work with a set of solutions, which are iteratively modified and/or combined to produce new solutions.

Trajectory Metaheuristics: Methods that work with a single solution, which is iteratively replaced by another one in its neighborhood.

An Essay on Denotational Mathematics

Giuseppe Iurato

University of Palermo, Italy

INTRODUCTION

Denotational Mathematics (in short, DM) is a new discipline which serves as a formal language to lay out and solve the various problems posed, amongst others, by cognitive informatics, software science and computational intelligence. It has been worked out in the last decades by Yingxu Wang (2008a-e, 2009a-c, 2010a-b, 2011, 2012a-b)¹, to whose many works and papers we refer for a wider and more complete information science contextualization of DM and its applications.

In this contribution, we wish to briefly outline, but in a rigorous manner, the main formal structures considered by DM, within the framework of classical algebra of relations. Our exposition is extremely technical and synthetic to allow a faster and clearer overall view of the pivotal mathematical structures of DM, which are *concept algebra*, *real-time process algebra*, *system algebra*, *visual semantic algebra*, *granular algebra*, and *inference algebra*.

In relation to its possible applications to the general context of information sciences, numerical mathematics, as regard continuous context, mainly deals with rational ($\mathbb{Q}$), real ($\mathbb{R}$) and complex ($\mathbb{C}$) number systems, while, as regard discrete context, it deals with natural ($\mathbb{N}$), integral ($\mathbb{Z}$) and binary ($\mathbb{B}$) numerical systems. DM may be considered either from the classical algebra viewpoint (as done here) or from the model theory one². In any case, it mainly deals with *hypercomplex* ($\mathbb{H}$) denotational algebraic systems, which will be defined in the next section 4.

BACKGROUND

Generally speaking, methods of mathematics may be roughly classified into *analytical*, *numerical*, and *denotational* approaches. The analytical approach deals with continuous mathematical objects defined in the real number field ($\mathbb{R}$); the numerical approach deals with discrete mathematical entities defined in $\mathbb{R}$ and in binary systems ($\mathbb{B}$) with iteratively and recursively approximation functions.

However, the domains of recent problems in the emerging disciplines of artificial intelligence, cognitive informatics, brain science, cognitive computing, knowledge science, and computational intelligence require rather new *complex hyperstructures* ($\mathbb{H}$), beyond that of pure numbers in ($\mathbb{B}$) and ($\mathbb{R}$), which just identify the denotational approach.

The requirement for reducing complicated formal problems and solutions to the low-level objects in conventional computing theories and the underpinning of analytical or numerical mathematical means, has greatly constrained the inference and computing ability of both human and intelligent systems.

This has motivated the introduction of transdisciplinary investigations provided by *denotational approach* based on novel mathematical structures for modelling and processing new complex hyperstructures required to be established by cognitive computing and computational intelligence, collectively known just as denotational mathematics (DM).

DM may be viewed as a rigorous formal approach to processing both complex architectures and intelligent formal behaviours in order to cope with the fundamental challenges in understanding, describing, and modelling natural and machine intelligence in general, as well as concept, knowledge and semantics, in particular.

As a counterpart of the classic analytic mathematics, DM encompasses new forms of mathematical structures for dealing with complex objects recently emerged in a very wide range of contemporary disciplines and fields such as cognitive informatics, cognitive computing, abstract intelligence, artificial intelligence, computational intelligence, semantic computing, software science, knowledge science, system science and computing with words.

In recognizing mathematics as the main methodology of all informatics sciences and engineering disciplines³, a set of DM structures have been created in the last decades. DM provides a coherent set of contemporary mathematical means and explicit expressive power for dealing with both complex mathematical objects and long-chains of serial and embedded mathematical operations emerged in multiple modern disciplines.

In DM, not only the mathematical entities are greatly complicated at a formal level, but also the mathematical structures and methodologies are significantly expanded from simple relations and independent functions to embedded relations and serially coupled functions.

In what follows, we outline the chief elements and aspects of the main DM structures, with a view toward their possible applications to information sciences and technology.

THE NOTION OF HYPERCOMPLEX ALGEBRAIC SYSTEM

Generally speaking, an *algebraic system* is a tuple $\Xi \stackrel{\text{def}}{=} (\{A_i\}_{i \in I}; \{R_j\}_{j \in J})$, where A_i is a non-empty set for each $i \in I$, and $\phi \neq R_j \subseteq \prod_{k \in K_j} A_k$ for each

$j \in J$, is a *relation*. If, for instance, $I = \{1, 2\}$ and J is a singleton, then $\phi \neq R \subseteq A_1 \times A_2$ is a *binary (external) relation* with *domain* in A_1 and *codomain* in A_2 ; if $A_1 = A_2 = A$, then R is a *binary (internal) relation* on the *support* A .

For example, a *semigroup* is an algebraic system of the type $(A; R)$, where A is a non-empty set (*support*) and R is a ternary internal relation $\phi \neq R \subseteq A^3$ such that

$$(a, b, c) \in R \wedge (a, b, c') \in R \Rightarrow (c = c') \forall a, b, c, c' \in A \quad (\text{univocity of } R),$$

which verifies the associative law. Often, such a ternary internal relation is written as a function of the type $R: A \times A \rightarrow A$, with $(a, b) \in R$ wrote as aRb .

An algebraic subsystem (or restriction) $\check{Z}'$ of Ξ , is an algebraic system of the type

$$\check{Z}' \stackrel{\text{def}}{=} (\{A'_i\}_{i \in I}; \{R'_j\}_{j \in J}), \quad \text{w h e r e} \\ \emptyset \neq A'_i \subseteq A_i \quad \forall i \in I, \text{ and } R'_j \subseteq \prod_{k \in K'_j} A'_k \text{ with} \\ R'_j = R_j \upharpoonright_{\prod_{k \in K'_j} A'_k} \text{ and } K'_j \subseteq K_j \text{ for each } j \in J. \text{ Like-}$$

wise, starting from $\check{Z}'$, Ξ may be considered as an algebraic extension of $\check{Z}'$. The notion of action of a set G on another set A , may be generalized considering it as an external relation of the type $R \subseteq G \times A$ (Artin, 1991).

Let $\mathcal{F} = (\{A_i\}_{i \in I}; \{R_j\}_{j \in J})$ be an algebraic system. A denotational algebraic hypercomplex system $\mathbb{H}_{\mathcal{F}}$ in the framework $\mathcal{F}$, is an algebraic

systems of the type $\mathbb{H}_{\mathcal{F}} \stackrel{\text{def}}{=} (\{B_k\}_{k \in K}; \{S_l\}_{l \in L})$, where $K = K_1 \cup K_2, L = L_1 \cup L_2$, with $K_1 \subseteq I, L_1 \subseteq J$, $K_1 \cap K_2 = L_1 \cap L_2 = \phi$, $B_k \in \{A_i\}_{i \in I}$ for each $k \in K$, $S_l \in \{R_j\}_{j \in J}$ for each $l \in L_1$, $\left(\bigcup_{k \in K_1} B_k \right) \cap \left(\bigcup_{k \in K_2} B_k \right) = \phi$, and $S_l \subseteq \prod_{k \in K_1^l \cup K_2^l} B_k$ with $K_1^l \subseteq K_1, K_2^l \subseteq K_2$ for each $l \in L_2$. Each $S_l, l \in L_1$, is said to be an internal relation of $\mathbb{H}_{\mathcal{F}}$, while

every $S_i, i \in L_2$, is said to be an external relation with input/output values in $B_k, k \in K_2^i \subseteq K_2$.

In what follows, typical instances of denotational algebraic hypercomplex systems ($\mathbb{H}$) are those of concept algebra, system algebra, and granular algebra, of which we shall give a brief formal explanation. $\mathbb{H}$ is a powerful modelling metaphor for abstract objects and systems, which can be used to elicit and model the essences of complex objects with hyperstructures beyond $\mathbb{H}$, and set.

Typical mathematical objects in $\mathbb{H}$ are such as multi-facet objects, complex relations, perceptual information, abstract concepts, formal knowledge, intelligent behaviours, behavioural processes, rational decisions, language semantics, visual semantics, causal inferences, and generic systems.

Almost all these complex objects with $\mathbb{H}$ are the targets objects and typical problems in contemporary disciplines such as cognitive informatics, cognitive computing, abstract intelligence, artificial intelligence, computational intelligence, semantic computing, software science, knowledge science, system science, and computing with words.

THE STRUCTURE OF CONCEPT ALGEBRA

Broadly speaking, a *concept* is the fundamental unit of human knowledge and reasoning. In cognitive informatics, a *concept* is a basic cognitive unit to identify and model a concrete entity in the external world or an abstract object in the perceived world. Its identification is possible only within a semantic framework defined as follows. Concepts are identified as the basic unit of both knowledge and reasoning in cognitive informatics, logic, linguistics, psychology, software engineering, knowledge engineering, artificial intelligence, and computational intelligence.

Formal treatments of concepts by rigorous mathematical models are at the centre of theories

for knowledge representation and manipulation. A concept in linguistics is a noun or noun-phrase that serves as the subject or object of a to-be statement, which can be extended to verbs or verb-phrases for denoting a behaviour. Concepts in denotational mathematics are an abstract structure that carries certain meaning in almost all cognitive processes such as thinking, knowledge representation, learning, and reasoning.

Let $\mathcal{O}$ be a finite and non-empty set of concrete or abstract *objects*, and $\mathcal{A}$ a finite and non-empty set of general *attributes*. Then, a *semantic framework* $\mathcal{F}_{sem}$, is a denotational algebraic system

$\mathcal{F}_{sem} \stackrel{def}{=} (\mathcal{O}, \mathcal{A}, \mathcal{R}_{sem})$, where $\mathcal{R}_{sem}$ is a finite and non-empty set of binary univocal or plurivocal relations, meaning-related, whose domains and codomains are $\mathcal{O}$ and/or $\mathcal{A}$, that is to say, each element $R \in \mathcal{R}_{sem}$ is such that $R \subseteq \mathcal{O} \times \mathcal{A}$ or $R \subseteq \mathcal{A} \times \mathcal{O}$.

So, for example, a binary relation of the type $\mathcal{O} \rightarrow \mathcal{A}$, assigns possible meanings, through attributes of $\mathcal{A}$, to objects of $\mathcal{O}$, while a binary relation of the type $\mathcal{A} \rightarrow \mathcal{O}$, identifies the set of objects of $\mathcal{O}$ having certain attributes belonging to $\mathcal{A}$; likewise, relations of the type $\mathcal{A} \rightarrow \mathcal{A}$, $\mathcal{O} \rightarrow \mathcal{O}$, identify respectively attributes and objects whose meanings are somehow related of each other.

An abstract concept $C_{\mathcal{F}_{sem}}$ in the semantic framework $\mathcal{F}_{sem} = (\mathcal{O}, \mathcal{A}, \mathcal{R}_{sem})$, is a denotational algebraic system of the type $C_{\mathcal{F}_{sem}} \stackrel{def}{=} (\mathcal{O}, A, R^c, R^i, R^o)$, where $\mathcal{O} \in \mathcal{P}(\mathcal{O})$ ($\stackrel{def}{=} \text{power set of } \mathcal{O}$), $A \in \mathcal{P}(\mathcal{A})$ ($\stackrel{def}{=} \text{power set of } \mathcal{A}$), and $R^c \in \mathcal{P}(\mathcal{R}_{sem})$ are finite and non-empty sets respectively of internal objects, internal attributes and internal relations of the abstract concept $C_{\mathcal{F}_{sem}}$ with respect to $\mathcal{F}_{sem}$, while $R^i \subseteq A' \times A$ and $R^o \subseteq A \times A''$, where $A', A'' \in \mathcal{P}(\mathcal{A})$ are assigned finite and non-empty sets of external attributes such that⁴

$A \cap A' = A \cap A'' = \phi$, are respectively input (A') and output (A'') relations of the abstract concept $C_{\mathcal{F}_{sem}}$ with respect to $\mathcal{F}_{sem}$.

A *concept algebra* $\mathfrak{C}_{\mathcal{F}_{sem}}$ in the semantic framework $\mathcal{F}_{sem} = (\mathcal{O}, \mathcal{A}, \mathcal{R}_{sem})$, is a denotational algebraic system of the type $\mathfrak{C}_{\mathcal{F}_{sem}} \stackrel{def}{=} (\mathcal{C}_{\mathcal{F}_{sem}}, \mathfrak{D})$, where $\mathcal{C}_{\mathcal{F}_{sem}}$ is a set of abstract concepts $C_{\mathcal{F}_{sem}}$ in the semantic framework $\mathcal{F}_{sem}$, and $\mathfrak{D} \stackrel{def}{=} \{\mathcal{R}_r, \mathcal{R}_p, \mathcal{R}_c\}$ is a set of *relational* ($\mathcal{R}_r$), *reproductive* ($\mathcal{R}_p$) and *compositional* ($\mathcal{R}_c$) computational relations among abstract concepts belonging to $\mathcal{C}_{\mathcal{F}_{sem}}$.

Among other things, concept algebra provides a denotational mathematical means for algebraic manipulations of abstract concepts. Concept algebra can be used to model, specify, and manipulate generic “to be” type problems, particularly system architectures, knowledge bases, and system refinement, in cognitive informatics, knowledge science, computational intelligence, computing science, software science, and computational linguistics. Learning is an high-level cognitive process according to the so-called *Layered Reference Model of the Brain* (LRMB) in cognitive informatics.

As a core capability of the natural and machine intelligence, learning gains knowledge of something or acquires skills in some actions by updating the cognitive models in *Long-Term Memory* (LTM). According to the *Object-Attribute-Relation* (OAR) model, the results of learning can be embodied by the updating of the existing OAR in the brain. In other words, learning is a dynamic composition of the currently created sub-OAR and the existing OAR in LTM.

THE STRUCTURE OF REAL-TIME PROCESS ALGEBRA

A central task of software science is the simulation of the behaviour of an arbitrary *intelligent system*

S . The most typical feature of an arbitrary intelligent system S , we deem to be its *memory* M_S .

A *behaviour* $\mathcal{B}$ of S is then operationally ruled by M_S , through a given set of computational operations $\mathcal{O}_S$, i.e., one or several variable mathematical functions input/output-producing, which governs a certain *action* $\mathcal{A}$ of S performed into a given *environment* $\mathcal{E}$, usually a *space-time* ST , in a certain time interval $\mathcal{T}$ specifying its *lifetime*. In dependence on $\mathcal{T}$ is a *compile* time $\mathcal{T}_c$ or a *run* time $\mathcal{T}_r$, $\mathcal{B}$ is respectively classified as a *static* behaviour or as a *dynamic* behaviour.

The basic units of system behaviours can be formalized through *processes*, which are defined as follows. Let $\mathfrak{V}$ be a finite and non-empty set of *variables* (or *operands*), $\mathfrak{S}$ a finite and non-empty set of *structures*, $\mathfrak{T}$ a finite and non-empty set of *types* (or *languages*), $\mathfrak{P}$ a finite and non-empty set of *meta-processes*, and $\mathfrak{R}$ a finite and non-empty set of *relational operators* or *process relations*. In general, $\mathfrak{S} \subseteq \mathfrak{V} \times \mathfrak{V}$, $\mathfrak{R} \subseteq \mathfrak{P} \times \mathfrak{P}$.

A meta-process of $\mathfrak{P}$ is a primitive or irreducible computational operation that cannot be reduced further to individual actions or behaviours, while a process relation of $\mathfrak{R}$ is an algebraic operation or a compositional rule between two or more meta-processes which gives rise to other more complex processes. The denotational algebraic system $\mathcal{F}_{pr} \stackrel{def}{=} (\mathfrak{V}, \mathfrak{S}, \mathfrak{T}, \mathfrak{P}, \mathfrak{R})$ is said to be a *process framework*.

An *abstract process* P in the process framework $\mathcal{F}_{pr} = (\mathfrak{V}, \mathfrak{S}, \mathfrak{T}, \mathfrak{P}, \mathfrak{R})$, is given by a finite, well-defined process iteration, i.e.,

$$P \stackrel{def}{=} \odot_{i=1}^{n-1} \gamma_{ii+1}(p_i, p_{i+1}) = p_1 \circ p_2 \circ \dots \circ p_n,$$
 which is a compendious notation for the sequence

of processes $\xrightarrow{p_1} \gamma_{12} \xrightarrow{p_2} \gamma_{23} \xrightarrow{p_3} \dots \xrightarrow{p_{n-1}} \gamma_{n-1n} \xrightarrow{p_n}$, in which $p_i \in \mathfrak{P}$ for each i , and $\gamma_{ij} \in \mathfrak{R}$ for each i, j . The big operator $\odot$ denotes a finite series of repetitive functions or composition of functions, which indicates that any i th function (statement)

given in a process P , is a relational composition with all cumulated $i-1$ previous functions.

A *real-time process algebra* $\wp_{\mathcal{F}_{pr}}$ in the process framework $\mathcal{F}_{pr} = (\mathfrak{V}, \mathfrak{S}, \mathfrak{T}, \mathfrak{P}, \mathfrak{R})$, is a denotational algebraic system of the type $\wp_{\mathcal{F}_{pr}} \stackrel{def}{=} (\mathcal{T}, \mathcal{P}, \mathcal{R})$, where $\mathcal{T} \subseteq \mathfrak{T}$ is a finite and non-empty set of primitive types for modelling computational entities, $\mathcal{P} \subseteq \mathfrak{P}$ is a finite and non-empty set of meta-processes for modelling fundamental system behaviours, and $\mathcal{R} \subseteq \mathfrak{R}$ is a finite and non-empty set of relational and compositional operators acting on $\mathcal{P}$ to build up complex system behaviours. There are applications of RTPA in human and system behavioural and functional modelling.

RTPA provides a coherent notation system and a formal engineering methodology for modelling software and intelligent systems as well as human behaviours. RTPA can be used to describe both logical and physical models of systems, where logic views of the architecture of a software system and its operational platform can be described using the same set of unified mathematical notations. When the system architecture is formally modelled, the process behaviours that perform on the structural models can be rigorously specified.

THE STRUCTURE OF SYSTEM ALGEBRA

Systemic approach to science is based on the general notion of *system* which, broadly speaking, is a coherent collection of interacting entities within a stable and well-defined environment. A system may be *open* or *closed* in dependence of its interactions with the related environment Θ . For the former, typical interactions with the environment occur with a wide range of inputs and outputs.

Let $\mathfrak{C}$ be a finite and non-empty set of *components*, $\mathfrak{B}$ a finite and non-empty set of *behaviours*, and $\mathfrak{R}$ a finite and non-empty set of binary relations between $\mathfrak{C}$ and $\mathfrak{B}$. Then, the

algebraic system $\mathcal{F}_{sys} \stackrel{def}{=} (\mathfrak{C}, \mathfrak{B}, \mathfrak{R})$ is said to be a *system framework*. An *abstract system* $\mathcal{S}_{\mathcal{F}_{sys}}$ in the system framework $\mathcal{F}_{sys} = (\mathfrak{C}, \mathfrak{B}, \mathfrak{R})$, is a denotational algebraic system of the type $\mathcal{S}_{\mathcal{F}_{sys}} \stackrel{def}{=} (C, B, R^c, R^i, R^o, \odot, \sim)$, where $C \subseteq \mathfrak{C}$ is a finite and non-empty set of *internal* components, $B \subseteq \mathfrak{B}$ is a finite and non-empty set of *internal* behaviours, $R^c \subseteq \mathfrak{R}$ is a finite and non-empty set of *internal* relations, $R^i \subseteq C' \times C$ is a finite and non-empty set of *input* (C') relations, $R^o \subseteq C \times C''$ is a finite and non-empty set of *output* (C'') relations, where $C', C'' \subseteq \mathfrak{C}$ are finite and non-empty subsets of *external* components of the *environment* $\sim = \{C', C''\} \subseteq \mathcal{P}(\mathfrak{C})$ ($=$ power set of $\mathfrak{C}$) such that⁵ $C' \cap C = C'' \cap C = \emptyset$, and $\odot \subseteq \mathfrak{B}$ is a finite and non-empty set of behaviours, said to be *constraints* on C and B , which provide further information about their elements.

A *system algebra* $\mathcal{SA}_{\mathcal{F}_{sys}}$ in the system framework $\mathcal{F}_{sys}$, is a denotational algebraic system of the type $\mathcal{SA}_{\mathcal{F}_{sys}} \stackrel{def}{=} (\mathfrak{G}_{\mathcal{F}_{sys}}, \mathcal{R})$, where $\mathfrak{G}_{\mathcal{F}_{sys}}$ is a set of system algebras of the type $\mathcal{S}_{\mathcal{F}_{sys}}$, while $\mathcal{R}$ is a finite set of computational relations among elements of $\mathfrak{G}_{\mathcal{F}_{sys}}$.

System algebra is a denotational mathematics for rigorous system modelling, computing, and manipulation. System algebra models any concrete system using the general abstract system model and manipulates complex system behaviours as algebraic operations on the abstract system. System algebra provides a denotational mathematical means for algebraic manipulations of all forms of abstract systems.

System algebra can be used to model, specify, and manipulate generic “to be” and “to have” type problems, particularly system architectures, system analyses, system refinements, and system syntheses in extremely wide areas. System algebra

is not only useful for system behavioural manipulations, but also efficient in system architectural modelling and refinement. In the latter case, a system can be integrated from the bottom up by a series of compositions level-by-level in a given system hierarchy.

THE STRUCTURE OF VISUAL SEMANTIC ALGEBRA

In order to model abstract visual objects, particularly their semantic representation, composition and manipulation as symbolic semantic entities, the structure of visual semantic algebra is introduced. In this regard, the basic 2-D and 3-D geometric units, collectively known as *geons*, will be the primary and fundamental elements to build up visual semantic algebra.

Let $\mathfrak{O}$ be a finite and non-empty set of *visual objects* (or *geons*), and $\mathfrak{R}$ a finite and non-empty set of *spatial relational operations*, which formalize the basic, elementary geometric transformations. Then, the denotational algebraic system $\mathcal{F}_{vis}^{def} = (\mathfrak{O}, \mathfrak{R})$ is said to be a *visual object framework*.

Usually, $\mathfrak{R} \subseteq \mathfrak{O} \times \mathfrak{O}$, while the set of *primitive* or *irreducible visual objects* O is given by the union of the sets of 2-D shapes (H), 3-D solids (S), generic figures (F) and spatial reference limits (L), that is, $O = H \cup S \cup F \cup L$, which will be related among them through algebraic operations whose set $R \subseteq \mathfrak{R}$ comprehends relational and behavioural computational operators.

The denotational algebraic system $\mathcal{VA}_{\mathcal{F}_{vis}}^{def} = (O, R)$ is said to be a *visual semantic algebra*.

The physical space is used to be formally modelled by the Euclidean space that is a collection of all abstract points in three coordinates. Therefore, the spatial properties of concrete and abstract objects, shapes, and their interrelations can be described by Euclidean plane and solid geometry.

However, it is discovered that the brain does not process visual objects and patterns as bit maps as a camera rather than representing them as abstract visual semantics, because the latter approach is more efficient in visual information processing.

In order to efficiently model the abstract visual objects, particularly their semantic representations, compositions, and manipulations, a new structure of denotational mathematics known as visual semantic algebra is introduced in this section for expressing the semantics of visual objects and patterns in cognitive computing.

THE STRUCTURE OF GRANULAR ALGEBRA

A *computing granule* is the basic mathematical structure of a *granular system*, with a stable topological setting and bearing, at least, a unit of computational capability or behaviour. If $\mathfrak{E}$ is a finite and non-empty set of *granules*, $\mathfrak{R}$ a finite and non-empty set of *relations* on $\mathfrak{E}$, and $\mathfrak{C}$ a finite and non-empty set of *constraints* on $\mathfrak{E}$ and $\mathfrak{R}$, then the denotational algebraic system

$\mathcal{F}_{gr}^{def} = (\mathfrak{E}, \mathfrak{R}, \mathfrak{C})$ is said to be a *granular framework*.

An *abstract granule* $\mathcal{G}_{\mathcal{F}_{gr}}$ in the granular framework $\mathcal{F}_{gr} = (\mathfrak{E}, \mathfrak{R}, \mathfrak{C})$, is a denotational algebraic system of the type $\mathcal{G}_{\mathcal{F}_{gr}}^{def} = (G, B, R^c, R^i, R^o, \odot, \sim)$, where $G \subseteq \mathcal{P}(\mathfrak{E})$ is a finite and non-empty set of *cells* of granules, $B \subseteq \mathfrak{R}$ is a finite and non-empty set of *behaviours* of G , $R^c \subseteq \mathfrak{R}$ is a finite and non-empty set of *internal relations*, $R^i \subseteq G' \times G$ is a finite and non-empty set of *input* (G') relations, $R^o \subseteq G \times G''$ is a finite and non-empty set of *output* (G'') relations, $\odot \subseteq \mathfrak{C}$ is a finite and non-empty set of *constraints* on G and B , and $\sim = \{G', G''\} \subseteq \mathcal{P}(\mathfrak{E})$ is the *environment* of $\mathcal{G}_{\mathcal{F}_{gr}}$ made by the *external cells* G', G'' , outside G , such that⁶ $G' \cap G = G'' \cap G = \phi$.

A granular algebra $\mathcal{GA}_{\mathcal{F}_{gr}}$ in the granular framework $\mathcal{F}_{gr} = (\mathfrak{C}, \mathfrak{R}, \mathfrak{E})$, is a denotational algebraic structure of the type $\mathcal{GA}_{\mathcal{F}_{gr}} \stackrel{def}{=} (\mathcal{G}, \mathcal{R})$, where $\mathcal{G}$ is a finite and non-empty set of abstract granules $\mathcal{G}_{\mathcal{F}_{gr}}$ in the granular framework $\mathcal{F}_{gr}$, and $\mathcal{R}$ is a finite and non-empty set of computational relations among elements of $\mathcal{G}$.

Granular algebra is a denotational mathematical structure for modelling granular systems and granular computing. The term granule is originated from Latin granum, i.e., grain, to denote a small compact particle in physics and in the natural world. However, the taxonomy of granules in computing can be classified into data granule, information granule, concept granule, computing granule, cognitive granule, and system granule.

A computing granule, shortly a granule, is a basic mathematical structure that possesses a stable topology and at least a unit of computational capability or behaviour. Granular computing is a new computational methodology that models and implements computational structures and functions by a granular system, where each granule in the system carries out a predefined function or behaviour by interacting to other granules in the system.

THE STRUCTURE OF INFERENCE ALGEBRA

Inferences are the basic elements of formal reasoning. *Inference algebra* formalizes usual, everyday and empirical cause-effect type reasoning processes in a rigorous fashion, trying to identify and reveal the possible causations in a thread of thought, so allowing their computational implementations in the cognitive informatics context.

Let $\mathfrak{S}$ be a finite and non-empty set of *states* or *facts*, given by the union of the sets of *causes* $\mathfrak{C}$ and *effects* $\mathfrak{E}$, and $\mathfrak{R}$ a finite and non-empty set of relations among causes and effects, i.e., $\mathfrak{R} \subseteq \mathfrak{C} \times \mathfrak{E}$, said to be *causations*. Then, the

denotational algebraic system $\mathcal{F}_{inf} \stackrel{def}{=} (\mathfrak{C}, \mathfrak{E}, \mathfrak{R})$ is said to be an *inference framework*.

An *abstract causation* ξ in the inference framework $\mathcal{F}_{inf} = (\mathfrak{C}, \mathfrak{E}, \mathfrak{R})$, is a functional relation $\xi \in \mathfrak{R}$ of the type $f_\xi: C \rightarrow E$ which maps a non-empty set of causes $C \subseteq \mathfrak{C}$ onto a set of effects $E \subseteq \mathfrak{E}$. If $\mathfrak{Z}_{\mathcal{F}_{inf}} \subseteq \mathfrak{R}$ is the category of abstract causations in the inference framework $\mathcal{F}_{inf}$, then $\mathfrak{Z}_{\mathcal{F}_{inf}}$ consists in four main types of formal causations in dependence on these are formally considered as functions in Boolean or numerical variables.

A *formal inference* κ in the inference framework $\mathcal{F}_{inf}$, is a formal reasoning process which identifies a causation $\xi \in \mathfrak{Z}_{\mathcal{F}_{inf}}$, say $f_\xi = \vdash$, between the elements of a certain pair (C, E) of non-empty sets of causes $C \subseteq \mathfrak{C}$ and effects $E \subseteq \mathfrak{E}$, that is to say, $\kappa = C \vdash E = f_\xi(C)$.

An *inference algebra* $\mathcal{IA}_{\mathcal{F}_{inf}}$ in the inference framework $\mathcal{F}_{inf}$, is a denotational algebraic system $\mathcal{IA}_{\mathcal{F}_{inf}} \stackrel{def}{=} (\mathfrak{Z}, \mathfrak{R})$, where $\mathfrak{Z}$ is a finite and non-empty set of formal inferences κ , and $\mathfrak{R}$ is a finite and non-empty set of computational relations between elements of $\mathfrak{Z}$.

Inference algebra is a denotational mathematics for cognitive computing and machine reasoning. Inference as the basic mechanism of thought is one of the gifted abilities of human beings, which is a cognitive process that creates rational causations between a pair of cause and effect based on empirical arguments, formal reasoning, and/or statistical norms. Inference algebra introduces a calculus of discrete causal differential and formal models of causations, based on them nine algebraic inference operators of inference algebra are created for manipulating the formal causations.

Inference algebra elicits and formalizes the common and empirical reasoning processes in a rigorous form, which enable artificial intelligence and computational intelligent systems to mimic

human inference abilities in cognitive computing. Inference algebra is one of the basic studies towards the next generation of intelligent computers known as cognitive computers. A wide range of applications of inference algebra are identified and demonstrated in cognitive informatics and computational intelligence towards novel theories and technologies for machine-enabled inferences and reasoning.

FUTURE RESEARCH DIRECTIONS

Among the typical paradigms of denotational mathematics, above all concept algebra has been designed for knowledge processing in cognitive computing, computational linguistics, and computational intelligence, which models concrete knowledge as abstract concepts and semantic relations, and manipulates complex knowledge and semantics as algebraic operations on the abstract concepts.

Real-Time process algebra (RTPA) has been designed for manipulating computational structures and behaviours, which models concrete data objects as structure models (SMs), represents system behaviours as process models (PMs), and manipulates complex behavioural processes as algebraic operations on the abstract SMs and PMs. System algebra (SA) has been designed for rigorous system manipulations, which represents real-world systems using the universal abstract system model and manipulates complex system behaviours as algebraic operations on the abstract systems.

Visual Semantic Algebra (VSA) has been designed for geometric analyses and compositions of visual patterns, which models visual objects and patterns as abstract and symbolic semantic entities and manipulates complex visual semantics as algebraic operations on the abstract semantic objects. Granular algebra (GA) has been designed for granular computing, which models concrete granules as an abstract mathematical structure

and manipulates complex granular behaviours as algebraic operations on the abstract granules.

Inference algebra (IA) has been designed for cognitive computing and machinable thought, which models human and machine inferences as abstract causations and manipulates complex inference processes as algebraic operations on the abstract causations. Based on these fundamental researches, denotational mathematics, as a metamethodology, has found a wide range of applications in formal inferences, knowledge manipulations, cognitive learning, as well as complex system modelling.

CONCLUSION

In the present contribution, we have outlined, in a very brief but rigorous manner, the basic formal structures of denotational mathematics, according to Y. Wang, from an elementary algebraic standpoint, making use of naïve set theory to be clearer in the explanation of their basic mathematical essence.

We have left out to quote the sources and motivations for these new algebraic structures, mainly coming from computer sciences, as well as their possible applications, referring to the original Wang's papers to this end, as reported, for example, in (Wang, 2012b).

In another place, instead, we will give a more exhaustive and complete formal exposition of the same above mentioned formal structures, but conducted within the wider and more general framework of model theory. The real-world expressional needs for denotational mathematics are such as abstraction of system structures and system behaviours, knowledge representation and manipulations, intelligence generation, effective complexity handling, rigorous system modelling, and expressive system denotation. These practical demands for denotational mathematics are driven by applications in the emerging fields of cognitive informatics, intelligence science, knowledge sci-

ence, brain science, computational intelligence, software science, system science, and cognitive computers.

The extension of conventional algebra onto more complicated mathematical objects beyond pure numbers leads to the contemporary denotational mathematics, which is applied to denote complex behaviours and long chains of inference processes of humans and intelligent systems. It is recognized that humans are not thinking and reasoning at the binary level in $\mathbb{B}$ or $\mathbb{R}$. Denotational mathematics extends the mathematical means for rigorous inference and modelling to a higher level, $\mathbb{H}$, closer to that of humans. Six paradigms of denotational mathematics such as concept algebra, RTPA, system algebra, VSA, granular algebra, and inference algebra will be described in the following sections.

A set of novel operators for each paradigm of denotational mathematics will be elaborated. Typical applications of denotational mathematics will be demonstrated in each section, which show how denotational mathematics may greatly improve the expressive power and efficiency in complex system modelling and intelligent behaviour manipulations. A general finding in computer science, system science, and denotational mathematics is that any complex system may be modelled by the interactions between its structures and functions (behaviours).

The six paradigms of denotational mathematics as presented in the following sections indicate that, although the general model of denotational mathematics is unique, the specific forms of each denotational mathematics and the operators on its structures vary due to the differences of the hyperstructural models of specific mathematical objects such as the abstract concept, system, granule, and causation. This is perfectly in line with the basic system principle that the structure determines or affects the behaviours and mathematical operations on it.

REFERENCES

- Artin, M. (1991). *Algebra*. Englewood Cliffs, NJ: Prentice Hall, Inc.
- Hoare, C. A. R. (1985). *Communicating Sequential Processes*. Englewood Cliffs, NJ: Prentice Hall International Series in Computer Science.
- Wang, Y. (2002). The Real-Time Process Algebra (RTPA). *Annals of Software Engineering*, 14(1/4), 235–274. doi:10.1023/A:1020561826073
- Wang, Y. (2008a). *Software Engineering Foundations. A Software Science Perspective*. Boca Raton, FL: Auerbach Publications, Taylor & Francis Group, LLC.
- Wang, Y. (2008b). RTPA: A Denotational Mathematics for Manipulating Intelligent and Computational Behaviours. *International Journal of Cognitive Informatics and Natural Intelligence*, 2(2), 44–62. doi:10.4018/jcini.2008040103
- Wang, Y. (2008c). On Contemporary Denotational Mathematics for Computational Intelligence. *Transactions of Computational Science*, 2, 6–29.
- Wang, Y. (2008d). On Concept Algebra: A Denotational Mathematical Structure for Knowledge and Software Modeling. *International Journal of Cognitive Informatics and Natural Intelligence*, 2(2), 1–19. doi:10.4018/jcini.2008040101
- Wang, Y. (2008e). On System Algebra: A Denotational Mathematical Structure for Abstract Systems and Modeling. *International Journal of Cognitive Informatics and Natural Intelligence*, 2(2), 20–43. doi:10.4018/jcini.2008040102
- Wang, Y. (2009b). On Visual Semantic Algebra (VSA): A Denotational Mathematical Structure for Modeling and Manipulating Visual Objects and Patterns. *International Journal of Software Science and Computational Intelligence*, 1(4), 1–16. doi:10.4018/jssci.2009062501

Wang, Y. (2009c). Paradigms of Denotational Mathematics for Cognitive Informatics and Cognitive Computing. *Fundamenta Informaticae*, 90(3), 282–303.

Wang, Y. (2010a). On Concept Algebra for Computing with Words (CWW). *International Journal of Semantic Computing*, 4(3), 331–356. doi:10.1142/S1793351X10001061

Wang, Y. (2010b). Keynote: Cognitive Informatics and Denotational Mathematics Means for Brain Informatics. *1st International Conference on Brain Informatics (ICBI '10)*. Toronto, Canada: Springer-Verlag.

Wang, Y. (2011). Inference Algebra (IA): A Denotational Mathematics for Cognitive Computing and Machine Reasoning. *International Journal of Cognitive Informatics and Natural Intelligence*, 5(4), 62–83. doi:10.4018/jcini.2011100105

Wang, Y. (2012a). On Granular Algebra: A Denotational Mathematical Structure for Modeling Granular Systems and Granular Computing. *International Journal of Advanced Mathematics and Applications*, 3(1), 60–73. doi:10.1166/jama.2014.1052

Wang, Y. (2012b). In Search of Denotational Mathematics: Novel Mathematical Means for Contemporary Intelligence, Brain, and Knowledge Science. *Journal of Applied Mathematics and Applications*, 1(1), 4–25. doi:10.1166/jama.2012.1002

Wang, Y., Zadeh, L. A., & Yao, Y. (2009a). On the System Algebra Foundations for Granular Computing. *International Journal of Software Science and Computational Intelligence*, 1(1), 64–86. doi:10.4018/jssci.2009010105

Zaitsev, D. A., & Sleptsov, A. I. (1997). State equations and equivalent transformations for timed Petri nets. *Cybernetics and Systems Analysis*, 33(5), 659–672. doi:10.1007/BF02667189

ADDITIONAL READING

Greniewski, H., Bochenek, K., & Marczyński, R. (1955). Application of bi-elemental boolean algebra to electronic circuits. *Studia Logica*, 2(1), 7–75. doi:10.1007/BF02124765

Gustaffson, B. (2011). *Fundamentals of Scientific Computing*. New York, NY: Springer-Verlag. doi:10.1007/978-3-642-19495-5

Jenkins, K. N. (2002). *Teach Yourself Algebra for Electronic Circuits*. New York, NY: The McGraw-Hill Companies, Inc.

Jenkyns, T., & Stephenson, B. (2013). *Fundamentals of Discrete Mathematics for Computer Science. A Problem-Solving Primer*. New York, NY: Springer-Verlag. doi:10.1007/978-1-4471-4069-6

Makinson, D. (2008). *Sets, Logic and Maths for Computing*. New York, NY: Springer-Verlag. doi:10.1007/978-1-84628-845-6

Moller, F., & Struth, G. (2013). *Modelling Computer Systems. Mathematics for Computer Science*. New York, NY: Springer-Verlag. doi:10.1007/978-1-84800-322-4

O'Regan, G. (2013). *Mathematics in Computing. An Accessible Guide to Historical, Foundational and Application Contexts*. New York, NY: Springer-Verlag.

Pace, G. J. (2012). *Mathematics of Discrete Structures for Computer Science*. New York, NY: Springer-Verlag. doi:10.1007/978-3-642-29840-0

ENDNOTES

¹ See also references therein.

² This treatment to be considered explained enough, will deserve another wider place than the present one.

³ See, for example, (Greniewsky et al., 1955), (Gustaffson, 2011), (Jenkins, 2002), (Jenkins & Stephenson, 2013), (Makinson, 2013), (Moller & Struth, 2012), (O'Regan, 2013), (Pace, 2012). See also (Hoare, 1985), (Zaitsev & Sleptsov, 1997), which have studied algebra of processes for verification of concurrent systems, as well as introduced an algebra unifying arithmetic, logic, and

time operations; they derived laws of this algebra, and applied them for equivalent transformations of Petri nets.

⁴ This condition may be weakened requiring only that A, A', A'' are distinct sets.

⁵ This condition may be weakened requiring only that C, C', C'' are distinct sets.

⁶ This condition may be weakened requiring only that G, G', G'' are distinct sets.

Quantum Computing and Quantum Communication

T

Göran Pulkkis*Arcada University of Applied Sciences, Finland***Kaj J. Grahn***Arcada University of Applied Sciences, Finland*

INTRODUCTION

Quantum computing and quantum communication are based on quantum physics. Information is represented by quantum states defined by energy levels of molecules, atoms, and photons. Quantum information technology includes quantum computers, other quantum information processing devices, quantum programming methodologies, and quantum communication applications such as quantum key distribution systems and quantum signatures. Error correction is a necessary quantum information technology property because of high error vulnerability caused by quantum physics characteristics. This article presents state-of-the-art and future perspectives of quantum computing and communication.

BACKGROUND

Feynman's (1982) observation, that a classical computer cannot efficiently simulate the stochastic parallelism of quantum states, started research on using quantum mechanical effects for efficient information processing. Operating principles and implementation possibilities of quantum computing were outlined in Oxford University (Deutsch, 1985).

Bennett and Brassard (1984) proposed a quantum protocol, BB84, for perfectly secret information transfer. Efficient quantum algorithms, for example integer factorization (Shor, 1994) and

unsorted search (Grover, 1996), were proposed. BB84 has been used for distribution of symmetric encryption/decryption keys (Quantum Key Distribution, QKD) in research networks (Elliot, Pearson, & Troxel, 2004; Quellette, 2004; Poppe, Momtchil, & Maurhart, 2008). Commercial QKD technology has been available over 10 years. Quantum digital signatures have been proposed and experimentally verified (Gottesman & Chuang, 2001; Lu & Feng, 2005; Clarke et al., 2012).

Small scale quantum circuit based computers have been built and successfully tested in research laboratories (Vandersypen et al., 2001; Monz et al., 2011). Since 2011 large scale commercial adiabatic quantum computers (Das & Chakrabarti, 2008) are available.

Achievements in quantum computing and quantum communication from 1970 till October 2015 are listed in (Timeline, 2015). Universities engaged in research and education on this topic are for example (qis.mit.edu, 2013; ...from Quantum, 2011; mathQI, 2015; Quantum, 2015a; Quantum, 2015b)

INFORMATION REPRESENTATION WITH QUANTUM STATES

Two quantum states, for which a state transition exists, can represent a bit called a quantum bit or qubit, if the energy level of both states is measurable. However, quantum states are probabilistic. A measured energy level is one of several possi-

bilities. Each possible outcome has a probability. The sum of all possible outcome probabilities is of course 1.

The *No Cloning* qubit property is the impossibility to clone an unknown quantum state. However, *Teleportation*, which changes the original qubit state, can transfer also an unknown quantum state.

Mathematical Treatment of Qubits

A qubit representation is a 2-dimensional vector. The orthogonal base vectors $(1,0)^T, (0,1)^T$ in *Dirac notation* $|0\rangle, |1\rangle$ represent binary values 0,1.

A qubit is a concurrent superposition of $|0\rangle, |1\rangle$. A measurement outcome is $|0\rangle$ or $|1\rangle$. In a qubit

$$|\psi\rangle = a \cdot |0\rangle + b \cdot |1\rangle. \quad (1)$$

a, b are complex numbers and

$$\langle \psi | \psi \rangle = (a^*, b^*) \cdot (a, b)^T = a^* \cdot a + b^* \cdot b = |a|^2 + |b|^2 = 1. \quad (2)$$

$\{a^*, b^*\}$ are complex conjugates of $\{a, b\}$. $\{|a|^2, |b|^2\}$ are measurement probabilities of $|0\rangle, |1\rangle$.

Multiple Qubits

A 2 *qubit* quantum state is a 2^2 component column vector. For

$$|\psi_1\rangle = a \cdot |0\rangle + b \cdot |1\rangle, |\psi_2\rangle = c \cdot |0\rangle + d \cdot |1\rangle \quad (3)$$

the state is

$$|\psi_1 \psi_2\rangle = |\psi_1\rangle \otimes |\psi_2\rangle = a \cdot c \cdot |00\rangle + a \cdot d \cdot |01\rangle + b \cdot c \cdot |10\rangle + b \cdot d \cdot |11\rangle. \quad (4)$$

$\otimes$ is the *tensor product*. $|00\rangle = |0\rangle \otimes |0\rangle = (1, 0, 0, 0)^T$, $|01\rangle = |0\rangle \otimes |1\rangle = (0, 1, 0, 0)^T$, $|10\rangle = |1\rangle \otimes |0\rangle = (0, 0, 1, 0)^T$, and $|11\rangle = |1\rangle \otimes |1\rangle = (0, 0, 0, 1)^T$ are the *base vectors*.

A N *qubit* quantum state $|\psi_1 \psi_2 \dots \psi_N\rangle$ is a superposition of 2^N base vectors. For a 3 qubit state $|\psi_1 \psi_2 \psi_3\rangle$ the base vectors are $\{|000\rangle, |001\rangle, |010\rangle, |011\rangle, |100\rangle, |101\rangle, |110\rangle, |111\rangle\}$, where $|001\rangle = |0\rangle \otimes |0\rangle \otimes |1\rangle = (0, 1, 0, 0, 0, 0, 0, 0)^T$, etc.

The state $|\psi_1 \psi_2 \dots \psi_N\rangle$ is *entangled* if no $|\psi_1\rangle, |\psi_2\rangle, \dots, |\psi_N\rangle$ exist for which $|\psi_1 \psi_2 \dots \psi_N\rangle = |\psi_1\rangle \otimes |\psi_2\rangle \otimes \dots \otimes |\psi_N\rangle$.

Example: $(1/\sqrt{2}) \cdot (|00\rangle + |11\rangle)$ is entangled.

Proof: $(a \cdot |0\rangle + b \cdot |1\rangle) \otimes (c \cdot |0\rangle + d \cdot |1\rangle) = a \cdot c \cdot |00\rangle + a \cdot d \cdot |01\rangle + b \cdot c \cdot |10\rangle + b \cdot d \cdot |11\rangle \neq (1/\sqrt{2}) \cdot (|00\rangle + |11\rangle)$, since one of $\{a, d\}$ must be 0.

Physical Implementation of Qubits

Qubits have been implemented by photon states, ion traps, cavity quantum electrodynamics (QED), nuclear magnetic resonance (NMR), quantum dots, superconducting circuits, and in silicon (Nielsen & Chuang, 2002; Devoret & Schoelkopf, 2013; Veldhorst et al., 2015).

A physical qubit property is decoherence time, during which a qubit state is maintained before collaps caused by interaction with the physical environment.

A qubit implementation in quantum communication is a photon *polarization state* consisting of all propagation planes of the electromagnetic wave of the photon. A random polarization is a superposition of an orthogonal state pair. Orthogonal state pair examples:

- Horizontal and vertical polarization.
- $+45^\circ$ and -45° diagonal polarization.

A photon polarization state is thus a qubit

$$|\psi\rangle = a \cdot |horis\rangle + b \cdot |vert\rangle = c \cdot | +45^\circ \rangle + d \cdot | -45^\circ \rangle \quad (5)$$

a, b, c, d are complex numbers and $|a|^2 + |b|^2 = |c|^2 + |d|^2 = 1$. One state in a chosen pair must be interpreted as $|0\rangle$, the other as $|1\rangle$. Pho-

ton polarization is measured with a filter. After measurement, only the component allowed by the filter proceeds.

The No Cloning qubit property prevents copying unknown polarization which, however, can be teleported between photons.

QUANTUM COMPUTING

This chapter outlines quantum computation models, quantum algorithms, quantum programming, and quantum computation realizations.

Quantum Computation Models

Proposed models:

- **Quantum Circuit:** Computations are sequences of quantum gate operations, which are reversible transformations on multiple bit quantum registers.
- **Adiabatic Quantum Computer:** Computing by quantum annealing (Das & Chakrabarti, 2008).
- **One-Way Quantum Computer:** An entangled resource state is prepared and thereafter qubits are measured (Raussendorf & Briegel, 2001).
- **Topological Quantum Computer:** Computation is decomposed into braiding of anyons in a 2D lattice (Sarma, Freedman, & Nayak, 2006).

This subchapter presents the quantum circuit and the adiabatic quantum computation models, which are shown to be equivalent (Aharonov et al., 2004).

Quantum Gates and Circuits

Quantum gates change qubit states. Single qubit gates:

- Identity I , $I \cdot |\psi\rangle = |\psi\rangle$
- Negation X (Pauli x), $X \cdot |0\rangle = |1\rangle$, $X \cdot |1\rangle = |0\rangle$
- Phase Shift Z (Pauli z), $Z \cdot |0\rangle = |0\rangle$, $Z \cdot |1\rangle = (-1) \cdot |1\rangle$
- Hadamard H , $H \cdot |0\rangle = 2^{-1/2} \cdot (|0\rangle + |1\rangle)$, $H \cdot |1\rangle = 2^{-1/2} \cdot (|0\rangle - |1\rangle)$

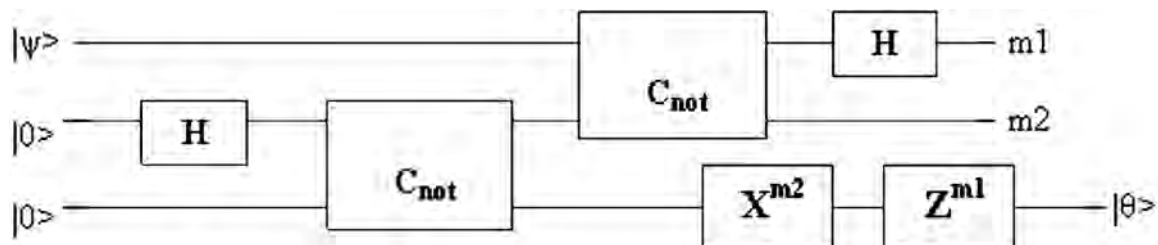
2 qubit gates:

- Identity I , $I \cdot |\psi_1 \psi_2\rangle = |\psi_1 \psi_2\rangle$
- Controlled-NOT C_{not} , $C_{not} \cdot |0\psi\rangle = |0\psi\rangle$, $C_{not} \cdot |1\psi\rangle = |1\rangle \otimes X \cdot |\psi\rangle$

$N \geq 2$ qubit gates are $2^N \times 2^N$ matrices G and $G \cdot G^* = I$. G^* is the complex conjugate matrix of G .

Quantum circuit: an interconnection of quantum gates. Figure 1: a teleportation circuit. The upper qubit is the control in both C_{not} gates. $m1, m2$ are measured qubit values. $X^{m2} = I, X$ for $m2 = 0, 1$. $Z^{m1} = I, Z$ for $m1 = 0, 1$. No Cloning occurs, since measurement $m1$ changes $|\psi\rangle$.

Figure 1. Quantum circuit teleporting the unknown qubit state $|\psi\rangle = a \cdot |0\rangle + b \cdot |1\rangle$ to $|\theta\rangle$



Adiabatic Quantum Computation (AQC)

AQC is based on Schrödinger equation

$$i\hbar \frac{d|\Psi(t)\rangle}{dt} = H|\Psi(t)\rangle \quad (6)$$

and the adiabatic theorem for calculations. i is the imaginary unit, $\hbar$ is the reduced Planck constant, $|\Psi(t)\rangle$ is the quantum system state at time t , and H is the Hamiltonian operator characterizing the total quantum system energy. In an n qubit system H is a $2^n \times 2^n$ hermitian matrix. The eigenvalues $\{E_i\}$, $i=0,1,2,\dots,n-1$ of H represent energies. Eigenvector(s) corresponding to the lowest eigenvalue represent *ground state(s)*.

In an n qubit system, the unknown ground state a Hamiltonian H_{final} represents a problem solution. The system is initialized to the ground state of a chosen Hamiltonian H_{init} , which is adiabatically evolved to H_{final} by applying

$$H(t) = A(t) \cdot H_{\text{init}} + B(t) \cdot H_{\text{final}} \quad (7)$$

during the time t , $0 \leq t \leq T$. The adiabatic theorem: the system remains in the ground state, if T is sufficiently long and the difference between the two smallest eigenvalues of $H(t/T)$ is strictly >0 when $0 \leq t \leq T$. The terminal system state represents the problem solution. (Farhi, Goldstone, Gutmann, & Sipser, 2000)

Quantum Algorithms

A quantum circuit computation algorithm: qubit state changes in quantum gates and qubit state measurements. Any classical algorithm can be realized in a quantum circuit based computer. Some quantum algorithms have lower computational complexity than corresponding classical algorithms. Best known classical algorithm to disclose a private RSA key has superpolynomial complexity. Shor's factorization quantum algo-

rithm (Shor, 1994) discloses a n bit private RSA key with polynomial (cubic) complexity in a $1.5 \times n$ qubit quantum computer. Best classical algorithms for unsorted database search and solving linear equation sets have linear complexity. In a quantum computer with sufficiently many qubits, unsorted size n search has $O(\sqrt{n})$ complexity (Grover, 1996) and solving linear equation sets has logarithmic complexity (Cai et al., 2013).

Quantum annealing is an AQC algorithm for minimizing a given object function represented by a Hamiltonian, whose ground state represents the minimum. Such object function:

$$E(\mathbf{s}) = -\sum_{i=1}^N h_i \cdot s_i + \sum_{i<j} K_{ij} \cdot s_i \cdot s_j \quad (8)$$

Binary values in $\mathbf{s} = [s_1, s_2, \dots, s_N]$ should be chosen to minimize $E(\mathbf{s})$. h_i , K_{ij} are dimensionless real numbers. No classical algorithm with polynomial complexity is known for finding the minimal of 2^N possible combinations. Faster minimization is expected from quantum annealing. The minimum is the ground state of the Hamiltonian

$$H_F = -\sum_{i=1}^N h_i \cdot \sigma_z^i + \sum_{i<j} K_{ij} \cdot \sigma_z^i \cdot \sigma_z^j \quad (9)$$

in a N qubit system, where σ_z^i , σ_z^j are Pauli z matrices $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ applied to qubits i, j . To find the ground state of H_F , the Hamiltonian

$$H(t) = A(t) \cdot H_I + B(t) \cdot H_F \quad (10)$$

is adiabatically evolved from $t=0$ to T for $A(0)/B(0) \gg 1$, $A(T)/B(T) \ll 1$. The Hamiltonian

$$H_I = -\sum_{i=1}^N \sigma_x^i \quad (11)$$

with $\sigma_x^i = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ applied to qubits i is in ground

state (all qubits in state $\frac{1}{\sqrt{2}} \cdot (|0\rangle + |1\rangle)$). The functions $A(t)$, $B(t)$ must guarantee adiabatic evolution from a state very close to the ground state of H_1 to a state very close to the ground state of H_F . (Harris et al., 2009)

Quantum Programming

Several imperative and functional programming languages, usually based on the Quantum Random Access Machine where a classical device controls quantum device computations, and also some different approaches have been proposed for quantum circuit based computers (Sofge, 2008; Miszczak, 2011).

Among the first imperative languages is QCL (Quantum Computing Language) with a syntax resembling C language: The basic data type *qreg* (quantum register) defines qubit arrays. The standard library provides quantum gate operators. The *measure* statement measures qubit states. Higher abstraction level algorithms are defined like C language functions. Other imperative languages are LanQ resembling QCL and Q (Bettelli, Calarco, & Serafini, 2003) based on C++. (Miszczak, 2011)

Lambda calculus and Haskell based functional quantum programming languages have been proposed. The first proposal, lambda-q calculus (Maymin, 1997) which expresses any quantum algorithm and efficiently solves problems with NP-hard complexity, is unfortunately

more expressive than is possible to realize as a physical quantum computer. Another quantum lambda calculus proposal is defined for quantum computations with no measurements (van Tonder, 2004). Altenkirch and Grattage (2005) propose a Haskell-like quantum language, QML, expressing both classical and quantum control and data. Quipper (Green, Lumsdaine, Ross, Selinger, & Valiron, 2013) is a recent embedded functional language providing Haskell libraries for quantum programs. An introduction to AQC programming is published in (Quantum, 2014).

Physical Realization of Quantum Computers

Properties restricting quantum circuit computation realizations (see Table 1):

- **Quantum Operation Time:** Qubit state change time in a basic quantum gate.
- **Decoherence Time**
- **Scalability:** Maximal qubit set in a quantum circuit.

The maximal number of quantum program operations is (decoherence time)/(quantum operation time).

Currently only small scale quantum circuit computation realizations exist. A basic building block of a quantum computer, a two qubit quantum gate, has been implemented in silicon (Veldhorst et al., 2015). Shor's quantum factorization has been executed in a 7 qubit NMR implemented

Table 1. Quantum circuit implementations properties

Technology	Quantum Operation Time	Decoherence Time	# Quantum Operations	Scalability (# Qubits)
Ion traps	10^{-7} s	10^{-1} s	10^6	50
Cavity QED	10^{-14} s	10^{-5} s	10^9	2-5
NMR	10^{-3} s	10^4 s	10^7	10-50
Quantum Dots	10^{-9} s	10^{-6} s	10^3	1000

Source: Brown, 2001

computer (Vandersypen et al., 2001). Monz et al., (2011) describe an implemented 14 qubit register.

AQC technology is magnetically coupled Josephson junction rf-SQUID (Radio Frequency Superconducting Quantum Interference Device) flux qubits (Devoret & Schoelkopf, 2013; Harris et al., 2009). Boixo et al. (2013) report experimental quantum annealing with 108 qubits in a 128 qubit adiabatic quantum computer. A 512 qubit adiabatic quantum computer has been delivered for artificial intelligence research (Jones, 2013). Also a 1000 qubit adiabatic quantum computer, D-Wave 2X (D-Wave, 2015), is available.

QUANTUM COMMUNICATION

Quantum communication is qubit transfer and use of entanglement and teleportation to transfer qubit states between physical locations.

Qubit transfer is photon transfer, for example polarized photons, in optical fibers or along free sight paths. Current transmission range is limited by photon state decoherence to about 300 km in optical fibers (Korzh et al., 2015) and about 150 km in free sight atmosphere paths (Schmitt-Manderbach et al., 2007). The range in free sight space is considerably longer because of negligible decoherence. Secure quantum code transmission between two ground stations via satellite has been demonstrated (Physicists, 2013).

Current quantum communication applications are Quantum Key Distribution (QKD), Quantum Key Agreement, Quantum Secret Sharing (QSS), and Quantum Signatures.

Quantum Repeaters

Quantum networks over long distances are confronted with loss and operation errors, which heralded entanglement generation, purification, and deterministic quantum error correction can suppress. Used correction methods classify quantum repeaters into three generations (Munro, 2015). The first generation uses entanglement generation

and purification. Communication rate overhead is reduced from exponential to polynomial scaling with distance. The second generation replaces entanglement purification with quantum error correction, which further reduces communication rate overhead from polynomial to poly-logarithmical scaling with distance. A third generation will eliminate two-way communication with both loss and operation error correction. Then only local quantum gate speed limits communication rate.

Current optical communication technology transfers polarized photons with reasonable error rates up to about 70...100 km (Stucki, Gisin, Guinnard, Ribordy, & Zbinden, 2002; MagiQ, 2007). In long-distance quantum communication, quantum repeaters are needed on the network's physical layer. Repeaters combine entanglement swapping and quantum memories. Transmission distance is divided into segments, entanglement is distributed independently within each segment, and entanglement swapping extends entanglement distance. Segments are mutually independent. Entanglement distribution is heralded, i.e. a signal indicates successful entanglement. Entanglement is stored in quantum memories for avoiding concurrent entanglement within each segment. (de Riedmatten, 2015)

Research efforts to implement quantum repeaters are reported in (Curcic et al., 2004; QuReP, 2014).

Quantum Key Distribution (QKD)

This subchapter surveys QKD protocols and their integration in the TCP/IP network protocol stack, presents current QKD technology and applications, and describes security threats.

QKD Protocols

BB84 (Bennet & Brassard, 1984):

1. Alice sends over an optical channel to Bob a sequence of photons randomly polarized to a state in $\{|horis\rangle, |vert\rangle, |+45^\circ\rangle, |-45^\circ\rangle\}$.

Bob guesses the polarization base (horizontal/vertical or diagonal) for each photon and measures polarization.

2. Bob sends over a classical channel the guess sequence to Alice.
3. Alice replies about guess correctness.

Bob's measurements for correct guesses represent a shared secret key. See Figure 2.

If eavesdropper Eve measures photon polarization in Alice's sequence in a guessed base and forwards measured photons to Bob, then a random choice of $|0\rangle$, $|1\rangle$ is forwarded for each incorrect guess. Thus successful key creation proves absence of eavesdropping (Yanofsky & Mannucci, 2008).

BB84 is actually a "Sifting Phase" creating raw keys excluding not received photons. In a practical BB84 the shared secret contains the raw key bits remaining after "Key Distillation" including "Error Correction" to remove bits different in both raw keys and "Privacy Amplification" to drop correct bits known by a possible eavesdropper. (Elliott et al., 2003)

BB84 is a 4-state protocol as it uses four different polarization states. However, a perfectly secret QKD protocol needs only 2 non-orthogonal polarization states (Yanofsky & Mannucci, 2008). Also perfectly secret 6-state protocols have been proposed and analysed (Bruss, 1998). EPR, proposed by Einstein, Podolsk, and Rosen, is a completely different QKD protocol type based on Bell's inequality (Yanofsky & Mannucci, 2008).

QKD Based Session Keys

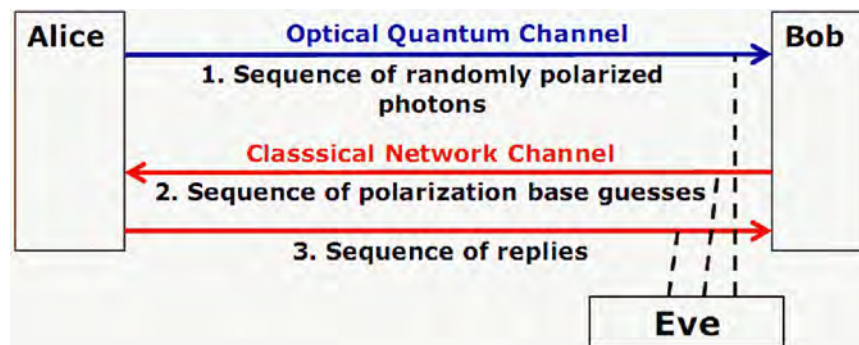
Elliott & al. (2003) describe QKD integration in IPSec. QKD integration in TLS/SSL requires new cipher suites defining key exchange. New ServerKeyExchange and ClientKeyExchange messages are needed.

Attacks Against QKD

Practical QKD security depends on environment noise, imperfect photon detection, and light sources properties. Some attacks are possible in ideal environments with single photon generators and 100% efficient detectors (Aggarwal, Sharma, & Gupta, 2011), while others stem from unsecure implementations (Gisin, Ribordy, Tittel, & Zbinden, 2002). Khan and Xu (2011) divide QKD communication into secret key generation, secret key storage and management, and encryption and transmission. Attack types are physical, quantum, and classical.

Quantum attacks on QKD are intercept-resend, Trojan horse, and photon number splitting (PNS) (Kahn & Xu, 2011). In *intercept-resend*, Eve guesses polarization base, measures the quantum states sent by Alice, and sends replacement states to Bob. In a *Trojan horse* attack, Eve occupies a part of the quantum channel (i.e. the spatial, temporal, and frequency modes) to probe Alice's apparatus. An auxiliary source is used and a detector analyses the backscattered signal. If Alice

Figure 2. The BB84 QKD protocol



transmits a multiple photon laser pulse, then Eve could split off photons with *PNS*. A QKD protocol, *SARG04*, using the same polarization states as BB84 with a different information encoding is less vulnerable to PNS than BB84 (Scarani, Acín, Ribordy, & Gisin, 2004).

Optical loophole attacks are strong light pulses (Bethune & Risk, 2000), high-power destruction of optical components, utilization of light emission from avalanche photodiodes, and faked states.

QKD Technology and Applications

A QKD channel is a free sight path or an optical fiber. Free sight transmission is faster. In free sight atmosphere paths distances exceed 140 km, but exchange rate is only 12 bit/s (Schmitt-Manderbach et al., 2007). Current QKD systems exchange 1 Mbit/s in 20 km fiber and 10 kbit/s when distance > 100 km (Dixon, Yuan, Dynes, Sharpe, & Shields, 2008). Modest exchange rates depend on limited single photon detection rate.

The first commercial quantum cryptography products, i.e. single photon detectors and random number generators, appeared in 2002. Encoding used photon phase.

The Swiss company ID Quantique offers for the QKD protocols BB84 and SARG04

- A QKD research platform (Clavis, 2014)
- A server combining QKD with AES (Advanced Encryption Standard), securely bridging two Fast Ethernet networks, and supporting refresh rates 1 key/minute for up to 12 encryptors and 100 km transmission distance in optical fibre (Cerberis, 2012).

QPN™ Security Gateway (MagiQ, 2007) available from the US company MagiQ Technologies, Inc. adds VPN security layers and classical data encryption to QKD. BB84, IPSec, and 256 bit AES are implemented. A refresh rate up to 100 keys/second, a transmission distance of 100 km,

and full-duplex 10/100 Ethernet ports are supported. Available is also a single photon BB84 based system (Q-BOX, 2003) for QKD research.

A QKD module with key transfer 10 kbit/s for 20 km and 100 bit/s for 80 km is available for QKD research from the French company SecCureNet (Cygnus, 2013).

Experimental QKD networks have been realized in USA, Europe, Japan, China, and Korea. The limited QKD range caused by photon state decoherence is extended by multi-hop networking with trusted nodes and QKD combined with the absolutely secure OTP (One Time Pad) encryption (Bellovin, 2011). China has in autumn 2015 almost completed installation of the world's hitherto largest quantum communications network, 2000 kilometers from Beijing to Shanghai. (Elliot, 2004; Poppe et al., 2008; Gisin, 2015; Fadilpašić, 2015)

Quantum Key Agreement

Subramaniam and Parakh (2014) propose a quantum Diffie-Hellman protocol. Two parties, Alice and Bob, agree on a basis set to use, the number m of qubits to exchange, and the initial qubit state $|\psi\rangle = |0\rangle$. Every basis consists of two rotation transformations, R for bit value 0 and $R+90^\circ$ for value 1, where $R=0^\circ$ or 30° or 42° etc. Then Alice and Bob independently and randomly choose m bases from the agreed set, generate a sequence of m random bits, and send a m qubit sequence to each other.

Each sent qubit is a rotation of $|\psi\rangle$ determined by the chosen basis and the generated bit. Alice and Bob then:

- Apply to each received qubit a rotation determined by the chosen basis and the generated bit,
- Apply the rotation 90° -(previous rotation angle),
- Record the bit from measuring the rotated qubit in the chosen basis.

Both Alice and Bob thus obtain an m bit sequence from received qubits and can now announce their chosen basis sequences to each other over a public channel. A bit in the obtained sequence is discarded, if Alice and Bob use a different basis. The resulting shared secret key is a joint random selection from the remaining bits. Third party interference in qubit transmission causes bit errors.

Quantum Secret Sharing (QSS)

The first QSS used the GHZ entanglement state (Hillery, Buzek, & Berthaume, 1999). QSS protocols have two types defined by the entanglement and product states. Zhang and Man (2005) propose a protocol based on entanglement swapping and local unitary operation. Tan, Feng, Jiang, and Fang (2013) present a verifiable quantum secret sharing protocol based on entanglement swapping. External eavesdroppers, internal cheaters, and cheating by two cooperating dishonest participants are prevented.

Information-Flow Security of Quantum Systems

Entanglement can cause Information leakage. A Trojan horse may exploit entanglement between itself and a user with sensitive information as a covert channel (Ying, Feng, & Yu, 2013). The noninterference formalism is a general framework for specifying and analyzing information-flow security.

Quantum Digital Signature

Classical digital signatures can be created with any one-way function (Lamport, 1979). A secure quantum one-way function based scheme for signing classical bit strings is proposed in (Gottesman & Chuang, 2001), patented in USA (Chuang & Gottesmann, 2002), and experimentally demonstrated in (Clarke et al., 2012). Lu and Feng

(2005) propose another quantum one-way function based scheme for signing general quantum states. Quantum message authentication can be based on quantum signature schemes (Barnum, Crepeau, Gottesman, Smith, & Tapp, 2002).

QUANTUM ERROR CORRECTION

Quantum state decoherence, quantum noise, faults in quantum gates and qubit memories, and faulty qubit state measurements cause errors. Quantum error correction methods, codes, and algorithms are therefore required. (Lidar & Brun, 2013)

Nielsen and Chuang (2002) describe quantum circuits for single qubit flip error correction. Reed et al. (2012) present superconducting circuits for three-qubit quantum error correction. A superconducting quantum computing chip, a 4 qubit square lattice, has been designed and successfully used to detect arbitrary single qubit errors (Córcoles et al., 2015).

FUTURE RESEARCH DIRECTIONS

Research aiming to build a large scale quantum circuit based computer will continue. Possible solutions are a multiple of two qubit gates in silicon (Veldhorst et al., 2015) and superconducting qubits in a two-dimensional grid (Córcoles et al., 2015; Simonite, 2015).

Future quantum computers will probably combine a classical and a quantum device. Computations are decomposed into a controlled set of subtasks. The quantum device executes subtasks, for which efficient quantum algorithms can be programmed, and some control, the classical device executes other subtasks and control. (Nielsen & Chuang, 2002)

Additional polynomial quantum algorithms are being investigated to solve so called hidden subgroup problems, to which integer factoriza-

tion belongs. New and more powerful quantum programming languages, mainly functional, will be developed. (Sofge, 2008)

A new technique using twisted light will increase the efficiency of quantum communication. Polarization encodes only one bit per photon but the new technique, Orbital Angular Momentum, encodes 2.05 bits per photon and may still be extended. (Mirhosseini et al., 2015)

Lack of fast photon detectors limits QKD performance. New detectors based on cryogenic niobium nitride superconductors may be a solution (Quellette, 2004). Research on ideal single-photon sources continues, since multiple photon emissions are vulnerable. Expected future QKD advances (Hogben, 2009):

- Increased QKD *distance* with quantum repeaters
- Increased transmission rate
- Decreased QKD *costs*
- Perfectly random OTP keys from quantum state measurements in *Quantum Random Number Generation*.

CONCLUSION

Future quantum computing can erode security of current public key cryptography. The vulnerability of RSA based cryptography to quantum integer factorization is evident. Also current Elliptic Curve Cryptography is vulnerable to quantum algorithm attacks (Maslov, Mathew, Cheung, & Pradhan, 2009). Quantum resistant public key cryptography is thus a future need. Proposed candidates are Lattice Based Public-Key Cryptography, Multivariate Public-Key Cryptography, Code-Based Public-Key Cryptography, Hash-Based Signatures, McEliece Cryptosystems and a revised Elliptic Curve Cryptography (Buchmann, 2015; Zhang & Xu, 2014; Perlner & Cooper, 2009; Jao & de Feo, 2011). Also a quantum resistant SSL/TLS protocol has been proposed (Chang, Chen, Wu, & Yang, 2014).

Future quantum computing also affects symmetric cryptosystems. Grover's quantum algorithm halves brute force search bit length. For example, quantum attack resistance of 256 bit AES is 128 bit.

Quantum cryptography is an emerging information security technology, which is presently already available but still neither fully integrated in current network technology nor standardized. Technological limitations regarding quantum transmission distance and quantum device technology need to be removed.

REFERENCES

- Aggarwal, R., Sharma, H., & Gupta, D. (2011). Analysis of Various Attacks over BB84 Quantum Key Distribution Protocol. *International Journal of Computers and Applications*, 20(3), 28–31. doi:10.5120/2454-3313
- Aharonov, D., van Dam, W., Kempe, J., Landau, Z., Lloyd, S., & Regev, O. (2004). Adiabatic Quantum Computation is Equivalent to Standard Quantum Computation. In *Proceedings of the Annual IEEE Symposium on Foundations of Computer Science* (vol. 45, pp. 42-51). IEEE Press. doi:10.1109/FOCS.2004.8
- Altenkirch, T., & Grattage, J. (2005). A Functional Quantum Programming Language. In *Proceedings of the Annual IEEE Symposium on Logic in Computer Science* (vol. 20, pp. 249-259). IEEE Press. doi:10.1109/LICS.2005.1
- Barnum, H., Crepeau, C., Gottesman, D., Smith, A., & Tapp, A. (2002). Authentication of Quantum Messages. In *Proceedings of the IEEE Symposium on Foundations of Computer Science* (vol. 43, pp. 449-458). IEEE Press. doi:10.1109/SFCS.2002.1181969
- Bellovin, S. M. (2011). Frank Miller: Inventor of the One-Time Pad. *Cryptologia*, 35(3), 203–222. doi:10.1080/01611194.2011.583711

- Bennett, C. H., & Brassard, G. (1984). Quantum Cryptography: Public Key Distribution and Coin Tossing. *Proceedings of IEEE International Conference on Computers, Systems & Signal Processing*, 175, 175–179.
- Bethune, D. S., & Risk, W. P. (2000). An auto-compensating fiber-optic quantum cryptography system based on polarization splitting of light. *IEEE Journal of Quantum Electronics*, 36(3), 340–347. doi:10.1109/3.825881
- Bettelli, S., Calarco, T., & Serafini, L. (2003). Toward an Architecture for Quantum Programming. *The European Physical Journal D*, 25(2), 181–200. doi:10.1140/epjd/e2003-00242-2
- Boixo, S., Rønnow, T. F., Isakov, S. V., Wang, Z., Wecker, D., Lidar, D. A., ... Troyer, M. (2013). *Quantum Annealing with More Than One Hundred Qubits*. Retrieved October 23, 2015, from <http://arxiv.org/abs/1304.4595>
- Brown, J. (2001). *Quest for the Quantum Computer*. Simon & Schuster.
- Bruss, D. (1998). Optimal eavesdropping in quantum cryptography with six states. *Physical Review Letters*, 81(14), 3018–3021. doi:10.1103/PhysRevLett.81.3018
- Buchmann, J. A. (2015). Post-Quantum Cryptography – an Overview. *Proceedings of 5th International Conference on Quantum Cryptography QCrypt2015*.
- Cai, X.-D., Weedbrook, C., Su, M.-C., Gu, M., Zhu, M.-J., Li, L., ... Pan, J.-W. (2013). *Experimental Quantum Computing to Solve Systems of Linear Equations*. Retrieved October 23, 2015, from <http://arxiv.org/abs/1302.4310>
- Cerberis Quantum Key Distribution (QKD) Server. (2012). ID Quantique SA. Retrieved October 19, 2015, from <http://www.idquantique.com/quantum-safe-crypto/qkd-server/>
- Chang, Y.-A., Chen, M.-S., Wu, J.-S., & Yang, B.-Y. (2014). Postquantum SSL/TSL for embedded systems. In *Proceedings of the 7th International Conference on Service-Oriented Computing and Applications* (pp. 266–270). IEEE Press.
- Chuang, I., & Gottesmann, D. (2002). *Quantum Digital Signatures*. US Patent US 2002/0199108 A1
- Clarke, P. J., Collins, R. J., Dunjko, V., Andersson, E., Jeffers, J., & Buller, G. S. (2012). Experimental Demonstration of Quantum Digital Signatures Using Phase-Encoded Coherent States of Light. *Nature Communications*, 3(1174). PMID:23132024
- Clavis the most versatile quantum key distribution research platform. (2014). ID Quantique SA. Retrieved October 19, 2015 from <http://www.idquantique.com/photon-counting/clavis-qkd-platform>
- Córcóles, A. D., Magesan, E., Srinivasan, S. J., Cross, A. W., Steffen, M., Gambetta, J. M., & Chow, J. M. (2015). Demonstration of a quantum error detection code using a square lattice of four superconducting qubits. *Nature Communications*, 25(4). doi: 10.1038/ncomms7979
- Curcic, T., Filipkowski, M. E., Chtchelkanova, A., D'Ambrosio, P. A., Wolf, S. A., Foster, M., & Cochran, D. (2004). Quantum Networks: From Quantum Cryptography to Quantum Architecture. *ACM SIGCOMM Computer Communications Review*, 34(5), 3–8. doi:10.1145/1039111.1039117
- Cygnus State-of-the-art Continuous-Variable Quantum Key Distribution Module. (2013). SecureNet. Retrieved October 26, 2015, from http://www.secrenet.com/datasheets/datasheet_cygnus.pdf
- D-Wave, The Quantum Computing Company™. (2015). *The D-Wave 2X™ Quantum Computer Technology Overview*. Retrieved October 7, 2015 from http://www.dwavesys.com/sites/default/files/D-Wave%202X%20Tech%20Collateral_0915F.pdf

- Das, A., & Chakrabarti, B. K. (2008). Quantum Annealing and Analog Quantum Computation. *Reviews of Modern Physics*, 80(3), 1061–1081. doi:10.1103/RevModPhys.80.1061
- de Riedmatten, H. (2015). *Quantum Repeaters. QCrypt2015 Tutorial*. Retrieved November 10, 2015, from http://2015.qcrypt.net/wp-content/uploads/2015/09/Tutorial3_Hugues-de-Riedmatten.pdf
- Deutsch, D. (1985). Quantum theory, the Church-Turing principle and the universal quantum computer. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, A400(1818), 97–117. doi:10.1098/rspa.1985.0070
- Devoret, M. H., & Schoelkopf, R. J. (2013). Superconducting Circuits for Quantum Information: An Outlook. *Science*, 339(6124), 1169–1174. doi:10.1126/science.1231930 PMID:23471399
- Dixon, A. R., Yuan, Z. L., Dynes, J. F., Sharpe, A. W., & Shields, A. J. (2008). Gigahertz decoy quantum key distribution with 1 Mbit/s secure key rate. *Optics Express*, 16(23), 18790–18799. doi:10.1364/OE.16.018790 PMID:19581967
- Elliot, C. (2004). Quantum Cryptography. *IEEE Security and Privacy*, 2(4), 57–61. doi:10.1109/MSP.2004.54
- Elliott, C. D., Pearson, D., & Troxel, G. (2003). Quantum Cryptography in Practice. *Proceedings of SIGCOMM*, 03, 227–238.
- Fadilpašić, S. (2015). *China's quantum communications network almost ready*. ITProPortal. Retrieved October 21, 2015 from <http://www.itproportal.com/2015/08/31/chinas-quantum-communications-network-almost-ready/>
- Farhi, E., Goldstone, J., Gutmann, S., & Sipser, M. (2000). *Quantum Computation by Adiabatic Evolution*. Retrieved October 23, 2015, from <http://arxiv.org/pdf/quant-ph/0001106v1.pdf>
- Feynman, R. (1982). Simulating physics with computers. *International Journal of Theoretical Physics*, 21(6&7), 467–488. doi:10.1007/BF02650179
- ...from Quantum Information to Biophysics. (2011). Faculty of Mathematics and Physical Sciences. University of Leeds. Retrieved October 23, 2015, from <http://www.qi.leeds.ac.uk/>
- Gisin, N. (2015). Quantum Cryptography: where do we stand? *Proceedings of 5th International Conference on Quantum Cryptography QCrypt2015*.
- Gisin, N., Ribordy, G., Tittel, W., & Zbinden, H. (2002). Quantum cryptography. *Reviews of Modern Physics*, 74(1), 145–195. doi:10.1103/RevModPhys.74.145
- Gottesman, D., & Chuang, I. (2001). *Quantum Digital Signatures*. Retrieved October 23, 2015, from <http://arxiv.org/abs/quant-ph/0105032>
- Green, A. S., Lumsdaine, P. L., Ross, N. J., Selinger, P., & Valiron, B. (2013). Quipper: A Scalable Quantum Programming Language. *ACM SIGPLAN Notices*, 48(6), 333–342. doi:10.1145/2499370.2462177
- Grover, L. K. (1996). A fast quantum mechanical algorithm for database search. In *Proceedings of the Annual ACM Symposium on the Theory of Computing* (vol. 28, pp. 212–219). New York: ACM Press. doi:10.1145/237814.237866
- Harris, R., Berkley, A. J., Johansson, J., Johnson, M. W., Lanting, T., Bunyk, P.,... Han, S. (2009). *Implementation of a Quantum Annealing Algorithm Using a Superconducting Circuit*. Retrieved October 23, 2015, from <http://arxiv.org/pdf/0903.3906v1.pdf>
- Hillery, M., Buzek, V., & Berthaume, A. (1999). Quantum Secret Sharing. *Physical Review A*, 59(3), 1829–1834. doi:10.1103/PhysRevA.59.1829

- Hogben, G. (2009). *Briefing: Quantum Key Distribution*. ENISA – European Network and Information Security Agency. Retrieved October 23, 2015, from <https://www.enisa.europa.eu/publications/archive/briefing-quantum-key-distribution>
- Jao, D., & de Feo, L. (2011). Towards Quantum-Resistant Cryptosystems from Supersingular Elliptic Curve Isogenies. In *Proceedings of the 4th International Conference on Post-Quantum Cryptography PQCrypto'11* (vol. 4, pp. 19-34). Berlin: Springer-Verlag doi:10.1007/978-3-642-25405-5_2
- Jones, N. (2013, May 16). Google and NASA snap up quantum computer D-Wave machine to work on artificial-intelligence problems. *Nature*. Retrieved October 23, 2015, from <http://www.nature.com/news/google-and-nasa-snap-up-quantum-computer-1.12999>
- Kahn, M. M., & Xu, L. (2011). A Taxonomy of Attacks on Quantum Key Distribution. *International Journal of Latest Trends in Computing*, 2(3), 358–363.
- Korzh, B., Lim, C. C. W., Houlmann, R., Gisin, N., Li, M. J., Nolan, D., & Zbinden, H. et al. (2015). Provably secure and practical QKD over 307 km of optical fibre. *Nature Photonics*, 9, 163–168. doi:10.1038/nphoton.2014.327
- Lamport, L. (1979 October). *Constructing digital signatures from a one-way function*. Technical Report CSL-98, SRI International.
- Lidar, D. A., & Brun, T. A. (Eds.). (2013). *Quantum Error Correction*. Cambridge University Press. doi:10.1017/CBO9781139034807
- Lu, X., & Feng, D. (2005). Quantum digital signatures based on quantum one-way functions. *Proceedings of the International Conference on Advanced Communication Technology ICACT*, 7, 514-517.
- MagiQ QPN 8505 Security Gateway Uncompromising VPN Security. (2007). MagiQ Technologies, Inc. Retrieved October 23, 2015, from http://www.magiqtech.com/MagiQ/Products_files/8505_Data_Sheet.pdf
- Maslov, D., Mathew, J., Cheung, D., & Pradhan, D. (2009). An $O(m^2)$ -Depth Quantum Algorithm for the Elliptic Curve Discrete Logarithm Problem over $GF(2^m)$. *Journal of Quantum Information & Computation*, 9(7), 610–627.
- mathQI Research group on mathematics and quantum information. (2015). Facultad de Ciencias Matemáticas, Universidad Complutense de Madrid. Retrieved November 2, 2015, from <http://www.mathqi.es/>
- Maymin, P. (1997). *The Lambda-q Calculus Can Efficiently Simulate Quantum Computers*. Retrieved October 23, 2015, from <http://arxiv.org/pdf/quant-ph/9702057v1.pdf>
- Mirhosseini, M., Magaña-Loaiza, O., Sullivan, M., Rodenburg, B., Malik, M., Lavery, M., & Boyd, R. et al. (2015). High-dimensional quantum cryptography with twisted light. *New Journal of Physics*, 17.
- Miszczak, J. A. (2011). Model of Quantum Computation and Quantum Programming Languages. *Bulletin of the Polish Academy of Sciences Technical Sciences*, 59(3), 305–324. doi:10.2478/v10175-011-0039-5
- Monz, T., Schindler, P., Barreiro, J. T., Chwalla, M., Nigg, D., William A. Coish, W. A., & Blatt, R. et al. (2011). 14-Qubit Entanglement: Creation and Coherence. *Physical Review Letters*, 106(13), 130506. doi:10.1103/PhysRevLett.106.130506 PMID:21517367
- Munro, B. (2015). *Quantum Repeaters: From the first generation to the third?* Retrieved November 10, 2015, from <http://wqrn.pratt.duke.edu/presentations>

- Nielsen, M. A., & Chuang, I. L. (2002). *Quantum Computation and Quantum Information*. Cambridge University Press.
- Perlner, R. A., & Cooper, D. A. (2009). Quantum Resistant Public Key Cryptography: A Survey. *Proceedings of 8th Symposium on Identity and Trust on the Internet IDTrust*, 8, 85-93. doi:10.1145/1527017.1527028
- Physicists Successfully Transmit Quantum Code Through the Atmosphere. (2013). SciTechDaily. Retrieved October 23, 2015, from <http://scitechdaily.com/physicists-successfully-transmit-quantum-code-through-the-atmosphere>
- Poppe, A., Momtchil, P., & Maurhart, O. (2008 April). Outline of the SECOQC Quantum-Key-Distribution Network in Vienna. *International Journal of Quantum Information*, 209-218.
- Q-BOX WORKBENCH Uncompromising QKD Research. (2003). MagicQ Technologies, Inc. Retrieved October 23, 2015, from http://www.magicqtech.com/Products_files/QBox%20Data-sheet-2011.pdf
- Quantum Computing Group. (2015a). The Laboratory for Physical Sciences, University of Maryland. Retrieved November 2, 2015, from <http://www.lps.umd.edu/Quantum%20Computing%20Group/QuantumComputingIndex.html>
- Quantum Computing Primer. (2014). D-Wave Systems Inc. Retrieved October 23, 2015 from <http://www.dwavesys.com/tutorials/background-reading-series/quantum-computing-primer>
- Quantum Information Laboratory. (2015b). M.V.Lomonosov Moscow State University. Retrieved October 23, 2015, from <http://qilab.phys.msu.ru/>
- qis.mit.edu quantum information science @ mit. (2013). Massachusetts Institute of Technology. Retrieved October 23, 2015, from <http://qis.mit.edu/>
- Quellette, J. (2004 December). Quantum Key Distribution. *The Industrial Physicist*, 22-25.
- QuReP Quantum Repeaters for Long Distance Fibre-Based Quantum Communication. (2014). Cordis Community Research and Development Information Service, European Commission. Retrieved October 23, 2015, from http://cordis.europa.eu/project/rcn/93257_en.html
- Raussendorf, R., & Briegel, H. J. (2001). A One-Way Quantum Computer. *Physical Review Letters*, 86(22), 5188–5191. doi:10.1103/PhysRevLett.86.5188 PMID:11384453
- Reed, M. D., DiCarlo, L., Nigg, S. E., Sun, L., Frunzio, L., Girvin, S. M., & Schoelkopf, R. J. (2012). Realization of Three-Qubit Quantum Error Correction with Superconducting Circuits. *Nature*, 482(7385), 382–385. doi:10.1038/nature10786 PMID:22297844
- Sarma, S. D., Freedman, M., & Nayak, C. (2006). Topological Quantum Computation. *Physics Today*, 59(July), 32–38. doi:10.1063/1.2337825
- Scarani, V., Acín, A., Ribordy, G., & Gisin, N. (2004). Quantum Cryptography Protocols Robust against Photon Number Splitting Attacks for Weak Laser Pulse Implementations. *Physical Review Letters*, 92(5), 057901. doi:10.1103/PhysRevLett.92.057901 PMID:14995344
- Schmitt-Manderbach, T., Weier, H., Fürst, M., Ursin, R., Tiefenbacher, F., Scheidl, T., & Weinfurter, H. et al. (2007). Experimental demonstration of free-space decoy-state quantum key distribution over 144 km. *Physical Review Letters*, 98(1), 010504. doi:10.1103/PhysRevLett.98.010504 PMID:17358463
- Shor, P. W. (1994). Algorithms for quantum computation: Discrete log and factoring. *Proceedings of the 35th Annual Symposium on Foundations of Computer Science*, 124–134. doi:10.1109/SFCS.1994.365700
- Simonite, T. (2015, April 29). IBM Shows off a Quantum Computing Chip. *MIT Technology Review*.

Sofge, D. A. (2008) A Survey of Quantum Programming Languages: History, Methods, and Tools. In *Proceedings of the Second International Conference on Quantum, Nano, and Micro Technologies ICQNM 2008* (pp. 66-71). IEEE Press. doi:10.1109/ICQNM.2008.15

Stucki, D., Gisin, N., Guinnard, O., Ribordy, G., & Zbinden, H. (2002). Quantum Key Distribution over 67 km with a PlugPlay System. *New Journal of Physics*, 4(41), 1–8.

Subramaniam, P., & Parakh, A. (2014). A Quantum Diffie-Hellman Protocol. In *Proceedings of the 11th International Conference on Mobile Ad Hoc and Sensor Systems* (pp. 523-524). IEEE Press.

Tan, X., Feng, Z., Jiang, L., & Fang, A. (2013). Verifiable Quantum Secret Sharing Protocol. *Proceedings of Fourth International Conference on Emerging Intelligent Data and Web Technologies (EIDWT)*. doi:10.1109/EIDWT.2013.44

Timeline of Quantum Computing. (2015). *Wikipedia*. Retrieved October 23, 2015, from http://en.wikipedia.org/wiki/Timeline_of_quantum_computing

van Tonder, A. (2004). A lambda calculus for quantum computation. *SIAM Journal on Computing*, 33(5), 1109–1135. doi:10.1137/S0097539703432165

Vandersypen, L., Steffen, M., Breyta, G., Yannoni, C., Sherwood, M. H., & Chuang, I. L. (2001). Experimental realization of Shors quantum factoring algorithm using magnetic resonance. *Nature*, 414(6866), 883–887. doi:10.1038/414883a PMID:11780055

Veldhorst, M., Yang, C. H., Hwang, J. C. C., Huang, W., Dehollain, J. P., Muhonen, J. T., & Dzurak, A. S. et al. A. S. (2015). A two-qubit logic gate in silicon. *Nature*. doi:10.1038/nature15263

Yanofsky, N. S., & Mannucci, M. A. (2008). *Quantum Computing for Computer Scientists*. Cambridge University Press. doi:10.1017/CBO9780511813887

Ying, M., Feng, Y., & Yu, N. (2015). *Quantum Information-Flow Security: Noninterference and Access Control*. Retrieved October 23, 2015, from <http://arxiv.org/pdf/1301.6804v1.pdf>

Zhang, X., & Xu, C. (2014). Efficient Identity-based Public Auditing Scheme for Cloud Storage from Lattice Assumption. In *Proceedings of the 17th International Conference on Computational Science and Engineering* (pp. 1819-1826). IEEE Press. doi:10.1109/CSE.2014.334

Zhang, Z.-J., & Man, Z.-X. (2005). Multiparty quantum secret sharing of classical messages based on entanglement swapping. *Physical Review A*, 72(2), 022303. doi:10.1103/PhysRevA.72.022303

ADDITIONAL READING

Bergou, J. A., & Hillery, M. (2013). *Introduction to the Theory of Quantum Information Processing. eBook*. Germany: Springer. doi:10.1007/978-1-4614-7092-2

de Falco, D., & Tamascelli, D. (2011). An Introduction to Quantum Annealing. *RAIRO – Theoretical Informatics and Applications*, 45(1), 99-116.

Gay, S. J. (2006). Quantum programming languages: Survey and bibliography. *Journal of Mathematical Structures in Computer Science*, 16(4), 581–600. doi:10.1017/S0960129506005378

Lidar, D. A., & Brun, T. A. (Eds.). (2013). *Quantum Error Correction. eBook*. United Kingdom: Cambridge University Press. doi:10.1017/CBO9781139034807

Perlner, R. A., & Cooper, D. A. (2009). Quantum Resistant Public Key Cryptography: A Survey. In *Proceedings of 8th Symposium on Identity and Trust on the Internet IDTrust* (vol. 8, pp. 85-93). Gaithersburg, MD. doi:10.1145/1527017.1527028

Quantum Information Science. (2015). *Wikipedia*. Retrieved October 26, 2015, from http://en.wikipedia.org/wiki/Quantum_information_science

van Assche, G. (2006). *Quantum Cryptography and Secret-Key Distillation*. England: Cambridge University Press. doi:10.1017/CBO9780511617744

Yanofsky, N. S., & Mannucci, M. A. (2008). *Quantum Computing for Computer Scientists*. England: Cambridge University Press. doi:10.1017/CBO9780511813887

KEY TERMS AND DEFINITIONS

AQC: Adiabatic quantum computation, calculation by quantum annealing.

Avalanche Photodiode: The semiconductor analog to a photomultiplier.

Decoherence Time: The time a qubit state prevails before collapse.

Entanglement: A multiple qubits state which cannot be created by combining single qubit states.

Photon: A discrete packet of electromagnetic energy.

Polarization: The propagation plane of an electromagnetic wave.

Quantum GATE: A device changing the state of one or multiple qubits.

Qubit: A two-state quantum-mechanical system which can be a concurrent superposition of both states.

Teleportation: Transfer of a qubit state.

Sleptsov Net Computing

T

Dmitry A. Zaitsev

International Humanitarian University, Ukraine

INTRODUCTION

Recently many researchers introduce new models of hyper-computations, such as quantum computations, computations on cell membranes, spiking P neurons and DNA (Cook & Neary, 2013), capable breaking through the obstacle of intractable tasks. Petri nets have been known for years as a model of concurrent systems (Murata, 1989) but their computationally universal extensions are exponentially slow comparing Turing machines, especially when implementing arithmetic operations. A Sleptsov net concept, suggested quarter a century ago, recently acquired its second birth (Zaitsev, 2016) due to its ability of fast implementation of basic arithmetic operations. Firing a transition in a few instances at a step leads to universal constructs which run in polynomial time (Zaitsev, 2017). In Sleptsov net computing (Zaitsev, 2014a; Zaitsev & Jürjens, 2016), a program, written in Sleptsov net language preserving concurrency of an application area, runs on Sleptsov net processor which implements concurrent firing of transitions in multiple instances providing computations having ultra-performance.

BACKGROUND

A concept of algorithm was formalized for the first time by Alan Turing in 1936 in the form of an abstract machine which is traditionally called a Turing machine. Universal Turing machine which runs a given Turing machine is considered a prototype of a traditional computer. Besides Turing machines, other computationally universal systems appeared: recursive functions of Kleene, normal

algorithms of Markov, tag rewriting systems of Post, register machines of Minsky. Variety of models is explained by controversial requirements of manifold application areas. Recent models employ facilities of massively parallel computations even in such simple constructs as elementary cellular automata which universality was proven in 2004 by Mathew Cook. Besides, smallest universal Turing machines were constructed in 2008 by Turlough Neary and Damien Woods which run in polynomial time. However, the way of programming cellular automata after Mathew Cook does not reveal their ability for massively parallel computing. Sleptsov net concept (Zaitsev, 2016) mends the flaw of Petri nets (Murata, 2013), consisting in incremental character of computations, which makes Sleptsov net computing (Zaitsev, 2014a; Zaitsev & Jürjens, 2016) a prospective approach for ultra-performance concurrent computing. In Zaitsev (2016), an overview of works, which refer to Sleptsov nets (Petri nets with multichannel transitions or multiple firing strategy), is presented.

MAIN FOCUS OF THE ARTICLE

Issues, Controversies, Problems

Definition of Sleptsov Net

A *Sleptsov net* (SN) is a bipartite directed multi-graph supplied with a dynamic process (Zaitsev, 2016). An SN is denoted as $N=(P,T,W,\mu_0)$, where P and T are disjoint sets of vertices called *places* and *transitions* respectively, the mapping F specifies *arcs* between vertices, and μ_0 represents the initial state (*marking*).

DOI: 10.4018/978-1-5225-2255-3.ch672

The mapping $W: (P \times T) \rightarrow \mathbb{N} \cup \{-1\}$, $(T \times P) \rightarrow \mathbb{N}$ defines arcs, their types and multiplicities, where a zero value corresponds to the arc absence, a positive value – to the *regular arc* with indicated multiplicity, and a minus unit – to the *inhibitor arc* which checks a place on zero marking. $\mathbb{N}$ denotes the set of natural numbers. To avoid nested indices we denote $w_{j,i}^- = w(p_j, t_i)$ and $w_{i,j}^+ = w(t_i, p_j)$. The mapping $\mu: P \rightarrow \mathbb{N}$ specifies the place marking.

In graphical form, places are drawn as circles and transitions as rectangles. An inhibitor arc is represented by a small hollow circle at its end, and a small solid circle represents the abbreviation of a loop. Regular arc's multiplicity, greater than unit, is inscribed on it and place's marking, greater than zero, is written inside it. Examples of SNs computing basic arithmetic and logic operations are shown in Figure 6.

To estimate *multiplicity of firability conditions* on each incoming arc of a transition, the following auxiliary operation is defined

$$x \succ y = \begin{cases} x / y, & \text{if } y > 0 \\ 0, & \text{if } y = -1, x > 0, \\ \infty, & \text{if } y = -1, x = 0. \end{cases}$$

To avoid inconsistency with infinite number of instances, here we prohibit transitions without input regular arcs.

The behavior (dynamics) of a SN could be described by the corresponding state equation similarly to (Zaitsev, 2012). The present work considers the behavior as result of applying the following *transition firing rule*:

- The number of instances of transition t_i firable at the current step is equal to

$$v_i = v(t_i) = \min_j (\mu_j \succ w_{j,i}^-), 1 \leq j \leq m, w_{j,i}^- \neq 0$$

- When transition t_i , $v_i > 0$ fires, for $u_i \leq v_i$ it
 - extracts $u_i \cdot w_{j,i}^-$ tokens from each its input place p_j for regular arcs $w_{j,i}^- > 0$;
 - puts $u_i \cdot w_{i,k}^+$ tokens into each its output place p_k , $w_{i,k}^+ > 0$;
- The net halts if firable transitions are absent.

When a transition, having a single regular incoming arc with multiplicity a from place p and a single regular outgoing arc with multiplicity b to place p' , fires with $u_i = v_i$, it implements the following computations $\mu(p) = \mu(p) \bmod a$; $\mu(p') = \mu(p') + b \cdot (\mu(p) \text{ div } a)$. Namely, it implements division by a with a remainder and multiplication by b . Choosing either a or b equal to unit we obtain either pure multiplication or pure division.

In Petri net, only one transition fires at a step while in Sawicki net (Burhard, 1981) a maximal set of firable transitions fires at a step. However in both Petri and Sawicki nets, only one instance of a transition fires at a step.

Transitions could be thought of as virtual actions. The number of really started actions depends on the amount of available resources represented by transitions' input places. Why we should restrict the number of transitions' instances to unit and fire them in sequence, as in Petri net, when available resources allow firing them simultaneously? Anyway, classical sequential order of transitions' firing could be obtained as a special case attaching a place to each transition connected with read arc and having marking equal to unit.

Various extensions of Petri nets are known such as priority, inhibitor, timed, loaded (colored), hierarchical, and nested Petri nets (Murata, 1989). Sometimes they sophisticate the basic model considerably. Our goal consists in obtaining hyper-performance at the cost of minimal modification which consist in the multiple firing

strategy. As for other concepts concurrent to Petri nets, multiset rewriting systems (Verlan, 2011) should be mentioned as well as vector addition and replacement systems. Moreover, for studying protocols, the concept of Hoare's communication sequential processes is close to Petri nets and mutual transformations are possible. An essential advantage of Petri (and Sleptsov) nets is graphical presentation of systems' structure and behavior.

Universal Sleptsov Net as a Processor

When programming in Sleptsov (Petri, Salwicki) net language, we use a concept of dedicated contact places. Before a net starts, we upload input data into its contact places, and when a net halts, we download output data from its contact places. Sometimes deeper specialization is more convenient when we distinguish separate subsets of input and output places. Note that data of various data types are encoded as nonnegative integer numbers to be processed by a net.

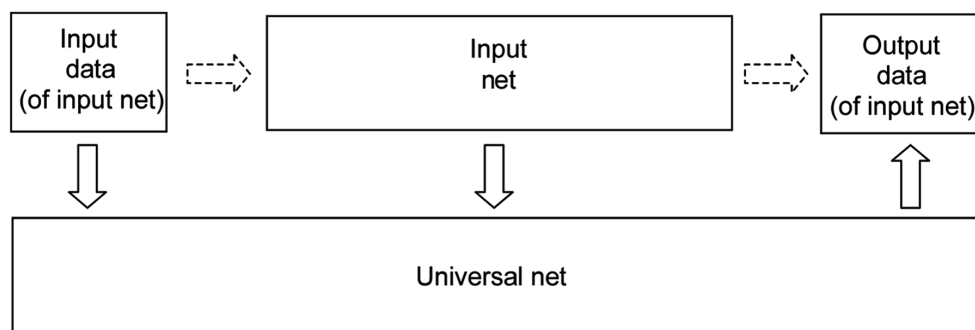
Turing-completeness of extended PNs means that a universal net, which runs a given net, exists. Recently a series of small universal Sleptsov nets was constructed (Zaitsev, 2017, 2014b) in which each net of the series can be considered a prototype of a processor in a Sleptsov net based computer (Zaitsev, 2014a). A program of such a computer, composed in the form of a (inhibitor) Sleptsov net and encoded, is given as an input to

a processor which represents a hardware implementation of a universal Sleptsov net; Figure 1 illustrates the approach.

The main obstacle for wide implementation of the Petri net paradigm of computation (Zaitsev, 2014a) has been the fact that all known small universal nets run in exponential time. In (Zaitsev, 2017), a small universal Sleptsov net was constructed which runs in polynomial time; the net contains 15 places and 29 transitions. Technique of a universal SN composition (Zaitsev, 2017) using reverse control flow and subnets for multiplication and division by a constant is handy for developing the technology of programming in SNs. The considerable complexity difference between universal PNs and universal SNs is easily explained: when processing the TM tape code, which is an exponent on the working zone size, a universal PN takes a single token at a step while a universal SN takes all of the tokens; eventually, we come to arithmetic operations, namely multiplication and division employed for encoding/decoding input PN.

In (Zaitsev, 2014a) we offer a computing memory concept for massively parallel asynchronous implementation of SN processors that promises ultra-performance when processing programs written in SN language. Apart from the problem of efficient hardware implementation of a SN computer, an issue arises regarding the technology of programming in SNs and preserving natural parallelism (concurrency) of an application area.

Figure 1. Universal net as a processor



Universal Sleptsov Net Runs in Polynomial Time

Similar to (Zaitsev, 2014b, 2017), we construct a universal SN via simulation of weakly-universal Turing machine of Neary and Woods with 2 states and 4 symbols denoted as WUTM(2,4). As far as Turing machine behavior is deterministic, we use deterministic SN where transitions are enumerated and the firable transition with the least index fires in the maximal number of instances. For visual reflection of the transitions order (priority), we use arcs connecting transitions; if there is an arc from transition t to transition t' , it means that index of transition t should be less than index of transition t' (lesser index means higher priority here). The peculiarity of programming in SN is application of reverse control flow represented with moving zero marking; thus the initial marking of control flow places is equal to 1. We check zero marking via an inhibitor arc which does not restrict the multiplicity of firability conditions.

The source information for simulation is the transition function of WUTM(2,4) and the encoding of its states and tape symbols (Zaitsev, 2017) given by Table 1. In WUTM(2,4), an infinite repetition of definite blank words is written on its tape: $b_l = 00\emptyset 1$ to the left and $b_r = 0\bar{1}\emptyset\emptyset 1$ to the right of the working zone. According to the function $s(x_{k-1}x_{k-2}\dots x_0) = \sum_{i=0} s(x_i) \cdot r^i$ for the tape words' encoding, the codes of the left and right blank words are: $s(b_l)=167$ and $s(b_r)=13596$.

The constructed universal Sleptsov net is denoted as USN(15,29) with respect to the total number of its places and transitions, respectively.

The general scheme of USN(15,29) is shown in Figure 2a. Subnets are depicted as rectangles with double line border. Some vertices have mnemonic names besides their numbers. Used subnets FS, MA5LR, and MD5LR are represented in Figure 2b, Figure 2c, and Figure 2d, respectively. Places with the same name (number) are considered as the same place and should be merged all over the components. The final assembly of USN(15,29) drawn in (Zaitsev, 2017) looks rather tangled.

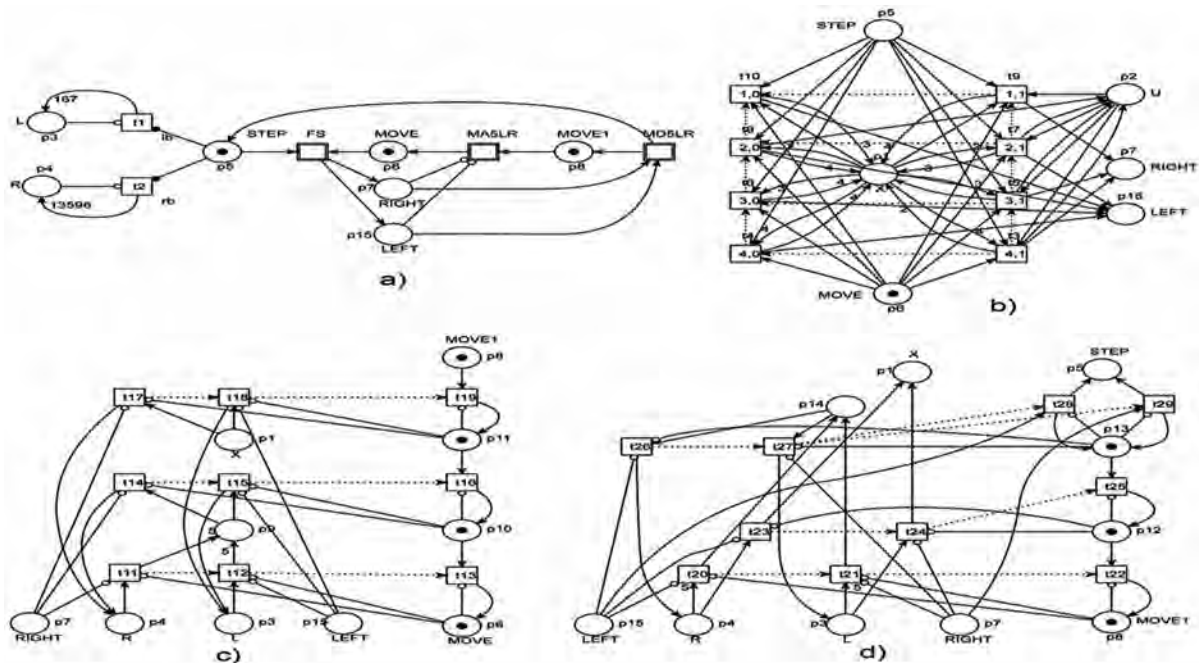
Place U contains encoded TM state $s(u)$, place X contains encoded current cell symbol $s(x)$, and places L and R contain encoded left and right parts of the tape working zone respectively regarding the current cell. At the beginning of each computation step, place STEP launches subnet FS, which simulates WUTM(2,4) transition function. Subnet FS produces the encoding $s(u')$ of the new state and the encoding $s(x')$ of the new symbol in places U and X respectively to simulate the Turing machine instruction. Subnet FS also puts a token into either place RIGHT or place LEFT to indicate the control head right or left moves correspondingly. A token is extracted from place MOVE after subnet FS has finished that launches the sequence of subnets MA5LR, MD5LR, which simulates the control head moves. At the end of a simulated computation step, a token is put into place STEP that allows the simulation of the next instruction to begin. Moreover, places LEFT and RIGHT are cleared.

The TM tape is represented by the places L, X, and R containing the encoded left part of the working zone, the current cell symbol and the encoded right part of the working zone correspondingly. The moves of the tape head on the tape are

Table 1. Weakly-universal Turing machine WUTM(2,4) and its encoding

Σ/Ω	$s(\Sigma)/\sigma(\Omega)$	u_1	u_2
		0	1
0	1	3,left,0	4,right,0
1	2	4,left,1	3,left,1
$\emptyset$	3	4,left,0	1,right,1
$\bar{1}$	4	4,left,0	2,right,1

Figure 2. Component-wise representation of USN(15,29)



simulated by the two connected subnets MA5LR and MD5LR shown in Figure 2c and Figure 2d, respectively. The meaning of subnets' names is the following: MA5 means multiplication and addition with radix 5 ($S := S \cdot 5 + X$); MD5 means modulo and division with radix 5 ($S := S \text{ div } 5$, $X := S \bmod 5$); LR means choice of places either L or R, where codes of the left and right parts of the tape, regarding the current cell symbol code X, are stored, depending on the marking of places LEFT and RIGHT.

Two transitions lb and rb simulate peculiarities of weakly universal TM work, they add the blank word codes $s(b_l)$ and $s(b_r)$ to the codes L and R of the left and right parts of the tape working zone correspondingly when its value is equal to zero.

Thus, the sequence of subnets MA5LR, MD5LR implements the following operations:

- To simulate a left move (when place RIGHT=0): $R := R \cdot 5 + X$, $L := L \text{ div } 5$, $X := L \bmod 5$;
- To simulate a right move (when place LEFT=0): $L := L \cdot 5 + X$, $R := R \text{ div } 5$, $X := R \bmod 5$.

In (Zaitsev, 2017), it was proven that USN(15,29) simulates WUTM(2,4) in time $O(14 \cdot k)$ and space $O(k)$, where k is the number of WUTM(2,4) steps. Taking into consideration polynomial complexity of running a TM on WUTM(2,4) and polynomial complexity of running an SN on TM, we obtain the final polynomial complexity of the constructed universal Sleptsov net.

Principles of Programming in Sleptsov Nets

Programming in SNs (Zaitsev & Jürjens, 2016) represents a composition of data and control flows supplied with transition substitution to describe hierarchical structure of routines (nets). We suppose that to combine control flow with a subnet which substitutes a transition, a special pair of places is used: *Start* – to launch the subnet and *Finish* – to indicate its completion.

Control flows are represented in an inverse form with a moving zero marking because a zero marking is easily checked via an inhibitor arc that does not restrict the number of fireable instances

of a transition. Thus, initially, all control flow places should have marking equal to one.

To simulate standard control flows of classic programming languages, we offer the patterns represented in Figure 3 for a sequence a), branching b), cycle c), and parallel execution d). Moreover, an arbitrary subnet with markings belonging to $\{0,1\}^m$ and with a pair of start/finish places could be considered as a control flow.

Each of transitions in Figure 3 could be substituted by a subnet using its input place to start it and its output place to indicate its completion. Moreover, additional incidental places could be attached which represent required input and output variables. In the patterns with alternative transitions (Figure 3 b, c), we suppose either an external control via conditional variables to make it deterministic or nondeterministic choice.

For computation of expressions, a data flow approach seems more efficient when a corresponding subnet is started and finished with an external control flow. It is a question of granulation whether to use strict composition of subnets launched via

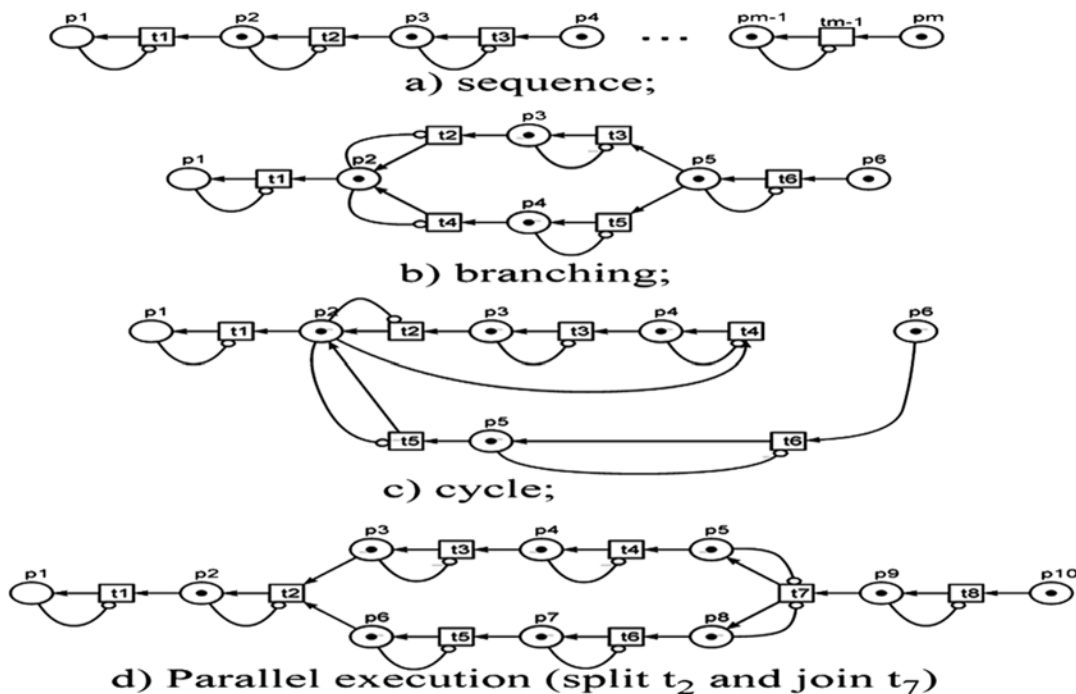
control flow or to write anew a closely connected tangle of control and data flows where each of them is not even clearly distinguished.

A program written in Sleptsov nets is obtained via composition of control flows with data flows using a modular approach. Substitution of a transition by a module is defined by the indication of the module name and connection (mapping) of its input and output places with places of a net which either represent variables or control flow.

It is also supposed that before a module starts, all of the input data is copied into its input places while after a module finishes and before its next start, all the output data is moved from its output places. For these purposes dashed arcs were introduced (Zaitsev, 2012) to designate a brief and convenient way of working with data.

A dashed input arc of a module denotes a subnet COPY for copying an input variable value into the corresponding input place of the module before launching its work via the place *Start*. A dashed output arc of a module denotes a sequence of subnets CLEAN, MOVE for replacing a variable

Figure 3. Standard control flow patterns



value with the result obtained in the corresponding output place of module. When a module has a few input (output) places, the subnets COPY (CLEAN, MOVE) are inserted using the parallel execution pattern (Figure 6). An example is shown in Figure 4 a,b where subnet a) having dashed arcs is expanded into the corresponding low-level subnet b).

Introduced for reduction in auxiliary subnet size, dotted arcs denote MOVE for both input and output variables; MOVE subnet is more concise and provides cleaning of its input variable. Thus, the value of the input variable is not preserved and the value of the output variable is incremented. When the same variable is both input and output, a dotted two-direction arc is used which is expanded as MOVE before launching a module and MOVE after finishing the module. The corresponding example is shown in Figure 4 c, d where subnet c) having a dotted arc is expanded into the corresponding low-level subnet d).

Thus, a module could be considered as a programming language procedure whose contact places correspond to its formal parameters. Copying variables with dashed and dotted arcs

correspond to the substitution of the actual input parameters and extracting its output parameters. The question remains open regarding the way a module call can be implemented either using an inline fashion via cloning entire modules or a call-return fashion via having a single copy of a module and switching control flows. The difference between the two mentioned ways is illustrated in Figure 5 for two references to the addition module ADD (without considering data). While the call-return fashion is more concise because of having a single copy of a module, the inline fashion is more attractive from the concurrent execution point of view (Zaitsev, 2014a).

A program composed of modules is finally compiled into a plain inhibitor priority Slepsov net that is loaded for execution on a SN computer. The best performance is achieved when all the transitions are fired independently based on their fireability conditions; to resolve conflicts a special arbiter is provided which blocks incidental places of a firing transition. This way of execution also requires writing programs which are invariant to the transitions' firing order.

Figure 4. Examples of dashed arcs expansion

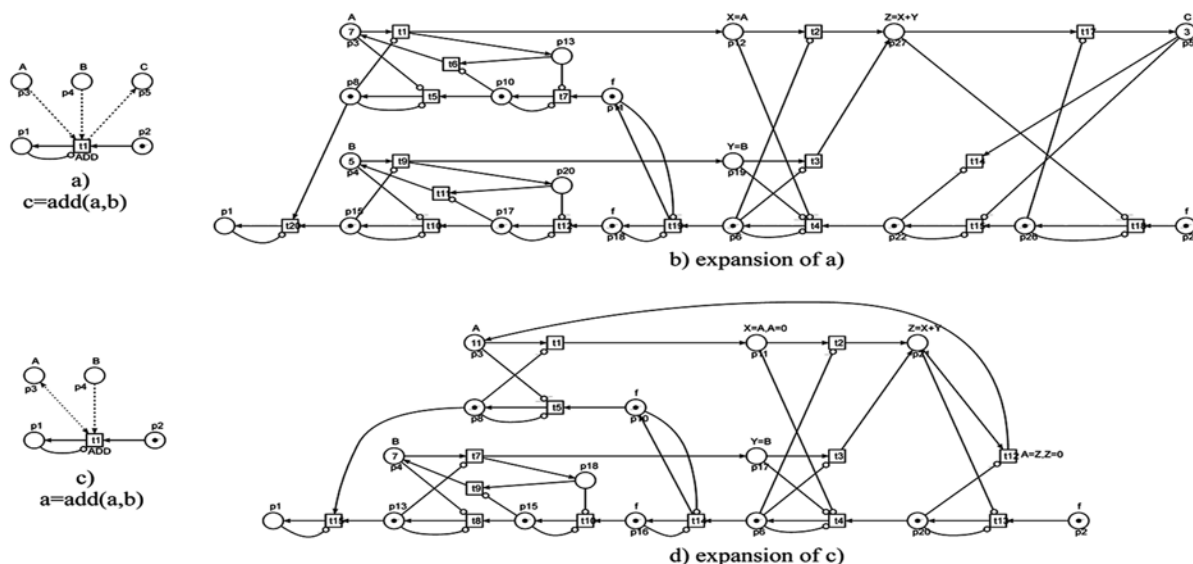
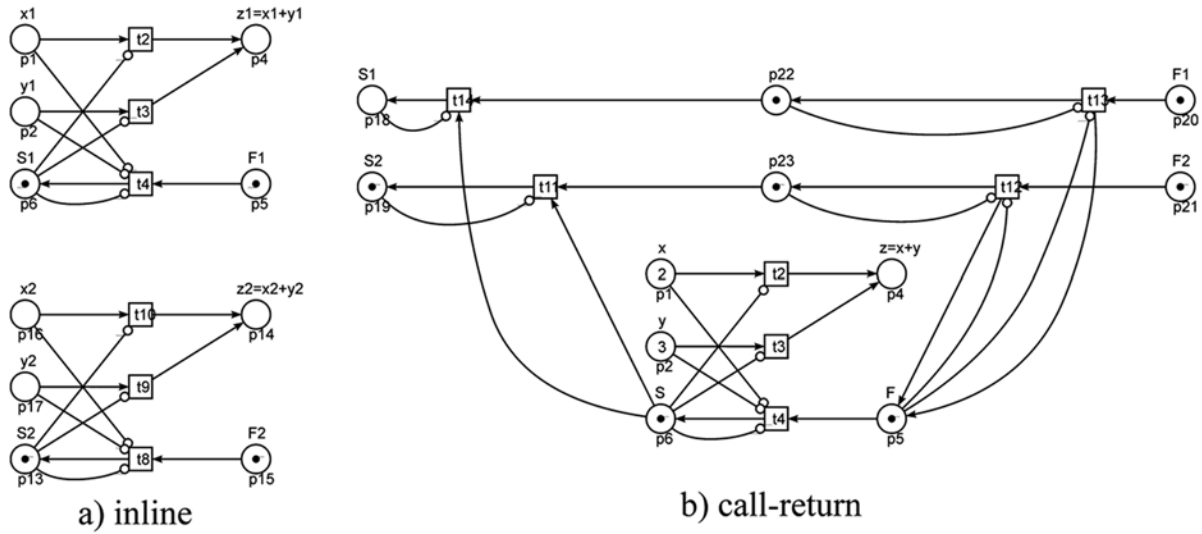


Figure 5. Example of transitions' substitution implementation for $z1=ADD(x1,y2)$; $z2=ADD(x2,y2)$ 

Efficient Implementation of Arithmetic and Logic Operations

To describe the transition substitution operation in a formal way, regarding an external control flow and input/output variables, a concept of *module* is introduced (Zaitsev, 2016). A module represents a subnet with contact (input and output) places which could be either the only connection with the outside world for closed modules or combined with the global variables usage for open modules. The module work is controlled via a special pair of contact places: place *Start* (*s*) starts work, and place *Finish* (*f*) indicates the completion of a module's work.

Extracting a token from the first control flow place *Start* launches the moving zero marking till it arrives to the last control flow place *Finish*. We put the following restrictions on the control flow graphs. It is supposed that all actions of a module are controlled by its control flows and inhibitor arcs are used to fire a transition. Thus, when all the control flow places contain a token, the module transitions are disabled. Therefore, before extracting a token from the place *Start* as well as after the arrival of zero to the place *Finish*, the module stands still.

To provide the re-enterability of a module, it is supposed that: when it starts, all its data places except input places have zero marking, and when it finishes, all its data places except for output places have zero marking. In Figure 6, nets are represented which implement basic copying, arithmetic and logic operations. Compared to (Zaitsev, 2012), subnets are reduced by one place and one transition under the assumption that we can use the input place *Start* to control transitions inside a subnet, but the place *Finish* is not used for the control inside a subnet.

Using the technique of analyzing all the permitted fireable transition, employed in (Zaitsev, 2014b), we can prove that the subnets represented in Figure 6 implement the corresponding operations.

For classic Petri nets, operations of multiplication and division are the most complex from the computational point of view; the corresponding subnets were studied in (Zaitsev, 2012). Namely their complexity influences the exponential slowness of computations on Petri nets. Considered in (Zaitsev, 2017), fragments of the universal Sleptsov net efficiently implement multiplication and division by a constant (equal to 5). Now we study multiplication and division of arbitrarily

given nonnegative integers using multiplication and division by a constant (equal to 2).

We choose simple standard algorithms of long multiplication and division. The best known efficient algorithms of multiplication and division for natural numbers could be encoded in SN as well. The multiplication algorithm represents the usual “long multiplication”. To find the current digit of the multiplier, the remainder of the division by two $d=y\%2$ is used; then the multiplier is recalculated via $y/=2$ for working with the next digit on the next pass of the main loop. The multiplicand is multiplied by two $x*=2$ which represents its shift to the left. When the current digit d of the multiplier is equal to 1, the shifted multiplicand is added to the result. The algorithm could be optimized to avoid recalculating $x*=2$ when the new value of y is equal to zero but it leads to a more sophisticated SN. The algorithm was encoded as a SN and is represented in Figure 7.

For Sleptsov nets shown in Figures 6 and 7, it has been proven (Zaitsev, 2016) that:

- CLEAN, MOVE, COPY, NOT, OR, AND, ADD, SUB, GT0, GT implement corresponding operations with time and space complexity equal to a constant;

- MUL implements multiplication of non-negative integer numbers x and y with time complexity $O(11 \cdot \log_2 y + 3)$ and constant space complexity (equal to 15);
- DIV implements division of nonzero integer numbers x by y with time complexity $39 \cdot (\log_2 x - \log_2 y) + 19$ and constant space complexity (equal to 48).

MUL(x, y): $z := x \cdot y$, $x := 0$, $y := 0$;

DIV(x, y): $z := x/y$, $r := x\%y$, $x := 0$, $y := 0$.

As an example, to calculate $3 \cdot 2 = 6$, the number 3 is loaded into place X (p_1) and the number 2 is loaded into place Y (p_2). To start computations, a token is removed from place S (p_3). Then the only enabled firing sequence

$$t_1 t_5 t_6 t_8 t_{10} t_{11} t_{12} t_{13} t_{14} t_{16} t_1 t_6 t_7 t_8 t_9 t_{11} t_{13} t_{14} t_{15} t_2 t_3 t_4$$

leads to obtaining the result 6 in place Z (p_4) which is indicated by the removal of a token from place F (p_5).

Note that, the multiplication SN module shown in Figure 7 multiplies given natural numbers X and Y , and uses for this purpose multiplication

Figure 6. Modules which implement basic operations: copying, logic, and arithmetic

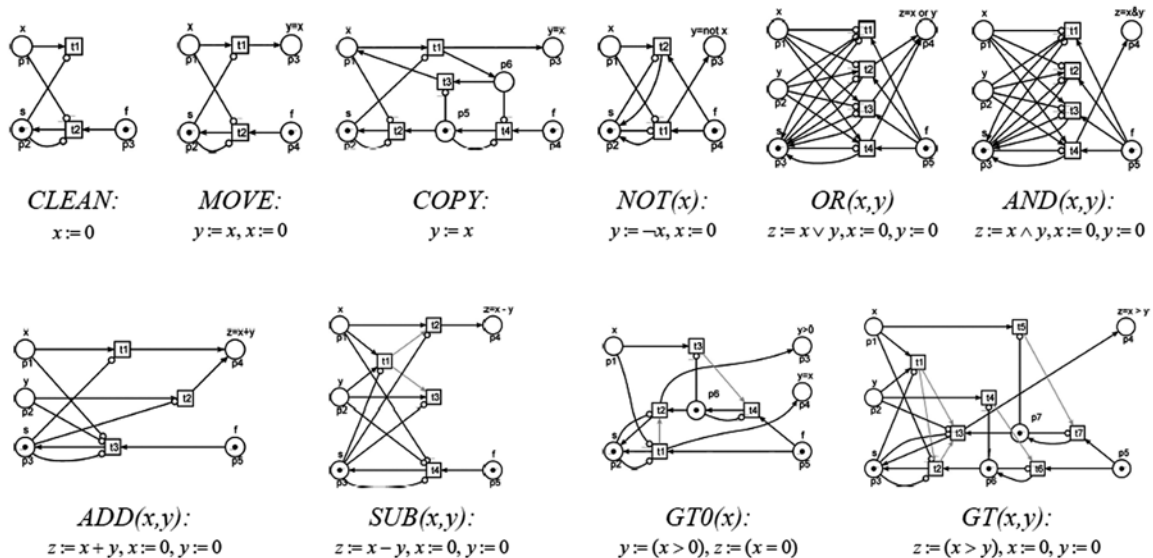
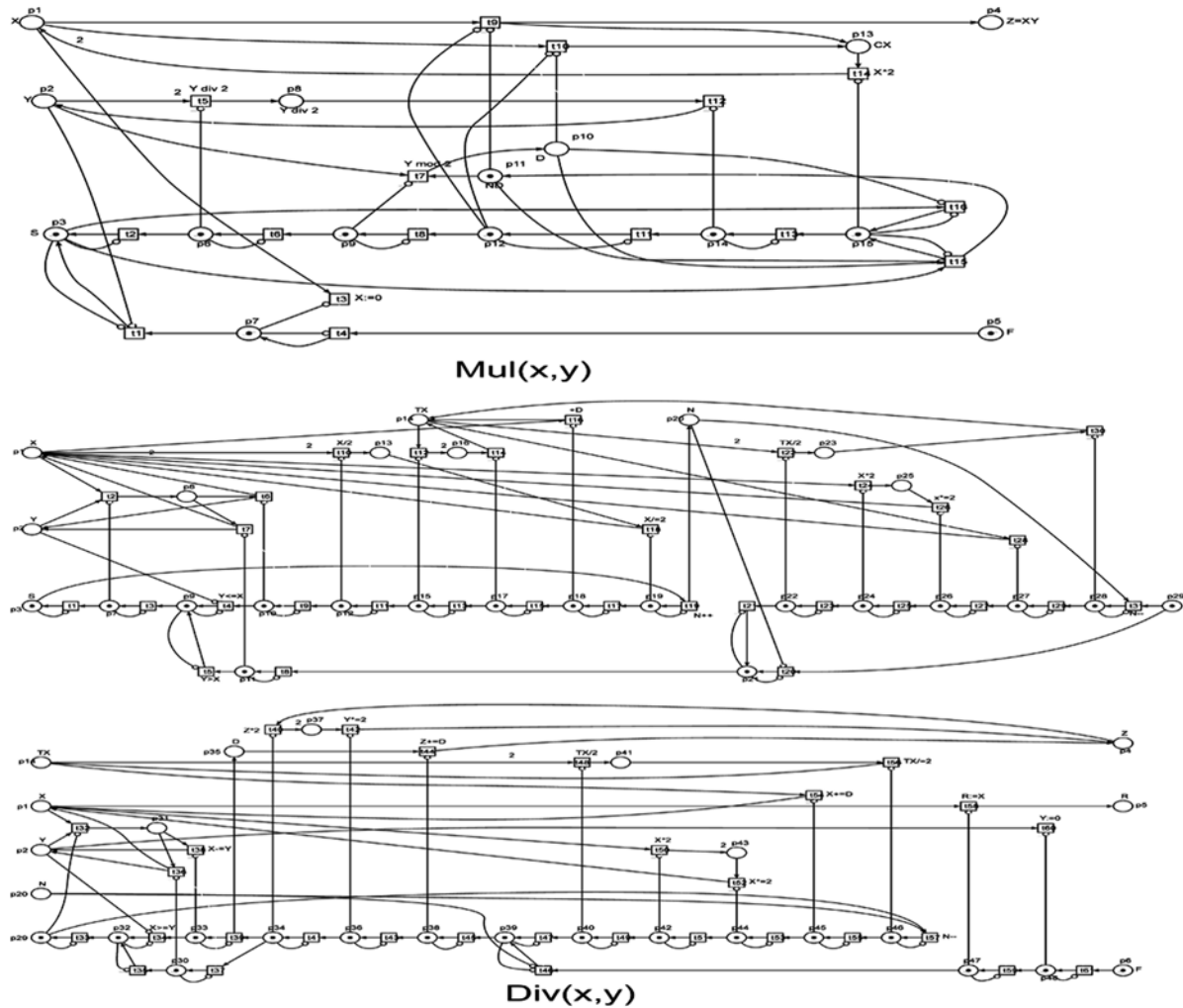


Figure 7. Modules which implement multiplication and division



and division by the constant 2 implemented with a single SN transition; multiplication and division by constant 2 are usually efficiently implemented as the left and right shift of binary code correspondingly.

The division algorithm represents the usual “long division”. An initial dividend x is split into two parts. Namely, x which represents its head comparable to y and its tail tx . The first loop *while*($y \leq x$) implements the splitting via obtaining separate digits d from the left of the tail using modulo operation $d = x \% 2$ while x is truncated of the last digit via division operation $x = x / 2$. The obtained digits are appended to the left of the tail using $tx = tx * 2 + d$. The branching *if*($n > 0$)

moves the separation one digit back when possible guaranteeing that x is the minimal possible head that is greater than y .

In a real-life implementation, either hardware or software, of SN as well as PN, we should consider the complexity of the above mentioned procedures. Supposition of a sequential computation device (Zaitsev, 2014b) yields the multiplier $O(|P| \cdot |T| \cdot \log_2 l)$ for PNs, where l is the maximal number of steps. For SNs, an extra complexity of multiplication and division by a constant equal to a power of 2 is also estimated as $O(\log_2 l)$; for an arbitrary constant the complexity is sublinear. If we suppose some independent computing memory device implements each transition (Zaitsev, 2014a),

we get rid of the multiplier $|P| \cdot |T|$ and obtain an ultra-parallel processor of Sleptsov nets with step complexity $O(\log_2 l)$.

The considered estimations are too meticulous for usual comparisons of arithmetic algorithms. Supposing that x and y contain about n binary digits $x \approx y < 2^n$ and using the logarithmic scale with the multiplier $\log_2(x \cdot y) \approx 2 \cdot n$, we obtain the multiplication complexity $O(2^{2n})$ for PNs and $O(n^2)$ for SNs. For the division operation, we suppose that x contains $2 \cdot n$ binary digits and y contains n binary digits. Then we obtain the division complexity $O(2^{2n})$ for PNs and $O(n^2)$ for SN. Thus, SNs run faster comparing PNs. For multiplication and division operations, SNs run in polynomial time while PNs require exponential time.

Note that some SNs can be interpreted as PNs giving the same result but with a certain slowdown of execution. The investigation of such equivalence is a direction of future work.

SOLUTIONS AND RECOMMENDATIONS

From the described above implementations of universal Sleptsov net and Sleptsov nets for arithmetic operations having polynomial time complexity, we conclude that they run exponentially faster than Petri nets and recommend their application as a

new kind of concurrent computing (Zaitsev, 2016). Basics of Sleptsov net programming perceived as a composition of data and reverse control flow were developed.

Computations on SN gain wide application. We complete the article with three examples described in (Zaitsev, 2016) from rather diverse areas such as data encryption with open key, soft computing with fuzzy logic, and parallel and distributed computing shown in Figure 8, 9, 10, respectively. The first net use studied subnets for arithmetic operations and additional subnet REV for reversing the bitwise representation of z while the second and third nets implement logic and arithmetic operations directly via SN transitions.

We recommend SN computing in the first place for application areas where concurrent style of programming is advised and brings considerable speed-up of computations.

FUTURE RESEARCH DIRECTIONS

Practical implementation of Sleptsov net computing requires ad-hoc CASE-systems development and hardware implementation of Sleptsov net processors. From theoretical point of view, issues of proving correctness of Sleptsov net programs and design of universal Sleptsov nets, which employ massively parallel approach, are in demand.

Figure 8. RSA encryption/decryption $y = x^z \bmod n$;

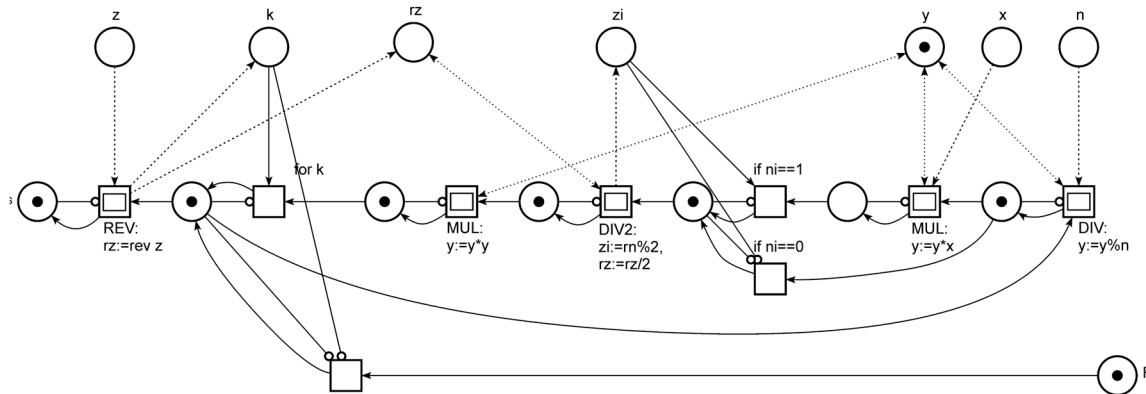


Figure 9. Calculating fuzzy logic function $\varphi = x_1 \bar{x}_2 \vee \bar{x}_1 x_2 = ((x_1 \bar{x}_2) \wedge (\bar{x}_1 x_2))$;

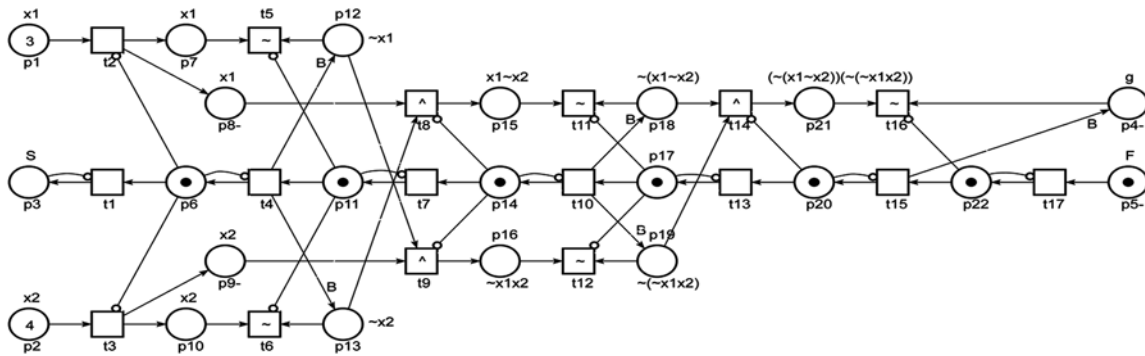
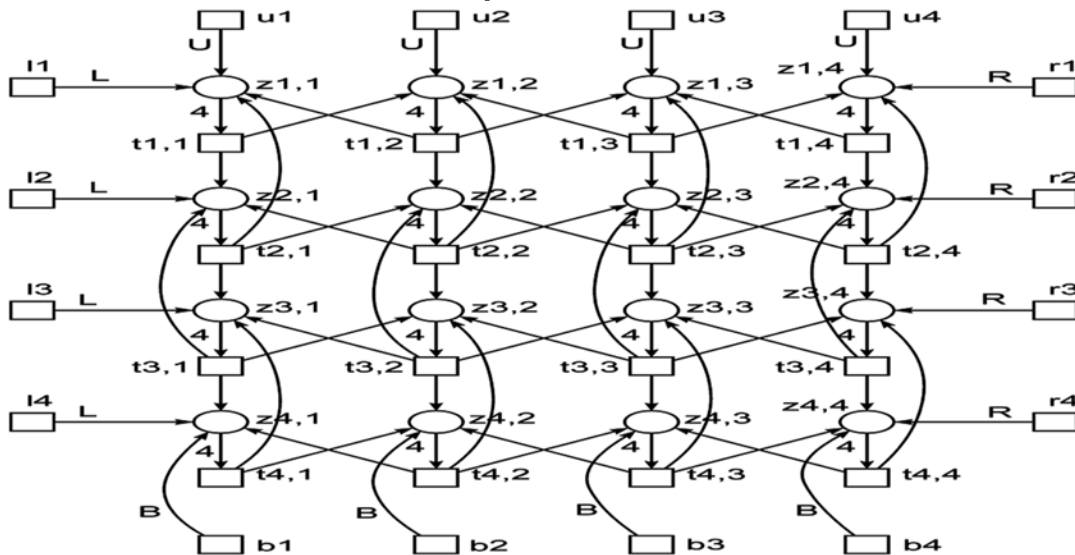


Figure 10. Solving Laplace equation $\frac{\delta^2 \varphi}{\delta x^2} + \frac{\delta^2 \varphi}{\delta y^2} = 0$.



CONCLUSION

The article studies a more general class of place-transition nets introduced by Anatoly Sleptsov, which allow a transition firing in multiple instances at a step. It was shown that Sleptsov nets run exponentially faster than Petri nets. A small universal inhibitor Sleptsov net runs in polynomial time and could be considered as a prototype of a SN processor devised for executing programs written in a SN language. Efficient massively parallel implementations of SN processors are discussed in (Zaitsev, 2014a).

Basics of Sleptsov net programming technology were developed including efficient implementation of arithmetic and logic operations, as well as rules of program composition via combination of data and control flows and substitution of transitions. Sleptsov net computing has such advantages as concurrent programming style, concise graphical language and small granulation of parallel processes on the level of separate events which promise hyper performance.

REFERENCES

- Alhazov, A., & Verlan, S. (2011). Minimization strategies for maximally parallel multiset rewriting systems. *Theoretical Computer Science*, 412(17), 1581–1591. doi:10.1016/j.tcs.2010.10.033
- Burkhard H.-D. (1981). Ordered Firing in Petri Nets. *Journal of Information Processing and Cybernetics*, 2, 71–86.
- Murata, T. (1989). Petri nets: Properties, analysis and applications. *Proceedings of the IEEE*, 77(4), 541–580. doi:10.1109/5.24143
- Neary, T., & Cook, M. (Eds.). (2013) *Proceedings Machines, Computations and Universality*. Zurich, Switzerland: Electronic Proceedings in Theoretical Computer Science 128.
- Zaitsev, D. A. (2012). Universal Petri Net. *Cybernetics and Systems Analysis*, 48(4), 498–511. doi:10.1007/s10559-012-9429-4
- Zaitsev, D. A. (2014a). Paradigm of Computations on the Petri Nets. *Automation and Remote Control*, 75(8), 1369–1383. doi:10.1134/S0005117914080025
- Zaitsev, D. A. (2014b). Toward the Minimal Universal Petri Net. *IEEE Transactions on Systems, Man, and Cybernetics. Systems*, 44(1), 47–58.
- Zaitsev, D. A. (2016). Sleptsov Nets Run Fast. *IEEE Transactions on Systems, Man, and Cybernetics Systems*, 46(5), 682–693.
- Zaitsev, D. A. (2017). Universal Sleptsov Net. *International Journal of Computer Mathematics*. doi:10.1080/00207160.2017.1283410
- Zaitsev, D. A., & Jürjens, J. (2016). Programming in the Sleptsov net language for systems control. *Advances in Mechanical Engineering*, 8(4), 1–11. doi:10.1177/1687814016640159

ADDITIONAL READING

- Aalst, W., & Stahl, C. (2011). *Modeling Business Processes: A Petri Net-Oriented Approach*. MIT Press.
- Li, Z. W., & Zhou, M. C. (2009). *Deadlock Resolution in Automated Manufacturing Systems: A Novel Petri Net Approach*. London, UK: Springer.
- Reisig, W. (2013). *Understanding Petri Nets: Modeling Techniques, Analysis Methods, Case Studies*. Springer. doi:10.1007/978-3-642-33278-4
- Zaitsev, D. A., & Sleptsov, A. I. (1997). State Equations and Equivalent Transformations for Timed Petri Nets. *Cybernetics and Systems Analysis*, 33(5), 659–672. doi:10.1007/BF02667189

KEY TERMS AND DEFINITIONS

Algorithm: A declaration of a stepwise process of computations; the order of steps precisely prescribed; at each step exactly known how to obtain a result from the source data.

Computationally Universal System: A system capable to execute (run) a given algorithm.

Place-Transition Net: A directed bipartite graph with a dynamic process defined on it.

Petri Net: A place-transition net where one transition fires at a step.

Salwicki Net: A place-transition net where the maximal set of firable transitions fires at a step – maximum firing strategy.

Sleptsov Net: A place-transition net where a transition fires at a step in a few copies – multiple firing strategy.

Turing Machine: An abstract machine consisting of infinite tape and a control head; was introduced by Alan Turing in 1936 as a formalization of the algorithm concept.

Category U

Ubiquitous and Pervasive Computing

Energy Conservation in the Era of Ubiquitous Computing



P. P. Abdul Haleem
Farook College, India

INTRODUCTION

The widespread availability of computing devices and easy access to the Internet has revolutionized the way people use computers. Computers and computing devices became so common and popular among the people of all walks of life. Advances in the technology pave way to computing devices that can be carried with users wherever they are – hence paving a new era called as “ubiquitous computing”. The barriers such as type of devices, location, time and format has become meaningless in this era (Mark, 1999). Other terminologies used for ubiquitous computing include pervasive computing, ambient intelligence (Hansmann, Merk, Nicklous, & Stober, 2003), ambient media (Artur, Thomas, Bjoern, Kari, & Juha, 2009) or everywhere (Adam 2006).

Ubiquitous computing resulted in a paradigm shift from the hitherto computing called desktop computing: (i) the traditional wired networks had to be extended to include wireless networks, and (ii) the concept of anywhere computing was materialized with the introduction of devices that are mobile and wireless. The extension of traditional wired network into wireless network resulted in a lot of challenges due to the reason that existing protocols and supporting software had to be extended to include wireless arena also. The introduction of wireless mobile devices also demanded a lot from the research and developer community for their smooth existence and interactions within a network.

When the facility of anywhere computing becomes the rule of the day, one of the main issues it raises is energy consumption. This is due to the

following reasons: (i) majority of the devices engaged in ubiquitous computing are mobile wireless devices, (ii) mobile wireless devices depend on the attached battery for their power requirements, (iii) the challenge of “packing” energy with the limited “real estate” available with a hand held device, and (iv) even if significant improvements have been achieved in the battery technology, the pace of development is not in par with the developments in the microprocessor design (Christopher 2014).

Due to the inherent deficiencies of battery power, devices engaged in ubiquitous computing have further headaches concerning their battery power. The energy consumed by the devices is depending on many factors such as the type of gadgets used in the assembly/make of the device, type of usage of the device (mobile or fixed), type of applications and networks onto which the device is getting connected (Internet Users, 2015).

In this chapter, the scope is limited to the energy consumption issues related to the web connectivity. This is due to the fact that Internet is the most viable and readily available infrastructure to hook the entire world into a web. In the era of ubiquitous computing, the majority of the users are getting connected to the Internet for their heterogeneous types of needs and applications. As of July 2016, the total number of Internet users in the world crossed three and half billion – 4% of the world’s population (Internet Users, 2016).

BACKGROUND

Main cause of energy consumption especially in hand held devices are usage of Internet. In this

section a brief account of the internet usage and energy consumption is outlined.

INTERNET USAGE

The popularity of smart devices used in the ubiquitous computing continues to grow at an exponential rate. This is due to the affordability of devices and availability of 3G, 4G and 5G networks. Currently there are as many mobile devices all over the world, as there are computers. The availability of smart phones is also increasing to the extent that global smart phone activations outnumber global child births by three to one (Tom, 2014). Trends reveal that smart phone customers are expected to reach 2.08 billion by 2016 (Statista, 2016).

This tremendous growth of devices (smart phones and feature phones with Internet connectivity) has its impact over the Internet traffic also. The Internet traffic is steadily increasing at the rate of 66% per year (Internet Live Stats, 2015). In the past five years Internet traffic has quintupled (World Wide Web, 2016). It is estimated that monthly global mobile data traffic will surpass 15 exabytes by 2018 (Cisco, 2016). Internet offers a bunch of services to the users at an affordable cost. The popularity of the services offered by the Internet is an indicative to the degree of the obsession the users across the world have developed with the Internet.

World wide web (WWW) is the foremost service offered by the Internet. The ease of use, easy access to the WWW and wide popularity of devices which can be used to access the Internet caused an exponential growth in terms of number of users logged on, devices used and data shared over the websites. WWW, one of the biggest repository of information in the world, is getting expanded in terms of its variety, veracity, velocity and volume - from just one web site in 1991 to 1 billion web pages in 2014 (World Wide Web, 2016).

When there is a variety of information available at WWW, the process of finding the required web pages will also increase. Search engines are

widely used to locate the pages of one's interest. It is estimated that Google processes over 40,000 search queries every second on average - over 3.5 billion searches per day and 1.2 trillion searches per year worldwide (Google Search Statistics, 2015). The types of queries being fired are unique in the sense that 16% to 20% of queries that get asked every day have never been asked before (Mitchell, 2012). Queries made using mobile devices are also increasing day by day. It is found that the time spent on Internet using mobile devices in USA exceeds the time spent using desktop PCs during February 2013 and January 2014 (Rebecca 2014). Studies reveal that more than 90 percent of Internet users will access online content through their phones in 2017 (Internet Usage Statistics, 2015).

The evolution of web technology from read-only static pages to readily updatable dynamic pages has contributed a lot to the increased traffic over Internet – this evolution gave birth to a class of web sites known as social media. Social media is being effectively used as a platform for many innovative ideas and needs such as political campaigns, job applications, business promotion and networking, religious campaigns, educational purposes, and customer services. One of the reasons behind the popular use of WWW is the emergence of social media.

As of July 2014, there are 106,000,000 active social media users in India. However this is not a big number considering the fact that out of the 1.25 billion people in India, only one-fifth of the people use Internet and only 50% of these are using social media (Bhavya, 2014). But when the worldwide activities are considered, the statistics are astonishing. Number of postings made by the users on Facebook reached nearly 2.5 million. Contents uploaded to YouTube amounts to 72 hours of content. Number of tweets on Twitter crossed 277,000, Vine users shared 8,333 clips, WhatsApp users share 347,222 photos, and Instagram users share 216,000 new photos per minute (Kimberlee, 2014). In addition to this people also make use of WWW for their needs such as shopping, attending seminars and online courses.

Email is another important service offered by Internet. Both text and multimedia attachments can be accompanied with an email message. An email message can be replicated to a large set of users. Internet users send more than 200 million emails every minute (Kimberlee, 2014). Number of mobile email users in the world is expected to reach 2.83 billion by the end of 2017 (Karl, 2013).

Short Messaging Service (SMS) is a service offered by mobile service providers as a cheap and user-friendly service to exchange messages. Internet based services are also becoming viral now a days that can be used to exchange multimedia information. Such services are called as over-the-top content (OTT) services. Estimates reveal that SMS traffic worldwide will peak at 8.3 trillion messages annually in 2015 (Kimberlee, 2014). But due to the popularity of OTT messaging apps, volume of messages sent by these apps are also increasing. It is predicted that by the end of 2017, there will be over 2.5 billion OTT messaging apps users. Subsequently the number of OTT messages are also increasing - OTT message traffic will exceed 32 trillion messages annually by 2017 (Kimberlee, 2014). Available trends indicate that eventually OTT messages will surpass SMS messages – it really happened in UK in 2014 (Luke & Sean, 2014).

It can be concluded that the explosive growth of Internet and expansion of ubiquitous computing have complemented each other – in terms of the number of devices, users and network traffic.

CONSUMPTION OF ENERGY

Energy being consumed is to be perceived from two angles – (i) the energy required to maintain the Internet infrastructure and WWW and (ii) the energy consumed by the device using a service offered by Internet.

Energy Consumed by Internet

Physical Internet consists of 1.6 billion connected PCs and notebooks and 6 billion mobile devices (Tessel 2013). Even if energy efficiency of com-

puting has doubled every 1.5 years, the combined use of energy consumption (80% at the time of use and 20% at the time of manufacturing) is a concern. The second largest component of Internet is the data centers. Data centers are the cloud infrastructure of the Internet used by the service providers to maintain the web servers and data. Estimates reveal that assuming low power cost a typical data center costs roughly \$7 million per megawatt to build, and \$9 million per megawatt for the cost of electricity (Mills, 2013). For instance, Google's data centers use 260 million watts - enough to power a city of 100,000 to 200,000 people (Daily Mirror, 2011). It is estimated that the Internet uses 84 to 143 gigawatts of electricity every year, which amounts to between 3.6 and 6.2 percent of all electricity worldwide (Jonathan, 2012) (Michael, 2014).

There are two perspectives of the energy consumed by the Internet. One is that considering the mammoth amount of the total energy consumed by the Internet and the global energy challenges, the energy consumed by the Internet is significant. The other perspective is that it is consuming only about a small fraction of the total—between 1.1% and 1.9% of the 16 TW used by humanity (Barath & Justin, 2011). The objective of this chapter is to consider how energy can be conserved from the point of view of end users and service providers, however small may be the contribution made by these conservation measures.

Energy Consumed by Devices When They Use Internet Services

This discussion can be sub divided into two categories: (i) energy consumed by the devices when they use Internet, and (ii) energy consumption by the Internet while a device makes use of an Internet service.

Energy Consumed by the Devices When They Use Internet

Several studies are available assessing the energy consumed by a device while it accesses a service offered by Internet. It is clear that considerable

time is spent by a user in getting connected with the Internet and naturally a browser will be one of the software that will be in use for the majority of these sessions. Studies reveal that the Internet browsers increased energy consumption by about 7% to 13% for notebooks and 3% to 5% for desktops. HTML 5 websites with flash videos also increased the power consumption considerably (Kurt, Shreyas & James, 2013). Energy consumed by the components of web pages such as images, client side scripts and cascading style sheets (CSS) also contribute to the net energy consumed by the machine. It is found that the amount of energy used to render images is proportional to the number and size of images on the page. For instance 24.22% of the total energy to render the YouTube page is spent on images, as there are 5 large images representing a screenshot from each video. Gmail, in contrast, take only 6.86% total energy as it contains only small GIFs (13 pixels wide) and images. Rendering energy needed to process the JavaScript also contribute to the energy consumption. Amazon (with a large complex script, which is to be downloaded to the users' devices when accessed by them) consumes 16.89% of its rendering energy for handling JavaScript. Yahoo's JavaScript code (amount of code is less, but the code is executed every time the page loads) processing takes only 6.79% of the total rendering energy. The rendering cost of CSS depends on the number of items styled using CSS. Amazon (with 104 items to be styled, complex styling in the page like a color fade out effect, horizontal bars to show product ratings, and huge wide buttons) has a high CSS rendering cost (17.57%) compared to a 3% of the total rendering energy cost for Gmail (simple styling, small size and defined in an internal style sheet) (Narendran, Gaurav, Angela, Dan, & Jatinder, 2014).

There are hidden energy costs when websites with advertisements are browsed. The energy consumption will have to be done by the client machine. It is found that online advertisements are highly CPU – intensive, even if they consume only a small space on the page. It is found that (i)

energy consumption is increased by 12% when websites with advertisements are browsed, (ii) when number of open tabs are more, more energy is consumed, (iii) both Apple Safari and Microsoft Internet Explorer 8 seem to require more energy than Firefox, and (iv) the total energy used for displaying web advertisements is equivalent to the total yearly electric energy consumption of 1891 Dutch households (Simons & Pras, 2013).

Internet connection devices like modems and routers also contribute a lot to the power consumption. It is estimated that in US, these devices consume 8.3 billion kilowatt-hours a year (Jonathan, 2012).

Energy Consumption by the Internet While a Device Makes Use of an Internet Service

When a device is hooked to Internet to avail a service, lot of energy consumptions are done by the Internet infrastructure, in addition to the power consumed by the device itself. Estimates revealed that per second spending of browsing a simple web site resulted in the generation of 20 mg of CO₂. Demand for data by the users is also getting increased – it is predicted that average demand for data for each person will be 32GB of data per day – 2,570 exabytes (1018 bytes) per year for the global population. The average power needed to sustain such activity would be 1175 gigawatts, equivalent of 1,175 large coal-red power plants (Ruzena, 2013).

When a search is made to access a desired page in WWW, considerable energy is consumed by the data centers to provide us the answer. For instance, each Google query has to travel on average 1,500 miles to a data center and back to return the answer to the user (Mitchell, 2012). Estimates done by the Google reveal that 100 searches on google.com uses equal power to light a 60 watt bulb for 28 minutes (Daily Mirror, 2011). Estimates reveal that Google spends about 0.0003kWh of energy on an average search query, translating to roughly 0.2g of carbon dioxide – hence one Google search

is equal to turning on a 60W light bulb for 17 seconds (Jared, 2011).

Email messages contribute a lot to the global energy consumption. An Email of 1 MB would result in the emission of 19 grams of CO₂. When the message is copied (cc'd) to 10 people, its impact is 73 grams of CO₂. Emails with attachments make energy consumption still worse – an e-mail attachment of 4.7 megabytes (MB) creates as much greenhouse gas as boiling a tea-kettle 17.5 times. This is due to the fact that when an email is sent it has to be maintained in the data centers and it is read/downloaded by the recipients concerned. It is found that in a 100-people company where each employee sends on average 33 e-mails a day and receives 58, the greenhouse gas emission linked to emails would be around 13.6 tons of CO₂ per year. If there is a 10% reduction in the Email for such a company, they would save CO₂ emissions equivalent to one round-trip flight between Paris and New York (Balasubramanian D., 2011) (Vita, 2011).

Apart from transmission of Email, tasks such as maintenance of Inboxes, handling of proper mails and sorting of spam mails also contribute to the carbon emissions. The energy consumed increases when the Email Inbox contains more number of mails. It is estimated that 80% of this electricity is consumed by the reading and deleting of spam and the identifying genuine emails that that ended up in the spam folder accidentally. The carbon footprint of a genuine email, spam email and an email with an attachment is estimated as 4g, 0.3g and 50g of CO₂ emissions, respectively (Mike & Duncan, 2010).

When the frequency of updations done thru social media sites are also considered, it can be seen that the per minute energy consumption and subsequent carbon emissions caused is quite alarming. It is found that three days of streaming YouTube video requires as much energy as making, packaging and delivering a DVD (Daily Mirror, 2011). Keeping in mind the energy requirements for email service, one can assume the energy consumption required to maintain social media sites.

For instance, in 2013 Facebook's carbon footprint per person on the site was 0.000311 metric tons of CO₂ and the total power consumed was 822M kWh (Carbone, 2010). The energy consumed in transmitting 1GB of data is estimated to be 19kW - considering the huge number of images that are shared through social media sites, net energy consumption due to this also become considerably large (Bryan, 2013). Hence it can be concluded that the requirements of for power is getting increased day by day, due to the wide popularity of Internet usage.

FUTURE RESEARCH DIRECTIONS

The demand for increased energy requirements in the era of ubiquitous computing and subsequent carbon emissions pose several challenges to the research community. A multifarious attempt is to be carried out both from the perspective of users and service providers in this regard: (i) even if the energy consumption statistics of search giants like Google look alarming, these companies are investing heavily in the generation of renewable energy sources. Considerable efforts are to be done to make the technology for renewable energy cost effective and viable, (ii) regarding the power requirements of wireless mobile devices, constant efforts are required to pack more energy to the limited size of battery – nanotechnology, biocomputing and computing using light may provide reprieve to these issues, (iii) search algorithms are to be fine tuned to provide the exact reply to the query being provided by the users, so that the user is tend to end up the query with the first attempt of the search itself, (iv) efforts to minimize the size of text messages and images without compromising the quality of the content is to be addressed, (v) energy aware methodologies in the web page development for desktop and mobile devices separately, and (v) efforts in developing components for computers and networks that consume less energy (Balasubramanian, 2011) (Narendran *et. al.*, 2014) (Ruzena 2013). It is

worth mentioning that several works are already reported in this regard. A detailed survey listing the contributions made so far is outside the scope of this chapter, hence it is not discussed.

Browsers have to be more power aware when dealing with the issues such as handling client side script, CSS and images.

From the user's side an awareness is to be generated among them to point out that everything won't go green when one goes for online services: (i) the memory space in the computer is to be freed regularly, (ii) e-mail boxes (in and out mails) are to be cleaned periodically, (iii) the number of recipients for each e-mail are to be limited, (iv) reducing the size of the attachments (boil less tea- water) and uploads, (v) entering the URL address directly rather than use a search engine for every access to the same website, (vi) don't leave computer and accessories on overnight (as many offices do), not even on 'sleep mode' (even if that eats up only 1-10 watts), laptops use 15-60 watts while desktops use 250W, (viii) reduce the power by doing more 'offline' work than online, and (ix) remember Facebooking and Twittering burn carbon and make CO₂ (Balasubramanian D., 2011) (Narendran T., Gaurav A., Angela N., Dan B., and Jatinder P. S., 2014) (Ruzena C., 2013). Users have to a judicious use of these blessings – only when needed, in the right context and in the appropriate measure.

CONCLUSION

Easy availability and affordability of devices and accessibility to wireless networks, there is an exponential growth in the number of devices in use. Popularity of feature rich mobile devices that can be hooked to the wireless networks among the users all over the world resulted in vanishing the barriers of locality and time – as far as computing is concerned. This gave birth to an era of computing known as ubiquitous computing.

Ubiquitous computing and availability of Internet is just revolutionizing the way people make

use of computing devices. Predications revealed that the monthly global mobile data traffic will surpass 15 exabytes by 2018. Popularity of mobile devices and increased use of Internet poses several issues and challenges to the research fraternity. One of the challenges is the energy consumption by the devices and the Internet backbone.

Energy consumed is to be perceived from two angles – consumption by the Internet infrastructure and devices being connected to the Internet by the users. It is found that Internet infrastructure uses 84 to 143 gigawatts of electricity every year, which amounts to between 3.6 and 6.2 percent of all electricity worldwide. Every activity on the Internet such as getting search results from search engines (Google processes over 40,000 search queries every second on average - 1.2 trillion searches per year), sending emails (200 million emails every minute), engaging in social media websites (number of postings made by the users on Facebook reached nearly 2.5 million, contents uploaded to YouTube amounts to 72 hours of content, and Instagram users share 216,000 new photos per minute) consumes considerable amount of energy (100 searches on google.com uses equal power to light a 60 watt bulb for 28 minutes, emails with attachment of 4.7 megabytes creates as much greenhouse gas as boiling a tea-kettle 17.5 times, for instance). In addition, the devices also consume energy while they are in use.

The growing energy consumption and the attempts to conserve energy poses several challenges to the research community such as (i) renewable energy is to be made cost effective and viable, (ii) more energy is to be packed to the limited size of battery space, especially on a mobile device – technologies such as nanotechnology, bio-computing and computing using light may provide efficient solutions, (iii) search algorithms are to be fine tuned to provide the exact reply to the query being provided by the users, so that the user is tend to end up the query with the first attempt of the search itself, (iv) size of messages (text, images, audio and video) are to be minimized without compromising the quality of the content, (v) design

of energy aware methodologies for developing web pages for desktop and mobile devices separately, and (v) efforts in developing components for computers and networks that consume less energy. Conscious efforts should be made from the users' part also to reduce the energy consumption by at least limiting the use of technology considering need, necessity and priority.

This work is indented to bring forth the issues and challenges in the area of energy consumption in the era of ubiquitous computing. A detailed study is to be undertaken to assess the progress of the works being reported in the energy conservation measures – this will help one to understand the state of the art and frame new solutions and approach. This is left as the future work.

REFERENCES

- Adam, G. (2006). *Everyware: The Dawning Age of Ubiquitous Computing*. New Riders Publishing.
- Artur, L., Thomas, R., Bjoern, S., Kari, L., & Juha, K. (2009). Semantic Ambient Media—an Introduction. *Multimedia Tools and Applications*, 44(3), 337–359. doi:10.1007/s11042-009-0282-z
- Balasubramanian, D. (2011). Speaking of science - E-mails not all that 'green'. *The Hindu*. Retrieved October 13, 2015, from <http://www.thehindu.com/sci-tech/energy-and-environment/speaking-of-science-emails-not-all-that-green/article2612323.ece>
- Barath, R., & Justin, M. (2011). The Energy and Emergy of the Internet. In *ACM SIGCOMM Workshop on Hot Topics in Networks (HotNets-X)*.
- Bhavya, S. (2014). *Mobile and Internet In India 2014: 349 Million Unique Mobile Phone Users, 70% Traffic From Mobile*. Retrieved August 17, 2014, from <http://dazeinfo.com/2014/07/11/mobile-internet-india-2014-349-million-unique-mobile-phone-users-70-traffic-mobile-india-shining-infographic/>
- Bryan, W. (2013). *The Surprisingly Large Energy Footprint of the Digital Economy*. Retrieved November 8, 2015, from <http://science.time.com/2013/08/14/power-drain-the-digital-cloud-is-using-more-energy-than-you-think/>
- Carbone. (2010). *Carbone and Energy Impact*. Retrieved October 13, 2015, from https://www.facebook.com/green/app_439663542812831
- Christopher, M. (2014). Tech World Vexed by Slow Progress on Batteries. *The Wall Street Journal*. Retrieved January 15, 2016, from <http://www.wsj.com/articles/tech-world-vexed-by-slow-progress-on-batteries-1412552377>
- Cisco. (2016). *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2013–2018*. Retrieved February 1, 2016, from http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white_paper_c11-520862.html
- Google Search Statistics. (2015). *Google Search Statistics*. Retrieved October 13, 2015, from <http://www.internetlivestats.com/google-search-statistics/>
- Hansmann, U., Merk, L., Nicklous, M. S., & Stober, T. (2003). *Pervasive Computing: The Mobile World*. Berlin: Springer-Verlag.
- Internet Live Stats. (2015). *Internet Live Stats*. Retrieved March 15, 2015, from <http://www.internetlivestats.com/total-number-of-websites/>
- Internet Usage Statistics. (2015). *Statistics and facts on Mobile Internet Usage*. Retrieved March 29, 2016, from <http://www.statista.com/topics/779/mobile-internet/>
- Jared, N. (2011). *6 Things You'd Never Guess About Google's Energy Use*. Retrieved November 13, 2015, from <http://techland.time.com/2011/09/09/6-things-you-d-never-guess-about-googles-energy-use/>

- Jonathan, S. (2012). *How much energy does the Internet use?* Retrieved May 10, 2015 from <http://computer.howstuffworks.com/internet/basics/how-much-energy-does-internet-use.htm>
- Karl, W. (2013). *17 Incredible Facts about Mobile Messaging that you should know*. Retrieved December 24, 2013, from <http://www.sentextsolutions.com/2015/09/17-incredible-facts-about-mobile-messaging-that-you-should-know/>
- Kimberlee, M. (2014). *What Happens in One Minute on the Internet? 70% Traffic From Mobile*. Retrieved May 2, 2014, from <http://www.adweek.com/socialtimes/one-minute-internet-infographic/148860>
- Kurt, R., Shreyas, P., & James, P. (2013). *The Impact of Internet Browsers on Computer Energy Consumption. Technical Report to Microsoft*. Fraunhofer Center for Sustainable Energy Systems.
- Luke, G., & Sean, P. (2014). *Number of text messages being sent falls for the first time ever as more people turn to Whatsapp and iMessage*. Retrieved March 8, 2015, from <http://www.dailymail.co.uk/sciencetech/article-2538488/SMS-takes-seat-IM-number-texts-sent-Britain-falls-time.html>
- Mark, W. (1999). The Computer for the 21st Century. *Mobile Computing and Communications Review*, 3(3), 3–11. doi:10.1145/329124.329126
- Michael, R. (2014). Internet science - energy consumption and optimization. *Proceedings of the Seminars Future Internet (FI) and Innovative Internet Technologies and Mobile Communications (IITM)*, 83 - 89.
- Mike, B., & Duncan, C. (2010). *What's the carbon footprint of... email?* Retrieved October 13, 2015, from <http://www.theguardian.com/environment/green-living-blog/2010/oct/21/carbon-footprint-email>
- Mills, M. P. (2013). How Green Is The Internet Wrong Question. How Big Is The Internet And Where Do I Invest? *Forbes*, Retrieved May 8, 2015, from <http://www.forbes.com/sites/mark-pmills/2013/08/11/how-green-is-the-internet-wrong-question-how-big-is-the-internet-and-where-do-i-invest/>
- Mirror, D. (2011). Google uses MORE power than Salt Lake City as vast data farms suck up electricity. *Mail Online*. Retrieved April 23, 2015, from <http://www.dailymail.co.uk/news/article-2035382/Google-discloses-energy-consumption--power-Salt-Lake-City.html>
- Mitchell, J. (2012). *How Google Search Really Works*. Retrieved February 28, 2012, from http://readwrite.com/2012/02/29/interview_changing_engines_mid-flight_qa_with_goog#awesm=%7EoiNkM4tAX3xhbP
- Narendran, T., Gaurav, A., Angela, N., Dan, B., & Jatinder, P. S. (2014). Who Killed My Battery: Analyzing Mobile Browser Energy Consumption. *Proceedings of WWW 2012 – Session: Mobile Web Performance*.
- Rebecca, M. (2014). *Mobile Now Exceeds PC: The Biggest Shift Since the Internet Began*. Retrieved July 8, 2014, from <http://searchenginewatch.com/sew/opinion/2353616/mobile-now-exceeds-pc-the-biggest-shift-since-the-internet-began>
- Ruzena, C. (2013). *Environmental cost of spending time on the Internet* (Master Thesis). Retrieved October 13, 2015, from <http://www.pagescroll.com/internetenergycost.pdf>
- Simons & Pras. (2013). *The Hidden Energy Cost of Web Advertising*. Technical Report TR-CTIT-10-24, Centre for Telematics and Information Technology, University of Twente, Enschede.
- Statista. (2016). *Number of Smartphone users worldwide from 2014 to 2019 (in millions)*. Retrieved September 9, 2016, from <http://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/>

Tessel, R. (2013). *How Much Electricity Does the Internet Use?* Retrieved October 13, 2015, from <http://www.techthefuture.com/technology/how-much-electricity-does-the-internet-use/>

Tom, V. (2014). *The planet has reached the mobile moment: Top 20 mobile growth stats*. Retrieved October 13, 2014, from <http://asmarterplanet.com/mobile-enterprise/blog/2014/01/planet-reached-mobile-moment-top-20-mobile-growth-stats.html>

Users, I. (2016). *Internet Users*. Retrieved September 8, 2016, from <http://www.internetlivestats.com/internet-users/>

Vita, S. (2011). *The true impact of email*. Retrieved October 13, 2015, from <http://www.earthtimes.org/scitech/true-impact-email/1147/>

Web, W. W. (2016). *World Wide Web*. Retrieved January 21, 2016 from http://en.wikipedia.org/wiki/World_Wide_Web

ADDITIONAL READING

Li, S., Da Xu, L., & Zhao, S. (2015). The internet of things: A survey. *Information Systems Frontiers*, 17(2), 243–259. doi:10.1007/s10796-014-9492-7

KEY TERMS AND DEFINITIONS

CO₂ Emission: Primary green house gas ejected as a result of human activities.

Green Computing: Usage of computers in environmentally responsible and eco-friendly manner.

Ubiquitous Computing: Anywhere computing.

From General Services to Pervasive and Sensitive Services

Mario Vega-Barbas

KTH-Royal Institute of Technology, Sweden & ESNE-Universidad Camilo José Cela, Spain

Iván Pau

Universidad Politécnica de Madrid, Spain

Fernando Seoane

KI-Karolinska Institutet, Sweden & University of Borås, Sweden

INTRODUCTION

The choice, development and use of technologies has very varied impacts on the progress of human beings and nature. In some cases, this impact leads to a radical change in the context of users of a specific technology, both in social and cultural terms. This technological innovation or invention is labelled enabling technology and is mainly characterized by the rapid development of subsequent technologies, often present in several scientific and social fields.

In this sense, ubiquitous computing, pervasive technology and smart spaces represent enabling technologies which have conditioned the evolution of the current Information Society (IS). These ICT boosting the development of Internet of Things (IoT), wearable sensing, Cloud Computing, Smart Textiles, etc., which are transforming society and particularly changing the interfacing methods between services and user as well as their expectations from services. Such changes are related to arising of new kind of pervasive and sensitive service such as new healthcare and well-being personalized services or the final irruption of the Smart Home.

In this article such kind of services are defined in a comprehensive manner including their general objective, limitations, potential interrelationship,

etc., including the supporting literature and reported related work but not limited to a mere review of the available literature. This sense, sensitive services refers to new IS services that manage data that is not only based on the physical disposition of the user in regards to the environment (such as geographical position, interior positioning, or patterns of movement) but also related to more personal aspects of their physical state, psyche, or behavior patterns. On its part, pervasive services are those developed by using pervasive technology such as cloud computing, Internet of Things approaches and therefore can be defined as a delocalized and persistent ICT service that allows its users to access the functionalities it offers at any time or place. These definitions will be supported by an analysis of the evolution of services of the IS both from a socio-economic and industrial focus.

BACKGROUND

The IS refers to the social model in which the means of wealth generation are based on the production, management and transmission of information (Sarell & Machlup, 1963). In the IS, information and knowledge are established as the main economic engine, giving greater importance to services over production and products. Infor-

mation and knowledge technologies (ICT) are an essential tool for the development of IS services. Their main objectives are the generation of knowledge and the exchange of information, through its encapsulation in the form of services to ensure its efficiency and accessibility to the citizen.

The services deployed in this society have undergone a continuous evolution to adapt to the needs of society and the specific characteristics of the technology used for its development. The incidence of these services in the IS goes beyond the technical field, and its evolution has been influenced, not only by the technology used in its development, but also by the socio-economic factors related to new forms of communication and global coexistence.

In (Achrol & Kotler, 2014) services are defined as the application of specialized competence by an entity through actions, processes and activities for the benefit of another entity or the own. This concept, in the context of the IS, refers to any set of features provided at a distance, electronically and by individual request from the recipient of the service (Directive 98/48/EC of the European Parliament and of the Council). To achieve this, ICTs are used as a means of developing the transmission, processing, and storage of the data associated with it.

In contrast to a specific product, the services deployed in the IS are specified at the time they are required by the user. The specifications or design requirements of the services can and must be defined a priori, but their implementation, deployment and behavior are subject to their use. This means that the services are tailored to the needs of the moment when they are required. This situation requires that the users can control and understand the elements involved in the use of the service at the time in which they are accessed, regardless of the technical complexity and underlying technology. In this way, the IS offer their users increasingly complex and useful services whose efficiency and effectiveness is subject to the relationship of communication between them. Given the complexity of nuances in many existing

services deployed, a detailed study and design of all of the elements of the service is necessary. Within these elements, the communication relationship between users and services is of particular importance for two basic reasons related to their implementation in society and the management of the inherent complexity of these services. On the one hand, the ever increasing inclusion of these services in the daily lives of people means that these affect more intimate and personal aspects of users, which could generate rejection or fears that did not exist initially. On the other hand, a more complex functionality that addresses closer needs to the user implies an increased degree of service sophistication and interconnection.

To better understand the importance of the relationships between users and the services of the IS it is necessary to analyze its evolution, from a more focused approach in the management of the information until a conception aligned with enabling technologies such as ubiquitous computing, pervasive technology or intelligent environments.

The Evolution of Services in the Information Society

As has already been mentioned, the evolution of services in the IS has been marked not only by the ICT used in its definition, but also by the needs and wants of the citizens that make up the society of the availability of the basic resources needed for its proper development.

Throughout this section, the evolution of services in the IS will be presented from two perspectives. The first analyses said evolution from a socio-economic focus in order to understand how key aspects of today's society and its needs have influenced the development of these new services in the IS. The second presents a view of the evolution of development efforts and investments made by the big ICT companies over different sectors of society in order to create new services in the IS.

Although development efforts for these services and their evolution are currently oriented

towards various areas of people's lives, book chapter is centered on two fundamental aspects, health and the intelligent home. Health and social well-being represent a group of services with an extremely high tendency to impact society and industry. Intelligent homes offer an important tool for enabling the spread of healthcare and personal wellness services.

Socio-Economic Focus

An ever-higher quality of life and an ageing population paint a picture where the number of users of health and wellness services is constantly increasing. Health assistance and wellness services play an increasingly critical role in consumer and, especially, State expenditures. This picture, along with the cyclical economic crisis situations that our society suffers from, forces States to redirect the distribution method for this type of services towards more sustainable methods in regards to time and resource use.

Enabling technologies, such as ubiquitous computing and pervasive technology, have allowed health assistance to be transferred into the daily context of users. This new paradigm of service distribution, along with the expected capacity of the Digital Home, are considered promising and more sustainable alternatives to traditional hospital-centered developments. Thus, the telehealth industry has turned into a competent complement to traditional health services and, as seen earlier, also into an opportunity for the development and distribution of new services by the ICT industry. In 2013, this industry generated US \$240 million in revenue through three primary means of care: store-and-forward, programmed real time, and emergency real time (GrowthCap, 2014). According to this, the telehealth market is about to grow to an annual cumulative rate of 56%, to \$1.9 billion in 2018. This implies that the number of patients using telehealth will grow from 250 thousand in 2013 to 3.2 million in 2018.

This evolution has created a change in vision for service development from a more complete and

complex concept to a simple and more accessible solution. In this way, the acceptance of services by users is fostered, while improving its distribution in society and, thus, its socio-economic impact. The application of mobile technologies, such as smart phones or tablets, as well as enabling technology, such as pervasive technology and ubiquitous computing, has favored the inclusion of companies that are foreign to the sector and the participation of governments as support entities.

If the expectations for telemedicine are, in fact, promising, the reality is that, up until now, the telemedical solutions distributed had achieved varying levels of success. In addition, this variability in the success of Telemedicine occurs in disparate development indexes. Even now, Telemedicine is used in a residual manner as a method for rendering routine services and only a few pilot projects have been able to sustain themselves once the development phase of the prototypes has finished.

The majority of studies related to the analysis of the true distribution of Telemedicine solutions show that patients and service providers are resistant to adopting service models that differ from the traditional focus or local practices, while others lack an ideal level of necessary knowledge and technological culture to use the new focus effectively. The most notable complex problems are related to linguistic and cultural differences between patients (particularly the underserved) and service providers (Al-Shorbaji, 2008; Craig & Patterson, 2005; Currell, Urquhart, Wainwright, & Lewis, 2010).

Another cause related to real Telemedicine service distribution is that it has not been a priority for most governments. In a 2010 global e-health study carried out by the World Health Organization (World Health Organization, 2010), only 52% of health and wellness services offered via Telemedicine had been adopted and established in countries by their governments. However, it seems that currently, governments have become more conscious of the need for this type of services. States are primarily concentrating their economic efforts on the investigation and creation of services

aimed at healthcare and wellness, including Telemedicine. In the European Union, Eurobarometer 419 (European Commission, 2014), researching European citizens' perceptions on the topics of science, research and innovation, established that the evolution of the services in the IS dedicated to the promotion and protection of health and wellness represent, along with job creation, the main priority for science and technological innovation for the next 15 years. Figure 1 shows an analysis about the priority of research and development-based on 13 basic aspects of current society and the expected positive impact of research and technology with regard to the expected impact of each citizen's individual action.

It becomes clear that current society is aware that the inclusion of ICTs to create effective and

useful services represents a necessary aspect of its development. Nonetheless, the manipulation of private and personal data performed by this type of services, together with its inclusion in the daily life of users in a pervasive manner (clothing, home, work spaces, schools), generates a certain level of rejection and fear that inhibits its complete adoption. Governmental agencies are aware of this and continue working on legislation for these sensitive services as a means of user protection since the mid-90s. As will be analyzed in the following sections, adopting this type of services is determined by the factors intervening in the development of sensitive services and how the enabling technologies presented in this chapter affect the storage and management of the information they control.

Figure 1. Comparative graph of the expected impact of science and technology and the individual actions of each citizen for the 13 most representative areas of current society's activity (European Commission, 2014)

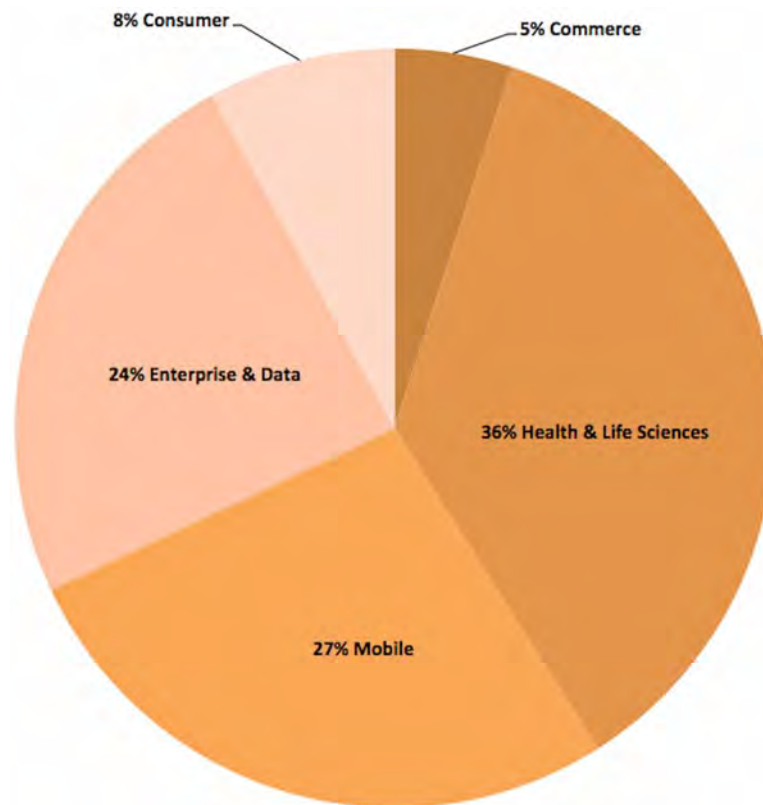
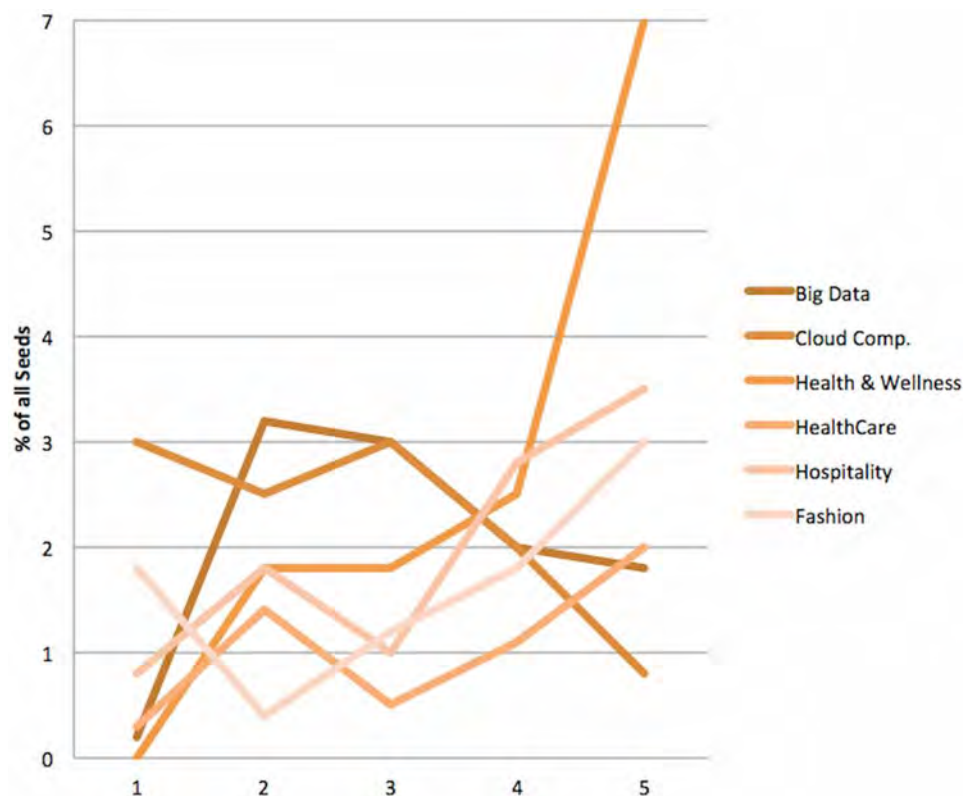


Figure 2. Trends in seed capital investment from 2010 to 2014 in the USA (Data source: Crunchbase)



Industrial Focus

Market movement and business investing from the largest companies in the ICT arena offers an addition perspective on understanding how services in the IS have evolved over the years. Figure 2 shows investment trends by business sector in the group of all American start-ups in their creation phase from 2010 to 2014. It can be seen that the traditional sectors related to Hardware and Software development, instant messaging, education, and mobile technology maintain a high investment index. However, movement towards an increased investment in other sectors can be discerned, marking an evolutionary trend in the services in the IS. The fashion sector, covering manufacture of intelligent fabrics, among other things (Fashion), E-Commerce, Hospitality and Digital Home development (Hospitality), and sectors related to health and wellness (Healthcare and Health and Wellness) have received a larger

rate of investment since 2010, putting them on par with the traditional sectors.

Large companies such as Google, Microsoft, Apple, etc., have followed this trend. In case of Google, the services developed originally were oriented towards the need and desire of its users to share and acquire knowledge, and it was managed and organized with a powerful search engine, the symbol and brand of the company. However, during the 2010-2015 period, Google has reoriented its development and acquisition of external companies towards the study of life sciences and development of Digital Homes. As shown in Figure 3, in 2014, investments made by Google Ventures, a sub-company dedicated to the acquisition of other companies as subsidiaries of Google, exceeded 35% in the Health and Life Sciences sector, putting it above even the mobile phone sector. This trend materializes in the creation of a highly protected division, Google[X], aimed at producing major technological advances,

Figure 3. Cont. Trends in seed capital investment from 2010 to 2014 in the USA (Data source: Crunchbase)

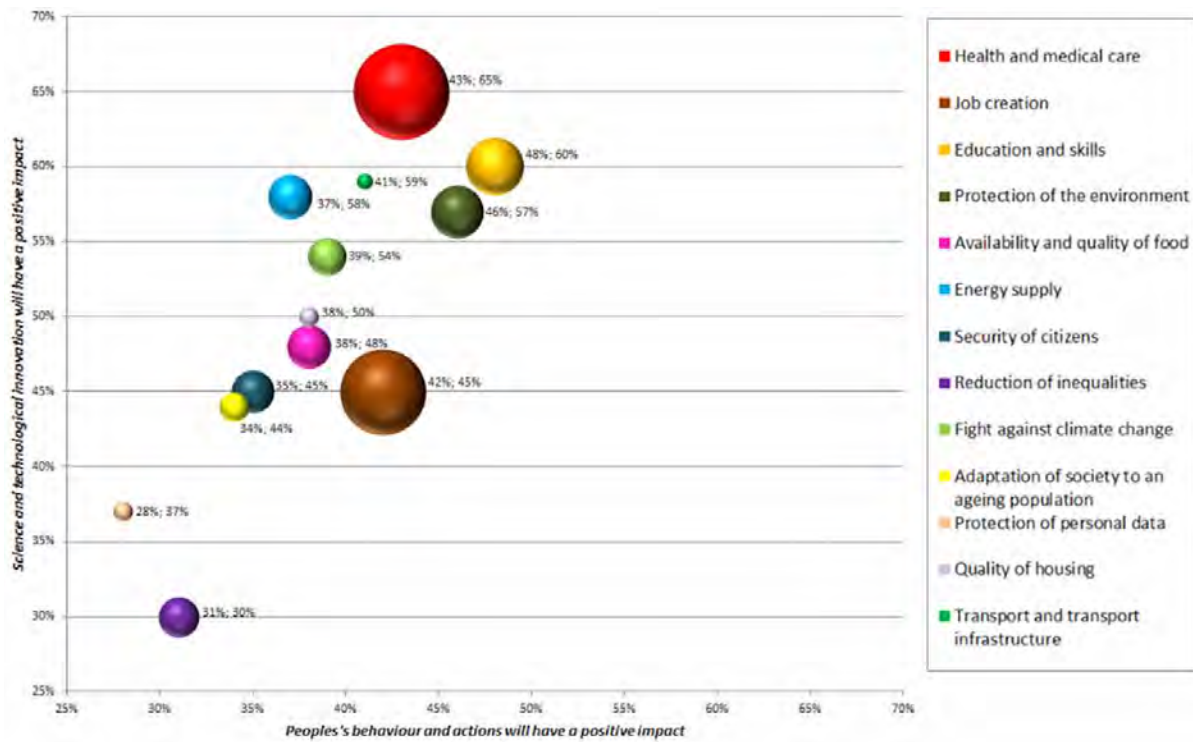
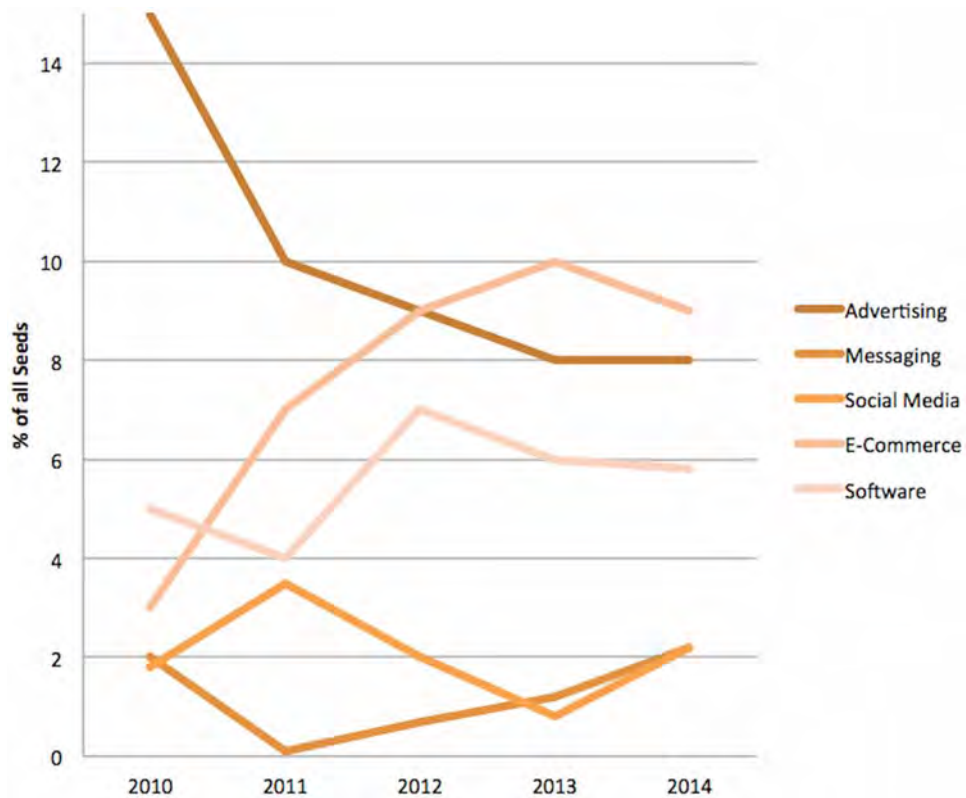


Figure 4. Investments in 2014 by Google Ventures (The New York Times)



emphasizing its line of research and innovation in the study of life sciences.

Other relevant ICT sector companies such as Microsoft, Apple, or IBM have also evolved their developments towards a convergence with healthcare. Specifically, Microsoft created the company Avanade alongside Accenture in 2010 with the objective of covering the needs of society focusing on health and wellness. With this initiative, Microsoft's technologies such as Kinect, Skype, and HealthValue (a tool based on services on the cloud for the exchange of patients' clinical histories) are joined together. In Apple's case, the development trend for this type of sensitive services materializes in their bet on information processing in regards to health through HealthKit, available on all its new devices. Lastly, IBM has aimed its technical and technological capacities towards clinical data analysis and development of technical tools for improving the healthcare process, an idea that has been captured in the interoperability framework IBM Watson Health.

Another sector where the biggest representatives of the ICT sector have overwhelmingly decided to invest since 2010 is the development of Digital Homes. In 2014, Google acquired the companies DropCam and NestLabs, cutting-edge enterprises in the development of solutions for the Digital Home. In this way, Google attempted to turn itself into a reference source in this field through development of systems and services for the Digital Home alongside adapting and using its mobile operating system, Android. In addition, Microsoft or Apple, who in the beginning were aimed towards the development of tools for knowledge processing or communications, have seen the opportunity that the Digital Home offers. In this way, new systems and technological developments aimed at this sector have appeared, such as HomeOs (Microsoft Research), an ad-hoc operating system aimed at device management, communication frameworks for Digital Home elements such as HomeKit (Apple Inc.), AllJoyn (Linux Foundation), and Thinking Things (Tele-

fónica), or research and development divisions such as Lab-of-Things (Microsoft). Amazon has also jump onto the Digital Home bandwagon with Echo, a coordinating device for home device networking and direct communication between the user and the Digital Home. As such, this means an important bet from the big ICT companies on the development of this technology that transforms the user's most private and personal spaces. Conscious of the business advantages and development possibilities that the Digital Home offers, they have decided to broaden their fields of business to all levels of the Digital Home, from a purely physical level with sensor and actuator development to a functional and logical level, represented by operating systems and interoperability, communication, and implementation platforms.

As observed in the development of investments by the big ICT sector and IS companies, the development of services for user health management and wellness and the transformation of the home into an intelligent platform for distribution of services represents a field of development with a wide margin of benefits, both economic and social (Balta-Ozkan, Boteler, & Amerighi, 2014; Martin, Kelly, Kernohan, McCreight, & Nugent, 2008). One of the most important implications of this shift is the emergence of two new types of services in the IS: sensitive services and pervasive services. In the following sections a clear and concise definition for both is offered.

SENSITIVE SERVICES

The evolution analyzed in the previous section shows a design and development trend towards services with a greater social impact in terms of quality of life. In addition, technologies giving shape to those services with a greater degree of autonomy, ubiquity, delocalization, and control over their users' personal aspects are evolving in parallel.

These new services manage data that is not only based on the physical disposition of the user in regards to the environment (such as geographical position, interior positioning, or patterns of movement) but also related to more personal aspects of their physical state, psyche, or behavior patterns (Ackerman, 2013).

More and more studies show the difficulties related to services that manage data or situations that are considered personal or of special importance to the user. In this sense, a sensitive service will be formally understood in this article as a service that fulfils the following conditions:

- The combination of its functionalities significantly affect its users' quality of life;
- It manages information that is considered private by its users. This information may be provided by users or service providers, learned via automations, or inferred via data mining processes;
- It involves the psychological responses of its users, being able to condition its own behavior or activities;
- They are present in the users' personal spaces, such as the home, school, or workplace, and as such affect the development of day-to-day activities.

Another defining trait of this type of service is where it is to be implemented. They not only affect their users' personal data, but also must be implemented in living spaces where daily life unfolds, such as the home, school, hospitals, or workplaces. Bettini & Rinobi propose, in (Bettini & Riboni, 2015), a classification of personal and sensitive data according to the category of the sensitive services provided. In this sense, for example, a health service deployed in the Digital Home, for its proper functioning, must record and analyze user activities in such an intimate and personal space as is the home. This represents a clear example of a sensitive service that fulfils

the requirements for a sensitive service that were previously expressed.

PERVASIVE SERVICES

The consolidation of the Internet as a means of communication in the IS has encouraged the evolution of computing systems from a centralized concept towards a viewpoint based on a shared information network where all users store and execute services. This situation is commonly known as Cloud Computing. Additionally, advances in the miniaturization of technological components linked to computing has enabled them to be embedded in everyday objects, creating computing networks based on common objects. This phenomenon is known as the Internet of Things (IoT).

These advances have also affected the development of IS services. At one point, these services were offered as a complex element that involved a higher level of control from their users. Currently, these services are offered by operators as simple functionality packets in objects that are already utilized and known by users such mobile phones, electrical appliances, furniture, etc. The orchestration of the capacities offered by these objects with computational capacity and the possibility of remotely processing information has contributed to the appearance of new ICT services called pervasive services. This way, *a pervasive service will be defined as a delocalized and persistent ICT service that allows its users to access the functionalities it offers at any time or place.* The delocalization of these services and their accessibility via the Internet independent of the user's location has facilitated their inclusion in individuals' daily life, thus offering more personalized ICT services. This simpler, delocalized, and persistent IS service concept promotes access by all users, independent of their location, technological culture and knowledge, social condition, or purchasing power.

ISSUES OF THE NEW SERVICES IN THE INFORMATION SOCIETY

Despite the fact that the number of laws and regulations directed at creating privacy and protection for users of this type of sensitive and pervasive services (i.e. the European Data Protection Supervisor (EDPS) or In the Health and Human Services (HHS) in USA) has grown over the last 20 years, the perception of control from the citizen's perspective has deteriorated. According to Eurobarometer 431 (European Commission, 2015), from 2010, researching Information Protection, it can be seen that only 15% of the European population believes that they have total control over data shared when using services in the IS. On the other hand, 31% of the population believes that they do not exercise any control over this information and finally, 50% believes that they exercise some control over shared data while using services in the IS.

A comparative analysis of the results offered by the 2015 Eurobarometer with the one created in 2010 show that the feeling of control over this data has varied. In 2010, the users of these services had a greater perception of control over data than in 2015. While in 2010, only 20% of users believed they did not exercise any control over information, in 2015 it increased by 12%, decreasing the percentage of users that were confident in their control over shared sensitive information. However, these users show a greater awareness, in comparison with the 2010 results, of daily activity monitoring with the use of IS services.

This shows that the sensitivity of the data managed by these services has increased, in large part because these services are more present in users' daily life. Citizens are more and more aware of the existence of risks related to the use of this specific type of services. These risks include fears that impede the acceptance of a technological solution and encourage its rejection in the presence of a viable alternative (Khadke & Chavan, 2014; Weiser, 1995).

In the other hand, the use of pervasive technology for the implementation the new services in the IS and their delocalized nature foster a state of rejection towards the information managed, generating insecurities about who manages such information and how. A study carried out by the CloudT project among European (Spain, France, and Italy) and Japanese entities, whose goal was to boost citizen influence in a future concept of an intelligent city, shows a global viewpoint on citizens' familiarity with the concepts of Cloud Computing and Internet of Things. While the Cloud Computing concept seems to be better-understood by society with 24% positive responses, the IoT turns out to be familiar for only 15% of the surveyed population. That means that 60% of the population is unaware of the nature and objectives of this type of technology, affecting final adoption of the services constructed with it.

FUTURE RESEARCH DIRECTIONS

The main future research arising from this work is the study of acceptance of pervasive and sensitive services. A deep analysis of how the acceptance affects the final adoption of these services is beyond the scope of this paper, however, the authors of this paper believe that this study represents a line of future research of utmost importance. In this sense authors have conducted several researches that provide different analysis approaches for this future line of work, establishing key indicators that can be used to model such acceptance (Vega-Barbas, Pau & Seoane, 2014).

Another line of future work related to this work is the development of systems for the visualization of complex data. Visualization represents an important means of communication between users and systems. No longer just to display complex data in a way that is understandable, but also to expedite the process of interaction with smart spaces and pervasive services. The rise of the deployment of Smart Cities means that the possibilities offered

to users grow uncontrollably. This situation can overwhelm users, preventing proper adoption as valid solutions.

CONCLUSION

The evolution of the IS has been marked by a shift of importance where services have supplanted almost entirely to specific products. As with traditional products, the services are conditioned by the underlying technology used for its development and definition. However, new needs and desires of society have acquired a vital importance in the evolution of these services, giving rise to two new concepts, services and Pervasive sensitive services. Although these concepts are present in industry and academia there is no clear definition. In this sense, this article provides a definition for both concepts with the aim of being used as a means of understanding and adapting to the new challenges that the IS must face. In addition, along of this article the main problems and challenges that these services must meet to be used as a means of developing good solutions have been highlighted. Thus, technology acceptance is presented as the main challenge associated with these new services, which must be adapted to new technological and socio-economic paradigms on which such services are developed.

REFERENCES

- Achrol, R. S., & Kotler, P. (2014). The service-dominant logic for marketing. *The Service-Dominant Logic of Marketing: Dialog, Debate, and Directions*, 320.
- Ackerman, L. (2013). *Mobile Health and Fitness Applications and Information Privacy*. San Diego, CA: Privacy Rights Clearinghouse.
- Al-Shorbaji, N. (2008). e-Health in the Eastern Mediterranean region: A decade of challenges and achievements. *Eastern Mediterranean Health Journal*, 14, S157–S173. PMID:19205616
- Balta-Ozkan, N., Boteler, B., & Amerighi, O. (2014). European smart home market development: Public views on technical and economic aspects across the United Kingdom, Germany and Italy. *Energy Research & Social Science*, 3, 65–77. doi:10.1016/j.erss.2014.07.007
- Bettini, C., & Riboni, D. (2015). Privacy protection in pervasive systems: State of the art and technical challenges. *Pervasive and Mobile Computing*, 17, 159–174. doi:10.1016/j.pmcj.2014.09.010
- Craig, J., & Patterson, V. (2005). Introduction to the practice of telemedicine. *Journal of Telemedicine and Telecare*, 11(1), 3–9. doi:10.1258/1357633053430494 PMID:15829036
- Currell, R., Urquhart, C., Wainwright, P., & Lewis, R. (2010). Telemedicine versus face to face patient care: Effects on professional practice and health care outcomes. *Cochrane Database of Systematic Reviews*, 1–35. doi:10.1002/14651858.CD002098
- European Commission. (2014). *Special Eurobarometer 419: Public Perceptions of Science, Research and innovation. Special Eurobarometer 419* (Vol. 419). Retrieved from http://ec.europa.eu/public_opinion/archives/ebs/ebs_419_en.pdf
- European Commission. (2015). *Special Eurobarometer 431: Data Protection. Special Eurobarometer* (Vol. 431). Retrieved from http://ec.europa.eu/public_opinion/archives/ebs/ebs_431_en.pdf
- GrowthCap. (2014). *The telehealth revolution: How technology companies are bridging quality and efficiency in medicine*. Retrieved from <https://growthcap.co/articles/?p=255>
- Khadke, K. S., & Chavan, U. B. (2014). Survey on distributed accountability for data sharing in the cloud. *International Journal of Computer Science and Mobile Computing-IJCSMC*, 3(1).
- Martin, S., Kelly, G., Kernohan, W. G., McCreight, B., & Nugent, C. (2008). Smart home technologies for health and social care support. *Cochrane Database of Systematic Reviews*, 4. PMID:18843715

Sarell, M., & Machlup, F. (1963). The Production and Distribution of Knowledge in the United States. *American Sociological Review*, 28(5), 840. doi:10.2307/2089942

Vega-Barbas, M., Pau, I., & Seoane, F. (2014, November). Confidence: Dependencies and their critical role in fostering user acceptance in pervasive applications. In *Wireless Mobile Communication and Healthcare (Mobihealth), 2014 EAI 4th International Conference on* (pp. 283-286). IEEE.

Weiser, M. (1995). The Technologist's Responsibilities and Social Change. *Computer-Mediated Communication Magazine*, 2(4), 17.

World Health Organization. (2010). *Telemedicine: opportunities and developments in Member States: report on the second global survey on eHealth*. World Health Organization.

ADDITIONAL READING

Chang, E., Dillon, T., & Calder, D. (2008). Human system interaction with confident computing. The mega trend. In *2008 Conference on Human System Interaction, HSI 2008* (pp. 1-11). doi:10.1109/HSI.2008.4581399

Connelly, K. (2007). On Developing a Technology Acceptance Model for Pervasive Computing. In *9th International Conference on Ubiquitous Computing (UBICOMP)-Workshop of Ubiquitous System Evaluation* (p. 520). USE; Retrieved from <http://www.cs.indiana.edu/~connelly/Use07.pdf>

Hoffman, D. L., & Novak, T. P. (2015). *Emergent Experience and the Connected Consumer in the Smart Home Assemblage and the Internet of Things*. Available at SSRN 2648786.

Vega-Barbas, M., Pau, I., Ferreira, J., Lebis, E., & Seoane, F. (2015). Utilizing Smart Textiles-Enabled Sensorized Toy and Playful Interactions for Assessment of Psychomotor Development on Children. *Journal of Sensors*.

Vega-Barbas, M., Pau, I., Martín-Ruiz, M. L., & Seoane, F. (2015). Adaptive software architecture based on confident hci for the deployment of sensitive services in smart homes. *Sensors (Basel, Switzerland)*, 15(4), 7294-7322. doi:10.3390/s150407294 PMID:25815449

KEY TERMS AND DEFINITIONS

Enabling Technology: Technological innovation or invention which is mainly characterized by the rapid development of subsequent technologies, often present in several scientific and social fields.

Interaction Design: An interdisciplinary vision that is fundamental to all disciplines, fields, or study and development approaches whose objective is to create systems based on computing technology for individuals such as Human-Computer Interaction, Ergonomics, Human Factors, Cognitive Engineering, Psychology or Design.

Pervasive Service: A delocalized and persistent ICT service that allows its users to access the functionalities it offers at any time or place.

Pervasive Technology: Potentially ideal environment, due to its characteristics, to support Ubiquitous Computing which fundamental characteristics are resilience and persistence.

Sensitive Service: An ICT service which as combination of its functionalities significantly affect its users' quality of life, manages information that is considered private by its users, involves the psychological responses of its users and are present in the users' personal spaces.

Technological Acceptance: Degree of adoption and use by the user of goods or services as the practical application of knowledge, whose objective is to satisfy a need or want of society.

Ubiquitous Computing: The third computing paradigm where the person relates physically with a large number of computers without temporal or spatial restrictions.

Home UbiHealth



John Sarivougioukas

“G. Gennimatas” Athens General Hospital, Greece

Aristides Vagelatos

CTI&P, Greece

Konstantinos Parsopoulos

University of Ioannina, Greece

Isaac E. Lagaris

University of Ioannina, Greece

INTRODUCTION

At the third computing era, users interact with many computing devices surrounding or implanted in them, in a natural way. These anytime and anywhere interactions implement the concept of *ubiquitous computing*, which provides the framework for computational awareness and personalization. These properties are precious in healthcare applications since the operating computing devices in the patient's environment can be aware about the evolving situations and actively participate in the medical treatment. In addition, ubiquitously supported healthcare services can be provided anywhere and at any time, allowing specific cases of the hospitalization model to be transferred to the home healthcare model.

The adoption of the home healthcare model in a ubiquitous computing environment provides the prerequisites for the development of the presented *Home UbiHealth* model. Extending medical services at home provides the capability to cover the medical needs of all population categories. Specifically, the Home UbiHealth model refers to all major population groups, namely the healthy population supported with prolepsis policies that retain health status; individuals that suffer a health crisis that requires recovery; and chronic patients who must maintain their quality of life coping with known health problems. Within the above

patient categories are included the special groups of infants, children, disabled, and pregnant, which have special healthcare needs.

Thus, the home environment is transformed to a reference starting point for the implementation of healthcare processes independently of location and time, thereby bringing together the various healthcare stakeholders and the market. In this framework, the Home UbiHealth model can change the perception about the structure of healthcare systems and the concerns about medically uncontrollable environments.

BACKGROUND

In recent years, medical research has offered tremendous developments. Within this framework, specialized personnel is required to carry out advanced processes within properly structured and controllable facilities, using state-of-the-art biomedical equipment. Unfortunately, the availability of medical resources hardly meets the current social demands for hospitalization in cities and rural areas. This is partially due to the long average hospitalization periods required to perform trivial medical and nursing procedures such as screenings, lab-tests, or follow-ups. The lack of adequate infrastructures leads to longer stay of the patient in hospital.

DOI: 10.4018/978-1-5225-2255-3.ch675

These deficiencies can be addressed through new healthcare models that are supported by modern computing technologies such as ubiquitous computing. Ubiquitous computing was introduced by Mark Weiser to describe the *third wave* or *calm computing* (Weiser, Gold & Brown, 1999), where computers are enweaved into every fabric supporting the end user. In this era, the users are supposed to subconsciously interact with many computers, concurrently, in such a natural way as one uses eye-glasses to restore vision problems.

The application of ubiquitous computing in healthcare systems introduced the term *ubiquitous health* (*UbiHealth*) to describe the use of inherited computing characteristics in healthcare models (Sarivougioukas & Vagelatos, 2015). UbiHealth refers to healthcare services that incorporate ubiquitous computing means. Such services can provide critical advantages to overcome limitations related to individualized care, medical personalized treatment, patient safety, economy of scale, as well as healthcare system efficiency, effectiveness, security, and scalability.

Controllable hospitalization is directly related to quality of treatment, continuity of care services, transparency of medical and nursing supportive activities, patient safety, and administration of the involved supply-chains and related costs. In principle, UbiHealth satisfies the requirements related to the quality of the provided medical and nursing services, the demands for continuity of the involved processes, the necessary conditions of transparency in the followed procedures, and the fundamental prerequisite of safety for medical professionals and patients. Hence, UbiHealth can highly contribute to overcome issues related to hospitalization by providing the ground for medical and nursing processes within ubiquitously performing environments, such as at home.

Treating and curing patients at home (see Figure 1) has been considered as beneficial for particular population groups and has been extensively tested (Madaris et al., 2016) to the point where the treatment is under complete control by the assigned medical personnel. Usually, this

refers to post-acute care or chronic patient cases. A typical example is the *Medical Home* paradigm (American Academy of Pediatrics, 1992), which was introduced by pediatricians to control infant mortality and children vaccination.

The significance of home treatment has been verified with respect to prolepsis, prognosis, testing, cure, and treatment. However, the continuously increasing costs of advanced cures, which are habitually applied in highly specialized facilities by specialized medical personnel, has turned the interest of the scientific community to home-based alternatives in order to address the lack of adequate treatment resources. Moreover, the associated risks during hospitalization, such as nosocomial internal infections, accidents, medical mistakes, and human errors, motivate researchers to delve into solutions alternative to hospitalization.

In addition, the bureaucracy imposed by long hospitalization increases the complexity of the carried processes, resulting in significant time and cost growth. In turn, this imposes substantial overhead to social security services, insurance funds, and the related market. Furthermore, limited available hospital resources are incapable to offer individualization of the provided services due to the applying workload. This is in contrast to the endeavors of medical practitioners and healthcare systems, which aim at personalization in the of-

Figure 1. Home Care main components and their relationships



ferred cures according to the patient needs and preferences (Snowdon, Scnarr & Alessi, 2014). Home UbiHealth offers a promising solution to this problem, given an adequate ubiquitous health ecosystem.

HOME UBIHEALTH ECOSYSTEM

Home UbiHealth systems operate in an environment where evolving compound processes follow complex workflows that involve stakeholders with diverging concerns and interests. The Home UbiHealth environment includes an underlying computing infrastructure that supports the patient's personal healthcare needs along with interaction with all participating stakeholders. Hence, Home UbiHealth can be considered as the common loci where computing and healthcare support personalized healthcare needs at home, eliminating time and space limitations.

The available technological developments provide the opportunity to resolve several issues in Home UbiHealth, which are related to three major problem areas. The first area is related to the lack of standardization in the home computing hardware infrastructure. This is tackled with the introduction of the *Internet of Things* (IoT), which allows overcoming problems such as the ones in local area networks applications (Riggins & Wamba, 2015). The second area is related to software infrastructures. This is addressed through *cloud computing*, which provides a variety of schemes such as software as a service (SaaS), platform as a service (PaaS), and infrastructure as a service (IaaS) (Ullrich, Vasileiadoy & Tamm, 2011). The third problem area is related to configurability that is necessary to address medical and social issues. This is handled by properties of ubiquitous computing such as the *computing awareness*. Thus, despite the raised problems, there is a variety of technological tools available to support the implementation of the Home UbiHealth paradigm.

The Home UbiHealth model aims at conditionally substituting parts of hospitalization (Leff, 2015), provided that prescribed prerequisites are fulfilled. Although the physical presence of medical professionals is irreplaceable, there are specific procedures that can be carried out in reformed fashion, assisted by computers and communication equipment. This can reduce dependability from specialized professionals. The Home UbiHealth model aims at sharing scarce medical resources by employing reliable and effective communication schemes. However, it may suffer from intolerable commuting overheads (Leff, 2009) since the supporting professionals are obliged to pay scheduled visits at patients' place.

The presence of medical professionals at home to consult family members and caregivers has a twofold purpose. First, the doctor at home audits and instructs about the appropriateness of treatment activities. Second, he coordinates all treatment activities regarding the in-house and out-of-house services. Therefore, the Home UbiHealth model relies on the cooperation and participation of the medical professionals, coordinated by an assigned medical doctor who determines the associated supporting services.

Medical treatment at home requires the availability of medical devices, consumables, and drugs. Also, the medical treatment involves approvals and reimbursements from social security organizations or insurance funds (see Figure 1). These procedures frequently produce delays, frustration, and significant overhead. However, the deficiencies can be eliminated through ubiquitous computing procedures. Additionally, by carrying ubiquitously performed processes, the necessary products, goods, drugs, and services can be organized on behalf of the patient, since all participating stakeholders are aware about the patient and the coverable treatment needs. Hence, multiple procedures can be concurrently activated involving all relevant stakeholders in order to meet the needs of the patient at home.

For instance, a doctor's prescription can receive the necessary approvals from the sup-

porting insurance fund, be executed by the local pharmacy store, and delivered at the patient's home. Therefore, the Home UbiHealth model can become a reference point for ubiquitously performed healthcare processes on behalf of the patient with minimal bureaucratic overheads and complete approval control.

DESIGN PRINCIPAL PRIMITIVES

The implementation of the Home UbiHealth model requires adequate computing infrastructure, capability to achieve continuity of care, and ability to adjust the functional performance of the designed system in order to meet personal needs and preferences.

The context in the Home UbiHealth model is produced from signals of dispersed sensors and computer-controlled devices in the patient's environment. These signals form the holding *situational context*. The devices form ad-hoc cooperative communication networks that support the patient's activities. They fall into three main categories with respect to their location or mobility, namely they can be stationary, mobile or carrying, and implanted (Latre *et al.*, 2011). Major issues of the devices are related to their participation in cooperative networks and their power supply. The formed networks must follow protocols that guarantee patient's safety, security, and privacy, along with support of the patient's medical treatment and personal preferences.

Thus, the computing infrastructure is distinguished into *in-house* and *out-of-house infrastructures*. These are characterized by scalability, which is offered by the ubiquitous computing abstraction of services (Loeser *et al.*, 2003). Both infrastructures are supported by IoT and cloud reproductions, ensuring continuous and uninterrupted feeding of data over the internet. The infrastructure facilitates pre-processing, processing, cognitive, and controlling phases. The pre-processing phase involves the operations of data cleansing, normalization, fuzzification, fusion, and correlation, which ex-

clude inaccuracies, ambiguities, and uncertainties from the obtained raw data. The processed raw data is adequate for the development of the situationally holding context, after the examination of existing relations among the obtained data through rule-based systems (Sarivougioukas & Vagelatos, 2014).

Then, the processing phase applies medical rules on the obtained context in order to produce useful information. The cognitive phase correlates information supported by medical knowledge representations, such as specialized medical taxonomies, in order to draw conclusions and develop a conceptual or knowledge model. Eventually, the controlling phase operates on networks of concepts, assisted by a decision-support system that makes decisions about the proper guidance of physical devices in the patient's environment. Therefore, the computing infrastructure analyzes and processes raw data from the surrounding sensors and returns particular data to each actuating device, thereby supporting the patient without requiring direct coupling between sensors and actuating devices.

Medical doctors can determine the objectives of the cure by prescribing, ordering, and instructing the individual patient's treatment activities through dedicated user interfaces. Thus, the treatment plan is scheduled along with the necessary procedures, within and outside of home. The involved nurses and caregivers receive specific instructions about the personalized treatment, allowing doctors to be regularly informed about the treatment progress, including alarms and warnings.

The model's software infrastructure is the reference or initiation point for all processes regarding the patient's support. Besides the direct support provided by the dispersed devices within the home environment, the patient's support requires the initiation of processes as instructed either by the doctor in charge or the patient's health conditions. For instance, the order of drugs for a chronic disease such as diabetes can be activated by one of the watching medical rules on behalf of the patient, requesting doctor's approval and

placing the order to a predetermined pharmacy store according to a predetermined payment type. The performing processes are circular, starting from the UbiHealth software infrastructure and ending, upon completion, to the same system. Thus, the supporting software infrastructure can involve all necessary stakeholders for the patient's support, following workflows that bring together medical authorities, government agencies, as well as services and products from the market. The entire operation of the Home UbiHealth model depends on the availability of the internet and the capability of the software infrastructure to operate from the residing virtual cyberspace.

The autonomy of the Home UbiHealth model depends on instructions and orders provided by the medical doctor in charge. In other words, it depends on the availability of information upon which, decisions can be made. The model can be used equally well for all three distinct cases of individuals, namely the healthy ones, individuals undergoing a health crisis, and the ones with a chronic disease. The Home UbiHealth paradigm allows doctors to perform remote audits besides the necessary in-person clinical examinations. Thus, the entire health history (healthcare records) of the individual is available for consultation by the assigned medical doctor with updated, reliable, and accurate data. Also, the use of the internet as the underlying communication medium that connects the patient with the software infrastructure that resides at the cloud, provides a uniformly and continuously available reference to the personal health data. Hence, the continuity of care is achieved with respect to documentation, time, and space.

On the other hand, continuity of care in the Home UbiHealth model depends on the participation of the involved stakeholders during health-related events. The development of systems and adequate software applications motivates stakeholders to contribute to the patient's treatment. Also, hospitals can share data and continue providing support to the patient who is transferred for recovery at home. The involved

social security organizations can support the home treatment of the patient by satisfying the financial issues on-time. Similarly, suppliers of medical products, consumables, and services can be proactively prepared, avoiding bureaucratic procedures regarding the administration, financing, and supply-chain support.

Therefore, the Home UbiHealth model possesses all the necessary properties that allow the patient's support at home in a holistic manner. It can be considered as the structural cell for building healthcare systems. Its parameters can be adjusted to match the desired conditions of an efficient healthcare system with respect to public health, social needs, environmental requirements, and financial constraints. The sustainable healthcare models meet operational requirements that satisfy individual medical needs at desirable level of quality. The operational requirements are satisfied by the prompt availability of data and information regarding the patient's personalized needs, while medical needs are satisfied through the provision of services of predefined quality. Given the desired policies, aimed purposes, operational objectives, and quality level of the offered products and services, the Home UbiHealth model can be optimized with respect to its parameters. This way, it promotes the development of optimal and sustainable designs.

CHALLENGING ISSUES

Contemporary research and development efforts on the extension of the Home UbiHealth model are focused on a number of challenging issues. First, the Home UbiHealth model relies on raw data that is transformed into medical knowledge to serve the medical needs of the supported patient. Second, the provision of services must be adjusted to support personalized medical needs of the patient. Third, the autonomy of the model is determined by the availability of information upon which, decisions are made. This underlines the need for adequately designed decision-making

systems. Fourth, the Home UbiHealth model requires efficient methodologies for its formal description. Finally, the design of the model must properly address all security issues. Significant research efforts are paid on the foundations of the model's building blocks that will eventually lead to *in vivo* applications.

The medical support of patients at home requires the adoption of medical knowledge, which must be adequately represented in order to be processed and properly applied. Knowledge is obtained from the available information as a result of processing raw data provided by dispersed devices and smart objects in the ubiquitous computing home environment. The applied data transformations must be accompanied by well-structured and formally defined ontologies. These make available the current level of medical knowledge in various forms, such as rule-based systems, case-based systems, and adapted medical protocols.

The performed data transformations are affected by the patient's characteristics and personalized needs. The ontological support provides the ability to perform inferences and eventually extract the involved discrete medical concepts from the obtained information. Ongoing research efforts try to achieve such representations of the medical knowledge that can be composed from raw data in order to make decisions and, then, be decomposed into low-level data to control the dispersed actuating devices in the ubiquitous computing home environment (Sarivougioukas & Vagelatos, 2015).

Moreover, the Home UbiHealth model has additional inherent restrictions, presented below in decreasing order of significance (Teijeiro *et al.*, 2013). The first and most significant source of constraints refers to the applied medical rules that govern the decision-making components of the model. The second source of constraints refers to the introduced instructions and orders placed by the treating medical doctor, taking into consideration the particular physical and medical characteristics of the patient. The third source

of restrictions is about the personal preferences declared by the patient. Current research efforts are focused on the accomplishment of formally defined uniform representations of constraints that nest the aforementioned sources of restrictions and eliminate conflicts and contradictions (Sarivougioukas, Vagelatos & Lagaris, 2015).

Another significant issue is the autonomy of the Home UbiHealth model, which depends on the availability of decision-making components (mechanisms). Decisions are the outcomes of processing operations that take into consideration the available options. The procedural processing in decision making proposed by Simon (Simon, 1979) approximates the practical limits. However, Simon's model requires restructuring since the identification of the problem upon which, a decision is pending, must be provided by the supporting ontology. Also, the process that provides the appropriate medical decision must be proactively available by the model, in order to minimize the necessary search time.

Thus, the model must be dedicated to synthesize actions and bridge the gap between the medically suggested solution and holding conditions. Also, it must be properly designed to provide either decision-making support to the medical professionals or decision making to follow the issued orders of the prescribing doctors. In other words, the decision-making component must follow and participate in the medically adapted workflow of operations carried by the patient and the assigned doctor. Therefore, the decision-making component must be designed to take advantage of the cognition effectiveness, efficiency of intelligence, and optimization.

The complexity of the presented Home UbiHealth model exceeds the efficiency of analytical mathematics to formally describe its components, their functionality, and their interactions. Set Theory and Algebra can hardly offer formal descriptions of the involved interactions and model's behavior. Thus, alternative mathematical methods must be employed to represent and manipulate

all the involved quantities in order to allow the handling of the static and the dynamic properties of the model.

To this end, *Denotational Mathematics* (Wang, 2012) can provide the necessary efficiency and flexibility for the formal representation of the Home UbiHealth model. The formally presented mathematical premises must be capable of approaching self-referenced systems, such as the ones formed in the Home UbiHealth model. Also, they shall include the patient as part within (i.e., observed) and outside (i.e., observer) of the formed system. The model must provide the means to administer formal knowledge representations, manipulate algebraically controlled operations, formally describe behavioral abstractions, and handle dynamically changing content.

Finally, the Home UbiHealth model must tackle security issues at design-time, satisfying the technical and medical requirements. Thus, security is approached as a framework consisted of complete processes, participating and rectifying the system's behavior. The security framework is distinguished into infrastructural and application parts that correspond to the designer's and the user's degrees of freedom, respectively. Infrastructural security includes processes that ensure proper accommodation and facilitation of the carried processes of the Home UbiHealth model. The application security framework consists of the interrelated procedures that act as features to the carried end-user's processes and reflect to the medical constraints. Therefore, security must be considered during both the design and operation of the Home UbiHealth model, in order to ensure the proper management and administration of human lives.

FUTURE RESEARCH DIRECTIONS

The ongoing technological progress offers the opportunity to include home healthcare services among the formal components of healthcare systems. In particular, Home UbiHealth can

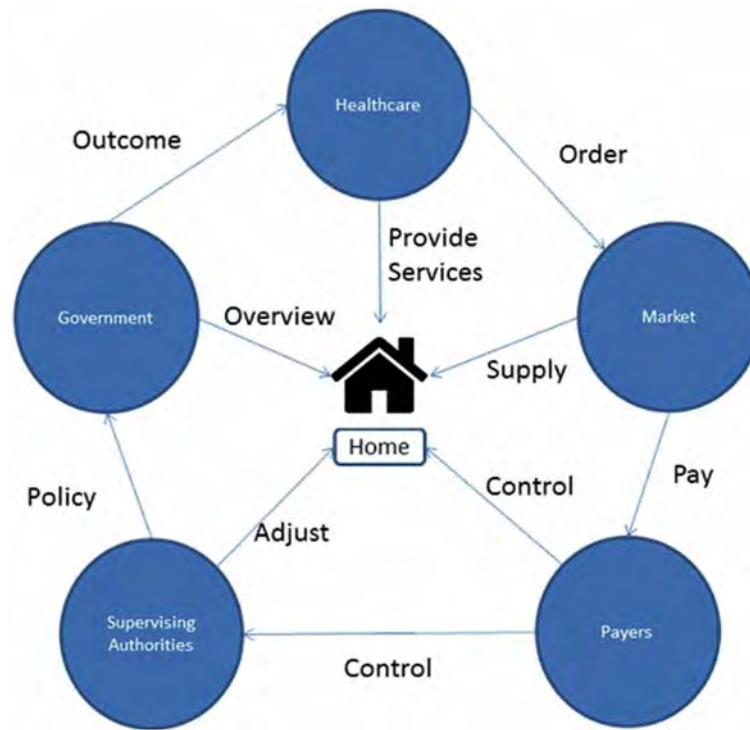
become an intermediate layer between primary and secondary healthcare systems (see Figure 2). As a consequence, novel medical professional specializations may appear in order to meet the emerging needs of healthcare services at home. For the realization of a formal home healthcare model, research efforts are required in medicine, nursing, computer science and engineering, oriented towards the Home UbiHealth paradigm. Nevertheless, computer simulation and optimization can accelerate scientific developments in home healthcare and bring it closer to real-world implementations. Thus, they appear to be fertile ground for future research.

Regarding technological aspects, the need for direct and indirect physiological signals, data, and information processing methods, reveals the necessity for efficient parallelization and data fusion applied upon configurable software applications capable of promoting personalization. Additionally, further research is needed on ad-hoc wireless networks, which suffer frequent interrupts due to the lack of continuous power supply, along with the unavailability of standardized communication protocols that meet the Home UbiHealth model requirements. Also, standardization of context-awareness is required in order to optimize the structure of future decision-support systems. Moreover, additional research is needed on the optimization of the underlying communication systems in order to be capable of supporting the increasing number of the dispersed sensors and actuators, as well as the carried mobile and implanted devices. This can reduce the time response of the entire system. Finally, cognitive research approaches must be considered to decrease the computational overheads aiming at the avoidance of analytical calculations to figure out the most adequate responses.

CONCLUSION

The Home UbiHealth model relies on the ubiquitous computing principles and characteristics

Figure 2. Home UbiHealth's place in the pyramid of healthcare



of context-awareness. The model processes contextual raw data that can be transformed into information which, in turn, produces knowledge supported by adequately designed medical ontologies. Knowledge representations and knowledge management provide the means to develop a formal model for treating patients of all categories at home. The model introduces the capability to control medical conditions at home, providing alternatives to the existing healthcare systems and hospitalization models. Thus, the Home UbiHealth model can have financial, social, and environmental impact through the personalization of the provided treatments and the transparent participation of healthcare stakeholders, including the related segments of the market.

The compound design of the Home UbiHealth model along with the complex workflows of its processes exceeds the capabilities of analytical

mathematics. Alternative mathematical approaches are required to ensure the necessary formality in the involved descriptions of the model, providing efficient and effective expressiveness of the occurring healthcare situations. Recent research has offered evidence that Denotational Mathematics provide the capability to efficiently represent and describe the model's static and dynamic behavior. Moreover, Denotational Mathematics provide a conceptual system framework to administer the involved entities with algebraic tools. This increases the efficiency of the model and facilitates the use of established Systems Theory and Control Theory principles and methods. Thus, the Home UbiHealth model simultaneously considers the patient acting as observer of the system or independently performing as controller of its underlying control system.

REFERENCES

- American Academy of Pediatrics. (1992). Ad Hoc Task Force on Definition of the Medical Home: The medical home. *Pediatrics*, 90(5), 774. PMID:1408554
- Latré, B., Braem, B., Moerman, I., Blondia, C., & Demeester, P. (2011). A survey on wireless body area networks. *Wireless Networks*, 17(1), 1–18. doi:10.1007/s11276-010-0252-4
- Leff, B. (2009). Defining and disseminating the hospital-at-home model. *Canadian Medical Association Journal*, 180(2), 156–157. doi:10.1503/cmaj.081891 PMID:19153385
- Leff, B. (2015). Hospital at home. In M. Malone, E. Capezuti, & R. Palmer (Eds.), *Geriatrics Models of Care* (pp. 163–171). Springer International Publishing. doi:10.1007/978-3-319-16068-9_14
- Loeser, C., Mueller, W., Berger, F., & Eikerling, H. (2003). Peer-to-Peer Networks for Virtual Home Environments. In *Proceedings of 36th Annual Hawaii International Conference on System Science*. IEEE.
- Madaris, L., Onyebueke, M., Liebman, J., & Martin, A. (2016). SCI Hospital in Home Program: Bringing Hospital Care Home for Veterans With Spinal Cord Injury. *Nursing Administration Quarterly*, 40(2), 109–114. doi:10.1097/NAQ.000000000000150 PMID:26938182
- Riggins, F., & Wamba, S. F. (2015). Research Directions on the Adoption, Usage, and Impact of the Internet of Things through the Use of Big Data Analytics. In *Proceedings of 48th Hawaii International Conference*. IEEE. doi:10.1109/HICSS.2015.186
- Sarivougioukas, J., & Vagelatos, A. (2014). Analyzing Medical Contexts in Ubiquitous Computing Home Environments with Denotational Mathematics. *Proceedings of 13th IEEE International Conference on Cognitive Informatics & Cognitive Computing*. doi:10.1109/ICCI-CC.2014.6921449
- Sarivougioukas, J., & Vagelatos, A. (2015). Use of Denotational Mathematics for the Formal Description of Autonomous Migration and Polymorphism as Prerequisites for Mobility in Home UbiHealth. *Proceedings of 14th IEEE International Conference on Cognitive Informatics & Cognitive Computing*. doi:10.1109/ICCI-CC.2015.7259412
- Sarivougioukas, J., Vagelatos, A., & Lagaris, I. (2015). Administration of Medical Contexts with Denotational Mathematics in Ubiquitous Computing Home Environments. *International Journal of Software Science and Computational Intelligence*, 7(2), 1–30. doi:10.4018/IJSSCI.2015040101
- Simon, H. (1979). Rational decision making in business organizations. *The American Economic Review*, 69(4), 493–513.
- Snowdon, A., Schnarr, K., & Alessi, C. (2014). *It's all about me, the personalization of health system*. London: Western University Canada.
- Teijeiro, T., Félix, P., Presedo, J., & Zamarrón, C. (2013). An open platform for the protocolization of home medical supervision. *Expert Systems with Applications*, 40(7), 2607–2614. doi:10.1016/j.eswa.2012.11.001
- Ullrich, S., Vasileiadou, E. & Tamm, G. (2011). *Cloud Computing Definitions and Approaches*. Whitepaper: SRH Hochschule Berlin.
- Wang, Y. (2012). In search of denotational mathematics: Novel mathematical means for contemporary intelligence, brain, and knowledge sciences. *Journal of Advanced Mathematics and Applications*, 1(1), 4–26. doi:10.1166/jama.2012.1002
- Weiser, M., Gold, R., & Brown, J. S. (1999). The origins of ubiquitous computing research at PARC in the late 1980s. *IBM Systems Journal*, 38(4), 693–696. doi:10.1147/sj.384.0693

KEY TERMS AND DEFINITIONS

Calm Computing: Describes the calm (quiet) participation of computing in the interaction with the user without claiming exclusive attention.

Denotational Mathematics: Refers to advanced mathematical entities utilizing objects and conceptual abstractions, defining and manipulating complex relations, representing and administering knowledge, and expressing and handling formally behavioral aspects of executing processes without needing analytical mathematics. The involved mathematical processing is performed on algebraic abstractions denoting mathematical functionalities that obey axioms and laws in order to support the denotational and expressive needs in cognitive informatics, computational intelligence, software engineering, and knowledge engineering.

Internet of Things (IoT): Used to depict the connection of physical objects to the internet with the capability to introduce their identity and communicate with other devices exchanging data. Using the producer-consumer scheme over the internet, the IoT paradigm allows machine produced data to be exploited by other machines.

Knowledge Representation: Required for expressing intelligence and it concerns views of explicit instantiations of conceptual abstractions

to facilitate knowledge sharing and exchange. The conceptual abstractions are formed by processed and interacting entities such as concepts, ideas, processes, and rules.

Pervasive Computing: Refers to the embodiment of computers into physical objects connected into networks, that are consistently available, exchanging information, existing everywhere and operating whenever and wherever needed. The term is used interchangeably with Ubiquitous Computing or UbiComp.

Personalized Healthcare Treatment: Refers to any biologic information that can assist in feasibly predicting on-time the occurrence of a disease or prognosticating a patient's response to a given treatment. It is used synonymously with Personalized Medicine which implies additional scientific aspects relying on the molecular basis of diseases with roots in genetics and genomics.

UbiHealth: Refers to the support of healthcare services taking place in a ubiquitous computing environment.

Ubiquitous Computing (UbiComp): The unobtrusively participation of computing at anytime and anywhere allowing users to interact naturally the way someone is using eyeglasses, without any additional effort.

Multifaceted Applications of the Internet of Things



Kijpokin Kasemsap

Suan Sunandha Rajabhat University, Thailand

INTRODUCTION

The Internet of Things (IoT) is a new world for connecting every object of the real world with the virtual space in the computer world (Erguler, 2015). The vision of the IoT predicts a future Internet incorporating smart physical objects that offer the hosted functionality as the IoT services (Dar, Taherkordi, Baraki, Eliassen, & Geihs, 2015). IoT refers to the uniquely identified computing devices that are connected to the Internet and embedded into things (Salim & Haque, 2015). The IoT paradigm envisions the pervasive interconnection and cooperation of smart things over the current and future Internet infrastructure (Ziegeldorf, Morchon, & Wehrle, 2014).

The IoT is enabled through several technologies ranging from communication systems to distributed intelligence (Moschakis & Karatza, 2015). The IoT is a new stage of intelligentization and informatization development (Ning & Hu, 2012). The IoT is envisioned as a natural evolution of the Internet, promising to enable the ubiquitous connections for pervasive objects (Ren, 2011). The IoT envisions some services that require both end-device mobility and high availability or high bandwidth (Sonntag & Suomi, 2015). The IoT envisions to connect many sensors to the Internet (Perera, Zaslavsky, Christen, & Georgakopoulos, 2014).

The IoT includes various kinds of devices (e.g., sensors, actuators, radio frequency identification (RFID) tags, smartphones, and backend servers), which are different in terms of size, capability, and functionality (Nguyen, Laurent, & Oualha, 2015). RFID solutions can be utilized

to reduce the operating costs through decreasing labor costs, enhancing automation, improving tracking and tracing, and preventing the loss of materials (Kasemsap, 2015a). Smart objects are generally added to the Internet using IPv6 over Low-power Wireless Personal Area Networks (6LoWPAN), which defines IP communication for resource-constrained networks (Raza, Duquennoy, Höglund, Roedig, & Voigt, 2014).

This article aims to bridge the gap in the literature on the thorough literature consolidation of the IoT. The extensive literature of the IoT provides a contribution to practitioners and researchers by describing the multifaceted applications of the IoT in order to maximize the business impact of the IoT perspectives.

BACKGROUND

The IoT is a paradigm where everyday objects can be equipped with identifying, sensing, networking, and processing the capabilities that will allow them to communicate with one another and with other devices and services over the Internet to accomplish the strategic goals (Whitmore, Agarwal, & Xu, 2015). Mashal et al. (2015) indicated that different types of smart devices are interconnected and communicate via Internet Protocol that creates a worldwide ubiquitous and pervasive network called the IoT.

With the emergence of Internet Protocol-related IoT devices (Shelby & Bormann, 2009) and the concept of embedded Web services (Shelby, 2010), enterprise level applications (e.g., business processes) are extended (Caracas, 2012) to opti-

mize their execution by collecting the real-time information provided by the IoT devices. Business process modeling (BPM) methodologies provide business users with the ability to model their business processes and to implement business models (Kasemsap, 2016a). Logistics efficiency (Guinard, Trifa, Karnouskos, Spiess, & Savio, 2010), safety processes for storing hazardous materials (Mahlknecht & Madani, 2007), and the remote patient monitoring (Yao, Chu, & Li, 2012) are the major IoT applications.

The IoT applications have been adopted in many different domains (Mashal et al., 2015). Several researches focus on the field of health care systems (Pang et al., 2015), rehabilitation systems (Fan, Yin, Xu, Zeng, & Wu, 2014), and systems assisting peoples with disabilities (Domingo, 2012). In order to improve the quality of human life, IoT technology is broadly utilized to create the smart environments, such as smart homes (Gao, Ling, & Yuan, 2011), smart buildings (Tao, Yajuan, Deyun, Junqi, & Hongke, 2010), and smart cities (Theodoridis, Mylonas, & Chatzigiannakis, 2013) which include intelligent transportation and logistics (Chen, Guo, & Hu, 2010), bus systems (Eberle, 2007), and smart parking management (Bechini, Marcelloni, & Segatori, 2014). The IoT paradigm promises to increase the visibility and awareness of energy consumption (Shrouf & Miragliotta, 2015).

CHALLENGES AND IMPLICATIONS OF THE INTERNET OF THINGS

This article highlights the overview of the IoT and the multifaceted applications of the IoT.

Overview of the Internet of Things

The Internet of Things (IoT) has attracted tremendous attention all over the world and is considered as the third wave of information industry (Mashal et al., 2015). In recent year, the IoT has drawn

significant research attention (Li, Xu, & Zhao, 2015). The effective strategy of the IoT can help firms reach the emerging opportunities from the IoT and improve their competitive advantage (Li, Hou, Liu, & Liu, 2012). The IoT refers to the information exchange and communication between anything and the Internet to realize the intelligent recognition, positioning, tracking, monitoring, and management (Wang, Xie, Wang, & Jia, 2011). The IoT, which is established over architectures of wireless sensor networks (WSNs), provides an actual platform for various applications of personal and ubiquitous computing (Guo, Zhang, Sun, & Bie, 2014).

Structure of a classical IoT is divided into three layers (Jia, Feng, Fan, & Lei, 2012). The first layer is the perception layer that recognizes objects and gather their information (Erguler, 2015). The second layer is the network layer which is responsible for the transmission of the collected data from the perception layer to the corresponding application systems through Internet and mobile telecommunication network. The last layer is called the application layer, where applications are executed to process information and serve to users. As IoT generates massive amount of useful data, data mining techniques play an important role in making IoT applications smart enough to provide the adequate services and enable improved environments (Mashal et al., 2015).

There have been several conducted studies and surveys that are relevant to the security in the IoT (Miorandi, Sicari, & Pellegrini, 2012). For instance, Wang et al. (2006) gave the detailed survey of security issues in WSNs, which can be considered as a reference for the IoT. Atzori et al. (2010) focused on authentication, data integrity, and privacy issues in the IoT, particularly in RFID systems and sensor networks. Kumar and Patel (2014) provided the general overview of security and privacy issues in the IoT. Kumar and Patel (2014) explained the description of different security threats and privacy concerns while processing, storing, and transmitting the data.

Multifaceted Applications of the Internet of Things

The IoT represents the future technology trend of sensing, computing, and communication (Guo et al., 2013). The IoT and its relevant technologies have been attracting the attention of researchers from academia, industry, and government in recent years (Tsai, Lai, & Vasilakos, 2014). The IoT is recognized as one of the most important areas of future technology and is gaining vast attention from a wide range of industries (Lee & Lee, 2015). The IoT extends the capabilities to identify products with the new technologies, such as near field communication, RFID, quick response code and with the existing identification technologies, such as barcodes (Jara, Parra, & Skarmeta, 2014).

Lee and Lee (2015) indicated that five IoT technologies used for the deployment of successful IoT-based products and services include RFID, WSNs, middleware, cloud computing, and the IoT application software. RFID technology constitutes an important part of what has become known as the IoT toward establishing the dynamic and complex environment (Rekleitis, Rizomiliotis, & Gritzalis, 2014). RFID allows automatic identification and data capture using radio waves, a tag, and a reader (Lee & Lee, 2015). The tag can store more data than traditional barcodes. The tag contains data in the form of the Electronic Product Code (EPC), a global RFID-based item identification system developed by the Auto-ID Center. Adaptation of RFID and wireless sensor networking technologies can make the revolution of the IoT (Gubbi, Buyya, Marusic, & Palaniswami, 2013). For RFID-based IoT, objects are identified by the RFID readers that are connected to the servers through Internet for processing data intelligently. Mobile handsets (e.g., netbooks, cellular phones, and tablet computers) act as RFID readers that efficiently scan and gather the information about RFID-tagged items (Erguler, 2015).

WSNs include the spatially distributed autonomous sensor-equipped devices to monitor the physical or environmental conditions and can

cooperate with the RFID systems to better track the status of things, such as their location, temperature, and movements (Atzori et al., 2010). WSNs consist of a large number of small, inexpensive, and resource-limited devices called sensors, which are used to collect the physical world data that is wirelessly transferred to a central location known as a Sink or Base Station (Mashal et al., 2015). These sensor devices communicate with each other using different proprietary protocols, such as ZigBee (Usman & Shami, 2013). WSNs have primarily been utilized in the cold chain logistics that employ the thermal and refrigerated packaging methods to transport the temperature-sensitive products (Hsueh & Chang, 2010).

Middleware is a software layer interposed between software applications to make it easier for software developers to perform the communication and input/output (Lee & Lee, 2015). Its feature of hiding the details of different technologies is fundamental to free IoT developers from software services that are not directly relevant to the specific IoT application. Many IoT applications require massive data storage, huge processing speed to enable the real-time decision making, and high-speed broadband networks to stream data, audio, and video. Regarding IoT application software, the IoT facilitates the development of industry-oriented and user-specific IoT applications (Lee & Lee, 2015).

Whereas devices and networks provide the physical connectivity, IoT applications enable device-to-device and human-to-device interactions in a reliable manner. The IoT applications on devices need to ensure that messages have been received in a timely manner. For example, transportation and logistics applications monitor the status of transported goods, such as fruits, fresh-cut produce, meat, and dairy products. FedEx uses SenseAware to keep tabs on the temperature, location, and other vital signs of a package. While device-to-device applications do not necessarily require data visualization, more and more human-centered IoT applications provide visualization to present information to end users in an intuitive way

and to allow the interaction with the environment (Lee & Lee, 2015).

The recent emerge of the IoT changes the teaching and learning process in global education (Uzelac, Gligoric, & Krco, 2015). Smart classrooms can be defined as intelligent environments equipped with an assembly of many different kinds of hardware and software modules, such as projectors, cameras, sensors, and face recognition module (Xie, Shi, Xu, & Xie, 2001). Another smart classroom environment that is based on the IoT technology presents a system that can detect the level of students' interest in near real-time with 80 percent accuracy (Gligoric, Uzelac, Krco, Kovacevic, & Nikodijevic, 2015).

FUTURE RESEARCH DIRECTIONS

The classification of the extensive literature in the domains of the IoT will provide the potential opportunities for future research. Human-computer interaction (HCI) is the study of how people design, implement, and utilize the interactive computer systems, and how computers affect individuals, organizations, and society through human-computer interfaces. Cloud computing includes network access to storage, processing power, development platforms, and software (Kasemsap, 2015b). Cloud computing is a model for on-demand access to a shared pool of configurable resources (e.g., compute, networks, servers, storage, applications, services, and software) that can be easily provisioned as Infrastructure as a Service (IaaS), Platform as a Service (Paas), and Software and applications (SaaS). SaaS represent the largest cloud market and are still growing quickly. SaaS utilizes the Web to deliver applications that are managed by a third-party vendor and whose interface is accessed on the clients' side. HCI, cloud computing, SaaS, and the IoT are certain to remain important topics for future research.

The adoption of Big Data becomes the foundation of competition and growth for individual

firms, enhancing productivity and creating significant value for modern business by reducing waste and increasing the quality of products and services. Web 2.0 describes a set of next-generation Internet technologies. These protocols and tools make it easier to create online applications that behave dynamically, much like traditional PC-based software. The Semantic Web (also known as Web 3.0) represents the next major evolution in connecting information, thus enabling data to be linked from a source to any other source recognized by computers. Leaders of virtual support teams may need to assume a coordinating role to ensure effective collaboration and communication among virtual team members through information technology (Kasemsap, 2016b). The relationships among the IoT, Big Data, Web 2.0, the Semantic Web, and virtual team will be the influential issues for future research directions.

CONCLUSION

The overview of the IoT and the multifaceted applications of the IoT are emphasized in this article. The IoT is made up of hardware and software technologies regarding WSNs, robotics, RFID, and cloud computing. The IoT describes the modern world in which everyday objects are connected to a network so that data can be effectively shared. The IoT refers to the networking of physical objects through the application of embedded sensors, actuators, and other devices that can collect and transmit information about the objects. Optimum utilization of energy and resources can be achieved by adopting the IoT-related technology. IoT proves to be very helpful to people in their daily routines by making the appliances communicate to each other in an effective manner, thus saving energy and cost.

The IoT enables the multifaceted applications ranging from the micro to the macro, and from the trivial to the critical perspectives. Due to physical objects getting connected and controlled digitally

and centrally with wireless infrastructure regarding the IoT, there is a large amount of automation and control in global operations. Without human intervention, the IoT-related machines are able to communicate with each other leading to faster and timely output. The IoT allows individuals to automate and control the tasks that are accomplished on a daily basis, avoiding human intervention. All the applications of the IoT accomplish in increased comfort, convenience, and better management, thus improving the quality of life.

However, there are many disadvantages of utilizing the IoT. The IoT presents a variety of potential security risks that could be exploited to harm consumers by enabling unauthorized access and misuse of personal information. Prominent technical limitations that may affect the growth and use of the IoT include a lack of new Internet addresses under the most widely used protocol, the availability of high-speed access, wireless communications, and lack of consensus on technical standards. A potential barrier to the development of IoT is the technical limitations of the version of the Internet Protocol that is widely utilized. To solve the IoT-related problems, organizational executives and government officials must work together to create practical standards, develop new technologies, and provide best practices for the Internet and Internet-enabled devices, toward promoting the utilization of the IoT.

REFERENCES

- Atzori, L., Iera, A., & Morabito, G. (2010). The Internet of Things: A survey. *Computer Networks*, 54(15), 2787–2805. doi:10.1016/j.comnet.2010.05.010
- Bechini, A., Marcelloni, F., & Segatori, A. (2014). Low-effort support to efficient urban parking in a smart city perspective. In S. Gaglio & G. Lo Re (Eds.), *Advances onto the Internet of Things* (pp. 233–252). Berlin, Germany: Springer-Verlag. doi:10.1007/978-3-319-03992-3_17
- Caracas, A. (2012). *From business process models to pervasive applications: Synchronization and optimization*. Paper presented at the 2012 IEEE 10th International Conference on Pervasive Computing and Communications (PerCom 2012), Lugano, Switzerland. doi:10.1109/PerComW.2012.6197503
- Chen, Y., Guo, J., & Hu, X. (2010). *The research of Internet of Things' supporting technologies which face the logistics industry*. Paper presented at the 2010 International Conference on Computational Intelligence and Security (CIS 2010), Nanning, China.
- Dar, K., Taherkordi, A., Baraki, H., Eliassen, F., & Geihs, K. (2015). A resource oriented integration architecture for the Internet of Things: A business process perspective. *Pervasive and Mobile Computing*, 20, 145–159. doi:10.1016/j.pmcj.2014.11.005
- Domingo, M. C. (2012). An overview of the Internet of Things for people with disabilities. *Journal of Network and Computer Applications*, 35(2), 584–596. doi:10.1016/j.jnca.2011.10.015
- Eberle, S. (2007). Adaptive Internet integration of field bus systems. *IEEE Transactions on Industrial Informatics*, 3(1), 12–20. doi:10.1109/TII.2006.890525
- Erguler, I. (2015). A potential weakness in RFID-based Internet-of-Things systems. *Pervasive and Mobile Computing*, 20, 115–126. doi:10.1016/j.pmcj.2014.11.001
- Fan, Y., Yin, Y., Xu, L., Zeng, Y., & Wu, F. (2014). IoT based smart rehabilitation system. *IEEE Transactions on Industrial Informatics*, 10(2), 1568–1577. doi:10.1109/TII.2014.2302583
- Gao, C., Ling, Z., & Yuan, Y. (2011). *The research and implement of smart home system based on Internet of Things*. Paper presented at the 2011 International Conference on Electronics, Communications and Control (ICECC 2011), Ningbo, China.

- Gligoric, N., Uzelac, A., Krco, S., Kovacevic, I., & Nikodijevic, A. (2015). Smart classroom system for detecting level of interest a lecture creates in a classroom. *Journal of Ambient Intelligence and Smart Environments*, 7(2), 271–284.
- Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. *Future Generation Computer Systems*, 29(7), 1645–1660. doi:10.1016/j.future.2013.01.010
- Guinard, D., Trifa, V., Karnouskos, S., Spiess, P., & Savio, D. (2010). Interacting with the SOA-based Internet of Things: Discovery, query, selection, and on-demand provisioning of web services. *IEEE Transactions on Services Computing*, 3(3), 223–235. doi:10.1109/TSC.2010.3
- Guo, B., Zhang, D., Yu, Z., Liang, Y., Wang, Z., & Zhou, X. (2013). From the Internet of Things to embedded intelligence. *World Wide Web (Bussum)*, 16(4), 399–420. doi:10.1007/s11280-012-0188-y
- Guo, J., Zhang, H., Sun, Y., & Bie, R. (2014). Square-root unscented Kalman filtering-based localization and tracking in the Internet of Things. *Personal and Ubiquitous Computing*, 18(4), 987–996. doi:10.1007/s00779-013-0713-8
- Hsueh, C. F., & Chang, M. S. (2010). A model for intelligent transportation of perishable products. *International Journal of Intelligent Transportation Systems Research*, 8(1), 36–41. doi:10.1007/s13177-009-0004-y
- Jara, A. J., Parra, M. C., & Skarmeta, A. F. (2014). Participative marketing: Extending social media marketing through the identification and interaction capabilities from the Internet of Things. *Personal and Ubiquitous Computing*, 18(4), 997–1011. doi:10.1007/s00779-013-0714-7
- Jia, X., Feng, Q., Fan, T., & Lei, Q. (2012). *RFID technology and its applications in Internet of Things (IoT)*. Paper presented at the 2012 IEEE 2nd International Conference on Consumer Electronics, Communications and Networks (CECNet 2012), Hubei, China.
- Kasemsap, K. (2015a). The role of radio frequency identification in modern libraries. In S. Thanuskodi (Ed.), *Handbook of research on inventive digital tools for collection management and development in modern libraries* (pp. 361–385). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-8178-1.ch021
- Kasemsap, K. (2015b). The role of cloud computing adoption in global business. In V. Chang, R. Walters, & G. Wills (Eds.), *Delivery and adoption of cloud computing services in contemporary organizations* (pp. 26–55). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-8210-8.ch002
- Kasemsap, K. (2016a). The roles of business process modeling and business process reengineering in e-government. In J. Martins & A. Molnar (Eds.), *Handbook of research on innovations in information retrieval, analysis, and management* (pp. 401–430). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-8833-9.ch015
- Kasemsap, K. (2016b). Examining the roles of virtual team and information technology in global business. In C. Graham (Ed.), *Strategic management and leadership for systems development in virtual spaces* (pp. 1–21). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-9688-4.ch001
- Kumar, J. S., & Patel, D. R. (2014). A survey on Internet of Things: Security and privacy issues. *International Journal of Computers and Applications*, 90(11), 20–26. doi:10.5120/15764-4454
- Lee, I., & Lee, K. (2015). The Internet of Things (IoT): Applications, investments, and challenges for enterprises. *Business Horizons*, 58(4), 431–440. doi:10.1016/j.bushor.2015.03.008

- Li, S., Xu, L. D., & Zhao, S. (2015). The Internet of Things: A survey. *Information Systems Frontiers*, 17(2), 243–259. doi:10.1007/s10796-014-9492-7
- Li, Y., Hou, M., Liu, H., & Liu, Y. (2012). Towards a theoretical framework of strategic decision, supporting capability and information sharing under the context of Internet of Things. *Information Technology & Management*, 13(4), 205–216. doi:10.1007/s10799-012-0121-1
- Mahlknecht, S., & Madani, S. (2007). *On architecture of low power wireless sensor networks for container tracking and monitoring applications*. Paper presented at the 2007 IEEE 5th International Conference on Industrial Informatics (INDIN 2007), Vienna, Austria. doi:10.1109/INDIN.2007.4384782
- Mashal, I., Alsaryrah, O., Chung, T. Y., Yang, C. Z., Kuo, W. H., & Agrawal, D. P. (2015). Choices for interaction with things on Internet and underlying issues. *Ad Hoc Networks*, 28, 68–90. doi:10.1016/j.adhoc.2014.12.006
- Miorandi, D., Sicari, S., & Pellegrini, F. D. (2012). Internet of Things: Vision, applications and research challenges. *Ad Hoc Networks*, 10(7), 1497–1516. doi:10.1016/j.adhoc.2012.02.016
- Moschakis, I. A., & Karatza, H. D. (2015). Towards scheduling for Internet-of-Things applications on clouds: A simulated annealing approach. *Concurrency and Computation*, 27(8), 1886–1899. doi:10.1002/cpe.3105
- Nguyen, K. T., Laurent, M., & Oualha, N. (2015). Survey on secure communication protocols for the Internet of Things. *Ad Hoc Networks*, 32, 17–31. doi:10.1016/j.adhoc.2015.01.006
- Ning, H., & Hu, S. (2012). Technology classification, industry, and education for future Internet of Things. *International Journal of Communication Systems*, 25(9), 1230–1241. doi:10.1002/dac.2373
- Pang, Z., Zheng, L., Tian, J., Kao-Walter, S., Dubrova, E., & Chen, Q. (2015). Design of a terminal solution for integration of in-home health care devices and services towards the Internet-of-Things. *Enterprise Information Systems*, 9(1), 86–116. doi:10.1080/17517575.2013.776118
- Perera, C., Zaslavsky, A., Christen, P., & Georgakopoulos, D. (2014). Sensing as a service model for smart cities supported by Internet of Things. *Transactions on Emerging Telecommunications Technologies*, 25(1), 81–93. doi:10.1002/ett.2704
- Raza, S., Duquennoy, S., Höglund, J., Roedig, U., & Voigt, T. (2014). Secure communication for the Internet of Things: A comparison of link-layer security and IPsec for 6LoWPAN. *Security and Communication Networks*, 7(12), 2654–2668. doi:10.1002/sec.406
- Rekleitis, E., Rizomiliotis, P., & Gritzalis, S. (2014). How to protect security and privacy in the IoT: A policy-based RFID tag management protocol. *Security and Communication Networks*, 7(12), 2669–2683. doi:10.1002/sec.400
- Ren, W. (2011). QoS-aware and compromise-resilient key management scheme for heterogeneous wireless Internet of Things. *International Journal of Network Management*, 21(4), 284–299. doi:10.1002/nem.794
- Salim, F., & Haque, U. (2015). Urban computing in the wild: A survey on large scale participation and citizen engagement with ubiquitous computing, cyber physical systems, and Internet of Things. *International Journal of Human-Computer Studies*, 81, 31–48. doi:10.1016/j.ijhcs.2015.03.003
- Shelby, Z. (2010). Embedded web services. *IEEE Wireless Communications*, 17(6), 52–57. doi:10.1109/MWC.2010.5675778
- Shelby, Z., & Bormann, C. (2009). *6LoWPAN: The wireless embedded Internet*. Chichester, UK: John Wiley & Sons. doi:10.1002/9780470686218

- Shrouf, F., & Miragliotta, G. (2015). Energy management based on Internet of Things: Practices and framework for adoption in production management. *Journal of Cleaner Production*, 100, 235–246. doi:10.1016/j.jclepro.2015.03.055
- Sonntag, S., & Suomi, H. (2015). Economic feasibility of multipath protocols in mobile Internet of Things applications. *Concurrency and Computation*, 27(8), 1913–1931. doi:10.1002/cpe.3127
- Tao, Z., Yajuan, Q., Deyun, G., Junqi, D., & Hongke, Z. (2010). *A practical deployment of Intelligent building wireless sensor network for environmental monitoring and air-conditioning control*. Paper presented at the 2010 IEEE 10th International Conference on Network Infrastructure and Digital Content (IC-NIDC 2010), Beijing, China.
- Theodoridis, E., Mylonas, G., & Chatzigiannakis, I. (2013). *Developing an IoT smart city framework*. Paper presented at the 4th International Conference on Information, Intelligence, Systems and Applications (IISA 2013), Piraeus, Greece.
- Tsai, C. W., Lai, C. F., & Vasilakos, A. V. (2014). Future Internet of Things: Open issues and challenges. *Wireless Networks*, 20(8), 2201–2217. doi:10.1007/s11276-014-0731-0
- Usman, A., & Shami, S. H. (2013). Evolution of communication technologies for smart grid applications. *Renewable & Sustainable Energy Reviews*, 19, 191–199. doi:10.1016/j.rser.2012.11.002
- Uzelac, A., Gligoric, N., & Krco, S. (2015). A comprehensive study of parameters in physical environment that impact students' focus during lecture using Internet of Things. *Computers in Human Behavior*, 53, 427–434. doi:10.1016/j.chb.2015.07.023
- Wang, C. L., Xie, H. D., Wang, Z. D., & Jia, S. Y. (2011). Application research on three-dimensional virtual mine in the framework of Internet of Things. *Journal of Coal Science and Engineering (China)*, 17(2), 212–216. doi:10.1007/s12404-011-0220-3
- Wang, Y., Attebury, G., & Ramamurthy, B. (2006). A survey of security issues in wireless sensor networks. *IEEE Communications Surveys and Tutorials*, 8(2), 2–23. doi:10.1109/COMST.2006.315852
- Whitmore, A., Agarwal, A., & Xu, L. D. (2015). The Internet of Things: A survey of topics and trends. *Information Systems Frontiers*, 17(2), 261–274. doi:10.1007/s10796-014-9489-2
- Xie, W., Shi, Y., Xu, G., & Xie, D. (2001). *Smart classroom: An intelligent environment for tele-education*. Paper presented at the 2001 IEEE 2nd Pacific Rim Conference on Multimedia, Beijing, China. doi:10.1007/3-540-45453-5_85
- Yao, W., Chu, C. H., & Li, Z. (2012). The adoption and implementation of RFID technologies in healthcare: A literature review. *Journal of Medical Systems*, 36(6), 3507–3525. doi:10.1007/s10916-011-9789-8 PMID:22009254
- Ziegeldorf, J. H., Morchon, O. G., & Wehrle, K. (2014). Privacy in the Internet of Things: Threats and challenges. *Security and Communication Networks*, 7(12), 2728–2742. doi:10.1002/sec.795

ADDITIONAL READING

- Bandyopadhyay, S., Sengupta, M., Maiti, S., & Dutta, S. (2011). Role of middleware for Internet of Things: A study. *International Journal of Computer Science and Engineering Survey*, 2(3), 94–105. doi:10.5121/ijcses.2011.2307
- Benssalah, M., Djeddou, M., & Drouiche, K. (2014). Security enhancement of the authenticated RFID security mechanism based on chaotic maps. *Security and Communication Networks*, 7(12), 2356–2372. doi:10.1002/sec.946
- Carretero, J., & Garcia, J. D. (2014). The Internet of Things: Connecting the world. *Personal and Ubiquitous Computing*, 18(2), 445–447. doi:10.1007/s00779-013-0665-z

- Chang, J. Y. (2015). A distributed cluster computing energy-efficient routing scheme for Internet of Things systems. *Wireless Personal Communications*, 82(2), 757–776. doi:10.1007/s11277-014-2251-8
- Clarke, J., Gritzalis, S., Zhou, J., & Roman, R. (2014). Protecting the Internet of Things. *Security and Communication Networks*, 7(12), 2637–2638. doi:10.1002/sec.1174
- Deng, H., Zhang, B., Li, C., Huang, K., & Liu, H. (2012). MAX–MIN aggregation in wireless sensor networks: Mechanism and modeling. *Wireless Communications and Mobile Computing*, 12(7), 615–630. doi:10.1002/wcm.1000
- Ding, Y., Zhou, X. W., Cheng, Z. M., & Lin, F. H. (2013). A security differential game model for sensor networks in context of the Internet of Things. *Wireless Personal Communications*, 72(1), 375–388. doi:10.1007/s11277-013-1018-y
- Forsstrom, S., & Kanter, T. (2014). Continuously changing information on a global scale and its impact for the Internet-of-Things. *Mobile Networks and Applications*, 19(1), 33–44. doi:10.1007/s11036-013-0479-2
- Granjal, J., Monteiro, E., & Silva, J. S. (2014). Network-layer security for the Internet of Things using TinyOS and BLIP. *International Journal of Communication Systems*, 27(10), 1938–1963. doi:10.1002/dac.2444
- Hsieh, H. C., Hsieh, W. H., & Chen, J. L. (2015). Mobile IMS integration of the Internet of Things in ecosystem. *Wireless Personal Communications*, 80(2), 819–836. doi:10.1007/s11277-014-2043-1
- Hu, W., & Cheng, T. (2015). Simulative research on the function of Internet of Things basing on the changing of topological structure. *Multimedia Tools and Applications*, 74(19), 8445–8454. doi:10.1007/s11042-013-1615-5
- Liu, Y., Chen, Z., Lv, X., & Han, F. (2014). Multiple layer design for mass data transmission against channel congestion in IoT. *International Journal of Communication Systems*, 27(8), 1126–1146. doi:10.1002/dac.2399
- Lopez, T. S., Ranasinghe, D. C., Harrison, M., & McFarlane, D. (2012). Adding sense to the Internet of Things. *Personal and Ubiquitous Computing*, 16(3), 291–308. doi:10.1007/s00779-011-0399-8
- Luo, M., Tu, M., & Xu, J. (2014). A security communication model based on certificateless online/offline signcryption for Internet of Things. *Security and Communication Networks*, 7(10), 1560–1569.
- Manches, A., Duncan, P., Plowman, L., & Sabeti, S. (2015). Three questions about the Internet of Things and children. *TechTrends*, 59(1), 76–83. doi:10.1007/s11528-014-0824-8
- Mustafee, N., & Bessis, N. (2015). The Internet of Things: Shaping the new Internet space. *Concurrency and Computation*, 27(8), 1815–1818. doi:10.1002/cpe.3483
- Najera, P., Roman, R., & Lopez, J. (2013). User-centric secure integration of personal RFID tags and sensor networks. *Security and Communication Networks*, 6(10), 1177–1197.
- Qiu, T., Sun, W., Bai, Y., & Zhou, Y. (2013). An efficient multi-path self-organizing strategy in Internet of Things. *Wireless Personal Communications*, 73(4), 1613–1629. doi:10.1007/s11277-013-1270-1
- Sun, Y., Bie, R., Thomas, P., & Cheng, X. (2014). Advances on data, information, and knowledge in the Internet of Things. *Personal and Ubiquitous Computing*, 18(8), 1793–1795. doi:10.1007/s00779-014-0789-9
- Tsai, C. W., Lai, C. F., Chiang, M. C., & Yang, L. T. (2014). Data mining for Internet of Things: A survey. *IEEE Communications Surveys and Tutorials*, 16(1), 77–97. doi:10.1109/SURV.2013.103013.00206

Wang, Y., Zhang, L., Huang, Y., Zhao, B., & Zhao, C. (2013). Service community construction method of Internet of Things based on semantic similarity. *Journal of Electronics (China)*, 30(1), 49–56. doi:10.1007/s11767-013-2121-7

Wortmann, F., & Fluchter, K. (2015). Internet of Things. *Business & Information Systems Engineering*, 57(3), 221–224. doi:10.1007/s12599-015-0383-3

Wu, T. Y., Liaw, G. H., Huang, S. W., Lee, W. T., & Wu, C. C. (2012). A GA-based mobile RFID localization scheme for Internet of Things. *Personal and Ubiquitous Computing*, 16(3), 245–258. doi:10.1007/s00779-011-0398-9

Xia, F., Yang, L. T., Wang, L., & Vinel, A. (2012). Internet of Things. *International Journal of Communication Systems*, 25(9), 1101–1102. doi:10.1002/dac.2417

Xu, L., He, W., & Li, S. (2014). Internet of Things in industries: A survey. *IEEE Transactions on Industrial Informatics*, 10(4), 2233–2243. doi:10.1109/TII.2014.2300753

Zhou, L., & Chao, H. C. (2011). Multimedia traffic security architecture for the Internet of Things. *IEEE Network*, 25(3), 35–40. doi:10.1109/MNET.2011.5772059

KEY TERMS AND DEFINITIONS

Computer: The electronic machine that is used for storing, organizing, and finding words, numbers, and pictures, for doing calculations, and for controlling other machines.

Internet: The worldwide computer network that provides information on many subjects and enables users to exchange messages.

Network: A number of computers that are connected together so that they can share information.

Radio Frequency Identification: The automatic identification of packages, products, machinery, etc., through attached transponders.

Sensor: The device that is utilized to record that something is present or that there are changes in something.

Technology: The use of scientific knowledge to solve practical problems, especially in industry and commerce.

The Internet of Things: All the different devices, including computers, phones, wearable technology, and smart systems, that are able to connect to each other utilizing the Internet.

The Role of U-FADE in Selecting Persuasive System Features



Isaac Wiafe

Ghana Institute of Management and Public Administration, Ghana

INTRODUCTION

Behavior Change Support Systems (BCSS) are systems that combine properties of interpersonal interaction and mass communication with the support of technology to change or alter human behavior or attitude (Oinas-kukkonen, 2013). It originated from Persuasive Technologies (PT) or Captology; which is the use of interactive technology for changing human attitude or behavior to a predetermined one (Fogg 1997). Due to its ability to adapt to individual differences, it is considered to have a greater potential for changing human behavior and/or attitude. This is because, it is capable of employing both animated and non-animated objects to persuaded its user. Thus it has been applied in areas of healthcare (Mateevitsi et al., 2014; Rana & Saleem, 2014), leisure and recreation (Sra & Schmandt, 2013; Tieben, Sturm, Bekker, & Schouten, 2013), energy saving (Emeakaroha, Ang, & Yan, 2012; Wunsch, Stibe, Millonig, & Seer, 2015) Information security (Kegel & Wieringa, 2015; Yeo, Rahim, & Ren, 2009), etc. Nonetheless, BCSS development is faced with a number of challenges and one of the most prominent is that majority of designers fail to use appropriate methods for its design (Wiafe, 2013). As a result, some BCSS applications become obsolete with time[2]. In addition, most of the existing frameworks or design approaches do not provide adequate information that can be used for analyzing and designing applications that address changing needs of users during use

ADDIN CSL_CITATION { "citationItems": [{ "id": "ITEM-1", "itemData": { "DOI": "10.1007/s00779-014-0782-3", "ISSN": "1617-4909", "author": [{ "dropping-particle": "", "family": "Wiafe", "given": "Isaac", "non-dropping-particle": "", "parse-names": false, "suffix": "" }, { "dropping-particle": "", "family": "Nakata", "given": "Keiichi", "non-dropping-particle": "", "parse-names": false, "suffix": "" }, { "dropping-particle": "", "family": "Gulliver", "given": "Stephen", "non-dropping-particle": "", "parse-names": false, "suffix": "" }], "container-title": "Personal and Ubiquitous Computing", "id": "ITEM-1", "issued": { "date-parts": [["2014", "6", "29"]] }, "title": "Categorizing users in behavior change support systems based on cognitive dissonance", "type": "article-journal" }, { "id": "ITEM-2", "itemData": { "ISBN": "9781605583761", "author": [{ "dropping-particle": "", "family": "Torning", "given": "Kristian", "non-dropping-particle": "", "parse-names": false, "suffix": "" }, { "dropping-particle": "", "family": "Hall", "given": "Cordura", "non-dropping-particle": "", "parse-names": false, "suffix": "" }, { "dropping-particle": "", "family": "Oinas-kukkonen", "given": "Harri", "non-dropping-particle": "", "parse-names": false, "suffix": "" }], "id": "ITEM-2", "issued": { "date-parts": [["2009"]] }, "title": "Persuasive System Design: State of the Art and Future Directions", "type": "article-journal" }, { "id": "ITEM-3", "itemData": { "DOI": "10.1007/s00779-014-0782-3", "ISSN": "1617-4909", "author": [{ "dropping-particle": "", "family": "Wiafe", "given": "Isaac", "non-dropping-particle": "", "parse-names": false, "suffix": "" }, { "dropping-particle": "", "family": "Nakata", "given": "Keiichi", "non-dropping-particle": "", "parse-names": false, "suffix": "" }, { "dropping-particle": "", "family": "Gulliver", "given": "Stephen", "non-dropping-particle": "", "parse-names": false, "suffix": "" }], "container-title": "Personal and Ubiquitous Computing", "id": "ITEM-3", "issued": { "date-parts": [["2014", "6", "29"]] }, "title": "Categorizing users in behavior change support systems based on cognitive dissonance", "type": "article-journal" }], "properties": { "noteIndex": 1 } }

DOI: 10.4018/978-1-5225-2255-3.ch677

0 }, "schema": "https://github.com/citation-style-language/schema/raw/master/csl-citation.json" }.

Recently, the Unified Framework for Analyzing, Designing and Evaluating persuasive systems (U-FADE) was proposed (Wiafe 2013). It incorporated key concepts from existing persuasive design models such as the Functional Triad (B. Fogg, 1998)(Wiafe et al., 2014), the Behavior Wizard (B. J. Fogg & Hreha, 2010)[4], the 3-Dimensional Relationship between Attitude and Behavior (3D-RAB) model (Wiafe, Nakata, & Gulliver, 2014) [1], [2] and the Persuasive System Design (PSD) model (Oinas-kukkonen & Harjumaa, 2009) ADDIN CSL_CITATION { "citationItems": [{ "id": "ITEM-1", "itemData": { "author": [{ "dropping-particle": "", "family": "Fogg", "given": "B J", "non-dropping-particle": "", "parse-names": false, "suffix": "" }, { "dropping-particle": "", "family": "Hreha", "given": "Jason", "non-dropping-particle": "", "parse-names": false, "suffix": "" }], "id": "ITEM-1", "issued": { "date-parts": [["0"]] }, "title": "Behavior Wizard: A Method for Matching Target Behaviors with Solutions 1. Overview of Behavior Wizard 3. The Need to Better Classify Behavior Types", "type": "article-journal" }, "uris": ["http://www.mendeley.com/documents/?uuid=8ba74f7f-e3ef-4ef5-83a2-4f19c9f8b4f6"] }], "mendeley": { "formattedCitation": "(B. J. Fogg & Hreha, n.d.)", "plainTextFormattedCitation": "(B. J. Fogg & Hreha, n.d.)", "previouslyFormattedCitation": "[5] }, "properties": { "noteIndex": 0 }, "schema": "https://github.com/citation-style-language/schema/raw/master/csl-citation.json" } to address design challenges associated with previous design methods. However, the framework has not been applied in developing any application to ascertain its practicality, although it has been demonstrated conceptually that it provides a more compressive approach for BCSS analysis and design (Wiafe 2013)(B. J. Fogg & Hreha, n.d.).

This chapter therefore presents an evaluation of the Unified Framework for Analyzing and Designing Persuasive Systems using the case of a weight management system. The various

stages of the U-FADE was followed to redesign an existing mobile application known as ObiMo Pet (Wiafe, 2013). The objective was to ascertain whether the framework is capable of facilitating design or enhancing the selection of persuasive system features.

The chapter is presented as follows: the next section presents a background discussion on U-FADE, this is followed by the method adopted for the study, and then the process in U-FADE is used to redesign the existing application. The findings, discussions, future research directions and conclusions are then presented.

BACKGROUND

Currently, the Unified Framework for Analyzing, Designing and Evaluating persuasive systems (U-FADE) is the only framework that seeks to formalize Behavior Change Support Systems (BCSS) design process with emphasis on the changing needs of users during the change intervention (B. Fogg, 2002)(Wiafe 2013). It comprises analysis of the Persuasion Event, selection of Persuasion Strategy, identification of System Features needed to promote persuasion, Development and Implementation of the persuasive system and the Evaluation of Change (See Figure 1)1. ADDIN CSL_CITATION { "citationItems": [{ "id": "ITEM-1", "itemData": { "DOI": "10.4018/ijcssa.2013070102", "ISSN": "2166-7292", "author": [{ "dropping-particle": "", "family": "Wiafe", "given": "Isaac", "non-dropping-particle": "", "parse-names": false, "suffix": "" }], "container-title": "International Journal of Conceptual Structures and Smart Applications", "id": "ITEM-1", "issue": "2", "issued": { "date-parts": [["2013", "6"]] }, "page": "6-16", "title": "U-FADE: A Unified Approach To Persuasive Systems Development", "type": "article-journal", "volume": "1" }, "uris": ["http://www.mendeley.com/documents/?uuid=8279677e-cf51-47d4-a16a-6fccdb5e387e"] }], "mendeley": { "formattedCitation": "(Wiafe, 2013)", "plain-

TextFormattedCitation": "(Wiafe, 2013)", "previouslyFormattedCitation": "[4]", "properties": { "noteIndex": 0 }, "schema": "https://github.com/citation-style-language/schema/raw/master/csl-citation.json" } The main benefit of using U-FADE is its ability to provide a thinking guide for both novice and expert persuasive system developers. It explores the problem space by identifying variations in users' cognitive dissonance state and the situational (or environmental) context to guide designers to select persuasive features for specific targets. This is made possible by employing activities in the Event Analysis which are the Use and User context analysis: concepts that were borrowed from the PSD model (Oinas-Kukkonen and Harjuma 2009). In User context analysis, the framework emphasized the need for the identification of internal and external factors that affect individuals to change or maintain behavior.

Here, the 3D-RAB model serves as a tool for the designer to group potential users into various categories based on inconsistencies in their attitude and behavior. The next step in Event analysis is Use context analysis. This considers BCSS application design in relation to formal and informal behavior change factors [2]. It is argued that, designers need to consider both formal and informal activities within their immediate environment that may interact or influence potential users to change their behavior. The persuasive technology onion provides steps that enable designers to categorize attitude and behavior change as planned or natural based on existing environmental factors. Natural Attitude or Behavior Change (NABC) takes into account activities within an environment that changes an individual's behavior or attitude naturally. Issues relating to culture, social norms, beliefs, ethics, commitments and values among others are of paramount interest. Arguably, these issues are exceptionally essential for BCSS designs since their key aim is to alter behavior or attitude. ADDIN CSL_CITATION { "citationItems": [{ "id": "ITEM-1", "itemData": { "DOI": "10.4018/ijcssa.2013070102", "ISSN":

"2166-7292", "author": [{ "dropping-particle": "", "family": "Wiafe", "given": "Isaac", "non-dropping-particle": "", "parse-names": false, "suffix": "" }], "container-title": "International Journal of Conceptual Structures and Smart Applications", "id": "ITEM-1", "issue": "2", "issued": { "date-parts": [["2013", "6"]] }, "page": "6-16", "title": "U-FADE: A Unified Approach To Persuasive Systems Development", "type": "article-journal", "volume": "1" }, "uris": ["http://www.mendeley.com/documents/?uuid=8279677e-cf51-47d4-a16a-6fccdb5e387e"] }], "mendeley": { "formattedCitation": "[5]", "plainTextFormattedCitation": "[5]", "previouslyFormattedCitation": "(Wiafe, 2013)", "properties": { "noteIndex": 0 }, "schema": "https://github.com/citation-style-language/schema/raw/master/csl-citation.json" }. Next, Planned Attitude or Behavior Change (PABC) is considered. PABC consists of existing approaches that are used on persuadees to change their behavior or attitude intentionally. This may include automated and non-automated approaches.

The Event Analysis provides designers the ability to identify existing activities that needs to be promoted to ensure that target behavior is achieved. Particularly, the U-FADE approach claims that designers must reinforce the existing overt or covert persuasive activities that are identified during Event analysis. Additionally, it proposes steps that guide designers to identify system features by focusing on incremental changes as suggested by Fogg (2009) and Oinas-Kukkonen (2009). Transitions that involve attitude-changing states must use logical arguments (elaborated messages) to ensure a long-term change, whereas heuristics (peripheral messages) must be used for short-term behavior change.

Designers are also encouraged to identify a list of familiar hardware and select those that the application can be deployed on [5]. In this process, they must pay attention to cost, environment and obtrusiveness of the technology they chose. Specifically, the introduction of a new hardware should be considered or used only when the ap-

plication cannot be administered effectively on an existing device or technology that is familiar to the potential persuadees.

The framework also provides a guide on how 28 system features (Oinas-kukkonen & Harjumaa, 2009; Torning, Hall, & Oinas-kukkonen, 2009) may support either elaboration or peripheral message routes (Wiafe et al., 2014). The approach is based on findings from the Use and User analysis. It is also emphasized that to enable designers to focus on design-relevant issues, a Transition Description Card should be used. The card provides a template for summarizing all the activities performed during the analysis.

The last step in U-FADE is the provision of methods for evaluating changes in persuadees to enable designers to modify system features during use.

From the above discussions it can be concluded that U-FADE's approach for developing BCSS applications seeks to gather and analyze user requirements by identifying: the distribution of cognitive dissonance levels of potential user, existing activities that supports or prevent potential users from performing the target behavior and select persuasive system features and device based on these identified activities.

As already mentioned, and to the best of our knowledge, the framework has not been applied for the design or development of any application, thus one cannot substantiate its appropriateness and effectiveness with any empirical evidence. Although it can be argued that the author conceptually demonstrated and evaluated its effectiveness, relevance and design implications, it is essential for it to be demonstrated practically.

The next section is a discussion on methods that were adopted to evaluate the benefits of using U-FADE for designing BCSS.

RESEARCH APPROACH

A designed artifact must map adequately to the real world, solve a problem and its relevance must

be demonstrated to justify the need for a design research (Hevner, March, Park, Ram, & Ram, 2004). Hence to practically demonstrate whether the framework is useful, it was used to develop a BCSS application. This is to say that U-FADE was applied to the development of an existing BCSS application that seeks to encourage its users to reduce calorie intake and manage their weight. The application considered in this case is known as ObiMo Pet (Wiafe & Nakata, 2012) and the system features that were realized after the analysis were compared to the existing one.

Potential Persuadees

Employees of Accra Polytechnic were considered for the study as potential users (potential persuadees). The Polytechnic is a tertiary institution located in Accra, the capital of Ghana. It was established in 1949 as the first polytechnic in the country. Currently, it offers Higher National Diploma (HND) programs in Science and Engineering, Fashion Design, Social Science, Hotel Catering and Institutional Management, etc. with a total staff population of over 600.

The Institute has keen interest in issues relating to staff healthcare thus it has recently built a state of the art gymnasium for staff. Additionally, healthier lifestyle seminars series are organized regularly to refresh staff on the need for maintaining good health by the Hotel Catering and Institutional Management department. It has a strong staff welfare association that also seeks to promote social welfare activities within the Institute. Amidst all these efforts, the ratios of overweight and obese persons continue to increase.

The ObiMo Pet

ObiMo Pet (Torning et al., 2009) is a weight management system designed to manage calorie intake and physical activities of its users. It uses persuasive strategies to motivate users to maintain healthier lifestyles as they cater for a virtual pet of their choice. It provides them with physical activities

and diet plans that seek to manage their weight. As persuadees adhere to suggestions provided by the application, they earn points to decorate their virtual pets. Also, ObiMo recognizes persuadees whose pet is the most decorated in the virtual community. This encourages them to compete with their peers as they seek to accomplish their target goals (i.e. manage their weight).

Six system features were used in the application to promote behavior change. These are reward, suggestion, tailoring, recommendation, competition and personalization. In addition, the semiotic approach to information systems development (Liu, 2004, p. 37) was adopted as the foundation for analysis and design.

Design Objectives

The study targeted a weight loss between 3 and 5 pounds a week, and consumption of 2300 calories a day. It is however acknowledge that the actual amount of calorie intake and calorie burnt is oversimplified in this study; since the values normally varies according to the individual's gender, type of work or activity level, age, body weight, etc.

Additionally, although U-FADE suggests the reassessment of user behavior to evaluate incremental behavior change, this was not considered as part of the objectives of this study. The next section reports findings from the analysis.

ANALYSIS AND DESIGN

Event Analysis

As apart of User Context analysis (the first step in Event Analysis), a questionnaire was developed to collect data and classify users into various cognitive dissonance states

```
ADDIN CSL_CITATION {
  "citationItems": [ { "id": "ITEM-1", "itemData": {
    "DOI": "10.4018/ijcssa.2013070102", "ISSN": "2166-7292", "author": [ { "dropping-particle": "", "family": "Wiafe", "given": "Isaac", "non-dropping-particle": "", "parse-names": false,
```

```
  "suffix": "" } ], "container-title": "International Journal of Conceptual Structures and Smart Applications", "id": "ITEM-1", "issue": "2", "issued": { "date-parts": [ [ "2013", "6" ] ] }, "page": "6-16", "title": "U-FADE: A Unified Approach To Persuasive Systems Development", "type": "article-journal", "volume": "1", "uris": [ "http://www.mendeley.com/documents/?uuid=8279677e-cf51-47d4-a16a-6fccdb5e387e" ] } ], "mendeley": { "formattedCitation": "(Wiafe, 2013)", "plainTextFormattedCitation": "(Wiafe, 2013)", "previouslyFormattedCitation": "[4]", "properties": { "noteIndex": 0 }, "schema": "https://github.com/citation-style-language/schema/raw/master/csl-citation.json" }. The classification was based on variations in cognitive dissonance (Festinger, 1957) of users' attitude and behavior (i.e. Attitude Towards Target Behavior, Attitude Towards Change or Maintaining Behavior, and Current Behavior (CB)) in relation to food menu goals (calorie intake) and workout goals (calories burnt). Target behavior was considered as maintaining a Body Mass Index (BMI) value between 18.5 and 25.
```

A pilot study was conducted to test the clarity and consistency of the questionnaire and the suggested amendments were made. Finally, the questionnaire was distributed to staff members of the Institution. One hundred and seventeen responses were received after one month of distribution. Three questions each were asked to measure respondents' Attitude Towards Target Behavior (ATTB) and Attitude Towards Change or Maintaining Behavior (ATCMB) in order to ensure reliable and consistent responses. Cronbach's alpha with reliability coefficient of 0.7 was used to check for reliability. Respondents were classified into positives or negative ATTB, ATCMB or CB by computing the averages of the numerical values they assigned to each question. The distribution of cognitive dissonance states for the 117 responses is presented in Table 1.

From Table 1, thirty-nine respondents (representing 33%) were identified to be in state 1 where all three values (ATTB, ATCMB and CB)

Table 1. Distribution of cognitive dissonance state observed

State	CB	ATTB	ATCMB	Response	Percent (%)
1	+	+	+	39	33
2	+	+	-	12	10
3	+	-	+	0	0
4	+	-	-	6	5
5	-	+	+	42	35
6	-	+	-	6	5
7	-	-	+	0	0
8	-	-	-	12	10
Total				117	100

are positive and users are considered to be in the “ideal” state. With all things being equal they are expected to continue to maintain a BMI that is between 18.5 and 25. However majority of the respondents were found to be in states that need a change in either attitude (ATTB/ATCMB) or behavior (CB). None of the respondents was in state 7, i.e. no staff was identified to have a BMI that is not between 18.5 and 25, and also do not believe that they should maintain a BMI value between 18.5 and 25 but would want to change. Likewise, no respondent was in state 3, where CB is positive, ATCMB is positive, but ATTB is negative. Forty-two respondents (representing 35%) were identified to be in state 5. These are individuals who do not have BMI between 18.5 and 25, but believe that there is a need for them to change their current BMI values. According to Wiafe, Nakata, and Gulliver (2014) these users experience a strong form of cognitive dissonance which serves as a motivating factor for changing their behavior. Table 1 indicates that, potential users of the system are experiencing different forms of cognitive dissonance levels. This therefore suggests that it will require different persuasive system features or properties to persuade them. Although in U-FADE it is argued that persuasive systems or technologies should ensure that they provide specific persuasive features for all states, for the purpose of this study only users in state 5 were considered.

The next stage of the analysis collected information regarding the Use context (i.e. PABC and NABC). Informal interactions, interviews and discussions were used to gather relevant information pertaining to this. Seven observations were identified to impact staff attitude and behavior in relation to maintaining a healthier BMI. Some of these are societal norms and culture, availability and access to technology, the existence of staff association, etc. The observations and associated suggested persuasive features identified to be appropriate for users in state 5 are summarized in Table 2.

Persuasive Strategy

The selection of technology must aim at identifying the most convenient hardware that is readily available, affordable and also familiar to potential users of the system[5]. Hence a mobile device was considered to be appropriate since each staff member owns a mobile device. This was realized during the Use analysis. Also since users with positive ATTB and ATCMB who perform negative behavior (not maintaining a between 18.5 and 25) are more likely to change if peripheral messages are used (Wiafe, 2013), the peripheral route was considered to be the most appropriate channel for the design.

Table 2. Selected system features identified to promote weight management for the target population

Observation	Description	System Feature (Peripheral route)	Implementation
Societal norms	Ghanaians perceive overweight people as wealthier and rich: this serves as a negative impact on the target behavior	Social facilitation	The system should provide a platform for discussion to counter the notion associated with being overweight
Hotel Catering and Institutional Management department	There is a Nutrition and Home Science unit at the institute with scholars in nutrition	Social role	An avatar of one of the renowned scholars in the unit should be used to suggest healthier menus
Intellectual and competitive environment	This is an educational institution made up of intellectuals who prefer to demonstrate that they can march their peers	Social comparison Competition	The system should provide a means for peers to compete
Staff association	There is an association for staff of Accra polytechnic employees: they meet regularly to discuss their welfare	Normative influence	The system should provide a means to make users know that the target behavior is a collective goal
Expensive and Poor Healthcare facilities	Healthcare cost is becoming expensive and in addition there are few healthcare facilities in the country	Reminders	The system should provide a means of reminding users of the increasing cost of healthcare and poor medical facilities
Rising cost of food prices	The cost of food prices continuous to increase due to high inflation	Suggestion	Suggest alternative cheaper and healthier menus
Gymnasium	The availability of a state of the art gym serves as a positive impact to the target behavior Some staff also believe that they are taxed indirectly to maintain the gym	Reminder	Prompt staff to use the gym regularly since they believe they are paying for its maintenance
		Suggestion	An avatar of the gym instructor should be used to suggest exercise schedules

Selection of System Features

Although a number of system features were identified during the analysis, only those that support heuristics (peripheral messages) were considered. This is because the study limited its design principles to users in state 5 (those with positive attitude), hence there is no need to change their attitude. Rather, emphasis is given to peripheral messages that promote behavior change. Social facilitation, Social role, Social comparison, Competition, Normative influence, Suggestions and Reminders were identified to be appropriate for promoting behavior change in this situation. Social role was identified based on the fact that there is a Nutrition and Home Science unit (see Table 2) within the Hotel Catering and Institutional Management department at the Institute. This unit has scholars in nutrition. Since the staff members

are familiar with their colleagues who are experts in areas of nutrition and wellbeing, they are more likely to accept persuasive messages from an avatar (a virtual nutrition scholar) of one of their colleagues who is an expert in the field.

Likewise, the existence of a staff welfare association (see table 2) at the Institute promotes normative influence; consequently, normative influence was also considered as an appropriate system feature.

Refer to Table 2 for the list of system features identified and its associated justification for the selection of these features.

As required by the framework, a Transition Description Card (TDC) was completed for transition 5 → 1 (i.e. changing users with the right attitude but a negative behavior). The assumptions, the type of change, the type of message, constraints, etc. are highlighted on the TDC for the specific

transition. Figure 2 is the TDC for transition 5 → 1 that was realized during analysis.

FINDINGS AND DISCUSSIONS

The analysis facilitated the selection of a specific target group that is “better defined” as compared to the original application. In ObiMo Pet the system was not designed for any particular target group. Indeed, it was assumed that the issue of weight management is generic and thus the designers did not deem it necessary to identify target specific issues. The use of the framework facilitated the identification of a particular group of users (those in state 5). Consequently, it focused its persuasive activities and system features on this type of users. Although one can argue that the other groups of users also need attention, this characteristic of the U-FADE enables the designer to plan or target a specific potential user needs.

Again, even though the analysis focused on users in state 5, it identified seven system features. However five of them were not the same as the ones used in ObiMo Pet. The seven features observed are social role, normative influence, competition, suggestion, social facilitation, social comparison and reminders as compared to tailoring, competition, rewards, suggestions, recommendation and reminders. Out of these, suggestion, competition and reminders were common to both, whereas social facilitation, social comparison, social role and normative influence were not present in ObiMo Pet.

One may therefore argue that ObiMo Pet’s failure to identify these features may be due to the fact that it did not adopt an appropriate approach to analyze the design of the BCSS application. Rather, the designers used a generic analysis and design methodology (the semiotic approach to information systems development). More importantly, the approach used for analyzing and designing of ObiMo Pet failed to identify these features although the application targeted a larger population as compared to the target users

for this analysis (potential users in state 5). This therefore suggests that using generic Information Systems design methods should not be encouraged in BCSS development.

Also, all the suggested system features seek to promote an existing condition that favors the target behavior or refute an existing belief or perception that impedes the performance of the target behavior. This was not the same in ObiMo Pet. This is to say that, the selection and use of system features was aimed at targeting only behavior change (see table 2). Thus they were all peripheral routes of persuasion.

ObiMo Pet did not provide any substantive justification for the selection and use of the persuasive features. It appears that the designers used an arbitrary approach in the selection process: a practice that is common in existing BCSS applications.

From the study it was observed that one crucial issue appears to be largely ignored in BCSS designs. This relates to how system features are implemented. It was observed that although some system features used in ObiMo Pet were also identified during the analysis, they were implemented differently. For example, In ObiMo Pet, reminders were used to remind users of their daily activities whereas in the analysis it was observed that reminders should be used to make staff use the gym regularly since they pay for the service indirectly. Hence in ObiMo Pet, reminders are used for elaborated messages (targeting attitude change) whereas they are used for peripheral messages (targeting behavior change) in the proposed system. Similarly, suggestion and competition are implemented differently.

With respect to hardware selection, the analysis advocated for the use of a mobile devices, however in ObiMo Pet both mobile and PC were used to implement the application. Again, there was no justification for the need of a PC interface for a weight management system. As observed from the analysis, majority of the users would be comfortable using an app on a mobile device. It is however important to state that ObiMo Pet did consider a

larger audience as compared to what was used for the study. Hence, there is the possibility that the use of a PC may be relevant for other targeted groups that were not considered in this study.

FUTURE RESEARCH DIRECTIONS

Our findings indicated that the framework fails to provide the exact messages that should be used for persuasion, rather it provides specification for the message: suggesting instance in which peripheral messages would be appropriate and those that will require elaboration. Nonetheless, the study cannot conclude that the introduction of additional system features to ObiMo Pet will make it more effective, since there is no empirical evidence to support such a claim. It is therefore recommended that, future research should investigate the relationship between the number of persuasive features and the persuadability of an application. This is to say that the proposed system must be implemented and evaluated in terms of its effectiveness in changing user behavior.

In addition, there is need for further studies to investigate into how system features should be implemented effectively. As observed above although some system features were common to both ObiMo Pet and the proposed application, they aim to present different persuasive messages. This raises a number of research questions relating to the appropriateness of how a system feature is implemented. Specifically will the manner in which a system feature is presented have an effective of the persuadability of a system?

Lastly, although there are a number of studies that have demonstrated that cultural and gender issues play a key role in the selection and design of persuasive system features (Orji, Vassileva, & Mandryk, 2014), the framework failed to establish how gender impacted the selection of persuasive features. Hence there is the need for further studies to identify how the framework caters for gender differences.

CONCLUSION

This chapter has demonstrated the practicality of the Unified Framework for Analyzing and Designing persuasive systems. The design principles of the framework were tested on the objectives of a persuasive weight management system (ObiMo Pet).

Our findings supported the claim that there are variations in cognitive dissonance of users. Users were found in six (6) out of the eight (8) listed states proposed by Wiafe, Nakata, and Gulliver (2014). It was also observed that by applying the U-FADE, newer system features emerge although there was no evidence that these system features will improve the persuadability of a system.

REFERENCES

- Emeakaroha, A., Ang, C. S., & Yan, Y. (2012). *Challenges in Improving Energy Efficiency in a University Campus Through the Application of Persuasive Technology and Smart Sensors*. <http://doi.org/10.3390/challe3020290>
- Festinger, L. (1957). *A theory of cognitive dissonance*. Evanston, IL: Peterson.
- Fogg, B. (1998). Persuasive computers: perspectives and research directions. In SIGCHI conference on Human factors in computing (Vol. 98, pp. 225–232). <http://doi.org/doi:10.1145/274644.274677>
- Fogg, B. J. (2009). Creating Persuasive Technologies : An Eight-Step Design Process. In *Persuasive Technology* (Vol. 91, pp. 1–6). <http://doi.org/10.1145/1541948.1542005>
- Fogg, B. J., & Hreha, J. (2010). Behavior Wizard : A Method for Matching Target Behaviors with Solutions. In *Persuasive Technology* (Vol. 6137, pp. 117–131). Springer Berlin Heidelberg. [doi:10.1007/978-3-642-13226-1_13](http://doi.org/10.1007/978-3-642-13226-1_13)

- Kegel, R. H. P., & Wieringa, R. J. (2015). *Behavior Change Support Systems for Privacy and Security*. Academic Press.
- Liu, K. (2004). *Semiotics in Information Systems Engineering*. Cambridge Press.
- Mateevitsi, V., Johnson, A. E., Mateevitsi, V., Reda, K., Leigh, J., & Johnson, A. (2014). *The Health Bar: A Persuasive Ambient Display to Improve the Office Worker's Well Being*. <http://doi.org/10.1145/2582051.2582072>
- Oinas-kukkonen, H. (2013). A foundation for the study of behavior change support systems A foundation for the study of behavior change support systems. *Personal and Ubiquitous Computing*, 17(6), 1223–1235. doi:10.1007/s00779-012-0591-5
- Oinas-kukkonen, H., & Harjumaa, M. (2009). Persuasive Systems Design : Key Issues, Process Model, and System Features. *Communications of the Association for Information Systems*, 24(1), 28.
- Orji, R., Vassileva, J., & Mandryk, R. L. (2014). Modeling the efficacy of persuasive strategies for different gamer types in serious games for health. *User Modeling and User-Adapted Interaction*, 24(5), 453–498. doi:10.1007/s11257-014-9149-8
- Rana, A. M., & Saleem, H. (2014). *Novel Integrated Sensor based Sleep Apnea Monitoring and Tracking System using Soft Computing and Persuasive Technology for Healthcare Support*. Academic Press.
- Sra, M., & Schmandt, C. (2013). Spotz : A location-based approach to self-awareness. In *Persuasive Technology*, (pp. 216–221). Springer Berlin Heidelberg. doi:10.1007/978-3-642-37157-8_26
- Tieben, R., Sturm, J., Bekker, M. M., & Schouten, B. (2013). An Active Lifestyle for Youths through Ambient Persuasive Technology. Implementing Activating Concepts in a School Environment. In *Games for Health* (pp. 293–308). Springer Fachmedien Wiesbaden; doi:10.1007/978-3-658-02897-8
- Torning, K., Hall, C., & Oinas-kukkonen, H. (2009). *Persuasive System Design: State of the Art and Future Directions*. Academic Press.
- Wiafe, I. (2013). U-FADE: A Unified Approach To Persuasive Systems Development. *International Journal of Conceptual Structures and Smart Applications*, 1(2), 6–16. doi:10.4018/ijcssa.2013070102
- Wiafe, I., & Nakata, K. (2012). A semiotic analysis of persuasive technology: An application to obesity management. In *International conference on informatics and semiotics in organisations*. Academic Press.
- Wiafe, I., Nakata, K., & Gulliver, S. (2014). Categorizing Users in Behavior Change Support Systems based on Cognitive Dissonance. *Personal and Ubiquitous Computing*, 18(7), 1677–1687. doi:10.1007/s00779-014-0782-3
- Wunsch, M., Stibe, A., Millionig, A., & Seer, S. (2015). What Makes You Bike? *Exploring Persuasive Strategies to Encourage Low-Energy Mobility*, 3, 53–64. doi:10.1007/978-3-319-20306-5
- Yeo, A. C., Rahim, M., & Ren, Y. Y. (2009). Use of Persuasive Technology to Change End- Users' IT Security Aware Behaviour: A Pilot Study. *International Journal of Behavioral, Cognitive, Educational and Psychological Sciences*, 1(1), 48–54.

ADDITIONAL READING

- Fogg, B. (1998). Persuasive computers: perspectives and research directions. In *SIGCHI conference on Human factors in computing* (Vol. 98, pp. 225–232). doi:10.1145/274644.274677
- Harjumaa, M., & Oinas-kukkonen, H. (2009). Understanding Persuasive Software Functionality in Practice : A Field Trial of Polar FT60. *Methodology*, 38, 1–9.

R Orji, RL Mandryk (2014) Developing culturally relevant design guidelines for encouraging healthy eating behavior in international Journal of Human-Computer Studies, 72 (2), 207-223

KEY TERMS AND DEFINITIONS

Captology: The study of computers as a persuasive technology.

Elaborate Message: A persuasive messages that are based on logics and they are relevant to the subject of discussion, that seek to alter a change in a receivers attitude.

Peripheral Message: Persuasive messages that uses cues and lacks relevant logic but however seeks to alter human behavior.

Persuadability: The ability of a persuasive technology, system or a behavior change support system to persuade a user to change his or her attitude and/or behavior.

Persuadee: Users of a persuasive technology of a behavior change support system that the designer intends to change.

Persuasive System Features: The features and functionality of a system that solely designed and implement to alter human cognition to a predetermined one.

Transition Description Cards: A template that is used to summarize the findings from a persuasive system design analysis. It aims at providing the designer with content relevant information for developing a persuasive system.

ENDNOTE

- ¹ For a detailed explanation of the processes in U-FADE refer to the cited article

Social Computing

Nolan Hemmatazad

University of Nebraska at Omaha, USA

BACKGROUND

Throughout the last few decades, computational technologies have grown increasingly more capable, useful, and connected at an exponential rate. However, while this general boon in computational power has occurred fairly recently, discussions relating to the ideas of interconnected computational systems and instantaneous, widespread information exchange began much earlier. As one example, we can look to the early efforts of Vannevar Bush, who helped facilitate cooperations between the United States government, business communities, and academicians for the advancement of military-centered scientific research initiatives. This formal cooperation would pave the way for later endeavors, such as the establishment of the Advanced Research Projects Agency (ARPA, and subsequently DARPA) and ARPANET, a precursor to the modern Internet.

The real fruits of early theoretical and engineering groundwork such as this, however, would become apparent on a much larger scale beginning in the 1980's and throughout the early 1990's, a timeframe that marks the development of several early communications technologies, including Usenet (a decentralized system of distributed discussions), Internet Relay Chat (IRC; a real-time, multiparty text communication system), and the World Wide Web (WWW), which would set a new standard for the electronic presentation and dissemination of text and media contents.

Though some of the fundamental characteristics of social computing had already been cemented even within these early technologies — real-time user content distribution, for example, was a

natural prerequisite of Internet-mediated chat — the Web and consumer technologies were still in their infancy. Technical limitations (such as a lack of bandwidth for widespread distribution of rich media, as well as limited processing power available to consumption devices), lack of user adoption, and delayed development of standards for how best to utilize new mediums for communication each constricted the advancement of more powerful social computing applications.

By the start of the new millennium, however, a movement known as “Web 2.0” was quickly gaining traction. The motivation behind this development was to acknowledge the evolving state of 1) consumer Web-enabled technologies, which boasted continuously increasing processing and display capabilities, 2) enhancements to the underlying network infrastructure which allowed for decreased latency and increased throughput of data transmissions across the internet, 3) more widespread adoption of Web technologies, and 4) increased user and developer activity surrounding collaborative and social technologies. As Fischer (2009) observed, this paradigm could be succinctly characterized by its objective of “fostering and supporting social production and mass engagement and collaboration”.

While the term (Web 2.0) itself may have been merely a label — considered little more than jargon even by Sir Tim Berners-Lee, the creator of the Web (Laningham, 2006) — the notions it represented provided the foundation for social computing as we know it today. Through the course of just over a decade, the Web had gone from the nascent realization of a technical dream, to a medium where users could stay constantly

connected via the exchange of text, images, audio, and video media, from the comfort of their home, or abroad with their mobile devices.

Today, social computing-related activities are among the most common uses of networked devices, and the evolution of mobile technologies has established social computing as an outstanding example of pervasive or ubiquitous computing technologies. In fact, recently the terms “pervasive social computing” (see, for example: Mokhtar and Capra, 2009) and “ubiquitous social computing” (for example: Motahari et al., 2007) have been coined to express exactly this dynamic, and to reflect the increasing prevalence of this relationship. Now more than ever, individuals are staying connected with one another via the use of social technologies, and new innovations are quickly being developed to help augment this pervasiveness and the facilitation of social interactions in new and imaginative ways.

With this understanding at hand, the definition of social computing employed within this chapter will be as follows: *The use of computational devices to facilitate or augment the social interactions and content sharing activities of their users, or to evaluate such interactions so as to obtain new information.* This aligns well with previous descriptions from the academic literature (see: Schuler, 1994; Charron et al., 2006; Parameswaran & Whinston, 2007a). Of note, however, is that this definition explicitly acknowledges the use of social computing as an analysis and prediction tool, does not impose limits upon the influence of institutions or providers over the social interactions of their users, and does not exclude anonymous or pseudonymous user interactions from its coverage. The definition is intentionally broad, so as to encompass social computing not only for what it is today, but for what it may become in the not-too-distant future.

PRACTICAL APPLICATIONS OF SOCIAL COMPUTING

The applications of social computing technologies are many and diverse in nature. While this is so,

perhaps the most prominent and widely used of these applications can be seen in general purpose, online social networks. These networks have been defined as “web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system” (Boyd and Ellison, 2007). Typically, such networks allow the exchange of short posts of information or other media contents to a timeline of activity visible to the public, their friends, or a more limited audience. Facebook, Twitter, and Google+ are each well-known examples of online social networks.

Web logs or *blogs* represent another application of social computing principles. Blogs allow users to create content that is typically of greater length than the concise updates that abound in social networks. These blog posts are often related to a specific interest or subject matter explicitly identified by the blog author. Though some blogging platforms allow users to interact on a limited basis — such as via comments, follower relationships, votes, or blogrolls (collections of links to related or otherwise notable blogs) — as a whole, these interactions tend to be much more localized, and their networked nature less explicit. A few well-known blogging platforms include Google’s Blogger, Yahoo’s Tumblr, LiveJournal, Medium, and WordPress.

Wikis and collaborative editing systems, such as Wikipedia, Basecamp, or Genius, allow users to create and edit articles, to-do lists, events, rich media such as images and videos, and a variety of other documents in shared contexts, and at least in some cases, in real-time. Though users are the driving force behind any collaborative effort, in these systems, the focus is on content and how users can help develop and curate those contents as an engaged, participatory community.

An array of rich media sharing platforms also exist which rely on user-contributed media to make up either a large portion or the entirety of their respective content bases. Some examples include

the video hosting and sharing services YouTube (owned by Google), Vimeo, Vine (from Twitter), and Justin.tv. Music and other auditory media can be shared through services like SoundCloud and MySpace. Image hosting and sharing services include Pinterest, Yahoo's Flickr, Google's Picasa, Imgur, 500px, Facebook's Instagram, and Snapchat (which has recently begun to expand into a more general online social network, through the deployment of new and broadened functionality).

Geolocation-based social services offer users the ability to share their location and "check-in" (an indication of one having visited a location of interest) activity, as well as to post comments or reviews, share pictures or other rich media, or to unlock rewards or other contents relating to a specific location. While many location-sharing services are built directly into or have been otherwise subsumed by larger social platforms, some yet exist as standalone ventures, such as Foursquare.

Link-sharing and social bookmarking services, which include, to various degrees, Reddit, Digg, Delicious, StumbleUpon, Slashdot, Hacker News, Fark, and others, offer as a primary feature the ability to share bookmarks or links to external content. These links may be associated with specific subjects, and are often able to be rated, saved, starred, flagged, or commented on by a larger community of viewers. In some instances, these platforms may allow text- or other direct media-based posts in addition to link submissions, though importantly, they often lack capabilities for networking among users, or place little emphasis on such functionality.

Social question answering platforms exist to help users locate relevant information based on questions they have or topics they want to learn more about. In these systems, other users with knowledge or expert experience relating to the subject matter in question can provide input to help steer the user along in their quest for information. Example providers within this domain include Quora, Stack Exchange, Ask MetaFilter, and Yahoo Answers.

Social shopping services span the gamut from online review portals such as Yelp, Epinions, and Angie's List, to deal sharing platforms like Slickdeals, Deals.Woot, FatWallet, and RetailMeNot, to actual user-driven marketplaces or online classifieds services, such as eBay, Etsy, and Craigslist, which distinguish themselves from traditional online shopping venues in part by allowing users to connect and market their goods or services to one another, directly.

Even outside of the range of services described above, social computing has had a dramatic impact. In some cases, whole other applications have been developed as a result of the growing popularity and reach of existing social platforms, such as reputation management or social analytics systems, including Netvibes and Klout; social media management systems like HootSuite and Twitter's TweetDeck; and community development or augmentation systems, including Disqus, Livefyre, and Ning, which allows customers to create their own social networks, tailored to a specific brand or interest.

Additionally, many specialized social services exist to address the needs of a specific niche or class of users, such as the widely used professional network, LinkedIn, enterprise social networks Yammer and Chatter (from Salesforce), or the social TV service, GetGlue, which provides "check-in" style functionality for television broadcasts. Still other services have been left outside of the scope of this review altogether, such as online podcasting and podcast sharing services, open source software collaboration systems, peer-to-peer (P2P) file sharing services, chat and video calling platforms, and more, if only for the sake of brevity.

It should be noted, too, that many sites that may not otherwise fit squarely into any of the categories outlined above have integrated a wide range of social computing features, such as comment systems, social voting or ranking systems, user networking functionality, social bookmarking or sharing links, customer reviews, or collaborative

filtering systems, such as Amazon's recommended product listings or Netflix's video recommendations system, which rely in part on the activities of other users to generate their product and video recommendations, respectively.

RESEARCH THEMES IN SOCIAL COMPUTING

Several existing works have already detailed many of the contemporary issues in social computing research (see, for example: Parameswaran & Whinston, 2007a; Parameswaran & Whinston, 2007b, Wang et al., 2007). Additional works have sought to classify the inherent themes of social computing research throughout the history of the field (e.g., Wang et al., 2014). In this section, we relay broad categories of contemporary research interest in social computing, offering specific examples of "hot spots" for researcher activity within each heading.

Engineering and Software Development

Engineering and software development comprise a crucial component of social computing, and although these research pursuits may seem more in alignment with the various reference disciplines of social computing, advances in social systems have often played a crucial role in shining a spotlight on engineering- and software development-related research interests.

Some examples of research issues within this theme include how best to represent networked relationships from an information storage and retrieval perspective; how to provide for the scalability needs of large, networked systems (including scaling data storage and retrieval facilities, information processing and session handling resources, content distribution, and so on); best practices and innovations in mining stream data or data at scale; load balancer design and optimization; optimal design and management of data

centers; how to implement and enforce access statutes relating to content visibility and shared resource ownership; how to detect anomalous user behaviors to prevent unauthorized access to user data, proprietary information, or intellectual property; the development of streamlined development processes and efficient team/project management capabilities; evaluation and selection of development methodologies; evaluations of feature testing and debugging practices; as well as various implementation details relating to specific social computing affordances, such as techniques for providing shared resource ownership, real-time collaboration, and granular privacy filtering mechanisms in shared contexts.

Social Psychology

Being inextricably intertwined with the social nature of human actors, social computing is also naturally concerned with issues stemming from its social psychology reference discipline, such as attitudes and change, social cognition, self and social identity, influence and reputation, interpersonal relationship development, and the wisdom of the crowd.

In addition, communications technologies have further highlighted more nascent areas of psychological interest, such as in determining the influence of mode of communication, sense of presence or social embedded-ness, and identity (including anonymity and pseudonymity) for a particular computer-mediated social interaction.

Researcher attention has also converged recently on evaluating the formation of online communities (for example, analyzing what conditions must exist to motivate the development of a community, or to procure involvement), as well as the long-term stability of online communities, and what factors play a part in maintaining the health and longevity of those communities. Participation itself is another dominant research interest in this context. For example, though it is widely accepted that the viability of a social application and its ability to establish communities

“depends on the willingness of the participants to share” (Nov and Wattal, 2009), motivations for participation may vary, and the same strategies for engaging one subset of users may not work for others. Some social studies also focus on investigating negative participant behavior, such as engaging in “trolling” (inciting), abusive, or criminal behaviors online. Studies in this area may examine the individual or collective propensity to offend, motivating factors for misconduct, or how best to moderate undesirable behaviors in online social communities.

Marketing and Design

Within the theme of marketing and design, social computing researchers may investigate the impact and motivators for virality (the rapid dissemination of contents or adoption of new services); consumer technology acceptance and adoption, and how this may influence the spread or use of a particular social platform across genders, age groups, locales, etc.; targeted marketing best practices; best practices for reputation management and brand development across social media platforms; user experience and human-computer interaction considerations, including best practices for optimizing for ease-of-use, readability, optimal click-through rates, and other factors across device types; how to maximize user engagement; and the use of design methods.

Social Computing and the Organization

As social computing has evolved and witnessed more mainstream adoption, it has become a significant technological cornerstone of the organization, as well. Research issues relating to organizations’ adoption and use of social computing systems include assessing social computing’s impact on organizational structure, the protection of intellectual property rights in systems of mass-scale distribution, the influence of social technologies on workplace productivity, organizational adoption of

social technologies, monitoring of social interactions, conflict resolution on social or collaborative systems, social media management, and others. Leonardi et al. (2013) provide a well-rounded account of social computing’s history, usage, and research issues in the enterprise context. R  th et al. (2012) have identified a number of obstacles to implementing social computing solutions within corporations, providing a list of recommendations or mediating actions to help ensure successful workplace adoption of such technologies.

Social Curation and Analysis

From an analytics perspective, large scale social computing platforms provide a wealth of potentially useful information made publicly available by their users. This has motivated research interests in natural language processing (NLP) and sentiment analysis to provide automated means of extracting meaning from user contents, estimating social capital (for example, influence or reputation metrics) for marketing purposes, knowledge mining, and data curation for separating the wheat from the chaff in automated tasks (such as identifying erroneous information or disambiguating subject matter in social question answering systems). In addition, personal analytics has recently emerged as a fledgling area of research interest in this area, seeking out means for quantifying personal data for evaluation or prediction tasks, as well as conducting automated knowledge extraction for the purpose of targeted marketing initiatives, catered content delivery, and other endeavors. Entirely new forms of social networking have even emerged around the idea of personal analytics, such as the web service, Gyroscope, which allows users to share monitored health and fitness metrics, app usage information, and more. The scale of these myriad forms of analysis, coupled with the immense availability of social data, has also spurred research interest in privacy and ethical considerations pertaining to social computing, in the same way that such concerns were previously expressed towards the

idea of ubiquitous computing (see, e.g., Bohn et al., 2005).

Clearly, the wide reach and diversity of social computing research issues offers this field as a noteworthy candidate for multidisciplinary or interdisciplinary studies, and one through which new knowledge discoveries may offer considerable reciprocal value to the various reference disciplines of the field.

FUTURE RESEARCH DIRECTIONS

While many of the emerging research themes in social computing have already been outlined in the previous section, there remain several areas of potential research interest that have, to this point, received comparatively little attention from the academic community. It is no mystery that we are moving toward a continuously more interconnected and mobile society. While there remain many challenges (whether technical, economic, cultural, or political in nature), researchers should look ahead to a time when these challenges are able to be resolved, or at least where they may loosen their restrictive hold on the progression of social computing technologies.

From a hardware perspective, we can expect continued advances in notable areas such as processing power and battery capacity. In addition, we can look forward to even greater coverage and throughput from wireless communications infrastructures, and fewer “dark spots” of wireless interconnectedness, even in regions that are as yet considered developing territories. In the course of this progression, it is important to consider the impact of such advancements in the present term, so that we may better prepare ourselves — as both the studiers and practitioners of social computing technologies — for the landscape we may soon be constituents of.

As mobile battery capacity continues to increase, it is not unlikely that the consumption of rich media (a typically battery-intensive activity)

will increase concomitantly. In addition, “always on” GPS-based location services will be much more feasible, along with more general wireless data transmissions, or sensor-based data collection activities. How might this new surge in location and data collection capabilities, as well as rich media consumption, alter the social computing landscape?

We’ve recently seen the development of many novel personal assistant technologies, such as Apple’s Siri, Google Now, Microsoft’s Cortana, and Evi. As systems such as these operate predominantly in the mobile application space, how will this new wealth of data augment their functionality? Will they become an increasingly prevalent force in our daily lives, perhaps going so far as to completely automate our scheduling and event planning activities, or managing social connections on our collective behalf? Perhaps remembering to purchase a birthday gift or to pick up flowers on an anniversary — for better or worse — will soon be but a memory, as these responsibilities are delegated to increasingly intelligent, socially-connected systems.

If we look at an even lower level, what new hardware devices might evolve as technologies with even the smallest of form factors develop more advanced capabilities? We’ve already witnessed several new developments in this area, such as smart watches or optical head-mounted displays (e.g., Google Glass), both of which boast at least moderately capable embedded computing platforms. When a device no larger than a marble offers more computing power than today’s high-end servers, coupled with sufficiently advanced battery life, what new applications and use cases for social computing technologies will unfold?

Lastly, in the more abstract sense, how will these things continue to alter how we operate as individuals and societies? This has certainly been a widely-discussed topic already, but when we look ahead at an increasingly global social footprint and sense of connectedness, will we witness only more of what we’ve seen already,

or will the impact change as social computing encompasses the daily lives of even more parts of the world and a more diverse array of cultures? Someday soon, it may well be the case that not a single person alive (barring the most extreme outliers) has existed in a time without advanced mobile communication devices. It could, in fact, be argued that we are still within a transitional time in the advancement of human capabilities. If so, what will this latest of “dark ages” look like through the eyes of our progeny, and how might they appear to us?

CONCLUSION

In this chapter, we’ve provided a concise and not overly technical overview of the field of social computing, so as to assist readers of various levels of expertise in becoming acquainted with the subject. In doing so, we have provided basic background information relating to the establishment of social computing as a field of academic and practitioner interest, defined social computing as it is today and may evolve in the future, provided an overview of a range of practical social computing applications, and outlined a number of research themes being actively investigated in the field.

Weiser (1999), in introducing his vision for 21st century computing, spoke of the idea of technology becoming so ingrained in everyday life that it seems to disappear from attention, entirely. In many ways, we find ourselves now entering the next stages of that evolution, where technology is no longer merely at home in our lives, but is indistinguishable from them, fundamentally reshaping our environments and decreasing our ability to demarcate how it is or is not influencing our collective behavior. As social computing continues to grow ever more ubiquitous, connected, and absorbed into even the most granular functions of our day-to-day existence, so too will it continue to take on new forms, and uncover new potential, owed to the efforts of practitioners and

academicians alike, who devote their energies to the upbringing of this young field.

REFERENCES

- Bohn, J., Coroamă, V., Langheinrich, M., Mattern, F., & Rohs, M. (2005). Social, Economic, and Ethical Implications of Ambient Intelligence and Ubiquitous Computing. In W. Weber, J. Rabaey, & E. Aarts (Eds.), *Ambient Intelligence* (pp. 5–29). Springer. doi:10.1007/3-540-27139-2_2
- Boyd, D. M., & Ellison, N. B. (2007). Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication*, 13(1), 210–230. doi:10.1111/j.1083-6101.2007.00393.x
- Charron, C., Favier, J., & Li, C. (2006, February). *Big Idea: Social Computing: How Networks Erode Institutional Power, And What to Do About It*. Forrester.
- Fischer, G. (2009). Cultures of Participation and Social Computing: Rethinking and Reinventing Learning and Education. 2009 Ninth IEEE International Conference on Advanced Learning Technologies (ICALT '09), 1-5. doi:10.1109/ICALT.2009.207
- Laningham, S. (Ed.). (2006, August 22). *developerWorks Interview: Tim Berners-Lee*. Retrieved from <http://www.ibm.com/developerworks/podcast/dwi/cm-int082206txt.html>
- Leonardi, P. M., Huysman, M., & Steinfield, C. (2013). Enterprise Social Media: Definition, History, and Prospects for the Study of Social Technologies in Organizations. *Journal of Computer-Mediated Communication*, 19(1), 1–19. doi:10.1111/jcc4.12029
- Mokhtar, S. B., & Capra, L. (2009). From Pervasive to Social Computing: Algorithms and Deployments. *ICPS '09 Proceedings of the 2009 International Conference on Pervasive Services*, 169-178.

Motahari, S., Manikopoulos, C., Hiltz, R., & Jones, Q. (2007). Seven Privacy Worries in Ubiquitous Social Computing. *SOUPS '07 Proceedings of the 3rd Symposium on Usable Privacy and Security*, 171-172.

Nov, O., & Wattal, S. (2009). Social Computing Privacy Concerns: Antecedents & Effects. *CHI '09 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 333-336.

Parameswaran, M., & Whinston, A. B. (2007). Research Issues in Social Computing. *Journal of the Association for Information Systems*, 8(6), 336-350.

Parameswaran, M., & Whinston, A. B. (2007). Social Computing: An Overview. *Communications of the Association for Information Systems*, 19, 762-780.

Räth, P., Urbach, N., Smolnik, S., & Butler, B. (2012). Corporate Adoption of Social Computing: A Process-Based Analysis. *Journal of Information Technology Case and Application Research*, 14(2), 3-27. doi:10.1080/15228053.2012.10845699

Schuler, D. (1994). Social Computing. *Communications of the ACM*, 37(1), 28-29. doi:10.1145/175222.175223

Wang, F.-Y., Zeng, D., Carley, K. M., & Mao, W. (2007). Social Computing: From Social Informatics to Social Intelligence. *IEEE Intelligent Systems*, 22(2), 79-83. doi:10.1109/MIS.2007.41

Wang, T., Liu, Z., Xiu, B., Mo, H., & Zhang, Q. (2014). Characterizing the Evolution of Social Computing Research. *IEEE Intelligent Systems*, 29(5), 48-56. doi:10.1109/MIS.2013.121

Weiser, M. (1991). The Computer for the 21st Century. *Scientific American*, 265(3), 94-104. doi:10.1038/scientificamerican0991-94

ADDITIONAL READING

Abelson, H., Ledeen, K., & Lewis, H. (2008). *Blown to Bits: Your Life, Liberty, and Happiness after the Digital Explosion*. Addison-Wesley Professional.

Bauerlein, M. (2011). *The Digital Divide: Arguments for and Against Facebook, Google, Texting, and the Age of Social Networking*. Tarcher.

Benkler, Y. (2007). *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. Yale University Press.

Botsman, R., & Rogers, R. (2010). *What's Mine Is Yours: The Rise of Collaborative Consumption*. Harper Business.

Brynjolfsson, E., & McAfee, A. (2014). *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. W. W. Norton & Company.

Eggers, W. D., & Macmillan, P. (2013). *The Solution Revolution: How Business, Government, and Social Enterprises Are Teaming Up to Solve Society's Toughest Problems*. Harvard Business Review Press.

Foth, M., & Forlano, L. (Eds.), Satchell, C. (Ed.), Gibbs, M. (Ed.), & Donath, J. (Afterword). (2011). *From Social Butterfly to Engaged Citizen: Urban Informatics, Social Media, Ubiquitous Computing, and Mobile Technology to Support Citizen Engagement*. The MIT Press.

Gansky, L. (2012). *The Mesh: Why the Future of Business Is Sharing* (Reprinted.). Portfolio Trade.

Gardner, H., & Davis, K. (2013). *The App Generation: How Today's Youth Navigate Identity, Intimacy, and Imagination in a Digital World*. Yale University Press.

Howe, J. (2009). *Crowdsourcing: Why the Power of the Crowd Is Driving the Future of Business*. Crown Business.

Jenkins, H., Ford, S., & Green, J. (2013). *Postmillennial Pop: Spreadable Media: Creating Value and Meaning in a Networked Culture*. New York University Press.

Kirkpatrick, D. (2011). *The Facebook Effect: The Inside Story of the Company That Is Connecting the World*. Simon & Schuster.

Lanier, J. (2011). *You Are Not a Gadget: A Manifesto* (Reprint ed.). Vintage.

Moe, W., & Schweidel, D. (2014). *Social Media Intelligence*. Cambridge University Press. doi:10.1017/CBO9781139381338

Nahon, K., & Hemsley, J. (2013). *Going Viral*. Polity.

Postman, N. (1993). *Technopoly: The Surrender of Culture to Technology*. Vintage.

Qualman, E. (2012). *Socialnomics: How Social Media Transforms the Way We Live and Do Business* (2nd ed.). Wiley.

Rainie, L., & Wellman, B. (2014). *Networked: The New Social Operating System* (Reprint ed.). The MIT Press.

Shirky, C. (2009). *Here Comes Everybody: The Power of Organizing Without Organizations* (Reprint ed.). Penguin Books.

Shirky, C. (2011). *Cognitive Surplus: How Technology Makes Consumers into Collaborators* (Reprint ed.). Penguin Books.

Standage, T. (2013). *Writing on the Wall: Social Media - The First 2,000 Years*. Bloomsbury USA.

Surowiecki, J. (2005). *The Wisdom of Crowds* (Reprint ed.). Anchor.

Tapscott, D., & Williams, A. D. (2010). *Wikinomics: How Mass Collaboration Changes Everything* (Expanded ed.). Portfolio Trade.

Thompson, C. (2013). *Smarter Than You Think: How Technology is Changing Our Minds for the Better*. The Penguin Press.

Turkle, S. (2012). *Alone Together: Why We Expect More from Technology and Less from Each Other*. Basic Books.

van Dijck, J. (2013). *The Culture of Connectivity: A Critical History of Social Media*. Oxford University Press. doi:10.1093/acprof:oso/9780199970773.001.0001

Vaynerchuk, G. (2011). *The Thank You Economy*. Harper Business.

KEY TERMS AND DEFINITIONS

Blog (Web Log): A medium for posting detailed text or rich media entries, often as part of a larger collection pertaining to some overarching subject or theme.

Check-In: A function for expressing a user's visiting of a physical location, or their participation in some event or occurrence, physical or otherwise.

Collaborative Filtering System: A system which evaluates the historic activity of its users in order to make predictions about the preferences of those users or others who are, by some criteria, viewed as being similar to them.

Peer-To-Peer: A connection existing directly from one user to another, with no intermediary.

Social Identity: How a user is perceived within a larger social community.

Social Network: A social computing application that allows users to establish friend or follower-type relationships with one another, and to exchange or broadcast text and rich media contents either directly or via aggregated timelines of activity.

Trolling: Incendiary behavior by users of a social system intending to incite outrage or otherwise evoke a negative response within a community.

Virality: The widespread propagation or adoption of new contents or services across an online community.

Ubiquitous Computing, Contactless Points, and Distributed Stores



Marco Savastano

Sapienza University of Rome, Italy

Eleonora Pantano

Middlesex University London, UK

Saverino Verteramo

University of Calabria, Italy

INTRODUCTION

Nowadays, retailing is subjected to constant changes due to the continuous advancements in technology (Pantano, 2014). In fact, the increasing computing capabilities, the mobile and wireless technologies improvements, as well as the development of flexible software architectures and automatic identification technologies support the ubiquitous access to data for both consumers and firms. As a consequence, these new systems modify the interface between clients and vendor, by providing new consumer-oriented services (Ngo and O’Cass, 2013; Pantano, 2014), as well as the customer behavior and corporate approach to retail process, by changing both the way customers access and consume the information, and the way firms and organizations reach clients and deliver the service (Kim, et al., 2009; Demirkan and Spohrer, 2014).

In this scenario, ubiquitous computing is supporting ubiquitous access to information for consumers and marketers, with emphasis on retail services (Pantano, 2013). It is based on the ubiquitous computing, which can be viewed as an extension of mobile computing characterized by portable accessing technologies (i.e. cameras, Location Based Service, Ubiquitous Sensor Network, etc.), always connected to a network, and linked to web-based multimedia content repositories that

adapt the provided contents according to users’ data (i.e. location, preferences, etc.) (Lin, et al., 2011). Therefore, these innovations are removing the boundaries of the physical space (Kourouthanassis et al., 2007; Bourlakis et al., 2009; Demirkan and Spohrer, 2014; Pantano, 2013; Pantano, 2014), while pushing retailers to redefine the traditional business model and practices. Similarly, a huge number of “contactless technologies” is emerging as the most promising direction for supporting retailing and shopping experience, by providing new modalities for automatic payment and self-checkout (Lai and Chuah, 2010). These are based on proximity sensor that allows payment without entering any pin. For these reasons, the number of contactless transactions is speeding up fast the payment process, increasing the convenience of shopping experience and reducing the queues. In fact, many retailers are soliciting users to adopt this system, such as McDonalds that offers a free drink on any menu purchased via contactless. For instance, MasterCard reports that the 10% of transactions made in Australia under \$100 is performed via contactless technologies, while VISA Europe reports more than 1 million of contactless purchases in August 2013 (VISA Europe, 2013).

The aim of this chapter is to provide a comprehensive view of the new retail competitive scenario, starting from the most innovative technologies in retail domain. Subsequently, it sheds

light on the possible consequences of these radical innovations in terms of store structure, and provides some indications for predicting the trends of the future (technology-enriched) shopping places.

BACKGROUND

As anticipated, the continuous progresses in information and communication technologies (ICT) impact on retail strategy and operations triggering the proliferation of new channels and new modalities through which customers may directly interact with firms. Due to the large amount of technological innovations, the necessity to classify them emerges as the starting point for a deep understanding of the current scenario. In this chapter, we propose a classification based on the functions of these systems for retail purposes as follows: (i) technologies for virtually trying the products, (ii) technologies for automatic product search, (iii) technologies for automatic payments, and (iv) full-service technologies (technologies integrating all these services).

Technologies for Virtually Trying Products

Current advances in 3D graphics and virtual reality tools offer novel and realistic interfaces that can be easily integrated in physical environments for enhancing the sensorial inputs and enriching consumers shopping experience (Bullinger et al., 2010). Augmented reality emerges from the integration of virtual reality, with emphasis on the most advanced graphics, in real contexts by enriching the human sensory perception with multimedia contents and information to the reality. (Papagiannidis et al, 2014). Hence, it includes a real-time view of the physical world enhanced (augmented) with virtual computer generated information, such as digital images or video stream, etc. (Azuma, 1997; Carmigniani et al., 2011).

The integration in retailing offers virtual mirrors, virtual fitting rooms and interacting shopping windows, which customers might experience by virtually trying the products through digital interfaces under the promise to save time while living a more exciting shopping experience (Dey and Sandor, 2014). In this direction, one of the most promising areas of research is the virtual garment try-on experience (or virtual fitting), which allow consumers to try clothes and evaluating the effect without really wearing them (Chen et al., 2011; Pereira et al., 2011; Wang et al., 2012).

Technologies for Automatic Product Search

Technologies included in this category support the search of a certain product throughout the store. They are especially exploited in the large department stores for helping consumers to easily find the favorite item (in terms of exact location including floor, area, etc.) without the direct assistance of a salesperson (Pantano and Timmermans, 2014).

RFID (Radio Frequency IDentification) is one of the first examples of contactless technology, consisting of an enriched version of the barcode. It is the most common technology used to identify and localize objects from a distance (Garfinkel and Rosenberg, 2006). In particular, it consists of a radio frequency transceiver (transmitter and receiver) attached to the item to be identified by a reader device (interrogator device) used to read the information stored on the tag. RFID tags are also used to improve the in-store security, to enhance the rapid checkout of consumers, and to manage the shelf-life of products (i.e. by evaluating in real-time the quantity of a certain product on the shelf).

QR Codes technology has become one of the most-used types of two-dimensional barcode. Developed in the automotive contest, it is currently used in a broader context, including both tracking for logistics purposes and convenience-oriented

applications aimed at enriching consumers shopping experience through smartphones.

Technologies for Automatic Payments

Contactless technologies support the spreading of mobile payments, by offering the shift from the traditional magnetic stripe or “Chip and Pin” to contactless cards, where tag brought into close proximity substitutes the card sweeping or insertion into the reader (and the subsequent request of PIN or authorization for the payment), with benefits for the financial transactions in terms of time saving and ease of process, and the consequent reduction of the queues. For these reasons, contactless payment is particularly important to retail settings where speed and convenience of payment are the basic elements of the success, such as fast food and kiosks. For instance, in the US market, contactless payment is an established method especially in the fast food, leisure and sport industries (Olsen, 2008).

The implementation of these technologies creates benefits for both customers and retailers. Consequently, innovative ways for transactions at the point-of-sale are proliferating and mobile payments are one of the most investigated technologies in retail industry from a both scholars and practitioners perspective (Pantano and Viasone, 2015).

The continuous research in wireless technologies provides further systems such as NFC (Near Field Communication) and beacons, which are driving mobile phones to facilitate the access execution of several monetary transactions and are successfully employed for automatic mobile payments (Olsen, 2008). In particular, beacons consist of small and low cost transmitting devices, which do not require a limited distance between the devices to set the transaction.

These devices might share data via Bluetooth connections, such as specific deals or product information, while tracking their location/route within the store (in-store location). For instance,

after downloading the mobile app, consumers may receive customized deals or special offers when they are within 50 meters of a specific shelf or mannequin. From a corporate perspective, retailers would obtain information about consumers behaviour such as time spent in a certain area of the stores, accessed information, consumers profile, etc...

Full-Service Technologies

The last group of technologies includes those that support the delivering of more services, including both product search and payment. Ubiquitous stores are a meaningful example of these technologies offering a “full-service” for consumers (Pantano, 2013). The term ubiquitous computing was created to describe the ubiquity of information technology and computer power that pervade almost all everyday objects (Pantano, 2013).

Communication technologies, and particularly mobile applications, can be considered as the fundamentals of the ubiquitous computing environment. From the systematic integration of the above mentioned technologies, the features of this new scenario arise as follows: (i) embedding of computer hardware and software in other equipment and objects; (ii) providing constant network accessibility and connectivity by user’s mobile device (e.g. via wireless, Bluetooth, etc.); (iii) supporting users through information services anytime and anywhere (accessing to the requested information via embedded input devices such as GPS or camera); (iv) context awareness and adaptation of the system to current information requirements (i.e. the system replies according to consumers’ location); (v) automatic recognizing and autonomous processing of repetitive tasks without user intervention (Kurkovski, 2005; Gabriel et al., 2006; Lin et al., 2011).

In a ubiquitous environment users are connected anytime and anywhere, automatically sharing information and receiving services or other contents related to the currently experienced context. Indeed one of the goals of such environ-

ment is to capture, use and manage a huge amount of information, on which companies can built up their competitive advantage with accessible, reliable, and relevant information (Turner, 1978).

Ubiquitous computing improves considerably companies' access to information by allowing them to acquire updated data at anytime and anywhere from their customers, by changing the interaction between businesses and customers in both time and space. Furthermore, information captured from the multiple interactions with customers can create value and increase transactions through the enhanced possibility to provide customized service and sell complementary items as well as be exploited for the development of more effective marketing strategies. Such information can also reduce the costs of subsequent transactions through more efficient communications programs and delivery methods (Kim et al., 2009). For these reasons, it can be successfully applied at the context of retailing, creating a new environment for selling and purchasing: ubiquitous retailing (Pantano, 2013). Within these settings, the products displayed appear as virtual full-size reconstructions, associated with a QR code, that consumers may scan by focusing a mobile camera in order to visualize the relevant information and have the possibility to collect them in their virtual baskets. After the effective order and mobile payment, purchases are delivered directly at home. Recent applications of "u-stores" can be found particularly in grocery industry, placed inside bus/tube stops for shopping daily food and beverage (Pantano, 2013). Thus, even perishable goods can be sold without a need for refrigeration within the store. Examples are Sorli Discau and Tesco "HomePlus" virtual stores, located respectively in Barcelona and Seoul, with the goal to turn waiting time into shopping time in consumers daily life.

MAIN FOCUS OF THE ARTICLE

The technologies presented in the previous section largely impact the traditional retailing activities.

While a problem of the traditional retail settings relies on the physical boundaries of the store (i.e. fixed and limited space), the new technologies allow consumers to buy anytime anywhere, without any need of further assistance. Hence, their main limit relies on the technological boundaries, due to the connection availability and the technological adequacy of the device used by the customer.

Therefore, ubiquitous computing and contactless technologies offer the possibility to totally reconsider the physical boundaries of the traditional stores for moving towards a new concept of store.

In fact, if a traditional point of sale is limited to the store location, spatial dimensions and opening hours, the new stores are not related to a specific location but *distributed*, in terms of access anytime and anywhere within an area enriched with the above mentioned technologies. As a consequence, these new stores are based on the higher level of technology integration and accessibility to products and related information for consumers.

Starting from Diamond's (2010) idea that retailer should be always ready to satisfy consumers' expectations, the traditional store concept breaks down as these technologies allow vendors to satisfy consumer's need as soon as it emerges through the high internet connectivity. For instance, if a consumer shows the need to buy a new bag while waiting for entering a theatre, he/she is able to find, compare and buy the favored items that retailer will delivery within an acceptable amount of time. Moreover, retailers might propose consumers new deals and offers while they are in queue for entering a certain exhibition. Likewise, a customized offer can be accessed by the in-car Internet connection, matching the user exact position with his/her product preferences (i.e. based on past purchases), personal profile (e.g. age, gender, education, address, etc.) and behaviour (car routes and destinations, etc.). This prompts consumer to be able to buy and retailer to sell continuously.

As a measure of the consequences of the new technology-based store, we could consider the spatial dimension, which synthetizes the changes

in physical surface triggered by the adoption of the advanced technologies and the idea of *distributed store* (in other words it represents the space that the innovation requires within the store for the correct technology access/functioning).

As emerged from previous sections, we identified few categories of technological innovations that are changing retail process. These emerging technologies tend to create a new store concept based on a mobile access to store, with consequences on the drastic reduction (or even *absence*) of the physical store floor (see ubiquitous computing), which might be accessed directly from consumer's own smart phone, without the need of a physical support tool in the point of sale. In fact, the ubiquitous store overcomes the physical surface of the traditional store, which is replaced by urban areas such tube stops, representing the current innovation frontier for retailing.

Therefore, the progresses in technologies provide new tools for moving traditional fixed stores towards a mobile and ubiquitous dimension, while the relevance of the technology in the store increases.

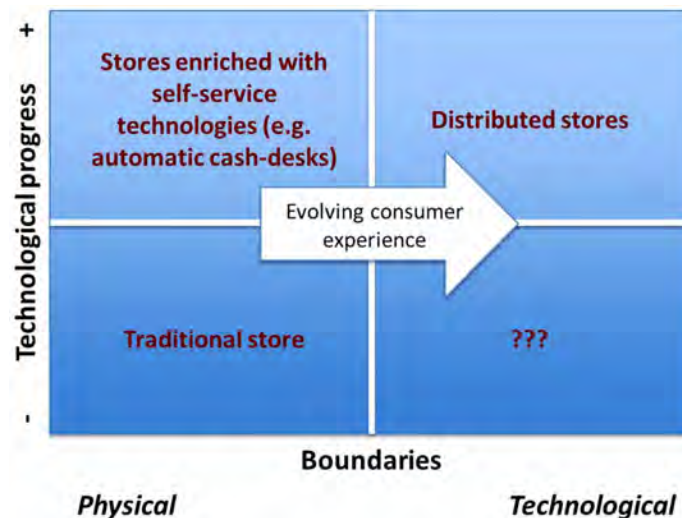
As anticipated, contactless technologies, RFID systems, and AR applications have been introduced as an added tool with a limited relevance into the physical store. In opposite, the stores built on ubiquitous and mobile computing do not allocate

any physical space in the “traditional store” (even if some stores can advise consumers about the available apps with posters, etc.), which might even take place within urban areas. This emerging store, the distributed one, is built on a software virtually reproducing the store, by presenting a new kind of point of sale that radically redefines the space and time of retailing, and modifies the role of the actors involved in the process (Evans and Hu, 2007; Kourouthanassis et al., 2007; Bourlakis et al., 2009; Walter et al., 2012; Pantano and Viassone, 2015).

SOLUTIONS AND RECOMMENDATIONS

To provide a comprehensive view of the existing retail scenario in order to better predict the future evolution, it is possible to summarize the possible retail scenario on the basis of technology progress/usage and related boundaries, by considering the actual shift from the physical boundaries to the technological ones. To this end, Figure 1 synthesizes the emerging retail landscape, where traditional stores are characterized by both a low level of technological integration and physical boundaries (i.e. the store in a fixed location and fixed opening hours) (down left quadrant).

Figure 1. From traditional to distributed stores



The subsequent introduction of technologies pushes towards a new enriched store, where the technology acts as support tool for consumers by delivering enriched services (i.e. stores offering interactive displays for accessing information, or automatic cash desks for self-checkout), thus it is still related to a fixed store location and fixed opening hours (up-left quadrant). The second quadrant of the matrix (up-right) represents a scenario characterized by a high access to the technology, where the extended usage of ubiquitous computing and contactless technologies allow consumers to purchase anytime and anywhere in a sort of “distributed store”. Thus, the main features of this new concept of store include (i) accessibility anytime (24/7) and anywhere within specific technology-enriched areas where Wi-Fi internet connection is available; (ii) enhanced and customized consumer services (these stores offer a larger service portfolio if compared to the traditional ones); and (iii) interactive and customizable products displaying (i.e. consumers might visualize only the favored item, rotate the product and change perspective, etc.).

In the one hand, the new retail settings enlarge the traditional boundaries of the fixed points of sale; whereas in the other one, they are limited by the availability and performance of these technologies, which characterize the new store boundaries.

As shown in the last quadrant of the matrix, due to the dichotomy of “no technological progress” and “technological boundaries”, a scenario characterized by a limited technology progress and technological barriers is not actually conceivable.

FUTURE RESEARCH DIRECTIONS

The development of an effective distributed retailing scenario is both a short and long-term program requiring well thought-out roadmaps, which retailers should start exploiting by deeply understanding their capacity to absorb technological innovation in terms of new business models, reconfiguration of internal resources (both financial and human

ones) and channel strategies. They should also deeply re-design the consumer experience taking place out of the traditional point of sale. Hence, new studies should focus on these elements, by providing new frameworks for understanding the phenomenon and new measures for evaluating the best strategy to adopt. Since this study is explorative in nature, it would be beneficial to complement it with further investigations concerning the shopping experience emerging from a technology-enriched retail environment, in order to identify the new dynamics of consumer behaviour and the evolution of the shopping habits.

Therefore, a set of new questions to be addressed emerges for the definition of more efficient marketing and retail strategies:

- To what extent will the technological progress for retailing be feasible and sustainable?
- How can retailers identify the best scenario for competing and the right moment for innovating?
- What would the consequences of a technology failure be for the retail process?

CONCLUSION

This chapter shows the extent to which the recent advances in technology provide an increasing integration of computing in consumers shopping activities through a pervasive penetration, starting from Evans and Hu (2006) reflections on the spread of pervasive and ubiquitous systems. These technologies represent the further steps of traditional point of sale development into an innovative shopping environment.

Past studies anticipated the trend towards ubiquitous stores based on a high level of connectivity and extensive usage of mobile devices (Kourouthanassis et al., 2007; Bourlakis et al., 2009; Pantano, 2013; Blazquez, 2014; Pantano, 2014). Our chapter extends these studies by exploiting the concept of distributed store, which

overcomes the conventional physical boundaries in a sort of evolution of the traditional stores. In fact, within these new retail settings, boundaries are no longer physical or temporal but technological: the implementation and success of distributed stores is strictly connected to the access and integration of the advanced technology. This evolution provides new benefits for customers as well as opportunities for practitioners. In this way, consumer experience and store concept are redefined, where the shopping experience begins before the consumer enters the traditional store and continues after the purchase, accompanying him/her during the various activities of daily life.

REFERENCES

- Azuma, R. T. (1997). A survey of augmented reality. *Presence (Cambridge, Mass.)*, 6(4), 355–385. doi:10.1162/pres.1997.6.4.355
- Blázquez, M. (2014). Fashion Shopping in Multichannel Retail: The role of Technology in enhancing the Customer Experience. *International Journal of Electronic Commerce*, 18(4), 97–116. doi:10.2753/JEC1086-4415180404
- Bourlakis, M., Papagiannidis, S., & Li, F. (2009). Retail spatial evolution: Paving the way from traditional to metaverse retailing. *Electronic Commerce Research*, 9(1-2), 135–148. doi:10.1007/s10660-009-9030-8
- Bullinger, H.-J., Bauer, W., Wenzel, G., & Blach, R. (2010). Towards user centred design (UCD) in architecture based on immersive virtual environments. *Computers in Industry*, 61(4), 372–379. doi:10.1016/j.compind.2009.12.003
- Carmigiani, J., Furht, B., Anisetti, M., Ceravolo, P., Damiani, E., & Ivkovic, M. (2011). Augmented reality technologies, systems and applications. *Multimedia Tools and Applications*, 51(1), 341–377. doi:10.1007/s11042-010-0660-6
- Chen, A., Kao, C.-Y., Chen, Y.-H., & Wang, W.-C. (2011). Applying augmented reality to consumer garment try-on experience. *Lecture Notes in Computer Science*, 6750, 169–190. doi:10.1007/978-3-642-22619-9_10
- Demirkan, H., & Spohrer, J. (2014). Developing a framework to improve virtual shopping in digital malls with intelligent self-service systems. *Journal of Retailing and Consumer Services*, 21(5), 860–868. doi:10.1016/j.jretconser.2014.02.012
- Dey, A., & Sandor, C. (2014). Lessons Learned: Evaluating Visualizations for Occluded Objects in Handheld Augmented Reality. *International Journal of Human-Computer Studies*, 72(10-11), 704–716. doi:10.1016/j.ijhcs.2014.04.001
- Diamond, D. (2010). Ubiquitous Retail. If being everywhere works for Verizon and Starbucks, why not for supermarket operators. *Progressive Grocer*, 89(8), 10.
- Evans, C., & Hu, B. (2006). E-commerce to U-business: a model for ubiquitous shopping mall. In *Proceedings of the 1st international Symposium on Pervasive Computing and Applications*. IEEE. doi:10.1109/SPCA.2006.297612
- Gabriel, P., Bovenschulte, M., Hartmann, E., Groß, W., Strese, H., Bayarou, K. M., & Oberweis, R. et al. (2006). *Pervasive Computing: Trends and Impacts*. Ingelheim, Germany: SecuMedia.
- Garfinkel, S., & Rosenberg, B. (2006). *RFID – Applications, Security, and Privacy*. Upper Saddle River, NJ: Addison-Wesley.
- Kim, C., Oh, E., Shin, N., & Chae, M. (2009). An empirical investigation of factors affecting ubiquitous computing use and U-business value. *International Journal of Information Management*, 29(6), 436–448. doi:10.1016/j.ijinfomgt.2009.06.003
- Kourouthanassis, P. E., Giaglis, G. M., & Vrechopoulos, A. P. (2007). Enhancing user experience through pervasive information systems: The case of pervasive retailing. *International Journal of Information Management*, 27(5), 319–335. doi:10.1016/j.ijinfomgt.2007.04.005

- Kurkovski, S. (2005). Using principles of pervasive computing to design m-commerce applications. In *Proceedings of the International Conference on Information Technology: Coding and Computing* (pp. 59-64). IEEE. doi:10.1109/ITCC.2005.287
- Lai, P. M., & Chuah, K. B. (2010). Developing an Analytical Framework for Mobile Payments Adoption in Retailing: A Supply-Side Perspective. In *Proceedings of International Conference on Management of e-Commerce and e-Government* (pp.356-361). IEEE. doi:10.1109/ICMeCG.2010.79
- Lin, P.-H., Huang, K.-S., Wu, C.-H., & Hong, Z.-W. (2011). A resource management model for ubiquitous computing environments. In *Proceedings of the 15th International Symposium on Consumer Electronics* (pp. 556-559). IEEE. doi:10.1109/ISCE.2011.5973892
- Ngo, L. V., & OCass, A. (2013). Innovation and business success: The mediating role of customer participation. *Journal of Business Research*, 66(8), 1134–1142. doi:10.1016/j.jbusres.2012.03.009
- Olsen, C. (2008). Is contactless payment a reality for the retail industry? *Card Technology Today*, 20(11-12), 10–11. doi:10.1016/S0965-2590(08)70232-9
- Pantano, E. (2013). Ubiquitous retailing Innovative Scenario: From the fixed point of sale to the flexible ubiquitous store. *Journal of Technology Management & Innovation*, 8(2), 84–92. doi:10.4067/S0718-27242013000200007
- Pantano, E. (2014). Innovation drivers in retail industry. *Journal of International Management*, 34, 344–350.
- Pantano, E., & Timmermans, H. (2014). What is smart for retailing? *Procedia Environmental Sciences*, 22, 101–107. doi:10.1016/j.proenv.2014.11.010
- Pantano, E., & Viassone, M. (2015). Engaging consumers on new integrated multichannel retail environments: Challenges for retailers. *Journal of Retailing and Consumer Services*, 25, 106–114. doi:10.1016/j.jretconser.2015.04.003
- Papagiannidis, S., See-to, E., & Bourlakis, M. (2014). Virtual test-driving: The impact of simulated products on purchase intention. *Journal of Retailing and Consumer Services*, 21(5), 877–887. doi:10.1016/j.jretconser.2014.02.010
- Pereira, F., Silva, C., & Alves, M. (2011). Virtual fitting room augmented reality techniques for e-commerce. *Communications in Computer and Information Science*, 220, 62–71. doi:10.1007/978-3-642-24355-4_7
- Piotrowicz, W., & Cuthbertson, R. (2014). Introduction to the Special Issue Information Technology in Retail: Toward Omnichannel Retailing. *International Journal of Electronic Commerce*, 18(4), 5–15. doi:10.2753/JEC1086-4415180400
- Turner, B. T. (1978). Senior Clayton fellowship final report: Information engineering design work. London, UK: The Institution of Mechanical Engineers.
- Visa Europe. (2013). *Boots UK goes contactless*. Retrieved April, 27, 2015, from <http://www.visaeurope.com/newsroom/news/boots-uk-goes-contactless>
- Wang, W.-C., Chen, Y.-H., & Kao, C.-Y. (2012). The digital design applied to consumer garment try-on experience integrated with augmented reality. *Lecture Notes in Electrical Engineering*, 124(1), 115–122. doi:10.1007/978-3-642-25781-0_18

ADDITIONAL READING

- Alexander, A. (2008). Format development and retail change: Supermarket retailing and the London Co-operative Society. *Business History*, 50(4), 489–508. doi:10.1080/00076790802106679

Friedewald, M., & Raabe, O. (2011). Ubiquitous computing: An overview of technology impacts. *Telematics and Informatics*, 28(2), 55–65. doi:10.1016/j.tele.2010.09.001

Hopping, D. (2000). Technology in retail. *Technology in Society*, 22(1), 63–74. doi:10.1016/S0160-791X(99)00042-1

Huth, C. L. (2015). A Privacy Primer on Beacon Technology. *Journal of Internet Law*, 18(8), 21–25.

Kang, J.-Y. M., Mun, J. M., & Johnson, K. K. P. (2015). In-store mobile usage: Downloading and usage intention toward mobile location-based retail apps. *Computers in Human Behavior*, 46, 210–217. doi:10.1016/j.chb.2015.01.012

Mitchell, I. (2010). Innovation in non-food retailing in the early nineteenth century: The curious case of the bazaar. *Business History*, 52(6), 875–891. doi:10.1080/00076791.2010.499427

Neslin, S. A., & Shankar, V. (2009). Key issues in multichannel customer management: Current knowledge and future directions. *Journal of Interactive Marketing*, 23(1), 70–81. doi:10.1016/j.intmar.2008.10.005

Pantano, E., & Migliano, G. (2015). Improving the quality of technology-based innovations selection: A Quality Function Deployment approach for retailers. *International Journal of Business Performance Management*, 16(2/3), 352–372. doi:10.1504/IJBPM.2015.068722

Papagiannidis, S., Pantano, E., See-to, E., & Bourlakis, M. (2013). Modelling the determinants of a simulated experience in a virtual retail store and users product purchasing intentions. *Journal of Marketing Management*, 29(13/14), 1462–1492. doi:10.1080/0267257X.2013.821150

Sorescu, A., Frambach, R. T., Singh, J., Rangaswamy, A., & Bridges, C. (2011). Innovations in retail business models. *Journal of Retailing*, 87(S1), s3–s16. doi:10.1016/j.jretai.2011.04.005

KEY TERMS AND DEFINITIONS

Beacon: Small and low cost transmitting device, which do not require a limited distance between the devices to set the transaction. This device might share data via Bluetooth connections, such as specific deals or product information, while tracking consumers location/route within the store (in-store location).

Contactless Technologies: New technologies based on proximity sensor that allows payment without providing any pin or signature.

Distributed Store: New stores not related to a specific location but distributed, in terms of access anytime and anywhere within an area enriched with contactless and ubiquitous technologies.

Near Field Communication (NFC): Technology providing mobile devices with wireless two-way short-range connectivity up to a maximum of 10 cm.

Quick Response Codes (QR): Bidimensional barcode able to augment the reality with contents and information by scanning it with a smartphone camera.

Radio Frequency Identification (RFID): One of the first examples of contactless technology, consisting of a modern version of the barcode. It is based on a radio frequency transceiver (transmitter and receiver) attached to the item to be identified by a reader device (interrogator device) used to read the information stored on the tag.

Ubiquitous Computing: A sort of extension of mobile computing based on the portable accessing technologies (i.e. cameras, Location Based Service, Ubiquitous Sensor Network, etc.), always connected to a network, and linked to web-based multimedia content repositories that adapt the provided contents according to users' characteristics (i.e. location) (Lin, et al., 2011).

Category U

**Urban and Regional
Development**

Climate Change as a Driving Force on Urban Energy Consumption Patterns



Mostafa Jafari

Agricultural Research Education and Extension Organization (AREEO), Iran

Pete Smith

University of Aberdeen, UK

INTRODUCTION

Climate change has impacted on Iranian natural ecosystems and urban area in various ways (Jafari, 2010). Climatic factors, including temperature, precipitation and humidity have changed in pattern in recent decades (Jafari, 2011). Changes in temperature and precipitation patterns could have impacts on urban areas as well as forests, rangelands and desert ecosystems (Jafari, 2008a). Changing climate patterns and increasing pollution may lead to changed production patterns (Jafari, 2012a) or may increase pressure on the environment (Jafari, 2012b). Environmental sustainability among two others, Energy security and Energy equity are the world energy trilemma (Wyman, 2013). Attempts to mitigate climate change need to be done without compromising food security or environmental goals (Smith *et al.*, 2013).

In this paper, we present a case study, from Rasht City in Iran, to show how changing climate is expected to have influenced energy consumption patterns. We use climatic data to determine the number of days when heating and cooling demands occurs, using Heating Degree Days (HDDs) and Cooling Degree Days (CDDs). These are based on daily temperature observations, with each month having at least 25 records and no less than 15 years of data (Anonymous, 2008a). HDD and CDD, which indicate the level of comfort, are based on the average daily temperature which is taken as

the mean of maximum and minimum daily temperature (the National Oceanic and Atmospheric Administration – US NOAA).

If the average daily temperature falls below comfort levels, heating is required and if it is above comfort levels, cooling is required. HDD is an index of the energy demand to heat buildings, and an analogous index for the energy demand for cooling is represented by cooling degree days (Sivak, 2013). The HDDs or CDDs are determined by the difference between the average daily temperature and the BASE (comfort level) temperature. The BASE values used are 12 and 18 degrees Celsius for heating and 18 and 24 degrees Celsius for cooling (Anonymous, 2008a). In this case, base degrees for heating are 18°C and for cooling is 21°C. For example, if heating is being considered to a temperature BASE of 18 degrees, and the average daily temperature for a particular location was 14 degrees, then heating equivalent to 4 degrees or 4 HDDs would be required to maintain a temperature of 18 degrees for that day. However if the average daily temperature was 20 degrees then no heating would be required, so the number of HDDs for that day would be zero. If cooling is being considered to a temperature BASE of 21 degrees, and if the average temperature for a day was 27 degrees, then cooling equivalent to 6 degrees or 6 CDDs would be required to maintain a temperature of 21 degrees for that day. However if the average temperature was 20

degrees, then no cooling would be required, so the number of CDDs for that day would be zero. Similar estimates have been made in the USA, mainly using the Fahrenheit temperature scale (Anonymous, 2008c; Anonymous, 2008d). Costs are calculated by multiplying the HDD or CDD by the average daily cost of heating or cooling (Anonymous, 2008e).

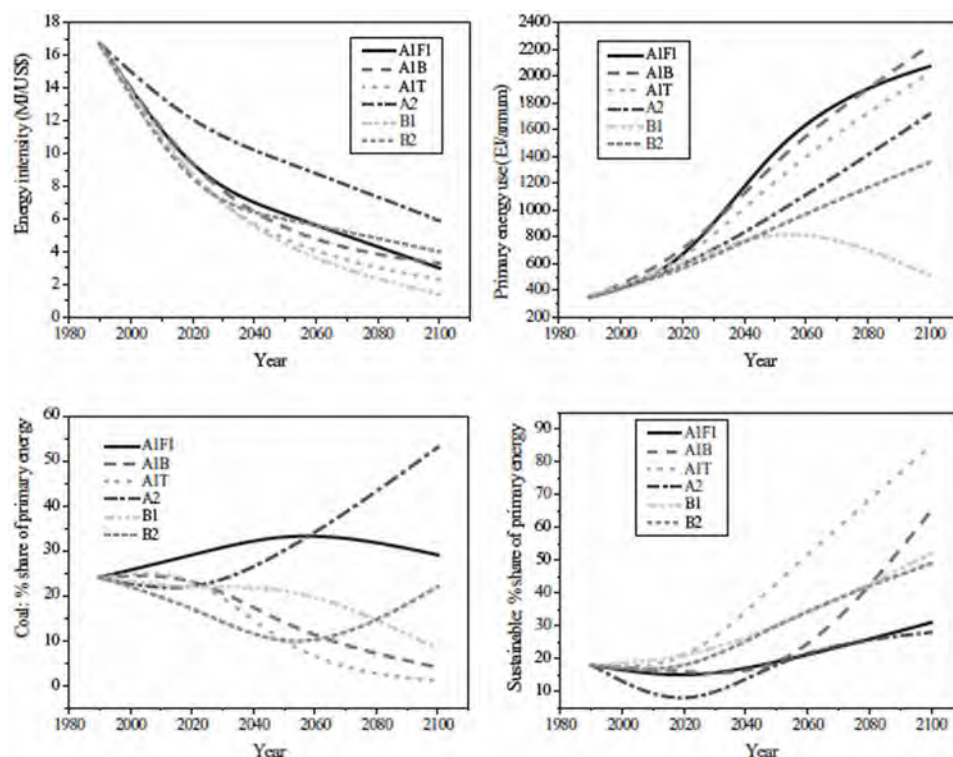
HDD can be added over periods of time to provide a rough estimate of seasonal heating requirements. In the course of a heating season, for example, the number of HDD for New York City is 5,050 whereas that for Barrow, Alaska is 19,990. Thus, one can say that, for a given home of similar structure and insulation, around four times the energy would be required to heat the home in Barrow than in New York. Likewise, a similar home in Los Angeles, California, where heating degree days for the heating season are

2,020, it would require around two fifths the energy required to heat the house in New York City (Anonymous, 2012).

The following figures (Figure 1) show some of the energy-based driving forces within the scenarios that are particularly important to us. The world faces a range of futures, with the possibility that energy intensity (efficiency) might improve by a factor of eight by 2100: primary energy use rise initially under all scenarios, then decrease slightly or increase by a factor of 3.3: coal use may increase slightly, or almost disappear completely: and alternative, non-carbon, energy sources (including non-commercial) may halve in importance or become the norm. The question is which future will we choose? (Coley, 2008).

Our civilization and our standard of living depend on an adequate supply of energy. We need energy to light and heat our homes, to cook

Figure 1. Energy-based driving forces of SRES illustrative scenarios, A1F1, A1B, A1T, A2, B1, B2: 1980-2100 [data from IPCC 00]; energy intensity (MJ/US\$): top left; primary energy use (EJ/ annum): top right; coal: % share of primary energy: down left; sustainable: % share of primary energy: down right
Source: adapted from data in Emission scenarios: summary for policy makers, IPCC, 2000, www.ipcc.ch/pub/synergy.htm



our food, to drive our transport and power our communications and to provide the motive force that drives the factories. Without energy all this would be impossible on the scale needed, and our civilization would soon collapse. Our dependence on energy is strikingly illustrated by the connection between average life expectancy and energy consumption (Hodgson, 2010). Climate is one of the determining features of civilization, so any change in the climate can have momentous consequences (Hodgson, 2010). The importance of climate for civilization is further discussed by Hodgson (2010).

BACKGROUND

An attempt has made to evaluate climate change impact on different sectors including natural ecosystem and urban areas. Outcomes obtain from research project in urban zone which is near to forest area may help to recognized its impact on natural ecosystem in one way or another (Jafari, 2013). According to the research outcome, consideration and analysis on Energy Consumption Pattern and GDP in global level in different developed and developing countries it has been proved that, GDP and Energy Consumption of developing countries are increasing exponentially, while GDP and Energy Consumption of developed countries are increasing linearly (Akhter Hossain, 2012).

ENERGY CONSUMPTION PATTERN

Energy consumption patterns have changed in urban environments during the last few decades. Climate and environmental changes can be considered as one of the main driving forces for the changing pattern. Climate change has impacted on Iranian natural ecosystems and urban area in various ways. Climatic factors, including temperature, precipitation and humidity have changed in

pattern in recent decades. Changes in temperature and precipitation patterns could have impacts on urban areas as well as forests, rangelands and desert ecosystems. Changing climate patterns and increasing pollution may lead to changed production patterns or may increase pressure on the environment. Environmental sustainability among two others, Energy security and Energy equity are the world energy trilemma. Attempts to mitigate climate change need to be done without compromising food security or environmental goals. This research investigated the phenomenon in Rasht city, Iran. HDD, in cases where temperatures are below 18°C, and CDD, in cases where temperatures are above 21°C, were used as energy consumption indices. During the last half century, mean annual temperatures have increased and as a consequence, CDD in the warm season have increased sharply. In the same time slice, HDD, even in the cool and cold season have declined steadily. This present work shows that energy consumption patterns have increased sharply, and with available projection scenarios, is projected to increase more rapidly, leading to higher energy costs.

STUDY METHOD

Electricity Energy and Heating Oil

Site-specific total electricity energy and heating oil consumption for individual residences show very high correlation with national weather temperature data when transformed to heating degree days (Quayle & Diaz, 1979). In this investigation, past climate conditions for temperature in Rasht City were considered. Energy demand for climate control was analyzed for Rasht city. The relevant parameters, climatological deviations from the desired indoor, were analyzed (Sivak, 2013). HDD, in cases where temperatures are below 18°C, and CDD, in cases where temperatures are

above 21°C, were used as energy consumption indices. Heating degree days and cooling degree days were analyzed (Jafari, 2013).

Location

Rasht is situated on the western part of the Albourz Mountain Range in the North West of Iran. The location of the study area is shown in Figure 2.

Temperature

Changing trends of temperature and precipitation in the last fifty years of the Caspian region have been investigated (Jafari, 2008a). In Rasht City, in the last 49 years (ended by 2005) precipitation increased by 56.4 mm and mean temperature increased by 1.28°C. Minimum temperature

increased by 2.45°C and maximum temperature by 0.08°C.

Past Climate Changes in Rasht

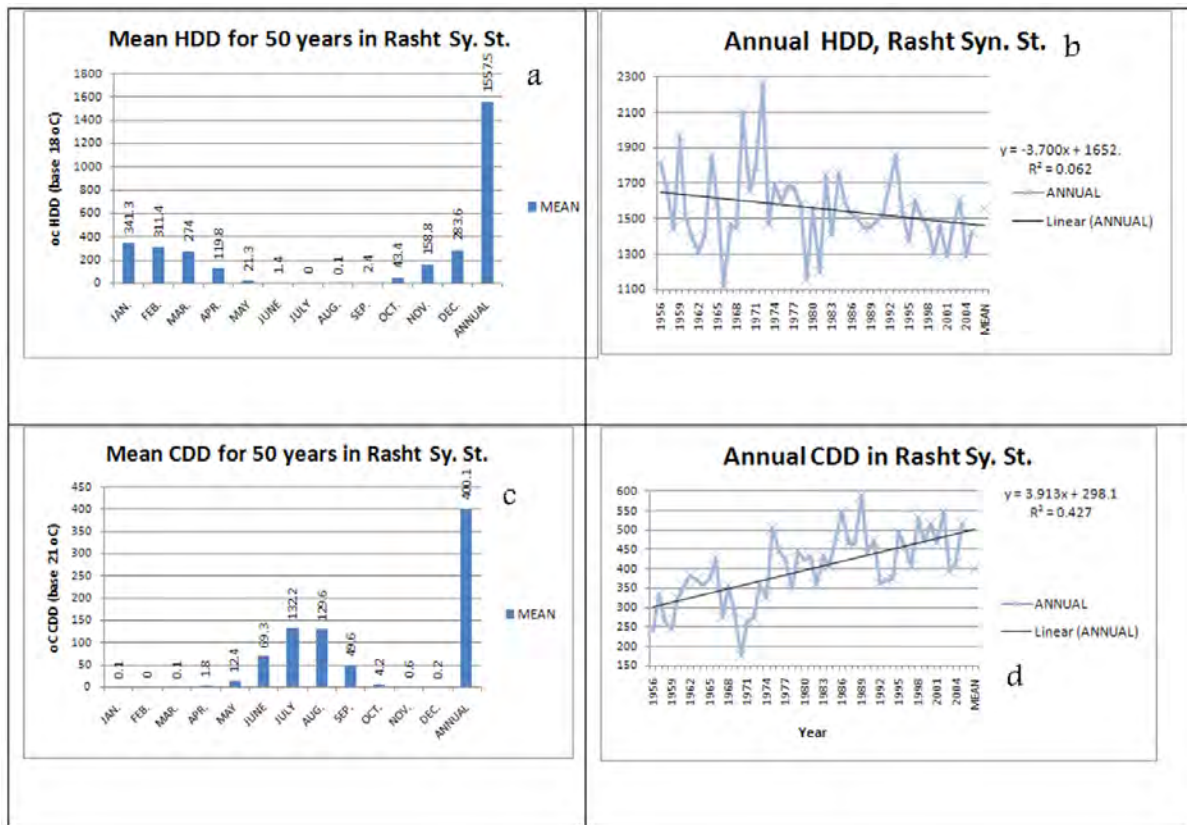
Past climate conditions for temperature in the region were analyzed in synoptic (49 years, 1956-2005) and climatology (Pasikhan climatology station which is nearest station to Rasht, 22 years, 1983-2005) stations. Two important factors, HDD and CDD, in Rasht, were considered. Mean annual daily temperature ($15 \pm 0.02^\circ\text{C}$) (1.28°C increase) and mean annual minimum ($10 \pm 0.05^\circ\text{C}$) (2.45°C increase) and maximum ($20 \pm 0.001^\circ\text{C}$) (0.08°C increase) temperatures have increased (Jafari 2008a). The HDD trend decreased in the last 49 years ($1652 \rightarrow 1470.7$) (Figures 3a and 3b), while at the same time, CDD has increased sharply (298

Figure 2. Study area location

Source: Maps from different sources: IPCC 2007 Asia WGII Chapter 10, Google, http://en.wikipedia.org/wiki/File:Locator_map_Iran_Gilan_Province.png



Figure 3. Rasht synoptic station, 1956 – 2005, total monthly mean: a) mean heating degrees days °C (based upon 18°C); b) linear trend of mean heating degrees days, °C; c) mean cooling degrees days, °C (based upon 21°C); d) linear trend of mean cooling degrees days, °C



→ 490) (Figures 3d and 3c), indicating a warming trend in the Rasht region.

Investigation on Climate Change Projections in Rasht

Based on IPCC data and models in global scale and national data and information, downscaling maps at national and regional scales (Figure 4) have been produced (Anonymous, 2008b). Temperature in study area of Rasht is projected to increase between 1.2 to 1.4°C. Mean temperature of Gilan province for the time slice of 1976-2005 was about 16.2°C, and is projected to increase by 1.3°C, to reach to 17.5°C for the period of 2010-2039. Based on the same results, projections for the future, the number of hot days in Rasht region will increase by

between 11-20 days, and the number of freezing days will decrease by between -11 and -8 days.

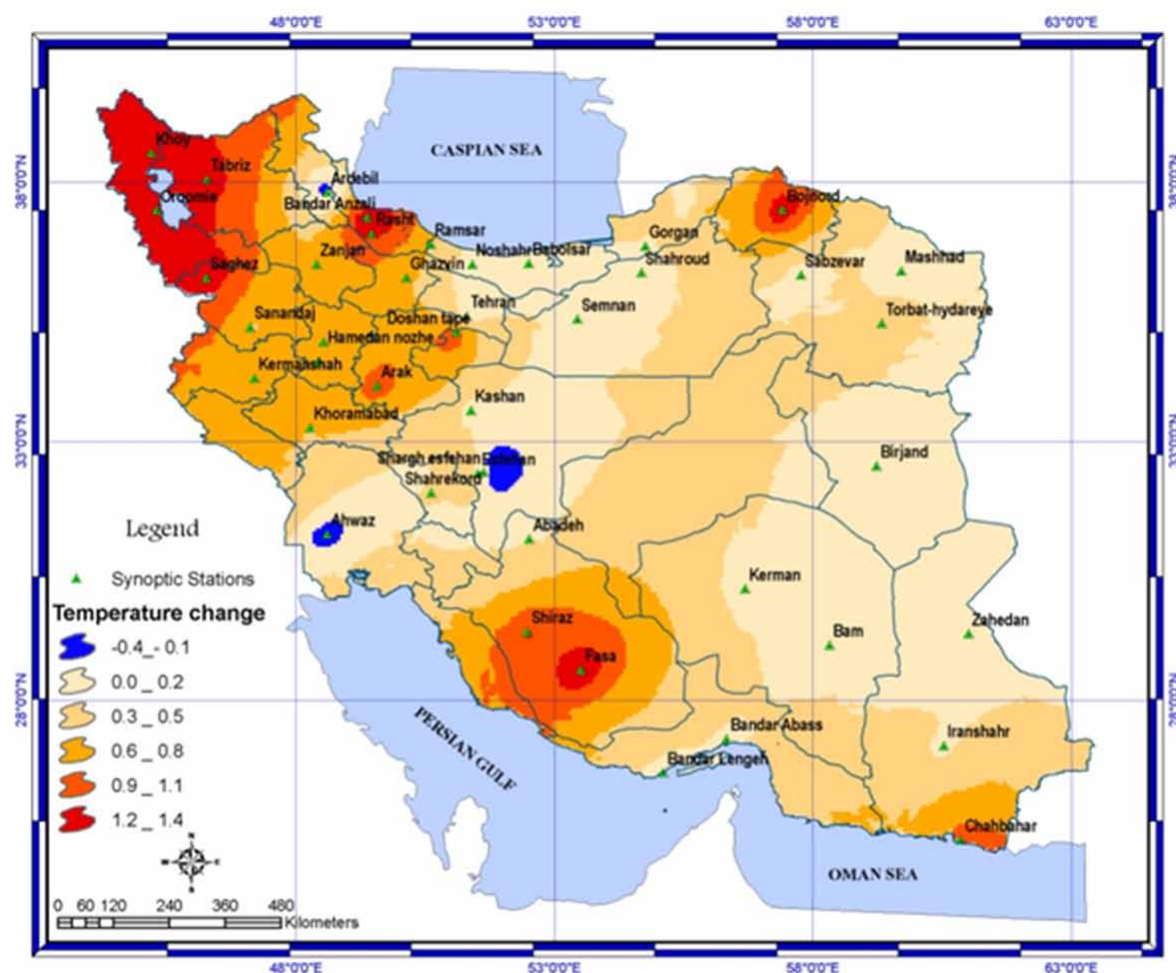
RESULTS

Heating Degree Days (HDD) and Cooling Degree Days (CDD)

In the Rasht case study, total monthly heating degrees recorded in the winter season reached 341.3 degrees of centigrade (January) which is about 11.37 degrees per day (Figure 3a). In the same period, cooling degrees was 0.1 (Figure 3c).

Total monthly cooling degrees recorded in the summer season reached 132.2 degrees of centigrade (July) which is about 4.40 degrees per day

Figure 4. Distribution of mean temperature differences comparing 2010-2039 with 1976-2005 according to the downscaling of ECHO-G model outputs
Source: Anonymous, 2008b



(Figure 3c). In the same period heating degree is about zero (Figure 3a).

The trend in HDD has decreased in last 49 years ($1652 \rightarrow 1470.7$) (Figures 3a and 3b), while in the same period CCD has increased at a sharper rate ($298 \rightarrow 489.7$) (Figures 3c and 3d). These factors certify the warming trend in the Rasht region.

Total annual heating degrees in years 1956 in Rasht station was 1652 degrees of centigrade (Figure 3b) and total annual cooling degrees was recorded as 298 degrees of centigrade (Figure 3d). During the 49 years between 1956 and 2005, the trend of mean annual temperature increased from 15°C by about 0.02°C to 15.02°C

This warming caused a decrease of total annual heating degrees by about 181.3 degrees of centigrade per year down to total annual of 1470.7°C (Figure 3b). Total annual cooling degrees increased by about 191.7°C per year up to total annual of 489.7°C (Figure 3d).

Decrease of 181.3°C per year of heating degrees and increase of 191.7°C of cooling degrees gives a difference of total of 10.4°C ($191.7 - 181.3 = 10.4^{\circ}\text{C}$) of extra energy consumption over the 49 year period.

Differences of HDD and CDD in year 1956 ($1652 - 298 = 1354$) and year 2005 ($1470.7 - 489.7 =$

981) shows about 373 degrees ($1354 - 981 = 373$), which is large.

The price of weather derivatives traded in the summer is based on an index made up of monthly CDD and HDD values. The settlement price for a weather futures contract is calculated by summing a month's CDD and HDD values and multiplying by \$20.

$$\text{HDD monthly cost} = \text{HDD} \times 30 \text{ (days of month)} \times 20 \text{ US\$}$$

In the Rasht case study a difference of annual total of 10.4°C calculated for an energy saving over 49 years' time. If we want to calculate cost as a rate of \$20 similar to USA conditions, after converting Celsius to Fahrenheit, it would be about \$374.4 as follow:

$$10.4 \times 9/5 = 18.72$$

$$18.72 \times \$20 = \$374.4$$

Temperature Changes

Based on recorded data in synoptic station, in Rasht during last forty-nine years, daily maximum (0.08°C), minimum (2.45°C) and mean (1.18°C) of temperature have increased.

Annual HDD with a base temperature of 18°C have reduced in the last 49 years (Figure 5e). HDD in January, which is a cold month, also slightly reduced. During the same period, annual CDD increased sharply (Figure 5f). CDD in July, August and September, which are warm months, also sharply increased (Figure 5d).

In Rasht City, total monthly heating degrees recorded in the winter season reached 341.3°C (January) which is about 11.4°C per day (Figure 3a). During the same period, cooling degrees were 0.1 (Figure 3c).

Total monthly cooling degrees recorded in the summer season reached 132.2°C (July) which is about 4.4°C per day (Figure 3c). During the same period, heating degrees were close to zero (Figure 3a).

SOLUTIONS AND RECOMMENDATIONS

Research work outcome on climate change impact on energy consumption pattern could help to plan better an adaptation action program.

It is a need to enhance much more research projects to consider climate change impact on natural resources including forest and firewood consumption.

FUTURE RESEARCH DIRECTIONS

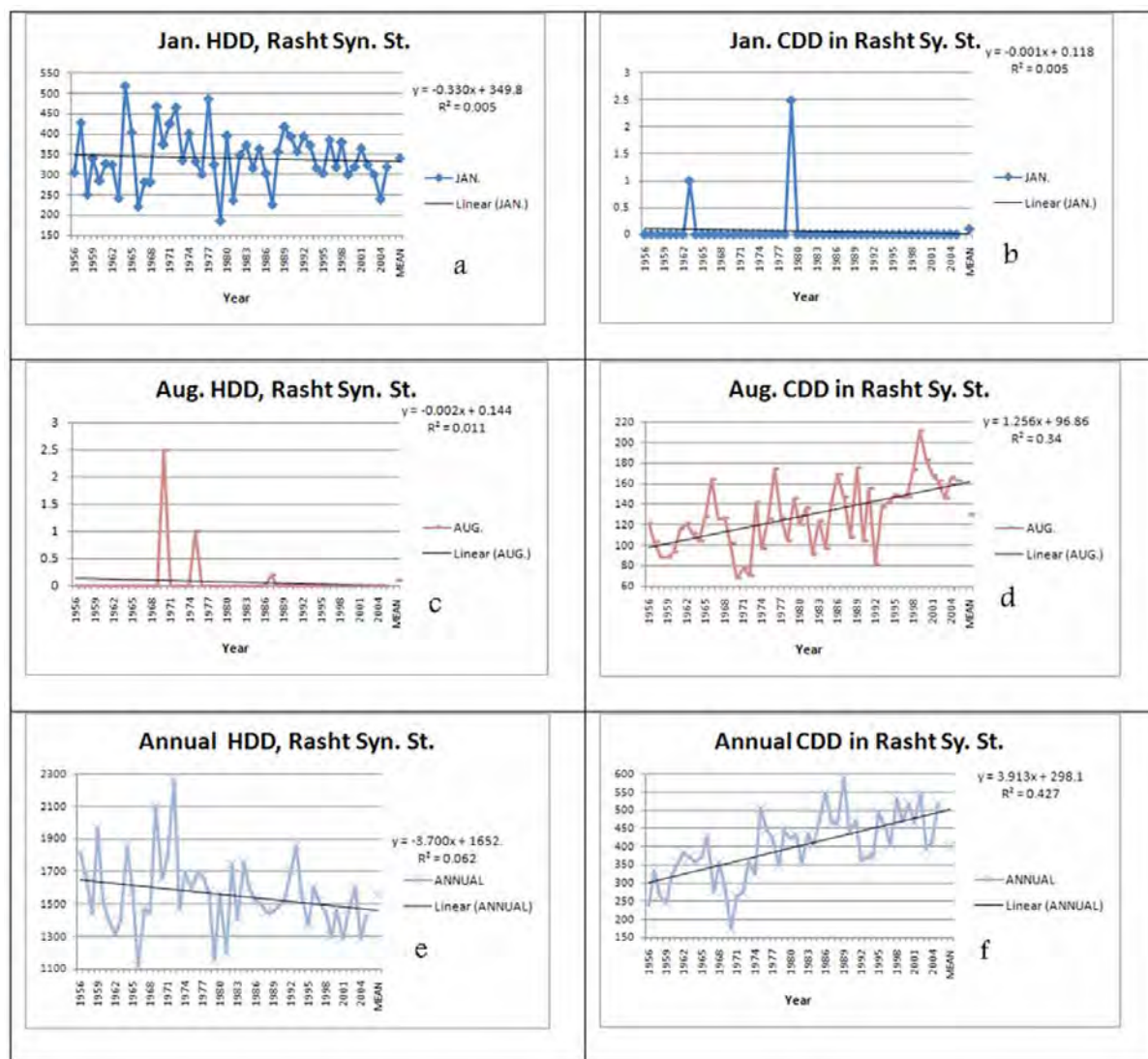
As it has been mentioned earlier, main field of research project should be identified for consideration of climate change impact on different sector in related to the energy consumption.

One direction is cooling and heating energy consumptions and other direction is using biomass as energy resources and it comparing with fossil fuel consumptions.

CONCLUSION

Climate change is driven by the cumulative amount of GHGs emitted to the atmosphere (Meinshausen, *et al.*, 2009; Solomon *et al.*, 2010), in particular of long-lived species such as carbon dioxide (CO_2) (Huber & Knutti, 2012). Therefore, to halt anthropogenic climate change, global GHG emissions need to be reduced to virtually zero in the long term (Matthews & Caldeira, 2008). Achieving this will require massive transformations of all GHG-emitting sectors. Given that the global energy system (which supplies fuels and electricity to the residential/commercial, industrial and transportation end-use sectors) is currently responsible for about 80% of global CO_2 emissions (RCP Database, 2009; Boden *et al.*, 2012), transformations in this sector will be essential for realizing a low-carbon future. Moreover, the need for transformational change in the energy system has also been advocated for other important reasons, for instance to spur sustain-

Figure 5. Rasht synoptic station, 1956–2005, monthly and annual total HDD ($^{\circ}\text{C}$, base 18°C) and CDD ($^{\circ}\text{C}$ base 21°C): a) HDD in January, b) CDD in January, c) HDD in August, d) CDD in August, $^{\circ}\text{C}$, e) HDD for total annual, and f) CDD for total annual



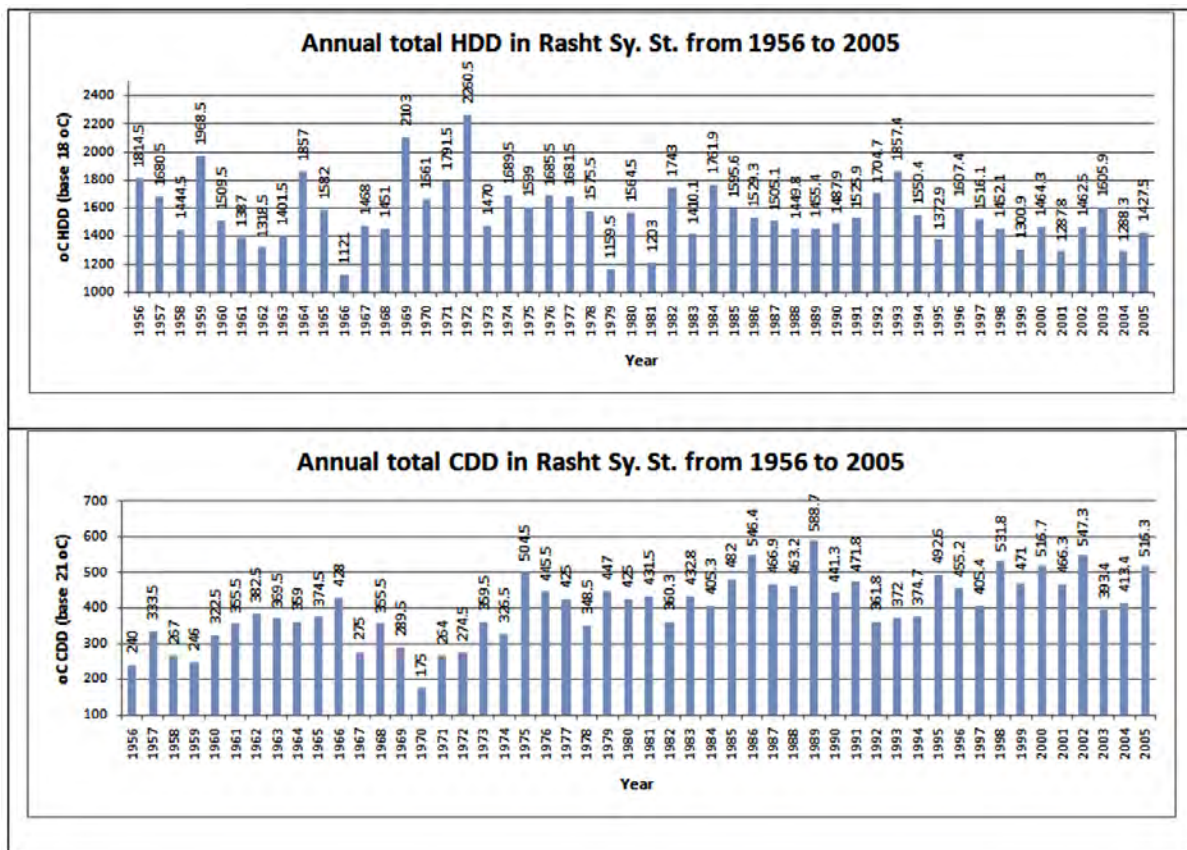
able development or to improve the well-being of the impoverished billions in our society who lack regular access to modern forms of energy to meet their basic needs. From this perspective, we look at how energy-related targets fostering the abovementioned objectives would, or would not, be consistent with global climate protection (Rogelj *et al.*, 2013).

Mean annual temperature and even minimum temperature in last half century in Rash city has

increased (Jafari, 2008a), Table 1). Projections for future temperature changes in Rasht are mainly documented for increasing temperature and the extent of temperature increase varies by scenario (Jafari, 2012c).

The data provided by Energy in Focus, BP Statistical Review of World Energy, BP Plc, 2004, bp.distributionservices@bp.com, and adapted by Coley, (2008) shows how energy is spilt across the world. Obvious inequalities can be observed in

Figure 6. Rasht synoptic station, 1956 – 2005, annual total HDD ($^{\circ}\text{C}$, base 18°C) (top), CDD ($^{\circ}\text{C}$ base 21°C) (bottom)



how the world's primary energy is shared out, with North America (mainly the USA) using almost ten times more primary energy than the whole of Africa. This point is re-emphasized if we divide national primary energy use by national population, to give the energy use per capita (Figure 7). Much of the world is seen to use less than 63 GJ (1.5 Toe) (see end note) per capita per annum, whereas others use many times this (Coley, 2008). The annual primary energy consumption per capita (toe) (ENE, 2004) for Iran is in the range of 1.5-3, which in the development process, may increase.

As reported by National Iranian Oil Co. (NIOC 2007), in comparison of years 2006 and 2007, consumption of white oil (petroleum) has decreased by – 5.7%. This decrease may be described of effect of warming as well as subsidy payment

policy (Table 1, data obtained from NIOC (2007) and modified by authors).

Rasht city, according to the IPCC assessments report, is located in both Central Asia and West Asia sub-regions (Cruz *et al.*, 2007). Based on the projection for Central Asia sub-region, the region will face with higher degrees of temperature change. Downscaled projections in local level project that Gilan province as a whole will have a higher degree of temperature change than the western part of the province. Westwards into the mountainous area, a reduction in temperature is projected. Downscaling and projection of seasonal extreme daily precipitation has been studied elsewhere (Wang & Zhang, 2008).

Based upon mean monthly temperature changes in Rasht city, Jan. & Feb. were recorded

Figure 7. Annual primary energy consumption per capita (toe)

Source: ENE (2004). Reprinted from *Energy in Focus*, BP Statistical Review of World Energy, BP Plc, 2004, bpdistribution-services@bp.com



Table 1. Comparison of energy consumption changes in Rasht city for years 2006 and 2007

Rasht City	Consumption (1000 Lit.)		Consumption Changing Rate (%)	Subsidy 2007	
	2006	2007		1000 US\$	Million Rials
Petroleum (White Oil)	126136	118906	- 5.7	66485	618311
Gas Oil	442099	594478	34.5	326004	3031838
Heating Oil	7128	8884	24.6	2961	27540

Source: Data obtained from NIOC (2007) and modified by authors

as coldest months and July & Aug. were recorded as warmest months (Figure 8 – top).

Monthly mean HDD and CDD changes are in high correlation with temperature changes pattern (Figure 8 – middle).

Marginal plot of annual total CDD vs HDD presents distributed dots as normal form for HDD, but for CDD is presented a little abnormality, which may refer to changing and increasing pattern of temperature (Figure 8 – bottom).

The amount of monthly and annual total of HDD (mean= 1556 Figure 9 bottom) are much higher than CDD (mean=400 Figure 9 top) in the investigated area. Even though, annual total HDD

and CDD have a negative correlation (Pearson correlation = - 0.493; $p < 0.001$).

HDD declines only a little and steadily ($R^2 = 0.062$ Figure 3b, $p = 0.099$ Figure 9 bottom), but CDD increases sharply during the same time slice ($R^2 = 0.427$ Figure 3d, $p = 0.813$ Figure 9 top). This means that the energy consumption pattern is expected to have increased sharply, and is expected to increase more rapidly in the future, which will result in increased costs.

Energy consumption for heating and cooling are driven by climate factors such as temperature, and is expected to change with a changing climate elements. Energy optimization solutions should

Figure 8. Recorded mean monthly temperature changes in Rasht city (top), Rasht synoptic station, 1956 – 2005, CDD (base 21°C) and HDD (base 18°C) for mean months for 50 years (middle), and marginal plot of annual total of CDD and HDD for 50 years (bottom)

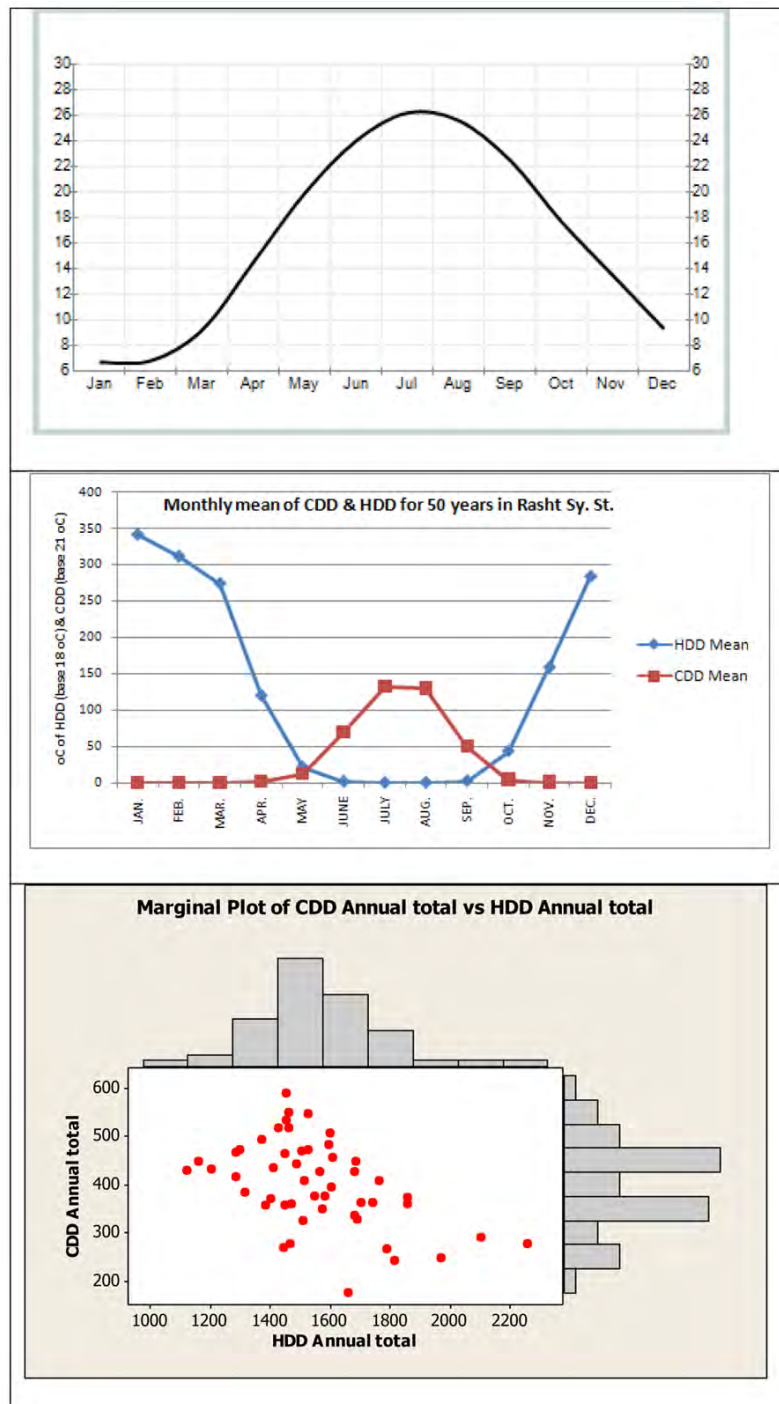
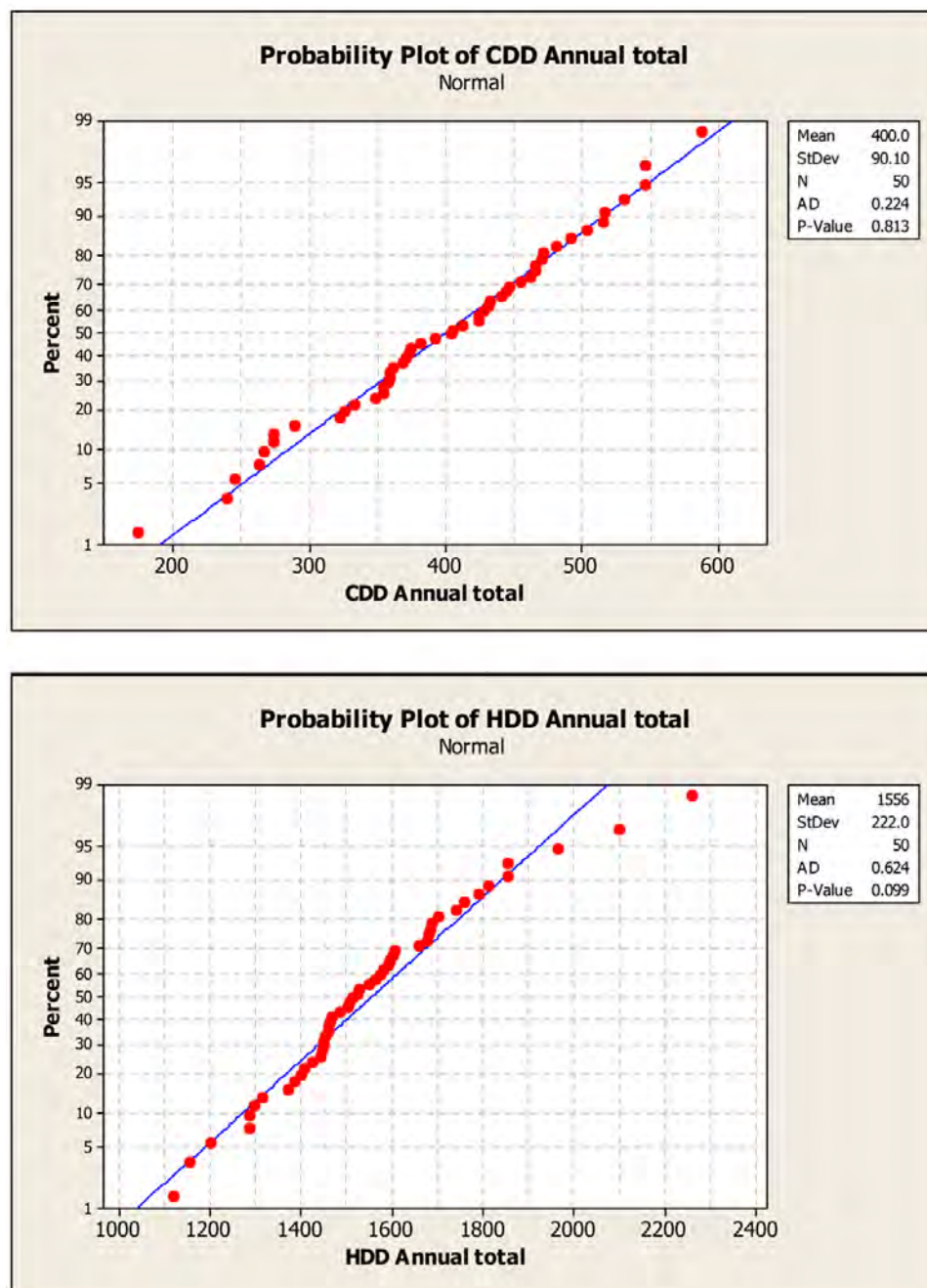


Figure 9. Rasht synoptic station, 1956 – 2005, probability plot of annual total of CDD (base 21°C) (top), and HDD (base 18°C) (bottom)



be considered, as an important energy related issue. It has been investigated in Nigeria at country level (Anayochukwu & Ndubueze, 2012). Results obtained by the Sivak (2013), in USA indicate that climate control in Minneapolis (the coldest large metropolitan area in the US) is about 3.5 times

as energy demanding as in Miami (the warmest large metropolitan area in the US). This finding suggests that, in the US, living in cold climates is more energy demanding than living in hot climates (Sivak, 2013).

Heating degrees, cooling degrees are factors which impact on energy consumptions and these have changed, and are projected to change further in the future in Rasht City, unless mitigated by improved building energy performance, as suggested for Greece (Maleviti 2012). Heating and cooling energy demands have changed in Rasht by 10.4°C between 1956 and 2005, representing an annual extra cost of ~\$374, calculated based upon US daily costs for heating and cooling.

REFERENCES

- Anayochukwu, A. V., & Ndubueze, N. A. (2012). Energy Optimization at GSM Base Station Sites Located in Rural Areas. *International Journal of Energy Optimization and Engineering*, 1(3), 1–31. doi:10.4018/ijeoe.2012070101
- Anonymous. (2008a). *Australian government, Bureau of meteorology*. Retrieved from <http://www.bom.gov.au/index.shtml>
- Anonymous (2008b). *CCIRIMO, National Centre of Climatology*. Islamic Republic of Iran Meteorological Organization, Downscaling Report.
- Anonymous. (2008c). *Nebraska Public Power District*. Retrieved from <http://www.nppd.com/>
- Anonymous. (2008d). *NOAA, The National Oceanic and Atmospheric Administration*. Retrieved from <http://www.noaa.gov>
- Anonymous. (2008e). *The Free Dictionary*. Retrieved from <http://financial-dictionary.thefreedictionary.com/Heating+Degree+Day+-+HDD>
- Anonymous. (2012). *Wikipedia, The Free Encyclopedia*. Retrieved from http://en.wikipedia.org/wiki/Heating_degree_day
- Boden, T. A., Marland, G. & Andres, R. J. (2012). *Regional, and National Fossil-Fuel CO2 Emissions*. 10.3334/CDIAC/00001_V2012
- Coley, D. A. (2008). *Energy and climate change: creating a sustainable future*. John Wiley & Sons Ltd.
- Cruz, R. V., Harasawa, H., Lal, M., Wu, S., Anokhin, Y., Punsalma, B.,... Huu Ninh, N. (2007). *Asia. Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.
- ENE. (2004). *Energy in Focus, BP Statistical Review of World Energy*. BP Plc.
- Hodgson, P. E. (2010). *Energy, the Environment and Climate Change*. Imperial College Press.
- Hossain, A. K. (2012). Global Energy Consumption Pattern and GDP. *International Journal of Renewable Energy Technology Research*, 1(1), 23 - 29. Available online www.ijretr.org
- Huber, M., & Knutti, R. (2012). Anthropogenic and natural warming inferred from changes in Earths energy balance. *Nature Geoscience*, 5(1), 31–36. doi:10.1038/ngeo1327
- Jafari, M. (2008a). Investigation and analysis of climate change factors in Caspian Zone forests for last fifty years. *Iranian Journal of Forest and Poplar Research*, 16(2), 314-326.
- Jafari, M. (2010). *Climate Change Impact on Iranian Ecosystems, with review on Climate Change Study Methods*. Research Institute of Forests and Rangelands Publication.
- Jafari, M. (2011). *Forest trees as good indicators for precipitation and humidity changes: A dendrochronological study in Hyrcanian forest ecosystems*. Presentation in the Workshop on “The Caspian Sea Level Change: Human Impact and Impact of Sea Level” Iranian National Centre for Oceanography, Tehran, Iran.
- Jafari, M. (2012a). Climate and environmental impacts on beech and oak wood production in the Hyrcanian forests. *Iranian Journal of Wood and Paper Science Research*, 27(3), 386-408.

- Jafari, M. (2012b). Aviation Industry and Environment Crisis: A Perspective of Impacts on the Human, Urban and Natural Environments. In *Technology Engineering and Management in Aviation: Advancements and Discoveries*. IGI-Global. Retrieved from <http://www.igi-global.com>
- Jafari, M. (2012c). Adaptation of global change projection for local level application: Case study, special forest ecosystem, Astara. *Iranian Journal of Natural Resources Research*, 1(1), 12-24. Retrieved from <http://www.rifr-ac.org/FA/userfiles/file/news/ijnrr.pdf>
- Jafari, M. (2013). Energy consumption impacted by climate changes application: Case study Astara. Lecture Note in Electrical Engineering, 239, 29-50.
- Maleviti, E. (2012). The Development of a Theoretical Model for Optimizing Energy Consumption in Buildings. *International Journal of Energy Optimization and Engineering*, 1(4), 1-14. doi:10.4018/ijeoe.2012100101
- Matthews, H. D., & Caldeira, K. (2008). Stabilizing climate requires near-zero emissions. *Geophysical Research Letters*, 35(4), L04705. doi:10.1029/2007GL032388
- Meinshausen, M., Meinshausen, N., Hare, W., Raper, S. C. B., Frieler, K., Knutti, R., & Allen, M. R. et al. (2009). Greenhouse-gas emission targets for limiting global warming to 2 °C. *Nature*, 458(7242), 1158-1162. doi:10.1038/nature08017 PMID:19407799
- NIOC. (2007). *National Iranian Oil Co.* Retrieved from <http://en.nioc.ir/Portal/Home/>
- Oliver, W. (2013). World Energy Trilemma, Time to get real – the agenda for change. World Energy Council.
- Quayle, R. G., & Diaz, H. F. (1979). Heating Degree days Data Applied to Residential Heating Energy Consumption. *Journal of Applied Meteorology*, 19(3), 241-246. doi:10.1175/1520-0450(1980)019<0241:HDDDAT>2.0.CO;2
- RCP Database. (2009). *RCP Database (version 2.0)*. Retrieved from <http://www.iiasa.ac.at/web-apps/tnt/RcpDb>
- Rogelj, J., McCollum, D. L., & Riahi, K. (2013). The UN's 'Sustainable Energy for All' initiative is compatible with a warming limit of 2 °C. *Nature Climate Change*, 3. Retrieved from www.nature.com/natureclimatechange
- Sivak, M. (2013). Air conditioning versus heating: climate control is more energy demanding in Minneapolis than in Miami. *Environ. Res. Lett.*, 8. Retrieved from stacks.iop.org/ERL/8/014050
- Smith, P., Bustamante, M., Ahammad, H., Clark, H., Dong, H. M., Elsiddig, E. A., & Tubiello, F. N. et al. (2013, August). How much land based greenhouse gas mitigation can be achieved without compromising food security and environmental goals? *Global Change Biology*, 19(8), 2285-2302. doi:10.1111/gcb.12160 PMID:23505220
- Solomon, S., Daniel, J. S., Sanford, T. J., Murphy, D. M., Plattner, G.-K., Knutti, R., & Friedlingstein, P. (2010). Persistence of climate changes due to a range of greenhouse gases. *Proceedings of the National Academy of Sciences of the United States of America*, 107(43), 18354-18359. doi:10.1073/pnas.1006282107 PMID:20937898
- Wang, J. F., & Zhang, X. B. (2008). Downscaling and projection of winter extreme daily precipitation over North America. *Journal of Climate*, 21(5), 923-937.

KEY TERMS AND DEFINITIONS

Cooling Degree Days (CDD): If the average daily temperature is above comfort levels, cooling is required. An analogous index for the energy demand for cooling is represented by CDD (Sivak, 2013). CDDs are based on daily temperature observations, with each month having at least 25 records and no less than 15 years of data (Anonymous, 2008a). CDD, which indicate the level of comfort,

are based on the average daily temperature which is taken as the mean of maximum and minimum daily temperature.

Energy Consumption Pattern: The way and total use of energy may be defined as its pattern. In general energy is classified into two main groups: renewable and non-renewable. Renewable energy is the cleanest sources of energy and non-renewable sources are not environmental friendly source of energy. According to (Akhter Hossain, 2012) GDP and energy consumption of developing countries are increasing exponentially, whereas GDP and energy consumption of developed countries are increasing linearly.

Heating Degree Days (HDD): If the average daily temperature falls below comfort levels, heating is required. HDD is an index of the energy demand to heat buildings (Sivak, 2013). HDDs are based on daily temperature observations, with each month having at least 25 records and no less than 15 years of data (Anonymous, 2008a). HDD, which indicate the level of comfort, are based on the average daily temperature which is taken as the mean of maximum and minimum daily temperature.

APPENDIX

Although the joule will be our standard unit, the importance of oil is such that the energy industry more often uses tonnes-of-oil-equivalent (toe). This is the energy content of one tonne of oil, i.e. 42 billion joules ($= 42 \times 10^9 \text{ J} = 42 \text{ GJ}$). (Coley, 2008).

Table 2. Common Energy Units

Unit	Pronunciation	Equivalence	Notes
kWh	kWh	$1 \text{ kWh} = 3.6 \times 10^6 \text{ J}$	Energy used by a single-bar electric fire left on for one hour
ft.	lb foot-pound	$1 \text{ ft.lb} = 1.356 \text{ J}$	Energy required to raise one pound in weight ($\approx 0.45 \text{ kg}$) by one foot ($\approx 0.3 \text{ m}$)
cal	calorie	$1 \text{ cal} = 4.19 \text{ J}$	Energy required to heat 1 g of water 1°C (at atmospheric pressure)
kcal or Cal	kilo-calorie	$1 \text{ Cal} = 4.19 \text{ kJ}$	Unit used by the food industry and easily confused with the cal
Therm	therm	$1 \text{ therm} = 100\,000 \text{ BTU}$	
BTU	British thermal unit	$1 \text{ BTU} = 1055 \text{ J}$	
Toe	tonne of oil equivalent	$1 \text{ toe} = 42 \text{ GJ}$	

Determination of Urban Growth by the Night-Time Images



Emre Yücer

Erzincan University, Turkey

Arzu Erener

Kocaeli University, Turkey

INTRODUCTION

Urbanization in Turkey and the world has continued to increase especially in the last 30 years. For developing countries, urbanization is a necessary phenomenon, especially in terms of the sustainability of urban growth. However, this growth must be controlled and in a planned manner, otherwise it could lead to many environmental problems. Therefore, urban areas, having a dynamic structure, should be monitored regularly and identified. Adequately accurate current data, which changes temporally and enables to monitor urban development and land use at regular intervals, are needed. Remote sensing technology is one of the most preferred methods in recent years to determine the urban areas developing fast and dynamically.

Within the scope of this study, using night-time images integrated with GIS, the spatial extent of urban areas in Turkey was identified and the methods used in the identification of the urban areas were compared. To determine the urban size of the cities in Turkey, DMSP-OLS (US Air Force Meteorological Satellite Program (Defense Meteorological Satellite Program) - Operational Line Scan System (Operational Linescan System) night-time images belonging to 2010 were used. While determining the urban size from night-time images, two different image analysis methods, object-based and pixel-based, were used. Object-based image analysis method involves image segmentation and pixel-based classification method involves natural breaks and Otsu method. By

comparing the results obtained through these three methods, the method that gives the most accurate urban area size has been determined. Through the comparison of these methods' features used in the determination of the urban areas, the differences, advantages, and disadvantages of these methods have been revealed.

BACKGROUND

Urbanization is quickly increasing in the world and according to data of the United Nations Population Fund it is estimated that the world's population will be 12 billion by 2050, 80% of this population will be living in cities. In our country where an intensely urban migration from villages has been observed, there has been a clear increase in the urban areas. The effects of this increase occurring in urban areas on the social, economic and cultural characteristics of a city and population are inevitable.

Observation and evaluation of these changes occurring in urban areas are crucial for making balanced and sustainable urban development. Today, through the use of remote sensing and information technologies, the assessment process of data is accelerating and the productivity is increasing. In recent years, night-time images have been often used in the studies that determine the urban areas. The fact that these images can be accessed quickly and their assessment process is fast increases the interest in the night images.

DOI: 10.4018/978-1-5225-2255-3.ch681

The method was used to detect a threshold value for Chicago, Miami and Sacramento states in the USA. Maps of urban areas were created from night-time images in accord with this threshold value (Imhoff et al., 1997).

It has been showed that a single threshold value in night-time images has been more appropriate to determine the urban areas for different metropolitan regions. Therefore, Henderson et al. have indicated that the threshold value of the data in night-time images should be determined separately for each year for which a temporal analysis needs to be carried out for urban growth detection (Henderson et al., 2003).

As there is very little fluctuation in light intensity in Italy's Sicilian city at night, Elvidg et al. have implemented a quadratic regression model to inter-calibrate among years by using the data of the city as the reference data. The purpose of using a single city for calibration in night-time images is to correct the changes according to the years in the satellite and detector sensors (Elvidg et al., 2009).

By examining the global economic and demographic differences of night-time images, Levin and Duke conducted a study that could be used as an indication of the demographic and socio-economic characteristics of the residential area at a local scale. In the study, the population data of the settlements Israel and West Bank (Arab and Jewish) were compared. It has been stated that the economic and geopolitical differences between Israel and the West Bank lead to pattern differences in night-time images (Levin & Duke, 2012).

Liu et al. carried out a study to uncover the relationship between making timely and right decisions about the dynamics of urban expansion in China between 1998-2002, urban growth, and the ecosystems and to optimize the ways of land use (Liu et al., 2012).

The study aiming to identify the variations of urban areas in major metropolitan areas in China between the years 2000-2010 consists of three steps; an average threshold for the reference year is determined in the first step, invariant features

(Pseudo Invariant Features) on the urbanization process are determined and then are normalized with respect to DMSP-OLS data PIF value. Linear regression model was applied in the normalization process. Furthermore an optimal threshold for urban area detection for each year has been designated with respect to reference threshold value and urban areas have been identified based on this (Wei et al., 2014).

MAIN FOCUS OF THE ARTICLE

The aim of this study is to obtain the urban sizes of 81 cities in Turkey by utilizing satellite imagery for 2010. Another object of the study is to apply different image analysis methods in order to determine the method which gives the optimal result in obtaining urban size and to compare these methods with each other. In this study, it is planned to compare two different methods, including object-based and pixel-based.

Used Data: What Are Night-Time Images

Night-time images for 2010 have been used to determine urban areas in Turkey. The NGDC started to develop algorithm in 1994 for global cloudless yearly night lights products collected by U.S. Air Force Meteorological Satellite Program (Defense Meteorological Satellite Program, DMSP) Operational Linescan System (Operational Linescan System, OLS). Temporal series of products are available at <http://ngdc.noaa.gov/eog/dmsp/downloadv4composites.html> and night-time images for the study area have been retrieved free of charge from this address.

By analyzing the location information, frequency and light image in temporal series of images, four kinds of main light sources in the world: settlements, fires, gas flares and fishing boats can be distinguished. Each pixel has 30x30 arc-second (or 0.86 km² at the equator) resolution and the images cover the area between +180 and

-180 latitudes and between -65 and +75 longitudes. Night-time images have been created as a composite presentation of two satellite images taken in cloudless nights. Each composite set has been labeled with satellite and year information.

These images used in the study contain data from fixed sources of light such as cities and towns. Moreover, background noise detection studies were carried out and these values were replaced with value of zero, and thus the image has been purged of these values.

Methods Used to Determine Urban Areas

Different types of features in digital images constitute combinations with different numerical values depending on the natural spectral reflectance and spreading properties. Earthly objects with the same spectral properties can be grouped benefiting from these differences. Assigning each pixel in the satellite images to the clusters corresponding on the earth with respect to the spectral properties and reflection values can be described as image classification. It is possible to divide the classification process into two main groups, namely pixel-based classification and object-based classification.

While determining the urban size from night-time images, two different image analysis methods, object-based and pixel-based, were used. Object-based image analysis method involves image segmentation and pixel-based classification method involves natural breaks and Otsu method. By comparing the results obtained through these three methods, the method that gives the most accurate urban area size has been determined.

Image segmentation can be described as dividing an image into meaningful regions each of which has different features in them. For example, there may be similar radiances in the image and these radiances can represent the objects in different regions of the relevant image. Pixels with similar reflectance properties during the segmentation stage are grouped and thus instead of working with millions of pixels, segments are formed by combining pixels with the common characteristics and at this stage the lowest heterogeneity level among segments will have been reduced. Depending on the importance of homogeneity or heterogeneity, users at this stage generate segmentation layers at various scale factors using spectral layers (Benz, et al, 2004). In addition, it is simplified to obtain many features during the classification using

Figure 1. Turkey night view of 2010

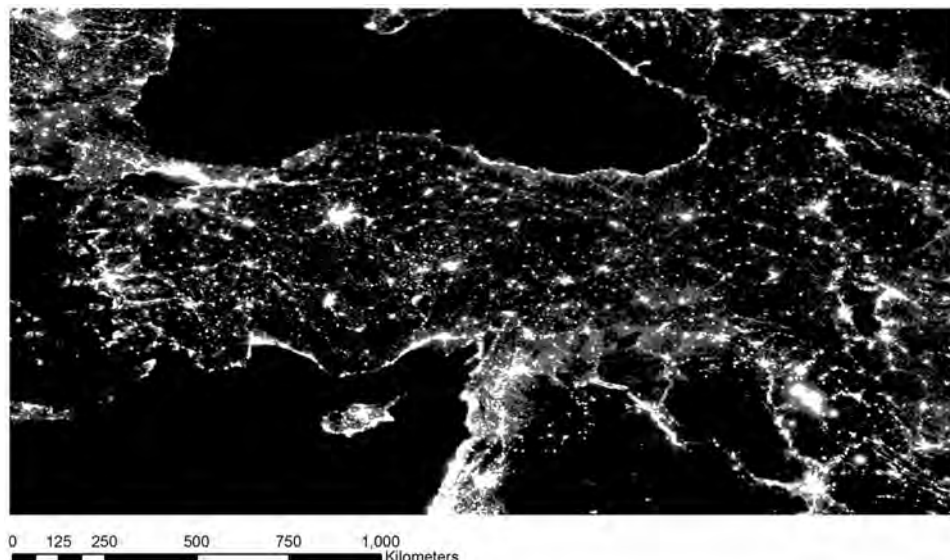
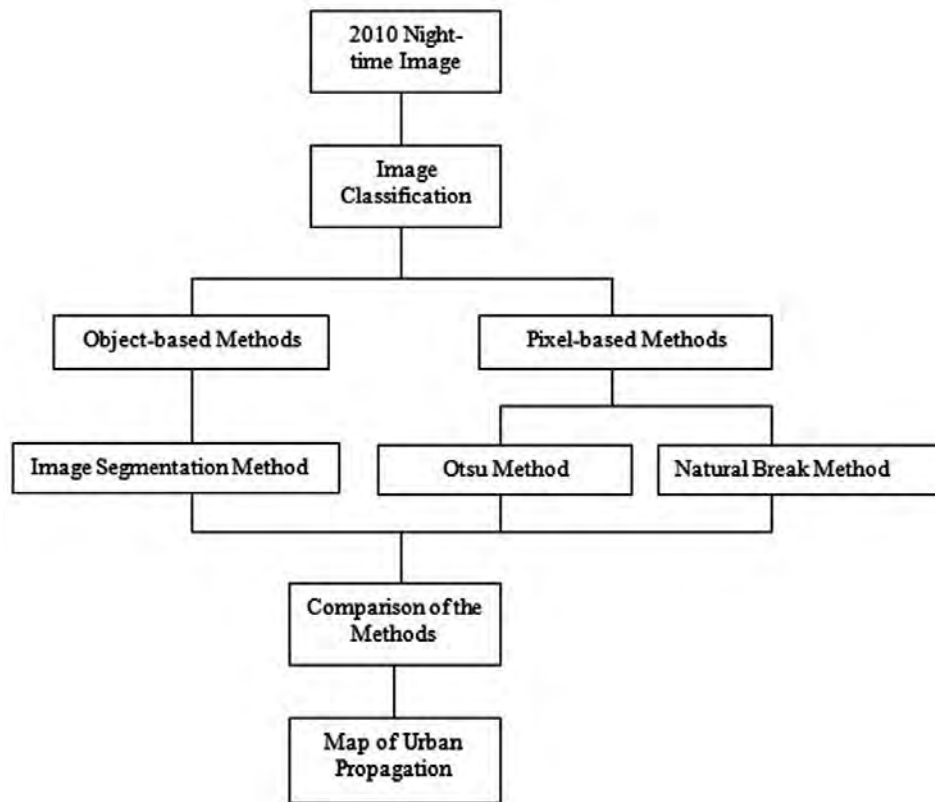


Figure 2. Application flowchart



features such as texture, length and area as well as the reflection features of resulting segments.

In this study, an image was divided into classes according to the edge detection (Edge Detection) method, an image segmentation method. In the edge detection method, the edge, lighting or surface reflections in an image correspond to significant changes occurring in the physical appearance of the image and this change shows itself as brightness, color and texture. In this study, in terms of edge detection, by looking at the changes in image brightness, the regions in which there are sudden changes in the gray levels of an image are called an edge (Aydin, 2014).

Maps obtained in the pixel-based image classification methods are produced so that the area of earth that image pixels correspond belongs to only one category. Classification algorithms produce a certain similarity function for assigning a value of any category to a pixel. In pixel-based

classification, an unknown pixel is assigned to the class label in which this pixel has the largest similarity value. The results of such a classification are univocal (Sunar et al., 2011).

In the study, image analysis method used as pixel-based contains two different applications, including the Otsu Method and Natural Breaks Method.

Otsu code method is a threshold setting method that allows determining the optimal threshold value of an image at a gray level that can be used while it is reduced to two groups. Otsu method is a threshold detection method that can be applied on gray level images. While this method is performed, it is assumed that the image consists of two color classes, including the background and the foreground. Then, in-class variance of these two color classes for all threshold values is calculated. The threshold value that enables this value to be the smallest is the optimum threshold

value. While in-class variance value is at minimum level, inter-class variance value becomes at maximum value. The method works on gray level images and only examines how often colors are on the image. So, the color histogram of the image is calculated firstly, and all operations are performed on the histogram array (Atasoy, 2015).

Natural Breaks method is a re-classification method. The working principle of this method is based on Jenks natural breaks algorithm. Jenks natural breaks algorithm is a classification procedure designed to determine the best way to organize the data in different classes. How many classes will be obtained in this method is described

and then pixel range values are obtained based on Jenks natural breaks.

SOLUTIONS AND RECOMMENDATIONS

Results Comparison of Images: Methods

By determining the urban areas according to the three methods, the urban area ratio of each province based on the total urban area across the country is presented as a percentage in Table 1.

Table 1. Percentage value of urban areas obtained by three methods in total urban area

City Name	Image Segmentation	Natural Break	Otsu
ADANA	3.42	2.99	3.01
ADIYAMAN	0.42	0.48	0.49
AFYON	0.73	0.89	0.91
AĞRI	0.33	0.34	0.35
AMASYA	0.10	0.31	0.30
ANKARA	7.15	8.59	8.78
ANTALYA	4.75	4.05	4.02
ARTVİN	0.20	0.19	0.19
AYDIN	0.81	0.61	0.57
BALIKESİR	0.52	0.69	0.73
BİLECİK	0.88	0.35	0.37
BİNGÖL	0.21	0.24	0.26
BİTLİS	0.30	0.25	0.24
BOLU	1.12	0.85	0.69
BURDUR	0.49	0.48	0.43
BURSA	3.63	3.52	3.60
ÇANAKKALE	0.35	0.48	0.44
ÇANKIRI	0.19	0.23	0.24
ÇORUM	0.62	0.66	0.62
DENİZLİ	2.12	1.38	1.29
DİYARBAKIR	0.85	1.05	1.04
EDİRNE	1.95	0.65	0.67
ELAZIĞ	0.71	0.87	0.88
ERZİNCAN	0.44	0.58	0.58

continued on following page

Table 1. Continued

City Name	Image Segmentation	Natural Break	Otsu
ERZURUM	0.95	1.20	1.19
ESKİŞEHİR	1.02	1.36	1.42
GAZİANTEP	2.23	2.16	2.15
GİRESUN	0.52	0.33	0.28
GÜMÜŞHANE	0.16	0.19	0.21
HAKKARİ	0.26	0.21	0.20
HATAY	1.41	1.27	1.25
ISPARTA	0.52	0.67	0.63
İÇEL	3.41	3.32	3.25
İSTANBUL	11.86	14.10	13.48
İZMİR	7.65	6.86	7.14
KARS	0.31	0.39	0.41
KASTAMONU	0.30	0.34	0.34
KAYSERİ	2.09	2.51	2.55
KIRKLARELİ	0.23	0.24	0.24
KIRŞEHİR	0.39	0.38	0.39
KOCAELİ	4.50	4.71	4.55
KONYA	2.68	2.53	2.48
KÜTAHYA	0.82	0.76	0.75
MALATYA	1.73	1.38	1.43
MANİSA	0.73	0.89	1.15
K.MARAŞ	0.76	0.92	0.94
MARDİN	0.77	1.01	1.05
MUĞLA	0.19	0.36	0.39
MUŞ	0.74	0.32	0.31
NEVŞEHİR	0.94	0.92	0.86
NİĞDE	0.83	1.04	1.04
ORDU	0.28	0.42	0.40
RİZE	0.97	0.52	0.47
SAKARYA	1.50	1.89	1.90
SAMSUN	2.03	1.51	1.50
SİİRT	0.28	0.22	0.27
SİNOP	0.22	0.21	0.25
SİVAS	0.64	0.68	0.69
TEKİRDAĞ	0.50	0.57	0.59
TOKAT	0.38	0.44	0.43
TRABZON	1.67	1.08	1.17

continued on following page

Table 1. Continued

City Name	Image Segmentation	Natural Break	Otsu
TUNCELİ	0.22	0.26	0.20
ŞANLIURFA	0.71	0.91	0.94
UŞAK	0.41	0.59	0.56
VAN	0.89	1.05	1.01
YOZGAT	0.35	0.27	0.28
ZONGULDAK	0.46	0.57	0.59
AKSARAY	0.55	0.57	0.60
BAYBURT	0.26	0.23	0.24
KARAMAN	0.52	0.49	0.50
KIRIKKALE	0.85	0.83	0.82
BATMAN	0.35	0.54	0.55
ŞIRNAK	0.11	0.15	0.16
BARTIN	0.10	0.21	0.23
ARDAHAN	0.21	0.13	0.13
IĞDIR	0.28	0.34	0.35
YALOVA	1.22	1.11	1.39
KARABÜK	0.40	0.53	0.55
KİLİS	0.37	0.37	0.38
OSMANİYE	0.97	0.75	0.67
DÜZCE	2.01	1.46	1.42

Correlation analysis method was used while comparing each method to each other. In probability theory and statistics, correlation indicates the direction and strength of the linear relationship between two random variables. The method used to determine the degree of the relationship between two variables or the correlation is called the correlation analysis. The correlation coefficient r takes values between $-1 \leq r \leq +1$. $r = -1$ means complete negative correlation, $r = +1$ means complete positive linear relationship, and $r = 0$ indicates that there is no relationship between the two variables. The positive relationship indicates a tendency of increase in the values of Y variable in case of an increase in the values of X variable, or tendency of decrease in the values of Y variable in case of an decrease in the values of X variable. Negative correlation (negative relationship) means that if there is an increase in

the value of one variable, there will be a decrease in the values of other variables. The correlation coefficient of “0” shows that there is not a linear relationship between variables.

When Table 2 is examined, it could be seen that correlations of the three methods with each other is high; and there appears to be a very high positive linear relationship especially when the correlation coefficient between Otsu method and

Table 2. Correlation coefficient matrix between urban areas obtained by three methods

	Image Segmentation	Natural Break	Otsu
Image Segmentation	1.00	0.981	0.982
Natural Break	0.981	1.00	0.999
Otsu	0.982	0.999	1

Table 3. The total urban area percentages of the cities with the largest and smallest urban areas in Turkey

Rank	Image Segmentation		Natural Break		Otsu	
1	İSTANBUL	11.85	İSTANBUL	14.10	İSTANBUL	13.49
2	İZMİR	7.65	ANKARA	8.59	ANKARA	8.78
3	ANKARA	7.15	İZMİR	6.86	İZMİR	7.14
4	ANTALYA	4.75	KOCAELİ	4.71	KOCAELİ	4.55
78	GÜMÜŞHANE	0.16	ARTVİN	0.19	TUNCELİ	0.20
79	ŞIRNAK	0.11	GÜMÜŞHANE	0.19	ARTVİN	0.19
80	AMASYA	0.10	ŞIRNAK	0.15	ŞIRNAK	0.16
81	BARTIN	0.10	ARDAHAN	0.13	ARDAHAN	0.13

Natural Break method is examined. This high positive correlation between the two methods could also be understood when distribution of urban areas as percentage in Table 1 is analyzed. The distribution of urban areas give values close to each other according to both methods and the results obtained indicate a high proportion of overlap.

According to the three methods analyzed in Table 3, the cities with the largest and smallest urban areas have been identified. According to the results obtained from Natural Break and Otsu method, the first four provinces with the largest urban areas are respectively Istanbul, Ankara, Izmir and Kocaeli. The ranking of cities with the smallest urban areas are largely overlapping. According to these two methods, urban area values can also be seen to be close to each other. Image segmentation method differs a little in the other two methods.

Spatial autocorrelation relationship of the urban areas obtained from three methods has been assessed according to Moran I method. The relationship of the concept of spatial autocorrelation with geography was first expressed by Tobler (1970) in his words “Everything is related to everything else, but near things are more related than distant things.” It could also be describes as a positive or negative evaluation process of the spatial properties of the object depending on the distance. Spatial autocorrelation is expressed in three different ways (O’Sullivan & Unwin, 2002).

A positive correlation indicates that objects exhibiting similar properties cluster together. Negative correlation indicates that objects not exhibiting similar properties cluster together. Zero correlation indicates that there is no clustering between objects depending on spatial effects.

Considering Moran I values ranging from -1 to +1, it could be seen that there is a positive clustering in each method (Table 4). However, when the significance level 0.05 is taken into account, z value must be lower than -1.96 or higher than +1.96, therefore, the results obtained according to any method show a statistical significance.

This means that there is a positive spatial autocorrelation in the distribution of urban areas, namely it shows that large urban areas cluster together and small urban areas cluster together.

After the urban areas obtained through three methods, urban propagation map has been created. Urban propagation maps of three largest cities have been illustrated in Figures 3, 4, and 5 so that

Table 4. Moran I spatial autocorrelation statistical values

	Image Segmentation	Natural Break	Otsu
Moran’s Index	0.135	0.135	0.137
z-score	1.981	2.095	2.089
p-value	0.047	0.036	0.036

urban propagation map could be better understood and methods could be compared visually better.

FUTURE RESEARCH DIRECTIONS

In this study, urban areas were identified by classifying night images obtained by remote sensing methods according to pixel and object based classification method. At the same time, the methods used in the classification have been compared to each other statistically. Using higher resolution night scenes in future implementations of the study will contribute to the higher accuracy of urban areas. Moreover, this will give more precise results in Moran I values and the correlation used in the comparison of the classification methods.

CONCLUSION

In this study in which urban areas have been determined from night-time images by using pixel-based and object-based classification methods and these methods have been compared, the advantages and disadvantages tested have been discussed.

Identifying the threshold pixel value that divides the pixel-based image classification into two classes is the most important factor affecting the accuracy of the classification process. To determine the threshold value, in-class and inter-class variance values which are statistical parameters have been calculated. In addition, the appropriate threshold value has been determined by utilizing the histogram.

The basic principle of pixel-based Natural Break and Otsu methods used in our study is to determine a threshold value according to the distribution of the image pixels and to divide the image into two different classes with respect to this threshold value. Therefore, when examining the results obtained, it could be seen that the urban

area values obtained from both methods were very close together. The most important indicator of this is that the coefficient of correlation between the results of both methods is $r = 0.999$.

Working with occurring segments instead of pixels provide more information on the implications for thematic classes, but this raises the question of how information will be used effectively (Mallinis et al, 2008). In the object-based classification, classes of the working area can be easily created in the segmentation parameters identification level, namely in the segmentation process of the image. Through this approach, segments of pixels that have similar reflectance values are obtained for each class. The identification step of the pixel threshold value in the Pixel-based classification process is determined by the segmentation step of the object-based classification process. Although this case reveals a more accurate classification process and results, it causes the decision limits to be transitive.

The results obtained in the image segmentation method are shown to differ slightly from the other two methods. The reason of this is that the image after the segmentation process does not divide into two classes as in the pixel-based methods and the limits become transitive. Therefore, it requires user intervention in the creation of the precise boundaries of the urban areas. Despite this difference, the fact that correlation coefficients between the results of the image segmentation method and pixel-based methods are $r=0.981$ with Natural Break and $r=0.982$ with Otsu indicates that resulting values are highly correlated and alike.

In the results obtained from the Moran I spatial autocorrelation analysis of the investigated methods, urban areas have been found to spread spatially correlated. In this sense, it is concluded that major urban areas cluster around major urban areas, and small urban areas cluster around small urban areas.

Determination of Urban Growth by the Night-Time Images

Figure 3. Urban area of Istanbul according to three methods: (a) natural break, (b) image segmentation, (c) otsu

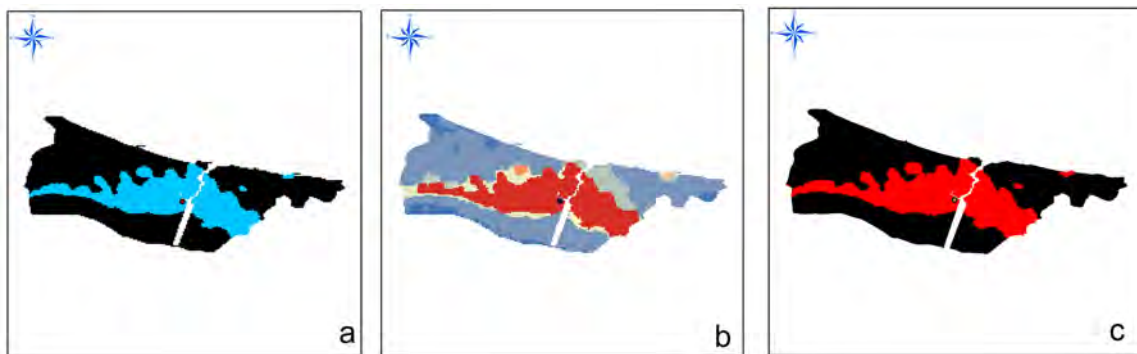


Figure 4. Urban area of Ankara according to three methods: (a) natural break, (b) image segmentation, (c) otsu

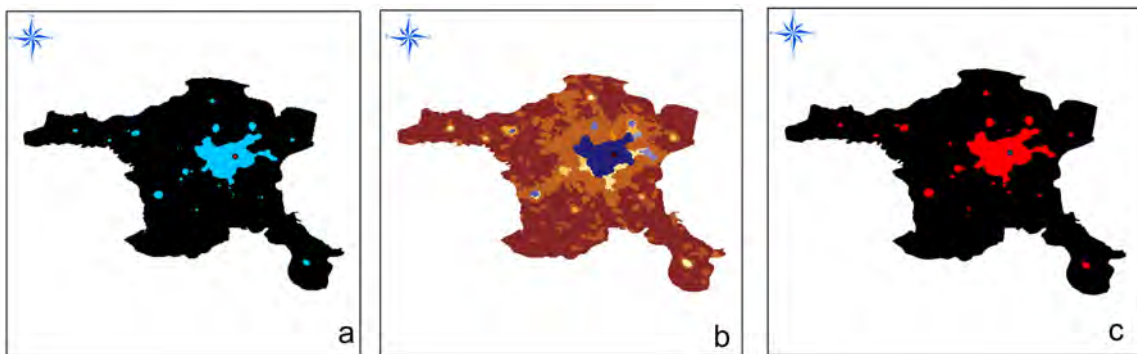
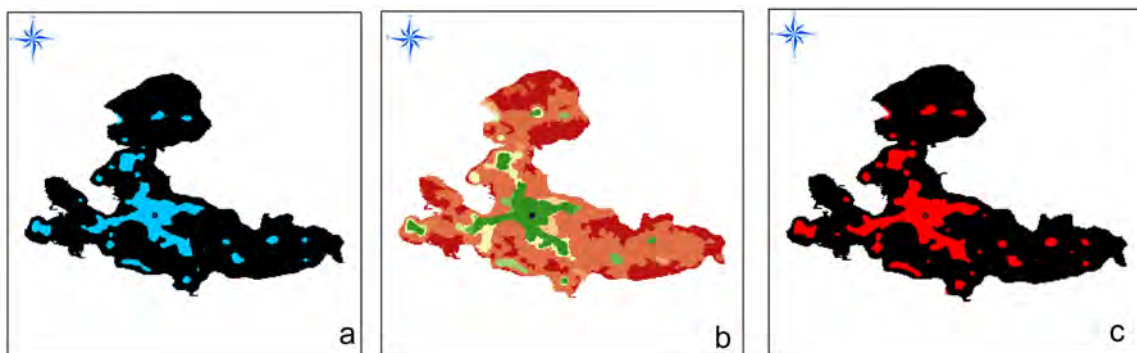


Figure 5. Urban area of Izmir according to three methods: (a) natural break, (b) image segmentation, (c) otsu



REFERENCES

- Atasoy, H. (2016). *Otsu Threshold Detection Method*. Retrieved February, 2016, from <http://www.atasoyweb.net/Otsu-Esik-Belirleme-Metodu>
- Aydın, İ. (2016). Firat University, Department of Computer Engineering. *Digital Image Processing and Image segmentation*. Retrieved February, 2016, from http://web.firat.edu.tr/iaydin/bmu357/bmu_357_bolum7.pdf
- Benz, U. C., Hofmann, P., Willhauck, G., Lingenfelder, I., & Heynen, M. (2004). Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS Journal of Photogrammetry and Remote Sensing*, 58(3–4), 239–258. doi:10.1016/j.isprsjprs.2003.10.002
- Elvidge, C. D., Erwin, E. H., Baugh, K. E., Ziskin, D., Tuttle, B. T., & Ghosh, T. (2009). Overview of DMSP night-time lights and future possibilities. Urban remote sensing event, 1–5.
- Henderson, M., Yeh, E. T., Gong, P., Elvidge, C., & Baugh, K. (2003). Validation of urban boundaries derived from global night-time satellite imagery. *International Journal of Remote Sensing*, 24(3), 595–609. doi:10.1080/01431160304982
- Imhoff, M. L., Lawrence, W. T., Stutzer, D. C., & Elvidge, C. D. (1997). A technique for using composite DMSP/OLS city lights satellite data to map urban area. *Remote Sensing of Environment*, 61(3), 361–370. doi:10.1016/S0034-4257(97)00046-1
- Levin, N., & Duke, Y. (2012, April). High spatial resolution night-time light images for demographic and socio-economic studies. *Remote Sensing of Environment*, 119(16), 1–10. doi:10.1016/j.rse.2011.12.005
- Liu, Z., He, C., Zhang, Q., Huang, Q., & Yang, Y. (2012, May 15). Extracting the dynamics of urban expansion in China using DMSP-OLS nighttime light data from 1992 to 2008. *Landscape and Urban Planning*, 106(1), 62–72. doi:10.1016/j.landurbplan.2012.02.013
- Mallinis, G., Koutsias, N., Tsakiri-Strati, M., & Karteris, M. (2008). Object-based classification using Quickbird imagery for delineating forest vegetation polygons in a Mediterranean test site. *ISPRS Journal of Photogrammetry and Remote Sensing*, 63(2), 237–250. doi:10.1016/j.isprsjprs.2007.08.007
- O’Sullivan, D., & Unwin, D. (2002). *Geographic Information Analysis*. John Wiley & Sons.
- Sunar, F., Özkan, C., & Osmanoğlu, B. (2011). *Uzaktan Algılama*. Eskişehir, Turkey: Publisher.
- Wei, Y., Liu, H., Song, W., Yu, B., & Xiu, C. (2014). Normalization of time series DMSP-OLS night-time light images for urban growth analysis with Pseudo Invariant Features. *Landscape and Urban Planning*, 128(August), 1–13. doi:10.1016/j.landurbplan.2014.04.015

KEY TERMS AND DEFINITIONS

Correlation: In statistics, dependence is any statistical relationship between two random variables or two sets of data. Correlation refers to any of a broad class of statistical relationships involving dependence, though in common usage it most often refers to the extent to which two variables have a linear relationship with each other.

Histogram: A graph showing the distribution of values in a set of data. Individual values are displayed along a horizontal axis, and the frequency of their occurrence is displayed along a vertical axis.

Moran’s I: In statistics, Moran’s I is a measure of spatial autocorrelation developed by Patrick Alfred Pierce Moran. Spatial autocorrelation is characterized by a correlation in a signal among nearby locations in space. Spatial autocorrelation is more complex than one-dimensional autocorrelation because spatial correlation is multi-dimensional and multi-directional.

Night-Time Images: Time series data of DMSP/OLS stable night-time light imagery span-

ning years 1992–2012 were derived from NOAA's National Geophysical Data Center. These image products provide gridded cell based annual cloud-free composited stable night-time lights with a digital number (DN) ranged from 0 to 63.

Object-Based Classification: Object-based image analysis, a technique used to analyze digital imagery, was developed relatively recently compared to traditional pixel-based image analysis. While pixel-based image analysis is based on the information in each pixel, object-based image analysis is based on information from a set of similar pixels called objects or image objects.

Pixel-Based Classification: The traditional pixel-based classification process automati-

cally categorizes all the pixels of an image into the thematic classes utilizing only the spectral information (relative reflectance) of each pixel in the image. The spectral information used by each pixel is also termed as a spectral signature, which is estimated by the relative reflectance in different wavelength bands.

Remote Sensing: Remote sensing is the acquisition of information about an object or phenomenon without making physical contact with the object and thus in contrast to on site observation.

Urban Area: An urban area is a location characterized by high human population density and many built environment features in comparison to the areas surrounding it.

Need for Rethinking Modern Urban Planning Strategies Through Integration of ICTs



Rounaq Basu

Indian Institute of Technology Bombay, India

Arnab Jana

Indian Institute of Technology Bombay, India

INTRODUCTION

Urban planners work towards achieving the three goals of environmental protection, economic development and social equity with the major focus being sustainable development (Campbell, 1996). Many practitioners feel that traditional urban practices have served their purpose, and it is time to search for a new approach which encourages openness and participation. However, conventional measures to enroll people in participatory planning activities have been found to be limited in both scope and reach. These findings coupled with recent advances in innovative technologies have led to opportunities to re-think the status quo in urban planning.

Information and Communications Technologies (ICT) have always been essential for new application opportunities. Recent progresses have paved the way for innovative communication services, and interactive models and tools for collaborative activities. Infrastructure can be made to dynamically and adaptively connect and collaborate in each other, thereby turning urban environments active and our cities into socio-technical organisms. Web 2.0, as suggested by O'Reilly (2007), has transformed the Web by turning it into a participatory platform where users can simultaneously consume (download) and share (upload) content. Web 2.0 techniques are linking users and data providers, thereby removing the barriers in direct communication that existed ear-

lier (Hudson-Smith & Crooks, 2008). Embedding these technologies into existing institutionalized processes still requires significant effort despite the collaborative and user-friendly nature of Web 2.0 services and applications.

ICT tools have the potential to radically impact modern urban planning strategies, thereby leading us into a new age of urban renaissance where we witness smart, digital, information and connected cities come to life. The applications of these tools vary widely, from a simple e-commerce service to complex data collection for decision making. These pervasive technologies have the power to impact dialogues between city managers and citizens, drive a shift to e-government services such as mobile health units and online payment solutions, and transform urban mobility through cloud communication and storage. The authors argue that there is potential for development of a new urban planning approach in which emphasis is given on active public participation through the use of Web 2.0 technologies. Thus, the role of urban planners needs to be redefined accordingly. Planners would benefit from integrating their current practices with modern ICT inventions to obtain socially inclusive long-term solutions. This chapter focuses on three of the most widely used ICT tools which have the potential for creating the highest impact in the field of urban planning: (1) Crowdsourcing, (2) Geographic Information Systems (GIS), and (3) Database Management Systems (DBMS).

DOI: 10.4018/978-1-5225-2255-3.ch682

The remainder of the article is organized as follows. The next section provides a detailed discussion about how crowdsourcing can be incorporated in the urban planning process. The associated challenges and benefits of its implementation have also been listed. Applications of a well-developed and vastly used tool in urban planning (GIS) are enumerated in Section 3, along with future research directions for further innovation. New ICTs have enabled us to collect and process large amounts of data. The importance of these tools (e.g. a DBMS or a web-based survey) is highlighted in Section 4. An application developed by the authors is also introduced in this section. Concluding remarks and future research directions are proposed in the final section.

BACKGROUND

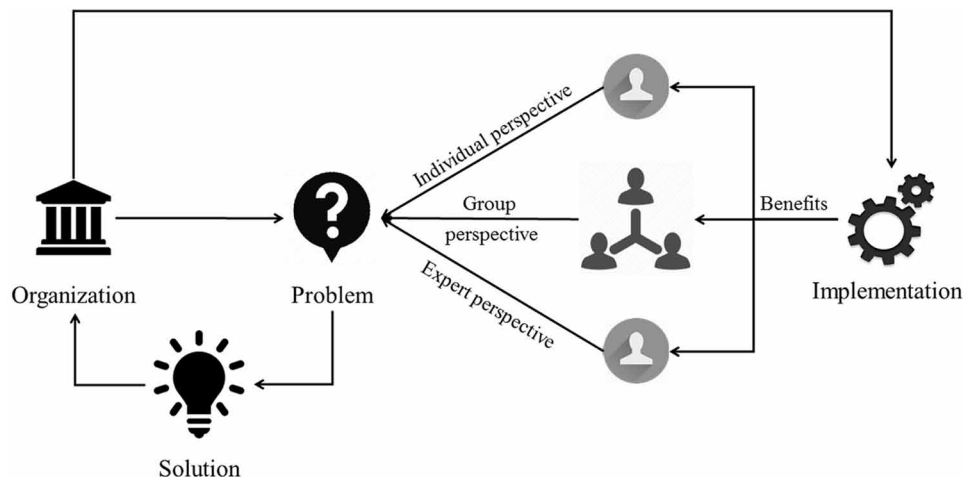
Crowdsourcing is a relatively new concept that incorporates a variety of services and activities. It can be characterized by any category of collaborative activity over the internet (Estellés-Arolas & González-Ladrón-de-Guevara, 2012). Crowdsourcing involves mechanisms that aim to leverage the collective intelligence of users for a productive outcome (Brabham, 2009). Jeffrey Howe coined the term ‘crowdsourcing’ in an

issue of *Wired Magazine* (Howe, 2006). Simply put, crowdsourcing harnesses the joint creativity of the public by creating an open call for proposals through Web 2.0 technologies in the hope of innovative and more robust solutions (Seltzer & Mahmoudi, 2013).

A classic example of crowdsourcing is Wikipedia due to its strong collaborative nature. Using the generalized crowdsourcing model shown in Figure 1, let us further analyze the Wikipedia example. The organization (Wikipedia.org) floats a topic and asks users to provide details about that topic (the problem). Due to its open-source nature where all users can freely edit the article, information is obtained from a variety of sources (individuals, groups and experts). Once all users are satisfied with the content on that topic (solution), Wikipedia finalizes that article and publishes it as a page (implementation). The entire world (not just those who contributed) can now benefit from the knowledge in that article. This is a continuous process, as more users can update or edit the information with progress of time, leading to continuous quality monitoring and improvement.

The crowdsourcing model asserts that online communities have ‘collective intelligence’ (Pierre, 1997) and ‘crowd wisdom’ (Surowiecki, 2004). It is quite intuitive that cognitively diverse communities would provide perspectives that a field

Figure 1. Generalized crowdsourcing model



expert might easily overlook. This was concretely proven by Van Herzele (2004) in the context of urban planning. Citizen participation and open innovation are essential tools for urban planners looking for fresh insights into their projects. It is quite evident that non-expert knowledge emerging from future beneficiaries of policies is indispensable to governance. However, urban planners need to give adequate thought to public involvement given the many inherent difficulties in the process. Brabham (2012) has found the crowdsourcing model to be appropriate for facilitating citizen participation in public planning projects. The crowdsourcing model has also proved to be useful in myriad urban planning applications such as:

- Citizen participation in public planning projects (Brabham, 2009).
- Collecting and sharing geocoded data (Hudson-Smith, Batty, Crooks, & Milton, 2009).
- Collecting pollution data using smart-phones (M. Stevens & D'Hondt, 2010).
- Environmental and public health surveillance as well as disaster informatics (Kamel Boulos et al., 2011).
- Identifying street accessibility problems (Hara, Le, & Froehlich, 2013).
- Calculating urban air temperatures from smartphone batteries (Overeem et al., 2013).
- Determining urban form and function (Crooks et al., 2015).
- Mapping urban emotion information for evaluation of ongoing planning processes (Resch, Summa, Sagl, Zeile, & Exner, 2015).

Extensive and systematic reviews of the literature focusing on crowdsourcing and related concepts have been attempted by Mandarano, Meenar, and Steins (2010) and Zhao and Zhu (2014). It has been concluded that this field of research is still emerging rather than established. Evans-Cowley (2010) too has written that the impact of such

technologies on individuals, communities and cities is yet to reach its full potential.

U

Challenges

In aiming to obtain solutions from the masses, crowdsourcing opens itself up to face several challenges in the path to successful implementation of the process. The authors provide a summary of the major issues plaguing the use of crowdsourcing as a tool for urban planning. The reader is invited to refer to Howe (2009), Evans-Cowley (2011), and Pedersen et al. (2013) for further details.

Challenges in System Governance

- **Selection of the Right Model:** Multiple forms of crowdsourcing exist and it is important that the form is selected according to the objective.
- **Selection of the Right Crowd:** Attracting and sustaining the crowd is essential for multiple iterations, which leads to a robust solution.
- **Definition of the Problem Statement:** The problem statement should be well-defined and simplified, along with decomposition and distribution of different steps for a more efficient process.
- **Providing the Right Incentives:** Fair incentives help in attraction and retention of participatory populace. It was found that “each new prize should have approximately twice higher marginal utility than the prize immediately above it” (Archak & Sundararajan, 2009).
- **Selection of the Best Solution:** Individual ideas might be the best solution for each individual sub-task. Therefore, a mechanism should be in place for identifying sub-solutions which can be integrated together to obtain the most optimal solution. It is important to remember that such open calls lead to a veritable flood of responses, most of which might not be related or useful to the problem.

Challenges in Implementation

- **Bridging the Digital Divide:** Digital seclusion, i.e. inability to access the online world, remains a harsh reality in most developing countries.
- **Need for State-of-the-Art Infrastructure:** A sound technical support system is indispensable for maximum participation and user retention.
- **Cost:** Crowdsourcing requires time, money and constant monitoring.
- **Providing User Feedback:** Users need constant feedback about how the entire process is functioning and how their suggestions are being used. Providing feedback might aid in encouraging people to participate multiple times, which is a challenge often mentioned in literature.

Benefits

Despite the challenges that exist in applying crowdsourcing to urban planning, there are several positives as well. The authors next discuss briefly about a few of the benefits that crowdsourcing has provided to the research and planning communities. The reader is invited to refer to Brabham (2009) and Janssen, Charalabidis, and Zuiderwijk (2012) for further details.

- **Network Democracy:** The social order is moving towards formation of organized networks that co-exist with (and are thereby enhanced by) digital technologies (Rossiter, 2006). Crowdsourcing activities enable the public to feel as if the solutions have been made by them, i.e. a home-grown or an organic solution, and promote community participation and involvement.
- **Collective Intelligence:** It is a “form of universally distributed intelligence, constantly enhanced, coordinated in real time, and resulting in the effective mobilization of skills” (Pierre, 1997). The Web provides

the perfect platform for harnessing this intellect for working towards a networked democratic process (Rossiter, 2006).

- **Crowd Wisdom:** Surowiecki (2004) finds that “under the right circumstances, groups are remarkably intelligent, and are often smarter than the smartest people in them”. Thus, it is more desirable to collect the aggregate solution, rather than the average one.
- **Political and Social Benefits:** These include greater transparency, democratic accountability, public engagement, innovative social services and improvement of citizen satisfaction and policy-making processes.
- **Operational and Technical Benefits:** These include optimization of administrative processes, access to external problem-solving capacity, fair decision-making, improvement of public policies, greater ease in accessing data, and sustainability of data.
- **Economic Benefits:** These include stimulation of innovation, economic growth, and availability of information for investors.

GIS: USED BUT NOT SPENT

Introduction

Finding community-driven solutions for local problems is enabled by crowdsourcing; however, spatial factors often come into play while analyzing urban planning issues. Franklin and Hane (1992) found that 80% of data contains geo-referenced information. Therefore, a spatial analysis tool is required for data processing and visualization of patterns and trends, which aids in fast and better identification of underlying problems in the study area. Developed in the late 1960s, GIS can serve both as a database and a toolbox for urban planning. Barriers for initial use by planning agencies include limited capabilities of the software, lack

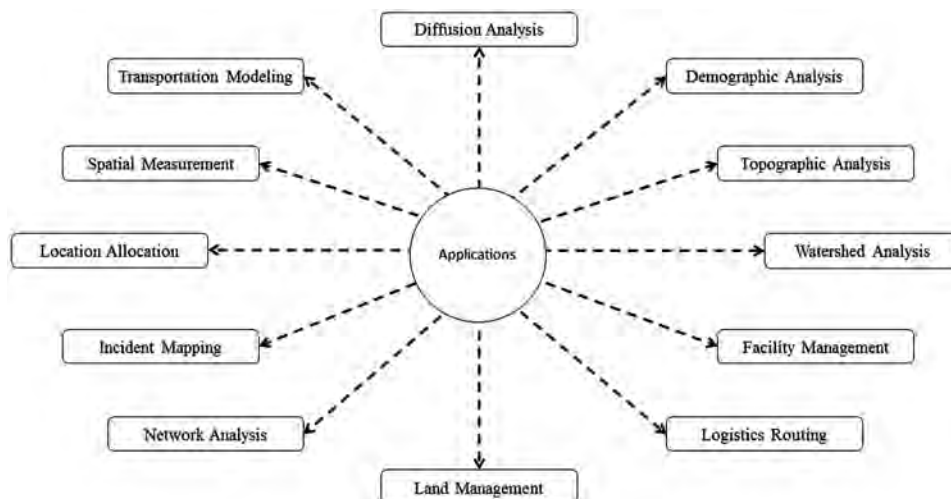
of skilled manpower and the extremely high cost of hardware. With advancements in technology leading to major improvements in both software and hardware performance, the use of GIS has become quite vital for urban planning. Visualization of databases and spatial modeling are the main uses of GIS in urban planning (Yeh, 1999). It has been pointed out by Douven, Grothe, Nijkamp, and Scholten (1993) that a major weakness of GIS in this sector is their integration with urban planning models.

Webster (1993) argues that the usefulness of GIS as a tool for planning can be best assessed with reference to the nature of the required data inputs. The potential contribution of GIS in predictive and prescriptive analyses for urban planning is discussed by Webster (1994). GIS technology offers opportunities to improve the quality of planning decisions through its support of both formal and informal analysis. The use of GIS over online forums can enhance citizen participation in the planning process by providing a platform for effective dissemination of information related to urban planning. GIS can be used for a wide array of diverse applications, some of which are outlined in Figure 2.

A few applications of GIS for urban planning over the years have been enumerated below:

- Open space planning for allocation of public facilities (Yeh & Chow, 1996).
- Evaluating neighborhood pedestrian accessibility (Aultman-Hall, Roorda, & Baetz, 1997).
- Modeling sustainable urban development (Li & Yeh, 2000).
- Simulating realistic urban dynamics for evaluation of development alternatives (Yeh & Li, 2003).
- Modeling the pattern and extent of urban sprawl (Sudhira, Ramachandra, & Jagadish, 2004).
- Evaluating urban expansion and land use change (Xiao et al., 2006).
- Hedonic price modeling of the amenity value of urban green space (Kong, Yin, & Nakagoshi, 2007).
- Predictive modeling of urban growth (Stevens, Dragicevic, & Rothley, 2007).
- Analyzing spatial patterns in metropolitan areas (Henderson, 2008).
- Determination of urban green space accessibility (Comber, Brunsdon, & Green, 2008).
- Creating a framework for landfill siting in a fast-growing urban region (Chang, Parvathinathan, & Breeden, 2008).

Figure 2. Diverse applications of GIS



Challenges

Technological advances with regard to software and hardware do not solely define the use of GIS in urban planning. The authors discuss briefly about some of the major challenges in this area, with particular regard to developing countries. The reader is invited to refer to Yeh (1999) for additional details.

- **Organization:** Separate studies have shown that factors related to the organizational structure and available staff are far more important when compared to technology for effective applications of GIS (Campbell, 1994; Maguire, Campbell, & Masser, 1997). Successful implementation of GIS requires (a) a dedicated information management strategy, (b) committed participation in the implementation of ICTs, and (c) a high degree of organizational stability.
- **Data:** Geographical data is readily available in developed countries, albeit at an expensive rate sometimes. However, developing countries face acute difficulties in collecting substantial data. Major inhibiting factors include (a) lack of financial resources, (b) lack of trained personnel, and (c) unavailability of modern data processing resources. Quality of data is an associated critical issue since data in developing countries are generally collected by external non-governmental agencies, which leads to a dearth of measures for quality checks (Batty, 1993).

Future Research Directions

Techniques for collection and use of data for participatory planning have seen major improvements over the past decade. However, their effective incorporation into GIS remains a goal to strive for (Dennis, 2006). Since GIS has been extensively used by researchers and planners all

over the globe, little remains to be explored in the technical domain. Therefore, formation of a community-based qualitative GIS should serve as a possible line of inquiry in the future. The major obstacles to this goal include the lack of community involvement and the absence of participation of external stakeholders or investors. Issues of local government and social service delivery have yet to be successfully handled through GIS, despite its widespread use by urban planners and policy makers. Since data pertaining to most urban studies are found in the form of maps, it is critical that modern mapping services are developed with integration of Web 2.0 tools in order to augment the visualization capacity of GIS models (Elwood, 2006). Web-based public participation GIS (PPGIS) practices were found to a valuable approach for fostering community engagement and promoting direct communication among users and decision-makers (Bugs, Granell, Fonts, Huerta, & Painho, 2010). It is evident that advances in information technology need to be integrated with existing methods like GIS to build a robust planning framework. Such a framework would also assist in assessing the desirability and suitability of alternative development plans.

DBMS: A BIG LEASH FOR BIG DATA

Introduction

As discussed in the previous section, one of the major challenges associated with GIS is the large amount of data required. Traditional means of data acquisition in such activities (on-site and photograph-based surveys) have severe disadvantages such as high costs or low participation rates. Developing countries often have to look for cost-effective alternatives, sacrificing quality for cost in the process. It has been found that web-based surveys have a higher response rate as compared to paper-based surveys (Kaplowitz, Hadlock, & Levine, 2004). Therefore, collecting such data within short durations at minimum cost

has now become possible through development of web-based surveys coupled with the use of DBMS. Applications of such surveys range from assessing built environment and visual landscape to monitoring the environmental impacts of urban sprawls and newly built transit networks. Reips (2002) provides a detailed list of the advantages (time and cost savings) and disadvantages (misunderstanding the questionnaire) of conducting online surveys.

Availability of vast amounts of data on a national scale, colloquially known as Big Data, is productive for understanding urban growth. Big Data facilitates flows of information and mechanisms of better understanding of heterogeneous individuals. Concepts such as ‘*smart urbanism*’ and ‘*real-time cities*’ are based on applications of Big Data in urban planning. Kitchin (2014) pinpoints emerging concerns in this regard such as the politics of big urban data, and corporatization of city governance and technological lock-ins. Batty (2013) describes how the emphasis is shifting from short-term measures for city management to long-term strategic planning due to the burgeoning growth of Big Data. Our capabilities to harness and analyze this kind of Big Data will lead to capturing information about every time horizon over longer periods of time.

DBMS in Developing Countries: Example

Although these concepts are used quite extensively in developed countries, developing countries are yet to embrace them for use in academic and industrial activities. The authors present an online survey technique packed with a latent DBMS structure that has been developed for collecting Big Data corresponding to citizen sciences, with a special focus on transportation, activity-time use and health care behavior.

Named ‘*Data Repository for Urban Models & Systems*’ (abbreviated as DRUMS), it provides complete freedom to the user (the academic institution in this case) to modify or delete existing

surveys, and even add new ones without any further efforts for writing new code. Although there are services such as SurveyMonkey, Typeform and Google Forms available on the internet, there are constraints such as: (a) limited number of free questions, (b) limited type and design of questions, and (c) data anonymity issues. Therefore, DRUMS came into being as a home-grown data collection and storage service that is expected to be a worthy alternative to similar paid services. DRUMS is currently hosted at this link: <http://www.cuse.iitb.ac.in/drums/>. The major advantage of DRUMS is that it enables citizen participation in urban planning approaches (see Figure 3). By linking end users with system developers, different surveys can be designed, modified and hosted to collect data with different exploratory questions in mind. All the collected data is stored in the database which is accessible only by the system administrator. In an effort to enhance the quality and use of open data, end users are encouraged to contact the system administrator in order to gain access to the data collected from their designed surveys. Citizens can thus perform exploratory analyses about the current community status in order to identify problems and propose solutions, all without the mandatory need of technical expertise (which is provided by DRUMS free of cost).

The benefits that DRUMS provides over manual methods of surveying are as follows:

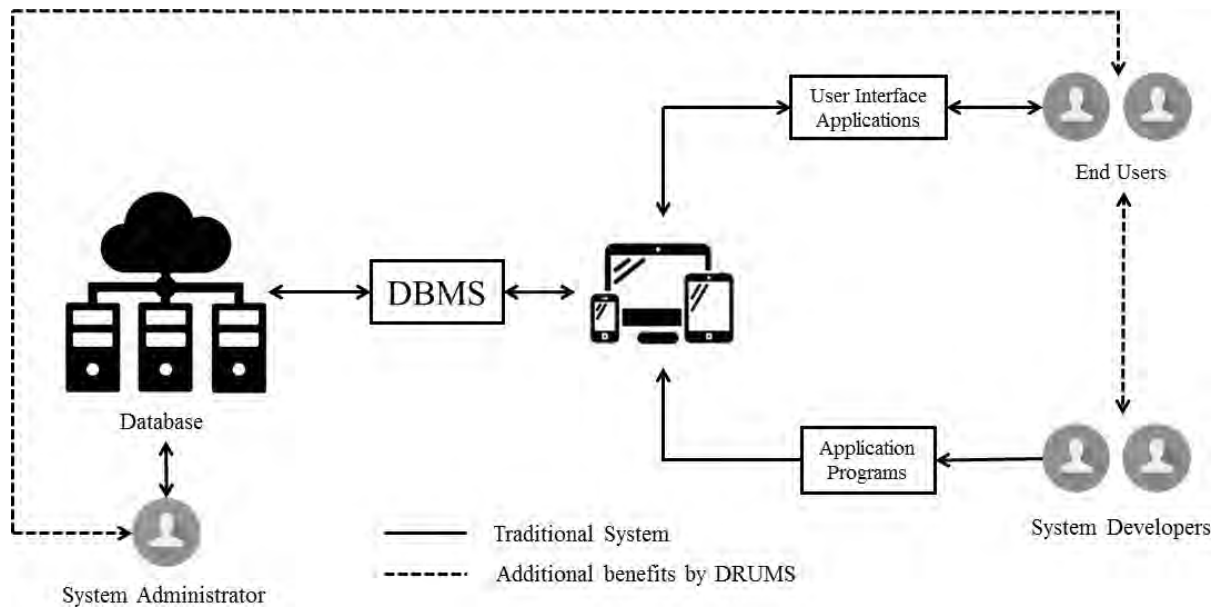
Surveyor Perspective

- Can design and host multiple surveys without writing code (using developer tools).
- Can use sets of pre-defined question types (Multiple-choice, Dropdown, Boolean, Text-Based, etc.).

Respondent Perspective

- Can switch back to previous sections without loss of data (choices get locked in).
- Can close the browser tab which has the survey and reopen it without loss of data.

Figure 3. Traditional DBMS systems vis-à-vis DRUMS



- No loss of data even if device is disconnected from network.
- Can fill in data for multiple modes without filling in individual's data multiple times.

System Perspective

- Extremely quick collection of large amounts of data (Single sample of Survey 1 takes 10 – 15 mins).
- Region-independent data collection (city, state, or even country level).
- Highly cost-effective (only server maintenance expenses).

Such methods of data collection pose several challenges, especially in developing countries like India. Despite its multiple benefits, DRUMS has a few drawbacks like all of the previously discussed ICT applications. They are as follows:

- **Failure to Capture Varied Socio-Economic Categories of Population:** The socio-economic heterogeneity of the population is the major obstacle contributing

to the inability of the survey to penetrate all segments of the population and thereby hindering collection of a statistically representative sample. Although DRUMS has been tested only in India, the authors expect this concern to remain valid for most developing countries.

- **Dependent on Infrastructure:** Online surveys are largely dependent on technological infrastructure. Potential respondents are unable to access the survey during server downtime (due to network or maintenance issues) leading to loss of such samples because most people do not return to fill the survey at a later stage.
- **Knowledge Seclusion:** Several segments of the society, especially the economically weaker sections and the low-income groups, lack understanding of ICT applications despite owning and operating smartphones.
- **Digital Seclusion:** There are other groups which do not have steady access to the internet and are thereby excluded from participating in such data collection activities.

FUTURE RESEARCH DIRECTIONS

Future work should focus on exploiting the benefits of ideas generated by multiple users. Better and more useful organization of the feedback for urban planning also needs to be explored. Besides, future research questions may center on trust and reputation issues. In conclusion, the authors argue that integration of urban policies, modern technologies and fundamental concepts of engineering will lead to discovery of new solutions to important age-old urban problems.

CONCLUSION

This chapter explored different aspects of ICTs with regard to their application in urban planning. The authors believe that a paradigm shift in urban planning may be required with modern advances enabling collection of vast amounts of high-resolution data, faster analysis, and formulation of more complex and robust models. Citizen participation and public involvement were quite difficult earlier, leading to urban planning policies being solely in the hands of the government or municipal organization. ICTs have entitled citizens to feel more involved by providing them opportunities to have a say in the planning process. Since the impacts will have to be borne by citizens, it is only fair that they be given a fair opportunity to shape these policies. The authors believe that integration of traditional planning methodologies with modern Web 2.0 tools will significantly strengthen civic participation and will eventually empower citizens.

These tools and technologies, however, are not as easy to implement as is proposed. While developed countries have an abundance of resources in these sectors, developing countries have yet to make a complete leap to information technology for data collection and management purposes. However, the reluctance in handling and using complex ICT systems is slowly fading, as there are numerous examples of self-designed systems currently in use.

REFERENCES

- Archak, N., & Sundararajan, A. (2009). Optimal design of crowdsourcing contests. *ICIS 2009 Proceedings*, 200.
- Aultman-Hall, L., Roorda, M., & Baetz, B. (1997). Using GIS for Evaluation of Neighborhood Pedestrian Accessibility. *Journal of Urban Planning and Development*, 123(1), 10–17. doi:10.1061/(ASCE)0733-9488(1997)123:1(10)
- Batty, M. (1993). Using geographic information systems in urban planning and policy-making. In M. M. Fischer & P. Nijkamp (Eds.), *Geographic Information Systems, Spatial Modelling and Policy Evaluation* (pp. 51–69). Berlin: Springer Berlin Heidelberg. doi:10.1007/978-3-642-77500-0_4
- Batty, M. (2013). Big data, smart cities and city planning. *Dialogues in Human Geography*, 3(3), 274–279. doi:10.1177/2043820613513390
- Boulos, , Resch, Crowley, Breslin, Sohn, Burtner,... Chuang. (2011). Crowdsourcing, citizen sensing and sensor web technologies for public and environmental health surveillance and crisis management: Trends, OGC standards and application examples. *International Journal of Health Geographics*, 10(1), 1–29. doi:10.1186/1476-072X-10-67 PMID:21214924
- Brabham, D. C. (2009). Crowdsourcing the Public Participation Process for Planning Projects. *Planning Theory*, 8(3), 242–262. doi:10.1177/1473095209104824
- Brabham, D. C. (2012). Motivations for Participation in a Crowdsourcing Application to Improve Public Engagement in Transit Planning. *Journal of Applied Communication Research*, 40(3), 307–328. doi:10.1080/00909882.2012.693940
- Bugs, G., Granell, C., Fonts, O., Huerta, J., & Painho, M. (2010). An assessment of Public Participation GIS and Web 2.0 technologies in urban planning practice in Canela, Brazil. *Cities (London, England)*, 27(3), 172–181. doi:10.1016/j.cities.2009.11.008

- Campbell, H. (1994). How effective are GIS in practice? A case study of British local government. *International Journal of Geographical Information Systems*, 8(3), 309–325. doi:10.1080/02693799408902002
- Campbell, S. (1996). Green Cities, Growing Cities, Just Cities?: Urban Planning and the Contradictions of Sustainable Development. *Journal of the American Planning Association*, 62(3), 296–312. doi:10.1080/01944369608975696
- Chang, N.-B., Parvathinathan, G., & Breeden, J. B. (2008). Combining GIS with fuzzy multicriteria decision-making for landfill siting in a fast-growing urban region. *Journal of Environmental Management*, 87(1), 139–153. doi:10.1016/j.jenvman.2007.01.011 PMID:17363133
- Comber, A., Brunsdon, C., & Green, E. (2008). Using a GIS-based network analysis to determine urban greenspace accessibility for different ethnic and religious groups. *Landscape and Urban Planning*, 86(1), 103–114. doi:10.1016/j.landurbplan.2008.01.002
- Crooks, A., Pfoser, D., Jenkins, A., Croitoru, A., Stefanidis, A., Smith, D., & Lamprianidis, G. et al. (2015). Crowdsourcing urban form and function. *International Journal of Geographical Information Science*, 29(5), 720–741. doi:10.1080/13658816.2014.977905
- Dennis, S. F. (2006). Prospects for Qualitative GIS at the Intersection of Youth Development and Participatory Urban Planning. *Environment & Planning A*, 38(11), 2039–2054. doi:10.1068/a3861
- Douven, W., Grothe, M., Nijkamp, P., & Scholten, H. (1993). Urban and regional planning models and GIS. *NATO ASI Series D Behavioural and Social Sciences*, 70, 317–317.
- Elwood, S. (2006). Critical Issues in Participatory GIS: Deconstructions, Reconstructions, and New Research Directions. *Transactions in GIS*, 10(5), 693–708. doi:10.1111/j.1467-9671.2006.01023.x
- Estellés-Arolas, E., & González-Ladrón-de-Guevara, F. (2012). Towards an integrated crowdsourcing definition. *Journal of Information Science*, 38(2), 189–200. doi:10.1177/0165551512437638
- Evans-Cowley, J. (2010). Planning in the Real-Time City: The Future of Mobile Technology. *Journal of Planning Literature*, 25(2), 136–149. doi:10.1177/0885412210394100
- Evans-Cowley, J. S. (2011). *Crowdsourcing the curriculum: public participation in redesigning a planning program*. Available at SSRN 1760525
- Franklin, C., & Hane, P. (1992). An Introduction to Geographic Information Systems: Linking Maps to Databases [and] Maps for the Rest of Us: Affordable and Fun. *Database*, 15(2), 12–15.
- Hara, K., Le, V., & Froehlich, J. (2013). *Combining crowdsourcing and google street view to identify street-level accessibility problems*. Paper presented at the SIGCHI Conference on Human Factors in Computing Systems, Paris, France. doi:10.1145/2470654.2470744
- Henderson, M. (2008). Mapping Global Cities: GIS Methods in Urban Analysis by Ayşe Pamuk and Quantitative Methods and Applications in GIS by Fahui Wang. *Journal of Regional Science*, 48(2), 474–478. doi:10.1111/j.1467-9787.2008.00559_9.x
- Howe, J. (2006). The Rise of Crowdsourcing. *Wired*. Retrieved July 20, 2016, from <http://www.wired.com/2006/06/crowds/>
- Howe, J. (2009). *Crowdsourcing: How the power of the crowd is driving the future of business*. New York: Three Rivers Press.
- Hudson-Smith, A., Batty, M., Crooks, A., & Milton, R. (2009). Mapping for the Masses: Accessing Web 2.0 Through Crowdsourcing. *Social Science Computer Review*, 27(4), 524–538. doi:10.1177/0894439309332299

- Hudson-Smith, A., & Crooks, A. (2008). *The renaissance of geographic information: Neogeography, gaming and second life*. Academic Press.
- Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management*, 29(4), 258–268. doi:10.1080/10580530.2012.716740
- Kaplowitz, M. D., Hadlock, T. D., & Levine, R. (2004). A Comparison of Web and Mail Survey Response Rates. *Public Opinion Quarterly*, 68(1), 94–101. doi:10.1093/poq/nfh006
- Kitchin, R. (2014). The real-time city? Big data and smart urbanism. *GeoJournal*, 79(1), 1–14. doi:10.1007/s10708-013-9516-8
- Kong, F., Yin, H., & Nakagoshi, N. (2007). Using GIS and landscape metrics in the hedonic price modeling of the amenity value of urban green space: A case study in Jinan City, China. *Landscape and Urban Planning*, 79(3–4), 240–252. doi:10.1016/j.landurbplan.2006.02.013
- Li, X., & Yeh, A. G.-O. (2000). Modelling sustainable urban development by the integration of constrained cellular automata and GIS. *International Journal of Geographical Information Science*, 14(2), 131–152. doi:10.1080/136588100240886
- Maguire, D., Campbell, H., & Masser, I. (1997). *GIS and Organisations: How Effective Are GIS in Practice?* JSTOR.
- Mandarano, L., Meenar, M., & Steins, C. (2010). Building Social Capital in the Digital Age of Civic Engagement. *Journal of Planning Literature*, 25(2), 123–135. doi:10.1177/0885412210394102
- O'Reilly, T. (2007). What is Web 2.0: Design patterns and business models for the next generation of software. *Communications & Strategies*, (1): 17.
- Overeem, A. R., Robinson, J. C., Leijnse, H., Steeneveld, G. J. P., Horn, B. K., & Uijlenhoet, R. (2013). Crowdsourcing urban air temperatures from smartphone battery temperatures. *Geophysical Research Letters*, 40(15), 4081–4085. doi:10.1002/grl.50786
- Pedersen, J., Kocsis, D., Tripathi, A., Tarrell, A., Weerakoon, A., Tahmasbi, N.,... de Vreede, G. J. (2013). *Conceptual Foundations of Crowdsourcing: A Review of IS Research*. Paper presented at the System Sciences (HICSS), 2013 46th Hawaii International Conference on.
- Pierre, L. (1997). *Collective Intelligence: Mankind's Emerging World in Cyberspace*. Cambridge, MA: Perseus Books.
- Reips, U.-D. (2002). Standards for Internet-based experimenting. *Experimental Psychology*, 49(4), 243–256. doi:10.1026//1618-3169.49.4.243 PMID:12455331
- Resch, B., Summa, A., Sagl, G., Zeile, P., & Exner, J.-P. (2015). Urban Emotions—Geo-Semantic Emotion Extraction from Technical Sensors, Human Sensors and Crowdsourced Data. In G. Gartner & H. Huang (Eds.), *Progress in Location-Based Services 2014* (pp. 199–212). Cham: Springer International Publishing. doi:10.1007/978-3-319-11879-6_14
- Rossiter, N. (2006). *Organized networks*. Rotterdam: NAI Publishers.
- Seltzer, E., & Mahmoudi, D. (2013). Citizen Participation, Open Innovation, and Crowdsourcing: Challenges and Opportunities for Planning. *Journal of Planning Literature*, 28(1), 3–18. doi:10.1177/0885412212469112
- Stevens, D., Dragicevic, S., & Rothley, K. (2007). iCity: A GIS-CA modelling tool for urban planning and decision making. *Environmental Modelling & Software*, 22(6), 761–773. doi:10.1016/j.envsoft.2006.02.004

- Stevens, M., & D'Hondt, E. (2010). *Crowdsourcing of pollution data using smartphones*. Paper presented at the Workshop on Ubiquitous Crowdsourcing.
- Sudhira, H. S., Ramachandra, T. V., & Jagadish, K. S. (2004). Urban sprawl: Metrics, dynamics and modelling using GIS. *International Journal of Applied Earth Observation and Geoinformation*, 5(1), 29–39. doi:10.1016/j.jag.2003.08.002
- Surowiecki, J. (2004). *The wisdom of crowds: why the many are smarter than the few and how collective wisdom shapes business, economics, society and nations*. Little, Brown.
- Van Herzele, A. (2004). Local Knowledge in Action: Valuing Nonprofessional Reasoning in the Planning Process. *Journal of Planning Education and Research*, 24(2), 197–212. doi:10.1177/0739456X04267723
- Webster, C. J. (1993). GIS and the Scientific Inputs to Urban Planning. Part 1: Description. *Environment and Planning. B, Planning & Design*, 20(6), 709–728. doi:10.1068/b200709
- Webster, C. J. (1994). GIS and the Scientific Inputs to Planning. Part 2: Prediction and Prescription. *Environment and Planning. B, Planning & Design*, 21(2), 145–157. doi:10.1068/b210145
- Xiao, J., Shen, Y., Ge, J., Tateishi, R., Tang, C., Liang, Y., & Huang, Z. (2006). Evaluating urban expansion and land use change in Shijiazhuang, China, by using GIS and remote sensing. *Landscape and Urban Planning*, 75(1–2), 69–80. doi:10.1016/j.landurbplan.2004.12.005
- Yeh, A. G.-O. (1999). Urban planning and GIS. *Geographical Information Systems*, 2, 877–888.
- Yeh, A. G.-O., & Chow, M. H. (1996). Decision Support Systems in Urban Planning: An integrated GIS and location-allocation approach to public facilities planning—An example of open space planning. *Computers, Environment and Urban Systems*, 20(4), 339–350. doi:10.1016/S0198-9715(97)00010-0
- Yeh, A. G.-O., & Li, X. (2003). Simulation of Development Alternatives Using Neural Networks, Cellular Automata, and GIS for Urban Planning. *Photogrammetric Engineering and Remote Sensing*, 69(9), 1043–1052. doi:10.14358/PERS.69.9.1043
- Zhao, Y., & Zhu, Q. (2014). Evaluation on crowdsourcing research: Current status and future direction. *Information Systems Frontiers*, 16(3), 417–434. doi:10.1007/s10796-012-9350-4

KEY TERMS AND DEFINITIONS

Big Data: Vastly large data sets that cannot be analyzed through traditional data processing techniques are collectively known as Big Data. Organizations in possession of Big Data (majorly business organizations) are attempting to capture and analyze even more samples in order to make sound decisions based on human behavior.

Crowdsourcing: The technique of presenting a problem statement to the public and floating an open call for solution proposals is known as crowdsourcing. The major advantage it possesses is that it uses modern Web technologies and tools for circulating the problem and collecting responses.

DBMS: Database management systems (DBMS) serve as an interface between the database and users, helping to ensure that data is organized in an easily accessible and comprehensible structure.

GIS: A geographical information system (GIS) is a computer system that enables storage and display of different varieties of spatial data so that underlying trends and patterns can be identified through analysis.

ICT: Information and communications technologies (ICTs) include any software, middleware, hardware or system that enables the transfer of information in an electronic or digital form.

Urban Governance: It can be defined as the system of processes through which a city is managed on a short-term scale, and the long-term development agendas are formulated and realized.

Urban Planning: It is a techno-political process that deals with the development of land, environment and public welfare along with design of relevant infrastructure, such as transportation and communications, which form the urban environment.

U

Reconstructive Architectural and Urban Digital Modelling

Roberta Spallone

Politecnico di Torino, Italy

INTRODUCTION

The digital reconstruction of contemporary architectural and urban complexes which were demolished, transformed or have been only theatrically conceived, remaining ‘on paper’, is now a tool of considerable heuristic value, allowing to preserve, interpret and create new images of cultural heritages that no longer exist in their original shape or never reached a material construction.

This chapter reviews methods and techniques of digital reconstruction, through an overview of the international theoretical statements, and then the author explores the potentials of three-dimensional reconstruction and the prerogatives of animation in the interaction with three-dimensional data.

The examination of several international case studies, and also some experiences personally conducted highlights the different strategies used for the preservation of the memory of such heritage.

BACKGROUND

The theme of the digital reconstruction of architectural and urban complexes is currently the subject of interest for many scholars and a field of convergence of different disciplinary viewpoints ranging from history of architecture, urban studies and environmental psychology.

This chapter privileges the point of view that takes as the central moment of the reconstruction the creation of interpretative digital models of a series of knowledge assumed through interdisciplinary researches. The relationship between architectural and urban dimensions is thought

seamless: the architectural complex, object of study, must be related to the urban context in which it was designed or it spent a special time in its life or it could live today, and the reciprocal transformations must be carefully recorded. In this way the object itself takes on an urban meaning. There could be several aims of reconstruction that, in this research ambit, goes beyond the mere iconographic documentation, to provide scientifically plausible readings of the analysed phenomena.

Several reconstruction techniques are inspired by Pagnano, Docci and Albisinni’s theories and methodologies of graphical analysis, applied to the design drawing as well to realize buildings, thanks to which the drawing can be identified with the model of understanding (Pagnano, 1975; Docci, 2009; Albisinni, 2011).

The scholars, adopting this method, re-draw and, or, re-model the space and use the graphic model as scientists reproducing artificially a certain phenomenon in the laboratory.

Reconstructive 3D modeling allows reviving an architecture and its relationships with the context, whether it is partly, or totally, lost, or hidden in the body of a building heavily altered, or, finally, it is going to be irreversibly transformed.

Representation and communication concern the knowledge data and their interpretation, resulting from the documentation phases. The digital models are the most useful database for collecting and synthesizing these analyses.

The modeling is therefore a cognitive strategy in which the idea of similarity, in comparison with the reality, plays a decisive role, a strategy that is utilized in different ways depending on which kind of model it is going to carry out (Gaiani, 2004)

There are several stages in the 3D reconstruction of a built complex. These include the gathering of the source information, the interpretation of this information, the comparison with contemporaneous examples, the development of 2D blueprints and/or 3D geometric (or parametric) models of the building and its context, the texture mapping, the addition of lights and, finally, the rendering.

Transforming several drawings into a digital model is a process which changes one model into another and deserves some attention. It is, as a matter of fact, not a simple variation without alteration of contents, but, on the contrary, it modifies, from time to time, “the wealth of the model, its expressive potentiality. In fact, the transmutations of the models are moved by the interpretative intent of the scholar, and they converge, therefore, toward an abstract model... that we can identify in the project idea” (Migliari, 2004).

The creation of a 3D digital model offers, as a result, infinite possibilities of observation and survey: from the objective visualization of a cylindrical projection, orthographic or isometric, to that subjective of a conical projection, perspective (Spallone, 2007). The model thus becomes an essential tool to check and control the validity of reconstructive hypotheses and the congruence between the building elements.

Moreover 3D computer models allow more enhanced and controlled interaction between users and models; they are able to cover the whole range of possible models in a single system of representation (Maldonado, 2005).

Conceived as real maquettes, which live in a virtual space perfectly corresponding to the real one, so much to offer all the four dimensions, the models carried out by means of computer are then observed through a screen..., with capacities to vary the point of view in order to simulate the mobility and the transformability in the time and in the appearance. (Gaiani, 2004)

In the operational strategy that ascertains the models attributes, a central moment is the

definition of the ratio between representation/visualization scale and modeling details. Since all the models are simplifications and schematizations of the reality, there is always a difference between the real and the details that are included in the model.

Even a very detailed graphic reconstruction, of an architectural product, aimed at giving the viewer a stimulating image of reality, will always necessarily be the result of a synthesis... A synthesis always implies the exploitation of specific aspects to the detriment of others and consequently a data loss. (Galifi, Moretti, & Aoyagi, 2002)

The increasingly enhancements of digital technologies also in the ambit of renderings, in many reconstructions, appear in the ability to produce high-level photorealistic imagery: digital processed images seem to be photographs of a real object, represented as if it had just been built. In this case, photorealism become sometimes hyper-realism overcoming the limits of truth likeness.

It is the reason why the author prefers the conceptual 3D models to which materials like clay are applied to the opaque surfaces, and glass to the transparent ones, for simulating physical plastic models. This latter kind of render allows to appreciate the shape generation, the geometric relationships between the building and its context, the perception of the interior and exterior spaces. When the model becomes part of an animated sequence, only some stop-images could be rendered by soft realistic imagery.

In any case the author shares with Ogleby the idea that “what is lacking presently is both an ontology for visual literacy in the area of virtual heritage, and some method of adding to the viewer’s understanding through the supply of supporting information” (Ogleby, 2007).

The ability to access the fourth dimension, through the construction of a sequence of images, constitutes a specific prerogative of the digital representation, which goes beyond the static constraint imposed by the conventional

methods of representation (Garzino, Spallone & Lo Turco, 2011).

Manovich dates back the rise of the movie camera as a universal paradigm for interaction with data represented on three dimensions to the 1980s and 90s, when the interaction between users and virtual models through an interface began to use actions like zooms, tilts, pans and tracks (Manovich, 2001).

The production of movie and animation of 3D models has to focus on well-organized sequences relating to space, event and movement.

As Engeli noted, the creation of specific messages relating to space requires an in-depth knowledge of the characteristics intrinsic to the different possibilities and to the aspects that are wished to emphasize (Engeli, 1999).

The animation, as a tool of the digital reconstruction, highlights its potential in enabling the exploration and dynamic perception of objects, spaces and contexts that no longer exist, or never really existed, but also to represent thematic readings such as periodization, decompositions, building sequences that find the most effective means of communication in the fourth dimension.

MAIN FOCUS OF THE ARTICLE

Issues, Controversies, Problems

As said, currently, the digital reconstruction of contemporary architectural and urban complexes relates to the artefacts that have undergone demolition or transformation or never realized ones. They have in common the fact that they exist today as a proof on paper as if they were or could be.

In any case, sketches and drawings, sometimes accompanied by the plastic models, are the main source for the reconstruction of the project.

Regarding to the sources, in the case of the demolished buildings some illustrative materials can witness the different design hypothesis up to the detailed design and even consist of surveys carried out afterwards, usually enriched by a

photographic repertoire. In the case of unrealized projects the graphic documentation, generally patchy and fragmented, often does not reach the level of detailed drawings. However, several materials, that lend themselves to numerous different interpretative readings, have to be put in relation. In addition, the hypothesis of considering the inevitable missing data is a common need.

Regarding to the reconstruction strategies, in the case of demolished or transformed buildings, several solutions can be reconstructed with different models, or even the final draft, that often deviates from the actual realization; in the case of ideal designs, different hypotheses can be followed and the missing elements can be integrated by the comparison with other existing works. In both cases, the author believes that digital tools, methods and techniques of representation should be carefully selected and adapted to the characteristics identified through the interpretation of the project and that is intended to communicate.

Several scientific approaches adopted by international scholars with respect to the case studies outlined above are addressed in the subsequent discussion, not aspiring to an exhaustive treatment of the topic.

Kahn's unbuilt masterworks and uncompleted projects are the subject of numerous and different studies, involving also a research conducted by the author, supervisor of some bachelor candidates.

Larson, in a book containing the reconstruction of six Kahn's unrealized buildings (Larson, 2000), raises some important questions related to the role of digital modeling for the reconstruction of unrealized projects, which can be extended to objects of the present chapter. He observes that the incomplete evidence left by Kahn could be the score of a performance, where personal interpretation and addition of details are permitted and then discusses about the representation methods and rendering techniques to be used in the reconstructive process. In fact, the synthetic nature of the image could be established through conceptual, non-perspectival sections and isometrics, or impossible views like the section-perspective.

Ambiguous imagery could be created, as Kahn did in his design process, with grainy black-and-white photographs of models that highlight the essential elements and obscure the missing details. These considerations, focusing on the interpretative potential of conceptual modeling, seem to be contradicted by the figurative choices consisting of hyper-realistic images chasing perfection and sharpness of architectural photographs available in contemporary magazines.

Despite Larson's ability to create imperfections in materials, some patterns are visibly repeated and the atmosphere seems too thin. Also it lacks an overall vision of the design that makes understanding the relationship of the building with the environment and its internal functioning.

Dotto develops a study on Kahn's unrealized design of Hurva Synagogue, documented in three versions (Dotto, 2012).

He argues that the shape of absent architectures can be investigated and represented disclosing the partially hidden aspects in the design drawings, right through new drawings making more

accessible the works that would not otherwise be available for study and research.

The attempt to reconstruct the path of Kahn's project also trying to tidy up the chronological sequence of undated drawings, leads to plot a story that highlights the uncertainties of interpretation.

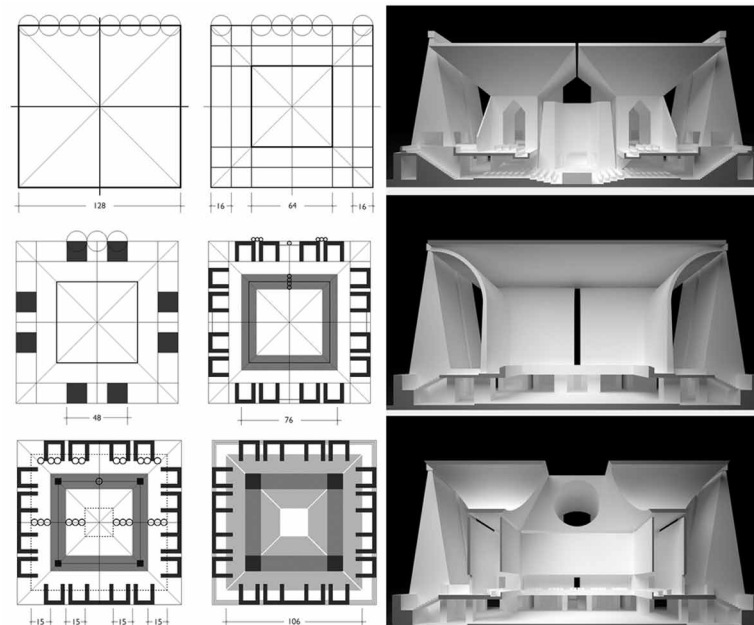
The construction of the virtual model has the advantage of allowing a check of the geometric, spatial and formal coherence of the available representations.

About the digital model display, Dotto raises the question of the use of drawing and modeling as tools of communication, different from the photograph, which triggers the processes of imagination and interpretation of spaces and forms. For this reason he promotes the use of orthographic projections, isometric cutaway, perspective sections, and graphic overlays. Similarly, surface treatments and lightings rather than pursuing photorealism and truth likeness should evoke the visible reality using the instruments of graphic abstraction.

Also a research conducted by the author relates some unrealized and incomplete Kahn's works.

Figure 1. Graphic analysis of the third design of Hurva Synagogue. Cut-away perspectives of the three versions of Hurva Synagogue design.

Source: Dotto, 2012



Among these, there is the unbuilt U.S. Consulate in Luanda (1959-'62) and the incompletely built design of the Salk Institute in San Diego, California (1959-'65), modelled respectively by the bachelor candidates Tancredi and Alberti, supervising by the author.

The design of the U.S. Consulate proposes the incorporation of unglazed, independent forms borrowed from the ruins of the ancient world, which Kahn saw during his stay in Rome, into a very modernist architecture.

In the building Kahn addresses the problems of climate by developing two design themes: the ruins wrapped around the buildings and the separated roofs for the sun and the rain.

Tancredi chose to display the model, mainly in parallel projections, with the aim to highlight the correspondence between the two buildings, the internal distribution, the structural system

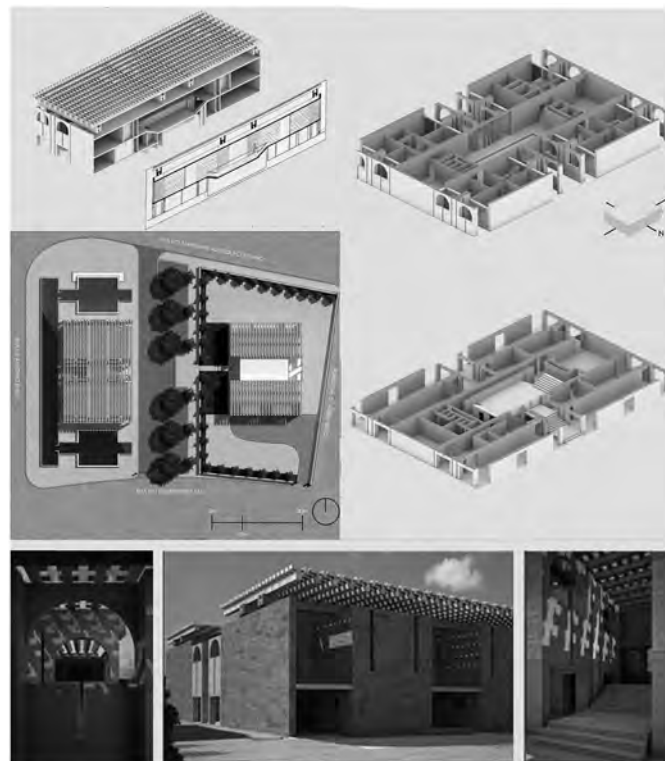
and the performance of the vertical diaphragms (the ruins) and horizontal (the double roof) with respect to the climatic conditions. The roof plan of the whole building shows correspondence in size, alignment and roof perforations to the sun and the rain; cutaways and exploded isometrics show functional partitions and load-bearing supports; the setting of shadows shows the solar behaviour. The application of photorealistic materials and lighting to the perspective scenes in which there are also daily life elements offers possible foreseeing of the artefacts and spaces designed by Kahn.

The Salk Institute for Biological Studies, along the Pacific coasts, designed by Kahn included laboratories, the only that were built, the meeting house and houses for the fellows.

The theme of ruins as devices to control glare appears also here as three-dimensional, complete forms that define and enclose space. Indeed, fully

Figure 2. Digital reconstruction of the unbuilt U.S. Consulate in Luanda. Roof plan of the complex, cut-away and exploded isometric of the Chancellery, renderings.

Source: From the bachelor's thesis by Marco Tancredi. Supervisor: prof. Roberta Spallone.



exterior and independent cylindrical ruins wrap around cubic inhabitable space, and cubic ruins wrap around cylindrical inhabited space.

Alberti aimed to represent Kahn's project in its context as if it were entirely built.

The relationship with an environment strongly characterized by the dry rocky slope overlooking the sea, the spatial articulation of the three architectural units and the careful choice of materials, are Kahn's design themes, which inform the choices of representing the model of the complex inserted in the environment as if it were a plastic model. The terrain has been shaped by extrusion of the contour lines giving the environment a degree of abstraction comparable to that of the artefacts. Even the colour attributed to all the natural elements through the application of a material that evokes the cork allows distinguishing them from the buildings clay rendered.

The final product of the representation is a 7 minutes video, which allows exploring dynamically the three units thanks to the fly and walk through. Only a few still images in photo-realistic rendering with fast fading effect can evoke a perceptual effect generated by materials and real lights.

Also the archival drawings of contemporary masters, currently the subject of renewed attention involving scholars in different disciplines and making use of the tools provided by digital world and convergence of its products, are object of a research line aimed to the digital reconstruction, conducted by the author.

In particular the research concerns several designs by Carlo Mollino, whose archive of drawings is preserved in the library of Architecture at the Politecnico di Torino, and characterizes a series of the master's theses, supervised with Sergio

Figure 3. Digital reconstruction of the latest design of Salk Institute. Three-dimensional model of the complex inserted in the environment. Rendering and screen-shots from the video.

Source: From the bachelor's thesis by Alessio Alberti. Supervisor: prof. Roberta Spallone.



Pace. The subjects until now studied are the Turin Horse-Racing Society Building, demolished in 1960, two ideal houses published respectively on the magazines *Domus* (1942) and *Stile* (1944), and the competition design for the Palazzo del Lavoro in Turin (1960), unrealized because it did not win.

The Horse-Racing, an early and acclaimed work by Mollino, had a short life (from 1937 to 1960). Its demolition was connected with the wider process of urban transformation, which involved the area along the west bank of the river Po, triggered by the celebration of the Unification centenary of Italy.

The digital reconstruction of the building has been realized by Canaj during her master thesis.

She also analyzed the relationships between the building and the context, less dense than the current, but strongly characterized by different buildings, reconstructed by means of the archival city maps and design drawings.

The Horse-Racing has been reconstructed in detail, once those archival drawings were identified to allow the most faithful reproduction of reality, after filling in the missing information and resolving ambiguities and inconsistencies of the documents. In this sense, the re-drawing of the project takes on the meaning of a true re-design that requires a deep understanding of the artefact and poetics of the architect.

The modeling phase has been carried out with the technique of 'blueprint', using CAD and 3D graphics software, arranging plans and elevations on orthogonal planes so as to foster the most appropriate control of the process.

The model has been lightened with sunlight while, about the materials, the opaque surfaces have been rendered using clay and the transparent ones with glass. In this way the model maintains the proper level of abstraction and avoids generating the sense of fake that characterizes the photorealistic reconstructions. The choice to represent the perspective views of the model in black and white goes to the same aim, allowing also a comparison with the vintage photographs taken under the guidance of the same Mollino.

Further processing, which aims to compare the digital model renderings with the original photomontages, and to search the same view points and solar illumination conditions, has been led to try a new photomontage of the model in the portrayed old context.

The Casa sulla collina (1943) and Casa sull'altura (1944) are two ideal houses designed by Mollino, digitally reconstructed by the master candidate Laudani.

The issue of the ideal houses animated the Italian architectural debate, during the period of the forced inactivity due to the II World War, so that several architectural magazines became promoters of the initiative to request and publish projects of the contemporary top architects on this subject.

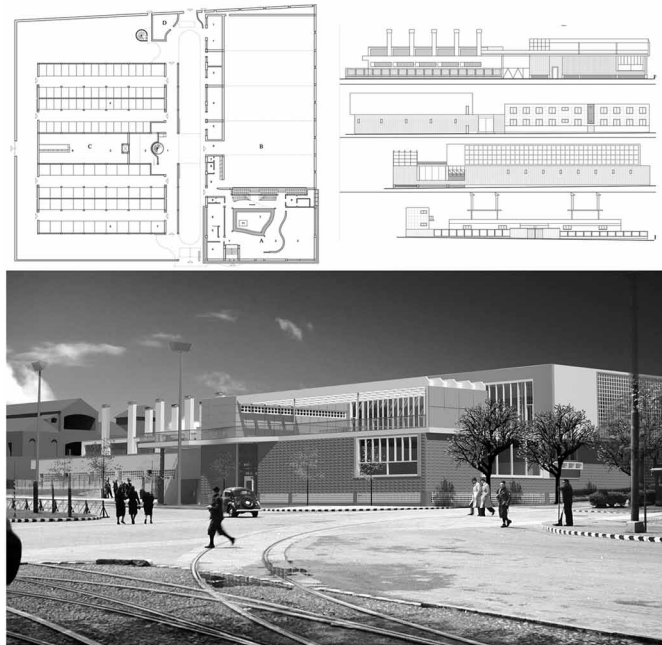
In both case studies, the publication of drawings was accompanied by Mollino's writings to the editors, which described the concept, his reasons and widens the speech stating his position with respect to the debate.

Carlo Mollino was an architect who combined the research of architectural quality to a strong knowledge and experience of building (Pace, 2006), because he worked, from the beginning and for about twenty years, with his father Eugenio, an engineer particularly productive especially in Turin area.

The relationship between the drawing and the constructive reality emerges also in Mollino's designs programmatically intended to remain on paper, of which also the technological details in large scale were graphically defined, the interior furnishings were designed as an integral part of the architecture and, even, the static schemes of some structural elements were traced.

The particularity of Mollino's *modus operandi*, which was expressed through hundreds of drawings for each project, led the author to propose the student to use a parametric three-dimensional modeller like BIM, aimed to the three-dimensional reconstruction, in order to assess the three-dimensional consistency of the two-dimensional archival drawings, checking possible variants and obtaining sections from the

Figure 4. Two-dimensional re-drawing of Horse-Racing Society, plan and elevations. Original scale 1:200. Photo-montage between rendered model of Horse-Racing Society and original photograph. Source: From the master's thesis by Florida Canaj. Supervisor: prof. Roberta Spallone, advisor: prof. Sergio Pace.



3D model, provided with the graphic standards of the architectural technical drawing and settable at different scales, selected also in order to compare them to the sources.

The reconstruction was completed by a video and a plastic model digitally fabricated: a demountable material model, made with a small 3D printer which extrudes fibres of polylactic acid.

Through these operations, for the first time the design has been freed from the two-dimensional support through the transformation in three-dimensional digital model (always accessible through the two-dimensional space of the screen), achieves the fourth dimension by the production of the video that allows to visit, offering new views, and finally truth imbued in a three-dimensional material object.

In addition, through the examination of some drawings by Mollino, who placed the Casa in collina in relationship to the centre of Turin, its location could be hypothesized, imagined by the architect on Monte dei Cappuccini, near the

XVII century church. The photo-montage of the 3D model in the current environment highlights a shocking antagonistic relationship between the house and the church for the conquest of the hill-top.

Other subjects of reconstructive modeling are some minor architectures realized in Italy during the Thirties on charge of the Fascist Party, today demolished or deliberately lost due to their symbolic means.

Francesco Maggio conducted researches on some 'forgotten' buildings of the fascist period.

In the case studies of the Fascist House (1929) by Enrico Del Debbio and the Post Office (1936) by Angiolo Mazzoni, both located in a square of Agrigento, the author compared the 3D model drawn out by the design drawings and the survey of the existent building, and underlined the relationships between the two buildings and the transformations of urban tissue (Maggio, 2008).

On the same subjects the research experience, carried out by the author and Bruno jr. with some

Figure 5. Screen-shots from BIM model of the Casa sulla collina. Scale material model of the first level. Horizontal sections of BIM model. Photo-montage of 3D digital model in the real environment.
Source: From the master's thesis by Antonio Laudani. Supervisor: prof. Roberta Spallone, advisor: prof. Sergio Pace.

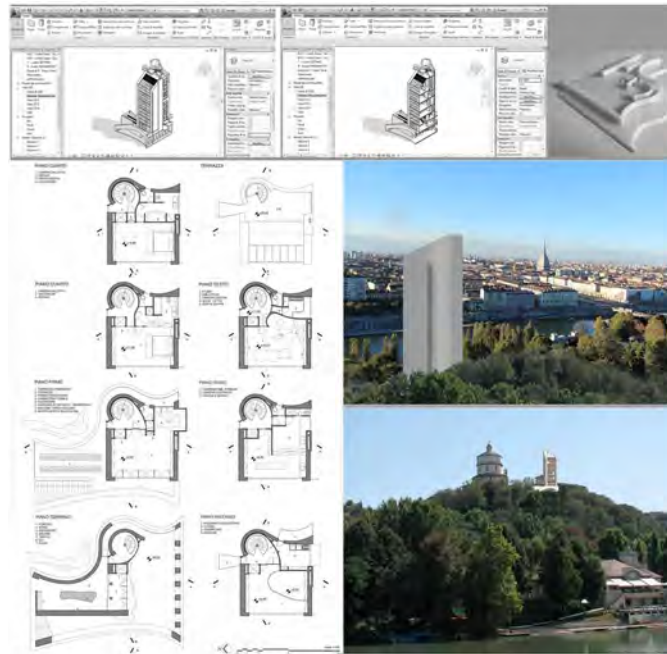
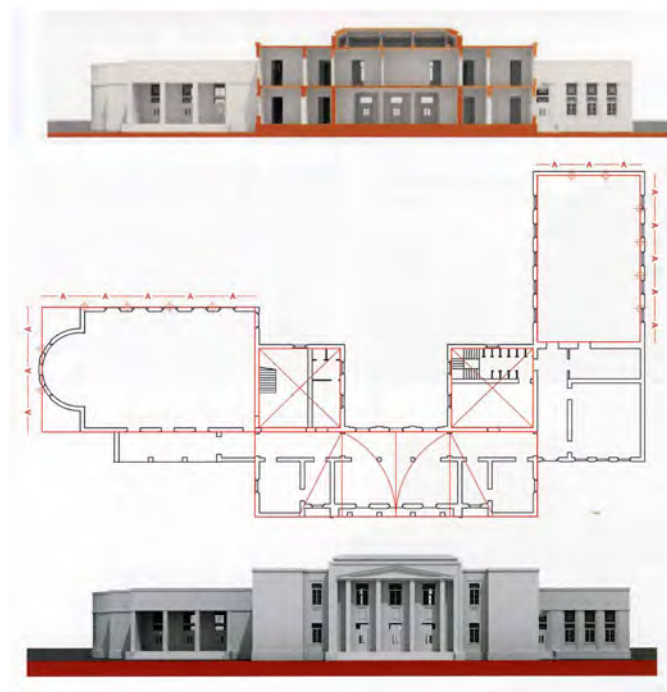


Figure 6. Digital reconstruction of the first version of the Fascist House in Agrigento. Graphical analysis of the ground floor plan, perspective view of the elevation, and perspective section.
Source: Maggio, 2008



bachelor candidate, concerns the digital reconstruction for the preservation of the architectural heritage memory (Spallone, Bruno jr., 2013).

The reconstructive operations started with the analysis of the archive documents, the search of old photographs, postcards and city maps, aimed to insert the building in the context, and a survey on the building materials utilized at that time. After this first step of data collection, the digital reconstruction of the building and the urban tissue began. The interaction among different disciplines, such as history of architecture, representation and also material technology guaranteed the experiment a higher scientific level.

One of the case studies analyzed, the Casa del Marinaretto, built in the Thirties and demolished without any consideration in the early sixties, shows such original and interesting design solu-

tions, that the preservation of its authentic memory and its valorisation seem justified.

The Casa del Marinaretto was designed by Costantino Costantini in Turin, close to the river Po, looking like a big 'urban anchored ship', constituting a high quality building with international inspiration. The work moves after the rigorous analysis of the archival original drawings and proposes the reconstruction of the building according to its original version as it was drawn by the architect for the first time, and so in a different manner respect to the built result. This process guarantees also the 3D perception, the authenticity of the materials and of the location into the urban context, using digital static and dynamic rendering systems. The 3D model of Costantini's project, inserted into the urban and natural context of our days, to simulate the perceptive effects that

Figure 7. Digital reconstruction of the design drawings of demolished Casa del Marinaretto inserted in the present context. Conceptual bird-eye view of the building in the environment and photorealistic rendering. Screen-shots from the video.

Source: From the bachelor's thesis by Francesco Carota. Supervisor: prof. Roberta Spallone, advisor: prof. Andrea Bruno jr.



the building could express today as if it was still standing, was realized by Carota.

The model, realized using 3D computer graphics and animations programs, was the basis for the production of a video-clip that dynamically explored the relationship between the actual image of Turin and the Casa del Marinaretto as if it still existed.

Again in Turin, Costantini built the Fascist house in via Guastalla aimed to accommodate the youth organizations and at the same time to ensure a gym to the adjacent school.

This building, now in a poor state of preservation, exemplifies the architect's ability to adopt the rationalist language, working on the sharpness of the shape, the absence of ornamentation, the use of materials and technical choices consistent with modern era, such as the use of glass bricks

and roofing flat. The digital model of this building, reconstructive of the original design, today heavily altered, shows the main elements which characterize a building of great balance and lightness, free from rhetoric of the regime or by self-complacency. It was realized by Businaro using BIM technologies with the aim of creating a database useful for managing the predictable actions of functional requalification and energy efficiency improvements.

SOLUTIONS AND RECOMMENDATIONS

The originality of the approach, here developed and proposed, is the thesis that the tools, methods and techniques of representation (graphical analysis, bi

Figure 8. Digital reconstruction of the original configuration of Fascist House in via Guastalla. Two-dimensional drawings obtained by 3D model and cut-away isometric. Identification of the key elements of the design. Comparison between rendered 3D model with the original material and the present condition of degrade.

Source: From the bachelor's thesis by Gabriele Businaro. Supervisor: prof. Roberta Spallone, advisor: prof. Andrea Bruno jr.



and three-dimensional modeling, animation, prototyping etc.) should be chosen with care, case by case, in order to properly interpret the knowledge data and generate interpretative readings, using as sources and ideas for research not only the archive drawings and any surviving vestiges, but also autographs writings and significant analyses of architectural critics.

FUTURE RESEARCH DIRECTIONS

The digital reconstructions are emerging as important tools for understanding, interpreting and sharing the cultural heritages. The wide and continually increasing availability of tools offered by the digital world makes thinking, on the one hand, the application of the latest game engines for interactive exploration of the reconstructed complex, on the other the use of augmented reality to enrich the informative content of the archival heritage.

CONCLUSION

This chapter develops some considerations about the issues raised by the digital reconstruction of contemporary architectures and urban context no longer existing in their original configurations.

Through the overview of the international theoretical positions, different scholars' critical approaches emerge.

Generally archival drawings are patchy and fragmented and refer to different ideation moments and paths of inspiration that lend themselves to numerous and different interpretative readings. The other sources of knowledge of the artefact, even when it was built and transformed or demolished, have similar characters of variability and fragmentation. Moreover a careful analysis of the author's poetics and significance of his work is necessary. Digital methods and techniques of representation, ranging from 3D modeling, video producing, digital fabrication and other future

application of digital technologies, should be carefully selected and adapted to the characteristics identified through the interpretation of the project. The path of knowledge, which should be as wide and deepened as possible, and the possible interpretations that follow, will suggest the reconstruction strategies of the complex.

REFERENCES

- Albisinni, P., & De Carlo, L. (Eds.). (2011). *Architettura disegno modello. Verso un archivio digitale dell'opera di maestri del XX secolo*. Roma, Italy: Gangemi.
- Docci, M. (2009). *Disegno e analisi grafica*. Roma-Bari, Italy: Laterza.
- Dotto, E. (2012). *Il progetto della Sinagoga di Hurva di Louis I. Kahn. Analisi grafica*. Roma, Italy: Aracne.
- Engeli, M. (1999). *Storie digitali. Poetiche della comunicazione*. Torino, Italy: Testo&Immagine.
- Gaiani, M. (2004). About the drawing and the model: representation for industrial design. In R. Migliari (Ed.), *Drawing as Model* (pp. 91–98). Roma, Italy: Edizioni Kappa.
- Galifi, P., Moretti, S., & Aoyagi, M. (2002). *The Archaeological Reconstructions, Methodologies and Instruments for a Graphic Representation of an Idea*. University of Tokyo, Faculty of Engineering. Retrieved June 17, 2009 from <http://www.altair4.it/>
- Garzino, G., Spallone, R., & Lo Turco, M. (2011). Digital strategies for knowledge based models. In G. Garzino (Ed.), *Drawing (and) information. Polytechnic drawing* (pp. 70–111). Sant'Arcangelo di Romagna, Italy: Maggioli.
- Larson, K. (2000). *Louis I. Kahn. Unbuilt Masterworks*. New York: The Monacelli Press.
- Maggio, F. (2008). Enrico Del Debbio e Angelo Mazzoni ad Agrigento. *Disegnare*, 37, 58–71.

Maldonado, T. (2005). *Reale e virtuale*. Milano, Italy: Feltrinelli.

Manovich, L. (2001). *The Language of New Media*. Cambridge, MA: MIT Press.

Migliari, R. (Ed.). (2004). *Drawing as Model*. Roma, Italy: Edizioni Kappa.

Ogleby, C. (2007). The 'Truthlikeness' of Virtual Reality Reconstructions of Architectural Heritage: Concepts and Metadata. *ISPRS Archives*, XXXVI-5/W47.

Pace, S. (Ed.). (2006). *Carlo Mollino architetto, 1905-1973: costruire la modernità*. Milano, Italy: Electa.

Pagnano G. (1975). *La lettura critica: analisi di cinque opere di Adolf Loos*. Catania, Italy: Istituto Dipartimentale di Architettura e Urbanistica dell'Università di Catania.

Spallone, R. (2007). 3D digital modeling as a method for the reconstruction of the historical image of the city: The case of piazza Bodoni in Turin (Italy) at the end of nineteenth century. *ISPRS Archives*, XXXVI-5(C53), 685–690.

Spallone, R., & Bruno, A., Jr. (2013). Heritage and memory. Digital reconstructions of minor architectures of the Thirties in Turin area. *Proceedings of XXXV International Congress UID*, 909-916.

KEY TERMS AND DEFINITIONS

3D Modeling: In 3D computer graphics, 3D modeling is the process of developing a mathematical representation of any three-dimensional surface of object via specialized software.

Animation: Animation is the process of creating a continuous motion and shape. In the architectural representations, animation can be the creation of a path inside and outside of buildings, or the putting in motion of static elements.

BIM Modeling: Building Information Modeling (BIM) is a digital representation of physical and functional characteristics of a facility. BIM involves representing a design as combinations of objects that carry their geometry, relations and attributes.

Blueprint Technique: It is a modeling technique which consists of having plans and elevations on orthogonal planes so as to ensure the selection of modeling operations and foster the most appropriate control during the process of creating three-dimensional digital model.

Digital Reconstruction: It is the process of reconstruction of a disappeared building or modified over time, by means of 3D computer graphics programs, typically starting from archival documents and surveys.

Regional Development and Air Freight Service Needs for Regional Communities



Tarryn Kille

Griffith University, Australia

Paul Bates

University of Southern Queensland, Australia

Patrick S. Murray

University of Southern Queensland, Australia

INTRODUCTION

This article is intended to assist in documenting the needs of regional communities with regards to airfreight services. In doing so, the article first considers the importance of air service access to small and regional communities with particular emphasis on those economies bolstered by agriculture and the production of high-value, low weight, perishable products. The article then explores the airfreight challenges facing regional communities and discusses the needs for: integration; infrastructure; and service reliability. The article provides an overview of potential solutions to assist in meeting these needs, including: improving the flow of information and communication; and applying electronic commerce strategies.

BACKGROUND

Over the past two decades, the air cargo industry has grown dramatically (Yuan, Low, & Ching Tang, 2010). Although world air cargo traffic stagnated from mid-2011 to 2013 (Boeing, 2014), growth returned in 2014 accounting for approximately 35% of global merchandise trade by value (International Air Transport Association, 2015). The dramatic growth in the global airfreight sector may be explained by three observations.

Firstly, the dramatic growth in airfreight can be explained by an apparent industry trend towards the production of high value, lightweight goods (Ari-Pekka & Hintsa, 2009). This includes the new economy associated with the transport of fresh, perishable, high value produce (Sim, Barry, Clift, & Cowell, 2007). Statistics support this emerging economy highlighting that the transport of perishable food accounts for 14% by volume of total global airfreight (Bridger, 2008). Secondly, producers and shippers are realizing that the higher costs associated with airfreight can potentially be offset by the costs savings associated with storage, and packaging when using other freight modes (Yuen et. al., 2010). Finally, on a global scale, air service costs have experienced a reduction due to the entry of large numbers of wide body freighters to the logistics industry (Gardiner, Ison, & Humphreys, 2005). As such, the overall effect is that airfreight is becoming a commonplace business decision in the distribution systems of many companies (Murphy, Dalenberg, & Daley, 1989).

Increases in the perishable airfreight task are linked to the emergence of a social focus on the consumption of foods produced from regions known for 'clean' production processes. This new social focus has called for research, which investigates the food chain in an effort to understand rural development patterns (Renting, Marsden, & Banks, 2003). Thus, sea, road, rail and air

transport of freight in the food production cycle from the farm to the consumer's table has come under scrutiny (Saunders & Hayes, 2007). The scrutiny has been applied in the pursuit of clarity surrounding the cost of transport to the environment, society and communities.

Regional communities are the first step of the food production cycle. For many regional communities, access to airfreight networks means an opportunity to penetrate and participate in markets that cannot be equally serviced by other modes of transport. For perishable agri-producers, access to market requires access to appropriate freight services with the ability to transport products efficiently and effectively from the farm to the consumer plate. Thus, understanding the needs of regional communities with regards to airfreight access is an important factor in fostering sustainable economies for these communities (Kille, Bates, & Murray, 2013).

Although air cargo decisions are becoming the norm in the distribution systems of many companies (Murphy et al., 1989), there is little recognition of the needs of regional communities with regards to airfreight access. There is a plethora of literature and research into international and domestic air cargo utilizing the cargo space of wide body passenger jets or dedicated aircraft freighters. However, there is a shortage of literature on another critical part of the perishable cargo supply chain (i.e., regional aviation and the associated airfreight service needs of regional communities). Furthermore, much of the current literature and studies focus on the needs of shippers in an international setting and provide little to further our understanding of how decisions associated with airfreight fit contextually within a regional community and their economic development. The authors contend that information and technology lie at the core of airfreight service needs of regional communities.

For the purposes of this article, a small or regional community is taken to be one that maintains a 'rural airport' as defined by the United States Internal Revenue Service (IRS). The IRS defines

a 'rural airport' as any airport that has fewer than 100,000 commercial passengers departing from the airport during a calendar year, and at least one of the following is true: a) the airport is not located within 75 miles of another airport from which 100,000 or more commercial passengers departed during the preceding calendar year; or b) the airport is not connected by paved roads to another airport (United States Department of Transportation, 2016). Regional aviation includes the air transport operators serving these rural airports.

This article investigates the needs of regional communities with regard to airfreight services. The author reveals that information and technological challenges may be overcome by addressing the key airfreight network areas of concern including: (1) integration; (2) infrastructure; and (3) service reliability.

THE CHALLENGE: AIRFREIGHT SERVICE NEEDS FOR REGIONAL COMMUNITIES

A number of challenges for the regional air transport sector of the perishable food supply chain have emerged. This section explores the impact of: integration, infrastructure, and service reliability on the provision of services to regional communities.

The Need for Integration

Access to air transport services and feeder services is often limited for rural communities. For example, the scarcity of freight forwarding providers presents significant challenges for growers who are left uncertain about availability of potential logistics providers in the region. Rundle (1998) explains that the limitations imposed on agricultural exporters is rooted in the passenger demand and scheduling policies administered by domestic airlines. Hanaoka and Phoosanabhongs (2010) support this and assert that reliability becomes a

key concern for shippers when the cargo capacity of an aircraft is used to accommodate both passenger services (i.e., baggage) as well as freight cargo. When an airline encounters difficulties associated with weight and balance, it is not passengers or baggage that are offloaded but the freight cargo.

Dedicated airfreighter services, however, are not common for regional communities. As such, it is often the regular passenger air service (where available) that offers the only method of freight transport by air. Batt and Morooka (2003) argue that cost is not the only decision element for producers of high value perishable products. When planning the mode of freight transport to use, it is usual for shippers of these products to equally consider reliability, strength of network and shipping elements that ensure the quality of the produce at delivery. These elements of logistics decision-making must be addressed by regional airfreight suppliers with consideration to the importance of integration requirements in strengthening service delivery. The transport of high value produce via air offers an additional revenue stream for passenger airlines and Popescu, Keskinocak, Johnson, LaDue, and Kasilingam (2006) contend that airlines, thus, must ensure that cargo capacity is managed effectively. The application of an effective cargo capacity forecasting model (similar to the passenger overbooking model utilized by many passenger airlines), should be considered by air transport providers handling cargo in an effort to enhance the utilization of cargo capacity (Popescu et al., 2006). Proficiency in this area offers the potential for regional airlines (often with thin margins) to achieve business success.

Fundamental elements of the 'farm to table' movement include prompt, reliable and quality transport services. It is apparent that a variety of transport modes are required to transport freight produce from the farm gate. Accordingly, it is essential for air service providers to collaborate with other freight transport modes. The advantages of integration in the supply chains of the food industry are explored by Stank, Crum, and Arango (1999) who support the coordination in processes

(such as communication, information technology, partnering, and performance monitoring) between firms. Such integration will encourage broad improvements to critical logistics performance areas (Stank, Crum, & Arango, 1999). For regional airfreight providers, the most significant obstacle will be the sharing of crucial business process information and operations with providers of other freight transport modes.

Lillie and Sparks (1993) concur that shippers of high value perishable products value equally both cost and total freight service offering. More recently, however, speed-to-market delivery requirements, and the management of the supply chain have expanded demand for complementary intermodal transport services and infrastructure. For those rural communities supplying the 'farm to table' movement, integration is a significant issue and a facility to engage complete door-to-door services through a single platform is most coveted for such shippers (Tseng, Yue, & Taylor, 2005). The pragmatic application of the concept of integration will be a challenge for regional airfreight providers. Hence, integration in freight-forwarding will need to be forged by key regional stakeholders able to communicate with and appreciate the scope of freight services available at one time (Kille, Bates, Murray, 2014).

The Need for Infrastructure

Specific climates within specific geographic locations often form one of the key attributes for many niche agri-products. Although the uniqueness of the geographic location is frequently the basis for the product success, the locality can also become a hindrance with respect to the time and distance the product travels to processing centers. Financial incentives and government supported research is necessary to attract investment in this area. With access to quality information, investors and producers can make more appropriate long term decisions associated with the development of infrastructure, innovative growing technologies and sustainable growing practices (Kille, Bates,

& Murray, 2014). For example, Gardiner et al., (2005) assert that airports offering marketing assistance (such as the funding of research activities) increase their chances of attracting airlines and associated investment opportunities.

Effective airfreight services are highly dependent on factors including the geographic location (Gardiner et al., 2005) and relative infrastructure of the airport. A recent study examining international cargo gateways found that airports within close proximity of shippers generally have lower costs and lower delivery times which assists those airports to develop into regional air cargo hubs (Zhang, 2003). While the regional airfreight sector is challenged by competing modes of transport, realizing the importance of airport location can make airfreight options competitively more attractive. Freight by air is generally considered a greater expense when compared to road, rail or even sea transport. However, the reduced transport time (often associated with airfreight) lends to reduced time and costs associated with storage. Balancing costs in this way may project airfreight as a reasonable mode of transport for certain transport purposes.

Further infrastructure related challenges impacting on the regional airfreight sector includes customs clearance efficiency (Zhang & Zhang, 2002) and access to skilled and well informed freight forwarders (Gardiner et al., 2005). Government and council imposed environmental restrictions imposed on airports, (such as curfews on night operations), also challenge the operations and development of airports (Shaw, 1993). Kille, Bates and Murray (2014) propose that environmental limitations such as these can be overcome with community cooperation and engagement, which aims to balance environmental concerns (such as noise restrictions) with economic limitations.

Availability and access to tangential facilities is an additional infrastructure challenge facing regional airfreight operations concerned with the movement of high value perishable products. Kille, Bates, and Murray (2014) suggest that appropriate access to tangential facilities such as cold storage, may strengthen the overall service quality and

reliability of the entire freight transport service. Yuan et al. (2010) concur, contending that the enhancement of airport infrastructure to include cold storage facilities nurtures the opportunity to improve the overall performance of airfreight services.

The Need for Service Reliability

The term 'service reliability' in logistics is established in a compact body of literature (Gardiner et al., 2005; Yuan et al., 2010). Park, Choi, and Zhang (2009) explain that service reliability includes attributes of accuracy, punctuality, dependability, and safety.

The challenge of service reliability is unavoidably linked to the two challenges discussed previously, as adequate service integration and suitable infrastructure are key characteristics of 'service reliability'. It is the speed and reliability requirements of modern supply chain management processes that has supported the rapid development of airfreight and express air cargo services (Park et al., 2009).

A study conducted by Foster and Strasser (1990) explored fundamental factors defining the carrier/modal selection of shippers. The inquiry uncovered that the most dominant selection variable for shippers was 'schedule reliability'. The significance of cost and service variables, before and after the impact of government imposed deregulation policies, was considered by Murphy and Hall (1995). The research revealed that when considering choices of freight transport, shippers predominantly rank 'service' higher than 'cost'.

The elements of competition impacting on regional airfreight are similar to those challenges recorded in the international studies (i.e., locality, cost and frequency). Thus, comparative conclusions can be applied to the regional airfreight sector. This sector will need to consider responding to the specific demands of shippers looking for complete door-to-door services as opposed to simple airport-to-airport services (Park et al., 2009).

SOLUTIONS AND RECOMMENDATIONS

Integration, infrastructure and service reliability are key challenges to be addressed by freight forwarders (including airfreight service providers in regional communities). It is apparent that these three challenges are inextricably bound and cannot be addressed in isolation. Over the past decade, similar challenges in the global freight industry have been overcome by: improvements in the flow of information, (both within an organization and external to the organization); and the application of technological tools and e-commerce. The following section discusses these solutions and their applicability in addressing the airfreight service needs of regional communities.

Improving the Flow of Information

International competition and the demand from shippers for timeliness and flexibility of delivery has forced companies at all stages of the supply chain to consider how they can enhance their competitiveness (Hesse, 2002). In the previous section, we discussed the importance for rural shippers (of high-value, perishable products) to have access and opportunities to book entire services door-to-door through a single platform. In an effort to realistically apply this concept, freight-forwarding companies (including regional air transport service providers) will need to consider integration where key regional stakeholders are able to communicate with and understand the range of freight services available. Thus, an important driver for the development of airfreight services is the requirement for logistics companies to improve the flow of information both internally and externally (Rutner, Gibson, & Williams, 2003).

Information systems have experienced steady growth in inventory management, production and logistics. In global freight networks, the expanding information requirements have facilitated the integration of logistics information systems (LIS) and supply chain information systems. Over the

past three decades, LIS has captured a significant area of study in the field of logistics and operations (House & Jackson, 1976; Whipple, Frankel, & Anselmi, 1999; Williams, Magee, & Suzuki, 1998). Considering the increased application of LIS, one area of research has focused on the evaluation of the ability to integrate logistics and LIS throughout the organization. Integrated logistics refers to the system-wide management of the complete logistics chain as a single entity, instead of separate management of the individual.

A study conducted by Gustin, Daugherty, and Stank (1995) found that companies that have higher levels of logistics integration throughout the organization, are also likely to be using more logistics information systems. Later, Rutner et al. (2003) supported this finding, asserting that companies that have successfully implemented the concept of integrated logistics are also companies that are carefully utilizing appropriate logistics information systems. For airfreight providers serving regional communities, this solution implies that air operators consider integrating businesses through acquisition and mergers with other modes of transport to ensure the door-to-door nature of freight service is effective. Furthermore, this option ensures that information and communication flow and management of the complete freight transport system are entirely centralized.

While this may offer an obvious solution in meeting the integration and service reliability needs of regional communities, this solution requires significant investment in additional transport modes, integrated information system management tools and associated infrastructure. This may be feasible for large freight companies. However, the smaller service providers operating in specific regional locations may find this option prohibitive (Kille et al., 2013). Berger and Schroder (2011) considered this limitation, appreciating that the transport of airfreight from the shipper's door to the door of the consignee is complex and involves a number of collaborating logistics companies (e.g., freight forwarders, road transporters, airfreighters etc). This process involves a number

of individuals who manage their own aspect of the delivery with their own perspectives of the logistics chain. These individual perspectives result in individual decisions which impact on the quality of the delivery and final delivery time.

Considering the reality of airfreight, Berger and Schroder (2011) maintained that information flow continues to be of great benefit to overall service quality. However, Berger and Schroder (2011) argue that efficiencies within the entire door-to-door transport system can be achieved by maintaining the system of decentralized decision-making. Effectively, the approach considers specific criteria that is relevant to each stage of the delivery process and encourages the sharing of critical information required for each aspect of the logistics chain. This allows sensitive information (including price, customer information etc) to be maintained by the local operators and only the most necessary information to be shared along the logistics chain allowing for more effective planning and improved service reliability.

Forster and Regan (2001) conducted a case study analysis of the international airfreight industry in the United States examining the impact of electronic integration in logistics supply chains. The study found that a technologically deterministic perspective of the impact of electronic integration, without considering the contextual influence is misguided in its assumption. In the body of supply chain literature, the discourse associated with relationships between companies is distinguished by language of coordination and cooperation that allow the seamless transfer of information. However, the research of Forster and Regan (2001) advocates that electronic integration must also consider issues of power (influence and domination one company has over the other) and conflict (i.e., competitive, technological or language differences) that influence the amount of information that needs to be transferred between companies.

Decentralized and centralized approaches are applicable in the context of airfreight service needs

for regional communities. The overriding solution, however, emphasizes the need for airfreight service providers to implement systems that allow the gathering and sharing of applicable information across the entire logistics process. While issues of power and conflict between companies will need to be addressed, the implementation of information sharing systems is further supported by the application of appropriate technological solutions and electronic commerce.

Technology: Application of Electronic Commerce

The integration of data and information systems within the entire logistics process has been supported by the development of modern information and communication technologies (Hesse, 2002). Distinct characteristics such as electronic data interchange (EDI), the automation of product flow in dedicated distribution centers and warehouses, or the use of computer-based tracking and tracing systems (offering online awareness of the status and location of the shipment) are only a sample of the principal sources of significant productivity gains over the last twenty five years (Golicic, Davis, McCarthy, & Mentzer, 2002).

The emergence of the Internet has allowed dramatic improvements in the sharing of information in logistics. The costs of web-based inventory management tools have reduced and thus the uptake of such systems has increased (Zillur, 2004). Furthermore, the web has become a common marketplace for internet-based suppliers. The Internet has developed into a significant management tool, not only for the individual companies, but also for controlling the complete producer-shipper relationship (Hesse, 2002).

The term 'electronic commerce' is principally concerned with the exchange of information and transactions between different commercial entities (i.e., business-business electronic commerce) and exchange of information between businesses and customers (i.e., business-to-customer elec-

tronic commerce). Electronic commerce (or e-commerce) transactions can be divided into customer related activities (e.g., receiving orders and marketing) and the business-related activities (e.g. the processing and shipment of the orders).

Rutner et al. (2003) argue that more advanced organizations are beginning to extend their logistics operations to the e-commerce environment through the implementation of internet-based purchasing and extranet-based supply chain management applications. For these companies that have progressed through the integration process, the application and implementation of systems that also support intranet-based activities and communication appears to be an important step towards achieving logistics integration.

Cost is an important driver in competitive industries such as transport and logistics. The enormous economic potential of business-to-business e-commerce to effectively lower costs is dependent on a single factor. That is, the complete break-down and re-organization of the supply chain with cost efficiencies at the heart of the examination (Reynolds, 2000). Many of the common e-commerce applications in freight distribution, manufacturing or retail relate to: order processing and completion; management of inventory; fleet management and vehicle routing; brokering of freight; and electronic payment.

In the case of regional communities, where integration of services is key to the success of the entire logistics process, the evolution of e-commerce is likely to further support the cost saving potential of any integrated logistics concepts. However, integration of systems in the regional airfreight network will continue to be a barrier unless the entire logistics chain is re-examined and restructured to allow essential information to be shared. Never-the-less, e-commerce as a part of the entire distribution system can provide positive outcomes for freight companies, shippers and consumers by allowing higher delivery frequencies over longer distances and enhanced air transport networks (Hesse, 2002).

FUTURE RESEARCH DIRECTIONS



The implementation of e-commerce offers increased levels of support for the integration of logistical operations by allowing improved access to and linkages between varied types of information that are critical to the logistics process. Rutner et al. (2003) argues that as the reliance on such systems continues to increase, it is even more imperative for managers in both the fields of logistics and information systems to strengthen working relationships between the two functions. Effectively, information systems will need to support logistics processes and logistics processes will need to feed in and support information systems.

To what extent business-to-business e-commerce applications are being implemented in freight and distribution systems is not entirely known. Considering the potential benefits, it is assumed that information management and integrated processing of the entire supply chain are used widely, particularly in medium sized and large firms (Hesse, 2002). Additionally, the contribution of e-commerce to a more sustainable and economically viable distribution system is highly dependent on a number of factors including regional circumstances, shipper habits, consumer habits, accessibility of various transport modes, geographic location of the airport and associated facilities, population density and political and governmental support. Future research is required to more comprehensively examine the application of e-commerce in regional freight networks and the impact that these factors may have on the provision of freight services to regional communities.

CONCLUSION

Access to airfreight networks allows regional communities the opportunity to participate in markets that cannot be equally serviced by other modes of transport. For agri-producers focused on the production of high-value, low weight products, this

means access to appropriate freight services with the ability to transport products efficiently and effectively from the shipper's door to the consumer's door. Thus, examining the needs of regional communities with regards to airfreight access is a principal concern in developing and fostering sustainable economies for these communities. This article has discussed the airfreight challenges facing regional communities and explored their needs for integration, infrastructure, and service reliability. Information sharing and the application of appropriate technological tools lie at the heart of these challenges. In discussing the shifting demands of consumers and the emergence of the Internet, the authors contend that these logistics challenges can be overcome by: improving the flow of information and communication between businesses in the entire distribution system; and applying e-commerce tools to assist in integrating systems and sharing of information.

REFERENCES

- Ari-Pekka, H., & Hintsa, J. (2009). Assessing the drivers of change for cross-border supply chains. *International Journal of Physical Distribution & Logistics Management*, 39(9), 741–761. doi:10.1108/09600030911008184
- Batt, P. J., & Morooka, R. (2003). Perceptual differences in offer quality between Western Australian rock lobster exporters and Japanese rock lobster importers. *Supply Chain Management*, 8(5), 476–484. doi:10.1108/13598540310500303
- Berger, M., & Schroder, M. (2011). Multicriteria decentralized decision making in logistic chains: A dynamic programming approach for collaborative forwarding of air cargo freight. *Logistics Research*, 3(2), 121–132. doi:10.1007/s12159-011-0055-8
- Boeing. (2014). *World Air Cargo Forecast*. Seattle, WA: Boeing World Air Cargo Forecast Team.
- Bridger, R. (2008). *Food air freight, the global infrastructure expansion*. Academic Press.
- Forster, P. W., & Regan, A. C. (2001). Electronic integration in the air cargo industry: An information processing model of on-time performance. *Transportation Journal*, 40(4), 46–61.
- Foster, J. R., & Strasser, S. (1990). Carrier/Modal Selection Factors: The shipper/carrier paradox. *Journal of the Transportation Research Forum*, 31(1), 206–212.
- Gardiner, J., Ison, S., & Humphreys, I. (2005). Factors influencing cargo airlines choice of airport: An international survey. *Journal of Air Transport Management*, 11(6), 393–399. doi:10.1016/j.jairtraman.2005.05.004
- Golicic, S. L., Davis, D. F., McCarthy, T. M., & Mentzer, J. T. (2002). The impact of e-commerce on supply chain management. *International Journal of Physical Distribution & Logistics Management*, 32(10), 851–871. doi:10.1108/09600030210455447
- Gustin, C. M., Daughterty, P. J., & Stank, T. P. (1995). The effects of information availability on logistics integration. *Journal of Business Logistics*, 16(1), 1–21.
- Hanaoka, S., & Phoosanabhongs, E. (2010). Estimation of the occurrence of short-shipping of air cargo. *Journal of Air Transport Management*, 16(6), 315–319. doi:10.1016/j.jairtraman.2010.03.007
- Helms, M. M., & Dileepan, P. (2005). Transportation Issues for Supply Chain Management. *Business Forum*, 27(1), 8–13.
- Hesse, M. (2002). Shipping news: The implications of electronic commerce for logistics and freight transport. *Resources, Conservation and Recycling*, 36(3), 211–240. doi:10.1016/S0921-3449(02)00083-6
- House, R. G., & Jackson, G. C. (1976). Trends in computer applications in transportation and distribution management. *International Journal of Physical Distribution & Logistics Management*, 24(4), 176–178. doi:10.1108/eb014396

- International Air Transport Association. (2015). *The Value of Air Cargo: An ode to air freight*. Retrieved from <http://www.iata.org/whatwedo/cargo/sustainability/Pages/benefits.aspx>
- Kille, T., Bates, P., & Murray, P. S. (2013). "Farm to table": Air transport issues for an island state. Paper presented at the Australasian Transport Research Forum 2013, Brisbane, Australia.
- Kille, T., Bates, P., & Murray, P. S. (2013). A systematic review of the role of regional aviation in the airfreight industry. *Science Based Technologies*, 4(24), 417–485.
- Lillie, M., & Sparks, L. (1993). The buying behaviour of air freight forwarders. *International Journal of Physical Distribution & Logistics Management*, 23(1), 14–22. doi:10.1108/09600039310025589
- Murphy, P., Dalenberg, D., & Daley, J. (1989). Improving international trade efficiency; airport and air cargo concerns. *Transportation Journal*, 32, 28–35.
- Murphy, P. R., & Hall, P. (1995). The relative importance of cost and service in freight transport choices before and after deregulation: An update. *Transportation Journal*, 35(1).
- Park, Y., Choi, J. K., & Zhang, A. (2009). Evaluating competitiveness of air cargo express services. *Transportation Research Part E, Logistics and Transportation Review*, 45(2), 321–334. doi:10.1016/j.tre.2008.09.004
- Popescu, A., Keskinocak, P., Johnson, E., LaDue, M., & Kasilingam, R. (2006). Estimating Air-Cargo Overbooking Based on a Discrete Show-Up-Rate Distribution. *Interfaces*, 36(3), 248–258. doi:10.1287/inte.1060.0211
- Renting, H., Marsden, T. K., & Banks, J. (2003). Understanding alternative food networks: Exploring the role of short food supply chains in rural development. *Environment & Planning A*, 35(3), 393–411. doi:10.1068/a3510
- Reynolds, J. (2000). E-commerce: A critical review. *International Journal of Retail & Distribution Management*, 28(10), 417–444. doi:10.1108/09590550010349253
- Rundle, T. (1998). *Tasmanian Government: Submission to the International Air Services Inquiry*. International Air Services Inquiry Retrieved from http://www.pc.gov.au/__data/assets/pdf_file/0018/22365/sub041.pdf
- Rutner, S. M., Gibson, B. J., & Williams, S. R. (2003). The impacts of the integrated logistics systems on electronic commerce and enterprise resource planning systems. *Transportation Research Part E, Logistics and Transportation Review*, 39(2), 83–93. doi:10.1016/S1366-5545(02)00042-X
- Saunders, C., & Hayes, P. (2007). *Air freight transport of fresh fruit and vegetables*. Canterbury, New Zealand: Lincoln University.
- Shaw, S. (1993). *Effective Airfreight Marketing*. London: Pitman Publishing.
- Sim, S., Barry, M., Clift, R., & Cowell, S. J. (2007). The Relative Importance of Transport in Determining an Appropriate Sustainability Strategy for Food Sourcing: A Case Study of Fresh Produce Supply Chains. *International Journal of Life Cycle Assessment*, 12(6), 422–431.
- Stank, T., Crum, M., & Arango, M. (1999). Benefits of interfirm coordination in food industry supply chains. *Journal of Business Logistics*, 20(2), 21–41.
- Tseng, Y.-y., Yue, W. L., & Taylor, M. (2005). *The role of transportation in logistics chain*. Paper presented at the Eastern Asia Society for Transportation Studies.
- United States Department of Transportation. (2016). *Rural Airports*. Retrieved 8 January, 2016, from http://www.rita.dot.gov/bts/sites/rita.dot.gov/bts/files/subject_areas/airline_information/rural_airports/index.html

Whipple, J. S., Frankel, R., & Anselmi, K. (1999). The effect of governance structure on performance: A case study of efficient consumer response. *Journal of Business Logistics*, 20(2), 43–62.

Williams, L. R., Magee, G. D., & Suzuki, Y. (1998). A multidimensional view of EDI: Testing the value of EDI participation to firms. *Journal of Business Logistics*, 19(2), 73–87.

Yuan, X.-M., Low, J. M. W., & Ching Tang, L. (2010). Roles of the airport and logistics services on the economic outcomes of an air cargo supply chain. *International Journal of Production Economics*, 127(2), 215–225. doi:10.1016/j.ijpe.2009.08.005

Zhang, A. (2003). Analysis of an international air-cargo hub: The case of Hong Kong. *Journal of Air Transport Management*, 9(2), 123–138. doi:10.1016/S0969-6997(02)00066-2

Zhang, A., & Zhang, Y. (2002). Issues on liberalisation of air cargo services in international aviation. *Journal of Air Transport Management*, 8(5), 275–287. doi:10.1016/S0969-6997(02)00008-X

Zillur, R. (2004). Use of Internet in supply chain management: A study of Indian companies. *Industrial Management & Data Systems*, 104(1/2), 31–41.

KEY TERMS AND DEFINITIONS

Electronic Data Interchange: The electronic interchange of business information within a standardized format. This process then allows one business to send information to another business electronically and without paper.

Logistics Information Systems: A subset of an organization's entire information system, and is focused on the particular issues associated with logistics decision-making.

Category V

Virtual Learning Environments

Creating Active Learning Spaces in Virtual Worlds

Reneta D. Lansiquot

City University of New York, USA

Tamrah D. Cunningham

New York University, USA

Zianne Cuff

City University of New York, USA

INTRODUCTION

Virtual reality can advance educational practices, as creating active learning spaces in virtual worlds offers a more natural way to promote interdisciplinary studies that integrate play to engage students (Lansiquot, 2015, 2012; Lansiquot, Cabo, & Cunningham, 2016; Lansiquot & Cunningham, 2016; Lansiquot, Liou-Mark, & Blake, 2015). Role-play can facilitate engagement in interdisciplinary courses for technology students. This chapter will detail how role-playing in virtual worlds can be used effectively to support interdisciplinary studies and place-based learning, extending learning in undergraduate urban technology classrooms.

As educational technology becomes more of an emergent field and more colleges are progressing towards implementing various forms of technology in their curriculum, it is to be expected that the use of virtual worlds will become more predominant. Many colleges are trying to find the perfect way to integrate virtual worlds in order to create a unique learning tool for students to grasp how vast the virtual space can be. However, virtual space has been used in many other ways for a multitude of purposes that are not limited to just teaching. For digital games, there have been many attempts to utilize virtual worlds in order to create an immersive game where the players can experience the game in a complete way. Virtual Boy, released in the United States

on August 14, 1995 by the Japanese gaming company, Nintendo, and discontinued March 2, 1996, provided a jarring, immersive virtual reality headset that was clunky and considered obsolete quickly. At the time, however, it proved to be a sufficient stepping stone in order to make the changes to create a better excursion into the virtual world. In the present, there is the Oculus Rift, released in 2016, with a newer virtual reality headset which is a strong foray into an all-around immersive look into a game in which the players are avatars who can manipulate the game as they would in real life.

In addition to its use in games, the virtual world and its massive space are used in training simulations for various professions. For example, the United States Army has implemented virtual training models in a bid to offer a more complete and effective training simulation. The benefit of this simulation is that it offers no harm to the soldiers, who can train as long as they want without risking the physical harm that real training can cause. Also, the simulation can be safely managed and edited so the training experience can be varied, and soldiers can learn how to deal with multiple scenarios. Because the virtual world is so vast and moldable, it can take whatever form its creators have in mind. Thus, the virtual world and its endless possibilities as an educational technology tool are the artist's virtual clay—if we can only find the right angle with which to approach it.

DOI: 10.4018/978-1-5225-2255-3.ch685

BACKGROUND

Student learning depends on the ability to create, reflect on experiences, construct understanding, generate rules, and make sense of experiences (Piaget, 1973; Anderson & Krathwohl, 2000; Bruner, 2007). Building on this constructivist perspective, a crucial component of dynamic interdisciplinary studies is problem-based learning, an educational approach that offers the potential to help students develop flexible understanding learning skills. Problem-based learning can also be an effective method for improving classroom engagement (Ahlfeldt, Mehta, & Sellnow, 2005; Duch, Groh, & Allen, 2001). Students learning by solving problems must understand and apply concepts while becoming proficient at new skills. They collaborate to identify what they need to learn in order to solve a problem. They engage in self-directed learning, apply their new knowledge to the problem, and reflect on what they learned as well as the effectiveness of the strategies employed. Rather than simply lecturing and testing students on what was said in lecture, instructors facilitate the learning process by guiding, monitoring, and supporting (Vygotsky, 2006), helping students develop flexible knowledge, successful problem-solving skills, self-directed learning skills, effective collaboration skills, and intrinsic motivation (Hmelo-Silver, 2004).

Student learning can be further improved by adding an interdisciplinary component to courses incorporating collaborative problem solving (Cabo & Lansiquot, 2014; Lattuca, Voigt, & Fath, 2004; Project Kaleidoscope, 2011). Interdisciplinary courses encourage students to connect purposively and integrate cross-discipline knowledge and skills to solve problems in ethically and socially responsible ways. In order to comprehend factors inherent in complex problems, students must synthesize and transfer knowledge across disciplinary boundaries while becoming comfortable with complexity, uncertainty, and varied perspectives. The process

demands that they think critically, communicate effectively, and work collaboratively. An innovative and effective way to address technology learning goals is to facilitate interdisciplinary studies via virtual worlds—multi-dimensional spaces composed of communities of practice in which people can establish a sense of presence, learn, socialize, and collaborate (Lave & Wenger, 1991; Schroeder, 2008; Spence, 2008; Downey, 2012).

These interdisciplinary virtual worlds are so effective because education and technology are continually interconnected, and both possess the ability to transform the world in which we live. Virtual worlds are allowing people from a variety of disciplines to create simulations for purposes that fall directly in line with their teachings. Medical virtual reality groups have devoted time to the study and advancement of simulation technology for clinical purposes, and they explore treating mental health as well as the improving motor and cognitive skills. Not only are virtual realities used for professional purposes, but they are also used to gauge patient progress in, for example, social skills. Research has found that students who struggled to become a part of a class group found it far easier to gain acceptance from their peers through the virtual realm, which made it easier for them to collaborate, instilling confidence in them and increasing their motivation. Where the real world has limitations on what can and cannot be done, the ability to introduce practical knowledge to the classroom through a means of fantastical three-dimensional effects makes the educational experience invaluable. Students are able to engage in a hazardous environment (like a construction site) and get a feel for what should and should not happen under drastic conditions, all in the safety and comfort of a classroom. The process is rewarding, and the reward in terms of self-improvement makes incorporating interdisciplinary virtual worlds all the more worthwhile (Abrosimova, 2014).

VIRTUAL PLACE-BASED LEARNING

Issues, Controversies, Problems

When the idea was first introduced, many instructors found the notion of using commercial virtual worlds appealing, but the excitement for these types of spaces died down as educators felt that their visions for the worlds could not come to fruition. Consequently, higher education institutions decided to design their own spaces in order to improve their contextualizations of collaborative learning in three-dimensional virtual worlds. Open Cobalt (2016), for example, is a free application that caters to the construction and facilitation of game-based learning spaces on which students and educators alike play guest to a network of interconnected collaborative spaces in which they can interact and learn from one another. While it is not nearly as popular as its counterpart virtual engine, Second Life (2016), a plethora of colleges have made use of the world and its facilities. OpenSimulator (2016), another virtual engine available for academic exchanges, provides colleges with the opportunity to access, set up servers, and provide virtual classrooms for learning (Young, 2010).

Second Life remains one of the most frequently used virtual worlds in the educational field because, “as the largest-ever 3D virtual world created entirely by its users,” it enables opportunities for distance learning and real-time communication. The effectiveness of this virtual world is supported by research: A study of student learning preferences in virtual worlds concluded that a great number of students are visual learners, and using Second Life is a viable way to cater to those students by implementing its vast space (Mabrito, 2012). And while this commercial space holds much promise for the promotion of distance learning and the processing of new information in an engaging and rewarding manner for the students, its potential can only be realized if the educators utilize the space properly. Treating the space as if it were just a recreation of a classroom

diminishes this potential by stifling the creative space that both the educators and students access.

Second Life can be a particularly effective virtual world for educating students. Launched in 2003, it is appealing in many ways. Because it is easy to learn, has highly customizable avatars, and users are able to manipulate objects and add interactivity using the Linden Scripting Language (LSL), it seems to be an ideal educational technology. However, the way that a majority of educators have used this space is very limited. Instead of using this virtual world as a means of teaching students by letting them create relevant objects, many educators have recreated their campuses and classrooms, where the most interaction that the students have is interacting with each other and reading lessons from in-world objects (e.g., slides placed on a virtual screen). In the real world, we are bound by natural laws to which we must abide; in virtual worlds, on the other hand, while there are still some laws that exist inherently by the creators of that space, students are otherwise not bound by natural laws. For example, in Second Life, student avatars can change their looks immediately, or they can teleport from one side of the world to the next. Students can even fly! The myriad possibilities seem almost endless in this space, only limited by the imagination of the ones who utilize it. Unfortunately, too often the only worlds that were created were fantasy settings of campuses. Because of this misunderstanding of the possible uses of Second Life, the number of educators who wanted to use it declined considerably. They looked to other virtual spaces that suit their needs or decided to build their own. Others just found that Second Life was too clunky, and the fact that it used its own created LSL instead of a pre-existing language—such as JavaScript or C#—educators decided that it was not worth the hassle of exploring the space. Further, the eclectic community itself was less than ideal for educational purposes. Still, there have been some who have been able to grasp the possibilities of Second Life and use it, for example, to facilitate virtual place-based learning. For example, geoscientists

can use it to allow students in urban classrooms to explore different environments and role-play as scientists climbing up a mountain, collecting data (see Lansiquot, Liou-Mark, & Blake, 2015). Thus, there are still educators who are effectively using this virtual world as a learning space, but recruiting and teaching them to use this is understandably difficult.

Solutions and Recommendations

To graduate with a bachelor's degree, students at New York City College of Technology (City Tech) must take an interdisciplinary team-taught course. Courses such as *Weird Science: Interpreting and Redefining Humanity* allow students to create their vision of the future of humanity. The course explores the enduring question, "What does it mean to be human?" It exposes students to a spectrum of ideas and attitudes from the sciences, technology, and engineering, familiarizing students with concepts from these disciplines. The course highlights complementary perspectives. For example, students are introduced to the anatomy and physiology of the cell while reading nonfiction literature designed to appeal to the imagination (Lansiquot, 2015; Lansiquot, Blake, Liou-Mark, & Dreyfuss, 2011). Students question how different disciplines define concepts of the self and disrupt or disturb seemingly stable identities. The course focuses on being virtually human.

Weird Science utilizes the *Visual Understanding Environment* (2016) for concept mapping, in addition to *Second Life*, which acts as a space for students to get a better understanding of multiple disciplines. By considering the question of what it means to be human, students are tasked with creating physical representations of their opinions. Initially, students were asked to present a possible idea of where humanity may end up in the distant future. Using the knowledge they gained from a dozen guest lecturers from various disciplines—from biology and computer science to economics and sociology—students come up with a plausible path that humanity would follow.

Students are then grouped and asked to build a physical representation of their ideas on City Tech Island, a space in *Second Life*. From there, students create interactive simulations. This is already much different than the way that many educators use virtual worlds. Instead of treating the island as a fantasy recreation of the City Tech campus and having the students sit and read a lecture in one of the many classrooms, essentially creating an online seminar, students are in charge of manipulating the space and using it as a way to express their visions on what they believe is humanity at its highest (or lowest) points in the future. Students test the boundaries of what they can do in this virtual world, and they treat the space as it is meant to be: moldable clay that can be shaped into whatever their imagination can come up with. And as such, City Tech Island has housed many creative ideas: From an underwater civilization to an interactive museum where the avatars are teleported to different recreations of historical sites, all of which occurs when students take full advantage of the virtual space. These projects help demonstrate that students were able to get a grasp of the essence of the class, but there was still a way to go before it could become a stronger and more integrated learning space for the students. Currently, student projects are designed in *Second Life* and spiral out in whatever direction that the student can justify. Although the fact that they could use their imagination was a boon, it was also a detriment, since the students had little restriction on what their futures could be.

FUTURE RESEARCH DIRECTIONS

The second iteration of the group project involves integrating the disciplinary perspectives of the guest lecturers more explicitly. The new project is composed of three parts: the proposal, the engineering of the idea, and the presentation. For the proposal, students present an idea of where they think humanity is going based on the combination of two or more disciplines. For example, one

project could combine the insights of psychology and physics into the idea that, in the future, humanity would develop a way to heighten visual perception to see more into the quantum world, the physics of which, presently, we are not able to see with the naked eye. This could be represented in Second Life by a pair of glasses that avatars can put on, enabling them to see molecules surrounding them as if they were as small as the various subatomic particles. By setting up the project this way, students are able to treat the activity more as interdisciplinary learning, thinking about each discipline and considering how each lecture can complement each other in a way to make a seamless project. Not only that, but the project also relates to their final paper, which is a paper that answers the same enduring question: What does it mean to be (virtually) human?

CONCLUSION

Students need a space where they can learn without fear of penalty, and effective pedagogies for interdisciplinary studies can supply it. An effective approach to teaching should facilitate active learning, foster effective problem-solving strategies, promote progressive collaboration, and provide an environment that requires creation. Virtual worlds offer telepresence, synchronous online learning, as well as opportunities for co-construction, collaboration, and exploring impossible environments. To help students to understand the big picture of complex applications, they should be engaged in process and application early and learn from play, as play involves simpler models of work. This is especially important in complex three-dimensional modeling. Alternatively, simpler models of existing technologies that teach process, collaboration, and creation can be used to engage students before attempting to teach rich, deep, powerful, but complex applications. As a result, students will be able to understand the big idea first. Using virtual worlds with simpler techniques empowers students

to understand process and implementation. An added bonus of this setting is that student success is immediately demonstrable.

These complex modes of thinking were stimulated by the scaffolding of assignments requiring collaborative interdisciplinary studies in Second Life, and simulating the future in Weird Science, a writing-intensive interdisciplinary team-taught course based on the enduring question of what it means to be human. Within this context, research opportunities for students include exploring how best to integrate disciplinary perspectives and represent these in virtual worlds. What are the implications for student engagement, retention, and graduation? Interdisciplinary collaborations enable educators to cultivate an ongoing relationship of learning, dialogue, and challenge. Promoting interdisciplinary studies necessitates creating active learning spaces in virtual worlds to support and engage students.

REFERENCES

- Abrosimova, K. (2014, September 7). 5 ways virtual reality will change education. *Hypergrid Business*. Retrieved May 31, 2016, from <http://www.hypergridbusiness.com/2014/09/5-ways-virtual-reality-will-change-education/>
- Ahlfeldt, S., Mehta, S., & Sellnow, T. (2005). Measurement and analysis of student engagement in university classes where varying levels of PBL methods of instruction are in use. *Higher Education Research & Development*, 24(1), 5–20. doi:10.1080/0729436052000318541
- Anderson, L. W., & Krathwohl, D. R. (Eds.). (2000). *A taxonomy for learning, teaching and assessing: A revision of Bloom's Taxonomy of educational objectives*. New York: Pearson.
- Bruner, J. (2007). *Acts of meaning*. Cambridge, MA: Harvard University Press. (Original work published 1990)

- Cabo, C., & Lansiquot, R. D. (2014). Synergies between writing stories and writing programs in problem-solving courses. In *Proceedings of the 2014 IEEE Frontiers in Education Conference* (pp. 888-896). New York: IEEE. doi:10.1109/FIE.2014.7044133
- Downey, S. (2012). Visualizing taxonomy for virtual worlds. *Journal of Educational Multimedia and Hypermedia*, 21(1), 53–69.
- Duch, B. J., Groh, S. E., & Allen, D. E. (2001). Why problem-based learning? In B. J. Duch, S. E. Groh, & D. E. Allen (Eds.), *The power of problem-based learning* (pp. 3–11). Sterling, VA: Stylus.
- Flower, L. (1993). *Problem-solving strategies for writing* (4th ed.). New York: Harcourt Brace Jovanovich.
- Hmelo-Silver, C. E. (2004). Problem-based learning: What and how do students learn? *Educational Psychology Review*, 16(3), 235–266. doi:10.1023/B:EDPR.0000034022.16470.f3
- Lansiquot, R. D. (2012). Real classrooms in virtual worlds: Scaffolding interdisciplinary collaborative writing. In A. P. Ayala (Series Ed.), *Smart innovation, systems and technologies: Vol. 17. Intelligent and adaptive educational-learning systems: Achievements and trends* (pp. 269-292). New York: Springer.
- Lansiquot, R. D. (2015). Towards open-source virtual worlds in interdisciplinary studies. In M. Khosrow-Pour (Ed.), *Encyclopedia of information science and technology* (3rd ed.; pp. 2647–2653). New York: Information Science Reference. doi:10.4018/978-1-4666-5888-2.ch258
- Lansiquot, R. D., Blake, R. A., Liou-Mark, J., & Dreyfuss, A. E. (2011). Interdisciplinary problem-solving to advance STEM success for all students. *Peer Review*, 13(3), 19–22.
- Lansiquot, R. D., Cabo, C., & Cunningham, T. D. (2016). Playing between the lines: Promoting interdisciplinary studies with virtual worlds. In A. Byers & F. Crocco (Eds.), *The role-playing society: Essays on the popular influence of role-playing games* (pp. 143–162). Jefferson, NC: McFarland.
- Lansiquot, R. D., & Cunningham, T. D. (2016). Conclusion: Integrating interdisciplinary pedagogy. In R. D. Lansiquot (Ed.), *Interdisciplinary pedagogy for STEM: A collaborative case study* (pp. 127–139). New York: Palgrave. doi:10.1057/978-1-137-56745-1_7
- Lansiquot, R. D., Liou-Mark, J., & Blake, R. A. (2015). Learning geosciences in virtual worlds: Engaging students in real-world experiences. In T. J. Bastiaens & G. H. Marks (Eds.), *Education and Information Technology Annual 2015: A Selection of AACE Award Papers* (pp. 223–240). Chesapeake, VA: AACE.
- Lattuca, L. R., Voigt, L. J., & Fath, K. Q. (2004). Does interdisciplinarity promote learning? Theoretical support and researchable questions. *The Review of Higher Education*, 28(1), 23–48. doi:10.1353/rhe.2004.0028
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511815355
- Mabrito, M. (2012, June). Student as avatar: A study of informational preferences in a virtual world class. *MERLOT Journal of Online Learning and Teaching*, 8(2). Retrieved May 31, 2016, from http://jolt.merlot.org/vol8no2/mabrito_0612.htm
- Open Cobalt. (2016). Retrieved May 31, 2016, from <http://www.opencobalt.net>
- OpenSimulator. (2016). Retrieved May 31, 2016, from <http://opensimulator.org>

- Piaget, J. (1973). *To understand is to invent: The future of education*. New York: Grossman.
- Project Kaleidoscope. (2011). *What works in facilitating interdisciplinary learning in science and mathematics: Summary report*. Washington, DC: Association of American Colleges and Universities.
- Schroeder, R. (2008). Defining virtual worlds and virtual environments. *Journal of Virtual Worlds Research*. Retrieved May 31, 2016, from <http://journals.tdl.org/jvwr/index.php/jvwr/article/view/294>
- Second Life. (2016). Retrieved May 31, 2016, from <http://www.secondlife.com>
- Spence, J. (2008). Demographics of virtual worlds. *Journal of Virtual World Research*. Retrieved May 31, 2016, from <http://journals.tdl.org/jvwr/index.php/jvwr/article/view/360>
- Visual Understanding Environment. (2016). Retrieved May 31, 2016, from <http://vue.tufts.edu>
- Vygotsky, L. S. (2006). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press. (Original work published 1978)
- Young, J. R. (2010, February 14). After frustrations in Second Life, colleges look to new virtual worlds. *The Chronicle of Higher Education*. Retrieved May 31, 2016, from <http://www.chronicle.com/article/After-Frustrations-in-Second/64137/>
- Cabo, C., & Lansiquot, R. D. (2013). Development of interdisciplinary problem-solving strategies through games and computer simulations. In R. D. Lansiquot (Ed.), *Cases on interdisciplinary research trends in science, technology, engineering, and mathematics: Studies on urban classrooms* (pp. 268–294). New York: Information Science Reference. doi:10.4018/978-1-4666-2214-2.ch011
- Kelley, T. (2010). Staking the claim for the T in STEM. *Journal of Technology Studies*, 36(1), 2–11. doi:10.21061/jots.v36i1.a.1
- Lansiquot, R. D. (Ed.). (2016). *Interdisciplinary pedagogy for STEM: A collaborative case study*. New York: Palgrave. doi:10.1057/978-1-137-56745-1
- Lansiquot, R. D. (Ed.). (2016). *Technology, theory, and practice in interdisciplinary STEM programs: Connecting STEM and non-STEM approaches*. New York: Palgrave. doi:10.1057/978-1-137-56739-0
- Lansiquot, R. D., Blake, R. A., Liou-Mark, J., & Dreyfuss, A. E. (2011). Interdisciplinary problem-solving to advance STEM success for all students. *Peer Review*, 13(3), 19–22.
- Lipponen, L. (2002). Exploring foundations for computer-supported collaborative learning. In *Proceedings of the Conference on Computer Support for Collaborative Learning* (pp. 72–81). Boulder, CO: International Society of the Learning Sciences.
- Padmanabhan, P. (2008). Exploring human factors in virtual worlds. *Technical Communication (Washington)*, 55(3), 270–276.
- Roush, W. (2007). Second earth. *Technology Review*, 110(4), 38–48.

ADDITIONAL READING

Atkinson, T. (2008). Second Life for educators: Myths and realities. *TechTrends*, 52(5), 26–29. doi:10.1007/s11528-008-0192-3

KEY TERMS AND DEFINITIONS

Digital Game: An electronic creation, recreation, or adaptation of a game; a type of play activity where the participants agree upon a set of rules that they follow and is decided by luck, skill, or strength.

Interdisciplinary Studies: Learning that focuses on questions, problems, and topics too complex or too broad for a single discipline or field to encompass adequately. These studies thrive on making connections among seemingly exclusive domains.

OpenSimulator: An open-source multi-platform, multi-user three-dimensional application server.

Second Life: A three-dimensional virtual world where every avatar is a real person and every place is created by residents.

Virtual Reality: A computer-simulated recreation of a real-world environment that allows users to more fully immerse themselves in the simulation.

Virtual World: Representation in an artificial system or a part of reality.

Visual Understanding Environment: An open-source tool for managing and integrating digital resources in support of teaching, learning and research; it provides a flexible visual environment for structuring, presenting, and sharing digital information.

V

Developments in MOOC Technologies and Participation Since 2012

Jeremy Riel

University of Illinois at Chicago, USA

Kimberly A. Lawless

University of Illinois at Chicago, USA

INTRODUCTION

Massive open online courses (MOOCs) are a recent approach to education, with much buzz generated around them by technology investors, educators, and the news media alike (Dolan, 2014). At the height of MOOC emergence, The New York Times proclaimed 2012 as “The Year of the MOOC,” with MOOCs hailed as the singular approach that would save higher education’s crises and educate the world using Internet-based technologies (Pappano, 2012). However, despite the hype and the subsequent enrollment of millions of learners into MOOCs, the following years experienced a reversal of energy as the results of studies on the effects and costs of MOOCs began to reveal that MOOCs were not a panacea to formal education’s challenges (Dolan, 2014; Liyanagunawardena, Adams, & Williams, 2013). Over the last two years, partnerships between universities and MOOC providers have become strained as outcomes of these courses remained ambiguous (Kolowich, 2013a, 2013b). Now, in 2016, as MOOCs continue to serve millions of students and there are more courses available than before, key questions persist about the efficacy, relevance, and value of MOOCs.

This chapter examines the state of MOOCs between 2012 and 2016 and discusses advances in knowledge around MOOCs that have been observed from empirical research during this period. We build upon our previous review of MOOCs

that was originally written in 2013 (Riel & Lawless, 2015), as a large number of studies focused on MOOC technologies, efficacy, and goals have been published in the short time between 2012 and 2016.

BACKGROUND

Characterizing MOOCs

MOOCs have gained notoriety primarily due to the massive number of participants (DeBoer et al., 2014; Dolan, 2014). While conventional college courses might host hundreds of students, MOOCs have the capability to host hundreds of thousands of students simultaneously with usually just one instructor. However, despite the benefits of reaching a large number of learners, educational researchers have been concerned with the loss of in-person interactions that are known to be beneficial to learning in higher education (Wiebe, Thompson, & Behrend, 2015; Kop, Fournier, & Mak, 2011).

Openness in the case of MOOCs refers to exclusivity or the ability of learners to access the course. In essence, anybody can “attend” the class without having been formally admitted to an educational program, which is one of the primary draws of the MOOC movement (Dillahunt, Wang, & Teasley, 2014; Morgan & Carey, 2009). The “open” policy does not mean all MOOCs are free of cost,

however. Some courses allow anyone to enroll, but charge a fee to receive certain services. For example, MOOC providers often offer premium certificates of completion and dedicated coaching services for a fee. All MOOCs operate on the Internet to some degree. While some MOOCs have been designed to incorporate in-person meetings of students, such as local study groups, all materials and interactivity within a course are generally facilitated online. Many formal college courses now also occur parallel to MOOC courses, with some activities performed in physical classrooms and some performed online (Agarwal & TED, 2014; Collins et al., 2013).

Perhaps the most defining element of a MOOC is that they are centered on a course of study. MOOCs are generally real-time, organized events in which participants are time bound on a course of study around a topic with deadlines and a syllabus (Spector, 2014). This is in contrast to other learning technologies on the Internet, such as articles and text resources, professional learning communities, and tutorials. Participants can ignore course trajectories based on their interests. The diversity of ways in which participants use MOOCs have led to research on course and curricular sequence in MOOCs (Collins et al., 2013; DeBoer et al., 2014).

Technologies Used in MOOCs

Although a MOOC is often referred to as a singular technology, they rely upon a combination of digital applications to achieve their intended functions. These functions serve to deliver content and information to participants, connect participants with other participants and the instructor, and facilitate activities and interactions that promote learning. The digital applications include:

- **Learning Management Systems:** Learning management systems (LMS), also called course management systems (CMS), are web-based applications that are designed to manage and deliver content to partici-

pants within an online course. Many LMS provide a place to host course content, provide an interface for interactive activities for participants, and keep a record of learners' participation.

- **Multimedia Information Delivery:**

MOOCs are predominantly comprised of video-based lectures and reading materials that can be accessed at a student's own pace within the timeframe of a course. Instructors often record lectures they would give in a conventional classroom and distribute these videos to the online course. Instructors also provide background readings or texts, links to websites, images, and other relevant videos or audio.

- **Interactive Elements / Activities for Creativity and Publication:**

Many MOOC instructors intend for participants to be actively engaged in courses instead of simply watching a series of video lectures. When instructors assign projects to students, students have better opportunity to "learn by doing." Active learning is when a student is engaged on creating or interacting with an authentic project related to the course instead of passively receiving information via text or video. Active learning strategies are supported by technologies that promote creating and publishing documents and media, writing blogs, and wikis. Active learning is supported by allowing students to pose questions to one another, share ideas, and collaboratively develop a better understanding of course topics.

- **Notifications:** It is important to provide students with regular reminders and notifications about new content within a MOOC. To address this, many MOOCs provide regular notifications to participants via multiple media (e.g., text/SMS, email, popups) whenever certain events occur within the course (e.g., new content uploads, comments on others' work, availability of a new activity is available).

- **Exams, Quizzes, and Problem Sets:** Students and instructors desire indicators of progress toward course objectives. Much like a conventional course, MOOCs employ quizzes, exams, and problem sets to assess whether participants understand course concepts or can perform certain skills. Many of these evaluations are automated, as well.
- **Progress Tracking and Badges:** Feedback on progress in a course is a helpful indicator for participants. MOOC designers include feedback on a participant's level of interaction in the course by providing badges, which are digital graphics that are "awarded" to participants to indicate completion of course objectives or the degree to which they have participated.
- **Learning Analytics:** Learning management systems typically employ learning analytics that track participant activity and provide statistics on how participants approach course content and activities. Sophisticated learning analytics can include analyses on how long participants view materials, the paths they typically take to complete a task, how frequently and for what purpose they communicate with other participants, and which resources they find valuable.

Interacting With MOOC Technologies

One way that researchers examine the design and efficacy of MOOCs is by investigating the structural aspects of a MOOC. In this common type of research, the focus is on the use and characterization of a particular design element, and researchers seek to explain how and why a design element is used. Researchers investigate the ways in which learners participate with various elements in a MOOC and how these participation patterns are related to psychological, demographic, or environmental variables and indicators. For instance, Milligan, Littlejohn, and Margaryan

(2013) identified similarities between MOOCs and other formats of online learning, with "lurker" or non-participatory behavior occurring in almost all cases of MOOC participants who had opportunities to post assignments, access course materials, or participate in discussions. Perna et al. (2014) emphasize that learner goals predict MOOC element use as well. In their study, they investigated 16 Coursera courses and identified that students' goals and completion predict their progression through courses. Higher numbers of interactions with MOOC elements are also positively correlated with learning skills, as participants who use interactive features are those who report a high level of self-responsibility and established personal goals, have good time management skills, and are consistent in checking the course for updates (Dearnley, 2003; Rodriguez, 2012).

Social interactions are a particular form of participation that are emphasized within MOOCs. It is difficult, however, for instructors to develop learning communities within MOOCs (Gillani & Enyon, 2014). The use of message boards and forums remains very low according to individual MOOC reports, with one study citing that only 3% of participants post to a public forum (Breslow et al., 2013). This number increases for students who complete course objectives, with 52% of course-completing students using public forums. In addition, there was an increased use of forums and message boards during course projects and exams, confirming that already-published material is used by more than just those who actively participate on forums. However, forums in courses are not the only tool for social interaction. An example of this is illustrated with how Shen & Kuo (2015) examined the social media use of MOOC students, particularly those who used Twitter during their courses. Using social media mining techniques, they found that participants' questions and restatements of key ideas were shared more frequently (retweeted), which signified importance. Also, students' sentiment toward a course, as expressed via their social media feeds (e.g., positive, neutral, or negative sentiment), play an important role in

MOOC participation. These studies illustrate that MOOC participation can be done in isolation, but is highly complimented by social interactions and the ability for students to express their questions, ideas, and concerns.

MOOCs are also natural experiment sites for new technologies and interactive features. MOOC platforms typically have the opportunity to collect significant user feedback and participation data as a result of thousands of simultaneous participants in a course. However, novel technologies are not the only interactive elements that have recently been examined. Guo, Kim, and Rubin (2014) explored the assumptions around publishing digital video for online courses, which are a staple of online course communications and activities. By using data from millions of video-watching sessions in MOOCs, they demonstrated that the length, composition, and production value of video matter in MOOC achievement. For instance, they found that the peak length that any video is watched is around 6 minutes, regardless of full length of a particular video segment. In addition, production value on video, such as the quality of filmmaking, lighting, and scripting, was related to achievement. However, they also found that large investments in film quality did not make a difference, instead recommending a middle ground in production quality. Similarly, Li, Kidzinski, Jermann and Dillenbourg (2015) examined fast forwarding and pausing with video. They identified many distinguishable patterns of pausing and fast forwarding based on students' level of expertise with the material under study, indicating that the ways in which students use video reference material might be different based on their incoming academic characteristics. These findings indicate that even the most common MOOC technologies have many opportunities for improvement and that additional supports can be identified, tested, and implemented with learners based on their needs, perhaps dynamically.

Interactions with course materials and elements are the drivers of learning in MOOCs. Scholars and designers agree that engagement in

MOOCs should be promoted in order for learning to occur, but finding the proper combination of interactive design feature, pedagogical approach, and course content remains a challenge (Hew, 2014; Hrastinski, 2009). However, there are many ways to measure engagement with online learning activities (Henrie, Halverson, & Graham, 2015). In a recent study, DeBoer et al. (2014) suggest a focus on specific research variables for studying MOOCs, such as level of participation in activities, achievement, curriculum sequence, and time. These suggested variables leverage the capabilities of MOOCs to capture digital logfiles of participant interactions to better understand how technologies are used for learning and how learners interact with MOOC technologies (DeBoer et al., 2014). With data that potentially spans millions of users, researchers can better identify connections between engagement, goals, and outcomes within MOOCs and other online courses (Henrie, Halverson, & Graham, 2015).

MOOCs and Achievement

Another method of studying MOOCs is to focus on students and the processes of learning within MOOCs, as opposed to examining the technologies themselves. Learning and achievement in MOOCs is most often measured by course completion. However, low completion rates have been a consistent headline concerning MOOCs. A prominent example of low levels of completion occurred in 2013 when San Jose State University partnered with Udacity to provide online-only courses for credit for their students. Low course passage rates in the 20-30% range for three courses in this pilot experiment prompted school officials to suspend any future MOOC courses (Collins et al., 2013). However, at the same institution, a partnership with edX proved to be fruitful. In a "blended course" model in which MOOC content was used in traditional, face-to-face courses in lieu of textbooks, a STEM course that had a 41% failure rate was reduced to 9% during a MOOC-blended pilot program (Agarwal & TED, 2014). While the issue

of high drop rates remains a persistent problem, the initial evidence for using MOOC materials complementarily in a blended, face-to-face course is encouraging to educators.

Completion, however, is not the only indicator of learning. Academic achievement in MOOCs is typically expressed as grades or percentiles in MOOC courses, similar to traditional higher education. In an effort to predict achievement outcomes, Greene, Oswald, and Pomerantz (2015) investigated connections between achievement and various affective and demographic factors in MOOC students. They found that the higher students' personal expectations of how much time they would spend on a course as well as students' level of prior educational experiences were predictors of achievement. Similarly, Terras and Ramsay (2015) examined psychological factors that influence students' achievement in MOOCs, finding that self-regulation skills, perception of time, and level of motivation are all critical factors that lead to achievement. To address limitations of studies that look at a single moment in time, Wang, Paquette, and Baker (2014) studied students' achievement in MOOCs as related to their career growth. They found that students' career goals were persistent indicators of student achievement over time. As students' goals change or are amended, the level of participation and achievement in MOOCs also changes. Wiebe, Thompson, and Behrend (2015) suggested that students' goals and motivations should be a primary focus in MOOC studies due to these factors' strong influences on learning and MOOCs' open invitation to learners with diverse interests.

These studies illustrate that cognitive skills such as goal setting, time keeping, and setting expected levels of effort, as well as affective factors such as motivation and interest all play significant roles in MOOC achievement. These findings also indicate that students who are used to traditional higher education methods and have done well in traditional higher education are more likely to succeed in a MOOC. This can be problematic as MOOCs attempt to reach potential students who

have not been highly represented in traditional education.

Who Uses MOOCs?

At their rise, the key promise of MOOCs was to open access to learning materials and courses to people who have otherwise been unable to participate in formal higher education. In the last two years, multiple studies have been published that shed light on the fulfillment of this goal. In early assessments of MOOC participation in individual courses, students from countries from all over the world were represented. In a study of edX's first STEM-focused MOOC by Breslow et al. (2013), students from 194 countries participated. In the same study, enrollment in this MOOC (a course on electronic circuits) was predominantly male (86%), the median student age was typically between 20 and 30, and no significant relationships were found between gender, age, and course completion. Additional studies of other STEM related MOOCs echo these patterns (Belanger & Thornton, 2013; Ho et al., 2014). However, in studies on non-STEM-oriented courses, higher levels of female participation were observed (Collins et al., 2013; Ho et al., 2014). This evidence suggests that course material follows general societal trends in levels of participation by gender, despite being openly available. As they mature, MOOCs may serve a role in increasing access to and interest of subject areas that traditionally see low levels of participation by female and older students.

In addition to increasing opportunities for broadening educational access based on gender, it is often cited that MOOCs have the potential to be particularly transformative for learners in countries other than America, where most MOOC providers are headquartered. Although students participate in MOOCs from all over the world, however, most MOOC content is provided only in English (Agarwal & TED, 2014; Breslow et al., 2013; Ho et al., 2014). To address the issue of language, the major MOOC providers integrate captioning technologies into their platforms. For

instance, edX's platform simultaneously displays a transcript alongside video as it is playing, and users can toggle between available languages. Morgan and Carey (2009) identified some benefits of international participation in a study of rural students who had difficulty accessing brick and mortar higher education campuses in Mexico, Russia, and Canada. They found that students in the study strongly benefited from interacting with students from other cultures and economic backgrounds. Dillahunt, Wang, and Teasley (2014) performed a similar study of underrepresented international students in a MOOC by asking questions about completion of MOOC courses. They found that although the students in underrepresented groups did not complete MOOC courses at a higher rate than students in traditionally represented groups, the students in underrepresented groups that did complete the course completed with distinction at the highest rates. These studies illustrate that international students are motivated to participate in MOOCs, but much work still needs to be done to achieve the potential of far-reaching MOOCs.

FUTURE RESEARCH DIRECTIONS

Studies on MOOCs have been published in prominent academic journals in recent years as data on them became available. Common among recent research are case studies and institutional reports of individual MOOC courses (e.g., Belanger & Thornton, 2013; Breslow et al., 2013; Collins et al., 2013; Ho et al., 2014; Liyanagunawardena, Adams, & Williams, 2013). Due to the large number of single-course case studies by 2013, the research community encouraged more robust investigations of relationships between student learning and the various design aspects of MOOCs (Velesianos & Sheperdson, 2015; Wiebe, Thompson, & Behrend, 2015). As technologies continue to improve each year, future studies should focus on understanding the behavioral and participation patterns of students (DeBoer et al., 2014; Perna et al., 2014), the learning processes of students in MOOCs

(Wiebe et al., 2015), and the design and efficacy of student supports in MOOCs (Dearnley, 2003; Kop, Fournier, & Mak, 2011).

In addition to researchers, grantmaking institutions also play a significant role in the future of MOOC research. To examine the priorities of grantmaking institutions, Gašević et al. (2014) studied 266 proposals submitted to the MOOC Research Initiative in 2013 to identify the directions that MOOC research was heading at the time. They found that the most frequent proposals were those that investigated social networks and communities underlying MOOCs. In the case of social technologies, MOOCs are an excellent context for research due to their online nature. They also identified other types of proposals for MOOC research, which included ideas for examining student behavioral patterns, technology improvements, learning processes, and the integration of MOOCs into higher education institutions. All of these trajectories continue to be productive areas of MOOC research today, as new interactive capabilities continue to be developed.

As a result of an emerging body of research around MOOCs, education scholars have suggested some guidelines in conducting future MOOC research. Raffaghelli, Cucchiara, and Persico (2015) examined the methodological strengths and shortcomings of recent peer-reviewed MOOC research and found that most studies over the last five years rely on theoretical analyses and case studies that span across multiple epistemological perspectives in a somewhat incoherent fashion. In order to promote consistency in future studies, they suggest that researchers consider approaches to MOOC research that consider how valid theoretical perspectives are connected, such as simultaneous foci on technology and on learners. However, Wiebe, Thompson, and Behrend (2015) argue that researchers must always consider the learning processes and experiences of the student, since learning is the primary intent of a MOOC. From their argument, studies should not only focus on singular variables, but also consider students as a whole unit in order to get a clearer picture. In other

words, there is no one way to study MOOCs and the research community will only get a full picture with multiple perspectives that are woven together. In particular, interdisciplinary approaches that examine MOOCs from multiple perspectives are growing and will be particularly valuable in the future (Veletsianos & Sheperdson, 2015).

CONCLUSION

MOOCs have been primarily studied in two ways. In one approach, researchers focus on the role, effects, and structural traits of MOOC technologies, and investigate participant interactions with particular technologies. In the second approach, researchers examine the processes associated with learning within MOOCs as a context. Researchers of this approach are concerned with psychological, social, and behavioral aspects of MOOC participants. Neither approach is the “right way” to conduct studies, and both are necessary to fully understand the complexity of MOOC learning (Weibe, Thompson, & Behrend, 2015).

Researchers have significantly advanced the field of MOOC scholarship in the last four years. The latest findings on learners and MOOC technologies suggest that MOOC developers and teachers should promote higher levels of participation in activities, as learning is an active process and students need meaningful interactions in order to learn (Hew, 2014; Kop, Fournier, & Mak, 2011; Milligan, Littlejohn, & Margaryan, 2013). However, designers should carefully consider the activities that participants will be expected to do in a MOOC. Activities should be purposefully chosen to match the content being studied and ensure that they are pedagogically appropriate to support student learning. Since behaviors are connected to psychological and social processes underlying learning, online learning participation patterns in their various forms provide an excellent, low-cost window to understanding the learner (Hrastinski, 2009).

MOOCs have come a long way in the last five years, but much work still remains to be done to ensure that technologies are developed to empirically improve learning in MOOCs. “The Year of The MOOC” has come and gone, and many benefits have been demonstrated in subsequent studies. However, MOOCs have not yet achieved the goal of increasing access of people from socio-economic backgrounds that do not traditionally have high levels of access to MOOCs. Course providers should increase their efforts to promote MOOCs to students other than those who traditionally attend higher education. Promoting MOOC access includes designing interfaces that meet the needs of underrepresented students. It is imperative that MOOCs do not simply serve as low-cost replicas of classrooms by using flashy, novel technologies. To simply mimic traditional classroom practices could risk the exclusion of the the best evidence-based practices of teaching and learning.

REFERENCES

- Agarwal, A., & TED. (2014). Anant Agarwal: Why massive open online courses (still) matter. *TED Talks*. Retrieved January 31, 2014, from http://www.ted.com/talks/anant_agarwal_why_massively_open_online_courses_still_matter.html
- Belanger, Y., & Thornton, J. (2013). *Bioelectricity: A quantitative approach*. Report. Duke University. Retrieved from http://dukespace.lib.duke.edu/dspace/bitstream/handle/10161/6216/Duke_Bioelectricity_MOOC_Fall2012.pdf
- Breslow, L., Pritchard, D. E., DeBoer, J., Stump, G. S., Ho, A. D., & Seaton, D. T. (2013). Studying learning in the worldwide classroom research into edX’s first MOOC. *Research & Practice in Assessment*, 8, 13–25.

- Collins, E. D., Firmin, R., Schiorring, E., Whitmer, J., Willett, T., & Sujitparapitaya, S. (2013). *Preliminary summary: SJSU+ augmented online learning environment pilot project*. Retrieved from [http://www.sjsu.edu/chemistry/People/Faculty/Collins_Research_Page/AOLE Report -September 10 2013 final.pdf](http://www.sjsu.edu/chemistry/People/Faculty/Collins_Research_Page/AOLE%20Report%20-%20September%202013%20final.pdf)
- Dearnley, C. (2003). Student support in open learning: Sustaining the process. *International Review of Research in Open and Distance Learning*, 4(1). doi:10.19173/irrodl.v4i1.132
- DeBoer, J., Ho, A. D., Stump, G. S., & Breslow, L. (2014). Changing course: Reconceptualizing educational variables for massive open online courses. *Educational Researcher*, 43(2), 74–84. doi:10.3102/0013189X14523038
- Dillahunt, T., Wang, Z., & Teasley, S. D. (2014). Democratizing higher education: Exploring MOOC use among those who cannot afford a formal education. *International Review of Research in Open and Distance Learning*, 15(5), 177–196. doi:10.19173/irrodl.v15i5.1841
- Dolan, V. L. B. (2014). Massive online obsessive compulsion: What are they saying out there about the latest phenomenon in higher education. *International Review of Research in Open and Distance Learning*, 15(2), 268–281. doi:10.19173/irrodl.v15i2.1553
- Gašević, D., Kovanovic, V., Joksimovic, S., & Siemens, G. (2014). Where is research on massive open online courses headed? A data analysis of the MOOC Research Initiative. *International Review of Research in Open and Distance Learning*, 15(5), 134–176. doi:10.19173/irrodl.v15i5.1954
- Gillani, N., & Eynon, R. (2014). Communication patterns in massively open online courses. *The Internet and Higher Education*, 23, 18–26. doi:10.1016/j.iheduc.2014.05.004
- Greene, J. A., Oswald, C. A., & Pomerantz, J. (2015). Predictors of retention and achievement in a massive open online course. *American Educational Research Journal*, 52(5), 925–955. doi:10.3102/0002831215584621
- Guo, P. J., Kim, J., & Rubin, R. (2014). How video production affects student engagement: An empirical study of MOOC videos. In *Proceedings of the first ACM conference on Learning @ scale conference - L@S '14*. doi:10.1145/2556325.2566239
- Henrie, C. R., Halverson, L. R., & Graham, C. R. (2015). Measuring student engagement in technology-mediated learning: A review. *Computers & Education*, 90, 36–53. doi:10.1016/j.compedu.2015.09.005
- Hew, K. F. (2014). Promoting engagement in online courses: What strategies can we learn from three highly rated MOOCs. *British Journal of Educational Technology*. doi:10.1111/bjet.1223
- Ho, A. D., Reich, J., Nesterko, S., Seaton, D. T., Mullaney, T., Waldo, J., & Chuang, I. (2014). *HarvardX and MITx: The first year of open online courses*. HarvardX and MITx working paper no. 1.
- Hrastinski, S. (2009). A theory of online learning as online participation. *Computers & Education*, 52(1), 78–82. doi:10.1016/j.compedu.2008.06.009
- Kolowich, S. (2013a, April 29). Why some colleges are saying no to MOOC deals, at least for now. *The Chronicle of Higher Education*.
- Kolowich, S. (2013b, May 2). Why professors at San Jose State won't use a Harvard professor's MOOC. *The Chronicle of Higher Education*.
- Kop, R., Fournier, H., & Mak, J. S. F. (2011). A pedagogy of abundance or a pedagogy to support human beings? Participant support on massive open online courses. *International Review of Research in Open and Distance Learning*, 12(7), 74–93. doi:10.19173/irrodl.v12i7.1041

- Li, N., Kidzinski, L., Jermann, P., & Dillenbourg, P. (2015). MOOC video interaction patterns: What do they tell us? In *Design for teaching and learning in a networked world* (pp. 197–210). New York: Springer International. doi:10.1007/978-3-319-24258-3_15
- Liyanagunawardena, T. R., Adams, A. A., & Williams, S. A. (2013). MOOCs: A systematic study of the published literature 20082012. *International Review of Research in Open and Distance Learning*, 14(3), 202–227. doi:10.19173/irrodl.v14i3.1455
- Milligan, C., Littlejohn, A., & Margaryan, A. (2013). Patterns of engagement in connectivist MOOCs. *Journal of Online Learning and Teaching*, 9(2).
- Morgan, T., & Carey, S. (2009). From open content to open course models: Increasing access and enabling global participation in higher education. *International Review of Research in Open and Distance Learning*, 10(5), 1–16. doi:10.19173/irrodl.v10i5.632
- Pappano, L. (2012, November 12). The year of the MOOC. *The New York Times*. Retrieved from http://www.nytimes.com/2012/11/04/education/edlife/massive-open-online-courses-are-multiplying-at-a-rapid-pace.html?pagewanted=all&_r=0
- Perna, L. L., Ruby, A., Boruch, R. F., Wang, N., Scull, J., Ahmad, S., & Evans, C. (2014). Moving through MOOCs: Understanding the progression of users in massive open online courses. *Educational Researcher*, 43(9), 421–432. doi:10.3102/0013189X14562423
- Raffaghelli, J. E., Cucchiara, S., & Persico, D. (2015). Methodological approaches in MOOC research: Retracing the myth of Proteus. *British Journal of Educational Technology*, 46(3), 488–509. doi:10.1111/bjet.12279
- Riel, J., & Lawless, K. A. (2015). Massive open online courses (MOOCs) and the technologies that support learning with them. In M. Khosrow-Pour (Ed.), *Encyclopedia of Information Science and Technology* (3rd ed.; pp. 7529–7537). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-5888-2.ch741
- Rodriguez, C. O. (2012). MOOCs and the AI-Stanford like courses: Two successful and distinct course formats for massive open online courses. *European Journal of Open, Distance, & E-Learning*, 2, 1–14.
- Shen, C., & Kuo, C.-J. (2015). Learning in massive open online courses: Evidence from social media mining. *Computers in Human Behavior*, 51, 568–577. doi:10.1016/j.chb.2015.02.066
- Spector, J. M. (2014). Remarks on MOOCs and Mini-MOOCs. *Educational Technology Research and Development*, 62(3), 385–392. doi:10.1007/s11423-014-9339-4
- Terras, M. M., & Ramsay, J. (2015). Massive open online courses (MOOCs): Insights and challenges from a psychological perspective. *British Journal of Educational Technology*, 46(3), 472–487; Advance online publication. doi:10.1111/bjet.12274
- Veletsianos, G., & Shepherdson, P. (2015). Who studies MOOCs? Interdisciplinarity in MOOC research and its changes over time. *International Review of Research in Open and Distributed Learning*, 16(3), 1–17. doi:10.19173/irrodl.v16i3.2202
- Wang, Y., Paquette, L., & Baker, R. (2014). A longitudinal study on learner career advancement in MOOCs. *Journal of Learning Analytics*, 1(3), 203–206. doi:10.18608/jla.2014.13.23
- Wiebe, E., Thompson, I., & Behrend, T. (2015). MOOCs from the viewpoint of the learner: A response to Perna et al. (2014). *Educational Researcher*, 44(4), 252–254. doi:10.3102/0013189X15584774

KEY TERMS AND DEFINITIONS

Active Learning: Experiences that require learners to actively participate in activities, to reflect on their experiences, and to interact with technology and other people in order to develop a deeper understanding of a subject.

Badges: Digital graphic icons that are “awarded” to participants in online spaces to indicate degrees of participation within the environment or completion of objectives within the environment.

Certificates of Completion: Documents issued by MOOC providers that certify that a

participant has successfully completed a course, usually for a fee.

Learning Analytics: Combined automatic data collection and analysis that tracks participant activity and provide statistics on how participants approach course content and activities.

Learning Management System (LMS): Applications that are designed to manage and deliver content to participants within an online course.

Massive Open Online Course (MOOC): An online educational course that is open to anyone to participate, often free of cost.

V

Massive Open Online Courses and Integrating Open Source Technology and Open Access Literature Into Technology-Based Degrees

Maurice Dawson

University of Missouri – St. Louis, USA

Sharon Burton

Grand Canyon University, USA

Dustin Bessette

National Graduate School of Quality Management, USA

Jorja Wright

University of Charleston, USA

INTRODUCTION

The purpose of this chapter is to (1) understand MOOCS, massive online open courses, and differentiate MOOCS from other like learning and training methodologies, (2) comprehend the manner in that MOOCS can be utilized, as well as the (3) lasting impacts of MOOCS on learning whether in academics or the corporate environment. MOOCs are aimed at large scale participation in online education (Bud, Smith, & Reisman, 2015). This chapter will identify for educators, administrators, and practitioners the staggering awareness of the outcomes of this learning modality. MOOCs are continuing to change the way individuals receive education and learn. One idea of MOOCs is that learners may learn through collective education using what is believed to be a form of experimentation wrapped in andragogy, adult learning, andragogy. Education through Knowles' adult learning, offers the notion of learners gaining knowledge and understanding through (a) self-directedness, (b) need to know, (c) use of experience in learning, (d) readiness to learn, (e) orientation to learning, and (f) internal moti-

vation (Knowles, 1979; Taylor & Kroth, 2009). Another view is that MOOCs could be the answer to the economic concerns faced by institutions of higher learning, the push to increase learner enrollment, and the drive to enhance graduation completion rates (Ng'ambi & Bozalek, 2015). Administrators, educators, and practitioners are faced with the increased popularity of MOOCs. The popularities of MOOCs include the positives as well as the concerns: their impersonal nature, numerous students enrolled into an individual section of a MOOC course, instructors serving as facilitators, as well as there being no instructors assigned to courses. MOOCs, post-date open courseware which was on the increase in 1990s and the sharing of courseware online. Specifically, open courseware stems from face-to-face instructions wherein instructors share aspects of their courses on the Internet which may be reading lists, assignments, recorded videos, audio lectures, or syllabi. One of the understood top aspects regarding MOOCs is that these courses fulfil knowledge gaps for learners. MOOCs propose for the learners the alternative of completing quality courses without a financial obligation. Some MOOCs

DOI: 10.4018/978-1-5225-2255-3.ch687

courses, like Coursera, offer certificates. With this said MOOCs offer inexpensive avenues to add certificates to learners' resumes and therefore ways to drive personal accomplishments. MOOCs have maintained a large following of users. They draw attention and remain the element learners need to study and become educated outside of the traditional classroom.

This chapter will delve into key significant areas. First, the researchers will gain understanding regarding the development of a method in which Open Source Software (OSS), open source technologies, and open access literature can be incorporated to strengthen the MOOC environment. Second, readers will garner comprehension of how to strengthen the MOOC environment and therefore increase retention in addition to increasing enrollment in higher education. Third, the lasting impacts of MOOCs will be described, mainly as the impacts relate to STEM, Science, Technology, Engineering, and Mathematics and how such programs are entering the online space. Through this chapter, educators and administrators will grasp an increased interpretation of the technologies that have influenced learning and development. Administrators and Practitioners will learn meaningful solutions about open technology solutions so that as-is and to-be MOOCs can be improved with minimal cost. Further readers, researchers, administrators, and practitioners will learn how to infuse MOOCs learning and training initiatives to ensure best outcomes in the academic and e business world.

BACKGROUND

MOOCs is another form of distance education and learning community; the term originated in 2008 by David Cornier and senior research fellow Bryan Alexander (Martin, 2010). At that time, MOOCs described a course entitled *Connectivism and Connective Knowledge* (Pence, 2012). According to Caulfield, Collier, and Halawa (2013), MOOCs

stem from the category of online connected learning communities. MOOCs are representative of Siemens' and Downes' connectivist learning theory, which provides that learning occurs through connections within networks (Pence, 2012). The model uses a network with nodes and associations to explain learning (Siemens, 2013). The name MOOC was derived because (a.) MOOCs are collaborative with the appearance of experimentation, (b.) the courses move along numerous paths, and (c.) MOOCs necessitate participation in "Massively Multiplayer Online Game (MMOG)" (Pence, 2012, p. 27).

MOOCs, a recent innovation in the distance education field, were labeled as disruptive innovation (Christensen, 1997). According to Flynn (2013), MOOCs describe the varied kinds of changes in an evolving business education environment. The concept bracing MOOCs began in the 1960s; however, this approach to distance education re-ignites because of the digital education age (Flynn, 2013). As posited by Ta'eed (2012), Khan Academy was the first to house the free lectures. The Bill Gates Foundation and Google supported Khan Academy, a non-profit, with substantial backing in 2006. Today Khan Academy has over 3000 short video lectures, which were stated to have over 160 million hits (Severns, 2014; Ta'eed, 2012).

Understanding MOOCs and their relevance is important as MOOCs meaningfully progressed online distance education in past years (Rodriguez, 2012). Standing as an online distance educational model, MOOCs, are supported by colleges and universities; therefore, the need exist for faculty to understand and augment their knowledge regarding MOOCs' past, current and future trends. MOOCs have several prescribed frameworks: cognitive-behaviorist, connectivist social and constructivist (Rodriguez, 2012), connectivism (Pence, 2012), constructivist, connectivist experiential, and social structure, (Koutropoulos et al., 2012). The postulations of these researchers are that MOOCs were formed by technologies

and reside beneath the online distance education environment.

FACULTY DEVELOPMENT IN MOOCs

Through the lens of MOOCs, readers, practitioners, and academicians can postulate the reasons MOOCs emerged on the distance education platform. As educators continue to concentrate on the online distance education environment, comprehending MOOCs is yet another lane in learning for meditation and conjecture. Faculty development with MOOCs has the potential to facilitate networking and permit instructors to converse and discuss concepts and experiences. Professors interested in MOOCs can add value to their repertoire of knowledge by engaging in MOOC mentoring, coaching, and job shadowing programs. Mentorships are associations encompassing a seasoned faculty member who guides faculty with minimal experience or faculty with no experience (Tareef, 2013). Coaching, as given by Rahim and Burrell (2013), is “having an ability to help an employee who has raw talent develop into a star performer”. Job shadowing involves investigating the actions of a professional systematically, and then including the desired actions into one’s work procedures (Leonard, Barton, & Barton, 2013).

Professional development within institutions of higher education remains a significant variable driving excellence of education learners’ academic experiences (Marshall, 2012; Mosley, 2007). Every institution of higher education should establish if, how, and when MOOCs will hold an academic space in their educational lineup. Faculty should comprehend the evolution of MOOCs into and around their institutions. According to Demirci (2014), the syllabus of MOOCs is very similar to other syllabus. Further learning activities are developed to accommodate large audiences. The lessons and activities can be organized as asynchronous and with flex-

ibility (Demirci, 2014). Faculty should look to faculty development programs, mentorships, and coaching to gain additional information regarding MOOCs and their impact.

HOW ARE MOOCs BEING SUPPORTED

Elite institutions of higher education are supporting MOOCs. According to Thrift (2013), an obsession exists with MOOCs because (a) institutions of higher learning are enticed by the promise of profit in educational environments, (b) middle-class aggravation over tuition costs is fueled, (c) economic downturn is propelling countries to seek better priced education that is valued, and (d) MOOCs seem to provide a more efficient mode of education. Pence (2012), provided, the thought remains that partaking in MOOCs may gather meaningful stature. Another advantage of MOOCs provided by Pence (2012) is that MOOCs can produce analytics about education. Analytics can support a drive for operational sustainability. Data regarding MOOCs is found primarily through blog posts (Yong, 2014).

OFFERING MOOCs

Led by Sebastian Thrun, in 2011 Stanford University offered a MOOC. Thrun’s postulation was that eventually participants would graduate with degrees with MOOCs on the transcripts (Krause, 2014). Stanford’s MOOC enrollment was over “160,000 students from 190 different countries” (Flynn, 2013, p. 152). According to Krause (2014), 28,000 completed the courses. The MOOC project expanded. Thrun left Stanford to start Udacity, a for-profit company (Flynn, 2013, Krause, 2014). The goal was to make education a democracy by providing it to learners at no cost (Flynn, 2013). Udacity offers STEM courses, Science, Technology, Engineering, and Math.

Motivated by Thrun, Andrew Ng, and Daphne Koller, two other Stanford faculty members co-founded Coursera, a for profit company (Krause, 2014). Coursera functioned akin to an educational cooperative between Stanford University, University of Pennsylvania, University of Michigan, and Princeton University (Flynn, 2013). According to Krause, Coursera and Stanford, now partners with “33 high caliber universities” (Decker, 2014, p. 7). Coursera courses start frequently and are available in a wide range of topics.

Articulated by Flynn, (2013), \$30 million each in institutional backing, funding of grants, and resources was donated by MIT and Harvard to launch edX, a MOOC. This not-for-profit initiative started with 32 charter members (PR, N, 2014). Harvard is the oldest institution of higher learning in the United States, established in 1636. Harvard’s support of MOOCs is significant. Kolowich (2013a), states that Harvard is very careful with its MOOCs partnerships. Since its beginning, edX has continuously progressed and offers courses on an array of subjects. MIT offers MOOC curricula (Kolowich (2013b).

Initially opposed to MOOCs, universities like Duke University decided against the MOOCs revolution. Disclosed by Kolowich (2013c), the faculty voted the measure down after Duke’s provost, Peter Lange, signed a contract with 2U, a MOOCs group. The *Chronicles of Higher Education* article provided that the faculty’s vote superseded the contract signed by Lange. Now Duke is a part of the MOOCs revolution and advertises MOOCs on its online website. Another institution that initially rejected MOOCs was Amherst College, a liberal arts institution. April 2013, Amherst College’s faculty rejected Harvard’s edX invitation to become a part of the elite association of higher education institutions providing MOOCs (Kolowich, 2013a). During this time, the academe world had mixed views about the MOOCs form of distance education. Today, when viewing Amherst’s web site, MOOCs are now advertised.

MOOCS CRITIQUED

Many elite colleges and universities have supported MOOCs. Some people see MOOCs as innovative and the future of education. Other people understand MOOCs to be all that is wrong with education. What is known is that MOOCs are receiving critiques. According to Rees (2014), MOOCs are akin to students going to the library to check out educational materials and then reading and studying the materials. Some believe that MOOCs will destroy today’s understanding of higher education (Fox, 2013). Others document that the course completion rate for MOOCs is too low (Guzdial, 2013). Questions are surfacing as to whether MOOCs will devalue instructors in the realm of education. The concern of sustaining revenue and not offering too many MOOCs is just another concern. What is understood about MOOCs is that they are perceived as a business model like other educational delivery models.

The MOOCs model was founded on the idea that institutions of higher of learning are the developing segments in line for the high-capacity, low-perimeter data and information-technology treatment after cost review, sales, and the media, as given by Thrift (2013). MOOCs are accessible and accommodate large numbers of learners. MOOCs support lifelong learning. In essence, MOOCs characterizes emerging technology whose educational impact has yet to be fully understood. As shown through Amherst’s and Duke’s later embracing of MOOCs, this form of online distance education will continue to undertake in the practices of higher education.

MOOCS IMPLEMENTATION AND CHANGE

The MOOCs phenomenon has driven an ever-increasing conversation regarding online courses whether in higher education, or business. These developments must be explored from an institutional policy standpoint, focusing on an examina-

tion of the different approaches to MOOCs and e-learning. MOOCs implementation represents change. This change points to (1) curriculum and course design and redesign, (2) a sound e-learning policy as a medium for curriculum development; (3) an emphasis on technology usage; (4) specific policies around licensing; (5) how MOOCs support student advancement; and (6) overall educational enhancement that drives student success.

The change to incorporate MOOCs requires a paradigm shift, a new approach to thinking about online educational processes and technologies. In order for MOOCs to be administered successfully, a top down approach is needed (Burton, 2016). Because people in like organizations will exhibit like behaviors, there must be a change leader to drive communication up and down the communication chain. This leader has to know the key organizational factors: understanding the vision of the implementation, ensuring accountability of key roles for the MOOCs implementation, ensuring the appropriate stakeholders' involvement, determining what skills need to be developed (e.g., developer, staff, students, employees, etc.), ensuring the appropriate metrics, policies and actions, as well as determining how to reinforce behaviors. Initiating the change will require an understanding of the MOOCs implementation range and scope of the project. All leaders and project participants must understand all difficulties and intricacies involved with MOOCs development, and implementations. Timeframe be known for the development and implementation. Most of all, to make the implementation work, all involved must buy-in to the idea.

MOOCs are a disruptive form of education. However; due technology driving learning to the anytime and anyplace environment, the conversation about MOOCs has change. The new conversation need to be about how this form of education will progress leaning institutions and organization toward business process improvement. MOOCs can be implemented as change to support the learning institutions and organizations broader strategic objectives.

OPEN ACCESS LITERATURE

Open access is digital, online, free of charge for everyone with an internet connection, free of most copyright and licensing restrictions (Suber, 2004). Technology has opened many doors and opportunities for the current needs for many academicians, researchers, and practitioners based on the need to have open access availability for literature. The possibility of restricting literature in the future will come at a cost that will be indefinitely become a nuisance for businesses and schools to utilize. MOOC programs of today already are facing the possibility of creating online learning environments that only enhance and filtrate open access literature and learning. Open access has even made it possible for a vast majority of scholars to share theories, research and information such that it can be primarily used in an online database of journals for future reference and research.

Proponents of open access publishing sometimes argue their case on the fairness of providing everyone with access to scholarly material, and by so doing, creating a scientific/academic level-playing field (Nicolas, Huntington, & Jamali, 2007). Open access publishing has become a well known publishing and research style that many scholars, faculty, and practitioners are beginning to position themselves around based on the possibility of other scholars referencing and citing their research and work. This is the ultimate goal of open access literature. With better availability and access to vital research, data can be disseminated to more outside sources to create a better understanding of global research. Therefore, the literature of today needs to be adaptable and accessible for communities and scholars of tomorrow. Creating a validation for open access literature is a primary need for many societies to have based on the utilization of tools in the virtual and traditional classrooms. The tools and styles used are practices of students in virtual learning environments (Bessette & Burton, 2013). The learning environments of the MOOCs are ideally using and utilizing online open access literature and references such that it is readily

available to more people today than ever before. Libraries with traditional literature are becoming outdated as technology for academic advancement is taking place.

The age of the electronic journal has also brought another problem to the library (Chang, 2006). Traditional libraries of today are being more advanced to replace the current physical copy of journals and textbooks and utilizing technology to the max. This adaptation is creating a faster advancement and movement for the retrieval process of literature. More scholars, students, researchers, and practitioners are requiring access 24/7. The increase of the electronic journal and the open access journal are beginning to develop a keen sense of adaption for research to be published and researched at any time in any place. The need for high speed internet access has also become part of the availability and usefulness for online open access literature.

Scholars, faculty, researchers, and practitioners need to coordinate and publish exclusively in all types of scholarly journals. These journals also need to be read and referenced accordingly in

order to successfully develop keen communication between scholars of different disciplines. Without scholastic publishing and referencing of open access work, scholarly research will be under-utilized and problems will persist. Costs of subscriptions for scholastic journals will increase, creating a crisis that is also known as serials crisis. To prevent the “Serials Crisis” situation from becoming worse, research organizations and libraries are starting to support open access publishing (Chang, 2006). Every academic discipline needs to currently outweigh the cost for using and utilizing open access literature based that all research needs to have an open accessible and ideal use based on the needs of scholars for the future.

The utilization of MOOCS into the open access literature will deviate cognitive learning and skill development for students. It is then possible to assume that students will accelerate their learning capabilities to such a level that is currently unmeasured. The need for the open accessibility and developmental change is in the hands of academicians today. MOOC classes will invite,

Figure 1. Screenshot of Directory of Open Access Journals

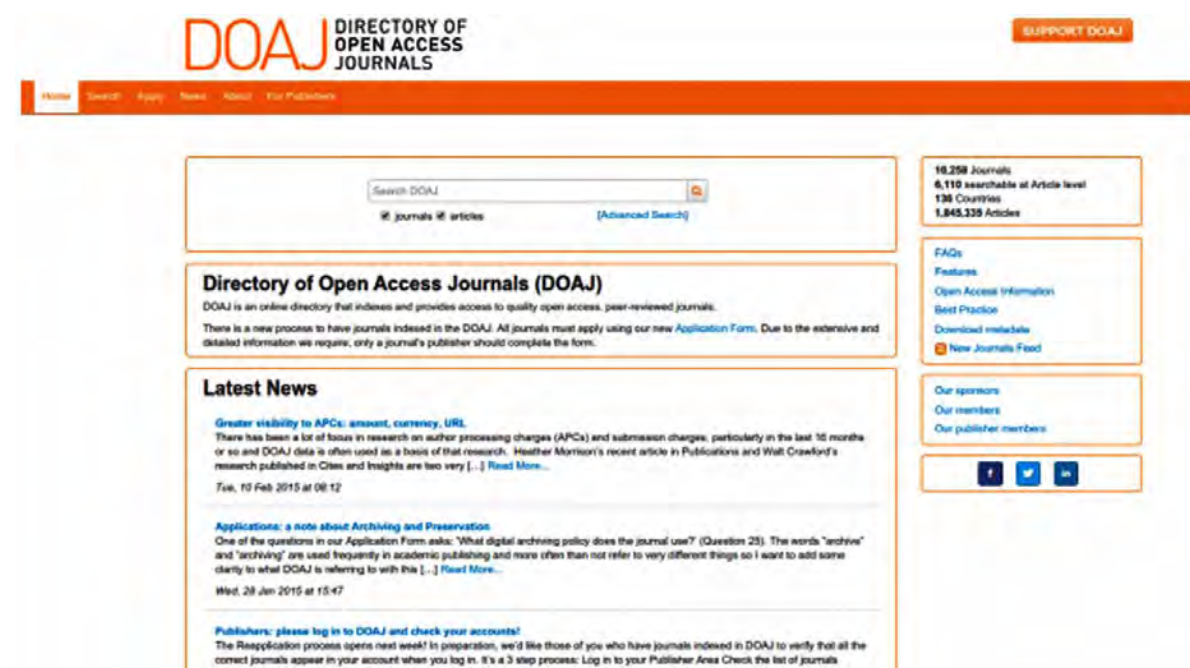


Table 1. Linux distributions and uses

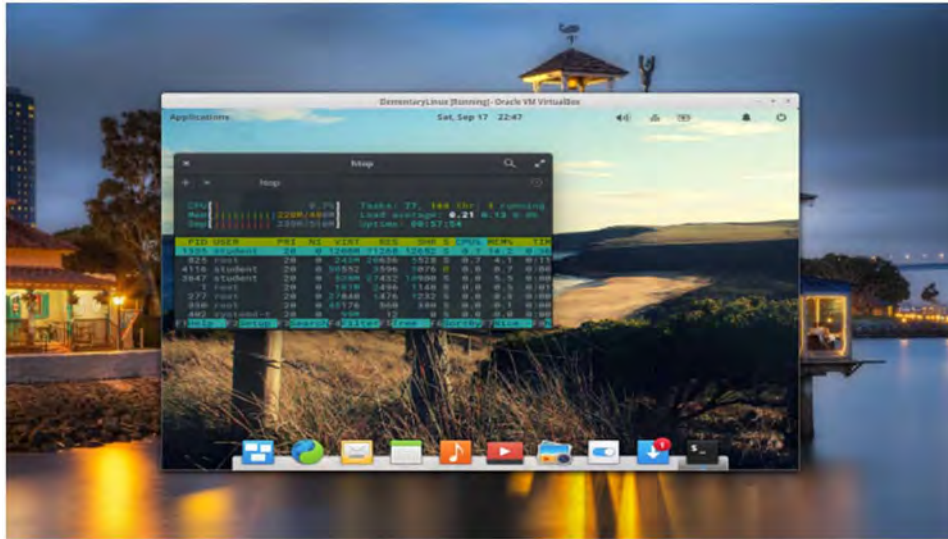
Linux Distributions	Description and Potential Use	Package
Ubuntu	One of the most popular Linux OS developed to be a complete OS that can be an easily replacement for other comparable OSs.	Debian-based
Edubuntu	OS targeted for grades k-12. Contained in OS are tons of software applications that is useful to those who are education majors.	Debian-based
Damn Small Linux	This OS is designed to as a small OS to be utilized on older hardware. This OS is great for institutions that have old computers and want to revitalize them for use. OS is also great for VMs as DSL requires a low amount of memory.	Knoppix-based
BackTrack	OS based on Ubuntu for digital forensics and penetration testing. Great tool for students majoring in technology fields. As cyber security is becoming a hot topic around the world this tool provides students the ability to learn from over thirty software applications that aid in penetration testing and more.	Debian-based
Fedora	This OS is supported by the Fedora Project and sponsored by Red Hat. This OS provides a great resource for learning Red Hat Enterprise Language (RHEL). As there are thousands of jobs requiring expertise specifically with Red Hat this OS is a great tool to prepare students for employment in IT. Fedora has over six Fedora Spins such as Design-suite, Scientific-KDE, Robotics, Electronic-lab, Games, and more.	RPM-based
Kali	An advanced forensics and penetration testing distribution that is the successor of BackTrack.	Debian-based
Elementary	OS is based on Ubuntu and showcases the Pantheon desktop. This OS is focused on usability, and design aesthetics. The desktop environment has been compared to Mac OS. This desktop origins from being an Ubuntu desktop theme.	Debian-based
CentOS	This OS derived entirely from RHEL. The source code is developed from Red Hat which allows a student to learn RHEL with a small number of differences. CentOS can be used for teaching IT students on how to setup, administer, and secure a server.	RPM-based
Ubuntu Studio	This OS is derived from Ubuntu. This OS is developed specifically for multimedia production such as audio, video, and graphics. Departments for multimedia could use this OS for multimedia instruction and the development of projects. As many of the tools for multimedia production are expensive this alleviates large license costs for institutions.	Debian-based
Lubuntu	OS is based on Ubuntu and uses the LXDE desktop environment. It replaces Ubuntu's Unity shell and GNOME desktop.	Debian-based

enhance, and protect the vital means of academic learning through learning techniques as part of program changes and alterations. Open access literature is needed for MOOC system improvements as well as academic and career changes nationwide. In the figure below is a screenshot of the Directory of Open Access Journals (DOAJ) which is of one of the most popular sites for open access refereed works.

STEM NEEDS FOR MOOCS

As STEM is continues to play a critical role in America, Linux could provide the ability an affordable large scale deployment. Even more important is the environment where individuals can practice skillsets and techniques learned from the classroom. Thus, Linux provides the ability for students to perform low level code analysis. In developing the class environments with cost in consideration Linux allows for this to be done with removing costly barriers with the exception of hardware. The barriers are proprietary software licenses, and software that operates on a particular

Figure 2. Elementary Linux in a VM



hardware configuration. The associated software licensing allows for applications to be used freely in MOOCs for enhancing the overall experience.

The addition of Virtual Machines (VM) in MOOCs enhance the overall courseroom experience. VMs can be loaded to the MOOC site which mimic the instructor's teaching environment for a particular course [See Figure 2]. The course tools such as programming environments, networking tools, offensive security applications, and more can already be included in the VM. Additionally, this environment can be exported to an ISO image file so that as this file format will work with multiple hypervisors.

SOFTWARE LICENSING

When considering software licensing it is essential to understand the license agreements for the applications that are used within the MOOC (Dawson, Leonard, & Rahim, 2015). This will allow for the use or redistribution of software. The use of OSS can greatly enhance the STEM environment (Dawson, Al Saeed, Wright, & Onyegbula, 2014).

GNU GPL v3

After a review of the terms and conditions provided by this license it appears to be more comprehensive in its requirements for use of the licensed software. It contains several more terms and appears to contain many more prohibitions than the previous version of the license terms contained. It contains the requirement to include appropriate notices for distribution of the code. It also contains specific prohibitions regarding restriction on the subsequent use of the code, including modified versions, by downstream users (Kumar, 2006).

GNU GPL v2

After review of the terms and conditions of this license, this version's license does not appear to have as many requirements and certainly is not as long as the newest version of this software's license appears to be. While considerably shorter than the subsequent version's license, this license does still maintain and include the requirement that appropriate notices accompany the distribution of the code (Kumar, 2006).

LGPLv3

After review of the terms and conditions of this license, this version's license does not appear to have as many requirements as either of the licenses under the GNUGPLv3 or v2, but it does maintain several requirements for compliance. Of note, is this license includes an exception to the GNUGPL license, namely that the work produced under this license may be reproduced without compliance with Section 3 of the GNUGPL, which relates to Protecting Users' Legal Rights from Anti-Circumvention Law.

LGPL v2

After review of the terms and conditions of this license, this version's license appears to somewhat longer than the terms and conditions of the subsequent version's license, but it appears to be closer to the GNUGPLv2's license terms than the LGPLv3's terms and conditions, and noticeably does not include the exception to the GNUGPL license as is contained in the subsequent version of this license.

LLGPL

After review of the Lisp Lesser General Public License (LLGPL), this version's license is like the LGPL but with a prequel. This prequel defines the effect in terms more typically used in Lisp programs. This license is grounded in the C programming language as the license specifically calls out functions not present in other languages that are not traditionally compiled (Greenbaum, 2013).

Apache 2.0

The previous Apache licenses were based on the GPL v2 however the Apache License v2 permits code that is covers to be subsumed into closed source projects (Rosen, 2015). It also explicitly

grants patent right where necessary to modify, operate, and distribute the software.

MIT License

The MT license provides permission, free of charge, to any individual obtain a copy of the software and associated documentation. Additionally, it (Rosen, 2015). Also provided without limitation is the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the software (Rosen, 2015).

Creative Commons

After review of the terms and conditions of this license, it appears that this license is very similar to that of Modified BSD. It is interesting of note that the license begins by indicating that the company is not a law firm. Additionally, this license appears to include a waiver of copyrights and related rights, and a fall-back in the event that the waiver is invalidated, which appears to be based upon the purpose of promoting the overall ideal of free culture. In addition this license includes a limitation to make sure that neither patent or trademark rights are being waived by this license.

Artistic License 2.0

After review of the terms and conditions of this license, this license appears to be very similar to that at issue in the Jacobsen case discussed above. Moreover, it appears that this license makes clear that the copyright holder intends to retain some creative control over the copyrighted work overall, while still trying to ensure that the copyrighted material remains as open and available to others as possible under the circumstances.

Modified BSD

After review of the terms and conditions of this license, these terms and conditions appear to be the shortest list of terms and conditions of all of

Figure 3. System Engineering Virtual Test



the licenses reviewed in this paper. Additionally this license appears to allow reproduction and modification of the copyrighted material provided certain conditions are met, which if subject to legal challenge, a court might construe as being subject to only protection as a contract, at best, and a bare license at worst. Moreover, based upon the legal authorities cited in this paper, it may be unclear whether this license may provide sufficient copyright protection.

Clear BSD License

After review of the terms and conditions of this license, this license appears to be very similar to the Modified BSD License, in that it is very short, and appears to allow reproduction only if certain conditions are met. This license does make clear that no patent rights are granted by this license.

VIRTUAL WORLDS

Virtual worlds can provide a platform to teach the labs in an environment that is dynamic and takes advantage of ubiquitous learning (Davis,

Dawson, & Omar, 2016; Dawson & Al Saeed, 2012). Access to a MOOC could occur in a manner that a student performs an engineering lab design in the virtual world environment. This would allow the individuals to have a low fidelity lab in which the activities provided in the MOOC can be done in an environment that allows for synchronous or asynchronous participation. The MOOC environment could be enhanced with the use of virtual environments with the use of virtual laboratories, virtual museums, augmented reality, collaborative e-learning, Neto, de Souza, & Gomes, 2016).

CONCLUSION

Despite the fact that many innovations in the field of educational technology have existed throughout the decades, undoubtedly the most conspicuous public conversations in mainstream media are pointed to the phenomenon, massive open online courses (MOOCs). Pappano (2012), in her New York Times November 2012 article branded the year 2012 as the “Year of the MOOCs”. This labeling help to propel MOOCs into being a significant educational technological catchphrase.

Conceptually, MOOCs appeared to be the utmost in learning in the online space; however, glitches appeared. These glitches include but are not limited to underwhelming involvement, lower than expected completion rates, and imbalanced content significance according to Marrapodi, Shimkus, and Onisk (2016). Notwithstanding the rising accessibility, attraction, and probability of MOOCs, their economic justification and their academic value required more exploration, and particularly the technological aspects.

FUTURE RESEARCH DIRECTIONS

More research is needed regarding MOOCs and quality of development, MOOCs and STEM needs, as software licensing. The data shows that the instructional quality of MOOCs varies. With this said instructional principals must be revisited to determine a best practice. Data needs to define the top principles of learning for MOOCs, and the specific types of learning environments. Some data exist about MOOCs in STEM education; however, more is needed. A key point of focus will be having the appropriate mix of academic information to learning exercises (e.g., videos, assignments, and activities). The key point is to make learning stick. Finally, when determining the type of MOOC course to develop, the development team must have a better understanding of the type of licenses required for MOOCs. It is important to know that key obstacle in terms of licenses are proprietary software licenses, as well as software that operates on a distinct hardware configuration. The associated software licensing permits the applications to be used freely in MOOCs for enhancing the overall experience

MOOCs encompass a host of different types of courses. These types of courses allow for global delivery. Delivery approached will continue to vary; however, will need to be studied and concentrated as to the needs of different student populations. Technology is important; thus, due to ubiquitous technological advancement,

software licensing will have to be consistently reviewed. MOOCs can prove useful as a new phenomenon of course delivery for students, faculty, and administrators to use. To effectively do this one must maximize the technical use OSS, open source technologies, and open access literature while minimizing the overall associated license costs. The use of virtual worlds with MOOCs would provide students an environment that allows student to be fully immersed in required technology labs (Calongne, 2008).

REFERENCES

- Bessette, D., & Burton, S. (2013). The Involvement of Professional and Managerial Practices and principles for Online Learners in Virtual Programs. In R. McBride & M. Searson (Eds.), *Proceedings of Society for Information Technology & Teacher Education International Conference 2013* (pp. 3040-3045). Chesapeake, VA: AACE.
- Burd, E., Smith, S., & Reisman, S. (2015). Exploring Business Models for MOOCs in Higher Education. *Innovative Higher Education*, 40(1), 37–49. doi:10.1007/s10755-014-9297-0
- Burton, S. L. (2016). *Transitioning to I can from I can't*. Raleigh, NC: Lulu Publications.
- Calongne, C. M. (2008). Educational Frontiers: Learning in a Virtual World. *EDUCAUSE Review*, 43(5), 36.
- Caulfield, M., Collier, A., & Halawa, S. (2013). Rethinking online community in MOOCs used for blended learning. *Educause Review Online*. Retrieved from <http://www.educause.edu/ero/article/rethinking-online-community-moocs-used-blended-learning>
- Chang, C. C. (2006). Business models for open access journals publishing. *Online Information Review*, 30(6), 699–713. doi:10.1108/14684520610716171

- Christensen, C. (1997). *The innovator's dilemma*. Boston, MA: Harvard Business School.
- Davis, L., Dawson, M., & Omar, M. (2016). Systems Engineering Concepts with Aid of Virtual Worlds and Open Source Software: Using Technology to Develop Learning Objects and Simulation Environments. In F. Neto, R. de Souza, & A. Gomes (Eds.), *Handbook of Research on 3-D Virtual Environments and Hypermedia for Ubiquitous Learning* (pp. 483–509). Hershey, PA: Information Science Reference; doi:10.4018/978-1-5225-0125-1.ch020
- Dawson, M., Al Saeed, I., Wright, J., & Onyegbula, F. (2014). Open Source Software to Enhance the STEM Learning Environment. In V. Wang (Ed.), *Handbook of Research on Education and Technology in a Changing Society* (pp. 569–580). Hershey, PA: Information Science Reference; doi:10.4018/978-1-4666-6046-5.ch042
- Dawson, M., Leonard, B., & Rahim, E. (2015). Advances in Technology Project Management: Review of Open Source Software Integration. In M. Wadhwa & A. Harper (Eds.), *Technology, Innovation, and Enterprise Transformation* (pp. 313–324). Hershey, PA: Business Science Reference; doi:10.4018/978-1-4666-6473-9.ch016
- Dawson, M. E., & Al Saeed, I. (2012). Use of Open Source Software and Virtualization in Academia to Enhance Higher Education Everywhere. *Cutting-edge Technologies in Higher Education*, 6, 283–313. doi:10.1108/S2044-9968(2012)000006C013
- Decker, G. L. (2014). MOOCology 1.0. In S. D. Krause & C. Lowe (Eds.), *Invasion of the MOOCs: The promise and perils of Massive Open Online Courses* (pp. 122–129). Anderson, SC: Parlor Press.
- Demirci, N. (2014). What is Massive Open Online Courses (MOOCs) and What is promising us for learning?: A Review-evaluative Article about MOOCs. (English). *Necatibey Faculty Of Education Electronic Journal Of Science & Mathematics Education*, 8(1), 231–256. doi:10.12973/nefmed.2014.8.1.a10
- Flynn, J. T. (2013). MOOCs: Disruptive Innovation and the Future of Higher Education. *Christian Education Journal*, 10(1), 149–162. doi:10.1145/2240236.2240246
- Fox, A. (2013). From MOOCs to SPOCs. *Communications of the ACM*, 56(12), 38–40. doi:10.1145/2535918
- Greenbaum, E. (2013). Lispering Copyleft: A Close Reading of the Lisp LGPL. *International Free and Open Source Software Law Review*, 5(1), 15–30. doi:10.5033/ifoosslr.v5i1.75
- Guzdial, M., & Adams, J. C. (2014). MOOCs need more work; So do CS Graduates. *Communications of the ACM*, 57(1), 18–19. doi:10.1145/2555813
- Harnad, S., & Brody, T. (2004). Comparing the Impact of Open Access (OA) vs. Non-OA Articles in the Same Journals. *D-Lib Magazine*, 10(6).
- Knowles, M. (1979). Malcom Knowles on... 'How I coped with fads in training'. *Training and Development Journal*, 33(9), 36. PMID:10240407
- Kolowich, S. (2013a). Why some colleges are saying no to MOOC deals, at least for now. *The Chronicles of Higher Education*. Retrieved from <http://chronicle.com.proxy.cecylibrary.com/article/Why-Some-Colleges-Are-Saying/138863>
- Kolowich, S. (2013b). *MIT will offer MOOC curricula, not just single course, on edX*. Retrieved from <http://chronicle.com/blogs/wiredcampus/mit-will-offer-mooc-curricula-not-just-single-courses-on-edx/46715>

- Kolowich, S. (2013c). Duke U.'s undergraduate faculty derails plan for online courses for credit. *The Chronicles of Higher Education*. Retrieved from <http://chronicle.com.proxy.cecylibrary.com/article/Duke-Us-Undergraduate/138895>
- Koutropoulos, A., Gallagher, M., Abajian, S. C., de Waard, I., Hogue, R., Keskin, N., & Rodriguez, C. (2012). Emotive vocabulary in MOOCs: Context & participant retention. *European Journal of Open, Distance, E-learning, 1*. Retrieved from <http://www.eric.ed.gov/PDFS/EJ979609.pdf>
- Krause, S. D. (2014). MOOC Assigned. In S. D. Krause & C. Lowe (Eds.), *Invasion of the MOOCs: The promise and perils of Massive Open Online Courses* (pp. 122-129). Anderson, SC: Parlor Press.
- Kumar, S. (2006). Enforcing the Gnu GPL. *U. Ill. JL Tech. & Pol'y*, 1.
- Leonard, D., Barton, G., & Barton, M. (2013). Make yourself an expert. *Harvard Business Review*, 91(4), 127-132. PMID:23593772
- Marrapodi, J., Shimkus, D., & Onisk, M. (2016, September). Do's and don'ts for MOOCS and SPOOCS. *Training (New York, N.Y.)*, 53(5), 20-27.
- Marshall, B., Cardon, P., Poddar, A., & Fontenot, R. (2013). Does sample size matter in qualitative research? a review of qualitative interviews in is research. *Journal of Computer Information Systems*, 54(1), 11-22. doi:10.1080/08874417.2013.11645667
- Martin, F. G. (2012). Education wills massive open online courses change how we teach? *Communications of the ACM*, 55(8), 26-28. doi:10.1145/2240236.2240246
- McAuley, A., Stewart, B., Siemens, G., & Cormier, D. (2010). *The MOOC model for digital practice*. Academic Press.
- Mosley, A. L. (2007). *An evaluation study of a training program to prepare community college faculty for online teaching* (Doctoral Dissertation). Retrieved from ProQuest Dissertations and Theses database (PQDT). (UMI Number: 3262899)
- Neto, F. M., de Souza, R., & Gomes, A. S. (2016). *Handbook of Research on 3-D Virtual Environments and Hypermedia for Ubiquitous Learning*. Hershey, PA: IGI Global; doi:10.4018/978-1-5225-0125-1
- Ngambi, D., & Bozalek, V. (2015). Editorial: Massive open online courses (MOOCs): Disrupting teaching and learning practices in higher education. *British Journal of Educational Technology*, 46(3), 451-454. doi:10.1111/bjjet.12281
- Nicholas, D., Huntington, P., & Jamali, H. R. (2007). Open access in context: A user study. *The Journal of Documentation*, 63(6), 853-878. doi:10.1108/00220410710836394
- Pappano, L. (2012, November 2). The year of the MOOC. *The New York Times*. Retrieved from <http://www.nytimes.com/2012/11/04/education/edlife/massive-open-online-courses-are-multiplying-at-a-rapid-pace.html>
- Pence, H. E. (2012). When will college truly leave the building: If MOOCs are the answer, what is the question? *Journal of Educational Technology Systems*, 41(1), 25-33. doi:10.2190/ET.41.1.c
- Perens, B. (1999). *The open source definition. Open Sources: Voices from the Open Source Revolution*. Academic Press.
- Rahim, E., & Burrell, D. N. (2013). *The value of an MBA in a post-recession world*. Colorado Technical University. Retrieved from <http://www.coloradotech.edu/resources/blogs/december-2013/the-value-of-an-mba-in-a-post-recession-world?form=degreetype>
- Rees, J. (2014). The MOOC racket. *Future Tense*. Retrieved from http://www.slate.com/articles/technology/future_tense/2013/07/moocs_could_be_disastrous_for_students_and_professors.html

Rodriguez, C. (2012). MOOCs and the AI-Stanford like courses: Two successful and distinct course formats for massive open online courses. *European Journal of Open, Distance And E-Learning*. Retrieved from <http://www.eric.ed.gov/PDFS/EJ982976.pdf>

Rosen, L. (2005). *Open source licensing* (Vol. 692). Prentice Hall.

Severns, M. (2014). *Math teacher's parody prompts project to critique Khan's Academy lessons*. Retrieved from http://www.slate.com/blogs/future_tense/2012/06/26/khan_academy_mystery_science_theater_300_parody_by_math_teachers_video_.html

Siemens, G. (2013). *2013 in MOOC s- Which event best defined the quest to solve education?* Retrieved from <https://allmoocs.wordpress.com/tag/george-siemens/>

Suber, P. (2004). *What is open access? An overview*. ALPSP-SSP Meeting, Washington, DC.

Ta'eed, C. (2012, September 15). Will the Internet replace traditional education? *The Next Web*. Retrieved from <http://thenextweb.com/insider/2012/09/15/will-internet-replace-traditional-education>

Tareef, A. (2013). The relationship between mentoring and career development of higher education faculty members. *College Student Journal*, 47(4), 703–710.

Taylor, B., & Kroth, M. (2009). Andragogy's transition into the future: Meta-analysis of andragogy and its search for a measurable instrument. *Journal of Adult Education*, 38, 1–11.

Thrift, N. (2013). To MOOC or not to MOOC. *The Chronicles of Higher Education*. Retrieved from <http://chronicle.com.proxy.cecybrary.com/blogs/worldwise/to-mooc-or-not-to-mooc/31721>

Yong, C. (2014). Investigating MOOCs through blog mining. *International Review of Research in Open and Distance Learning*, 15(2), 85–106.

KEY TERMS AND DEFINITIONS

GNU Public License: A widely used free software license that is managed under the GNU Not Linux Project (Stallman, 1991).

Linux: An open source version of the UNIX OS (Perens, 1999).

MOOC: An online course with the option of free and open registration, publicly shared curriculum, and open ended outcomes (McAuley, Stewart, Siemens, & Cormier, 2010).

Open Source Software: Software that allows the original source code to be free available which may be freely redistributed or modified (Perens, 1999).

Software License: Legal instrument for governing the use or redistribution of software (Dawson, Leonard, & Rahim, 2015).

Open Source Software Virtual Learning Environment (OSS–VLEs) in Library Science Schools

Rosy Jan

University of Kashmir, India

INTRODUCTION

The cost of higher education with the attrition of formal learning and the rise and evolution of informal learning presents a unique challenge to academic institutions. Students can no longer afford the cost of textbooks, and the use and embracement of proprietary software limits the freedom of knowledge exchange. The situation is further stimulated by considerably changed attitude and skills of students in IT over the last few years. The transformation of web from static to dynamic has changed the perception of the younger generation towards information presentation and availability. The younger generation now expect to be in ‘constant connectivity’ with friends and family taking the benefit of participatory and interactive nature of the web. They increasingly expect technology to have a significant role in their learning also. Learning styles are growing fast as the technology grows in the world. New techniques are forthcoming within the higher education system to get the better student’s presentation, tutors philosophy and the institution enlargement. Information and Communication Technology innovates the new technologies for conversing the information and knowledge to have a different transform for teachers to teach and for students to learn Ruíz, Martínez & López (2016). So, need emerged to rethink how we teach and how we use learning technologies and how education is delivered. In response to this changing environment, e-learning is being implemented more and more frequently in higher education, creating new and exciting opportunities for both educational institutions and

students. E-learning means “instructional content or learning experience delivered or enabled by electronic technologies” (Ong, Lai & Wang, 2004). Which more recently, include mobile and wireless learning applications. E-learning is increasingly adopted in the workplace for supporting professional development and continuing education; however, in higher education, the use of e-learning is predominantly used as a tool support teaching (King & Boyatt, 2015).

The variations in the configuration of e-learning offerings can be described through a number of attributes. One of such attributes is the use of VLEs. A Virtual Learning Environment (VLE) is a software system designed to facilitate teachers in the management of educational courses. The system can often track the learners’ progress, which can be monitored by both teachers and learners. While often thought as primarily tools for distance education, they are most often used to supplement the face-to-face, classroom as well as blended learning. VLEs are defined as “computer-based environments that are relatively open systems, allowing interactions and knowledge sharing with other participants and instructors” and providing access to a wide range of resources (Wilson, 2006). “The Joint Information Systems Committee (JISC) considers a Virtual Learning Environment (VLE) an online environment in which learners and tutors participate in “on-line” interactions of various kinds” (JISC, 2000). Besides many more names VLEs are also referred as online learning environments, learning management systems or collaborative learning software (Britain & Liber, 2000). The first systems that fitted the criteria of

DOI: 10.4018/978-1-5225-2255-3.ch688

VLEs as we know them now started to emerge between 1995 and 1997 (Stiles, 2007). VLE is a boxed system that mediates between teacher(s) and student(s). They have a consistent and customizable interface and a clear navigational structure (Stiles, 2007). Access to the environment is often ubiquitous and supports anywhere learning with the support of internet connection (Jacobsen & Kremer, 2000).

The value of a VLE is to fully bring out the characteristics of both “Learning Any Where” and “Learning Any Time,” and the purpose of a VLE is to emphasize and diffuse thinking models, diverse viewpoints, independent thinking, etc. (Chou & Liu, 2005). Plethora of research conducted on VLEs has reported positive impacts from various contexts. They can; increase enthusiasm and confidence, improved readiness to learn, promote reflection, accommodate the needs of students and broadly reported improved course assessment performance (Means, Toyama, Murphy, Bakia and Jones, 2009). More recently, the research targets at using on Web 3.0 – based personalisation of learning objects (LOs) while learning in virtual learning environments (Kurilovas, Kubilinskiene & Dagiene, 2014).

Keeping in view the benefits, the work has been initiated to review concept, features and issues of virtual learning environments. Some of the most used OSS VLEs are discussed. Further it determines the suitability of a VLE for higher education. The chapter also explore and identify the recent contributions to the concept by analyzing ongoing virtual learning initiatives and projects by different organizations and information centres to stimulate future Research and development trend in the field.”

BACKGROUND

Technologies are affecting almost all facets of life be it business, entertainment or social activity. Commentators and futurists suggest that there are profound implications for education as well.

They argue that because the “Net Generation” of learners is so engrossed in a networked world of digital technology, they behave differently, have different social characteristics, different ways of using and making sense of information, different ways of learning, and different expectations about life and learning. So, need emerged to rethink how we teach and use learning technologies and how education is delivered.

In response to this changing environment, one of the means of education that has been propagated with internet revolution is e-learning. e-learning is being implemented more and more frequently in higher education, creating new and exciting opportunities for both educational institutions and students. E-learning means “instructional content or learning experience delivered or enabled by electronic technologies” (Ong, Lai & Wang, 2004). Naresh and Reddy (2015) identified the key factors of successful implementation of e-learning in both developing and developed countries and concluded that financial support from governments, students’ motivation and well trained tutors are important for such implication. In addition, user perception and readiness also play an important role in e-learning effectiveness. The barriers faced to a larger extent by developing countries than developed ones are as follows: lack of infrastructure, trained instructors, lack of financial support, government policies and less student readiness. Developed countries, on the other hand, have a strong infrastructure and face the following challenges: student engagement, student motivation, and high student drop out ratio. It has been observed that the variations in the configuration of e-learning offerings can be described through a number of attributes. One of such attributes is the use of VLEs.

A VLE is a web-based communications platform that allows students, without limitation of time and place, to access different learning tools, such as program information, course content, teacher assistance, discussion boards, document sharing systems, and learning resources Martins and Kellermanns (2004). VLEs electronically

support training and development in higher educational and vocational training settings. These are frequently used and preferred over the traditional systems keeping in view the benefits like ubiquity, timeliness, efficiency, individuality and learning task orientation. They do so by offering a repository for course documents, discussion forums, chat boxes, mass communication options, etcetera.

Within the overwhelming amount of technology acceptance studies Sun and Zhang (2006). The number of those studying the acceptance and use of VLEs is small but growing (Martins and Kellermanns 2004; Ngai, Poon & Chan, n. d.; Ong et al, 2004; Pituch & Lee, 2006 & Selim, 2003). VLEs are rapidly becoming an integral part of the teaching and learning process (Pituch & Lee, 2006). Ahmed and Morley (2010) qualitative and quantitative explored how a Virtual Learning Environment (VLE) can be supported in a Higher Education setting. Findings revealed that holistic support strategies were capable of targeting larger groups of teachers effectively via staff development workshops and strategies. Study by Mimirinis and Bhattacharya (2007) presents the results of a case study which explore the relationship between approaches to learning and studying, and perceptions of use of a VLE in a Higher Education taught module. Further the study investigates the requirements for appropriate design of VLEs. Recommendations are aiming to highlight the importance of specific elements in the design and delivery of online courses through VLEs. Graham, Woodfield & Harrison (2013) observed that there has been rapid growth in blended learning implementation and research focused on course-level issues such as improved learning outcomes, but very limited research focused on institutional policy and adoption issues.

Comparisons of traditional classroom learning and studying with a VLE have been carried out by Piccoli, Ahmad and Ives (2001), Kekkonen-Moneta and Moneta (2001), Marandi and Luik (2003), Zhang et al. (2004) and McDonald et al. (2004). VLEs have advantages compared with traditional teaching. When different teachers lecture

on the same topic, no two lectures are ever exactly the same. Furthermore, the same teacher hardly ever repeats the same lecture in exactly the same form or with exactly the same content (Walker & Harrington, 2004). Lecturers can de-motivate students with their routines. On the other hand, VLEs are a flexible way of teaching (Walker & Harrington, 2004) because VLEs can be used at the most convenient time (Clarke, 2001). The learners learn beyond the timetable routines and guidelines, and therefore chances of missing a lecture is rare (Walker & Harrington, 2004). Lee et al. (2002) found in their study that positive attitudes towards using computers were the key factor in the VLE's success. Students' positive relationship to computers helps also the learning process (Lee et al., 2002). Costello (2014) highlighted that Virtual learning environments (VLEs), such as Moodle, have become an integral part of the complex infrastructure of higher education systems throughout the world. However, Babic (2012) argued that both educational practice and scientific research indicate that teachers' insufficient use of VLEs in higher education teaching still represents a problem. It would not be honest to claim that virtual learning environments will improve the quality of education or reduce the costs of educational systems. These environments have some potential effects. However, the past tells us that it is very difficult to set up the conditions that turn potential into actual effects. Nonetheless, even if there were no proof of superiority in terms of learning outcomes, the evolution would not stop. The issue is not to prove the effects but to understand them.

Coppola (2005) highlighted and focused the e-learning platforms (both LMS and LCMS). Author concluded that the platforms currently available are based on either proprietary e-learning software (PES) or open source e-learning software (OSS). Further it was observed that OSS usage in implementing e-learning systems is more emphasized due to the challenges faced when implementing the PES. Open source software offers the potential to reduce the cost of the software while providing

the universities greater control over its destiny. Elimination or reduction of license leaves more budgets available to invest in adapting and managing the software; offers reliability, performance and security over proprietary software due to the availability of the source code, which allows vulnerabilities to be identified and resolved by third parties and it is easy to customize. Some of the widely used open-source e-learning software programs are the Claroline and Moodle. Littlejohn et al (2006) found that with the merger of Blackboard and WebCT, the selection of the open source system Moodle is an increasingly attractive alternative for many institutions. The Open University (OU) recently choose Moodle as a core component of its virtual learning environment after an extensive requirement-gathering process and evaluations of commercial and open source products demonstrate the way of using the Moodle in the education process at the Gaia Vocational School (GVS). The platform was characterized and promoted among the School members of GVS. The impact of this new platform (e-Courses) on the learning institution activities was evaluated through an online questionnaire, which showed high satisfaction levels with e-Courses. Thus, regarded MOODLE an excellent example of learning management system (Syed, 2009). Report from the Croy, Smelser and McAlpin (2009) describes the results of an evaluative comparison of the Moodle Learning Management System with the Blackboard system. Recommends that the University of North Carolina adopt Moodle as its sole Learning Management System; and Outlines the consequences of doing so in conjunction with a plan for making the transition from Blackboard to Moodle. Hodhod et al. (2010a) found that current market place for VLEs in HEIs is dominated by two products Blackboard and WebCT, each offering a variety of tools and functionality. Many other products are available and in wide use such as First-Class and LearnWise. More recently the advent of open source has offered the academic community Moodle. This VLE is fast superseding the proprietary products as being the most popular

and easy to use system of choice apart from many licensed VLEs like BlackBoard, desire2learn, CyberExtension or It's Learning; there are few open source VLEs as well in the market. Adoption to open source VLE is very much dependent on requirements of the organizations. Famous open source VLEs are Bodington & Moodle. A comparison between the available free VLEs Claroline, Moodle and Atutor by Hodhod et al (2010b) to assure that Moodle is really the suitable choice and it holds the suitable features for a good learning environment. White paper by Luzet (2010) provides brief overview of the usability of the main features of Moodle and Fronter. The results of this work reveal that Fronter wins out from a usability perspective in its simplicity: it offers fewer options and is therefore easier to grasp for a novice user. On the other hand, Moodle's help system scores highly for its comprehensive and content-specific help documentation. Further Moodle is more flexible and scalable which makes it popular as the educational institutions become more demanding in their needs. A comparative study of Moodle with other e-learning systems by Kumar, Gankotiya & Dutta (2011) focused on the Moodle Architecture and presents some authentication plug-in that Moodle supports. LMS Moodle has been adopted by many people and organizations around the world because it offers a tightly integrated set of tools said to be designed from a social constructive perspective. Moodle has been developed under the general public license and many of its components were developed without a specific design documentation including its security services. Ajlan (n. d.) in a book chapter study, analyze, and explore the right decision when choosing a suitable VLE platform to meet the requirements of Qassim University. It focused on a comparison between Moodle and other VLE systems, and is based on two kinds of comparison. This study has proved that the best platforms are Moodle and Sakai, which have missed just two out of forty features. Other than the comparative studies recent studies focused on the development process different

experiences and perspectives of the staff involved. Study by Buus (2016) highlighted the development process and some of the didactic considerations undertaken for the implementation of VLE. The evaluations undertaken during the process was presented, along with the results collected in the use of Moodle to highlight the educational changes. While as Sinclair & Aho (2016) presents initial, qualitative research aimed at understanding how Moodle is being used and the different experiences and perspectives of the staff involved. To generate themes and areas of interest for future investigation the paper uses interview data from two “expert witnesses” having a deep understanding of how the platform is used. Emergent themes include: divergence between confident and basic users; the spread of usage within an academic community; lack of progression to innovative teaching methods.

VLE IN LIS SCHOOLS

The field of library and information science is presently passing through a phase of rapid changes. This situation demands an educational system of continuing or life-long learning for new technologies, methods and service procedures. The augmentation of traditional classroom activities with electronic learning objects along with the rise of the free and open source software based course management systems is changing the way faculty and students access, create and use learning objects, information and knowledge FERL (2005). It is providing new opportunities for libraries to design and to disseminate new services. Library schools are also investigating the use of VLE in imparting traditional classroom based courses as well focused and continuing education and/or training programmes for LIS professionals. In fact, the relationship between VLE and LIS profession may be viewed from two angles – application of LIS services in virtual learning environment and use of VLE in design & delivery of traditional as well as focused courseware. OCLC in its white

paper McLean (2003) prescribes following library supported services for a typical VLE: Consecutive display and integration of a variety of information windows as part of a learning activity, Aggregation of access to content in any given learning context, Provision of bibliographic tools, Integration of plagiarism software into course management systems, Integration of library resources in course management systems, Customization of portal facilities for storing personal preferences. Paper by (Singh & Devi, 2009) highlighted VLE, Virtual Community, Characteristic features, objectives, Issues, Essentialities, Choice of content, etc. LIS Education in VLE emphasizes on the American Experience, IGNOU initiatives in India concluded that Indian LIS Schools should adopt virtual learning system. However, paper by Mukhopadhyay (2006) concentrates on use of VLE in managing LIS courseware. The author has experimented the design of vidyaOnline, the prototype web-based modular and interactive learning system, aimed to produce a Virtual Learning Environment (VLE) for library and information science courses as well as for other distance and traditional courses of Vidyasagar University. The structure of Vidya-Online extends support for all three forms of VLE – web-based training, supported online learning and informal e learning. Chowdhury and Chowdhury (2006) describe the findings of the initial phase of the eLLIS that aims to study the current practices and support for e-learning available in the Library and Information Science departments in UK. The paper provides a snapshot of the current practices, highlighting the most interesting ones, of e-learning supports provided by the LIS departments and the libraries of the corresponding universities. This paper shows that all the LIS departments and the corresponding university libraries provide some supports for e-learning, though the nature of these support services varies significantly. Most LIS departments in UK currently offer a blended learning environment. Islam, Chowdhury and Islam (2009) examine the LIS education in e-learning environment and its implications in other areas. This study also gives

an overview of e-learning education systems in Bangladesh and finds out problems of e-learning in LIS educations in Bangladesh. The findings reveal that Bangladesh needs a proposal to acquaint with this system. It provides a proposal to apply distance and e-learning modes adopted by different LIS institutions in Bangladesh making some recommendations for implementing the proposal. Paper by Lihitkar (2010) attempts to provide a conceptual idea to build-up a prototype model for VLE@DLISc,RTMNU in ICT environment by integrating the two software i.e. Moodle and GSDL. Paper by Arora and Lihitkar (2015) endeavours to discuss the need of VLE in the present LIS education, its pros and cons and also some issues that one will face while implementing it for university, college, or any educational institute. While as Qutab, Shafi-Ullah, Safdar & Khan (2016) explores the use of Virtual Learning Environment tools by professional library associations for career development, collaboration, networking, and building communities across the globe. The paper further emphasised that Continuing Professional Development is the dire obligation of the practising library and information professional in the challenging information society, ever-changing user needs, and advancing support tools. The national and international library associations carry the responsibility of the CPD for the librarians. Traditional face-to-face training, discussions and conferences are supplementary with eLearning platforms, making acquisition of knowledge easy, accessible, economical and thus possible across the geographical limits for career advancement. While as Watkins & Sheu (2016) presented the concept of open knowledge diffusion and the effects of social inertia in preserving constrained learning environments. Authors identified and explained Open Knowledge Diffusion Tools (OKDT) such as Open Educational Resources (OER), Open Textbooks (OT), Massive Open Online Course (MOOC), and Free and Open Source Software (FOSS).

FUTURE RESEARCH DIRECTIONS



The future of learning is outside the traditional campus and classroom. There are profound advances and changes in technology resulting in rapid and unparalleled changes in learning and teaching. The default access points will be mobile phones, laptops and tablets keeping in view the plethora of new devices and APIs as well as super-fast internet speeds available around the world. The future will turn and twist the whole gamut of educational system. Thus, the changes and disruptions to traditional form of university teaching in general and teaching in LIS School in particular will be wide and profound. The LIS schools that mainly rely on disseminating information in traditional classrooms will have fewer resources to such an evolution and therefore will be pressured to bring changes. So, the only way to deal with that will be by investing in blended learning, using the flipped classroom model. The future lies in the use of Open Knowledge Diffusion Tools and the potential of Three-Dimensional Multi-User Virtual Environments (3-D MUVES) such as Second Life (SL) in the service of learning.

That leads towards infinite imaginative educational possibilities. The future lies in extending classrooms practice, curriculum enhancement, engagement of students, teachers and administrators to support learning, promote collective creativity and shared leadership and unite learning groups with shared values, vision and practices in a global perspective.

CONCLUSION

Open Knowledge Diffusion Tools, MUVES and other latest leaning technologies of the future will offer a very different set of features and tools. The recent merger and ongoing development of various VLEs like SAKAI and other open source products, which offer the promise of

easily shared learning tools, and new techniques such as AJAX, will drive the change process. Eventually, a more responsive set of learning tools will be a demand of upcoming generation of students. Just as they are quickly moving beyond the traditional institutional environment to create their own virtual social realms, they may one day create and utilize their own learning tools and communities. Of course, institutions will always hold the upper hand in these developments as they maintain the formal credentials students require for future careers. But this exercise of power is an illusion in a rapidly changing society and we want to engage our students through experiences that are both empowering and transformative. Doing this will require that we rethink about our teaching model, and grasp the reality that the underlying paradigms are, indeed, changing faster than we think.

REFERENCES

- Syed Adnan Adil. (2009). Virtual Learning Environment (VLE) at a Glance. *PLAP News Online: A Bimonthly Newsletter of the Pakistan Library Association*, 12.
- Ahmed, J., & Morley, G. (2010). VLE a blessing or a curse: VLE use by HE Academic Staff. In *Global Learn Asia Pacific 2010 - Global Conference on Learning and Technology*. Retrieved from <http://eprints.hud.ac.uk/8901/>
- Arora, D., & Lihitkar, S. (2015). *Need of Virtual Learning Environment for Educating Library & Information Science*. Retrieved from <http://ir.inflibnet.ac.in/handle/1944/1880>
- Britain, S., & Liber, O. (2004). *A framework for pedagogical evaluation of virtual learning environments*. Retrieved from <https://hal.archives-ouvertes.fr/hal-00696234/>
- Chou, S. W., & Liu, C. H. (2005). Learning effectiveness in a Web-based virtual learning environment: A learner control perspective. *Journal of Computer Assisted Learning*, 21(1), 65–76. doi:10.1111/j.1365-2729.2005.00114.x
- Chowdhury, G., & Chowdhury, S. (2006). *E-learning support for LIS education in UK*. Retrieved from <http://strathprints.strath.ac.uk/40628/>
- Clarke, A. (2001). *Designing Computer-Based Learning Materials*. Aldershot Gower Company.
- Coppola, C. D. (2005). *Will open source unlock the potential of e-learning?*. Campus Technology. Retrieved from eprints.hud.ac.uk/8901/
- Croy, M., Smelser, R., & McAlpin, V. (2009). *Report to the provost from the learning management system evaluation committee*. Retrieved from <http://lmseval.uncc.edu/index.php>
- FERL. (2005). *Designing online courses – good practice guidelines*. Retrieved from <http://ferl.ngfl.gov.uk>
- Hodhod, R., Ibrahim, M., Khafagy, M., & Abdel-Wahab, M. S. (2010). Issues of Choosing the Suitable Virtual Learning Environment. *Research Journal of Information Technology*, 2(1), 24–29.
- Islam, S., Chowdhury, S., & Islam, A. (2009). LIS education in e-learning environment: problems and proposal for Bangladesh. In *Asia-Pacific Conference on Library & Information Education & Practice*. Retrieved from <http://ir.inflibnet.ac.in/dxml/handle/1944/1470>
- Jacobsen, M., & Kremer, R. (2000). *Online testing and grading using WebCT in computer science*. Academic Press.
- Kekkonen-Moneta, S., & Moneta, G. B. (2001). Online Learning in Hong Kong: A Preliminary Comparison of the Lecture and Online Versions of a Computing Fundamentals Course. In *WWW Posters*. Retrieved from <http://www10.org/cdrom/posters/1081.pdf>

- King, E., & Boyatt, R. (2015). Exploring factors that influence adoption of e-learning within higher education. *British Journal of Educational Technology*, 46(6), 1272–1280. doi:10.1111/bjet.12195
- Kumar, S., Gankotiya, A. K., & Dutta, K. (2011). A comparative study of moodle with other e-learning systems. In *2011 3rd International Conference on Electronics Computer Technology (ICECT)* (Vol. 5, pp. 414–418). IEEE. doi:10.1109/ICECTECH.2011.5942032
- Kurilovas, E., Kubilinskiene, S., & Dagiene, V. (2014). Web 3.0–Based personalisation of learning objects in virtual learning environments. *Computers in Human Behavior*, 30, 654–662. doi:10.1016/j.chb.2013.07.039
- Lee, J., Hong, N. L., & Ling, N. L. (2002). An analysis of students' preparation for the virtual learning environment. *The Internet and Higher Education*, 4(3), 231–242. doi:10.1016/S1096-7516(01)00063-X
- Lihitkar, S. R. (2010). Virtual Learning Environment for DLISc, RTMNU, Nagpur: A Prototype Design. *Proceeding of one day National Workshop on Digital Library Initiative and Applications in Indian Context*.
- Littlejohn, A., Cook, J., Campbell, L., Sclater, N., Currier, S., & Davis, H. (2006). Managing educational resources. In *Contemporary perspectives in e-learning research* (pp. 134–146).
- Marandi, T., & Luik, P. (2003). Teacher Training – With or Without Computers?. In *Proceedings of the 2nd European Conference on eLearning*. Academic Conferences International.
- Martins, L. L., & Kellermanns, F. W. (2004). A model of business school students acceptance of a web-based course management system. *Academy of Management Learning & Education*, 3(1), 7–26. doi:10.5465/AMLE.2004.12436815
- McDonald, M., Dorn, B., & McDonald, G. (2004). A Statistical Analysis of Student Performance in Online Computer Science Courses. In *Proceedings of the 35th SIGCSE technical symposium on Computer science education*. ACM Press. doi:10.1145/971300.971327
- McLean, N. (2003). OCLC e-learning task force: libraries and the enhancement of e learning. OCLC Online Computer Library Center. Retrieved from <http://www5.oclc.org/downloads/community/elearning.pdf>
- Means, B., Toyama, Y., Murphy, R., Bakia, M., & Jones, K. (2009). *Evaluation of Evidence-Based Practices in Online Learning: A Meta-Analysis and Review of Online Learning Studies*. US Department of Education.
- Mimirinis, M., & Bhattacharya, M. (2007). Design of Virtual Learning Environments for Deep Learning. *Journal of Interactive Learning Research*, 18(1), 55–64.
- Mukhopadhyay, P. (2006). VidyaOnline: Design and Development of a FOSS based Virtual Learning Environment on Library and Information Science. *DRTC – ICT Conference on Digital Learning Environment*.
- Ngai, T., Poon, L., & Chan, C. (n.d.). Empirical examination of the adoption of WebCT using TAM. *Computers & Education*. doi:10.1016/j.compedu.2006.11.007
- Ong, L., Lai, J.-Y., & Wang, Y.-S. (2004). Factors affecting engineer's acceptance of asynchronous e-learning systems in high-tech companies. *Information & Management*, 41(6), 795–804. doi:10.1016/j.im.2003.08.012
- Piccoli, G., Ahmad, R., & Ives, B. (2001). Web-based virtual learning environments: A research framework and a preliminary assessment of effectiveness in basic IT skills training. *Management Information Systems Quarterly*, 25(4), 401–426. doi:10.2307/3250989

Pituch, K. A., & Lee, Y.-K. (2006). The influence of system characteristics on e-learning use. *Computers & Education*, 47(2), 222–244. doi:10.1016/j.compedu.2004.10.007

Selim, H. M. (2003). An empirical investigation of student acceptance of course websites. *Computers & Education*, 40(4), 343–360. doi:10.1016/S0360-1315(02)00142-2

Singh, I., & Devi, M. (2009). *Virtual learning environment: issues and challenges before LIS schools and libraries*. Retrieved from <http://hdl.handle.net/1944/1470>

Stiles, M. (2007). Death of the VLE?: A challenge to a new orthodoxy. *Serials*, 20(1), 31–33. doi:10.1629/20031

Sun, H., & Zhang, P. (2006). The role of moderating factors in user technology acceptance. *International Journal of Human-Computer Studies*, 64(2), 53–78. doi:10.1016/j.ijhcs.2005.04.013

The Joint Information Systems Committee (JISC). (2000). *Circular 7/00: MLEs in Further Education: Progress report*. Retrieved from http://www.jisc.ac.uk/index.cfm?name=news_circular_7_00

Venkataraman, S., & Sivakumar, S. (2015). Engaging students in Group based Learning through e-learning techniques in Higher Education System. *International Journal of Emerging Trends in Science and Technology*, 2(01).

Walker, B. L., & Harrington, S. S. (2004). Can nursing facility staff with minimal education be successfully trained with computer-based training? *Nurse Education Today*, 24(4), 301–309. doi:10.1016/j.nedt.2004.02.004 PMID:15110440

Wilson, S. (2006). Personal Learning Environments: Challenging the dominant design of educational systems. *Journal of e-Learning and Knowledge Society*, 3(2), 27–38. Retrieved from www.sciencedirect.com/science/article/pii/S0378720603001265

Zhang, D., Zhao, J. L., Zhou, L., & Nunamaker, J. F. Jr. (2004). Can e-learning replace classroom learning? *Communications of the ACM*, 47(5), 75–79. doi:10.1145/986213.986216

ADDITIONAL READING

Britain, S. (1999). A Framework for Pedagogical Evaluation of Virtual Learning Environments. JISC Technology Applications Programme: Report. 41.

Dawley, L., & Dede, C. (2014). Situated learning in virtual worlds and immersive simulations. In *Handbook of research on educational communications and technology* (pp. 723–734). Springer New York. doi:10.1007/978-1-4614-3185-5_58

Hamada, M. (2007). Web-based active e-learning tools for automata theory. Paper published in the Proceedings of Seventh IEEE International Conference on Advanced Learning Technologies, Niigata, Japan, July 18–20, 2007. Doi:10.1109/ICALT.2007.283

Haythornthwaite, C., Kazmer, M. M., Robins, J., & Shoemaker, S. (2000). Community development among distance learners: Temporal and technological dimensions. *Journal of Computer-Mediated Communication*, 6(10). Retrieved from <http://jcmc.indiana.edu/vol6/issue1/haythornthwaite.html>

Hölbl, M., & Welzer, T. (2015). Students' feedback and communication habits using Moodle. *Elektronika ir Elektrotechnika*, 102(6), 63–66.

Hotline, W. (2014). Top trends in academic libraries: A review of the trends and issues affecting academic libraries in higher education. *College & Research Libraries News*, 75, 294–302.

Islam, M. S., Kunifuji, S., Miura, M., & Hayama, T. (2014). How library and information science academic administrators perceive e-learning in LIS schools A qualitative analysis. *IFLA Journal*, 40(4), 254–266. doi:10.1177/0340035214541401

Niemi, H., Harju, V., Vivitsou, M., Viitanen, K., Multisilta, J., & Kuokkanen, A. (2014). *Digital Storytelling for 21st-Century Skills in Virtual Learning Environments*. Creative Education.

Secker, J. (2004). *Electronic Resources in the Virtual Learning Environment*. Chandos Publishing. doi:10.1533/9781780630700

Stichter, J. P., Laffey, J., Galyen, K., & Herzog, M. (2014). iSocial: Delivering the social competence intervention for adolescents (SCI-A) in a 3D virtual learning environment for youth with high functioning autism. *Journal of Autism and Developmental Disorders*, 44(2), 417–430. doi:10.1007/s10803-013-1881-0 PMID:23812663

KEY TERMS AND DEFINITIONS

E-Learning: Education via the Internet, network, or standalone computer. It is essentially the network-enabled transfer of skills and knowledge. E-learning applications and processes include Web-based learning, computer-based learning, virtual classrooms and digital collaboration.

Massive Open Online Course (MOOCs): An online course aimed at unlimited participation and open access via the web. In addition to traditional course materials such as filmed lectures, readings, and problem sets, many MOOCs provide interactive user forums to support community interactions among students, professors, and teaching assistants (TAs). MOOCs are a recent and widely researched development in distance education which were first introduced in 2008 and emerged as a popular mode of learning in 2012.

Moodle: Moodle is free and open-source software learning management system written in PHP and distributed under the GNU General Public License. Developed on pedagogical principles. It is used for blended learning, distance education, flipped classroom and other e-learning projects in schools, universities, workplaces and other sectors.

Open Source Software: Software with source code that anyone can inspect, modify, and enhance.

Virtual Learning Environment (VLEs): A system for delivering learning materials to students via the web. These systems include assessment, student tracking and collaboration and communication tools.

Teacher Presence

Caroline M. Crawford

University of Houston – Clear Lake, USA

INTRODUCTION

The impact of the online instructor, whom one may label an instructional facilitator, within distributed learning environments has become a topic of interest over the previous decades. The impact of the instructor is not only clear within a traditional face-to-face and blended learning environment, but is also vitally important within a distributed learning environments. Although the impact of the instructor upon the learner's experience and subject matter understanding has not been fully understood, the needs revolving around instructor presence within distributed learning environments are not merely motivational in nature. Instead, instructor presence also serves as a facilitative guide, self-regulatory maven, cognitive load support system, and the mentor-focused instructional effort towards not merely cognitive understanding but also towards higher order thinking skills and associated subject matter engagement.

BACKGROUND

Within the bounds of a discussion related to the concept of teacher presence within an instructional environment, there has been intriguing work revolving around the concept of the teacher's impact upon the instructional environment for decades. Of interest is the distinct engagement within this subject during the 1980s and again arising in the later 1990s due to the introduction of instructional technologies and distributed learning environments into the instructional environment. A discussion revolving around understandings related to teacher presence and subsequent impacts upon the instructional environment are addressed. Se-

DOI: 10.4018/978-1-5225-2255-3.ch689

cifically, focusing upon the subjects of discourse, reflective practices, and motivational needs of the learner.

Towards supporting this understanding, Laidyschewsky (2013) engaged in the impact of the instructor's presence upon the learner satisfaction, and is in strong company with current instructor presence engagement in online courses (Bowers & Kumar, 2015; Hodges & Cowan, 2012; Lowenthal, 2009; Sheridan & Kelly, 2010) and associated multimedia-based efforts towards the suggestion of instructor presence (Young, Hicks, Villa-Lobos & Franklin, 2014), as well as professional associations' efforts towards supporting the engagement of teachers within online learning environments (NEA, n. d.). The teacher presence efforts and associated student perceptions surrounding teacher presence are currently intriguing and worthy of further consideration.

Discourse

The concept of discourse within an instructional event is of significant interest. The importance of discourse is not only due to the inherent social engagement of the learner within an instructional environment, but also upon the understanding of the teacher's instructional and facilitative impact upon the learner. As well, the communal understandings revolving around practices within the institutions that undergird the instructional efforts directly impact the concept of discourse within the instructional environment.

Social Discourse: Words We Use

The instructional process is a socially framed effort, no matter whether learning occurs within a

face to face instructional environment, a blended learning approach, a distributed learning approach or even a mobile instructional environment. The ways that people learn new knowledge and information is framed through a socially mediated environment. For this reason, an introductory discussion that focuses upon the words that are implemented throughout the social discourse effort are vitally important. An initial point of interest is Wittgenstein's (1961) research that engaged in an understanding about the words that people implement within an instructional environment, not only to frame an understanding of the knowledge but also perceptions revolving around the nuanced understandings of the knowledge. The word choices of the instructors, the word choices of the learners and perceptions of understanding as framed through these word choices directly impact the learning environment's social discourse. Word choices reflect one's understanding of the knowledge, as well as the nuanced subject matter understandings. The simplistic shift of the instructor's word choice, from "student" to "learner colleague" or merely "colleague", directly impacts the perception of the learners within the instructional environment. One example is offered by a master's student towards the end of a semester-long course, when stated "I also appreciate that you view us as colleagues and not just students. Thank You!!" (personal communication, 27 April 2016), suggesting that the labels and terminology articulated by the instructor not only reflects upon the community endeavor and style of engagement but also upon the perception of the instructor's presence within the course environment.

Further, one may consider Vygotsky's (1933/1966, 1935, 1981) work as pertains to the ways that people understand the knowledge within the instructional environment as conceptually framed through social understandings. Ways that instructors and learners talk about the knowledge within an instructional environment is important, just as the socially relevant engagement with the subject matter more fully frames the learner's understandings of the information through

knowledge checks and corrections in developing understandings of the information. Following on this thought process is Cherryholmes' (1988) structural understanding of discourse as it is bound by the conception of time and place. Meaning, the instructionally relevant discourse that occurs within a socially relevant and engaged environment is supported by the short-term memory understandings that are bound by the instructor's and collegial learner's efforts. Discourse is inherently framed through a cognitively conceptual engagement within a specific time and place cognitive understanding. The usage of the information learned falls within a specific locale and a specific point in time; as such, understanding this information is also bound by the developmental appropriateness of the learner working with this information, as well as the engagement of the instructor within the social discourse that further frames the learner's understanding of the information. Of interest is the shift from social discourse towards patterns of behavior, distinctly framed within the practices of the instructional institutions; although each is directly relevant and supports the other, these are distinctly separate events and practices.

Patterns of Behavior and Practices of Institutions

The language of the instructional environment and the language choices of the instructor directly impact the learner within any type of instructional environment. Not only does the type of discourse occurring reflect differentiated patterns of behavior directly related to the subject matter understanding within the field. However, the instructor's discourse style and discourse engagement with the learners may directly impact the learner's understanding of the subject matter as well as the learner's success associated with knowledge attainment and holistic engagement with the subject matter (Smith & Zantiotis, 1988). The terms implemented by the instructor, the ways through which the chosen texts represent the information and even the multimedia products

that digitally represent the subject matter integrate styles of discourse and subject matter framing of thought processes that can support the learner. On the other hand, the disengagement of the learner from understanding or even motivation towards learning more about the subject matter presented is a realistic concern. The instructor's curricular choices, the instructor's instructional design efforts, and the instructor's style of discourse and engagement support between the subject matter and the learner each leads towards a recognition of the importance of the teacher throughout the instructional process. The teacher is present throughout each instructional choice, analyzing the appropriateness of the subject matter presented to the learner, framing the learner's understanding of the knowledge base and subsequent working with the new information in different cognitively creative manners. As such, all levels of engagement revolve around the teacher presence throughout the subject matter discourse efforts.

Reflective Practices

During the latter half of the last century, a developing interest in the teacher's ability to reflect upon their instructional abilities and engagement support between the subject matter and the learner's cognitive understandings became a prime area of recognition within a constructivist understanding of learning and engagement. As such, Cruikshank's Reflective Teaching Model (1988), Schon's Reflection in Action (1983, 1987) and Zeichner's Three Levels of Reflection (1981, 1983; Zeichner & Liston, 1987) were important models and theories representing the shifting understandings of teacher presence within the instructional environment, and the importance revolving around the instructor's professional analytical efforts as regards the instructional process.

Cruikshank's Reflective Teaching Model

Cruikshank's recognition of the importance of reflective practices within the teaching profes-

sion developed into a Reflective Teaching Model (1988) that purported to support the continual professional development of the instructor through continuously reflecting upon the instructional practices inherent within the learning environment. The ability of an instructor to continuously and critically analyze one's own instructional practices and instructional impact upon the learners and the learning environment were inherent needs. The ability of an instructor to analyze one's own effectiveness within an instructional environment, not only within the holistic instructional environment but also within a learner-by-learner impact analysis, are suggested as being reflective analytical efforts that suggest an integrated understanding of one's own instructional impact beyond themselves as teacher individuals and taking into account the instructional community impact.

Schon's Reflection in Action

Schon's Reflection in Action (1983, 1987) work is less interested in a strictly framed reflective analytical engagement on the part of the course instructor, and more fully engages the instructor in a holistic understanding of instructional impact. As suggested by Schon's statement wherein the focus should be upon the "competency and artistry already embedded in skillful practice" (1987, p. xi), the inherent professionalism of the instructor and the subsequent instructional presence throughout the learning process is more fully acknowledged by Schon's work.

Also, recognized within Schon's work are three primary areas of focus: tacit knowledge; knowing in action; and, reflection in action (1987). The concept of tacit knowledge is the reflection of the instructor's instructional presence through the experience of the professional instructor's ability to engage the learner with the subject matter. Again, as suggested by Schon (1987), "We reveal it by our spontaneous, skillful application of the performance; and we are characteristically unable to make it verbally explicit" (p. 25), suggesting the actively engaged instructor's presence with the

knowledge and “knowing” within the instructional environment, suggested as being actively vibrant and framed within a circumstance or condition that is naturally occurring within an instructional environment. The final consideration is the ability of an instructor to be reflective in instructional action and presence; specifically, this is the ability of an instructor to unconsciously and consciously reflect as one is progressing through an instructional event or an instructional design effort, while at the same time being able to recognize the verbal and nonverbal cues of the learners as pertains to the level of understanding and engagement with the subject matter. An example of this effort is towards an instructor’s ability to showcase their analytical thought process, either through the word processed communicative style, a video-based presentation of an analytical reflection, or perhaps a synchronous video conference event wherein the learners have the cognitive efforts of the instructor’s expertise on display so as to reflect a way through which to analyze and think about a process. This style of communication may initially suggest an instructor’s vulnerability, as an analytical thought process is different for each person and the reflective efforts of the instructor on display may also be initially uncomfortable to make public within an instructional venue. However, allowing the learners to engage in this style of thoughtful understanding supports Schon’s work (1987) as framed through not only tacit knowledge, knowing in action and reflection in action.

As concerns the verbal and nonverbal cues of the learners, these efforts will represent themselves differently as based upon the learner personality and strengths, weaknesses and vulnerabilities of the learners. Meaning, the instructor’s presence within the course can support and strengthen the ability of the learners towards reflecting verbal and nonverbal cues of the learners, through innumerable differentiated means. One example may be through publicly thanking a learner for asking questions “early” and “often”, with special reference to the fact that this learner proactively engaged in querying the instructor for clarification

so that the rest of the course colleagues will have this support readily available when engaging in the course effort. Another example, may be to thank a learner for a private communicative request for explanation and support, with a follow-up request that the learner please consider sharing the question-answer effort with the other course colleagues so as to support everyone else within the course. These efforts may appear simplistic in nature, but engaging in a community effort towards course understanding and proactive effort not only highlights the proactive learner as knowledgeable, on task and reflective, but also as a collegial learner who is willing to share and support learner colleagues. This style of engagement and instructional presence more fully supports the learner’s ability towards representing verbal and nonverbal cues within a “safe” and supportive environment wherein questions and concerns are welcomed, appreciated and fully addressed due to the instructor presence and interest in the learners’ success. As well, instructor presence is also more fully engaged by addressing differentiated verbal and nonverbal cues of the learners through considering the non-stated concerns of learners. An example of this effort occurred recently, due to city-wide floods that impacted learners living within the area. One learner’s home was partially flooded and the vulnerability was reflected within the course environment through difficulty with scheduling team-based efforts. Instead of the issues focusing upon the scheduling angst, what was actually occurring was the learner’s upset by the team’s inability to emotionally support and differentiate previously scheduled meeting dates and times so as to support the learner who was in the midst of personal and professional chaos. The instructor presence and knowledge of the learners in the team reflected upon the angst-ridden communications by the one learner, considering nonverbalized word processed cues that required the instructor to not only offer a nurturing response and recognition to the current learner situation, while also attempting to subtly reflect upon the other team member’s differentiated responses to

difficult situations. Meaning, when different personalities find themselves in difficult situations, some persons “go with the flow” while others become more structured and display a need to maintain order, control and structure. Although this style of response to learner cues requires the instructor presence to move beyond a knowledge-based course understanding of the learners, the reflective efforts and understandings of differentiated learner needs and efforts within a course environment mandate the ability to more fully understand the differentiated meanings behind learner verbal and nonverbal cues.

The instructor who is a skilled professional has the ability to ensure their instructional presence to analyze the immediate context within the instructional environment and learner understanding, with the subsequent ability to seamlessly shift instructional efforts towards more fully supporting the learner’s subject matter understanding and engagement. The teacher’s ability to ensure their own cognitive and professional understandings as the lesson progresses, supports the learner’s through the ability to reflect in action. One’s own instructional practice and prowess supports the teacher presence.

Zeichner’s Three Levels of Reflection

Zeichner supports both Cruikshank’s and Schon’s understandings of reflective practice, but method, performance and practical, methodical proficiency may be considered necessary but not sufficient. Zeichner’s (1981, 1983; Zeichner & Liston, 1987) understanding of reflective practices is threefold, beginning with technical elements such as method and performance efforts with the primary impact being the instructional strategies and successes within the instructional environment. Reflecting upon these basic masterful elements of style and strategy are necessary and appropriate. Secondly, situational and instructional contexts are addressed within a reflective manner. Reflecting upon and analyzing why instructional contexts were implemented, as well as why situational contexts were

inherently appropriate within the instructional process are necessary. This also includes not only the straightforward curricular efforts and subject matter instructional efforts, but this also includes analytical reflections upon to what extent any hidden curricula or hidden agendas might be in play. A third level within Zeichner’s levels of reflection revolve around social justice and equity efforts, stated as moral and ethical issues upon which an instructor must analytically reflect. As stated by Adler (1991), “Reflections at this level asks that teachers become, in Henry Giroux’s (1988) terms, ‘transformative intellectuals,’ who are capable of examining the ways in which schooling generally, and one’s own teaching specifically, contribute or fail to contribute, to a just and humane society” (p. 11). Zeichner’s expectation that reflective instructors are fully present within the instructional process, not only towards addressing the instructional needs of the learners as regards working with the subject matter, but also towards social justice and equity concerns.

Motivational Needs of the Learner

The instructor’s presence has an inherent impact upon the learner’s sense of motivation throughout the instructional process. Maslow directly addressed the needs associated with motivational efforts on the part of the learner, specifically as framed through Maslow’s Hierarchy of Needs and associated understandings (1943, 1954, 1962, 1968, 1964/1970a, 1970b). The learners are supported by the instructor’s instructional presence and recognition of the inherent needs of the learners, towards engaging in addressing levels of deficiency needs (physiological, safety, social and esteem) and overcoming these needs, with the expectation being to strive towards ensuring that each learner reaches the learner’s level of growth needs (self-actualization). This hierarchy of needs, as delineated by Maslow, reflects the impact of the instructional presence within the learning environment; more specifically stated, the instructor’s presence and engagement with the learners

and within the instructional environment directly supports the deficiency needs of each learner more directly, with the desire towards supporting each learner's efforts towards reaching the self-actualization level that offers an engaged learner who may deftly engage with the subject matter, the learning process and the collegial learners at a level of growth and strength.

TEACHER PRESENCE: IMPLICATIONS WITHIN THE DISTRIBUTED LEARNING ENVIRONMENT

Within a distributed learning environment, the instructional presence of the teacher-facilitator is even more necessary than what may be perceived within a face to face traditional instructional environment or even a blended learning approach towards learning. Defining distributed learning environments as online and mobile learning environments, the instructional presence of the teacher-facilitator is enhanced due to the time and location concerns of presence. Towards addressing the teacher presence implications within the distributed learning environments, discussions revolving around motivational and cognitive support are offered; specifically, a focus upon facilitative instructor guidance, self-regulatory instructional efforts, cognitive load support systems and mentor-focused instructional efforts engage in motivational and cognitive support efforts more fully frame aspects of teacher presence with more specific implications within the distributed learning environment are fully engaged.

The motivational and cognitive support inherent within an instructional environment is even more pronounced within a distributed learning environment. As previously mentioned, Maslow's Hierarchy of Needs (1943, 1954, 1962, 1968, 1964/1970a, 1970b) supports an understanding of the deficiency needs and growth needs of each learner. However, the instructor's presence within the distributed learning environment is

necessary as *perceived* and *actual* presence support systems. Within this section's discussion, each aspect is inherently *perceived* instructional presence referring to the instructional facilitator's prior, asynchronous efforts towards supporting the learner's cognitive perception of teacher engagement; this may be perceived as previously recorded and available video, audio, slideshow and other forms of subject-focused materials that offer the perception of the instructor's presence and immediacy within the instructional environment for purposes of learner motivational efforts and perceived cognitive support structures, yet the instructor is not actually synchronously available "live" to the learner at that point in time. On the other hand, the *actual* instructional facilitator's presence in the course may be realized by the learners through synchronous efforts within the course structure such as video conferencing, telephone conversations, texting, chats, and other synchronous events that reflect the true timely availability and structured teacher presence of the instructional facilitator. Even though the learner and instructor have differentiated time and location realities, the availability of the instructional facilitator within an easily available and less structured environment offers motivational and cognitive support for the learner who desires or requires non-traditional teacher presence so as to perceive a sense of well-being and engagement within the bounds of the distributed learning environment.

Facilitative Guide

Within an instructional environment, the instructor's presence is offered as a facilitative guide. The instructor engages in the efforts towards being a "touch point" within the instructional process, towards ensuring that each learner progresses naturally and within a supported manner through the instructional environment. Facilitating the learner's efforts within a course environment may be perceived as not only motivational in nature, but also towards supporting the learner's ease of environmental understanding which may be per-

ceived as supporting the learning environment's usability on the part of the learner. As a facilitative guide, the instructional presence also supports a stronger understanding effort revolving around the subject matter, introducing new subject-specific information within a timely fashion.

Self-Regulatory Maven

The instructor's presence within the instructional environment more fully supports the concept of a self-regulatory maven. Self-regulation is an area of study that is fully framed within all types of instructional environments (Schunk, 1994; Schunk & Zimmerman, 1997, 2003; Zimmerman, 1989, 1990, 1994) and strongly reflects the needs of the learner towards fully framing the learner's ability to naturally control and progress through an instructional structure. As Zimmerman describes a self-regulated learner, "They approach educational tasks with confidence, diligence, and resourcefulness. Perhaps most importantly, self-regulated learners are aware when they know a fact or possess a skill and when they do not" (1990, p. 4). Although a self-regulated learner is recognized as controlling one's own environment and controlling one's own learning, recognition must be offered towards the instructional facilitator's presence towards also supporting the learner's efforts towards self-regulation. A self-regulatory maven not only designs support structures into the instructional environment towards guiding and supporting the learner, but the strength of environmental usability and clarity of expectations is also a supportive self-regulatory maven's strength.

Cognitive Load Support System

Within distributed learning environments, the teacher presence is inherently integral. The instructional facilitator's efforts towards supporting the learner's cognitive load efforts works towards ensuring that the learner's efforts towards understanding the subject matter are not overwhelming in nature. The introduction and representation of

information within a learner's working memory can be overwhelmed with too much information offered too soon within a learner's progressive understanding of the information, reflecting a cognitive load issue (Chandler & Sweller, 1991, 1992, 1996; Sweller, 1988, 1989, 1991, 1994; Sweller & Chandler, 1994; Sweller, Chandler, Tierney & Cooper, 1990; Sweller & Cooper 1985; Sweller & Levine, 1982; Sweller, Mawer & Ward, 1983; Sweller, Van Merriënboer, & Pass, 1998; Ward & Sweller, 1990). As quoted from Sweller (1994), "The difficulties we face when learning new intellectual tasks can fluctuate dramatically. Learning can vary from being trivially easy to impossibly hard. Some of the reasons for variations in ease of acquisition, such as changes in amount of information, are obvious" (p. 195). The instructional presence within the distributed learning environment supports the learner's efforts, so that the "trivially easy" cognitive efforts are opportunities are initial opportunities for the instructor to engage in appropriate cognitive efforts through teacher presence efforts, so that the "impossibly hard" cognitive efforts of the learners have been previously supported by prior teacher presence efforts and the learners may have developed a level of comfortableness with the instructional facilitator so that communicative and relational efforts towards understanding the subject matter will have already become natural progressions in understanding on the part of the learners.

Mentor-Focused Instructional Efforts

The instructional facilitator within the distributed learning environment engages in a community of practice environment (Lave & Wenger, 1991; Wenger, 1998, 2009, 2010; Wenger, McDermott & Snyder, 2002) wherein an instructor's presence may be reflected as a mentor-style relationship wherein the instructor holds a mentorship role and the learner holds an apprenticeship role. Within this instructional environment, the instructional facilitator engages in a style of presence wherein the learner's efforts revolving around the subject

matter are guided and supported by the instructor's efforts. Within this framework of instructional effort, three areas of emphasis are recognized as imperative: cognitive understanding; higher order thinking skills; and, subject matter engagement.

Cognitive Understanding

Within a mentor-focused instructional effort environment, the learner's cognitive understanding revolving around the subject matter is constructed and supported by the instructional facilitator's presence, inherent within the instructional design process, the subject matter discussions that may integrate additional subject matter discussions, ways of looking at the information through different lenses of understanding and thought process, as well as a Socratic questioning effort wherein a style of thoughtful yet disciplined dialogue ensues, towards logically reviewing the subject matter in different ways of meaning and validating those understandings through checks, balances and engaging in new and different understandings of the subject matter.

Higher Order Thinking Skills

Integrally important within the instructional environment is the instructional facilitator's presence and engagement with the subject matter and the learners, towards slowly working with the subject matter from lower order thinking skills engagement (knowledge, comprehension and application) towards higher order thinking skills of analysis, synthesis and evaluation, as framed through Bloom's Taxonomy of the Cognitive Domain (Bloom, 1956, 1984; Bloom, Englhart, Furst, Hill & Krathwohl, 1956; Bloom & Krathwohl, 1956; Krathwohl, Bloom & Masia, 1964). Further, Anderson and Krathwohl's revised Cognitive Taxonomy for the Digital Age (2001) suggested a shift from the lower order thinking skills designations of remember, understand and apply, with the higher order thinking skills designations framed

as analyze, evaluate and create. The instructional facilitator's presence within the learning environment supports the learner's efforts from the basic level of knowledge acquisition and remembering, slowly progressing from the novice subject matter understanding through the differentiated levels of subject matter cognitive efforts, with the desired engagement of the learner with the subject matter at the higher order thinking skills levels. The teacher presence further supports and engages the learner's efforts within the instructional environment, shifting the learner's understanding over a longitudinal period from novice towards developing an enhanced understanding of the subject matter.

Subject Matter Engagement

The subject matter engagement within the distributed learning environment is multifold for the learner, as well as for the instructional facilitator. Much as was described within the previous discussion, the ability to work with the subject matter in novice levels of engagement and then slowly engaging with the subject matter in new and cognitively differentiated manners is vitally important towards more fully reflecting the learner's successful attainment of information from a learning environment.

SOLUTIONS AND RECOMMENDATIONS

Further understandings revolving around the teacher presence and need within an instructional environment is vital. Towards addressing different recommendations towards supporting teacher presence within the instructional environment, whether face to face or blended or online learning environment in nature, suggestions towards solutions are necessary. One recommendation is towards supporting a novice instructor's teacher presence through a mentorship style engagement.

One style of mentorship may reflect a teaching assistant style of training effort wherein the expert instructor supports and engages the novice instructor in an appropriately timed engagement with the learners. The mentor would not only model differentiated teacher presence engagement to the novice instructor, but also engage in extensive reflections and explanations as to the reasoning and past experiences that have led to each approach. Another style of mentorship could be a “guide on the side” approach, wherein the expert instructor would take a voyeur’s stance towards supporting the novice instructor’s teacher presence engagement. The expert instructor may engage in online courses as merely a silent administrative access role, to “check in” and look around the course environments to see what is happening and consider ways to support the novice instructor. On the other hand, the expert instructor and novice instructor may desire a less proactive effort, wherein the expert instructor may choose engage in a support style that requires the novice instructor to communicate and request the mentor’s support and guidance when requested. The primary concern with this style of “hands off” approach is the lack of teacher presence understanding that may be unrealized by the novice instructor; understanding the ability to recognize what is not yet understood is imperative within this style of mentor effort, as the concept of “I don’t know what I don’t know” may be detrimental to the course experience of the learner as well as the novice instructor’s developing understanding of the ebb and flow of a learning environment and the associated development of the learning community. The concept of discourse throughout an instructional environment is necessary and appropriate, while the teacher’s needs towards delving into reflective practices will more so engage instructors towards understanding the needs of the learners, the engagement with the subject matter, as well as the motivational and cognitive support needs of the learners.

FUTURE RESEARCH DIRECTIONS

Emerging trends revolving around distributed learning environments, especially within the realm of mobile learning, articulate a need for strengthened understanding of the teacher presence impact upon learners. Instructional facilitators may naturally engage in aspects of teacher presence, but the implications towards future research directions more fully articulate impactful research efforts and associated professional development needs of all instructional facilitators. An initial consideration towards future research directions is focused upon a differentiated thought process, wherein a “one training event” episode is not sufficiently adequate towards developing a strong instructor who can implement teacher presence at a high level of quality engagement. Of course, there are innumerable talented individuals who have the ability to develop their own understanding of successful teacher presence within an instructional environment; however, a focus upon continuous professional development is necessary and appropriate. Continuous support and knowledge base extension through the sharing of ideas, talents and tips by instructor colleagues may be offered as inherently quick and useful short professional development vignettes that develop important reflective efforts and instructional design thoughts of instructors who desire to enhance their instructional presence as well as learner satisfaction.

CONCLUSION

The background understandings of discourse, reflective practices and motivational needs of the learners are merely a beginning understanding of teacher presence within a distributed learning environment. More fully understanding teacher presence through motivational and cognitive support of the learners is integrally important,

especially as distributed learning environments become more viable instructional environments wherein a strong teacher presence is necessary and appropriate.

REFERENCES

- Adler, S. A. (1991). The Reflective Practitioner and the Curriculum of Teacher Education. *Journal of Education for Teaching: International Research and Pedagogy*, 17(2), 139–150. doi:10.1080/0260747910170203
- Anderson, L.W., & Krathwohl. (Eds.). (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman.
- Bloom, B. S. (Ed.). (1956). *Taxonomy of educational objectives: The classification of educational goals: Handbook I, cognitive domain*. New York: Longmans, Green.
- Bloom, B. S. (1984). *Taxonomy of educational objectives*. Boston, MA: Allyn and Bacon.
- Bloom, B. S., Englhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (Eds.). (1956). *Taxonomy of educational objectives, the classification of educational goals, Handbook I: cognitive domain*. New York: Longmans.
- Bloom, B. S., & Krathwohl, D. R. (1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals, by a committee of college and university examiners. Handbook I: Cognitive Domain*. New York: Longman, Green.
- Bowers, J., & Kumar, P. (2015). Students perceptions of teaching and social presence: A comparative analysis of face-to-face and online learning environments. *International Journal of Web-Based Learning and Teaching Technologies*, 10(1), 27–44. doi:10.4018/ijwl.2015010103
- Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, 8(4), 293–332. doi:10.1207/s1532690xc0804_2
- Chandler, P., & Sweller, J. (1992). The split attention effect as a factor in the design of instruction. *The British Journal of Educational Psychology*, 62(2), 233–246. doi:10.1111/j.2044-8279.1992.tb01017.x
- Chandler, P., & Sweller, J. (1996). Cognitive load while learning to use a computer program. *Applied Cognitive Psychology*, 10(2), 151–170. doi:10.1002/(SICI)1099-0720(199604)10:2<151::AID-ACP380>3.0.CO;2-U
- Cherryholmes, C. H. (1988). *Power and criticism: Poststructural investigations in education*. New York: Teachers College Press.
- Cruikshank, D. R. (1987). *Reflective teaching: The Preparation of Students of Teaching*. Reston, VA: Association of Teacher Educators.
- Giroux, H. A. (1988). Teachers as transformative intellectuals. In H. A. Giroux (Ed.), *Teachers as Intellectuals: Toward a Critical Pedagogy of Learning*. Westport, CT: Bergin & Garvey Publishers, Inc.
- Hodges, C. B., & Cowan, S. F. (2012). Preservice teachers views of instructor presence in online courses. *Journal of Digital Learning in Teacher Education*, 28(4), 139–145. doi:10.1080/21532974.2012.10784694
- Krathwohl, D. R., Bloom, B. S., & Masia, B. B. (1964). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: The affective domain*. New York: David McKay.
- Ladyshewsky, R. K. (2013). Instructor presence in online courses and student satisfaction. *International Journal for the Scholarship of Teaching and Learning*, 7(1), 13. doi:10.20429/ijstl.2013.070113

- Lave, J., & Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511815355
- Lowenthal, P. R. (2009). Social presence. In P. Rogers, G. Berg, J. Boettcher, C. Howard, L. Justice, & K. Schenk (Eds.), *Encyclopedia of distance and online learning* (2nd ed.; pp. 1900–1906). Hershey, PA: IGI Global.
- Maslow, A. H. (1943). A Theory of Human Motivation. *Psychological Review*, 50(4), 370–396. doi:10.1037/h0054346
- Maslow, A. H. (1954). *Motivation and personality*. New York: Harper and Row.
- Maslow, A. H. (1962). *Towards a psychology of being*. Princeton, NJ: D. Van Nostrand Company. doi:10.1037/10793-000
- Maslow, A. H. (1964/1970a). *Religions, values, and peak experiences*. New York: Penguin.
- Maslow, A. H. (1968). *Toward a Psychology of Being*. New York: D. Van Nostrand Company.
- Maslow, A. H. (1970b). *Motivation and personality*. New York: Harper & Row.
- National Education Association (NEA). (n.d.). *Guide to teaching online courses*. Retrieved from <http://www.nea.org/assets/docs/onlineteachguide.pdf>
- Schon, D. A. (1983). *The reflective practitioner: How professionals think in action*. New York: Basic Books, Inc.
- Schon, D. A. (1987). *Educating the reflective practitioner: Toward a new design for teaching and learning in the professions*. San Francisco: Jossey-Bass.
- Schunk, D. H. (1994). Self-regulation of self-efficacy and attributions in academic settings. In D. H. Schunk & B. J. Zimmerman (Eds.), *Self-regulation of learning and performance: Issues and educational implications* (pp. 75–99). Hillsdale, NJ: Erlbaum.
- Schunk, D. H., & Zimmerman, B. J. (1997). Social origins of self-regulatory competence. *Educational Psychologist*, 32(4), 195–208. doi:10.1207/s15326985ep3204_1
- Schunk, D. H., & Zimmerman, B. J. (2003). Self-regulation and learning. In W. M. Reynolds, G. E. Miller, & I. B. Weiner (Eds.), *Handbook of Psychology* (Vol. 7, pp. 59–78). Hoboken, NJ: John Wiley & Sons, Inc. Retrieved from http://t5303.oceanikpsi.org/downloads/handbook_of_psychology_vol_.pdf#page=83
- Sheridan, K., & Kelly, M. A. (2010, December). The indicators of instructor presence that are important to students in online courses. *Journal of Online Learning and Teaching*, 6(4). Retrieved from http://jolt.merlot.org/vol6no4/sheridan_1210.htm
- Smith, R., & Zantiotis, A. (1988). Practical teacher education and the avant garde. *Journal of Curriculum Theorizing*, 8(2), 77–106.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285. doi:10.1207/s15516709cog1202_4
- Sweller, J. (1989). Cognitive technology: Some procedures for facilitating learning and problem solving in mathematics and science. *Journal of Educational Psychology*, 81(4), 457–466. doi:10.1037/0022-0663.81.4.457

- Sweller, J. (1991). Some modern myths of cognition and instruction. In J. B. Biggs (Ed.), *Teaching for Learning: The view from cognitive psychology*. ACER, Radford House.
- Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, 4(4), 295–312. doi:10.1016/0959-4752(94)90003-5
- Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction*, 12(3), 185–233. doi:10.1207/s1532690xc1203_1
- Sweller, J., Chandler, P., Tierney, P., & Cooper, M. (1990). Cognitive load and selective attention as factors in the structuring of technical material. *Journal of Experimental Psychology. General*, 119, 176–192. doi:10.1037/0096-3445.119.2.176
- Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction*, 2(1), 59–89. doi:10.1207/s1532690xc10201_3
- Sweller, J., & Levine, M. (1982). Effects of sub-goal density on means-ends analysis and learning. *Journal of Experimental Psychology. Human Learning and Memory*, 8(5), 463–474. doi:10.1037/0278-7393.8.5.463
- Sweller, J., Mawer, R., & Ward, M. (1983). Development of expertise in mathematical problem solving. *Journal of Experimental Psychology. General*, 112(4), 634–656. doi:10.1037/0096-3445.112.4.639
- Sweller, J., Van Merriënboer, J., & Paas, F. (1998). Cognitive Architecture and Instructional Design. *Educational Psychology Review*, 10(3), 251–296. doi:10.1023/A:1022193728205
- Vygotsky, L. S. (1933/1966). Play and its role in the mental development of the child. *Social Psychology*, 12(6), 62–76.
- Vygotsky, L. S. (1935). *Mental development of children during education*. Moscow: Uchpedgiz.
- Vygotsky, L. S. (1981). The genesis of higher mental functions. In J. V. Wertsch (Ed.), *The concept of activity in Soviet psychology*. Armonk, NY: Sharpe.
- Ward, M., & Sweller, J. (1990). Structuring effective worked examples. *Cognition and Instruction*, 4, 137–166.
- Wenger, E. (1998). *Communities of Practice: Learning, Meaning, and Identity*. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511803932
- Wenger, E. (2009). *Digital Habitats*. Portland, OR: CPsquare.
- Wenger, E. (2010). Communities of practice and social learning systems: the career of a concept. In C. Blackmore (Ed.), *Social Learning Systems and communities of practice*. Springer Verlag and the Open University. doi:10.1007/978-1-84996-133-2_11
- Wenger, E., McDermott, R., & Snyder, W. M. (2002). *Cultivating Communities of Practice*. Harvard Business Press.
- Wittgenstein, L. (1961). *Tractatus logico-philosophicus* (D. F. Pars & B. F. McGuinness, Trans.). New York: Humanities Press.
- Young, I. I., Hicks, W. A., Hl, B., Villa-Lobos, D., & Franklin, T. J. (2014, December). Using student feedback and professor-developed multimedia to improve instructor presence and student learning. *Journal of Teaching and Learning with Technology*, 3(2), 12–30. doi:10.14434/jotlt.v3n2.12990
- Zeichner, K. M. (1981, December). Reflective teaching and field based experience. *Interchange*, 12(4), 1–22. doi:10.1007/BF01807805
- Zeichner, K. M. (1983, May-June). Alternative paradigms of teacher education. *Journal of Teacher Education*, 34(3), 3–9. doi:10.1177/002248718303400302

Zeichner, K. M., & Liston, D. P. (1987, April). Teaching student teachers to reflect. *Harvard Educational Review*, 57(1), 23–49. doi:10.17763/haer.57.1.j18v7162275t1w3w

Zimmerman, B. J. (1989). A social cognitive view of self-regulated academic learning. *Journal of Educational Psychology*, 80(3), 329–339. doi:10.1037/0022-0663.81.3.329

Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist*, 25(1), 3–17. doi:10.1207/s15326985ep2501_2

Zimmerman, B. J. (1994). Dimensions of academic self-regulation: A conceptual framework for education. In D. H. Schunk & B. J. Zimmerman (Eds.), *Self-regulation of learning and performance: Issues and educational implications* (pp. 3–21). Hillsdale, NJ: Erlbaum.

KEY TERMS AND DEFINITIONS

Cognition: The way that people think, work with and understand information.

Discourse: Discussions within a social environment, towards more fully understanding the knowledge, towards checking and double-checking one's understanding of information within a social context.

Mentor: A guiding force, meant to support the needs of an apprentice.

Motivation: The desire to move forward, to engage more fully.

Presence: Engagement within the environment, engagement with the subject matter, and availability for the learner.

Reflective Practice: The ability of a person, specifically an instructor, to analyze their actions and understandings within an instructional environment.

Self-Regulation: The ability to control one's actions and reactions.

Virtual Worlds in the Educational Context



Felipe Becker Nunes

Federal University of Rio Grande do Sul, Brazil

Fabício Herpich

Federal University of Rio Grande do Sul, Brazil

Leo Natan Paschoal

University of Cruz Alta, Brazil

INTRODUCTION

Information and Communication Technologies (ICTs) are changing the daily activities of human beings, promoting the process of teaching and learning, facilitating the construction of knowledge. In recent years, ICTs are becoming stable in the educational context, as they allow different ways of presenting content to students (Quintana and Fernández, 2015). Technological resources to support the transmission of knowledge can be represented by heterogeneous formats, such as simulations, games, virtual worlds, videos, animated agents and avatars, web pages, among others.

Virtual worlds appear as technological tools to support construction of knowledge and for this reason has been the subject of study by different researchers over the past few years. There are many educational institutions that use virtual worlds, as they offer to the student an interaction similar to the real world, permits the student to get involved with various activities and learn in an interactive way, acquiring new skills (Mateu et al. 2013). Hulsey, Pence and Hodges (2014) point out that, students represented by avatars can get a better learning experience watching and/or participating in practical activities in virtual worlds.

Abreu et al. (2012) argue that the process of application development focused on education, in addition to the use of principles stemming from

Software Engineering (e.g. extraction requirements, requirements analysis and modeling), should be concerned with the incorporation of learning theories.

There is scientific evidence that shows promising results when researchers link pedagogical concepts in the construction and application of educational systems. Virtual learning environment Moodle, for example, was designed and developed from social constructivist epistemology (Moodle, 2016). Currently, numerous educational institutions, to support the process of teaching and learning, use this environment. In the field of virtual worlds, Nunes et al. (2016) show that use of Mastery Learning theory in the context of virtual worlds for education contributes to reveal optimistic results regarding performance advancement to the students.

Seeking to identify learning theories used in virtual worlds applied in the educational context and obtain an overview of what is being researched and developed in this area, as for in order to investigate the scientific production related to the theme and highlight possibilities for future research, this chapter has order to present a systematic review of literature. Thus, primary studies were researched in article format, indexed by journals and international conferences, considering the period of last five years (2010-2015). To achieve this, research questions was created aiming to consolidating quantitative and qualitative information found.

To present the research conducted; this chapter was organized with a descriptive section about the concepts inherent in virtual worlds, contextualizing the reader about this area. It is also presented some of the currently existing major difficulties and problems in this area, and later presented the listed research and analysis of the results obtained in this study. Finally, future directions for the application of virtual worlds in education and final considerations of this chapter is presented.

BACKGROUND

Also known as, immersive virtual environments, virtual worlds are tools that permits reproducing aspects of the real world or fantasy in a virtual space, providing its users a controlled environment with numerous opportunities and experiences. Griol et al. (2014) corroborate stating that three-dimensional spaces are simulated by computer, which can be exploited in first or third person, through a graphical representation called avatar.

Increased use of virtual worlds in different application contexts, allowed their inclusion in the educational field, from which emerge new opportunities for use of these computing resources as support and even motivation in the learning process of its members. In this sense, is possible to claim that these immersive environments, besides offer a combination of simulation tools, which enhance user interactions and consequently the way they learn, also provide a sense of immersion, as well as the opportunity for communication and collaboration between its users, justifying its application in activities related to the educational field.

There are several possibilities to be explored using virtual worlds applied to education, e.g., as the extension trend of real laboratories for virtual environments, which according to Nunes et al. (2014), is possible offer new opportunities without costs and risks to users, besides to offer opportunity to operating virtual equipment's in controlled environments. It also offers easy access feature and available at all times; immediate

feedback; expertise distributed among users, where exchange of experiences with each other is possible; student-centered learning; and encouraging for practice of authorship.

In virtual worlds, besides users interact with three-dimensional and multimedia objects, they have the possibility to explore the interaction with other connected users using chat to communicate via instant messaging or through voice communication. This feature enables users to build their knowledge jointly, where each student can cooperate with a colleague, e.g., assisting in the resolution of a particular activity or simply discussing their understanding of a topic addressed during their interactions.

Several investigations in different areas has shown the use of virtual worlds with promising results, as well for building scenarios and attractive three-dimensional objects to improve user interactions and promote learning. Implementation of these environments demand that several factors be considered for development, such as: educational objectives, teaching strategies based on learning theories, friendly and instructional design, objects, experiments and exercises that encourage interaction and collaboration among users, besides passing the feeling of being immersed in the environment (Herpich et al., 2016).

ISSUES, CONTROVERSIES, AND PROBLEMS

Taking into account the theoretical review presented above and expertise of the authors of this chapter, it was possible to make a number of considerations about some of the main problems that are included in the scope of this study, specifically the implementation of educational activities in virtual worlds. First, it becomes possible to identify an important point when comparing to a virtual learning environment like Moodle, where these have been developed with specific features for educational use with students, while the virtual worlds need to be adapted to use with this propose.

Because of this adaptation, it becomes important to use programming features through OSSL and LSL language in the virtual world, so they can be inserted into different types of learning materials such as videos, slides, text, links and questions. In addition, simulation of experiments about a particular phenomenon also become complex, as there is a demand for a deeper level of knowledge on the part of the developer to create such resources.

At this point it becomes important to highlight the user's learning curve to use the virtual world, where its basic use can be considered fast to be learned and enjoyed by its users. Regarding the creation of more complex features that involve programming scripts, this process has become costlier, from the moment that makes necessary the use of programming logic and algorithms, as well as other aspects relating to the manipulation of 3D objects in the environment.

VIRTUAL WORLDS APPLIED IN EDUCATION

This study is characterized as a systematic review of the literature. In this perspective, seeks to evaluate and interpret primary research to achieve secondary research objectives (Kitchenham et al., 2007). It was conducted in four phases: (i) definition of the scope of this review; (ii) search and selection of primary studies; (iii) data extraction; (iv) analysis and synthesis.

First phase consisted of the design of the period, in which, the primary research was socialized, covering the range of 2010 to 2015. In addition, the languages of primary research were defined as being Portuguese and English. Five research questions were elaborated, the main bases of scientific data were selected, was constructed a string search, and inclusion and exclusion criteria were established. Seeking to achieve the review objectives, the following research questions (RQ) were structured:

- RQ1:** How are distributed and what are the developments of the investigations over the years?
- RQ2:** What conferences and journals that have the highest number of published articles?
- RQ3:** What are the trends in the context of Virtual Worlds in Education?
- RQ4:** What are the most used learning approaches in the context of virtual worlds applied to education?

To answer these questions, were chosen the following databases (the names in Portuguese was maintained):

- Congresso Internacional de Informática Educativa or International Conference of Educational Computing;
- Workshop de Informática na Escola or Computer Workshop in Schools;
- IEEE Transactions on Learning Technologies;
- IEEE Transactions on Education;
- Congresso da Sociedade Brasileira de Computação (Workshop sobre Educação em Computação) or Congress of the Brazilian Computer Society (Workshop on Computer Education);
- Journal of Virtual Worlds Research;
- Revista Latinoamericana de Tecnología Educativa or Latin American Journal of Educational Technology;
- Simpósio Brasileiro de Informática na Educação or Brazilian Symposium of Information Technology on Education;
- Computers & Education;
- Revista Novas Tecnologias na Educação or Journal of New Technologies in Education;
- Creative Education;
- Informática na Educação: teoria & prática or Computers in Education: Theory & Practice;
- IEEE Computer Graphics and Applications;
- IEEE Revista Iberoamericana de Tecnologías del Aprendizaje or Iberoamerican Journal of Technology Apprenticeship.

To locate the primary research in the stipulated databases, the string search, based on the research questions, was established. String search was composed of keywords considering its versions in Portuguese and English. In some databases where was not include the automatic search, manual search method was applied.

The string consists of the following operators: “AND”, “OR”. Thus, the logical OR operator is used to indicate that there are alternative terms and “AND” is used to connect two terms, as it can be seen in the following:

- (“mundos virtuais” OR “ambientes virtuais imersivos” OR “ambientes imersivos” OR metaversos OR “mundos digitais virtuais 3D” OR “ambientes virtuais multi usuá-rios” OR “virtual worlds” OR “immersive virtual environments” OR “immersive environments” OR metaverses OR “digital virtual worlds 3D”) AND (educação OR ensino OR aprendizagem OR education OR teaching OR learning).

To determine the inclusion of articles, considering some factors that identify the relevance of primary research to address the research questions, inclusion and exclusion criteria were created (Table 1). They were used as parameters to aid in the screening of publications.

In the second phase of the research, there was the implementation of the search string in the selected research sources, aiming to collect primary research. 419 articles were founded in the initial search and databases of IEEE Transactions on Education did not return results. Localized items passed to the third stage of the research.

In the third phase, were used the inclusion and exclusion criteria defined during the planning (phase I), were classified as accepted or rejected. This phase consisted of reading the titles, abstracts and keywords of 419 articles, were at end of this process only remained 113 included and 306 excluded articles. Later, still in the third phase, 113 articles were fully analyzed and criteria for

Table 1. Inclusion and Exclusion Criteria

Inclusion Criteria	Exclusion Criteria
IC - 1: Presents the virtual worlds applied to education in various fields of study and platform independent.	EC - 1: Not addresses the actual application of virtual worlds in education, i.e., only has one of those terms in the article.
IC - 2: Presents results that are consistent with the objective listed in the article and the instruments used in the evaluation of virtual worlds applied to education.	EC - 2: Technical reports, documents that are available in the form of summaries or presentations and secondary literature reviews.
	EC - 3: Only addresses theoretical and philosophical aspects (without technical definition, methodology, etc.).
	EC - 4: Not presents the results clearly and instruments used in the evaluation of virtual worlds applied to education.
	EC - 5: Papers written in a language other than English or Portuguese.

inclusion and exclusion were applied. 58 articles were included and 55 were excluded. A complete re-reading of the publications taken in the final phase was conducted, i.e., 58 articles were read again considering the inclusion and exclusion criteria. Finally, remaining 24 included articles and 34 were excluded. Is important to emphasize that the reading held at the end of the third stage, sought to remove the items that did not mention the use of learning theories in virtual worlds.

In the fourth stage of this research, the remaining articles were analyzed and categorized, in order to answer the research questions and therefore the objective of this research. This step is detailed in the next section.

ANALYSIS OF RESULTS

Initially, a comparative analysis considered the 24 articles included and excluded 34 at the end of the third stage. This analysis seeks to highlight the databases that have studies about virtual worlds

Table 2. Data from results

EB	IR	Selection Phase		Extraction Phase		Re-Reading	
		I	E	I	E	I	E
Congreso Internacional de Informática Educativa or International Conference of Educational Computing	40	4	36	3	1	1	2
Workshop de Informática na Escola Computer Workshop in Schools	15	0	15	-	-	-	-
IEEE Transactions on Education	0	-	-	-	-	-	-
IEEE Transactions on Learning Technologies	9	3	6	1	2	1	0
Congresso da Sociedade Brasileira de Computação (Workshop sobre Educação em Computação) or Congress of the Brazilian Computer Society (Workshop on Computer Education)	5	3	2	3	0	0	3
Journal of Virtual Worlds Research	130	21	109	7	14	3	4
Revista Latinoamericana de Tecnología Educativa or Latin American Journal of Educational Technology	4	2	2	1	1	1	0
Simpósio Brasileiro de Informática na Educação or Brazilian Symposium of Information Technology on Education	26	7	19	6	1	2	4
Computers & Education	149	44	105	23	21	11	12
Revista Novas Tecnologias na Educação or Journal of New Technologies in Education	26	19	7	10	9	3	7
Creative Education	8	4	4	3	1	2	1
Informática na Educação: teoria & prática or Computers in Education: Theory & Practice	1	1	0	0	1	-	-
IEEE Computer Graphics and Applications	4	3	1	0	3	-	-
IEEE Revista Iberoamericana de Tecnologías del Aprendizaje or Iberoamerican Journal of Technology Apprenticeship	2	2	0	1	1	0	1
Total	419	113	306	58	55	24	34

Table Legend: EB – Electronic Base; IR – Initial Result; I – Included; E – Excluded.

applied in education that uses learning theories. Figure 1 presents a comparative distribution of the publications, which highlights the number of articles that used educational theories in their studies and the number of articles that do not cover educational theories.

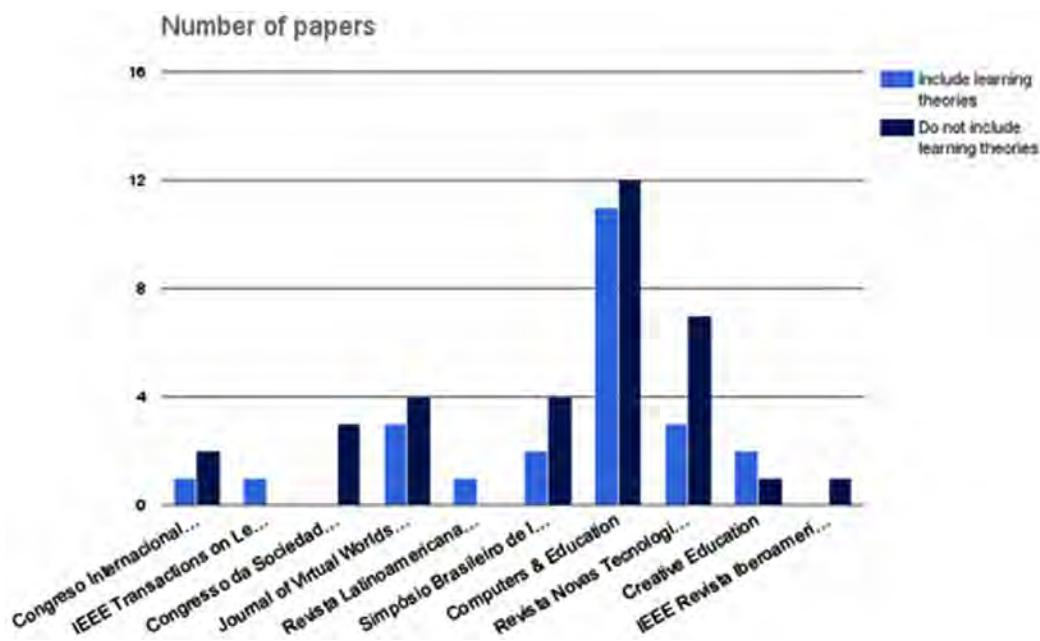
In Figure 1, you can see that most of the publications found are from Computers and Education, which represents 41% of articles that mention the use of learning theories. Articles from IEEE Transaction on Learning Technologies and Latin American Journal of Educational Technology differs from others surveyed, because it contains only articles that discuss the use of learning theories to explain the process of learning in virtual worlds. Analyzing the articles selected, was possible to trace some influences of the major figures of the

field, like theories of Papert's, Piaget's, Ausubel's, Vygotsky's and Kolb's.

One of the current challenges in the area of virtual worlds is related to authorship of users who do not have specific technical knowledge, to build and manipulate multimedia features in this type of environment. Researches related to learning of foreign languages and teaching of computing topics, such as algorithms and programming, can use features presented in virtual worlds to instigate the user to become more active and exercise his critical through by the manipulation and construction of knowledge on their own.

It is in this scope that some authors of the articles selected sought to anchor the constructionist theory proposed by Papert (1997), in which the user performs the construction of knowledge through

Figure 1. Data from articles included and excluded



concrete actions, resulting in an actual product and from your interest. Features such as creating objects, manipulating educational activities and exchange of information through the chat are some of the points present in these environments, which fits into this educational approach, allowing its use in several articles found.

Another great thinker related to construction of knowledge is Piaget (1973), in which his psychogenic theory has also been used as an educational base for the construction of activities in virtual worlds. The use of virtual worlds for teaching physics with simulations of phenomena that occur in the virtual world makes room for the application of this theory.

The dynamicity of simulations in virtual worlds, where its parameters can be modified to change its mode of operation and therefore the end result, forces users to exercise their knowledge of this topic to understand what happened during the process. Thus, based on the theory proposed by Piaget, the individual acts on objects in a physical or mental form, causing the occurrence of an imbalance of the knowledge that he

had acquired previously, which is established as cognitive conflict.

To resolve this conflict, the author describes that student must carry out the process of assimilation and accommodation of new knowledge, which affects the balance of the same, thus allowing to suffer another imbalance and thus restart the process. In the case of simulations in virtual worlds, it can encourage the user to perform new actions that were being analyzed based on the concepts that held, in order to cause an imbalance, to be organized again and there is an accommodation of knowledge.

Other items were also found that seeks to instigate users to learn content related to the areas of physics and chemistry, through the creation of virtual laboratories with simulations of various types of experiences. The possibility of creating virtual 3D animated objects, which reflect experiments that occur in the real world, allows complementary practical theory that is seen in the classroom.

Thus, authors have used as an educational basis of the theory of experiential learning of

Kolb (1984), which explains learning as the process by which knowledge is created through the transformation of experience, the knowledge results from the combination of understanding and transforming experience. In virtual worlds, it can be carried out this cyclical process in an organized and structured way in teaching activities done by students. The concepts are explored and viewed through 3D simulations, which can be further discuss issues involving these experiences.

At this point, it can also be driven together with this approach the learning processes of problem based on an experiment in field of algorithms and programming. Such an approach can come to cover the entire cyclic process present in experiential learning and also enter in problem solving to assist in student learning in this research topic, using an auxiliary software to create codes and view the results of the algorithm on the objects in the virtual world.

With regard to articles that addressed the use of serious games in virtual worlds, challenges combined with existing knowledge in cognitive user structure and those who should be acquired with the game opens the possibility for the use of meaningful learning proposed by Ausubel (1963). From the moment that the game can be planned so that the user has to have certain knowledge about a topic to meet the proposed challenges and advance of phase, it exercises the use of previous concepts to learn new topics that are presented on the challenges.

An example of serious game of an article analyzed seeks to help student's knowledge in computer networks, in which, after viewing a particular content in the classroom, students can play in the virtual world seeking to strengthen the process of acquiring a significant learning about a topic, as well as to use this knowledge to solve challenges on topics which does not have much knowledge.

In the case of Vygotsky's theory (1978), which discusses concepts of the zone of proximal development (ZPD), the author explains that there is a potential for cognitive development in guid-

ing the gap between the growth capacity of the individual point view of an individual, with the resolution of problems through the help of people or groups who have a more advanced knowledge of the subject. The issues proposed by this thinker can be embedded in the context of virtual worlds because of the possibility of communication and collaboration in the environment.

Features, such as the use of chat and the possibility of building objects and activities in joint authorship in virtual worlds fall on the principles defended by Vygotsky. The interaction between avatars with conversations using chat can be useful for the development of activities that are based on existing provisions in the ZPD, as well as its teacher or educational tutor can come to assist in solving problems during an interaction in the virtual world.

Reported several features present in virtual worlds and theories surrounding the use of these that have been proposed by the great thinkers of the educational area, it is important to also highlight an important concern related to cognitive load level of user's in virtual worlds. The theory of cognitive load proposed by Sweller (2003) explains that learning occurs in a better manner when the information's are aligned with human cognitive processes, that is, when the volume of information provided to the student is compatible with the ability of human understanding. Thus, it was found that some of the selected studies have examined in the context of virtual worlds certain aspects of the user perception, didactic instruction and comfort level that can be considered appropriate for the context of the user, so that there is not a cognitive overload, especially due to various possibilities of interaction and visualization present in virtual worlds.

Analysis in this section creates the possibility to draw a certain line of influence of some of the leading thinkers of education mentioned above with certain areas, educational topics and forms of exploitation of the resources present in virtual worlds. It is noteworthy that such links and analyzes are only based on selected articles in this

research, not seeking to generalize and draw labels that define the use of a particular educational theory in an educational area. Thus, it seeks to make clearer to readers of this chapter as certain situations and research problems are likely to fit into certain educational foundation for building experiments in virtual worlds.

Explained the main points, concepts and applications of virtual worlds in various educational theories identified during the analysis of the selected studies, it was possible to make a brief description about these issues and clarify to readers of this chapter some of the peculiarities present in each of the cases described. The following section presents some of the future directions contained within virtual worlds applied to education.

FUTURE RESEARCH DIRECTIONS

Virtual worlds have had their expansion in the last two decades, which in recent years has been carried out a significant number of experiences in various fields of education, such as chemistry, physics, engineering, computer science and many others. The analysis in this chapter has shown an important variety of options regarding the choice of an educational theoretical basis to support formulation and implementation of teaching strategies and teaching activities with students of different age groups.

Conducting simulations of experiments and construction of learning activities in virtual worlds tend to continue growing, to the extent that results obtained from research has been positive and instigators. The favorable scenario for the application of educational theories in virtual worlds opens up a broad field of possibilities in this scope, allowing increasingly exploit the use of computing resources present in these environments, to perform activities with students during the class period, at distance or as complementary to reinforce contents seen in classroom, where the application of flipped and blended learning can be explored within this context.

Throughout this chapter, the authors sought to meet the objectives previously established, as well as develop a theoretical collection about virtual worlds applied to education, in order to contribute to the researchers who are starting or looking for the best use of strategies of these virtual immersive environments in the educational sector. For this, an analysis in various contexts was performed by discussing the different educational approaches used in various areas of knowledge and education levels.

Based on evidenced findings, the authors qualify as future directions of research in this area, the need to establish a line of investigation seeking evidence that motivational elements and of engagement can be implemented for the involvement of users in these virtual worlds with resources for learning thereof, analyzing from primary strategies, e.g., use of videos, slides, images, sounds, objects and 3D animations, even applications involving the use of theories as flow, gamification, learning trajectories, among other strategies were also revealed in this systematic review.

CONCLUSION

This chapter presented a systematic review, in which, it was possible to identify different educational theories that have been used as a theoretical basis for the formulation and implementation of educational activities in virtual worlds. It was found that research has been applied in different areas of education, especially in the fields of chemistry, physics, engineering and computer science, which can be developed various types of simulations in 3D environment representing real-world phenomena.

An important point to note is the existence of a wide range of possibilities to be deployed in virtual worlds concerning the use of educational theories, in which, was established that characteristics of immersion, active learning, communication, collaboration and interaction, that are present in this environment, open a very wide range to be

worked in different educational concepts present in several theories.

This possibility has resulted in the development of a wide range of research in the past few years, in which different approaches and applications of methodologies informed by one or more educational theories have been the subject of research by authors. Results have been positive, instigators, and help to improve the learning process of students, mainly due to the integration of these theories to the elements present in virtual worlds.

For the final considerations of this chapter, the authors of this study believe that this type of approach can be considered valid and with a great potential, to be explored in different research that can be carried out. It is expected that readers have been able to improve their knowledge in the areas of virtual worlds applied to education, including its theoretical foundations, difficulties and applications. Thus, the authors envision that this chapter has provided a better understanding about the integration of these approaches, explaining the steps taken to carry out such processes and how to carry out this approach in new research that may arise from this chapter.

REFERENCES

- Abreu, F., Almeida, A., Barreiros, E., Saraival, J., Soares, S., Araújo, A. & Henrique, G. (2012). *Métodos, Técnicas e Ferramentas para o Desenvolvimento de Software Educacional: Um Mapeamento Sistemático*. Anais do 23º Simpósio Brasileiro de Informática na Educação.
- Ausubel, D. P. (1963). *The psychology of meaningful verbal learning*. New York: Grune & Stratton.
- Bloom, B. S. (1956). *Taxonomy of Educational Objectives, the classification of educational goals. Handbook 1: Cognitive Domain*. New York: McKay.
- Griol, D., Molina, J. M., & Callejas, Z. (2014). An approach to develop intelligent learning environments by means of immersive virtual worlds. *Journal of Ambient Intelligence and Smart Environments*, 6(2), 237–255.
- Herpich, F., Nunes, F. B., Voss, G. B., & Medina, R. D. (2016). Three-Dimensional Virtual Environment and NPC: A Perspective about Intelligent Agents Ubiquitous. In F. M. Neto, R. de Souza, & A. S. Gomes (Eds.), *Handbook of Research on 3-D Virtual Environments and Hypermedia for Ubiquitous Learning* (pp. 510–536). Hershey, PA: IGI Global. doi:10.4018/978-1-5225-0125-1.ch021
- Hulsey, C., Pence, T. B., & Hodges, L. F. (2014). Camp CyberGirls: Using a Virtual World to Introduce Computing Concepts to Middle School Girls. *Proceedings of the 45th ACM technical symposium on Computer science education*. doi:10.1145/2538862.2538881
- Jonassen, D. H., Peck, K. L., & Wilson, G. B. (1999). *Learning with Technology—A Constructivist Perspective*. Prentice Hall.
- Kitchenham, B., Charters, S., Budgen, D., Brerton, P., Turner, M., Linkman, S., & Visaggio, G. et al. (2007). *Guidelines for performing Systematic Literature Reviews in Software Engineering. EBSE Technical Report*. Keele University and University of Durham.
- Kolb, D. A. (1984). *Experiential Learning: Experience as the Source of Learning and Development*. Englewood Cliffs, NJ: Prentice-Hall.
- Mateu, J., Lasala, M. J., & Alamán, X. (2013). Tangible Interfaces and Virtual Worlds: A New Environment for Inclusive Education. *Lecture Notes in Computer Science*, 8276, 119–126. doi:10.1007/978-3-319-03176-7_16
- Moodle. (2016). *Philosophy*. Available at: <https://docs.moodle.org/24/en/Philosophy>

Nunes, F. B., Herpich, F., Voss, G. B., Medina, R. D., Lima, J. V., & Tarouco, L. M. R. (2014). *Laboratório Virtual de Química: uma ferramenta de estímulo à prática de exercícios baseada no Mundo Virtual OpenSim*. Anais do XXV Simpósio Brasileiro de Informática na Educação.

Nunes, F. B., Herpich, F., Zunguze, M. C. & Lima, J. V. (2016). *Utilização do Mastery Learning em sala de aula: uma abordagem integrada aos Mundos Virtuais*. Anais do 24º Workshop sobre Educação em Computação.

Papert, S. M. (1997). *A Família em Rede*. Lisboa: Relógio D'Água Editores.

Piaget, J. (1973). *Problemas de Psicologia Genética*. Editora Forense.

Quintana, M. G. B., & Fernández, S. M. (2015). A pedagogical model to develop teaching skills. The collaborative learning experience in the Immersive Virtual World TYMMI. *Computers in Human Behavior*, 51, 594–603. doi:10.1016/j.chb.2015.03.016

Spiro, R. J., & Jehng, J. (1990). *Cognitive Flexibility and Hypertext: Theory and Technology for the nonlinear and multidimensional Traversal of Complex Subject Matter*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Sweller, J. (2003). *Cognitive Load Theory: A Special Issue of educational Psychologist*. LEA, Inc.

Vygotsky's, L. S. (1978). *Mind in Society: the development of higher psychological processes*. Cambridge, UK: Academic Press.

KEY TERMS AND DEFINITIONS

Educational Theories: A formal thinking that searches to resolve and instruct, as to explain how students can understand and construct knowledge in education.

Virtual Learning Environment: Environment that seeks to provide support to teachers, managers and student in planning, construction and execution of educational activities, as for the insertion and make available educational materials.

Virtual Worlds: Are tools that simulate the real world in a three-dimensional environment, providing users the sense of immersion in a controlled environment with many opportunities and experiences.

Category W

Web Technologies

Anger and Internet in Japan

Hiroko Endo

Rissho University, Japan

Kei Fuji

University of Tsukuba, Japan

INTRODUCTION

The world has changed a lot due to the Internet. Japan is no exception and has become a society where people can freely share their opinions, thoughts, experiences, and personal emotions to a large number of people all over the world. It is reported that 94.6% of people in Japan use a cellphone or smart phone (Ministry of Internal Affairs and Communications, 2015), and the spread of mobile devices which allow easy access to the Internet no matter the time or location is fueling people's active sharing of opinions and expression of emotion.

However, this has given rise to a new problem unlike anything seen before. An unspecified large number of people's anger, complaints, criticisms, and offensive language not like anything seen in the real world is spreading on the Internet and continuing to cause uproar.

First off, this chapter will first focus on anger expression on the Internet in Japan by presenting a few examples. It will also highlight the results of large-scale studies relating to anger expression in Japan. Second, it will explain in terms of "Social Sharing" what sort of psychological effect sharing one's anger with others has, as well as go over an empirical study that demonstrated where social sharing of anger on the Internet has no positive effect. Third, it will show the results of studies done in Japan which described emotional experiences to look at how we should control our anger. It will also discuss what care should be taken when expressing emotions and experiences of anger on the Internet. Finally, it

will consider what main factors are needed to prevent the problem as above by re-examining the relationship between the Internet and anger from a new perspective.

BACKGROUND

Many readers out there likely took pleasure in watching the 2016 Rio de Janeiro Olympics. All of us were excited by the sight of top athletes gathering at the Olympic flame and showing their inflamed passion for sports. But that excitement can also easily lead to flaming on the Internet. Such a situation had already occurred during the 2000 Sydney Olympics. There was an incident where the athlete Shinichi Shinohara from Japan was judged to have been defeated in the final match for Men's Judo in the over 100Kg class. Shinohara had countered his opponent with an Uchi-mata-sukashi at one minute and 35 seconds from the start of the match and taken a victory pose, but judges did not award the Ippon victory that he had deserved and instead judged in favor for his opponent. Immediately afterward people took to the Internet in Japan and there was an outpour of rage against the chief judge for what was referred to as the "greatest misjudgment of the century". But the flaming did not stop there and spread to other places that were not involved in any way. Since the chief judge for the match was from New Zealand, people ended up targeting the Embassy of Japan in a fit of rage and caused the site to become inaccessible (Tashiro & Orita, 2012).

DOI: 10.4018/978-1-5225-2255-3.ch691

One person's opinion can be shared between others who share the same opinion or thoughts and lead to an increase of supporters, which can garner much attention. This phenomenon has been likened to a waterfall by Sunstein (2011) and referred to as a "cyber cascade". In Japan this is more likened to that of a small spark suddenly erupting and causing an inextinguishable fire and is generally referred to as "flaming". Aside from the difference between water and fire, either one is an extremely appropriate metaphor for anger that should have been constricted to the emotions of individual suddenly transforming into the anger of a group or whole society.

ANGER EXPRESSION ON THE INTERNET IN JAPAN

There are many other examples of flaming on the Internet that also cover a range of other topics. One dramatic example in particular was for the 2011 Great East Japan Earthquake. The unprecedented major earthquake and tsunami on March 11 as well as the damage incurred from the continuing aftershocks brought much shock and grief to the whole nation. The damage was not confined to natural disaster and also caused the Fukushima Daiichi nuclear disaster, forcing neighboring residents to evacuate to far away land. The grief and anxiety that people experienced in this earthquake was also clearly expressed on Twitter. The number of tweets on that day rose to 33 million which was an increase of 1.8 times that of a normal day, and 70% to 80% of all tweets within the country for the following week after March 11 were related to the earthquake (NEC BIGLOBE, 2011). Many of the tweets showed people's concerns for the earthquake and tsunami and fluctuated to a constant cycle that reached a peak at late hours of the night (Miura, Komori, Matsumura, & Maeda, 2015). In short, it is surmised that people affected by the disaster were sharing their concerns with one another through Twitter whenever the night came.

What also deserves attention is that fact that a characteristic fluctuating pattern was observed in relation to people's anger regarding this disaster. This pattern matched with the timing of news reports concerning the nuclear power plant accident (Miura et al., 2015). The number of angry tweets was increasing whenever the mass media exposed the worsening of the situation with the nuclear accident, the government's delayed response, and unclear reports of the situation from the power company. From these examples we see that even when, or indeed because of being in the midst of a large-scale disaster, people's anger reaches massive levels through the Internet and can have various effects on society.

In recent years, a large-scale web survey study was conducted based on the present condition of flaming on the Internet. The study, conducted by Tanaka & Yamaguchi (2016), surveyed 19,992 participants regarding their awareness of flaming on the Internet and their experience participating in such. Only 7.09% of respondents said they had not heard of flaming, while a majority said that while they had heard of it, they had never seen it (71.35%). We therefore find that the issue of flaming is being recognized as a serious problem in society even by those who have never seen it for themselves. Meanwhile, it was reported that only 1.52% of respondents said that they had either flamed once or more (Tanaka & Yamaguchi, 2016). When considering the bias of being a web survey, that number is assumed to not surpass 1.11%. In other words, contrary to the general image that may exist, only an extremely small amount of participants in flaming intimidate others online.

When we look at it this way, it would seem that flaming is not that serious of a problem. However, focus should be brought to a follow-up study done by Tanaka & Yamaguchi (2016) concerning the attributes and psychological factors behind the participants in flaming. Results showed that men were significantly more like than women to participate in flaming, the younger they were, and the more frequently they used social media (see Table 1). What is notable here is that it was also

found that those with higher personal or household income or those with children were more likely to participate in flaming. These results are understood to reflect the fact that, for example, families with children or families with a financial base for supporting childcare would have a strong interest in protecting their children and thus be more prone to transitioning to such radical behavior.

Another important point is that those who have had an unpleasant experience being spoken to in a rude manner on the Internet or those who have no problem criticizing others with harsh language on the Internet are more likely to participate in flaming (Tanaka & Yamaguchi, 2016). To put it simply, those who have been met with slander on the Internet do not refrain from doing the same based on that experience, but instead become an active participant. In short, we can say that there is a negative chain of slander that is occurring on the Internet. This negative chain is reinforced by the thought that slander on the Internet is permissible, and can be linked to the process of new aggressors being produced with the increase of other victims. In this sense, even though the number of people who participate in flaming on the Internet is small as previously mentioned, that number is likely to increase. 2011 was a turning point that saw a tremendous increase in the number of incidents of flaming on the Internet in Japan, and is likely to continue that way (Yamaguchi, 2015).

EFFECT OF SHARING ANGER ON THE INTERNET

So what sort of psychological effect or benefit is there for people by expressing anger on the Internet and sharing that anger with a large number of people? Does anyone calm their anger by flaming on the Internet the previous night and wake up the next morning feeling refreshed?

To consider this issue, we will reference the social sharing of emotion. Social sharing of emotion refers to the inter-personal action of someone who has experienced an emotion sharing their

Table 1 Result of logit model concerning participation to flaming

Variables (* Yes=1, No=0)	Marginal Effect	Odds Ratio	p
Personal Attributes			
sex (male=1, female=0)	0.0066	0.43	.05
age	-0.0190	-1.24	.00
parenthood*	0.0080	0.52	.01
income (individual)	0.0006	0.04	.08
income (household)	0.0033	0.22	.09
usage time of radio	0.0022	0.14	.00
usage time of social media	0.0019	0.12	.01
usage of general BBS*	0.0130	0.85	.00
usage of personal or commercial BBS*	0.0086	0.56	.01
Psychological Factor			
those who have had an unpleasant experience being spoken to in a rude manner on the Internet	0.0052	0.34	.06
those who have no problem criticizing others with harsh language on the Internet	0.0163	1.06	.00

Excerpt from Tanaka & Yamaguchi, 2016.

experience relating to that emotion with others (Rimé, Mequita, Philippot, & Boca, 1991). Rimé, Mequita, Philippot, & Boca (1992) conducted a detailed analysis regarding this phenomenon through a series of studies which indicated the following various opinions. First, a vast majority making up 90% of people who have had an emotional experience have done social sharing. Although this survey was carried out in Belgium, Southern France, Germany, Spain, and Italy, no large cultural differences were seen in the act of social sharing, which demonstrated how much of a cross-cultural phenomenon it is. No gender or age difference was found either, confirming that it was broadly occurring. In short, people cannot resist talking with others when having experienced some kind of emotion. Second, such social sharing of emotion did not differ between varying types of emotion and was found to have occurred the same way whether it was a positive or negative

emotion. In other words, people feel the need to share with others regardless of whether they are happy, sad, or anger. Third, approximately 60% of people shared their emotions with others within the day that they experienced an emotional episode. The extent of social sharing of emotion (the number of times sharing was done and the number of people shared with) was found to be more likely to occur the stronger the degree of emotion which occurred in that episode.

What is particularly interesting is results indicate that the social sharing of such emotions goes against what is generally believed and does not contribute to recovering from negative emotions. An *expeRimént* conducted by Zech & Rimé (2005) demonstrated this. Zech & Rimé (2005) conducted this *expeRimént* by randomly assigning groups to those who recalled negative episodes, which was divided between those who shared the reality of the incident with others, those who shared an emotional perspective, and those who did not share anything. They then checked to see what effect social sharing of emotion had for recovering from negative emotions by having participants visit the laboratory both one week following the *expeRimént* and once more two months later to ask them to respond to a question sheet that measured the state of their negative emotions in response to the incident that they recalled during the *expeRimént*. Results of the *expeRimént* did not demonstrate any significant difference in relation to scores for negative emotions between the three conditions. The only thing that had a significant influence on reducing one's negative emotions was time, and whether or not they shared their emotions or what the details of their experience were had no contribution. In other words, Zech & Rimé (2005) showed that whether or not one shared their emotions played no part in decreasing negative emotion.

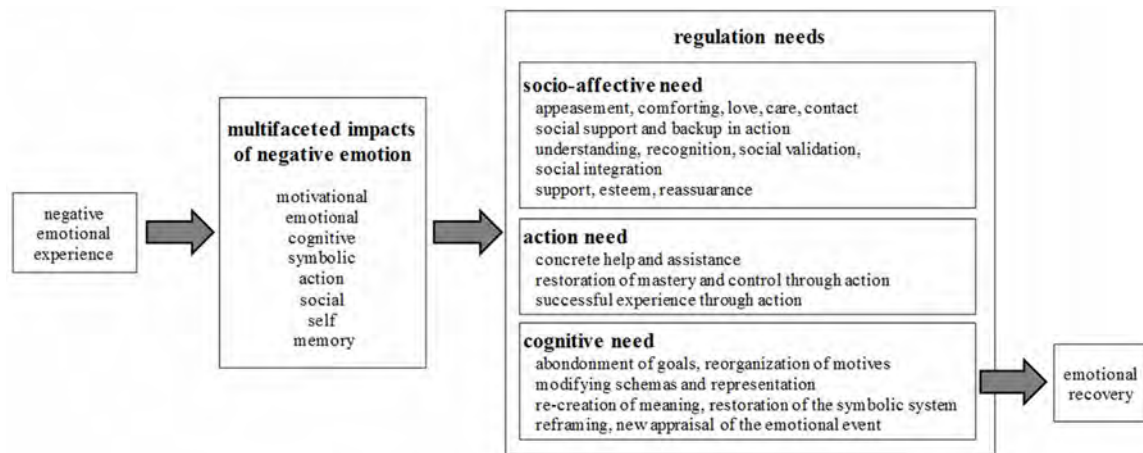
Other studies have exhibited the same opinion. Finkenauer & Rimé (1998) have conducted a study that inquires into emotional episodes that were either shared or not shared with others. By

comparing the two, they found that there was no significant difference in the intensity of an evoked emotion when recalling either. Rimé et al. (1991) and Stroebe, Stroebe, Schut, Zech, & van den Bout (2002) also reported that there is no relationship between social sharing of emotion and recovering from an emotional state.

So why doesn't social sharing contribute to recovering from negative emotions? Rimé (2007) has given a theoretical model for this (Figure 1). First, he purports that a negative emotional experience has a psychological effect from varying aspects, including motivational (frustration from not being able to achieve a goal), emotional (strong social desire and venting of emotion), cognitive (changing of schema and enforcement of unification), symbolic (enforcement of change of view in response to various things), action (suppression of action due to feeling helpless), social (desire for social approval and validation), self (reduced self-esteem), and memory (re-experiencing incidents). Either of these is considered to be unpleasant for someone who has had an emotional experience.

At the same time, Rimé (2007) contends that people also desire to regulate the various influences that come from these negative emotions. The regulation need is categorized between three needs, including 1) the desire for love, acceptance, social support, and approval (socio-affective needs), 2) the desire for concrete support, attention to one's surroundings, and regulation of one's own actions (action needs), and 3) the desire to reorganize objectives, revise schemas, re-evaluate incidents, and reconstruct meaning (cognitive needs). Reducing the negative emotions that occur due to an incident means that cognitive needs are being fulfilled. That is to say, unless cognitive needs are fulfilled, negative emotions won't decline and will continue to repeatedly enter one's conscious (see also Rimé, 2009). The importance of this cognitive reappraisal in controlling emotions has attracted attention in recent years in fields of study concerning the of emotions (e.g. Garnefski & Kraaij, 2007).

Figure 1. Theoretical model of regulation needs from negative emotional experience
Rimé, 2007.



In addition, Rimé (2007) conducted a series of studies which looked at what motivates people to share socially in order to see if these desires are being fulfilled in every day social settings. These studies looked into how subjects were motivated to share by having them recall emotional experiences which they had shared with others and indicate what motivated them to share their story. Through the study, Rimé (2007) asserts that first of all, social sharing allows people to obtain various reactions and support from the listener, which easily satisfies our socio-affective needs. Second, however, he argues that there are no motivations among these that satisfy our cognitive needs, and that even when we share out negative emotions socially in our general personal relationships, this sharing is not done based on our cognitive needs and therefore does not satisfy them. Social sharing of emotions thus is said to not contribute to reducing negative emotions.

Based on this theory by Rimé (2007), we consider that perhaps sharing our negative emotions with others, including anger on the Internet, does not reduce those emotions. Endo & Fuji (2015) focused on this and held a web survey to see the effect that social sharing of negative emotions on the Internet has. Their findings demonstrated that while negative emotions such as anger, anxiety, and depression that we evoke in our daily lives

push us to share them on the Internet, this social sharing does not reduce them, but instead maintains them. Fuji & Endo (2015) also considered what needs drive us to share negative emotions on the Internet and found that it is primarily our socio-affective needs and action needs, while our cognitive needs are ignored. Also, sharing of negative emotional episodes is more likely to be done immediately on the Internet than face-to-face (Endo & Fuji, 2015), as well as likely to be done without much cognitive processing or deliberation. Thus, social sharing on the Internet is hard to promote cognitive reappraisal in response to incidents and does not help reduce negative emotions. In addition, as previously mentioned, social sharing of emotion itself revitalizes emotions related to these incidents, which tells us all the more that negative emotions are difficult to reduce.

It is also understood that when negative emotions are shared, this also evokes largely negative emotion from the listener and encourages them to also share negative emotions with others (secondary social sharing; Christophe & Rimé, 1997). Specifically on the Internet, since it is easy to take something said by someone else and share it with others by either forwarding, quoting, or retweeting others comments, this secondary social sharing is easier to carry out.

Supporting this is a survey study conducted in China which targeted social networking services. Roughly 70 million posts made by the 200,000 users on Weibo, which is like China's version of Twitter, were collected and analyzed over a span of six months. Results showed that emotions of anger were more likely to be spread from user to user than emotions of happiness. Users connected with those that had submitted an angry posts also became angered and made their own angry submission (Fan, Zhao, Chen, & Xu, 2013).

To put it simply, an individual's anger is not reduced by being shared with others through the Internet, but instead ends up dragging others in and creates a large outpouring. Social sharing of emotion through the Internet does not reduce anger emotions for neither the individual who shared nor for those that it was shared with, and instead gives them an opportunity to spread anger.

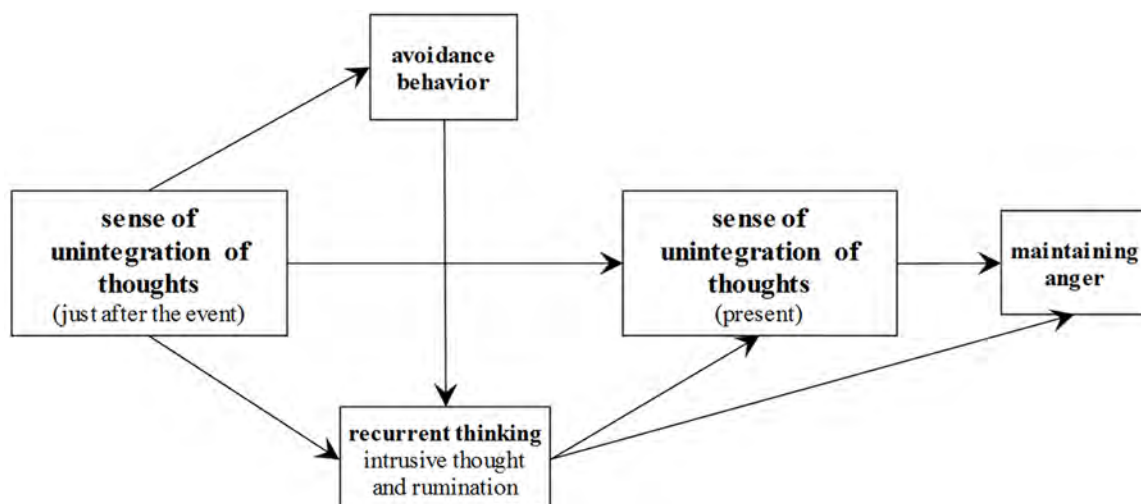
EFFECTIVE WAY TO REGULATE ANGER EMOTION

This also poses the following questions: what can we do to reduce our anger? What do we need to do in order to not hold onto this anger?

A vital lead in considering this issue is a study done by Endo & Yukawa (2011, 2012) concerning the process of maintaining anger. While anger is easily experienced in our daily lives, it is an emotion that is short term and vanishes quickly. However, Endo & Yukawa (2012) conducted a study and found that over half of the respondents still were maintaining that anger over two weeks later. Based on this, Endo & Yukawa (2012) analyzed the factors that cause this anger and found that it being maintained is due to rumination repetitive thoughts about the incident that caused their anger and intrusive thoughts unwelcome involuntary thought, image, or unpleasant idea relating to an incident. In addition, these thoughts were demonstrated to be a result of the feeling that no solution regarding an incident can be found, causing unintegration with one's intentions, or an incident still threatening someone, leaving them unable to accept it. In brief, as long as this sense of unintegration of thought remains, anger will continue to be maintained due to continued rumination and intrusive thinking (Figure 2).

What does this mean? Normally, when people are feeling angry or stressed, they will first share what they are feeling with others or avert their attention elsewhere and attempt to turn their mood

Figure 2. The process of maintaining anger
Endo & Yukawa, 2012.



around. But these actions do nothing more than attempt to deal with the emotion that is caused by an incident and does not reduce the sense of unintegration of thought which is at the foundation of the incident. The emotion of anger thus continues to be maintained. In fact, Endo & Yukawa (2012) report that through avoidance behavior where we attempt to forget these incidents, we instead end up holding onto our emotions of anger. They suggest that in order to be released from this anger, people must confront these incidents, re-evaluate their significance (Rimé, 2007) and come to resolve and accept them through their own understanding.

In support of this, Endo (2009) conducted a further experimental study of expressive writing of anger emotion. This experiment asked participants to recall and write down experiences which they still presently felt stressed about due to anger deriving from past experiences. One week later, these participants were randomly divided among the following four conditions to tackle the issue: 1) that they verbalize and continue to note everything including the details and facts of the incident and their feelings, 2) that they continue taking notes by focusing on the emotions that they felt from the incident, 3) that they continue making notes by only focusing on the details and facts of the incident, and 4) that they continue to take notes concerning details that have no relation with the incident. Participants in each group were instructed to focus on tackling their tasks for three days straight.

In a span of one week, two weeks, and one month following these tasks, measurements continued to be taken concerning to what extent the participants still felt anger in response to their incidents. Results of this study proved very interesting. Among the four conditions, participants under the condition that they confront the incident which made them angry and verbalize it without making room for any objective facts or subjective emotions were shown to have the most reduced anger. In addition, participants under this condition had the most significantly improved score of the extent to which they felt clear about their

incident in terms of their thoughts and emotions. Conversely, those in the control group which did not confront their incidents showed no such results, nor was their anger reduced. We therefore find that Endo's (2009) experiments show that confronting an incident and resolving it through finding your own meaning for it plays an essential role in reducing anger.

FUTURE DIRECTIONS

Based on these study results, let us re-examine the act of expressing anger and sharing it on the Internet. The majority of anger expressed on the Internet for the purpose of sharing it with others is written in extremely short and fragmented writing. At the very least, communication which describes both one's subjective emotions and the objective facts in detail and is mutually shared is a rare occurrence. It is plausible that this trend has become more evident due to the appearance of social networking services that allow real-time communication, such as Twitter and Facebook. Thus, we presume from study findings concerning the regulation of anger emotions that the expression and sharing of anger on the Internet is difficult to link to the reduction of and recovery from angry emotions.

In summarizing our arguments, it would be appropriate to consider that expressing and sharing anger on the Internet cannot be expected to reduce anger emotions. Further social sharing also occurs through the evoking of anger by the listener who has been shared these emotions. In this way, anger is shared which is then maintained without being reduced, and the process repeats where this causes further sharing, which leads to a large outpour of flaming on the Internet.

In order to resist this outpour, Rimé (2007, 2009) suggests that we must do a cognitive reappraisal of the incidents that cause these emotions from a new perspective, which will lead to our own resolution and settlement (Endo & Yukawa, 2012). Only then will anger be reduced. For this

purpose, it is essential to take the time to refer to the emotions which we have experienced and the facts and put them into words (Endo, 2009). However, while such communication does not generally occur due to such media tending to focus on real-time interaction, it is deeply rooted to the background of this problem. Flaming on the Internet is not only bound to happen, but is an inevitable phenomenon. This is something that we must bear in mind.

CONCLUSION

The society in which we live today allows us to use the Internet as a means of expressing and sharing out anger. To put it another way, it has given us a new means for controlling the various anger we experience in our daily lives. However, questions still remain concerning the effective use of this new method for our benefit. What we need in order to calm our emotions through our own submissions and comments and prevent a spark that ignites large-scale flaming on the Internet is perhaps a calmness like that of water, to not be influenced by the characteristics of the media, and instead confront our emotions and ourselves and put our thoughts into words without making room for emotions and facts.

REFERENCES

- Christophe, V., & Rimé, B. (1997). Exposure to the social sharing of emotion: Emotional impact, listener responses and secondary social sharing. *European Journal of Social Psychology*, 27(1), 37–54. doi:10.1002/(SICI)1099-0992(199701)27:1<37::AID-EJSP806>3.0.CO;2-1
- Endo, H. (2009). Effects of writing about anger experiences on mental health. *Japanese Journal of Research on Emotion*, 17(1), 3–11. doi:10.4092/jsre.17.3
- Endo, H., & Fuji, K. (2015). The adverse effect of social sharing of negative emotional events on a social networking service. *Paper presented at Conference of the International Society for Research on Emotions 2015*, Geneva.
- Endo, H., & Yukawa, S. (2011). Relationship between sense of unintegration of thoughts and maintaining anger among high school students. *Japanese Journal of Counseling Science*, 44, 92–100. doi:10.11544/cou.44.2_92
- Endo, H., & Yukawa, S. (2012). The mediating roles of cognition and behavior in the process of maintaining anger. *Japanese Journal of Psychology*, 82, 505–513. doi:10.4992/jjpsy.82.505 PMID:22514902
- Fan, R., Zhao, J., Chen, Y., & Xu, K. (2013). Anger is more influential than joy sentiment correlation in Weibo. *PLoS ONE*, 9(10), e110184. doi:10.1371/journal.pone.0110184 PMID:25333778
- Finkenauer, C., & Rimé, B. (1998). Socially shared emotional experiences vs. emotional experiences kept secret: Differential characteristics and consequences. *Journal of Social and Clinical Psychology*, 17(3), 295–318. doi:10.1521/jscp.1998.17.3.295
- Fuji, K., & Endo, H. (2015). An examination of regulation needs after negative emotional experiences. *Paper presented at Conference of the International Society for Research on Emotions 2015*, Geneva.
- Garnefski, N., & Kraaij, V. (2007). The Cognitive Emotion Regulation Questionnaire: Psychometric features and prospective relationships with depression and anxiety in adults. *European Journal of Psychological Assessment*, 23(3), 141–149. doi:10.1027/1015-5759.23.3.141
- Ministry of Internal Affairs and Communications. (2015). *2015 White paper information and communications in Japan*. Tokyo: Ministry of Internal Affairs and Communications.

- Miura, A., Komori, M., Matsumura, N., & Maeda, K. (2015). Expression of negative emotional responses to the 2011 Great East Japan Earthquake: Analysis of big data from social media. *Japanese Journal of Psychology*, 86, 102–111. doi:10.4992/jjpsy.86.13076 PMID:26182486
- NECBIGLOBE. (2011). Higashi nihon daishinsai ni okeru twitter no riyō joukyō ni suite [The survey report of the usage situation of twitter in the great east Japan earthquake]. Retrieved from <http://www.biglobe.co.jp/pressroom/release/2011/04/27-1>
- Rimé, B. (2007). Interpersonal emotion regulation. In J. J. Gross (Ed.), *Handbook of emotion regulation* (1st ed., pp. 466–485). New York: Guilford Press.
- Rimé, B. (2009). Emotion elicits the social sharing of emotion: Theory and empirical review. *Emotion Review*, 1(1), 60–85. doi:10.1177/1754073908097189
- Rimé, B., Mesquita, B., Philippot, P., & Boca, S. (1991). Beyond the emotional event: Six studies on the social sharing of emotion. *Cognition and Emotion*, 5(5-6), 435–465. doi:10.1080/02699939108411052
- Rimé, B., Philippot, P., Boca, S., & Mesquita, B. (1992). Long-lasting cognitive and social consequences of emotion: Social sharing and rumination. In W. Stroebe & M. Hewstone (Eds.), *European review of social psychology* (pp. 225–258). Chichester: John Wiley & Sons. doi:10.1080/14792779243000078
- Stroebe, M., Stroebe, W., Schut, H., Zech, E., & van den Bout, J. (2002). Does disclosure of emotion facilitate recovery from bereavement? Evidence from two prospective studies. *Journal of Consulting and Clinical Psychology*, 70(1), 169–178. doi:10.1037/0022-006X.70.1.169 PMID:11860043
- Sunstein, C. R. (2001). *Republic.com*. Princeton: Princeton University Press.
- Tanaka, T., & Yamaguchi, S. (2016). *Net-enjo no kenkyū* [The research of flaming]. Tokyo: Keiso Shobo.
- Tashiro, M., & Orita, A. (2012). Consideration concerning generation and setting process of flamingwar. *Information Processing Society of Japan SIG Notes*, 57, 1–6.
- Yamaguchi, S. (2015). Net-enjo no jittai to seisakuteki taiō no kōsatsu: Jisshō bunseki kara miru shakaiteki eikyō to meiyo kisonzai, seigenteiki honninkakunin seido, Internet literacy kyouiku no arikata [The reality of flaming and the discussion about political response: Consideration from the empirical data analysis of social impact, defamation, restrictive identification procedure, and Internet literacy education]. *Information and Communications Policy Review*, 11, 52–74.
- Zech, E., & Rimé, B. (2005). Is talking about an emotional experience helpful? Effects on emotional recovery and perceived benefits. *Clinical Psychology & Psychotherapy*, 12(4), 270–287. doi:10.1002/cpp.460

ADDITIONAL READING

- Boyd, D. (2014). *It's complicated: The social lives of networked teens*. New Haven: Yale University Press.
- Duprez, C., Christophe, V., Rimé, B., Congard, A., & Antoine, P. (2015). Motives for the social sharing of an emotional experience. *Journal of Social and Personal Relationships*, 32(6), 757–787. doi:10.1177/0265407514548393
- Gross, J. J. (2013). *Handbook of emotion regulation* (2nd ed.). New York: Guilford Press.
- Lepore, S. J., & Smyth, J. M. (2002). *Writing cure: How expressive writing promotes health and emotional well-being*. Washington, DC: American Psychological Association. doi:10.1037/10451-000

Sunstein, C. R. (2009). *Republic.com 2.0*. Princeton: Princeton University Press.

KEY TERMS AND DEFINITIONS

Cyber Cascade: A form of group-polarization on Internet.

Expressive Writing: Writing about past emotional experiences, including one's thoughts, feeling, and perceived facts.

Flaming: Large-scaled and offensive posting behaviors on the Internet.

Sense of Unintegration of Thoughts: The states when people could not integrate their thoughts by unacceptable events or unattained goals (see also Endo & Yukawa, 2012).

Social Sharing of Emotion: Communications about the emotional experiences in socially-shared language by the person who experienced them to other people.

W

Determine Democracy in Web Design

Rowena Li

Bayside High School Library, USA

INTRODUCTION

Surveys are the traditional and most widely used research instrument for measuring democracy. It is often used to measure the progress and decline of freedom and democracy in political rights and civil liberties experienced by individuals. As the Internet has become one of the most important vehicles of communication, and websites one of the most popular channels for information dissemination, a question has often been asked: in what way, if any, does a website, especially its homepage, carry its country's cultural traits and represent its nation's democracy level?

Evidently, web interface design reflects not only the linguistic aspects of a nation, but also its cultural characteristics, such as values, norms, and ethics. When we examine a country's cultural and social attributes represented on the web, one of the most important areas to consider is a country's democracy level, since power and authority create a special social structure for a society's culture.

Hofstede (1980) defined five primary cultural dimensions for measuring cultural differences. Power distance became the first dimension. Subsequently, Marcus (2005) and Marcus and Gould (2000) extended Hofstede's cultural theory to web interface design by identifying online indicators for the five cultural dimensions. Power distance received seven cultural indicators. These seven indicators, as well as three others (Gould, Zakaria, & Yusof, 2000; Singh, Zhao, & Hu, 2003, 2005), were statistically analyzed and validated in Li's (2009) study. Li concluded that special title, monumental building, authority figure, symbol of nationalism or religion, link to information about the leaders of the organization, information arranged according to management hierarchy, and

symmetric layout are valid indicators for measuring democracy on web interface design.

However, how exactly can web interface design be measured to detect a nation's democracy level with these seven indicators?

This article serves as an introduction to apply these seven indicators in examining democracy on web interface design. It introduces a new measuring instrument to assist in determining a nation's democracy level, so that democracy can be measured not only by traditional methods (surveys, case studies, questionnaires, interviews, and observations), but also through the study of web interface design. As a result, it extends cultural and political studies into the fields of human-computer interaction and user interface design.

BACKGROUND

Democracy and Its Measures

Over the years, the concept of democracy has been defined and redefined many times. For a long period of time, democracy has been associated with the demand of political and social equality (Laski, 1931). Some definitions for democracy place more emphasis on elections, examining voter participations and equal voting rights (Dahl, 1956; Lipset, 1963); others on the existence of political liberties (Lenski, 1966). Bollen (1980) defines democracy as "the extent to which the political power of the elite is minimized and that of the nonelite is maximized" (p.372). He argues that democracy should not be measured by voter participation, political stability, or multiparty political system, but by political rights and political liberties.

DOI: 10.4018/978-1-5225-2255-3.ch692

A growing number of studies concentrated on democracy measures and indices have been proposed and evaluated. First of all, whether democracy should be measured on a dichotomy approach (Lipset, 1959; Przeworski et al., 2000) or on a continuous scale (Bollen, 2009; Cutright, 1963) has been a major debate. Bollen (1990) believes democracy is continuous and should be evaluated in degrees. Although Bollen provided democracy indices for more than 100 countries, his studies only cover the years of 1960, 1965, and 1980 (Bollen, 1980, 1993). The Polity IV Democracy Scale, however, covers the years from 1800 to 2010 and “examines concomitant qualities of democratic and autocratic authority in governing institutions” (Marshall & Jaggers, 2012). It places a country’s democratic values on a 21-point scale. Freedom House Index of Political Freedom also places a country’s democracy on an ordinal scale. Freedom is measured by the progress and decline of freedom and democracy in political rights and civil liberties experienced by individuals. Each country is classified by the status of Free (Level 1.0 to 2.5), Partly Free (Level 3.0 to 5.0), or Not Free (Level 5.5 to 7.0) (Freedom House, 2015). Since its publication in 1972, this freedom rating remains as the standard in trans-national democracy evaluations (McClintock & Lebovic, 2006). Together with Polity scheme, it has become one of the two most widely used measures for democracy across countries (Foweraker & Krznaric, 2002).

Cultural Dimensions

In recent years, an increasing number of studies have focused on defining cultural dimensions. Hofstede’s (2001) five cultural dimensions have become the most quoted in cross-cultural studies and have been applied to a variety of research fields. After conducting two large surveys with 116,000 questionnaires, Hofstede concluded that four cultural dimensions (power distance, collectivism vs. individualism, masculinity vs. femininity, and uncertainty avoidance) can be used to measure cultural differences. The fifth cultural

dimension, long-term vs. short-term orientation, was added in 1991. In 2010, based on Michael Minkov’s analysis of the World Values Survey data for 93 countries, the sixth cultural dimension, indulgence vs. restraint, was included.

Hofstede’s cultural dimensions have been used to examine the impact of cultural differences of Asian-Americans on marketing strategies (Rallapalli & Montgomery, 2015), tax compliance (Putnam, Abdelfattah, Bagchi, & Braun, 2016), and market reception on capital structure (Arosa, Richie, & Schuhmann, 2015). The reliability of Hofstede’s cultural dimension was also validated in Bakir’s study (Bakir et al., 2015). Marcus’s study went further and made great contributions to cross-cultural study by applying Hofstede’s cultural dimension theory to web interface design (Marcus, 2005; Marcus & Gould, 2000). Marcus (2005) mapped Hofstede’s five cultural dimensions to user interface components and defined cultural indicators for each. Marcus and Gould (2000) also pointed out that power distance may influence several aspects of user-interface design, such as symmetric layout, information highly structured, hierarchies in mental model, nationalism or religion, focus on authority, official stamp, restricted security to access, and restricted managerial sections. At the same time, Gould, Zakaria, and Yusof (2000) examined three Malaysian and three US websites and concluded that prominent organizational charts, special title on members of the organization, and information arranged according to the management hierarchy are strong power distance indicators. Subsequently, Singh’s (Singh, Kumar, & Baack, 2005; Singh, Zhao, & Hu, 2003, 2005) studies also applied Hofstede’s cultural dimension theory to web content. They conducted considerable amount of scientific research to systematically validate Marcus and Gould’s framework in measuring cultural adaptation on the web. In their studies, six indicators were singled out as indicators for power distance.

Callahan (2007) examined cultural similarities and differences in terms of webpage organizations

and graphic designs. Her study found that cultural differences exist across countries and significant statistical correlations were found for the dimensions of power distance and individualism/collectivism between Hofstede's cultural dimension scores and the frequency counts of web indicators. Recently, while analyzing cultural elements in English-language and Chinese-language website designs, Chang (2011) found that cultural differences exist in power distance, uncertainty avoidance, individualism/collectivism, and long-term/short-term dimensions. Hsieh (2014) measured effects of cultural preferences on the use of web interface design and concluded that proper use of cultural elements on web interface design can promote the effectiveness of web communication.

Democracy and Web Interface Design

Built on previous studies, Li (2009) focused on statistical validation of national political freedom (democracy) indicators. Two coders performed content analysis on 156 college/university websites selected from 39 countries. The results indicated that six out of the ten proposed indicators could be used to measure a country's democracy on web interface design. They are, special title conferred on members of the organization, monumental building, authority figure, symbol of nationalism or religion, link to information about the leaders of the organization, and information arranged according to the management hierarchy. The seventh indicator, symmetric layout, demonstrated a negative correlation between the freedom level and the web representation of power distance.

Later on, Li (2010) used these seven indicators to determine if any differences exist between government-based websites and business-oriented websites in representing national political freedom. The result indicated that great differences exist in web interface design, which in turn reflects the national political freedom. The research concluded that web interface design correlates with a country's democracy level and government-based

websites embody more of a nation's authority and supremacy than business-oriented websites do.

Recently, Li (2014) used these seven democracy indicators in investigating the influence of geospatial factors on democracy and whether these associations are represented on web interface design. She concluded that democracy correlates positively with annual precipitation and latitude, but negatively with land boundaries and annual temperature. After examining 130 university/college websites collected from 65 countries, the study concluded that these four geospatial variables correlate with democracy represented on web interface design in the same manner, although these associations were not statistically significant. The study further suggested that the four geospatial factors be considered together as dependent variables to predict democracy more accurately.

MEASURING DEMOCRACY ON WEB INTERFACE DESIGN

As indicated in Li's (2009) study, democracy might be detected and measured not only in forms of surveys and questionnaires, but also through web interface design. Seven variables have been statistically verified in her study, which might be used as democracy indicators. They are:

1. Symmetric layout
2. Special title conferred on members of the organization
3. Monumental building
4. Authority figure
5. Symbol of nationalism or religion
6. Link to information about the leaders of the organization
7. Information arranged according to the management hierarchy

A content analysis method can be used to collect data from websites. The analysis is to carry out mainly on homepage with extensions to the

Figure 1. Homepage of University of Gadarif in Sudan at <http://www.gaduniv.edu.sd>



second and third level web pages. The seven power distance indicators are to be evaluated on a 0 (not present) to 4 (present to the most extent) scale with a coding sheet (Appendix). With some of the websites in languages other than English, a native speaker or an online translation service, such as Google Translate, should be consulted. If a website is presented in both English and its native language, the coding should be done based on its native language, since culture elements are best presented in their native languages.

In the course of selecting websites for sampling, it is necessary to consider the nature of the websites. College/university websites stand on a relatively neutral ground in providing information to their general audience in comparison with other types of websites, such as government-based or business-orientated websites. A government agency represents the power of its government and may reflect more of the authoritative aspects of an agency. On the other hand, the function of a

Table 1. Content analysis for University of Gadarif

Indicators	Coding
Symmetric layout (homepage): 75% symmetric	3
Monumental building: on the homepage as the focus	4
Symbol of nationalism/religion: on homepage and beyond	4
Link to information about the leaders: on every webpage	4
Information arranged according to the management hierarchy: On navigation bar, the Administration link is the first link or in the first group of links	4
Authority figure: One on the homepage	3
Special title: on none of the webpages	0

business website is to introduce its products and service to customers and may target specific constituencies, so that localization may play a part in web interface design. College/university websites, on the other hand, target the same constituencies: future students, current students, visitors, faculty and staff, and alumni and donors both nationally and internationally. Therefore, both globalization and localization have to be weighted in consideration of their web interface design. With further consideration, museum websites, post office websites, and news/media websites might share the same neutral stand and achieve the same research results as college/university websites. Following are examples of sampling and coding two college/university websites in measuring democracy.

University of Gadarif from Sudan at <http://www.gaduniv.edu.sd> (Figure 1), for example, scored comparatively high on its coding sheet (22 points, Table 1). It features a 75% symmetric homepage (scored 3) with a monumental building on its homepage as a focus (scored 4). The symmetric feature is evaluated by comparing the left half with the right half of the webpage and the percentile is calculated by setting up visual horizontal lines to the webpage in four quarters. Neither left menu bar nor right menu bar is considered within the parameter when calculating the symmetric layout. The second indicator,

monumental building, is coded according to the webpage level on which a monumental building is displayed (homepage, 2nd level webpage, and/or 3rd level webpage), and whether it is the focus of the webpage or is only a background in a picture. Symbol of nationalism includes any national flags and/or country maps appeared on the website. Religious remarks, pictures, paintings, and web links to external religious websites are considered as symbols of religion. The basic coding criterion for this indicator is focused on the website level on which the symbol appears. For this website, a symbol of religion (“In the name of the God the Merciful”) is found on the homepage and on Management Service page via University Administration link (scored 4). Information on organizational charts, titles and ranks of the administrators is another indicator for democracy. It aims to find out the emphasis the web designers (or rather the administrators) place on the leaders (or administrators themselves) according to the location and the transparency of the information. It is believed that the navigation theme is achieved through the proper arrangement of the links on and between webpages. The information placed on the homepage is usually the most important of all. This is examined horizontally (the arrangement of the administration information on the navigation bar) and vertically (the website level is it placed on). The indicator, information arranged according to management hierarchy, is coded to examine the link arrangement on the navigation bar of the homepage. Navigation bars, in most cases, are place on the left/right side, at the top and/or at the bottom of a webpage. Sometimes, two navigation bars appear on the same page. Since the vertical navigation bar bear the same weight in importance as the horizontal navigation bar, the two bars weight the same in the coding theme. Generally, links on a navigation bar are arranged according to their importance from left to right or from top to bottom, except websites in languages written from right to left, such as Arabic. We might find out how important the web designers (or the administrators) view information on leaders

by examining the location of the Administration link on the navigation bar on the homepage. If the Administration link is not presented directly on the homepage, any link providing information on administrators or any link leading to administration is considered as the link to be examined for this purpose. Generally, Home link points back to the homepage and is placed as the first on the navigation bar. It is simply a navigation tool instead of an authoritative indicator. Therefore, it is not counted as one of the links displayed before the Administration link. On this website, the first link from the left - University Administration is the link to information about the leaders (scored 4) and is displayed on every page (scored 4). The indicator, authority figures, should be evaluated by the frequency of its appearance and the webpage level on which it is displayed. Authority figures include administrators and any prominent visitors. These authority figures can usually be found on the homepage, in News and/or Administration sections. They can also be found in Introduction section, under History link, which introduces former administrators of the organization. For University of Gadarif website, one authority figure is found on the homepage and again on a 3rd level page (scored 3). A special title placed in front of a person’s name usually shows a certain degree of respect and represents the importance of his/her

Table 2. Content analysis for University of Melbourne

Indicators	Coding
Symmetric layout (homepage): 75% symmetric	3
Monumental building: on none of the webpage	0
Symbol of nationalism/religion: on none of the webpage	0
Link to information about the leaders: on the 2nd level webpage and beyond	1
Information arranged according to the management hierarchy: On navigation bar, the Administration link is the 4th link or in the fourth group of links	1
Authority figure: on none of the webpage	0
Special title: on none of the webpage	0

Figure 2. Homepage of The University of Melbourne in Australia at <http://www.unimelb.edu.au>

social status. It can be of religion, political party, or social class, including but not limited to titles such as Honorable, H.E., His/Her Majesty, His/Her Excellency, His/Her Highness, Rev., Leader, etc. However, academic titles, such as Dr., President, and Dean, are not considered in this category. No special title appears on this homepage (scored 0).

On the other hand, University of Melbourne in Australia at <http://www.unimelb.edu.au> (Figure 2) scored only 5 in its web content analysis (Table 2). It has 75% symmetrical layout from left to right (scored 3). Only the top banner part is not quite symmetrical. Neither monumental buildings nor religious remarks/national flags are found on this website (scored 0 respectively). Link to information about leaders can be found via Contact Us link, through the links of About the University, Governance, and About Academic Board (scored 1 for the 2nd level webpage and beyond category). For information arranged according to the management hierarchy, Contact Us link is the 7th link on the homepage (scored 1). Neither authority figure nor special title is found on University of Melbourne website (scored 0 respectively).

The examples above suggest that the website of University of Melbourne in Australia shows more democracy than that of University of Gadarif in Sudan does. This result is consistent with the freedom ratings provided by Freedom House for 2012 at <http://bit.ly/zA5MaG>, which indicates that Australia is in the free-country group with its freedom level as 1 (1 is the most free and 7 the least free), while Sudan is in the not-free-country group with its freedom level as 7. This shows that the democracy levels detected from these two websites correlate with the freedom levels of their corresponding countries generated from surveys and questionnaires conducted by Freedom House. However, it is still too early to conclude that democracy measure for each website reflects its country's democracy level. Li indicated in her studies (2009, 2010) that the medians of the total scores for the three freedom groups did not spread out as they should and most of their inter quartile ranges overlapped with each other, although the means of the total scores for the three freedom group were statistically significant. This indicates some of the websites in the lower

quartile of free-country group scored very close to some of the websites in the upper quartile of partly-free-country group. As a result, it would be a challenge to draw a scientific conclusion of a country's democracy level based on the democracy measures of only one website. Therefore, it is necessary to increase the sample size to at least 10% of the population in order to get an accurate democracy level assessment for a country.

FUTURE RESEARCH DIRECTIONS

With technology transformation, the diffusion of innovations, and globalization and localization of the web, measuring democracy on web interface design will become an alternative assessment instrument in detecting democracy along with surveys, questionnaires, case studies, and interviews. As we have introduced the method and identified the indicators of measuring democracy on web interface design, web designer will be more aware of democracy issues in designing culturally adapted websites for international users.

Future study may focus on the evaluation of different website sectors in measuring democracy. These website sectors may include but not limited to college/university websites, news/media websites, entertainment websites, sports websites, postal office websites, museum websites, and so on. In this way, a new framework for measuring democracy on web interface design will be established.

CONCLUSION

This article has demonstrated a new approach to assess democracy on web interface design. It has introduced seven democracy indicators, which can be used to measure a country's democracy level through web interface design. Examples are given with illustrations to demonstrate the higher and lower ends. Detailed explanations of the essential elements to look for in the examination process are accompanied with each indicator. This approach

has been used in two recent studies to evaluate democracy on web interface design. Li (2010) used this content analysis approach and coded 312 government and business websites. The results then underwent two-way analysis of variance to examine if any statistical significance exists among the three freedom groups defined by Freedom House Index of Political Freedom and also between the two different types of websites. With this measuring approach, Li's study concludes that web interface design correlates with a country's political freedom level and government-based websites embody more of a nation's authority and supremacy than business-oriented websites do.

This measuring approach was once again used in the most recent study. Li (2014) collected data from 130 college/university websites selected from 65 countries using this coding sheet and conducted both linear and multiple regression analyses. The study concludes that democracy correlates positively with annual precipitation and latitude, but negatively with land boundaries and annual average temperature.

Since local websites embed more cultural values and are culturally sensitive, this new approach provides a better understanding for those who study the essence of cross-cultural interface design. It also helps researchers gain insights into the cultural characteristics of a nation. Furthermore, it has increased the awareness of cultural elements represented on web interface design in a way that general web users understand not only their own cultures, but other cultures as well. In addition, these democracy indicators will be an asset in investigating the impact of localization on web interface design by comparing an organization's local website targeting national audience with its mirror website designed for its international audience.

In general, measuring democracy on web interface design opens up a new perspective in the study of democracy. Sociologists, anthropologists, and information scientists will find this approach helpful in assessing democracy and promoting understanding of different cultures.

REFERENCES

- Arosa, C. M. V., Richie, N., & Schuhmann, P. W. (2015). The impact of culture on market timing in capital structure choices. *Research in International Business and Finance*, 35(C), 180–196. doi:10.1016/j.ribaf.2014.05.005
- Bakir, A., Blodgett, J. G., Vitell, S. J., & Rose, G. M. (2015, January). A preliminary investigation of the reliability and validity of Hofstede's cross cultural dimensions. *Proceedings of the 2000 Academy of Marketing Science (AMS) Annual Conference* (pp. 226-232). Springer International Publishing. doi:10.1007/978-3-319-11885-7_55
- Bollen, K. A. (1980). Issues in the comparative measurement of political democracy. *American Sociological Review*, 45(3), 370–390. doi:10.2307/2095172
- Bollen, K. A. (1990). Political democracy: Conceptual and measurement traps. *Studies in Comparative International Development*, 25(1), 7–24. doi:10.1007/BF02716903
- Bollen, K. A. (1993). Liberal democracy: Validity and method factors in cross-national measures. *American Journal of Political Science*, 37(4), 1207–1230. doi:10.2307/2111550
- Bollen, K. A. (2009). Liberal democracy series I, 1972-1988: Definition, measurement and trajectories. *Electoral Studies*, 28(3), 368–374. doi:10.1016/j.electstud.2009.05.009
- Callahan, E. (2007). Cultural differences in the design of human-computer interfaces: A multi-national study of university websites [Doctoral dissertation]. Indiana University, Indianapolis, IN. Retrieved from ProQuest Dissertations and Theses database.
- Chang, H.-J. (2011). Multinationals on the web: Cultural similarities and differences in English-language and Chinese language website designs. *Journal of the American Society for Information Science and Technology*, 62(6), 1105–1117. doi:10.1002/asi.21515
- Cutright, P. (1963). National political development. *American Sociological Review*, 28(2), 253–264. doi:10.2307/2090612
- Dahl, R. (1956). *A preface to democratic theory*. IL. Chicago: University of Chicago Press.
- Foweraker, J., & Krznaric, R. (2002). The uneven performance of third wave democracies: Electoral politics and the imperfect rule of law in Latin America. *Latin American Politics and Society*, 44(3), 29–60. doi:10.2307/3177046
- Freedom House. (2015). Methodology. Retrieved from <https://freedomhouse.org/report/freedom-world-2015/methodology>
- Gould, E. W., Zakaria, N., & Yusof, S. A. (2000). Applying culture to website design: A comparison of Malaysian and US websites. Proceedings of 2000 joint IEEE international and 18th annual conference on computer documentation symposium conducted at Professional Communication Conference, Cambridge, MA.
- Hofstede, G. (1980). *Culture's consequences: International differences in work-related values*. Beverly Hills, CA: Sage Publications.
- Hofstede, G. (2001). *Culture's Consequences: Comparing values, behaviors, institutions, and organizations across nations* (2nd ed.). Thousand Oaks, CA: Sage Publications.
- Hsieh, H. (2014). Measuring the effects of cultures on Website Use. In S. Schatz, J. Cohn, & D. Nicholson (Eds.), *Proceedings of the 5th AHFE Conference: Advances in Cross-Cultural Decision Making* (pp. 52-60).
- Hsieh, H. (2014). Measuring the effects of cultures on Website Use. In S. Schatz, J. Cohn, & D. Nicholson (Eds.), *Proceedings of the 5th AHFE Conference: Advances in Cross-Cultural Decision Making* (pp. 52-60).

- Laski, H. J. (1931). Democracy. In E. R. A. Seligman & A. Johnson (Eds.), *Encyclopedia of the Social Sciences* (Vol. 5, pp. 76–85). London: Macmillan.
- Lenski, G. (1966). *Power and privilege*. New York: McGraw-Hill.
- Li, R. (2009). The representation of national political freedom on web interface design: The indicators. *Journal of the American Society for Information Science and Technology*, 60(6), 1222–1248. doi:10.1002/asi.21046
- Li, R. (2010). The representation of national political freedom on web interface design: A comparison of government-based and business-oriented websites. In J.-R. Park & E. Abels (Eds.), *Interpersonal relations and social patterns in communication technologies: Discourse norms, language structures and cultural variables* (pp. 368–404). Hershey, PA, USA: IGI Global; doi:10.4018/978-1-61520-827-2.ch020
- Li, R. (2014). The influence of geospatial factors on democracy: Its representation on web interface design. *Journal of the American Society for Information Science and Technology*, 65(2), 313–333. doi:10.1002/asi.22964
- Lipset, S. (1963). *Political man: The social bases of politics*. Garden City, NY: Doubleday & Company, Inc.
- Lipset, S. M. (1959). Some social requisites of democracy: Economic development and political legitimacy. *The American Political Science Review*, 53(1), 69–105. doi:10.2307/1951731
- Marcus, A. (2005). User interface design and culture. In N. Aykin (Ed.), *Usability and internationalization of information technology* (pp. 51–78). NJ: Lawrence Erlbaum Associates.
- Marcus, A., & Gould, E. W. (2000). Crosscurrents: Cultural dimensions and global web user-interface design. *Interaction*, 7(4), 32–46. doi:10.1145/345190.345238
- Marshall, M. G., & Jaggers, K. (2012). Polity IV Project: Political regime characteristics and transitions, 1800–2010. Retrieved from <http://www.systemicpeace.org/polity/polity4.htm>
- McClintock, C., & Lebovic, J. H. (2006). Correlates of levels of democracy in Latin America during the 1990s. *Latin American Politics and Society*, 48(2), 29–59. doi:10.1353/lap.2006.0021
- Przeworski, A., Alvarez, M., Sheibub, J. A., & Limongi, F. (2000). *Democracy and development: Political institutions and well-being in the world*. Cambridge: Cambridge University Press; doi:10.1017/CBO9780511804946
- Putnam, K., Abdelfattah, B., Bagchi, K., & Braun, G. P. (2016). Tax Compliance and International Cultural Factors: A Multilevel Empirical Study. In *Smart Technologies for Smart Nations* (pp. 19–30). Singapore: Springer. doi:10.1007/978-981-287-585-3_2
- Rallapalli, K. C., & Montgomery, C. D. (2015). Marketing strategies for Asian-Americans: guidelines based on Hofstede's Cultural Dimensions. In *Minority Marketing: Research Perspectives for the 1990s* (pp. 73–77). Springer International Publishing. doi:10.1007/978-3-319-17386-3_15
- Singh, N., Kumar, V., & Baack, D. (2005). Adaptation of cultural content: Evidence from B2C e-commerce firms. *European Journal of Marketing*, 39(1/2), 71–86. doi:10.1108/03090560510572025
- Singh, N., Zhao, H., & Hu, X. (2003). Cultural adaptation on the web: A study of American companies domestic and Chinese websites. *Journal of Global Information Management*, 11(3), 63–80. doi:10.4018/jgim.2003070104
- Singh, N., Zhao, H., & Hu, X. (2005). Analyzing the cultural content of web sites: A cross-national comparison of China, India, Japan, and US. *International Marketing Review*, 22(2), 129–146. doi:10.1108/02651330510593241

ADDITIONAL READING

Bollen, K. A. (1986). Political rights and political liberties in nations: An evaluation of human rights measures. *Human Rights Quarterly*, 8(4), 567–591. doi:10.2307/762193

Burgmann, I., Kitchen, P. J., & Williams, R. (2006). Does culture matter on the web? *Marketing Intelligence & Planning*, 24(1), 62–76. doi:10.1108/02634500610641561

Cyr, D., & Trevor-Smith, H. (2004). Localization of web design: An empirical comparison of German, Japanese, and United States web site characteristics. *Journal of the American Society for Information Science and Technology*, 55(13), 1199–1208. doi:10.1002/asi.20075

Hall, E. T., & Hall, M. R. (1989). *Understanding cultural differences: Germans, French and Americans*. Yarmouth, ME: Intercultural Press.

Hall, S. (2003). The work of representation. In S. Hall (Ed.), *Representation: Cultural representations and signifying practices* (pp. 13–74). Thousand Oaks, CA: Sage Publications.

Hofstede, G., & Hofstede, G. J. (2005). *Cultures and organizations: Software of the mind*. New York: McGraw-Hill.

Hofstede, G., Hofstede, G. J., & Minkov, M. (2010). *Cultures and organizations: Software of the mind* (3rd ed.). New York: McGraw-Hill.

Hoft, N. (1996). Developing a cultural model. In E. M. del Galdo & J. Nielsen (Eds.), *International user interfaces* (pp. 41–73). NY: John Wiley & Sons.

Nath, R., & Murthy, N. V. (2004). A study of the relationship between Internet diffusion and culture. *Journal of International Technology and Information Management*, 13(2), 123–132. Retrieved from <http://scholarworks.lib.csusb.edu/jiim/vol13/iss2/5/>

Robbins, S. S., & Stylianou, A. C. (2003). Global corporate web sites: An empirical investigation of content and design. *Information & Management*, 40(3), 205–212. doi:10.1016/S0378-7206(02)00002-2

Trompenaars, F., & Hampden-Turner, C. (1998). *Riding the waves of culture: Understanding diversity in global business* (2nd ed.). New York: McGraw-Hill.

Zahed, F., Van Pelt, W. V., & Song, J. (2001). A conceptual framework for international web design. *IEEE Transactions on Professional Communication*, 44(2), 83–103. doi:10.1109/47.925509

Zakour, A. B. (2004). Cultural differences and information technology acceptance. Proceedings of the 7th Conference of the Southern Association for Information Systems. Retrieved from <http://sais.aisnet.org/2004/Zakour.pdf>

Zhao, W., Massey, B. L., Murphy, J., & Fang, L. (2003). Cultural dimensions of website design and content. *Prometheus*, 21(1), 75–84. doi:10.1080/0810902032000051027

KEY TERMS AND DEFINITIONS

Collectivism vs. Individualism: Measures the relationship among individuals. In an individualistic culture, individuals value more on individualism; while in a collectivist culture, individuals view more on community and society.

Culture: A set of values, norms, and ethics shared by a group of people. It distinguishes one group from another.

Indulgence vs. Restraint: Measures the gratification for enjoyment and fun. Indulgence indicates the degree at which a society allows for relatively free gratification of basic and natural human drives related to enjoying life and having fun, while restraint stands for the degree of a society that suppresses gratification of needs and regulates it by means of strict social norms.

Long-Term vs. Short-Term Orientation:

Measures the tendency of perseverance or quick results. Societies with a long-term orientation show a strong tendency to invest and thriftiness to achieve long-term results, while societies with a short-term orientation have great respect for traditions and tend to achieve quick results.

Masculinity vs. Femininity: Measures the degree to which a culture views its gender roles.

Masculine cultures: tend to view males as tough conquerors and females as tender homemakers, while feminine cultures tend to reduce these gender differences.

Power Distance: Measures the degree of equality or inequality between people in a society.

Uncertainty Avoidance: Measures the extent people feel about uncertain or ambiguous situations.

APPENDIX: CODING SHEET

W

Please examine each webpage carefully and circle the correspondent “yes”.

Table 3.

Code	Present -----> Not Present				
	4	3	2	1	0
Symmetric Layout (Homepage Only)					
The webpage is 100% symmetric	yes				
The webpage is 75% symmetric		yes			
The webpage is 50% symmetric			yes		
The webpage is 25% symmetric				yes	
The webpage is not symmetric					yes
Monumental Building					
On the homepage as the focus?	yes				
in the background?		yes			
On a 2nd level webpage as the focus?		yes			
in the background?			yes		
On the 3rd level webpage/pages or beyond				yes	
On none of the webpages					yes
Symbol of Nationalism or Religion					
on homepage and beyond	yes				
on homepage only		yes			
on the 2nd level page/pages			yes		
on the 3rd level webpage/pages or beyond				yes	
on none of the webpages					yes
Link to Information about the Leaders					
on every webpage	yes				
on homepage and 2nd level webpages		yes			
on homepage only			yes		
on the 2nd level webpage/pages and beyond				yes	
no such link is found					yes
Information Arranged According to the Management Hierarchy. Home Link Is Not Counted.					
On navigation bar, is the Administration link					
the first link or in the first group of links?	yes				
the 2nd link or in the second group of links?		yes			
the 3rd link or in the third group of links?			yes		
the 4th link or in the fourth group of links?				yes	
no such link found					yes
Authority figure					
More than one on the homepage	yes				

Determine Democracy in Web Design

Code	Present -----> Not Present				
	4	3	2	1	0
One on the homepage		yes			
More than one on the 2nd level webpage/pages		yes			
One on a 2nd level webpage			yes		
On the 3rd level webpage(s) and/or beyond				yes	
on none of the webpages					yes
Special Title Conferred on Members (Religious or Political, like Rev., Majesty, Honorable)					
More than one on the homepage	yes				
One on the homepage		yes			
More than one on the 2nd level webpage(s)		yes			
One on the 2nd level webpage(s)			yes		
More than one on the 3rd level webpage(s)			yes		
One on the 3rd level webpage and/or beyond				yes	
on none of the webpages					yes

Discussion Processes in Online Forums



Gaowei Chen

The University of Hong Kong, Hong Kong

Ming M Chiu

The Education University of Hong Kong, Hong Kong

INTRODUCTION

Using online discussions to facilitate learning is a major issue in information science and technology, especially with the increasing number of massively open, online courses (MOOCs). Online discussions involve a group of participants exchanging ideas by posting messages on an electronic medium (e.g., discussion forum, knowledge building environment). Due to its information transparency, communication flexibility and opportunities for reflection, online discussions in both independent forums and forums linked to school courses offer students additional opportunities for information processing, higher order thinking and learning (Chen, Chiu, & Wang, 2012a, 2012b; Gillani & Eynon, 2014; Qiu & McDougall, 2013).

However, an online discussion forum does not necessarily guarantee engagement, effective interactions or substantial learning (Hew & Cheung, 2014). For instance, despite the widespread use of MOOC forums, often only a small proportion of the students are active participants (Onah, Sinclair, & Boyatt, 2014).

This article discusses the advantages and disadvantages that online discussions offer compared to face-to-face discussions. Specifically, individual characteristics and message attributes can influence participants' thinking and social relationships (Chen et al., 2012a, 2012b). By understanding the discussion processes through which students create new ideas and develop social relationships in online forums, designers can improve online forum interfaces. Likewise, educators can capitalize on this information to

help students participate, cooperate and learn in online forums more effectively.

BACKGROUND

While online discussions have several advantages over face-to-face discussions, they also have some drawbacks. Online discussions' advantages include information transparency, communication flexibility and reflection opportunities. As online messages are explicit, relatively permanent and organized, they are more transparent than face-to-face talk. Online messages are written explicitly and stored, so group members and teachers/facilitators can access them later. Furthermore, authors can organize online discussion messages to highlight their relationships to other messages by responding along a specific thread or via quotes of previous messages (Chiu & Chen, 2013). The interface designs of some online discussion forums constrain each message to respond to a single previous message, which helps establish clear connections and avoid ambiguous relationships among messages. Readers who heed these explicit relationships can read the related messages in the authors' preferred sequence, which can facilitate their understanding of the messages' content.

As a result of their greater permanence, online discussions offer greater communication flexibility across time and space compared to face-to-face discussions. Face-to-face discussants must be in the same place at the same time to engage in a shared conversation. In contrast, synchronous online discussants can communicate with one an-

other from any location. In asynchronous forums, participants can review the relevant information or post messages at any time from any location.

Moreover, the greater permanence of online discussions also allows participants to take more time to reflect before responding, in comparison to face-to-face discussions, especially during asynchronous discussions (Hew, Cheung, & Ng, 2010). During face-to-face discussions, people respond in real time to one another, so they are less likely to spend much time editing their responses. In contrast, posting asynchronous, online discussion messages on a permanent online forum provides convenient access to participants, so they can spend minutes, hours, even days gathering more information from other sources, contemplating their relationships, and evaluating competing claims and justifications before writing a suitable response.

Online discussions also have some disadvantages compared to face-to-face discussions. For example, face-to-face discussion participants can use nonverbal facial expressions and social cues to clarify and reinforce their meaning. In contrast, online discussion participants cannot use them, which can lead to misunderstandings among participants (Walther, Loh, & Granka, 2005). Also, while multi-threaded discussions allow greater time flexibility, their demands are also less immediate. As these students do not need to respond immediately, they are more likely to ignore the messages and not respond at all (Hewitt, 2005; Thomas, 2002). Instead, they may initiate off-topic discussions (Wu & Hou, 2015). As participants can respond later, the group often requires more time to make group decisions, which can reduce their efficiency (Baltes, Dickson, Sherman, Bauer, & LaGanke, 2002).

Online discussion forums can be independent or linked to school courses. An independent academic forum is a bulletin board on a specific subject (e.g., high school geometry) but not related to any class or school. In such forums, peers communicate with one another as they wish, without instructor moderation or inference (Chen, 2004;

Chen & Chiu, 2008). In a course-related forum (e.g., a MOOC) however, an instructor may structure, scaffold, or moderate the discussions (Coll, Rochera, & de Gispert, 2014; Park et al., 2015).

ONLINE DISCUSSION PROCESSES

Like face-to-face discussions, online discussions include both problem content and social relations (Chiu, 2008). This section explicates the processes by which online discussants create correct, new ideas (micro-creativity) and develop social relationships. First, a theoretical framework characterizes online discussions at the message level, including a message's content and author. Then, it shows how specific message attributes and individual characteristics influence participants' micro-creativity and use of social cues during online discussions. By understanding students' micro-creativity and uses of social cues during online discussions, educators can help students engage in beneficial discussion processes that improve cognitive outcomes and positive social relationships.

Characterizing Online Discussion Messages

An online discussion message can be characterized along four dimensions: knowledge content, social metacognition, social cues, and individual characteristics (See Table 1; Chen & Chiu, 2008; Chiu & Chen, 2013; Hara, Bonk, & Angeli, 2000; Wong, Pursel, Divinsky, & Jansen, 2015).

As acquiring useful information is often a key discussion goal, the knowledge content dimension characterizes the information displayed regarding the focal topic: new ideas, old repetitions, and null content (Chiu, 2000). The validity of an idea is clear in some contexts (e.g., arithmetic), but not others (e.g., poetry). A justification provides evidence, an explanation or citation of an authority to support the validity of an idea (e.g., Neuman, Leibowitz, & Schwarz, 2000). Online discussants

Table 1. A framework characterizing online discussion messages

Dimension Categories	Description
Knowledge Content	
Correct, new idea (Micro-creativity)	An idea that is both correct (consistent with both subject content and problem constraints) and new relative to the discussion
Wrong, new idea	A new idea that is inconsistent with at least the subject content or problem constraint
Justification	An action that supports a new idea by linking it to data, using a warrant, or backing (Toulmin, 2003)
Repetition	An idea that has been mentioned earlier in the topic being discussed
No academic content	A message without any problem-related information (e.g., simple evaluations ["No"], simple questions ["What?"], or off-topic information ["Where are you from?"])
Social Metacognition	
Agree	Agree with a previous message
Disagree	Disagree with a previous message
Question	A query that expects a response
Command	A directive to demand/stop an action
Social Cues	
Positive	Words, symbol, or emoticon expressing a positive attitude or affective state toward others
Negative	Words, symbol, or emoticon expressing a negative attitude or affective state toward others
Individual Characteristics	
Masculine	A masculine participant, as indicated by nickname, photo, personal statement, etc.
Feminine	A feminine participant, as indicated by nickname, photo, personal statement, etc.
Past forum experience	The number of past messages posted by a participant
Topic initiator	A discussant who initiates the current topic

can monitor and try to control one another's actions (social metacognition, Chiu & Kuo, 2009). For example, online discussants can use evaluations (e.g., agree, disagree) to assess one another's ideas and identify flaws. Also, they can invite audience participation to varying degrees with an invitational form (e.g., question, command).

Discussions can contain positive or negative social cues. These include verbal or nonverbal actions in face-to-face discussions and words, symbols, or emoticons in written messages in online discussions (Derks, Bos, & von Grumbkow, 2007). Igniting emotional responses, social cues can help build up or tear down social relationships, along with the collaborative fabric of their discussion (Sproull & Kiesler, 1986; Swan & Shih, 2005).

Lastly, individual characteristics denote a message author's displayed identity. During online discussions, participants usually do not know one

another's actual individual characteristics. Instead, they often portray a public image by displaying information such as online gender or past forum experience. By initiating a topic or responding to one, they also announce their role.

Micro-Creativity Processes During Online Discussions

Micro-creativity is an expressed idea that is both correct (consistent with the problem situation and the subject content) and new relative to the topic discussion. Like face-to-face discussions, online discussions with greater micro-creativity are more likely to yield correct solutions to problems (e.g., Chiu, 2008). Hence, understanding micro-creativity processes can help educators improve online students' understanding, group problem solving and learning. Specifically, new

ideas and justifications (knowledge content), disagreements and questions (social metacognition) in recent messages and authors' individual characteristics can facilitate or hinder participants' micro-creativity (Chen et al., 2012a).

New Ideas and Justifications

As online discussants are often geographically and culturally diverse, they are more likely than face-to-face discussants to have diverse views and sources of knowledge (Swann, Kwan, Polzer, & Milton, 2003). Capitalizing on this diversity, heterogeneous participants can both generate diverse ideas and build on one another's micro-creativity through processes such as sparked ideas and jigsaw pieces (Paulus & Brown, 2003). One person's micro-creativity (e.g., a key word) might spark another person's micro-creativity by activating related concepts in his or her semantic network (Nijstad, Diehl, & Stroebe, 2003). For example, when responding to a micro-creative act, " $5 \times 9 = 5 \times (10 - 1)$ ", a participant might continue the thread and add another micro-creative act: " $5 \times (10 - 1) = 5 \times 10 - 5 \times 1 = 50 - 5 = 55$." Like fitting jigsaw pieces together, participants also can combine micro-creative acts into a new micro-creative act (Milliken, Bartel, & Kurtzberg, 2003). Hence, recent micro-creativity can increase subsequent micro-creativity.

A justification suggests that an idea is reasonable and hence, often enhances its perceived validity (e.g., Neuman et al., 2000). A justification can support an idea's validity by linking it to data, using a warrant, or supporting a warrant with backing (Toulmin, 2003). Furthermore, online discussants justify new ideas in explicit writing during online discussions, so they are more likely to recognize errors and specify relationships among ideas more precisely than during face-to-face discussions (Jonassen & Kwon, 2001). Hence, justifications can support the validity component of new ideas, identify errors, and clarify relationships to aid micro-creativity.

Disagreements and Questions

Unlike agreements, a disagreement tries to alter the discussion trajectory by identifying obstacles/flaws of the previous action or by developing alternatives. A respondent who recognizes a flaw or has a conflicting view of the previous message's understanding of terms, concepts, or schemas is likely to disagree with the previous message (Gunawardena, Lowe, & Anderson, 1997). To support his or her disagreement, the respondent might add a new idea. Or, other participants might address the disagreement with subsequent micro-creativity, according to socio-cognitive conflict theory (Buchs, Butera, Mugny, & Darnon, 2004; Piaget, 1985). In both cases, disagreements can aid micro-creativity, either immediately or in subsequent actions.

When a discussant has both supportive and conflicting information regarding a previous message, online discussants are more likely than face-to-face discussants to disagree. During a face-to-face discussion, participants often seek to enhance their relationship with other group members by agreeing while withholding the conflicting information (Chiu & Khoo, 2003). During an online discussion however, concerns about one's public self-image (face) are less salient (politeness theory; Brown & Levinson, 1987), so online discussants disagree more often than face-to-face discussants do (Lu, Chiu, & Law, 2011). As noted above, this disagreement can facilitate micro-creativity, either immediately or in subsequent messages (according to socio-cognitive conflict theory).

As in face-to-face discussions, questions during online discussions can facilitate discussions by inviting new ideas. When discussants recognize problematic ideas, they can ask questions to identify these knowledge gaps and elicit more information (e.g., "Could you explain the steps further?"). These questions can invite further elaborations, justifications, and micro-creativity. Furthermore, online discussants have a permanent

record of their discussion, no conversation turn structure, more time to find related information and more time for reflection, so online discussants are more likely than face-to-face discussants to reflect on a question, which can increase micro-creativity (Hara et al., 2000). Without a record to support their limited working memory, face-to-face discussants who cannot recall relevant information often ask other group members to review old information rather than create new information, resulting in less micro-creativity (Chiu, 2008). As face-to-face discussants wait their turn to respond to a question, they might forget their ideas or decline to share them if the topic has changed. In contrast, online discussants can write their new ideas at any time and attach their message to the relevant prior message.

Individual Characteristics

Gender differences in micro-creativity might be smaller in online discussions than in face-to-face discussions. As online discussants are less certain of one another's sex, their gender identities are less salient (Palomares, 2004). Also, as the anonymity of online discussions obscures individual differences, online discussants are more likely to perceive the group as an entity, thereby enhancing the salience of the shared social identity (social identity model of deindividuation effects; Postmes, Spears, & Lea, 1998). As a result, behaviors of males vs. females are less likely to differ in online discussions than face-to-face discussions.

Online discussants' past activities in online forums might facilitate their micro-creativity (Wong et al., 2015). Experienced participants with many past posts are more likely to carefully consider others' messages, use time to reflect, and search for more information, all of which might aid micro-creativity. Furthermore, experienced participants are less likely to be frustrated by technological barriers of online discussions, thereby asking fewer technology-related questions (e.g., "How do I type mathematical equations in this forum?"). This in turn allows them to invest more

time and effort in content-focused discussions, which can yield more micro-creativity.

In the forums where discussants respond to an initiator's questions, initiators often ignite discussion topics because they have knowledge gaps in the subject matter and seek to fill the gap by interacting with others (e.g., "Can you explain how to solve this problem?"). Hence, an initiator often shows less micro-creativity than the respondents during discussions.

The Processes of Using Social Cues During Online Discussions

Online discussants' use of social cues can reflect their social relationships. A social cue is a participant's expressed emotion or attitude toward others during a discussion. Specifically, positive social cues refer to a discussant's expression of positive affective states (e.g., "Oh, I get it now") or positive attitudes toward others (e.g., "Thanks for your explanation"). In contrast, negative social cues express negative affective states (e.g., "☹") or negative attitudes toward others (e.g., "No, I'm not, YOU are wrong!!!"; McKee, 2002). While positive social cues signal creation and maintenance of a mutually respectful, supportive, and encouraging climate that promotes positive social relationships and collaboration in the problem content space (Garrison, Anderson, & Archer, 2000), negative social cues often create a discouraging, aggressive or hostile environment that harms participants' social relationships and collaborative learning (Paulus & Roberts, 2006; Sproull & Kiesler, 1986); ritual insults are a notable exception (Progovac & Locke, 2009).

Agreements and Disagreements

When agreeing with a previous idea, online discussants who feel strongly about it might add affective words (e.g., "Aha, I agree with you") or emoticons (e.g., "Yes ☺"), which parallel smiling agreements in face-to-face discussions. In contrast, participants who agree but do not feel

strongly are less likely to post a message at all. By expressing an agreement with positive social cues, the respondent support the previous participant's face, enhance their common ground, and make the discussion more enjoyable, all of which promote positive social relationships (Vinagre, 2008).

When face-to-face discussants disagree with a previous idea, they often disagree politely with positive, verbal and non-verbal social cues to mitigate the threat to the speaker's face and show shared positioning (Brown & Levinson, 1987; Chiu & Khoo, 2003). The benefits of face-to-face polite disagreements are so strong that it is the accepted norm among peers; lack of redress during a face-to-face disagreement is noticeably rude and unacceptable (Holtgraves, 1997). During online discussions however, anonymity and reduced face concerns allow participants to disagree with one another more freely. Thus, online discussants are less likely to use positive social cues to soften or redress disagreements. Moreover, they often add negative social cues to disagree rudely due to the reduced normative constraints of online discussions (e.g., adding exclamation marks, "You made it too complicated!!!!").

New Ideas and Justifications

Online discussants are less likely to use social cues when providing new ideas and justifications. Unlike face-to-face discussions, online discussions lack nonverbal channels' rich interpersonal information (Derks et al., 2007). Furthermore, online discussants usually do not know each other, have weaker social commitments, and are separated by greater psychological distance (Chen & Chiu, 2008; Hiltz & Goldman, 2005). Due to narrowed social cue channels, reduced immediacy and weaker past relationships, online discussants are less likely to attend to others' social presence when sharing new ideas (Sproull & Kiesler, 1991). If participants focus more on problem content and less on social relationships, they are less likely to use social cues when contributing

new ideas. Furthermore, justifications facilitate calm reason-based communication (Chiu, 2008). When justifying new ideas, online discussants often focus on the problem content rather than on their social relationships, so social cues are less likely to accompany justifications.

Earlier Social Cues

While the effects of earlier social cues on subsequent social cues are similar in face-to-face and online discussions, their intensities differ. People tend to reciprocate positive social cues. According to social information processing theory, online discussants often spontaneously post positive social cues to develop positive and meaningful relationships (Vinagre, 2008). Likewise, negative social cues harming social relationships might elicit more negative social cues. For example, a negative social cue, "Can you be more WRONG?!" can threaten the face of the previous participant (Brown & Levinson, 1987). The previous participant can retaliate, with another rude, emotionally-loaded response, "Can you be more STUPID?!" to protect his or her face and attack the previous participant's face (Gottman & Krokoff, 1989). With less social presence in the online discussion environment, negative social cues occur more often there than in face-to-face discussions (cues-filtered-out perspective, Sproull & Kiesler, 1986).

Individual Characteristics

As with micro-creativity, social cue use differs less across gender in online discussions than face-to-face discussions. Initiating an online discussion topic often indicates both a knowledge gap in the subject and motivation to fill that gap by asking others. By using positive social cues and being polite, an initiator is more likely to receive satisfactory responses from others, especially in online discussions in which many messages receive no responses (Hewitt, 2005; Thomas, 2002).

RECOMMENDATIONS FOR TEACHERS AND FORUM DESIGNERS

This chapter highlights how new ideas, justifications, disagreements, and questions can affect micro-creativity and social cues during online discussions. To aid students' micro-creativity and social relationships, teachers can encourage them to prepare their messages carefully, brainstorm new ideas, justify them, and evaluate their ideas carefully (Chen et al., 2012a, 2012b). Meanwhile, forum designers can design the interactive environments that hinder the use of negative social cues (Chen et al., 2012b).

Unlike real-time face-to-face discussions, asynchronous online forums allow delayed responses, so teachers can encourage students to spend more time thinking about their messages, gathering more information, and polishing their messages, which can foster their micro-creativity. Students can brainstorm many ideas before writing out their justifications to choose the best ideas. As micro-creativity can foster subsequent micro-creativity, these micro-creativity chain reactions might be more likely to yield a solution. Meanwhile, students can consider whether they can justify their ideas, which would increase their likelihood of being correct. New ideas and justifications during online discussions reduce the likelihood of social cues, suggesting that when students are reasoning about the content, they can devote less effort to social cues. Hence, when students contribute new ideas and justifications, teachers need not be concerned about the absence of social cues and need not to try to promote them.

Negative social cues during disagreements can tear apart a group's cohesiveness, so teachers often worry about them. As negative social cues and disagreements often occur together in online forums, teachers can encourage students to evaluate one another's ideas slowly and carefully, thereby reducing impulsive, false disagreements that might yield negative social cues and face attacks. Furthermore, the prevalence of such dis-

agreements during online discussions suggests an important role for a teacher: encouraging positive social cues during disagreements to help save face and discouraging face attacks, especially with negative social cues (Savvidou, 2013).

As online discussants often use negative social cues when disagreeing, forum designers can create online discussion environments that hinder their use. For example, forum designers can offer fewer ready-to-use negative emoticons (e.g., emoticons expressing anger or attack) on the toolbar. Also, they can consider adding an automatic detection function to the forums, which would remind the participants to remove inappropriate emoticons/words before they post their messages. By doing so, online discussants might be less likely to use negative emoticons/words to express their impulsive negative feelings when disagreeing.

FUTURE RESEARCH DIRECTIONS

In addition to testing the validity of the above discussion processes, future research can investigate the impact of teacher messages on student messages in course-related forums and compare how students' cognitive and affective styles affect their responses to teachers' messages (Park et al., 2015; Wu & Hou, 2015). Future research can also explore how techniques such as visual analytics and intelligent support can be designed and used by teachers and students to facilitate their online discussion processes (Hsiao & Awasthi, 2015; Kumar & Kim, 2014). Addressing these questions can improve the understanding of students' discussion processes in online forums.

CONCLUSION

This article discusses the processes that influence online discussants' micro-creativity and use of social cues during their discussions. Specifically, new ideas and justifications, disagreements and questions in recent messages facilitate micro-

creativity during online discussions. New ideas and justifications are likely to reduce both positive and negative social cues during online discussions. While agreements facilitate positive social cues, disagreements facilitate negative social cues. Meanwhile, more experienced participants (who have more past posts) are likely to show micro-creativity. Topic initiators are less likely to show micro-creativity but more likely to display positive social cues. Together, these mechanisms show how messages in online discussions create a local context that influences participants' correct outcomes and social relationships.

REFERENCES

- Baltes, B. B., Dickson, M. W., Sherman, M. P., Bauer, C. C., & LaGanke, S. (2002). Computer-mediated communication and group decision-making. *Organizational Behavior and Human Decision Processes*, 87(1), 156–179. doi:10.1006/obhd.2001.2961
- Brown, P., & Levinson, S. C. (1987). *Politeness*. New York: Cambridge University Press.
- Buchs, C., Butera, F., Mugny, G., & Darnon, C. (2004). Conflict elaboration and cognitive outcomes. *Theory into Practice*, 43(1), 23–30. doi:10.1207/s15430421tip4301_4
- Chen, G. (2004). Online discussion in an independent academic BBS forum. *Educational Research Journal*, 19(2), 281–305.
- Chen, G., & Chiu, M. M. (2008). Online discussion processes. *Computers & Education*, 50(3), 678–692. doi:10.1016/j.compedu.2006.07.007
- Chen, G., Chiu, M. M., & Wang, Z. (2012a). Social metacognition and the creation of correct, new ideas: A statistical discourse analysis. *Computers in Human Behavior*, 28(3), 868–880. doi:10.1016/j.chb.2011.12.006
- Chen, G., Chiu, M. M., & Wang, Z. (2012b). Predicting social cues during online discussions. *Computers in Human Behavior*, 28(4), 1497–1509. doi:10.1016/j.chb.2012.03.017
- Chiu, M. M. (2000). Group problem solving processes. *Journal for the Theory of Social Behaviour*, 30(1), 27–50. doi:10.1111/1468-5914.00118
- Chiu, M. M. (2008). Flowing toward correct contributions during group problem solving. *Journal of the Learning Sciences*, 17(3), 415–463. doi:10.1080/10508400802224830
- Chiu, M. M., & Chen, G. (2013). Statistical discourse analysis. In H. L. Lim & F. Sudweeks (Eds.), *Innovative methods and technologies for electronic discourse analysis* (pp. 285–303). Hershey, PA, USA: IGI Global.
- Chiu, M. M., & Khoo, L. (2003). Rudeness and status effects during group problem solving. *Journal of Educational Psychology*, 95(3), 506–523. doi:10.1037/0022-0663.95.3.506
- Chiu, M. M., & Kuo, S. W. (2009). From metacognition to social metacognition. *The Journal of Educational Research*, 3(4), 1–19.
- Coll, C., Rochera, M. J., & de Gispert, I. (2014). Supporting online collaborative learning in small groups. *Computers & Education*, 75, 53–64. doi:10.1016/j.compedu.2014.01.015
- Derks, D., Bos, A. E. R., & von Grumbkow, J. (2007). Emoticons and social interaction on the Internet. *Computers in Human Behavior*, 23(1), 842–849. doi:10.1016/j.chb.2004.11.013
- Garrison, D. R., Anderson, T., & Archer, W. (2000). Critical inquiry in a text-based environment. *The Internet and Higher Education*, 2(2-3), 87–105. doi:10.1016/S1096-7516(00)00016-6
- Gillani, N., & Eynon, R. (2014). Communication patterns in massively open online courses. *The Internet and Higher Education*, 23, 18–26. doi:10.1016/j.iheduc.2014.05.004

- Gottman, J. M., & Krokoff, L. J. (1989). The relationship between marital interaction and marital satisfaction. *Journal of Consulting and Clinical Psychology*, 57, 47–52. doi:10.1037/0022-006X.57.1.47 PMID:2487031
- Gunawardena, C. N., Lowe, C. A., & Anderson, T. (1997). Analysis of a global online debate and the development of an interaction analysis model for examining social construction of knowledge in computer conferencing. *Journal of Educational Computing Research*, 17(4), 395–429. doi:10.2190/7MQV-X9UJ-C7Q3-NRAG
- Hara, N., Bonk, C. J., & Angeli, C. (2000). Content analysis of online discussion in an applied educational psychology course. *Instructional Science*, 28(2), 115–152. doi:10.1023/A:1003764722829
- Hew, K. F., & Cheung, W. S. (2014). Students and instructors use of massive open online courses. *Educational Research Review*, 12, 45–58. doi:10.1016/j.edurev.2014.05.001
- Hew, K. F., Cheung, W. S., & Ng, C. S. L. (2010). Student contribution in asynchronous online discussion. *Instructional Science*, 38(6), 571–606. doi:10.1007/s11251-008-9087-0
- Hewitt, J. (2005). Toward an understanding of how threads die in asynchronous computer conferences. *Journal of the Learning Sciences*, 14(4), 567–589. doi:10.1207/s15327809jls1404_4
- Hiltz, S. R., & Goldman, R. (2005). *Learning together online*. Mahwah, NJ: Lawrence Erlbaum.
- Holtgraves, T. (1997). YES, BUT ... Positive politeness in conversation arguments. *Journal of Language and Social Psychology*, 16(2), 222–239. doi:10.1177/0261927X970162006
- Hsiao, I. H., & Awasthi, P. (2015). Topic facet modeling. *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge* (pp. 231–235). ACM.
- Jonassen, D. H., & Kwon, H. I. II. (2001). Communication patterns in computer mediated and face-to-face group problem solving. *Educational Technology Research and Development*, 49(1), 35–51. doi:10.1007/BF02504505
- Kumar, R., & Kim, J. (2014). Special issue on intelligent support for learning in groups. *International Journal of Artificial Intelligence in Education*, 24(1), 1–7. doi:10.1007/s40593-013-0013-5
- Lu, J., Chiu, M. M., & Law, N. (2011). Collaborative argumentation and justifications. *Computers in Human Behavior*, 27(2), 946–955. doi:10.1016/j.chb.2010.11.021
- Milliken, F. J., Bartel, C. A., & Kurtzberg, T. R. (2003). Diversity and creativity in work groups. In P. B. Paulus & B. A. Nijstad (Eds.), *Group creativity* (pp. 32–62). New York: Oxford. doi:10.1093/acprof:oso/9780195147308.003.0003
- Neuman, Y., Leibowitz, L., & Schwarz, B. (2000). Patterns of verbal mediation during problem solving: A sequential analysis of self-explanation. *Journal of Experimental Education*, 68(3), 197–213. doi:10.1080/00220970009600092
- Nijstad, B. A., Diehl, M., & Stroebe, W. (2003). Cognitive stimulation and interference in idea generating groups. In P. B. Paulus & B. A. Nijstad (Eds.), *Group creativity* (pp. 137–159). New York: Oxford University Press. doi:10.1093/acprof:oso/9780195147308.003.0007
- Onah, D. F., Sinclair, J., & Boyatt, R. (2014). Exploring the use of MOOC discussion forums. *Proceedings of London International Conference on Education* (pp. 1–4).
- Palomares, N. A. (2004). Gender schematicity, gender identity salience, and gender-linked language use. *Human Communication Research*, 30(4), 556–588. doi:10.1111/j.1468-2958.2004.tb00745.x

- Park, J. B. H., Schallert, D. L., Sanders, A. J., Williams, K. M., Seo, E., Yu, L. T., & Knox, M. C. et al. (2015). Does it matter if the teacher is there? *Computers & Education*, 82(C), 315–328. doi:10.1016/j.compedu.2014.11.019
- Paulus, P. B., & Brown, V. (2003). Ideational creativity in groups. In P. B. Paulus & B. A. Nijstad (Eds.), *Group creativity* (pp. 110–136). New York: Oxford University Press. doi:10.1093/acprof:oso/9780195147308.003.0006
- Paulus, T., & Roberts, G. (2006). Learning through dialogue. *Journal of Technology and Teacher Education*, 14(4), 731–754.
- Piaget, J. (1985). *The equilibrium of cognitive structures* (T. Brown & K. L. Thampy, Trans.). Chicago: University of Chicago Press.
- Postmes, T., Spears, R., & Lea, M. (1998). Breaching or building social boundaries? *Communication Research*, 25(6), 689–715. doi:10.1177/009365098025006006
- Progovac, L., & Locke, J. L. (2009). The urge to merge. *Special Issue of Biolinguistics*, 3(2-3), 337–354.
- Qiu, M., & McDougall, D. (2013). Foster strengths and circumvent weaknesses. *Computers & Education*, 67, 1–11. doi:10.1016/j.compedu.2013.02.005
- Savvidou, C. (2013). Thanks for sharing your story. *Technology, Pedagogy and Education*, 22(2), 193–211. doi:10.1080/1475939X.2013.787267
- Sproull, L., & Kiesler, S. (1986). Reducing social context cues. *Management Science*, 32(11), 1492–1512. doi:10.1287/mnsc.32.11.1492
- Sproull, L., & Kiesler, S. (1991). *Connections*. Cambridge, MA: MIT Press.
- Swan, K., & Shih, L. F. (2005). On the nature and development of social presence in online course discussions. *Journal of Asynchronous Learning Networks*, 9(3), 115–136.
- Swann, W. B. Jr, Kwan, V. S. Y., Polzer, J. T., & Milton, L. P. (2003). Fostering group identification and creativity in diverse groups. *Personality and Social Psychology Bulletin*, 29(11), 1396–1406. doi:10.1177/0146167203256868 PMID:15189577
- Thomas, M. J. W. (2002). Learning within incoherent structures. *Journal of Computer Assisted Learning*, 18(3), 351–366. doi:10.1046/j.0266-4909.2002.03800.x
- Toulmin, S. E. (2003). *The uses of argument*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511840005
- Vinagre, M. (2008). Politeness strategies in collaborative e-mail exchanges. *Computers & Education*, 50(3), 1022–1036. doi:10.1016/j.compedu.2006.10.002
- Walther, J. B., Loh, T., & Granka, L. (2005). Let me count the ways. *Journal of Language and Social Psychology*, 24(1), 36–65. doi:10.1177/0261927X04273036
- Wong, J. S., Pursel, B., Divinsky, A., & Jansen, B. J. (2015). An analysis of MOOC discussion forum interactions from the most active users. In *Social Computing, Behavioral-Cultural Modeling, and Prediction* (pp. 452–457). Springer International Publishing. doi:10.1007/978-3-319-16268-3_58
- Wu, S. Y., & Hou, H. T. (2015). How cognitive styles affect the learning behaviors of online problem-solving based discussion activity. *Journal of Educational Computing Research*, 52(2), 277–298. doi:10.1177/0735633115571307

KEY TERMS AND DEFINITIONS

Micro-Creativity: An idea that is both correct (consistent with both subject content and problem constraints) and new relative to a discussion.

Multi-Threaded Discussion: A discussion in which messages proceed along multiple threads.

Negative Social Cue: A discussant's expressed negative affective state or negative attitude toward others.

Online Discussion Message: The content that an online discussant posts at a time.

Online Discussion Topic: A topic or problem that initiates a discussion.

Online Discussion: A discussion in which a group of participants exchanging ideas by posting

messages on an electronic medium (e.g., online forum).

Positive Social Cue: A discussant's expressed positive affective state or positive attitude toward others.

Social Relationship: The relationship between two or more participants during an online discussion, as indicated by their use of social cues.

Topic Initiator: A discussant who initiates the current online discussion topic.

W

The Economics of Internetization

Constantine E. Passaris
University of New Brunswick, Canada

INTRODUCTION

Internetization is a new word and concept that has gained currency with the advent of the new global economy of the 21st century. Prior to the ascendance of Internetization economists had embraced the term globalization to describe the operational parameters of the new economy. The problem with the word globalization is that it is neither a new concept nor is it an appropriate descriptive for the contemporary transformational change precipitated by the spectacular technological inventions on the international economic landscape.

Globalization does not portray the enabling powers of contemporary technology. This electronic capacity has empowered the information technology revolution which is a defining feature of the new global economy. Globalization is a throwback to a previous age prior to electronic connectivity and with more limited means of information accessibility and rapid communication.

The new word, Internetization, describes more succinctly the transformative powers of the world-wide-web and the electronic information highway on the evolving dynamics of interconnectivity for the new global economy of the 21st century. Indeed, Internetization captures the pervasive influence of technological change and electronic innovations on the global economy and all aspects of human endeavour for our civil society in the 21st century.

BACKGROUND

Globalization is not a modern concept or a new theoretical construct. Indeed, it has been a constant feature of international economic outreach since

time immemorial. Globalization has evolved and mutated over the centuries to reflect the priorities and ambitions of different generations. The global outreach of nations for geopolitical, economic, military and trade benefits has transgressed the centuries and embraced almost every country in the world (Erllichman, 2013).

From time immemorial, the process of globalization has taken different forms and proceeded in different directions. Through the discovery and exploitation of new found lands, through the military conquest and annexation of adjacent territories and through the signing of contemporary multilateral free trade agreements, the process of globalization has been an uninterrupted continuum in the evolving history of humankind.

Examples of globalization in ancient history include the seafaring voyages of Odysseus recorded by Homer in *The Odyssey*. The Babylonian Empire that stretched over Mesopotamia in western Asia between the rivers Tigris and Euphrates from 1894 BC to 1595 BC, and again from 625 BC to 539 BC when its grasp reached as far as Palestine. The conquests of Alexander the Great (356 BC to 323 BC) forged an empire that included parts of Europe, Africa, and the Asian continent as far as India. In the late 3rd century BC, the Romans began their conquest of the Balkan Peninsula in search of iron, copper, precious metals, agricultural crops and slaves. This marked the beginning of the Roman Empire, which lasted from 27 BC until 476 AD, and blended unity and diversity across Sicily, Spain, Macedonia, Greece, Egypt, North Africa, Syria, parts of Asia Minor, Gaul and Britain. The Byzantine Empire lasted from 395 AD to 1453 AD and spanned the Middle East, North Africa and Spain. The British Empire from 1583 AD to 1931 AD included such a large collection

DOI: 10.4018/978-1-5225-2255-3.ch694

of countries around the world that it sparked the familiar phrase “the sun never sets on the British Empire” (Passaris, 2006a).

This short and selective geographical survey of the history of globalization attests to the permanence of humankind’s international ambitions. The steady progression of globalization has found expression in the geopolitical and economic ambitions of military, economic and political superpowers. Their globalization ambitions have been achieved by means of wars, mercantilism, colonization, political and economic supremacy, and more recently, through international economic liaisons and multilateral trade agreements.

NEW ECONOMY

The new global economy of the 21st century has transformed the economic, social, educational and political landscape in a profound and indelible manner. Never before in human history has the pace of structural change been more pervasive, rapid and global in its context.

The new economy has become a catalyst for geopolitical symbiosis, economic integration and financial interconnectedness. The new economy is composed of a trilogy of interactive forces that include globalization, trade liberalization and the information technology and communications revolution. Globalization has melted national borders and redefined economic policy. Free trade has enhanced economic integration and extended the economic architecture. The information and communications revolution has made geography and time irrelevant and enhanced the reach of economic parameters (Passaris, 2015).

The advent of the new economy has resulted in the fundamental restructuring of economic society. Electronic interconnectedness is the glue that holds the contemporary global economy together. The new economy is built on a culture of innovation and an emphasis on creativity. Indeed, the signature mark of the new global economy is new ideas, new technologies and new initiatives.

ECONOMIC GLOBALIZATION



A working definition of economic globalization in its contemporary form can be summarized as the global integration of economies through trade and investment flows as well as the internationalization of the production of goods and services. The economic profile of globalization includes the development of global corporations and global networks. It also includes the widespread internationalization of all forms of economic activity in production, marketing, consumption, capital, standards and tastes. Economic governance has resulted in a rapid growth in intra-firm and intra-network trade of components and sub-assemblies as well as finished products leading to a much higher level of specialization (Rodrik, 2008).

Globalization has become the catalyst for the re-orientation of large-scale production in high wage economies from economies of scale to economies of scope. It has contributed to the shortening of product cycles, placing a high premium on innovation, product quality and niche marketing. Globalization has also witnessed the rapid growth and diffusion of service and knowledge-intensive activities, for both products and processes (Huwart & Verdier, 2013).

The process of globalization has been driven by technological change and financial liberalization and sustained by an appreciation among policy makers that an open and rules-based international trading and financial system is essential to global economic progress (Stiglitz, 2008). The new economy has become truly global in scope and substance. The free flow of capital, labour, goods and services within free trade regions, the development of new financial instruments and institutions, instantaneous access to information and communication through the new digital networks, have created a fully integrated global economic system of tremendous scope and opportunity. Economic globalization has achieved a higher level of international economic interdependence and linkages than ever before. For example, the speed and rapidity of economic and financial

contamination across national borders during the 2008 financial crisis is a case in point.

TRADE LIBERALIZATION

The second axiom of the new economy is trade liberalization. The prevailing philosophy in favour of trade liberalization is based on the export led growth model which espouses the economic benefits of exports to the national economy in the form of employment creation, income generation and as a contributor to economic growth (Wacziarg & Welch, 2008).

In the contemporary context, most countries around the world have endorsed the principle and signed on to the potential economic rewards from global trade liberalization. The contemporary vision of the new global economy embraces the promotion of a free trade environment that encourages trade across national borders of goods and services, the legal transfer of intellectual property and the unregulated flow of financial capital. The economic benefits of this modern paradigm along with improvements in the international transportation network have resulted in the expansion of the range of tradable products and services and empowered the global integration of the domestic resources and the production capacity of national economies (Michael, 2009).

One of the most striking differences between the new economy and the one that preceded it is found in the magnitude and rapid movement of international capital flows. Capital account liberalization, the development of new financial instruments and the emergence of new digital technologies have created a fully integrated capital market of tremendous scope and composition. Indeed, a major force driving the growth of international trade and investment has been the liberalization of capital controls, exchange regulations and global financial transactions (Passaris, 2011a).

In short, both deregulation and technological innovation are driving the modern globalization process by tearing down the barriers that

have separated economic markets and reducing transportation and production costs. This has contributed to a greater willingness and ease to trade and invest in foreign countries.

INFORMATION TECHNOLOGY

Information and communication technologies play a central role in the new global economy of the 21st century. The information technology revolution has profoundly altered the structural parameters and the modus operandi of most national economies. More specifically, information and communications technologies have altered the production function, enhanced productivity growth, facilitated the collection of data, spearheaded the transmission of ideas and extended the reach of economic and social interactions. Scientific advances and technological breakthroughs have contributed to the success of the information and communication technology in effectively shrinking the time and distance that separate economic regions and international markets around the world.

The role of information and communications technology in the new economy has been pivotal. This is particularly true of the changing structure of international production. In this context, firms are integrating the production and marketing of goods and services across national borders. International economic transactions that were formerly conducted between independent entities are now being internalized within a single firm or multinational corporation. The new technological infrastructure has empowered services to be delinked from production and traded or performed remotely. In this contemporary venue the market for a growing number of internationally integrated but geographically dispersed business enterprises is global, rather than national or regional (Passaris, 2006a). Indeed, information and communications technologies have displaced the physical market with the virtual market of the Internet for business to business and business to consumer transactions.

At the very heart of the information and communications revolution is the vital process of the commercialization of scientific discoveries and new inventions. There is no denying that the road well-travelled from invention to innovation is long and fraught with many obstacles (Jones, 2009). It is not unusual for many inventions to be left behind because of obstacles in securing the necessary financial capital or adapting an invention to the economic realities of mass production. Indeed, an invention that is the product of a new idea, extensive research and a successful laboratory controlled experiment does not guarantee that it will result in the launch of an innovation. An idea for a new product, a better product or a new process that meets all its specifications as a blueprint and results in a successful invention in a controlled environment may turn out to be an unprofitable undertaking in the world of mass production and global competition. Furthermore, in this modern era individual inventors like Graham Bell, Thomas Edison and Guglielmo Marconi who endowed us with path breaking inventions practically single-handed, are few and far between. Inventions today are more likely to be the product of a team effort and a concerted research and development initiative of some government laboratory, academic institution or a major corporation.

Economists are divided into two schools of thought regarding the process that leads to inventions. The first school subscribes to the notion that inventions are an incremental and marginal process. The second school of thought argues that some inventions are the catalyst for abrupt structural change that permeates the economic landscape in a tidal wave of production realignments and technological clustering. Regardless of what school one subscribes to, there is no denying that the great inventions that took place during the industrial revolution between 1860 and 1900 had a profound impact on economic productivity and personal lifestyle. These inventions included electricity, the internal combustion engine, radio, the telephone, phonograph, motion pictures, the chemical and pharmaceutical indus-

tries, advances in entertainment, communications, urban sanitation and travel in the form of air and motor transportation. The information revolution has also triggered a new spurt of inventions with expansive structural changes and a cluster of innovations with a far reaching economic and social impact. The list of inventions ascribed to the information revolution is still in its infancy but already it includes such significant inventions as computers, the Internet and wireless telecommunications devices.

The role of innovation as a catalyst that drives the engine of economic growth is a fundamental postulate of the new global economy. Contemporary economies have recognized the important role of innovation in nation building and have devoted their resources and public policy focus towards supporting and enhancing the advancement of innovation. Accelerating the process of innovation is vital to ensuring that national economies create and sustain wealth in a swiftly changing global economy.

INTERNETIZATION

The ancient Greek philosopher Heraclitus observed that there is nothing permanent except change. In this regard, recent structural changes on the economic landscape and the collateral requirements for rapid economic recalibration are becoming essential prerequisites for successfully navigating the new global economy.

The shortcomings and limitations of the term globalization requires the introduction of a new word and a new conceptual framework. A new word that is an appropriate descriptor for the operational parameters of the new global economy of the 21st century. Indeed, the word globalization does not embrace the intricacies and nuances of the new world economic order. Furthermore, globalization does not project the extensive digitalization and the electronic connectivity that permeates the new global economy. The word Internetization is new to the economic

lexicon. It was coined for the singular purpose of replacing the word globalization. Internetization captures the importance of the electronic and digital transformation that has defined the new global economy and the pervasive influence of the World Wide Web and the information super highway on all aspects of human endeavour for our society in the 21st century. The transformation of globalization into Internetization was facilitated by the creation of a digital infrastructure that enabled global outreach through electronic connectivity.

The process of Internetization is not static. It is constantly evolving, mutating and transforming. The capacity for Internetization took a giant leap forward with the transformation of wired electronic technology into wireless devices. New technological frontiers have been reached through nanotechnology, Internet of Things, cloud computing and virtual networks.

The word Internetization has gained currency and visibility since it was first introduced into the economic lexicon during the first decade of the 21st century. It has become widely adopted by the academic community and has appeared in scholarly publications (Passaris, 2001; Passaris, 2006b; Kwan, 2011; Tan, 2013; Ankit, Khandelwal, Sinha and Alex, 2014; Iakobidze & Turashvili, 2014). Internetization has been used as a Twitter hashtag since 2009.

Internetization has precipitated a global communications network that connects billions of people to data, machines, computers and each other. The Internetization process has extended social contact, facilitated economic liaisons, enhanced the transmission of services, disseminated knowledge and ideas and truly made the world a global village. In essence, Internetization refers to how people, businesses and cultures have increased their capacity to interact on multiple levels through revolutionary advances in digital technology.

On the social front, Internetization has broken the ties of neighbourhood based on geographical proximity and replaced them with electronic communities. Similarly, family ties based on blood

relatives have been extended through the medium of the Internet and social media. This domain now includes a broader social community of likeminded people with similar interests who socialize on Facebook, Skype, LinkedIn, Instagram and Twitter. In consequence creating a new code of civil values and communications etiquette.

Internetization has had a significant influence on the scope and magnitude of the new global economy of the 21st century. Indeed, Internetization is the catalyst that binds and connects the three pillars of the new global economy and empowers the synergies that are the signature mark of the modern economic landscape. Internetization is a process that is empowered by the information and communications technology revolution in a borderless world with a tremendous capacity for virtual connectivity. In short, Internetization combines the modern version of globalization with the economic empowerment of the Internet.

All in all, Internetization covers more ground, has more contemporary descriptive features and is a more comprehensive concept than globalization for the new economic landscape. This is especially the case as an ever growing number of people from around the world use the Internet as a daily means of accessing new information resources, conducting more convenient financial transactions, experiencing new means of social interaction and indulging in instantaneous means of communication.

Hardly a day goes by when our individual and collective lives are not touched by some aspect of the information technology and communications revolution. From the way we shop, eat, dress, invest, travel, entertain ourselves, communicate with each other, access health care, or pay our bills. These are just a few of our routine daily functions that have been profoundly influenced by the process of Internetization. We shop on-line, we access government services on-line, we book our travel itinerary on-line, attend church services on-line, we pay our bills on-line, we read our newspapers on-line and we do our banking on-line. The electronic prefix that is appearing before an

increasing number of our daily interactions such as e-commerce, e-mail, e-learning, e-banking and e-government is a tangible expression of the pervasive influence of the information technology revolution and the Internet (Passaris, 2006b).

Governance and government services have experienced a foundational transformation as a result of Internetization. Public policy in particular has been subjected to a higher degree of transparency and accountability through the digitalization of government documents. Open government has become the new standard for modern governance (Passaris, 2011b). Furthermore, Internetization has facilitated the digital interface between the public and government services. This interface covers a broad range of services including accessing, completing and submitting forms and applications online, bidding for government contracts, acquiring motor vehicle licenses, filing income tax returns, the issuing of birth and death certificates, marriage licenses, applications for retirement benefits, filing copyright documents, applying for veterans benefits and a host of other public services.

The process of Internetization describes the role of modern technological enablers and the capacity of the Internet in empowering the new global economy of the 21st century. This has been achieved through the advent of computerization, the technology of instant communication, by extending the capacity of economic outreach and by enhancing the speed for completing financial transactions. In short, the word Internetization combines a more comprehensive dimension of the modern era of global outreach and the economic empowerment of the Internet.

Internetization has had a significant influence on the scope and magnitude of the new global economy of the 21st century. Electronic interconnectedness is the glue that holds the contemporary global economy together. The new economy is built on a culture of innovation and an emphasis on creativity. In effect, Internetization is a more comprehensive and contemporary descriptor of the new economic landscape. Particularly since

the public and private sectors, small, medium and large businesses, non-profit organizations and community associations rely on the Internet.

The information technology revolution has profoundly altered the structural parameters and modus operandi of most national economies. Internetization has become the catalyst and the engine that drives the information and communication revolution in the context of the new economy. This takes the form of the digitization of information and the creation of the information super highway. Furthermore, the industrial age created territorial communities whereas the information age is creating electronic communities.

Internetization is multifaceted and multidimensional. This has become abundantly clear in the contemporary knowledge driven economy. At the very heart of the information technology applications for the knowledge based sector is the widespread use of computers and robotics. A collateral benefit of this transformation has been the extraordinary scale of research and development in the quest for new applications for the advances in information and communications technology. This has triggered the phenomenal growth of the software industry and related business services. Indeed, the scale of investment in computerized equipment and in the telecommunications infrastructure is unprecedented. In addition, the rapid growth of niche markets for satellite and peripheral industries supplying information and communications technology products and specialized components and services have catapulted the knowledge based sector into the leading sector of the new economy of the 21st century.

The word global has taken on a new meaning since the emergence of the Internet more than three decades ago. The Internet has opened up countless opportunities for businesses and individuals worldwide. It has eliminated restrictions and borders regarding communication and interaction. It has archived an immense amount of information at our fingertips. Internetization has triggered an age of individual and collective empowerment that is unprecedented in our history

of civilization. It provides individuals, businesses and governments with a global influence and outreach. In short, Internetization has made time and geography irrelevant.

The ascendance of Internetization was propelled by the prominent role of human capital as a foundational characteristic of the new economy. The new global economy of the 21st century has given birth to a new and vibrant knowledge sector. More specifically, the knowledge based economy is fuelled by technology, human capital and research and development. This synergistic approach contributes to accelerating levels of productivity and economic performance. In short, the fuel of the new economy is technology and its currency is human capital. The product of the new economy is knowledge and its market is the virtual marketplace facilitated by the Internet.

FINANCIAL INTERNETIZATION

The most pervasive impact of Internetization on the economic landscape has been in the empowerment of electronic financing. Internetization is responsible for spearheading new economic, financial and commercial transactions. This takes the form of e-commerce, e-banking, electronic business transactions, instantaneous credit card purchases, debit cards, Internet bill payments, mobile phone financial transactions, electronic cash and the electronic transfer of capital. There is no denying that Internetization has facilitated the instantaneous convenience of digital financial transactions.

In most advanced industrialized countries, the use of cash and paper currency as a medium of exchange and a means for payment has declined dramatically. All of this as a direct result of the advent of Internetization in the form of electronic connectivity and capacity. There has also been a marked decline in the public demand for cash services along with a deliberate switching of financial transactions from paper cash to digital currency.

The economic impact of Internetization in terms of financial transactions is most discernible in Sweden. In this regard, Sweden has embraced and is acknowledged as a leader in the transformational process for switching from conventional currency into the digital medium. Sweden has become the role model for the effective application of Internetization in regard to conducting the modern era of financial transactions.

Swedish banks have been particularly enthusiastic about this transformation. They have eliminated ATM cash dispensing machines and promoted the use of digital currency. More specifically, Sweden's largest banks such as, SEB, Swedbank and Nordea Bank, have stopped manual cash-handling services in a large number of their local branches. This action is a consequence of their assessment that Swedes rely on credit cards, the Internet and mobile phones to make a significant portion of all their payments.

The Swedish Bankers Association has noted that "over the past few decades, the use of paper-based payments such as giro forms, cheques and cash payments have rapidly been replaced by electronic payments of various types. As an example, the use of different kinds of cards has increased from 100 million transactions in the middle of the 1990s to 1,956 million transactions in 2011. During the same period, the use of cheques has in practice ceased. While the share of electronic giro payments, mainly online, has increased, the share of paper-based giro transactions has decreased. The share of Swedes older than 15 years, who pay their bills through an Internet bank, has increased from 9 percent in 1999 to 79 percent in 2012. In the younger age groups (aged up to 34 years), up to 96 percent pay their bills through the Internet. During the same period, the use of paper based payments, such as giro forms, diminished from almost 79 percent in 1999 to 12 percent in 2012." (Swedish Bankers Association, 2013, p. 13). In consequence, the end of cash is a natural next step in an evolutionary process that has already led to the extinction of chequebooks.

It is worth noting that the transformational impact of Internetization on the scope and substance of financial transactions has exposed several vulnerabilities and fault lines on the contemporary digital landscape. The element of financial risk takes the form of identity theft, digital financial fraud, unauthorized electronic currency transfers, misappropriation of payments in the course of electronic consumer to business transactions and electronic credit card theft.

CYBER SECURITY

The rapid pace of technological innovation that has been the hallmark of the Internetization process has exposed the digital super highway to distinctive vulnerabilities. These take the form of hacking, malware infection, identity theft, electronic espionage, cyber terrorism and financial misappropriation.

It is becoming increasingly clear that the contemporary digital safeguards, firewalls and electronic locks do not offer a high level of personal privacy and cyber protection. Indeed, cybersecurity has become one of the foremost threats and negative side effects of Internetization. In particular, lapses in cybersecurity have exposed the inherent systemic risk of Internetization in the conduct e-commerce, e-banking, e-entertainment and e-communication.

On the contemporary landscape, advances in technological innovations are moving faster than the adoption of security protocols to police, control and protect digital connectivity. It is becoming increasingly prevalent on the contemporary digital network that security vulnerability in electronic devices has become systemic. This vulnerability may take the form of a successful cyber-attack that results in an invasion of privacy, access to personal emails, the remote high jacking of personal e-banking functions and the resetting of one's passwords. Ultimately all of this results in stealing a person's electronic identity.

The frequency of cyberespionage between countries has increased recently as a result of the availability of spy software and cyberespionage products. FinFisher, one of the best-known purveyors of spyware and cyberespionage training courses, has a clientele that includes countries and government agencies around the world. Modern cyberespionage products have the capacity to infect their targets' computers and phones, copy messages, record conversations and even activate webcams. The international media make headlines when they report on countries using cyberespionage to gain illicit electronic access and eavesdrop on their national neighbors, competitors, enemies and even members of their own political alliances.

The capacity of Internetization has been a great benefactor to the social media. Indeed, the impact of the social media on personal connectivity, business promotion, enhanced NGO's networks and government transparency have been massive. In some cases it has also created a big electronic headache. For example, the emergence of WikiLeaks, an international non-profit journalistic organization, whose mission is to expose secret information, news leaks and classified government documents has created considerable public embarrassment for some governments.

Electronic vulnerabilities and exposure are found in devices that do not offer protection from phishing emails, allow a backdoor access, can be remotely monitored and permit the tampering of an electronic smart security system. In consequence, there is an urgent need for a more effective partnership between government and manufacturers of electronic devices to enhance the security features of their devices in order to prevent the sophisticated hacking and the remote high jacking of those devices.

The next generation of cybersecurity will require an elevated level of personal identification, safer digital locks and the creation of impenetrable electronic firewalls. It has become abundantly clear that a PIN, username and password are simply insufficient and inadequate. The new generation

of cybersecurity will probably take the form of codified electronic locks, voice identification, finger print matching and eye retina protocols. All of this for the purpose of adopting fail safe cybersecurity systems that will protect individual identity, national security and corporate secrets in the digital age.

INTELLECTUAL PROPERTY

A cornerstone of the new economy has been the emergence of spectacular inventions, mass consumption innovations and digital connectivity. A consequence of this global electronic outreach has been the vulnerability of Internetization in regard to the protection of intellectual property. Indeed, the enforcement of intellectual property rights has become increasingly porous and elusive in a virtual and electronically borderless world.

The operational definition of intellectual property rights is the awarding of an economic monopoly to the inventor of an intellectual product or service. Intellectual property rights are foundational for the knowledge economy of the 21st century. In this regard, the conferring of intellectual property rights has legal stature and protection within most national jurisdictions. Examples of intellectual property rights are copyrights, patents, industrial designs and trademarks. The areas that are directly affected by intellectual property rights are music, literature, movies, different forms of artistic expression as well as mechanical, electronic and scientific inventions.

Intellectual property law became increasingly more prevalent in countries around the world in the latter half of the 20th century. Intellectual property rights have two foremost objectives. First, to reward inventors for their intellectual discoveries. Second, to provide an incentive for future inventors who will push the boundaries of intellectual achievement and in turn will propel economic growth and development.

The World Intellectual Property Organization (WIPO) was established in 1967 by treaty as an agency of the United Nations. It is worth noting that the WIPO Intellectual Property Handbook articulates the *raison d'être* of property laws in this manner:

“One is to give statutory expression to the moral and economic rights of creators in their creations and the rights of the public access to those creations. The second, is to promote, as a deliberate act of Government policy, creativity and the dissemination and application of its results and to encourage fair trading which would contribute to economic and social development.” (WIPO, 2015).

While the protection and enforcement of intellectual property rights prevails in most national jurisdictions, its enforceability across national borders is more problematic. There is no denying that Internetization has diluted the protection and enforcement of intellectual property rights. This is especially the case where the advancement of scientific knowledge has spawned new industries in the fields of biotechnology and nanotechnology and the originators of these new technologies have sought intellectual property protection.

On the contemporary digital landscape, electronic search engines like Kazaa and Gnutella represent a challenge for copyright policy. The Recording Industry Association of America has been at the forefront of the battle against electronic piracy and the enforcement of copyright laws. Modern legislation such as the Digital Millennium Copyright Act in the USA is an example of national attempts to thwart the consequences of malicious software and to enforce digital rights. Despite the contemporary legal protection and enforcement of intellectual rights, in 2011, trade in counterfeit copyrighted and trademarked products was a \$600 billion industry on an international scale and accounted for up to 7% of global trade (Bitton, 2012).

REDISCOVERING SCHUMPETER

The ascendance of the new economy requires a rediscovery of Schumpeterian theories. Schumpeterian theories have the potential of linking microeconomic derivatives to the macroeconomic postulates for economic growth and development. Indeed, Schumpeter's intellectual and theoretical legacy on the pivotal role of entrepreneurship and innovation is a vibrant analysis and laudable framework for determining the causal factors that promote economic prosperity and contribute to the wealth of nations in the 21st century (Passaris, 2001).

Schumpeter emphasized the predominance of sectoral economic analysis and the paramount importance of the entrepreneur as a catalyst for innovation and as the engine that drives economic growth and development. In this microeconomic scenario, innovation in the Schumpeterian model consisted of new products, new processes, new qualities of products, new sources of supply and new forms of business and industry organization. Essentially, it is a "process of industrial mutation.... that incessantly revolutionizes the economic structure from within, incessantly destroying the old one, incessantly creating a new one" (Schumpeter, 1942, p. 79). It is in this context that he gives birth to his most famous phrase- "creative destruction" by which he means the replacement of old products, old enterprises and old organizational forms by new ones. The process of Internetization and its consequential impact on the new economy is a stellar example of the modern relevance of Schumpeterian theories.

Economic growth can only be sustained by finding new and better ways to utilize our limited and finite resources. In this respect, technological progress has always been an ingrained feature of national economies. Schumpeter's explanatory model for economic growth is focused primarily on the role of technological innovation. He proposed a model that postulated growth through the interaction of bursts of technological development

and competition between firms. Schumpeter saw capitalism as moving in long waves; every 50 years or so, technological revolutions would cause "gales of creative destruction" in which old industries would be swept away and replaced by new ones. Each wave of technology would fuel an upsurge in investment and create an impressive amount of job opportunities in new industries (Schumpeter, 1935). It is worth noting that the duration of the innovation cycle appears to be contracting over time from a 50 to 60 year duration to a shorter 30 to 40 year period. What has changed in recent times is the more rapid pace of technological innovation.

The role of the entrepreneur in the progression of the technological cycles was paramount. Schumpeter explains that "the function of entrepreneurs is to reform or revolutionize the pattern of production by exploiting an invention or, more generally an untried technological possibility for producing a new commodity or producing an old one in a new way, by opening up a new source of supply of materials or a new outlet for products, by reorganizing an industry and so on." (Schumpeter, 1942, p. 132).

Internetization is a vivid example of what Schumpeter describes as the process of creative destruction. In this context, Internetization has replaced old technologies with new and improved technologies. This is the process by which wealth and economic growth are achieved. It is also the process by which entrepreneurship creates its most important economic impact in the form of new investment, job creation and economic prosperity. In short, Internetization is the 21st century wave of Schumpeter's innovation cycles and provides contemporary confirmation that the technology life cycle is contracting over time.

The extrapolation of Schumpeter's microeconomic theories has the potential to form the theoretical construct for linking microeconomics with macroeconomics in the context of the new global economy of the 21st century. In this context, innovation and technological progress are an inherently microeconomic phenomenon, which are in turn a

consequence of optimum resource allocation and the profit oriented pursuits of economic activity. However, special mention should be made of the fact that they also result in the Schumpeterian waves of innovation and technological progress that have long term macroeconomic impact and consequences (Passaris, 2003).

In its microeconomic constructs, Schumpeter's concept of creative destruction embraces a multitude of features. These include innovation and obsolescence, the rise and fall of products and processes, entrepreneurial vision, risk and uncertainty, short term market advantage and abnormal profits. All of these microeconomic antecedents resonate with the process of Internetization and the creation of avant garde innovations. They also lead to positive long term macroeconomic consequences in the form of economic growth and development.

CONCLUSION

The advent of the new global economy of the 21st century has been empowered by momentous advances in technological innovation and the unprecedented outreach through digital capacity. In consequence, a new word, Internetization, is a more appropriate portrayal of the contemporary economic system with its virtual connectivity.

Internetization is the catalyst for transformative change on the modern economic landscape. It has precipitated unprecedented structural changes and redefined economic linkages on a global scale. Indeed, Internetization is playing a central role in the new global economy of the 21st century.

The economic profile of the new global economy is driven by technology, fueled by innovation and based on new ideas, new perspectives and new business strategies. Internetization is a process that is empowered by the information and communications technology revolution in a borderless world with a tremendous capacity for virtual connectivity.

FUTURE RESEARCH DIRECTIONS

The economic landscape of the 21st century is changing rapidly. Transformational change is the signature mark of the dynamic internetization process. Indeed, the impact of Internetization is not static; it is evolving and pervasive. Technological innovations are being introduced at record speed. The foundational structural changes that Internetization has precipitated have also triggered several intellectual fault lines that require focused research efforts in the near future.

Foremost among future research endeavors is the causal relationship between Internetization and the contemporary face of structural unemployment. Workers are being replaced as a result of technological innovations. Furthermore, they are having difficulty re-entering the workforce and transitioning to new employment opportunities. During the Industrial Revolution workers who were replaced by machines found work in the factories that made the machines. In the contemporary phase, chronic structural unemployment appears to have become a negative side effect of the Internetization process. The reason for the stubborn resistance of worker re-deployment during the Information Age is that the new jobs have raised the bar with respect to educational requirements, technical expertise and modern workplace competencies. This effectively renders the technologically unemployed ineligible for workplace re-integration.

Future research should also be directed towards ascertaining the scope of the evolving and pervasive nature of Internetization. The impact of technological advances and the consequences of electronic connectivity are both diverse and diffused. Indeed, the interactive nature of governance, economy and society in the context of a virtually interconnected world is a worthwhile research undertaking. This research endeavor will require an interdisciplinary and multidisciplinary approach. For example, the impact of Internetization has not been limited to the economic landscape, but has

also affected the political and governance process, the social values and communications etiquette of the new Information Age.

Another area for future research is in the domain of economic development. The hypothesis that requires analytical examination is the extent to which developing countries have benefited from Internetization. This research endeavor should explore whether Internetization is the great economic equalizer that has created a level playing field for economic development for both developed and developing countries. Alternatively, this research project will expose the extent to which developing countries have been adversely affected as a result of the lack of digital infrastructure and the cost of electronic connectivity. In consequence, this research endeavor will answer the question if Internetization has widened the gap between developed and developing countries or bridged the divide between developing and developed countries in regard to economic growth and individual prosperity.

REFERENCES

- Ankit, A. P., Khandelwal, A., Sinha, A., & Alex, S. A. (2014). Epidemic Analysis (Web Application for Epidemic Analysis and Prediction). *International Journal of Scientific & Engineering Research*, 5(4), 900–904. Retrieved from <http://www.ijser.org/paper/Epidemic-Analysis-Web-Application-for-Epidemic-Analysis-and-Prediction.html>
- Bitton, M. (2012). Rethinking the Anti-Counterfeiting Trade Agreement's Criminal Copyright Enforcement Measures. *The Journal of Criminal Law & Criminology*, 102(1), 67–117.
- Erlichman, H. J. (2013). *Conquest, Tribute and Trade: The Quest for Precious Metals and the Birth of Globalization*. Amherst, N.Y.: Prometheus Books.
- Huwart, J.-Y. & Verdier, L. (2013). *Economic Globalization: Origins and Consequences*. Paris: OECD. Retrieved from http://www.oecd-ilibrary.org/economics/economic-globalisation_9789264111905-en
- Iakobidze, T., & Turshvili, T. (2014). *Internet Freedom in Georgia*, Tbilisi, Georgia: Institute for Development of Freedom of Information. Retrieved November 7, 2015 from <https://idfi.ge/public/upload/pdf/Internet-Freedom-Report-3-4.pdf>
- Jones, B. F. (2009). The Burden of Knowledge and the Death of the Renaissance Man: Is Innovation Getting Harder? *The Review of Economic Studies*, 76(1), 283–317. doi:10.1111/j.1467-937X.2008.00531.x
- Kwan, L. H. (2011). *Competing Globally with Cost Accounting*. University of Tennessee Honors Thesis Projects. Retrieved from http://trace.tennessee.edu/cgi/viewcontent.cgi?article=2417&context=utk_chanhonoproj
- Michael, M. (2009). *Trade Liberalization and Trade Preferences*. Hackensack, N.J.: World Scientific. doi:10.1142/6918
- Passaris, C. E. (2001). Schumpeter's Legacy of Technological Innovation in the Context of the Twenty-First Century. In S. B. Dahiya (Ed.), *Economic Theory in the Light of Schumpeter's Scientific Heritage* (pp. 349–361). Rohtak: Spellbound Publications.
- Passaris, C. E. (2003). *Schumpeter and Globalization: Innovation and Entrepreneurship in the New Economy*. Italy: International Institute of Advanced Economic and Social Studies.
- Passaris, C. E. (2006a). The Business of Globalization and the Globalization of Business. *Journal of Comparative International Management*, 9(1), 3–18.

Passaris, C. E. (2006b). Immigration and Digital Government. In A.-V. Anttiroiko & M. Malkia (Eds.), *The Encyclopaedia of Digital Government* (pp. 988–993). London: Idea Group Reference; doi:10.4018/978-1-59140-789-8.ch148

Passaris, C. E. (2011a). Redesigning Financial Governance for the New Global Economy of the 21st Century. *Journal of Comparative International Management*, 14(1), 1–15.

Passaris, C. E. (2011b). Economic Governance and Full Employment. In M. Ugur & D. Sutherland (Eds.), *Does Economic Governance Matter* (pp. 168–183). Cheltenham, U.K & Northampton, Mass: Edward Elgar Publishing. doi:10.4337/9780857931771.00014

Passaris, C. E. (2015). A New Economic Governance Model for Greece in the 21st Century. *Journal of Heterodox Economics*, 1(3), 163–190.

Rodrik, D. (2008). *One Economics, Many Recipes: Globalization, Institutions, and Economic Growth*. Princeton, N.J: Princeton University Press.

Schumpeter, J. A. (1935). The Analysis of Economic Change. *The Review of Economics and Statistics*, 17(4), 2–10. doi:10.2307/1927845

Schumpeter, J. A. (1942). *Capitalism, Socialism and Democracy*. London: Allen and Unwin.

Stiglitz, J. E. (2008). Making Globalization Work. *The Economic and Social Review*, 39(3), 171–190.

Swedish Bankers Association. (2013). *Banks in Sweden*. Retrieved from [http://www.swedishbankers.se/web/bfmm.nsf/lupGraphics/1302Banker%20i%20Sverige-ENG.pdf/\\$file/1302Banker%20i%20Sverige-ENG.pdf](http://www.swedishbankers.se/web/bfmm.nsf/lupGraphics/1302Banker%20i%20Sverige-ENG.pdf/$file/1302Banker%20i%20Sverige-ENG.pdf)

Tan, S. G. (2013). *Internet Economics: A Regression Analysis on the Factors that Affect the Growth Rate of Internet Usage Around the World*. Retrieved from http://www.academia.edu/5626689/ECONMET_Paper

Wacziarg, R., & Welch, K. H. (2008). Trade Liberalization and Growth: New Evidence. *The World Bank Economic Review*, 22(2), 187–231. doi:10.1093/wber/lhn007

WIPO. (2015). *Intellectual Property Handbook: Policy, Law and Use*. Retrieved from <http://www.wipo.int/about-ip/en/iprm/>

ADDITIONAL READING

Abouzeedan, A., & Busler, M. (2007). Internetization Management: The Way to Run the Strategic Alliances in the E-globalization Age. *Global Business Review*, 8(2), 303–321. doi:10.1177/097215090700800208

Acocella, N. (2005). *Economic Policy in the Age of Globalisation*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511753947

Bhagwati, J. (2008). *Termites in the Trading System: How Preferential Agreements Undermine Free Trade*. Oxford: Oxford University Press; doi:10.1093/acprof:oso/9780195331653.001.0001

Canada. (2014). *Canada's Cybersecurity Strategy*. Ottawa: Public safety Canada, Government of Canada. Retrieved from <http://www.publicsafety.gc.ca/cnt/rsrscs/pblctns/cbr-scrt-strtgty/index-eng.aspx>

Castells, M. (2010). *The Rise of the Network Society*. Malden, MA, USA: Wiley-Blackwell Publishers.

Chesbrough, H. W. (2006). *Open Business Models: How to thrive in the New Innovation Landscape*. Boston, USA: Harvard Business School Press.

Etemad, H., Wilkinson, I., & Dana, L. P. (2010). Internetization as the Necessary Condition for Internationalization in the Newly Emerging Economy. *Journal of International Entrepreneurship*, 8(4), 319–342. doi:10.1007/s10843-010-0062-x

Farah, P. D., & Tremolada, R. (2014). Intellectual Property Rights, Human Rights and Intangible Cultural Heritage. *Journal of Intellectual Property Law*, 2(I), 21–47.

Fortunati, L. (2005). Mediatization of the Net and Internetization of the Mass Media. *International Communication Gazette*, 67(1), 27–44. doi:10.1177/0016549205049177

Freeman, C., & Louca, F. (2001). *As Time Goes by: The Information Revolution and the Industrial Revolutions in Historical Perspective*. Oxford: Oxford University Press.

Grossman, G. M., & Helpman, E. (2015). Globalization and Growth. *The American Economic Review*, 105(5), 100–104. doi:10.1257/aer.p20151068

Hamel, G. (2002). *Leading the Revolution: How to Thrive in Turbulent Times by Making Innovation a Way of Life*. New York, USA: Plume Books.

Irwin, D. A. (1996). *Against the Tide: An Intellectual History of Free Trade*. Princeton, NJ: Princeton University Press.

Irwin, D. A. (2009). *Free Trade Under Fire*. Princeton, NJ: Princeton University Press.

Jones, A. (2010). *Globalization: Key Thinkers*. Cambridge: Polity Press, John Wiley & Sons.

LeClair, M. S. (2002). Fighting the Tide: Alternative Trade Organizations in the Era of Global Free Trade. *World Development*, 30(6), 949–958. doi:10.1016/S0305-750X(02)00017-7

Osle, R. D. (2010). *The New Global Law*. Cambridge: Cambridge University Press.

Osterhammel, J., & Niels, P. P. (2005). *Globalization: A Short History*. Princeton, NJ, USA: Princeton University Press.

Potrafke, N. (2015). The Evidence on Globalisation. *World Economy*, 38(3), 509–552. doi:10.1111/twec.12174

Sassen, S. (2000). *Cities in a World Economy*. Los Angeles, USA: SAGE Publications.

Schechter, R. E., & Thomas, J. R. (2003). *Intellectual Property: The Law of Copyrights, Patents and Trademarks*. New York: West/Wadsworth.

Smith, C. (2010). *International Trade and Globalisation*. Stocksfield: Anforme.

Stiglitz, J. E. (2002). *Globalization and Its Discontents*. New York: W.W. Norton.

Trott, P. (2008). *Innovation Management and New Product Development*. London: Financial Times Prentice Hall.

Wilson, E. J. (2004). *The Information Revolution and Developing Countries*. Massachusetts, USA: Massachusetts Institute of Technology.

KEY TERMS AND DEFINITIONS

Economic Globalization: The global integration of economies through trade and investment flows as well as the internationalization of the production of goods and services.

Financial Internetization: Internetization is responsible for spearheading new economic, financial and commercial transactions such as e-banking, e-commerce, instantaneous credit card purchases, debit cards, Internet bill payments, mobile phone financial transactions, digital cash and the electronic transfer of capital.

Globalization: A traditional concept that embraces international outreach and global linkages.

Internetization: A new word and concept that describes the empowerment of electronic connectivity on the new global economy.

IT Revolution: The information technology revolution has served as a catalyst for electronic connectivity, altered the production function, enhanced productivity growth, facilitated the collection of data, spearheaded the transmission of ideas and extended the reach of economic and social interactions.

Joseph A. Schumpeter: A 20th century economist who postulated the importance of entrepreneurship and innovation as the foundations for wealth creation.

New Global Economy: Describes the 21st century economy that is composed of a trilogy of interactive features that include globalization,

trade liberalization and the information technology and communication revolution.

Trade Liberalization: Also known as free trade encourages trade across national borders of goods and services, the legal transfer of intellectual property and the unregulated flow of financial capital.

An Efficient and Effective Index Structure for Query Evaluation in Search Engines

W

Yangjun Chen

University of Winnipeg, Canada

INTRODUCTION

Indexing the Web for fast keyword search is among the most challenging applications for scalable data management. In the past several decades, different indexing methods have been developed to speed up text search, such as inverted files, signature files and signature trees for indexing texts (Anh and Moffat, 2005; Chen et al., 2004; Chen et al. 2006; Faloutsos, 1985; Faloutsos et al., 1988); and suffix trees and tries (Knuth, 1975) for string matching. Especially, different variants of inverted files have been used by the Web search engines to find pages satisfying a query (Arasu, 2001; Lemple et al., 2003).

A text database can be roughly viewed as a collection of documents and each document is stored as a list of words. Over the documents, there are two kinds of Boolean queries, that is, queries that can be constructed from query terms by conjunction ($\wedge$) or disjunction ($\vee$). A document D is an answer to a conjunctive query $w_1 \wedge w_2 \wedge \dots \wedge w_k$ if it contains every w_i for $1 \leq i \leq k$ while D is an answer to a disjunctive query $w_1 \vee w_2 \vee \dots \vee w_l$ if it contains any w_i for $1 \leq i \leq l$. Conjunction and disjunction can be nested to arbitrary depth, but can always be transformed to a conjunctive normal form:

$$(w_{11} \dots \vee w_{1l_1} \dots) \wedge (w_{k1} \dots \vee w_{kl_k} \dots)$$

In this chapter, we discuss a new method to evaluate both conjunctive and disjunctive queries by decomposing an inverted list into a collection of disjoint sub-lists. The decomposition is based

on the construction of a trie structure T over documents and then associating each document word with an interval sequence generated by labeling T by using a kind of tree encoding.

With this method, we can improve the efficiency of traditional methods by an order of magnitude or more.

BACKGROUND

In order to efficiently evaluate such queries, indexes need to be established. It is well known that English texts typically contain many different variants of basic words, by using variant word endings such as ‘ing’, ‘ed’, ‘ses’, and ‘ation’. All the variants of a word should be regarded as a match and therefore it is efficient for an index only include these basic words, or say, stems. Different algorithms have been developed to extract stems from documents. Among them, the algorithm proposed by Lovins (1968) is widely used.

By the signature file, a word is hashed to a bit string (called a signature) and all the words’ signatures of a document are superimposed (bit-wise OR operation) into a document signature. When a query arrives, its signature will be created using the same hash function and the document signatures are scanned and many nonqualifying documents are discarded. The rest are either checked (so that the ‘false drops’ are removed) or they are returned to the user as they are. The main disadvantage of this method is the false drop (Kitagawa et al., 1997), which needs extra time to check. The signature file is greatly improved by the so-called signature tree (Chen et al. 2006),

by which a set of signatures is organized into a binary tree structure and a sequential search of signatures is replaced with a search of binary trees. However, signature-based methods can be used only for evaluating conjunctive queries. For disjunctive queries, they are not efficient.

As pointed out by many researchers (Anh et al., 2005; Ao et al., 2011; Zobel et al., 2006), the inverted file is a more competitive indexing method than signature-based approaches. It is extensively used by different web search engines due to its efficiency and simplicity. Structurally, it contains two parts: a search structure or vocabulary, containing all the distinct words to be indexed, and a set of inverted lists with each constructed for a distinct word w , storing the identifiers of all those documents containing w . Queries are evaluated by fetching the inverted lists for the query terms, and then intersecting them for conjunctive queries, or merging them (by a set union operation) for disjunctive queries. According to (Zobel et al., 1998), the inverted file is superior to the signature file in almost every respect, including functionality, query time, and space overhead.

Since it was first proposed in mid-1960s, the inverted file has been adopted in information retrieval, database systems, distributed systems (Büttcher et al., 2005; Camel et al., 2001), and different search engines. Also, much effort has been spent on the improvement of its performance by using integer coding (Golomb, 1966), bitmap compression (Apaydin et al., 2006; Bjørklund et al., 2009), caching (Saraiva et al., 2001), and parallelism (Ao et al., 2011). For a static environment, the bitmap compression is most efficient. However, it is obviously not suitable for a dynamical environment like the Web. So, different set intersection algorithms have been proposed to directly manipulate sorted arrays (Barbay et al., 2009), with caching (Barbay et al., 2009; Saraiva et al., 2001), parallelism (Ao et al., 2011), interpolation (Demaine et al., 2004), and even hardware characteristics (Tsirogiannis et al., 2009) being utilized to enhance performance. The method discussed in (Ding et al., 2011) is an in-

memory algorithm for set intersection, by which an inverted list is partitioned into a collection of sublists. Then, generate a hash image for each of them, by which a number in an inverted list is mapped to a single bit in an image. In this way, a set intersection is done by a series of hash value intersections.

INDEX STRUCTURES

In this section, we mainly discuss our index structure, by which each word with high frequency will be assigned an interval sequence. We will then associate intervals, instead of words, with inverted sub-lists. To clarify this mechanism, we will first discuss interval sequences for words in Subsection 1. Then, in Subsection 2, how to associate inverted lists with intervals will be addressed.

1. Interval Sequences Assigned to Words

Let $D = \{D_1, \dots, D_n\}$ be a set of documents. Let $W_i = \{w_{i1}, \dots, w_{in_i}\} (i = 1, \dots, n)$ be all of the words appearing in D_i , to be indexed. Denote $W = \bigcup_{i=1}^n W_i$, called the vocabulary. We define the word appearance frequency by the following formula:

$$f(w) = \frac{\text{num. of documents containing } w}{\text{num. of documents}}, (w \in W).$$

We then define a frequency threshold ζ . For any word w with $f(w) < \zeta$, we will associate it with an inverted list in a normal way, denoted as $\delta(w)$, exactly as in the method of inverted files. However, for all those with $f(w) \geq \zeta$, we will create a new index. For this, we will represent each D_i as a sequence containing all those words w with $f(w) \geq \zeta$, decreasingly sorted by $f(w)$. That is, in such a sequence, a word w precedes another w' if w is more frequent than w' in all documents.

Table 1. Documents and word sequences

DocID	Words	Sorted Word Sequence
1	<i>c, a, f, m, p</i>	<i>c, f, a, m, p</i>
2	<i>c, f, b, a</i>	<i>c, f, a, b</i>
3	<i>b, a, c, d</i>	<i>c, a, d, b</i>
4	<i>f, d, p, m</i>	<i>f, d, m, p</i>

In addition, for any subset of words that have the same appearance frequency a global ordering is defined so that in each sorted word sequence this global ordering is followed. In addition, we maintain a hash table H that maps each word w to its inverted list $\delta(w)$ or to its new index.

Example 1: In Table 1, we show a set of four documents, their words w with $f(w) \geq \zeta = 0.4$, and the corresponding sorted word sequences, where we use a character to represent a word for simplicity.

Notice that the global order on $\{f, a, c\}$ (with $f(w) = 0.75$) is set to be $c \rightarrow f \rightarrow a$ while the global order on $\{m, b, p, d\}$ (with $f(w) = 0.5$) is $d \rightarrow b \rightarrow m \rightarrow p$.

In practice, however, we can adjust ζ to include more words in the new index; and in the extreme case, all words can be included.

For each document $D_i (i = 1, \dots, n)$, we will use s_i to represent its sorted word sequence. Over all such sequences $S = \{s_1, \dots, s_n\}$, we will construct a digit tree, called a trie, as follows.

Assume that $W = \{w_1, \dots, w_m\}$. If $|S| = 0$, the trie is, of course, empty. For $|S| = 1$, $trie(S)$ is a single node. If $|S| > 1$, S is split into m (possibly empty) subsets $S_1, S_2, \dots, S_m$ so that a string is in S_j if its first word is w_j ($1 \leq j \leq m$). The tries $trie(S_1), trie(S_2), \dots, trie(S_n)$ are constructed in the same way except that at the k th step, the splitting of sets is based on the k th words in the sequences. They are then connected from their respective roots to a single node to create $trie(S)$. In Figure 1, we show a trie T constructed over the sorted word sequences in Table 1.

In the trie, v_0 is a virtual root, labeled with an empty word ε while any other node is labeled with a *real* word. Therefore, all the words on a path from the root to a leaf spell a sorted word sequence for a certain document. For instance, the path from v_0 to v_{13} corresponds to the sequence: *c, f, a, m, p*. Then, to check whether two words w_1 and w_2 are in the same document, we need only to check whether there exist two nodes v_1 and v_2 such that v_1 is labeled with w_1 , v_2 with w_2 , and v_1 and v_2 are on the same path. This shows that the reachability needs to be checked for this task, by which we ask whether a node v can reach another node u through a path. If it is the case, we denote it as $v \Rightarrow u$; otherwise, we denote it as $v \not\Rightarrow u$.

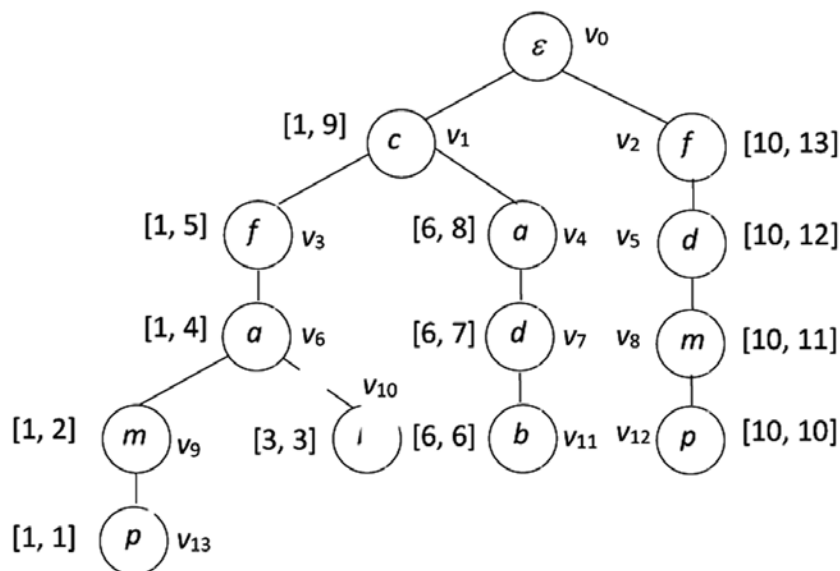
The reachability problem on tries can be solved very efficiently by using a kind of tree encoding, which labels each node v in a trie with an interval $I_v = [\alpha_v, \beta_v]$, where β_v denotes the rank of v in a post-order traversal of the trie. Here the ranks are assumed to begin with 1, and all the children of a node are assumed to be ordered and fixed during the traversal. Furthermore, α_v denotes the lowest rank for any node u in $T[v]$ (the subtree rooted at v , including v). Thus, for any node u in $T[v]$, we have $I_u \subseteq I_v$ since the post-order traversal enters a node before all of its children, and leaves after having visited all of its children. In Figure 1, we also show such a tree encoding on the trie, assuming that the children are ordered from left to right. It is easy to see that by interval containment we can check whether two nodes are on a same path. For example, $v_3 \Rightarrow v_{10}$, since $I_{v_3} = [1, 5]$, $I_{v_{10}} = [3, 3]$, and $[3, 3] \subset [1, 6]$; but $v_2 \not\Rightarrow v_9$, since $I_{v_2} = [10, 13]$, $I_{v_9} = [1, 2]$, and $[1, 2] \not\subset [10, 13]$.

Let $I = [\alpha, \beta]$ be an interval. We will refer to α and β as $I[1]$ and $I[2]$, respectively.

Lemma 1: For any two intervals I and I' generated for two nodes in a trie, one of four relations holds: $I \subset I'$, $I' \subset I$, $I[2] < I'[1]$, or $I'[2] < I[1]$.

Proof: The lemma can be directly derived from the definition of intervals.

Figure 1. A trie



However, more than one node may be labeled with the same word, such as nodes v_9 , and v_8 in Figure 1. Both are labeled with word m . Therefore, a word may be associated with more than one node (or say, more than one node's interval). Thus, to know whether two words are in the same document, multiple checking may be needed. For example, to check whether p and d are in the same document, we need to check v_{13} and v_{12} each against both v_7 and v_5 , by using the node's intervals.

In order to minimize such checks, we associate each word w with a word sequence of the form: $L_w = I_w^1, I_w^2, \dots, I_w^k$, where k is the number of all those nodes labeled with w and each $I_w^i = [I_w^i[1], I_w^i[2]]$ ($1 \leq i \leq k$) is an interval associated with a certain node labeled with w . In addition, we can sort L_w by the interval's first value such that for $1 \leq i < j \leq k$ we have $I_w^i[1] < I_w^j[1]$, which will greatly reduce the time for the reachability checking. We illustrate this in Figure 2, in which each word in Table 1 is associated with an interval sequence.

From this figure, we can see that for any two intervals I and I' in L_w we must have $I \not\subset I'$, and $I' \not\subset I$ since in any trie no two nodes on a path are labeled with the same word.

Finally, we point out that constructing a trie as above can be used as a general method for document classification. To see this, let's have a look at an example shown in Figure 3, which is part of an animal classification.

c :	$[1, 9]$
f :	$[1, 5]$ $[10, 13]$
a :	$[1, 4]$ $[6, 8]$
d :	$[6, 7]$ $[10, 12]$
b :	$[3, 3]$ $[6, 6]$
m :	$[1, 2]$ $[10, 11]$
p :	$[1, 1]$ $[10, 10]$

However, it is also a trie established using our method by checking the first several words in the corresponding document word sequences, where *animal* has the highest frequency while *mammalia*, *birds*, *fish*, and *amphibians* are less frequent.

2. Assignment of DocIDs to Intervals

Another important component of our index is to assign document identifiers to intervals. An interval I can be considered as a representative of

Figure 2. Sorted interval sequences

$c:$ [1, 9]
 $f:$ [1, 5][10, 13]
 $a:$ [1, 4][6, 8]
 $d:$ [6, 7][10, 12]
 $b:$ [3, 3][6, 6]
 $m:$ [1, 2][10, 11]
 $p:$ [1, 1][10, 10]

some words, i.e., all those words appearing on a *prefix* in the trie, which is a path P from the root to a certain node that is labeled with I . Then, the document identifiers assigned to I should be those containing all the words on P . For example, the words appearing on the prefix: $v_1 \rightarrow v_3 \rightarrow v_6$ in the trie shown in Figure 1 are words c, f , and a , represented by the interval [1, 4] associated with v_6 . So, the document identifiers assigned to [1, 4] should be {1, 2}, indicating that both documents D_1 and D_2 contain those three words. See the trie shown in Figure 4 for illustration, in which each node v is assigned a set of document identifiers that is also considered to be the set assigned to the interval associated with v .

Let v be the ending node of a prefix P , labeled with I . We will use $\delta(I)$, interchangeably $\delta(v)$, to

represent the set of document identifiers containing the words appearing on P . Thus, we have $\delta(v_6) = \delta([1, 4]) = \{1, 2\}$.

Lemma 2: Let u and v be two nodes in a trie T . If u and v are not on the same path in T , then $\delta(u)$ and $\delta(v)$ are disjoint, i.e., $\delta(u) \cap \delta(v) = \Phi$.

Proof: It is easy to prove.

Proposition 1: Assume that $v_1, v_2, \dots, v_j$ be all the nodes labeled with the same word w in T . Then, $\delta(w)$, the inverted list of w (i.e., the list of all the documents identifiers containing w) is equal to $\delta(v_1) \uplus \delta(v_2) \uplus \dots \uplus \delta(v_j)$, where $\uplus$ represents *disjoint union* over disjoint sets that have no elements in common.

Proof: Obviously, $\delta(w)$ is equal to $\delta(v_1) \cup \delta(v_2) \cup \dots \cup \delta(v_j)$. Since $v_1, v_2, \dots, v_j$ are labeled with the same word, they definitely appear on different paths as no nodes on a path are labeled with the same word. According to Lemma 2, $\delta(v_1) \cup \delta(v_2) \cup \dots \cup \delta(v_j)$ is equal to $\delta(v_1) \uplus \delta(v_2) \uplus \dots \uplus \delta(v_j)$.

As an example, see the nodes v_2 and v_3 in Fig. 5. Both are labeled with word f . So the inverted list of f is $\delta(v_2) \uplus \delta(v_3) = \{4\} \uplus \{1, 2\} = \{1, 2, 4\}$.

QUERY EVALUATION

Based on the new index structure, we design our algorithms. First, in Subsection 1, we introduce

Figure 3. Illustration for document classification

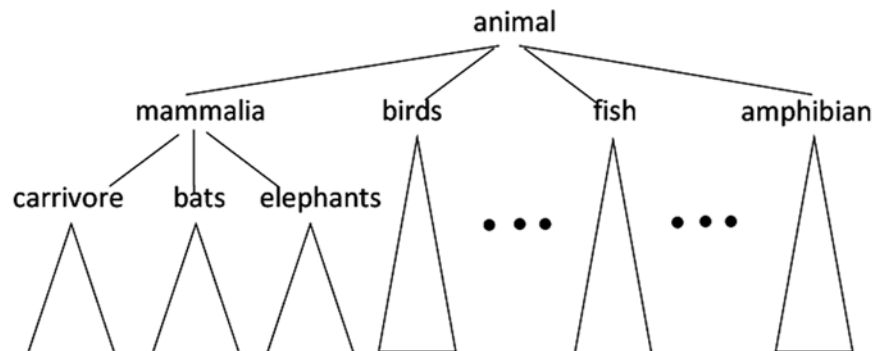
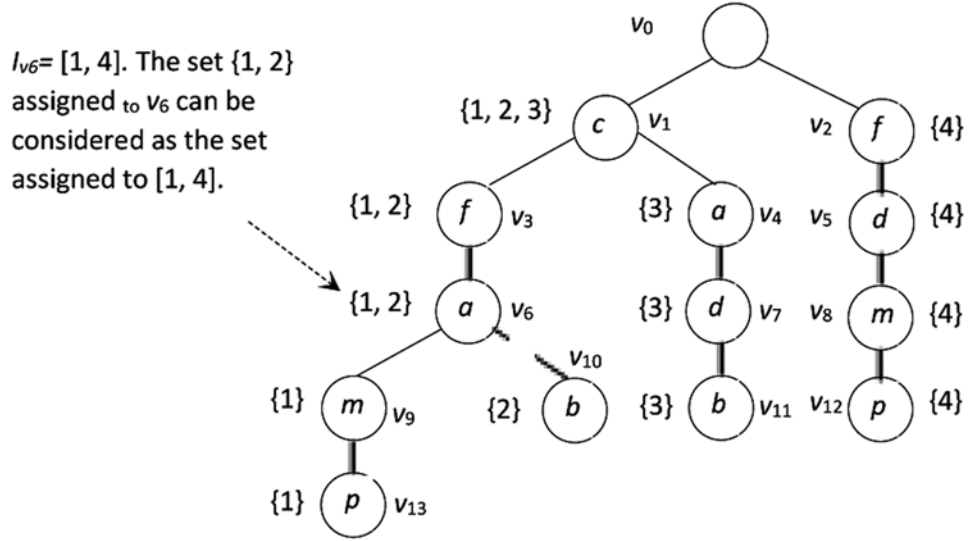


Figure 4. Illustration for assignment of document identifiers



the concept of word topological order, which is important to our query evaluation approach. Then, we present two algorithms to evaluate conjunctive and disjunctive queries in Subsections 2 and 3, respectively.

1. Word Topological Sequence

As shown below, using such interval sequences, the checking of whether two words are in the same document can be done in a very efficient way.

Definition 1: (Word topological order) Let $S = \{s_1, s_2, \dots, s_n\}$ be a set of n sorted word sequences. A word topological order over S is a sequence $\vartheta = w_1, w_2, \dots, w_m$, which contains all the words appearing in S such that for any two words w and w' if w appears before w' in some s_j ($1 \leq j \leq n$) then w appears before w' in ϑ , denoted as $w < w'$.

In Figure 2, the words are also listed (from top to bottom) in a word topological order with respect to the sorted word sequences given in Table 1. To find a word topological order over $S = \{s_1, s_2, \dots, s_n\}$ with $W = \{w_1, \dots, w_m\}$, we will transform the corresponding trie T to an acyclic directed graph

(DAG) G by splitting the node set of T (except for the virtual root) into m groups such that all the nodes in a group are labeled with the same word, and then collapsing each group g to a single node u . There is an edge in G from u (standing for a group g) to u' (for another group g') if T contains (x, y) with $x \in g$ and $y \in g'$. For example, the trie shown in Figure 1 will be transformed to a DAG shown in Figure 5(a).

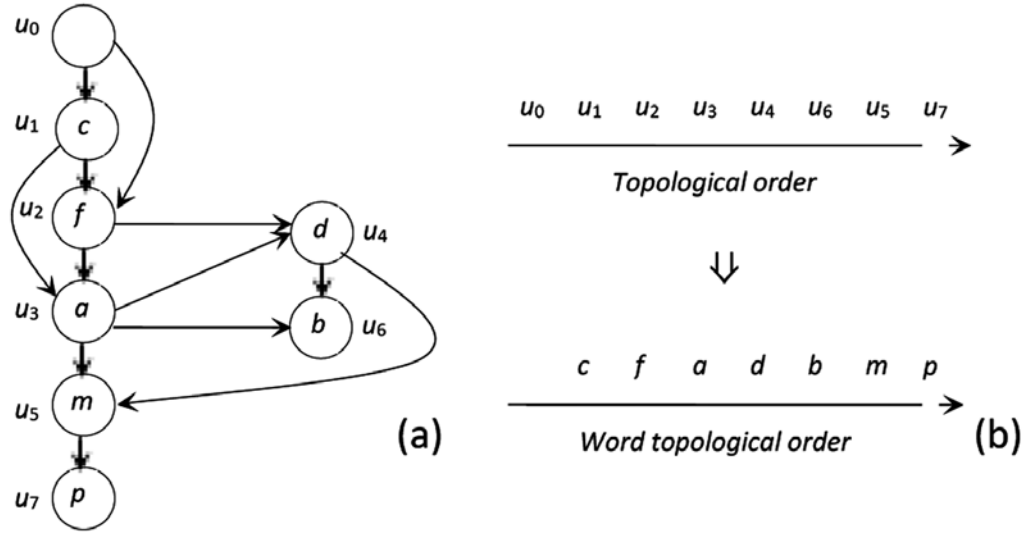
Using a hash function H on the words in W , the transformation can be done in $O(|W|)$ time, by which all those nodes labeled with the same word w will be mapped to a single node identified by $H(w)$.

Let $G(V, E)$ be such a DAG. It is well known that only $O(|V| + |E|)$ time is required to find a *topological order* of G , which is a linear ordering of all its nodes such that if $u \rightarrow v \in E$, then u appears before v in the ordering. Replacing each node in the ordering with the corresponding word, we will obtain a word topological sequence, as illustrated in Figure 5(b).

2. Evaluation of Conjunctive Queries

We first consider a query containing only two words $w \wedge w'$ with $w < w'$. It is easy to see that

Figure 5. A transformed DAG



any interval in L_w cannot be contained in any interval in $L_{w'}$. Thus, to check whether w and w' are in the same document, we need only to check whether there exist $I \in L_w$ and $I' \in L_{w'}$ such that $I \supset I'$. Therefore, such a query can be evaluated by running a process, denoted as $conj(L_w, L_{w'})$, to find all those intervals in $L_{w'}$ with each being contained in some interval in L_w , stored in a new sequence L .

- Let $L_w = I_1^1, I_2^2, \dots, I_k^k$. Let $L_{w'} = I_1^{k'}, I_2^{k'}, \dots, I_{k'}^{k'}$. $L \leftarrow \emptyset$.
- Step through L_w and $L_{w'}$ from left to right. Let I_w^p and $I_{w'}^q$ be the intervals currently encountered. We will do one of the following checking:

- If $I_w^p \supset I_{w'}^q$, append $I_{w'}^q$ to then end of L . Move to $I_{w'}^{q+1}$ if $q < k'$ (then, in a next step, we will check I_w^p against $I_{w'}^{q+1}$).
- If $I_w^p[1] > I_{w'}^q[2]$, move to $I_{w'}^{q+1}$ if $q < k'$. If $q = k'$, stop.
- If $I_w^p[2] < I_{w'}^q[1]$, move to I_w^{p+1} if $p < k$ (then, in a next step, we will check I_w^{p+1} against $I_{w'}^q$). If $p = k$, stop.

Assume that $L = I_1, I_2, \dots, I_l$ ($l \leq k'$). Then, for each $1 \leq l_j \leq l$, there exists an interval $I \in L_w$ such that $I_j \subset I$, and we can return $\delta(I_1) \uplus \dots \uplus \delta(I_l)$ as the result. In Figure 6, we illustrate the working process on L_d and L_m shown in Figure 2.

Figure 6. Illustration of two-word checking

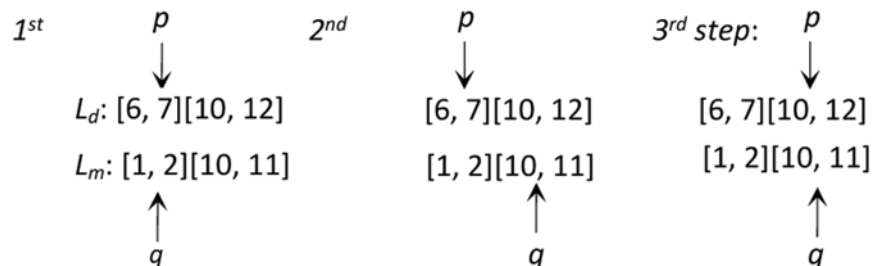
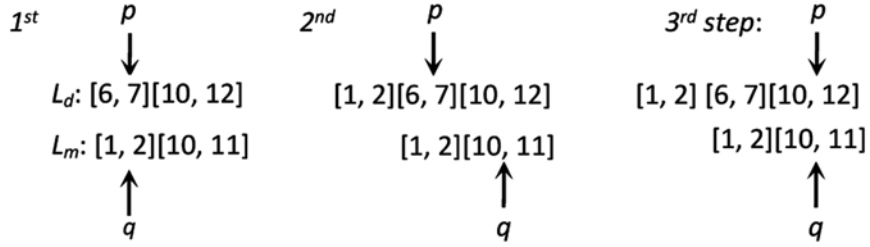


Figure 7. A merging process



Since in this process, each interval in both L_w and $L_{w'}$ is accessed only once, the time complication of this process is bounded by $O(|L_w| + |L_{w'}|)$.

The above approach can be easily extended to evaluate general queries of form $Q = w_1 \wedge w_2 \wedge \dots \wedge w_l$ with $w_1 < w_2 < \dots < w_l$ and $l \geq 1$ based on the transitivity of intervals:

$$I \supseteq I' \supseteq I'' \rightarrow I \supseteq I''.$$

What we need to do is to repeatedly apply *conj()* to the corresponding interval sequences associated with the query words one by one. Algorithm 1 is a formal description of this process.

It is easy to see that the time complexity of the algorithm is bounded by $O\left(\sum_{w \in Q} |L_w|\right)$, but can be further improved by using the binary, or galloping search, as well as the interpolation probing.

Example 2: Continued with Example 1. Let $Q = f \wedge m \wedge p$. Then, the execution of *containment**(Q)

Algorithm 1. Con-evaluation(Q)

```

begin
1.      let  $|Q| = l$ ; assume that
 $Q[1] < Q[2] < \dots < Q[l]$ ;
2.       $L := Q[1]$ ;
3.      for ( $j = 2$  to  $l$ ) do
4.      {  $L := \text{conj}(L, L_{Q[j]})$ ; }
5.      let  $L = I_1, \dots, I_k$ ;
6.      return  $\delta(I_1) \uplus \dots \uplus \delta(I_k)$ .
end
    
```

will find two containment sequences: $I_1 = [1, 5], [1, 2], [1, 1]$ and $I_2 = [10, 13], [10, 11], [10, 10]$. The results is then $R = \delta([1, 1]) \uplus \delta([10, 10]) = \{1\} \uplus \{4\} = \{1, 4\}$.

3. Evaluation of Disjunctive Queries

Similarly, we first consider a query containing only two words $w \vee w'$ with $w < w'$. For this, we define another process, denoted as *disj*($L_w, L_{w'}$), to merge L_w into $L_{w'}$, by which all those intervals I from L_w will be inserted into $L_{w'}$ if there is no interval in $L_{w'}$ is contained by I .

The operation can be efficiently conducted as follows.

Again, we will step through L_w and $L_{w'}$ from left to right. Let I_w^p and $I_{w'}^q$ be the intervals currently encountered. We will do the following checks:

1. If $I_w^p \supset I_{w'}^q$, move to $I_{w'}^{q+1}$ if $q < k'$. If $q = k'$, stop.
2. If $I_w^p[2] < I_{w'}^q[1]$, insert I_w^p into $L_{w'}$ just before $I_{w'}^q$. If $p < k$, move to I_w^{p+1} ; otherwise ($p = k$), stop.
3. If $I_{w'}^q[2] < I_w^p[1]$, move to I_w^{p+1} if $q < k'$. If $q = k'$, append $I_w^p, \dots, I_w^k$ to the end of $L_{w'}$ and then stop.

Example 3: Continued with Example 1. Let $Q = d \vee m$. We have $d < m$. By using the above procedure to merge $L_m = [1, 2][10, 11]$ into $L_d = [6, 7][10, 12]$, we will get a new sequence: $[1, 2][6, 7][10, 11]$. So, the result is

$\delta([1, 2]) \uplus \delta([6, 7]) \uplus \delta([10, 12]) = \{1\} \uplus \{3\} \uplus \{4\} = \{1, 3, 4\}$. In the first step, we compare $I_d^1 = [6, 7]$ and $I_m^1 = [1, 2]$. Since $I_d^1[1] = 6 > I_m^1[2] = 2$, I_m^1 will be inserted into L_d just before I_d^1 . Then, in the second step, we will compare I_d^1 and I_m^2 . Since $I_d^1[2] = 7 < I_m^2[1] = 10$, we will move to I_d^2 . Next, in the third step, we compare I_d^2 and I_m^2 , and find $I_d^2 \supset I_m^2$. Since I_m^2 is the last interval in L_m , we terminate the merging process and return the result.

Figure 7 shows the entire merging process.

By using the operation $disj(L_w, L_{w'})$, the algorithm to evaluate disjunctive queries of the form $Q = w_1 \vee w_2 \vee \dots \vee w_l$ with $w_1 < w_2 < \dots < w_l$ and $l \geq 1$ can also be efficiently implemented.

In Algorithm 2, we use $disj()$ to merge $L_{Q[i]}$ for $i = l - 1, \dots, 1$ into $L_{Q[1]}$ one by one. The running time is obviously bounded by $O(l \cdot r)$, where r is the largest number of intervals in all $L_{Q[i]}$'s which are not contained in each other. Again, the time requirement can be improved by using the binary search, the galloping search, and the interpolation probing.

CONCLUSION

In this chapter, a new index structure is discussed. It associates each word w with a sequence of intervals, which partition the inverted list $\delta(w)$ into a set of disjoint subsets. In this way, both the intersection (for conjunctive queries) and union (for disjunctive queries) of inverted lists can be done by checking the containment of intervals. This is much more efficient than the traditional inverted file method. Also, how to maintain such an index is described in great detail. Although the index is of a more complicated structure, the cost of maintaining it in the cases of addition and deletion of documents is (theoretically) comparable to the inverted file. Extensive experiments have

Algorithm 2. Dis-evaluation(Q)

```

begin
1.      let  $|Q| = l$ ; assume that
 $Q[1] < Q[2] < \dots < Q[l]$ ;
2.       $L := L_{Q[1]}$ ;
3.      for ( $i = l - 1$  downto 1)
do {
4.      {  $L := disj(L_{Q[i]}, L)$  };
5.      let  $L = I_1, \dots, I_k$ ;
6.      return  $\delta(I_1) \uplus \dots \uplus \delta(I_k)$ ;
end

```

been conducted, which show that our method outperforms the inverted file and the signature tree by an order of magnitude or more.

FUTURE WORK

As the future work, two issues need to be addressed: the binary search of interval sequences, and the use of *SIMD* (single instruction/multiple data) (Inoue et al., 2015) instructions. By the former, we will not scan an interval sequence linearly, but by the binary search. Especially, by using the *lowest common ancestors* (LCAs) of intervals (in a trie) in a binary search, the running time can be dramatically improved (Chen and Shen, 2015). By the latter, we will study how to arrange multiple processors to do the containment checking over multiple intervals simultaneously. That is, how to use SIMD instructions to do parallel checks of containments. In this way, the running time can also be significantly reduced.

REFERENCES

Anh, V. N., & Moffat, A. (2005). Inverted index compression using word-aligned binary codes, *Kluwer Int. Journal of Information Retrieval*, 8(1), 151–166. doi:10.1023/B:INRT.0000048490.99518.5c

- Ao, N., Zhang, F., Wu, D., Stones, D., Wang, G., Liu, X., ..., Lin, S. (2011). Efficient Parallel Lists Intersection and Index Compression Algorithms using Graphics Processing Units. Proceedings of the PVLDB '11, Seattle, USA.
- Apaydin, T., Canahuat, G., Ferhatosmanoglu, H., & Saman, T. A. (2006). Approximate Encoding for Direct Access and Query Processing over Compressed Bitmaps. Proceedings of VLDB '06 (pp. 846–857).
- Arasu, A., Cho, J., Garcia-Molina, H., Paepcke, A., & Raghavan, S. (2001, August). Searching the Web. *ACM Transactions on Internet Technology*, 1(1), 2–43. doi:10.1145/383034.383035
- Bjørklund, T. A., Grimsø, N., Gehrke, J., & Torbjørnsen, Ø. (2009). Inverted indexes vs. bit-map indexes in decision support systems. *CIKM, 2009*, 1509–1512. doi:10.1145/1645953.1646158
- Büttcher, S., & Clarke, C. L. A. (2005). Indexing time vs. query time trade-offs in dynamic information retrieval systems. *Proc. Intl. Conf. on Information and Knowledge Management*, Bremen, Germany (pp. 317-318).
- Carmel, D., Cohen, D., Fagin, R., Farchi, E., Hercovici, M., Maarek, Y. S., & Soffer, A. (2001). Static index pruning for information retrieval systems. *Proc. 24th Annual Intl. Conf. on Research and Development information Retrieval*, New Orleans, LA, USA (pp. 43-50). doi:10.1145/383952.383958
- Chen, Y. (2004). Building Signature Trees into OODBs. *Journal of Information Science and Engineering*, 20, 275–304.
- Chen, Y., & Chen, Y. B. (2006, September). On the Signature Tree Construction and Analysis. *IEEE TKDE*, 18(9), 1207–1224.
- Chen, Y., & Shen, W. (2015, July 27-30) On the Set Intersection. *Proceedings of the International Conference on Foundation of Computer Science*, Las Vegas, Nevada, USA (pp. 1007 – 1013).
- Demaine, K. D., López-Ortiz, A. and Munro, J. I. (2000). Adaptive set intersections, unions, and differences. *Proc. 11th ACM-SIAM Symposium on Discrete Algorithms*, Philadelphia (pp. 743-752).
- Ding, B., & König, A. C. (2011). Fast set intersection in memory. *Proc. of the VLDB Endowment*, 4(4), 255-266.
- Faloutsos, C. (1985). Access Methods for Text. *ACM Computing Surveys*, 17(1), 49–74. doi:10.1145/4078.4080
- Faloutsos, C., & Chan, R. (1988). Fast text access methods for optical and large magnetic disks: designs and performance comparison. *Proc. 14th Int'l Conf. Very Large Data Bases* (pp. 280-293).
- Golomb, S. W. (1966). Run-length encodings. *IEEE Trans. Inform. Theory*, 3(July), 399-401.
- Inoue, H., Ohara, M., & Taura, K. (2015). Faster set intersection with SIMD instructions by reducing branch mispredictions. *Proc. of VLDB* (pp. 293 – 304).
- Kitagawa, H., & Ishikawa, Y. (1997). False Drop Analysis of Set Retrieval with Signature Files. *IEICE Trans. Inf. & Syst.*, 80-D(6).
- Knuth, D. E. (1975). *The Art of Computer Programming* (Vol. 3). Massachusetts: Addison-Wesley Publish Com.
- Lovins, J. (1968). Development of a Stemming Algorithm. *Mechanical Translation and Computational Linguistics*, 11, 22–31.
- Saraiva, P. C. et al.. (2001). Rank-preserving two-level caching for scalable search engines. *Proc. 24th Annual Intl. Conf. on Research and Development in Information Retrieval*, New Orleans, LA, USA (pp. 51-58). doi:10.1145/383952.383959
- Tsirogiannis, D., Guha, S. and Koudas, N. (2009). Improving the Performance of List Intersection. *Proceedings of PVLDB '09*.

Zobel, J., & Moffat, A. (2006, July). Inverted Files for Text Search Engines. *ACM Computing Surveys*, 38(2), 1–56. doi:10.1145/1132956.1132959

ADDITIONAL READING

Chen, Y., & Chen, Y. B. (2008). An Efficient Algorithm for Answering Graph Reachability Queries. *Proc. 24th Int. Conf. on Data Engineering (ICDE '08)* (pp. 892-901). doi:10.1109/ICDE.2008.4497498

Chen, Y., & Chen, Y. B. (2011). Decomposing DAGs into spanning trees: A new way to compress transitive closures. *Proc. 27th Int. Conf. on Data Engineering (ICDE '11)* (pp. 1007-1018). doi:10.1109/ICDE.2011.5767832

Deppisch, U. (1986). S-Tree: A Dynamic Balanced Signature Index for Office Retrieval. *Proc. ACM SIGIR conf.* (pp 77 – 87). doi:10.1145/253168.253189

Willett, P. (2006). The Porter stemming algorithm: then and now. *Program: electronic library and information systems*, 40(3), 219 – 233.

Zobel, J., Moffat, A., & Ramamohanarao, K. (1998). Inverted Files Versus Signature Files for Text Indexing. *ACM Transactions on Database Systems*, 1998(4), 453–490. doi:10.1145/296854.277632

KEY TERMS AND DEFINITIONS

Interval: A pair of integers $[a, b]$ where b denotes the rank of v in a post-order traversal of a trie. Here the ranks are assumed to begin with 1, and all the children of a node are assumed to

be ordered and fixed during the traversal. In addition, a denotes the lowest rank for any node u in the subtree rooted at v . Its purpose is to check the reachability. Let $[a, b]$ and $[c, d]$ be the intervals associated with v and u , respectively. If $[c, d] \subset [a, b]$, then u is reachable from v through a path in the tree.

Inverted List: (Also referred to as postings file or inverted file) an index data structure associated with a key word w , storing a set of document identifiers, which contain w . Its purpose is to allow fast full text searches, at a cost of increased processing when a document is added to the database.

Search Engine: A software used by the Internet to search data (as text or a database) for specified information; also, a server on the World Wide Web that uses such software to locate key words in other sites.

Set Disjunction: (Also called set union, denoted as $A \cup B$, where A and B are two sets) the set of elements which are in A , in B , or in both A and B .

Set Intersection: (Denoted as $A \cap B$, where A and B are two sets) the set that contains all elements of A that also belong to B (or equivalently, all elements of B that also belong to A), but no other elements.

Signature File: A set of signatures (bit strings) with each created for a document by superimposing (bitwise OR) all the word signatures. To find all the documents that contain a set of key words, we will first generate a query signature by superimposing all the query word signatures and then search the signature file to find its matching ones.

Trie: (Also called digital tree and sometimes radix tree or prefix tree as they can be searched by prefixes) an ordered tree data structure that is used to store a dynamic set of strings like texts and documents.

Improving Usability of Website Design Using W3C Guidelines

G. Sreedhar

Rashtriya Sanskrit Vidyapeetha (Deemed University), India

INTRODUCTION

Over the last few years there has been a remarkable increase in use of the World Wide Web (WWW) for a wide and variety of purposes. There was also a fast growth in its applications. This led the Internet users to realize the importance and the benefits gained from a globally interconnected hypermedia system. On the other hand it causes a larger number of useless, meaningless and badly designed websites on the Internet world causing unwanted additional traffic; this is all because of an unorganized non-planned websites development processes. Due to the unceasing growth of web sites and applications, developers and evaluators have interesting challenges not only from the development but also from the quality assurance point of view.

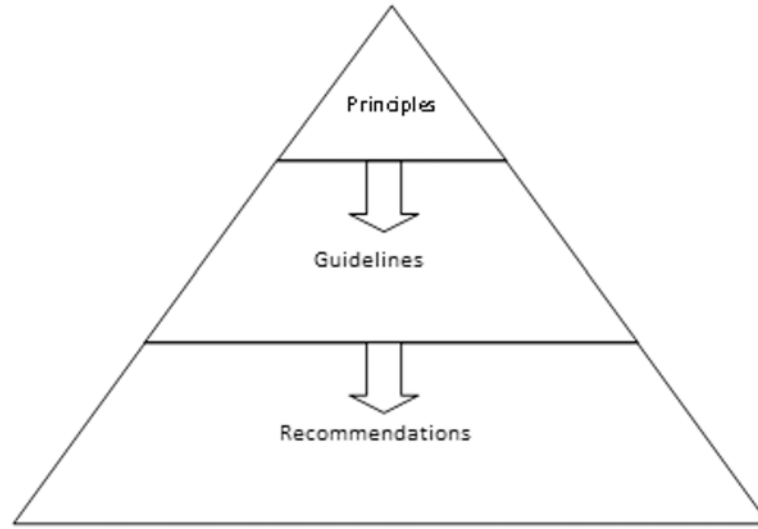
BACKGROUND

As we know, the quality assurance was and is one of the challenging processes in software engineering as well as for the web engineering, as a new discipline. Although there exists many design guidelines, and metrics for the evaluation of web sites and applications, most of them lack a well-defined specification framework and even worse a strategy for consultation and reuse. Some initial efforts have been recently made to classify metrics for some entity type as for example metrics for software products. Particularly, in last few years a set of web site metrics were defined and specified based on the data collection point

of view. The quality model must be able to assess the quality of each and every aspect of the website and it should cover the process of all web engineering activities. A set of guidelines are evolved to build a qualitative model of a website. According to Drefus P (1998) a guideline consists of a design and evaluation principle to be observed to get and to guarantee a usable user interface [1]. Guidelines can be found in many different formats with contents varying both in quality and level of detail, ranging from ill-structured common sense statements to formalized rules ready for automatic guidelines checking. Certain rules are validated by experimental results provided by user tests, experiments in laboratory or other techniques. Guidelines can be classified (Figure 1) by type ranging from the most general to the most specific: principles, guidelines and recommendations.

Principles are general objectives guiding conceptual User Interface (UI) decisions. They reflect the knowledge around human perception, learning and behavior and are generally expressed in generic terms like “Use images and metaphors consistent with real world” so that they can be applied for a wide range of cases. *Guidelines* are based on principles specific to a particular design domain. For example, a web design rule can stipulate to “use a consistent look and a visual language inside the site”. Some guidelines have to be interpreted more and altered to reflect the needs of a particular organization or a design case. *Recommendations* determine conceptual decisions specific to a particular domain of application and should reflect the needs and the terminology of a given organization. They are unambiguous state-

Figure 1. Website Guidelines and Sources



ments so that no place for interpretation is left. Recommendations include ergonomic algorithms, user interface patterns and design rules. Design rules are functional and operational requirements specifying the design of a particular interface, e.g. “Every web page needs an informative title”. Beirekdar A et al (2002) developed a framework to define a Guideline Definition Language (GDL) to investigate quality evaluation procedure. The GDL expresses guideline information in a sufficiently rich manner so that evaluation engine can perform GDL-compliant guideline.

$$\begin{aligned}
 U(p) &= f_{kwaremi}(Web_page, UES_{i,j}) \\
 &= EXEC(EC_{i,j} \{INST_UES_{i,j}\}) \\
 &= \{"Respected" | "Violated" | "Partially Respected"\}
 \end{aligned}
 \tag{1}$$

Where $UES_{i,j}$ be the set of evaluation sets associated to the guideline i in the source j and that will be used for the evaluation of the evaluated web page. $EC_{i,j}$ be the set of evaluation conditions associated to $UES_{i,j}$. $INST_UES_{i,j}$ is the set of captured instances of $UES_{i,j}$ in the evaluated page. In practice, the $f(Web_page, UES_{i,j})$ executes each $EC_{i,j}$ condition and then it combines the results to have the overall result for the guideline

i. We say that a web page satisfies a guideline $G_{i,j}$, if the execution of all $EC_{i,j}$ on all the $INST_UES_{i,j}$ is true. Using the above evaluation parameters allows us to define a kind of quality model to balance the evaluation result. In the accessibility field, Bobby, Valet, and EvalIris define a set of accessible evaluation tools. All these tools are based on accessibility guidelines. It does this through automatic checks as well as manual checks. It also analyzes web pages for compatibility with various browsers (eq. 3.2). Accessibility tools use a binary model to evaluate the accessibility of web pages (eq. 3.2).

$$Accessibility\ errors = \sum_{i=1}^{guidelines} a_i x_i \tag{2}$$

where a_i is 0 when guideline is violated and 1 when guideline is not violated and x_i is a guideline. A set of guidelines are considered to establish the procedure for Correctness of the Website. The World Wide Web Consortium (W3C) is an open source organizations and it defines various web standards for designing a website. The W3C is led by web inventor Tim Berners-Lee and CEO. The standards defined by W3C are considered as guidelines and these guidelines help in assessing

the quality of website content in presenting the web content. The guidelines are as summarized as follows.

Guideline 1: *Provide a text equivalent for every non-text element.* This includes images, graphical representations of text, image map regions, animations, applets and programmatic objects, frames, scripts, spaces, audio and video files.

Guideline 2: *Do not rely on color scheme only.* The content of web page must match with foreground and background color. Also provide sufficient contrast to the content for visibility.

Guideline 3: *Use markup and style sheets instead of images to convey information.* Style sheets controls the layout and presentation of the web page and decreases the download time of the web page.

Guideline 4: *Clearly mention the text information of web page with natural language.* Specify the expansion of each abbreviation or acronym in the document.

Guideline 5: *Use tables properly in the web document.* For data tables, clearly specify row and column headers and number of rows and columns exactly.

Guideline 6: *Ensure that web pages featuring new technologies transform gracefully.* When dynamic contents are updated, ensure that content is changed. Ensure that pages are available and meaningful when scripts, applets or other programmatic objects are not supported by the browsers. If this is not possible, provide equivalent information as alternative in the web page.

Guideline 7: *Ensure user control of time sensitive content changes.* Until user agents provide the ability to stop the refresh, do not create periodically auto-refreshing pages.

Guideline 8: *Ensure direct accessibility of embedded user interfaces.* Make programmatic elements such as scripts and applets directly

accessible or compatible with assistive technologies.

Guideline 9: *Design for device-independence.*

Ensure that any element that has its own interface can be operated in a device-independent manner.

Guideline 10: *Provide context orientation information.* Title each frame to facilitate frame identification and navigation. Divide large blocks of information into more manageable groups wherever appropriate.

Guideline 11: *Provide clear navigation mechanisms.* Clearly identify the target of each link. Provide information about the general layout of a site such as site map or table of contents.

Guideline 12: *Ensure that documents are clear and simple.* Create a style of presentation that is consistent across pages.

METHODOLOGY

In optimizing the correctness of website content different qualitative measures need to be investigated. These measures derived from the Web page errors that are generated using the W3C Validation Service. This process uses the standard web tool W3C HTML Validator to validate and identify the number of different errors according syntax errors of HTML tags, properties of web page and standards mentioned by various organizations such as W3C. Most pages on the World Wide Web are written in computer languages (such as HTML) that allow

Web authors to structure text, add multimedia content, and specify what appearance or style, the result should have. As for every language, these have their own *grammar*, *vocabulary* and *syntax*, and every document written with these computer languages are supposed to follow these rules. Markup languages are defined in *technical specifications*, which generally include a *formal grammar*. The tool compares HTML document

to the defined syntax of HTML and reports any discrepancies. The outputs of the Markup Validator are a list of error messages and their interpretation. W3C HTML Validator helps to ensure that documents are free of potential problems that can result in unexpected output when users view the bad documents with different browsers. A screenshot of W3C HTML Validator is shown in figure 2.

The errors related to website content cause incorrect display of some components of Web pages. These errors include:

1. **Table Tag Errors (TTE):** All the sub tags in table tag should be properly used in the web page design. Errors in table tag cause for display problems of web page.
2. **Body Tag Errors (BTE):** Body Tag Errors cause the errors in displaying the contents of the web page.
3. **Image Tag Errors (ITE):** Image Tag Errors cause for errors in downloading the image in a website.
4. **Head Tag Errors (HTE):** Head Tag Errors cause for errors in displaying heading and title of the web page.
5. **Font Tag Errors (FoTE):** Font Tag Errors cause the errors in textual display of the web page.
6. **Script Tag Errors (STE):** Script Tag Errors cause the errors in programming at client side scripting.
7. **Style Tag Errors (StTE):** Style Tag Errors cause errors in dynamic display features of the web page.
8. **Form Tag Errors (FmTE):** Form Tag Errors cause errors in input and output display of the script programming in a web page.

Figure 2. W3C HTML Validator

Browser	Internet Explorer	Firefox	Safari	Opera	Chrome	iOS	Android	BlackBerry
Version	6.0 7.0 8.0 9.0 10.0 11.0 12.0 13.0 14.0 15.0 16.0 17.0 18.0 19.0 20.0 21.0 22.0 23.0 24.0 25.0 26.0 27.0 28.0 29.0 30.0 31.0 32.0 33.0 34.0 35.0 36.0 37.0 38.0 39.0 40.0 41.0 42.0 43.0 44.0 45.0 46.0 47.0 48.0 49.0 50.0 51.0 52.0 53.0 54.0 55.0 56.0 57.0 58.0 59.0 60.0 61.0 62.0 63.0 64.0 65.0 66.0 67.0 68.0 69.0 70.0 71.0 72.0 73.0 74.0 75.0 76.0 77.0 78.0 79.0 80.0 81.0 82.0 83.0 84.0 85.0 86.0 87.0 88.0 89.0 90.0 91.0 92.0 93.0 94.0 95.0 96.0 97.0 98.0 99.0 100.0	1.0 2.0 3.0 4.0 5.0 6.0 7.0 8.0 9.0 10.0 11.0 12.0 13.0 14.0 15.0 16.0 17.0 18.0 19.0 20.0 21.0 22.0 23.0 24.0 25.0 26.0 27.0 28.0 29.0 30.0 31.0 32.0 33.0 34.0 35.0 36.0 37.0 38.0 39.0 40.0 41.0 42.0 43.0 44.0 45.0 46.0 47.0 48.0 49.0 50.0 51.0 52.0 53.0 54.0 55.0 56.0 57.0 58.0 59.0 60.0 61.0 62.0 63.0 64.0 65.0 66.0 67.0 68.0 69.0 70.0 71.0 72.0 73.0 74.0 75.0 76.0 77.0 78.0 79.0 80.0 81.0 82.0 83.0 84.0 85.0 86.0 87.0 88.0 89.0 90.0 91.0 92.0 93.0 94.0 95.0 96.0 97.0 98.0 99.0 100.0	1.0 2.0 3.0 4.0 5.0 6.0 7.0 8.0 9.0 10.0 11.0 12.0 13.0 14.0 15.0 16.0 17.0 18.0 19.0 20.0 21.0 22.0 23.0 24.0 25.0 26.0 27.0 28.0 29.0 30.0 31.0 32.0 33.0 34.0 35.0 36.0 37.0 38.0 39.0 40.0 41.0 42.0 43.0 44.0 45.0 46.0 47.0 48.0 49.0 50.0 51.0 52.0 53.0 54.0 55.0 56.0 57.0 58.0 59.0 60.0 61.0 62.0 63.0 64.0 65.0 66.0 67.0 68.0 69.0 70.0 71.0 72.0 73.0 74.0 75.0 76.0 77.0 78.0 79.0 80.0 81.0 82.0 83.0 84.0 85.0 86.0 87.0 88.0 89.0 90.0 91.0 92.0 93.0 94.0 95.0 96.0 97.0 98.0 99.0 100.0	1.0 2.0 3.0 4.0 5.0 6.0 7.0 8.0 9.0 10.0 11.0 12.0 13.0 14.0 15.0 16.0 17.0 18.0 19.0 20.0 21.0 22.0 23.0 24.0 25.0 26.0 27.0 28.0 29.0 30.0 31.0 32.0 33.0 34.0 35.0 36.0 37.0 38.0 39.0 40.0 41.0 42.0 43.0 44.0 45.0 46.0 47.0 48.0 49.0 50.0 51.0 52.0 53.0 54.0 55.0 56.0 57.0 58.0 59.0 60.0 61.0 62.0 63.0 64.0 65.0 66.0 67.0 68.0 69.0 70.0 71.0 72.0 73.0 74.0 75.0 76.0 77.0 78.0 79.0 80.0 81.0 82.0 83.0 84.0 85.0 86.0 87.0 88.0 89.0 90.0 91.0 92.0 93.0 94.0 95.0 96.0 97.0 98.0 99.0 100.0	1.0 2.0 3.0 4.0 5.0 6.0 7.0 8.0 9.0 10.0 11.0 12.0 13.0 14.0 15.0 16.0 17.0 18.0 19.0 20.0 21.0 22.0 23.0 24.0 25.0 26.0 27.0 28.0 29.0 30.0 31.0 32.0 33.0 34.0 35.0 36.0 37.0 38.0 39.0 40.0 41.0 42.0 43.0 44.0 45.0 46.0 47.0 48.0 49.0 50.0 51.0 52.0 53.0 54.0 55.0 56.0 57.0 58.0 59.0 60.0 61.0 62.0 63.0 64.0 65.0 66.0 67.0 68.0 69.0 70.0 71.0 72.0 73.0 74.0 75.0 76.0 77.0 78.0 79.0 80.0 81.0 82.0 83.0 84.0 85.0 86.0 87.0 88.0 89.0 90.0 91.0 92.0 93.0 94.0 95.0 96.0 97.0 98.0 99.0 100.0	1.0 2.0 3.0 4.0 5.0 6.0 7.0 8.0 9.0 10.0 11.0 12.0 13.0 14.0 15.0 16.0 17.0 18.0 19.0 20.0 21.0 22.0 23.0 24.0 25.0 26.0 27.0 28.0 29.0 30.0 31.0 32.0 33.0 34.0 35.0 36.0 37.0 38.0 39.0 40.0 41.0 42.0 43.0 44.0 45.0 46.0 47.0 48.0 49.0 50.0 51.0 52.0 53.0 54.0 55.0 56.0 57.0 58.0 59.0 60.0 61.0 62.0 63.0 64.0 65.0 66.0 67.0 68.0 69.0 70.0 71.0 72.0 73.0 74.0 75.0 76.0 77.0 78.0 79.0 80.0 81.0 82.0 83.0 84.0 85.0 86.0 87.0 88.0 89.0 90.0 91.0 92.0 93.0 94.0 95.0 96.0 97.0 98.0 99.0 100.0	1.0 2.0 3.0 4.0 5.0 6.0 7.0 8.0 9.0 10.0 11.0 12.0 13.0 14.0 15.0 16.0 17.0 18.0 19.0 20.0 21.0 22.0 23.0 24.0 25.0 26.0 27.0 28.0 29.0 30.0 31.0 32.0 33.0 34.0 35.0 36.0 37.0 38.0 39.0 40.0 41.0 42.0 43.0 44.0 45.0 46.0 47.0 48.0 49.0 50.0 51.0 52.0 53.0 54.0 55.0 56.0 57.0 58.0 59.0 60.0 61.0 62.0 63.0 64.0 65.0 66.0 67.0 68.0 69.0 70.0 71.0 72.0 73.0 74.0 75.0 76.0 77.0 78.0 79.0 80.0 81.0 82.0 83.0 84.0 85.0 86.0 87.0 88.0 89.0 90.0 91.0 92.0 93.0 94.0 95.0 96.0 97.0 98.0 99.0 100.0	1.0 2.0 3.0 4.0 5.0 6.0 7.0 8.0 9.0 10.0 11.0 12.0 13.0 14.0 15.0 16.0 17.0 18.0 19.0 20.0 21.0 22.0 23.0 24.0 25.0 26.0 27.0 28.0 29.0 30.0 31.0 32.0 33.0 34.0 35.0 36.0 37.0 38.0 39.0 40.0 41.0 42.0 43.0 44.0 45.0 46.0 47.0 48.0 49.0 50.0 51.0 52.0 53.0 54.0 55.0 56.0 57.0 58.0 59.0 60.0 61.0 62.0 63.0 64.0 65.0 66.0 67.0 68.0 69.0 70.0 71.0 72.0 73.0 74.0 75.0 76.0 77.0 78.0 79.0 80.0 81.0 82.0 83.0 84.0 85.0 86.0 87.0 88.0 89.0 90.0 91.0 92.0 93.0 94.0 95.0 96.0 97.0 98.0 99.0 100.0
Critical Issues	●	●	●	●	●	●	●	●
Major Issues	●	●	●	●	●	●	●	●
Minor Issues	●	●	●	●	●	●	●	●

Priority	URL / Description
●	This image is corrupt. This can be caused by the following: • running out of disk space quota during image file upload • transferring the image using FTP text mode • saving the image with the wrong extension (e.g. saving as JPEG, but using a .gif)
●	http://test.impressa.com/usa/base/hoofound.png
●	This link is broken. The page could not be found on the target web server.
●	This link is broken. The target anchor does not exist or is commented out.
●	This page contains some spelling errors. Click the word to add it to the dictionary or ignore it.
●	http://test.impressa.com/privacy.htm
●	This link requires a username and password.
●	5 issues on 17 pages

Browser	Internet Explorer	Firefox	Safari	Opera	Chrome	iOS	Android	BlackBerry
Version	6.0 7.0 8.0 9.0 10.0 11.0 12.0 13.0 14.0 15.0 16.0 17.0 18.0 19.0 20.0 21.0 22.0 23.0 24.0 25.0 26.0 27.0 28.0 29.0 30.0 31.0 32.0 33.0 34.0 35.0 36.0 37.0 38.0 39.0 40.0 41.0 42.0 43.0 44.0 45.0 46.0 47.0 48.0 49.0 50.0 51.0 52.0 53.0 54.0 55.0 56.0 57.0 58.0 59.0 60.0 61.0 62.0 63.0 64.0 65.0 66.0 67.0 68.0 69.0 70.0 71.0 72.0 73.0 74.0 75.0 76.0 77.0 78.0 79.0 80.0 81.0 82.0 83.0 84.0 85.0 86.0 87.0 88.0 89.0 90.0 91.0 92.0 93.0 94.0 95.0 96.0 97.0 98.0 99.0 100.0	1.0 2.0 3.0 4.0 5.0 6.0 7.0 8.0 9.0 10.0 11.0 12.0 13.0 14.0 15.0 16.0 17.0 18.0 19.0 20.0 21.0 22.0 23.0 24.0 25.0 26.0 27.0 28.0 29.0 30.0 31.0 32.0 33.0 34.0 35.0 36.0 37.0 38.0 39.0 40.0 41.0 42.0 43.0 44.0 45.0 46.0 47.0 48.0 49.0 50.0 51.0 52.0 53.0 54.0 55.0 56.0 57.0 58.0 59.0 60.0 61.0 62.0 63.0 64.0 65.0 66.0 67.0 68.0 69.0 70.0 71.0 72.0 73.0 74.0 75.0 76.0 77.0 78.0 79.0 80.0 81.0 82.0 83.0 84.0 85.0 86.0 87.0 88.0 89.0 90.0 91.0 92.0 93.0 94.0 95.0 96.0 97.0 98.0 99.0 100.0	1.0 2.0 3.0 4.0 5.0 6.0 7.0 8.0 9.0 10.0 11.0 12.0 13.0 14.0 15.0 16.0 17.0 18.0 19.0 20.0 21.0 22.0 23.0 24.0 25.0 26.0 27.0 28.0 29.0 30.0 31.0 32.0 33.0 34.0 35.0 36.0 37.0 38.0 39.0 40.0 41.0 42.0 43.0 44.0 45.0 46.0 47.0 48.0 49.0 50.0 51.0 52.0 53.0 54.0 55.0 56.0 57.0 58.0 59.0 60.0 61.0 62.0 63.0 64.0 65.0 66.0 67.0 68.0 69.0 70.0 71.0 72.0 73.0 74.0 75.0 76.0 77.0 78.0 79.0 80.0 81.0 82.0 83.0 84.0 85.0 86.0 87.0 88.0 89.0 90.0 91.0 92.0 93.0 94.0 95.0 96.0 97.0 98.0 99.0 100.0	1.0 2.0 3.0 4.0 5.0 6.0 7.0 8.0 9.0 10.0 11.0 12.0 13.0 14.0 15.0 16.0 17.0 18.0 19.0 20.0 21.0 22.0 23.0 24.0 25.0 26.0 27.0 28.0 29.0 30.0 31.0 32.0 33.0 34.0 35.0 36.0 37.0 38.0 39.0 40.0 41.0 42.0 43.0 44.0 45.0 46.0 47.0 48.0 49.0 50.0 51.0 52.0 53.0 54.0 55.0 56.0 57.0 58.0 59.0 60.0 61.0 62.0 63.0 64.0 65.0 66.0 67.0 68.0 69.0 70.0 71.0 72.0 73.0 74.0 75.0 76.0 77.0 78.0 79.0 80.0 81.0 82.0 83.0 84.0 85.0 86.0 87.0 88.0 89.0 90.0 91.0 92.0 93.0 94.0 95.0 96.0 97.0 98.0 99.0 100.0	1.0 2.0 3.0 4.0 5.0 6.0 7.0 8.0 9.0 10.0 11.0 12.0 13.0 14.0 15.0 16.0 17.0 18.0 19.0 20.0 21.0 22.0 23.0 24.0 25.0 26.0 27.0 28.0 29.0 30.0 31.0 32.0 33.0 34.0 35.0 36.0 37.0 38.0 39.0 40.0 41.0 42.0 43.0 44.0 45.0 46.0 47.0 48.0 49.0 50.0 51.0 52.0 53.0 54.0 55.0 56.0 57.0 58.0 59.0 60.0 61.0 62.0 63.0 64.0 65.0 66.0 67.0 68.0 69.0 70.0 71.0 72.0 73.0 74.0 75.0 76.0 77.0 78.0 79.0 80.0 81.0 82.0 83.0 84.0 85.0 86.0 87.0 88.0 89.0 90.0 91.0 92.0 93.0 94.0 95.0 96.0 97.0 98.0 99.0 100.0	1.0 2.0 3.0 4.0 5.0 6.0 7.0 8.0 9.0 10.0 11.0 12.0 13.0 14.0 15.0 16.0 17.0 18.0 19.0 20.0 21.0 22.0 23.0 24.0 25.0 26.0 27.0 28.0 29.0 30.0 31.0 32.0 33.0 34.0 35.0 36.0 37.0 38.0 39.0 40.0 41.0 42.0 43.0 44.0 45.0 46.0 47.0 48.0 49.0 50.0 51.0 52.0 53.0 54.0 55.0 56.0 57.0 58.0 59.0 60.0 61.0 62.0 63.0 64.0 65.0 66.0 67.0 68.0 69.0 70.0 71.0 72.0 73.0 74.0 75.0 76.0 77.0 78.0 79.0 80.0 81.0 82.0 83.0 84.0 85.0 86.0 87.0 88.0 89.0 90.0 91.0 92.0 93.0 94.0 95.0 96.0 97.0 98.0 99.0 100.0	1.0 2.0 3.0 4.0 5.0 6.0 7.0 8.0 9.0 10.0 11.0 12.0 13.0 14.0 15.0 16.0 17.0 18.0 19.0 20.0 21.0 22.0 23.0 24.0 25.0 26.0 27.0 28.0 29.0 30.0 31.0 32.0 33.0 34.0 35.0 36.0 37.0 38.0 39.0 40.0 41.0 42.0 43.0 44.0 45.0 46.0 47.0 48.0 49.0 50.0 51.0 52.0 53.0 54.0 55.0 56.0 57.0 58.0 59.0 60.0 61.0 62.0 63.0 64.0 65.0 66.0 67.0 68.0 69.0 70.0 71.0 72.0 73.0 74.0 75.0 76.0 77.0 78.0 79.0 80.0 81.0 82.0 83.0 84.0 85.0 86.0 87.0 88.0 89.0 90.0 91.0 92.0 93.0 94.0 95.0 96.0 97.0 98.0 99.0 100.0	1.0 2.0 3.0 4.0 5.0 6.0 7.0 8.0 9.0 10.0 11.0 12.0 13.0 14.0 15.0 16.0 17.0 18.0 19.0 20.0 21.0 22.0 23.0 24.0 25.0 26.0 27.0 28.0 29.0 30.0 31.0 32.0 33.0 34.0 35.0 36.0 37.0 38.0 39.0 40.0 41.0 42.0 43.0 44.0 45.0 46.0 47.0 48.0 49.0 50.0 51.0 52.0 53.0 54.0 55.0 56.0 57.0 58.0 59.0 60.0 61.0 62.0 63.0 64.0 65.0 66.0 67.0 68.0 69.0 70.0 71.0 72.0 73.0 74.0 75.0 76.0 77.0 78.0 79.0 80.0 81.0 82.0 83.0 84.0 85.0 86.0 87.0 88.0 89.0 90.0 91.0 92.0 93.0 94.0 95.0 96.0 97.0 98.0 99.0 100.0
Critical Issues	●	●	●	●	●	●	●	●
Major Issues	●	●	●	●	●	●	●	●
Minor Issues	●	●	●	●	●	●	●	●

Priority	URL / Description
●	Absolute CSS positioning can make pages unreadable when style sheets are turned off.
●	Document site must not be blank.
●	IMG tags must have an ALT attribute.
●	Add an ALT attribute describing each image, which screen readers read aloud. SpI should use ALT="".
●	http://test.impressa.com/products/yfftemplates.htm
●	Include a TITLE attribute for MAP elements.
●	People with low vision often turn off style sheets, so they won't see content that move the content from CSS to the HTML page so that all users can see it.
●	http://test.impressa.com/usa/base/textContent.asp
●	W3C HTML/XHTML Validation - Some pages fail validation.
●	W3C CSS Validation - All pages valid.
●	W3C Style Guide - No issues found.
●	W3C Deprecated Features - Issues found.
●	An attribute specification must start with a name or name token.
●	http://test.impressa.com/products/obastore.htm
●	http://test.impressa.com/products/berfestipub.htm
●	An attribute value specification must be an attribute value literal unless SHORTTAG Character 'E' is the first character of a delimiter but assumed as data.
●	This often occurs in text or JavaScript code when a bare single bracket is used instead
●	http://test.impressa.com/products/obastore.htm
●	http://test.impressa.com/products/berfestipub.htm
●	http://test.impressa.com/products/yfftemplates.htm
●	Document type does not allow element "div" here.

Priority	URL / Description
●	This image is corrupt. This can be caused by the following: • running out of disk space quota during image file upload • transferring the image using FTP text mode • saving the image with the wrong extension (e.g. saving as JPEG, but using a .gif)
●	http://test.impressa.com/usa/base/hoofound.png
●	This link is broken. The page could not be found on the target web server.
●	This link is broken. The target anchor does not exist or is commented out.
●	This page contains some spelling errors. Click the word to add it to the dictionary or ignore it.
●	http://test.impressa.com/privacy.htm
●	This link requires a username and password.
●	5 issues on 17 pages

9. **Link Tag Errors (LTE):** Link Tag Errors cause errors in linking various web components.

The web content errors are occurred due to non-standards of web site design. The developer must be attentive in using HTML tags so that appropriate tags should be used in web design process. All the tags and their attributes properly set and closed accordingly. This will reduce the problem of web page display and avoids the problem in downloading of the web page.

Website optimization is required in the following areas in order to improve the correctness of the website.

1. **Text Presentation:** Text Presentation is an important issue in display of the web content. These issues should be properly handled in presenting 100% correct text presentation. Several literature sources provide guidance about appearance of text on the web page. To format the text on web page the developer should consider following properties in text presentation.
 - Fonts must be chosen among the most readable ones.
 - Font size must be defined as relative size
 - Use fonts designed for computer screens rather than fonts designed for print
 - In a single page, the number of different fonts must be limited.
 - When using different fonts and font sizes, they should have some specific meaning (e.g. notes, links, and navigation location).
 - Avoid italicizing and underlining text.

These properties can be detected and measured by parsing both the text and CSS. The above properties are defined as attributes in various HTML tags. These tags include <HEAD>, <BODY> and tags. The errors in attributes of tag

errors cause incorrect or semi correct display of text on web page. The algorithm QAPM identifies the errors related to text presentation. Thus the Head Tag Errors (HTE), Font Tag Errors (FoTE) and Body Tag Errors (BTE) identify the problems in the text presentation of web page.

2. **Link Presentation:** The link presentation is important aspect in organization of web pages. The links in a website may be internal or external links. There are differences among the internal and external links. While internal links must all be valid and links pointing to external domains are out of control of the webmaster, but can be checked. Link topology is an often neglected aspect. Some sites are just trees of nodes, with links from a node pointing to children and to ancestors. Some others have a much more complex link topology, with many horizontal or traverse links. To format the links on the web page perform following functions.
 - Use moderate levels of breadth with minimal in the information architecture.
 - Minimize depth.
 - Avoid broken links.
 - Use corresponding text links.
 - Redundant links may cause confusion and avoid them.
 - Effective navigation requires small pages, few clicks between pages and strong scent.

Thus Link Tag Errors (LTE) and broken links identify the problems in link presentation.

3. **Page Layout:** The page layout is probably the principal characteristic perceived by the user. Layout must be clean, and the whole content should be well structured. A page layout is designed using tables, <div> tag or <frame> tag. Layout must adaptable to different devices. This implies that page must

avoid making reference to specific device settings, like screen resolution or fixed size page components. An automated analysis of CSS usage and coding can supply information about the layout and the adoption of an organization wide standard. The algorithm QAPM generates table tag errors (TTE), frame tag errors (FTE), style tag errors (StTE), font tag errors (FoTE), frame tag usage errors and document type declaration errors if any attribute of tag element deviate the properties of page formatting. The Page Formatting Measures assess the following features of website.

- Use browser-safe colors.
 - Use no more than 6 discriminable colors.
 - Use 256 (8-bit) color palettes.
 - Avoid using black backgrounds.
 - Use high contrast between background and text.
 - Keep line lengths to 40-60 characters
 - Keep text between 9 to 15 words per line.
 - Avoid using framesets.
 - Text should cover no more than 25-30% of the screen.
 - Greater text density facilitates page scanning.
4. **Graphics Presentation:** The graphic presentation is the important issue in presenting pictorial and multimedia components. *To format Graphics on the web page* perform following properties should be followed.
- Avoid using graphical text links.
 - Use corresponding text links instead of graphical links.
 - Avoid using animation unless it is appropriate.
 - Proper contrast between foreground image and background (color or image).

The image tag error (ITE), body tag errors (BTE) and image load errors related to image

identifies the errors in display of images and hence Graphic Element Measures to be evaluated. Graphics Element Measures developed for assessing the following features of web interfaces.

5. **Page Performance:** Correctness is a merely technical aspect, which can be easily checked. In many cases inconsistent behaviour with different browsers can be originated by lack of conformance to the published grammars (HTML, XHTML) and the default actions taken by the browsers themselves. Correctness is easily checked as an internal quality factor. The important aspects to consider in some environments are the professionalism and effectiveness of the web site that could be measured through how many different platforms are supported and it supports adaptivity and adaptability for a personalization Etnoteam et al (2000), M.Y. Ivory (2001). The form tag errors (FmTE), script tag errors (STE) and title tag with no keyword errors identify the need of page performance measure. The page performance measures developed to answer the following questions related to the website. To increase the performance of the web page following guidelines are considered.
- Minimize the use of video.
 - Avoid using sound files.
 - Effective navigation requires small pages.
 - Avoid using 'Click Here' for link text.

The contents of website must be presented in such a way that the user has to get 100% satisfaction in viewing all web pages of the website so that complete content of website visible and all pages are accessible. At this part of the research paper, it is tried to evolve 10-point scale. Thus 10-point scale is a metric towards defining quality of web content. In this connection it is interpreted the 10-point scale indicates such that '0' always represent poorer side and '10' always represent the best side of quality aspect. The 10-point scale metrics

for various qualitative measures are formulated using empirical evaluation. Here 10-point scale of web content depends on the value computed using measures and its level competence based on performance.

$$k = \left(\frac{n}{10} \right) * \text{No. of Pages with errors} \quad (3)$$

$$\text{Correctness} = (10 - k) \quad (4)$$

where n is number of Web Pages in a Website

Very Good if Correctness ≥ 9

Good if Correctness ≥ 8

Better than Average if Correctness ≥ 7

Average if Correctness ≥ 6

Below Average if Correctness ≥ 5

Poor if Correctness ≥ 4

Very Poor otherwise

(5)

Table 1. Quality Status of University Websites in India

S. No	University	No. of Pages With Errors (Per 100 Pages)	Correctness	Performance
1	English and Foreign Language University, Hyderabad	29	7.1	Better than Average
2	Moulana Azad National Urdu University, Hyderabad	93	0.7	Very Poor
3	University of Hyderabad, Hyderabad	96	0.4	Very Poor
4	Rajiv Gandhi Nagar University, Itanagar	14	8.6	Good
5	Assam University	38	6.2	Average
6	Tezpur University	45	5.5	Below Average
7	Central University of Bihar	87	1.3	Very Poor
8	Nalanda University	13	8.7	Good
9	Central University of Haryana	45	5.5	Below Average
10	Central University of Himachal Pradesh	8	9.2	Very Good
11	Central University of Jammu	12	8.8	Good
12	Central University of Kashmir	7	9.3	Very Good
13	Central University of Jharkhand	13	8.7	Good
14	Central University of Karnataka	19	8.1	Good
15	Central University of Kerala	87	1.3	Very Poor
16	Guru Ghasidas Viswavidyalaya, Bilaspur	17	8.3	Good
17	Dr. Harisingh Gour Viswavidyalaya, Sagar	54	4.6	Very Poor
18	Indira Gandhi National Tribal University	36	6.4	Average
19	Mahatma Gandhi Antharashtriya Viswavidyalaya	18	8.2	Good
20	Central Agricultural University, Manipur	22	7.8	Better than Average

EVALUATION

The study was conducted on 20 University websites using the methodology discussed in section 3 and the results are shown in Table 1.

CONCLUSION

The main theme of the chapter is to provide optimization techniques to improve the correctness of the website. It is observed that website must be informative and all contents of the website must be accommodated in page layout according to standard guidelines. An attempt is made to enhance the quality of website content and layout so that web designer shall follow the quality of content in designing a web site.

FUTURE RESEARCH DIRECTIONS

The chapter shows the way for the development of quality web designing in all aspects of websites. This is very much necessary to provide Quality of Service (QoS) to the online users. The chapter is well illustrated with a case study so that one can easily identify the defects in their own websites. The research work done in this chapter is the basis path for further extending the research in web engineering.

REFERENCES

- Beirekdar, A., Vanderdonckt, J., & Noirhomme-Fraiture, M. K. (2002). Knowledge based Web Automated Evaluation with reconfigurable guidelines optimization. LNCS, 2545, 362-376.
- Bobby. (n.d.). Retrieved from <http://webxact.watchfire.com/>

Dreyfus, P. (1998). *Principles of usability*. Retrieved from http://www.devedge.netscape.com/viewsource/arcive/editor_98_3_23.html

Etnoteam, S. P. A. (2000). Retrieved from <http://www.etnoteam.it/webquality>

EvalIris. (n.d.). Retrieved <http://www.sc.ehu.es/acwbbpke/evaliris.html>

Ivory, M. Y. (2001). *An Empirical foundation for Automated Web Interface Evaluation, doctoral dissertations, University of California*. Berkeley, CA: Computer Science Department.

Valet. (n.d.). Retrieved from <http://valet.webthing.com/access/url.html>

KEY TERMS AND DEFINITIONS

Correctness: A merely technical aspect, which can be easily checked. Correctness is easily checked as an internal quality factor. The important aspects to consider in some environments are the professionalism and effectiveness of the web site that could be measured through how many different platforms are supported and it supports adaptivity and adaptability for a personalization.

Graphics Presentation: The graphic presentation is the important issue in presenting pictorial and multimedia components.

Guidelines: Consists of a design and evaluation principle to be observed to get and to guarantee a usable user interface.

Link Presentation: The link presentation is important aspect in organization of web pages. The links in a website may be internal or external links.

Text Presentation: Text Presentation is an important issue in display of the web content. These issues should be properly handled in presenting 100% correct text presentation.

W3C HTML Validator: The outputs of the Mark up Validator are a list of error messages and their interpretation. W3C HTML Validator helps to ensure that documents are free of potential problems that can result in unexpected output when users view the bad documents with different browsers.

Website Content: The collection of web pages organized on a Web server. Web pages are of two types, static and dynamic.

Website Errors: Website errors related to website content cause incorrect display of some components of Web pages.

Internet Phenomenon



Lars Konzack

University of Copenhagen, Denmark

INTRODUCTION

Internet phenomenon is a new field of research. An internet phenomenon is an occurrence on the internet about somebody, a website, or a picture that for some reason captures the attention of numerous internet users and develops a craze that fast-spreads through the internet. The most common internet phenomenon is an internet meme, but also internet celebrities, political campaigns, or simply something out of the ordinary.

BACKGROUND

Before the internet there were the existence of folk tales and urban legends, folk songs and oral poetry as a way share content (Duggan, Haase, & Callow, 2016). Internet phenomena have often been compared to folklore and urban legends; however there is one significant difference in that folklore was passed on in an oral culture of illiterates. Internet on the other hand sharing are mostly done among 21st Century literates and often stored on servers for other people to see. In this sense sharing internet phenomena are closer to chain letters except the internet technology makes the process a lot easier and faster and may spread globally within minutes.

“Ideas are transmitted, often without critical assessment, across a broad array of minds and this uncoordinated flow of information is associated with “bad ideas” or “ruinous fads and foolish fashions.” (Jenkins, Ford, & Green, 2013, p. 307)

With 21st Century computer technology any idea or creation has the potential to spread like wildfire globally on the internet and if they do they become internet phenomena.

DOI: 10.4018/978-1-5225-2255-3.ch697

INTERNET MEMES

The most common internet phenomenon is an internet meme. The idea of memes takes it root in the memetics of Richard Dawkins but the concept of internet memes have evolved since then. Internet memes has become part of everyday life on the internet. Research has been done to understand this internet phenomenon as regards the development of internet memes, categorization of memes, and how they work.

An Internet meme is defined as a motif that is virally disseminated through the Internet. The motif often undergoes lots of variations (mash-ups) and may consist of sound, picture, movie clip, game and written text, or as is mostly the case, a combination by two or more modalities. Moreover the motif can be connected to only one of these modalities but need not be and in such case may enter different kinds of modalities.

It is difficult to pinpoint the first internet meme. One could argue that the emoticon introduced as the smiley in September 19th 1982 with all the variations of the theme is in fact the first internet meme (Rosenträger, 2008).

The term meme stems from Richard Dawkins controversial work *The Selfish Gene* referring partly to gene and partly to mimeme, which means to imitate. In his use of the term it is considered as any cultural idea or behavior such as fashion, language, religion, science and sports – cultural DNA reproducing itself (Dawkins, 1976). It is unclear whether Richard Dawkins comprehends the meme as an objective structure, or a metaphor for cultural practices. However, recent use of the term of internet meme has outgrown Richard Dawkins and has become a phenomenon in its own right (Stryker, 2011). According to Mole Empire the

ten most famous internet memes as of 2011 are as follows Keyboard Cat, Three Wolf Moon, Om Nom Nom, Auto Tune, The David After Dentist, Penaut Butter Jelly Time, Christian Bale Rant, Fail, O RLY, and Numa Numa (Smith, 2011). While this of course is by no means based on real academic research, it still gives a clue as to what these internet memes are.

A more systematic approach comes from Know Your Meme (<http://knowyourmeme.com/>), given that they try to accommodate a database of all known internet memes, and as of 2015 they have collected more than 13,000 meme entries of which at least 2,400 are confirmed, and they have categorized them as regards to confirmation status, what year it came to be, and where on the internet it originated. Wikipedia has descriptions of some of the most famous internet memes. An internet researcher may likewise find descriptions of internet memes on Oh Internet (<http://ohinternet.com>). On a far more chaotic scale it is possible to find information about internet memes on Encyclopedia Dramatica (<http://encyclopedia-dramatica.se>) although it requires skill to understand the in-jokes and to select the right bits of information and knowledge about internet memes. However, with the skill to comprehend Encyclopedia Dramatica, there is indeed information as regards to origins and explanations to a lot of these memes, information that may be difficult if not impossible to achieve by other means, and Encyclopedia Dramatica provides the right context and attitude for these memes, which the other catalogues do not.

Observation of how memes develop can be done, at websites such as YouTube, 4chan, 9gag, reddit, YTMND and Tumblr. While a lot of these memes originates from YouTube or 4chan the key to their success is the viral dissemination through e.g. E-mails, Facebook or Twitter. Memes can also be made by the use of so-called meme generators. These are internet services on which the user can upload images or use a wide range of ready-made images and put in a text-caption.

Since internet memes are a new phenomenon only few studies of have been made and they have all been experimental in their approach, some of which are presented here.

STUDYING MEMES

A study of Internet memes was conducted by Michele Knobel and Colin Lankshear (Knobel & Lankshear, 2007). They suggest an approach to memes based on three characteristics of memes that according to Richard Dawkins is the key to a successful meme: fidelity, fecundity, and longevity. Fidelity is how replicable the meme is. Fecundity is the dissemination speed. And longevity is the staying power of the meme. Furthermore the memes have been analyzed using three general axes: referential or ideational system, contextual or interpersonal system, and ideological or worldview system. During their research, they found that successful memes had three key components: humor, rich intertextuality (references to other works of art), and anomalous juxtaposition (mash-ups of deliberate provocative or off-guard instances of absurdity). Furthermore they made a categorization of internet memes primarily dividing into three groups: 1) social commentary purposes, 2) absurdist or humor purposes, and 3) otaku and manga fan purposes.

Another approach comes from Colin Stryker that conducted his own research of the meme life cycle that he lays out as seven stages (Stryker, 2011). The seven stages are as follows: birth, discovery, aggregation, word of mouth, blog pickup, mainstream exposure, commercialization, and death. It must be added that Colin Stryker says that not all memes goes through this exact life cycle. Some never become mainstream and other jump over certain stages, or follow a different path in order of sequence.

The birth of the meme is where and how it originated. A discovery of the meme could be that if someone posts the meme on a web-com-

munity like 4chan and is immediately picked up by the other users with mash-ups and comments. Aggregation begins when the meme jumps from the web-community that first discovered the meme to other web-communities like Reddit. Word of Mouth is actually more like dissemination through means of electronic written text and uploading images or movies, using blogs, tweets, status updates and instant messaging. It only reminds of actual word of mouth in the sense that it is distributed via informal means. Blog pickup happens when the specialized internet sites like 9gag, YTMND, Tumblr, and Know Your Meme receives and propagates the meme. It is at this stage the meme becomes a part of the general internet culture with self-referential jokes and clever mash-ups. Mainstream exposure of the meme occurs when old media like television, radio and newspapers distributes the story of the meme. It is at this stage the main stream population outside the internet culture learns about the internet meme, and it is paradoxically at that point where the inner circle of the internet culture may lose interest in the meme. Then there is only commercialization left (e.g. turning the meme into some sort of merchandise) before the meme is almost death, disappeared and forgotten.

In a study of YouTube memes conducted by Limor Shifman various meme features are recognized (Shifman, 2011). First of all Shifman discusses the difference between a viral video and a memetic video, in which a viral video is a video clip that spreads without significant change, while a memetic video is a popular video clip that 'lures extensive creative user engagement in the form of parody, pastiche, and other derivative work' (p. 190). The recognized Limor Shifman key features of YouTube memes are as follows: They relate to 'ordinary' people, and are often about flawed masculinity. They use humor, simplicity and repetitiveness, and portray whimsical content.

Patrick Davison argues that in order to understand internet memes, we will have to specify what exactly is being replicated (Davison, 2012).

According to Davison a meme can be separated into three components: manifestation, behavior, and ideal. The manifestation is the actualized as an observable, external phenomenon in time and space. The behavior is the action taken by users and producers of the meme. And the ideal of the meme is the idea conveyed as a conceptual purpose. The ideal determines the behavior which again resolves into the manifestation.

All of these approaches to internet memes demonstrate first of all that it is a new field that needs to be further studied but it also shows that these people take it very seriously and come up with ways to grasp and comprehend internet memes. They are trying to categorize them, understand how they function, how they develop over time, and how we should perceive them as cultural expression.

INTERNET CELEBRITIES

An internet celebrity is a celebrity known on the internet, and maybe even in old media as well, through at least one internet meme. In many cases this celebrity status has been notorious or ill-famed rather than the more positive kind of renown and fame. Due to the spreadability of internet memes, these meme celebrities have got lots of attention (Jenkins, Ford, & Green, 2013).

In November 2002, a Canadian high school student, Ghyslain Raza, made a video of himself playing with a golf ball retriever as if he was fighting like Darth Maul from Star Wars Episode 1: The Phantom Menace (Knobel & Lankshear, 2007). The video was uploaded to Kazaaby some of his classmates without his knowledge. The video became known as Star Wars Kid, went viral as a global hit, because people found it hilarious to see how a heavy-built kid was using all his energy to immerse himself into the role of a Sith Lord. The video was manipulated in different ways with sound and light effects, and lots of images were made based on the original video. Ghyslain Raza

W

had unknowingly become an Internet sensation. Ghyzlan became the laughing stock of millions and did not get anything else out of it (Stryker, 2011).

Numa Numa is the popular name of the music video *Dragostea din tei* by the Moldovan pop-group O-Zone, and it became notorious when overweight 19 year old Gary Brolsma in December 2004 lip synched to the video, waving his arms around. He later appeared in *Good Morning America*, and in 2006 he made his own website with merchandise and a remix of the original song (Stryker, 2011). Although Brolsma became laughed at he could at least cash in his fame.

David After Dentist is a video clip by his seven year old son David after his visit to the dentist, filmed by his father in May 2008. Seven months later he uploaded the video to YouTube (Prout, 2010). It became an instant sensation, and David was considered funny as well as cute. David along with his father appeared in *Today Show*, *Tyra Banks Show* and *The O'Reilly Factor*. The father detected a business opportunity, got a website, and made more video clips of his son, selling T-shirts and other merchandise. There has been controversy as to whether or not it was ethical for his father to exploit his own son like that (Linkins, 2009).

Another internet phenomenon was Boxxy. She made a couple of videos about herself to internet friends from a website called Gaia, and uploaded the videos to YouTube as BoxxyBabee in January 2008. This wasn't dramatic until 4chan came across the videos a year later and decided it was either the worst they had ever seen or the cutest girl on the internet. Memes were made in her honor as either cancer or as queen. She was portrayed as a girl version of Che Guevara and sold as T-shirts by meme merchandise internet shops. Little did she know what was going on until her mail account was hacked and all of her personal files was spread across the internet. Later she tried with mixed luck to cash in her sudden fame (Stryker, 2011).

Rebecca Black on the other hand tried to be famous by making a music video. On February 2011, the 13 year old girl uploaded the music video

to YouTube and it turned viral. And people hated it. The song *Friday* came to be the most hated song on the internet and viewed by millions. The ill-famed girl was ridiculed through countless of memes (Stryker, 2011). She did however make an appearance as Katy Perry's friend in Katy Perry's music video for *Last Friday Night (T.G.I.F.)*.

The face of Laina Morris became known as *Overly Attached Girlfriend* (Lagore, 2015). It was based on a video uploaded June 6th 2012 named *JB Fanvideo* for a Justin Bieber song competition. The starring eyes of Laina Morris and the song about an overly attached girlfriend made her face known to the internet. She has coped with the fame by posting other videos, and on March 11th 2013 she appeared on the show *Late Night with Jimmy Fallon*.

But not only people become famous. Products may become well-known as well. *Three wolf Moon* was a black T-shirt with three wolves howling at the Moon sold on Amazon. The T-shirt sold well but wasn't anything in particular, when suddenly a guy made a false review of the T-shirt stating that you would get all the girls you wanted if you wore it. The review was written in a humorous tone and others joined in, making their own reviews. After a while it spread across the internet with images superimposing the T-shirt to famous people and some music videos were made as well, and the T-shirt became a bestseller (Stryker, 2011).

For this reason or just for the fun of it many internet users have tried to create their own internet memes, hoping they will catch on. These are made on purpose and for the most part they fail. It is much more fun if the internet meme seems to come out of nowhere, like a coincidence or accident. These pushed memes are considered unfunny and are called forced memes. Within the depths of internet culture forced memes are frowned upon (Nissenbaum & Shifman, 2015).

Since internet phenomena have had such a tremendous impact, they have been used for propaganda purposes as well. Anonymous, the freedom fighters from 4chan, have used internet memes countless of times to spread their point of

view across the internet. On Project Chanology, their ongoing revolt against Scientology, they used the Guy Fawkes Mask (Stryker, 2011). It is correct that it originates from Guy Fawkes and the comic book and the later movie *V for Vendetta*. However, it was actually used at the time as a meme called Epic Fail Guy, which featured a dancing cartoon-like figure wearing the mask. The message to Scientology was that they were epic fail. They used other known memes like Rickrolling, and Longcat as well. As Anonymous progressed into support for WikiLeaks, Arab Spring, and the Occupy Movement it became a symbol of an internet youth rebellion. It became a sign of the times (Coleman, 2014).

Internet memes have been used in the American election for either candidate as far back as 2008, and have become a part of the campaign strategy in which each candidate produce forced memes to either a positive portrayal of their own candidate or a negative portrayal of the other candidate (Tay, 2015).

OTHER INTERNET PHENOMENONA

There are lots of internet phenomena and it is not possible to show all them here. This is just a few examples to show the range of diversity of internet phenomena.

Beany Babies have been said to be the first internet phenomenon. In the 1990s Beanie Babies attracted bidders on eBay more than any other item category. In May 1997, the average price for a Beany Baby was \$30, generating a total Beany Babies turnover of \$500.000 on eBay. By the end of the decade the sales plummeted to a more reasonable price around 50 cents apiece (Bissonnette, 2015).

Creepypasta are horror stories spread on the internet that are creepy. They are very close to urban legends except they are spread not through word-of-mouth but via internet as written or au-

diovisual communication that can be saved. The term creepypasta comes from a portmanteau of creepy and cospypasta – cospypasta meaning stories that are copied and pasted. The most common creepypasta is Slender Man which is a character of a tall man abducting children. It took the world by storm and made a lot of people scared or anxious. Slender Man was originally created by Erik Knudsen who in June 2014 apologized that it was taken too seriously (Delaney & Madigan, 2016).

Another internet phenomenon is Let's Play (or LP). LP was promulgated by the website Something Awful and is video footage of someone playing and commenting a videogame – often shared on YouTube or. The focus is not on the player but what happens in the game. It's a mixture of review, general commenting on the game and how it is played and advice on how to play the game. Most notably is PewDiePie, whose real name is Felix Arvid Ulf Kjellberg, with almost 50.000 subscribers (Berry, 2015).

Kony 2012 was a failed attempt at arresting Joseph Kony, an Ugandan war criminal, by the end of 2012. Kony 2012 was the title of a short documentary that was shared on the internet with over 100 million views. Although the video got a lot of attention and a resolution condemning the terror of Joseph Kony from American senators, and even President Obama offering a reward of \$5 million for his arrest, Joseph Kony has still not been captured (Mahoney & Tang, 2016).

In 2014, Ice Bucket Challenge arose among people to donate money to amyotrophic lateral sclerosis (ALS). They should take the challenge of having ice cold water poured over them and challenge three others to donate and be poured over by cold water. This chain letter approach meant that a lot of people came in contact with the Ice Bucket Challenge and eventually a lot of celebrities participated. The Ice Bucket Challenge was often shared as YouTube videos on the internet and raised \$114 million dollars (Hutchins & Tindall, 2016).

FUTURE RESEARCH DIRECTIONS

There is still lot to be done as far as future research in the field of internet meme, and we have yet to see their cultural, sociological and political impact. Internet phenomena are used, reproduced and part of a participating internet culture. One of the most interesting future research directions would be to see the development of political internet memes and the link between geek culture and popular culture in relations to internet phenomena. There is a lot of internet phenomena that cannot be analyzed by a single individual. Future research will demand digital humanities solutions to delve into the many aesthetic and cultural aspects of internet phenomena.

CONCLUSION

As we have seen internet phenomena can be analyzed as far as categorization, functionality, development, and as a socio-cultural phenomenon. Research has shown how internet memes are born; how they develop over time; and how they eventually die. It has shown how internet memes may be categorized into three major subgroups: 1) social commentary purposes, 2) absurdist or humor purposes, and 3) otaku and manga fan purposes. An internet meme consists of three components: 1) manifestation, 2) behavior, 3) and ideal. The spreadability of internet phenomena means that indeed internet celebrities have arisen, and we have seen some examples of these. Furthermore a range of different internet phenomena has spawned and disappeared again.

REFERENCES

Berry, J. (2015). *PewDiePie: The Ultimate Unofficial Fan Guide to The World's Biggest Youtuber*. London: Hachette.

Bissonnette, Z. (2015). *The Great Beanie Baby Bubble: Mass Delusion and the Dark Side of Cute*. New York, NY: Penguin.

Blackmore, S. (1999). *The Meme Machine*. Oxford, UK: Oxford University Press.

Coleman, G. (2014). *Hacker, Hoaxer, Whistle-blower, Spy: The Many Faces of Anonymous*. New York, NY: Verso.

Culture, L. L. (n.d.). Tim Delaney. *Tim Madigan*.

Davison, P. (2012). The Language of Internet Memes. In M. Mandiberg (Ed.), *The Social Media Reader* (pp. 120–134). New York: New York University Press.

Dawkins, R. (1976). *The Selfish Gene*. Oxford, UK: Oxford University Press.

Delaney, T., & Madigan, T. (2016). *Lessons Learned from Popular Culture*. Albany, NY: SUNY Press.

Duggan, A. E., Haase, D., & Callow, H. J. (2016). *Folktales and Fairy Tales: Traditions and Texts from around the World, 2nd Edition: Traditions and Texts from around the World*. Santa Barbara, CA: ABC-CLIO.

Hutchins, A., & Tindall, N. (2016). *Public Relations and Participatory Culture: Fandom, Social Media and Community Engagement*. New York, NY: Routledge.

Jenkins, H., Ford, S., & Green, J. (2013). *Spreadable Media: Creating Value and Meaning in a Networked future*. New York: New York University Press.

Knobel, M., & Lankshear, C. (2007). Online Memes, Affinities, and Cultural Production. In M. Knobel & C. Lankshear (Eds.), *A New Literacy Sampler* (pp. 199–227). New York: Peter Lang Publ.

Lagore, J. (2015). Self-Promotion for All! Content Creation and Personal Branding in the Digital Age. In D. C. Sarver & S. Collister (Eds.), *Debates for the Digital Age: The Good, the Bad, and the Ugly of Our Online World* (pp. 221–240). Santa Barbara, CA: ABC-CLIO.

Linkins, J. (2009, July 9). Bill O'Reilly Goes After Father Of 'David After Dentist' For Some Reason. *Huffington Post*. Retrieved from http://www.huffingtonpost.com/2009/06/08/bill-oreilly-goes-after-f_n_212635.html

Mahoney, L. M., & Tang, T. (2016). *Strategic Social Media: From Marketing to Social Change*. Oxford, UK: John Wiley & Sons. doi:10.1002/9781119370680

Nissenbaum, A., & Shifman, L. (2015, October 9). Internet memes as contested cultural capital: The case of 4chan's /b/ board. *New Media & Society*, 1–19.

Prout, S. (2010). *The Power of Influence: The Easy Way to Make Money Online*. Milton: Wrightbooks.

Rosenträger, S. (2008). *Emoticons as a New Means of Communication in Italy and Germany*. Munich: GRIN Verlag.

Shifman, L. (2011, October 3). An anatomy of a YouTube meme. *New Media & Society*, 14(2), 187–203. doi:10.1177/1461444811412160

Smith. (2011, December 14). The History Of The Most Famous Internet Memes So Far. *Mole Empire*. Retrieved from <http://molempire.com/2011/12/14/the-history-of-the-most-famous-internet-memes-so-far/>

Stryker, C. (2011). *Epic Win for Anonymous: How 4chan's Army Conquered the Web*. London: Overlook Duckworth.

Tay, G. (2015). Binders full of LOLitics: Political humour, internet memes, and play in the 2012 US Presidential Election (and beyond). *The European Journal of Humour Research*, 2(4), 46–73. doi:10.7592/EJHR2014.2.4.tay

ADDITIONAL READING

Bauckhage, C. (2011, May). Insights into Internet Memes. In ICWSM. <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/download/2757/3304>

Burgess, J. (2008). All Your Chocolate Rain Are Belong to Us. Viral Video, Youtube and the Dynamics of Participatory Culture. In G. Lovink & S. Niederer (Eds.), *Video Vortex Reader: Responses to YouTube* (pp. 101–109). Amsterdam: Institute of Network Culture.

Dibbell, J. (2010). Radical opacity. *Technology Review*, 113(5), 82–86.

Guadagno, R. E., Rempala, D. M., Murphy, S., & Okdie, B. M. (2013). What makes a video go viral? An analysis of emotional contagion and Internet memes. *Computers in Human Behavior*, 29(6), 2312–2319. doi:10.1016/j.chb.2013.04.016

McConnell, B., & Huba, J. (2007). *Citizen marketers: When people are the message*. Berkshire: Kaplan Pub.

Shifman, L., & Thelwall, M. (2009). Assessing global diffusion with Web memetics: The spread and evolution of a popular joke. *Journal of the American Society for Information Science and Technology*, 60(12), 2567–2576. doi:10.1002/asi.21185

Wang, L., & Wood, B. C. (2011). An epidemiological approach to model the viral propagation of memes. *Applied Mathematical Modelling*, 35(11), 5442–5447. doi:10.1016/j.apm.2011.04.035

Zimmer, B., & Carson, C. E. (2011). Among the new words. *American Speech*, 86(4), 454–479. doi:10.1215/00031283-1587259

KEY TERMS AND DEFINITIONS

Internet Meme: A motif that is virally disseminated through the Internet. The motif often undergoes lots of variations (mash-ups) and may consist of sound, picture, movie clip, game and

written text, or as is mostly the case, a combination by two or more modalities. Moreover the motif can be connected to only one of these modalities but need not be and in such case may enter different kinds of modalities.

Internet Phenomenon: An occurrence on the internet about somebody: a website, or a picture that for some reason captures the attention of numerous internet users and develops a craze that fast-spreads through the internet.

Manga: Japanese comics style.

Otaku: A Japanese term for obsessive collector – often manga-related items. It is commonly used as the Japanese term for a geek or nerd.

Spreadability: The ability to disseminate information.

An Overview of Crowdsourcing



Eman Younis

Minia University, Egypt

INTRODUCTION

During the past decade, there were rapid developments in the Internet, computing technologies, wide-spread and use of location-aware technologies such as GPS and mobile phones.

These developments influenced how people communicate and share their opinions, views, knowledge, maps, and many others throughout software platforms. These technologies have participated in the creation of what is now called Web 2.0, which is a new era of web technologies enabling users to play an active role in adding contents to the web in a collaborative way, instead of just consuming the web contents.

People are now easily sharing social media posts, blog posts, product reviews, ideas, opinions, maps and others. Crowdsourcing is characterized as a phenomenon that appeared due to enabling web users to contribute to the web.

There are many online Crowdsourcing applications such as Amazon Mechanical Turk, Open Street Map, and Yahoo Answers amongst others. Thus, crowdsourcing is a collaborative process which involves four main components (requester, crowd, open call and platform). People might benefit financially or intellectually from participation in crowdsourcing. This chapter serves as a general overview of crowdsourcing research and envisages future research directions.

BACKGROUND

Due to the invasion of pervasive computing and Web 2.0 technologies (Brabham, 2013), users can not only consume information on the web

by search and navigation but also, contribute their own contents. Crowdsourcing is also called fan sourcing, crowd casting, mass collaboration, socio-technical systems (Geiger, D. et al. 2012), collective intelligence, smart mobs, peer production, citizen science and user generated content (Haythornthwaite, 2009).

Many applications have been created to make use of the online crowds such as Question Answering, Mapping, Software Engineering, Healthcare and many more. Examples of Crowdsourcing systems are yahoo answers, Open Street Map, Amazon Mechanical Turk, among others. Crowdsourcing can be classified according to (Howe, J. 2008) into co-creation, crowd creation, crowd voting, crowd wisdom, and crowdfunding.

There is always confusion between the terms crowdsourcing and outsourcing. Here we distinguish between these terms. Wikipedia has created a debate; some researchers do not consider it a crowdsourcing site. Because it is not satisfying their 8 criteria listed in (Estellés-Arolas et al. 2012). On the other hand (Haythornthwaite, 2009) considers Wikipedia a crowdsourcing platform, contributed by a collection of volunteer contributors.

This chapter presents a general overview of the crowdsourcing field. It is structured as follows; it begins with the analysis of definitions of the term from literature followed by a classification of crowdsourcing activities. Then, a comparison between crowdsourcing and outsourcing is presented. After that a discussion of crowd-user motivations is presented. It also presents the benefits and problems of crowdsourcing, its applications, online crowdsourcing systems, current and future research avenues.

CROWDSOURCING DEFINITIONS

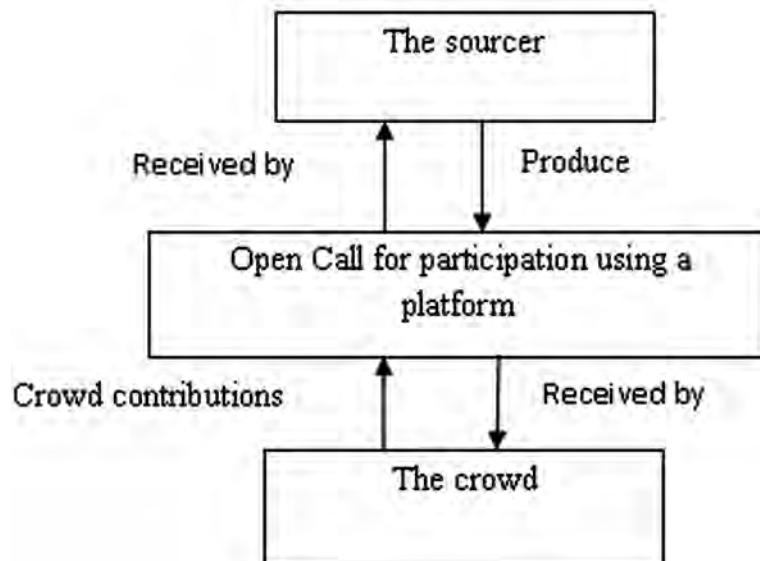
Crowdsourcing is not a new phenomenon; it has been used in the past in many cases for collecting users' participations (Howe, J. 2008). But, until now there is no standard agreed upon definition of crowdsourcing. The term was first proposed by Jeff Howe's in a Wired Magazine article (Howe, J. 2008) and defined as:

the act of a company of institution taking a function once performed by employees and outsourcing it to an undefined and generally large network of people in the form of open call.

After that, there were several other attempts to define it in the literature such as (Ambati et al. 2012 Azzam, T. et al. 2013, Von Ess 2010, Bell, 2009, Doan et al. 2011, Lebraty et al. 2013, Brabham, 2013, Estellés-Arolas et al. 2012, Sharma et al. 2014 and many more). These definitions are different in their view regarding the targeted application of crowdsourcing. The definitions of Crowdsourcing are unified in terms of the general concepts and components, but, targeting different applications.

Figure1 depicts the crowdsourcing process, showing various components. The first component is the sourcer or the requester, which can be an organization or an individual. The requester wants to accomplish a specific goal or solve a problem. This goal can be to create, test, or rate a product or a service or it can even be to collect capital investment to initiate a project. The second component is an open call (this is the call for participants to contribute to achieving the requester's goal). The third component is the crowd (the anticipated contributors to the open call). The last component is the platform that facilitates the process of interaction between the requester and the crowd. Figure1 depicts the crowdsourcing process, showing various components. It shows that the crowdsourcing process is initiated by the requester, which may be a company, an institution or an individual, having some problem to solve. This requester sends an open call for participations to the crowd. The diversity and heterogeneity of the crowd capabilities are an advantage, which produces variations of the presented solutions. The crowd participates in the open call by providing solutions to the problem. The requester receives and evaluates the presented solutions.

Figure 1. The Crowdsourcing process



CLASSIFICATION OF CROWDSOURCING ACTIVITIES

Crowdsourcing activities can be classified into the following categories (Howe, J. 2008):

- **Crowd Wisdom:** There are some tasks which require human intelligence, such as text or image classification, tagging or annotation. The crowdsourcing system collects several human-annotated examples from the crowd to be used for computer learning. In such applications, the crowd wisdom has been used to help the computer solving a hard problem. As an example, Amazon Mechanical Turk (www.mturk.com) is offering paid work for the crowd to solve human Intelligence tasks (HITs) called micro-tasks. Another example is (<https://www.google.com/recaptcha>), where users are acting voluntarily to support Google performing intelligent tasks such as text and image recognition automatically. These kinds of tasks are essential for Machine Learning applications.
- **Crowd Creation:** In this type of crowdsourcing, the crowd is asked to contribute some kind of artifacts, designs, maps or even a question and its answer. For example (threadless.com), is a company offering its customers a share of profit for the best T-shirt design. Another example in mapping is (Open Street Map), which is an online free mapping platform, where anyone can share their own maps or use existing maps. It is mainly based on the volunteer contributions from the crowd. Moreover, Question Answering Services such as (<https://answers.yahoo.com/>) are depending on the crowd for creating questions and answers. Wikipedia is another example of crowd creation tasks, where the crowd collaborates to create an encyclopedia.
- **Crowdfunding:** It is a recent phenomenon depending on collecting financial funding

for establishing a project or financing an idea. The idea behind its success is finding the interested crowd in the suggested project or idea. An example is (sellaband.com) asking for donations from their crowd-fans to sponsor their music albums.

- **Crowd Opinion or Crowd Rating:** In these systems, the crowd worker is required to provide his/her feedback or opinion in an artifact, design or even rate a piece of software. This type of systems can be used to evaluate the outputs of the previous types of systems using the crowd opinions.
- **Crowd Solving:** This type of crowdsourcing usually involves a contest for the crowd to solve a specific problem. For example, (www.kaggle.com) is an online crowdsourcing system for scientific problem solving. It encourages participants by giving financial prizes for the best solutions. Another example in the scientific field is (www.researchgate.net), where any researcher can send his problem and receive answers from his/her peers.

Although Howe, J. 2008 first presented crowdsourcing as an outsourcing activity, there are potential differences between those two concepts. Despite being aiming to get a task or a job done by people outside the organization boundary, there are still differences between the two terms. The next section shows the differences between them.

CROWDSOURCING VERSUS OUTSOURCING

Although Jeff How's definition of Crowdsourcing presented earlier as an outsourcing activity, outsourcing is a different concept. Outsourcing is the process of hiring an external company to provide a product or a service. Here we want to clarify that outsourcing is different from crowdsourcing in many aspects. First, outsourcing is the process throughout which the company contract

with other external company to provide a product or a service, whereas in the crowdsourcing process the company sends an open call of participations to the crowd. Thus, the differences between both processes involve three points (Sharma and Padmanaban, 2014):

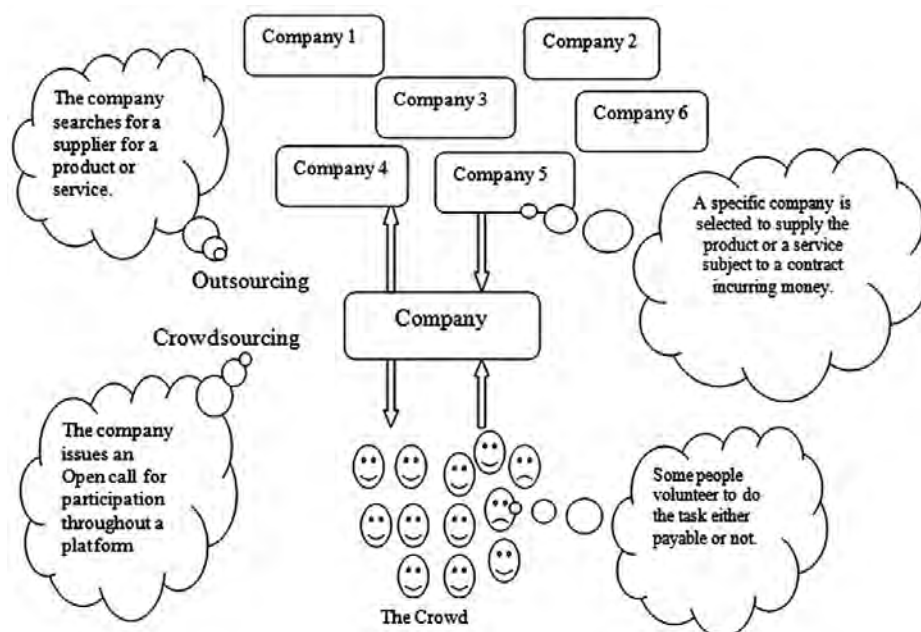
1. **Contract:** In the outsourcing process, there is a contract between the main company and the specified outsourcing company. Whereas, in the crowdsourcing process, there is no contract between the sourcer and the crowd.
2. **Cost:** The contract entitles the sourcer company to pay a pre-determined amount of money according to the contract. The crowdsourcing could involve volunteering work from the crowd or pay a nominal amount of money.
3. **Commitment:** The company involved in the outsourcing is committed to abiding by all the contract conditions by delivering product or service on time and specifications required. But, the crowd is free of any obligations to participate or give good quality outputs.

Figure 2 shows the difference between crowdsourcing and outsourcing.

USER MOTIVATIONS FOR CONTRIBUTION

It is important to know why people are engaged in crowdsourcing activities. Motivations are the stimulating factors that encourage people to do actions. Ryan et al. 2000 presented the crowd motivations to participate to Intrinsic and extrinsic motivations. The intrinsic motivations are defined as those leading to enjoyment or any type of satisfaction for the participant. Extrinsic motivations are those triggered by a separate outcome such as a reward. Pilz et al. 2013 discussed different incentives of participants in crowdsourcing. They further divided the user incentives more fine-grained classes out of the essential intrinsic and extrinsic proposed by (Ryan et al. 2000). They further divided the intrinsic motivations into enjoyment-based containing (skill variety, task identity, task autonomy, time killing and feedback) and community- based including (community

Figure 2. Crowdsourcing versus outsourcing



identification and social contact). Whereas, the extrinsic motivations were further sub-divided into immediate payoff (payment), delayed pay-off (Signaling and human capital advancement) and social motivation (action significance by external values, action significance by external obligations and norms and Indirect feedback from the job). Regarding the company, there are also gains out of crowdsourcing which will be discussed in the next section.

BENEFITS OF CROWDSOURCING

There are a lot of benefits for the company to use crowdsourcing. These benefits as presented by (Stol and Fitzgerald, 2014) are as follows:

1. **Cost Reduction:** Using crowdsourcing for doing a task is financially effective, as the cost of hiring online crowd is much less than hiring internal employees for performing the same task. The crowd may even do the work voluntarily. In addition to the cost reduction, using the crowd gives indications of the actual needs of the customers.
2. **Faster Time-to-Market:** Using the crowd makes the production process more effective, as the task can be done in less time.
3. **Higher Quality:** Crowdsourcing allows harnessing the diversity of experiences and knowledge of the crowd. This leads to higher quality solutions.
4. **Higher Creativity and Innovation:** By receiving many contributions from the crowd, which are having a wide variety of experiences, mentalities and geographically distributed different cultures. This gives more opportunities to have more creative and innovative ideas.

PROBLEMS OF CROWDSOURCING

1. It is difficult to estimate the project budget, as the participants will vary in size.

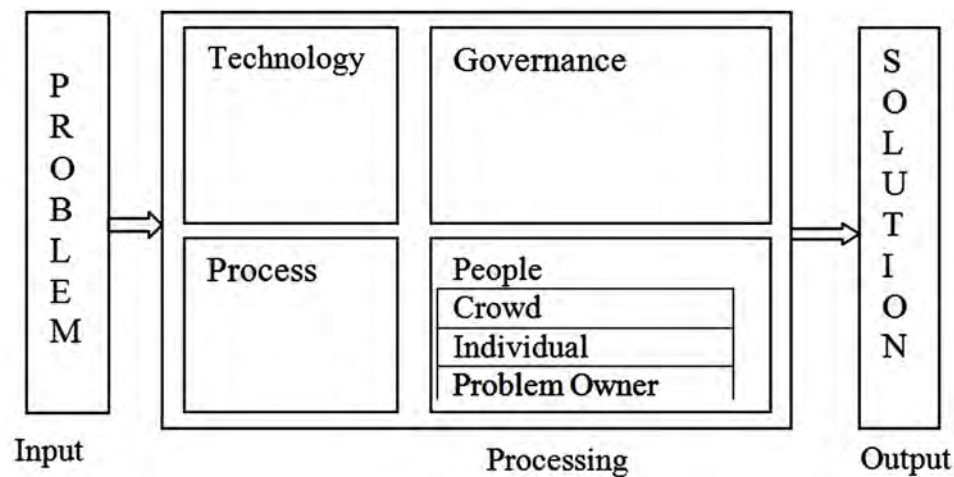
2. It is challenging to produce a precise specification of the problem, to present it to the crowd.
3. The crowd structure is missing.
4. Legal considerations, regarding who has the intellectual property of the contents.
5. Balancing the incentive motivations of the crowd (Extrinsic and intrinsic).
6. Problems regarding using crowdsourcing for software development are discussed in (Stol et al. 2014) as task decomposition, coordination and communication, planning and scheduling, quality assurance, Knowledge, and intellectual property.

CONCEPTUAL MODELS OF CROWDSOURCING

Conceptual models are abstractions and simplifications of the real world systems. Conceptual models are different from software design and implementation. An example of a conceptual model is the E-R (entity - relationship) diagram in database design, which is the logical design of the database, and then it is used for creating the physical database. Pedersen, J. et al. 2013, presented a conceptual model for crowdsourcing from the information systems perspectives. This conceptual model presents the basic input processing output components of Crowdsourcing. It is important to classify the research relating it to the information systems research community. The basic components of this proposed model are presented in Figure 3.

1. **Problem:** The specifications of the current situation and the tasks required to obtain a solution. The problem can be co-creation, crowd creation, crowd wisdom or crowd funding (Howe, J. 2008). Research related to the problem is concentrating on goals and tasks (Wiggins et al. 2012), and mechanisms for the evaluation of ideas (Blohm et al. 2011)

Figure 3. Crowdsourcing Conceptual Model (Pedersen, J. et al. 2013)



and user selection of a problem to contribute (Peddibhotla, 2013).

2. **Process:** Is the actual plan to solve the problem. The problem can be simple, one stage solution, such as product design or voting. It also can be a complex problem that requires a multiple-stage solution such as software development. Research targeting the processes includes collaborative e-science design (Farooq et al. 2009) and modeling information sharing process (Grant et al. 2010).
3. **Technology:** Acts as the enabling component to the crowdsourcing process. The technology involves not only the software platform, which facilitates the interaction between the sourcer and the crowd but also, tools supporting the online community creation. ICT technology and crowdsourcing (Davis, 2011), (Tilson et al. 2010).
4. **Governance:** These are the rules and regulations that organize the whole process of crowdsourcing. It also includes the rules for consistency preserving and quality assurance. Research involves quality control (Lasecki et al. 2012), (Lease, 2011) and crowdsourcing activity coordination (Zhang et al. 2011).

5. **People:** People participating in the crowdsourcing process are divided into three classes. The first is the crowd, all the possible participants. Second, the individual is the person who participates in the crowdsourcing process. Third, the problem owner can be an individual, an organization or a company having a problem aiming to solve. Research tackled the types of participants (Cullen et al. 2011), user motivations (Pilz et al. 2013)
6. **Solution:** The output of the crowdsourcing process. It can be a product, a design, a piece of software, or a rating of any of the previous artifacts. Research address the sourcer feedback (Moon et al. 2008), (Bao et al. 2011) and solution trustworthiness (Zhai et al. 2012).

Additionally, Malone et al. (2010) suggested a gene-based model for studying crowdsourcing systems. His model is composed of the following questions:

1. **What? (Goal or Mission)**
In this question the requester decides which tasks need to be performed, either create a new product or service or decide between different available options.

2. **Who? (Staff)**
The answer to the first question helps to decide the audience. Who will be the targeted participants and also determine the suitable software platform for the required task. If the task can be done by organization employees, the tasks can be assigned to them. Otherwise, it will be better to use crowdsourcing, where there are more varieties of tools and experiences.
3. **How? (Process)**
This refers to how the task will be accomplished, the useful tools and the results evaluation criteria. Considering the way in which the problem will be divided into smaller tasks. Moreover, deciding on the integration of the results to form a solution to the problem.
4. **Why? (Incentives or Motivations)**
This involves incentives of the crowd to participate. There should be a balance between intrinsic motivations, which are related to the participants and extrinsic, which is something external such as prize or money.
3. **Geospatial Mapping:** The term volunteered geographic information (VGI) suggested by Goodchild, 2007 has become very popular, with the spread of geographic mapping platforms, which allows the public to contribute their own maps and GPS traces. Examples are Open Street Map and Google My Map.
4. **Scientific Research:** There are many platforms that facilitate the research and collaboration in scientific problem solving such as (innocetive.com) and (researchgate.net).
5. **Decision Support:** Organizations learning from the crowd (Schlagwein et al, 2014) and (Chiu et al 2014) presented a framework for using the crowd to enhance the decision support for organizations.
6. **New Product Design:** Poetz et al, 2012 compared between using the crowd and employees for designing tasks. He concluded that crowdsourcing is a promising approach to product design.
7. **Medical Applications:** King et al. 2013 held a comparison between conventional methods and crowdsourcing methods for Skin self-examination. He argued that using crowdsourcing has a potential in the cancer detection task. Brabham et al. 2014 presented a framework for using crowdsourcing in public health applications. Meyer et al. 2016 proposed the use of crowdsourcing for medical diagnosis.
8. **Machine Learning and Prediction:** Tong et al. 2014 presented an approach to use crowdsourcing for race results prediction. Demartini, et al. 2013 proposed an approach to use the crowd for identifying similar items in the Linked Data. Scharl et al. 2012 presented an approach for the use of crowdsourcing in the acquisition of multi-lingual language data sets. Zhang et al. 2012 demonstrated using the crowd for web page clipping, which is useful for pattern mining applications. Sprugnoli, et al,

CROWDSOURCING APPLICATIONS

Crowdsourcing has been applied in many applications

1. **Software Development, Testing, and Engineering:** Using crowdsourcing for software development is a challenging process as it is a complex process, comprising of multiple stages and activities. It has been extensively discussed in (Stol et al. 2014), (LaToza et al, 2016) and Sharma et al, 2014).
2. **Question Answering:** Questions answering Systems (QAS) are software systems accepting user questions in natural language and presenting exact answer. They are becoming more popular with crowdsourcing user questions and answers. Yahoo Answers and Google Answers are examples of community QAS.

- 2013 presented the use of crowdsourcing for speech translation.
9. **Sales and Design:** Crowdsourcing has been used in product design to boost sales. This has been the idea used by (Threadless.com) by giving its customers the chance to send their own T-shirt designs. Then, different designs were evaluated and the designs with most sales are entitled to share profit with the company (Howe, J. 2008).
 10. **Crowdsourcing in Media:** Nylund, et al. 2014 demonstrated the role of crowdsourcing in the media industry. The authors claimed that the spread of social media sites among the crowds has created the potential for growing the citizen journalism. In other words, it has become easier for the public to capture the events as soon as it happens and spread it out throughout social media sites.
 11. **Marketing:** Doing marketing research and collecting crowd ideas and opinions about products or services during or after development is essential for business. (Whitla, 2009)
 12. **Optimization:** (Yi et al. 2011) demonstrated the use of crowdsourcing for the traveling salesman optimization problem.
 13. **Disaster Management:** (Gao et al, 2012), suggested the use of social media as a crowdsourcing platform for disaster relief.

Figure 4. Crowdsourcing platforms

PLATFORM	TASKS	MOTIVATIONS
TopCoder. (http://www.topcoder.com)	• Crowd Creation and Innovation. • Users participate in solving, designing or creating software.	• Extrinsic Motivations (Users are paid for the completed tasks)
Amazon Mechanical Turk (www.mturk.com)	• Micro-tasks (Crowd creation) • Acquire user input to evaluate a design or measure a particular criterion that requires human intelligence. • Purpose: Crowd rating, Crowd tagging and Crowd creation.	• Extrinsic motivations used to encourage participants. • People are paid a nominal amount of money for every task completed.
Crowdflower (http://www.crowdflower.com/)	• An online platform for collecting data from people around the world. • Purpose: Micro-task (Crowd creation and rating tasks)	• Extrinsic motivations used by paying contributors.
DesignCrowd (http://www.designcrowd.co.uk/)	• An online service to harness user innovations in designing a logo, web page, T-shirt, Brochure or business card. • Crowd creation.	• Extrinsic motivated by payment per design.
InnoCentive (http://www.innocentive.com/)	• Online based challenges for users to solve.	• Extrinsic motivations by giving awards for best solutions.
Istockphoto (http://www.istockphoto.com/)	• Open challenge for photographers to send their own photos and videos. • Crowd creation.	• Extrinsic motivations by giving credits that can be used by users to download their favorites.
(mobileworks) www.mobileworks.com	• An online crowdsourcing platform for business data collection. • Crowd creation task.	• Extrinsic Motivations. Participants are paid for their online work.
Kaggle (www.kaggle.com)	• An online crowdsourcing platform for data science competitions. • Crowd creation.	• Extrinsic Motivations. The best solvers are paid for their work.
http://www.crowdsourcing.com	• An online crowdsourcing platform for data science competitions. • Crowd creation.	• Extrinsic motivations. Users are given money for their participations.
www.utest.com	• An online crowdsourcing platform for testing software. • Crowd rating.	• Extrinsic motivations.

CROWDSOURCING PLATFORMS ON THE WEB

There are a plethora of Crowdsourcing platforms available online. Table 1 shows different available crowdsourcing platforms with their associated tasks and user motivations. Crowdsourcing platforms are different from crowdsourcing systems also in the user motivations. All the crowdsourcing platforms have extrinsic motivations.

FUTURE RESEARCH DIRECTIONS

Crowdsourcing has been an active research area for ten years. Although a lot of research has been done in many areas such as business, behavioral science, and computer science, there is still more to be done. First regarding the *problem*, it is critical to present the calls for participations in a way suitable for the intended audience and finding the suitable audience for the problem. This needs to be per application customization. A possible research question can be what are the criteria that determine the suitability of the individual to participate in certain crowdsourcing application? Second, regarding the *process*, there is a need for a unified set of processes for each crowdsourcing application context. Third, *technology*, it is important to integrate web2.0 technologies such as wikis blogs and social media with new technologies for crowdsourcing such as sensors, which are now widely available in mobile phones. Fourth, *governance*, it is crucial to control the whole process throughout presenting evaluation methods for every element in the systems starting with the problem specification to the final output. Fifth, *people*, studying user motivations for contributions is an important aspect, it needs to be studied per application, as it helps in future systems design. Finally, *the output* is the final solution to the problem presented by the crowd. It is essential to quantify the quality of the proposed solutions, in order to, select the best one. Thus, a concise set of procedures for measuring the quality of the

solution are needed. Moreover, it is essential for the organization or the individual to ensure that the intended goals of crowdsourcing have been achieved. Another issue is how to detect and avoid cheating from the crowd and create trust among the crowd members and between the requester and the crowd. Regarding applications, it is of interest to investigate the use of crowdsourcing with the semantic web and showing the differences imposed by these technologies.

W

CONCLUSION

Crowdsourcing is a promising research area which has proven success in many applications. It has been an active research area over the past ten years and it is anticipated to grow in the future, due to its endless applications. Using the crowd thinking and experience could give the organization new and innovative ideas that could never be possible using only the internal organization resources. Thus, crowdsourcing presents great opportunities for value creation in areas such as business, Information systems development, machine learning data collection and many more.

REFERENCES

- Ambati, V., Vogel, S., & Carbonell, J. (2012, February). Collaborative workflow for crowdsourcing translation. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work* (pp. 1191-1194). ACM.
- Azzam, T., & Jacobson, M. R. (2013). Finding a Comparison Group Is Online Crowdsourcing a Viable Option? *The American Journal of Evaluation*, 34(3), 372-384. doi:10.1177/1098214013490223
- Bao, J., Sakamoto, Y., & Nickerson, J. V. (2011). Evaluating design solutions using crowds. *Seventeenth Americas Conference on Information Systems*, 2013-5.

- Bell, D. (2009). *The Crowdsourcing Handbook—THE How To on Crowdsourcing, Complete Expert's hints and tips Guide by the leading experts, everything you need to know about Crowdsourcing*. Emereo Pty Ltd.
- Brabham, D. C. (2013). *Crowdsourcing*. MIT Press.
- Brabham, D. C., Ribisl, K. M., Kirchner, T. R., & Bernhardt, J. M. (2014). Crowdsourcing applications for public health. *American Journal of Preventive Medicine*, 46(2), 179–187. doi:10.1016/j.amepre.2013.10.016 PMID:24439353
- Chiu, C. M., Liang, T. P., & Turban, E. (2014). What can crowdsourcing do for decision support? *Decision Support Systems*, 65, 40–49. doi:10.1016/j.dss.2014.05.010
- Cullen, R., & Morse, S. (2011, January). Who's contributing: Do personality traits influence the level and type of participation in online communities. In *System Sciences (HICSS), 2011 44th Hawaii International Conference on* (pp. 1–11). IEEE.
- Davis, J., & Lin, H. (2011). *Web 3.0 and Crowdservicing*. AMCIS.
- Demartini, G., Difallah, D. E., & Cudré-Mauroux, P. (2013). Large-scale linked data integration using probabilistic reasoning and crowdsourcing. *The VLDB Journal*, 22(5), 665–687. doi:10.1007/s00778-013-0324-z
- Doan, A., Ramakrishnan, R., & Halevy, A. Y. (2011). Crowdsourcing systems on the world-wide web. *Communications of the ACM*, 54(4), 86–96. doi:10.1145/1924421.1924442
- Estellés-Arolas, E., & González-Ladrón-de-Guevara, F. (2012). Towards an integrated crowdsourcing definition. *Journal of Information Science*, 38(2), 189–200. doi:10.1177/0165551512437638
- Farooq, U., Ganoë, C. H., Carroll, J. M., & Giles, C. L. (2009). Designing for e-science: Requirements gathering for collaboration in CiteSeer. *International Journal of Human-Computer Studies*, 67(4), 297–312. doi:10.1016/j.ijhcs.2007.10.005
- Gao, H., Barbier, G., Goolsby, R., & Zeng, D. (2011). *Harnessing the crowdsourcing power of social media for disaster relief*. Arizona State Univ Tempe.
- Geiger, D., Rosemann, M., Fielt, E., & Schader, M. (2012). *Crowdsourcing information systems—definition, typology, and design*. Academic Press.
- Geiger, D., & Schader, M. (2014). Personalized task recommendation in crowdsourcing information systems—Current state of the art. *Decision Support Systems*, 65, 3–16. doi:10.1016/j.dss.2014.05.007
- Goodchild, M. F. (2007). Citizens as sensors: The world of volunteered geography. *GeoJournal*, 69(4), 211–221. doi:10.1007/s10708-007-9111-y
- Grant, T. J., & van den Heuvel, G. G. A. (2010, May). Modelling the Information Sharing Process in Military Coalitions: A work in progress. *Proceedings of the 7th International ISCRAM Conference*.
- Haythornthwaite, C. (2009, January). Crowds and communities: Light and heavyweight models of peer production. In *System Sciences, 2009. HICSS'09. 42nd Hawaii International Conference on* (pp. 1–10). IEEE.
- Howe, J. (2006). The rise of crowdsourcing. *Wired Magazine*, 14(6), 1–4.
- Howe, J. (2008). *Crowdsourcing: How the power of the crowd is driving the future of business*. Random House.

- King, A. J., Gehl, R. W., Grossman, D., & Jensen, J. D. (2013). Skin self-examinations and visual identification of atypical nevi: Comparing individual and crowdsourcing approaches. *Cancer Epidemiology*, 37(6), 979-984.
- Lasecki, W. S., & Bigham, J. P. (2012, October). Online quality control for real-time crowd captioning. In *Proceedings of the 14th international ACM SIGACCESS conference on Computers and accessibility* (pp. 143-150). ACM. doi:10.1145/2384916.2384942
- LaToza, T. D., & van der Hoek, A. (2016). Crowdsourcing in Software Engineering: Models, Motivations, and Challenges. *Software, IEEE*, 33(1), 74-80. doi:10.1109/MS.2016.12
- Lease, M. (2011). On Quality Control and Machine Learning in Crowdsourcing. *Human Computation*, 11, 11.
- Lebraty, J. F., & Lobre-Lebraty, K. (2013). *Crowdsourcing: One step beyond*. John Wiley & Sons. doi:10.1002/9781118760765
- Malone, T. W., Laubacher, R., & Dellarocas, C. (2010). The collective intelligence genome. *MIT Sloan Management Review*, 51(3), 21.
- Meyer, A. N., Longhurst, C. A., & Singh, H. (2016). Crowdsourcing Diagnosis for Patients With Undiagnosed Illnesses: An Evaluation of CrowdMed. *Journal of Medical Internet Research*, 18(1), e12. doi:10.2196/jmir.4887 PMID:26769236
- Moon, J. Y., & Sproull, L. S. (2008). The role of feedback in managing the Internet-based volunteer work force. *Information Systems Research*, 19(4), 494-515. doi:10.1287/isre.1080.0208
- Nylund, M., & Gopalkrishnan, A. (2014). *Crowdsourcing in Media*. Academic Press.
- Peddibhotla, N. (2013). How individuals choose topics to contribute at an online context. *Electronic Markets*, 23(3), 241-250. doi:10.1007/s12525-013-0125-7
- Pedersen, J., Kocsis, D., Tripathi, A., Tarrell, A., Weerakoon, A., Tahmasbi, N.,... De Vreede, G. J. (2013, January). Conceptual foundations of crowdsourcing: A review of IS research. In *System Sciences (HICSS), 2013 46th Hawaii International Conference on* (pp. 579-588). IEEE. doi:10.1109/HICSS.2013.143
- Pilz, D., & Gewald, H. (2013). Does Money Matter? Motivational Factors for Participation in Paid-and Non-Profit-Crowdsourcing Communities. *Wirtschaftsinformatik*, 37.
- Poetz, M. K., & Schreier, M. (2012). The value of crowdsourcing: Can users really compete with professionals in generating new product ideas? *Journal of Product Innovation Management*, 29(2), 245-256. doi:10.1111/j.1540-5885.2011.00893.x
- Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, 25(1), 54-67. doi:10.1006/ceps.1999.1020 PMID:10620381
- Scharl, A., Sabou, M., Gindl, S., Rafelsberger, W., & Weichselbraun, A. (2012). *Leveraging the wisdom of the crowds for the acquisition of multilingual language resources*. Academic Press.
- Schlagwein, D., & Bjørn-Andersen, N. (2014). Organizational learning with crowdsourcing: The revelatory case of LEGO. *Journal of the Association for Information Systems*, 15(11), 754.
- Sharma, M., & Padmanaban, R. (2014). *Leveraging the Wisdom of the Crowd in Software Testing*. CRC Press. doi:10.1201/b17483
- Sprugnoli, R., Moretti, G., Fuoli, M., Giuliani, D., Bentivogli, L., Pianta, E.,... Brugnara, F. (2013, May). Comparing two methods for crowdsourcing speech transcription. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on* (pp. 8116-8120). doi:10.1109/ICASSP.2013.6639246

Stol, K. J., & Fitzgerald, B. (2014, May). Two's company, three's a crowd: a case study of crowdsourcing software development. In *Proceedings of the 36th International Conference on Software Engineering* (pp. 187-198). ACM. doi:10.1145/2568225.2568249

Tilson, D., Lyytinen, K., & Sørensen, C. (2010). Digital Infrastructures: The Missing IS Research Agenda. *Information Systems Research*, 21(5).

Tong, Y., Cao, C. C., & Chen, L. (2014, August). Tcs: efficient topic discovery over crowd-oriented service data. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 861-870). ACM. doi:10.1145/2623330.2623647

Von Ess, H. (2010). *Crowdsourcing: How to find a crowd*. Berlin: ARD/ZDF Medien Akademien.

Whitla, P. (2009). Crowdsourcing and its application in marketing activities. *Contemporary Management Research*, 5(1).

Wiggins, A., & Crowston, K. (2012, January). Goals and tasks: Two typologies of citizen science projects. In *System Science (HICSS), 2012 45th Hawaii International Conference on* (pp. 3426-3435). IEEE.

Wu, B., Zhong, E., Tan, B., Horner, A., & Yang, Q. (2014, August). Crowdsourced time-sync video tagging using temporal and personalized topic modeling. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 721-730). ACM. doi:10.1145/2623330.2623625

Yi, S. K. M., Steyvers, M., Lee, M. D., & Dry, M. (2011). Wisdom of the crowds in traveling salesman problems. *Memory & Cognition*, 39, 914-992. PMID:21264578

Zhai, Z., Hachen, D., Kijewski-Correa, T., Shen, F., & Madey, G. (2012, January). Citizen engineering: Methods for "crowdsourcing" highly trustworthy results. *System Science (HICSS), 2012 45th Hawaii International Conference on*, 3406-3415.

Zhang, H., Horvitz, E., Miller, R. C., & Parkes, D. C. (2011). *Crowdsourcing general computation*. Academic Press.

Zhang, L., Tang, L., Luo, P., Chen, E., Jiao, L., Wang, M., & Liu, G. (2012, August). Harnessing the wisdom of the crowds for accurate web page clipping. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 570-578). ACM. doi:10.1145/2339530.2339621

ADDITIONAL READING

Brabham, D. C. (2008). Crowdsourcing as a model for problem solving an introduction and cases. *Convergence (London)*, 14(1), 75-90. doi:10.1177/1354856507084420

Doan, A., Ramakrishnan, R., & Halevy, A. Y. (2011). Crowdsourcing systems on the world-wide web. *Communications of the ACM*, 54(4), 86-96. doi:10.1145/1924421.1924442

KEY TERMS AND DEFINITIONS

Crowd: A group of people from geographically distributed locations, different professions, and different cultures.

Crowdsourcing Platform: A more generic online crowdsourcing system, designed for multi-purpose. It can be used for multiple tasks such as product design, rating or voting and data collection tasks. An example is Amazon Mechanical Turk.

Open Source: The act of developing software for free usage and distributions. It is not only for free usage, but also, for free modifications.

Outsourcing: The process used by companies to get a product or a service done externally by other people or companies.

User Incentives: Motivational factors that are used to encourage the crowd to participate in a crowdsourcing activity.

Web 2.0: A set of tools such as social media, wikis and blogs, which made it possible for people to create and share their own contents on the web. The further advancements of using the Semantics and linkage are called Web 3.0.

W

Revisiting Web 2.0

Michael Dinger

University of South Carolina Upstate, USA

Varun Grover

Clemson University, USA

INTRODUCTION

Though well over a decade old now, the term Web 2.0 was intended to capture a wave of technologies that indicate advancement beyond the original Web 1.0, which was dominated by tightly controlled, relatively static websites. As such, Web 2.0 represents a set of technologies that enable high levels of interactivity and participation via the internet. It is an umbrella term that describes a variety of dynamic and community-based web initiatives that place value on the power of distributed knowledge, leverage data, and provide users with rich multimedia experiences (O'Reilly, 2005). Web 2.0 technologies have given rise to the dominance of social media which over 65% of U.S. adults use (Perrin, 2015). Furthermore, Web 2.0 technologies are now inherent throughout the modern web, but are also increasingly embedded throughout the milieu of digital technology, including mobile platforms and cloud services. Web 2.0 technologies and their associated changes in internet usage enable new forms of data collection and data analytics, including social network analysis and social media analytics (Chen, Chiang, & Storey, 2012). As a result, it is useful to revisit the core technological components of the Web 2.0 wave and to consider how these elements have become embedded throughout modern business applications and how the technologies can add value for consumers and businesses. Cutting edge firms can leverage such technologies to create integrated solutions that improve their relationships with customers and enrich customer experiences.

Many businesses create additional value for users and customers by leveraging the voluntary participation of users through interactive technologies. For example, Amazon offers extensive product support through an interactive set of tools supporting customer interactions. Customers can leave product reviews for their past purchases. However, Amazon also provides a question and answer system, wherein prospective customers can ask questions about a product and previous buyers are able to respond, with the answers posted on the product listing page. The application of interactive technologies enables Amazon to create value without extensive organizational labor, but instead harnesses the power of their existing customer base to create valuable content.

Businesses continue to capitalize on this set of technologies in a variety of ways. Many companies leverage interactive technologies by capturing customer data and leveraging it to generate instantaneous, custom-tailored customer experiences (Bodoff & Ho, 2015). For example, Netflix aggregates and analyzes subscriber movie preferences in order to provide accurate movie recommendations. Similarly, Pandora, the internet music service, creates customized radio stations for individuals based on their expressed preferences. However, Pandora also mines user data to customize targeted advertising to each user (Singer, 2014). Furthermore, businesses can leverage Web 2.0 technologies in order to dynamically cooperate with customers and partners in efforts to generate new design innovations (Roberts & Dinger, 2016; Roberts & Grover, 2012; Wong, Peko, Sundaram,

& Piramuthu, 2015). Both online and traditional businesses must understand how to navigate and capitalize on the changing internet terrain to stay competitive in the Web 2.0 era.

BACKGROUND

Web 2.0 thinking emphasizes the distributed and interactive nature of information technologies. Therefore, the core concept of a web page is altered to allow for quick and efficient interaction from users. This mindset is represented in the way that users can create, remove or edit informational content on wikis, comment on blogs or content aggregation sites like Reddit, or drive the content of media-sharing sites like YouTube. The distributed nature of Web 2.0 technologies allows many users to create and participate while needing little technical knowledge.

Characteristics of Web 2.0

Web 2.0 technologies can be identified by a number of common characteristics. These technologies generally capitalize on the ability of website users to actively participate, including the capacity to dynamically contribute content and network with other users. Web 2.0 initiatives are dynamic in nature, enabling constant change and updates. Also, Web 2.0 technologies regularly include social networking elements which enable users to form connections with one another. Finally, these endeavors are noted for their reliance on the distributed contributions of many participants.

Distributed Contributions

The primary driver behind Web 2.0 technologies are the ability of firms to harness the value of distributed contributions from many users. Wikis inherently rely on the contributions and efforts of many users. The intent of wikis is to represent a culmination of the knowledge of all participating users. Media platform sites like YouTube, Instagram, Tumblr and Flickr entirely consist of user

contributions. Facebook has opened up its software platform so that users can create and contribute original applications (developers.facebook.com). Similarly, Apple has enabled a wide range of developers to create new and innovative applications for the iOS ecosystem (developer.apple.com) as has Google with the Android ecosystem (developer.android.com). The ability to harness the distributed contributions of many participants plays a significant role in a firm's ability to generate value from these initiatives.

Dynamic Nature

The dynamic nature of Web 2.0 technologies is driven by their ability to be quickly changed. A core design element of wikis is the ability to add, remove or change content quickly. On social networking sites, users are able to add, modify and remove content ease. Content platforms like Twitter, Instagram, and Vine thrive on the speed of content creation and sharing. By focusing on a small piece of content, users can rapidly create and share content, such as a 140 character post, an image, or a 6 second video, respectively. The speed of movement can be a double-edge sword, as it enables companies to quickly interact with customers to resolve problems, or for upset users to share stories of bad customer service that go viral. For example, a United Airlines customer created a YouTube video to share how the airline broke his guitar and would not reimburse any of the damages, and a \$1,200 guitar ended up costing the airline far more in bad publicity (Sawhney, 2009). In another case, a Comcast customer service representative would not simply comply with a customer's request to cancel their service. The customer recorded the call and posted it online on SoundCloud, where it went viral with bad publicity for Comcast (Diamond, 2015).

Rich Media

Rich media is a common characteristic of Web 2.0 technologies. Many online sites and app platforms are compiled solely of user-generated

content, including YouTube, Instagram, Tumblr and Flickr, but rich media can enhance any user experience. For years, many news sites, such as CNN.com and ESPN.com, have embedded video in conjunction with print stories. However, the technology has become ubiquitous, with local TV channels and newspapers commonly posting video in conjunction with news stories.

Social Networking Elements

Social networking leverages people-to-people interactions. For example, blog users can form connections to other bloggers. Analysis suggests that users of media sharing websites, specifically YouTube, can engage in social networking activity through the manner in which they manipulate access to their contributed media (Lange, 2007). Finally, social networking platforms can be embedded in a variety of different websites.

ARCHETYPES OF WEB 2.0

The Web 2.0 trend popularized three general archetypes of technology: blogs, wikis, and social networking sites (SNS). Blogs, short for weblogs, are perhaps the most traditional of the technologies and are essentially websites that can be dynamically and regularly updated. Wikis are information-distribution sites that exist through the collective efforts of many users. Wikis are intended to efficiently harness the collective knowledge of all willing participants. In general on a wiki, all users can generate original content and edit or remove published content. The core concept of the wiki is that all changes can be made, or undone, quickly. Social networking sites represent the ability of internet technologies to connect users and provide a platform that enables these connections.

Blogs

Blogs are simple websites that are largely defined by the fact that they are updated easily and regu-

larly. Blogs are inherently flexible and can be used for a variety of purposes, ranging from knowledge management initiatives (Sultan, 2013) to customer engagement tools (Brodie, Ilic, Juric, & Hollebeek, 2013) or potential sources of business intelligence (Chau & Xu, 2012). In fact, blogs have long been challenging traditional media outlets for the attention of many online users (Singh, Veron-Jackson, & Cullinane, 2008).

Blogs can serve as rallying points for people of similar interests, such as consumer advocates at the Consumerist (www.consumerist.com), and organizations can also offer their own blogs as a platform on which to interact directly with customers. Blogs can serve as a platform to gain the attention of an organization (Roberts & Dinger, 2016), as they not only enable individuals to publically voice their own concerns, but entire communities can form regarding a common cause. For example, in 2005, a writer for BusinessWeek chronicled his complaints with Dell on his own personal blog, called Dell Hell (Jarvis, 2007). Eventually, he penned an open letter to Michael Dell and challenged him to respond to the public complaints and concerns of bloggers. The following year, Dell capitalized on the chance to positively interact with their customers by creating their own blog (Fournier & Avery, 2011).

Within-firm blogs offer a platform for individual employees to express themselves inside the firm. These repositories can be used to store thoughts or voice opinions, and as such, internal blogs can be used as a knowledge repository and to enhance knowledge sharing (Baxter & Connolly, 2013). Executives have been regularly using internal blogs to publicize statements within firms and connect with their own employees (Gaines-Ross, 2013). Firms and CEOs can use a public blog in order to provide a 'face' for the organization (Vidgen, Mark Sims, & Powell, 2013). The publication of a blog enables the firm to interface directly with consumers. By allowing customers to interact directly with the firm, the organization may seem more human and down-to-earth (Singh et al., 2008). While the blog can be used for impression management, it can also be

used to gauge consumer reactions to changes or investigate the potential for new products. For instance, Facebook makes public announcements on their blog and updates users on potential changes to Facebook (blog.facebook.com).

Wikis

Wikis are a type of technology that allows many users to combine their collective knowledge (Gaines-Ross, 2013). The wiki is a type of technology that also embodies a specific mindset towards the accumulation of knowledge. Wikis are designed such that participants can quickly and easily create new wiki pages, edit existing wiki pages, and trim unnecessary wiki pages. Wikis offer the ability for many users to converge on a given topic. As such, wikis represent an excellent technological platform for knowledge generation and can be harnessed to inspire innovations and growth (Biasutti & Heba, 2012). Wiki technology enables users to search for information, link other articles, author new articles, and tag articles (McAfee, 2006). Furthermore, wikis can use extensions to enhance user ability to find related content and set up signals which notify users of new content (McAfee, 2006).

A popular source of online information, Wikipedia hosts over 5.1 million articles in English alone. However, the collaborative nature of public wikis raises concern about the validity of information contained in such a source (Reavley et al., 2012). Therefore, a number of checks are implemented to insure the validity of contributions to a given wiki topic. A common wiki tenet is to make it easy to correct mistakes. Changes are tracked and edits are easily undone. This makes it a simple task to undo vandalism. Firms have the opportunity to derive value from the collective of the public that participates in wikis. Wikis can form around any given subject. For example, users of the Linux-based operating system Ubuntu operate a wiki to assist all Ubuntu users (wiki.ubuntu.com).

Firms have the opportunity to harness the power of their consumers by creating and driving

wikis that feature the firm's products. By allowing consumers to create content revolving around the firm's products, the firm is both developing the core consumer base in a positive manner and enabling the creation of additional value for any consumer that can benefit from this knowledge and experience. For instance, fans of Marvel properties have created the Marvel Database (marvel.wikia.com), which hosts over 150,000 articles created and moderated by the fan community.

Internal wikis can be applied to a number of possibilities. Wikis offer value as tools for collaborative knowledge management (Gaines-Ross, 2013). The wiki becomes a repository of knowledge that users can improve over time. For example, one report suggests that the U.S. Government leverages wikis in a variety of ways, including an Army-based wiki on Afghanistan, an intelligence wiki full of non-classified information for government personnel (Intellipedia), and a wiki for diplomats (Diplopedia) (Bell, 2009). Wikis are relatively easy to set up and modify, which makes them ideal for projects and routine use. Wikis can be created at the beginning of a project and then modified according to project progress. Alan Mansfield, a small business owner, estimates that using wikis to manage projects cuts down time spent on each project by 25 percent (Miller, 2006). In running his literary agency, Mansfield uses wikis to "store drafts, e-mails, proposals, contracts, contacts, and anything else to do with the project" (Miller, 2006).

Social Network Sites

Social network sites (SNS) use internet-enabled technologies to allow users to form or maintain social connections. Boyd and Ellison define social networks as "web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system" (2007 p 211). The core factors driving SNS are the platforms on which many users can

W

engage and interact with one another. SNS have become so extensive that firms are having trouble with employees who insist on accessing social networking sites during regular working hours (Andreassen, Torsheim, & Pallesen, 2014).

While firms may view social networking as a distraction to employees, firms can seek to engage with social networking software in a value-generating manner. Many companies leverage social networking software internally in order to facilitate connections and exchanges between employees (Leonardi, Huysman, & Steinfield, 2013) and enhance knowledge sharing (Ellison, Gibbs, & Weber, 2014). Social networks also facilitate connecting potential participants for distributed innovation processes (Lee, Olson, & Trimi, 2012). As recruitment tools, social networks present a valuable tool that can be leveraged by recruiters to more thoroughly assess potential recruits (Davison, Bing, Kluemper, & Roth, 2016). Organizations can also use social networks and social media to engage in direct market research (Chen et al., 2012; Fan & Gordon, 2014). Also, social networks enable firms to create higher levels of customer engagement through positive discourse with customers (Sashi, 2012).

A FRAMEWORK FOR WEB 2.0 INITIATIVES

This framework attempts to simplify and direct Web 2.0 initiatives. Due to the broad range of potential Web 2.0 initiatives, we propose a framework for understanding the value generated from these initiatives.

Value Derived within the Firm

This part of the framework addresses the ability of the firm to generate value through Web 2.0 initiatives. These applications may take a variety of forms, but the focus is on the benefits of the internal applications of the Web 2.0 applications. These applications may take form in terms of

knowledge management initiatives (Gaines-Ross, 2013; Sultan, 2013), project management efforts (Miller, 2006), and social networks that connect employees (Leonardi et al., 2013). Firms must understand the type of benefit they intend to achieve before installing Web 2.0 technologies. The application of multiple Web 2.0 endeavors allows for firms to integrate across technology, thus potentially giving rise to super-additive value due to the complex interactions between multiple Web 2.0 platforms (see Figure 1).

Value Derived from External Sources

The firm should consider how to derive value from Web 2.0 instances that exist in the public space. The firm should evaluate all potential sources of access to these instances. Once the sources are identified, the firm can approach the issue of deriving value from these websites. For instance, firms can leverage social network sites to evaluate job applicants or to actively recruit individuals (Davison et al., 2016). Firms can also capitalize on externally-run blogs in order to understand the mindset and concerns of customers, as external sites provide access to the unfiltered complaints of customers (Fournier & Avery, 2011). Due to the open and simple nature of Web 2.0 technologies, many users are participants on multiple platforms. Furthermore, users, and Web 2.0-savvy businesses, form connections between multiple Web 2.0 initiatives. For instance, Facebook enables Instagram users to link their account to their Facebook account and post content to both Instagram and Facebook simultaneously. The challenge for firms is to identify potential sources of value and to effectively funnel value from those sources into the firm (see Figure 2).

Value Derived from the Interaction Between the Firm and Public

Firms should evaluate how to generate value from dynamic interactions with their customers or other public sources. This proposition deals

Figure 1. Value derived within the firm

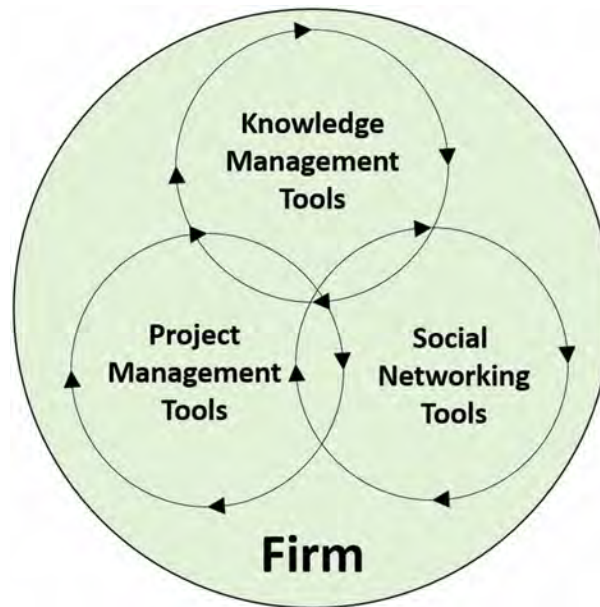
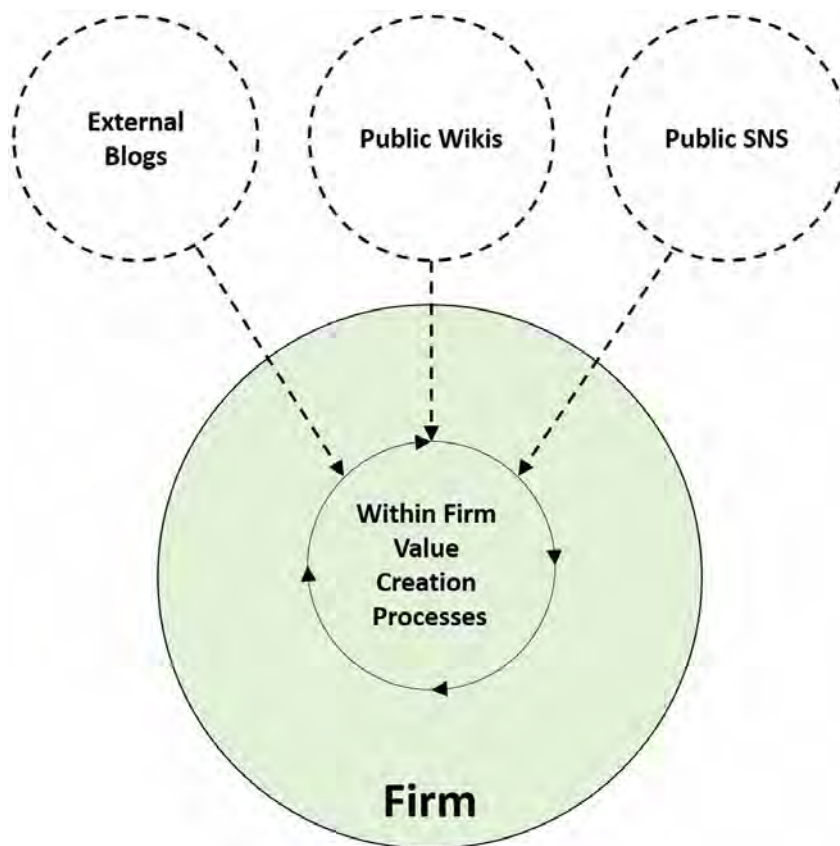


Figure 2. Value derived from external sources



with two fronts: social and technical. The social aspect of this interaction deals with the actual exchange between firm and public. For instance, Dell responded to customer complaint blogs directly (Jarvis, 2007). Beyond interacting directly with customers and other external parties, a firm may choose to create a platform to facilitate this interaction. For instance, Dell chose to publish a blog (a technical issue) for the purpose of interacting with customers (a social issue) (Fournier & Avery, 2011; Jarvis, 2007). This suggests that firms must first understand and engage Web 2.0 technologies from a technical perspective before progressing to capitalizing on potential social interactions via firm-driven technologies.

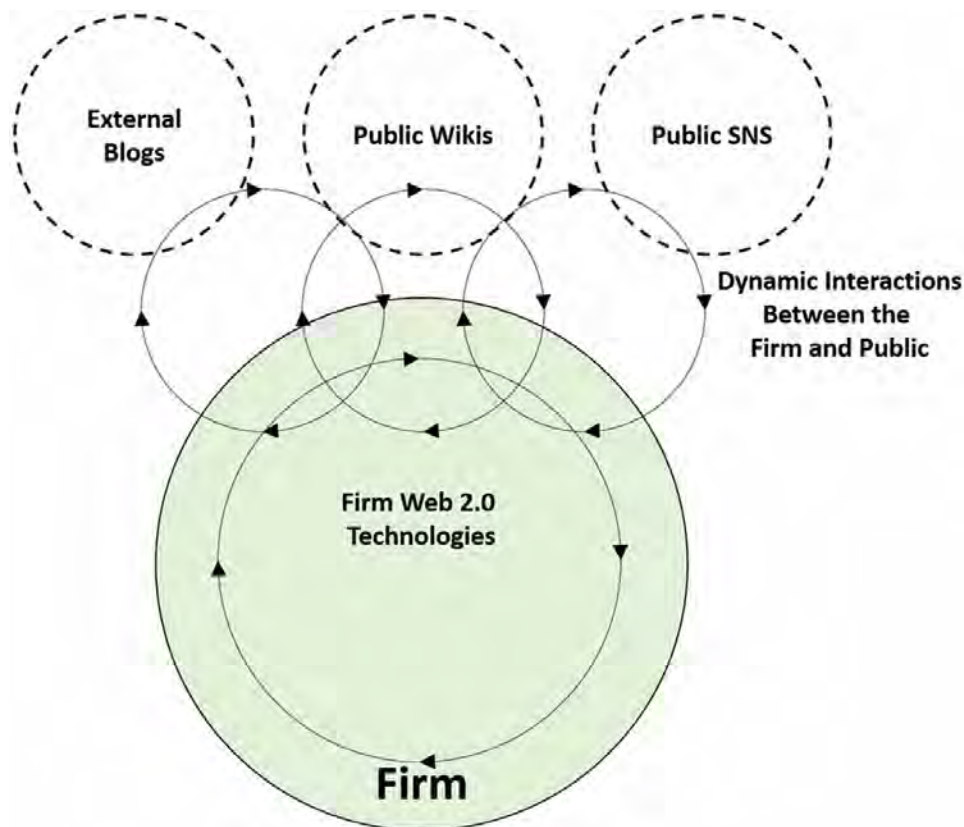
Furthermore, firms can capitalize on direct involvement with customers and third parties by creating a platform to drive distributed innovation (Lee et al., 2012). Many design initiatives focus

on distributing design and innovation processes within the firm. However, firms can leverage these technologies and allow customers to become involved in the creation processes (Frow, Nenonen, Payne, & Storbacka, 2015). In doing so, firms can engage in dynamic value generation in conjunction with customers and third parties. Firms can seek to integrate technologies within the firm and with external sources in order to reach out and provide a platform for dynamic social interactions and distributed innovation processes (see Figure 3).

Value Derived from Innovative Business Models

Firms must stay current amidst the constant change of business models. For instance, firms are attempting to develop retail environments within online virtual worlds (Frow et al., 2015).

Figure 3. Value derived from the interaction between the firm and public



Furthermore, firms may consider how to leverage social media and mobile applications to advance firm goals and create additional value for consumers and the firm (Aksoy et al., 2013). Online firms may focus on implementing aspects of Web 2.0 technologies into their existing business models. For instance, online retailing environments may enable individual sellers to connect their web stores, thus leading to a virtual mall for online shoppers (Stephen & Toubia, 2010). Firms, both online and traditional, must stay on top potential changes to their business model in order to be competitive in this fast-paced environment.

Risks Posed by Web 2.0 Initiatives

In seeking to gain value from Web 2.0 endeavors, firms must not ignore the risks associated with these technologies. Primarily, firms must control their exposure regarding information technology security. While these technologies inherently connect individuals within the firm and may connect the firm to outsiders, firms must maintain tight security standards in order to prevent unauthorized access or the introduction and spread of harmful computer viruses. Similarly, firms should be wary of potentially unreliable information available from external sources. The potential for bad information in online settings is always a concern and is particularly relevant regarding wikis (Reavley et al., 2012), however, recent trends have tilted towards analytical study of large data sets derived from social media, which should iron out small fluctuations in data accuracy (Chen et al., 2012; Fan & Gordon, 2014).

CONCLUSION

Web 2.0 technologies have not only changed the landscape of the internet, but continue to create a dynamic and changing competitive environment for many firms. Firms that aggressively implement Web 2.0 and social media technologies may

successfully develop competencies that create additional value and drive competitive advantage.

W

REFERENCES

- Aksoy, L., van Riel, A., Kandampully, J., Larivière, B., Joosten, H., Malthouse, E. C., & Huang, M.-H. et al. (2013). Value fusion: The blending of consumer and firm value in the distinct context of mobile technologies and social media. *Journal of Service Management*, 24(3), 268–293. doi:10.1108/09564231311326996
- Andreassen, C. S., Torsheim, T., & Pallesen, S. (2014). Predictors of Use of Social Network Sites at Work-A Specific Type of Cyberloafing. *Journal of Computer-Mediated Communication*, 19(4), 906–921. doi:10.1111/jcc4.12085
- Baxter, G. J., & Connolly, T. M. (2013). The state of art of organisational blogging. *The Learning Organization*, 20(2), 104–117. doi:10.1108/09696471311303755
- Bell, S. (2009). WIKIS for Reference, Enthusiasts, and Government Information. *Online*, 33(1), 20–23.
- Biasutti, M., & Heba, E.-D. (2012). Using Wiki in teacher education: Impact on knowledge management processes and student satisfaction. *Computers & Education*, 59(3), 861–872. doi:10.1016/j.compedu.2012.04.009
- Bodoff, D., & Ho, S. Y. (2015). Effectiveness of Website Personalization: Does the Presence of Personalized Recommendations Cannibalize Sampling of Other Items? *International Journal of Electronic Commerce*, 20(2), 208–235. doi:10.1080/10864415.2016.1087821
- Boyd, D. M., & Ellison, N. B. (2007). Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication*, 13(1), 210–230. doi:10.1111/j.1083-6101.2007.00393.x

- Brodie, R. J., Ilic, A., Juric, B., & Hollebeek, L. (2013). Consumer engagement in a virtual brand community: An exploratory analysis. *Journal of Business Research*, 66(1), 105–114. doi:10.1016/j.jbusres.2011.07.029
- Chau, M., & Xu, J. (2012). Business intelligence in blogs: Understanding consumer interactions and communities. *Management Information Systems Quarterly*, 36(4), 1189–1216.
- Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *Management Information Systems Quarterly*, 36(4), 1165–1188.
- Davison, H. K., Bing, M. N., Kluemper, D. H., & Roth, P. L. (2016). *Social Media as a Personnel Selection and Hiring Resource: Reservations and Recommendations*. In *Social Media in Employee Selection and Recruitment* (pp. 15–42). Springer. doi:10.1007/978-3-319-29989-1_2
- Diamond, M. L. (2015). *Viral power: Negative social media bad for business*. Retrieved June 27, 2016, from <http://www.usatoday.com/story/money/business/2015/01/10/viral-power-negative-social-media-bad-for-business/21570851/>
- Ellison, N. B., Gibbs, J. L., & Weber, M. S. (2014). The use of enterprise social network sites for knowledge sharing in distributed organizations the role of organizational affordances. *The American Behavioral Scientist*.
- Fan, W., & Gordon, M. D. (2014). The power of social media analytics. *Communications of the ACM*, 57(6), 74–81. doi:10.1145/2602574
- Fournier, S., & Avery, J. (2011). The uninvited brand. *Business Horizons*, 54(3), 193–207. doi:10.1016/j.bushor.2011.01.001
- Frow, P., Nenonen, S., Payne, A., & Storbacka, K. (2015). Managing Co-creation Design: A Strategic Approach to Innovation. *British Journal of Management*, 26(3), 463–483. doi:10.1111/1467-8551.12087
- Gaines-Ross, L. (2013). Get Social: A Mandate for New CEOs. *MIT Sloan Management Review*, 54(3), 1.
- Jarvis, J. (2007). Dell Learns to Listen. *BusinessWeek*.
- Lee, S. M., Olson, D. L., & Trimi, S. (2012). Co-innovation: Convergenomics, collaboration, and co-creation for organizational values. *Management Decision*, 50(5), 817–831. doi:10.1108/00251741211227528
- Leonardi, P. M., Huysman, M., & Steinfield, C. (2013). Enterprise social media: Definition, history, and prospects for the study of social technologies in organizations. *Journal of Computer-Mediated Communication*, 19(1), 1–19. doi:10.1111/jcc4.12029
- McAfee, A. P. (2006). Enterprise 2.0: The Dawn of Emergent Collaboration. *Sloan Management Review*, 47(3), 20–28.
- Miller, D. I. (2006). Collaborating through wikis. *CNet*.
- O'Reilly, T. (2005). *What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software*. Retrieved from <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>
- Perrin, A. (2015). *Social Media Usage: 2005-2015*. Retrieved May 15, 2016, from <http://www.pewinternet.org/2015/10/08/social-networking-usage-2005-2015/>
- Reavley, N. J., Mackinnon, A. J., Morgan, A. J., Alvarez-Jimenez, M., Hetrick, S. E., Killackey, E., & Jorm, A. F. et al. (2012). Quality of information sources about mental disorders: A comparison of Wikipedia with centrally controlled web and printed sources. *Psychological Medicine*, 42(08), 1753–1762. doi:10.1017/S003329171100287X PMID:22166182

Roberts, N. H., & Dinger, M. (2016). The Impact of Virtual Customer Community Interactivity on Organizational Innovation: An Absorptive Capacity Perspective. *IEEE Transactions on Professional Communication*, 59(2), 110–125. doi:10.1109/TPC.2016.2561118

Roberts, N. H., & Grover, V. (2012). Leveraging Information Technology Infrastructure to Facilitate Firms Customer Agility and Competitive Activity: An Empirical Investigation. *Journal of Management Information Systems*, 28(4), 231–270. doi:10.2753/MIS0742-1222280409

Sashi, C. (2012). Customer engagement, buyer-seller relationships, and social media. *Management Decision*, 50(2), 253–272. doi:10.1108/00251741211203551

Sawhney, R. (2009). *Broken Guitar Has United Playing the Blues to the Tune of \$180 Million*. Retrieved August 3, 2015, from <http://www.fastcompany.com/1320152/broken-guitar-has-united-playing-blues-tune-180-million>

Singer, N. (2014). *Listen to Pandora, and It Listens Back*. Retrieved June 2, 2016, from http://www.nytimes.com/2014/01/05/technology/pandora-mines-users-data-to-better-target-ads.html?_r=0

Singh, T., Veron-Jackson, L., & Cullinane, J. (2008). Blogging: A new play in your marketing game plan. *Business Horizons*, 51(4), 281–292. doi:10.1016/j.bushor.2008.02.002

Stephen, A. T., & Toubia, O. (2010). Deriving value from social commerce networks. *JMR, Journal of Marketing Research*, 47(2), 215–228. doi:10.1509/jmkr.47.2.215

Sultan, N. (2013). Knowledge management in the age of cloud computing and Web 2.0: Experiencing the power of disruptive innovations. *International Journal of Information Management*, 33(1), 160–165. doi:10.1016/j.ijinfomgt.2012.08.006

Vidgen, R., Mark Sims, J., & Powell, P. (2013). Do CEO bloggers build community? *Journal of Communication Management*, 17(4), 364–385. doi:10.1108/JCOM-08-2012-0068

Wong, T. Y., Peko, G., Sundaram, D., & Piramuthu, S. (2015). Mobile environments and innovation co-creation processes & ecosystems. *Information & Management*.

W

KEY TERMS AND DEFINITIONS

Blogs: Short for ‘weblogs,’ blogs are simple, content-driven sites that are updated regularly.

Distributed Contributions: The practice of leveraging the willing participation of users.

Dynamic Content: Internet content that can be modified and uploaded quickly which keeps users ‘up-to-the-minute.’

Rich Media: Media, particularly images, video and sound, conveyed via internet technologies that provides a deeper user experience than simple text.

Social Networking Elements: Web 2.0 applications embedded in web sites that enable users to uniquely identify and form connections with one another.

Social Networking Sites (SNS): Web-based platforms which enable many individuals to create individual profiles, find and connect with other users.

Web 2.0: An umbrella term that refers to an assortment of advances in internet technologies, marked by increases in rich media, dynamic content, social networking elements, and distributed contributions.

Wikis: Content-driven sites which are editable by all participants and focus on harnessing the collective knowledge of all users.

Search Engine Optimization

Dimitrios Giomelakis

Aristotle University of Thessaloniki, Greece

Andreas A. Veglis

Aristotle University of Thessaloniki, Greece

INTRODUCTION

The extensive range of information resources and services is certainly one of the most important features of the Internet while at the same time, web search is considered as a crucial application for managing the massive volumes of distributed web content. Beyond argument, search engines have made an enormous contribution to the web by making the process of finding information online a very quick and easy process. Today, major search engines are considered to be the most common and trusted tool or service to retrieve information from the Internet (Spink & Xu, 2000). Also, they are the primary method used for navigation for hundreds of millions of users worldwide and one of the most common online activities (Purcell, 2011; Purcell, Brenner, & Rainie, 2012). The majority of Internet traffic depends largely on them (Safran, 2013) and thus, web search is one of the best sources of traffic for any website. However, it is true that the vast majority of all search traffic comes from the first or the first pages of search results as users usually focus on the top ranks.

There are two ways an online user – customer will find a business website via a search engine: through a pay-per-click campaign (PPC) or through an organic result listing that is based essentially on what is called Search Engine Optimization or briefly, SEO. The latter can be defined as the process of affecting - improving the visibility of a website (or a web page) so that it ranks well for particular keywords in a search engine's "natural" or "organic" (un-paid) search results (Ledford,

2009; Potts, 2007). Generally, the earlier, and more frequently a site appears in the search engine results page, the more visitors it will receive from the search traffic. In other words, it is a set of techniques that take into account the evaluation criteria of search engines regarding website content and structure (Giomelakis & Veglis, 2015a).

There have been plenty studies regarding online users' click behavior on search engine results pages. According to the results, 90 percent of search engine users never read beyond the third page of search results (iProspect, 2006). Also, the top listing in Google's organic search results receives 32.5 percent of the traffic, compared to 17.6 percent for the second and 11.4 percent for the third position. Finally, websites listed on the first page in Google's results generate 92 percent of all traffic from an average search (Chitika, 2013). From all the above, it is evident that if a website is not in the first search results page or even worse is absent from the top 30, it has almost no chance of being read by a user (Clay, 2006). As a consequence, and while more and more websites are indexed by search engines and compete one another to ensure their own market share, it is clear that factors as the highest ranking and top of the results page become increasingly essential for businesses of all kinds (Enge, Spencer, & Stricchiola, 2015; Giomelakis & Veglis, 2015a).

This chapter provides an overview of Search Engine Optimization, with a focus on its different characteristics as well its history and how it has evolved over the years. In order to give a better understanding of the importance of SEO in the

current state of the Internet and in information search, basic knowledge of how search engines operate along with their recent updates are also provided.

BACKGROUND

Search engines are software that catalogs the World Wide Web and provides search using keywords into their vast databases containing full-text indexes of web pages. Users actually search this database of retrieved web pages, not the World Wide Web itself. As a consequence, they manage to take rapid results something that would be impossible if engines were trying to search billions of pages on the web in real time (Veglis, Pomportsis, & Avraam, 2004). When users click on search results they retrieve the current version of a web page. It is worth mentioning that search engines consist of three parts: the web crawler or spider, the indexer and also the query processor (Mudgil, Sharma, & Gupta, 2013). The crawler systematically browses the World Wide Web, looks at every URL (Uniform Resource Locator) collecting keywords and phrases on each page, which are then included in a massive database. The crawler is also responsible for keeping indexed pages up to date (Ledford, 2009). Search engines start with a set of very high quality sites and then visit the links on each page (of those sites) to discover other web pages. This complex process repeats over and over again until the crawling is complete. Through links, web crawlers (i.e. automated robots) can reach the many trillions of interconnected documents (Enge, Spencer, & Stricchiola, 2015).

Search engines use algorithms so as to find and collect information about web pages. Generally, a search algorithm can be characterized as a problem-solving procedure that takes the problem (i.e. the users' word or phrase), sifts through a vast database of cataloged keywords with their URLs, and then returns a listing of best-matching web pages according to the search criteria. Search results (i.e. page ranking) usually depend on the

perceived quality-importance of the page (quality score - links from other sites) conforming to the algorithm that's being used as well as on relevance (relevant terms/links) and several other factors such as frequency of keywords, location, age or click-through rates (Ledford 2009; Enge, Spencer, & Stricchiola, 2015). It is worth mentioning that in the case of Google, the famous PageRank (one of its oldest algorithms) is a quality metric used in ranking (one out of many) that measures the importance of web pages by counting the number and quality of links to a page (Page, Sergey, Rajeev, & Winograd, 1998; Grappone & Couzin 2008). PageRank (PR) is a numerical measurement (0–10, the higher the better) and in the past it was one of the most important factors. Google has continuously improved the way it uses links to impact rankings (one reason is to deal with link spam), and its current algorithm is not based on PR as it was originally defined (Enge, Spencer, & Stricchiola, 2015). However, it still has some value in SEO (Grappone and Couzin 2008; Slegg 2013; Giomelakis & Veglis, 2015b). Also, it is true that PR had become an obsession for many SEO workers worrying about their numbers and examining closely the quality of every new link. The last Pagerank update was in December 2013 and on April 15, 2016 the search engine has officially shut down the PageRank data to public (Southern, 2016; Schwartz 2016). In brief, search algorithms can be classified into three broad categories: on-page algorithms that measure on-page factors looking at the elements of every page (e.g. keywords in content or meta tags), whole-site algorithms that focus on the relationship of site elements (e.g. anchor text, linking between pages or architecture of pages) and finally, off-site algorithms that explore the links between the websites. All these three types are generally part of a larger algorithm (Ledford, 2009).

Search Engines have evolved dramatically over time and they constantly "crawl" the web to update their databases with new information. According to statistics, Google is the world's most popular search engine having the vast majority

(over 2/3) of the global search engine market followed by Bing, Yahoo and Baidu, China's most popular search engine (Netmarketshare, 2016). In a comScore survey (2015) data showed that Google sites led the U.S. explicit core search market with 64.5 percent market share, followed by Microsoft sites with 19.8 percent and Yahoo! sites with 12.8 percent.

HISTORY OF SEO

SEO constitutes a part of Search Engine Marketing (SEM) and it is considered one of the leading and most influential activities in the field of online marketing, hence there is a wealth of information and tools designed to train and support the community (Potts, 2007; Nigro, Balduzzi, Cuesta, & Cisaró, 2012). Historically, the actual phrase "Search Engine Optimization" was possibly first coined in 1997 by an unknown person. Search guru Danny Sullivan found for the first time the term used in May 1997 in a meta tag on his website, Search Engine Watch. However, it is quite possible that it was used before that. Sullivan acknowledges Bruce Clay as one of the first people in the industry to popularize the term (Sullivan, 2004; Dover, 2011).

SEO has always been associated with the influence of search engine results and its creation is rationally connected with the creation of the first search engines in the early-mid 90's. Danny Dover describes (2011) that early SEO experts used to take advantage of alphabetical ("AAA", "1ForU") and chronological (submit websites at certain times) order to get to the top of rankings. Since then, things have changed considerably. Search engines improved using more complicated algorithms and at the same time, SEO became a more profitable business. Over time, site owners started to identify the worth of having their websites highly ranked and visible in online search results. They also noticed that the higher their website ranked, the more people would click on their content. From a simple URL submis-

sion initially and use of unsophisticated tactics, companies started to implement strategies such as keywords insertion, link building (process of obtaining hyperlinks from other websites), and recently social media utilization.

From the early days of search engines until today many SEO users have attempted to deceive search engines in order to artificially improve their visibility. That is usually called as "black-hat SEO" and pertains to the use of strategies or tactics (e.g. excessive use of keywords, hidden text or pages) that are used to get higher rankings in an unethical manner, doesn't obey search engines guidelines for webmasters or involve deception. In general, it is a practice that violates the search engines' terms of service and possibly results in a website being banned from a search engine and affiliate sites (Grappone & Couzin 2008; Ledford 2009). A remarkable example of penalizing was Google removal of BMW Germany and Ricoh Germany for using misleading tactics in 2006 (Cutts, 2006). As stated, search engines have evolved tremendously in recent years and one of the most basic reasons that their algorithms are in a constant change is due to the need to address "Black Hat SEO" and provide the best possible results. Thus, they developed more complex ranking algorithms, considering additional factors that were more difficult for webmasters to manipulate.

SEO CHARACTERISTICS

There are definitely myths about search engines. For example, that attaining a top spot in the organic, natural results requires actually paying Google and other search sites. On the contrary, a high position in search results requires actually worthy content coupled with time, effort, and skill (Potts, 2007). SEO can be applied to many different websites to some degree and can target different types of search, including image - video search, local, news or academic search (Beel, Gipp, & Wilde, 2010; Giomelakis & Veglis, 2015b; Wikipedia, 2016). Arguably, it is very closely connected to

e-Commerce websites (Malaga, 2007). According to a relevant study, 89 percent of consumers turn to search engines to find information on products, services or businesses before making purchases (Fleishman-Hillard & Harris Interactive, 2012). Given that it takes into account how search engines work and the actual terms that people search for, there are several factors that affect SEO but unfortunately, there is no magic recipe for the first page and certainly no guaranteed results. By contrast, there is a variety of things to consider that in conjunction can bring good results. Some of them can directly affect SEO while others can have an indirect impact on search rank. Everything has its value and Google, Bing or other search engines use them to evaluate and rank web pages in their results. Although there are differences in the ways various search engines work, the basic rules remain similar and a well-planned SEO strategy can usually bring good results in all major search engines (Giomelakis & Veglis, 2015a).

SEO strategies can be divided into four main categories: keyword research/selection, search engine indexing, on-page optimization and off-page optimization (Malaga, 2008). *On-page optimization* refers to the management of all factors related directly to someone's website (content – appropriate keywords, architecture and internal link structure as well as html elements). More specifically, this category contains among others, page titles (or HTML titles tag that usually appear in search results), on-page headlines, description of web pages (or meta description tag) and URL's. Using main, appropriate words/phrases that people often use in their searches in all the above elements is crucial to SEO. However, it should be noted that the use of keywords to a great extent in the actual content of a web page is by all means not a key requirement. In contrast, it is effective when this happens naturally without boring the readers. Besides, keyword stuffing (i.e. the practice of overloading a web page's content with keywords) can bring unpleasant results to a website (Giomelakis & Veglis, 2015a). Further-

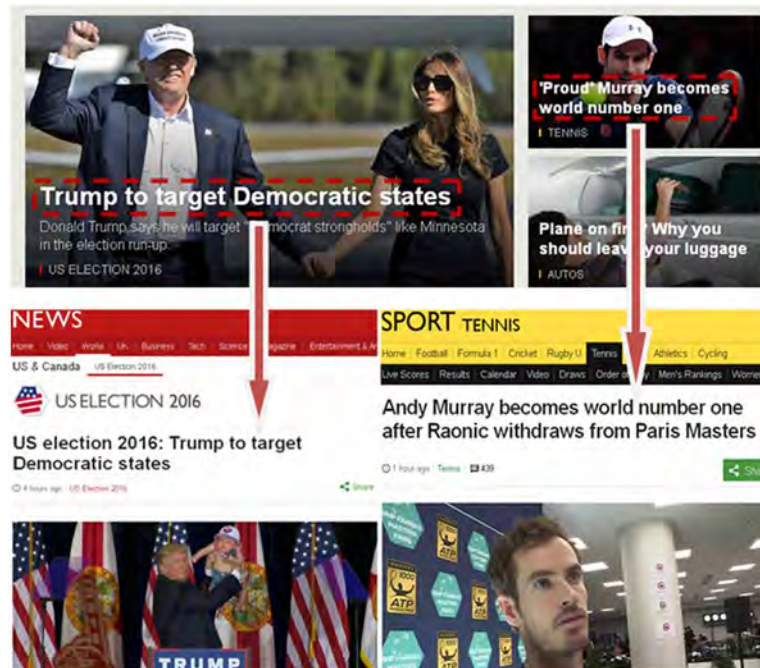
more, as mobile devices become more prevalent in daily life, a good user experience across all devices (desktop, tablet or smartphone) has become a significant consideration in SEO strategy (mobile SEO). Google has updated its algorithm in this direction (it took the name "Mobilegeddon") given that in many countries around the world including the US and Japan, more searches take place on mobile devices than on computers (Makino, Jung, & Phan, 2015; Dischler, 2015).

Examples of SEO Practices

1. Page titles/headlines
 - a. World Cup 2014: Brazil humiliated by Germany 7-1 in semi-final / Novak Djokovic beats Roger Federer in epic Wimbledon 2014 men's final (SEO-Friendly)
 - b. Germany's triumph in the semi-final / Djoko did it again! (Non-SEO Friendly)
2. Url's
 - www.newsandseo.com/2014/seo-best-practices-for-structuring-urls.html (SEO-Friendly)
 - www.newsandseo.com/2014/how-to-create-seo-friendly-urls.html (SEO-Friendly)
 - www.newsandseo.com/2014/new/item.aspx?itemid=3051.html (Non-SEO Friendly)
 - www.newsandseo.com/2014/forum/viewtopic/m-159327.html (Non-SEO Friendly)
3. Different titles/headlines

This practice is used by several news organizations (e.g. BBC News, Guardian, Huffington Post or New York Times) and the most common is the different headlines between home page or other Web site indexes and the story page itself (see figure 1). The latter that appears in search results, is usually more specific including more keywords for the topic.

Figure 1. The dual headline system in BBC news



Regarding *off-page optimization*, it includes all the efforts made away from someone's website like link building or social signal strategy. Links from other websites (called backlinks or inbound links) have great importance, serve as votes for the value and are a basic requirement for a web page to be found both from search engines and online searchers. In addition, it is highly important for inbound links to be qualitative, coming from well-known, trusted websites in order to give much more value and better evaluation. There are many different ways to gain inbound links such as articles/domain submission to directories (e.g. Dmoz), forum, blogs, and news aggregators as well as building relationships and links exchange with other websites. It is true that for a website with quality content, it is usually a matter of time until it gains many natural and qualitative backlinks. However, not only inbound links but also external links that lead to other relevant web pages (preferably credible and high-ranking sites) can contribute positively especially if a website is new (Giomelakis & Veglis, 2015a).

SEO, SOCIAL MEDIA AND CONTENT

The development of the World Wide Web (Web 2.0) over the last decade has affected and brought significant changes in web content such as user-generated content and social web. Furthermore, search technology itself evolved over the last years with the growth of contemporary search engines something that has arguably affected SEO. Nowadays, the optimization for search engines is not anymore only about link-building and keywords. Social media and social signals (such as Facebook Shares/likes) have increasingly become one of the many factors search engines take seriously into account. In order to enhance their perfect reach, many search engines incorporate information from Web 2.0 services into their searchable indexes (Zimmer, 2008). For instance, active Facebook profile pages have many ranking possibilities for branded, subject-based and personal name-based queries at Google results. Also, due to Twitter's domain authority, its profile pages also rank highly for relevant searches. Results from Twit-

ter can be used for queries where real-time data are helpful (Dover, 2011). Regarding Google+, its characteristics make it a far superior platform for SEO and sharing content on it can possibly influence search rankings in significant ways (Shepard, 2013). In this context, recent studies from Searchmetrics (2013) and Moz's experts (Peters, 2013) confirmed how important social media signals are to SEO ranking, especially Google +1s. Social signals, in general, have a positive impact on websites and are considered as the new link building metric as search engines increasingly search for social signals to help the ranking of pages (Rayson, 2013; Frasco, 2013). The more referrals someone has across the web, the more search engine spiders notice and categorize their content (Ehrlich, 2013). Social media and SEO do overlap, and the former can contribute to the total organic success of websites in several ways (Pascale, 2014).

Today, recent algorithm changes from search engines especially in Google (e.g. Penguin, Panda, and Hummingbird) have added more value to quality as well as content marketing that many experts have called it as the new SEO (DeMers, 2013). For example, the change of Panda algorithm (February 2011) was made in order to eliminate low-quality websites in favor of those with frequently updated, in-depth content. Also, Google Penguin (April 2012) was the response to "Black-hat" SEO and web spam and finally Hummingbird (late 2013) added the contextual search, a way to interpret content by looking at the relationship between terms (Rampton & DeMers, 2014). The world of SEO is constantly evolving and thus, SEO workers have to be always up to date on new developments and also utilize useful tools and services for their work. Tools such as Google Analytics (the most widely known web analytics package), Google PageSpeed Insights (analyzes the performance of a web page providing information on how to improve the load time), Google Search Console (previously Google Webmaster Tools that enable webmasters to monitor and maintain their presence in search results), Google Trends (a tool to iden-

tify "hot" topics), SEMrush (software for online marketing) or finally MozBar (for the page and domain authority of a website) are very popular in SEO industry. It is crucial to understand that SEO is not only about engines but also making someone's website (and its content) user-friendly and better for people. The content is still the most important thing for any website and this has to be reliable and interesting. Even search engines look now for unique, qualitative and useful content that will satisfy and motivate readers to share it through social networks, blog posts or forums. In spite of how good is SEO, it is not likely for any website with useless content to get ranked well in the long term because no one is ever going to visit or look for it. Besides, Google and other search engines got much smarter and can distinguish between content with real value and content that exists only for SEO purposes.

FUTURE RESEARCH DIRECTIONS

Beyond argument, there are several important aspects in this field that require further study. For instance, the increasing involvement of social media in SEO strategy is an area of research that has to be studied more in order to better clarify the relation between them. And that's because there is a strong indication that social networks and social media, in general, will continue to increase their significance and presence in the internet population. It is true that web search is increasingly becoming more personalized and user influenced. For many years, SEO and search results were all about keywords and link building. Nowadays, things are changed, the search results are arguably better and the search experience is much more advanced. SEO experts need to take the audience, their interests, user experience or what device they are using more into account when developing their strategies. All the above are issues that deserve more careful attention in future research because these are elements intimately related with search engines. In this context

and as the web ecosystem evolves, the so-called mobile, local or vertical search-SEO has grown considerably gaining significant attention from field experts. Also, we would say that the mobile landscape has evolved to such an extent that it has become a completely unique (mobile-first) search environment. All these new types of SEO along with the opportunities they offer in online industry are areas that need to be addressed in future studies.

CONCLUSION

Undeniably, search engines have come a long way since their appearance in the early 90's. In this context, SEO grew out of the development of search engines and the World Wide Web. As the years pass, there is no question that search engines will continue to evolve their algorithms and make improvements in their ability to crawl and index new types of content such as multimedia or data coming from social media. Over time, there have been fundamental shifts in the SEO landscape and field experts see a shift away from traditional ranking factors to deeper analysis and factors such as quality, multi-form content and social signals (DeMers, 2013). Within this progress, social media have increasingly become one of the many factors search engines take into account and content is still the king (Zimmer, 2008; Dover, 2011).

It might be thought that SEO will always have a place in a web strategy as long as Google and other search engines continue to drive a great amount of traffic on the Internet. The tactics involved in optimization may change, but the need will always remain. It is evident that every business that has a web presence (or wants to get clients from the web) would be wise to think about a SEO strategy as well as the potential it creates in order to increase or maintain its position within the search results. Implementing SEO into a website is becoming more and more necessary in today's digitalized world. Given that people will always be on a quest for information, and given the market-

ing imperative to get your information seen and keep ahead of the competition, companies will continue to figure out ways to use SEO-related techniques for higher rankings. The good use of SEO certainly does not guarantee high traffic and top rank. Nevertheless, an effective strategy can play its role by helping companies to obtain the best results and take advantage of the benefits provided through the utilization of search engines and web search in general.

REFERENCES

- Beel, J., Gipp, B., & Wilde, E. (2010). Academic Search Engine Optimization (ASEO): Optimizing Scholarly Literature for Google Scholar and Co. *Journal of Scholarly Publishing*, 41(2), 176–190.
- Chitika Online Advertising Network. (2013, June 7). *The Value of Google Result Positioning*. Retrieved February 24, 2016 from <http://chitika.com/google-positioning-value>
- Clay, B. (2006). Put SEO in your site design. *Bruce Clay*. Retrieved March 13, 2016, from http://www.bruceclay.com/eu/seo/articles/seo_sitedesign.htm
- ComScore. (2015, March 17). *comScore analysis of the U.S. search marketplace (February 2015)*. Retrieved February 15, 2016 from <http://www.comscore.com/Insights/Market-Rankings/comScore-Releases-February-2015-US-Desktop-Search-Engine-Rankings>
- Cutts, M. (2006, February 4). *Ramping up on international webspam* [Web log comment]. Retrieved February 10, 2016 from <https://www.matcutts.com/blog/ramping-up-on-international-webspam>
- DeMers, J. (2013). SEO in 2013: 7 Surprisingly Simple Factors That Will Take The Lead. *Search Engine Journal*. Retrieved March 13, 2016 from <http://www.searchenginejournal.com/seo-in-2013-7-surprisingly-simple-factors-that-will-take-the-lead/57092>

- Dischler. (2015, May 5). *Building for the next moment* [Web log comment]. Retrieved February 10, 2016 from <http://adwords.blogspot.gr/2015/05/building-for-next-moment.html>
- Dover, D. (2011). *Search Engine Optimization Secrets*. Indianapolis, IN: Wiley.
- Ehrlich, S. (2013). SEO Best Practices: The Impact of Social Media on Search Engine Optimization. *Bulldog reporter*. Retrieved March 10, 2016 from <https://www.bulldogreporter.com/seo-best-practices-the-impact-of-social-media-on-search-engine-opti/>
- Enge, E., Spencer, S., & Stricchiola, J. (2015). *The Art of SEO. Mastering Search Engine Optimization* (3rd ed.). O'Reilly.
- Fleishman-Hillard International Communications and Harris Interactive. (2012, January 31). *Digital Influence Index*. Retrieved March 5, 2016 from <http://fleishmanhillard.com/2012/01/31/2012-digital-influence-index-shows-internet-as-leading-influence-in-consumer-purchasing-choices/>
- Frasco, S. (2013). 6 Reasons Social Media is Critical to Your SEO. *Search Engine Watch*. Retrieved March 13, 2016 from <http://www.socialmediatoday.com/content/6-reasons-social-media-critical-your-seo>
- Giomelakis, D., & Veglis, A. (2015a). Employing Search Engine Optimization Techniques in Online News Articles. *Studies in Media and Communication*, 3(1), 22–33. doi:10.11114/smc.v3i1.683
- Giomelakis, D., & Veglis, A. (2015b). Investigating Search Engine Optimization Factors in Media Websites. *Digital Journalism*, 4(3), 379–400. doi:10.1080/21670811.2015.1046992
- Grappone, J., & Couzin, G. (2008). *Search Engine Optimization - an Hour a Day* (2nd ed.). Wiley Publishing.
- Iprospect. (2006, April). *Search Engine User Behavior Study*. Retrieved March 8, 2016 from http://district4.extension.ifas.ufl.edu/Tech/TechPubs/WhitePaper_2006_SearchEngineUserBehavior.pdf
- Ledford, J. L. (2009). *Search Engine Optimization Bible* (2nd ed.). Indianapolis, IN: Wiley.
- Makino, T., Jung, C., & Phan, D. (2015, February 26). *Finding more mobile-friendly search results* [Web log comment]. Retrieved February 11, 2016 from <https://webmasters.googleblog.com/2015/02/finding-more-mobile-friendly-search.html>
- Malaga, R. A. (2007). The Value of Search Engine Optimization: An Action Research Project at a New E-Commerce Site. *Journal of Electronic Commerce in Organizations*, 5(3), 68–82. doi:10.4018/jeco.2007070105
- Malaga, R. A. (2008). Worst Practices in Search Engine Optimization. *Communications of the ACM*, 51(12), 147–150. doi:10.1145/1409360.1409388
- Mudgil, P., Sharma, A. K., & Gupta, P. (2013). An Improved Indexing Mechanism to Index Web Documents. *Proceedings of the 5th International Conference on Computational Intelligence and Communication Networks (CICN)*. doi:10.1109/CICN.2013.101
- Net Market Share. (2016). *Desktop Search Engine Market Share - February 2016, Market Share Statistics for Internet Technologies*. Retrieved March 2, 2016 from <https://www.netmarketshare.com/search-engine-market-share.aspx?qprid=4&qpcustomd=0>
- Nigro, H. O., Balduzzi, L., Cuesta, I. A., & Císaro, S. E. G. (2012). Knowledge Based System for Intelligent Search Engine Optimization. In F. L. Gaol (Ed.), *Recent Progress in Data Engineering and Internet Technology* (pp. 65–72). Berlin: Springer. doi:10.1007/978-3-642-28798-5_10

- Page, L., Sergey, B., Rajeev, M., & Winograd, T. (1998, January 29). *The PageRank Citation Ranking: Bringing Order to the Web*. Technical Report, Stanford InfoLab. Retrieved March 5, 2016 from <http://ilpubs.stanford.edu:8090/422>
- Pascale, A. (2014). 7 Legitimate Ways That Social Media Impacts SEO. *Clickz*. Retrieved March 10, 2016 from <https://www.clickz.com/clickz/column/2342211/7-legitimate-ways-that-social-media-impacts-seo>
- Peters, M. (2013, July 9). *2013 Search Engine Ranking Factors (Moz's Study)* [Web log comment]. Retrieved February 10, 2016 from <http://moz.com/blog/ranking-factors-2013>
- Potts, K. (2007). *Web Design and Marketing Solutions for Business Websites*. New York: Apress.
- Purcell, K. (2011). *Search and email still top the list of most popular online activities*. Pew Research Center. Retrieved March 2, 2016 from <http://www.pewinternet.org/2011/08/09/search-and-email-still-top-the-list-of-most-popular-online-activities/>
- Purcell, K., Brenner, J., & Rainie, L. (2012). *Search Engine Use 2012*. Pew Research Center. Retrieved March 2, 2016 from <http://www.pewinternet.org/2012/03/09/search-engine-use-2012/>
- Rampton, J., & DeMers, J. (2014). The 2014 Beginner's Guide to SEO. *Search Engine Journal*. Retrieved from <https://www.searchenginejournal.com/seo-guide/>
- Rayson, S. (2013). The Social Media Optimization (SMO) of SEO: 7 Key Steps. *Social Media Today*. Retrieved March 11, 2016 from <http://www.socialmediatoday.com/content/social-media-optimization-smo-seo-7-key-steps>
- Safran, N. (2013, June 25). *Natural Web Site Traffic Accounts for Nearly Half of All Traffic* [Web log comment]. Retrieved February 6, 2016 from <http://www.conductor.com/blog/2013/06/web-traffic-natural-search-data/>
- Schwartz, B. (2016). Google Toolbar PageRank officially goes dark. *Search Engine Land*. Retrieved November 05, 2016 from <http://searchengineland.com/google-toolbar-pagerank-officially-goes-dark-247553>
- Search Engine Optimization. (n.d.). In *Wikipedia*. Retrieved from https://en.wikipedia.org/wiki/Search_engine_optimization
- Searchmetrics. (2013). *Search Metrics Ranking Factors and Rank Correlation*. Retrieved March 2, 2016 from <http://www.searchmetrics.com/en/knowledge-base/ranking-factors-us-2013/>
- Shepard, C. (2013, August 20). *Amazing Correlation between Google +1s and Higher Search Rankings* [Web log comment]. Retrieved February 12, 2016 from <http://moz.com/blog/google-plus-correlations>
- Slegg, J. (2013). Is Google Getting Ready to Retire PageRank? *Search Engine Watch*. Retrieved October 28, 2016 from <http://searchenginewatch.com/article/2299954/Is-Google-Getting-Ready-to-Retire-PageRank>
- Southern, M. (2016). Google Confirms Toolbar PageRank is No More. *Search Engine Journal*. Retrieved November 05, 2016 from <https://www.searchenginejournal.com/google-pagerank/159112/>
- Spink, A. & Xu, J. L. (2000). Selected results from a large study of Web searching: The Excite study. *Information Research*, 6(1).
- Sullivan, D. (2004, June 14). Who Invented the Term "Search Engine Optimization"? *Search Engine Watch*. Message posted to <http://forums.searchenginewatch.com/showpost.php?p=2119&postcount=10>
- Veglis, A., Pomportsis, A., & Avraam, E. (2004). *Searching and gathering information on the Internet*. Thessaloniki: Tziola. (In Greek)

Zimmer, M. (2008). The Externalities of Search 2.0: The Emerging Privacy Threats When the Drive for the Perfect Search Engine Meets Web 2.0. *First Monday*, 13(3). doi:10.5210/fm.v13i3.2136

ADDITIONAL READING

Adams, R. (2015). *SEO 2016: Learn Search Engine Optimization*. CreateSpace Independent Publishing Platform.

Clarke, A. (2015). *Search engine optimization 2016: Learn SEO with smart internet marketing strategies*. CreateSpace Independent Publishing Platform.

Dick, M. (2011). Search Engine Optimisation in UK News Production. *Journalism Practice*, 5(4), 462–477. doi:10.1080/17512786.2010.551020

Google. (2010). *Search Engine Optimization Starter Guide*. Retrieved February 15, 2016 from <http://static.googleusercontent.com/media/www.google.com/en/webmasters/docs/search-engine-optimization-starter-guide.pdf>

McDonald, J. (2015). *SEO Fitness Workbook, 2016 Edition: The Seven Steps to Search Engine Optimization Success on Google*. CreateSpace Independent Publishing Platform.

Moz. (2015). *Moz's Search Engine Ranking Factors 2015. Expert Survey and Correlation Data*. Retrieved March 5, 2016 from <https://moz.com/search-ranking-factors>

KEY TERMS AND DEFINITIONS



Click-Through Rate (CTR): The ratio of users who click on a specific link compared to the number of total users who view an email, a web page or an advertisement. It is used to measure the success of an advertisement campaign. For example, if an advertisement had 10 clicks and 1000 impressions (number of times it was served), the CTR would be 10% (clicks / impressions x 100).

Local Seo: The optimization process that focuses on results that are relevant to a user based on its current location.

Organic Results: Also called «natural» or un-paid. The results that appear because of their relevance to the search terms.

Social Signals: A kind of recommendations through social media. Social signals are the Facebook shares or likes, the tweets, Google +1s, pins or some other way of social media bookmarking or sharing.

User-Generated Content (UGC): Any form of digital content such as images, video, status updates or blogs that is produced and shared by users of an online service or website, often made available via social media websites.

Vertical Search: The search accessibility of specific segments and formats of online content. Vertical search services focus on specific topics such as news, restaurants or products. Google provides also vertical search such as Google news or Google images.

Toward Trustworthy Web Services Coordination

Wenbing Zhao

Cleveland State University, USA

INTRODUCTION

Many Web-based companies such as amazon.com, yahoo.com and twitter.com have been offering Web services to their partners and customers. Through such Web services, new value-added services have been provided and hence higher revenues have been generated. Essentially, the Web services technology is transforming the World Wide Web from a predominantly publishing platform to a programmable platform, which undoubtedly will make it easier to conduct business online, and enable automated business-to-business communications (Papazoglou, 2003; Zhao et al., 2008). The Web services technology is particularly useful for Application Service Providers that offer various on-demand services and software-as-a-service (SAAS) to their customers (Chakrabarty, 2007). Such service-oriented computing is attractive to many businesses because they can save valuable resources and money by avoiding installing and maintaining sophisticated enterprise software on-site. Furthermore, most of business interactions are transactional, which require well-defined standardized coordination support. To meet this requirement, a number of specifications have been proposed and rectified by OASIS (Feingold & Jeyaraman, 2009; Little & Wilkinson, 2009; Freund & Little, 2009).

Inevitably, the dependability requirement of these Web services has been increasing because of the key role they play in e-businesses. One of the most well-known approaches to enhancing the trustworthiness of Web services coordination is Byzantine fault tolerance (BFT) (Zhao, 2014). The BFT technique employs space redundancy

(i.e., replication) and a replication algorithm that can tolerate malicious faults (often referred to as Byzantine fault) (Lamport et al., 1982; Castro & Liskov, 2002).

In this article, we present an overview of our recent works on enhancing the trustworthiness of Web services coordination for business activities and transactions. The approach is based on what we call application-aware Byzantine fault tolerance. We argue that it is impractical to apply general-purpose Byzantine fault tolerance algorithms for such systems in a straightforward manner. Instead, by exploiting the application semantics, much lighter weight solutions can be designed to enhance intrusion tolerance, and hence the trustworthiness of systems that require Web services coordination.

BACKGROUND

Web services interactions are becoming more and more complex in structure and relationships. More complex means we need longer time to execute them, because of business latencies and user interactions. The Web Services Coordination specification (WS-Coordination) (Feingold & Jeyaraman, 2009) describes an extensible framework for plugging in protocols that coordinate the actions of Web services applications. Such coordination protocols can be used to support a variety of business applications, including those that require strict consistency and those that require agreement of a proper subset of the participants. The framework enables a Web service to create a context needed to propagate an activity to other

Web services and to register for a particular coordination protocol.

There are two types of business transactions. One follows the traditional atomic transaction semantics, and the other is referred to as business activities, which implies that the atomicity property may be relaxed. The former is suitable for short transactions that require strong atomicity, such as a fund transfer transaction. The latter is more suitable long running transactions, such as those used in supply chain management. Based on WS-Coordination, two specifications, namely Web Service Atomic Transaction (WS-AtomicTransaction) (Little & Wilkinson, 2009) and Web Service Business Activity (WS-BusinessActivity) (Freund & Little, 2009), were standardized by OASIS to address the coordination needs for common types of business transactions.

Web Services Atomic Transactions

The Web Services Atomic Transactions specification defines a coordination framework for Web services atomic transactions. In a distributed atomic transaction, all participants must reach the same final agreement as to whether the transaction has succeeded or not. This is ensured by the coordination mechanisms specified in WS-AtomicTransaction. In WS-AtomicTransaction, there are three actors for each transaction: Completion Initiator, Coordinator and Participants. Each provides a different set of services for the atomic transaction, and they interact with each other via two protocols, the Completion Protocol and the Two-Phase Commit Protocol (Gray & Reuter, 1993) (Tanenbaum & Steen, 2002).

The completion initiator is responsible to start and terminate a transaction. It also provides the Completion Initiator Service so that the coordinator can inform it the final outcome of the transaction, as part of the completion protocol. The coordinator provides the following services:

- **Activation Service:** At the beginning of a transaction, the initiator invokes the

Activation Service for creating a coordinator object, which will generate a new coordination context for the transaction and return it to the initiator. The coordination context contains a unique transaction identifier and an endpoint reference for the Registration Service. This coordination context will be included in every request messages within the transaction boundary.

- **Registration Service:** The participants and the completion initiator use this service to register their endpoint references for other associated participant-side services. Later these endpoint references will be used by the coordinator to contact the participants during the two-phase commit.
- **Coordinator Service:** When a participant gets a Prepare request from the coordinator, it places its vote by invoking the coordinator service. The participants also use this service to notify the coordinator their acknowledgments to the commit/abort request. The participants obtain the endpoint reference of the Coordinator Service during the registration step.
- **Completion Service:** The initiator invokes this service to notify the coordinator to start a distributed commit. The Completion service, together with the Completion Initiator service on the participant side, implements the WS-AtomicTransaction completion protocol. The endpoint reference of the Completion Service is returned to the initiator during the registration step.

The participant provides the Participant Service, which allows the coordinator to solicit votes from, and to send the transaction outcome to the participants according to the two-phase commit protocol.

The Completion Protocol is used by the completion initiator to initiate the atomic termination of a transaction. When the initiator decides that it is time to commit the transaction, it sends a Commit request to the coordinator. The coordinator

will then launch an instance of the Two-Phase Commit (2PC) protocol to carry out the coordination for atomic commitment of the transaction. When the 2PC completes, the coordinator notifies the initiator the outcome of the transaction (i.e., Committed or Aborted). If the request from the initiator is Rollback instead, the coordinator will abort the transaction directly.

The 2PC Protocol is used by the coordinator and participants to guarantee atomic commitment of a transaction, and it executes in two phases. During the first phase, i.e., the prepare phase, the coordinator sends a Commit request to all registered participants soliciting their votes. When the coordinator receives votes from all participants, or a timeout has occurred, it starts the second phase, i.e., the commit phase, to notify the participants the outcome of the transaction.

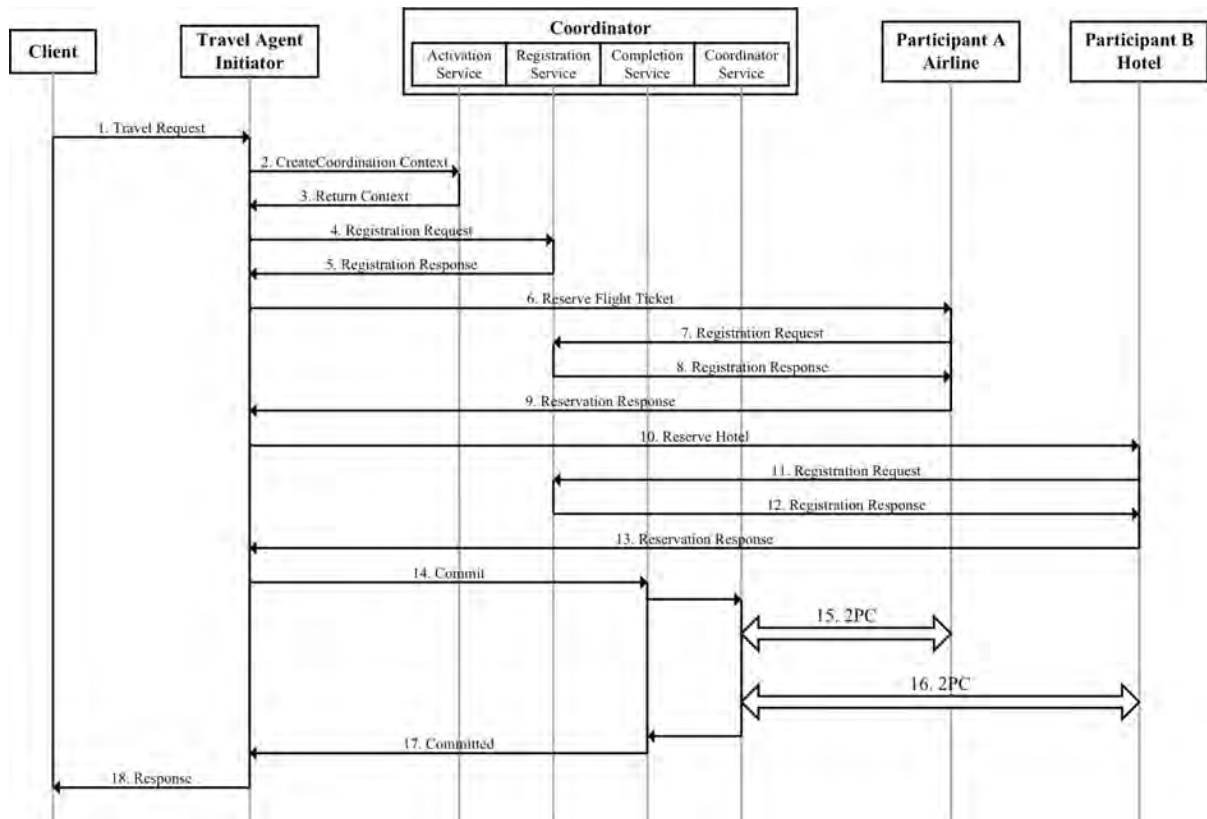
2PC has two variants used for different resources, Volatile 2PC and Durable 2PC. Volatile 2PC is used for volatile resources such as caches and Durable 2PC focuses on durable resources like a database. Participants must register in an appropriate protocol before the termination of the transaction. A participant can register in more than one protocol. Upon receiving a Commit request from the initiator in the completion protocol, the coordinator begins the prepare phase first for every participant who has registered in the Volatile 2PC protocol by sending a Prepare request to them before it begins the prepare phase for Durable 2PC. The participant that gets the request must respond with a Prepared, Aborted or ReadOnly message based on its own decision. During this period, other participants can continuously register with the coordinator for Durable 2PC, but the registration progress has to be done by the coordinator before the start of the first phase for durable resources. After the prepare phase for Volatile 2PC, the coordinator begins to take care of the prepare phase for Durable 2PC participants. Same as the prepare phase of Volatile 2PC, all participants in the Durable 2PC protocol have to

respond appropriately before the protocol advances to the second phase, in which the coordinator will issue the Commit requests to all participants for both Volatile and Durable 2PC protocols if all participants have provided positive feedbacks. If there are any negative votes, even if only one, the coordinator has to abort the transaction. After the participants get a Commit notification, they will commit the transaction and send the Committed acknowledgement back to the coordinator.

Figure 1 shows an example of a Web services atomic transaction for a travel reservation coordinated by WS-AtomicTransaction. In this example, a client contacts a Travel Agent to make travel arrangement. The Agent, which acts as the completion initiator, is responsible to make flight and hotel reservations in the context of an atomic distributed transaction, on behalf of the client. We assume that the Agent relies on a Flight reservation Web service and a Hotel reservation Web service, for booking a plane ticket and a hotel room for the traveler.

To begin this transaction, the client sends a request to the initiator (step 1). The initiator invokes the Activation Service on the Coordinator. This will create a unique coordination context for the transaction (step 2). This context contains the Endpoint Reference for the Registration Service. After the Activation step, the initiator gets the reply back with the transaction context (step 3). In next step, the initiator registers the Completion Initiator Service with the coordinator so that the coordinator could inform the initiator after it obtains the outcome of the transaction (steps 4 and 5). Now the booking service offered by the initiator carries out the reservations for the plane ticket and a hotel room (steps 6 and 10). The flight and hotel booking Web services must register their participant endpoint references with the coordinator (steps 7 & 8 and 11 & 12). After the registration step, these services will send their responses for the reservation requests (steps 9 and 13). Subsequently, the initiator asks the Comple-

Figure 1. An example Web services atomic transaction



tion Service to commit the transaction (steps 15 – 17). Finally, the Travel Agent sends the result back to the client (step 18).

Web Services Business Activities

WS-BusinessActivity differs from WS-Atomic-Transaction in that it focuses on the coordination of long running business activities where the atomic transaction model is not appropriate. WS-BusinessActivity is also built on top of the WS-Coordination framework. The WS-Business-Activity specification describes two coordination types, Atomic-Outcome and Mixed-Outcome, and two coordination protocols, Business-Agreement-with-Participant-Completion and Business-Agreement-with-Coordinator-Completion. Either protocol can be used with either coordination type. If the Atomic-Outcome Coordination type is used, all participants must reach an agreement

about the activity outcome (i.e., either to close or to compensate). If the Mixed-Outcome coordination type is used, some participants may be directed to close while others to compensate. All WS-BusinessActivity frameworks must implement the Atomic-Outcome coordination type.

In a WS-BusinessActivity framework, a participant registers either one of the two protocols, which are managed by the coordinator of the business activity. The only difference between the two protocols is that the Business-Agreement-with-Participant-Completion protocol assumes that the participants know when the coordinator has completed all the processing related to a business activity and the Business-Agreement-with-Coordinator-Completion protocol notifies the participants when the coordinator has received all requests.

A participant who has registered the Business-Agreement-with-Participant-Completion protocol

informs its coordinator by sending a Completed notice when it has done its entire works for a business activity. The coordinator should reply with either Close or Compensate message depends on the circumstance. The participant receives a Close instruction if the activity has completed successfully. If it gets a Compensate instruction instead, the coordinator will undo the completed work and will have to restore the data recorded from the initial condition. The participant may encounter a problem or fail during the processing of the activity, in which case, it must signal the coordinator with an error message. If it gets the Fail message, the coordinator will acknowledge the participant by a Failed notification.

Upon receiving a CannotComplete notification, the coordinator learns that the participant cannot successfully finish its work. On sending out this message, the participant discards all its pending work and cancels all related executions, and exits the current business activity. On receiving such a message, the coordinator is required to notify the participant with a NotCompleted message. In the active state, the coordinator could cancel any transaction by using the Cancel notification, and the participant will respond with a Canceled message if it receives the message.

In the Business-Agreement-with-Coordinator-Completion protocol, the completion decision comes from the coordinator. The coordinator sends a Complete message to the participants informing them that they won't receive any new requests within the current business activity and it is time to complete the processing. The participant then replies with a Completed message if it could successfully finish its work

BYZANTINE FAULT TOLERANT WEB COORDINATION

In this section, we first introduce the general Byzantine fault tolerance technique. It is followed by a discussion on how to apply the BFT technique for Web services coordination within minimum

runtime overhead by exploiting the semantics of the Web services coordination.

Byzantine Fault Tolerance

A general purpose BFT replication algorithm is designed for a client-server architecture where the clients issue requests to the replicated server replicas. A Byzantine faulty replica may use behave arbitrarily to prevent the normal operations of a replica. In particular, it might propagate conflicting information to other replicas or components that it interacts with. To tolerate f Byzantine faulty replicas in an asynchronous environment, we need to have at least $3f+1$ number of replicas (Castro & Liskov, 2002). One replica is designated as the primary and the remaining replicas are backups.

The client first sends its request to the primary replica. The primary then broadcasts the request message to the backups and also determines the execution order of the message. To prevent a faulty primary from intentionally delaying a message, the client starts a timer after it sends out a request. It waits for $f+1$ identical replies from different replicas. Because at most f replicas are faulty, at least one reply must come from a correct replica. If the timer expires before it receives a correct reply, the client broadcasts the request to all server replicas. This enables the correct replicas to detect the primary failure so that a new primary can be elected (this is often called a view change).

All correct replicas must agree on the same set of input messages with the same execution order. In other words, the request messages must be delivered to the replicas reliably in the same total order. A BFT replication algorithm, such as PBFT (Castro & Liskov, 2002) is designed to meet this objective.

During the normal operation, the PBFT algorithm uses three phases to commit the total order of a request. During the first phase, often referred to as the pre-prepare phase, the primary multicasts to the backups a pre-prepare message containing the client's request, the current view number, and the sequence number of the request.

A backup verifies the pre-prepare message and control information proposed by the primary for the purpose message ordering and possibly for controlling nondeterministic operations. If the backup accepts the message, it starts the prepare phase (i.e., the second phase) by multicasting to the backups a prepare message containing the ordering information and the digest of the request message being ordered. A replica (including the primary) waits until it has collected matching prepare messages from $2f$ other replicas and the corresponding pre-prepare message. It then starts the commit (i.e., the third phase) by multicasting a commit message to other replicas. The commit phase ends when a replica has received matching commit messages from $2f+1$ replicas (including the one it has sent). At this point, the request message is totally ordered. The message can be delivered to the server application if the replica has delivered all previously ordered requests.

Applying the BFT Techniques Smartly to Web Services Coordination

Because the interactions between the coordinator (replicated) and other components in Web services atomic transactions and business activities must conform to the WS-AtomicTransaction and WS-BusinessActivity standards, there is no reason not to exploit the semantics specified in these standards to significantly reduce the runtime overhead of achieving Byzantine fault tolerance. The specific semantics of the Web services coordination enables us to apply the following optimizations (Chai et al., 2013; Zhang et al., 2012):

Partitioning of the requests. Without the knowledge of application semantics, all requests received at the replicated coordinator would have to be totally ordered and executed sequentially according to that total order. This would severely limit the system throughput and increase the end-to-end latency experienced by clients under heavy load. With the knowledge of the Web ser-

vices coordination context, however, we can see that the requests that belong to different atomic transactions or business activities are handled by different coordinator objects, hence, they can be executed in parallel without causing inconsistency among the replicas.

Using source ordering instead of total ordering for business activities. For business activities, requests within different business activities are handled independently by different coordinator objects. Hence, their relative ordering does not result in conflicting state transitions of the Coordination service. These include all requests for the Registration and Coordinator services, such as the Registration requests. For the Activation requests, even though they are handled by the same object, their relative ordering does not affect how the coordinator objects are created. One concern is the activity context id, which must be unique for each business activity. We modified the implementation such that the coordinator would use a client-supplied unique id to ensure replica consistency. Hence, the requests are commutative with each other even if they belong to the same business activity. Hence, they can be processed in parallel.

Deferred Byzantine agreement. Even for commutative requests, all nonfaulty replicas must deliver the same set of requests to ensure replica consistency. Even though the mechanism to ensure this is lightweight, it would nevertheless incur a performance penalty if the number of commutative requests is large. For a session-oriented application, such as an application uses WS-AtomicTransaction or WS-BusinessActivity, it is possible to defer the Byzantine agreement for the set of commutative requests at the end of the session to reduce the runtime overhead. More specifically, in a Web service atomic transaction, the agreement on the registration requests, which are commutative with each other, are deferred until the transaction commitment time. Furthermore, the Byzantine agreement on the participants set is combined with that for the transaction outcome.

FUTURE TRENDS

The BFT mechanisms described in the previous section are specifically handcrafted for Web services coordination. Although the mechanisms are optimal as a way to enhance the trustworthiness of Web services coordination, they cannot be directly applied to other distributed systems. We see two future trends with respect to developing techniques to enhance Web services and distributed applications in general: (1) automating the application semantics discovery; and (2) reconstructing applications using commutative data structures or software transactional memory.

We have been exploring methods to automate the discovery of application semantics, including the dependency analysis of interface specifications, such as that written by the Business Process Execution Language (Fu et al., 2004; Ouyang et al., 2007) for business transactions, source code static analysis, and modeling and pattern recognition of messages exchanges. We have documented a number of interaction patterns as the foundation for automated analysis of application semantics for Byzantine fault tolerance (Chai and Zhao, 2012).

An alternative approach to implementing highly efficient solution for Byzantine fault tolerance is by using commutative data types (Shapiro et al., 2011; Prego et al., 2009). Because all operations on the replicas are commutative, the need for carrying out Byzantine agreement is minimized. We have designed a system that requires only periodic synchronization of operations (Chai and Zhao, 2014). Similarly, the server application can be constructed using software transactional memory (Shavit & Touitou, 1997), which enables automatic analysis for possible concurrent processing of requests without application semantics (Zhang and Zhao, 2012). The biggest benefit for this approach is that the solution is ignorant to application semantics, which means any solution designed will be readily applicable to other applications written using commutative data types. The down side of this approach is that the applications would have to be redesigned and implemented using commutative data types.

CONCLUSION

In this article, we introduced Web services coordination and its applications to Web services atomic transactions and business activities, and presented a overview of our recent works on enhancing the trustworthiness of Web services coordination for business activities and transactions. We argue that even though it is possible to use a general purpose BFT replication algorithm for Byzantine fault tolerant replication of a Web services coordination service, doing so naively would incur unacceptable runtime overhead. Instead, by exploiting the application semantics, much lighter weight mechanisms can be designed to enhance intrusion tolerance, and hence the trustworthiness of systems that require Web services coordination.

REFERENCES

- Castro, M., & Liskov, B. (2002). Practical Byzantine fault tolerance and proactive recovery. *ACM Transactions on Computer Systems*, 20(4), 398–461. doi:10.1145/571637.571640
- Chai, H., Zhang, H., Zhao, W., Melliar-Smith, P. M., & Moser, L. E. (2013). Toward Trustworthy Coordination of Web Services Business Activities. *IEEE Transactions on Services Computing*, 6(2), 276–288. doi:10.1109/TSC.2011.57
- Chai, H., & Zhao, W. (2012). Interaction Patterns for Byzantine Fault Tolerance Computing. In *Computer Applications for Web, Human Computer Interaction, Signal and Image Processing, and Pattern Recognition* (pp. 180-188). Springer Berlin Heidelberg. doi:10.1007/978-3-642-35270-6_25
- Chai, H., & Zhao, W. (2014, June). Byzantine fault tolerance for services with commutative operations. In *Proceedings of the 2014 IEEE International Conference on Services Computing* (pp. 219-226). IEEE. doi:10.1109/SCC.2014.37

Chakrabarty, S. (2007). Strategies for business process outsourcing: An analysis of alternatives, opportunities and risks. In *E-Business Process Management: Technologies and Solutions* (pp. 204-229). Hershey, PA: IGI Publishing.

Freund, T., & Little, M., (2007). *Web Services Business Activity (WS-BusinessActivity) Version 1.1*. OASIS Standard.

Fu, X., Bultan, T., & Su, J. (2004, May). Analysis of interacting BPEL web services. In *Proceedings of the 13th international conference on World Wide Web* (pp. 621-630). ACM.

Gray, J., & Reuter, A. (1993). *Transaction Processing: Concepts and Techniques*. Morgan Kaufmann Publishers.

Lamport, L., Shostak, R., & Pease, M. (1982). The Byzantine generals problem. *ACM Transactions on Programming Languages and Systems*, 4(3), 382-401. doi:10.1145/357172.357176

Little, M., & Wilkinson, A. (Eds.). (2007). *Web Services Atomic Transaction (WS-AtomicTransaction) Version 1.1*. OASIS Standard.

Ouyang, C., Verbeek, E., Van Der Aalst, W. M., Breutel, S., Dumas, M., & Ter Hofstede, A. H. (2007). Formal semantics and analysis of control flow in WS-BPEL. *Science of Computer Programming*, 67(2), 162-198. doi:10.1016/j.scico.2007.03.002

Papazoglou, M. P. (2003). Web Services and Business Transactions. *World Wide Web. Internet and Web Information Systems*, 6(1), 49-91.

Preguica, N., Marques, J. M., Shapiro, M., & Letia, M. (2009, June). A commutative replicated data type for cooperative editing. In *Proceedings of the 29th IEEE International Conference on Distributed Computing Systems* (pp. 395-403). IEEE. doi:10.1109/ICDCS.2009.20

Shapiro, M., Preguica, N., Baquero, C., & Zawirski, M. (2011). *A comprehensive study of convergent and commutative replicated data types. Technical report*. INRIA.

Shavit, N., & Touitou, D. (1997). Software transactional memory. *Distributed Computing*, 10(2), 99-116. doi:10.1007/s004460050028

Zhang, H., Chai, H., Zhao, W., Melliar-Smith, P. M., & Moser, L. E. (2012). Trustworthy coordination of Web services atomic transactions. *IEEE Transactions on Parallel and Distributed Systems*, 23(8), 1551-1565. doi:10.1109/TPDS.2011.292

Zhang, H., & Zhao, W. (2012). Concurrent Byzantine Fault Tolerance for Software-Transactional-Memory Based Applications. *International Journal of Future Computer and Communication*, 1(1), 47-50. doi:10.7763/IJFCC.2012.V1.14

Zhao, W., Moser, L. E., & Melliar-Smith, P. M. (2008). A reservation-based extended transaction protocol. *IEEE Transactions on Parallel and Distributed Systems*, 19(2), 188-203. doi:10.1109/TPDS.2007.70727

ADDITIONAL READING

Benatallah, B., Dumas, M., & Sheng, Q. Z. (2005). Facilitating the rapid development and scalable orchestration of composite web services. *Distributed and Parallel Databases*, 17(1), 5-37. doi:10.1023/B:DAPD.0000045366.15607.67

Chai, H., & Zhao, W. (2013). Byzantine fault tolerance for session-oriented multi-tiered applications. *International Journal of Web Science*, 2(1), 113-125. doi:10.1504/IJWS.2013.056578

Deswarte, Y., & Powell, D. (2006). Internet security: An intrusion-tolerance approach. *Proceedings of the IEEE*, 94(2), 432-441. doi:10.1109/JPROC.2005.862320

- Fensel, D., Lausen, H., Bruijn, J., & Polleres, A. (2007). *Enabling semantic web services: the web service modeling ontology*. Springer. doi:10.1007/978-3-540-34520-6
- Garcia, M., Bessani, A., Gashi, I., Neves, N., & Obelheiro, R. (2011, June). OS diversity for intrusion tolerance: Myth or reality? In *Proceedings of the IEEE/IFIP 41st International Conference on Dependable Systems & Networks* (pp. 383-394). IEEE.
- Mahajan, S., & Singhal, R. (2009, December). Azvasa: Byzantine Fault Tolerant Distributed Commit with Proactive Recovery. In *Proceedings of the 2nd International Conference on Emerging Trends in Engineering and Technology* (pp. 659-663). IEEE. doi:10.1109/ICETET.2009.44
- Malik, Z., & Bouguettaya, A. (2009). Ratweb: Reputation assessment for trust establishment among web services. *The VLDB Journal—The International Journal on Very Large Data Bases*, 18(4), 885-911.
- Min, B. J., & Choi, J. S. (2004). An approach to intrusion tolerance for mission-critical services using adaptability and diverse replication. *Future Generation Computer Systems*, 20(2), 303-313. doi:10.1016/S0167-739X(03)00146-8
- Montagut, F., Molva, R., & Golega, S. T. (2008). Automating the composition of transactional web services. *International Journal of Web Services Research*, 5(1), 24-41. doi:10.4018/jwsr.2008010102
- Pal, P., Webber, F., Schantz, R. E., & Loyall, J. P. (2000, October). Intrusion tolerant systems. In *Proceedings of the IEEE Information Survivability Workshop* (pp. 24-26).
- Rajan, H., & Hosamani, M. (2008). Tisa: Toward trustworthy services in a service-oriented architecture. *IEEE Transactions on Services Computing*, 1(4), 201-213. doi:10.1109/TSC.2008.18
- Sousa, P., Bessani, A. N., Correia, M., Neves, N. F., & Verissimo, P. (2007, December). Resilient intrusion tolerance through proactive and reactive recovery. In *Proceedings of the 13th Pacific Rim International Symposium on Dependable Computing* (pp. 373-380). IEEE. doi:10.1109/PRDC.2007.52
- Sousa, P., Bessani, A. N., Dantas, W. S., Souto, F., Correia, M., & Neves, N. F. (2009, June). Intrusion-tolerant self-healing devices for critical infrastructure protection. In *Proceedings of the IEEE/IFIP International Conference on Dependable Systems & Networks*. (pp. 217-222). IEEE. doi:10.1109/DSN.2009.5270333
- Yang, S. J., Lan, B. C., & Chung, J. Y. (2005, February). A trustworthy Web services framework for business processes integration. In *Proceedings of the 10th IEEE International Workshop on Object-Oriented Real-Time Dependable Systems* (pp. 186-193). IEEE. doi:10.1109/WORDS.2005.13
- Young, M., Kate, A., Goldberg, I., & Karsten, M. (2013). Towards Practical Communication in Byzantine-Resistant DHTs. *Networking, IEEE/ACM Transactions on*, 21(1), 190-203.
- Zhao, W. (2007, October). BFT-WS: A Byzantine fault tolerance framework for web services. In *Proceedings of the Eleventh International IEEE EDOC Conference Workshop* (pp. 89-96). IEEE. doi:10.1109/EDOCW.2007.6
- Zhao, W. (2007) A Byzantine fault tolerant coordination for Web services atomic transactions. *Proceedings of the 3rd IEEE International Symposium on Dependable, Autonomic and Secure Computing* (pp. 37-44). Columbia, MD. Chai, H., & Zhao, W. (2012). Byzantine Fault Tolerance as a Service. In *Computer Applications for Web, Human Computer Interaction, Signal and Image Processing, and Pattern Recognition* (pp. 173-179). Springer Berlin Heidelberg. doi:10.1109/DASC.2007.10

Zhao, W. (2009). Design and implementation of a Byzantine fault tolerance framework for Web services. *Journal of Systems and Software*, 82(6), 1004–1015. doi:10.1016/j.jss.2008.12.037

Zhao, W. (2013, April). Towards practical intrusion tolerant systems. In *Proceedings of the IET International Conference on Information and Communications Technologies* (pp. 280-287). IET.

Zhao, W. (2014). *Building dependable distributed systems*. John Wiley & Sons. doi:10.1002/9781118912744

Zhao, W. (2015). Optimistic Byzantine fault tolerance. *International Journal of Parallel, Emergent and Distributed Systems*, 1-14 (preprint).

Zhao, W., Moser, L. E., & Melliar-Smith, P. M. (2005). A reservation-based coordination protocol for Web services. *Proceedings of the IEEE International Conference on Web Services* (pp. 49-56). Orlando, FL. doi:10.1109/ICWS.2005.14

KEY TERMS AND DEFINITIONS

Atomic Transaction: An atomic transaction in the context of Web services refers to a distributed transaction to be executed atomically. It should exhibit the atomicity, consistency, isolation and durability properties, just like a local transaction.

Business Activity: A business activity is usually long running and requires flexible outcomes. A Web service business activity must conform to the

WS-BusinessActivity specification and adopts one of two two coordination types, Atomic-Outcome and Mixed-Outcome, and one of two coordination protocols, Business-Agreement-with-Participant-Completion and Business-Agreement-with-Coordinator-Completion. Either protocol can be used with either coordination type.

Business Process Execution Language (BPEL): It is a computer language used to describe the actions and execution order of these actions within a business process. It is designed specifically for the Web services paradigm. A BPEL for Web services was standardized by OASIS.

Byzantine Fault Tolerance: A replication-based technique used to ensure high availability of an application subject to Byzantine fault.

Byzantine Fault: It is used to model arbitrary fault. A Byzantine faulty process might send conflicting information to other processes to prevent them from reaching an agreement.

Distributed System: A distributed system is a computer network system, shown to end users as a single machine but actually work with a set of independent computers connected.

Replica Consistency: The states of the replicas of an application should remain to be identical at the end of the processing of each request. Replica consistency is necessary to mask a fault in some replicas.

Web Services: The Web services technology refers to the set of standards that enable automated machine-to-machine interactions over the Web. The core standards include XML, HTTP, SOAP, WSDL and UDDI.

Usability of CAPTCHA in Online Communities and Its Link to User Satisfaction

Samar I. Swaid

Philander Smith College, USA

INTRODUCTION

CAPTCHA- Completely Automatic Public Turing test to Tell Computers and Humans Apart- is a security application that is used by many websites to avoid spamming and hacking. CAPTCHA provides a test that humans are able to solve, but computer programs cannot, to defeat websites and stop spammers and hackers. CAPTCHA test can be viewed as a function of random input that generates a challenge test and a solution. Previous research found that website users spend on average 10 seconds per CAPTCHA (Sutherland, 2012), which suggests for 200 million CAPTCHA that are solved by human users every day, more than 5000,000 hours on daily basis are lost productivity (Sutherland, 2012). CAPTCHA is developed as an instrument to limit misuse of websites that offer free services of email creation, weblogs posting, social networking, online voting, online games online banking, or chat rooms, etc. With this aim in mind, several CAPTCHAs (Boshmaf et al., 2012; Coats and Baird, 2001; Chellapilla and Simard, 2005a; Von Ahn and Blum, 2005) are developed. CAPTCHAs share a number of common characteristics: (i) automated test; (ii) publicly available; (iii) separability of test- generation and solution; and (iv) defeating attacks (Fidas et al., 2011, Swaid, 2013). A number of different techniques for generating CAPTCHAs have been developed, each satisfying the criteria of CAPTCHA test described above.

CAPTCHA types can be categorized to five main types: (i) text CAPTCHA such as reCAPTCHA and Gimby CAPTCHA (ReCaptcha, 2015);

DOI: 10.4018/978-1-5225-2255-3.ch702

(ii) image-based CAPTCHA such as ASIRRA (Elson et al., 2007), Bongo (Bongo, 2015) and Pix (Pix, 2015) that rely on image-recognition techniques (Vikram et al., 2011); (iii) animation-based CAPTCHA; (iv) audio CAPTCHA (e.g., ReCaptcha Audio Captcha, Digg, and (v) other (e.g., mathematical functions, games, multiple choice, and cognitive based CAPTCHA) (Hernandez-Castro and Ribagorda, 2010; Yamamoto et al., 2010). Text CAPTCHA is the most common type where the user is presented with a challenge-response test in a form of numbers and letters, and user needs to strike in the right characters in the given text box. In this chapter, the author uses text CAPTCHA and CAPTCHA alternatively focusing on usability of text-based CAPTCHA that is formed of randomly generated sequence of letters and/or numbers that appear as a distorted image.

Cracking CAPTCHAs

Cracking CAPTCHA and bypassing the CAPTCHA tests have attracted a number of studies to examine methodologies used in solving CAPTCHA. Generally, text CAPTCHA cracking needs three steps. First phase is pre-processing CAPTCHA by removing background, color and added noise. Second, a segmentation process is used to locate individual character in CAPTCHA challenge. Finally, a classification or recognition phase is applied to solve CAPTCHA, such as using standard OCR software (Serrao, et al., 2013; Sutherland, 2012), CAPTCHA farms (Serrao, et al., 2013), or CAPTCHA smuggling (Egele et al., 2010). In addition, there are other alternatives of

commercial software that are developed to crack CAPTCHA such as Death BY CAPTCHA, CAPTCHA Sniper, Automated CAPTCHA Bypass, CAPTCHA Monster or PWNTCHA (Serro et al., 2013). Exploring mechanisms to crack CAPTCHA and its relation to usability is also important and of practical relevance, but beyond the scope of this chapter.

CAPTCHA Usability and Online Communities

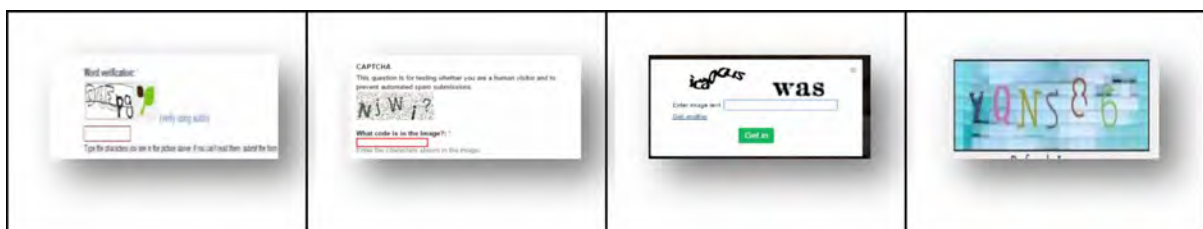
In general, we may define the concept of usability as the effort required to use a computer system. For instance, Nielsen (2003) suggests that usability concerns several aspects such as the ease with which the user is capable of learning to manage the system, the ease of memorizing the basic functions, the grade of efficiency with which the site has been designed, the degree of error avoidance and the general satisfaction of the user in terms of manageability. Therefore, greater levels of usability will be associated to lower levels of difficulty to manage that functionality (Davis, 1993) and, as a result, usability is traditionally considered a key factor for predicting intentions to use a system (e.g. Davis, 1993). More specifically, text CAPTCHA usability might reflect perceived ease of solving CAPTCHA test (Swaid, 2013) explaining satisfaction. Designing successful CAPTCHA that is unsolvable by CAPTCHA solving software, yet, a user friendly one, is a challenging task (Ahmad and Yan, 2010). Several proprieties to make CAPTCHA resistant to Optical Character Recognition (OCR) software can make CAPTCHA unusable

(Sutherland, 2012) and therefore, suggest users' dissatisfaction.

Today, millions of people join online social gathering spheres to chat, to debate, to play games, to ask for information, or to find social support (Preece, J. and Maloney-Krichmar, 2003). Some forms of such online gatherings come in the form of forums, discussion groups, and social networking sites among others. These online social gatherings are known by a variety of names including 'online community', that is described as 'cultural aggregations that emerge when enough people bump into each other often enough in cyberspace' (Rheingold, 1994, p. 57). To eliminate automated spammers and other computer agents who are attempting to violate community norms, online communities ask applicants to complete a CAPTCHA test before subscribing. Studies showed that new comers to online communities are unable to join due to the entry barrier posed by visual CAPTCHAs (Chandrasheka, 2010).

A CAPTCHA that is robust, yet, unsolvable by humans is useless. The issue of usability has been studied focusing on the functional level (Ahmad and Yan, 2010; Burstein et al., 2010), however, the ultimate effect of CAPTCHA usability on legitimate goal-oriented users is not well documented in literature (Motoyama et al. 2010). According to Ellson et al. (2007), even relatively simple challenge can drive away a substantial number of potential customers to websites. Thus, a methodological analysis is needed for CAPTCHA schemes used in websites for online communities to understand the correlations between design characteristics and users satisfaction.

Figure 1.



Background

Over the past decade, a number of CAPTCHAs have been created to differentiate humans from computer programs. Unsurprisingly, CAPTCHAs have been evaluated in terms of its robustness (Azad and Jain, 2013; Chew and Baird, 2003; Chellapilla et al., 2005a, 2005b), its economic context (Motoyama et al., 2010), but limited number of studies focused on usability issues of text CAPTCHA. One study by Ahmad and Yan (2008) analyzed CAPTCHA technically to find attributes of content, presentation and distortion to explain usability. In another study by the same authors they shed light on the role of color on CAPTCHA

usability (Ahmad and Yan, 2010) (see Table 1). A study by Sutherland (2012) focused on analyzing different properties of layout CAPTCHA in terms of background color and integration with the website. In an attempt to maximize the usability of CAPTCHA, security features of visual, anti-segmentation and anti-recognition were identified and applied to create two new CAPTCHA schemes for Google (Bursztein et al., 2014; 2011). In the work of Swaid (2013), a think-a-loud protocol and a sorting task are conducted to understand the dimensionality of CAPTCHA usability. The study applied a formative approach in operationalizing the latent construct of usability (see Table 1).

Table 1.

Reference	Focus/Research	Usability Issues in CAPTCHA Design
Yan and El-Ahamad (2008)	Text-CAPTCHA/Content Analysis and	-Distortion (distortion method and level, confusing characters) -Content (Character set, string length, predictability, randomness of letters, , offensiveness of words) -Presentation (front type, and size, image size, use of color, integration with web pages.
Ahmad and Yan (2010)	Text-CAPTCHA color use/Content analysis and Experiment	Complicated background color effects negatively both security and usability Using few colors (e.g., two colors) is recommended for both usability and security of CAPTCHA
Bursztein (2011)	Text-based Captcha/ content analysis and experimentation	Anti-Recognition -Use of Multi font: Use of multiple font-face - Use of character-set: Character-set scheme used -Use of font size: Using variable font size -Use of Distortion: attractor fields - Use of blurring: Use of blurring letters - use of tilting: rotating characters with various angles - use of waving: Rotating characters in a wavy fashion Anti-Segmentation -Complex background: try to hide text in complex background -Lines: adding lines to prevent solver from knowing the real character segments - Collapsing: Remove Space between characters to prevent segmentation
Sutherland (2012)	Text CAPTCHA/content analysis of several CAPTCHA	Character set size(size of character set, types of characters; lower and upper case characters; String length(number of characters, dictionary word; fixed length vs. variable length scheme);distortion (variety of fonts styles, complex background, stray lines types, collapsing)
Swaid (2013)	Text-CAPTCHA/Think-Al-Loud Protocol and Card-Sorting task	Content :Extent contents embedded in CAPTCHA is understandable Format: Extent CAPTCHA content is formatted and presented to users Distortion: Extent distortion is applied to CAPTCHA's text; Services: Extent CAPTCHA provides services to facilitate solving CAPTCHA
Burztein et al.,(2014)	Text CAPTCHA/ Content analysis and experiment	-Visual: character set, character count, font size, font families, foreground color, background color, -Character overlap, random dot size, random dot count, line types, lines count, line positions, similar foreground/background colors - Rotated character count, rotated character degrees, vertical shifting sizes, character size variations, character distortion

CAPTCHA USABILITY: THE STUDY

To understand the relationships among dimensions of CAPTCHA usability and behavioral variables of user satisfactions, the study is designed as two-phase one. First phase would focus on surveying the top 50 most popular online communities' websites to identify CAPTCHA use; while second phase would collect users' perceptions on CAPTCHA usability and their satisfaction with the registration process.

CAPTCHAs in Top 50 Online Communities: A Content Analysis Approach

To gather exiting text CAPTCHAs exit in top 50, most visited online communities websites, the author uses Alexa (Alexa.com, 2015), a web tracking and statistics gathering service, to determinate the top 50 most popular websites for online communities visited as of February 2015. Of the top 50 online communities, the ones that follow the criteria below are selected:

1. Website is one of the online communities' platforms.
2. Website offers registration, blogging, or somehow updating activity.
3. No membership fee is required.
4. English-speaking website.
5. Challenge users with static text CAPTCHA. Text CAPTCHAs with animation or other types of CAPTCHAs (e.g., CAPTCHA used by Yahoo) are excluded from the study.

Applying the criteria listed above, the author confirmed findings of previous study on CAPTCHA usability using qualitative approach (Swaide, 2013). Analysis revealed that there are four main text CAPTCHAs used by the most popular online communities websites (see Table 2). Unsurprisingly, ReCAPTCHA was the dominant CAPTCHA used by most of the online communities, while SolevMedai CAPTCHA and Mollom were realized in less number of online communities. This is in line with other market analysis reports that found ReCAPTCHA market share is about 99% (WebAnalyzer, 2016). Some of the CAPTCHAs were excluded for reasons of animation added, non-text interactive CAPTCHA or image orientation CAPTCHA (see Figure 2). These mainly were found at Yahoo.com, Namepros.com and Care2.com. Next, a content analysis to CAPTCHA is applied, as follows.

Content analysis as a technique is used to examine artifacts of social communications that might come in the form of written documents, multimedia content and any form of transcriptions of recorded verbal. As Holsti noted, content analysis is "any technique for making inferences by systematically and objectively identifying special characteristics of messages". Based on this definition, content analysis can be applied on messages that come in form of written text, photographs, and/or multimedia content, to identify specific characteristics of the message. Similarly, Krippendorff defined content analysis as a research technique for making valid and reliable inferences from data to their context (Krippendorff, 2013). Weber (1990) defined content analysis as research methodology that utilizes a set of procedures to

Figure 2.



make valid inferences from text. Content analysis is most conveniently used with textual types of data such as online comments, online content, blogs, and open-ended survey questions (Krippendorff, 2013).

Dimensions of CAPTCHA Usability

Text CAPTCHA found in online communities exhibits four main dimensions that can be used to measure the usability of CAPTCHA. Based on previous research (i.e., Al-Ahmad and Yah, 2010, Burztein et al., 2014; Swaid, 2013), it is suggested that usability of CAPTCHA is a function of four dimensions: Content, Visual Layout, Distortion and Service. To understand the relationships among dimensions of CAPTCHA usability and perceived usability and satisfaction with the registration process, dimensions of CAPTCHA usability are developed as reflective latent constructs represented by at least three measures (Hatcher,

1994) (see Table 3). The conceptual model of the study also examines the effect of perceived usability on satisfaction (see Figure 3).

Following the recommendation of Churchill (1979), this study adopts the following steps: (i) latent constructs are conceptualized by defining the domain of the constructs; (ii) dimensions are operationalized focusing on the content validity of the dimensions, (iii) data are collected using the developed scale, (iv) reliability analysis and factor analysis are applied to assess the reliability and validity of the developed model. Endogens variable of overall CAPTCHA usability is adopted from the System Usability Scale (SUS) modified by Lewis and Sauro (2009). Satisfaction with registration process of website is adopted from Flavia et al. (2006). Confirmatory factor analysis (CFA) and Structural Equation Modeling (SEM) are conducted to examine the reliability and validity of the measurement model and to test associations hypothesized in research model.

Table 2.

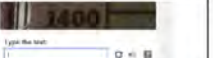
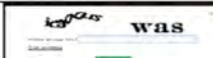
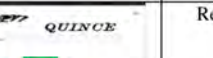
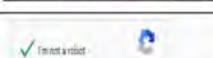
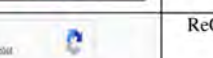
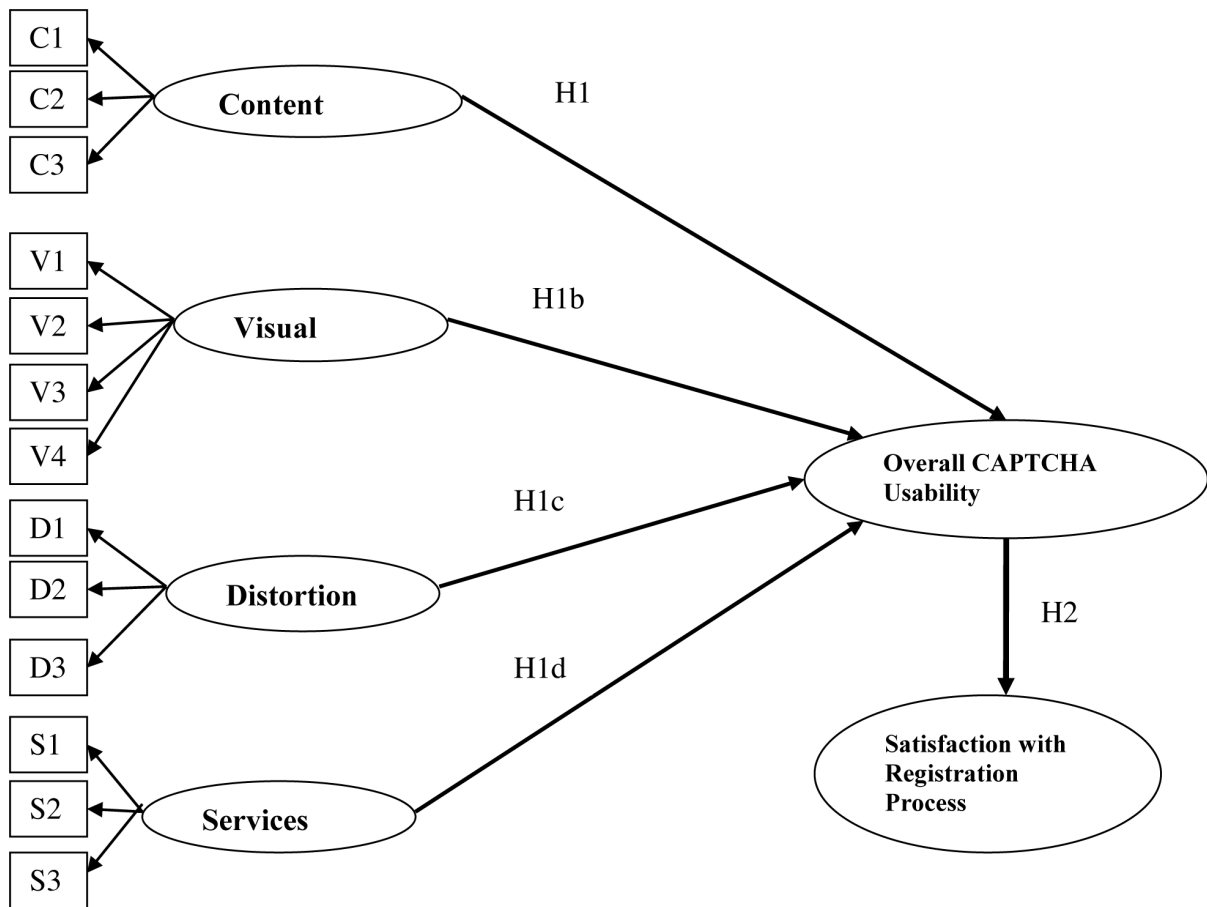
Website	Test1	Test2	Test3	Developed
www.google.com				ReCaptcha
www.Livejournal.com				SolveMedia
www.oxwall.com				OXWALL
www.Nexopia.com				ReCaptcha
Twitter.com				Mollom
www.xing.com				ReCaptcha
www.fotolog.com				Developed by SolveMedia
www.Wayn.com				ReCaptcha
www.Fotic.com				ReCaptcha v2

Figure 3.



RESULTS

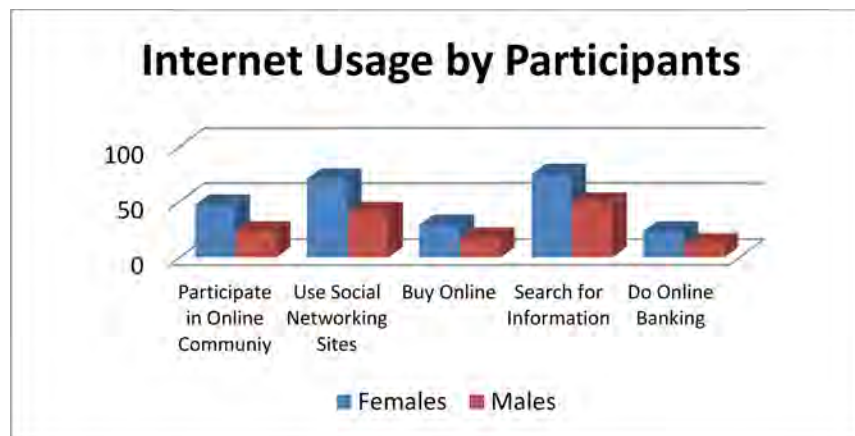
The 13-Item battery of CAPTCHA usability is arranged in a survey instrument to be measured using a seven-point Likert scale ranging from (1) strongly disagree to (7) strongly agree. In addition, a 5-item construct representing usability and 3-item construct representing satisfaction are embedded in the survey. Total of 124 college students have been invited to take part in the study for a credit in one of the universities in the Mid-South of the United States. About 60% of the participants are females, whom ages range from 19 to 35 years old. All participants have some type of participation in online communities such as forums, discussion groups, and social networking websites (see Figure 4).

The participants are first presented with questionnaire asking their experience with web browsing, experience with CAPTCHA, and level of engagement in online communities. Then, participants are asked to solve survey questions after they are challenged with three CAPTCHAs to solve.

Measurement Model

The measurement model is assessed by CFA to confirm the dimensionality of CAPTCHA usability. Chi-Square value is 2.18 less than the three recommended value by Bagoozi and Yi (1988). Other fit indexes show good fit for the measurement model. The GFI is 0.92, greater than the cut-off value of 0.90 recommended (Hatcher,

Figure 4.



1994). The adjusted goodness of fit (AGFI) is 0.90 (Bagozzi and Yi, 1988). The root mean square error of approximation (RMSEA) is 0.049, which also suggests a good fit with data collected. Reliability was assessed using Cronbach's alpha which ranges from 0.711 to 0.889, exceeding the 0.70 threshold, recommended by Nunnally and Bernstein (1994) (see Table 3). As factor loadings exceed 0.70, this indicates acceptable item convergence on its intended constructs, indicating validity (Bagozzi and Yi, 1988).

Structural Model

Hypotheses of research model are tested using the structural equation modeling. Overall fit statistics suggest that the model has adequate model fit (GFI = 0.91; AGFI = 0.90; CFI = 0.91; RMSEA = 0.043). The model fit indexes all exceeded their respective acceptance levels, indicating that the structural model fits the data well. In order to test research hypotheses, structural coefficients for model paths are calculated (see Table 4).

Discussion

This study examined the dimensionality of text CAPTCHA usability and examined how CAPTCHA usability dimensions affect perception of CAPTCHA usability users' satisfaction with

registration process. The analytical results of this study are discussed below.

First, the analytical results shows that CAPTCHA usability can be explained by CAPTCHA content, visual layout, distortion and service. Usability is most strongly affected by visual layout. Online communities thus need to ensure that visual layout introduced by CAPTCHA text is appropriate and solvable to users. Second, the dimension of content significantly affects CAPTCHA usability perception. This finding might be explained by the fact that users expect CAPTCHA with content from character set and symbols they are familiar with. Third, the distortion dimension is a significant predictor of text CAPTCHA usability. Other studies that analyzed the CAPTCHA suggest that distortion might affect the perception of CAPTCHA usability. Therefore, to enhance users satisfaction with the registration process, online communities should use CAPTCHA that applies distortion technique in a way that would not affect CAPTCHA usability of CAPTCHA, and still can be used to stop spam.

Next, although service dimension had only a mild effect on CAPTCHA usability perception in this study, its importance should not be underestimated. Managers of online communities should pay careful attention to this aspect. Particularly, CAPTCHA should be designed with features that facilitate its solvability such as explaining reasons

Table 3.

Measures	Mean	Std	Factor	Construct Cronbach's Alpha
Content: C1- CAPTCHA Text is of appropriate content C2- CAPTCHA Text is something I expected C3- I feel comfortable with CAPTCHA content	4.88 4.55 4.23	0.80 0.91 0.88	0.808 0.767 0.751	0.889
Visual Layout: V1- CAPTCHA text is of appropriate visual look V2- CAPTCHA text is presented in a way that I can understand V3- CAPTCHA text is of appropriate color V4- CAPTCHA text is of appropriate design	4.44 4.34 4.01 4.09	0.78 0.89 1.01 0.78	0.883 0.834 0.821 0.818	0.878
Distortion: D1- Text CAPTCHA is distorted with appropriate level of noise D2- Text CAPTCHA is distorted with appropriate type of noise D3- Text CAPTCHA is manipulated in a way that does not make it difficult to solve it	4.80 4.56 4.03	0.80 0.77 0.86	0.854 0.834 0.831	0.823
Service: S1- CAPTCHA test provides help to solve CAPTCHA test S3- CAPTCHA test provides functions to bypass CAPTCHA test	3.89 4.33	1.22 0.98	0.721 0.801	0.711
Overall CAPTCHA Usability CU1: I think that I would like to use this CAPTCHA system again CU2: I thought the system was easy to use. CU3: I found the various functions in this system were well integrated CU4: I would imagine that most people would learn to use this CAPTCHA system very quickly CU5: I felt very confident using the system.	4.23 4.11 3.89 4.30 4.09	1.01 1.23 .099 1.14 1.15	0.891 0.867 0.877 0.801 0.859	0.801
Satisfaction with Website: SAT1: I think that I made the correct decision to use this website SAT2 The experience that I have had with this website during registration has been satisfactory SAT3: In general terms, I am satisfied with the way that this website has carried out registration process	3.99 4.33 4.29	0.89 1.12 1.09	0.800 0.779 0.780	0.811

Table 4.

Path To	Path From	Hypothesis	Structural Coefficient	t-value
Overall CAPTCHA usability	Content	H1a	0.40	2.05*
	Visual Layout	H1b	0.45	3.12*
	Distortion	H1c	0.33	3.70**
	Service	H1d	0.13	2.86*
Satisfaction with Registration Process	Overall CAPTCHA Usability	H2	0.47	5.60*
Note: * $p < 0.01$; and ** $p < 0.001$				

behind using CAPTCHA and providing users with the ability to switch to other non-text CAPTCHA (e.g., audio CAPTCHA) to enhance accessibility (Kuzma et al., 2011). Additionally, this study found a positive relationship among CAPTCHA usability and satisfaction with registration when using online communities.

IMPLICATIONS AND FUTURE RESEARCH DIRECTIONS

This study has the following implications for practitioners initiating or currently using CAPTCHA in their online communities. First, this study suggests that to enhance users' satisfaction, online communities should employ CAPTCHAs that are carefully tested in terms of its usability, addressing each of CAPTCHA dimensions of content, visual layout, distortion and service. Practitioners are advised to use the scale to measure each dimension while testing the appropriate mix of features in an iterative fashion to identify aggressive distortion, unnecessary characters or inefficient collapsing to exclude from CAPTCHA test. The scale might be used as a starting point to understand the non-linear effect of interactions (Bursztein et al., 2014) on CAPTCHA's usability. Secondly, there are variety of commercial CAPTCHA. Managers are advised to use CAPTCHAs after careful testing. Thirdly, practitioners are advised to employ CAPTCHAs that offer features enhancing accessibility such as switching to audio CAPTCHA for users with vision problems or contacting a technical person to help in bypassing CAPTCHA test. Such services would improve the perception of responsiveness in web-based service-oriented systems (Swaid and Wigand, 2009) such as online communities.

Although the study develops a scale following a systematic approach and validity testing, limitations need to be addressed. First, participants in this study are college students. Although online communities' population resembles the college students, replicating the study with non-college students might suggest additional findings. Sec-

ond, in this study, the author focused on perception of usability and satisfaction as behavioral variables. Future research can test relationships between CAPTCHA usability and other variables such as intentions to recommend and loyalty to online communities. Third, the growth of the internet and online communities will continue, and future research can replicate the study solely involving objective measures to assess solvability and usability of CAPTCHA such as time consumed and number of errors while solving CAPTCHA test. Forth, evidence exists that the relative importance and possible meaning of CAPTCHA usability may differ across cultures (Zhu and Keough, 2014). Thus, the study can be replicated to provide cross-cultural comparisons. Similarly, considering the cognitive characteristics of individuals, their learning styles (Belk et al., 2013, Germanakos et al. 2009; Grafs et al., 2005) or demographic variables (Bursztein et al., 2010) might point out other fruitful findings in terms of how to improve usability. Finally, as smartphones are widely used to access web services offered by online communities, it is worthy to address usability issues of text CAPTCHAs on smart phones.

This chapter is intended to provide an essential step in quantifying usability of text CAPTCHA contributing to the limited knowledge in this area. There is more work to be done to understand CAPTCHA usability considering a number of socio-technical and cultural challenges that relates to CAPTCHA engineering.

CONCLUSION

The conclusions drawn from this study make contributions in two main areas. First, this study is the first one to develop an dimension-based instrument to measure text CAPTCHA usability applying the reflective approach in dimension construction. Second, this study identified text CAPTCHA usability dimensions that affect CAPTCHA usability perception, which in turn is significantly related to users' satisfaction. This chapter is intended to

provide an essential step in quantifying usability of text CAPTCHA contributing to the limited knowledge in this area. Clearly, there is much more work to be done to understand CAPTCHA usability considering the multiple aspects of its dimensions that would support creating not only robust CAPTCHAs but also usable and effective ones. This, however, boils down to solving a number of socio-technical challenges that relate to CAPTCHA engineering.

REFERENCES

- Acquire. (2014). *Embrace Social Content with Confidence*. Retrieved from: <http://www.acquire.com/sites/default/files/library/attachment/mollom-data-sheet.pdf>
- Ahmad, A., & Yan, J. (2010). *Color, Usability, and Security: A Case Study*. Retrieved from: <http://www.computer.org/csdl/mags/ic/2012/02/mic2012020044-abs.html>
- Ahn, L., Blum, M., Hopper, N., & Longford, J. (2013). CAPTCHA: Using Hard AI problems for Security. *The 22 International Conference on Theory and Applications on Cryptographic*, 249-311.
- Alexa. (2015). *Actionable Analytics for the Web*. Retrieved from: <http://www.alexa.com/>
- Azad, S., & Jain, K. (2013). CAPTCHA: Attacks and Weaknesses against OCR Technology. *Global Journal of Computer Science and Technology Neural & Artificial Intelligence*, 13(3).
- Bagozzi, R., & Yi, Y. (1988). On the Evaluation of Structural Equation Models. *Journal of the Academy of Marketing Science*, 16(Spring), 74-94. doi:10.1007/BF02723327
- Belk, M., Germanakos, P., Fidas, C., Holzinger, A., & Samaras, G. (2013). Towards the Personalization of CAPTCHA Mechanisms Based on Individual Differences in Cognitive Processing. *Proceedings of the International Conference on Human Factors in Computing & Informatics*, 409-426. doi:10.1007/978-3-642-39062-3_26
- Bongo. (2015). *The CAPTCHA Project*. Retrieved from: <http://www.captcha.net/captchas/bongo/>
- Boshmaf, Y., Muslukhov, I., Beznosov, K., & Ripeanu, M. (2012). *Key Challenges in Defending Against Malicious Socialbots*. Retrieved from: <https://www.usenix.org/system/files/conference/leet12/leet12-final10.pdf>
- Burstein, E., Bethard, S., Fabry, C., Mitchell, J., & Jurafsky, D. (2010). *How Good Humans at Solving CAPTCHAs? A Large Scale Evaluation*. Retrieved from: http://web.stanford.edu/~jurafsky/burszstein_2010_captcha.pdf
- Burstein, E., Moscicki, A., Fabry, C., Bethard, S., Mitchell, J., & Jurafsky, D. (2014). *Easy Does It: More Usable CAPTCHA*. CHI 2014, Toronto, Ontario, Canada.
- Bursztein, E. (2011). *Text-Based CAPTCHA Strengths and Weaknesses*. ACM Computer and Communication Security (CSS'2011). Retrieved from: <https://cdn.elie.net/publications/text-based-captcha-strengths-and-weaknesses.pdf>
- Chellapilla, C., Larson, K., Pimard, P., & Czerwinski, M. (2005a). Building Segmentation Based Human-friendly Human Interaction Proofs. *LNCS*, 3517.
- Chellapilla, K., Larson, K., Simard, P., & Czerwinski, M. 2005. Computers beat Humans at Single Character Recognition in Reading based Human Interaction Proofs (HIPs). *Second Conference on Email and Anti-Spam (CEAS)*.

- Chellapilla, K., & Simard, P. (2005b). *Using Machine Learning to Break Visual Human Interaction Proofs (HIPs)*. Retrieved from: <http://papers.nips.cc/paper/2571-using-machine-learning-to-break-visual-human-interaction-proofs-hips.pdf>
- Chew, M., & Baird, H. S. (2003). Baffletext: A Human Interactive Proof. *Proceedings of the Society for Photo-Instrumentation Engineers*, 305–316. doi:10.1117/12.479682
- Churchill, G. (1979). A paradigm for developing better measures of marketing constructs. *JMR, Journal of Marketing Research*, 16(February), 64–73. doi:10.2307/3150876
- Coats, A., Baird, H., & Fateman, R. (2001). Pessimistic Print: A Reverse Turing Test. *Proceedings of 6th International Conference on Document Analysis and Recognition*, 1154–1158.
- Davis, F. (1993). User Acceptance of Information Technology: Systems Characteristics, User Perceptions and Behavioral Impacts. *International Journal of Man-Machine Studies*, 38(3), 475–487.
- Egele, M., Bilge, L., Kirda, E., & Kruegel, C. (2010). CAPTCHA Smuggling: hijacking Web Browsing sessions to Create CAPTCHA Farms. *Proceedings of the 2010 ACM Symposium on Applied Computing*. Retrieved from: <https://iseclab.org/papers/manuel-captcha.pdf>
- Ellson, J., Douceur, J., Jowell, J., & Saul, J. (2007). Asirra: A CAPTCHA that Exploits Interest-Aligned Manual Image Categorization. CCS'07, Alexandria, VA.
- Fidas, C., Voyiatzis, A., & Avouris, N. (2011). On the Necessity of User-Friendly CAPTCHA. *CHI 2011*, 2623–2626.
- Flavián, N., Guinalíu, M., & Gurrea, R. (2006). The Role Played by Perceived Usability, Satisfaction and Consumer Trust on Website Loyalty. *Information and Management*, 43, 1–14.
- Gossweiler, R., Kamvar, M., & Basluja, S. (2009). *What's Up CAPTCHA? A Captcha Based on Image orientation*. The International World Wide Web Conference Committee (IW3C2), Madrid, Spain.
- Grafs, S., & Kinshuk. (2005). Improving Student Modeling: The Relationship between Learning Styles and Cognitive Traits. *Proceedings of the International Conference on Cognition and Exploratory Learning in Digital Age*, 37–44
- Grafs, S., & Kinshuk. (2009). Advanced Adaptivity in Learning Management Systems by Considering Learning Styles. *International Workshop on Social and Personal Computing for Web-Supported Learning Communities*, 235–238.
- Hatcher, L. (1994). *A step-by-step approach to using the SAS® System for factor analysis and structural equation modeling*. Cary, NC: SAS Institute Inc.
- Holsti, O. (1968). Content Analysis. In G. Lindzey & E. Aronson (Eds.), *The Handbook of Social Psychology* (2nd ed.; vol. 2, pp. 596–692). New Delhi: Amerind Publishing Co.
- Hwang, K., Haung, C., & You, G. (2013). A Spelling CAPTCHA Systems By Using Click. *International Symposium on Biometrics and Security Technologies*, 2(4), 1–8.
- Krippendorff, K. (2013). *Content Analysis: An Introduction to Its Methodology*. Sage.
- Kuzma, J. (2011). *CAPTCHA Accessibility Study of Online Forums*. Retrieved from: <https://eprints.worc.ac.uk/1414/2/captchaapril10.pdf>
- Kvalseth, T. (1989). Note on Cohen's kappa. *Psychological Reports*, 65(1), 223–226. doi:10.2466/pr0.1989.65.1.223
- Lewis, J., & Sauro, J. (2009). *The Factor Structure of the System Usability Scale*. Retrieved from: https://www.measuringu.com/papers/Lewis_Sauro_HCII2009.pdf

- Mollom. (2010). *Mollom CAPTCHA are intelligent*. Retrieved from: <https://mollom.com/blog/mollom-captchas-are-intelligent>
- Motoyam, M., Levchenko, K., Kanich, C., & McCoy, D. (2010). *Re: Captcha- Understanding CAPTCHA Solving services in an Economic Context*. Retrieved from: <http://cseweb.ucsd.edu/~savage/papers/UsenixSec10.pdf>
- Nielsen, J. (2003). *Usability 101: Introduction to Usability*. Retrieved from: <http://www.useit.com/alertbox/20030825.html>
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.
- Oxwall. (2013). *Smart Captcha Plug-in*. Retrieved from: <http://www.oxwall.org/store/item/583>
- Pix. (2015). *The CAPTCHA Project*. Retrieved from: <http://www.captcha.net/captchas/pix/>
- Pope, C., & Kaur, K. (2005). Is It Human or Computer? Defeating E-Commerce with Captchas. *IT Professional*, 7(2), 43–49. doi:10.1109/MITP.2005.37
- Preece, J., & Maloney-Krichmar, D. (2003). Online Communities. In J. Jacko & A. Sears (Eds.), *Handbook of Human-Computer Interaction*. Lawrence Erlbaum Associates Inc.
- ReCaptcha. (2015). *Google ReCaptcha*. Retrieved from: <https://www.google.com/recaptcha/intro/index.html>
- Rheingold, H. (1994). *The Virtual Community. Homesteading on the Electronic Frontier*. Reading, MA: Addison-Wesley Publishing.
- Serro, M., Salunke, S., & Mathur, A. (2013). Carcaking CAPTCHAs for Cash: A Review of CAPTCHA Crackers. *International Journal of Research and Technology*, 2(1).
- SolveMedia. (2013). *SolveMedia Launches in Europe*. Retrieved from: <http://news.solvemedia.com/post/55617465910/solve-media-europe>
- Sutherland, C. (2012). *Usability and Security of Text-Based CAPTCHAs UMM, CSci Senior Seminar Conference*. Morris.
- Swaid, S. (2013a). *Scale Development to Measure the Usability of Text-Based CAPTCHA*. International HCI 2013, Las Vegas, NV. doi:10.1007/978-3-642-39473-7_33
- Swaid, S. (2013b). *Usability Measures of Text-based CAPTCHA: Application of Think-Aloud Protocol*. Academic Business World International Conference, Nashville, TN.
- Swaid, S., & Wigand, R. (2009). Measuring the quality of e-service: Scale development and initial validation. *Journal of Electronic Commerce Research*, 10(1), 13–28.
- Vikram, S., Fan, Y., & Gu, G. (2011). *DEMAGE: A New Image-based Two-Factor CAPTCHA*. ACSAC'11, Orlando, FL.
- Von Ahn, L., Blum, M., & Langford, J. (2005). Telling Humans and Computers Apart Automatically. *Communications of the ACM*, 47(2), 57–60.
- Weber, R. (1990). *Basic content analysis*. London: Sage. doi:10.4135/9781412983488
- Yamamoto, T., Tyger, J., & Nishigaki, M. (2010). CAPTCHA Using Strangeness in machine Translation. *Proceedings of the 24th IEEE International Conference on Advanced Information Networking and Application*, 430–437.
- Yan, J., & El-Ahmad, A. (2008). Usability of CAPTCHA or Usability Issues in CAPTCHA Design. *Proceedings of the 4th symposium on Usable Privacy and Security*, 4–52. doi:10.1145/1408664.1408671
- Zhu, Q., & Keough, E. (2014). *Using CAPTCHAs to Index Cultural Artifacts*. Retrieved from: http://www.cs.ucr.edu/~eamonn/CAPTCHAs_Rock.pdf

ADDITIONAL READING

Germankos, P., Tsianos, N., Lekkas, Z., Mourlas, C., Belk, M., & Samaras, G. (2009). Towards an Adaptive and Personalized Web interaction using human factors. In M. Angelides, P. Mylonas, & M. Wallace (Eds.), *EDS. advances in semantic media adaptation and personalization* (pp. 247–282). Taylor and Francis. doi:10.1201/9781420076653-c12

Hernandez-Castro, C., & Ribogorda, A. (2010). Pitfalls in CAPTCHA Design and Implementaion: The Math CAPTCHA, A Case Study. *Computers & Security*, 29(1), 141–157. doi:10.1016/j.cose.2009.06.006

KEY TERMS AND DEFINITIONS

CAPTCHA: An acronym for “Completely Automated Public Turing test to tell Computers and Humans Apart”. CAPTCHA is a program that protects websites against bots and hackers by

generating and grading tests that humans can pass but current computer programs cannot.

CAPTCHA Usability: Ease of use CAPTCHA scheme and solve its test.

Equation Modeling: Structural equation modeling refers to a set of mathematical models, algorithms, and statistical methods that fit networks of constructs to data. SEM includes confirmatory factor analysis, path analysis, partial least squares path analysis, and latent growth modeling, among others.

Factor Analysis: Factor analysis is a statistical method for data reduction that describes variability among observed, correlated variables in terms of a potentially lower number of unobserved variables called factors.

Online Community: A virtual network of people who communicate with one another through interactive tools such as forums, discussion groups of social networking sites.

User Satisfaction: The attitude of a user to the CAPTCHA Scheme adopted by the website.

Visual Identity Design for Responsive Web



Sunghyun Ryoo Kang

Iowa State University, USA

Debra Satterfield

California State University – Long Beach, USA

INTRODUCTION

Because the web has moved beyond desktop environments, responsive web design requires reconsideration with regard to the interface design and development workflow (Carver, 2015). It is hard to create the same feeling and look across all devices (Moss, 2015), and maintaining a strong, consistent visual identity in a responsive web design is challenging because of this flexible or liquid layout and navigation. This flexible and dynamic design system is known as responsive web design (Marcotte, 2011). Responsive web design has changed the web eco system. Therefore, fundamental changes in web design are necessary to accommodate this variety in interfaces and platforms.

The first impression of a website is affected by the visual design elements of color, typography, and graphic elements (Park, 2016). If the visual design does not immediately connect with a user the likelihood that they will stay on the site is greatly diminished. Therefore, the website must capitalize on a familiar and trusted brand identity or the site must quickly establish a visual identity that is professional and immediately establishes a sense of trust or familiarity in the user (Kang and Satterfield, 2010).

The purpose of this chapter is to identify design strategies for establishing and managing the visual identity of a responsive web design. This chapter revisits the article, *Design Elements and Principles for Maintaining Visual Identity on*

Websites (Kang and Satterfield, 2010), to reflect the changed Web environments with flexible and dynamic web technologies.

BACKGROUND

Responsive Web Design

To create a successful responsive web design, a layout that can easily be customized for different screen sizes is key (Voutilainen, Salonen, and Mikkonen, 2015). According to Voutilainen, Salonen, and Mikkonen, it is important to consider “proportion-based grids and flexible images, where element sizing takes place using relative units instead of absolute ones, and CSS3 media queries, where different styles can be used for different devices (2015).” Since the web page layout is liquid and flexible, the exact structure of the layout would not be an element that can be used to maintain the web site identity. Using images and design elements that can be modified using CSS media queries and script language while still maintaining the essence of the visual characteristics of the site are essential to a successful responsive web design.

Visual Identity Design

According to Perry and Wisnom, the role of visual identity is to translate a company’s brand positioning and verbal identity into a visible vi-

sual representation (Perry and Wisnom III, 2003, p14). Perry and Winsnom outlined the purpose of visual identity as fourfold: “to bring the brand to life by giving character and personality to the positioning and name,” “to enhance brand recognition and recall,” “to differentiate the brand from competition”, and “to tie all the disparate brand elements together with the same look and feel” (Perry and Wisnom III, 2003, pp95-96).

The visual elements such as logos and colors in the identity system can create an immediate sense of recognition for the company, express its character, and build familiarity and trust (Haig and Haper, 1977, p14). According to Kang and Satterfield (2010), “an identity system can be thought of as a tool that is used to maintain a company’s corporate visual image. This system must be flexible enough for different applications, at the same time, tight enough to maintain visual consistency across media.” The most recognizable elements of corporate visual images are the brand logo, corporate colors and the visual identity (or corporate philosophy) as represented through photos, images and other familiar graphic symbols associated with that brand.

A logo or corporate symbol is the predominant visual cue for a business and it conveys the company’s philosophy and message. The roles of corporate symbols are to create awareness; to trigger recognition of an organization; and to activate a stored image of the organization (Dowling, 2001, p 167). Logos are used as a powerful tool for marketing. Customers often make buying decisions based on the brand recognition rather than on the specific characteristics of the product itself. Apple, Google, Coca-Cola, and McDonald’s logos are examples of these types of familiar and trusted corporate symbols.

Color is another important element in the identity system. Examples of color associations are Coca-Cola red, Kodak yellow, and Barbie pink. Thus, color is an important element in any identity system. According to Dowling (2001), color is easier to read than form or shape and it holds the viewer’s attention longer. Color also has

cultural and psychological associations. Therefore, the same color can be interpreted in different ways depending on the culture and context of the situation.

THE VISUAL ELEMENTS OF RESPONSIVE WEBSITE IDENTITY

Maintaining website identity for a responsive web design can be challenging and requires different design strategies from those used for print or traditional desktop web design. The use of logos, icons, symbols, colors, images, and typography for responsive web design need to maintain the visual identity of a site across a variety of sizes and screen configurations.

Use of Web Site Identifier

A site identifier such as the company name, logo, or sometimes both, and a web address (URL) is a unique feature identifying a website and the role is more important in the responsive web design. These site identifiers mainly logos are generally located in the top left corner with fixed location no matter in the flexible or dynamic layouts. Making the brand easy to see and identify is critical in insuring that a user will quickly identify the site and associate it with other sites on other devices and platforms.

According to Steve Douglas (2015), 93% of logos from the top 100 brands are simple enough to recognize at small size. For example, Apple and Nike’s logo is very simple but yet unique. It can be recognized with very small size even without the company name. In 2005, Google changed its typeface from serif to sanserif. One of reason for this change is for “scalable mark that could convey the feeling of the full logotype in constrained spaces (<https://design.google.com/articles/evolving-the-google-identity/>).”

Therefore, a logo as a site identifier should have a unique appearance to identify the characteristics of a particular web site, easier to recognize and recall, and scalable for different devices.

Use of Color

Color becomes the more powerful tool to build web site identity in the responsive web design. Color can be displayed slightly different in each device but it is much easier keeping the same color on the screen compared to print. According to the top 100 brand in 2016 from Forbes (<http://www.forbes.com/powerful-brands/list/>), blue and red are still the most popular color for cooperate logos. However, Nike uses a black logo and recently their web site (http://www.nike.com/us/en_us/) used this black and a monochromatic color scheme to integrate their responsive web site designs. The black and monochromatic color scheme enhanced the memorability and made Nike stand out among the numerous more colorful web sites. This emphasizes that the fact that the use of color as “a ubiquitous perceptual experience” (Elliot and Maier, 2007, p250) can be used in unique or distinctive ways to attract a viewer’s attention.

Use of Image

Graphics and images can be used as visual cues to forecast the characteristics of a website. According to Donald Norman (2004), photography has a strong appeal to human emotions (p50) and photographic images can be used to strengthen the visual appearance of an interface. For example, photographs showing families give a feeling of comfort, security, and trustworthiness and are perceived to be more friendly than photos that feature single men, single women, or couples (Ha and Kang, 2010). Another study (Park, 2016) showed that web users found websites that included photographic images were perceived to be more attractive than websites without photographic images. Thus, photos, graphics, illustrations, or icons can enhance the visual appearance of an interface and give users a more positive impression of the site. In the responsive web design, images must be scalable and set with a percentage, not as a fixed width. This allows images fit into the grid and allows the images to resize without sacrificing any of their aspect ratios.

Use of Layout

The layout of an interface is a visible form of its information architecture. Most web pages are divided into four main areas: site identifier; navigation; content; and footer (Kang & Lee, 2003) and these four areas have not changed in responsive web environments.

According to Kang and Satterfield (2010), “the layout is designed based on the quantity of content information, numbers of organizational categories, and marketing and design strategies.” However, in responsive web design, the layout is flexible and dynamic based on the screen size. Therefore, before a designer begins coding the web site, a layout style should be carefully considered to create an appropriate and consistent visual appearance across all screen configurations and platforms. Figure 1 shows how the layout and design elements translate between the horizontal desktop web screen and the smaller vertical smartphone screen.

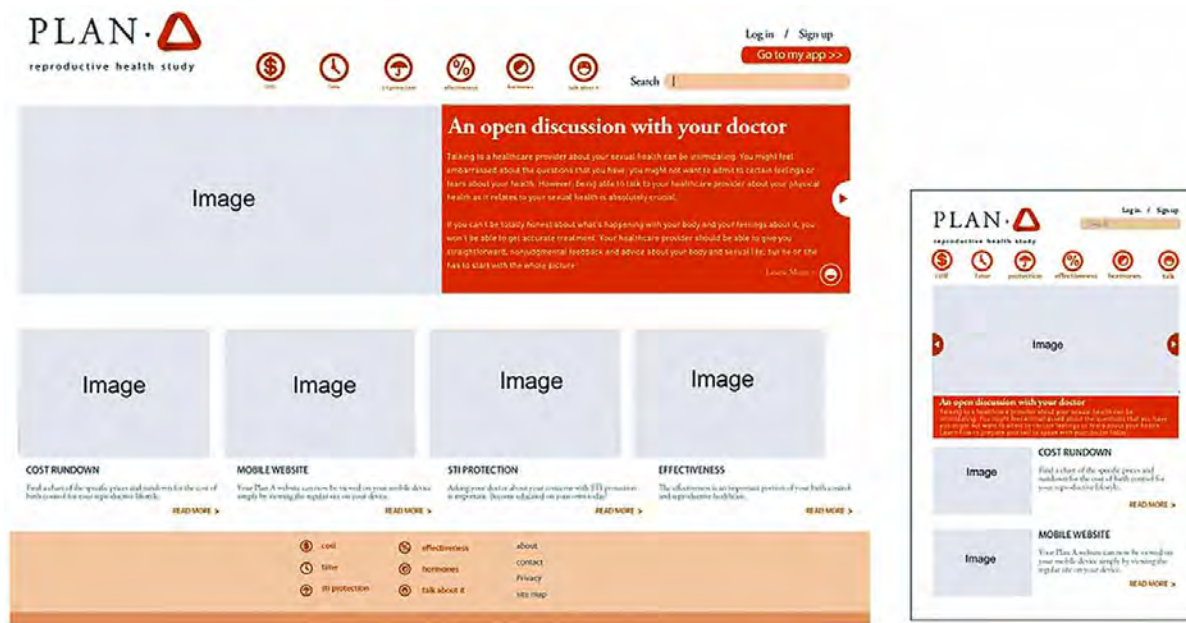
Use of Design System

The four main areas of a web; site identifier; navigation; content; and footer need to be organized in a system to maintain web site identity across various screen formats and sizes. The navigation bar plays an important role in maintaining visual consistency across multiple devices. Accordingly, design style is also an important element in responsive web design. A consistent visual style gives users a unified feeling between formats. A unified look and feel across the various platforms can also easily create a strong web site identity.

The design system should be flexible enough to be applied dynamically to a variety of devices yet strict enough in its use of a design system to maintain visual consistency. It is also important to consider the limitations and variability of web sites. For example, CSS 3 and media queries make it possible to control the web page layout.

A grid is a good way to improve usability and give a consistent look and feel to the page (Watzman 2003, pp268-285). However, in re-

Figure 1. Desktop Web Screen and Smart Phone Screen



sponsive web design the number of columns in a grid differs between devices and design elements will dynamically move around the web page. In order to compensate for these differences in layout, the use of color, typeface, and design style will be important in maintaining a consistent visual identity. A very tight systematic approach and design strategies are required to maintain an effective visual identity across all platforms in responsive web design.

Strategies for an Effective Responsive Web Site Identity Design

The basic strategies required to maintain a strong visual identity in web design remain the same. As Kang and Satterfield (2010) note, “the most important consideration in the creation of web site identity is an understanding of a company’s goals for having the web site and the users’ goals for visiting the web site. The design strategy for a web site comes from an understanding of the business and how to provide appropriate information in this media. It is also determined by the

target audience and an understanding of the web’s limitations and strengths.”

An effective identity design will use visual identifiers such as logos, icons, symbols, colors, images, and typography to maintain the brand identity of a responsive site across a variety of sizes and screen configurations. It will also capitalize on a user’s familiarity and trust in the brand’s identity. In order for the site to gain this advantage it must quickly align itself with its brand through the use of these visual identifiers and through the cohesive use of the four main areas of organization of a responsive web; the site identifier; navigation bar; site content; and the footer. For responsive web design the following are needed for each of these design and organizational elements.

Interaction and User Experience

The best web site identity combines effective usability and user experience. As Norman mentioned (2004), “attractive things make people feel good (p19)” and create positive emotions. Thus, a combination of good usability and providing

positive experience is essential with the appropriate web site identity elements. “Appropriate visual elements such as logos, images, colors, and typography will create a positive user experiences and build good brand impressions. Understanding the balance between usability and user experience will help designers create successful web site identity systems.” Therefore, usability and user experience are vital to building a successful responsive web site. “In a broad sense, usability can be considered a part of user experience. User experience depends on visual elements, which in turn affect the final appearance of web site design (Kang and Satterfield, 2010). Designing appropriate interactions for a responsive web site is essential to successful usability. Each device requires different interactions. For example, typing on the keyboard of a laptop or desktop computer is a very different experience from typing on a touch screen. According to Kang and Satterfield (2010), “to build a successful web site identity, a designer must understand not only web site usability but also users’ emotions and experiences. Users will experience a web site through navigational interfaces. An interface is a communication tool between the web site and the users. The interaction that takes place as facilitated by the interface is a personalized experience for each individual. Symbols, icons, colors, text, and images are the tools that create this user experience. When a web site has unique icons for navigation, people will remember the icons and identify the web site through them.”

Visual Identifiers and Branding

- **Logos or Branding:** Brand symbols and logos must be used consistently and must scale to fit the screen in the site identifier area of the screen.
- **Icons and Symbols:** The use of screen icons should remain consistent and the icons should be easily visible and readable in a variety of sizes. The use of complicat-

ed symbols or low contrast colors should be avoided to maintain a consistent look and an effective readability in all sizes.

- **Color:** Color should be selected for best readability in content areas and to create brand recognition. Colors should be used to create emphasis or visual hierarchy in similar ways in a variety of screen formats. Text, icons and symbols should use colors that are highly visible and readable in small sizes on small devices and not overpowering in larger screens.
- **Typography:** Type should be in colors that maintain readability in a variety of sizes. The typefaces used should have a variety of weights and styles to support readability and visual hierarchy in various sizes. Color should be used to maintain and reinforce the visual hierarchy of the type such as using darker colors for body copy and lighter colors for headings and subheads. Brand colors may be used to create brand identity or for emphasis.

Web Site Organization Elements

- **Site Identifier:** A site identification area must be clearly established through the use consistent use of brand symbols and logos. It should have the top level of visual hierarchy on the screen and the users should be able to quickly associate this web site with its brand affiliation.
- **Navigation Bar:** The navigation bar should quickly and easily orient the user to the site’s structure or information architecture. The location and visual style should be easy to find and readable across all devices even though the location and screen orientation may vary. The navigation system should be clear and consistent in its use of verbal language, symbols and categories of information. The number of categories should be visually compatible with

the amount of space available on a variety of devices and consistent with how the user interprets the information in each category.

- **Site Content:** Site content should be designed and written to be easily accessible at a variety of sizes and scales. Writing should be concise and use headings to help users orient themselves in the content and have a clear expectation of what they are reading or seeing. Content should be developed to be accessed with the fewest barriers such as excessive click throughs or long scrolling pages.

Visual hierarchy should be reinforced by selecting a clean, bold font for headings and an easily readable font for content areas and descriptions. The visual hierarchy can be further reinforced by adding color to major headings and using moderate to high contrast between the body copy and the background. Symbols and icons can be added to link text areas to the navigation categories or to create a visual contrast that will draw attention to key information in the body copy and help user's better understand the content or more quickly find an element of text.

Photography and illustration can be used in the content area to establish a sense of connection with the user by enhancing the relatability of the content to their lifestyle or to clarify the verbal content of a site such as including a construction diagram along side a list of instructions. A robust and sensitive combination of visual and verbal information can greatly enhance the perceived value and visual appeal of a responsive web site and should be used to create a strong connection with a company's corporate branding and philosophy.

- **Footer:** The footer should be consistently located at the bottom of the screen or device. It should include information that will allow the user to solve problems or access the company in appropriate ways. The footer should be clearly identified and separate from the content area. Color and

icons or symbols may be used to distinguish the area visually and to identify key associations with other organizations.

CONCLUSION

Responsive web design involves teamwork in various areas and it is critical that each individual who involved in the process must understand the roles of usability and user experience in order to build a successful responsive web site identity. For the responsive web, a lean UX (Gothelf and Seiden, 2013) method is recommended for the development process. Testing design ideas rapidly with end users during the design process to adjust the design and usability problems will help to achieve the same look and feel of the visual information in a responsive web environment.

Responsive web usability should consider the different ways of navigating the web site across multiple devices, while still maintaining clear information hierarchy, readability and legibility, and consistency in design styles. A consistent user experience is created through the effective use of visual elements such as the brand logo, colors, images, and a unique design style. In order to create successful responsive web site identity, defining the company's goals and users' goals for the web site with a team and testing the every step with end users are strongly recommended. "Just like corporate identity systems in traditional media, responsive web site identity must be considered critical to an effective branding effort on the web. At the same time, to design effective web site identity systems, a designer must take into consideration the web's unique dynamic and flexible characteristics (Kang and Satterfield, 2010)" for multiple screen sizes and platforms.

FUTURE RESEARCH DIRECTIONS

In this chapter, visual identity for responsive web design is examined through a variety of sources in a literature review. There are many usability

and UX studies for responsive web. However, no study has been conducted to determine the role of web site identity in the effectiveness of responsive web design. Future studies to identify best practices for usability and the effective use of visual elements in the responsive web design would be beneficial to both visual web designer and developers.

REFERENCES

- Carver, M. (2015). *The responsive web*. Shelter Island, NY: Manning Publications.
- Douglas, S. (2011). *Design secrets of the top 100 brands*. Retrieved June 15, 2016 from <http://www.thelogofactory.com/design-secrets-top-100-brands/>
- Dowling, G. (2001). *Creating Corporate Reputation: identity, image, and performance*. Oxford, UK: Oxford University Press.
- Elliot, A. J., & Maier, M. A. (2007). Color and psychological functioning. *Current Directions in Psychological Science*, 16(5), 250–254. doi:10.1111/j.1467-8721.2007.00514.x PMID:17324089
- Gothelf, J. and Seiden, J. (2013). *Lean ux applying lean principles to improve user experience*. Beijing: O'Reilly Media, Inc.
- Haig, W., & Harper, L. (1997). *The Power of Logos: how to create effective company logos*. New York: Van Nostrand Reinhold.
- Kang, S. R., & Lee, E. (2003). Investigating Elements on the E-commerce Homepage: Focus on Business to Customer Web sites. *Proceeding CD in The 6th Asian Design Conference*.
- Kang, S. R., & Satterfield, D. (2010). Design Elements and Principles for Maintaining Visual Identity on Websites. In I. Lee (Ed.), *Encyclopedia of E-Business Development and Management in the Global Economy* (pp. 955–1001). IGI Global. doi:10.4018/978-1-61520-611-7.ch100
- Marcotte, E. (2011). *Responsive Web Design*. New York: A Book Apart.
- Moss, B. (2015). *Is it time to embrace responsive branding?* Retrieved June 12, 2016 from <http://www.webdesignerdepot.com/2015/03/is-it-time-to-embrace-responsive-branding/>
- Norman, D. (2004). *Emotional Design: Why we love (or hate) everyday things*. New York: Basic books.
- Park, J. H. (2016). *A study of healthcare website's visual information in service design* (Unpublished MFA thesis). Iowa State University, Ames, IA.
- Perry, A., & Wisnom, D. III. (2003). *Before the Brand: Creating the unique DNA of an ensuring brand identity*. New York: McGraw-Hill.
- Voutilainen, J.-P. Salonen, J., & Mikkonen, T. (2015). On the design of a responsive user interface for a multi-device web service. *Proceedings of the Second ACM International Conference on Mobile Software Engineering and Systems*, 60–63.
- Watzman, S. (2003). Visual Design Principles for Usable Interfaces. In J. A. Janco & A. Sears (Eds.), *The Human-Computer Interaction Handbook*. Lawrence Erlbaum Associates.

KEY TERMS AND DEFINITIONS

Brand Identity: The visual elements and company's identity can distinguish the brand in consumer's mind and create brand identity.

Corporate Identity: Taglines, logos, and marketing elements which could build company's identity and brand image.

Experience Design: The way that the target audience experiences the brand from the first knowledge of the company through the complete consumer process.

Grid: Organizational structure used to create information templates.

Logo: A graphical element composed of a symbol and/or organization name.

Monochromatic: Color schemes uses only one color (hue) with various shades and tints.

Signatures: Combination of a logo mark and a company name.

Visual Identity: Visual elements such as a corporate logo, system elements, color system, typography to build company's identity.

Web Site Mobilization Techniques



John Christopher Sandvig

Western Washington University, USA

INTRODUCTION

The introduction of the first Apple iPhone in January 2007 started a historic shift in web browsing from desktop and laptop computers to smartphones and other mobile digital devices. By 2014 two of the largest global shopping sites, Amazon.com and Target.com, reported that the majority of their traffic originated from mobile devices (Sterling, 2015). During the first quarter of 2015 34% of global e-commerce transactions were conducted via mobile devices (“State of Mobile Commerce,” 2015). The number of smartphone users worldwide is forecast to grow from 2.08 billion in 2016 to 2.5 billion by 2019. Over 36% of the world’s population is projected to use a smartphone by 2018 (Statista, 2016). As the number of smartphone uses grows it is becoming increasingly important that web sites provide a good user experience on mobile devices.

BACKGROUND

Mobile devices have several physical limitations that require the use of special web design techniques to produce a mobile-friendly user interface. These physical limitations include small screens, virtual keyboards, slow download speeds, and high network latency. To accommodate these physical limitations mobile-friendly web sites often display less information than desktop sites, have larger tap targets such as buttons and form fields, use smaller images, and offer fewer navigation options. Sites that are developed primarily for desktop and laptop computers often require mobile users to zoom in and scroll horizontally, providing a poor user experience. Mobile-friendly web sites

are designed to render well on all devices, not just mobile devices.

MOBILIZATION TECHNIQUES

Currently there are three popular techniques for developing mobile-friendly web sites: responsive, dynamic serving, and redirect to separate URL. These techniques are referred to by a variety of names so to avoid confusion this article will adopt the nomenclature utilized in Google’s Mobile Developer Guide (“Mobile SEO Overview,” 2016). All three techniques detect the size of the user’s screen and modify the page content and layout to accommodate the device. The techniques differ in how they detect screen size and the mechanisms by which they modify page content.

Responsive Web Design

The responsive design technique delivers the same content (HTML, CSS, JavaScript, images, etc.) to all devices and “responds” to screen size by modifying the layout. For instance, a page design that utilizes a three column layout on a desktop computer may be displayed as one long vertical column on a mobile device. Tabbed navigation on a desktop computer may be displayed as a hamburger menu on a mobile device, and large images may be reduced in size for mobile devices.

Responsive is a relatively new technique. The phrase “responsive web design” was originally defined by Ethan Marcotte in 2007 (Els, 2015). Marcotte states that web pages should be designed with the flexibility to respond and adapt to the capabilities of various devices. At the time Marcotte proposed this approach CSS did not have the

DOI: 10.4018/978-1-5225-2255-3.ch704

ability to detect screen size but the approach could be implemented using JavaScript. This capability was included in the final recommendations for CSS3 and adopted in 2012 by the World Wide Web Consortium (Hargreaves, 2015). It then took a few years for web browser software to implement new standards and for web site developers to employ the new capabilities in web sites.

CSS media queries have the ability to detect screen size, thus allowing developers to create different layouts for different size screens. For example the following media query located in a site's css file would allow a developer to define styles for devices with screen widths between 320px and 400px.

```
@media screen and (min-device-width:
320px) and (max-width: 400px) {
    /* styles for small screens
go here */
}
```

The responsive technique is often implemented using any one of several open-source grid frameworks. These frameworks provide a basic structure for page layout, eliminating the need for developers to create designs from scratch. There are over two dozen grid frameworks include Bootstrap, Foundation 3, Skeleton, 1140 CSS Grid, and Less Framework 4 (Jain, 2015).

CSS frameworks allow developers to design different layouts for different size screens. For instance, Bootstrap 4.0 defines breakpoints at five different screen widths: extra small (less than 34em), small (34-47em), medium (48-61em), large (62-74em) and extra large (75em and larger) (Oliver 2015). Web designers can then create specific layouts for each of these five screen size ranges. Other styles classes, such as Bootstrap's. hidden-*-up and.hidden-*-down (where * is xs, sm, md, lg, and xl) allow developers to target specific content to different size screens. Many grid frameworks also provide JavaScript files that

work in conjunction with built-in CSS classes to provide dynamic features such as dropdown navigation, breadcrumbs, and hamburger menus (Lambert, 2016).

The responsive mobilization technique can also be implemented using JavaScript. JavaScript is a programming language that is supported by all modern browsers. It has the ability to detect screen sizes and apply formatting styles. For example the following JavaScript would hide an HTML element for screens widths between 320px and 400px.

```
if(screen.width >= 320 && screen.
width <= 400) {
    document.getElementById("nav").
style.display = "none";
}
```

While JavaScript is capable of implementing the responsive technique CSS media queries are designed specifically for page formatting and are easier to implement (Knight, 2011).

Dynamic Serving

The dynamic serving technique sends different page content and layout to different devices. The web server examines information which is contained in the web request and sends content and layout appropriate to the capabilities of the device. The dynamic serving technique has the advantage that it can send assets which are optimized for the device. For instance, if the server detects a mobile device it may send smaller and fewer images, thus reducing data usage and page load time.

To identify the requesting device the dynamic serving technique examines metadata that is included in the headers of each web request. Metadata includes information about the request, such as character encoding, language, cookies, caching, and the user-agent. The user-agent attribute includes information about the user's

browser and operating system. For example, the user-agent data sent from an iPhone 6 contains the following information:

```
user-agent:Mozilla/5.0 (iPhone; CPU
iPhone OS 9_1 like Mac OS X) App-
leWebKit/601.1.46 (KHTML, like
Gecko) Version/9.0 Mobile/13B143 Sa-
fari/601.1
```

A server can examine this data to determine that the requesting device is an iPhone using operating system 9_1. While the user-agent does not specifically include information about the screen size the server can lookup the typical screen sizes for the device and send appropriate content and formatting.

The weakness of the dynamic serving technique is in correctly identifying the user's device ("Mobile SEO Overview," 2016). The technique depends upon the server having a current list of user-agent strings to match against. Given the large number of mobile devices and the frequent introduction of new devices the user-agent lookup tables must be frequently updated. In practice many website administrators do not update the lookup table on a regular basis and servers do not correctly identify devices.

It is also easy for servers to misidentify the requesting device by matching incorrectly. For example, the server may match the word "Android" and determine that the request came from an Android phone when the request actually originated from a Chromebook laptop which also utilizes the Chrome operating system.

Redirect to Separate URL

The third popular mobilization technique redirects mobile devices to a separate URL. The separate URL then delivers content and layout appropriate to mobile devices. This technique is similar to the dynamic serving technique in that it can send images, layout, and other content that is

optimized for mobile devices. Like the dynamic serving technique, this technique depends upon correctly identifying the requesting device from the user-agent string.

Selecting a Mobilization Technique

The responsive technique has the advantage that it is relatively easy to implement. A number of free open-source CSS grid frameworks are available that provide developers with convenient tools for designing pages that render well on all devices. A weakness of the responsive technique is that exactly the same content is delivered to all devices, using excessive amounts of bandwidth and slowing page load times. Mobile layouts often display less content and smaller images than desktop computers and can benefit from using less data and having faster load times, an opportunity that is lost with the responsive technique. A second weakness of the response technique is that the CSS and JavaScript files used by the grid frameworks can be fairly large, using more data and increasing page render time. Despite these disadvantages the responsive technique is Google's recommended design pattern ("Mobile SEO Overview" 2016).

The dynamic server and separate URL techniques both have the advantage of serving content that is optimized to the capabilities of the device. Smaller images and less content can be sent to mobile devices, thus speeding download times and using less data. A recent empirical study of large global e-commerce sites found that sites using the separate URL technique used 67.7% as much data for their mobile sites as for their desktop sites, while dynamic serving used 80.5% as much data (Sandvig, 2016). Surprisingly the reduction in data usage did not translate into faster page render times.

The disadvantage of these techniques is that they both rely upon examining the user-agent string and correctly identifying the device. If this is done incorrectly the techniques may send content and layout that is not appropriate for the device.

A second disadvantage of the redirect to a separate URL technique is that redirecting the browser to a different site increases page load time. The amount of time varies from a fraction of a second to several seconds depending upon network latency, network congestion, and other factors (Doswell, Kendall, Aslam, & Sexton (2015).

The previously mentioned empirical study by Sandvig (2016) explored the utilization of the three techniques on large global e-commerce sites. The study examined the landing pages of Alexa's 527 top global shopping sites. It found that 89.3% of top global shopping sites utilized mobile-friendly designs. The most popular mobile technique was to redirect to a separate URL, utilized by 44.1% of sites, followed by dynamic serving, utilized by 23.8%, and responsive, utilized by 23.0% of sites.

Identifying Mobilization Techniques

A web site's behavior can be used to identify which mobilization technique, if any, is used. Modern desktop browsers, such as Chrome, Firefox, and Internet Explorer have built-in mobile emulators that can be used to compare how a site displays on desktop versus a mobile device. The following sequence of tests can be used to identify a site's mobilization technique or lack of technique. These instructions are specific to the Chrome browser but the steps are similar with other major browsers. Steps to identify a site's mobilization technique:

1. On a desktop computer use the Chrome browser to open the web site being examined.
2. Press the F12 key to open the developer console. In the top left of the developer console is an icon with a small box superimposed over a larger box. This icon toggles the view between desktop and mobile views. Initially it will be in desktop view.
3. **Separate URL Technique:** Click the mobile emulation icon to toggle to mobile view. Also reload the page by clicking the page reload icon left of the site's URL (or press the F5 key). If the address changes to a dif-

ferent URL then the site uses the separate URL mobilization technique. For example, at the time of this writing www.facebook.com changes to m.facebook.com.

4. **No Mobilization Technique:** If the mobile site displays poorly with small text or is too wide for the mobile screen the site is not mobile friendly. At the time of this writing the site discoverelks.org is not mobile friendly.
5. **Responsive Technique:** The layout of responsive sites changes without reloading the page. Click the mobile emulation icon a few times and observe if the page layout changes back and forth between a mobile and desktop layout. If it does then the site utilizes the responsive technique. At the time of this writing site www.wvu.edu uses the responsive technique.
6. **Server Adaptive Technique:** This technique sends different content to the browser depending upon the type of device that made the request. To identify the server adaptive technique we need to see if the content sent to the browser changes between desktop and mobile views. In the Chrome developer console click the network tab (top of window), check the "disable cache" box, and reload the page. At the bottom the developer console the number of requests and the kilobytes (KB) of data downloaded is displayed. Compare these numbers for the desktop and mobile views (remember to reload the page after toggling the view). If they are significantly different the technique is server adaptive. At the time of writing Amazon.com's desktop site made 309 request and transferred 6.1 MB (megabytes) of data and its mobile site made 61 request and transferred 925 KB of data. Because these numbers are significantly different we can conclude that Amazon uses the server adaptive technique. The numbers may changes slightly with each load so it may be necessary to repeat the above steps a few times.

Emerging Mobilization Technique

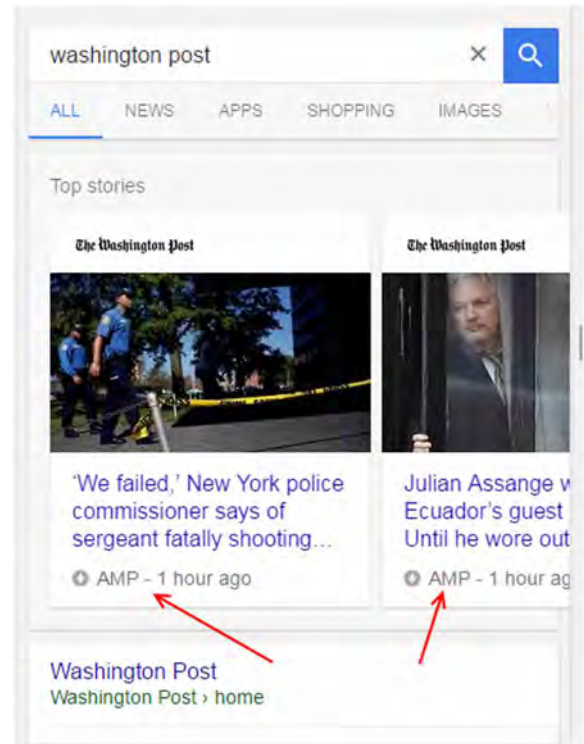
The limitations of current mobilization techniques have motivated the development of a new approach called Accelerated Mobile Pages (AMP). The open source project was launched February 2016 and is sponsored by Google and Twitter. The goal of AMP is “dramatically improve the performance of the mobile web” (Mizroch, 2015). Usability studies of mobile user have found that the average mobile user will wait only three seconds for a mobile page to load before abandoning the page (Vijayan 2016). AMP is a new web design standard that prioritizes page render speed over other features. It disallows many design elements that slow the rendering of web pages on mobile devices, such as JavaScript code and third-party scripts. It also requires the use of certain design features that allow pages to load more quickly, such as height and width attributes on images. The only JavaScript allowed in AMP pages is the AMP JS library itself. This library optimizes resource loading, provides custom AMP HTML tags, disables slow CSS selectors, and provides other speed optimization features. To further increase speed Google will cache valid AMP pages and serve them directly from its data centers. According to Google AMP pages load on mobile devices in approximately one quarter the time and use about 10% as much data as regular web pages (Vijayan 2016).

To encourage rapid adoption AMP supports have enlisted several major web sites, media outlets, and content management systems to adopt AMP standards. A variety of media outlets including the *New York Times*, BuzzFeed, Twitter, *The Guardian*, Ebay and Vox Media have pledged support for the project (Vanian, 2016).

Web pages that use AMP are identified as such in Google mobile search results as shown in figure 1.

A downside of the AMP standard is that it disallows many rich features that are common on desktop sites. Many publishers that utilize AMP for their mobile sites will likely elect to maintain

Figure 1.



a separate site for desktop users (Vijayan 2016). Maintaining multiple web sites increases expenses and has the other disadvantages discussed earlier for the separate URL mobilization technique.

FUTURE RESEARCH DIRECTIONS

The rapid evolution of mobilization techniques presents researches with a host of research opportunities. While there has been a plethora of research of measuring mobile usability (Coursaris & Kim, 2011; Venkatesh, Ramesh, & Massey, 2003; Kjeldskov & Stage, 2004; Zhang & Adipt, 2005) there is little research on implementing mobilization techniques. Practitioners who are faced with the decision of selecting a mobilization strategy need to guidance to answer the following questions:

- What are the relative implementation and maintenance costs for each mobilization technique?
- Does the technique provide a good user experience for both desktop and mobile users?
- What are the performance tradeoffs between the techniques in terms of speed, reliability, user experience, and costs?
- Which techniques are more popular within different industries and why?
- Which mobilization techniques are growing and which are declining?
- How does an organization's existing infrastructure and human resources impact its choice of mobilization strategy?
- Are new and improved mobilization techniques on the horizon that are going to make existing techniques obsolete?

Several research approaches could be used by researchers to address these issues. Questions relating to implementation and maintenance costs could be answered by surveying webmasters of different organizations in different industries. Webmaster surveys could also be used to evaluate satisfaction with end user experience and reliability. The questions related to technique popularity and performance could be answered by evaluating actual web performance, similar to the survey of large global e-commerce sites by Sandvig (2016). The questions related to infrastructure, human resources, and future directions could be addressed using focus groups of industry experts.

CONCLUSION

The rapid adoption of mobile devices over the past decade has created a need web for mobile-friendly web sites. This article has described the three most popular web mobilization techniques: response, server adaptive, and redirect to a separate URL. All three techniques modify web page content

and layout to accommodate all types of devices, including mobile.

A completely new approach to mobilization, Accelerated Mobile Pages, was introduced in early 2016. While current mobilization techniques utilize standard HTML and CSS to adapt web pages to various screen sizes, AMP utilizes a new technical standard. The new standard consists of a subset of HTML and CSS and utilizes JavaScript to optimize speed performance. While it is too early to know how widely AMP will be adopted it is being widely promoted and several major web publishers are experimenting with it.

REFERENCES

- Coursaris, C. K., & Kim, D. J. (2011). A meta-analytical review of empirical mobile usability studies. *Journal of Usability Studies*, 6(3), 117–171.
- Doswell, S., Kendall, D., Aslam, N., & Sexton, G. (2015, August). A longitudinal approach to measuring the impact of mobility on low-latency anonymity networks. In *2015 International Wireless Communications and Mobile Computing Conference (IWCMC)* (pp. 108–113). IEEE. doi:10.1109/IWCMC.2015.7289066
- Els, D. (2015). *Responsive Design High Performance*. Packt Publishing Ltd.
- Hargreaves, B. (2015). *HTML5 & CSS3-The Complete Guide to Modern Day Web Development*. Academic Press.
- Jain, N. (2015). Review of different responsive CSS Front-End Frameworks. *Journal of Global Research in Computer Science*, 5(11), 5–10.
- Kjeldskov, J., & Stage, J. (2004). New techniques for usability evaluation of mobile systems. *International Journal of Human-Computer Studies*, 60(5), 599–620. doi:10.1016/j.ijhcs.2003.11.001
- Knight, K. (2011). Responsive web design: What it is and how to use it. *Smashing Magazine*, 12.

Lambert, M. (2016). *Bootstrap Site Blueprints* (Vol. 2). Packt Publishing Ltd.

Mizroch, A. (2015, Oct. 7). Google's 'Accelerated Mobile Pages' Promises Faster Loads. *Wall Street Journal*.

Mobile S. E. O. Overview. (2016, May 16). Retrieved July 12 from <https://developers.google.com/webmasters/mobile-sites/mobile-seo/>

Sandvig, J. (2016). Mobilization Techniques Utilized by Leading Global E-commerce Sites. *International Journal of E-Business Research*, 12(2), 1-16.

State of Mobile Commerce. (2015). Retrieved Dec. 12, 2015 from <http://www.criteo.com/media/1894/criteo-state-of-mobile-commerce-q1-2015-ppt.pdf>

Statista. (2016). *Number of smartphone users worldwide from 2014 to 2019*. Retrieved 7/15/2016 from <http://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/>

Sterling, G. (2015). *The Mobile Majority: 60 Percent Of Holiday Traffic Came From Mobile, Says Target*. Retrieved 12/24/2015 from <http://marketingland.com/mobile-majority-60-percent-holiday-traffic-came-mobile-says-target-113873>

Vanian, J. (2016, Jan 21). *Google's Plan for Faster Mobile Web May Reshuffle Search Rankings*. Fortune.com.

Venkatesh, V., Ramesh, V., & Massey, A. P. (2003). Understanding usability in mobile commerce. *Communications of the ACM*, 46(12), 53–56. doi:10.1145/953460.953488

Vijayan, J. (2016, Feb. 24). Accelerated Mobile Pages Go Live in Google Mobile Search. *eWeek*, p. 1.

Zhang, D., & Adipat, B. (2005). Challenges, methodologies, and issues in the usability testing of mobile applications. *International Journal of Human-Computer Interaction*, 18(3), 293–308. doi:10.1207/s15327590ijhc1803_3

ADDITIONAL READING

Bootstrap Get Started. (2016) Available at http://www.w3schools.com/bootstrap/bootstrap_get_started.asp

Bootstrap is the most popular HTML, CSS, and JS framework for developing responsive, mobile first projects on the web (2016) Available at <http://getbootstrap.com/>

Getting Started with AMP HTML. (2016) Available at <https://www.ampproject.org/>

Mobile, S. E. O. Overview (2016, May 16) Available at <https://developers.google.com/webmasters/mobile-sites/mobile-seo/>

Mobile Friendly Websites. (2016, April 6) Available at <https://developers.google.com/webmasters/mobile-sites/>

KEY TERMS AND DEFINITIONS

CSS: Cascading style sheets is a language used to describe the presentation of a web page. CSS is used to define style attributes such as fonts, colors, spacing, positioning, and backgrounds. CSS works in conjunction with HTML to define and format page content.

HTML: Hypertext markup language is a standard used for describing web page content. HTML elements are delineated by dozens of different tags which are enclosed by angle brackets. For

example, a paragraph of text may be surrounded by paragraph `<p>` `</p>` tags, and an image included using an `<image>` tag. HTML works in conjunction with CSS to define and format page content.

JavaScript: A programming language that is supported by all modern web browsers. It can be used to add dynamic features such as image slide shows, accordion menus, rollover effects, and asynchronous server requests.

Media Queries: A feature of CSS that is used to define different style rules for different media types. Media types include screen, print, and speech. When developing pages that are mobile-friendly media queries are often used to detect screen size and apply styles appropriate to the device.

What Accounts for the Differences in Internet Diffusion Rates Around the World?

W**Ravi Nath***Creighton University, USA***Vasudeva Murthy***Creighton University, USA*

INTRODUCTION

The literature is replete with studies that indicate a strong link between the adoption and use of the Internet and economic growth not only for individual citizens but also for national economies. Despite the irrefutable benefits of the Internet, in many countries around the world, the Internet adoption rates remain low. In Africa, according to the 2015 data, 10.7 percent of the households enjoy Internet access at home and about one individual out of five (20.7%) uses the Internet (International Telecommunications Union (ITU), 2016). In contrast, in Europe, these figures are 82.1% and 77.6%, respectively. Figure 1 clearly highlights the gross disparity that exists between the Internet diffusion rates for the Organization for Economic Cooperation and Development (OECD) and Sub Saharan Africa (SSA) countries. Thus, the question of why do these differences exist is worth investigating. And, an understanding of the factors that promote and/or hamper the adoption of the Internet is warranted as the insights gleaned from this can assist policy makers, economic developmental agencies, and the political leaderships in formulating appropriate strategies and policies.

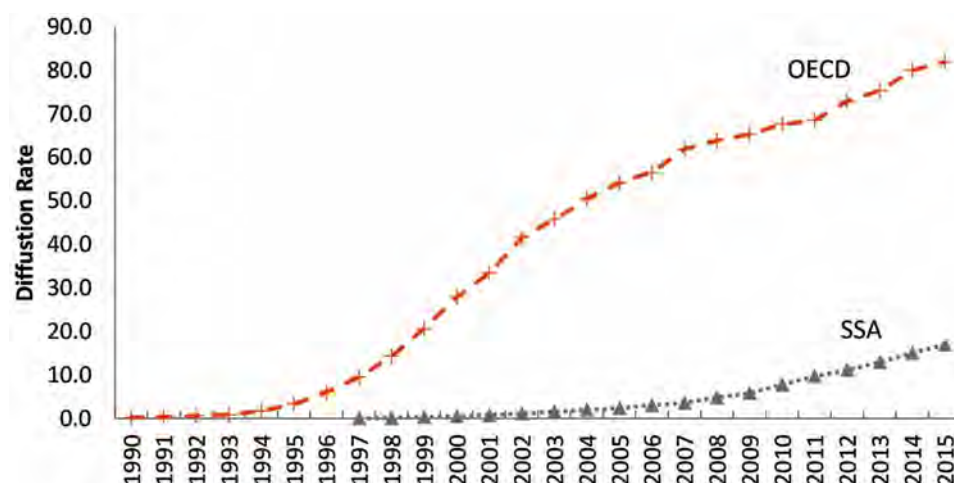
BACKGROUND

The Internet and the associated technologies are vital in boosting the economic wellbeing of nations and their citizens. Applications such

as e-business, voice over IP (VoIP), mobile commerce, and integrated supply chains have become the primary drivers of the growth of economic activities in many countries (Albirini, 2008; Dedrick, Gurbaxani, & Kraemer, 2003; Kenny, 2003; Koh & Chong, 2002). Further, the Internet is providing revenue-generating and skill-enhancing opportunities to individuals across the globe. For individuals, it avails the opportunities to sell things online, collaborate with others at far-flung places, learn new skills, access data/information quickly, and communicate rapidly (Chavula, 2013; James, 2008; Larson & Murray, 2008; Laguerre, 2013). Adoption and use of the Internet and other technologies have boosted trade and globalization leading to improved GDP (Albirini, 2008; Kuppusamy, M. & Santhapparaj, A.S., 2005; Lawrence, 2002; Raheel, Karim, Saleem & Bharwani, 2012). In fact, the 2015 Global Information Technology Report asserts that digital technologies can boost global economic output and creates new economic opportunities for individuals particularly in countries with very low technology penetration rates. The report also claims that a 10% increase in a country's telecommunications infrastructure boosts the GDP by nearly 2.8 percentage points provided the country first achieved a certain minimum threshold of digitization. The threshold is estimated to be about 24 of the population. The 2013 Global Information Technology Report had claimed that a doubling of mobile data usage increases GDP per capita by ½ a percent point. Accelerated economic growth of

DOI: 10.4018/978-1-5225-2255-3.ch705

Figure 1. Internet Diffusion rates for OECD and Sub Saharan Africa (SSA) – 1990-2015



India and China in the last decade is also a prime example of how Information and Communication Technologies (ICT) in concert with appropriate economic, intellectual property protection, and infrastructure improvement policies promote rapid economic development (Also see, Global Information Technology Report, 2015).

GLOBAL INTERNET ADOPTION FACTORS

At the national level, there exist many factors that influence the Internet adoption rate. These factors include the availability of reasonably-priced technology/telecommunications infrastructure, access to personal computers, educational and training opportunities for individuals, income levels, innovative capability, culture, competition in the information technology industry, and the availability of local Internet content (Beilock & Dimitrova, 2003; Birba & Diagne, 2012; Chinn & Fairlie, 2007; Choi, Williams & Ha, 2014; Dewan, Ganley & Kraemer, 2010; Dholakia, Dholakia & Kshetri, 2003; Feng, 2015; Galagedarage & Salman, 2015; Huang & Chen, 2010; Kiiski & Pohjola, 2002; McCoy, Cha & Durcikova, 2012; Meijers, 2006; Murthy, 2004; Murthy, Nath & Soleimani, 2015; Nath & Murthy, 2003, 2004;

Ojuloge & Awoleye, 2012; Oyelaran-Oyeyinka & Lal, 2005; Soleimani, 2015; Touray, Saminen & Mursu, 2015; Wunnava & Leiter, 2009). Further, the rule of law (e.g., property rights, strong legal system) governing the country's trading system, government regulations and market liberalization policies, and credible payment systems (e.g., credit cards, digital wallet and cash) are necessary for expanding into mobile and digital commerce.

Human Capital Development Factors

It is virtually impossible to fully reap the benefits of the Internet if people are not able to read or write or have a basic understanding of computers, Internet and their applications. Therefore, an individual has to possess a minimum level of computer/technology and language literacy to use the Internet for beneficial purposes. At the national/regional level, one way to assess this is to consider factors such as:

1. Adult literacy rate
2. Percent of school age children enrolled in schools
3. Per capita spending on education

Many studies including those by Balamoun-Lutz (2003), McCoy, Cha and Durcikova (2012),

Nath and Murthy (2003), Noce and McKeown (2008), and Wunnava and Leiter (2009) have concluded that literacy and tertiary enrollment rates have a strong positive influence on Internet adoption. Also, literate citizens are more accepting of information and communication technology (ICT) innovations. And this results in higher acceptability and utilization of the Internet and its applications in commerce, personal productivity tools, education, health, and social networking.

Technology Factors

As one would expect, a reliable and cost effective network infrastructure is a prerequisite for people to use the Internet. This includes the bandwidth, the number of Internet hosts, the reliability of electric power, and the percent of the population of a nation that have access the Internet. In addition, people who are familiar and comfortable with using other technologies such as a phone, a mobile phone and a personal computer are more likely to adopt and use the Internet. Specific variables that one may consider here are:

1. Number of personal computers per 100 inhabitants.
2. Telephone lines per 100 inhabitants.
3. Cell phone subscribers per 100 inhabitants.
4. Number bandwidth (bits) per capita.
5. Reliability of electrical power.

One key aspect of the new economy is the availability of cost-effective information and communication technologies (ICT) and the above listed items represent leading indicators of ICT (Balioune-Lutz, 2003). Several studies have established that countries with higher availability rates of personal computers, tablets, telephones, and mobile phones, sufficient bandwidth and reliable supply of electricity that powers most ICT devices, also have more Internet users (Nath & Murthy, 2003, 2004; Soleimani, 2015; Wunnava & Leiter, 2009). Further, the availability of online digital content, its quality and relevance also seem

to play a role on the adoption of the Internet (Viard and Economides, 2015).

Economic Factors

How expensive is it to get online? If this cost is prohibitively high relative to the income, then the Internet penetration rate is bound to be low. The two variables that are relevant here are:

1. Real Gross domestic product per capita (in US purchasing parity \$).
2. Average monthly cost of 20 hours of Internet access.

Kiiski and Pohjola (2002), and McCoy, Cha and Durcikova (2012), found that, across countries, GDP per capita and Internet access costs were good predictors of the growth of the Internet. Further, Nath and Murthy (2004) using data from 62 countries, demonstrated that higher “average monthly cost of 20 hours of Internet access” had a significant negative impact on the Internet adoption rates.

Political Factors

A nation’s government policies and regulations are also important in promoting Internet adoption among its population. Many nations restrict the use of the Internet (e.g., China, Iran and North Korea). Also, the economic policies of the government can affect the degree to which its citizens use the Internet. Variables that measure “political and economic” policies include:

1. **Economic Freedom Index (EFI):** Beach and O’Driscoll (2003) define this index as the “.... absence of government coercion or constraint on the production, distribution, or consumption of goods and services beyond the extent necessary for citizens to protect and maintain liberty itself.” This index aggregates several factors covering broad issues such as corruption, non-tariff

barriers to trade, the fiscal burden of government, the rule of law and efficiency of the judiciary, regulatory hurdles for businesses, labor market restriction, and black market activities. Complete details regarding the development and description of this index can be found in Beach and O'Driscoll (2003). The values of EFI can vary from 1 to 5. A value of 1 indicates set of national policies that promote economic freedom and a value of 5 signifies policies that are least conducive to economic freedom.

2. **Innovation Capability of the Country:** This variable is calculated as the product of the number of patents granted per million inhabitants and gross tertiary enrollment rate. Note that the number of patents reflects the nation's innovation intensity and the enrollment rates denote the degree of investment in human capital. Thus, this measure reflects a country's capability, ability to provide a conducive environment for augmenting technological advance and its capacity for innovation in technologies and products (McArthur & Sachs, 2000).

In a study of the adoption of the Internet across countries, Nath and Murthy (2003) have shown that both the economic freedom index and the innovation capability of a country play a positive role towards the adoption of the Internet. These findings have been further supported by Balamoun-Lutz (2003) and Nath and Murthy (2004). Also, Choi, Williams & Ha (2014) report that the State influence – policies that promote the Internet and education – have significant impact on the diffusion of the broadband Internet.

Cultural Factors

Cultural factors do impact the rate of diffusion of the Internet. Culture is defined as: "the collective programming of the mind which distinguishes

the members of one human group from another (Hofstede, 1991, p.5)." In his book titled *Culture's Consequences*, Hofstede (1980) suggested four dimensions of culture.

1. **Power Distance (PD):** Power distance is defined as "the extent to which the less powerful members of institutions and organizations within a country expect and accept that power is distributed unequally" (Hofstede, 1991, p.27). In cultures with high power distance, decisions are centralized and subordinates are often fearful of disagreeing with their superiors. On the other hand, cultures with low power distance are more participative and have less tolerance for the lack of autonomy. In a society with low power distance, we expect people to be more innovative and willing to try new things.
2. **Individualism vs. Collectivism (IND):** This dimension relates to the way people live together. Individualism "pertains to societies in which the ties between individuals are loose: everyone is expected to look after himself or herself and his or her immediate family" (Hofstede, 1991, p.51). On the other hand, collectivism "pertains to societies in which people from birth onwards are integrated into strong, cohesive groups, which throughout people's lifetime continue to protect them in exchange for unquestioning loyalty" (Hofstede, 1991). Cultures with high individualism value personal time and achievement. Such societies are expected to be more innovative and open to new ideas.
3. **Uncertainty Avoidance (UA):** Uncertainty avoidance refers to "the extent to which the members of a culture feel threatened by uncertain or unknown situations" (Hofstede, 1991, p. 113). A culture high in uncertainty avoidance is rule oriented, has less tolerance for opinions and behaviors different from its own, and avoids taking risks. There is also

resistance to change. Cultures with high uncertainty avoidance are expected to be less innovative and less accepting of new things.

4. **Masculinity vs. Femininity (MAS):** This is possibly the most controversial dimension of culture advocated by Hofstede. In highly masculine cultures “men are supposed to be assertive, tough and focused on material success; women are supposed to be more modest, tender and concerned with the quality of life” (Hofstede et al., 1998). The more modern and popular perspective on this dimension is to view the masculine and feminine culture in terms of competitiveness and material success versus nurturing behavior and quality of life, as opposed to gender roles for the sexes.

There is considerable evidence supporting that cultural factors influence the use of various technological innovations. Yaveroglu and Donthu (2002) indicate that the use of cell phones, home computers, and microwave ovens is high in countries with low power distance, low uncertainty avoidance, and high individualism. Yenyurt and Townsend (2003) investigated the role that culture plays in the acceptance of new products. Their findings indicate that lower acceptance of new products is related to power distance and uncertainty avoidance. With respect to the adoption of Enterprise Resource Planning (ERP) software, national culture is shown to have a significant influence on its adoption rate (van Everdingen & Waarts, 2003). Cultural considerations even play a role in how information systems, in general, are designed, implemented, and used (Gallupe & Tan, 1999; Nelson & Clark, 1994; Watson, Ho, & Raman, 1994; Straub, 1994; Montealegre, 1997; Jarvenpaa & Leidner, 1998; Thatcher, Srite, Stepina, Liu, 2003). With respect to the role of culture on Internet diffusion, Nath and Murthy (2004) have shown that “uncertainty avoidance” and “masculinity” are positively associated with the diffusion of the Internet.

Individual User Factors



An individual user's needs and cognition are important in the Internet adoption decision. Zhu and He (2002) assert that there are three user cognition factors that one needs to consider:

1. **Perceived Characteristics of the Internet:** This includes perceived ease of use, compatibility, and the perceived advantage of the Internet.
2. **Perceived Popularity of the Internet:** This refers to the percent of Internet users among the user's close network (family, friends, and acquaintances).
3. **Perceived Need for the Internet:** To what extent the user thinks he/she needs the Internet for things like the news, information, entertainment, etc.

Using surrogate variables such as literacy ratios, number of PCs, number of TV sets (discussed above in other factors) for McCoy, Cha, and Durcikova (2012) show that user cognitive factors are important in the diffusion of the Internet.

FUTURE RESEARCH DIRECTIONS

The future for the Internet adoption and its use is very promising. First, the broadband access to the Internet will increase rapidly allowing people very high-speed access to the Internet. Consequently, it will be possible for users to download rich content at a very high speed. Second, Internet access costs will decline due to global competition and economic reforms taking place in countries around the world. Third, as many nations become receptive to economic reforms and provide additional economic freedom opportunities to its citizens, entrepreneurship will thrive, incomes will rise, information economies will thrive and more people will use the Internet and become fully vested in the information economy. Fourth, smart phones will play an increasingly important

role in providing Internet access to citizens around the world in anytime, anywhere environment. Further, mobile phone access to the Internet will yield unprecedented economic benefits to citizens at the “bottom of the pyramid” around the world. For example, mobile money services have proliferated in Sub-Saharan Africa providing financial services to millions of previously unbanked citizens (The mobile economy: Sub-Saharan Africa, 2014). Finally, individual privacy issues will become more prominent as there will be fierce debates on striking a balance between protecting individual privacy and going fully-digital via the Internet, mobile phone, and other accompanying technologies. However, in spite of these limiting concerns, people and governments around the world will increasingly look to the Internet to better their lives.

CONCLUSION

A clear and vivid understanding of the factors that determine the adoption of the Internet is important for researchers, practitioners, and policy makers. Researchers who wish to study international information technology issues and their economic implications must be cognizant of these factors in determining Internet user behaviors and technology adoption patterns. Also, an understanding of inter-country differences with respect to these adoption factors can improve IT/IS research design strategies and enhance the conduct of global information technology research studies.

At the national level, policy makers, political leadership, and think tanks can benefit by incorporating the insights gleaned from this article into their deliberations and policy plans – leading to more informed macro-level strategic policies. Without a cohesive and carefully crafted national development strategies and policies, nations are likely to miss out on the tremendous benefits the Internet has to offer to its citizens and national development.

REFERENCES

- Albirini, A. (2008). The Internet in developing countries: A medium of economic, cultural and political domination. *International Journal of Education and Development using Information and Communication Technology*, 4(1), 49-65.
- Baliamoune-Lutz, M. (2003). An analysis of the determinants and effects of ICT diffusion in developing countries. *Information Technology for Development*, 10(3), 151–169. doi:10.1002/itdj.1590100303
- Beach, W. W., & O'Driscoll, G. P. (2003). *Explaining the factors of the index of economic freedom*. Retrieved from www.heritage.org
- Beilock, R., & Dimitrova, D. V. (2003). An exploratory model of inter-country Internet diffusion. *Telecommunications Policy*, 27(3-4), 237–252. doi:10.1016/S0308-5961(02)00100-3
- Birba, O., & Diagne, A. (2012). Determinants of adoption of Internet in Africa: Case of 17 sub-Saharan countries. *Structural Change and Economic Dynamics*, 23(4), 463–472. doi:10.1016/j.strueco.2012.06.003
- Chavula, H. K. (2013). Telecommunications development and economic growth in Africa. *Information Technology for Development*, 19(1), 5–23. doi:10.1080/02681102.2012.694794
- Chinn, M. D., & Fairlie, R. W. (2006). The determinants of the global digital divide: A cross-country analysis of computer and Internet penetration. *Oxford Economic Papers*, 59(1), 16–44. doi:10.1093/oep/gpl024
- Choi, S. B., Williams, C., & Ha, S. H. (2014). Institutions and broadband internet diffusion in emerging economies: Lessons from Korea and China. *Innovation: Management Policy & Practice*, 16(1), 2–18. doi:10.5172/impp.2014.16.1.02

- Dedrick, J., Gurbaxani, V., & Kraemer, K. L. (2003). Information technology and economic performance: A critical review of the empirical evidence. *ACM Computing Surveys*, 35(1), 1–28. doi:10.1145/641865.641866
- Dewan, S., Ganley, D., & Kraemer, K. L. (2010). Complementarities in the diffusion of personal computers and the Internet: Implications for the global digital divide. *Information Systems Research*, 21(4), 925–940. doi:10.1287/isre.1080.0219
- Dholakia, N., Dholakia, R. R., & Kshetri, N. (2003). *Internet Diffusion. The Internet Encyclopedia*. New York: Wiley.
- Feng, G. F. (2015). Factors affecting Internet diffusion in China: A multivariate time series analysis. *Telematics and Informatics*, 32(4), 681–693. doi:10.1016/j.tele.2015.02.009
- Galagedarage, T. M., & Salman, A. (2015). Issues in Internet adoption and rural development: Theory of the diffusion of innovations. *The Journal of Development Communication*, 26(1), 14–23.
- Gallupe, R. B., & Tan, F. (1999). A research manifesto for global information management. *Journal of Global Information Management*, 7(3), 5–18. doi:10.4018/jgim.1999070101
- The global information technology report. (2015). Retrieved from http://www3.weforum.org/docs/WEF_Global_IT_Report_2015.pdf
- Hofstede, G. H. (1980). *Culture's consequences: International differences in work-related values*. Beverly Hills, CA: Sage Publications.
- Hofstede, G. H. (1991). *Cultures and Organizations: Software of the Mind*. New York: McGraw-Hill.
- Hofstede, G. H., Arrindell, W. A., Best, D. L., De Mooij, M., Hoppe, M. H., & Wan de Vliert, E. et al. (1998). Masculinity and femininity: The taboo dimension of national cultures. In W. Lonner & J. Berry (Eds.), *Cross-Cultural Psychology Series* (Vol. 3). Newbury Park, CA: Sage Publications.
- Huang, C., & Chen, H. (2010). Global digital divide: A dynamic analysis based on the Bass model. *Journal of Public Policy & Marketing*, 29(2), 248–264. doi:10.1509/jppm.29.2.248
- International Telecommunication Union. (2016). Retrieved February 22, 2016 from www.itu.int/itu-d/ict/statistics/pages/stat/default.aspx
- James, J. (2008). Time-intensive information technology and human welfare in developing countries. *Prometheus*, 26(2), 165–177. doi:10.1080/08109020802029760
- Jarvenpaa, S., & Leidner, D. (1998). An information company in Mexico: Extending the resource-based view of the firm to a developing country context. *Information Systems Research*, 9(4), 342–361. doi:10.1287/isre.9.4.342
- Kenny, C. (2003). The Internet and economic growth in less-developed countries: A case of managing expectations. *Oxford Development Studies*, 31(1), 99–113. doi:10.1080/1360081032000047212
- Kiiski, S., & Pohjola, M. (2002). Cross-country diffusion of the Internet. *Information Economics and Policy*, 14(2), 297–310. doi:10.1016/S0167-6245(01)00071-3
- Koh, C. E., & Chong, H. (2002). Does the Internet improve business? An empirical inquiry into the perceived strategic value and contribution of the Internet. *Journal of International Technology & Information Management*, 11(1), 81–97.
- Kuppusamy, M., & Santhapparaj, A. S. (2005). Investment in information and communication technologies (ICT) and its payoff in Malaysia. *Perspectives on Global Development and Technology*, 4(2), 147–168. doi:10.1163/1569150054739014
- Laguerre, M. S. (2013). Information technology and development: The Internet and the mobile phone in Haiti. *Information Technology for Development*, 19(2), 100–111. doi:10.1080/02681102.2012.690170

- Larson, R. C., & Murray, M. E. (2008). Distance learning as a tool for poverty reduction and economic development: A focus on China and Mexico. *Journal of Science Education and Technology*, 17(2), 175–196. doi:10.1007/s10956-007-9059-1
- Lawrence, S. (2002, February). Technology and the global economy. *Red Herrings*.
- McCoy, S., Cha, H. S., & Durcikova, A. (2012). Modeling Internet diffusion in developing countries. *Australasian Journal of Information Systems*, 17(2), 5–23. doi:10.3127/ajis.v17i2.618
- Meijers, H. (2006). Diffusion of the Internet and low inflation in the information economy. *Information Economics and Policy*, 18(1), 1–23. doi:10.1016/j.infoecopol.2005.02.005
- Montealegre, R. (1997). The interplay of information technology and the social milieu. *Information Technology & People*, 10(2), 106–131. doi:10.1108/09593849710174977
- Murthy, N. R. (2004). Internet diffusion: An econometric analysis. *Asian-African Journal of Economics and Econometrics*, 4(1), 45–54.
- Murthy, N. R., Nath, R., & Soleimani, M. (2015). Long term factors of Internet diffusion in sub-Saharan Africa: A panel co-integration analysis. *Journal of International Technology & Information Management*, 24(4), 79–94.
- Nath, R., & Murthy, N. R. (2003). An examination of the relationship between digital divide and economic freedom: An international perspective. *Journal of International Technology & Information Management*, 12(1), 15–23.
- Nath, R., & Murthy, N. R. (2004). A study of the relationship between Internet diffusion and culture. *Journal of International Technology & Information Management*, 13(2), 123–132.
- Nelson, K., & Clark, T. Jr. (1994). Cross-cultural issues in information systems research: A research program. *Journal of Global Information Management*, 2(4), 19–29. doi:10.4018/jgim.1994100102
- Noce, A. A., & McKeown, L. (2008). A new benchmark for Internet use: A logistic modeling of factors influencing Internet use in Canada. *Government Information Quarterly*, 25(3), 462–476. doi:10.1016/j.giq.2007.04.006
- Ojuloge, B., & Awoleye, M. (2012). Analysis of Internet diffusion and adoption in selected African countries. *International Journal of Internet of Things*, 1(3), 5–11.
- Oyelaran-Oyeyinka, B., & Lal, K. (2005). Internet diffusion in sub-Saharan Africa: A cross-country analysis. *Telecommunications Policy*, 29(7), 507–527. doi:10.1016/j.telpol.2005.05.002
- Raheel, H., Karim, M., Saleem, S., & Bharwani, S. (2012). Information and communication technology use and economic development. *PLoS ONE*, 7(11), 1–7. PMID:23133658
- Soleimani, M. (2015). *A longitudinal study of co-diffusion between personal computers and global diffusion: An approach to bridging the digital divide* (Master's Thesis). Creighton University, Omaha, NE.
- Straub, D. (1994). The effect of culture on IT diffusion: E-mail and FAX in Japan and the U.S. *Information Systems Research*, 5(1), 23–47. doi:10.1287/isre.5.1.23
- Thatcher, J. B., Srite, M., Stepina, L. P., & Liu, Y. (2003). Culture overload and personal innovativeness with information technology: Extending the nomological net. *Journal of Computer Information Systems*, 44(1), 74–81.
- The mobile economy: Sub-Saharan Africa. (2014). Retrieved from http://www.gsmamobileeconomyafrica.com/GSMA_ME_SubSaharanAfrica_Web_Singles.pdf
- Touray, A., Salminen, A., & Mursu, A. (2015). Internet adoption at the user level: Empirical evidence from the Gambia. *Information Technology for Development*, 21(2), 281–296.

van Everdingen, Y. M., & Waarts, E. (2003). The effect of national culture on the adoption of innovations. *Marketing Letters*, 14(3), 217–232. doi:10.1023/A:1027452919403

Viard, V. B., & Economides, N. (2015). The effect of content on global internet adoption and the global digital divide.. *Management Science*, 61(3), 665–687. doi:10.1287/mnsc.2013.1875

Watson, R. T., Ho, T. H., & Raman, K. S. (1994). Culture: A fourth dimension of group support systems. *Communications of the ACM*, 37(10), 44–55. doi:10.1145/194313.194320

Wunnava, P. V., & Leiter, D. B. (2009). Determinants of intercountry Internet diffusion rates. *American Journal of Economics and Sociology*, 68(2), 413–426. doi:10.1111/j.1536-7150.2009.00634.x

Yaveroglu, I. S., & Donthu, N. (2002). Cultural influences on the diffusion of new products. *Journal of International Consumer Marketing*, 14(4), 49–63. doi:10.1300/J046v14n04_04

Yeniyurt, S., & Townsend, J. D. (2003). Does culture explain acceptance of new products in a country? *International Marketing Review*, 20(4), 377–396. doi:10.1108/02651330310485153

Zhu, J. J. H., & He, Z. (2002). Perceived characteristics, perceived needs, and perceived popularity: Adoption and use of the Internet in China. *Communication Research*, 29(4), 466–495. doi:10.1177/0093650202029004005

ADDITIONAL READING

Andrés, L., Cuberes, D., Diouf, M., & Serbrisky, T. (2010). The diffusion of the Internet: A cross-country analysis. *Telecommunications Policy*, 34(5–6), 323–340. doi:10.1016/j.tel-pol.2010.01.003

Atkin, D. J., Jeffres, L. W., & Neuendorf, K. A. (1998). Understanding Internet adoption as telecommunications behavior. *Journal of Broadcasting & Electronic Media*, 42(4), 475–490. doi:10.1080/08838159809364463

Billon, M., Marco, R., & Lopez, F. R. (2009). Disparities in ICT adoption: A multidimensional approach to study the cross-country digital divide. *Telecommunications Policy*, 33(10–11), 596–610. doi:10.1016/j.telpol.2009.08.006

Chang, H. H., & Just, D. R. (2009). Internet access and farm household income – empirical evidence using a semi-parametric assessment in Taiwan. *Journal of Agricultural Economics*, 60(2), 348–366. doi:10.1111/j.1477-9552.2008.00189.x

Crenshaw, E. M., & Robison, K. K. (2006). Globalization and the digital divide: The roles of structural conduciveness and global connection in internet diffusion. *Social Science Quarterly*, 87(1), 190–207. doi:10.1111/j.0038-4941.2006.00376.x

Dewan, S. (2005). Across the digital divide: A cross-country multi-technology analysis of the determinants of IT penetration. *Journal of the Association for Information Systems*, 6(12), 409–432.

Erumban, A. A., & de Jong, S. B. (2006). Cross-country differences in ICT adoption: A consequence of culture? *Journal of World Business*, 41(4), 302–314. doi:10.1016/j.jwb.2006.08.005

Fuchs, C., & Horak, E. (2008). Africa and the digital divide. *Telematics and Informatics*, 25(2), 99–116. doi:10.1016/j.tele.2006.06.004

Goldfarb, A., & Prince, J. (2008). Internet adoption and usage patterns are different: Implications for the digital divide. *Information Economics and Policy*, 20(1), 2–15. doi:10.1016/j.infoecopol.2007.05.001

Guillén, M. F., & Suárez, S. L. (2005). Explaining the global digital divide: Economic, political and sociological drivers of cross-national Internet use. *Social Forces*, 84(2), 681–708. doi:10.1353/sof.2006.0015

Hanafizadeh, M. R., Saghaei, A., & Hanafizadeh, P. (2009). An index for cross-country analysis of ICT infrastructure and access. *Telecommunications Policy*, 33(7), 385–405. doi:10.1016/j.telpol.2009.03.008

Kim, S. (2011). The diffusion of the Internet: Trend and causes. *Social Science Research*, 40(2), 602–613. doi:10.1016/j.ssresearch.2010.07.005

Li, R., & Shiu, A. (2012). Internet diffusion in China: A dynamic panel data analysis. *Telecommunications Policy*, 36(10–11), 872–887. doi:10.1016/j.telpol.2012.06.004

Orviska, M., & Hudson, J. (2009). Dividing or uniting Europe? Internet usage in the EU. *Information Economics and Policy*, 21(4), 279–290. doi:10.1016/j.infoecopol.2009.06.002

Oyeleran-Oyeyinka, B. O., & Adeya, C. N. (2004). Internet access in Africa: Empirical evidence from Kenya and Nigeria. *Telematics and Informatics*, 21(1), 67–81. doi:10.1016/S0736-5853(03)00023-6

Pohjola, M. (2003). The adoption and diffusion of ICT across countries: Patterns and determinants. In D. C. Jones (Ed.), *New Economy Handbook* (pp. 77–100). San Diego, CA: Academic Press.

Zhao, H., Kim, S., Suh, T., & Du, J. J. (2007). Social institutional explanation of global Internet diffusion: A cross country analysis. *Journal of Global Information Management*, 15(2), 28–55. doi:10.4018/jgim.2007040102

KEY TERMS AND DEFINITIONS

Economic Freedom Index: A measure of the absence of government constraints on the economic activities of a nation.

Hofstede's Cultural Dimensions: Four factors proposed by Hofstede along which cultures might differ.

Internet Adoption Rate: Percent of population of a country having access to the Internet.

Power Distance (PD): A measure of the unequal distribution of power in a society. Cultures with high power distance are centralized whereas cultures with low power distance are more participative and individuals like more autonomy.

Uncertainty Avoidance (UA): A measure of how well a culture accepts uncertainty. A culture with high uncertainty avoidance is mostly rule oriented and is less tolerant of divergent opinions and behaviors.

About the Contributors

For full author biographies, please visit the book's webpage at:

<http://www.igi-global.com/book/encyclopedia-information-science-technology-fourth/173015>

Index

'Taggable' Nature of Media 7109, 7115
 1G Technology 6274
 21 Inc. 2865-2866, 2868
 21st Century Learning 1421, 1426, 1512, 2240, 3240, 7180
 21st Century Skills 1421-1422, 1427, 2236, 2240, 2253, 2293, 2299, 2492, 2643, 2646, 3248-3249, 3252-3253, 3255, 3259, 3310, 6390, 6859
 2G technology 6274
 360 Degree Feedback 4339, 4346
 3D Modeling 542, 982, 987, 1278, 7856, 7867-7868
 3D Printing 305, 505, 521, 523-524, 528, 545, 683, 2626-2630, 2632-2634, 2636-2638, 2640, 4382-4387, 5108, 5110, 5291, 5866
 3D stereoscopic display 2417
 3D systems 4382
 3G technology 6274
 4G Technology 6274

A

AACSB 3174-3177, 3182
 AADL 6848, 6858
 AASS 3052, 3055
 ABC Analysis 3699-3700, 3702, 3707, 3709
 AC and EC Motors 5649
 Academia 100, 230-231, 402, 575, 761, 805, 1131, 1362, 2567, 2580-2582, 2584-2585, 2587, 3177, 3341, 3473, 3479, 3963, 4004, 4077, 4327, 4680, 5079, 5245, 5251, 5313, 5540, 5546, 5865, 6882-6883, 7044-7047, 7056, 7150, 7155, 7266, 7383, 7408, 7763, 7777
 Academic Analytics 5159-5160, 5168
 Academic Business Ethicists 3177, 3182
 Academic Community 8, 1479-1480, 1486, 1883, 2385, 4416, 4724, 5082, 5726, 5730-5731, 5733, 5870, 7205, 7801, 7915-7916, 7984
 academic development center 1880, 1882, 1892, 1894
 Academic learning cycle 3853

academic librarians 5196, 5198, 5203, 5210, 5252, 5254, 5257, 5268
 Academic Library 5194-5197, 5203, 5205-5206, 5209-5210, 5726-5727, 5730-5731, 5733, 5735
 academic library director 5203
 Academic Performance 1540-1541, 1543-1545, 1547, 1728, 2605, 3296, 3881, 4168, 5122, 5161, 5170, 5177, 6404, 7088-7092, 7095, 7153, 7682, 7689
 Academic Social Networks 4416, 7044-7049, 7051-7052, 7056
 ACBSP 3174-3176, 3182
 Accelerated Mobile Pages 8091-8092
 Acceptance Criteria 6827, 7587
 Acceptance of Political Regimes 3540, 3548
 Access Control 373, 376, 651-653, 655-656, 658, 660, 890, 1030, 1357, 1670, 1677, 2156, 2160-2163, 2491, 4927, 4985, 6561, 6618, 7475
 Access issue 7250
 Accordance with Rules and Social Contract 3548
 accountability culture 1880, 1883, 1892, 1894
 Accreditation 760, 1107, 3173-3175, 3182, 3881, 3884-3885, 3889
 achievement motivation 1540-1545, 1547
 Ackoff DIKW hierarchy 946
 acquiring bank 2709, 2722
 Action Potential 6594
 Active Learning 1426, 1437, 1453, 1530, 1883, 2305, 2413, 2418, 2559-2561, 2563, 2567-2568, 2629, 2641, 3847, 4360, 5161, 5801, 5803, 5808, 5812, 5822, 5853-5854, 6362, 7880, 7884, 7897, 7942
 Active Mediation 7098-7099, 7102-7103, 7105
 Active Users of the ICTs 2315, 2325
 Activism 1642, 2292, 3027, 3372, 3570-3574, 3577, 3668, 6278, 6280, 6282
 Activists 1510, 2064, 3569-3574, 3669, 4163, 6276-6277, 6280-6282, 6700, 7194, 7296
 Activity Theory 2943-2944, 2947, 2950, 2952, 3290, 4482, 6352-6353
 Activity-Based Costing (ABC) 2176

- Activity-Based Management (ABM) 2176
 Actor Network Theory (ANT) 2949, 7190, 7199, 7312-7313, 7321, 7325
 Actors of Open Data 3041
 Ad Hoc Network 6106, 6108-6109, 6111, 6113, 6530, 6535-6536, 6540, 6543, 6545
 ADALINE 121, 6580
 Adaptable Hypermedia System 6433
 Adaptation Methods 4555, 6426, 6428
 adaptation problems 2616-2617
 Adaptation Techniques 4554-4555, 6427-6428
 Adaptive Buffering 4554, 4558, 6436
 Adaptive Education (AE) 2381, 2391
 Adaptive Educational Hypermedia (AEH) 2380-2381, 2391, 2458-2459
 Adaptive Educational Hypermedia Systems (AEHS) 2380-2381, 2391
 Adaptive Hypermedia (AH) 2357-2359, 2361-2363, 2368, 2380, 2382-2383, 6424-6425, 6429-6430, 6477, 6479
 Adaptive Hypermedia Systems 2380, 2382-2383, 6424-6425, 6479
 Adaptive Management Information System 363
 adaptive network 4554
 Adaptive Network-On-Chip 4558
 Adaptive Routers 4550-4552, 4558
 Adaptive Routing 4549-4555, 4558
 adaptive skills 2439
 Adaptive Switching 4553-4554, 4559
 Adaptive Topology 4555, 4559
 Adaptivity 180-181, 1519, 1521, 1527, 5540, 8013
 add-compare-select 6309-6312, 6318
 Add-Compare-Select Unit (ACSU) 6309, 6311, 6318
 Adder 4022, 4608-4611, 4614, 6017, 6244, 6312-6313
 ADDIE Model 1562, 3289, 6344, 6346
 Adding Object 196
 Additive 222, 298, 305-308, 313, 506, 520-521, 545, 1298, 1738, 2086, 2189, 2640, 3874, 4382-4383, 4385, 4388, 4884-4885, 4887, 4896, 4977, 5108, 6674
 Additive Manufacturing (AM) 298, 305-308, 313, 520-521, 2640, 4382-4383, 4385, 4388
 Adhesion 431-432, 434, 436-437, 439
 Admission Control 1166, 6266-6272, 6610, 6612-6613, 6615-6616
 Adoption And Use 2516, 2724, 2728, 2731, 2734-2735, 4089, 6162, 6166, 7764, 7800, 8095
 Adsorption 439, 2680, 3156, 3161, 3165, 3169-3170
 Adult Education 1432, 1440, 1464, 1520, 1527, 2228, 2231, 2357, 2421, 2523, 5821, 6188, 7680
 Adult Learners 1432-1433, 1435, 1439, 1453, 1463-1470, 2264, 2269, 2422, 2503-2509, 3832, 3947, 5095, 5833, 7636
 Adult Learning Theory 5832, 5837, 5840
 Adult motivation 2508
 Advance Fee Scam 6221
 Advanced Imaging Technology 6710
 Advanced Persistent Threats (APT) 4848, 5986, 5990
 Advanced Recommendation 1742, 1744
 Advanced Technology 458, 2628, 2640, 3398, 4320, 7636, 7645, 7811
 Adversary 1243, 1647, 2828, 4910, 4927, 4930-4931, 4976, 4981-4982, 4984
 Adverse Selection 4855, 6917, 6926
 Advertising Model 7129, 7139
 Advocacy Coalition 6713, 6725
 Aesthetic Classification 158, 165
 Aesthetic Scoring 158, 165
 Aesthetics 156, 158-161, 165, 573, 732, 2468-2472, 2476, 2883, 3241-3242, 3244, 3247, 3331, 3333, 4459, 4465, 4693, 5040, 5106-5108, 5146, 5767-5772, 6025, 6378, 6387, 7604
 Affection and Play 6377-6378
 Affective Computing 132, 143, 161, 4124-4130, 4134
 Affective Human-Computer Interaction 4124, 4126, 4134
 Affective Promotion 2663-2665, 2667-2668, 2671
 Affective State 4125-4130, 4134, 6379, 7979
 Affinity Distance 743, 750, 5989
 Affordability 425, 2508, 2726, 2769, 4628, 5257, 6416, 7222, 7228, 7746, 7750
 Affordable Care 2098, 2508, 3806, 3815
 AFO 298-304, 306-308, 313
 AFOLU 3127, 3130
 Agile Methodology 7508, 7578, 7587
 Agile Processes 7513
 Agile Software Development Methodology 1406, 7401, 7410
 Agility 574, 626, 639, 688-689, 873, 878, 880, 883, 1106, 1146, 1152, 1155, 1157, 1398-1400, 1405, 2079, 2774, 3840, 3883, 5285, 5306, 5540-5541, 5559, 5610, 5612, 6727, 7424, 7578
 Agnostic 4329, 6184, 6194, 6503
 agogos 1467, 1470, 1472
 Agree to Disagree 73, 76
 AHP 587, 2006, 2012, 3132-3133, 3137, 3143, 5693, 7283
 AIA 3156, 3160-3161, 3165, 3170, 5479
 AICTE 4724, 4728, 4734
 Air Freight 7869

Index

- Air Pollution 3076, 3373-3374, 3380, 3639, 5465, 5737, 5743, 6981
Air Wars 3612, 3617, 3620
Akerlof 62, 2833
Alan November 2579
Alan Turing 98, 106, 229, 238, 255, 4744, 7731, 7743
Algebraic Expressions 6848-6850, 6855
Algebraic Specification 7612, 7621
Algorithmic Discrimination 4951, 4955, 4957
Algorithmic Intelligence 220-221, 225, 228
algorithmic thinking 2393, 2401
Aliasing 4746-4747, 4758-4759, 6043, 6054, 6171, 6179, 6183, 6901
Aliasing Rejection 4747, 6054
Alive Mutant 7469
allied health 5810-5812, 5816, 5819, 5822-5823, 5917
ALM 132, 886, 2792, 2834-2838, 2845
altcoins 2876, 2878, 2880
Alternative Dispute Resolution (ADR) 2825, 2832
Alternative Set Theory 2211
Altmetrics 7049, 7051, 7056
Altruism 89, 5095, 5632, 6020, 7164, 7167, 7285
Alvin Toffler 1640
Ambient Intelligence (Ami) 1994, 4951-4952, 4961, 7745
Ambiguity 39, 186, 553, 577, 1147, 1487, 1535, 1788, 1838-1839, 2124, 2230, 2266, 2394, 2398, 2922, 3226, 3353, 3405, 3625, 3842-3843, 3845, 4474, 4490, 4492, 4494, 4498, 4519, 4529, 4883, 5127, 5129, 5135, 5663, 5682, 5694, 6058, 6086, 6698, 6705, 6784, 7072, 7283, 7313-7314, 7317, 7401-7402, 7555, 7564
American Library Association 2227, 2245-2246, 2305, 2480, 2483, 3871, 6685-6686, 6689, 6691, 6750, 7294-7296
American Recovery and Reinvestment Act (ARRA) 3679, 3806, 3814
amicable settlement 2825-2826, 2832
Analysis Model 590, 1991, 3289, 4821, 6790
Analytical Hierarchy Process 734, 2006, 2019, 3132, 5693
analytics process 1883-1884, 1889, 1892, 1894
Anderson And Krathwohl 1475, 5150-5152, 7929
Andragogy 1464-1468, 1472-1473, 1520, 5840, 6338, 7673, 7898
Android 266, 268, 2396, 2418, 3275-3276, 3615, 4238, 5280, 5887, 6035, 6037, 6042, 6057, 6060-6069, 6083, 6154, 6201, 6207, 6210-6213, 6245, 6302, 6392, 7369-7377, 7380, 7395-7396, 7548, 7638, 7640, 7760, 8037, 8089
Android Studio 6057, 6061-6062, 6069, 7395-7396
Animation 133, 808, 973-982, 2335, 2368, 2398, 2411, 2443, 2446, 2454, 2599-2608, 2610-2615, 3287, 4191, 4215, 4257, 5104, 6074, 6341, 6414, 6422, 7856, 7858, 7867-7868, 8069
anime 2326, 2329, 2335
ANN 4-5, 8, 112-114, 120-127, 131, 148-150, 232-233, 408, 493, 496-498, 500-501, 2085, 2089, 2608, 3107, 3355, 5172-5173, 5178, 5182, 6351, 6579, 6581, 6585-6586, 6588-6590, 6631-6635, 6640, 6642-6651, 6657, 6660-6666, 6669-6670, 7352
Annotation 235, 422-424, 443, 988, 1209, 1522, 1788, 2058, 2065, 2067, 2071, 2074, 2547, 2644, 4528, 5859, 6286, 6291, 6428-6430, 6476, 6478, 6480, 6483, 7628
Anonymity 663, 1133, 1608, 1695, 1728, 2123-2124, 2296, 2534-2535, 2541, 2787-2788, 2863, 2870, 2876, 2878, 3767-3769, 3771, 3843, 4158-4159, 4161-4162, 4167, 4169-4173, 4901, 6561, 6950, 6956, 6960, 7072, 7079, 7127, 7164, 7220, 7318, 7320, 7799, 7849, 7973-7974
Anonymous 255, 565, 1220, 1359, 1678, 1695, 1697, 1724, 1728, 2122, 2534, 2536, 2538, 2540-2542, 2544, 2548, 2716, 2788, 2868, 2870, 2875-2876, 2878, 3114, 3117-3119, 3768, 3771, 3972, 4164, 4167, 4169-4171, 4198, 4922, 5391, 6000, 6020, 6024, 6276, 6391, 6953, 6955, 7211, 7296, 7674, 7797, 7815-7816, 7819-7820, 7828-7829, 8018-8019
Anonymous Social Media 4167
antenarrative 5660, 5665, 5668
Antenna 2019, 5491, 5495-5497, 5520, 5889, 5967, 5970-5972, 5976, 5980, 6095, 6105, 6256-6260, 6263-6266, 6319-6320, 6325-6326, 6328, 6330, 6332, 6334, 6499-6501, 6505-6506, 6595, 6597-6598, 6602-6603, 6712
anthropological spectrum 1467, 1470, 1472
Anthropology 382, 425, 3117, 3574, 4189, 4404, 4493, 4676, 4702, 4762, 5770, 5889
anthropometry 2626-2627, 2630, 2640
Anthropomorphic 202, 211, 1012, 4128
anti-aliasing filter 6014
Anti-Forensics 2918
Anti-Monotonic Constraint 343, 348
Anti-monotonicity 343
Anxiety Disorders 3262, 3271
Anxiously Attached Person 7199
AofABE 635
APA Practice Organization 3807
Aphasia 2439, 2441

- app developers 6057, 6090, 6165, 6170, 6381
 Appearance Descriptor 6895, 6898, 6904
 Apple 101, 266, 276, 1104, 1659-1660, 2749, 3518, 3615, 3747, 3828, 4542, 4631, 5280, 5750-5751, 5756, 5981-5987, 6057, 6060, 6069, 6119, 6141, 6152, 6154, 6166-6167, 6210-6211, 6224, 6302, 6392, 6449-6450, 6696, 7118, 7124, 7141, 7273, 7369, 7371, 7375-7377, 7381, 7748, 7758, 7760, 7801, 8037, 8080, 8087
 Application Domain 421, 657, 1350, 1377, 1874, 1908, 1921, 1985, 1990, 2776, 5127-5132, 5135, 6034, 6930, 7411-7413, 7419, 7421-7422
 Application Programming Interface (API) 464, 4017, 7018
 Application-to-Application (A2A) 2783
 Applied Cybermetrics 3479, 3482
 Applied Research 123, 613, 4560, 6769-6770, 6777-6778
 Appraisal Systems 4337, 5789-5790, 5799
 Appraise 3300, 4339, 4346, 4792
 Appraiser 4338, 4346
 Appropriation process 3795-3799, 3801
 AQC 7718-7720, 7730
 Arabization 7263-7264, 7268
 Arbor 6582-6583, 6594
 Archaeology 1070, 3419, 3433, 3500, 7005
 Archetypes 3638, 3645, 8038
 Architectural Document 5224
 Architectural Principles 573, 578-580, 583
 Architectural Representation 975, 980, 5213-5214, 5218
 Architecture Driven Modernization (ADM) 7457
 Archives 1412, 1786, 2265, 2293, 3035, 3119, 3550, 3691, 4191, 4307, 4311, 4437-4439, 4442, 5106, 5207, 5213-5218, 5223-5224, 5226, 5229, 6147, 6300, 6447, 6689, 6755, 7004
 Arduboy 3277, 3286
 Arduino Board 3274-3277, 3279-3280, 3282, 3286
 Argumentation 2643, 4632, 4637, 4920, 5739, 7271
 Aristotle University of Thessaloniki 3374
 Art Practice 4146
 Artifacts 493, 583, 653-654, 789, 966, 1359, 1365, 1662, 2020, 2156, 2393, 2395, 2444, 2450, 2779-2780, 2910-2911, 2914, 2924, 3243, 3292, 3592, 3794-3801, 4115, 4120, 4191, 4255-4257, 4260-4263, 4279, 4287, 4301, 4310, 4629, 4638, 4642, 4666, 4670, 4672, 4676, 4679, 4810, 4817, 5226, 5229, 5438, 6185, 6354, 6779, 6782, 6785, 6863, 6933, 7089, 7168, 7293, 7295, 7314, 7448, 7452, 7454, 7458, 7487, 7530, 7584, 8069
 Artificial Ethics 88-90, 93-94, 97
 Artificial Intelligence (AI) 1-8, 14, 94, 98, 106-108, 114-115, 120, 124, 126, 144, 154, 158, 180-181, 183, 203, 211, 226, 229, 231, 238, 256, 453, 497, 643, 683, 690, 963, 1057, 1511, 1814-1815, 1824, 1920, 1994, 2003, 2020, 2108, 2370, 2720, 2810, 2818, 2824, 2933, 3284, 4004, 4013, 4326, 4349, 4528, 4543, 4547, 4736-4737, 4744-4745, 5561, 5748-5750, 5755-5756, 5834, 5889, 5934, 5940, 6205, 6378, 6459, 6487, 6579-6581, 6696, 6833, 6879, 6907, 7334, 7336, 7704-7706, 7710, 7720
 Artificial Moral Agents 88-89, 91, 97
 Artificial Neural Network Architecture 6640
 Artificial Neural Networks 3-4, 6-8, 100, 108, 112, 120, 144, 148, 151, 181, 205, 232, 441, 450, 493, 496-497, 1519, 1795-1796, 2085, 3107-3108, 4533-4534, 5169, 5171-5172, 5175, 5177-5178, 5182, 6582, 6590-6591, 6631, 6633, 6640, 6642-6643, 6660, 7352
 Artificial Sport Trainer 7357
 Asset Management Systems 3113
 Asset Valuation 1-3, 14, 5670
 Assistive Listening Device 323-324
 Assistive Technologies (AT) 266-271, 276-281, 286-287, 297, 314-316, 318, 320-321, 323, 492, 501, 504, 522, 528, 2522, 2531, 3517, 3525, 6392, 6397, 6464, 6471, 6474
 Assistive Technology Device 267, 276, 528
 associative network 4041, 4050, 4066-4068
 Associative Network Technique 4050
 Assurance 374, 380, 810, 820-825, 827-828, 830, 1107, 1132, 1134, 1335, 1455, 1670, 1683, 1892, 2561, 2763, 2883, 3087, 3173, 3175, 3182, 3210, 3529, 3581, 3587, 3782, 3793, 3843, 3881, 3887, 3899, 4810, 4901, 5158, 5297, 5367, 5383, 5459, 5461, 5475, 5516, 5583, 5654, 5699, 5742, 6038, 6928, 7401, 7566, 7592, 7615, 7617, 8006
 Assuring Quality 3173, 3934, 3939-3940, 3944
 ASTM 7448-7451, 7453-7454
 Astroturf 3574, 3577
 asymmetry of information 2833-2834, 2836-2837, 2840
 Asynchronous Communications 7108, 7124
 Asynchronous Concurrent Processes 6836, 6843
 Asynchronous e-Learning 5808
 Asynchronous Interaction 3903-3904, 5833, 5840
 Asynchronous Learning 1461, 1491, 3902, 3909
 Athletic Trainers (ATs) 5810-5811, 5813-5816, 5819
 Atkinson 133, 2319-2320, 2444, 3020, 3506, 3606, 5698

Index

- Atlas Transformation Language (ATL) 7457
ATM 78-79, 81-85, 166, 660, 2870, 4987, 4992, 5707, 6219-6220, 6813, 7369, 7986
Atom 2673-2674, 2898, 3079-3080, 4437, 4742, 4745
Atomic Transaction 1091, 8057-8059, 8061, 8065
Attachment Theory 4763, 7190-7191, 7199
Attack Scenario 1693
Attention Deficit Hyperactivity Disorder (ADHD) 3262, 3271
Attention-based IT Infrastructure 4822, 4827-4830, 4835
Attribute evaluation 2094
AUC (C-Statistic) 2223
Auditing 3, 107, 376, 755, 820-825, 827, 830, 1938, 4852, 5372, 5617, 5741, 6935
Auditory Effects 6135
Augmentative And Alternative Communication 267-269, 276, 6464, 6474
Augmentative and Alternative Communication Device 276
Augmented Reality (AR) 391, 544, 983, 985, 987-989, 993, 2552, 2857, 3463, 3466-3467, 3471, 4031-4032, 4037, 4073, 4114-4115, 4202, 4297, 5104, 5108, 5110, 5116, 5218, 5768, 5889, 6186, 6363-6364, 6393, 6397, 6470, 6480, 6489, 6942, 7806, 7867, 7907
Authentication 50-51, 174, 373, 653-654, 661, 868, 890, 1033, 1068, 1240, 1247, 1359, 1659, 1670, 1676-1680, 1683, 1840, 2644, 2657, 2660, 2716, 2754, 2756, 2784, 2815, 2908-2915, 3565, 3691, 4643-4644, 4651, 4927, 4981, 4984-4987, 4992, 4994, 4996-4997, 5003, 5008, 5296, 5313, 5522, 5545, 5575, 5581, 5604, 5928, 5986, 5991, 6063, 6137, 6144, 6207, 6258, 6953, 7590-7591, 7723, 7776, 7915
Authorities 334, 523, 1018, 1022, 1545, 1547, 1560, 1698, 1837, 2276, 2351, 2423, 2439, 2724, 2804, 2863, 2945, 2947, 2990, 3374, 3377-3378, 3380, 3394, 3461, 3579, 3582-3583, 3592, 3594, 3604-3608, 3624, 3637, 3644, 3910, 4159, 4385, 4525-4526, 4704, 4709, 4816, 4856, 4899, 4969, 4971, 4975, 5381, 5383, 5465, 5852, 5904, 6101, 6163-6165, 6276-6277, 6280, 6672, 6714, 6859, 6952, 7295-7296, 7769, 7907
Autism 289, 2434, 2525, 3271, 3374, 4128, 6393, 6464
Auto-Composition 2832
Automated modelling 7488, 7490, 7500
Automated Readability Index 1502, 1507
Automatic Emotion Recognition 132, 143
Automatic Identification And Data Capture Technologies (AIDC) 4841, 5519, 5576, 5581, 5597, 5604, 5928, 7777
Automatic Identification And Data Capture Technologies (AIDC) Technologies 4841, 5280, 5576, 5581, 5604, 5918, 5928
Automatic Item Generation 2369-2370, 2376, 2378
Automatic Processes 1824
Automation 106, 158, 184, 349, 541, 543, 621, 628, 637-638, 821, 823-824, 886, 1030, 1075, 1104, 1115, 1227, 1318-1319, 1330, 1340, 1354, 1509, 1597, 1619, 1624, 1783-1784, 2114-2115, 2200, 2629, 2977, 3149, 3151, 3274-3276, 3403, 3520, 3691, 3694, 3916, 3930, 4127, 4201, 4216, 4736, 5322, 5372, 5375, 5491, 5519, 5531, 5858, 5911, 6106, 6478, 6480, 6500, 6522, 6529, 6533, 6833, 6848, 6855-6856, 7436, 7438, 7448, 7453-7454, 7463, 7506-7508, 7568, 7574, 7696, 7775, 7779, 7874
Automotive Industry 535, 965-966, 1686, 5378-5381, 5383-5384, 5391, 5441, 6676-6677, 6680
Automotive original equipment manufacturers 5386
Autonomic Networking 1985, 6504
Autonomic System 6499-6500, 6506
Autonomous agents 4126, 4267, 4273, 4798
Autonomous Learner 2589, 2598
Autonomy 89, 461, 651, 689, 743, 802, 1223, 1280, 1407, 1490, 1518, 1696, 2321, 2445, 2456, 2532, 2534-2536, 2544, 2548, 2552, 2665, 2773, 2775, 2777, 3000, 3009, 3012, 3035, 3239, 3244, 3249, 3252, 3383, 3643, 3646, 3678, 3683-3684, 3786, 3808, 3840, 3946, 4127, 4151, 4162, 4269, 4320, 4461, 4892, 5096, 5116, 5198, 5292, 5401, 5477, 5630, 5634, 5853, 6919, 7059, 7102, 7218-7219, 7249, 7305, 7336, 7359, 7443, 7760, 7769-7770, 8026, 8104
Avalanche Photodiode 7730
Avatar 133, 803, 980, 2410, 2531, 3242, 3262, 3350, 3353-3354, 3358, 4148, 4151-4153, 4234, 4237-4240, 4244, 4300, 4302, 7791, 7887, 7936
Aviation Safety 666
awareness education 4965, 4968-4971, 4973
Away Messages 7076
Axiology 97, 4189, 4336
Axon 205, 6582-6583, 6594
- ## B
- Backlash 676, 3265, 3383, 3391, 5231, 6163, 6166
Backpropagation 121, 123, 131, 149, 183, 185, 6581-6582, 6647, 6657
Badges 997, 1000, 1477, 3241-3242, 5277, 7897
Bagging 410, 1826, 1829, 1834, 1909, 2085, 4947
Bag-of-Words model 1906
Balanced Scorecard 1596, 2171, 2176, 3530, 5638, 5714, 5716-5719

- Ballistic Conduction 2680, 2684
band gap 2673-2675, 2678, 2684
Bandwidth 170, 745, 1037, 1108, 1164-1168, 1170, 1241, 1503, 1774, 1947, 1951-1953, 2901, 3693, 4004, 4009-4013, 4019-4021, 4024, 4152, 4319, 4549, 4551-4553, 4628, 4991, 5575, 5972, 6010, 6058, 6069, 6095, 6101, 6109, 6199, 6236, 6253, 6264, 6267-6269, 6271, 6348, 6351, 6481, 6537, 6547-6552, 6558, 6561, 6611-6614, 6616, 6621, 6676, 6678, 6680, 7229, 7251, 7564-7565, 7575, 7603, 7775, 7796, 8089, 8097
Bandwidth Management Server 6549
Bank Profit 47
Bank Risk 47
Bank Size 47
Bankruptcy 1, 3, 6, 8, 802, 5552, 6813-6816, 6818, 6821-6823
Barcoding Technology 5284, 5526, 5604, 5928
Barriers of Open Data 3041
Baseline Predictors 1761-1763, 1770
Bash Boards 4247, 4253
basic ethical principles 4884, 4886, 4895-4896
basic personalized recommendation 1736-1737, 1745
Basic Research 6101, 6769, 6778
BAT 183, 3156-3157, 3160-3161, 3163, 3165, 3168-3170, 4371, 7357, 7697
Bat Miner 7357
Bayes Theorem 4940, 5684, 5689, 6977
Bayesian Approach 2188, 2717, 5679-5681, 5684-5687, 6895
BCI 5887, 5893-5894, 5896, 5899
BD 1667-1668, 1670, 1672, 1674
Beacon 6218, 7813
Beebots 2494, 2496-2497, 2501, 6863-6865, 6869
Beehive Effect 1058, 1061
Behavioral Effects 6135, 7311
Behavioral Intelligence 286
Behavioral Intention (BI) 531-533, 832, 841, 2253, 2762-2763, 3583-3584, 3586, 3590, 3654-3662, 3665, 3787, 4281, 6071, 6403-6404
Behaviorally Anchored Rating Scales (BARS) 4339-4340, 4346
Behavior-Based Robots 5940
behavior-score mapping 1764, 1768, 1770
Benchmark 234, 251, 257, 610, 923, 925, 1772-1777, 1782, 1788, 1888, 1891, 1953, 1955-1956, 2225, 3377, 3688, 3713, 3825, 3885, 5011, 5126, 5434, 5816, 6538, 6543, 6596, 6633, 7131, 7562, 7696
Benchmarking 351, 543, 587-589, 592, 595, 600, 609, 615, 734, 932, 1772, 1774-1777, 1782, 1891, 2170, 2172, 2176, 2688, 2694, 3190, 3529, 3531, 3534, 3758, 3765, 3789, 3885, 4709, 4711, 4829, 4922, 5299, 5402, 5434, 5444, 5541, 5583, 5763, 6536-6537, 6543
Benders decomposition 5411-5418, 5420, 5511
Benefits Management 3529, 3756, 3765
Benefits Of Open Data 3032, 3034, 3041, 4444
Bernoulli vacation schedule 1570, 1573, 1576, 1583
Between-Methods Triangulation 6757, 6765
Betweenness 1746-1747, 1751-1754, 1756, 6507, 6509-6510, 6516-6517, 6520, 6964, 6972, 7151, 7160
Betweenness Centrality 1747, 1751-1754, 1756, 6507, 6509-6510, 6516-6517, 6520, 7151, 7160
beyond STEM 6865
Bhattacharjee 831, 835, 841, 894, 1400-1401, 6024
BI portal 5544-5546, 5549
Bibliographic Database 4547
bibliographical repository 5245
Bibliometric Indexes 4406-4407, 4420, 5246
Bibliometrics 1328, 1330, 3482, 4416, 6739, 6747, 7046, 7051
Bidding 535, 538, 541, 1169, 3530, 5697, 6837, 7985
Bi-Directional Business/IT Alignment 601-605, 607-608, 2965
Big “L” Literacy 1516
Big Data Analytics 329, 333-335, 349, 373, 407, 862-864, 866, 870-872, 952, 955, 1774-1775, 2008, 2019, 2687, 2692, 2810, 2813, 2817, 2822, 4013, 4089, 4844, 4955, 5296, 5375, 7141-7142
Big data mining 340, 342, 344-345, 406, 414
Big Databases 340-345, 1808, 1908, 2200
Big Society 6695, 6699, 6702-6706
Big-Data Technology 2006, 2019
BIM modeling 7868
BIM+AR 987, 989-990
Binary Image 213, 219, 4931, 4997
Binary System 6835
Binary-Coded Decimal (BCD) 4607, 4618
Binding Constraint 4606
Bioelectromagnetics 1249-1250, 1259
Bio-Imaging 419, 425, 5111
Bioinformatics 184, 419-425, 440, 453, 455-458, 460-461, 851, 1680, 1775, 1825, 1908, 1965, 1976, 5108, 7696
Biomechanics 302, 7362
Biomedical Engineering 506, 518, 528, 1318-1319, 1330, 1354
Biomedical Object 506, 508, 518, 520
Biomedical Ontology 424, 455, 464
Biomedical Portals 456, 460
Biometric Systems 166-167, 2918, 4651

Index

- Biometrics 166-167, 1677, 2918, 3694, 4937, 4985-4987, 4991-4998, 5003, 6739
- Biometry 4651
- Biomimetics 439
- bionics 5889, 5894
- Biostratigraphic Markers 3489, 3500
- bipolar analysis 7282, 7284
- Bipolarity 7283, 7289-7290
- Bitcoin 2749, 2784-2790, 2815, 2861-2876, 2878-2880
- bit-for-bit copy 1384, 1394
- Bitpay 2865, 2868
- Black Swan 5704
- Black-hat SEO 8048
- Black-Scholes PDE 15-18, 20-24, 26, 28-30, 35
- Blended Delivery 7394-7395, 7397-7398
- Blended Learning 1476, 1487-1488, 1490-1493, 1498, 2423-2425, 2430, 2459, 3238-3239, 3242-3244, 3247, 3340, 3839, 3845, 3849, 3860, 5154, 5788-5789, 5792-5793, 5799, 5802, 5811, 5822, 5828, 6347, 6391, 7225, 7642, 7912, 7914, 7916-7917, 7921-7923, 7927, 7942
- Blended Learning Environment 1498, 3238, 3242-3244, 3340, 7916, 7922
- Bletchley Park 255, 258, 260
- Blind Spots 899-901, 905-907, 909-910, 1684-1685, 1687, 1690-1691, 1693
- Block Chain 2785, 2790, 2815, 2863-2868, 2870-2876
- Block-based programming 2396-2397
- Block-Decomposed Mathematical Model 5420
- Blogs 350-351, 395, 461, 464, 563, 707, 745, 783, 834, 1167, 1492, 1533, 1535, 1695, 1697, 1710, 1874, 2069, 2254, 2256, 2292, 2325, 2329, 2481, 2485, 2493, 2523-2524, 2539, 3035, 3239, 3404, 3409, 3613, 3667, 3794, 3934, 4047, 4054, 4071, 4088, 4092, 4098, 4161, 4190, 4215, 4279, 4295, 4359, 4362, 5049, 5252, 5255-5258, 5262-5264, 5266-5267, 5559, 5562-5565, 5569, 5776, 5778-5779, 5782, 5785, 5808, 5909, 6019, 6252, 6302, 6360, 6391, 6393, 6492, 6756, 6759-6761, 6781, 6784, 6941, 6945, 6974, 6994, 6996-6997, 7006, 7025, 7036, 7043, 7056, 7096, 7106, 7112, 7128, 7130, 7161, 7182, 7273, 7629, 7797, 8017, 8031, 8035, 8037-8038, 8040, 8042, 8045, 8050, 8055, 8070
- Blood Flow 1006, 5856, 6117-6118, 6120, 6129
- Blood Pressure (BP) 1006, 1012, 1015-1016, 2373-2374, 3452, 3693-3694, 3707, 5908, 5911, 6119
- Blueprint 573-575, 578, 584, 586, 914, 926, 956, 3614, 3939, 5216, 7471, 7862, 7868, 7983
- Blueprint Technique 7868
- Board Game 147, 154
- BOCR Analysis 7284, 7290
- body ownership 4235-4237, 4239, 4244
- Body-Area-Network (BAN) 3697, 6116-6117, 6122, 6319, 6334
- Book Banning 7292
- Book Value 5056, 5060, 5066
- Books of Knowledge 5704
- Bootstrapping 837, 1828, 1834, 1909, 1923, 2959
- Bop 5015, 6995, 7001
- Botnets 4838-4839, 6215
- Bottom-up Requirements 602-604
- Bottom-Up Tree Search 4514
- Boundary Control 16-17, 20, 23-24, 28-30, 35
- Boundary lubrication 436-437
- Boundary Permeability 6986, 6988, 6992
- Boundary Turbulence 6986, 6992
- Bounded Rationality 1017, 2076-2077, 2117, 2989-2990, 3228, 4915, 6634, 6713, 6725
- BPMN 766-769, 771-774, 788-789, 794, 895, 2927, 4815, 5434, 5438, 5440-5441
- Brain Games 7035
- Branch And Bound 4371, 6508, 6511, 6520
- Branch Metric Unit (BMU) 6309-6310, 6318
- Brand Awareness 712, 2817, 2858, 3369, 3861, 4053, 4098, 7126, 7131-7132, 7134, 7139
- Brand Community 708, 719, 7139
- brand element mix 4038, 4047-4048, 4051
- Brand Identity 4038-4039, 4051, 8079, 8082, 8086
- Brand Image 719, 3863, 4038-4039, 4051, 4056, 5293, 5768-5769, 8086
- Brand Reputation 7132, 7139
- Breadboard 3280, 3284, 3286
- Breadth-first search 1978, 1983
- Breast Cancer 413-414, 471, 1794-1802, 1804-1805, 2089, 2218, 3448-3449, 3451-3455, 3459, 3770, 3775, 7079
- BREF 3159, 3170
- Bring Your Own Device (BYOD) 374, 5981, 6382, 6390
- Broadband 1101, 1167, 1552, 1657, 1707, 2007, 2239, 2246, 2342, 2769, 2804, 3150, 3184, 3549-3551, 3553-3555, 3558, 3560, 3693, 4212, 4619, 4622-4623, 4628, 4709-4711, 5018, 5934, 6094-6096, 6105, 6266, 6268, 6319, 6330, 6351, 6690, 6735, 7028, 7214, 7225, 7228-7229, 7253, 7258, 7777, 8098-8099
- Broadcast Search Approach 572
- Bsc 1388, 2170, 2172, 3889, 4452, 4727, 5638-5640, 5644, 5646-5647, 5649, 5714, 5716-5719, 7203-7204, 7211
- Build Security In (BSI) 7526, 7538
- Building Blocks 396, 411, 464, 477, 621-622, 635-636, 640, 650, 1443, 1529, 1630, 1849, 2756, 2921,

- 3289, 4015, 4287, 5264, 5401, 5541, 5607-5608, 5614, 5619, 5625, 6064, 6436, 6480, 6487, 6499, 6837, 6870, 7128, 7223, 7227, 7268, 7770
- Building Information Modelling (BIM) 540, 542, 983-984, 987-989, 993, 7868
- Bulicide 4167
- Bullying 1695-1699, 1702-1703, 1706-1710, 1723-1724, 1726-1728, 1733, 2316, 2616-2617, 2619-2621, 2624, 4157-4163, 4167-4173, 4177, 4245-4249, 4253-4254, 7320
- bundled payment 2095, 2098, 2106-2107
- Bundling 341, 545, 1158, 1160, 1627, 1629-1630, 1636, 2098, 6731
- Business Activity 565, 568, 711, 2344, 2740, 2747, 2930, 2999, 3184, 4348, 4399, 4854, 5549, 8057, 8059-8061, 8065
- Business Analytics 327, 862, 864, 872, 930, 933-934, 936-938, 945-947, 950, 1146, 1581, 2213, 2220, 3077, 5158, 5164, 5292, 5296, 5504, 5511-5512
- Business Applications 869, 2563, 2752, 2794, 4278, 5346, 6631, 6635, 6642-6643, 6651, 7229, 8036, 8056
- Business Architecture 573-577, 579-580, 582-584, 622, 627-628, 637, 640, 701, 5610, 5614, 5619
- Business Ecosystem 575, 2350, 2354, 2743, 2747, 5305
- Business Engineering 620-621, 626, 631, 636-645, 650, 5609, 5625
- Business Entity (BE) 653, 768, 788, 790-791, 795-796, 2774, 2783, 2804, 4784, 4824, 5401, 5617, 5678, 5910
- Business Ethics 675, 3173-3178, 3182, 4318, 5293
- Business Group 61-62, 70, 7261
- Business Intelligence (BI) 333, 349, 363, 576, 847-849, 851, 856-857, 859-860, 862, 864, 900, 943-945, 947, 951-952, 954, 958-959, 963, 1378, 1383, 1400, 1587, 1618-1619, 1623-1624, 1911, 2085, 2932-2934, 2937-2938, 2941, 3530, 3863, 3883, 3898, 4073, 4829, 5158, 5160, 5292, 5519, 5538-5541, 5546, 5549, 5887, 6907-6909, 6971, 8038
- Business Model 563, 565-566, 568, 572, 577, 582, 603, 610-611, 618, 673, 676, 694-696, 703, 706, 874, 930, 1026, 1034, 1102, 1141-1143, 1146-1147, 1149, 1152, 1157, 1160-1161, 1167, 1356, 1630, 1660, 2345, 2348, 2739, 2741-2746, 2885, 2925, 4560-4566, 4569, 4837-4838, 5306, 5316, 5382, 5388, 5390, 5428, 6482, 6547, 7121, 7273, 7525, 7532, 7583, 7805, 7901, 8043
- Business Model Framework 4563-4565, 6547
- Business Model Innovation 4565, 4569
- business modeling 895, 1141-1143, 1146-1149, 1152
- Business Modelling with Cloud Computing (BMCC) 1152
- Business Model-Oriented Technology Management 4565, 4569
- Business Planning 573, 576, 1581, 1691
- Business Process as a Service (BPaaS) 7538
- Business Process Execution Language (BPEL) 8062, 8065
- Business Process Modeling (BPM) 765, 768, 2775, 4815, 5434, 5438, 5444, 7526, 7531, 7589, 7776
- Business Process Outsourcing (BPO) 5478, 5484, 5488, 7261, 7268
- Business Process Reengineering 892-893, 2966, 3621, 5054, 5402, 5444, 7429
- Business Processes 352-353, 357, 360, 363, 554, 574-575, 577-578, 582, 639, 650, 755, 765-768, 774, 788, 790, 795, 816, 821, 825, 828, 874-875, 878-880, 888-896, 898, 913-914, 922, 924, 934, 956, 1071, 1147, 1399, 1570, 2154, 2157, 2159, 2162-2163, 2348, 2354, 2566, 2773-2776, 2780, 2783, 2923, 2928, 2933-2934, 2941, 2943, 2945, 2947-2948, 2950-2951, 2953, 2956, 2962, 2965-2966, 2968-2970, 2974-2976, 2979, 2986, 3177, 3193, 3527, 3603, 3606-3607, 3913, 3916, 4279, 4284, 5046, 5048-5052, 5054, 5287, 5291, 5297, 5310, 5322, 5371, 5387, 5410, 5433-5434, 5436-5441, 5444-5445, 5484, 5490, 5493, 5544, 5547, 5549, 5560, 5613, 5617, 5625, 5638, 5644, 5646-5647, 5705, 5775, 6546, 7403, 7422, 7526, 7532, 7589, 7775-7776
- Business Strategy 611, 671, 681, 689, 816, 825, 875, 883, 912, 914, 918, 921, 2640, 2745, 2806, 2922, 2930, 2965, 2969, 3021, 3025, 3031, 4348-4349, 4575, 4686, 5324, 5529, 5549, 5909, 7118, 7373
- Business Sustainability 609-615, 617-618
- Business Transaction 2154-2155, 2157-2159, 2161, 2165, 5620-5621
- Business Transformation Projects 621, 631-632, 638, 645, 5607, 5609, 5623
- Business Value 602, 875, 923-926, 928, 952, 954, 1397, 1586, 2921, 3757, 3963, 5432, 5776
- Business-To-Business (B2B) 2740, 2747, 2772, 2783, 2792, 4200, 5285, 5311, 5367, 5380, 5388, 5397, 6094, 7875, 7982, 8056
- Business-To-Consumer (B2C) 638, 2740, 2747, 2886, 4088, 5367
- Buyer price 532-533, 535
- Bystanders 1696, 1698, 1702, 1725, 2620, 4158, 4162, 4167, 6714, 6726
- Byzantine Fault 1238-1240, 1243-1244, 1247, 1647-

Index

1648, 1650-1651, 1655, 3056-3058, 3060, 3062, 3065, 4928-4931, 4935, 8056, 8060-8062, 8065
Byzantine Fault Tolerance 1238-1240, 1243-1244, 1247, 1647-1648, 1650-1651, 1655, 3056-3058, 3060, 3062, 3065, 4928, 4930-4931, 4935, 8056, 8060-8062, 8065
Byzantine quorum system 4928, 4931, 4935

C

C# 368, 7395, 7459-7460, 7462, 7464-7465, 7570-7571, 7573, 7882
C++ 410, 1969, 2047, 3273, 3275, 4608, 5511, 6057, 6068, 6260, 7448, 7454, 7459-7460, 7464-7465, 7699, 7719
Cache Oblivious 1950-1951, 1961
Cache Sensitive/Conscious 1950, 1952-1953, 1961
Caching 1095, 4009, 6557-6559, 6561, 7996, 8088
Cacophony 3562, 5146
CAD Assessment 7382, 7385-7389, 7391
CAD Design File 4382, 4388
CAD Education 7384-7385, 7387-7389, 7391
CALL 38, 191, 230, 409, 412-414, 563-564, 683, 715, 734, 756, 760-761, 929, 933, 1000, 1022, 1045, 1050, 1054, 1059, 1117, 1189, 1300-1301, 1512, 1597, 1624, 1638, 1737, 1763, 1964, 2031, 2039, 2225, 2456, 2516, 2704, 2714, 2716, 2777, 2921, 2970, 3009, 3062, 3331-3332, 3356, 3385, 3393, 3488, 3496, 3587, 3616, 3642, 3672, 3692-3693, 3715-3716, 3721, 3851, 3854-3856, 3929, 3996, 4000, 4023, 4058, 4077, 4151, 4283, 4298, 4310, 4317, 4328, 4373, 4392, 4422, 4429, 4504, 4506-4509, 4511, 4536, 4542, 4632, 4697, 4884, 4896, 4909, 5040-5041, 5226, 5372, 5424, 5518, 5628, 5652, 5726, 5730, 5790, 6031, 6062, 6096, 6099, 6253, 6267-6270, 6272, 6278, 6354, 6377, 6609-6610, 6612-6613, 6615-6616, 6660, 6691, 6696, 6756, 7117, 7144, 7192, 7219, 7251, 7261, 7315, 7320, 7472, 7476, 7564, 7665, 7669, 7737, 7844, 7854, 8023-8024, 8026, 8037, 8056
Call Admission Control 6267-6269, 6272, 6610, 6612-6613, 6616
Calm Computing 7766, 7774
Calvin Mooers 4537
camera phone 6149, 6296-6301, 6303, 6306
campus bridging 1063, 1069-1071, 1074
Canter 1468-1469, 1605
Capability Maturity Model (CMM) 1405, 3526, 3528-3530, 3537, 5050, 5480, 6928, 7553-7555, 7583
Capability Maturity Model Integration (CMMI) 1405, 3529, 6928, 755

Capitalization 45, 47, 2742, 2748, 2870, 2876, 2992, 5058, 5060-5061, 5675, 7118
CAPTCHA 8066-8072, 8074-8075, 8078
CAPTCHA Usability 8067-8072, 8074-8075, 8078
Captology 7785, 7795
Carbon Footprint 3085, 3371, 5378, 5380, 5446, 5455, 7749
carbon nanomaterials 2675, 2679-2681, 2684, 2897-2899, 2901-2903, 2906
carbon nanotubes 2673-2675, 2677-2681, 2684, 2897-2903, 2906
career anchor 5633, 5637
Cartesianism 1855
Cartograms 1200, 3479, 3482
Case History 5663, 5668
Case Management 789, 798, 5440
Case Study 510, 513, 694, 700-701, 703, 771, 780-781, 811-813, 978, 1208, 1277, 1551, 1605, 1607, 1611, 1615, 1630, 2121-2125, 2145-2148, 2494, 2558, 2626, 2629, 2920, 2923, 2925-2927, 2971, 3361, 3365-3366, 3420, 3428, 3818, 3882, 3884, 3904-3905, 3982, 4375, 4481-4482, 4484, 4635, 4775, 5217, 5319, 5324, 5411, 5451, 5491-5492, 5497, 5500, 5517, 5521, 5571, 5596-5597, 5600, 5638-5639, 5647, 5661-5664, 5668, 5698, 5855-5856, 5918, 5921, 5923, 5985, 6165, 6167, 6492, 6819, 6862, 7044-7045, 7185, 7496, 7526, 7539, 7542-7544, 7550, 7650, 7815, 7819, 7821, 7874, 7914, 8013
Case-Based Reasoning 1, 4, 7-8, 14, 184, 186, 4573, 4797, 4800, 4809
Cash Dispenser 78, 84
Cash Flow 3, 63, 67, 803, 1584, 2166, 2734, 3018, 5398, 5461, 6820-6821, 6928
Casuistry 4921, 4925
Catalytic Oxidation 3169-3170
Categorical Data 1841, 1844
Catheter 1006, 1008, 3797, 3799, 3804
Causal Map 2404, 4571
Causality 39, 951, 1044-1045, 1848, 1920, 2960, 3770, 4768, 5642, 5646-5647, 5996, 6635, 6735, 6917, 7578
CAVE 1068, 1845, 2359, 2442, 3118, 3130, 4114, 4536, 4695, 5106
CCEGK 1668, 1670-1672, 1674
CDMA 5572, 6138, 6267, 6682
Celebrity on the Web 7121, 7124
Celje Regional Museum 4077
Cell Differentiation Counters 5853, 5862
Cell Phone 952, 2315, 2319, 2327, 2726, 3414, 3612, 3614, 3620, 4967, 5021, 5027, 5106, 5983, 5992,

- 5995-5996, 5999-6000, 6004-6005, 6151, 6217, 6245, 6277-6280, 6282, 6284, 6302, 6306, 6348, 6354, 6362, 6375, 6712, 6987, 7098-7100, 7102, 7216-7217, 7369, 7373
- cell phone conversation 5992, 5995-5996, 5999-6000, 6004-6005
- cell towers 6287, 6289
- Cellular Phone 2335, 3961, 5572, 5767, 6300, 6306, 6382, 6716, 7216-7217
- Censorship 3494, 4196, 4761-4762, 4764, 4766, 4768-4770, 4773, 4922, 4924, 6482, 6950-6952, 6956, 6960, 7292-7298, 7301
- Center for Medicare and Medicaid Services 3806
- Center for Multiple Attention 2435, 2441
- Central Focus 7687, 7691
- Centrality 975, 1746-1756, 3647, 3976, 3978-3980, 3983, 3985, 4360, 5038, 5098, 6507-6510, 6513-6517, 6520, 7151, 7160
- Centre of Gravity (COG) 3224
- Certificates of Completion 7889, 7897
- Certification 373, 583, 620, 638, 641, 1107, 1469, 2424, 2835-2836, 3133, 3189, 3437, 3626, 3810, 3814, 3833, 3961, 5313, 5423, 5428, 5458, 5460-5462, 5464, 5531, 5533, 5670, 5813-5814, 5816, 5834-5835, 6927-6928, 6935, 6952-6953, 7402, 7625, 7634, 7682-7683, 7687-7688, 7691
- CeWebrity Business 7121, 7124
- CGRA 4021, 4023
- Challenged Projects 5668
- Change Leadership 3621, 3625, 3627-3630, 5194, 5199
- Change Management 620-622, 629, 656, 810-811, 913, 922, 956, 2048, 2155-2156, 2980, 3624, 3626, 3628-3630, 3633, 3822-3823, 3849, 4328, 5274, 5540, 7425, 7515
- Channel of Interaction 3963, 3965, 3967, 3971, 3975
- Charity 2320, 4923, 5461, 6218
- CHEA 3173-3174, 3182
- Check-In 5521, 7798, 7804
- Checkpoint algorithm 2143, 2145-2146, 2153
- Checkpointing 1116-1117, 1119-1126, 1128, 1955, 4928
- chemistry learning 3248, 3259
- Child Molester 1375
- Child Pornography 1366-1369, 1371, 1375, 1706, 4320, 4924, 7294
- choice function 3226, 3230, 3235-3236
- Choreography 637, 2778, 6569, 6578
- Choropleth Maps 1200, 3475, 3483
- Chromosome 472, 475-476
- Cinematic Cyborg 3358
- Ciphertext 1675, 4978-4979, 4981-4982, 4984
- Circadian Rhythm 3452-3453, 3459
- CIS System 3749
- citation histogram 5073-5074
- Citizen Science 1070-1071, 1074, 8023
- Citizenship 672, 676, 759, 1520, 1697, 2228-2230, 2234-2235, 2237-2238, 2240, 2242, 2247, 2325, 2492, 2502, 2520, 2546, 2548, 2571, 3242, 3638, 3676, 4255, 4715, 5578, 6714, 6729, 6870, 7249, 7625, 7673
- civic duty 3208-3209
- Civic Engagement 2325, 2481, 3578, 3676, 6188, 6689, 6703-6704, 7060
- Civic Participation 2245, 3578, 3667-3668, 3670-3673, 3676, 4283, 6184, 7851
- Civil unrest 6275
- CK 6308, 7624, 7662, 7665-7666, 7670
- Class Mutation Operator 7469
- Classification Context 194, 196-197
- Classification Learning 191-192, 197
- Classification Models 135, 441, 1922, 5170
- Classification Test 200
- classification tree 1941, 2132-2133, 2136, 2141
- Classifier 4, 7, 137-138, 160, 411-412, 496, 498-502, 1825-1831, 1900, 1902, 1923, 1926, 2086, 2088, 2134, 2136, 2201, 2216, 2219, 2716, 4644, 4941, 4950, 6197-6200, 6203, 6206, 6288-6289, 6590, 6633
- Claude Shannon 144, 551, 1187, 1973, 2121
- Click-Through Rate (CTR) 5752, 8055
- client-side handheld programming 6060, 6068-6069
- Climate Change 671, 3114, 3116-3119, 3127, 3130, 3183, 3371-3372, 3374, 3377-3378, 3380, 3491, 3528, 4404, 5081, 5423, 5446, 5450-5451, 5465, 7815, 7817, 7819, 7821
- climate change mitigation 5446, 5451
- clinical decision support systems 5864
- Clinical Effectiveness 3807, 3814
- Clinical Research 3678, 6776, 7181-7183, 7186-7187, 7189
- Clinical Trials 406, 3788, 4227, 5016, 5788, 5941, 6776, 7181-7183, 7186-7187, 7189
- clip transform 1318
- Clique 421, 1746, 3639, 4181, 6507-6508, 6510-6517, 6520
- Clique Size 1746, 6507-6508, 6512-6517, 6520
- CLIR 4471-4472, 4479, 4490-4498
- closed loop supply chain 5378-5384, 5386
- Closed loop supply chain management 5380, 5386
- Closeness 245, 441, 721, 1746-1747, 1752-1754, 1756, 2317, 3136-3137, 5098, 5403, 5994, 5999, 6185, 6507-6508, 6510, 6516-6517, 6520, 7074, 7154

Index

- Closeness Centrality 1747, 1752-1753, 1756, 6507-6508, 6510, 6516-6517, 6520
- Cloud (Based) Computing 187, 235, 238, 374, 380, 456-459, 461, 517, 725, 880, 886-887, 895, 925-926, 985-986, 989, 1026-1027, 1029, 1033-1037, 1039, 1041, 1044, 1053, 1058-1059, 1061, 1063, 1067, 1069, 1071, 1074, 1076, 1078, 1101-1110, 1114-1116, 1119, 1128-1132, 1135, 1137, 1140-1149, 1152-1153, 1163, 1167, 1174-1175, 1181, 1184, 1356-1363, 1365, 1517, 1627-1631, 1634, 1636-1637, 1673, 1675, 1985, 1989, 1991-1992, 2220, 2303, 2459, 2465, 2591, 2742-2743, 2810, 2813, 2817, 2821-2822, 2824, 2965, 2968-2969, 2971, 3038, 3188, 3275-3276, 3455, 3462, 3464-3465, 3530, 3594, 3726, 3732, 3810, 4203, 4361, 4584, 4593, 4732-4733, 4839, 4884, 4891, 4894, 4896, 4909, 5105, 5108, 5291, 5307, 5373, 5376, 5709, 5711, 5713, 5981, 6339, 6481, 6566, 6658, 6727-6728, 6730-6735, 6737, 6807, 6812, 7170, 7447, 7525-7526, 7530, 7532, 7534, 7566, 7588, 7700, 7754, 7761-7762, 7767, 7777-7778, 7984
- Cloud Benchmarking 1782
- Cloud Computing Based Technology 1129, 1140
- Cloud Computing Business Models (CCBM) 1141, 1146-1149, 1152
- Cloud computing governance 1108
- Cloud Computing Services 925-926, 1101, 1105, 1129, 2743, 2971, 5373, 6727-6728, 6730-6735, 6737, 7532
- Cloud Computing System 1035, 1041, 1101-1102, 1106, 1358, 1675
- cloud customer 1029, 1090, 1166
- Cloud Data Security (CDS) 7528, 7530, 7538
- Cloud Deployment Models 886, 1028, 1032, 1105, 1356, 1990, 6728
- cloud ERP 2965, 2968-2971, 2974
- Cloud Migration 1164, 1166-1170, 1172, 7530
- Cloud Models 1090
- Cloud Provider 1027-1029, 1032, 1090, 1107-1108, 1166-1170, 1173, 1357, 1628, 1668, 1776, 2969
- cloud service models 1032, 1356
- Cloud Services 457-458, 925, 989, 1029, 1032, 1090-1091, 1095, 1119, 1157, 1167, 1358, 1628, 1631, 1782, 1990, 2908, 2920-2921, 2970, 3692, 4311, 4460, 4839, 5981, 6730, 7525-7526, 7528, 7530, 7532, 7534, 7538, 8036
- Cloud Solutions 1163, 2965, 4013
- Cloud Storage 1107, 1109, 1167, 1412, 5860, 6733
- cloud suppliers 1159
- Cloud-Based Applications 923, 928
- Cloud-IoT 1991-1992
- Cluster Analysis 408, 440-441, 453, 851, 933, 1628, 1963, 1973
- Clustering 148, 181, 184, 186-187, 338, 340, 350-351, 359, 375, 406-414, 423, 440-441, 443-445, 447-448, 450, 453, 851-852, 860, 1019, 1091, 1093-1094, 1096, 1333, 1354, 1617, 1774, 1788, 1809, 1827-1828, 1896, 1898, 1901-1902, 1909, 1921, 1923-1925, 1927, 1929-1930, 1932, 1962-1966, 1968-1969, 1973, 2071, 2986, 3436, 3438, 3440, 3447, 3538, 3542, 3544, 3548, 4006, 4012, 4443, 4574-4575, 4990, 5104, 5108, 5160, 5163, 5170-5171, 5176, 5665, 6199, 6203, 6430, 6523-6524, 6528-6531, 6595, 6632, 6647, 6961, 6966, 7348, 7351, 7696, 7838, 7983
- CMDB 2920, 2930
- C-n-tuple 1996, 1998-1999, 2002, 2004-2005
- CO₂ absorption 3044-3053, 3055
- CO₂ Emission 7753
- CO₂ stripping 3046, 3050-3052, 3055
- Coalition 1065, 2096-2102, 2105, 2107, 2354, 2484, 4143, 5374, 6105, 6713, 6725, 6748
- Coarse-Grain 4019, 4022-4025
- Coarse-Grained Granularity 4029
- COBIT 601-604, 608, 630, 811, 819, 2155-2156, 2920, 3529
- Cochlear Implants 2433-2434, 5888
- Co-Construction Theory 7081, 7087
- Co-Creation 563, 707-710, 712-715, 719, 2598, 4146, 4150-4151, 4153, 5775, 5777, 6490, 6493, 7003-7004, 7131, 7168, 7588, 7601, 8023
- Coding 206, 234, 290, 422, 424, 460, 466, 472, 474, 477, 483, 490, 556, 1318, 1330, 1354, 1424, 2076-2077, 2392, 2394-2399, 2401, 2491, 2584, 2619, 3078, 3080, 3341, 3466, 3717, 3956, 4052, 4092, 4391, 4448, 4452, 4746, 4849, 5086-5087, 5089, 5108, 5274, 5278, 5280, 5284, 5526, 5604, 5928, 6065, 6095, 6245, 6403, 6448, 6451, 6463, 6488, 6501, 6771, 7340, 7392-7398, 7625, 7959-7960, 7962, 7967, 7996, 8081
- Co-Evolution 97, 4295, 6913
- COG defuzzification technique 3215, 3223
- Cognitive Constraints 6411
- Cognitive Dissonance Theory 3821, 3825
- Cognitive Distraction 6004
- cognitive effects 6126, 6130-6131, 6135
- cognitive engagement 1475, 1480, 1489, 1491, 1498
- Cognitive Feedback 205, 4574, 7516, 7523
- Cognitive Friction 5005, 5011, 5013
- Cognitive Intelligence 286

- Cognitive Load 1401, 1422, 1474, 1480, 2360, 2383, 2444, 4214, 5106-5107, 5998, 6004, 6130, 6402, 7922, 7927-7928, 7941
 Cognitive Map 4570-4571, 4573-4574, 4579-4580
 Cognitive Model 832, 2370-2375, 2378-2379, 4201, 4491
 Cognitive Presence 2507, 2665, 3239, 3976-3977, 3979-3980, 3982-3983, 3985
 Cognitive Process 203, 392, 1442, 2076-2077, 2392, 2450, 3301, 4423, 4495, 5964, 7516, 7707, 7710
 Cognitive Radio 6536, 6540-6541, 6543, 6545
 cognitive schemata 3794-3795, 3798, 3804
 Cognitive Scheme 3804
 Cognitive Science 181, 662, 1353, 2433, 4267, 4449, 4672, 7599, 7608
 Cognitive Skills 999, 1433, 2254-2255, 2393, 2413, 2532-2534, 2542, 3242, 3799, 4362, 4787-4790, 4793, 4796, 5832, 5854, 6340, 6363, 6860, 6862, 6865, 6869, 7180, 7601, 7881, 7892
 Cognitive Vulnerability 1474, 5149-5154, 5156
 Cohesion 777, 783, 966, 1850, 1924, 2228, 2246, 2520, 2575, 3786, 3977-3978, 3980, 3983, 3985, 4321, 4622, 4853, 5577, 6184-6185, 6299, 6715, 7151, 7274, 7674
 cointegration 5639-5640, 5642, 5649
 Cold Storage 1950, 2790, 7872
 Coleman-Liau Index 1502, 1507
 Collaborative Filtering System 7804
 Collaborative Health Information Seeking 3750, 3755
 Collaborative Information Seeking System 3755
 collaborative innovations 5559, 5562, 5564-5565, 5569
 Collaborative Learning 1490, 1512, 1530, 2304, 2402, 2404, 2407, 2459, 2548, 2558, 2589-2590, 2598, 2641, 2643, 2647, 2650, 3250, 3255, 3259, 3749, 3851-3852, 3976-3977, 4359, 4363, 4822, 4825, 5029, 5097, 5102, 5210, 5789, 5830, 5833, 6336-6338, 6344, 6346, 6351, 6415, 7168, 7171-7175, 7180, 7384, 7645, 7882, 7912, 7973
 Collaborative Participatory Design (CPD) 6343, 6346
 Collaborative Problem Solving 1436, 2402-2407, 2410, 7881
 collaborative recommender systems 1759, 1872, 1878
 Collaborative Storytelling 2454
 Collaborative Technology 2591, 2947, 5713
 Collaborative Tools 2524, 7161
 Collective Choice 6714, 7282, 7284, 7286, 7289, 7291
 Collective I/O 1117-1118, 1120, 1123, 1125-1126, 1128
 Collective Intelligence 565, 572, 1528, 2281, 2286, 2288, 2485, 2489, 3571, 3578, 6337, 7038, 7042, 7109, 7111, 7114, 7146, 7162, 7170-7171, 7410, 7844, 8023
 Collectivism vs. Individualism 7957, 7965
 Collegial Learner 1479, 5151, 5154, 5156, 7923, 7925
 Colonial Lag 1207, 1215
 Color Deficiency 3525
 Colorado Student Assessment Program 5183
 Comb Compensation Filter 6056
 Comb Corrector Filters 6183
 Comb Filter 4746-4750, 4753-4760, 6007-6014, 6043-6044, 6047, 6052, 6054-6056, 6171, 6174-6175, 6177-6180, 6183
 Comb Folding Bands 4746, 4754, 4756, 6171, 6173, 6183
 comb order 6055, 6178, 6183
 Comb Passband 6056
 Combinatorics 4631, 4642
 Comfortableness 5149-5153, 5157, 7928
 Commercial Video Games 3261, 3271
 Commodification 4722, 6161-6162, 6166-6167, 6170, 6459, 7194
 Common Core Standards 665, 669
 Common Knowledge 76, 1450, 3228, 5027
 Common Random Numbers 1300, 1306
 Common Sense 100-102, 847, 951, 957, 1051, 1544, 3083, 3215, 4051, 4404, 4451, 4745, 5027, 7014-7015, 7022, 7025, 7628, 8006
 Common Sense Reasoning 100-102, 4745
 Communication Aid 289, 2523, 6474-6475
 Communication Privacy Management Theory 6985, 6989, 6992
 Communication Process 975, 2125, 2237, 2432, 2806, 2948, 4535, 6740, 7002, 7004, 7008
 Communication Technology 419, 550, 567, 787, 984, 1065, 1411, 1686, 1733, 1935, 2239, 2281, 2296, 2303-2305, 2313, 2411, 2421, 2423, 2457, 2559, 2761-2762, 2791, 2810, 3066, 3144, 3403, 3407, 3484, 3561, 3573, 3579, 3621, 3644, 4112, 4157, 4178, 4188, 4261, 4350, 4490, 4619, 4679, 4691, 4695, 4704, 4724, 4920, 5015, 5025, 5078, 5258, 5565, 5711, 5713, 5855, 5909, 5982, 6222, 6362, 6412, 6680, 6776, 6988, 7180, 7223, 7233, 7244, 7259, 7268-7269, 7369, 7720, 7912, 7982, 8097
 communication theory 1409
 Communication Topology 6536-6541, 6543, 6545
 Communicative Language Teaching (CLT) 7666, 7684, 7691
 Communities of Champions (CoC) 5080, 5084
 communities of inquiry 3976, 3978
 Communities Of Interest (Coi) 4282, 5080, 5084, 6490, 7089
 communities of learners 2256
 Communities Of Practice (Cop) 944, 947, 1219, 1221, 1883, 2594, 2598, 3737, 3795, 3797-3798, 3804,

Index

- 3939, 4178, 4180-4181, 4188, 4283, 4328, 4351, 5027-5029, 5032, 5035, 5079, 5084, 5093-5095, 5150, 5698-5699, 5704, 6337, 6352, 6413, 7057, 7063, 7672, 7674-7675, 7881, 7928
- Community Broadband Networks 3549, 3553, 3555, 3560
- Community Cloud 1028, 1033-1037, 1039, 1104, 6737
- Community Informatics 4282-4283, 6690, 6694, 7205-7206
- Community Of Learners 3239, 5150-5151, 5156
- Community Science 7201-7208, 7211
- Company Competitiveness 5366
- Competencies 281, 568, 695, 803, 805-806, 878, 893, 911, 946, 1147, 1420, 1422-1423, 1427, 1455, 1511, 1528-1529, 1566, 2226-2228, 2234, 2243-2246, 2254-2256, 2281, 2289, 2292, 2295-2297, 2299, 2302-2303, 2305, 2307-2308, 2311, 2314, 2330, 2351, 2364, 2537, 2556, 2643, 2650, 2802-2803, 2805-2806, 2809, 2923, 2930, 2935, 2950, 2969, 3001, 3010, 3240, 3587, 3593, 3692, 3829-3830, 3832, 3834, 3838, 3872, 3881, 4195, 4329, 4786, 4796, 5066, 5197, 5307, 5328-5332, 5562, 5708, 5716, 5719, 5733, 5857, 6413, 6687, 6689, 7154, 7252, 7262, 7314, 7540, 7548, 7623, 7661, 7672, 7680, 7990, 8043
- Competitive Advantage 49, 281, 406, 551, 555, 568, 602, 607-608, 708, 714, 730, 735, 739, 758, 760, 847, 875, 911-913, 918, 922, 926, 929, 943, 945, 1159, 1161, 1570, 1594, 1598, 1616, 1661, 2266, 2343, 2347, 2689, 2794, 2801, 2805, 2817, 2947, 2975, 2977, 3000-3001, 3025, 3031, 3077, 3756, 3827, 3969, 4282, 4327, 4348-4349, 4351, 4562, 4565, 4569, 4688, 4776, 5049, 5051, 5056, 5061, 5275, 5292-5293, 5306-5307, 5310, 5313, 5330, 5346, 5354, 5367, 5369, 5425, 5436, 5456, 5467, 5470, 5479, 5484, 5570, 5578, 5582, 5596, 5675-5676, 5691, 5714-5715, 5725, 5744, 5760-5761, 5929, 5985, 6416, 7133, 7146, 7369, 7776, 7808, 8043
- Competitive Intelligence 735, 944-947, 950, 962, 970, 6907
- Competitive Scanning 910
- Competitiveness 554, 629, 643, 760, 820, 913, 929, 952, 961, 1165, 1630, 2257, 2794, 2803, 2999, 3009, 3032, 3034, 3041, 3316, 3356, 3435, 3440, 4091, 4191, 4331, 4357, 4565, 4777-4778, 5024, 5285, 5346, 5360, 5362, 5364, 5366, 5423-5424, 5427, 5446, 5460, 5464, 5467, 5470, 5481, 5518, 5528-5529, 5644, 5719, 5762, 6927, 6935, 6995, 7223, 7226-7227, 7230, 7514, 7696, 7873
- Competitors 334, 349, 351, 556, 573, 611, 672, 695, 874, 899, 909-910, 922-923, 925-926, 938, 966, 1339-1340, 1376, 1775, 1976, 2177, 2347, 2489, 2687-2688, 2801, 3012, 3175, 3685, 4039, 4093, 4561, 4680, 4778-4779, 4794, 4829, 4839, 5061, 5358, 5360-5362, 5364, 5366, 5373, 5388-5389, 5392, 5394, 5481, 5490, 5493, 5528-5529, 5533, 5575, 5709, 5725, 5738, 5752, 5763, 5765, 5982, 5986, 7164, 7277, 7987
- Complete Cybernetic Connection 97
- Completeness 50, 384, 392, 758, 762, 1045, 1380, 1825, 1848, 1975, 2024, 2469, 2476, 2488, 3023, 3811, 4459, 4688, 4911, 5067, 5069, 5072, 5108, 5129, 5135, 6088, 7503, 7616
- Complex Exponential Signal 6017, 6234, 6244
- Complex Function 6017, 6234, 6244, 6623
- Complex Networks 498, 6508, 6521, 7321-7322
- complex number 6017-6018, 6244
- Complex Systems 220-221, 223, 225, 227-228, 795, 854, 925, 1398, 1581, 1611, 1685-1686, 2189, 2225, 2380, 2382, 2385, 3107, 4391, 4469, 5339, 5341, 5404, 5436, 5439-5440, 5538, 5583-5584, 5694, 6568, 6651, 6825, 6827, 6847, 6858, 7436
- Compliance 376, 613, 676, 760, 789, 805, 815, 820, 823-825, 827, 830, 887, 1028-1029, 2043, 2162-2163, 2380, 2487-2488, 2761, 2768-2770, 2772, 2833-2834, 2836, 2864, 3067, 3081, 3083, 3191, 3193, 3279, 3516-3517, 3519-3522, 3528, 3533, 3594, 3624, 4013, 4349, 4383, 4385-4386, 4438, 4816, 4852-4853, 4856, 4912, 5007-5008, 5146, 5158, 5227, 5231, 5296-5297, 5313, 5380, 5460, 5464, 5467-5468, 5478, 5552, 5646, 5985, 6095, 6480, 6570-6571, 6574-6575, 6578, 6815, 6930, 7144, 7329, 7483, 7485, 7505, 7906, 7957
- Composites 304, 2680-2681, 2897, 2900, 2902-2903, 2906, 7832
- Compositing 982
- Comprehensibility 183-185, 756-757, 760, 2094, 2134
- Comprehensive Evaluation 587, 5357, 6790
- Compressibility 1847, 1855
- Computational Aesthetics 156-158, 160-161, 165
- Computational Biology 419, 425, 4503
- Computational Complexity 195, 242, 247-248, 365, 1667, 1675, 1969, 1973, 2000, 2133, 3548, 4448, 4452-4453, 4998, 5880, 5974, 5976, 6010, 6014, 6017, 6236-6238, 6241, 6308, 6318, 6964, 7718
- Computational Environments 187, 2392, 2394
- Computational Intelligence 150, 156, 158-161, 165, 180-181, 183, 3107, 3109, 4951, 6907-6909, 7348, 7704-7707, 7711-7712, 7774
- Computational Techniques 2, 165, 180, 184, 455, 1249, 1259, 3107, 3483

- Computational Thinking 2287, 2296-2297, 2299, 2302, 2392-2394, 2397-2399, 2401, 2492, 2497, 3311, 3316, 6859
- Computationally Lightweight Metric 6507-6508, 6520
- Computationally Universal System 7743
- Computer Access 267-268, 270-271, 276, 2526, 6391
- computer agent 2402-2405, 2407, 2410
- Computer Animation 2335, 2599, 2602, 2605, 2615
- Computer Approach 7384, 7386-7387, 7391
- Computer Assisted Diagnosis (CAD) 1017, 1024, 4144
- computer competencies 2303, 2305
- Computer Ethics 4886-4887, 4889-4890, 4894-4896
- Computer Ethics Society 4887, 4889, 4896
- Computer Forensics 1384-1386, 2918
- Computer Forensics Science 2918
- Computer Information Library Clusters (CILC) 4399-4403
- Computer Instruction 2421, 7029
- Computer Mediated Communication 4282, 6490, 7076
- Computer Platform 464, 2342
- Computer Reservation Systems (CRS) 363, 3417
- Computer Supported Collaborative Learning 5102
- Computer Supported Ubiquitous Learning (CSUL) Environment 2546, 2556, 7672, 7680
- Computer Vision (CV) 99, 143, 166, 1344, 1352, 1354, 4200, 5880, 6195, 6201-6203, 6205, 6894
- Computer-Aided Design (CAD) 506, 1772, 4388, 7382, 7391
- Computer-Assisted Instruction 2303, 2590, 2598
- Computer-Based Conferencing 2590, 2598
- Computer-based information retrieval 4538
- Computerized Maintenance Management System (CMMS) 3698-3699, 3709
- computerized testing 2369-2370, 2376
- Computer-Mediated Communication (CMC) 1220, 1223, 1454, 2121, 2123, 2131, 2457, 4283, 4294, 6284, 7071, 7172, 7182
- Computer-Supported Collaborative Learning (CSCL) 5830, 6337, 6346
- Concentration of Media Ownership 7270, 7272, 7279
- Concept Algebra 7704, 7706-7707, 7711-7712
- concept discovery 1974-1979, 1981, 1983
- Concept Map 2448, 4572, 5114
- Concept of Boundedness 220-222, 227-228
- Conceptual Framework Of Understanding 1475, 1477-1478, 1481-1482, 5149-5153, 5156
- Conceptual Modeling 1376, 1380, 4573, 7859
- Conceptual Models 378, 3760, 4810, 8027
- Conceptualization 388, 554, 651, 654-655, 828, 961, 963, 967, 1135, 1141, 1148, 1477, 1642, 2227-2228, 2230, 2862, 2915, 3289, 3292, 3786, 3818-3819, 3882, 3887, 4169, 4172-4173, 4279, 4328, 4541, 5466, 5676, 5777, 5793, 6478, 6480-6481, 6487, 7073, 7091, 7193
- concurrent programming 7742
- Confederate Effect 257, 263
- Confidants 6986, 6992
- Confidentiality 374, 376, 745, 1037, 1087, 1107, 1247, 1359, 1676, 1683, 2488, 2717, 3678, 3684, 3806, 4340, 4362, 4887, 4927, 4930-4931, 4987, 4994, 5860, 5903, 6087, 6155, 6499, 6953, 6986, 6992, 7186, 7316, 7528, 7590-7592
- Confidentiality Breaches 6986, 6992
- Configurable 1026, 1067, 1102, 1114, 1144, 1153, 1235, 1627, 1990, 3464, 4022-4023, 5353, 7771, 7778
- Confluence 1509, 1516, 4139, 4144, 4257, 5311
- Conformance 732, 2686, 3520, 4608, 5944-5945, 5948, 6570-6571, 6574-6575, 6578, 7563
- conjunctive queries 7996, 8000
- Connected E-Government 3602-3608, 3611
- Connected Passenger 4053, 4055-4056, 4058-4059, 4063
- Connected Presence 4620, 6194
- Connectionism 232, 238
- Connotation 4069, 4325, 5128, 6350, 6705, 7340, 7346, 7413
- Consistency Matrix 2435, 2441
- Constitutive Entanglement 706
- constrained nonlinear optimization 4594
- Construction Industry 389, 539-542, 544-546, 550, 983-985, 989-990, 5050, 5491
- Constructionism 2392, 2395, 3250, 3259, 3310, 6340, 6380, 6861, 7313, 7315
- Constructive Cost Model (COCOMO) 1405, 4852
- COConstructive COst Model II (COCOMO II) 1405
- Constructive Engagement 287, 291, 297
- constructive logical database design 2024, 2026, 2028
- Constructive Recreation 4316, 4320-4321, 4325
- Constructivism 811, 1426, 2588, 2598, 2620, 2624, 3250, 3259, 5832, 6340, 6362, 6758, 6861, 7577
- Constructivist Learning Model 2642, 2650
- constructivist teaching 5854
- Constructivist Theory 2401, 3250, 5832
- Consultative Leadership 5195, 5203
- Consumer Behavior 334, 962, 1597, 1638, 2346, 2817, 2844, 4088, 4090, 4092, 4717, 5750, 6071, 6077, 7143, 7145
- Consumer Intention 6020, 6022, 6024, 6026, 6030
- Consumer Price Index (CPI) 47
- Consumer Referral Behavior 6030
- Consumer Sovereignty 1641, 1644

Index

- Consumerism 908, 1638, 1640-1642, 5531
Consumerization of IT (CoIT) 5986, 5991
Consumer-To-Consumer (C2C) 638, 2741, 2747, 2833-2834, 2836-2838, 4057
Consumption System 5360, 5362
Contact area 434-435, 439
Contactless Technologies 7805, 7807-7810, 7813
Container Corporation of India 588, 600
Contemporaneity of Communications 7108, 7114
Contemporary Architectures 7867
Contemporary Governance 6695, 6699, 6705-6706
Content Analysis 279, 1715, 1896, 1900, 1903, 3243, 3396, 3976-3977, 4092, 5111, 5162, 5359, 5447, 5451-5452, 5455, 5738, 5740, 5744, 6164, 6994, 7014, 7018, 7358, 7958, 7961-7962, 8069-8070
content authentication 2908-2910, 2913, 2915
Content Centric Networking (CCN) 6565
Content Diversity 5234, 7271-7272, 7274, 7276, 7279
content tampering 2908, 2915
Content-Based Recommender Systems 1872, 1878
context recognition 6032, 6035, 6037, 6042, 6287
Context-Aware Personalized Collaborative Recommendation 6042
Context-Aware Personalized Content-based Collaborative Recommendation 6042
Context-Aware Personalized Content-based Recommendation 6042
context-aware personalized recommendation 6037-6038, 6042
Context-Aware Recommender System 1738, 1745, 1770
Context-Awareness 4216, 6032, 6402, 6411, 6480, 6489, 7771-7772, 7807
Contextual Links 6428, 6433
contextual user profile 6035-6037, 6039, 6042
contextual user profiling 6032, 6035-6036, 6039
contingency planning 5507-5508
Contingency Theories of Leadership 5195, 5203
Continual Improvement 3930, 3933, 3939-3940, 3944, 4092
Continuance Intention 833-835, 838, 846
Continuation 1040, 1606, 3682, 4191-4192, 4484, 4489, 4719, 5493, 5626, 5635, 6174, 6180, 6348, 6472, 7308, 7639
Continuing Education 2254, 2604, 3692, 3832, 5811-5816, 5819-5825, 6148, 6160, 6415, 6690, 7078, 7912, 7916
Continuous Assurance 820-825, 827-828, 830
Continuous Auditing 821-823, 830
Continuous Improvement 690, 714, 729, 731, 767, 1934-1935, 2169, 3131, 3137-3138, 3528, 3623, 3707, 3939, 3944, 5299, 5358, 5360, 5518-5519, 5717, 7518, 7550
Continuous Integration 1405, 2592, 7506, 7513
Continuous Monitoring 821-823, 828, 830, 1007, 1936, 1938, 3158, 3440, 3691, 3975, 4164, 4864, 4869, 6820, 7186
Continuous Review System 5344
contrast ratio 6070, 6072-6073, 6076, 6082
Conventional Level 3197, 3203
conventional raster 171-172, 177
Conventional Raster-Stereography 172, 177, 179
Convergence 4, 6, 15, 24, 29, 231, 241, 663, 756, 963, 980, 1146, 1206, 1208-1209, 1212, 1215, 1575, 1577, 1766, 1928, 2134, 2180, 2193, 2293, 2297, 2345, 2455, 2459, 2567, 2588, 2591, 2756, 3105, 3108, 3506, 3551, 3560, 3638, 3798, 3831-3832, 3871, 4048, 4113, 4210, 4215-4216, 4255-4257, 4262, 4397, 4606, 4688, 4709, 5379, 5413-5418, 5421, 5571, 5628, 5669, 5683, 5737, 5771, 5870, 6097-6098, 6355, 6580, 6596, 6599-6602, 6757, 6784, 7014, 7098, 7234, 7253, 7261, 7284, 7287, 7313, 7440, 7695, 7698, 7760, 7856, 7861, 8072
Conversion 349, 489, 523, 742, 944, 1039, 1287, 1331, 1339, 1376, 2052, 2063, 2201, 2266, 2664, 2680, 2854, 2897, 2900-2901, 3170, 3525, 3713, 3715, 3795, 3869, 4282, 4575-4576, 4612, 4630, 4719, 5094, 5305, 5361, 5583, 5613, 6014, 6619, 6814, 6850, 7161, 7602
convolution encoder 6307-6308
convolutional codes 6308, 6318
Convolutional Encoder 6307, 6318
Convolutional Neural Networks 4529-4530, 4534
Co-Occurrence Matrix 4645, 4651
Cooling Degree Days (CDD) 7815, 7818-7819, 7828
cooperating participants 2095, 2098, 2107
Cooperative Behaviors 6836
Cooperative Relaying 6499-6500, 6505-6506
coordination improvement 2954, 2956, 2960, 2962-2963
Coordination Mechanism 7282, 7291
COPIS 730, 732
Copreneurs 3002, 3007
Co-Presence 6184, 6194
Copyright 565, 898, 1657-1659, 1661, 2239, 2477-2491, 3834, 4310, 4383-4384, 4386, 4667-4668, 4671, 4691, 4872, 4883, 4885, 4988, 5207, 5210, 5217, 5226-5228, 5231, 5234, 5236, 6248, 6749, 6755, 7108, 7236, 7252, 7369, 7902, 7906-7907, 7985, 7988
Core Team 3683, 5713
core vectors 1684, 1687, 1693

- Co-Regulation 3934-3936, 3939, 4964
- corporate blogs 5559, 5562-5565, 5569
- Corporate disclosure (CD) 1896, 1900-1903, 1906
- Corporate Identity 695, 2581, 8084, 8086
- Corporate Organizations 3396, 3401
- Corporate Reputation 671-672, 680, 7297, 7565
- Corporate Social Responsibility (CSR) 671-673, 680, 754, 3177, 3183, 3193, 4912, 4969, 5286, 5292-5293, 5305, 5426, 5718-5719, 5737-5738, 5740, 5742-5743, 5747, 6276
- Correction Step 219
- Correctness 93-94, 1668, 2000, 2027, 2057, 2087, 2393, 3022, 3106, 4852, 4931, 5879, 5935, 5942-5943, 5947, 5949-5950, 6339, 6541, 7436, 7452, 7461, 7503, 7505, 7508, 7588, 7610, 7616-7617, 7741, 8007-8008, 8010, 8013
- corrector filters 6174, 6178, 6180-6181, 6183
- Correlation Analysis 734, 933, 2610, 5062, 7837
- Correlation Coefficient 593, 1009, 1750-1753, 1756, 1828, 4880, 6508, 6513-6514, 6516-6517, 7837
- Correlation Graph 444, 446-449
- Correspondence Learning 2465
- Cosine Filter 4751, 4753-4755, 4759
- Cost Allocation 1170, 2095-2096, 2098, 2101, 2103, 2105-2107, 2170
- Cost Estimation 4849-4853, 4857-4858, 4860, 5698
- cost matrix 1834
- Cost of Caring 3747-3748, 3755
- Cost Sharing 1164-1165, 1168-1170, 1172, 2098, 2107
- Coulson 3767-3770, 3775, 7079
- Coupling 224, 664, 1056, 1400, 1405, 2412, 2417, 2419, 4139, 4234, 5279, 5318-5320, 5326, 5411, 6328, 6332, 6571, 6574-6575, 7015, 7412, 7572, 7604, 7768
- courier service 2802-2807, 2809
- Course Buddies 5824, 5828
- Course Management System 808, 2666, 5812, 5822, 5840
- Covariance Measure 219
- Covariation 2966, 5187, 5192
- cover media 4985, 5003
- COX proportional hazard ratio 1802
- CPU 326, 1069, 1077, 1080, 1082, 1178-1179, 1230, 1629, 1651, 1806-1807, 1810-1811, 1947-1948, 1951-1952, 1954, 1956, 1961, 2872, 3275, 3693, 3991-3992, 3994, 3996-3997, 4005, 4009, 4011, 4016, 4019, 4025, 4453, 5415, 5417-5418, 6669, 6907, 7575, 7748
- CPU Utilization 7575
- Cracking 6696-6699, 6704-6706, 8066
- crash risk 5992, 5994, 5997, 5999, 6005
- Creative Commons 2481, 2485-2488, 2490-2491, 6248, 7906
- Creative Industries for Cultural Heritage 5225
- Credit Card Fraud 2713-2714, 4838, 4937, 4948, 6705
- Credit Card Processing 2710, 2712-2713, 4931
- Credit Card Processor 2710, 2722
- Credit Risk 1, 4, 8, 15, 37-38, 373, 2219-2220, 7141, 7143-7144
- Credit Score 2219, 7140-7145, 7149
- Crisis Informatics 6687-6688, 6690, 6694
- Crisis Management 1398, 1407-1410, 1414, 2006-2008, 2012, 2014-2016, 2019
- Crisis Response 1396, 1407-1409, 1412, 1414, 2012, 2016
- crisis response ontologies 2012
- Crisis/Disaster Response 1414-1415
- critical constant 1863-1866, 1868
- Critical Section 3990-3992, 3994-3995, 3997-3998, 4000, 4003
- Critical Success Factors (Csfs) 579, 620-621, 627-628, 631, 635-636, 644-645, 1935, 2994, 2996, 3435, 3561, 3564, 3765, 4284, 5050, 5309, 5313, 5460, 5619, 5691, 5695, 5903, 7268
- Critical Thinking 1434, 1469, 1530, 2228, 2244-2245, 2293, 2445, 2447, 2467, 2492-2496, 2502, 2526, 2574, 2579, 2588, 2591, 2594, 2616, 2643, 2664, 2994, 3240, 3978, 4359-4360, 4363, 4367, 4637, 4699, 5832, 5855, 6362, 6399, 6859, 6870, 7090, 7175, 7180, 7251, 7254, 7543-7545
- Critical Thinking Skills 1469, 2293, 2467, 2496, 2574, 2579, 4359, 4363, 6362, 6399
- Critical to Quality (CTQ) Characteristics 739
- Cross Layer Modelling 6618, 6622-6623, 6629
- Cross-Domain Recommender System 1740, 1745
- cross-functional drivers 5350, 5354
- Cross-Language Information Retrieval 4471, 4490-4491, 4495, 4497-4498, 4501
- Cross-Language QAS 4471, 4473, 4479
- crowd creation 8023
- Crowd Wisdom 565, 572, 7844, 8023
- Crowdfunding 3286, 7078, 7080, 7082, 8023
- Crowding Distance 242-243, 247, 249
- Crowdsourcing 159, 563-568, 572, 707-708, 710-712, 715, 719, 2007, 2019, 2067, 2477, 2486, 2491, 2910, 3464, 3471, 3750, 4092, 6187, 6194, 6290, 6974, 7005, 7036, 7038, 7162, 7843-7846, 7854, 8023-8031, 8034-8035
- Crowdsourcing Applications 564, 8023, 8029
- Crowdsourcing Platform 8023, 8034

Index

- Crowdsomers 1644
crypto-currency 2790, 2815, 2862-2866, 2868-2870, 2876, 2878-2880
Cryptography 375, 1243-1244, 1247, 1297, 1304, 1650-1651, 1667-1668, 1675, 1849, 2862, 2870, 2872, 2876, 2880, 4448, 4451-4453, 4898-4901, 4903, 4906-4908, 4927-4928, 4930-4931, 4936, 4975-4977, 4984-4987, 4996, 5003, 6560, 7722, 7724
CSA 374-375, 380, 5205
CSR tools 3187, 3191
CSS 1228, 1233, 2660, 6071, 6074, 7748, 7750, 8010, 8079, 8081, 8087-8089, 8091-8094
C-system 1996-1997, 1999-2000, 2002-2003, 2005
CT 507, 523, 1255, 2296, 2392-2394, 2396-2398, 3311-3312, 5880, 5892-5893, 5899-5900, 6662
Cuckoo Search Algorithm 2201, 2205, 2208, 2211, 4369-4374, 4376-4377, 4381
Culmination 235, 1616, 4484, 4489, 4635, 8037
Cultural Capital 286, 395, 3924, 7061
Cultural Communication 5330, 7002-7004, 7008, 7013
Cultural Dimensions 5637, 7545, 7956-7957, 8104
Cultural Heritage 2442, 4082, 4087, 4189-4191, 4196, 4438, 4717, 5106, 5214-5215, 5217-5219, 5221-5225, 7002, 7004-7005
Cultural Heritage Tourism 4087
Cultural Intelligence (CQ) 286
Cultural Organization 2292, 7002, 7005, 7007-7008, 7013
Cultural Patterning 4139, 4144
Cultural Plans 7332
Cultural Structure 4722
Cultural Studies 2513, 4189-4190, 4404, 6306
Cultural Systems 3926, 3932-3934, 3938-3939, 3944
Culture Industry 4716-4719, 4721-4722
Current Density 2675, 2678, 2902, 2906, 6322-6323, 6326-6328
Current ERP Performance 2955, 2960, 2962-2963
Customer Awareness 5380
Customer Capital 715, 5057, 5066, 5671
Customer Engagement 712, 1587, 3359, 3862, 4054, 5775-5778, 5780-5783, 5785, 6088, 7126, 7131, 7134, 7139, 8038, 8040
Customer Equity 713, 1584-1585, 1593
Customer Feedback 356-357, 360, 604, 4088, 4090-4091, 4094, 4098, 4199, 5780, 7126, 7131
Customer Lifetime Value 1584-1586, 1593, 1595, 2860
Customer Loyalty 672, 712-713, 1584-1585, 1593-1594, 1596-1597, 1604-1605, 2817, 3360, 4217, 5361-5362, 5364, 5423, 5532, 5719, 5983, 7131
Customer Relationship Management (CRM) 350, 923, 929, 945, 1405, 1584, 1594-1596, 1604-1605, 1615, 1616, 1619, 1626, 1880, 2086, 2169, 2854, 2860, 2953, 2977, 4054, 4909, 5353, 5444, 5559-5562, 5569, 5775, 7131
Customer Relationship Management Solution 5569
Customer Retention 412, 414, 834, 929, 1585-1586, 1593-1595, 1604, 1606, 2883, 5293
Customer Satisfaction 49, 51-53, 55, 765, 833-834, 914, 1130, 1570, 1587, 1593-1594, 1596-1597, 1604, 2742, 2803, 2978, 3089, 3092, 3527, 4092, 5346, 5354, 5363, 5369, 5380, 5387, 5390, 5423, 5433, 5490, 5516, 5532, 5748, 5921, 6928, 7143, 7551
Customer Service Life Cycle 932, 938, 941, 2688, 2696
Customer-Based Data 349, 363
customer-driven innovation 4092, 4094
CVV 77, 85
Cyber Bully 4172, 4247, 4253
Cyber Cascade 7947, 7955
Cyber Physical System 1227, 1236
Cyber Politics 3538, 3541, 3548
Cyber Predator 1721
Cyber Security 1647, 1685, 1705-1706, 1715-1716, 2693-2694, 2866, 2878, 3530, 4857, 5308, 5312, 7525, 7528, 7534, 7987
cyber supply chain 5306-5308, 5310, 5313, 5316
cyber threat 1706
Cyber Victim 1715, 1721
Cyber Victimization 1697, 1725-1728, 1733, 4245-4249
Cyber Violence 4157-4158
Cyberactivism 3569
Cyberattack 1684-1685, 1687, 1693, 1697, 4838, 4848, 5991
Cyberbludging 4316, 4325
Cyberbullying 1366, 1695-1700, 1703-1716, 1721, 1723-1724, 1726-1729, 1733, 2423, 4157-4164, 4168-4173, 4177, 4245-4249, 4254, 4964, 4974, 6391, 7173, 7175
Cybercrime 1684, 2714, 4837-4839, 4841, 4843-4844, 4848, 5307, 5311, 5313, 5316, 5991
Cyber-Criminal 2722
Cyberdeviance 5988, 5991
Cyberfeminism 3350, 3358
cyberinfrastructure 1063-1071, 1074
Cybersecurity 1684-1687, 1690-1691, 1693, 3530, 4711, 4837-4838, 4841, 4843, 4888, 5311, 5313, 5316, 7988
Cybersex 1370, 4224-4225, 4233
Cyberslacking 4317, 4325
Cyberspace 252, 258-260, 263, 551, 565, 1453, 1657, 1660-1662, 1684-1686, 1691, 1693, 1704-1705, 1708, 1721, 1728, 2326-2327, 2330, 2335, 2512,

2593, 2826-2827, 2830, 2870, 3349-3350, 3352-3355, 3358, 3562, 3572, 3642, 3669, 4139, 4146, 4157-4158, 4167, 4198, 4234, 4296, 4623, 4887, 4909, 4919, 6481, 7028, 7088, 7106, 7249, 7769, 8067

Cyberspace Relative Anonymity 4158, 4167

Cyberspace Threats 4157

Cyberstalking 1366, 1695-1696, 1702, 1709

Cyclic Redundancy Check 1384, 1394

Cycling 2900, 3371, 3623, 7348-7349

Cynicism Theory 3820, 3825

Cypher Query Language (CQL) 2032, 2035, 2042

D

Dante Work's Source 5072, 5076

Dante work's structure 5067, 5069, 5071-5072, 5076

Dark Cultural Heritage 4082, 4087

Dark Data 4310, 4313

Dark Ride 4692, 4702

Dark Tourism 4077-4084, 4087, 4767-4768

Data Acquisition 133, 507, 862, 1359, 1383, 1399, 1406, 1814, 2059, 3065, 3113, 3277, 3491-3492, 3498, 5880, 6122, 6802, 6974-6976, 7848

Data Analysis 4, 39, 187, 338-339, 345, 350, 355, 375, 378, 410, 420, 440, 453, 472, 542, 593, 781, 837, 851, 933, 1129, 1131, 1133, 1197, 1205, 1249, 1333, 1378-1379, 1383, 1532, 1608, 1712, 1794-1795, 1800-1802, 1804, 1814, 1824, 1836, 1866, 1889, 1909, 1984, 1987-1989, 1991, 2199-2200, 2286, 2306, 2609, 2886, 2959, 2990, 3089, 3122, 3340, 3426, 3450, 3462, 3479, 3482, 3569, 3583, 3659, 3673, 3882-3883, 4040, 4066, 4262, 4340, 4399, 4402-4403, 4487, 4829, 5121, 5139, 5158, 5163-5164, 5168, 5184, 5245, 5369, 5407, 5522, 5549, 5620, 5950, 6490, 6907-6909, 7018, 7317, 7650, 7665, 7760

Data Analytics 182, 329, 333-335, 349, 373, 375, 387, 407, 862-864, 866, 870-872, 952, 955-956, 1774-1775, 1806, 1813, 1835, 1837, 1839, 1880-1883, 1889, 1891-1892, 1894, 1919, 1934, 2687, 2692, 2810, 2813, 2817, 2822, 2941, 3088, 3711, 3713, 3721, 3723, 4012-4013, 4089, 4217, 4844, 4951, 4955, 5296, 5375, 5889, 6038, 7141-7142, 7273, 7396, 7527, 8036

data building blocks 5607, 5614

Data Cleaning/Scrubbing 1198, 1205, 1757, 1765-1766

Data Envelopment Analysis 587, 589-590, 600

Data Flow 204, 208, 789, 795, 798, 863, 901, 1250, 1398, 4023, 5889, 7508-7510, 7736

Data Formats 354, 420, 459-460, 1109, 1622, 2048, 3461, 3476, 4442, 5076, 6565-6566, 7119

Data Governance 954-957, 959, 5612

Data Integration 205, 353, 420, 455, 461, 847, 849, 851, 854, 862-864, 933, 959, 1379, 1621, 1624, 1773, 1921, 2776-2777, 3450, 5538, 5541, 5613, 6759

Data Intensive Computing 336, 1950, 4012

data journalism 1196-1198, 1200-1202, 1205

Data Management 326-327, 335, 374-376, 407, 420, 459, 868, 872, 875, 887, 954-957, 959, 1091, 1772, 1776, 1806-1807, 1810, 1951, 1989, 1991, 2020, 2059, 2346, 2483, 3106, 3521, 3689, 3697, 3898, 4307-4308, 4439, 4441-4442, 4455, 4733, 4884, 4896, 5223, 5229, 5609, 5617, 5619, 5870, 5903, 7995

Data Mart 941, 3885

Data Mining 101, 108, 133, 181-182, 184, 333, 339-340, 342, 344-345, 348, 350, 355, 358, 363-366, 374-375, 406-407, 409-414, 417, 422, 824, 847, 851, 854, 859, 872, 879, 887, 933, 935, 938, 941, 956, 1360, 1376, 1378, 1383, 1400, 1447, 1595, 1617-1618, 1620, 1622-1623, 1788, 1795, 1825, 1831, 1835, 1837-1838, 1842, 1844, 1875, 1897-1898, 1907, 1909-1911, 1918-1924, 1930, 1932, 1934-1935, 1943, 1945, 1962, 1974-1975, 1985, 1988-1989, 1992, 2064, 2066, 2071, 2074, 2089, 2094, 2132, 2137, 2169, 2200-2201, 2779, 2794, 2933, 2941, 3275, 3479, 3543, 3571, 4349, 4420, 4440, 4503, 4513, 4829, 4866, 5055, 5105-5108, 5111, 5159-5160, 5168-5170, 5177, 5182, 5407, 5539, 5541-5542, 5544, 5549, 6430-6431, 6478, 6634, 6739, 6908, 7264, 7348-7353, 7355, 7411, 7527, 7696, 7776

Data Mining Algorithms 101, 339, 345, 410-411, 1795, 1918, 1921-1922, 1932, 1975

Data Mining Language 1918

Data Mining Models 1911, 1918, 2089, 5544

data mining task 1918, 1922

Data Mining Tools 406, 410, 414, 1620, 1622, 2933, 5407

Data Modelling 933, 1835, 1838, 5607, 5613, 5615, 5617, 5625

Data Models 328, 420, 864, 874, 1775-1776, 1956, 2020, 2023, 2027, 3461, 3496, 4811, 5571, 5607, 5612-5614, 5617-5618, 5625

Data Privacy 374-375, 1069, 1360, 4439, 4849, 4854, 4862, 4866, 4887, 4909-4913, 4916, 4918-4919, 5164, 6084, 6089

Data Privacy Protection Problem 4910, 4919

Data Protection 373, 642, 868, 878, 1157, 1163, 2717, 4013, 4267-4268, 4711, 4838, 4841, 4863-4864, 4866, 4909, 4911, 4918, 5986, 6089, 6151, 6727, 6951, 7762

Data Protection Principles (DPP) 4918

Index

- Data Quality 699, 849, 955, 959, 1616-1617, 1619-1621, 1623-1626, 1776, 3497, 4483, 5612, 6645, 6910
- Data Science 8, 1746, 2086, 2091, 2773, 6908-6909, 7397
- Data Scraping 1205
- Data Selection 3, 798, 1919, 1921
- Data Sieving 1117-1120, 1123, 1126, 1128
- Data Snooping 6640
- data steward 956, 959
- Data Stream 208, 364-366, 369-370, 372, 406, 409, 411-412, 417, 1386, 1394, 2047, 2086, 6101
- Data Structure 862-863, 1242, 1333, 1342, 1950-1951, 2145, 2148, 4585, 4588, 6062, 6452, 6828, 6910, 7479, 7611, 8005
- Data Validation 1198, 1205
- data virtualization 868-869, 872
- Data Visualization 354, 388, 404, 422, 933, 956, 1198, 1200, 1205, 1249, 3419, 3450, 3794, 5105, 5111, 5114, 5162, 6189, 6941, 7777
- Data Warehouse (DW) 350-355, 357, 360, 363, 854, 860-861, 863-864, 867, 870, 933, 941, 1376-1378, 1773-1775, 1921, 2008, 2955, 3883, 3885-3886, 3898, 4437, 4829, 5541, 5543-5545, 5549, 6908, 6914
- Data Warehouse Bus Matrix 363
- Database Benchmark 1782
- database cluster 1094-1096, 1099
- Database Conversion/Migration 2052, 2063
- Database Management Systems (DBMS) 326, 337, 1091-1094, 1096, 1636, 1773, 1782, 1808, 1912, 1955, 2054, 2977, 4528, 7843, 7854
- Database Recovery 1099
- database replication 1091-1093, 1096, 1099
- Database Schema 657, 1773, 1782, 2043-2046, 2052-2056, 2063
- Datacenter 377, 1027-1028, 1078, 1095-1096, 1120, 1142, 1144-1145, 1163, 1774, 3080, 3085, 3122, 3464, 3530, 4012, 4015, 4442, 5483, 5520-5521, 5985, 6615, 6711, 7747-7748, 7842
- Datalog 2032, 2035-2036, 2039-2042
- Dataset 192, 201, 215-216, 233, 365-366, 389, 408-410, 412, 414, 448, 499, 924-925, 1012, 1205, 1622, 1757, 1761-1763, 1765-1768, 1770, 1772-1773, 1776, 1798-1800, 1885-1888, 1948, 1964, 1968, 2067, 2133, 2137, 2141, 2143, 2146-2147, 2149, 2201, 2835, 3410, 3464, 3615, 4633, 4787, 4942, 4947, 4950, 5451, 5843, 5847, 5996, 5999, 6033, 6646-6647, 6649, 6657, 6663, 6665-6668, 6759, 6765, 6977, 7111, 7350-7351, 7353
- Davis 52, 114, 157, 531-533, 554, 835, 840, 1562, 1640, 1831, 2303, 2346, 2383, 2424, 2468, 2762, 2791, 2998, 3067, 3178, 3382, 3489, 3503, 3571, 3581-3582, 3613, 3616, 3655, 3712, 3786-3787, 3821, 4223, 4225, 4281, 4742, 4785, 5263, 5273, 5552, 5835, 5837, 6022, 6025, 6715, 7302, 7422, 7426, 7874, 7907, 8067
- DDR 4009-4010, 4016-4017, 7463
- DDR4 4009, 4017
- De Rosa 4038-4039, 4047-4048, 4050-4051, 4064, 4066, 4072-4073, 4076, 4404-4407, 4409-4414, 4416, 4420, 5237-5240, 5242, 5245-5246, 5251, 7014-7017, 7021-7022, 7046-7049, 7051-7052, 7056
- Dead Mutant 7469
- dead reckoning 6287, 6289-6290, 6295
- Deadlock 2704, 4550, 4553, 6713, 7435, 7439, 7445
- debit card fraud 2711, 2713-2714, 2716, 2718
- debit card processing 2707-2708, 2710
- Debt Crisis 2342, 6918
- Decentralized architecture 6950, 6953
- decimal arithmetic 4607-4608, 4614-4615
- Decimal Fixed-Point Multiplication 4609, 4618
- decimal floating-point 4608, 4612-4614, 4618
- Decimal Floating-Point Multiplication 4608, 4612, 4618
- decimal multiplication 4607-4608, 4610, 4612-4615, 4618
- Decimation 4746, 4749, 4752, 4754-4757, 6014, 6043-6044, 6046, 6054-6056, 6171, 6173, 6177-6180, 6183
- Decimator 4749
- Decision Field Theory (DFT) 2108, 2111, 2116, 2120
- decision fusion 4937
- Decision Rules 1839-1840, 1842, 2085-2086, 2204, 2208, 2211, 2213, 2218, 6710, 6715, 6719
- Decision Support 102, 184, 352, 355, 363, 545, 621, 627-628, 637, 642-643, 734, 847, 859, 930, 933, 962-963, 1875, 2012, 2108, 2112-2115, 2117, 2121-2122, 2131, 2168, 2177-2178, 2181, 2200-2202, 2208, 2211, 2932, 2934-2935, 2937-2938, 2977, 3460, 4573-4574, 5400, 5503-5504, 5509, 5511, 5515, 5561, 5864, 5870, 6025, 6634
- Decision Support Systems 102, 184, 643, 933, 1875, 2108, 2115, 2121-2122, 2131, 2177, 2201, 2932, 2934, 4573, 5515, 5561, 5864, 6025
- Decision Tree 4, 14, 105, 137, 366, 407, 409, 412-414, 643, 933, 1011, 1446, 1795, 1827-1829, 1928-1929, 2132-2134, 2137, 2141, 3229, 6633, 6829-6833

- Decision Tree Pruning 2133, 2141
- Decision Variables 239-240, 2080, 2145, 2148, 2178, 2182, 2187, 4598, 4606, 4915, 5411-5413, 5416-5417, 5421, 6829
- Decoherence Time 7716, 7719, 7730
- decomposition technique 4877, 5411, 5420
- deductive databases 1995, 2042
- Deep Architectures 4528, 4530, 4533-4534
- Deep Belief Network 4531-4534
- Deep Blue 98, 100, 145, 230, 234-235, 238
- Deep Learning (DL) 126, 859, 1354, 1530, 1538, 2443, 2445, 3249, 3311, 4008, 4013, 4528, 4531, 4533-4534, 5172, 6120, 6635, 6640, 6651
- Deep Neural Networks 4529-4530, 4534
- Defuzzification 109-110, 3215-3216, 3223-3224, 6798, 6800
- defuzzification technique 3215, 3223-3224
- Deindividuation 3769, 3781, 7973
- De-Industrialization 7234, 7247
- Delay Jitter 6445
- Deliberate Confidants 6986, 6992
- Delphi 589, 662-664, 669, 825, 910, 967, 2806, 2809, 5197, 6189
- Delphi Method 825, 2806, 2809, 6189
- DEM 1134, 3422, 3485, 3496, 3504-3506, 3514
- Demand and Supply Risk 5555, 5558
- Demand Response (DR) 3104, 3113, 3144, 3155
- Demand-Driven View of Knowledge 5054
- Dematerialization 888-891, 893-896, 898, 3188
- Demonstration Video 5862
- Dendrite 6594
- Dendritic Algorithm 107, 119, 231
- Dendrochronology 3116-3117, 3119, 3121, 3126-3127, 3130
- Denotation 2190, 4461, 5128, 6350, 7340, 7344, 7346, 7413, 7610, 7711
- Denotational Mathematics 7704, 7706, 7708-7712, 7771-7772, 7774
- Densely Packed Decimal 4612, 4618
- dependable system 1655
- Deployment Models 886, 1028, 1032, 1104-1105, 1356, 1359, 1990, 2983, 6728
- Depth-First Search 1983
- Descriptive Analytics 1809, 1813
- Descriptive Approaches 2077, 3177, 5434
- Descriptive Gender Stereotypes 3383, 3391
- Descriptive Metadata 5216, 6447-6448, 6450, 6452, 6455, 6458, 6463
- Descriptive Unit 5225
- Design and Applied Technology 7385, 7391
- design element hierarchy 6343-6344
- Design Model 1562, 2857, 3290, 3340, 4525, 4821
- Design Pattern 1406, 1444, 1446, 1450, 2156, 3290, 3411, 3413, 8089
- Design Requirements 2634, 6240, 6307, 6316, 6382, 7755
- Design Rule 5321-5322, 5326, 8006
- Design Suggestions 1441, 1444, 1446
- Desk Research 906-907
- Desktop Computers 1110, 2427, 2636, 2652, 2745, 2757, 3857, 4210, 4214, 5575, 5981, 6059, 6061, 6360, 7119-7120, 7124, 8089
- Desktop VR 6233
- DESTEP 907
- Destination Management Information System (DMIS) 354-355, 360, 363
- Destination Management Organization (DMO) 351, 363
- Destination@-Branding 4038, 4048, 4051
- Destination-Branding 4051
- deterministic algorithm 1238, 1240-1242, 1297, 1306
- Deterministic Sequence 1297, 1306
- developing context 4178-4179, 4181-4183
- Developing Countries 50, 77, 277, 673, 1659, 2425, 2439, 2514, 2604, 2739, 2746, 2791, 2805, 2993, 3066-3067, 3079, 3086, 3184, 3187, 3451, 3549-3550, 3553-3555, 3558, 3579-3580, 3585, 3587, 3591-3592, 3621, 3654, 3658, 3670, 3679, 4136, 4180-4182, 4184, 4439, 4685, 4710, 4794, 5016, 5253, 5307, 5320, 5337, 5340, 5357, 5363, 5436, 5870, 5902, 5940, 5960, 6147, 6161, 6216-6218, 6221, 6280, 6916-6918, 7144, 7209, 7211, 7214, 7216, 7227, 7234, 7236, 7251, 7258, 7665, 7817, 7829, 7831, 7848-7851, 7913, 7991
- Developing Economy 676, 3579, 7233
- Development Banks 5015, 5022, 6918, 6926
- Development Economics 7214, 7222
- Development Stakeholder 5084
- Development Studies 6164, 7201
- Developmental Disabilities 287-288, 290-292, 297, 5940
- Developmental Stages of Digital Literacy 2325
- Device Technology 1410, 7375, 7724
- Diachronic 1206, 1208, 1212, 1215, 5068
- Diagnostic And Statistical Manual Of Mental Disorders 995, 4233, 4306
- Diagnostic Online Reading Assessment 5183
- Diagnostic Process 1017, 1019, 1024, 2698, 2701-2702, 4141, 4145, 5843
- Diagnostics Test 2223
- Dialogical and Network Nature 7109
- Diamond-Like Carbon (DLC) 431-432, 439

Index

- Differential Flatness Theory 15, 23, 29, 35
Differential Gene Expression 442
Differentiated Service (Diffserv) 3077, 6613
DiffServ with MPLS 6614
Diffusion of Information Technology 7268
Diffusion Rate 3050, 4561-4562, 4569
Diffusion Theory 3655
diffusive transport 2679, 2681, 2684
digit multiplication 4609-4610
digital access 2727
digital advertising 4255
Digital Age 683-684, 686, 690, 874, 1474-1476, 1511, 1594, 1706, 2234, 2264, 2292, 2357, 2411, 2794, 2830, 3484, 3591-3595, 3602, 3630, 3957, 4189, 4351, 4515, 5150, 5154, 5204-5206, 5209, 5211-5212, 5267, 5627, 5854, 6336, 6718, 6751, 7145, 7292, 7637, 7929, 7988
digital audio 5972, 5980, 6447-6449, 6451-6452, 6458-6459, 6463
Digital Audio Broadcasting (DAB) 5972, 5980
Digital Buyers 2896
Digital Citizenship 2235, 2237-2238, 2240, 2242, 3242, 7625
Digital Communications 2121, 2295-2296, 4017, 5377, 6014, 6449
Digital Consumption 5575, 6394
Digital Content 1422, 1427, 1936, 2229, 2235-2237, 2240, 2242, 2484, 2487, 2908, 2914-2915, 3516, 4161, 4439, 4447, 4872, 6072, 6363, 6475, 6750, 7037, 7109, 7127, 7273, 7623, 8055, 8097
digital corpus 5232, 5236
Digital Curation 2242, 6752, 6755
digital currencies 2866, 2869, 2878, 2880
digital development 2282-2283, 2285-2287, 2289, 2291
Digital Divide 1426, 1431, 1513, 2230-2231, 2234, 2238, 2242-2247, 2251, 2282-2283, 2287-2288, 2291, 2330, 2515, 2520, 2526-2527, 3187, 3522, 3525, 3561, 3565, 3567, 3573, 3670, 3747, 4181-4184, 4619-4625, 4628, 5754, 5859, 6394, 7015, 7059, 7209, 7211, 7223, 7225-7226, 7230, 7233, 7248-7254, 7258
Digital dualism 4296, 4298, 4302
Digital Economy 688, 1422, 3594, 3930, 4278, 4705, 4710-4711, 4796, 5285, 6776, 7264
Digital Environments 1186, 2235, 2240, 2242, 2297, 2326-2327, 2330, 2335, 2368, 4156
Digital Equity 6688, 6690-6691, 6694
Digital Exclusion 2228, 2283, 2503-2504, 2509
Digital Fabrication 505-510, 513, 515-518, 520, 5108, 7867
Digital Filter 4759, 6007, 6017-6018, 6174, 6234-6235, 6241, 6244
Digital Footprint 2242, 4171, 4177, 7100
Digital Forensics 1357-1358, 1362, 1365, 2919, 3531
Digital Frequency 4746, 6017, 6234
Digital Game-Based Learning (DGBL) 3238, 3240, 3247, 3249, 3259
Digital Games 1422-1423, 2425-2426, 3238-3240, 3244, 3247-3249, 3253, 3255, 3259, 3287, 3310, 3317-3319, 3321-3322, 3326, 6380-6381, 7359, 7880
Digital Generation 2513, 2520, 7101, 7103, 7105, 7193, 7540
Digital Generation Gap 2513, 2520, 7101, 7103, 7105
Digital Health 3692-3693, 3695
Digital Humanities (DH) 5069, 5076, 5236, 7062, 8020
Digital Image Forensics (DIF) 2911, 2919
Digital Immigrants 1509, 1511, 1513, 1516, 1706, 2329, 2331, 2506, 2512-2513, 2520, 6389
Digital Inclusion 2228, 2245, 2280, 2283-2285, 2287-2289, 2503, 2505-2506, 2509, 2517, 2520, 3191, 4621-4622, 6382, 6695, 6704, 7251, 7253-7254, 7258
Digital Inequality 4623, 7249, 7251, 7540
Digital Information 326, 987, 1383, 1656, 2243, 2245, 2252, 2254-2255, 2293, 2299, 2305, 2314, 2325, 2478, 2480, 2908, 4017, 4198, 4453, 4458, 4735, 4769-4770, 4796, 5217, 6465, 7211, 7887
Digital Innovation Contest 3041
Digital Interface 4212, 4215, 4222, 7985
Digital Learning Materials 2547, 2556, 7681
Digital Literacy 1422-1423, 1426, 1431, 2225-2231, 2235-2240, 2242-2248, 2252-2257, 2280-2284, 2286-2287, 2289, 2291, 2294, 2307, 2314, 2316, 2320, 2325-2331, 2335, 2397-2398, 2492, 2502-2503, 2505-2506, 2516, 2571, 2579, 2650, 2912, 3830, 6405, 6859, 6866, 6870, 7061, 7248, 7251-7252, 7258, 7542, 7545, 7548
Digital Literacy Maturity Model 2291
digital microscopy 5853-5854, 5860, 5862
Digital monism 4296, 4298-4299
Digital Natives 1463, 1465, 1467, 1508-1510, 1512, 1516, 1706, 2230, 2239, 2248, 2253, 2320, 2329-2331, 2395, 2419, 2422, 2506, 2512-2515, 2520, 3249, 4358, 4367, 6389-6390, 6394, 7539-7540, 7544, 7665, 7670
Digital Object 2485, 2638, 4438, 4442, 5225, 5266
digital participation 2280, 2286
digital production 2234
Digital Public Art 6695, 6699

- Digital Public Library of America (DPLA) 5226, 5229
 Digital Quotient 684-685, 690, 692
 Digital Reconstruction 206, 5220, 5222, 7856, 7858, 7860-7862, 7864-7868
 digital religion 6161-6162, 6164, 6166-6168, 6170
 Digital Repositories 2483, 4437, 4447
 Digital Representation 542, 973, 975, 985, 3446, 5213, 5225, 7857, 7868
 Digital Rights Management 1658, 2487-2488, 2491, 6559
 Digital Scholarship 2484, 4439, 6749, 6780, 6785, 6787
 Digital Signal 4008, 4017, 6007, 6017, 6094, 6118, 6235, 6256-6258, 6447
 Digital Signal Processing 4017, 6007, 6094, 6118, 6235, 6256-6258, 6447
 Digital Signature 641, 653, 1033, 1385, 1649-1651, 1655, 2715, 2754, 2871, 2880, 3058, 4872, 4883, 4900, 4928, 6560, 7723
 Digital Stethoscope 1015
 Digital Storytelling 2442-2444, 2446-2448, 2450, 2454, 2616-2618, 2620-2622, 2624, 2641-2642, 2646, 3309-3310, 3316, 6393, 6477
 Digital Surveillance 3653, 4773, 5226
 Digital Technologies 801, 886, 980, 1421, 1426, 1516, 2226-2227, 2229-2231, 2234, 2251, 2280, 2291, 2325, 2331, 2513, 2515, 2520, 3354, 3573, 3602, 3605, 3645, 3794, 4157, 4184, 5204-5205, 5208, 5316, 6339, 6360, 6380, 6787, 6996, 7062, 7111, 7127, 7250, 7252, 7254, 7258, 7312-7314, 7320, 7322, 7325, 7358, 7539-7540, 7677, 7857, 7867, 7982, 8095
 Digital Tools 395, 874, 975-976, 1512, 2238-2239, 2244, 2246, 2254, 2314, 2325, 2328, 2488, 2504, 2507, 2552, 2559, 2643, 3798, 7249, 7539-7540, 7858
 Digital Transformation 388, 601, 682-685, 690, 692, 873-874, 1422, 1424, 1426-1427, 1431, 2346, 2350, 7984
 Digital Video 1696, 1891, 2357, 2449, 2641-2647, 2650, 2660, 4872, 4880, 5972, 5980, 6476, 7891
 Digital Video broadcasting (DVB) 5972, 5980
 digital video storytelling 2641-2644, 2646-2647, 2650
 digital video watermarking 4872, 4880
 Digitalization 873-875, 877-880, 882-883, 886, 1196-1197, 1657, 5207, 5306-5307, 5316, 7983, 7985
 Digitization 373, 682, 692, 873-874, 886, 2726, 3622, 5207, 5212-5213, 5216-5217, 5223, 5228-5230, 5232, 5291, 5311, 6781, 8095
 Dimensions of Interoperability 6578
 Dioxins and Furan 3076
 Direct Bullying 4254
 direct cyberattacks 5986, 5988
 Disability-Aware Information System 323
 Disaster Response 1409-1412, 1414-1415, 2065, 2069, 2467-2469, 2472
 Disciplinary Literacy 1509, 1516
 discipline integration 6869
 Disclosure Of Information 755-756, 759, 762-764, 4871
 Discourse Analysis 1017-1018, 3289, 5160, 5162, 5666
 discourse community 1219-1220, 1223, 1225
 Discrete Event Model 7488, 7502
 Discrete event processes 7488, 7499
 Discrete Event Systems 6836, 7435-7437, 7439, 7445, 7489, 7502
 Discrete-Event Simulation (DES) 1862, 5335, 5337, 5401, 5440
 Discrimination 184, 220, 224, 228, 278, 286, 412, 500-501, 1314, 2204, 2208, 2433, 2849, 3206, 3383, 3394-3395, 3397, 3400-3401, 3507, 4951, 4955, 4957, 5892, 6128, 6994, 7214, 7219-7220
 Discussion Board 1455, 2575, 2579, 3903, 3935, 5264, 7088
 Disease Mapping 3451
 Disengagement 1523, 1561, 1726, 2549, 3567, 7190-7195, 7199, 7516, 7924
 disjunctive queries 7995-7996, 8000, 8002-8003
 Disk I/O Utilization 7575
 Disney, Walt 4702
 Disposition to Trust (DTT) 3657, 3659, 3661-3662, 3665
 disruption risk 5553, 5555-5556, 5558
 Disruptive Innovation 2348, 3624-3626, 5194, 7899
 Disruptive Technologies 4774-4775, 4777-4781, 4784
 Dissemination 381, 398, 420, 730, 755, 759, 761, 776, 1407, 1550, 1911, 2064-2065, 2243, 2484, 2490, 2580, 2591, 2908, 2912, 2915, 3132, 3648, 3872, 3880, 3899, 3923, 3929, 4210, 4216, 4257, 4310, 4348, 4395, 4397, 4406, 4415, 4438, 4450, 4472, 4490, 4536, 4707-4708, 4711, 4724, 4761, 4822, 4827, 4829, 5024, 5085-5086, 5090, 5217, 5238, 5246, 5256, 5312, 5617-5618, 5661, 5676, 5757, 5761, 5819, 5960, 6149, 6478, 6483, 6501, 6526, 6556-6557, 6562, 6686, 6711, 6781-6782, 6940, 6997, 7005, 7044-7045, 7048-7049, 7052, 7071, 7109, 7113, 7126, 7160, 7206, 7223-7225, 7230, 7272, 7296, 7555, 7796, 7800, 7847, 7956, 7988, 8016-8017
 Distance Education 1451-1452, 1455, 1491, 2455-2456, 2478, 2503-2505, 2507-2509, 2588, 2590-2591, 2594, 2661, 2667, 3832, 3937, 5808,

Index

- 5811-5813, 5820-5823, 5829-5834, 5837, 5840, 5854, 6380, 6412-6416, 6422, 6429, 7117, 7899-7901, 7912, 7921
- Distance Learning 1462, 1677, 1883, 1887-1888, 2455-2460, 2465, 2477, 2505, 2507, 2550, 2589-2592, 2594, 3832, 3834, 3850, 3852, 3900, 3937, 3940, 3944, 4720, 4824, 5634-5635, 5789, 5808, 5824, 6413, 6416, 6438, 7882
- distance measure 1788, 1827, 1840-1841, 1844, 1962
- Distractor Pool Method with Random Selection 2372, 2378
- Distributed Authorship 4146, 4150, 4155
- Distributed Autonomous Control 6837, 6844
- Distributed Contributions 8037, 8045
- Distributed Cooperative Relaying 6499, 6505-6506
- Distributed Human Intelligence Tasking 572
- Distributed Interoperability 6566, 6578
- Distributed Knowledge 930, 936, 2593, 2598, 5400, 8036
- Distributed Learning Environment 1498, 1522, 7674, 7927-7930
- distributed parameter systems 15, 17, 21, 28
- Distributed Programming Frameworks 375, 380
- Distributed protocols 6524
- Distributed Store 7809-7810, 7813
- Distributed Strategies 6523, 6535
- Distributed System 337, 456, 1077, 6580, 6837, 8065
- Distribution Centers 5275, 5356, 5369, 5404, 5507, 5509, 5520, 7874
- Distribution Management System 3113
- Diversification 8, 61, 730, 4881, 5215, 6488, 7693, 7700, 7703
- DMADV 729-730, 733, 735, 739
- DMAIC 729-730, 733, 735, 739
- DNA 186, 206, 221, 419-424, 440, 443-444, 450, 453-454, 466-467, 474-478, 482-483, 489, 682, 687, 2901, 3694, 3798, 4643, 4986, 4993-4994, 5003, 7731, 8015
- DNA Microarray 443-444, 450, 453-454
- D-n-tuple 1996, 1999, 2003, 2005
- Domain Expert 669, 3339, 7400
- Domain Knowledge 2, 100, 102, 180, 183, 461, 1736, 1783, 2143, 2236, 2550, 2705, 5127, 6341, 7163, 7399-7400, 7402, 7407, 7423, 7672
- Domain Level Explanation 2094
- Domain Specific Language 765
- Domotics 3275
- Dosimetry 1249, 1259
- dot code 6464-6468, 6470, 6474-6475
- dot code technology 6464-6465
- dotted raster 172-174, 179
- Dotted Raster-Stereography 166, 172-175, 177, 179
- double spend 2790
- DPMO 731
- DRAM 1809, 1956, 4010, 4016-4017, 5572
- Drawing Graphs Methods 7421
- Driver Attention 6005
- Driver Distraction 6005
- Dropshipper 2805, 2809
- DSM 1949, 1961, 4224, 4233, 4810
- DSP 4008, 4017, 5413, 5415-5416, 6007, 6235, 6240-6241
- dsRNA 478, 483-484, 489
- DSS 933, 1029, 1875, 2168, 2932, 2934, 4573, 5504, 5511, 6458
- D-system 1997, 1999-2000, 2005
- DTM 2048, 3434, 3440, 3446, 3496
- Dual watermark 4873
- Dynamic Capabilities 936, 3845, 5425, 5715, 5725
- Dynamic Content 4215, 6993, 8045
- Dynamic Creative Optimization (DCO) 5750
- dynamic decision making 2108-2111, 2115, 2120
- Dynamic Fault Trees 6848, 6856
- Dynamic Pricing 350-351, 1166, 2346, 3113, 5337, 6548, 6550, 6554
- Dynamic Statistic 4635, 4642
- Dynamic Time Warping (DTW) 5879, 5885

E

- E01 1384, 1386-1388, 1390-1394
- E-Activism 3569-3571, 3573-3574
- E-Applications 3560
- Earned Value Management 5680, 5682, 5689
- Earned Value Management System (EVMS) 5682, 5689
- Ease Of Use 531-533, 755, 760, 833-835, 838, 840-841, 846, 1063, 1069, 2123, 2358, 2360-2361, 2468-2469, 2471, 2476, 2494, 2659, 2734, 2855, 2886, 2896, 2956, 3582, 3587, 3681, 3787, 3793, 3821, 4089, 4134, 4182, 4203, 4281, 5005-5006, 5008, 5013, 5586, 6022, 6096, 6149-6150, 6210, 6403, 6803, 6847, 7031, 7164, 7193, 7506, 7608, 7746, 8078
- EAST-ADL 6848, 6858
- Ebay 1167, 1658, 2741, 2743, 2833-2837, 2841, 2865, 2886, 7798, 8019, 8091
- Echo Chambers 6783, 6999, 7001
- Eco-Efficiency 3088, 5357, 5360-5361, 5366, 5382
- E-Collaborative Learning (E-CL) 6336, 6346
- Ecological Performance 730, 5357-5358, 5360, 5362-5364, 5366
- E-Commerce Platform 2740, 2748, 2857, 5370, 5388

- Economic Crisis 801, 1559, 2344, 2844, 3806, 3931, 4786-4787, 5546, 5553, 5630, 5642, 7003-7004, 7275, 7756
- Economic Development 676, 2149, 2171, 2341, 2479, 2802, 2994, 3000, 3010, 3013-3014, 3018-3019, 3021, 3376, 3403-3404, 3413, 3434, 3436, 3440, 4119, 4283, 4350, 4774-4775, 5015-5016, 5036, 5041, 5293, 5305, 5368, 5479, 6216, 6915, 7205, 7220, 7224-7225, 7227, 7233, 7235-7236, 7259-7260, 7265-7266, 7843, 7870, 7991, 8096
- Economic Freedom Index 47, 8098, 8104
- Economic Globalization 7981, 7993
- Economic of Information Security 4860
- Economic Recovery 2341, 2343-2344, 5479
- Economics of Scale 886
- Economics of Trust 260, 263
- e-CRM systems 1616-1621, 1624-1626
- E-Culture 4189-4194, 4196, 4198, 7273
- E-Culture/Digital Culture 2292-2293, 2299, 4189-4194, 4196, 4198, 6161, 7174, 7273
- Edge Betweenness 6972
- e-Discovery 5987, 5991, 6095, 6104
- EDM 5160, 5170, 5182
- edTPA 7682-7689, 7691
- Educational Data Mining 1945, 4440, 5159-5160, 5168, 5170, 5182
- Educational Digital Storytelling 3309, 3316
- educational ontology 1441, 1445-1446
- Educational Robotics 2492-2495, 2497, 2501, 6860, 6862-6863, 6865, 6870
- Educational Robotics Packages 2494-2495, 2497, 6862
- Educational Technology 1464, 2357, 2421-2422, 2426-2427, 2466, 2477, 2487, 2512, 2517, 2532, 2571, 2579, 2658, 2660, 3334, 4330, 5834, 5836, 6377, 6398, 6412, 6414, 6861, 6863-6864, 7880, 7882, 7907, 7939
- Educational Theories 1518, 5800, 6355, 7939, 7942-7944
- educational toys 2626-2630, 2637-2638
- Edutainment 3287, 3328, 3330, 3336
- E-Engagement 3564, 3568
- Effect-Based Operations 7306, 7311
- Effective Average Gradient (EAG) 1308, 1312, 1318
- Effective Ethics 94
- Effective Isotropic Radiated Power (EIRP) 6260, 6265
- Effectiveness of Animation 2615
- Efficient Market Hypothesis 72-73, 5060, 5066
- Effort Expectancy (EE) 2253, 2762-2763, 2765, 2769, 3581-3582, 3584, 3586, 3590, 3655-3656, 3659-3662, 3666, 3787
- e-filing 831-832, 835, 840-841, 846
- EFLOPS 4015-4016
- e-fulfillment 5369, 5371
- E-Governance 888, 2229, 3603-3604, 4099, 4101-4102, 4105-4106, 4110, 4112, 4255, 6710, 7245
- E-Government Services 841, 2342, 3549-3550, 3558, 3560, 3579-3587, 3591-3595, 3603, 3605-3606, 3608, 3654, 5022, 6566, 7843
- EHEA 3870-3871, 3880
- eHealth technologies 5864-5866, 5868-5870, 5875
- eHealth Technology Feature 5875
- eHealth Technology Functionality 5875
- EHR Adoption Barrier 3688
- EHR Diffusion 3688
- EHR Resistance 3688
- EHR Technology Adoption 3678, 3688
- Eigenvector 1746-1747, 1749-1751, 1753-1754, 1756, 6507-6510, 6516-6517, 6520, 6596, 6602, 7718
- Eigenvector Centrality 1746-1747, 1749-1751, 1754, 1756, 6507-6510, 6516-6517, 6520
- Elasticity 1030, 1067, 1077, 1184, 1356, 1467, 1776, 6550, 6730, 7241, 7525
- E-Learning 184, 314, 319-320, 613-614, 618, 808, 833-835, 1443, 1676, 1874, 1883, 2254, 2303-2304, 2380-2381, 2383-2386, 2391, 2417, 2421, 2427, 2430, 2455, 2465, 2559-2560, 2562, 2582, 2585, 2587, 2592, 2637-2638, 2640, 2667, 2671, 3572, 3831, 3840, 3900, 3915, 3921, 4122, 4130, 4593, 4798, 5161, 5168-5171, 5184, 5192, 5245, 5634, 5787-5790, 5792-5794, 5799-5800, 5808, 5811, 5820, 5822, 5853-5855, 5858-5860, 6188, 6339, 6348-6352, 6402, 6415-6416, 7545, 7902, 7907, 7912-7917, 7921, 7985
- E-Learning Platforms 2585, 2592, 2638, 3915, 3921, 5788, 7545, 7914
- electric system 3144, 3146, 3151, 3155
- electrical conductivity 2673-2676, 2678, 2680-2681, 2684, 2897-2898, 2900-2902, 2906
- Electrocardiogram (ECG) 1006, 1015, 3707
- Electroencephalography (EEG) 1824, 5887, 5899
- Electromagnetic Fields 1249, 1994, 6124, 6135, 6710, 6712-6713, 6718
- electromagnetic models 6321-6322, 6334
- Electromagnetic Shielding 2907
- Electromagnetic Waves Energy 3487, 3501
- Electromedical Department 3698, 3707, 3710
- Electromedical Equipment 3698-3702, 3707, 3710
- Electronic Advocacy 3569
- Electronic and Information Technology 3516-3517, 3520, 3525

Index

- Electronic Commerce (EC) 49, 51, 184, 259-260, 314, 334, 349, 352, 388, 643, 712, 831, 834, 870, 938, 1616, 1622, 1735, 1757, 1775, 1869, 1874, 1876, 1878, 2108, 2694, 2707-2709, 2711-2712, 2716-2718, 2720, 2722, 2739-2748, 2750, 2752, 2761, 2773, 2779, 2783, 2789, 2791-2794, 2801-2807, 2810-2813, 2815-2818, 2821-2822, 2824-2830, 2832, 2841, 2844-2845, 2851-2858, 2860, 2869, 2885-2886, 2890-2891, 2953, 2956, 3606, 3654, 3793, 3869, 4189-4190, 4278, 4284, 4709, 4798, 4830, 4839, 4863-4866, 4931, 4937, 5021, 5345, 5367-5377, 5388-5390, 5392, 5394, 5397, 5865, 6067, 6070-6073, 6096, 6341, 7118, 7120-7121, 7124, 7264, 7588, 7590, 7758, 7843, 7869, 7873-7876, 7985-7987, 7993, 8049, 8087, 8089-8090, 8092
- Electronic Connectivity 7980, 7983-7984, 7986, 7990-7991, 7993
- Electronic Data Interchange 2168, 2792, 4573, 4575, 5352, 7874, 7878
- Electronic Devices 287-288, 297, 1296, 1319, 1994, 2673, 2675, 2681, 2715, 2794, 2897-2898, 2902-2903, 2907, 3057, 3066, 3103, 3149, 3956-3957, 3959-3960, 4122, 4170, 5210, 5856, 5908, 5982, 6115, 6136, 6284, 6319, 6337, 6346, 6349, 6360, 6387, 6415, 6697, 7987
- electronic Document and Records Management System (eDRMS) 3884, 3898
- Electronic Documents Flow 3916, 3921
- electronic documents management 889-891, 893, 895, 898
- Electronic Government (E-Government) 698, 831, 841, 1037, 2342, 2943, 3516-3519, 3521-3522, 3530, 3549-3553, 3555, 3557-3558, 3560, 3563-3564, 3569, 3579-3587, 3590-3595, 3601-3608, 3611, 3654-3655, 3657, 3659, 3662, 3809, 4709, 5018, 5022, 6566, 6575, 7843, 7985
- Electronic Health (E-Health) 314, 326, 333, 523, 3678-3680, 3685, 3688, 3690-3691, 3782-3786, 3788-3790, 3793, 3805, 3807, 3809-3810, 3814, 3817, 3823, 4689, 5021, 5820, 5828, 5864-5865, 5870, 5902, 5904-5905, 5908, 5910, 5913, 5916-5917, 5930, 5932, 7756
- Electronic Health Records (EHR) 333, 3679, 3688, 3788, 3805, 3807, 3814, 3817, 3823, 5864
- electronic IQA (eIQA) 3881, 3898
- Electronic Management Information System (eMIS) 3883, 3899
- Electronic Medical Record 102, 3656-3657, 3680, 3806, 5901, 5913
- electronic payment frameworks 2749-2750, 2757
- Electronic Payment Protocol 2760
- Electronic Payment System 60, 2760
- Electronic Payments (E-Payments) 2760, 2815, 5397, 7214, 7986
- Electronic Product Code (EPC) 4952, 4961, 5520, 5526, 5604, 7777
- electronic publication 5830, 5840
- Electronic Resources 2271, 2487, 4191, 4196, 4525, 5204-5207, 5209, 5212, 5215, 7628
- Electronic Theses And Dissertations 4439, 6748-6749, 6755
- Electronic Waste (E-Waste) 3070, 3076
- Electronic Word-Of-Mouth (Ewom) 712, 4032, 4091, 5982, 6019, 6030
- Electronics Industry 2897, 2899, 2902-2903, 5572
- Elephant Foot 4383, 4388
- Elevation Analysis 3436, 3447
- Elitist Mutation 243, 249
- Eliza Effect 257, 263
- Elliptic Curve Cryptography 1668, 4977, 7724
- Ellis Model 4522
- E-Marketing 3869
- Embed Code 2657, 2660
- Embedded Metadata 6448, 6463
- Embeddedness 723, 2557, 3923, 3926, 6162, 6972, 7059, 7155, 7313, 7321, 7671, 7681
- Embodiment 877, 3351, 3354-3355, 4234-4240, 4244, 5911, 6704, 7774
- E-Mentoring 3852, 3854, 3860
- Emergent Technologies 4560-4562, 4565-4566, 6784
- Emerging Economies 61-62, 966, 2791, 3002, 4774-4775, 5015, 5026, 5317, 5560, 6918, 7265
- emerging solutions 5559, 5561-5563, 5565, 5569
- Emerging Tourism Economy 4032-4033, 4037
- EMFs 6124-6131, 6135
- Emoticons 1510, 3769, 7073-7074, 7076, 7318, 7971, 7973, 7975
- Emotion Recognition 132, 143, 4202, 5749
- Emotional Arousal 139, 143, 6942
- Emotional Engagement 1488, 1498, 6413
- Emotional Intelligence 286, 1534
- Emotional Regulation 3318, 3326
- Empirical Analysis 382, 2835, 2869, 4785, 4787, 7238
- Empirical Cumulative Distribution Function (ECDF) 3713, 3715, 3723
- Empirical Distribution Functions 1857, 1861
- Employees' Performance 4337-4338, 4343-4344
- Employment 277-281, 564, 677, 683, 740, 1260, 1469, 1473, 1540, 1559, 1566, 1568, 1697, 2245, 2247, 2305, 2422, 2501, 2989, 2991-2994, 3014, 3043, 3374, 3376, 3385, 3395, 3397, 3434, 3770, 3831,

- 3865-3866, 3931, 4156, 4267, 4621-4622, 4708, 4763, 4785-4787, 4790, 4843, 5016, 5018, 5423, 5477, 5630, 5810, 5858, 5931, 5935-5936, 6221, 6354, 6412, 6415, 6727, 6735, 6908, 6916, 6918, 6988, 7026, 7031, 7224, 7227, 7229, 7234-7235, 7238-7241, 7245, 7261, 7269, 7982, 7990
- Empty Core 2107
- Enabling Technologies 1026, 3150, 3155, 7754-7757
- ENC 707, 719, 4975-4977
- Encyclopedia Dramatica 8016
- End User Piracy 1658, 1665
- End Users 132, 318-319, 353, 355, 381, 390, 778, 1153, 1158, 1161, 1261, 1609, 1615, 2032-2033, 2035, 2380, 2384-2385, 3144, 3185, 3409, 3412-3413, 3475, 3654, 4438, 4472-4473, 4543, 4785, 5226, 5321, 5361, 5383, 5387, 5546, 5886, 6465, 6546, 6618, 6908, 7168, 7170, 7372, 7375, 7377, 7505, 7617, 7747, 7777, 7849, 8065, 8084
- Energy Consumption 299, 501, 543, 3043, 3052, 3077-3079, 3085, 3104, 3187-3188, 3371, 3374, 3464, 4018, 4549, 5006, 5362-5363, 6253, 6289, 6522, 6524-6526, 6528-6531, 6533, 6537, 6561, 7745-7751, 7776, 7815, 7817-7818, 7820-7821, 7823-7824, 7829
- Energy Consumption Pattern 7817, 7821, 7824, 7829
- Energy Management System 3113
- Energy Poverty 3374, 3381
- Engineering Design 239-241, 244, 247, 430, 550, 2626, 2633-2634, 2636, 3795-3797, 3804, 7383, 7543
- English as Foreign Language (EFL) 7661
- English Auction 2835, 2840
- English usage 1213, 1217
- Enhanced Data Rates for Global Evolution (EDGE) 6265
- Enanglement 7730
- enquiry-based approach 5116, 5118, 5126
- Enquiry-Based Learning 5116-5117, 5122, 5124, 6399
- ensemble methods 1829, 1962, 1969, 2085
- Enterprise 2.0 765, 2775, 2779
- Enterprise Agility Matrix 1405
- Enterprise Architecture 574-580, 620-621, 631, 638-639, 810-811, 819, 2920-2928, 2930, 2943, 2952, 3530, 4482, 4810, 5607-5610
- Enterprise Data Architecture 5618, 5625
- Enterprise Model 4821
- Enterprise Network 376, 380, 6616
- Enterprise Resources Planning (ERP) 349, 387, 554, 652, 834, 923-924, 929, 2168, 2933, 2974-2975, 2986, 5288, 5305, 5340, 5392, 5519, 5561, 5564, 5649, 8099
- Enterprise Resources Planning (ERP) System 925, 929, 2934, 2953-2957, 2962, 2969, 2974-2975, 2977-2979, 3915, 5564
- Enterprise System 657, 827, 1405, 2160, 2933, 2935, 2937-2938, 2941, 2969, 3820, 4392, 7040
- Entrée system 1757-1759, 1768
- Entrepreneur 63, 803, 2989, 2993, 2998-3001, 3009-3011, 3016, 3019-3022, 3025-3026, 3031, 3413, 5533, 6813, 7989
- Entrepreneurship 62, 800-806, 808, 2424, 2794, 2989-2994, 2998-3003, 3007, 3009-3016, 3019-3023, 3026-3027, 3031-3032, 3177, 3830-3831, 4347, 4710, 5259, 5460, 5464, 5533, 5671, 6916, 6995-6996, 7000, 7063, 7248, 7261, 7263, 7989, 7994, 8099
- ENTRExplorer 802, 804
- Entropy Combination 1242-1243, 1655
- Entropy Extraction 1242, 1649, 1655
- environment management 5357-5358, 5361
- Environment Mapping 6846
- Environmental Assessment 812, 3081, 3131-3133, 3137
- Environmental Concern 1901, 5382, 5475, 7359
- Environmental Hazard 3076
- Environmental Impact 223, 611, 1936, 3131-3133, 3190, 3699, 5358, 5362, 5379-5380, 5384, 5466-5467, 5597, 6718, 7772
- Environmental Management 425, 1935, 3087, 3094, 3133, 3186-3187, 3190-3191, 3381, 5357, 5360, 5381-5383, 5446, 5450, 5455, 5531
- Environmental Monitoring 1677, 1679-1680, 3276, 3694
- Environmental Performance 615, 618, 3133, 3189, 5358, 5366, 5427, 5447, 5467, 5470
- environmental scanning 899-900, 905-907, 909-910
- Environmental Sustainability 609, 615, 618, 3076, 3131-3133, 3190, 3435, 3437, 4895, 5428, 5573, 7815, 7817
- E-Participation 3561-3563, 3565, 3569, 4267, 4276
- Ephemeral Communications 4310, 4313
- Epigenetics 423-424, 430
- episode of care 2098, 2101
- Epistemology 1845, 1855, 3925, 4065, 4336, 4459, 4465, 4672, 6760, 7935
- Epoch 97, 131, 1574, 1856, 4300, 4459, 6004, 6647, 6657, 6665, 6670, 6739
- E-Portfolio 833, 5854, 5856-5857, 5859-5860, 5862
- E-Procurement 349, 538, 3594, 3606, 5296, 5390-5392, 5397
- Equation Modeling 1133-1134, 2954, 3584, 3586, 3660, 3666, 6020-6022, 6025-6026, 8070, 8072, 8078

Index

- Equity Premium Puzzle 71-72, 76
Equivalency 2189, 2191, 2456, 2465
Equivalent Mutant 7460, 7469
ERASMUS 1560, 4070
E-Retailer 2804, 2896
E-Reverse Auction 530-535, 538
Ergonomic Design 1260, 1263, 2640, 7030
Ergonomics 383, 393, 662-663, 1263, 2123, 2626, 3795, 3804, 4200, 5008, 5889, 6074, 7599, 7764
Erich Fromm 1638
ERP alignment 2966-2968, 2970-2971, 2974
ERP benefits 2953-2954, 2956-2957, 2960
ERP intervention 2981-2982
ERP-enabled application integration 2953-2957, 2960, 2964
error correction mechanism 5638, 5640, 5642, 5647
Error Correction Model 5639, 5649
error correction term 5640, 5649
Error Covariance Matrix (ΔP) 219
Error Function 6045, 6048, 6056, 6670
E-Science 461, 1065, 2393-2394, 2401, 6341
E-Service Quality 2882-2883, 2888, 2890, 2896, 6074
E-Services 3560, 3662, 4798, 6547
e-shoppers 2802, 2809, 2891
E-Shopping 834, 2342, 2761, 2764-2766, 2768-2770, 2772
e-shopping compliance 2761, 2768-2770
estimate at completion 5680, 5682-5683
Estimate Development Costs 7513
Estimate to Complete (ETC) 5679-5680, 5687, 5689
Estimate-Talk-Estimate (ETE) 662-663, 669
e-Supply Chain Management 5308
E-Tailers 2708, 2714, 2761-2762, 2767, 2769-2770, 2772, 5368, 5375
E-Tailing 2804, 2809
ETD 6748, 6750-6753, 6755
Ethical Computing 4884, 4890, 4894-4896
Ethical Dilemmas 3637, 3648, 4895, 5922, 5928
Ethical Risks 568, 6153
Ethical Standards 2293, 3203, 4196, 4895-4896, 6221, 7673
Ethics of Information Society 4198
Ethics of Norms 4921, 4926
European ATM Security Team (EAST) 78, 85
Evaluation Instrument 2438, 4871
Evaluation Measures 4473-4475, 4479
Evidence Based Policy 4962-4963, 4973
Evidence-Based Practice 269, 5810, 5813, 5819
Evolutionary Algorithm 159-160, 186, 241-242, 2134-2135, 4676, 5967, 6595
Evolutionary Art 158-160, 165
Evolutionary Computation 4, 100, 159, 180-182, 185-186, 3107, 7699-7700
Evolutionary Computing 8, 14, 159, 165, 182, 2134
Evolutionary Trajectory 4139, 4144-4145
e-voting 4898, 4900-4901, 4903, 4906-4907, 4997
Exascale 1069, 4015
Exclusive Climate 3391
Executive Control 3297, 3308
Executive Education 3833-3834, 3838
Executives 49, 61, 811, 917-918, 926, 953-957, 1147, 1595, 1597, 2085, 2694, 2957, 4348, 4776, 5160, 5393, 5477, 5482-5483, 5511, 5635, 5705-5706, 5708-5711, 5775, 5778, 7779, 8038
Exergame 7358-7360, 7362-7363, 7367
Existential Philosophy 4189, 4925
Exogenous 17, 824, 2157, 3026, 3429, 4336, 4922, 5509, 5670, 6020, 6022, 6403
Expanded Filter By N 4759
Expected Utility Theory 2189, 2199
Experience Design 4222, 7578, 7600, 8086
Experiential Learning 1529, 1538, 2411, 2418, 3289, 3831-3832, 3834, 3838, 4079, 4360, 7940-7941
Expert Systems 2, 100, 107, 112, 124, 183, 229-231, 235, 238, 669, 1875, 2201, 3108, 4573, 4743, 5402, 6478, 6590, 7110, 7336
Explanatory Capacity 6635-6636, 6640
Explicit Knowledge 552, 555, 943-944, 946, 2266, 4530, 4574-4576, 4580, 5026, 5046-5047, 5049, 5054, 5080, 5094, 5102, 7161-7162
Explicit Route 6612, 6614, 6616
Export Performance 7234-7235, 7244
Expressive Power of Explanation 2094
Expressive Writing 7952, 7955
Extended Literal 4738, 4745
Extensible Markup Language 2777, 5613, 6463, 7405
Extensive Reading 3945-3953
extent of ERP implementation 2955-2957, 2960, 2962-2963
External Environment 206, 208, 497, 540, 698, 879, 900, 903, 905, 907, 910, 963, 2177, 2345, 2945, 2952, 3087, 4775, 4855, 5318, 5390, 5715, 5747, 6073, 6320, 6840
external integrity 5287-5288, 5305
External Processes 4196, 5444, 5492
Extract, Transform and Load (ETL) 849, 851, 860
Extrapolation 5682, 5689, 6718, 7989
Extreme Programming (XP) 1398, 1406, 7504, 7578
Extrinsic Motivation 5102
Extrude 1290, 1295
Eysenbach 3770, 5865, 5868, 5902, 5908, 7181

F

- face identification 172-173, 175-176
- Face Recognition 166-168, 171, 174, 1678, 1826, 4986, 4992, 5278, 7778
- Facebook ads 3360-3362, 3366
- Face-To-Face Bullying 1723-1724, 1726-1728, 1733, 4168-4173, 4177, 4246
- Face-To-Face Classes 1550-1552, 1554-1556, 3903, 5787, 6413
- Face-To-Face Instruction 2455, 2478, 3244, 3247, 5789, 5813, 5823
- Face-To-Face Learning 2457, 2465, 3238-3239, 3844, 5799-5800, 5811, 5822, 5828
- Face-to-Face Victimization 1727, 1733
- Facilitating Conditions (FC) 3655, 3657, 3659-3660, 3662, 3666
- Factor Analysis 53, 1131, 1134-1135, 1711, 2886, 4787, 7649-7650, 8070, 8078
- Failed Projects 5668
- Failure Mode and Effects Analysis 6858
- Fair Use 2477-2480, 2483, 2489-2491, 2856, 5231, 5234, 5236, 6749
- Fall Detection 5880, 5885
- False Negative Rate 2214, 2223
- false positive rate 2214, 2223
- Fan Fiction 2329, 2335
- FANET 6106-6110, 6112-6113
- FANET Applications 6107
- Fault Tolerance 180, 1078, 1084-1086, 1088-1089, 1092-1093, 1096, 1116, 1119-1120, 1122, 1126, 1128, 1238-1240, 1243-1244, 1247, 1396, 1647-1648, 1650-1651, 1655, 1808, 3056-3058, 3060, 3062, 3065, 4550, 4928, 4930-4931, 4935, 6852-6853, 8056, 8060-8062, 8065
- Fault Tolerance Techniques 1085-1086, 1088, 1651
- Fault Tree Analysis 6847, 6856, 6858
- Fault-Tolerance Systems 6858
- Fear of Missing Out 4170, 4177
- Feasibility 23, 31, 159, 233, 244, 306, 631, 646, 651, 768, 827-828, 985, 1049, 1311, 1556, 2047, 2057, 2407, 3026, 3056, 3062, 3479, 3727, 3874, 4094, 4893, 4957, 5413, 5415-5417, 5492, 5609, 5902, 5905, 5942-5943, 5950, 6037, 6138, 6151, 6153, 6241, 6548, 6661, 6803, 6807, 6900, 7184, 7388, 7449
- Feasible Region 243, 2181-2182, 4596, 4606, 5416
- Feature Classification 496-497, 504
- Feature Contribution 2094
- Feature Extraction (FE) 135, 158, 186, 492-493, 496, 504, 1354, 2201, 2208, 6195, 6197, 6199-6203, 6205-6206
- Feature Matching (FM) 6198, 6205-6206
- Feedback Control 15-16, 20, 23, 29-30, 35, 828, 1400, 2156, 3047
- Feedback Neural Networks 6594
- Feed-Forward Neural Networks 6594
- Feminine gender 3364, 3383
- FHS 4475-4476, 4479
- FICO 7140, 7143-7144
- Fictional Texts 3948, 3953
- Fidelity 160, 366, 1052, 2094, 4004, 4202, 4872-4873, 4875, 4877, 4883, 5018, 5693, 5877, 6233, 6758, 6878, 7359, 7907, 8016
- Figurative Language 976, 1043, 5086, 5092
- financial exclusion gap 2734
- Financial Inclusion 2724-2726, 2729, 2733-2735, 2738, 6921, 7214-7216, 7218-7219, 7221-7222
- Financial Index 372
- Financial Internetization 7986, 7993
- Financial Investing 1-2, 5, 7-8
- Financial Statement 5066, 6633
- financially excluded 2724, 2733, 2738
- financially served 2724, 2735, 2738
- Financially Under-Served 2738
- Financing Performance 36
- Find and Make Test Data 7513
- Fine Mapping 471, 473
- Fine-Grain 4022
- Fine-Grained Granularity 4029
- Fingerprint 1385, 1678-1679, 2918, 4986-4988, 4992, 4994, 4996-4997, 5003, 5017, 5278, 6137, 6139, 6141, 6143, 6146, 6289
- Fingerprinting 2911, 6137-6139, 6143-6144, 6287, 6289-6290, 6295
- Finite Impulse Response (FIR) filter 6007-6008, 6183, 6235-6237, 6241
- Finite-Difference Time-Domain (FDTD) 1252, 1259
- Firefly Algorithm 6, 183, 2178, 2182, 2187, 4371, 7696-7697
- Firm Performance 912-914, 917, 922, 1596, 2793, 2980, 2989, 5715, 5725, 7154
- First Heart Sound 1008, 1016
- First-Time Visitors 4064, 4066-4067, 4072, 4076
- Fiscal Year 6820, 7259, 7336
- Flaming 1709, 1724, 3769, 4169, 4247, 7946-7948, 7952-7953, 7955
- Flesch Kincaid Grade Level Formula 1507
- Flesch Reading Ease Formula 1499-1500, 1502, 1507
- Flesch-Kincaid Grade Level 1499-1502, 1505, 1507
- Flipped Classroom 1476, 1498, 5803, 6340, 7917, 7921
- Flipped Learning 1510, 1516, 1551, 3839, 3849-3854, 3860, 5801-5805, 7643, 7645
- flipped lecture 5801, 5808

Index

- Floor Clipping Plane 5942, 5954
Floor Mat 2501, 6870
FLOPS 4017
Floss 1788, 6252
fluctuation theory 1573
Fluidized Bed 1342
Focal Firm 5348, 5982
Folklore 6695-6697, 6708, 8015
Folksonomy 2487, 7109-7110, 7114
Food and Nutrition 6940, 7203-7204
Force directed 7412
Force feedback devices 4654, 4656, 4658
Force-Directed Algorithm 7415, 7421
Forecasting 4-7, 15, 115, 120, 123, 181, 183, 186-187, 350-351, 406, 543, 557, 847-850, 854-859, 900, 905, 908, 929, 1376, 1606, 1608, 1813, 1835, 2008, 2091, 2170, 2779, 3108, 3491-3492, 3495, 3498, 3640, 3712, 4569, 4763-4764, 4773, 5290, 5296, 5309, 5336, 5375, 5516, 5519, 5522, 5541-5542, 5596, 5644, 5679-5682, 5684-5687, 5698, 5766, 6591, 6632-6633, 6659, 7871
Foreign Students 3900, 3909
Forensic Data Analysis (FDA) 1378, 1383
Forensic Investigation 376, 1356-1357, 1359, 2915, 2918, 4855, 5660
forensic pattern warehouse 1376, 1378-1379, 1381
Forensic systems engineering 5661, 5668
foresight process 907
Formal And Informal Learning 1517-1520, 1523, 2548, 4178, 6351, 7174
Formal Approaches 5434, 5438
Formal Concept 192-193, 197, 1929-1930
Formal Learning 1519-1521, 1528, 2248, 2256, 2317, 2556, 3842, 3849, 6337, 6355, 6399, 6403, 6416, 7542, 7625, 7680, 7912
Formal Methods 3109, 3924, 6878, 7445, 7450, 7452, 7455, 7609-7610, 7615, 7621
Formal Notations 7610-7611
Formal Proof 7621
Formal Specification 823, 5869, 7451-7452, 7454-7455, 7610, 7615, 7621
Formal Specification Languages 7610, 7621
Formal Verification 7610, 7615-7617, 7621
Formative Assessment 1469, 1560, 2230, 2369, 2376, 2507, 5117-5118, 5122-5124, 5126, 5183, 5192, 7384-7385
formative evaluation 1880, 1882, 1890, 1892, 1894
FOSS 3013, 3333, 5715, 6252, 7917
Fourth Industrial Revolution 682-683, 690, 692, 886
FPGA 1956, 4008, 4017-4023, 4025, 4029, 4551-4552, 4554, 4611-4612, 4614, 6241, 6258, 6313
Fragmentation Redundancy Scattering 4936
Franchise Relationship 2688
Franchisee 930-936, 938, 941, 2686-2692, 2694, 2696, 4822, 4824, 4826-4829, 4831, 4835
Franchisee Life Cycle 931, 936, 941, 2687, 2690, 2696, 4835
Franchising 930, 933-934, 937-938, 2686-2692, 2694, 2696, 2977-2978, 2991, 4822-4824, 4826, 4828-4831, 4835
Franchisor 930-938, 941, 2686-2688, 2691-2694, 2696, 4822-4829, 4831, 4835
Franchisor/Franchisee Learning Process 941, 2696, 4835
Franchisor/Franchisee Relationship Management 941, 2696, 4835
Fraud 78, 258, 260, 333, 409, 412-413, 615, 636, 642, 821, 824, 827, 874, 1366, 1686-1687, 1806, 1828, 2088, 2157, 2219, 2707, 2709, 2711, 2713-2714, 2716-2718, 2720, 2722, 2730, 2755, 2784, 2856, 4837-4839, 4841-4842, 4937-4943, 4945-4948, 4962, 5311, 6038, 6221, 6279, 6633-6634, 6705, 6802, 6925, 7143, 7987
Frederick University 2494, 2497, 6862
Free Container Application 7645
Free Information 4189, 4459, 4461, 4469
Free Software (FS) 854, 1109, 2863, 3463, 4602, 6069, 6248, 7369, 7371, 7377, 7380-7381, 7911
Frequency Response 2901, 5972, 6008, 6017-6018, 6234-6239, 6241-6242, 6244, 6600, 6603
Frequency Response Masking 6235, 6237, 6242
Frequent itemset 348
Frequent Pattern 338-341, 348, 409-411
Frequent Pattern Mining 338-340, 348, 410-411
Fresh Fruit Bunches (FFB) 5422, 5424, 5428, 5431
Friction 431-432, 434-437, 439, 4201, 4388, 4645, 4653-4655, 4665, 4851, 5005, 5011, 5013, 5651, 6160, 7306
Friction-Weld 4383, 4388
Fritzing 3277-3278, 3282, 3286
frontal plane 5941, 5945-5948, 5954
Fully Immersive VR 6233
Fully Rational 71-73, 76
Functional Dependencies 191, 852, 860-861, 1923-1924, 1927
functional digital literacy 2286-2287, 2291
Functional Requirement 2942, 7590
Fundamental Ordinance on Local Autonomy 7336
Fundamentalist 72, 76
furans 3068
Fusion of Scores 4941, 4948, 4950
Future Combat Systems 7304-7305, 7307, 7311

Future Internet 6557, 6560, 6565, 7775
 Future Surgical Robots 6878, 6885
 Futures Approaches 910
 fuzzy arithmetic 3217, 3224
 Fuzzy Assessment Methods 3222, 3224
 Fuzzy Cognitive Map 4573-4574, 4580
 Fuzzy Inference System (FIS) 7, 186, 6792, 6805
 Fuzzy Information 4398
 Fuzzy Logic (FL) 4-7, 106, 108-109, 111, 180-183,
 185-186, 204, 208, 493, 496, 498-500, 1814, 1824,
 2094, 2200-2201, 3107, 3215, 3222, 3224, 4799,
 5879, 5885, 6595, 6789-6791, 6793-6794, 6800,
 6802-6803, 6805, 7741-7742
 Fuzzy Membership 186, 6805, 6962
 Fuzzy Number (FN) 3135, 3224-3225
 Fuzzy Set (FS) 108, 110, 185, 589, 1814-1818, 1838,
 1981, 3215-3216, 3224, 3226, 3235-3236, 4398,
 5447, 6790, 6805
 Fuzzy Systems 1, 4-5, 14, 150, 180, 183-186, 498,
 500-501, 1814, 1822, 6790, 6794, 6909
 Fuzzy-evolutionary systems 185-186

G

Gait 133-135, 137-139, 166, 169-170, 298-301, 303,
 307, 313, 2134, 4986, 4992, 4994, 5881
 Galvanically Coupled IBC 6319-6321, 6334
 Gambling 615, 1366, 1845, 2188, 2191-2192, 2197,
 3299-3301, 3308, 4158, 4224-4225, 4318, 4320,
 5460-5461, 5464
 Game Design 2468-2469, 2476, 3241, 3244, 3249,
 3252, 3287-3293, 3295, 3327, 3334, 3338-3341,
 6377-6378, 6382, 6387, 7359, 7363
 Game Design Models 3287-3288, 3295
 Game Development 3287, 3309-3311, 3316, 3328,
 3338-3339, 3341, 3344, 6377, 6382
 Game Development Framework 3309, 3316
 Game Development-Based Learning 3309, 3316
 Game Elements 997, 3238, 3241, 3289, 3292, 3337-
 3339, 3341, 6430
 Game Environment 997, 999, 3341, 3344-3345, 3347,
 4128
 Game Feature 2476, 3336
 Game Play 804, 997, 3242, 3260, 3262, 3338, 3341,
 3345, 3347, 6387, 6489, 7367
 Game Theory 2095, 2106-2108, 3151, 3226-3228,
 3230-3233, 3235-3236, 4853-4857, 5447, 5451,
 6546, 6634-6635, 7283
 Game Tree 105, 144-147, 149-151, 155, 3229
 Game User Research 7367
 Game-Based Learning (GBL) 801, 2240, 2257, 3238,

3240, 3244, 3247, 3249, 3259, 3287, 3289, 3337,
 3339-3341, 3344-3345, 6361-6362, 6364, 6375-
 6377, 6416, 7359, 7367, 7882
 Games and Affection 6377, 6387
 Games And Learning 2523, 3330, 6361-6362, 6387
 Games for Change 3336
 Games For Learning 3287, 3327, 6377, 6379-6380
 Gamification 995, 997-998, 1001, 1476-1477, 2425,
 2430, 2581, 3238, 3240-3244, 3247, 3287, 3330,
 4316, 4321, 4325, 6085, 6376, 6430-6431, 6862,
 7358, 7942
 Gaming 266, 388, 805, 835, 1109, 1167, 1647-1648,
 1651-1652, 1723, 1726, 1849, 2227, 2327, 2329-
 2330, 2423, 2425, 2427, 2515, 2558, 2563, 2567,
 2570, 2851, 3238-3239, 3241, 3262, 3289-3290,
 3296-3300, 3317-3322, 3328-3330, 3333-3334,
 3338, 3340, 3571, 4202, 4223-4225, 4298, 4321,
 4325, 4629, 4658, 5229, 5879, 5881, 5900, 5988,
 6095, 6185-6186, 6367, 6375, 6480, 6614, 6698,
 7030-7031, 7078, 7118-7119, 7122, 7130, 7358-
 7359, 7362-7363, 7563, 7625, 7880
 Garbage Can Model 1017, 4136-4139, 4143-4145
 gardening implements 2626, 2634, 2636-2638
 Gazetteer 6978-6979, 6983
 GCF Computing 4913, 4918
 GDH 1668-1670, 1673-1674
 Gender Empowerment 3400-3401
 Gender Equality 909, 3371-3374, 3376-3378, 3381,
 3400-3401
 Gender Identity 3349, 3354, 3358
 Gender Imbalance 1161, 3394-3396
 Gender Mainstreaming 3372, 3376, 3380
 Gender Oppression 3401
 Gender Sensitive 3400-3401, 4184
 Gender Stereotype 3384, 3401
 Gene Expression 419, 423, 430, 440-443, 445, 447,
 454, 477-480, 483-485, 489, 1826, 6712
 gene ontology 424
 Gene Silencing 477-478, 480, 483-484, 489
 General Intelligence (IQ) 231, 286, 1541, 5897
 General Morphological Analysis 2976, 2979-2980,
 2983
 General Trust 4797, 4799, 4801, 4803, 4805, 4807, 4809
 Generalized Comb Filter 4748, 6180
 generalized operations and relations 1998-1999, 2003,
 2005
 Generational Cohorts 4358, 4361, 4363, 4367
 Genetic Algorithms 1, 4, 6-8, 100, 106, 114, 180-183,
 185, 2134, 2180, 4595, 4600, 5968, 5980, 6595-
 6596, 6608, 6879, 6909
 Genetic Predisposition 3326

Index

- Genome 419-420, 422, 424-425, 430, 443, 453-455, 466-467, 471, 473, 476, 482, 2043, 2045, 2047, 2056, 4021, 5107, 7696
- Genomics 419, 422, 425, 430, 455, 459, 7774
- Geo Spatial Science (GSS) 3501
- Geocoding 3412, 3475, 3479, 6973, 6975, 6978-6979, 6982-6983
- Geo-Cultural Context 4076, 4420
- Geographic Information Science (GIScience) 3463, 3483
- Geographical Information Systems (GIS/Mgis) 1837, 1841, 2047, 3404, 3407, 3417, 3434, 3448-3449, 3459-3461, 3466, 3483-3484, 3501, 3503, 4299
- Geographically Dispersed Team 7042
- Geo-Mapping 4092, 4406, 4408, 4413, 4415, 7049, 7056
- Geospatial Mapping 3475, 3483
- Geospatial Mashups 3403-3404, 3409-3410, 3412, 3414-3415, 3417
- Geo-Stationary 3501
- Geo-Synchronized 3490, 3501
- GFLOPS 234, 4008, 4017
- Ghosting 3280, 3286
- Gilligan's Moral Development Theory 3203
- GIS in the Cloud 3463-3464, 3471
- GIS Mapping 3434, 3436, 3438, 3440-3441, 3443-3444, 3447
- GIS Server 3408-3409, 3472
- GIS trends 3463
- Gist 2297, 4541, 4912-4913, 4919, 6896, 6899-6901, 6904
- GIT 2656, 3407, 3414-3415, 3417, 7376
- Glass Ceiling 3382, 3391
- GLMC 3043-3050, 3052-3053, 3055
- Global Capital 286
- Global Digital Divide 2238, 4620-4621, 7250-7251, 7258
- Global Distribution Systems (GDS) 349-350, 363
- Global Hyperspace Maps 6428, 6434
- Global IDs 5986, 5991
- global induction 2132, 2134, 2138, 2141
- global manager 5328-5329, 5334
- Global Position System (GPS) 6205
- global sharing pedagogy 2642-2643, 2647, 2650
- Global Society 672, 2592, 3695, 4769, 5578, 6908, 6910-6911, 6915, 7224
- global supply chain management 5328-5332, 5334
- global supply chain manager 5327-5328, 5330-5331
- Global Virtual Teams 720, 722, 724-726, 728
- Global Warming 3079, 3103, 3114, 3130, 3371, 3373-3374, 3377, 3380, 3489, 3642, 4698, 5465, 6489
- Glocalization 3022, 3031
- Glycemic Expert Query 2153
- GNU Public License 7911
- Gold Standard 134, 1006, 1680, 2215, 2217, 2223, 5879
- Good Classification Test 200
- Good Formal Concept 197
- Good Irredundant Classification Test (GIRT) 200
- Good Maximally Redundant Classification Test 200
- Google Advanced Search 7017, 7024
- Google App Engine 459, 1147, 3465
- Google Books 5227-5232, 5234
- GoVenture 802-803
- Government as a Service 1033, 1036-1037, 1039
- Government Linked Corporations (GLC) 671, 681
- Government to Citizen (G2C) 3579, 3603
- Government-Private Sector Partnership 7269
- GPA index 3219-3223
- GPGPU 1947, 1961
- GPU 1947, 1952-1953, 1961, 4008, 4017, 4586
- Grade Point Average (GPA) 1540, 3215, 3220, 3224, 7095
- Graded Reading 3953
- Gradient Descent 1761, 1763, 1771, 4531, 5178, 5974, 6664-6665, 6670
- Granloup 6968-6970
- Granular Algebra 7704, 7706, 7709-7712
- Granularity 376, 639, 788, 1357, 1489, 1851, 4019, 4021, 4023, 4029, 4671, 6138, 6352, 6636, 6755
- Granulation 1836-1837, 1839-1841, 1844, 2317, 7736, 7742
- Granule 1836, 1839, 1841, 6963, 6970, 6972, 7709-7710, 7712
- Graph Construction 1981, 7412, 7414-7415, 7418-7419
- Graph Databases 657, 1978, 1981, 2031-2036, 2039, 2041-2042
- Graph Partitioning 2020, 2026-2027, 6961
- graphene 2673-2678, 2680-2681, 2684, 2897-2903, 2906-2907
- Graphical Interface 411, 1291, 1334, 1342, 2036, 2396, 5005
- Graphical Programming Language 1249, 2401
- Graphical Representation 1861-1862, 1868, 1921, 2159, 2631, 3339, 5067, 5069, 5073, 5108, 5114, 6342, 7047-7048, 7936
- Graphics Presentation 8013
- Graphology 4651
- Gray Gap 7027, 7032, 7035
- Grayscale Images 219, 1314, 2133, 4644-4649, 4651, 7834-7835
- Great East Japan Earthquake 7947
- Green Computing 3077, 3082-3083, 4733, 4896, 4912, 7753
- green IT 3077-3085, 3188

green procurement 3086-3090, 3092, 3094, 3096
 green shipping 5467
 Green Supply Chain Management (GSCM) 5357, 5364, 5366, 5378, 5380, 5382, 5386, 5455, 5465, 5467, 5475
 Greenhouse Gasses 5446, 5455
 Grid Computing 1064, 1076-1077, 1102-1104, 1109, 1115, 1147, 4013
 GridOnput 6464-6469, 6471, 6475
 Griffith University 2634, 5853-5857, 5860
 Gross Domestic Product (GDP) 2337, 2727, 3805, 3814, 3816, 4787, 5423, 5527, 5841-5842, 7223-7224, 7260
 Gross Enrolment Ratio 3921
 Gross Merchandise Volume (GMV) 2743, 2747
 Ground War 3612-3614, 3617, 3620
 Grounded Theory 779-780, 787, 1360, 2990, 4481, 6022, 6189
 Group Decision Support Systems 2121-2122, 2131
 Group Dynamics 68, 160, 553, 663-665, 669, 1532, 5117, 5119, 7068
 Group Recommender System 1737, 1745
 group signature scheme 4898-4901, 4903, 4906-4908
 Group synchronization 6435, 6438-6440, 6443
 Groupthink 669, 905, 7165
 Growth Domestic Product (Δ gdpt) 40, 47
 GSM 4680, 4987, 5021, 5465, 6094, 6126-6130, 6135, 6137-6139, 6141, 6144, 6207, 6265, 6274, 6276, 6287, 6289, 6297
 Guglielmino 1469
 Gunes 133, 137, 1746, 1754
 Gunning fog index 1502, 1507
 Gunsmithing 4388
 Gurus 5707, 5834
 GWAS 466-467, 469-474

H

hackathons 3032-3036, 3038-3039, 3041
 Hacker 52, 2565-2567, 2712, 4620, 4837-4841, 4843, 4848, 4906, 4908, 5991, 6246, 6696-6697, 6705, 7078, 7372, 7798
 Hacking 166, 1037, 1366, 1724, 2563, 2565, 2863, 3209, 4247, 4317, 4837-4838, 4840, 4843, 4894, 5311, 6695-6702, 6704-6706, 6708, 6813, 7296, 7987, 8066
 Hadoop 235, 326, 328, 333-334, 406, 410, 457, 460, 862-863, 865, 867-872, 1064, 1358, 1774-1775, 1809, 2137, 5158, 5161, 6908, 7149
 Halal supply chain 5456-5462, 5464
 handmade content 6466-6472, 6475

Haptic Device 4653-4656, 4658, 4665
 Haptic Rendering 4653-4655, 4665
 Haptics-Based System 4652-4653, 4655, 4665
 Haptics-based system architecture 4652-4653
 hash algorithms 1385
 Hash Functions 1385-1387, 1390, 2870, 4927, 4975, 4984
 Hashtag 2071, 2074, 4052, 6186, 7984
 Hathitrust 2479, 5226, 5228-5232, 5234
 HCI-Sec 5005-5006, 5008-5009
 Head Mounted Display (HMD) 4114, 4122, 4235, 6233
 Headcanons 2335
 Health Care 279, 327, 333, 378, 413, 492, 528, 1017-1018, 1022, 1411, 1507, 1673, 1794, 1984-1985, 1988-1989, 1991-1992, 2220, 2508, 3131-3134, 3136-3138, 3261-3262, 3450, 3678, 3693, 3695, 3697-3698, 3701, 3734-3737, 3743, 3747, 3750, 3759-3760, 3776, 3783-3784, 3786-3789, 3793, 3805-3810, 3815-3816, 4817, 5016, 5136, 5794, 5800, 5810-5811, 5813, 5819-5822, 5824, 5852, 5864-5866, 5869-5870, 5875, 5903, 5910, 5917, 5933, 5941, 6083, 6089-6090, 6147-6148, 6160, 6190, 6216, 6229-6230, 6415, 6688, 7035, 7062, 7776, 7849, 7984
 Health care IT 3806
 Health Care Monitoring 3697
 Health Care Organization 3131-3134, 3136-3138, 3698
 Health Education 2495, 2497, 3691, 3734, 5916, 5934, 6415, 6863, 7035
 Health Informatics 3697, 3810, 4727, 4732
 Health Information for Economic and Clinical Health (HITEACH) 3814
 Health Information for Economic and Clinical Health (HITEACH) Act of 2009 3814
 Health Information Seeking (HIS) Behavior 3746-3748, 3755, 7616
 Health Information Systems (HIS) 3750, 5916
 Health information Systems Research (HISR) 3750, 3755
 health information technology 3678-3679, 3681, 3787-3788, 3805-3806, 3816, 3822
 Health Insurance Portability And Accountability Act (HIPA) 1029, 3808, 3815, 6083
 Health Policy 2213, 5852, 5902, 7131
 Health Promotion 3450, 5820-5821, 5823-5825, 5828, 5910, 6940
 Health Providers 995-997, 1002, 7062
 Healthcare Business process 5916
 Healthcare Delivery 1984, 3725, 3732, 3789, 5924, 6806, 6812
 Healthcare Facility 3727, 5279, 5852, 6808

Index

- Healthcare Organization 3698, 3700, 3707, 3710, 5909, 5911
Healthcare Professionals' Satisfaction 3782
Healthcare Reform 3725, 3815, 3817, 5820, 6806
Healthcare Service Strategies 5928
Healthcare Stakeholders 3688, 7765, 7772
Health-IoT 1984
Health-Related Professions 5821, 5828
Health-Related Quality of Life (HRQoL) 528
heart sound features 1011, 1013
Heating Degree Days (HDD) 7815-7819, 7829
Hermeneutic Circle 3347, 3647
Hermeneutics Method 812, 3345, 3347
Heterogeneity 225, 352, 354, 378, 406, 564, 804, 1381, 1560, 2773-2775, 2777-2779, 2836, 2949, 4005, 4009, 4300, 4531, 4794, 4956, 5214-5215, 6096, 6098, 6105, 6130, 6443, 6481, 7238, 7245, 7274, 7322, 7401, 7833, 8024
Heterogeneous agent models 72
Heterogeneous Belief 76
Heterogeneous Data Sources 350, 5613
Heterogeneous Pattern 1383
Heuristic Algorithm 2153, 4600, 6965
Heuristics 72, 101, 120, 150, 231, 239, 421, 423, 555, 629, 643, 1232, 1875, 1925, 1967-1968, 1981, 2076-2077, 2133, 2187, 3747, 4200, 4370-4371, 5418, 5511, 5754, 6033, 6355, 6499-6502, 6505-6508, 6530, 6646, 7697, 7700, 7787, 7791
heutagogical approaches 1469-1470, 1472
heutagogy 1468-1469
hexa-dimension criteria 4884-4885, 4896
Hierarchical Planning Time Horizons 5515
Hierarchical Structure 555, 1351, 1444, 3290, 4214, 4461, 4536, 4666, 4671, 5094, 5196, 5437, 6245, 6248-6249, 6846, 6964-6965, 7438, 7735
Hierarchical Supply Chain Planning 5505, 5515
High Availability (HA) 1041, 1091-1094, 1096, 1099, 1238, 1647, 1652, 1655, 2817, 2822, 3057, 3594, 4928, 4931, 4935, 7775, 8065
High Impact Practices 3933, 3944
High Performance Computing 1064, 1071, 1074, 4000, 4004, 4018-4019
High Speed Packet Access (HSPA) 6096, 6256, 6265
high throughput computing 1064, 1071, 1074
Higher Education (HE) 286, 316, 801, 1130-1131, 1137, 1153, 1163, 1432-1434, 1437, 1439, 1451, 1453, 1455-1456, 1466, 1487, 1492-1493, 1504, 1523, 1544, 1551, 1880-1881, 1883, 1885, 1892, 1894, 1935, 2304-2305, 2383, 2423, 2457, 2503-2504, 2506, 2508-2509, 2514, 2560, 2627-2629, 2642, 2661, 3000, 3173-3174, 3177, 3182, 3238, 3240, 3242, 3244, 3382, 3475, 3827-3834, 3838-3839, 3861-3863, 3866-3867, 3870-3872, 3880-3881, 3909-3912, 3916, 3918, 3920, 3922, 3924, 3926, 3929-3930, 3945-3947, 3952, 3954, 3976, 3982-3983, 4319, 4330, 4405, 4817, 5159, 5197, 5203, 5626, 5820-5821, 5823, 5857, 6188, 6361, 6391, 6403, 6405, 6413, 6415-6416, 6748, 6784-6785, 6870, 7088, 7253, 7623, 7882, 7888, 7892-7894, 7899-7901, 7912-7914
Higher education marketing 3861, 3863, 3866-3867
Higher Learning Commission 3182
Higher Order Learning 2426, 3239, 3976, 5812, 5822
High-Hazard Industries 6858
High-Performance Computing (HPC) 2008, 4004-4005, 4015, 4017-4018, 4021, 4024-4025, 4029
High-Performance Reconfigurable Computing (HPRC) 4018, 4024-4025, 4029
H-Index 6741-6743, 6747
Histogram 212, 1312, 1314, 1862, 1868, 4644-4645, 4943, 4945, 4987, 4989, 5067, 5069-5070, 5073-5074, 6199, 6896, 6901, 7835, 7839, 7841
Historical Ecology 3433
History of Animation 2615
Hobby-ism 6696-6697, 6708
Hofstede's Cultural Dimensions 7957, 8104
Hogenauer 4746, 6043, 6171
Holarchies 4667-4669, 4672, 4676-4677
hollow fiber membrane 3051
Holon 551-552, 557, 561, 4666-4672, 4677
Home Automation 3151, 3274, 3276, 3694, 4216, 6529, 6533
Home UbiHealth 7765, 7767-7772
Homophobia 4163, 4167
horizon scanning 900, 907, 909-910
Horizontal E-Commerce 2741, 2747
Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) 3784, 3793
Hot Shot Business 801, 803
hotel sector 5778, 5780, 5782
HTML 321, 354, 1220, 1512, 2480, 2660, 2753-2754, 3126, 3407-3408, 3412, 3518, 3568, 3854, 3917, 4405, 4918, 4980-4981, 5170, 5215, 5833, 5840, 6352, 7018, 7506, 7628, 7748, 7832, 8008-8010, 8014, 8049, 8087-8088, 8091-8094
Human Arm Prostheses 504
human body antenna 6325-6326, 6328, 6334
Human Capitals 277-278, 281, 286
Human Computer Interaction for Development (HCI4D) 4178, 4181, 4188
Human Factor 873, 1398-1400, 1815, 5004
Human Interpreter 3732, 6812

Human Judgement 160
 human preferences 2188-2189, 2198
 Human Resource Management 743, 2169, 3526, 4337-4338, 4347-4349, 4357, 5329-5330, 5916, 7519
 Human Resources 281, 352, 554, 611, 802, 887-889, 891, 929, 961, 1618, 1880, 1902-1903, 2347, 2548, 2717, 2930, 2933, 2977, 3025, 3189-3190, 3430, 3554, 3587, 3607, 3622-3623, 3967, 3970, 4033, 4282, 4318, 4339, 4362, 4740, 4776, 4823-4824, 4841, 4913, 5062, 5097, 5239, 5293, 5305, 5328, 5356, 5403, 5651, 5676, 5985-5986, 6088, 6155, 6776, 6794, 6908, 7008, 7223, 7229, 7266, 7268, 7673-7674, 8092
 Human Tasks 788, 791-792
 Human Values 2283-2284, 2286, 2288-2289, 3027, 4198, 4277
 human/non-human attachment 7190, 7193, 7195, 7199
 Human-Computer Interaction (HCI) 89, 132, 167, 381-383, 393, 662, 3746, 4101, 4124-4127, 4129-4130, 4134, 4178, 4181, 4188, 4199-4202, 4222, 4283, 4294, 4574, 5004, 5886, 5900, 6025, 6077, 6206, 6225, 6339, 6378, 7599, 7602, 7764, 7778, 7800, 7956
 Human-Machine Interaction 251, 260, 4127, 6476, 6891
 Humanoid Robots 3354, 4127, 4239-4240
 Human-Related Factor 1401, 1405
 Hybrid Cloud 1028, 1032, 1104, 6737
 Hybrid Course 1498
 Hybrid Flow Shop 4369-4372, 4377
 Hybrid intelligence 183-184
 Hybrid IT Infrastructure 886
 Hybrid Learning 412-413, 1487, 2459, 5789, 5799
 Hybrid Models 184, 2201, 6477, 6632-6633, 6636, 7695
 Hybrid Pricing 1169, 6550, 6552, 6554
 Hybrid Recommender Systems 1878, 6033
 Hybrid Refrigerator 1277, 1295
 Hybrid System 5-7, 14, 184, 1294
 Hydrophobic 432, 434, 436-437, 3044-3046, 3051-3052, 3055
 hyper-computations 7731
 Hyperlinks 1516, 2357-2359, 2362, 2368, 2484, 2857, 5223, 5262, 5271, 7429, 8048
 Hyperliteracy 2234
 Hypermedia 1511, 2255, 2357-2364, 2368, 2380-2383, 2391, 2458-2459, 2616, 2624, 3351-3353, 4213, 4359, 6424-6426, 6429-6431, 6433, 6476-6482, 6487, 8006
 Hyper-sensitivity 720, 723-726, 728
 Hyper-Sensitivity Perspective 728

Hypertext 1508, 1510, 1516, 2244, 2368, 2381, 2391, 4214, 4362, 4437, 4547, 5105, 5128, 5135, 5173, 5182, 6424, 6433, 6477-6478, 6487, 7411-7413, 7415, 7421, 8093
 Hypoacoustic 2441
 Hypoglycemic Expert Query 2153
 hypothesis test 1861, 1864, 2215

I

IACBE 3174-3176, 3182
 IBM'S Watson 101, 145, 230, 234-235, 5750-5751, 5756
 IC Design 2907
 ICN 6556-6562
 ICT Adoption 539, 2763, 4182, 7253
 ICT in the Construction Context 550
 ICT Literacy 2226-2227, 7228, 7665-7666, 7670
 ICT Skills 1196-1197, 2226, 2228, 4184, 4794, 7638, 7671
 ICT specialists 4785-4794
 ICT Sphere 4706, 4714
 ICT Usage Statistics 4711, 4714
 ICT4DEV 5025-5026
 ID-AGKA 1668
 IDC 373, 380, 957, 4013
 Ideal Solution 3135, 3143, 4885, 4887, 4895-4896, 5531, 5859, 6266
 IDEF 5434, 5436-5438, 5440-5441
 Identification (With Characters) 4244
 Identification Methods 7488-7489, 7500, 7502
 Identity Based Cryptography 4975, 4984
 Identity Formation 2316, 2321, 3320, 5626, 6941, 7097
 Ideology 455, 1513, 1659, 1661, 3356, 3396, 3561, 3640, 3646-3649, 3653, 3843, 4059, 4404, 4718, 4762, 5109
 IEEE 802.11 1227, 1233-1234, 1236-1237, 5980, 6618-6619, 6628
 IEEE 802.3 1229, 1237
 IL-HUMASS Questionnaire 3872, 3880
 IL-HUMASS survey 3870, 3872
 ILP 1919-1920, 1974-1975, 4008, 4017, 7489
 Image Analysis (IA) 166, 851, 1308, 1318, 1320, 1328, 1330, 1332, 1359-1360, 7831-7834, 7842
 Image Classification 122, 1354, 4092, 5173, 7833-7834, 7839
 Image Engineering (IE) 1308, 1318-1320, 1322, 1325, 1330, 1344, 1354
 Image File 1385, 1391, 1393-1394, 2912, 7905
 Image Filtering 219

Index

- Image Processing (IP) 166, 173, 176, 186, 517, 1308, 1320, 1328, 1330-1333, 1336, 1339, 1343, 2008, 2416, 2911, 2919, 3467, 4202, 4529, 4724, 5749, 6836, 7696
- Image Segmentation 1308, 1310-1312, 1314, 1316, 1318, 1354, 7831, 7833-7834, 7838-7840
- image techniques 1318-1320, 1330, 1354
- Image Understanding (IU) 1320, 1328, 1330, 1344, 1352, 1354
- imbalanced data 1825-1827, 1829-1831
- IME method 2089
- IMELS 3156, 3170
- Imitation Game 106, 251-252, 254-255, 259-260, 263
- Immediate Feedback 997, 2507, 4359, 5116, 5118-5121, 5123, 5257, 5801, 6362, 7936
- Immersive Environments 2591, 7936, 7942
- Immersive MVR 6228, 6233
- immigrant students 2616, 2621
- Impact Factor 3309, 4405, 4409, 4414-4415, 4420, 5246, 6742, 6747
- imperceptibility 4873, 4875, 4880, 4883, 4985, 4997
- implant communication 6319-6323, 6332, 6334
- Impression Management 6943-6944, 6949, 6987, 7073, 8038
- Impulse Response 39, 6007-6008, 6014, 6018, 6045, 6183, 6234-6239, 6241-6242
- Incentive-based Demand Response 3148, 3155
- Incidence and Mortality Rates 3459
- Incivility 3384-3385, 3392
- Inclusive Climate 3386, 3391
- Inclusive Education 2515, 2521-2522, 2524-2528, 2531, 2545, 2552, 7253, 7677
- Inclusive Learning 2545, 2548, 2556, 7673
- Incomplete Information 906, 1444, 1844, 2200, 2211, 2271, 2841, 3595, 3749, 4137, 6941
- Incremental Learning 193-194, 197, 201, 1831
- Incubator Programs 7269
- Index of Polarity 4044-4045, 4051
- Indexing 1787, 1793, 1810, 1950-1951, 2484, 4528, 4536-4538, 4542-4543, 4547, 4727, 4807, 5207, 6212, 6458-6459, 6750, 6771, 6953, 7995-7996, 8049
- Indian Railways 588-589, 600
- Indices of Happiness 288, 290, 297
- indifference zone 1862, 1868
- Indirect Bullying 4254
- indiscernibility relation 1839, 2212
- Indoor Localization 6286-6287, 6289-6291, 6675, 6680
- Indoor/Outdoor Detection 6286-6289
- Induction Loop 316, 324
- Inductive Database 1918
- inductive learning 1920-1921, 1924, 1932
- Inductive Logic Programming 184, 1919-1920, 1974, 1983, 2086
- Indulgence vs. Restraint 7957, 7965
- Industry 4.0 886, 1227, 2774, 5291-5292, 5300, 5311, 5375
- Industry Ecology 5362, 5366
- Inertial Measurement Unit (IMU) 6295
- Inference Algebra 7704, 7710-7712
- Inference Mechanism 1814-1815, 1817, 1824, 4574, 4576
- Influence Limitation 6999, 7001
- Influencing Factors 817, 1397, 2264, 2943, 2947, 2950, 3434, 3440, 3581, 3585
- Infographics 402, 974, 1200, 3832, 5106, 5108, 5114
- Informal Collaboration 3975
- Informal Learning 1517-1521, 1523, 1528, 2246, 2328, 2515, 2548, 2556, 4178, 4181, 4184, 5027, 5035, 6337, 6351-6352, 6403, 7174, 7539-7540, 7543-7545, 7671, 7676-7677, 7680, 7912
- Informatics 421-422, 424-425, 1041, 1622, 1815, 1835, 2394, 2397-2398, 2590, 2598, 2934, 3679, 3697, 3758, 3805-3806, 3810, 4228, 4282-4283, 4404, 4437, 4452, 4455, 4482, 4725, 4727, 4732, 5106, 5206, 5889, 6459, 6687-6688, 6690, 6694, 7051, 7202, 7205-7206, 7269, 7693, 7704-7707, 7710-7711, 7774
- Informatics Projects 7269
- information abilities 3875-3876, 3878
- Information And Communication Technology For Development (ICT4DEV) 5025, 7223
- Information Architecture 574-575, 579-580, 583-584, 2920, 2927, 8081
- Information Asymmetry 63, 71, 535, 2793, 4856-4857, 6916-6918, 6926
- Information Behavior 2268, 2275, 2279, 3746, 3755, 4392, 4515, 4522, 4527
- Information Centric Networking 6556, 6558, 6560, 6562, 6565
- Information Communication and Dissemination 3872, 3880
- Information competence 4465-4466
- Information Content 422, 1192, 1787-1788, 1973, 2269, 2276, 4101, 4191, 4222, 4391-4393, 4395-4398, 4423, 5098, 6137
- Information Co-Ownership 6992
- Information Coping 6690, 6694
- Information Delivery 849, 1434, 2588, 4542, 5261, 5549, 5823, 5909, 6388, 6558
- Information Disclosure 756-757, 4865
- Information Ecology 4178, 4180, 4184, 4188

- Information Evaluation 2247-2248, 3872, 3880
- Information Flow 422, 660, 891, 2160-2161, 2804, 2948, 2962, 3146, 3289-3290, 4306, 4308, 4313, 4391-4393, 4395-4398, 4462, 4469, 4829, 5277, 5305-5306, 5308-5310, 5345, 5349, 5367, 5490, 5570, 5908, 7110, 7146, 7162, 7874
- Information Highway 7233
- Information Industry Statistics 4714
- Information Literacy 1423, 2225-2227, 2234, 2243-2244, 2246, 2255-2256, 2273, 2279, 2281, 2292-2295, 2297-2299, 2302, 2305, 2308, 2310-2311, 2313, 2325, 3310, 3870-3872, 3874, 3878, 3880, 4525, 4527, 4963, 6415, 6688, 7206, 7541-7542, 7548
- Information Management 235, 387, 601-603, 608, 875, 1065, 1074, 1129, 2560, 2717, 3530, 3811, 3881, 4310, 4458, 4723, 4727, 4735, 5108, 5317, 5911, 6476, 6689, 7206, 7875
- information meaning 2265
- Information Models 389
- Information Needs 349, 914, 2264, 2268-2271, 2274-2276, 2664, 3683, 3747-3749, 3751, 3755, 3863, 3872, 4065, 4199, 4490, 4492, 4496, 4515-4516, 4525-4526, 4574, 5079, 5255, 5666, 5870, 6752, 7162
- Information permeability 4465
- Information Power 1064, 4391, 4393, 4395-4396, 4398
- Information Privacy 387, 957, 4862-4868, 4871, 6988
- information privacy concerns 4862-4866, 4868, 4871
- Information Processing 112, 155, 384, 682, 692, 757, 778, 886, 1049, 1595, 2076-2077, 2079-2080, 2125, 2231, 2244, 2255, 2329, 2444, 2581, 2652, 3297, 3450, 3746, 3872, 3880, 4124, 4195, 4214, 4236, 4423, 4425, 4448, 4451, 4453, 4459, 4743, 4849, 5308, 5864, 6355, 6400, 6403, 6405, 6478, 6524, 6580, 6710, 6790, 7307, 7311, 7696, 7709, 7715, 7760, 7771, 7799, 7969, 7974
- Information Recovery 6976-6977, 6981
- Information Retrieval 193, 235, 851, 1446, 1785-1788, 1793, 1898, 1966, 2276, 3746, 3748-3750, 4443, 4471, 4476, 4479, 4490-4491, 4495-4498, 4501, 4525, 4528, 4534-4543, 4547, 5228, 5231, 5233, 5727, 6033, 6424, 6556-6557, 6752, 6956, 7049, 7206, 7602-7604, 7996
- Information Retrieval Systems 1786-1787, 1898, 3748, 4479, 4497-4498, 4535, 4537-4541, 4543, 6033, 7206
- Information School 4451-4452, 4458, 4734
- Information Science 338, 345, 419, 425, 662, 961, 967, 1064, 1187-1188, 1319, 1330, 1407-1408, 1411-1412, 1415, 1642, 1676, 1845, 2264, 2268, 2439, 2503, 3274, 3377, 3463, 3465, 3473, 3477, 3483, 3558, 3746, 4064, 4066, 4105, 4267, 4269, 4280, 4391, 4397, 4422-4423, 4425-4426, 4431, 4448-4455, 4458, 4460, 4465, 4516, 4537-4538, 4594, 4599, 4602, 4666, 4672, 4723-4733, 4735, 4822, 4828, 4922, 5203, 5206, 5226, 5254, 5257, 5262, 5411, 5700, 6089-6090, 6189, 6686, 6690, 6739-6740, 6750, 6971, 7051, 7057, 7201, 7204-7209, 7211, 7602, 7605, 7704, 7916, 7969
- Information Search 2882, 3861, 3872, 3880, 4064, 4211, 4491, 4516, 4519-4520, 4525-4526, 4535, 8047
- Information Security 51, 60, 373-374, 379, 535, 1041, 1107, 1359, 1361, 1678, 1793, 2561, 2908, 3695, 4689, 4730, 4840, 4848, 4851, 4853-4855, 4860, 4884-4885, 4909, 4911-4913, 4916, 4997, 5481, 5552, 5985-5986, 6092, 6731, 7528, 7724, 7785
- Information Seeking 2226, 2243, 2264, 2268, 2272-2274, 2276, 2279, 3745-3750, 3755, 4213, 4490-4492, 4494-4498, 4515-4520, 4522-4527, 4535, 5105, 5197, 5266, 6690, 6694
- Information Seeking Behavior 2272-2274, 2276, 2279, 3745, 3748, 4213, 4492, 4495, 4497, 4515-4518, 4520, 4522-4527, 5197
- Information seeking model 4517
- Information Services 834, 2480, 2485, 2752, 3034, 3187, 4210, 4212, 4217-4218, 4453, 4458, 4735, 4862, 5757, 5763, 5856, 6186, 7211, 7807
- Information Society 1319, 1421, 1985, 2244, 2246, 2253, 2266, 2268, 2273, 2397-2398, 2481, 3550, 3594, 3870, 3872, 4194-4196, 4198, 4276, 4459, 4466, 4621, 4680, 4704-4707, 4710-4712, 4714, 4785, 4952, 4961, 5019, 5078, 5234, 6907-6910, 6913-6915, 7224-7230, 7233, 7260, 7262, 7303, 7754-7755, 7762, 7917
- Information Society Statistics 4704-4705, 4712, 4714
- Information System 50, 60, 103, 107, 315, 317-319, 321, 323-324, 354-355, 360, 363, 545, 577, 582, 630, 788, 823, 825, 827, 831-832, 898, 1077, 1376, 1835, 1841, 1844, 1935-1936, 1938-1939, 2020, 2203-2204, 2208, 2211-2212, 2268, 2270, 2335, 2570, 2792-2793, 2804, 2806, 2937-2938, 2942, 2962, 2965, 2975, 3088, 3092, 3274, 3277, 3403-3404, 3407, 3414, 3417, 3434, 3448-3449, 3455, 3459, 3461, 3466, 3485, 3590, 3750, 3755, 3782, 3817, 3819-3820, 3883, 3899, 3912, 3918, 4391-4392, 4422, 4448, 4451-4453, 4498, 4810, 4828, 4850, 4852-4853, 4856, 4858, 4860-4861, 4927, 5279, 5287, 5305, 5311, 5403-5405, 5434, 5436, 5613, 5646, 5709, 5916, 6678, 7447, 7855, 7873, 7878

Index

- Information System Security (ISS) 2570, 4861
Information Systems Success 49, 52, 60, 4281
Information Systems Success Model 49, 60, 4281
Information Technology (IT) Projects 777-781, 783-784, 787, 3756, 3759, 3783, 5698
Information Technology Governance 810
Information Technology Industry 4688, 7261, 7269, 8096
Information Technology Outsourcing 5483, 7268
Information Technology Revolution 2746, 4138, 4448-4449, 7234, 7980, 7982, 7985, 7993
Information Technology Sector 3183
Information Technology Solutions 1405, 5559-5562, 5565, 5569
Information Technology Startups 7268
Information Theory 157, 159, 1186-1188, 1191, 1793, 1973, 2121-2122, 2125, 2131, 4391-4392, 4398, 4423, 4426, 4448, 4451-4452, 4455
Information Visualization 1788, 3748, 4262, 5105-5107, 5109, 5114, 7414, 7421
Informational Literacy 1511, 1516
Information-Based Revolution 7302, 7311
Information-Theoretic 1787, 1793
Informativity 4465, 4469
Informatization 3918, 3921, 7775
Infrastructure As A Service (IaaS) 887, 1027, 1102, 1104, 1129, 1148, 1184, 1356, 1365, 1627, 1636, 6730, 7532, 7767, 7778
Initial Values 1302, 4606
Inland Revenue Board 841
In-Memory 1806-1811, 1813, 1949, 1952, 1955-1956, 5543, 5546, 7996
In-Memory Computing 1808, 1813
Inner Circle 1218, 1220, 1225, 1964, 5046, 8017
Innovation in Higher Education 3827, 3838
Innovation Process 708, 713, 2345, 4350, 5841, 7150, 7160, 7224, 7371
Inorganic 3156, 3160-3161, 3358
Input-Output Identification 7490-7491, 7502
inquiry learning 2642-2643, 2647, 2650
In-Service Training 7635, 7645
Insolvency 2337-2338, 6633, 6813-6816, 6818, 6820, 6823
Installation Art 4692-4693, 4695, 4697-4698, 4700, 4702
instance explanation 2088-2090
Instance Level Explanation 2094
Institution Based Trust (IBT) 3658-3659, 3661-3662, 3666
Institutional Repository 2483, 4438, 4447, 5229
Instructional Design 1420-1421, 1425, 1432-1433, 1436, 1455, 1474, 1486, 1560, 1562, 2361, 2382, 2444, 2664, 3291-3292, 3933, 3939, 4328, 5152, 5821, 5824, 6377, 6380, 6402, 7924-7925, 7929-7930, 7936
Instructional Facilitator 5150-5154, 5156-5157, 7922, 7927-7929
Instructional Preparation 2663-2664, 2667, 2671
Integrated Circuits 2684, 2897, 2899, 2901, 2907, 3078, 6581-6582, 6827
Integrated Development Environment (IDE) 3273, 3275, 3316, 6057, 6061, 6069, 7398
Integrated Electronic IQA 3881
Integrated Health Systems 5869, 5875
Integrated IT System 3921
Integrated Performance Assessment (IPA) 7683-7684, 7689, 7691
Integrated Quality Management System (IQMS) 3899
Integrated Shopping 2845, 2849
Intellectual Capital 943-945, 947, 950, 3023, 3031, 4349, 4776, 4778, 5056-5058, 5060-5062, 5064, 5066, 5078-5080, 5084, 5669-5671, 5674-5676, 5761, 7064, 7261
Intellectual Property (IP) 377, 425, 613, 943, 1028, 1157, 1656-1657, 1661-1662, 1665, 1684, 2318, 2477, 2479-2483, 2486-2488, 2490-2491, 3186, 3594, 4267, 4306-4307, 4309, 4313-4314, 4383, 4679, 4691, 4776, 4840, 4997, 5010, 5079, 5552, 5558, 5573, 5714, 5716, 5744, 6252, 6458, 6749, 6755, 6915, 7264, 7296, 7370, 7799-7800, 7982, 7988, 7994, 8096
Intellectual Property Rights (IPR) 2481, 2483, 2486, 2490-2491, 3186, 4267, 4679, 4691, 4997, 6755, 7264, 7800, 7988
Intelligent Agent (IA) 89, 5749, 5756
Intelligent Banknote Neutralization System (IBNS) 81, 86
Intelligent Electronic Device (IED) 3065
Intelligent Systems 94, 98, 100, 180-185, 187, 221, 425, 1875, 2003, 2812, 2934, 4269, 4736, 5225, 5291, 6478, 6481, 6580, 7699, 7704, 7708, 7710, 7712
Intelligent Techniques 183-187, 2810, 2813-2814, 2818, 2822, 2824
Intelligent Transportation Systems (ITS) 6675-6677, 6682
Intensification 3354, 3421, 4193, 6918, 7693, 7700, 7703
Intensive Reading 3946-3947, 3951, 3953
Interaction Design 389, 3287, 4134, 4199-4201, 5887, 6225, 7401-7402, 7599, 7608, 7764

- Interaction Theory 3817-3818, 3825
- Interactive Communication 2663-2664, 2668, 3851, 4091, 6489, 7006, 7127
- Interactive Learning 1512-1513, 2460, 2559, 3243, 3851, 4360, 5126, 5829-5830, 5835, 5837, 5840, 5862, 7916
- Interactive Learning Environment 3851, 4360, 5840
- Interactive Multimedia (IMM) 2357, 2359, 2368, 4218, 5789, 5830, 6270, 6557
- Interactive Technologies 345, 389, 7097-7098, 7100-7101, 8036
- Interactive Walkthrough 389, 393
- Interactivity 89, 263, 388-389, 715, 719, 805, 980, 1197, 1510, 1883, 2242, 2359, 2442, 2444-2445, 2484, 2487, 2525, 2546, 2591, 2657-2659, 3289, 3407, 3571, 3578, 3603, 3668, 4006, 4052-4054, 4102, 4113, 4122, 4179, 4210, 4213-4214, 4697, 5110, 5116, 5118, 5223, 5401, 5793, 5801, 5829-5832, 5834-5837, 5840, 5909, 5914, 5959, 6074, 6165, 6189, 6230, 6348, 6441-6442, 6477-6478, 6488, 6993, 7001, 7313, 7672, 7882, 7889, 8036
- Interconnects 1228, 2673, 2899, 2901, 4011-4012
- Inter-Dependencies 5584, 7589-7590, 7592-7593
- inter-destination media synchronization 6435, 6438
- Interdisciplinarity 2493, 2495, 2501, 2548, 6869
- Interdisciplinary 90, 94, 382, 455, 522-524, 551, 805, 910, 1869, 2433, 2493, 2497, 2548-2550, 2552, 2640, 3287, 3290, 3374, 3641, 3759, 3798, 3839-3840, 3847, 3922, 4064, 4189-4190, 4278, 4282-4284, 4287, 4294-4295, 4310, 4318, 4422, 4449, 4451-4452, 4455, 4458-4459, 4467, 4476, 4537, 4689, 4723, 4726, 4728, 4735, 4775, 4780, 4920, 5111, 5546, 5676, 5889, 6150, 6166-6167, 6225, 6377, 6387, 6405, 6481, 6546, 6750, 6753, 6759, 6821, 7046, 7057, 7059, 7063, 7068, 7201-7202, 7204, 7211, 7345, 7643, 7675-7677, 7764, 7801, 7856, 7880-7881, 7883-7884, 7887, 7894, 7990
- Interdisciplinary Research 94, 455, 524, 2433, 3641, 3922, 4295, 4723, 5676, 6166, 6546, 7643
- Interdisciplinary Studies 3798, 7801, 7880-7881, 7884, 7887
- Interface Design 985, 1335, 2857, 3669, 3748, 4155, 4181, 4188, 4213, 4217-4218, 4222, 5009, 5886, 6351, 7070, 7073-7074, 7402, 7956-7959, 7962, 8079
- inter-media synchronization 6435-6436, 6438, 6443
- internal integrity 5288, 5293, 5295, 5305
- Internal Processes 1617, 1687, 2792, 3591, 3881, 3915-3916, 5288, 5309, 5444, 5543, 5560, 5569, 5595, 7499
- Internal Quality Assurance (IQA) 3881, 3887
- International Classification of Diseases 4233
- International Data Privacy Principles 4909, 4911, 4916, 4918
- International Nemetics Institute 6490, 6492, 6495
- International Statistical Standards 4705, 4714
- International Statistics 2315, 4714
- International Students 3828, 3900-3906, 3909, 7893
- Internet Addiction 4223-4228, 4233, 4321, 4325
- Internet Addiction Treatment 4227, 4233
- Internet Adoption Rate 8096, 8104
- Internet Affinity 7071, 7076
- Internet Archive 5226, 5228-5229, 5234
- Internet Banking 49-53, 55-57, 60, 2731, 2763, 3656-3657, 5708
- Internet Community 4052, 4191, 6478, 7211
- Internet crimes against children 1366-1367, 1370-1371
- Internet Literacy Assessment Indicator for Students (ILAS) 4962, 4964, 4966, 4971, 4973
- Internet Meme 8015, 8017-8018, 8020, 8022
- Internet of Medical Things (IoMT) 1994
- Internet Of Things (Iot) 333, 461, 876, 886-887, 995, 1002, 1984-1986, 1989, 2775, 2810, 2815, 2822, 2824, 2863, 2865, 2868, 3077, 3082-3083, 3274-3275, 3286, 3689-3691, 3695, 3697, 4004, 4013, 4017, 4203, 4210, 4216, 4262, 4310, 4314, 4689, 4710-4711, 4884, 4891, 4894, 4896, 4909, 4951-4954, 4956-4957, 5291, 5307, 5311, 5575, 5581, 5911, 5914, 6120, 6200, 6908, 7122, 7124, 7225, 7227, 7230, 7233, 7392, 7397, 7525-7526, 7754, 7761-7762, 7767, 7774-7777, 7784, 7984
- Internet Payments 4899, 5397
- Internet Phenomenon 8015, 8018-8019, 8022
- Internet piracy 1657-1658, 1665
- Internet Technology 936, 1101, 1375, 2254, 2687, 2690, 2856, 5345, 5367, 5397, 6557, 7068, 7125, 7161, 7525, 8015
- internet-based business ecosystem 2743, 2747
- Internet-Facilitated Child Sexual Exploitation 1366, 1375
- Internetization 7980, 7983-7991, 7993
- Interoperability Framework 644, 2779, 4438, 6566-6567, 6569, 6573-6575, 6578, 7760
- interpersonal behavior 5632-5633
- Interpersonal Conflict 720, 723-726, 728
- Interpersonal Skills 2967, 2974, 2993, 6687
- Interpretivism 811-812, 3342, 3345, 3347
- Intertext 6497
- Intra-Body Communication (IBC) 6319, 6321, 6334
- intra-media synchronization 6435-6436, 6440
- Intrinsic Motivation 2666, 3265, 3321, 3326, 3945, 3947, 3951, 5095, 5102, 7359, 7516, 7881

Index

- Introductory to Programming 7398
Introspection 1532-1533, 2085, 4736, 5813, 5823
Intrusion Tolerance 3056, 4927-4928, 4930-4932, 4936, 8056, 8062
Inventory Management 1400, 2933, 2977, 3700, 5280, 5335-5337, 5340-5341, 5344, 5372, 5375, 5519-5520, 7873-7874
Inverse Distribution Function 1304, 1306, 1858, 1868
inverted files 7995-7996
Inverted List 7995-7997, 7999, 8003, 8005
Investigative Site Personnel 7182-7183, 7186-7187, 7189
Investor sentiment 72
IOPS 4010, 4017
Ios 268-269, 276, 3475, 5373, 5887, 6042, 6057, 6069, 6154, 6202, 6207, 6210, 6213, 6302, 7371-7377, 7381, 7545, 7548, 7638, 7640, 8037
IP Address 3283, 4017, 6086, 6556, 6610, 6612, 6616, 7294, 7301
IP theft 1657-1659, 1661, 1665
Ipad 268-269, 3747, 5027, 5756, 5896, 5983-5985, 6069, 6120, 6467, 6469-6470, 7375-7377, 7381, 7626, 7637
IPPC 3156-3166, 3168-3170
IR Camera (IR-Cam) 6206
IS failures 5660-5661, 5664, 5666
Is Project 5317-5319, 5321, 5324-5325, 5651, 5660, 5700
IS Success Model 50-51, 55, 57, 6025
IS/IT investments 3757
Is-A Relationship 2027, 2030, 7473
is-aC Relationship 2030
is-ao Relationship 2030
iSchools 4723, 4727
Islamic Banking 37-45, 47
Islamic financial system 5461
Islamic Financing 36, 39, 44, 5459, 5461-5462, 5464
Islamic Interbank Investment Rate (IIBRT) 47
ISO Standards 6458, 6933
ISO/IEC 811, 2930, 2935, 3521, 3529, 3533-3534, 3793, 4685-4686, 5007-5008, 5013, 5436, 5440, 6458-6459, 6566, 6569-6570, 6927-6930, 6932-6935, 7557-7558, 7562, 7565, 7584
ISO/IEC 25010 2935, 3793, 5007-5008, 5013
Isomorphism of Polynomials Problem 4901, 4908
ISPRA 3156, 3159-3160, 3170
Issuing Bank 2708, 2710, 2722
IST 90, 94, 3382, 4267, 4269, 4273-4274, 4452, 4723-4733, 7205-7207
IT Applications 882, 891, 917, 2307, 3188, 3759, 3810, 4629, 5484, 5570, 5576, 5578, 5982
IT Architecture 575-577, 579, 810-811, 2970, 3607-3608, 4828-4829, 5985
IT Artifact 4279, 4286, 4295
IT for Socioeconomic Development 7233
IT Governance 601-602, 604-605, 608, 810-813, 815-817, 819, 886, 2922, 2969
IT Infrastructure 378, 407, 534, 810, 875, 879-880, 886, 926, 1147, 1159, 1629, 2341, 2560, 2954, 3088, 4822, 4827-4830, 4835, 5158, 5297, 5520, 5902-5903, 6730
IT Innovation Management 604
IT Investments 923, 2342, 3757, 5483
IT organization 873-874, 878, 883
IT Professionals 777-778, 784, 1637, 2969, 3083, 4886, 4912, 5204, 5476, 5478-5479, 5924, 5985
IT Revolution 7993
IT Security 1154, 1157, 1163, 2561, 5856
IT Service 880, 1597, 2920-2925, 2927-2928, 2930, 5477-5478, 7238
IT Service and Product Portfolio 880
IT Service Catalogue 2924-2925, 2930
IT Service Management 2920-2921, 2925, 2927, 2930
IT sourcing strategy 880-881
IT Strategy 811, 813-814, 873-879, 882-883, 2965, 2969, 3014, 5714
IT transfer 7233
IT Transfer to Developing Nations 7233
IT User Resistance 3823, 3825
IT value chain 3185-3187, 3191, 3193
IT4IT 2928
item model 2370-2371, 2373, 2375, 2378
item-based collaborative filtering 1757, 1760, 1762, 1766-1768, 1771
Itemset 348
item-to-item collaborative filtering 1745, 1872-1873, 1878
Iterative Decimal Multiplication 4618
ITIL 603-604, 608, 811, 819, 2156, 2920-2924, 2927-2928, 2930, 3529
- ## J
- J. Peter Kincaid 1499-1500, 1507
Janssen 133, 892, 2178, 2943, 3032, 3034-3035, 3604, 7846
Java 411, 414, 1181, 1238, 1332, 1334-1335, 1339, 1621-1622, 1969, 2057, 2065, 2753, 2778, 3252, 4608, 5106, 5553, 5555-5556, 6037, 6057, 6061, 6065-6066, 6068, 6541, 6956, 7372-7373, 7448, 7454, 7459-7460, 7464, 7470, 7472, 7475-7476, 7478, 7480, 7507-7508, 7566-7567, 7573, 7616, 7699

Javascript 2395, 5834, 6993, 7506, 7748, 7882, 8087-8089, 8091-8092, 8094
 Jet Lag 3453-3454, 3459
 Jingping Xu 1007, 1010
 Jitter 1230-1231, 3060-3062, 3065, 6116, 6267, 6435-6436, 6445, 6548, 6609-6610, 6613, 6616
 Job Polarization 4786, 4796
 John Traxler 6377, 6399
 Jörmungandr 7342-7344, 7346
 Joseph A. Schumpeter 7994
 Jump Ahead (Back) v Steps 1306
 Justification 952, 1411, 2574, 2652, 3062, 3173, 3206, 3208, 3210, 3255, 3637, 3950, 4307, 4431, 4596, 6073, 6163, 6713, 6927, 7108, 7295, 7791-7792, 7908, 7970, 7972

K

Kalman Filter 212, 214, 218-219
 Kalman Gain Matrix 219
 Kano Customer Needs Model 732, 739
 Kaprow, Allan 4702
 KDD Process 1907-1908, 1911, 1918-1919, 1921-1922, 1932, 1988, 1992, 6651
 KDM 7448-7454
 Kedai Rakyat 1 Malaysia 5536
 key features cognitive model 2370-2371, 2373, 2378
 Key Performance Indicators (KPI) 357, 359, 581, 891, 1882, 2169, 2854, 2955, 3885, 4937, 5540, 5542, 5544, 5549, 5638
 Kickstarter 3273, 3277, 3286
 Kinect 133-135, 138-139, 5876-5882, 5885, 5941-5950, 5954, 7360, 7760
 Kinect SDK 5885, 5942, 5947, 5950, 5954
 Kinematic Design 6882-6883, 6885, 6888-6889, 6891
 Kinematics 300, 498, 524, 1078, 5878, 6875, 6882-6886, 6888, 6890-6891, 6893
 Kinesthetic Sense 4652-4653, 4665
 Klout 7143, 7149, 7798
 Klout Score 7143, 7149
 Knowcosters 1644
 Knowledge Application 351, 354, 4572-4573, 5089
 Knowledge Artifacts 2779, 7168
 Knowledge Base 102, 105, 231, 281, 1442, 1450, 1474, 1482, 1788, 2003, 2450, 2561, 2593, 3244, 3839, 4331, 4539, 4541, 4547, 4575, 4741, 4811, 5149, 5151, 5401, 5403, 5410, 5706, 5801, 7059, 7154-7155, 7162-7163, 7659-7661, 7664-7665, 7924, 7930
 Knowledge Building 2481, 2663-2665, 2671, 3570, 6336-6337, 6488, 7015, 7169, 7969
 Knowledge Communities 3922, 4283, 7037, 7042, 7163
 Knowledge Construction 1422, 1427, 1454, 2492, 2502, 2506, 2546, 2642, 2650, 3250, 3978, 6362, 6780, 6859, 6870
 Knowledge Creation 359-360, 363, 1149, 2225, 2305, 2641, 3922-3927, 3929, 3970, 4349, 4357, 4436, 4574-4576, 4580, 5027, 5046-5047, 5086, 5088, 5093-5094
 Knowledge Discovery 182, 186, 235, 338, 348, 572, 1783, 1875, 1896, 1907, 1919-1921, 1932, 1988, 2085, 2091, 2132, 2201, 3760, 5549, 6642, 7448, 7451
 Knowledge discovery in datasets 6642
 Knowledge Economy (KE) 1065, 2253, 2284, 2291, 4350-4351, 5017, 5026, 5033, 5669, 5676, 7180, 7988
 Knowledge Extraction 461, 1919-1921, 6651, 6657, 7800
 Knowledge Harvesting 5709, 5711, 5713
 Knowledge Management (KM) 184, 349-350, 554, 943-945, 947, 950, 961-962, 967, 970, 1443, 1517, 1523, 1596-1597, 2008, 2020, 2483-2484, 2488, 2688, 2694, 2793, 2933, 2992, 3527, 3531, 3746, 4073, 4089, 4282-4283, 4347-4348, 4350, 4357, 4427, 4571, 4723-4724, 4729, 4732, 5027, 5046-5050, 5052, 5054-5055, 5077-5082, 5084-5090, 5092, 5095, 5265, 5410, 5670, 5706-5707, 5911, 5913, 5917, 6412, 6956, 7042, 7131, 7772, 8038-8040
 Knowledge Management for Development (KM4D) 5077, 5079, 5084
 Knowledge Management Software 5055
 Knowledge Management System 2484, 5049, 5055
 Knowledge Networks 2593, 3922, 3924-3926, 3929, 4829
 Knowledge Performance Indicators 1892, 1894
 Knowledge Representation 99-101, 105-106, 108, 183, 1442-1443, 1975, 2200, 4736-4737, 4740, 4742-4743, 4745, 6341, 6426, 7541, 7706, 7711, 7774
 knowledge retrieval 1378
 Knowledge Sharing 239, 756, 765, 812, 913, 917, 936, 1158, 1520, 2548, 2773, 3969, 4090, 4278, 4282-4283, 4351, 4822, 4829, 5047-5048, 5079, 5086, 5089, 5093-5099, 5102, 5261, 5266, 5529, 5706, 5869, 5912, 7036-7040, 7045, 7080, 7090, 7092, 7161-7165, 7774, 7912, 8038, 8040
 Knowledge Society (KS) 806, 1518, 1521, 2283-2289, 2291, 2293, 3066, 4436, 5015-5016, 5019, 5025-5026, 5634, 7223, 7677
 Knowledge Storage 4573
 Knowledge Transfer 998, 1397, 1401-1402, 1741, 2467,

Index

2981, 3621, 3831, 3963, 3965, 4245, 4574, 4580,
5047, 5098, 5103, 7163, 7359
knowledge translation 5811, 5813, 5819
knowledge visualization 5103, 5105-5107, 5110-5111,
5114
Knowledge Workers 741, 743, 788, 943, 2560-2561,
2568, 2570, 4362, 5198, 5440
Knowledge-Based Economy 3930, 5046, 5056, 5066,
6776
Knowledge-Based View 5715, 5725
Knowles 241, 1464-1465, 1518, 1520, 3822, 3830,
4328, 7673, 7898
Kohlberg's Moral Development Theory 3203
Kohonen 122-123, 442, 6581
Kuhn-Tucker 4595, 4606
KW 4016-4017, 7749

L

Labview 1249-1253, 1255-1257, 1259, 5492
Lagrange 499, 4594-4598, 4606
Lagrange Multiplier 4595-4596, 4598, 4606
Land Cover Change 3489-3490, 3497-3498, 3501
Landmark 979, 2121, 4032, 4116-4117, 5479, 5835,
5879, 6279, 6287, 6295, 6839, 6897-6899, 7682
LANDSAT 3491-3493, 3495-3496, 3498, 3507-3508
Landscape Archaeology 3433
Landslide Susceptibility 3503-3504, 3508-3512, 3514
language contact 1206, 1213, 1215
Language Extended Lexicon 5128, 5135, 7411-7413,
7421
Language Motivation 3953
Language Translation 3732, 4496, 6812
Large Datasets 406-407, 409-410, 456-457, 1774,
1894, 1963, 5519, 6649
Large Organizations 281, 1102, 2792, 3401, 4492, 6928
Latency 1119, 1123-1124, 1128, 1227-1229, 1232,
1237, 1807, 1809-1810, 1948, 1951, 2059, 2774,
2783, 3996, 4004, 4009-4013, 4549-4552, 4555,
4614, 6095, 6109, 6199, 6269, 6309, 6313, 6315,
6349, 6533, 6558, 6561, 7564, 7575-7576, 7796,
8061, 8087, 8090
latent factor model-based collaborative filtering 1745,
1873, 1878
Latent Query Features 4533-4534
Latent Semantic Analysis 1783, 1786, 1793, 5172-
5173, 5175-5177, 5179, 5182
Latour 7191, 7193, 7312-7315, 7317, 7322
Law Enforcement 166, 1358, 1366-1367, 1370, 1378,
1656, 1697, 2238, 2717, 3646, 4114, 4385-4386,
4844, 6276, 6971

Layer Model of Information Management 603, 608
Layers of Interoperability 6578
LCGFT 6750-6751, 6755
LCMPT 6752, 6755
LCSH 6750-6752, 6755
Lean Six Sigma 729-730, 732-733, 735, 739
learner characteristics 2591
Learner Experience 6351-6352, 7602
Learner-Centered Instruction 1433, 3954
Learning Analytics 1498, 2644, 2647, 5158-5164,
5168-5169, 6342-6343, 6415-6416, 6430, 7897
learning analytics processor 5164
Learning Design 3239, 3291, 3293, 4178-4179, 4182-
4184, 4188, 6339, 6341-6343, 6346, 6399, 7254
learning gain 2467, 2470-2471, 2473
Learning Games 802, 3328, 3330-3333, 3336, 5837,
6377
Learning Management System (LMS) 802, 805, 808,
1441, 1490, 1528, 1558, 1676, 2382, 2574, 2579,
3828, 3851, 3854, 3902, 3958, 3960, 5159, 5161,
7174, 7626, 7630, 7643, 7645, 7897, 7915, 7921
Learning Organization 2689, 4357
learning personalization 1560, 1563, 1568
Learning Segment 7686-7687, 7691
Learning Style 1519, 1548, 1556, 1558, 2357, 2425,
2538, 3935, 4360, 5801
Learning Theories 1476, 2382, 2546, 2560, 2588,
3250-3251, 3341, 3344-3345, 6336, 6339-6340,
6342, 6344, 6362, 6367, 6863, 7672, 7935-7936,
7938-7939
Learning Through Designing 3248, 3250, 3252, 3255,
3259
Legal-Rationality 3548
Legislation 277, 281, 318, 610, 638-639, 641-643,
645-646, 741-742, 761, 815-816, 830, 2098,
2246, 2346, 2477, 2508-2509, 2791, 2849, 3034,
3066-3067, 3070-3072, 3081, 3084, 3087, 3134,
3160, 3165, 3189, 3372, 3374, 3376, 3381, 3912,
4679, 4862, 4864, 4866, 4868, 4958, 5037, 5381,
5428, 5479, 6699, 6705, 6951, 7271, 7274-7275,
7294, 7682-7683, 7689, 7757, 7988
Legitimacy 1020, 2327, 2329, 2331, 2866, 3540, 3542,
3548, 3581, 3646-3647, 3967, 4761, 4763, 5039,
6698, 7293, 7517
Legitimation Crisis 3539-3540, 3542-3543, 3548
Lego Mindstorms NXT 6863, 6865, 6870
Lego Wedo 2497, 6863-6864, 6870
LEL Graph 7412, 7415, 7421
Lenddo 7144-7145
lessons-learned 5698

- Level Of Abstraction 388, 573, 586, 632, 1345, 1551, 2393, 6567, 7438-7439, 7447, 7862
- Level Of Engagement 720, 722, 728, 1000, 1468, 1491, 1626, 3337, 3563, 4156, 5031, 8071
- Lexical Analysis 4101-4103, 4112
- LGPL 2656, 2660, 7906
- Library And Information Science 1330, 4105, 4516, 4538, 4726, 5206, 5254, 5257, 5262, 6686, 6690, 7916
- library education 5206
- Library Leaders 5195, 5730-5731, 5733, 5735
- Library Services 2245, 4360, 5255-5259, 5261, 5264, 5727, 6685-6688, 6690
- library user 4105
- Life Cycle Costing 2176
- Life Cycles 788-790, 793-794, 796, 2405, 3013, 3963, 4560, 4688-4689, 6298
- Life Publishing 2650
- Lifelong Learners 1517, 1519-1523, 1528, 2545, 2547, 2552, 3872, 5189, 7062, 7671-7672, 7680
- Lifelong Learning 1421, 1463, 1512, 1517-1519, 1528, 2228, 2238, 2244-2246, 2273, 2545, 2556, 3827, 3832-3833, 3872, 4350, 6351, 6398, 6412-6413, 7171, 7269, 7642, 7672-7673, 7680, 7901
- Light-Emitting Diode 3080, 6682
- Lightweight Formal Methods 7615
- Limited English Proficiency (LEP) 3725, 6806
- Limor Shifman 8017
- Lineal Time-Invariant (LTI) 6244
- linear congruential generators 1298-1299
- Linear L-Checkpoint Algorithm 2144, 2148, 2153
- Linear Phase 4751, 6045, 6174, 6183, 6235, 6241
- linear-phase filter 6241
- lingua franca 1207, 1217-1218, 1225
- Linguistics 1195, 1215, 1217, 1219-1220, 1788, 1793, 2121, 2433, 2497, 2919, 3956, 4103, 4434, 4493, 4702, 5089, 5146, 5160, 6379, 6567, 6860, 6863, 7339, 7403, 7412, 7608, 7706-7707, 7711
- Link building 8048, 8050-8051
- Link Expected Results 7505, 7513
- Link Presentation 8010, 8013
- linkage discovery 1783, 1785, 1787, 1789, 1793
- Linkage Disequilibrium 467, 472, 476
- Linked Data 1200, 1333, 2485, 2914, 3477, 4953, 5068, 5076, 6458, 6753
- Linsear Write calculation 1507
- Linus 7372
- Linux 327, 566, 868, 1084, 1104, 1227, 1229-1231, 1233-1235, 2397, 2656, 3060-3061, 3280, 3420, 4004, 4014-4015, 4021, 5373, 5856, 5862, 6057, 6252, 7078, 7117-7118, 7369-7372, 7508, 7760, 7904-7905, 7911
- literature statistics 1344
- Literature Survey 50, 168, 1540, 2981, 3261, 3404, 4101, 4340, 4986, 4996, 5345, 7528
- Lithography 434, 437
- Little "I" Literacy 1516
- Ljubljana Castle 4077
- load and performance testing tools 7506
- Load Balancing 331, 1086, 1125-1126, 1165, 1981, 4584, 4586, 4591, 6527, 6530
- Load Factor 2849, 5373
- load flow analysis 3103-3109
- Local E-Commerce 2745, 2747
- Local Government Webs 7330-7331
- Local Governments 3087, 3594, 3603-3605, 3607-3608, 3611, 5554, 6219, 7327, 7330-7334, 7336
- Local Non-Contextual Links 6428, 6433-6434
- Local Private Companies 671, 681
- Local Seo 8055
- Localization 985-986, 1234, 2231, 3519, 3522, 4943, 5225, 5582, 5959, 6026, 6071, 6200, 6286-6291, 6295, 6675, 6680, 6836, 6841-6843, 6846, 6894-6895, 6897-6898, 6901, 6904-6905, 7406, 7959, 7962
- Localsumers 1644
- Location-Based Service 2019
- Locus 476, 566, 1051, 4218, 5093
- Logic Programming 184, 208, 1919-1920, 1974, 1983, 2032, 2035, 2039, 2042, 2086, 4736-4737, 4739-4740, 4743, 4745, 7451
- logical inference 1995, 2000, 2003
- logistic regression model 2146-2147, 2149, 2219
- Logistics 543, 584, 587-589, 598, 876, 891, 1144, 1408, 1412, 1585, 1815, 2487, 2665, 2668, 2743, 2802-2806, 2809, 2930, 3087, 3132, 3843, 4267, 5291-5292, 5305, 5309, 5327, 5350, 5354, 5356-5357, 5362, 5366-5376, 5378-5379, 5381, 5386-5387, 5407, 5415, 5422-5429, 5432, 5434, 5446, 5456-5462, 5464, 5467, 5490-5494, 5500, 5504, 5516, 5529, 5532, 5553, 5560, 5570-5571, 5574, 5576, 5578, 5582-5583, 5595, 5599-5600, 6100, 7305, 7697, 7776-7777, 7806, 7869-7876, 7878
- logistics capabilities 5356, 5425
- Logistics Capacities 5356
- Logistics Clusters 2803, 2809, 5428
- Logistics Competencies 2805, 2809
- Logistics Information Systems 7873, 7878
- Logistics Protocol 5356
- Logo 1639, 2392, 2394-2395, 3081, 3252, 4038, 4044-4048, 4872, 4883, 6380, 7184, 8080-8081, 8084, 8086
- Loki 7342-7343, 7346
- Long Tail 7109, 7111-7112, 7114

Index

Long Term Evolution (LTE) 6096, 6105, 6256, 6266-6269
Long-Term vs. Short-Term Orientation 7957, 7966
Lotus Effect 434, 439
Louvain 1754, 6961, 6965, 6968
Low Carbon Supply Chain Management 5446, 5455
Low Carbon Supply Chain Management Measurement 5455
Low Density Benders Cut 5421
Lower Approximation 1838, 2204, 2212
Loyalty 611, 672, 712-713, 715, 945, 1131, 1168-1169, 1584-1585, 1587, 1593-1594, 1596-1597, 1604-1605, 1642, 2749, 2756, 2817, 2883, 3194, 3360, 3613, 3616, 4057, 4217-4218, 4284, 4768, 5361-5362, 5364, 5423, 5528, 5532, 5719, 5778, 5983, 6281, 6483, 7003, 7131, 8074
LSA 1783, 1786, 5171, 5182
LTE Technology 6265
LU 300, 408, 542, 707, 719, 834, 1227, 1234, 1359, 1454, 1783, 1785, 1874, 1953, 2791, 4022, 4281, 4574, 4986, 5465, 6364, 6447, 6582, 6716, 7715, 7723, 7972
Ludic Elements 3290, 3295
LUNAR Technology 482, 489
Lurking 1716, 3771-3772, 3781, 4910, 6987
Lyapunov Function 23, 28, 35

M

Machine Intelligence 98, 100, 251, 1919, 4737, 4951, 4953, 7705, 7707
Machine Learning (ML) 1, 3, 5, 7, 14, 99-101, 103, 120, 122, 124, 127, 132, 134-135, 137, 139, 143-145, 148, 151, 155, 158, 160, 180-186, 201, 229, 231-232, 235, 238, 252, 350, 406, 410-411, 414, 421, 424, 455, 499, 501, 504, 851, 854, 867, 1011-1012, 1234, 1354, 1763, 1775, 1786, 1806, 1808-1810, 1813, 1831, 1898, 1919-1921, 1925, 1930, 1932, 1962-1963, 1988-1989, 1994, 2069, 2074, 2085, 2089, 2091, 2188, 2193, 2198, 2201, 2720, 2910, 2915, 2919, 3541, 3713, 4008, 4012-4013, 4092, 4349, 4529, 4531, 4575, 4937, 5170, 5749, 5877, 5909, 5950, 6042, 6137, 6144, 6146, 6205-6206, 6430, 6640, 6642, 6670, 6976-6977, 6981, 7063, 8031
Machine Translation 107, 233, 1786, 4492-4493, 5136, 7327, 7329-7330, 7334, 7336
Macro Environment 910
magical sheet 6466-6467, 6475
Magnitude Characteristic 4746, 4750, 6014, 6043-6044, 6046-6047, 6054-6056, 6171, 6175, 6178, 6180, 6183

Magnitude Response 4752-4754, 4756, 4759, 6007, 6009-6012, 6014, 6018, 6043, 6047, 6171, 6174-6175, 6177, 6183, 6234-6235, 6241, 6244
MAIN DRIVERS 1560, 3605, 3963, 3965, 3971-3972, 5004, 5307, 5715
Main Memory Utilization 7575
Mainlobe Beamwidth 5970, 5980, 6599
Maintenance Datasheet 3710
majority instances 1826-1827, 1829
Majority Voting 7282-7283, 7291
Makespan 4369-4372, 4377, 4381
Malaysia Supply Chain 5527-5529, 5532, 5536
Malaysian Government Securities (MGSt) 47
MALL 2745, 2857, 5021, 5116, 5118, 5124, 5126, 5771, 7133, 8043
Manage Changing Requirements 7513
Manage Software Outsourcing 7513
Management by Objectives (MBO) 4339, 4346
Management Education 1529-1530, 1536, 4734, 5628
managerial tools 2166, 2169-2170
Manga 2326, 2329, 2335, 8016, 8020, 8022
Manipulator Workspace 6876
Manmade Disaster 1415
Manufacturing Machines 3649, 7435, 7443, 7445
Manufacturing Sector 2954, 3089-3090, 3094, 5051, 5307, 5311, 5527, 7235, 7239, 7244, 7247
MAP 329, 389, 4020, 4475, 4479
map matching 6287, 6290, 6295
Mapreduce 326-328, 331-333, 335, 337-343, 345, 348, 374-375, 457, 867, 1774-1775, 1809, 2048
Marc 1510, 4301, 4768, 4961, 5204, 6389, 6491, 6750-6752, 6755
marginal contribution 2097-2098, 2107
marginal distributions 4939, 4945, 5584-5588, 5592
Marginal Effect 2094, 2838
Marginalization 2230, 2503, 3354, 4628, 6710, 7272, 7540
Maritime Green Supply Chain Management (MGSCM) 5465, 5467, 5475
maritime supply chain 5465-5467, 5470, 5475
Market Capitalization 2742, 2748, 2870, 2876, 5058, 5060-5061, 7118
market intelligence 961-967, 970
Market Value 923, 2479, 2748, 2956, 4114, 5056, 5060, 5062-5063, 5066
Marketing Decisions 961-963, 968, 1586, 5780
marketing engagement 5754
marketing plan 5757, 5761-5763, 5765
Marketing Planning 962, 964, 967, 970-971, 4573
Marketing Recommendation System (MRS) 5756
Marketing Strategy 1586, 1596, 1604, 2763, 3360, 4032, 4094, 5759, 5771-5772, 5983

- Marketplace Fairness Act 2845, 2849
 Markov Chain 421, 423, 1574-1575, 1581, 3989, 3993
 Markov Techniques 6858
 Masculinity vs. Femininity 7957, 7966
 Maslow's motivation theory 2534
 Mass Culture 4718, 4722
 Mass Customization (MC) 6697, 6708
 Mass Media 402, 754, 2292, 3350, 3571, 3612, 3617, 3620, 3638, 3640, 3648, 3655, 4066, 4189-4191, 4195, 4260, 4715-4716, 4718, 4721, 5137, 6188, 6338, 6488, 7112, 7947
 mass-digitization 5226, 5229, 5234, 5236
 Massive Digital Libraries (MDLs) 5226, 5231, 5234, 5236
 Massive Open Online Courses (MOOC) 1477, 2230, 2424, 2459, 3833, 5836, 5840, 6404, 7888, 7897-7898, 7907, 7917, 7921
 Master Data Management (MDM) 887, 955, 959
 master problem 5412-5413, 5416, 5418
 Material Feminism 3351, 3358
 Material Requirements Planning 923, 929, 2168, 2976, 5352
 Material Requirements Planning II 929
 Mathematics Education 2629, 5855, 6377, 6380
 Matrimonial Portals/Websites 4145
 matrix factorization 1745, 1761-1763, 1771, 1873, 1878
 matrix multiplicative congruential generator 1301
 Maturity Model 583, 954-955, 1405, 2280, 2285, 2288-2289, 2291, 3526, 3528-3531, 3537, 5050, 5286, 5290, 5293, 5297, 5300, 5305, 5480, 6928, 7553-7555, 7583
 Maturity Stages 3533, 3537, 5290, 5295, 5298
 maximally different solutions 2181, 2183, 2187
 Maximize Test Coverage 7513
 Maxwell 1252-1253, 1492, 1562, 4020-4021, 4480-4481, 5186, 6322, 6363
 MD5 1386, 1388, 1390-1394, 4976, 4987, 7735
 Me Generation 4358, 4367
 Mean Flow Time 4369, 4371-4372, 4377, 4381
 Meaningful Management 2663, 2666, 2671
 Meaningful Use 3679, 3685, 3688, 3725, 3732, 3805, 3815, 3817, 3825, 6566, 6806, 6812, 7174, 7248, 7251-7253, 7317
 Meaning-Making Process 1535, 5188
 Measurement Noise 219
 Measurement Residue 219
 Measurement Theory 2199, 3874
 Measurement Vector 219
 Media And Information Literacy 2244, 2246, 2292, 2294-2295, 2297-2298, 2302, 7548
 Media Literacy 1423, 1512, 1516, 2225-2226, 2243-2245, 2293-2294, 2298, 2325, 2480, 3309, 3670, 4963, 6398, 6404-6405, 6411
 Media Pluralism 7270-7277, 7279
 Media Queries 8079, 8081, 8088, 8094
 Media Services 2641, 2650, 2908, 6477, 6479, 6487, 7168-7169, 7172-7175
 Médiance 4296, 4298-4299, 4301, 4305
 Mediated Embodiment 4234-4240, 4244
 Medical Education 523-524, 2371, 3692, 5104, 5800-5805, 5829-5831, 5834-5837, 6152-6153
 Medical Laboratory Science 5853, 5855, 5860, 5862
 medical translation 5136, 5143-5144
 Medical-Based Geographic Information System (mGIS) 3459
 mega-project 5691-5693, 5697, 5699-5700
 membrane technology 3043
 Memes 1516, 2326, 2329, 4666-4668, 4672, 7111-7112, 8015-8020
 Memetics 4666, 4671, 8015
 Memex 4537, 4539
 Memorability 383, 5008-5009, 5011, 5013, 8081
 Memory Barrier 3997, 4003
 Mendeley 7044-7047, 7056, 7785-7787, 7789
 Mental Health 995-1002, 1369, 1697, 1710, 2413, 3261-3263, 3265, 3736, 4223, 4306, 4321, 5878, 6185, 6223, 6228-6230, 7032, 7080, 7881
 Mental Illness 3736, 4384
 Mental Model 555, 557, 1276, 2080, 3254, 5005, 7601, 7957
 Mentality 314, 318, 321, 324, 671, 2620, 3021, 3393-3395, 4448, 4912, 5048, 5628, 6651, 7340
 Mentoring 277, 281, 688, 1401, 2581, 2992, 3377, 3382, 3385-3386, 3527, 3852, 3860, 3938, 4358, 4361, 4363, 6404, 6787, 7077-7080, 7082, 7087, 7647, 7651-7652, 7900
 Mentors 1533, 3385-3386, 3844, 7079, 7081, 7652
 merchant account 2707-2708, 2710-2712, 2722
 Mesocriticism 4296, 4302
 Message Authentication Code (MAC) 1240, 1247, 4927
 Message Passing 337, 3997-3998, 4000, 4005, 4014, 4017, 4585, 6535
 Meta_Body 4151-4153
 Meta-Cognition 1529, 1538, 2550, 3239, 3983, 4360, 6400, 6405, 6411, 7673, 7675, 7970-7972
 Metacognitive Intelligence 286
 meta-data mining 4404
 Metadata Schemas 6463
 Metadata Standards 424, 5613, 6430, 6447-6448, 6455, 6459, 6463
 meta-data systems 2064, 2068-2069, 2071
 Metagenomics 422, 430

Index

- Metagogical Approaches 1472
metagogy 1468-1469
Meta-Heuristics 2183, 2187, 4370-4371, 5583, 7693-7701, 7703
metal poisoning 3076
Metal toxicity 3068
Metamodel 355, 3529, 5402-5404, 5410, 5545-5546, 7448-7449, 7451-7454, 7458, 7483-7484
meta-modeling 7449, 7451-7452, 7454-7455, 7483
Meta-Object Facility (MOF) 7453, 7458
Metaphor 203, 207-208, 211, 235, 241, 384, 442, 609, 976, 2160, 2589, 3940, 4124, 4146, 4517, 5078, 5086-5087, 5089, 5092, 5103-5106, 5108-5109, 5684, 5887, 6494, 6697, 6728, 6940, 7303, 7314, 7706, 7947, 8015
Metascanning 911
Meta-Theoretical Analysis 4406, 4416, 5237-5238, 5245, 5251, 7051
Metaverse 4146-4147
Method of LaGrange 4596-4597
methodological consistency 2435, 2438, 2441
Metrical Map 6905
Micro and Small Firms 6924
Micro Business 7121, 7124
Microarray 420, 440, 443-444, 447, 450, 453-454, 1969
Microblog 394, 7088, 7301
Microblogging 2065, 2855, 3572, 3620, 5776, 5778-5779, 5782, 5785, 6982, 6984, 6996-6997, 7096, 7119, 7168, 7293
Microblogging Service 2855, 3620, 6982, 6984, 7119
Micro-Creativity 7970-7976, 7978
Micro-Electro-Mechanical Systems (MEMS) 431, 439
Microelectronics 2684, 2907, 4704, 4714, 6114
Micropayment 2751, 2753, 2760
Microsoft Cortana 6658
Microsoft Kinect 133, 5876-5877, 5880, 5941-5942, 5950, 5954
Microsoft Solution Framework (MSF) 1398, 1405
Microswitches 288, 290-291, 297
Middle States Association of Colleges and Schools 3182
Middleware 375, 421, 810, 1064, 1068, 1103, 1115, 1129, 1359, 1400, 1952, 2757, 2777-2778, 2780, 2783, 3689, 5520, 5713, 5878, 6061, 6063, 6069, 6557, 6955, 7568, 7590, 7777, 7855
Milieu 100, 384, 611, 1529, 1716, 2243, 2992, 3622, 3627, 4298-4299, 4301-4302, 4305, 5199, 5262, 5268, 8036
Millennials (Y-Generation) 1508-1510, 1516, 2422, 2513, 2694, 3083, 4358-4363, 4367, 5197, 5982, 7145, 7540
Millennium Development Goals 4442, 5015, 5019, 5079, 5081, 7260
Mimicking 100, 434, 2460, 2618, 2624, 7360
Mindfulness 2327, 4325, 6490
Mindlessness 4202
Mindset 324, 555, 558-559, 692, 802, 805, 905, 2225, 2235, 2561, 2925, 2992, 3082, 3395-3396, 3399, 3621, 3842, 3844, 3849, 5006, 5330-5331, 5699, 5896, 7038, 7074, 7163, 7565, 7627, 7676, 8037, 8039-8040
Miner 411, 1622, 2790, 2870-2872, 2875, 3385, 3392, 6826, 7357
Minimally Invasive Surgery (MIS) 6882-6883, 6893
MiniMax Algorithm 144, 146-147, 151, 155
mining big data 406-407, 409-414
Mining Data Stream 417
minority instances 1826-1827, 1829
Minsky 121, 232, 5172, 6580-6581, 7731
Misinformation 2915, 4768, 5205, 6993-6994, 6997-7001
mission statements 5204, 5208
mission-assurance 5698-5699
Mixed Decision Tree 2137, 2141
Mixed Methods Research 6756-6759, 6765, 7689
Mixing Data 6765
Mix-Mode Learning 5862
Mjöltnir 7343, 7346
Mlearning 2465, 6194
mMedical 5912-5914, 5917
MMORPG 4233
Mobbing 1696, 1698, 1702, 6277
mobile ad hoc network (MANET) 6113
Mobile Applications 266, 744, 1294, 2043, 2047, 2636, 3587, 3828, 4054, 5376, 5988, 6035, 6060, 6067, 6084, 6090, 6092, 6147, 6152, 6154, 6161, 6195-6197, 6200, 6202-6203, 6207, 6245, 6338, 6673, 7263, 7545, 7574, 7638, 7641, 7807, 8043
Mobile Commerce (M-Commerce) 2794, 5492, 6067, 6071-6072, 6076, 6185, 6189, 8087, 8095
mobile crowdsourcing 6194
Mobile Device State 6658-6660, 6662, 6669-6670
Mobile Electronic Word-of-Mouth (mWOM) 6030
Mobile Enhanced Language Learning 7637, 7645
Mobile Game-Based Learning (mGBL) 6361-6362, 6364, 6375-6377
Mobile Games 2467-2468, 6067, 6302, 6361, 6363-6364, 6375, 6377-6378, 6380
Mobile Graph 6538, 6541-6543, 6545
mobile handheld devices 6057-6059, 6069
Mobile Health (Mhealth) 3691, 3694, 3697, 5908-5911, 5913-5914, 5917, 6083-6084, 6088-6089, 6092, 6114, 6122, 7843
Mobile Learning 1109, 1464, 1476-1477, 1519, 1521, 2303, 2466-2468, 2592, 3960, 5116-5118, 5121,

- 5126, 5154, 6184, 6189, 6194, 6347-6353, 6356, 6360, 6362, 6377, 6379, 6381, 6387-6389, 6398-6406, 6411, 6415-6416, 7637-7638, 7927, 7930
- mobile malware 6207-6210, 6213, 6215
- Mobile Marketing 2814, 2822, 2824, 6019
- Mobile Media Technologies 7013
- Mobile medicine 6152
- mobile messaging 6184-6186, 6190, 6194
- Mobile Money 2724, 2726-2735, 2738, 6216, 8100
- mobile money adoption 2724, 2727-2728, 2731, 2733-2735, 2738
- Mobile Money Operators (MMO) 2724, 2738
- Mobile Money Services 2724, 2726-2727, 2730-2731, 2733-2734, 2738, 8100
- Mobile Phone Culture 6306
- Mobile Phone Technology 1143, 6150-6151, 6153-6154, 6160, 6166, 6277, 6282
- Mobile Phones 382, 1723-1724, 1887, 2229, 2341, 2726, 2730, 2738, 2794, 2886, 2900, 3001, 3066, 3184, 3351, 3471, 3649, 3691, 3747, 3827, 3857, 4161, 4188, 4247, 4261, 4263, 4297, 4652, 4709, 4843, 4953, 4963, 5017, 5019, 5126, 5576, 5933-5934, 5983, 6019, 6024, 6096, 6100, 6105, 6124, 6130, 6137, 6141, 6147-6153, 6155, 6160, 6167, 6184, 6186-6187, 6194, 6207, 6217, 6219, 6233, 6277, 6279-6282, 6286, 6288, 6291, 6296, 6298-6300, 6306, 6338, 6348, 6353, 6388-6393, 6397, 6399, 6411, 6704, 6987, 6998, 7003, 7061, 7117, 7215-7216, 7224, 7301, 7375, 7637-7638, 7640, 7642-7643, 7761, 7807, 7917, 7986, 8023, 8031, 8097
- Mobile Point of Care (MPOC) 6222
- Mobile Robot 6836, 6838-6839, 6841-6842, 6844, 6895, 6898, 6905
- Mobile Social Commerce 2860
- Mobile Social Media 2642, 6194
- Mobile Technology 268, 1687, 2465, 2592, 2866, 3522, 3530, 3571-3572, 3691, 3954, 3956, 4090, 5291, 5908, 5981, 6101, 6148-6149, 6151, 6161-6163, 6167, 6185, 6187, 6216-6218, 6220, 6222, 6278, 6285, 6297, 6306, 6355, 6361, 6364, 6367, 6388-6392, 6397, 6399, 6404-6405, 6987, 7068, 7295, 7321, 7636, 7642, 7758
- Mobile Telemedicine 6154, 6160
- mobile testing system 5116-5117, 5124, 5126
- Mobile Users 1102, 1761, 3522, 6031-6032, 6034-6035, 6038, 6058, 6162, 6266, 6269, 6272, 7120, 8087
- mobile viral marketing 6019-6024, 6030
- Mobile Virtual Reality (MVR) 6223, 6225, 6233
- Mobile Vision (MV) 6195, 6206
- Mobile/Handheld Computing 554, 3691, 4319, 4896, 5909, 5984, 6057, 6069, 6189, 6286, 6361, 6363-6364, 6367, 6476, 6480, 6658, 7376, 7447, 7805, 7809, 7813
- mobile-based eWOM services 6026, 6030
- Model Development 5447, 5516, 7387
- Model Driven Architecture (MDA) 2046, 3339, 4810, 7448, 7457
- Model Driven Development (MDD) 7447
- model explanation 2089-2090
- Model Level Explanation 2094
- Model Transformation 7449, 7457
- model tree 2132, 2141
- Model-Based on SPI 7562
- Model-Based Specification 7613, 7621
- Model-Driven Engineering 2020-2021, 3338, 3344, 3347
- Modeling Language 772, 3340, 3344, 4821, 7436, 7449, 7475, 7481-7485, 7487
- Modelling Approach 352-353, 1010, 1740, 1801, 2187, 2372, 2927, 3089, 4038-4039, 4051, 4065, 4071-4073, 4076, 4420, 5398, 5437, 5440-5441, 5504-5505, 5541, 5543, 6479, 6827
- Modelling Technique 3337, 3339, 5436, 5438, 7500
- Modelling to Generate Alternatives (MGA) 2178, 2187
- Modernity 3642-3643, 4301, 4431, 4716, 4722, 4761-4763, 6343, 7312-7313, 7318, 7321-7322
- Modernization 888, 895, 2021, 2717, 3591, 3593, 3595, 3621, 4014, 5036, 5258, 7201, 7206, 7209, 7311, 7447-7448, 7450-7452, 7454-7455, 7457
- Modes Of Communication 2576, 7684, 7691
- MoDisco 7449, 7454, 7457
- Modular Dependency 5326
- Modularity 564, 5163-5164, 5317-5326, 6120, 6836, 6846, 6890-6891, 6961-6962, 6964-6965, 6968, 6970, 6972, 7470-7471, 7473, 7698-7699
- Moiré Technique 169-170, 179
- Money Supply 40, 44, 47, 2787
- Monochromatic 8081, 8086
- Montage 973, 982
- Monte Carlo Simulation 3109, 5440, 5445, 5689, 6820
- Monthly Active Users (MAU) 2748, 5255, 7118, 7120, 7124, 7190
- monthly profit 5641
- Moodle 805, 1518-1521, 1523, 1552, 1554-1555, 1558, 1562, 1676, 1881, 2405, 2525, 2581, 2763, 3572, 3828, 3833, 3979, 5790, 5792, 5836, 7674-7675, 7914-7917, 7921, 7935-7936
- Morality 88-90, 92, 94, 3194, 3203, 3206, 3208, 3298, 3569, 3637, 4078, 4195, 4360, 4887, 5040, 5203, 7292, 7301, 7315

Index

- Morphing Organization 694, 702, 706
Morphological Analysis 2976, 2979-2980, 2983, 3132, 4564, 5146
morphological framework 2980-2982, 2986
Morphosemantics 5146
Moscovici 3320, 4039, 4064-4065, 4072, 4076, 4404, 4420, 5237, 5251, 7014-7016, 7019, 7021, 7046-7047, 7051, 7056
motion detection 5882
Motion Envelope 6881
Motivational Context 2431, 2441
Motivational Intelligence 286
Motor Disabilities 3525
Motorola 729, 3528, 4608, 4774, 5050, 6154, 7262, 7373, 7376
Moven 7145
Movie Makers 2454
mPersonal 5912, 5914, 5917
MPI 2137, 4005, 4012, 4014, 4017, 4584-4586, 4588, 4590
MPOB 5422-5424, 5426
m-quotient 6743
MRI 507, 523, 1255, 2080, 4994, 5880, 5892, 5900, 5902
MRR 4474-4476, 4479
MSC 831, 4724, 4727, 5527-5534, 7203-7204, 7211
mSocial 5912-5914, 5917
MSPO 5423, 5428
Multi-Agent based Models 3433
Multi-Agent Systems (MAS) 114, 160, 4798, 4801, 4807, 4809, 5401, 6839, 6846, 6953-6954, 6956, 6960
multi-asset Black-Scholes PDE 16, 23-24, 26, 28-30
Multi-Axis Mechanisms 7445
Multi-Channel 351, 500, 2303, 2359, 2841, 3605, 4089, 4094, 5371, 5908, 5912
Multichannel Retailers 2849
Multicollinearity 2958, 5781, 6635, 6640
Multicore Architecture 3992, 4003
Multi-Core CPU 1947, 1961, 4025
Multi-Criteria Decision Analysis 1761, 3132, 3143
Multi-Criteria Recommender System 1739, 1745
Multi-culture residents 7327
Multi-Dimensional Data Modeling (MDM) 354, 363
Multidimensional Frameworks 6568
Multidimensional M-Checkpoint Algorithm 2153
Multi-Domain Framework 3337, 3341-3342, 3345
Multi-Event Expert Query 2153
multi-generation of cuts 5416-5417, 5421
Multi-Interactive Collaborative Higher-Order Thinking Skills (MICH) 6346
Multi-Interactive Collaborative Higher-Order Thinking Skills (MICH) Activities 6340, 6343, 6346
Multi-Jet Modeling 4383
Multilateral Development Banks 5015
Multilayered Method 1250, 1256, 1259
Multilevel Modeling 5183, 5185, 5192
Multilingual Communication of Information 7331-7332, 7334
Multi-Material Layered Manufacturing 505, 520
Multi-Material Object 520
Multi-material virtual prototyping 506, 518, 520
Multimedia Assets 6480, 6487
Multimedia Cognitive Theory of Learning (MCTL) 2454
Multimedia Learning 2443-2444, 2454, 2603, 2652, 3292, 6402
Multimedia Messaging Service (MMS) 6297, 6306
Multimedia semantics 6478, 6480
Multimedia Services 6031, 6266, 6269-6270, 6274, 6477, 6481, 6487, 6611, 6615
multimedia synchronization 6436, 6443
multimodal literacy 1508, 1511-1513, 1516
Multimodal Media Agents (M3A) 6479, 6487
Multimodal Media Assets (MMA) 2908, 2919, 6477, 6487
Multimodal Metaphor 5092
Multimodal Student 1558
Multinational Corporation (MNC) 671, 681, 2791, 4408, 5050, 5334, 6935, 7155, 7261, 7982
Multi-Objective Optimization 239-241, 249, 2132, 2137-2138, 4857, 5976, 6595, 6603
Multi-Platform Social Media Advertising Campaign 4063
Multiple Input, Multiple Output (MiMo) Technology 6105
Multiple Media 1422, 2235-2236, 2292, 2357, 2368, 2459, 2589, 2592, 4700, 4702, 6435, 6487, 7248
Multiple Objectives 239, 241, 4369
multiple recursive random number generator 1298-1299
Multiple Tenancy 1184
Multiplierless Design 4758, 6056
Multi-Point Relays 6506
Multi-protocol Label Switching (MPLS) 6609-6611
Multi-Sink Environments 6522, 6524, 6531, 6535
Multisourcing 5478, 5489
Multi-Spectral 3492, 3501
Multitasking 1508, 1510, 2316, 2327, 3360, 4316, 4319, 4321, 4325, 4362, 5215, 5223, 6222, 6404, 6843, 7030
Multitenancy 1356-1357, 1365
Multi-Threaded Discussion 7978

multivariate cryptography 4900, 4908
 Multivariate Decision Tree 2133, 2141
 multivariate non-normal 5584-5586, 5592
 multivariate normal 5584, 5586-5587, 5592
 Museum of Ribnica 4077
 Mutable 4326, 4336, 6557
 Mutation Analysis 7469
 Mutation Operator 241, 2136, 7462, 7469
 Mutation Score 7461-7462, 7469
 Mutation Testing 7459-7465, 7469
 MVCC 1954, 1961
 MW 3148, 4015-4017
 MyKimDG 3248, 3250-3256, 3259
 Myspace 1705, 2316-2317, 2319, 4844, 5252, 5255,
 5267, 5840, 6185, 6961, 6988, 6994-6995, 7088,
 7096, 7128, 7130, 7250, 7798
 Myth 317, 754, 761, 1658, 3397, 3939, 4767, 7022,
 7338, 7344-7346

N

Named Data Networking (NDN) 6557, 6565
 Naming / Word Designation Process 5146
 naming guidelines 5136-5140, 5146
 NAND 1948, 2899, 3338, 4011, 4017
 nanomaterials 2675, 2679-2681, 2684, 2897-2899,
 2901-2903, 2906
 Narb 396, 398, 404
 Narrative 207, 389, 396-402, 404, 676, 756, 804, 974-
 975, 977, 980, 982, 999, 1197, 1892, 2286, 2608,
 2616, 2621, 2624, 3239, 3242, 3266, 3287, 3290,
 3310, 3316, 3334, 3350, 3502, 3646, 3653, 4055,
 4234, 4339-4340, 4358, 4669-4671, 4677, 4692,
 4695-4700, 4702, 5077, 5086, 5135, 5664-5666,
 5668, 5835, 6167, 6476, 6479, 6488-6489, 6494,
 6695, 6716, 6719, 6726, 7293, 7316, 7338-7339,
 7342-7343, 7345-7346
 Narrative inquiry 5664
 Narrative Map 398, 400, 402, 404
 narrative therapy 2616, 2621, 2624
 Nash Equilibrium 3231, 3236, 4857
 National Athletic Trainers' Association (NATA) 5810,
 5819
 National Cash Register 86
 national crisis 2006, 2008, 2016, 2019
 National Culture 4779, 5627-5628, 5630, 5637, 8099
 National Development Incubator 5530, 5536
 National Institute of Standards and Technologies 1032
 National Population Census 7336
 Natural Disaster 1415, 2008, 2068, 2468, 2472, 5551-
 5552, 5555, 5558, 5656, 6106, 6690, 6730, 7947

Natural Gas 3043, 3055, 3375
 Natural Language Model 5135
 Natural Language Processing (NLP) 1-3, 99, 101, 106,
 184, 233-235, 1209, 1442, 1444, 1447, 1783,
 1897, 2370, 2910, 2919, 4092, 4472, 4528-4529,
 4531, 4533, 5749, 5869, 5935, 6199, 7696, 7800
 Natural Science So Far 4400-4401, 4403
 Natural User Interface (NUI) 985-986, 5876, 5942,
 5954, 6206
 Naturalistic Perspective 1194
 Nature-inspired Algorithms 7352
 Nature-Inspired Meta-Heuristics 2187
 NDVI 3503-3504, 3508-3509, 3514
 Near-Field Communication (NFC) 2007, 2019, 4951,
 4961, 6166, 7777, 7807, 7813
 Needs Assessment 5922, 7645
 NEET 1559-1562, 1566, 1568
 Negative Social Cue 7974, 7979
 Neighborhood systems 1838, 1840, 1842, 1844
 NEME 6488, 6490-6491, 6494, 6497
 Nemetics 6488, 6490-6495, 6497
 Neo4j 655, 657, 1978-1980, 2031, 2034-2036, 2041,
 2066
 Neoclassical 3021, 3639, 4282
 Nested terms 1898, 1906
 Net Neutrality 7294, 7297, 7301
 net-based education 1220, 1222-1223, 1225
 Netiquette 7071, 7076
 Netpreneurs 3001, 3007
 Network Architecture 206, 1228, 2752, 4530, 6265,
 6499-6500, 6502-6503, 6505, 6557, 6560, 6562,
 6565, 6586, 6588-6590, 6640
 Network Centralization 3978, 3985
 Network Effects 2281, 2349-2350, 2354, 2728, 6941,
 7410
 Network-Centric Warfare 7304-7305, 7311
 Networked Individuals 2318, 7060-7061, 7068
 Networked Organizations 2773, 7062-7063, 7068
 Networked Participatory Scholarship 6780, 6784-
 6785, 6787
 Networked Workers 7063, 7068
 Network-Function Virtualization (NFV) 6615-6616
 Network-On-Chip 4551, 4554, 4558-4559
 Networks Epistemics 3929
 Neural Networks 1-8, 14, 100, 106, 108, 112-113, 120,
 125, 144, 148-151, 155, 166, 180-181, 184-186,
 205, 232, 238, 350, 406, 408, 423, 441-442, 447-
 448, 450, 453, 493, 496-498, 1519, 1795-1796,
 1808, 1827, 1899, 1909, 1918, 1924, 2085, 2716,
 3107-3108, 4012, 4529-4530, 4533-4534, 4644,
 5169, 5171-5172, 5175, 5177-5178, 5182, 6579-

Index

- 6582, 6584, 6588-6591, 6594, 6631, 6633, 6640,
6642-6644, 6658-6660, 6670, 6909, 7352
neuroaesthetics 5767-5768, 5770-5772, 5774
Neurode 120, 122, 131
neuroeconomics 6634
Neuro-fuzzy systems 183, 185
Neurologist 1819, 1824, 4193
neuromarketing 5767-5768, 5770-5772, 5774
Neuron 112, 131, 149, 204, 206, 232-233, 477, 483,
497, 5178, 6580-6584, 6589, 6594, 6631, 6633,
6643-6644, 6647, 6657
Neuroscience 181, 207, 2413, 2433, 3002, 4236, 4267,
4414, 5089, 5770-5772, 5774, 5886, 5889, 6378,
6634, 6942
New Economic Model 5527, 5536
New England Association of Schools and Colleges,
Inc. 3182
New Global Economy 7980-7986, 7989-7990, 7993-
7994
New Keynesian 3022-3023
New Literacies 1509, 2226, 2326-2329, 2331, 2335,
7252
New Media 389, 1705, 1708, 2281, 2294, 2318-2319,
2327, 3355, 3551, 3560, 3667, 3672, 4073, 4161,
4163, 4234, 4257, 4259, 4261-4263, 4404, 5237,
5763, 5775, 5887, 6167, 6278, 6281, 6303, 6306,
6442, 7051, 7069-7070, 7110, 7214, 7270, 7273-
7274, 7322
New Media Studies 4073, 6167, 6306, 7051
New Public Management 3592-3593, 6710, 6714,
6716, 6725
New Technologies as a Marketing Tool 3828, 3838
New Work World's Fair 4702
Newsvendor Framework 3711-3713, 3720-3721, 3723
Next Generation Network (NGN) 6267, 6546, 6554
Nexus 2647, 3540, 4051
Ngram Viewer 5228, 5232-5234, 5236
Nigerian Library Association 5255
Night-Time Images 7831-7833, 7839, 7841
Noise traders 71-73
nominalization 5128-5129, 5131-5132, 5135
Nomophobia 3001, 3007, 4194
Non Recursive Form 4749, 4751, 4757-4758, 6171
Non-Cognitive Skills 2492, 6859, 6861-6862, 6870
Non-contractibility 531-534
Non-Fictional Texts 3953
non-financial reporting 5738, 5740, 5742, 5747
Non-Formal Learning 7680
Nonfunctional Requirements 2942
Non-Places 4301, 4305
Non-Recursive Form 4757, 4760, 6014
Non-Uniform Memory Access 1948, 4005
Non-Volatile Memory 1810, 1947-1948, 1950, 1955-
1956, 1961, 4009-4011, 4017
Normal Operation 1240-1241, 1648, 1651, 3057-3061,
3065, 6337, 8060
Normalization 53, 851-852, 861, 864, 1009, 1044, 1408,
1991, 3475, 4317, 4613, 4752, 4875, 4990, 5612-
5613, 6312, 6645, 7286, 7659-7660, 7768, 7832
Normative Theory 6945
Norming 5656, 5659
Norms 90-92, 532-533, 552, 557, 598, 744, 814, 823,
1217-1223, 1261, 1707, 1733, 2314, 2318, 2321,
2515, 2944, 2986-2987, 2990, 3023, 3027, 3088,
3158, 3177, 3320, 3351, 3355, 3383-3384, 3391-
3393, 3401, 3489, 3541, 3638, 3676, 3771, 3923,
3925, 3929, 3935, 4051, 4053, 4184, 4190-4192,
4195-4196, 4198, 4248, 4280, 4285, 4680, 4719,
4733, 4921-4922, 4925-4926, 5037, 5095-5096,
5596, 5664, 6249, 6354, 6382, 6403-6404,
6705-6706, 6779-6780, 6941, 6943-6945, 6988,
7038, 7080, 7154, 7163, 7194, 7219, 7223, 7292,
7327, 7333, 7602, 7710, 7787, 7790, 7956, 7965,
8027, 8067
Norns, The 7346
Northwest Accreditation Commission (NWAC) 3182
Nosql 326-329, 332-333, 335, 375, 378, 459, 464,
862-863, 872, 1095, 1100, 1774, 2031-2032,
2035, 2042-2043, 2045, 2047-2049, 2054, 2056,
2058-2060, 3274, 3278, 3286, 3462, 7568
NoSQL Approaches 3286
Nosql Database 375, 459, 464, 1100, 1774, 2047-2049,
2059, 3278
Novice Teachers 7647-7648, 7651-7653, 7655, 7686
NP-Hard 421, 1973, 4369-4370, 4381, 4898, 4908,
6507-6508, 6511, 6825, 7719
NP-Hard Problems 1973, 4370, 4381, 4898, 4908, 6511
NSM 1949, 1961
N-Tuple Algebra 1995-1996, 2003, 2005
Nuclear Power 1399, 3653, 4268, 6847, 7947
nucleolus 2095-2097, 2106-2107
Nucleus 206, 467, 476-477, 513-514, 4047, 4051,
5890, 7414
Numerical Methods 3105, 3107-3109, 4594, 4606
NVM 4009-4011, 4017
NVRAM 1810, 1948, 1954-1956, 1961
- ## O
- OA 4438, 6749, 6755
OAI-PMH 4437-4438, 6750, 6755
Object Classification (OC) 191, 6206

- Object Detection 212-213, 219, 1351, 6200
- Object Management Group 4765, 5436, 5438, 7481-7483, 7487
- Object Recognition (OR) 6195, 6199-6200, 6206
- Object Representation 219
- Object Tracking 212-213, 216, 219, 1344, 1348
- Object-Based Classification 7833, 7839, 7842
- Objective Information and Subjective Information 4434
- Objective Relevance 4496, 4501
- Object-Oriented Analysis And Design (OOAD) 1405, 2937, 7475
- Object-Oriented Paradigm 1405, 7465, 7469
- Object-Oriented Programming 1343, 2395, 3289, 7444, 7460, 7464, 7470, 7482, 7487
- Observation Matrix 219
- Occlusion 218, 5880-5881, 5950, 5954
- Occupational Hazard 3076
- Occupational Therapy 2413, 3261, 5812, 5822, 5828, 5931, 5935, 5940
- OCD 2160-2162, 4306, 4313
- odds ratio 5995, 6005
- Odin 7342-7343, 7346
- OER 5428, 5432, 7917
- Office Mini-Storage Products 2626
- off-peak 3104, 3147, 3155
- Offshore Handoffs 5478, 5489
- Offshoring 5476, 5478, 5480, 5485, 5489, 5518, 7259, 7261-7263, 7265
- OLPC 2229, 4623, 6252
- Olweus 1707, 1724, 4159-4160, 4169-4171, 4245
- Omics 430
- Omni-Channel 887, 5368, 5371-5372, 5375-5377
- Omnidirectional Vision 6905
- on premise 887, 928, 1028, 1030
- On-chip communication 4549, 4559
- On-Demand Economy 2346, 2354
- online academia 2580-2582, 2585, 2587
- Online Asynchronous Learning 3902, 3909
- Online Behavior 1367, 2315, 2317, 4842
- Online Bullying 4159, 4162
- Online Communities 565, 1217, 1219, 1454, 2237, 2281, 2523, 2851, 4052-4055, 4058-4059, 4063, 4146, 4194, 4278-4279, 4282-4284, 4286-4287, 5777, 6024, 6413, 6761, 7003-7004, 7081, 7090, 7161, 7168, 7181, 7274, 7394-7395, 7799, 7844, 8066-8067, 8069-8072, 8074
- Online Consultation (E-Consultation) 3562, 3568, 5912-5913
- Online Course 1443, 1451-1454, 1456, 2483, 2506-2508, 2583, 2594, 2661, 2664-2666, 2671, 3900-3902, 3905-3906, 3909, 3930-3931, 3933-3934, 3936-3938, 3940, 3944, 3979, 5789, 5812, 5822, 5824-5825, 5840, 7891, 7897, 7911, 7917, 7921
- Online Dating 7069-7074, 7076
- Online Digital Technologies 7312-7314, 7320, 7322, 7325
- Online Discussion 2303, 2575, 4295, 5098, 5258, 6019, 7022, 7028, 7088, 7969-7970, 7972, 7974-7975, 7979
- Online Discussion Message 7970, 7979
- Online Discussion Topic 7974, 7979
- Online Dispute Resolution (ODR) 2825, 2827, 2832
- Online Donating/Fundraising 7082, 7087
- Online Education 326, 1146, 1452, 1455, 1499, 1676, 2457, 2503-2504, 2506, 2579, 2593, 2661, 3083, 3930, 3937, 5788, 5824-5825, 6149, 7396, 7898
- Online Exam 1679, 1683
- Online Friendship 7173
- Online Instruction 1451, 1488, 1491, 2584, 2661, 2667, 3238, 5792, 5811, 5820, 5823, 5840
- online knowledge sharing 5093, 5096-5099, 5102
- Online Learning 834, 1451, 1453, 1455-1456, 1462, 1476, 1490, 2256, 2384, 2507, 2560, 2570, 2574, 2580-2585, 2588, 2590-2594, 2661-2663, 2665-2666, 2668, 3239, 3242, 3827-3828, 3834, 3850-3852, 3854, 3857, 3900, 3902-3903, 3930-3931, 3933-3934, 3937-3938, 3976-3978, 3983, 4130, 4283, 4437, 4623, 5080, 5093, 5096-5098, 5154, 5159, 5789-5790, 5792, 5808, 5810-5814, 5816, 5820-5825, 5828, 5834, 5860, 6415, 7251, 7253, 7602, 7604, 7884, 7890-7891, 7894, 7902, 7912, 7916, 7922, 7929
- Online Learning Environment 2570, 2590, 2661-2663, 2668, 3242, 3978, 5093, 5097, 7929
- Online Library 3953
- online mediation 2825-2830, 2832
- Online Mentoring 3852, 7077-7080, 7082, 7087
- Online Payments 2348, 5397
- Online Platform 572, 1201, 1660, 2354, 3579, 3587, 3794, 5102, 7074, 7096, 7181
- Online Privacy Policy 6093
- Online Product Review 7139
- Online Programs 1451, 1455, 2504-2505, 3400, 3937, 5821
- Online Prosocial Behaviors 7077-7079, 7081-7082, 7087
- Online Reading Comprehension 2240, 2242
- online recreation 4316, 4320-4321, 4325
- online reputation 5782
- online reputational damage 1695-1696, 1698, 1703

Index

- Online Reputational Mechanism 2840
- Online Retail 2769, 2882, 2885-2887, 5369, 5371
- Online Risk Behavior 1721
- Online Risks 1716, 1726, 4962, 4964-4966, 4969, 4973-4974, 7097, 7100
- Online Safety 1705, 1715-1716, 1721, 2318, 4963
- Online Shopping 57, 1616, 2707, 2745, 2748, 2761-2765, 2769, 2792, 2794, 2806, 2851-2853, 4318-4319, 4986, 5768, 7798
- Online Social Networking 1143, 2320, 2854-2855, 5256, 7088-7092, 7096, 7169
- Online Social Networking Sites 7088, 7091-7092, 7096
- Online Social Networks 2325, 2580, 2856, 4053, 4266, 5253-5254, 5258, 6943, 6945, 6949, 7063, 7088, 7091, 7797
- Online Support 3767-3772, 3775-3776, 3781, 7079, 7081-7082, 7087
- Online Support Groups 3767-3772, 3776, 7079, 7081-7082, 7087
- Online Synchronous Learning 3909
- On-line Teaching and Learning 3944
- Online Transaction Processing (OLTP) 337, 1376, 1807, 1949
- Online Video Platform 7280
- Online Voluntarism 7080
- Online-to-Offline (O2O) 2745, 2748
- on-Premise ERP 2965, 2968-2969, 2971, 2974
- ontology development 1441-1447
- ontology evaluation 1441, 1446
- Ontology-Based 1447, 1787, 1793, 2778, 4954
- Entrepreneurs 3001, 3007
- OOPSLA 7482, 7487
- Opaque 66, 203, 758, 762-763, 2085, 2091, 6642, 6673, 6680, 7037, 7857, 7862
- Open Access 1509, 2483-2485, 2491, 3032, 3041, 3461, 3833, 4317, 4405, 4436-4441, 4447, 4938, 5208, 5228-5229, 5231, 5234, 5236, 5239, 5246, 5253, 5268, 5308, 5815, 5820, 5840, 6343, 6748-6749, 6755, 6781-6782, 6911, 7045, 7297, 7892, 7898-7899, 7902-7904, 7908, 7921
- Open Data 420, 761, 1196, 1200-1201, 1205, 2067, 3032-3036, 3038-3039, 3041, 3530, 4405, 4436-4444, 4447, 5215, 5571, 5613, 7849
- Open Data Ecosystem 3041
- Open Data Repositories 4436-4440, 4444
- Open Government 756, 761, 3033-3034, 4439-4441, 4444, 4447, 7985
- Open Government Data 761, 4439-4441, 4444, 4447
- Open Innovation 567-568, 711, 4282, 5027, 5711, 7369-7371, 7373-7374, 7376-7377, 7381, 7845
- Open Learning 1434, 1517, 1519, 6415
- open learning environment 1434
- Open Repositories 4437-4439, 4443
- Open Research 461, 1652, 1741, 1911-1912, 1930, 2049, 2060, 2163, 3035, 3473, 3782, 4024, 4439, 4444, 4555, 4615, 5882, 6562
- Open Research Data 3035, 4439
- Open Scholarship 6780, 6783-6785, 6787
- Open Science 1064-1066, 4442, 7153, 7160
- Open Science Grid 1064-1066
- Open Shortest Path First (OSPF) 6610, 6616
- Open Simulator 2525, 4151, 4156, 7882, 7887
- Open Source 327-328, 355, 410-411, 414, 566, 711, 854, 859, 872, 895, 1083-1084, 1109, 1517, 1558, 1617, 1807, 1809, 2342, 2481, 2483, 2486, 2491, 2651, 2654, 2656, 2660, 2870, 2879, 2977, 3107, 3275-3276, 3284, 3286, 3462-3463, 3534, 3934, 4014-4015, 4437, 4444, 4624, 4923, 5108, 5163, 5218, 5234, 5880, 6062, 6245-6246, 6248-6249, 6252, 6260, 6452, 6874, 6956, 7077-7078, 7081-7082, 7087, 7130, 7149, 7252, 7369-7374, 7376-7377, 7380-7381, 7407, 7503, 7507-7508, 7566-7567, 7697, 7798, 7898-7899, 7908, 7911-7912, 7914-7917, 7921, 8007, 8035, 8091
- Open Source Software (OSS) 410, 711, 893-895, 898, 1109, 1227, 1233-1234, 2651, 2660, 3276, 3534, 3828, 4014-4015, 4923, 5218, 6252, 6956, 7077-7078, 7081-7082, 7087, 7149, 7252, 7369-7372, 7374, 7377, 7381, 7407, 7798, 7899, 7911-7912, 7914, 7916-7917, 7921
- OpenCL 4005, 4008, 4014, 4017
- OpenCV 214
- OpenGL 2420
- Opening Price 2835, 2840
- Openmp 2137, 4005, 4008, 4014, 4017
- Openness 91, 754, 756, 759, 762, 764, 976, 1029, 1563, 3056, 3464, 3603, 3840, 3844, 3952, 4151, 4191, 4320, 4327, 4436, 4685, 4710, 4855, 4918, 5098, 5163-5164, 5424, 6210, 6783, 6787, 7059, 7195, 7297, 7375, 7410, 7843, 7888
- open-source technologies 895
- Operability 5007-5008, 5013
- Operating System 208, 234, 276, 375, 464, 703, 849, 868, 1027, 1032, 1078-1081, 1083-1084, 1176-1180, 1227-1230, 1234, 1250, 1333-1334, 1365, 1443, 1629, 1658, 2318, 2487, 2752, 2757, 2899, 3276, 3991, 4129, 4553, 5373, 5834, 5856, 5859, 5862, 5917, 6042, 6057, 6060-6061, 6063, 6069, 6245, 6252, 6392, 6435, 6443, 6836, 6843, 7064, 7078, 7369-7370, 7372-7374, 7376-7377, 7380-7381, 7760, 8039, 8089
- Operational Distance 750, 5989

Operational Planning Horizon 5515
 operationalization scheme 4909, 4912, 4918
 Operationalize 2886, 3338, 4170, 4236, 4336, 4786, 4794, 4909, 4913
 Operations Efficiency 5581, 5595, 5604, 5929
 Operations Research 2, 600, 2196, 3143, 3230, 5411, 5700, 6908
 optical sensory technology 2459, 2465, 2591
 Optimal Management Policy 1572, 1577, 1581, 1583
 Optimal nursing staffing 3712-3713, 3717-3718
 Optimal Portfolio Utility Allocation 2194
 optimal security investment 4849, 4854, 4856-4857, 4860
 Optimal Staffing 3711-3713, 3716, 3718-3721, 3723
 Optimism Bias 5697, 5704
 optimization model 5338, 5400, 5505
 Optimize Software Tests 7513
 Oral/Aural Communication 2431, 2436, 3876, 5127, 5887, 5889, 5892-5893, 5896
 ORB 6198, 6200
 Orchestration 788, 1040, 1053-1057, 1141, 3976, 5373, 6499, 7761
 order statistics 1856-1862, 1864, 1866, 1868
 Ordered Tree Inclusion 4503, 4505, 4512, 4514
 Organic Results 8055
 Organization Architecture 694-699, 701-704, 706
 Organization Identity 694, 697-698, 701, 706
 Organization Structure 61, 65, 697, 706, 892, 2986, 3626, 5309
 Organization Transformation 706
 Organizational Capital 5057, 5066, 5671
 Organizational Change 694, 696, 698, 701, 956, 2966, 2978, 2982-2983, 2986-2987, 3817, 3822, 4329, 5199
 Organizational Commitment 3382, 3385, 3793, 3883, 4344, 7079
 Organizational Culture 553, 559, 755, 2274, 2584, 2960, 3088, 3759, 3827, 4308, 4350, 4363, 5031, 5087, 5627, 5637, 7555
 Organizational Design 697
 Organizational Factors 532, 534, 699, 3758, 7040, 7551, 7902
 Organizational Learning 351, 357, 555, 559, 900, 934, 2160, 2694, 3922, 4278, 4347-4351, 4357, 5027, 5094, 5330, 7515, 7517
 Organizational Management 690, 954, 2157, 5277, 5465-5466
 Organizational Trust 755
 Orthogonal Frequency-Division Multiplexing (OFDM) 5980

Orthosis 298-299, 304-305, 313
 Oscillator Circuit 1405
 Otaku 8016, 8020, 8022
 Other World 7076
 Outer Radiation Boundary Condition (ORBC) 1252, 1259
 Outsourcing 563-564, 821, 893, 1101-1103, 1105, 1109, 1115, 1157, 1159, 2341, 2805, 2920-2921, 2950, 2971, 4822, 4854-4855, 4858, 5306, 5317-5322, 5324-5325, 5407, 5476-5485, 5488-5489, 5518, 5527, 5532, 5552, 5562, 5577, 5919, 5922, 6717, 6725, 7229, 7259, 7261-7263, 7265, 7268, 7505, 7513, 8023-8026, 8035
 Over Fitting 498, 1763, 1829, 1834, 2133, 2136, 2141, 6645-6647, 6649, 6657, 6664-6665
 Overlapping Communities 6337, 6962, 6970
 Oversampling 1827-1828, 5973
 Overstaffing (Overage) 3712-3713, 3715-3716, 3718-3721, 3723-3724
 oximetry 6115, 6117-6118, 6122

P

PaaS Application 460
 Pace-Time 3472
 Pairing Based Cryptography 4977, 4984
 Paleoagriculture 3428, 3433
 Panchromatic 3493-3496, 3498, 3501
 Paper-And-Pencil Approach 7384-7387, 7391
 Papert 121, 232, 2392-2395, 2397, 2493, 2512, 3249-3250, 6380, 6581, 6859, 6861, 7939
 Parallel Computing 235, 238, 336, 338, 4014, 4024, 4583, 4593, 7731
 Parallel Decimal Multiplication 4618
 Parallel Metaheuristics 7695, 7697, 7700, 7703
 Parallel Programming 328, 337, 2008, 4014-4015
 Paralysis by Analysis 911
 parameter learning 2143
 Parametric Estimation Model 2153
 Parental Involvement 1733, 7098, 7102
 Parental Mediation 1733, 7097-7103, 7105
 Parental Mediation of Technology Use 1733
 Pareto Analysis 662, 664, 669, 3702
 Pareto Front 240-242, 250
 Pareto Optimal Solution 240, 250
 Parser 2420, 7452
 Partial and Temporary Unified Theory 4400-4401, 4403
 Partial and Temporary Unified Variation Principle 4401, 4403
 Participant Recruitment 7186, 7189

Index

- Participatory Culture 563, 565, 2284, 2293, 2302, 2316, 2319, 2321, 2325, 2857, 7004-7005
- Participatory Leadership 5196, 5203
- Particle Sphericity 1343
- Particle Swarm Optimization 6-7, 114, 239, 241-242, 247, 2180, 3107-3108, 4371-4372, 4377, 4595, 4600, 5968, 5974, 6595, 6913
- Parton 4667-4672, 4677
- Pass Band Droop 4746, 6043, 6055, 6171, 6177-6179, 6181, 6183
- pass band edge 6043-6044, 6173, 6183
- Passband 6009-6014, 6056, 6234, 6237, 6241, 6600
- passband droop 6009-6010, 6013-6014
- Passive Stereo 2415, 2420
- Path Computational Element (PCE) 6615
- Path Loss 6254-6256, 6259-6261, 6263, 6265, 6319, 6324-6325, 6332, 6624, 6673
- Pathfinding 1977-1979
- Pathological Gambling 3300-3301, 3308, 4224-4225
- Patient census 3711, 3713, 3715, 3719
- Patient Classification System 3713, 3723
- Patient Empowerment 5917
- Patient Protection And Affordable Care Act (PPACA) 2508, 3806, 3815
- Patient Satisfaction 3707-3708, 3784, 3789, 3793, 5813, 5864
- Patient Satisfaction Questionnaire Short Form (PSQ-18) 3784, 3793
- pattern matching algorithm 1019
- Pattern Recognition 4, 102, 120, 127, 166, 175, 180, 197, 232, 492, 494, 496-497, 500-501, 504, 851, 1008, 1074, 1333, 1340, 1354, 1838, 1988, 2188, 2191, 2193, 2201, 2393, 4130, 4937, 4942, 4950, 5108, 6042, 6205, 6478, 6580-6581, 6632, 6642, 6646, 6651, 7336, 8062
- Pattern Warehouse Management System (PWMS) 1377, 1379-1381, 1383
- Pattern-Warehouse 1383
- PAX 1949, 1961
- Payoff Function 17, 3230, 3233, 3236
- Payout Policy 61-62, 64-65, 67-68
- Paywall 2760
- PCI 1029, 3507, 4008, 4010, 4017, 4020
- PDPO 4909, 4918
- PDSA 731
- Peak-Load 3155
- Peak-Shifting 3155
- Pedagogical Efficacy 2663-2665, 2667-2668, 2671
- Pedagogical Elements 3295
- Pedagogical Models 2641-2642, 3983
- Pedophilia 1368, 1375
- Peer Attachment 1727, 1733
- Peer Coaching 7652, 7655
- Peer-To-Peer 1064, 1366, 1371, 1652, 1660, 2219, 2346, 2350-2351, 2355, 2784, 2815, 2869-2870, 2876, 3541, 3767, 3776, 4090, 4307, 4956, 5018, 5909, 6479, 6614, 6951-6952, 6954-6956, 6960, 7092, 7100, 7126, 7303, 7798, 7804
- Peer-to-Peer Platform 2351, 2355
- Peer-to-Peer System 6954, 6960
- Peer-Vetted Creative Production Approach 572
- Peltier cells 1283-1284, 1294-1295
- Perceived Ease Of Use 531-533, 833-835, 838, 840-841, 846, 2855, 3582, 3787, 3821, 4089, 4182, 4281, 6403
- Perceived Service Quality (PSQ) 1132, 1134, 1137, 1140, 2882, 2884-2885, 2887-2888, 2890-2891
- Perceived Usefulness 531-533, 833-835, 838, 840-841, 846, 1143, 1145, 2468, 2734, 3786-3787, 3821, 4182, 4281, 6022, 6024, 6184, 6403-6404, 7031
- Perception (P) 1140
- Perceptions of the Task 722-724, 726, 728
- Perceptron 8, 120-121, 232-233, 497, 1011-1012, 5177, 6580-6582, 6590, 6631-6632, 6643-6644, 6661
- Percolation 2673, 2680-2681, 3542-3543, 3548
- Performance Appraisal (PA) 4337-4344, 4346, 4913
- Performance Evaluation 7, 587-589, 593, 598, 600, 1008, 1772, 1774, 1777, 2149, 2439, 3060, 3509, 3784, 3882, 4337, 4777, 5032, 5646, 6622, 6629
- Performance Expectancy (PE) 2253, 2762-2763, 2765, 2768, 2770, 3581, 3584, 3586, 3590, 3655-3656, 3659-3661, 3666, 3787, 5095
- Performance Feedback 5596, 7516-7517, 7524
- Performance Indicators 357, 359, 581, 587, 593-594, 596, 598, 891, 933, 1541, 1882, 1892, 1894, 2169, 2854, 2955, 3133, 3191, 3885, 4779-4780, 4937, 5360, 5540, 5542, 5544, 5549, 5638, 5649, 5731, 6499, 6741, 6747
- Performance Measurement 44, 954, 1772, 1892, 2167, 2170-2172, 3882, 3971, 4774-4775, 4777-4781, 4784, 5358, 5360, 5363-5364, 5410, 5425, 5428, 5447, 5450-5451, 5538, 5540-5541, 5545, 5549, 5638, 5676, 5683, 5716
- Performance Metrics 356, 559, 1088, 1772, 1774, 1782, 3362, 3885, 3889, 5346, 5403, 5444, 6555, 6618, 6679, 7563-7564, 7573, 7576
- Performance Monitoring 1058, 3888, 5544, 5549, 7568, 7871
- Performance Testing 7506-7507, 7563, 7565-7569, 7572-7574, 7576

- performance testing tools 7506, 7567
 perfusion 6115, 6117-6118, 6122, 6150
 period length 1297-1299, 1302, 1304, 1306
 Periodic Review System 5339, 5344
 Peripheral Message 7788, 7795
 Perishable Product 5344
 Permanency 2546, 4169, 4171, 4173, 7672
 Perpetrator 78, 1695, 4163, 4169-4173, 4177, 4245, 4887
 personal community 1480-1481, 1486
 Personal Computers 382, 412, 414, 1075-1076, 1229, 1552, 2314, 2325, 2413, 2468, 2642, 3079, 3353, 3550, 4199, 4247, 4622, 5006, 5984, 6353, 6360, 6404, 6448, 7061, 7069-7070, 7108, 7116, 7248, 7258, 8096-8097
 Personal Data 2320, 2828, 2851, 3758, 4711, 4848, 4862-4869, 4871, 4909, 4911, 4916, 4918, 4931, 4961, 6069, 6083-6086, 6088-6090, 6222, 7112, 7195, 7528, 7757, 7761, 7800
 personal data intensive services 4862-4863, 4865, 4868, 4871
 Personal Identifiers Information (PII) 377, 1029, 3809, 4840, 4848, 5284, 5986, 5991, 6086, 6093
 Personal Learning Environment (PLE) 1528, 2546
 Personal Media 4307, 6348
 Personality Disorders 3272
 Personalized Healthcare Treatment 7774
 Personalized Learning 1517-1519, 1528, 2357-2358, 2384, 2546, 2556, 5792, 7673-7674, 7676, 7680
 Personalized Medicine 471, 1984, 2220, 3760, 5901-5905, 7774
 Personalized Recommendation Problem 1869, 1878
 Persuasive System Features 7785-7786, 7788, 7790, 7793, 7795
 Persuasive Technology 7787, 7795
 Pervasive Computing 1994, 3339, 4951, 5909, 5917, 6287, 6291, 7349, 7745, 7774, 8023
 pervasive knowledge 5908, 5911-5912, 5914, 5917
 Pervasive Service 7761, 7764
 Pervasive Technology 1071, 7754-7756, 7762, 7764
 PEST 911
 Petri Nets 789, 5434, 5436, 5438-5441, 6841, 6844, 7435-7437, 7439, 7444, 7488, 7490, 7502, 7731-7733, 7738, 7741-7742
 Pew Internet 1102, 1707, 3612, 7027
 PFLOPS 4016-4017
 Phantom Omni 4658, 4660, 6874
 phase response 6018, 6234, 6244
 Phasor Measurement Unit (PMU) 3065
 Phishing 78, 1685, 2565, 2876, 4838, 4842-4843, 4845, 5010, 5311, 6215, 7987
 Phonetics 1220, 2432, 2439
 Phonocardiogram (PCG) 1007, 1009, 1016
 photo messaging 6149, 6160, 6296-6298
 Photogrammetry 1277-1278, 1283, 1295
 Photogrammetry Images 1295
 Photon 6680, 7716, 7720-7722, 7724, 7730
 Photoplethysmography 6117, 6122
 Phreaking 6696-6697, 6699
 Phylogenetics 421, 423, 430
 Physical Attack 78, 86
 Physical Distance 743, 750, 1462, 2065, 2405, 2460, 5788, 5799, 5989, 7111, 7115
 Physical Therapy 300, 528, 3261, 3700, 5816, 5876, 5878, 5881, 5941-5942, 5944, 5947, 5950, 5956, 5958-5960
 Physician Resistance 1017, 1022, 3678, 3681-3682, 3684-3685
 Physiology 206, 1007, 1249, 1261, 1318-1319, 1330, 1354, 1540, 2121, 2405, 4125, 4134, 5110, 5751, 5835, 6070, 7883
 PIAAC Survey 4786-4789, 4792, 4796
 Pictorial Metaphor 5092
 Picture-Sharing Websites 5785
 Pillowing 4383, 4388
 Pilot 52, 80, 233, 544, 666, 1608, 2407, 2495, 2552, 2978, 3195, 3197, 3200, 3580, 3727, 3945, 3951, 5016, 5018, 5023, 5110, 5170, 5492, 5544, 5576-5577, 5958, 6106, 6113-6114, 6152, 6807-6808, 6933, 7185, 7216, 7219, 7683, 7756, 7789, 7891
 Piracy 1656-1662, 1665, 4838, 7265, 7988
 Pixel-Based Classification 7831, 7833-7834, 7839, 7842
 PK 1964, 1973, 7203, 7624, 7660, 7662, 7665-7666, 7670
 Place-Identity 4038, 4064-4065, 4071, 4076, 4420
 Place-Transition Net 7743
 plaintext 4927, 4982, 4984
 Planning Agent (PA) 2698-2699, 2705
 planning management 3407, 3460, 3881, 3896
 plastic surgery 6147-6149, 6151-6155, 6160
 Platform-As-A-Service (Paas) 887, 1027, 1102, 1104, 1129, 1148, 1356, 1365, 1627, 1637, 6730, 7532, 7767, 7778
 PMO 477, 480-482, 485, 489, 625
 PNPVillage 802-803
 Podcast 2446, 2651, 2660, 2665, 5265-5266, 5987, 6974, 7798
 Point Cloud 301, 1278, 1283-1287, 1295
 Point Density Analysis 3436, 3438, 3440-3441, 3443, 3447
 Poisson process 1572

Index

- Polarization 3349, 3354, 3541, 4786, 4796, 6680, 6758, 7063, 7717, 7721-7722, 7724, 7730
- Political Engagement 2316, 2319-2320, 3563, 3676, 6280
- Political Narratives 3647
- Political Participation 761, 1641, 2325, 3561-3562, 3565, 3570, 3613, 3676, 4715, 4719
- Pollutants 2405, 3068, 3076, 3157-3158, 3160-3161, 3163, 3165, 3373, 3380, 3489, 5363, 5470, 6119
- Polygonal Net 1295
- Population Census 3460, 7327-7328, 7331, 7336
- Population Metaheuristics 7695, 7703
- Portability of Explanation 2094
- Porter's View 5725
- Positive Learning Outcomes 3978
- Positive Social Cue 7979
- Post-Contract Problem 4892, 4918
- Post-Conventional Level 3203
- Posterior State 219
- Posthuman Body 3350-3351, 3355-3356, 3358
- Post-Implementation Problem 4891, 4919
- Postmodernism 3922, 6716, 6719, 6725
- Post-Operative 298, 524, 1838, 6153
- Post-Processing 501, 1284, 1332, 1760, 1988, 2911, 4383, 4389, 6657
- Post-Production 978-979, 982, 6447
- Postsecondary Education 3900, 3909, 3930-3931
- Potential Difference 2673, 6319, 6322, 6583, 6594
- Potentiator 1468, 1473
- Power Consumptions 4552, 6255, 6257-6258
- Power Distance (PD) 1530, 3021, 7956-7959, 7966, 8099, 8104
- Power Spectrum 221-222, 495-496, 1009, 1819, 1821
- PPG 6115-6118, 6122, 6676
- Pragmatics 1186-1188, 1190-1192, 1194-1195, 1518, 3934, 4103, 6567
- Pre-Conventional Level 3197, 3203
- Prediction Models 1795, 2085, 2087, 2091, 2213, 2216, 6624
- Prediction Step 219
- Predictive Analytics 334, 351, 387, 410, 847, 850, 861, 876, 887, 957, 1813, 2074, 2692, 3571, 5160, 5163-5164, 5541, 6806, 6907
- Predictive Modelling 182, 4642
- Pre-Processing 1760, 1765-1766, 1768, 1988, 2201, 4491, 4533, 4938, 6037, 6647, 6657, 7768, 8066
- prescriptive analytics 1806, 1813
- Prescriptive Gender Stereotypes 3383, 3391
- Pre-Service Teachers 3961, 6391, 6394, 7634, 7687
- Presidential Election 3612-3614, 3616-3617
- Presumers/Preconsumers 1644
- Presynaptic Membrane 6594
- Preventive Control Techniques 6443, 6446
- Prezi 1423, 1425, 1431
- Price Discrimination 2849
- Price Dispersion 2841-2842, 2844-2845, 2849, 7215
- Price Partitioning 2849
- Price-based Demand Response 3145, 3155
- Price-S 1398, 1405
- Pricing Model 71, 1627-1631, 1633-1634, 5062, 6547, 6731
- Primary User (PU) 6545
- Principal Challenges 3288, 3290-3291, 3622
- Principal Investigator 7189
- Prior State 219, 1086
- Privacy Act of 1974 3809, 3815
- Privacy Boundaries 6986, 6992
- Privacy Concerns Evaluation Instrument 4871
- Private Cloud 1028, 1032, 1104, 1144, 1154, 3465, 6737
- Private Information Ownership 6992
- Private Key 1240, 1650-1651, 1655, 2784, 2790, 2870-2873, 2880, 4907, 4928, 4930-4931, 4975-4981, 4983-4984, 6953
- Probabilistic Localization 6905
- Probing 62, 1051, 1531, 1534, 2284, 2450, 2572, 4483-4484, 4486-4487, 4489, 7503, 8002-8003
- Problem Based Learning (PBL) 2493, 2495, 2501, 5853-5854, 5862
- Problem Solving 98-103, 181-183, 202, 553, 563, 567, 579-580, 627, 629, 643, 1142, 1152, 1407, 1414, 1436, 1536, 1624, 2228, 2256, 2279, 2293, 2296, 2310, 2371, 2393-2394, 2398, 2401-2407, 2410, 2425, 2442-2444, 2467, 2492-2494, 2497, 2502, 2534-2535, 2537, 2548, 2560, 2591, 2630, 2633, 2643, 2647, 2811, 2825, 2827-2830, 2832, 2994, 3239-3240, 3243, 3248, 3250, 3261, 3311, 3403, 3622, 3794-3798, 3800-3801, 3804, 3871, 3880, 3964, 4362, 4518, 4527, 4583-4584, 4591, 4593, 4787-4788, 4891, 5033, 5652, 5656, 5756, 5800, 5802, 5832, 5940, 6336, 6489-6490, 6492, 6837, 6859, 6861, 6870, 7171, 7382, 7387-7389, 7391, 7393, 7396, 7398, 7514, 7541, 7548, 7630, 7673, 7675, 7696, 7699, 7881, 7941, 7971
- Problem Solving Environment (PSE) 4584, 4591, 4593
- Problem Solving Process 102, 580, 2403, 2828-2829, 3801, 5940
- procedural pain 5956-5959, 5965
- Process Assessment 6935, 6938, 7514-7515, 7517-7519, 7524
- Process Block 7427-7429, 7434

- Process Commonality 7431, 7434
 Process Improvement (SPI) 540, 1401, 1620, 3527, 3529-3530, 3537, 5050, 5574, 6927-6928, 6938, 7514, 7524, 7549-7552, 7555, 7557-7558, 7562, 7902
 Process Integration 3606, 3608, 5308, 5441, 5445, 5528
 Process Model 642, 645-646, 765, 773, 788-790, 792, 2982, 3317-3322, 3623, 3819, 5541, 5543, 6022, 6024, 6929, 7425, 7562
 Process Re-Engineering 887, 5434
 Process Transparency 755, 757, 764
 Process Variability 7431, 7434
 Process-Oriented Usability 5013
 Procurement 352, 530-531, 534-535, 587, 642, 789, 893, 1028, 3086-3090, 3092, 3094, 3096, 3435, 3529-3530, 3551, 3623, 3626-3627, 3683, 4851, 5296, 5305, 5389-5394, 5457, 5459-5462, 5464, 5504, 5552, 5555-5556, 5571, 5583, 5696-5697, 5704, 5985, 6266-6267, 6715, 6717, 6719
 Product Design And Development 505, 5292, 5297, 7132, 7134
 Product Substitution 5335, 5344
 Product Sustainability 617
 Productive audiences 7108
 Productivity 230, 235, 266, 278, 542, 550, 553-554, 557, 568, 587, 589-590, 598, 642, 688, 707, 714, 735, 743-745, 774, 820, 827, 864, 880, 888, 923, 925, 952, 983, 990, 1065, 1074, 1376, 1407, 1699, 1883, 1923, 2238, 2245, 2256-2257, 2403, 2739, 3000, 3022, 3184, 3253, 3338-3339, 3422, 3429, 3602, 3678, 3681-3682, 3699, 3721, 3748, 3793, 3809, 3963, 3966, 3969, 4014, 4128, 4201, 4316, 4318-4321, 4326, 4328, 4339-4340, 4344, 4349, 4351, 4359, 4363, 4369, 4622, 4709-4710, 4785, 4838, 4864, 4881, 5006, 5018, 5024, 5029, 5051, 5147, 5273, 5352, 5422, 5434, 5477, 5527, 5571, 5596, 5598-5600, 5641, 5651, 5659, 5676, 5689, 5727, 5733, 5852, 5985, 5988, 6095, 6222, 6389, 6678, 6682, 6725, 6739-6741, 6744, 6776, 7215, 7220, 7224, 7234-7236, 7240-7241, 7244, 7247, 7259, 7382, 7454, 7508, 7549, 7573, 7605, 7625, 7696, 7778, 7800, 7831, 7874, 7982-7983, 7986, 7993, 8066, 8097
 Product-Oriented Usability 5013
 Produsage 4149, 4151, 4153, 4156
 Professional Community 1476, 1481, 1486, 5095, 7348
 Professional Development 269, 1420, 1439, 1883, 1885, 1894, 2238, 2247, 2255-2257, 2298-2299, 2380, 2406, 2424-2425, 2427, 2495, 2561, 2604, 2667, 3386, 3798, 3830, 3839, 3844, 3938, 3956, 4188, 4351, 5047, 5258, 5834, 6413, 6490, 6859, 6863, 7079, 7173, 7252-7253, 7630, 7635-7637, 7640, 7642, 7648, 7652, 7660, 7663, 7665, 7900, 7912, 7917, 7924, 7930
 proficiency test 5183-5184, 5190, 5192
 Profiling 440, 454, 589, 1519, 1564, 1620-1621, 1976, 2383, 2487, 3450, 3861-3863, 3866, 3869, 5170, 6032, 6034-6037, 6039
 Profitability 2, 62-63, 73, 505, 578, 715, 914, 923-925, 929, 1165, 1168, 1584, 1594, 2170, 2978, 3020, 3464, 3637-3638, 4434, 4717, 4858, 5062, 5306, 5337, 5346, 5362, 5392, 5394, 5427, 5460, 5719, 5743, 5991, 6546-6547, 6711, 6725, 6820, 7143
 progenitor 1049-1052, 1061
 Prognosis 413, 1794-1802, 1804-1805, 2220, 3455, 5870, 7766
 prognostic factor 1795-1796, 1801-1802, 1805
 prognostic tool 1795, 1797, 1799, 1802, 1805, 2218
 Program Mutation 7469
 Programming Language 204, 368, 410, 1249-1250, 1257, 1259, 1406, 1995, 2032, 2039, 2042, 2047, 2392, 2397, 2401, 3273, 4008, 4736, 6061, 6260, 6429, 6596, 6793, 6837, 7392-7393, 7395, 7398, 7442, 7451, 7459, 7473-7474, 7476, 7478, 7480-7481, 7508, 7565, 7737, 7906, 8088, 8094
 programming technology 7742
 project configuration 1939
 Project Control 734, 981, 1938, 5679, 5682-5683, 5689, 5693
 Project Culture 784, 787
 Project Implementation 621, 3846, 5118, 5321, 6938, 7517
 Project Lifecycle 625, 989, 1936-1937, 5698, 5704
 Project Management 620-621, 624-625, 630, 632, 635-637, 690, 692, 734, 777, 780-781, 784, 800, 1405, 1596, 1934-1936, 1938-1939, 1943, 2157, 3529, 3531, 3621-3630, 3633, 3635, 3756-3760, 3766, 3840, 3846, 3849, 4566, 4849, 4851, 4858, 4860, 5478, 5607, 5650-5653, 5679, 5682-5683, 5689-5690, 5692-5694, 5699-5700, 5704, 6088, 6927, 6930, 6932-6934, 6938, 7551, 7556, 7584, 7799, 8040
 Project Management Success 3766
 project monitoring 1938
 Project Performance 1934-1935, 1939, 5679-5680, 5689-5690
 Project Success 784, 1934-1935, 3622, 3624, 3626-3627, 3629-3630, 3757-3759, 3766, 5650, 5690-5693, 5695, 5698-5700, 6929, 7423, 7431, 7485
 Project Success Criteria 3766, 5695
 Project Team 543, 622, 630, 635, 730, 777, 781-784, 787, 1938, 1940, 1946, 2966, 3626-3627, 3830,

Index

- 3840, 3844-3845, 4860-4861, 5048, 5650-5652, 5656-5657, 5680, 5684-5685, 5687, 5695, 7372, 7403-7404
- Proof System 7611, 7621
- Proof-of-Work 2815, 2870, 2872, 2880
- Propagation Delay 6609, 6616, 6678-6679
- Property-Oriented Specification 7621
- Proprietary Software 893-895, 898, 1012, 3463, 4145, 5983, 7369, 7373, 7377, 7381, 7904, 7908, 7912, 7915
- Proprietary Technology 7153, 7160
- Prosocial Behaviors 7077-7079, 7081-7082, 7087
- prospect decision-making factors 3864, 3866
- Prosthetics 492-494, 496-497, 499-502, 524, 5110, 5886, 5889, 5893-5894, 5897, 5931
- Prosumerism 563, 6696-6697, 6708
- Protection of Children Online 4964
- Protein Structure 420, 424, 7696
- Protest 3645, 6187, 6275-6281, 6285, 6995, 7296
- Protester 6285
- Prototyping 301, 505-507, 517-518, 520, 689, 2418, 2420, 2627, 3286, 3327, 3839, 4382, 4389, 5108, 5609, 6890, 7384-7385, 7867
- Prototypology 5079, 5084
- proxy data 3114, 3118, 3130
- Proxy re-encryption 1668, 1675
- Proxy signature 1668, 1675
- Pseudo-Dynamics Characteristics 4651
- Pseudorandom Number Generator (PRNG) 1239-1240, 1243, 1247, 1647, 1655
- PSO Algorithm 241-243, 247, 249-250
- PSP 5412-5418, 6594
- psychiatry 7017-7019, 7022
- Psychoanalysis 3023, 7014, 7016-7020, 7022, 7046
- Psychobiology 5771, 5774
- Psychological Acceptability 5006, 5013
- Psychological Appraisal 4346
- psychological challenges 6398-6401, 6405, 6411
- psychological reactance 3821, 3825
- Psychological Reactance Theory 3821, 3825
- Psychological Time 6411
- Psychological Well-Being 1707, 3770, 3772, 7029, 7035
- Psychology 132, 156, 181, 382-383, 1318-1319, 1330, 1353-1354, 1685, 1699, 1723, 1783, 2108, 2121, 2273, 2413, 2439, 2493, 2603, 2615, 2852, 2856, 3000, 3023, 3260, 3262, 3290, 3292, 3297, 3638, 3647, 3819, 3821, 3823, 3831, 3833, 4031, 4042, 4064, 4072, 4125, 4134, 4153, 4180, 4188-4189, 4200, 4203, 4236, 4278, 4280, 4318, 4327, 4363, 4404-4405, 4407, 4414, 4422-4423, 4434, 4460, 4493, 4516, 4537, 4653, 4676, 4717, 4720, 5086, 5149, 5156, 5160, 5237, 5251, 5626, 5634, 5651, 5770, 5774, 5889, 6131, 6185, 6225, 6353, 6378, 6398, 6400, 6411, 6634, 6860, 6945, 7019-7020, 7026, 7052, 7091, 7219, 7503, 7519, 7608, 7706, 7764, 7799, 7856, 7884
- Psycho-Pedagogy 2603, 2615
- Psychophysiology 4127, 4134
- Psychosis 3272
- Psychosocial Adjustment 4173, 4254
- Psychotherapy 99, 2621, 2624, 3260-3262, 3265, 3272, 4226
- PTS 3170, 5814, 5967-5968, 5972-5976, 6850-6852, 7623-7627, 7629-7630
- Public Administration 759, 761, 2861, 3461, 3528, 3574, 3580, 3591-3592, 3833, 4318, 4714, 4817, 5635, 6479, 6711, 6725, 7204
- Public Choice Theory 6714, 6726
- Public Cloud 377, 887, 1027-1030, 1032, 1104-1105, 1144, 1629, 1668, 2817, 2822, 3465, 4013, 6731, 6737
- public data resources 420
- Public Domain 398, 2481, 2485-2486, 2488, 3461, 4427, 4441, 4693, 5094, 5228-5229, 5231-5232, 5236, 6354, 6783, 6941, 6987
- Public Key 1651, 1655, 2715, 2722, 2784, 2815, 2870-2873, 2880, 4899, 4903, 4930, 4975-4976, 4980-4984, 6560, 6953, 7724
- Public Key Cryptography 2870, 2880, 7724
- public key encryption 4899, 4975, 4982-4984
- Public Libraries 2248, 4709, 5205, 5229, 6685, 6687-6688, 6690
- Public Policy 115, 1037, 2220, 2238, 2330, 2339, 3329, 3601, 3639, 5024, 6710-6711, 6714, 6716, 6719, 6726, 6729, 6737, 7236, 7983, 7985
- Public Scholarship 6780, 6785, 6787
- Public Sector 40, 61, 588, 600, 698, 828, 888, 895, 900, 1028, 1033-1034, 1037, 1541, 1544, 2337, 2805, 3086, 3528, 3532, 3550-3551, 3591-3595, 3601, 3633, 3654, 3783, 4101, 4439, 4871, 5019, 5319, 5675, 6186, 6710, 6919, 6921-6923, 7220, 7260
- Public Service 280, 1698, 2230, 3531, 3593, 3601, 3654, 4711, 4839, 5041, 6818, 7220, 7229, 7271, 7275, 7629
- Publish/Subscribe Paradigm 6565
- Pulmonary Arterial Hypertension (PAH) 1006, 1016
- Pulse Oximetry 6117-6118, 6122
- Purposeful Commitment 2663, 2665-2667, 2671
- Purposeful Organization 2663-2664, 2666, 2671
- PVDF 3046, 3050-3051, 3055

Q

- Qassim University 7915
 QGDH 1668, 1670, 1674
 QIP 7562
 QKD 7715, 7720-7722, 7724
 Q-Q Plot 1861, 1868
 Qrcode Format 2546-2550, 2552, 2556, 7673, 7675-7676, 7681
 qualitative “measure” 2030
 Qualitative Data 1424, 1607, 1715, 2494, 2496, 2654, 3956, 4480-4482, 4484-4485, 4487, 6758-6759, 6761, 6766, 6862-6863, 7519
 Qualitative Methods 1607, 2514, 3670, 4480-4481, 4484, 4487, 5451, 5639, 6756, 6759, 6770, 7649
 Qualitative Research 811-812, 1607, 2517, 3344, 3672, 3900, 3905, 3923, 4480-4482, 5319, 5451, 5646, 5661, 6744, 6756-6761, 6766, 7051, 7916
 Quality Attribute 759, 1405, 7590
 Quality Evaluation 2020, 2024, 2887, 4416, 5251, 8007
 Quality Function Deployment (QFD) 730, 739, 2006, 2019
 Quality Improvement 564, 1625, 2098, 3173, 3182, 3526-3528, 3530, 3532, 3534, 3806-3807, 5574, 7562
 Quality Management 735, 1619, 2169, 3527-3528, 3627, 3694, 3737, 3806, 3881-3882, 3884-3885, 3899, 4851, 5049, 5274, 5280, 5361, 5707, 7573
 Quality Management System (QMS) 3882, 3884, 3899
 Quality of Benders Cut 5421
 Quality of Business (QoBiz) 6546, 6549, 6554
 Quality Of Experience (Qoe) 1164, 6270, 6274, 6442, 6446, 6477, 6546
 Quality Of Life (Qol) 279, 287-288, 290, 297, 525, 528, 996, 998, 1001, 1006, 1071, 1984, 2527, 3082, 3187, 3580, 3640, 3694, 4128, 4180, 4225, 4621, 4712, 4765, 5293, 5305, 5357, 5852, 5864, 5890, 5933, 6185, 6188, 6190, 6320, 6393, 6776, 7028, 7031, 7092, 7226, 7756, 7760, 7764-7765, 7779
 Quality Of Service (Qos) 349, 765, 1096, 1129-1131, 1134, 1138, 1164, 1181-1182, 1184, 1229, 2142, 2150, 2350, 2921, 3554, 3582, 3654, 4098, 4550-4551, 4553, 5479, 6097, 6255, 6266-6267, 6269-6270, 6272, 6274, 6442-6443, 6546, 6555, 6557, 6561-6562, 6609, 6616, 6618, 7588, 7593, 7598, 7696, 8013
 Quality of the Explanation 2094
 Quantified Intuition 682, 689-690, 692
 Quantified Self 6122
 quantile estimation 1856-1857, 1866
 Quantile Function 1304, 1306, 1858, 1868
 quantitative “measure” 2030
 quantitative content analysis 3976
 Quantitative Data 587, 1423, 2286, 2496, 2514, 4484, 5120, 5122, 6758, 6766, 6863
 Quantitative Research 1131, 1536, 1607, 3086, 3658, 3760, 3900, 3905, 4480, 4963, 5646, 6756-6758, 6760-6761, 6766, 7689
 quantum algorithm 7718-7719, 7724
 Quantum Bit (Qubit) 1044, 1046, 1048-1052, 1056-1057, 1061, 4398, 4451, 4458, 7715-7720, 7722-7723, 7730
 Quantum Circuit 1053, 7715, 7717-7719, 7723
 Quantum Computers 1056, 4453, 4898-4899, 4906, 4908, 7715, 7719, 7723
 Quantum Computing 186, 1040, 1045, 1050-1051, 1053, 1058, 4009, 4448-4449, 4451-4453, 4455, 4458, 4899, 7715, 7717, 7719, 7723-7724
 Quantum Cryptography 4451, 4453, 4899, 7722, 7724
 quantum entanglement 1040, 1044, 1046, 1050-1052, 1057-1059, 1061
 Quantum Error 4451, 4455, 7720, 7723
 Quantum GATE 1049, 7719-7720, 7730
 Quantum Information 1045, 1051, 4391, 4394, 4398, 4448-4455, 4458, 4725, 4727, 7715
 Quantum Information Science 4391, 4448-4450, 4452-4455, 4458, 4725, 4727
 Quantum Machine 1061
 quantum phenomena 1057
 quantum programming 7715, 7717, 7719, 7724
 quantum protocol 7715
 quantum teleportation 1044-1045, 1055, 1061
 Quantum Tunneling 2684
 Query by Example (QBE) 2033, 2042
 Query Processing 332, 1376, 1948, 1951-1952, 1955-1956, 2057, 3409, 4540
 Query, View, Transformation (QVT) 7448, 7458
 Question Answering Systems 4471, 4474, 4479, 7800
 Questioned Document Examination 4651
 queueing system 1570-1572, 1576, 1581, 1583
 Quick Response (QR) 2546, 2556, 3521, 3525, 6701, 6709, 7672, 7681, 7777, 7813
 Quick Response (QR) Code 2547-2548, 2550-2551, 2556, 3521-3522, 5284, 5520, 5526, 5604, 6465, 6709, 7674, 7681, 7777, 7808, 7813
 Quorum 1095-1096, 1238, 1240, 1244, 1247, 1571, 4928, 4931, 4935

R

- Race Condition 4003
 Radio Access Network (RAN) 6100, 6105, 6253, 6265

Index

- Radio Frequency Identification (RFID) 1994, 2007, 2019, 2814, 3689, 3691, 4201, 4216, 4951, 4961, 5340, 5344, 5352, 5490, 5516-5517, 5519, 5559, 5562-5563, 5565, 5569, 5575, 5918, 6676, 7674, 7775, 7784, 7806, 7813
- Radio Resource Management (RRM) 6266-6267, 6269, 6272, 6274
- Radiometric Resolution 3485, 3501-3502
- radix-2 trellis 6314
- Ragnarök 7343-7344, 7346
- RAISE Method 7611-7612, 7614
- Random Network 6516-6517, 6521, 6972
- Random Number 1240-1244, 1247, 1297-1302, 1304, 1306, 1648-1649, 1827, 1845, 1848, 2872, 3107, 3995, 4374, 4979, 5969, 7722
- Random Number Generators (RNG) 1240-1241, 1243, 1297-1299, 1301-1302, 1304, 1306, 1845, 2872, 3107, 7722
- Random Oracle Model 1668, 4975, 4983-4984
- Randomized Decision Forest (RDF) 5885
- range statistics 1857, 1859, 1868
- ranking and selection 1856-1857, 1861, 1866, 1868
- Ransomware 6215
- Rapid Prototyping 689, 2418, 4382, 4389, 5108, 6890, 7384
- Rasch Measurement Theory 3874
- Raspberry Pi 5854, 5856, 5862
- raster grid 170-174, 179
- Raster-Stereography 166, 169, 171-175, 177, 179
- Rational Choice 4719, 6710, 6714, 6726, 7195
- Rational Unified Process (RUP) 1398, 1405, 7402
- RDA 6750-6751, 6753, 6755
- Reaction Time 1230, 1294, 6126-6130, 6135
- Reactive Control Techniques 6446
- Reading Grade Level 1503-1504
- Reading Level 1499-1505, 1507, 3948, 3953
- Real Contact Area 434, 439
- Real Hypertext Links 6433
- Real-Time Networks 1237
- Real-Time Process Algebra 7704, 7707-7708, 7711
- Real-Time Translation 3732, 6812
- Real-World Problems 181, 240, 7693, 7695-7698, 7701
- Receiving Operating Characteristic (ROC) Curve 2213, 2215-2220, 2223
- Receptor Sites 6594
- Reciprocity 89, 1595, 3035, 3771, 3923, 3925, 3967, 4202, 5095, 6190, 6336, 6698, 6704-6705, 7078, 7163
- Recommender Systems 359, 1735-1740, 1745, 1757-1762, 1765, 1767-1768, 1770, 1869-1873, 1875, 1879, 2819-2820, 6032-6035, 6038, 6430, 6659, 7117
- Reconfigurability 4018, 4024, 4554, 6054
- Reconfigurable Computing (RC) 4018-4019, 4021-4025, 4029
- Reconfigurable Hardware 4018, 4029
- Reconstruction of Historical Landscape (ROHL) 3485, 3489-3490, 3501
- Reconstructive Surgery 6149, 6154, 6160
- Record Linkage 354, 1617, 1793
- Recovery of Function 528
- Recurrence or Recursion 1333, 1343
- Recurrent Neural Networks 6581, 6594, 6658-6659
- Recursive Form 4749, 4751, 4757-4758, 4760, 6008, 6014, 6171
- Recycling 77, 543, 2627, 3046-3047, 3066-3067, 3070-3072, 3076, 3079, 3081-3082, 3088, 3094, 3132, 3371, 5360, 5362, 5379, 5381, 5383, 5428
- red-blue eyewear 2416, 2420
- RedLeaf Communications 812, 819
- Reduce Software Testing Costs 7513
- reduced form 1220-1221, 1223, 1225
- Redundancy 365, 441, 852, 861, 1085-1086, 1094-1095, 1106-1107, 1227, 1231-1232, 1234, 1237, 1239, 1384, 1394, 1840, 1903, 1991, 2031, 2059, 2121-2122, 2124-2125, 2131, 2264, 2268, 2619, 2866, 3898, 4549, 4576, 4610, 4885, 4890, 4928, 4936, 5009, 6120, 6502, 6523, 6531, 6548, 6853-6854, 6952, 8056
- Reengineering 874, 888-889, 891-893, 895, 898, 1399, 2966, 2978-2979, 3533, 3592, 3605, 3621, 5054, 5402, 5433-5434, 5444, 6996, 7402, 7422, 7429, 7447-7449, 7452
- Reference Architecture 581-582, 586, 636, 1229, 7572
- Reference Library 5068, 6141, 6146, 6688
- Reference Model 360, 573-574, 579-580, 582-584, 586, 1774, 3527, 3532, 3537, 5402, 5434, 6260, 6429, 6566-6567, 6619, 6688, 6933, 7436, 7562, 7707
- Reflective Light 6116-6117, 6122
- Reflective Practices 1530-1536, 5153, 7922, 7924, 7926, 7930
- Reflective Practitioner 5153, 5157
- reflective vision 6116
- Refrigeration 1295, 7808
- Reggio Emilia Approach (REA) 7625, 7634
- Regional Communities 7869-7871, 7873-7876
- Regression Analysis 2, 476, 1134, 1544, 2307, 2472, 2611-2612, 3452, 3459, 3980, 4851, 5400, 5776, 6640
- regression tree 2094, 2132-2134, 2136, 2141
- Rehabilitation Engineering 528
- rehabilitation exercises 5878-5879, 5885, 5950
- Relationship Management 350, 604, 923, 929, 932, 941, 945, 1405, 1584, 1594-1596, 1604-1605, 1615,

- 1616, 1619, 1626, 1880, 2086, 2169, 2687-2688, 2696, 2854, 2860, 2933, 2953, 2977, 4054, 4441, 4826, 4835, 4909, 5353, 5444, 5559-5562, 5569, 5750, 5775, 6985, 7131
- Relative Risk 1371, 5998-6000, 6005-6006
- Relaxed R-Checkpoint Algorithm 2153
- Reliability Engineering 6826-6827, 6858
- Reliable/Prompt Response 2886, 2896
- religious communities 6161-6164, 6167, 6170
- Reluctant Confidants 6986, 6992
- REM Sleep 205
- Remote Access Trojan (RAT) 4839, 4848
- Remote Center of Motion (RCM) 6874-6876, 6881, 6884
- remote consultation 5016
- Remote Deletion/Wiping 4310, 4314
- Remote Sensing 1318-1319, 1330, 1354, 1414, 1825, 1835, 2008, 3467, 3475, 3484-3485, 3487-3488, 3490-3491, 3494, 3496-3498, 3501-3503, 3514, 5932, 6117, 7831, 7839, 7842
- Renewable Energy Integration 3113
- Renewable Energy Sources 1294, 3103-3104, 3113, 3147, 7749
- replenishment policy 5335, 5341
- Replica Consistency 1093, 1095, 1099, 1238-1240, 1244, 1247, 1651-1652, 4936, 8061, 8065
- reproduction literacy 2226-2227, 2234, 2281
- Request for Quotation 530, 538
- Requirements Engineering 1874-1875, 4811, 5127-5130, 5132, 5135, 5318, 7400, 7402, 7407, 7411-7412, 7419, 7421-7422, 7424-7425, 7434, 7528-7529, 7531, 7577-7578, 7583-7584
- Requirements Engineering Process 5127, 5130, 5132, 5135, 7411-7412, 7422, 7424-7425, 7434
- Requirements Model 4821
- requisite competence 4884-4885, 4890, 4896
- Research Approaches 3322, 3672, 4404, 6770, 6777-6778, 7771, 8092
- Research Methodology 52, 126, 1129, 1551, 1901, 2286, 2494, 2806, 3851, 4102, 4190, 5609, 6767, 6772, 6777-6778, 6862, 8069
- Research Methods 811, 964, 1684, 2078, 2577, 2957, 3340, 3344, 3587, 4416, 4480-4482, 4518, 4817, 5183, 5199, 5238, 5661, 5771, 6027, 6077, 6756-6757, 6759, 6761, 6772-6773, 6777, 7098, 7689
- Research Model 51, 530-532, 535, 631, 645, 720-721, 837, 2857, 2954-2955, 2957, 2959-2960, 2968, 3581, 3584, 3586-3587, 3659, 5428, 8070, 8072
- Research Paradigm 126, 3900, 3905, 4961, 6756, 6758, 6760
- Research Problems 100, 120, 1332, 1667, 1930, 6522, 6778, 7488, 7942
- Research Trends 1695, 2064, 2070-2071, 2110, 4316, 4857, 5407, 6195, 6631
- ResearchGate 7044, 7056
- Reseller Piracy 1658, 1665
- Reserve Prices 535, 538
- Residual Risk 629, 4849, 4853-4857, 4861
- Resilience Engineering 1411
- resistance theories 3821
- Resource Block (RB) 6268, 6274
- Resource Efficiency 3159, 4550, 5360, 5366
- Resource Management 743, 1880, 1952, 2169, 3419, 3430, 3447, 3494, 3501, 3526, 3529, 3756, 4337-4338, 4347-4349, 4357, 4954, 5207, 5329-5330, 5916, 6266-6267, 6269, 6272, 6274, 6490, 7519
- Resource-Based View (RBV) 4349, 4776, 5714, 5725
- Response Time 50, 1126, 1181, 1772, 1774-1775, 1952, 2805, 3060, 3898, 4372, 4529, 4800, 4804, 4885, 4998, 5022, 5372, 5997-5998, 6004, 6033, 6126, 6128-6129, 6150, 7564-7565, 7575, 7591, 7593, 7598, 7603
- Responsible Innovation 4273, 4277
- Responsive Design 315, 1476, 2660, 4215, 4222, 6072, 8087
- Responsive Pricing 6546-6547, 6549-6550, 6552, 6555
- Responsive Web 8079-8085, 8087
- Responsive Web Design 8079-8082, 8084-8085, 8087
- Restful Services 464, 2778
- Restful Web Services 460, 3286
- Restriction Mediation 7105
- Result Rate 3369
- Revealed Comparative Advantage Index 7247
- Revenge Porn 4163, 4167
- revenue model 7139
- Revenue Sharing 1164-1165, 1168-1170, 1173, 7130, 7139
- Reverse Engineer Requirements 7513
- Reverse Engineering 301, 424, 521-525, 528, 1284-1285, 1287-1289, 1294, 2059, 7447-7450, 7452-7455, 7457-7458, 7500
- reverse logistics 5362, 5366, 5378-5379, 5381, 5386, 5600
- review sites 5776, 5778-5780, 5782-5783, 5785, 6019
- RF Emissions 6136, 6146
- RFID Technologies 5284, 5492, 5526, 5581, 5596, 5605, 5918, 5921, 5923, 5929
- RFID-Embedded Technologies 5284, 5526, 5577-5578, 5581, 5597, 5601, 5605, 5924, 5929
- RGB Camera (RGB-Cam) 6197, 6206

Index

Rich Media 4150, 4542, 5263-5264, 5855, 6477, 7007, 7796-7798, 7801, 7804, 8037-8038, 8045
Richard Clark 6377, 6380
Richard Dawkins 7111, 8015-8016
Risk Assessment 1, 15, 620, 629, 825, 827, 1019, 1359, 1369, 1796, 2069, 2170, 2718, 4573, 5511, 5691, 6789-6792, 6800, 6803, 6805, 7530
Risk Categories 4974, 6085
Risk Communication 5087, 6212-6213
Risk Management 1-2, 4, 14, 38, 78, 620, 642, 666, 783, 823-825, 968, 1397, 1402, 1411, 2171-2172, 2693, 2717, 2922, 3088, 3092, 3447, 3528-3530, 3641, 4573, 4853, 4962, 5296, 5308, 5312, 5316, 5398, 5450-5452, 5542, 5550-5551, 5553, 5556, 5699, 6633, 6789-6791, 6802, 6820, 6923, 7531
Risk Monitoring 825
Risk Perception 3653, 4573, 5459, 6918, 6923
Risk Regulation Regime 6726
Risk Society 4404, 4764, 4766, 4773
RNA Interference 477-478, 489
RNA Therapeutics 477, 480, 482, 485, 490
RNAi Therapeutics 477, 484-485, 490
ROA 5061-5063, 5066, 5642, 5644, 5672-5673, 6148
Robert Kozma 6377, 6380
Robot Kinematics 6883, 6893
Robotics 99, 105, 107, 202, 234, 383, 683, 1278, 1407, 1920, 2492-2495, 2497-2498, 2501-2502, 3276, 3284, 3349, 4203, 4236, 4267, 4653, 5897, 6847-6848, 6856, 6859-6861, 6863-6866, 6870, 6872-6875, 6877-6879, 6882-6883, 6885, 6887, 6889-6891, 6895, 7026, 7778, 7985
Robotics Academy 2494, 2497, 6862-6863
Robotics Package 2493-2494, 6860, 6863-6865, 6870
Robotics Packages 2494-2495, 2497, 2502, 6862-6863, 6865-6866, 6870
Robustness 114, 181, 371, 375, 637, 730, 1240, 1276, 1298, 1311, 1376, 1669-1670, 1963, 1968-1969, 2057, 3062, 3109, 4010, 4139, 4873, 4880, 4883, 4947, 4985, 4991, 4997-4998, 5003, 5692, 5697, 5881, 6131, 6198-6199, 6500, 6506, 6524, 6531, 6561, 6632, 6647, 6883, 7443, 7506, 7696, 7700, 8068
Role Players 782, 6966, 6972
Role-Playing Games 803, 808, 3265, 4223, 4233, 4298
ROMIO 1117-1121, 1123-1126, 1128
Ross Ulbricht 2862, 2868
Rotated Comb 4760
Rough Relations 2212
rough set theory 1835, 1837-1838, 1840, 2200-2202, 2208, 2212
Rough Sets 184, 587, 1835-1840, 1842, 1844, 2200-2201, 2206

Routing Protocol 6499-6501, 6505-6506, 6526, 6529, 6537, 6610
RS 234, 593, 1735, 2215, 2392, 2763, 2885, 3088, 3197, 3279, 3490, 3503, 3510, 3512, 3514, 4747-4754, 4760, 5378, 5381, 5913, 5974, 6260, 6504
RS Filter 4747-4749, 4751-4753, 4760
Rss Service 2486, 6973, 6976, 6984
RSVP-TE 6612, 6617
Rubrics 270-271, 2237, 3255, 3396, 6863-6864, 7683-7684
Rudolf Franz Flesch 1499

S

Safora 6848-6849, 6858
Sage 1398, 1405, 1513, 2589, 5157, 5800, 5804
Sagittal Plane 299, 5942, 5948, 5954
Salwicki Net 7743
Sampling Rate 3062, 3065, 4746, 6008, 6014, 6043, 6056, 6138, 6171
Samsung 215, 4010, 4116, 5570-5578, 5982, 6101, 6141, 6154, 6663, 7369-7370, 7372-7374, 7376-7377, 7380
Sandra Petronio 6985
SAP ERP 5644, 5646, 5649
SAS 854, 953, 1107, 1910, 2215, 2702, 6648, 7140, 7144
Satisficing Game 7283-7284, 7291
Scalability 326, 329, 335, 365, 376-377, 406, 457, 876, 1030, 1053, 1077-1078, 1091, 1094-1095, 1108, 1119, 1121, 1125, 1179, 1234, 1356, 1358, 1408, 1677, 1772, 1774-1775, 1806-1810, 1813, 1873, 1954, 1981, 1989, 2008, 2350, 2817, 2822, 2824, 2870, 2921, 3109, 3409, 4014-4015, 4549, 4930, 4957, 5226, 5307, 5311, 5527, 5543, 6097-6098, 6106, 6108-6109, 6211, 6533, 6549, 6556-6557, 6559-6560, 6562, 6677, 6730-6731, 6951, 6993, 7226, 7259, 7265, 7454-7455, 7506, 7563, 7575-7576, 7699, 7766, 7768, 7799
Scalable Production 4383, 4389
Scale-Free Network 6521
Scaling 328, 1776, 2189, 4012, 4210, 4214, 4299-4300, 4302, 4339, 6008, 6018, 6770, 6879, 6951, 7720, 7799
scanner pen 6464-6466, 6475
Scene Graph 2420
Scene Modeling 1347-1348
Scheduling 186, 340, 540, 542, 622, 625, 629, 725, 868, 984-985, 1181, 1232, 1570, 1952, 2953, 2956, 3108, 3404, 3621, 3624, 3630, 3697, 3699, 3713, 3721, 3786, 3995-3996, 4369-4372, 4374, 4376-4377, 4381, 4851, 5287, 5400, 5411, 5415, 5418,

- 5436, 5505, 5511, 5515, 5518, 5552, 5573-5574, 5582, 5788, 5920-5921, 5991, 6268, 6270, 6446, 6531, 6611, 6613, 7508, 7696, 7801, 7870, 7925
- Schema Change Propagation 2052, 2057, 2063
- Schema Change Semantics 2053, 2063
- schema changes 2043-2046, 2048-2049, 2053-2056, 2058, 2063
- schema evolution 2026, 2043-2050, 2053-2054, 2057-2058, 2060, 2063
- schema version 2052, 2054-2058, 2063
- schema versioning 2054-2061, 2063
- Scholarly Networks 7059, 7068
- Scholarship 251-252, 944, 946-947, 2227, 2231, 2484, 2504, 3176, 3672, 4439, 5184, 5228, 5232-5233, 5236, 5254, 5860, 6161, 6355, 6748-6749, 6779-6781, 6783-6785, 6787, 7059, 7894
- School Activity 6471, 6475
- School Adjustment 2618-2619, 2625
- School Climate 1727, 1733, 7647, 7652
- School Integration 2616-2617
- School Readiness 6691
- Schumpeterian Entrepreneurship 3007
- Science and Technology Institution 4784
- Science Animation 2599, 2605, 2613, 2615
- Science Education 1422-1423, 3291, 4698, 5103, 5108, 5802
- science gateways 1063, 1069-1071, 1074
- Science mapping 3473-3475, 3478-3480
- Science, Technology, Engineering, And Mathematics (STEM) 2643, 3311, 3316, 5109, 6376, 7899
- Scientific Literacy 3248
- Scientific Method 1851, 1855, 2492, 4004, 4209, 6767, 6859
- Scientific Simulation 4593
- Scientific Visualization 5106-5107, 5109, 5114
- Scientometric Indicators 6739-6741, 6744, 6747
- Scientometrics 3483, 6739-6740, 6747, 7051
- SCOR model 5400, 5402-5403, 5434-5436, 5440-5441
- Scorecard 1596, 2171, 2176, 2219, 3530, 5058, 5545, 5638, 5714, 5716-5719, 7140-7142, 7146
- Screen Culture 4234, 4255-4256, 4260, 4262
- Screen MVR 6233
- Screen Reader 316, 321, 2523, 3517-3518, 3525
- Screencast 1548, 1550-1551, 1556, 1558, 2660
- ScreenCode 6465, 6467, 6469-6471
- Scrum 627, 635, 683, 689, 692, 1398, 1406, 7401, 7550, 7573, 7578
- Scrum Master 635
- SDI 3465, 4012, 4014, 4017, 5256, 5265
- seamless redundancy 1231-1232, 1234, 1237
- Search Engine 101, 207, 1036, 1788, 2851, 3078, 3680, 4065, 4472, 4495, 4536, 4541, 5227-5228, 5230, 5262, 5264, 6072, 6651, 6771, 7018, 7024, 7117-7118, 7273, 7293, 7750, 7758, 8005, 8046-8049, 8051
- Search Engine Marketing 8048
- SEBI 70
- Second Heart Sound (S2) 1007-1008, 1010-1013, 1016
- Second Law 224, 228
- Second Life (SL) 745, 2525-2526, 4146, 4156, 4194, 4296, 4298-4299, 4301-4302, 4305, 7128, 7882-7884, 7887, 7917
- secondary animations 978-980, 982
- secondary task 5996, 5998-5999, 6005
- Secondary User (SU) 4105, 6543, 6545
- Secure Authentication 2657, 2660
- security attack 4854, 4861, 5310
- security economics 4856
- security policy 4852-4854, 4856-4857, 4861, 4912
- security project 4849-4853, 4858, 4861
- Security Risk 1156, 2693, 4854-4856, 4861, 4971, 5986
- Seer 1398, 1406, 1795, 1798-1801, 7785
- Segmentation 135, 350-351, 359, 364-366, 370-372, 945, 1007-1008, 1013, 1308, 1310-1314, 1316, 1318, 1354, 1585, 1594, 1617, 1783, 1786-1787, 1789, 1965, 2882-2883, 2912, 3413, 3474, 3519, 4215, 5542, 5750, 5759-5760, 5763, 5765, 5774, 6459, 6478, 6818, 7016, 7020, 7831, 7833-7834, 7838-7840, 8066
- Selective Laser Sintering 305-306, 313, 4383
- Self Evaluation 1568
- Self-Assembled Monolayers (SAM) 431, 439
- Self-Authorship 1434-1435, 1440
- Self-directed learning 5862
- Self-Efficacy 834, 1423-1424, 1426, 1452, 1474, 1478-1480, 1482, 1486, 1561, 1727, 2534, 3000, 3239, 3587, 3737, 3771, 3853, 3878, 4280, 5096, 5150-5151, 5153, 5691, 5727, 5860, 6022, 6403-6405, 6863-6864, 6944, 7163-7164, 7167, 7253
- Self-Help Mechanisms 3772, 3781
- self-leadership 5199
- Self-leveraging network 7112
- Self-Location 4235-4237, 4239, 4244
- Self-Organisation 1520, 6535
- Self-Organizing Map (SOM) 8, 122, 408, 442, 6581, 6644
- Self-Presence 4235-4236, 4244
- Self-Regulated Learning 1522, 2256, 2383, 3238, 3243, 3852, 3860, 3983, 7172, 7174
- Self-Regulation 1452, 1478, 1486, 1528, 1530, 2494, 2497, 3239, 3935-3936, 3944, 4957, 4963, 4971, 5997-5999, 6004-6005, 6400, 6405, 7089, 7516,

Index

- 7680, 7892, 7928, 7934
Semantic Analysis 235, 1783, 1786, 1793, 2070, 2817, 5172-5173, 5175-5177, 5179, 5182, 6478, 6481, 7156, 7416
Semantic Audio 6459, 6463
Semantic Cluster 7421
Semantic Information 456, 1209, 1764, 2755-2756, 4391, 4398, 4541, 6291
Semantic Intelligence 220-228
Semantic Interaction 6477-6480, 6482, 6487
Semantic Map 4571
Semantic Mismatch 2705
Semantic Network 4152, 5067, 5074, 7972
semantic systems 2064, 2067-2068, 2071
Semantic Web 235, 1036, 1039, 1200, 1441, 1443, 1450, 1517-1519, 1523, 1528, 2012, 2021, 2485, 2490, 2697-2698, 2773, 2777, 2780, 2811, 3038, 3462, 3475, 3477, 4540-4541, 4543, 4547, 4573, 4576, 4953, 5076, 5106-5107, 5111, 5215, 5225, 5239, 5251, 5613, 6406, 6430, 6458-6459, 6463, 6477, 6752, 6783, 7778, 8031
Semantically-Enhanced Media Authoring Model (SeMAM) 6487
Semantically-Enhanced Multimedia (SeMM) 6479-6480, 6483, 6487
Semantically-Enhanced Multimedia Assets (SeMMA) 6480, 6487
Semantically-Enhanced Multimedia Storytelling Services (SeMMSS) 6480, 6483, 6487
Semantics 181, 221, 224-225, 329, 1187, 1190-1192, 1194, 1208, 1344, 1352, 1443, 1517, 1523, 1788, 1840, 1980, 1995, 2022, 2028, 2043, 2046-2047, 2053-2054, 2058, 2063, 2267, 2755, 2775, 2777, 2779, 2914, 3519, 4103, 4423, 4430, 4543, 4573, 4576, 4736-4739, 4742-4743, 5068, 5268, 5410, 5613, 6291, 6448, 6459, 6477-6478, 6480, 6566-6567, 6571, 6848, 7450-7452, 7475, 7484, 7610-7611, 7614, 7616-7617, 7621, 7705-7706, 7709, 7711, 8035, 8056-8057, 8060-8062
Sematic Web 4540
Semiotic Analysis 4102-4103, 4107-4108, 4110, 7339
Semiotics 1511, 1513, 2237, 4103, 4112, 4404, 4426, 4430, 6567, 7313, 7339-7340
semi-regenerative process 1575
Semi-Structured 459, 692, 812, 1155, 1369, 1423, 1607, 2549, 3898, 4481-4486, 5638, 5640, 6189, 7636, 7674
Sense Making 722, 4485, 4487, 4517-4518, 7040, 7042
Sensitive Service 7754, 7761, 7764
Sensitivity Analysis 1572, 1578, 1633, 1796, 1801-1802, 1815, 2086, 2170, 2208, 2212, 4602, 4606, 5399
Sensor Networks 182, 409, 1065, 1227-1228, 1991, 3697, 4203, 4951, 4953, 6106, 6320, 6522, 6524, 6526, 6528, 6530, 6533, 6535, 6537, 7525-7526, 7776
Sensor Node 3697, 6530
Sensory Marketing 2584, 5767, 5770, 5772, 5774
sensory modalities 1548-1550, 1553, 1558
sensory preferences 1549-1551, 1555, 1558
Separate URL 8087, 8089-8092
Serial Cyberbullying 1697, 1703
serial FRM 6238-6239
Serious Game Design 3287-3288, 3290, 3295, 3340
Serious Game Design Models 3287-3288, 3295
Serious/Applied Game 801-803, 805, 2466-2467, 2471-2472, 2476, 3272, 3287-3288, 3290, 3295, 3329, 3333-3334, 3336, 3340, 7941
Server Adaptive 8092
Server Vacation 1583
Server-Side Handheld Programming 6069
Service Gateway 2722
Service Improvement 3529, 5562, 5569
Service Learning 3829, 6689-6690, 6694
Service Level Agreement (SLA) 630, 644, 887, 1029, 1032, 1359, 1362, 1365, 2690, 2920-2921, 4013, 5320, 6546, 6555, 6613, 6617, 6735
Service Oriented Architecture (SOA) 586, 637, 639, 653, 879, 887, 1061, 1077, 2774, 2922, 3604, 5544, 5546, 5612
Service Provider (SP) 459, 531, 586, 604, 887, 933, 1026, 1028, 1067, 1101-1102, 1106, 1114, 1144, 1147, 1153, 1164-1165, 1167-1170, 1181, 1357-1359, 1365, 1407, 1627, 1629-1630, 1633-1634, 1652, 1687, 1990, 2286, 2347, 2645, 2690, 2694, 2708-2709, 2712, 2731, 2738, 2803-2804, 2806, 2809, 2829, 2922-2923, 2925, 2928, 2930, 2969, 2977, 2993, 3033, 3150, 3186, 3581, 3586, 3768, 4056-4057, 4063, 4217, 4628, 4822-4824, 4829, 4839, 5305, 5319, 5368, 5427, 5477-5478, 5489, 5492, 5570, 5583, 5598, 5909, 5911, 6087, 6098, 6101-6102, 6154, 6184, 6189, 6280, 6297, 6301, 6546, 6548, 6552, 6555, 6691, 6700-6701, 6704, 6709, 6716, 6727, 6951, 7228, 7293, 7525, 7528, 7532, 7593, 7747, 7749, 7756, 7871, 7873-7874, 8056
Service Providing Agent (SA) 2699, 2705
Service Quality 51, 816, 1129-1132, 1134-1135, 1137, 1140, 1175, 1593, 1604, 1622-1623, 2735, 2761-2763, 2767, 2769-2770, 2806, 2857, 2882-2888, 2890-2891, 3024, 3186, 3594, 4091, 4212, 5293, 5360-5361, 5489, 5493, 5518, 6414, 6546, 7872, 7874
Service Reliability 2735, 7869-7870, 7872-7874, 7876

- Service Reuse 7538
 Service Science 7203, 7205, 7211
 Service Sector 2806, 4717, 5477, 7202, 7239, 7245, 7247
 Service Virtualization Testing 7513
 Service-Oriented Computing 7588, 7593, 7598, 8056
 Service-Oriented System 7588-7592, 7598
 SERVQUAL 1129-1133, 1137, 2761, 2763-2764, 2883
 Set Disjunction 8005
 Set Intersection 7996, 8005
 Set Union 348, 7996, 8005
 sex trafficking 1366
 Sexting 1697, 1710, 4247, 4318, 6391
 Shallow Reading 1508, 1512, 1516
 Shannon Entropy 157, 1008, 1187, 4398
 Shapley Value 1165, 2088, 2095-2098, 2101-2107
 Shared Identity 720, 723-724, 728, 6706
 Shared Leadership 5198, 5203, 7917
 Shared Service Centers 887, 5559, 5562-5563, 5565, 5569
 Shareholder 61, 65, 672, 675, 923, 4858, 5379, 5428, 7168
 Sharing Economy 2347, 2351, 2355, 3570
 Sharpening 663, 669, 4755, 6012-6014, 6052, 6174-6178, 6180-6181, 6183
 Sharples 2466, 5117, 6185, 6347, 6349-6352, 6354, 6363-6364, 6398-6400, 6405, 7003
 Shopbot 2849
 Short Cycle Time and Low Inventory (SLIM) Methodology 5574, 5582
 Short Message Service (SMS) 6150, 6210, 6222, 6275, 6285, 6985
 Shortest Path Trees 6539-6542
 SIAM 2920, 2927, 2930
 Sidelobe Level (SLL) 5970, 5980, 6598-6599, 6608
 SIEM 376, 380
 SIFT 256, 2237, 2912, 4535, 6198-6200, 6783
 Signal Processing 143, 184, 493, 498, 1007, 1011-1012, 1249, 1816, 4017, 4022, 4129-4130, 4872, 4875-4876, 4937-4939, 4941-4942, 4947, 6007, 6042-6043, 6094, 6118, 6137, 6144, 6235, 6256-6258, 6447, 6458, 6581, 6659, 6677
 Signal To Noise Ratio (SNR) 1190, 4875, 4880, 6141, 6143, 6621, 6629, 6683
 Signature File 7995-7996, 8005
 significant learning experiences 1435-1436, 1439-1440
 significant learning taxonomy 1432
 Silk Road 2861-2863, 2868
 Siloed 1467, 1473
 Silver Tsunami 7028, 7035
 SIMD 1955-1956, 4008, 4014, 4017, 8003
 Simulated Annealing 114, 183, 186, 2178, 3108, 4371, 4377, 4600
 Simultaneous Comparison Turing Test 263
 Simultaneous Localization And Mapping (SLAM) 6290, 6295, 6894
 SimVenture 801-803
 single-asset Black-Scholes PDE 16-18, 20, 22, 28-30
 Single-Event Expert Query 2145, 2153
 Singular Configurations 6893
 Singular Value Decomposition (SVD) 1740, 1762, 1771, 4875-4876
 Singularity 211, 1043, 1850, 4014, 5751
 SIPOC 730, 732
 Siri 101, 4542, 5750-5751, 5753, 5756, 6185, 6392, 6413, 6658, 7801
 siRNA 478, 480, 482, 484, 489-490
 Situational Awareness 983, 989, 1411-1412, 1691, 2064-2065, 2067-2069, 2071, 2074, 7304
 Situational Factor 7434
 Situational leadership 5195, 5203
 Situational Method Engineering 7422, 7434
 Situational Trust 4797, 4799-4801, 4803-4805, 4807, 4809
 SJR 4407, 4409, 4414, 4420, 5246
 Skill Acquisition 2226, 2228, 7358-7360, 7367
 Skill Shortage 4785, 4796
 Skimming 78, 1510, 1512, 2715, 2722
 skin tone 4986-4987
 Skunkworks 551, 555, 562
 Sleptsov Net 7731, 7733-7735, 7737-7738, 7741-7743
 Sliding block method 6315-6316
 Sliding Window 364-366, 368, 370, 372, 410
 Slope Analysis 3436, 3447
 Small And Medium Enterprises (SME) 893, 1673, 2220, 5308, 5529, 6553, 6737, 6916, 6935
 Small Group Delphi Paradigm (SGDP) 663, 669
 Small Worlds 7153
 Small-World Property 6516, 6521
 Smart Environment 7122, 7124
 Smart Grids 1227, 3103-3104, 3107-3109, 3113, 3147
 Smart Meters 3104, 3146, 3149-3150, 3155, 6683, 7227
 smart self-learning systems 2064, 2069, 2071
 Smart Supply Chain 5305
 Smart Systems 4736, 4953, 6483, 7784
 Smart Village 7261, 7269
 smart watch 6116, 6123, 6154
 Smart Weapon 7303, 7311

Index

- smartphone application 6085, 6152, 6160
smartphone built-in sensors 6035, 6037, 6042
SMOG Grade 1507
Snapchat 1143, 2295, 2641, 2743, 3612, 3620, 4310, 4313, 5226, 5909, 6185, 6249, 6302, 7120, 7798
Snow World 5965
SNPs 467, 470-474, 476
So.Re.Com. "A.S. De Rosa" @-Library 4406, 4420, 5237-5239, 5245-5246, 5251, 7051, 7056
Soap Operas 7218-7219, 7222
Social Business 765-769, 773-774, 776, 2558, 2570
Social Business Gaming 2558, 2570
Social Business Process 765-766, 768-769, 773, 776
Social Capital 778, 944, 947, 2281, 2288, 2325, 2592, 3027, 3540, 3925, 4280, 4283, 4969, 5095, 6698, 6703, 6943, 7059-7061, 7080, 7106, 7153, 7800
Social Change 2319, 2328, 3020, 3336, 3539, 3569-3571, 3573-3574, 3577-3578, 4715-4716, 4722, 5739, 6281-6282, 7106, 7109, 7113, 7303
Social Cognitive Theory 1478, 3290, 3655, 4280, 7081, 7087
Social Commerce 2851-2858, 2860, 7128
Social Connectedness 2318, 4245, 6184-6185, 6188-6190, 6194, 6703
Social Construction of Technology 7312, 7325
Social Constructionist Perspective 721, 728
Social Context 384, 726, 1186-1187, 1192, 1194-1195, 2123, 2280, 2361, 2643, 2982, 3871, 3880, 4162, 4358-4359, 4486, 5149, 5156-5157, 6038, 6403, 6698, 7059-7060, 7078, 7314, 7934
Social CRM 2779, 2854, 2860, 5775, 5782-5783
social cue 7973-7974, 7979
Social Determinism 7325
Social Dimension 611, 768, 778, 2226, 2966-2967, 2970-2971, 2974, 2978, 3035, 3184, 3189, 3747-3748, 4300, 5093, 5466, 7121
Social Discourse 1474-1476, 3250, 4047-4048, 4189, 5117, 5152, 7015, 7022, 7922-7923
Social Disinhibition 4177
Social E-Commerce 2745, 2748
Social Educator 5940
Social Engagement 1475, 1477, 5153, 6695, 6705-6706, 7922
Social Exchange 713, 6714, 6719, 6726
Social Facilitation 6943, 6949, 7100, 7791-7792
Social Identity 723, 3320, 3326, 4194, 4871, 7799, 7804, 7973
Social Inclusion 279, 1559, 1566, 5017, 5019, 7031, 7223-7224, 7233, 7251
Social Influence (SI) 778, 2253, 2733, 2762-2763, 2765, 2768, 2770, 3539, 3541, 3581-3582, 3584, 3586, 3590, 3655-3657, 3659-3662, 3666, 3787, 6038, 6945, 7063, 7194, 7283
Social Information 778, 3370, 4462, 4469, 7974
Social Integration 2417, 2419, 2616, 2625, 7274
Social Intelligence 286
Social Interaction 720-721, 723, 728, 768, 1187, 1191-1192, 1195, 1491, 1715, 2316, 2436, 2444, 2528, 2546, 2617, 2643, 2650, 2851, 2860, 3239, 3249-3250, 3265, 3271, 3292, 3649, 3747-3748, 3834, 3905, 4179, 4190, 4194, 4262, 4281, 4284, 4761, 5004, 5093-5094, 5096-5097, 5099, 5102, 5253, 5268, 5776, 6354, 6393, 6442, 6705, 6837, 6861, 6949, 7015, 7037, 7088, 7096, 7132, 7162, 7169, 7360, 7799, 7890, 7984
Social Learning 1562, 2230, 2359, 2562, 4180-4181, 4188, 4359, 4517, 5158-5159, 5161, 5164, 5168, 5199, 6342
Social Learning Analytics 5161, 5164, 5168, 6342
social machine 2779
Social Marketing 3360, 5737-5741, 5743-5744, 5747, 5757, 6635
Social Media Credit Score 7140, 7143, 7149
Social Media Marketing 2852, 2860, 3360, 3369, 4053, 4098, 5763, 7139, 7143
Social Media Network Services 6184
Social Media Platform 2064-2065, 4053, 6153, 7063, 7119, 7169
Social Media Sites 357, 2341, 2423, 2860, 4830, 5710-5711, 5750, 6085, 6278, 6940, 6993, 7001, 7018, 7099, 7143, 7145, 7182-7183, 7749
Social Media Technologies 702, 1594, 1596, 2558-2562, 2567-2568, 2570, 3572, 5782-5783, 6994, 7090, 7164, 7168, 7170, 7174-7175, 8043
Social Modeling 6943-6945, 6949
Social Movement 1188, 1642, 3538-3539, 3542, 3570, 3573, 5263, 6276
Social Network Analysis (SNA) 345, 937, 2034, 2860, 3923, 3976-3977, 5098, 5160, 5162, 6961-6962, 6970-6971, 7015, 7058, 7068, 7150-7151, 7156-7157, 7160, 8036
Social Network Theory (SNT) 6999, 7001
Social Network(Ing) Sites (SNS) 1715, 2486, 2559, 3359-3360, 3612-3614, 3667, 4138, 4145, 4192, 5102, 5256, 5713, 6941, 6944, 6995, 7006, 7014, 7036, 7088, 7092, 7096, 7127, 7129, 7185, 7190, 7199, 8039-8040
Social Networking 266, 326, 333, 344, 387, 459, 783, 1103, 1109, 1143, 1367, 1370, 1533, 1642, 1705-1706, 1710, 1724, 2065, 2237, 2247, 2316-2320, 2345, 2417, 2419, 2479, 2481, 2486-2487, 2558, 2745, 2748, 2769, 2772-2773, 2851-2852, 2854-

- 2858, 2860, 2908, 3572, 3671, 3852, 3854, 3922, 4053-4054, 4122, 4145, 4190-4193, 4198, 4223, 4225, 4247, 4279, 4320, 4361, 4460, 4830, 4872, 5098-5099, 5102, 5105, 5110, 5126, 5206, 5252-5258, 5261-5262, 5264, 5266-5267, 5546, 5713, 5776, 5778-5779, 5782-5783, 5785, 5803, 5840, 5908, 5910, 6019, 6030, 6095, 6249, 6278, 6393, 6404, 6479, 6781, 6784-6785, 6909, 6942, 6944, 6950-6953, 6955-6956, 6960-6961, 6985, 6987-6989, 6994-6995, 7006, 7016, 7024, 7027-7028, 7036, 7049, 7058, 7078, 7088-7092, 7096, 7101, 7116-7122, 7124-7125, 7128-7129, 7161-7165, 7168-7170, 7175, 7180, 7182-7183, 7185, 7189, 7273, 7369, 7671, 7800, 7951-7952, 8037-8038, 8040, 8045, 8066-8067, 8071, 8078, 8097
- Social Networking Applications 2481, 2857, 5253
- Social Networking Elements 8037-8038, 8045
- Social Networking System 3854, 6950-6951, 6955, 6960
- Social Networks 333, 344, 388, 461, 464, 653, 657, 707, 719, 767, 778, 782-783, 787, 947, 952, 1512, 1517, 1522, 1639, 1642, 1644, 1708, 1746, 1775, 1806, 2244, 2292, 2319, 2325, 2516, 2524, 2579-2580, 2773, 2779, 2851, 2854, 2856-2857, 2860, 3353, 3359-3360, 3366, 3369, 3645, 3670, 3691, 3750-3751, 3768, 3922-3924, 3926, 3929, 4047, 4051, 4053, 4071, 4073, 4088, 4147, 4151, 4161, 4164, 4168, 4266, 4278, 4295, 4329, 4349, 4359, 4361, 4416, 4716, 4718-4720, 4829, 5084, 5098, 5102, 5106, 5159, 5161, 5168, 5253-5255, 5258, 5262-5263, 5267, 5376, 5749, 5779, 5783, 5865, 5891, 5909-5914, 6038, 6186, 6189, 6248, 6252, 6339, 6478, 6489, 6508, 6635, 6783, 6909, 6941-6943, 6945, 6949-6951, 6953-6954, 6956, 6960-6962, 6965, 6967-6968, 6971, 6973-6974, 6989, 6996, 7006, 7008, 7014, 7024, 7028, 7044-7049, 7051-7052, 7056-7061, 7063-7064, 7068, 7088-7089, 7091, 7110, 7128, 7143-7144, 7149, 7165, 7252, 7317, 7350, 7411, 7628, 7797-7798, 7893, 8039-8040, 8051
- SOCIAL ONLINE PRACTICES 2315
- Social Platforms 1167, 2065, 4090, 5777, 6489, 7128, 7162-7163, 7798
- Social Presence 1454, 2507, 2665, 3239, 3861, 3866, 3976-3977, 3980, 3982, 3985, 5833, 7974
- Social Promotion 7133, 7139
- Social Reach 3370
- Social Relationships 777-778, 780-784, 787, 1704, 3420, 3540, 3772, 3905, 3922-3923, 3926, 4193, 4266, 5095, 5334, 5457, 5936, 6185, 6557, 6783-6784, 6943, 6953, 6988, 7015, 7031-7032, 7064, 7072, 7097, 7139, 7319, 7969-7971, 7973-7976
- Social Representations 3320, 4038-4039, 4043-4045, 4048, 4050-4051, 4064-4068, 4071-4073, 4076, 4404-4409, 4413-4416, 4420, 5237-5239, 5244-5246, 5250-5251, 7014, 7025, 7044-7049, 7051-7052, 7056, 7321
- Social Responsibilities 675, 2227, 2505, 5382, 5552
- Social Scholarship 6780, 6785, 6787
- Social Semantic Web 1517-1519, 1523, 1528
- Social Shaping Of Technology 6163, 7312, 7325
- social sharing 7946, 7948-7952, 7955
- Social Shopping 2852, 2860, 7798
- Social Signals 8050-8052, 8055
- Social Skills 1518, 1520, 2297, 2317, 3262, 3838, 3871, 6363, 6393, 7173, 7881
- Social Structure 2791, 2851, 3430, 3639, 3648, 3923, 3976, 3979-3980, 3983, 4280, 4424, 4715-4716, 4722, 7014, 7068, 7150, 7190, 7899, 7956
- Social Support 2575-2577, 3770-3772, 3781, 7028, 7059, 7079, 7081, 7249, 7253, 7949, 8067
- social telerehabilitation 5930, 5932-5937
- Social Ties 724, 2318, 4763, 4922, 6184, 6941, 7038-7039, 7042, 7059, 7162, 7167
- Social Transparency 755
- Social Validation 288-289, 297
- Social Web 709, 1528, 2779, 5264, 5271, 6430-6431, 6783, 6950-6951, 7407, 7410, 8050
- Social World 2773-2774, 3833, 3932, 4196, 4254, 4279, 4481, 4717, 6185, 6567
- societal vulnerability 4267
- Socio Economic Processes 7270-7271, 7280
- socio technical change 2976, 2980-2983, 2986
- Socio Technical Change Perspective 2986
- Sociocultural 2298, 2328, 2444, 4180, 4182, 6341, 6399-6400
- Socioeconomic 278, 800, 907, 1661, 2238, 3021-3022, 3086, 3454, 3538, 3770, 4110, 4301, 4704, 5022, 7092, 7223-7224, 7226, 7228, 7233, 7260
- Socio-Economic Digital Literacy 2287, 2291
- Socio-Emotional 726, 2227, 2255, 2281, 2327, 6184, 6188, 6190, 6194
- Sociolinguistic Area 1206, 1215
- Sociomateriality 699, 706
- Socioscientific Issues 1423, 1431
- Socio-Technical Perspective 4279-4280, 4287, 4295
- Socio-Technical Research 4280
- Socratic Method 1439, 2571-2572, 2574, 2576-2579
- Soft Computing 181, 184, 186, 1837, 6971, 7288, 7693, 7700, 7741
- Soft Function (S-Function) 1250, 3226, 3230-3231, 3236

Index

- soft game 3226, 3231, 3236
- soft set 3226-3227, 3230-3232, 3235-3236
- Soft Skills 1476, 1561-1562, 1564, 1566, 2422, 4786-4788, 4790, 4793-4794, 4796
- Software Agent 107, 6821, 6839, 6960
- Software Application 886, 1385, 1388, 1406, 1629, 1869, 3620, 4156, 7503, 7538, 7542-7543, 7548, 7598
- Software Architecture 91, 1365, 1398, 5095, 5106, 6838, 7527
- Software As A Service (Saas) 887, 1027, 1033, 1102, 1104, 1129, 1144, 1148, 1356, 1365, 1627, 1637, 2968, 2983, 3464, 5263, 6730, 7525, 7532, 7598, 7767, 8056
- Software Control 6838, 6843, 6874, 7435-7436, 7443, 7446
- software crisis 1396-1397
- Software Defined Networks (SDN) 6617, 7392, 7397
- software defined radio 6054, 6138, 6141, 6146
- Software Design 1398, 1406, 3035, 7394, 7481, 7532, 7541, 7579, 8027
- Software Design Pattern 1406
- Software Development 133, 460, 574, 622, 669, 689, 724, 777, 1332, 1335, 1396-1402, 1405-1406, 2020-2021, 2481, 2491, 2923, 3339, 3347, 3463, 3526, 3529, 4124, 4852, 4931, 5127, 5478, 5518, 5660, 5666, 5854, 5876, 5942, 5954, 6061, 6248, 6913, 6927-6930, 6935, 7259-7266, 7269, 7370-7372, 7377, 7392-7395, 7398, 7400-7401, 7405, 7407, 7410-7411, 7422-7425, 7431, 7442, 7455, 7482, 7504, 7508, 7512-7513, 7525, 7528, 7531, 7549-7550, 7553-7554, 7557, 7562-7563, 7566, 7577-7578, 7584, 7587, 7609, 7616, 7758, 7799
- Software Development Kit (SDK) 133, 5876, 5942, 5954, 6061
- Software Development Life Cycle 777, 7398, 7423, 7425, 7513, 7528, 7563, 7566
- Software Engineering 317-318, 894, 1396-1398, 1401-1402, 1405-1406, 1444, 1450, 1772, 1874-1875, 2020-2021, 2313, 2380, 3309, 3316, 3338, 3526, 3528, 3784, 3793, 4200, 4209, 4725, 4730, 4810, 5106, 5436, 6929, 6933, 7392, 7398-7400, 7402-7403, 7407-7408, 7410, 7421-7422, 7481-7482, 7484-7485, 7487, 7512, 7525, 7529, 7534, 7553, 7555, 7577-7579, 7584, 7587, 7609, 7617, 7706, 7774, 7935, 8006, 8023
- Software Engineers 377, 1397, 3337-3338, 4138, 5480, 6067, 6188, 7394, 7401-7402, 7408, 7528, 7551, 7603
- Software Evaluation 3782, 3784, 6910
- Software Industry 614, 5388, 5436, 5546, 6927-6928, 7118, 7238, 7259-7260, 7262-7266, 7269, 7455, 7556, 7985
- Software License 7382, 7911
- Software Literacy 7539, 7541-7545, 7548
- Software Piracy 1658, 1665, 7265
- Software Platform 583, 4742, 5853, 7548, 8037
- Software Process 3529, 6927-6929, 6938, 7422, 7425, 7514-7515, 7517-7519, 7524, 7549-7555, 7557-7558, 7562
- Software Process Assessment 7514-7515, 7517-7519, 7524
- Software Process Improvement (SPI) 3529, 6938, 7514, 7524, 7549-7552, 7555, 7557-7558
- Software Process Model 7562
- Software Product 1396-1397, 1402, 1405, 2975, 3793, 5011, 6913, 6927-6928, 6934, 7399, 7402-7403, 7431, 7565
- Software Programming 6479, 6481, 7392-7394, 7396-7398
- Software Quality 1333, 1406, 2935, 3793, 6935, 7555, 7566, 7603
- Software Security Engineering (SSE) 7528, 7531, 7534, 7538
- Software System 578, 654-655, 808, 1064, 1398-1399, 1401, 1745, 1995, 2057, 2157, 2978, 3405, 3960, 4584, 5127, 5135, 5393, 5545-5546, 7402, 7421, 7423, 7435, 7447, 7453-7455, 7505, 7578-7579, 7587, 7598, 7609, 7611, 7616, 7708, 7912
- solar cells 2897-2898, 2900-2901
- Solar Energy 1296, 3491
- solid state disks 1807, 1810, 1948, 1956
- Solid Works 2637
- Solidarity 1640-1641, 1644, 3027, 4767, 5018, 5039
- Soloway 2546, 2572, 6347, 6349, 6353, 6355, 7672
- Soma 4805, 6582, 6594
- sound pen 6464-6466, 6468-6469, 6471, 6475
- Sound Relationships 778-779
- Source Language 4493-4498, 4501, 5136, 5139, 7329
- Southern Association of Colleges and Schools (SACS) 3182
- Sovereign risk 5556, 5558
- Space-Time 1851, 2065, 3463, 3465, 3467, 3472, 3640, 4191, 4466, 5218, 6501, 6506, 7707
- Spatial Analysis 3420, 3461, 3463, 3467, 3472-3474, 3479, 3483, 3501, 7846
- spatial auto correlation 1835-1836, 1838, 1840-1841, 1844
- spatial data mining 1835, 1837-1838, 1842, 1844, 1909
- Spatial Mechanism 6893

- Spatial Resolution 206, 3485, 3487, 3492, 3494-3498, 3502, 3507
- Spatio-Temporal Processing 6506
- Spearman's Rank 441, 3197, 5586
- Special Education 2431, 2434, 2439, 6341, 6397, 6863, 7647
- specialized logical class 2026, 2030
- Spectral Radius 1748, 1755, 6513-6514, 6521
- Spectral Resolution 3485, 3498, 3502
- Spectroscopy 2901, 3488, 5111, 6123
- Speedup 1809, 1956, 7576
- SPI Models and Standards 7558
- Spin Glass 114, 119
- Spinning 3994-3996, 4000, 4003
- Splintering Effect 7235, 7247
- Sport activity 7357
- Sport Tracker 7357
- Sport Training 7357
- SPOT 29, 35, 546-547, 901-903, 905, 908, 1164, 1166, 1169, 1173, 1319, 1686-1687, 1691, 1837, 2059, 2548, 2622, 3413-3414, 3435, 3440, 3485, 3491-3494, 3496, 3498, 3799, 4009, 4077, 4337, 5373, 5546, 5652, 5656, 5983, 6523, 6527, 6529, 6632, 7673, 8048
- Spot Pricing 35, 1164, 1166, 1169, 1173
- Spring embedders 7414
- SSD 1948-1949, 4011, 4017
- SSL 2711, 2722, 3492, 3494, 7721, 7724
- Stackelberg Game 1165, 6552, 6555
- Staffan Björk 6377
- Stakeholder Relations 671, 681
- Standards Setting Organization 4691
- Stand-Off Weapon 7311
- Stanford Yeast Database 440
- Startups 230, 1955, 2347, 2863, 3032, 3034, 3038-3039, 7259, 7261, 7263, 7266, 7268
- Stasis 6715-6716, 6719, 6726
- State Automata 6858
- State Transformation Matrix 219
- State Vector 17-19, 24-25, 30, 35, 214, 219, 1048, 4391-4392
- Static Graph 6538, 6545
- Stationarity 372, 5639-5640, 6641
- Statistical Features 4950, 6137-6138, 6141, 6143-6144, 6146
- Status Quo Bias Theory 3820, 3825
- steady-state probabilities 1571, 1575
- steganography 4985-4989, 4991, 4994, 4996-4998, 5003
- stego image 4986, 4991, 4997, 5003
- STEM Education 2626-2627, 3248, 3316, 6376-6377, 6381, 6387, 7908
- Stemming 459, 997, 1906, 2349, 2505, 3382, 3541, 3933, 4529, 4542, 5054-5055, 5680, 5684, 6499, 6567, 6570, 7799, 7935
- Stent 3797, 3799, 3804
- step detection 6289-6290, 6295
- stereo visualization 2411, 2413-2414, 2418, 2420
- Stereolithography (STL) 305-306, 313, 2640, 4382-4383
- Stereo-Photogrammetry 168, 179
- Stereotype Threat 3382-3384, 3392
- Stereotypes 3350, 3354, 3360, 3382-3384, 3386, 3391, 3396-3397, 3638, 4202, 4239, 4358-4359, 6706, 7078, 7540
- Stochastic MGA 2183, 2187
- Stochastic Modelling 4642
- Stochastic Programming 2188-2189, 2199, 5412, 5418
- Stock Market 6-8, 17, 73, 681, 2111, 2142, 2145-2147, 2151, 7027
- store of value 2784, 2786-2787, 2861, 2866
- Storming 4520, 5656, 5659, 5707
- Storytelling 944, 997, 2442-2444, 2446-2448, 2450, 2454, 2616-2618, 2620-2622, 2624, 2641-2647, 2650, 3309-3311, 3316, 3502, 4257, 4696, 4698, 4700, 5077, 5086, 5104, 5107-5108, 5110, 5114, 5664-5665, 5668, 6393, 6477, 6480-6481, 6483, 6487, 6489, 7338, 7341-7342, 7345, 7359
- Storytelling Services 6480, 6483, 6487
- STR 1668-1673
- Strategic Analysis 922, 1400, 4624, 4777
- Strategic Awareness 912-913, 917, 922
- Strategic Human Resource Management 4347-4349, 4357
- Strategic Information Systems Planning 912, 922
- Strategic Management Accounting (SMA) 2177
- strategic map 5640-5641
- Strategic Marketing 961-964, 967-968, 970-971, 5767, 5770, 5774
- strategic marketing planning 962, 964, 967, 970-971
- Strategic Planning 575, 580-581, 584, 639, 900, 909, 912, 914, 917-918, 961-962, 966, 971, 1408, 2085, 2170, 2246, 3380, 3414, 3451, 4058, 5107, 5424, 5505, 5507-5508, 5511, 5515, 7652, 7849
- Strategic Planning Horizon 5511, 5515
- Strategic Scanning 907, 911
- strategy conception 912-913, 917-918, 922
- Strategy Formulation 577, 902, 912-913, 922, 961, 963, 3025, 4574, 4780
- strategy implementation planning 912-913, 918, 922

Index

- Strength of the relationship 707, 713
Stress Reduction 4325
Stress Testing 6813-6816, 6818-6821, 6823, 7506
Stringing 4383, 4389
Stripping 3044, 3046-3048, 3050-3053, 3055, 4533
Strong Negation 4738, 4745
strong replica consistency 1093, 1238-1239, 1244, 1247
Strong/Weak Ties 766-767, 2999, 3768, 3925, 7039, 7059-7061, 7152-7153, 7160-7161, 7194
Structural Change 7234-7236, 7981, 7983
Structural Equation Modeling (SEM) 54, 837, 1133-1134, 2954, 3584, 3586, 3660, 3666, 6020-6022, 6025-6026, 6403, 8070, 8072, 8078
Structural Holes 3924-3925, 7062, 7152-7153, 7160
Structural Information 4459-4462, 4469
structural mismatch 2699, 2702, 2705
Structural Nativization 1208, 1215
Structurality 4465, 4469
Structured Analysis and Design Technique (SADT) 1397, 1405
Structured Data 335, 354, 407, 851, 864, 872, 1620, 1773, 2031-2032, 2059, 4953, 5076, 5105, 7149
Structured Data Source 7149
Structured Ontological Term 2705
Structured Query Language (SQL) 2031, 2042
Student Assessment and Learning 2570
student autonomy 2535
Student Community 2412, 3834, 3976
Students With Disabilities 266-271, 314, 319, 321, 2441, 6363, 6389, 6392-6394, 6397, 6471
Students With Special Needs (SEN) 2521, 2531, 2547
Subject Matter 233, 663, 669, 821, 956, 1435, 1445, 1454-1455, 1463-1464, 1468, 1474-1482, 1486, 2493, 2501, 2532-2533, 2541, 3173, 3310, 3624, 3627, 3629, 3840, 3851, 4491, 5116, 5149-5154, 5157, 5232, 5234, 5493, 6363, 6771, 6803, 6860, 6870, 7057, 7403, 7481, 7660-7662, 7797-7798, 7800, 7922-7930, 7934, 7973
Subject Matter Expert (SME) 669, 1454
Subjective Norms 532-533, 5096, 6403, 7080, 7194
Subjective Probability 51, 2197, 4637, 5684, 5689
Subjective Relevance 4496-4497, 4501
Subjective Well-Being 6223-6224, 6228-6230, 7253
Subject-Matter 3341, 3344-3345, 3347, 6862
Subscription Model 7129, 7139
substitutable 5336-5337
substructure 1976, 1978-1979
Subsumption Architecture 6846
Subtask of the First Kind 201
Subtask of the Second Kind 201
Success Factors 579, 620-621, 627-628, 631, 635-636, 643-645, 777, 1108, 1935, 2994, 2996, 3435, 3561, 3564, 3765, 4283-4284, 4287, 5050, 5309-5310, 5312-5313, 5460, 5619, 5691, 5695, 5720, 5902-5904, 7268, 7377
Succinct Constraint 348
Succinctness 344
Summative Assessment 1522, 3933, 5120, 5122, 5126, 5183, 5192
Summative Evaluation 1880, 1894
Sun-Synchronous 3494, 3501
supercapacitors 2673, 2675, 2680, 2897, 2900-2902
Super-Efficiency 588, 592-594, 596-597, 600
Supervised Learning 4, 14, 120-123, 131, 148, 183, 350, 406, 408, 498, 1831, 1933, 1966-1967, 1973, 4940-4941, 5173, 5182, 6141, 6146, 6633, 6641, 6657
Supervisory Control And Data Acquisition (SCADA) 1406, 2059, 3065, 3113
Supplier Network 5389, 5394, 5397
Supply Chain 582, 584, 588, 611, 684, 863, 873, 923-924, 929, 945, 1587, 1686, 2110, 2692, 2694, 2773, 2791, 2793, 2802, 2805-2806, 2810, 2813, 2933, 2953, 2955, 2977, 3086-3088, 3133, 3144, 3150, 3187, 3711, 4600, 4826, 4951-4952, 5050, 5273, 5275, 5280, 5284-5285, 5287-5288, 5295, 5300, 5305-5313, 5316, 5327-5332, 5334-5337, 5339-5340, 5344-5354, 5356-5361, 5364, 5366-5369, 5371, 5373, 5375, 5378-5384, 5386-5387, 5390, 5392, 5394, 5397-5404, 5406-5408, 5411, 5427-5428, 5433-5441, 5446-5447, 5450-5451, 5455-5462, 5464-5468, 5470, 5475, 5477, 5490-5493, 5503-5507, 5510-5511, 5515-5517, 5519-5522, 5526-5530, 5532-5533, 5536-5544, 5546, 5549-5553, 5556, 5558-5565, 5569-5571, 5574-5578, 5581-5584, 5590, 5592, 5595-5601, 5605, 5699, 5918-5924, 5929, 5981-5982, 5987, 6094, 6821, 7870-7875, 8057
supply chain activities 5312, 5357, 5366-5367, 5381, 5383, 5387, 5402, 5455, 5460, 5464, 5515, 5519, 5537, 5551, 5570, 5581, 5605
Supply Chain Efficiency 2806, 5306, 5310, 5576, 5597
Supply Chain Facilities 5356
supply chain flow 5464, 5521
Supply Chain Integration 5288, 5307-5309, 5313, 5316, 5468, 5477, 5527-5528, 5537, 5919-5920, 5981
supply chain intelligence 5538-5539, 5541, 5543, 5549
Supply Chain Management (SCM) 923, 929, 945, 2110, 2810, 2813, 2933, 2977, 3086-3088, 3711, 4951-4952, 5273, 5280, 5306-5308, 5328-5332, 5334-5335, 5337, 5345, 5357, 5364, 5366-5367, 5378-5380, 5382, 5384, 5386-5387, 5392, 5397-

- 5398, 5401-5402, 5407, 5433, 5446, 5450, 5455-5462, 5464-5467, 5475, 5490, 5492, 5503-5504, 5516, 5519, 5528, 5538-5541, 5551, 5559-5563, 5565, 5569-5570, 5582-5583, 5590, 5595-5597, 5600, 5918-5919, 5924, 5929, 5987, 6094, 7872, 7875, 8057
- Supply Chain Management Solution 5569
- Supply Chain Optimization Models 5515
- Supply Chain Performance 5284, 5351, 5382, 5400, 5428, 5434, 5436, 5526, 5528-5529, 5533, 5537-5541, 5551, 5581-5582, 5601, 5605, 5919-5920, 5922, 5924, 5982
- supply chain risk 5308, 5310, 5511, 5550-5551, 5553, 5556
- supply network 5291, 5399-5401, 5403-5405, 5407, 5410, 5433, 5539, 5541, 5543, 5545, 5547, 5549
- Supply-Driven View of Knowledge 5055
- Support Group 3769-3770, 3772, 3775, 3781, 6439, 7079
- Support Vector Machines (SVM) 4, 8, 14, 184, 406, 423, 496, 499, 854, 1012, 1808, 1829, 1909, 1925, 2085-2086, 4942, 5173, 5177, 5179, 5182, 7352
- Supporting Facility 5356
- SURF 2247, 2810, 6198-6200, 6895, 7098, 7381
- Surface Electromyography 492, 504, 7362
- Surface Learning 1538, 2584
- Surgical Manipulators 6877, 6890
- Surgical Robots 6878-6879, 6882-6883, 6885-6886, 6889, 6891
- Surprising Questions 2705
- Surrogate Body 4234, 4239-4240
- surrogate data 4937-4938, 4942-4943, 4945-4946, 4948, 4950
- survivability rate 1795, 1801-1802, 1805
- Survival Analysis 1794-1795, 1801, 1805, 5192
- Survival of the Fittest 159, 165, 226-228
- Sustainability Balanced Scorecard 5718
- sustainability development 5533
- Sustainability Reporting 3131, 5737-5740, 5742, 5744, 5747
- Sustainable Competitive Advantage 730, 739, 926, 943, 1159, 1161, 1598, 2347, 2794, 4776, 5061, 5425, 5570, 5675, 5714, 5725, 6416
- Sustainable Development (SD) 227-228, 910, 2280, 2284-2286, 2291, 2629, 3014, 3177, 3183, 3190, 3248, 3372, 3377, 3381, 3435, 3438, 3580, 4034, 4267, 4277, 4302, 4621, 4710, 5015, 5020, 5026, 5079, 5081, 5293, 5357, 5362, 5366, 5387, 5529, 5532, 5537, 5718-5719, 5742-5743, 6718, 6907, 6913-6915, 7204, 7821, 7843
- Sustainable Development Goals (SDG) 3580, 5015, 5026, 5079, 5081
- sustainable knowledge societies 2286, 2288-2289
- Swarm Intelligence 1, 7, 14, 106, 114, 144, 149-151, 180, 239-240, 247, 250, 6530, 6595, 6907-6909, 6913-6915, 7353, 7699-7700
- swimming 7348, 7350, 7359-7363
- SWOT 580, 734, 901-903, 1408, 1522, 2549, 4724, 6567, 7264
- Synapse 206, 6594
- Synchrodipity 6492, 6497
- Synchronic 1215, 3426, 6492
- Synchronization 621, 849, 879, 891, 1095, 1180, 1227-1229, 1232-1234, 1237, 1297, 1300, 1810, 1954-1955, 2059, 3452, 3539, 3642, 3989-3992, 3994, 3996-3998, 4000, 4003, 4238, 4240, 4549, 4988, 5286-5287, 5291, 5309, 5569, 5574, 5751, 6059, 6063, 6125, 6141, 6143, 6315-6316, 6435-6436, 6438-6443, 6446, 6451, 6506, 7111, 7435-7436, 7440, 7442, 7444, 7446, 7566, 8062
- Synchronization Protocols 1227-1229, 1232-1234, 1237, 6435, 6438
- Synchronous Communications 7108, 7117, 7125
- Synchronous e-Learning 5808
- Synchronous Learning 1462, 3909
- Synchronous Shared Experiences 6442, 6446
- synergetic aggregation 7287
- Synoecism 3419, 3430, 3433
- Syntactic Information 4398
- Synthetic Benchmark 1782
- Synthetic Biology 420, 430
- synthetic data 1774, 1782, 1834
- Synthetic Synchronization 6446
- System Algebra 7704, 7706, 7708, 7711-7712
- System Design 2, 186, 320, 657, 739, 2108, 2937, 2942, 3748-3749, 3751, 3818, 4200-4201, 4337, 4957-4958, 5445, 5494, 6448, 6499-6500, 6836-6837, 6878, 7436-7437, 7444, 7785-7786, 7795
- System Dynamics 2156, 5438, 5440, 5445
- system integrity 890, 1239-1240, 1247
- System Noise 219
- System of Health Care 5852
- System of Records Notices (SORN) 3809, 3815
- System Performance Testing 7563, 7565-7567, 7569, 7572-7574, 7576
- System Performance Testing Framework 7563, 7565-7567, 7569, 7572
- System Testing 7504, 7566, 7568-7569
- systematic distractor generation 2372-2373, 2379
- Systematic Generation with Rationales Method 2372, 2379
- Systematic Review 3287, 3581, 3680, 3684, 3787, 3967-3968, 4227, 4474, 5447, 6152, 7359, 7935, 7937, 7942

Index

Systemic Information 4462, 4469
Systems Biology 419-420, 424, 430, 1680
Systems Development 317, 777, 1399, 2932, 3793,
7481, 7526, 7593, 7786-7787, 7789, 7792, 8031
Systems Theory 551-552, 555, 2155-2156, 2439, 3899,
4327-4329, 4667, 4671-4672, 4677, 5694, 7772
Systems Thinking 551-555, 557-559, 562, 3312,
3882, 7515

T

Tablets 266, 276, 382, 412, 414, 1678-1679, 1715,
1887, 2229, 2237, 2295, 2325, 2427, 2459, 2528,
2545-2546, 2745, 2841, 3078, 3398, 3463, 3517,
3525, 3693, 3810, 3830, 3832, 3960, 4066, 4210,
4263, 4310, 4536, 4623, 4662, 4837, 4839, 4843,
5104, 5106, 5120, 5210, 5886, 5908, 5983, 5985,
5987, 6058, 6071-6073, 6098-6099, 6101, 6170,
6197, 6206-6207, 6222, 6245, 6249, 6349, 6379,
6388, 6399, 6411, 6415, 6430, 6470, 7013, 7028,
7227, 7376, 7671-7672, 7674-7675, 7677, 7756,
7917, 8097
Tacit Knowledge 533, 943, 946, 2266, 2484, 3925, 4283,
4357, 4571, 4573-4575, 4580, 5029, 5046-5048,
5055, 5080, 5094, 5102, 5681, 5684-5686, 7037,
7043, 7161-7162, 7167, 7924-7925
Tactical Decision-Making Simulation 3336
Tactical Planning Horizon 5515
Tactile feedback devices 4654, 4656
Tactile Sense 4652-4653, 4665
Tailored-Applications 7269
Talk RR 5994-5997, 6005
Target Costing 2172, 2177
Target Language 3946-3947, 4492-4498, 4501, 5137,
5139, 7329, 7684-7685, 7688, 7691
Target Transparency 755, 760, 764
Target User 1735-1741, 1745, 1869-1873, 1878-1879,
2813, 3605, 6033-6035, 6037, 6042
task efficiency 2954-2955, 2957, 2959-2960, 2962-
2963
Task Selection 792, 798
tax inversion 5489
Taxpayers 831, 835, 840-841, 846, 3173
TBL 611, 5427, 6313
teacher educator 2607
Teacher Preparation 1427, 2329, 3954-3956, 3958-
3959, 3961, 7623-7626, 7630-7631, 7682, 7685-
7686, 7689, 7691
Teacher Preparation Program 3961, 7691
Teacher Presence 1454, 2661-2664, 2671, 5153, 7922,
7924, 7926-7931

Teacher Training 1420, 1423, 2229, 2298, 2572, 5016,
6391, 7174, 7624, 7635-7637, 7643, 7648
Teacher-Made Digital Stories 2454
Teaching And Learning 1065, 1130, 1420-1421,
1423-1427, 1433, 1439, 1466, 1882, 2226, 2236-
2237, 2276, 2303-2304, 2329, 2369, 2381, 2391,
2412, 2430, 2450, 2455, 2457, 2460, 2483, 2493,
2497, 2501, 2517, 2521, 2526-2527, 2533, 2545,
2552, 2558-2561, 2576-2577, 2579-2581, 2583,
2587-2588, 2604, 2641, 2652, 2662, 2664-2668,
3248-3249, 3253, 3289, 3843, 3865, 3882, 3931-
3940, 3944, 3954-3956, 3959, 3961, 3979, 4525,
5086, 5123, 5183, 5190, 5192, 5259, 5788, 5792,
5800, 5811-5812, 5821-5822, 5824, 5854, 5860,
6349, 6380, 6388, 6392, 6399, 6412-6413, 6489,
6859-6862, 6864-6865, 6870, 7225, 7392-7394,
7397-7398, 7539-7545, 7624-7625, 7629, 7631,
7635-7640, 7642-7643, 7645, 7659, 7662, 7665,
7669, 7671, 7677, 7684, 7778, 7894, 7914, 7935
Teaching Presence 2662, 2664-2665, 3905, 3976-
3977, 3985
Teamwork 557-558, 1397, 1517, 1562, 1934, 2245,
2394, 2417, 2537, 2575, 2993, 3240, 3383, 3830-
3834, 3838, 3978, 4338, 4360, 5040, 5196, 5652,
5656, 6377, 8084
Technical Dimension 2966, 2974, 3035
technical metadata 6449, 6452, 6458, 6463
technical security 1679
technique application 1322
Techno Organizational Change 2987
techno-ethical risk 4884, 4887-4888, 4896-4897
Technological Acceptance 7193, 7764
Technological Affordances 6163, 7313, 7321-7322,
7325, 7660
Technological Change 2561, 3930, 3963, 4143, 4775,
4862, 5046, 5307, 5477, 5551, 6405, 7302, 7311,
7980-7981
Technological Determinism 166, 6165, 6355, 7059,
7193, 7312, 7315, 7325
Technological Innovation 1160, 2227, 2345, 2730,
3654, 3830-3831, 4692, 4700, 4774, 4778-4780,
4784, 5982, 6249, 6296, 6306, 7015, 7224, 7235,
7304, 7322, 7754, 7757, 7764, 7810, 7982, 7987,
7989-7990
technological paradigms 1142-1143, 1152
Technological Pedagogical Content Knowledge
(TPACK) Model 3961, 5116, 6338, 6346, 6391,
7623-7624, 7659, 7662
Technological Progress 2328, 2827, 3184, 4267-4268,
4696, 5163, 7329, 7771, 7810, 7989-7990
technological trajectories 1141-1142, 1146, 1152

- Technology Acceptance Model 389, 531-532, 832, 841, 2253, 2791, 3655, 3786-3787, 3821, 4281, 6025, 6403, 6405, 7601
- Technology Adoption 699, 2283-2284, 2761, 2791, 3655, 3658, 3678, 3688, 4037, 5902, 5904, 7234, 7243-7245, 8100
- Technology Assessment (TA) 1064, 4267-4269, 4274, 4277, 7302
- Technology Enhanced Learning (TEL) 800, 808, 1423, 2380-2382, 2391, 2558-2559, 7670
- Technology Integration 1421, 1476, 2303, 2512, 3954-3956, 3958-3959, 3961, 5375, 6389-6390, 6394, 6861, 7092, 7623, 7625, 7630-7631, 7659, 7661, 7666, 7808
- Technology Management 620, 1141, 1408, 2483, 3530, 4560-4561, 4565, 4569
- Technology Transfer 805, 3830-3832, 3834, 3838, 3970, 4560, 4679-4680
- Technology Ventures 4774-4780, 4784
- technology-driven information-intensive 4884, 4895-4897, 4909, 4916, 4919
- Technology-Driven Information-Intensive Phenomenon 4884, 4909, 4919
- telecare 6150
- Telecommunication 107, 406, 412, 414, 550, 877, 1189-1190, 1656, 1715, 1835, 1837, 1985, 2007, 2019, 2341, 3617, 3633, 4216, 4234, 4237, 4622, 4679, 4681-4682, 4705, 4708, 4710, 4725, 4952, 4956-4957, 5869, 5886, 5932, 6032, 6061, 6095, 6136, 6256, 6265, 6275, 6546, 6548, 6552, 6554-6555, 6878-6879, 7035, 7225, 7227-7228, 7238, 7241-7242, 7250, 7380, 7776
- Telehealth 1984, 3692-3693, 5930, 5932, 7031, 7035, 7756
- Telemedicine 1984, 3691-3692, 3697, 3784-3785, 3788, 4709, 5021, 5930, 6147-6155, 6160, 7207, 7756-7757
- Teleoperation 6877
- Teleportation 1044-1045, 1055-1056, 1059, 1061, 7716-7717, 7720, 7730
- Telepresence 4114, 4299, 6615, 7059, 7884
- telerehabilitation 5930, 5932-5937
- Telesurgical robotics 6872-6873, 6875, 6877-6879, 6882-6883, 6885
- Telework 740-745, 750, 5029, 5035, 7059
- Temporal Database 2053-2054, 2063
- Temporal Resolution 3484, 3492, 3494, 3497, 3502
- Temporal Schema Versioning 2058, 2063
- Tension-based devices 4658-4659
- TeraGrid 1064, 1066
- Terminal and Network Heterogeneity 6096, 6105
- Territory Control 3422, 3426, 3433
- Terrorism 166, 1685, 3637-3638, 3644-3645, 3648, 3653, 4115, 4766, 5552, 7987
- Tertiary Education 1163, 2651, 2653, 3395, 3397-3398, 3839, 3909, 3911, 3921, 4786, 5841, 5853, 5860, 5862, 7239
- test design tools 7505
- test development 1030
- test management tools 7507
- Test Of Independence 1856-1857, 1862, 1868, 3197-3198
- Tethered Technologies 6347, 6360
- Text Mining (TM) 184, 350, 354-355, 410, 872, 1783, 1789, 1896-1898, 1900-1901, 1903-1904, 1906, 1909, 1930, 1969, 2817, 3543, 5170-5171, 5228, 6973, 6975-6976, 6981, 7063
- Text Presentation 8010, 8013
- textchat 1219-1220, 1225
- TFLOPS 4008, 4016-4017, 4023, 4025
- TGDH 1667-1674
- Thana-Capitalism 4761-4762, 4764-4767, 4769-4770, 4773
- Thanatourism 4077-4078, 4087
- The Frankfurt School 4716-4718, 4721
- Theory of Active Learning 5808
- Thermal Oxidation 3170
- Third Age Learners 1473
- Third-Generation Programming Language 1406
- Thor 7342-7344, 7346
- Thread Scheme 4936
- Three Laws of Robotics 105
- Three-Dimensional Virtual Learning 2465
- Threshold Cryptography 1243-1244, 1247, 1650, 4927-4928, 4930-4931, 4936
- threshold digital signature 1650-1651, 1655
- Thresholding 213-214, 1318, 1354
- tiered learning 1476-1478, 1482, 1486
- Tiki Toki 1423, 1431
- Tim Berners-Lee 1200, 2780, 4540, 4547, 7796, 8007
- Time Division Multiple Access (TDMA) 1227, 1237, 6267
- Time Series 6, 39, 123-124, 221-222, 355, 364-372, 381, 407, 409-410, 412, 854, 858, 2109, 2142, 2144-2146, 2148, 3114, 3116, 3124, 3126, 3423, 3712, 3788, 4574, 4938, 5639, 5644, 5646, 5885, 6035, 6116, 6635, 6641, 6659, 6734-6735, 7841
- Time Theft 4316, 4325
- Timelines 1422-1423, 1427, 1431, 3087, 3814, 3843, 4851, 5651, 6100, 7804
- time-place dependent 2458, 2465
- Time-Place Independent 2458, 2465

Index

- time-varying covariate 5185-5187, 5189, 5192
Tinkering 2492, 2911, 6859, 7625, 7627, 7629-7630, 7634
TK 1847, 4502, 4504-4505, 4508-4510, 4512, 7662, 7665-7666, 7670
TOC 2168, 3170
TOEIC 3853-3857, 3860
TOGAF 574, 576, 621, 625, 630, 635, 810-811, 819, 2922, 2924, 5614, 5617-5619
Tool Appropriation 3794, 3797, 3801, 3804
Top-Down Induction 2133, 2141
Top-Down Tree Search 4505, 4514
Topic Initiator 7979
Topological Map 6896-6897, 6905
Topology 169, 1052-1054, 1119-1122, 1125-1126, 1835, 1838, 1840, 1949, 2026, 2059, 4530, 4549-4555, 4559, 6110-6112, 6501-6502, 6521, 6523-6524, 6527-6528, 6535-6543, 6545, 6560, 6640, 6828, 6873-6875, 7710
Topology Lifetime 6545
TOPSIS 2934, 3131-3132, 3134-3135, 3137-3138, 3143
Total Financing (TF) 37, 47
Total Interpretive Structural Modeling (TISM) 3089, 3096
Tourism Cluster 3434-3438, 3440-3444
Tourism Destination 350, 352-353, 360, 363, 3414, 4033-4035, 4113
Tourism Icons 4099-4103, 4105, 4108, 4110
Tourism Knowledge Destination 350, 359-360, 363
Tourism Product 349, 363, 3403, 4033, 4082, 4087, 4113
Tourism Providers 4088-4094, 4098, 4115
Tourism Services 3441, 4088-4091, 4094, 4098
Tourist Destination Image Formation 4031, 4034, 4037
Tourist Feedback 363
Tourist Perception 4037
Toyota Production System 729, 5517
Trace-Back Unit (TBU) 6313, 6318
Trade Liberalization 7981-7982, 7994
Traditional Ecological Knowledge (TEK) 3485, 3489, 3502
Traditional Study 2465
Training Set 114, 125, 160, 183, 197, 407, 1762, 1766, 1827, 1829, 1899, 1902, 1906, 1923-1924, 1926, 1973, 1983, 2219, 4530, 4941, 5172-5173, 5182, 6647, 6657, 6670
Trajection/Trajectivity 4305
Trajectory Metaheuristics 7694-7695, 7703
Transaction Processing and Concurrency Control 1953
Transactional Distance Learning 5808
transactional leadership 5195, 5197, 5203
Transcendence Into the Virtual World 4198
Transcriptomics 422-423, 430
Transdisciplinary 1468, 1473, 4476, 4672, 5854, 6089, 7005
Transfer Function 112, 122, 497, 4746-4747, 4749-4754, 4760, 6007-6008, 6018, 6043-6044, 6052, 6056, 6171, 6174-6175, 6177-6179, 6657
transfer pricing 5504, 5506, 5515
Transform And Load 849, 851, 860, 5612
Transformation Society 6495
Transformational Digital Literacy 2284, 2291
Transformational Government 3633
Transformational Leadership 3634, 5196-5197, 5203, 5627, 5726-5728, 5730, 5733, 5735
Transformational Learning 7636, 7645
Transformative Learning 1509, 1529, 1534-1536, 1538, 1883, 5808
Transition Description Cards 7795
transition region 1308, 1311-1316, 1318
translation procedures 5136-5147
Transliteracy 2243, 2294, 6488-6490, 6492-6495, 6497
Transmedia 4262, 4669, 4677, 6488-6490, 6492-6495, 6497
Transparency 63, 675, 754-762, 764-765, 879-880, 945, 975-976, 1106, 1212, 1335, 1565, 1629, 1634, 1637, 1901, 2085-2086, 2091, 2236, 2284, 2459, 2878, 2900-2901, 2920, 2922, 3016, 3025, 3032, 3034-3036, 3041, 3149, 3191, 3550-3551, 3579, 3581, 3592, 3594-3595, 3601, 3603, 3606, 3628, 3655, 3662, 3814, 3843, 4190, 4255, 4257, 4263, 4343, 4383, 4436-4437, 4439, 4441, 4444, 4453, 4685, 4858, 4909, 5306, 5308, 5310, 5313, 5368-5369, 5424, 5612, 5737, 5741, 5747, 5803, 5920-5921, 6591, 6716, 6781, 6911, 7005, 7037, 7194, 7220, 7766, 7960, 7969, 7985, 7987
Transparent Organizations 764
Transportation risk 5552-5553, 5556, 5558
Trapezoidal Fuzzy Number (TpFN) 3225
Travel Industry 4052-4053, 4055-4056, 4058-4059, 4063, 4115
Traxler 6347-6351, 6356, 6377, 6379, 6382, 6399-6400, 6405
Tree Encoding 4504, 7995, 7997
Tree Induction 366, 413, 2132-2134, 2137
Tree Matching 4514
tree rings 3114, 3118-3119, 3127, 3130
trellis decoding 6307-6308, 6310, 6313
trellis diagram 6308-6311, 6313-6314, 6318
Trend Analysis 5682, 5689
Triangular Fuzzy Number (TFN) 3135, 3225

Triangulation 1283, 1889, 2592, 2618, 3672, 3956,
 6189, 6756-6757, 6765-6766, 7316-7317
 Tribology 431, 439
 Tries 78, 81, 98, 137, 146, 155, 185, 243, 382, 1011,
 1190, 1195, 1551, 1639, 1815, 2086, 2094, 2479,
 2825, 2828, 2882, 2885, 3366, 3475, 3760, 3929,
 3994, 4073, 4203, 4257, 4505, 4541-4542, 4551-
 4552, 4838, 4842, 4908, 5223, 5374, 5489, 5554,
 5609, 5662-5663, 6144, 7194, 7392, 7394, 7503,
 7972, 7995, 7997
 Trimline 300-301, 304-305, 313
 Triple Bottom Line 610, 3177, 3186, 5423, 5425,
 5428-5429, 5432, 5466, 5738
 Triple Constraint 5659
 Triple Revolution 7064, 7068
 Trojan 4839, 4848, 6210, 6215, 7721, 7723
 Trolling 1709, 4247, 7800, 7804
 Trophy Generation 4367
 troubled economy 2342-2343
 TRR 4475-4476, 4479
 Trust Boundary 1029, 1032
 Trust Rating 4797, 4799-4805, 4807, 4809
 Trusted Authority 4975, 4978-4979, 4983-4984
 Trusting Belief (TB) 3658-3659, 3661-3662, 3666
 Trysumers 1644
 Turing Education Day 258-260
 Turing Machine 224, 7731, 7734, 7743
 Turing Test 1, 98-99, 106, 229, 238, 251, 254, 258-260,
 263-264, 4744, 8066, 8078
 Turing100 255, 258-260
 Tweet 394, 402, 1513, 2065-2069, 2071, 2074, 2855,
 3613, 5750, 6976, 7149, 7183, 7185
 Twenty-First Century Skills 2302, 2502, 2641, 2647,
 2650
 TWI 3503-3504, 3506, 3508-3509, 3514
 Twig Matching 4514
 Twitter 269, 333, 373, 407, 412, 765, 1104, 1143,
 1533, 1535, 1596, 1698, 1706-1707, 2064-2071,
 2074, 2237, 2326, 2423, 2524, 2563, 2570, 2794,
 2814, 2816, 2855, 2910, 3369, 3398, 3477, 3548,
 3572, 3612-3613, 3615-3616, 3620, 3648, 3667,
 3747, 4052, 4054-4056, 4063, 4088, 4090-4091,
 4212, 4214, 4318, 4329, 4830, 5226, 5252, 5255-
 5256, 5267, 5713, 5748, 5775-5780, 5782, 5785,
 5840, 5870, 5909, 5913, 6038, 6088, 6152-6153,
 6185-6186, 6276-6278, 6280, 6302-6303, 6491-
 6492, 6908, 6911, 6941, 6951, 6961, 6973-6976,
 6981, 6985, 6994-6999, 7001, 7006, 7016-7018,
 7020, 7025-7027, 7036, 7056, 7058, 7063, 7081,
 7088, 7096, 7106, 7112, 7119, 7121, 7127, 7130,
 7132-7134, 7142-7143, 7145, 7149, 7169, 7172,

7182-7184, 7192, 7531, 7541, 7626, 7628, 7746,
 7797-7798, 7890, 7947, 7951-7952, 7984, 8016,
 8037, 8050, 8056, 8091
 Twitter Apis 2066-2067, 2074, 7530
 Two-Phase Commit (2PC) Protocol 1093, 1095-1096,
 1100, 8057-8058
 type of school 1540, 1543-1545, 1547
 Types of e-Culture Expression 4191, 4198
 Types of UIC 3975
 Typology 609, 611-613, 618, 713-714, 1369, 1657,
 3340, 3423-3424, 3474, 3875-3876, 4077, 4080-
 4081, 5080, 5084-5085, 6164

U

U.S. Department of Health and Human Services 3809
 UAV 6106, 6108-6109, 6113
 Uber 683, 1160, 2345, 2347-2351, 2355, 5226
 UbiHealth 7765-7772, 7774
 Ubiquitous Computing (UbiComp) 386-387, 1994,
 3541, 4951, 4961, 5110, 5440, 6779, 7745-7747,
 7749-7751, 7753-7756, 7764-7768, 7770-7771,
 7774, 7776, 7785, 7797, 7801, 7805, 7807-7810,
 7813
 Ubiquitous Health 7766-7767
 Ubiquitous Learning 1517, 1519, 1523, 2545-2546,
 2556-2557, 7672-7673, 7677, 7680-7681, 7907
 ubiquitous world 2777
 UI 40, 240, 386, 390-391, 1298, 1303, 3518, 4598-
 4599, 5585, 5887, 5889, 5900, 6036, 6064, 6066,
 7402, 8006
 ultra-performance 7731, 7733
 Unbanked 2725, 2730, 2738, 2878, 8100
 Uncertain Data 338-339, 342-343, 1837-1838, 2200-
 2201
 Uncertainty Avoidance (UA) 4306, 4308, 7957-7958,
 7966, 8099, 8104
 Unconventional Monetary Policy 6919, 6926
 under sampling 1827-1829
 under-banked 2725, 2738
 Under-Reporting 3194-3195, 3199-3200, 4170
 Underrepresentation 3382, 3386
 Understaffing (Underage) 3715-3716, 3718-3719,
 3723-3724, 4140, 5106
 Unified Modeling Language (UML) 3340, 7436, 7449,
 7475, 7481-7482, 7484-7485, 7487
 Unimodal 1550-1552, 1554-1556, 1558, 1749
 Unimodal Student 1558
 Unintegration of Thoughts 7955
 Unique Clicks 3370
 Unique CTR 3370

Index

- Unit of culture 4666, 4668, 4671-4672, 4677
- United States Federal Aviation Administration (FAA) Advisory Circulars (AC) 670
- United States Federal Aviation Administration Regulations (FAR) 670
- United States Navy 1500, 1507
- Univariate Decision Tree 2134, 2141
- Universal Design 318, 528, 2630, 4360, 7030
- Universal Mobile Telecommunication System (UMTS) 6265
- University Education 2421, 3870
- University Microfilms 6748
- University-Industry Collaboration (UIC) 3963, 3975
- University-Industry Relations 7151, 7153-7157, 7160
- unmodelled issues 2178, 2181, 2183, 2187
- Unordered Tree Inclusion 4512, 4514
- unquantified issues 2181, 2187
- unreliable server 1570-1571, 1583
- Unstructured Data 351, 353-354, 373, 377-378, 380, 692, 851, 862, 864, 869, 871-872, 951-952, 960, 1911, 2822, 5749, 5751, 7146, 7149
- Unstructured Data Source 7149
- Unsupervised Learning 122-124, 131, 148, 350-351, 406, 411, 1354, 1809, 1928, 1933, 1963, 1973, 4532, 4575, 6146, 6352, 6657
- Unsupervised Neural Network 442, 448-449
- upper approximation 1838, 2202, 2204, 2212
- Urban Area 6261, 6264, 6981, 7815, 7817, 7831-7833, 7835, 7838-7840, 7842
- Urban Governance 7855
- Urban Planning 373, 573, 985, 3419, 4573, 7843-7849, 7851, 7855
- Urban Tissue 976, 7863, 7865
- Urban Tourism 4064-4066, 4069, 4072-4073, 4076
- Usability Evaluation 4099, 4101-4103, 4105, 4110
- Usability Testing 4102, 4105-4106, 4108, 4112
- usable security 5004-5006, 5011, 5013
- Use Case 1066, 2870, 3338, 5436, 6287, 7402, 7453, 7476, 7484, 7577-7579, 7587
- User Acceptance 360, 890, 3786, 3793, 3821
- User Experience 345, 684, 878, 997, 1334, 2751, 2855, 3519, 4034, 4125, 4130, 4213-4214, 4217-4218, 4222, 4542, 5752, 6070-6071, 6230, 6301, 6481, 6548, 6658, 6752, 6950, 7045, 7363, 7402, 7422, 7506, 7599-7600, 7602-7605, 7608, 7800, 8038, 8045, 8049, 8051, 8082-8084, 8087, 8092
- User Generated Content (UGC) 350, 464, 776, 2292-2293, 2295, 2570, 2641, 2908, 2919, 3571, 3578, 3828, 4052-4053, 4059, 4089, 4098, 4279, 5776, 6355, 6360, 6478, 6974, 7003, 7006-7008, 7013, 7017, 7025, 7036, 7043, 7126-7131, 7133-7134, 7139, 7161, 7167, 7169-7170, 7270-7271, 7273-7275, 7277, 7280, 8023, 8037, 8050, 8055
- User Incentives 8026, 8035
- User Interface 105, 360, 363, 385-386, 389, 391, 421, 460, 544, 554, 854, 985-986, 1074, 1250, 1256-1257, 1261, 1264-1265, 1274, 1334-1336, 1340, 1622, 2380, 2438, 2745, 2857, 3260-3261, 3477, 3516, 3565, 3681, 3748-3749, 4125, 4134, 4200, 4202, 4656, 5005, 5009, 5106, 5876, 5887, 5899-5900, 5942, 5954, 6061, 6064, 6066, 6069, 6206, 6414, 6627, 6659, 7045, 7402, 7563, 7570, 7579, 7599-7600, 7608, 7956-7957, 8006-8007, 8013, 8087
- User Model 1442-1443, 2384-2385, 6424-6426, 6430-6431
- User Payment Agent 2760
- User resistance to IT 3816, 3822
- User Satisfaction 50, 52-54, 56-57, 383, 833-834, 914, 2468, 2750, 2938, 2978, 4200, 4885, 5006-5008, 5727, 6609, 8066, 8078
- User Story 7401, 7410, 7577-7579, 7587
- User-Centered Design 317, 4112, 4125, 4134, 7402
- User-Experience Design 4222, 6164, 6170, 7600
- user-to-user collaborative filtering 1745, 1872-1873, 1878-1879
- UTAUT 2253, 2761-2764, 3579, 3581, 3585, 3587, 3655-3656, 3659, 3661-3662, 3666, 3787, 7193
- UTAUT Model 2762-2763, 3581, 3585, 3587, 3655
- Utilitarianism 1641, 3176, 4886, 4894, 4896, 4921, 4926
- Utility Computing 1077, 1101, 1103-1104, 1147, 6547
- Utility Function 71, 2145, 2149, 2188-2189, 2191-2194, 2197, 3228, 6550
- Utilization 103, 123-124, 139, 203, 208, 223, 457, 546, 614, 729, 849, 925, 956, 999, 1002, 1077, 1145, 1164, 1166, 1175, 1181, 1196, 1344, 1369-1370, 1455, 1565, 1595-1597, 1810, 1949-1950, 1952-1953, 1985, 2253-2255, 2532, 2567, 2817, 2841, 2849, 2969, 3144, 3217, 3275, 3385, 3461, 3517, 3593-3594, 3633, 3655, 3736, 3936, 3954-3955, 3961, 4024, 4202-4203, 4282, 4328, 4348, 4382, 4436, 4549-4550, 4552-4554, 4776, 4799, 4862, 4871, 4913, 4919, 5030, 5079, 5082, 5108, 5310, 5328, 5362, 5470, 5490, 5500, 5503, 5511, 5516, 5552, 5559, 5562-5565, 5598, 5802, 5904, 5908, 5911, 6094-6095, 6266-6269, 6412, 6443, 6479, 6537, 6549-6550, 6552, 6610, 6613, 6713-6715, 6847, 6887, 6933, 7150, 7202-7203, 7205-7206, 7209, 7485, 7575, 7630, 7722, 7778-7779, 7871, 7902-7903, 8048, 8052, 8090, 8097
- Utopia 4299-4301, 6281

V

- Vacuum Molding 304-305, 313
- Valid Inequality 5421
- Value Based Decision 2191, 2199
- Value Capture 1143, 1152
- Value Creation 61, 350, 551-552, 554, 562, 564, 568, 612, 676, 708-709, 729, 735, 879, 954, 1143, 1146, 1149, 1152, 1159-1160, 1585, 2350, 3013, 3881, 4282, 4330, 5056, 5292-5293, 5297, 5378, 5386-5387, 5671, 5675-5676, 5716, 5719, 5777, 8031
- Value Creation System 551, 562
- Value Stream 729, 734-735, 739
- Value Systems 88-90, 97, 3684, 5037
- Value-Chain 553-554, 582, 703, 707, 709, 739, 821, 824, 966, 968, 1157, 1596, 1686, 2953-2955, 2975, 2977-2978, 2987, 3146, 3183, 3185-3187, 3190-3191, 3193, 3594, 4282, 4284, 4560-4561, 4680, 4826, 5291, 5297, 5388, 5457, 5533, 5539, 5546, 5574, 5595-5596
- Value-Driven Modelling 2199
- Vannevar Bush 4537, 4539, 7796
- variance reduction 1300, 1306
- Variation Point 7427, 7429-7431
- Variation Principle of Library 4400-4403
- VARK model 1549-1550, 1558
- vector of extrusion 1289-1290, 1296
- Vendor-Managed Inventory Systems (VMIS) 5929
- VERB NOMINALIZATIONS 5129
- Verbal Metaphor 5092
- Verdoorn's Law 7247
- Vernacular Universals 1206, 1215
- vertical e-commerce 2741, 2748
- Vertical Integration 1164-1165, 1167-1170, 1173, 2806
- Vertical Search 8055
- Very Small Entities 6927-6929, 6938
- Victimisation 4245-4249, 4254
- Video File Format 2660
- video inquiry learning 2642-2643, 2647, 2650
- Video Recording Strategies 7688
- Videogames 801-802, 974, 1510-1511, 1516, 2255, 2396, 2430, 3240, 3244, 3260-3262, 3265, 3271, 3274, 3277, 3280, 3296-3297, 3299, 3301, 3327-3328, 3355, 4115, 4234, 4297-4298, 4317, 4325, 4362, 4652, 4658, 4676, 5089, 5105, 5576, 5878, 6060, 6069, 6361-6363, 6375, 6379, 6389, 6489, 7035, 7358-7359, 7363, 7367
- Video-Sharing Websites 5785
- View Change 1239-1240, 1649, 3059, 3062, 3065, 4929, 8060
- Viewshed 3422-3423, 3426, 3428-3430, 3433, 3446
- viral marketing 6019-6024, 6030
- Virality 7109-7112, 7115, 7800, 7804
- Virgin Consumers 1644
- Virtual Asset Trackers (VAT) 5919, 5929
- Virtual Body 3354, 3358, 4237
- Virtual Brand Community (VBC) 708, 719
- Virtual Classroom 1158, 2304, 2533, 2535-2536, 2538-2539, 2544, 2664, 2671, 5097, 6415
- Virtual Classroom Usage 2544
- virtual collaboration 2594
- Virtual Community (VC) 707, 715, 1676, 4071, 4088, 4091, 4094, 4098, 4122, 4146, 4153, 4190, 4196, 4278, 4280, 4282-4284, 4286, 4295, 5029, 5097, 5102, 5251, 5253, 6285, 6300, 6303, 6705, 7067, 7116-7118, 7121, 7125-7126, 7162, 7167, 7201, 7211, 7789, 7916
- Virtual Design 540
- Virtual Distance 750, 5988-5989, 5991
- Virtual Education 2421-2422, 2424, 2426, 2430, 5087
- Virtual Enterprise (VE) 2778, 2783
- Virtual Environments 1068, 1365, 2412, 2548, 3001, 3262, 3356, 3366, 4114, 4146-4147, 4150, 4153, 4199, 4234, 4296, 4653, 5104, 5106, 5881, 5956, 6233, 6942, 7360, 7673, 7907, 7917, 7936
- Virtual Goods 4306-4310, 4314, 7130
- Virtual Learning 2275-2276, 2303, 2414, 2423-2424, 2427, 2465, 2521, 2524, 2531, 2560, 2592, 2661-2662, 2664, 2667-2668, 3829, 3937, 5161, 5836, 6339, 6343, 6348, 6404, 7902, 7912-7917, 7921, 7935-7936, 7944
- Virtual Learning Environments (VLE) 2275, 2414, 2521, 2524, 2531, 2560, 2592, 2661-2662, 2664, 2667-2668, 3829, 3937, 6339, 6348, 6404, 7902, 7912-7917, 7921, 7935-7936, 7944
- Virtual Library 1019, 5203
- Virtual Machine 89, 411, 1027, 1032, 1078-1080, 1084, 1088, 1090, 1119, 1166, 1175-1181, 1238, 1357-1358, 6061-6062
- virtual machine migration 1078, 1084, 1090, 1181, 1357
- Virtual Manufacturing 505, 507, 517, 520
- Virtual Objectivization 4193, 4198
- Virtual Organization 1033, 1039, 2981
- Virtual Personal Agent (VPA) 5756
- Virtual Reality (VR) 391, 505-506, 543, 800, 1278, 1696, 2411-2413, 2416, 2418, 2420, 2459, 2465, 2581, 3261, 3298, 3333, 3352-3355, 3358, 3418, 4037, 4113, 4122, 4152, 4202, 4234-4240, 4244, 4255, 4308, 4325, 4920-4924, 4926, 5104, 5225, 5268, 5709-5711, 5713, 5830, 5835, 5853, 5858, 5860, 5878, 5885, 5888-5889, 5893-5894, 5897, 5900, 5955-5958, 5961, 5965, 6223, 6225, 6233,

Index

6338, 6756, 6759-6761, 6878, 7005, 7122, 7125,
7358-7359, 7806, 7880-7881, 7887
virtual reality analgesia 5956, 5958, 5965
Virtual Reality Film 3352, 3358
Virtual Socialization 7116, 7121-7122, 7125
Virtual sport 7122
Virtual Teams 720-726, 728, 777, 783, 787, 1935
Virtual Tourism 3407, 4113, 4115, 4122
Virtual Voluntarism 7077-7078, 7080-7082, 7087
Virtual Worlds (VW) 2412, 2458, 2489, 2521, 2524-
2525, 3350, 3355, 4114, 4146-4147, 4149-4153,
4156, 4296, 4298-4302, 4305, 4920, 4926, 4951,
5896, 6225, 6339, 7008, 7128, 7602, 7880-7884,
7907, 7935-7944, 8042
Virtual-Time Rendering (VTR) Algorithm 6446
Virtue Ethics 90, 4921, 4926
Visible Light Communication (VLC) 6672, 6676, 6683
Visual Cypher 2035-2036, 2041-2042
Visual Identity 8079-8080, 8082, 8084, 8086
Visual Identity Design 8079
Visual Literacy 1423, 2244, 2294, 2325, 2594, 5103,
5106-5108, 5110, 7857
Visual Modeling Language 4821
Visual Programming Environments 2397, 3309, 3312,
3316
Visual Semantic Algebra 7704, 7709, 7711
visual transformation 2608, 2610-2611, 2615
Visual Understanding Environment 7883, 7887
Viterbi algorithm 6307-6309, 6315-6316, 6318
Viterbi Decoder (VD) 6307, 6309-6310, 6314-6316,
6318
Viv 5753
Viva Voce Turing Test 264
VLC Applications 6673-6674
VOC 729, 3156, 3158-3160, 3163, 3169-3170
Vodcast 2651-2652, 2660, 5265-5266
vodcasting 2651-2654, 2657-2659
Voice Output Communication Aid 289, 6475
Voice Over IP (Voip) 6271, 6617, 8095
Voluntarism 6705, 7077-7078, 7080-7082, 7087
Voluntary Disclosure 1897, 1901, 1904, 5675, 5739,
5747
voluntary reporting 5737-5739, 5742-5744
Volunteered Geographic Information 6974
Von Neumann Entropy 4391, 4398
Voronoi Regions 1962, 1973
Voter Turnout 3613-3617
VR analgesia 5957-5960
VRML 2412, 2414-2415, 2420
VSE 6928-6929, 6933
Vygotsky 1475, 1520, 1540, 1544, 2328, 2534, 2642-

2643, 3250, 5094, 5117, 5149, 5156, 5803, 7627,
7673, 7881, 7923, 7939, 7941

W

W3C HTML Validator 8008-8009, 8014
WAI 3517-3519, 3522, 3565, 6077
Wallet 2715, 2753-2754, 2760, 2785-2786, 2788, 2870,
2874, 2876, 6207, 8096
Warehouse Management 1377, 1379-1381, 1383, 2804,
2806, 5490-5493, 5502
Warping 4383, 4389, 5879, 5885
Warsaw 3910-3911, 3916, 4044, 4065-4067, 4070-
4072, 4076, 7302
WASC 3182
Wasserman 2214, 6950, 7015, 7150-7151
Waste Management 614, 3067, 5274, 5359, 5378,
5383-5384
Waterfall Project Methodology 692
watermark 4872-4875, 4877-4881, 4883, 4986
Watermarking 2911, 4872-4875, 4877-4878, 4880,
4883, 4985-4986, 5003
Watson 51, 101-102, 145, 207, 230, 234-235, 238,
397, 734, 933, 1022, 2739, 2883, 4126, 4310,
4829-4830, 4839, 5310, 5319, 5750-5751, 5756,
6073, 7088, 7172, 7760, 8099
Wavelet Transform 496, 4645, 4875
WCAG 321, 3518, 3520-3522, 6073
WCAG 2.0 321, 3520-3522
Weak Negation 4738, 4745
weapons research 3205-3206, 3208-3210
Wearable Computers 4127-4128, 4135
Wearable Devices 133, 1985, 3697, 4201, 5052, 5886,
5897, 6115, 6154
Wearable Technology 4322, 5110, 5291, 5935, 5940,
6123, 7784
Web 2.0 Tools 2236-2237, 2242, 2604, 2659, 3243,
3571, 5254, 5262, 5264, 5763, 5910, 7848, 7851
Web Accessibility 3516-3518, 3520, 3525, 6077
Web Crawler 2074, 8047
Web GIS 3403-3404, 3407-3415, 3417, 3463-3464
Web Interface Design 7956-7959, 7962
Web Mining 184, 410, 872, 1036, 1587, 1596, 1909,
1930, 2794
Web Payments 2750-2751, 2755, 5397
Web Portal 460, 2772, 3475-3477, 3605, 4099, 4101,
5104, 5541, 5544, 5547, 5549, 6188, 6690, 7117
Web Services 387, 421, 455, 460, 653, 1027, 1153,
1174, 1562, 2012, 2021, 2754, 2773, 2777-2778,
3058, 3275, 3282, 3286, 3410, 3462, 3604, 3657,
4013, 5068, 5161, 5263-5264, 5271, 5390, 5543-

- 5544, 5547, 6037, 6088, 6477, 6547, 6571, 6974,
7170, 7396, 7532, 7567, 7588, 7590, 7775, 8057-
8062, 8065, 8074
- Web Technologies 1200, 1441, 2524, 2531, 2773, 2775,
2777, 3038, 3407, 5544, 5613, 6752, 7003, 7088,
7169, 7277, 7796, 7854, 8023, 8079
- web traffic 1885-1886, 1888
- web-based courseware 3855
- Web-Based GIS 3472
- Web-Based Instruction 1455, 1462, 5812, 5822
- Web-Based Learning 1451-1452, 1491, 1517, 3828,
5811, 5814, 5820-5821, 6415-6416, 7174, 7921
- Web-based Project 7550, 7553-7558
- Web-based surveys 7848-7849
- Web-Documentary 6487
- Website Content 8008-8009, 8013-8014, 8046
- Website Errors 8014
- weighted cardinal fuzzy measure 7284-7285
- Weighted Connections 120, 125
- Weighting 219, 497, 587, 761, 2094, 2219, 2913, 3503,
3509, 3511, 3514, 4797, 4800, 5686, 5790, 5794,
6661, 7286
- Weighting Factors 3509, 3511, 3514
- Wetting 434, 437, 3049-3052, 3055
- White-Box Testing Technique 7459
- Wholesale Price 1164-1165, 1168-1170, 1173, 3147
- Wicked Problem 6497
- wide pass band 6043, 6051
- Wi-Fi 1227-1228, 1231-1237, 1656, 1994, 2506, 3414,
3553, 3689, 4842, 4848, 5021, 5120, 5575, 5921,
5980, 5984, 5988, 5991, 6032, 6037, 6042, 6058,
6096-6097, 6101, 6138, 6258, 6287-6290, 6471,
6621, 6676, 7810
- Wikileaks 1217, 3644, 7296-7297, 7301, 7987, 8019
- Wikis 461, 464, 1104, 1492, 1706, 2254, 2481, 2523-
2524, 2548, 3239, 3404, 3409, 3794, 3934, 4071,
4088-4089, 4279, 4282, 4295, 5252, 5256, 5258,
5262-5265, 5267, 5776, 5794, 5909, 5913, 6252,
6413, 6974, 7036, 7043, 7077-7078, 7082, 7111,
7130-7131, 7628, 7797, 8031, 8035, 8037-8039,
8043, 8045
- Wikisumers 1645
- Willingness To Pay 713, 715, 2834-2838, 2840, 6546,
7129
- Willingness To Share 3737, 5095
- Wilson Model 4516
- Wimax 5020, 6096, 6105, 6138, 6254, 6256-6261,
6263-6264, 6268, 6603, 6608
- Windows Phone 6069, 6202
- Wi-Red 1232, 1237
- wireless access networks 6253-6254
- Wireless Body Area Network (WBAN) 6319, 6334
- Wireless Communications 413, 4937, 5967, 5970,
5976, 6094-6096, 6105, 6595, 6598, 6603, 6608,
7779, 7801
- wireless device 5995, 6004, 6136-6137, 6141, 6143-
6144, 6146
- Wireless Local Area Network (WLAN) 1994, 5575,
5577, 5980, 6608, 6676
- Wireless Networks 1228, 1231, 1234, 1237, 3417,
3553, 3689, 4202, 5018, 5575, 5976, 6095-6097,
6102, 6105, 6255, 6260, 6268, 6270, 6272, 7252,
7745, 7750, 7771
- Wireless Sensor Network (WSN) 1227-1228, 1994,
3697, 4203, 4951, 4953, 6522, 6524-6525, 6533,
7525-7526, 7776
- Wireless Sensors 1278, 6533, 6535
- Wireless Technology 1228, 1235, 1657, 2506, 3571,
5019-5020, 5022, 6097, 6274, 6676
- Wisdom 565, 572, 943-944, 1195, 2006, 2265-2266,
2286, 2592, 3242, 3340, 3570, 4328, 4715, 5005,
5078, 5700, 5709, 5711, 5775, 6164, 6246, 6336,
6355, 6705, 6918, 7005, 7111, 7342, 7345, 7799,
7844, 8023
- Witchcraft 3642, 4077, 4082-4084
- Within-Methods Triangulation 6757, 6766
- Women Empowerment 3376-3377, 3381
- word formation processes 5136, 5138-5139, 5145-5147
- Word-Of-Mouth (WOM) 712, 715, 2853, 2860, 4032,
4058, 4091, 4284, 5775, 5777, 5780, 5982, 6019,
6030, 8019
- Work Breakdown Structure (WBS) 1939, 1946, 3630
- Work Load 1086, 1090, 3786
- Work Orders (WO) 3698, 3702, 3705-3706, 3710
- Work Performance 4340, 4343-4344, 4346, 7040, 7162
- Work Station 6835
- Workaround 706, 2163, 2165
- Work-Family Conflict 3382, 3385
- workflow management 888-889, 891-893, 895, 898
- workflow management system 891-893, 898
- Workflows 420-421, 889, 895-896, 1069, 2773, 3471,
4150, 5051, 5277, 5921, 7767, 7769, 7772
- Workforce Management 2348, 3527, 3711
- Working Experience 3026, 7666
- Work-Life Balance 4361-4363, 4367
- workload model 1772, 1775, 1782
- Workpackage 1939, 1946
- Workplace Incivility 3384, 3392
- Workspace 744, 2570, 3383, 3750, 4656, 6875-6876,
6881, 6884-6885, 6889, 6893

Index

Workspace Technology 2570
World Bank (WB) 278, 1260, 2284-2285, 2521, 2725, 2989, 4117, 4436, 4442, 5077-5078, 5082, 5932, 5940, 6217, 6276, 7216
World Health Organization (WHO) 278, 307, 317, 398, 2521, 3448, 3782, 3793, 4233, 5931-5932, 5940, 6115, 6124, 6218, 6776, 7756
World Trade Organization (WTO) 4685, 4691, 6277
World Wide Web 350, 887, 1039, 1721, 1875, 1985, 2012, 2303, 2305, 2335, 2457, 2825, 3286, 3404, 3407, 3516, 3521, 3571, 4535-4536, 4539-4541, 4547, 4619, 4628, 4681, 4953, 5076, 5097, 5205, 5212, 5262-5263, 6252, 6556, 6752, 7001, 7107, 7115, 7549, 7746, 7796, 7984, 8005-8008, 8047, 8050, 8052, 8056, 8088
Worldview 3484-3485, 3495, 5039, 6759, 8016
Worldwide Interoperability for Microwave Access (WiMAX) standard 6105, 6138
Worms 2490, 3684, 4838, 4843, 4848, 5991, 6215
Worst Case Aliasing 6179, 6183
Wrongdoing 754, 3194-3196, 3199-3201, 3203

X

XML-RPC 3568
XSEDE 1065-1070

Y

Yahoo! Answers 7016, 7019-7020, 7025
Ydria Motors LL Company 5638

Young People 800, 806, 1368, 1421, 1510, 1559, 1562, 1564-1566, 1696, 1699, 1703-1704, 1708, 1710, 1733, 2246-2247, 2256, 2315-2321, 2327, 2396, 2512, 2514-2517, 2520, 2522, 2524, 2989, 2991-2994, 3215, 3317, 3667-3673, 3676, 3736, 4157, 4159, 4162-4163, 4168-4173, 4179, 4245-4246, 4248-4249, 4308, 4923, 4962-4966, 4971, 5016, 5627, 6127, 6277, 6300, 6940-6941, 7090, 7140, 7142, 7253
Youtube 388, 976, 1104, 1167, 1510, 1642, 1698, 1883, 1888, 1891-1892, 2255, 2295, 2326, 2422, 2458, 2486, 2643-2644, 2647, 2817, 2856, 2911, 3555, 3572, 3612-3615, 3667, 3794, 3828, 3851, 4054, 4065, 4071, 4073, 4088, 4162, 4318, 5252, 5257, 5267-5268, 5775-5777, 5780, 5782, 5785, 5983-5984, 6038, 6152, 6355, 6404, 6478, 6615, 6941, 6993, 6996-6998, 7001, 7006, 7088-7089, 7096, 7109, 7112, 7121, 7128-7130, 7143, 7149, 7182, 7343, 7396, 7541, 7544, 7628, 7630, 7642, 7746, 7748-7750, 7798, 8016-8019, 8037-8038

Z

Zero Latency Enterprise (ZLE) 2774, 2783
Zero-Day Attack 4848
Zero-Sum Game 3, 155
Zombie 207-208, 7358
Zone Of Proximal Development 2328, 2534, 2643, 3250, 5808, 7941
ZYNQ 4021